



Ecole doctorale  
« des Génomes Aux Organismes »



THE UNIVERSITY OF  
WESTERN AUSTRALIA  
*Achieving International Excellence*

School of Chemistry  
and Biochemistry

## THESE EN COTUTELLE

Présentée pour l'obtention du grade de  
DOCTEUR DE L'UNIVERSITE D'EVRY VAL D'ESSONNE

Presented for the degree of  
DOCTOR OF PHILOSOPHY  
AT THE UNIVERSITY OF WESTERN AUSTRALIA

Caractérisation moléculaire et fonctionnelle de la protéine  
DYW1 dans le complexe d'édition chloroplastique d'*Arabidopsis*  
*thaliana*

Molecular and functional characterization of the DYW1 protein  
in the chloroplast editing complex of *Arabidopsis thaliana*

Par :  
**Clément BOUSSARDON**

Jury composé de:

**Claire LURIN**

**Ian SMALL**

**Francis-André WOLLMAN**

**Christopher HERBERT**

**Renée SCHROEDER**

**Bénédicte STURBOIS**

Co-directeur de thèse

Co-directeur de thèse

Rapporteur UEVE et UWA

Rapporteur UEVE et UWA

Rapporteur UWA, Examinatrice UEVE

Examinatrice

## **Remerciements** :

Ma thèse a, notamment, été un travail d'équipe. J'aurai voulu citer toutes les personnes qui y ont participé (de près ou de loin) mais ceci paraît un défi beaucoup trop ardu. Néanmoins, je me dois de remercier un maximum de personnes tant leur contribution a été importante et bénéfique pour le bon déroulement de ma thèse.

Tout d'abord, un grand merci à mes mentors Claire et Ian sans qui rien de tout cela n'aurait pu arriver. Ma carrière aurait très bien pu prendre un autre chemin. Ca a été un honneur d'avoir pu travailler avec vous.

## **URGV** :

Mes anges gardiens à Evry (juste pour citer : brouh ! brouh !), Laure à qui je souhaite tout le bonheur du monde dans son nouvel environnement et Alex sans qui rien n'aurait été possible pendant ma thèse et qui a été mes mains à Evry lorsque j'étais en Australie. J'ai une dette envers vous deux !

Andéol et Dario, mes collègues de café.

Aymeric pour les discussions sur nos sujets de thèse... ou pas.

Merci à Eddy, enfin quelqu'un qui parle foot à l'URGV (ma santé mentale en dépendait) !

Les membres de l'équipe FGA qui m'ont soutenu et supporté pendant mes années de thèse : Richard, Mauricio, Etienne, Sansan, Lulu, Delphine, les Steph et José.

Tous les gens qui ont participé au volley du mercredi et au foot du mardi, pour mes nerfs, mention spéciale à Mathieu « ours » Charles qui est revenu d'Orléans pour jouer.

Mes ex-camarades de GPV dont les réunions me manquent.

Bénédicte pour m'avoir donné l'opportunité de travailler à la fac.

Héribert pour m'avoir laissé évoluer dans son unité.

Tous les membres de l'URGV qui m'ont aidé à m'intégrer et à mettre en place mes expériences que ce soit l'équipe de la serre l'équipe administrative et les membres des équipes de l'URGV.

Patricia et Anne-Françoise pour m'avoir offert mes premiers contrats après l'échec de ma première inscription en thèse à l'UEVE.

Merci à Jean Colcombet, Philippe Giegé et Dominique Rumeau (mon comité de thèse) pour m'avoir prodigué leurs précieux conseils.

### UWA-ARC CoE PEB:

Merci à Cathie pour... tout : la science, les discussions sur le temps qu'il fait, le sport. Le labo serait bien triste sans toi !

Peter qui m'a initié aux arts subtils de l'expression de PPR ainsi qu'au football suédois.

Sabine et Anne-Laure qui ont été, l'espace de quelques mois, mes co-tuteurs de thèse ce qui a été suffisant pour me guider dans le labo.

Aaron, Tegan et Sheng pour leur constante bonne humeur.

Les « gens de passage » dont l'expérience a été inestimable : Oren, Mareike, Tom et Hans-Peter.

Jenny, Jude et Deb pour leur soutien, leur gentillesse, leur bienveillance et leur efficacité (qui a dit que l'administration était laborieuse? Pas à Perth apparemment).

A toute l'équipe « Small » avec qui j'ai pu échanger : Kamel, Sandra, Kate, Michael, Ian C., Will, Jess, Connie et Joanna.

Olivier et Phoebe pour m'avoir hébergé lors de mon arrivée à Perth.

Merci à tous les membres du laboratoire pour leur aide et leurs séminaires passionnant.

Merci à Rohan, Ben et Charlie pour m'avoir initié à la protéomique et à la biochimie des PPR.

Merci pour l'aide fourni par Michael Challenor et Swaminathan Iyer (Perth, Australie) ainsi que Kenji Okuda et Toshiharu Shikanai (Tokyo, Japon)

Merci à l'UWA pour leur soutien financier et matériel.

### Pour moi :

Merci à mes parents, Raymond et Martine, qui m'ont soutenu tout au long de ma thèse et qui ont toujours été là pour moi.

Tom et Gui, mes frangins ainsi qu'à leurs charmantes épouses Aurore et Karine.

Cédric, Olivier et Zoé mes petits bouts qui m'ont manqué lors de mes voyages à Perth.

Toute ma famille qui a pris de mes nouvelles sans arrêt...

A mes amis de Cheptainville (du quartier, du foot, du badminton ou du tennis), de Perth et d'ailleurs. Plus particulièrement à Gar'n, Jess, Fred et Jhon.

A Laure, mon écureuil.

Je dédie cette thèse à ma tante Claudine et à ma nièce Alexie.

# Abréviations

ADN = Acide DésoxyriboNucléique

ADNc = ADN Complémentaire

ADNp = ADN Plasmidique

ADN-T = ADN de Transfert

APOB = apolipoprotéine B

ARN = Acide RiboNucléique

ARNg = ARN Guide

ARNm = ARN Messenger

ARNr = ARN Ribosomique

ARNt = ARN de Transfert

ARNi = ARN Interférent

BiFC = Bimolecular Fluorescence Complementation

CATMA = Complete Arabidopsis Transcriptome MicroArray

Col-0 = *Arabidopsis thaliana*, écotype « Columbia »

CoIP = Co-ImmunoPrecipitation

CRR4 = ChloroRespiratory Reduction 4

DO = Densité Optique

*E. coli* = *Escherichia coli*

EMS = EthylMéthane Sulfonate

GST: Glutathion S-transferase

HA = hémagglutinine

His = Histidine

ICP-MS = Inductively coupled plasma-mass spectrometry (Spectrométrie par torche à plasma)

kDa = Kilo Dalton

MBP = Maltose Binding Protein (Protéine de liaison au maltose)

MORF = Multiple Organellar RNA editing Factor (Facteur d'édition multiple des organites)

NDHD = sous unité 4 de la NADH déshydrogénase

NdhD-1 = Premier site d'édition du transcrit *ndhD*

ORF = Open Reading Frame (Phase ouverte de lecture)

Pb = Paire de base

PCR = Polymerase Chain Reaction (Réaction de polymérase en chaîne)

PEB = « Plant Energy Biology »

PPE = Poisoned Primer Extension (Extension d'oligonucléotides empoisonnés)

PPR = Pentatricopeptide Repeat

PPR-DYW = PPR contenant un domaine E et DYW

PPR-E = PPR "étendue" (domaine E)

RFP = Red Fluorescent Protein (Protéine fluorescente rouge)

TAP = Tandem Affinity Purification (purification en tandem par affinité)

TEV = Tobacco Etch Virus

TILLING = Targeting Induced Local Lesions IN Genomes

URGV = « Unité de Recherche en Génomique Végétale »

UTR = UnTranslated Region (région non traduite)

YFP = Yellow Fluorescent Protein (Protéine fluorescente jaune)

Bases azotés :

A = Adénine

T = Thymine

G = Guanine

C = Cytidine

U = Uridine

# Sommaire

<b>Chapitre I : Introduction</b> .....	<b>9</b>
Préambule.....	9
I. L'édition de l'ARN.....	11
I.1. Généralités sur l'édition.....	11
I.1.1. Comment définir l'édition de l'ARN ?.....	11
I.1.2. L'édition : un changement dans la séquence des transcrits (Knoop, 2011).....	13
I.2. Caractérisation de l'édition chez les plantes.....	16
I.2.1. Un phénomène spécifique des organites.....	16
I.2.2. Différents types d'édition sur différents substrats.....	18
I.2.3. Positionnement des sites d'édition dans les transcrits des organites.....	19
I.2.4. Evolution de l'édition dans le règne végétal.....	20
II. Les facteurs d'édition chez les plantes.....	24
II.1. Analogie entre le modèle APOBEC chez l'Homme et l'édition chez les plantes.....	24
II.1.1. Le modèle APOBEC.....	24
II.1.2. Des éléments en <i>cis</i> dans les transcrits des organites chez les plantes.....	25
II.1.3. Des facteurs agissant en <i>trans</i> impliqués dans l'édition.....	26
II.1.4. Quelle enzyme pour l'édition de l'ARN ?.....	27
II.2. Identification des facteurs d'édition.....	28
II.2.1. La famille des PPR.....	28
II.2.2. Les protéines MORF.....	32
II.3. Interaction PPR-ARN : déchiffrement du code.....	37
II.3.1. Parallèle avec les protéines PUF.....	38
II.3.2. Parallèle avec les protéines TALE.....	38
II.3.3. Les protéines PPR.....	39
II.4. Les PPR contiennent-elles une activité cytidine déaminase ?.....	40
II.4.1. Implication du domaine E/E+ ?.....	40
II.4.2. Implication du domaine DYW ?.....	41
III. Projet de thèse.....	44
<b>Chapitre II : Résultats</b> .....	<b>46</b>
I. Etude génétique du candidat AtDYW1.....	46
I.1. DYW1 : une étrange PPR.....	46

I.1.1. Annotation du gène DYW1 .....	46
I.1.2. Alignement DYW1/PPR-DYW .....	47
I.2. Localisation de DYW1 .....	48
I.2.1. Prédiction <i>in silico</i> .....	48
I.2.2. Localisation subcellulaire .....	49
I.2.3. Etude de la localisation de DYW1 par western blot.....	50
I.3. Etude de DYW1 par génétique inverse : obtention de mutants TILLING.....	50
I.3.1. Obtention de lignées TILLING.....	50
I.3.2. Caractérisation des lignées TILLING.....	53
I.3.3. Conclusions.....	53
I.4. Etude des sites d'édition dans le génome chloroplastique du mutant <i>dyw1-1</i> ..	54
I.4.1. Séquençage des ADNc chloroplastiques .....	54
I.4.2. Complémentation fonctionnelle du mutant <i>dyw1-1</i> .....	55
I.4.3. Quantification de l'édition .....	56
I.4.4. Analyse de l'intégrité du transcrit <i>ndhD</i> .....	57
I.4.5. Conclusion .....	58
I.5. Effet de la mutation <i>dyw1-1</i> sur l'activité NDH : comparaison avec le mutant <i>crr4-3</i> .....	58
I.6. Analyse du transcriptome de <i>crr4-3</i> et <i>dyw1-1</i> .....	60
I.7. Recherche d'une interaction entre CRR4 et DYW1.....	61
I.7.1. Purification sur colonne cobalt de CRR4 .....	62
I.7.2. Co-immunoprécipitation de CRR4 et DYW1.....	63
I.7.3. « Split-YFP »: technique <i>in planta</i> .....	63
I.7.4. TAP-tag : recherche d'un complexe d'édition.....	65
I.8. Reconstitution d'une PPR-DYW à partir de CRR4 et DYW1 .....	68
I.8.1 Alignement des domaines E de CRR4...et de DYW1 .....	68
I.8.2. Complémentation du double mutant <i>crr4-3xdyw1-1</i> .....	69
I.9. Recherches d'orthologues du complexe minimal d'édition d'AtDYW1 .....	70
I.9.1. Orthologues chez des monocotylédones.....	70
I.9.2. Orthologues chez des dicotylédones.....	71
II. Etude fonctionnelle du candidat AtDYW1 .....	74
II.1. Etude des mutants TILLING.....	74
II.1.1. Obtention de mutants EMS .....	74
II.1.2. Analyse de l'édition du site <i>ndhD-1</i> chez les mutants EMS .....	76
II.2. Mutagenèse par PCR et complémentation de <i>dyw1-1</i> .....	77

II.2.1. Choix des mutants .....	77
II.2.2. Analyse de l'édition du site <i>ndhD-1</i> pour les mutagénèses dirigées .....	80
II.3. Etude de l'interaction CRR4-DYW1 par complémentation du double mutant <i>crr4-3x dyw1-1</i> .....	83
II.4. Recherche de l'interaction entre DYW1 et les ions métaux .....	84
II.4.1. Préparation de l'expérience .....	84
II.4.2. Difficultés rencontrées et résolution.....	85
II.4.3. Résultat de l'ICP-MS .....	87
II.5. Reconstitution de l'édition <i>in vivo</i> et <i>in vitro</i> .....	88
II.5.1. <i>In vivo</i> .....	89
II.5.2. <i>In vitro</i> .....	91
<b>Chapitre III : Discussion .....</b>	<b>94</b>
I. Rappel du modèle d'étude.....	94
II. L'enzyme d'édition contenue dans le domaine DYW chez les plantes : des données en faveur du modèle.....	95
II.1. Premières observations et changements dans l'hypothèse : DYW1 essentielle à l'édition d'un site dans les chloroplastes .....	95
II.2. Découverte d'une interaction spécifique CRR4-DYW1.....	95
II.3. Implication de protéines « DYW1-like » dans l'édition ? .....	96
II.3.1. Recrutement en <i>trans</i> d'un domaine DYW : une caractéristique commune à toutes les PPR-EE+ ?.....	96
III.3.2. Le candidat DYW2 .....	96
II.3.3. Les candidats DYW3 à DYW6 .....	97
II.4. Des caractéristiques communes avec les cytidines déaminases (CDA).....	98
II.4.1. La signature CDA indispensable à l'édition.....	98
II.4.2. Structure du domaine DYW .....	99
III. D'autres candidats à l'édition de l'ARN chez les plantes .....	100
III.1. Intégration des protéines MORF au modèle d'étude .....	100
III.1.1. Un nouveau modèle d'étude chez <i>Arabidopsis thaliana</i> .....	100
III.1.2. ... invalide chez <i>Physcomitrella patens</i> .....	100
III.2. Implication des cytidines déaminases nucléaires dans l'édition ?.....	101
IV. Cas particuliers de PPR-PLS non impliquées dans l'édition .....	103
IV.1. Une diversification de la fonction ?.....	103
IV.2. Cas de la PPR596.....	103
V. Identification du complexe d'édition et futures expériences.....	104



V.1. Identification du complexe d'édition.....	104
V.2. Reconstitution de l'édition de l'ARN <i>in vitro</i> .....	104
V.3. Reconstitution de l'édition dans un système hétérologue.....	105
VI. Pourquoi éditer les transcrits des organites? .....	106
VI.1. Augmentation de la « capacité codante » du génome des organites .....	106
VI.2. Complication du transfert du génome des organites vers le noyau .....	106
VI.3. Augmentation de la diversité.....	106
VI.4. Réparation du génome.....	107
VI.5. Régulation de l'expression des gènes.....	107
<b>Chapitre IV : Matériels et méthodes .....</b>	<b>109</b>
I. Matériels biologiques.....	109
I.1. Espèces végétales .....	109
I.1.1. <i>Arabidopsis thaliana</i> .....	109
I.1.2. <i>Nicotiana benthamiana</i> .....	109
I.2. Souches bactériennes.....	109
I.2.1. DH5 $\alpha$ (Invitrogen™, 18265-017).....	109
I.2.1. DH10B (Invitrogen™, 18290-015) .....	109
I.2.3. DB3.1 (Invitrogen™, 11782-018) .....	109
I.2.4. BL21(DE3) (Invitrogen™, C6020-03).....	110
I.2.5. C58C1 .....	110
I.3. Souches de levure .....	110
I.3.1. INVSc1 (Invitrogen™, C810-00).....	110
I.3.2. MAV203 (Invitrogen™, 11281011).....	110
II. Méthodes relatives à l'ADN .....	112
II.1. Extraction d'ADN .....	112
II.1.1 Extraction d'ADNg d' <i>Arabidopsis thaliana</i> .....	112
II.1.2. Extraction d'ADNg en plaque d' <i>Arabidopsis thaliana</i> .....	112
II.1.3. Extraction d'ADN plasmidique.....	115
II.1.4. Quantification de l'ADNg ou l'ADNp.....	115
II.2. Génotypage de l'ADN.....	115
II.2.1. Amplification de l'ADN par « Polymerase Chain Reaction (PCR) ».....	115
II.2.2. Electrophorèse sur gel d'agarose.....	116
II.2.3. Séquençage des produits de PCR .....	117
II.3. Clonage par enzyme de restriction.....	117

II.3.1. Synthèse et purification des inserts .....	117
II.3.2. Digestion de l'ADN par enzyme de restriction .....	117
II.3.3. Ligation.....	118
II.3.4. Clonage dans pETDuet-1, pCOLADuet.....	118
II.4. Clonage Gateway™ (Invitrogen™).....	119
II.4.1. Amplification par PCR du fragment à cloner.....	119
II.4.2. Purification du produit PCR.....	120
II.4.3. Recombinaison BP (Invitrogen™, 11789-013).....	120
II.4.4. Recombinaison LR (Invitrogen™, 11791-019) .....	120
II.5. Clonage pGEMT-easy.....	121
II.5.1. Ligation.....	121
II.5.2. Sélection des transformants .....	121
II.6. Mutagenèse par PCR.....	121
II.6.1. Délétion de région .....	121
II.6.2. Remplacement de nucléotides .....	122
III. Méthodes relatives aux bactéries.....	123
III.1. DH5α.....	123
III.1.1. Préparation de bactéries DH5α thermo-compétentes.....	123
III.1.2. Transformation de bactéries DH5α thermo-compétentes et sélection ...	123
III.2. Transformation de bactéries BL21/Rosetta2 thermo-compétentes.....	124
III.3. Induction de souches d'E. Coli BL21/Rosetta2.....	124
III.4. C58C1 .....	125
III.4.1. Préparation de bactéries C58C1 thermo-compétentes .....	125
III.4.2. Transformation de bactéries C58C1 thermo-compétentes.....	125
III.5. DH10B .....	126
III.5.1. Préparation de bactéries DH10B électro-compétentes.....	126
III.5.2. Transformation de bactéries DH10B électro-compétentes .....	126
IV. Méthodes relatives aux ARN .....	127
IV.1. Extraction d'ARN total et traitements.....	127
IV.1.1. Extraction par le biais du kit RNeasy plant Minikit (Qiagen, 74903) ...	127
IV.1.2. Extraction au Trizol «RNA Now »™ (Ozyme, BX-102).....	127
IV.1.3. Traitement de l'ARN total à la Turbo™DNase .....	127
IV.1.4. Quantification de l'ARN total.....	128
IV.2. Réaction de transcription inverse .....	129

IV.3. Northern Blot.....	129
IV.3.1. Transcription <i>in vitro</i> : synthèse de la sonde .....	130
IV.3.2. Préparation des ARN à tester.....	130
IV.3.3. Northern blot: transfert .....	130
IV.3.4. Northern blot: hybridation .....	131
IV.3.5. Northern blot: détection .....	131
IV.4. Extension d'oligonucléotides « empoisonnés » (PPE).....	132
IV.4.1. Purification de l'ADN complémentaire amplifié par PCR.....	132
IV.4.2. Réaction de PPE.....	132
IV.4.3. Analyse de la réaction de PPE sur gel de séquençage .....	133
IV.5. Edition <i>in vitro</i> (adapté de Hayes et <i>al.</i> , 2007).....	134
IV.5.1. Transcription <i>in vitro</i> : synthèse de l'ARN à éditer.....	134
IV.5.2. Réaction d'édition <i>in vitro</i> .....	135
IV.5.3. Transcription « inverse » et réaction de PPE.....	135
IV.6. Analyse du transcriptome chez <i>Arabidopsis thaliana</i> par CATMA.....	135
IV.6.1 Généralités sur la plateforme et l'utilisation d'une puce CATMA.....	135
IV.6.2. La puce CATMA .....	136
IV.6.3. Préparation des échantillons .....	136
IV.6.4. Procédure expérimentale.....	136
V. Méthodes relatives aux protéines .....	137
V.1. Extraction de protéines totales.....	137
V.1.1. De feuilles d' <i>Arabidopsis thaliana</i> .....	137
V.1.2. De cellules d' <i>E. Coli</i> .....	137
V.2. Migration des protéines sur gel dénaturant.....	137
V.2.1. Préparation du gel et migration .....	137
V.2.2. Coloration au bleu de coomassie.....	138
V.2.3. Coloration au nitrate d'argent .....	138
V.3. Western blot.....	139
V.3.1. Transfert sur membrane de PVDF .....	139
V.3.2. Saturation de la membrane et application des anticorps I et II .....	140
V.3.3. Révélation du western blot.....	141
V.4. Dosage des protéines au réactif de Bradford (Sigma, B6916).....	141
V.5. Clivage au TEV.....	142
V.6. TAP-tag (effectué dans le laboratoire de Geert de Jaeger, Gand, Belgique). 142	

V.7. « Split-YFP » .....	142
V.8. Purification de protéines sur colonne cobalt/nickel .....	143
V.9. Co-immunoprécipitation (coIP) .....	144
V.10. Spectrométrie par torche à plasma (ICP-MS) .....	146
V.10.1. Préparation des échantillons .....	146
V.10.2. ICP-MS .....	146
VI. Méthodes relatives aux levures .....	147
VI.1. Préparation de levures compétentes .....	147
VI.2. Transformation de levures compétentes .....	147
VI.3. PCR sur colonies .....	148
VII. Méthodes relatives aux plantes .....	149
VII.1. Synchronisation de la germination des graines d' <i>Arabidopsis thaliana</i> .....	149
VII.2. Culture <i>in vitro</i> .....	149
VII.2.1. Stérilisation des graines .....	149
VII.2.2. Culture <i>in vitro</i> .....	149
VII.2.3. Transfert des plantes sur terreau .....	150
VII.2.4. Semis de graines non stérilisées sur milieu silice, ¼ MS sans sucre .....	150
VII.3. Croisement de plantes d' <i>Arabidopsis thaliana</i> .....	150
VII.4. Transformation de plantes .....	151
VII.4.1. Expression transitoire chez <i>Nicotiana benthamiana</i> .....	151
VII.4.2. Préparation de protoplastes .....	151
VII.4.3. Transformation stable d' <i>Arabidopsis thaliana</i> .....	151
VIII. Obtention de mutants : étude par génétique inverse .....	152
VIII.1. Mutants TILLING .....	152
VIII.2. Mutants d'insertion .....	152
IX. Outils bioinformatiques utilisés .....	153
IX.1. Alignement de séquences .....	153
IX.2. Création d'amorces pour PCR et PPE .....	153
IX.3. Alignement de séquences sur un génome .....	153
IX.4. Recherche de motifs PPR .....	153
IX.5. Enzymes de restriction .....	153
IX.6. Recherches bibliographiques .....	154
IX.7. Bases de données .....	154
IX.8. Prédiction de la localisation sub-cellulaire .....	154

IX.9. Prédiction de la structure secondaire et tertiaire d'une protéine .....	154
IX.10. Recherche d'interactions protéine/protéine chez Arabidopsis .....	155
<b>Références bibliographiques .....</b>	<b>156</b>
<b>Annexe.....</b>	<b>170</b>
I. Publication .....	170
II. Liste des principaux vecteurs utilisés .....	182
III. Liste des principaux oligonucléotides utilisés .....	187
IV. Rapport de l'ATP sur le TILLING de DYW1 .....	194
V. Table des figures.....	195
VI. Table des tableaux.....	197
<b>Résumé/Summary .....</b>	<b>198</b>
I. Résumé (français) .....	198
II. Summary (english).....	199

# Chapitre I : Introduction

## Préambule

Les cellules eucaryotes sont composées de compartiments appelés organites, essentiels au bon fonctionnement de la cellule. Parmi ces organites, les mitochondries et les plastes ont évolué tous deux à partir d'une cellule procaryote. En effet, ils ont pour origine une endosymbiose entre une cellule eucaryote et une  $\alpha$ -protéobactérie, pour les mitochondries, ou une cyanobactérie, pour les plastes (Gray, 1999; Gray et al., 1999; Dyall et al., 2004). Ces deux organites ont donc la particularité de posséder leur propre génome, dont le nombre de gènes s'est réduit au cours du temps. Ceci est majoritairement dû à une perte directe des gènes mais aussi au transfert des gènes vers le noyau (Adams and Palmer, 2003).

Il y a plus de 50 ans, la découverte de la structure en double hélice de l'ADN a permis de mieux comprendre le langage « parlé » par les gènes. Les quatre bases azotées composant l'alphabet de l'ADN, adénosine (A), cytidine (C), guanosine (G) et thymidine (T) ont été identifiées et font maintenant partie des connaissances du grand public. L'ADN est transcrit en ARN, des molécules copiant le brin codant de l'ADN à l'exception des thymidines remplacées par des uridines (U).

Au contraire de ce que l'on pourrait penser, il semblerait que l'uridine soit plus ancienne que la thymidine. En effet, le génome des premières cellules vivantes était probablement composé d'ARN et non d'ADN. Même si elle est spécifique de l'ARN, l'uracile peut être formée par désamination oxydative d'une cytosine impliquant une mutation dans l'ADN après répllication, l'uridine s'associant à l'adénosine et non à la guanosine comme le ferait la cytosine (A-U au lieu de C-G). Si une uridine est par erreur incorporée dans l'ADN, l'organisme répare de lui-même sa séquence pour former une cytosine à la place.

Concernant l'origine de l'uridine, une équipe de la NASA est parvenue à créer, en laboratoire, de l'uracile à partir d'une molécule de pyrimidine par bombardement d'ultraviolet. Le but de l'expérience ici était de recréer les conditions trouvées dans l'espace (pression, température et radiation). L'équipe a proposé que l'uracile soit éventuellement apparue pendant la panspermie, une période hypothétique où la vie serait apparue sur Terre par bombardement de météorites (<http://www.nasa.gov/centers/ames/news/features/2009/urasil.html#>).

Une grande quantité d'ARN est produite pour synthétiser des protéines. Pour cela, chaque triplet de bases azotées présent dans l'ARN, appelé codon, est lu et traduit en un alphabet composé de 20 caractères, appelés acides aminés, qui sont les composants de la protéine. Jusqu'aux années 70, il était proposé que chaque gène impliquait la formation d'un seul transcrit et donc d'une seule protéine. Ce dogme a été réfuté après la découverte des introns et du phénomène d'épissage/épissage alternatif. Ainsi pour

obtenir un ARN fonctionnel, les introns doivent être excisés et les exons reliés les uns aux autres. Ajouté à l'épissage, un autre phénomène de modification post-transcriptionnelle permet le changement d'une base en une autre augmentant encore la complexité des génomes et des transcrits. Ce phénomène est appelé «édition de l'ARN».

Dans cette thèse, je me suis focalisé sur le phénomène d'édition chez les plantes, correspondant à une modification d'une cytidine en uridine. Cependant, il arrivera que je fasse le parallèle avec l'édition chez l'Homme ou encore le trypanosome.

En effet, les études effectuées sur l'Homme ont été pionnières dans ce domaine, des facteurs d'édition étant impliqués notamment dans la réponse immunitaire. Un exemple parmi tant d'autres est celui d'APOBEC3G impliqué dans la lutte contre le virus de l'immunodéficience humaine (Malim and Bieniasz, 2012). Le nombre de publications sur cette dernière indique un très fort intérêt de la communauté scientifique pour l'édition.

# I. L'édition de l'ARN

## I.1. Généralités sur l'édition

### I.1.1. Comment définir l'édition de l'ARN ?

L'édition de l'ARN a été assez difficile à définir pendant des années car cette expression a été utilisée pour décrire différents processus d'altération de l'ARN. Avant de commencer à parler de l'édition chez les plantes, il est donc important de définir ce processus.

Le terme « édition de l'ARN » a été introduit en 1986 après une étude sur le trypanosome, *Trypanosoma brucei* (Benne et al., 1986). Dans cette publication, il a été montré que le transcrit *coxII* possédait quatre nucléotides qui n'étaient pas présents dans l'ADNg. Après avoir écarté l'hypothèse de la présence d'une deuxième copie du gène *COXII* par Southern blot, les auteurs se sont rendus à l'évidence que le transcrit était modifié, par insertion de nucléotides, pendant ou après la transcription (Figure 1).

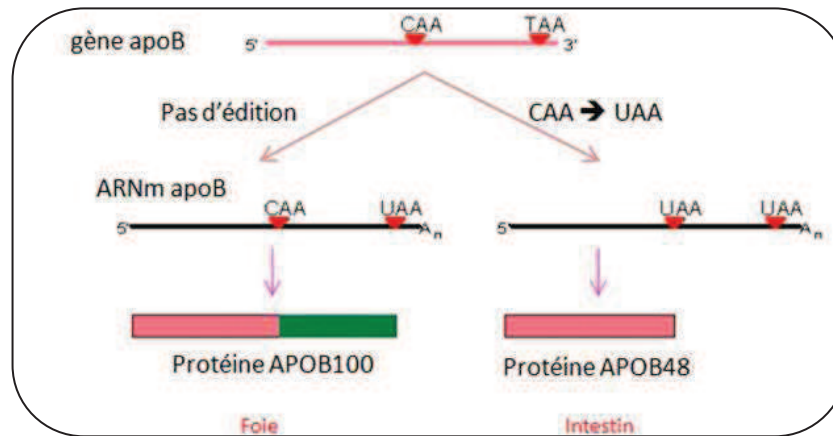
Cox-II	
CAAGTTTAGGTATAAAAAGTAGATTGTATACCTGGTAGCTGTAATGAAATAATTTTGTTCCTAAA	ADN
***** * * *****	
CAAGTTTAGGTATAAAAAGTAGA G A ACCTGGTAGCTGTAATGAAATAATTTTGTTCCTAAA	ARN

**Figure 1 : Edition de l'ARNm *coxII* chez le trypanosome (d'après Benne et al., 1986).**  
Quatre uridines sont insérées dans l'ARNm de *coxII*. Un ARNm différent de l'information génétique de base est alors synthétisé. Les astérisques (\*) représentent les identités de séquence entre l'ADN et l'ARN.

Très peu de temps après cette publication, un autre exemple d'édition est observé chez l'Homme après transcription du gène *APOB*. Ainsi, deux formes du transcrit *apoB* sont produites, *apoB-100* qui est exprimé dans le foie et *apoB-48* dans l'intestin. Malgré la très forte homologie entre ces deux transcrits, il n'a jamais pu être mis en avant deux gènes différents permettant la production de ces deux formes. Il est en fait nécessaire au transcrit primaire d'*apoB* d'être édité de sorte à pouvoir produire deux transcrits : alors que le transcrit *apoB-100* est similaire à l'ADNg, le transcrit *apoB-48* est modifié au milieu de sa séquence (Powell et al., 1987). Ce changement a pour conséquence la formation d'un codon STOP, raccourcissant la protéine de moitié. Malgré cette importante modification, la protéine APOB-48 est toujours fonctionnelle, étant impliquée dans le transport et l'absorption du cholestérol et des triglycérides. APOB100 est impliquée dans l'assemblage des lipoprotéines de très petites tailles (VLDL en anglais) (Figure 2). Dans cette publication de 1987, tout à la fin de l'article, le parallèle est fait avec la modification de *coxII* découverte en 1986.

A nouveau chez le trypanosome, l'édition par délétion de nucléotide a été décrite deux ans après la découverte de l'édition par insertion de nucléotide (Shaw et al., 1988). L'insertion et la délétion de nucléotides impliquant l'apparition de codon méthionine et de nouveaux transcrits, il a été avancé que l'édition avait une fonction centrale chez le trypanosome.





**Figure 2 : Edition de l'ARNm *apoB* chez l'Homme (extrait de web-books.com).**

L'édition de l'ARN *apoB* (CAA→UAA) permet la formation de deux transcrits de taille différente permettant la traduction de APOB100 et APOB48.

Par la suite, l'édition des ARN a été définie tout d'abord comme un événement post-transcriptionnel différent de l'épissage modifiant l'ARN (Cattaneo, 1991). Il a été aussi inclus dans le terme édition de l'ARN, les événements de « balbutiement » (stuttering en anglais) de l'ARN polymérase chez la famille des paramyxovirus, ou encore la polyadénylation en 3' des ARN (Brennicke et al., 1999; Knoop, 2011).

C'est en 1998 que la définition la plus acceptée par la communauté scientifique a été émise par le biais du manuscrit de Price et Gray et qui a été repris dans bon nombre de publications. Je cite ici une phrase définissant l'édition de l'ARN :

« RNA editing describes processes of nucleotide alterations which result in different or additional nucleotides in the RNA which could as well have been encoded in the genomic sequence » (Brennicke et al., 1999).

En 2011, Volker Knoop donne une définition qui, à mon sens, complète celle citée précédemment :

« RNA editing [...] describes sequence changes introduced through selective nucleotide insertions, deletions, and substitutions, which could alternatively be directly encoded by the four standard nucleotides in the gene » (Knoop, 2011).

Il s'agit donc d'une altération des ARN due à une insertion, délétion ou substitution à un site très spécifique qui change l'information codée dans l'ADNg sur un ou plusieurs nucléotides. Les nucléotides insérés sont les nucléotides standards trouvés dans les gènes et les transcrits. Cette définition exclut donc les modifications post-transcriptionnelles chimiques comme la méthylation qui ne change pas l'information codée par le génome.

De même, les modifications de A en I observées chez les métazoaires ne sont pas reconnues comme de l'édition « vraie », l'inosine n'étant pas présente dans les gènes. Néanmoins comme l'inosine est reconnue comme une guanosine, ce type de modification est généralement considéré comme de l'édition (Sommer et al., 1991;

Knoop, 2011). Ce processus est catalysé par les protéines de type ADAR (Bass and Weintraub, 1988).

### I.1.2. L'édition : un changement dans la séquence des transcrits (Knoop, 2011)

L'édition de l'ARN est observée chez différentes espèces d'eucaryotes dans les ARNm, les ARNt et les ARNr. Chez les procaryotes elle ne concerne que les ARNt et ARNr. Enfin, de l'édition a même été décrite dans les microARN chez l'Homme et la souris (Ohman, 2007) (Tableau 1).

**Tableau 1 : Edition de l'ARN dans différents organismes suivant leur date de découverte (d'après Knoop et al., 2011).**

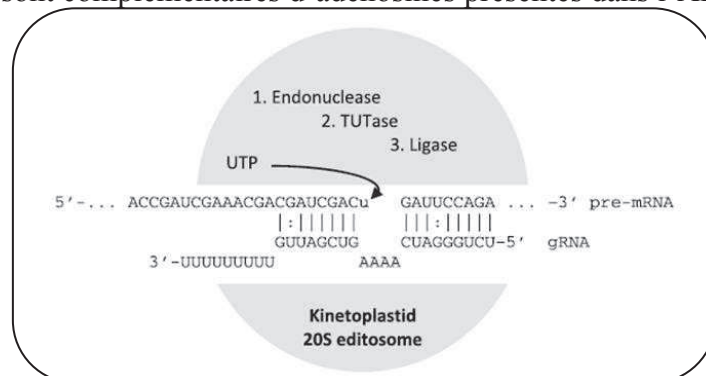
Date	Organismes	Type d'édition de l'ARN	ARN ciblé(s)	Localisation
1986	Kinétoplastes	Insertions et délétions de U	ARNm	Mitochondriale
1987	Mammifères	C en U	ARNm et ARN viraux	Nucléaire
1988	Paramyxovirus	G et A en U	ARNm	Virale
1989	Plantes	C en U, U en C	ARNm, ARNt	Mitochondriale
1991	Plantes	C en U, U en C	ARNm, ARNt	Chloroplastique
1991	Myxomycètes	Insertions de C, insertions de U, C en U	ARNm, ARNt, ARNr	Mitochondriale
1991	Métazoaires	A en I	ARNm, ARNt, miARN et ARN viraux	Nucléaire
1993	Acanthamoeba (protiste) et Chytridiomycetes (champignons)	N en N'	ARNt en 5'	Mitochondriale
1993	Marsupiaux	C en U	Anticodon ARNt	Mitochondriale
2002	Dinoflagellés	A en G, G en A, C en U, U en C, U en R, G en C	ARNm	Mitochondriale
2004	Dinoflagellés	A en G, G en A, C en U, U en C, R en C, U en G	ARNm, ARNr	Chloroplastique
2009	Archéobactéries	C en U	ARNt	Bactérienne
2009	Placozoaires	U en C	ARNm	Mitochondriale

Comme décrit précédemment, deux grands types d'édition de l'ARN peuvent être distingués :

- (1) l'édition par délétion ou insertion de nucléotides,
- (2) l'édition par conversion de nucléotides.

(1) L'édition par délétion ou insertion de nucléotides concerne essentiellement les organismes tels que les trypanosomes (Benne et al., 1986), les virus (Thomas et al., 1988) ou encore les champignons (Mahendran et al., 1991). La découverte de l'édition a souvent été associée au fait que quelque chose était incohérent entre le gène et son expression.

Comme décrit précédemment, le gène mitochondrial *COXII* chez le trypanosome possède un changement de cadre de lecture corrigé uniquement dans le transcrit lors de l'apparition de 4 uridines dans l'ARNm *coxII* (Benne et al., 1986). L'événement inverse est aussi observé chez les trypanosomes. En effet, des uridines bien spécifiques peuvent être supprimées de la séquence des ARNm (Shaw et al., 1988). Dans le transcrit *coxIII* de *Trypanosoma brucei*, l'édition affecte plus de 50% du transcrit soit 550 uridines insérées et 41 supprimées (Feagin et al., 1988). A ce stade, nous parlons de « pan-editing » (Simpson and Shaw, 1989). Dans le cas du trypanosome, les ARN guides (ARNg) permettent la reconnaissance de l'ARN à éditer (Blum et al., 1990). Les ARNg se lient au pré-ARNm et indiquent le nombre d'uridines à ajouter. Les uridines insérées dans l'ARNm sont complémentaires d'adénosines présentes dans l'ARNg (Figure 3).



**Figure 3 : Schéma simplifié de l'édition par insertion de nucléotides dans les mitochondries des kinéoplastes (Extrait de Knoop 2011).**

L'ARNg contenant des uridines en 3' se lie avec sa cible pré-ARNm. Le nombre d'Adénines supplémentaires dans la séquence indique le nombre d'uridines à insérer dans le pré-ARNm.

Le complexe d'éditosome 20S contient 3 enzymes différentes permettant l'édition:

- l'endonucléase clive le pré-ARNm au niveau du site d'édition,
- la TUTase (« terminal uridylyl transferase ») permet l'insertion d'UTP en 3',
- la ligase relie les deux extrémités de l'ARNm.

Dans cet exemple (hypothétique), le pré-ARNm a été édité et 4 uridines ont été insérées pour former un ARNm dont la séquence est différente de l'ADN.

De la même façon, l'ARNg est aussi impliqué dans la délétion d'uridines, cette fois en ne possédant pas les adénosines complémentaires aux uridines présentes dans le pré-ARNm. Ainsi, les ARNg agissent en *trans* dans la plupart des cas. Néanmoins, l'ARNg permettant l'édition du transcrit *coxII* chez les trypanosomes est présent dans sa région 3'UTR (pour « UnTranslated Region »), le pré-ARNm porte donc une région « guide » qui agit en *cis* du même ARNm (Golden and Hajduk, 2005). D'un point de vue biochimique, le complexe d'édition chez les trypanosomes, appelé éditosome, contient trois enzymes majeures permettant l'ajout de nucléotides. Tout d'abord, une endonucléase qui coupe l'ARNm au niveau du site d'édition, puis une TUTase (pour « terminal uridine transferase ») ajoutant les uridines au site d'intérêt et enfin une ligase permettant la reconstitution de l'ARNm incluant les modifications dues à l'édition (Knoop, 2011).

Chez les virus, l'introduction d'un nucléotide non codé par le génome a été démontrée dans différentes espèces de virus tel le virus simiiforme SV5 (Thomas et al., 1988), la rougeole (Cattaneo et al., 1989) ou encore le virus de Sendai (Vidal et al.,

1990a). Peu de temps après la découverte de l'édition chez les virus, le mécanisme a été identifié comme étant dû à un balbutiement de la polymérase, c'est-à-dire que la polymérase transcrit le même nucléotide plusieurs fois (Vidal et al., 1990b). Dans le cas du virus Ebola, une adénosine a été insérée dans le transcrit codant la glycoprotéine en plein milieu de sa région codante (Volchkov et al., 1995). L'insertion change le cadre de lecture et permet la production d'une protéine plus longue et active. Il est donc envisageable que, dans ce cas, l'édition régule l'infection par le virus Ebola (Knoop, 2011). Il a été montré chez un mutant n'exprimant que la forme longue de la glycoprotéine une augmentation de la cytotoxicité par le virus (Volchkov et al., 2001).

L'édition par insertion de nucléotides a aussi été observée chez les myxomycètes : ainsi, par exemple, plusieurs cytidines sont insérées dans l'ARNm mitochondrial codant pour la sous unité 1 de l'ATPase (Mahendran et al., 1991).

En plus des insertions très fréquentes de cytidines, un cas d'insertion d'une uridine a été découverte dans le transcrit *coxI* de *Physarum polycephalum* (Gott et al., 1993). Dans le cas des myxomycètes, il ne semble pas que l'édition se fasse, comme chez le trypanosome, à l'aide d'un ARNg (aucun ADN codant pour des ARNg n'a pu être identifié). Mais il est montré que l'édition dans ce cas est cotranscriptionnelle (Visomirski-Robic and Gott, 1995, 1997). L'étude du transcrit codant pour l'ARN polymérase montre que l'édition par addition de nucléotides en 3' (Miller and Miller, 2008) est dépendante de 9 pb entourant le site d'insertion de nucléotides dans l'ADNg (Rhee et al., 2009).

(2) L'édition par conversion de nucléotides est observée chez un grand nombre d'espèces comme les métazoaires (Chen et al., 1987; Powell et al., 1987; Sommer et al., 1991; Janke and Paabo, 1993; Burger et al., 2009), les plantes (Covello and Gray, 1989; Gualberto et al., 1989; Hiesel et al., 1989; Hoch et al., 1991), les myxomycètes (Gott et al., 1993), les champignons (Laforest et al., 1997), certains protistes (Lin et al., 2002; Zauner et al., 2004), ou encore les bactéries (Randau et al., 2009).

Chez les métazoaires, la première preuve d'édition a été apportée en 1987 avec la découverte des différentes formes du transcrit *apoB*, produites par le même gène. Le transcrit est soumis à l'édition par conversion d'une cytidine en uridine, transformant un codon CAA en codon STOP UAA de manière tissu spécifique : le foie pour APOB100 et les intestins pour APOB48 (Chen et al., 1987; Hospattankar et al., 1987; Powell et al., 1987). D'un point de vue moléculaire, ce système d'édition est plus simple que celui d'ajout ou de délétion de nucléotides car il ne touche pas au squelette nucléotidique composant l'ARN. En effet, il s'agit d'une conversion par déamination due à une cytidine déaminase dépendante du zinc. APOBEC-1 a été identifiée comme étant l'enzyme catalysant cette réaction d'édition (Teng et al., 1993). Il s'agit du cas le plus documenté de conversion de nucléotides chez les mammifères, d'autres sites (peu nombreux) ayant été découverts mais dont les fonctions restent obscures. A noter, quelques cas de désamination de l'ADN chez les vertébrés par des membres de la famille APOBEC, notamment APOBEC-3 qui semble avoir une action antivirale en désaminant l'ADN des provirus. Cette protéine semble alors capable d'affecter fortement la réplication ainsi que l'activité de rétrotransposition des virus (Wiegand and Cullen, 2007; Chen et al., 2008).

Toujours chez les métazoaires, plus spécifiquement chez les marsupiaux, l'anticodon ARNt GCC (décodant l'aspartate) est édité en GUC (décodant la glycine) à hauteur de 50% par un évènement de désamination de C en U (Janke and Paabo, 1993). A l'inverse, dans les mitochondries d'*Euhadra herklotsi*, il a été observé un phénomène d'édition, non pas dans les anticodons des ARNt, mais dans leurs accepteurs 3'. Ainsi, des bases C, G ou U sont modifiées en A de sorte à corriger les problèmes d'association avec les nucléotides de l'accepteur 5' (Yokobori and Paabo, 1995). Le même mécanisme a été observé chez la pieuvre (*Loligo blekeri*) et le poulet (*Gallus gallus*) (Tomita et al., 1996; Yokobori and Paabo, 1997).

Chez les placozoaires, les organismes les plus simples connus chez les métazoaires, un évènement d'édition de U en C permet la restauration d'un codon CAU, à partir d'un codon UAU, extrêmement conservé dans l'ARNm mitochondrial de *coxI* (Burger et al., 2009). Il s'agit ici du premier évènement d'édition de U en C découvert dans les mitochondries des métazoaires.

Chez les myxomycètes, les premières évidences d'édition de l'ARN par substitution de nucléotides ont été obtenues en 1993 avec la découverte de quatre cytidines dans l'ARNm de *coxI* modifiées en uridines (Gott et al., 1993).

Chez le champignon *Spizellomyces punctatus*, les premières bases composant l'accepteur en 5' de l'ARNt sont éditées par modification de U en A, U en G et de A en G (Laforest et al., 1997). Il s'avère que le mécanisme d'édition observé dans les ARNt des champignons est extrêmement similaire à celui des protistes ce qui est assez fascinant (Lonergan and Gray, 1993). Le mécanisme d'édition de l'accepteur en 5' des ARNt semble ainsi être très ancien.

Chez les bactéries, un cas d'édition « vraie » a été découvert assez récemment dans les ARNt. Dans l'accepteur en 5', une uridine très conservée est manquante (une cytidine la remplace) dans 30 des 34 ARNt présents chez *Methanopyrus kandleri* (Randau et al., 2009). Les cytidines sont converties en uridines chez tous ces ARNt par l'enzyme CDAT8 (Cytidine DeAminase Type 8).

Enfin, l'édition chez les plantes, faisant l'objet de ma thèse, sera développée dans les paragraphes suivants.

## I.2. Caractérisation de l'édition chez les plantes

### I.2.1. Un phénomène spécifique des organites

Le premier manuscrit rendant compte de ce que l'on appelle dorénavant l'édition de l'ARN dans les organites a été publié en 1981, un évènement antérieur à la découverte du phénomène d'édition en 1986 (Fox and Leaver, 1981). Dans cette publication, le gène mitochondrial *cox2* du maïs a été séquencé. Il s'agissait du premier séquençage d'un gène codant pour une protéine mitochondriale chez les plantes. La comparaison des séquences en acides aminés avec COX2P chez des espèces non-végétales a montré des codons CGG (arginine) s'alignant avec des résidus tryptophanes (codon TGG), ces derniers étant extrêmement conservés chez un grand nombre d'espèces étudiées à l'époque. Il a été proposé que ce codon CGG ne codait pas une arginine mais un tryptophane dans le génome mitochondrial du maïs (en plus du codon TGG). A

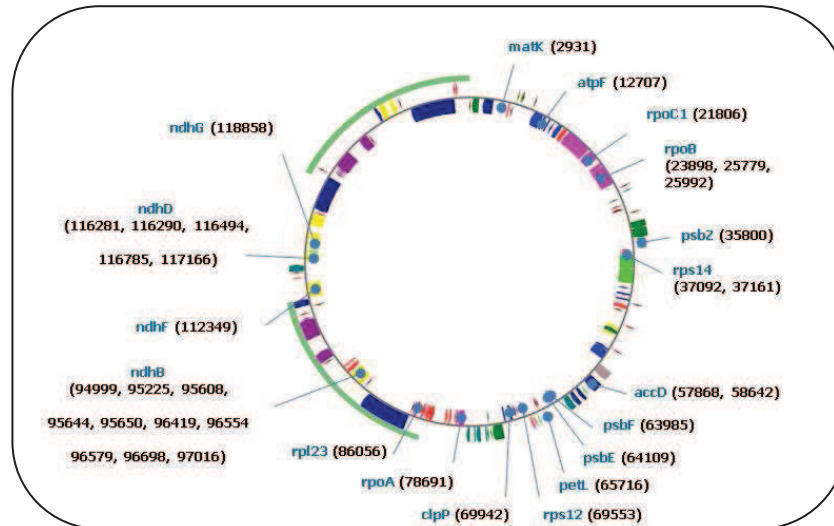
l'époque, on avait découvert que le code génétique du génome mitochondrial chez les animaux et les levures était différent du code standard (Barrell et al., 1979; Macino et al., 1979). Il était possible qu'il en soit de même chez les plantes. Néanmoins, il est à noter qu'aucune explication n'a pu être fournie sur comment un codon pouvait coder deux acides aminés différents (tryptophane et arginine). De même, aucun ARNt décodant un triplet CGG en tryptophane n'avait été découvert.

Cette problématique a été résolue 8 années plus tard avec la découverte de l'édition des transcrits mitochondriaux (Covello and Gray, 1989; Gualberto et al., 1989; Hiesel et al., 1989). En effet, le codon CGG, que l'on pensait coder pour un tryptophane dans les mitochondries à l'époque, est en fait édité pour être transformé en un codon UGG (tryptophane).

C'est donc en 1989 que sont parues trois publications décrivant des événements d'édition dans des transcrits mitochondriaux des plantes. Deux ans plus tard, il a été montré que l'édition avait également lieu dans les chloroplastes du maïs (Hoch et al., 1991). A l'heure actuelle, plus de 500 sites ont été identifiés chez *Arabidopsis thaliana* dans les ARN mitochondriaux (Bentolila et al., 2008; Zehrmann et al., 2008) (Tableau 2) et 34 sites dans les chloroplastes (Chateigner-Boutin and Small, 2007) (Figure 4). Le nombre de sites d'édition mitochondriaux peut aller jusqu'à 1000 chez le gymnosperme *Cypripedium acaule* (Chaw et al., 2008), 1782 sites chez la fougère *Isoetes engelmannii* (Grewe et al., 2011) et 2139 sites d'édition mitochondriaux (pour 18 gènes) chez le lycophyte *Selaginella moellendorffii* (Hecht et al., 2011).

**Tableau 2 : Sites d'édition présents dans les transcrits mitochondriaux d'*Arabidopsis thaliana* (adapté de Bentolila et al., 2008).**

Transcrit	Nb de sites édités	Transcrit	Nb de sites édités
<i>atp1</i>	16	<i>nad6</i>	10
<i>atp6-1</i>	1	<i>nad7</i>	32
<i>atp9</i>	4	<i>nad9</i>	7
<i>ccb203</i>	13	<i>orf114</i>	1
<i>ccb256</i>	34	<i>orf240A</i>	1
<i>ccb382</i>	13	<i>orf25</i>	11
<i>ccb452</i>	28	<i>orfB</i>	0
<i>cob</i>	15	<i>orfX</i>	52
<i>cox2</i>	11	<i>rpl16</i>	5
<i>cox3</i>	7	<i>rpl2</i>	7
<i>matR</i>	20	<i>rpl5</i>	9
<i>nad1</i>	20	<i>rps12</i>	8
<i>nad2</i>	29	<i>rps14</i>	2
<i>nad3</i>	7	<i>rps3</i>	30
<i>nad4</i>	45	<i>rps4</i>	16
<i>nad4L</i>	10	<i>rps7</i>	3
<i>nad5</i>	37	<b>Total</b>	<b>504</b>



**Figure 4 : Positionnement des sites d'édition dans le génome chloroplastique d'*Arabidopsis thaliana*.** A ce jour, 34 sites d'édition ont été découverts dans les transcrits d'*Arabidopsis thaliana*.

Il a été montré que l'édition est essentielle à la production de protéines fonctionnelles car elle restaure des acides aminés extrêmement conservés (Covello and Gray, 1990). Les exemples d'*atp9* et de *psbF* étayent cette thèse. L'accumulation dans les mitochondries de protéines ATP9, provenant d'ARNm non édité, résulte en un dysfonctionnement des mitochondries impliquant une stérilité male dans les mutants de tabac et *Arabidopsis* (Hernould et al., 1993; Gomez-Casati et al., 2002). Ce phénotype peut être aboli par inhibition de l'expression des protéines Atp9 (Busi et al., 2006). De même, la synthèse de l'ARN *psbF* non édité dans les chloroplastes résulte en une déficience du photosystème II (Bock et al., 1994).

### I.2.2. Différents types d'édition sur différents substrats

A ce jour, deux types d'édition ont été identifiés chez les plantes :

- modification de cytidine en uridine (C → U) (Figure 5),
- modification d'uridine en cytidine (U → C).

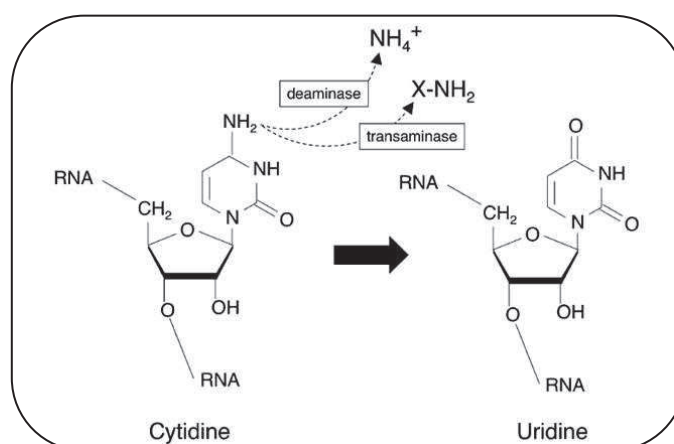
L'édition de C en U, la plus commune, a été identifiée dans les organites de toutes les plantes terrestres étudiées à l'exception des marchantiales (Groth-Malonek et al., 2007). L'édition de C en U des ARNt des organites est très rare mais est représentée chez *Isoetes engelmannii* (Grewe et al., 2011).

La réaction inverse de U en C est absente chez les mousses et les plantes à graine et les marchantiales mais est très répandue dans les ARNm et ARNt des organites des fougères et des lycophytes (Steinhauser et al., 1999; Grewe et al., 2011). Aucune édition retrouvée chez les plantes n'a été identifiée chez les algues, pas même chez les algues vertes pourtant assez proches des plantes terrestres (Steinhauser et al., 1999).

Plus récemment, il a été montré que les ARNr mitochondriaux de *Selaginella moellendorffii* semblent aussi être soumis à l'édition (Hecht et al., 2011).

De même, par séquençage haut débit, il a été montré que les miARN sont édités (modification de C en U ou de U en C) chez *Arabidopsis thaliana* et *Oryza sativa* (Ebhardt et al., 2009).

Il a été observé que toutes les molécules d'ARN contenant un site d'édition n'étaient pas forcément éditées, c'est ce que l'on appelle l'édition partielle. Dans les transcrits des végétaux, la moitié des sites d'édition sont partiellement édités dans les transcrits chloroplastiques et 14% dans les transcrits mitochondriaux d'*Arabidopsis thaliana* (Zehrmann et al., 2008; Takenaka et al., 2012). La découverte des sites partiels d'édition a été le premier argument permettant de dire que l'édition était un processus séparé de la transcription (Takenaka et al., 2008).



**Figure 5 : Edition de l'ARN par conversion de nucléotides de cytidine en uridine (Extrait de Takenaka et al., 2008).**

L'édition de l'ARN a lieu par déamination d'une cytidine. Le groupement amine pourrait être modifié par le biais d'une cytidine déaminase ou transféré à une autre molécule grâce à une transaminase. Néanmoins, il est peu probable qu'une transaminase soit impliquée, les expériences d'édition *in vitro* montrant qu'une molécule acceptrice du groupement amine n'était pas nécessaire.

### I.2.3 Positionnement des sites d'édition dans les transcrits des organites

L'édition de l'ARN change l'information contenue dans le génome. L'édition de C en U chez les ARNm des plantes est généralement retrouvée dans les régions codantes du génome.

Pour les transcrits mitochondriaux, la majorité des sites d'édition (89% chez *Arabidopsis thaliana*) est localisée dans la première et seconde base du codon, changeant ainsi l'identité de l'acide aminé (Giege and Brennicke, 1999; Jobson and Qiu, 2008). Dans certains cas, l'édition de l'ARN peut conduire à la formation de codons d'initiation ou de terminaison de la traduction et ainsi restaurer une phase ouverte de lecture absente du gène (Grewe et al., 2011). Dans ce cas précis, l'édition affecte directement le processus de traduction.



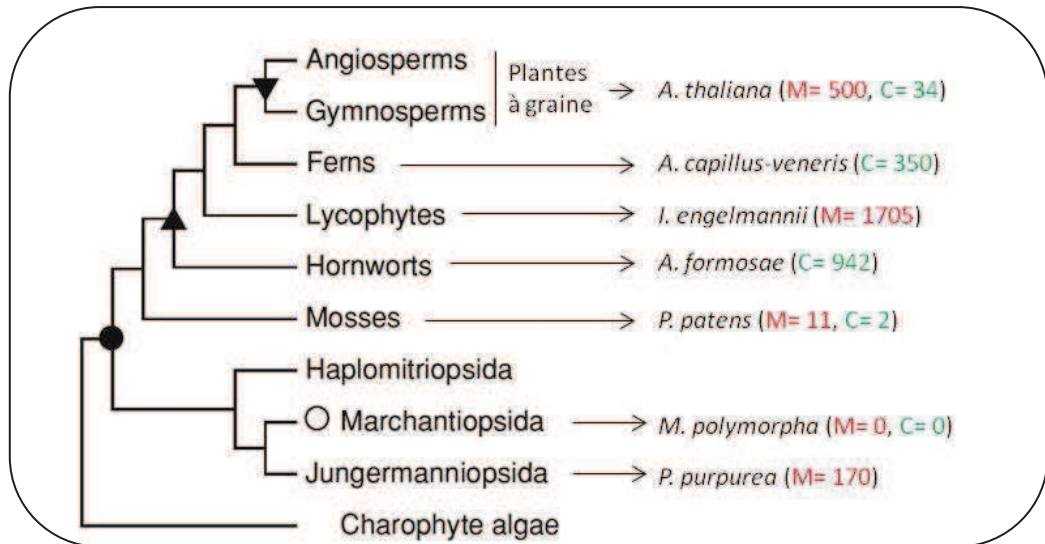
Même si la grande majorité des sites d'édition est observée dans les régions codantes, certains sites sont localisés dans des régions non-codantes (en plus des ARNt décrits précédemment). Chez *Arabidopsis thaliana* par exemple, huit sites d'éditions mitochondriaux sont situés dans les introns (Giege and Brennicke, 1999). L'édition dans les régions non-codantes ne semble pas correspondre à l'idée selon laquelle l'édition est essentielle à la synthèse de protéines fonctionnelles. Néanmoins, il est maintenant clair que l'édition de certains sites dans des introns est nécessaire à l'épissage. Elle permet ainsi la correction d'erreurs dans des régions double-brin indispensables à la conformation des introns de groupe II (Grewe et al., 2011) mais aussi chez les introns de groupe I (Begu et al., 2011). De plus, l'édition de l'ARN est aussi observée dans les régions non codantes (UTR) en 5' et en 3' des ORF (Giege and Brennicke, 1999; Kudla and Bock, 1999). Même si aucune preuve expérimentale n'a été mise en évidence, il est possible que ces sites aient un effet sur la traduction de l'ARNm.

Il est assez étrange de voir que l'édition est aussi présente à des positions silencieuses des codons (Castandet and Araya, 2011). En effet, ces changements n'induisent aucune modification de la séquence en acides aminés de la protéine ce qui est aberrant. La plupart de ces modifications silencieuses a lieu au niveau de la troisième base du codon. L'implication de ces événements silencieux, que ce soit dans la régulation ou la maturation des ARN, reste à élucider. Néanmoins, il est possible que la séquence du transcrit puisse être assez proche de celle d'un transcrit devant être édité et qu'ainsi un facteur d'édition reconnaisse ce transcrit malgré tout.

En ce qui concerne les transcrits chloroplastiques, le même schéma se dessine que pour les transcrits mitochondriaux, pour beaucoup moins de sites d'édition. L'édition de l'ARN a lieu surtout dans les premières et secondes bases du codon et les modifications silencieuses restent assez rares (Tsudzuki et al., 2001). L'édition permet dans un cas de restaurer une ORF en modifiant un codon ACG en codon AUG (Kotera et al., 2005). C'est le cas du transcrit *ndhD* qui est édité à la deuxième base du premier codon ACG. Ceci conduit à la formation d'un codon d'initiation de la traduction AUG et ainsi à la traduction de la sous unité NdhD du complexe NDH (Kotera et al., 2005).

#### I.2.4. Evolution de l'édition dans le règne végétal

L'édition de l'ARN de C en U se fait dans les chloroplastes et les mitochondries chez toutes les plantes terrestres connues et non chez les algues vertes. Il est donc très probable que l'origine de l'édition remonte à l'apparition des premières plantes terrestres (Groth-Malonek et al., 2007) (Figure 6). Les études effectuées par séquençage des génomes mitochondriaux et chloroplastiques ont montré que la fréquence de l'édition de l'ARN variait fortement de plante à plante (Maier et al., 1995; Giege and Brennicke, 1999; Kugita et al., 2003; Grewe et al., 2011) (Figure 6).



**Figure 6 : Distribution phylogénétique de l'édition chez les plantes (adaptée de Knoop et al., 2011).** Quatre événements majeurs concernant l'édition sont représentés ici.

- (1) L'édition chez les plantes est apparue chez les plantes terrestres (cercle noir).
- (2) Une perte secondaire de l'édition est observée chez les marchantiales (cercle blanc).
- (3) Chez les trachéophytes/Anthocérothes (« Hornworts »), le nombre de sites d'édition, lorsqu'il est comparé aux bryophytes (« mousses »), a explosé aussi bien dans les mitochondries que dans les chloroplastes (triangle vers le haut).
- (4) Les plantes à graine montrent une diminution de l'édition dans les transcrits mitochondriaux et chloroplastiques (triangle vers le bas).

M (en rouge) = nombre de sites d'édition mitochondriaux.

C (en vert) = nombre de sites d'édition chloroplastiques.

*A. thaliana* = *Arabidopsis thaliana* (Chateigner-Boutin and Small, 2007; Bentolila et al., 2008);

*A. capillus-veneris* = *Adiantum capillus-veneris* (Wolf et al., 2004);

*I. engelmannii* = *Isoetes engelmannii* (Grewe et al. 2011);

*A. formosae* = *Anthoceros formosae* (Kugita et al. 2003);

*P. patens* = *Physcomitrella patens* (Miyata and Sugita 2004, Rüdinger et al. 2009);

*M. polymorpha* = *Marchantia polymorpha* (Oda et al. 1992);

*P. purpurea* = *Pleurozia purpurea* (Wang et al. 2009).

Il apparaît que la fréquence d'édition chez les transcrits des chloroplastes est inférieure à celle des transcrits mitochondriaux. Ceci peut être observé chez des plantes qui éditent leurs organites à une fréquence moyenne, comme *Arabidopsis thaliana* qui contient 34 sites dans les chloroplastes (Tillich et al., 2005) et plus de 500 sites dans les mitochondries (Bentolila et al., 2008; Zehrmann et al., 2008), mais aussi chez les plantes ayant une fréquence d'édition faible, comme *Physcomitrella patens* qui contient 2 sites d'édition dans les chloroplastes (Miyata and Sugita, 2004) et 11 sites dans les mitochondries (Rüdinger et al., 2009).

L'anthocérophyte *Anthoceros formosae* montre à ce jour le plus grand nombre de sites d'édition dans les chloroplastes avec 509 sites édités de C en U et 433 sites de U en C (Kugita et al., 2003). Au sein des anthocérophytes, l'étude du transcrit *nad5* démontre que cette division possède une haute fréquence d'édition également dans les mitochondries (Duff, 2006).

Plus de 100 sites d'édition sont observés dans les chloroplastes de plantes terrestres ayant divergé assez précocement tels *Anthoceros*, *Alsophila*, *Adiantum* ou *Selaginella* (Freyer et al., 1997; Kugita et al., 2003; Wolf et al., 2004; Tillich et al., 2006; Smith, 2009) en revanche, les angiospermes possèdent moins de sites d'édition chloroplastiques.

Il a été proposé que les transcrits chloroplastiques des angiospermes perdent leurs sites d'édition plus vite qu'ils n'en gagnent (Freyer et al., 1997; Tillich et al., 2006). Cette perte d'édition peut être reliée à une variation dans le taux de mutation de l'ADN, le taux de transition de C en T étant supérieur à celui de T en C (Sloan et al., 2010). L'inverse est observé pour les plantes ayant divergé précocement (Fujii and Small, 2011).

Des sites d'édition ont été comparés à d'autres C non édités. Il a été montré que les sites impliqués dans l'édition de l'ARN ont tendance à être modifiés en T plus rapidement que les C non édités (Fujii and Small, 2011). Ces mutations peuvent être dues à un phénomène de rétrotranscription de l'ARNm impliquant une réversion du site d'édition (Sloan et al., 2010) ou encore au coût de l'édition pour la cellule qui serait assez important pour impliquer une réversion du site d'édition (Fujii and Small, 2011). A noter que la réversion d'un site d'édition peut impliquer la perte du facteur d'édition correspondant (Hayes and Mulligan, 2011).

D'après ces observations, les transcrits chloroplastiques des angiospermes semblent éliminer au fur et à mesure les sites d'édition.

Les monilophytes, les lycophytes et les anthocérophytes ont de très fortes fréquences d'édition de C en U (Yoshinaga et al., 1996; Grewe et al., 2011). Il est intéressant de noter que ces plantes terrestres montrent aussi un taux très important d'édition inverse de U en C dans les transcrits des organites. Cette édition « inverse » proviendrait de l'ancêtre commun des trachéophytes et des anthocérophytes. Néanmoins, ce type d'édition est très rarement observé chez les spermatophytes (Chaw et al., 2008). *Selaginella*, un lycophyte sans aucun site d'édition inverse, est assez intéressante car, d'une part, elle est composée d'un très fort taux de GC dans le génome de ses organites et d'autre part, elle contient un très grand nombre de sites édités de C en U dans ses transcrits mitochondriaux (Hecht et al., 2011).

Jusqu'à maintenant, aucune édition inverse n'a pu être détectée chez les mousses ou les marchantiophytes (Rudinger et al., 2012). En général, les mousses ont une fréquence d'édition de C en U assez faible dans leurs transcrits mitochondriaux, à l'exception du genre *Takakia* – « la mousse impossible » au Japon (Rudinger et al., 2012). Ce genre n'est pas seulement une exception dans sa fréquence d'édition mais aussi dans sa morphologie et dans sa biochimie (Asakawa, 2004).

On ne peut parler d'évolution de l'édition des ARN sans parler des Marchantiophytes, une division proche du clade des plantes terrestres. Les Marchantiophytes sont composés de trois grands sous-groupes :

- les haplomitrides, composés des genres *Haplomitrium* (haut taux d'édition), *Apotreubia* (taux d'édition moyen) et *Treubia*, ayant un faible taux d'édition (Rudinger et al., 2012),

- les jungermanniales ayant un taux d'édition comparable à celui des angiospermes,

- les marchantiales qui sont les seules plantes terrestres qui n'éditent pas leurs ARN (Groth-Malonek et al., 2007; Salone et al., 2007; Rudinger et al., 2012).

Considérant les différentes tendances de réduction et d'expansion de l'édition chez les plantes terrestres, il est clair que l'évolution de l'édition de l'ARN semble être assez dynamique.

De plus, la distribution des sites d'édition dans les régions codantes étant bien supérieure aux régions non codantes, il a été suggéré que l'édition soit assujettie à une pression de sélection. En effet, si un site n'est pas nécessaire d'un point de vue fonctionnel, celui-ci sera sûrement perdu au cours du temps (Takenaka et al., 2008).

## II. Les facteurs d'édition chez les plantes

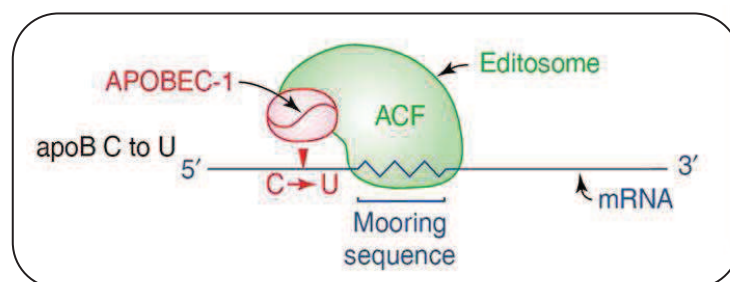
### II.1. Analogie entre le modèle APOBEC chez l'Homme et l'édition chez les plantes

Parmi les dizaines de milliers de pyrimidines dans les transcrits des organites, seulement quelques centaines sont éditées. La question posée ici est comment une telle précision peut-elle être obtenue et quels facteurs (*cis*- ou *trans*-) sont impliqués dans la reconnaissance du site puis son édition ?

#### II.1.1. Le modèle APOBEC

D'un point de vue biochimique, l'édition des organites des plantes est très similaire à l'édition identifiée dans les transcrits nucléaires des mammifères (Chen et al., 1987; Wedekind et al., 2003; Meier et al., 2005; Legendre et al., 2009), où des cytidines sont pareillement désaminées en uridines, l'exemple le plus connu chez les mammifères étant celui de l'édition de C en U du transcrit codant l'apolipoprotéine B (*apoB*) (Hospattankar et al., 1987). L'édition de ce transcrit repose sur l'interaction entre trois éléments majeurs que sont le site de reconnaissance situé dans l'ARN (élément en *cis*), le facteur de spécificité à l'ARN (élément en *trans*) et l'enzyme permettant de catalyser la réaction d'édition (élément en *trans* également).

La protéine APOBEC-1 a été identifiée comme étant l'enzyme permettant la désamination du site d'édition présent dans le transcrit *apoB*. Une séquence de 11 nt, la séquence d'ancrage ou « mooring », est reconnue par la protéine ACF (Wedekind et al., 2003). Les protéines ACF, permettant la reconnaissance du site à éditer, et dans une moindre mesure ASP, « stimulant » l'édition du transcrit, ont été identifiées comme étant impliquées dans l'édition d'*apoB* (Dance et al., 2002) (Figure 7).



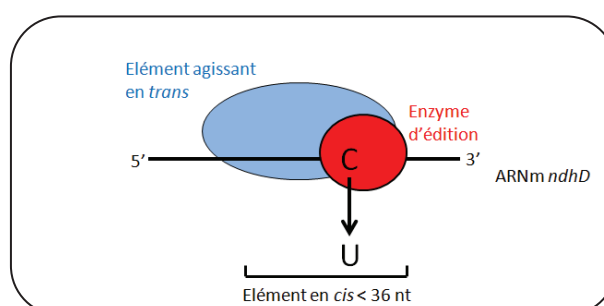
**Figure 7 : Mécanisme de l'édition du site *apoB* par APOBEC-1 (Adapté de Wedekind et al., 2003).** L'enzyme d'édition APOBEC-1 recrute un facteur de spécificité ACF pour reconnaître la séquence de « Mooring » présente dans l'ARNm *apoB*. Une désamination a alors lieu permettant la formation d'un codon STOP (UAA) chez la moitié des transcrits, permettant alors la traduction de deux protéines différentes APOB-48 et APOB-100 provenant du même gène.

Plusieurs observations suggèrent que le mécanisme d'édition chez les plantes serait plus proche de celui des mammifères que celui des kinétoplastes basés sur un système d'ARN guide (Blanc et al., 1995; Yu and Schuster, 1995). En effet, aucun nucléotide exogène n'est requis pour effectuer la réaction d'édition *in vitro*, la réaction semblant correspondre à une désamination de la cytidine (Araya et al., 1992).

## II.1.2. Des éléments en *cis* dans les transcrits des organites chez les plantes

Chez les plantes, les sites d'édition ne sont pas choisis au hasard. En effet, tout comme la séquence de « mooring » chez le transcrypt *apoB*, un élément *cis* est indispensable à la sélection du site. Depuis 1989, plusieurs études sur les éléments *cis* dans les chloroplastes (Chaudhuri et al., 1995; Bock et al., 1996; Chaudhuri and Maliga, 1996; Okuda et al., 2006; Tillich et al., 2006) et dans les mitochondries (Mulligan et al., 1999; Takenaka et al., 2004; Verbitskiy et al., 2008; Hammani et al., 2011) chez les angiospermes ont été publiées. Dans ces deux organites, 20 à 25 nucléotides en amont du site d'édition et généralement de 1 à 5 nucléotides en aval sont importants pour la reconnaissance du site d'édition. De plus, des régions en amont et en aval plus lointaines influencent l'efficacité de l'édition dans les chloroplastes et les mitochondries (Sasaki et al., 2006; Hayes and Hanson, 2007a). Chez le chou, 20 nucléotides en amont du site d'édition *atp9* sont suffisants pour l'édition mais l'efficacité « optimale » d'édition est atteinte lorsque la région en amont est étendue de 20 à 50 nucléotides (Neuwirt et al., 2005; van der Merwe et al., 2006). L'augmentation de l'efficacité par l'augmentation de la taille de la séquence en 5' a aussi été observée chez le transcrypt chloroplastique *rpoB* chez le tabac (Hayes et al., 2006). L'influence des séquences « éloignées » pourrait être expliquée par un changement éventuel de la conformation et de la structure du transcrypt, améliorant l'accessibilité au facteur agissant en *trans*. Les facteurs de stabilisation telles les ribonucléoprotéines influencent aussi l'efficacité de l'édition (Tillich et al., 2009). Néanmoins, il est possible que l'accumulation de protéines agissant en *trans* sur un site puisse gêner l'accessibilité à un autre site proche (Castandet and Araya, 2011). Il serait donc nécessaire que les protéines en *trans* agissent et se dissocient rapidement de la molécule d'ARN éditée pour éviter tout encombrement. Il a été montré pour les transcrits *psbE* et *petB* du tabac que les nucléotides les plus proches du site d'édition (-1 à -5) ainsi que la cytidine à éditer ne sont pas nécessaires à l'interaction mais sont indispensables à la déamination (Miyamoto et al., 2002, 2004).

Des expériences d'interaction *in vitro*, avec le transcrypt, *ndhD-1*, montrent que 20 nucléotides en amont et un peu moins de 10 nucléotides en aval sont suffisants à l'interaction et que le fait que la cytidine soit éditée ou non ne modifie pas l'interaction (Okuda et al., 2006) (Figure 8).



**Figure 8 : Modélisation complexe d'édition dans les plastides (adaptée de Okuda et al., 2006).**

Un modèle ressemblant à celui d'APOBEC-1 a été proposé. Un facteur de spécificité (en bleu) permet l'interaction avec l'ARN dans une région inférieure à 36 nt pour le site *ndhD-1*. L'enzyme d'édition (en rouge) recrutée permet la déamination de la cytidine.

En ce qui concerne le facteur d'édition Phypaff181369 chez *Physcomitrella patens*, cette protéine se fixe sur une sonde *ccmF<sub>C</sub>* de 46 nucléotides (40 en amont, 5 en aval) (Tasaki et al., 2010). Il est intéressant de noter que l'édition de ce site (position 122) est dépendante de l'édition d'un site se trouvant en amont (position 103).

### II.1.3. Des facteurs agissant en *trans* impliqués dans l'édition

L'existence d'un facteur agissant en *trans* a été évoquée pour la première fois chez les plastes du tabac (Chaudhuri et al., 1995). En effet, l'expression du transcrite contenant le site d'édition *psbL* a eu pour conséquence une diminution spécifique de l'édition de l'ARN endogène. Il a été proposé un effet de compétition entre l'ARN endogène et exogène pour un même facteur *trans* (Hirose and Sugiura, 2001). Il a été montré que le facteur agissant en *trans* n'était certainement pas un ARN comme l'ARNg du trypanosome (Hirose and Sugiura, 2001). Par contre, il semblait être plus probable qu'une protéine soit impliquée dans l'édition dans les transcrits chloroplastiques (Hirose and Sugiura, 2001; Miyamoto et al., 2002). En effet, des protéines ayant différents poids moléculaires interagissent très spécifiquement à la séquence en *cis* des sites d'édition *psbL*, *psbE* et *petB*. Il a donc été suggéré que le facteur *trans* soit une protéine plutôt qu'un ARN.

Le modèle proposé pour le complexe d'édition est ainsi très similaire à APOBEC (séquence en *cis* interagissant avec une protéine agissant en *trans*).

Une analyse du consensus d'éléments *cis* nécessitant des facteurs d'édition suggèrent que ces facteurs d'édition pourraient reconnaître spécifiquement la séquence d'intérêt (Hammani et al., 2009).

