Biogénèse du chloroplaste : Voies d'import alternatives

Imen Bouchnak

▶ **To cite this version:**

Imen Bouchnak. Biogénèse du chloroplaste : Voies d'import alternatives. Biologie végétale. Université Grenoble Alpes, 2018. Français. NNT : 2018GREAV074 . tel-02954347

HAL Id: tel-02954347
https://theses.hal.science/tel-02954347

Submitted on 1 Oct 2020

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

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

Communauté
UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTE UNIVERSITE GRENOBLE ALPES

Spécialité : **Biologie Végétale**

Arrêté ministériel : 25 mai 2016

Présentée par

Imen BOUCHNAK

Thèse dirigée par **Norbert ROLLAND, DR1, CNRS**
et codirigée par **Marcel KUNTZ, DR2, CNRS**

préparée au sein du **Laboratoire de Physiologie Cellulaire et Végétale**

dans **l'École Doctorale Chimie et Sciences du Vivant**

Biogénèse du chloroplaste : voies d'import alternatives

Thèse soutenue publiquement le **01 Octobre 2018**,

devant le jury composé de :

Mme Christelle BRETON
Professeur à l'Université Grenoble Alpes, Président
Mme Anne-Marie DUCHENE
Professeur à l'Université de Strasbourg, Rapporteur
M. Yves CHOQUET
Directeur de recherche CNRS, Université Pierre et Marie Curie, Rapporteur
M. Thierry GAUDE
Directeur de recherche CNRS, Ecole Normale Supérieure de Lyon, Examinateur
M. Marcel KUNTZ
Directeur de recherche CNRS, Université Grenoble Alpes, Invité
M. Norbert ROLLAND
Directeur de recherche CNRS, Université Grenoble Alpes, Directeur de thèse

J'adresse mes plus vifs remerciements à Madame Christelle Breton, Professeur à l'Université Grenoble Alpes, Madame Anne-Marie Duchêne, Professeur à l'Université de Strasbourg, Monsieur Yves Choquet, Directeur de Recherche au CNRS, ainsi qu'à Monsieur Thierry Gaude, Directeur de Recherche au CNRS, pour l'honneur qu'ils me font en acceptant de juger cette thèse.

Je remercie tout particulièrement Monsieur Norbert Rolland pour m'avoir témoigné sa confiance en m'accueillant dans son laboratoire de Physiologie Cellulaire et Végétale. Je le remercie également pour avoir dirigé ces recherches. Je lui exprime toute ma reconnaissance pour sa disponibilité, ses conseils, ses discussions, ses encouragements, son aide et son soutien tout au long de ces trois années.

J'exprime aussi toute ma gratitude à Monsieur Marcel Kuntz pour avoir co-encadré ce travail. Je lui exprime ici toute ma reconnaissance pour sa disponibilité et son aide.
J'associe à ces remerciements Monsieur Eric Maréchal qui m'a renouvelé cette confiance après avoir pris la direction de ce laboratoire.

Je tiens à attribuer dans mes remerciements une place privilégiée à Lucas Moyet et Daniel Salvi pour leur contribution à ce travail, pour leurs précieux conseils techniques et pour leur bonne humeur. Je leur exprime ici toute ma sympathie.

Je tiens également à remercier Madame Marianne Tardif, Chercheur CEA, ainsi que Madame Sabine Brugière, Ingénieur CEA à l'équipe Etude de la Dynamique des Protéomes (EDyP) pour leur collaboration et leur participation à cette étude.

J'exprime toute ma sympathie à tous les membres du laboratoire de Physiologie Cellulaire et Végétale.

Je remercie tout particulièrement ma famille et mes amis qui m'ont soutenue au cours de ces années.

A mes tès chers parents…

Sommaire

Introduction-données bibliographiques

Importation des protéines dans les chloroplastes: entre voie d'import classique et voies d'import alternatives **1**

Introduction 1

I. Chloroplaste : origine, structure et fonctions 2

1. Origine et évolution du chloroplaste 2
2. Structure du chloroplaste 3

a) L'enveloppe 3

b) Le stroma 4

c) Les thylakoïdes 4

3. Principales fonctions du chloroplaste (Rolland et *al*., 2018) 5

II. Voie d'import classique des protéines dans le chloroplaste : la machinerie TOC/TIC 5

1. Les peptides de transit 6
2. Le complexe TOC 7
3. Le complexe TIC 9

a) Le modèle classique du système d'import TIC 9

b) Le modèle complexe1 MDa du système d'import TIC 11

4. Le mécanisme général d'import des protéines canoniques 12

a) Protéines adressées à la membrane interne, l'espace intermembranaire et le stroma 12

b) Protéines destinées vers les thylakoïdes 13

c) Protéines adressées à la membrane externe 15

III. Voies d'imports alternatives des protéines dans le chloroplaste 16

1. Protéines non-canoniques 17

a) La protéine ceQORH 17

b) La protéine TIC32 21

c) La protéine NQR 22

d) La protéine FSD1 22

2. Système endomembranaire et transport vésiculaire 23
3. Le système d'import PRAT 25
4. Le système d'import Sec2 25
5. Un système d'import Tat-like ? 27

IV. Protéome de l'enveloppe du chloroplaste 28

1. Protéome du chloroplaste 28
2. Protéome de l'enveloppe 29
3. Principales bases de données 30
4. AT-CHLORO database (Salvi et *al*., 2018) 31

V. Objectifs de la thèse 32

Matériel et méthodes **35**

I. Préparation de fractions subcellulaires d'*Arabidopsis thaliana* 36
1. Choix du matériel végétal 36
2. Préparation des chloroplastes intacts d'*Arabidopsis thaliana* 36
3. Purification de l'enveloppe des chloroplastes 37

II. Techniques d'étude des protéines 37
1. Extraction des protéines totales de feuilles d'*Arabidopsis thaliana* 37
2. Dosage des protéines 38
3. Analyse des protéines par électrophorèse sur gel de polyacrylamide 38
a) Séparation par électrophorèse en conditions dénaturantes 38
b) Coloration des protéines au bleu de Coomassie 39
4. Transfert des protéines sur membrane de nitrocellulose et immunomarquage 39
a) Transfert des protéines 39
b) Fixation des anticorps 39
c) Révélation par mesure de la chimiluminescence 40

III. Techniques d'étude des acides nucléiques 40
1. Extraction de l'ADN à partir d'un gel d'agarose 40
2. Synthèse d'oligonucléotides 40
3. Amplification *in vitro* de fragments d'ADN 41
4. Séquençage de l'ADN 41
5. Digestion d'ADN par des endonucléases 41

IV. Techniques d'expression procaryote 41
1. Construction de vecteurs d'expression des protéines recombinantes 41
2. Préparation de l'ADN plasmidique 42
a) Minipréparation d'ADN plasmidique 42
b) Midipréparation d'ADN plasmidique 42

3. Préparation et transformation des bactéries compétentes 43
4. Conditions d'expression 43
5. Vérification de la solubilité des protéines 44
6. Technique de purification des protéines recombinantes contenant une extension *C*-terminale poly-histidine 44
a) Préparation des protéines 44
b) Préparation de la colonne 45
c) Chromatographie d'affinité 45
V. Techniques d'identification des récepteurs à la surface des chloroplastes 45
1. Pontage chimique de protéines recombinantes avec les protéines de l'enveloppe de chloroplastes 45
a) Préparation de l'enveloppe de chloroplastes 46
b) Interaction protéine recombinante-enveloppe de chloroplaste 46
c) Réaction de pontage chimique 46
2. Technique de purification des complexes protéiques issus de pontage chimique contenant une étiquette GFP 47
a) Solubilisation des complexes protéiques et fixation des billes magnétiques 47
b) Préparation de la colonne 47
c) Chromatographie d'affinité 47
VI. Expression stable de protéines chez *Arabidopsis thaliana* 48
1. Le système *Agrobacterium tumefaciens* 48
2. Construction des vecteurs d'expression des protéines fusionnées à la GFP 49
3. Techniques de transformation des plantes et sélection des transformants 49
a) Transformation d'*Agrobacterium tumefaciens* 49
b) Transformation des plantes par trempage d'inflorescences 50
4. Observation par microscope confocale des plantes transformées 51

Chapitre I

Pontage chimique et approche protéomique pour identifier les récepteurs d'import alternatif a la surface des chloroplastes **52**

I. Production de protéines de chloroplaste non canoniques à utiliser comme appâts dans les études d'interaction 53

1. Validation de la séquence primaire de la protéine FSD1 prédite par l'analyse du génome d'*Arabidopsis thaliana* 54

2. Construction de vecteurs d'expression des protéines recombinantes 55

3. Surexpression des protéines recombinantes 55

4. Vérification de la solubilité des protéines recombinantes produites 56

5. Purification des protéines recombinantes 57

II. Pontage chimique : vers l'identification de récepteurs d'import alternatif 58

1. Optimisation de la méthode de pontage chimique 59

a) Réaction de pontage en présence de l'AMAS 59

b) Réaction de pontage en présence d'une nouvelle molécule de pontage (BS^3) 61

2. Purification par affinité des complexes d'import 62

3. Identification des peptides spécifiques des récepteurs d'import par spectrométrie de masse 63

4. Identification de récepteurs candidats 65

III. Validation de la nature des récepteurs putatifs 68

1. Isolement de mutants d'insertions d'*Arabidopsis thaliana* déficients dans l'expression des candidats récepteurs 69

2. Expression *in planta* des versions GFP-étiquetées des protéines chloroplastiques non canoniques dans les mutants des candidats récepteurs d'import 71

a) Production de plantes transgéniques d'Arabidopsis surexprimant les protéines non canoniques 71

b) Sélection de transformants 72

3. Analyse de l'impact des mutations sur la localisation subcellulaire des protéines non canoniques par microscope confocal 74

a) Mutants homozygotes 74

b) Mutants hétérozygotes (avec phénotype embryon létal) 84

IV. Conclusion et perspectives 86

Chapitre II

Bio-analyse du protéome de l'enveloppe du chloroplaste

Article 1 **88**

Chapitre III

Contrôle de l'adressage des protéines au chloroplaste

Article 2 **138**

Conclusion générale et perspectives **175**

Références bibliographiques **186**

Annexes **197**

Liste des abréviations

ALB3: protéine Albino3

ATCHLORO: *Arabidopsis thaliana* chloroplast proteome database

ATP: adénosine triphosphate

BFA: bréfeldine A

BiFC: Bimolecular fluorescence complementation

CAH1: anhydrase carbonique 1

Cam5: calmoduline 5

CCV: Clathrin-Coated Vesicles

ceQORH: chloroplast quinone oxidoreductase homologue

ClpD: une protéine chaperone du complex Clp

DDM: dodécylmaltoside

DTT: dithiothreitol

EDYP: Exploring the Dynamics of Proteomes

FSD1: superoxyde dismutase à Fer 1

FTSH12: protéine Filamentous temperature sensitive H

GFP: Green Fluorescent Protein

GTP: guanosine triphosphate

Hcf106: protéine haute fluorescence chlorophyllienne 106

IPTG: Isopropyl β-D-1-thiogalactopyranoside

KO: Knock-out

LHC: Light-Harvesting Chlorophyll

MS: spectrométrie de masse

Mut2: protéine ceQORH mutée n'interagissant plus avec Cam5

NPP1: Nucléotide Pyrophosphatase Phosphodiestérase 1 de riz

NQR: NADPH Quinone Réductase

pFd: précurseur de la protéine ferrédoxine

plprot: plastid protein database

pOE33 : précurseur de la sous-unité du complexe de l'évolution de l'oxygène

PPDB: Plant Proteomics Database

PRAT: putative preprotein amino acid transporters

RE: réticulum endoplasmique

Rubisco: ribulose-1,5-bisphosphate carboxylase oxygenase

Sec: Secretory dependent pathway

SPP: stomal processing protease

SSU: la petite sous-unité de la Rubisco

SUBA: SUBcellular localisation database for Arabidopsis proteins

TAT: Twin Arginine Translocase

TIC: Translocon at the Inner Chloroplast envelope

TIM: Translocon at Mitochondrial Inner envelope membrane

TOC: Translocon at the Outer Chloroplast envelope

TOM: Translocon at Mitochondrial Outer envelope membrane

TPP: Thylakoïdal Processing Peptidase

TPT: triose phosphate transporter

WT: Wild Type

XIC: extracted Ion Chromatograms

αAmy3: α-amylase 3 de riz

Introduction-données bibliographiques

Importation des protéines dans les chloroplastes: entre voie d'import classique et voies d'import alternatives

Introduction

Les végétaux sont des organismes autotrophes, capables, grâce à la photosynthèse, de synthétiser leur propre matière organique à partir de substances minérales qu'ils puisent dans l'atmosphère ou dans le sol. La réaction de photosynthèse est réalisée dans les chloroplastes, qui sont des organites semi-autonomes, présents dans les cellules des tissus foliaires. Ces organites sont le fruit d'une endosymbiose, survenue entre une cellule eucaryote et une cyanobactérie [1-3]. Ainsi, l'évolution endosymbiotique a entrainé le transfert des gènes codant pour la majorité des protéines plastidiales vers le génome nucléaire [4, 5]. En conséquence, la plupart des protéines chloroplastiques sont aujourd'hui codées par le noyau, synthétisées dans le cytosol sous forme de précurseurs dotés d'une une extension *N*-terminale clivable (le "peptide de transit") et ensuite importées sans les chloroplastes via le système TOC/TIC (Translocons localisés au niveau des membranes externe et interne de l'enveloppe des chloroplastes) [4, 5]. Jusqu'à récemment, toutes les protéines destinées aux compartiments chloroplastiques internes étaient censées posséder une séquence d'adressage *N*-terminale clivable et engager la machinerie d'import général TOC/TIC. Cependant, des études récentes reposant sur des approches protéomiques ont a révélé l'existence de plusieurs protéines chloroplastiques dépourvues de la séquence additionnelle clivable. La première évidence de telles protéines dites non canoniques a été fournie par notre équipe, étudiant le protéome de l'enveloppe du chloroplaste d'Arabidopsis en collaboration avec la plateforme de protéomique «EDyP», qui a conduit à l'identification d'une protéine quinone oxidoréductase homologue nommée «ceQORH» [6, 7]. Plus tard, une deuxième protéine non canonique, désignée Tic32, a été identifiée par une équipe allemande [8]. Bien que dépourvues de peptide de transit clivable, il s'est avéré que ces protéines sont associées à la membrane interne de l'enveloppe du chloroplaste [7, 8]. D'autre part, il a été également montré que l'import de ces protéines dans le chloroplaste n'est pas médiée par la machinerie de translocation générale TOC/TIC [7-9]. De plus, il s'est avéré que la protéine ceQORH a la particularité d'être multilocalisée : Dans les cellules du parenchyme de la feuille, la protéine est localisée dans les chloroplastes, alors que dans les cellules de l'épiderme, en particulier dans les cellules de garde, sa localisation est cytosolique [9]. Cependant, les mécanismes moléculaires qui contrôlent la localisation sub-cellulaire de telles protéines chloroplastiques contenant un peptide de transit non canonique demeurent encore inconnus.

Dans la suite de ce chapitre, je présenterai dans un premier temps le chloroplaste, son origine ainsi que sa structure et ses principales fonctions. Par la suite, je décrirai la voie d'import classique des protéines dans les chloroplastes. Je présenterai ensuite en détail les différentes protéines chloroplastiques non canoniques et citerai les différents mécanismes d'import alternatifs potentiels précédemment proposés. Dans une dernière partie, je décrirai le protéome du chloroplaste avec un focus sur le répertoire protéique de l'enveloppe du chloroplaste.

I. Chloroplaste : origine, structure et fonctions

1. Origine et évolution du chloroplaste

Les chloroplastes sont des organites semi-autonomes présents à l'intérieur des cellules végétales au niveau des tissus foliaires. Ces plastes, comme les mitochondries, sont le fruit d'une endosymbiose, survenue il y a 1,6 milliards d'années environ, entre une cellule eucaryote et une cyanobactérie [1-3]. Des études évolutives, ont montré l'existence de deux événements endosymbiotiques successifs conduisant à l'apparition des chloroplastes (figure 1). Le premier événement endosymbiotique a eu lieu entre un organisme de type α-protéobactérium (l'ancêtre des mitochondries) et une cellule hôte [10]. Cette cellule eucaryote a ensuite acquis un endosymbionte supplémentaire, une cyanobactérie, l'ancêtre des chloroplastes. Par conséquent, au cours de l'évolution, bien que certains traits cyanobactériens soient conservés, la majorité des gènes initialement présents dans le génome chloroplastique a été transférée vers le noyau de la cellule hôte [4, 5]. En effet, les analyses les plus récentes montent une divergence nette entre le nombre de protéines actuellement codées par le génome plastidial (~ 100-150 gènes) et le nombre total de protéines que contiennent les plastes (~ 3000 protéines) [4]. Ainsi, la grande majorité des protéines chloroplastiques de la plante (qui assurent le fonctionnement des chloroplastes) est aujourd'hui codée par le génome nucléaire et doit être importée dans le plaste après traduction dans le cytosol sous forme de précurseurs dotés d'une séquence d'adressage clivable en *N*-terminal [4, 5].

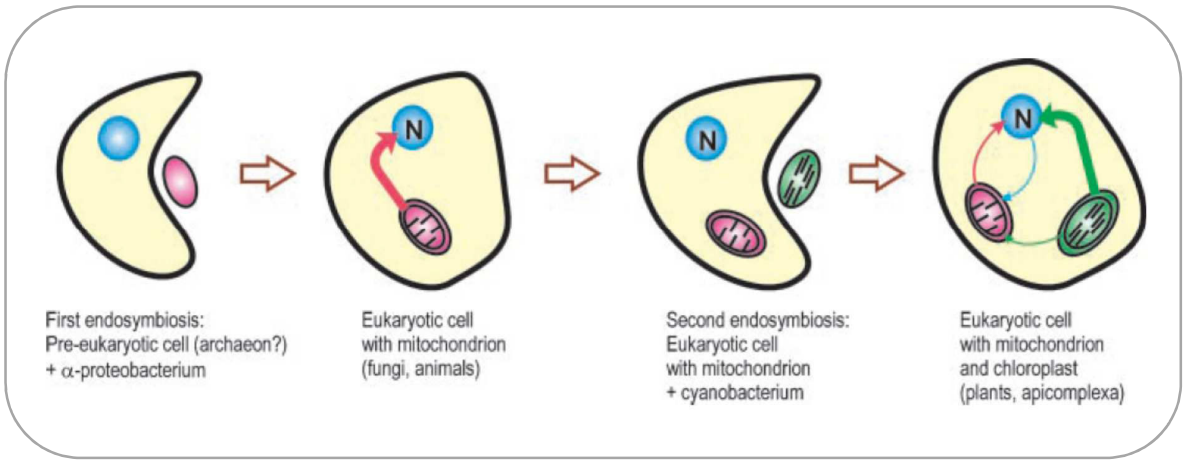


Figure1. Représentation schématique de deux évènements endosymbiotiques conduisant à l'apparition du chloroplaste et transfert intracellulaire des gènes au cours de l'évolution (d'après Bock et *al.*, 2008).

2. Structure du chloroplaste

Les chloroplastes, distribués dans le cytoplasme de la cellule végétale, mesurent généralement entre 4 et 10 micromètres de long. Ils sont limités par une double membrane, ou enveloppe, qui contrôle les échanges avec les autres compartiments de la cellule végétale [11]. Le compartiment ainsi délimité contient une phase aqueuse, le stroma, dans laquelle baigne un système membranaire complexe, les thylakoïdes (figure 2).

a) **L'enveloppe :** composée de deux membranes (membrane externe et membrane interne) séparées par un espace inter-membranaire (figure 2). L'enveloppe joue un rôle central dans le maintien de la communication du chloroplaste avec le milieu cellulaire et le cytosol [11]. En effet, la membrane externe est perméable aux molécules de faible taille (ions, métabolites, précurseurs hormonaux…) du fait de la présence de plusieurs porines, canaux ioniques et transporteurs localisés dans la membrane externe. Cette membrane contient également, une partie du complexe d'import classique des protéines canoniques du chloroplaste, le complexe TOC, appelé également «Translocon at the Outer Chloroplast envelope» [12]. De la même manière, la membrane interne de l'enveloppe régule les échanges entre le cytoplasme et le stroma par l'intermédiaire de nombreux transporteurs comme, par exemple, le transporteur de phosphate (triose phosphate transporter (TPT)) qui représente la protéine majeure de l'enveloppe des chloroplastes, ou le complexe TIC pour

«Translocon at the Inner Chloroplast envelope» [12] qui constitue la seconde partie du système d'import des protéines. Il est important de noter que l'enveloppe (l'ensemble de deux membranes externe et interne) représente seulement 1 à 2% des protéines du chloroplaste entier [13].

b) **Le stroma :** le composant majeur du chloroplaste (~ 50%). Cette phase aqueuse du plaste constitue un milieu réactionnel extrêmement complexe, principalement composé de protéines solubles et enzymes impliquées dans les différentes voies métaboliques spécifiques des chloroplastes (assimilation du carbone, métabolisme des acides aminés, vitamines, lipides…). La Rubisco (ribulose-1,5-bisphosphate carboxylase oxygénase) est la protéine la plus abondante du stroma, qui peut représenter jusqu'à 50% des protéines solubles.

c) **Les thylakoïdes :** forment un système membranaire interne hautement organisé formé de vésicules compressées plates formant les grana et de structures inter-granaires, beaucoup plus longues, reliant les différents grana, appelées lamelles (figure 2). Ce compartiment sub-plastidial est le siège de la photosynthèse oxygénée. Toutes les vésicules thylakoïdiennes sont connectées entre elles par l'intermédiaire de leur lumière interne, formant ainsi un autre compartiment soluble appelé lumen des thylakoïdes (figure2). Les thylakoïdes, comme le stroma, représente ~ 50% du volume plastidial dont les LHCP (Light-Harvesting Complex Proteins) sont les protéines membranaires les plus abondantes.

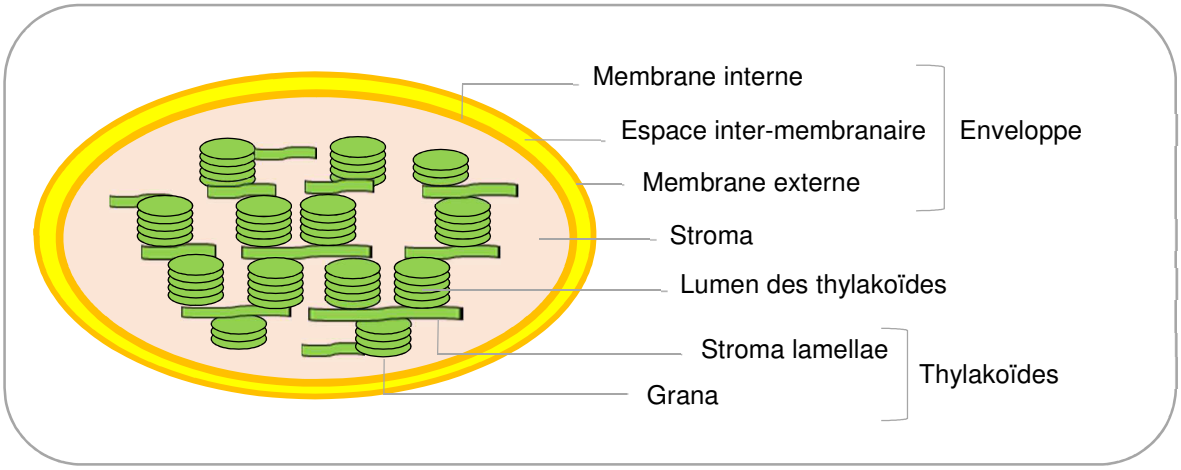


Figure2. Représentation schématique de l'organisation du chloroplaste (d'après Bouchnak et *al*., 2018).

1. Principales fonctions du chloroplaste (voir Rolland et *al*., 2018)

Les chloroplastes sont de véritables usines énergétiques de la cellule végétale. En effet, ces organites photosynthétiques sont responsables de la synthèse de la totalité de la biomasse de la plante et constituent une «pompe à CO_2», agissant ainsi sur l'environnement et le climat terrestre. Les chloroplastes sont aussi le siège de la synthèse de nombreux métabolites secondaires incluant des acides aminés, des pigments, des vitamines, des précurseurs des hormones et des acides gras et des lipides [14]. Les chloroplastes jouent aussi un rôle primordial dans le contrôle de différentes voies métaboliques au sein de la cellule végétale, du fait de la localisation de l'enveloppe à l'interface entre le stroma des chloroplastes et le cytosol de la cellule, favorisant donc de nombreux échanges entre ces deux compartiments [11]. L'enveloppe est ainsi un compartiment sub-plastidial extrêmement complexe et dynamique [11]. Néanmoins, les systèmes de transport associés à ce système membranaire sont encore peu étudiés et voire inconnus. Ainsi, pour une meilleure compréhension du fonctionnement intégré de la cellule végétale, il est nécessaire de décrypter les mécanismes moléculaires qui contrôlent la dynamique du chloroplaste.

II. Voie d'import classique des protéines dans le chloroplaste : la machinerie TOC/TIC

Il est admis que la plupart des protéines chloroplastiques (~ 95%) sont synthétisées dans le cytosol à partir du génome nucléaire [4]. Leur transit, du cytosol vers leurs compartiments sub-plastidiaux (espace intermembranaire, membrane interne, stroma, membrane des thylakoïdes et lumen), est une étape clé. En effet, c'est à ce stade que les protéines chaperonnes interviennent afin d'assurer une orientation spécifique des protéines vers leurs destinations finales en les maintenant dans une conformation dénaturée, condition requise par le système d'import TOC/TIC. Ainsi, les protéines chloroplastiques sont synthétisées sous forme de précurseurs dotés d'une séquence d'adressage, appelée aussi «peptide de transit», qui porte l'information nécessaire à l'adressage spécifique des protéines [15-17].

1. Les peptides de transit

Les peptides de transit chloroplastiques sont nécessaires et suffisants pour permettre un adressage spécifique des précurseurs vers leurs compartiments sub-plastidiaux [18]. Ces pré-séquences additionnelles sont localisées à l'extrémité *N*-terminale et sont requises pour la reconnaissance et la prise en charge des précurseurs par les récepteurs de la machinerie d'import TOC/TIC [16, 17]. En effet, les peptides de transit sont constitués de sites de liaison spécifique favorisant l'interaction des précurseurs chloroplastiques avec les chaperonnes 14-3-3 et Hsp70, guidant ainsi le complexe jusqu'à la surface des chloroplastes [19-21]. La première évidence d'une telle séquence d'adressage chloroplastique a été prouvée par une étude menée sur l'une des protéines chloroplastiques les plus connues, la petite sous-unité de la Rubisco (SSU) [22]. Il a été démontré que la SSU synthétisée dans le cytosol possède une taille supérieure à celle qui est présente dans le stroma du chloroplaste [16, 17, 22]. Ainsi, l'hypothèse émise, était l'existence d'un lien entre la séquence d'acides aminés supplémentaire et le transfert de la protéine dans le chloroplaste. De la même manière, d'autres études similaires ont été reproduites sur de nombreuses autres protéines chloroplastiques, et il s'est avéré que la majorité de ces protéines sont synthétisées avec des extensions *N*-terminales sous forme de «pré-protéines» et que ces pré-séquences sont nécessaires à la translocation des protéines à l'intérieur du chloroplaste. Les peptides de transit, contiennent généralement entre 20 à 120 acides aminés et possèdent une charge nette positive. Bien que leurs séquences primaires soient différentes d'une protéine à une autre, trois régions distinctes du peptide de transit présentent des caractéristiques physico-chimiques très conservées [16, 23]: la partie *N*-terminale est majoritairement composée d'acides aminés non chargés. La région centrale de la séquence d'adressage est généralement enrichie en acides aminés hydroxylés (sérine et thréonine). Enfin, l'extrémité *C*-terminale du peptide de transit semble enrichie en acides aminés chargés positivement. Cette région porte le site de clivage de la séquence d'adressage par une protéase du stroma de type «stromal processing protease» (SPP) qui mature les protéines après leur passage au travers du canal formé par les complexes TOC/TIC [24].

2. Le complexe TOC

Comme pour les mitochondries, pour assurer la spécificité de l'adressage des protéines canoniques, les chloroplastes possèdent un système d'import spécifique. Le complexe TOC constitue la première partie d'un tel système d'import. Ce complexe est fonctionnellement similaire au complexe TOM «Translocon at Mitochondrial Outer envelope membrane» chez les mitochondries [25-27]. Au cours des dernières décennies, plusieurs études ont été réalisées dans le but d'isoler les composants du complexe TOC. Les premières tentatives, réalisées au début des années 1990 en utilisant les chloroplastes de pois isolés comme système modèle [28], ont abouti à l'identification des trois composants majeurs du complexe TOC : Toc34, Toc75 et Toc159, nommés ainsi en fonction de leur poids moléculaire [29]. Plus tard, deux autres protéines, Toc12 [30] et Toc64, ont été identifiées comme de nouveaux composants putatifs du complexe TOC. Des homologues de toutes ces sous-unités ont été par la suite caractérisés chez *Arabidopsis thaliana*. En effet, il existe quatre homologues du composant Toc159 chez Arabidopsis : atToc159, atToc132, atToc120 et atToc90 ; et deux homologues du composant Toc34 : atToc34 et atToc33 [31, 32]. Les récepteurs Toc34 sont caractérisés par un seul domaine GTPase cytosolique (domaine G) ancré à la membrane externe de l'enveloppe du chloroplaste. Les membres de la famille Toc159 sont structurellement beaucoup plus complexes, comprenant trois domaines distincts : un domaine *N*-terminal (domaine A) très acide, un domaine central GTP-binding (domaine G) et un domaine d'ancrage membranaire *C*-terminal (domaine M). Bien qu'ils partagent une similarité élevée dans leurs domaines G et M, ces récepteurs Toc159 diffèrent grandement dans la longueur et la séquence de leurs domaines A [33, 34]. Les membres de la famille Toc159 et Toc34 jouent un rôle central dans le processus d'import des protéines dans les chloroplastes. En effet, ils constituent les premiers acteurs de la machinerie d'import, directement impliqués dans la reconnaissance du peptide de transit des pré-protéines cytosoliques à la surface des chloroplastes [35]. Toc75 est le troisième composant essentiel du complexe d'import TOC, et il constitue un composant majeur du canal de translocation de la pré-protéine de la membrane externe [35]. Ainsi, les trois composants Toc159, Toc34 et Toc75 forment un complexe stable dans la membrane externe de l'enveloppe du chloroplaste, dédié à l'import des pré-protéines cytosoliques (figure 3). Des études génétiques et biochimiques sur les différents isoformes de Toc159 et Toc34 ont révélé l'existence de deux combinaisons de complexes TOC présentant une sélectivité différentielle pour la reconnaissance et la translocation des protéines [36-39]. En effet, des expériences de co-immunoprécipitation effectuées par Ivanova et

ses collaborateurs ont démontré que atToc159 s'associe préférentiellement à atToc33, tandis que atToc120 et atoc132 forment préférentiellement un complexe avec atToc34 [40] (figure 3). Cette étude a été soutenue par des études génétiques, étudiant le mutant knock-out toc34 (ppi3) qui n'a pas révélé de phénotype visible [37] tandis que le mutant toc33 (ppi1) présente un phénotype vert pâle [38] avec un défaut de biogenèse chloroplastique similaire à celui du mutant toc159 (ppi2). Ces résultats supportent l'hypothèse selon laquelle ces deux derniers récepteurs fonctionnent dans la même voie d'import de pré-protéines avec une spécificité pour les protéines photosynthétiques [32, 36-38, 41]. En effet, alors que l'accumulation des protéines photosynthétiques est affectée dans le mutant ppi2, les protéines non photosynthétiques semblent s'accumuler à des niveaux normaux [41]. Cependant, le rôle de Toc64 dans l'import de protéines chloroplastiques reste discutable. Bien que son rôle de récepteur pour les pré-protéines délivrées par des facteurs cytosoliques soit validé par des études biochimiques *in vitro* [42], des études génétiques basées sur des mutations affectant l'expression de Toc64 ont montré que l'absence de cette protéine ne conduit à aucun défaut évident [43], suggérant que Toc64 n'est pas essentiel à l'import de protéines *in planta* [44]. Une hypothèse possible est que Toc64 joue un rôle accessoire dans l'import des protéines dans les chloroplastes et que sa fonction n'est importante que dans certaines circonstances spécifiques. Une étude récente, réalisée par Kessler et ses collaborateurs [45], a abouti à l'identification d'une kinase nommée KOC1 qui semble être associée à la membrane externe de l'enveloppe du chloroplaste. KOC1 interagit avec Toc159 lors de l'import des précurseurs à la surface du chloroplaste, sans pour autant faire partie du translocon d'import TOC. En effet, une association de KOC1 avec Toc159 a été montrée par la co-purification de cette kinase avec la protéine Toc159 marqué par un tag (TAP) inséré à son extrémité *N*-terminale (NTAP-Toc159) [45]. Pour confirmer ces résultats, la même technique a été reproduite en marquant la protéine KOC1 (NTAP-KOC1). Toc159, Toc75 et Toc33 ont été détectés dans l'éluât de NTAP-KOC1. Des approches génétiques, utilisant le mutant *koc1*, ont montré l'implication de la protéine kinase KOC1 dans le processus d'import des pré-protéines dans les chloroplastes sans interférer avec la composition du complexe TOC [45]. Elle est essentiellement directement impliquée dans la régulation de l'activité des récepteurs d'import à savoir la phosphorylation du domaine A de la protéine Toc159 et ses homologues.

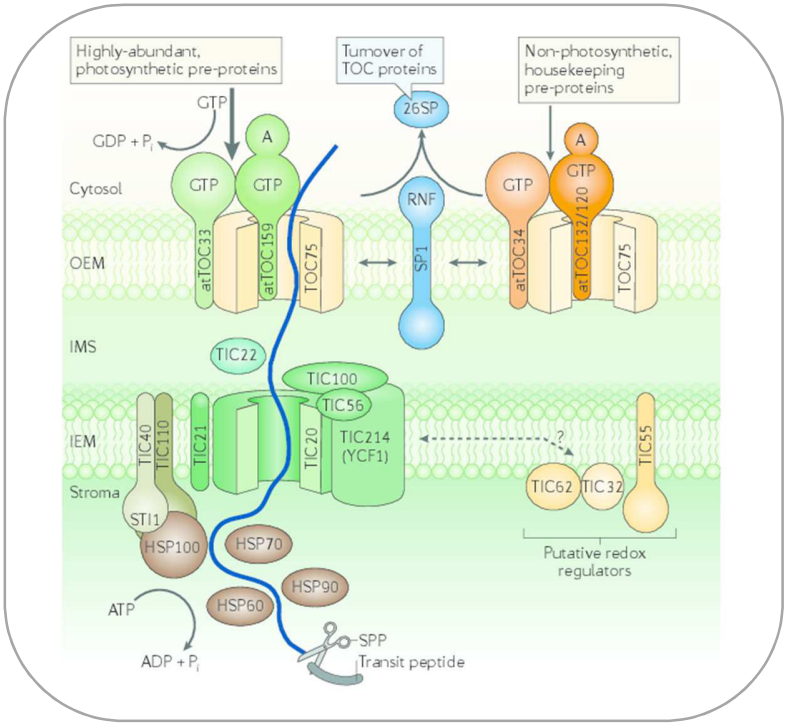


Figure 3. Modèle du mécanisme d'action de deux combinaisons du complexe TOC (d'après Jarvis, P. and E. Lopez-Juez., 2013).

3. Le complexe TIC

Le complexe TIC constitue la deuxième partie du système d'import des protéines canoniques dans les chloroplastes. Ce complexe est fonctionnellement similaire au complexe TIM «Translocon at Mitochondrial Inner envelope membrane» présent chez les mitochondries [25, 27]. Des études antérieures ont révélé l'existence d'au moins 8 protéines majeures constitutives de ce complexe : Tic110, Tic62, Tic55, Tic40, Tic32, Tic22, Tic20 et Tic21 [29, 46, 47]. Sur la base de ces nombreuses études, un modèle classique de translocation de protéines chloroplastiques à travers la membrane interne de l'enveloppe a été construit et accepté jusqu'à récemment [24, 48-50].

a) Le modèle classique du système d'import TIC

Il a été toujours admis que le système d'import de protéines canoniques TIC fonctionne comme suit : le segment *N*-terminal de la pré-protéine émerge du complexe TOC vers l'espace inter-membranaire, il interagit avec Tic22 associé à la face externe de la membrane interne de l'enveloppe et qui a pour rôle de faciliter le transit des précurseurs [35] (figure 4). Ensuite, le

précurseur est pris en charge par la protéine Tic110 qui forme un canal de translocation du complexe TIC avec les deux protéines Tic20 et Tic21 [35]. Le domaine soluble de cette protéine Tic110 semble interagir avec la protéine Tic40 qui, à son tour, ancrée à la face stromale de la membrane interne (figure 4), joue un rôle dans la régulation de l'activité des chaperonnes du stroma (Hsp93, cpHsp70…) en coordination avec Tic110, durant le processus d'import des précurseurs [35]. Enfin, ces chaperons moléculaires coopèrent afin de conduire la translocation complète de la pré-protéine entière à travers la membrane interne de l'enveloppe chloroplastique. Quant aux protéines Tic55, Tic62 et Tic32, elles semblent assurer un rôle de régulation en réponse aux variations de l'état redox des chloroplastes, du fait de la présence de sites de fixations du NADPH [35] (figure 4).

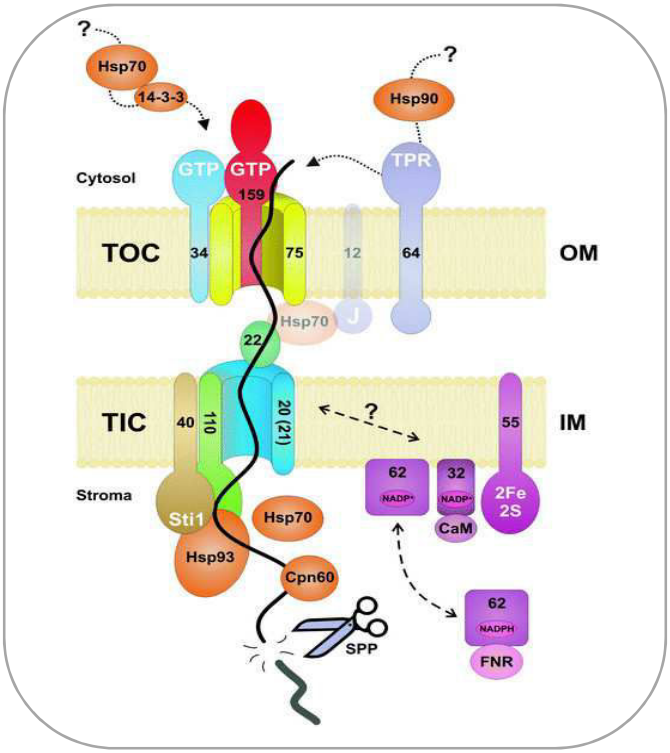


Figure 4. Modèle du système d'import des protéines chloroplastiques : les complexes TOC et TIC (d'après Ling et *al*., 2013).

Ce modèle classique du système d'import TIC a longtemps été le plus répandu. Cependant, la question la plus controversée qui demeurait non résolue depuis le tout début de l'identification d'un tel complexe d'import, concernait l'association étroite réelle entre les différentes sous-unités TIC mentionnées ci-dessus. En effet, toutes les études biochimiques antérieures, reposant sur des

méthodologies classiques combinant la réticulation chimique non spécifique et l'immunoprécipitation, n'ont pas révélé la présence évidente d'une association étroite observable entre les protéines connues du complexe TIC telle que mentionnée ci-dessus. Pour cette raison, plusieurs modèles d'assemblage des différents composants TIC ont été proposés [24, 48-50]. Une étude récente, réalisée par Nakai et ses collaborateurs [51, 52], a montré l'existence d'un nouveau modèle du complexe de translocation des pré-protéines à travers la membrane interne de l'enveloppe : le complexe de 1-megadalton (1 MDa).

b) Le modèle complexe 1 MDa du système d'import TIC

En raison de l'incertitude dans le processus d'identification de certaines protéines TIC proposées par les techniques aspécifiques de réticulation chimique et d'immunoprécipitation, l'équipe japonaise de Nakai s'est concentrée sur l'identification d'un véritable complexe d'import à travers la membrane interne en appliquant des expériences d'importation *in vitro* utilisant des pré-protéines radiomarquées et des chloroplastes intacts isolés, suivies d'une séparation BN-PAGE des complexes de protéines membranaires solubilisés par un détergent doux, la digitonine, sans aucune réticulation chimique préalable. Les analyses des signaux radioactifs correspondant aux pré-protéines accumulées ont abouti à l'observation d'un complexe de très haut poids moléculaire (1 MDa). La localisation de ce complexe au niveau de la membrane interne du chloroplaste a été ensuite confirmée par des approches biochimiques. Il s'est avéré que la protéine Tic20 constitue le composant central du complexe de 1 MDa (figure 5). Tic21 semble associée à la périphérie de ce complexe, du fait de sa dissociation complète en présence d'une forte concentration en sels pendant la solubilisation du complexe [53]. Ces expériences ont également montré que les deux protéines Tic110 et Tic40 sont absentes de ce complexe de translocation de 1 MDa. Pour déterminer la composition réelle de ce complexe, une plante transgénique d'Arabidopsis exprimant une forme étiquetée de Tic20 a été construite et le complexe contenant cette forme modifiée de Tic20 a pu être purifié [51]. Ce complexe purifié conservait sa taille de 1 MDa et contenait trois protéines essentielles, précédemment non caractérisées, de 214 kDa, 100 kDa et 56 kDa qui interagissent avec la protéine Tic20. Ces protéines ont été nommées respectivement Tic214, Tic100 et Tic56. Toutes ces protéines, y compris Tic20, forment un complexe de 1 MDa stable dans les chloroplastes d'Arabidopsis (figure 5). Ces résultats ont été supportés par l'étude de Köhler et ses collaborateurs [54] qui ont caractérisé le rôle du composant Tic56 du complexe de 1 MDa dans l'import des

précurseurs dans les chloroplastes en se basant sur des approches *in planta*. Toutefois, cette étude démontre que l'import des protéines dans le mutant *tic56* reste fonctionnel suggérant l'existence de translocases alternatives dans la membrane interne [54].

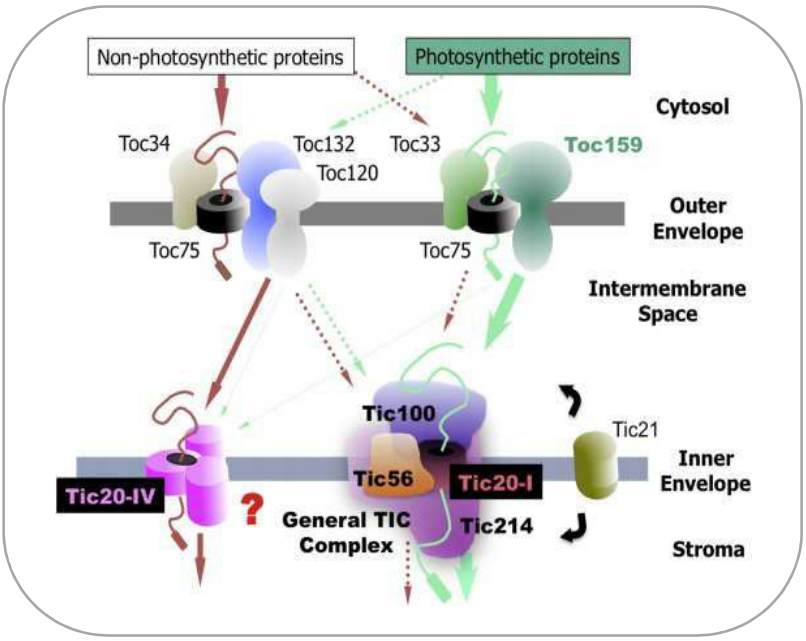


Figure 5. Modèle du mécanisme d'action du complexe de 1 MDa du système d'import TIC (d'après Nakai, 2015).

4. Le mécanisme général d'import des protéines canoniques

La première étape de l'import dans le chloroplaste a lieu dans le cytosol : après la prise en charge des précurseurs dotés d'un peptide de transit clivable en *N*-terminal par les molécules chaperonnes Hsp70 [20], ces protéines sont soit reconnues par la protéine 14-3-3 [21] pour former un complexe qui guide le précurseur jusqu'à la membrane externe de l'enveloppe, soit s'associent simplement à une deuxième protéine chaperonne Hsp90. La suite de l'import est commune aux différents précurseurs de même destination sub-plastidiale.

a) Protéines adressées à la membrane interne, l'espace intermembranaire et le stroma

Sur la membrane externe de l'enveloppe des chloroplastes, la protéine précurseur interagit avec les récepteurs Toc159 et Toc34 par l'intermédiaire de leur domaine GTPase. Lors de l'hydrolyse du GTP par Toc34, le précurseur est transféré dans le canal conducteur de la protéine Toc75. La translocation à travers la membrane de l'enveloppe externe est entraînée par de faibles

concentrations d'ATP (<100 μM) [55]. Une fois arrivé dans l'espace intermembranaire, le précurseur est pris en charge par un complexe constitué par Toc64, Toc12, Tic22 et Hsp70. Dans l'étape suivante, le précurseur est engagé dans le canal conducteur de la protéine formé par Tic20 et Tic21 dans la membrane interne de d'enveloppe. Du côté trans de la membrane interne, la région *N*-terminale émergente du précurseur interagit avec le domaine stromal de la protéine Tic110. Ce dernier composant interagit également avec Tic40 et recrute les chaperones stromales Hsp93 et Hsp70 au complexe TIC. Lors de l'hydrolyse de l'ATP par ces deux chaperones, le précurseur est importé dans le stroma. Les trois composants régulateurs Tic62, Tic55 et Tic32 interviennent dans la régulation de l'importation du précurseur. Pendant, ou peu après, l'arrivée dans le stroma, le peptide de transit est clivé par la peptidase stromale SPP qui permet le clivage du peptide et la libération de la protéine mature. La protéine mature se replie ensuite dans le stroma, spontanément ou sous l'influence de chaperones de type Hsp60 [56]. La majorité des protéines destinées à la membrane interne (cas des protéines Tic110 et Tic40) et l'espace inter-membranaire (cas des protéines MGD1 « monogalactosyldiacylglycerol synthase 1 » et Tic22) transite également dans le stroma pour être maturées avant de rejoindre leur destination finale (membrane interne ou espace intermembranaire) [26, 57, 58].

b) Protéines destinées vers les thylakoïdes

Une fois dans le stroma, une partie des protéines va ensuite pouvoir rejoindre les sous-compartiments des thylakoïdes (la membrane et le lumen). Quatre voies sont décrites (figure 6) dont deux voies permettent l'insertion des protéines dans la membrane des thylakoïdes : la voie SRP-dépendante « Signal Recognition Particle » et la voie spontanée. Les deux autres voies permettent le transport des protéines au sein du lumen des thylakoïdes (voies TAT1 « Twin Arginine Translocase » et Sec1 « Secretory dependent pathway »). Toutes les protéines qui empruntent ces voies possèdent un peptide de transit composé de deux parties : la partie *N*-terminale pour l'import dans le stroma, et une séquence additionnelle, pour l'import dans les thylakoïdes. Une fois la protéine importée au sein du lumen, cette séquence d'adressage additionnelle est éliminée par une endopeptidase liée à la surface interne de la membrane des thylakoïdes : TPP « Thylakoïdal Processing Peptidase ».

i) La voie SRP-dépendante : sert à l'intégration de protéines comme les LHCPs « Light Harvest Complexe Proteins ». L'insertion, dépendante de l'hydrolyse du GTP, est médiée par trois facteurs localisés dans le stroma des chloroplastes (cpSRP54, cpSRP43 et cpFtsy) et par une protéine de la membrane des thylakoïdes Alb3 « Albino3 » [59, 60].

ii) La voie spontanée : dépend de la présence de deux domaines hydrophobes localisés dans la partie *N*-terminale de la protéine. Elle concerne un nombre limité de protéines, parmi lesquelles on retrouve les protéines responsables de l'assemblage de l'ATP synthase chloroplastique [61].

iii) La voie Sec1 : est utilisée par des protéines telles que les sous-unités du photosystème I. Trois protéines sont constitutives de ce complexe d'import : les protéines SecA1, SecY1 et SecE1 homologues du système de transport bactérien SecYEG qui exporte les protéines vers la membrane plasmique et l'espace périplasmique [62]. SecA1, localisée dans le stroma des chloroplastes, constitue le premier acteur du complexe Sec1 qui lie le précurseur avant de le transférer aux protéines SecY1 et SecE1, enchâssées dans la membrane des thylakoïdes [63].

iv) La voie Tat1 : est utilisée par les protéines qui possèdent deux résidus arginine en amont du domaine central hydrophobe. Cette voie, dépendante du gradient de protons générés par la photosynthèse, implique une translocase constituée de trois sous-unités TatA1, TatB1 et TatC1 [64] également homologues au système de translocation bactérien.

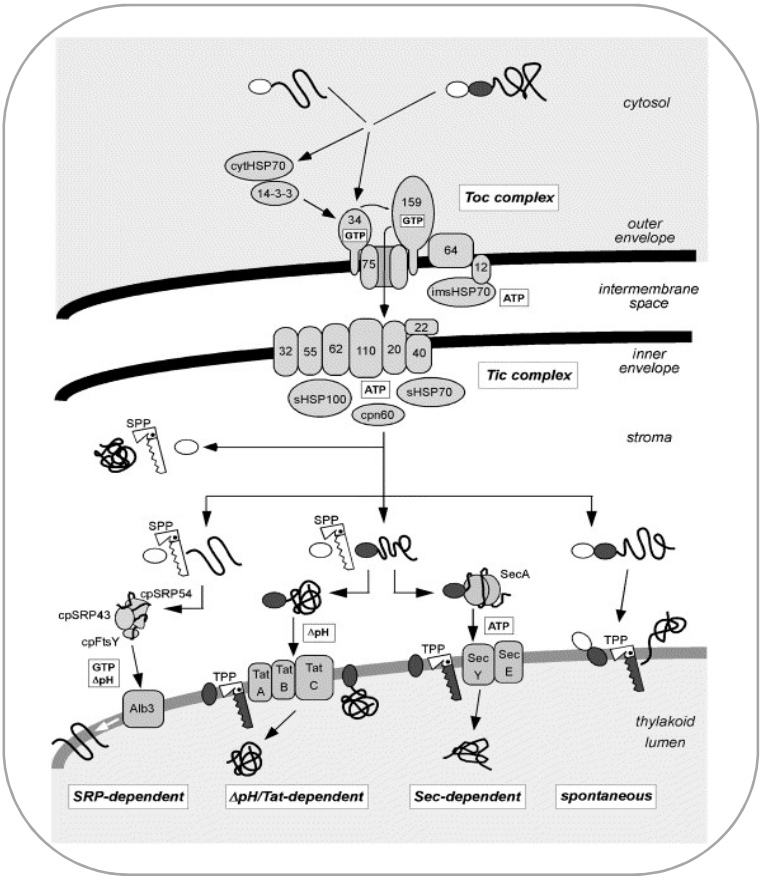


Figure 6. Représentation schématique des machineries d'importation de protéines dans le thylakoïde (d'après Gutensohn et *al*., 2006).

c) Protéines adressées à la membrane externe de l'enveloppe

Il est admis que les protéines adressées à la membrane externe de l'enveloppe du chloroplaste ne possèdent pas de peptide de transit clivable. Leur insertion dans la membrane se fait via un signal interne hydrophobe indépendamment du système d'importation classique TOC/TIC [65]. Plusieurs mécanismes ont été discutés pour l'adressage des protéines à la membrane externe ; le plus important parmi ceux-ci est celui probablement utilisé par les protéines OEP14 [66, 67] et Toc64 / OEP64 qui possèdent un domaine transmembranaire localisé en *N*-terminal de leur séquence primaire et qui semble être responsable de leur adressage spécifique dans la membrane externe. De la même manière, il a été montré par Muckel et Soll [68] et Li et Chen [69] que la spécificité d'adressage des protéines Toc159 et Toc34 dans la membrane externe de l'enveloppe est strictement dépendante d'une région intrinsèque localisée cette fois dans l'extrémité *C*-terminale de la séquence primaire de ces protéines. Cette localisation variable du signal d'adressage rend

difficile la compréhension du mécanisme d'ancrage dans la membrane externe de l'enveloppe. Cependant, il semble que ces domaines transmembranaires hydrophobes sont capables d'interagir avec la bicouche lipidique de la membrane externe. Plusieurs autres interrogations sur le mode d'insertion des protéines dans la membrane externe de l'enveloppe ont été soulevées, à savoir si cette insertion est médiée par le système endomembranaire (réticulum endoplasmique). Bien qu'elle soit une protéine de la membrane externe de l'enveloppe du chloroplaste, Toc75 est particulièrement adressée à la membrane externe via une séquence d'adressage chloroplastique clivable en *N*-terminal. Cette séquence d'adressage dite bipartite est constituée de deux parties distinctes : une partie *N*-terminale permettant au précurseur de traverser la membrane externe et la membrane interne de l'enveloppe et ensuite clivée dans le stroma par une protéase spécifique ; une partie *C*-terminale, quant à elle, semble servir de signal pour stopper le transfert de la protéine Toc75 à travers le complexe TIC de la membrane interne. Ainsi, cette partie *C*-terminale assure l'insertion de cette protéine vers la membrane externe de l'enveloppe [70]. Elle est clivée lorsque la protéine Toc75 est insérée dans la membrane externe par un mécanisme qui demeure inconnu [71].

III. Voies d'imports alternatives des protéines dans le chloroplaste

Pendant de nombreuses années, à l'exception de protéines de la membrane externe de l'enveloppe du chloroplaste, l'import de protéines via la machinerie d'import TOC/TIC a été considéré comme la seule voie d'adressage à destination des différents compartiments sub-plastidiaux. Cependant, des études récentes ont révélé l'existence de plusieurs protéines qui ne possèdent pas de séquence de transit clivable en *N*-terminal [7, 26, 72, 73]. Bien que dépourvues de séquence d'adressage clivable, ces protéines dites « non canoniques » atteignent néanmoins les compartiments internes du chloroplaste. La première évidence de telles protéines non canoniques a été fournie par notre équipe, étudiant le protéome de l'enveloppe du chloroplaste d'Arabidopsis en collaboration avec la plateforme de protéomique « EDyP », qui a conduit à l'identification d'une protéine homologue aux quinones oxidoréductases nommée « ceQORH » [6, 7]. En effet, malgré l'absence d'une pré-séquence de transit clivable, ceQORH est associée à la face stromale de la membrane interne de l'enveloppe du chloroplaste [74]. De plus, il a été également montré que l'import de cette protéine

ceQORH dans le chloroplaste n'est pas médié par la machinerie d'import générale TOC/TIC [9]. La même étude a également montré que la protéine ceQORH est multilocalisée dans les cellules des différents tissus de la plante [9]. Plus récemment, une étude menée par Nada et Soll [8] a abouti à l'identification d'une deuxième protéine membranaire « Tic32/IEP32 » qui possède aussi un peptide de transit non canonique. De la même manière, malgré l'absence d'une séquence d'adressage clivable dans la séquence primaire de cette protéine, Tic32 est toutefois associée à la membrane interne de l'enveloppe du chloroplaste. La même étude a également montré que, contrairement à la protéine ceQORH, l'import de Tic32 n'exige pas de fortes concentrations en ATP et se fait aussi indépendamment du système d'import classique TOC/TIC [8]. Ce résultat suggère que Tic32 et ceQORH ne sont pas en compétition pour le même mécanisme d'import. Plus récemment, en se basant toujours sur des expériences de protéomique, notre équipe a pu identifier une troisième protéine non canonique nommée « NQR » (résultats non publiés). Ainsi, cette protéine ne possède pas de peptide de transit clivable à l'extrémité *N*-terminale, mais elle est partagée entre le stroma du chloroplaste et d'autres compartiments de la cellule (résultats non publiés). Sur la base de toutes ces études, une équipe allemande a eu l'idée de généraliser l'existence d'une seconde classe de protéines chloroplastiques pourvues d'une séquence d'adressage non canonique en se basant sur des outils de prédictions bio-informatiques [73]. Environ 12 % de l'ensemble de protéines du chloroplaste ne sont pas dotées d'une séquence de transit classiques et ont été classées dans la liste des protéines non canoniques [73]. Il est supposé, mais non démontré que l'ensemble de cette catégorie de protéines adoptent des voies d'import dites « alternatives », indépendantes de la machinerie générale TOC/TIC, pour leur translocation vers leur destination sub-plastidiale [73, 75]. Cependant, toutes ces machineries d'import alternatives ainsi que les mécanismes qui contrôlent la localisation sub-cellulaire des protéines chloroplastiques contenant un peptide de transit non canonique demeurent encore inconnus.

1. Protéines non canoniques

a) La protéine ceQORH

Dans le cadre d'une étude protéomique sur l'enveloppe du chloroplaste d'Arabidopsis [76, 77] visant à identifier de nouveaux composants protéiques de l'enveloppe chloroplastique dans le but de mieux comprendre la biogenèse et la dynamique du chloroplaste, notre équipe en collaboration

avec la plateforme de protéomique « EDyP » a pu identifier la protéine quinone oxidoréductase homologue nommée « ceQORH » [7, 9, 76, 78-80]. Alors que la plupart des protéines de l'enveloppe sont très hydrophobes, cette protéine ceQORH est structurellement apparentée aux protéines solubles de la famille des quinones oxydoréductases qui sont conservées chez les bactéries, les champignons et les animaux. Des expériences préliminaires de fractionnement des différents compartiments du chloroplaste et d'immuno-marquage ont démontré la localisation exclusive de la protéine ceQORH dans l'enveloppe du chloroplaste [7], excluant ainsi sa localisation potentielle dans le stroma du chloroplaste et la contamination potentielle de la fraction d'enveloppe analysée en protéomique par cette protéine hydrosoluble. La même étude [7] a ensuite confirmé l'association de cette protéine ceQORH à la membrane interne de l'enveloppe des chloroplastes. En effet, la combinaison d'un traitement protéolytique et le fractionnement de l'enveloppe en membranes interne et externe suivi d'un immuno-marquage en utilisant l'anticoprs anti ceQORH, a montré que cette protéine était insensible au traitement des chloroplastes intacts par la thermolysine (une protéase qui n'a pas la capacité de traverser la membrane externe). Bien qu'elle soit associée à la membrane interne de l'enveloppe, il s'est avéré que plusieurs caractéristiques de la protéine ceQORH allaient à l'encontre d'une telle localisation sub-plastidiale. En effet, toujours dans la même étude menée par notre équipe [7], il a été démontré que la protéine ceQORH ne subit pas de clivage post-traductionnel d'un peptide de transit *N*-terminal suite à son import dans le chloroplaste. Ce résultat a été déduit du fait de la conservation de la taille de la protéine native ceQORH observée sur gel (et par spectrométrie de masse) par rapport à la taille prédite de la séquence primaire codée par l'ADNc d'*Arabidopsis thaliana*. Cette évidence a été également validée par des expériences d'import *in vitro* montrant que la protéine ceQORH n'est pas maturée après son processus d'import. De plus, il a été également démontré que les deux régions *N*-terminale et *C*-terminale de cette protéine ne sont pas essentielles à son adressage dans le chloroplaste. En effet, afin de déterminer le domaine responsable de l'adressage chloroplastique de cette protéine, différentes constructions codant pour des formes tronquées de la protéine ceQORH fusionnées à la GFP ont été produites et testées *in planta* [7], puis lors d'expériences d'import *in vitro* [9]. La localisation sub-cellulaire de chaque construction a été ensuite analysée par microscopie confocale [7, 9] ou immunodétection. Il s'est avéré qu'une région interne de 40 acides aminés (positions 60-100) dans la séquence primaire de la protéine ceQORH est essentielle et suffisante pour son adressage *in vitro* dans le chloroplaste. Cependant, *in planta* cette séquence

s'avère insuffisante pour permettre l'adressage plastidial de cette protéine. La délétion des domaines *N*-terminal ou *C*-terminal de la protéine ceQORH semble être préjudiciable à l'import de cette protéine [7, 9]. En effet, bien que n'étant pas essentielles à l'adressage chloroplastique de la protéine ceQORH, les régions *N*-terminale et *C*-terminale de cette protéine sont responsables de l'interaction de la protéine ceQORH avec les membranes biologiques [7, 9] (figure 7). Dans le chloroplaste, la région *N*-terminale de la protéine ceQORH (positions 6-100) permet une interaction aspécifique de cette protéine avec l'enveloppe et les membranes des thylakoïdes. La région *C*-terminale de la protéine ceQORH (résidus 100 à 323), quant à elle, assure la localisation spécifique de cette protéine dans l'enveloppe des chloroplastes [9] (figure 7). Ainsi, du fait de cette organisation tripartite unique (figure 7), la protéine ceQORH se distingue des autres protéines chloroplastiques dites canoniques.

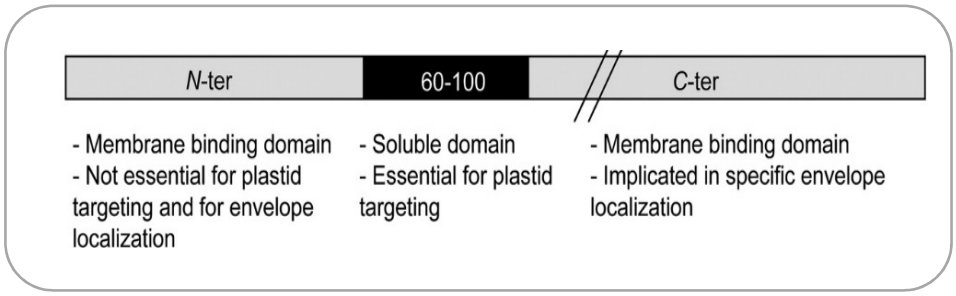


Figure 7. Modèle résumant les fonctionnalités des différents domaines de la protéine ceQORH (d'après Miras et *al*., 2007).

La protéine ceQORH n'utilise pas la voie générale d'import des précurseurs chloroplastiques TOC/TIC [9, 81]. En effet, des expériences d'import *in vitro* ont montré que cette protéine n'est pas en compétition avec les pré-protéines chloroplastiques à savoir la petite sous-unité de la Rubisco (pSSU) et la protéine ferrédoxine (pFd), pour l'utilisation de la machinerie d'import TOC/TIC [9]. Ce résultat a été également supporté par des expériences de blocage des voies d'import TOC/TIC en utilisant des anticorps dirigés contre les protéines récepteurs constitutives du complexe d'import général TOC/TIC (anticorps anti Toc159 et anticorps anti Toc75). En effet, bien que la voie TOC soit inaccessible, l'import de la protéine ceQORH est toujours conservé [9]. Ainsi, cette protéine utilise une voie d'import originale qui demeure inconnue à ce jour. Ce résultat a été récemment validé par une équipe allemande montrant que la protéine ceQORH passe par une voie indépendante de la voie TOC [81]. De plus, la protéine ceQORH a la particularité d'être

multilocalisée notamment dans les cellules de l'épiderme de la feuille où la protéine est essentiellement localisée dans la membrane plasmique, alors qu'elle est localisée dans l'enveloppe du chloroplaste dans les cellules du parenchyme [9] (figure 8). Les mécanismes moléculaires expliquant cette multilocalisation cellulaire de la protéine ceQORH demeurent aussi inconnus; notre équipe a émis l'hypothèse d'un lien probable entre cette localisation sub-cellulaire variable et l'interaction de cette protéine avec un partenaire cytosolique ou sous l'influence de différents stress biotiques et abiotiques. Plus récemment, afin de mieux comprendre son rôle ainsi que sa structure originale, la protéine ceQORH a été étudiée sur le plan structural et fonctionnel [78-80]. Il s'avère que cette protéine participe à la détoxication de la plante en métabolisant des longues chaines carbonées toxiques issues du stress oxydatif [78, 80]. En effet, cette protéine a un rôle dans la réduction de la double liaison des oxylipines électrophiles réactives de type γ-ketols issus de la peroxydation enzymatique des lipides synthétisés comme sous-produits de la voie métabolique du jasmonate [78, 80]. Afin de mieux caractériser sa fonction, une étude structurale a été réalisée récemment en se basant sur une approche de cristallographie et diffraction des rayons X. Ainsi, les données cristallographiques de la protéine ceQORH sont maintenant disponibles à une résolution de 2.34 Å [79].

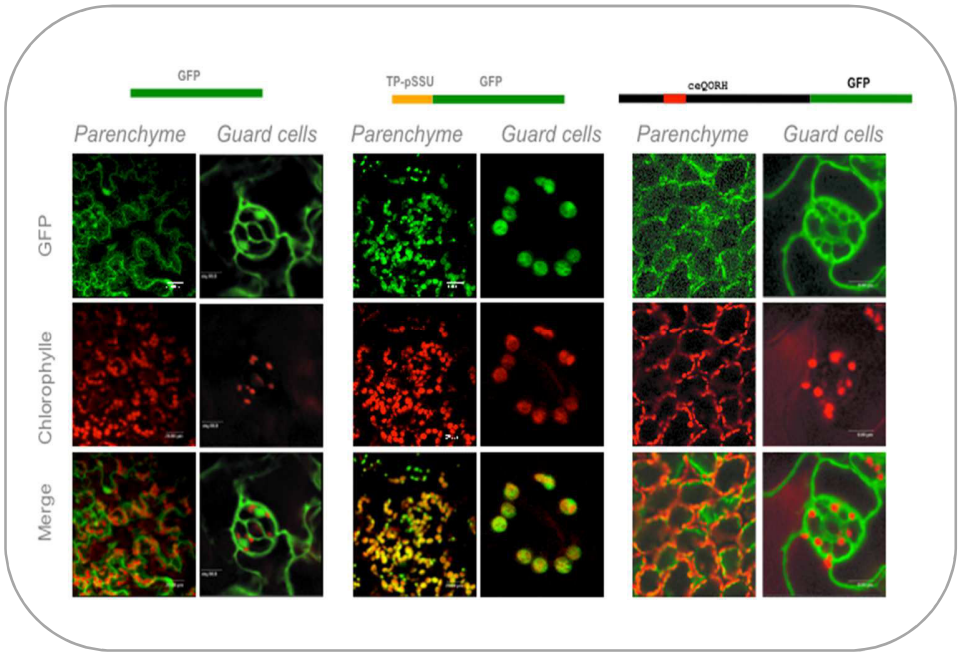


Figure 8. Localisation sub-cellulaire variable de la protéine ceQORH. Dans les cellules du parenchyme de la feuille, la protéine est localisée dans les chloroplastes, alors que dans les cellules de l'épiderme, en particulier dans les cellules de garde, sa localisation est cytosolique (d'après Miras et *al.*, 2007).

b) La protéine TIC32

Tic32, comme ceQORH, est une protéine membranaire localisée dans l'enveloppe des plastes. Similaire à Tic62, Tic32 appartient à la famille des déshydrogénases réductases à chaîne courte [46]. Il a été démontré que les mutants d'Arabidopsis affectés dans l'expression de Tic32 sont viables et ne présentent pas de phénotype notable d'import des protéines dans les chloroplastes. Ainsi, Tic32 n'est pas essentiel pour la biogenèse des chloroplastes [46, 82]. Comme pour la protéine ceQORH, la protéine Tic32 a été identifiée dans le cadre d'une analyse protéomique des membranes de l'enveloppe du pois réalisée par une équipe allemande en 2004 [8]. Il s'agit de la seconde protéine non canonique de la membrane interne de l'enveloppe du chloroplaste. En effet, Tic32 ne possède pas de séquence de transit clivable [8, 24] mais l'information nécessaire à son adressage dans le chloroplaste réside dans l'extrémité *N*-terminale de sa séquence primaire [8]. Des mutants de délétion progressive des premiers acides aminés (10-40) de la protéine Tic32 ont été produits. Des essais d'import *in vitro* de la protéine Tic32 sauvage et des protéines tronquées ont été réalisés. Contrairement à la forme sauvage de Tic32, toutes les protéines tronquées sont devenues accessibles à la thermolysine, ce qui démontre que la suppression des dix premiers acides aminés est suffisante pour abolir l'import, indiquant que cette partie constitue une partie indispensable du signal [8]. Comme pour la protéine ceQORH, la protéine Tic32 n'utilise pas la voie générale d'import des précurseurs chloroplastiques TOC/TIC. En effet, des expériences d'import *in vitro* en présence d'un compétiteur, le précurseur de la sous-unité du complexe d'oxydation de l'eau (Oxygen-Evolving Complex) « pOE33 », ont montré que contrairement à la pSSU dont l'import a été inhibée, l'import de Tic32 n'a pas été affecté [8]. Cependant, contrairement à la protéine ceQORH, la concentration d'ATP requise pour l'import de Tic32 est beaucoup plus faible, suggérant que cette protéine utilise plus probablement un système d'import indépendant de celui de la protéine ceQORH. Plus tard, il a été démontré que la protéine Tic32 possède un domaine de liaison à la calmoduline à l'extrémité *C*-terminale de sa séquence primaire [83]. Il a été démontré, *in vitro,* que Tic32 se lie à la calmoduline de manière dépendante du calcium [83]. Tic32 suit donc une voie d'importation distincte qui demeure inconnue. Le principal défi pour l'avenir sera d'identifier les composants de l'enveloppe extérieure impliqués dans la reconnaissance et la translocation de la protéine Tic32.

c) La protéine NQR

Toujours dans le cadre des analyses protéomiques, notre équipe a récemment identifié une troisième protéine chloroplastique non canonique de type NADPH Quinone Réductase nommée « NQR » [84, 85]. Bien qu'elle soit dépourvue de séquence d'adressage chloroplastique clivable, cette protéine est partagée entre le stroma du chloroplaste et d'autres compartiments de la cellule (résultats non publiés). Il s'agit d'une protéine soluble de la famille des NQR, qui protègent les plantes contre divers stress oxydatifs. En particulier, il a été démontré que NQR joue un rôle important dans le maintien de l'homéostasie des plantes en relation avec l'état redox cellulaire et la signalisation de défense suite à une attaque de pathogène fongique [84, 85]. Comme la protéine ceQORH [78], cette protéine NQR pourrait ainsi participer aux processus de signalisation et de défense suite à des attaques pathogènes. Sa localisation multicellulaire pourrait donc résulter soit de fonctions cellulaires non redondantes, soit d'exigences de mécanismes de défense en divers endroits de la cellule végétale. Le mystère de cette localisation variable de la protéine NQR ainsi que son mécanisme d'import dans les chloroplastes restent à identifier.

d) La protéine FSD1

La superoxyde dismutase « FSD1 » est une protéine non canonique sans le peptide de transit clivable classique. Comme les protéines ceQORH et NQR, cette protéine semble aussi être multilocalisée au sein de la cellule végétale. En effet, cette protéine a été détectée dans la membrane plasmique [86] et aussi localisée dans le stroma du chloroplaste [81]. Bien qu'elle soit dépourvue de séquence d'adressage clivable, il s'est avéré que cette protéine est importée dans les plastes par la voie d'import générale, de manière similaire à la majorité des pré-protéines plastidiales. En effet, en réalisant des réactions d'interaction avec l'enveloppe du chloroplaste de pois reposant sur des approches de pontage chimique, il a été démontré que FSD1 lie les deux membres de la famille Toc159 à savoir Toc120 et Toc132. Ce résultat a été conforté par des expériences d'import *in vitro* qui ont montré une diminution forte de l'import de FSD1 en présence de pSSU [81], suggérant que cette protéine est en compétition avec les pré-protéines plastidiales pour l'utilisation de la voie d'import TOC/TIC (figure 9). La même étude a montré que la protéine FSD1 se comporte différemment de la protéine ceQORH dans presque tous les tests effectués [81]. Il s'est avéré que cette protéine FSD1 partage certaines caractéristiques avec les substrats de la voie d'import générale comme la dépendance à l'ATP lors de la translocation et la nécessité de l'extrémité *N*-

terminale pour l'adressage et l'import [81]. En effet, il a été démontré que l'import de la protéine FSD1 est clairement stimulé à des concentrations supérieures à 0,1 mM d'ATP de manière similaire à pSSU. En utilisant des mutants de délétion, l'extrémité *N*-terminale ainsi qu'un domaine C-proximal s'avèrent essentiels pour l'adressage. Ainsi, FSD1 fait partie d'une troisième classe de protéines qui utilisent Toc120 et probablement d'autres composants TOC tels que Toc75 malgré l'absence de l'information de reconnaissance canonique du complexe TOC dans leur séquence primaire (figure 9).

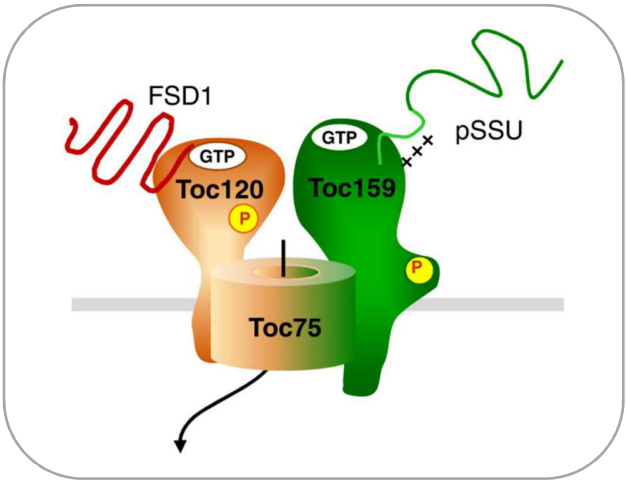


Figure 9. Modèle hypothétique du complexe d'import de la protéine FSD1 (d'après Chang et *al*., 2014).

2. Système endomembranaire et transport vésiculaire

Un article de revue rédigé par Jarvis et ses collègues [87] a posé des questions originales concernant le protéome chloroplastique d'Arabidopsis. En effet, cette analyse a révélé l'existence d'un grand nombre de protéines détectées dans les plastes qui possèdent un peptide signal prédit pour la translocation via le réticulum endoplasmique (RE) [75]. Tandis que certaines de ces protéines étaient clairement des contaminants issus d'autres compartiments de la cellule, l'hypothèse a été émise que certaines protéines interagissent avec le RE du fait de la proximité de la membrane externe de l'enveloppe du chloroplaste et du RE [88, 89]. Ces observations suggéraient l'existence d'un lien probable entre l'interaction du RE avec la surface du chloroplaste et la translocation de certaines protéines chloroplastiques codées par le génome nucléaire via ce système endomembranaire. La première évidence indirecte d'un tel mécanisme d'import impliquant le RE et l'appareil de Golgi résulte de l'étude d'une glycoprotéine qui avait été détectée dans les chloroplastes en 2005 [90]. Dans cette étude impliquant notre équipe, Villarejo *et al* ont démontré

que certaines glycoprotéines utilisent la voie sécrétoire, transitent à travers le RE et le Golgi avant d'être adressées aux plastes. Cette première démonstration concerne l'anhydrase carbonique « CAH1» [90] qui contrôle la concentration du CO_2 dans la cellule et dont la localisation dans le stroma du chloroplaste a été démontrée en dépit du fait qu'elle possède un peptide de sécrétion fonctionnel et correctement prédit par TargetP [91]. En 2005 puis 2009, il a été démontré que l'α-amylase de riz « αAmy3» **[92, 93]**, impliquée dans la dégradation de l'amidon, et qui possède également un peptide de sécrétion, possède une double localisation plastidiale et extracellulaire [92]. Enfin, en 2014, des collègues ont démontré que la nucléotide pyrophosphatase phosphodiestérase de riz « NPP1» [94] est *N*-glycosylée suite à son passage par le RE et l'appareil de Golgi avant son import dans le chloroplaste. Ce résultat a été très récemment conforté par la révélation de la localisation de NPP1 dans les compartiments trans-Golgi qui participeraient donc probablement au mécanisme de trafic et de ciblage du Golgi vers les plastes de ces nucléotides pyrophosphatase phosphodiestérases. Cette voie vésiculaire est supposée dériver des tous premiers stades (post-endosymbiose) de l'évolution des chloroplastes [95, 96]. En conséquence, à la lumière de tous ces travaux, il semble probable que certaines protéines comme CAH1 ou NPP1 sont effectivement transportées vers les chloroplastes au travers du système endomembranaire de sécrétion. Toutefois, les mécanismes moléculaires impliqués restent totalement inconnus (figure 10).

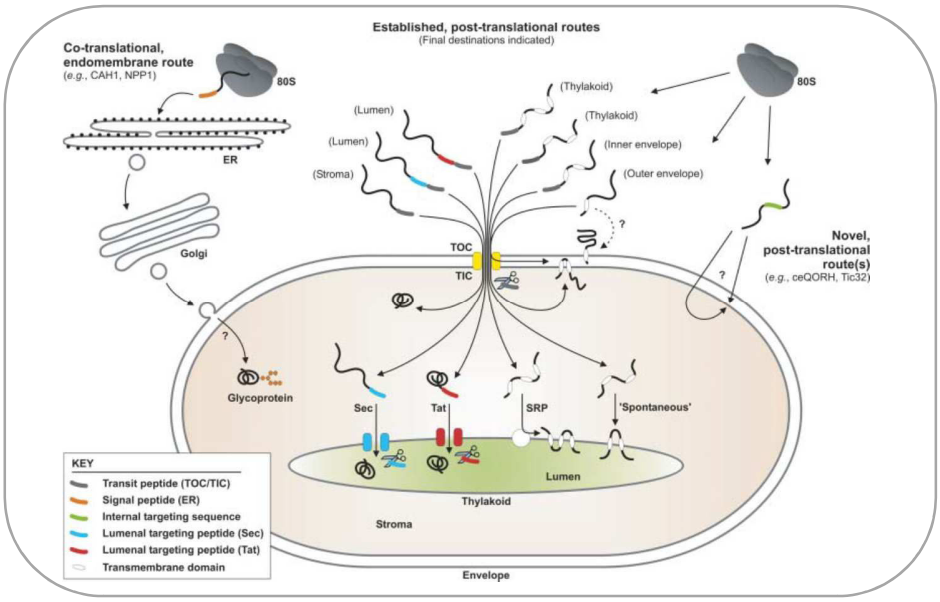


Figure 10. Aperçu des différentes voies d'import des protéines dans les différents compartiments du chloroplaste (d'après Jarvis et *al.*, 2008). Notez les deux points d'interrogation pour les voies dites « alternatives ».

3. Le système d'import PRAT

Une famille de 17 transporteurs putatifs d'acides aminés et de pré-protéines a été identifiée chez *Arabidopsis thaliana* désignée « PRAT» (putative preprotein amino acid transporters). Ces membres de la famille PRAT sont distribués dans les mitochondries ainsi que dans les chloroplastes de la cellule végétale [97, 98]. Dans les mitochondries, ces protéines membranaires à savoir les protéines Tim22, Tim23 et Tim17 jouent un rôle important dans la translocation et l'importation des pré-protéines mitochondriales [99]. Cependant, la fonction précise des composants PRAT associés aux chloroplastes n'est pas encore claire. Une étude récente menée par Reinbothe et ses collègues [100] a abouti à l'identification de trois protéines hypothétiques distinctes de la famille PRAT qui coopèrent lors du transit des protéines non canoniques dans les chloroplastes à savoir la protéine ceQORH [100]. Ces protéines HP20, HP30 et HP30-2 ont été nommées ainsi en fonction de leur poids moléculaires. HP20 est associée à la membrane externe de l'enveloppe du chloroplaste et HP30 et HP30-2 sont deux protéines constitutives de la membrane interne de l'enveloppe du chloroplaste [6, 101]. Le plus proche parent de HP20 est HP22 qui partage 79% d'identité de séquence d'acides aminés avec HP20. En se basant sur des techniques biochimiques reposant sur une réticulation chimique en utilisant la protéine ceQORH comme modèle, il a été conclu que cette protéine interagissait avec les trois protéines HP20, HP30 et HP30-2 lors de son import à l'intérieur du chloroplaste [100]. Cependant, aucun impact n'a été observé sur la localisation de la protéine ceQORH quand elle est exprimée dans deux lignées différentes du mutant hp20 [100]. Ainsi, l'implication de cette voie « PRAT » dans l'import des protéines non canoniques dans les chloroplastes n'est pas validée *in planta*.

4. Le système d'import Sec2

Comme indiqué dans la section II.4.b, un sous ensemble des protéines chloroplastiques sont dirigées vers les membranes des thylakoïdes où elles sont intégrées dans la bicouche membranaire ou transportées dans le lumen des thylakoïdes via les systèmes Sec ou Tat. Le système Sec a été conservé au cours de l'évolution. Il est présent chez les archées, les bactéries et les eucaryotes (figure 11). Il constitue la plus importante voie d'export de protéines vers le périplasme chez les bactéries. Sa présence est indispensable puisque son absence est létale pour la bactérie. Ce complexe Sec est constitué de trois protéines : SecY, SecE et SecG localisées au niveau de la membrane interne de la bactérie (figure 11). De plus, le système comprend également la protéine

cytoplasmique SecA impliquée dans l'apport d'énergie par l'hydrolyse d'ATP indispensable au fonctionnement de la machinerie [102]. Le système Sec a été aussi identifié dans les chloroplastes, à la surface des thylakoïdes, et est responsable de l'import des protéines dans le lumen [103-106]. Ce complexe est constitué de deux composants membranaires : SecY1 qui se compose de dix domaines transmembranaires et forme le canal principal du système d'import ; SecE1 avec un seul domaine transmembranaire, contribue à la stabilité du complexe. Une troisième protéine constitutive du complexe Sec1 désignée SecA1 est un composant soluble relargué dans le stroma et constitue le premier acteur du complexe d'import directement impliqué dans la reconnaissance du précurseur. Il s'est avéré que ces trois composants du complexe Sec jouent un rôle important dans la biogénèse des thylakoïdes [106, 107]. En effet, des mutants affectés dans l'expression des gènes seca1, secy1 et sece1ont tous montré un phénotype albino [106] suggérant que ces translocases sont indispensables pour le maintien des thylakoïdes. Récemment, il a été démontré l'existence d'un deuxième complexe Sec associé à la membrane interne de l'enveloppe du chloroplaste [104-107] désigné Sec2. Ce complexe est structuralement similaire au complexe Sec1 des thylakoïdes, constitué de trois composants : SecA2, SecY2 et SecE2 [104, 108]. Des mutants affectés dans l'expression des composants du complexe Sec2 ont présenté un phénotype embryon létal sévère suggérant que ce système est essentiel pendant les premiers stades de développement de la plante [106]. Pour étudier la fonction du système Sec2 dans les plantes, une étude récente [108] reposant sur l'inhibition spécifique de SecY2 en utilisant des ARN interférents a montré que la protéine SecY2 a un rôle important dans la biogénèse du chloroplaste au-delà du développement de l'embryon. En effet, des chloroplastes aberrants avec des thylakoïdes non développés ont été révélés suite à l'inhibition de l'expression de SecY2. Chez les mêmes plantules qui sont soumises à une régulation négative de l'expression de SecY2, les teneurs de plusieurs protéines membranaires thylakoïdiennes, dont SecY1, Alb3 et TatC, ainsi que des protéines membranaires de l'enveloppe interne, notamment Tic40, Tic110 et FTSH12, ont été considérablement réduites, suggérant qu'elles pourraient être des substrats interagissant avec le complexe Sec2 [108]. Une étude supplémentaire a été réalisée par analyse de l'import *in vitro* des protéines Tic40, FTSH12, SecY1 et Alb3 en présence d'un inhibiteur puissant de la protéine SecA2 ou des anticorps dirigés contre les protéines SecY2 et SecE2 [108]. Ainsi, il s'est avéré que contrairement à SecY1 et Alb3, l'intégration de Tic40 et FTSH12 dans l'enveloppe du chloroplaste est médiée par le complexe Sec2 [108]. Plus précisément, les anticorps dirigés contre SecY2 et Sec2 ont inhibé l'intégration de

Tic40 et l'azide de sodium, un inhibiteur des protéines SecA, a inhibé l'intégration de FTSH12. En conséquence, cette étude fournit une identification initiale des substrats du système Sec2 et il semble probable que plusieurs protéines sont importées dans les chloroplastes via le complexe Sec2 exactement comme Tic40 et FTSH12 et bien d'autres protéines qui restent à identifier.