Les sites d'édition *psbZ*, *ndhB*, et *ndhF* sont reconnus par le même facteur. Une simple recherche de consensus donne une identité de 5 bases sur 16 (U-----UA-U----C, les tirets correspondant à des bases ambiguës) contenant donc peu d'information et correspondant à 444 accessions dans le génome (Hammani et al., 2009). Ainsi, un consensus de séquence est insuffisant pour expliquer la spécificité des facteurs d'édition à l'élément *cis*.

Néanmoins, il est connu que les nucléotides composants l'ARN ont tendance à interagir spécifiquement avec les protéines, soit par interaction avec le noyau aromatique, soit par le biais de liaisons hydrogène (Wang et al., 2002; Morozova et al., 2006). Il a été proposé que les facteurs d'édition puissent reconnaître :

- les purines (A et G) des pyrimidines (C et U),
- les bases suivant leur liaison hydrogène (G-C et A-U).

Un consensus plus poussé des sites d'édition *psbZ*, *ndhB*, et *ndhF* indiquant les bases ambiguës est représenté (UWRYWWYUAYUWYRYC ; W= A ou U, Y= C ou U, R= A ou G) simplement 4 fois dans le génome, une correspondant à une région non transcrite et les trois autres correspondant à *psbZ*, *ndhB*, et *ndhF* (Tableau 3). Il a été proposé que les facteurs en *trans* reconnaissent les purines des pyrimidines et sont capables de discriminer les nucléotides (Hammani et al., 2009).

**Tableau 3 : Consensus des sites d'édition nécessitant le même élément *trans* (adapté de Hammani et al., 2009).**

Element <i>trans</i>	Consensus final	Nb d'occurrences	Localisation
OTP82	UAGYU-YUYRGSWUC	2	<b><i>ndhB, ndhG</i></b>
CRR22	YW-YYYYARY-YYWYC	8	<i>psbK, ycf1, 3'UTR atpH, rpoB, ndhB, ndhD, ndhG, ndhA</i>
OTP84	UWRYWWYUAYUWYRYC	3	<b><i>psbZ, ndhB, ndhF</i></b>
CLB19	WYASA-GYSCAAR-UC	2	<b><i>clpP, rpoA</i></b>
CRR28	YWAUYUWUG-AGCU-C	2	<b><i>ndhB, ndhD</i></b>

Les consensus ont été produits à partir des 15 nucléotides se situant en amont des sites d'édition. Le consensus final provient de la combinaison entre les consensus suivants:

- conservation exacte des nucléotides,
- conservation des purines (A ou G = R) et des pyrimidines (U ou C = Y),
- liaisons hydrogènes : A ou U = W (2 liaisons) G ou C = S (3 liaisons).

Le consensus final est ensuite comparé aux séquences de transcrits. Les résultats positifs sont listés dans la colonne « localisation ».

En gras, les sites d'édition publiés pour être reconnus par les éléments en *trans*.

#### II.1.4. Quelle enzyme pour l'édition de l'ARN ?

Quelle enzyme est responsable de l'édition de l'ARN chez les plantes ? Des suppositions sont faites en l'absence de preuve formelle sur l'identité de l'enzyme (Takenaka et al., 2008). Cette enzyme pourrait être :

- une cytidine déaminase,
- une transaminase,
- une CTP synthase,

Les expériences d'édition *in vitro* utilisant des extraits de chloroplastes sont dépendantes du zinc car inactives si un chélateur est ajouté (Hegeman et al., 2005; Hayes and Hanson, 2007b). Or, les cytidines désaminases fixent un atome de zinc dans leur site catalytique lors de la désamination (Carter, 1995). De plus, lors des tests d'édition *in vitro*, il a été montré que ce processus procède par désamination de cytidines en uraciles dans les transcrits chloroplastiques et mitochondriaux, et non pas par remplacement de nucléotides (Yu and Schuster, 1995; Hirose and Sugiura, 2001). Ainsi, il est possible que l'enzyme responsable de l'édition dans les organites soit une cytidine déaminase. Néanmoins, l'expérience de chélation du zinc n'a pas fonctionné dans les extraits mitochondriaux. Nous pouvons supposer que le zinc puisse être difficile à extraire et ainsi que les chélateurs n'aient pas d'effet (Spath et al., 2007).

Il a été montré que les facteurs protéiques agissant en *trans* sont synthétisés en dehors des chloroplastes. En effet, des mutants albinos contenant des chloroplastes incapables



de synthétiser des protéines sont capables d'éditer les transcrits chloroplastiques (Halter et al., 2004). La même observation a été faite pour les mitochondries (Hinrichsen et al., 2009). Ces expériences indiquent que les facteurs d'édition sont codés dans le noyau.

Neuf cytidines désaminases putatives ont été prédites chez *Arabidopsis thaliana* et aucune d'entre elles ne se lie à l'ARN (Faivre-Nitschke et al., 1999). Par ailleurs, sept de ces cytidines déaminases semblent être localisées dans le cytosol (Faivre-Nitschke et al., 1999) et deux pourraient potentiellement être localisées dans les organites. Il est donc peu probable que ces enzymes agissent sur l'édition des ARN des organites. Chez *Physcomitrella patens*, trois cytidines déaminases sont présentes dans le génome nucléaire mais n'ont pas encore été étudiées (Mareike Rüdinger, communication personnelle). Ainsi, aucune cytidine déaminase, codée par le génome nucléaire d'*Arabidopsis thaliana* ou de *Physcomitrella patens*, n'a été à ce jour identifiée dans l'activité de déamination des ARN.

En 2005, le premier facteur agissant en *trans* a été identifié par génétique inverse (Kotera et al., 2005). Cette protéine, une PPR (pour Pentatricopeptide Repeat) appelée CRR4 cible un site d'édition très spécifique dans le transcrit *ndhD* d'*Arabidopsis thaliana*. A partir de cette observation, un grand nombre de facteurs d'édition ont été identifiés. Ils peuvent être classés en deux familles : les PPR et plus récemment, les protéines MORF (pour « Multiple Organellar RNA editing Factor ») aussi appelées RIP (pour « RNA-editing factor interacting protein »).

## II.2 Identification des facteurs d'édition

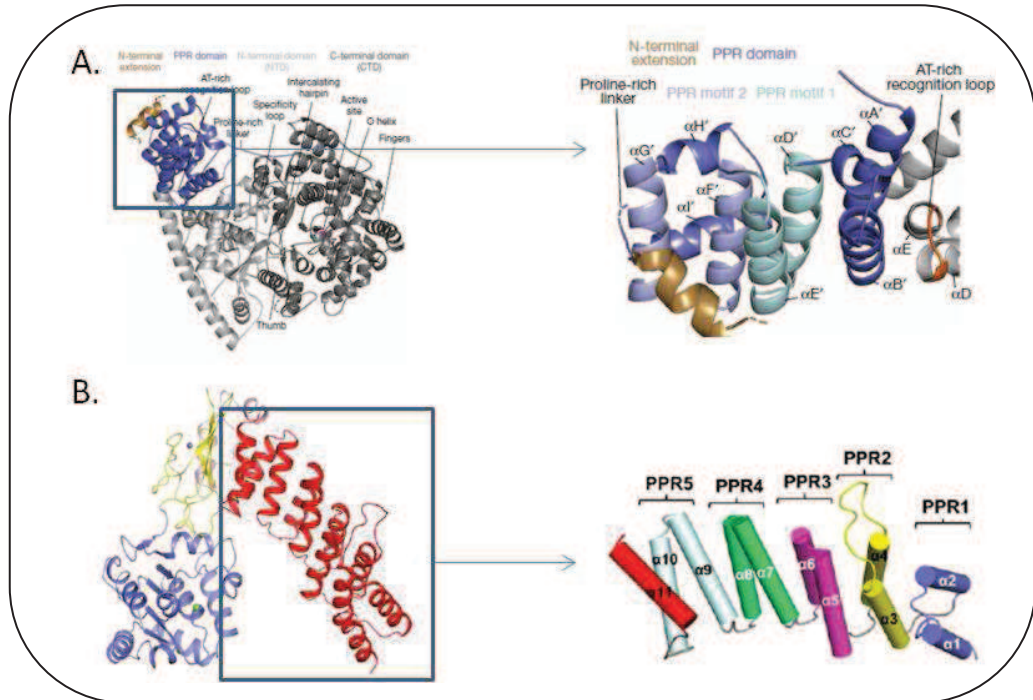
### II.2.1 La famille des PPR

#### (1) Composition et structure de la famille des protéines PPR

Le séquençage complet du génome de la plante modèle *Arabidopsis thaliana* en 2000 a permis la découverte de l'une des plus imposantes familles de protéines chez les plantes terrestres, la famille des « Pentatricopeptide Repeat » ou PPR (Aubourg et al., 2000; Small and Peeters, 2000). Les PPR sont des protéines caractérisées par des motifs de 35 acides aminés dégénérés répétés en tandem (de 2 à 26 fois) (Small and Peeters, 2000). Ces répétitions sont similaires à celles trouvées chez les protéines TPR (pour « Tetratricopeptide Repeat ») qui contiennent 34 acides aminés (Blatch and Lassle, 1999) et permettent l'interaction entre protéines. Les motifs PPR, tout comme les motifs TPR, ont été prédits comme étant constitués de deux hélices- $\alpha$  antiparallèles (Small and Peeters, 2000). Les PPR et les TPR forment une « super-hélice » constituant un sillon. Chez les motifs PPR, les chaînes proches du sillon central sont hydrophiles et positivement chargées ce qui leur permettrait d'interagir avec des éléments chargés négativement tels les acides nucléiques (Small and Peeters, 2000).

Les premières cristallisations d'un motif PPR ont été effectuées à partir d'une ARN polymérase humaine, mtRNAP (2 motifs PPR) (Figure 9A) et d'une ribonucléase chez *Arabidopsis thaliana* (5 motifs PPR et demi), PRORP1 (Ringel et al., 2011; Howard et

al., 2012) (Figure 9B). La structure du domaine PPR montre que les résidus conservés du motif PPR constituent le cœur hydrophobe d'une structure en hélice-tour-hélice.

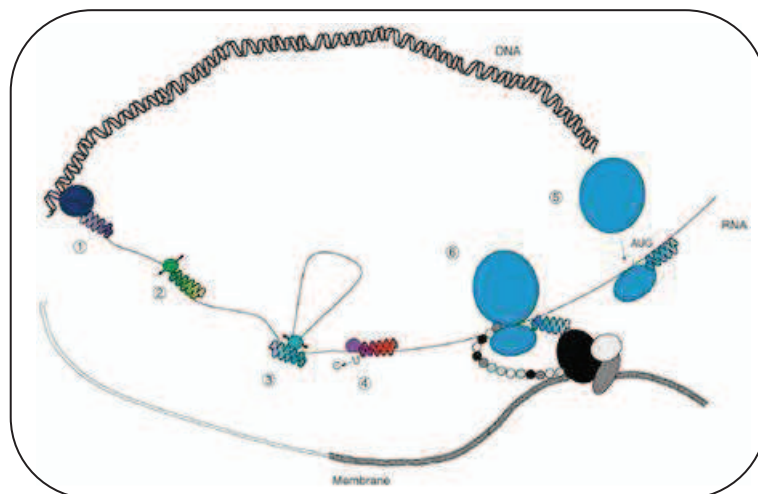


**Figure 9 : Structure des motifs PPR (adaptée de Ringel et al., 2011; Howard et al., 2012).**

A: structure de la protéine mtRNAP et de ses motifs PPR (en bleu). Deux motifs PPR ont été identifiés formant une structure en hélice-boucle-hélice antiparallèles.

B: structure de la protéine PRORP1 et de ses motifs PPR (en rouge). 5 motifs PPR et ½ sont identifiés formant une structure en hélice-boucle-hélice antiparallèles.

La grande majorité des PPR caractérisées sont impliquées dans des modifications post-transcriptionnelles des ARN comme le clivage, l'épissage, la stabilisation de l'ARN, le « processing » de la partie 5' de l'ARN ou encore l'édition dans les chloroplastes et les mitochondries des plantes terrestres (Fujii and Small, 2011) (Figure 10).



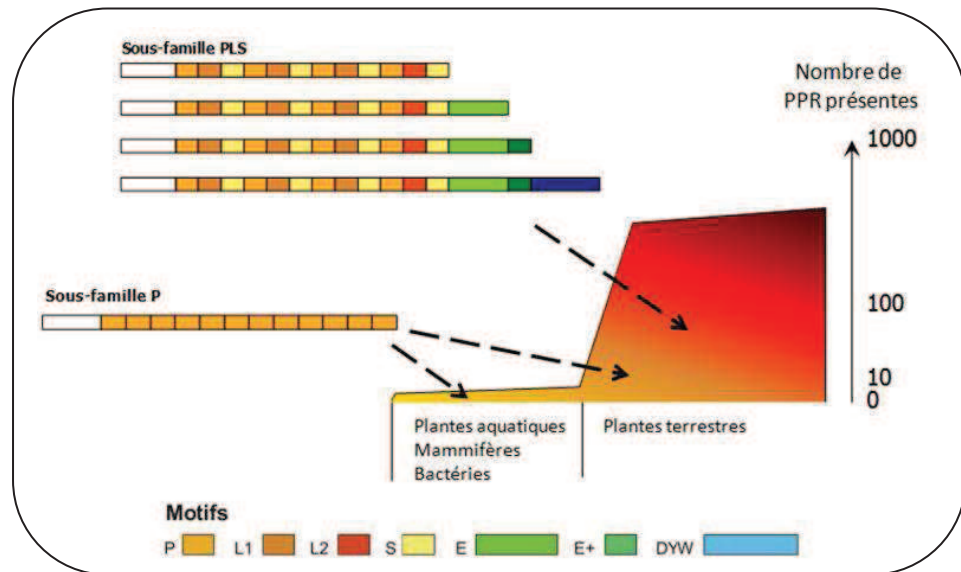
**Figure 10 : Fonctions connues des protéines PPR (Extrait de Andrès et al., 2006).**

Six rôles sont associés aux protéines PPR à l'heure actuelle.

- 1 : association à l'ARN polymérase,
- 2: clivage de l'ARN,
- 3: épissage de l'ARN,

- 4: édition de l'ARN,
- 5: initiation de la traduction,
- 6: régulation de la traduction.

Les PPR sont composées de deux sous familles : la sous famille P et la sous famille PLS (Figure 11).



**Figure 11 : Structure et expansion des protéines PPR chez les êtres vivants (adaptée de Lurin et al., 2004).**

Chez la sous-famille P, des motifs « Pure » de 35 aa exactement se succèdent. Cette sous-famille est la plus répandue et est observée chez un grand nombre d'êtres vivants.

La sous-famille PLS est composée de motifs PPR classiques mais aussi dégénérés L (Long) de 36 aa et S (Short) de 31 aa. En plus de leurs motifs PPR, cette sous-famille se divise en quatre sous-groupes caractérisés par leurs domaines C terminaux. Les PLS-P ne possèdent pas de domaine particulier, alors que les PLS-E portent un domaine E, les PLS-E+ un domaine E suivi d'un domaine E+ et enfin les DYW un domaine E suivi d'un domaine E+ puis d'un domaine DYW. Le nombre de protéines de cette sous-famille a explosé chez les plantes terrestres.

Les PPR-P possèdent uniquement des motifs de type P constitués de 35 acides aminés exactement. Ce motif est appelé « pure » car il s'agit du motif décrit à l'origine et qui a donné son nom à la famille PPR. Il est retrouvé aussi bien chez les plantes terrestres qu'en dehors.

Chez les champignons (*Saccharomyces cerevisiae* : 15 membres de la famille identifiés), les animaux (*Homo sapiens* : 6) et les algues (*Cyanidioschyzon merolae* : 10), la famille des PPR contient un faible nombre de membres toutes de la sous famille P (Manthey et al., 1998; Lurin et al., 2004; Lipinski et al., 2011). Le protozoaire *Trypanosoma brucei* contient 28 PPR-P différentes (Pusnik et al., 2007). Chez les procaryotes, quelques rares protéines PPR ont été identifiées (Salanoubat et al., 2002; Cazalet et al., 2010).

Chez les plantes terrestres, la famille s'est étendue et diversifiée (Lurin et al., 2004) (Figure 11). Chez *Arabidopsis thaliana*, la famille des protéines PPR est représentée par 450 membres. La grande majorité des PPR sont sans introns et distribués dans le génome sur les cinq chromosomes. Chez le monocotylédone *Oryza sativa*, la famille des protéines PPR consiste en 477 membres, alors que la plante modèle *Physcomitrella patens* (mousse) code 103 PPR dans son génome (O'Toole et al., 2008).

Mis à part la sous-famille P des PPR, une deuxième sous-famille est trouvée presque exclusivement chez les plantes terrestres, la sous-famille PLS (Aubourg et al., 2000; Lurin et al., 2004; Rivals et al., 2006; Andres et al., 2007). Elle a été très récemment identifiée chez le protiste *Naegleria gruberi* (Rudinger et al., 2011a). Ces protéines contiennent, en plus des motifs P habituels, des motifs L (long) et S (short) qui diffèrent dans la conservation et la longueur de leur séquence. En plus de ces nouveaux motifs, les PPR PLS contiennent différents domaines C-terminaux, le domaine E (~90 acides aminés), le domaine E+ (~30 acides aminés) et le domaine DYW (~100 acides aminés). Ces extensions se succèdent dans cet ordre (Figure 11).

## (2) Répartition des facteurs d'édition PPR-PLS dans le règne végétal

Ces dernières années, beaucoup de protéines PPR ont été identifiées comme étant des facteurs de spécificité nécessaires à l'édition des ARN à des sites spécifiques dans les chloroplastes et les mitochondries (Fujii and Small, 2011). Les 44 PPR publiées à ce jour comme étant impliquées dans l'édition sont toutes de type PLS, contenant des domaines C-terminaux caractéristiques, E et parfois DYW.

La majorité des membres de la sous-classe PLS pourrait être impliquée dans l'édition de l'ARN, si on prend en compte le grand nombre de mutants d'édition affecté dans l'expression de ces gènes.

Néanmoins, il est à noter trois exceptions connues de protéines de la sous-famille PLS pour lesquelles une fonction différente de l'édition a été montrée. La première la PPR-DYW CRR2 qui est impliquée dans le clivage d'un ARN dans les chloroplastes (Hashimoto et al., 2003; Okuda et al., 2009). La seconde, la PPR de la sous classe E OTP70 est impliquée dans l'épissage du transcrite chloroplastique *rpoC1* (Chateigner-Boutin et al., 2011). La troisième, la PPR-DYW PpPPR\_43 impliquée dans l'épissage du pré-ARNm *cox1* chez *Physcomitrella patens* (Ichinose et al., 2012).

Chez toutes les plantes terrestres étudiées éditant l'ARN, des PPR de type PLS/E/E+/DYW peuvent être identifiées (Salone et al., 2007; Rudinger et al., 2008). Le génome de *Physcomitrella* ne code que pour des PPR-PLS de type DYW et aucune PPR de type E ou E+ n'a été identifiée. Il a été proposé que les PPR-DYW seraient des facteurs d'édition anciens et que les autres protéines PPR avec des domaines C-terminaux plus court (E-E+) seraient plus récentes (Rudinger et al., 2011b). Chez les mousses et les marchantiophytes, la fréquence de l'édition des ARN est corrélée à la diversité des domaines C-terminaux des PPR (Rudinger et al., 2012). Chez les marchantiales, qui n'éditent pas les ARN, aucune extension E/E+/DYW n'a pu être observée chez les PPR (Rudinger et al., 2008). L'absence d'édition et de PPR de type E/E+/DYW pourrait être expliquée par une perte secondaire (Groth-Malonek et al., 2007; Rudinger et al., 2008). En effet, *Haplomitrium*, contenant un nombre important de sites d'édition et de protéine PPR-PLS, est un groupe très proche des marchantiales et des jungermanniales (Groth-Malonek et al., 2007).

Une question fascinante reste de savoir pourquoi (et comment) un groupe est capable de perdre toute la machinerie d'édition alors que d'autres plantes terrestres continuent à

l'étendre, comme par exemple *Selaginella moellendorffi* dont une cytidine sur cinq est changée en uridine dans le transcriptome mitochondrial (Hecht et al., 2011).

Il a été montré que le nombre de sites d'édition et le nombre de PPR sont corrélés (O'Toole et al., 2008). Les orthologues des gènes PPR de type DYW de *Physcomitrella patens* ont été identifiés chez *Funaria hygrometrica*, un groupe proche de *Physcomitrella* qui édite aussi ses ARN (Rudinger et al., 2011b). Néanmoins, chez *Funaria*, aucun orthologue du gène PPR de type PLS, Phypa\_169414, n'est observé. Les deux sites d'édition dépendants de ce facteur (*nad3* and *nad4*) sont aussi absents de *Funaria* montrant une corrélation entre la perte du site d'édition et la disparition du facteur d'édition (Rudinger et al., 2011b).

Chez *Oryza sativa*, des orthologues de PPR-PLS d'*Arabidopsis thaliana* peuvent être identifiés (O'Toole et al., 2008). De la même façon, un orthologue de la PPR AtCRR4 (Kotera et al. 2005) a été observé chez *Nicotiana benthamiana* et a été assigné au même site d'édition *ndhD-1* (Okuda et al., 2008). CRR4 est une PPR de type PLS ne présentant pas de domaine DYW. Il est possible que certains facteurs d'édition de l'ARN aient perdu leur domaine DYW pendant l'évolution chez les angiospermes. Néanmoins, lorsqu'il est présent, ce domaine est extrêmement conservé tant dans sa séquence que dans sa structure (Lurin et al., 2004).

Jusqu'à très récemment, aucune protéine PPR-DYW n'avait été découverte en dehors du règne végétal (Fujii and Small, 2011). Néanmoins, des gènes codant des PPR-DYW ainsi que des sites d'édition ont été découverts chez le protiste *Naegleria gruberi*. Les 1.5 milliards d'années de divergence entre les protistes et les plantes terrestres laissent à penser qu'il pourrait y avoir eu un transfert horizontal des facteurs d'édition ou que le phénomène d'édition soit encore plus ancien qu'on ne le pensait (Rudinger et al., 2011a).

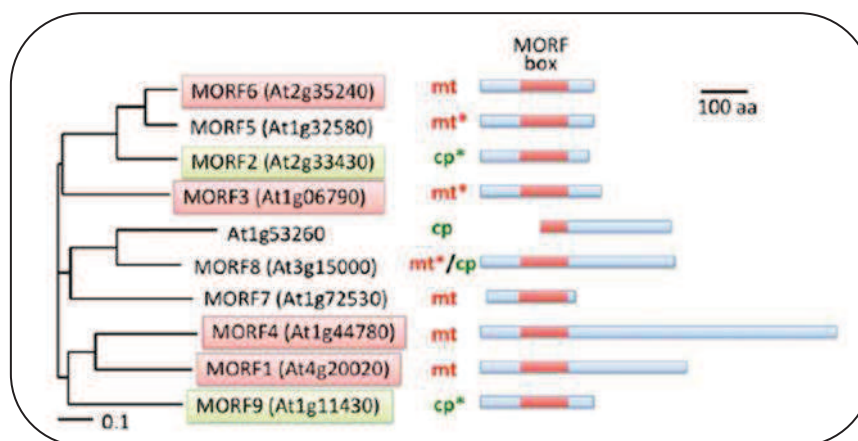
## II.2.2 Les protéines MORF

### (1) Identification et localisation des protéines MORF

En 2012, une nouvelle famille de protéines, appelée MORF, a été identifiée comme étant nécessaire à l'édition de nombreux sites dans les transcrits des organites chez *Arabidopsis thaliana* (Takenaka et al., 2012). A noter qu'une protéine de cette famille, MORF8, a été identifiée, dans le même temps, sous le nom de RIP1 (Bentolila et al., 2012).

La famille des protéines MORF, codée par des gènes nucléaires, est composée de neuf membres (notés MORF1 à MORF9) (Figure 12). Parmi ces membres, deux sont prédits comme étant localisés uniquement dans les chloroplastes (MORF2 et MORF9). Les autres protéines sont localisées dans les mitochondries (Takenaka et al., 2012). Il semblerait que MORF8 soit doublement localisée dans les chloroplastes et les mitochondries (Bentolila et al., 2012).

A l'heure actuelle, les protéines MORF ont été détectées chez les angiospermes uniquement et n'ont pas été identifiées chez la mousse *Physcomitrella patens*.



**Figure 12 : Cladogramme de similarité entre les différentes protéines MORF (Extrait de Takenaka et al., 2012).**

Les séquences des protéines MORF ont été comparées entre elles. Plus les protéines sont proches, plus elles seront proches sur le cladogramme. En rouge, la boîte MORF, une zone de 100 aa très bien conservée caractéristique des protéines MORF. Certaines localisations ont été confirmées par localisation GFP (seulement MORF2) ou par spectrométrie de masse (astérisques).  
 mt = localisation mitochondriale; cp = localisation chloroplastique.

## (2) Les protéines MORF sont nécessaires à l'édition de l'ARN

Le mutant EMS du gène *MORF1* modifiant une proline en sérine, *morfl-1*, montre une forte diminution de l'édition de plusieurs sites dans les mitochondries d'*Arabidopsis thaliana* (Takenaka et al., 2012). Après comparaison avec différents sites d'édition déjà étudiés auparavant, il a été montré que *morfl-1* affectait des sites déjà connus comme étant dépendant de protéines PPR. Néanmoins, le mutant *morfl-1* contient des défauts d'édition dans un grand nombre de sites dans les mitochondries (diminution dans 54 sites, augmentation dans 2 sites) alors que les mutants de PPR affectent rarement plus deux sites d'édition en même temps. Il est intéressant de noter que le mutant d'insertion ADN-T, *morfl-2*, n'est pas viable. En conclusion, la protéine MORF1 semble donc être vitale pour la plante.

A l'inverse du mutant *morfl-2*, le mutant d'insertion *morf3-1* est viable à l'état homozygote et montre un défaut d'édition dans les transcrits mitochondriaux dans plusieurs cas, majoritairement différents des sites retrouvés chez *morfl-1* (Takenaka et al., 2012).

De façon étonnante, des lignées d'insertion dans *MORF4* et *MORF6* ont montré une diminution de l'édition au niveau d'un site unique pour chacun (Takenaka et al., 2012). Néanmoins, lorsque les protéines MORF sont comparées en se basant sur les similarités de séquence, il est clair que MORF1 est similaire à MORF4 et MORF6 est similaire à MORF5. L'hypothèse posée est qu'il pourrait y avoir une certaine redondance de fonction entre les différentes protéines MORF (cette hypothèse n'étant pas valable pour MORF1 dont le mutant d'insertion est non-viable).

Dans le cas de MORF8 (RIP1), le mutant d'insertion montre un défaut d'édition dans des transcrits mitochondriaux (108 sites affectés) ainsi que dans des transcrits chloroplastiques (14 sites affectés), néanmoins le défaut dans les sites d'édition

chloroplastiques pourrait résulter d'un effet secondaire de la mutation (Bentolila et al., 2012).

Le taux d'édition des 34 sites chloroplastiques a été caractérisé chez les mutants *morf2-1* et *morf9-1* (Tableau 4). A peu près tous les sites sont affectés dans au moins un de ces deux mutants, l'édition étant perdue ou diminuée dans la plupart des cas. Si certains sites sont spécifiques de l'une des deux protéines, d'autres en revanche semblent nécessiter les deux protéines pour être édités (comme *ndhD-1*). Ainsi, il est clair que les protéines MORF2 et MORF9 sont essentielles à l'édition des transcrits chloroplastiques (Takenaka et al., 2012). Il est possible que, pour les sites montrant une légère réduction, une protéine MORF compense l'autre.

**Tableau 4 : Impact des mutations *morf2-1* et *morf9-1* sur l'éditotype d'*Arabidopsis thaliana* (adapté de Takenaka et al., 2012).**

Transcrit	Position du site	Col0	<i>morf2-1</i>	<i>morf9-1</i>	Transcrit	Position du site	Col0	<i>morf2-1</i>	<i>morf9-1</i>
<i>ndhB</i>	149	90%	0%	10%	<i>clpP</i>	559	80%	60%	40%
	467	90%	5%	15%	<i>accD</i>	794	89%	4%	4%
	586	100%	20%	10%		58642 3'UTR	20%	40%	40%
	746	100%	50%	50%	<i>matK</i>	706	70%	45%	50%
	830	100%	0%	50%	<i>ndhF</i>	290	100%	0%	40%
	836	100%	0%	0%	<i>ndhG</i>	50	100%	40%	0%
	872	85%	0%	0%	<i>petL</i>	5	100%	65%	0%
	1255	100%	10%	15%	<i>rpoB</i>	338	90%	70%	70%
	1481	100%	10%	10%		551	90%	80%	60%
				2432		100%	80%	80%	
<i>ndhD</i>	2	45%	0%	0%	<i>rps14</i>	80	100%	0%	50%
	383	90%	50%	40%		149	95%	50%	50%
	674	80%	50%	50%		<i>rpoA</i>	200	50%	80%
	878	80%	0%	10%	<i>psbZ</i>	50	100%	0%	30%
	887	100%	5%	10%	<i>rps12</i>	intron 69,553	20%	15%	10%
<i>psbE</i>	214	100%	30%	100%	<i>rpl23</i>	89	70%	0%	0%
<i>psbF</i>	65	100%	0%	70%					
<i>atpF</i>	92	100%	10%	10%					

Col0 = pourcentage d'édition observé dans la lignée sauvage Columbia pour le site donné.

*morf2-1* = pourcentage d'édition observé dans la lignée mutante du gène *MORF2*.

*morf9-1* = pourcentage d'édition observé dans la lignée mutante du gène *MORF9*.

### (3) Interaction MORF/MORF

Il a été proposé que les protéines MORF pourraient fonctionner en hétérodimères et dans certains cas en homodimères. Pour le tester, des tests d'interactions entre MORF1, MORF2, MORF3 et MORF9 ont été réalisées en double hybride chez la levure (Takenaka et al., 2012). Ainsi, il a ainsi été montré que MORF1, MORF2 et MORF9 peuvent interagir en homodimères dans ces conditions (Tableau 5).

De plus, les protéines chloroplastiques MORF2 et MORF9 interagissent dans ce système, tout comme les protéines MORF1 et MORF3. Ceci montre que ces protéines semblent avoir des affinités et ainsi former des hétérodimères. Etonnamment, la

protéine MORF1 semble être capable d'interagir avec MORF2 et MORF9 alors que ces protéines sont localisées dans des compartiments différents de la cellule. Il est possible qu'il y ait une certaine flexibilité de l'interaction entre les protéines MORF pouvant former différentes combinaisons dans un organe donné (Tableau 5).

Dans la mitochondrie, MORF1 et MORF3 ciblent moins de 100 sites à eux deux. Ainsi, il est probable que les autres protéines MORF mitochondriales, dont la fonction n'est pas identifiée encore, soient impliquées dans les 400 (environ) autres sites d'édition non affectés dans *morf1-1* et *morf3-1*. Pour les sites partiellement affectés ou non affectés, il est possible que les protéines se substituent l'une l'autre. Ceci pourrait être expliqué par le fait que les protéines MORF formeraient des homodimères et des hétérodimères plus ou moins efficaces selon que l'une ou l'autre protéine MORF soit mutée (Takenaka et al., 2012).

**Tableau 5 : Interaction des protéines MORF en homo- et hétérodimères (extrait de Takenaka et al., 2012).**

	Empty	MORF1	MORF3	MORF2	MORF9
MORF1	-	+++	+++	+++	+++
MORF3	-	+++	-	+	+
MORF2	-	+++	+	+++	+++
MORF9	-	+++	+	+++	+++

Expérience d'interaction menée en double hybride chez la levure pour quatre protéines MORF différentes. Les MORF de la colonne verticale représentent les appâts et horizontale les proies. « Empty » est le contrôle d'autoactivation. La qualité de l'interaction est représentée par :

- +++ : forte interaction,
- +: colonies dont la croissance est plus faible,
- : aucune colonie présente.

La protéine mitochondriale MORF1 interagit avec le panel entier de protéines même celles situées dans les chloroplastes.

MORF2 et MORF9 interagissent en homodimères et en hétérodimères.

MORF3 interagit avec MORF1. Aucune homodimérisation n'est observée.

#### (4) Interaction MORF/PPR

Dans les transcrits chloroplastiques et mitochondriaux, les sites d'édition sont affectés par les protéines PPR mais aussi par les protéines MORF. Dans les plastes, le site d'édition *ndhD-1* est dépendant, pour être édité, des protéines MORF2 et MORF9 (Takenaka et al., 2012) mais aussi de la PPR CRR4 (Kotera et al., 2005). Dans les mitochondries, MORF1 (Takenaka et al., 2012) et MEF19 (Takenaka et al., 2010) sont nécessaires à l'édition du transcrit *ccmB-556* et MORF1 (Takenaka et al., 2012) et MEF21 (Takenaka et al., 2010) sont indispensables à l'édition de *cox3-257*. Ceci indique que la fonction des protéines PPR et des protéines MORF, citées ci-dessus, est liée et « additionnelle ». En effet, les deux fonctions des PPR et des MORF ne se complètent pas et sont nécessaires à l'édition.

L'interaction entre les protéines PPR et les protéines MORF a été étudiée par double hybride chez la levure. La protéine MORF1 semble avoir une forte affinité pour MEF9

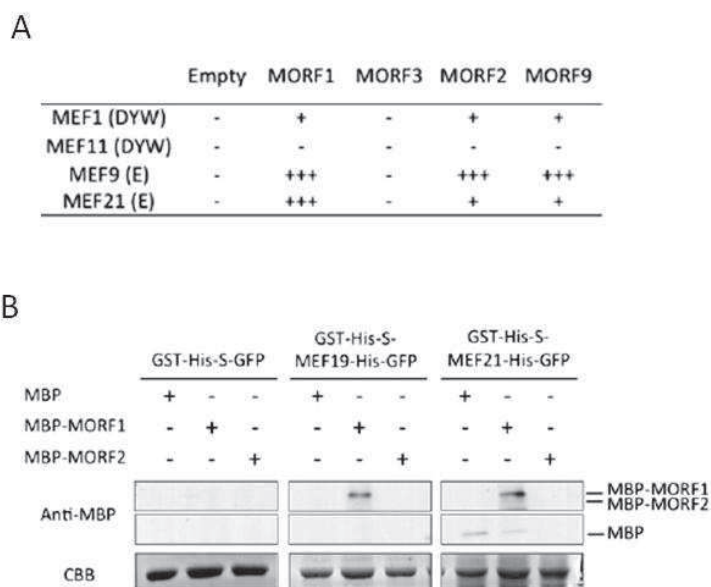


et MEF21, tandis que MORF2 et MORF9 interagissent avec MEF9 (Tableau 6). Aucune interaction n'a été observée par cette méthode pour MEF11.

Il est intéressant de noter l'interaction entre les protéines MEF19 et MORF1, ces deux protéines étant toutes deux nécessaires à l'édition du site *ccmB-556*.

L'interaction observée entre MORF1 et MEF1, MEF9, MEF21 montre une certaine spécificité de l'interaction entre les MORF et les PPR. Néanmoins, il a aussi été montré que les protéines MORF chloroplastiques interagissaient en double hybride avec la protéine mitochondriale MEF9. Les MORF pourraient donc potentiellement se lier à un grand nombre de protéines PPR de type PLS-E et DYW.

**Tableau 6 : Interactions entre protéines MORF et protéines PPR (Extrait de Takenaka et al., 2012).**



(A) Expérience d'interaction menée en double hybride chez la levure entre quatre protéines MORF différentes contre quatre protéines PPR. « Empty » est le contrôle d'autoactivation. La qualité de l'interaction est représentée par:

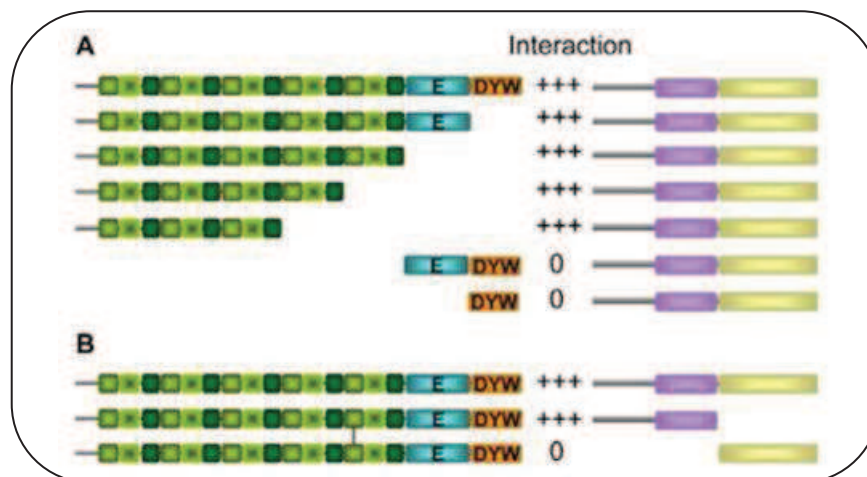
- +++ : forte interaction,
- +: colonies dont la croissance est plus faible,
- : aucune colonie présente.

Les protéines chloroplastiques MORF2 et MORF9 interagissent fortement avec la protéine mitochondriale MEF9. MORF1 interagit spécifiquement avec MEF9 et MEF21.

(B) Expérience d'interaction menée en co-immunoprécipitation. Les protéines marquées à la GST ont été fixées sur colonne de glutathione. Les protéines MORF marquées au MBP sont passées sur la colonne afin de tester si elles sont retenues par la PPR. L'interaction MEF21/ MORF1 est confirmée. Une nouvelle interaction est montrée ici entre MEF19 et MORF1.

La caractérisation de la protéine RIP1 (MORF8) est liée à l'étude de la protéine PPR RARE1 (Robbins et al., 2009). En effet, RIP1 a été identifiée suite à l'immunoprécipitation de RARE1, cette interaction a été confirmée par la suite en double hybride chez la levure. Pour identifier les domaines de RARE1 et RIP1 indispensables à l'interaction, ces deux protéines ont été tronquées et testées par double hybride. Ces expériences ont montré que quelques motifs PPR sont nécessaires à

l'interaction avec RIP1 et que la moitié de RIP1 en N-terminal est suffisante à l'interaction avec RARE1 (Bentolila et al., 2012) (Figure 13).



**Figure 13 : Des domaines spécifiques de RIP1 et RARE1 sont nécessaires à leur interaction *in vivo* (Extrait de Bentolila et al., 2012).**

Expérience d'interaction par double-hybride chez la levure menée sur différentes versions de la protéine PPR-DYW RARE1 et la protéine RIP1 (=MORF8).

(A) Protéine RARE1 tronquée en C-terminal puis en N-terminal. L'interaction semble avoir lieu au niveau des motifs PPR.

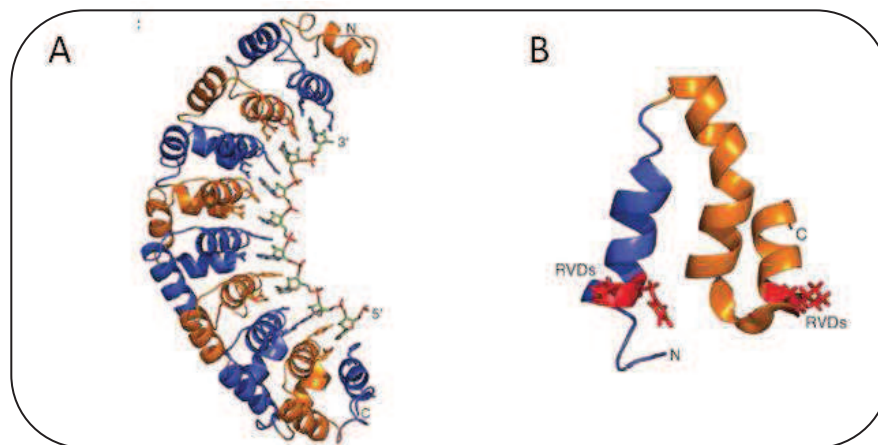
(B) Protéine RIP1 tronquée en C-terminal et en N-terminal. La boîte MORF semble être nécessaire à l'interaction PPR/MORF.

Ainsi, il semble que le modèle de base impliquant une PPR, un ARN et une enzyme doit être plus compliqué. La fonction des protéines MORF reste mystérieuse mais il a été montré que MORF3 et MORF8 ont une affinité avec les ions cobalt (Heazlewood et al., 2004). Il semble peu probable que ces protéines constituent l'enzyme permettant l'édition car elles sont absentes chez *Physcomitrella patens* (où l'édition est présente). De plus, aucun site catalytique ressemblant à celui découvert chez les cytidines déaminases n'a été mis en évidence dans la séquence des MORF.

### II.3. Interaction PPR-ARN : déchiffrement du code

Le challenge a été de découvrir les acides aminés nécessaires à l'interaction PPR-ARN et ainsi découvrir un code permettant de « déchiffrer » cette interaction. Récemment, les résidus 1, 3 et 6 ont été identifiés comme déterminants pour l'interaction impliquant les PPR-P (Fujii et al., 2011).

Auparavant les protéines TALE (Transcription-Activator-Like Effectors) et les protéines PUF (Pumilio and FBF homology), des protéines appartenant à la superfamille des solénoïdes à hélice alpha (en anglais « alpha-solenoides ») avaient été caractérisées (Figure 14). Ces protéines ont la particularité d'interagir avec des acides nucléiques (Filipovska and Rackham, 2012).



**Figure 14: Structure tridimensionnelle des motifs PUF et TALE (Extrait de Filipovska et al., 2012).**

(A) Les motifs PUF sont composés de trois hélices alpha permettant l'interaction avec l'ARN (représenté en vert et rouge).

(B) 1,5 motifs TALE sont représentés (0,5 en bleu + 1 en orange). Chaque répétition est composée de 2 hélices alpha. En rouge, les sites dont la composition en acides aminés varie (RVDs pour « repeat variable di-residues ») conférant la spécificité à la séquence d'ADN.

### II.3.1. Parallèle avec les protéines PUF

Les protéines PUF possèdent huit répétitions de 36 acides aminés en C-terminal nécessaires à l'interaction avec les ARNm de manière antiparallèle (ratio 1:1). Une expérience de cristallisation a permis de montrer que ces répétitions prenaient la forme d'un arc, chaque répétition étant composée de trois hélices alpha (Edwards et al., 2001) (Figure 14A). De plus, il a été montré que chaque répétition se fixait à une base spécifique dans l'ARNm par le biais des acides aminés 12 et 16 (Wang et al., 2002). Cette interaction est extrêmement spécifique (Tableau 7).

Le changement d'un seul acide aminé peut modifier l'interaction et permettre de cibler un autre transcrit (Cheong and Hall, 2006).

### II.3.2. Parallèle avec les protéines TALE

Les protéines TALE quand à elles contiennent des répétitions de 34 acides aminés nécessaires à l'interaction avec l'ADN de manière parallèle (ratio 1:1) (Boch and Bonas, 2010). Chaque répétition est composée de deux hélices alpha (Figure 14B). Il y a peu de variation dans les motifs TALE à l'exception des positions 12 et 13. Ces résidus sont en fait les acides aminés spécifiques à l'interaction avec l'ADN (Filipovska and Rackham, 2012). Les différentes bases sont reconnues par différentes combinaisons (Tableau 7).

Au contraire des protéines PUF chez lesquelles chaque paire de résidus reconnaît spécifiquement une base avec une grande affinité (comparée à une autre paire), certaines répétitions TALE semblent plus « flexibles » (Filipovska and Rackham, 2012). Par exemple, les répétitions contenant deux glutamines reconnaissent une adénine ou une guanine et les répétitions contenant une glutamine et une sérine reconnaissent les

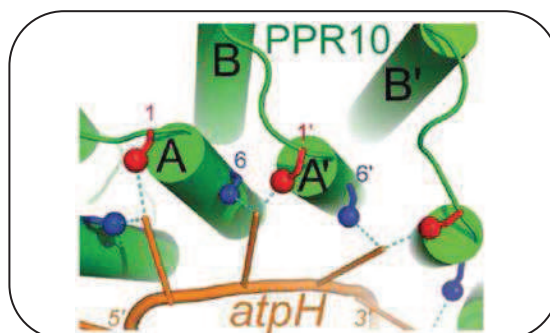
quatre bases. De façon intéressante, les protéines TALE semblent avoir des similitudes avec les protéines TPR. En effet, tout comme les TPR, les répétitions TALE sont composées de deux hélices alpha séparées par une boucle.

**Tableau 7 : Code d'interaction des protéines PUF, TALE et PPR (adapté de Filipovska et al., 2012; Barkan et al., 2012).**

<b>PUF/ARN</b>	<b>Position 12</b>	<b>Position 16</b>	<b>Base</b>
	Cystéine	Glutamine	Adénine
	Asparagine	Glutamine	Uracile
	Sérine	Glutamate	Guanine
	Glycine, alanine, sérine, thréonine ou cystéine	Arginine	Cytosine
<b>TALE/ADN</b>	<b>Position 12</b>	<b>Position 13</b>	<b>Base</b>
	Glutamine	Isoleucine	Adénine
	Glutamine	Glutamine	
	Glutamine	Sérine	
	Histidine	Aspartate	Cytosine
	Glutamine	Sérine	
	Glutamine	Lysine	Guanine
	Glutamine	Glutamine	
	Glutamine	Sérine	Thymine
	Glutamine	Glycine	
Glutamine	Sérine		
<b>PPR/ARN</b>	<b>Position 6</b>	<b>Position 1'</b>	<b>Base</b>
	Thréonine	Aspartate	Guanine
	Thréonine	Asparagine	Adénine
	Sérine	Asparagine	
	Asparagine	Aspartate	Uracile
	Asparagine	Asparagine	
Asparagine	Asparagine	Cytosine	

### II.3.3. Les protéines PPR

La protéine PPR10 a été étudiée dans le cadre de l'étude des interactions PPR-ARN (Barkan et al., 2012). Cette protéine est une PPR de type P interagissant avec *atpH*, pouvant former un dimère en absence d'ARN et un monomère en présence d'ARN (3 $\mu$ M PPR10 pour 3 $\mu$ M d'ARN). Dix-sept nucléotides d'*atpH* sont nécessaires à l'interaction avec PPR10. Une modélisation a été effectuée de telle sorte que l'ARN soit parallèle à PPR10 (5' terminal s'alignant avec l'extrémité N-terminal) et que chaque motif PPR puisse contacter une base ARN. Une forte corrélation a été montrée entre l'ARN et les acides aminés se trouvant aux positions 1 et 6 (Figure 15).



**Figure 15 : Modèle structural montrant l'interaction entre PPR10 et *atpH* (Extrait de Barkan et al., 2012).**

Les résidus 6 et 1', respectivement des hélices A et A', de PPR 10 interagissent avec une base d'*atpH*. Chaque motif est composé de deux hélices nommées A et B.

Un alignement a permis d'identifier que toutes les asparagines en position 6 (N<sub>6</sub>) s'alignent avec des pyrimidines, chaque purine s'aligne avec une sérine et une thréonine en position 6 (S<sub>6</sub> et T<sub>6</sub> respectivement) et que tous les aspartates en position 1 (D<sub>1</sub>) s'alignent avec une uridine. Ceci a été confirmé par l'étude de HCF152 et CRP1 deux PPR du sous-groupe P.

Le modèle semble également fonctionner pour d'autres PPR telles que des PPR-PLS, par exemple OTP82, CRR22 et CRR4 (Barkan et al., 2012). Des séquences construites à partir de paires de motifs PPR ont été alignées avec les bases A, G, C et U et ont montré qu'un code de reconnaissance pouvait être établi (Tableau 7).

Ce code a été confirmé expérimentalement par gel retard pour PPR10 (Barkan et al., 2012). En effet, les motifs 6 et 7 de PPR10 ont été modifiés afin qu'ils puissent reconnaître de nouvelles bases. Alors que PPR10 sauvage (répétition 6=ND / répétition 7=NN) reconnaît une uridine et une cytidine, les mutants de PPR10 (respectivement répétition 6 et répétition 7, en position 6 et 1') peuvent reconnaître d'autres bases (Tableau 7).

Ces résultats montrent qu'il devrait être possible de modifier une protéine PPR afin qu'elle interagisse avec de nouvelles cibles. Néanmoins, le code reste relativement dégénéré lorsqu'il est comparé à celui des protéines TALE, l'alignement n'étant pas parfait d'une PPR à une autre. Chez les PPR-P, des trous peuvent être observés dans l'alignement du duplex PPR/ARN dans les cas de PPR10, CRP1 et HCF152. Cet événement a aussi été observé dans le duplex PUF/ARN. Cependant, aucun trou n'a encore été observé concernant les duplex PPR-PLS/ARN mais il a été proposé que les motifs L n'interagissent pas avec les acides nucléiques et qu'ils permettent la formation d'un trou (« mini-gap ») tous les trois nucléotides permettant un relâchement de la structure.

Le modèle d'interaction spécifié par Barkan et al (2012) consiste en un duplex de 9 motifs PPR et de 8 acides nucléiques ce qui est très proche de ce que l'on peut trouver chez les protéines PUF (Barkan et al., 2012).

## II.4. Les PPR contiennent-elles une activité cytidine déaminase ?

Les cytidines déaminases connues possèdent généralement un site de fixation au zinc ainsi qu'une structure secondaire en hélice/feuillet/hélice.

Les motifs PPR étant impliqués dans l'interaction spécifique avec l'ARN par le biais des acides aminés en position 6 et 1', il ne reste que les domaines E/E+ et le domaine DYW comme candidats potentiels à une activité cytidine déaminase directement contenue par les protéines PPR.

### II.4.1. Implication du domaine E/E+ ?

CRR4 le premier facteur d'édition identifié dans les chloroplastes est une PPR-PLS du sous-groupe E+ ne possédant pas de domaine DYW (Kotera et al., 2005) tout comme CRR21 responsable de l'édition du site *ndhD-2* (Okuda et al., 2007).

Ni le domaine E ni le domaine E+ ne semblent contenir d'activité catalytique. En effet, ces domaines étant assez variables en taille et en séquence, il est peu probable qu'ils soient associés à une fonction nécessitant une forte conservation de séquence (Figure 16). De plus, ce domaine ne contient pas la signature HxE(x)<sub>n</sub>CxxC à laquelle on s'attend pour expliquer une activité cytidine déaminase.

Ainsi, il a été avancé que CRR4 et CRR21 seraient des facteurs de spécificité et qu'un autre facteur serait nécessaire à la modification de la cytidine en uridine (Okuda et al., 2007). Il a été montré que le domaine E/E+ est essentiel à l'édition de l'ARN chez toutes les protéines PPR testées mais ne possède a priori pas d'activité catalytique (Okuda et al., 2007; Okuda et al., 2010; Takenaka et al., 2010). La nouvelle hypothèse proposée avance que le domaine E/E+ permettrait de recruter une cytidine déaminase.

```

E
1  VWgaLLgACrCriHgnvnellaeaAakkLl 30
31  eLePendgsYeeGensGnYVLLSNiYAsaG 60
61  rWeevakvRklMkdrekGvkKePGcSWIEv 90
91  k

E+
1  GkvHeFvaGDkkeSHPeseeIyekLeeLie 30
31  emk 33

```

**Figure 16 : Séquence consensus des domaines E et E+ (Lurin et al., 2004).**

Les consensus des domaines E et E+ ont été obtenus par le biais de l'alignement de 184 et 148 séquences, respectivement. Les acides aminés bien conservés sont indiqués en lettres capitales. Aucun acide aminé invariant n'a pu être identifié.

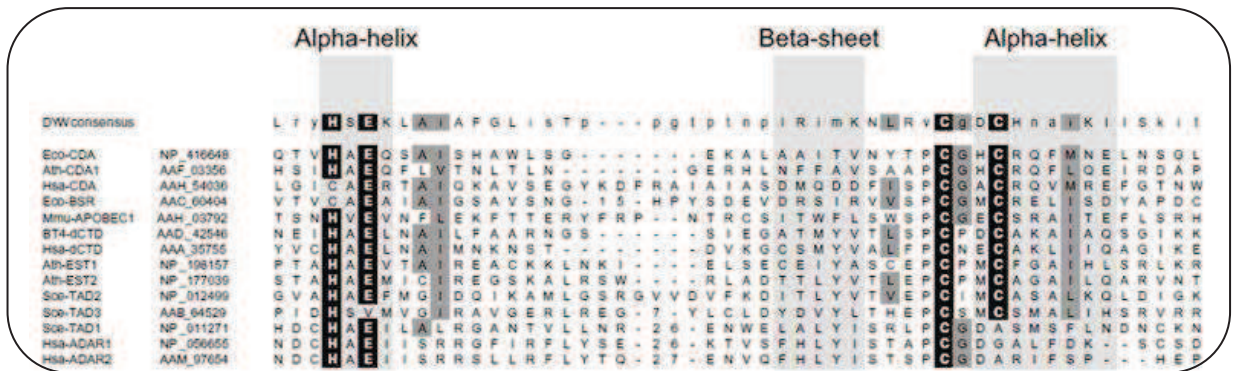
#### II.4.2. Implication du domaine DYW ?

Le domaine DYW a été nommé ainsi à cause du triplet d'acides aminés final le composant (aspartate-tyrosine-tryptophane), ces derniers étant extrêmement conservés. La très forte conservation des acides aminés de ce domaine suggère une fonction catalytique (Aubourg et al., 2000). Le domaine DYW des PPR-PLS semble avoir des similarités avec des cytidines déaminases connues dans :

- sa structure secondaire en hélice-feuillet-hélice (Salone et al., 2007) (Figure 17),

- sa séquence primaire avec la conservation d'acides aminés HxE(x)<sub>n</sub>CxxC (Figure 17). Tandis que les résidus histidine et cystéines sont impliqués dans la fixation de zinc, le glutamate est indispensable à la catalyse de la réaction d'édition chez les cytidines déaminases (Salone et al., 2007).

De plus, la présence des domaines DYW est corrélée avec la distribution de l'édition de l'ARN chez les embryophytes.



**Figure 17 : Alignement entre une partie du consensus du domaine DYW et la séquence de cytidines déaminases connues (Extrait de Salone et al., 2007).**

Le consensus DYW a été obtenu à partir de l'alignement de 87 domaines DYW. Ce consensus a été aligné avec 14 cytidines déaminases ou enzymes d'édition de l'ARN. Les acides aminés impliqués dans la fonction d'édition (signature CDA HxE(x)nCxxC) sont marqués en noir. Les lettres capitales montrent les acides aminés les mieux conservés dans le consensus DYW.

La structure secondaire est conservée entre le domaine DYW et les CDA. De plus, les acides aminés indispensables à l'activité catalytique de déamination sont très bien conservés.

La corrélation phylogénétique entre domaine DYW et édition de l'ARN, la similarité de séquence entre le domaine DYW et les cytidines déaminases et l'absence de candidats pour contenir la fonction d'édition chez les plantes terrestres sont les raisons pour lesquelles la communauté scientifique s'est focalisée sur les protéines PPR-PLS de type DYW pour étudier l'édition de l'ARN chez les plantes (Zhou et al., 2008; Cai et al., 2009; Hammani et al., 2009; Kim et al., 2009; Okuda et al., 2009; Robbins et al., 2009; Verbitskiy et al., 2009; Zehrmann et al., 2009; Bentolila et al., 2010; Ohtani et al., 2010; Takenaka et al., 2010; Tasaki et al., 2010; Tseng et al., 2010; Verbitskiy et al., 2010; Uchida et al., 2011; Zehrmann et al., 2011; Sosso et al., 2012) (Tableau 8).

**Tableau 8 : PPR-DYW (publiées) impliquées dans l'édition chez les plantes terrestres.**

Mutant	Transcrits affectés	Localisation	Référence	Organisme
<i>lpa66</i>	<i>psbF</i>	Chloroplaste	Cai et al., 2009	<i>A. thaliana</i>
<i>otp81</i>	<i>rps12</i>	Chloroplaste	Hammani et al., 2009	<i>A. thaliana</i>
<i>otp84</i>	<i>ndhB, ndhF, psbZ</i>	Chloroplaste	Hammani et al., 2009	<i>A. thaliana</i>
<i>otp85</i>	<i>ndhD</i>	Chloroplaste	Hammani et al., 2009	<i>A. thaliana</i>
<i>otp86</i>	<i>rps14</i>	Chloroplaste	Hammani et al., 2009	<i>A. thaliana</i>
<i>ogr1</i>	<i>nad2, ccmFc, cox2, cox3, cox4</i>	Mitochondrie	Kim et al., 2009	<i>O. sativa</i>
<i>otp82</i>	<i>ndhG, ndhB</i>	Chloroplaste	Okuda et al., 2009	<i>A. thaliana</i>
<i>crr28</i>	<i>ndhB, ndhD</i>	Chloroplaste	Okuda et al., 2009	<i>A. thaliana</i>
<i>crr22</i>	<i>ndhB, ndhD, rpoB</i>	Chloroplaste	Okuda et al., 2009	<i>A. thaliana</i>
<i>rare1</i>	<i>accD</i>	Chloroplaste	Robbins et al., 2009	<i>A. thaliana</i>
<i>mef11</i>	<i>cox3 et nad4</i>	Mitochondrie	Verbitskiy et al., 2009	<i>A. thaliana</i>
<i>mef1</i>	<i>Rps4, nad7, nad2</i>	Mitochondrie	Zehrmann et al., 2009	<i>A. thaliana</i>
<i>ys1</i>	<i>rpoB</i>	Chloroplaste	Zhou et al., 2009	<i>A. thaliana</i>
<i>reme</i>	<i>nad2-558</i>	Mitochondrie	Bentolila et al., 2010	<i>A. thaliana</i>
<i>ppr77</i>	<i>cox2 and cox3</i>	Mitochondrie	Ohtani et al., 2010	<i>P. patens</i>
<i>ppr56</i>	<i>nad3 and nad4</i>	Mitochondrie	Ohtani et al., 2010	<i>P. patens</i>
<i>ppr91</i>	<i>nad5</i>	Mitochondrie	Ohtani et al., 2010	<i>P. patens</i>
<i>mef22</i>	<i>nad3</i>	Mitochondrie	Takenaka et al., 2010	<i>A. thaliana</i>
<i>ppr71</i>	<i>ccmFc</i>	Mitochondrie	Tasaki et al., 2010	<i>P. patens</i>
<i>vac1</i>	<i>accD</i>	Chloroplaste	Tseng et al., 2010	<i>A. thaliana</i>
<i>mef14</i>	<i>matR</i>	Mitochondrie	Verbitskiy et al., 2010	<i>A. thaliana</i>
<i>ppr78</i>	<i>rps14 et cox1</i>	Mitochondrie	Uchida et al., 2011	<i>P. patens</i>
<i>ppr79</i>	<i>nad5</i>	Mitochondrie	Uchida et al., 2011	<i>P. patens</i>
<i>mef7</i>	<i>cob, ccb, nad2, nad4L</i>	Mitochondrie	Zehrmann et al., 2011	<i>A. thaliana</i>
<i>ppr2263</i>	<i>nad5 and cob</i>	Mitochondrie	Sosso et al., 2012	<i>Z. mays</i>
<i>mef8</i>	<i>nad5 and nad6</i>	Mitochondrie	Verbitskiy et al., 2012	<i>A. thaliana</i>
<i>mef8S</i>	<i>nad5 and nad6</i>	Mitochondrie	Verbitskiy et al., 2012	<i>A. thaliana</i>

Toutefois, la responsabilité du domaine DYW dans la réaction d'édition est, toujours aujourd'hui, sujette à beaucoup d'interrogations. En effet, plusieurs points vont à l'inverse de cette hypothèse :

(1) L'implication du domaine DYW n'a jamais été confirmé par réaction d'édition *in vitro* ou *in vivo* (Nakamura and Sugita, 2008; Okuda et al., 2009).

(2) Des protéines PPR ne contenant pas de domaine DYW sont nécessaires à l'édition. En effet parmi les 44 PPR connues pour être impliqués dans l'édition, 15 s'avèrent être des PPR de type E/E+ (Kotera et al., 2005; Okuda et al., 2007; Chateigner-Boutin et al., 2008; Hammani et al., 2009; Sung et al., 2010; Takenaka et al., 2010; Hammani et al., 2011; Verbitskiy et al., 2012a; Yuan and Liu, 2012) (Tableau 9).

(3) Et surtout, des études montrent que le domaine DYW de certaines PPR peut être délété sans aucune conséquence sur leurs fonction (Okuda et al., 2009; Okuda et al., 2010). En effet, des constructions contenant les gènes codant les PPR CRR22 et CRR28, dont le domaine DYW a été préalablement éliminé, complètent les mutants *crr22-1* et *crr28-1*.

(4) Les PPR-DYW CRR2 et PpPPR\_43 ne sont pas impliqués dans l'édition chez les plantes (Hashimoto et al., 2003; Ichinose et al., 2012). Dans le cas de CRR2, le domaine DYW est indispensable à la maturation du transcrit *ndhB* (Okuda et al., 2009).

**Tableau 9 : PPR-E/EE+ (publiées) impliquées dans l'édition chez les plantes terrestres.**

Mutant	Transcrits affectés	Localisation	Référence	Organisme
<i>crr4</i>	<i>ndhD</i>	Chloroplaste	Kotera et al., 2005	<i>A. thaliana</i>
<i>crr21</i>	<i>ndhD</i>	Chloroplaste	Okuda et al., 2007	<i>A. thaliana</i>
<i>clb19</i>	<i>rpoA</i> et <i>clpP</i>	Chloroplaste	Chateigner-Boutin et al., 2008	<i>A. thaliana</i>
<i>otp80</i>	<i>rpl23</i>	Chloroplaste	Hammani et al., 2009	<i>A. thaliana</i>
<i>slo1</i>	<i>nad4</i> et <i>Nad9</i>	Mitochondrie	Sung et al., 2010	<i>A. thaliana</i>
<i>mef18</i>	<i>nad4</i>	Mitochondrie	Takenaka et al., 2010	<i>A. thaliana</i>
<i>mef19</i>	<i>ccb206</i>	Mitochondrie	Takenaka et al., 2010	<i>A. thaliana</i>
<i>mef20</i>	<i>rps4</i>	Mitochondrie	Takenaka et al., 2010	<i>A. thaliana</i>
<i>mef21</i>	<i>cox3</i>	Mitochondrie	Takenaka et al., 2010	<i>A. thaliana</i>
<i>mef9</i>	<i>nad7</i>	Mitochondrie	Takenaka et al., 2010	<i>A. thaliana</i>
<i>otp87</i>	<i>nad7</i> et <i>atp1</i>	Mitochondrie	Hammani et al., 2011	<i>A. thaliana</i>
<i>slg1</i>	<i>nad3</i>	Mitochondrie	Yuan et al., 2011	<i>A. thaliana</i>
<i>ahg11</i>	<i>nad4</i>	Mitochondrie	Murayama et al., 2012	<i>A. thaliana</i>
<i>mpr25</i>	<i>nad5</i>	Mitochondrie	Toda et al., 2012	<i>O. sativa</i>
<i>mef3</i>	<i>atp4</i>	Mitochondrie	Verbitskiy et al., 2012	<i>A. thaliana</i>
<i>slo2</i>	<i>mttB</i> , <i>nad1</i> , <i>na4L</i> et <i>nad7</i>	Mitochondrie	Zhu et al., 2012	<i>A. thaliana</i>

A l'inverse des observations rapportées ci-dessus, d'autres études menées chez *Arabidopsis thaliana* ont montré que certaines PPR-DYW perdent leur fonction lorsque le domaine DYW est éliminé ou muté à des positions bien spécifiques (Zehrmann et al., 2010, 2011). Il semble donc que dans certains cas uniquement, la perte du domaine DYW soit sans effet et puisse être complétée directement par la cellule végétale.

De plus, chez *Physcomitrella patens* où 13 sites doivent être édités, aucune PPR de type E/E+ n'est détectée (O'Toole et al., 2008). Seules 10 PPR-PLS, toutes du sous-groupe DYW, sont présentes (Rudinger et al., 2009). Le domaine DYW semble être le candidat le plus probable pour catalyser la réaction d'édition chez *Physcomitrella patens*.



### III. Projet de thèse

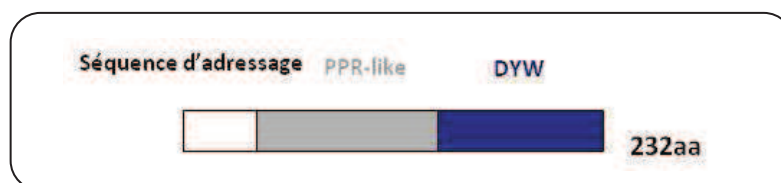
Véronique Salone, étudiante en thèse en cotutelle entre l'UWA (ARC-PEB, équipe de Ian Small) et l'UEVE (INRA-URGV, équipe de Claire Lurin) de 2004 à 2008, a étudié le domaine DYW des protéines PPR et a proposé qu'il serait impliqué dans la réaction d'édition (Salone et al., 2007).

Un gène dans le génome d'*Arabidopsis thaliana* consiste exclusivement en un domaine DYW sans domaine E ou E+, il s'agit de At1g47580 pour lequel aucune fonction n'a été proposée (Lurin et al., 2004).

Comme décrit dans l'introduction, nous savions que plusieurs PPR, toutes de type PLS, étaient impliquées dans l'édition de l'ARN. Parmi celles-ci, nous retrouvions des PPR du sous groupe E, sans motif DYW. Ces PPR-E reconnaissent un (des) site(s) d'édition bien spécifique et agissent en *trans*. Mais, comme le suggèrent plusieurs études, ces PPR ne semblaient pas avoir une activité cytidine déaminase, qui aurait pu être portée par le domaine DYW des PPR-DYW.

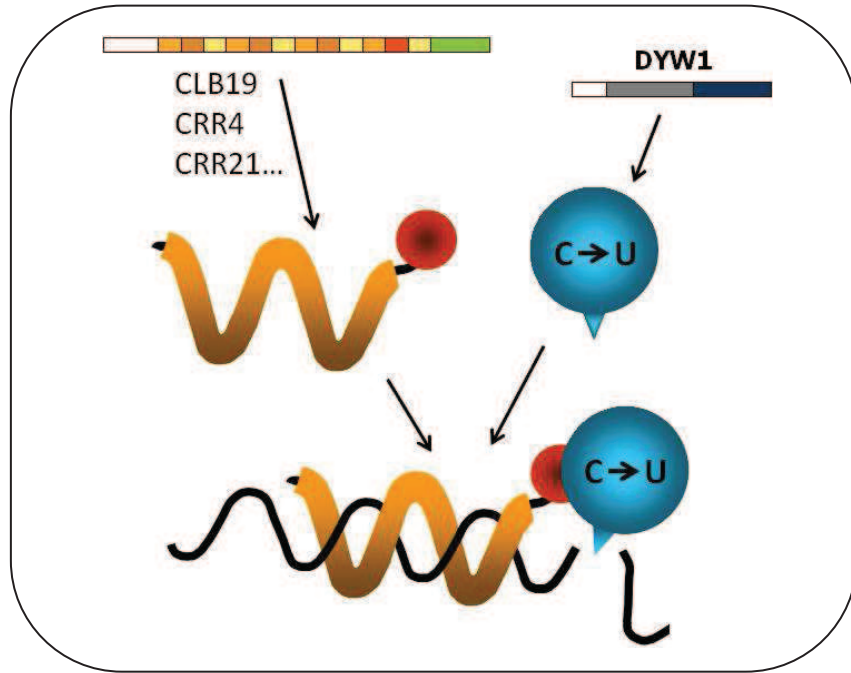
L'hypothèse posée était que les PPR-E interagiraient avec des protéines contenant une activité catalytique. At1g47580, appelée DYW1, semblait être le candidat idéal pour tester cette possibilité. En effet, cette dernière est composée d'une séquence d'adressage, d'un domaine assez éloigné des PPR (« PPR-like ») ainsi que d'un domaine DYW (Figure 18).

Le domaine PPR-like ne semblant pas suffisant à l'interaction protéine-ARN, DYW1 ne paraissait pas être en mesure d'éditer un site par lui-même. Il a été proposé que DYW1 puisse interagir avec une (ou plusieurs) PPR de type E, la PPR étant le facteur de spécificité et DYW1 donnant l'activité enzymatique au complexe (Figure 19).



**Figure 18 : Structure de la protéine AtDYW1.**

AtDYW1 est composée d'une séquence d'adressage, d'un domaine « PPR-like » ne contenant pas de vrai motif PPR et d'un domaine DYW.



**Figure 19 : Hypothèse de ma thèse : Interaction PPR EE+ / DYW1.**

Les PPR-E/E+ ne contenant pas de domaine DYW, donc pas de potentielle activité cytidine déaminase, pourraient recruter une protéine apportant le domaine DYW (contenant l'activité d'édition?). DYW1, ne possédant pas de motifs PPR mais un domaine DYW, semble être un candidat idéal pour tester cette hypothèse.

## Chapitre II : Résultats

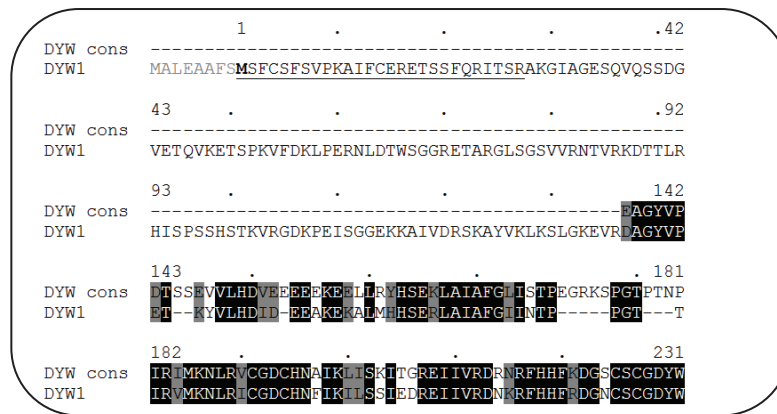
### I. Etude génétique du candidat AtDYW1

#### I.1. DYW1 : une étrange PPR

##### I.1.1. Annotation du gène DYW1

Si l'on suivait l'annotation AGI (v8) disponible au début de ma thèse, le gène At1g47580 codait pour une lipoyltransférase (ORF de 1251 pb). Néanmoins, les prédictions faites par Eugene (Foissac et al., 2003) proposaient que ce locus pourrait contenir deux gènes bien distincts, le premier ayant une homologie avec une lipoyltransférase et le second étant un gène codant une protéine contenant un domaine DYW. En 2011, une nouvelle version de l'annotation AGI (v9) est sortie incluant un changement dans l'ORF At1g47580. Depuis, At1g47580v8 est séparé en 2 gènes distincts : At1g47580v9, le gène codant pour un domaine DYW et At1g47590v9, un pseudogène étant possiblement le résultat d'un événement de duplication d'un exon de LIP2. Selon le TAIR, At1g47580 serait une protéine PPR. Cependant, lorsque nous soumettons la séquence dans le logiciel de prédiction de PPR/TPR du Max Planck Institute (<http://toolkit.tuebingen.mpg.de/tpred>), aucun motif PPR n'est identifié. Cette protéine ne contient en fait qu'une séquence d'adressage, une région qui a des similarités en Blast avec des motifs PPR (nommée PPR-like) et un domaine DYW. C'est ainsi que le gène At1g47580 a été baptisé AtDYW1.

Selon les premières observations, le gène DYW1 comprendrait 717 pb, le transcrit serait sans intron et donc coderait pour une protéine de 239 acides aminés. En 2004, une équipe du Génoscope a généré des clones d'ADNc GSLT (GenoScope/LifeTechnologies) d'Arabidopsis dans le but de déterminer la structure des gènes et améliorer l'annotation des gènes (Castelli et al., 2004). Lorsque nous avons regardé la base de données (<http://www.genoscope.cns.fr/Arabidopsis/>), deux ADNc provenant du transcrit DYW1 ont été clonés. Il est très intéressant de noter que ces deux ADNc commencent juste après le codon ATG prédit par le modèle AGI (ATG<sub>1</sub>) montrant que le deuxième ATG (ATG<sub>2</sub>), se situant 24 pb après ATG<sub>1</sub>, est probablement le codon initiant la traduction de DYW1. Le gène DYW1 coderait donc pour une protéine plus courte, de 8 acides aminés, que prévu. Nous proposons donc un modèle de gène codant pour une protéine de 231 acides aminés correspondant à une ORF de 696 pb (Figure 20).

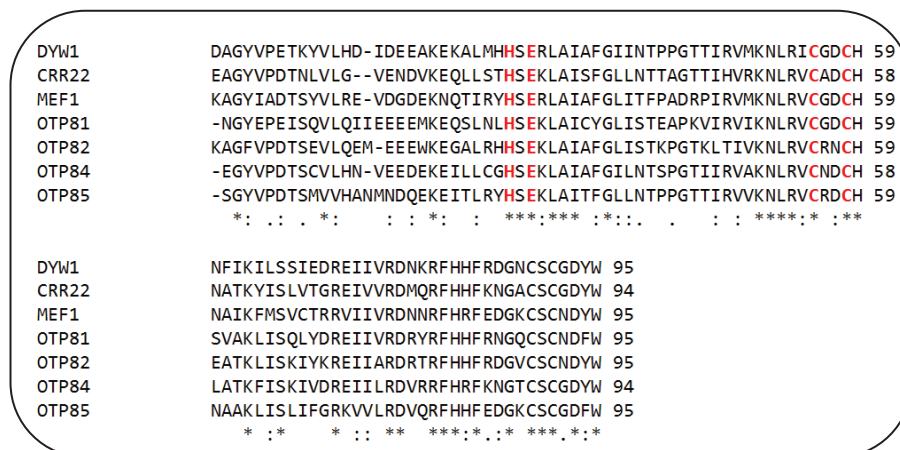


**Figure 20 : Aligement de la séquence de DYW1 avec le consensus du domaine DYW (Lurin et al., 2004).**

La séquence de DYW1 a été alignée avec le consensus du domaine DYW (DYW cons). Les résidus identiques sont surlignés en noir, les résidus similaires en gris.

### I.1.2. Aligement DYW1/PPR-DYW

En alignant la séquence protéique de DYW1 et des PPR-DYW à l'aide du logiciel CLUSTALW, nous avons pu souligner l'excellente conservation du domaine DYW (Figure 21). En effet, nous observons une conservation parfaite d'au moins 58% pour chaque alignement (Tableau 10). Cette excellente conservation du domaine DYW souligne la forte pression de sélection exercée sur ce domaine. Il est alors probable que ce domaine soit capital pour l'activité de la protéine DYW1. De plus, le domaine DYW est caractéristique des protéines impliquées dans l'édition de l'ARN ce qui tend à penser que DYW1 serait impliqué dans l'édition de l'ARN.



**Figure 21 : Aligement des domaines DYW de DYW1 et de PPR impliquées dans l'édition des ARN mitochondriaux et chloroplastiques (Lurin et al., 2004).**

Le domaine DYW de DYW1 a été aligné avec les domaines DYW de 6 protéines PPR à l'aide du logiciel CLUSTALW2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>). Les résidus identiques pour tous les alignements sont indiqués par une astérisque (\*), les résidus similaires sont indiqués par deux points (:), et les résidus proches sont indiqués par un point (.). La signature cytidine déaminase HxE(x)<sub>n</sub>CxxC est surlignée en rouge.

**Tableau 10 : Pourcentage d'identité des domaines DYW de DYW1 et de PPR impliquées dans l'édition.**

	DYW1vsCRR22	DYW1vsMEF1	DYW1vsOTP81	DYW1 vsOTP82	DYW1vsOTP84	DYW1vsOTP85
Résidus identiques (*)	58%	60%	58%	61%	63%	60%
Résidus dont les propriétés sont très similaires (:)	21%	19%	16%	17%	16%	25%
Résidus dont les propriétés sont faiblement similaires (.)	5%	9%	12%	11%	6%	3%
Résidus non conservés	16%	12%	14%	11%	15%	12%

Les domaines DYW des 6 PPR ont été alignés un par un à DYW1 par le logiciel CLUSTALW2. Le nombre de résidus identiques, similaires et faiblement similaires ont été calculés et convertis en pourcentages d'identité entre les différentes occurrences.

Le domaine PPR-like est, quand à lui, très différent des motifs PPR trouvés chez d'autres PPR-DYW. Comme mentionné dans l'introduction, les motifs PPR permettent l'interaction entre la protéine et les acides nucléiques (ici ARN). La non-conservation des motifs PPR suggère que DYW1 ne se lie pas directement à l'ARN et aurait besoin d'un facteur de reconnaissance (PPR ?) de l'ARN pour agir sur un site d'édition.

Notre hypothèse de travail de départ a été que DYW1 interagissait avec les protéines PPR-E pour former un complexe d'édition, la PPR-E reconnaissant l'ARN à éditer et DYW1 catalysant la réaction d'édition. Cette hypothèse a été renforcée dans un premier temps par l'analyse de la séquence de DYW1. L'étape suivante consistait à savoir si DYW1 était réellement impliquée dans l'édition de l'ARN. Toutefois, nous ne pouvons pas nous lancer dans l'analyse des sites d'édition sans savoir où est localisée la protéine DYW1.