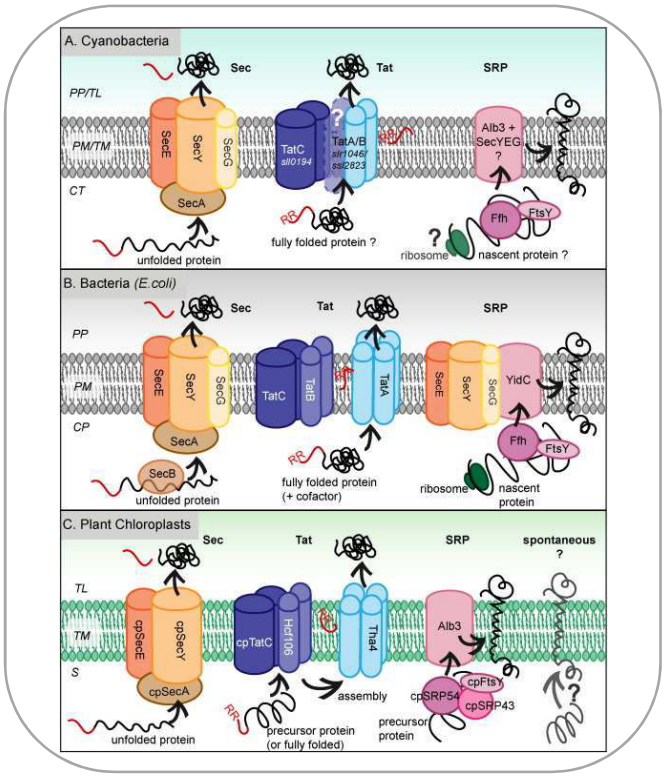


Figure 11. Comparaison de différents mécanismes de translocation dans les membranes des thylakoïdes fonctionnant chez les cyanobactéries, les bactéries et les chloroplastes végétaux (d'après Frain et *al.*, 2016).

5. Un système d'import Tat-like ?

Comme évoqué dans la section II.4.c le système Tat fonctionne en parallèle avec le système Sec au niveau de la membrane des thylakoïdes des chloroplastes ainsi que la membrane plasmique de la plupart des bactéries [109, 110]. Initialement, ce mécanisme a été appelé la voie ΔpH-dépendante parce que l'énergie de translocation dépend du ΔpH transmembranaire [111]. Il s'est avéré que les substrats de cette voie de translocation présentent un motif de type « double arginine » dans leurs peptides signal. Pour cette raison, ce système a été renommé Tat pour Twin Arginine Translocation [112]. Le premier composant identifié de la machinerie Tat était Hcf106 (haute fluorescence chlorophyllienne 106) qui avait été trouvé dans les thylakoïdes des chloroplastes de maïs au début

des années 1990 [113], suivi de l'identification des composants restants, Tha4 et cpTatC. Chez les bactéries et les cyanobactéries, des homologues à ces trois composants ont été identifiés et désignés TatA, TatB et TatC qui correspondent à Tha4, Hcf106 et cpTatC, respectivement dans les chloroplastes. Ces trois composants sont intégrés à la surface des thylakoïdes sous forme de complexe avec cpTatC et Hcf106 qui forment les récepteurs de reconnaissance avec les substrats. Le troisième composant Tha4 constitue le canal conducteur qui vient s'assembler avec le complexe récepteur lié au précurseur pour assurer sa translocation. Plus tard, il a été démontré que la plupart de génomes mitochondriaux des plantes supérieures contiennent une protéine TatC-like [114]. Récemment, ce résultat a été validé par une étude menée par une équipe allemande qui a abouti à l'identification d'une sous-unité de type TatB localisée dans la membrane interne des mitochondries [115]. Il s'est avéré également que le gène correspondant est essentiel pour la croissance et le développement des plantes suggérant qu'une voie Tat est potentiellement requise pour la biogénèse mitochondriale chez les plantes [115]. Toutes ces évidences posent la question de l'existence probable d'un complexe Tat-like alternatif, localisé à la surface des chloroplastes, qui comme le complexe Sec2 de l'enveloppe, serait potentiellement impliqué dans l'import des protéines dans les chloroplastes.

IV. Protéome de l'enveloppe du chloroplaste

1. Protéome du chloroplaste

Comprendre la diversification fonctionnelle et l'évolution des organites cellulaires nécessite l'identification de leurs répertoires protéiques complets, ce qui est le but ultime de la protéomique sub-cellulaire. La prédiction du protéome des chloroplastes a fait l'objet de nombreux débats. Ce débat résulte à l'origine de l'écart entre les protéines potentiellement localisées dans le chloroplaste et les données disponibles à ce jour. Bien que des progrès significatifs aient été réalisés au cours des dernières années pour établir expérimentalement le protéome des chloroplastes, son répertoire reste toujours incomplet. En effet, il est clair que plusieurs protéines clés manquent encore dans le répertoire des protéines chloroplastiques, comme, pour ne citer qu'eux, divers transporteurs membranaires déduits de la compartimentation du métabolisme cellulaire [14].
Les premières études visant à explorer le protéome complet du chloroplaste d'*Arabidopsis thaliana* avaient été réalisées à partir de fractions entières de chloroplastes [75, 116] et avaient abouti à

l'identification d'environ 700 protéines. Cela ne représentait que le cinquième des protéines totales qui étaient prédites pour être associées aux plastes (environ 3500 protéines au total) [26]. En outre, ces données ne fournissaient pas de détails sur la localisation des protéines dans les différents sous compartiments des chloroplastes. Ainsi, pour avoir une meilleure couverture du protéome chloroplastique et accéder aux protéines mineures, des analyses protéomiques indépendantes ont été réalisées en utilisant des fractions sub-plastidiales purifiées. Des études ciblant l'analyse de la composition de l'enveloppe par notre équipe, en collaboration avec la plateforme de protéomique « EDyP» [117], ont abouti à l'identification de nombreuses protéines nouvelles associées à ces membranes. Plusieurs autres études ciblant les thylakoïdes [118-120], le stroma [121], les plastoglobules [122] ou les membranes de l'enveloppe [123, 124] ont été réalisées pour une détection accrue des protéines mineures spécifiques de chaque compartiment sub-plastidial. Cependant, cette sensibilité accrue des appareils de détection s'est accompagnée de la détection d'un plus grand nombre de contaminants mineurs des fractions purifiées. D'autre part, l'identification d'un nombre croissant de protéines mineures de localisation sub-cellulaire et de fonction totalement inconnues a posé la question de la réelle localisation sub-cellulaire de ces protéines. Il a donc été important de développer de nouvelles approches permettant de valider la localisation sub-cellulaire ou sub-plastidiale des protéines identifiées et d'exclure définitivement les contaminants provenant d'autres compartiments du plastes ou de la cellule.

2. Protéome de l'enveloppe

Comme mentionné ci-dessus (section I.2), l'enveloppe du chloroplaste est situé à l'interface entre le chloroplaste et les autres compartiments de la cellule, formant ainsi une barrière stratégique contrôlant les échanges d'ions, de métabolites et de protéines et régulant les fonctions métaboliques essentielles (synthèse des précurseurs d'hormones, acides aminés, pigments, sucres, vitamines, lipides, nucléotides ...) de la cellule végétale. Ainsi, pour mieux comprendre la fonction et la biogenèse du chloroplaste, il est essentiel de caractériser le protéome de l'enveloppe chloroplastique afin de définir de nouveaux mécanismes moléculaires qui contrôlent la dynamique du chloroplaste. Bien qu'une part importante du protéome chloroplastique a été identifiée et caractérisée ces dernières années, l'enveloppe du chloroplaste reste la face cachée du protéome chloroplastique. En effet, en raison de leur faible abondance à l'échelle cellulaire (0,25% du protéome de la cellule entière) ou même chloroplastique (1-2% des protéines chloroplastiques)

[13], les protéines de l'enveloppe étaient historiquement, et restent, beaucoup moins connues que les autres protéines plastidiales. La protéomique sub-cellulaire et sub-plastidiale, reposant sur l'enrichissement préalable des compartiments sub-plastidiaux, a largement aidé à déchiffrer la composition protéique de l'enveloppe chloroplastique d'*Arabidopsis thaliana* et de quelques autres plantes [117, 125, 126]. Cependant, en raison de la contamination croisée des sous-compartiments chloroplastiques entre eux et par d'autres compartiments cellulaires, le fractionnement sub-cellulaire ne permettait pas de déterminer une localisation définitive pour de nombreuses protéines de l'enveloppe. En effet, l'enveloppe du chloroplaste est principalement contaminée par les protéines du stroma (environ 10%) ainsi que les protéines des thylakoïdes (environ 3%). Des approches réalisées par notre équipe (voir plus bas, base de données AT_CHLORO) ont permis d'aboutir à la localisation de plus d'un millier de protéines dans les trois compartiments majeurs du chloroplaste (envelope, strom et thylakoides) [117]. Toutefois, les protéines identifiées à ce jour par toutes les approches citées plus haut ne représentent encore qu'une fraction du protéome plastidial théorique.

3. Principales bases de données

Comme mentionnées ci-dessus, de nombreuses recherches sur le protéome plastidial se sont concentrées sur des fractions de chloroplastes ou sur des fractions sub-plastidiales indépendantes au cours des dernières années. En conséquence, plusieurs bases de données ont été créées au cours de la dernière décennie, telles que la base de données des protéines de plantes PPDB (http://ppdb.tc.cornell.edu) [127], créée dès 2004 et qui fournit une ressource intégrée pour les protéines identifiées expérimentalement chez Arabidopsis et le maïs, la base de données sur les protéines plastidiales plprot (http://www.plprot.ethz.ch) [128] créée en 2006 et qui rassemblent toutes les informations sur le protéome des plastes, les bases de données sub-cellulaire SUBA (http://suba3.plantenergy.uwa.edu.au/) et MASCP Gator (http://gator.masc-proteomics.org/) [129-132], créées respectivement en 2007 et 2011 et qui indiquent si les protéines ont déjà été détectées par MS dans des compartiments cellulaires spécifiques, si leur localisation sub-cellulaire a été déterminée en utilisant l'expression de protéines de fusion à la GFP *in planta* et si une localisation sub-cellulaire consensus peut être déduite de l'utilisation de tous les outils de prédiction sub-cellulaires disponibles, et finalement la base de données AT_CHLORO (http://at-chloro.prabi.fr/at_chloro/) [117, 120, 133] créée en 2010 et qui reste la seule base de données à

proposer des informations concernant la localisation sub-plastidiale (enveloppe, stroma, grana ou lamelle des thylakoïdes) des protéines chloroplastiques d'*Arabidopsis thaliana*.

4. AT-CHLORO database (voir Salvi et *al*., 2018)

Comme cité plus haut, AT_CHLORO est une base de données qui stocke des informations sur la localisation sub-plastidiale des protéines chloroplastiques d'*Arabidopsis thaliana*. Cette base de données a été créé par notre équipe en collaboration avec la plateforme de protéomique «EDyP» à partir des analyses protéomiques approfondies ciblant les différents compartiments sub-plastidiaux (enveloppe, stroma et thylakoïdes) du chloroplaste d'*Arabidopsis thaliana*. AT_CHLORO rassemble plusieurs informations complémentaires qui donnent un aperçu complet des connaissances actuelles sur la localisation sub-plastidiale et la fonction des protéines chloroplastiques identifiées. Une attention particulière a été accordée aux protéines identifiées dans les fractions de l'enveloppe chloroplastique, pour lesquelles une annotation fonctionnelle exhaustive a été réalisée. Depuis la première version de la base de données en 2010, le contenu de la base AT_CHLORO a été considérablement amélioré par des études complémentaires qui ont été réalisées pour approfondir la cartographie du protéome chloroplastique. En effet, depuis 2012, la base de données fournit un aperçu graphique des sous-ensembles de protéines identifiées à partir de recherches spécifiques. Depuis 2013, AT_CHLORO stocke également des informations sur les protéines liant la calmoduline identifiées dans les fractions chloroplastiques en utilisant des analyses protéomiques [134]. Enfin, depuis 2015, AT_CHLORO fournit également des informations sur l'enrichissement spécifique des protéines dans les principaux compartiments (grana et stroma-lamellae) des membranes des thylakoïdes d'*Arabidopsis thaliana* [120]. La base de données AT_CHLORO est donc régulièrement mise à jour avec des données expérimentales internes et des informations liées aux protéines chloroplastiques, avec une attention plus spécifique pour les protéines de l'enveloppe. Sachant qu'il existe un lien direct entre la localisation sub-cellulaire des protéines et leurs fonctions respectives, l'information sur la localisation sub-plastidiale des protéines ouvre de nouvelles possibilités pour sonder des hypothèses fonctionnelles sur ces composants plastidiaux encore non caractérisés. Au-delà des informations concernant la localisation sub-cellulaire et sub-plastidiale des protéines, la base de données AT_CHLORO stocke des informations donnant accès à des études quantitatives à savoir, la comparaison de plantes mutantes et de plantes sauvages, ou de plantes cultivées en conditions de croissance variables.

IV. Objectifs de la thèse

L'objectif de mon projet de thèse était d'identifier des composants des systèmes d'import alternatifs de protéines chloroplastiques « non canoniques » (indépendants des systèmes TIC/TOC), dont l'existence a été mise en évidence par notre équipe dès 2007. Deux stratégies ont alors été envisagées.

Une approche directe (biochimie) qui reposait sur des techniques biochimiques combinant le pontage chimique, la purification par affinité et la spectrométrie de masse. Cette stratégie avait pour objectif d'identifier des récepteurs d'import localisés à la surface des chloroplastes et requis pour l'import de protéines contenant des peptides de transit non canoniques et étude de leur rôle dans le contrôle de la physiologie de la plante.

Une approche alternative (gènes candidats) qui portait sur la bio-analyse du protéome de l'enveloppe du chloroplaste et qui impliquait de revisiter la composition du protéome de l'enveloppe du chloroplaste à l'aide d'approches nouvelles reposant sur la quantification des protéines grâce à une plus grande sensibilité des équipements actuels, et a été étendu à identifier des nouveaux mécanismes moléculaires qui pourraient être impliqués dans le contrôle de la localisation sub-cellulaire des protéines des plastes adressées à plusieurs compartiments sub-cellulaires.

Le premier volet de mon travail (chapitre I) a été consacré à l'identification des candidats récepteurs à la surface des chloroplastes qui sont nécessaires pour l'import des protéines chloroplastiques contenant des peptides de transit non canoniques. Pendant la première année de thèse, j'ai réussi à produire les protéines chloroplastiques contenant des peptides de transit non canoniques (Tic32, FSD1, ceQORH et NQR). Ces protéines ont ensuite été utilisées comme appâts dans les études d'interaction protéine/récepteur. Ensuite, après optimisation et expérimentations préliminaires, une purification par affinité des récepteurs d'import (complexes générés par la réaction de pontage chimique) a été réalisée. Une approche reposant sur de la spectrométrie de masse quantitative a été ensuite adoptée afin d'identifier des peptides de récepteurs spécifiquement enrichis dans les fractions éluées à partir de la matrice d'affinité. Après traitement de données protéomiques générées par la spectrométrie de masse, plusieurs candidats « partenaires » ont été identifiés.

La deuxième année de thèse a été donc principalement consacrée à la validation *in planta* de la nature de ces candidats sélectionnés (issus de la réaction de pontage) par étude de l'impact des mutations sur la localisation sub-cellulaire des protéines étudiées (Tic32, FSD1, ceQORH et NQR) par microscopie confocale. Pour cela, des lignées d'insertions d'Arabidopsis affectées dans l'expression de ces récepteurs putatifs ont été isolées. Les mutants homozygotes ont été ensuite transformés par les versions étiquetées par la protéine GFP des protéines chloroplastiques non canoniques correspondantes, afin de valider l'implication effective des candidats récepteurs identifiés dans le contrôle de l'adressage alternatif des protéines au chloroplaste en étudiant l'impact de ces mutations sur la localisation sub-cellulaire des protéines étudiées par microscopie confocale. Il s'est avéré que, parmi tous les candidats sélectionnés et déjà analysés, une protéine désignée « ClpD » avait un impact sur la localisation de la protéine ceQORH. La troisième année de thèse a été donc principalement consacrée à la validation du rôle de cette protéine « ClpD » sur la physiologie de la plante par des analyses phénotypiques sur le mutant correspondant. Cette étude sera suivie par l'application d'approches de protéomique différentielle à l'échelle du chloroplaste entier de plantes WT versus des mutants (ΔClpD) visant à déchiffrer la nature des protéines dont les niveaux d'équilibre seront affectés chez le mutant par rapport au WT.

Des échantillons d'enveloppe sans pontage chimique ont été aussi analysés (qui ont servi pour contrôles). En conséquence, en parallèle à la recherche des récepteurs potentiels par l'approche du pontage chimique, les études protéomiques réalisées ont également généré une masse de données originales qui permettaient d'avoir une vision quantitative de la composition protéique de l'enveloppe des chloroplastes.

Le deuxième volet de mon travail (chapitre II) portait ainsi sur la mise à jour de la composition du protéome de l'enveloppe du chloroplaste en reposant sur *i*) l'enrichissement des protéines dans des fractions purifiées de l'enveloppe des chloroplastes par rapport à l'extrait cellulaire total et *ii*) sur la sensibilité accrue des appareils de détection. Ces travaux ont permis de compléter les bases de données actuelles et mettre en évidence un grand nombre de protéines qui semblent partagées entre l'enveloppe du chloroplaste, le réticulum endoplasmique et le Golgi. Ces dernières pourraient être impliquées dans des interactions entre ces divers compartiments cellulaires ouvrant des pistes de recherche très intéressantes.

Sur la base de ce protéome revisité, a été initiée une bioanalyse de l'enveloppe du chloroplaste afin d'identifier des nouvelles protéines mineures spécifiques de l'enveloppe et qui pourraient être compatibles avec un rôle de récepteurs d'import des protéines non canoniques des chloroplastes. Cette approche est encore en cours.

Le troisième volet de mon travail (chapitre III) concerne mon implication, sur les deux dernières années de ma thèse, dans un autre projet de l'équipe qui cible toujours sur le contrôle de l'adressage des protéines au chloroplaste. Comme évoqué ci-dessus (section III.1.a), depuis plusieurs année l'équipe étudie une protéine chloroplastique très originale : la protéine ceQORH. Cette protéine est multilocalisée notamment dans les cellules de l'épiderme de la feuille où la protéine est essentiellement localisée dans la membrane plasmique. Afin d'expliquer la localisation sub-cellulaire variable de la protéine ceQORH, les membres de l'équipe avaient émis l'hypothèse d'une interaction probable de cette protéine ceQORH avec un partenaire cytosolique. Ils ont ensuite validé l'interaction, *in planta*, de la protéine chloroplastique ceQORH avec cette protéine cytosolique, une calmoduline, puis déterminé le domaine de la protéine ceQORH qui est responsable de l'interaction avec cette calmoduline. La construction d'une version mutée (« Mut2») de la protéine ceQORH n'interagissant plus avec la calmoduline leur a ensuite permis de démontrer *in planta* le rôle direct de cette interaction dans le contrôle de l'adressage de ceQORH aux chloroplastes.

Ma participation à ce projet a reposé sur une approche de génétique inverse qui a consisté à confirmer l'identité de l'isoforme de calmoduline impliquée dans ce mécanisme. Pour ce faire, j'ai isolé un mutant KO d'Arabidopsis affecté dans l'expression de cette isoforme. Le mutant homozygote (*cam5*) a été ensuite transformé par les versions étiquetées par la protéine GFP des protéines ceQORH sauvage et ceQORH mutée (Mut2, qui ne lie plus la calmoduline) et la localisation de ces protéines de fusion a été visualisée par microscopie confocale. Ces expériences ont permis de confirmer que c'est bien le produit du gène *cam5* qui régule la localisation sub-cellulaire de la protéine ceQORH chez Arabidopsis.

Matériel et méthodes

I. Préparation de fractions subcellulaires d'*Arabidopsis thaliana*

1. Choix du matériel végétal

Les travaux décrits dans ce rapport ont été réalisés à partir de feuilles d'*Arabidopsis thaliana* sauvage d'écotype Wassilevskija (Ws). La crucifère *Arabidopsis thaliana* est une plante modèle. Son petit génome est entièrement séquencé (The AGI, 2000). Elle possède un temps de génération court et de nombreux mutants d'insertion sont disponibles. Les graines d'*Arabidopsis thaliana* sont directement semées en terre dans des pots qui sont placés en chambre de culture à 22°C, sous lumière, avec une photopériode de 10 (jours courts) ou 16 (jours longs) h de jour.

2. Préparation des chloroplastes intacts d'*Arabidopsis thaliana* (Seigneurin-Berny *et al.,* 2008)

Les cultures d'*Arabidopsis thaliana* (écotype Ws) sont réalisées à une température de 22°C, avec une photopériode fixée à 10 h de jour et sous une intensité lumineuse de 150 µmoles/m^2/s. Les plantes sont récoltées pour le fractionnement subcellulaire après 5 à 6 semaines de croissance. Les chloroplastes d'*Arabidopsis thaliana* sont purifiés selon Seigneurin-Berny [2008] : les chloroplastes sont obtenus à partir de 400 à 500 g de feuilles d'*Arabidopsis thaliana* après purification par centrifugation sur gradient continu de Percoll. Les feuilles d'*Arabidopsis thaliana* sont hachées dans 2 L de milieux de broyage (sorbitol 0,4 M, Tricine KOH 20 mM pH 8,4, EDTA 100 mM pH 8, $NaHCO_3$ 40 mM, SAB 1 g/L, ajusté à pH 8,4) grâce à un broyeur à hélice (type Waring Blendor, volume 4 L, broyage pendant 2 à 3 s). Le broyat est ensuite filtré sur 4 épaisseurs de gaze et une toile à bluter de 50 µm de vide de maille. Cette opération est effectuée deux fois de suite. Le broyat est alors filtré et centrifugé à 2070 *g* pendant 2 min (Sorvall RC-5, rotor SLA3000). Les culots sont délicatement resuspendus dans le milieu de lavage (sorbitol 0,2 M, Tricine/KOH 20 mM pH 7,6, EDTA 100 mM pH 8, $MgCl_2$ 10 mM, ajusté à pH 7,6). Les gradients continus de Percoll (Percoll 50% [v/v], sorbitol 0,2 M, Tricine KOH 20 mM pH 7,6, EDTA 100 mM pH 8, $MgCl_2$ 10 mM, ajusté à pH 7,6) sont réalisés par centrifugation à 38700 *g* pendant 55 min (Sorvall RC-5, rotor SS34). La suspension précédemment obtenue, enrichie en chloroplastes, est déposée sur les gradients de Percoll. Les gradients sont alors centrifugés à 13240 *g* pendant 10 min (Sorvall RC-5, rotor HB-6). Les chloroplastes intacts sont collectés dans le tiers le plus bas des gradients de

Percoll. Pour éliminer le Percoll, les chloroplastes intacts sont dilués par 3 à 4 volumes de milieu de lavage (sorbitol 0,2 M, Tricine/KOH 20 mM pH 7,6, EDTA 100 mM pH 8, $MgCl_2$ 10 mM, ajusté à pH 7,6) et centrifugés à 2070 *g* pendant 2 min (Sorvall RC-5, rotor SLA3000). Toutes les opérations sont effectuées à 4°C. Le culot obtenu après cette étape correspond aux chloroplastes intacts purifiés.

3. Purification de l'enveloppe des chloroplastes (voir Bouchnak *et al.,* 2018)

Les chloroplastes intacts et purifiés sont lysés dans un milieu hypotonique (MOPS/KOH 10 mM, pH 7,8, $MgCl_2$ 4 mM) qui contient des inhibiteurs de protéases (acide ε-aminocaproïque 0,5 mM, benzamidine 1 mM et PMSF 1 mM). La suspension de chloroplastes cassés est déposée sur un gradient discontinu de saccharose (saccharose 0,3 M, 0,6 M et 0,93 M dans du MOPS/KOH 10 mM, pH 7,8, $MgCl_2$ 4 mM) et centrifugée à 68500 *g* pendant 1 h (Beckman LE 80K, rotor SW41). Après centrifugation, les protéines du stroma des chloroplastes sont récupérées en haut du gradient, l'enveloppe des chloroplastes est collectée à l'interface des couches de saccharose 0,6 M et 0,93 M et le culot correspond aux thylacoïdes.

Les fractions correspondant à l'enveloppe et les thylacoïdes sont diluées 3 fois avec du milieu de lavage (MOPS/KOH 10 mM pH 7,8, acide ε-aminocaproïque 0,5 mM, benzamidine 1 mM et PMSF 1 mM), puis concentrées par centrifugation à 107170 *g* pendant 1 h (Beckman LE 80K, rotor SW41). Les thylacoïdes et l'enveloppe ainsi obtenus sont mis en suspension (MOPS/KOH 10 mM, pH 7,8, acide ε- aminocaproïque 0,5 mM, benzamidine 1 mM et PMSF 1 mM) et conservés dans l'azote liquide. Toutes les opérations sont effectuées à 4°C.

II. Techniques d'étude des protéines

1. Extraction des protéines totales de feuilles d'*Arabidopsis thaliana*

Deux feuilles sont broyées dans 200 µL de tampon de broyage (pyrophosphate de sodium 50 mM, Tris, pH 6,8). Le broyat obtenu est centrifugé 5 min à 130000 *g* (Eppendorf 5415 D). Le surnageant ainsi obtenu contient les protéines totales de la feuille.

2. Dosage des protéines

La concentration des protéines est déterminée selon la méthode de Bradford (1976). Cette méthode est basée sur la modification de la longueur d'onde d'absorption du bleu de Coomassie G-250 en milieu acide après sa fixation sur les acides aminés. Cela s'effectue à l'aide du kit Bio-Rad Protein assay (Bio-rad) suivant les indications du fournisseur. 200 µL de réactif de Bradford (BioRad) sont mélangés à 800 µL d'eau et à 2 µL de l'extrait protéique à quantifier et l'absorbance est mesurée à 595 nm. Cette mesure de DO permet de déterminer la quantité de protéines présentes en se référant à une gamme étalon de BSA établie avec des quantités connues. Il faut noter que les échantillons contenant des détergents tels que le SDS ou le Triton X-100 en quantité importante ne peuvent être quantifiés par cette méthode car ces derniers interfèrent avec le réactif de Bradford. La quantité de protéines d'enveloppe des chloroplastes utilisées lors des différentes expériences a été évaluée par cette méthode.

3. Analyse des protéines par électrophorèse sur gel de polyacrylamide

a) Séparation par électrophorèse en conditions dénaturantes

La séparation électrophorétique des protéines est effectuée selon la méthode de Laemmli (1970). Cette technique permet de séparer les protéines, ou leurs sous-unités constitutives, selon leur taille et d'estimer ainsi leur poids moléculaire par comparaison avec la mobilité électrophorétique de protéines de référence. Les protéines ont été séparées par migration sur un gel de polyacrylamide constitué d'un gel de concentration contenant 5% [v/v] d'acrylamide et d'un gel de séparation contenant 10% [v/v] ou 12% [v/v] d'acrylamide. Pour l'électrophorèse SDS-PAGE, les protéines à analyser sont dénaturées pendant 5 min à 95°C dans le tampon de dénaturation (Tris/HCl 50 mM pH 6,8, glycérol 10% [v/v], SDS 1% [v/v], bleu de bromophénol 0,01% [m/v] et DTT 25 mM), avant d'être déposées dans les puits du gel de polyacrylamide. La migration électrophorétique des protéines est réalisée dans un tampon de migration (Tris 25 mM, glycine 0,192 M pH 8,3, SDS 0,1% [v/v]) à voltage constant (120 V). La migration est arrêtée lorsque le bleu de bromophénol atteint le bas du gel. Après l'électrophorèse, les protéines peuvent alors être colorées directement dans le gel de polyacrylamide ou transférées sur une membrane de nitrocellulose.

b) Coloration des protéines au bleu de Coomassie

Le gel de polyacrylamide est coloré par incubation d'une heure dans la solution de coloration : acide acétique 10% [v/v], isopropanol 25% [v/v], bleu de Coomassie 2,5 g/L (Coomassie brilliant-blue R250, Sigma). Il est ensuite décoloré plusieurs fois dans un mélange éthanol 40% [v/v], acide acétique 7% [v/v] pour révéler les protéines présentes dans le gel.

4. Transfert des protéines sur membrane de nitrocellulose et immunomarquage

Cette technique consiste à révéler une protéine particulière à l'aide d'anticorps spécifiquement dirigés contre cette protéine. Elle nécessite la fixation des protéines sur une membrane.

a) Transfert des protéines

Cette technique permet de transférer électrophorétiquement les protéines séparées par électrophorèse sur gel de polyacrylamide en condition dénaturante sur une membrane de nitrocellulose (BA 85 (0,45 µm), GE Healthcare Life Sciences Whatman). Le transfert sur membrane est une étape préalable pour la réalisation de l'immunomarquage. Après migration, le gel est incubé dans le tampon de transfert (Tris 25 mM, glycine 0,192 M pH 8,3, SDS 0,1% [v/v], éthanol 20% [v/v]). La membrane de nitrocellulose est mouillée dans le tampon de transfert et déposée sur 3 feuilles de papier Whatman, préalablement trempées dans le même tampon. Le gel est ensuite déposé sur la membrane et recouvert par 3 feuilles de papier Whatman imbibées de tampon. Une pipette Pasteur est utilisée comme rouleau sur l'ensemble du montage afin de chasser les bulles d'air qui affectent le transfert. Le tout est placé dans une cassette et installé dans l'appareil de transfert rempli de tampon de transfert. Le transfert s'effectue à 80 V pendant 1 h 30 à 4°C.

b) Fixation des anticorps

Après transfert des protéines, la membrane de nitrocellulose est rincée dans du tampon TBST (TBS (NaCl 150 mM, Tris/HCl 50 mM pH 7,5) + Triton X-100 0,05% [m/v]) puis saturée dans le même tampon contenant du lait écrémé déshydraté 5% [m/v] pendant 1 h, sous agitation. La membrane est ensuite incubée avec le premier anticorps dilué dans du TBST-Lait, 3 h à température ambiante ou pendant la nuit à 4°C. La membrane est rincée 3 fois 10 min dans du tampon TBST, puis incubée

avec le deuxième anticorps (anticorps dirigé contre les immunoglobulines G de lapin ou de souris, couplé à la peroxydase (Jackson ImmunoResearch, dilution 1/10000ème) pendant 1 h à température ambiante et sous agitation.

c) Révélation par mesure de la chimiluminescence

Après incubation en présence du deuxième anticorps, la membrane est lavée 3 fois 10 min dans du tampon TBST. L'activité de la peroxydase est révélée dans un milieu alcalin (Tris/HCl 100 mM, pH 8,5) auquel sont ajoutés extemporanément du luminol (3-aminophalhydrazine) 12,5 mM et de l'acide p-coumarique 0,2 mM, en présence d'une goutte d'eau oxygénée. En milieu alcalin, la peroxydase catalyse l'oxydation du luminol. Le luminol oxydé est dans un état excité et revient à son état fondamental en émettant de la lumière bleue (chimiluminescence). L'acide p-coumarique, composé phénolique, joue le rôle de catalyseur et améliore l'émission de lumière d'un facteur 100 (Durrant, 1990). La membrane de nitrocellulose est incubée 1 min avec la solution de révélation, puis est exposée par l'utilisation d'un ChemiDoc ™ (BioRad) pendant des temps d'exposition variables (de 5 s à 15 min) suivant l'intensité du signal.

III. Techniques d'étude des acides nucléiques

1. Extraction de l'ADN à partir d'un gel d'agarose

Afin de purifier un fragment d'ADN issu d'une amplification par PCR ou d'une digestion, les différents fragments sont séparés par électrophorèse sur gel d'agarose. Les fragments sont localisés sous UV et celui à purifier est excisé. Nous avons utilisé le système NucleoSpin Gel and PCR Clean-up (Macherey-Nagel) pour purifier l'ADN. Cette purification repose sur l'interaction spécifique de l'ADN avec une matrice d'affinité (NucleoSpin Gel and PCR Clean-up column).

2. Synthèse d'oligonucléotides

Des oligonucléotides de synthèse ont été utilisés comme amorces pour amplifier les fragments d'ADN clonés dans les différents plasmides par réaction de polymérisation en chaine (PCR). La synthèse des oligonucléotides a été réalisée par la société Eurofins Genomics.

3. Amplification *in vitro* de fragments d'ADN

L'amplification *in vitro* de fragments d'ADN à partir d'une banque d'ADNc obtenue au laboratoire à partir de feuilles de jeunes plantes d'*Arabidopsis thaliana* a été réalisée par PCR. L'ADN matrice (50 ng) est dilué dans 50 µL de milieu de PCR contenant Tris-HCl 10mM pH 8,3, KCl 50 mM, $MgCl_2$ 1,5 mM, DTT 1 mM, Tween-20 0,05% [v/v], des désoxyribonucléotides triphosphate (dNTP) 200 µM, 14 nmoles de chaque amorce spécifique et une unité d'ADN polymérase Phusion™ High-Fidelity (Thermo Scientific). L'amplification est arrêtée après 35 cycles. La réaction commence par 8 min de pré-dénaturation à 95°C. Pour une raison de spécificité, la réaction commence par 10 cycles « touch down » comprenant 30 s de dénaturation à 95°C, 30 s d'hybridation en allant de 65°C à 55°C en raison de 1°C par cycle suivie d'une minute d'élongation à 72°C. Puis 25 cycles, comprenant chacun 30 s de dénaturation à 95°C, 30 s d'hybridation de 55°C, 1 min d'élongation à 72°C. L'efficacité de l'amplification est testée par analyse de 10 µL du mélange réactionnel par électrophorèse sur gel d'agarose 1,5%.

4. Séquençage de l'ADN

Les séquençages ont été réalisés par la société Eurofins Genomics. Les séquences obtenues sont ensuite analysées à l'aide de banques de données et de différents programmes accessibles sur Internet (Expasy tools (https://web.expasy.org/translate/), Multalin (http://multalin.toulouse.inra.fr/multalin/), …).

5. Digestion d'ADN par des endonucléases

L'ADN à digérer (quelques centaines de ng) est mélangé à 20 U d'enzymes de restriction d'intérêt (New England Biolabs) placées dans un tampon compatible avec leur activité. La digestion se fait à 37°C ou à 25°C pendant 1 à 16 h. Le produit de digestion est ensuite analysé par électrophorèse sur gel d'agarose.

IV. Techniques d'expression procaryote

1. Construction de vecteurs d'expression des protéines recombinantes

Les régions codantes pour les protéines Tic32, ceQORH, NQR et FSD1 ont été obtenues à partir d'amplification par PCR sur de l'ADN génomique d'*Arabidopsis thaliana* en utilisant des

oligonucléotides (Eurofins Genomics) spécifiques. Ces oligonucléotides ont permis d'introduire les sites de restriction *Nde*I et *Xho*I aux extrémités 5' et 3' des ADNc codant pour les protéines Tic32, ceQORH, NQR et FSD1. Les fragments amplifiés ont été insérés par un clonage à bouts francs dans un plasmide intermédiaire, le plasmide pBlueScript (pKS), afin d'être entièrement séquencés. La digestion de ce plasmide par les enzymes de restriction *Nde*I et *Xho*I a permis d'isoler les inserts afin de vérifier leurs séquences.

La région codante pour la fusion GFP a été obtenue à partir d'amplification par PCR sur le plasmide commercialisé pUC-GFP [135] en utilisant des amorces contenant les sites de restriction *Xho*I et *BamH*I. Le fragment amplifié a été inséré par un clonage à bouts francs dans le plasmide pKS. La digestion de ce plasmide par les enzymes de restriction *Xho*I et *Bam*HI a permis d'isoler l'insert qui code pour la protéine de fusion GFP afin de l'introduire dans le vecteur d'expression.

La construction des recombinaisons Tic32::GFP, ceQORH::GFP, NQR::GFP, et FSD1::GFP a été réalisée par clonage à bouts cohésifs, dans le plasmide d'expression pET-20b qui a été préalablement digéré par *Nde*I et *Bam*HI pour la forme fusionnée à la séquence poly-histidine.

La souche DH5α a été utilisée pour la production de plasmides recombinants. Cette souche contient la partie α de la β-galactosidase et permet d'avoir une sélection blanc/bleu avec des plasmides contenant la partie complémentaire de la β-galactosidase.

2. Préparation de l'ADN plasmidique

a) Minipréparation d'ADN plasmidique

Le kit de minipréparation NucleoSpin Plasmid (Macherey-Nagel) a été utilisé. Il repose sur le principe de la lyse alcaline des bactéries suivie d'une purification sur colonne. Les termes mini et midi désignent le volume de départ de la culture bactérienne. Pour une minipréparation, il s'agit de 3 mL de culture bactérienne. Les bactéries recombinantes sont cultivées (sous agitation à 250 rpm, à 37°C) pendant 12 h dans du milieu LB contenant 100 mg/mL de carbénicilline. Le protocole expérimental fourni avec le kit a été suivi à la lettre.

b) Midipréparation d'ADN plasmidique

Le kit NucleoBond Xtra Midi/Maxi (Macherey-Nagel) a été utilisé pour la préparation d'un stock

de l'ADN plasmidique (pET-20b et pKS) à partir de cultures bactériennes de 50 mL. Toutes les étapes de purification sont réalisées dans les conditions préconisées par le fabriquant.

3. Préparation et transformation des bactéries compétentes

Les bactéries compétentes (c'est à dire aptes à intégrer l'ADN étranger) sont préparées par un traitement au $CaCl_2$ (Maniatis et *al*., 1982).

A partir d'une colonie bactérienne isolée sur boite LB agar (Carl Roth), les bactéries (DH5α, ROSETTA) sont cultivées dans 5 mL de milieu LB (250 rpm, 37°C) pendant la nuit, suivie d'une dilution au 1:100 dans 500 mL de milieu LB (250 rpm, 37°C) jusqu'à une absorbance à 600 nm de 0,5 à 0,6. Les bactéries sont ensuite centrifugées 10 min à 4000 *g* (SLA 3000) et remises en suspension dans 100 mL de $MgCl_2$ 100 mM froid. Après 20 à 30 min dans la glace, une nouvelle centrifugation de 10 min à 4000 *g* a été réalisée, le culot de bactéries est repris dans 10 mL de $CaCl_2$ 100 mM froid. A ce stade, les cellules sont considérées comme compétentes. Des aliquotes de 50 µL ont été préparés et immédiatement congelés dans l'azote liquide puis stockés à -80°C.

Les plasmides (environ 100 ng) sont mélangés à 100 µL de bactéries compétentes pendant 25 min dans la glace. Un choc thermique de 1,5 min à 42°C permet la transformation des bactéries par les plasmides. La suspension bactérienne est alors incubée 1 h (250 rpm, 37°C) dans 0,6 mL de milieu LB, puis étalée sur boîte de Pétri contenant du milieu LB et de la carbénicilline (concentration finale : 100 mg/mL). Dans ces conditions, seules les bactéries ayant incorporé un plasmide se multiplient.

4. Conditions d'expression

Les bactéries recombinantes (ROSETTA) sont mises en culture dans du milieu LB à 37°C jusqu'à ce que l'absorbance à 600 nm soit comprise entre 0,4 et 0,8. L'induction de la synthèse de la protéine recombinante (Tic32::GFP, ceQORH::GFP, NQR::GFP, et FSD1::GFP) est déclenchée par ajout d'IPTG 1mM. Lorsque la protéine synthétisée est insoluble (corps d'inclusion) en condition standard, diverses conditions d'induction sont testées. Ces conditions incluent des variations de température (18, 30, 37 et 42°C), diverses concentrations d'IPTG (0, 0,1 et 0,4 mM) et des durées d'induction variables.

5. Vérification de la solubilité des protéines

Après induction de l'expression dans les bactéries ROSETTA et avant purification, une étape de vérification de la solubilité de toutes les protéines recombinantes (Tic32::GFP, ceQORH::GFP, NQR::GFP, et FSD1::GFP) a été réalisée. Diverses conditions d'induction sont testées. Au temps optimal déterminé pour l'expression, la suspension de cellules est centrifugée 20 min à 180 *g*. Le culot de cellules est remis en suspension dans un tampon contenant (Tris/HCl 20 mM pH 7,9, imidazole 5 mM, NaCl 0,5 M) et lysé pendant trois cycles successifs de sonication à 0°C (120 V, 1 min) afin de casser la totalité des cellules.

Cette suspension est ensuite centrifugée 40 min à 16000 *g* afin de distinguer une fraction membranaire et une fraction soluble. Les deux fractions sont ensuite analysées par électrophorèse en conditions dénaturantes et par marquage immunologique.

6. Technique de purification des protéines recombinantes contenant une extension *C*-terminale poly-histidine

Les protéines recombinantes exprimées sous forme de fusion avec une séquence polyhistidine en *C*-terminal peuvent être purifiées par chromatographie d'affinité. En effet, lorsque les histidines sont déprotonées à pH 7,9, la séquence poly-histidine se lie, avec une forte affinité, aux cations bivalents (Ni^{2+}) immobilisés sur des colonnes de chélation des métaux.

Les protéines non fixées sont alors éliminées par des lavages successifs et la protéine étiquetée est éluée de la colonne par des tampons contenant des concentrations croissantes d'imidazole.

a) Préparation des protéines

Un culot de bactéries contenant la protéine recombinante conservé à -80°C (correspondant à 500 mL de culture induite) est dilué dans du tampon de fixation froid (Tris/HCl 20 mM pH 7,9, imidazole 5 mM, NaCl 0,5 M) (30 mL de tampon pour 500 mL de culture bactérienne). Les bactéries sont lysées et leur ADN est détruit par des cycles successifs de sonication à 0°C (120 V, 1 min). La sonication est poursuivie jusqu'à ce que le milieu devienne limpide. Le lysat est ensuite centrifugé à 39000 *g* pendant 40 min à 4°C (Sorvall RC-5, rotor SS34) afin d'éliminer les débris cellulaires. Le surnageant qui contient les protéines solubilisées est immédiatement déposé sur la colonne d'affinité.

b) Préparation de la colonne

La matrice (Novagen) est rincée dans 3 volumes (volume de colonne) d'eau, chargée en nickel par 5 volumes de tampon de charge ($NiSO_4$ 50 mM) et équilibrée par 3 volumes de tampon de fixation (Tris/HCl 20 mM pH 7,9, imidazole 5 mM, NaCl 0,5 M).

c) Chromatographie d'affinité

L'échantillon protéique est déposé sur la matrice, le débit utilisé est de l'ordre de 10 volumes de colonne/h. Les protéines peu ou non affines pour les ions Ni^{2+} sont éluées par des concentrations croissantes d'imidazole allant de 20 mM à 50 mM. La colonne est lavée successivement par 10 volumes de tampon de fixation, puis 10 volumes de tampon de lavage (Tris/HCl 20 mM pH 7,9, imidazole 60 mM, NaCl 0,5 M). L'élution s'effectue par addition de 6 volumes de tampon d'élution (Tris/HCl 20 mM pH 7,9, imidazole 1 M, NaCl 0,5 M). Les fractions d'élution sont collectées (0,2 mL par fraction). Les protéines contenues dans les différentes fractions sont ensuite analysées par électrophorèse en conditions dénaturantes et par marquage immunologique.

V. Techniques d'identification des récepteurs à la surface des chloroplastes

1. Pontage chimique de protéines recombinantes avec les protéines de l'enveloppe de chloroplastes

Cette technique consiste à piéger les récepteurs à la surface des chloroplastes avec leurs substrats en protéines recombinantes (Tic32::GFP, ceQORH::GFP, NQR::GFP, et FSD1::GFP) en utilisant un agent chimique (crosslinker). Diverses conditions sont testées afin d'optimiser les conditions idéales de la réaction. Ces conditions incluent des changements de type de crosslinker utilisé (AMAS, BS^3 et DMS), diverses concentrations de crosslinker (0, 1 mM), des quantités en protéines variables et des variations de l'origine de l'enveloppe (Arabidopsis ou épinard). Toutes les opérations sont effectuées à 4°C avec agitation (roue) et à l'abri de la lumière. Pour la validité statistique des résultats, trois répliques biologiques ont été réalisés pour chaque protéine recombinante.

a) Préparation de l'enveloppe de chloroplastes

L'enveloppe de chloroplastes d'*Arabidopsis thaliana* initialement purifiée et conservée dans le tampon (MOPS/KOH 10 mM, pH 7,8, acide ε- aminocaproïque 0,5 mM, benzamidine 1 mM et PMSF 1 mM) est lavée dans le PBS-EDTA 1mM. La suspension est ensuite ultracentrifugée 15 min à 50000 *g* (SORVALL-Discovery M150SE, S55A2). Le culot est de nouveau repris dans le même tampon PBS-EDTA 1 mM, puis traité à l'ATP 0,1 mM pendant 30 min à 4°C avec agitation (roue) et à l'abri de la lumière, afin de favoriser les conditions idéales pour assurer l'interaction protéine recombinante-enveloppe de chloroplaste.

b) Interaction protéine recombinante-enveloppe de chloroplaste

Après incubation en présence d'ATP, la suspension homogène est ensuite répartie dans 4 tubes à raison de 60 µg de protéines d'enveloppe dans chaque tube, puis incubée avec les protéines recombinantes (Tic32::GFP, ceQORH::GFP, NQR::GFP, et FSD1::GFP) pendant 15 min, à 4°C et à l'abri de la lumière. Après interaction, une ultracentrifugation est réalisée pendant 15 min à 50000 *g* (SORVALL-Discovery M150SE, S55A2) afin d'éliminer l'excès de protéines recombinantes qui n'ont pas interagi avec l'enveloppe de chloroplastes. Les complexes protéiques ainsi formés sont ensuite récupérés dans le PBS-EDTA 1 mM avant d'être utilisés pour les expériences de pontage.

c) Réaction de pontage chimique

Le réactif chimique utilisé est le BS^3 ou bis (sulfosuccinimidyl) subérate (Thermo Scientific), il crée un lien covalent entre groupements amines des acides aminés (hors ceux impliqués dans la liaison peptidique) et le groupement amine des protéines.

Les complexes protéiques ainsi dilués dans du tampon PBS-EDTA 1 mM sont de nouveau incubés en présence d'ATP 0,1 mM et par ajout de BS^3 à une concentration finale de 1 mM. Le BS^3 est préparé extemporanément dans l'eau. Après 2 h d'incubation à 4°C avec agitation et à l'abri de la lumière, la réaction est arrêtée par l'ajout de la glycine 50 mM afin de neutraliser l'excès de BS^3.

2. Technique de purification des complexes protéiques issus de pontage chimique contenant une étiquette GFP

Le kit de purification µMACS™ Epitope Tag Protein Isolation Kits (Miltenyi Biotec) a été utilisé. Il utilise le principe de l'immunopurification à l'aide d'anticorps greffés sur billes magnétiques. Le protocole expérimental fourni avec le kit a été suivi à la lettre.

a) Solubilisation des complexes protéiques et fixation des billes magnétiques

Après 15 min d'incubation à 4°C en présence de glycine, la suspension est solubilisée dans du dodécylmaltoside (DDM) 1% [m/v] pendant 45 min, puis traitée par ajout des anticorps sous formes de billes magnétiques (µMACS™ MicroBeads) dirigés contre l'étiquette GFP. Après 35 min d'incubation dans la glace, le mélange est immédiatement déposé sur la colonne d'affinité.

b) Préparation de la colonne

La colonne (µcolumns, MACS Molecular) est rincée extemporanément dans 200 µL de tampon de suspension (PBS-EDTA 1 mM ; DDM 1% [m/v]).

c) Chromatographie d'affinité

L'échantillon protéique est déposé sur la colonne, les contaminants protéiques sans affinité pour les anticorps greffés sur les billes magnétiques sont élués par deux lavages successifs de la colonne en utilisant deux solutions de lavage fournies dans le kit (WB1 contenant NaCl 150 Mm, Igepal CA-630 1% [m/v], deoxycholate de sodium 0,5% [m/v], SDS 0,1% [v/v], Tris HCl 50 mM pH 8,0 et WB2 contenant du Tris HCl 20 mM pH 7,5. L'élution s'effectue dans les conditions réductrices par addition de 20 µL de tampon d'élution (Tris HCl 50 mM pH 6,8, DTT 50mM, SDS 1% [v/v], EDTA 1 mM, Bleu de bromophénol 0,005% [m/v], glycérol 10% [v/v]) préalablement chauffé pendant 5 min à 95°C. Après incubation pendant 5 min, la fraction d'élution est collectée par addition de 50 µL de tampon d'élution. A l'issue de ces purifications, quelques microlitres des élutions de la colonne d'affinité ont été analysées par western-blot afin de visualiser les complexes protéiques formés. Le reste de ces fractions d'élution est ensuite analysé par spectrométrie de masse afin d'identifier les récepteurs à la surface des chloroplastes.

VI. Caractérisation de mutants d'insertion par PCR

1. Extraction de l'ADN à partir de feuilles d'Arabidopsis

Deux jeunes feuilles de la même rosette du mutant d'Arabidopsis sont broyées dans l'azote liquide dans 400 µL de tampon de broyage (tampon Edwards : 20 mM Tris HCl, pH 7,5, 250 mM NaCl, 25 mM EDTA, 0,5% SDS). Le broyat obtenu est centrifugé 5 min à 130000 *g* à 12°C (Eppendorf 5415 D). Le surnageant ainsi obtenu (300 µL) est transféré dans un nouveau tube eppendorf et précipité par ajout de 300 µL d'isopropanol. Après précipitation pendant 2 min à température ambiante, le contenu est centrifugé 5 min à 130000 *g* à 12°C (Eppendorf 5415 D). Le petit culot blanc ainsi obtenu contient l'ADN. Une étape de lavage du culot obtenu avec 150 µL d'éthanol est ensuite réalisée afin d'enlever les restes du tampon Edwards et d'isopropanol. Après séchage, l'ADN est repris dans 25 µL de Tris-EDTA, pH 8.

2. Analyse des mutants d'insertion par PCR

L'ADN précédemment obtenu est utilisé pour des réactions d'amplifications en utilisant toutes les combinaisons possibles entre à les couples d'amorces spécifiques du gène d'intérêt (gène muté) (G3= antisens et G5= sens) et de l'ADN-T (LB= left border et RB= right border) pour révéler la présence de l'insertion. Les produits de PCR sont ensuite analysés par électrophorèse sur gel d'agarose. Après avoir vérifié que les amorces G3 et G5 amplifient un fragment d'ADN sauvage, l'absence d'amplification d'un fragment de l'ADN du mutant en utilisant cette combinaison G3 et G5 (fragment devenu trop grand pour être amplifié à cause de la présence d'une insertion) et la visualisation de bandes avec au moins une des autres combinaisons (G3-LB, G3-RB, G5-LB, G5-RB) signifie qu'il y a bien une insertion dans ce gène et son homozygotie dans le mutant. Le mutant est ensuite prêt à être transformé par *Agrobacterium tumefaciens*.

VII. Expression stable de protéines chez *Arabidopsis thaliana*

1. Le système *Agrobacterium tumefaciens*

Agrobacterium tumefaciens est une α-protéobactérie de la famille des *Rhizobiaceae*. Cette bactérie est pathogène pour les plantes supérieures et induit une tumeur chez la plante parasitée. Elle

possède un génome de 5,67 mégabases constitué d'un chromosome circulaire, d'un chromosome linéaire et de deux plasmides (le plasmide Ti et le plasmide At). Le plasmide Ti, de 214 kb, permet à la bactérie de transférer un fragment d'ADN (ADN-T) portant des gènes qui induisent la tumeur, vers les cellules eucaryotes. Ce fragment d'ADN pourra alors s'intégrer de manière aléatoire dans le génome nucléaire de la cellule hôte. Les gènes portés par cet ADN-T pouvant facilement être remplacés par des gènes d'intérêt. *Agrobacterium tumefaciens* représente un outil idéal pour la production de plantes transgéniques. Dans ce but, des plasmides plus petits portant les deux bordures de l'ADN-T et des sites uniques de clonage sont communément utilisés.

2. Construction des vecteurs d'expression des protéines fusionnées à la GFP

Les deux plasmides pEL103 (créé par Eric Lam, Université du New Jersey, USA) et pt-gb sont utilisés au laboratoire pour l'expression stable d'une protéine d'intérêt sous le contrôle d'un promoteur constitutif fort (promoteur 35S). Le plasmide pEL103 est un vecteur binaire, capable de se répliquer dans *E. coli* et dans Agrobactérium, et qui possède un gène de résistance à la kanamycine. Le vecteur pt-gb est également un vecteur binaire qui possède un gène de résistance à l'herbicide Basta. Ces vecteurs comportent entre les bords droit et gauche de l'ADN-T une construction destinée à l'expression de protéines chez les plantes. Les ADNcs codant pour les protéines ceQORH, Tic32, NQR et la forme mutée (Mut2) de ceQORH ont été intégrés dans ces vecteurs d'expression. L'ADNc de chaque protéine a été d'abord cloné dans le plasmide pKS. Les fragments obtenus à partir de la digestion de ces plasmides ont été ensuite insérés dans le vecteur puc19-35Ω-sGFP(S65T) préalablement digéré par les mêmes enzymes. Ces plasmides ont été digérés par les enzymes de restriction *Eco*RI et *Hind*III, et les fragments isolés correspondant aux séquences codantes des protéines ceQORH, Tic32, NQR et Mut2 ont été introduits dans les vecteurs binaires pEL103 et pt-gb.

3. Techniques de transformation des plantes et sélection des transformants

a) Transformation d'*Agrobacterium tumefaciens*

La souche d'*Agrobacterium tumefaciens* utilisée est la souche C58, résistante à la rifampicine et à la carbénicilline, et ne possédant pas de plasmide Ti avec ADN-T. Des cellules compétentes (50 µL) sont incubées en présence d'environ 120 ng de plasmide recombinant dans une cuve d'électroporation, pré-refroidie dans la glace. Les agrobactéries sont transformées par

électroporation, induite par un choc électrique de 2,5 kV. Immédiatement après ce choc électrique, 1 mL de LB est ajouté au mélange. La suspension bactérienne est ensuite incubée pendant 1 h à 28°C sous agitation. Afin de sélectionner les bactéries ayant intégré le plasmide d'intérêt, la suspension est étalée sur boite de Pétri, contenant du milieu LB et les antibiotiques (rifampicine (50 µg/mL), carbénicilline (100 µg/mL), kanamycine (50 µg/mL), et les boites sont incubées à 28°C pendant 2 jours pour permettre le développement des colonies.

b) Transformation des plantes par trempage d'inflorescences

Les souches *A. tumefaciens* précédemment transformées avec le plasmide d'intérêt sont utilisées pour transformer des plantes d'Arabidopsis (mutants et WT). Pour cela, une préculture contenant plusieurs colonies d'agrobactéries transformées est réalisée dans 10 mL de milieu LB contenant 100 µg/mL de carbénicilline, 50 µg/mL de rifampicine, et 50 µg/mL de kanamycine. Les bactéries sont mises en culture pendant 8 h à 28°C sous agitation. Cette préculture est utilisée pour ensemencer 500 mL de milieu LB contenant 100 µg/mL de carbénicilline, 50µg/mL de rifampicine, et 50 µg/mL de kanamycine. La culture est incubée pendant la nuit à 28°C sous agitation. Le lendemain, un échantillon de cette culture est dosé au spectrophotomètre à une longueur d'onde de 600 nm. Les agrobactéries sont alors concentrées par centrifugation 20 min à 5000 *g* et remises en suspensions dans du saccharose 5% [m/v] et Silwett L-77 0,035% [v/v] afin d'obtenir une suspension ayant une absorbance voisine de 0,8 à 600 nm. Le Silwett L-77 (LEHLE Seeds) est un surfactant qui favorise l'infection des plantes par les agrobactéries. La transformation d'*A. thaliana* se fait par trempage d'inflorescences (Clough & Bent, 1998) avec la suspension d'agrobactéries préparée précédemment. Les fleurs sont trempées pendant 5 sec dans un bécher contenant le milieu d'inoculation. Les plantes sont ensuite placées 24 à 48 h en lumière faible et sous cloche afin d'éviter leur dessèchement. Elles sont ensuite replacées en condition de culture standard. Une deuxième transformation, identique à la première, est effectuée la semaine suivante afin d'augmenter le rendement de transformation. Les plantes poussent pendant 3 à 5 semaines jusqu'à ce que les graines arrivent à maturité. Les graines sont récoltées et conservées dans un grainier. Les plantes transformées en utilisant le vecteur pEL103 sont sélectionnées par germination des graines sur milieu MS contenant 50 µg/mL de kanamycine. Les plantes transformées en utilisant le vecteur pt-gb sont sélectionnées par germination des graines en terre suivie d'un traitement à l'herbicide Basta (deux à trois traitements Basta sont nécessaires pour éliminer les faux transformants). Une

fois les transformants identifiés il faudra s'assurer de l'intégrité du gène d'intérêt en vérifiant qu'il n'y a pas eu de recombinaison dans l'ADN-T d'intérêt après son transfert dans les agrobactéries. Cette étape de vérification repose sur des réactions d'immunodétection (western blot) en utilisant l'anticorps dirigé contre la GFP.

4. Observation par microscope confocale des plantes transformées

L'observation au microscope confocal a été réalisée à partir de feuilles de jeunes plantes d'*Arabidopsis thaliana.* Pour cette étude, la fluorescence de la GFP (pour les plantes exprimant de manière stable les protéines ceQORH, Tic32, NQR et Mut2 fusionnées à la GFP) a été observée avec le système Leica (TCS SP2 AOBS, BIG, CEA Grenoble). La GFP est excitée à 488 nm d'un laser Argon. Après excitation, la GFP émet une fluorescence comprise entre 500 nm et 540 nm. Pour visualiser des chloroplastes, l'autofluorescence de la chlorophylle a été utilisée. Dans ce cas, la chlorophylle est excitée à 633 nm et la fluorescence émise recueillie entre 640 et 725 nm.
La coloration des tissus avec la rhodamine B hexyl ester a été réalisée comme décrit précédemment [136]. Des solutions mères de rhodamine B (Invitrogen) ont été préparées à 10 mg / mL en dissolvant de la rhodamine dans du DMSO. Des aliquotes de ce stock ont été diluées dans du Tris-HCl 50 mM, pH 8,0, jusqu'à une concentration finale de 1 mg / ml de rhodamine. Cette aliquote a été utilisée pour infiltrer des feuilles d'Arabidopsis âgées de 2 à 3 semaines. Après incubation de quelques minutes à température ambiante avec de la rhodamine, les feuilles ont été observées par microscopie confocale.
De même, le marquage du noyau avec le DAPI (diamidino phénylindole) a été réalisé par infiltration des feuilles d'Arabidopsis à une concentration de 1/50000. Après incubation de 5 min à température ambiante avec le DAPI, les feuilles ont été observées par microscope confocale.

Chapitre I

Pontage chimique, purification et approche protéomique pour identifier les récepteurs d'import alternatif à la surface des chloroplastes

Pontage chimique, purification et approche protéomique pour identifier les récepteurs d'import alternatif à la surface des chloroplastes

L'analyse protéomique de l'enveloppe des chloroplastes d'*Arabidopsis thaliana* développée au laboratoire [6]suivie de l'étude réalisée par Stéphane Miras (ancien doctorant de l'équipe) et Daniel Salvi [7, 9], avait montré l'existence dans ce sous–compartiment, d'une protéine nommée ceQORH, pour chloroplast envelope quinone oxido-reductase homologue. Bien que dépourvue de séquence d'adressage chloroplastique *N*-terminale clivable, cette protéine est localisée dans la membrane interne de l'enveloppe des chloroplastes. Il a été également démontré que l'import de cette protéine ceQORH dans la membrane interne de l'enveloppe chloroplastique ne nécessite pas l'intervention du système de translocation TOC-TIC [9] précédemment décrit (voir introduction).
L'existence d'un nouveau système d'adressage et d'import, dit « alternatif », des protéines codées par le génome nucléaire à destination des chloroplastes n'impliquant pas de séquence d'adressage *N*-terminale clivable, de type « canonique », est également supportée par l'étude de la protéine Tic32 par des collègues de l'équipe de Jürgen Soll [8]. Comme ceQORH, Tic32 est localisée au niveau de la membrane interne de l'enveloppe des chloroplastes, dépourvue de séquence d'adressage chloroplastique *N*-terminale clivable et son import est indépendant du système TOC-TIC.
Ces premières observations ont encore été confortées, suite à un projet de collaboration entre notre équipe et la plateforme de protéomique EDyP, qui a permis l'identification d'une troisième protéine chloroplastique non canonique nommée NQR (voir introduction).
L'objectif de cette première partie de mon projet de thèse, visait à identifier des récepteurs, localisés à la surface des chloroplastes et requis pour l'import de ces trois protéines (ceQORH, Tic32 et NQR) utilisant potentiellement des voies d'import non canoniques. A titre de contrôle, nous avons également utilisé la protéine FSD1, une protéine dépourvue de peptide de transit canonique clivable qui utilise néanmoins la voie d'import classique (voir introduction). En effet, la première étape de notre approche reposait sur la production de protéines recombinantes chez *E. coli*. Il était donc préférable de sélectionner des protéines dépourvues de séquences de transit chloroplastiques

classiques, les protéines possédant une séquence de transit clivables étant très instables lorsque produites chez *E. coli*.

I. Production des protéines non canoniques à utiliser comme appâts dans les études d'interaction

La première étape de ce travail a consisté à produire les protéines chloroplastiques ceQORH, Tic32, NQR et FSD1. Afin de pouvoir les utiliser ultérieurement lors des études d'interaction avec l'enveloppe de chloroplaste, ces protéines recombinantes devaient être produites en bactéries. Enfin, la stratégie mise en place pour la purification ultérieure des complexes impliquait d'utiliser une étiquette commune pour les quatre protéines. Les constructions permettant de produire les protéines ceQORH, Tic32 et NQR fusionnées à la protéine GFP (Green Fluorescent Protein) étaient déjà disponibles dans le laboratoire.

1. Validation de la séquence primaire de la protéine FSD1 prédite par l'analyse du génome d'*Arabidopsis thaliana*

L'analyse du génome d'*Arabidopsis thaliana* montre l'existence du gène (AT4G25100.1) codant pour la protéine chloroplastique FSD1. L'ADNc correspondant a été amplifié par PCR à partir d'une banque d'ADNc d'*Arabidopsis thaliana* (figure 1). Le fragment ainsi amplifié a été cloné dans un plasmide pBlueScript(KS) et a été séquencé. La séquence du fragment obtenu s'est avérée être parfaitement en accord avec la protéine théorique déduite de l'analyse du génome d'*Arabidopsis thaliana* (figure 1).

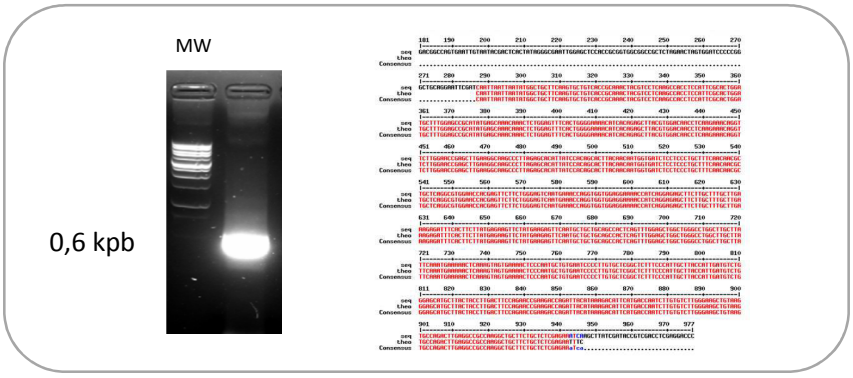


Figure 1. Alignement de la séquence nucléotidique codant la protéine FSD1 d'*Arabidopsis thaliana*. L'ADNc codant pour la protéine FSD1 a été amplifié à partir d'une banque d'ADNc d'A*rabidopsis thaliana* et séquencé. Les nucléotides identiques entre la séquence nucléotidique issue du séquençage et la séquence

du gène AT4G25100.1 d'*Arabidopsis thaliana* sont indiqués en rouge. Les nucléotides non identiques sont indiqués en bleu.

2. Construction de vecteurs d'expression des protéines recombinantes

Selon le protocole décrit dans la partie « Matériel et méthodes », la construction des fusions de séquences codantes pour Tic32::GFP, ceQORH::GFP, NQR::GFP et FSD1::GFP a été réalisée par clonage de bouts cohésifs, dans le plasmide d'expression pET-20b qui avait été préalablement digéré par *Nde*I et *Bam*HI dans le cas de protéines fusionnées à la séquence poly-histidine codée par ce vecteur.

Afin de valider chaque construction, une digestion de ce plasmide par les enzymes de restriction *Nde*I et *Bam*HI a été réalisée. Les produits issus de cette digestion ont été séparés par électrophorèse sur gel d'agarose (figure 2). Les bandes observées sur gel se sont avérées être parfaitement en accord avec la construction FSD1::GFP attendue dans pET-20b.

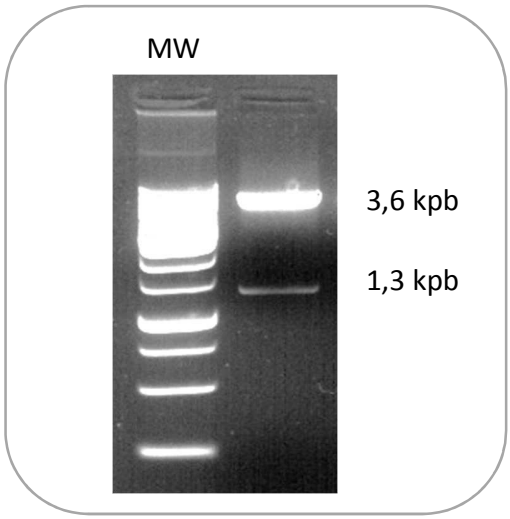


Figure 2. Analyse des produits issus de la digestion du vecteur pET-20b contenant l'insert FSD1::GFP. La digestion a été réalisée en utilisant le couple d'enzymes de restriction *Nde*I/*BamH*I afin de vérifier que l'insertion du fragment codant FSD1::GFP avait bien eu lieu. Deux bandes ont été observées, la première à une taille de 1,3 kpb correspondant à la cassette codant FSD1::GFP et l'autre, plus haute (3,6 kpb), qui correspond au vecteur pET-20b.

3. Surexpression des protéines recombinantes

Afin de pouvoir les utiliser dans les études ultérieures, il était nécessaire de disposer d'une quantité suffisante de chacune des protéines recombinantes. Pour cela, nous avons produit ces protéines dans un système procaryote (*E. coli*) selon le protocole décrit dans la partie «Matériels et

méthodes». La figure 3 montre que le système utilisé permet d'exprimer efficacement la protéine recombinante FSD1::GFP chez *E. coli*.

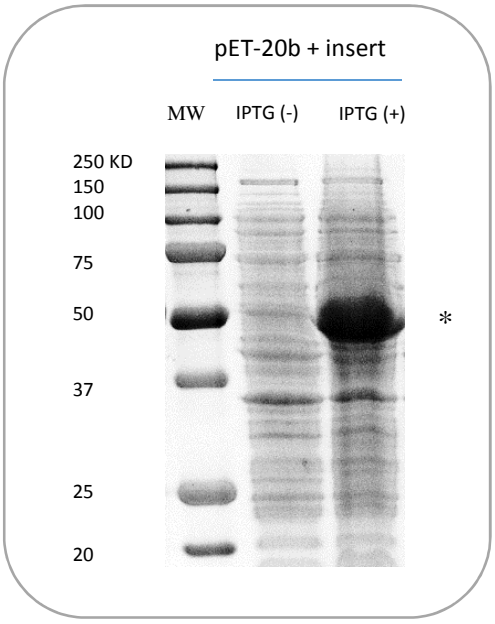


Figure 3. Expression de la protéine FSD1::GFP (*) chez *E. coli visualisée après séparation sur SDS-PAGE et coloration des protéines au bleu de Coomassie.* La transformation, les conditions de culture et les conditions d'induction de la synthèse de la protéine recombinante sont décrites dans la partie «Matériel et méthodes». Le contrôle négatif correspond à l'analyse des protéines présentes dans les bactéries avant induction par l'IPTG (-).

4. Vérification de la solubilité des protéines recombinantes produites

En fin d'induction, les bactéries sont lysées et les protéines bactériennes sont centrifugées afin de séparer les protéines solubles des protéines membranaires (ou agrégées). Les protéines sont ensuite analysées, après séparation par électrophorèse sur gel en conditions dénaturantes (SDS-PAGE), par marquage immunologique (western blot ; figure 4). Comme pour les autres protéines recombinantes déjà produites par les collègues (Lucas Moyet et Daniel Salvi) de l'équipe (Tic32, ceQORH et NQR), la protéine FSD1 ainsi synthétisée se répartit également en deux fractions distinctes : une fraction majeure, insoluble, correspondant probablement à la protéine FSD1 agrégée et une fraction mineure soluble. Cette fraction protéique soluble, dont l'identité a été validée par western blot (figure 4), peut être ensuite purifiée par une étape de purification en conditions non dénaturantes.

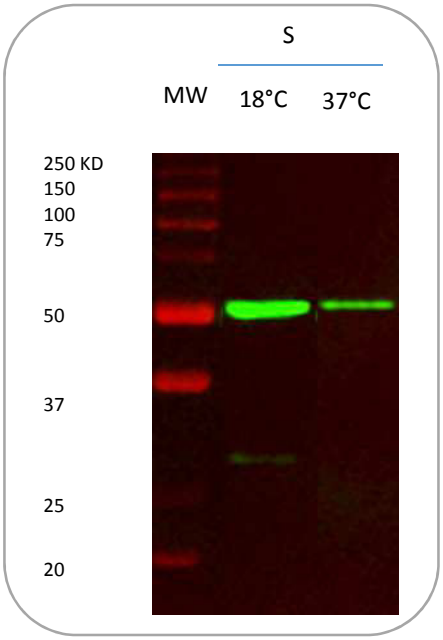


Figure 4. Test de solubilité de la protéine recombinante FSD1::GFP par marquage immunologique (western blot). Après induction par l'IPTG suivie d'une mise en culture pendant la nuit à 18°C ou 3 h à 37°C, les bactéries sont concentrées par centrifugation, mises en suspension dans 1 ml de Tris/HCl 20 mM pH 7,9, imidazole 5 mM, NaCl 0,5 M puis lysées par sonication. Une centrifugation (40 min, 13000 rpm) permet de récupérer les protéines solubles (**S**). Les protéines insolubles sont mises en suspension dans 1 ml du même tampon. Les deux fractions sont analysées par électrophorèse en conditions dénaturantes. Seule la fraction soluble (**S**) est analysée par marquage immunologique.

5. Purification des protéines recombinantes

Le clonage des cassettes codant les protéines Tic32::GFP, ceQORH::GFP, NQR::GFP, et FSD1::GFP dans le vecteur d'expression pET-20b a permis de produire des protéines recombinantes fusionnées à une séquence poly-histidine à l'extrémité *C*-terminale. La purification de la fraction soluble de la protéine recombinante a donc été effectuée à l'aide d'une colonne d'affinité comme décrit dans la partie «Matériel et méthodes».

Cette étape a permis d'obtenir plusieurs µg de protéines recombinantes pures (figure 5). Une aliquote provenant de cette purification a ensuite été utilisée pour les études d'interaction avec la surface des chloroplastes.

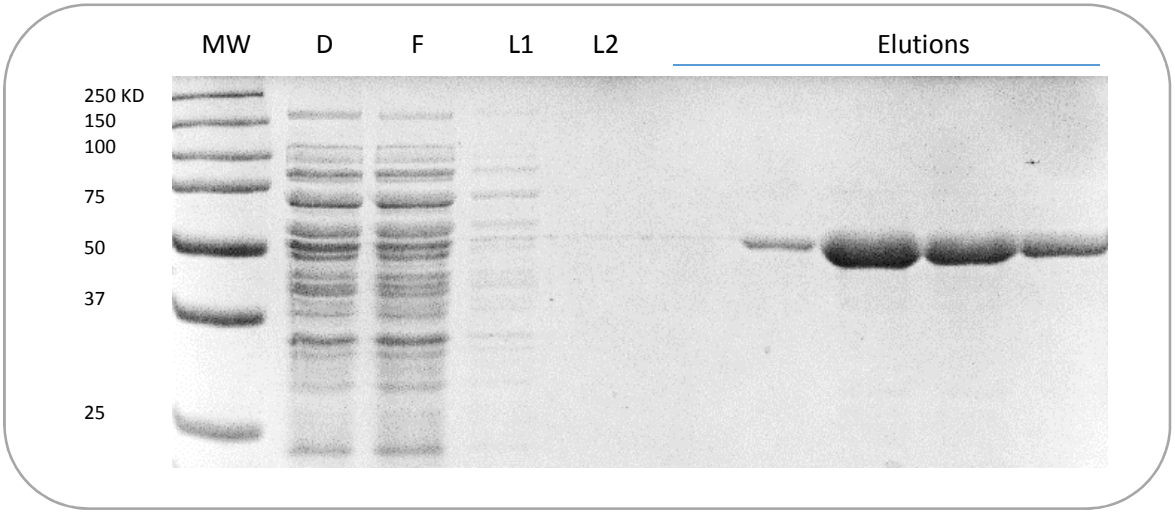


Figure 5. Purification de la protéine recombinante FSD1::GFP sur matrice de nickel. La purification de la protéine est réalisée comme décrit dans la partie «Matériel et méthodes». **D** = protéines bactériennes solubles déposées sur la colonne; **F** = protéines non retenues sur la matrice; **L1** et **L2** = lavages de la matrice avec deux tampons de lavage contenant 5 mM et 60 mM d'imidazole respectivement. L'élution de la protéine a été obtenue par utilisation de tampon contenant 250 mM d'immidazole. Chaque fraction est alors analysée par électrophorèse en condition dénaturante (SDS-PAGE, coloration au bleu de Coomassie).

II. Pontage chimique : vers l'identification de récepteurs d'import alternatif

A ce stade de travail, toutes les protéines non canoniques du chloroplaste fusionnées à la GFP étaient disponibles de manière satisfaisante (pureté et quantité). Nous avons aussi accumulé de nombreuses préparations d'enveloppe de chloroplastes d'*Arabidopsis thaliana* (plusieurs semaines de travail) et préparé des chloroplastes intacts d'*Arabidopsis thaliana* (voir partie « Matériel et méthodes »). Tout d'abord, nous avons réalisé des tests préliminaires afin d'optimiser les conditions d'interaction des protéines à la surface des organites, en utilisant des chloroplastes intacts fraichement purifiés. Des observations au microscope confocal ont été réalisées qui nous ont permis de valider que la présence de concentrations suffisantes d'ATP permettait un import rapide des protéines testées à l'intérieur des chloroplastes. A l'issue de ces expériences, nous avons choisi de partir de l'enveloppe des chloroplastes pour la suite du travail.

Selon le protocole décrit dans la partie «Matériel et méthodes», la deuxième étape de cette partie de travail a consisté à piéger les récepteurs membranaires présents à la surface des chloroplastes

avec leurs ligands, les protéines recombinantes (Tic32::GFP, ceQORH::GFP, NQR::GFP, et FSD1::GFP) en utilisant un agent chimique de pontage (cross-linker).

1. Optimisation de la méthode de pontage chimique

L'objectif visé à travers cette méthode est de purifier les récepteurs d'imports alternatifs à la surface des chloroplastes. Nous avons testé diverses conditions afin d'optimiser les conditions idéales de la réaction. Ces conditions incluent des changements de types de cross-linker utilisés (AMAS, BS3 et DMS), diverses concentrations de ces cross-linkers (en absence du cross-linker (0 mM), 1 mM…) (figure 8), des quantités de protéines variables et diverses préparations d'enveloppe (Arabidopsis ou épinard).

a) Réaction de pontage en présence d'AMAS

Les premières expériences de pontage ont été réalisées avec l'enveloppe des chloroplastes d'*Arabidopsis thaliana* et en présence d'AMAS 1mM comme agent de pontage chimique. Cette molécule génère des liaisons (-SH) réversibles. A l'issue de ces expériences, quelques microlitres des élutions de la colonne d'affinité ont été analysées par western-blot (figure 6). Il apparait que seules 3 des 4 protéines recombinantes ont été retrouvées dans les éluâts issus de ces colonnes d'affinité (Tic32::GFP, NQR::GFP, et FSD1::GFP). Comme les vésicules d'enveloppe avaient été préalablement lavées avant l'étape de pontage (voir partie « Matériel et méthodes »), il est probable que la protéine ceQORH::GFP présentait moins d'affinité pour son récepteur et avait été perdue lors de cette étape de lavage.

Il faut noter ici que les différentes expériences ont été réalisées à partir de plusieurs préparations d'enveloppe d'*Arabidopsis thaliana,* qui ont été mélangées puis homogénéisées et ensuite séparées en 4 tubes, pour réaliser les étapes de purifications sur les 4 protéines recombinantes. Cela a été fait afin de limiter l'impact de la variabilité éventuelle liée aux différentes préparations d'enveloppe. Ainsi, pour chacune des quatre expériences, seules les 4 protéines recombinantes étaient différentes.

Les résultats présentés dans la figure 6, nous permettent d'observer également des bandes de haut poids moléculaire dans les trois premières fractions, qui peuvent être des complexes protéiques ou simplement des multimères de Tic32::GFP, NQR::GFP, et FSD1::GFP, qui ont été générés par le pontage.

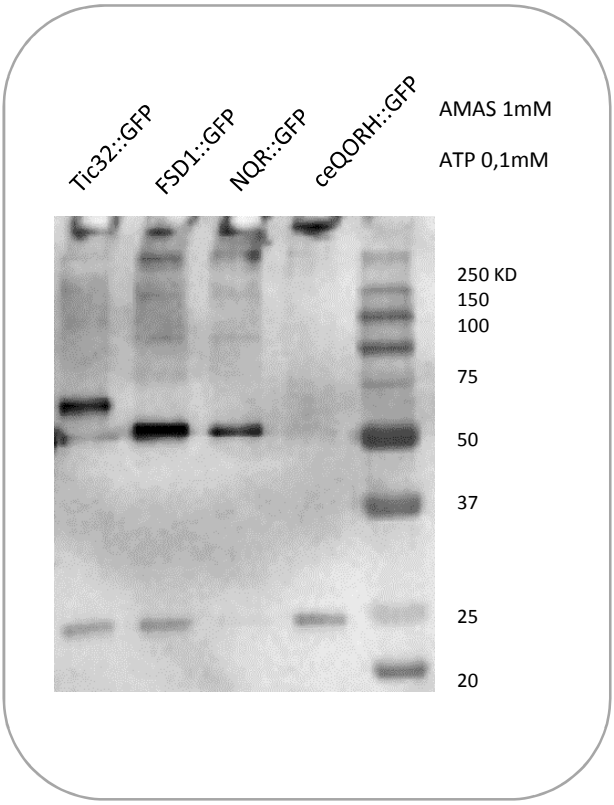


Figure 6. Analyse des élutions de la colonne d'affinité afin de visualiser les complexes protéiques issus de la réaction de pontage chimique, en présence de l'AMAS et de l'enveloppe des chloroplastes d'*Arabidopsis thaliana,* par SDS-PAGE et marquage immunologique (western blot). Après solubilisation par le dodécylmaltoside (DDM) 1% [m/v] pendant 45 min, puis incubation pendant 35 min en présence des anticorps dirigés contre l'étiquette GFP et liés à des billes magnétiques, le mélange est immédiatement déposé sur la colonne d'affinité. L'élution est réalisée en présence de DTT. Quelques microlitres d'éluât de la colonne d'affinité ont été analysés.

Afin d'optimiser les étapes d'interaction et de valider l'efficacité du pontage en présence de l'AMAS, des expériences similaires mais reposant sur de plus grandes quantités de protéines d'enveloppe d'épinard sont réalisées.

Comme l'AMAS génère des liaisons réversibles, et afin de maintenir la structure des complexes cross-linkés, nous avons essayé de visualiser, sur des SDS-PAGE en conditions non réductrices, la présence de complexes protéines/récepteurs (figure 7).

D'après les résultats présentés dans la figure 7, nous avons constaté qu'il n'existe pas de différence nette en ce qui concerne l'enrichissement en composés de haut poids moléculaire en présence ou en absence de l'AMAS. Cela est également valable en conditions réductrices (présence de DTT)

ou non réductrices (absence de DTT). En conclusion, le cross-linker AMAS ne semblait pas être efficace dans le contexte.

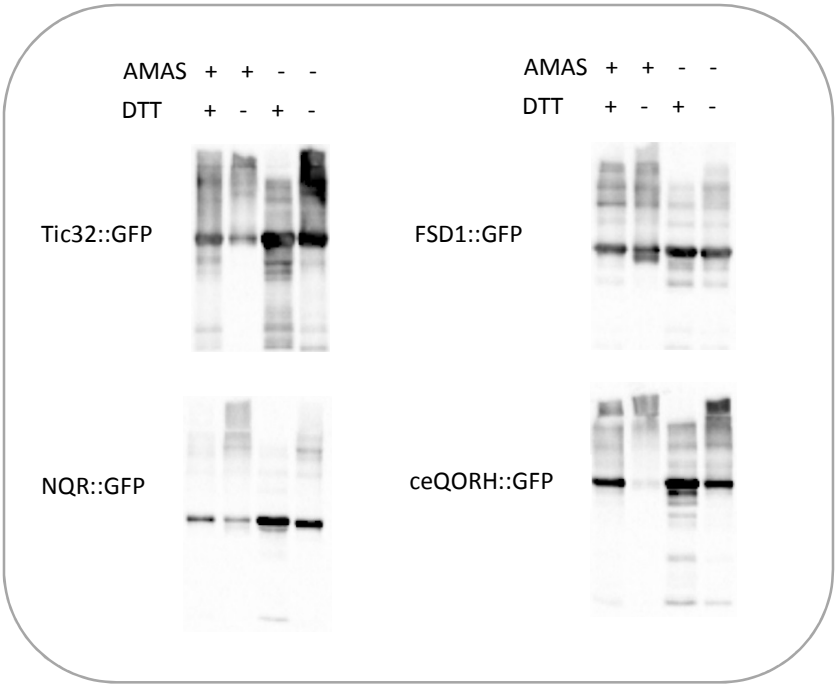


Figure 7. Efficacité de la réaction de pontage chimique en présence de l'AMAS en utilisant l'enveloppe des chloroplastes de l'épinard. Les échantillons contenant des complexes protéiques pontés (AMAS(+)) ou non (AMAS(-)) sont solubilisés dans le SDS, puis séparés par SDS-PAGE en conditions réductrices (DTT(+)) ou en conditions non réductrices (DTT(-)). Un marquage immunologique (western blot) est réalisé.

b) Réaction de pontage en présence d'une nouvelle molécule : BS^3

Comme montré par les résultats d'expériences précédentes, l'AMAS ne semblait pas être adapté au contexte de nos expériences. Nous avons donc décidé de tester d'autres molécules chimiques de pontage. Toutes les expériences de pontage ont été reproduites avec deux nouveaux cross-linkers, le BS^3 et le DMS. A l'issue de ces expériences, quelques microlitres des échantillons ont été analysés par western-blot (figure 8). Il apparait que le pontage a été efficace en présence de BS^3, au moins avec trois (Tic32::GFP, ceQORH::GFP et FSD1::GFP) des quatre protéines recombinantes utilisées. En effet, contrairement aux fractions issues des expériences réalisées en présence de DMS, on observe une accumulation en matériel protéique de haut poids moléculaire pour trois des quatre expériences réalisées en présence de BS^3.