## I.2. Localisation de DYW1

### I.2.1. Prédications *in silico*

D'après les logiciels de localisation PREDOTAR (Small et al., 2004) et TargetP (Emanuelsson et al., 2007), DYW1 serait localisée dans les chloroplastes quel que soit le codon d'initiation utilisé (ATG<sub>1</sub> et ATG<sub>2</sub>). Toujours d'après les prédictions de TargetP, un site de clivage du peptide signal est prédit après le 27ème acide aminé, et ainsi la protéine DYW1 mature aurait une taille de 22,8 kDA (Tableau 11).

**Tableau 11 : Prédiction de l'adressage de DYW1 par les programmes Predotar et TargetP.**

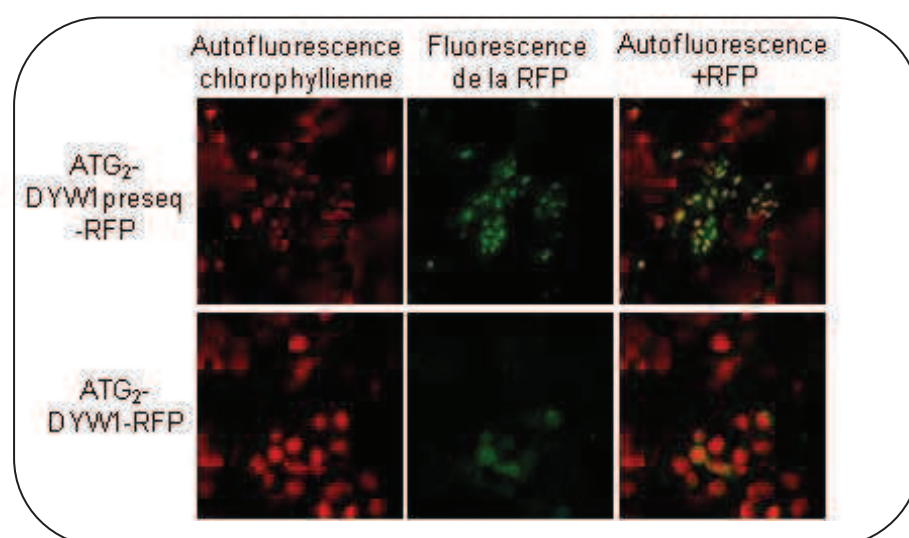
	Programme	Mitochondrie	Chloroplastes	Autres	Prédiction	Site de clivage
ATG <sub>2</sub> DYW1	Predotar	0,02	0,22	0,77	Chloroplastes	N/A
	TargetP	0,082	0,465	0,224	Chloroplastes	27 acides aminés
ATG <sub>1</sub> DYW1	Predotar	0.01	0.84	0.17	Chloroplastes	N/A
	TargetP	0.038	0.457	0.145	Chloroplastes	35 acides aminés

Les séquences protéiques de DYW1 (ATG<sub>1</sub> et ATG<sub>2</sub>) ont été soumises aux logiciels Predotar (<http://urgi.versailles.inra.fr/predotar/predotar.html>) et TargetP (<http://www.cbs.dtu.dk/services/TargetP/>). Les chiffres obtenus correspondent à la probabilité d'adressage dans les différents compartiments de la plante. Le site de clivage définissant la séquence d'adressage est prédit par le logiciel TargetP uniquement.

## I.2.2. Localisation subcellulaire

Afin de tester ces prédictions, l'équipe de Claire Lurin a, préalablement à ma thèse, cloné les séquences codant les 108 premiers acides aminés à partir de l'ATG<sub>1</sub> ou les 100 premiers acides aminés à partir de l'ATG<sub>2</sub> dans le vecteur d'expression pGreen 0229, contenant le gène codant pour la RFP et le promoteur 35S du virus de la mosaïque du chou-fleur (ou CaMV). La souche C58C1 pSOUP d'*Agrobacterium tumefaciens* a été transformée par le vecteur pGreen 0229 et utilisée pour effectuer une transformation d'*Arabidopsis* (Marion et al., 2008). L'observation des plantules a été faite 3 jours après transformation en microscopie confocale permettant la détection de la RFP. Alors que la transformation effectuée avec ATG<sub>1</sub> n'a émis aucun signal détectable dans les cellules, celle contenant l'ATG<sub>2</sub> a clairement montré un signal RFP dans les chloroplastes montrant une localisation chloroplastique (Figure 22).

Dans un même temps, une nouvelle manipulation d'expression transitoire de la forme complète de la protéine DYW1 dont la traduction est initiée à l'ATG<sub>2</sub> (ATG<sub>2</sub>-STOP) a été effectuée, la séquence codant DYW1 étant une nouvelle fois clonée dans le vecteur pGreen 0229. Cette construction donne le même résultat que pour la localisation utilisant les 100 premiers acides aminés de l'ORF (Figure 22). Les deux constructions émettant un signal dans les chloroplastes co-localisent avec l'autofluorescence de la chlorophylle. Il est donc très probable que DYW1 soit localisée dans les chloroplastes et que ATG<sub>2</sub> soit le codon d'initiation de traduction de la protéine DYW1.

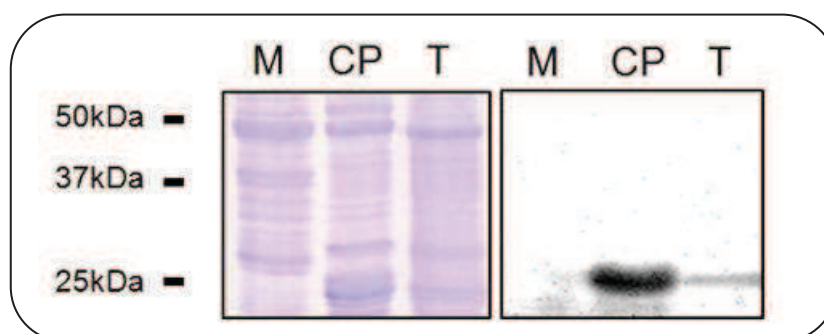


**Figure 22 : Localisation subcellulaire de DYW1 *in planta*.**

Les 100 premiers acides aminés de DYW1 (DYW1preseq) puis la séquence entière de DYW1 ont été fusionnés à la RFP en C-terminal, les deux séquences commençant à la deuxième méthionine (ATG<sub>2</sub>). Les plantes ont été observées 4 jours après transformation. La fluorescence RFP est montrée en vert et la fluorescence naturelle de la chlorophylle est colorée en rouge. La superposition des deux images montre un signal jaune/orangé correspondant à une localisation dans les chloroplastes.

### I.2.3. Etude de la localisation de DYW1 par western blot

Lors de sa thèse, Véronique a produit un anticorps dirigé contre DYW1. Cet anticorps a été utilisé pour détecter DYW1 dans des fractions subcellulaires d'*Arabidopsis thaliana*. Des extraits chloroplastiques et mitochondriaux ont été préparés. Par western blot, elle a pu observer un signal d'environ 25kDa (la taille attendue pour DYW1 étant de 23kDa) dans la fraction enrichie en chloroplastes et non dans la fraction enrichie en mitochondries (Figure 23).



**Figure 23 : Identification de la protéine DYW1 dans différents extraits cellulaires (thèse de Véronique Salone).**

Des chloroplastes et mitochondries d'*Arabidopsis thaliana* ont été isolés et un western blot a été effectué sur les échantillons en utilisant un anticorps dirigé contre DYW1 (25kDa). Un signal est observé dans les chloroplastes (CP) et dans les extraits totaux (T) mais pas dans les mitochondries (M).

Ces résultats, obtenus avec des méthodes indépendantes, ont montré que DYW1 était localisée uniquement dans les chloroplastes et qu'une éventuelle localisation mitochondriale était à exclure. Cette donnée était très importante pour la suite de la thèse et l'étude de la fonction moléculaire de DYW1, fonction qui devait être recherchée dans les plastes.

### I.3. Etude de DYW1 par génétique inverse : obtention de mutants TILLING

Plusieurs méthodes de génétique inverse étaient envisageables pour l'étude de DYW1 : l'analyse de lignées ADN-T provenant du SALK Institute (thèse de Véronique Salone) puis de lignées ARN interférent (ARNi ; thèse de Véronique Salone) et enfin de lignées TILLING.

Seule l'analyse des lignées TILLING utilisées au cours de cette thèse sera détaillée.

#### I.3.1. Obtention de lignées TILLING

Aucune lignée d'insertion ADN-T dans l'ORF *DYW1* n'étant disponible et les lignées ARNi ne donnant pas de résultat réellement exploitable, nous avons commandé des lignées TILLING (Till et al., 2003). La mutation TILLING semblait être la meilleure option car aucune insertion ADN-T n'ayant pu être isolée, une des hypothèses était qu'une mutation « totale » pourrait causer un phénotype létal dès le stade gamète,

rendant impossible d'identification d'une insertion dans le gène. Une recherche de mutants EMS (EthylMéthane Sulfonate) dans le gène *DYW1* a été commandée à la plate-forme TILLING Arabidopsis, Seattle TILLING Project (STP), aux Etats-Unis. Mes premières études ont consisté à analyser par génotypage et par ségrégation génétique ces lignées TILLING.

Le TILLING (Targeting Induced Local Lesions IN Genomes) est une technique de génétique inverse qui permet la sélection de mutations ponctuelles (générées par mutagenèse aléatoire EMS) grâce à la recherche systématique d'hétéroduplexes pour un gène donné dans des pools de mutants. En général, la détection des hétéroduplexes s'effectue sur des produits PCR par le biais d'une nucléase (CELI ou EndoI par exemple). Cette technique permet de découvrir une ou plusieurs mutations dans un gène d'intérêt (Till et al., 2003).

Le STP est une plate-forme aux Etats-Unis proposant d'effectuer du TILLING sur une région indiquée par le commanditaire chez Arabidopsis ou la drosophile. Dans le but d'étudier le gène *DYW1* chez Arabidopsis, nous avons fait appel à la « branche » Arabidopsis de cette plate-forme, l'Arabidopsis TILLING Project (ATP).

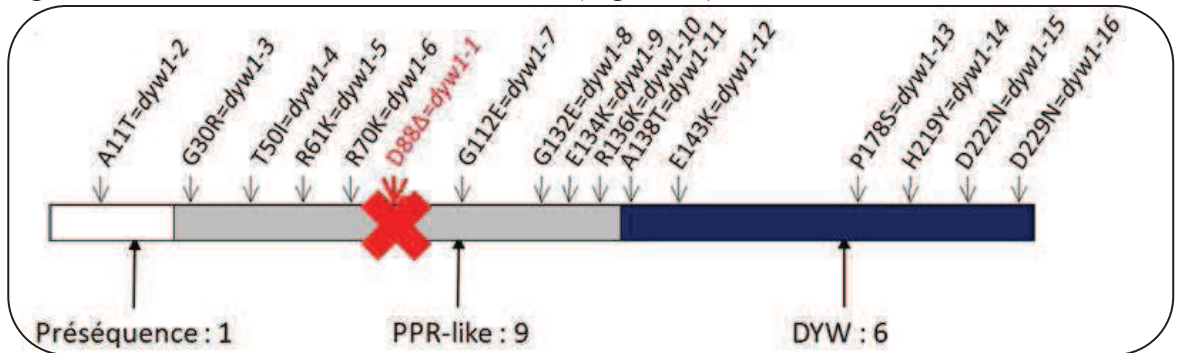
Le TILLING à l'ATP se décompose en cinq étapes :

- (1) Pour établir la plateforme et créer la ressource, une mutagenèse massive EMS, causant un changement de G/C en A/T, a été effectuée sur un lot de graines. Les graines mutées M1 ont été mises à pousser et les graines M2 ont été récoltées. Des extractions d'ADN ont été effectuées sur des pools de 8 familles M2.
- (2) le commanditaire doit choisir une région de 1500 pb à cibler pour la recherche de mutations et les amorces spécifiques à utiliser pour amplifier la zone choisie. Dans le cas du gène *DYW1*, l'ORF d'intérêt ne mesure que 696 pb ce qui nous a permis de cibler une région d'ADN génomique plus large encadrant bien le gène.
- (3) Lors d'une commande, les mutations sont détectées à partir d'ADN provenant de tissus des plantes M2. Le fragment de 1500 pb d'intérêt est amplifié en utilisant l'ADN des pools et des amorces marquées. Des hétéroduplexes sont formés suite à une dénaturation et renaturation de l'ADN. La nucléase CEL I est utilisée pour couper les hétéroduplexes. Les produits de digestion sont ensuite visualisés sur gel Li-Cor par électrophorèse (Till et al., 2003). Les ADN mutés sont alors identifiés.
- (4) l'ATP séquence les mutations pour identification de la plante d'intérêt parmi le pool de 8 et intègre les résultats dans leur base de données.
- (5) un rapport est finalement envoyé au commanditaire pour résumer les différentes mutations trouvées. L'ATP indique la position de la mutation, le changement que cela va impliquer sur la protéine (changement d'un acide aminé, changement de cadre de lecture, sans effet), le numéro de la lignée dans laquelle la mutation a été détectée et la prédiction Sorting Intolerant From Tolerant (SIFT) qui estime l'effet de la mutation sur la protéine (suivant les propriétés de l'acide aminé obtenu après mutation) avec un taux de confiance de 75% ( $SIFT < 0,05$  : effet important) (Ng and Henikoff, 2003). Les



graines M3 des lignées intéressantes doivent ensuite être commandées directement au Nottingham Arabidopsis Stock Center (NASC).

Au total, 50 mutations ont été identifiées dans les 1500 pb couvrant la zone du gène *DYW1* (cf Annexe IV). Trente d'entre elles sont situées dans l'ORF codant pour *DYW1* chez *Arabidopsis*. Parmi elles, 20 modifient la séquence finale de la protéine. Certaines mutations étant redondantes, au final 16 mutations différentes ont été isolées dans l'ORF *DYW1*. Les différentes mutations ont été nommées de *dyw1-1* à *dyw1-16* (nom des lignées). 1 mutation a été identifiée dans la pré-séquence d'adressage, 9 dans la région PPR-like et 6 dans le domaine *DYW* (Figure 24).



**Figure 24 : Ensemble des mutations EMS affectant la séquence de la protéine *DYW1*.**

Pour chaque position présentant une mutation, la nature du changement en acide aminé et le nom de la lignée sont indiqués selon le code ci-dessous.

Nomenclature : exemple de A11T = *dyw1-2*.

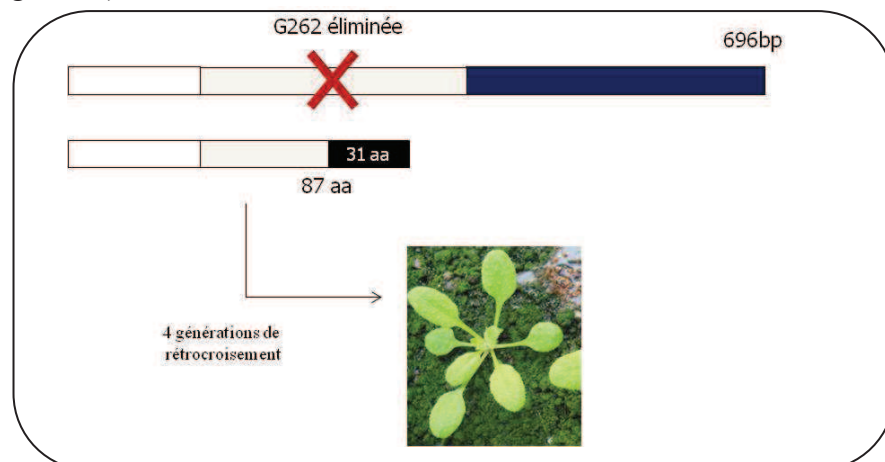
A : l'acide aminé trouvé chez la protéine originale,

11 : position de l'acide aminé,

T : nouvel acide aminé après EMS,

*dyw1-x* : nom de la lignée.

Parmi ces mutations, une nous a particulièrement intéressés : la mutation *dyw1-1* à la 262ème base qui implique un changement de cadre de lecture («frameshift») dans l'ORF codant la protéine et qui conduit à l'apparition d'un codon stop à la 119ème position (Figure 25).



**Figure 25 : Changement de cadre de lecture dans le gène *DYW1* (*dyw1-1*).**

La guanine en position 262 est éliminée impliquant un changement de cadre de lecture et l'apparition d'un codon STOP 31aa après la délétion dans la protéine *DYW1*. Après quatre générations de rétrocroisements effectuées pour «nettoyer» le fond génétique, un phénotype sauvage, totalement différent de celui de la lignée ARNi, est observé.

### I.3.2. Caractérisation des lignées TILLING

Comme décrit précédemment, nous avons obtenu 16 lignées TILLING d'intérêt. Parce que le traitement à l'EMS provoque des mutations partout dans le génome d'*Arabidopsis*, il a été nécessaire de « nettoyer » le fond génétique des mutants obtenus, c'est-à-dire se débarrasser des mutations ponctuelles ailleurs dans le génome tout en gardant la mutation dans *DYW1*. Pour cela, les mutants, arborant des phénotypes plus ou moins marqués, ont été croisés avec des plantes Col-0 de type « sauvage » (technique dite du rétro-croisement). Une seule génération de rétro-croisement n'étant pas suffisante au nettoyage, cette action a été répétée 3 fois pour chaque lignée. A chaque génération, il a été indispensable de génotyper la lignée pour récupérer les plantes hétérozygotes pour la mutation. Le génotypage de l'ADNg a été effectué avec les primers *DYW1\_gen-F* et *DYW1\_gen-R*.

Après rétro-croisement, nous avons généré la génération suivante par autofécondation (descendance des plantes hétérozygotes) afin d'obtenir des plantes homozygotes pour la mutation dans *DYW1*. A notre grande surprise, les plantes homozygotes obtenues possédaient toutes un phénotype équivalent à une plante Col-0 de type « sauvage », y compris les lignées arborant le changement de cadre de lecture (Figure 25). C'est après le 4<sup>ème</sup> rétrocroisement que nous avons obtenu une plante ayant le phénotype Col-0.

Nous avons avancé l'hypothèse selon laquelle ce mutant induisait une perte de fonction à cause du changement de cadre de lecture. Néanmoins, un phénotype sauvage a été observé chez le mutant *dyw1-1*. Notre hypothèse de départ supposant que *DYW1* interagisse avec toutes les protéines PPR de type E/E+, le phénotype devrait être au moins aussi marqué que les mutants « perte de fonction » affectant ces PPR. En effet, chez des mutants présentant des défauts d'édition, certains montrent des phénotypes marqués (Chateigner-Boutin et al., 2008; Zhou et al., 2008; Cai et al., 2009; Hammani et al., 2009; Verbitskiy et al., 2009; Yu et al., 2009; Sung et al., 2010; Tseng et al., 2010).

Toutefois, d'autres mutants perte de fonction ont aussi des phénotypes sauvages notamment ceux affectant le complexe NDH (Kotera et al., 2005; Okuda et al., 2007; Okuda et al., 2009).

### I.3.3. Conclusions

L'étude des lignées ARNi étant très difficile, nous avons opté pour la génération de mutant TILLING afin d'obtenir des « mutations douces » dans le génome. A l'inverse de ce que l'on attendait, le mutant *dyw1-1*, a priori un mutant perte totale de fonction, ne montre aucun phénotype particulier.

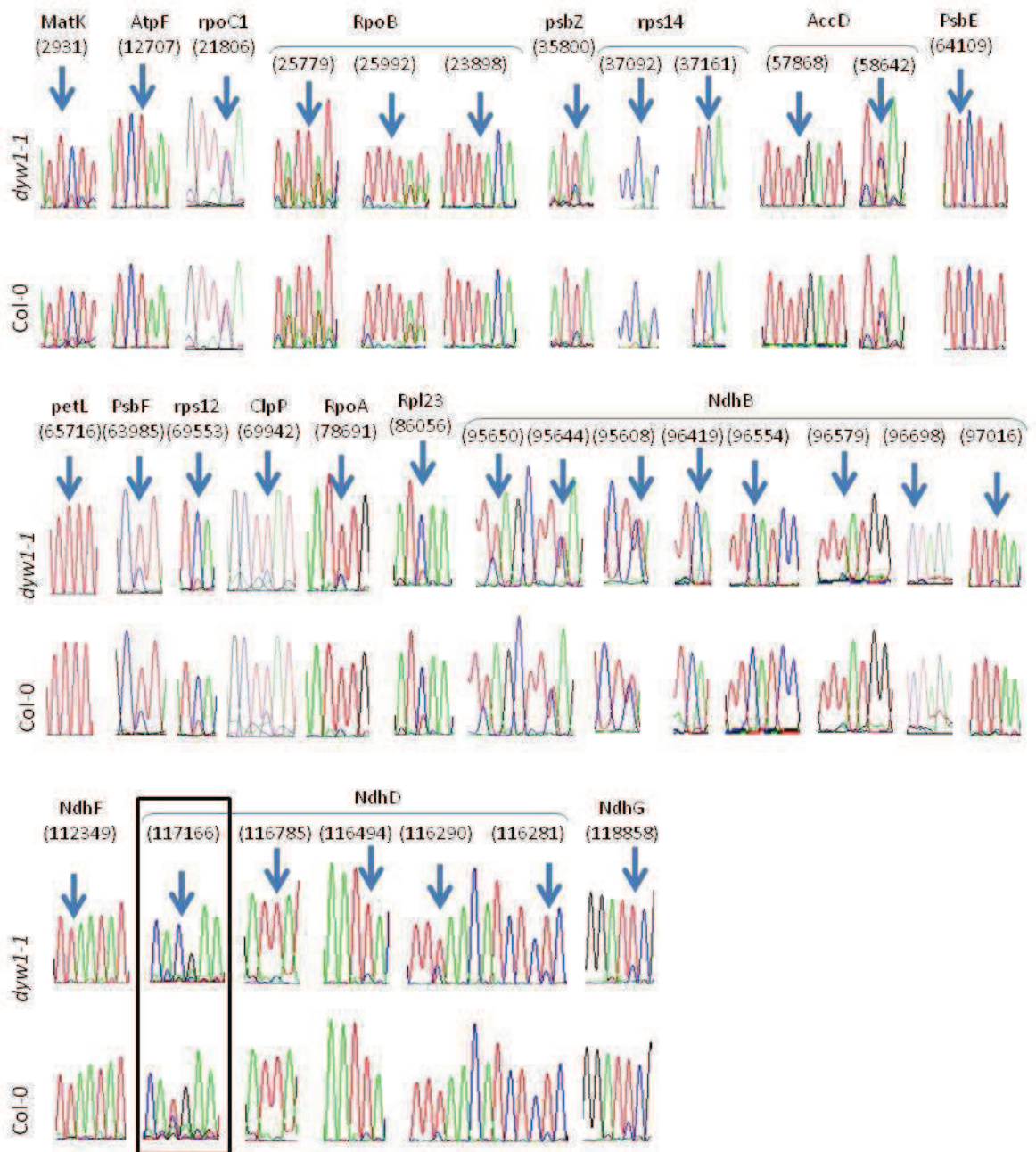
## I.4. Etude des sites d'édition dans le génome chloroplastique du mutant *dyw1-1*

D'après les expériences de localisation décrites précédemment, la protéine DYW1 est localisée dans les chloroplastes d'*Arabidopsis thaliana*. De plus, parce qu'elle contient un domaine DYW, nous avons fait l'hypothèse qu'elle était impliquée dans le mécanisme moléculaire d'édition et pour cela nous avons logiquement caractérisé l'édition dans les chloroplastes des plantes mutantes. Cette caractérisation de la fonction de DYW1 a été essentiellement réalisée sur le mutant *dyw1-1*.

### I.4.1. Séquençage des ADNc chloroplastiques

Pour étudier l'édition des 34 sites chloroplastiques présents dans *Arabidopsis*, nous avons extrait les ARN totaux de *dyw1-1* et de plantes « sauvages » Col-0 (témoin). Ensuite, un traitement à la DNase a été effectué pour éliminer toute trace de contamination par de l'ADNg (qui conduirait à une surestimation du nombre de molécules d'ARN non éditées). Ensuite, les ADN complémentaires ont été synthétisés par transcription inverse et une PCR a été effectuée pour chacun des 34 sites d'édition (cf. Annexe III). Les produits de PCR ainsi obtenus ont été séquencés et les chromatogrammes ont été analysés. Connaissant chaque site d'édition, il est très facile de comparer le taux d'édition du mutant par rapport au sauvage.

Par séquençage, il a été montré que, chez le mutant *dyw1-1*, 33 sites sur 34 se comportent de la même façon que chez les plantes « sauvages » (Figure 26). Seul le site d'édition *ndhD-1* est affecté montrant une absence probable d'édition comparée aux 55% d'édition observés chez le « sauvage » (Figure 26). L'édition à ce site permet la modification d'un ACG en codon d'initiation de traduction AUG. Pour cette raison, l'édition de ce codon est indispensable à la synthèse de la protéine NdhD faisant partie du complexe NDH du photosystème I (Shikanai et al., 1998; Kotera et al., 2005).



**Figure 26 : Editotype chloroplastique du mutant *dyw1-1*.**

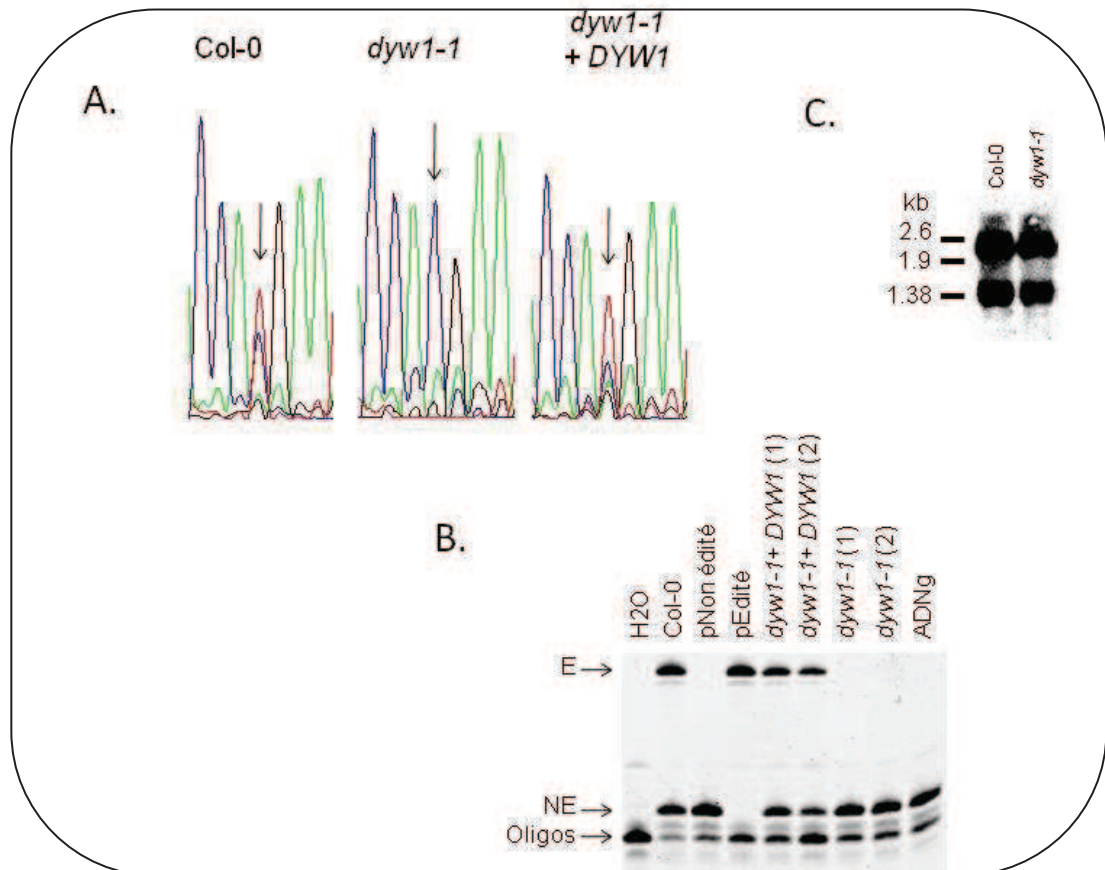
L'ADNc de *dyw1-1* synthétisé à partir d'ARN total a été amplifié par différents oligonucléotides spécifiques des 34 sites d'édition chloroplastique. Les produits de PCR ont été séquencés et comparés à l'ADNc d'une plante sauvage Col-0 (même écotype que *dyw1-1*). Les pics bleus représentent une cytidine et le pic rouge une thymidine. Les flèches bleues indiquent les sites d'édition. Le site d'édition aboli dans le transcrit *ndhD* est encadré en noir.

#### I.4.2. Complémentation fonctionnelle du mutant *dyw1-1*

Pour confirmer que le défaut observé était dû à la mutation dans *DYW1* et non à une autre mutation dans le génome, le mutant *dyw1-1* a été transformé avec une version étendue de *DYW1* contenant, potentiellement, le promoteur et le terminateur natifs. Pour ce faire, *DYW1* a été amplifié avec les oligonucléotides compF et compR donnant un fragment de 1502 pb. Ce fragment a été cloné en Gateway dans pGWB1, un vecteur

binaire sans promoteur constitutif. Cette construction a été insérée dans *Agrobacterium tumefaciens* et *dyw1-1* a été transformé par « floral dipping ». Après sélection sur milieu ½ MS contenant 1% de saccharose et de la kanamycine, l'édition au site *ndhD-1* a été analysée par séquençage des ADNc.

L'analyse de ces séquences a montré la complémentation fonctionnelle du mutant *dyw1-1* par le fragment DYW1 de 1502 pb, le site d'édition du mutant complémenté exhibant environ 55% d'édition, tout comme la plante sauvage (Figure 27A). Ceci confirme que le défaut d'édition dans *ndhD-1* est bien la conséquence du changement de cadre de lecture dans *DYW1*.



**Figure 27 : Edition du site *ndhD-1* avant (*dyw1-1*) et après complémentation (*dyw1-1+DYW1*).**

- A. Le mutant *dyw1-1* a été complémenté avec une version de *DYW1* contenant son propre promoteur. la flèche noire indique la position du site d'édition. Les ADNc ont été amplifiés par PCR et analysés par séquençage.
- B. Analyse de l'édition par PPE sur gel de séquençage. E = transcrit édité, NE = non édité.
- C. Analyse de l'intégrité du transcrit *NdhD* par northern blot. Le signal à 2,6 kb correspond à la version du transcrit non clivé (*psaC-ndhD*) et celui à 1,38 kb correspond à la version clivée.

#### I.4.3. Quantification de l'édition

Afin d'obtenir une quantification précise du défaut d'édition de ce site, une réaction de PPE (pour Poisoned Primer Extension) a été effectuée sur les ADNc de Col-0 sauvage, *dyw1-1* et *dyw1-1* complémenté. La PPE repose sur l'intégration de didéoxyribonucléotides lors d'une réaction d'élongation permettant la discrimination

entre des sites édités et des sites non édités. L'ADN est amplifié par un couple d'amorces encadrant le site étudié puis une élongation (avec une amorce marquée) permet l'incorporation de ddCTP ou ddGTP arrêtant l'extension de l'ADN portant un C ou un G (~non édité) quelques bases avant le site d'arrêt sur une molécule portant un T ou un A (~édité) (Figure 27B). Les échantillons sont analysés sur un gel de séquençage et les produits d'élongation sont détectés par un scanner (LiCor à Evry, Typhoon à Perth).

Par cette méthode, nous avons, d'une part, confirmé le défaut d'édition dans le mutant *dyw1-1* et la restauration de l'édition chez *dyw1-1* complémenté, et d'autre part, calculé le taux d'édition dans chacun des échantillons (Tableau 12). Alors que le site *ndhD-1* de la plante Col-0 de type « sauvage » est édité à hauteur de 55-60%, le transcrite *ndhD* n'est pas édité (0% d'édition) à ce site chez le mutant *dyw1-1* ; après complémentation fonctionnelle, le taux d'édition de *ndhD-1* est de 48 à 52% dans cette expérience.

**Tableau 12 : Quantification du taux d'édition dans chaque échantillon analysé par PPE.**

	Col-0	Plasmide non édité	Plasmide édité	<i>dyw1-1+DYW1</i> (1)	<i>dyw1-1+DYW1</i> (2)	<i>dyw1-1</i> (1)	<i>dyw1-1</i> (2)
%Édité	56	0	100	52	48	0	0
%Non édité	42	100	0	48	52	100	100

La quantification est effectuée après scan sur Typhoon Trio Imager (GE Healthcare) par le biais du logiciel IQtools permettant l'analyse de l'intensité du signal de la fluorescence des oligonucléotides.

#### I.4.4. Analyse de l'intégrité du transcrite *ndhD*

Dans le but de confirmer que l'abolition de l'édition observée n'est pas un effet secondaire d'une modification post-transcriptionnelle déficiente ou d'un problème de stabilité dans l'ARN, le transcrite *ndhD* chez le mutant *dyw1-1* a été analysé en comparaison avec celui de plantes Col-0. Cette expérience a été réalisée par *northern blot* en utilisant une sonde ARN marquée à la biotine. La sonde antisens a été transcrite *in vitro* à partir d'un plasmide pGEMT-easy contenant une portion de *ndhD*. Chez les plantes de type « sauvage » et mutantes, deux types d'ARN sont observés grâce à la sonde *ndhD*. Le premier transcrite est monocistronique et correspond à l'ARN mature de 1,38 kb. Le deuxième transcrite, polycistronique, correspond à la forme précurseur du transcrite comprenant les transcrits *psaC* et *ndhD* (2,3 kb) avant clivage (Del Campo et al., 2002).

Aucune différence dans l'expression des deux formes d'ARN (polycistronique et monocistronique) n'est observée chez le mutant comparé aux plantes « sauvages » (Figure 27C). Ce résultat confirme un rôle spécifique et direct de DYW1 dans l'édition du transcrite *ndhD*.

#### I.4.5. Conclusion

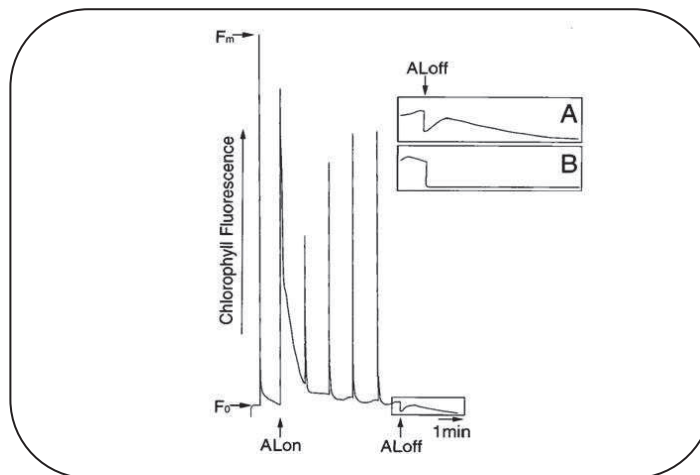
DYW1 étant localisée dans les chloroplastes, il a été facile d'identifier le défaut moléculaire des plantes *dyw1-1* car les transcrits chloroplastiques contiennent beaucoup moins de sites d'édition que les transcrits mitochondriaux. L'ensemble des sites d'édition était connu et les couples d'oligonucléotides étaient aisés à construire. Au vu du phénotype des plantes *dyw1-1*, nous nous attendions à n'observer que peu de sites affectés. Nos résultats montrent que DYW1 est nécessaire à l'édition du site *ndhD-1* permettant la traduction du transcrit *ndhD*. L'unique défaut d'édition concorde avec le phénotype observé. Au vu de l'absence d'édition du site *ndhD-1*, nous pouvons dire que le mutant *dyw1-1* est un mutant « perte de fonction ».

Nous savons que l'édition du site *ndhD-1* requiert également la protéine PPR CRR4 (Kotera et al., 2005). CRR4 est une protéine PPR de type E, qui ne possède pas de domaine DYW en C-terminal. Notre hypothèse de travail (page 41) est compatible avec cette donnée : DYW1 et CRR4 pourraient interagir pour éditer le site *ndhD-1* grâce aux motifs PPR de CRR4 reconnaissant le site d'édition *ndhD-1* et au motif DYW apporté par la protéine DYW1. Pour caractériser plus en détails le phénotype du mutant *dyw1-1*, nous avons caractérisé l'activité NDH chez les plantes mutantes, pour cela une collaboration avec le groupe de Toshiaru Shikanai a été mise en place au vu de leur longue expérience sur le complexe NDH. Ensuite, il était intéressant de comparer les transcriptomes des deux mutants *crr4* et *dyw1-1*. En parallèle à cette caractérisation fine du mutant, nous avons cherché à montrer l'interaction moléculaire entre CRR4 et DYW1.

#### I.5. Effet de la mutation *dyw1-1* sur l'activité NDH : comparaison avec le mutant *crr4-3*

Cette étude a été effectuée par Kamel Hammani, à l'époque étudiant en thèse à Perth dans le laboratoire de Ian Small, et par Kenji Okuda post-doc dans l'équipe de Toshiaru Shikanai, à l'université de Kyoto au Japon.

L'activité du complexe NDH peut être mesurée par l'analyse de l'augmentation transitoire de la fluorescence chlorophyllienne due à la réduction du pool de plastoquinone lorsque la lumière actinique est éteinte (Figure 28) (Shikanai et al., 1998). Cependant, au moment où la lumière s'éteint, il est observé une très légère augmentation de la fluorescence chlorophyllienne due à la réduction (dans le noir) du pool de plastoquinone (Figure 28A). Chez les mutants affectés dans l'activité NDH, tels que *ndhB* par exemple, cette légère augmentation est absente signifiant que la réduction du pool est dépendante du complexe NDH (Shikanai et al., 1998) (Figure 28B).



**Figure 28 : Observation de la fluorescence chlorophyllienne chez le tabac (Shikanai et al., 1998).**

La plante est soumise à une lumière actinique (ALon) durant 5 minutes. la lumière est ensuite éteinte (ALoff).

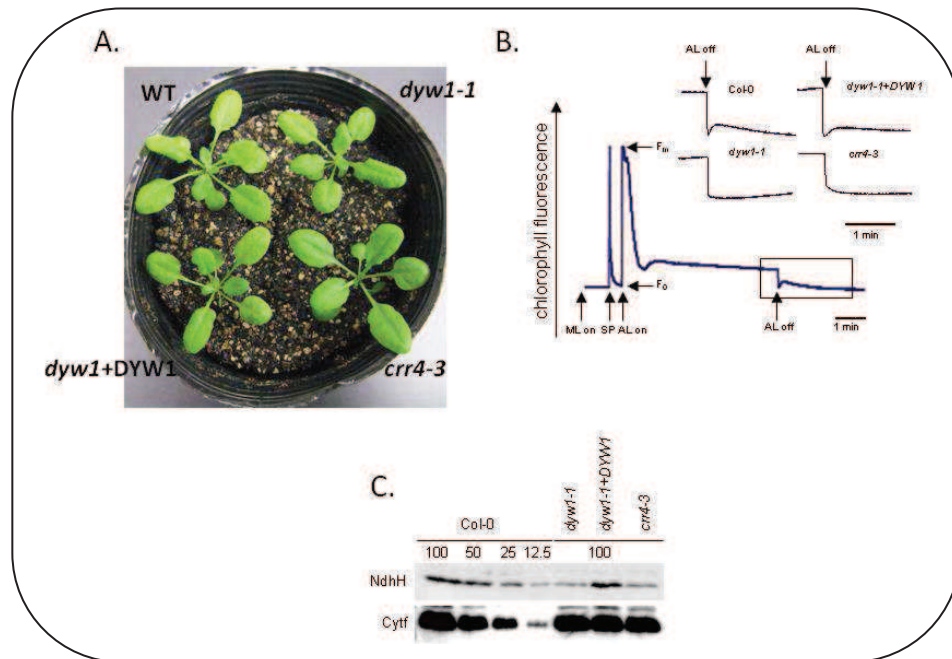
- A. La légère augmentation de la fluorescence après extinction est due à la réduction du pool de plastoquinones.
- B. Le mutant *ndhB* n'affiche pas d'augmentation de la fluorescence, montrant une dépendance du pool de plastoquinones au complexe NDH.

L'édition du site *ndhD-1* permet la conversion d'un codon thréonine (ACG) en codon méthionine (AUG) initiateur de traduction de la protéine NdhD. En absence d'édition chez *dyw1-1*, l'hypothèse est qu'aucune protéine NdhD n'est produite et que le complexe NDH est très affecté. L'absence de d'activité NDH aurait pour conséquence une non-réduction du pool de plastoquinone, comme observé précédemment pour des mutants *ndhB* (Shikanai et al., 1998) mais aussi *crr4* (Kotera et al., 2005). Nous avons donc décidé de comparer la fluorescence chlorophyllienne entre Col-0, *crr4-3* et *dyw1-1*.

Le phénotype des deux mutants étant similaire à celui des plantes « sauvages », il était impossible de les discriminer visuellement (Figure 29A). Le changement de fluorescence a donc été analysé et une absence de réduction du pool de plastoquinone a été observée chez les mutants montrant une perte d'activité du complexe NDH (Figure 29B). Par western blot, un anticorps dirigé contre la sous-unité NdhH a montré une diminution dans la formation du complexe NDH chez le mutant *dyw1-1* (Figure 29C).

Nous avons ainsi montré que le complexe NDH était inactif chez les mutants *dyw1-1* et *crr4-3*, renforçant l'idée que les phénotypes des mutants sont très similaires, seul le génotypage des deux gènes *DYWI* et *CRR4* permettant de distinguer les mutants.





**Figure 29 : Comparaison de l'activité NDH chez *dyw1-1*, *crr4-3* (Kotera et al., 2005), *dyw1-1+DYW1* et Col-0.**

- Comparaison des phénotypes des différentes plantes observées.
- Fluorescence chlorophyllienne chez 4 plantes. SP, saturation; Fo, fluorescence minimale; Fm, fluorescence maximale.
- Western blot effectuée sur des protéines thylakoidiennes. Détection à l'aide d'anticorps des complexes NDH (par NdhH) et du cytochrome b6f (cytF). Les échantillons Col-0 ont été chargés à différentes concentrations afin d'estimer si des changements en terme de quantité de protéines sont visibles.

## I.6. Analyse du transcriptome de *crr4-3* et *dyw1-1*

Dans l'optique de comparer les deux phénotypes mutants, la plateforme transcriptome de l'URGV a effectué une comparaison entre les transcriptomes de plantes *crr4-3* et Col-0 puis de plantes *dyw1-1* et Col-0. Nous étions curieux de savoir si l'on pouvait trouver une différence entre les deux mutants et Col-0 d'un point de vue transcriptomique mis à part l'édition. Par ailleurs, il était intéressant de rechercher un éventuel impact du défaut d'activité NDH sur le transcriptome nucléaire ; un tel défaut signifierait l'existence d'un signal chloroplastique indiquant un dysfonctionnement du complexe NDH.

La puce CATMA (pour « Complete *Arabidopsis* Transcriptome MicroArray ») contient plus de 73 000 sondes spécifiques de gènes (GST) issues de l'annotation EuGene du génome. Elle permet d'analyser le niveau d'expression des gènes chez *Arabidopsis*.

La puce CATMA a été utilisée pour comparer les transcriptomes suivant:

- *crr4-3* vs Col-0 ; 83 gènes différemment exprimés ont été identifiés,
- *dyw1-1* vs Col-0 ; 159 gènes différemment exprimés ont été identifiés.

Pour *crr4-3* vs Col-0, 43 gènes sur 83 sont sous-exprimés et 40 gènes sur 83 sont surexprimés. Concernant *dyw1-1* vs Col-0, 17 gènes sont sous-exprimés et 142 sont surexprimés. Les deux comparaisons de transcriptome n'affichent pas les mêmes nombres gènes sur/sous-exprimés. Il y'a donc une différence notable entre les deux

expériences. Les deux mutants ne semblent pas équivalents du point de vue de la différence de transcriptome malgré leur similarité phénotypique (plantes « sauvages ») et moléculaire (abolition de l'édition du site *ndhD-1*).

A l'aide du logiciel MapMan, j'ai regardé dans quelle voie les gènes différentiellement exprimés étaient impliqués. La fonction des gènes dont les statistiques (« Wilcoxon Rank SumTest ») étaient les plus significatives (de l'ordre de  $10^{-3}$ ) semblent être impliqués dans :

- la résistance au stress et dans la régulation de la transcription pour la comparaison *crr4-3*vsCol0,
- la régulation de la transcription pour la comparaison *dyw1-1*vsCol0.

Les statistiques semblent néanmoins assez faibles et nous sommes en droit de nous demander si nous avons à faire à des artéfacts ou à un réel changement dans ces voies.

A noter que les deux mutants *dyw1-1* et *crr4-3* ont seulement sept gènes différentiellement exprimés en commun (dont la variation va dans le même sens) lorsqu'ils sont comparés à Col-0.

Cinq gènes sous-exprimés ont été observés conjointement dans les deux comparaisons:

- "ADS1 (DELTA 9 DESATURASE 1); oxidoreductase" = AT1G06080,
- "ILL6 (IAA-leucine resistant-like gene 6); metallopeptidase" = AT1G44350,
- "MATE efflux family protein" = AT1G66760,
- "JAI3 JASMONATE-INSENSITIVE 3" = AT3G17860,
- "PHYTOCHROME RAPIDLY REGULATED1 (PAR1)" = AT2G42870.

Deux gènes surexprimés ont été observés conjointement dans les deux comparaisons:

- "PDF1.2b (plant defensin 1.2b)" = AT2G26020,
- "gypsy-like retrotransposon family" = AT2G09187.

En l'état, il est difficile de conclure sur l'effet des mutations sur les gènes ci-dessus, ces derniers ayant des fonctions dans diverses voies métaboliques sans relation évidente avec le phénotype primaire des deux mutants (respectivement métabolisme des lipides, métabolisme de l'auxine, transport transmembranaire, métabolisme du jasmonate, répression transcriptionnelle de l'auxine, résistance au stress, rétrotransposition).

## I.7. Recherche d'une interaction entre CRR4 et DYW1

Comme nous l'avons vu précédemment, les mutants *dyw1-1* et *crr4-3* ne peuvent être distingués ni par leur phénotype macroscopique, ni par leur phénotype moléculaire primaire (cf I.4 des résultats), ni par leur activité NDH (cf I.5 des résultats) ce qui montre que CRR4 et DYW1 jouent un rôle essentiel dans le même processus, celui de l'édition du transcrit *ndhD-1*. De plus, l'absence d'édition chez les deux mutants indique que DYW1 ne peut pas compléter la fonction de CRR4 chez le mutant *crr4-3* et que de la même façon, CRR4 est incapable de compléter la fonction de DYW1 chez le mutant *dyw1-1*.

Ces résultats confortent notre hypothèse selon laquelle CRR4, possédant des motifs PPR, permettrait la reconnaissance et la fixation à l'ARN (sans porter d'activité catalytique permettant l'édition du site) alors que DYW1 contiendrait tout ou partie du site catalytique permettant la conversion de C en U dans son domaine DYW C-terminal (mais ne posséderait aucun motif PPR clair permettant une fixation à l'ARN). C'est dans cette optique que nous avons initié des expériences d'interaction *in vivo*. Au total, quatre expériences ont été réalisées afin de montrer une éventuelle interaction entre CRR4 et DYW1.

#### I.7.1. Purification sur colonne cobalt de CRR4

Les régions codant pour les protéines CRR4 et DYW1 ont été clonées, à l'aide d'enzymes de restriction, dans le vecteur d'expression pETDUET-1 (Novagen). Ce vecteur contient deux sites multiples de clonage permettant l'expression de deux protéines conjointement. Les deux inserts clonés sont chacun placés sous le contrôle d'un promoteur T7 et les protéines sont produites par le biais d'un site de fixation aux ribosomes (RBS). Les protéines CRR4 et DYW1 ont été exprimées dans un système bactérien (BL21). En parallèle de la transformation de la souche BL21 avec le vecteur pETDUET-1, le vecteur pRARE, apportant les ARNt se trouvant en faible abondance chez les bactéries pour l'expression de protéines végétales, a été ajouté.

Dans la construction utilisée, la protéine CRR4 a été étiquetée en N-terminal par une étiquette histidine (6xHis), codée par le plasmide pETDUET-1. L'idée était de purifier la protéine CRR4 (éventuellement en complexe) grâce à cette étiquette sur colonne cobalt. Ainsi, si DYW1 était co-purifiée dans ces expériences avec CRR4, elle pourrait être détectée dans l'élution, par western blot, grâce à l'anticorps anti-DYW1.

Le clonage dans pETDUET-1 et la transformation dans BL21 effectués, les bactéries ont été mises en préculture puis en culture avant que l'expression des protéines recombinantes ne soit induite par traitement à l'IPTG (0,4mM) pendant 3h à 5h. Pour les premières expériences effectuées, nous avons induit à 37°C pendant 3h ce qui donnait une très bonne expression mais augmentait le risque de formation d'agrégats de protéines recombinantes. Nous avons ensuite réduit la température à 15°C et augmenté le temps d'incubation jusqu'à 5h.

Lors de l'élution des protéines CRR4 à l'aide d'imidazole (un compétiteur de l'histidine permettant une élution en conditions non dénaturantes), un signal DYW1 a été obtenu dans l'élution finale. Néanmoins, le signal était assez faible et DYW1 semblait être toujours présente dans les différents lavages. Malgré l'intensification des lavages (quatre au lieu de deux) et l'augmentation de la concentration en imidazole lors des lavages, nous avons été incapables d'éliminer DYW1 de l'élution.

Afin de vérifier l'affinité des colonnes cobalt pour la protéine DYW1, un témoin bactérien n'exprimant que DYW1 et non CRR4 a été produit. Cette expérience a montré que DYW1 est retenue sur la colonne même en absence de CRR4, résultat remettant en question les observations précédentes de co-purification de CRR4 et DYW1 sur ces mêmes colonnes.

Il a donc été impossible de reconstituer l'interaction CRR4-DYW1 dans ces conditions, l'interaction entre DYW1 et la colonne n'étant pas spécifique à CRR4. Néanmoins ce résultat inattendu a permis de montrer que DYW1 interagissait spontanément avec la colonne cobalt. Le domaine DYW contient une signature cytidine déaminase HxE(x)nCxxC (Salone et al., 2007). Parmi les résidus conservés se trouvent une histidine et deux cystéines qui seraient impliquées dans la fixation des ions zinc ( $Zn^{2+}$ ). Il est donc possible que le domaine DYW ait une affinité avec non seulement les ions zinc mais aussi d'autres cations divalents tel que le cobalt. Ce résultat pourrait être un nouvel argument en faveur d'une similarité de DYW1 avec les cytidines déaminases qui ont une affinité pour les ions métaux. Toutefois, cette hypothèse reste à confirmer avec des méthodes spécifiques permettant de tester l'interaction DYW1 avec les ions métaux (cf II.4 des résultats).

### I.7.2. Co-immunoprécipitation de CRR4 et DYW1

Suite à l'échec du test de l'interaction CRR4-DYW1 sur colonne cobalt, une expérience utilisant un autre type de colonne (constituée d'anticorps) a été tentée. Par ailleurs, dans cette expérience, les protéines ont été produites *in planta*.

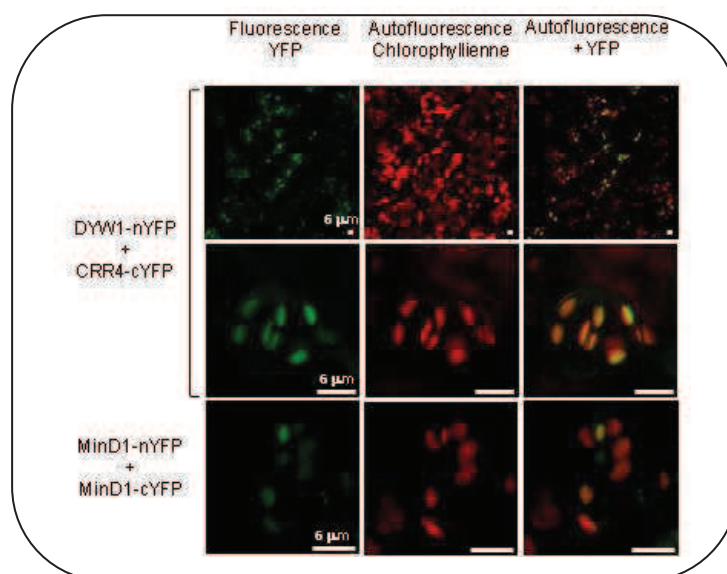
CRR4 a été cloné dans le vecteur binaire Gateway pGWB20 et DYW1 dans le vecteur pGWB15. Ces deux vecteurs permettent respectivement l'étiquetage des protéines avec les étiquettes Myc (10xMyc) et hémagglutinine (3xHA), toutes deux en position en C-terminale. Dans ces vecteurs, les inserts sont placés sous le contrôle du promoteur 35S. Les deux vecteurs ont permis de transformer une souche d'*Agrobacterium tumefaciens* contenant le vecteur p19, codant pour un inhibiteur de « silencing ». Les souches de bactéries ont été inoculées par agro-infiltration de feuilles de *Nicotiana benthamiana*. Trois à cinq jours après infiltration, les feuilles ont été prélevées et les protéines totales extraites. Les premiers essais ont été infructueux car aucune expression de CRR4 ou de DYW1 n'a été obtenue dans les feuilles de *Nicotiana benthamiana*. Nous avons donc laissé cette expérience de côté pour nous consacrer pleinement aux suivantes.

Ainsi par co-immunoprécipitation, il a été impossible de montrer l'interaction entre CRR4 et DYW1. J'aurais eu besoin de plus de temps pour affiner mon protocole mais je me suis concentré sur deux autres expériences d'interaction.

### I.7.3. « Split-YFP »: technique *in planta*

Jusque là, les tentatives pour tester l'interaction entre CRR4 et DYW1 s'étaient révélées infructueuses. Il a fallu trouver un système d'interaction *in planta* et de préférence dans *Arabidopsis thaliana*. Ainsi, nous avons proposé d'étudier l'interaction par BiFC (pour « bimolecular fluorescence complementation » ou Split-YFP). Le principe est de fusionner les protéines d'intérêt à une moitié de l'YFP (pour « Yellow fluorescent protein »). Si les deux protéines interagissent, la YFP est reconstituée et émet un signal visible en microscopie confocale, les moitiés de YFP n'étant pas fluorescentes de façon indépendante et ne se reconstituant pas spontanément (Citovsky et al., 2006; Marion et al., 2008).

Les clonages ont été effectués par Alexandra Avon et moi-même. L'analyse en microscopie confocale a été faite par Richard Berthomé et Alexandra Avon. Différentes combinaisons ont été effectuées : fusion YFP en N-terminal ou en C-terminal des protéines et la partie N-terminale de l'YFP fusionnée à une protéine, la partie C-terminale de l'YFP fusionnée à l'autre, et inversement. Après transformation des plantules sous vide, les feuilles sont observées au microscope confocal. Ainsi, un signal pour la construction CRR4-cYFP x DYW1-nYFP montrant une possible interaction entre les deux protéines a pu être observé dans les chloroplastes (Figure 30). A noter que ce signal n'est pas observé si les moitiés du gène rapporteur sont inversées (CRR4-nYFP x DYW1-cYFP) ainsi que lorsque les protéines sont fusionnées à la YFP en N-terminal. Dans ce dernier cas, les protéines ne sont certainement pas correctement localisées ce qui peut affecter leur interaction. Aucun signal n'est également observé lors de la coexpression DYW1-cYFP x DYW1-nYFP (Figure 30).



**Figure 30 : Etude de l'interaction entre CRR4 et DYW1 par split-YFP.**

Des plantules d'*Arabidopsis thaliana* ont été co-transformées de manière transitoire avec DYW1-nYFP et CRR4-cYFP. Cette co-transformation génère une fluorescence jaune qui est colocalisée avec l'autofluorescence des plastes. Ce signal est aussi observé chez l'homodimère MinD1xMinD1, utilisé comme contrôle positif (Maple et al., 2007). Aucun signal n'est observé pour les plantules transformées avec DYW1-nYFP et MinD1-cYFP ainsi que DYW1-nYFP et CLB19-cYFP

Pour montrer que l'interaction CRR4-DYW1 observée est bien spécifique des deux protéines, différents contrôles négatifs ont été effectués (Tableau 13). Nous avons ainsi testé l'interaction entre CLB19 et DYW1 dans les mêmes conditions. CLB19 est une protéine PPR de type E agissant comme facteur de spécificité dans l'édition des sites *rpoA* et *clpP* dans les chloroplastes (Chateigner-Boutin et al., 2008). La structure de CLB19 est très similaire à celle de CRR4. Dans ce cas, aucun signal n'a été détecté confirmant que DYW1 n'interagit pas avec toutes les protéines PPR-E des chloroplastes. Comme contrôle positif, nous avons également testé des constructions exprimant des protéines connues pour interagir dans les chloroplastes (interactions déjà observées par BiFc) : MinD1 X MinD1 (Maple et al., 2007) et FSD2 X FSD3 (Myouga et al., 2008). Alors que nous avons confirmé que ces protéines interagissent entre elles, celles-ci sont par contre incapables d'interagir avec DYW1 confirmant une spécificité

d'interaction entre CRR4 et DYW1 (contrôle négatif). La surexpression de deux protéines dans les chloroplastes ne permet donc pas d'obtenir un signal observable au microscope confocal.

Ainsi, grâce à la technique du BiFc, nous avons pu montrer une interaction spécifique entre CRR4 et DYW1. Notre modèle de départ s'est au fur et à mesure affiné, passant de l'hypothèse selon laquelle toutes les PPR-E interagissaient avec DYW1 à celle où seulement CRR4 interagit avec DYW1.

**Tableau 13 : Constructions testées par split-YFP.**

P1 (nYFP)	P4 (cYFP)	Résultat
DYW1	CRR4	<b>Interaction</b>
CRR4	DYW1	Pas d'interaction observée
DYW1	DYW1	Pas d'interaction observée
DYW1	CLB19	Pas d'interaction observée
DYW1	FSD2	Pas d'interaction observée
DYW1	MIND1	Pas d'interaction observée

#### I.7.4. TAP-tag : recherche d'un complexe d'édition

Nous avons contacté puis collaboré avec l'équipe de Geert de Jaeger (VIB, Gand, Belgique) afin d'initier une purification de complexe protéique par une approche de TAP-tag (pour «Tandem Affinity Purification») (Van Leene et al., 2011). L'intérêt ici est de confirmer l'interaction CRR4-DYW1 mais aussi d'identifier les autres membres du complexe, s'il y'en a.

L'approche a été menée dans un premier temps en étiquetant la protéine DYW1 puis dans un second temps en ciblant la protéine CRR4. Les séquences codant pour les deux protéines ont été clonées dans un vecteur permettant l'expression d'une fusion en position C-terminale de l'insert d'une étiquette SBP (Streptavidine Binding Protein), d'un site de clivage TEV (pour «Tobacco Etch Virus») et de deux protéines G (Figure 31). Des cellules d'*Arabidopsis thaliana* ont été transformées, cultivées et sélectionnées afin d'obtenir une culture cellulaire importante. Les protéines marquées (et le complexe dans lequel elles sont éventuellement impliquées) ont été purifiées sur une colonne de sépharose contenant des IgG. Ceci permet à la colonne de retenir la protéine marquée par son étiquette constituée de protéines G. Par la suite, les complexes ont été élués de la colonne par clivage du peptide TEV grâce à une protéase spécifique. Les protéines ont été ensuite purifiées sur une colonne de streptavidine permettant la rétention des étiquettes SBP. Après ces deux étapes de purification, l'éluat obtenu contient la protéine d'intérêt étiquetée ainsi que toutes d'éventuelles protéines impliquées dans le même complexe. Les complexes ont été séparés sur SDS PAGE et

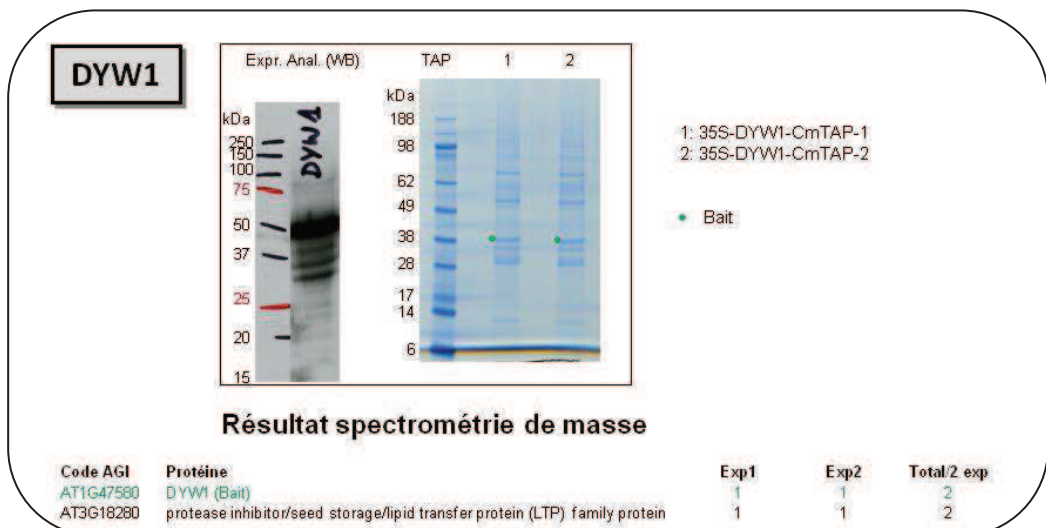
chaque ligne de migration a été divisée en 48 morceaux qui ont été analysés par spectrométrie de masse, dans le but d'identifier les protéines membres du complexe. Deux répétitions indépendantes ont été effectuées.



**Figure 31 : Construction effectuée pour l'étude de PPR en TAP-tag.**

La protéine d'intérêt est liée à un enchainement d'étiquettes permettant les différentes étapes de purification du complexe. Tout d'abord, l'extrait cellulaire est passé sur une colonne liant les protéines G (protG) puis le produit est clivé par une enzyme (TEV) de sorte à libérer le complexe. Ensuite, une deuxième purification sur colonne de streptavidine par le biais du marqueur SBP (pour « streptavidine binding protein »). Les protéines sont séparées sur gel dénaturant et analysées par spectrométrie de masse. Une répétition est effectuée pour comparer les résultats et éliminer les faux positifs.

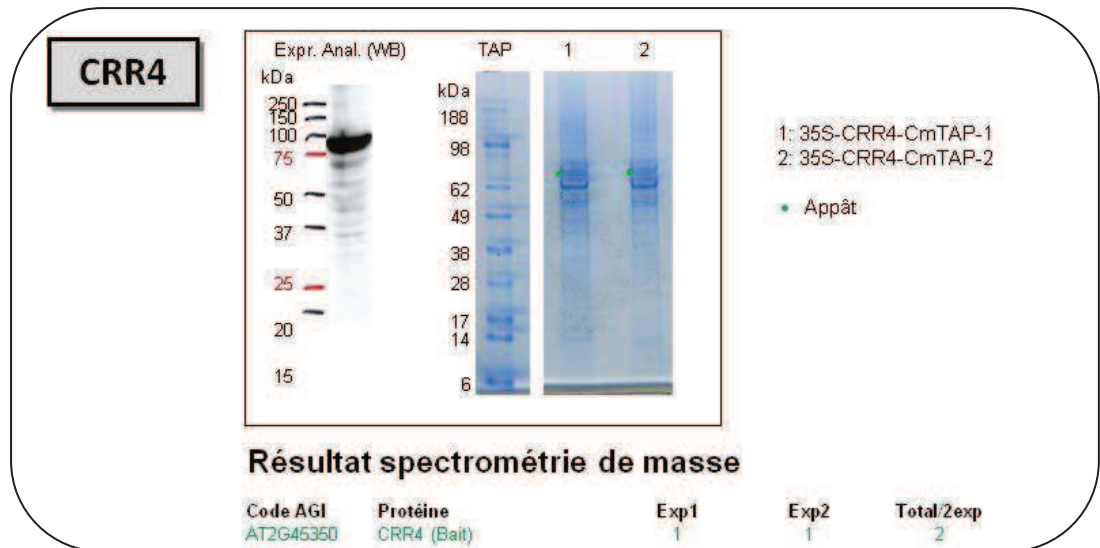
Il a été montré que l'ORF contenant DYW1, dont la masse moléculaire attendue est de 50 kDa, était très bien exprimée (Figure 32). Le complexe contenant la protéine DYW1 a été purifié et déposé sur gel SDS PAGE afin de séparer les éventuelles protéines contenues dans le complexe DYW1. A l'œil, nous pouvons observer la protéine DYW1 à un poids moléculaire de 38 kDa après clivage à la protéase (Figure 32). Cependant par spectrométrie de masse, il a été impossible d'identifier des protéines purifiées autre que DYW1 à l'exception de la protéine codée par le gène At3g18280, une protéine inhibitrice de protéase (Figure 32). Par cette méthode, nous n'avons pas pu confirmer les résultats obtenus en Split-YFP. Il faut remarquer que l'étiquette TAP-tag utilisée fait environ le même poids moléculaire que DYW1 (soit 25 kDa). Cette étiquette est ajoutée en C-terminal et il est possible que l'étiquette entraîne un encombrement stérique et/ou un problème dans la conformation de DYW1, empêchant l'interaction avec ses interacteurs, de façon générale, et CRR4, en particulier.



**Figure 32 : Résultat de la purification TAP-tag effectuée sur DYW1 (Geert de Jaeger, Gand).**

L'extrait est analysé par western blot afin de vérifier si l'appât (bait) est présent (colonne de gauche). Ensuite, les échantillons sont extraits d'un gel d'acrylamide. Chaque ligne est divisée en petites fractions et qui sont analysées par spectrométrie de masse. Après analyses statistiques, les faux positifs sont éliminés. DYW1 est bien identifiée par spectrométrie de masse dans l'extrait purifié mais seul un inhibiteur de protéase a été co-purifié.

Afin de tester l'approche TAP-tag dans des conditions différentes, nous avons choisi dans un deuxième temps d'étiqueter la protéine CRR4. La protéine CRR4 étiquetée s'est parfaitement exprimée dans les cellules, exprimant une protéine de masse moléculaire supérieure à 75k Da (masse moléculaire attendue 94 kDa) (Figure 33). A nouveau, alors que la protéine CRR4 a été purifiée sans problème et pouvait être observée à l'œil sur gel SDS PAGE, la spectrométrie de masse n'a permis d'identifier aucun partenaire protéique de CRR4.

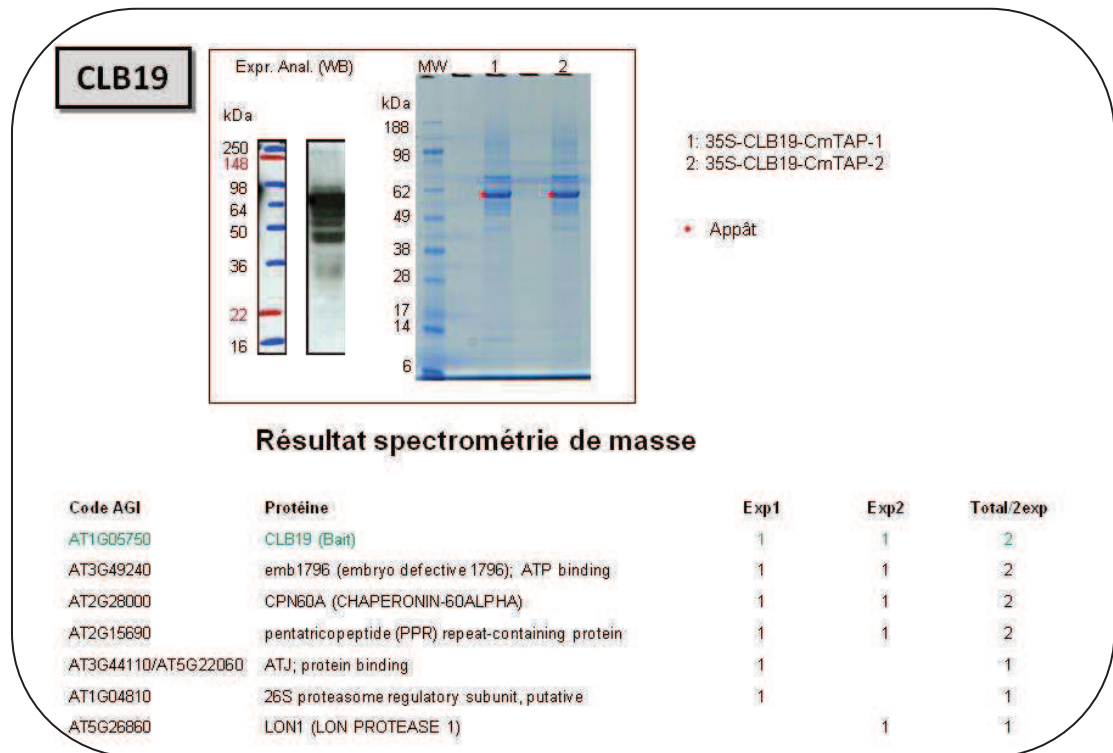


**Figure 33 : Résultat de la purification TAP-tag effectuée sur CRR4 (Geert de Jaeger, Gand).**  
Aucun interacteur n'a été identifié avec CRR4 dans cette expérience.

Par TAP-tag, nous avons donc été incapables de trouver des interacteurs de CRR4 et DYW1 et de prouver en particulier l'interaction entre les deux protéines. Ceci ne veut pas dire que l'interaction entre les deux protéines soit remise en cause, mais que dans ces conditions là, il est impossible de purifier le complexe.

Toutefois, nous avons dans le même temps purifié un complexe contenant la protéine PPR-E CLB19. A l'inverse des essais réalisés avec DYW1 et CRR4, la purification de CLB19 a permis de co-purifier deux protéines PPR : la première At3g49240 une PPR pure baptisée OTP100 et la deuxième At2g15690, une PPR de type DYW ne contenant que quatre motifs PPR baptisée DYW2 (Figure 34). Ce résultat particulièrement excitant est mentionné dans cette thèse mais fait partie du projet de thèse de Mauricio Lopez Obando, à l'URGV sous la direction de Claire Lurin. DYW2 a des similarités avec DYW1 en raison de la présence d'un domaine DYW ainsi qu'un nombre très faible de motifs PPR. Nous avons décidé de considérer DYW2 comme une protéine DYW1-like. En effet, la structure commune qu'elle a avec DYW1 fait penser qu'elle contiendrait une activité cytidine déaminase et qu'elle interagirait avec CLB19 pour éditer les transcrits *rpoA* et *clpP*. Le modèle pourrait être calqué sur celui de DYW1.





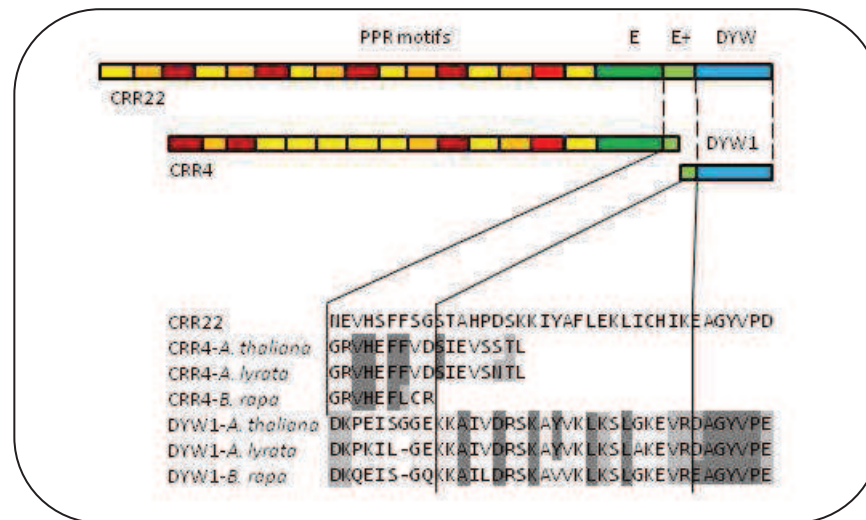
**Figure 34 : Résultat de la purification TAP-tag effectuée sur CLB19 (Geert de Jaeger, Gand).**

CLB19 interagit avec un groupe de protéines notamment deux PPR AT2G15690 (une PPR « DYW1-like » appelée DYW2) et AT3G49240 (une PPR-P appelée OTP100).

## I.8. Reconstitution d'une PPR-DYW à partir de CRR4 et DYW1

### I.8.1 Alignement des domaines E de CRR4...et de DYW1

Nous avons montré en utilisant l'approche de « Split-YFP » que CRR4 et DYW1 interagissent. Par ailleurs, nous avons remarqué que DYW1 possède, en amont de son domaine DYW, 16 acides aminés correspondant potentiellement à une fin de domaine E. Par exemple, l'alignement de DYW1 avec la protéine PPR-DYW CRR22 confirme la présence d'une partie C-terminale de domaine E dans DYW1 (Figure 35). De la même façon, l'alignement du domaine E de CRR4 avec le domaine E de CRR22 montre que le domaine E de CRR4 est légèrement tronqué en C-terminal (Figure 35). Nous avons alors émis l'hypothèse simple que DYW1 pouvait tout simplement être la partie manquante de CRR4 pour agir sur le site d'édition *ndhD-1*.



**Figure 35 : Alignement des domaines E et DYW de CRR22 avec ceux de CRR4 et DYW1.**

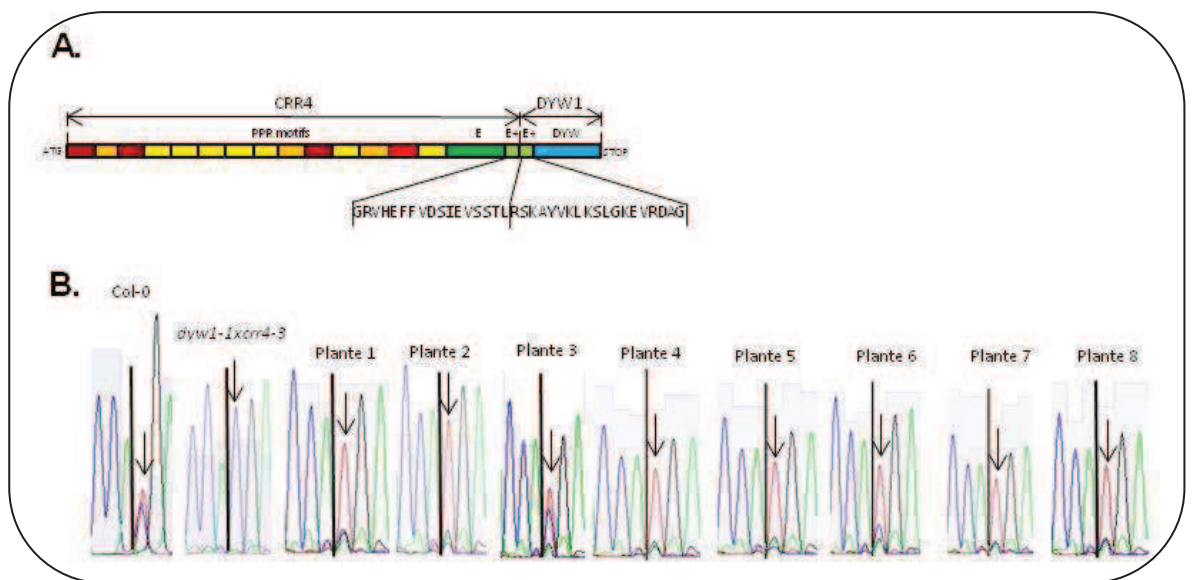
Le domaine E+ partiel de CRR4 (et de ses orthologues), de DYW1 (et de ses orthologues) et de CRR22 (Okuda et al., 2009) ont été alignés à l'aide du logiciel CLUSTALW2. Les résidus identiques à CRR22 sont surlignés en gris foncé, les résidus similaires en gris clair. CRR4 contient la région N-terminale du domaine E+ tandis que la région C-terminale est située dans DYW1.

### I.8.2. Complémentation du double mutant *crr4-3xdyw1-1*

Pour tester la possibilité que DYW1 soit la partie manquante de CRR4, ces deux protéines ont été fusionnées, la jonction entre les deux se faisant entre les deux domaines E. Nous avons ainsi reconstruit une PPR-DYW contenant les motifs PPR de CRR4 permettant la fixation au transcrit *ndhD-1*, le domaine E de CRR4 et de DYW1 permettant d'obtenir un domaine E complet et le domaine DYW de DYW1 permettant hypothétiquement la conversion de la cytidine en uridine (Figure 36A).

Cette fusion a été clonée en Gateway dans le vecteur pGWB2 contenant un promoteur 35S. Dans le même temps, un double mutant homozygote pour les mutations *crr4-3* et *dyw1-1* a été produit. Comme attendu, ce mutant est parfaitement viable et ne contient qu'un défaut d'édition dans le site *ndhD-1*, le même que pour les simples mutants *crr4-3* et *dyw1-1*. Le double mutant a été complété avec la fusion de CRR4-DYW1. Après sélection sur milieu MS contenant de la kanamycine, les ARN ont été extraits et l'édition du site *ndhD-1* a été analysée. Dans ces plantes, nous avons pu montrer que l'édition est restaurée à un niveau au moins égal à celui observé dans des plantes sauvages (et même augmentée dans 7 lignées sur 8) (Figure 36B).