Les résultats présentés dans la figure 8, nous permettent d'observer une accumulation de complexes de haut poids moléculaires en présence de BS^3 qui pourraient être compatibles avec des complexes d'import des protéines. Nous constatons également que le profil protéique obtenu en présence de DMS est presque le même qu'en absence de toute molécule de pontage (-), ce qui suggère que le DMS est inefficace pour ce type de pontage.

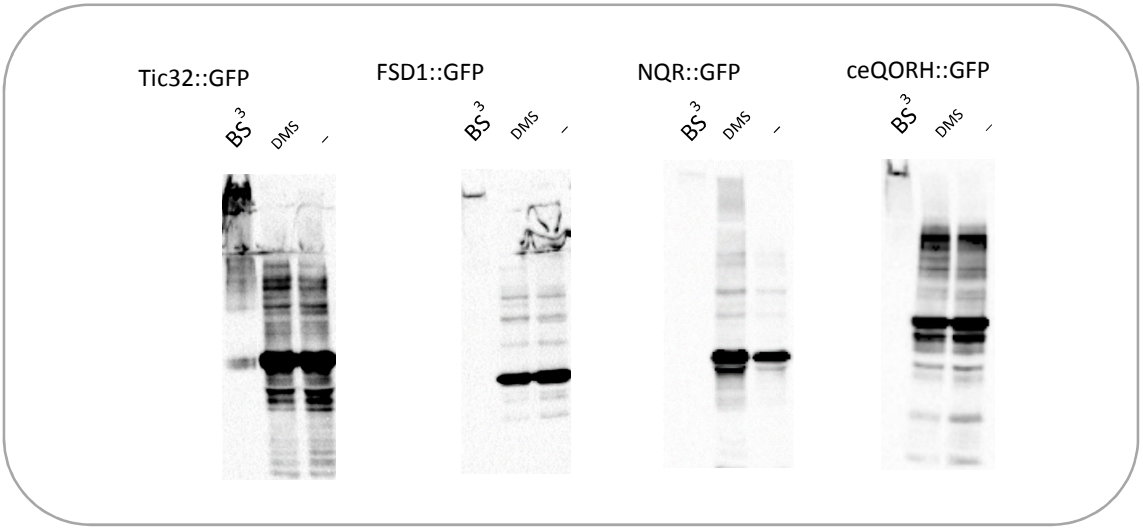


Figure 8. Visualisation après SDS-PAGE et immunomarquage du pontage chimique en présence de BS^3 et de DMS en utilisant l'enveloppe des chloroplastes de l'épinard. Le BS3 génère des liaisons irréversibles, toutes les séparations sont réalisées en conditions réductrices. Les profils protéiques en absence de cross-linker (-) et en présence de DMS sont les mêmes. Les résultats obtenus en présence de BS3 semblent être meilleurs, une accumulation en matériel protéique de haut poids moléculaire étant observée avec les protéines Tic32::GFP, ceQORH::GFP et FSD1::GFP.

2. Purification par affinité des complexes d'import

Après ces différentes étapes de mise au point, la même expérience a été reproduite sur l'enveloppe de chloroplaste d'*Arabidopsis thaliana*. La taille des complexes (protéines / récepteurs réticulés) observée étant compatible avec celle de récepteurs d'import potentiels couplés à nos protéines cibles, nous avons choisi de poursuivre l'approche. Ainsi, après la réaction de pontage, les complexes protéiques ont été solubilisés en présence d'un détergent (DDM 1% [m / v]) et ensuite purifiés en utilisant des anticorps anti-GFP fusionnés à des billes magnétiques (MicroBeads). Les fractions éluées à partir de la matrice d'affinité ont été analysées par western blot en utilisant un anticorps anti-GFP (figure 9). La présence de protéines contenant de la GFP a été validée dans au

moins 3 des 4 fractions éluées (protéine Tic32::GFP, ceQORH::GFP et NQR::GFP). Il est important de noter ici que tout le processus s'est avéré très efficace en utilisant NQR::GFP (puisque les signaux liés à la présence de GFP étaient enrichis dans la fraction d'élution par rapport à la fraction de dépôt) alors qu'il semble que l'expérience réalisée avec FSD1::GFP n'a pas donné les résultats attendus (figure 9).

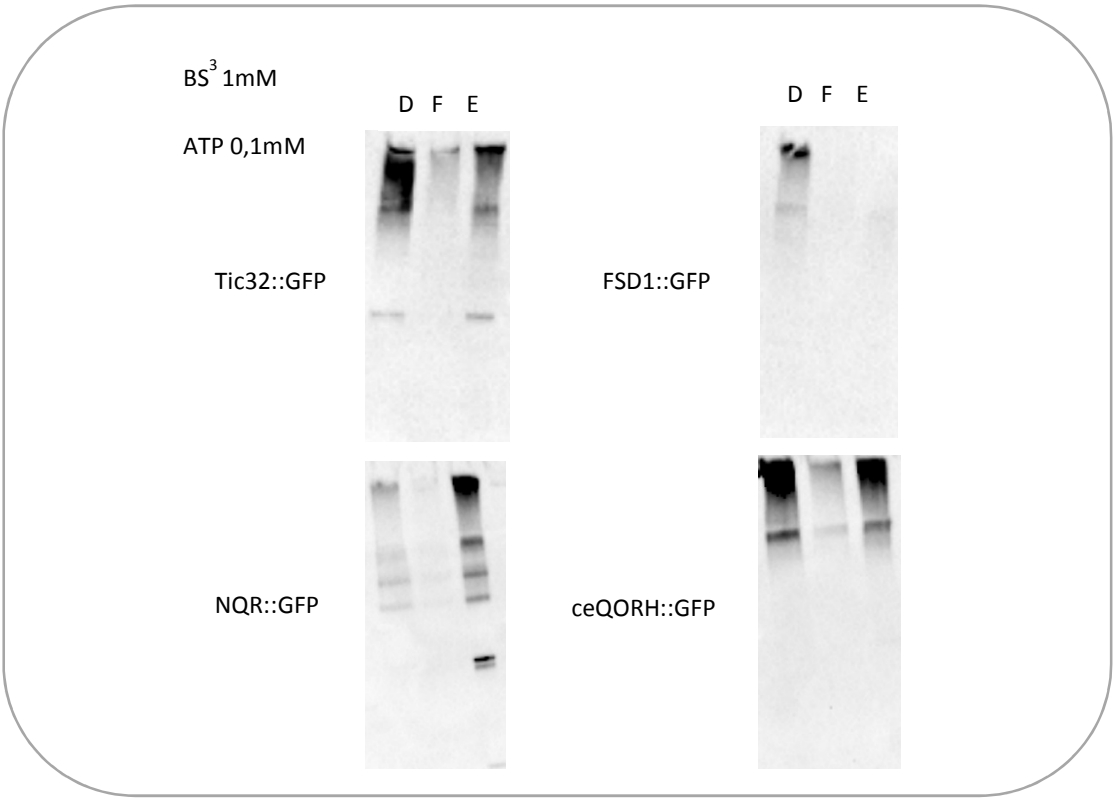


Figure 9. Analyse des élutions (E) de la colonne d'affinité afin de visualiser les complexes protéiques issus de la réaction de pontage chimique, en présence de BS^3 et de l'enveloppe des chloroplastes d'*Arabidopsis thaliana,* par marquage immunologique (western blot). La purification des complexes protéiques est réalisée comme décrit dans la partie «Matériels et méthodes». **D** = extrait protéique déposé sur la colonne; **F** = protéines non retenues sur la matrice. **E** = élution faite en conditions réductrices. La présence de matériel protéique de haut poids moléculaire dans la fraction d'élution est observée avec les protéines Tic32::GFP, ceQORH::GFP et NQR::GFP.

3. Identification des peptides spécifiques des partenaires protéiques (récepteurs d'import potentiels) par spectrométrie de masse

Une approche reposant sur de la spectrométrie de masse quantitative a été ensuite adoptée afin d'identifier les protéines qui avaient été spécifiquement enrichies dans les quatre fractions d'élution

issues de la chromatographie d'affinité. L'approche quantitative a été nécessaire pour différencier les peptides spécifiquement enrichis dans un mélange de milliers d'autres peptides résultant de la contamination par des protéines co-purifiées lors de la purification par affinité des complexes protéiques. En théorie, seuls les peptides issus des protéines partenaires devaient être plus abondants dans les quatre échantillons analysés par rapport aux échantillons contrôles (même réaction d'interaction enveloppe-protéine recombinante sans l'agent de pontage). Cependant, comme la nature moléculaire des récepteurs interagissant avec les protéines ceQORH, Tic32 et NQR était inconnue, et comme l'expérience réalisée avec la protéine FSD1 (qui devrait interagir avec les sous-unités TOC) n'avait pas fonctionné, cette étape a nécessité de combiner l'expertise MS / MS, des statistiques complexes et beaucoup de bioanalyse et d'examen de la littérature. Notre équipe collabore avec la plateforme de protéomique « EDyP» depuis de nombreuses années. La stratégie adoptée par cette plateforme est le "shotgun" d'un échantillon déposé sur un gel SDS-PAGE. Cette stratégie vise à étudier l'ensemble du matériel protéique en une seule analyse. Pour la validité statistique, trois réplicats biologiques (trois purifications indépendantes, pour les quatre protéines cibles) et deux réplicats techniques (MS / MS) ont été réalisés. L'identification des protéines et des peptides a été faite par les collègues de l'équipe EDyP (Sabine Brugière et Marianne Tardif) en utilisant une stratégie de "Mass Matching". Cette approche vise à comparer les données de fragmentation de peptides théoriques (via une digestion *in silico* d'une base de données de protéines d'Arabidopsis) à des spectres de fragmentation expérimentaux. Chaque protéine a été validée par la présence de deux peptides uniques au moins. Tous les contaminants tels que les kératines ont été retirés de l'analyse. Pour prédire *in silico* la localisation subcellulaire et la fonction de chaque protéine, plusieurs bases de données d'Arabidopsis ont été utilisées (http://chlorokb.fr, http://at-chloro.prabi.fr/at_chloro, https://www.arabidopsis.org, http://www.uniprot.org). La majorité des protéines détectées dans les élutions étaient de véritables protéines de l'enveloppe du chloroplaste, mais certaines protéines issues du stroma et des thylakoïdes ont également été détectées. Une quantification relative "sans marquage" des protéines a été réalisée, en utilisant le logiciel MaxQuant. Deux stratégies ont été adoptées : la stratégie de comptage spectral qui vise à compter le nombre de spectres protéiques, et la stratégie XIC (Extracted Ion Chromatograms), permettant l'intégration du courant ionique MS1. Cette étape permet d'évaluer l'abondance relative des peptides en intégrant le signal MS1 (zone de pic issue des ions MS1). Ainsi, il est possible de comparer l'abondance relative de la même protéine et du même peptide présents dans nos quatre

échantillons protéiques indépendants. Après élimination des contaminants et analyse des données, plus de 600 protéines et 4000 peptides ont été identifiés (figure 10).

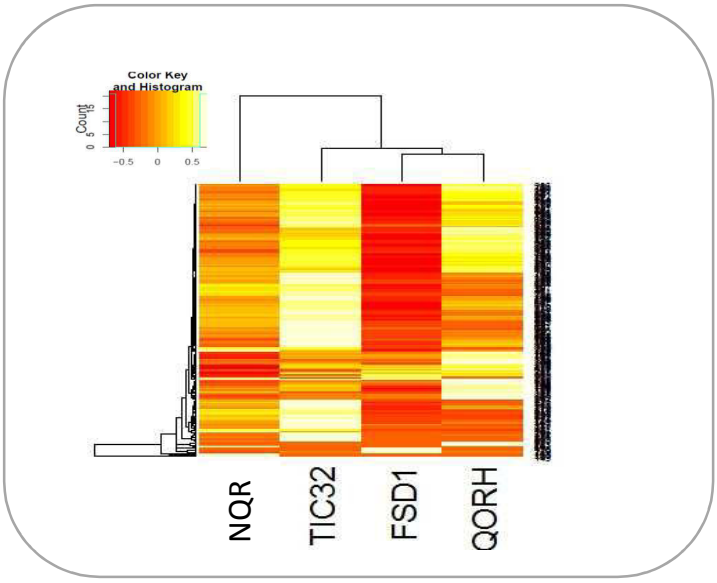


Figure 10. Heatmap des protéines détectées par spectrométrie de masse dans les quatre fractions issues de la purification d'affinité. Les nuances allant du rouge vers le jaune désignent l'abondance de chaque protéine détectée dans chaque fraction dans l'ordre croisant.

Cette complexité des échantillons est due au fait que de nombreuses autres protéines ont été co-purifiées au cours de cette étape de chromatographie d'affinité, contaminant ainsi les fractions éluées. Pour cela, afin d'identifier les partenaires de nos protéines (candidats récepteurs), nous nous sommes concentrés sur des peptides et des protéines spécifiquement enrichis dans l'une des quatre fractions éluées à partir de la matrice d'affinité (échantillons ceQORH, NQR, Tic32 et FSD1).

1. Identification de récepteurs candidats

Après traitement de données protéomiques générées par la spectrométrie de masse, plusieurs partenaires potentiels (candidats récepteurs) ont été identifiés. Outre les récepteurs potentiels, ces études protéomiques réalisées ont également généré une masse de données originales qui permettaient d'avoir une vision quantitative de la composition protéique de l'enveloppe des chloroplastes (voir chapitre II).

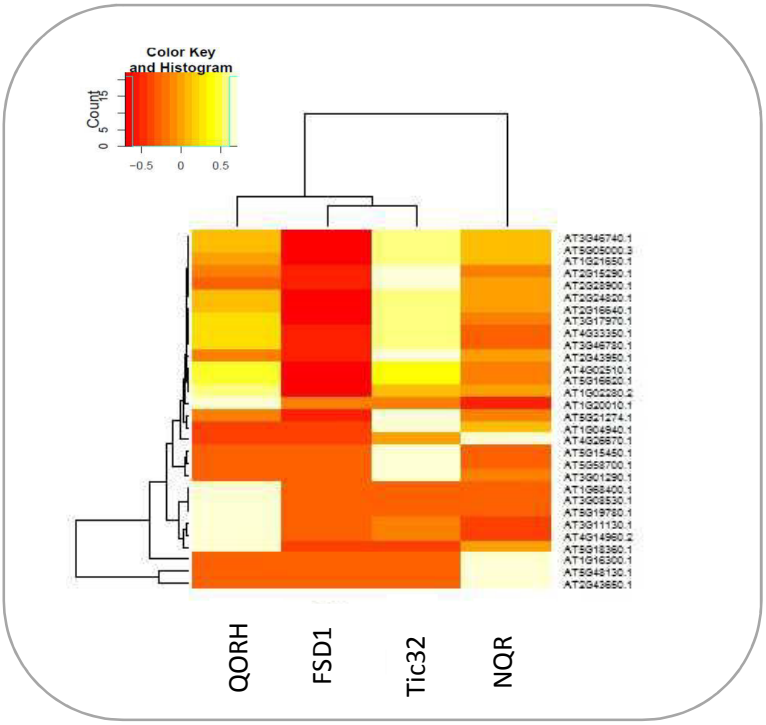


Figure 11. Heatmap des protéines spécifiquement enrichies dans les fractions issues de la purification d'affinité.

La figure 11 montre que certaines protéines n'ont été détectées que dans un seul échantillon, alors que d'autres protéines sont plus abondantes dans un échantillon par rapport aux trois autres. Après avoir criblé la littérature et utilisé divers sites de recherche et bases de données pour évaluer la nature ou la localisation connue des protéines, une liste de partenaires potentiels (candidats récepteurs) a été sélectionnée (Tableau 1).

Tableau 1 : **liste de partenaires potentiels issus de la réaction de pontage**

Protéine	AGI	Interaction
HP20	AT4G26670	NQR
NPH3	AT5G48130	NQR
EMB2777	AT2G43650	NQR
LRRIII	AT1G68400	ceQORH
ClpD	AT5G51070	ceQORH
ClpB3	AT5G15450	ceQORH
ClpR3	AT1G09130	ceQORH
HIR2	AT3G01290	Tic32

Comme le montre le tableau 1, il apparait que HP20 (AT4G26670.1), qui est une protéine membranaire de la membrane externe de l'enveloppe du chloroplaste [6, 101], était enrichie dans la fraction obtenue en utilisant NQR comme appât. Comme cité ci-dessus (voir introduction), HP20 appartient à la famille des transporteurs de pré-protéine et d'acides aminés (PRAT) [97, 98]. Il a été récemment proposé que cette protéine interagissait plutôt avec ceQORH [100] et serait nécessaire pour l'adressage chloroplastique de ceQORH *in vitro* [100]. Les résultats obtenus par notre approche n'ont pas permis de confirmer ce rôle (HP20 n'était pas enrichi dans l'élution de la protéine ceQORH). Deux autres protéines ont été spécifiquement enrichies dans la fraction d'élution NQR : 1) la protéine désignée NPH3 «non-phototropic hypocotyl 3» (AT5G48130.1) [137] qui est prédite pour être localisée dans le chloroplaste (possédant une séquence d'adressage chloroplastique prédite). Il s'agit d'une protéine de la famille des partenaires de photorécepteurs (photoreceptor-interacting protein), impliquée dans la voie de signalisation suite à une réponse à la lumière bleue [138] et qui contient plusieurs domaines d'interaction protéine-protéine [138], 2) la protéine EMB2777 (AT2G43650.1) qui appartient à la famille des ribonucléoprotéines Sas10 / U3 (Utp) [139], impliquée dans la conversion d'un transcrit d'ARNr primaire en une ou plusieurs molécules d'ARNr matures. Ces deux protéines n'avaient jamais été identifiées par spectrométrie de masse dans des fractions d'enveloppe du chloroplaste purifiées (elles ne sont pas présentes dans la base de données AT_CHLORO). En d'autres termes, ces protéines ont été spécifiquement enrichies dans l'échantillon NQR. Il est important de noter ici que la détection de nouvelles protéines mineures dans l'enveloppe chloroplastique pourrait également résulter de l'amélioration de la sensibilité des équipements de spectrométrie de masse au cours des dernières années. Cependant, comme ces protéines n'avaient pas été détectées dans les fractions d'élution des trois autres protéines, nous les avons considérées dans la liste des partenaires potentiels de la protéine NQR. Nous avons donc choisi d'explorer le rôle de ces trois protéines sélectionnées dans le contrôle de l'adressage chloroplastique de la protéine non canonique NQR.

Une autre pré-protéine chloroplastique a été détectée uniquement dans la fraction éluée en utilisant ceQORH comme appât. De même, cette protéine n'a jamais été détectée dans les fractions d'enveloppe par spectrométrie de masse et elle est absente de notre base de données de référence AT_CHLORO. Cette protéine notée LRRIII (AT1G68400.1) possède un peptide de transit prédit et appartient à la famille des kinases transmembranaires de type «leucine-rich repeat receptor kinase» [140-142]. Cette famille de protéines est caractérisée par son domaine récepteur kinase et

son domaine transmembranaire [140] et semble fournir un cadre structurel afin de favoriser les interactions protéine-protéine [141, 143]. Pour des raisons de spécificité et d'enrichissement et en raison de ses propriétés structurelles, nous avons également choisi d'explorer le rôle de cette protéine dans l'import de la protéine non canonique ceQORH dans les chloroplastes.

Cinq autres protéines, ClpD (AT5G51070), ClpB3 (AT5G15450), ClpP6 (AT1G11750), ClpR2 (AT1G12410) et ClpR3 (AT1G09130), étaient spécifiquement enrichies dans la fraction d'élution ceQORH. Ces protéines appartiennent au complexe Clp [144-148] qui est connu pour jouer un rôle important (protéase) dans l'homéostasie des protéines des chloroplastes [144-148] (figure 12). Ce complexe est localisé à l'interface entre l'enveloppe chloroplastique et le stroma. Cependant, un rôle direct de ce complexe dans l'adressage des protéines vers les chloroplastes (récepteur ou canal d'importation) n'a jamais été suggéré. En revanche, son association au complexe TIC/TOC a été proposé (voir discussion ci-dessous).

Comme le montre le tableau 1, la protéine AT3G01290.1 est spécifiquement enrichie dans l'échantillon Tic32. Il s'agit d'une préprotéine, prédite pour être localisée dans le chloroplaste, qui appartient à une famille de protéines membranaires contenant un domaine SPFH/bande7/PHB [86]. Ces protéines sont décrites dans la littérature comme contenant une courte hélice transmembranaire dans leur région *N*-terminale [149].

Outre les partenaires potentiels sélectionnés, les données générées par cette approche protéomique ont montré l'enrichissement de trois protéines : la clathrine, la tubuline et la calmoduline dans la fraction d'élution correspondant à la protéine ceQORH. Ce résultat s'est avéré être en accord avec des expériences antérieures basées sur la technique de double hybride réalisées par notre équipe (non publiés) et qui ont démontré que la protéine ceQORH interagit avec la clathrine et la tubuline en système levure. Il a également été démontré par notre équipe, en se basant sur des approches biochimiques, que la protéine ceQORH interagit avec la calmoduline [134]. Ces résultats nous ont donc encouragés également à déterminer si ces trois protéines participent au contrôle de l'adressage de la protéine ceQORH dans le chloroplaste (voir chapitre III).

III. Validation de la nature des récepteurs potentiels

Comme mentionné ci-dessus, nous avons identifié des candidats possédant des caractéristiques (localisation, structure, domaines…) compatibles avec une fonction de récepteurs d'import des protéines non canoniques ou qui pourraient participer à des mécanismes régulateurs contrôlant la

biogenèse des chloroplastes. Cependant, le rôle hypothétique de ces candidats dans ces processus devait être validé en utilisant des approches *in planta*. Nous avons donc initié les approches requises dans le contexte.

1. Isolement de mutants d'insertions d'*Arabidopsis thaliana* déficients dans l'expression des candidats récepteurs

De nombreux mutants d'insertion (insertions aléatoires d'un ADN-T modifié d'*Agrobacterium tumefaciens* dans le génome d'*Arabidopsis thaliana*) ont été générés et sont disponibles dans différentes banques de graines, dont le « European Arabidopsis Stock Centre » (NASC ; http://arabidopsis.info/BasicForm). Afin de valider le rôle des candidats sélectionnés dans le mécanisme d'import alternatif des protéines dans les chloroplastes, des mutants d'Arabidopsis affectés dans l'expression de ces candidats récepteurs ont été présélectionnés à partir du site du NASC (Tableau 2).

Tableau 2 : **liste de mutants d'insertion déficients dans l'expression des candidats récepteurs**

Protéine	AGI	Interaction	Mutant	Resistance	Analyse PCR
HP20	AT4G26670	NQR	N656245	Kanamycine	Homozygote
NPH3	AT5G48130	NQR	N862390	Basta	Homozygote
EMB2777	AT2G43650	NQR	N536872	Kanamycine	Hétérozygote
LRRIII	AT1G68400	ceQORH	N717997	Kanamycine	Homozygote
ClpD	AT5G51070	ceQORH	N686643	Kanamycine	Homozygote
ClpB3	AT5G15450	ceQORH	N559520	Kanamycine	Hétérozygote
ClpR3	AT1G09130	ceQORH	N669071	Kanamycine	Hétérozygote
HIR2	AT3G01290	Tic32	N657289	Kanamycine	Hétérozygote

Le criblage des plantes homozygotes pour les différentes insertions a été réalisé au laboratoire par la confirmation de la présence de l'insertion de l'ADN-T dans le génome des plantes mutantes en utilisant des amorces spécifiques pour amplifier les bordures de l'ADN-T ou le gène étudié par PCR (figure 12).

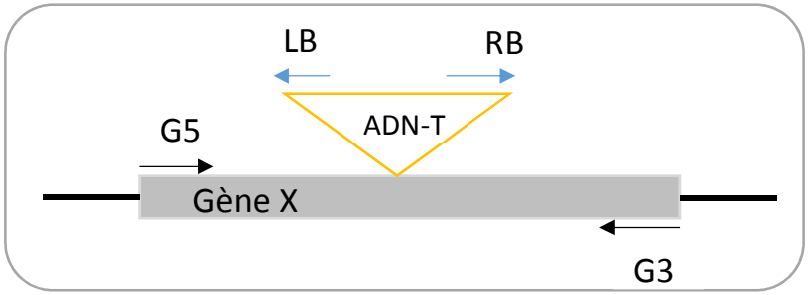


Figure 12. Représentation schématique de l'amplification par PCR des bordures d'insertion dans les mutants d'*Arabidopsis thaliana*. G3 et G5 sont les amorces spécifiques du gène muté de part et d'autre de l'insertion de l'ADN-T. LB et RB sont les amorces spécifiques de l'ADN-T inséré dans le génome d'*Arabidopsis thaliana*. La position de l'insertion de l'ADN-T dans le gène d'intérêt a été repérée sur le site https://seqviewer.arabidopsis.org/

Afin d'isoler des mutants homozygotes, des amplifications ont été réalisées en utilisant toutes les combinaisons possibles entre les amorces spécifiques des gènes (Tableau 3) et de l'insertion.

Parmi les gènes candidats sélectionnés, quatre mutants correspondants ont révélé un phénotype embryon létal suggérant qu'ils ne pouvaient être viables à l'état homozygote (voir Tableau 2, colonne « analyse PCR »). Il est important de noter que les quatre mutants homozygotes isolés ne présentaient pas de phénotype particulier par rapport à la plante sauvage.

Tableau 3 : liste des oligonucléotides utilisés pour l'isolation de mutants homozygotes

Protéine	Mutant	Nom de l'amorce	Séquence (5'-3')
HP20	N656245	Sens	GGCCTGTGTAGACTTCATTGGTG
		Antisens	GAGACAGAGTGAGCCAAAGCCG
NPH3	N862390	Sens	CAAGGCATATAGGTTTGTTCCAGTTGG
		Antisens	GTAAAGCTTACAGAGCCAATACAAGCC
EMB2777	N536872	Sens	CCAATGGAGTCCTTGGTTCTACTG
		Antisens	GGGTCAATCCTCTGTTACTCACC
LRRIII	N717997	Sens	CCAGATTCCACCGGATTTAACCG
		Antisens	CACATGTTTTGGTCCTACTCCTACC
ClpD	N686643	Sens	GTTCTGTGTGGATCTGACAGCAC
		Antisens	CCAACAAAATCTTCACCGCATCTTCC
ClpB3	N559520	Sens	GAGCTAATGGCGACGGCTACG
		Antisens	CAACACAGGGTTGTTCTTTGTTCTCC
ClpR3	N669071	Sens	GAGCATGTTGGTTCTTATTTGTGCAGG
		Antisens	CCTCTGAAGGAACCACGTCCG
HIR2	N657289	Sens	CTTCAGCAACTCGATGTTCAGTGTG
		Antisens	GAATCTGCGCTGCCACGTCAG

2. Expression *in planta* des versions GFP-étiquetées des protéines chloroplastiques non canoniques dans les mutants des candidats récepteurs d'import

a) Production de plantes transgéniques d'Arabidopsis surexprimant les protéines non canoniques :

Le système d'*Agrobactérium tumefaciens* permet également l'intégration aléatoire, dans le génome d'*Arabidopsis thaliana,* d'un ADN-T fusionné à la séquence codante d'une protéine d'intérêt associée à un promoteur fort. Cette combinaison permet la surexpression de la protéine codée par cet ADNc via l'obtention d'une plante dite transgénique. Ainsi, pour étudier l'impact de l'absence des gènes sélectionnés sur la localisation subcellulaire des protéines chloroplastiques non canoniques étudiées, des versions GFP-étiquetées de ces protéines non canoniques ont été produites dans un système de clonage binaire (pourvu d'un marqueur de sélection) permettant l'intégration et l'expression de ces protéines *in planta* (voir partie Matériel et méthodes). Les mutants d'*Arabidopsis thaliana* ont été ensuite transformés à l'aide d'*Agrobactérium tumefaciens* portant la séquence codante des fusions de la protéine non canonique correspondante et de la GFP (ceQORH::GFP, NQR::GFP ou Tic32::GFP) (Tableau 4). Outre la transformation par ceQORH::GFP, les plantes mutantes clpD et LRRIII ont été également transformées avec la construction Mut2::GFP (Tableau 4). En effet, comme mentionné dans l'introduction, il a été démontré que la protéine ceQORH interagit avec les protéines de la famille des calmodulines [134]. Cette interaction est à l'origine de la localisation variable de la protéine ceQORH dans la cellule (plutôt localisée dans la membrane plasmique dans l'épiderme) (voir chapitre III). Mut2 représente une version mutée de la protéine ceQORH, qui a perdu sa capacité d'interaction avec les calmodulines (et donc toujours localisée par défaut dans les chloroplastes). Comme le montre le tableau 4, une sélection des plantes transformées a été ensuite réalisée sur la base de leur résistance à la kanamycine ou à l'herbicide Basta. Des plantes sauvages ont été également transformées par les mêmes constructions et dans les mêmes conditions, qui nous ont servi de contrôles.

Tableau 4: aperçu des transformations effectuées

Protéine	Interaction	Mutant	Resistance	Transformation	sélection
HP20	NQR	N656245	Kanamycine	NQR::GFP-pt-gb	Basta
NPH3	NQR	N862390	Basta	NQR::GFP-pEL	Kanamycine
EMB2777	NQR	N536872	Kanamycine	NQR::GFP-pt-gb	Basta
LRRIII	ceQORH	N717997	Kanamycine	ceQORH::GFP-pt-gb Mut2::GFP-pt-gb	Basta
ClpD	ceQORH	N686643	Kanamycine	ceQORH::GFP-pt-gb Mut2::GFP-pt-gb	Basta
ClpB3	ceQORH	N559520	Kanamycine	ceQORH::GFP-pt-gb Mut2::GFP-pt-gb	Basta
ClpR3	ceQORH	N669071	Kanamycine	ceQORH::GFP-pt-gb Mut2::GFP-pt-gb	Basta
HIR2	Tic32	N657289	Kanamycine	Tic32::GFP-pt-gb	Basta

b) Sélection de transformants :

Après sélection des plantes transgéniques produites reposant sur la résistance à la kanamycine ou à l'herbicide Basta (figure 13, A), nous avons pu identifier plusieurs transformants dont la majorité étaient en fait de faux transformants (plantes résistantes mais qui n'expriment pas ou très peu la protéine d'intérêt). Ainsi, dans une seconde étape de sélection, nous avons utilisé des anticorps dirigés contre la protéine GFP pour identifier les plantes qui surexpriment effectivement la protéine de fusion recombinante (figure 13, B).

La figure 13-B montrent que le niveau d'expression des protéines est variable dans les plantes transgéniques analysées (cette variabilité est visible par exemple dans le cas de la protéine Tic32 en comparant les plantes 3, 4 et 6). Cette variation d'expression peut être expliquée par l'intégration de l'ADN-T dans des régions plus ou moins compatibles avec la transcription des transgènes. Pour étudier l'impact des mutations sur la localisation subcellulaire des protéines chloroplastiques non canoniques, nous avons utilisé les plantes qui présentent un niveau d'expression fort de ces protéines (le cas des pistes 4, 5 et 7 pour la protéine Tic32, figure 13-B).

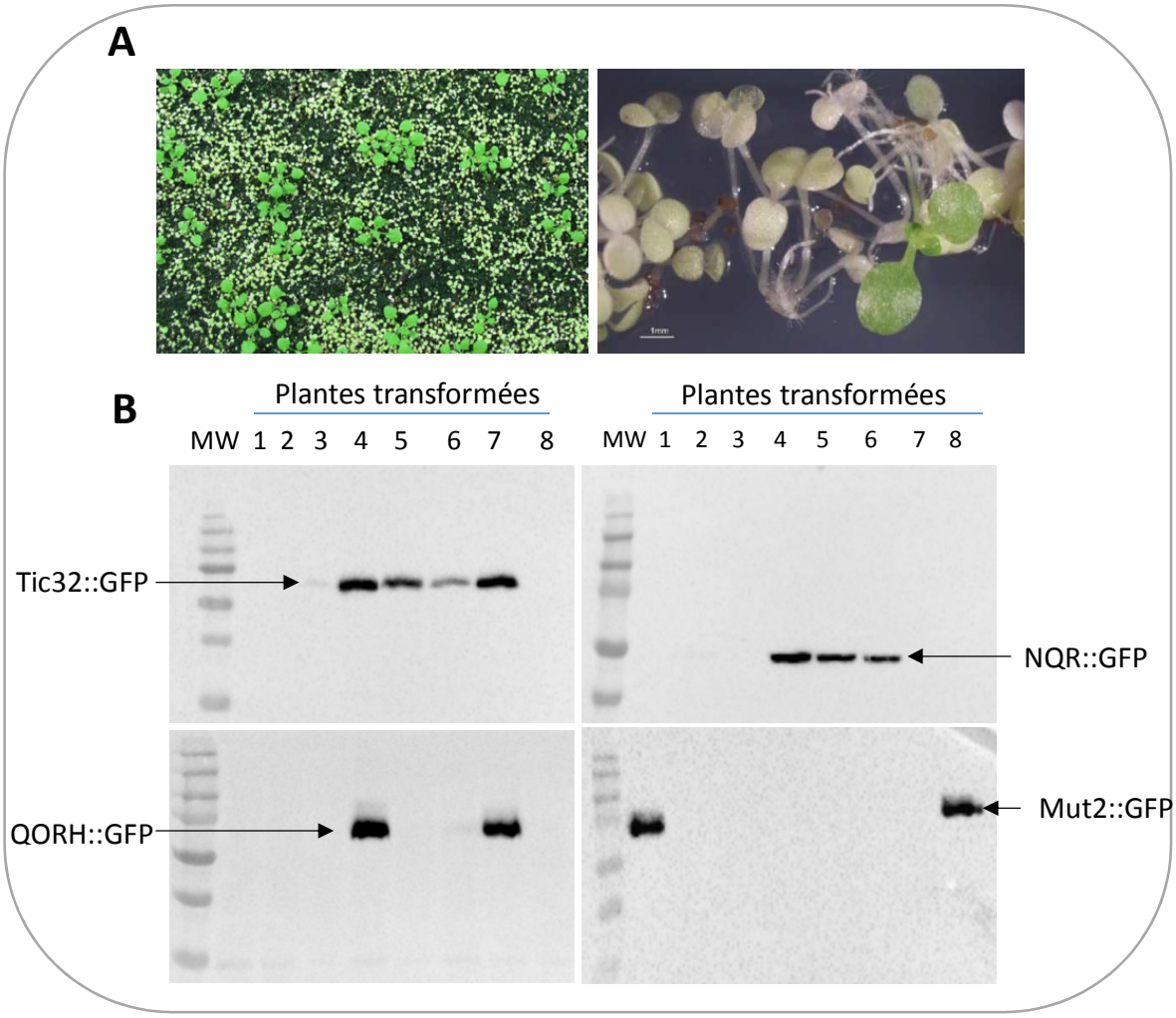


Figure 13. Stratégie de sélection des plantes transformées. A) Sélection des plantes résistantes après traitement à l'herbicide Basta (à gauche) ou à la kanamycine sur milieu MS (à droite). B) Visualisation par western blot des protéines non canoniques Tic32, NQR, ceQORH ou ceQORH muté (Mut2) fusionnées à la GFP dans des feuilles d'*Arabidopsis thaliana* de différentes plantes transformées. Les pistes 1 à 8 correspondent aux extraits totaux de feuilles (10 µg) de 8 plantes mutantes analysées par immunomarquage en utilisant l'anticorps anti-GFP. Le protocole d'extraction des protéines totales est décrit dans la partie « Matériels et méthodes ».

3. Analyse de l'impact des mutations sur la localisation subcellulaire des protéines non canoniques par microscope confocal

a) Mutants homozygotes

Dans le cas des plantes homozygotes, l'analyse de la localisation des protéines étudiées a été directement faite sur les feuilles des rosettes des transformants déjà validés (observations de la résistance à la kanamycine ou au Basta et de la surexpression de la protéine fusionnée à la GFP).

- Impact de l'absence de HP20 sur la localisation de la protéine NQR :

Comme cité plus haut, nous avons révélé de façon inattendue que HP20 (AT4G26670) interagit avec la protéine NQR. En effet, une étude récente [100], reposant également sur des approches de pontage chimique, a montré que cette protéine membranaire de l'enveloppe chloroplastique (HP20) interagit avec la protéine ceQORH et qu'elle est nécessaire à son adressage chloroplastique *in vitro* [100]. Outre HP20, deux autres protéines de la même famille « PRAT » supplémentaires ont été identifiées, désignées HP30 (At3g49560) et HP30-2 (At5g24650) [6, 97, 101], qui sont présentes dans la membrane interne de l'enveloppe chloroplastique et coopèrent pour l'import de la protéine ceQORH dans les chloroplastes selon la même étude [100]. Pour cette raison, nous avons testé l'implication potentielle de HP20 dans l'import de la protéine NQR, mais aussi dans l'adressage de la protéine ceQORH dans les chloroplastes. Autrement dit, le mutant hp20 a été transformé pour produire chacune des trois protéines de fusion ceQORH::GFP, Mut2::GFP et NQR::GFP. Comme le montre la figure 14, nous n'avons noté aucun impact de la mutation HP20 sur la localisation subcellulaire des trois protéines ceQORH, Mut2 et NQR. En effet, la co-localisation de l'auto-fluorescence de la chlorophylle et la fluorescence de la GFP (figure 14) montrent que similairement à la plante sauvage, la protéine NQR est localisée dans le stroma dans le mutant hp20. De même, la protéine Mut2 semble être importée dans les chloroplastes dans le mutant hp20, tandis que la protéine native ceQORH est localisée dans la membrane plasmique au niveau de l'épiderme. Ces résultats préliminaires ne nous permettent donc pas de confirmer que HP20 est impliquée dans le processus d'import des protéines ceQORH, ni de valider un rôle de la protéine HP20 dans l'import de NQR.

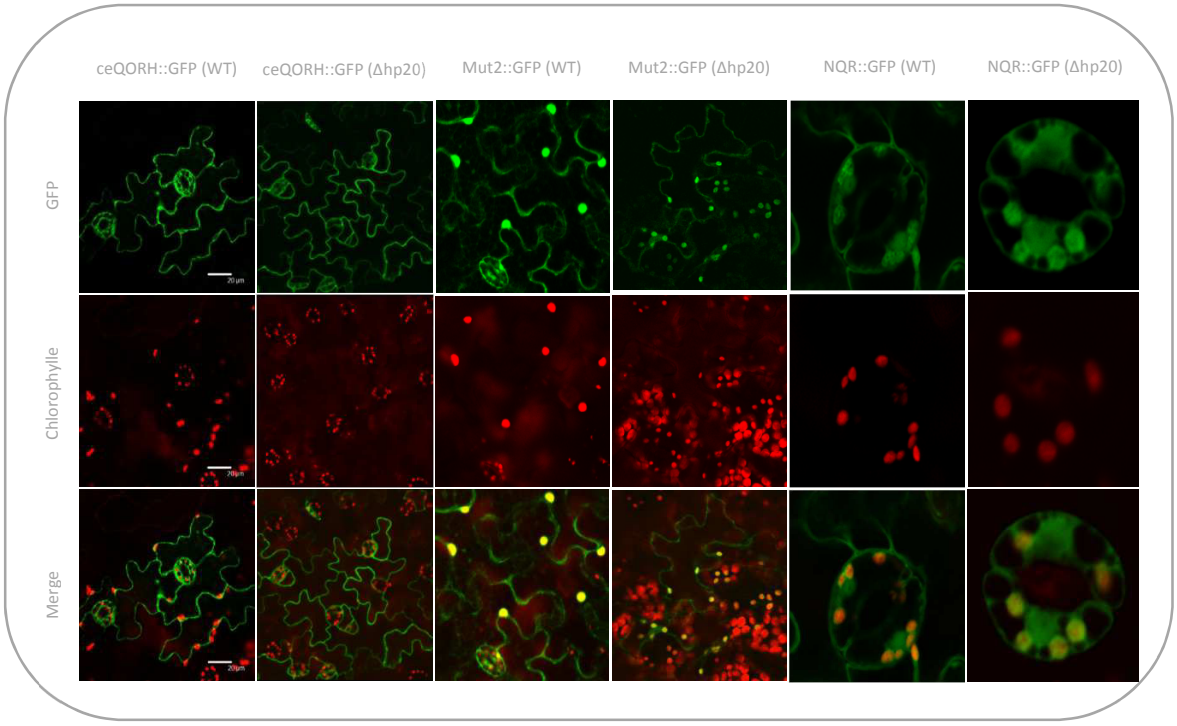


Figure 14. Analyse de la localisation des protéines ceQORH et NQR dans le mutant hp20. La localisation a été déduite grâce à l'autofluoresence de la chlorophylle (rouge) et la fluorescence de la GFP (vert).

- Impact de l'absence de NPH3 sur la localisation de la protéine NQR :

De même, comme mentionné ci-dessus NPH3 [137] a été également sélectionnée comme récepteur potentiel susceptible d'être impliqué dans l'import alternatif de la protéine NQR. Pour cela, nous avons testé l'implication putative de ce récepteur via transformation du mutant d'Arabidopsis déficient dans l'expression de ce candidat récepteur par la version GFP-étiqueté de la protéine NQR. L'analyse de l'impact de l'absence de NPH3 sur la localisation de la protéine NQR au sein de la cellule a été faite par microscopie confocale. Le résultat obtenu a montré que la protéine NPH3 n'est pas nécessaire à l'import de NQR dans les plastes (résultat non montré).

- Impact de l'absence de LRRIII sur la localisation de la protéine ceQORH :

Comme mentionné ci-dessus la kinase transmembranaire LRRIII [143] a interagi avec la protéine ceQORH lors de la réaction du pontage chimique. Du fait de ses propriétés structurales [140], cette protéine a été retenue comme récepteur d'import potentiel. L'implication putative de ce candidat récepteur dans l'import de la protéine ceQORH a été testée par production de plantes transgéniques

déficientes dans l'expression de LRRIII et qui surexpriment les protéines ceQORH::GFP/ Mut2::GFP. Les résultats de l'imagerie confocale ont montré que l'absence de la kinase LRRIII n'affecte pas l'import de la protéine ceQORH dans les chloroplastes (résultats non montrés).

- Impact de l'absence de ClpD sur la localisation de la protéine ceQORH :

  i) Résultats :

Comme mentionné plus haut (voir Tableau 1), la chaperonne ClpD [150], et d'autres composants du même complexe Clp, sont spécifiquement enrichis dans la fraction éluée en utilisant ceQORH comme appât. Il est très connu que le système protéasique Clp joue un rôle central dans le fonctionnement et le maintien de l'homéostasie des protéines plastidiales [151, 152], grâce à l'élimination sélective de protéines indésirables. Comme le montre la figure 15 le système Clp d'Arabidopsis consiste en un noyau protéolytique tétradécamère avec des sous-unités catalytiquement actives (ClpP3-6) et des sous-unités inactives (ClpR1-4) [144], des chaperons hexamères dépendants de l'ATP (ClpC1-2 et ClpD) [146, 153], une protéine adaptatrice (ClpS1) qui favorise la livraison de substrats et de deux protéines accessoires (ClpT1-2) [154] (figure 15). Seul ClpB3 n'a pas de domaine IGF essentiel pour l'interaction avec le complexe Clp et ainsi ne jouerait pas un rôle dans le système protéolytique Clp [147]. Cependant, cette sous-unité ClpB3 joue un rôle crucial dans le développement des plastes et intervient notamment en condition de stress thermique [155]. Le premier composant Clp localisé dans le chloroplaste a été découvert en 1993 par l'équipe de Kenneth Keegstra en tant que protéine principalement localisé dans le stroma des plastes d'épinard mais également présent dans l'enveloppe de chloroplaste [156]. Cela a été rapidement suivi par la découverte de ClpP1 et ClpC dans les chloroplastes d'*Arabidopsis thaliana* [157]. En 2001, il a été démontré que les protéases Clp dans le chloroplaste d'Arabidopsis forment un complexe protéolytique similaire aux protéases Clp présentes chez les procaryotes [158]. De plus, il s'est avéré que la protéine ClpP1 est essentielle au développement du chloroplaste [159]. Plus tard, du fait de la disponibilité du génome d'Arabidopsis, de l'émergence de la collection de mutants d'insertion d'Arabidopsis et de l'amélioration de la spectrométrie de masse, des études approfondies de l'organisation du système Clp dans les chloroplastes des plantes supérieures ainsi que de son rôle physiologique ont été réalisées [144-148, 153, 154]. Au cours de la dernière décennie, des études basées sur la génétique inverse [160-164] ont montré l'implication du complexe Clp dans l'embryogenèse, la croissance cellulaire et la biogenèse des chloroplastes. En effet, il s'est avéré que la perte de ClpP4 ou ClpP5 entraîne un arrêt du développement de l'embryon

au stade globulaire. La perte de ClpP1 a entraîné une perte rapide de la fonction plastidiale et une perte de la viabilité et ainsi la mort cellulaire [161, 164]. De même, les mutants affectés dans l'expression de ClpR2, ClpR4 ou ClpP3 ont montré une embryogenèse retardée par rapport aux plantes sauvages [165-167]. Un phénotype très faible à fertilité normale a été observé suite à une mutation au niveau de ClpR1. Ce phénotype modéré est due à une substitution partielle par ClpR3 [160, 166]. Toutes ces observations montrent que la plupart, des sous-unités centrales du complexe protéolytique Clp ont des contributions structurelles et fonctionnelles distinctes. Les mutants des composants accessoires ClpT1 et ClpT2 ne montrent pas un phénotype visible. Cependant, le double mutant (ClpT1 et ClpT2) présente un phénotype embryon létal [168]. Comme mentionnée ci-dessus, le complexe Clp des plastes est formé de trois protéines chaperonnes de la famille HSP 100, désignées ClpC1, ClpC2 et ClpD [153]. Ces composants sont censés se fixer aux pores axiaux du système Clp et ensuite transférer les polypeptides non repliés destinés à la dégradation. Les études de génétique inverse faites sur ces composants chaperons ont montré que le mutant ClpC2 ne montre pas de phénotype particulier contrairement au mutant ClpC1 qui a entrainé une croissance retardée et une chlorose modérée à tous les stades de développement de la plante [169, 170]. La perte complète de deux protéines (double mutant ClpC1×ClpC2) a entrainé une létalité de l'embryon [153]. Les phénotypes visibles distincts entre les deux protéines ClpC1 et ClpC2 peuvent être attribués à leurs niveaux d'accumulation différents et à une certaine complémentarité fonctionnelle. Il a été également noté que la perte de la fonction de ClpC1 a entrainé une diminution de l'import des précurseurs dans les chloroplastes [171, 172]. De plus, il s'est avéré que la protéine ClpC1 est localisée dans le stroma des plastes de *Pisum sativum* à proximité de la membrane interne de l'enveloppe chloroplastique, interagissant ainsi avec les composants de la machinerie d'import TIC des protéines précurseurs (notamment avec Tic110 et Tic40) [171]. Bien que cette interaction reste encore énigmatique, des chercheurs proposent l'hypothèse d'une stimulation de l'activité ATPase de la machinerie d'import par ClpC1, fournissant ainsi l'énergie nécessaire pour la translocation [173]. Plusieurs autres études ont également montré l'interaction directe entre les chaperonnes HSP100 et HSP70 et les peptides de transit des précurseurs chloroplastiques lors de leur translocation à travers la machinerie d'import TIC [49, 174-176]. Ainsi, des certains chercheurs soupçonnent l'implication du complexe Clp dans le mécanisme d'import des précurseurs canoniques dans les chloroplastes [176]. Néanmoins cette hypothèse est rejetée par d'autres équipes de recherche [144].

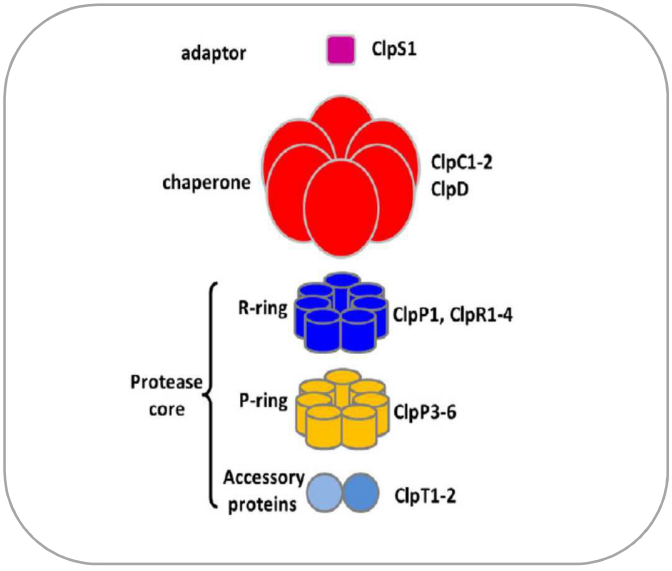


Figure15. Représentation schématique du complexe Clp du chloroplaste d'Arabidopsis (d'après Kenji Nishimura et Klaas J. van Wijk., 2014).

Contrairement aux autres sous-unités du complexe Clp plastidial, ClpD est une sous-unité spécifique des plantes [153]. En effet, ce composant Clp désigné aussi ERD1 ou SAG15 n'existe pas dans les systèmes bactériens [150, 153]. Cette protéine a été localisée dans le stroma des plastes [150], s'accumule non seulement dans les feuilles sénescentes mais aussi dans les jeunes feuilles [144, 150] et joue un rôle important dans l'acclimatation de la plante notamment en conditions de stress [148]. En effet, il s'est avéré que la protéine ClpD est sous le contrôle d'un promoteur qui est induit dans les conditions de stress à savoir la sécheresse et la concentration élevée de la salinité et de l'acide abscissique [177]. Bien que considérée comme exclusivement localisée dans le stroma [178, 179], ClpD est partagée entre le stroma et l'enveloppe du chloroplaste dans la base de données AT_CHLORO. Comme la protéine ClpC2, il s'est avéré aussi que la sous-unité ClpD a la capacité de s'associer avec un peptide de transit [153]. Cependant, aucune étude n'a montré leur implication effective dans l'import des précurseurs et leur rôle exact dans la translocation demeure toujours inconnu.

Comme indiqué ci-dessus, cette protéine est susceptible d'être impliquée dans l'import de la protéine chloroplastique non canonique ceQORH du fait de son enrichissement spécifique dans la fraction éluée en utilisant ceQORH comme appât pendant la réaction de pontage chimique. Son rôle connu de chaperonne semble compatible avec un rôle potentiel dans la voie d'import non-canonique. Ainsi, afin d'examiner l'implication potentielle de la protéase ClpD dans le contrôle de

la localisation de la protéine ceQORH dans la cellule, des plantes transgéniques qui sont affectées dans l'expression de la protéine ClpD et qui surexpriment la protéine ceQORH et Mut2 ont été produites. Là aussi, l'impact de la mutation ClpD sur la localisation de la protéine ceQORH a été analysé par microscopie confocale (figure 16). Il est important de noter que l'imagerie confocale a été réalisée sur trois transformants différents. Comme le montre la figure 16, un profil de fluorescence identique entre les trois plantes transgéniques et différent de celui de la plante sauvage a été observé. En effet, contrairement à la plante sauvage où la protéine ceQORH apparait préférentiellement localisée dans la membrane plasmique, sa localisation subcellulaire est affectée dans le mutant clpD : une suraccumulation de ceQORH à la surface des chloroplastes a été observée. Cependant, l'absence de ClpD n'a aucun effet sur la localisation de la protéine Mut2 (résultat non montré) qui a perdu la capacité d'interagir avec son partenaire de type calmoduline (voir chapitre III). Ces résultats suggèrent que ClpD joue un rôle central dans le contrôle de l'adressage de la protéine ceQORH.

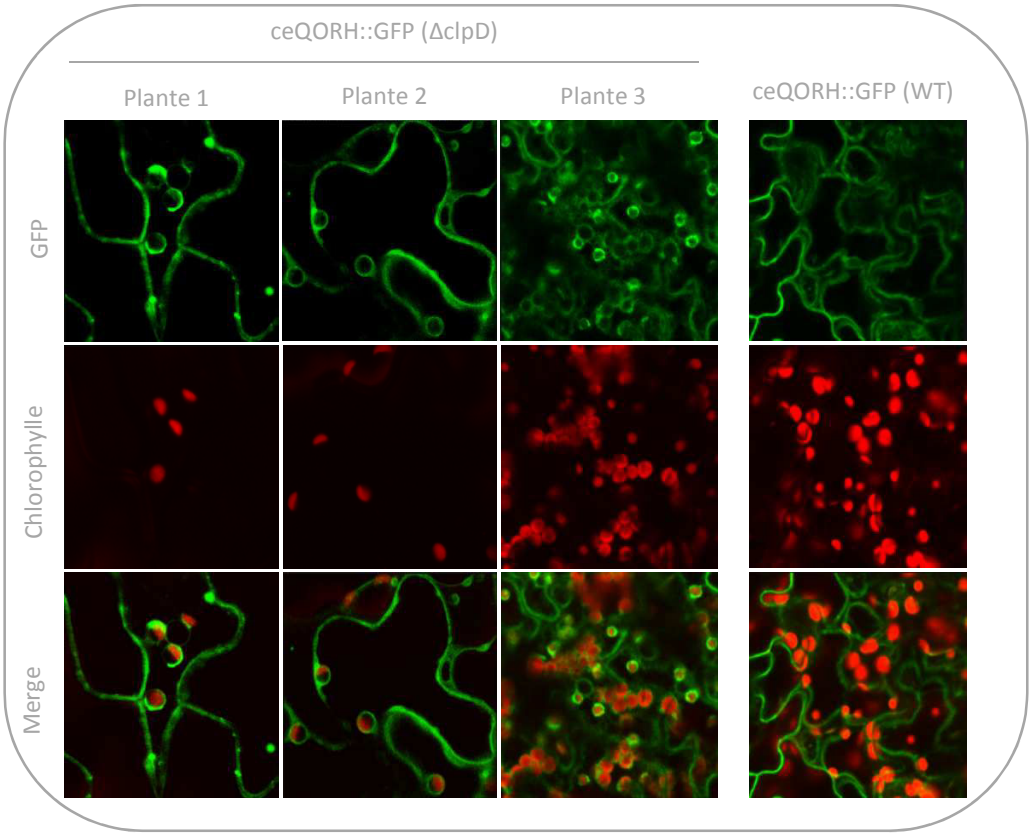


Figure 16. Analyse de la localisation de la protéine ceQORH dans le mutant *clpD*. De gauche à droite, transformant 1, transformant 2, transformant 3 et plante sauvage qui surexpriment la protéine ceQORH::GFP.

Comme la montre la figure 16, nous constatons aussi l'apparition d'amas marqués par la GFP à la surface des chloroplastes et notamment à l'interface entre les chloroplastes et le noyau (figure 17). Cela nous a conduits à soupçonner une corrélation de la fluorescence observée avec le réticulum endoplasmique (RE) dans le mutant clpD. Ainsi, une localisation spécifique du RE a été testée sur des feuilles des trois transformants utilisés pour étudier l'impact de l'absence de ClpD sur la localisation de ceQORH (figure 16). Une coloration des tissus avec de la Rhodamine (un marqueur du RE) a été réalisée par infiltration des feuilles d'Arabidopsis. Après incubation de quelques minutes à température ambiante avec de la Rhodamine, une analyse de la localisation a été ensuite réalisée par microscopie confocale (figure 17). Comme le montre la figure 17, une parfaite corrélation entre la fluorescence de la GFP et celle de la Rhodamine a été observée. Ce résultat a été validé par un marquage du noyau avec le DAPI (molécule fluorescente capable de se lier fortement aux bases adénine et thymine de l'ADN). Une fluorescence à la surface du noyau et à l'interface des plastes a été aussi observée (résultat non montré). Ce résultat s'est avéré être parfaitement en accord avec notre hypothèse. Sur la base de ces observations, les questions qui se posent est de savoir quel lien entre le composant protéolytique ClpD et le trafic vésiculaire vers le chloroplaste ? La protéine ceQORH joue-t-elle un rôle dans le trafic vésiculaire ou elle est simplement transportée dans ces vésicules ?

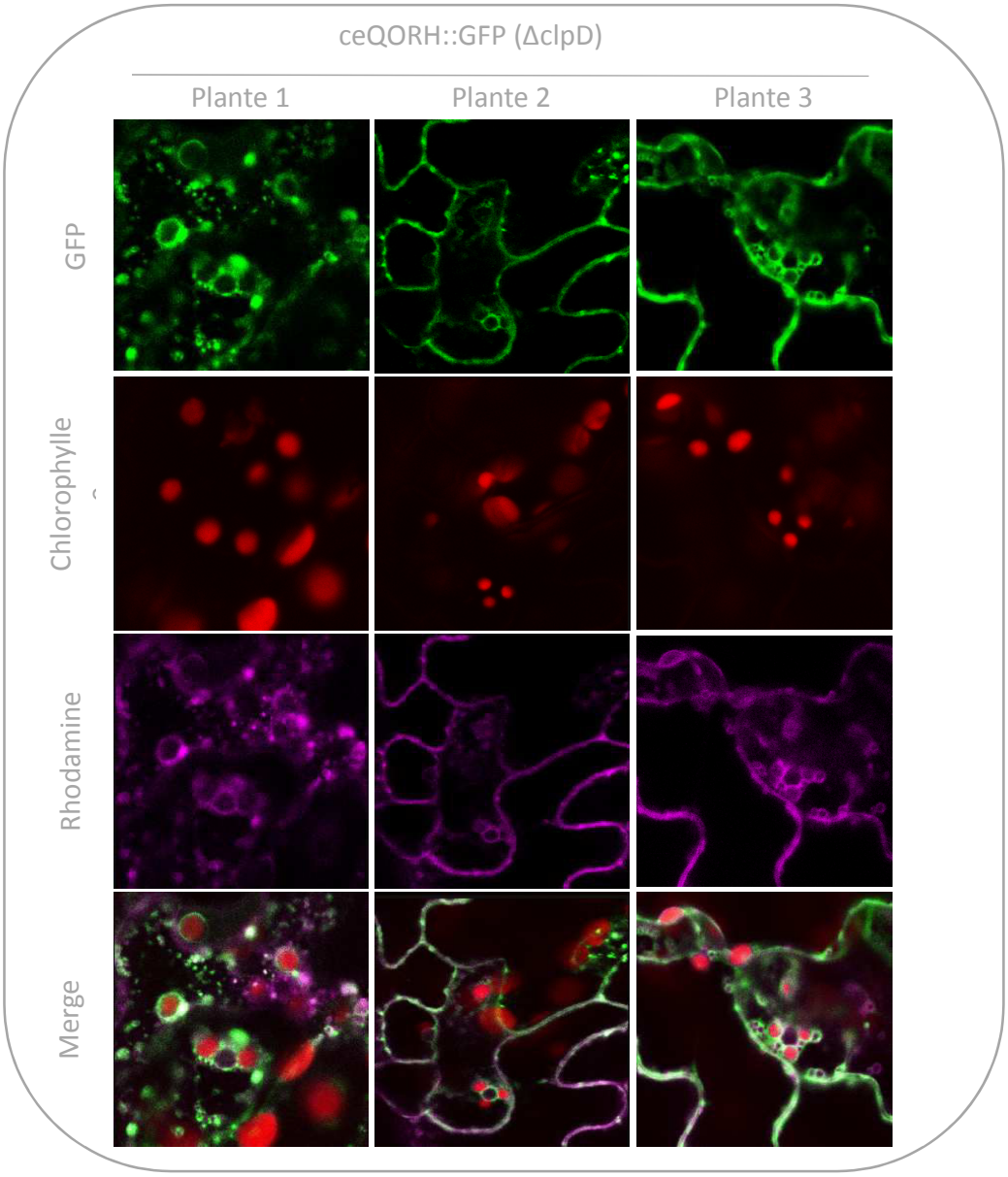


Figure 17. Analyse de la localisation du RE dans le mutant clpD qui surexprime la protéine ceQORH::GFP. Le RE a été marqué par la rhodamine (violet). La corrélation entre RE et la protéine ceQORH::GFP a été déduite de la superposition de la fluorescence (vert/violet) de la GFP et la rhodamine, respectivement.

Pour vérifier que l'insertion supposée inactiver le gène ClpD a bien comme résultat l'absence de la protéine ClpD, un anticorps dirigé contre la protéine ClpD a été ensuite utilisé pour détecter cette protéine. Le mutant clpD décrit ci-dessus, ainsi qu'un deuxième mutant ClpD ont été utilisés. Deux

autres lignées d'insertions ont été isolées afin de reproduire l'analyse de l'impact de l'absence de ClpD sur la localisation de ceQORH dans un second mutant clpD.

Comme le montre la figure 18, la mutation au niveau de clpD a bien été validée. En effet, deux bandes (ClpD et ClpC2) ont été visualisées dans la piste 1 qui contient l'extrait protéique total d'une plante sauvage contre une seule bande (ClpC2 uniquement) dans la piste 2 qui contient l'extrait protéique total du mutant clpD. Le même profil a été observé dans la piste 3 qui contient l'extrait total de la deuxième lignée d'insertion clpD. Cependant, le gène clpD semblait ne pas être inactivé dans le mutant 3. Ainsi, la même expérience a été reproduite en transformant le deuxième mutant (mutant 2) avec la protéine ceQORH::GFP.

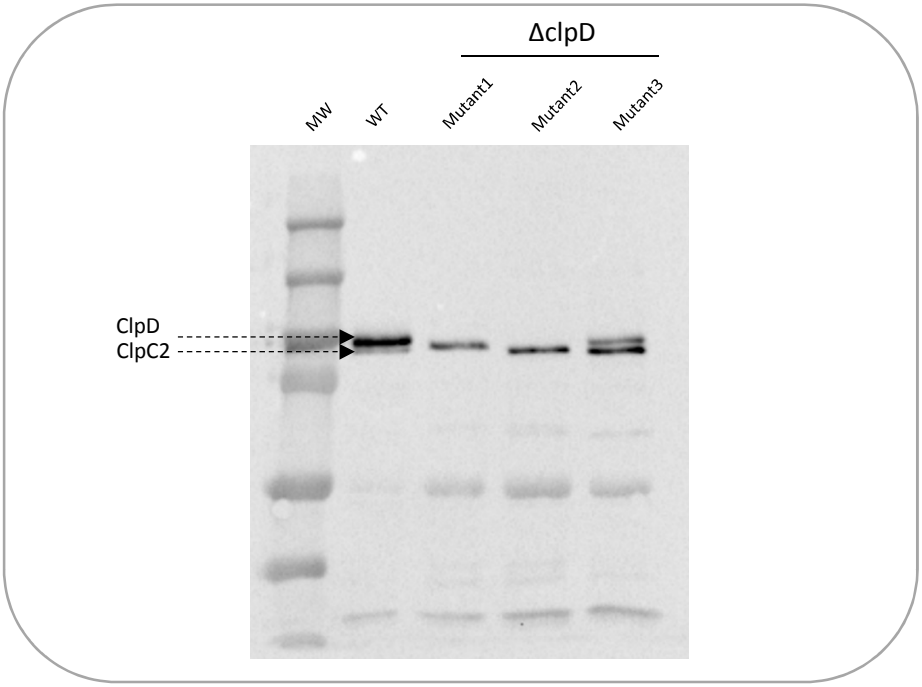


Figure 18. Validation des mutants clpD par immunomarquage. Chaque colonne contient 10 µg d'extrait protéique total de feuille d'Arabidopsis. L'immunomarquage a été réalisé en utilisant l'anticorps anti-ClpD (1:5000).

ii) Discussion :

Comme indiqué dans l'introduction (section III.2), certains chercheurs soupçonnent l'existence d'un lien probable entre l'interaction du RE avec la surface du chloroplaste et la translocation de certaines protéines chloroplastiques codées par le génome nucléaire via ce système endomembranaire. La première évidence indirecte d'un tel mécanisme d'import impliquant le RE

et l'appareil de Golgi résulte de l'identification de glycoprotéines adressées aux chloroplastes par Villarejo et *al* [90].

Pour revenir à nos résultats (figure 17), une parfaite corrélation a été observée entre la fluorescence de la GFP associée à la protéine ceQORH et la fluorescence de la Rhodamine (marqueur du RE). Ce profil observé (figure 17) est similaire à celui révélé par notre équipe (voir chapitre III) en bloquant le trafic vésiculaire du système endomembranaire en traitant des feuilles des plantes qui expriment la ceQORH::GFP par la bréfeldine A (BFA) (une parfaite corrélation entre les vésicules accumulées après traitement BFA et la GFP associée à la protéine ceQORH). Dans ce contexte, il est important de rappeler qu'une étude récente [180] a détecté la protéine ceQORH dans le réseau trans-Golgi (TGN) en utilisant la purification par affinité et un marqueur TGN comme appât. Dans cette étude ciblant le Golgi, ceQORH était considérée comme un contaminant du chloroplaste. Ces résultats nous ont incités à une première hypothèse : i) la protéine ceQORH se déplace par le système vésiculaire via le réticulum endoplasmique avant son transit dans le chloroplaste. Cette hypothèse est aussi attractive du fait de l'enrichissement spécifique de la clathrine (protéine impliquée dans le trafic vésiculaire dans le système CCV) et les microtubules dans la fraction éluée de la purification par affinité réalisée en utilisant ceQORH comme appât. Cette interaction directe entre la clathrine (acteur majeur de transport vésiculaire) et ceQORH nous incitent à réfléchir sur le statut exact de la protéine ceQORH dans le système vésiculaire : est-ce qu'elle joue un rôle dans le trafic vésiculaire ou est-elle simplement transportée dans des vésicules ?

A noter que dans le mutant clpD, l'accumulation de la protéine ceQORH est observée à la surface des chloroplastes. Cela suggère l'existence d'un lien probable entre la chaperonne ClpD et le trafic vésiculaire vers les plastes. En effet, il semblerait que l'absence de ClpD ralentisse fortement le transport vésiculaire vers les plastes, d'où la possibilité de visualisation du phénomène d'accumulation par imagerie confocale. Dans ce contexte, il est important de rappeler qu'il a été démontré que la perte de l'activité de la protéase ClpC1 (faisant partie de la même famille que ClpD) a entrainé une diminution de l'import « classique » des précurseurs dans les plastes [41, 42]. Ainsi, par analogie, l'accumulation de ceQORH à la surface des chloroplastes (figure 16) peut être expliquée par un ralentissement d'une étape précoce dans le mécanisme d'import « non-canonique » dans le mutant clpD. En conséquence, toutes ces données nous ont permis d'avancer une seconde hypothèse : ii) la protéine ClpD affecterait indirectement la localisation sub-cellulaire

de ceQORH en déstabilisant l'interaction entre ceQORH et les vésicules du RE. Dans ce cas, la protéine ceQORH intègre le RE-Golgi avant son transit dans les chloroplastes via des vésicules et ClpD est en contact direct avec la protéine ceQORH à la surface du plaste. Cependant, avant que cette hypothèse soit validée, il faudrait montrer que Clp est aussi localisée dans ces vésicules, ce qui n'est pas le cas pour le moment. En effet, bien qu'elle soit considérée par certains comme uniquement localisée dans le stroma [178, 179], ClpD est partagée entre le stroma et l'enveloppe du chloroplaste dans la base de données AT_CHLORO, mais sans preuve d'une localisation dans la membrane externe de l'enveloppe. Il est envisageable que ClpD agisse seule, indépendamment du complexe Clp, ou au contraire reste associée à ce complexe. La mise en évidence par les expériences de pontage d'autres sous-unités Clp (voir Tableau 1) est en faveur de cette seconde possibilité. Si cela est le cas, il est possible d'envisager une troisième hypothèse : le complexe agit dans une phase plus tardive de l'import de ceQORH (au niveau de la membrane interne de l'enveloppe) et les phénomènes observés à la surface du chloroplaste résulteraient d'un « engorgement » du système d'import non-canonique.

Le fait que le mutant MUT2 soit correctement importé dans les plastes indique que le phénomène d'import non-canonique est plus complexe et implique la participation d'une calmoduline (voir Chapitre III).

b) Mutants hétérozygotes (avec phénotype embryon létal)

Parmi les candidats récepteurs sélectionnés, certains mutants correspondants ont révélé un phénotype embryon létal suggérant que ces gènes sont essentiels au développement de la plante. En conséquence, ces mutants ne peuvent pas être viables à l'état homozygote. Ainsi, il est difficile de tester le rôle de ces récepteurs putatifs dans l'import des protéines non canoniques d'intérêt dans les chloroplastes. Alternativement, afin de tester l'impact des candidats récepteurs dans le contrôle de l'adressage des protéines étudiées, nous avons eu recours à une approche alternative reposant sur l'imagerie confocale dans les graines homozygotes. Ainsi, pour savoir si les protéines ceQORH, Tic32 ou NQR sont affectées dans leur import dans le chloroplaste dans les mutants correspondants (mutants qui révèlent un phénotype embryon létal, où par conséquent le stade de la plante homozygote avec des feuilles ne peut pas être obtenu) une stratégie alternative a été examinée. Nous avons ainsi essayé de localiser les protéines non canoniques d'intérêt dans les tissus chlorophylliens de graines homozygotes pour ces mutations. Pour cela, nous avons d'abord vérifié

la présence de tissus chlorophylliens dans les graines blanches qui sont à priori les graines homozygotes (figure 19, A). Bien que le développement embryonnaire ait été arrêté dans la graine blanche du mutant Δemb2777 par rapport à la graine verte où l'embryon se développe en intégralité, une partie des tissus chlorophylliens et des plastes reste présente dans la graine blanche (figure 19, B). Cela peut s'avérer suffisant pour une localisation plastidiale des protéines étudiées. Ainsi, comme mentionné précédemment, des plantes hétérozygotes transgéniques qui expriment les protéines non canoniques correspondantes ont été produites dans le fond génétique du mutant Δemb2777 hétérozygote. Après sélection des transformants résistants à l'herbicide Basta et exprimant la fusion GFP, les plantes ont été cultivées jusqu'au stade graine. Les graines issues de ces plantes transformées ont été ensuite observées par microscopie confocale afin de localiser la protéine non canonique grâce à l'autofluorescence de la chlorophylle et la fluorescence de la GFP. A la date de rédaction de ce rapport, les résultats obtenus ne sont pas convaincants et il semble difficile de les exploiter pour tirer des conclusions définitives sur le rôle de ces protéines.

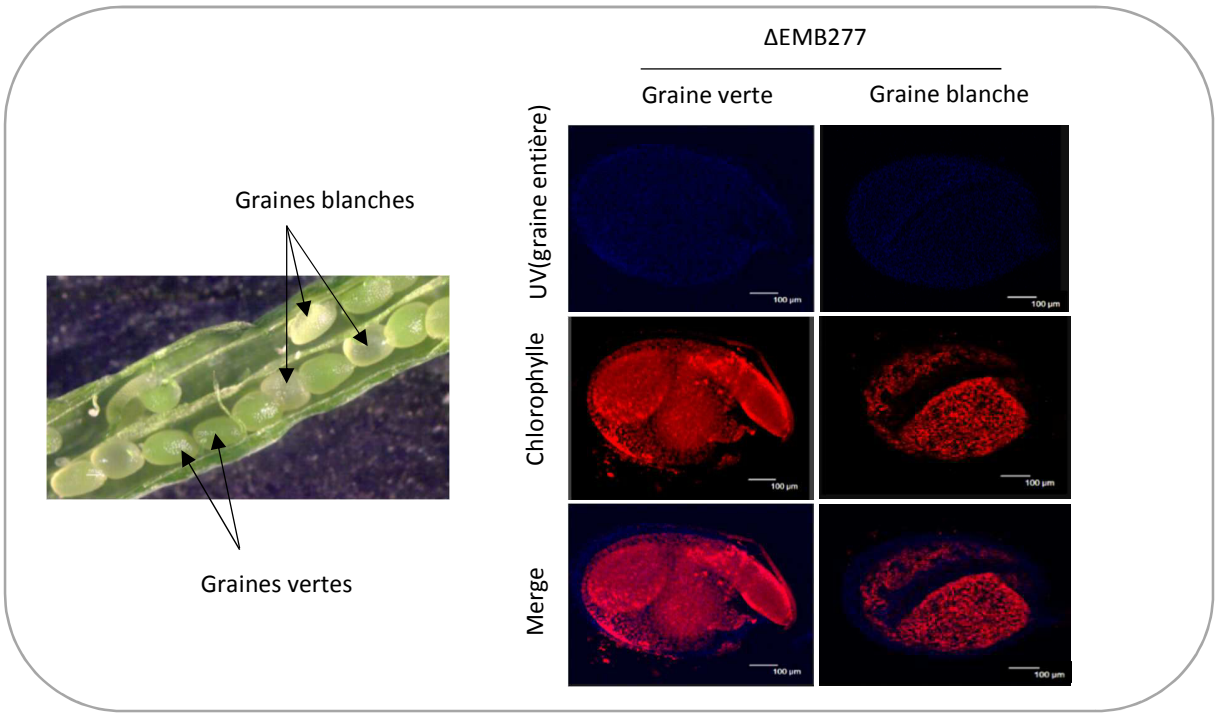


Figure 19. Vérification de la présence de tissus chlorophylliens dans les graines homozygotes des mutants présentant un phénotype embryon létal. A) photo d'une silique du mutant emb2777 avec des graines blanches homozygotes. B) imagerie confocale d'une graine verte versus graine blanche du mutant emb2777.

IV. Conclusion et perspectives

Ces travaux nous ont permis d'identifier un partenaire, la chaperonne ClpD, qui parait jouer un rôle central dans le mécanisme d'import alternatif de la protéine ceQORH. Cette protéine a la particularité de former un complexe avec d'autres sous-unités de la même famille (les autres chaperonnes Clp), qui elles-mêmes s'associent au complexe protéasique Clp dans le stroma du chloroplaste d'Arabidopsis. Ce mécanisme d'import semble aussi potentiellement dépendant de l'intervention du système vésiculaire du réticulum endoplasmique.

Les perspectives de cette partie de ce travail consisteront donc à valider l'implication putative du candidat identifié dans le contrôle de l'adressage de la protéine ceQORH dans les chloroplastes. Pour cela, nous avons initié une approche reposant sur la protéomique différentielle à l'échelle du chloroplaste sur des plantes sauvages et mutants affectés dans le récepteur potentiel ClpD, afin d'étudier le niveau d'import de la protéine ceQORH et des autres protéines chloroplastiques dans la plante sauvage versus mutant ClpD.

Comme mentionné ci-dessus, l'hypothèse d'un contact direct entre ClpD et ceQORH implique que ClpD soit localisée à la surface du plaste, outre le stroma du chloroplaste. Pour cela, nous avons initié la production de la construction recombinante de la protéine ClpD fusionnée à la GFP destinée à caractériser la protéine et à l'exprimer dans la plante.

Malgré la forte probabilité que les mécanismes d'import des deux protéines Tic32 et NQR soient indépendants de celui de la protéine ceQORH (voir introduction section III.1.b) et même si ClpD n'est pas enrichie dans les fractions éluées en utilisant Tic32 et NQR comme appâts, il sera aussi intéressant de tester la localisation de Tic32 et NQR dans le mutant clpD.

De même, une analyse phénotypique du mutant clpD est en cours afin d'étudier son rôle dans le contrôle de la physiologie de la plante par analyse des propriétés d'acclimatation aux changements environnementaux.

Comme indiqué ci-dessus, outre le composant ClpD, deux autres sous-unités ClpB3 et ClpR3 qui font partie du même complexe Clp ont été enrichies dans la fraction ceQORH suite à la réaction de pontage chimique. Ainsi, il est important de tester la localisation de ceQORH dans les mutants clpR3 et clpB3 (expériences en cours).

Pour approfondir nos résultats obtenus suggérant la non implication de la protéine HP20 dans l'import de ceQORH et NQR, nous avons aussi analysé la protéine la plus proche de HP20, nommée HP22 (AT5G55510) [97] qui est susceptible de complémenter l'absence de HP20 dans le mutant hp20. Pour cela, le mutant d'*Arabidopsis thaliana* affecté dans l'expression de HP22 a été également isolé. Nous avons généré le double mutant (Δhp20 × hp22) en croisant les deux mutants homozygotes hp22 et hp20 qui porte déjà la fusion GFP de NQR, ceQORH ou Mut2. De la même manière, l'analyse de l'impact de ces deux délétions sur la localisation subcellulaire de ces trois protéines sera étudiée par imagerie confocale.

Comme indiqué ci-dessus, certains mutants des candidats récepteurs ont montré un phénotype embryon létal. En conséquence, ces mutants ne peuvent pas être viables à l'état homozygote. Ainsi, une seconde approche alternative visera à limiter l'expression des candidats récepteurs dans les mutants hétérozygotes correspondants pendant les phases très précoces du développement de la plante : il s'agit de complémenter les mutants avec une construction qui permet d'exprimer le gène candidat seulement pendant les phases très précoces du développement sous la dépendance d'un promoteur inductible, puis d'arrêter l'induction et d'observer l'impact sur l'import de protéines non-canoniques. Nous avons ainsi initié la production de mutants transgéniques qui portent la construction qui code pour le gène d'intérêt associé à un promoteur inductible par l'éthanol. Ce système est composé de deux éléments : un facteur de transcription ALCR, qui suite au traitement de la plante par l'éthanol sera induit et activera un promoteur cible *palcA* sous le contrôle duquel est placé notre gène candidat que l'on désire exprimer dans le fond génétique où le gène endogène est inactivé. Cette approche permettra aussi d'analyser le phénotype des mutants complémentés pendant différentes phases de croissance afin d'identifier à quel stade le candidat récepteur est indispensable au développement de la plante (et éventuellement à la biogenèse du chloroplaste).

Chapitre II

Bio-analyse du protéome de l'enveloppe du chloroplaste

Article I

Unravelling hidden components of the chloroplast envelope proteome: opportunities and limits of better MS sensitivity

Authors : Imen Bouchnak[‡], Sabine Brugière[§], Lucas Moyet[‡], Sophie Le Gall[‡*], Daniel Salvi[‡], Marcel Kuntz[‡], Marianne Tardif[§], Norbert Rolland[‡¶]

Complete names of Institutions:

‡ Univ. Grenoble Alpes, INRA, CNRS, CEA, BIG-LPCV, 38000 Grenoble, France

§ Univ. Grenoble Alpes, CEA, Inserm, BIG-BGE, 38000 Grenoble, France

¶ To whom correspondence should be addressed:

Norbert Rolland, Laboratoire de Physiologie Cellulaire & Végétale, 17 rue des Martyrs, 38054 Grenoble cedex 9, France. Phone: +33 4 38 78 49 86; Fax: +33 4 38 78 50 91; E-Mail: norbert.rolland@cea.fr

*Present address: INRA-UR1268 BIA, Plate-forme BIBS, rue de la Géraudière, 44316 Nantes, France

Running Title

Chloroplast envelope genuine and shared components

Abbreviations

EF, enrichment factor; MS, mass spectrometry; SC, spectral count; SSC, specific spectral count; WSC, weighted spectral count; CCE, crude cell extract; IEM, inner envelope membrane; OEM, outer envelope membrane; THY, thylakoids, STR; Stroma.

Summary

Chloroplast is a major plant cell organelle that fulfills essential metabolic and biosynthetic functions. Located at the interface between the chloroplast and other cell compartments, the chloroplast envelope system is a strategic barrier controlling exchanges of ions, metabolites and proteins, thus regulating essential metabolic functions (synthesis of hormones precursors, amino acids, pigments, sugars, vitamins, lipids, nucleotides…) of the plant cell. However, establishing

the chloroplast envelope proteome remains a specific challenge and many proteins constituting this functional double membrane system still remain to be identified. Indeed, the envelope contains only 1% of the chloroplast proteins (*i.e.* 0.25% of the whole cell proteome). In other words, most envelope proteins are so rare at the cell, chloroplast or even envelope level, that they remained undetectable using targeted MS studies. Besides, due to the cross-contamination of chloroplast sub-compartments by each other and by other cell compartments, cell fractionation hardly allowed to provide accurate localization of many envelope proteins. The aim of the present study was to take advantage of present-days better MS sensitivity towards a better definition (differentiate genuine envelope proteins from contaminants) of the chloroplast envelope proteome. This MS-based analysis relied on an enrichment factor calculated for each protein identified in purified envelope fractions when compared to the value obtained for the same protein in crude cell extracts. Using this approach, a total of 1269 proteins was detected in purified envelope fractions, of which, 462 could be assigned an envelope localization combining MS-based spectral counts analyses and manual annotation using data from the literature or prediction tools. Interestingly, many of such proteins being unknown or unexpected genuine (or shared) envelope components, these data constitute a new resource of significant value to the broader plant science community aiming to define principles and molecular mechanisms controlling fundamental aspects of plastid biogenesis and functions.

Data are available via ProteomeXchange with identifier PXD010545.

1. Introduction

Understanding the functional diversification and evolution of the cell organelles requires the identification of their complete protein repertoires, which is the ultimate goal of organellar proteomics. Proteomics-based data with analyses of nuclear genes that include predictions of subcellular location is the method of choice to perform these analyses (1). Until now, it is technically not feasible to generate an organelle fraction totally free from other cellular components and increased sensitivity of mass spectrometry (MS) allows detecting even minimal impurities (2). The development of new software tools and improvement of mass spectrometry equipment and sensitivity have paved the way for a burst of MS data. Manual examination of these MS data is of great help to address the problem of contamination and to evaluate the validity of organelle proteome data. However, identification of a growing number of minor proteins with unknown (and

sometimes unpredictable) subcellular localization and function still raises the question of the genuine subcellular localization of these uncharacterized proteins.

Plastids are major components of plant cells. They derive from a cyanobacterial ancestor that lost most of its genes after the establishment of endosymbiosis and eventually evolved as an organelle during evolution (3). As a consequence, only one hundred of the three thousands different plastid proteins are still organelle-encoded, and most are now specified by nuclear genes (4). These ~3000 proteins, encoded by such relocated genes, must be imported into chloroplast compartments following synthesis in the cytosol (5-7). Chloroplasts are consisting of several key sub-compartments including: *i*) the envelope, a double membrane system surrounding the organelle and controlling the communication of the chloroplast with the rest of the cell (8), *ii*) the stroma, the soluble phase of the chloroplast and the main site for the conversion of carbon dioxide into carbohydrates, and *iii*) the thylakoid membrane, which is a widely organized internal membrane network where oxygenic photosynthesis takes place. These organelles are the place where essential metabolic pathways occur, such as synthesis of lipids, pigments, amino acids, vitamins, starch, precursors of plant hormones, reduction of nitrite and sulfate and photosynthesis (9). Thereby, deciphering new regulatory mechanisms that control the chloroplast dynamics and physiology, will allow generating important generic data on regulation of these essential metabolic pathways in vascular plants. For that, it is critical to characterize the chloroplast proteome as an important complementary approach allowing to obtain new information that will help to define novel guidelines to better understand the chloroplast function and biogenesis.