La fusion entre les protéines CRR4 et DYW1 est donc fonctionnelle et complémente le double mutant, démontrant ainsi que DYW1 serait la partie manquante de CRR4.



**Figure 36 : Complémentation du double mutant *crr4-3xdyw1-1* avec une fusion de CRR4 et DYW1.**

A. Structure de la fusion CRR4-DYW1. Les motifs PPR et les domaines E et E+ de CRR4 ont été fusionnés aux domaines E+ et DYW de DYW1.

B. Séquençage des ADNc obtenus à partir de plantes *crr4-3xdyw1-1* complétementées avec la fusion de CRR4 et DYW1. Les flèches indiquent la position du site d'édition *ndhD-1*. Les courbes rouges représentent une thymidine (édition), les bleus représentent une cytidine (molécules non éditées). 8 transformants primaires indépendants ont été analysés.

## I.9. Recherches d'orthologues du complexe minimal d'édition d'AtDYW1

D'après les résultats obtenus précédemment, il est maintenant clair que CRR4 et DYW1 agissent pour éditer le site *ndhD-1* chez *Arabidopsis thaliana*. Afin de savoir s'il pourrait se produire la même chose chez d'autres plantes, les orthologues putatifs de DYW1 et CRR4 ont été comparés par bioinformatique. De même, l'état de conservation du transcrit *ndhD* à travers les différentes plantes étudiées a été examinées. Mes recherches se sont limitées aux angiospermes.

### I.9.1. Orthologues chez des monocotylédones

Nous savons que l'édition du transcrit *ndhD* est très variable à travers les espèces. Chez le riz, le transcrit *ndhD* n'a pas besoin d'être édité pour être traduit. En effet, il contient d'origine le codon d'initiation de la traduction (ATG) dans son génome, au contraire de ce qui est observé chez *Arabidopsis* (Tsudzuki et al., 2001). Dans ce cas, des orthologues de CRR4 et DYW1 ont été cherchés. Aucun orthologue d'AtCRR4 n'a pu être trouvé mais un orthologue putatif d'AtDYW1 a été découvert chez le riz (OsPPR\_03g55840) (<http://www.plantenergy.uwa.edu.au/applications/osatppr/>) (O'Toole et al., 2008). Il est donc possible que la présence d'un orthologue d'AtCRR4 soit dépendante du besoin d'éditer le transcrit *ndhD*. Le constat est le même chez le maïs (Hayes and Mulligan, 2011). Faute d'autres données, il a été impossible de généraliser à l'ensemble des monocotylédones.

## I.9.2. Orthologues chez des dicotylédones

En ce qui concerne les dicotylédones, des orthologues d'AtCRR4 ont pu être identifiés chez plusieurs espèces. Les orthologues d'AtCRR4 sont extrêmement bien conservés ce qui a rendu leur identification assez simple. Des orthologues d'AtCRR4 ont été obtenus pour 25 espèces dont 4 de la classe des Astéridés, 4 de la classe des Rosidées et enfin 17 de la famille des Brassicacées (Figure 37). Il est possible de montrer un lien entre la conservation des motifs PPR de ces orthologues et la conservation de séquence des différents transcrits *ndhD* et ce dans toutes les plantes observées. Cette observation est très cohérente avec le rôle d'interaction séquence spécifique des motifs PPR. Ainsi, si la séquence du *ndhD* est bien conservée autour du site d'édition, alors son interacteur, en l'occurrence CRR4, sera particulièrement bien conservé dans la région des répétitions PPR. En revanche, l'extrémité C-terminale des protéines CRR4 est assez variable. Alors que les protéines CRR4 de la famille des Brassicacées ont un domaine E assez court, en règle générale les domaines E de CRR4 chez d'autres espèces sont plus longs d'une dizaine d'acides aminés (Figure 37).

	Domaine E	Domaine E+		
<i>Acordi folium</i>	IPGCSWIELDGNVHEFFV-----DSNTL-----	587		
<i>Agrandiflorum</i>	IPGCSWIELDGNVHEFFVQDK----FVDSNTL-----	594		
<i>Dnemorosa</i>	IPGCSWIELDGNVHEFFVDT-----MELSENT-----	612		
<i>Mincana</i>	IPGCSWIELDGNVHEFFVDT-----MELSENT-----	584		
<i>Ahirsuta</i>	IPGCSWIELDGNVHEFFVDS-----MELSENTL-----	615		
<i>Athaliana</i>	IPGCSWIELDGRVHEFFVDS-----IEVSSTL-----	613		
<i>Alyrata</i>	IPGCSWIELDGRVHEFFVDS-----IEVSNTL-----	606		
<i>Cbursa-pastoris</i>	IPGCSWIELDGRVHEFFVDS-----IEVSNTL-----	617	Dicots-rosids-malvids -brassicaceae	
<i>Opumila</i>	IPGCSWIELDGRVHEFFVDS-----IEVSNTL-----	579		
<i>Lsativum</i>	IPGCSWIELDGRVHEFFVDS-----IEVSKTL-----	597		
<i>Lvirginicum</i>	IPGCSWIELDGRVHEFFVDS-----IEVSNTL-----	594		
<i>Cwallichii</i>	VPGCSWIELDGRVHEFFVDS-----IEVSNTL-----	617		
<i>Nofficinale</i>	VPGCSWIELDGRVHEFFVDS-----IEVSNTL-----	614		
<i>Everna</i>	VPGCSWIELDGRVHEFFVDS-----IEVSNTL-----	588		
<i>Erapa</i>	IPGCSWIELDGRVHEFFLCR-----	612		
<i>Eoleracea</i>	IPGCSWIELDGRVHEFFLCR-----	590		
<i>Rsativus</i>	LPGCSWIELDGRVHEFFLCR-----	577		
<i>Lmaritima</i>	VPGCSWIELDGRVHEFFLCR-----	570		
<i>Csativus</i>	VPGCSWIELGIVVHEFFLVDRDKSHFYVSEIYSVLDGFGASNLQVIDCKC---	515		Dicots-asterids
<i>Gmax</i>	IPGCSWIELGGIVVHQFS-----	545		
<i>Ptrichocarpa</i>	IPGCSWIELGSHVYAFVQDKSHFQFSGIYSLDLSMTDSEHSYCKNVKM	643		
<i>Rcommunis</i>	IPGCSWIDLEGITHEPSVQDKSHFKVTBIYSLLDSPVTPNSELSYSGNYG-	639		
<i>Nsylvestris</i>	IVGSSRIELQGI VHEFSVQDKSHDQVBEIYSTLQMVCV-----	626		Dicots-rosids
<i>Ntomentosiformis</i>	IVGSSRIELQGI IHEFSVQDKSHDQVBEIYSTLQMVCM-----	626		
<i>Mguttatus</i>	VPGCSWIELEGTVH-----	568		
<i>Vvinifera</i>	IPGCSQIELEGNVHEFFVDRDKSHFQVREIYMLDLSLALNSEVAYCKH---	635		

Figure 37 : Domaines E et E+ d'orthologues potentiels de AtCRR4.

Les orthologues d'AtCRR4 ont été identifiés par BLAST de la séquence CRR4 protéique au NCBI (<http://blast.ncbi.nlm.nih.gov/>), sur le site du JGI (<http://www.jgi.doe.gov/>), sur phytozome (<http://www.phytozome.net/>) et la plateforme des brassicacées (<http://brassica.bbsrc.ac.uk/>). Alors que les domaines E semblent assez conservés entre les espèces, le domaine E+ semble beaucoup plus variable avec notamment des différences de taille assez significatives.

Dans le cas d'AtDYW1, il est difficile de l'aligner sur les génomes d'autres espèces sans obtenir toutes les protéines PPR contenant un domaine DYW. Il a donc été nécessaire d'aligner DYW1 et de filtrer les tailles des protéines obtenues. J'ai pu identifier par cette méthode 7 protéines potentiellement orthologues d'AtDYW1 dont 3 chez des Monocotylédones, 2 chez des Rosidées et 2 chez des Brassicacées.

Chez les Rosidées, tous les orthologues potentiels d'AtDYW1 contiennent des motifs PPR mais pas de motif E. La protéine CRR4 est également identifiée facilement chez les Rosidées. Néanmoins, elle possède un domaine E+ plus long que chez Arabidopsis. Ainsi, il semble y avoir un lien entre la longueur du domaine E+ et structure de DYW1. A l'inverse, chez les Brassicacées, les orthologues de DYW1 ne possèdent pas de motifs PPR, contiennent un petit fragment de domaine E+ et un domaine DYW. Chez toutes ces plantes, la protéine CRR4 possède un domaine E+ relativement court.

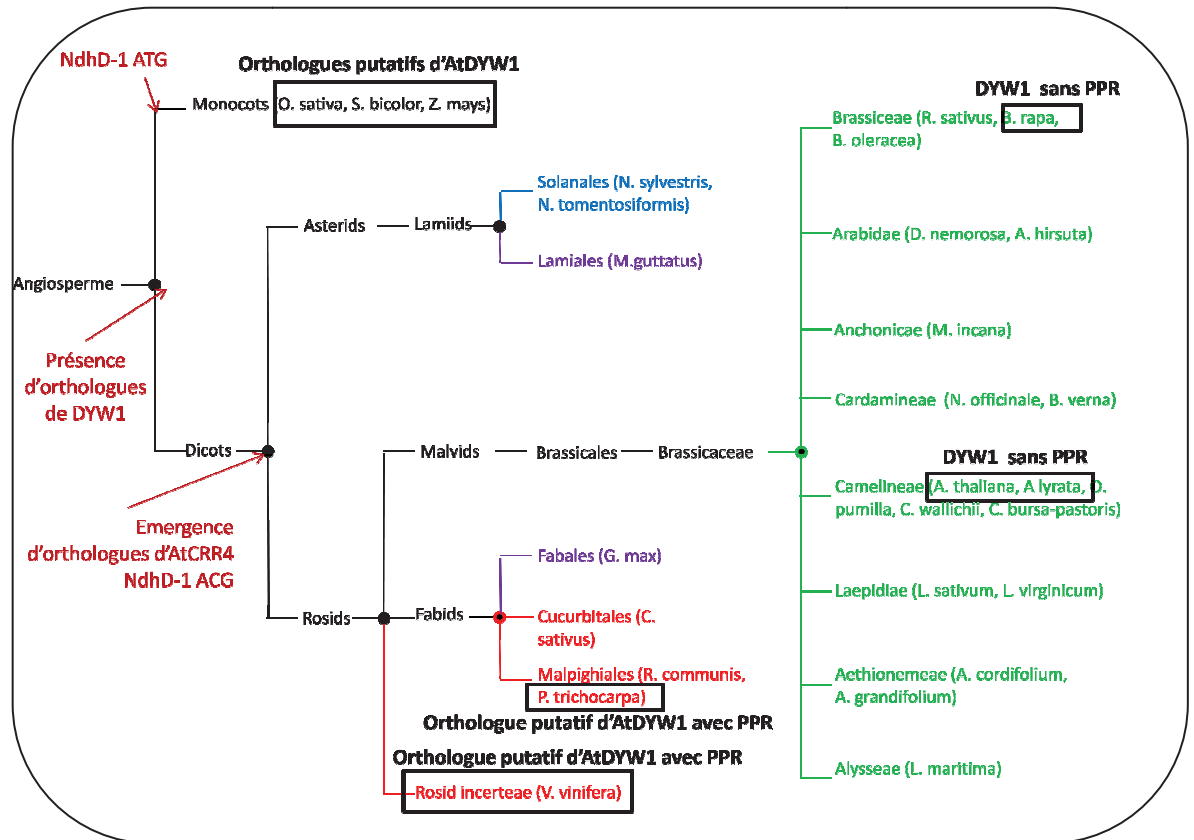
Cette étude phylogénétique combinée aux résultats expérimentaux précédents permet de formuler plusieurs hypothèses.

L'édition du site *ndhD-1* est probablement apparue dans un ancêtre commun des végétaux et a été perdue chez les monocotylédones (les monocotylédones ne possédant pas ce site d'édition). Il est également possible que ce site soit apparu au moment de l'émergence des dicotylédones mais cette dernière hypothèse est moins probable car le site *ndhD-1* existe chez *Anthoceros formosae* mais pas chez la fougère *Adiantum capillus-veneris* (Tillich et al., 2006).

Les orthologues potentiels d'AtCRR4 semblent avoir évolué en parallèle du site *ndhD-1* (Figure 38).

Des orthologues potentiels d'AtDYW1 semblent être présents chez les monocotylédones alors qu'il n'y pas d'édition de *ndhD-1*, leur rôle y serait donc différent.

Les domaines E+ des orthologues potentiels d'AtCRR4 semblent être assez variables chez les dicotylédones.



**Figure 38 : Reconstitution schématique de l'apparition du « complexe » d'édition de *ndhD-1* chez les Angiospermes.**

Chez les monocotylédones, aucun orthologue de CRR4 n'a pu être identifié mais un orthologue potentiel de DYW1 a été identifié (O'Toole et al., 2008). A noter que le codon de la méthionine du transcrit *ndhD* est codé par une ATG et n'est donc pas soumis à l'édition.

Chez les dicotylédones, le site *ndhD-1* est un ACG et doit donc être édité afin de produire la protéine NdhD. Ceci est à corrélérer avec l'apparition d'orthologues potentiels d'AtCRR4. A noter que des orthologues d'AtDYW1 sans motifs PPR n'ont été identifiés que chez les Brassicaceae.

## II. Etude fonctionnelle du candidat AtDYW1

Dans la première partie de cette thèse, nous avons vu que DYW1 était impliquée dans l'édition de l'ARN dans les chloroplastes. Seul le site *ndhD-1* est affecté chez le mutant *dyw1-1*. La protéine PPR-E CRR4 est également indispensable à l'édition de ce site (Kotera et al., 2005). Nous avons également montré l'interaction entre CRR4 et DYW1 dans les chloroplastes montrant ainsi la présence d'un complexe d'édition. La fusion entre CRR4 et DYW1 par leur domaine E a permis la complémentation du double mutant *crr4-dyw1*.

Après cette étude génétique, nous nous sommes penchés sur la caractérisation de la fonction de la protéine DYW1. Alors que CRR4 est le facteur de spécificité permettant la reconnaissance à l'ARN, DYW1 semble être le candidat idéal pour catalyser la réaction d'édition par déamination. En effet, le domaine DYW de cette protéine possède des motifs communs aux cytidines déaminases trouvées chez d'autres espèces ou dans d'autres règnes. Il est question dans cette seconde partie d'analyser, d'une part, la protéine DYW1 afin d'étudier les acides aminés essentiels à l'édition et, d'autre part, reconstituer le complexe d'édition de *ndhD-1 in vivo* et *in vitro*.

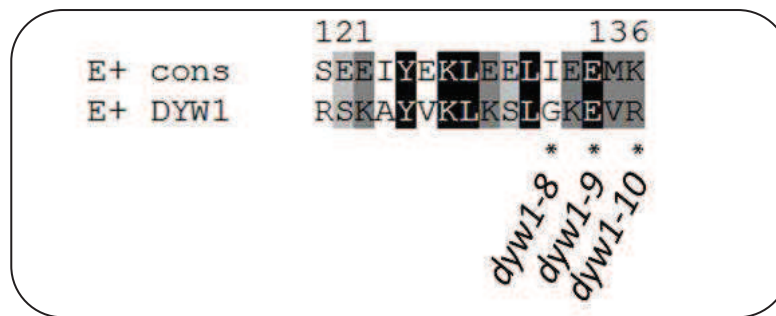
### II.1. Etude des mutants TILLING

#### II.1.1. Obtention de mutants EMS

Comme indiqué dans la première partie de cette thèse, celle-ci a débuté en étudiant les mutants TILLING issus de la mutagenèse EMS à l'ATP. Pour l'étude génétique de DYW1, nous nous sommes focalisés sur *dyw1-1*, ce mutant « perte de fonction » étant parfaitement viable. Cependant, les autres mutants EMS ont été gardés de côté pour les analyser par la suite.

Au total, 15 mutants affectant la séquence protéique ont été obtenus dans l'ORF *DYW1*, baptisés *dyw1-2* à *dyw1-16* selon leur positionnement sur *DYW1* (Figure 24). La mutation *dyw1-2* se situe dans la séquence d'adressage, les mutations *dyw1-3* à *dyw1-7* dans la région PPR-like, *dyw1-8* à *dyw1-10* dans le fragment de domaine E et *dyw1-11* à *dyw1-16* dans le domaine DYW (cf Figure 24, partie I des résultats). Il s'agit ici de mutations ponctuelles ayant modifié des bases G/C en A/T, modifiant la séquence en acides aminés et potentiellement affectant la fonction de DYW1.

Le domaine PPR-like étant assez dégénéré, aucun acide aminé conservé n'est affecté par ces mutations et il semble probable que les mutations dans ce domaine n'affectent pas la fonction de DYW1. Les mutations *dyw1-9* (acide glutamique en lysine) et *dyw1-10* (arginine en lysine) affectent des acides aminés conservés dans le domaine E (Lurin et al., 2004) (Figure 39).

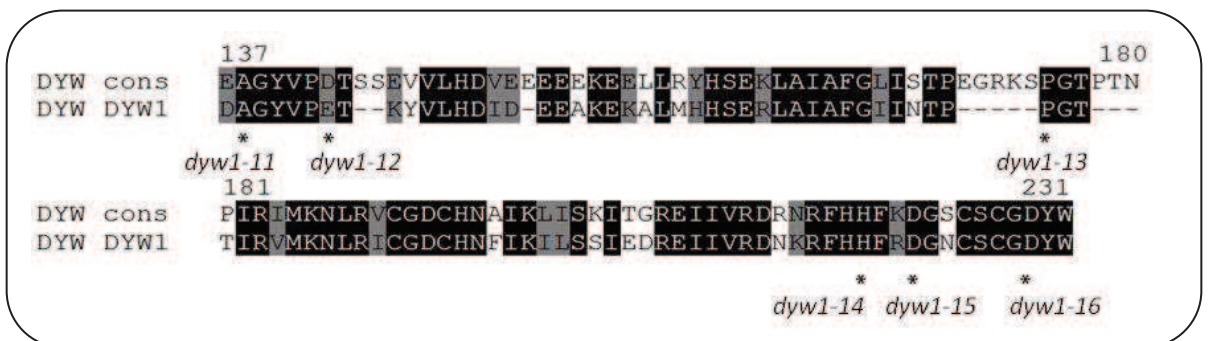


**Figure 39 : Conservation des acides aminés affectés par les mutations EMS dans le domaine E.**

La séquence du domaine E+ de DYW1 a été alignée avec le consensus du domaine E+ de protéines PPR (E+ cons, Lurin et al, 2004) à l'aide du logiciel CLUSTALW2. Les résidus identiques sont surlignés en noir, les résidus similaires en gris et les résidus proches en gris clair.

Les mutations *dyw1-9* et *dyw1-10* affectent des acides aminés présents dans le consensus E+ tandis que *dyw1-8* affecte un acide aminé dont les propriétés sont différentes de l'acide aminé présent dans le consensus.

Dans le domaine DYW, *dyw1-11* (alanine en thréonine), *dyw1-13* (proline en serine), *dyw1-14* (histidine en tyrosine), *dyw1-15* (acide aspartique en asparagine) et *dyw1-16* (acide aspartique en asparagine) affectent des acides aminés très conservés (Figure 40).



**Figure 40 : Conservation des acides aminés affectés par les mutations EMS dans le domaine DYW.**

La séquence du domaine DYW de DYW1 a été alignée avec le consensus du domaine DYW de protéines PPR (DYW cons, Lurin et al, 2004) à l'aide du logiciel CLUSTALW2. Les résidus identiques sont surlignés en noir, les résidus similaires en gris et les résidus proches en gris clair.

Les mutations *dyw1-11*, *dyw1-13*, *dyw1-14*, *dyw1-15* et *dyw1-16* affectent des acides aminés identiques au consensus DYW tandis que *dyw1-12* affecte un acide aminé dont les propriétés sont similaires à l'acide aminé présent dans le consensus.

Nous avons ici des mutants EMS ayant affecté le génome dans son ensemble. Plusieurs générations de rétro-croisements avec Col-0 ont été effectuées, tout comme pour le mutant *dyw1-1*, pour nous assurer que les défauts potentiels observés étaient dûs à la mutation dans le gène et non à une autre mutation dans le génome. Ainsi quatre rétro-croisements ont été effectués et les mutants homozygotes pour la mutation ont été analysés. A noter que tous les mutants analysés ici ont le même phénotype macroscopique que *dyw1-1* et *crr4-3*, à savoir un phénotype de type « sauvage » comme Col-0.



### II.1.2. Analyse de l'édition du site *ndhD-1* chez les mutants EMS

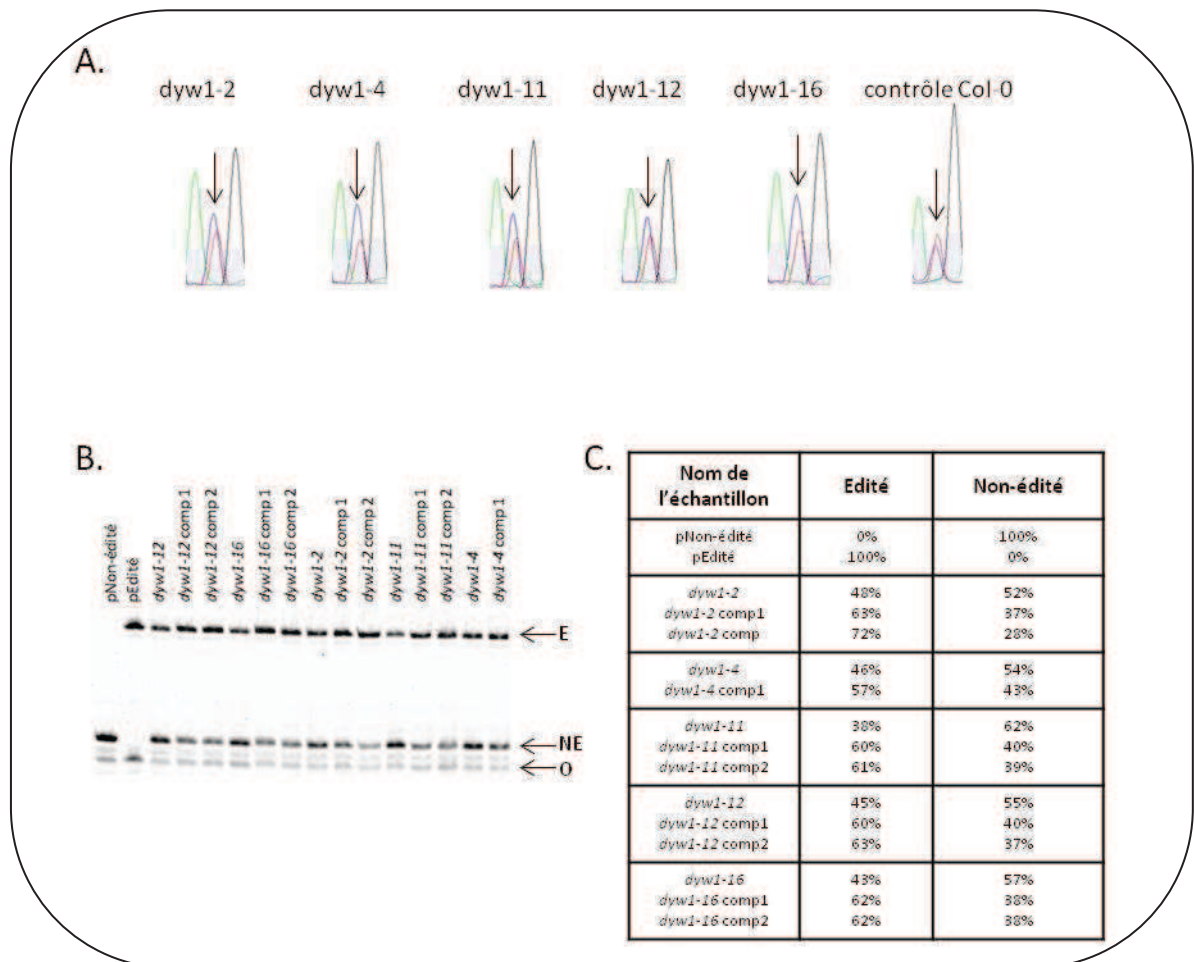
Les ARN des 15 mutants ont été extraits puis les ADNc ont été synthétisés. Une PCR a été effectuée en utilisant les oligonucléotides 214F/214R, permettant l'amplification d'un produit de 450 pb contenant le site d'édition *ndhD-1*. Les produits de PCR ont été séquençés afin de caractériser le niveau d'édition au site *ndhD-1* chez les différents mutants. Les ARN d'au moins trois plantes de chaque mutant ont été extraits afin de pouvoir conclure facilement.

Parmi les 15 mutants étudiés, cinq d'entre eux présentent un défaut d'édition au niveau du site *ndhD-1*. Les dix autres ont un niveau d'édition similaire à une plante Col-0 sauvage (environ 55%). Parmi ces cinq mutants, on trouve *dyw1-2* muté dans la préséquence d'adressage, *dyw1-4* dans le domaine PPR-like, *dyw1-11*, *dyw1-12* et *dyw1-16* mutés dans le domaine DYW. Pour chacun de ces mutants, une diminution de l'édition (et non une abolition comme vu chez *dyw1-1*) a pu être observée (Figure 41A). A l'inverse des plantes « sauvage », il semble ainsi y avoir plus de molécules d'ARN non éditées que de molécules éditées chez ces mutants. Dans le but de quantifier mes observations, j'ai effectué une expérience de PPE sur les mêmes échantillons puis j'ai analysé les produits de la réaction sur un gel de séquençage Li-Cor (Figure 41B). Cette expérience a confirmé que le site *ndhD-1* est toujours édité dans chacun des échantillons mais à un taux inférieur à ce que l'on peut observer chez une plante Col-0 « sauvage ». Le mutant *dyw1-2* (affecté dans la préséquence) montre une très légère diminution de l'édition (48%) tout comme le mutant *dyw1-4* dans le domaine PPR-like (46%). En ce qui concerne les mutants affectant le domaine DYW, *dyw1-11*, *dyw1-12* et *dyw1-16*, la diminution est plus marquée avec respectivement 38%, 45% et 43% d'édition.

Afin de confirmer ces observations, les différents mutants ont été complémentés avec la construction génomique DYW1 ayant permis la complémentation de *dyw1-1* (compF/compR). L'analyse de l'édition de *ndhD-1* a été faite par PPE chez ces mutants complémentés. Une restauration de l'édition du site *ndhD-1* a été observée pour tous les mutants (Figure 41B). Les mutants *dyw1-2*, *dyw1-4*, *dyw1-11*, *dyw1-12* et *dyw1-16* sont restaurés à hauteur respectivement de 63%, 57%, 60%, 60% et 62% soit, dans tous les cas à un niveau supérieur à celui observé chez la plante Col-0 sauvage (55%) (Figure 41C).

Sur un lot de 15 mutations EMS aléatoires et différentes, cinq ont montré une diminution de l'édition du site *ndhD-1* sans jamais l'abolir. Les mutants *dyw1-11*, *dyw1-12* et *dyw1-16* affectent des acides aminés conservés dans le consensus du domaine DYW. Il n'est pas étonnant que ces mutants affectent l'édition de *ndhD-1*.

La fonction de ces acides aminés importants pour l'édition reste inconnue (interaction ? activité ? conformation ?), mais nous savons dorénavant qu'ils ont une influence sur l'édition du site *ndhD-1* par DYW1. Il est clair que chaque acide aminé conservé ne peut avoir une importance capitale pour l'édition, en témoignent les mutants *dyw1-13*, *dyw1-14*, *dyw1-15*, parfaitement édités.



**Figure 41 : Etude de l'édition au site *ndhD-1* dans les lignées EMS affectées dans le gène *DYW1*.**

- Chromatogrammes représentant le site d'édition *ndhD-1* (indiqué par une flèche). Les 5 mutants présentent un défaut d'édition lorsqu'ils sont comparés au contrôle Col-0. Les 10 autres mutants EMS ont été testés de la même façon mais ne montrent aucun défaut d'édition.
- PPE effectuées sur des ADNc provenant des mutants EMS d'intérêt. Les mutants (*dyw1-X*) et les complémentations (*dyw1-X comp*) ont été analysés sur gel d'acrylamide afin de quantifier le défaut d'édition ainsi qu'une éventuelle restauration de l'édition. E: édité, NE: non-édité, O: oligonucléotide.
- Quantification de l'édition pour chaque échantillon effectuée à l'aide de l'outil IQtools. L'intensité des signaux édités et non-édités est quantifiée et un ratio est calculé afin d'obtenir un pourcentage d'édition.

## II.2. Mutagenèse par PCR et complémentation de *dyw1-1*

### II.2.1. Choix des mutants

Il était assez intéressant d'étudier les mutants EMS car ils étaient disponibles à l'URGV. Néanmoins, ces mutants sont peu nombreux et affectent des acides aminés de façon complètement aléatoire.

La méthode de mutagenèse par PCR m'a permis de générer des mutants de *DYW1* à des endroits définis au préalable. Deux types de mutations ont été réalisés : des mutations ponctuelles et des délétions.

Pour les mutations ponctuelles, deux produits de PCR pour chaque mutant ont été synthétisés, le premier contenant le promoteur natif de *DYW1* (compF) et une mutation

contenue dans son primer (compM<sub>x</sub>R), et le deuxième contenant le terminateur de DYW (compR) et une mutation contenue dans son primer (compM<sub>x</sub>F). La fusion de ces deux fragments par PCR a permis la formation d'un gène chimérique muté à l'endroit désiré. Ces formes mutantes de *DYW1* sous le contrôle de leur propre promoteur sont ensuite clonées dans le vecteur pGWB1 de la même façon que pour la complémentation fonctionnelle de *dyw1-1* (cf I.4.2 des résultats). Elles ont ensuite été introduites dans le mutant *dyw1-1* via transformation par la souche C58C1 d'*Agrobacterium tumefaciens*.

Mutations ponctuelles dans DYW1 : 8 mutants ponctuels ont été synthétisés. M1 à M3 ont été mutés dans la signature cytidine déaminase HxE(x)<sub>n</sub>CxxC :

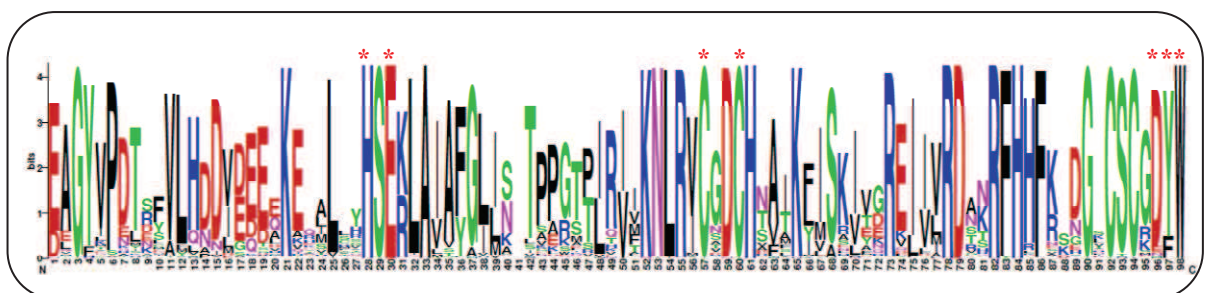
- \*M1 en HxA(x)<sub>n</sub>CxxC,
- \*M2 en AxA(x)<sub>n</sub>CxxC,
- \*M3 en HxE(x)<sub>n</sub>AxxA.

DYW1 a aussi été muté dans des acides aminés très conservés :

- \*M4, le triplet final DYW a été changé en AYW,
- \*M5, le triplet final DYW a été changé en DAW,
- \*M6, le triplet final DYW a été changé en DYA,
- \*M8, une lysine changée en alanine dans le domaine E en position 128,
- \*M9, une sérine changée en alanine dans le domaine DYW en position 202.

Les mutations choisies ont été sélectionnées parmi des acides aminés extrêmement conservés dans le domaine DYW (Figure 42). Parmi les 48 domaines alignés, la signature cytidine déaminase est totalement invariante. Ainsi si le domaine DYW de DYW1 possède bien une activité cytidine déaminase, il est très probable qu'une mutation dans l'un des quatre acides aminés parfaitement conservés de la signature ait des effets importants sur l'édition de *ndhD-1*.

D'autre part, le triplet DYW final est tellement conservé qu'il a donné son nom au domaine DYW (Figure 42). Une telle pression de sélection sur ces trois acides aminés ne peut être synonyme que d'un rôle central dans la fonction du domaine DYW. En effet, le tryptophane final est invariant et l'acide aspartique ainsi que la tyrosine sont très bien conservés.



**Figure 42 : Variation naturelle des domaines DYW dans 500 protéines.**

Le programme « WebLogo » a été utilisé pour aligner les domaines DYW de 500 accessions obtenues à partir du BLAST du domaine DYW de DYW1 sur le site du NCBI.

WebLogo est une représentation graphique de la conservation d'une séquence (nucléique ou protéique)

Plus le score est important (hauteur de chaque lettre), plus l'acide aminé est conservé dans les 500 accessions.

Les astérisques représentent la signature cytidine déaminase HxE(x)<sub>n</sub>CxxC ainsi que le triplet terminal DYW. Programme disponible sur le site: <http://weblogo.berkeley.edu/logo.cgi>

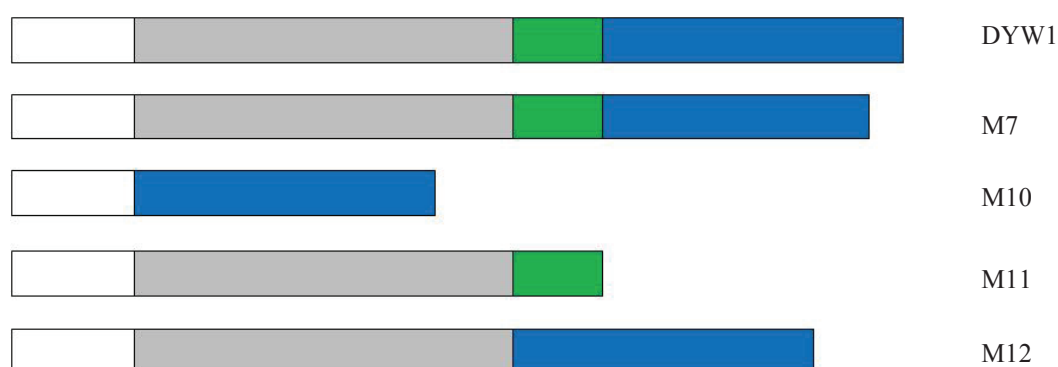
L'idée de la mutation M9 provient d'une publication sur MEF1 une PPR-DYW (Zehrmann et al., 2009). Cette mutation est similaire à celle observée dans le mutant *mef1-2* qui montre une abolition de l'édition aux sites *rps4-956 nad7-963* et un défaut dans *nad2-1160*. Cette mutation ponctuelle de la sérine pourrait supprimer la fonction potentielle d'édition du domaine DYW. Dans DYW1, cette sérine est aussi présente dans le domaine DYW ce qui indique une conservation à travers ces deux protéines.

La dernière mutation touche un acide aminé très conservé dans le consensus du domaine E. Une simple mutation ponctuelle dans le domaine E pourrait-elle déstabiliser la fonction de DYW1?

Dans tous les cas, une alanine a été utilisée pour remplacer l'acide aminé d'intérêt. Dans des études précédentes, chaque changement effectué dans la séquence de déaminases ont été faites en remplaçant l'acide aminé en question par une alanine (Carlow et al., 1996; Sideraki et al., 1996; Opi et al., 2006). De plus, les propriétés physico-chimiques de l'alanine sont éloignées de celle des acides aminés qu'elle remplace. En effet, l'alanine est un acide aminé apolaire au contraire du glutamate (E) et l'aspartate (D) chargés négativement. A la différence de l'alanine, la tyrosine (Y) et le tryptophane (W) sont deux acides aminés aromatiques.

Délétions dans DYW1 (M7, M10-12): 15 à 20 nucléotides ont été sélectionnées de part et d'autre de la zone à cliver pour construire deux oligonucléotides sens et antisens. Ensuite deux produits de PCR ont été synthétisés, l'un à partir des oligonucléotides compF/compM<sub>x</sub>R l'autre des oligonucléotides compM<sub>x</sub>F/compR. Ainsi, 4 mutants ont été générés.

Pour M7, il y a une délétion du triplet DYW, M10 une délétion du domaine PPR-like et du domaine E, M11 une délétion du domaine DYW et M12 une délétion du domaine E (Figure 43).



**Figure 43 : Représentation schématique des délétions effectuées dans DYW1.**

Quatre mutants, dont les différents domaines composant DYW1 ont été supprimés, ont été synthétisés par PCR et clonés par Gateway. Le mutant *dyw1-1* a ensuite été transformé par *Agrobacterium tumefaciens*.

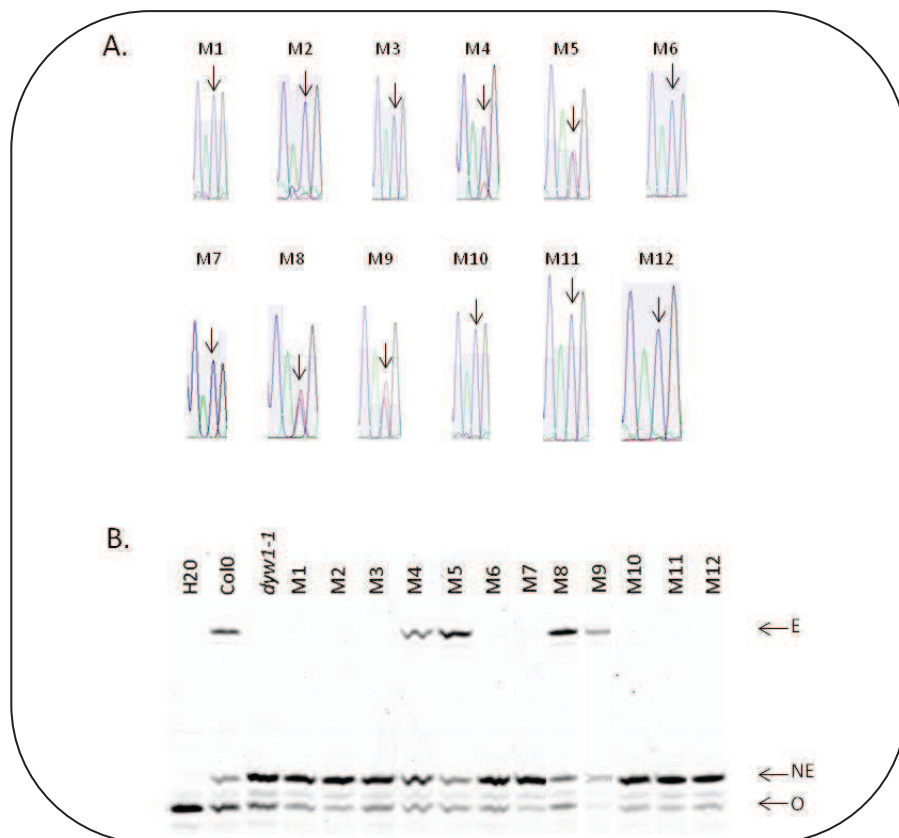
Toutes ces constructions ont été générées de façon à ce qu'elles contiennent le promoteur et le terminateur natifs de DYW1 (fragment identique à celui utilisé pour la complémentation de *dyw1-1*). Le mutant *dyw1-1* a été complétement par *Agrobacterium tumefaciens* contenant les 12 mutants de DYW1. Les constructions ont été effectuées en

Gateway avec pGWB1 comme vecteur de destination. Les graines issues de la transformation ont été sélectionnées sur milieu MS contenant de la kanamycine.

### II.2.2. Analyse de l'édition du site *ndhD-1* pour les mutagénèses dirigées

Les ARN des mutants *dyw1-1* complémentés par les 12 constructions de DYW1 et sélectionnés sur milieu MS ont été extraits puis les ADNc ont été synthétisés. Une PCR a été effectuée en utilisant les oligonucléotides 214F/214R, permettant l'amplification d'un produit de 450pb contenant le site d'édition *ndhD-1*. Les produits de PCR ont été séquencés afin de vérifier le niveau d'édition des différents mutants. Pour chaque complémentation, plusieurs plantes indépendantes ont été sélectionnées et analysées par séquençage et PPE.

Nous savons qu'une plante *dyw1-1* transformée par *Agrobacterium tumefaciens* contenant FL\_DYW1 est complémentée à chaque observation (après sélection sur kanamycine). En effet, sur les 2 plantes analysées nous avons toujours observé une restauration de l'édition du site *ndhD-1*. Pour caractériser la capacité à compléter d'une construction mutante, plusieurs transformants indépendants ont été étudiés. Une absence de complémentation a été conclue quand, après analyse d'au moins 4 transformants, aucune complémentation n'a pu être observée.



**Figure 44 : Etude de l'édition dans la lignée *dyw1-1* complémentée avec des versions mutées de DYW1.**

- A. Chromatogrammes représentant le site d'édition *ndhD-1* (indiqué par une flèche) pour chaque complémentation.
- B. PPE effectuées sur des ADNc provenant des complémentations. E: édité, NE: non-édité, O: oligonucléotide.

Les lignées constructions M10, M11 et M12 portent toutes d'importantes délétions dans DYW1. Aucune des 4 lignées indépendantes pour ces trois constructions n'est capable de compléter le défaut d'édition de *dyw1-1* (Figure 44A). Ceci a été confirmé et quantifié par PPE pour chaque plante (Figure 44B, Tableau 14). Ce résultat implique que chaque domaine de DYW1 possède une importance dans la fonction cytidine déaminase putative ou dans l'interaction protéine-protéine.

Chose assez intéressante, le domaine DYW de DYW1 semble indispensable à l'édition ce qui est très différent des résultats concernant les PPR-PLS CRR22 et CRR28 qui montrent que le domaine DYW de ces deux PPR n'est pas nécessaire à l'édition (Okuda et al., 2009). Il est possible que dans les cas de CRR22 et CRR28 sans domaine DYW, une autre PPR de type DYW (DYW1-like ?) soit recrutée pour catalyser la réaction d'édition (toujours dans l'hypothèse où le domaine DYW contient bien une activité cytidine déaminase). Il semble que dans le cas de DYW1 aucune protéine ne pourrait compléter sa fonction.

Dans les constructions M1, M2 et M3, la signature cytidine déaminase de DYW1 a été affectée. Pour chacune de ces constructions, 6 lignées indépendantes ont été analysées. En utilisant ces trois constructions, nous avons été incapables de restaurer l'édition du site *ndhD-1* (Figure 44B).

Si nous calquons notre modèle à celui de cytidines déaminases trouvées chez d'autres espèces, l'histidine et les deux cystéines seraient impliquées dans la fixation de zinc et le glutamate dans l'activité cytidine déaminase (Betts et al., 1994; Salone et al., 2007). La déamination a lieu par déprotonation de l'eau liée au zinc par le glutamate, suivie par une attaque nucléophile de la cytidine en C4 conduisant à la séparation du groupe NH<sub>2</sub> (Yao et al., 2005). Au vu de nos résultats, nous pouvons dire que ces acides aminés sont indispensables à l'édition et qu'ils supportent l'hypothèse selon laquelle le domaine DYW est une cytidine déaminase. Néanmoins, cette découverte n'est qu'un argument supplémentaire en faveur de notre hypothèse mais ne peut pas être considéré comme une preuve. Des tests d'activité d'édition *in vitro* seront nécessaires pour pouvoir conclure sur la fonction de DYW1.

Dans les constructions M4, M5, M6 et M7, le triplet DYW final a été muté. Le mutant M7, possédant une délétion de ce triplet, est incapable de compléter le mutant *dyw1-1* dans 4 lignées indépendantes indiquant la nécessité de conserver le triplet (Figure 44B).

Les trois acides aminés terminaux ont été modifiés un par un. Chez M4, l'acide aspartique a été modifié en alanine. Cette dernière construction complète partiellement l'édition du site *ndhD-1* du mutant *dyw1-1* montrant une certaine importance du résidu pour l'édition mais pas une nécessité absolue. En effet, la restauration atteint un taux de 20 à 30% pour 3 lignées indépendantes contrairement aux 55% observés généralement chez Col-0 (Tableau 14). Ce résultat peut être corrélé au taux d'édition observé pour le mutant EMS *dyw1-16*, dont l'acide aspartique est mutée en asparagine, qui présente un taux d'édition de variant de 20 à 30%. Deux mutations Pour M5, la tyrosine a été mutée. Dans ce cas, l'édition du site *ndhD-1* est restaurée chez trois plantes indépendantes montrant qu'un changement de tyrosine en alanine à cette position est sans effet pour l'édition et donc que la tyrosine n'est pas nécessaire à

l'édition dans le cas de DYW1. Enfin dans le cas de M6, c'est le tryptophane qui a été changé. La transformation avec cette construction ne permet pas de restaurer l'édition de *ndhD-1* dans les 3 plantes testées ce qui montre que le dernier acide aminé de DYW1 est indispensable à l'édition (Tableau 14).

Enfin, chez les mutants *dyw1-1* complémentés par M9 et M8, nous pouvons observer une complémentation de l'édition du site *ndhD-1*, pour respectivement 3 plantes et 5 plantes indépendantes (Figure 44A). La mutation de la sérine en alanine chez M9 n'affecte pas la fonction de DYW1 qui édite toujours le site *ndhD-1*. Ceci est contradictoire avec les résultats obtenus chez *mef1-2*. Néanmoins, il semble que le choix de changer une sérine en alanine n'ait pas été le meilleur. Il aurait été plus judicieux de changer la sérine en phénylalanine tout comme pour *mef1-2*, cette modification constituant un changement majeur dans la propriété de l'acide aminé. Pour M8, un crible plus important sur le domaine E devra être fait pour modifier chaque acide aminé de ce domaine. En effet, la modification d'un seul acide aminé sur 16 n'a pas affecté la fonction même en se focalisant sur un acide aminé relativement bien conservé.

Nous avons identifié par cette méthode plusieurs acides aminés qui semblent indispensables à l'édition du site *ndhD-1* chez DYW1, les constructions contenant la mutation de l'un des résidus formant la signature cytidine déaminase ou du tryptophane en C-terminal étant incapables de restaurer l'édition.

**Tableau 14 : Quantification de l'édition pour les 12 complémentations.**

Mutation	Nom	Pourcentage d'édition U/C du site <i>ndhD-1</i>					
		Plante 1	Plante 2	Plante 3	Plante 4	Plante 5	Plante 6
HxA(x)nCxxC	M1	0	0	0	0	0	0
AxA(x)nCxxC	M2	0	0	0	0	0	0
HxE(x)nAxxA	M3	0	0	0	0	0	0
DYW => AYW	M4	30	20	20			
DYW => DAW	M5	55	55	50			
DYW => DYA	M6	0	0	0	0	0	0
Del. DYW triplet	M7	0	0	0	0	0	0
L <sub>128</sub> => A (E)	M8	35	55	48	56	58	
S <sub>202</sub> => A (DYW)	M9	65	55	68			
Del. PPR-like	M10	0	0	0	0		
Del. DYW	M11	0	0	0	0		
Del. E+	M12	0	0	0	0		
-	FL_DYW1	55	50				

Plusieurs plantes indépendantes ont été sélectionnées pour chaque transformation:

- au moins 4 pour les plantes où aucune complémentation n'a pu être observée.
- au moins 3 pour les plantes où l'édition de *ndhD-1* est restaurée.

### II.3. Etude de l'interaction CRR4-DYW1 par complémentation du double mutant *crr4-3x dyw1-1*

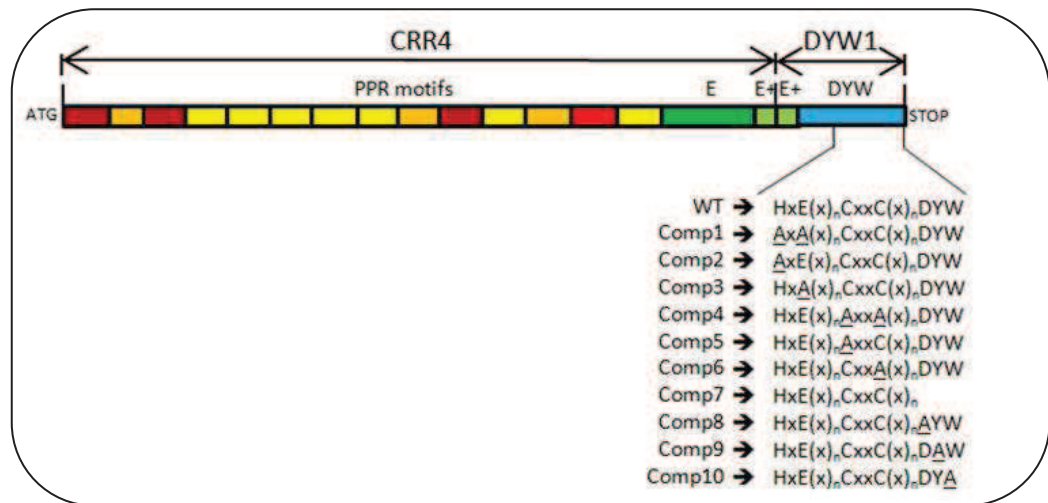
Nous avons montré précédemment qu'une fusion effectuée entre CRR4 et les domaines E-DYW de DYW1 permettait la complémentation du double mutant *crr4-3x dyw1-1*. De plus, nous savons dorénavant que certains acides aminés sont indispensables à la fonctionnalité de DYW1. L'idée ici a été de coupler ces deux résultats dans le but d'étudier si ces acides aminés étaient impliqués dans l'interaction avec CRR4.

Avec la fusion CRR4-DYW1, nous forçons l'interaction entre ces deux protéines, c'est-à-dire que les acides aminés indispensables à l'interaction ne sont plus nécessaires dans ce cas. J'ai donc utilisé la fusion CRR4-DYW1 et introduit des mutations ponctuelles afin de modifier les acides aminés essentiels à la fonction de DYW1. Pour ce faire, j'ai modifié :

- la signature cytidine déaminase HxE(x)<sub>n</sub> CxxC (comp1 à comp6),
- le triplet terminal DYW (comp7 à comp10) (Figure 45).

Si les acides aminés ciblés sont impliqués dans l'interaction entre CRR4 et DYW1, les modifier ne devrait pas poser de problème à la restauration de l'édition. Au contraire, si ces derniers ne sont pas impliqués, la construction ne devrait pas être capable de restaurer l'édition.

A l'heure où j'écris ces lignes, la complémentation du double mutant est effectuée pour chaque construction mais la sélection n'est pas encore faite.



**Figure 45 : Mutants de la fusion CRR4-DYW1 produits pour complémenter *crr4-3xdyw1-1*.**

Le double mutant *crr4-3xdyw1-1* produit précédemment a été complémenté avec 10 mutants de fusion CRR4-DYW1 différents.

Les inserts ont été clonés par Gateway dans le vecteur pGWB2 (promoteur 35S présent). Le double mutant a été transformé par *Agrobacterium tumefaciens* (C58C1).

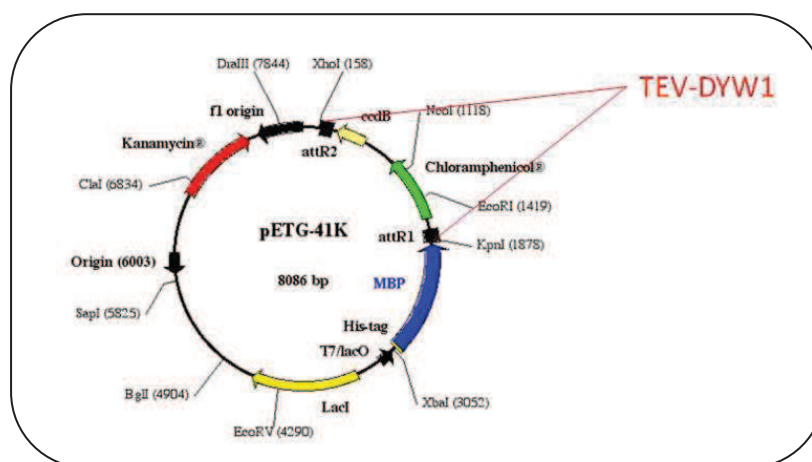


## II.4. Recherche de l'interaction entre DYW1 et les ions métaux

### II.4.1. Préparation de l'expérience

Nous avons vu précédemment que la signature cytidine déaminase HxE(x)<sub>n</sub>CxxC était indispensable à la restauration de l'édition dans le mutant *dyw1-1*. D'après la bibliographie, ce domaine chez les cytidines déaminases contient des acides aminés permettant la fixation des ions zinc permettant la catalyse de la réaction d'édition (Betts et al., 1994; Yao et al., 2005). Nous avons décidé de tester si la protéine DYW1 avait effectivement une affinité pour le zinc voire pour d'autres ions métaux. Lors des expériences de chromatographie d'affinité, nous avons vu que DYW1 restait fixée sur les colonnes cobalt ce qui pourrait signifier que DYW1 a une affinité pour un certain nombre d'ions métaux (cations divalents).

Pour tester cette hypothèse, nous avons analysé des fractions de protéines DYW1 solubles par ICP-MS (spectrométrie par torche à plasma). Le gène *DYW1* a été cloné dans le vecteur pETG41K et la protéine a été exprimée dans la souche Rosetta d'*E. coli*. L'induction de l'expression a été effectuée sur une culture de 500 ml à 16°C pendant 20h avec de l'IPTG. Le vecteur pETG41K permet l'ajout en N-terminal de la protéine d'intérêt de 40kDa supplémentaires correspondant aux deux étiquettes MBP (pour « maltose binding protein ») et His (hexahistidine) (Figure 46). Généralement, l'étiquette MBP placée en N-terminal de la protéine aide à la solubilisation de cette dernière. Le site de clivage TEV a aussi été cloné en N-terminal de la protéine à exprimer, entre la protéine et les étiquettes MBP-His. Ceci permet de pouvoir cliver la protéine après purification afin d'éliminer les étiquettes dont la très grande taille pourrait gêner l'interaction avec les ions.



**Figure 46 : Vecteur d'expression pETG41K.**

L'induction à l'IPTG permet la production de la protéine DYW1 (dont la séquence codante a été préalablement clonée en Gateway), marquée en N-terminal par une étiquette 6xHistidine et MBP. Un site de clivage TEV a été inséré de sorte à éliminer les marqueurs après purification.

La protéine DYW1 fusionnée aux étiquettes a été caractérisée sur gel d'acrylamide. Une protéine soluble de 65 kDa correspondant à la taille attendue pour une fusion DYW1-MBP-His (40 kDa d'étiquettes + 25 kDa DYW1) a été obtenue. Les étiquettes sont ensuite clivées à l'aide de la protéase TEV. Après clivage, les échantillons sont à

nouveau caractérisés sur gel pour d'une part vérifier que le clivage a eu lieu et d'autre part tester la solubilité de la protéine DYW1 après clivage de la MBP. L'échantillon est ensuite dialysé afin d'éliminer les contaminants en solution.

Le zinc étant un composant chimique essentiel aux être vivants, celui-ci se trouve en quantité importante dans les organismes. Au lieu d'ajouter le zinc directement en solution ou dans le milieu de culture ce qui peut ralentir à fortes concentrations la croissance des bactéries (Yao et al., 2005), nous avons proposé que les ions zinc puissent être apportés directement par la bactérie. Si les ions zinc sont fixés à DYW1, nous devrions pouvoir les identifier par ICP-MS.

L'étape d'analyse des échantillons a été effectuée dans le laboratoire de TSW Analytical Pty Ltd par Sven Fjastad (Level 2, Myers Street Building, University of Western Australia, Myers Street, CRAWLEY, Western Australia, 6009).

#### II.4.2. Difficultés rencontrées et résolution

Plusieurs questions se sont posées lors de la préparation des échantillons pour effectuer une ICP-MS :

- (1) comment produire des protéines de fusion solubles ?
- (2) quels contrôles utiliser ?
- (3) comment purifier DYW1 après clivage ?

(1) Les protéines PPR sont souvent obtenues de façon insoluble lorsqu'elles sont surexprimées en bactéries. La protéine DYW1 ne contenant pas de domaine PPR, elle est néanmoins moins difficile à solubiliser. Nous avons effectué l'élution de la colonne nickel avec un tampon contenant une forte concentration en sel (500 mM) et effectué l'induction à 16°C sur des cultures ayant une densité optique comprise entre 0,3 et 0,4.

(2) Dans cette expérience, plusieurs contrôles ont été ajoutés :

- DYW1 mutée dans la signature cytidine déaminase  $HxE(x)_nC_{xx}C$ ,
- cytidine déaminase d'*E. coli* (contrôle positif),
- élution d'une souche Rosetta non induite pour estimer les contaminants en solution (contrôle négatif).

(3) Après clivage avec la protéase TEV, quatre différents produits sont attendus (par ordre de poids moléculaire décroissant) :

- MBP-His-DYW1 non clivée de 62 kDa, car la protéase a une efficacité de 80% d'efficacité seulement,
- MBP-His, étiquettes séparées de DYW1 ayant un poids de 40 kDa,
- la protéase His-TEV ayant un poids moléculaire de 27 kDa,
- la protéine d'intérêt DYW1, clivée ayant un poids moléculaire de 22 kDa.

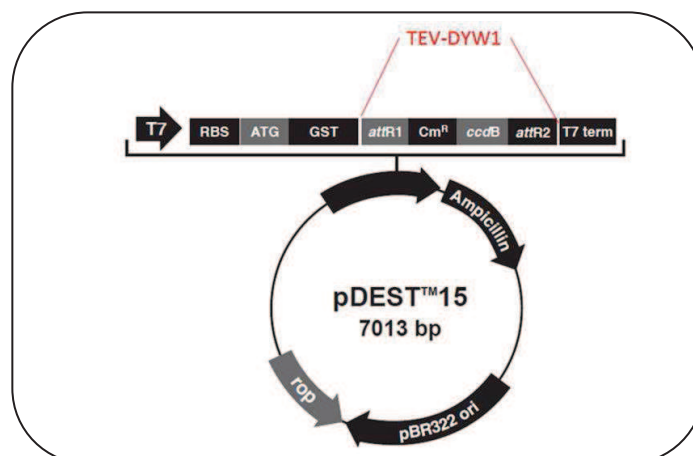
Il a donc fallu éliminer les trois autres produits car ils possèdent tous une étiquette 6xHis et donc une affinité avec les cations divalent. Nous avons tout d'abord pensé à effectuer une chromatographie d'exclusion stérique, séparant les différents produits composant la solution en fonction de leur taille. Il serait assez aisé d'éliminer les deux produits de haut poids moléculaire de 62 et 40 kDa. Il nous a été impossible de séparer par cette méthode le 6xHis-TEV et la protéine DYW1 qui n'ont que 5 kDa de différence.

J'ai donc tenté d'éliminer les trois produits par purification sur billes de nickel dans un tube de 1,5ml. DYW1 (ne contenant pas d'étiquette) ne devrait pas être retenue par les billes et serait obtenue dans le surnageant. Je n'ai pas réussi à obtenir DYW1 dans le surnageant, cette dernière étant présente dans les billes. Il est possible que DYW1 se soit fixée à la colonne nickel en raison de son affinité potentielle pour les cations divalents ou alors que DYW1 soit précipitée au fond du tube.

Même si cette dernière expérimentation avait marché, la purification sur billes de nickel aurait inévitablement libéré dans la solution des ions métaux contaminant  $Ni^{2+}$  qu'il aurait fallu éliminer par des étapes supplémentaires de chélation avec de l'EDTA puis de dialyse pour éliminer l'EDTA. Ces nombreuses étapes auraient risqué d'affecter l'activité et l'intégrité de DYW1 (réduction de la solubilité, diminution de la concentration...).

Nous sommes parvenus à la conclusion que l'utilisation d'une étiquette histidine cause plus de problème qu'elle n'en résout dans le cas de DYW1. Il a donc été nécessaire de réfléchir à de nouvelles constructions mettant en jeu d'autres étiquettes.

J'ai donc modifié les constructions afin d'utiliser le marqueur GST pour purifier la protéine (Figure 47). Cette étiquette, n'ayant pas d'affinité particulière pour les ions métaux, peut être conservée après élution. Ceci permet de diminuer le nombre d'étapes de purification et ainsi éviter une perte de la solubilité ou d'intégrité de la protéine. Par ailleurs, je savais que le marqueur GST avait déjà été utilisé avec succès pour exprimer une protéine PPR dans la littérature. En effet, la protéine de maïs MPPR6 fusionnée à la GST a été utilisée pour effectuer une expérience de retard sur gel, ce qui nécessite une protéine soluble (Manavski et al., 2012). L'ORF *DYW1*, clonée dans pDEST15, a permis d'exprimer la protéine dans la souche bactérienne Rosetta2. La purification a été réalisée sur billes de glutathion. Aucun clivage protéique n'a été nécessaire par la suite.



**Figure 47 : Vecteur d'expression pDEST15.**

L'induction à l'IPTG permet la production de la protéine DYW1 (dont la séquence codante a été préalablement clonée en Gateway), marquée en N-terminale par un marqueur GST. Un site de clivage TEV a été inséré de sorte à éliminer les marqueurs si nécessaire après purification.

En utilisant ce nouveau vecteur, les protéines suivantes ont été produites et purifiées :

- \* GST-DYW1 (47,3 kDa),

\* GST-DYW1 dont la signature cytidine déaminase est mutée (AxA(x)nCxxC) (47,3 kDa),

\* GST-DYW1 dont toute la signature cytidine déaminase a été supprimée (32aa supprimés) (43,8 kDa),

\* GST-NP\_416648, une cytidine déaminase présente dans *E. coli* (témoin positif) (56,8 kDa).

Ces protéines ont été purifiées sur colonnes de glutathion puis dialysées dans un tampon PBS pour éliminer le glutathion utilisé pour l'élution (interagissant potentiellement avec les ions métaux). J'ai obtenu pour chaque purification 3 ml de protéines purifiées à une concentration d'environ 0,3 mg/ml. Les échantillons ont été envoyés à Michael Challenor pour analyse en ICP-MS.

#### II.4.3. Résultat de l'ICP-MS

L'ICP-MS est utilisée notamment pour l'analyse de métaux à des concentrations de l'ordre de la partie par trillion (ppt) présents dans une solution donnée. Les échantillons sont ionisés par une torche à plasma et analysés par un spectromètre de masse.

L'avantage de cette technique est sa très grande sensibilité mais, de ce fait, elle est aussi extrêmement sensible aux contaminants provenant des échantillons et de la verrerie. De plus, l'utilisation de certains sels à des concentrations de plus de 0,1 M peuvent interférer avec le signal et minimiser la quantité de zinc détectée (Olivares and Houk, 1986). Cette technique est « destructive » c'est-à-dire que les échantillons doivent être dissous dans l'acide préalablement à l'ionisation ce qui les rend inutilisables par la suite.

Dans un premier temps, le tampon de PBS seul a été analysé afin d'avoir une idée de la quantité de zinc contaminant présente dans l'eau filtrée. Le PBS a été préparé dans un bécher en verre ou un bécher en plastique puis mis dans un Falcon en plastique. La quantité de zinc mesurée est restée inférieure à 80 ppt dans tous les échantillons témoins. L'eau et les produits chimiques sont donc purs et le type de bécher utilisé n'a pas d'influence sur les résultats.

Les quatre protéines à tester ont été purifiées et leurs concentrations ont été estimées à l'aide d'un Nanodrop mesurant la densité optique à 280 nm en utilisant le coefficient d'extinction de la protéine. La concentration des protéines mesurée dans les 4 échantillons était comprise entre 0,13 mg/ml et 0,24 mg/ml.

Par ICP-MS, nous avons caractérisé la quantité de zinc de chaque échantillon. Tandis que les formes mutées de DYW1 contiennent une quantité de zinc inférieure à 7 ppb (respectivement 6,58 ppb et 4,52 ppb), le contrôle positif NP\_416648 et DYW1 contiennent des quantités supérieures à 40 ppb (respectivement 83,1 ppb et 45,3 ppb) (Tableau 15). Nous pouvons donc voir une différence (d'un facteur 10) entre la quantité de zinc dans l'échantillon DYW1 et dans ceux correspondants à ses formes mutées suggérant une fixation du zinc à la protéine DYW1 par le biais de sa signature CDA.

Le ratio molaire zinc/protéine a ensuite été calculé. Chez DYW1 et NP\_416648, le ratio est supérieur aux mutants de DYW1 (Tableau 15). Les ratios de DYW1, NP\_416648, DYW1\_HE et DYW1ΔCDA sont respectivement de 0,16, 0,29, 0,024 et 0,019.

NP\_416648 étant une CDA connue fixant un atome de zinc, nous attendions un ratio proche de 1 or le ratio calculé ici est trois fois inférieur à ce qui est prévu (0,29).

**Tableau 15 : Résultat de l'ICP-MS.**

Echantillon	MM (Da)	Concentration en Zn (ppb)	Zn (mol/L)	Protéine (mol/L)	Ratio molaire [Zn]/[Protéine]
GST DYW1 ΔCDA	43800	6,58	1,00596E-07	5,17504E-06	0,019
GST-DYW1 HE	47300	4,52	6,91026E-08	2,88936E-06	0,024
GST-NP CDA	56800	83,1	1,27045E-06	4,34272E-06	0,293
GST-DYW WT	47300	45,3	6,92555E-07	4,22833E-06	0,164
PBS	0	<0.100	1,52882E-09	0	—

Les échantillons (étiquetés avec la GST) analysés sont les suivants :

GST DYW1 ΔCDA : DYW1 dont toute la signature cytidine déaminase a été supprimée

GST-DYW1 HE: DYW1 dont la signature cytidine déaminase est mutée (AxA(x)nCxxC)

GST-NP CDA: cytidine déaminase présente dans *E. coli* (NP\_416648)

GST-DYW WT: DYW1 forme sauvage

PBS : tampon de dialyse PBS 1X

MM : Masse moléculaire en Dalton (Da).

Il semble donc qu'il y'ait un problème dans le calcul du ratio molaire final ou que seulement un tiers des protéines aient fixé le zinc.

En effet, des volumes égaux (15µl) de protéines purifiées ont été analysés sur gel d'acrylamide et il est clair que la quantité de protéines NP\_416648 est largement supérieure à celle des autres protéines. Il serait judicieux de recalculer la concentration en protéines. Néanmoins, la technique d'ICP-MS étant destructive, il m'a été impossible de récupérer les échantillons pour tester à nouveau les concentrations.

De plus, une perte de signal du produit à analyser peut être observé lorsque les concentrations en sel sont trop importantes (Olivares and Houk, 1986). Or, il a été utilisé dans cette expérience un tampon composé de 0,1 M de NaCl. Il semblerait que cette concentration en sel pourrait diminuer de moitié la qualité du signal, sous-évaluant ainsi la quantité de zinc en solution.

## II.5. Reconstitution de l'édition *in vivo* et *in vitro*

Mon travail de thèse, ainsi que des résultats antérieurs de l'équipe ou des données publiées, apportent de très nombreux arguments en faveur de l'hypothèse selon laquelle le domaine DYW des protéines PPR porterait le site catalytique permettant l'édition.

Ainsi, nous savons que :

- le domaine DYW contient des structures similaires aux cytidines déaminases connues chez d'autres espèces (Salone et al., 2007; Iyer et al., 2011),

- CRR4 (une PPR-E) nécessite une interaction avec la protéine DYW1 (contenant un domaine DYW) pour éditer sa cible,

- les résidus correspondant à la signature cytidine déaminase chez DYW1 sont nécessaires à l'édition du site *ndhD-1*,

- le domaine DYW de DYW1 est nécessaire à l'édition de *ndhD-1*,

- par TAP-tag, nous avons vu que CLB19 pouvait interagir avec OTP100, une PPR-P sans domaine cytidine déaminase, et DYW2, une protéine PPRDYW ressemblant à DYW1.

- le domaine DYW semble fixer des molécules de zinc (cf II.4.3 des résultats).

A priori, tout porte à croire que la présence d'un domaine DYW, associé en cis (PPR-DYW) ou en trans (par exemple CRR4 et -DYW1), est nécessaire à l'édition des transcrits. Néanmoins, malgré les arguments cités plus haut, nous n'avons pas de preuve directe et indiscutable que DYW1 est l'enzyme permettant l'édition de *ndhD*-1.

Seule une reconstitution (*in vivo* ou *in vitro*) de l'activité d'édition pourrait apporter cette preuve. C'est avec cet objectif ambitieux que j'ai mené des expériences de reconstitution d'activité dans la dernière partie de ma thèse. De telles expériences avaient déjà été tentées par Véronique Salone au laboratoire mais nous pensions que les données nouvelles sur le sujet pouvaient permettre de faire de meilleurs choix stratégiques pour ces essais. En particulier, grâce à mes résultats, je disposais de la connaissance de la nécessité de 3 composants pour reconstituer l'édition : les protéines CRR4 et DYW1 et la cible ARN *ndhD*.

### II.5.1. *In vivo*

Pour réaliser ces essais *in vivo*, deux systèmes d'expression ont été utilisés :

(1) Expression en bactérie : dans cet essai, *CRR4* et *DYW1* ont été clonés dans le vecteur d'expression pETDUET-1 contenant deux sites multiples de clonage. D'autre part, une partie du gène *ndhD* chevauchant le site *ndhD*-1 (123 pb avant et 150 pb après le site d'édition) a été clonée dans le vecteur pCOLADUET-1 dans lequel le site de fixation des ribosomes RBS a été supprimé pour éviter toute traduction. Le but ici était de reconstituer le complexe minimal d'édition dans un autre système. A noter que lorsque j'ai réalisé ces constructions, la publication concernant les protéines MORF n'était pas publiée. Les trois membres du complexe ont été placés dans ces vecteurs sous le contrôle du promoteur T7. La souche BL21(DE3) a été transformée par ces constructions. Cette souche contient le lysogène  $\lambda$ DE3 qui permet l'induction de l'expression de la T7 ARN polymérase en réponse à un traitement à l'IPTG.

L'induction à l'IPTG a été réalisée à 16°C afin de produire des protéines solubles (des agrégats de protéines sont souvent produits lors d'une surexpression à température plus haute). Les protéines totales de BL21 ont été extraites pour vérifier l'expression des deux protéines sur gel d'acrylamide. Les ARN totaux de BL21 ont été extraits et les ADNc ont été synthétisés. Une PCR a été effectuée afin d'amplifier les ADNc recombinants correspondant à *ndhD*. Le niveau d'édition des transcrits a été analysé par PPE à l'URGV. Malheureusement, dans ces expériences, aucune édition de la cible n'a pu être mise en évidence. Les hypothèses pour expliquer cet échec sont très nombreuses. Parmi elle, on peut proposer l'absence d'un ou de plusieurs membre(s) du complexe indispensable à l'édition (par exemple une ou des MORF). D'autre part, il est possible que les protéines exprimées en bactéries ne soient pas fonctionnelles (mauvaise conformation des protéines, absence de maturation post-traductionnelle, etc).

(2) Expression en levure : la reconstitution d'un complexe d'édition des mammifères a été effectuée en levure il y a 10 ans (Lellek et al., 2002). Dans cette étude, l'activité d'édition a été reconstituée en exprimant l'enzyme d'édition APOBEC-1 avec la protéine ACF et une partie d'APOB dont le transcrit doit être édité. Chez l'homme, *apoB* est partiellement édité et conduit à la formation d'un codon stop dans l'ARN. *APOB* a été cloné en aval du gène rapporteur *GAL4* et en amont de *GAL80* (inhibiteur de *GAL4*) en phase. Ainsi, *GAL4* est toujours exprimée mais lorsqu'*apoB* n'est pas édité (sans codon stop), la protéine *GAL80* est produite et inhibe *GAL4*. Si *apoB* est édité, un codon stop est introduit de sorte à ce que *GAL80* ne soit pas traduite. Les auteurs ont ainsi réussi à montrer une activité d'édition *in vivo* d'*apoB* lorsqu'*APOBEC-1* et *ACF* sont exprimés.

J'ai décidé de reproduire cette expérience en exprimant *CRR4*, *DYW1* et *ndhD-1* dans la souche de levure MAV203. A l'inverse d'*apoB*, dans le cas de *ndhD*, l'édition conduit à la formation d'un codon d'initiation de la traduction. Soixante nucléotides de *ndhD* contenant la cytidine à éditer ont été clonés en amont d'un gène rapporteur *GFP* (pour « Green Fluorescent Protein »). Si *ndhD-1* est édité, un site d'initiation de la traduction est créé permettant l'expression d'une fusion *NdhD-GFP* (Tableau 16). Cette protéine fluorescente sera recherchée par microscopie.

Tableau 16 : Principe de l'expression *in vivo* en système levure.

Expression en levure	Résultat attendu
<p><u>*témoins négatifs :</u></p> <p>CRR4 ACG ———— NdhD-1 ———— GFP ———— ARN</p> <p>DYW1 ACG ———— NdhD-1 ———— GFP ———— ARN</p>	<p>Pas de traduction de la GFP =&gt; pas de fluorescence</p>
<p><u>*témoin positif:</u></p> <p>AUG ———— NdhD-1 ———— GFP ———— ARN</p>	<p>Traduction de la GFP =&gt; fluorescence (100% GFP traduite)</p>
<p><u>*Test :</u></p> <p>CRR4 DYW1 ACG ———— NdhD-1 ———— GFP ———— ARN ↓ AUG</p>	<p>Traduction de la GFP ? =&gt; fluorescence ? (% GFP traduite dépendant de l'efficacité de l'édition)</p>

La traduction de la GFP est dépendante de la méthionine potentiellement présente dans le transcrit *ndhD-1*. Dans le cas des témoins négatifs, seule l'une des deux protéines d'intérêt est présente ce qui rend théoriquement impossible l'édition de l'ACG en AUG.

Le contrôle positif possède un transcrit d'ores et déjà « édité » (l'AUG est présent).

Le test possède les deux protéines d'intérêt pour éditer l'ACG en AUG. Une fois le site *ndhD-1* édité, la GFP est traduite. La quantité de protéine sera très certainement inférieure au contrôle positif.

Différents plasmides me permettant l'expression chez les levures ont été obtenus : pYES2-Dest52 (Ura), pYes2-Dest52 (Leu) and pYes3/CT. Ces trois plasmides sont compatibles en levure et contiennent un promoteur inductible au galactose. *CRR4* a été cloné dans pYES2-Dest52 (URA), *DYWI* dans pYes2-Dest52 (Leu) et *ndhD* dans pYes3/CT. La souche MAV203 a été transformée par ces trois plasmides (test) ou par pYes3/CT avec pYES2-Dest52 (URA) ou pYES2-Dest52 (LEU) (contrôles négatifs). Une version de *ndhD* édité a été clonée dans pYes3/CT ce qui a pour but d'avoir directement un codon d'initiation de la traduction en phase avec le gène codon la GFP (contrôle positif de fluorescence).

J'ai aussi reçu des plasmides de la part de Marc Blondel (INSERM 1078, Brest) : p413, p414, p415, p416. Ces plasmides sont compatibles et contiennent l'un ou l'autre de ces deux promoteurs constitutifs, ADH (promoteur « moyen ») et GPD (promoteur « fort »). Ici, *CRR4* a été cloné dans p416 (ADH/GPD), *DYWI* dans p414 (ADH/GPD) et *ndhD* dans p415 (ADH/GPD). La souche MAV203 a été transformée par les trois plasmides (un test avec le promoteur ADH, un autre test avec le promoteur GPD) en même temps ou seulement avec p415 *ndhD* et p416 *CRR4* ou p414 *DYWI* (contrôles négatifs). De même que pour les plasmides pYES, une version de *ndhD* éditée a été clonée dans p415 (ADH/GPD) (contrôle positif).

Aucune fluorescence des levures n'a pu être observée au microscope à l'exception des contrôles positifs de fluorescence p415 *ndhD* GPD et pYes3/CT *ndhD* (p415 *ndhD* ADH n'a pas fonctionné). Ce résultat laisse penser qu'il nous manque un ou plusieurs facteurs pour permettre l'édition de *ndhD-1 in vivo*. En particulier, à nouveau, les protéines MORF n'étaient pas connues au moment où ces expériences ont été réalisées et elles n'ont pas été ajoutées au système.

### II.5.2. In vitro

Même si les expériences d'édition *in vivo* fonctionnaient, ce ne serait pas la preuve irréfutable que le complexe DYW1 est, à lui seul, responsable de l'édition observée. En effet, il serait toujours possible qu'un élément composant la bactérie ou la levure s'associe au complexe et permette l'édition de l'ARN. Pour cette raison, mais également pour tenter un autre système de test de l'activité, des essais d'édition *in vitro* ont été effectués. Il s'agit de faire agir des protéines purifiées directement sur l'ARN à éditer en présence d'un tampon.

Pour guider nos choix, nous ne disposons d'aucun protocole similaire (même chez un autre modèle). En revanche, nous avons utilisé les protocoles décrits dans quelques publications ayant reconstitué l'édition grâce à un tampon composé d'extraits de chloroplastes ou de mitochondries contenant tous les facteurs nécessaires à l'édition (Hirose and Sugiura, 2001; Hegeman et al., 2005; Hayes and Hanson, 2007b; Takenaka and Brennicke, 2007).



La première tentative effectuée a été réalisée à partir de l'élution du complexe TAP-tag contenant CLB19, DYW2 et OTP100 (VIB, Gand, Belgique). *RpoA* contenant le site d'édition *rpoA* a été cloné dans un vecteur pGEMT-easy et transcrit *in vitro* afin d'obtenir un ARN de plus de 250pb (123 pb avant le site d'édition et 150 pb après). Le complexe a été mis en présence de l'ARN à éditer et du tampon utilisé pour l'édition *in vitro* en présence du complexe TAP-tag. Après la réaction, l'ADNc comprenant le site *rpoA* a été synthétisé et une PCR utilisant des amorces spécifiques de l'ADNc a été effectuée. Les produits de RT-PCR ont été soumis à une réaction de PPE et observés sur gel de séquençage. Les conditions utilisées n'ont pas permis de montrer une activité d'édition du complexe CLB19-DYW2-OTP100 sur le transcrit *rpoA in vitro*. Il est possible que ceci soit dû à une instabilité de l'activité du complexe. En particulier, nous avons dû aller chercher le complexe en Belgique ce qui implique qu'il n'a pas forcément été stocké dans des conditions optimales.

Aucun complexe TAP-tag n'étant disponible pour DYW1, nous avons choisi de produire les différentes protéines du complexe séparément dans la souche Rosetta d'*E. coli*, une souche dérivée de BL21 permettant l'expression de protéines recombinantes eucaryotes. Cette souche permet l'induction de gènes placés sous le contrôle du promoteur T7 avec de l'IPTG. *CRR4*, *DYW1*, *MORF2* et *MORF9* ont été clonés par Gateway dans le vecteur pETG41K, contenant les marqueurs MBP et His. L'étiquette MBP placée en N-terminal de la protéine aide généralement à la solubilisation de cette dernière ce qui peut être un avantage considérable lorsque l'on exprime des PPR. Le site de clivage TEV a aussi été cloné en N-terminal de la protéine à exprimer, entre la protéine et les étiquettes MBP-His. Ceci permet de pouvoir cliver la protéine après purification afin d'éliminer les étiquettes dont la très grande taille pourrait gêner l'activité des protéines.

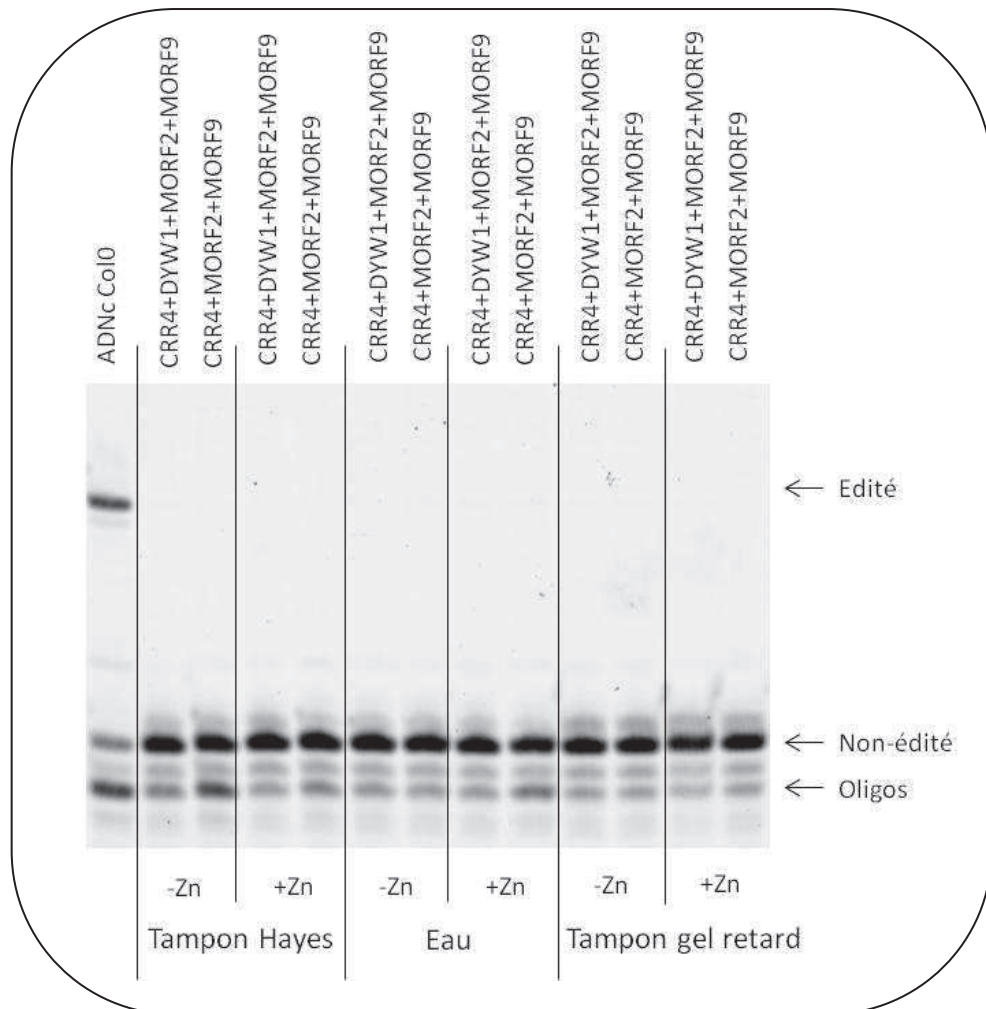
Différents tampons ont été testés lors de cette expérience :

- \* tampon d'édition décrit dans (Hayes and Hanson, 2007b),
- \* tampon d'édition (Hayes and Hanson, 2007b) auquel du zinc a été additionné,
- \* tampon utilisé pour les expériences de retard sur gel par Peter Kindgren,
- \* tampon utilisé pour les expériences de retard sur gel, auquel du zinc a été additionné.

DYW1 contient un domaine DYW dont la signature cytidine déaminase contient quelques acides aminés importants dans la fixation de zinc. De plus, nous savons que MORF2 et MORF9 ont une affinité pour le cobalt (Heazlewood et al., 2004). Une petite quantité de zinc a alors été ajoutée dans chaque tampon. Le zinc est apporté afin de reproduire les concentrations en zinc dans la cellule végétale soit 100 ppm (0,7 mM de ZnCl<sub>2</sub> en solution).

Dans les tampons où le zinc n'a pas été ajouté, je faisais l'hypothèse que le zinc contaminant pouvait être suffisant pour initier l'édition. La contamination en zinc dépend de la pureté de l'eau et des produits chimiques. En général, elle serait de 100 pM en solution (Kay, 2004) même s'il semble qu'à Perth l'eau soit très peu contaminée (cf II.4.3 des résultats).

Peter Kindgren a effectué une grande quantité d'expériences de gels retard impliquant des protéines PPR-E fusionnées au marqueur MBP. Grâce au tampon utilisé, il est parvenu à montrer l'interaction *in vitro* entre les PPR et leur cible ARN (Peter Kindgren, communication personnelle). L'idée en choisissant ce tampon était donc de maximiser les chances d'interaction entre CRR4 et *ndhD-1*. Pour cela, les tests ont été effectués à partir de CRR4 étiqueté ou non avec la MBP. Cette fois encore, il m'a été impossible de reconstituer l'édition (Figure 48). D'autres conditions auraient pu être testées (interaction CRR4-ARN, changement de température, durée d'incubation) mais faute de temps, il m'a été impossible de poursuivre cette expérience.



**Figure 48 : PPE effectuée sur les produits d'édition *in vitro*.**

Deux tampons ont été utilisés dans cette expérience :

- tampon d'édition (Hayes et al., 2007),
- tampon utilisé pour les gels retards (Peter Kindgren, communication personnelle).

Dans chaque cas, 0,7mM zinc a été ajouté (+Zn) ou non (-Zn).

De l'eau a été utilisé comme témoin négatif.

Aucun signal d'édition n'apparait dans aucun des tampons utilisés avec ou sans DYW1.

## Chapitre III : Discussion

### I. Rappel du modèle d'étude

A partir 2005, il a été montré que des protéines PPR étaient nécessaires à l'édition chez les plantes (Kotera et al., 2005). Toutes ces protéines nécessaires à l'édition, à l'exception de la PPR-P DNR596 qui sera discutée plus tard, appartiennent au sous-groupe PLS possédant un domaine C-terminal. Sur la base de similarités de séquence et d'une étude phylogénique, il a été proposé que le domaine DYW réalise la déamination de la cytidine (Salone et al., 2007).

Ainsi, selon cette hypothèse, les PPR de type DYW seraient impliquées dans l'édition des ARN dans les organites par le biais d'une activité enzymatique du domaine DYW. Néanmoins si tel est le cas, comment les PPR de type E/E+, telles CRR4 ou CLB19, peuvent-elles être impliquées dans l'édition (Kotera et al., 2005; Chateigner-Boutin et al., 2008) ? L'hypothèse posée dans la thèse de Véronique Salone et dans la mienne était que les protéines PPR de type E/E+ recrutent une autre protéine portant un motif DYW contenant l'activité cytidine déaminase.

En parallèle, il a été découvert une nouvelle protéine, unique chez *Arabidopsis thaliana*, la protéine DYW1. Cette protéine ne possédant pas de réel enchaînement de motifs PPR, il semblait peu probable qu'elle interagisse directement avec un substrat ARN. Néanmoins, la présence du domaine DYW laissait à penser que DYW1 était impliquée dans l'édition. Il a été proposé que cette protéine pouvait interagir avec les PPR de type E/E+ pour former un complexe permettant l'édition de l'ARN, le facteur de spécificité étant la PPR E/E+ et l'enzyme étant DYW1. De plus, sachant que de nombreuses cytidines déaminases agissent sous forme de dimères (Wang et al., 2008; Salter et al., 2009), il a été proposé que 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. L'hypothèse était donc que DYW1 puisse avoir un rôle central dans l'édition de l'ARN.

## II. L'enzyme d'édition contenue dans le domaine DYW chez les plantes : des données en faveur du modèle

### II.1. Premières observations et changements dans l'hypothèse : DYW1 essentielle à l'édition d'un site dans les chloroplastes

Parmi les 16 lignées TILLING obtenues au cours de ma thèse, une présentait un changement de cadre de lecture dans le domaine PPR-like impliquant la formation d'un codon STOP en plein milieu de la séquence (mutant *dyw1-1*) (Boussardon et al., 2012). Nous savions que certains défauts d'édition dans les chloroplastes ou les mitochondries pouvaient induire des phénotypes extrêmement marqués. Or, si DYW1 avait eu un rôle général important dans l'édition des sites dépendant des PPRE/E+, le phénotype aurait au moins été équivalent à celui du mutant *clb19* (plantule jaune pâle). Ce n'est pas le cas, le phénotype étant semblable de celui observé chez les mutants PPR impliqués dans l'édition des transcrits *ndh*, CRR4, CRR21, CRR22 ou encore CRR28 (Kotera et al., 2005; Okuda et al., 2008; Okuda et al., 2009).

Notre nouvelle hypothèse s'est donc orientée vers un rôle beaucoup moins central de DYW1 dans la machinerie d'édition, affectant éventuellement quelques sites dans les chloroplastes et/ou les mitochondries.

L'analyse des sites d'édition chloroplastique du mutant *dyw1-1* a montré une abolition de l'édition dans le premier site du transcrit *ndhD*. Lorsqu'il est édité, ce site d'édition implique la formation d'un codon d'initiation de la traduction de la sous unité NdhD. Ce site d'édition est connu pour être dépendant de CRR4, une PPR de type E/E+ donc sans motif DYW.

La protéine DYW1, une protéine étant composée d'un domaine DYW et d'un domaine E+ tronqué, est donc nécessaire à l'édition de l'ARN au site *ndhD-1*.

### II.2. Découverte d'une interaction spécifique CRR4-DYW1

Des expériences d'interaction *in vivo* par split-YFP entre CRR4 et DYW1 ont confirmé l'hypothèse de départ selon laquelle DYW1 s'associait avec une PPR de type E/E+ (Boussardon et al., 2012). Cette interaction semblait extrêmement spécifique puisqu'elle n'impliquait pas d'autres PPR.

CRR4 et DYW1 sont toutes les deux indispensables à l'édition de *ndhD-1*. Il est donc nécessaire à CRR4, une PPR-EE+, de recruter DYW1 pour éditer le site d'édition *ndhD-1*.

Notre hypothèse propose donc que CRR4 interagisse avec DYW1 pour former un complexe protéique fonctionnel, équivalent à une PPR-PLS DYW, permettant la reconnaissance et l'édition du site *ndhD-1*. Ce modèle est similaire à celui des mammifères dans lequel ACF, le facteur de spécificité, interagit avec APOBEC-1, l'enzyme, pour éditer le transcrit *apoB*. L'hypothèse selon laquelle le domaine DYW contient une activité catalytique d'édition repose sur la présence d'une signature HxE(x)<sub>n</sub>CxxC trouvée dans les enzymes d'édition dans d'autres organismes (Salone et

al., 2007). Nous avons montré que ce domaine pouvait être apporté en *trans* à la protéine CRR4 et que ce phénomène était nécessaire à l'édition de *ndhD-1*.

### II.3. Implication de protéines « DYW1-like » dans l'édition ?

#### II.3.1. Recrutement en *trans* d'un domaine DYW : une caractéristique commune à toutes les PPR-EE+ ?

Le domaine DYW est dispensable chez certaines PPR-PLS DYW (Okuda et al., 2009). En revanche, le domaine E, qui lui est indispensable, ne semble pas contenir d'activité catalytique permettant la conversion de cytidine en uridine. Il est donc possible que les PPR-PLS dont le domaine DYW est absent puissent recruter en *trans* un domaine DYW et interagir avec des protéines « DYW1-like » pour restaurer l'édition.

Quel moyen de reconnaissance est utilisé dans le cas de CRR4 et DYW1? En quoi CRR4 est-elle différente d'une autre PPR-EE+ ? D'autres protéines ressemblant à DYW1 sont-elles impliquées ?

Il est possible que cette reconnaissance soit dépendante de la longueur du domaine E. Chez CRR4, le domaine E est tronqué d'une quinzaine d'acides aminés et DYW1 possède une quinzaine d'acides aminés dont la séquence est très proche de l'extrémité C-terminale du domaine E. L'interaction entre les deux protéines permet la formation d'un domaine E « complet ». Ceci a été confirmé lors de ma thèse par la fusion entre CRR4 et DYW1, la jonction se faisant entre les deux domaines E.

Il a été identifié chez *Arabidopsis thaliana* cinq protéines « DYW1-like » contenant très peu de motifs PPR et contenant un domaine DYW. Ces protéines ont été nommées DYW2 à DYW6.

#### III.3.2. Le candidat DYW2

Nous avons vu précédemment que l'une de ces protéines de la petite sous-famille proche de DYW1 interagissait avec CLB19, cette protéine a été nommée DYW2. La caractérisation fonctionnelle de DYW2 est effectuée à l'INRA-URGV dans le cadre du projet de thèse de Mauricio Lopez Obando. Au vu de la létalité de la mutation d'insertion *dyw2-1*, il semble que cette protéine affecte beaucoup plus de sites d'édition que DYW1 ou qu'elle cible des sites indispensables à *Arabidopsis*. En effet, le mutant *clb19* affiche un phénotype très marqué (jaune et stérile dans des conditions de croissance classique en serre mais fertile en condition de faible éclaircissement). Il est donc clair que DYW2 affecte au moins un autre site indispensable à la viabilité d'*Arabidopsis thaliana*. De plus, il semble que DYW2 soit doublement adressée. Selon notre hypothèse, elle interagirait avec des protéines PPR-E/E+ chloroplastiques et mitochondriales.

Afin d'étudier DYW2 grâce à des mutants viables, des mutants EMS pourraient être synthétisés de la même façon que pour DYW1 ou bien la quantité de transcrit dans la cellule pourrait être modifiée par ARNi inductible (l'induction à l'éthanol est testée en ce moment). En effet, ces mutations sont plus douces et ont plus de chance de conduire

à un mutant viable qu'une perte de fonction due à une insertion en plein milieu du gène. Il a été observé dans des lignées hétérozygotes pour l'insertion *dyw2-1* une légère diminution de l'édition aux sites *rpoA* et *clpP* non édités chez *clb19*. Néanmoins, il est toujours difficile de spéculer sur une simple diminution de l'édition. Il serait important d'obtenir des mutants ayant une forte diminution de l'édition à certains sites pour prouver son implication dans l'édition.

### II.3.3. Les candidats DYW3 à DYW6

A l'inverse de DYW1 et DYW2, les domaines DYW des protéines DYW3 à DYW6 sont plus difficiles à définir. Il semble ainsi que les domaines aient des structures particulières :

- DYW3 (At2g34370) possède le triplet DYW terminal mais n'est pas conservé dans la signature cytidine déaminase. En effet, l'histidine, le glutamate et la deuxième cystéine sont absentes. La conservation du domaine DYW est plus faible que celle observée pour d'autres PPR-DYW. Le gène *DYW3* est transcrit (un ADNc au TAIR) et deux mutants sont disponibles. Selon TargetP, la protéine DYW3 est prédite pour être mitochondriale même si le « score » chloroplastique n'est pas négligeable (double adressage ?).

- pour DYW4 (At1g29710), tout comme DYW3, le domaine DYW semble assez dégénéré. En effet, la signature cytidine déaminase n'est pas conservée, seul le glutamate et la première cystéine le sont, et le triplet final correspond à un NLW et non un DYW. Néanmoins, nous avons vu dans cette thèse qu'il semble que seul le tryptophane était nécessaire à l'édition de *ndhD-1* par DYW1 (la tyrosine n'est pas nécessaire et l'aspartate est remplacée par une asparagine ayant des propriétés proches). Le gène *DYW4* est transcrit (9 EST au TAIR) et six mutants sont disponibles. Cette protéine est prédite pour être localisée dans les mitochondries.

- DYW5 (At4g32450), d'après les prédictions, est mitochondriale. Cette protéine est beaucoup mieux conservée que DYW3 et DYW4 sur l'ensemble de son domaine DYW. Le site actif du domaine DYW est parfaitement conservé et le triplet final correspond à un EYW (l'aspartate étant remplacée par un glutamate ayant des propriétés proches). Le gène *DYW5* est transcrit (24 EST au TAIR) et neuf mutants sont disponibles. Tout comme DYW2, cette protéine semble être un excellent candidat pour un rôle dans l'édition dans les mitochondries. Très récemment, cette protéine a été caractérisée fonctionnellement. Dans cette publication, elle est baptisée « MEF8S » et il est montré qu'elle est importante pour l'édition des transcrits *nad5-676* et *nad6-95* dans le pollen (Verbitskiy et al., 2012b). A noter que le mutant *mef8s* est parfaitement viable.

- DYW6 (At2g25580) a été caractérisée sous le nom de MEF8 (Verbitskiy et al., 2012b). Cette protéine est mitochondriale et impliquée dans l'édition des transcrits *nad5-676* et *nad6-95* dans les feuilles. Le domaine DYW de cette protéine est très bien conservé, le site actif et le triplet DYW étant parfaitement conservés. Le mutant *mef8* est viable.

Le double mutant *mef8xmef8s* (*dyw5xdyw6*) n'est en revanche pas viable ce qui n'est pas explicable simplement par la perte d'édition dans les sites *nad5-676* et *nad6-95* (Arabidopsis étant viable sans complexe I). Une explication pourrait être que MEF8 et

MEF8s se complètent l'une l'autre et qu'elles possèdent une fonction bien plus centrale que laisse supposer l'étude des simples mutants. Il est donc possible que d'autres sites d'édition soient dépendants de l'activité de ces deux protéines.

Le nombre de motif PPR de ces protéines étant trop faibles pour interagir de façon séquence spécifique avec l'ARN, il est probable que ces protéines agissent de concert avec des PPR de type E/E+ de la même manière que CRR4 et DYW1. Toutefois, aucune preuve n'existe à ce jour pour appuyer cette hypothèse. Des expériences d'interaction protéine-protéine pourraient être effectuées avec des PPR candidates identifiées au préalable. De plus, une étude approfondie des sites d'édition semble nécessaire dans ce cas.

## II.4. Des caractéristiques communes avec les cytidines déaminases (CDA)

### II.4.1. La signature CDA indispensable à l'édition

Le domaine DYW contient une signature invariante HxE(x)<sub>n</sub>CxxC trouvée aussi chez les CDA suggérant une activité catalytique de ce domaine (Salone et al., 2007). De plus, d'après des prédictions de structures, ce domaine possède des similarités structurales avec des désaminases connues (Iyer et al., 2011). Chez les désaminases connues, la signature permet la fixation d'un atome de zinc essentiel à la fonction de désamination (MacGinnitie et al., 1995). Le glutamate contenu dans la signature est impliqué dans la fonction de désamination de la cytidine. Si le domaine DYW est réellement une CDA, alors on s'attend à ce que chaque résidu contenu dans la signature soit essentiel à l'édition de l'ARN.

Les expériences de complémentation de *dyw1-1* par des versions de DYW1 dont la signature CDA a été substituée montrent que cette signature est indispensable à l'édition tout comme chez les CDA connues. Il sera nécessaire de muter individuellement tous les acides aminés composant la signature pour savoir si chaque acide aminé est indispensable à l'édition.

De plus, l'ICP-MS effectuée sur DYW1 suggère que le domaine DYW interagisse avec un (des) atome(s) de zinc. Cette interaction est caractéristique des cytidines déaminases. La majorité des CDA ne fixe qu'un seul ion zinc dans leur site actif par le biais des quatre acides aminés présents dans la signature.

Il semble que lorsque des acides aminés composant la signature sont éliminés, DYW1 soit incapable de se lier aux atomes de zinc. La signature CDA du domaine DYW serait donc indispensable à la fixation des atomes de zinc. Même si ces résultats restent à confirmer, nous pourrions avoir ici la première preuve d'interaction entre un facteur d'édition avec un atome de zinc.

De nouvelles expériences d'ICP-MS devraient être reconduites, incluant des mutations individuelles dans la signature cytidine déaminase du domaine DYW ainsi que dans le contrôle (cytidine déaminase d'*E. coli*).

Par ces différentes expériences, nous pourrions confirmer que le domaine DYW a des caractéristiques communes avec les CDA malgré l'absence d'activité *in vitro* à partir de protéines purifiées.

Le domaine DYW contient aussi d'autres acides aminés très conservés qui ne semblent pas avoir de lien avec les CDA comme la séquence HHxxxGxCSC en C-terminal ou le triplet final DYW. Ces séquences, de fonction inconnue, pourraient être testées de la même façon que la signature CDA par mutagenèse dirigée puis par ICP-MS.

#### II.4.2. Structure du domaine DYW

En comparant les domaines DYW et les CDA, il apparaît que le domaine DYW correspondrait en réalité à une moitié de CDA. Nous pouvons penser que le domaine DYW pourrait fonctionner en hétérodimère avec une autre protéine pour former une CDA complète.

Peu de choses sont connues sur la fonction du domaine C-terminal des CDA. Chez les cytidines déaminases de type AID (pour « Activation-induced cytidine deaminase ») agissant sur l'ADN, l'extrémité C-terminale est nécessaire à la commutation de classe (ou « Class Switch Recombination ») permettant le changement d'une classe (isotype) d'anticorps en une autre chez les mammifères (Doi et al., 2009). Chez *E. coli*, il semble que cette extrémité isole l'atome de zinc du solvant « l'emprisonnant » dans l'homodimère (site web PDB : <http://www.rcsb.org/pdb/home/home.do>, exemple de la CDA NP\_416648, numéro d'accèsion au PDB : 1CTU).

Il semble donc clair que dans certains cas, ce domaine C-terminal soit essentiel. En est-il de même chez les plantes ? Est-il possible d'identifier une seconde moitié de CDA ?

A l'heure actuelle, il m'a été impossible de retrouver une protéine chez *Arabidopsis thaliana* ressemblant à la moitié C-terminale d'une CDA mis à part les neuf cytidines déaminases codées par le génome nucléaire étudiées préalablement (Faivre-Nitschke et al., 1999) (cf II.1.4 de l'introduction). Il n'est néanmoins pas certain qu'une telle protéine existe, le fonctionnement d'une CDA étant peut être différent d'une espèce à une autre.

L'identification du complexe d'édition serait une exceptionnelle source d'information sur l'édition chez les plantes et permettrait potentiellement d'identifier, si elle existe, une protéine homologue à l'extrémité C-terminale des CDA.

En ce qui concerne la structure, nous ne disposons pour l'instant que de prédictions de la structure tridimensionnelle du domaine DYW basées sur un alignement « forcé » avec une cytidine déaminase humaine. Ce type de prédiction apporte donc un fort biais en faveur d'une structure similaire à celle des CDA. Il serait nécessaire de caractériser la structure tridimensionnelle de DYW1 et plus spécifiquement de son domaine DYW. Le site actif a-t-il vraiment la même conformation qu'une cytidine déaminase ? Sa conformation donne-t-elle une indication sur une éventuelle interaction avec une autre protéine en C-terminal ? La cristallisation de DYW1 pourrait nous permettre de répondre à ces questions. Néanmoins, il faut que la protéine DYW1 soit synthétisée en grande quantité et de façon soluble ce qui n'est pas évident pour les protéines PPR et ce qui a été tenté sans succès pendant la thèse de Véronique Salone.



### III. D'autres candidats à l'édition de l'ARN chez les plantes

#### III.1. Intégration des protéines MORF au modèle d'étude

##### III.1.1. Un nouveau modèle d'étude chez *Arabidopsis thaliana* ...

Une des explications avancées pour expliquer l'échec des différentes tentatives de reconstitution *in vitro* de l'édition est que CRR4 et DYW1 ne sont pas les seules protéines nécessaires pour éditer *ndhD-1*. Récemment, les protéines MORF ont été identifiées (Bentolila et al., 2012; Takenaka et al., 2012). Ces protéines, dont la fonction est inconnue, sont impliquées dans l'édition d'un grand nombre de sites dans les chloroplastes et les mitochondries. Cette famille de protéines est composée de neuf membres, deux d'entre eux sont adressés aux chloroplastes (MORF2 et MORF9). Ces deux protéines sont nécessaires à l'édition du site *ndhD-1* (0% d'édition chez les deux mutants). Nous pouvons supposer que ces deux protéines interviennent à un moment donné dans le complexe CRR4-DYW1. De même, il est possible qu'elles interviennent dans le complexe impliquant CLB19 et DYW2 même si aucune protéine MORF n'a été retrouvée par TAP-tag (cf résultats I.7.4). Comment interviennent-elles dans le complexe d'édition ?

Il semble que les protéines MORF ne possèdent pas de domaine permettant l'interaction avec l'ARN et qu'elles ne possèdent pas de résidus ayant une activité cytidine déaminase. Néanmoins, une étude a permis de montrer que ces protéines avaient une affinité avec les ions métaux (Heazlewood et al., 2004). Nous savons que les domaines DYW possèdent la signature cytidine déaminase contenant un domaine de fixation au zinc (Salone et al., 2007). Il a été envisagé que ces protéines recrutent des ions métaux pour ensuite les « donner » au domaine DYW.

Néanmoins, les expériences de fixation des ions métaux *in vitro* montrent que le domaine DYW peut fixer les ions zinc sans le concours des protéines MORF.

De son côté, le domaine E de DYW1 est indispensable à l'édition de *ndhD-1*. Il semble que le domaine E ne contienne pas d'activité catalytique. Sa fonction est encore inconnue mais nous pouvons éventuellement penser à un rôle dans des interactions protéine-protéine qui, par exemple, permettrait le recrutement de cofacteur comme les protéines MORF2 et MORF9 (Takenaka et al., 2012). Cette dernière option est peu probable en raison des résultats d'interaction entre MORF8 et RARE1 montrant une interaction au niveau des motifs PPR par double-hybride chez la levure (Bentolila et al., 2012).

Une mutagenèse dirigée pourrait être effectuée pour connaître l'importance de chaque acide aminé dans la fonction ou l'interaction avec les protéines PPR.

##### III.1.2. ...invalide chez *Physcomitrella patens*

Alors qu'il est clair que les protéines MORF sont nécessaires à l'édition des ARN chez *Arabidopsis thaliana*, il semble que ce ne soit le cas que chez les plantes à fleurs. En

effet, chez la plante modèle *Physcomitrella patens*, aucune protéine MORF n'est identifiée dans le génome. Pour comprendre la fonction des MORF, il faudrait donc probablement rechercher une spécificité de l'édition des plantes à fleurs.

Sachant que chez *Physcomitrella*, seules des protéines PPR de type DYW sont impliquées dans les quelques événements d'édition, une hypothèse pourrait être que les protéines MORF compensent la perte du domaine DYW dans les PPR-PLS E/E+ chez les plantes à fleur (Takenaka et al., 2012). Néanmoins, d'après l'étude de DYW1, il semblerait plutôt que la perte du domaine DYW soit compensée par l'interaction entre une PPRE/E+ avec une protéine « DYW1-like ».

Une autre hypothèse pourrait être que les protéines MORF servent à recruter soit:

- une cytidine déaminase, dans le cas où le domaine DYW ne contienne pas d'activité,

- la deuxième moitié de la cytidine déaminase (qui n'est pas présente dans le domaine DYW), dans le cas où le domaine DYW aurait une activité.

Toutefois, dans le cas de *Physcomitrella*, les domaines DYW sont très proches de ceux retrouvés chez *Arabidopsis* ce qui suppose qu'ils ont une fonction similaire.

### III.2. Implication des cytidines déaminases nucléaires dans l'édition ?

Même si l'alignement du domaine DYW avec les cytidines déaminases ne montre que des similarités partielles, ce domaine semble être le meilleur candidat pour contenir l'activité catalytique.

Toutefois, dans le génome nucléaire d'*Arabidopsis thaliana*, neuf cytidines déaminases (dont huit sont positionnées ensemble dans le génome) ont été identifiées. Sept d'entre elles ne possèdent *a priori* pas de séquences d'adressage et seraient donc cytosoliques (Faivre-Nitschke et al., 1999). A moins d'une mauvaise annotation, il semble que ces enzymes ne soient pas adressées vers les organites et ne participent donc pas à l'édition des ARN dans les chloroplastes ou les mitochondries.

En revanche, deux d'entre-elles, CDA2 et CDA7, possèdent une séquence d'adressage potentielle. Les prédictions ne permettent pas de définir clairement leur localisation mais des expériences de localisation subcellulaire de protéines de fusion avec la GFP ont été effectuées pour tous les membres de cette famille et ont montré que CDA2 serait localisée dans les chloroplastes et les mitochondries alors que CDA7 serait adressée aux chloroplastes uniquement (José Gualberto, communication personnelle).

Des mutants de CDA2 et CDA7 ont été étudiés. Le mutant SALK de *cda2* est viable et ne semble pas montrer un phénotype particulier. La même observation a été effectuée pour le mutant *cda7* possédant une insertion de transposon (José Gualberto, communication personnelle). Comme les protéines CDA2 et CDA7 sont très proches, il est possible que la fonction de CDA2 soit redondante avec celle de CDA7 et qu'une cytidine déaminase puisse compléter la fonction de l'autre. Néanmoins, il semble peu probable que CDA7 (qui aurait une localisation chloroplastique uniquement) puisse compléter la fonction mitochondriale de CDA2. Il serait très intéressant de produire

un double mutant *cda2xcda7* pour tester cette hypothèse mais du fait de la proximité physique des deux gènes (quelques kb de différences), ce double mutant est très difficile à obtenir.

L'étude de la famille des 9 CDA d'*Arabidopsis* datant de près de 10 ans (Faivre-Nitschke et al., 1999), il serait intéressant d'étudier la fonction de ces gènes avec les outils actuels, obtenir des mutants de chaque gène, synthétiser des double-mutants, vérifier l'éditotype chloroplastique et mitochondrial et éventuellement tenter de localiser à nouveau les protéines CDA tout en utilisant les données fournies par l'équipe de José Gualberto sur les CDA.

De nouvelles données sont en effet présentes sur TAIRv10 pour CDA1 (mutants et ADNc), CDA2 (mutants et ADNc), CDA3 (ADNc), CDA4 (ADNc), CDA5 (ADNc), CDA6 (mutants et ADNc), CDA7 (mutants et ADNc), CDA8 (ADNc) et CDA9 (mutants et ADNc). De plus, des CDA ont été identifiées lors d'expériences d'interaction protéine-protéine. Ces expériences de protéomique peuvent donner des indications quant à la localisation mais aussi à la fonction des CDA chez *Arabidopsis thaliana*.

Néanmoins, il est important de noter que CDA-1 agit sur des nucléotides libres et non sur des ARN (Faivre-Nitschke et al., 1999). Il est donc improbable que ces CDA soient impliquées dans l'édition de l'ARN chez *Arabidopsis*.

## IV. Cas particuliers de PPR-PLS non impliquées dans l'édition

### IV.1. Une diversification de la fonction ?

Il faut aussi noter que le domaine DYW d'OTP85, PPR impliquée dans l'édition, est connu pour avoir une activité endoribonucléase (Nakamura and Sugita, 2008). Sachant que le domaine DYW d'OTP85 est aussi conservé que les autres domaines DYW connues chez *Arabidopsis thaliana*, il n'est pas impossible que d'autres PPR de ce type permettent le clivage d'un ARN à un site bien spécifique. La protéine CRR2 (PPR-DYW) permettant le clivage des transcrit *rps7* et *ndhB* semble être de ce type (Hashimoto et al., 2003).

De plus, chez *Physcomitrella patens*, PpPPR\_43 (PPR-DYW) est impliquée dans l'épissage du transcrit *cox1* (Ichinose et al., 2012). Néanmoins, il apparaît qu'une délétion du domaine E et DYW n'affecte pas l'épissage. De même, OTP70 (PPR-E) chez *Arabidopsis thaliana* est impliquée dans l'épissage de l'ARNm *rpoC1* (Chateigner-Boutin et al., 2011). Encore une fois, le domaine E est dispensable à l'épissage. Alors que l'on pensait que les PPR-PLS étaient réservées à l'édition, il semble qu'il y ait une diversification des PPR impliquées dans la maturation de l'ARN et ce même si les domaines E ou DYW ne semblent pas forcément impliqués, dans les cas de PpPPR\_43 et d'OTP70 du moins.

### IV.2. Cas de la PPR596

En ce qui concerne les PPR-P, seule PPR596 a été publiée comme étant impliquée dans l'édition de l'ARNm (Doniwa et al., 2010). Il s'agit en fait ici d'une augmentation de l'édition. Le phénotype extrêmement marqué du mutant peut laisser supposer que les changements dans l'édition obtenus soient dû à un effet indirect de la mutation. L'influence de PPR596 sur l'édition est comparable à celle des ribonuléoprotéines chloroplastiques (cpRNPs) qui sont principalement impliquées dans la maturation et la stabilisation de l'ARNm (Hirose and Sugiura, 2001; Tillich et al., 2009). Elle pourrait résulter par exemple d'une modification de l'expression de l'ARN précurseur chez le mutant ou bien d'une absence de compétition pour un site de fixation d'une PPR d'édition.

## V. Identification du complexe d'édition et futures expériences

Dans ma recherche de la fonction exacte de DYW1, j'ai été amené à effectuer des essais de réactions d'édition *in vitro* qui se sont révélés jusque-là infructueux. Les causes peuvent être multiples, il semble que la plus plausible soit que le complexe que je propose soit incomplet ou que les conditions d'édition ne soient pas appropriées. La récente découverte des protéines MORF, requises pour l'édition de l'ARN, supporte cette hypothèse.

Je propose dans ce paragraphe une expérience pouvant permettre l'identification du complexe d'édition ainsi que la reconstitution de l'édition *in vitro*.

### V.1. Identification du complexe d'édition

Afin de découvrir de nouveaux facteurs requis pour l'édition, les protéines PPR pourraient être utilisées comme cibles dans des expériences d'interaction protéine-protéine *in planta* (par TAP pour « Tandem Affinity Purification »). Généralement, l'étiquette est ajoutée en C-terminal de la protéine. Dans le cas des PPR impliquées dans l'édition, l'extrémité C-terminale des protéines DYW est très conservée et requise pour la fonction d'édition. Il serait donc nécessaire d'insérer une étiquette à l'extrémité N-terminale de la protéine pour éviter que l'étiquette bloque la fonction ou l'interaction protéine-protéine. Une construction pourrait être effectuée afin de placer l'étiquette à l'extrémité C-terminale de la séquence d'adressage aux organites (après le site de clivage).

La fonctionnalité de la protéine de fusion sera vérifiée par complémentation du mutant « perte de fonction » approprié.

Le complexe sera purifié sur colonne en utilisant l'étiquette de la protéine de fusion exprimée *in planta* et analysé par spectrométrie de masse dans le but d'identifier chaque protéine le composant.

### V.2. Reconstitution de l'édition de l'ARN *in vitro*

Pour tester que nous avons pu identifier tous les éléments composant le complexe d'édition grâce à la protéine PPR étiquetée, il faudrait effectuer une reconstitution de l'édition *in vitro*. Pour ce faire, les protéines devront être exprimées et incubées avec le substrat ARN (non édité).

Le tampon d'édition pourra être celui utilisé dans les expériences d'édition *in vitro* effectués à partir d'extrait chloroplastique (Hegeman et al., 2005).

Les protéines recombinantes seront étiquetées avec l'étiquette MBP qui a été utilisée avec succès pour la production, la purification et les expériences d'interaction ARN-protéine *in vitro* de la protéine CLB19 (projet post-doctorat de Peter Kindgren). En parallèle de la protéine PPR, les facteurs retrouvés lors de l'identification du complexe d'édition seront produits et purifiés de la même façon. En parallèle, le complexe purifié en TAP pourra directement être utilisé sur l'ARN.

La reconstitution de l'édition de l'ARN pourra être évaluée par la technique de PPE, qui donne des résultats très précis du taux d'édition dans le substrat.

### V.3. Reconstitution de l'édition dans un système hétérologue

En parallèle de la reconstitution de l'édition *in vitro*, la reconstitution de l'édition dans des systèmes génétiques où l'édition n'a pas lieu pourra être tentée. La réussite de cette reconstitution confirmera que tous les facteurs nécessaires à l'édition ont été retrouvés et pourra éventuellement ouvrir la voie vers de nouvelles applications biotechnologiques.

Plusieurs conditions doivent être prises en compte dans la sélection de l'hôte. En effet, il faut que l'hôte soit incapable d'éditer l'ARN et simple à transformer. De plus, il faudrait évaluer s'il est simple d'exprimer les facteurs d'édition dans le système choisi. L'introduction de la cible pourrait aussi être un problème c'est pourquoi il pourrait être possible de modifier le facteur d'édition pour cibler une séquence endogène comme dans le cas de la protéine PPR10 (Barkan et al., 2012).

Chez *Marchantia* par exemple, les transcrits ne sont pas édités. Maintenant que l'on sait comment les PPR interagissent avec l'ARN (Barkan et al., 2012), il devrait être possible de modifier une PPR-PLS DYW afin qu'elle interagisse avec un ARNm de *Marchantia*. Mais cela nécessite de transformer le génome nucléaire afin de pouvoir éditer les transcrits chloroplastiques. D'autres systèmes peuvent être envisagés pour reconstituer l'édition tels *E. coli* (transformation effectuée en routine), les levures (transformation du génome nucléaire pour éditer les transcrits mitochondriaux), des cellules humaines contenant une mutation dans leur génome mitochondrial pouvant être corrigé par une conversion de C en U (transformation du génome nucléaire possible).

L'association des différents membres du complexe pourra être évaluée par co-immunoprécipitation, les protéines pouvant être étiquetées. Le taux d'édition de la cible sera calculé par PPE. Dans le cas d'une utilisation d'un ARN synthétique, ce dernier pourra être produit de sorte à ce que l'édition puisse conduire à une altération importante dans l'ARN conduisant l'expression d'une protéine comme la luciférase, un peu de la même manière que l'expérience d'édition *in vivo* tentée en levure (cf résultats II.5.1, expression de la GFP).

## VI. Pourquoi éditer les transcrits des organites?

Les raisons pour lesquelles des êtres vivants ont recours à l'édition de l'ARN sont encore très mal comprises. Selon le modèle de la bureaucratie, l'édition serait même une aberration car il semblerait, à première vue, qu'elle soit une perte d'énergie et de temps (Takenaka et al., 2008). En effet, il semble qu'il serait beaucoup plus simple pour les génomes des organites d'être directement à l'état « édité » au niveau de l'ADNg. Néanmoins, plusieurs hypothèses ont été proposées pour répondre à cette problématique.

### VI.1. Augmentation de la « capacité codante » du génome des organites

Cette hypothèse a été réfutée dans une publication parue en 2008 (Takenaka et al., 2008). Le génome mitochondrial étant déjà assez grand et pouvant certainement coder des gènes supplémentaires rend obsolète l'idée que l'édition pourrait être utilisée pour synthétiser plusieurs formes d'un transcrite et ce uniquement pour économiser de la place dans le génome mitochondrial.

De plus, l'édition a besoin pour fonctionner de centaines d'autres facteurs. Dans ce sens, l'édition occuperait plus de place dans le génome qu'elle n'en libérerait.

### VI.2. Complication du transfert du génome des organites vers le noyau

Il est maintenant admis par la communauté scientifique que des gènes sont transférés des organites vers le noyau. Pour « contrer » cette tendance, le génome des organites aurait recours à l'édition des transcrits pour protéger son génome. Ce serait un mécanisme similaire à la modification des codons dans le génome mitochondrial des mammifères.

Il semble qu'aucun système d'édition de l'ARN de C en U n'existe dans le noyau des plantes. Ainsi, un gène « non-édité » ayant été transféré vers le noyau ne serait pas fonctionnel. Si l'intégration des gènes se faisait par le biais d'une transcription inverse à partir de l'ARNm édité, le génome nucléaire posséderait un gène capable de coder une protéine fonctionnelle. Or, la plupart des gènes transférés du génome mitochondrial vers le noyau ne sont pas édités indiquant un transfert majoritairement fait à partir de l'ADNg (Takenaka et al., 2008). Donc, les transcrits produits à partir de gènes transférés ne devraient pas être édités dans le noyau, les protéines pouvant ainsi ne pas être fonctionnelles.

### VI.3. Augmentation de la diversité

L'hypothèse de l'augmentation de la diversité est surtout avancée dans les cas d'édition nucléaire chez les mammifères (comme le transcrite *apoB*). Elle semble beaucoup moins plausible chez les plantes car aucun site d'édition, publié à ce jour, ne permettrait

d'augmenter la diversité d'une protéine dans les organelles. La seule exception dans les organites est celle de la protéine mitochondriale, AEP1, chez *Trypanosoma brucei* (Ochsenreiter and Hajduk, 2006).

Chez les plantes, il est clair que les protéines codées par un ARNm totalement édité ont été sélectionnées et conservées pour leurs propriétés biochimiques (Takenaka et al., 2008). L'étude de l'ADNg d'atp9 « non-édité » (contenant le signal d'adressage mitochondrial) intégré dans le génome nucléaire chez le tabac a montré que la protéine importée, d'une part, n'était pas fonctionnelle et, d'autre part, perturbait le bon fonctionnement de la mitochondrie, certainement en entrant en compétition avec la protéine Atp9 « éditée » (Hernould et al., 1993). Les protéines non fonctionnelles comme Atp9 « non-éditée » sont, pour la majorité, rapidement dégradées (Takenaka et al., 2008).

#### VI.4. Réparation du génome

Dans la majorité des cas, l'édition de l'ARN restaure des séquences conservées et ainsi permet la production d'une protéine fonctionnelle. Ainsi, l'édition pourrait être un mécanisme visant à réparer les erreurs dues aux mutations accumulées dans le génome des organites au cours de l'évolution. Dans la majorité des cas, le génome des chloroplastes et des mitochondries est transmis de manière monoparentale diminuant fortement les chances de recombinaison et ainsi maintenant dans le génome des mutations négatives (Muller, 1964; Chateigner-Boutin and Small, 2011). Néanmoins, il semble que la recombinaison biparentale soit aussi présente chez les plantes (dans les mitochondries) même si elle est très rare (Barr et al., 2005).

Y a-t-il un avantage à éditer l'ARN plutôt qu'à corriger la mutation directement dans le génome ? Chez les plantes, une réversion des mutations vers une version non mutée semble plus difficile rendant l'édition de l'ARN plus stable (Tillich et al., 2006).

Néanmoins, pourquoi conserver l'édition à des sites silencieux ou qui divergent d'un consensus ?

De même, pourquoi seules les mutations de T en C sont-elles corrigées chez les plantes ? Pour cette dernière question, il pourrait y avoir une explication. Chez les plantes, il y a une corrélation entre la composition en GC et l'édition. En effet, il est clair que l'édition de C en U permet de contrecarrer la préférence qu'a le génome pour les résidus cytidines (Fujii and Small, 2011).

#### VI.5. Régulation de l'expression des gènes

Dans le cas du site *ndhD-1*, le transcrit est édité à hauteur de 50-55% (Kotera et al., 2005). Ce phénomène est appelé édition partielle, c'est-à-dire que l'on retrouve des molécules d'ARN matures éditées et des molécules non éditées. Dans ce cas précis, le transcrit non édité n'est pas traduit en protéine car l'édition introduit un codon d'initiation de la traduction absent du pré-ARNm. Néanmoins, d'autres publications montrent que les transcrits non édités peuvent être traduits. C'est le cas des transcrits *rps12* dans les mitochondries, *ndhD-2* chez le mutant *crr21* ou encore *rpoA* chez le mutant *clb19* qui ont été détectés (Lu et al., 1996; Okuda et al., 2007; Chateigner-



Boutin et al., 2008). Ainsi, il est possible qu'à partir d'un seul pré-ARNm plusieurs protéines soient synthétisées (Okuda et al., 2010). Ces protéines pourraient avoir des fonctions ou des caractéristiques différentes (Chateigner-Boutin and Small, 2010). De plus, il a été montré que les sites d'édition partiels diffèrent selon les organes étudiés (Chateigner-Boutin and Hanson, 2003; Shikanai, 2006), les conditions de culture et les stress environnementaux (Karcher and Bock, 1998; Nakajima and Mulligan, 2001). Par exemple, le transcrit *ndhD-1* est édité à hauteur de 55% dans les feuilles d'*Arabidopsis thaliana* mais n'est pas édité dans les racines. Ainsi, dans les racines, le complexe NDH est absent (Shikanai, 2006). Il est possible que les variations présentes dans l'édition permettent de contrôler la quantité de complexes actifs, voire même de changer les propriétés d'un complexe (Chateigner-Boutin and Small, 2011).

# Chapitre IV : Matériels et méthodes

## I. Matériels biologiques

### I.1. Espèces végétales

#### I.1.1. *Arabidopsis thaliana*

L'essentiel de mon travail de doctorat a été effectué sur la plante modèle *Arabidopsis thaliana*. L'écotype *Columbia* (Col-0) a été principalement utilisé (mutagénèse, génotypage,...). L'écotype *Landsberg erecta* (noté Ler) a été utilisé dans le cadre de l'expression protéique en split-YFP.

#### I.1.2. *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. Souches bactériennes

#### I.2.1. DH5 $\alpha$ (Invitrogen™, 18265-017)

La souche d'*Escherichia coli* DH5 $\alpha$  est utilisée dans de nombreux laboratoires de biologie moléculaire 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 ou par électroporation. Au cours de cette thèse, cette souche a été transformée uniquement par choc thermique.

Le génotype : *fhuA2*  $\Delta$ (*argF-lacZ*)*U169 phoA glnV44*  $\Phi$ 80  $\Delta$ (*lacZ*)*M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*

#### I.2.1. DH10B (Invitrogen™, 18290-015)

Tout comme DH5 $\alpha$ , DH10B est une souche d'*Escherichia coli* utilisée pour les clonages et la multiplication plasmidique. Cette souche est utilisée pour effectuer des transformations par électroporation.

Le génotype : F- *mcrA*  $\Delta$ (*mrr-hsdRMS-mcrBC*)  $\phi$ 80*lacZ* $\Delta$ *M15*  $\Delta$ *lacX74 recA1 endA1 araD139 $\Delta$ (*ara, leu*)*7697 galU galK*  $\lambda$ - *rpsL nupG**

#### I.2.3. DB3.1 (Invitrogen™, 11782-018)

DB3.1 est une souche bactérienne d'*Escherichia coli* modifiée génétiquement. Cette souche est insensible à l'effet toxique du gène *ccdB* (Controlling Cell Death B ; contenu

dans les cassettes Gateway). Nous utilisons cette souche pour multiplier les vecteurs donneurs et destination du système Gateway.

Le génotype : F- gyrA462 endA1 glnV44  $\Delta$ (sr1-recA) mcrB mrr hsdS20(r<sub>B</sub><sup>-</sup>, m<sub>B</sub><sup>-</sup>) ara14 galK2 lacY1 proA2 rpsL20(Sm<sup>r</sup>) xyl5  $\Delta$ leu mtl1

#### I.2.4. BL21(DE3) (Invitrogen™, C6020-03)

La souche d'*Escherichia coli* BL21 utilisée est une souche commerciale (BL21 Star (DE3) One Shot®) permettant l'expression de protéines. Cette souche contient le lysogène DE3 permettant de réguler les gènes contrôlés par le promoteur T7 ainsi qu'une mutation (par rapport à la version précédente de BL21) *rne131* empêchant la dégradation des ARNm par la RNase E et impliquant donc une meilleure expression des protéines recombinantes. Ainsi, cette souche a été transformée par des vecteurs d'expression contenant le promoteur T7 et par un vecteur « helper » (possédant des ARNt rare chez *E. Coli*) facilitant la traduction de protéines uniquement retrouvées chez les végétaux.

Le génotype : F- ompT hsdSB (r<sub>B</sub>-m<sub>B</sub>-) gal dcm rne131 (DE3)

#### I.2.5. C58C1

*Agrobacterium tumefaciens* est une agrobactérie provoquant la maladie de la galle du collet chez les plantes. L'ADN-T est transféré dans les cellules végétales pour ensuite être intégré au génome (formant ainsi une galle à l'endroit de l'infection). C58C1 est une souche « domestiquée » d'*A. tumefaciens* permettant la transformation transitoire des plants de tabac (*N. benthamiana*) ainsi que la transformation stable d'*Arabidopsis thaliana*. L'insert d'intérêt est ainsi inséré dans le génome de la plante. La souche C58C1 peut être utilisée avec les plasmides « helper » pSOUP, pCH32 et/ou pMP90.

Le génotype : pTi<sup>-</sup>, Ery<sup>R</sup>, Chl<sup>R</sup>

### I.3. Souches de levure

#### I.3.1. INVSc1 (Invitrogen™, C810-00)

La souche INVSc1 de *Saccharomyces cerevisiae* est une souche de levure permettant l'expression de protéines recombinantes dans un système eucaryote. Il s'agit d'une souche auxotrophe pour l'histidine, la leucine, le tryptophane et l'uracile (c'est-à-dire qu'elle ne peut pas pousser sans ces composés).

Le génotype : MATa his3D1 leu2 trp1-289 ura3-52 MAT $\alpha$  his3D1 leu2 trp1-289 ura3-52

#### I.3.2. MAV203 (Invitrogen™, 11281011)

La souche MAV203 de *Saccharomyces cerevisiae* est une souche de levure généralement utilisée pour le double hybride. Cette souche est auxotrophe pour l'histidine, le tryptophane, la leucine et l'uracile.

Le génotype: *MAT $\alpha$* , *leu2-3, 112*, *trp1-901*, *his3 $\Delta$ 200*, *ade2-101*, *gal4 $\Delta$* , *gal80 $\Delta$* ,  
*SPAL10::URA3*, *GAL1::lacZ*, *HIS3<sub>UAS</sub> GAL1::HIS3@LYS2*, *can1<sup>R</sup>*, *cyh2<sup>R</sup>*

## II. Méthodes relatives à l'ADN

### II.1. Extraction d'ADN

#### II.1.1 Extraction d'ADNg d'*Arabidopsis thaliana*

Afin d'extraire l'ADN génomique, il est possible d'utiliser un protocole d'extraction (dit « rapide ») assez simple permettant le génotypage des plantes. Ce protocole a été utilisé sur feuilles et cotylédons mais peut aussi être utilisé à partir de boutons ou de fleurs.

Ce protocole peut-être utilisé en tube ou en plaque 96 puits.

Les tissus sont collectés et broyés dans l'azote liquide puis 400µl de tampon de lyse sont ajoutés (SDS 0,5 % ; NaCl 250 mM ; EDTA 25 mM; Tris-HCl (pH: 7,5) 200 mM). L'échantillon est ensuite vortexé, incubé 15 minutes à 50°C et centrifugé 15 minutes à 15000g et 4°C. 250µl de surnageant sont prélevés et 250µl d'isopropanol sont ajoutés. L'échantillon est ensuite vortexé, incubé 20 minutes à température ambiante et centrifugé 30 minutes à 15000g à 4°C. Le surnageant est éliminé et le culot est rincé avec 125µl d'éthanol à 70%. L'échantillon est centrifugé 15 minutes à 15000g à 4°C, le surnageant est éliminé et le culot est séché à température ambiante. Le culot est resuspendu dans 125µl de TE 1X ou d'eau.

TE 10X : 100ml Tris HCl 1M, 20ml EDTA 0,5M pH8, 880 ml d'eau.

L'ADNg ainsi extrait est assez pur pour effectuer une PCR par la suite.

Pour obtenir un ADNg de meilleure qualité, une étape de séparation phénol/chloroforme peut être ajoutée à la suite de l'extraction « rapide », le phénol permettant la séparation protéines/acides nucléiques et le chloroforme permettant la séparation avec le phénol. Des traces de phénol peuvent gêner les manipulations à venir.

Un volume d'un mélange de phénol/chloroforme (v/v 1 : 1) est ajouté à l'échantillon. L'échantillon est ensuite vortexé et centrifugé 10 minutes à vitesse maximale à température ambiante. La phase supérieure est extraite et un volume de chloroforme est ajouté. L'échantillon est vortexé et centrifugé 10 minutes à vitesse maximale à température ambiante.

La phase supérieure est prélevée et 1/10<sup>e</sup> de volume de NaAc (5M) et 2 volumes d'ETOH 95°C froid sont ajoutés. L'échantillon est centrifugé 20 minutes à vitesse maximale à 4°C, le surnageant est éliminé et le culot est rincé avec 125µl d'éthanol à 70%. L'échantillon est centrifugé 15 minutes à vitesse maximale à 4°C, le surnageant est éliminé et le culot est séché à température ambiante. Le culot est resuspendu dans 125µl de TE 1X ou d'eau.

#### II.1.2. Extraction d'ADNg en plaque d'*Arabidopsis thaliana*

Une autre méthode d'extraction consiste à utiliser des filtres, assemblés en plaque 96, pour purifier les ADN. Cette méthode est bien évidemment plus couteuse car les plaques doivent être commandées chez un fournisseur (Whatman, 7700-1801) mais l'ADN est de bien meilleure qualité.

1. Placer une bille dans chaque puits de la plaque
2. Prélever le matériel végétal (50-100 mg. de tissu foliaire) et le mettre dans chaque tube de la plaque (il est conseillé d'avoir la plaque dans l'azote liquide le temps du prélèvement)
3. Mettre les plaques encore congelées dans le mélangeur (broyer), agiter 30 secondes et replonger les plaques dans l'azote.
4. Vérifier que le broyage est complet, en cas d'un mauvais broyage recommencer l'étape 3 jusqu'à broyage complet du matériel végétal.
5. Centrifuger les plaques à 1000g pour faire descendre la poudre.
6. Préchauffer le tampon AP1 à 65°C
7. Déposer 300µl de tampon AP1 + RNase dans chaque tube. (Pour deux plaques 60ml d'AP1 +150µl de RNase A)
8. Fermer et Agiter énergétiquement pendant 15 secondes de haut en bas, retourner légèrement, puis centrifuger à 1000g
9. Ajouter 100µl de tampon AP2 par tube
10. Agiter énergétiquement pendant 15 secondes plus 5 secondes de retournement léger, puis centrifuger à 1000g
11. Mettre les plaques à -20°C pendant minimum 10 minutes. (Laisser la nuit pour favoriser la précipitation des protéines est conseillé)
12. Centrifuger pendant 5 minutes à 2000g
13. Transférer 300µl de surnageant dans une nouvelle plaque de collection
14. Ajouter 1.5 en volume (450µl) de tampon AP3/E à chaque échantillon. (Ne pas oublier l'éthanol dans AP3)
15. Fermer les tubes avec des nouveaux bouchons
16. Agiter énergétiquement pendant 15 secondes plus 5 secondes de retournement léger, puis centrifuger à 1000g
17. placer la plaque de colonnes sur le S-Block
18. Transférer 600µl d'échantillon sur les colonnes
19. Sceller avec une feuille spéciale perméable aux gaz
20. Centrifuger pendant 5 minutes à 4000g (max vitesse en cas de pas avoir une centrifugeuse pour 6000)
21. Vider le S-Block et remettre la plaque de colonne dessus
22. Ajouter doucement 600µl de tampon AW/E à chaque colonne (Ne pas oublier l'éthanol dans AW/E)
23. Fermer avec la feuille
24. Centrifuger pendant 15 minutes à 4000g ou max vitesse
25. Placer la plaque de colonnes sur la plaque d'élution.
26. Ajouter 100µl de Tampon AE à chaque colonne
27. Incuber 1 minute à température ambiante
28. Centrifuger pendant 3 minutes à 4000g ou max vitesse
29. Sceller la plaque de colonnes avec un film métallique
30. Stocker l'ADN élué à -20°C

Composition des tampons :

**Tampon AP1**

(Quantité des réactifs pour 500ml de solution)

Réactifs	Vol (ml)	Quantité (g)
Tris 1 M pH 8.0	50	
EDTA 0.5M	50	
NaCl 2.5 M	100	
SDS 20 %	31.5	
PVP 40.000		5
Bisulfite de sodium		5
H <sub>2</sub> O	268	

**Tampon AP2 (K Ac. 3M – A. Ac. 5M)**

(Quantité des réactifs pour 200ml de solution)

Réactifs	Vol (ml)
Acétate de potassium 5M	120
Acide Acétique Glacial	23
H <sub>2</sub> O	57

**Tampon AP3**

(Quantité des réactifs pour 150ml de solution)

Réactifs	Quantité
Guanidine Chloride	37.25 g
Ethanol 96%	100ml
H <sub>2</sub> O	x

Dissoudre le Guanidine Chloride dans H<sub>2</sub>O pour attendre un volume final de 50ml de solution puis ajouter l'éthanol.

**Tampon AW**

(Quantité des réactifs pour 1 L de solution)

Réactifs	Vol (ml)
Tris 1 M pH 8.0	22.5
EDTA 0.5M	0.2
Acétate de potassium 5M	32
H <sub>2</sub> O	945.3

Le tampon AW/E est préparé en mélangeant 100ml de tampon AW et 170ml d'éthanol 96%

### II.1.3. Extraction d'ADN plasmidique

Une culture liquide bactérienne est culottée par centrifugation (3000g pendant 15 minutes). Le surnageant est éliminé et le culot est conservé. Pendant cette thèse, deux kits d'extraction d'ADN plasmidique ont été utilisés: « Purelink quick plasmid miniprep kit » (Invitrogen™) et « GeneJet plasmid miniprep kit » (Fermentas). Les protocoles sont disponibles en ligne.

### II.1.4. Quantification de l'ADNg ou l'ADNp

Durant cette thèse, l'ADNg a été quantifié sur NanoDrop™ ND-1000 spectrophotometer. Pour tester nos échantillons, 1µl du produit d'extraction est appliqué sur la base de la machine. Trois valeurs sont à observer en priorité :

-la concentration en acides nucléiques (ADN et ARN) de l'échantillon calculée en ng/µl.

-le ratio 260/230 correspondant au degré de pureté de l'échantillon. Les acides nucléiques absorbent à 260nm et certains contaminants d'extraction à 230nm. Plus ce ratio est proche de 2, moins l'échantillon sera contaminé. Il est généralement considéré que 1,8 est le maximum toléré.

-le ratio 260/280 donne une indication sur la quantité de protéine dans l'échantillon. En effet, les protéines absorbent à 280nm. Les protéines sont des contaminants pouvant gêner l'efficacité de futures manipulations, tel que la PCR, obligeant le manipulateur à diluer les échantillons. Encore une fois, plus ce ratio est proche de 2, moins l'échantillon sera contaminé.

## II.2. Génotypage de l'ADN

### II.2.1. Amplification de l'ADN par « Polymerase Chain Reaction (PCR) »

A partir d'une matrice ADN, il est possible d'amplifier, en très grand nombre, les régions d'intérêt par PCR. Le principe repose sur la fixation d'oligonucléotides, spécifiques de la séquence à amplifier, qui seront « allongés » par la Taq (du nom de la souche *Thermus aquaticus*) polymérase à l'aide de dNTP pour copier exactement la portion d'ADN. Les oligonucléotides ont été créés à l'aide du logiciel Primer3 disponible sur internet (<http://frodo.wi.mit.edu/primer3/>).

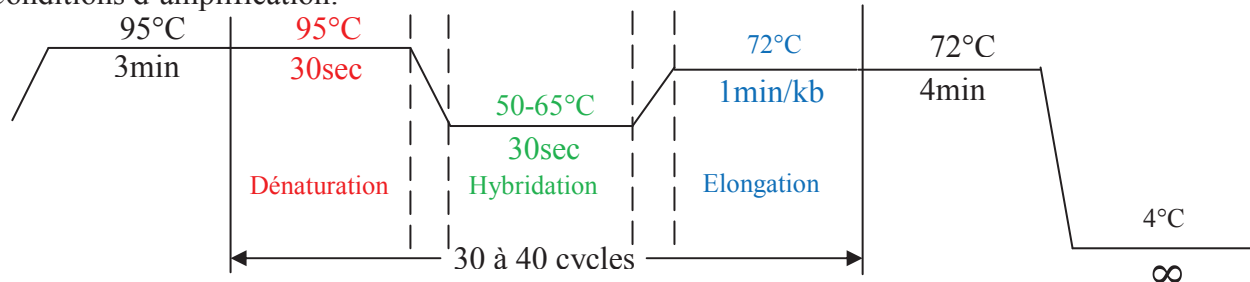
Plusieurs types de Taq polymérases étaient disponibles lors de ma thèse :

- Taq DNA polymérase (Biolabs, M0273) utilisée pour les génotypages et contrôles.

Réactifs	Dans 25µl	Dans 50µl
Tampon (10X)	2,5µl	5µl
Oligonucléotide sens (10µM)	1µl	2µl
Oligonucléotide antisens (10µM)	1µl	2µl
dNTPs (10mM chacun)	0,5µl	1µl
Taq polymérase (5u/µl)	0,125µl	0.3µl
ADNg 50ng/µl	1µl	2µl
H <sub>2</sub> O	18,375µl	37,7µl



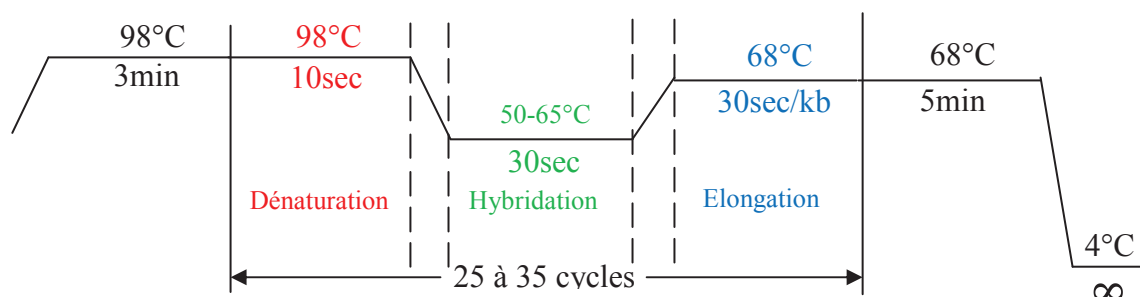
Conditions d'amplification:



- iProof™ high fidelity PCR kit (Biorad, 172-5300) utilisée pour les PCR de clonage.

Réactifs	Dans 20µl	Dans 50µl
Tampon (5X)	4µl	10µl
Oligonucléotide sens (10µM)	1µl	2µl
Oligonucléotide antisens (10µM)	1µl	2µl
dNTPs (10mM chacun)	0,5µl	1µl
Taq polymérase (2u/µl)	0,2µl	0,5µl
ADNg 25ng/µl	1µl	2µl
H <sub>2</sub> O	12,3µl	32,5µl

Conditions d'amplification:



## II.2.2. Electrophorèse sur gel d'agarose

Les produits de PCR (notamment) doivent être vérifiés préalablement à leur utilisation (clonage, séquençage,...). Pour cela, des échantillons sont analysés sur un gel d'agarose. Ce gel permet la séparation des échantillons en fonction de leurs poids moléculaires. Sous l'effet d'un champ électrique, les acides nucléiques (chargés négativement) vont se déplacer vers la borne positive, les éléments de petites tailles migrant plus vite.

Le gel d'agarose liquide est préparé en présence de bromure d'éthidium (BET), un intercalant d'acides nucléiques qui fluoresce aux UV. La concentration des gels le plus souvent utilisé dans cette thèse est de 1% (1g d'agarose pour 100ml de TBE 1X). La migration est effectuée dans du TBE 1X (89 mM TRIS, 89 mM Acide borique, 2 mM EDTA ; pH8) généralement de 50 à 200V (le temps et le voltage variant en fonction des échantillons testés). Le bleu de bromophénol 6X (4g de sucrose, 25mg de bleu de

bromophénol dans 10ml d'eau) est utilisé comme marqueur coloré à une concentration finale de 1X.

### II.2.3. Séquençage des produits de PCR

Le séquençage a été effectué dans une société privée spécialisée (GATCbiotech : [http://www.gatc-biotech.com/en/home.html?gclid=CO\\_E0sKeyqQCFWlc4wods11VUA](http://www.gatc-biotech.com/en/home.html?gclid=CO_E0sKeyqQCFWlc4wods11VUA)). Les résultats sont reçus 2-3 jours après envoi et les séquences sont analysées par bioinformatique.

## II.3. Clonage par enzyme de restriction

### II.3.1. Synthèse et purification des inserts

Les vecteurs utilisés dans cette thèse possèdent un ou plusieurs sites multiples de clonage. Pour cloner les inserts, il a d'abord fallu analyser les différents sites de restriction pour savoir s'il n'y avait pas de site de restriction à l'intérieur du gène. Lorsque les sites de restrictions ont été choisis, des oligonucléotides ont été produits permettant la synthèse du produit d'intérêt avec les sites de restriction aux extrémités 5' et 3'. Les inserts sont synthétisés par PCR (cf II.2.1 du matériel et méthodes).

Avant d'appliquer les enzymes de restriction coupant aux extrémités de l'insert et dans l'ADNp, le produit de PCR est purifié à l'aide du kit de purification « Qiaquick PCR purification kit » (Qiagen, 28104).

### II.3.2. Digestion de l'ADN par enzyme de restriction

Les enzymes de restriction utilisées ont été fournies par New England Biolabs. (NEB : <http://www.neb.com/nebecomm/EnzymeFinder.asp>).

Deux types de digestion ont été effectués :

- Simple digestion = dans le cas où une seule enzyme a été sélectionnée

Réactifs	Dans 30µl
Produit PCR	10µl
Tampon (type à vérifier sur le site NEB)	2µl
Enzyme de restriction (10u/µl)	1,5µl
H <sub>2</sub> O	16,5µl

La réaction est effectuée à 37°C toute la nuit (sans dépasser 16h d'incubation).

- Double digestion = dans le cas où deux enzymes de restriction ont été sélectionnées

Réactifs	Dans 30µl
Produit PCR	10µl
Tampon (type à vérifier sur le site NEB)	2µl
Enzyme de restriction 1 (10u/µl)	1,5µl
Enzyme de restriction 2 (10u/µl)	1,5µl
H <sub>2</sub> O	15µl

La réaction est effectuée à 37°C toute la nuit (sans dépasser 16h d'incubation).

Remarque : il ne faut pas que le volume d'enzyme dépasse 1/10<sup>e</sup> de la réaction de digestion (ici 3µl dans 30µl = 1/10 → OK). A trop haute concentration, le glycérol contenu dans les enzymes peut inhiber la réaction.

### II.3.3. Ligation

La ligation permet le clonage de l'insert dans le plasmide d'intérêt. Pour cela, la T4 DNA ligase (Promega) est utilisée dans les conditions prévues par le fournisseur

Réactifs	Dans 10µl
Vecteur digéré	100ng
Insert digéré	X ng*
Tampon ligase 10X	1µl
T4 DNA ligase	1µl
H <sub>2</sub> O	Qsp 10µl

\* : X est la valeur pour laquelle le ratio molaire vecteur/insert est égal à 3:1. Cette valeur est calculée comme suit : ((ng vecteur x taille en kb d'insert)/ taille en kb du vecteur) x 3/1

La réaction est laissée à 4°C la nuit ou 3h à température ambiante.

### II.3.4. Clonage dans pETDuet-1, pCOLADuet

Les vecteurs pETDUET-1 (Novagen, 71146) et pCOLADUET-1 (Novagen, 71406) sont des vecteurs d'expression contenant deux sites multiples de clonage (MCS). Ils permettent l'expression de protéines recombinantes dans *E. Coli*. Les inserts clonés dans les MCS sont sous le contrôle de deux promoteurs T7 différents. Lorsque le gène de la polymérase T7 est induit par l'IPTG, les gènes insérés vont être transcrits dans *E. Coli*. Les sites de fixation de ribosome (rbs), situés en aval des promoteurs T7, permettent la traduction des transcrits.

Les inserts sont d'abord synthétisés avec, à leurs extrémités, des sites de restriction compatibles aux MCS des vecteurs (dans cette thèse EcoRI/HindIII, AvrII/FseI pour les inserts dans pETDUET-1 et BsrGI pour pColaDUET). Certains de ces sites de restriction ont besoin de quelques paires de bases en amont et en aval du site de restriction pour que le clivage puisse se faire dans des conditions optimales (par exemple, EcoRI a besoin d'une base avant et une base après le site de reconnaissance pour que l'efficacité du clivage soit de 100%. Pour référence: site web NEB, « cleavage close to the end »). Les inserts et les vecteurs sont digérés par double ou simple digestion (cf II.3.2 du matériel et méthodes) puis les inserts sont insérés dans les vecteurs par ligation (cf II.3.3 du matériel et méthodes). Les vecteurs sont insérés dans DH5α pour multiplication (cf III.1.2 du matériel et méthodes).

## II.4. Clonage Gateway™ (Invitrogen™)

Il s'agit d'une technique de clonage permettant l'insertion d'un produit PCR dans un vecteur d'entrée, puis dans un vecteur d'expression, par recombinaison. Les séquences *attB1* et *attB2* clonées aux extrémités du fragment vont permettre l'intégration dans un vecteur donneur par recombinaison avec des sites *attP*. Ce vecteur possède une origine de répllication pour sa multiplication dans *Escherichia coli* et un gène de résistance à la gentamycine permettant la sélection des bactéries transformées. La seconde recombinaison va permettre l'insertion dans un vecteur de destination, et une origine de répllication, par recombinaison entre les sites *attL* et les sites *attR*.

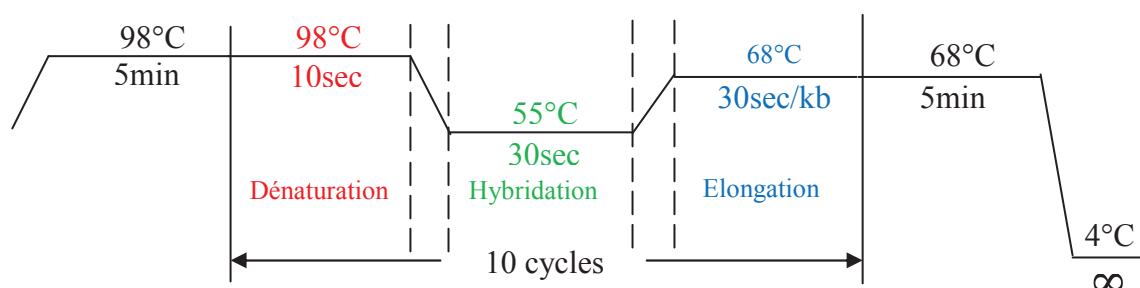
### II.4.1. Amplification par PCR du fragment à cloner

Avant de commencer les étapes de recombinaison, il faut produire un produit PCR portant à ses extrémités les sites *attB1* et *attB2*. Pour cela, deux PCR successives sont effectuées.

La première (PCR1) permet d'amplifier le gène d'intérêt et une partie d'*attB1* et *attB2*. La deuxième (PCR2) va permettre d'ajouter *attB1* et *attB2* grâce à des amorces universelles en prenant comme matrice la première réaction PCR.

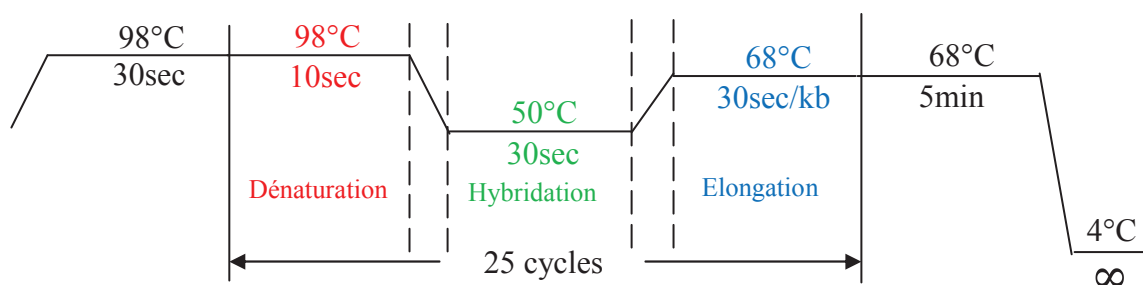
Réactifs PCR1	Dans 10µl
Oligonucléotide sens (10µM)	1µl
Oligonucléotide antisens (10µM)	1µl
dNTP (10mM chacun)	0,2µl
Tampon iProof 5X	2µl
iProof polymerase (2u/µl)	0,1µl
ADN génomique	1µl
Eau	4,7µl

Conditions d'amplification PCR1:



Réactifs PCR2	Dans 40µl
Oligonucléotide universel sens (10µM)	2µl
Oligonucléotide universel antisens (10µM)	2µl
dNTP (10mM chacun)	1µl
Tampon iProof 5X	8µl
iProof polymerase (2u/µl)	0,4µl
PCR1	10µl
Eau	16,6µl

Conditions d'amplification PCR2:



#### II.4.2. Purification du produit PCR

Cette étape permet d'éliminer les oligonucléotides et les dimères d'amorces. Il faut ajouter à 45µl de produit PCR2, 225µl de solution PEG/Mg/TE (2 volumes de PEG 30% 8000, 30mM MgCl<sub>2</sub> et 3 volumes de TE) et 1µl de glycogène azur (favorise la précipitation et colore en bleu les culots). Le culot est resuspendu dans 20µl de TE.

#### II.4.3. Recombinaison BP (Invitrogen™, 11789-013)

Par cette réaction, le produit de PCR *attB1/attB2* est cloné dans un vecteur d'entrée pDONR207 contenant les sites *attP1/attP2*. Le clonage permet la formation des sites *attL1/attL2* dans pDONR207. La cassette contient le gène CCDB, un poison de l'ADN gyrase, permettant la sélection positive des bactéries. En effet, s'il y a recombinaison, le gène CCDB est éliminé. Dans le cas contraire le gène CCDB est maintenu dans la cassette et la réplication bactérienne ne peut se faire.

Réactifs BP	Dans 5µl
Produit de PCR <i>attB</i>	1-2µl
Vecteur d'entrée pDONR207 (150ng/µl)	1µl
Tampon BP clonase 5X	1µl
BP clonase	1µl
Eau	0-1µl

La réaction a lieu à 25°C pendant 1h minimum.

Les 5µl de produit BP sont mis en contact avec 100µl de bactéries compétentes DH5α. Le mélange est placé 20 minutes sur la glace, 45 secondes à 42°C et 5 minutes sur la glace. Le mélange est ensuite ajouté à 900µl de SOC et les bactéries sont incubées à 37°C pendant 1h sous agitation. 100µl et 900µl de bactéries sont étalés sur boîte LB gentamicine (50mg/l). Les boîtes sont incubées la nuit à 37°C. Les colonies sont testées par PCR (cf III.1.2 du matériel et méthodes) et l'ADN plasmidique est extrait (cf II.1.3 du matériel et méthodes).