Chloroplast envelope membranes remains the hidden part of the chloroplast proteome and many envelope proteins remain to be characterized (known but uncharacterized envelope proteins) or identified (orphan known envelope-associated functions carried by unidentified proteins) (9). Indeed, due to their low abundance at the cell or even chloroplast scale (1–2% of the chloroplast proteins) (10), envelope proteins were historically, and remain much less-known than other plastid proteins. Subcellular and subplastidial proteomics, combining prior enrichment of subplastidial compartments was of great help to decipher the basic membrane protein composition of the envelope proteome from *Arabidopsis thaliana* and few other plants (11-13). However, due to earlier lower sensitivity of MS, these studies targeting the envelope proteome did not rely on quantitative analyses comparing crude cell extracts as a reference (see (12, 14)). Consequently, the quantitative parameter (enrichment factor or EF) resulting from the strong enrichment of the

chloroplast envelope during subcellular and subplastidial fractionation was not available when analyzing MS data and annotating identified proteins. Furthermore, contaminating proteins are often detected because they are among the most abundant proteins of other cell compartments. Increasing knowledge about the subcellular localization of more and more proteins and the establishment of sophisticated proteome processing databases (*e.g.* MASCP Gator (15), SUBA (16) …) are now expected to enhance and support evaluation of chloroplast envelope proteome results.

The aim of the present study was to take advantage of the establishment of the above-cited novel subcellular proteome databases, and of the improvement of mass spectrometry sensitivity to perform a quantitative proteome study comparing purified envelope fractions and crude cell extracts. We anticipated that these analyses would help *i*) identify specific envelope proteins with increasing accuracy *ii*) make available a novel parameter to clarify the subcellular or subplastidial localization of identified proteins and definitely exclude contaminants deriving from other plastid or cell compartments. As expected, many more protein were detected, including genuine but previously undetected chloroplast envelope proteins but also large amounts of proteins deriving from other cell compartments. Accessibility to the quantitative approach (EF parameter), thanks to better MS sensitivity, allowed revisiting the chloroplast envelope proteome and to easily exclude most contaminants from other plastid or cell compartments.

2. Experimental procedures

2.1 Plant Material and Growth Conditions

Arabidopsis thaliana plants, Wassilewskija background (Ws), were grown in culture chambers at 23°C (12-h light cycle) with a light intensity of 150 µmol/m^2/s in standard conditions (17).

2.2 Purification of Arabidopsis Chloroplast and its sub-compartments

All operations were carried out at 0–5°C. Intact chloroplasts were obtained from 400–500 g of Arabidopsis (*Arabidopsis thaliana Ws*) leaves (*i.e.* 1,800 five weeks old plants) and purified by isopyknic centrifugation using Percoll gradients (17). Purified intact chloroplasts were lysed in hypotonic medium containing protease inhibitors (10 mM MOPS-NaOH, pH 7.8, 4 mM $MgCl_2$, 1 mM PMSF, 1 mM benzamidine, and 0.5 mM amino caproic acid), and envelope was purified from the lysate by centrifugation on sucrose gradients as previously described (17). To recover the

envelope proteins, the yellow band of the sucrose gradient containing the envelope proteins was then carefully aspirated with a pipette. Recovered envelope proteins were then diluted in 10 mM MOPS–NaOH pH 7.8 buffer (containing protease inhibitors), and concentrated as a pellet by centrifugation at 110,000 *g* for 1 h (Beckman SW 41 Ti rotor) (17). Envelope proteins were diluted in 100 µL of the same medium containing protease inhibitors and stored in liquid nitrogen (an average of ~100 µg of envelope proteins was obtained from each such preparation).

2.3 Extraction of Crude Leaf Extract (CCE)

All operations were carried out at 0-5°C. Total leaf proteins were obtained from three leaves of three independent Arabidopsis (*Arabidopsis thaliana Ws*) plantlets. Leaf material was homogenized in 200 µL of extraction buffer (30 mM tetrasodium pyrophosphate, 50 mM Tris pH 6.8, SDS 1% [v/v]) and then centrifuged at 16,000 *g* for 5 minutes. Crude leaf extract (CE) was then recovered by carefully aspirating the supernatant. Total proteins concentration was measured using Bradford dye-binding method (Bio-Rad protein assay). Note that due to the risk of proteolysis in crude cell extracts, these protein samples were not stored in liquid nitrogen, but separated on SDS-PAGE for further MS analyses immediately after their preparation.

2.4 Purification of Arabidopsis membrane fractions enriched in microsomal and tonoplast membranes

Fractions enriched in microsomal and tonoplast membranes from Arabidopsis plants were purified according to a previously described protocol (18). All operations were carried out at 0–5°C. Starting material was 13 g of Arabidopsis (*Arabidopsis thaliana* Ws) leaves (from four weeks old plants). These leaves were first incubated in 25 mL of grinding buffer (0.3 M mannitol; 20 mM MOPS; 2 mM EGTA; 0.5% (w/v) polyvinylpyrrolidone 25; 0.5% (w/v) BSA (Bovine Serum Albumin); 5 mM DTT; 1 mM PMSF; 5 mM α-aminocaproic acid; 1 mM benzamidine; pH 8.0). Leaves were then ground using a pestle in a mortar containing 10 mL of micro silica sand from Fontainebleau. After addition of 200 mL of grinding buffer, the suspension was centrifuged for 10 min at 2200 *g* and the supernatant was stored for further purification steps. Then, light membrane vesicles (microsomes) were enriched from this supernatant following 4 successive centrifugation steps: 2 centrifugations of 5 min at 9500 *g* and 1 centrifugation of 15 min at 25,000 *g*. The resulting supernatant was then centrifuged for 45 min at 100,000 *g*. The resulting pellet (microsomal

membranes) was then stored on ice. This pellet was then diluted in 2 mL of washing buffer (0.3 M Mannitol; 10 mM MOPS; 1 mM EDTA; 1 mM PMSF; 5 mM α-aminocaproic acid; 1 mM benzamidine; pH 7.4) and loaded on the top of a discontinuous gradient made of 4 layers of sucrose (25 % (w/v); 30 % (w/v); 38 % (w/v); 45 % (w/v) sucrose in 50 mM MOPS; 2 mM EGTA; 2 mM PMSF; 10 mM α-aminocaproic acid; 2 mM benzamidine; pH 7.4). After a centrifugation for 4 h at 80,000 *g*, the tonoplast enriched fraction was collected at the interface of the 25-30% layers of the gradient. To remove sucrose, this fraction was then diluted 4 times in washing buffer lacking mannitol and then centrifuged for 1 h at 175,000 *g*. The resulting pellet (tonoplast) was then diluted in a minimal volume of washing buffer lacking mannitol and stored at -80°C until use.

2.5 SDS-PAGE and Western Blot Analyses

SDS-PAGE analyses were performed as described by Chua (19). For western blot analyses, gels were transferred to a nitrocellulose membrane (BA85, Schleicher & Schuell). Various markers of cell compartments were detected using several antibodies (in-house, commercially available, or gifts from colleagues): GFP (GFP-2A5, obtained from Euromedex, 67458 Mundolsheim, France), HMA1 (20) as an envelope marker, LHCP (provided by Dr. Olivier Vallon, IBPC Paris) as a thylakoid marker, KARI (provided by Dr. Renaud Dumas, LPCV Grenoble) as a marker from the stroma, Bip (obtained from Agrisera) as a marker from the ER, V-ATPase (21) as a tonoplast marker, bobTIP (provided by Prof. Nathalie Leborgne-Castel (INRA/Université de Bourgogne) directed against tonoplast TIP1.1 and FtsY (provided by Dr. Laurent Nussaume, CEA Cadarache) directed against chloroplast FtsY.

2.6 Experimental design

Total extracts and envelope fractions were prepared in triplicate. For envelope sample preparation, these triplicate are to be considered as technical replicates since the potential biological variability is already averaged accounting for the 1,800 five weeks old plants needed to obtain one sample of purified envelope fraction. Total leaf proteins were extracted from three leaves coming from three independent Arabidopsis plantlets that were used to further extract envelope fractions.

2.7 Proteins digestion for proteomics Analyses

Each protein sample (10 µg) was stacked in the top of a SDS-PAGE gel (NuPAGE 4–12%, Invitrogen) before Coomassie blue staining (R250, Bio-Rad). Gel bands of concentrated proteins were manually excised and cut in pieces before being washed by 6 successive incubations of 15 min in 25 mM NH_4HCO_3 containing 50% (v/v) acetonitrile. Gel pieces were then dehydrated in 100% acetonitrile and incubated at 53°C with 10 mM DTT in 25 mM NH_4HCO_3 for 45 min and in the dark with 55 mM iodoacetamide in 25 mM NH_4HCO_3 for 35 min. Alkylation was stopped by adding 10 mM DTT in 25 mM NH_4HCO_3 and mixing for 10 min. Gel pieces were then washed again by incubation in 25 mM NH_4HCO_3 before dehydration with 100% acetonitrile. Modified trypsin (Promega, sequencing grade) in 25 mM NH_4HCO_3 was added to the dehydrated gel pieces for an overnight incubation at 37°C. Peptides were then extracted from gel pieces in three 15-min sequential extraction steps in 30 µL of 50% acetonitrile, 30 µL of 5% formic acid and finally 30 µL of 100% acetonitrile. The pooled supernatants were then vacuum-dried.

2.8 Nano-LC-MS/MS analyses

The dried extracted peptides were resuspended in 5% acetonitrile and 0.1% trifluoroacetic acid and analyzed by online nanoLC–MS/MS (NCS, and Q-Ex_HF, Thermo Fischer Scientific). Peptides were sampled on a 300 µm x 5 mm PepMap C18 precolumn and separated on a reprosyl 25 cm 1.9 µm (Cluzeau). The nanoLC method consisted in a 140-min gradient ranging from 4% to 40% acetronitrile in 0.1% formic acid in 123 min and wash to 90% and equilibration at 4% at a flow rate of 300 $nL.min^{-1}$. MS and MS/MS data were acquired using Xcalibur (Thermo Fischer Scientific). Spray voltage and heated capillary were set at 2 kV and 270°C, respectively. Survey full-scan MS spectra (m/z = 400–1600) were acquired in the Orbitrap with a resolution of 60,000 after accumulation of 1^{e6} ions (maximum filling time: 200 ms). The twenty most intense ions from the preview survey scan delivered by the Orbitrap were fragmented by collision induced dissociation (collision energy 30%) in the LTQ after accumulation of 1^{e5} ions (maximum filling time: 50 ms).

2.9 Database searches and results processing of MS data

Peaklists files were generated using Mascot Daemon. MS/MS spectra were searched using Mascot 2.6.0 (Matrix Science) against the target-decoy version of a compilation of the *A. thaliana* protein database (nuclear, mitochondrial and plastid genome; TAIR v10.0; December 14, 2010; 35,386

entries) and a home-made list of contaminants, frequently observed in proteomics analyses (249 entries). Trypsin/P was chosen as the enzyme and maximum of 2 missed cleavage allowed. Precursor and fragment mass error tolerances were set at 10 ppm and 0.025 Da, respectively. Peptide modifications allowed during the search were: carbamidomethyl (C, fixed) acetyl (Protein N-term, variable) and oxidation (M, variable).

The Proline software (http://proline.profiproteomics.fr/) was used to filter the results (filters at replicate level were conservation of only rank 1 peptides, peptide identification FDR < 1% as calculated on peptide scores by employing the reverse database strategy, minimum peptide score of 25, peptide length ≥ 7, and minimum of 1 specific peptide per identified protein group). Next identifications were merged over the whole experiment and the minimum of 1 specific peptide per protein group filter was applied again. Spectral count (SC), specific spectral count (SSC) and weighted spectral count (WSC, calculated as described in (22)) were computed within each replicate. The mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (23) with the dataset identifier PXD010545 and 10.6019/PXD010545. The identifications resulting from this validation pipeline are also available in the form of excel tables (supplemental table S1).

Further filters were applied manually: contaminants as well as protein groups identified by a single peptide and with a total sum of SWC of 1 were discarded. The WSC values were normalized by the total WSC within each replicate. The ratio of the sum of normalized WSC over triplicates (WSC_Env / WSC_CCE) was used as an estimator of the enrichment of each protein in the envelope fraction compared to the total extract. In cases where the sum of WSC_CCE equals zero (protein groups identified only in the envelope fractions), a count of 1 was added to both the sum of WSC-Env and the sum of WSC_CCE.

2.10 Construction of Vectors for Stable Expression in Arabidopsis

To construct the vectors for FtsY::GFP, UP1::GFP, VTE1::GFP and eIF-5A::GFP overexpression in Arabidopsis, the coding region of Arabidopsis correspondent protein was PCR-amplified using the two flanking primers *Sal*I-*N*-ter and *Nco*I-*C*-ter (see supplemental table S2). The amplified fragments were inserted by cloning a blunt-ended in an intermediate plasmid, the plasmid pBluescript (pKS) to be fully sequenced. The *Sal*I-*Nco*I fragment cleaved from each such plasmid was inserted into the *Sal*I-*Nco*I digested GFP reporter plasmid (puc19-35Ω-sGFP(S65T)) (24).

Digestion of this plasmid with the restriction enzymes *Hind*III and *EcoR*I allowed to isolate the insert which encodes both the interest protein and the fusion protein GFP in order to be introduced in the *Hind*III-*EcoR*I digested pEL103 binary vector (kan resistance to transform wild-type plants). SFR2::CFP and TSP9::CFP were constructed by recombination using the Gateway system and the 35S-CFP construction.

2.11 Arabidopsis transformation and confocal microscopy analyses

Wild-type Arabidopsis plants ecotype WS were transformed by dipping floral buds of 4-week-old plants into an *Agrobacterium tumefaciens* (C58 strain) solution containing a surfactant (Silwet L-77) according to Clough and Bent (25). Primary transformants were selected on Murashige and Skoog medium (Murashige and Skoog MS5519, 1% [w/v] sucrose, and 1% [w/v] agarose) containing 100 $mg.l^{-1}$ kanamycin. Only lines segregating 3:1 for the resistance to kanamycin and expressing the recombinant protein were selected for further analyses (proteins subcellular localization using confocal microscopy). Fluorescence microscopy was then performed with a confocal laser-scanning microscope (TCS-SP2, Leica, Deerfield, IL).

3. Results

Towards a deep-rooted and authentic investigation of the chloroplast envelope proteome: overview of the complete strategy

Chloroplast proteins are major components of plant cells representing 40% of the whole cell proteins. Such organelles are surrounded by two envelope membranes (the chloroplast envelope) that contain only 1-2% of the chloroplast proteins. Therefore, chloroplast envelope proteins represent less than 1/250 of the whole cell proteins (Fig. 1). When comparing the composition crude leaf extract and purified envelope vesicles, a high theoretical enrichment factor (EF) of specific envelope proteins is thus expected.

Sensitivity of MS technique has been greatly improved during the last decade. While only 100 proteins (including contaminants from other chloroplast sub-compartments) could be identified in purified chloroplast envelope fractions 15 years ago (26), 700 proteins were detected in 2010 (12). Today, thanks to the continuous improvement of MS techniques and instruments, we are able to detect more components in a complex sample, to generate quantitative data and to statistically

validate above-cited enrichment factors. The aim of the present study was thus to analyze, using a MS-based approach, crude cell extracts and envelope fractions in order to *i*) detect more components in chloroplast envelope samples, and *ii*) differentiate genuine (highly enriched) chloroplast envelope proteins from components deriving from other cell or plastid compartments by comparing quantitative data for each protein in both crude cell extracts and envelope fractions. For that, crude cell extracts and envelope fractions were prepared in independent triplicates. Proteins (15 µg) of each sample were separated by 12% SDS-PAGE gel with Coomassie blue staining in order to assess reproducibility of the preparation and homogeneity of the samples (see supplemental Fig. S1). As expected, similar protein profile was observed over triplicates. Well-known abundant markers associated to the chloroplast were revealed in the appropriate fractions. Indeed, the large subunit of Rubisco (RbcL) and the light harvesting complex protein (LhcP) were highly and proportionally detected in the tree total cell extracts (supplemental Fig. S1A, lanes CCE1-3). The phosphate/triose-phosphate transporter (TPT), the major protein of the envelope sub-chloroplast fraction, was largely enriched in each purified envelope fraction (supplemental Fig. S1B, lanes E1-3). These triplicate biological samples were then analyzed by MS, after concentration of the protein samples (10 µg) in the top of a NuPage 4-12% (supplemental Fig. S1C).

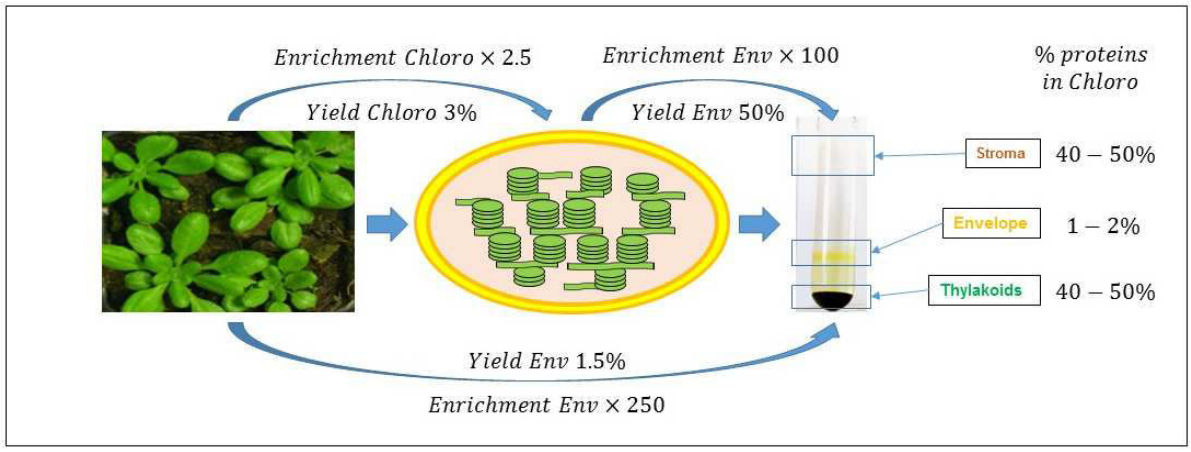


Fig. 1: Theoretical enrichment factor (EF) of envelope proteins, from crude cell extract to purified envelope fractions. Chloroplasts contain 40% of plant cell proteins. Thus, an enrichment factor of chloroplast proteins in Percoll-purified chloroplast fraction compared to total extract would be 2.5. Since chloroplast envelope represent the minor chloroplast component (*i.e.* only 1-2% of the chloroplast proteins), envelope proteins should around 50-100 times more abundant in purified envelope fraction when compared to whole cell extract. This enrichment factor (EF) could be used to discriminate genuine envelope proteins from contaminants.

Crude data analysis: distribution and predicted subcellular classification

Almost 2500 non-redundant proteins were identified from the all analyses performed (supplemental table S3). These 2480 Arabidopsis proteins result from the cross of 2222 proteins identified in the crude cell extracts (minimum of 1 specific peptide in the CCE triplicate) and 1269 proteins identified in envelope fractions (minimum of 1 specific peptide in the Env triplicate). In the present work, proteins present in purified chloroplast envelope fractions were thus detected at an unprecedented high level of sensitivity, when compared to the protein number (ca 700 proteins) previously identified in comparable envelope fractions, using the same numbers of biological samples (12).

As a preliminary analysis, subcellular localizations of the whole detected proteins from either crude cell extract or purified envelope fraction were predicted using information (consensus localization) available in the SUBA3 database (see supplemental table S3). According to this analysis, 1017 proteins (41% of the 2479 detected proteins) were assigned to the plastid. More than half (618 proteins, 61%) of these plastid proteins were shared by envelope and crude extract lists (Fig. 2). These shared plastid proteins might represent major envelope proteins or abundant plastid proteins (stroma/thylakoids) cross-contaminating purified envelope fractions. Part (169 proteins, 17%) of these plastid proteins were only detected in envelope fractions: these proteins are expected to represent minor envelope proteins that are too poorly abundant to be detected in crude cell extracts. Altogether, 77% (787 proteins) of the plastid proteins were identified in the envelope fraction. On the other hand, 23% (230 proteins) of the plastid proteins were only detected in crude cell extracts. These last proteins are anticipated to be stroma or thylakoid proteins that are abundant enough to be detected in crude cell extracts, but too poorly abundant to be detected in the list of stroma or thylakoid proteins contaminating purified envelope fractions.

Beyond plastid proteins, many identified proteins were predicted as components deriving from other cell compartments (cytosol, mitochondria, nucleus, plasma membrane…). This MS/MS analysis identified 593 proteins from the cytosol, 292 from the secretory system, 204 from the mitochondria, 104 from the nucleus, 117 from the plasma membrane, 75 from the peroxisome and 77 from the vacuole (see supplemental table S4). Interestingly, out of these 593 cytosolic proteins, 74% of them (436 proteins, see supplemental table S4), were only detected in the crude cell extract, while only 4% (22 proteins) were only detected in the envelope fractions. The same is true for other cell compartments, since most ExtraC/ER/golgi (64%), mitochondria (72%), nucleus (83%),

plasma membrane (52%) and peroxisomal (56%) proteins were only detected in crude cell extracts (see supplemental table S4). Surprisingly, only vacuolar proteins show a more balanced repartition between envelope fractions and crude cell extracts, 31% of them being only detected in crude cell extracts and 17% of them being only detected in envelope fractions, thus suggesting that cross-contamination of the envelope preparations with vacuolar proteins is more pronounced (see supplemental table S4 and Fig. 2).

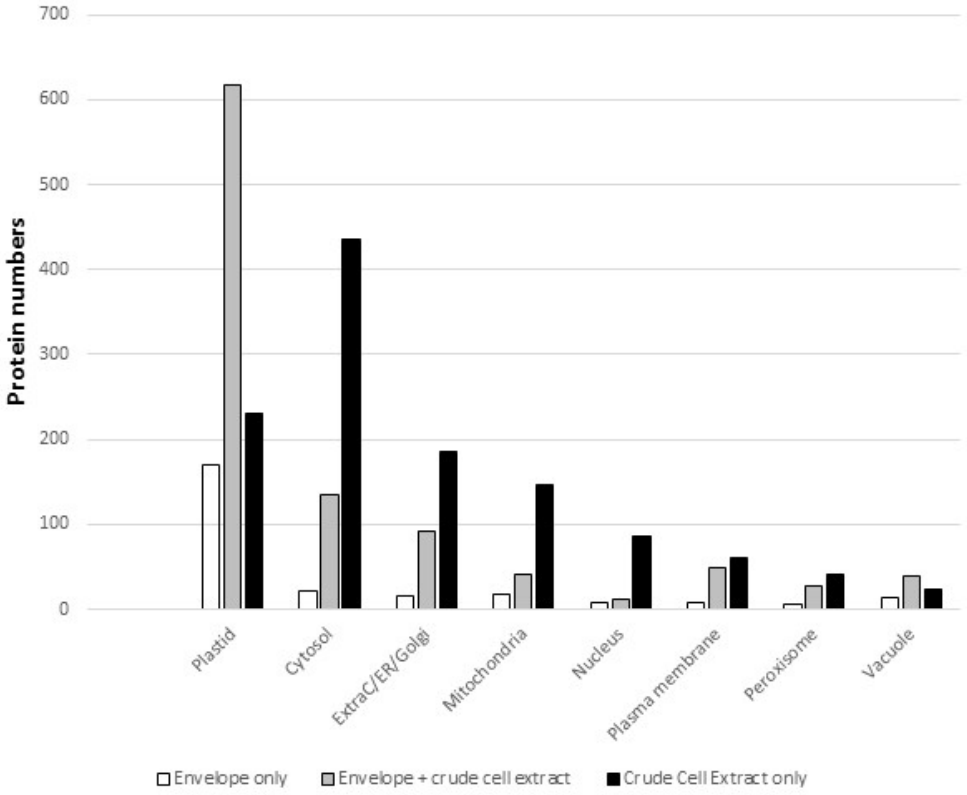


Fig. 2: Predicted subcellular localization of proteins identified in purified envelope fractions and crude cell extracts according to the SUBA3 database (SUBAcon, see Hooper et al., 2014). Note that only 1 (AT5G12170) of the almost 2500 proteins detected during this work were not present in the SUBA3 database. Proteins identified in envelope fractions are enriched in plastid predicted proteins while crude cell extracts (CCE) contain more proteins predicted to be localized in other cell compartments. Envelope only (white): proteins only detected in the envelope fraction. Envelope + CCE (gray): proteins detected in both envelope and crude cell extract. CCE only (black): proteins detected in crude cell extract and absent from the list of envelope proteins.

Annotation of the 1269 proteins detected in purified envelope fractions

Since our objective was to concentrate on the chloroplast envelope proteome, we then targeted our analysis to the 1269 proteins detected, at least once, in purified envelope fractions (see supplemental table S5,"This work").

To explore these data, manual annotation of either subcellular and subplastidial localization and accurate description or function was performed (supplemental Fig. S2). To do so, information about protein description, function and subcellular localization were sequentially and manually searched for each protein in several public databases (TAIR, Uniprot, NCBI, MAPMAN…) and the appropriate literature (PubMed, searching for either AGI numbers or protein names…). Then, the 1269 proteins were tested for the presence of a predictable chloroplast transit peptide using ChloroP (27) and TargetP (28) prediction tools. Furthermore, as shown in supplemental Fig. S2, two reliable and effective databases termed "SUBA3" (29) and "MASCP Gator" (30) were used to determine if some proteins were known components of other cell compartments (mitochondria, plasma membrane, cytosol…). Indeed, both "SUBA3" and "MASCP Gator" provide substantial information by indicating whether each protein has already been previously detected by MS in specific cell compartments, if its subcellular localization was determined using *in planta* expression of GFP-tagged forms and if a consensus subcellular localization can be deduced from the use of all subcellular prediction tools available. Finally, present data were also overlapped with the list of proteins previously identified in the envelope fractions, and their subplastidial localization, present in the AT_CHLORO database (12) (see supplemental table S6).

In this context, we first considered previous MS-based detection of each protein in specific cell compartments (using MASCP-Gator and SUBA3) as a key criterion (supplemental Fig. S2). For example, when a protein was only and repeatedly detected in the plastid and never or rarely in another cell compartment, association of this protein to the plastid compartment was considered as highly plausible (even if this protein was lacking a classical and predictable chloroplast transit peptide). On the contrary, when a protein was predicted to contain a chloroplast transit peptide but repeatedly detected in mitochondria or any other cell compartment, it was excluded from the list of plastid proteins.

Validation of the EF value to predict the subcellular and subplastidial localization of the 1269 proteins detected in purified envelope fractions

To further support the known or predicted localization of the 1269 identified proteins, spectral count (SC), specific spectral count (SSC) and weighted spectral count (WSC) were computed for each detected protein within each replicate. For each sample independently (crude extract or envelope), the WSC were normalized by the total of SWC of each replicate related to the sum of

one replicate (supplemental table S2). A ratio (EF = Enrichment Factor) of the sum of normalized WSC over each triplicate (EF = norm_WSC_Env / norm_WSC_CCE) was then calculated that was tentatively used as an estimator of the enrichment of each protein in the envelope fraction compared to the crude cell extract. To calculate this ratio for proteins that were only detected in envelope fractions (*i.e.* norm_WSC_CCE = 0), plus one was added to both norm_WSC_Env and norm_WSC_CCE (see supplemental table S5, "This work", column EF estimator).

In order to test the reliability of this EF parameter, we first created a "safe set" of proteins. These proteins were selected as follows: *i*) their subcellular localization was experimentally determined *in planta* using GFP-fusion experiments and confocal microscopy. Alternatively, 42 chloroplast- and 2 mitochondria-encoded proteins were considered as genuine proteins from these respective organelles, *ii*) SUBA.con, *i.e*. the consensus of all localization prediction tools present in the SUBA3 database, supports the data issued from the *in planta* experiments, *iii*) the subplastidial localization of the proteins was previously (12) and unambiguously determined (thus excluding proteins only detected during the present work). Combining the above-cited parameters, only 175 proteins could be selected, representing 14% of the 1269 proteins detected in the purified envelope fractions. Interestingly, this set of 175 proteins contains protein groups from the all cell and chloroplast compartments (see supplemental Fig. S3). In good agreement with their expected enrichment in purified envelope fractions when compared to crude cell extracts (see Fig. 1), envelope proteins (inner or outer membrane) proteins from the "safe set" were highly enriched (EF > 10) in purified envelope fractions (see supplemental Fig. S3). This enrichment was less pronounced for envelope proteins previously shown to be shared between envelope and other plastid compartments (stroma and thylakoids). Interestingly, if one excludes proteins from the vacuole, all other cell- and plastid-associated components showed an average EF value far below (< 1) the one of genuine envelope proteins.

Having confirmed the good correlation between genuine envelope localization and effective enrichment (high EF values), we then used (when required) this parameter to help predicting the subcellular and subplastidial localization of the "Negative of the safe set", *i.e.* the 1094 proteins whose localization could not be deduced from previous analyses (see bottom of supplemental Fig. S2). We especially considered this EF value as a crucial parameter, determinant to classify proteins as genuine envelope proteins (*i.e.* EF value above 5, highly enriched in the purified envelope fraction when compared to crude cell extract). When the EF value was lower (*i.e.* $1 < EF < 5$) for

some envelope proteins, data from AT_CHLORO were essential to discriminate proteins shared by envelope and other plastidial compartments (stroma or thylakoids) from proteins most probably contaminating (very low EF value, *i.e.* EF < 1) purified envelope fractions. Finally, we combined rare data from the literature and useful data recently published by Schleiff and colleagues (14) to classify the proteins as either inner or outer envelope proteins (see supplemental table S5, "This work", column "Simm *et al*., 2014").

Combining all information present in existing databases and the literature, and following the strategy depicted in supplemental Fig. S2, we were able to classify (see supplemental table S5, "This work", columns "Curated function" and "Curated description") all identified proteins in various categories ranging from their subcellular and subplastidial localization, specific activity or main function. As depicted in Fig. 3A, out of the 1269 proteins, detected at least once in purified envelope fractions, 462 proteins (36%) could be classified as envelope proteins (see Fig. 3A, Env-All), including 383 (30%) as inner envelope proteins (see Fig. 3A, IEM), 52 (4%) as outer envelope proteins (see Fig. 3A, OEM) and 27 (2%) as envelope candidates (see Fig. 3A, Env?), these last proteins being mostly uncharacterized proteins, never or rarely detected in any other cell compartments, or enriched in the envelope and containing a predicted cTP…). Out of these 462 above-cited envelope proteins (Fig. 3A, Env-All), 254 inner and 41 outer envelope proteins (see Fig. 3A, IEM and OEM only) were highly enriched (or were only detected) in the purified envelope fractions (see Fig. 3B), while 130 others, being envelope components most-probably shared with other compartments (see Fig. 3A, IEM-stroma, IEM-thylakoids, IEM-other, OEM-other) were less enriched (see Fig. 3B).

However, due to highly improved detection sensitivity of current spectrometers, not only several minor envelope proteins were revealed but also many other contaminants from other chloroplast and cell compartments. As expected from previous studies, main contaminants (Fig. 3A) derive from the stroma (237 proteins, *i.e.* 191 + 38 + 1 + 7) and the thylakoids (170 proteins, *i.e.* 143 + 17 + 1 + 9). Being well-known or predicted plastid components, these proteins were however easily excluded from the envelope list due to their very low EF value (EF < 0.5; see supplemental table S5, "This work", and Fig. 3B).

All other identified proteins are well-known (abundant markers) or expected contaminants deriving from other cell compartments. Many (10%, 125 proteins) derive from the cytosol most of them being linked to cytosolic translation of proteins (ribosomal proteins) with an average EF value of

0.3. An equivalent number (9%, 111 proteins) are either ER/golgi proteins or additional cytosolic proteins linked to intracellular trafficking functions (average EF value of 1.5), 45 proteins (4%) are vacuolar components with a surprisingly non negligible EF value of 3.3 (see Fig. 3B), 36 (3%) were assigned to the plasma membrane (average EF value of 0.8), 32 (3%) to the mitochondria (average EF value of 0.7), 6 (<1%) to the nucleus (average EF value of 1.0) and 7 (< 1%) to extracellular proteins (average EF value of 0.6). Finally, it was not possible to identify or predict a subcellular localization for 23 proteins (2%). According to this classification, the average EF for the 462 envelope proteins (see Fig. 3B, Env-All) was 13 while most other categories were given a value around 10 times lower (see Fig. 3B). The only exceptions were the 27 envelope candidates (see Fig. 3B, Env?) with an average EF value of 3 and the 45 vacuolar components (see Fig. 3B, vacuole) with the same average EF value of 3.

Altogether, following the strategy depicted in supplemental Fig. S2, a total of 869 (462 env + 237 stroma + 170 thylakoids) proteins (69%) were thus classified as well-known or good candidates for a chloroplast localization, only 53% of these chloroplast proteins being predicted envelope components.

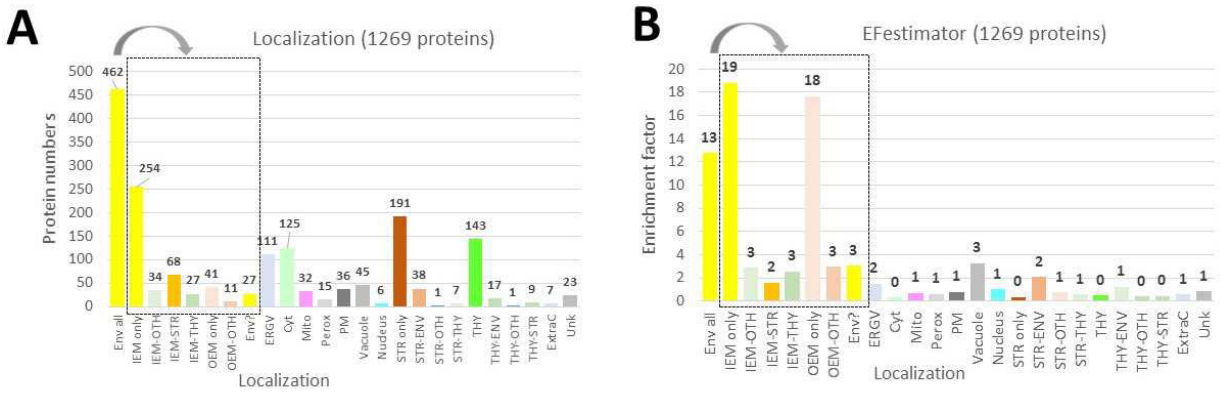


Fig. 3: Subplastidial and subcellular localization of the 1269 proteins detected in purified envelope fractions as deduced from manual annotation (see supplemental table S5, "This work"). The black square contains subcategories of envelope proteins (Env All): proteins only detected in envelope (IEM or OEM only), proteins shared with other undefined plastid compartments (IEM- or OEM-OTH), stroma (IEM-STR) or thylakoid (IEM-THY) or envelope candidates (Env?). **A.** As deduced from information present in the literature, databases and prediction tools (see Fig. 3), only 462 (383 Env + 52 OEM + 27 Env?) of the 1269 proteins are predicted to be associated to envelope membranes. **B.** As expected (see Fig. 1 and supplemental Fig. S3, "safe set"), the average enrichment factor of predicted envelope proteins is far above the ones of proteins associated to other plastid or cell compartments. Note the relatively lower EF values of proteins shared between envelope and other plastid compartments, and the surprisingly high EF value of vacuolar proteins. IEM; inner envelope membrane, OEM; outer envelope membrane, Env?; Envelope candidates; ERGV: endoplasmic reticulum/Golgi, Cyto; cytosol, Mito; mitochondria, Perox; peroxisome, PM; plasma membrane, Vacuole; vacuole, Nucleus; nucleus, Str; stroma, thy; thylakoid, Ext; extracellular localization, Unk; unknown and unpredictable localization, OTH; other undefined localization.

Comparison to recent studies targeting the chloroplast envelope proteome

Previously undetected proteins in any MS/MS-based study: We first compared proteins detected in purified envelope fractions during this work with all previous proteomics studies referenced in MASCP-Gator and SUBA3. According to these databases, 70 proteins (of the 1269 proteins identified in envelope fractions) were detected for the first time using a MS-based approach by the present study (see supplemental table S5, "This work", proteins classified as "nd" in column "Experimental evidence"). In good agreement with the targeted subcellular fraction and with a better sensitivity of the present MS-based study, most (53) of these previously undetected 70 proteins were known or predicted envelope components (29 IEM, 9 OEM, 15 Env?), while less (17) were contaminants deriving either from ER or Golgi (5), thylakoids (4), cytosol (3), stroma (2), mitochondria (2), plasma membrane (1) and none from the peroxisome, the nucleus, the vacuole or the extracellular compartment. In other words, since most other cell compartments were deeply analyzed by previous MS-based targeted studies, a large part of proteins present in other cell compartments, and detected during this work, were previously detected using targeted MS-based studies. One might even be skeptic about the classification (subcellular localization) of the very few (17) above-cited non-envelope proteins knowing that the probability to detect them in a purified envelope fraction is expected to be lower than the probability to detect them in the cell compartments were they reside.

Previously undetected proteins in MS/MS-based study targeting envelope fractions: As one of the main goal of the present study was get a better overview of the chloroplast envelope composition, present data were also overlapped with our previous data stored in AT_CHLORO (12). Out of the 1269 proteins identified in the purified envelope fractions during the present study, 433 proteins were not previously detected in any chloroplast sub-compartment (envelope, stroma, thylakoid) during our previous work (12) (see supplemental table S5, "this work", column "Env/Str/Thy, AT_CHLORO"). Out of these 433 newly detected proteins, 87 could be assign an envelope localization (including 24 envelope candidates, *i.e.* "Env?"), 98 were classified ER/Golgi components, 90 cytosol, 29 mitochondria, 22 thylakoid, 29 plasma membrane, 29 vacuole, 12 stroma, 5 extracellular compartment, 5 nucleus, 5 peroxisomal, and for 22 of them, a subcellular localization could not be proposed (see supplemental table S5, "This work", Column Simplified location this work, category "unknown"). Again, as suggested above using prediction tools (see

Fig. 2), taking advantage of better MS detection, not only several new minor envelope proteins were detected, but also many other contaminants deriving from other chloroplast and cell compartments. Indeed, as shown in Fig. 4B, the present study provided a better overview of the chloroplast envelope composition (92 proteins) when compared to our previous work (12).

On the other hand, 91 proteins, present in AT_CHLORO (12), were not detected in the purified envelope fractions during the present work (see supplemental table S6) and Fig. 4C. Out of these 91 proteins, only 16 proteins were associated to the chloroplast envelope while most (75) other proteins were contaminants from other plastid or cell compartments. It is also important to note here that part (38 proteins) of these 91 proteins previously detected by Ferro *et al*. (12) were also detected during this work, but only in crude cell extracts (see supplemental table S6), in good agreement with their known or predicted subcellular localization (13 thylakoids, 12 cytosol, 6 stroma, 3 mitochondria, 2 peroxisome, 1 plasma membrane, 1 envelope candidate).

We next compared our results to a more recent study by Simm and colleagues (14) who aimed to identify the core proteome of the chloroplast envelope through analysis of the envelope proteome from three plant species. As shown in Fig. 4B and supplemental table S5, "This work", 98 (*i.e.* 90 + 8 in Fig. 4C) proteins were classified as envelope components in both studies and only 90 proteins were detected in all three studies. Many proteins (116, *i.e.* 82 + 34 in Fig. 4C) identified during the study of Simm and colleagues (14) were detected during the present work, but assigned other subcellular or subplastidial localizations (82 for thylakoids, 14 for stroma and few components from the mitochondria, peroxisome, ER/Golgi, plasma membrane and cytosol). Finally, 57 proteins (*i.e*. 54 + 3 in Fig. 4C) detected during the study of Simm and colleagues [14] were not detected here, but only one protein was classified as an envelope component (14) while most others were classified as "unknown", "thylakoid", "plastid" or "stroma".

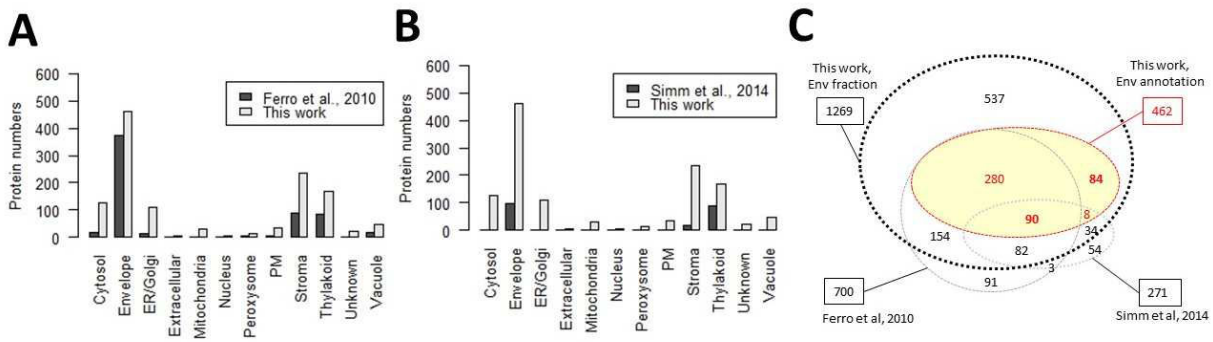


Fig. 4: **Evaluation of coverage of the chloroplast envelope proteome when comparing present data with earlier analyses targeted to the same membrane system**. This work versus (A) Ferro et *al*. [12] or (B) Simm *et al*. [14].

Note that the better MS sensitivity allows detection of additional (probably minor) envelope proteins when compared to previous analyses. On the other hand, this is also the case for non-plastid compartments, cytosol and ER/Golgi proteins being specifically detected during this work. PM; plasma membrane. **C. Evaluation of coverage of the chloroplast envelope proteome when combining present data with earlier analyses targeted to the same membrane system.** Venn diagram indicating the weight of protein identified during this work when compared with previous data obtained by Ferro *et al*. (12) (see sup data 10 in (12)) or Simm *et al*. (14) (see sup data S8 in (14)). Indicated values are numbers of proteins identified throughout the diverse studies. For the present study, were considered both the all 1269 proteins identified in the purified envelope fractions and the 462 proteins classified as envelope components (thus excluding proteins classified as envelope "contaminants" and suspected to derive from non-plastid cell compartments (see supplemental table S5, "This work"). When combining all three studies, 84 envelope proteins were only identified during this work, 370 envelope proteins are shared between the present work and data obtained by Ferro *et al*. (12), but only 90 proteins were detected in all three studies. Note that most of the unique proteins identified by Ferro *et al*. (12) (*i.e.* 91 proteins) or Simm *et al*. (14) (*i.e.* 54 proteins) were detected during the present work in the crude cell extracts, and thus assigned other subcellular or subplastidial localizations.

To conclude, when considering data in Fig. 4, it appears that the present study provides a better overview of the chloroplast envelope composition, when compared to earlier, but relatively recent studies (12, 14). To cite a few examples (see Table 1), besides components of the protein-conducting channels, a novel isoform of the solute conducting channel OEP24 (31), termed OEP24A (AT1G45170) was identified during this work as an outer envelope pore protein (see supplemental table S5, "This work", and Table 1). This protein, previously detected by Simm *et al.* (14), was not present in the AT_CHLORO database. Two components of the TOC translocon, annotated TOC90 (AT5G20300) (32) and TOC120 (AT3G16620) (33) were detected during the present work while both proteins were not detected during our previous study (12) and only TOC120 was previously detected during the study performed by Simm *et al*. (14) (see supplemental table S5, "This work"). Additionally, the present work (see Table 1) identified TIC20-I (AT1G04940) and TIC20-V (AT5G55710) that are involved in protein precursor import into chloroplasts at the inner envelope membrane (34, 35). TIC20-I, which was recently described to be specific for photosynthesis-related pre-proteins, and partially redundant with TIC20-IV but not with TIC20-II or TIC20-V (34, 35), was not identified by both earlier proteomic studies (12, 14) but detected during the present work (see Table 1).

Table 1: Detection of representative inner and outer envelope membrane proteins. Protein accessions are provided (AGI numbers) together with short names of proteins (Name). Detection (X) of these proteins in the present work, in the work published by Simm *et al*. (14) or Ferro *et al*. (12) are indicated. The enrichment factor (EF) of each protein in the envelope fraction compared to the crude cell extract is indicated in the EF estimator column. na, means that the protein was detected in the whole chloroplast fraction but not in a the purified envelope fractions. Bold accessions are known envelope proteins that were not detected in previous studies targeting the envelope proteome.

AGI	Name	This work (2018)	Core proteome (2014)	AT_CHLORO (2010)	EF estimator (this work)
AT1G04940	**TIC20-I**	X	-	-	10.8
AT2G47840	**TIC20-II**	X	-	-	7
AT5G55710	TIC20-V	X	-	na	0.7
AT3G23710	TIC22-III	X	-	X	10.9
AT4G33350	TIC22-IV	X	X	X	12
AT4G23430	TIC32-IVa	X	-	X	10
AT5G16620	TIC40	X	X	X	17.4
AT2G24820	TIC55-II	X	X	X	8.8
AT4G25650	TIC55-IV	X	X	X	94.3
ATCG01130	TIC214	X	X	X	41.8
AT5G22640	TIC100	X	X	X	90.2
AT1G06950	TIC110	X	X	X	10
AT5G20300	**TOC90**	X	-	-	8.9
AT3G16620	TOC120	X	X	-	5.2
AT1G21650	**SECA2**	X	-	-	27.8
AT2G31530	SCY2	X	-	X	7.9
AT1G45170	OEP24A	X	X	-	1.9
AT3G52230	OEP24	X	X	X	4.1
AT5G42960	OEP24-II	X	-	X	1.9

Functional classification of the 1269 proteins identified in purified envelope fractions

The in depth manual analysis (literature, databases, predictions…) also allowed us to assign functional categories to 1269 proteins identified in purified envelope fractions, including 462 chloroplast envelope proteins (see supplemental table S5, "This work", columns "Curated function" and "Curated description"). This allowed to get an overview of the functional profile of the newly detected proteins, and to actualize functional description of previously identified envelope proteins (12) using post-2010 data from the literature. As shown in Fig. 5, 100 envelope proteins were classified as metabolism actors (mostly involved in metabolic pathways linked to synthesis of lipids, hormones, vitamins and pigments) or as transporters (almost 90 members including inner and outer envelope membranes). While the function of most envelope proteins could be deduced from sequence similarities but still await functional characterization, the function

of around 16% of envelope proteins (more than 70 proteins) remain totally unknown and unpredictable (Fig. 5). Remarkably, "chaperone and protease", "translation stroma" and "protein targeting" were also among the main functional categories for the envelope proteins (Fig. 5). Indeed, around 30% of total genuine envelope proteins were assessed to such three functional groups. Finally, many envelope proteins were classified as "redox", "stress", "DNA/RNA interacting proteins", "signaling" or "others", knowing again, that while classified according to similarities with known proteins, or predicted functional domains, most of these proteins remain uncharacterized envelope proteins.

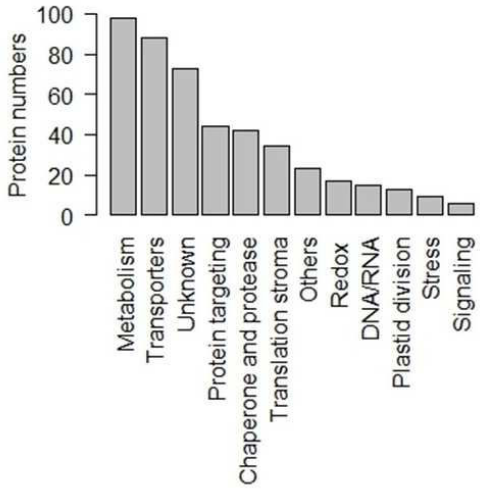


Fig. 5: Functional categories of the chloroplast envelope proteome as deduced from manual annotation (see supplemental table S5, "this work", column "Curated function"). Functional annotation of the 462 envelope proteins was collected from the appropriate literature (PubMed) and databases (Uniprot, TAIR, MapManBins…). Note that while 20% of envelope proteins were classified as "unknown", the function of the vast majority of proteins assigned to other functional groups are deduced from sequence similarity or predictions and remain to be demonstrated.

Variation of the enrichment factor (EF) allows distinguishing genuine envelope proteins from components shared with other chloroplast or cell compartments

The average EF of the 462 envelope proteins is clearly above the one of all non-envelope proteins detected in the purified envelope fractions (Fig. 6A). This is also the case when comparing envelope proteins with proteins associated to any other plastid or cell compartments (if one excludes vacuolar proteins, or stroma and thylakoid proteins shared with the envelope, see Fig. 6B). From this Fig. 6B, it is also noticeable that the average EF of genuine stroma of thylakoid proteins is far below the one of stroma and thylakoid proteins shared with the envelope.

However, we observed that, while assigned to a unique subcellular or subplastidial localization, the EF values were highly variable for proteins of the same group (see Fig. 6B). This observation could result from quantification artefacts resulting themselves from proteins situated at the threshold level of MS detection (*i.e.* where very low values of spectral counts have a strong impact on EF values, see supplemental table S5, "This work", column "EFestimator").

We thus analyzed these data in more details, not only considering envelope localization, but also functional categories (Fig. 7A) at the scale of the whole envelope. Again, while the median EF was in good agreement with enrichment of these functional groups in the chloroplast envelope ("transporters" being at the top level while "translation stroma" is at the lowest), the EF values were highly variable (*i.e.* high degrees of dispersion) for proteins of the same group. On reason for that could derive from the various subplastidial locations of the proteins from the same groups (the genuine envelope proteins having a higher EF value than the proteins shared with stroma or thylakoid sub-compartments).

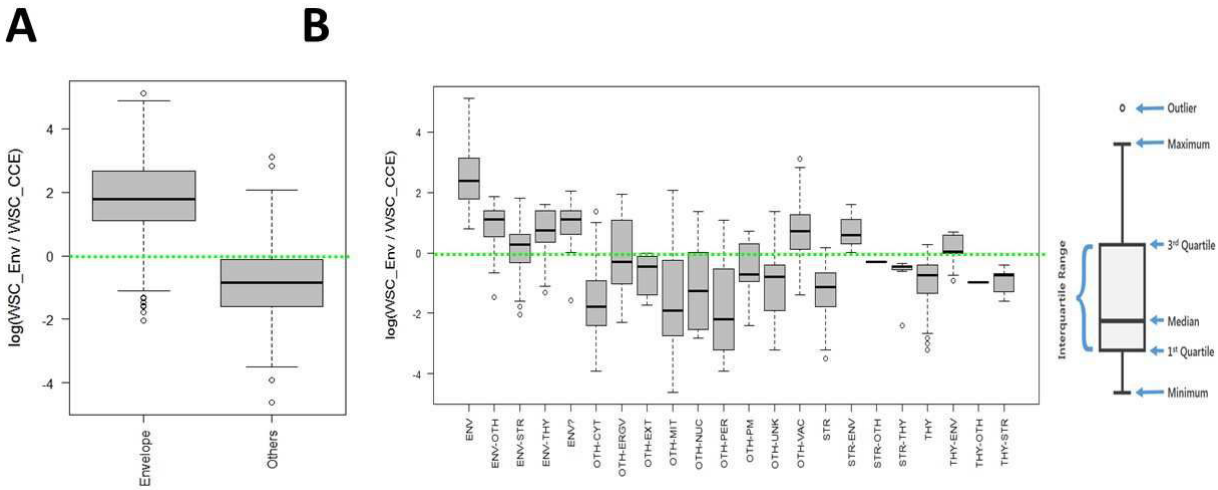


Fig. 6: Descriptive statistics depicting groups of EF values for cell proteins identified in purified envelope fractions (see supplemental table S5, "this work", columns "EFestimator" and "Simplified location (this work)"). **A.** Box plots displaying EF variations in the chloroplast envelope proteome (average EF = 30) when compared to other proteins (average EF = 0.1). **B.** Most non-envelope cell components have average lower EF. Surprisingly, vacuolar components are slightly enriched in the purified envelope fraction with an average of 2. Env; All envelope proteins (including outer envelope membrane components), Env-Oth; proteins shared with other undefined plastid compartments, Env-Str: envelope proteins shared with the stroma, Env-Thy; envelope proteins shared with the thylakoid, Env?; Envelope candidates; Oth-Cyt; cytosol, Oth-ERGV: endoplasmic reticulum/Golgi, Oth-Ext; extracellular localization, Oth-Mit; mitochondria, Oth-Nuc; nucleus, Oth-Per; peroxisome, Oth-PM; plasma membrane, Oth-Vac; vacuole, Str; stroma, Thy; thylakoid.

We then concentrated on specific proteins known to be part of the same family or the same multi protein complexes within one of these functional groups. As an example, starting from the 40 members (see supplemental table S5, "This work") of the group "chaperone and proteases" (Fig. 7A) we calculated the median EF value (and its variation) of the 9 FtsH and 14 ClpP proteins (Fig. 7B). In good agreement with previous observations, and according to AT_CHLORO (12), the FTSH (36) complex is exclusively localized in the envelope. As a support, this protein complex appear to be highly enriched in the envelope fraction when compared to crude cell extract (see supplemental table S5, "This work", column "EF estimator"). The median EF value of the FTSH complex is even stronger than the one of the 88 transporters known to be associated to the

chloroplast envelope (see Fig. 7B). On the other hand, it was suggested in AT_CHLORO (12) that the subunits of the Clp protease complex (37, 38) are most probably shared between stroma and envelope compartments (see supplemental table S5 "This work", columns "ENV %" and "STR %"). Interestingly, the EF values of most of these subunits are very low when compared to the ones of FtsH proteins. With the exception of ClpC2 (AT3G48870) which is highly enriched in the envelope fraction (EF > 50, see supplemental table S5, "This work", column "EF estimator", and the top outlier/red square in Fig. 7B), all other Clp proteins share the same intermediate EF value (average = 2) with a reduced degree of dispersion (EF between 1.3 and 3.5), *i.e.* just above a threshold level corresponding to the 34 "ribosomal proteins" from the stroma (see Fig. 7B). To cite other examples, in the functional category "lipid metabolism", two members of the Acetyl-CoA carboxylase complex (ACCDa, AT2G38040 and ACCDb, ATCG00500) were mostly detected within envelope membranes while two others (BCCP1, AT5G16390 and BCCP2, AT5G15530) were previously shown to be shared between envelope and stroma (12). In good agreement with these previous observations, the EF values of ACCDa and ACCDb are higher (24.3 and 13.6, respectively), when compared to the ones of BCCP1 and BCCP2 (3.5 and 0, respectively, BCCP2 being even not detected during the present work, see supplemental table S5, "This work", lanes 73 to 75). Again, in the functional category "lipid metabolism", the six members of the pyruvate dehydrogenase multi-enzyme complex, were previously shown (12) to be shared by envelope and stroma (see supplemental table S5, "This work", lanes 97 to 102). In good agreement, out of the six subunits of this complex, five share the same intermediate EF (average of 4), and only one member (ptLPD1, AT3G16950), only detected in purified envelope fractions, has a higher EF of 24 (see supplemental table S5, "This work", lane 118).

In other words, even at the envelope scale, a good correlation is observed between high EF values and unique envelope localization or intermediate EF values and previously observed sharing of the proteins between envelope and other plastid compartments.

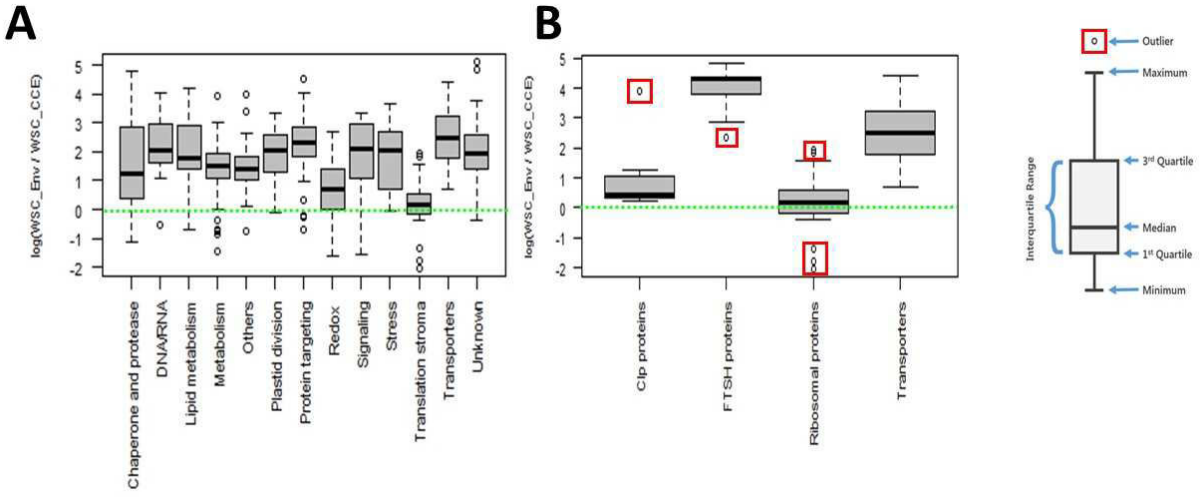


Fig. 7: Descriptive statistics depicting groups of EF values for functional categories at the chloroplast envelope scale. A. Box plots displaying EF variations in functional categories (see Fig. 5) of the chloroplast envelope proteome. **B**. Note that when targeting protein complexes (*e.g.* 14 Clp subunits 9 FtsH subunits) within one functional category (*e.g.* here, "Chaperone and protease"), both the interquartile ranges and the degree of dispersion are reduced (see A).

Outliers (here, ClpC2 and Ftsh7) are plotted (red squares) as individual points easily identified in supplemental table S5, "this work", columns "EFestimator" and "Curated function (this work)"). Ribosomal proteins = 34 proteins, transporters = 88 proteins.

Defining EF threshold levels to exclude putative plastid and cell contaminants

From the above-cited observation, the EF value appears fairly reliable. It was thus tempting to define a tentative threshold level for EF that would allow discriminating genuine envelope proteins from contaminants from other plastid or cell compartments. As discussed above, most envelope proteins are characterized by a high enrichment factor with an average value of 13 (Fig. 3B). Most other plastid or cell components have an average lower EF value of 1 except for vacuolar proteins that were revealed to be slightly enriched in the envelope fraction, having a higher (average of 3.3) EF value than other cell components (Fig. 3B). Remarkably, when increasing the EF threshold to 1 or 2 (see supplemental table S5, "Only EF>1" and "Only EF>2"), most (85% for EF>1 or 95% for EF>2) of the 244 proteins from the cytosol, mitochondria, peroxisome, plasma membrane, nucleus, extracellular and unknown proteins are excluded (see Fig. 8B and 8C, light gray square). This is also true for most (84% for EF>1 or 95% for EF>2) of the 406 plastid proteins localized in stroma and thylakoid that are excluded using a threshold of 1 or 2 when compared to whole data (see Fig. 8B and 8C, right black square). However, these tentative threshold levels, which allow eliminating most envelope contaminants, reduce, to a lesser extend, the number of the 129 envelope protein (33% for EF>1 or 57% for EF>2) shared with other plastid compartments (see Fig. 8B and 8C, left black square). This observation thus raises the question of the genuine envelope localization of these proteins shared by envelope and other plastid compartments, and of the relevance of these intermediate EF values.

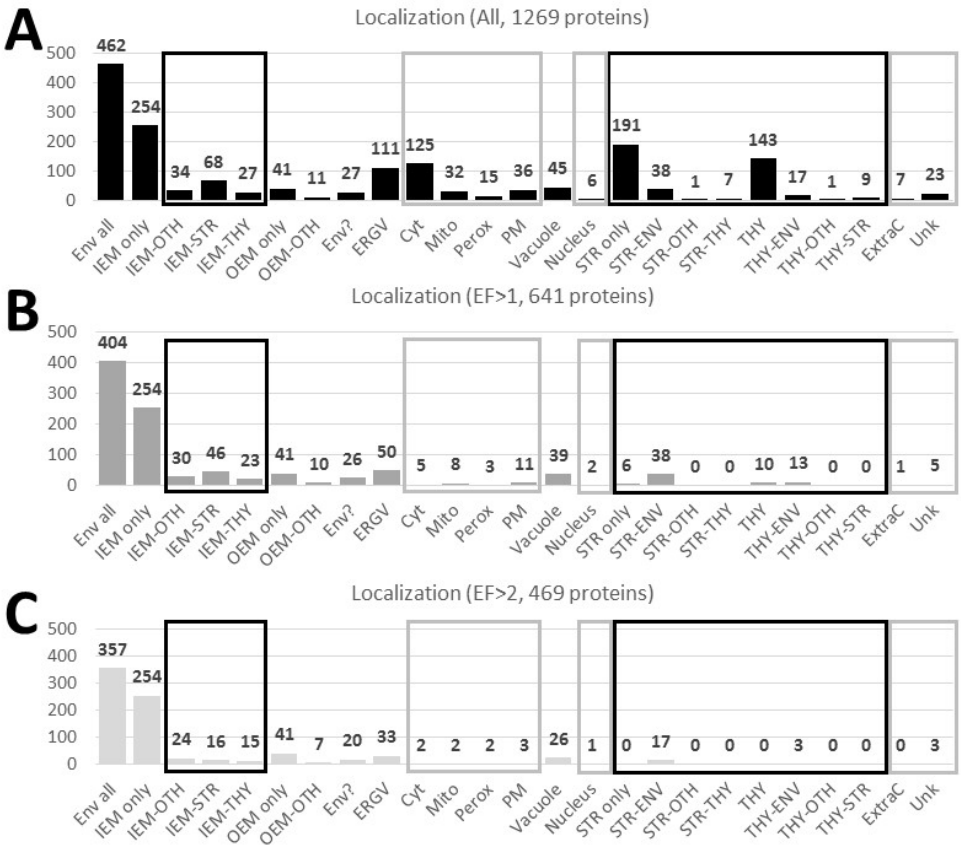


Fig. 8: Increasing the EF threshold level to 1 or even 2 reduces envelope protein shared with other compartments and excludes most other plastid and cell components (see supplemental table S5, **A.** "This work" and **B.** "Only EF>1" or **C.** "Only EF>2"). Note that most (84% for EF>1 or 95% for EF>2) of the 406 plastid proteins (right black squares) localized in stroma and thylakoid are excluded using a threshold of 1 or 2 when compared to whole data. This is also true for most (85% for EF>1 or 95% for EF>2) of the 244 proteins (gray squares) from the cytosol, mitochondria, peroxisome, plasma membrane, nucleus, extracellular and unknown proteins. However, note that even an EF threshold of 2 does not remove non-negligible parts of ER/Golgi (30%) and vacuolar (58%) components.

Intermediate EF values are compatible with genuine envelope localization

To answer this question, we chose to analyze the subcellular and subplastidial localization of several proteins.

The protein FtsY (AT2G45770) (39), was previously shown by Ferro et *al.* (12) to be distributed (34% in envelope, 32% in stroma and 33% in thylakoid) between the three chloroplast sub-compartments (see supplemental table S5, "This work", lane 146). In good agreement with these previous observations, this protein has an intermediate EF value of 2.7 quite compatible with its shared localization. A western blot was performed on Arabidopsis crude cell extract and the three purified main plastid sub-compartments, using a polyclonal antibody raised against FtsY (Fig. 9). As expected, FtsY was detected in all three chloroplast compartments, *i.e.* envelope, stroma and thylakoid, and its signal in the purified envelope was enriched when compared to the CCE. As internal controls, HMA1 (AT4G37270) used as an envelope marker (20), was exclusively localized

in purified envelope fractions (EF = 34.7), the ketol-acid reductoisomerase (KARI, AT3G58610, EF = 0.6) was associated to the stroma, and the light harvesting complex family of proteins (LHCP) (average EF of LHCPs = 0.5) the well-known abundant thylakoid markers, were only associated to the thylakoid (Fig. 9). Altogether, these observations confirm the good correlation between respective EF values and subplastidial localization of these proteins. In other words, while the EF value (2.7) of FtsY is far below the average EF value of envelope proteins (*i.e.* above 13, see Fig. 3B), this protein is unambiguously a component of the envelope membrane.

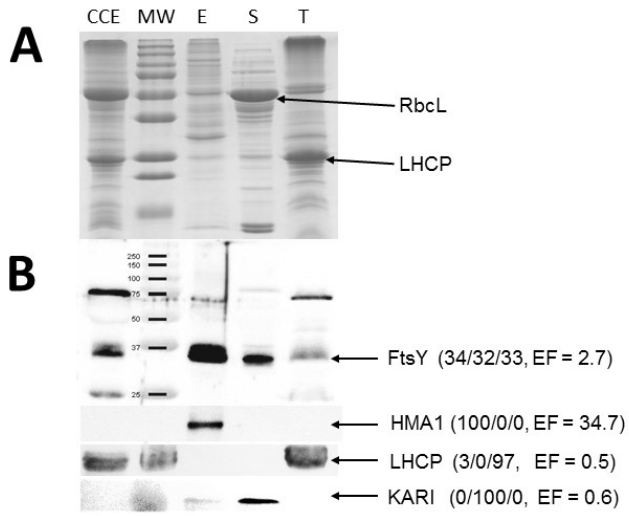


Fig. 9: Intermediate EF values are compatible with genuine envelope localization. A. Representative SDS-PAGE analysis. Each lane contains 15 µg of proteins. CE, crude leaf extract; MW, molecular weight; E, envelope fraction, S, stroma; T, thylakoid. RbcL, Large Subunit of Rubisco: LHCP, Light harvesting complex proteins. **B**. Validation of the multiple subplastidial localization of FtsY protein by western blot. Experiment performed on Arabidopsis plants, using FtsY antibody. The expected MW of the mature form (*i.e.* after cleavage of its plastid transit peptide) of FtsY is 35.1 kDa. Values between brackets are % of abundance in envelope, stroma and thylakoids according to Ferro *et al*. [12] and EF value (this work). As expected from previous data (see AT_CHLORO database (12)), FtsY is shared between the three chloroplast sub-compartments (34% in Envelope, 32% in Stroma and 33% in Thylakoid). Note that while unambiguously present in the envelope membrane, the EF value (2.7) of FtsY is far below the average EF value of envelope proteins (*i.e.* 13, see Fig. 3).

As an alternative to immunodetection of proteins, we also examined the subcellular and subplastidial localization of several other proteins *in planta*. WT Arabidopsis plants stably expressing UP1::GFP (AT1G11320, EF = 8.0) (40), SFR2::GFP (AT3G06510, EF = 7.7) (41), VTE1::GFP (AT4G32770, EF = 1.7) (42), TSP9::CFP (AT3G47070, EF = 0.2) (43) and eIF-5A::GFP (AT1G26630, EF = 0.1) (44, 45) were constructed in order to assess the compatibility between localization of the proteins and their EF values. As expected for a negative control, eIF-5A which has a very low EF value (0.1) is distributed in the cytosol. The thylakoid soluble phosphoprotein TSP9 (43) was distributed in the thylakoid subplastidial compartment as evidenced by its perfect co-localization with the chlorophyll auto-fluorescence (Fig. 10) in agreement with its very low EF value (0.2). Likewise, the protein encoded by AT1G11320, annotated UP1 (40) was, with no doubt, localized in the chloroplast envelope (12) (Fig. 10). Therefore, these results are in

good agreement with the respectively low and high EF values of TSP9 and UP1 (Fig. 10). VTE1 (42), the tocopherol cyclase involved in tocopherol synthesis, was suggested to be dually localized in the chloroplast envelope and the thylakoid (12). The EF value of VTE1 is 1.7, thus 8 times higher than TSP9 (EF = 0.2) which is only localized only in the thylakoid, but far below the EF of UP1 (EF = 8.0). Analyzing the subcellular localization of VTE1::GFP in Arabidopsis plant tissues (Fig. 10) revealed the existence of long extensions into the cytosol called stromules (46). According to Breuers *et al.* (46), these stromules are membrane proliferations derived from the envelope membrane, suggesting that VTE1 is potentially localized in this membrane system. However, when compared to UP1 images, besides stromules, VTE1 fluorescence also correlates with chlorophyll auto-fluorescence, in turn suggesting that VTE1 is also localized in the thylakoid compartment. Finally, we also analyzed the localization of SFR2 (Fig. 10), an outer envelope glycosyl hydrolase with a demonstrated role in protecting chloroplasts against freeze induced damage in Arabidopsis (41). Interestingly, in good agreement with its expected envelope localization, images obtained for SFR2 are close to the ones of UP1 (Fig. 10). Indeed, plastids are surrounded by SFR2::CFP fluorescence and there is a poor, if any, correlation between the chlorophyll autofluorescence and the fluorescence of SFR2::CFP. Accordingly, the high EF value (7.7) of SFR2 when compared to the one of VTE1 (1.7) suggest that SFR2 might not be shared by plastid envelope and other plastid sub-compartments.

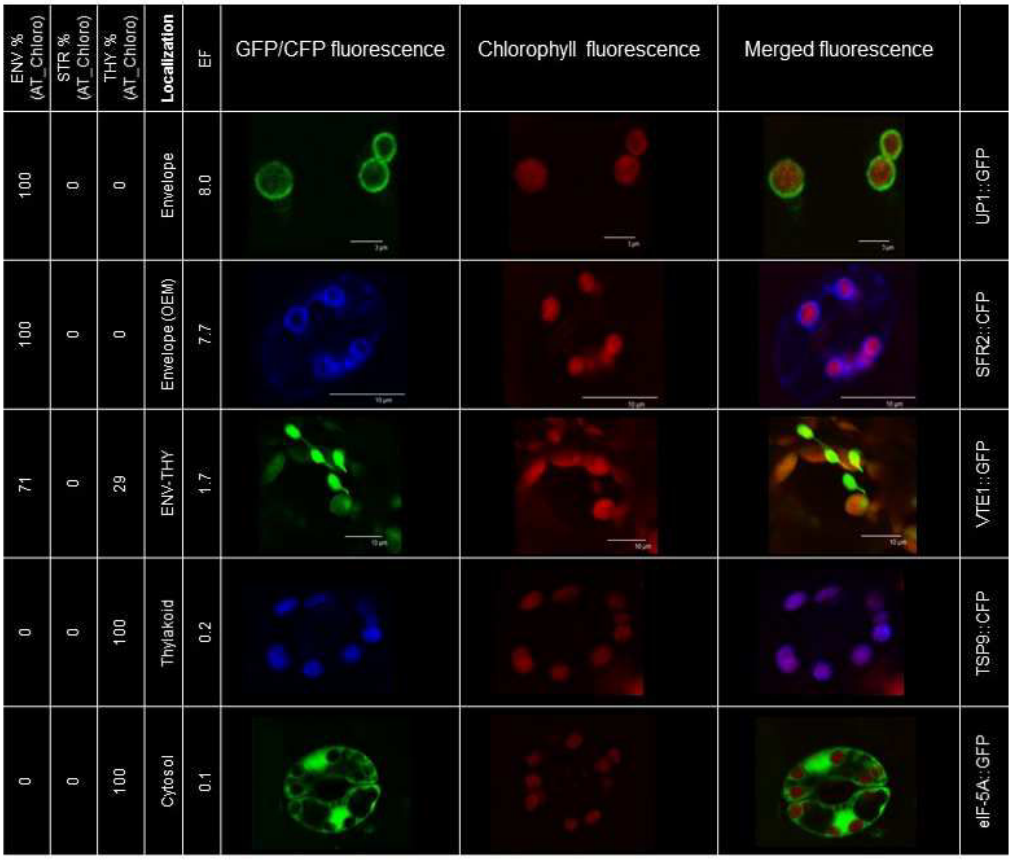


Fig. 10: Illustration of the good correlation between enrichment factor (EF) and subcellular or subplastidial localization of proteins deduced from *in planta* expression of GFP/CFP fusions. To investigate protein

localizations, Arabidopsis plants stably expressing UP1::GFP (envelope only, high EF), SFR2::CFP (outer envelope membrane, intermediate EF), VTE1::GFP (shared between envelope and thylakoid, intermediate EF), TSP9::CFP (thylakoid only, low EF) and eIF-5A::GFP (cytosol, low EF) were constructed. From left to right: data from AT_CHLORO (% in envelope, stroma and thylakoid) (12), deduced subcellular and subplastidial localization, EF deduced from this work, fluorescence and GFP or CFP fusion proteins, chlorophyll auto-fluorescence, merge of GFP:CFP and chlorophyll fluorescences, protein names. Note that the fluorescence of SFR2::CFP reveals that SFR2, which is localized in the outer envelope membrane, might also be shared with non-plastid structures, thus explaining the lower EF value when compared to the one of UP1. The perfect correlation between TSP9::CFP fluorescence and chlorophyll autofluorescence is in good agreement with the thylakoid localization of the protein and supports the low EF value.

High EF threshold levels exclude most envelope proteins shared with other compartments, but does not remove some ER/Golgi and vacuolar components.

A discussed above, increasing the EF threshold level excludes most other plastid and cell contaminants. However, we noticed that while high EF threshold levels (*i.e.* EF > 2) exclude most proteins from the stroma, thylakoid, cytosol, mitochondria, peroxisome, plasma membrane, nucleus and extracellular proteins, they do not exclude non-negligible parts of ER/Golgi (30%) and vacuolar (58%) components (see Fig. 8 and supplemental table S5, "Only EF>1", "Only EF>2"). Furthermore, when calculating the impact of higher EF thresholds on average EF values of remaining proteins, we noticed the high average EF values of the 33 remaining ER/Golgi and the 26 remaining vacuolar proteins (see supplemental Fig. S4, black squares). These average EF values are respectively 4 and 3 for remaining ER/Golgi and vacuolar proteins, *i.e.* above genuine chloroplast envelope proteins (like FTSY, see above) shared with other compartments (see Fig. 9). We thus questioned whether the presence of some vacuolar and ER/Golgi components detected in purified envelope fractions could be explained by specific high levels of cross-contaminations. A shown in Fig. 11A, we thus purified crude protein extracts from leaf (CCE), microsomes, tonoplast, Percoll-purified chloroplasts, envelope, stroma and thylakoids. These fractions were then tested for their cross-contamination using markers from envelope, ER/Golgi and tonoplast components (Fig. 11B). As expected, HMA1 (envelope marker) was strongly enriched when compared to CCE, and only long time exposure (Long ECL detection, see Fig. 11B) allowed detecting its signal within microsomes and tonoplast lanes. The same was true for Bip (ER marker) which was strongly enriched in microsomal fraction when compared to CCE, and contaminates purified tonoplast fractions. Similarly, V-ATPAse (tonoplast marker) is strongly enriched in the purified tonoplast fraction when compared to CCE. Bip and V-ATPAse signals were undetectable in purified envelope fractions, and only long time exposures (Long ECL detection, see Fig. 11B) allowed detecting their signals within envelope lane. At this detection level, we could notice that both Bip and V-ATPAse signals are not enriched in the envelope when compared to CCE (Fig. 11B). In other words, if one expects to detect minute amounts of ER/Golgi and vacuolar proteins in purified envelope fractions thanks to present-days high sensitivity of MS, these proteins should not be as enriched as proteins only localized in envelope membranes. However, from these observations, we cannot rule out that some abundant markers of ER/Golgi and vacuole could get higher EF than envelope proteins shared with other plastid or cell compartments. In other words, only specific

assays (immuno-localization or *in planta* analyses) should allow getting more detailed information about the putative sharing of these ER/Golgi or vacuole proteins with the chloroplast envelope, even if their repetitive detection in purified ER and vacuole fractions strongly support their expected localization in these compartments.

Surprisingly, during the course of these analyses, the signal detected using the bob-TIP antibody (raised against *N*- and *C*-ter of the cauliflower TIP1.1, another tonoplast marker, see Fig. 11C) was, as expected, enriched in the tonoplast when compared to CCE, but was also detected at a similar level in purified envelope fractions (Fig. 11B). It is important to note here that TIP1.1 (At2g36830) was the only TIP isoform previously detected in envelope fraction by both Simm *et al*. (14) and Ferro *et al*. (12). It was repeatedly localized in the tonoplast using *in planta* expression of GFP fusions, but was hardly detected in purified tonoplast fractions (see supplemental table S5, "This work", lane 793, columns "Location GFP" and "Experimental evidence"). Altogether, these observations suggest that either specific cross-contamination or dual localization of this vacuolar protein in the chloroplast envelope cannot be excluded for this expected tonoplast marker.

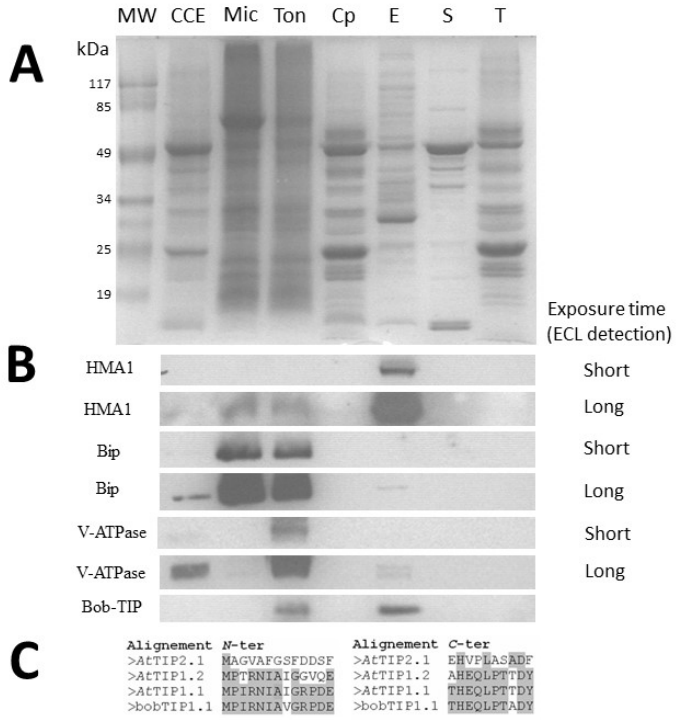


Fig. 11: Detection of vacuolar and ER/Golgi components in purified envelope fractions is partially explained by cross-contamination. A. **SDS-PAGE analysis**. Each lane contains 15 µg of proteins. MW, molecular weight; CCE, crude leaf extract; Mic, microsomal fraction, Ton, tonoplast fraction, Cp, Percoll-purified chloroplast, E, envelope fraction, S, stroma; T, thylakoid. **B**. **Analysis of envelope contamination with markers from ER/Golgi and tonoplast components.** Validation of the enrichment of HMA1 (envelope marker) when compared to CCE. As expected, Bip (ER marker) is enriched in microsomal fraction and contaminates purified tonoplast fractions. The same is true for V-ATPAse (tonoplast marker) which is strongly enriched in the tonoplast fraction. Note that both Bip and V-ATPAse signals are barely detectable in purified envelope fractions, but that these signals are not enriched when compared to CCE. Surprisingly, the signal detected using the bob-TIP antibody (raised against N- and C-ter of the

cauliflower TIP1.1, another tonoplast marker), is similar in the tonoplast and purified envelope fractions thus suggesting specific cross-contamination or dual localization. Note that this protein was the only TIP isoform previously detected in envelope fraction by both Simm et al. [33] and Ferro et al. [9]. **C.** Alignment of *N*- and *C*-ter of cauliflower bobTIP1.1 supporting detection of the TIP1.1 protein in Arabidopsis samples.

4. Discussion

The aim of the present study was to revisit the chloroplast envelope proteome, taking advantage of the establishment of ever-actualized subcellular proteome databases, and of the improvement of MS sensitivity to perform a quantitative proteome study comparing purified envelope fractions and crude cell extracts. As expected, many more protein were detected when compared to previous studies, including well-known but previously undetected chloroplast envelope proteins (see Table 1 and supplemental table S5 "This work"). As also expected, the present investigation detected large amounts of proteins deriving from other cell compartments (*i.e.* putative contaminants, see Fig. 3) with a specific overrepresentation of proteins from the cytosol, the ER/Golgi and the vacuole when compared to previous studies (see Fig. 3 and 4, and Fig. 8). Accessibility to a quantitative approach (EF parameter), thanks to better MS sensitivity, allowed revisiting the chloroplast envelope proteome and to easily exclude most contaminants from other plastid or cell compartments (see Fig. 3 and Fig. 6 to 8). The overall quality of the present proteome was thus not downgraded by detecting proteins that are highly abundant in other cell compartments.

Improved annotation of proteins - During this work, more than identifying novel envelope components (Fig. 4) we also provide a detailed manual annotation based on recent characterization of some of these proteins (see supplemental table S5, "This work", column "curated description"). These recent annotations (*i.e.* up to early 2018) strongly revisited the information present in AT_CHLORO database inventorying the MS-based list of identified envelope proteins (12). Indeed, 185 of the 413 envelope proteins (ENV), 22 of the 54 outer envelope membrane proteins (OEM) and 42 of the 46 envelope candidates (ENV?) categories could be redefined thanks to recent publications related to functional characterization of the protein (or of one member of its family), or similarities with characterized orthologous proteins from other species, or even identification of functional domains predicting specific functions. To cite a few, chloroquine-resistance transporter-like proteins (AT5G19380 and AT5G12170) were shown to be required for glutathione homeostasis and stress responses (47). BASS2 (At2g26900, former called IEP36) is now characterized as a plastidial sodium-dependent pyruvate transporter (48). Members of the FAX family (At3g57280, At2g38550, At3g43520, former called HP26-like proteins) were recently characterized as novel membrane proteins mediating plastid fatty acid export (49). Even if their definitive role in targeting proteins to plastids is still a matter of debate (50, 51), Tic214 (AtCg01130, former called Ycf1) and Tic56 (At5g01590, former called HP65) were recently shown to be involved in a 1-MDa protein complex protein comprising Tic214, Tic56, Tic100 (At5g22640) and the protein conducting channel protein Tic20-I (At1g04940), at the inner chloroplast envelope membrane (52, 53). VTE6 (At1g78620, former called HP34) was demonstrated to be a phytyl-phosphate kinase catalyzing the conversion of phytyl-monophosphate

to phytyl-diphosphate (54). ceQORH (At4g13010), a former putative quinone oxidoreductase was recently demonstrated to reduce long-chain, stress-related oxidized lipids (55) and its structure was solved (56).

As previously observed, SecA2 (At1g21650) and SCY2 (At2g31530) were highly enriched in the envelope fraction compared to the total extract. While SecA2 was previously detected in the envelope fraction by proteomic approach (12), SCY2 was identified for the first time during the present work. Both proteins were recently described as components of a recently identified Sec translocon, localized in the inner envelope membrane, which likely integrates into the envelope a subset of inner envelope membrane proteins (57-59).

Altogether, these data now limit the number of envelope components or expected functions that were missing in previous analyses targeted to envelope fractions. However, while many new or expected but previously undetected envelope proteins were identified during this work (see supplemental table S5, "This work", column "curated description"), expected envelope components are still lacking. This is, for example, the case of numerous envelope transporters that are expected to control exchanges of ions and metabolites across the chloroplast envelope (for a review, see (9)), but whose molecular identity remains to be determined. During this work, we also detected several chloroplast encoded proteins that are enriched in envelope fractions: the protease ClpP1 (ATCG00670), the beta chain of the acetyl-CoA carboxylase ACCDb (ATCG00500), Ycf1 (ATCG01130), and Ycf2 (ATCG00860) proteins. However, the chloroplast encoded protein Ycf10 or CemA (AtCg00530) (60), which, to our knowledge, has never been detected using a MS-based approach targeting envelope or any other chloroplast sub-compartment was, once more, not detected here. To cite few more, some expected components of the TIC (*i.e.* Tic20-IV, At4g03320) and Toc (Toc75-iV, At4g09080) complexes were also not detected during this work.