#### II.4.4. Recombinaison LR (Invitrogen™, 11791-019)

Cette dernière étape va permettre le transfert des inserts clonés d'un vecteur pDONR207 à un vecteur d'expression par recombinaison des sites *attL* du clone d'entrée et les sites *attR* du clone de destination pour donner les sites *attB*. Les plasmides les plus utilisés dans cette thèse sont pGWB1 et pGWB2.

Réactifs LR	Dans 10µl
Vecteur de destination (150ng/µl)	1µl
pDONR207 (300ng/µl)	1-5µl
Tampon LR clonase 5X	2µl
LR clonase	2µl
Eau	0-4µl

La réaction a lieu à 25°C pendant 1h minimum.

Les souches DH5α sont transformées avec 5µl de produit de réaction LR puis étalées sur boîte contenant 50mg/l de kanamycine. Les colonies sont ensuite testées par PCR de façon identique à ce qui a été décrit précédemment.

## II.5. Clonage pGEMT-easy

### II.5.1. Ligation

Les produits PCR purifiés peuvent être directement clonés dans le vecteur pGEMT-Easy® (Promega, A1360) en suivant les instructions du fournisseur. Ce vecteur linéarisé possède en 3'-terminal une thymidine limitant la re-linéarisation du vecteur. Ces thymidines permettent le clonage de produits PCR générés par certaines polymérases qui insèrent une adénosine en dernière base.

Cependant toutes les polymérases n'effectuent pas cet ajout en 3' (comme la polymérase iProof® utilisée dans cette thèse). Il faut ajouter en 3' une adénosine par le biais de la technique du « A-tailing ». A 7µl de produit de PCR purifié, il faut ajouter 1µl de tampon polymérase 10X, 0.2mM de dATP, 5 unités de Taq polymérase et compléter à 10µl à l'eau déionisée. La réaction est incubée à 70°C pendant 30 minutes et 1µl de cette réaction est utilisé pour la ligation pGEMT-easy®.

### II.5.2. Sélection des transformants

Des bactéries DH5α thermocompétentes (cf III.1.2 du matériel et méthodes) sont transformées avec les produits de ligation. La sélection a lieu sur milieu LB agar contenant l'antibiotique ampicilline ainsi que du X-gal (80µg/ml) et l'IPTG (0.5mM), ceci pour induire une sélection « blanc/bleu ». Les colonies blanches sont, à priori, positives et les bleues négatives.

## II.6. Mutagénèse par PCR

### II.6.1. Délétion de région

Il s'agit de produire des versions tronquées d'une portion d'ADN par PCR. Pour cela, 15 nucléotides avant et après le site à éliminer, en sens et en antisens, ont été sélectionnés afin de produire par PCR deux fragments (PCR1). Ces deux fragments purifiés sont « fusionnés » grâce à une deuxième PCR utilisant un oligonucléotide

chevauchant l'extrémité 3' du premier fragment et l'extrémité 5' du second fragment (PCR2).

A noter que la première PCR se fait en général à une température d'hybridation de 55°C.

#### II.6.2. Remplacement de nucléotides

Par PCR, il est aussi possible de muter ponctuellement un gène pour modifier la séquence en acides aminés de la protéine. Il faut pour cela sélectionner des amorces s'hybridant sur le site à muter et les modifier selon l'acide aminé à produire par la suite.

Deux fragments sont ainsi produits (l'un avec la mutation en 3' l'autre en 5') et fusionnés grâce à une deuxième PCR utilisant un oligonucléotide à l'extrémité 5' du premier fragment et un oligonucléotide à l'extrémité 3' du second fragment (PCR2).

La première PCR se fait en général à une température d'hybridation de 55°C.

### III. Méthodes relatives aux bactéries

#### III.1. DH5 $\alpha$

##### III.1.1. Préparation de bactéries DH5 $\alpha$ thermo-compétentes

Une préculture de 10ml dans du LB est lancée sur la nuit à 37°C à 200rpm. La préculture estensemencée dans 500ml de milieu SOB à 19°C sous agitation jusqu'à l'obtention d'une DO (600nm) comprise entre 0,5 et 0,6. La culture est incubée 10 minutes sur la glace et centrifugée à 5000g pendant 10 minutes à 4°C (en 2 flasques de 1l). Chaque culot est repris dans 80ml de TB froid et la resuspension est incubée sur glace 10 minutes. Les resuspensions sont centrifugées 5000g pendant 10 minutes à 4°C. Chaque culot est repris dans 20ml de TB froid et 2,8ml de DMSO sont ajoutées puis incubées 10 minutes sur la glace. Les aliquots sont congelés dans l'azote liquide et conservés à -80°C.

Milieu SOB : 20g/l bacto-tryptone, 5g/l extrait de levure, 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).

##### III.1.2. Transformation de bactéries DH5 $\alpha$ thermo-compétentes et sélection

Le produit de ligation est transféré dans des bactéries DH5 $\alpha$  thermo-compétentes préparées par nos soins (à décongeler sur glace). Le principe repose sur l'application de deux chocs thermiques successifs permettant l'entrée du vecteur dans la bactérie.

De 10 à 100ng de vecteur sont mis en contact avec 100 $\mu$ l de bactéries et laissé sur glace 30 minutes (homogénéisation par léger tapotement). Le tout est placé à 42°C pendant 45 secondes (premier choc thermique) puis mis immédiatement sur glace 3 minutes (second choc thermique). Après transformation, les bactéries sont mises en culture dans 900 $\mu$ l de milieu LB liquide ou 900 $\mu$ l de SOC, milieu plus riche que le LB et se contaminant plus facilement. La culture a lieu à 37°C pendant une heure sous agitation (700rpm). 100 $\mu$ l ou la totalité de la culture sont ensuite placées sur milieu LB solide (agarose) contenant l'antibiotique associé au vecteur. Les boîtes sont ensuite placées à 37°C pendant la nuit.

Milieu LB liquide	Dans 1L
Bactotryptone	10g
Extrait de levure	5g
NaCl	10g

Milieu LB solide (pH7.5)	Dans 1L
Bactotryptone	10g
Extrait de levure	5g
NaCl	10g
Agarose	15g



Milieu SOC (pH7)	Dans 1L
Bactotryptone	20g
Extrait de levure	5g
NaCl	0.5g
KCl	0,186g
MgCl <sub>2</sub>	0,952g
Glucose	3,603g

Le lendemain, les colonies sont piquées avec une pointe de cône de 1-10µl et analysées par PCR.

Réactifs PCR	Dans 25µl
Oligonucléotide sens (10µM)	1µl
Oligonucléotide antisens (10µM)	1µl
dNTPs (10mM chacun)	0,5µl
Tampon de PCR 10X	2,5µl
Rouge de crésol	2.5µl
Taq polymérase	0,125µl
H <sub>2</sub> O	17,375µl

La PCR est effectuée selon les conditions des amorces. La PCR est testée sur gel agarose (BET) et observée aux UV.

Les colonies ayant réagi positivement à la PCR sont mises à pousser en milieu LB liquide et stockées à -80°C avec du glycérol (300µl glycérol 50% + 700µl de culture).

L'ADN plasmidique est ensuite extrait (cf II.1.3 du matériel et méthodes) et l'insert est séquencé.

### III.2. Transformation de bactéries BL21/Rosetta2 thermo-compétentes

Lorsque les vecteurs ont été vérifiés par séquençage, ceux-ci sont insérés dans BL21 Star (DE3) one Shot<sup>®</sup> (Invitrogen<sup>™</sup>, C6010-03) ou Rosetta<sup>™</sup> 2(DE3) (EMDMillipore, 71397-3) par choc thermique. Dix nanogrammes de vecteurs sont mis en contact avec 100µl de bactéries et laissés sur glace pendant 30 minutes. Le premier choc thermique a lieu à 42°C pendant 30 secondes, le second sur glace pendant 3 minutes. Une préculture est réalisée à 37°C pendant 1h sous agitation (700rpm) après ajout de 250µl de SOC. Les bactéries sont étalées sur milieu LB agar contenant l'antibiotique approprié. La sélection des colonies positives est réalisée par PCR (cf III.1.2 du matériel et méthodes).

### III.3. Induction de souches d'E. Coli BL21/Rosetta2

Deux protocoles ont été utilisés lors de cette thèse, le premier à l'URGV et le deuxième à Perth. Dans le cas de l'expression de PPR, le deuxième protocole donne un meilleur taux de protéines solubles.

(1) Les bactéries transformées précédemment sont striées sur milieu LB agarose afin d'obtenir des colonies isolées. Une colonie est alors sélectionnée et mise en culture avec

5ml de LB liquide. Cette préculture est incubée une nuit à 37°C jusqu'à obtenir une DO de 0.6 à 1. La culture est alors diluée 20 fois (1ml de préculture environ + 19ml de LB liquide). Cette culture est incubée à 37°C pour obtenir une DO de 0.4 (~2h30).

La souche contient le promoteur inductible à l'IPTG *lacUV5* contrôlant le gène de l'ARN polymérase. Pour initier l'induction, 0.5mM final d'IPTG doit être ajouté à la culture. Cette culture peut être ensuite incubée à 37°C pendant 3h mais augmente le risque d'obtenir des agrégats de protéines. Une induction pendant 5h à 15°C sous agitation a été préférée pour ainsi mieux « contrôler » le niveau d'expression des protéines.

Enfin, les bactéries sont centrifugées, le surnageant est éliminé et le culot est conservé à -70°C jusqu'à utilisation.

(2) Les bactéries transformées précédemment sont striées sur milieu LB agarose afin d'obtenir des colonies isolées. Une colonie est alors sélectionnée et mise en culture avec 5ml de LB liquide. Cette préculture est incubée la nuit à 37°C. La culture est alors diluée 100 fois (5ml de préculture + 500ml de LB liquide). Cette culture est incubée pendant 3h à 37°C puis placée 30 min à 16°C. Pour initier l'induction, 0.4mM final d'IPTG doit être ajouté à la culture. Cette dernière est ensuite incubée la nuit (environ 16h) à 16°C. Enfin, les bactéries sont centrifugées pour être lysées le jour même.

### III.4. C58C1

#### III.4.1. Préparation de bactéries C58C1 thermo-compétentes

Un glycérol d'*agrobacterium tumefaciens* est étalé sur boîte. Une préculture est effectuée en ensemençant 8ml de LB agro avec une colonie d'*Agrobacterium tumefaciens*. La préculture est incubée pendant la nuit à 28°C. Dans un erlenmeyer, contenant 192ml de milieu LB agro, ajouter 8ml de cette préculture et incubé 3h-4h à 28°C sous agitation afin d'obtenir une DO à 600 nm comprise entre 0,4 et 0,6. Centrifuger 15 min à 4000 rpm à 4°C. Chaque culot est repris dans 10ml de Tris-HCl 10 mM pH 7,5 froid puis centrifugé 15 min à 4000 rpm à 4°C. Chaque culot est resuspendu dans 4ml de LB agro froid. Les bactéries sont congelées dans de l'azote liquide et conservées à -80°C.

Milieu LB agro	Dans 1L	Milieu LB agro agar	Dans 1L
Bactotryptone	10g	Bactotryptone	10g
Extrait de levure	5g	Extrait de levure	5g
NaCl	5g	NaCl	5g
NaOH 2N	pH7.5	NaOH 2N	pH7,5
-	-	Agarose	15g
Autoclaver	-	Autoclaver	-

#### III.4.2. Transformation de bactéries C58C1 thermo-compétentes

Un tube de 1.5ml de bactéries compétentes est décongelé sur la glace. Les tubes contenant 1 µg de plasmide sont placés dans un bain d'eau glacée pendant 5 min.

Ajouter 200µl de bactéries compétentes par tube. L'incubation est laissée 5 min sur glace.

Les chocs thermiques sont effectués en plaçant d'abord le(s) tube(s) 5 min dans de l'azote puis en plaçant le(s) tube(s) 5 min à 37°C. Ajouter 800µl de LB Agro et laisser incuber 2h à 28°C sous agitation douce. Etaler 200µl sur LB Agro agar. Incuber 48h à 28°C.

### III.5. DH10B

#### III.5.1. Préparation de bactéries DH10B électro-compétentes

Ce protocole a été utilisé ici pour une souche de DH10B mais ce protocole est adapté aussi pour des souches telles BL21, Rosetta, ...

Avant de commencer, il faut préparer 30ml de glycérol 10% froid (4°C). Inoculer 5ml de milieu LB avec la une colonie de la souche à préparer. La préculture est incubée sur la nuit à 37°C. Inoculer 50ml de LB avec 5ml de préculture et laisser pousser à 37°C jusqu'à ce que la DO soit égale à 0,8 (environ 2h30). La culture est centrifugée à 4000g pendant 10min à 4°C dans deux tubes Falcon de 50ml.

Le surnageant est éliminé et le culot est resuspendu dans 2ml de glycérol 10% froid. Les bactéries sont centrifugées à 8000g pendant 2min à 4°C dans des tubes de 2ml. Cette étape est répétée 4 fois. Les cultures sont finalement resuspendues dans 400µl maximum de glycérol 10%.

#### III.5.2. Transformation de bactéries DH10B électro-compétentes

Mélanger 20µl de cellules avec 2.5µl de BP ou LR dessalées par dialyse sur colonne Sepharose. Les cellules sont placées entre deux électrodes dans un bain de glace. Le choc électrique est effectué à 400V.

Les cellules sont prélevées et mises en culture dans 400µl de LB à 37°C pendant 45min sous agitation. La culture est finalement étalée sur boîte contenant l'antibiotique approprié.

## IV. Méthodes relatives aux ARN

### IV.1. Extraction d'ARN total et traitements

L'ARN est une molécule extrêmement fragile car les RNases sont très stables. Afin de protéger les échantillons, il est impératif de travailler avec des gants, des cônes « RNase free » d'utiliser de l'eau « nuclease free » (ou DEPC) et de manipuler au maximum à 4°C.

#### IV.1.1. Extraction par le biais du kit RNeasy plant Minikit (Qiagen, 74903)

L'utilisation du kit RNeasy plant Minikit facilite grandement l'extraction (gain de temps et peu de manipulation de produits dangereux) mais est onéreux et n'extrait pas la totalité des ARN. En effet, les ARN de taille inférieure à 100pb ne sont pas retenus par la colonne. Néanmoins, ce kit a été utilisé à maintes reprises. Le protocole est disponible en ligne sur le site de Qiagen :

<http://www.qiagen.com/products/rnastabilizationpurification/rneasysystem/rneasyplantmini.aspx#Tabs=t2>.

#### IV.1.2. Extraction au Trizol «RNA Now »™ (Ozyme, BX-102)

Au contraire de l'extraction avec le kit, l'extraction au Trizol permet d'extraire tous les ARN mais nécessite l'utilisation de substances dangereuses (phénol, chloroforme). De plus, certaines étapes sont critiques et peuvent, si elles sont mal faites, gêner les manipulations ultérieures. Le Trizol a été commandé chez Ozyme sous le nom « RNA Now ».

Immédiatement après broyage dans l'azote liquide, les échantillons sont homogénéisés dans 1ml de Trizol pour 50-100 mg de tissu. Les échantillons sont incubés 5 minutes à température ambiante et 0,2ml de chloroforme pour 1ml de Trizol est ajouté. Les tubes sont inversés 15 secondes et mis à incuber 2 à 3 minutes à température ambiante. Les échantillons sont centrifugés à 12 000 g maximum 15 minutes à 4°C.

Après centrifugation, le mélange est séparé en une phase inférieure rouge, la phase de phénol-chloroforme, une interphase, et une phase aqueuse supérieure sans couleur. L'ARN reste exclusivement dans la phase aqueuse. Le volume de la phase aqueuse est environ 60% du volume de Trizol utilisé pour l'homogénéisation.

La phase aqueuse est donc transférée dans un nouveau tube et les ARN sont précipités avec 0.5ml d'isopropanol pour 1ml de Trizol. Les échantillons sont incubés 10 minutes à température ambiante et centrifugés à 12 000 g maximum 15 minutes à 4°C. Le surnageant est éliminé et le culot est rincé avec 1ml d'éthanol à 75% pour 1ml de Trizol. Une dernière centrifugation est nécessaire à 7 500g maximum 5 minutes à 4°C. Les culots sont resuspendus dans de l'eau « nucléase free » (30µl environ pour une jeune feuille).

#### IV.1.3. Traitement de l'ARN total à la Turbo™DNase

Les échantillons sont enrichis en ARN mais on ne peut empêcher une contamination par de l'ADNg. Au vu des expérimentations futures, l'ADNg doit être éliminé. Pour cela, la

Turbo<sup>TM</sup>DNase (Ambion, AM2238) a été utilisée. Le protocole suivant fonctionne pour 15µg d'ARN.

Réactifs DNase	Dans 50µl
ARN	15µg
Tampon 10X	5µl
Turbo <sup>TM</sup> DNase 2u/µl	1µl
Eau	Qsp50µl

Les échantillons sont placés à 37°C pendant 30 minutes puis 1µl de Turbo<sup>TM</sup>DNase est à nouveau ajouté. Les ARN sont incubés à 37°C pendant 30 minutes.

L'ARN est ensuite précipité avec 1/10<sup>e</sup> de volume AcNa 5M et 2.5 volumes d'EtOH 95 froid. Les échantillons sont placés à -80°C pour une nuit puis centrifugés à 20000 g pendant 15 minutes à 4°C. Le culot est rincé avec 200µl d'EtOH 70 et les tubes sont centrifugés 5 minutes à 6000 g à 4°C. Le culot est repris dans de l'eau « nucléase free ».

#### IV.1.4. Quantification de l'ARN total

De la même façon que pour l'ADN, l'ARN peut être quantifié au Nanodrop (cf II.1.4 du matériel et méthodes).

Les mesures effectuées au Nanodrop sont néanmoins peu précises ce qui peut être un problème pour les expériences de transcriptomique quantitative (cf IV.6 du matériel et méthodes).

Pour plus de précision, les ARN sont dosés au Ribogreen (Molecular Probes, R-11490). Le Ribogreen est une molécule qui fluoresce peu à l'état libre mais qui, lorsqu'il est complexé aux acides nucléiques, fluoresce de manière très importante. La fluorescence est détectée par une sonde et les ARN peuvent être quantifiés de manière très précise. Le protocole donné ici est celui utilisé par la plateforme transcriptome, dirigée par Sandrine Balzergue à l'URGV.

- Préparation de la gamme :

Prendre un aliquot de 4µl (Ladder ARN 0,5-10kb dilué) à -20°C.

Ajouter 76µl de TE et préparer la gamme directement dans une plaque noire selon le tableau suivant :

Puit de la plaque noire	Volume de TE (µl)	Volume d'ARN gamme (µl)	Quantité (ng)
A1	30	20	100
B1	34	16	80
C1	38	12	60
D1	42	8	40
E1	46	4	20
F1	48	2	10
G1	49	1	5
H1	49,5	0.5	2,5

- Préparation des échantillons :

Diluer les ARN à quantifier avec du TE (en barrettes PCR), prévoir une concentration d'environ 50ng/μl.

Déposer 49μl de TE dans le nombre de puits correspondant au nombre d'échantillons à quantifier à partir du puits A2 (inutile de compléter la dernière colonne jusqu'au puits Hx).

Ajouter 1μl d'échantillon dilué par puits.

- Préparation du Ribogreen :

Le Ribogreen ne décongèle pas sur glace, le laisser à température ambiante.

Calcul du volume de Ribogreen dilué à utiliser :  $V = (\text{Nb d'échantillons} + 8 \text{ pour la gamme } + 1) \times 50$ . Préparer la solution dans un tube 1,5ml noir: V de TE + V/200 de ribogreen pur.

Ajouter au dernier moment 50μl de Ribogreen dilué dans tous les puits (gamme + échantillons).

Couvrir la plaque du couvercle opaque et agiter pendant 5 minutes sur l'agitateur à plaques.

Lecture sur ordinateur à l'aide du logiciel BIOLISE.

## IV.2. Réaction de transcription inverse

Il s'agit de la synthèse d'un brin d'ADN complémentaire (ADNc) à partir d'une matrice ARN. Le but est d'obtenir un équivalent ADN d'un transcrit. A partir de ce brin d'ADNc, il est possible d'effectuer une PCR pour séquencer un transcrit (et ainsi voir le niveau d'édition relatif d'un transcrit). La synthèse d'ADN complémentaire est effectuée en utilisant la reverse transcriptase SuperscriptII™ (Invitrogen™, 11904-018).

La synthèse est effectuée à partir de 150ng à 5μg d'ARN traité à la DNase (cf IV.1.3 du matériel et méthodes) mélangé à 1μl de dNTP (10mM), 100ng d'hexanucléotides ou 2μL oligonucléotide dT (10μM), le tout devant être ajusté à un volume final de 12μl (avec de l'eau DEPC). Ce mix est chauffé 5 minutes à 65°C puis placé 1 minute sur glace. Il est ajouté 4μl de tampon de réaction 5X, 2μl de DTT, 1μl de RNase OUT™ (Invitrogen™, 10777-019). Le tout est incubé à 25°C pendant 2 minutes puis 1μl de transcriptase inverse est ajouté. La réaction complète est placée à 25°C pendant 10 minutes (hybridation) puis 50 minutes à 42°C (élongation) et enfin 20 minutes à 80°C (inactivation).

En général, des témoins négatifs sont ajoutés ; il s'agit de la réaction complète sans transcriptase inverse. Ceci permet de voir s'il reste une contamination ADN génomique dans les ARN.

## IV.3. Northern Blot

Le Northern blot est une technique permettant une quantification relative des ARN ainsi que l'observation de différentes formes de maturation d'un ARN (épissage, clivage). La détection se fait par hybridation d'une sonde marquée sur les acides nucléiques (ARN).

#### IV.3.1. Transcription *in vitro* : synthèse de la sonde

Un produit PCR correspondant à la sonde est cloné dans le vecteur pGEM®T-easy, contenant le promoteur SP6. L'insertion doit produire par transcription un ARN antisens pour pouvoir s'hybrider au transcrit d'intérêt.

La transcription *in vitro* est obtenue par le biais d'un kit (Ambion, Maxiscript). Un produit de PCR est amplifié à partir du plasmide pGEM®T-easy grâce à un couple d'oligonucléotides se situant de part et d'autre de l'insert. Le produit de PCR est ensuite purifié et précipité. Le culot est dissout dans 10µl d'eau. Le mélange qui suit est réalisé à température ambiante.

Réactifs Transcription <i>in vitro</i>	Dans 20µl
Produit PCR	5µl
Tampon 10X	2µl
dATP (10 mM)	1µl
dGTP (10 mM)	1µl
dUTP (10mM)	1µl
dCTP (10mM)	0,6µl
biotin-14 CTP (10 mM)	0,4µl
SP6 polymerase mix	2µl
H2O	7µl

L'incubation a lieu à 37°C et la réaction est arrêtée après ajout de 1µl de Turbo™DNase à 37°C pendant 30 minutes. 1µl d'EDTA 0.5M est ajouté pour inhiber la DNase.

#### IV.3.2. Préparation des ARN à tester

L'ARN est extrait par le kit RNeasy plant Minikit (Qiagen) ou par Trizol (Ozyme). La quantité d'ARN est estimée au Nanodrop. Les ARN sont ensuite traités à la turbo DNase (cf IV.1.3 du matériel et méthodes) et quantifiés à nouveau. Dans cette thèse, 15µg d'ARN ont été soumis au « speedvac » permettant la concentration des acides nucléiques par évaporation. Le culot est resuspendu dans 3µl d'eau. Deux volumes de tampon ARN (65µl de 10 X MOPS, 110µl 35 % formaldéhyde, 320µl formamide pour 500µl) sont ajoutés à l'ARN ainsi qu'à 3µl de marqueur ARN (Promega, G3191). Les échantillons sont incubés à 55°C pendant 15 minutes avant d'être mis sur glace. Enfin, 2µl de tampon de charge sont ajoutés.

#### IV.3.3. Northern blot: transfert

Les échantillons doivent d'abord être séparés sur gel d'agarose contenant du formaldéhyde. La cuve d'électrophorèse est traitée avec du RNaseZap (Ambion, AM9780) et rincée avec de l'eau DEPC. Pour un gel à 1.2%, 1.2g d'agarose sont dissous dans 70ml d'eau DEPC puis porté à ébullition. 20ml de formaldéhyde et 10ml de MOPS 10X sont ajoutés sous une hotte et le gel est coulé dans un portoir préalablement traité au RNaseZap. Les échantillons sont mis à migrer dans un tampon

MOPS 1X à 70V pendant 1h30. Le gel est ensuite rincé dans 2 bains d'eau DEPC pendant 5 minutes puis dans un bain de 10X SSC pendant 45 minutes.

Pour effectuer le transfert, 4 papiers Whatman et un morceau de membrane de nylon sont coupés à la taille du gel. Un « pont » de Whatman de la largeur du gel est coupé afin que le papier soit en contact avec le tampon 20X SSC. Le transfert est effectué comme suit (de la première couche à la dernière couche) :

- le pont Whatman trempant dans le 20X SSC,
- un papier Whatman,
- le gel face avant vers le sol (les puits sont enlevés mais marqués sur la membrane),
- un Whatman humidifié,
- deux Whatman,
- 10cm de papier absorbant,
- un poids permettant le plaquage des différentes strates.

Le transfert est effectué pendant la nuit grâce au tampon SSC.

#### IV.3.4. Northern blot: hybridation

La membrane est rincée dans un bain de 5X SSC pendant 5 minutes et elle est soumise à un « cross-linker » à  $0.12\text{J}/\text{cm}^2$  pour fixer les ARN. La membrane est soumise à 3 bains différents : acide acétique 5% pendant 15 minutes, NaAc 0.5M pH5.2 avec 0.04% de bleu de méthylène, 3 rinçages à l'eau DEPC. Les bandes correspondant à l'ARN apparaissent et la membrane peut être prise en photo. La membrane est conservée à 4°C jusqu'à utilisation.

L'hybridation se fait en plusieurs étapes. La première est un rinçage de la membrane dans un bain de 5X SSC puis la membrane est placée dans un tube d'hybridation avec 10ml de tampon d'hybridation (2,5ml 20x SSC (5x), 5ml formamide (50 %), 10µl 10 % heparin (100ug/ml), 500µl 10 % SDS (0,5 %), ajouter de l'eau à 10ml). La préhybridation est effectuée à 68°C pendant 1 heure. La sonde (5µl) est dénaturée 5 minutes à 75°C et placée sur glace. La sonde est ajoutée au tampon d'hybridation après l'heure de préhybridation. L'hybridation est incubée pendant la nuit à 68°C.

#### IV.3.5. Northern blot: détection

La membrane hybridée est lavée dans 3 bains successifs :

- 3x15min 1X SSC / 0,5% SDS (température ambiante),
- 20min 0.1X SSC / 0,1% SDS (60°C),
- 1h 0.1X SSC / 0,1% SDS (60°C).

Le signal est ensuite détecté par le biais du kit « Pierce Chemiluminescent Nucleic Acid Detection » (Pierce, 89880). La membrane est exposée grâce à un ordinateur couplé à la fonction « image Quant RT-ECL » (GE Healthcare) pendant 1h.



#### IV.4. Extension d'oligonucléotides « empoisonnés » (PPE)

La PPE (Poisoned Primer Extension) repose sur l'intégration de didéoxyribonucléotides lors d'une réaction d'élongation permettant la discrimination entre des sites édités (ADNc) et des sites non édités (ADNg). L'ADN est amplifié par un couple d'amorces encadrant le site étudié puis une élongation (avec une amorce marquée) permet l'incorporation de ddCTP/ddGTP arrêtant l'extension de l'ADN non édité. Le protocole de départ a été mis au point par Anne-Laure Chateigner-Boutin à Perth.

##### IV.4.1. Purification de l'ADN complémentaire amplifié par PCR

Afin d'éliminer les primers et les dNTP en excès, le kit ExoSAP-IT® (USB, 78200) est utilisé. Il est composé de deux enzymes réunies dans un même tube : l'exonucléase I (élimine les primers) et la phosphatase alcaline de chimpanzé (hydrolyse les dNTPs).

Réactifs de purification PCR	Dans 7µl
Produit PCR	5µl
ExoSAP-IT®	2µl

Cette réaction est incubée 15 minutes à 37°C puis 15 minutes à 80°C.

##### IV.4.2. Réaction de PPE

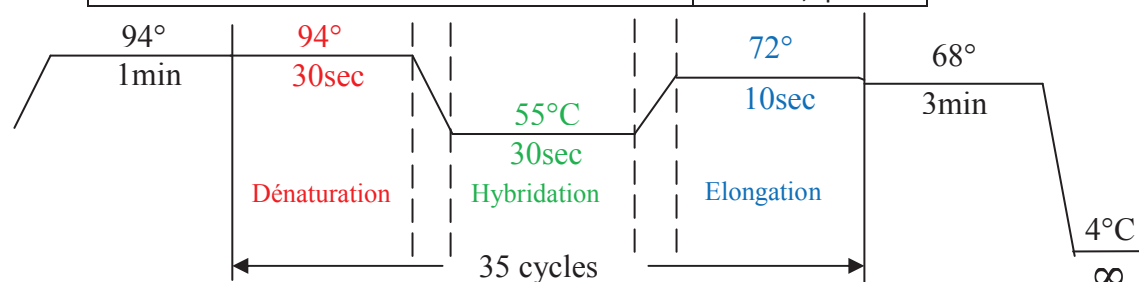
L'extension de l'ADNc est réalisée grâce à :

-une amorce marquée à l'IRD-700 (Evry) ou 5-FAM (Perth) en 5', située le plus proche possible du site d'édition à analyser (en 3' ou en 5' du site).

-un mix de 3dNTPs + 1 ddGTP ou 1 ddCTP (dépendant de l'endroit où se fixe le primer). Lorsque ces ddCTP/ddGTP sont intégrés, l'extension s'arrête. Dans le cas où il y a édition, le ddCTP/ddGTP est intégré plus loin que le site d'édition car la cytidine est « transformée » en uridine, ce sera à la prochaine cytidine que le ddCTP/ddGTP sera intégré.

-une enzyme utilisée spécialement pour le séquençage, la ThermoSequenase® (Thermo Sequenase Cycle Sequencing Kit, USB, 78500).

Réactifs de PPE	Dans 20µl
Produit PCR traitée à l'ExoSAP-IT®	2µl
Tampon de réaction 10X	2µl
Oligonucléotide marqué (1µM)	0,5µl
Mix 3dNTP-1ddCTP/ddGTP (1mM chacun)	2µl
Thermosequenase®	1µl
Eau	12,5µl



A partir du moment où le primer marqué est ajouté, la réaction doit être manipulée à l'abri de la lumière.

Après l'extension au thermocycler, la réaction est incubée avec 5µl de solution « stop » (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF) à 70°C pendant 2 minutes. La réaction est laissée sur glace jusqu'à analyse.

#### IV.4.3. Analyse de la réaction de PPE sur gel de séquençage

(1) Méthode de détection à Perth (Typhoon Trio imager, GE Healthcare)

- *Monter le support :*

Nettoyer les plaques à l'aide d'eau puis d'éthanol (lavage au Kimwipes). Ajouter les espaceurs et mettre les plaques les unes sur les autres. Ajouter les serre-joints et placer les plaques légèrement inclinées (le haut étant surélevé). Ajouter l'adaptateur en bas du gel et le relier au tuyau.

- *Préparation du gel :*

Préparer un gel d'acrylamide à 12% permettant une meilleure séparation des composés compris entre 10 et 50 nucléotides. 15ml d'acrylamide/bis-acrylamide (19 :1) 40%, 21g d'urée, 5ml de TBE 10X et qsp 50ml H<sub>2</sub>O. Homogénéiser avec un barreau aimanté et ajouter 20µl de TEMED et 400µl d'APS 10% → passer **rapidement** à l'étape suivante.

- *Couler le gel :*

Avec une seringue reliée au tuyau, couler le gel entre les deux plaques en évitant la formation de bulles. Ajouter le peigne au sommet et vérifier la qualité du gel (bulles et fuites). Attendre 2h la polymérisation.

- *Mise en place du gel et lancement du logiciel de visualisation :*

Enlever la seringue, l'adaptateur et le peigne et ajouter immédiatement de l'eau dans les puits. Préparer 1L de TBE 1X préchauffé 3 minutes 30 au micro-onde. Ajouter 1cm de TBE 1X dans la cuve inférieure, placer le gel dedans et ajouter le reste de TBE entre les deux plaques. Nettoyer les puits avec une pipette de 200µl. Effectuer la pré-migration 5 minutes à 55W puis ajouter 5µl de chaque produit de PPE. Laisser migrer à 55W pendant 2h.

Allumer le scanner permettant la visualisation et calibrer la fluorescence avec un filtre à 488nm. Scanner la partie du gel d'intérêt. Le scan prend quelques minutes. L'ADNc édité aura une taille supérieure à l'ADNc non édité.

(2) Méthode de détection à Evry (Licor 4300, Licor)

- *Monter le support :*

Ajouter les « espaceurs » de 0.25mm de chaque côté de la plaque et mettre la plaque à encoche sur l'autre plaque. Vérifier que rien n'a bougé et ajouter les serre-joints aux extrémités.

- *Préparation du gel :*

Pour un grand gel : 12.6g d'urée (7M) + 3ml TBE 10X + 9ml acrylamide 50% (15% final) et qsp H<sub>2</sub>O 30ml. Dissoudre le tout dans un « bain » d'eau chaude (bécher). Sous

la hotte, ajouter 180µl d'APS et 18µl de TEMED → passer **rapidement** à l'étape suivante.

- *Couler le gel :*

Avec une seringue, couler le gel entre les 2 plaques assez rapidement pour éviter que le gel ne polymérise dans la seringue (les plaques doivent être sur un plan incliné). Pendant que le gel coule, tapoter la plaque pour améliorer la coulée et éviter les bulles. Remettre les plaques à l'horizontal une fois que le gel est coulé, ajouter le peigne et attendre 1h30 la polymérisation.

- *Mise en place du gel et lancement du logiciel de visualisation :*

Enlever le peigne et nettoyer les plaques à fond (eau puis alcool technique 70%). Aucune trace ne doit rester sur plaque sous peine que l'analyse ne se fasse pas. Mettre le bac en bas. Remplir de tampon TBE 1X et ajouter le montage du gel en tirant sur les serre-joints. Ajouter le bac en haut (penser au joint d'étanchéité) et remplir de tampon TBE 1X. Nettoyer la fente où l'on dépose les échantillons. Placer les électrodes en bas et en haut et faire le branchement Licor-gel. Se connecter par le biais des adresses IP et lancer la pré-migration (20 minutes). Une fois effectué, faire monter par capillarité les échantillons (0.6µl) dans des peignes absorbants et charger le gel. Laisser le peigne 5 minutes pendant la migration et enlever. Laisser les échantillons migrer pendant 4h environ et regarder le gel (« Run » → image). L'ADNc édité aura une taille supérieure à l'ADNc non édité.

#### IV.5. Edition *in vitro* (adapté de Hayes et *al.*, 2007)

L'édition, chez les plantes, correspond principalement à une déamination de la cytidine pour former de l'uridine. Peu de choses sont encore comprises quand à l'activité de cytidine déaminase mais un protocole d'édition *in vitro* a été proposé par Hegeman et *al.*, d'abord, en 2005 puis par Hayes et *al.* en 2007 (Hegeman et *al.*, 2005; Hayes and Hanson, 2007a). Ces protocoles ont été mis au point à partir de fractions chloroplastiques, de l'ARN non édité et marqué ayant été ajouté à la fraction avant la réaction. L'efficacité de ce protocole est très fluctuante (de 0 à 85% de réussite). Dans cette thèse, ce test a été effectué à partir de complexes protéiques purifiés par TAP-tag ou expression de protéines recombinantes (et non de fraction chloroplastique).

##### IV.5.1. Transcription *in vitro*: synthèse de l'ARN à éditer

L'ADN correspondant au transcrit contenant le site d'édition désiré est cloné à l'intérieur d'un vecteur contenant le promoteur T7 (pGEMT-easy). Le vecteur est linéarisé par coupure simple avec une enzyme de restriction (attention à ce qu'elle ne coupe pas dans l'insert).

La transcription *in vitro* est effectuée par le biais du kit MEGAscript T7 (Ambion).

Réactifs de transcription <i>in vitro</i>	Dans 20µl
dATP	2µl
dUTP	2µl
dGTP	2µl
dCTP	2µl
Tampon 10X	2µl
T7 enzyme mix	2µl
Vecteur	1µg

La réaction est incubée toute la nuit à 37°C puis est traitée à la DNaseI (ou turbo<sup>TM</sup>DNase) et précipitée à l'AcNa 3M. L'ARN est ensuite testé sur gel agarose (BET) et nanodrop pour estimer la concentration.

#### IV.5.2. Réaction d'édition *in vitro*

Tampon d'édition <i>in vitro</i> dans 12.5µl
45mM HEPES-KOH
67,5mM KAc
45mM NH4Ac
5% glycérol
6mM DTT
1% PEG8000
1mM ATP
2.4U/µl RNaseOUT

Il faut ajouter au tampon d'édition 8pM d'ARN transcrit *in vitro* (1µg ARN de 1000 pb ~ 2.95pmole) et différentes quantités de protéines (la quantité de complexe actif à mettre étant inconnue).

Le mix est incubé 4h à 30°C puis 5 minutes à 65°C et enfin 5 minutes sur glace.

#### IV.5.3. Transcription « inverse » et réaction de PPE

L'ARN est ensuite rétrotranscrit (1µl par réaction) (cf. IV.2 du matériel et méthodes) et analysé par PPE, cette méthode étant beaucoup plus sensible qu'un simple séquençage (cf. IV.4.2 du matériel et méthodes).

### IV.6. Analyse du transcriptome chez *Arabidopsis thaliana* par CATMA

L'analyse du transcriptome d'*Arabidopsis* est effectuée par la plateforme de l'URGV. Les informations, données ici, résument les grandes lignes de la préparation des échantillons. Je ne suis intervenu qu'au moment de la préparation des échantillons d'ARN par extraction sur colonne (RNeasy plant minikit).

#### IV.6.1 Généralités sur la plateforme et l'utilisation d'une puce CATMA

La puce CATMA (pour "Complete Arabidopsis Transcriptome MicroArray") est disponible à travers la plateforme de l'URGV (contact : Sandrine Balzergue). Les étapes de marquage et d'hybridation sont réalisées à l'URGV. Les échantillons sont extraits

par le laboratoire collaborant avec la plateforme. Les programmes statistiques sont disponibles à l'URGV (développés par Marie-Laure Martin-Magniette).

Au début de ma thèse, la version 5 de CATMA était utilisée, celle-ci ayant été supplantée en 2011 par la version 6, plus complète.

#### IV.6.2. La puce CATMA

La puce CATMA\_v6 (disponible en 2011) contient environ 73192 sondes spécifiques de gènes (GST) issues de l'annotation EuGene du génome (gènes codant notamment pour les petits ARN, éléments transposables, ARN des organites,...). C'est une puce à haute densité commandée à la société Nimblegen (Roche, Platform GPL15719).

Les informations sur les GSTs sont disponibles sur le site de FLAGdb++

(<http://urgv.evry.inra.fr/projects/FLAGdb++/HTML/index.shtml>).

#### IV.6.3. Préparation des échantillons

Les ARN sont analysés sur puce Agilent (cf. IV.1.5). Généralement, 4µg d'ARN sont suffisants pour mener à terme l'expérience.

Il est indispensable d'effectuer deux répétitions biologiques (répétitions indépendantes, effectuées à quelques jours d'intervalles par exemple). Deux répétitions techniques par échange des marqueurs fluorescents Cyanine 3 (Cy3) et Cyanine 5 (Cy5) (communément appelé « Dye swap ») :

- Contrôle Cy3 vs mutant Cy5,
- Mutant Cy3 vs contrôle Cy5.

#### IV.6.4. Procédure expérimentale

Ce qui est présenté ici, point par point, est un résumé des étapes permettant l'analyse du transcriptome d'*Arabidopsis thaliana* par puce CATMA:

-Transcription inverse de 2µg d'ARN avec un oligodT contenant un promoteur T7 (Kit AMBION MessageAmpII aRNA Amplification #1751),

-Purification et transcription *in vitro* avec la polymérase T7 pour obtenir 80µg d'ARN amplifié (aARN),

-5µg d'aARN sont purifiés puis rétrotranscrits à partir de Cy3-dCTP (550nm/570nm) ou Cy5-dCTP (650nm/670nm),

-Les fluorochromes sont purifiés (kit Qiaquick, Qiagen, cat 28106) et dosés, 30pmoles sont utilisés par lame,

-Les lames sont préhybridées puis hybridés à 42°C sur la nuit avec du formamide,

-Les lames sont lavées au SSC et séchées,

-L'analyse se fait sur un scanner Innoscan700

L'étape de normalisation et les statistiques sont réalisées par la plateforme (statistiques développées par M-L Martin-Magniette).

Après cette étape, la liste de gène est fournie avec la valeur des log<sub>2</sub> (mesure relative d'intensité) ainsi que les ratios log<sub>2</sub> et les valeurs-p (corrections de Bonferroni ou FDR). Ceci permet de déduire un seuil statistique permettant la caractérisation des gènes différentiellement exprimés.

## V. Méthodes relatives aux protéines

### V.1. Extraction de protéines totales

#### V.1.1. De feuilles d'*Arabidopsis thaliana*

De 10 à 50mg de feuilles sont broyées dans l'azote liquide (au mortier et pilon) et 1ml d'acétone froid, 0,2% B-mercaptoéthanol sont ajoutés à la poudre. La solution est placée à -20°C pendant 1h minimum. Les tubes sont centrifugés 10 minutes à vitesse maximale, le surnageant est vidé et le culot est séché sur la pailleasse.

Le culot est enfin resuspendu dans 100-200µl de tampon de Laemmli 2X (4% SDS, 20% glycerol, 10% 2-mercaptoéthanol, 0,004% bleu de bromophénol, 0,125 M Tris HCl).

#### V.1.2. De cellules d'*E. Coli*

Centrifuger 2ml de culture à 6000g pendant 10 minutes puis resuspendre le culot dans 100µl de tampon d'extraction (DTT 0,5M, SDS 1%) et incubé 5 minutes à 94°C. Centrifuger à 6000g pendant 5 minutes et stocker à -20°C.

### V.2. Migration des protéines sur gel dénaturant

#### V.2.1. Préparation du gel et migration

Les protéines sont séparées sur gel d'acrylamide dénaturant. Les protéines migrent selon leur poids moléculaire (en kDa).

Après les avoir soigneusement nettoyées à l'eau puis à l'éthanol 70%, les plaques sont encastrées dans un portoir permettant de couler les différents gels, le gel de séparation (« separating gel ») puis le gel de concentration (« stacking gel »). Les gels utilisés dans cette thèse étaient de 1mm d'épaisseur.

<b>Gel de séparation 12%</b>	<b>Dans 15ml</b>
Acrylamide/bis-acrylamide 30%	6ml
1.5M trisHCl pH8,8	3,8ml
SDS 20%	0,075ml
APS 10% (persulfate d'ammonium)	0,15ml
TEMED (Tetramethylethylenediamine)	0,006ml
Eau	4,969ml

<b>Gel de concentration</b>	<b>Dans 4ml</b>
Acrylamide/bis-acrylamide 30%	0,67ml
1.5M trisHCl pH6,8	0,5ml
SDS 20%	0,02ml
APS 10%	0,04ml
TEMED	0,004ml
Eau	2,766ml

Le gel de séparation ne polymérise pas à l'air libre. Il faut donc rajouter une fine couche d'éthanol 100% après avoir coulé le gel entre les deux plaques (45 minutes polymérisation).

Le peigne, permettant de déposer les échantillons, est placé juste après avoir coulé le gel de concentration tout en évitant de faire des bulles (1h polymérisation).

Les plaques sont ensuite défaites de leurs supports et placées dans un portoir possédant les électrodes. Le tout est placé dans une cuve contenant du tampon de migration. Les échantillons protéiques, dans du Laemmli 4X (5µl de Laemmli pour 10µl de protéines) et chauffés à 95°C pendant 10 minutes, sont chargés dans le gel sans oublier le marqueur de taille adapté à la taille du fragment attendu. La migration a lieu à 30mA par gel pendant 1h environ. Il est important de vérifier la migration régulièrement.

Tampon de migration	Pour 1L
Tris Base	30,3g
Glycine	144g
SDS	10g

Ajuster le tampon si besoin à pH8,3

#### V.2.2. Coloration au bleu de coomassie

Le bleu de Coomassie est un colorant se fixant préférentiellement à l'arginine et aux acides aminés aromatiques (la phénylalanine, la tyrosine et le tryptophane). L'avantage de cette technique est que le gel est réutilisable une fois lavée.

La solution de teinte est composée de 2,5g de bleu de Coomassie (Thermo Scientific, 20278), 455ml d'éthanol, 455ml d'eau et enfin 90ml d'acide acétique glacial. Le gel est incubé une nuit dans cette solution et observé le lendemain sous un néon placé en dessous du gel.

La solution de lavage est composée de 10% d'éthanol et 10% d'acide acétique glacial.

#### V.2.3. Coloration au nitrate d'argent

Le nitrate d'argent permet la détection des protéines présentes dans un gel acrylamide. Cette technique est de 10 à 100 fois plus sensible que le mode de détection au bleu de Coomassie. L'inconvénient de cette technique est que le gel n'est pas réutilisable.

Le gel est fixé avec une solution d'éthanol 40%, HAc (acide acétique) 10% pendant une heure sous agitation puis est rincé dans une solution d'éthanol 30% pendant 20 minutes, deux fois. Le lavage est effectué avec de l'eau pendant 20 minutes et le gel est mis dans une solution de Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub> (Sodium thiosulfate) 0,02% pendant 1 minute 30 secondes. Le gel est lavé 3 fois 20 secondes dans l'eau puis est incubé dans une solution d'AgNO<sub>3</sub> (nitrate d'argent) 0,1% froid pendant 20 minutes à 4°C. Le gel est ensuite rincé avec de l'eau 3 fois 20 secondes en ayant changé de chambre d'incubation puis il est développé dans une solution de Na<sub>2</sub>CO<sub>3</sub> (carbonate de soude) 3%, 0,05% formaldéhyde. Lorsque

le développement est suffisant, la solution doit être rapidement enlevée pour pouvoir ensuite rincer à l'eau et la réaction est arrêtée dans une solution HAc 5%.

### V.3. Western blot

#### V.3.1. Transfert sur membrane de PVDF

Afin de rendre les protéines accessibles aux anticorps, il est nécessaire de les transférer du gel vers une membrane de PVDF (pour polyfluorure de vinylidène) (GE Healthcare, RPN303F). Sous l'application d'un courant électrique, les protéines chargées migrent du gel vers la membrane (en gardant la disposition sur le gel).

Deux types de transfert ont été effectués pendant cette thèse : le transfert classique (Evry) et le transfert semi-sec (Perth).

Le transfert classique se fait dans une cuve d'électrophorèse remplie de tampon de transfert (voir composition ci-après) et contenant un bloc froid pour éviter une surchauffe du tampon. Un barreau aimanté est utilisé pour faire circuler le tampon.

Le montage se fait dans l'ordre suivant :

- Ouvrir le portoir face noir vers le bas et ajouter une « éponge » imbibée dans du tampon de transfert,
- Ajouter trois papiers Whatman (coupés un peu plus large que le gel) (GE Healthcare, TE46) imbibé dans du tampon de transfert,
- Poser délicatement le gel mis dans du tampon de transfert face vers le plafond,
- Poser très délicatement la membrane de PVDF (coupée à la dimension du gel) préalablement imbibée quelques secondes dans du méthanol puis dans du tampon de transfert,
- Ajouter trois papiers Whatman (coupés un peu plus large que le gel) imbibé dans du tampon de transfert,
- Ajouter une « éponge » imbibée dans du tampon de transfert et refermer le portoir avec la face rouge.

Le montage est ensuite placé dans la cuve. Le transfert se fait à 100V constant pendant une heure. Le transfert a fonctionné si le marqueur de taille est passé du gel à la membrane.

Le transfert semi-sec (SD-trans blot semi-dry, Biorad) est utilisé avec très peu de tampon de transfert (quelques millilitres contre plus d'un litre en mode classique).

Le montage se fait dans l'ordre suivant :

- Ouvrir la machine et placer trois papiers Whatman (coupés un peu plus large que le gel) imbibé dans du tampon de transfert,
- Poser très délicatement la membrane de PVDF (coupée à la dimension du gel) préalablement imbibée quelques secondes dans du méthanol puis dans du tampon de transfert,
- Poser délicatement le gel imbibé dans du tampon de transfert face vers le plafond,
- Ajouter trois papiers Whatman (coupés un peu plus large que le gel) imbibé dans du tampon de transfert.



La machine est refermée et le transfert est effectué à 10V pendant 30 minutes. Tout comme dans le transfert classique, le transfert a fonctionné si le marqueur de taille est passé du gel à la membrane.

<b>Tampon de transfert 10X</b>	<b>Pour 1L</b>
Tris Base	30,3g
Glycine	144g

<b>Tampon de transfert 1X</b>	<b>Pour 1L</b>
Eau	700ml
Ethanol absolu	200ml
Tampon de transfert 10X	100ml

### V.3.2. Saturation de la membrane et application des anticorps I et II

Afin d'éviter une interaction entre anticorps et membrane (et ainsi éviter un signal partout sur la membrane), il est impératif de « bloquer » la membrane avec une solution concentrée en protéines comme du lait écrémé à 5% dilué dans du TBS-T 1X (tampon tris-salin tween) pendant 1 heure.

<b>TBS 10X</b>	<b>Pour 1L</b>
Tris HCl pH 7.4 500mM	500ml
NaCl 1,5M	300ml
Eau	200ml

<b>TBS-T 1X</b>	<b>Pour 1L</b>
TBS 10X	100ml
Tween	1ml
Eau	899ml

La membrane est rincée dans trois bains successifs de TBS-T 1X pendant 5 à 10 minutes puis l'anticorps primaire (produit chez la souris ou chez le lapin en général) est ajouté. L'anticorps est dilué dans du TBS-T 1X suivant les instructions du fournisseur. Les anticorps primaires principaux, utilisés en western blot, ont été l'anti-DYW1 (1/2000<sup>e</sup>), l'anti-cMyc (1/5000<sup>e</sup>) et l'anti-HA (1/4000<sup>e</sup> + 5% lait). En général, la membrane est incubée la nuit à température ambiante sous agitation mais peut être aussi laissée le week-end à 4°C sous agitation.

La membrane est ensuite rincée trois fois avec du TBS-T 1X pendant 10 minutes puis l'anticorps secondaire est ajouté. Cet anticorps doit être dirigé contre l'anticorps primaire. Dans cette thèse, les anticorps secondaires utilisés étaient couplés à la phosphatase alcaline ou à la peroxydase (ces deux systèmes de révélation seront développés dans le paragraphe suivant). L'anticorps secondaire (1/20000<sup>e</sup>) est appliqué pendant une heure à température ambiante sous agitation. Avant détection, la membrane est une nouvelle fois rincée trois fois avec du TBS-T 1X pendant 10 minutes.

Les anticorps peuvent être déshybridés dans un tampon contenant 62,5mM de tris-HCl pH6,8, 2% de SDS et 100mM de  $\beta$ -mercaptoéthanol incubé à 60°C pendant 30 minutes.

### V.3.3. Révélation du western blot

(1) Détection d'un anticorps secondaire couplé à la phosphatase alcaline :

L'anticorps secondaire couplé à la phosphatase alcaline est fixé à la protéine d'intérêt. L'ajout du substrat colorimétrique à la phosphatase alcaline va permettre une coloration de la membrane (Biorad, kit 170-6432). Cette révélation est irréversible.

Le tampon de coloration (25X AP reagent buffer) est dilué à 1X (1ml dans 25ml final pour une membrane). Avant d'effectuer la réaction, ajouter 250µl de réactif A et 250µl de réactif B (AP color reagent A/B) pour 25ml de tampon 1X. La membrane est incubée avec le substrat jusqu'à l'apparition d'une coloration violette des protéines. Cette réaction est extrêmement sensible. Il est important de rester proche au moment de l'incubation. En cas d'incubation trop longue la membrane se colorera aussi, la rendant inexploitable. La réaction est arrêtée en enlevant le tampon et en rinçant plusieurs fois la membrane dans de l'eau.

(2) Détection d'un anticorps secondaire couplé à la peroxydase :

L'anticorps secondaire couplé à la peroxydase est fixé à la protéine d'intérêt. La réaction se fait par chimioluminescence c'est-à-dire par formation d'une molécule excitée émettant à une longueur d'onde donnée. La détection se fait grâce à une caméra CCD (pour « charge coupled device ») permettant la conversion d'un signal lumineux en un signal électrique.

Le kit utilisé dans ce protocole permet la détection de l'anticorps secondaire par un mélange de 1ml de solution A et 1ml de solution B (GE Healthcare, ECL plus RPN2132).

### V.4. Dosage des protéines au réactif de Bradford (Sigma, B6916)

La méthode de Bradford est un dosage colorimétrique. Lorsqu'il est ajouté à un mélange protéique, la solution se colore en bleu. Plus la solution est bleue plus la solution est concentrée en protéine. La mesure se fait à 595nm (changement d'absorbance).

Une gamme de BSA est préparée dans les conditions suivantes :

<b>BSA à 2mg/ml</b>	<b>H2O</b>	<b>Réactif</b>	<b>Concentration finale</b>
0µl (0µg)	5µl	95µl	0µg/ml
0.5µl (1µg)	4.5µl	95µl	10µg/ml
1µl (2µg)	4µl	95µl	20µg/ml
2µl (4µg)	3µl	95µl	40µg/ml
3µl (6µg)	2µl	95µl	60µg/ml
4µl (8µg)	1µl	95µl	80µg/ml
5µl (10µg)	0µl	95µl	100µg/ml

Les échantillons à tester sont préparés afin de couvrir différents panels de concentration (5µl, 2,5µl et 1µl dans 95µl de réactif de Bradford, qsp 100µl H<sub>2</sub>O). Les échantillons sont agités 1 minute à température ambiante puis laissés à température ambiante 5 à 45 minutes suivant la coloration. La gamme est testée sur NanoDrop™ ND-1000 spectrophotometer. La concentration des échantillons est calculée grâce à la gamme et l'absorbance. Suivant la dilution selon laquelle on obtient une valeur de concentration, il faut recalculer la concentration initiale des échantillons.

### V.5. Clivage au TEV

Le TEV est une protéase initialement trouvée chez le Virus de la gravure du tabac (TEV en anglais). Il reconnaît une séquence protéique Glu-Asn-Leu-Tyr-Phe-Gln-(Gly/Ser) (ENLYFQ(G/S)). Le premier TEV utilisé (40kDa) a été produit en bactérie et contient en N-terminal une région polyhistidine permettant la purification de l'échantillon traité (Invitrogen, 12575-015). Le second TEV utilisé (81kDa) contient en N-terminal un marqueur Halo (Promega™, G6601).

Protéine purifiée	20µg
DTT 0,1M	1,5µl
Tampon 20X	7,5µl
TEV (10u/µl)	1µl
H <sub>2</sub> O	Qsp 150µl

Les échantillons sont incubés à 30°C pendant 4h. Des aliquots sont prélevés à 1h, 2h et 4h d'incubation pour être testés sur gel dénaturant. Sur celui-ci, il devrait apparaître un signal correspondant à la protéine d'intérêt et un autre correspondant à ce que l'on veut éliminer (marqueurs et protéines non clivées).

### V.6. TAP-tag (effectué dans le laboratoire de Geert de Jaeger, Gand, Belgique)

Cette purification a été effectuée dans une plateforme dédiée au TAP-tag, au département de biotechnologie et de génétique végétale à Gand en Belgique.

La méthodologie de cette manipulation est tirée d'un article paru dans « Plant Transcription Factors, Methods in Molecular Biology » (Van Leene et al., 2011).

### V.7. « Split-YFP »

Des PCR amplifiant les gènes d'intérêt ont été purifiées avec du PEG à 30% (cf II.4.2 du matériel et méthodes). 2µl de produit PCR purifié ont été clonés dans le vecteur pDONR207 par BP (cf II.4.3 du matériel et méthodes).

Après clonage dans un vecteur d'entrée, les inserts ont été introduits dans pBIFC1 et pBIFC4 par recombinaison LR. Les plasmides pBIFC1 et pBIFC4 ont été obtenus à

l'INRA de Versailles. Ces plasmides contiennent un promoteur 35S, un site de recombinaison Gateway LB/RB pour le clonage LR, ainsi que la moitié de la « YFP » fusionnée en C-terminal. La sélection des bactéries transformées a été faite sur du milieu contenant de la spectinomycine (50mg/L final). 5µl de vecteur ont été ajoutés à 200µl de C58C1 (pCH32) pour transformation.

Dans des plaques à six puits, 4ml de milieu solide Arabidopsis Duchefa ont été ajoutés (Saatec, PA500/38). Des graines de l'écotype Landsberg, stérilisées et maintenues dans l'eau pendant 2 à 3 jours, ont été semées sur le milieu recouvert d'une grille et gardées dans la chambre de culture Arabidopsis.

Quatre jours après, une préculture de C58C1-pBiFC est initiée et 1ml de cette préculture est utilisé pour inoculer 30ml de milieu LB agro (spectinomycine/rifampycine). Les cultures sont incubées jusqu'à ce que la DO atteigne une valeur de 2. Les cultures sont centrifugées 15min à 5000rpm et resuspendues dans un milieu liquide pour Arabidopsis contenant de l'acétylsyringone afin d'obtenir une DO de 2.

Les plantules sont alors transformées avec un mélange d'agrobactéries (un tiers de culture contenant pBIFC1, un tiers avec pBIFC4 et un tiers avec une culture contenant le plasmide p19, inhibant le « silencing » des ARN). La transformation est effectuée à l'aide d'une pompe à vide (deux fois 1 minute). La solution est éliminée et les plantules sont placées dans la chambre de culture. Trois jours après transformation, les cotylédons sont observés grâce à un microscope confocal (LEICA SP2 AOBS diode 405 avec deux filtres: n°GFP, filtre d'excitation BP 450-490, filtre d'arrêt BP 500-550; n°I3, filtre d'excitation BP 450-490, filtre d'arrêt BP 515).

Les observations ont été faites à l'INRA de Versailles par Alexandra Avon et Richard Berthomé.

## V.8. Purification de protéines sur colonne cobalt/nickel

Ce protocole a été utilisé pour purifier des protéines recombinantes, marquées à l'histidine et produites en *E. Coli*. L'histidine se lie spécifiquement aux billes d'ions cobalt ( $\text{Co}^{2+}$ ) ou nickel ( $\text{Ni}^{2+}$ ) ce qui permet la rétention des protéines marquées. L'imidazole, un compétiteur de l'histidine, est utilisé pour éluer les protéines de façon native.

Les bactéries induites sont resuspendues dans un tampon de lyse (30mM Tris, pH7.5, 500mM NaCl, 0.1% Triton X-100 ; 3ml de tampon pour 50ml de culture). Il faut ensuite effectuer 6 sonications de 12 secondes avec 10 secondes d'intervalle à 4°C afin de lyser les bactéries en conservant les protéines sous forme native. Les échantillons sont centrifugés à 4°C pendant 20 minutes et à 10000g.

Une autre solution pour lyser les cellules bactériennes est d'utiliser un homogénéiseur de bactérie à pression. Ceci m'a été conseillé par Charlie Bond à Perth qui a souligné le fait que dans cette condition, la température de la solution ne changeait pas (lyse effectuée conjointement avec Rohan Bythell-Douglas). En effet, en utilisant un

sonicateur, même avec de la glace, la température au contact de l'outil augmente considérablement. Cette technique a été utilisée à Perth pour lyser les cellules en vue d'effectuer une ICP-MS sur les protéines purifiées.

Pendant ce temps, 250µl de billes de cobalt centrifugées 2 minutes à 800g, le surnageant est éliminé et les billes sont équilibrées avec 250µl de tampon de lyse. Le surnageant est à nouveau éliminé après centrifugation.

Le surnageant issue de la lyse des cellules est incubé avec les billes à 4°C en rotation constante pendant 30 minutes à 1h (20µl de surnageant sont prélevés pour analyse ainsi qu'un peu de culot).

Les billes sont ensuite placées dans des colonnes de polypropylène. Le surnageant est éliminé au travers de la colonne (20µl prélevés pour analyse) puis la colonne est lavée avec 4x2ml de tampon de lavage (30mM Tris, pH7,5, 300mM NaCl, 5 à 20 mM imidazole) (20µl de chaque étape de lavage sont récupérés pour analyse). Six éluions (125µl chacune) sont effectuées dans un tampon concentré en imidazole (Tp phosphate pH6,8, 250mM KCl, 150mM imidazole).

### V.9. Co-immunoprécipitation (coIP)

Ce protocole a été utilisé pour purifier des protéines recombinantes, marquées à l'histidine au c-Myc ou au HA (Human influenza hemagglutinin), produites en *E. coli* ou en *Nicotiana benthamiana*. Le principe repose sur le fait que les protéines d'intérêt sont reconnues grâce à leur marqueur par un anticorps. L'anticorps se lie quand à lui à des billes de protéines A (spécifique des IgG, molécules très abondantes chez les mammifères).

Les protéines sont extraites par le biais d'un tampon d'extraction différent que l'on soit chez *E. Coli* ou *N. benthamiana*.

(1) Chez *N. benthamiana* :

Tampon d'extraction feuilles	Pour 200mL
Tris HCl 1M pH7,6	10ml
MgCl <sub>2</sub> 1M	3ml
EGTA 1M (ethylene glycol tetraacetic acid)	1ml
NaCl 5M	6ml
pNO <sub>2</sub> PhePO <sub>4</sub> (4-Nitrophenylphosphate)	1,384g
Beta-glycerophosphate	2,592g
Na <sub>3</sub> VO <sub>4</sub> (sodium orthovanadate)	3,6mg
NaF	8,4mg
NP-40 (nonyl phenoxypolyethoxyethanol)	200µl
DTT 0,1M	2ml
PMSF (extemporanément)	1,74mg pour 10ml
Inhibiteur de protéase (extemporanément)	1 gélule pour 10ml

Les feuilles sont d'abord broyées dans de l'azote liquide dans un mortier jusqu'à obtention d'une poudre fine et deux volumes de tampon d'extraction froid sont ajoutés. Le lysat est mélangé au vortex trois fois et est centrifugé 15min à 4°C à vitesse

maximale pour culotter les débris. Le surnageant est récupéré et la centrifugation est répétée deux fois.

(2) Chez *E. coli*:

Tampon de lyse <i>E. Coli</i>	Pour 100mL
Tris HCl 1M	3ml
NaCl 2,5M	20ml
PMSF (extemporanément)	1,74mg pour 10ml
Inhibiteur de protéase (extemporanément)	1 gélule pour 10ml

Une culture de 40mL est resuspendue dans le tampon de lyse. Il faut ensuite effectuer 6 sonications de 12 secondes avec 10 secondes d'intervalle à 4°C. Les échantillons sont centrifugés à 4°C pendant 20 minutes à 10000g. Le surnageant est prélevé pour effectuer la coIP.

A partir de cette étape, les protéines sont traitées de la même façon selon qu'elles aient été produites en plante (1) ou en bactérie (2).

Les protéines sont dosées avec du réactif de Bradford (cf V.4 du matériel et méthodes) afin d'ajuster la concentration des solutions de protéines à 2mg/ml (avec du tampon d'extraction/de lyse). Pour la coIP il faut travailler avec 2mg de protéines.

Lavage des billes : prélever 50µl de billes de protéines A (Invitrogen, 10-1042) et centrifuger ceux-ci à 16000g pendant 10 secondes. Le surnageant est éliminé et les billes sont resuspendues dans 1ml de tampon d'extraction. Les billes sont bien mélangées et à nouveau centrifugées à 16000g pendant 10 secondes. Le surnageant est à nouveau éliminé. Ce lavage est recommencé deux fois.

Pré-lavage (pre-cleaning) : pour savoir s'il y a des interactions non spécifiques entre les billes et les protéines, il faut mélanger les 2mg de protéines avec les billes lavées auparavant. L'extrait est incubé 1h à 4°C en rotation constante puis centrifugé à 16000g pendant 10 secondes. Le surnageant est prélevé pour la suite de la manipulation. Les billes sont conservées afin de vérifier si la protéine marquée n'est pas restée fixée à la colonne.

Incubation de l'anticorps : Le surnageant est incubé avec 5µg d'anticorps pendant 2h à 4°C en rotation constante. Ce mélange est transféré 50µl de billes de protéine A (préalablement lavées) et incubées 1h à 4°C en rotation constante.

Lavage des échantillons : Les billes sont récupérées après centrifugation à 16000g pendant 10 secondes et lavées dans 1ml de tampon d'extraction ne contenant ni PMSF ni inhibiteur de protéase. Les billes sont centrifugées à 16000g pendant 10 secondes. Cette étape de lavage est répétée une fois. Les billes sont alors resuspendues dans une solution de Laemmli 4X et chauffées à 95°C.

Les échantillons sont analysés par western blot (cf V.3 du matériel et méthodes). Il est nécessaire de tester la fraction de protéines brutes, les billes du pré-lavage, le surnageant issu du pré-lavage, les deux lavages et enfin la resuspension finale.

## V.10. Spectrométrie par torche à plasma (ICP-MS)

La spectrométrie par torche à plasma (ou ICP-MS en anglais) est une technique permettant le dosage des différents éléments présents dans l'échantillon. L'avantage principal de cette technique est sa sensibilité, sa rapidité et le fait qu'elle peut détecter tous les éléments du tableau périodique. Dans notre cas, cette technique a été utilisée pour identifier les ions métaux interagissant avec DYW1.

### V.10.1. Préparation des échantillons

Les protéines ont été exprimées dans une souche d'*E. coli* Rosetta grâce au clonage des ORFs correspondant dans un vecteur pDEST15. Ce vecteur permet d'ajouter en N-terminal de la protéine à exprimer un marqueur : GST. En plus de l'insert, il est pratique d'ajouter un site de clivage au TEV pour pouvoir par la suite éventuellement éliminer ce marqueur. Les protéines exprimées (16°C pendant 20h) sont purifiées sur colonne histidine (cf section V.8 du matériel et méthodes) et analysées sur gel d'acrylamide dénaturant. Il est important d'effectuer une centrifugation des éluats à 4°C pendant 15min afin de culotter les protéines insolubles.

Les protéines sont dialysées dans le tampon de lyse (PBS) utilisée au départ afin d'éliminer toute trace de glutathion qui aurait un effet sur le résultat final (le glutathion fixant les ions métaux). Les protéines sont dosées à 280nm ou grâce à un dosage Bradford.

Ainsi, nous obtenons en théorie la protéine d'intérêt qui a fixé (ou non) l'ion métal appliqué.

### V.10.2. ICP-MS

Pour les besoins de l'expérience, il est impératif de renseigner:

- la concentration des échantillons,
- le tampon utilisé,
- la taille en kDa des protéines à étudier.

Il est impératif d'avoir 5ml de solution pour effectuer l'expérience. Néanmoins, la technique est extrêmement sensible ce qui implique que les échantillons peuvent être fortement dilués.

Les échantillons ont été donnés à TSW Analytical Pty Ltd (University of Western Australia, Myers Street, CRAWLEY, Western Australia, 6009, contact : Sven Fjastad).

Le tampon seul a d'abord été analysé afin d'évaluer la quantité de zinc (ou autre) contaminant en solution. Les échantillons sont ensuite analysés.

Les concentrations sont données en partie par million ( $10^{-6}$ )/billion ( $10^{-9}$ )/trillion ( $10^{-12}$ ).

## VI. Méthodes relatives aux levures

### VI.1. Préparation de levures compétentes

Une colonie est inoculée dans 5ml de milieu YPD. La préculture est incubée sur la nuit à 30°C sous agitation afin d'atteindre la phase de croissance stationnaire (DO>1.5). La culture est transférée dans 50ml d'YPD afin d'obtenir une DO de 0.1. La culture est incubée à 30°C pendant 3 heures pour avoir une DO de 0.4-0.6 (phase exponentielle) et est divisée en 2 tubes de 50ml puis centrifugée à 700g pendant 5 minutes et à température ambiante.

Le surnageant est éliminé et le culot est resuspendu soigneusement dans 50ml d'eau stérile. Les tubes sont à nouveau centrifugés à 700g pendant 5 minutes, le surnageant est éliminé et le culot est resuspendu dans 100-500µl de solution 1X TE/ 1X LiAc. Les cellules sont prêtes à être transformées (elles peuvent être gardées à température ambiante pendant quelques heures).

<b>YPG/D</b>	<b>Dans 1L</b>	<b>YPG/D agar</b>	<b>Dans 1L</b>
Extrait de levures	10g	Extrait de levures	10g
Bacto-peptone	20g	Bacto-peptone	20g
-	-	Agarose	20g
Eau	Qsp900ml	Eau	Qsp900ml
Autoclaver	-	Autoclaver	-
Galactose ou D-Glucose 20%	100ml	Galactose ou D-Glucose 20%	100ml

TE 10X : 0,1M tris-HCl, 10mM EDTA, pH 7.5 (autoclave).

LiAc 10X : 1M acétate de lithium. Ajuster à pH 7.5 avec de l'acide acétique (autoclave).

### VI.2. Transformation de levures compétentes

Ajouter 100ng de plasmide et 50µg d'ADN de sperme de hareng dénaturé (fragment « intégratif ») à un tube de 2ml puis ajouter 50µl de cellules de levure préparées ci-dessus. Le mélange est homogénéisé par pipetage et 300µl de solution PEG/LiAc/TE sont ajoutés. La solution est incubée à 30°C pendant 30 minutes sous agitation.

40µl de DMSO sont ajoutés et les cellules sont mélangées doucement par inversion. Le choc thermique est effectué en incubant les cellules à 42°C pendant 15 minutes puis à 4°C pendant 2 minutes. Les cellules sont centrifugées 5 secondes à 20000g à température ambiante, le surnageant est éliminé soigneusement et le culot est resuspendu dans 500µl de TE 1X stérile.

100µl de produit de transformation sont déposés sur des boîtes contenant du milieu SC (Synthetic Complete) et les acides aminés adaptés à la sélection (dépendant du marqueur d'auxotrophie présent dans le plasmide)

Un milieu ne contenant pas de leucine, tryptophane, uracile a été préparé (US biological, D9539-05). Ces acides aminés (préparés à 100X) peuvent être rajoutés par la suite directement dans le milieu chaud.



<b>PEG/TE/LiAc</b>	<b>Dans 1ml</b>
TE 10X	100µl
LiAc 10X	100µl
PEG 3350 50%	800µl

<b>Milieu SC agar</b>	<b>Dans 1l</b>
Yeast Nitrogen Base (YNB) sans acides aminés (Sigma, Y1251)	1,7g
Sulfate d'ammonium	5g
Mélange d'acides aminés sans leucine-uracile-tryptophane	2g
Agarose	20g
Qsp eau	900ml
Autoclaver	-
D-glucose ou Galactose 20%	100ml

100X uracile (Sigma-Aldrich, U0750) : 0,2g dans 100ml (autoclaver, conservation à température ambiante).

100X tryptophane (Sigma-Aldrich, T0254) : 0,5g dans 100ml (filtrer, conservation à 4°C).

100X leucine (Sigma-Aldrich, L8000) : 0,72g dans 100ml (autoclaver, conservation à 4°C).

Le glucose et le galactose apportent le carbone nécessaire au développement des levures. Certains plasmides utilisés dans cette thèse possèdent un promoteur GAL1, activé en présence de galactose. Au contraire, en présence de glucose, ce promoteur est fortement réprimé. Le choix de la source de carbone est donc fonction de la souche et du plasmide utilisés.

### VI.3. PCR sur colonies

Au contraire de la PCR sur colonies bactériennes, dans le cas des levures il est nécessaire lyser les cellules pour effectuer la PCR sur l'ADNp.

Dans un tube de 0,5 ml, il faut ajouter 10µl de NaOH 0,02M et piquer une colonie. La solution doit être trouble. Incuber ensuite à 99°C pendant 10 minutes.

Pour effectuer la PCR, il faut utiliser 1µl de plasmide pour une réaction de 10µl.

## VII. Méthodes relatives aux plantes

### VII.1. Synchronisation de la germination des graines d'*Arabidopsis thaliana*

Afin qu'elles germent au même moment, les graines sont resuspendues dans l'eau et placées les tubes à 4°C pendant 3 jours. Les graines sont ensuite semées directement sur terre.

### VII.2. Culture *in vitro*

#### VII.2.1. Stérilisation des graines

Afin d'éliminer les organismes qui pourraient contaminer les milieux de culture (champignons, bactéries), les graines sont soumises à une stérilisation préalablement à la synchronisation. Deux techniques ont été utilisées au cours de cette thèse.

Sous une hotte à flux laminaire (« hotte végétale »), les graines à stériliser sont mises dans un tube stérile et resuspendues dans une solution d'éthanol 70% + SDS 0,05% pendant 10 minutes. La solution est ensuite éliminée et de l'éthanol 95% est ajouté. Les graines sont laissées 10 minutes toujours sous la hotte. Les temps sont à respecter très précisément.

L'éthanol est éliminé et les graines sont laissées à l'air libre pour permettre au reste d'éthanol de s'évaporer (toujours sous la hotte). Une fois sèches, les graines sont resuspendues dans de l'eau stérile et placées à 4°C pendant 3 jours (vernalisation).

L'autre méthode est une stérilisation des graines en phase gazeuse. Dans un dessiccateur sous hotte chimique, placer les tubes, contenant les graines, ouvertes. Placer ensuite sur le côté un bécher contenant 50ml d'eau de javel et 1,5ml HCl 32%. Les vapeurs de chlore vont stériliser les graines (**attention très toxique**). La stérilisation peut durer entre 3 et 16h (dépendant de la qualité des graines, viabilité des mutants,...). Le couvercle est enlevé et les graines sont placées assez rapidement sous hotte stérile (la vapeur de chlore est un gaz assez lourd). Les graines sont laissées à l'air libre pendant 15 min le temps que la chlorine se dissipe et sont ensuite prêtes à être semées.

#### VII.2.2. Culture *in vitro*

Une fois stériles, les graines sont semées à la pipette sur un milieu de culture nutritif. Il faut ajouter au milieu l'antibiotique permettant la sélection des graines transformées par *Agrobacterium tumefaciens*.

<b>½ MS + 3% sucrose</b>	<b>Dans 1L</b>
Mix Murashige et Skoog	2.2g
Sucrose	30g
MES	0,5g
KOH	pH5,7
Agar type E	8g

Les graines sont déposées à la pipette (cônes autoclavés) sur le milieu sous une « hotte végétale ». Les milieux contenant les graines sont placés dans une rotonde en atmosphère contrôlée (25°C, 45% d'hygrométrie, photopériode de 16h).

#### VII.2.3. Transfert des plantes sur terreau

Sept à quatorze jours après le début de la culture *in vitro* des graines d'*Arabidopsis thaliana*, les plantules sont transférées sur terreau et cultivées en serre sur des barquettes 4x6 à raison d'une plantule par pot. Les barquettes sont humidifiées et les plantules sont maintenues sous cloche pendant 3 jours. Les plantules poussent ensuite à 21°C le jour et 19°C la nuit, 16 heures de lumière et 65% d'hygrométrie.

#### VII.2.4. Semis de graines non stérilisées sur milieu silice, ¼ MS sans sucre

Malgré toutes les précautions prises lors de la stérilisation et du semis, le risque de contamination est toujours présent. Afin de contourner ce problème, il est possible de remplacer le milieu MS par un milieu contenant de la silice (Sigma-Aldrich, S5631). Les risques de contamination sont limités (voir nuls). La préparation des boîtes se fait sur la paillasse et les graines ne sont pas stérilisées. Cette préparation est surtout utilisée pour la sélection de plantes transformées par *Agrobacterium tumefaciens*.

La silice est déposée dans des boîtes de Pétri afin de recouvrir le fond. Une solution de ¼ MS (avec l'antibiotique approprié : dilution au centième) préparée préalablement est ajoutée pour rendre la silice pâteuse. Une fois la silice assez homogène, attendre 15 minutes. Il faut ensuite éliminer à la pipette l'excès de MS qui n'a pas été absorbé par le milieu (refaire cette étape autant de fois qu'il le faut pour qu'il n'y ait plus de liquide). Les graines sont semées directement sur le milieu et les boîtes sont placées à 4°C pendant trois jours avant de rejoindre la rotonde en atmosphère contrôlée.

A chaque fois que le milieu est sec, il est nécessaire d'ajouter à la pipette une solution de ¼ MS avec l'antibiotique approprié (tout en évitant de « noyer » les plantules). Les plantules sont repiquées après 3 semaines.

### VII.3. Croisement de plantes d'*Arabidopsis thaliana*

Les siliques et les bourgeons floraux ouverts d'une hampe florale de la plante désignée comme pied femelle sont dans un premier temps éliminés afin d'éviter tout risque de sélection d'une silique non produite par nos soins ou d'essayer de féconder un style déjà autofécondé par ses étamines. Les pétales, les sépales et les étamines du bourgeon floral sont ensuite délicatement enlevés sous loupe binoculaire pour ne laisser que le style. Toujours sous loupe binoculaire, le pollen des étamines de la plante désignée comme pied mâle est déposé sur le stigmate. Si le croisement a fonctionné, une silique est observable 3 jours après le croisement. Les siliques matures (après 2 semaines environ) sont ensuite prélevées et mises à sécher dans un tube ouvert.

## VII.4. Transformation de plantes

### VII.4.1. Expression transitoire chez *Nicotiana benthamiana*

Une préculture de 5ml puis une culture d'Agrobactérium de 7ml (réalisée à partir de 500µl de préculture) sont effectuées en milieu LB agro liquide à 28°C. Les bactéries sont culotées et resuspendues dans un milieu d'agroinfiltration (10mM MES pH5.6, 10mM MgCl<sub>2</sub>, 150µM d'acétosyringone) afin d'obtenir une densité optique (DO) à 600nm comprise entre 0,2 et 0,3 dans le cas d'une localisation subcellulaire et entre 0,8 et 1,2 pour une coexpression en vu d'une coIP. L'agroinfiltration est effectuée grâce à une seringue sans aiguille de 1ml sur la face inférieure d'une feuille de *Nicotiana benthamiana* en serre (l'espace intercellulaire se remplit).

### VII.4.2. Préparation de protoplastes

Les feuilles sont prélevées à trois temps différents (J+3, J+4, J+5) et sont lacérées pour être placées dans un milieu de digestion des parois cellulaires. La digestion est incubée 2 à 6 heures à 30°C à l'obscurité. Les protoplastes sont ensuite prélevés et observés au microscope inverse ou confocal couplé à une caméra et un ordinateur. Le microscope possède une lampe UV et différents filtres GFP (BP530/30) et RFP (BP580/80).

Milieu de digestion	Dans 1L
Maceroenzyme R10	0,4g
Driselase	1g
Cellulase « Onusuka »	2g
Murashige et Skoog Salt Mix	4,3g
MES	0,5g
Sucrose	20g
Mannitol	80g

### VII.4.3. Transformation stable d'*Arabidopsis thaliana*

Une préculture de 5ml d'*Agrobacterium tumefaciens* est préparée et est inoculée dans 200ml de LB agro contenant les antibiotiques appropriés. La culture est incubée à 28°C jusqu'à obtention d'une DO (600nm) de 0,8 (environ 9 heures). Il est possible de d'incuber la culture sur la nuit en inoculant seulement 1ml de préculture.

La culture est ensuite séparée dans un récipient de 200ml et centrifugée 30 min à 5000g. Le surnageant est éliminé et le culot est resuspendu dans 200ml de milieu de transformation. Les inflorescences des plantes à transformer sont trempées 2 minutes dans la solution.

Milieu de transformation	Dans 2L
Sucrose	100g
Silwet	400µl
Eau	Qsp 2L

## **VIII. Obtention de mutants : étude par génétique inverse**

### **VIII.1. Mutants TILLING**

Les lignées TILLING (Targeting Induced Local Lesions IN Genomes) ont été commandées au Seattle Tilling Project (STP : <http://tilling.fhcrc.org/>).

Les plantes M1 sont autofécondées et les plantes M2 obtenues sont utilisées pour la détection des mutations. L'ADN de 8 plantes est mélangé pour améliorer l'efficacité de la détection de la mutation. Une PCR est effectuée avec des primers marqués de sorte à cibler le locus d'intérêt et des hétéroduplexes sont formés par dénaturation/renaturation. L'enzyme CEL I est utilisée pour cliver les sites hétéroduplexes et les produits de digestion sont analysés par électrophorèse sur gel acrylamide dénaturant (Till et al., 2003).

### **VIII.2. Mutants d'insertion**

Les mutants d'insertion ont été obtenus au Salk Institute Genomic Analysis Laboratory (SIGnAL : <http://signal.salk.edu/>). Les commandes se font auprès de l'European Arabidopsis Stock Center (NASC : <http://arabidopsis.info/>) ou de l'Arabidopsis Biological Resource Center (ABRC : <http://abrc.osu.edu/>).

L'ADN-T d'*Agrobacterium tumefaciens* a été inséré dans le génome d'*Arabidopsis thaliana* (Col-0). Les régions flanquantes de l'insertion sont séquencées de sorte à identifier l'endroit de l'insertion.

## IX. Outils bioinformatiques utilisés

### IX.1. Alignement de séquences

-Ape (A Plasmid Editor) par Wayne Davis:

<http://biologylabs.utah.edu/jorgensen/wayned/ape/>

-Genalys par Masazumi Takahashi:

<http://software.cng.fr/docs/genalys.html>

-MEGA3 par Koichiro Tamura *et al.* :

<http://www.megasoftware.net/>

-MultAlin par Corpet *et al.* :

<http://multalin.toulouse.inra.fr/multalin/>

-ClustalW par Kyoto University Bioinformatics Center:

<http://www.genome.jp/tools/clustalw>

### IX.2. Création d'amorces pour PCR et PPE

-Primer3 Rozen *et al.*, 2000 :

<http://frodo.wi.mit.edu/primer3/>

### IX.3. Alignement de séquences sur un génome

- National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST):

[http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE\\_TYPE=BlastHome](http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome)

-Phytozome par une collaboration JGI/CIG/Université de Californie:

<http://www.phytozome.net/>

### IX.4. Recherche de motifs PPR

-TPR pred par le Max Planck Institute, prédiction de motifs PPR dans une protéine :

<http://toolkit.tuebingen.mpg.de/tpred>

-Arabidopsis PPR genome browser par l'ARCPEB-UWA, identification de PPR par leur code AGI :

<http://www.plantenergy.uwa.edu.au/cgi-bin/gbrowse/atbrowser/>

### IX.5. Enzymes de restriction

-NEB enzyme Finder, recherche d'enzyme de restriction :

<http://www.neb.com/nebecomm/EnzymeFinder.asp>

-NEB cleavage close to the end, recherche de clivage proche des extrémités 5'/3':  
[http://www.neb.com/nebecomm/tech\\_reference/restriction\\_enzymes/cleavage\\_oligonucleotides.asp](http://www.neb.com/nebecomm/tech_reference/restriction_enzymes/cleavage_oligonucleotides.asp)

## IX.6. Recherches bibliographiques

-Biblioview portail du CNRS:  
<http://biblioview.inist.fr/>

-Pubmed par l'US National Library of Medicine:  
<http://www.ncbi.nlm.nih.gov/pubmed/>

## IX.7. Bases de données

Regroupement d'information sur *Arabidopsis thaliana* :

-TAIR (The Arabidopsis Information Resource):  
<http://www.arabidopsis.org/>

Annotation des génomes d'*Arabidopsis thaliana*, *Oryza sativa*, *Vitis Vinifera* et *Populus trichocarpa* :

-Flagdb++ :  
<http://urgv.evry.inra.fr/projects/FLAGdb++/HTML/index.shtml>

Expression des gènes chez *Arabidopsis thaliana* :

-Genevestigator :  
<https://www.genevestigator.com/gv/index.jsp>

## IX.8. Prédiction de la localisation sub-cellulaire

-TargetP par Emanuelson et *al.*, 2007 :  
<http://www.cbs.dtu.dk/services/TargetP/>

-Predotar par Small et *al.*, 2004 :  
<http://urgi.versailles.inra.fr/predotar/predotar.html>

## IX.9. Prédiction de la structure secondaire et tertiaire d'une protéine

-I-TASSER Zhang lab, université du Michigan (Roy et *al.*, 2010):  
<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>

- POLYVIEW\_3D, université de Cincinnati (Porollo and Meller, 2007)  
<http://polyview.cchmc.org/polyview3d.html>

- *Jmol*: an open-source Java viewer for chemical structures in 3D

<http://www.jmol.org>

<http://jmol.sourceforge.net/>

## IX.10. Recherche d'interactions protéine/protéine chez Arabidopsis

- Arabidopsis Interactions Viewer :

[http://bar.utoronto.ca/interactions/cgi-bin/arabidopsis\\_interactions\\_viewer.cgi](http://bar.utoronto.ca/interactions/cgi-bin/arabidopsis_interactions_viewer.cgi)



## Références bibliographiques

- Adams, K.L., and Palmer, J.D.** (2003). Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Molecular phylogenetics and evolution* **29**, 380-395.
- Andres, C., Lurin, C., and Small, I.D.** (2007). The multifarious roles of PPR proteins in plant mitochondrial gene expression. *Physiologia Plantarum* **129**, 14-22.
- Araya, A., Domec, C., Begu, D., and Litvak, S.** (1992). An in vitro system for the editing of ATP synthase subunit 9 mRNA using wheat mitochondrial extracts. *Proc Natl Acad Sci U S A* **89**, 1040-1044.
- Asakawa, Y.** (2004). Chemosystematics of the hepaticae. *Phytochemistry* **65**, 623-669.
- 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.
- Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y.S., Bond, C.S., and Small, I.** (2012). A combinatorial amino Acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS genetics* **8**, e1002910.
- Barr, C.M., Neiman, M., and Taylor, D.R.** (2005). Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytol* **168**, 39-50.
- Barrell, B.G., Bankier, A.T., and Drouin, J.** (1979). A different genetic code in human mitochondria. *Nature* **282**, 189-194.
- Bass, B.L., and Weintraub, H.** (1988). An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell* **55**, 1089-1098.
- Begu, D., Castandet, B., and Araya, A.** (2011). RNA editing restores critical domains of a group I intron in fern mitochondria. *Curr Genet* **57**, 317-325.
- 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., Elliott, L.E., and Hanson, M.R.** (2008). Genetic architecture of mitochondrial editing in *Arabidopsis thaliana*. *Genetics* **178**, 1693-1708.
- Bentolila, S., Knight, W., and Hanson, M.** (2010). Natural variation in *Arabidopsis* leads to the identification of REME1, a pentatricopeptide repeat-DYW protein controlling the editing of mitochondrial transcripts. *Plant Physiol* **154**, 1966-1982.
- Bentolila, S., Heller, W.P., Sun, T., Babina, A.M., Friso, G., van Wijk, K.J., and Hanson, M.R.** (2012). RIP1, a member of an *Arabidopsis* protein family, interacts with the protein RARE1 and broadly affects RNA editing. *Proc Natl Acad Sci U S A* **109**, E1453-1461.
- 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.
- Blanc, V., Litvak, S., and Araya, A.** (1995). RNA editing in wheat mitochondria proceeds by a deamination mechanism. *FEBS Lett* **373**, 56-60.
- Blatch, G.L., and Lassle, M.** (1999). The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays* **21**, 932-939.
- Blum, B., Bakalara, N., and Simpson, L.** (1990). A model for RNA editing in kinetoplastid mitochondria: "guide" RNA molecules transcribed from maxicircle DNA provide the edited information. *Cell* **60**, 189-198.

- Boch, J., and Bonas, U.** (2010). Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annual review of phytopathology* **48**, 419-436.
- 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.
- Boussardon, C., Salone, V., Avon, A., Berthome, R., Hammani, K., Okuda, K., Shikanai, T., Small, I., and Lurin, C.** (2012). Two Interacting Proteins Are Necessary for the Editing of the NdhD-1 Site in Arabidopsis Plastids. *Plant Cell*.
- Brennicke, A., Marchfelder, A., and Binder, S.** (1999). RNA editing. *FEMS Microbiol Rev* **23**, 297-316.
- Burger, G., Yan, Y., Javadi, P., and Lang, B.F.** (2009). Group I-intron trans-splicing and mRNA editing in the mitochondria of placozoan animals. *Trends Genet* **25**, 381-386.
- Busi, M.V., Gomez-Casati, D.F., Perales, M., Araya, A., and Zabaleta, E.** (2006). Nuclear-encoded mitochondrial complex I gene expression is restored to normal levels by inhibition of unedited ATP9 transgene expression in Arabidopsis thaliana. *Plant physiology and biochemistry : PPB / Societe francaise de physiologie vegetale* **44**, 1-6.
- Cai, W., Ji, D., Peng, L., Guo, J., Ma, J., Zou, M., Lu, C., and Zhang, L.** (2009). LPA66 is required for editing psbF chloroplast transcripts in Arabidopsis. *Plant Physiol* **150**, 1260-1271.
- 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.
- Carter, C.W., Jr.** (1995). The nucleoside deaminases for cytidine and adenosine: structure, transition state stabilization, mechanism, and evolution. *Biochimie* **77**, 92-98.
- Castandet, B., and Araya, A.** (2011). RNA editing in plant organelles. Why make it easy? *Biochemistry. Biokhimiia* **76**, 924-931.
- Castelli, V., Aury, J.M., Jaillon, O., Wincker, P., Clepet, C., Menard, M., Cruaud, C., Quetier, F., Scarpelli, C., Schachter, V., Temple, G., Caboche, M., Weissenbach, J., and Salanoubat, M.** (2004). Whole genome sequence comparisons and "full-length" cDNA sequences: a combined approach to evaluate and improve Arabidopsis genome annotation. *Genome Res* **14**, 406-413.
- Cattaneo, R.** (1991). Different types of messenger RNA editing. *Annu Rev Genet* **25**, 71-88.
- Cattaneo, R., Kaelin, K., Baczko, K., and Billeter, M.A.** (1989). Measles virus editing provides an additional cysteine-rich protein. *Cell* **56**, 759-764.
- Cazalet, C., Gomez-Valero, L., Rusniok, C., Lomma, M., Dervins-Ravault, D., Newton, H.J., Sansom, F.M., Jarraud, S., Zidane, N., Ma, L., Bouchier, C., Etienne, J., Hartland, E.L., and Buchrieser, C.** (2010). Analysis of the Legionella longbeachae genome and transcriptome uncovers unique strategies to cause Legionnaires' disease. *PLoS genetics* **6**, e1000851.
- Chateigner-Boutin, A.L., and Hanson, M.R.** (2003). Developmental co-variation of RNA editing extent of plastid editing sites exhibiting similar cis-elements. *Nucleic Acids Res* **31**, 2586-2594.
- 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.

- Chateigner-Boutin, A.L., and Small, I.** (2010). Plant RNA editing. *RNA Biol* **7**, 213-219.
- Chateigner-Boutin, A.L., and Small, I.** (2011). Organellar RNA editing. Wiley interdisciplinary reviews. *RNA* **2**, 493-506.
- Chateigner-Boutin, A.L., des Francs-Small, C.C., Delannoy, E., Kahlau, S., Tanz, S.K., de Longevialle, A.F., Fujii, S., and Small, I.** (2011). OTP70 is a pentatricopeptide repeat protein of the E subgroup involved in splicing of the plastid transcript rpoC1. *Plant J* **65**, 532-542.
- Chateigner-Boutin, A.L., Ramos-Vega, M., Guevara-Garcia, A., Andres, C., de la Luz Gutierrez-Nava, M., Cantero, A., Delannoy, E., Jimenez, L.F., Lurin, C., Small, I., and Leon, P.** (2008). CLB19, a pentatricopeptide repeat protein required for editing of rpoA and clpP chloroplast transcripts. *Plant J* **56**, 590-602.
- 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.
- Chaw, S.M., Shih, A.C., Wang, D., Wu, Y.W., Liu, S.M., and Chou, T.Y.** (2008). The mitochondrial genome of the gymnosperm *Cycas taitungensis* contains a novel family of short interspersed elements, Bpu sequences, and abundant RNA editing sites. *Mol Biol Evol* **25**, 603-615.
- 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.
- Cheong, C.G., and Hall, T.M.** (2006). Engineering RNA sequence specificity of Pumilio repeats. *Proc Natl Acad Sci U S A* **103**, 13635-13639.
- Citovsky, V., Lee, L.Y., Vyas, S., Glick, E., Chen, M.H., Vainstein, A., Gafni, Y., Gelvin, S.B., and Tzfira, T.** (2006). Subcellular localization of interacting proteins by bimolecular fluorescence complementation in planta. *J Mol Biol* **362**, 1120-1131.
- Covello, P.S., and Gray, M.W.** (1989). RNA editing in plant mitochondria. *Nature* **341**, 662-666.
- Covello, P.S., and Gray, M.W.** (1990). Differences in editing at homologous sites in messenger RNAs from angiosperm mitochondria. *Nucleic Acids Res* **18**, 5189-5196.
- Dance, G.S., Sowden, M.P., Cartegni, L., Cooper, E., Krainer, A.R., and Smith, H.C.** (2002). Two proteins essential for apolipoprotein B mRNA editing are expressed from a single gene through alternative splicing. *J Biol Chem* **277**, 12703-12709.
- 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.
- Doi, T., Kato, L., Ito, S., Shinkura, R., Wei, M., Nagaoka, H., Wang, J., and Honjo, T.** (2009). The C-terminal region of activation-induced cytidine deaminase is responsible for a recombination function other than DNA cleavage in class switch recombination. *Proc Natl Acad Sci U S A* **106**, 2758-2763.

- Doniwa, Y., Ueda, M., Ueta, M., Wada, A., Kadowaki, K., and Tsutsumi, N.** (2010). The involvement of a PPR protein of the P subfamily in partial RNA editing of an *Arabidopsis* mitochondrial transcript. *Gene* **454**, 39-46.
- Duff, R.J.** (2006). Divergent RNA editing frequencies in hornwort mitochondrial nad5 sequences. *Gene* **366**, 285-291.
- Dyall, S.D., Brown, M.T., and Johnson, P.J.** (2004). Ancient invasions: from endosymbionts to organelles. *Science* **304**, 253-257.
- Ebhardt, H.A., Tsang, H.H., Dai, D.C., Liu, Y., Bostan, B., and Fahlman, R.P.** (2009). Meta-analysis of small RNA-sequencing errors reveals ubiquitous post-transcriptional RNA modifications. *Nucleic Acids Res* **37**, 2461-2470.
- Edwards, T.A., Pyle, S.E., Wharton, R.P., and Aggarwal, A.K.** (2001). Structure of Pumilio reveals similarity between RNA and peptide binding motifs. *Cell* **105**, 281-289.
- Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H.** (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* **2**, 953-971.
- 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.
- Feagin, J.E., Abraham, J.M., and Stuart, K.** (1988). Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. *Cell* **53**, 413-422.
- Filipovska, A., and Rackham, O.** (2012). Modular recognition of nucleic acids by PUF, TALE and PPR proteins. *Molecular bioSystems* **8**, 699-708.
- 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.
- Fox, T.D., and Leaver, C.J.** (1981). The *Zea mays* mitochondrial gene coding cytochrome oxidase subunit II has an intervening sequence and does not contain TGA codons. *Cell* **26**, 315-323.
- 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 U S A* **94**, 6285-6290.
- Fujii, S., and Small, I.** (2011). The evolution of RNA editing and pentatricopeptide repeat genes. *New Phytol* **191**, 37-47.
- Fujii, S., Bond, C.S., and Small, I.D.** (2011). Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. *Proc Natl Acad Sci U S A* **108**, 1723-1728.
- Giege, P., and Brennicke, A.** (1999). RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs. *Proc Natl Acad Sci USA* **96**, 15324-15329.
- Golden, D.E., and Hajduk, S.L.** (2005). The 3'-untranslated region of cytochrome oxidase II mRNA functions in RNA editing of African trypanosomes exclusively as a cis guide RNA. *RNA* **11**, 29-37.
- Gomez-Casati, D.F., Busi, M.V., Gonzalez-Schain, N., Mouras, A., Zabaleta, E.J., and Araya, A.** (2002). A mitochondrial dysfunction induces the expression of nuclear-encoded complex I genes in engineered male sterile *Arabidopsis thaliana*. *FEBS Lett* **532**, 70-74.
- Gott, J.M., Visomirski, L.M., and Hunter, J.L.** (1993). Substitutional and insertional RNA editing of the cytochrome c oxidase subunit I mRNA of *Physarum polycephalum*. *J Biol Chem* **268**, 25483-25486.
- Gray, M.W.** (1999). Evolution of organellar genomes. *Curr Opin Genet Dev* **9**, 678-687.

- Gray, M.W., Burger, G., and Lang, B.F.** (1999). Mitochondrial evolution. *Science* **283**, 1476-1481.
- Grewe, F., Herres, S., Viehover, P., Polsakiewicz, M., Weisshaar, B., and Knoop, V.** (2011). A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte *Isoetes engelmannii*. *Nucleic Acids Res* **39**, 2890-2902.
- 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. *Molecular Biology and Evolution*.
- 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.
- Halter, C.P., Peeters, N.M., and Hanson, M.R.** (2004). RNA editing in ribosome-less plastids of iojap maize. *Curr Genet* **45**, 331-337.
- Hammani, K., Okuda, K., Tanz, S.K., Chateigner-Boutin, A.L., Shikanai, T., and Small, I.** (2009). A study of new *Arabidopsis* chloroplast RNA editing mutants reveals general features of editing factors and their target sites. *Plant Cell* **21**, 3686-3699.
- Hammani, K., Colas des Francs-Small, C., Takenaka, M., Tanz, S.K., Okuda, K., Shikanai, T., Brennicke, A., and Small, I.** (2011). The Pentatricopeptide Repeat Protein OTP87 Is Essential for RNA Editing of *nad7* and *atp1* Transcripts in *Arabidopsis* Mitochondria. *J Biol Chem.* **286**, 21361-21371.
- 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.** (2007a). Identification of a sequence motif critical for editing of a tobacco chloroplast transcript. *RNA* **13**, 281-288.
- Hayes, M.L., and Hanson, M.R.** (2007b). Assay of editing of exogenous RNAs in chloroplast extracts of *Arabidopsis*, maize, pea, and tobacco. *Methods Enzymol* **424**, 459-482.
- Hayes, M.L., and Mulligan, R.M.** (2011). Pentatricopeptide repeat proteins constrain genome evolution in chloroplasts. *Mol Biol Evol* **28**, 2029-2039.
- Hayes, M.L., Reed, M.L., Hegeman, C.E., and Hanson, M.R.** (2006). Sequence elements critical for efficient RNA editing of a tobacco chloroplast transcript in vivo and in vitro. *Nucleic Acids Res* **34**, 3742-3754.
- Heazlewood, J.L., Tonti-Filippini, J.S., Gout, A.M., Day, D.A., Whelan, J., and Millar, A.H.** (2004). Experimental analysis of the *Arabidopsis* mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins. *Plant Cell* **16**, 241-256.
- Hecht, J., Grewe, F., and Knoop, V.** (2011). Extreme RNA editing in coding islands and abundant microsatellites in repeat sequences of *Selaginella moellendorffii* mitochondria: the root of frequent plant mtDNA recombination in early tracheophytes. *Genome biology and evolution* **3**, 344-358.
- 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.
- 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.

- Hiesel, R., Wissinger, B., Schuster, W., and Brennicke, A.** (1989). RNA editing in plant mitochondria. *Science* **246**, 1632-1634.
- Hinrichsen, I., Bolle, N., Paun, L., and Kempken, F.** (2009). RNA processing in plant mitochondria is independent of transcription. *Plant Mol Biol* **70**, 663-668.
- 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.
- 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.
- Hospattankar, A.V., Higuchi, K., Law, S.W., Meglin, N., and Brewer, H.B., Jr.** (1987). Identification of a novel in-frame translational stop codon in human intestine apoB mRNA. *Biochemical and biophysical research communications* **148**, 279-285.
- Howard, M.J., Lim, W.H., Fierke, C.A., and Koutmos, M.** (2012). Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5' processing. *Proc Natl Acad Sci U S A* **109**, 16149-16154.
- Ichinose, M., Tasaki, E., Sugita, C., and Sugita, M.** (2012). A PPR-DYW protein is required for splicing of a group II intron of cox1 pre-mRNA in *Physcomitrella patens*. *Plant J* **70**, 271-278.
- Iyer, L.M., Zhang, D., Rogozin, I.B., and Aravind, L.** (2011). Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. *Nucleic Acids Res.*
- Janke, A., and Paabo, S.** (1993). Editing of a tRNA anticodon in marsupial mitochondria changes its codon recognition. *Nucleic Acids Res* **21**, 1523-1525.
- Jobson, R.W., and Qiu, Y.L.** (2008). Did RNA editing in plant organellar genomes originate under natural selection or through genetic drift? *Biol Direct* **3**, 43.
- Karcher, D., and Bock, R.** (1998). Site-selective inhibition of plastid RNA editing by heat shock and antibiotics: a role for plastid translation in RNA editing. *Nucleic Acids Res* **26**, 1185-1190.
- Kay, A.R.** (2004). Detecting and minimizing zinc contamination in physiological solutions. *BMC physiology* **4**, 4.
- Kim, S.R., Yang, J.I., Moon, S., Ryu, C.H., An, K., Kim, K.M., Yim, J., and An, G.** (2009). Rice OGR1 encodes a pentatricopeptide repeat-DYW protein and is essential for RNA editing in mitochondria. *Plant J* **59**, 738-749.
- Knoop, V.** (2011). When you can't trust the DNA: RNA editing changes transcript sequences. *Cell Mol Life Sci* **68**, 567-586.
- Kotera, E., Tasaka, M., and Shikanai, T.** (2005). A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature* **433**, 326-330.
- Kudla, J., and Bock, R.** (1999). RNA editing in an untranslated region of the Ginkgo chloroplast genome. *Gene* **234**, 81-86.
- Kugita, M., Yamamoto, Y., Fujikawa, T., Matsumoto, T., and Yoshinaga, K.** (2003). RNA editing in hornwort chloroplasts makes more than half the genes functional. *Nucleic Acids Res* **31**, 2417-2423.
- Laforest, M.J., Roewer, I., and Lang, B.F.** (1997). Mitochondrial tRNAs in the lower fungus *Spizellomyces punctatus*: tRNA editing and UAG 'stop' codons recognized as leucine. *Nucleic Acids Res* **25**, 626-632.
- Legendre, P., Forstera, B., Juttner, R., and Meier, J.C.** (2009). Glycine Receptors Caught between Genome and Proteome - Functional Implications of RNA Editing and Splicing. *Frontiers in molecular neuroscience* **2**, 23.

- 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.
- 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.
- Lipinski, K.A., Puchta, O., Surendranath, V., Kudla, M., and Golik, P.** (2011). Revisiting the yeast PPR proteins--application of an Iterative Hidden Markov Model algorithm reveals new members of the rapidly evolving family. *Mol Biol Evol* **28**, 2935-2948.
- Lonergan, K.M., and Gray, M.W.** (1993). Editing of transfer RNAs in *Acanthamoeba castellanii* mitochondria. *Science* **259**, 812-816.
- Lu, B., Wilson, R.K., Phreaner, C.G., Mulligan, R.M., and Hanson, M.R.** (1996). Protein polymorphism generated by differential RNA editing of a plant mitochondrial rps12 gene. *Mol Cell Biol* **16**, 1543-1549.
- 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.
- 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.
- Macino, G., Coruzzi, G., Nobrega, F.G., Li, M., and Tzagoloff, A.** (1979). Use of the UGA terminator as a tryptophan codon in yeast mitochondria. *Proc Natl Acad Sci U S A* **76**, 3784-3785.
- Mahendran, R., Spottswood, M.R., and Miller, D.L.** (1991). RNA editing by cytidine insertion in mitochondria of *Physarum polycephalum*. *Nature* **349**, 434-438.
- Maier, R.M., Neckermann, K., Igloi, G.L., and Kossel, H.** (1995). Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J Mol Biol* **251**, 614-628.
- Malim, M.H., and Bieniasz, P.D.** (2012). HIV Restriction Factors and Mechanisms of Evasion. *Cold Spring Harbor perspectives in medicine* **2**, a006940.
- Manavski, N., Guyon, V., Meurer, J., Wienand, U., and Brettschneider, R.** (2012). An Essential Pentatricopeptide Repeat Protein Facilitates 5' Maturation and Translation Initiation of rps3 mRNA in Maize Mitochondria. *Plant Cell* **24**, 3087-3105.
- 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.
- Maple, J., Vojta, L., Soll, J., and Moller, S.G.** (2007). ARC3 is a stromal Z-ring accessory protein essential for plastid division. *EMBO Rep* **8**, 293-299.
- Marion, J., Bach, L., Bellec, Y., Meyer, C., Gissot, L., and Faure, J.D.** (2008). Systematic analysis of protein subcellular localization and interaction using high-throughput transient transformation of Arabidopsis seedlings. *Plant J* **56**, 169-179.
- Meier, J.C., Henneberger, C., Melnick, I., Racca, C., Harvey, R.J., Heinemann, U., Schmieden, V., and Grantyn, R.** (2005). RNA editing produces glycine

- receptor alpha3(P185L), resulting in high agonist potency. *Nature neuroscience* **8**, 736-744.
- Miller, M.L., and Miller, D.L.** (2008). Non-DNA-templated addition of nucleotides to the 3' end of RNAs by the mitochondrial RNA polymerase of *Physarum polycephalum*. *Mol Cell Biol* **28**, 5795-5802.
- 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 USA* **101**, 48-52.
- 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.
- Morozova, N., Allers, J., Myers, J., and Shamoo, Y.** (2006). Protein-RNA interactions: exploring binding patterns with a three-dimensional superposition analysis of high resolution structures. *Bioinformatics* **22**, 2746-2752.
- Muller, H.J.** (1964). The Relation of Recombination to Mutational Advance. *Mutation research* **106**, 2-9.
- Mulligan, R.M., Williams, M.A., and Shanahan, M.T.** (1999). RNA editing site recognition in higher plant mitochondria. *The Journal of heredity* **90**, 338-344.
- Nakajima, Y., and Mulligan, R.M.** (2001). Heat stress results in incomplete C-to-U editing of maize chloroplast mRNAs and correlates with changes in chloroplast transcription rate. *Curr Genet* **40**, 209-213.
- Nakamura, T., and Sugita, M.** (2008). A conserved DYW domain of the pentatricopeptide repeat protein possesses a novel endoribonuclease activity. *FEBS Lett* **582**, 4163-4168.
- 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.
- 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.
- Ochsenreiter, T., and Hajduk, S.L.** (2006). Alternative editing of cytochrome c oxidase III mRNA in trypanosome mitochondria generates protein diversity. *EMBO Rep* **7**, 1128-1133.
- Ohman, M.** (2007). A-to-I editing challenger or ally to the microRNA process. *Biochimie* **89**, 1171-1176.
- Ohtani, S., Ichinose, M., Tasaki, E., Aoki, Y., Komura, Y., and Sugita, M.** (2010). Targeted gene disruption identifies three PPR-DYW proteins involved in RNA editing for five editing sites of the moss mitochondrial transcripts. *Plant Cell Physiol* **51**, 1942-1949.
- Okuda, K., Habata, Y., Kobayashi, Y., and Shikanai, T.** (2008). Amino acid sequence variations in *Nicotiana CRR4* orthologs determine the species-specific efficiency of RNA editing in plastids. *Nucleic Acids Res* **36**, 6155-6164.
- 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.



- Okuda, K., Chateigner-Boutin, A.L., Nakamura, T., Delannoy, E., Sugita, M., Myouga, F., Motohashi, R., Shinozaki, K., Small, I., and Shikanai, T.** (2009). Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in Arabidopsis chloroplasts. *Plant Cell* **21**, 146-156.
- Okuda, K., Hammani, K., Tanz, S.K., Peng, L., Fukao, Y., Myouga, F., Motohashi, R., Shinozaki, K., Small, I., and Shikanai, T.** (2010). The pentatricopeptide repeat protein OTP82 is required for RNA editing of plastid *ndhB* and *ndhG* transcripts. *Plant J* **61**, 339-349.
- Olivares, J.A., and Houk, R.S.** (1986). Suppression of analyte signal by various concomitant salts in inductively coupled plasma mass spectrometry. *Anal. Chem.* **58**, 20-25.
- 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. *Journal of virology* **80**, 4673-4682.
- Porollo, A., and Meller, J.** (2007). Versatile annotation and publication quality visualization of protein complexes using POLYVIEW-3D. *BMC bioinformatics* **8**, 316.
- 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.
- Pusnik, M., Small, I., Read, L.K., Fabbro, T., and Schneider, A.** (2007). Pentatricopeptide repeat proteins in *Trypanosoma brucei* function in mitochondrial ribosomes. *Mol Cell Biol* **27**, 6876-6888.
- Randau, L., Stanley, B.J., Kohlway, A., Mechta, S., Xiong, Y., and Soll, D.** (2009). A cytidine deaminase edits C to U in transfer RNAs in Archaea. *Science* **324**, 657-659.
- Rhee, A.C., Somerlot, B.H., Parimi, N., and Gott, J.M.** (2009). Distinct roles for sequences upstream of and downstream from *Physarum* editing sites. *RNA* **15**, 1753-1765.
- Ringel, R., Sologub, M., Morozov, Y.I., Litonin, D., Cramer, P., and Temiakov, D.** (2011). Structure of human mitochondrial RNA polymerase. *Nature* **478**, 269-273.
- Rivals, E., Bruyere, C., Toffano-Nioche, C., and Lecharny, A.** (2006). Formation of the Arabidopsis pentatricopeptide repeat family. *Plant Physiol* **141**, 825-839.
- Robbins, J.C., Heller, W.P., and Hanson, M.R.** (2009). A comparative genomics approach identifies a PPR-DYW protein that is essential for C-to-U editing of the Arabidopsis chloroplast *accD* transcript. *RNA* **15**, 1142-1153.
- Roy, A., Kucukural, A., and Zhang, Y.** (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* **5**, 725-738.
- Rudinger, M., Polsakiewicz, M., and Knoop, V.** (2008). Organellar RNA editing and plant-specific extensions of pentatricopeptide repeat proteins in jungermanniid but not in marchantiid liverworts. *Mol Biol Evol* **25**, 1405-1414.
- Rudinger, M., Fritz-Laylin, L., Polsakiewicz, M., and Knoop, V.** (2011a). Plant-type mitochondrial RNA editing in the protist *Naegleria gruberi*. *RNA*.
- Rudinger, M., Szovenyi, P., Rensing, S.A., and Knoop, V.** (2011b). Assigning DYW-type PPR proteins to RNA editing sites in the funariid mosses *Physcomitrella patens* and *Funaria hygrometrica*. *Plant J* **67**, 370-380.
- Rudinger, M., Funk, H.T., Rensing, S.A., Maier, U.G., and Knoop, V.** (2009). RNA editing: only eleven sites are present in the *Physcomitrella patens* mitochondrial transcriptome and a universal nomenclature proposal. *Mol Genet Genomics* **281**, 473-481.

- Rudinger, M., Volkmar, U., Lenz, H., Groth-Malonek, M., and Knoop, V.** (2012). Nuclear DYW-type PPR gene families diversify with increasing RNA editing frequencies in liverwort and moss mitochondria. *J Mol Evol* **74**, 37-51.
- Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arlat, M., Billault, A., Brottier, P., Camus, J.C., Cattolico, L., Chandler, M., Choisne, N., Claudel-Renard, C., Cunnac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schiex, T., Siguier, P., Thebault, P., Whalen, M., Wincker, P., Levy, M., Weissenbach, J., and Boucher, C.A.** (2002). Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature* **415**, 497-502.
- 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.
- Salter, J.D., Krucinska, J., Raina, J., Smith, H.C., and Wedekind, J.E.** (2009). A hydrodynamic analysis of APOBEC3G reveals a monomer-dimer-tetramer self-association that has implications for anti-HIV function. *Biochemistry* **48**, 10685-10687.
- 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.
- Shaw, J.M., Feagin, J.E., Stuart, K., and Simpson, L.** (1988). Editing of kinetoplastid mitochondrial mRNAs by uridine addition and deletion generates conserved amino acid sequences and AUG initiation codons. *Cell* **53**, 401-411.
- Shikanai, T.** (2006). RNA editing in plant organelles: machinery, physiological function and evolution. *Cell Mol Life Sci* **63**, 698-708.
- Shikanai, T., Endo, T., Hashimoto, T., Yamada, Y., Asada, K., and Yokota, A.** (1998). Directed disruption of the tobacco *ndhB* gene impairs cyclic electron flow around photosystem I. *Proc Natl Acad Sci U S A* **95**, 9705-9709.
- Sideraki, V., Wilson, D.K., Kurz, L.C., Quiocho, 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., and Shaw, J.** (1989). RNA editing and the mitochondrial cryptogenes of kinetoplastid protozoa. *Cell* **57**, 355-366.
- Sloan, D.B., MacQueen, A.H., Alverson, A.J., Palmer, J.D., and Taylor, D.R.** (2010). Extensive loss of RNA editing sites in rapidly evolving *Silene* mitochondrial genomes: selection vs. retroprocessing as the driving force. *Genetics* **185**, 1369-1380.
- 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, D.R.** (2009). Unparalleled GC content in the plastid DNA of *Selaginella*. *Plant Mol Biol* **71**, 627-639.
- 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.
- Sosso, D., Mbello, S., Vernoud, V., Gendrot, G., Dedieu, A., Chambrier, P., Dautzat, M., Heurtevin, L., Guyon, V., Takenaka, M., and Rogowsky, P.M.** (2012).

- PPR2263, a DYW-Subgroup Pentatricopeptide repeat protein, is required for mitochondrial nad5 and cob transcript editing, mitochondrion biogenesis, and maize growth. *Plant Cell* **24**, 676-691.
- 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.
- Steinhauser, S., Beckert, S., Capesius, I., Malek, O., and Knoop, V.** (1999). Plant mitochondrial RNA editing. *J Mol Evol* **48**, 303-312.
- Sung, T.Y., Tseng, C.C., and Hsieh, M.H.** (2010). The SLO1 PPR protein is required for RNA editing at multiple sites with similar upstream sequences in Arabidopsis mitochondria. *Plant J*.
- Takenaka, M., and Brennicke, A.** (2007). RNA editing in plant mitochondria: assays and biochemical approaches. *Methods Enzymol* **424**, 439-458.
- 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., Zehrmann, A., and Brennicke, A.** (2010). Reverse genetic screening identifies five E-class PPR-proteins involved in RNA editing in mitochondria of Arabidopsis Thaliana. *J Biol Chem*.
- 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., Zehrmann, A., Verbitskiy, D., Kugelmann, M., Härtel, B., and Brennicke, A.** (2012). Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. *Proc Natl Acad Sci U S A*.
- Tasaki, E., Hattori, M., and Sugita, M.** (2010). The moss pentatricopeptide repeat protein with a DYW domain is responsible for RNA editing of mitochondrial ccmFc transcript. *Plant J*.
- Teng, B., Burant, C.F., and Davidson, N.O.** (1993). Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science* **260**, 1816-1819.
- Thomas, S.M., Lamb, R.A., and Paterson, R.G.** (1988). Two mRNAs that differ by two nontemplated nucleotides encode the amino coterminal proteins P and V of the paramyxovirus SV5. *Cell* **54**, 891-902.
- Till, B.J., Reynolds, S.H., Greene, E.A., Codomo, C.A., Enns, L.C., Johnson, J.E., Burtner, C., Odden, A.R., Young, K., Taylor, N.E., Henikoff, J.G., Comai, L., and Henikoff, S.** (2003). Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Res* **13**, 524-530.
- Tillich, M., Lehwark, P., Morton, B.R., and Maier, U.G.** (2006). The evolution of chloroplast RNA editing. *Mol Biol Evol* **23**, 1912-1921.
- 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.
- Tillich, M., Hardel, S.L., Kupsch, C., Armbruster, U., Delannoy, E., Gualberto, J.M., Lehwark, P., Leister, D., Small, I.D., and Schmitz-Linneweber, C.** (2009). Chloroplast ribonucleoprotein CP31A is required for editing and stability of specific chloroplast mRNAs. *Proc Natl Acad Sci U S A* **106**, 6002-6007.
- Tomita, K., Ueda, T., and Watanabe, K.** (1996). RNA editing in the acceptor stem of squid mitochondrial tRNA(Tyr). *Nucleic Acids Res* **24**, 4987-4991.

- Tseng, C.C., Sung, T.Y., Li, Y.C., Hsu, S.J., Lin, C.L., and Hsieh, M.H.** (2010). Editing of *accD* and *ndhF* chloroplast transcripts is partially affected in the *Arabidopsis thaliana* *cream1* mutant. *Plant Mol Biol* **73**, 309-323.
- Tsudzuki, T., Wakasugi, T., and Sugiura, M.** (2001). Comparative analysis of RNA editing sites in higher plant chloroplasts. *J Mol Evol* **53**, 327-332.
- Uchida, M., Ohtani, S., Ichinose, M., Sugita, C., and Sugita, M.** (2011). The PPR-DYW proteins are required for RNA editing of *rps14*, *cox1* and *nad5* transcripts in *Physcomitrella patens* mitochondria. *FEBS Lett* **585**, 2367-2371.
- 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.
- Van Leene, J., Eeckhout, D., Persiau, G., Van De Slijke, E., Geerinck, J., Van Isterdael, G., Witters, E., and De Jaeger, G.** (2011). Isolation of transcription factor complexes from *Arabidopsis* cell suspension cultures by tandem affinity purification. *Methods Mol Biol* **754**, 195-218.
- Verbitskiy, D., van der Merwe, J.A., Zehrmann, A., Brennicke, A., and Takenaka, M.** (2008). Multiple specificity recognition motifs enhance plant mitochondrial RNA editing in vitro. *J Biol Chem* **283**, 24374-24381.
- Verbitskiy, D., Zehrmann, A., van der Merwe, J.A., Brennicke, A., and Takenaka, M.** (2009). The PPR protein encoded by the *LOVASTATIN INSENSITIVE 1* gene is involved in RNA editing at three sites in mitochondria of *Arabidopsis thaliana*. *Plant J*.
- Verbitskiy, D., Zehrmann, A., van der Merwe, J.A., Brennicke, A., and Takenaka, M.** (2010). The PPR protein encoded by the *LOVASTATIN INSENSITIVE 1* gene is involved in RNA editing at three sites in mitochondria of *Arabidopsis thaliana*. *Plant J* **61**, 446-455.
- Verbitskiy, D., Merwe, J.A., Zehrmann, A., Hartel, B., and Takenaka, M.** (2012a). The E-class PPR protein *MEF3* of *Arabidopsis thaliana* can also function in mitochondrial RNA editing with an additional DYW domain. *Plant Cell Physiol* **53**, 358-367.
- Verbitskiy, D., Zehrmann, A., Haertel, B., Brennicke, A., and Takenaka, M.** (2012b). Two related RNA editing proteins target the same sites in mitochondria of *Arabidopsis thaliana*. *J Biol Chem*.
- Vidal, S., Curran, J., and Kolakofsky, D.** (1990a). Editing of the Sendai virus P/C mRNA by G insertion occurs during mRNA synthesis via a virus-encoded activity. *Journal of virology* **64**, 239-246.
- Vidal, S., Curran, J., and Kolakofsky, D.** (1990b). A stuttering model for paramyxovirus P mRNA editing. *Embo J* **9**, 2017-2022.
- Visomirski-Robic, L.M., and Gott, J.M.** (1995). Accurate and efficient insertional RNA editing in isolated *Physarum* mitochondria. *RNA* **1**, 681-691.
- Visomirski-Robic, L.M., and Gott, J.M.** (1997). Insertional editing in isolated *Physarum* mitochondria is linked to RNA synthesis. *RNA* **3**, 821-837.
- Volchkov, V.E., Becker, S., Volchkova, V.A., Ternovoj, V.A., Kotov, A.N., Netesov, S.V., and Klenk, H.D.** (1995). GP mRNA of Ebola virus is edited by the Ebola virus polymerase and by T7 and vaccinia virus polymerases. *Virology* **214**, 421-430.
- Volchkov, V.E., Volchkova, V.A., Muhlberger, E., Kolesnikova, L.V., Weik, M., Dolnik, O., and Klenk, H.D.** (2001). Recovery of infectious Ebola virus from complementary DNA: RNA editing of the GP gene and viral cytotoxicity. *Science* **291**, 1965-1969.

- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B.** (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470-476.
- Wang, X., McLachlan, J., Zamore, P.D., and Hall, T.M.** (2002). Modular recognition of RNA by a human pumilio-homology domain. *Cell* **110**, 501-512.
- 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.
- Wiegand, H.L., and Cullen, B.R.** (2007). Inhibition of alpharetrovirus replication by a range of human APOBEC3 proteins. *Journal of virology* **81**, 13694-13699.
- Wolf, P.G., Rowe, C.A., and Hasebe, M.** (2004). High levels of RNA editing in a vascular plant chloroplast genome: analysis of transcripts from the fern *Adiantum capillus-veneris*. *Gene* **339**, 89-97.
- Yao, L., Sklenak, S., Yan, H., and Cukier, R.I.** (2005). A molecular dynamics exploration of the catalytic mechanism of yeast cytosine deaminase. *The journal of physical chemistry. B* **109**, 7500-7510.
- Yokobori, S., and Paabo, S.** (1995). Transfer RNA editing in land snail mitochondria. *Proc Natl Acad Sci U S A* **92**, 10432-10435.
- Yokobori, S., and Paabo, S.** (1997). Polyadenylation creates the discriminator nucleotide of chicken mitochondrial tRNA(Tyr). *J Mol Biol* **265**, 95-99.
- 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, Q.B., Jiang, Y., Chong, K., and Yang, Z.N.** (2009). AtECB2, a pentatricopeptide repeat protein, is required for chloroplast transcript *accD* RNA editing and early chloroplast biogenesis in *Arabidopsis thaliana*. *Plant J* **59**, 1011-1023.
- 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.
- Yuan, H., and Liu, D.** (2012). Functional disruption of the pentatricopeptide protein SLG1 affects mitochondrial RNA editing, plant development, and responses to abiotic stresses in *Arabidopsis*. *Plant J* **70**, 432-444.
- 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.
- Zehrmann, A., van der Merwe, J.A., Verbitskiy, D., Brennicke, A., and Takenaka, M.** (2008). Seven large variations in the extent of RNA editing in plant mitochondria between three ecotypes of *Arabidopsis thaliana*. *Mitochondrion* **8**, 319-327.
- Zehrmann, A., Verbitskiy, D., van der Merwe, J.A., Brennicke, A., and Takenaka, M.** (2009). A DYW domain-containing pentatricopeptide repeat protein is required for RNA editing at multiple sites in mitochondria of *Arabidopsis thaliana*. *Plant Cell* **21**, 558-567.
- Zehrmann, A., Verbitskiy, D., Hartel, B., Brennicke, A., and Takenaka, M.** (2010). RNA editing competence of trans-factor MEF1 is modulated by ecotype-specific differences but requires the DYW domain. *FEBS Lett* **584**, 4181-4186.
- Zehrmann, A., Verbitskiy, D., Hartel, B., Brennicke, A., and Takenaka, M.** (2011). PPR proteins network as site-specific RNA editing factors in plant organelles. *RNA Biol* **8**, 67-70.
- Zhou, W., Cheng, Y., Yap, A., Chateigner-Boutin, A.L., Delannoy, E., Hammani, K., Small, I., and Huang, J.** (2008). The *Arabidopsis* gene YS1 encoding a

DYW protein is required for editing of rpoB transcripts and the rapid development of chloroplasts during early growth. Plant J.

# Annexe

## **I. Publication**

Boussardon C, Salone V, Avon A, Berthomé R, Hammani K, Okuda K, Shikanai T, Small I, Lurin C (2012) Two Interacting Proteins are Necessary for the Editing of the NdhD-1 Site in *Arabidopsis* Plastids. **PlantCell**, 24(9):3684-94.

## Two Interacting Proteins Are Necessary for the Editing of the *NdhD-1* Site in *Arabidopsis* Plastids<sup>□□</sup>

Clément Boussardon,<sup>a,b</sup> Véronique Salome,<sup>a,b,1</sup> Alexandra Avon,<sup>a</sup> Richard Berthomé,<sup>a</sup> Kamel Hammani,<sup>b,2</sup> Kenji Okuda,<sup>c,3</sup> Toshiharu Shikanai,<sup>c</sup> Ian Small,<sup>b</sup> and Claire Lurin<sup>a,4</sup>

<sup>a</sup>Unité de Recherche en Génomique Végétale, Unité Mixte de Recherche, Institut National de la Recherche Agronomique/Université Evry Val d'Essonne/Equipe de Recherche Labellisée, Centre National de la Recherche Scientifique 91057, 91057 Evry cedex, France

<sup>b</sup>Australian Research Council Centre of Excellence in Plant Energy Biology, University of Western Australia, Crawley, Western Australia 6009, Australia

<sup>c</sup>Department of Botany, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan

After transcription, mRNA editing in angiosperm chloroplasts and mitochondria results in the conversion of cytidine to uridine by deamination. Analysis of *Arabidopsis thaliana* mutants affected in RNA editing have shown that many pentatricopeptide repeat proteins (PPRs) are required for specific cytidine deamination events. PPR proteins have been shown to be sequence-specific RNA binding proteins allowing the recognition of the C to be edited. The C-terminal DYW domain present in many editing factors has been proposed to catalyze C deamination, as it shows sequence similarities with cytidine deaminases in other organisms. However, many editing factors, such as the first to be discovered, CHLORORESPIRATORY REDUCTION4 (CRR4), lack this domain, so its importance has been unclear. Using a reverse genetic approach, we identified DYW1, an RNA editing factor acting specifically on the plastid *ndhD-1* editing site recognized by CRR4. Unlike other known editing factors, DYW1 contains no identifiable PPR motifs but does contain a clear DYW domain. We were able to show interaction between CRR4 and DYW1 by bimolecular fluorescence complementation and to reconstitute a functional chimeric CRR4-DYW1 protein complementing the *crr4 dyw1* double mutant. We propose that CRR4 and DYW1 act together to edit the *ndhD-1* site.

### INTRODUCTION

RNA editing is a sequence-specific posttranscriptional modification leading to an insertion, deletion, or conversion of one or more nucleotides in a precursor RNA. Such modifications are observed in many organisms, including mammals, plants, bacteria, and protists. The first editing event to be described was the post-transcriptional addition of four nucleotides to the mitochondrial *coxII* transcript in trypanosomes (Benne et al., 1986). In plant organelles, RNA editing occurs as a pyrimidine exchange, resulting in a conversion of a cytidine into a uridine nucleotide (C to U) in mitochondria and plastids of virtually all land plants (Chateigner-Boutin and Small, 2010; Knoop, 2011). More rarely, uridine-to-cytidine (U to C) conversions are observed in organelles of

some hornworts, lycopods, and ferns (Chateigner-Boutin and Small, 2010; Knoop, 2011). Thirty-four editing sites have been found in *Arabidopsis thaliana* chloroplasts (Chateigner-Boutin and Small, 2007), and more than 500 sites have been described in *Arabidopsis* mitochondria (Giegé and Brennicke, 1999; Bentolilla et al., 2008; Zehmann et al., 2008). RNA editing often restores conserved codons indispensable for synthesis of functional proteins (Bock et al., 1994).

Many thousands of editing events have been reported, but few editing factors have been identified at the molecular level. In mammals, the *apoB* transcript undergoes a C-to-U deamination that generates a stop codon (CAA to UAA) (Chen et al., 1987; Powell et al., 1987). The APOBEC-1 editing enzyme that catalyzes this modification contains a signature [C/HxE(x)<sub>n</sub>PCxxC] characteristic of a family of nucleotide deaminases (Teng et al., 1993; Wedekind and McKay, 2003; Iyer et al., 2011). The 11-nucleotide recognition sequence in the *apoB* transcript is named the "mooring sequence" and is localized 5 nucleotides after the edited cytidine. This sequence is highly conserved in mammals. The APOBEC1 complementation factor (ACF) was shown to specifically bind the mooring sequence (Mehta and Driscoll, 2002). The core editosome complex is constituted by interaction of the editing enzyme APOBEC-1 with the RNA/ACF complex (Blanc et al., 2001; Mehta and Driscoll, 2002).

Although first described over 20 years ago, mRNA editing in plant organelles is still not fully understood. It is observed in almost all land plant groups, but so far, not in algae (Steinhauser et al., 1999). A model similar to the mammalian model of *apoB* editing has been proposed, in which a specificity factor targeting the appropriate site in the RNA molecule and a catalytic factor

<sup>1</sup> Current address: Unité Mixte de Recherche, ARN-RNP, structure-fonction-maturation, Enzymologie Moléculaire et Structurale 7214, Centre National de la Recherche Scientifique-Université de Lorraine, Faculté des Sciences et Technologies, Nancy University, BP70239, 54506 Vandœuvre-Lès-Nancy Cedex, France.

<sup>2</sup> Current address: Institute of Molecular Biology, University of Oregon, Eugene, OR 97403.

<sup>3</sup> Current address: Department of Life Science, Faculty of Science and Engineering, Chuo University, Tokyo 112-8551, Japan.

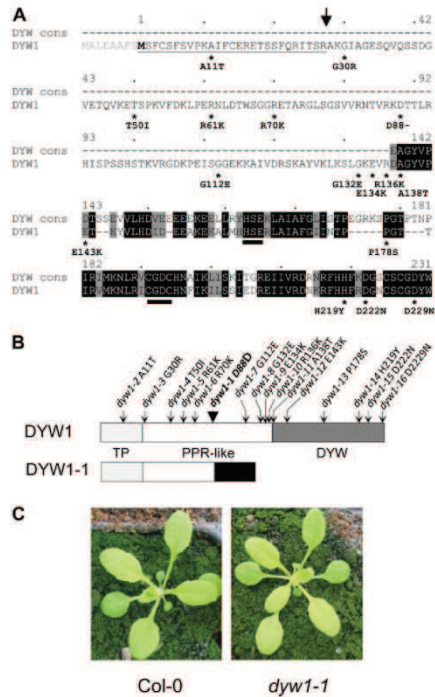
<sup>4</sup> Address correspondence to [lurin@evry.inra.fr](mailto:lurin@evry.inra.fr).

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors ([www.plantcell.org](http://www.plantcell.org)) is: Claire Lurin ([lurin@evry.inra.fr](mailto:lurin@evry.inra.fr)).

□ Some figures in this article are displayed in color online but in black and white in the print edition.

□□ Online version contains Web-only data.  
[www.plantcell.org/cgi/doi/10.1105/tpc.112.099507](http://www.plantcell.org/cgi/doi/10.1105/tpc.112.099507)





**Figure 1.** Structure of the Wild-Type DYW1 Protein and the Mutant Proteins Encoded by the Different Alleles.

**(A)** Alignment of the DYW1 sequence with the DYW domain consensus (Lurin et al., 2004). The DYW1 sequence was aligned against the DYW domain consensus (DYW cons) using the pairwise alignment software at <http://pir.georgetown.edu/pirwww/search/pairwise.shtml>. The sequence numbering is shown above the sequence according to  $M_0$ . Identical residues are shaded black, and similar residues are shaded gray. The putative DYW1 targeting peptide is underlined, and the position of a potential cleavage site indicated by an arrow. The HxE<sub>x</sub>C<sub>x</sub>C deaminase signature identified in DYW domains (Salone et al., 2007) is indicated by two black bars. Positions of the 16 substitutions identified in *DYW1* coding sequence are indicated by stars under the DYW1 protein sequence. The sequence modification resulting from each mutation is indicated at the protein sequence level (amino acid identity in wild-type sequence/position of the amino acid in the protein sequence/ amino acid identity in the mutant protein).

**(B)** Structure of the 231-amino acid protein encoded by the *DYW1* gene and positions of the EMS mutations. The 27-amino acid targeting peptide (TP) is indicated in light gray and the 95-amino acid DYW domain in dark gray. The PPR-like region in the middle is indicated in white. The PPR-like region of DYW1 is not recognized by any PPR detection software but presents a low similarity to PPR proteins when searching by BLAST. Positions of the mutations are indicated by arrows on the DYW1 protein. The sequence modification resulting from each mutation is

carrying a C-to-U deamination catalytic activity could act together (Miyamoto et al., 2002). In tobacco (*Nicotiana tabacum*), it was demonstrated that 22 nucleotides around the editing site are essential for psbL editing (Chaudhuri and Maliga, 1996) and 14 nucleotides are necessary for ndhB editing (Bock et al., 1996). The first editing specificity factor identified in plants was CHLOROPLAST BIOGENESIS19 (CLB19), necessary for the editing of two sites (Chateigner-Boutin et al., 2008) and many other proteins (Fujii and Small, 2011). In parallel, similar proteins have been found to be involved in mitochondrial RNA editing (Tasaki et al., 2010; Fujii and Small, 2011; Hammani et al., 2011b; Uchida et al., 2011).

All these proteins belong to the pentatricopeptide repeat (PPR) family (Small and Peeters, 2000). PPR proteins are involved in almost all stages of organellar gene expression, from transcription to translation (Andres et al., 2007; Schmitz-Linneweber and Small, 2008). Whereas it has been experimentally demonstrated only for very few of them, it is currently accepted that they act as sequence-specific RNA binding adaptors (Delannoy et al., 2007), and, more hypothetically, that they recruit effector enzymes to the target RNA (Okuda et al., 2006). Strikingly, all the PPR proteins required for specific editing events belong to the PLS subfamily, a particular plant-specific subfamily distinguished by arrays of characteristic variants of the canonical PPR motif and by additional C-terminal domains (Lurin et al., 2004). The PLS PPR proteins can be classified into two subgroups according to the last domain of the protein: the E/E+ subgroup and the DYW subgroup (Lurin et al., 2004; O'Toole et al., 2008). The DYW domain is generally preceded by the E/E+ domain (Lurin et al., 2004). The DYW consensus contains the cytidine deaminase signature HxE (x)nCxxC (Salone et al., 2007). In addition, the phylogenetic distribution of the DYW domain appears to exactly match the phylogenetic distribution of plant organellar RNA editing (Salone et al., 2007; Knoop, 2011; Rüdinger et al., 2011). These results supported the hypothesis that the DYW domain might contain the catalytic activity for editing in plant organelles (Salone et al.,

indicated at the protein sequence level (amino acid identity in wild-type sequence/position of the amino acid in the protein sequence/amino acid identity in the mutant protein). The structure of the putative DYW1-1 is indicated below the wild-type DYW1 protein. The additional protein sequence unrelated to wild-type DYW1 protein is indicated in black.

**(C)** Phenotype of a *dyw1-1* mutant and Col-0 wild-type plant after 18 d in soil.

[See online article for color version of this figure.]

2007). This model is still a matter for debate, and no RNA editing activity of the DYW domain has been demonstrated (Okuda et al., 2009). Moreover, not all plant editing factors contain a DYW domain. For example, CRR4, CRR21, and CLB19 belong to the PLS-E/E+ subgroup comprising 107 proteins in *Arabidopsis*, most of which are suspected to be editing factors, and all of which lack a terminal DYW domain (Kotera et al., 2005; Okuda et al., 2007; Chateigner-Boutin et al., 2008). Even more detrimentally for the hypothesis of the involvement of the DYW domain in editing, it was shown that deleting the DYW domains of the two editing factors CRR22 and CRR28 does not affect these factors' ability to restore editing when expressed in the corresponding mutants (Okuda et al., 2009).

Here, we identified a unique *Arabidopsis* PPR protein (DYW1) composed of a plastid targeting sequence and a DYW domain without any identifiable intervening PPR motifs. The *dyw1-1* knockout mutant completely lacks editing at the *ndhD-1* site (i.e., is molecularly and physiologically identical to the *crr4* mutant). We show that the DYW1 protein interacts in vivo with the CRR4 protein. Our findings suggest that DYW1, by interaction with CRR4, provides in trans an essential function required for editing of the *ndhD-1* site.

## RESULTS

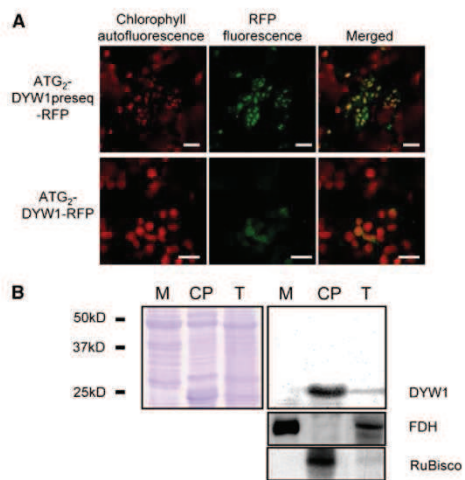
### At1g47580 Encodes a Unique DYW Protein

The At1g47580 gene (that we have named *DYW1*) attracted our attention because it encodes a protein unique in *Arabidopsis* with a clear DYW domain, characteristic of the PLS-DYW subgroup, but lacking any identifiable PPR motifs. The intronless At1g47580 locus consists of a 717-bp open reading frame (ORF) encoding a putative 239-amino acid protein. This protein is composed of a 110-amino acid region with weak similarity to PPR proteins and a 95-amino acid C-terminal DYW domain (Figure 1A). About 90 *Arabidopsis* proteins carry a DYW domain at the C terminus, but DYW1 is the only one with no identifiable PPR, E, or E+ motifs. By contrast, its DYW domain is highly conserved and is close to the DYW domain consensus (Figure 1A). This implies that this domain is under strong selection and that *DYW1* is not simply a degenerating pseudogene. Two putatively full-length GSLT (GenoScope/LifeTechnologies) cDNAs that correspond to the At1g47580 model have been described (Castelli et al., 2004). Both cDNAs start just after the ATG codon at position +24 (ATG<sub>2</sub>), indicating that the ATG codon at position +24 (ATG<sub>2</sub>) is likely to be the translation start used in the plant cell.

### The *dyw1-1* Mutant Has a Frameshift in the First Half of the DYW1 Coding Sequence

Two apparent insertion mutants in the At1g47580 locus are listed in the T-DNA Express database of the Salk Institute (Alonso et al., 2003). After sequence verification of the insertions, we found that the T-DNA in Salk\_012425 is inserted 97 bp upstream of the ATG<sub>2</sub> codon, whereas the T-DNA in Salk\_123655 is inserted ~150 bp downstream of the ORF. As neither insertion disrupts the *DYW1* ORF, we requested a screen for ethyl methanesulfonate (EMS)-induced mutations in the At1g47580 locus from the

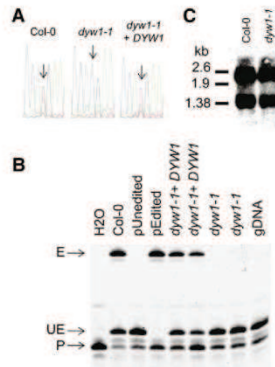
Seattle *Arabidopsis* TILLING Project (Till et al., 2003; <http://tilling.fhcrc.org/>). Forty-three different mutations across the 1500-bp locus were obtained. Among them, 19 were in noncoding regions, eight were in the ORF but did not change the protein sequence, and 16 mutations (*dyw1-1* to *dyw1-16*) did alter the predicted protein sequence (Figure 1; see Supplemental Table 1 online). One of these, the *dyw1-1* mutation, was a deletion of a guanidine at the position + 262 (after ATG<sub>2</sub>) and caused a frameshift in the ORF. The putative 118-amino acid protein encoded by this allele would be identical to the wild-type protein for the first 87 amino acids, then differ with a premature stop codon 31 amino acids after the mutation (Figure 1B). In particular, this truncated protein would completely lack the DYW domain. This mutant and the other alleles were backcrossed three times to *Arabidopsis* ecotype Columbia-0 (Col-0) to segregate other EMS-induced mutations. Homozygous mutant plants were identified after self-pollination of heterozygous plants obtained after backcrossing. All of them exhibited a macroscopic



**Figure 2.** DYW1 Is Localized in Chloroplasts of *Arabidopsis* Cells.

(A) Fluorescence images of *Arabidopsis* plantlets transiently expressing DYW1-RFP fusion proteins. Either the first 100-amino acid polypeptide or the full-length DYW1 protein, starting both at the M<sub>2</sub> Met encoded by the ATG<sub>2</sub> codon, were expressed as a fusion with the RFP at the C terminus. Plantlets were observed by confocal microscopy 4 d after transformation. Bars = 10 μm.

(B) Immunoblot analysis of total extract (T), mitochondrial (M), and chloroplast (CP) protein fractions with antibodies directed against DYW1. Protein fractions from *Arabidopsis* were analyzed by immunoblots. Loading control (Coomassie blue staining; left panel) shows that equal amounts of total protein were loaded. Black bars represent molecular mass marker positions. The purity of the two fractions was tested with antibodies directed against the mitochondrial formate dehydrogenase (FDH) and the chloroplast large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo).



**Figure 3.** The *dyw1-1* Mutant Lacks Editing of the *ndhD-1* Site.

(A) Nucleotide sequences of RT-PCR products obtained from Col-0, *dyw1-1*, and complemented *dyw1-1* cDNA are shown as sequencing chromatograms. An arrow pointing to the corresponding peak indicates the *ndhD-1* editing site.

(B) PPE assays were conducted on the *ndhD-1* (117166) editing site. RT-PCR products were obtained from Col-0, *dyw1-1*, and complemented *dyw1-1* cDNA with primers surrounding the editing sites and these served as templates for the extension reaction from a 5'-labeled 6-carboxy-fluorescein primer (P) that anneals next to the target editing site. The extension was stopped by the incorporation of 2',3'-dideoxycytidine-5'-triphosphate at the location of the editing site for unedited molecules, producing a short unedited product (UE). The extension was stopped at the next C/G for the edited molecules, producing a longer edited product (E). PPE samples obtained on cloned edited (pEdited) and unedited (pUnedited) *ndhD* cDNA fragments were loaded as controls. gDNA, genomic DNA.

(C) RNA gel blot analysis of leaves of *Arabidopsis* Col-0 and *dyw1-1* mutant using an *ndhD* probe. Polycistronic (~2.6 kb) and monocistronic (~1.38 kb) forms of the *ndhD* transcript were identified.

phenotype identical to wild-type Col-0 plants, as shown for homozygous *dyw1-1* mutants (Figure 1C).

#### DYW1 is Targeted to Plastids

Using either ATG<sub>1</sub> or ATG<sub>2</sub> (encoding M<sub>1</sub> and M<sub>2</sub>, respectively) as the start of the protein coding sequence, both TargetP (<http://www.cbs.dtu.dk/services/TargetP/>; Emanuelsson et al., 2007) and Predotar (<http://urgi.versailles.inra.fr/predotar/predotar.html>; Small et al., 2004) predict that the DYW1 protein is targeted to chloroplasts. The cleavage site is predicted to be after the 27th amino acid (counting from M<sub>2</sub>) leading to a protein of 22.8 kD after cleavage. To test this localization prediction, two fusion proteins were transiently expressed in *Arabidopsis* plantlets. The sequences encoding the first 108 or 100 amino acids of the protein (starting from M<sub>1</sub> and M<sub>2</sub>, respectively) were fused in frame upstream of the red fluorescent protein (RFP) coding sequence. Expression of the M<sub>1</sub> fusion protein did not lead to detectable fluorescence. By contrast, expression of the M<sub>2</sub> fusion protein led to RFP-specific fluorescence signals in plastids (Figure 2A). This observation was further confirmed by expression of the full-length

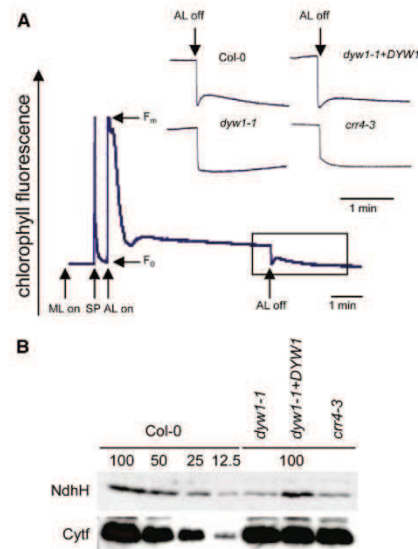
M<sub>2</sub> protein fused to RFP (Figure 2A). The emission spectrum of RFP was confirmed for these signals and they were shown to colocalize with chlorophyll autofluorescence (Figure 2A).

A polyclonal antibody directed against DYW1 was used to detect the protein in subcellular fractions. Figure 2B shows that an ~25-kD signal was observed in a total extract of leaf proteins of *Arabidopsis* and strongly increased in a subcellular fraction enriched in chloroplasts. By contrast, this signal was not observed in a mitochondrial protein extract.

Both the transient expression of RFP fusions and the analysis of subcellular fractions by immunoblot using a DYW1-specific antibody confirmed the bioinformatics predictions that the DYW1 protein is targeted to plastids.

#### *dyw1-1* Is Not Able to Edit the *ndhD-1* Site

Most DYW proteins investigated to date have been shown to be involved in RNA editing (Fujii and Small, 2011; Hammani et al., 2011b; Uchida et al., 2011), so we screened all 34 known



**Figure 4.** The *dyw1-1* Mutant Is Impaired in NDH Activity.

(A) Monitoring NDH activity using chlorophyll fluorescence analysis after turning off AL. The bottom curve indicates a typical trace of chlorophyll fluorescence in the wild-type Col-0. The transient rise in fluorescence ascribed to NDH activity was monitored by chlorophyll fluorimetry. Insets are magnified traces from the boxed area. F<sub>0</sub>, minimum fluorescent yield; F<sub>m</sub>, maximum fluorescent yield; ML, measuring light; SP, saturating pulse. (B) Immunoblot analysis of thylakoid proteins. Immunodetection of NDH (NdhH) and cytochrome b<sub>6</sub>f (Cyt f) complexes. The lanes were loaded with a series of dilutions as indicated.

[See online article for color version of this figure.]

chloroplast editing sites (Chateigner-Boutin and Small, 2007) in *dyw1-1* by sequencing of RT-PCR products surrounding these sites. Among the 34 sites, a very strong defect was observed at the first editing site of the *ndhD* transcript (Figure 3A). None of the other sites were affected in the mutant (see Supplemental Figure 1 online). A complete lack of editing of *ndhD-1* was confirmed by a quantitative and very sensitive poisoned primer extension (PPE) assay (Figure 3B). The editing of the *ndhD-1* site was also characterized by RT-PCR product sequencing in the other 15 *dyw1-x* alleles. A defect in three lines (*dyw1-11*, *dyw1-12*, and *dyw1-16*) was observed with a reduction of 25 to 40% when compared with Col-0 (see Supplemental Figure 2 online).

To confirm that the editing defect observed in *dyw1-1* was a consequence of the frameshift in *DYW1*, the mutant was complemented with a 1502-bp genomic fragment comprising a region of 357 bp before ATG<sub>0</sub>, the *DYW1* ORF (717 bp) and 428 bp downstream. After *Agrobacterium tumefaciens* transformation of *dyw1-1* with this construct, two complemented lines were shown to be restored in the editing of *ndhD-1* by the PPE assay (Figures 3A and 3B). This complementation experiment and the defects observed in weaker alleles demonstrated that the abolition of *ndhD-1* editing, shown by sequencing and PPE, is due to the frameshift in *DYW1*.

To examine whether the loss of editing was a secondary effect of altered RNA processing or RNA stability, the *ndhD* transcript was analyzed by RNA gel blot analysis of the *dyw1-1* mutant compared with the wild type. No variation in the pattern or level of the monocistronic (mature *ndhD* RNA) or polycistronic (precursor form of *ndhD* RNA) transcripts was observed (Figure 3C). This result supports a specific and direct role of *DYW1* in the editing of the *ndhD-1* site.

#### ***dyw1-1* Is Impaired in NDH Activity and Phenocopies the *crr4* Editing Mutant**

The *ndhD-1* site is partially edited in the wild-type plant (Figures 3A and 3B, Col-0), and editing at this site converts a Thr codon (ACG) into a Met codon (AUG) that is the *ndhD* putative translation initiation codon. The chloroplast NDH complex catalyzes electron donation probably from ferredoxin to plastoquinone (Yamamoto et al., 2011). When NDH activity is present, a transient increase of chlorophyll fluorescence can be observed when actinic light (AL) is switched off (Shikanai et al., 1998). We analyzed this change of fluorescence to examine whether or not the NDH complex was affected in *dyw1-1*. Unlike the Col-0 control, the mutant showed no transient increase in fluorescence after the AL was turned off

(Figure 4A). This defect was reverted in the complemented transgenic plants (Figure 4A). Immunoblot analysis using an antibody against NdhH showed a decrease in the level of the NDH complex in *dyw1-1* (Figure 4B). The editing site impaired in the *dyw1-1* mutant is exactly the same site that has been shown to be unedited in *crr4* mutants (Kotera et al., 2005). To perform a fine comparison of the *crr4* and *dyw1-1* mutants, *crr4-3* (encoding a truncated CRR4 protein) and *dyw1-1* mutants were grown and characterized in parallel. The primary molecular defect (absence of editing of the *ndhD-1* site), the resulting lack of NDH activity (Figure 4A), the decreases in the level of NDH complex (Figure 4B), and the macroscopic phenotypes of adult plants compared with the wild type (Figure 1C; Kotera et al., 2005) were identical in *dyw1-1* and *crr4-3*.

#### **DYW1 Interacts with CRR4**

The *dyw1-1* and *crr4* mutants are indistinguishable by phenotype, indicating that the *DYW1* and *CRR4* proteins have essential roles in the same process. However, the common phenotype also indicates that *CRR4* is not able to complement the *DYW1* function in the *dyw1* background, and, equally, *DYW1* is not able to complement the *CRR4* function in the *crr4* background. This observation could suggest that the *DYW1* protein interacts with the *CRR4* protein to form a protein complex that functions in *ndhD-1* editing. To test this hypothesis, we used bimolecular fluorescence complementation (BiFC) assays (Marion et al., 2008) to visualize any interaction between *CRR4* and *DYW1* in planta. In these experiments, one protein is fused to the N-terminal half of the yellow fluorescent protein (nYFP) and the potential partner is fused to the C-terminal half of YFP (cYFP). *Arabidopsis* plantlets were transiently cotransformed with pairs of plasmids encoding YFP fusions (Table 1). Because each half of the YFP is not intrinsically fluorescent, YFP fluorescence is observed only when intermolecular interactions occur between nYFP- and cYFP-tagged proteins (Citovsky et al., 2006; Marion et al., 2008). The emission spectrum of YFP was confirmed for each positive interaction (see Supplemental Figure 3 online). The results presented in Table 1 and Figure 5 show a positive interaction between *CRR4*-cYFP and *DYW1*-nYFP fusion proteins that was not observed when the opposite interaction (*CRR4*-nYFP X *DYW1*-cYFP) was assayed. Furthermore, no BiFC signal was observed for coexpressed *DYW1*-cYFP and *DYW1*-nYFP, suggesting that *DYW1* does not dimerize, unlike known cytidine deaminases (Prochnow et al., 2007). To test the specificity of the interaction between *DYW1* and *CRR4*, we tested the interaction

**Table 1.** BiFC Analysis of in Vivo Interactions between *DYW1*, *CRR4*, and Control Proteins

Interaction Tested	<i>DYW1</i> -cYFP	<i>CRR4</i> -cYFP	CLB19-cYFP	FSD2-cYFP	MinD1-cYFP
<i>DYW1</i> -nYFP	–	+	–	–	–
<i>CRR4</i> -nYFP	–	–	–	–	–
FSD3-nYFP	–	–	–	+	–
MinD1-nYFP	–	–	–	–	+

*Arabidopsis* plantlets were cotransformed with *DYW1*, *CRR4*, CLB19, FSD, and MinD1 fused to either the N- (rows) or C-terminal (columns) halves of YFP. Interactions were scored based on either the presence (+) or absence (–) of a BiFC (YFP) signal in the plastids. For each pair of plasmids tested, BiFC results were scored from at least two independent experiments. Blank cells correspond to untested interactions.

between DYW1-nYFP and CLB19-cYFP. CLB19 is a plastid PLS-E protein necessary for the editing of *clpP* and *rpoA* sites (Chateigner-Boutin et al., 2008) and functionally and structurally very similar to CRR4. CLB19 did not interact with DYW1 in this assay as shown by the absence of YFP fluorescence in plant cells (Figure 5). In addition, several negative controls testing for interactions between either DYW1-nYFP or CRR4-cYFP and unrelated proteins that have been shown to interact in plastids using BiFC were tested: MinD1  $\times$  MinD1 (Maple et al., 2007) and FSD2  $\times$  FSD3 (Myouga et al., 2008). The consistent lack of YFP fluorescence in these control cells (Figure 5, Table 1) supports that the fluorescence observed in the BiFC assays resulted from a specific CRR4-DYW1 interaction.

#### DYW1 Provides the Amino Acid Sequences Missing from CRR4

Approximately 50% of the editing factors identified so far contain the E, E+, and DYW motifs as defined by Lurin et al. (2004), whereas the other 50% lack these C-terminal domains to varying degrees. CRR4 lacks a DYW domain and the 14 C-terminal amino acids of the E+ motif. Interestingly, DYW1 contains not only a complete DYW domain, but also the C-terminal segment of the E+ motif (Figure 6). Thus, a complex of CRR4 and DYW1 would reconstitute all of the amino acid motifs found in full-length editing factors, such as CRR22 (Okuda et al., 2009).