The present analysis also raises numerous questions about the genuine subcellular localization of some identified proteins. Due to space constraints, it was not possible to describe all peculiar cases that were classified during manual annotation of the 1269 proteins detected in purified envelope fractions. In the following paragraphs, we will however discuss some few interesting cases that might deserve specific attention.

Few examples of genuine envelope proteins whose EF values help predicting subplastidial localization - Unexpectedly, during this analysis, we noticed that some genuine plastid and expected envelope proteins were not enriched in the envelope fraction (Fig. 3B). Most of these proteins are most probably shared between envelope and other subcellular or subplastidial compartment. For example, four allene oxide cyclase (AOC1, AOC2, AOC3 and AOC4) were previously described as enzymes of the jasmonate pathway that are required for the jasmonate biosynthesis (61, 62). Together with AOS (Allene oxide synthase), the four AOC isoforms were previously demonstrated to be targeted to plastids (62) and the three AOC1, 2 and 3 isoforms were previously detected in purified envelope fractions (12). AOC4, which was not previously identified by MS-based approaches in the chloroplast envelope fraction, was also detected during this work. During the present study, we were able to show that AOC3 is the only isoform of AOC that is

enriched in the envelope (see supplemental table S5, "This work", lanes 65-68). This observation suggest that all three other isoforms might be shared with other plastid compartments.
Transporters were also deeply analyzed here. Interestingly, as cited above, several members of this functional category were given a substrate recently. The present work also helped solving question about their subplastidial locations. As controls, KEA3 (At4g04850), a recently described thylakoid K(+) efflux antiporter (63) has an expected low EF value of 0.3 (see supplemental table S5, "This work", lane 1220). On the other hand, its close homologues, KEA1 (At1g01790) and KEA2 (At4g00630), being also described envelope transporters (64, 65) have EF values of 8 and 15 respectively (see supplemental table S5, "This work", lanes 286-287). We previously provided data (12) in contradiction with previous information describing TAAC (AT5G01500), as a thylakoid ATP/ADP carrier. Indeed, this protein was previously detected in purified envelope fraction and absent from purified thylakoid membranes (12). In the present work, as before, the protein was only detected in envelope fractions (see supplemental table S5, "This work", lanes 292) with a high EF value of 37.1, thus definitively ruling out its association with thylakoid membranes. The same is true for the thylakoid transporter ANTR1 (AT2G29650). Again, this protein was described as a thylakoid Na+-dependent phosphate transporter from Arabidopsis (see (12)) while our previous data (proteomics and western blot analyses using specific polyclonal antibodies, see (12)) clearly demonstrated that this protein was associated to the envelope. Again, the present work, strongly support our previous data since this protein was only detected in envelope fractions (see supplemental table S5, "This work", lanes 312) with an EF value of 8.9, thus also ruling out its association with thylakoid membranes. The function of this protein was recently revisited, and it was demonstrated to catalyze ascorbate transport (66). This ascorbate transporter is thus, without a doubt, associated to the chloroplast envelope.

The good correlation between EF values of proteins being involved in the same multisubunit complexes was discussed above for Clp and FtsH subunits (Fig. 7B). In a model proposed by Nakai (67), Toc34 (At5g05000), Toc132 (At2g16640), Toc120 (At3g16620) are components of a first TOC complex required for import of "non-photosynthetic" proteins, while Toc33 (At1g02280) and Toc159 (At4g02510) are part of a second "photosynthetic" complex devoted to import of proteins linked to photosynthesis. Interestingly, Toc33 (EF = 16.5) and Toc159 (EF = 11.5) that are part of the same "photosynthetic-type" TOC complex have similar intermediate EF values, while the three components of the "non-photosynthetic-type" TOC complex Toc34 (EF = 22.7), Toc132 (EF = 38.4) and Toc120 (EF = 5.2) show less homogenous EFs values. In the same model proposed by Nakai (67), the "photosynthetic type" major TIC complex would be made of Tic56 (At5g01590), Tic100 (At5g22640), Tic20-I (At1g04940) and Tic2014/Ycf1 (AtCg01130). Interestingly, Tic56 (EF = 56.1), Tic100 (EF = 90.2), Tic20-1 (EF = 10.9) and Tic 214 (EF = 41.9) have higher EF values when compared to all other detected Tic subunits (see supplemental table S5, "This work", lanes 150-163).

It is however important to note here that the crude cell extract used during this study contains plastids proteins deriving from all leaf tissues (including epidermal cells), while purified envelope fractions mostly derive from photosynthetic chloroplasts issued from parenchymal cell. While

probably limited, this small bias might impact on EF values of some genuine envelope proteins, being part of the same complex. However, if one excludes the higher EFs of Toc34 (22.7) and Toc132 (38.4), most other Toc subunit share similar EF values with a relatively low degree of dispersion (around 14) (EF Toc33 = 16.5, EF TOC64-III = 14.6, EF Toc75-III = 9.6, EF Toc75-V = 16.9, EF Toc90 (9.0), EF Toc120 = 5.2, EF Toc159 = 11.5).

Cross-contamination by mitochondria components - Qualitative proteome analyses identified 32 mitochondrial proteins (Fig. 3) with an expected low EF (average EF = 0.7). With very few exceptions (At3g27930 and At1g28690), these proteins were also repeatedly detected in purified mitochondria during previous MS-based studies (see supplemental table S5, "This work", lanes 700-731). The protein with the highest EF value (EF = 7.9, only detected in purified envelope fractions) was an ethanolamine-phosphate cytidylyltransferase called PECT1 (AT2G38670), a rate-limiting enzyme in phosphatidylethanolamine biosynthesis. Using enzymatic measurement, MS-based analyses, GFP fusions and *in planta* analyses (see (68) and references therein), this protein was previously associated to the mitochondrial periphery, most likely at the interface between the cytosol and the outer membrane of mitochondria. Interestingly, this protein was also detected in purified envelope membranes from *Medicago Sativa* (14), thus supporting our data and suggesting that PECT1 might also interact with the plastid surface. We however chose to maintain this protein in the list of mitochondrial components. The same was true for a NADH-ubiquinone oxidoreductase (AT4G16450) with the subsequent highest EF value (EF = 3.9) for a mitochondrial protein. Interestingly, PECT1 and this NADPH-ubiquinone oxidoreductase were also both detected in purified envelope membranes during the recent study of Simm *et al.* (14).

All other known mitochondrial proteins, known to be shared by mitochondria and chloroplasts (dual targeted), appear as envelope components in the supplemental table S5. We even chose to include in the list of envelope components, few previously known mitochondria proteins that were shown to be highly enriched in purified envelope fractions. For example, MSL1 (At4g00290, also called MscC) was recently shown to be a mechanosensitive ion channel that dissipates mitochondrial membrane potential and maintains redox homeostasis in mitochondria during abiotic stress (69). During this work, this protein was only detected in the chloroplast envelope (see supplemental table S5, "This work", lane 277) with a very strong EF of 27.8, *i.e.* far above the 0.7 average impact factor of mitochondrial proteins. The same is true for ATM3 (At5g58270, also called ABCB25, STA1 or STARIK1). This protein was recently proposed to export glutathione polysulfide, containing glutathione and persulfide, from mitochondria, for iron-sulfur cluster assembly in the cytosol (70). Again, this protein was only detected in the chloroplast envelope during this work (see supplemental table S5, "This work", lane 266) with an EF of 66.4, thus far above the average impact factor of genuine envelope proteins).

The few above-cited examples thus raise the question of the genuine localization of these proteins or at least, open the question of the possible sharing of these mitochondrial proteins with the chloroplast envelope. In other words, the application of the present analyses opens up interesting possibilities to answer targeted questions about subcellular localization of proteins.

Cross-contamination by vacuolar components - Qualitative proteome analyses identified various vacuolar and particularly, tonoplast proteins with a surprisingly high EF (see supplemental Fig. S4). This can be explained by specific co-purification of the two light membrane systems (envelope and tonoplast). Alternatively, this could also result from sharing of specific tonoplast proteins with envelope membranes. The TIP1.1 (At2g36830, see supplemental table S5, "This work", lane 793) might be one example even if its EF (2.0) is relatively low. Similarly, autoinhibited Ca2+-ATPases ACA4 (At2g41560) and ACA11 (At3g57330, see supplemental table S5, "This work", lanes 797-798) were only detected or highly enriched in envelope fractions during this work (respective EF values of 17 and 7.7). While repeatedly detected in purified tonoplast fractions, ACA11 is the only tonoplast protein containing a predicted (TargetP) chloroplast transit peptide. The vacuolar H+-ATPase (V-ATPase) is a membrane-bound multisubunit enzyme complex composed of at least 14 different subunits. Arabidopsis V-ATPase has three B subunits (AtVAB1, AtVAB2, and AtVAB3). These B subunits, including VATB3 (At1g20260, see supplemental table S5, "This work", lane 805), the tonoplast protein with the highest EF value (EF = 22.3, see highest outlier of the "Vacuole" group in Fig. 6B), were recently shown, besides their known role as vacuolar ATPAse subunits, to play distinct roles in Arabidopsis cells (71). Indeed, AtVABs subunits were demonstrated to bind to and co-localize with F-actin, bundle F-actin to form higher order structures, and stabilize actin filaments *in vitro*. In addition, the AtVABs also show different degrees of activities in capping the barbed ends of actin filaments, but no nucleating activities, and these activities were not regulated by calcium. AtVAB1 and AtVAB3 inhibited depolymerization significantly, whereas AtVAB2 weakly inhibited actin depolymerization. In plant cells, the formation of higher order actin structures, such as bundles and cables, is crucial to stabilize the organization of actin strands and maintain the overall cellular architecture. A subset of overlapping activities of AtVABs may be involved in stabilizing long actin filaments. Interestingly, in plant cells, plastids are closely associated with actin microfilaments. A direct interaction of plastids with the actin cytoskeleton has been postulated to anchor chloroplasts at appropriate intracellular positions, to support chloroplast light-intensity dependent movement, to facilitate plastid stromule mobility and to participate in gravity perception. The known proteins implicated in plastid-actin interaction are CHUP1, an outer envelope membrane protein that is essential for chloroplast anchorage to the plasma membrane, myosin XI proteins that play a role in stromule movement and in gravitropism, and Toc159, a component of the TOC complex (for a review, see (72)). Interestingly, actin (At5g09810), myosin (At1G64330) or tubulin subunits (At5g19770 and At5g62690), CHUP1 (At3g25690) and Toc159 (At4g02510) were also detected during this work (see supplemental table S5, "This work", lanes 661-664, and lanes 437 and 451, respectively). While actin and tubulin subunits share the same low EF of 0.1, values for myosin, CHUP1 and Toc159 are intermediate to high (2.0, 3.6 and 11.5, respectively), *i.e.* below the average EF of envelope proteins but above the average EF of other cytosol components, and thus, in agreement with a putative shared localization between envelope and other cell compartments.

Cross-contamination by ER/Golgi components – Intriguingly, as shown in Fig. 3 and 4, many of the newly detected proteins in the purified chloroplast envelope fraction compared to previous

studies targeted to envelope membranes (12, 14), were assigned to the endoplasmic reticulum (ER) and the Golgi. While their presence in purified envelope fractions might result from isodensity of light membrane vesicles (in the original sucrose gradient) and thus cross contamination, recent data suggest that part of these proteins could be potential interactors with the outer chloroplast envelope membrane surface rather than cell contaminants. Indeed, few chloroplast proteins were shown to transit through the ER en route to the chloroplast (73). The same is true for lipid vesicular trafficking which was proposed to transit through membrane contact sites connecting both organelles (for a review, see (74). Thus, some of the proteins detected during this work could be required for such processes. Glycerolipid synthesis in plant cells is characterized by an intense trafficking of lipids between the ER and chloroplasts. A list of the main expected or putative Arabidopsis proteins playing a role in lipid transfer between plastid and ER was recently established (74). Interestingly, this short list is composed of proteins from the chloroplast IEM or OEM, the cytosol and the ER. During this work, we detected members from all involved cell compartments: *i*) the Phospholipase D PLDα1 (At3g15730, EF = 0.04) from the cytosol, *ii*) the two long chain acyl-CoA synthetases LACS4 (At4g23850, EF = 0.8) and LACS8 (At2g04350, EF = 0.1) from the ER, *iii*) the fatty acid translocator FAX1 (At3g57280, EF = 4.3), lipid translocators TGD1 (At1g19800, EF = 21.8), TGD2 (At3g20320, EF = 7.8) and TGD3 (At1g65410, EF = 28.5), the MGDG synthase MGD1 (At4g31780, EF = 37.8), and the three lipid desaturases FAD6 (At4g30950, EF = 68.3), FAD7 (At3g11170, EF = 43.2) and FAD8 (At5g05580, EF = 5.3) from the chloroplast IEM (but not FAD2 and FAD3 from the ER), *iv*) the DGDG synthase DGD1 (At3g11670, EF = 5.3) and the lipid translocator LPTD1 or TGD4 (At3g06960, EF = 6.7) from the chloroplast OEM (whose subcellular localization was unclear, chloroplast associated or ER in (74)) but which was classified as an OEM protein during this work thanks to its EF value of 6.7, in good agreement with recently published data (75). In other words, the correlation between EF values of known or predicted location of the main actors of the ER to chloroplast lipid trafficking process was impressively good.

Cross-contamination by cytosol components - All cytosol proteins identified during this work showed a very low EF with an average of 0.4 (Fig. 3). One exception could be the ALBA protein (AT1G76010, EF = 3.96) which was previously identified during the determination of the mRNA-binding proteome of arabidopsis etiolated seedlings (76). However, since this protein was previously detected several times in the cytosol or as a cytoskeleton binding protein using MS-based approaches, it was maintained in the list of cytosol components. Interestingly, 49 members of the cytosol translation machinery were also detected in purified envelope fractions during this work. This includes several initiation and elongation factors, but also 42 subunits of the ribosome. None of these proteins were enriched in purified envelope membranes. However, the presence of cytosolic ribosomes on the surface of mitochondria was recently demonstrated, highlighting how protein synthesis in the cytosol may be coupled with protein targeting to the organelle and strongly supporting a role for cotranslational import (77). Presence of this mechanism at the chloroplast surface would explain the detection of these ribosomal proteins, their sharing with other cell compartments explaining their low EF. We also detected HPL1 (At4g15440), a hydroperoxide

lyase that, while cytosolic, was previously proposed to be involved in the degradation of lipid catabolites produced in the chloroplast envelope (78, 79). In good agreement with the previous work of Simm *et al.* (14), this protein was also detected in purified envelope fractions during this work (see supplemental table S5, "This work", lane 425). The low EF value of HPL1 (0.9) might result from a transient interaction of this protein with the outer surface of the chloroplast, this protein being most probably shared by the cytosol and the outer chloroplast envelope membrane.

Cross-contamination by peroxisomal components – During this work, 15 proteins annotated as peroxisomal components were identified. Their function range from lipid metabolism to oxidative stress or photorespiration (see supplemental table S5, "This work", lanes 812-826). Experimental evidences from MASCP gator or SUBA show that most of these proteins are known markers of peroxisomes. As expected for abundant markers of the peroxisome, some of these proteins were however repeatedly detected in purified envelope, stroma or thylakoid fractions (see supplemental table S5, "This work", from column "ENV %" to column "Simm *et al*., 2014"). From previous data analyzing the subcellular localization of members of the monodehydroascorbate reductase family (80), MDAR6 (At1g63940) was classified as a protein from the chloroplast stroma (EF = 0.3) and MDAR1 (At3g52880) was located in peroxisomes (EF = 0.02). MDAR5 (mitochondria) and MDAR2 and 3 (cytosol) members were not detected during this work. Surprisingly, the MDAR4 member (At3g27820), previously detected in outer envelope membrane fractions purified from Pea chloroplasts (14) was also enriched [EF = 2.6] in our purified envelope fractions (see supplemental table S5, "This work", lane 826). This MDAR6 isoform was previously shown to be localized in peroxysomes (transient expression of GFP fusions in tobacco or Arabidopsis cells), and to be the only membrane associated member of the MDAR family thanks to a 60 residues extension located at its *C*-terminus (80). Peroxisomes are highly motile organelles that can be juxtaposed to chloroplasts, which has led to the hypothesis that they are physically interacting. Recently, Gao and colleagues (81) showed that chloroplasts and peroxisomes are physically tethered through peroxules, a poorly described structure in plant cells. Their observation suggests that peroxules might have a role in maintaining peroxisome-organelle interactions that could be important for fatty acid mobilization and photorespiration through the interaction with oil bodies and chloroplasts. These recent observations and the described novel role of these structures highlighting a fundamentally important role for organelle interactions for essential biochemistry and physiological processes might also provide some explanation (alternative to cross-contamination) for the detection of some of these peroxisomal components in the purified envelope fractions during the present work.

Many investigations of the chloroplast proteome have focused on either whole chloroplast fractions or on independent subplastidial fractions within the last two decades (11, 12). However, while highly informative, most of these databases do not contain accurate information on protein localization within the different subplastidial compartments. Therefore, the accurate localization of many envelope proteins remains mostly hypothetical. During this work, the use of MS-based parameters and manual annotation result in a fairly good correlation (see supplemental table S7) between subcellular localization deduced from our analyses and SUBAcon (82) or MASCP Gator

(16). However, the present data complement these databases with information about subplastidial location of shared proteins. Now that the level of MS sensitivity is sufficient to in-depth analyze the proteome of minor cell compartments and compare it to the proteome of crude cell extracts, the approach completed during this work should allow revisiting other cell compartments. However, the success of these studies will rely on a prerequisite: only a strong enrichment factor will allow distinguishing minor components from cross-contaminants, *i.e.* abundant markers from other cell compartments.

To conclude, although work on chloroplast biology is progressing rapidly, the envelope compartment remains the hidden part of the organelle. Homology searches of many envelope proteins fail to identify homologues in other prokaryotic or eukaryotic species suggesting that there are fundamental specificities in the controlled exchanges of ions, metabolites, proteins, metabolic pathways and other mechanisms involved in regulation of the morphology and the dynamics in higher plant chloroplasts. This is reflected in the results of our manual annotation where many envelope proteins identified during this work remain unknown or uncharacterized proteins, some of them with previously unknown or erroneous subcellular or subplastidial localization. Altogether, protein identified here constitute a powerful resource that will help colleagues, in search of gene candidates associated to this specific compartment, to delineate the principles and mechanisms controlling fundamental aspects of plastid biogenesis and functions.

References

1. Leister, D. (2016) Towards understanding the evolution and functional diversification of DNA-containing plant organelles. *F1000Research* 5
2. Trentmann, O., and Haferkamp, I. (2013) Current progress in tonoplast proteomics reveals insights into the function of the large central vacuole. *Frontiers in plant science* 4, 34
3. Gould, S. B., Waller, R. F., and McFadden, G. I. (2008) Plastid evolution. *Annual review of plant biology* 59, 491-517
4. Cline, K., and Dabney-Smith, C. (2008) Plastid protein import and sorting: different paths to the same compartments. *Current opinion in plant biology* 11, 585-592
5. Jarvis, P., and Lopez-Juez, E. (2013) Biogenesis and homeostasis of chloroplasts and other plastids. *Nature reviews. Molecular cell biology* 14, 787-802
6. Li, H. M., and Chiu, C. C. (2010) Protein transport into chloroplasts. *Annual review of plant biology* 61, 157-180
7. Sakamoto, W., Miyagishima, S. Y., and Jarvis, P. (2008) Chloroplast biogenesis: control of plastid development, protein import, division and inheritance. *The Arabidopsis book* 6, e0110
8. Block, M. A., Douce, R., Joyard, J., and Rolland, N. (2007) Chloroplast envelope membranes: a dynamic interface between plastids and the cytosol. *Photosynthesis research* 92, 225-244
9. Rolland, N., Curien, G., Finazzi, G., Kuntz, M., Marechal, E., Matringe, M., Ravanel, S., and Seigneurin-Berny, D. (2012) The biosynthetic capacities of the plastids and integration between cytoplasmic and chloroplast processes. *Annual review of genetics* 46, 233-264

10. Joyard, J., Block, M., Pineau, B., Albrieux, C., and Douce, R. (1990) Envelope membranes from mature spinach chloroplasts contain a NADPH:protochlorophyllide reductase on the cytosolic side of the outer membrane. *The Journal of biological chemistry* 265, 21820-21827
11. Agrawal, G. K., Bourguignon, J., Rolland, N., Ephritikhine, G., Ferro, M., Jaquinod, M., Alexiou, K. G., Chardot, T., Chakraborty, N., Jolivet, P., Doonan, J. H., and Rakwal, R. (2011) Plant organelle proteomics: collaborating for optimal cell function. *Mass spectrometry reviews* 30, 772-853
12. Ferro, M., Brugiere, S., Salvi, D., Seigneurin-Berny, D., Court, M., Moyet, L., Ramus, C., Miras, S., Mellal, M., Le Gall, S., Kieffer-Jaquinod, S., Bruley, C., Garin, J., Joyard, J., Masselon, C., and Rolland, N. (2010) AT_CHLORO, a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins. *Molecular & cellular proteomics : MCP* 9, 1063-1084
13. Lilley, K. S., and Dupree, P. (2007) Plant organelle proteomics. *Current opinion in plant biology* 10, 594-599
14. Simm, S., Papasotiriou, D. G., Ibrahim, M., Leisegang, M. S., Muller, B., Schorge, T., Karas, M., Mirus, O., Sommer, M. S., and Schleiff, E. (2013) Defining the core proteome of the chloroplast envelope membranes. *Frontiers in plant science* 4, 11
15. Mann, G. W., Calley, P. C., Joshi, H. J., and Heazlewood, J. L. (2013) MASCP gator: an overview of the Arabidopsis proteomic aggregation portal. *Frontiers in plant science* 4, 411
16. Heazlewood, J. L., Verboom, R. E., Tonti-Filippini, J., Small, I., and Millar, A. H. (2007) SUBA: the Arabidopsis Subcellular Database. *Nucleic acids research* 35, D213-218
17. Salvi, D., Rolland, N., Joyard, J., and Ferro, M. (2008) Purification and proteomic analysis of chloroplasts and their sub-organellar compartments. *Methods in molecular biology (Clifton, N.J.)* 432, 19-36
18. Yamaryo, Y., Dubots, E., Albrieux, C., Baldan, B., and Block, M. A. (2008) Phosphate availability affects the tonoplast localization of PLDzeta2, an Arabidopsis thaliana phospholipase D. *FEBS letters* 582, 685-690
19. Chua, N.-H. (1980) [40] Electrophoretic analysis of chloroplast proteins. *Methods in Enzymology* 69, 434-446
20. Seigneurin-Berny, D., Gravot, A., Auroy, P., Mazard, C., Kraut, A., Finazzi, G., Grunwald, D., Rappaport, F., Vavasseur, A., Joyard, J., Richaud, P., and Rolland, N. (2006) HMA1, a new Cu-ATPase of the chloroplast envelope, is essential for growth under adverse light conditions. *The Journal of biological chemistry* 281, 2882-2892
21. Betz, M., and Dietz, K. J. (1991) Immunological characterization of two dominant tonoplast polypeptides. *Plant physiology* 97, 1294-1301
22. Fermin, D., Basrur, V., Yocum, A. K., and Nesvizhskii, A. I. (2011) Abacus: a computational tool for extracting and pre-processing spectral count data for label-free quantitative proteomic analysis. *Proteomics* 11, 1340-1345
23. Vizcaino, J. A., Csordas, A., Del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q. W., Wang, R., and Hermjakob, H. (2016) 2016 update of the PRIDE database and its related tools. *Nucleic acids research* 44, 11033

24. Chiu, W.-l., Niwa, Y., Zeng, W., Hirano, T., Kobayashi, H., and Sheen, J. (1996) Engineered GFP as a vital reporter in plants. *Current Biology* 6, 325-330
25. Clough, S. J., and Bent, A. F. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *The Plant journal : for cell and molecular biology* 16, 735-743
26. Ferro, M., Salvi, D., Brugiere, S., Miras, S., Kowalski, S., Louwagie, M., Garin, J., Joyard, J., and Rolland, N. (2003) Proteomics of the chloroplast envelope membranes from Arabidopsis thaliana. *Molecular & cellular proteomics : MCP* 2, 325-345
27. Emanuelsson, O., Nielsen, H., and von Heijne, G. (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein science : a publication of the Protein Society* 8, 978-984
28. Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of molecular biology* 300, 1005-1016
29. Tanz, S. K., Castleden, I., Small, I. D., and Millar, A. H. (2013) Fluorescent protein tagging as a tool to define the subcellular distribution of proteins in plants. *Frontiers in plant science* 4, 214
30. Joshi, H. J., Hirsch-Hoffmann, M., Baerenfaller, K., Gruissem, W., Baginsky, S., Schmidt, R., Schulze, W. X., Sun, Q., van Wijk, K. J., Egelhofer, V., Wienkoop, S., Weckwerth, W., Bruley, C., Rolland, N., Toyoda, T., Nakagami, H., Jones, A. M., Briggs, S. P., Castleden, I., Tanz, S. K., Millar, A. H., and Heazlewood, J. L. (2011) MASCP Gator: an aggregation portal for the visualization of Arabidopsis proteomics data. *Plant physiology* 155, 259-270
31. Pohlmeyer, K., Soll, J., Grimm, R., Hill, K., and Wagner, R. (1998) A high-conductance solute channel in the chloroplastic outer envelope from Pea. *The Plant cell* 10, 1207-1216
32. Hiltbrunner, A., Grunig, K., Alvarez-Huerta, M., Infanger, S., Bauer, J., and Kessler, F. (2004) AtToc90, a new GTP-binding component of the Arabidopsis chloroplast protein import machinery. *Plant molecular biology* 54, 427-440
33. Kubis, S., Patel, R., Combe, J., Bedard, J., Kovacheva, S., Lilley, K., Biehl, A., Leister, D., Rios, G., Koncz, C., and Jarvis, P. (2004) Functional specialization amongst the Arabidopsis Toc159 family of chloroplast protein import receptors. *The Plant cell* 16, 2059-2077
34. Kasmati, A. R., Topel, M., Patel, R., Murtaza, G., and Jarvis, P. (2011) Molecular and genetic analyses of Tic20 homologues in Arabidopsis thaliana chloroplasts. *The Plant journal : for cell and molecular biology* 66, 877-889
35. Hirabayashi, Y., Kikuchi, S., Oishi, M., and Nakai, M. (2011) In vivo studies on the roles of two closely related Arabidopsis Tic20 proteins, AtTic20-I and AtTic20-IV. *Plant & cell physiology* 52, 469-478
36. Lu, X., Zhang, D., Li, S., Su, Y., Liang, Q., Meng, H., Shen, S., Fan, Y., Liu, C., and Zhang, C. (2014) FtsHi4 is essential for embryogenesis due to its influence on chloroplast development in Arabidopsis. *PloS one* 9, e99741

37. Nielsen, E., Akita, M., Davila-Aponte, J., and Keegstra, K. (1997) Stable association of chloroplastic precursors with protein translocation complexes that contain proteins from both envelope membranes and a stromal Hsp100 molecular chaperone. *The EMBO journal* 16, 935-946
38. Moreno, J. C., Tiller, N., Diez, M., Karcher, D., Tillich, M., Schottler, M. A., and Bock, R. (2017) Generation and characterization of a collection of knock-down lines for the chloroplast Clp protease complex in tobacco. *Journal of experimental botany* 68, 2199-2218
39. Tzvetkova-Chevolleau, T., Hutin, C., Noel, L. D., Goforth, R., Carde, J. P., Caffarri, S., Sinning, I., Groves, M., Teulon, J. M., Hoffman, N. E., Henry, R., Havaux, M., and Nussaume, L. (2007) Canonical signal recognition particle components can be bypassed for posttranslational protein targeting in chloroplasts. *The Plant cell* 19, 1635-1648
40. Dunkley, T. P., Hester, S., Shadforth, I. P., Runions, J., Weimar, T., Hanton, S. L., Griffin, J. L., Bessant, C., Brandizzi, F., Hawes, C., Watson, R. B., Dupree, P., and Lilley, K. S. (2006) Mapping the Arabidopsis organelle proteome. *Proceedings of the National Academy of Sciences of the United States of America* 103, 6518-6523
41. Fourrier, N., Bedard, J., Lopez-Juez, E., Barbrook, A., Bowyer, J., Jarvis, P., Warren, G., and Thorlby, G. (2008) A role for SENSITIVE TO FREEZING2 in protecting chloroplasts against freeze-induced damage in Arabidopsis. *The Plant journal : for cell and molecular biology* 55, 734-745
42. Maeda, H., Sage, T. L., Isaac, G., Welti, R., and Dellapenna, D. (2008) Tocopherols modulate extraplastidic polyunsaturated fatty acid metabolism in Arabidopsis at low temperature. *The Plant cell* 20, 452-470
43. Friso, G., Giacomelli, L., Ytterberg, A. J., Peltier, J. B., Rudella, A., Sun, Q., and Wijk, K. J. (2004) In-depth analysis of the thylakoid membrane proteome of Arabidopsis thaliana chloroplasts: new proteins, new functions, and a plastid proteome database. *The Plant cell* 16, 478-499
44. Ren, B., Chen, Q., Hong, S., Zhao, W., Feng, J., Feng, H., and Zuo, J. (2013) The Arabidopsis eukaryotic translation initiation factor eIF5A-2 regulates root protoxylem development by modulating cytokinin signaling. *The Plant cell* 25, 3841-3857
45. Teng, Y. B., Ma, X. X., He, Y. X., Jiang, Y. L., Du, J., Xiang, C., Chen, Y., and Zhou, C. Z. (2009) Crystal structure of Arabidopsis translation initiation factor eIF-5A2. *Proteins* 77, 736-740
46. Breuers, F. K., Brautigam, A., Geimer, S., Welzel, U. Y., Stefano, G., Renna, L., Brandizzi, F., and Weber, A. P. (2012) Dynamic Remodeling of the Plastid Envelope Membranes - A Tool for Chloroplast Envelope in vivo Localizations. *Frontiers in plant science* 3, 7
47. Maughan, S. C., Pasternak, M., Cairns, N., Kiddle, G., Brach, T., Jarvis, R., Haas, F., Nieuwland, J., Lim, B., Muller, C., Salcedo-Sora, E., Kruse, C., Orsel, M., Hell, R., Miller, A. J., Bray, P., Foyer, C. H., Murray, J. A., Meyer, A. J., and Cobbett, C. S. (2010) Plant homologs of the Plasmodium falciparum chloroquine-resistance transporter, PfCRT, are required for glutathione homeostasis and stress responses. *Proceedings of the National Academy of Sciences of the United States of America* 107, 2331-2336
48. Furumoto, T., Yamaguchi, T., Ohshima-Ichie, Y., Nakamura, M., Tsuchida-Iwata, Y., Shimamura, M., Ohnishi, J., Hata, S., Gowik, U., Westhoff, P., Brautigam, A., Weber, A. P., and Izui, K. (2011) A plastidial sodium-dependent pyruvate transporter. *Nature* 476, 472-475

49. Li, N., Gugel, I. L., Giavalisco, P., Zeisler, V., Schreiber, L., Soll, J., and Philippar, K. (2015) FAX1, a novel membrane protein mediating plastid fatty acid export. *PLoS biology* 13, e1002053
50. Agne, B., Kohler, D., and Baginsky, S. (2017) Protein import-independent functions of Tic56, a component of the 1-MDa translocase at the inner chloroplast envelope membrane. *Plant signaling & behavior* 12, e1284726
51. Bolter, B., and Soll, J. (2017) Ycf1/Tic214 Is Not Essential for the Accumulation of Plastid Proteins. *Molecular plant* 10, 219-221
52. Kikuchi, S., Bedard, J., Hirano, M., Hirabayashi, Y., Oishi, M., Imai, M., Takase, M., Ide, T., and Nakai, M. (2013) Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science (New York, N.Y.)* 339, 571-574
53. Nakai, M. (2015) The TIC complex uncovered: The alternative view on the molecular mechanism of protein translocation across the inner envelope membrane of chloroplasts. *Biochimica et biophysica acta* 1847, 957-967
54. Vom Dorp, K., Holzl, G., Plohmann, C., Eisenhut, M., Abraham, M., and Weber, A. P. (2015) Remobilization of Phytol from Chlorophyll Degradation Is Essential for Tocopherol Synthesis and Growth of Arabidopsis. 27, 2846-2859
55. Curien, G., Giustini, C., Montillet, J. L., Mas, Y. M. S., Cobessi, D., Ferrer, J. L., Matringe, M., Grechkin, A., and Rolland, N. (2016) The chloroplast membrane associated ceQORH putative quinone oxidoreductase reduces long-chain, stress-related oxidized lipids. *Phytochemistry* 122, 45-55
56. Mas, Y. M. S., Curien, G., Giustini, C., Rolland, N., Ferrer, J. L., and Cobessi, D. (2017) Crystal Structure of the Chloroplastic Oxoene Reductase ceQORH from Arabidopsis thaliana. *Frontiers in plant science* 8, 329
57. Li, Y., and Martin, J. R. (2017) Identification of Putative Substrates of SEC2, a Chloroplast Inner Envelope Translocase. 173, 2121-2137
58. Li, Y., Singhal, R., Taylor, I. W., McMinn, P. H., Chua, X. Y., Cline, K., and Fernandez, D. E. (2015) The Sec2 translocase of the chloroplast inner envelope contains a unique and dedicated SECE2 component. *The Plant journal : for cell and molecular biology* 84, 647-658
59. Skalitzky, C. A., Martin, J. R., Harwood, J. H., Beirne, J. J., Adamczyk, B. J., Heck, G. R., Cline, K., and Fernandez, D. E. (2011) Plastids contain a second sec translocase system with essential functions. *Plant physiology* 155, 354-369
60. Rolland, N., Dorne, A. J., Amoroso, G., Sultemeyer, D. F., Joyard, J., and Rochaix, J. D. (1997) Disruption of the plastid ycf10 open reading frame affects uptake of inorganic carbon in the chloroplast of Chlamydomonas. *The EMBO journal* 16, 6713-6726
61. Stenzel, I., Hause, B., Miersch, O., Kurz, T., Maucher, H., Weichert, H., Ziegler, J., Feussner, I., and Wasternack, C. (2003) Jasmonate biosynthesis and the allene oxide cyclase family of Arabidopsis thaliana. *Plant molecular biology* 51, 895-911
62. Schaller, F., Zerbe, P., Reinbothe, S., Reinbothe, C., Hofmann, E., and Pollmann, S. (2008) The allene oxide cyclase family of Arabidopsis thaliana: localization and cyclization. *The FEBS journal* 275, 2428-2441

63. Armbruster, U., Carrillo, L. R., Venema, K., Pavlovic, L., Schmidtmann, E., Kornfeld, A., Jahns, P., Berry, J. A., Kramer, D. M., and Jonikas, M. C. (2014) Ion antiport accelerates photosynthetic acclimation in fluctuating light environments. *Nature communications* 5, 5439
64. Aranda-Sicilia, M. N., Cagnac, O., Chanroj, S., Sze, H., Rodriguez-Rosales, M. P., and Venema, K. (2012) Arabidopsis KEA2, a homolog of bacterial KefC, encodes a K(+)/H(+) antiporter with a chloroplast transit peptide. *Biochimica et biophysica acta* 1818, 2362-2371
65. Zheng, S., Pan, T., Fan, L., and Qiu, Q. S. (2013) A novel AtKEA gene family, homolog of bacterial K+/H+ antiporters, plays potential roles in K+ homeostasis and osmotic adjustment in Arabidopsis. *PloS one* 8, e81463
66. Miyaji, T., Kuromori, T., Takeuchi, Y., Yamaji, N., Yokosho, K., Shimazawa, A., Sugimoto, E., Omote, H., Ma, J. F., Shinozaki, K., and Moriyama, Y. (2015) AtPHT4;4 is a chloroplast-localized ascorbate transporter in Arabidopsis. *Nature communications* 6, 5928
67. Nakai, M. (2015) YCF1: A Green TIC: Response to the de Vries et al. Commentary. *The Plant cell* 27, 1834-1838
68. Mizoi, J., Nakamura, M., and Nishida, I. (2006) Defects in CTP:PHOSPHORYLETHANOLAMINE CYTIDYLYLTRANSFERASE affect embryonic and postembryonic development in Arabidopsis. *The Plant cell* 18, 3370-3385
69. Lee, C. P., Maksaev, G., Jensen, G. S., Murcha, M. W., Wilson, M. E., Fricker, M., Hell, R., Haswell, E. S., Millar, A. H., and Sweetlove, L. J. (2016) MSL1 is a mechanosensitive ion channel that dissipates mitochondrial membrane potential and maintains redox homeostasis in mitochondria during abiotic stress. *The Plant journal : for cell and molecular biology* 88, 809-825
70. Schaedler, T. A., Thornton, J. D., Kruse, I., Schwarzlander, M., Meyer, A. J., van Veen, H. W., and Balk, J. (2014) A conserved mitochondrial ATP-binding cassette transporter exports glutathione polysulfide for cytosolic metal cofactor assembly. *The Journal of biological chemistry* 289, 23264-23274
71. Ma, B., Qian, D., Nan, Q., Tan, C., An, L., and Xiang, Y. (2012) Arabidopsis vacuolar H+-ATPase (V-ATPase) B subunits are involved in actin cytoskeleton remodeling via binding to, bundling, and stabilizing F-actin. *The Journal of biological chemistry* 287, 19008-19017
72. Jouhet, J., and Gray, J. C. (2009) Is chloroplast import of photosynthesis proteins facilitated by an actin-TOC-TIC-VIPP1 complex? *Plant signaling & behavior* 4, 986-988
73. Jarvis, P. (2008) Targeting of nucleus-encoded proteins to chloroplasts in plants. *New Phytologist* 179, 257-285
74. Block, M. A., and Jouhet, J. (2015) Lipid trafficking at endoplasmic reticulum-chloroplast membrane contact sites. *Current opinion in cell biology* 35, 21-29
75. Hsueh, Y. C., Ehmann, C., Flinner, N., Ladig, R., and Schleiff, E. (2017) The plastid outer membrane localized LPTD1 is important for glycerolipid remodelling under phosphate starvation. 40, 1643-1657
76. Reichel, M., and Liao, Y. (2016) In Planta Determination of the mRNA-Binding Proteome of Arabidopsis Etiolated Seedlings. 28, 2435-2452

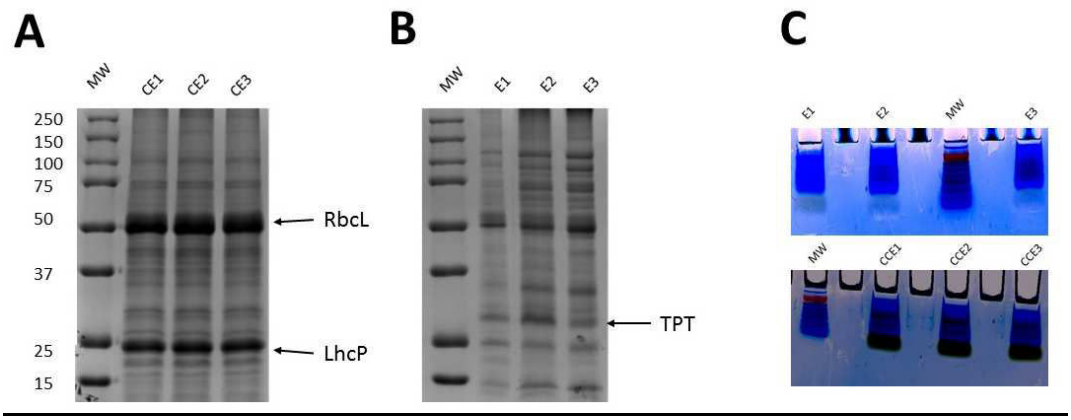
77. Gold, V. A., Chroscicki, P., Bragoszewski, P., and Chacinska, A. (2017) Visualization of cytosolic ribosomes on the surface of mitochondria by electron cryo-tomography. *EMBO reports* 18, 1786-1800
78. Blee, E., and Joyard, J. (1996) Envelope Membranes from Spinach Chloroplasts Are a Site of Metabolism of Fatty Acid Hydroperoxides. *Plant physiology* 110, 445-454
79. Froehlich, J. E., Itoh, A., and Howe, G. A. (2001) Tomato allene oxide synthase and fatty acid hydroperoxide lyase, two cytochrome P450s involved in oxylipin metabolism, are targeted to different membranes of chloroplast envelope. *Plant physiology* 125, 306-317
80. Lisenbee, C. S., Lingard, M. J., and Trelease, R. N. (2005) Arabidopsis peroxisomes possess functionally redundant membrane and matrix isoforms of monodehydroascorbate reductase. *The Plant journal : for cell and molecular biology* 43, 900-914
81. Gao, H., and Metz, J. (2016) In Vivo Quantification of Peroxisome Tethering to Chloroplasts in Tobacco Epidermal Cells Using Optical Tweezers. 170, 263-272
82. Hooper, C. M., Tanz, S. K., Castleden, I. R., Vacher, M. A., Small, I. D., and Millar, A. H. (2014) SUBAcon: a consensus algorithm for unifying the subcellular localization data of the Arabidopsis proteome. *Bioinformatics (Oxford, England)* 30, 3356-3364
83. Tomizioli, M., Lazar, C., Brugiere, S., Burger, T., Salvi, D., Gatto, L., Moyet, L., Breckels, L. M., Hesse, A. M., Lilley, K. S., Seigneurin-Berny, D., Finazzi, G., Rolland, N., and Ferro, M. (2014) Deciphering thylakoid sub-compartments using a mass spectrometry-based approach. *Molecular & cellular proteomics : MCP* 13, 2147-2167

FOOTNOTES

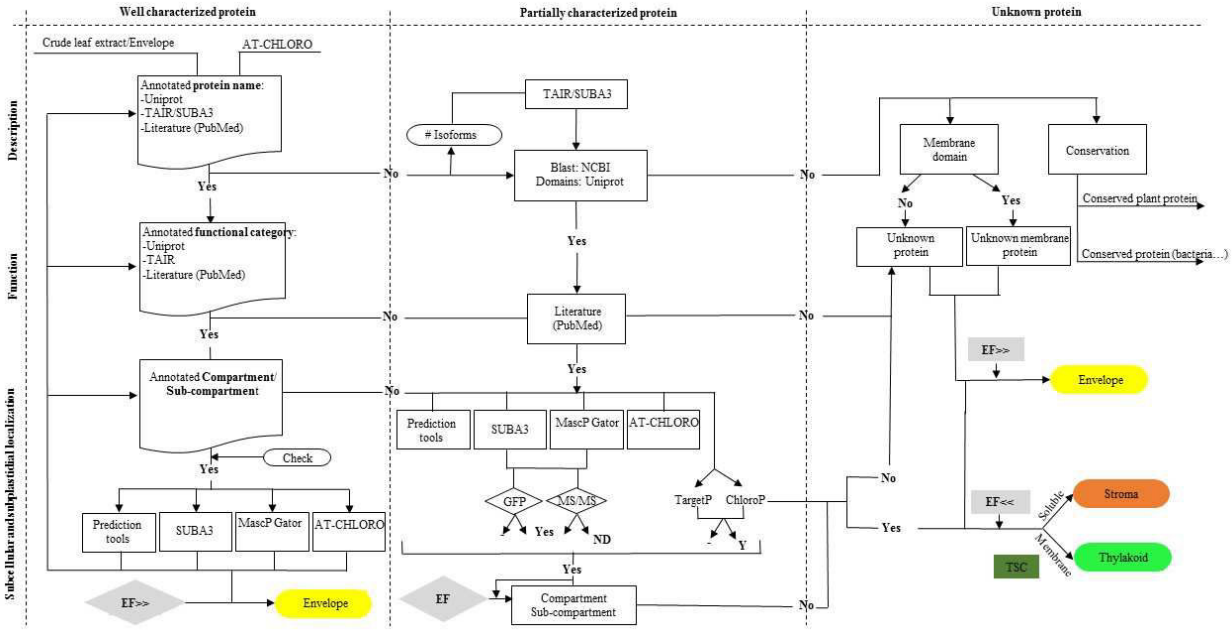
We would like to acknowledge the ANR project ANR-15-IDEX-02. This work was supported by a joint PhD fellowship to IB from the INRA Plant Biology and Breeding Division, from the Labex GRAL (ANR-10-LABX-49-01), the “Investissement d’Avenir Infrastructures Nationales en Biologie et Santé” program (ProFI project, ANR-10-INBS-08), CNRS and INRA, and access to the BIG facilities (confocal microscopy). We would also like to thank Dr. Laurent Nussaume (CEA Cadarache) for antibodies against cpFtsY, Prof. Nathalie Leborgne-Castel (INRA/Université de Bourgogne) for antibodies against bobTIP, Dr. Olivier Vallon (IBPC Paris) for antibodies against LHCP and Dr. Renaud Dumas (LPCV, Grenoble) for antibodies against KARI.

SUPPLEMENTAL DATA

Supplemental Figures

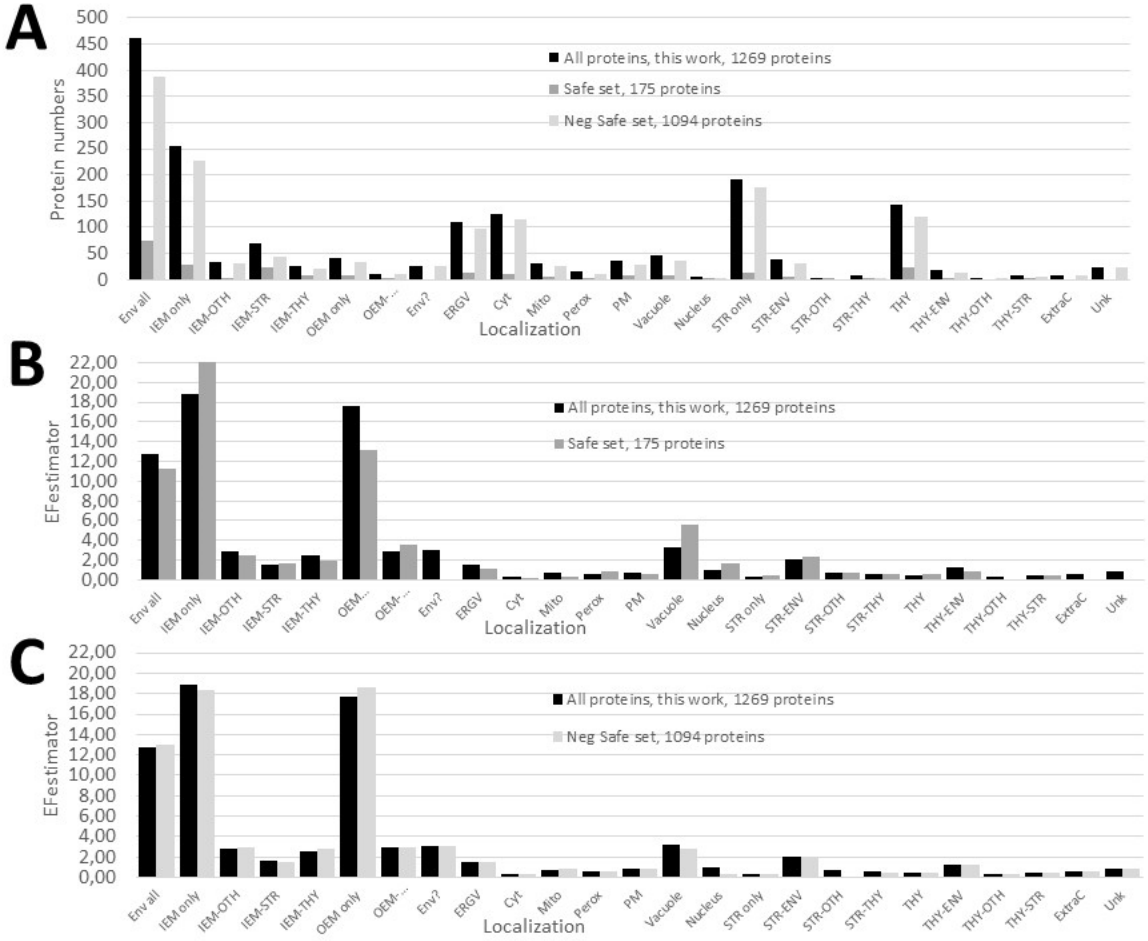


Supplemental Fig. S1: Similar protein profile over independent crude cell extracts and purified Arabidopsis chloroplast envelope fractions. The migration of the molecular weight standard (MW) is indicated on the left of each gel. Crude cell extract (CCE) and purified chloroplast envelope (E) were prepared in triplicate as respectively indicated in lanes CCE1, CCE2, CCE3 and E1, E2, and E3. Proteins (15 µg of each sample) were separated on a 12% SDS-PAGE further stained with Coomassie blue staining in order to assess reproducibility and homogeneity over triplicate. RBCL: Large subunit of Rubisco (marker from the stroma), TPT: Phosphate-triose-phosphate translocator (envelope marker) and LHCP: Light harvesting complex protein (thylakoid marker) were detected in each sample. C) Concentration of each protein sample (10 µg) at the top of NuPAGE 4-12% before MS analysis.



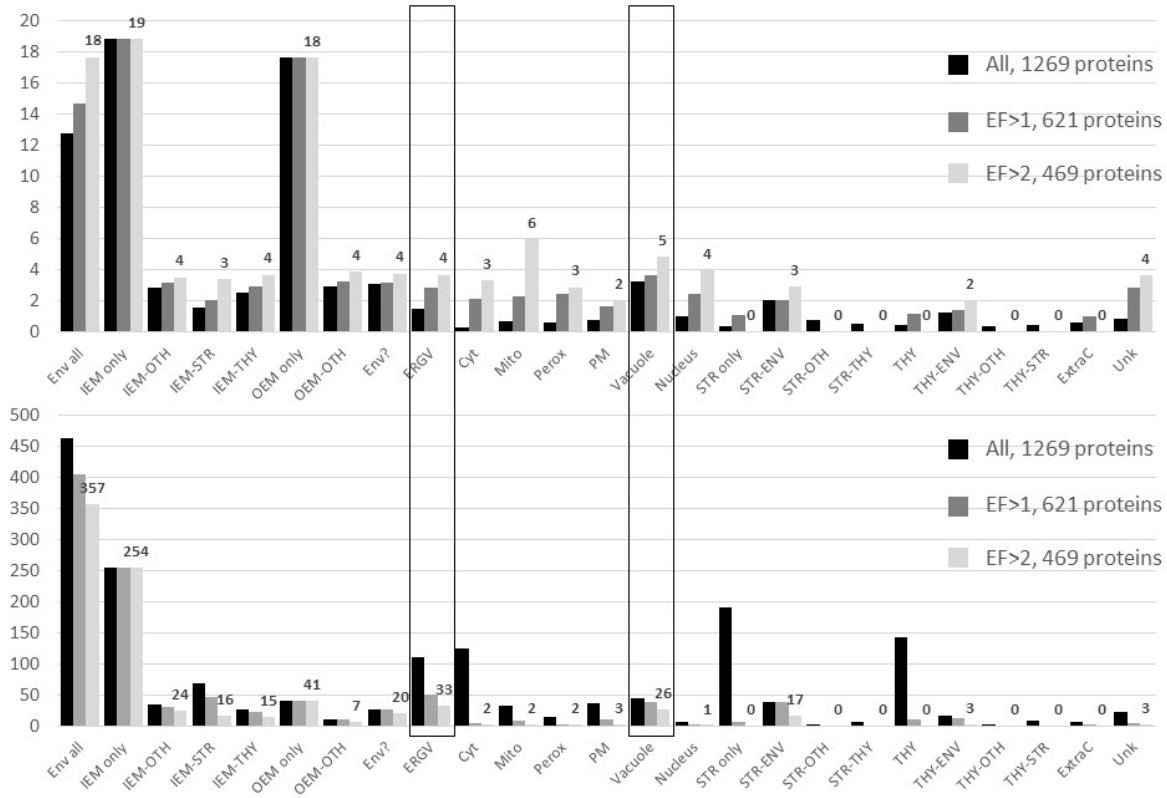
Supplemental Fig. S2: Overview of manual annotation strategy. Subcellular and subplastidial localization,

description and functional annotation were manually performed using several databases and prediction tools. EF, the ratio of the sum of weighted spectral count (WSC) over each triplicate (EF estimator) was used to estimate the enrichment of each protein in the envelope fraction (E) compared to the total extract (CCE) and therefore to differentiate genuine chloroplast envelope proteins from contaminants.



Supplemental Fig. S3: Validation of the EFestimator to associate the 1269 proteins detected in purified envelope fractions to specific subcellular and subplastidial localizations. A. Numbers of proteins attributed to the various subplastidial and subcellular localizations. **B.** Average enrichment (EFestimator) of proteins attributed to the various subplastidial and subcellular localizations for a "safe set" of 175 proteins. **Black columns** indicate the all **1269 proteins** identified during this work in purified envelope fractions (see supplemental table S5, "This work"). **Dark gray columns** indicate the "**Safe set**", *i.e.* the 175 proteins whose localization was predicted from the consensus of all prediction tools in SUBA3 (SUBA.con) and confirmed using expression of GFP fusions *in planta*. Note that plastid-encoded or mitochondria-encoded proteins were also considered as "Safe" in their respective cell compartment and that the 14 remaining proteins whose subplastidial location could not be deduced from AT_CHLORO were also removed from this "Safe set". **A and C. Light gray columns** indicate the "**Negative of the Safe set**", *i.e.* the 1094 proteins (1269 – 175) whose localization was not strongly supported by previous data. Note that in both the "**Safe set**" and the "**Negative of the Safe set**", the enrichment of predicted envelope proteins is far above the ones of proteins associated to other plastid or cell compartments. Note the relatively lower EF values of proteins shared between envelope and other plastid compartments, and the surprisingly high EF of vacuolar proteins. IEM; Inner envelope

membrane, OEM; outer envelope membrane, Env?; Envelope candidates; ERGV: endoplasmic reticulum/Golgi, Cyt; cytosol, Mito; mitochondria, Perox; peroxisome, PM; plasma membrane, Str; stroma, Thy; thylakoid, ExtraC; extracellular, OTH; other, Unk; unknown and unpredictable localization.



Supplemental Fig. S4: Impact of enhancing of EF thresholds on average EF values of proteins from plastid and other cell compartments. Note the high average EF values of the 33 remaining ER/Golgi and the 26 remaining vacuolar proteins (black squares) when the EF threshold is increased to 2.

Supplemental Tables

Supplemental Table S1: MS/MS identifications from analysis of Arabidopsis crude cell extracts (CCE1 to 3) and purified chloroplast envelope fractions (Env1 to 3). The lists of proteins and peptides result from automated validation pipeline as described in materials and methods. The corresponding mass spectrometry data are available at the PRIDE repository (23) with the dataset identifier PXD010545.

Protein sets: list of all protein sets (2964 including 16 contaminants) identified in the whole experiment. Proteins retained must have at least one SSC in one replicate. "sameset" and "subset" accessions indicate proteins that share the same set of peptides, or a subset of the peptides, respectively. **Protein matches**: list of all (representative, sameset and subset) proteins belonging to protein sets (5562 including 136 contaminants). **Best PSMs**: list of best (highest score) Peptide Spectrum Matches of peptides identified. Unique peptides (15460) are defined by a unique combination of primary sequences + modifications and positions.

Supplemental Table S2: Oligonucleotides used to generate constructs (GFP and CFP fusions) to validate subcellular and subplastidial localization of TSP9, SFR2, UP1, eIF5A and VTE1 proteins (see Fig. 10).

A. Oligonucleotides primers used for Gateway cloning strategy

Number	Accession	Nickname	Oligonucleotides	Sequence 5'-3'
1	AT3G47070	TSP9	TSP9 TOPO fwd	*ccac*ATGGTTTCTTCGCTTCTTATG
			TSP9 TOPO rev	TTTCTTGAAGAGGCTTCCTAAG
2	AT3G06510	SFR2	SFR2 TOPO fwd	*ccac*ATGGAATTATTCGCATTGTTA
			SFR2 TOPO rev	GTCAAAGGGTGAGGCTAA

B. Oligonucleotides primers used for classical cloning strategy

	Accession	Nickname	Oligonucleotides	Sequence 5'-3'
3	AT1G11320	UP1	UP1 SalI fwd	GTCGACATGGACCCAATTGCTTCGG
			UP1 NcoI rev	CCATGGACAGCGACCAGTGAGACTTTAG
4	AT1G26630	eIF-5A	eIF5A SalI fwd	GTCGACATGTCTGACGACGAGCACC
			eIF5A BspHI rev	TCATGAACTTGCCACCACCAACTTCC
5	AT4G32770	VTE1	VTE1 SalI fwd	TCTGTCGACATGGAGATACGGAGCTTG
			VTE1 NcoI rev	ATCCCATGGACAGACCCGGTGGCTTG

Supplemental Table S3: List of proteins (2480) identified in crude cell extracts (2222) and purified *Arabidopsis* chloroplast envelope (1269) triplicates. Crude cell extract (CCE1 to 3) and purified chloroplast envelope fractions (E1 to 3) were prepared in triplicates as indicated in materials and methods. Composition of these triplicates were analyzed by MS. Accession: AGI numbers. Description: annotations extracted from databases. Sequence specific: number of sequences not shared by another protein set. Pep: number of identified peptides. SC: spectral counts. SSC: specific spectral counts. WSC: weighted spectral counts. WSC-CCE: specific spectral counts in crude cell extracts. WSC-Env: specific spectral counts in envelope fractions. Norm-WSC-CCE, normalized WSC over the sum of replicate CCE3. Norm-WSC-Env, normalized WSC over the sum of replicate Env3. Location SUBAcon: known or predicted subcellular localization extracted from SUBA3 (see (82)) to design Fig. 2. Description SUBA3: protein description extracted from the SUBA3 database.

Supplemental Table S4. Predicted subcellular localization of proteins identified in purified envelope fractions and crude cell extracts according to the SUBA3 database (SUBAcon, see (82)). Numbers (**Upper part)** and % (**Lower part)** of proteins identified in various cell compartments. **A** Note that only 10% (257) of the 2479* detected proteins were only detected in purified envelope fractions (not in the crude cell extract). Note that 77% (787) of the 1017 predicted plastid proteins were identified in purified envelope fractions. Conversely, CCE contains more proteins predicted to be localized in other cell compartments. Surprisingly, purified envelope fractions contain 53 (69%) of the 77 detected proteins that are predicted to be vacuolar components.
B. Graphical representation of data from **A**. Note that only vacuolar proteins are almost as abundant in purified envelope fractions when compared to CCE. **C**. Note that 62% (787) of the 1269 proteins identified in purified envelope fractions are predicted to be plastid proteins (only 38% for proteins detected in CCE). Conversely, CCE contains more proteins predicted to be localized in other cell compartments. Again, note that predicted vacuolar proteins were enriched in purified envelope fractions (4%) when compared to their relative abundance in CCE (3%). * Note that one protein (AT5G12170.2) has no SUBACon

Supplemental Table S5: List of proteins identified in purified *Arabidopsis* chloroplast envelope triplicates. All proteins: manual annotation of the 1364 non-redundant proteins identified in purified envelope fractions during this work and in Ferro *et al*., 2010 (12). **This work:** list of the 1269 proteins identified in purified envelope fractions during this work. **Only EF>1**: list of the 641 proteins identified in purified envelope fractions during this work, with a WSC_Env/WSC_CCE ratio > 1. **Only EF>2**: list of the 469 proteins identified in purified envelope fractions during this work, with a WSC_Env/WSC_CCE ratio > 2.
Protein group (AGI numbers), Protein number according to Sup Data AT_CHLORO (12), accession number (AGI numbers), accession in UNIPROT, classification in MapManBin (-sept 2017), Curated function, curated description. Simplified location AT_CHLORO: subplastidial localization in AT_CHLORO (12) (nd: not detected in Ferro *et al*., 2010 (12)). Simplified location This work: revised subplastidial localization according to new manual annotation. WSC_CCE: specific spectral counts in crude cell extracts. WSC_Env: specific spectral counts in envelope fractions. WSC_Env/WSC_CCE: ratio of weighted spectral counts in envelope fractions to weighted spectral counts in crude cell extracts. Env %, STR % and THY % refer to detection of the protein in the three main chloroplast compartments in AT_CHLORO (12). TSC refer to detection of the protein in thylakoid sub-compartments (83). TargetP: prediction of subcellular localization (28). cTP: prediction of chloroplast localization using ChloroP (27). cTP (loc): predicted maturation site according to ChloroP (27). Env proteome: detection of the protein by Simm *et al*., 2014 (14). Experimental evidence MASCP Gator / SUBA3: numbers of detection of the protein using MS-based approaches in specific cell compartments. Predicted MASCP Gator & SUBA: numbers of independent tools predicting specific subcellular localizations. Location GFP (MASCP GAtor & SUBA): experimental evidences for the specific subcellular localization of a protein. Location SUBAcon: consensus subcellular localization of the protein extracted from SUBA3 (82). Sequence Specific in ENV: number of specific sequences detected in purified envelope fractions (or only in CCE for "All proteins" since some proteins detected in Ferro *et al*., 2010 (12) were only detected in CCE during this work). Crude cell extract (CCE_T1 to T3) and purified chloroplast envelope fractions (Env_E1 to E3) are triplicates as indicated in materials and methods. Norm-WSC, normalized Weighted Spectral Count. SWSC, sum of the Norm-WSC in each sample. Imp_SWC, imputated SWSC *i.e.* when this sum equals zero in the CCE fraction, an extra count of 1 was added to both sum SWSC_Env and SWSC_CCE.

Supplemental Table S6: **Overlap of the 1269 proteins identified in purified envelope fractions with the list of 700 proteins previously identified in the envelope fractions and present in the AT_CHLORO database** (12) (see supplemental table S10 of (12)).

Supplemental Table S7: **Comparison of manual annotation (this work) with SUBAcon** ((82))

Chapitre III

Contrôle de l'adressage des protéines au chloroplaste

Article 2

Calmodulin 5 is a sensor for environmental control of organellar protein import involving vesicular trafficking

Lucas Moyet[1]ǂ, Daniel Salvi[1]ǂ, Imen Bouchnak[1], Stéphane Miras[1†], Laura Perrot[1†], Daphné Seigneurin-Berny[1†], Marcel Kuntz[1], Norbert Rolland[1]*

[1] Laboratoire de Physiologie Cellulaire & Végétale, Univ. Grenoble Alpes, INRA, CNRS, CEA, BIG-LPCV, 17 rue des Martyrs, 38000 Grenoble, France. ǂ These authors contributed equally to this work. † Present addresses: Kelly Scientific, Rue du Rhône 114, 1204 Geneva, Switzerland (S.M.); Génétique et Amélioration des Fruits et Légumes, INRA, 67 allée des Chênes, 84140 Montfavet, France (L.P.); Institut pour l'Avancée des Biosciences, Univ. Grenoble Alpes, CNRS, INSERM, La Tronche, France (D.S.-B.). *Correspondence and requests for materials should be addressed to N.R. (norbert.rolland@cea.fr).

Abstract

Most proteins from organelles such as plastids are imported from the cytosol via a specific machinery (Toc-Tic), as precursors harboring a cleavable transit peptide (cTP). In contrast, non-canonical proteins lacking a cTP use Toc-Tic-independent pathways to reach these organelles. However, nothing is known about actors and cellular events regulating non-conventional trafficking events. Here, we show that non-canonical protein trafficking involves the vesicular pathway, indicating unsuspected similarities with glycoprotein trafficking to plastids. While calmodulins are well known to regulate catalytic properties of their targets, we now show that a plasma membrane calmodulin (CaM5) also regulates non-conventional trafficking. Our model incorporates observations resulting from the loss of CaM-binding properties of targeted proteins; *cam5* gene knock-out; and environmental constraints affecting CaM5 stability. CAM5 is able to trap plastid proteins to the plasma membrane, while its response to environmental constraints spatially and temporally uncouples retention of these chloroplast proteins allowing them to reach plastids.

Introduction

Cell compartmentalization is an essential process by which eukaryotic cells separate and control biological processes. Plant metabolism, development and adaptation to environmental stresses rely on highly dynamic processes involving the targeting of thousands of proteins to specific cell organelles, but also to secretory routes from the endoplasmic reticulum (ER) via the Golgi[1], allowing proteins and glycans to reach the plasma membrane (PM) (the interface between the cell and its environment) and the cell wall.

Protein targeting to the chloroplast through the classical Toc-Tic pathway has been largely decrypted during these last decades (for reviews, see refs [2,3,4]). Recent considerations related to this classical pathway concentrate on the composition and respective roles of several Toc and Tic complexes (*e.g.*, refs [5,6]). However, some underlying mechanisms regulating protein targeting to plastids remain unknown. It is, however, largely accepted that import pathways that target proteins into and across the plastid envelope are critical for plant development by regulating the response to physiological and metabolic changes within the cell[2,3].

In addition, chloroplast proteins targeted to the organelle through Toc-Tic-independent pathways have also been described (for reviews, see refs [1,2,3,6]). There is a growing literature on the existence of ER to plastids communication in plants[1]. Indeed, a few glycoproteins were shown to traffic along the secretory pathway through the ER and Golgi before being targeted to plastids (a chloroplast carbonic anhydrase[7], a rice α-amylase[8,9], a manganese superoxide dismutase[10] and nucleotide pyrophosphatase/phosphodiesterases[11,12]). Recently, trans-Golgi compartments were shown to participate in Golgi-to-plastid trafficking of these nucleotide pyrophosphatase/phosphodiesterases[13]. This vesicular pathway is believed to originate from the early stages of chloroplast evolution[14,15].

In addition to glycoproteins, a few chloroplast proteins (*e.g.*, ceQORH, Tic32) containing non-canonical transit sequences have been described[16,17,18]. While the majority of canonical chloroplast preproteins feature an *N*-terminal cleavable transit peptide (cTP), which is necessary for targeting and recognition at the chloroplast surface by Toc receptors, these non-canonical proteins lack predictable and cleavable transit peptides and are imported into chloroplasts through Toc-Tic-independent pathways[16,17,18]. These alternative pathways are depicted in review articles targeting plastid biogenesis (*e.g.*, ref [3]) as fully independent routes, and recent investigations by experts in the field have confirmed initial observations (*e.g.*, ref [19]).

The diversity of protein targeting pathways to plastids and their regulation in response to developmental and metabolic status is a key issue in the regulation of cellular function in plants. However, the mechanisms allowing trafficking to plastids via these non-conventional pathways remain largely unknown. Additionally, we know comparatively little about the proteins regulating these dynamic trafficking events when compared to the classical Toc-Tic dependent pathway. To address this problem at the cellular and molecular levels, we used the ceQORH protein from *Arabidopsis thaliana* as an experimental system for a non-canonical chloroplast protein. This protein reduces long-chain, stress-related oxidized lipids that are spontaneously produced in the chloroplast from the unstable allene oxide formed in the biochemical pathway leading to 12-oxo-phytodienoic acid, a precursor of the defence hormone jasmonate[20,21]. It was demonstrated to reach the chloroplast while lacking a canonical (predictable and cleavable) cTP[16] via a non-conventional targeting pathway, independent of the Toc machinery[17,19]. ceQORH was originally found in the chloroplast envelope, where it represents 2,5% of envelope proteins[22]. Here, we first demonstrate that this "non-canonical" chloroplast protein, en route to the chloroplast, follows the vesicular pathway, thus demonstrating unsuspected similarities between the "vesicular" (glycoproteins) and the "non-canonical" (lack of cTP) alternative pathways for trafficking to the chloroplast. Then, we identify and characterize a key player regulating this vesicular traffic: the calmodulin isoform CaM5. Surprisingly, CaM5, which is located in the plasma membrane (PM) of specific leaf cells, traps ceQORH at the level of this membrane. Environmental stresses can then release this trapped component allowing it to reach plastids.

Results

In epidermal cells from Arabidopsis leaves, ceQORH shows a complex subcellular multilocation pattern. Targeted studies[16] and proteomic data[22,23] unambiguously localized ceQORH (AT4G13010) in the inner membrane of the chloroplast envelope. However, while analysing its *in vivo* subcellular location in epidermal cells, we formerly observed that this protein was not targeted to plastids[17] and was also partly located in small structures with sizes similar to those of mitochondria. Due to the relatively high frequency (over 100 known cases) of proteins dual targeted to chloroplasts and mitochondria[24], the question raised of whether ceQORH is located in the mitochondria in epidermal cells. However, we exclude this hypothesis since colocalization of ceQORH::GFP and MitoTracker fluorescence was not observed *in planta* (see Supplementary

Fig. 1c-d). Alternatively, we found an almost perfect correlation between ER and ceQORH signals (see Fig. 1a), thus strongly supporting the ER location of ceQORH in specific tissues. To confirm this observation, we used brefeldin A (BFA) to treat plant leaves expressing ceQORH::GFP. Vesicles accumulating after BFA treatment in Arabidopsis are known to result from the continual input of vesicles from both secretory and endocytic pathways[29]. Strikingly, ceQORH also accumulates in BFA bodies (see Fig. 1b and Supplementary Fig. 2).

ceQORH was also repeatedly detected in purified PM fractions using large-scale proteomic analyses targeting this membrane[25,26,27]. More recently, ceQORH was also detected in the Golgi[27] and more precisely in the trans-Golgi network (TGN) using affinity purification and a TGN marker (RABF2b/ARA7) as a bait[28]. However, in these previous studies targeting the PM or Golgi, ceQORH was considered a contaminating plastid protein. However, we now show that *in vivo* ceQORH is indeed frequently located at the periphery of plant cells, most likely in the PM (see Fig. 1a), which further complexifies its subcellular trafficking.

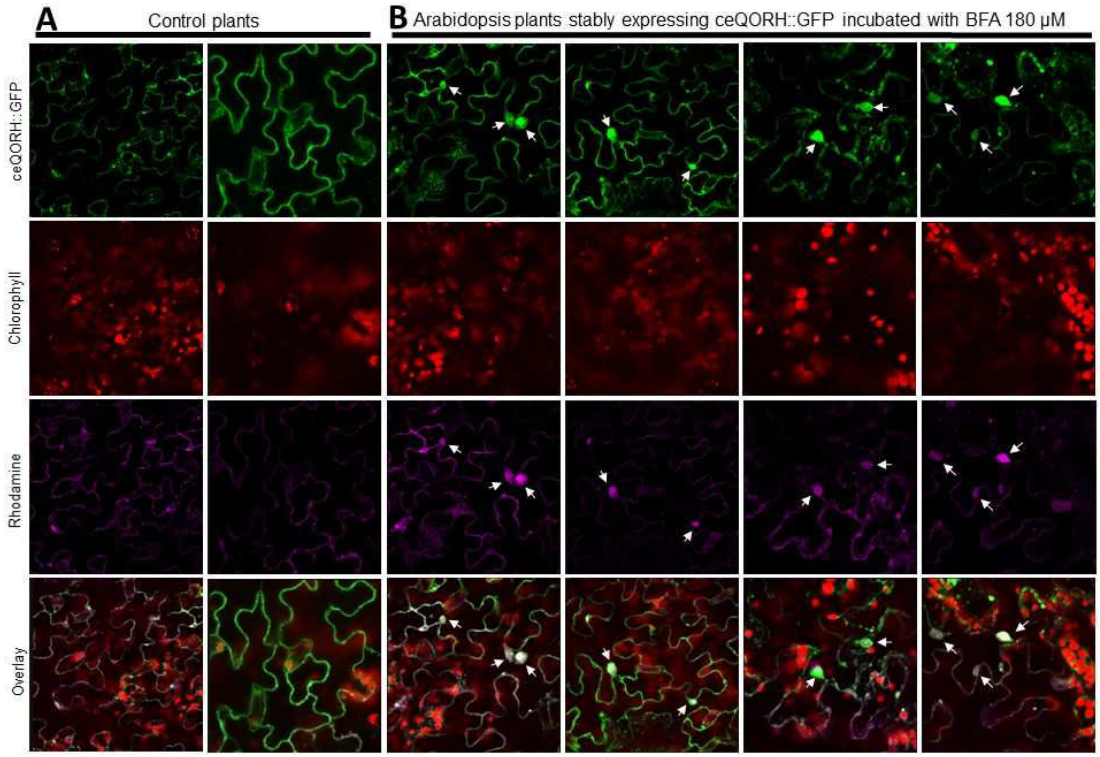


Figure 1: ceQORH accumulates in in BFA bodies in epidermal cells. Confocal microscopy was performed on plants stably expressing ceQORH::GFP (ceQORH fused to GFP). **A**. Control plants treated with rhodamine (1 µg/ml). **B**. Plants treated with both BFA (180 µM) and rhodamine (1 µg/ml). ceQORH::GFP; GFP fluorescence, Chlorophyll;

Chlorophyll autofluorescence. Rhodamine; Rhodamine fluorescence, Overlay: Overlay of all three channels. Arrows indicate BFA bodies.

Interaction of the natural plant ceQORH protein with Calmodulin. Interestingly, ceQORH is present in the list of CaM-binding chloroplast proteins after affinity purification and large-scale MS-based identification[30]. In the present study, we demonstrate the CaM-binding property of ceQORH. First, ceQORH was enriched in the fraction eluted from a calmodulin-affinity resin when compared to its level in crude cell extracts (see Supplementary Fig. 3a). Then, we compared the CaM-binding properties of the recombinant ceQORH protein produced in *E. coli* to those of the endogenous protein ecQOR (as a negative control), the closest bacterial homologue of plant ceQORH (see Supplementary Fig. 3b). As demonstrated in Supplementary Fig. 3c, while the plant ceQORH is detected in the elution fraction of CaM affinity chromatography, its bacterial homologue is only detected in the pass-through fraction. In other words, while both proteins are highly conserved (and share a similar charge and size as well as other biophysical properties; see refs [16,21]), they do not share CaM-binding properties. In addition, we previously showed that this bacterial ecQOR protein was not targeted to chloroplasts *in planta*[16]. Thus, during the evolution of this protein from bacteria to eukaryotic cells, ceQORH has acquired internal plastid targeting domain and CaM-binding properties.

The CaM-binding domain is located in the *C*-terminus of ceQORH. Using successive deletions of the ceQORH protein and expression of these constructs in *E. coli*, we were first able to demonstrate that the CaM-binding domain is located in the *C*-terminus of ceQORH (see Supplementary Fig. 4). Further targeted mutagenesis was then performed to identify essential residues for the CaM-binding properties of ceQORH (see Fig. 2). One mutated form, termed Mut2-ceQORH, in which only three residues abolishes the interaction of ceQORH with CaM (see Fig. 2b) was considered as the best tool for further approaches.

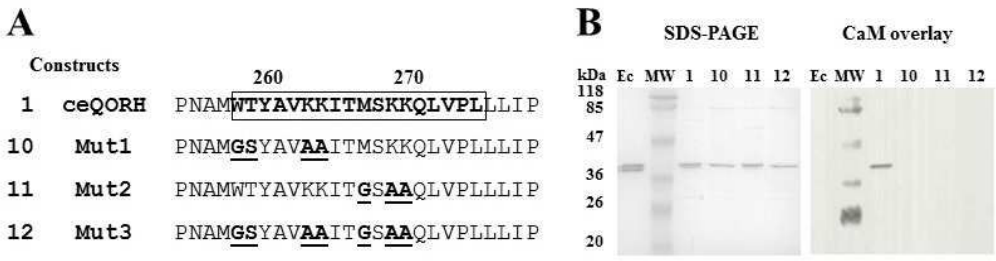


Figure 2: Targeted mutagenesis identifies essential residues required for CaM-binding properties of ceQORH. **A.** Scheme of the residues selected for targeted mutagenesis of the ceQORH sequence. The CAM-binding region in

the ceQORH sequence is deduced from successive deletions of ceQORH domains (see Supporting Figure 4). **B.** SDS-PAGE. **C**. CaM Overlay. Ec: recombinant QOR from *Escherichia coli* K12 used as a negative control, lane 1: recombinant WT form of ceQORH from *A. thaliana* used as a positive control, lane 10: mutagenized form of Arabidopsis ceQORH mutant 1, lane 11: mutagenized form of Arabidopsis ceQORH mutant 2 and lane 12, mutagenized form of Arabidopsis ceQORH mutant 3. Note that, in Mut2, mutagenesis of only three residues is required to abolish interaction of ceQORH with CaM.

CaM-binding is neither essential for ceQORH targeting to parenchymal cell chloroplasts nor required for the specific localization of ceQORH to the plastid envelope. As previously observed[17], the *C*-terminus domain of ceQORH is essential *in planta* for its specific location in the envelope membrane in mesophyll cells. This is confirmed by our observation that, while the full-length ceQORH is enriched in purified chloroplast envelope fractions, ceQORH lacking its *C*-terminus also accumulates in thylakoid membranes at a similar level (see Fig. 3a). Interestingly, this specific envelope location does not rely on the ceQORH CaM-binding domain located in this *C*-terminal region. Indeed, Mut2-ceQORH is targeted to the chloroplast, and only detected in the purified envelope fraction and not in thylakoids (see Fig. 3a), similarly to the endogenous ceQORH protein. This clearly demonstrates that the CaM-binding domain is not essential *in planta* for the targeting of ceQORH to the envelope. To examine the possibility whether ceQORH could nevertheless binding calmodulin within the envelope, we then assessed whether CaM isoforms were present in the purified envelope fractions. In good agreement with previous MS-based analyses targeting purified chloroplast envelope fractions[22], we could not detect signals associated with CaM isoforms at the envelope level. Through dilution and western blotting analyses, we further observed that while representing approximately 1/2000 (~ 10 ng CaM in 20 □g crude leaf extract) of all cell proteins (see Fig. 3b), CaM should represent less than 1/600,000 (less than 0.1 ng CaM in 60 □g envelope proteins) of all envelope proteins (see Fig. 3b). Knowing that ceQORH represents 2.5% of envelope proteins[22], a ceQORH/CaM ratio above 2000 would thus be poorly compatible with a meaningful role for CaM binding to ceQORH at the envelope level. To further elucidate the role of CAM-binding, it was necessary to identify which CaM polypeptide interacts with ceQORH *in planta* because plant cells contain 7 classical CaM isoforms and tens of CaM-like proteins[31,32].

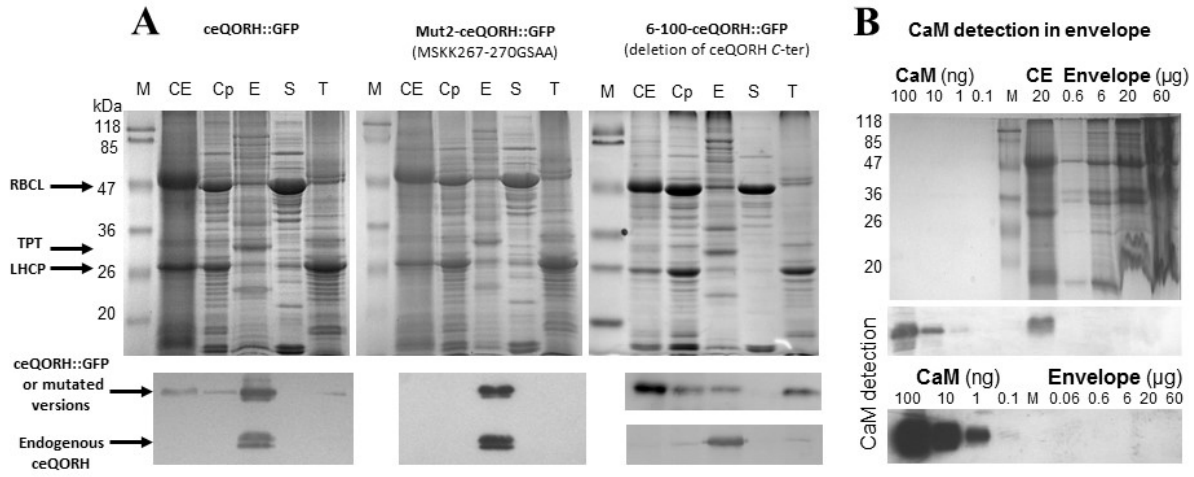


Figure 3. The ceQORH CaM-binding domain is neither essential for ceQORH targeting to parenchymal cell chloroplasts nor required for specific localization of ceQORH to the plastid envelope. A. Lack of CaM-binding domain does not affect envelope localization of ceQORH. Cell fractionation was performed on plants stably expressing ceQORH::GFP (ceQORH fused to GFP), Mut2-ceQORH::GFP (ceQORH mutant affected in CaM-binding properties fused to GFP) and 6-100-ceQORH::GFP (ceQORH lacking its 197 *C*-ter residues, including the CaM-binding domain, fused to GFP). M, prestained protein molecular weight markers, CE; Crude cell extract, Cp; Chloroplast extract. E; Envelope. S; Stroma, T; thylakoids. Note that, to limit artefacts resulting from overexpression of the various ceQORH constructs, transgenic plants, were selected for expression levels of recombinant proteins similar to endogenous ceQORH level. Each lane contains 20 µg proteins. RBCL, large subunit of RuBisCO (stroma marker). LHCP, Light Harvesting Complex Proteins (thylakoid membrane marker). Western blot analyses were performed using the antibody raised against ceQORH. **B**: **Purified chloroplast envelope fractions do not contain detectable levels of CaM**. CaM, 100 to 0.1 ng of purified recombinant CaM1 from Arabidopsis. M, prestained protein molecular weight markers, CE; Crude cell extract, Envelope; 0,6 to 60 mg envelope proteins. Note that a second CaM detections experiment is shown, using longer exposure time to improve sensitivity of ECL detection.

A high-molecular-weight CaM isoform is enriched in membrane fractions of epidermal cells compared to crude leaf extract. Having observed that ceQORH was present at the periphery of plant cells in epidermal tissues (see Fig. 1a), we then assessed the abundance of CaM isoforms in epidermal tissue compared to crude leaf extracts. As seen in Fig. 4a, the CaM signal was enriched in epidermal tissue compared to whole leaf extract. This signal in epidermal cells was of higher molecular weight (Mw; *i.e.,* above 20 kDa) than classical (shorter) CaM isoforms (*i.e.,* 16 kDa). This high-Mw signal was also enriched in the membrane fraction of epidermal cells (see Fig. 4b), while the soluble fraction contained CaM isoforms of lower Mw. Finally, this high-Mw CaM signal was 10 to 20 times more enriched in epidermal cells (~ 5 ng CaM in 4 □g crude epidermal extract, *i.e*., 1/1000 of total epidermal proteins) compared to crude leaf extract (~ 1 ng CaM in 20 □g crude leaf extract, *i.e.,* 1/20,000 of total leaf proteins) (see Fig. 4c). In screening the literature for a

membrane-bound CaM isoform of higher Mw than classical (short) CaM isoforms, we identified CaM53 from petunia (CaM53_PETY), which contains an additional *C*-terminal sequence compared to short CaM isoforms (*e.g.*, CaM1, AT5G37780) (see Fig. 4d). In addition to having an extended *C*-terminal basic domain, this long CaM isoform is also isoprenylated in specific environmental conditions[33], thus contributing to its higher Mw. Interestingly, this isoprenylation allows the protein to be anchored to the PM in leaf cells[33,34]. A close homologue of this protein is present in Arabidopsis and named CaM5 (AT2G27030.3). Compared to other CaM isoforms, this CaM5 protein contains an additional *C*-terminus sequence that is highly conserved with CaM53 from petunia (CALM_PETHY). Notably, using AT2G27030 as a query in Uniprot, only a short version (149 amino acids, Mw 16,820 Da) of CaM5 is indicated (Q682T9 or TCH1 or CALM5_ARATH). However, in TAIR, a query using the term AT2G27030 results in the gene model AT2G27030.3, which encodes a 181-amino acid protein with a Mw of 20,575.9 Da (*i.e.,* similar to CaM53 from petunia).

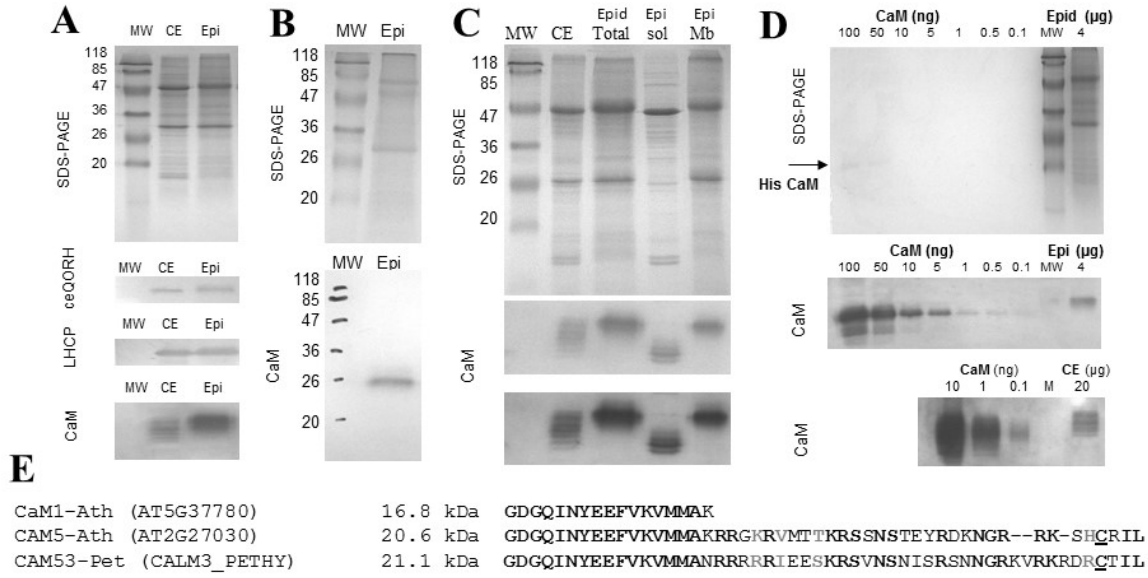


Figure 4. Membranes fractions of epidermal cells from Arabidopsis leaves are enriched in a high-molecular weight CaM isoform when compared to crude leaf extract. A. Detection of CaM isoforms in crude leaf extracts (CE) and epidermal tissue (Epi) from Arabidopsis leaves. Western-blots were performed using antibodies raised against ceQORH, LHCP and CaM-767. **B**. Mw of the CaM isoform enriched in epidermal cells is above (~25 kDa) the one expected for classical (short) CaM isoforms (~17 kDa). Mw, prestained Mw markers. Epi: 5 mg of proteins from crude epidermal extract from Arabidopsis leaves. **C**. Fractionation of membrane and soluble fractions of epidermal tissue reveal that the high molecular weight (> 20 kDa) CaM isoform is bound to membranes. Mw, prestained Mw markers. CE: 15 mg of proteins from crude leaf extracts from Arabidopsis; (Epi) Epi: 15 mg of proteins from crude epidermal extract from Arabidopsis leaves. Sol: 15 mg of soluble proteins from epidermal tissue; Mb: 15 mg of

membrane proteins from epidermal tissue. Note that a second CaM detections experiment is shown using longer exposure time to improve sensitivity of ECL detection. **D**. CaM is more abundant in epidermal cells (1/400 to 1/800 of all epidermal proteins) when compared to crude leaf extract (1/2000 to 1/20000 of all cell proteins). CaM, 100 to 0.1 ng of purified recombinant CaM1 from Arabidopsis. M, prestained protein molecular weight markers, Epi; 4 mg of proteins from crude epidermal extract, CE; 20 mg of proteins from crude leaf extract. **E**. Alignment of *C*-termini from classical short CaM isoform (e.g. CaM1) with the CaM isoform identified in the plasma membrane from petunia (CaM53-Pet) and its closest homologue in Arabidopsis (CaM5-Ath). Note that CaM5 from Arabidopsis and CaM53 from petunia contain an additional *C*-terminus sequence when compared to short CaM isoforms (CaM1). Conserved residues are bold (black for identity, and gray for similarity). Underlined C residue is the CaM53 isoprelynation site.

CaM5 and ceQORH colocalize and interact in epidermal cells. Having identified CaM5 (AT2G27030.3) as a candidate for the most abundant (and high-Mw CaM) isoform in epidermal cells, we were encouraged to pursue its analysis by confocal imaging. Indeed, ceQORH and CaM5 were shown to colocalize in epidermal cells (see Supplementary Fig. 5). We then used bimolecular fluorescence complementation (BiFC) as a method to determine the interaction of these two proteins *in planta* (see Fig. 5). Again the ceQORH mutant with affected CaM-binding properties (*i.e.,* Mut2-ceQORH; see above) was greatly helpful in performing these experiments. Indeed, while both ceQORH and Mut2-ceQORH accumulate at similar levels in plant cells (see Fig. 5b), BiFC was only detected using the WT version of ceQORH (*i.e.* containing the CaM-binding domain). Finally, it is striking that BiFC was mostly detected in the PM, *i.e*., where CaM5 and ceQORH were colocalized within epidermal cells (see Supplementary Fig. 5).

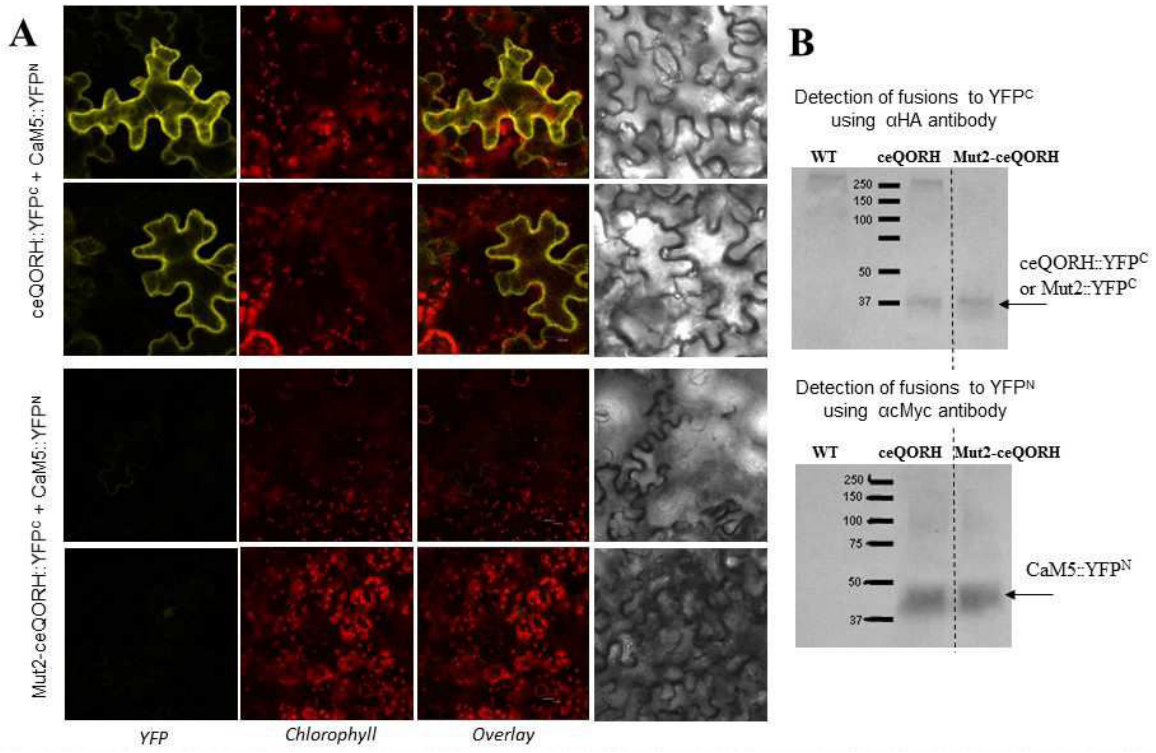


Figure 5. Identification of the cytosolic CAM isoform interacting with ceQORH: CaM5 and ceQORH interact in epidermal cells. A. Confocal imaging (**BiFC**) performed on plants transiently expressing either ceQORH::YFP^C (ceQORH fused to the *C*-terminus of YFP) or Mut2-ceQORH::YFP^C (Mut2 version of ceQORH fused to the *C*-terminus of YFP) and CaM5::YFP^N (CaM5 sequence fused to the *N*-terminus of YFP). YFP: fluorescence of YFP resulting from BiFC. Chlorophyll; Chlorophyll autofluorescence. Overlay; Overlay of the two channels. Bar equals 40 mm. **B**. Western blots analyses performed to validate expression of YFP fusions. Crude cell extracts from WT plants were included as negative controls. Note that while both ceQORH::YFP^C and Mut2-ceQORH::YFP^C accumulate at a similar level in plant cells, BiFC is only detected using the WT ceQORH version (*i.e.* containing its CaM-binding domain). Note that BiFC is mostly detected in the plasma membrane, *i.e.* where CaM5 and ceQORH were shown to colocalize within epidermal cells (see Supporting Figure 5).

ceQORH lacking its CaM-binding domain is targeted to chloroplasts in epidermal cells. Having previously demonstrated that ceQORH and CaM5 interact in epidermal cells, we examined the role of this interaction. We thus expressed Mut2-ceQORH *in planta* and analysed its subcellular localization compared to that of WT ceQORH. While ceQORH::GFP is never detected in plastids from epidermal cells (see Fig. 6c), Mut2-ceQORH was mostly detected in plastids (see Fig. 6d) at the periphery of organelles, thus most likely in their envelope membrane. This result strongly supports the idea that the CaM-binding domain of ceQORH is essential *in planta* for the alternative

localization of ceQORH in specific tissues such as epidermal cells, *i.e*. its retention in the PM by CaM5.

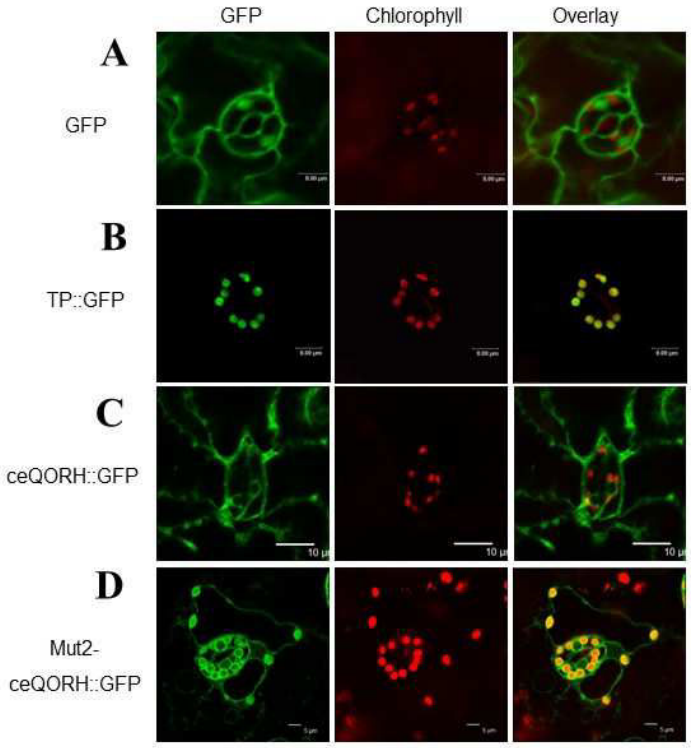


Figure 6. ceQORH lacking its CaM-binding domain is targeted to chloroplasts in epidermal cells. Confocal microscopy was performed on plants stably expressing GFP (GFP alone as a negative control of plastid localization, bar equals 8 mm), TP::GFP (Transit sequence of the small subunit of RuBisCO fused to GFP as a positive control of plastid localization, bar equals 8 mm), ceQORH::GFP (plant ceQORH fused to GFP, bar equals 10 mm) and Mut2-ceQORH::GFP (ceQORH lacking its CaM-binding domain, bar equals 5 mm). GFP; GFP fluorescence, Chlorophyll; Chlorophyll autofluorescence. Overlay: Overlay of all three channels.

The CaM-binding domain of ceQORH is not required for its ER trafficking. Having shown that ce-QORH also associates to the ER (see above), we then examined the putative role of the CaM-binding domain in the transport of ceQORH through the secretory pathway. However, similar to ceQORH, Mut2-ceQORH also accumulated in BFA bodies in BFA-treated Arabidopsis plants (see Fig. 7 and Supplementary Fig. 6), thus demonstrating the CaM-binding domain of ceQORH has no impact on its trafficking from the ER. The main difference from previously analysed plants expressing WT ceQORH (see Fig. 1) was that while accumulating in BFA bodies, Mut2-ceQORH was also detected in plastids, as shown in Fig. 7 and Supplementary Fig. 6.

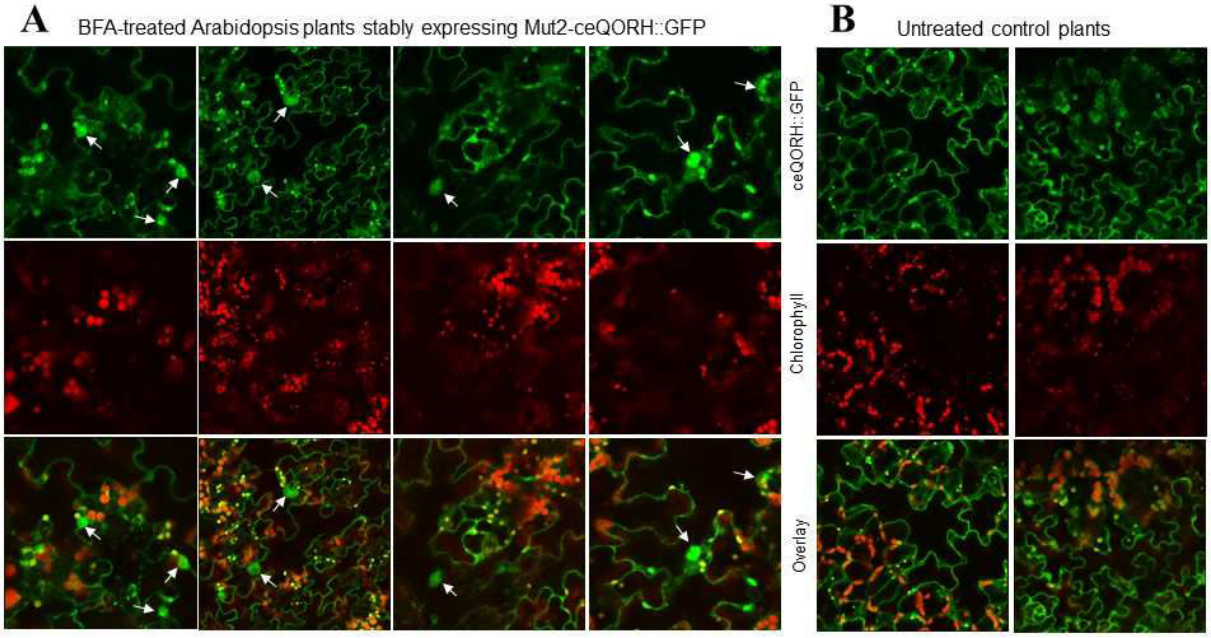


Figure 7: The CaM-binding domain of ceQORH is not required for trafficking through ER: Mut2-ceQORH::GFP accumulates in BFA bodies. A. Arabidopsis plants stably expressing Mut2-ceQORH::GFP were treated with BFA (180 µM). **B.** Untreated control plants. **ceQORH::GFP**; GFP fluorescence, **Chlorophyll**; Chlorophyll autofluorescence. **Overlay:** Overlay of the two channels. Arrows indicate BFA bodies. Note that Mut2-ceQORH::GFP is also detected in plastids. Bar equals 40 mm.

CaM control of chloroplast targeting is not limited to ceQORH. The observation that other plastid proteins apart from ceQORH contain CaM-binding properties[30], prompted us to examine whether a more global role for this regulatory mechanism exists in controlling the subcellular localization of proteins. We used Tic32, an inner envelope membrane protein previously demonstrated to contain a CaM-binding domain located in its *C*-terminus[18], as a candidate to answer this question. Interestingly, when expressed *in planta*, Tic32 was also detected at the periphery of epidermal cells and in a few dotted structures (see Fig. 8a). Similar to Mut2-ceQORH, Tic32 lacking its *C*-terminus (*i.e.* its CaM-binding domain) was mostly detected in plastids (see Fig. 8b). In other words, like ceQORH, the subcellular localization of Tic32 is also CaM dependent in specific tissues, such as epidermal cells.

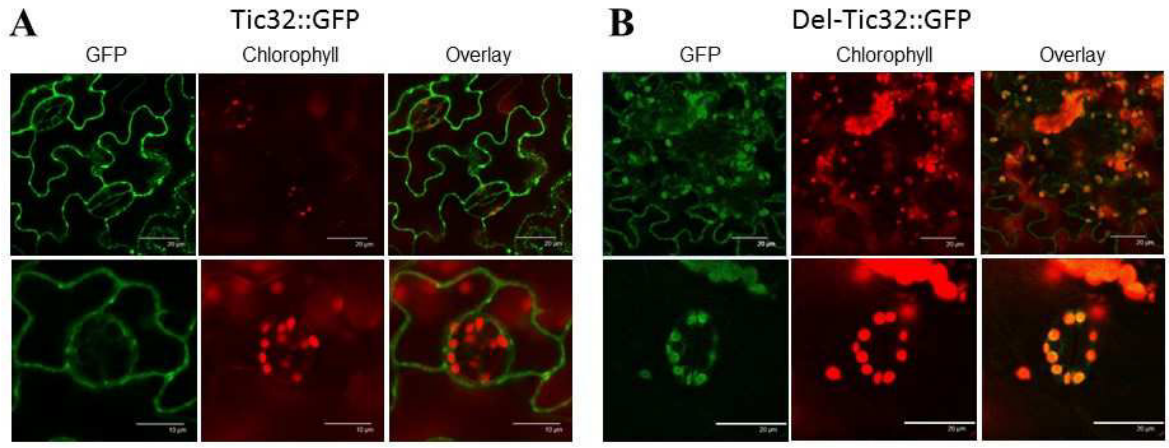


Figure 8. CaM control of chloroplast targeting is not limited to ceQORH. A. Like ceQORH, Tic32 accumulates in the plasma membrane of epidermal cells when containing its CaM-binding domain. **B.** Tic32 lacking its CaM binding domain is targeted to chloroplasts in epidermal cells. Confocal microscopy was performed on plants stably expressing Tic32::GFP (Tic32 fused to GFP, bar equals 10 or 20 mm) and Del-Tic32::GFP (Tic32 protein lacking its *C*-terminus, bar equals 20 mm). GFP; GFP fluorescence, Chlorophyll; Chlorophyll autofluorescence. Overlay: Overlay of the two channels.

Chemically induced release of ceQORH from the PM *in planta*. Mevinolin, which is known to impact the isoprenylation of some PM proteins such as CaM53_PETY[34] and MUBS[35], was expected to block the isoprenylation of Arabidopsis CaM5 based on the similarities among CaM5 homologs. Indeed, such a treatment had a strong impact on ceQORH::GFP localization in epidermal cells. However, while ceQORH was definitively released from the PM, it was not relocated to plastids in treated plants (see Supplementary Fig. 7).

Environmental conditions release ceQORH from the PM. CaM53 from petunia was previously demonstrated to be regulated in its subcellular localization by specific environmental stresses. The role of its isoprenylation status was also studied deeply by colleagues[33,34] who showed that CaM53_PETY is regulated in its subcellular location through cellular metabolic status (*e.g.*, dark-grown plants). Because adverse light conditions were previously shown to impact CaM53 isoprenylation status, we reproduced the original experimental conditions and followed ceQORH location. Compared to plants grown in standard conditions (100 $\mu mol.m^{-2}.s^{-1}$), in dark-grown plants and plants exposed to adverse light conditions (500 $\mu mol.m^{-2}.s^{-1}$), ceQORH was released from the PM but did not localize to plastids in treated plants (see Supplementary Fig. 7). As previously observed for CaM53[34], ceQORH is detected in vesicular structures, probably associated with the endomembrane system.

Heat stress releases ceQORH from CaM5 *in planta*. Interestingly, the expression level of the *cam5* gene was recently demonstrated to considerably increase in response to heat stress, while expression of other *cam* genes was not affected[36]. We thus tested the impact of the same heat stress on the CaM5 protein level *in planta*. Interestingly, after only 2 hours of heat treatment, the high-Mw CaM isoform detected in a whole-cell extract from MS-grown Arabidopsis plants disappeared (see Supplementary Fig. 8a). When Arabidopsis plants were grown on soil, the same high-Mw CaM isoform almost disappeared in only 1 hour (see Supplementary Fig. 8b). We then analysed the subcellular location of ceQORH::GFP in plants after heat stress. Strikingly, following 2 hours of heat stress, ceQORH began to accumulate in plastids of epidermal cells in WT plants (see Fig. 9). This observation strongly suggests that CaM5 plays an important role in controlling ceQORH subcellular location *in planta* and that this control is temperature regulated.

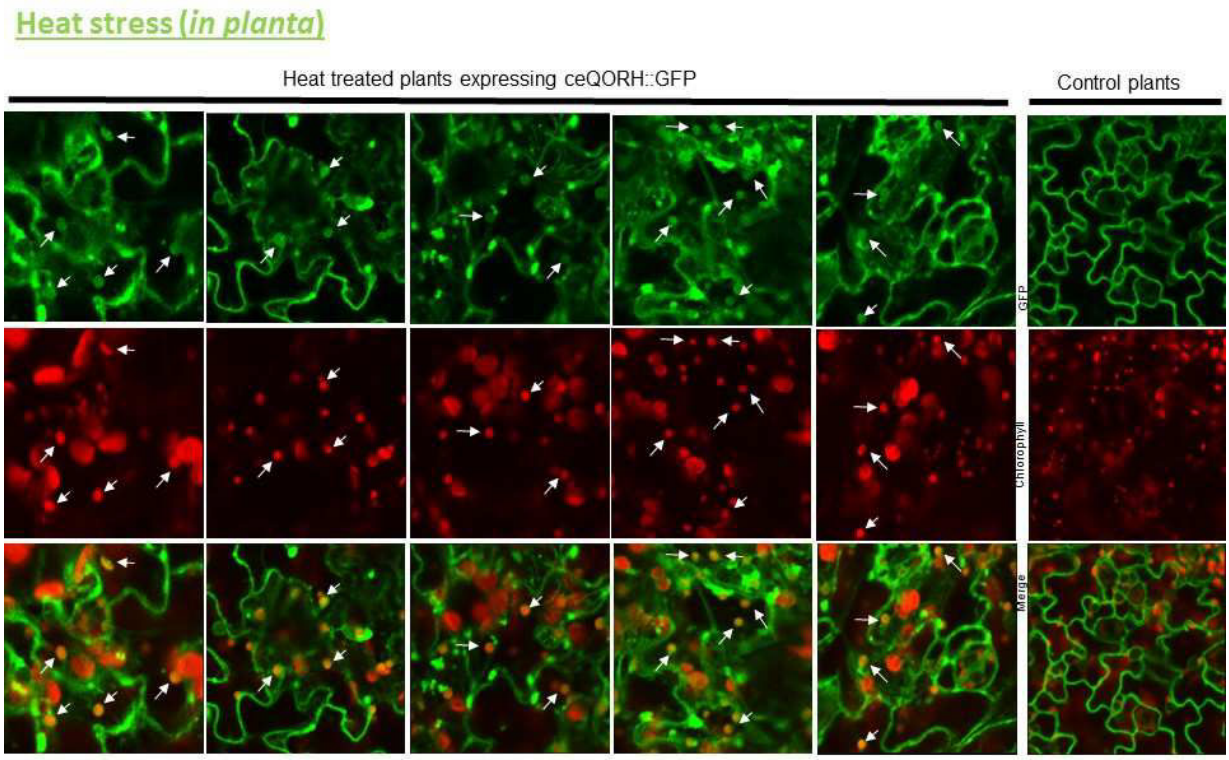


Figure 9. Following heat stress, ceQORH is detected in chloroplasts of epidermal cells in WT plants. Arabidopsis plants stably expressing ceQORH::GFP were incubated for 2 hours at 42°C. Control (untreated plants) are shown as a control. Confocal microscopy was performed on both untreated WT plants stably expressing **ceQORH::GFP** (plant ceQORH fused to GFP) and on the same plants after 2 hours of heat stress at 42°C. **GFP**; GFP fluorescence, **Chlorophyll**; Chlorophyll autofluorescence. **Overlay:** Overlay of the two channels. Arrows indicate plastids accumulating ceQORH::GFP after 2 hours of heat treatment. Bar equals 40 mm.

CaM5 is the only CaM isoform that controls ceQORH retention in the plasma membrane. To confirm that a high-Mw CaM (most likely CaM5) plays a unique role in controlling ceQORH subcellular localization, excluding the involvement of other CaM isoforms in regulating this process, we used a loss-of-function mutant of CaM5[36]. We thus tested the expression of ceQORH (and Mut2-ceQORH as a control) in this *cam5* mutant. As demonstrated in Fig. 10, in contrast to the PM localization in control WT cells, ceQORH was mostly targeted to the chloroplast envelope in epidermal cells of the *cam5* mutant. As expected, Mut2-ceQORH was also detected in plastids, mostly at the periphery of organelles, thus most likely in their envelope membrane (see Fig. 10). This result supports that the CaM-binding domain of ceQORH is essential *in planta* for the alternative localization of ceQORH in specific tissues and confirms that CaM5 is indeed directly involved in the retention of ceQORH in the PM.

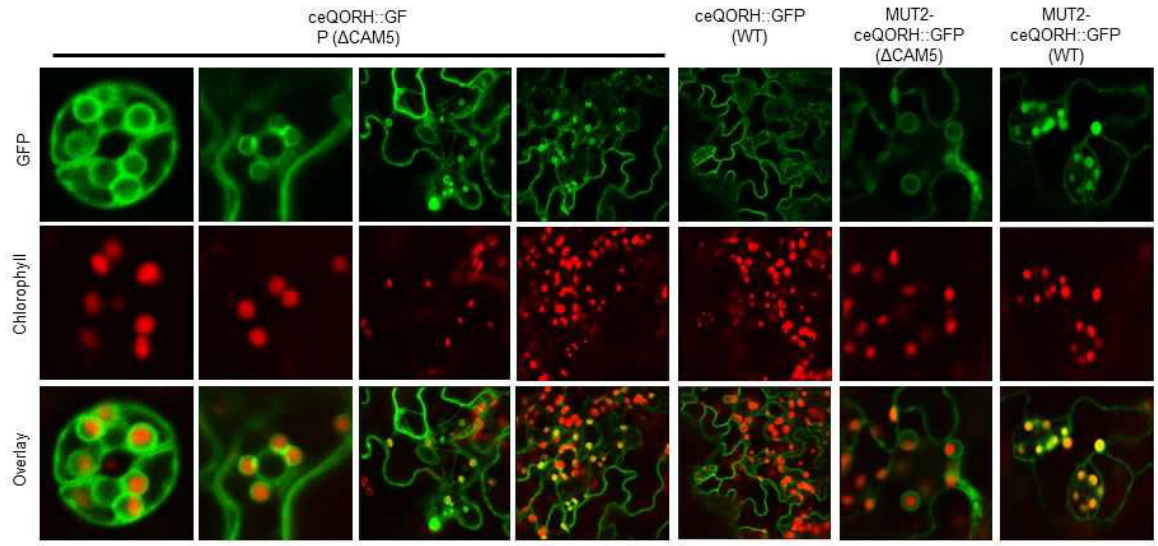


Figure 10. CAM5 is the CaM isoform that controls retention of ceQORH in the plasma membrane: ceQORH is targeted to the chloroplast envelope in epidermal cells of the *cam5* mutant). Confocal microscopy was performed on both WT plants and *cam5* mutant stably expressing **ceQORH::GFP** (plant ceQORH fused to GFP) or **Mut2-ceQORH::GFP**. **GFP**; GFP fluorescence, **Chlorophyll**; Chlorophyll autofluorescence. **Overlay:** Overlay of the two channels. Bar equals 40 mm.

Discussion

Having previously demonstrated that ceQORH was not targeted to plastids through the classical Toc-dependent import machinery[17], a possible hypothesis was that the ceQORH-targeting machinery was absent in epidermal cells, thus explaining its alternative localization in this leaf tissue. However, we show that such an original alternative localization of ceQORH results from a much more complex regulatory process (see Supplementary Fig. 9 for a working model depicting the regulatory mechanism for chloroplast targeting of CaM-binding proteins as deduced from present data and schematic diagrams of the classical secretory pathway in plants).

First, ceQORH cellular movement was shown to involve vesicular trafficking. This finding sheds new light on non-canonical protein import into plastids by suggesting that such an alternative targeting pathway to plastids shares features with glycoproteins en route to the chloroplast despite being considered separate pathways[2,3].

Second, we showed that ceQORH interacts with CaM5, the Arabidopsis orthologue of CaM53 from petunia, in the PM. From earlier work performed by colleagues, we knew that the trafficking of CaM53 to the PM is regulated by its isoprenylation status. Following prenylation, CaM53 is associated with the endomembrane system, where an ER-localized prenyl-cysteine methyltransferase activity methylates the prenylated protein prior to targeting it to its final destination in the PM[32,33]. The simplest interpretation of our results is that ceQORH follows the same route as CAM5 (*i.e.* bound to a prenylated form of CAM5) *in planta.*

Loss of CAM5 prenylation resulting from mevinolin treatment or dark and adverse light conditions lead to the release of ceQORH from the PM. However, none of these treatments allowed ceQORH to reach the chloroplast, most likely because ceQORH remained tightly bound to CaM5. Based on earlier work from other colleagues[36], we then tested the heat resistance of CaM5 and showed that heat stress strongly impacts CaM5 stability. Interestingly, this stress was the only identified environmental condition that allowed ceQORH to accumulate into plastids in epidermal tissues, in contrast with the systematic PM localization observed in control cells. This supports the model where ceQORH, when tightly bound to CAM5, cannot reach the plastid. Further support is provided by the fact that the loss of the CaM-binding properties of ceQORH (*i.e.* Mut2-ceQORH) abolishes PM retention and instead leads to targeting to plastids. However, only the expression of

ceQORH in the *cam5* mutant definitively proves that CaM5 is the sensor of environmental conditions that regulates ceQORH targeting plastids, and not another CaM isoform.

Interestingly, while vesicular trafficking appears as an early evolutionary process for proteins to reach endosymbiotic organelles[14,15], proteins such as ceQORH and Tic32 had to co-evolve to acquire both internal plastid targeting sequences and regulatory CaM-binding properties when evolving from bacteria to eukaryotic cells.

Based on these data, the current model can be summarized as follows. Some non-canonical plastid proteins, en route to the chloroplast, first associate to the ER to reach the PM when two conditions are met. Firstly, they must be CaM-binding proteins. Secondly, CaM5 must be sufficiently abundant to trap them. The release of these trapped components is then regulated by environmental stresses (heat stress in the case of heat-sensitive CaM5), thus allowing these plastid proteins to reach plastids. Thus, CaM5 can be considered as an environmental and metabolic sensor for the subcellular location of these proteins.

This is probably the case for other non-canonical plastid proteins such as Tic32[18] and other already identified CaM-binding chloroplast proteins[34]. However, it cannot be concluded that all plastid proteins lacking a typical transit peptide will follow the same route. Indeed, while recent observations[19] strongly supported that ceQORH does not engage with the trimeric TOC159/75/34 complex for translocation into the chloroplasts, a result that is consistent with previous work[16,17], this recent work also demonstrated that, en route into the plastid, FSD1, which lacks a cleavable cTP), recognizes a specific receptor (Toc132 and Toc120), and does engages with components of the general import pathway, as do the majority of plastid preproteins[19]. In other words, while large-scale proteomic studies identified many plastid proteins lacking a predicted cTP[22,23,37,38], the use of such alternative trafficking pathways to the plastid might not be a general rule when a cTP is missing. Other tools are thus required to answer these questions, including the isolation of mutants affected in vesicular trafficking to plastids in order to depict the role of these non-conventional pathways at the organelle level using large-scale approaches.

CaM proteins have no catalytic activity but rather are well known to act as sensor relays that regulate the catalytic properties of downstream targets[32]. This work demonstrates that specific CaM isoforms might also be involved in the regulation of the subcellular localization of their targets. Many questions emerge from this work. First, CaM retention might not be limited to proteins

targeted to plastids, and this regulatory process might involve biogenesis and functions of other cell compartments. When released from the PM, CaM53 is targeted to the nucleus, as another response to environmental stress[33], and the role of this nuclear localization remains to be identified. Finally, whether such a subcellular control of protein localization involving CaM isoforms exists in non-plant species remains to be investigated.

Materials and Methods

Plant material and standard growth conditions. Arabidopsis plants, Wassilevskija background (WS), were germinated in Petri dishes containing solidified medium (Murashige and Skoog, 1% [w/v] sucrose, and 1% [w/v] agarose) for 2 weeks before being transferred to soil. Plants were then grown in growth rooms at 23°C (12-h light cycle) with a light intensity of 100 $\mu mol.m^{-2}.S^{-1}$.

Purification of chloroplasts and chloroplast subfractions from Arabidopsis. All operations were carried out at 0-5°C. Percoll-purified chloroplasts were obtained from 200-300 g of *Arabidopsis thaliana* leaves. Chloroplast fractions were purified from these purified chloroplasts as previously described[39] and stored in liquid nitrogen until use. The protein content of fractions was estimated using the BIO-RAD protein assay reagent[40].

Purification of crude cell extracts and epidermal tissue from Arabidopsis leaves. Whole-cell proteins from Arabidopsis leaves were extracted from 2- to 4-week-old leaves using a small plastic potter and an Eppendorf tube in 150 µl of extraction buffer containing 30 mM tetrasodium pyrophosphate, 100 mM Tris-HCl pH 6.8 and 1% [w/v] SDS. The crude extract was centrifuged for 5 min at 13,200 rpm and 4°C. Due to the presence of a high concentration of detergent, the supernatant resulting from this centrifugation step contains whole-cell proteins (including highly hydrophobic membrane proteins). Purification of epidermal tissues from Arabidopsis leaves was performed by peeling the epidermis of Arabidopsis leaves using a thin forceps used for plant crosses. Crude epidermal extracts were obtained using the procedure described above and the same extraction buffer. Membrane and soluble fractions of crude cell extracts were obtained as previously described[17].

Production of recombinant His-CaM1 and of rabbit polyclonal antibodies directed against Arabidopsis CaM1 protein. The CaM1 isoform (AT5G37780) was selected due to its high

sequence conservation with other CaM isoforms (CaM1 is 100% identical to CaM4; 97% identical to CaM2, CaM3, CaM6, CaM7 and the first 149 residues of CaM5; and 76% identical to CaM8 and CaM-Like11). Cloning of the Arabidopsis cDNA coding for the CAM1 protein was performed by RT-PCR using pFU Ultra High-Fidelity DNA Polymerase (Stratagene) starting from a homemade Arabidopsis cDNA library. The primer GAAGAACATATGGCGGATCAACTC was designed to introduce an *Nde*I restriction site (underlined residues) at the 5' end of the DNA fragment coding for CAM1. The primer ATCACCTGGATCCAATCACTTAGC was designed to introduce a *Bam*HI restriction site (underlined residues) at the 3' end of the DNA fragment coding for CaM1. The amplified fragments were cloned in pBluescriptSK$^-$. The insert was then digested with *Nde*I and *Bam*HI and inserted into the expression vector pET-15b (Novagen). The resulting expression plasmids allowed the expression of *N*-terminal His-tagged versions of the CaM1 protein. The purified protein was desalted (PD-10 column, Pharmacia) and stored at - 80°C. The purified recombinant His-CaM1 was used as a control (from 100 to 0.1 ng) to estimate the content of the natural CaM proteins in crude leaf extracts, epidermal fractions or purified chloroplast envelope membranes. Two independent rabbit polyclonal antibodies (including CaM-767) raised against the Arabidopsis recombinant CaM1 protein were obtained from the Elevage Scientifique des Dombes (F-01400 Châtillon sur Chalaronne, France).

SDS-PAGE and western blot analyses. SDS-PAGE analyses were performed as described by Chua (ref [41]) using prestained protein molecular weight markers (Fermentas). For western blot analyses, gels were transferred to a nitrocellulose membrane (BA85, Schleicher & Schuell). ceQORH was detected using a previously described ceQORH rabbit polyclonal antiserum raised against the recombinant Arabidopsis ceQORH produced in *E. coli*[16] at a 1/5000 dilution. The antibody raised against LHCP was provided by Dr. Olivier Vallon (IBPC Paris) and used at a 1/20,000 dilution. The secondary antibody for ceQORH and LHCP was an anti-rabbit HRP-conjugated secondary antibody (Interchim) used at a 1/10,000 dilution. GFP was detected using a primary antibody raised against GFP (Euromedex) at a 1/70,000 dilution and an anti-mouse HRP-conjugated secondary antibody (Interchim). Detection of the proteins was performed using ECL. CaM transfer and detection require specific conditions. First, the transfer requires 80 minutes (at 80 V) in a specific buffer (20 mM Tris-HCl, 15 mM glycine, 2 mM $CaCl_2$, 20% [v/v] ethanol). The membrane was then washed for 5 min in TBST and incubated for 1 hour in TBST containing 5% [w/v] defatted milk. Incubation with the primary antibody at a 1/1500 dilution requires an overnight incubation at

4°C. The membrane was then washed 3 times for 10 minutes in TBST and incubated for 1 hour in TBST containing the secondary antibody (anti-rabbit fused to HRP, Interchim) at a 1/10,000 dilution. The membrane was then washed 3 times for 10 minutes in TBST before the proteins were detected using ECL. Notably, due to the 97% sequence identity between the two isoforms, the rabbit polyclonal antibody CaM-767 raised against the Arabidopsis CaM1 isoform cross-reacts with all Arabidopsis CaM isoforms, including CaM5.

Production of vectors for expression of the full-length and truncated forms of ceQORH in *E. coli* to identify its CaM-binding domain. Several constructs used to express ceQORH and several truncated forms in *E. coli* were previously described[16,17]: Construct 1, ceQORH; construct 2, ceQORH::GFP (full-length ceQORH protein fused to GFP); construct 3, Δ(1-59)ceQORH::GFP (ceQORH lacking 60 residues in the *N*-terminus fused to GFP); construct 4, Δ(1-99)ceQORH::GFP (ceQORH lacking 100 residues in the *N*-terminus fused to GFP); construct 5, (6-100)ceQORH::GFP (ceQORH lacking both 6 residues in the *N*-terminus and 229 residues in *C*-terminus fused to GFP); and construct 6, (60-100)ceQORH::GFP (ceQORH lacking both 60 residues in the *N*- terminus and 229 residues in the *C*-terminus fused to GFP). Additional constructs were produced, as follows. Construct 7, Δ(280-329)ceQORH::GFP (ceQORH lacking 49 residues in the *N*-terminus fused to GFP), was amplified using the following primers: ceQORH1-175 (TATGCAGTCCAGCTGTAAGAGCGTTGGATC) and ceQORH 1-175 reverse (GATCCAACGCTCTTACAGCTGGACTGCATA). Construct 8, Δ(229-329)ceQORH::GFP (ceQORH lacking 100 residues in the *N*-terminus fused to GFP), was amplified using the following primers: ceQORH 1-229 (GTGGTCCATTGTGCATAAGAGCGTTGGATC) and ceQORH1-229-reverse (GATCCAACGCTCTTATGCACAATGGACCAC). Construct 9, Δ(175-329)ceQORH::GFP (ceQORH lacking 146 residues in the *N*-terminus fused to GFP), was amplified using the following primers: ceQORH1-280 (CTCTTGTTGATCCCAAAATAAGAGCGTTGGATC) and ceQORH1-280-reverse (GATCCAACGCTCTTATTTTGGGATCAACAAGAG). For targeted mutagenesis and the creation of construct 10 (Mut1), construct 11 (Mut2) and construct 12 (Mut3), vectors were created from targeted mutagenesis of available vectors using the QuikChange® Site-Directed Mutagenesis Kit from Stratagene. Mut1-ceQORH was obtained using the following primers: mut1qorBamHI (GGGCCTAATGCAATGGGATCCTATGCGGTTGCTGCAATAACCATGTCAAAG) and

mut1qorBamHIreverse (CTTTGACATGGTTATTGCAGCAACCGCATAGGATCCCATTGCATTAGGCCC). Mut2-ceQORH was obtained using the following primers: mut2qorBamHI (GTTAAGAAAATAACCGGATCCGCAGCTCAGTTAGTGCCACTC) and mut2qorBamHIreverse (GAGTGGCACTAACTGAGCTGCGGATCCGGTTATTTTCTTAAC). Mut3-ceQORH was obtained using the following primers: mut3qorBamHI (GGGCCTAATGCAATGGGATCCTATGCGGTTGCTGCAATAACCGGGTCAGCAGCTCAG TTAGTGCCACTC) and mut3qorBamHIreverse (GAGTGGCACTAACTGAGCTGCTGACCCGGTTATTGCAGCAACCGCATAGGATCCCA TTGCATTAGGCCC).

Production of recombinant full-length and truncated forms of ceQORH in *E. coli*. Several constructs used to express full-length ceQORH and truncated forms in *E. coli* were previously described[16,17]. The additional Δ(175-329), Δ(229-329) or Δ(280-329) truncated forms and the constructs corresponding to mutants 1 to 3 (targeted mutagenesis) were obtained during this work. Briefly, PCR-amplified fragments were digested with *Nde*I and *Bam*HI and inserted into the expression vector pET-15b (Novagen). For all constructs, after the induction of protein expression for 3 hours using IPTG, bacterial pellets (equivalent to 1 ml of cell culture) were solubilized in 100 µl of 10 mM Tris-HCl, pH 6.8. Because GFP was present in all protein fusions, it was possible to estimate the relative expression levels (using western blotting and GFP antibody detection) obtained for all recombinant proteins. Variation in expression levels was thus controlled before CaM overlay experiments. As a negative control, the bacterial ecQOR protein (from *E. coli* K12) was also produced and tested.

Affinity purification of recombinant CaM-binding proteins from Arabidopsis or *E. coli*. Affinity purification of Arabidopsis ceQORH from crude cell extracts was performed on crude solubilized plant proteins diluted in CaM-binding buffer (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM Mg-acetate, 1 mM imidazole, 2 mM $CaCl_2$) containing 0.1% [w/v] NP-40. CaM-binding proteins were purified using the CaM Affinity Resin (Stratagene) and the small-scale quick batch method according to the manufacturer's instructions with minor modifications. Purification from bacteria started with bacterial pellets (equivalent to 100 ml of cell culture) that were solubilized in

5 ml of CaM-binding buffer. After 3 rounds (for 2 minutes) of sonication, cell extracts were centrifuged for 20 minutes at 5500 rpm at 4°C. Then, NP40 (0.1% [w/v] final concentration) was added to the resulting supernatant (5 ml). CaM Affinity Resin (Stratagene; 40 µl, binding capacity equivalent to 60-100 µg proteins) was then added to the solution before incubation (stirring wheal) for 1.5 h at 4°C. The resin was then washed with 5 ml of CaM-binding buffer, and bound proteins were eluted in 250 µl of elution buffer (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM Mg-acetate, 1 mM imidazole, 2 mM EGTA, 0.1% [w/v] NP40). Purified recombinant proteins were then quantified using the BIO-RAD protein assay reagent[40] before SDS-PAGE analyses and CaM overlay.

CaM overlay. After SDS-PAGE and transfer, the resulting membrane was first dried and further incubated for 45 min in TBS buffer containing 0.1% [w/v] Triton X-100. The membrane was then incubated for 1 h in TBS buffer containing 0.1% [w/v] Triton X-100 and 5% [w/v] defatted milk followed by 1 h in saturation buffer containing 3% [w/v] gelatin, 50 mM Tris-HCl pH 7.5, 50 mM $MgCl_2$, 150 mM NaCl and 5 mM $CaCl_2$. Hybridization was then performed for 2 h in hybridization buffer containing 1% [w/v] gelatin, 50 mM Tris-HCl pH 7.5, 50 mM $MgCl_2$, 150 mM NaCl, 5 mM $CaCl_2$ and 0.1% [w/v] Tween 20) + 0.1 µg/ml biotinylated CaM (Sigma). Detection of the bound CaM protein was then performed using a streptavidin-HRP conjugate (Sigma) and ECL using the manufacturer's instructions. Alternatively, after transfer, washing and saturation, the membrane was incubated in the hybridization solution containing 0.3 µg/ml of a homemade CaM::HRP conjugate (the recombinant Arabidopsis CaM1 (AT5G37780) was covalently coupled to horseradish peroxidase (HRP)). Detection of the bound CaM protein was then performed using ECL according to the manufacturer's instructions.

Arabidopsis transformation. Wild-type and transgenic *Arabidopsis* plants (ecotype WS) were transformed by dipping the floral buds of 4-week-old plants into an *Agrobacterium tumefaciens* (C58 strain) solution containing a surfactant (Silwett L-77), according to Clough and Bent (ref [42]). Primary transformants were selected on Murashige and Skoog medium containing 100 mg/l kanamycin. Resistant lines expressing the recombinant protein (as controlled with western blots using GFP antibody) were selected for further analyses. Primary transformants were then self-pollinated to obtain plants homozygous for insertion.

Construction of vectors for stable expression of GFP fusions in Arabidopsis. To construct vectors for the overexpression in Arabidopsis, all cDNAs were PCR amplified using pFU Ultra High-Fidelity DNA Polymerase (Stratagene) starting from a homemade Arabidopsis cDNA library. Correct orientation and sequence of the inserted fragment were controlled. Plasmids used for *Agrobacterium tumefaciens* transformation were prepared using the QIAfilter Plasmid Midi Kit (Qiagen Laboratories; Germany). **ceQORH**: The coding region of Arabidopsis ceQORH (AT4G13010) was PCR amplified using the two flanking primers ceQORH*BamH*I-*N*-ter (CCTGGATCCATGGCTGGAAAACTCATG) and ceQORH*Sac*I-*C*-ter (ACAGAGCTCTTATGGCTCGACAATGATCTTC), and the PCR product was cloned into the pBluescript SK$^-$ vector (Stratagene) to create the ceQORH-pKS vector. The *BamH*I-*Sac*I fragment cleaved from this plasmid was inserted into the *BamH*I-*Sac*I-digested pEL103 binary vector (kanamycin resistance to transform wild-type plants), resulting in a vector allowing stable expression of ceQORH::GFP. **Mut2-ceQORH** was produced using targeted mutagenesis of pre-existing constructs using the following primers: mut2qorBamHI (GTTAAGAAAATAACCGGATCCGCAGCTCAGTTAGTGCCACTC) and mut2qorBamHIrevers (GAGTGGCACTAACTGAGCTGCGGATCCGGTTATTTTCTTAAC). **TIC32**: The coding region of Arabidopsis Tic32 (At4g23430) cDNA was PCR amplified using the two flanking primers Tic32*Sal*I-*N*-ter (GCTAGTCGACATATGTGGTTTTTTGGATC) and Tic32*Nco*I-C-ter (TCCTCCATGGAACTGCTTTCTCCTGATTG). This fragment was cloned into the pKS vector (Stratagene) and then transferred to pUC-GFP (see ref [16]) using the *Sal*I (5') and *Nco*I (3') restriction sites. Then, the *Eco*RI/*Hind*III cassette was transferred to pEL103, resulting in a vector allowing stable expression of Tic32::GFP. **Del-Tic32**: Tic32 lacking its *C*-terminus (*i.e.,* lacking 27 residues in the *C*-terminus, including the CaM-binding domain, see ref [18]) was constructed as for Tic32::GFP, except that the primer used to amplify the Tic32 cDNA in 3' was Del-Tic32*Nco*I-C-ter (GTATCCATGGCAAGTGGTAATGGTTTAGC).

Construction of vectors for BiFC. Construction of the plasmids for the expression of ceQORH, Mut2-ceQORH and CaM5 proteins fused to YFP *C*-ter or *N*-ter was performed as follows. The coding regions of candidate proteins were PCR amplified using two flanking primers, *XbaI-N*-ter (TCATCTAGAATGGCTGGAAAACTC for **ceQORH** and **Mut2-ceQORH** or TCATCTAGAATGGCAGATCAGCTC for **CaM5**) and *SalI-C*-ter

(CCAGTCGACTGGCTCGACAATGATC for **ceQORH** and **Mut2-ceQORH** or CCAGTCGACGAGAATACGGCAGTG for **CaM5**). The PCR products were cloned into the pBlueScript SK$^-$ Vector. *XbaI-Sal*I fragments cleaved from this plasmid were inserted into the *XbaI-Sal*I-digested YFP *C*-ter or YFP *N*-ter reporter plasmid pUC-SPYCE and pUC-SPYNE[43] to create the 35Ω-candidate-YFP *C*-ter/*N*-ter vectors. From these constructs, the *Eco*RI-*Hind*III fragments were extracted and inserted into the *Eco*RI-*Hind*III-digested pEL103 binary vector (kanamycin resistance as a selection marker for transformed plants). The correct orientation and sequence of the inserted fragments were controlled. Constructs were transferred to *Agrobacterium tumefaciens* strain C58 and used to transform tobacco leaves using agroinfiltration. Plasmids used for *Agrobacterium tumefaciens* transformation were prepared using the NucleoSpin Plasmid "Kit" (Macherey-Nagel). Localization of the GFP and GFP fusions in tobacco leaves was analysed by confocal fluorescence microscopy three to four days after agroinfiltration.

Confocal microscopy. Fluorescence microscopy was performed with a confocal laser-scanning microscope (TCS-SP2, Leica, Deerfield, IL). Treatment with **Brefeldin A** (BFA) was performed as previously described[7]. Briefly, stock solutions of BFA (Sigma) were prepared at 36 mM by dissolving BFA in DMSO. Aliquots of this stock were diluted with 50 mM Tris-HCl, pH 8.0 to a final concentration of 180 µM BFA. This dilution was used for infiltration of 2- to 3-week-old Arabidopsis leaves. After incubation for 3 h at room temperature and continuous agitation with BFA, leaves were observed by confocal microscopy. The coloration of tissues with **rhodamine B hexyl ester** was performed as previously described[44]. Briefly, stock solutions of rhodamine B (Invitrogen) were prepared at 10 mg/ml by dissolving rhodamine in DMSO. Aliquots of this stock were diluted with 50 mM Tris-HCl, pH 8.0, to a final concentration of 1 µg/ml rhodamine. This dilution was used to infiltrate 2- to 3-week-old Arabidopsis leaves. After incubation for a few minutes at room temperature with rhodamine, leaves were observed by confocal microscopy. Treatment with **mevinoline** was performed as previously described[35]. Briefly, stock solutions of mevinolin (Sigma) were prepared at 1 M by dissolving mevinolin in acetone. Aliquots of this stock were diluted with 50 mM Tris-HCl, pH 8.0, to a final concentration of 25 µM. This dilution was used to infiltrate 2- to 3-week-old Arabidopsis leaves. After incubation for a few minutes at room temperature, leaves were observed by confocal microscopy.

Growth of plant material under environmental stress conditions. Arabidopsis plants, Wassilevskija background (WS), were first germinated in Petri dishes containing solidified medium (Murashige and Skoog, 1% [w/v] sucrose, and 1% [w/v] agarose) for 2 weeks before being transferred to soil and placed in growth rooms at 23°C (12-h light cycle) with a light intensity of 100 $\mu mol.m^{-2}.S^{-1}$. Before imaging, plants grown in standard conditions were either stored in the dark (**dark-grown** plants) for 0.5 and 3 days or were transferred to **high light** conditions (500 $\mu mol.m^{-2}.S^{-1}$) for 0.5 and 3 days. After treatment, plants were immediately analysed using fluorescence microscopy. Note that dark-grown plants were maintained in the dark until imaging to limit the residual impact of ambient light. **For heat stress**, Arabidopsis plants were either pre-grown on MS medium or on soil in growth rooms at 23°C (12-h light cycle) with a light intensity of 100 $\mu mol.m^{-2}.S^{-1}$. For heat-stress conditions, plants were transferred to 42°C but maintained at the same light intensity. At various periods (at 0 h, 2 h, 5 h and 8 h for MS-grown plants and at 15 min, 30 min, 1 h, and 2 h before cooling at 23°C for 1 h and 2 h for soil-grown plants), samples were taken and either immediately analysed using confocal microscopy or stored in liquid nitrogen upon use for western blot detection of CaM. For western blot analyses, whole-cell proteins from Arabidopsis leaves were extracted from heat-treated leaves using the protocol and buffers described for crude cell extracts (see above). Then, 15 µg of proteins from each crude cell extract were run on an SDS-PAGE, transferred to membranes using CaM-specific conditions (see above), and CaM was detected using the CaM-767 antibody.

Genotyping of the *cam5* mutant from Arabidopsis: The Arabidopsis homozygous mutant (SALK_027181) affected in the cam5 gene (AT2G27030.3) was controlled using the following flanking primers: CaM5fwd (CTCGGCAGCTGAGTTAAGACATG) and CaM5rev (GCTGCCATAACTCTCTTCCCTC) coupled to T-DNA primers LB (GGCAATCAGCTGTTGCCCGTCTCACTGGTG) and RB (GCTCATGATCAGATTGTCGTTTCCCGCCTT).

Acknowledgements

NR wishes to dedicate this article to the memory of Dr. Arsenio Villarejo, our respected colleague and friend, who passed away in May 2011, for his pioneering work and inspiring contribution to this topic. This study was financially supported by the French National Research Agency. L.M.

was funded by fellowships from the French National Research Agency (ANR-06-GPLA-0003, ANR Génomique végétale - Genoplante 2010 *Glyco-chloroplast*; ANR-2010-BLAN-1610-01, ANR Blanc *Chloro-Pro*; ANR-10-LABX-49-01, *Labex GRAL*). I.B. was funded by a PhD fellowship from the INRA Plant Biology and Breeding Division and from the French National Research Agency (ANR-10-LABX-49-01; *Labex GRAL*). S.M. was funded by a PhD fellowship from the CEA. We would also like to acknowledge the ANR project ANR-15-IDEX-02. Additional support was provided by specific ANR Labex GRAL, CNRS and INRA funding and by access to the local BIG-CEA facilities for confocal microscopy.

Competing interests

The authors declare no competing interests.

References

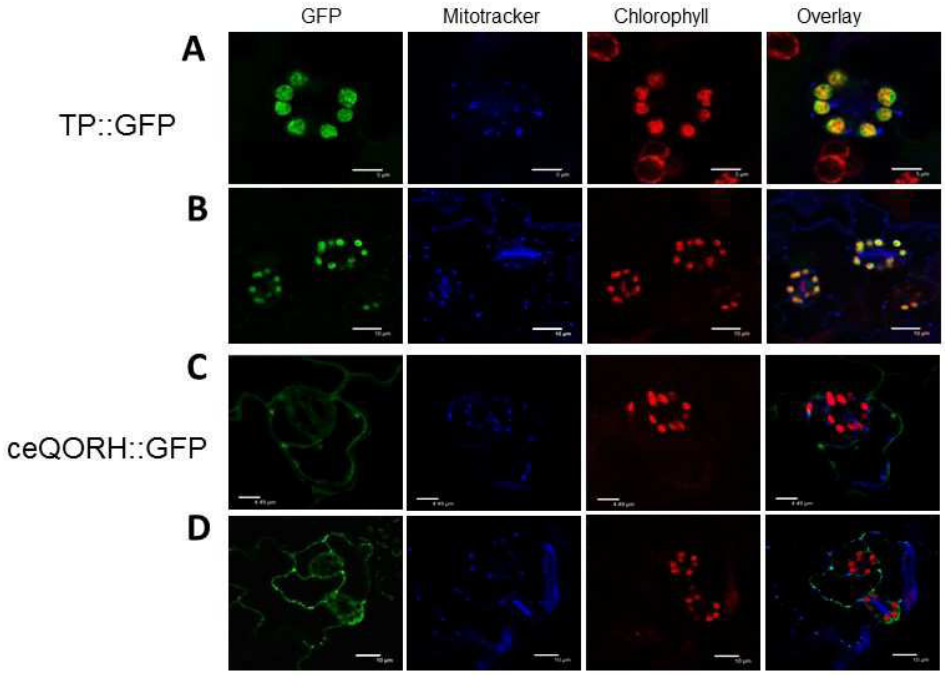
1. Bellucci M, De Marchis F, Pompa A. The endoplasmic reticulum is a hub to sort proteins toward unconventional traffic pathways and endosymbiotic organelles. *J. Exp. Bot.* **69**(1), 7-20 (2018).
2. Jarvis P. Targeting of nucleus-encoded proteins to chloroplasts in plants. *New Phytol.* **179**(2), 257-285. 2008.
3. Jarvis P, López-Juez E. Biogenesis and homeostasis of chloroplasts and other plastids. *Nat. Rev. Mol. Cell. Biol.* **14**(12), 787-802 (2013).
4. Lee DW, Lee J, Hwang I. Sorting of nuclear-encoded chloroplast membrane proteins. *Curr. Opin. Plant. Biol.* **40**, 1-7 (2017).
5. Sjuts I, Soll J, Bölter B. Import of soluble proteins into chloroplasts and potential regulatory mechanisms. *Front. Plant Sci.* **8**, 168 (2017).
6. Tian L, Okita TW. mRNA-based protein targeting to the endoplasmic reticulum and chloroplasts in plant cells. *Curr. Opin. Plant Biol.* **22**, 77-85 (2014).
7. Villarejo A, Burén S, Larsson S, Déjardin A, Monné M, Rudhe C, Karlsson J, Jansson S, Lerouge P, Rolland N, von Heijne G, Grebe M, Bako L, Samuelsson G. Evidence for a protein transported through the secretory pathway en route to the higher plant chloroplast. *Nat. Cell. Biol.* **7**(12), 1224-1231 (2005).
8. Asatsuma S, Sawada C, Itoh K, Okito M, Kitajima A, Mitsui T. Involvement of alpha-amylase I-1 in starch degradation in rice chloroplasts. *Plant Cell Physiol.* **46**(6), 858-869 (2005).
9. Kitajima A, Asatsuma S, Okada H, Hamada Y, Kaneko K, Nanjo Y, Kawagoe Y, Toyooka K, Matsuoka K, Takeuchi M, Nakano A, Mitsui T. The rice alpha-amylase glycoprotein is targeted from the Golgi apparatus through the secretory pathway to the plastids. *Plant Cell.* **21**(9), 2844-2858 (2009).

10. Shiraya T, Mori T, Maruyama T, Sasaki M, Takamatsu T, Oikawa K, Itoh K, Kaneko K, Ichikawa H, Mitsui T. Golgi/plastid-type manganese superoxide dismutase involved in heat-stress tolerance during grain filling of rice. *Plant Biotechnol. J*. **13**(9), 1251-1263 (2015).
11. Nanjo Y, Oka H, Ikarashi N, Kaneko K, Kitajima A, Mitsui T, Muñoz FJ, Rodríguez-López M, Baroja-Fernández E, Pozueta-Romero J. Rice plastidial N-glycosylated nucleotide pyrophosphatase/phosphodiesterase is transported from the ER-golgi to the chloroplast through the secretory pathway. *Plant Cell.* **18**(10), 2582-2592 (2006).
12. Kaneko K, Takamatsu T, Inomata T, Oikawa K, Itoh K, Hirose K, Amano M, Nishimura S, Toyooka K, Matsuoka K, Pozueta-Romero J, Mitsui T. N-Glycomic and Microscopic subcellular localization analyses of NPP1, 2 and 6 strongly indicate that trans-Golgi compartments participate in the Golgi to plastid traffic of Nucleotide Pyrophosphatase/Phosphodiesterases in rice. *Plant Cell Physiol*. **57**(8), 1610-28 (2016).
13. Baslam M, Oikawa K, Kitajima-Koga A, Kaneko K, Mitsui T. Golgi-to-plastid trafficking of proteins through secretory pathway: Insights into vesicle-mediated import toward the plastids. *Plant Signal. Behav*. **11**(9), e1221558 (2016).
14. Nowack EC, Grossman AR. Trafficking of protein into the recently established photosynthetic organelles of Paulinella chromatophora. *Proc. Natl. Acad. Sci. U. S. A.* **109**(14), 5340-5345 (2012).
15. Singer A, Poschmann G, Mûhlich C, Valadez-Cano C, Hänsch S, Hüren V, Rensing SA, Stühler K, Nowack ECM. Massive Protein Import into the Early-Evolutionary- Stage Photosynthetic Organelle of the Amoeba Paulinella chromatophora. *Curr. Biol*. **27**, 2763-2773 (2017).
16. Miras S, Salvi D, Ferro M, Grunwald D, Garin J, Joyard J, Rolland N. Non canonical transit peptide for import into the chloroplast. *J. Biol. Chem.* **277**, 47770-47778 (2002).
17. Miras S, Salvi D, Piette L, Seigneurin-Berny D, Grunwald D, Reinbothe C, Joyard J, Reinbothe S, Rolland N. Toc159- and Toc75-independent import of a transit sequence-less precursor into the inner envelope of chloroplasts. *J. Biol. Chem.* **282**(40), 29482-29492 (2007).
18. Nada A, Soll J. Inner envelope protein 32 is imported into chloroplasts by a novel pathway. *J. Cell. Sci.* **117**(Pt 17), 3975-3982 (2004).
19. Chang W, Soll J and Bölter B A new member of the psToc159 family contributes to distinct protein targeting pathways in pea chloroplasts *Front. Plant Sci*. **5**, 239 (2014).
20. Curien G, Giustini C, Montillet JL, Mas-Y-Mas S, Cobessi D, Ferrer JL, Matringe M, Grechkin A, Rolland N. The chloroplast membrane associated ceQORH putative quinone oxidoreductase reduces long-chain, stress-related oxidized lipids. *Phytochemistry*. **122**:45-55 (2016).
21. Mas Y Mas S, Curien G, Giustini C, Rolland N, Ferrer JL, Cobessi D. Crystal structure of the chloroplastic oxoene reductase ceQORH from *Arabidopsis thaliana*. *Front. Plant Sci*. **8**, 329 (2017).

22. Ferro M, Brugière S, Salvi D, Seigneurin-Berny D, Court M, Moyet L, Ramus C, Miras S, Mellal M, Le Gall S, Kieffer-Jaquinod S, Bruley C, Garin J, Joyard J, Masselon C, Rolland N. AT_CHLORO, a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins. *Mol. Cell. Proteomics*. **9**(6), 1063-1084 (2010).
23. Ferro M, Salvi D, Brugière S, Miras S, Kowalski S, Louwagie M, Garin J, Joyard J, Rolland N. Proteomics of the chloroplast envelope membranes from *Arabidopsis thaliana*. *Mol. Cell. Proteomics.* **2**: 325-345 (2003).
24. Carrie C, Small I. A reevaluation of dual-targeting of proteins to mitochondria and chloroplasts. *Biochim. Biophys. Acta*. **1833**(2), 253-259 (2013).
25. Marmagne A, Rouet MA, Ferro M, Rolland N, Alcon C, Joyard J, Garin J, Barbier-Brygoo H, Ephritikhine G. Identification of new intrinsic proteins in Arabidopsis plasma membrane proteome. *Mol. Cell. Proteomics.* **3**(7), 675-691 (2004).
26. Alexandersson E, Saalbach G, Larsson C, Kjellbom P. Arabidopsis plasma membrane proteomics identifies components of transport, signal transduction and membrane trafficking. *Plant Cell Physiol*. **45**(11), 1543-1556 (2004).
27. Nikolovski N., Rubtsov D., Segura M. P., Miles G. P., Stevens T. J., Dunkley T. P., Munro S., Lilley K. S., Dupree P., Putative glycosyltransferases and other plant Golgi apparatus proteins are revealed by LOPIT proteomics. *Plant Physiol*. **160**, 1037-1051 (2012).
28. Heard W, Sklenář J, Tomé DF, Robatzek S, Jones AM. Identification of Regulatory and Cargo Proteins of Endosomal and Secretory Pathways in Arabidopsis thaliana by Proteomic Dissection. *Mol. Cell. Proteomics*. **14**(7), 1796-1813 (2015).
29. Robinson DG, Jiang L, Schumacher K. The Endosomal System of plants: charting new and familiar territories. *Plant Physiol*. **147**, 1482-1492 (2008).
30. McCormack E, Tsai YC, Braam J. Handling calcium signaling: Arabidopsis CaMs and CMLs. *Trends Plant Sci.* **10**(8), 383-389 (2005).
31. La Verde V, Dominici P, Astegno A. Towards understanding plant calcium signaling through calmodulin-like proteins: A biochemical and structural perspective. *Int. J. Mol. Sci.* **19**(5), pii: E1331 (2018).
32. Rodríguez-Concepción M, Yalovsky S, Zik M, Fromm H, Gruissem W. The prenylation status of a novel plant calmodulin directs plasma membrane or nuclear localization of the protein *EMBO J*. **18**(7), 1996-2007 (1999).
33. Rodríguez-Concepción M, Toledo-Ortiz G, Yalovsky S, Caldelari D, Gruissem W. Carboxyl-methylation of prenylated calmodulin CaM53 is required for efficient plasma membrane targeting of the protein. *Plant J*. **24**(6), 775-784 (2000).

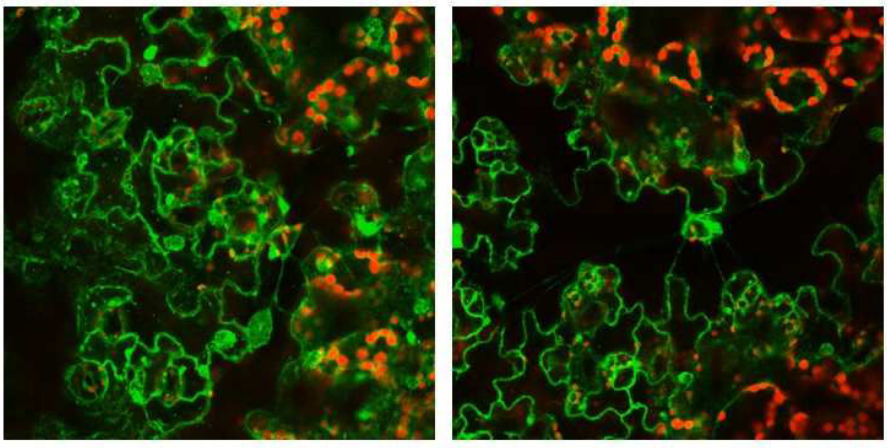
34. Dell'Aglio E, Giustini C, Salvi D, Brugière S, Delpierre F, Moyet L, Baudet M, Seigneurin-Berny D, Matringe M, Ferro M, Rolland N, Curien G. Complementary biochemical approaches applied to the identification of plastidial calmodulin-binding proteins. *Mol. Biosyst*. **9**(6), 1234-1248 (2013).
35. Downes BP, Saracco SA, Lee SS, Crowell DN, Vierstra RD. MUBs, a family of Ubiquitin-fold proteins that are plasma membrane-anchored by prenylation. *J. Biol. Chem*. **281**, 27145-27157 (2006).
36. Al-Quraan NA, Locy RD, Singh NK. Expression of calmodulin genes in wild type and calmodulin mutants of Arabidopsis thaliana under heat stress. *Plant Physiol. Biochem.* **48**(8), 697-702. (2010).
37. Zybailov B, Rutschow H, Friso G, Rudella A, Emanuelsson O, Sun Q, van Wijk KJ. Sorting signals, *N*-terminal modifications and abundance of the chloroplast proteome. *PLoS One*. **3**(4), e1994 (2008).
38. Armbruster U, Hertle A, Makarenko E, Zühlke J, Pribil M, Dietzmann A, Schliebner I, Aseeva E, Fenino E, Scharfenberg M, Voigt C, Leister D. Chloroplast proteins without cleavable transit peptides: rare exceptions or a major constituent of the chloroplast proteome? *Mol. Plant*. **2**(6), 1325-1335 (2009).
39. Salvi D, Rolland N, Joyard J, Ferro M. Purification and proteomic analysis of chloroplasts and their sub-organellar compartments. *Methods Mol. Biol*. **432**, 19-36 (2008).
40. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem*. **72**, 248-254 (1976).
41. Chua NH. Electrophoretic analysis of chloroplast proteins. Methods Enzymol **69**: 434-436 (1980).
42. Clough SJ, Bent AF. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J*. **16**, 735-743 (1998).
43. Walter M, Chaban C, Schütze K, Batistic O, Weckermann K, Näke C, Blazevic D, Grefen C, Schumacher K, Oecking C, Harter K, Kudla J. Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. *Plant J.* **40**(3), 428-438 (2004).
44. Boevink P, Santa Cruz, Hawes C, Harris N, Oparka KJ. Virus-mediated delivery of the green fluorescent protein to the endoplasmic reticulum of plant cells. *Plant J.* **10**, 935-941 (1996).
45. Rose JK, Lee SJ. Straying off the highway: trafficking of secreted plant proteins and complexity in the plant cell wall proteome. *Plant Physiol*. **153**(2), 433-436 (2010).
46. Robinson DG, Pimpl P. Clathrin and post-Golgi trafficking: a very complicated issue. *Trends Plant Sci*. **19**(3), 134-139 (2014).

Supplemental figures:

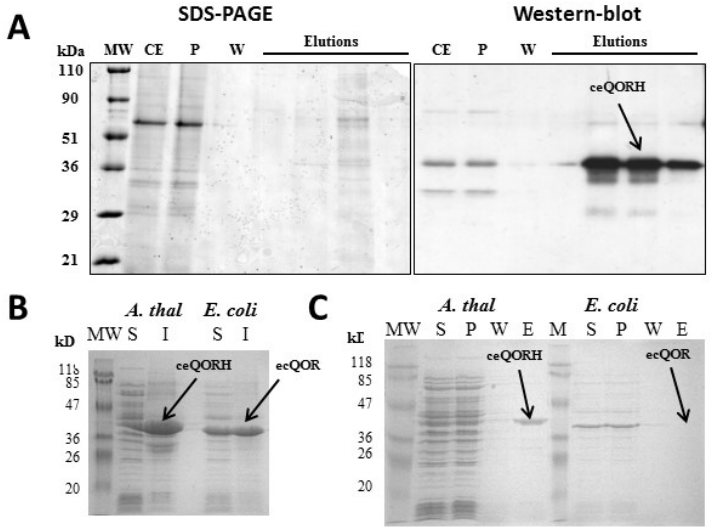


Supporting Figure 1. ceQORH is not addressed to plastids or mitochondria in epidermal cells. Confocal microscopy was performed on plants stably expressing **TP::GFP** (transit sequence of the small subunit of RuBisCO fused to GFP) for **A** (bar equals 5 mm) and B (bar equals 10 mm) or **ceQORH::GFP** (plant ceQORH fused to GFP) for **C** (bar equals 4.5 mm) and **D** (bar equals 10 mm). **GFP**; GFP fluorescence, **Mitotracker**; Mitotracker fluorescence, **Chlorophyll**; Chlorophyll autofluorescence. **Overlay:** Overlay of all three channels.

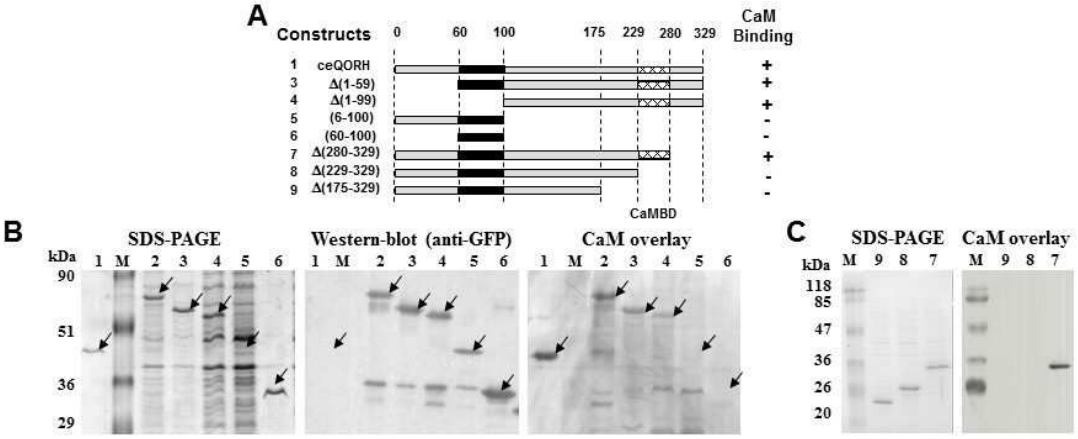
Video: ceQORH accumulates in ER (BFA-bodies) in BFA treated plants



Supporting Figure 2: ceQORH::GFP accumulates in BFA bodies. Arabidopsis plants stably expressing ceQORH::GFP were treated with BFA (180 µM). Images result from overlay of GFP fluorescence (ceQORH::GFP) and Chlorophyll autofluorescence.

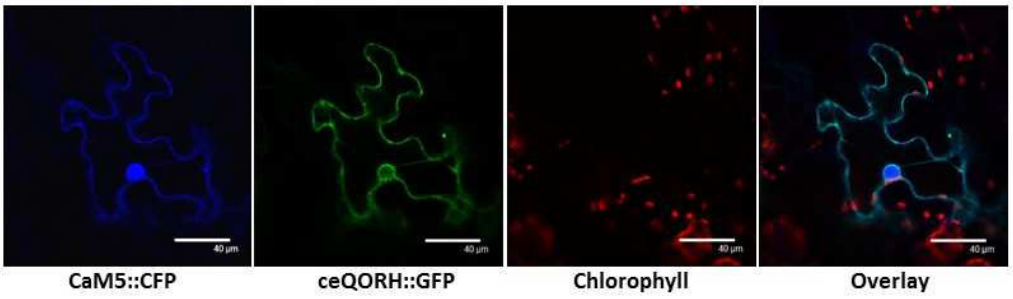


Supporting Figure 3. Interaction of the natural plant ceQORH protein with Calmodulin. **A. Affinity purification of Arabidopsis ceQORH from crude cell extracts.** Purification was performed on a CaM-Affinity Resin (Stratagene). **CE**; crude solubilized plant proteins diluted in CaM-binding buffer containing 0.1% NP-40, **P**; unbound proteins, **W**; wash and **E**; elution. **B. Expression of Arabidopsis (ceQORH) and *E. coli* K12 QOR (ecQOR) proteins in *E. coli*. S**: soluble fraction of the recombinant plant and bacterial proteins. **I**: insoluble fraction of the recombinant proteins (inclusion bodies). **C. Interaction of the recombinant Arabidopsis ceQORH protein with CaM**. Purification was performed on a CaM-Affinity Resin (Stratagene). As a control, the bacterial ecQOR protein was also tested. **S**; soluble bacterial proteins diluted in CaM-binding buffer, **P**; unbound proteins, **W**; wash and **E**; elution.



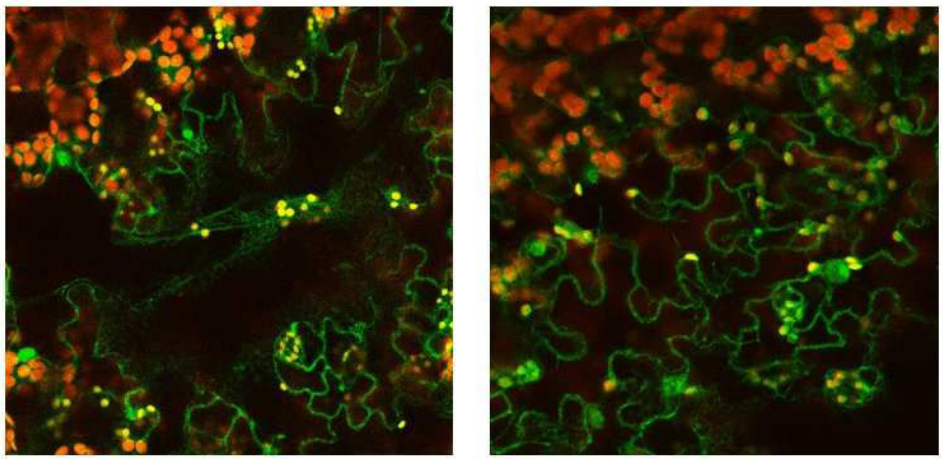
Supporting Figure 4. The CAM-binding domain is located in the *C*-terminus of ceQORH. A. Scheme of the successive deletions and constructs used to localize the CAM-binding domain in the ceQORH sequence. Sequences of the various deletions in ceQORH are indicated. 1, ceQORH protein, 2: ceQORH::GFP (full-length ceQORH protein fused to GFP), 3. Δ(1-59)ceQORH::GFP (ceQORH lacking 60 residues in *N*-terminus fused to GFP), 4: Δ(1-99)ceQORH::GFP (ceQORH lacking 100 residues in *N*-terminus fused to GFP), 5: (6-100)ceQORH::GFP (ceQORH lacking both 6 residues in *N*-ter and 229 residues in *C*-ter fused to GFP), 6: (60-100)ceQORH::GFP (ceQORH lacking both 60 residues in *N*-ter and 229 residues in *C*-ter fused to GFP), 7: Δ(280-329)ceQORH::GFP (ceQORH lacking 49 residues in *N*-ter fused to GFP), 8: Δ(229-329)ceQORH::GFP (ceQORH lacking 100 residues in

N-ter fused to GFP), 9: Δ(175-329)ceQORH::GFP (ceQORH lacking 146 residues in *N*-ter fused to GFP). The CAM-binding domain in the ceQORH sequence is deduced from ability of the various constructs to interact with CaM ("+" means CaM-binding and "-" means lack of CaM-binding) and supported by results of experiments shown in B and C. **B. SDS-PAGE:** Lanes 1 to 6 are crude bacterial extracts containing recombinant ceQORH fusions as described in **A**. M, prestained protein molecular weight markers (Fermentas). **Western blot:** Performed using a primary antibody raised against GFP (1/70000), a anti-rabbit HRP-conjugated secondary antibody (1/10000) and ECL. **CaM overlay:** After transfer, washing and saturation, the membrane was incubated in the hybridization solution containing 0.1 µg/ml biotinylated CAM. Detection of the bound CaM protein was performed using a Streptavidin-HRP conjugate and ECL. **C. Successive deletions in the *C*-terminus of ceQORH to localize its CaM-binding domain. SDS-PAGE:** Lanes 7 to 9 are crude bacterial extracts containing recombinant ceQORH fusions as described in **A**. M, prestained protein molecular weight markers (Fermentas). **CaM overlay:** After transfer, washing and saturation, the membrane was incubated in the hybridization solution containing 0.3 µg/ml of CAM::HRP conjugate. Detection of the bound CaM protein was performed using ECL.



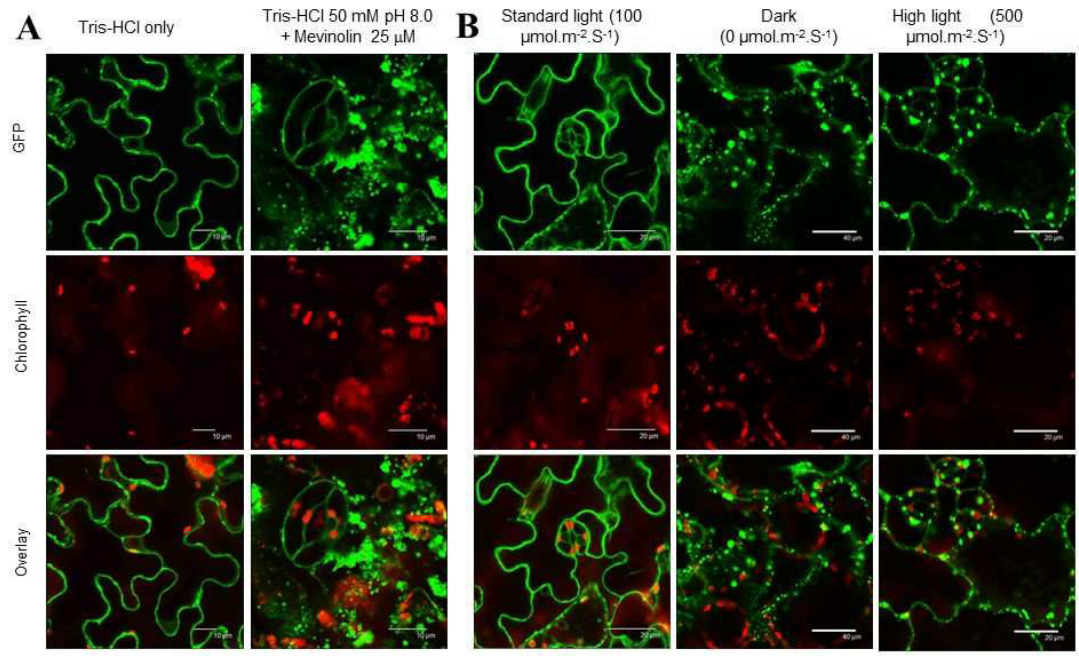
Supporting Figure 5. CaM5 and ceQORH co-localize in epidermal cells. Confocal microscopy was performed on plants stably expressing ceQORH::GFP (ceQORH fused to GFP) and transiently expressing CaM5::CFP (CaM5 fused to CFP). Chlorophyll; Chlorophyll autofluorescence. Overlay; Overlay of all three channels. Bar equals 40 mm. Note that while CaM5 and ceQORH are co-located in the plasma membrane, only CaM5 accumulates within the nucleus.