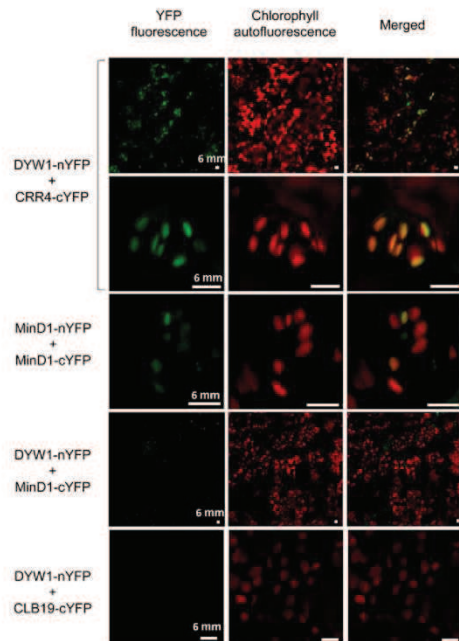
To test this model, we crossed the *crr4-3* mutant with the *dyw1-1* mutant to obtain a homozygous double mutant and then attempted to complement it with a chimeric construct consisting of the CRR4 PPR and E-E+ domains and the DYW1 E+-DYW domain (Figure 7). As one would expect, the double mutant completely lacks *ndhD-1* editing. Eight lines complemented with the chimeric construct were tested and all were shown to be restored for the editing of *ndhD-1* by cDNA sequencing (Figure 7). Seven of the eight plants tested showed a higher proportion of edited transcripts (more than 80%) than the wild type. Thus, a fusion of CRR4-DYW1 is functional, further evidence in favor of interaction between these two proteins.

## DISCUSSION

### Editing of *ndhD-1* Requires at Least Four Proteins

Through genetic studies, we have shown that DYW1 is necessary for editing of the *ndhD-1* site in chloroplast transcripts of *Arabidopsis*, the same site that was shown previously to require CRR4. As the molecular process requiring these two proteins is the same and because they are both necessary for this function, the simplest model is that they interact to achieve this function. Indeed, we demonstrated an interaction between DYW1 and CRR4 using transgenic fusion proteins in *Arabidopsis* and showed that a fusion of CRR4 and part of DYW1 is capable of complementing a double mutant lacking both proteins.

Although PPR proteins have often been observed to be components of high molecular weight protein or RNA-protein complexes (Uyttewaal et al., 2008; Olinas et al., 2010; Klodmann et al., 2011), only a few specific protein-PPR interactions have been identified. The PPR protein GLUTAMINE-RICH

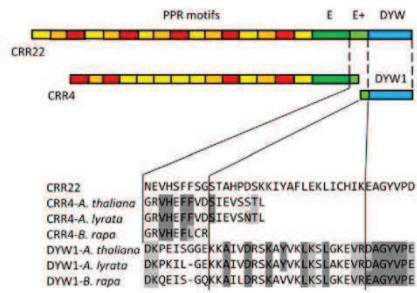


**Figure 5.** In Planta Protein Interaction of DYW1 and CRR4 as Shown by BiFC.

BiFC of YFP in transiently transformed *Arabidopsis* plantlets. Left column, YFP fluorescent signal detection by confocal microscopy; middle column, chlorophyll autofluorescence; right column, merge of fluorescent signal and autofluorescence. Cotransformation of *Arabidopsis* plantlets with DYW1-nYFP and CRR4-cYFP (top lines) generates yellow fluorescence that colocalizes with autofluorescence of chlorophylls in plastids. Similar signals are observed when using the homodimerization of MinD1 (third panel) as a positive control (Maple et al., 2007). No signal was observed when plantlets were cotransformed with either DYW1-nYFP and MinD1-cYFP or DYW1-nYFP and CLB19-cYFP (bottom panels). Bars = 6  $\mu$ m.

PROTEIN23 was reported to interact with RNA polymerase II in the nucleus (Ding et al., 2006); similarly, PPR PROTEIN LOCALIZED TO THE NUCLEUS AND MITOCHONDRIA was shown to interact with two nuclear proteins, NUCLEOSOME ASSEMBLY PROTEIN and the transcription factor TCP8 (Hammani et al., 2011a). DELAYED GREENING1, a PPR protein, interacts with SIGMA FACTOR6, a cofactor required for the transcription of plastid-encoded RNA polymerase-dependent chloroplast genes in *Arabidopsis* cotyledons (Chi et al., 2010). In none of these cases has the functional relevance of the observed interactions been elucidated.

Unlike these prior examples, the CRR4/DYW1 interaction involves two PPR proteins implicated in the same molecular event



**Figure 6.** CRR4 and DYW1 Each Contain a Different Part of the E+ Domain.

The partial E+ domains of CRR4 and its putative orthologs, DYW1 and its putative orthologs, and the PPR-DYW editing factor CRR22 (Okuda et al., 2009) were aligned using ClustalW. Residues identical to those in CRR22 are shaded in dark gray, and residues similar to those in CRR22 are shaded in light gray. CRR4 contains the N-terminal segment of the E+ domain, whereas the C-terminal segment of the domain is found in DYW1. There is no apparent overlap of conserved sequence.

on the same target RNA molecule, making it easier to understand the functional implications of the interaction. However, although clearly both proteins are required for editing to occur, we have not formally shown that their interaction is required for their function, and it is still possible that they could act sequentially. The proteins MORF2 and MORF9 are also required for *ndhD-1* RNA editing (Takenaka et al., 2012), whereas a third member of the MORF family, RIP1 (=MORF8), facilitates editing at this site but is not strictly required (Bentolila et al., 2012). Based on the fact that several MORF-PPR interactions have been demonstrated (Bentolila et al., 2012; Takenaka et al., 2012), one or more of these MORF proteins are likely to interact with CRR4 and/or DYW1. The role of MORF proteins in editing is not clear yet.

#### Conservation of CRR4 and DYW1 across Species

In *Arabidopsis*, editing at the *ndhD-1* site is necessary for correct synthesis and assembly of the NDH complex. This editing site is conserved across many dicots (Tsudzuki et al., 2001), suggesting that the editing factors should also be conserved. Hayes and Mulligan (2011) have recently shown a strict correlation between the presence of the *ndhD-1* editing site in chloroplast transcripts and of an apparent CRR4 ortholog in the corresponding nuclear genome. However, the length of the E domain of these CRR4 homologs is not conserved (see Supplemental Figure 4 online); in Brassicaceae, this sequence is shorter than in the other dicots. In parallel, DYW1 proteins containing the missing E domain sequence and no PPR domain are only found in Brassicaceae (see Supplemental Table 2 and Supplemental Figure 5 online).

#### Is Association with DYW Proteins a General Feature of E-Class Editing Factors?

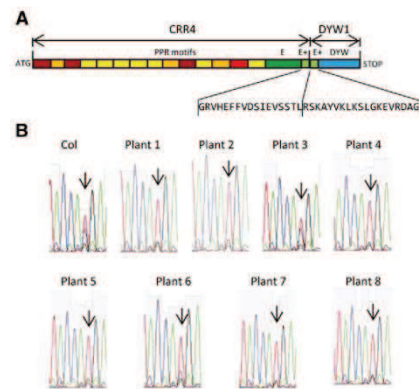
As discussed in the Introduction, 50% of known editing factors lack DYW domains. Do all these DYW-less proteins associate

with DYW-containing partners, or is the CRR4/DYW1 association an isolated case? Currently, there are no published data that can help us decide either way, but it is tempting to speculate that other similar examples of such associations will be discovered.

Although the *Arabidopsis* genome contains 87 DYW proteins, DYW1 is unique in containing no E, E+, or PPR motifs. However, five other *Arabidopsis* DYW proteins contain only a few PPR motifs and a poorly conserved E domain; these make ideal candidates for interaction partners for E/E+ proteins. If all E/E+ proteins have DYW partners, then these associations would have to be less specific than that between CRR4 and DYW1. Our present results show that the *dyw1-1* mutant is affected only in *ndhD-1* editing and that the physical interaction between CRR4 and DYW1, which was not observed between CLB19 and DYW1, seems also to be specific.

#### What Does the Association of CRR4 and DYW1 Imply for the Editing Enzyme?

Our hypothesis is that the CRR4 protein interacts with the DYW1 protein to form a protein complex functionally equivalent to a PPR PLS-DYW protein that contributes to both specifically binding and editing the *ndhD-1* site. This model is similar to the model in mammals in which ACF, the specificity factor, interacts with APOBEC-1, the enzyme, to edit the *apoB* transcript. The hypothesis that the DYW domain harbors the RNA editing activity is based on the presence of the conserved signature HxE (x)nCxXC, found in editing enzymes in other organisms (Salone



**Figure 7.** Complementation of the *crr4-3 dyw1-1* Double Mutant by a CRR4-DYW1 Fusion.

(A) Structure of the CRR4-DYW1 construct. The PPR motifs and E/E+ domain of the CRR4 protein were fused to the E+ domain and DYW domain of DYW1.

(B) Sequencing chromatograms of RT-PCR products obtained from complemented *crr4-3 dyw1-1* double mutants expressing the CRR4-DYW1 fusion. Each trace is from a different independent transformant. Arrows indicate the *ndhD-1* editing site.

et al., 2007), and broader similarities to a wide class of nucleotide deaminases (Iyer et al., 2011). This hypothesis was weakened by the discovery that many editing factors lack the DYW domain, and even those that contain it do not always require it in vivo (Okuda et al., 2009). However, we have shown that the DYW domain can be supplied in trans to the CRR4 protein and that in this case, it is essential for editing to occur. If, in the future, this phenomenon could be generalized to the other E/E+ editing factors, it would effectively eliminate one of the arguments against the DYW domain being the editing enzyme.

## METHODS

The complete list of oligonucleotides used in this study is summarized in Supplemental Table 3 online.

### Plant Material, Growth Conditions, and Complementation Analysis

*Arabidopsis thaliana* ecotype Col-0 was used in this study. Seeds were surface sterilized, vernalized at 4°C for 3 d, and grown on half-strength Murashige and Skoog media containing 3% Suc in vitro. Plates were placed in growth chambers under 16 h light/8 h dark, at 25°C, and with 45% humidity. Two-week-old seedlings were transferred onto soil and grown under 16 h light/8 h dark at 21°C and 65% humidity. A screening for EMS-induced mutations was ordered at the Seattle *Arabidopsis* TILLING Project (Till et al., 2003; <http://tilling.fhcrc.org/>) using two specific primers (TILLING\_LP and TILLING\_RP) surrounding a 1509-bp region containing the At1g47580 locus. The identified EMS lines were obtained from the Nottingham Arabidopsis Stock Centre (see Supplemental Table 1 online). EMS mutants were backcrossed three times to Col-0 ecotype. EMS lines were genotyped by amplification of a 1502-bp product using DYW1\_compF and DYW1\_compR primers and sequencing using DYW1\_genF and DYW1\_genR primers. The *crr4-3* mutant was provided by Toshiharu Shikanai (Kotera et al., 2005).

For complementation analysis, the complete DYW1 locus with its native promoter and terminator was amplified by PCR using DYW1\_compF and DYW1\_compR primers on genomic *Arabidopsis* Col-0 DNA, cloned into the pDNR207 vector by Gateway BP reaction (Invitrogen), and subcloned into pGWB1 vector (Nakagawa et al., 2007) by LR reaction. For the double mutant, CRR4 was amplified with CRR4\_ATG\_F and CRR4-DYW1\_R and DYW1 was amplified with CRR4-DYW1\_F and DYW1\_STOP\_R on *Arabidopsis* Col-0 genomic DNA. The fusion was achieved by pooling the purified PCR products and amplifying by PCR with CRR4\_ATG\_F and DYW1\_STOP\_R. The fusion product was cloned into the pDNR207 vector by Gateway BP reaction (Invitrogen) and subcloned into pGWB2.

### Protein Expression and Antibody Production

The DYW1 coding region without its presequence was amplified using DYW1\_internal and DYW1\_stop primers and cloned into the pDNR207 vector by Gateway BP reaction (Invitrogen). LR recombination was done with the pDEST 17 destination vector (Invitrogen) allowing an N-terminal fusion with a 6His tag. Proteins were expressed for 5 h at 37°C in salt-inducible BL21-SI *Escherichia coli* induced with 0.3 M NaCl and purified by affinity to nickel-nitrilotriacetic acid agarose in denaturing conditions according to the manufacturer's instructions (Qiagen). Purified DYW1 protein was used to immunize rabbits to produce polyclonal antibodies (Eurogentec).

### Subcellular Localization

The full-length ORF or the 300 bp encoding the first 100 amino acids of DYW1 were amplified using DYW1\_start, DYW1\_end, and DYW1\_preseq primers and subsequently cloned into the pDNR207 vector by BP reaction.

The LR recombination was done with the pGreen 0229 destination vector containing the green fluorescent protein (GFP) or RFP gene (Lurin et al., 2004). C58C1 pSOUP *Agrobacterium tumefaciens* transformed with these binary plasmids was used for *Arabidopsis* transformation (Marion et al., 2008). Plantlets were observed 3 d after transformation by confocal microscopy (Leica SP2 AOBS 405) allowing the detection of GFP (excitation/emission 488/509nm, filter BP500/550), RFP (excitation/emission 543/583nm, filter BP580/620), and autofluorescence (excitation/emission 543/680nm, filter BP580/620).

For immunoblot analysis, *Arabidopsis* chloroplast and mitochondrial protein extracts were prepared as previously described (Hegeman et al., 2005; Sweetlove et al., 2007). Proteins were separated in 12.5% SDS-PAGE gel, transferred onto polyvinylidene difluoride membrane by electrotransfer, and incubated with the DYW1-specific antibody diluted in TBS-T (1/5000), the formate dehydrogenase-specific antibody (1/6000 in TBS-T) (Colas des Francs-Smal et al., 1993), or the chloroplast large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase-specific antibody (1/20,000 in TBS-T) (supplied by Spencer Whitney, Australian National University). After incubation with the anti-rabbit IgG antibody (1/10,000 in TBS-T), immunoblots were analyzed using the ECL Western Blotting Analysis System reaction kit (GE Healthcare) and visualized with LAS-1000 (Fujifilm).

### Analysis of RNA Editing

RNA from leaves of 18-d-old plantlets was extracted with the RNeasy plant mini kit (Qiagen). RNA was treated twice with DNase I (2 units/μL; Ambion) for 30 min at 37°C, and cDNA was synthesized using Superscript II (Invitrogen). RT-PCR products were obtained with NdhD\_AT\_For and NdhD\_AT\_rev primers surrounding the *NdhD-1* editing site (117,166) and used as template for sequencing using the NdhD\_AT\_For primer.

Poisoned primer extension of RT-PCR products was performed as described by Chateigner-Boutin and Small (2007). RT-PCR products were obtained with NdhD\_AT\_For and NdhD\_AT\_rev primers and serve as templates for the extension reaction from 5'-Carboxyfluorescein-labeled ndhD\_PPE\_C primer purified on reverse phase cartridge (Sigma Genosys) that anneals next to the editing site. The extension was stopped by the incorporation of 2',3'-dideoxycytidine-5'-triphosphate at the location of the editing site for unedited molecules, producing a short unedited product. The extension was stopped at the next G/C for the edited molecules, producing a longer edited product.

### RNA Gel Blot Analysis

The *ndhD* RNA probe was labeled with biotinylated cytidine in vitro transcription of a PCR product obtained and cloned in pGEM-T Easy vector (Promega) as described by Hammani et al. (2009). The PCR product served as a template for in vitro transcription with SP6 polymerase following the manufacturer's instructions (Maxiscript Ambion). Ten micrograms of total RNA extracted from leaves of 14- and 28-d-old plantlets (RNeasy plant mini kit; Qiagen) were separated on 1.2% (w/v) formaldehyde agarose gel and transferred onto Hybond N<sup>+</sup> nylon membranes (GE Healthcare). After transfer, the membrane was stained with 0.04% methylene blue to check RNA integrity, loading, and transfer and subsequently hybridized with the biotinylated *ndhD* antisense RNA probe according to Hammani et al. (2009).

### Chlorophyll Fluorescence Analysis

Chlorophyll fluorescence was measured using a MINI-PAM portable chlorophyll fluorometer (Waltz). The transient increase in chlorophyll fluorescence after turning off AL was monitored as previously described (Shikanai et al., 1998). Leaves were exposed to AL (50 μmol photons m<sup>-2</sup> s<sup>-1</sup>) for 5 min. AL was turned off and the subsequent transient rise in fluorescence ascribed to NDH activity was monitored by chlorophyll fluorimetry.

**Split-YFP Assay**

The full-length ORFs of *DYW1*, *CRR4*, *CLB19*, *FSD2*, *FSD3*, and *MIND1* were amplified without their stop codon using corresponding ORF\_start and ORF\_end primers and cloned into the pDNR207 vector. LR recombinations were done with split-YFP destination vectors. pBiFC1 and pBiFC4, coding for the N- and C-terminal YFP moieties, respectively, cloned at the 3' end (C-terminal fusions) of the Gateway recombination sequence, were used (Azimzadeh et al., 2008). C58C1 *Agrobacterium* pCH32 containing these vectors was used to transform 15 to 20 seedlings of Landsberg *erecta* *Arabidopsis* seedlings grown in six-well plates as described previously (Marion et al., 2008). Cotyledons of transformed seedlings were observed 3 to 4 d after transformation using a confocal microscope (Leica SP2 AOBS diode 405 with two filters: GFP [band-pass], excitation filter band-pass 450 to 490, stop filter band-pass 500 to 550; I3 [long pass], excitation filter band-pass 450 to 490, stop filter band-pass 515).

**Accession Numbers**

Sequence data from this article can be found in the Arabidopsis Genome Initiative or GenBank/EMBL databases under the following accession numbers: AT1G47580 (*DYW1*), AT2G45350 (*CRR4*), ATCG01050 (*NDHD*), AT5G24020 (*MIND1*), AT5G51100 (*FSD2*), AT5G23310 (*FSD3*), and AT1G11290 (*CRR2*).

**Supplemental Data**

The following materials are available in the online version of this article.

**Supplemental Figure 1.** Analysis of the Editotype of the *dyw1-1* Mutant.

**Supplemental Figure 2.** Analysis of the Editing Rate of the *ndhD-1* Site in *dyw1-x* Mutants.

**Supplemental Figure 3.** Fluorescence Emission Spectral Analysis of DYW1-nYFP X CRR4-cYFP BiFC Signals.

**Supplemental Figure 4.** Alignment of CRR4 Orthologs.

**Supplemental Figure 5.** Alignment of DYW1 Orthologs.

**Supplemental Table 1.** Complete List of EMS Mutations Identified in the *DYW1* Locus by the TILLING Approach.

**Supplemental Table 2.** DYW1 Orthologs.

**Supplemental Table 3.** Complete List of Oligonucleotides Used in This Study.

**ACKNOWLEDGMENTS**

We thank Etienne Delannoy (Unité de Recherche en Génomique Végétale, Evry, France) and Andéol Falcon de Longevialle (Unité de Recherche en Génomique Végétale, Evry, France) for valuable discussions and comments on the article and Olivier Grandjean (Institut Jean Pierre Bourgin, Versailles, France) and Lionel Gissot (Institut Jean Pierre Bourgin, Versailles, France) for their help with confocal microscopy. Research at the University of Western Australia was supported by Australian Research Council Grant CE0561495. We thank Tsuyoshi Endo (Kyoto University, Kyoto, Japan) and Amane Makino (Tohoku University, Sendai, Japan) for giving us anticytochrome *b<sub>f</sub>* and *NdhH* polyclonal antibodies, respectively. Collaboration between the University of Western Australia and Unité de Recherche en Génomique Végétale was supported by Australian Government International Science Linkages Grants FR060030 and CG120098. K.H. is supported by a postgraduate scholarship from the University of Western Australia.

**AUTHOR CONTRIBUTIONS**

C.B. characterized the EMS lines, analyzed RNA editing, and drafted the article. V.S. characterized the T-DNA lines and localized DYW1 by immunoblot. C.B. and A.A. complemented the *dyw1-1* mutant and the *crr4 dyw1* double mutant. A.A. and R.B. localized DYW1 by confocal analysis and analyzed CRR4-DYW1 interaction by split-YFP assay. K.H., K.O., and T.S. measured the chlorophyll fluorescence of the *dyw1-1* mutant and complemented plants. I.S. helped design the research, analyze the data, and write the article. C.L. conceived and coordinated the project and drafted the article.

Received April 12, 2012; revised July 30, 2012; accepted August 30, 2012; published September 21, 2012.

**REFERENCES**

- Alonso, J.M., et al. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**: 653–657.
- Andres, C., Lurin, C., and Small, I.D. (2007). The multifarious roles of PPR proteins in plant mitochondrial gene expression. *Physiol. Plant.* **129**: 14–22.
- Azimzadeh, J., Nacry, P., Christodoulidou, A., Drevensek, S., Camilleri, C., Amour, N., Parcy, F., Pastuglia, M., and Bouchez, D. (2008). *Arabidopsis* TONNEAU1 proteins are essential for preprophase band formation and interact with centrin. *Plant Cell* **20**: 2146–2159.
- 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., Elliott, L.E., and Hanson, M.R. (2008). Genetic architecture of mitochondrial editing in *Arabidopsis thaliana*. *Genetics* **178**: 1693–1708.
- Bentolila, S., Heller, W.P., Sun, T., Babina, A.M., Friso, G., van Wijk, K.J., and Hanson, M.R. (2012). RIP1, a member of an Arabidopsis protein family, interacts with the protein RARE1 and broadly affects RNA editing. *Proc. Natl. Acad. Sci. USA* **109**: E1453–E1461.
- Blanc, V., Henderson, J.O., Kennedy, S., and Davidson, N.O. (2001). 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.
- Bock, R., Hermann, M., and Kössel, H. (1996). In vivo dissection of cis-acting determinants for plastid RNA editing. *EMBO J.* **15**: 5052–5059.
- Bock, R., Kössel, 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.
- Castelli, V., et al. (2004). Whole genome sequence comparisons and "full-length" cDNA sequences: A combined approach to evaluate and improve Arabidopsis genome annotation. *Genome Res.* **14**: 406–413.
- Chateigner-Boutin, A.L., Ramos-Vega, M., Guevara-García, A., Andrés, C., de la Luz Gutiérrez-Nava, M., Cantero, A., Delannoy, E., Jiménez, L.F., Lurin, C., Small, I., and León, P. (2008). CLB19, a pentatricopeptide repeat protein required for editing of *rpoA* and *clpP* chloroplast transcripts. *Plant J.* **56**: 590–602.
- 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.



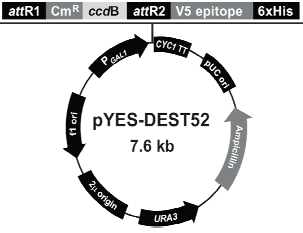
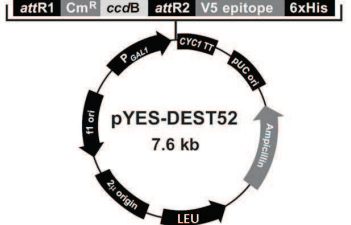
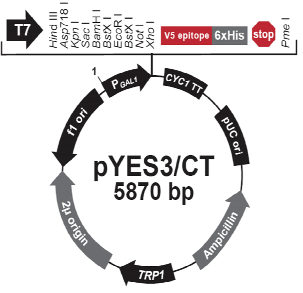
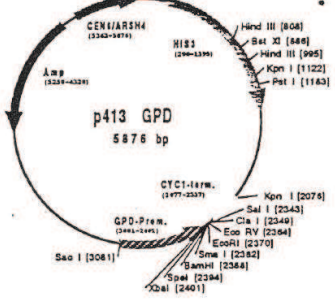
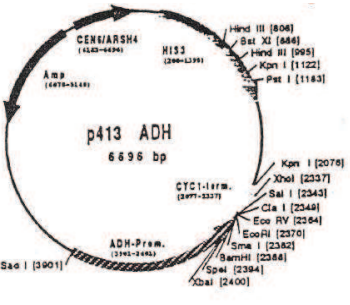
- Chateigner-Boutin, A.L., and Small, I.** (2010). Plant RNA editing. *RNA Biol.* **7**: 213–219.
- Chaudhuri, S., and Maliga, P.** (1996). Sequences directing C to U editing of the plastid psbL mRNA are located within a 22 nucleotide segment spanning the editing site. *EMBO J.* **15**: 5958–5964.
- Chen, S.H., 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.
- Chi, W., Mao, J., Li, Q., Ji, D., Zou, M., Lu, C., and Zhang, L.** (2010). Interaction of the pentatricopeptide-repeat protein DELAYED GREENING 1 with sigma factor SIG6 in the regulation of chloroplast gene expression in *Arabidopsis* cotyledons. *Plant J.* **64**: 14–25.
- Citovsky, V., Lee, L.Y., Vyas, S., Glick, E., Chen, M.H., Vainstein, A., Gafni, Y., Gelvin, S.B., and Tzfira, T.** (2006). Subcellular localization of interacting proteins by bimolecular fluorescence complementation in plants. *J. Mol. Biol.* **362**: 1120–1131.
- Colas des Francs-Small, C., Ambard-Bretteville, F., Small, I.D., and Rémy, R.** (1993). Identification of a major soluble protein in mitochondria from nonphotosynthetic tissues as NAD-dependent formate dehydrogenase. *Plant Physiol.* **102**: 1171–1177.
- 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.
- 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.
- Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H.** (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2**: 953–971.
- Fujii, S., and Small, I.** (2011). The evolution of RNA editing and pentatricopeptide repeat genes. *New Phytol.* **191**: 37–47.
- Giegé, P., and Brennicke, A.** (1999). RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs. *Proc. Natl. Acad. Sci. USA* **96**: 15324–15329.
- Hammani, K., des Francs-Small, C.C., Takenaka, M., Tanz, S.K., Okuda, K., Shikanai, T., Brennicke, A., and Small, I.** (2011b). The pentatricopeptide repeat protein OTP87 is essential for RNA editing of *nad7* and *atp1* transcripts in *Arabidopsis* mitochondria. *J. Biol. Chem.* **286**: 21361–21371.
- Hammani, K., Gobert, A., Hleibieh, K., Choulier, L., Small, I., and Giegé, P.** (2011a). An *Arabidopsis* dual-localized pentatricopeptide repeat protein interacts with nuclear proteins involved in gene expression regulation. *Plant Cell* **23**: 730–740.
- Hammani, K., Okuda, K., Tanz, S.K., Chateigner-Boutin, A.L., Shikanai, T., and Small, I.** (2009). A study of new *Arabidopsis* chloroplast RNA editing mutants reveals general features of editing factors and their target sites. *Plant Cell* **21**: 3686–3699.
- Hayes, M.L., and Mulligan, R.M.** (2011). Pentatricopeptide repeat proteins constrain genome evolution in chloroplasts. *Mol. Biol. Evol.* **28**: 2029–2039.
- 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.
- Iyer, L.M., Zhang, D., Rogozin, I.B., and Aravind, L.** (2011). Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. *Nucleic Acids Res.* **39**: 9473–9497.
- Klodmann, J., Senkler, M., Rode, C., and Braun, H.P.** (2011). Defining the protein complex proteome of plant mitochondria. *Plant Physiol.* **157**: 587–598.
- Knoop, V.** (2011). When you can't trust the DNA: RNA editing changes transcript sequences. *Cell. Mol. Life Sci.* **68**: 567–586.
- 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., et al.** (2004). Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* **16**: 2089–2103.
- Maple, J., Vojta, L., Soll, J., and Moller, S.G.** (2007). ARC3 is a stromal Z-ring accessory protein essential for plastid division. *EMBO Rep.* **8**: 293–299.
- Marion, J., Bach, L., Bellec, Y., Meyer, C., Gissot, L., and Faure, J.D.** (2008). Systematic analysis of protein subcellular localization and interaction using high-throughput transient transformation of *Arabidopsis* seedlings. *Plant J.* **56**: 169–179.
- 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.
- Myamoto, T., Obokata, J., and Sugura, 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.
- Myouga, F., Hosoda, C., Umezawa, T., Iizumi, H., Kurumori, T., Motohashi, R., Shono, Y., Nagata, N., Ikeuchi, M., and Shinozaki, K.** (2008). A heterocomplex of iron superoxide dismutases defends chloroplast nucleoids against oxidative stress and is essential for chloroplast development in *Arabidopsis*. *Plant Cell* **20**: 3148–3162.
- Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., Toyooka, K., Matsuoka, K., Jinbo, T., and Kimura, T.** (2007). Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J. Biosci. Bioeng.* **104**: 34–41.
- Okuda, K., Chateigner-Boutin, A.L., Nakamura, T., Delannoy, E., Sugita, M., Myouga, F., Motohashi, R., Shinozaki, K., Small, I., and Shikanai, T.** (2009). Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in *Arabidopsis* chloroplasts. *Plant Cell* **21**: 146–156.
- 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.
- Olinares, P.D., Ponnala, L., and van Wijk, K.J.** (2010). Megadalton complexes in the chloroplast stroma of *Arabidopsis thaliana* characterized by size exclusion chromatography, mass spectrometry, and hierarchical clustering. *Mol. Cell. Proteomics* **9**: 1594–1615.
- 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.
- 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.
- 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.
- Rüdinger, M., Fritz-Laylin, L., Polsakiewicz, M., and Knoop, V.** (2011). Plant-type mitochondrial RNA editing in the protist *Naegleria gruberi*. *RNA* **17**: 2058–2062.
- Salone, V., Rüdinger, M., Polsakiewicz, M., Hoffmann, B., Groth-Malonek, M., Szurek, B., Small, I., Knoop, V., and Lurin, C.** (2007).

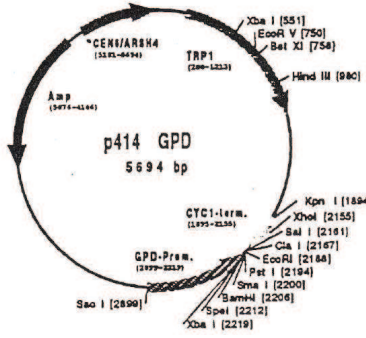
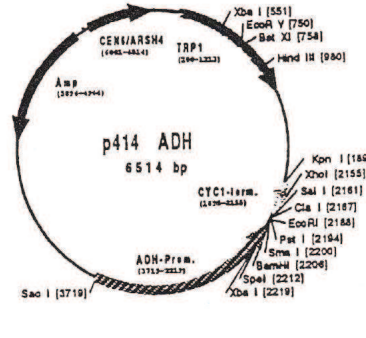
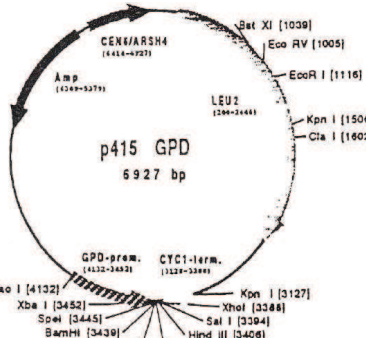
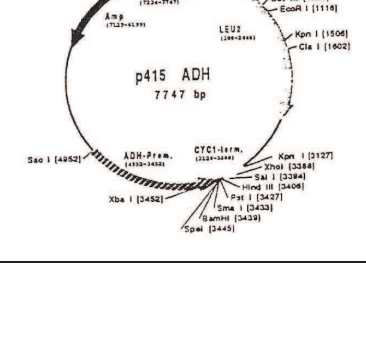
- A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Lett.* **581**: 4132–4138.
- Schmitz-Linneweber, C., and Small, I.** (2008). Pentatricopeptide repeat proteins: A socket set for organelle gene expression. *Trends Plant Sci.* **13**: 663–670.
- Shikanai, T., Endo, T., Hashimoto, T., Yamada, Y., Asada, K., and Yokota, A.** (1998). Directed disruption of the tobacco *ndhB* gene impairs cyclic electron flow around photosystem I. *Proc. Natl. Acad. Sci. USA* **95**: 9705–9709.
- 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.
- Steinhauser, S., Beckert, S., Capesius, I., Malek, O., and Knoop, V.** (1999). Plant mitochondrial RNA editing. *J. Mol. Evol.* **48**: 303–312.
- Sweetlove, L.J., Taylor, N.L., and Leaver, C.J.** (2007). Isolation of intact, functional mitochondria from the model plant *Arabidopsis thaliana*. *Methods Mol. Biol.* **372**: 125–136.
- Takenaka, M., Zehrmann, A., Verbitskiy, D., Kugelmann, M., Härtel, B., and Brennicke, A.** (2012). Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. *Proc. Natl. Acad. Sci. USA* **109**: 5104–5109.
- Tasaki, E., Hattori, M., and Sugita, M.** (2010). The moss pentatricopeptide repeat protein with a DYW domain is responsible for RNA editing of mitochondrial *ccmF*c transcript. *Plant J.* **62**: 560–570.
- Teng, B., Burant, C.F., and Davidson, N.O.** (1993). Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science* **260**: 1816–1819.
- Till, B.J., et al.** (2003). Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Res.* **13**: 524–530.
- Tsudzuki, T., Wakasugi, T., and Sugiura, M.** (2001). Comparative analysis of RNA editing sites in higher plant chloroplasts. *J. Mol. Evol.* **53**: 327–332.
- Uchida, M., Ohtani, S., Ichinose, M., Sugita, C., and Sugita, M.** (2011). The PPR-DYW proteins are required for RNA editing of *rps14*, *cox1* and *nad5* transcripts in *Physcomitrella patens* mitochondria. *FEBS Lett.* **585**: 2367–2371.
- Uyttewaal, M., Mireau, H., Rurek, M., Hammani, K., Arnal, N., Quadrado, M., and Giegé, P.** (2008). PPR336 is associated with polysomes in plant mitochondria. *J. Mol. Biol.* **375**: 626–636.
- Wedekind, J.E., and McKay, D.B.** (2003). Crystal structure of the leadzyme at 1.8 Å resolution: Metal ion binding and the implications for catalytic mechanism and allo site ion regulation. *Biochemistry* **42**: 9554–9563.
- Yamamoto, H., Peng, L., Fukao, Y., and Shikanai, T.** (2011). An Src homology 3 domain-like fold protein forms a ferredoxin binding site for the chloroplast NADH dehydrogenase-like complex in *Arabidopsis*. *Plant Cell* **23**: 1480–1493.
- Zehrmann, A., van der Merwe, J.A., Verbitskiy, D., Brennicke, A., and Takenaka, M.** (2008). Seven large variations in the extent of RNA editing in plant mitochondria between three ecotypes of *Arabidopsis thaliana*. *Mitochondrion* **8**: 319–327.

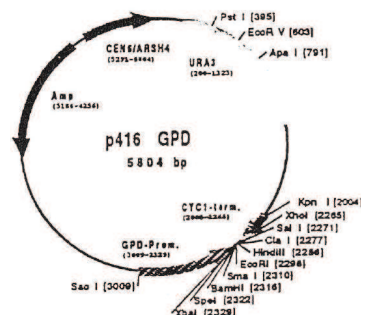
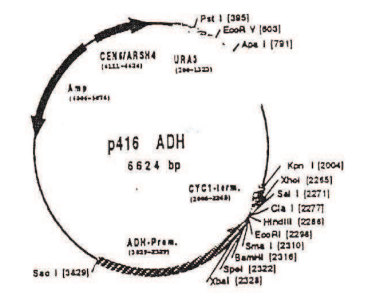
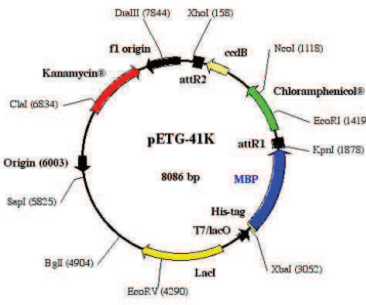
## II. Liste des principaux vecteurs utilisés

Nom	Fournisseur	Clonage	Résistance	Carte
pGEMT-easy	Promega A1360	Recombinaison de produit PCR	Cb <sup>R</sup>	
pDONR207	Invitrogen 12213-013	Gateway-BP	Genta <sup>R</sup>	
pDONR221	Invitrogen 12536-017	Gateway-BP	Kan <sup>R</sup>	
pGWB1	INRA-Versailles	Gateway-LR	Kan <sup>R</sup>	Sans promoteur, sans tag (--R1-CmR-ccdB-R2-).
pGWB2	INRA-Versailles	Gateway-LR	Kan <sup>R</sup>	35S, sans tag (--35S promoteur-R1-CmR-ccdB-R2--).
pGWB4	INRA-Versailles	Gateway-LR	Kan <sup>R</sup>	Sans promoteur, C-sGFP (--R1- CmR-ccdB-R2-sGFP--).
pGWB15	INRA-Versailles	Gateway-LR	Kan <sup>R</sup>	35S, N-3xHA (--35S promoteur-3xHA-R1- CmR- ccdB-R2--)
pGWB20	INRA-Versailles	Gateway-LR	Kan <sup>R</sup>	35S, C-10xMyc (--35S promoteur-R1- CmR-ccdB-R2-10xMyc)
pDEST14	Invitrogen 11801-016	Gateway-LR	Cb <sup>R</sup>	

pDEST15	Invitrogen 11802-014	Gateway-LR	Cb <sup>R</sup>	<p>pDEST<sup>TM</sup>15 7013 bp</p> <p>Ampicillin</p> <p>PBR322 ori</p> <p>T7</p> <p>RBS ATG GST attR1 Cm<sup>R</sup> ccdB attR2 T7 term</p>
pBiFc1	INRA- Versailles	Gateway-LR	Spec <sup>R</sup>	<p>gateway Nter YFP</p>
pBiFc2	INRA- Versailles	Gateway-LR	Spec <sup>R</sup>	<p>Nter YFP gateway</p>
pBiFc3	INRA- Versailles	Gateway-LR	Spec <sup>R</sup>	<p>Cter YFP gateway</p>
pBiFc4	INRA- Versailles	Gateway-LR	Spec <sup>R</sup>	<p>gateway Cter YFP</p>
pETDuet-1	Novagen 71146	Enzyme de restriction	Cb <sup>R</sup>	<p>pETDuet-1 (5420 bp)</p>
pColaDuet-1	Novagen 71406	Enzyme de restriction	Kan <sup>R</sup>	<p>pCOLADuet-1 (5719 bp)</p>
pRARE	Novagen 71402 (Rosetta)	X	Cam <sup>R</sup>	<p>pRARE 4694 bp</p> <p>Cam</p> <p>pLysSRARE 7393 bp</p> <p>pLacIRARE 6320 bp</p> <p>proL tRNA leuW tRNA met tRNA argW tRNA thrT tRNA glyT tRNA tyrU tRNA thrU tRNA argU tRNA</p>

<p>pYES-Dest52 URA</p>	<p>Invitrogen 12286-019</p>	<p>Gateway-LR</p>	<p>Cb<sup>R</sup></p>	
<p>pYES-Dest52 LEU</p>	<p>Invitrogen modifié à l'INRA 12286-019</p>	<p>Gateway-LR</p>	<p>Cb<sup>R</sup></p>	
<p>pYES3/CT</p>	<p>Invitrogen V825320</p>	<p>Enzyme de restriction</p>	<p>Cb<sup>R</sup></p>	
<p>p413_GPD</p>	<p>Inserm-Groupe de Marc Blondel</p>	<p>Enzyme de restriction</p>	<p>Cb<sup>R</sup></p>	
<p>p413_ADH</p>	<p>Inserm-Groupe de Marc Blondel</p>	<p>Enzyme de restriction</p>	<p>Cb<sup>R</sup></p>	

p414_GPD	Inserm- Groupe de Marc Blondel	Enzyme de restriction	Cb <sup>R</sup>	
p414_ADH	Inserm- Groupe de Marc Blondel	Enzyme de restriction	Cb <sup>R</sup>	
p415_GPD	Inserm- Groupe de Marc Blondel	Enzyme de restriction	Cb <sup>R</sup>	
p415_ADH	Inserm- Groupe de Marc Blondel	Enzyme de restriction	Cb <sup>R</sup>	

p416_GPD	Inserm- Groupe de Marc Blondel	Enzyme de restriction	Cb <sup>R</sup>	
p416_ADH	Inserm- Groupe de Marc Blondel	Enzyme de restriction	Cb <sup>R</sup>	
pETG-41K	EMBL	Gateway-LR	Kan <sup>R</sup>	

Résistance :

Cb<sup>R</sup> : Carbénicilline

Kan<sup>R</sup> : Kanamycine

Genta<sup>R</sup> : Gentamycine

Cam<sup>R</sup> : Chloramphénicol

Spec<sup>R</sup> : Spectinomycine

### III. Liste des principaux oligonucléotides utilisés

<b>Oligonucléotides vecteurs-spécifiques</b>		
<b>Nom</b>	<b>Séquence (5'→3')</b>	<b>Description</b>
Easy_T7_MCS F	GGCGAATTGGGCCCGACG	Amplification T7 pGemTeasy F
Easy_Sp6_MCS R	CTCAAGCTATGCATCCAACG	Amplification Sp6 pGemTeasy R
ColaDuet_MCS2_F	ATCATCACCACAGCCAGGAT	Amplification insert pColaDuet F
ColaDuet_MCS2_R	CCCCAAGGGGTTATGCTAGT	Amplification insert pColaDuet R
Yes3-F	AACCCCGGATCGGACTACTA	Amplification de l'insert pYES3 F
Yes3-R	GGGAGGGCGTGAATGTAAG	Amplification de l'insert pYES3 R
DNR5	CTGGCAGTTCCCTACTCTCG	Amplification de l'insert pDNR207 F
DNR3	GATGGTCGGAAGAGGCATAA	Amplification de l'insert pDNR207 R

<b>Clonage BP</b>		
<b>Nom</b>	<b>Séquence (5'→3')</b>	<b>Description</b>
CompF	AAAGCAGGCTTCTGGAAGCGTAAAGGGTTTGA	attb1 + DYW1 promoteur natif (378pb avant l'ATG)
CompR	GAAAGCTGGGTCTTGGATGGTTTAAGGGAGGA	attb2 + DYW1 terminateur natif (428pb après le STOP)
U3	GGGGACCACTTTGTACAAGAAAAGCTGGGTC	Amplification totale attb2 (Universel)
U5	GGGGACAAGTTTGTACAAAAAAGCAGGCTTC	Amplification totale attb1 (Universel)

<b>Séquençage</b>		
<b>Nom</b>	<b>Séquence (5'→3')</b>	<b>Description</b>
DYW1_GST-F	GCCTGAAAGAAATTTAGATACTGGAGC	Amplification interne DYW1
DYW1_GST-R	ATTCAGCTTTACGTAGGCCTTGCTCTAT	Amplification interne DYW1
DYW1_gen_F	GGAAGCTGCTTTCTCCATG	Amplification ATG DYW1
DYW1_gen_R	TGGAAGTTGCATAAATGGCTTAT	Amplification STOP DYW1
CRR4_seq_F	TTGGTAATAGTTCTGCCTGCAA	Amplification interne CRR4



<b>Mutagenèse par PCR</b>		
<b>Nom</b>	<b>Séquence (5'→3')</b>	<b>Description</b>
M1F	CATGCTAGTGAACGGTTA	Mutation ponctuelle DYW1 : H <sub>163</sub>
M1R	ACTAGCATGCATCAAGGC	Mutation ponctuelle DYW1 : H <sub>163</sub>
M2F	AGTGCACGGTTAGCTATT	Mutation ponctuelle DYW1 : E <sub>165</sub>
M2R	CCGTGCACTATGATGCATC	Mutation ponctuelle DYW1 : E <sub>165</sub>
M3F	CATGCTAGTGCACGGTTAGCT	M1+M2
M3R	CCGTGCACTAGCATGCATCAAG	M1+M2
M4F	ATAGCTGGAGATTGC	Mutation ponctuelle DYW1 : C <sub>191</sub>
M4R	TCCAGCTATCCTCAG	Mutation ponctuelle DYW1 : C <sub>191</sub>
M5F	GATGCCATAACTTCA	Mutation ponctuelle DYW1 : C <sub>194</sub>
M5R	ATGGGCATCTCCACATAT	Mutation ponctuelle DYW1 : C <sub>194</sub>
M6F	ATAGCTGGAGATGCCATAACTT	M4+M5
M6R	ATGGGCATCTCCAGCTATCCTCAG	M4+M5
M7F	TTCGCTGCTTTAGAGAT	Mutation motif conservé : FHHF
M7R	AAAAGCAGCGAACCTCTT	Mutation motif conservé : FHHF
M8F	AACGCCTCTTGTTG	Mutation ponctuelle DYW1 : C <sub>225</sub>
M8R	AGAGGCGTTACCATC	Mutation ponctuelle DYW1 : C <sub>225</sub>
M9F	TCTGCTGGAGATTAT	Mutation ponctuelle DYW1 : C <sub>227</sub>
M9R	TCCAGCAGAGCAGTT	Mutation ponctuelle DYW1 : C <sub>227</sub>
M10F	AACGCCTCTGCTGGAGATTATTG	M8+M9
M10R	TCCAGCAGAGGCGTTACCATC	M8+M9
M11R	GAAAGCTGGGTCTTATCCACAAGAGCAGTTACCATC	attb2 + STOP + DYW1 sans triplet DYW
M12F	TGTGGAGCTTATTGGTAGACTTTTG	Mutation ponctuelle DYW1 : D <sub>229</sub>
M12R	CCAATAAGCTCCACAAGAGCAG	Mutation ponctuelle DYW1 : D <sub>229</sub>
M13F	TGTGGAGATGCTTGGTAGACTTTTG	Mutation ponctuelle DYW1 : Y <sub>230</sub>
M13R	TCTACCAAGCATCTCCACAAGAGC	Mutation ponctuelle DYW1 : Y <sub>230</sub>
M14F	TTGTGGAGATTATGCGTAGACTTTTG	Mutation ponctuelle DYW1 : W <sub>231</sub>
M14R	TCTACGCATAATCTCCACAAGAGC	Mutation ponctuelle DYW1 : W <sub>231</sub>
M15F	TTGCGGGTGAGAGTTATGTACCAGAGACT	Délétion PPR like
M15R	AGTCTCTGGTACATAACTCTCACCCGCAA	Délétion PPR like
M16R	GAAAGCTGGGTCTTAGCCTGCGTCTCTCACTTCTT	attb2 + STOP (TTA) + DYW1 sans le domaine DYW
M17F	CCTCCAGGGACGACAAATCTGAGGATATGT	Délétion feuillet Bêta
M17R	ACATATCCTCAGATTGTGTCCTGGAGG	Délétion feuillet Bêta
M18-F	GATATTAGCGAGTATCGAAGATAGAGAG	Mutation ponctuelle DYW1 : S <sub>202</sub> (Zehrmann et al., 2009)
M18-R	TCTTCGATACTCGCTAATATCTTTATGA	Mutation ponctuelle DYW1 : S <sub>202</sub> (Zehrmann et al., 2009)
M19-F	AAAGCTATTGTAGATGACGCAGGCTATGTA	Délétion RSKAYVKLKSGLGKEVR (16 aa E+)

M19-R	TACATAGCCTGCGTCATCTACAATAGCTTT	Délétion RSKAYVKLKLKSLGKEVR (16 aa E+)
M20-F	AAAGAAGTGAGAGACGAGACTAAGTATGTT	Délétion AGYVP
M20-R	AACATACTTAGTCTCGTCTCTCACTTCTTT	Délétion AGYVP
M21-F	TCGAGTATCGAAGATAACAAGAGGTTCCAT	Délétion REIIVRD (région conservée dans DYW)
M21-R	ATGGAACCTCTTGTATCTTCGATACTCGA	Délétion REIIVRD (région conservée dans DYW)
M22-F	CATCATAGTGAACGGTTCGGGATTATAAAC	Délétion LAIA (région conservée dans DYW)
M22-R	GTTTATAATCCCGAACCGTTCACTATGATG	Délétion LAIA (région conservée dans DYW)
M23-F	GCTATTGTAGATAGGCGCAAGGCCT	Mutation ponctuelle DYW1 : K <sub>123</sub>
M23-R	AGGCCTTGCGCCTATCTACAATAGC	Mutation ponctuelle DYW1 : K <sub>123</sub>
M24-F	TAGATAGAAGCAAGGCCGCGTAAA	Mutation ponctuelle DYW1 : Y <sub>125</sub>
M24-R	TTACGGCGGCCTTGCTTCTATCTA	Mutation ponctuelle DYW1 : Y <sub>125</sub>
M25-F	AGGCCTACGTAAAGCGCAAATCT	Mutation ponctuelle DYW1 : L <sub>128</sub>
M25-R	AGATTTCGCCTTTACGTAGGCCT	Mutation ponctuelle DYW1 : L <sub>128</sub>

<b>Editotype</b>		
<b>Nom</b>	<b>Séquence (5'→3')</b>	<b>Description</b>
matK_F	AGCGGCGTATCCTTTGTTCG	Editotype MatK
matK_R	CGTTACCGGGTAAAAGATGC	Editotype MatK
atpF_F	TTAATAGTACGTAAATGTAACCTC	Editotype atpF
atpF_R	AGAAGAGGAGATTATATGAAA	Editotype atpF
rpoC1_F	GTCCACGGATTCCATTATCAA	Editotype RpoC1
rpoC1_R	TTGGACCTATAAAGAGTGGAATTTG	Editotype RpoC1
rpoB_A_F	CAAGTACTTCTTGGCGAGCTCTA	Editotype RpoB
rpoB_A_R	CCTTTATACTGACACTGAGAAGATCG	Editotype RpoB
rpoB_B_F	GTCTCCAATTAATATTTGCGCG	Editotype RpoB
rpoB_B_R	GAAAACCAGTAGGAATATGC	Editotype RpoB
psbZ_F	AGAACATAGCCCTATGAGTTAATACGA	Editotype PsbZ
psbZ_R	GATAAGAGAATTAAGGATACCCACCA	Editotype PsbZ
rps14_F	AAAGTTGCGGAGACAGGATTT	Editotype Rps14
rps14_R	TTATAGGGAGAAGAAGAGGC	Editotype Rps14
accD_A_F	TTCATTTGTAGTAAAAGCGG	Editotype AccD
accD_A_R	AGAGGTAAACATTGATTGGTAGCAT	Editotype AccD
accD_B_F	TATGGGATCCGTAGTAGGCCG	Editotype AccD
accD_B_R	AGAATCTGATCTAACAACAGGGAA	Editotype AccD
psbEF_F	CGTTGGATGAACTGCATTGC	Editotype PsbE et PsbF
psbEF_R	ACAGGAGAACGTCTTTTTCG	Editotype PsbE et PsbF
petL_F	GCCAAAGACGATGAATTGAGTT	Editotype PetL

petL_R	GAGAGGTAATCTATCTATTCCTAGCTCTTAATC	Editotype PetL
rps12_F	CTTGTACAATTCACATTCTTTGGC	Editotype Rps12
rps12_R	CAAGACAGCCAATCCGAAAC	Editotype Rps12
clpP_F	TGAACCGCTACAAGATCAAC	Editotype ClpP
clpP_R	GTAATGATCCATCAACCCGC	Editotype ClpP
rpoA_F	GGAGTTAAACTTCCATTGTCCA	Editotype RpoA
rpoA_R	GCGATGCGAAGAGCTTTACT	Editotype RpoA
rpl23_F	TCCAATTACGTAATCAAATAGTTCAA	Editotype Rpl23
rpl23_R	AATTCCTACTGGATGCACGC	Editotype Rpl23
ndhB_A_F	AGGTCAGATTTGCCTATTCC	Editotype NdhB
ndhB_A_R	TCATTGCTATTACTCAAACAAGCA	Editotype NdhB
ndhB_B_F	TCCGATTTGACCTATGGACG	Editotype NdhB
ndhB_B_R	GTGCTCCTACGGAACCAAGG	Editotype NdhB
ndhB_C_F	TCTCAGAGATAGATAGAGAGGTAGGAATTT	Editotype NdhB
ndhB_C_R	CATTGAATGTACAGAAATGGCTATAA	Editotype NdhB
ndhB_D_F	ATTCTTGAAGCTCAATCTCTCCCCGGAT	Editotype NdhB
ndhB_D_R	TCATGATCTGGCATGTACAG	Editotype NdhB
ndhF_F	GCATTCGCTGCAATAGGTCG	Editotype NdhF
ndhF_R	AAAACCTTCGCCGCATGTGG	Editotype NdhF
ndhD_A_F	CAAGCCTAATTCTATCATAACTCG	Editotype NdhD
ndhD_A_R	ACCTGGTGTATCTTGTC	Editotype NdhD
ndhD_B_F	TCGATCCATTTATAATCTTCGGA	Editotype NdhD
ndhD_B_R	GACTGTGTTGGTTGTAAGAGATGTG	Editotype NdhD
ndhG_F	AGAATTATTGAAATGAGTTC	Editotype NdhG
ndhG_R	TATAATGGATTTGCCTGGAC	Editotype NdhG

### Oligonucléotides de PPE (marquage 5'FAM)

Nom	Séquence (5'→3')	Description
ndhD_PPE_C	TCTTTGGACCTGGTGTATCTTGCTTTTACC	Produit non édité 32bp Produit édité 43bp
clpP_PPE	ATGTTTTTATGTCAGCAACAGAAGCCCAAGC	Produit non édité 33bp Produit édité 48bp
RpoA_PPE	CGAAATAGAAGGAACATGTATTACACGTGC	Produit non édité 36bp Produit édité 43bp

### Coexpression en levure

Nom	Séquence (5'→3')	Description
GFP_EcoRI_R	GGAATTCCTTACTTGTACAGCTCGTCCATG	GFP-stop fusionné à EcoRI en 3'

GFP_NdhD_C_PC R1F	ACGAATGATTTTCCTTGGTTAACAATAATTGTGAGCAAGGGCGAGGA	GFP fusionné à 30nt NdhD-1 forme non éditée en 5'
NdhD_C_HindPC R2F	CCCAAGCTTGGGTCTTTGGACCTGGTGTATCTTGTCTTTACCACGAATGATT TTCCTTGGTTAAC	Reste de NdhD-1 à cloner (60nt final) forme non éditée pour faire une PCR2 HindIII 5'
GFP_NdhD_U_P CR1F	ATGAATGATTTTCCTTGGTTAACAATAATTGTGAGCAAGGGCGAGGA	GFP fusionné à 30nt NdhD-1 forme éditée en 5'
NdhD_U_HindPC R2F	CCCAAGCTTGGGTCTTTGGACCTGGTGTATCTTGTCTTTACCATGAATGATTT TTCCTTGGTTAAC	Reste de NdhD-1 à cloner (60nt final) forme éditée pour faire une PCR2 HindIII en 5'
CRR4 IS Eco-F	GGAATTCATGGCTTTTGCCTCTTCTCGACG	CRR4 interne stop avec EcoRI en 5'
CRR4 IS Hind-R	CCCAAGCTTGGGCTACAATGTACTGGAAACTTCAATGCTAT	CRR4 interne stop avec HindIII en 3'
DYW1 IS Bam-F	CCCGGATCCGGGATGCAAGTTCAAAGCTCTGATGGAGT	DYW1 interne stop avec BamHI en 5'
DYW1 IS Eco-R	GGAATTCCTACCAATAATCTCCACAAGAGCAG	DYW1 interne stop avec EcoRI en 3'
NdhD C Bam-F	CCCGGATCCGGGTCTTTGGACCTGGTGTATCTTGTCTTTACCACGAATGATT TTCCTTGGTTAAC	Reste de NdhD-1 à cloner (60nt final) forme non éditée pour faire une PCR2 BamHI en 5'
NdhD U Bam-F	CCCGGATCCGGGTCTTTGGACCTGGTGTATCTTGTCTTTACCATGAATGATT TTCCTTGGTTAAC	Reste de NdhD-1 à cloner (60nt final) forme éditée pour faire une PCR2 BamHI en 5'
GFP_XhoI_R	CCCCTCGAGGGGTTACTTGTACAGCTCGTCCATG	GFP-stop fusionné à XhoI en 3'
GFP STOP Hind- R	CCCAAGCTTGGGTTACTTGTACAGCTCGTCCATG	GFP-stop fusionné à HindIII en 3'

<b>Coexpression dans <i>E. coli</i></b>		
<b>Nom</b>	<b>Séquence (5'→3')</b>	<b>Description</b>
DYW1Avr- STOP	GGCCTAGGCCCTACTACCAATAATCTCCACAAGAGCAGT	AvrII + DYW1 STOP
CRR4Hind- STOP	CCCAAGCTTGGGCTACTACAATGTACTGGAAACTTCAATGC	HindIII + CRR4 STOP
DYW1Fse- notarg	CGGGCCGCGCATGCAAGTTCAAAGCTCTGATGGAGT	FseI + DYW1 sans la séquence d'adressage
CRR4Eco- notarg	GGAATTCATGGCTTTTGCCTCTTCTCGACG	EcoRI + CRR4 sans la séquence d'adressage
DYW1Eco- notarg	GGAATTCATGCAAGTTCAAAGCTCTGATGGAGT	EcoRI + DYW1 sans la séquence

		d'adressage
NdhD BsrG -123	GGTGTACACCTAATACGACTCACTATAGGGAATTTGGAATACATTTTATTTTTTTT ATTGACA	BsrGI+ promoteur T7
NdhD BsrG -25	GGTGTACACCTAATACGACTCACTATAGGGGACCTGGTGTATCTTGTCTTTACC	BsrGI+ promoteur T7
NdhD Avr +18	GGCCTAGGCCAACCAAGGAAAATCATTCGTG	AvrII + 18pb après le site d'édition
NdhD Avr+150	GGCCTAGGCCAAAGCGTAGGTCGTTAGAAGAAGT	AvrII + 150pb après le site d'édition
DYW1Eco- ATG	GGAATCCATGAGTTTTTGTCTCTTCTCTGTC	EcoRI + DYW1 avec la séquence d'adressage
DYW1Hind -STOP	CCCAAGCTTGGGCTACTACCAATAATCTCCACAAGAGCAGT	HindIII + DYW1 ave le codon STOP

<b>Gène chimérique CRR4-DYW1</b>		
<b>Nom</b>	<b>Séquence (5'→3')</b>	<b>Description</b>
CRR4DYW1 F	GTTTCCAGTACATTGCAAGTCAAAGCTCT	15nt CRR4 sans STOP + 15nt DYW1 sans adressage
CRR4DYW1R	AGAGCTTTGAACTTGCAATGTACTGGAAAC	15nt CRR4 sans STOP + 15nt DYW1 sans adressage
CRR4DYW1Extend_ F	GTTTCCAGTACATTGGCCTACGTAAAGCTG	15ntCRR4 sans STOP + 15ntDYW1 16aa avant le domaine DYW
CRR4DYW1Extend_ R	CAGCTTTACGTAGGCCAATGTACTGGAAAC	15ntCRR4 sans STOP + 15ntDYW1 16aa avant DYW
CRR4DYW1Extend (40)_F	TTGAAGTTTCCAGTACATTGGCCTACGTAAAGCTGAAATC	20ntCRR4 sans STOP + 20ntDYW1 16aa avant DYW
CRR4DYW1Extend (40)_R	GATTTCCAGCTTTACGTAGGCCAATGTACTGGAAACTTCAA	20ntCRR4 sans STOP + 20ntDYW1 16aa avant DYW
DYW1 STOP R	GAAAGCTGGGTCTACCAATAATCTCCACAAGAGCA	Attb2 + DYW1 contenant le codon STOP

<b>Split-YFP</b>		
<b>Nom</b>	<b>Séquence (5'→3')</b>	<b>Description</b>
CRR4_start	GGAGATAGAACCATGCTTGTCTTCAAGTCAACCATG	CRR4 ATG
CRR4_end	TCCACCTCCGGATCMCAATGTACTGGAACTTCAATGCTATCTAC	CRR4 sans STOP
CLB19_start	AAAAAGCAGGCTCCACCATGGGTCTCCTTCCCGTCGTCG	CLB19 ATG
CLB19_end	AGAAAGCTGGGTCAGCATTGAGGAGATCACCAGC	CLB19 sans STOP
FSD2_start	GGAGATAGAACCATGATGAATGTTGCAGTGACAGC	Témoin FSD2 ATG
FSD2_end	TCCACCTCCGGATCCGTCAACCTCAGATACATCGATG	Témoin FSD2 sans STOP
FSD3_start	GGAGATAGAACCATGAGTCTTGTGTTGTGACGAC	Témoin FSD3 ATG

FSD3_end	TCCACCTCCGGATCCAGCGATTGGGATGTTGGGTTC	Témoin FSD3 sans STOP
MinD1_start	GGAGATAGAACCATGGCGTCTCTGAGATTGTTCT	Témoin MIND1 ATG
MinD1_end	TCCACCTCCGGATCCGCCGCCAAAAGAAAGAGAAGAAG	Témoin MIND1 sans STOP

<b>Localisation subcellulaire DYW1</b>		
<b>Nom</b>	<b>Séquence (5'→3')</b>	<b>Description</b>
DYW1_start	GGAGATAGAACCATGAGTTTTTGTCTCTCTGT	DYW1 ATG
DYW1_end	CAAGAAAGCTGGGTCCCAATAATCTCCACAAGAGCAG	DYW1 sans STOP
DYW1_stop	CAAGAAAGCTGGGTCTACCAATAATCTCCACAAGAG	DYW1 STOP
DYW1_preseq	CAAGAAAGCTGGGTCTCGTAAAGTAGTATCTTTTCGTACAG	300bp après l'ATG

<b>Expression dans pETG41K-pDEST15 (ICP-MS, <i>In vitro</i> editing)</b>		
<b>Nom</b>	<b>Séquence (5'→3')</b>	<b>Description</b>
MORF2_ATG_int_a ttb_TEV_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAGAATCTTTATTTTCAGGG CATGTTCTCAATTCGATGTGGAGCTAAC	MORF2 + attb complet + site de clivage TEV
MORF2_STOP_attb R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCATCTTGTGTTTTCTCTGCG	MORF2 + attb complet
DYW1_full_attb_int _TEV_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAGAATCTTTATTTTCAGGG CATGCAAGTTCAAAGCTCTGATGGAGT	DYW1 + attb complet + site de clivage TEV
DYW1_full_attb_S TOP_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTACCAATAATCTCCACAAG AGCAGT	DYW1 + attb complet
CRR4_full_attb_int _TEV_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAGAATCTTTATTTTCAGGG CATGGCTTTTGCCTTCTCTCGACG	CRR4 + attb complet + site de clivage TEV
CRR4_full_attb_ST OP_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTACAATGTACTGGAAACTTC AATGC	CRR4 + attb complet
MORF9_ATG_int_a ttb_TEV_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAGAATCTTTATTTTCAGGG CATGAATCGTTTCTTACTACTCTCTGGT	MORF9 + attb complet + site de clivage TEV
MORF9_STOP_attb R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAAGAGGAATCAGAGGCTG CT	MORF9 + attb complet
NP_416648_attb_T EV-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAGAATCTTTATTTTCAGGG CATGCATCCACGTTTTCAAACC	Cytidine déaminase <i>E. coli</i> + attb complet + site de clivage TEV
NP_416648_attb-R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAAGCGAGAAGCACTCGGT C	Cytidine déaminase <i>E. coli</i> + attb

## IV. Rapport de l'ATP sur le TILLING de DYW1

[http://tilling.fhcrc.org/display\\_results.html](http://tilling.fhcrc.org/display_results.html)

Nom de la lignée à l'ATP	Effet sur la séquence protéique	Zygotie	Stock à l'ABRC
i104D1	F13=	HO	CS86407
i110C4	G112E	HT	CS86873
i112A1	H219Y	HT	CS86982
i129A1	Non-codant	HO	CS87977
i132A3	N83=	HT	CS88125
i135C3	Non-codant	HO	CS88342
i136A4	E143K	HO	CS88412
i136C5	D222N	HT	CS88422
i143E3	Non-codant	HO	CS90703
i144C1	Non-codant	HO	CS90755
i147E7	G132E	HO	CS91040
i149G8	A11T	HO	CS91193
i150C1	Non-codant	HO	CS91197
i150E7	D229N	HT	CS90619
i163E8	Non-codant	HT	CS91308
i163G7	Non-codant	HO	CS91302
i164E6	Non-codant	HT	CS91359
i168H3	A159=	HO	CS91624
i169G3	Non-codant	HT	CS91688
i172B1	R136K	HT	CS91871
i172B3	Non-codant	HT	CS91887
i174E4	L188=	HT	CS91425
i181C5	Non-codant	HT	CS91427
i184G4	E109=	HT	CS93389
i188E6	Non-codant	HT	CS93711
i192C1	E134K	HT	CS94013
i192D2	Non-codant	HT	CS94022
i192E5	Non-codant	HT	CS94048
i194C3	A138T	HT	CS92117
i199H5	E134K	HT	CS93968
i200H6	A159=	HT	CS93010
<b>i201G8</b>	<b>Changement de cadre de lecture</b>	<b>HT</b>	<b>CS94700</b>
i202E7	R61K	HO	CS94754
i202H4	R61K	HT	CS94733
i204A1	F13=	HT	CS94833
i205B8	Non-codant	HT	CS94956
i205H1	T50I	HT	CS94906
i205H5	Non-codant	HT	CS94938
i205H7	R70K	HO	CS94954
i206A7	Non-codant	HT	CS95015
i207E5	Non-codant	HT	CS95060
i207G8	Non-codant	HT	CS95096
i210F3	A32=	HT	CS95324
i214F7	Non-codant	HT	CS95719
i216B2	S5=	HT	CS95820
i216F5	G30R	HO	CS95879
i220A5	P178S	HT	CS96115
i220B2	R70K	HT	CS96092
i220H4	I94=	HO	CS96114

## V. Table des figures

Figure 1 : Edition de l'ARNm <i>coxII</i> chez le trypanosome (d'après Benne et al., 1986).	11
Figure 2 : Edition de l'ARNm <i>apoB</i> chez l'Homme (extrait de web-books.com).	12
Figure 3 : Schéma simplifié de l'édition par insertion de nucléotides dans les mitochondries des kinétoplastes (Extrait de Knoop 2011).	14
Figure 4 : Positionnement des sites d'édition dans le génome chloroplastique d' <i>Arabidopsis thaliana</i> .	18
Figure 5 : Edition de l'ARN par conversion de nucléotides de cytidine en uridine (Extrait de Takenaka et al., 2008).	19
Figure 6 : Distribution phylogénétique de l'édition chez les plantes (adaptée de Knoop et al., 2011).	21
Figure 7 : Mécanisme de l'édition du site <i>apoB</i> par APOBEC-1 (Adapté de Wedekind et al., 2003).	24
Figure 8 : Modélisation complexe d'édition dans les plastes (adaptée de Okuda et al., 2006).	25
Figure 9 : Structure des motifs PPR (adaptée de Ringel et al., 2011; Howard et al., 2012).	29
Figure 10 : Fonctions connues des protéines PPR (Extrait de Andrès et al., 2006).	29
Figure 11 : Structure et expansion des protéines PPR chez les êtres vivants (adaptée de Lurin et al., 2004).	30
Figure 12 : Cladogramme de similarité entre les différentes protéines MORF (Extrait de Takenaka et al., 2012).	33
Figure 13 : Des domaines spécifiques de RIP1 et RARE1 sont nécessaires à leur interaction <i>in vivo</i> (Extrait de Bentolila et al., 2012).	37
Figure 14: Structure tridimensionnelle des motifs PUF et TALE (Extrait de Filipovska et al., 2012).	38
Figure 15 : Modèle structural montrant l'interaction entre PPR10 et <i>atpH</i> (Extrait de Barkan et al., 2012).	39
Figure 16 : Séquence consensus des domaines E et E+ (Lurin et al., 2004).	41
Figure 17 : Alignement entre une partie du consensus du domaine DYW et la séquence de cytidines déaminases connues (Extrait de Salone et al., 2007).	42
Figure 18 : Structure de la protéine AtDYW1.	44
Figure 19 : Hypothèse de ma thèse : Interaction PPR EE+ / DYW1.	45
Figure 20 : Alignement de la séquence de DYW1 avec le consensus du domaine DYW (Lurin et al., 2004).	47
Figure 21 : Alignement des domaines DYW de DYW1 et de PPR impliquées dans l'édition des ARN mitochondriaux et chloroplastiques (Lurin et al., 2004).	47
Figure 22 : Localisation subcellulaire de DYW1 <i>in planta</i> .	49
Figure 23 : Identification de la protéine DYW1 dans différents extraits cellulaires (thèse de Véronique Salone).	50
Figure 24 : Ensemble des mutations EMS affectant la séquence de la protéine DYW1.	52
Figure 25 : Changement de cadre de lecture dans le gène DYW1 ( <i>dyw1-1</i> ).	52
Figure 26 : Editotype chloroplastique du mutant <i>dyw1-1</i> .	55



Figure 27 : Edition du site <i>ndhD-1</i> avant ( <i>dyw1-1</i> ) et après complémentation ( <i>dyw1-1+DYW1</i> ).....	56
Figure 28 : Observation de la fluorescence chlorophyllienne chez le tabac (Shikanai et al., 1998). .....	59
Figure 29 : Comparaison de l'activité NDH chez <i>dyw1-1</i> , <i>crr4-3</i> (Kotera et al., 2005), <i>dyw1-1+DYW1</i> et Col-0. ....	60
Figure 30 : Etude de l'interaction entre CRR4 et DYW1 par split-YFP. ....	64
Figure 31 : Construction effectuée pour l'étude de PPR en TAP-tag. ....	66
Figure 32 : Résultat de la purification TAP-tag effectuée sur DYW1 (Geert de Jaeger, Gand).....	66
Figure 33 : Résultat de la purification TAP-tag effectuée sur CRR4 (Geert de Jaeger, Gand).....	67
Figure 34 : Résultat de la purification TAP-tag effectuée sur CLB19 (Geert de Jaeger, Gand).....	68
Figure 35 : Alignement des domaines E et DYW de CRR22 avec ceux de CRR4 et DYW1. ....	69
Figure 36 : Complémentation du double mutant <i>crr4-3xdyw1-1</i> avec une fusion de CRR4 et DYW1. ....	70
Figure 37 : Domaines E et E+ d'orthologues potentiels de AtCRR4. ....	71
Figure 38 : Reconstitution schématique de l'apparition du « complexe » d'édition de <i>ndhD-1</i> chez les Angiospermes.....	73
Figure 39 : Conservation des acides aminés affectés par les mutations EMS dans le domaine E.....	75
Figure 40 : Conservation des acides aminés affectés par les mutations EMS dans le domaine DYW. ....	75
Figure 41 : Etude de l'édition au site <i>ndhD-1</i> dans les lignées EMS affectées dans le gène DYW1.....	77
Figure 42 : Variation naturelle des domaines DYW dans 500 protéines.....	78
Figure 43 : Représentation schématique des délétions effectuées dans DYW1. ....	79
Figure 44 : Etude de l'édition dans la lignée <i>dyw1-1</i> complétement avec des versions mutées de DYW1. ....	80
Figure 45 : Mutants de la fusion CRR4-DYW1 produits pour compléter <i>crr4-3xdyw1-1</i> . ....	83
Figure 46 : Vecteur d'expression pETG41K. ....	84
Figure 47 : Vecteur d'expression pDEST15. ....	86
Figure 48 : PPE effectuée sur les produits d'édition <i>in vitro</i> . ....	93

## VI. Table des tableaux

<b>Tableau 1</b> : Edition de l'ARN dans différents organismes suivant leur date de découverte (d'après Knoop et al., 2011).	13
<b>Tableau 2</b> : Sites d'édition présents dans les transcrits mitochondriaux d' <i>Arabidopsis thaliana</i> (adapté de Bentolila et al., 2008).	17
<b>Tableau 3</b> : Consensus des sites d'édition nécessitant le même élément <i>trans</i> (adapté de Hammani et al., 2009).	27
<b>Tableau 4</b> : Impact des mutations <i>morf2-1</i> et <i>morf9-1</i> sur l'éditotype d' <i>Arabidopsis thaliana</i> (adapté de Takenaka et al., 2012).	34
<b>Tableau 5</b> : Interaction des protéines MORF en homo- et hétérodimères (extrait de Takenaka et al., 2012).	35
<b>Tableau 6</b> : Interactions entre protéines MORF et protéines PPR (Extrait de Takenaka et al., 2012).	36
<b>Tableau 7</b> : Code d'interaction des protéines PUF, TALE et PPR (adapté de Filipovska et al., 2012; Barkan et al., 2012).	39
<b>Tableau 8</b> : PPR-DYW (publiées) impliquées dans l'édition chez les plantes terrestres.	42
<b>Tableau 9</b> : PPR-E/EE+ (publiées) impliquées dans l'édition chez les plantes terrestres.	43
<b>Tableau 10</b> : Pourcentage d'identité des domaines DYW de DYW1 et de PPR impliquées dans l'édition.	48
<b>Tableau 11</b> : Prédiction de l'adressage de DYW1 par les programmes Predotar et TargetP.	48
<b>Tableau 12</b> : Quantification du taux d'édition dans chaque échantillon analysé par PPE.	57
<b>Tableau 14</b> : Constructions testées par split-YFP.	65
<b>Tableau 15</b> : Quantification de l'édition pour les 12 complémentations.	82
<b>Tableau 16</b> : Résultat de l'ICP-MS.	88
<b>Tableau 17</b> : Principe de l'expression <i>in vivo</i> en système levure.	90

# Résumé/Summary

## I. Résumé (français)

Dans les organites des plantes, l'édition de l'ARN consiste le plus souvent en une conversion de cytidines en uridine à des sites spécifiques de l'ARNm. Trente-quatre sites d'édition ont été découverts dans les transcrits chloroplastiques d'*Arabidopsis thaliana* et plus de 500 dans les transcrits mitochondriaux. Depuis 2005, beaucoup de facteurs d'édition ont été découverts. La majorité de ces protéines appartiennent à la famille des «PentatrigoPeptide Repeat» (PPR). Parmi ces PPR, certaines contiennent un domaine DYW possédant de faibles similarités avec les cytidines désaminases (CDA), alors que d'autres en sont dénuées, générant un doute sur le fait que le domaine DYW ait une fonction dans l'édition.

Le gène At1g47580 (*DYW1*) nous a intéressés car il code une protéine unique chez *Arabidopsis thaliana* contenant « seulement » un domaine DYW. L'hypothèse initialement avancée propose que *DYW1* puisse interagir avec les PPR ne contenant pas de domaine DYW, de sorte à former un hétérodimère, capable d'éditer spécifiquement un site. En accord avec cette hypothèse, nous avons trouvé que *DYW1* agissait sur le même site d'édition que *CRR4*, une PPR sans domaine DYW, et que ces protéines interagissaient *in vivo*. De plus, nous avons montré que *DYW1* remplaçait les parties manquantes de *CRR4* pour l'édition.

Néanmoins, *DYW1* est extrêmement spécifique à *CRR4*. Nous avons posé l'hypothèse que d'autres facteurs devaient exister pour les autres PPR ne possédant pas de domaine DYW.

Pour obtenir plus d'information sur la fonction du domaine DYW, des mutations ont été introduites dans *DYW1*. Nous avons montré que les résidus très conservés formant la signature CDA dans les protéines DYW étaient essentiels à l'édition de l'ARN et qu'ils étaient nécessaires à l'interaction avec les ions zinc, caractéristique connue des CDA. Les données collectées sont en accord avec l'hypothèse que le domaine DYW contient une activité CDA. Néanmoins, aucune activité CDA n'a pu être découverte *in vitro* ou après expression dans des bactéries ou des levures. Il est vraisemblable qu'au moins un cofacteur doit encore être identifié.

## **II. Summary (english)**

In plant organelles, RNA editing mostly takes the form of conversions of cytidines to uridines at specific sites in mRNAs. Thirty-four editing sites have been found in *Arabidopsis thaliana* chloroplast transcripts and more than 500 sites in mitochondrial transcripts. Since 2005, lots of proteins have been found to act as RNA editing factors. Most of these proteins belong to the Pentatricopeptide Repeat (PPR) family. Amongst these PPR proteins, some contain a DYW domain with weak similarity to cytidine deaminases (CDA), whilst others lack such a domain, creating doubts about whether this domain could have CDA activity and whether it is required for editing.

The gene At1g47580 (named *DYW1*) encodes a protein in *Arabidopsis thaliana* that contains “only” a DYW domain. Our initial hypothesis was that the DYW1 protein might interact with PPR proteins that lack a DYW domain, in order to form a heterodimer, able to perform site-specific editing. In accordance with this hypothesis, we discovered that DYW1 is involved in editing the same site as CRR4, a PPR lacking a DYW domain, and that these two proteins interact together *in vivo*. Moreover, we showed that DYW1 replaces all the missing parts of CRR4 for editing.

However, DYW1 is highly specific for CRR4, so other partners need to be hypothesized for the dozens of other DYW-lacking editing factors if this hypothesis is to be generalized.

In an attempt to learn more about the function of the DYW domain, point mutations were introduced into DYW1. The highly conserved residues making up the CDA signature in DYW proteins were found to be essential for RNA editing and are also required for zinc binding, which is a known characteristic of CDAs. All the data so far are consistent with the DYW domain being (part of) a CDA activity; nevertheless, no CDA activity could be detected *in vitro*, or following expression in yeast or bacteria. It is likely that at least one required cofactor remains to be identified.