Video: Mut2-ceQORH accumulates in ER (BFA-bodies) in BFA treated plants

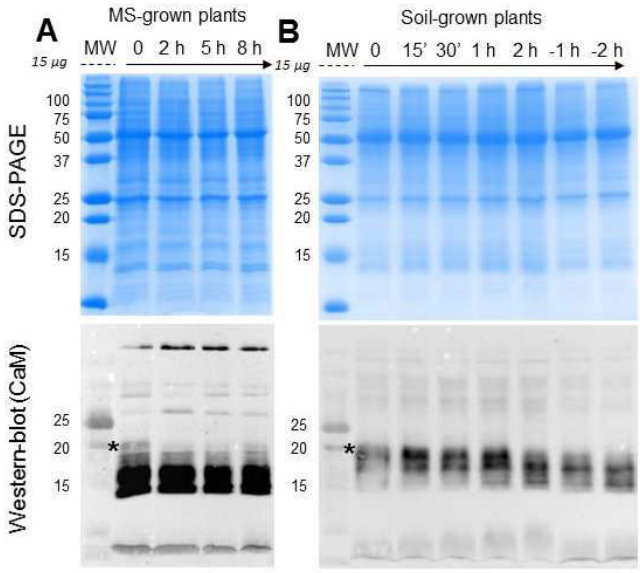


Supporting figure 6: Mut2-ceQORH::GFP accumulates in BFA bodies: the CaM-binding domain of ceQORH is not required for trafficking through ER. Arabidopsis plants stably expressing Mut2-ceQORH::GFP were treated

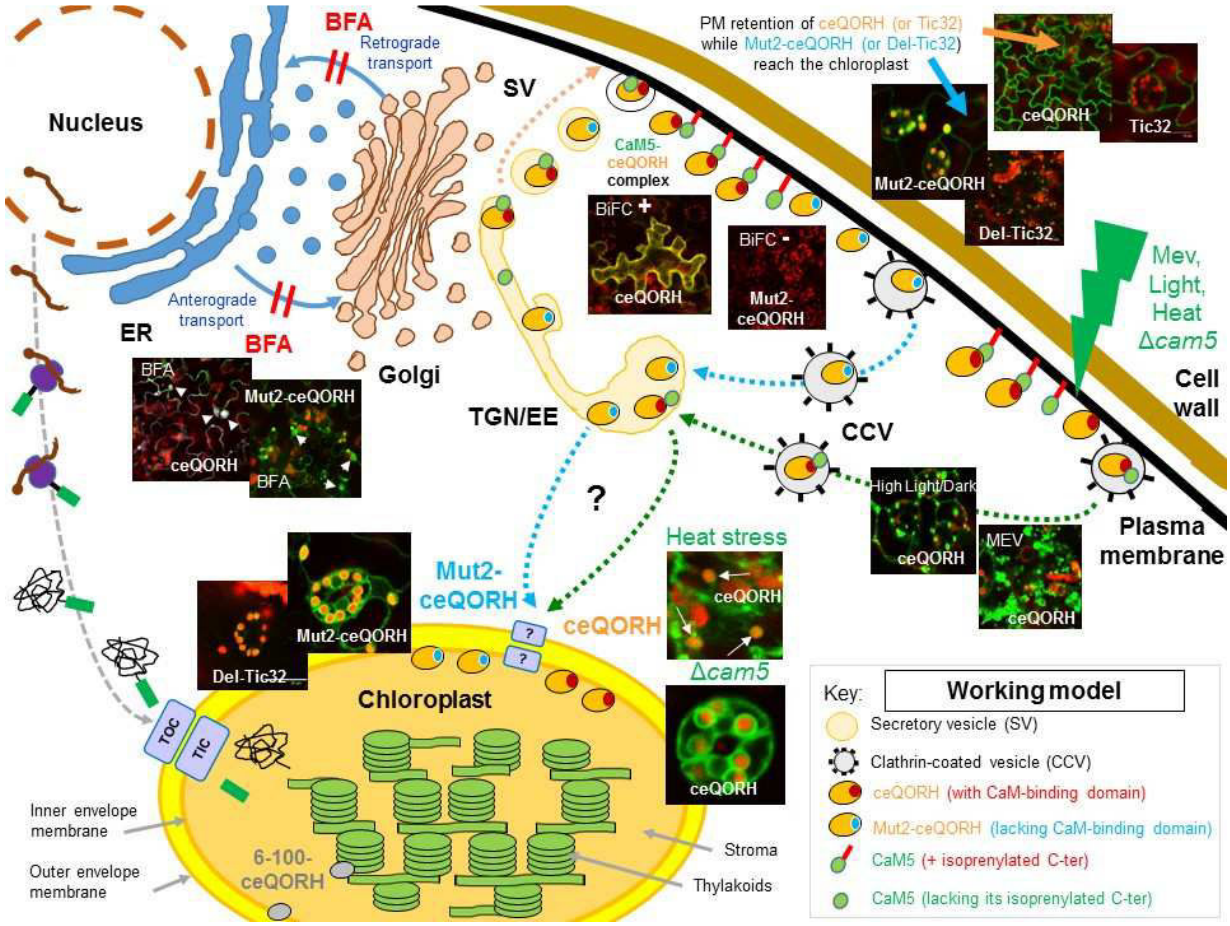
with BFA (180 µM). Imaging results from overlay of GFP fluorescence (ceQORH::GFP) and Chlorophyll autofluorescence.



Supporting Figure 7: Chemical and environmental control of ceQORH localization. A. Mevinolin treatment. Arabidopsis plants stably expressing ceQORH::GFP (plant ceQORH fused to GFP) were infiltrated with 25 mM Mevinolin in Tris-HCl 50 mM pH 8.0. As a control, the same plants were infiltrated with buffer only (Tris-HCl 50 mM pH 8). Blocking isoprenylation of CaM53 and its anchoring to the plasma membrane, Mevinolin has a strong impact on ceQORH::GFP localization in epidermal cells. However, note that while released from the plasma membrane ceQORH do not localize to plastids. Bars equal 10 mM. **B. Adverse light conditions**. Arabidopsis plants stably expressing ceQORH::GFP (plant ceQORH fused to GFP) were incubated for 0.5 or 3 days in the dark or grown under hight light (500 µmol.m^{-2}.S^{-1}, bars equal 20 mM). As a control, plants were maintained in standard light conditions (100 µmol.m^{-2}.S^{-1}, bars equal 20 mM). Again, note that while released from the plasma membrane ceQORH does not localize to plastids. **GFP**; GFP fluorescence, **Chlorophyll**; Chlorophyll autofluorescence. **Overlay:** Overlay of two channels.



Supporting Figure 8: Heat stress affects stability of CaM5 *in planta*. Arabidopsis plants were either pre-grown on MS-medium (**A**), or on soil (**B**), in growth rooms at 23°C (12-h light cycle) with a light intensity of 100 µmol.m^{-2}.S^{-1}. Plants were transferred to 42°C at the same light intensity. At 0 h, 2 h, 5 h et 8 h for MS-grown plants (**A**) and 15 min, 30 min, 1 h, 2 h, before cooling at 23°C for 1 h and 2 h for soil-grown plants (**B**), samples were taken and either immediately analyzed using confocal microscopy (see Figure 10), or stored in liquid nitrogen upon use for SDS-PAGE and western-blot detection of CaM. **SDS-PAGE**: whole cell proteins from Arabidopsis leave were extracted from heat-treated leaves, using protocol and buffers described for crude cell extracts (see Methods). Then, 15 µg of proteins from each crude cell extract were run on a SDS-PAGE. **Western-blot**: proteins were transferred to membranes using CaM transfer specific conditions (see Methods) and the CaM isoforms were detected using the CaM-767 antibody (1/5000 dilution). The secondary antibody (anti rabbit fused to HRP) was used at a 1/10,000 dilution, before detection using ECL. * indicates expected Mw of CaM5 (AT2G27030.3), 20,576 Da.



Supporting Figure 9: Working model depicting the regulatory mechanism for chloroplast targeting of CaM-binding proteins as deduced from schematic diagrams of the plant classical secretory pathway. Note that dotted arrows are known secretory pathways but their role in targeting of specific chloroplast proteins remain hypothetical. Indeed, unconventional protein secretion routes to the chloroplast from the ER cannot be excluded by similarity with examples of tonoplast proteins being delivered to the vacuole via different transport routes (Rose and Lee, Plant Phys 2012, Robinson and Pimpl, TIPS 2014) or through plastid-associated membrane (PLAM) sites of ER and chloroplast contacts (Belluci et al., J Exp Bot 2018). In this working model, ceQORH protein first transits through the endoplasmic reticulum (ER) as demonstrated from its colocalization with ER structures (rhodamine) and accumulation in BFA bodies (see Figure 1 and Supporting Figure 2). It may then leave the ER and traverse the Golgi stack (as a protein cargo). ceQORH was indeed detected in the Golgi using MS-based experiments (Nikolovski et al., Plant Physiol 2012). The next stage in this membrane flow process is the release of the trans-most cisternae, which becomes the trans-Golgi network (TGN) where ceQORH was also detected using MS-based experiments following affinity purification and a TGN marker (RABF2b/ARA7) as a bait (Heard et al., MCP 2015). As the TGN matures, ceQORH secretory cargo is sequestered into secretory vesicles (SV), which move to the plasma membrane (PM), plausibly in a coat protein II-vesicle-dependent manner. ceQORH was also repeatedly detected in the plasma membrane using MS-based experiments (*e.g.* Alexandersson et al, Plant cell Physiol 2004; Marmagne et al., MCP 2007). Owing to its CaM-binding property, ceQORH then interacts with CaM5 and is trapped at the PM (see Figure 5 and Supporting Figure 5). The same is true for Tic32, another CaM-binding plastid protein (see Figure 8). The TGN also acts as the early endosome, and receives incoming clathrin-coated vesicles (CCV) from the PM. Mut2-ceQORH, follows the same trafficking pathway, through the ER to the PM (see Figure 7 and Supporting Figure 6). However, lacking CaM5-binding properties, Mut2-ceQORH is not trapped at the PM and plausibly recycled to TGN through these CCV to reach

the plastids (see Figure 6). The same is true for Del-Tic32 lacking CaM-binding properties which is mostly detected in the plastids (see Figure 8). When isoprenylation of CaM5 is affected of after treatment with adverse light conditions, ceQORH is released from the PM and possibly reaches the secretory (as demonstrated for CaM53). However, while ceQORH accumulates in cellular vesicles, it does not localize to plastids (see Supporting Figure 7) thus suggesting that its CaM-binding properties limits further trafficking to plastids. After heat stress, affecting CaM5 stability (see Supporting Figure 8), or when CaM5 is deleted (*cam5* mutant), ceQORH is also released from the PM (see Figures 9 and 10, respectively) and, like Mut2-ceQORH or Del-Tic32, may follow the same hypothetical trafficking pathway, back to plastids. The following steps remains totally hypothetical. First, ceQORH or Mut2-ceQORH could use the CCV-dependent pathway back to the TGN. Indeed, Baslam et al. (Plant Signal Behav. 2016) have recently shown that TGN is involved in ER-Golgi-plastid trafficking. There, fragmentation of the TGN are known to separates the pre-multivesicular body (MVB) from CCV-forming domains, which release some CCVs for transport (recycling) of membrane proteins to the PM and other CCVs for the transport of proteins to other compartments (including the vacuole). Hypothetical fusion of the MVB with the plastid envelope then could deliver soluble proteins and some membrane proteins to the chloroplast envelope

Conclusion générale et perspectives

Conclusion générale et perspectives

Il existe deux catégories de protéines chloroplastiques codées par le génome nucléaire : *i*) des protéines plastidiales classiques, qui sont synthétisées dans le cytosol sous forme de précurseurs dotés d'une extension *N*-terminale clivable, et sont ensuite importées sans les chloroplastes via le système TOC/TIC [4, 5]; *ii*) des protéines dites « non-canoniques », plus récemment identifiées, qui sont dépourvues de peptide de transit clivable en *N*-terminal. La première évidence de telles protéines dites « non canoniques » a été apportée par notre équipe, lors d'études ciblant l'analyse du protéome de l'enveloppe du chloroplaste d'Arabidopsis. Ces études ont conduit à l'identification d'une protéine homologue à une quinone oxidoréductase nommée « ceQORH » [6, 7]. Bien que dépourvue de séquence de transit clivable, il s'est avéré que cette protéine est adressée à la membrane interne de l'enveloppe des plastes. D'autre part, il a également été montré que l'import de cette protéine dans le chloroplaste est médié par des machineries d'import alternatives indépendantes du système TOC/TIC [9]. De plus, il s'est avéré que cette protéine a la particularité d'être multilocalisée dans les cellules de différents tissus de la feuille [9]. Cependant, toutes ces études n'ont apporté aucune information quant à la machinerie d'import alternative dédiée à l'adressage de ceQORH aux chloroplastes, ni sur les mécanismes moléculaires qui contrôlent la localisation sub-cellulaire de cette protéine chloroplastique originale. En parallèle, l'équipe de Jurgen Soll a identifié une autre protéine (Tic32) de la membrane interne de l'enveloppe des plastes et démontré que cette protéine était aussi dépourvue de séquence de transit clivable en *N*-terminal. Des expériences similaires à celles réalisées en local ont aussi permis de montrer que cette protéine n'utilise pas la machinerie d'import TOC/TIC [8]. Enfin, une troisième protéine (NQR) a été identifiée plus récemment par notre équipe, qui est aussi dépourvue de séquence de transit clivable en *N*-terminal et pourtant importée dans le stroma des plastes (résultats non-publiés).

Les objectifs de ma thèse étaient, *i*) d'une part de contribuer à caractériser les composantes des systèmes d'import alternatifs de ces protéines chloroplastiques (ceQORH, Tic32, NQR) dites « non canoniques », et *ii*) d'autre part d'identifier les mécanismes moléculaires qui contrôlent la

multilocalisation cellulaire de ces protéines « non canoniques ». En parallèle de ces études ciblées, un 3ème objectif, visait à revisiter la composition du protéome de l'enveloppe du chloroplaste, à la recherche de gènes candidats pouvant être impliqués dans le contrôle de l'adressage de ces protéines « non canoniques » aux plastes.

Identification des récepteurs d'import alternatifs à la surface des chloroplastes impliqués dans l'import des protéines non canoniques : afin d'identifier les composantes des systèmes d'import alternatifs des protéines chloroplastiques « non canoniques », nous avons d'abord réalisé des constructions permettant de synthétiser quatre protéines « non canoniques » modèles fusionnées à la protéine GFP (ceQORH::GFP, Tic32::GFP, NQR::GFP et FSD1::GFP). Nous avons ensuite produit ces protéines chez *Escherichia coli* et les avons purifiées. Nous avons également accumulé de nombreuses préparations d'enveloppe de chloroplastes d'Arabidopsis pouvant être utilisées pour les réactions de pontage chimique avec les protéines recombinantes produites chez *Escherichia coli*. L'objectif de la réaction de pontage chimique était de piéger les récepteurs membranaires à la surface des chloroplastes avec leurs ligands (les quatre protéines recombinantes) en utilisant un agent chimique de pontage. Après la réaction de pontage, les complexes protéiques ont été solubilisés par des détergents et ensuite purifiés sur colonne d'affinité. Les fractions éluées à partir de la matrice d'affinité ont alors été analysées par spectrométrie de masse afin d'identifier les peptides spécifiques des partenaires co-purifiés (récepteurs d'import potentiels). Une complexité surprenante des échantillons analysés a été révélée. En effet, outre les complexes protéines/récepteurs attendus, de nombreuses autres protéines ont été aussi co-purifiées au cours de l'étape de chromatographie d'affinité. Il était donc difficile d'identifier, sans études supplémentaires, les récepteurs potentiels que nous espérions détecter spécifiquement dans chacune des quatre fractions éluées. Après traitement des données protéomiques générées (sélection des protéines les plus enrichies dans les fractions d'élution), une dizaine de partenaires potentiels ont été identifiés sur la base de plusieurs caractéristiques (localisation, fonction, structure…). Ensuite, nous avons testé l'implication potentielle des candidats sélectionnés dans le processus d'import alternatif des protéines chloroplastiques « non canoniques ».

Pour ce faire, des mutants d'insertions d'*Arabidopsis thaliana* déficients dans l'expression de ces partenaires/récepteurs candidats ont été isolés. Parmi les gènes candidats sélectionnés, quatre

mutants correspondants ont révélé un phénotype embryon létal suggérant qu'ils ne pouvaient pas être viables à l'état homozygote.

Pour les mutants viables à l'état homozygote, une transformation de ces lignées a été réalisée permettant une expression stable des protéines chloroplastiques « non canoniques » fusionnées à la protéine GFP. Des plantes sauvages ont été également transformées avec les mêmes constructions et dans les mêmes conditions, au titre de contrôles. L'analyse de l'impact des mutations sur la localisation subcellulaire des protéines « non canoniques » a été réalisée par microscopie confocale directement sur des feuilles des rosettes des lignées mutantes transformées. Il s'est avéré que **la localisation subcellulaire de la protéine ceQORH était fortement affectée dans un mutant (KO) possédant une insertion dans le gène codant une protéine nommée « ClpD ».** En effet, contrairement aux plantes sauvages dans lesquelles la protéine ceQORH est préférentiellement localisée dans la membrane plasmique des cellules d'épiderme de la feuille, nous avons observé une suraccumulation de la protéine ceQORH à la surface des chloroplastes dans le mutant clpD. Il est important de rappeler ici que le même profil de localisation subcellulaire a été observé sur trois transformants indépendants. La **suraccumulation de ceQORH à la surface des chloroplastes et notamment à l'interface avec le noyau** (apparition d'amas révélés par la fluorescence de la GFP) nous a conduits à soupçonner une corrélation de la fluorescence observée avec le réticulum endoplasmique (RE) dans le mutant clpD. Cette hypothèse a été ensuite validée par imagerie confocale à l'aide d'un marquage spécifique du RE à l'aide de la Rhodamine. Ces résultats suggèrent d'une part, que la protéine chloroplastique **ClpD joue un rôle central dans le contrôle de la localisation subcellulaire de ceQORH** et d'autre part, conforte nos autres observations suggérant que **la protéine ceQORH intègre le RE avant son transit dans les chloroplastes**. Ces observations suggèrent aussi que **l'absence de ClpD ralentit fortement le trafic vésiculaire de la protéine ceQORH, qui s'accumule donc en surface des plastes**. Dans ce contexte, il est important de rappeler que la protéine ClpD fait partie du complexe protéolytique Clp (constitué de ClpC1, ClpC2, ClpB…) localisé dans le stroma des chloroplastes. Il a été récemment démontré que l'absence du composant ClpC1 (faisant partie de la famille des protéines de type ClpD) entraine une diminution de l'import, dans les plastes, des précurseurs de protéines plastidiales « classiques » utilisant la voie TOC/TIC [171, 172]. Il a aussi été récemment démontré que, tout comme la protéine ClpC2, la sous-unité ClpD a la capacité de s'associer avec les peptides de transit de précurseurs chloroplastiques [153]. Il existe donc des données indiquant un lien entre

des composants du complexe Clp et l'import des protéines dans les chloroplastes. Toutefois, ce lien n'est établi que pour les protéines utilisant la voie classique pour leur import dans les plastes. L'originalité de nos résultats vient du fait qu'ils concernent le transport de protéines « non canoniques » et que ces données nous ont permis de soupçonner l'implication indirecte de ClpD dans le contrôle de l'adressage de ceQORH à savoir son intervention au niveau du trafic vésiculaire vers les plastes. Cependant, cette hypothèse implique soit *i*) un effet indirect, comme cité plus haut, lié au ralentissement du trafic vers les plastes, soit *ii*) une localisation de la protéine ClpD à l'extérieur du chloroplaste qui participe au processus d'import. Bien que la protéine ClpD soit considérée par certains comme uniquement localisée dans le stroma [178, 179], ClpD est partagée entre le stroma et l'enveloppe du chloroplaste dans la base de données AT_CHLORO [12]. Il est ainsi envisageable que ClpD soit localisée dans plusieurs compartiments de la cellule et cette hypothèse est en cours d'analyse dans l'équipe.

Nous avons également poursuivi l'analyse de l'impact de l'absence de certains partenaires/récepteurs potentiels sur la localisation des protéines « non canoniques » d'intérêt sur les mutants présentant un phénotype embryon létal et cités plus haut. Parmi ces candidats, se trouvent deux autres composants du complexe Clp : ClpB3 et ClpR3 qui étaient aussi enrichis dans la fraction issue du pontage chimique avec la protéine ceQORH et dont les mutants correspondants ne sont pas viables à l'état homozygote. Il était donc difficile de tester leur implication potentielle dans l'import de la protéine ceQORH.

Dans ce contexte, nous avons eu recours à une première approche alternative reposant sur des approches d'imagerie confocale réalisées sur les graines issues de mutants hétérozygotes (en ciblant les graines plus claires, correspondant probablement à un état homozygote pour la mutation). Cependant, la co-localisation de la fluorescence de la protéine ceQORH fusionnée à la GFP et de l'autofluorescence de la chlorophylle n'a pas été facile à établir. En conséquence, il était difficile d'exploiter les résultats obtenus à la date de la rédaction du présent rapport et d'en tirer des conclusions définitives sur le rôle de ces protéines. Pour ces mutants létaux à l'état homozygote, nous avons donc initié une approche reposant sur l'utilisation de promoteurs inductibles visant à les complémenter pendant les phases initiales du développement de la plante. Ces approches ont été initiées et seront poursuivies par l'équipe.

Au-delà de l'identification de récepteurs potentiels, ces études protéomiques réalisées sur des fractions d'enveloppe purifiée ont également généré des données originales qui permettaient de

revisiter la composition protéique de l'enveloppe des chloroplastes. Sur la base de ces résultats, nous avons initié une approche gène candidats qui visait également à identifier les récepteurs d'import des protéines « non canoniques » du chloroplaste. Cette approche reposait sur la bio-analyse de l'enveloppe du chloroplaste d'*Arabidopsis thaliana.*

Etude de protéome de l'enveloppe du chloroplaste d'*Arabidosis thaliana* : L'enveloppe des plastes constitue un site privilégié de communication entre le chloroplaste et le cytosol [11]. Cependant, ce sous-compartiment chloroplastique reste la partie la moins bien connue du protéome chloroplastique et la fonction de très nombreuses protéines de l'enveloppe est encore inconnue. De plus, les connaissances de la compartimentation du métabolisme cellulaire suggèrent que l'enveloppe contient de nombreuses protéines qui n'ont pas été identifiées par les approches protéomiques ciblant ce compartiment [14].

En effet, l'ensemble de deux membranes externe et interne de l'enveloppe ne contient que 1 à 2% des protéines du chloroplaste [13]. En conséquence, les protéines de l'enveloppe sont très peu abondantes à l'échelle du chloroplaste comme à l'échelle de la cellule végétale. Par conséquence, les protéines de l'enveloppe des plastes sont très enrichies dans les fractions d'enveloppe purifiée comparativement à leur abondance dans un extrait cellulaire total. Ce facteur d'enrichissement n'avait encore jamais été exploité dans les études antérieures ciblant le protéome de l'enveloppe du chloroplaste. Il était donc difficile de distinguer les vraies protéines de l'enveloppe du chloroplaste de contaminants dérivant d'autres compartiments cellulaires. Dans cette partie du projet, nous avons tiré profit de la sensibilité accrue de la spectrométrie de masse et de la disponibilité de plusieurs bases de données de protéomique subcellulaire en réalisant **une étude de protéomique quantitative basée sur la comparaison des fractions d'enveloppe purifiée à des extraits cellulaires totaux.** Pour cela, nous avons préparé trois extraits cellulaires totaux et purifié trois fractions d'enveloppe d'Arabidopsis. Nous avons ensuite analysé ces fractions par spectrométrie de masse. Afin d'explorer les données protéomiques générées, une annotation manuelle de la localisation sub-cellulaire et sub-plastidiale, ainsi que de la fonction des protéines identifiées, a été réalisée en se basant sur plusieurs bases de données (PubMed, TAIR, NCBI, Uniprot, SUBA3, MASCP GATOR…) et plusieurs outils de prédiction de localisation (ChloroP, TargetP…). Le nombre de spectres de chaque protéine, dans chaque répliquât biologique, a été utilisé pour calculer **un facteur d'enrichissement EF** qui correspond au rapport de la somme des

nombres de spectres dans les trois fractions de l'enveloppe par rapport à la somme des nombres de spectres dans les trois extraits cellulaires totaux (EF = norm_WSC_Env / norm_WSC_CCE). Ce paramètre s'est avéré très utile pour **estimer l'enrichissement des protéines détectées dans les fractions d'enveloppe par rapport à l'extrait cellulaire total et par conséquent pour l'annotation des vraies protéines de l'enveloppe**. Afin de tester la fiabilité de ce paramètre d'enrichissement utilisé, un « safe set » a été créé, qui contenait des protéines dont la localisation était connue et reposait sur la base de données expérimentales, comme des publications fournissant des informations de localisation (imagerie confocale après expression d'une protéine fusionnée à la protéine fluorescente GFP), le fait que ces protéines soient codées par le génome chloroplastique (ou mitochondrial), l'identification préalable et sans ambiguïté d'une localisation sub-plastidiale (issue de la base de données AT_CHLORO), ou encore un consensus des outils de prédiction de localisation issu de la base de données SUBA3 qui devait être en accord avec les données d'imagerie confocale (fusions GFP). Comme attendu, les vraies protéines de l'enveloppe présentes dans ce « safe set » étaient fortement enrichies dans la fraction enveloppe purifiée par rapport à l'extrait cellulaire total. Cet enrichissement était moins important pour les protéines partagées entre l'enveloppe et les autres sous-compartiments du chloroplaste (stroma et thylakoïde). Par comparaison, les protéines issues des autres compartiments cellulaires ou plastidiaux présentaient un faible facteur d'enrichissement (à l'exception de certaines protéines de la vacuole et du réticulum endoplasmique). Ainsi, après avoir validé la bonne corrélation entre la localisation subcellulaire des protéines et leurs niveaux enrichissement respectifs sur ce « safe set », **ce facteur d'enrichissement « EF » a été utilisé pour classifier l'ensemble des protéines détectées** et plus particulièrement les protéines dont la localisation ne pouvait être déduite des études antérieures (protéines nouvellement identifiées…). A l'issue du traitement des données générées, environ 1269 protéines ont été détectées au moins une fois dans l'enveloppe dont **433 protéines nouvellement identifiées**.

Parmi les 433 nouvelles protéines, **87 protéines spécifiques de l'enveloppe ont été identifiées. Parmis ces 87 protéines, certaines ont été considérées comme de bons candidats pouvant être impliqués dans l'adressage au chloroplaste de protéines « non canoniques » par des voies d'import alternatives**. Il s'agit de protéines membranaires homologues de composants des systèmes d'import TAT et SEC (localisés dans les membranes des thylakoïdes, voir introduction) [62-64]. Malheureusement, là encore les mutants d'Arabidopsis affecté dans l'expression des gènes

correspondants se sont avérés létaux à l'état homozygote. En conséquence, la même approche que celle qui est décrite ci-dessus (complémentation par des promoteurs inductibles) a été reproduite sur ces mutants et les résultats ne sont pas encore disponibles à la date de la rédaction de ce rapport.

Cette analyse du protéome de l'enveloppe nous a aussi permis de révéler une **contamination importante de nos fractions d'enveloppe purifiées par des composants du RE-Golgi. Cette contamination est assez importante comparativement aux autres compartiments** de la cellule. Ces résultats suggèrent **qu'il existe une interaction potentielle du RE avec la surface des chloroplastes** ouvrant ainsi des pistes de recherche très intéressantes. Ces données sont parfaitement en accord avec les observations issues de la première partie de la thèse montrant que la protéine ceQORH s'accumule dans le RE (dans le mutant ClpD) avant son import dans les chloroplastes. Les données générées par la présente étude ont également permis **de mettre à jour la base de données « AT_CHLORO », créée par notre équipe en 2010** et actualisée régulièrement.

Contrôle de l'adressage plastidial de la protéine ceQORH : au cours des dix dernières années, l'équipe s'est intéressée à étudier les mécanismes moléculaires impliqués dans le contrôle de la localisation subcellulaire de la protéine ceQORH, projet dans lequel je me suis impliquée depuis la deuxième année de ma thèse. Initialement, les collègues de l'équipe avaient observé (imagerie confocale) que la protéine ceQORH avait la particularité d'être multilocalisée dans la cellule. Ceci était notamment vrai dans les cellules de l'épiderme de la feuille où la protéine n'est pas adressée aux chloroplastes, alors qu'elle est majoritairement localisée dans l'enveloppe du chloroplaste dans les cellules du parenchyme [9]. Afin d'expliquer cette multilocalisation dans l'épiderme, notre équipe avait d'abord émis l'hypothèse d'une double localisation entre le chloroplaste et la mitochondrie. Cette hypothèse avait été rejetée suite à l'utilisation d'un marqueur spécifique de la mitochondrie qui avait montré l'absence de colocalisation entre la fluorescence de la protéine GFP fusionnée à la protéine ceQORH et celle du marqueur mitochondrial (MitoTracker). Dans le même temps, la protéine ceQORH avait été détectée à plusieurs reprises dans la membrane plasmique [181-183] lors d'analyses protéomiques ciblant des fractions purifiées de membrane plasmique. Une étude de protéomique plus récente [184] ciblant le système endoplasmique avait également permis de détecter la protéine ceQORH dans le Golgi [184]. Cependant, il est important de noter que, dans toutes ces études antérieures, la détection de la protéine ceQORH avait été considérée

comme résultant de la contamination des fractions subcellulaires analysées. Ces résultats nous ont conduits à soupçonner une interaction probable entre le système endomembranaire, et plus particulièrement le RE et la protéine ceQORH. Nous avons donc utilisé un marqueur spécifique du RE (la Rhodamine) sur des plantes transgéniques exprimant la protéine ceQORH fusionnée à la GFP. Cette analyse, réalisée par imagerie confocale a révélé **une parfaire corrélation entre la localisation du RE et de la protéine ceQORH**. Il est très important de noter que **ces résultats ont fortement orienté les analyses détaillées dans les deux premières parties de mon travail de thèse (en particulier, en questionnant la co-localisation du RE et de la protéine ceQORH à la surface des plastes dans le mutant ClpD, ou en nous orientant vers une bioanalyse plus fine des composants détectés dans l'enveloppe par analyse protéomique et potentiellement localisés dans le RE)**.

Les membres de notre équipe avaient aussi préalablement démontré que la protéine ceQORH possède un domaine de liaison à la calmoduline. Il avait été aussi démontré que le domaine de liaison de la protéine ceQORH à la calmoduline est localisé dans l'extrémité *C*-terminale de la séquence primaire de la protéine ceQORH. Une version mutante (Mut2-ceQORH) de la protéine ceQORH, qui ne possède plus la capacité de se lier à la calmoduline avait aussi été créée. Afin d'expliquer la localisation de la protéine ceQORH dans la membrane plasmique des cellules de l'épiderme (observée en microscopie confocale), notre équipe a émis l'hypothèse d'une interaction potentielle entre la protéine ceQORH et une isoforme de la famille des calmodulines qui pouvait être localisée dans la membrane plasmique. Diverses analyses (incluant des expériences de complémentation de type « BiFC ») ont ensuite été réalisées qui ont permis de démontrer que cette isoforme était la calmoduline CAM5. Il a ensuite été montré que contrairement à la protéine de fusion ceQORH::GFP qui est localisée dans la membrane plasmique dans les cellules de l'épiderme, la version mutante de la protéine ceQORH (Mut2::GFP) est adressée à l'enveloppe du chloroplaste. Ces résultats ont permis de démontrer que **la protéine CAM5 joue un rôle essentiel dans le contrôle de la localisation subcellulaire de la protéine ceQORH dans les cellules de l'épiderme**. Comme il existe de nombreuses (environ 50) isoformes de la calmoduline dans les cellules végétales, il était nécessaire de valider que la protéine CAM5 est l'unique isoforme de calmoduline qui contrôle la rétention de ceQORH dans la membrane plasmique. Pour ce faire, nous avons analysé l'impact de l'absence de CAM5 sur la localisation subcellulaire des protéines ceQORH::GFP et Mut2::GFP. Contrairement à sa localisation dans la membrane plasmique dans

les plantes sauvages, ceQORH::GFP est localisée dans l'enveloppe des chloroplastes dans le mutant cam5. Ce résultat valide le rôle prépondérant de **CAM5 dans le contexte, et suggère que CAM5 est la seule isoforme de calmoduline qui soit directement impliquée dans le contrôle de la localisation subcellulaire de la protéine ceQORH**.

En conclusion, les différentes approches initiées se sont avérées parfaitement complémentaires et se sont nourries les unes des autres. Les résultats obtenus à ce jour convergent toutes vers l'idée que la voie d'import alternative utilisée par la protéine ceQORH lors de son adressage aux plastes implique un trafic vésiculaire par le RE, et que ce trafic est régulé par deux acteurs moléculaires, la protéine CaM5 et la protéine ClpD.

Perspectives…

Les perspectives de ce travail sont nombreuses, tant sur la validation de l'implication potentielle de la protéine ClpD dans le contrôle de l'adressage de la protéine ceQORH, que sur l'acheminement de la protéine ceQORH via le trafic vésiculaire avant son import dans le chloroplaste.

Afin de valider le rôle de ClpD dans le contrôle de l'adressage de ceQORH, nous avons initié une approche reposant sur la protéomique différentielle à l'échelle du chloroplaste sur des plantes sauvages et mutantes affectés dans l'expression de la protéine ClpD. Ces analyses permettront de déterminer l'impact de cette mutation sur l'import de la protéine ceQORH, et d'autres protéines chloroplastiques, dans la plante sauvage et le mutant ClpD.

L'hypothèse d'un contact direct entre ClpD et ceQORH (issue des expériences de pontage chimique) implique que ClpD puisse être aussi localisée à la surface du plaste (et pas seulement dans le stroma). Pour cela, nous avons réalisé une construction permettant d'expression de la protéine ClpD fusionnée à la protéine GFP afin d'analyser la localisation subcellulaire de la protéine ClpD *in planta*.

En parallèle, il sera également nécessaire de comprendre la nature de l'interaction entre le RE et la protéine ceQORH. Il est important ici de signaler que les données générées par les expériences de

pontage chimique et la purification par affinité, ont révélé la présence de la clathrine (protéine impliquée dans le trafic vésiculaire) dans la fraction éluée en utilisant la protéine ceQORH comme appât. Cela nous incite à réfléchir sur le statut exact de la protéine ceQORH dans le système vésiculaire : cette protéine joue-t-elle un rôle actif dans le trafic vésiculaire ou est-elle simplement transportée passivement dans les (ou à la surface des) vésicules ?

Il sera enfin nécessaire de généraliser les observations obtenues avec le mutant ClpD en testant également l'impact de cette mutation sur la localisation subcellulaire des deux autres protéines « non canoniques », NQR et Tic32. De même, l'analyse de l'impact de l'absence de deux autres sous-unités du complexe Clp (ClpB3 et ClpR3), devra être réalisée. Pour cela, nous avons initié une approche de complémentation des mutants embryons létaux à l'aide des promoteurs inductibles par éthanol devra être poursuivie. Cette même approche sera aussi appliquée aux candidats récepteurs issus de l'approche gène candidats, et codant les protéines TAT et Sec.

Références bibliographiques

1. Bogorad, L., *Evolution of early eukaryotic cells: genomes, proteomes, and compartments.* Photosynth Res, 2008. **95**(1): p. 11-21.
2. Reyes-Prieto, A., A.P. Weber, and D. Bhattacharya, *The origin and establishment of the plastid in algae and plants.* Annu Rev Genet, 2007. **41**: p. 147-68.
3. Yoon, H.S., et al., *A molecular timeline for the origin of photosynthetic eukaryotes.* Mol Biol Evol, 2004. **21**(5): p. 809-18.
4. Jarvis, P. and E. Lopez-Juez, *Biogenesis and homeostasis of chloroplasts and other plastids.* Nat Rev Mol Cell Biol, 2013. **14**(12): p. 787-802.
5. Lee, D.W., J. Lee, and I. Hwang, *Sorting of nuclear-encoded chloroplast membrane proteins.* Curr Opin Plant Biol, 2017. **40**: p. 1-7.
6. Ferro, M., et al., *Proteomics of the chloroplast envelope membranes from Arabidopsis thaliana.* Mol Cell Proteomics, 2003. **2**(5): p. 325-45.
7. Miras, S., et al., *Non-canonical transit peptide for import into the chloroplast.* J Biol Chem, 2002. **277**(49): p. 47770-8.
8. Nada, A. and J. Soll, *Inner envelope protein 32 is imported into chloroplasts by a novel pathway.* J Cell Sci, 2004. **117**(Pt 17): p. 3975-82.
9. Miras, S., et al., *Toc159- and Toc75-independent import of a transit sequence-less precursor into the inner envelope of chloroplasts.* J Biol Chem, 2007. **282**(40): p. 29482-92.
10. Bock, R. and J.N. Timmis, *Reconstructing evolution: gene transfer from plastids to the nucleus.* Bioessays, 2008. **30**(6): p. 556-66.
11. Block, M.A., et al., *Chloroplast envelope membranes: a dynamic interface between plastids and the cytosol.* Photosynth Res, 2007. **92**(2): p. 225-44.
12. Vojta, A., et al., *The protein translocon of the plastid envelopes.* J Biol Chem, 2004. **279**(20): p. 21401-5.
13. Joyard, J., et al., *Envelope membranes from mature spinach chloroplasts contain a NADPH:protochlorophyllide reductase on the cytosolic side of the outer membrane.* J Biol Chem, 1990. **265**(35): p. 21820-7.
14. Rolland, N., et al., *The biosynthetic capacities of the plastids and integration between cytoplasmic and chloroplast processes.* Annu Rev Genet, 2012. **46**: p. 233-64.
15. Chotewutmontri, P., K. Holbrook, and B.D. Bruce, *Plastid Protein Targeting: Preprotein Recognition and Translocation.* Int Rev Cell Mol Biol, 2017. **330**: p. 227-294.
16. Bruce, B.D., *Chloroplast transit peptides: structure, function and evolution.* Trends Cell Biol, 2000. **10**(10): p. 440-7.
17. Bruce, B.D., *The paradox of plastid transit peptides: conservation of function despite divergence in primary structure.* Biochim Biophys Acta, 2001. **1541**(1-2): p. 2-21.
18. Douwe de Boer, A. and P.J. Weisbeek, *Chloroplast protein topogenesis: import, sorting and assembly.* Biochim Biophys Acta, 1991. **1071**(3): p. 221-53.
19. Zhang, X.P. and E. Glaser, *Interaction of plant mitochondrial and chloroplast signal peptides with the Hsp70 molecular chaperone.* Trends Plant Sci, 2002. **7**(1): p. 14-21.
20. Beckmann, R.P., L.E. Mizzen, and W.J. Welch, *Interaction of Hsp 70 with newly synthesized proteins: implications for protein folding and assembly.* Science, 1990. **248**(4957): p. 850-4.
21. May, T. and J. Soll, *14-3-3 proteins form a guidance complex with chloroplast precursor proteins in plants.* Plant Cell, 2000. **12**(1): p. 53-64.

22. Jarvis, P. and F. Kessler, *Mechanisms of Chloroplast Protein Import in Plants.* 2014: p. 241-270.
23. Karlin-Neumann, G.A. and E.M. Tobin, *Transit peptides of nuclear-encoded chloroplast proteins share a common amino acid framework.* Embo j, 1986. **5**(1): p. 9-13.
24. Shi, L.X. and S.M. Theg, *The chloroplast protein import system: from algae to trees.* Biochim Biophys Acta, 2013. **1833**(2): p. 314-31.
25. Bohnert, M., N. Pfanner, and M. van der Laan, *A dynamic machinery for import of mitochondrial precursor proteins.* FEBS Lett, 2007. **581**(15): p. 2802-10.
26. Jarvis, P., *Targeting of nucleus-encoded proteins to chloroplasts in plants.* New Phytologist, 2008. **179**(2): p. 257-285.
27. Neupert, W. and M. Brunner, *The protein import motor of mitochondria.* Nat Rev Mol Cell Biol, 2002. **3**(8): p. 555-65.
28. Schnell, D.J., F. Kessler, and G. Blobel, *Isolation of components of the chloroplast protein import machinery.* Science, 1994. **266**(5187): p. 1007-12.
29. Schnell, D.J., et al., *A consensus nomenclature for the protein-import components of the chloroplast envelope.* Trends Cell Biol, 1997. **7**(8): p. 303-4.
30. Becker, T., et al., *Toc12, a novel subunit of the intermembrane space preprotein translocon of chloroplasts.* Mol Biol Cell, 2004. **15**(11): p. 5130-44.
31. Gutensohn, M., et al., *Functional analysis of the two Arabidopsis homologues of Toc34, a component of the chloroplast protein import apparatus.* Plant J, 2000. **23**(6): p. 771-83.
32. Jarvis, P., et al., *An Arabidopsis mutant defective in the plastid general protein import apparatus.* Science, 1998. **282**(5386): p. 100-3.
33. Hiltbrunner, A., et al., *Protein translocon at the Arabidopsis outer chloroplast membrane.* Biochem Cell Biol, 2001. **79**(5): p. 629-35.
34. Jackson-Constan, D. and K. Keegstra, *Arabidopsis genes encoding components of the chloroplastic protein import apparatus.* Plant Physiol, 2001. **125**(4): p. 1567-76.
35. Jarvis, P. and J. Soll, *Toc, tic, and chloroplast protein import.* Biochim Biophys Acta, 2002. **1590**(1-3): p. 177-89.
36. Kubis, S., et al., *Functional specialization amongst the Arabidopsis Toc159 family of chloroplast protein import receptors.* Plant Cell, 2004. **16**(8): p. 2059-77.
37. Constan, D., et al., *An outer envelope membrane component of the plastid protein import apparatus plays an essential role in Arabidopsis.* Plant J, 2004. **38**(1): p. 93-106.
38. Kubis, S., et al., *The Arabidopsis ppi1 mutant is specifically defective in the expression, chloroplast import, and accumulation of photosynthetic proteins.* Plant Cell, 2003. **15**(8): p. 1859-71.
39. Demarsy, E., A.M. Lakshmanan, and F. Kessler, *Border control: selectivity of chloroplast protein import and regulation at the TOC-complex.* Front Plant Sci, 2014. **5**: p. 483.
40. Ivanova, Y., et al., *Members of the Toc159 import receptor family represent distinct pathways for protein targeting to plastids.* Mol Biol Cell, 2004. **15**(7): p. 3379-92.
41. Bauer, J., et al., *The major protein import receptor of plastids is essential for chloroplast biogenesis.* Nature, 2000. **403**(6766): p. 203-7.
42. Qbadou, S., et al., *Toc64--a preprotein-receptor at the outer membrane with bipartide function.* J Mol Biol, 2007. **367**(5): p. 1330-46.
43. Aronsson, H., et al., *Toc64/OEP64 is not essential for the efficient import of proteins into chloroplasts in Arabidopsis thaliana.* Plant J, 2007. **52**(1): p. 53-68.
44. Rosenbaum Hofmann, N. and S.M. Theg, *Toc64 is not required for import of proteins into chloroplasts in the moss Physcomitrella patens.* Plant J, 2005. **43**(5): p. 675-87.

45. Zufferey, M., et al., *The novel chloroplast outer membrane kinase KOC1 is a required component of the plastid protein import machinery.* J Biol Chem, 2017. **292**(17): p. 6952-6964.
46. Hormann, F., et al., *Tic32, an essential component in chloroplast biogenesis.* J Biol Chem, 2004. **279**(33): p. 34756-62.
47. Kessler, F. and G. Blobel, *Interaction of the protein import and folding machineries of the chloroplast.* Proc Natl Acad Sci U S A, 1996. **93**(15): p. 7684-9.
48. Kessler, F. and D. Schnell, *Chloroplast biogenesis: diversity and regulation of the protein import apparatus.* Curr Opin Cell Biol, 2009. **21**(4): p. 494-500.
49. Li, H.M. and C.C. Chiu, *Protein transport into chloroplasts.* Annu Rev Plant Biol, 2010. **61**: p. 157-80.
50. Schwenkert, S., J. Soll, and B. Bolter, *Protein import into chloroplasts--how chaperones feature into the game.* Biochim Biophys Acta, 2011. **1808**(3): p. 901-11.
51. Kikuchi, S., et al., *Uncovering the protein translocon at the chloroplast inner envelope membrane.* Science, 2013. **339**(6119): p. 571-4.
52. Nakai, M., *The TIC complex uncovered: The alternative view on the molecular mechanism of protein translocation across the inner envelope membrane of chloroplasts.* Biochim Biophys Acta, 2015. **1847**(9): p. 957-67.
53. Kikuchi, S., et al., *A 1-megadalton translocation complex containing Tic20 and Tic21 mediates chloroplast protein import at the inner envelope membrane.* Plant Cell, 2009. **21**(6): p. 1781-97.
54. Kohler, D., et al., *Characterization of chloroplast protein import without Tic56, a component of the 1-megadalton translocon at the inner envelope membrane of chloroplasts.* Plant Physiol, 2015. **167**(3): p. 972-90.
55. Scott, S.V. and S.M. Theg, *A new chloroplast protein import intermediate reveals distinct translocation machineries in the two envelope membranes: energetics and mechanistic implications.* J Cell Biol, 1996. **132**(1-2): p. 63-75.
56. Flores-Perez, U. and P. Jarvis, *Molecular chaperone involvement in chloroplast protein import.* Biochim Biophys Acta, 2013. **1833**(2): p. 332-40.
57. Inaba, T. and D.J. Schnell, *Protein trafficking to plastids: one theme, many variations.* Biochem J, 2008. **413**(1): p. 15-28.
58. Vojta, L., J. Soll, and B. Bolter, *Protein transport in chloroplasts - targeting to the intermembrane space.* Febs j, 2007. **274**(19): p. 5043-54.
59. Nussaume, L., *Chloroplast SRP takes another road.* Nat Chem Biol, 2008. **4**(9): p. 529-31.
60. Yi, L. and R.E. Dalbey, *Oxa1/Alb3/YidC system for insertion of membrane proteins in mitochondria, chloroplasts and bacteria (review).* Mol Membr Biol, 2005. **22**(1-2): p. 101-11.
61. Michl, D., et al., *Targeting of proteins to the thylakoids by bipartite presequences: CFoII is imported by a novel, third pathway.* Embo j, 1994. **13**(6): p. 1310-7.
62. Eichacker, L.A. and R. Henry, *Function of a chloroplast SRP in thylakoid protein export.* Biochim Biophys Acta, 2001. **1541**(1-2): p. 120-34.
63. Asakura, Y., S. Kikuchi, and M. Nakai, *Non-identical contributions of two membrane-bound cpSRP components, cpFtsY and Alb3, to thylakoid biogenesis.* Plant J, 2008. **56**(6): p. 1007-17.
64. Braun, N.A. and S.M. Theg, *The chloroplast Tat pathway transports substrates in the dark.* J Biol Chem, 2008. **283**(14): p. 8822-8.

65. Hofmann, N.R. and S.M. Theg, *Protein- and energy-mediated targeting of chloroplast outer envelope membrane proteins.* Plant J, 2005. **44**(6): p. 917-27.
66. Krogh, A., et al., *Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.* J Mol Biol, 2001. **305**(3): p. 567-80.
67. Li, H.M. and L.J. Chen, *Protein targeting and integration signal for the chloroplastic outer envelope membrane.* Plant Cell, 1996. **8**(11): p. 2117-26.
68. Muckel, E. and J. Soll, *A protein import receptor of chloroplasts is inserted into the outer envelope membrane by a novel pathway.* J Biol Chem, 1996. **271**(39): p. 23846-52.
69. Li, H. and L.J. Chen, *A novel chloroplastic outer membrane-targeting signal that functions at both termini of passenger polypeptides.* J Biol Chem, 1997. **272**(16): p. 10968-74.
70. Inoue, K. and K. Keegstra, *A polyglycine stretch is necessary for proper targeting of the protein translocation channel precursor to the outer envelope membrane of chloroplasts.* Plant J, 2003. **34**(5): p. 661-9.
71. Tranel, P.J., et al., *A component of the chloroplastic protein import apparatus is targeted to the outer envelope membrane via a novel pathway.* Embo j, 1995. **14**(11): p. 2436-46.
72. Jarvis, P. and C. Robinson, *Mechanisms of protein import and routing in chloroplasts.* Curr Biol, 2004. **14**(24): p. R1064-77.
73. Armbruster, U., et al., *Chloroplast proteins without cleavable transit peptides: rare exceptions or a major constituent of the chloroplast proteome?* Mol Plant, 2009. **2**(6): p. 1325-35.
74. Bandaranayake, P.C., et al., *A single-electron reducing quinone oxidoreductase is necessary to induce haustorium development in the root parasitic plant Triphysaria.* Plant Cell, 2010. **22**(4): p. 1404-19.
75. Kleffmann, T., et al., *The Arabidopsis thaliana chloroplast proteome reveals pathway abundance and novel protein functions.* Curr Biol, 2004. **14**(5): p. 354-62.
76. Ferro, M., et al., *Integral membrane proteins of the chloroplast envelope: identification and subcellular localization of new transporters.* Proc Natl Acad Sci U S A, 2002. **99**(17): p. 11487-92.
77. Seigneurin-Berny, D., et al., *Technical Advance: Differential extraction of hydrophobic proteins from chloroplast envelope membranes: a subcellular-specific proteomic approach to identify rare intrinsic membrane proteins.* Plant J, 1999. **19**(2): p. 217-228.
78. Curien, G., et al., *The chloroplast membrane associated ceQORH putative quinone oxidoreductase reduces long-chain, stress-related oxidized lipids.* Phytochemistry, 2016. **122**: p. 45-55.
79. Mas y mas, S., et al., *Analytical ultracentrifugation and preliminary X-ray studies of the chloroplast envelope quinone oxidoreductase homologue from Arabidopsis thaliana.* Acta Crystallogr F Struct Biol Commun, 2015. **71**(Pt 4): p. 455-8.
80. Mas, Y.M.S., et al., *Crystal Structure of the Chloroplastic Oxoene Reductase ceQORH from Arabidopsis thaliana.* Front Plant Sci, 2017. **8**: p. 329.
81. Chang, W., J. Soll, and B. Bolter, *A new member of the psToc159 family contributes to distinct protein targeting pathways in pea chloroplasts.* Front Plant Sci, 2014. **5**: p. 239.
82. Kovacs-Bogdan, E., J. Soll, and B. Bolter, *Protein import into chloroplasts: the Tic complex and its regulation.* Biochim Biophys Acta, 2010. **1803**(6): p. 740-7.
83. Chigri, F., et al., *Calcium regulation of chloroplast protein translocation is mediated by calmodulin binding to Tic32.* Proc Natl Acad Sci U S A, 2006. **103**(43): p. 16051-6.
84. Heyno, E., N. Alkan, and R. Fluhr, *A dual role for plant quinone reductases in host-fungus interaction.* Physiol Plant, 2013. **149**(3): p. 340-53.

85. Sparla, F., G. Tedeschi, and P. Trost, *NAD(P)H:(Quinone-Acceptor) Oxidoreductase of Tobacco Leaves Is a Flavin Mononucleotide-Containing Flavoenzyme.* Plant Physiol, 1996. **112**(1): p. 249-258.
86. Marmagne, A., et al., *Identification of new intrinsic proteins in Arabidopsis plasma membrane proteome.* Mol Cell Proteomics, 2004. **3**(7): p. 675-91.
87. Jarvis, P., *Organellar proteomics: chloroplasts in the spotlight.* Curr Biol, 2004. **14**(8): p. R317-9.
88. Andersson, M.X., M. Goksor, and A.S. Sandelius, *Membrane contact sites: physical attachment between chloroplasts and endoplasmic reticulum revealed by optical manipulation.* Plant Signal Behav, 2007. **2**(3): p. 185-7.
89. Crotty, W.J. and M.C. Ledbetter, *Membrane continuities involving chloroplasts and other organelles in plant cells.* Science, 1973. **182**(4114): p. 839-41.
90. Villarejo, A., et al., *Evidence for a protein transported through the secretory pathway en route to the higher plant chloroplast.* Nat Cell Biol, 2005. **7**(12): p. 1224-31.
91. Emanuelsson, O., et al., *Predicting subcellular localization of proteins based on their N-terminal amino acid sequence.* J Mol Biol, 2000. **300**(4): p. 1005-16.
92. Asatsuma, S., et al., *Involvement of alpha-amylase I-1 in starch degradation in rice chloroplasts.* Plant Cell Physiol, 2005. **46**(6): p. 858-69.
93. Kitajima, A., et al., *The rice alpha-amylase glycoprotein is targeted from the Golgi apparatus through the secretory pathway to the plastids.* Plant Cell, 2009. **21**(9): p. 2844-58.
94. Kaneko, K., et al., *Nucleotide pyrophosphatase/phosphodiesterase 1 exerts a negative effect on starch accumulation and growth in rice seedlings under high temperature and CO2 concentration conditions.* Plant Cell Physiol, 2014. **55**(2): p. 320-32.
95. Nowack, E.C. and A.R. Grossman, *Trafficking of protein into the recently established photosynthetic organelles of Paulinella chromatophora.* Proc Natl Acad Sci U S A, 2012. **109**(14): p. 5340-5.
96. Singer, A., et al., *Massive Protein Import into the Early-Evolutionary-Stage Photosynthetic Organelle of the Amoeba Paulinella chromatophora.* Curr Biol, 2017. **27**(18): p. 2763-2773.e5.
97. Murcha, M.W., et al., *Characterization of the preprotein and amino acid transporter gene family in Arabidopsis.* Plant Physiol, 2007. **143**(1): p. 199-212.
98. Rassow, J., et al., *The preprotein translocase of the mitochondrial inner membrane: function and evolution.* J Mol Biol, 1999. **286**(1): p. 105-20.
99. Neupert, W. and J.M. Herrmann, *Translocation of proteins into mitochondria.* Annu Rev Biochem, 2007. **76**: p. 723-49.
100. Rossig, C., et al., *Three proteins mediate import of transit sequence-less precursors into the inner envelope of chloroplasts in Arabidopsis thaliana.* Proc Natl Acad Sci U S A, 2013. **110**(49): p. 19962-7.
101. Froehlich, J.E., et al., *Proteomic study of the Arabidopsis thaliana chloroplastic envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis.* J Proteome Res, 2003. **2**(4): p. 413-25.
102. Cao, T.B. and M.H. Saier, Jr., *The general protein secretory pathway: phylogenetic analyses leading to evolutionary conclusions.* Biochim Biophys Acta, 2003. **1609**(1): p. 115-25.
103. Frain, K.M., et al., *Protein translocation and thylakoid biogenesis in cyanobacteria.* Biochim Biophys Acta, 2016. **1857**(3): p. 266-73.

104. Li, Y., et al., *The Sec2 translocase of the chloroplast inner envelope contains a unique and dedicated SECE2 component.* Plant J, 2015. **84**(4): p. 647-58.
105. Singhal, R. and D.E. Fernandez, *Sorting of SEC translocase SCY components to different membranes in chloroplasts.* J Exp Bot, 2017. **68**(18): p. 5029-5043.
106. Skalitzky, C.A., et al., *Plastids contain a second sec translocase system with essential functions.* Plant Physiol, 2011. **155**(1): p. 354-69.
107. Williams-Carrier, R., et al., *Use of Illumina sequencing to identify transposon insertions underlying mutant phenotypes in high-copy Mutator lines of maize.* Plant J, 2010. **63**(1): p. 167-77.
108. Li, Y. and J.R. Martin, *Identification of Putative Substrates of SEC2, a Chloroplast Inner Envelope Translocase.* 2017. **173**(4): p. 2121-2137.
109. Cline, K., *Mechanistic Aspects of Folded Protein Transport by the Twin Arginine Translocase (Tat).* J Biol Chem, 2015. **290**(27): p. 16530-8.
110. Dabney-Smith, C., H. Mori, and K. Cline, *Oligomers of Tha4 organize at the thylakoid Tat translocase during protein transport.* J Biol Chem, 2006. **281**(9): p. 5476-83.
111. Aldridge, C., et al., *Tat-dependent targeting of Rieske iron-sulphur proteins to both the plasma and thylakoid membranes in the cyanobacterium Synechocystis PCC6803.* Mol Microbiol, 2008. **70**(1): p. 140-50.
112. Palmer, T. and B.C. Berks, *The twin-arginine translocation (Tat) protein export pathway.* Nat Rev Microbiol, 2012. **10**(7): p. 483-96.
113. Settles, A.M., et al., *Sec-independent protein translocation by the maize Hcf106 protein.* Science, 1997. **278**(5342): p. 1467-70.
114. Unseld, M., et al., *The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides.* Nat Genet, 1997. **15**(1): p. 57-61.
115. Carrie, C., S. Weissenberger, and J. Soll, *Plant mitochondria contain the protein translocase subunits TatB and TatC.* J Cell Sci, 2016. **129**(20): p. 3935-3947.
116. Zabrouskov, V., et al., *A new approach for plant proteomics: characterization of chloroplast proteins of Arabidopsis thaliana by top-down mass spectrometry.* Mol Cell Proteomics, 2003. **2**(12): p. 1253-60.
117. Ferro, M., et al., *AT_CHLORO, a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins.* Mol Cell Proteomics, 2010. **9**(6): p. 1063-84.
118. Eichacker, L.A., et al., *Hiding behind hydrophobicity. Transmembrane segments in mass spectrometry.* J Biol Chem, 2004. **279**(49): p. 50915-22.
119. Friso, G., et al., *In-depth analysis of the thylakoid membrane proteome of Arabidopsis thaliana chloroplasts: new proteins, new functions, and a plastid proteome database.* Plant Cell, 2004. **16**(2): p. 478-99.
120. Tomizioli, M., et al., *Deciphering thylakoid sub-compartments using a mass spectrometry-based approach.* Mol Cell Proteomics, 2014. **13**(8): p. 2147-67.
121. Peltier, J.B., et al., *The oligomeric stromal proteome of Arabidopsis thaliana chloroplasts.* Mol Cell Proteomics, 2006. **5**(1): p. 114-33.
122. Ytterberg, A.J., J.B. Peltier, and K.J. van Wijk, *Protein profiling of plastoglobules in chloroplasts and chromoplasts. A surprising site for differential accumulation of metabolic enzymes.* Plant Physiol, 2006. **140**(3): p. 984-97.
123. Brautigam, A. and A.P. Weber, *Proteomic analysis of the proplastid envelope membrane provides novel insights into small molecule and protein transport across proplastid membranes.* Mol Plant, 2009. **2**(6): p. 1247-61.

124. Simm, S., et al., *Defining the core proteome of the chloroplast envelope membranes.* Front Plant Sci, 2013. **4**: p. 11.
125. Agrawal, G.K., et al., *Plant organelle proteomics: collaborating for optimal cell function.* Mass Spectrom Rev, 2011. **30**(5): p. 772-853.
126. Lilley, K.S. and P. Dupree, *Plant organelle proteomics.* Curr Opin Plant Biol, 2007. **10**(6): p. 594-9.
127. Sun, Q., et al., *PPDB, the Plant Proteomics Database at Cornell.* Nucleic Acids Res, 2009. **37**(Database issue): p. D969-74.
128. Kleffmann, T., et al., *plprot: a comprehensive proteome database for different plastid types.* Plant Cell Physiol, 2006. **47**(3): p. 432-6.
129. Heazlewood, J.L., et al., *SUBA: the Arabidopsis Subcellular Database.* Nucleic Acids Res, 2007. **35**(Database issue): p. D213-8.
130. Hooper, C.M., et al., *SUBAcon: a consensus algorithm for unifying the subcellular localization data of the Arabidopsis proteome.* Bioinformatics, 2014. **30**(23): p. 3356-64.
131. Tanz, S.K., et al., *SUBA3: a database for integrating experimentation and prediction to define the SUBcellular location of proteins in Arabidopsis.* Nucleic Acids Res, 2013. **41**(Database issue): p. D1185-91.
132. Joshi, H.J., et al., *MASCP Gator: an aggregation portal for the visualization of Arabidopsis proteomics data.* Plant Physiol, 2011. **155**(1): p. 259-70.
133. Bruley, C., et al., *AT_CHLORO: A Chloroplast Protein Database Dedicated to Sub-Plastidial Localization.* Front Plant Sci, 2012. **3**: p. 205.
134. Dell'Aglio, E., et al., *Complementary biochemical approaches applied to the identification of plastidial calmodulin-binding proteins.* Mol Biosyst, 2013. **9**(6): p. 1234-48.
135. Chiu, W.-l., et al., *Engineered GFP as a vital reporter in plants.* Current Biology, 1996. **6**(3): p. 325-330.
136. *<Boevink_et_al-1996-The_Plant_Journal.pdf>*.
137. Wan, Y., et al., *The signal transducer NPH3 integrates the phototropin1 photosensor with PIN2-based polar auxin transport in Arabidopsis root phototropism.* Plant Cell, 2012. **24**(2): p. 551-65.
138. Kansup, J., et al., *Arabidopsis G-protein beta subunit AGB1 interacts with NPH3 and is involved in phototropism.* Biochem Biophys Res Commun, 2014. **445**(1): p. 54-7.
139. Chen, Y.C., H.J. Wang, and G.Y. Jauh, *Dual Role of a SAS10/C1D Family Protein in Ribosomal RNA Gene Expression and Processing Is Essential for Reproduction in Arabidopsis thaliana.* 2016. **12**(10): p. e1006408.
140. Enkhbayar, P., et al., *Structural principles of leucine-rich repeat (LRR) proteins.* Proteins, 2004. **54**(3): p. 394-403.
141. Gay, N.J., et al., *A leucine-rich repeat peptide derived from the Drosophila Toll receptor forms extended filaments with a beta-sheet structure.* FEBS Lett, 1991. **291**(1): p. 87-91.
142. Rothberg, J.M., et al., *slit: an extracellular protein necessary for development of midline glia and commissural axon pathways contains both EGF and LRR domains.* Genes Dev, 1990. **4**(12a): p. 2169-87.
143. Kobe, B. and A.V. Kajava, *The leucine-rich repeat as a protein recognition motif.* Curr Opin Struct Biol, 2001. **11**(6): p. 725-32.
144. Nishimura, K. and K.J. van Wijk, *Organization, function and substrates of the essential Clp protease system in plastids.* Biochim Biophys Acta, 2015. **1847**(9): p. 915-30.
145. Olinares, P.D., J. Kim, and K.J. van Wijk, *The Clp protease system; a central component of the chloroplast protease network.* Biochim Biophys Acta, 2011. **1807**(8): p. 999-1011.

146. Stanne, T.M., et al., *Identification of new protein substrates for the chloroplast ATP-dependent Clp protease supports its constitutive role in Arabidopsis.* Biochem J, 2009. **417**(1): p. 257-68.
147. Peltier, J.B., et al., *Clp protease complexes from photosynthetic and non-photosynthetic plastids and mitochondria of plants, their predicted three-dimensional structures, and functional implications.* J Biol Chem, 2004. **279**(6): p. 4768-81.
148. Zheng, B., et al., *Characterization of Chloroplast Clp proteins in Arabidopsis: Localization, tissue specificity and stress responses.* Physiol Plant, 2002. **114**(1): p. 92-101.
149. McClung, J.K., et al., *Prohibitin: potential role in senescence, development, and tumor suppression.* Exp Gerontol, 1995. **30**(2): p. 99-124.
150. Weaver, L.M., J.E. Froehlich, and R.M. Amasino, *Chloroplast-targeted ERD1 protein declines but its mRNA increases during senescence in Arabidopsis.* Plant Physiol, 1999. **119**(4): p. 1209-16.
151. Kato, Y. and W. Sakamoto, *New insights into the types and function of proteases in plastids.* Int Rev Cell Mol Biol, 2010. **280**: p. 185-218.
152. van Wijk, K.J., *Protein maturation and proteolysis in plant plastids, mitochondria, and peroxisomes.* Annu Rev Plant Biol, 2015. **66**: p. 75-111.
153. Rosano, G.L., E.M. Bruch, and E.A. Ceccarelli, *Insights into the Clp/HSP100 chaperone system from chloroplasts of Arabidopsis thaliana.* J Biol Chem, 2011. **286**(34): p. 29671-80.
154. Colombo, C.V., E.A. Ceccarelli, and G.L. Rosano, *Characterization of the accessory protein ClpT1 from Arabidopsis thaliana: oligomerization status and interaction with Hsp100 chaperones.* BMC Plant Biol, 2014. **14**: p. 228.
155. Myouga, F., et al., *An Arabidopsis chloroplast-targeted Hsp101 homologue, APG6, has an essential role in chloroplast development as well as heat-stress response.* Plant J, 2006. **48**(2): p. 249-60.
156. Moore, T. and K. Keegstra, *Characterization of a cDNA clone encoding a chloroplast-targeted Clp homologue.* Plant Mol Biol, 1993. **21**(3): p. 525-37.
157. Shanklin, J., N.D. DeWitt, and J.M. Flanagan, *The stroma of higher plant plastids contain ClpP and ClpC, functional homologs of Escherichia coli ClpP and ClpA: an archetypal two-component ATP-dependent protease.* Plant Cell, 1995. **7**(10): p. 1713-22.
158. Peltier, J.B., et al., *Identification of a 350-kDa ClpP protease complex with 10 different Clp isoforms in chloroplasts of Arabidopsis thaliana.* J Biol Chem, 2001. **276**(19): p. 16318-27.
159. Shikanai, T., et al., *The chloroplast clpP gene, encoding a proteolytic subunit of ATP-dependent protease, is indispensable for chloroplast development in tobacco.* Plant Cell Physiol, 2001. **42**(3): p. 264-73.
160. Koussevitzky, S., et al., *An Arabidopsis thaliana virescent mutant reveals a role for ClpR1 in plastid development.* Plant Mol Biol, 2007. **63**(1): p. 85-96.
161. Kuroda, H. and P. Maliga, *The plastid clpP1 protease gene is essential for plant development.* Nature, 2003. **425**(6953): p. 86-9.
162. Nishimura, K., et al., *ClpS1 is a conserved substrate selector for the chloroplast Clp protease system in Arabidopsis.* Plant Cell, 2013. **25**(6): p. 2276-301.
163. Olinares, P.D., et al., *Subunit stoichiometry, evolution, and functional implications of an asymmetric plant plastid ClpP/R protease complex in Arabidopsis.* Plant Cell, 2011. **23**(6): p. 2348-61.

164. Ramundo, S., et al., *Conditional Depletion of the Chlamydomonas Chloroplast ClpP Protease Activates Nuclear Genes Involved in Autophagy and Plastid Protein Quality Control.* Plant Cell, 2014. **26**(5): p. 2201-2222.
165. Kim, J., et al., *Modified Clp protease complex in the ClpP3 null mutant and consequences for chloroplast development and function in Arabidopsis.* Plant Physiol, 2013. **162**(1): p. 157-79.
166. Kim, J., et al., *Subunits of the plastid ClpPR protease complex have differential contributions to embryogenesis, plastid biogenesis, and plant development in Arabidopsis.* Plant Cell, 2009. **21**(6): p. 1669-92.
167. Rudella, A., et al., *Downregulation of ClpR2 leads to reduced accumulation of the ClpPRS protease complex and defects in chloroplast biogenesis in Arabidopsis.* Plant Cell, 2006. **18**(7): p. 1704-21.
168. Sjogren, L.L. and A.K. Clarke, *Assembly of the chloroplast ATP-dependent Clp protease in Arabidopsis is regulated by the ClpT accessory proteins.* Plant Cell, 2011. **23**(1): p. 322-32.
169. Park, S. and S.R. Rodermel, *Mutations in ClpC2/Hsp100 suppress the requirement for FtsH in thylakoid membrane biogenesis.* Proc Natl Acad Sci U S A, 2004. **101**(34): p. 12765-70.
170. Sjogren, L.L., et al., *Inactivation of the clpC1 gene encoding a chloroplast Hsp100 molecular chaperone causes growth retardation, leaf chlorosis, lower photosynthetic activity, and a specific reduction in photosystem content.* Plant Physiol, 2004. **136**(4): p. 4114-26.
171. Kovacheva, S., et al., *In vivo studies on the roles of Tic110, Tic40 and Hsp93 during chloroplast protein import.* Plant J, 2005. **41**(3): p. 412-28.
172. Kovacheva, S., et al., *Further in vivo studies on the role of the molecular chaperone, Hsp93, in plastid protein import.* Plant J, 2007. **50**(2): p. 364-79.
173. Chou, M.L., et al., *Stimulation of transit-peptide release and ATP hydrolysis by a cochaperone during protein import into chloroplasts.* J Cell Biol, 2006. **175**(6): p. 893-900.
174. Nielsen, E., et al., *Stable association of chloroplastic precursors with protein translocation complexes that contain proteins from both envelope membranes and a stromal Hsp100 molecular chaperone.* Embo j, 1997. **16**(5): p. 935-46.
175. Su, P.H. and H.M. Li, *Stromal Hsp70 is important for protein translocation into pea and Arabidopsis chloroplasts.* Plant Cell, 2010. **22**(5): p. 1516-31.
176. Richardson, L.G.L., R. Singhal, and D.J. Schnell, *The integration of chloroplast protein targeting with plant developmental and stress responses.* BMC Biol, 2017. **15**(1): p. 118.
177. Tran, L.S., et al., *Co-expression of the stress-inducible zinc finger homeodomain ZFHD1 and NAC transcription factors enhances expression of the ERD1 gene in Arabidopsis.* Plant J, 2007. **49**(1): p. 46-63.
178. Akita, M., E. Nielsen, and K. Keegstra, *Identification of protein transport complexes in the chloroplastic envelope membranes via chemical cross-linking.* J Cell Biol, 1997. **136**(5): p. 983-94.
179. Chu, C.C. and H.M. Li, *The amino-terminal domain of chloroplast Hsp93 is important for its membrane association and functions in vivo.* Plant Physiol, 2012. **158**(4): p. 1656-65.
180. Heard, W., et al., *Identification of Regulatory and Cargo Proteins of Endosomal and Secretory Pathways in Arabidopsis thaliana by Proteomic Dissection.* Mol Cell Proteomics, 2015. **14**(7): p. 1796-813.

181. Alexandersson, E., et al., *Arabidopsis plasma membrane proteomics identifies components of transport, signal transduction and membrane trafficking.* Plant Cell Physiol, 2004. **45**(11): p. 1543-56.
182. Marmagne, A., et al., *Identification of new intrinsic proteins in Arabidopsis plasma membrane proteome.* Mol Cell Proteomics, 2004. **3**(7): p. 675-91.
183. Heard, W., et al., *Identification of Regulatory and Cargo Proteins of Endosomal and Secretory Pathways in Arabidopsis thaliana by Proteomic Dissection.* Mol Cell Proteomics, 2015. **14**(7): p. 1796-813.
184. Nikolovski, N., et al., *Putative glycosyltransferases and other plant Golgi apparatus proteins are revealed by LOPIT proteomics.* Plant Physiol, 2012. **160**(2): p. 1037-51.

Annexes

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

Standard Manuscript Template
Please Remove all Grey Text before Submitting

TITLE: (Instructions)
Preparation of chloroplast subcompartments from Arabidopsis for the analysis of protein localization by immunoblotting or proteomics and other studies

AUTHORS & AFFILIATIONS: (Instructions)
Imen Bouchnak[1], Lucas Moyet[1], Daniel Salvi[1], Marcel Kuntz[1], Norbert Rolland[1]
[1]Univ. Grenoble Alpes, INRA, CNRS, CEA, BIG-LPCV, 38000 Grenoble, France

Corresponding Author:
Norbert Rolland
Email Address: norbert.rolland@cea.fr
Tel: + 33 (0)4 38 78 49 86

Email Addresses of Co-authors:
Imen Bouchnak (imen.bouchnak@cea.fr)
Lucas Moyet (lucas.moyet@cea.fr)
Daniel Salvi (daniel.salvi@cea.fr)
Marcel Kuntz (marcel.kuntz@cea.fr)
Norbert Rolland (norbert.rolland@cea.fr)

KEYWORDS: (Instructions)
Plant, Arabidopsis, chloroplast, organelle, plastid, envelope, stroma, thylakoid, cell fractionation, Percoll, localization, immunoblotting

SHORT ABSTRACT: (Instructions)
Here, we describe a method to purify intact chloroplasts from Arabidopsis leaves and their three main sub-compartments (envelope, stroma and thylakoids), using a combination of differential centrifugations, continuous Percoll gradients and discontinuous sucrose gradients. The method is valuable for subplastidial and subcellular localization of proteins by immunoblotting and proteomics.

LONG ABSTRACT: (Instructions)
Chloroplasts are major components of plant cells. Such plastids fulfill many crucial functions, such as assimilation of carbon, sulfur and nitrogen as well as synthesis of essential metabolites. These organelles consist of three key sub-compartments. The envelope, characterized by two membranes, that surrounds the organelle and controls the communication of the plastid with other cell compartments, the stroma, the soluble phase of the chloroplast and the main site where carbon dioxide is converted into carbohydrates, and the thylakoid membrane (internal membrane network consisting of grana (flat compressed sacs) and lamellae (less dense structures)), where oxygenic photosynthesis takes place. The present protocol describes step by step procedures required for the purification, using differential centrifugations and Percoll gradients, of intact chloroplasts from Arabidopsis, and their fractionation, using sucrose gradients, in three sub-compartments (*i.e.* envelope, stroma and thylakoids). This protocol also

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

Instructions for Authors

provides instructions on how to assess the purity of these fractions using markers associated to the various chloroplast sub-compartments. The method described here is valuable for subplastidial localization of proteins using immunoblotting, but also for subcellular and subplastidial proteomics and other studies.

INTRODUCTION: (Instructions)

Chloroplasts are major components of plant cells. They derive from a cyanobacterial ancestor that has undergone an endosymbiosis and eventually evolved as an organelle during evolution[1,2]. Such organelles contain three main compartments (Figure 1). The envelope system, is made of an inner and an outer membranes surrounding the organelle. This double membrane system contains various enzymes involved in the metabolism of lipids and pigments, and is mostly devoted to the control of the communication between plastids and the cytosol. It contains various transport systems that allows the import of nuclear-encoded proteins, and the exchange of ions and metabolites between the cytosol and the chloroplast thus regulating essential metabolic functions of the plant cell [3,4]. The stroma, the soluble phase of the chloroplast, contains enzymes of the Calvin cycle (CO_2 assimilation), the synthesis of various metabolites including amino acids and vitamins, and the transcription and translation machineries of the plastid. The thylakoid membrane, is a widely organized internal membrane network where the light phase of photosynthesis takes place. Thereby, chloroplasts are the place where essential metabolic pathways occur [5].

In order to decipher new regulatory mechanisms that control the chloroplast dynamics and physiology, defining the sub-plastidial localization of chloroplast proteins is thus critical to support targeted studies aiming to better understand proteins functions in model organisms[6] . In order to get access to the genuine subplastidial localization of these proteins, it is thus essential to start from highly pure subplastidial fractions (envelope membranes, stroma, and thylakoids). In this context, the aim of the present protocol is to purify intact chloroplasts from Arabidopsis leaves using differential centrifugations and continuous Percoll gradients, and to fractionate them using discontinuous sucrose gradients, in three sub-compartments (i.e. envelope, stroma and thylakoids). The method described here also provides instructions to assess the purity of purified sub-organellar fractions using markers associated to the various chloroplast subcompartments. This protocol is valuable for subplastidial localization of proteins using immunoblotting and for further analysis of purified fractions using MS-based proteomic studies.

PROTOCOL: (Instructions)

1. Preparation of buffers, stock solutions and gradients

1.1. Prepare the following stock solutions that can be stored up to 6 months at 4°C:

1.1.1. Tricine buffer (1 M, pH 8.4) and Tricine buffer (1 M, pH 7.6). Adjust pH by adding KOH pellets.

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

1.1.2. 1 L of ethylenediaminetetraacetic acid (EDTA, 0.5 M, pH 8) and 3-(N-morpholino) propane sulfonic acid (MOPS) buffer (1 M, pH 7.8). Adjust pH by adding NaOH pellets.

1.1.3. 50 mL of $MgCl_2$ (1 M).

1.1.4. 50 mL of protease inhibitors solutions: phenylmethylsulfonyl fluoride (PMSF set up in isopropanol, 100 mM), benzamidine hydrochloride hydrate (100 mM), and ε-amino caproic acid (50 mM). Note that while PMSF and amino caproic acid are stable in solution for months at 4°C, benzamidine solution should be stored at -20°C.

1.2. Prepare the following solutions the day before the experiment:

1.2.1. 4 L of grinding medium pH 8.4 containing Tricine-KOH (20 mM, pH 8.4), sorbitol (0.4 M), EDTA (10 mM, pH 8), and $NaHCO_3$ (10 mM). Adjust pH by adding NaOH pellets. Add bovine serum albumin (BSA) at 0.1% (w/v) just before use and mix well.

1.2.2. 500 mL of washing medium (2x) pH 7.6 containing Tricine-KOH (20 mM, pH 7.6), sorbitol (0.8 M), $MgCl_2$ (5 mM), EDTA (2.5 mM). Adjust pH by adding NaOH pellets. Do not forget to dilute such solution after preparation of Percoll gradient solution to obtain washing medium (1x).

1.2.3. 200 mL of Percoll gradient solution for chloroplast purification by mixing Percoll with washing medium (2x) at equal volume to get a final solution at 50% (v/v) Percoll / 0.4 M sorbitol.

1.2.4. 50 mL of sucrose solutions for chloroplast fractionation by mixing MOPS (10 mM, pH 7.8), $MgCl_2$ (4 mM), and different concentrations of sucrose (0.3 M, 0.6 M and 0.93 M).

Store all solutions at 4°C.

1.3. Prepare the following gradients and buffers prior to starting the experiment:

1.3.1. Six tubes of Percoll gradients (each containing 30 mL of a 50% Percoll / 0.4 M sorbitol) by centrifugation at 38,700 *g* for 55 min at 4°C (Beckman JA-20 rotor). The brake should be off to prevent blending of the gradients. After centrifugation, the tubes containing the preformed gradients are stored in a cold room until use.

1.3.2. Four tubes (Beckman SW41-Ti rotor) of sucrose gradients. Each gradient is formed of three following sucrose layers: 3 mL of 0.93 M, 2.5 mL of 0.6 M, and 2 mL of 0.3

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com
Instructions for Authors

M sucrose. Carefully overlaid each layer, using a peristaltic pump starting with 0.93 M, at the bottom and finishing with 0.3 M at the top

1.3.3. 50 mL of hypotonic medium for chloroplast lysis containing MOPS (10 mM, pH 7.8), $MgCl_2$ (4 mM), PMSF (1 mM, set up in isopropanol), benzamidine hydrochloride hydrate (1 mM), and ε-amino caproic acid (0.5 mM). Store the buffer on ice until use.

1.3.4. 50 mL of membrane washing buffer containing MOPS (10 mM, pH 7.8), PMSF (1 mM), benzamidine hydrochloride hydrate (1 mM), ε-amino caproic acid (0.5 mM). Store the buffer on ice until use.

2. Growth and harvesting of Arabidopsis leaves

2.1. Growth of Arabidopsis plants: prepare 4 large plastic pans (for a total surface of 0.5 to 1 m^2) of Arabidopsis plants by sowing 30 mg of seeds in each pan. Grow Arabidopsis plants for 5 weeks at 12-h light cycle at 23°C (day) / 18°C (night) with a light intensity of 150 $\mu mol.m^{-2}.s^{-1}$.

2.2. Incubate plants in a dark and cold room (4°C) overnight prior to the experiment (to reduce the amount of starch granules in chloroplasts).

2.3. Pre-weigh a 1 L beaker and then place it on ice before starting harvesting of leaf material.

2.4. Harvest Arabidopsis leaves by avoiding soil (compost). Reweigh the beaker and record the tissue weight. 400 to 500 g of leaf material are expected from four pans.

2.5. Homogenize leaves in a cold room with 2 L of grinding buffer (do not forget to add BSA before use) three times/2 s each time, in a Waring blender at high speed.

2.6. Filter the homogenate in a cold room using 4 layers of muslin and one layer of nylon blutex. Gently squeeze the homogenate leaves inside the muslin/nylon blutex, to extract all the liquid.

2.7. Recover the remaining tissue in the blender cup for a second extraction. Repeat steps 2.5 and 2.6 using 2 L of grinding medium and new 4-5 layers of muslin (in a cold room).

3. Purification of crude chloroplasts using differential centrifugation

3.1. Equally distribute the crude cell extract into six 500 mL bottles and centrifuge for 2 min as soon as the max speed (2,070 *g*) is reached (Beckman JLA-10.500, max acceleration and brake on, 4°C). Place the bottles on ice before centrifugation.

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com
Instructions for Authors

3.2. Gently discard the supernatant.

3.3. Aspirate the remaining supernatant using a water pump and keep the pellets containing concentrated crude chloroplasts on ice.

3.4. Gently resuspend pellets, by adding a minimal volume of washing medium (1x) (final volume of the combined chloroplast suspensions = 36 mL) using a paintbrush or a curved plastic spatula. Use a 10 mL pipet to add 3 ml of washing medium in each bottle (do not use Pipetman with very fine tips, to avoid breakage of chloroplasts. Alternatively, cut the blue tip of a Pipetman with a razor blade to generate a larger hole).

3.5. Collect the resuspended chloroplasts in one tube by using a 10 mL pipet. Gently mix (by inverting the tube) to obtain a homogenous suspension prior to loading on Percoll gradients.

4. Purification of intact chloroplasts on continuous Percoll gradient

4.1. Slowly load 6 ml of the chloroplast suspension on top of each of the six Percoll gradients using a 10 mL pipet to avoid breakage of chloroplasts.

4.2. Centrifuge the gradients for 10 min at (13,300 *g*, 4°C) using a Beckman Swinging-bucket JS-13.1 rotor. The acceleration should be slow and the brake should be disconnected (brake off or slow deceleration) to prevent blending of the Percoll gradients.

4.3. Aspirate the upper phase that contains broken chloroplasts and intact mitochondria using a water pump, and then retrieve intact chloroplasts present in the lower phase (the broad dark-green band) with a 10 mL pipet. Be careful to not aspirate nuclei and cell debris (found at the bottom of the tube) with the intact chloroplasts (Figure 2A).

4.4. Dilute 3-4-fold the intact chloroplast suspension with washing buffer (1x). Centrifuge for 2 min as soon as the max speed (2,070 *g*, 4°C) is reached (Beckman JLA-10.500 rotor, max acceleration and brake on).

4.5. Carefully discard the supernatant.

4.6. Completely aspirate the remaining supernatant with a water pump and keep the pellet of concentrated intact chloroplasts on ice.

4.7. Before chloroplast lysis, keep an aliquot of intact chloroplast fraction in ~1ml of washing medium (1x) for further analyses using SDS-PAGE and western blotting. Keep a small aliquot of these intact chloroplasts for determination of protein concentration. Store the intact chloroplast fraction in liquid nitrogen for further experiments.

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com
Instructions for Authors

5. Lysis of intact chloroplasts using a hypotonic buffer and purification of chloroplast sub-compartments on discontinuous sucrose gradients

5.1. Lyse the purified intact chloroplasts by resuspending the pellet in hypotonic medium that contains protease inhibitors (the final volume should not exceed 12 mL). From this step, the use of Pipetman with fine tips (blue tips) is possible since intactness of chloroplasts is no more essential (pipetting chloroplasts up and down as long as pellet is not entirely resuspended). Arabidopsis chloroplasts are very fragile (when compared to pea chloroplasts, for example) and their lysis is almost immediate after incubation in hypotonic medium.

5.2. Slowly load 3 mL of the lysed chloroplasts on the top of each preformed sucrose gradients using a peristaltic pump.

5.3. Ultracentrifuge the gradients for 1 h at (70,000 *g*, 4°C) (SW41-Ti rotor). Do not forget to balance pairs of tubes using hypotonic medium buffer prior to centrifugation.

5.4. Carefully recover the soluble stromal proteins by pipetting the upper phase of the gradient (3 mL from each gradient) (Figure 2B). Take an aliquot for determination of protein concentration [7]. Store the stroma in liquid nitrogen for further experiments.

5.5. Aspirate the remaining upper phase of each gradient up to the yellow band, using a water pump.

5.6. Retrieve the yellow band (the envelope) with a Pipetman (~ 1mL from each gradient). Pool the envelopes in one tube (Beckman SW41).

5.7. Remove the remaining phase of each gradient up to the thylakoid pellet, using a water pump.

6. Washing and concentration of thylakoid and envelope membrane systems

6.1. Resuspend the thylakoid pellets (green pellets) in a minimum volume (2 mL) of membrane washing buffer (1x) (with protease inhibitors).

6.2. Dilute the envelope and thylakoid suspensions 3-4-fold in membrane washing medium (adjust volume to 10 mL) and ultracentrifuge for 1 h at (110,000 *g*, 4°C) (SW41-Ti rotor). Do not forget to balance pairs of tubes using membrane washing buffer prior to centrifugation.

6.3. Carefully aspirate the supernatants using a water pump.

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com
Instructions for Authors

6.4. Add ~100 µL of membrane washing buffer (with protease inhibitors) to the envelope pellet. Take an aliquot for determination of protein concentration [7]. Store the purified envelope membrane preparation in liquid nitrogen.

6.5. Resuspend thylakoids pellet in 3 mL of membrane washing buffer (with protease inhibitors). Take an aliquot for determination of protein concentration [7]. Store thylakoid membrane fraction in liquid nitrogen.

REPRESENTATIVE RESULTS: (Instructions)

Successive steps of the procedure resulting in purified chloroplast and their subcompartments are resumed in Figure 2. The Percoll gradient (Figure 2A) allows distinguishing intact chloroplasts from broken chloroplast and mitochondria (top of the gradient) or nuclei and cell debris (bottom of the gradient). After rupture of the Percoll-purified organelles thanks to an osmotic shock, the resulting fractions are separated on a sucrose gradient (Figure 2B). The stroma (soluble part of the chloroplast) is floating at the surface of the sucrose gradient. The light envelope membrane vesicles are recovered as a discrete yellow band at the 0.6/0.93 M sucrose interface. The heaviest thylakoid membranes vesicles are concentrated at the bottom of the tube. After recovery, washing and concentration of the two membrane fractions, proteins are quantified and the composition of all four fractions is analyzed on a SDS-PAGE (Figure 2C). The lanes are loaded on an equal protein basis (20 µg of each purified fraction). Knowing that chloroplasts contain only 1% of envelope proteins and 50% of proteins from the stroma or from the thylakoids, this tends to overestimate cross-contamination of purified envelope preparations with other chloroplast subcompartments. However, this method allows to detect minute amounts of proteins cross-contaminating the envelope fraction. Markers from each compartment (*i.e.* abundant proteins) are very helpful in evaluating the cross contamination of the fractions. Indeed, the thylakoid and envelope membrane fractions are expected to contain very low amounts of the large subunit of RuBisCO (RBCL), the most abundant protein from the stroma (50 kDa). Broken chloroplasts can easily be distinguished from intact chloroplast due to the loss of this stromal protein [8]. The light harvesting complex proteins (LHCP) are 25-kDa abundant thylakoid components that should barely (less than 3%) contaminate envelope membranes [9]. Finally, the phosphate-triose-phosphate translocator (TPT) is a 30-kDa protein that is only visible in the purified envelope fraction due to its strong enrichment (*i.e.* 50 to 100x) in the envelope fraction when compared to whole chloroplast extracts. Using the method described here, chloroplast subcompartments are generally poorly cross-contaminated as confirmed using western-blot analyses (Figure 2D) relying on antibodies directed against known markers of all three compartments : the soluble ketol-acid reductoisomerase (KARI) from the stroma, the chloroplast envelope copper ATPase (HMA1) and the light harvesting complex proteins (LHCP) from the thylakoid membranes. Cross-contamination of the three compartments can be quantified using both immunoblotting and mass spectrometry analyses [9]. While the stroma is usually not contaminated by envelope or thylakoid fractions, the purified envelope fractions contain 3% of thylakoid proteins and up to 10% of proteins from the stroma. Proteins from the stroma poorly contaminate the thylakoid membranes (less than 1%) but thylakoids contain up to 3% of envelope membrane proteins. More than having a crucial role in identifying the genuine subplastidial location of chloroplast proteins, the present method thus also limits erroneous conclusions about subplastidial

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com
Instructions for Authors

localization of proteins resulting from cross contaminations.

FIGURE AND TABLE LEGENDS: (Instructions)

Figure 1: Representative scheme of chloroplast sub-compartments

Figure 2: Purification of intact chloroplasts and their three main subcompartments using Percoll and sucrose gradients. A. Percoll gradient allowing separation of broken and intact chloroplasts. **B. Sucrose gradient** allowing separation of stroma, envelope and thylakoid fractions. **C. Representative SDS-PAGE of proteins from intact chloroplasts and their three main subcompartments** allowing to visualize abundant markers from each compartments**.** Each lane contains 10 µg of proteins. MW, molecular weight markers. RBCL, Large Subunit of RuBisCO (marker for the stroma); TPT, Phosphate/triose-phosphate translocator (marker for the envelope); LHCP, Light harvesting complex proteins (marker for the thylakoid). **D. Western blot experiments** allowing to detect (using specific antibodies) specific markers from each compartment: the chloroplast envelope copper ATPase HMA1 [10], the light harvesting complex proteins LHCP from the thylakoid membranes [11] and the ketol-acid reductoisomerase KARI from the stroma [9].

DISCUSSION: (Instructions)

The present article aims to detail the step by step protocol used to purify chloroplasts (and their subcompartments) from *Arabidopsis thaliana*. Since the availability of its complete genome sequence, almost two decades ago, and of large collections of insertion mutants made available to the community, Arabidopsis is now widely accepted as a model plant. However, while this plant was perfectly adapted for genetic approaches, plant scientists needed to adapt biochemical and physiological tools to this emerging model. Protocols allowing to purify photosynthetically active chloroplasts from leaves of well-established biochemical models like spinach [12] or pea [13] thus had to be adapted. The first method describing purification of Arabidopsis chloroplasts was published in 1998[14], just before the release of the Arabidopsis genome sequence. Several years later, simple methods for isolating Arabidopsis chloroplasts compatible with studies aiming to analyze *in vitro* import of proteins in purified organelles were made available [15,16]. However, these methods did not allow to combine high level of purity and preservation of photosynthetic activity of the purified chloroplasts. More recently [17], a rapid method was established, that both relies on the use of Percoll gradients, and allows to retain almost 90% of the photosynthesis rate measured in the starting leaves of Arabidopsis.

The protocol described here allows to purify Arabidopsis chloroplasts at an excellent level of purity. Indeed, immunological detection of contaminants from other cell compartments demonstrated that the purified organelles are devoid of mitochondrial and plasma membrane markers [9, 10]. This protocol was also efficient to purify chloroplasts from several Arabidopsis ecotypes [18], like Columbia (Col) or Wassilewskija (WS), *i.e.* the ecotypes that were used for genome or ESTs sequencing projects but also to generate T-DNA insertion mutants in Arabidopsis. In other words, when proteomics studies have to be performed, the present protocol is compatible with these two reference ecotypes from Arabidopsis. Finally, the yield of

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com
Instructions for Authors

chloroplasts using the present protocol is similar to the one obtained when starting from spinach or pea leaves (*i.e.* 3%, as measured from the chlorophyll content in the Percoll-purified chloroplast when compared to the total chlorophyll amount present in starting leaves). In terms of proteins, the yield is close to 50 mg of chloroplast proteins, when organelles are purified from 500 g of 5-week-old Arabidopsis leaves.

To reach such a good yield (and chloroplast integrity), one should however pay special attention to several steps when using the present protocol. The chloroplast is, in Arabidopsis, an extremely fragile structure (this is not the case for pea chloroplasts, for example). Specific attention is thus required in order to avoid large-scale rupture of the organelles during purification. The number and size of starch granules present in chloroplasts is critical for the preparation of intact chloroplasts. Indeed, chloroplasts containing large starch grain will generally be broken during initial differential centrifugations steps aiming to concentrate the crude chloroplast fractions [12]. Therefore, the plants should be kept overnight in a dark and cold room (4°C) prior to the experiment, to reduce the amount of starch.

New users of the present protocol could be tempted to start from larger amounts of leaf material (huge rosettes from old Arabidopsis plants with larger leaves) trying to enhance the recovery of purified chloroplasts. However, in our hands, starting from young leaves (5-week-old) is the best compromise to combine yield, purity, and integrity of the purified organelles. Indeed, too old leaves are highly enriched in phenolic compounds that were shown to have a negative impact on chloroplast integrity [19].

Finally, the initial extraction step (grinding of the tissue) is another critical step. The blending process must be limited to few seconds. As stated above, new users might be tempted to use longer blending, thus expecting to strongly improve the yield of purified organelles. However, if longer blending effectively releases more material from leaves, it appears that the proportion of broken chloroplasts rapidly increases in the crude cell extract. Due to this high ratio of broken to intact chloroplasts in the medium, further purification steps (separation on Percoll gradients) are strongly affected and the yield of the purification is unexpectedly lower.

Availability of specific protocols to purify organelles have allowed a series of high throughput proteomics-based experiments to be conducted on chloroplast samples. These data were made available in several public databases [6], thus providing to biologists in the field an accurate subcellular (and subplastidial) localization for many chloroplast proteins. This was especially true for envelope proteins whose identity and location remained mostly unknown before these analyses, since envelope membranes represent a minor chloroplast component (1–2% of the chloroplast proteins) while playing a key role in chloroplast metabolism and biogenesis [5,20]. Using the protocol described here, we recently analyzed the composition of the three main chloroplast compartments from Arabidopsis (*i.e.* the stroma, the thylakoids, and the envelope membrane system) [9]. Based on a semi-quantitative proteomics approach (spectral counting), we were able to assess the partitioning of hundreds of proteins in these three chloroplast compartment.

While the present protocol allows to purify the three main compartments of the chloroplast from Arabidopsis, it is also possible to distinguish additional sub compartments in the chloroplast. Indeed, the envelope membrane system is made of the inner and the outer envelope membranes (Figure 1). However, to the best of our knowledge, a method to purify inner and outer envelope membranes from Arabidopsis chloroplasts remains to be established. Inner and outer envelope membranes can be purified from spinach [21] or pea [22] chloroplasts. The main limit of Arabidopsis

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com
Instructions for Authors

mostly results from the limiting amounts of starting material. Starting from 500 g of Arabidopsis leaves (which already requires 1 m^2 surface in a growth chamber) allows purifying only 100 µg of envelope proteins. On the other hand, it is easy to start with 5-10 kg of spinach leaves from the market, to purify large amount of chloroplasts [8] and to end with a yield of 3 to 10 mg of envelope proteins from this material.

The same is true for thylakoid sub-compartments. Indeed, thylakoids are made of light membrane vesicles (lamellae) and dense structures (grana) (Figure 1). Specific protocols are available to distinguish these two compartments in Arabidopsis [23,24]. Again, based on a quantitative proteomics analysis, we recently inventoried the proteins present in these two subcompartments [24]. These approaches, together with an in-depth investigation of the literature, allowed validating, or propose hypotheses for, the subplastidial location of hundreds of thylakoid proteins. However, it is important to note that additional membrane microdomains are present at the curved margins of thylakoids. These lipoprotein subcompartments, or plastoglobules, are permanently coupled to thylakoid membranes and contain a specific set of proteins [25]. Using the present protocol, it is thus not possible to distinguish these specific proteins from other thylakoid components.

Some genuine (well known) envelope, stroma or thylakoid components are still lacking from the lists of detected proteins. Together with targeted biochemical and immunological analyses, the continual improvement of MS sensitivity will be of great help to revisit the chloroplast content towards a complete repertoire of the composition of its various subcompartments.

ACKNOWLEDGMENTS: (Instructions)

This work was supported by a joint PhD fellowship to IB from the INRA Plant Biology and Breeding Division and from the Labex GRAL (ANR-10-LABX-49-01). We would also like to acknowledge the ANR project ANR-15-IDEX-02, Dr. Olivier Vallon (IBPC Paris) for antibodies against LHCP and Dr. Renaud Dumas (LPCV, Grenoble) for antibodies against KARI.

DISCLOSURES: (Instructions)

Authors have nothing to disclose.

REFERENCES: (Instructions)

1 Zimorski, V., Ku, C., Martin, W. F. & Gould, S. B. Endosymbiotic theory for organelle origins. *Curr Opin Microbiol.* **22** 38-48, doi:10.1016/j.mib.2014.09.008, (2014).

2 Gould, S. B., Waller, R. F. & McFadden, G. I. Plastid evolution. *Annu Rev Plant Biol.* **59** 491-517, doi:10.1146/annurev.arplant.59.032607.092915, (2008).

3 Linka, N. & Weber, A. P. Intracellular metabolite transporters in plants. *Mol Plant.* **3** (1), 21-53, doi:10.1093/mp/ssp108, (2010).

4 Block, M. A., Douce, R., Joyard, J. & Rolland, N. Chloroplast envelope membranes: a dynamic interface between plastids and the cytosol. *Photosynth Res.* **92** (2), 225-244, doi:10.1007/s11120-007-9195-8, (2007).

5 Rolland, N. *et al.* The biosynthetic capacities of the plastids and integration between cytoplasmic and chloroplast processes. *Annu Rev Genet.* **46** 233-264, doi:10.1146/annurev-genet-110410-132544, (2012).

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

Instructions for Authors

6 Agrawal, G. K. *et al.* Plant organelle proteomics: collaborating for optimal cell function. *Mass Spectrom Rev.* **30** (5), 772-853, doi:10.1002/mas.20301, (2011).
7 Chua, N.-H. [40] Electrophoretic analysis of chloroplast proteins. *Methods in Enzymology.* **69** 434-446, doi:http://dx.doi.org/10.1016/S0076-6879(80)69042-9, (1980).
8 Seigneurin-Berny, D., Salvi, D., Joyard, J. & Rolland, N. Purification of intact chloroplasts from Arabidopsis and spinach leaves by isopycnic centrifugation. *Curr Protoc Cell Biol.* **Chapter 3** Unit 3 30, doi:10.1002/0471143030.cb0330s40, (2008).
9 Ferro, M. *et al.* AT_CHLORO, a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins. *Mol Cell Proteomics.* **9** (6), 1063-1084, doi:10.1074/mcp.M900325-MCP200, (2010).
10 Seigneurin-Berny, D. *et al.* HMA1, a new Cu-ATPase of the chloroplast envelope, is essential for growth under adverse light conditions. *J Biol Chem.* **281** (5), 2882-2892, doi:10.1074/jbc.M508333200, (2006).
11 Vallon, O. *et al.* Lateral redistribution of cytochrome b6/f complexes along thylakoid membranes upon state transitions. *Proc Natl Acad Sci U S A.* **88** (18), 8262-8266 (1991).
12 Douce, R. J., J. Purification of the chloroplast. *M. Edelman, R. Hallick, N. H. Chua (Eds.), Methods in Chloroplast Molecular Biology, Elsevier Science Publishers BV, Amsterdam.* 239–256 (1982).
13 Cerovic, Z. G. & Plesnicar, M. An improved procedure for the isolation of intact chloroplasts of high photosynthetic capacity. *Biochem J.* **223** (2), 543-545 (1984).
14 Kunst, L. Preparation of physiologically active chloroplasts from Arabidopsis. *Methods Mol Biol.* **82** 43-48 (1998).
15 Aronsson, H. & Jarvis, P. A simple method for isolating import-competent Arabidopsis chloroplasts. *FEBS Lett.* **529** (2-3), 215-220 (2002).
16 Ling, Q. & Jarvis, P. Analysis of Protein Import into Chloroplasts Isolated from Stressed Plants. *J Vis Exp.* (117), doi:10.3791/54717, (2016).
17 Seigneurin-Berny, D., Salvi, D., Dorne, A. J., Joyard, J. & Rolland, N. Percoll-purified and photosynthetically active chloroplasts from Arabidopsis thaliana leaves. *Plant Physiol Biochem.* **46** (11), 951-955, doi:10.1016/j.plaphy.2008.06.009, (2008).
18 Salvi, D., Rolland, N., Joyard, J. & Ferro, M. Purification and proteomic analysis of chloroplasts and their sub-organellar compartments. *Methods Mol Biol.* **432** 19-36, doi:10.1007/978-1-59745-028-7_2, (2008).
19 Walker, D. The use of the oxygen electrode and fluorescence probes in simple measurements of photosynthesis. *Oxygraphics Ltd, University of Sheffield, South Yorkshire, UK.* (1990).
20 Jarvis, P. & Lopez-Juez, E. Biogenesis and homeostasis of chloroplasts and other plastids. *Nat Rev Mol Cell Biol.* **14** (12), 787-802, doi:10.1038/nrm3702, (2013).
21 Block, M. A., Dorne, A. J., Joyard, J. & Douce, R. Preparation and characterization of membrane fractions enriched in outer and inner envelope membranes from spinach chloroplasts. II. Biochemical characterization. *J Biol Chem.* **258** (21), 13281-13286 (1983).
22 Soll, J. Phosphoproteins and protein-kinase activity in isolated envelopes of pea (Pisum sativum L.) chloroplasts. *Planta.* **166** (3), 394-400, doi:10.1007/bf00401178, (1985).
23 Moyet, L., Salvi, D., Tomizioli, M., Seigneurin-Berny, D. & Rolland, N. Preparation of Membrane Fractions (Envelope, Thylakoids, Grana, and Stroma Lamellae) from

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

Instructions for Authors

Arabidopsis Chloroplasts for Quantitative Proteomic Investigations and Other Studies. *Methods Mol Biol.* **1696** 117-136, doi:10.1007/978-1-4939-7411-5_8, (2018).

24 Tomizioli, M. *et al.* Deciphering thylakoid sub-compartments using a mass spectrometry-based approach. *Mol Cell Proteomics.* **13** (8), 2147-2167, doi:10.1074/mcp.M114.040923, (2014).

25 Spicher, L. & Kessler, F. Unexpected roles of plastoglobules (plastid lipid droplets) in vitamin K1 and E metabolism. *Curr Opin Plant Biol.* **25** 123-129, doi:10.1016/j.pbi.2015.05.005, (2015).

Figure 1

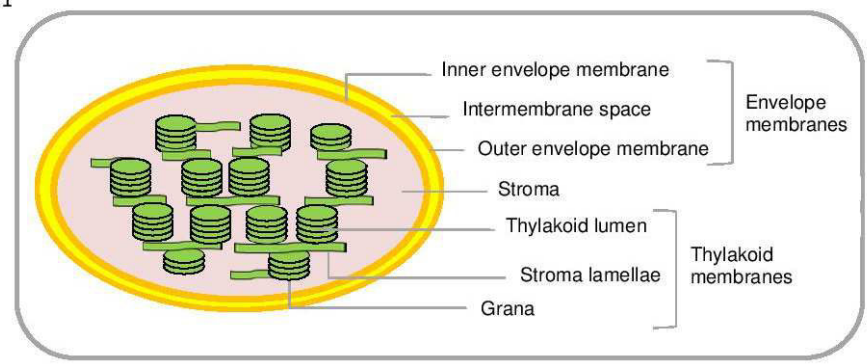
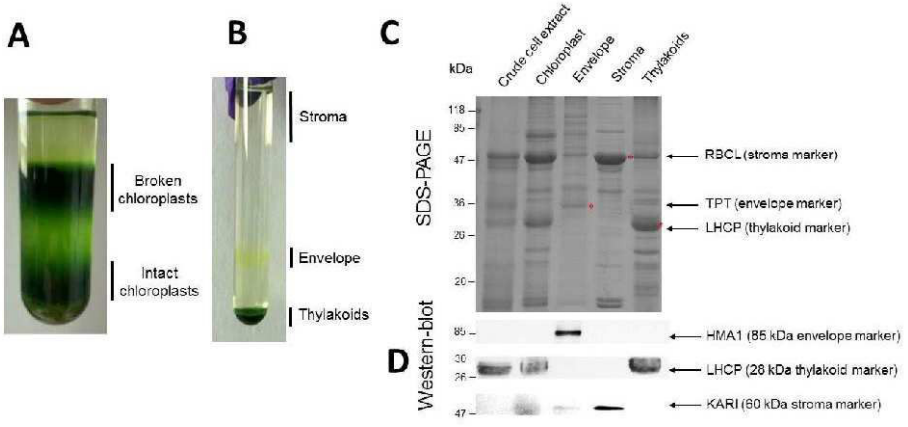


Figure 2



jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

Instructions for Authors

508 Table of Materials

Name of Material/ Equipment	Company	Catalog Number	Comments/Description
Percoll	GE Healthcare	17089101	
Tricine	Roth	6977.2	
Sorbitol	Roth	6213.1	
Ethylenediaminetetraacetic acid (EDTA)	Promega	H5032	
$NaHCO_3$	Roth	8551.1	
Bovine serum albumin (BSA)	Roth	8076.5	
$MgCl_2$	Roth	2189.1	
Phenylmethylsulfonyl fluoride (PMSF)	Sigma	P7626	
Benzamidine	Sigma	B6506	
ε-amino caproic acid	Fluka	21530	
3-(N-morpholino) propane sulfonic acid (MOPS)	Roth	6979.3	
Sucrose	Roth	9286.2	
KOH	Sigma	1.05012	
NaOH	CARLO ERBA	480507	
Anti-HMA1 antibody		Seigneurin-Berny et *al,* 2006	Used at a 1:1000 dilution
Anti-KARI antibody		Ferro et *al*, 2010	Used at a 1:1000 dilution
Anti-LHCP antibody		Vallon et *al*, 1991	Used at a 1:25,000 dilution
Large (30 cm × 45 cm) plastic cases	Puteaux	162135	
A. thaliana seeds			Around 30 mg of seeds for a whole case
Compost "Floragard"	Puteaux	16311770	
Growth rooms			12-h light cycle, set at 23°C (day) / 18°C (night) with a light intensity of 150 µmol/m^2/s.
Muslin or cheesecloth	Raffin	70116	80-cm-large
Nylon blutex 50 µm aperture	Tripette et Renaud, Sailly Saillisel		50 µm aperture
Motor-driven blender, three speeds, 1 gallon (4 L) capacity	Waring Blender		
Fixed-angle rotors JLA-10.500 (6 × 500-mL plastic bottles)	Beckman Coulter		
Beckman JA-20 rotor	Beckman Coulter		
JA-20 (6 × 50 mL polypropylene tubes)	Sorvall instruments		
Swinging-bucket rotor JS-13.1 (6 × 50 mL polycarbonate tubes)	Beckman Coulter		
SW 41 Ti rotor (6 × 13.2 mL ultraclear tubes)	Beckman Coulter		
SW 41 Ti rotor tubes (13.2 mL ultraclear tubes)	Beckman Coulter		

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

Instructions for Authors

Ultracentrifuge (Beckman L7)	Beckman Coulter		
Centrifuge (Beckman JSE-06D18)	Beckman Coulter		
Microcentrifuge	Eppendorf 5415D or equivalent		
Water pump connected to a Pasteur pipette via a plastic tube.			
Liquid nitrogen			
Peristaltic pump	Gilson		

509

Editor's Proof

Chapter 5

The Main Functions of Plastids

Norbert Rolland, Imen Bouchnak, Lucas Moyet, Daniel Salvi, and Marcel Kuntz

Abstract

Plastids are semiautonomous organelles like mitochondria, and derive from a cyanobacterial ancestor that was engulfed by a host cell. During evolution, they have recruited proteins originating from the nuclear genome, and only parts of their ancestral metabolic properties were conserved and optimized to limit functional redundancy with other cell compartments. Furthermore, large disparities in metabolic functions exist among various types of plastids, and the characterization of their various metabolic properties is far from being accomplished. In this review, we provide an overview of the main functions, known to be achieved by plastids or shared by plastids and other compartments of the cell. In short, plastids appear at the heart of all main plant functions.

Key words Chloroplast, Function, Metabolism, Plant, Proteomics

Uncorrected Proof

1 Origin and Evolution of Plastids

Endosymbiotic theory postulates that mitochondria and plastids were originally free-living prokaryotes and became organelles of eukaryotic cells [1]. Gene trees and other evolutionary studies indicate that eukaryotes arise from a first endosymbiotic event between an α-proteobacterium-like organism (the ancestor of mitochondria) and a host cell. Photosynthetic cells have later acquired an additional endosymbiont, a cyanobacterium-like prokaryote (the ancestor of plastids) more than 1.6 billion years ago [2–4]. During evolution, these organelles have conserved some traits of their origin. Present-day plastids from plant cells still contain a small genome that only codes for few percents of the protein content of the free-living cyanobacterial ancestor and of the present organelle. The divergence between the number of the remaining organelle-encoded proteins and the total number of proteins that plastids contain is generally explained by a consequence of gene transfer, from the organellar genome to the nucleus of the host cell during the course of evolution. As a consequence, the vast

Eric Maréchal (ed.), *Plastids: Methods and Protocols*, Methods in Molecular Biology, vol. 1829,
https://doi.org/10.1007/978-1-4939-8654-5_5, © Springer Science+Business Media, LLC, part of Springer Nature 2018

Editor's Proof

majority of plastid proteins from plant is now nucleus-encoded and must be imported to the plastid after translation in the cytosol [5, 6]. Prediction of the chloroplast proteome has been the subject of many debates. However, most recent analyses converge, and the plastid is estimated to contain ca. 3000 proteins while contemporary plastid genomes only contain ~100–150 genes [5].

2 Various Plastids for Diverse Functions

In higher plants, plastids are in reality a diverse group of organelles (Fig. 1) that are, for some of them, interconvertible during development [7]. All plastids share house-keeping functions (transcription/translation of the plastid genome), and basal metabolic functions that are essential to the cell metabolism. However, differentiation of plastids is always coupled to acquisition of specific roles.

Nongreen plastids are found in roots, seeds, tuber and some fruits. Proplastids, present in developing meristems (e.g., root tips, cauliflower buds), are undifferentiated nongreen plastids. Etioplasts accumulate in dark-grown seedlings and are rapidly converted to chloroplasts when exposed to light. Chloroplasts are undoubtedly the most characterized plastids. They are found in the cytoplasm of leaf cells, carry out the light-driven carbon fixation in the photosynthetic tissues. Gerontoplasts are senescent chloroplasts mostly

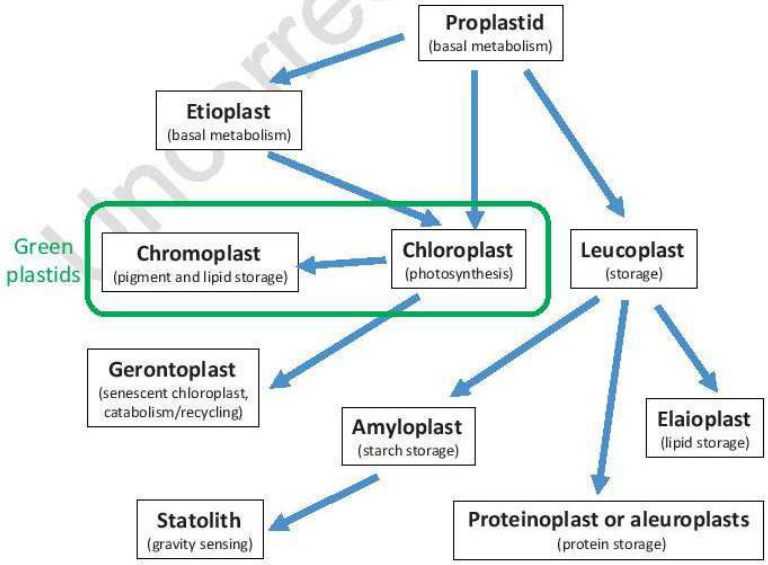


Fig. 1 Plastids are a diverse group of organelles that are, for some of them, interconvertible during development. These various plastids catalyze basic and more specific functions

Uncorrected Proof

Editor's Proof

devoted to catabolism and recycling of their components. Presenescing chloroplasts are sometimes positioned between chloroplasts and gerontoplasts. Chromoplasts are usually derived from chloroplasts and present in some fruits and flowers where they mostly accumulate carotenoids and lipids.

Leucoplasts are a family of plastids devoted to storage of various compounds. They can be converted to elaioplasts whose main function is lipid storage. They can also evolve to proteinoplasts (or aleuroplasts) who accumulate crystals of proteins. Leucoplasts can also be converted to amyloplasts, which are nonphotosynthetic plastids that contain large amounts of starch and play roles in storage of reduced carbon. In roots, gravity perception is sometimes explained by the starch-statolith hypothesis that states that dense starch-filled amyloplasts or statoliths within columella cells sediment in the direction of gravity, resulting in the generation of a signal that controls root growth [8].

3 Structure and Composition of the Plastid

Plastids contain several key subcompartments. The two limiting envelope membranes (inner and outer membrane of the plastid envelope with an intermembrane space between) are the only structure shared by all plastids [9]. The plastid envelope plays central roles in sustaining the communication (ions, metabolites, proteins, hormone precursors) of the chloroplast with the other compartments of the plant cell [10]. The aqueous phase of the plastid, also termed stroma, is mainly composed of soluble proteins whose role is devoted to most plastid-specific metabolic functions (carbon assimilation, metabolism of amino acids, vitamins, lipids, etc.). The thylakoid membrane system present in chloroplasts is a highly organized internal membrane network formed of flat compressed vesicles (grana) and more light structures (lamellae). This subcompartment is the center of oxygenic photosynthesis (light energy assimilation and conversion in ATP and reduced redox compounds). The thylakoid vesicles delimit another discrete soluble compartment, the thylakoid lumen, mostly devoted to the fueling of the ATPase activity in compartmenting protons resulting from cracking of the water molecule and from the linear photosynthetic electron flow (Fig. 2).

As mentioned above, the prediction of the plastid proteome has been the subject of many debates. This debate originally resulted from the discrepancy between proteomes predicted using various approaches (plastid protein targeting prediction tools versus protein similarities with the prokaryote ancestor of the plastid). Targeted and large-scale proteomic approaches have provided considerable knowledge about the compartmentation of the plant cell metabolism. Significant progresses have been made to establish the

Uncorrected Proof

Editor's Proof

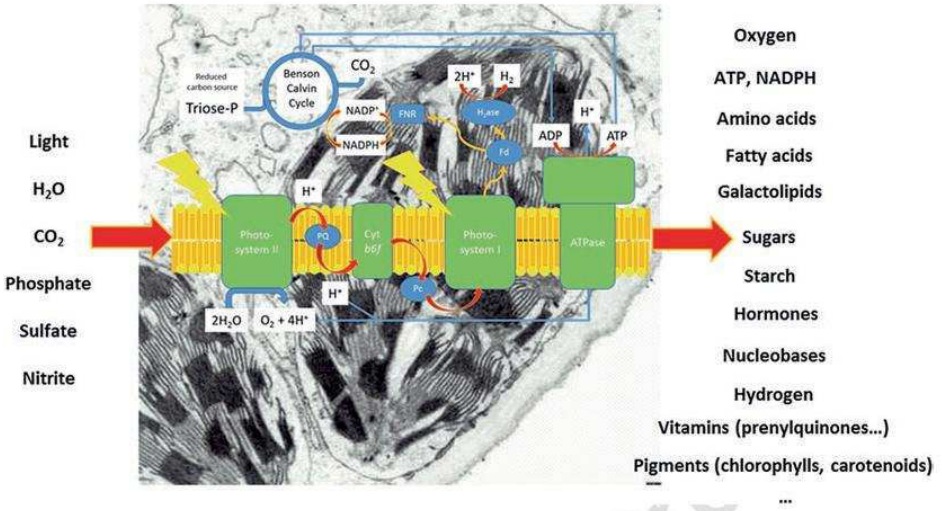


Fig. 2 The chloroplast, the energy factory of the photosynthetic cells, imports inorganic carbon, phosphate, and water for photosynthetic reactions. In turn, it produces reduced carbon, reducing power and ATP, which are used to fuel the cellular metabolism. From almost nothing (light, water, inorganic carbon, phosphate, sulfate, and nitrite), the plastid synthesizes almost everything (most essential organic compounds). Altogether, these compounds are the source of the biomass production for food, feed, energy, etc.

proteome of plastids using subcellular and subplastidial proteomics. Many investigations of the plastid proteome have focused on either whole chloroplast fractions or on independent subplastidial fractions within the last years [11]. As a result, several databases were created during the last decade, such as the Plastid Protein Database (plprot, http://www.plprot.ethz.ch) gathering plastid proteomics data [12], the Plant Protein Database (PPDB, http://ppdb.tc.cornell.edu) [13], the subcellular proteomics database SUBA (http://suba3.plantenergy.uwa.edu.au/) [14–16], the AT_CHLORO database (http://at-chloro.prabi.fr/at_chloro/) [17–19] that stores subplastidial (envelope, stroma, grana, or lamellae of thylakoids) localization of *Arabidopsis thaliana* chloroplast proteins, or the MASCP Gator database (http://gator.masc-proteomics.org/) combining all cell proteomics databases [20].

Altogether, these in-depth studies targeting both the composition and the subcellular localization of the plastid proteome now provide details on protein localization within the different cell and subplastidial compartments. These studies, complementing more specific studies targeted to single plastid proteins, were of great help to inventory the main functions carried by plastids.

However, the proteins identified to date using subcellular proteomics still represent a fraction of the complete plastid proteome. Indeed, some plastid proteins with highly particular functions are only present in specific plastid types (Fig. 1) within plant tissues

Uncorrected Proof

Editor's Proof

that are not accessible to subcellular fractionation approaches and proteomics. On the other hand, subcellular proteomics has identified hundreds of plastid proteins whose function remains unknown.

4 Main Metabolic Functions of the Plastids

While frequently only identified as the place where photosynthesis occurs, plastids, producing molecular oxygen that we breathe, also catalyze many other essential functions, including assimilation of nitrogen and sulfur, synthesis of carbohydrates (sugars and starch), amino acids, fatty acids and specific lipids, nucleotides, pigments, alkaloids, isoprenoids, hormone precursors, and vitamins (Fig. 2). During evolution, functions of the former cyanobacteria have been optimized to limit redundancy of metabolic pathways with other cell compartments. Consequently, plastids communicate and coordinate their various functions with other cell compartments, and a wide range of metabolites and ions now have to traffic between plastids and other cell compartments. Here, we highlight main plastid functions that can be deduced from our present knowledge of the compartmentation of the plant cell metabolism.

4.1 Photosynthesis

In plant leaves, the chloroplast imports CO_2, Pi, and H_2O as precursors for photosynthetic reactions. In turn, it produces reduced carbon in the form of triose-P through the Benson Calvin Cycle (BCC), reducing power (NADPH) and ATP, which are used to fuel the plastid and the cellular metabolisms (Fig. 2). In this context, it is important to note that, when compared to unicellular algae, plant chloroplasts lack most of the CO_2 concentrating mechanisms, or CCM [21]. Indeed, due to the lack of bicarbonate transporters similar to the ones found in cyanobacteria and algae, it is generally admitted that passive diffusion of CO_2 across biological membranes is sufficient to supply photosynthesis in higher plants. However, the chloroplast genome from higher plants contains a gene (*ycf10 or cemA*) whose deletion was shown to be detrimental for efficient uptake of inorganic carbon (CO_2 and HCO_3^-) into the chloroplast of *Chlamydomonas reinhardtii* [22]. This suggests that such a system could also be present in higher plant chloroplasts. Furthermore, another study performed in *Nicotiana tabacum* has suggested that CO_2 could be imported through an aquaporin (NtAQP1) present in the chloroplast envelope [23]. Finally, CO_2 assimilation is facilitated by carbonic anhydrases (catalyzing interconvertion of CO_2 into bicarbonate) present in the chloroplast stroma. Indeed, equilibrium of the CO_2 to bicarbonate conversion is displaced in favor of bicarbonate at alkaline pH. Since a ΔpH exist between the cytosol (7.4) and the stroma (up to 8) in the light, CO_2 diffusion within the chloroplast is driven by a difference in CO_2 concentration between these two compartments

Uncorrected Proof

Editor's Proof

resulting from the constant conversion of CO_2 into bicarbonate, catalyzed by stromal carbonic anhydrases. After its import, CO_2 is converted to reduced carbon through the BCC (Fig. 2). This reduced carbon, in the form of triose-P, is exported from the chloroplast (in a counter exchange with inorganic phosphate or Pi) for sucrose synthesis in the cytosol through the triose-P transporter [24]. Alternatively, this reduced carbon can be allocated to starch biosynthesis within the chloroplast. At night, stored carbohydrates are mostly exported via a maltose transporter [25]. Finally, chloroplasts not only fuel the cell metabolism with sugars, but can also export reducing power (NADPH) and energy (ATP) through dedicated envelope transporters [10, 26].

4.2 Synthesis of Amino Acids

The chloroplast was demonstrated to be able to synthesize 17 of the 20 proteinaceous amino acids (Fig. 2) [for a review, see ref. 10]. Ten (Arg, Lys, Thr, Leu, Ile, Val, Trp, Phe, Tyr, His) of these amino are only synthesized within the chloroplast while seven (Asp, Cys, Gln, Glu, Gly, Ser, Met) of them are also synthesized in other cell compartments. Chloroplasts thus provide many amino acids to the other cell compartments (at least for protein biosynthesis in the cytosol and the mitochondria) and may have to import Asn, Ala, and Pro, but the subcellular localization of the metabolic pathways involved in the synthesis of the latter still remains unclear. However, the nature of most of the chloroplast envelope transporters that would catalyze these import or export reactions remains unclear [10, 27, 28]. In this context, knowing that chloroplasts are the place where nitrogen assimilation occurs (through reduction of plastid-imported nitrite, nitrate being already reduced in the cytosol), it is important to note that export of reduced nitrogen to the cytosol in the form of the amino acid glutamate occurs through the glutamate/malate antiporter present in the chloroplast envelope [29].

4.3 Synthesis of Pigments

Carotenoids are C40 terpenoids synthesized in plant plastids [30] and by certain micro-organisms, and which function as light-harvesting and photoprotective pigments [31]. Carotenoids are responsible for the distinctive colors of chromoplasts. They also provide precursors for signaling molecules such as abscisic acid, a phytohormone which has roles in stress tolerance, seed development, and sugar sensing, or strigolactones which regulate various developmental processes including plant shoot branching [32]. Apocarotenoids (derived from carotenoids by oxidative cleavage) are organic compounds which occur widely in all living organisms and are also important animal diet component (e.g., vitamin A retinoids retinal, retinoic acid, and retinol). Like all terpenoids, carotenoids are synthesized through condensations of isopentenyl diphosphate and its isomer dimethylallyl diphosphate. The most visible chloroplast pigments, chlorophylls, are specifically inte-

Uncorrected Proof

Editor's Proof

grated into protein pigment complexes embedded within photosynthetic membranes or thylakoids. Although devoid of chlorophyll, plastid envelope membranes contain low amounts of chlorophyll precursors, namely chlorophyllide and protochlorophyllide.

Most proteins involved in the biosynthesis of soluble precursors (isopentenylpyrophosphate, geranylgeranyl pyrophosphate, tetrapyrroles, etc.) for chlorophyll, carotenoids, or prenylquinones are unequivocally localized in the plastid stroma. Then, protoporphyrin IX is formed and is directed into different metabolic pathways: (1) a ferrochelatase (required for heme synthesis) tightly linked to thylakoids, and (2) chlorophyll biosynthesis which takes place in both thylakoids and envelope membranes. In contrast, today proteomic data found a complete set of carotenogenic enzyme only in the enveloppe [17, 30]. This envelope localization would be in agreement with the observation that envelope membranes are the only permanent membrane structure of all types of plastids, and to the carotenoids and prenylquinones requirement of all plastids at least for their protection against oxidative stress.

4.4 Synthesis of Fatty Acids and Lipids

De novo biosynthesis of fatty acids occurs in the stroma of plastids [33, 34]. These fatty acids are precursors required for the immense demand of cells, to synthesize membrane glycerol- and sphingolipids, triacylglycerols, or precursors for very-long-chain-fatty acids, epicuticular waxes, or oxylipins, like jasmonates. Export of plastid-synthesized fatty acids to the cytosol feeds the cytosolic pool of acyl-CoAs, which, with glycerol-3-phosphate, are precursors for the synthesis of lyso-phosphatidic acid and phosphatidic acid in the endoplasmic reticulum. From phosphatidic acid and diacylglycerol, endoplasmic reticulum enzymes synthesize all phospholipids (phosphatidylglycerol, phosphatidylinositol, phosphatidylserine, phosphatidylethanolamine, and phosphatidylcholine). The plastids membranes have a unique lipid composition, mostly containing nonphosphated glycoglycerolipids (monogalactosylglycerol, digalactosyldiacylglycerol, sulfoquinovosyldiacylglycerol) and low proportions of phospholipids (phosphatidylglycerol, phosphatidylcholine and phosphatidic acid). Phosphatidylinositol is strictly located in the outer leaflet of the outer envelope membrane bilayer [35].

The plastid envelope assembles glycerolipids, using either prokaryotic substrates freshly synthesized in the plastid (thus harbor a diacyl structure known as "prokaryotic," similar to that observed in cyanobacteria glycerolipids) or imported eukaryotic lipid precursors derived from extraplastidial phospholipids after their hydrolysis by phospholipases [34, 36, 37]. Interestingly, while galactolipids have long been considered to be plastid-specific lipids, it was recently shown that when plants meet inorganic phosphate starvation, synthesis of galactolipids increases and digalactosyldiacylglycerol is exported to the plasma membrane, the

Uncorrected Proof

Editor's Proof

tonoplast, and the mitochondria, where it can substitute to phospholipids [38–40]. Finally, plastid galactolipids provide polyunsaturated fatty acids precursors for the biosynthesis of triacylglycerols, oxylipins (including jasmonic acids) and cyclooxylipin-galactolipids (like "arabidopsides") that accumulate under specific conditions including pathogen attacks [41].

4.5 Synthesis of Vitamins and Vitamin-Derived Cofactors

Vitamin deficiencies are major forms of micronutrient deficiencies, and are associated with huge economic losses as well as severe physical and intellectual damages to humans. Plastids play a unique role in the synthesis of several vitamins and their derived cofactors (Fig. 2). They synthesize tocopherols and tocotrienols (antioxidant molecules, limiting lipid peroxidation and collectively termed vitamin E), phylloquinone (electron transfer cofactor, termed vitamin K1), and carotenoids (pro-vitamin A, see above). Like carotenoids, enzymes involved in vitamins E and K1 biosynthesis are almost exclusively localized within the plastid envelope [30]. Thiamine diphosphate is the active form of thiamine (vitamin B1), which is produced in plastids [42]. Vitamin B1 derivatives should be exchanged across the chloroplast envelope. While thiamin diphosphate carriers were recently identified in mitochondria, the same is not true for plastids [43]. Lipoate, an antioxidant sometimes called a vitamin-like chemical, is also synthesized in plastids [44]. Flavin nucleotides (FMN and FAD, organic redox cofactors) originate from riboflavin (vitamin B2) that is also synthesized in plastids. Plastids have to export riboflavin to the cytosol through yet unknown flavin transporters of the plastidial envelope [45].

While plastids that are autonomous for the synthesis of the above cited coumpounds, they rely on the import of others. Although ascorbate (vitamin C) plays a major role in scavenging reactive oxygen species and is the most abundant antioxidant in chloroplasts [46], it is synthesized outside the chloroplast and is imported into plastids through a recently identified transporter [47]. The same is true for three cofactors specifically involved in transfer of one-carbon (C1) units, folates (vitamin B9), adenosylmethionine (AdoMet), and biotin (vitamin B8), that are produced in the cytosol, although the synthesis of folates is initiated in the plastid stroma. To date, specific envelope transporters to import folates, AdoMet and biotin from the cytosol remain unidentified. Again, Pyridoxal 5′-phosphate, a vitamin B6 derivative that is synthesized in the cytosol, should be imported into plastids thanks to a yet unidentified envelope transporter [48]. The synthesis of CoA from pantothenate (vitamin B5) is a five-step process located in the cytosol and must be imported into plastids through an envelope transporter, so far not identified. As for vitamin C, a mitochondrial transporter was recently identified [49], but the plastid one remains unknown. Finally, while the first two reactions involved in synthesis of pyridine nucleotides NAD(P) which are derivatives of nicotinic

Uncorrected Proof

Editor's Proof

acid (vitamin B3) occur in plastids, the subsequent steps leading to NAD synthesis are likely cytosolic [50]. Compartmentation of pyridine nucleotides synthesis thus relies on plastid envelope transporters to export quinolinate and to import NAD [51].

4.6 Ionic Cell Homeostasis

The plastids contain and store a large diversity of ions, which are essential for their development and functions. These ions play key roles in processes like signaling (calcium), generation of membrane potential (potassium, proton), electron transports (copper, magnesium, manganese, iron, and chloride), ATP biosynthesis (phosphate), Benson Calvin cycle (bicarbonate), biosynthesis of cysteine and methionine (sulfate), and ammonium assimilation (nitrite). The two limiting envelope membranes (inner and outer membrane of the plastid envelope) thus contain a series of transporters that are not only a remnant of the bacterial ancestor, but are of primary importance for the integrated metabolisms of plastids. Earliest studies had suggested the existence of other transporters for which ion transport activities have been measured physiologically but the corresponding proteins are either still not identified of characterized. Furthermore, it is important to note that several transporters with identical ionic specificity (e.g., copper, phosphate) are found in the plastid envelope. Although these transporters could differ in their substrate affinities, association to specific plastid types or differential expression during development, this apparent functional redundancy suggests that plastids need various similar transporters to participate to the ionic cell homeostasis. Despite the importance of these transporters for the chloroplast metabolism, only few of them have been recently characterized [10, 52–54]. During the last decade, several targeted proteomic-based studies were performed to identify envelope proteins including ion transporters (*see* ref. 17 and references therein). During these studies, most previously known transporters were identified, but also dozens of proteins that share structural similarity with transporters for which no functional data are available. Based on either reverse genetic studies or biochemical approaches, an active field of research aims to allow identifying the ionic specificity and function of these still uncharacterized plastid envelope transporters [55–61].

5 Conclusion

As depicted in the simplified view in Fig. 3, plastids are at the heart of all main plant functions. Light, water, inorganic carbon, starch, oxygen, phosphate, sulfate, and nitrite are all assimilated, stored, converted, or produced, thanks to the plastid metabolism. From these inorganic coumpounds, plastids produce most essential organic compounds, but also reducing power and ATP, which are used to fuel the cellular metabolism. Altogether, these

Uncorrected Proof

Editor's Proof

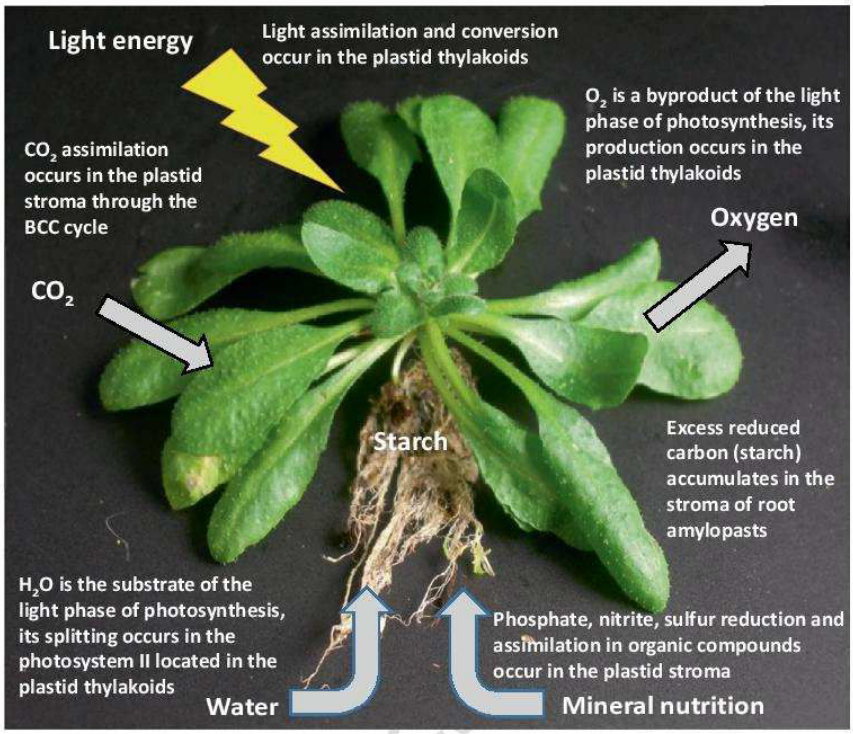


Fig. 3 In a simplified view, plastids are at the heart of all main plant functions. Light, water, inorganic carbon, starch, oxygen, phosphate, sulfate, and nitrite are all assimilated, stored, converted, or produced, thanks to the plastid metabolism

compounds are the source of the biomass production for food, feed, fiber, energy, etc.

Enormous progress has been achieved in understanding the compartmentation of the plant metabolism, thanks to both targeted studies and large-scale approaches. Efforts to perform the inventory of plastid proteins have progressed rapidly. The size and content of the plastid proteome are variable depending on the plant tissues or the environmental conditions. The proteome of different plastid types has been described, although not to the same extent as chloroplasts [11]. However, when searching to analyze the subcellular localization of plastid proteins, it appeared that around 30% of the experimentally localized proteins detected in the Arabidopsis chloroplast have no predictable transit peptides [5, 12, 17] and thus escape prediction tools. Furthermore, a number of plastid proteins have been reported not to follow a canonical import pathway [5] and the physiological meanings of alternative protein targeting pathways that accumulate specific proteins into plastids remain unknown. Based on all these observations, the evaluation of a plas-

Uncorrected

Editor's Proof

tid proteome, on the basis of software-based protein localization prediction, yields partial and unreliable data. These arguments strongly support further subcellular proteomics analyses, and more targeted traditional biochemical approaches, to provide experimental data about the proteome dynamics of various plastids.

As described in this review, many plastid envelope transporters by which plastids exchange ions and metabolites remain to be identified, and dozens of transporter homologues were identified in the chloroplast envelope through targeted proteomic approaches whose function remain undefined [10]. Reverse genetic studies, combined with other approaches, will be of great help to identify the specificity of these yet uncharacterized envelope transporters and to allow defining their respective function in connecting metabolic pathways present in the plastid and other cell compartments.

There is also a strong need to further integrate metabolic functions [e.g., [62]] of the various plastid types present in different plant tissues helping understanding metabolism compartmentation, its evolution following endosymbiosis of the plastid cyanobacterial ancestor, and its dynamics at the scale of the whole cell metabolism. Furthermore, additional analyses are required to identify mechanisms (protein interactions, protein posttranslational modifications) regulating complete metabolic pathways, including pathways that are shared by the plastid and other cell compartments. Altogether, these studies are expected to make the plastid organelle a prime target for systems biology research in plants.

Uncorrected Proof

Acknowledgments

Authors regret omission of many relevant citations due to space constraints. N.R., I.B., L.M., D.S. and M.K. acknowledge support from the ANR project ANR-15-IDEX-02. I.B. is supported by a joint PhD fellowship from the INRA Plant Biology and Breeding Division and from the Labex GRAL (ANR-10-LABX-49-01).

References

1. Zimorski V, Ku C, Martin WF et al (2014) Endosymbiotic theory for organelle origins. Curr Opin Microbiol 22:38–48
2. Bogorad L (2008) Evolution of early eukaryotic cells: genomes, proteomes, and compartments. Photosynth Res 95:11–21
3. Reyes-Prieto A, Weber AP, Bhattacharya D (2007) The origin and establishment of the plastid in algae and plants. Annu Rev Genet 41:147–168
4. Yoon HS, Hackett JD, Ciniglia C et al (2004) A molecular timeline for the origin of photosynthetic eukaryotes. Mol Biol Evol 21:809–818
5. Jarvis P, L pez-Juez E (2013) Biogenesis and homeostasis of chloroplasts and other plastids. Nat Rev Mol Cell Biol 14:787–802
6. Lee DW, Lee J, Hwang I (2017) Sorting of nuclear-encoded chloroplast membrane proteins. Curr Opin Plant Biol 40:1–7
7. Inaba T, Ito-Inaba Y (2010) Versatile roles of plastids in plant growth and development. Plant Cell Physiol 51:1847–1853
8. Sato EM, Hijazi H, Bennett MJ et al (2015) New insights into root gravitropic signaling. J Exp Bot 66(8):2155–2165

Editor's Proof

9. Block MA, Douce R, Joyard J et al (2007) Chloroplast envelope membranes: a dynamic interface between plastids and the cytosol. Photosynth Res 92:225–244
10. Rolland N, Curien G, Finazzi G et al (2012) The biosynthetic capacities of the plastids and integration between cytoplasmic and chloroplast processes. Annu Rev Genet 46:233–264
11. Agrawal GK, Bourguignon J, Rolland N et al (2011) Plant organelle proteomics: collaborating for optimal cell function. Mass Spectrom Rev 30(5):772–853
12. Kleffmann T, Hirsch-Hoffmann M, Gruissem W et al (2006) Plprot: a comprehensive proteome database for different plastid types. Plant Cell Physiol 47(3):432–436
13. Sun Q, Zybailov B, Majeran W et al (2009) PPDB, the plant proteomics database at Cornell. Nucleic Acids Res 37(Database issue):D969–D974
14. Heazlewood JL, Verboom RE, Tonti-Filippini J et al (2007) SUBA: the Arabidopsis subcellular database. Nucleic Acids Res 35(Database issue):D213–D218
15. Tanz SK, Castleden I, Hooper CM et al (2013) SUBA3: a database for integrating experimentation and prediction to define the SUBcellular location of proteins in Arabidopsis. Nucleic Acids Res 41:1185–1191
16. Hooper CM, Castleden IR, Tanz SK et al (2017) SUBA4: the interactive data analysis Centre for Arabidopsis subcellular protein locations. Nucleic Acids Res 45(D1):D1064–D1074
17. Ferro M, Brugi re S, Salvi D et al (2010) AT_CHLORO, a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins. Mol Cell Proteomics 9(6):1063–1084
18. Bruley C, Dupierris V, Salvi D et al (2012) AT_CHLORO: a chloroplast protein database dedicated to sub-Plastidial localization. Front Plant Sci 3:205
19. Tomizioli M, Lazar C, Brugi re S et al (2014) Deciphering thylakoid sub-compartments using a mass spectrometry-based approach. Mol Cell Proteomics 13(8):2147–2167
20. Joshi HJ, Hirsch-Hoffmann M, Baerenfaller K et al (2011) MASCP gator: an aggregation portal for the visualization of Arabidopsis proteomics data. Plant Physiol 155(1):259–270
21. Raven JA (2010) Inorganic carbon acquisition by eukaryotic algae: four current questions. Photosynth Res 106:123–134
22. Rolland N, Dorne A-J, Amoroso G et al (1997) Disruption of the plastid *ycf10* open reading frame affects uptake of inorganic carbon in the chloroplast of Chlamydomonas. EMBO J 16:6713–6726
23. Uehlein N, Otto B, Hanson DT et al (2008) Function of Nicotiana tabacum aquaporins as chloroplast gas pores challenges the concept of membrane CO_2 permeability. Plant Cell 20:648–657
24. Fl gge UI, Fischer K, Gross A et al (1989) The triose phosphate-3-phosphoglyceratephosphate translocator from spinach chloroplasts: nucleotide sequence of a full-length cDNA clone and import of the in vitro synthesized precursor protein into chloroplasts. EMBO J 8:39–46
25. Niittyl T, Messerli G, Trevisan M et al (2004) A previously unknown maltose transporter essential for starch degradation in leaves. Science 303:87–89
26. Facchinelli F, Weber AP (2011) The metabolite transporters of the plastid envelope: an update. Front Plant Sci 2:50
27. Weber APM, Schwacke R, Flugge UI (2005) Solute transporters of the plastid envelope membrane. Annu Rev Plant Biol 56:133–164
28. Oh YJ, Hwang I (2015) Targeting and biogenesis of transporters and channels in chloroplast envelope membranes: unsolved questions. Cell Calcium 58(1):122–130
29. Renne P, Dressen U, Hebbeker U et al (2003) The Arabidopsis mutant dct is deficient in the plastidic glutamate/malate translocator DiT2. Plant J 35:316–331
30. Joyard J, Ferro M, Masselon C et al (2009) Chloroplast proteomics and the compartmentation of plastidial isoprenoid biosynthetic pathways. Mol Plant 2:1154–1180
31. Sun T, Yuan H, Cao H et al (2017) Carotenoid metabolism in plants: the role of plastids. Mol Plant (in press). https://doi.org/10.1016/j.molp.2017.09.010
32. Siddiqi KS, Husen A (2017) Plant response to strigolactones: current developments and emerging trends. Appl Soil Ecol 120:247–253
33. Brown AP, Slabas AR, Rafferty JB (2010) Fatty acid biosynthesis in plants - metabolic pathways, structure and organization. In: Wada H, Murata N (eds) Lipids in photosynthesis. Springer, Dordrecht, pp 11–34
34. Jouhet J, Dubots E, Mar chal E et al (2010) Lipid trafficking in plant photosynthetic cells. In: Wada H, Murata N (eds) Lipids in photosynthesis. Springer, Dordrecht, pp 349–372
35. Dorne AJ, Joyard J, Block MA et al (1985) Localization of phosphatidylcholine in outer envelope membrane of spinach-chloroplasts. J Cell Biol 100:1690–1697

Uncorrected Proof

Editor's Proof

36. Benning C (2009) Mechanisms of lipid transport involved in organelle biogenesis in plant cells. Annu Rev Cell Dev Biol 25:71–91
37. Joyard J, Ferro M, Masselon C et al (2010) Chloroplast proteomics highlights the subcellular compartmentation of lipid metabolism. Prog Lipid Res 49:128–158
38. Jouhet J, Marechal E, Baldan B et al (2004) Phosphate deprivation induces transfer of DGDG galactolipid from chloroplast to mitochondria. J Cell Biol 167:863–874
39. Sz kely G, Abrah m E, Cs plo A et al (2008) Duplicated P5CS genes of Arabidopsis play distinct roles in stress regulation and developmental control of proline biosynthesis. Plant J 53:11–28
40. Tjellstrom H, Andersson MX, Larsson KE et al (2008) Membrane phospholipids as a phosphate reserve: the dynamic nature of phospholipid-to-digalactosyl diacylglycerol exchange in higher plants. Plant Cell Environ 31:1388–1398
41. Andreou A, Brodhun F, Feussner I (2009) Biosynthesis of oxylipins in non-mammals. Prog Lipid Res 48:148–170
42. Goyer A (2010) Thiamine in plants: aspects of its metabolism and functions. Phytochemistry 71:1615–1624
43. Frelin O, Agrimi G, Laera VL et al (2012) Identification of mitochondrial thiamin diphosphate carriers from Arabidopsis and maize. Funct Integr Genomics 12(2):317–326
44. Rebeille F, Alban C, Bourguignon J et al (2007) The role of plant mitochondria in the biosynthesis of coenzymes. Photosynth Res 92:149–162
45. Azevedo RA, Lancien M, Lea PJ (2006) The aspartic acid metabolic pathway, an exciting and essential pathway in plants. Amino Acids 30:143–162
46. Foyer CH, Noctor G (2011) Ascorbate and glutathione: the heart of the redox hub. Plant Physiol 155:2–18
47. Miyaji T, Kuromori T, Takeuchi Y et al (2015) AtPHT4;4 is a chloroplast-localized ascorbate transporter in Arabidopsis. Nat Commun 6:5928
48. Mooney S, Hellmann H (2010) Vitamin B6: killing two birds with one stone? Phytochemistry 71:495–501
49. Zallot R, Agrimi G, Lerma-Ortiz C et al (2013) Identification of mitochondrial coenzyme a transporters from maize and Arabidopsis. Plant Physiol 162(2):581–588
50. Noctor G, Queval G, Gakiere B (2006) NAD(P) synthesis and pyridine nucleotide cycling in plants and their potential importance in stress conditions. J Exp Bot 57:1603–1620
51. Palmieri F, Rieder B, Ventrella A et al (2009) Molecular identification and functional characterization of *Arabidopsis thaliana* mitochondrial and chloroplastic NAD^+ carrier proteins. J Biol Chem 284:31249–31259
52. Finazzi G, Petroutsos D, Tomizioli M et al (2015) Ions channels/transporters and chloroplast regulation. Cell Calcium 58(1):86–97
53. Xu H, Martinoia E, Szabo I (2015) Organellar channels and transporters. Cell Calcium 58(1):1–10
54. L pez-Mill n AF, Duy D, Philippar K (2016) Chloroplast iron transport proteins - function and impact on plant physiology. Front Plant Sci 7:178
55. Kunz HH, Gierth M, Herdean A et al (2014) Plastidial transporters KEA1, −2, and −3 are essential for chloroplast osmoregulation, integrity, and pH regulation in Arabidopsis. Proc Natl Acad Sci U S A 111(20):7480–7485
56. Boutigny S, Sautron E, Finazzi G et al (2014) HMA1 and PAA1, two chloroplast-envelope PIB-ATPases, play distinct roles in chloroplast copper homeostasis. J Exp Bot 65(6):1529–1540
57. Maeda S, Konishi M, Yanagisawa S et al (2014) Nitrite transport activity of a novel HPP family protein conserved in cyanobacteria and chloroplasts. Plant Cell Physiol 55(7):1311–1324
58. Sautron E, Mayerhofer H, Giustini C et al (2015) HMA6 and HMA8 are two chloroplast cu+-ATPases with different enzymatic properties. Biosci Rep 35(3):e00201
59. Goetze TA, Patil M, Jeshen I et al (2015) Oep23 forms an ion channel in the chloroplast outer envelope. BMC Plant Biol 15:47
60. Aranda-Sicilia MN, Aboukila A, Armbruster U et al (2016) Envelope K+/H+ Antiporters AtKEA1 and AtKEA2 function in plastid development. Plant Physiol 172(1):441–449
61. Sautron E, Giustini C, Dang T et al (2016) Identification of two conserved residues involved in copper release from chloroplast PIB-1-ATPases. J Biol Chem 291(38):20136–20148
62. Gloaguen P, Bournais S, Alban C et al (2017) ChloroKB: a web application for the integration of knowledge related to chloroplast metabolic network. Plant Physiol 174(2):922–934

Unedited Proof

Editor's Proof

Chapter 26

AT_CHLORO: The First Step When Looking for Information About Subplastidial Localization of Proteins

Daniel Salvi, Sylvain Bournais, Lucas Moyet, Imen Bouchnak, Marcel Kuntz, Christophe Bruley, and Norbert Rolland

Abstract

Plastids contain several key subcompartments. The two limiting envelope membranes (inner and outer membrane of the plastid envelope with an intermembrane space between), an aqueous phase (stroma), and an internal membrane system terms (thylakoids) formed of flat compressed vesicles (grana) and more light structures (lamellae). The thylakoid vesicles delimit another discrete soluble compartment, the thylakoid lumen. AT_CHLORO (http://at-chloro.prabi.fr/at_chloro/) is a unique database supplying information about the subplastidial localization of proteins. It was created from simultaneous proteomic analyses targeted to the main subcompartments of the chloroplast from *Arabidopsis thaliana* (i.e., envelope, stroma, thylakoid) and to the two subdomains of thylakoid membranes (i.e., grana and stroma lamellae). AT_CHLORO assembles several complementary information (MS-based experimental data, curated functional annotations and subplastidial localization, links to other public databases and references) which give a comprehensive overview of the current knowledge about the subplastidial localization and the function of chloroplast proteins, with a specific attention given to chloroplast envelope proteins.

Key words Chloroplast, Subcellular localization, Envelope, Stroma, Thylakoid, Plant, Proteomics

1 Introduction

It is now widely accepted that plastids derive from endosymbiosis of an originally free-living cyanobacteria [1]. Evolutionary analyses of cyanobacteria, plant nucleus and chloroplast genomes, have suggested that gene transfer of ~4500 genes occurred from this cyanobacterial ancestor to the nuclear genome of present plant cells [2]. Surprisingly, less than half of the 4500 genes transferred to the host genome now code for proteins that are still targeted to plastids while other proteins of prokaryotic origin are addressed to other compartments of the host cell. Unexpectedly, present-day plastids also contain hundreds of other proteins that are coded by nonendosymbiont genes (i.e., genes from the original host cell). Altogether, ~3000 proteins are now nuclear encoded and targeted to plastids through

Eric Maréchal (ed.), *Plastids: Methods and Protocols*, Methods in Molecular Biology, vol. 1829,
https://doi.org/10.1007/978-1-4939-8654-5_26, © Springer Science+Business Media, LLC, part of Springer Nature 2018

Editor's Proof

yet highly studied plastid targeting machineries [3], while less than 100–150 genes are still present in the plastid genome.

After their import into plastids, these 3000 proteins are targeted to various plastidial subcompartments, and their ending subplastidial localization is directly linked to their function within plastids. Starting from the plastid surface, the outer envelope membrane contains porins and channels for ions and metabolites, receptors of protein targeting machineries (Toc translocon), members of the plastid division or positioning machineries, but also enzymes that are essential for lipid metabolism or synthesis of hormone precursors. Similarly, the inner envelope membrane contains more than 70 transporters for ions and metabolites, a protein targeting machinery (Tic translocon), more than 20 chaperones and proteases, more than 60 enzymes of the lipid, vitamin, pigment, and hormone metabolism, still few members of the plastid division machinery, and more than hundred proteins of yet unknown (and unpredictable) function [4]. Altogether, these limiting outer and inner envelope membranes thus play central roles in sustaining the communication (ions, metabolites, proteins, lipids, hormone precursors) of the chloroplast with the other compartments of the plant cell, but also play key roles in metabolic functions [5]. The stroma, the aqueous phase of plastids, is mainly composed of soluble proteins whose role is devoted to most plastid-specific metabolic functions: carbon assimilation (the "dark" phase of photosynthesis), metabolism of amino acids, vitamins, pigments, lipids, nucleotides, etc. It also contains the plastid transcription and translation machineries, various DNA- or RNA-binding proteins, many chaperones and proteases and many proteins of yet uncharacterized functions. Green plastids contain additional compartments. The thylakoid membrane system is a highly organized internal network formed of flat compressed (grana) and more light (lamellae) vesicles delimiting an additional soluble compartment, the thylakoid lumen. Thylakoids are the place where the "light" phase of photosynthesis occurs: assimilation of light energy and its conversion into ATP and reduced redox compounds. However, thylakoid membranes also contain lots of chaperones and proteases, regulatory proteins (kinases, phosphatases, stress and signaling proteins), enzymes involved in lipid and pigment metabolism, a few ion transporters, components of protein targeting machineries and again, many proteins of yet uncharacterized functions [4].

As mentioned above, the subcellular localization of proteins is directly linked to their function within the cell. Defining the composition of the plastid proteome was thus of major interest to support studies aiming to help characterize identified proteins whose function remain undefined (e.g., an essential kinase, whose localization was unknown) or to identify proteins carrying specific functions, but whose primary sequence remained unknown, (e.g., a specific step in a metabolic pathway). During the last two decades,

Uncorrected Proof

Editor's Proof

major progresses have been made to establish the subcellular localization of plastid proteins using protein similarities with the prokaryote ancestor of the plastid [2], plastid protein targeting prediction tools (e.g., ChloroP [6], TargetP [7], or subcellular and subplastidial proteomics [8]. Consequently, and due to the extent of protein lists, issued from these analyses, specific databases needed to be generated, that are still highly useful for present-day studies targeted to single plastid proteins.

Among the various plant protein databases devoted to the subcellular proteomes, are the Plastid Protein Database (plprot, http://www.plprot.ethz.ch) [9], the Plant Protein Database (PPDB, http://ppdb.tc.cornell.edu) [10], the MASCP Gator database (http://gator.masc-proteomics.org/) [11], the SUBA database (http://suba3.plantenergy.uwa.edu.au/) [12–14], and the AT_CHLORO database (http://at-chloro.prabi.fr/at_chloro/) [4, 15, 16]. AT_CHLORO specifically stores information about the subplastidial localization of plastid proteins. It was created from in depth proteomic analyses targeted to the various subcompartments (i.e., envelope, stroma, thylakoid grana and lamellae) of the chloroplast from *Arabidopsis thaliana*. AT_CHLORO assembles several complementary information (MS-based experimental data, curated functional annotations, public information) which gives a comprehensive overview of current knowledge about the subplastidial localization and the function of identified chloroplast proteins. A specific attention was given to proteins identified in chloroplast envelope fractions, for which expert functional annotation was performed. Some data from AT_CHLORO can also be retrieved from the MASCP Gator Portal [11], the aggregation portal for the visualization of Arabidopsis proteomics data. Information related to the subplastidial localization was also submitted to TAIR (https://www.arabidopsis.org/), The Arabidopsis Information Resource database.

The present chapter aims to provide a short overview of specificities and usefulness of the AT_CHLORO database with respect to stored information, data extraction, and recent improvements.

2 Materials

2.1 MS/MS Data Obtained from Purified Plastidial Subcompartments to Feed AT_CHLORO

1. The AT_CHLORO database stores information about proteins that have been identified in Percoll-purified chloroplasts from *Arabidopsis thaliana*. Additional steps (fractionation on sucrose gradients, and/or detergent solubilization and differential centrifugations) allowed to isolate and analyze different subplastidial compartments (*see* Fig. 1) from these chloroplasts. Before MS/MS analyses of the triplicate biological samples, cross-contamination was evaluated using western blot analyses [17–20].

Uncorrected Proof

Editor's Proof

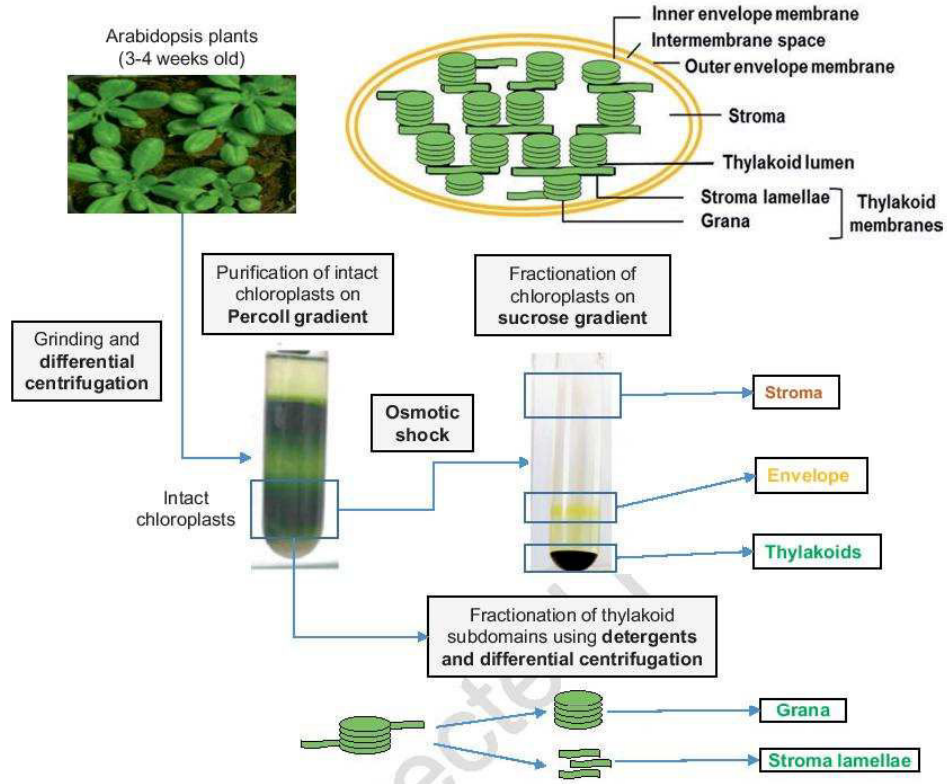


Fig. 1 Upstream experimental strategy used to feed the AT_CHLORO database

2. During the first experiment, subplastidial localization of plastid proteins was deduced from MS-based analyses of the three main chloroplast compartments (i.e., envelope, thylakoids, and stroma) analyzed in the same experiment, and starting from the same chloroplast samples [4]. It is important to note here that these three main chloroplast compartments were isolated with a low level of cross-contamination, as determined by western blot analyses [4]. From the 500 LC-MS/MS analyses performed in this context (1323 identified proteins), normalized and semi-quantitative spectral count data were used to assess relative abundances of 819 proteins in the three subcompartments. This study allowed for identification of numerous novel envelope proteins and provided information about the distribution of transporters or metabolic pathways at the chloroplast scale [4, 5, 21, 22]. It was also used as a substance to perform a graphic representation of the chloroplast metabolism [23].
3. The objective of the second experiment was to gain further information about the protein segregation in the subdomains

Uncorrected

Editor's Proof

of thylakoid membranes [16]. As for the first experiment, MS-based analyses of the two purified compartments of thylakoids (i.e., grana and stroma lamellae purified in the same experiment) were performed. Again, the level of cross-contamination (less than 5%) of the two purified fractions was quantified using western blot analyses, and a semiquantitative approach was completed to analyze the differential distribution of the 1290 thylakoid proteins identified in these two subcompartments. This study not only confirmed previous information about the localization of photosynthetic proteins but also identified approximately hundred new minor thylakoid or chloroplast proteins and revealed differential distribution of some proteins in the grana or the stroma-lamellae subcompartments of the thylakoid [16].

2.2 Map of the AT_CHLORO Database

1. AT_CHLORO is organized around several types of pages (*see* Fig. 2). The "Home page" (http://at-chloro.prabi.fr/at_chloro/) provides an overview of MS/MS experiments that were performed to implement the database, the list of methods used for subcellular and subplastidial fractionation of the chlo-

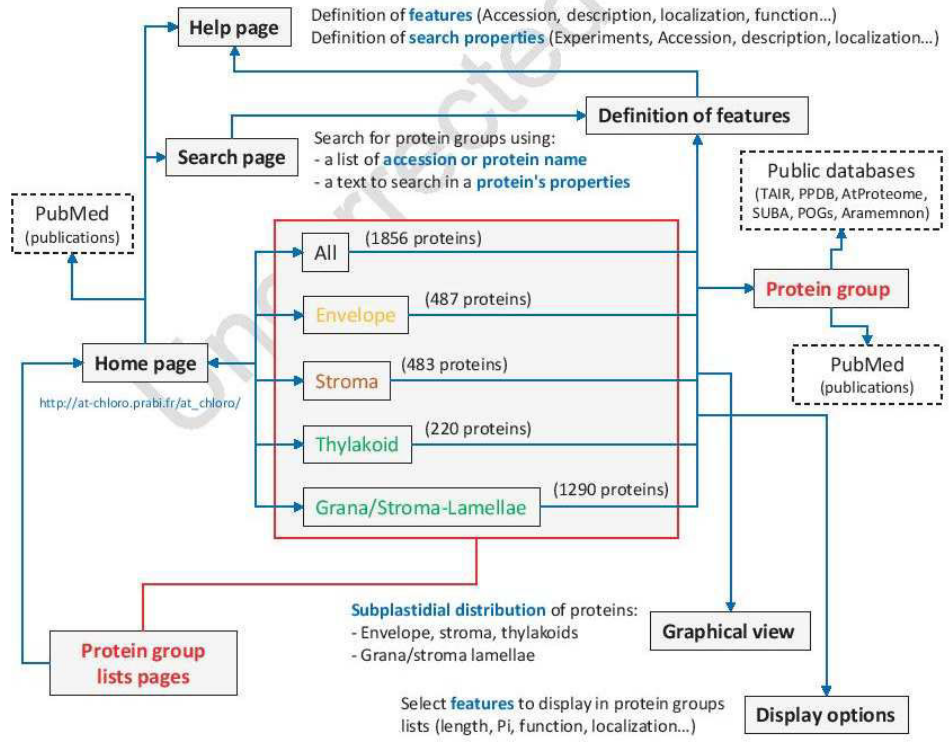


Fig. 2 Map of the AT_CHLORO database

Editor's Proof

roplast from *Arabidopsis thaliana* and a short publication list (with links to PubMed) targeting the direct application of AT_CHLORO to revisit the subplastidial localization of specific metabolic pathways (*see* **Note 1**).

From this "Home page," users have access to the "Help page" (see "How to use the AT_CHLORO database") (*see* **Note 2**). This page first provides every information about "features" that are listed in the "Protein group lists" and "Protein group" pages. This page also defines "search properties" that can be used to screen the database searching for specific groups of proteins (see below).

2. This "Home page" also links to several "Protein group lists" pages: All proteins (1856), Envelope proteins (487), Stroma proteins (483), Thylakoid proteins (220), and Grana/Stroma-Lamellae proteins (1290). It is important to note here that these group lists not only contain proteins that were only detected in corresponding subcellular fractions, but also proteins that are shared between this compartment and another one, and some abundant cell components cross-contaminating the analyzed fractions (*see* **Note 3**).

   From these "Protein group lists" pages, access is allowed to "Protein group" pages (*see* Fig. 3) that provide specific information about every entry. These "Protein group" pages contain links to other public databases (TAIR, PPDB, AtProteome, SUBA, POGs and Aramemnon), thus allowing to get additional information (*see* **Note 4**) about each entry. The top left of the page provides information about the AGI number and some physicochemical properties of the protein: theoretical molecular weight (MW), isoelectric point (pI), and number of residues (Length) (*see* **Note 5**). A functional annotation is available that provides information about the specific activity and the role of the protein. When available, the curated subplastidial localization of the protein is also provided (*see* **Note 6**). Publication lists are included that link to PubMed entries. Sequences of the detected peptides are provided together with the corresponding spectral counts. At the bottom of the page, are also included a reference to the experiment that allowed to detect these peptides and the subplastidial compartment in which these peptides were detected.

3. The "Home page" is also linked to a "Search page" that allows to select a set of proteins with their AGI accession numbers (Searching option) or recover protein lists according to specific properties (name, function, localization, experiment, spectral count…). From the "Search result" list, a "Display options" page allow to select features that will be included in the table, and a "Graphical view" option allows to get access to more detailed information about the selected entries (subplastidial

Uncorrected Proof

Editor's Proof

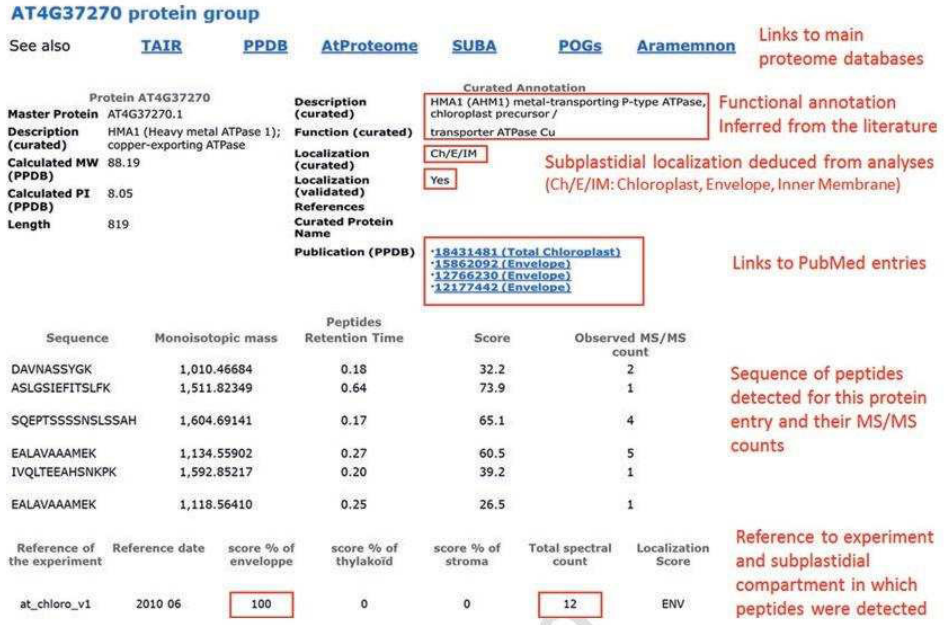


Fig. 3 A representative protein group page

Uncorrected

localization, % of occurrence in the envelope, stroma, and thylakoid, or a Volcano plot of proteins differentially distributed between grana and stroma-lamellae domains of the thylakoid membranes).

3 Methods

3.1 Search Tools

The chloroplast is commonly recognized as the place where photosynthesis occurs. However, plastids also catalyze many other essential functions, including assimilation of nitrogen and sulfur, synthesis of carbohydrates (sugars and starch), amino acids, fatty acids and specific lipids, nucleotides, carotenoids, chlorophylls, alkaloids, isoprenoids, hormone precursors, vitamins, etc. While the subplastidial localization of well-known enzymes and metabolic pathways is defined, important questions remain to be elucidated concerning the exact topology of uncharacterized plastid proteins. Knowing that there is a direct link between the subcellular localization of proteins and their respective functions, the information about the subplastidial localization of proteins opens new opportunities to probe functional hypotheses about these yet uncharacterized plastid components. On the other hand, in addition to current knowledge regarding subplastidial localization, AT_CHLORO provides curated information regarding known or

Editor's Proof

predicted functional properties of proteins. Having access to a short list of candidates, associated to a specific compartment, also provides alternative opportunities (such as biochemical characterization or analyses of the impact of gene knockout or silencing) to test for a link between one of these candidates and a missing step in a metabolic pathway or a regulatory mechanism.

3.1.1 Searching AT_CHLORO with AGI Accession Numbers

1. The link to access the search page is http://at-chloro.prabi.fr/at_chloro/searchProteinGroup/show. AT_CHLORO users can screen the whole database to retrieve a subset of proteins with their AGI numbers and identify their subplastidial localization or any other stored functional information.
2. The search can either target a single protein, or a list of accessions (*see* **Note 7**). The result of the search is a specific protein list ("Search result protein groups list") presented as a table that delivers a rapid overview of the data. Then, access to individual "Protein ID pages" (*see* Fig. 3) is granted by clicking on the AGI accessions (the leftmost column of the results table).

3.1.2 Searching AT_CHLORO with Specific Features

1. Starting from the above-cited search page, AT_CHLORO users can search for candidates with specific properties or functions in a defined subplastidial compartment. A description of the features can be found in the help page http://at-chloro.prabi.fr/at_chloro/menu/help (*see* Fig. 2).
2. In this context, users have access to a wide range of options. They can search for proteins encoded by a specific chromosome of Arabidopsis using the keyword "At1g," for example, as an "Accession" property. They can search for the keyword 'transport" in the "Curated Annotation's Description" or "Description" and get information about transporters whatever their subplastidial localization. They can look for envelope proteins using the keyword "Ch/E" within "Curated Annotation's Localization Curated," etc. More interestingly, users can combine several properties to simultaneously identify, for example, specific transporters (keyword "transport") that are associated to the chloroplast envelope (keyword "Ch/E") (*see* **Note 8**).

3.2 Display Options and Graphical View

1. The entries present in the "Protein group lists" pages issued from links present at the top of the "Home page" (All, Envelope, Stroma, …), or resulting from a search for a specific property or a list of entries, can be displayed using the "Graphical view" option. For example, starting from the protein group list "Envelope" from the "Home page," users have access to a long list of 487 proteins.
2. The "Graphical view" option allows to get a better view of these envelope components. Indeed, graphical views show that 62%

Uncorrected Proof

Editor's Proof

(303 proteins) of these envelope proteins were only detected in the envelope fraction, while 23% and 15% of these proteins were also recovered in (shared with) the stroma and thylakoid membranes, respectively. As another example, this "Graphical view" option also advises users that purified envelope membrane vesicles are more prone to contaminate stroma-lamellae vesicles (and vice versa) but that envelope proteins are poorly detected in the grana vesicles deriving from thylakoid membranes. As another example, a search option using "protease" in the "Curated Annotation's Function" finds 121 proteins.

3. A graphical view of these proteins reveals that a large portion of these proteins are shared between several plastid subcompartments.
4. Finally, at the bottom of "Protein group lists" tables, data extraction is allowed in various formats (CSV, EXCEL, ODS, PDF, and XML), thus allowing users to download the information about all proteins present in the whole database, a specific plastid compartment, or only the results of their specific quests.

3.3 Ongoing Developments

As revealed by the still increasing number of citations of the original publications [4, 15, 16], information derived from the AT_CHLORO database is of valuable interest for researchers in the field of chloroplast biogenesis and physiology.

Since the first release of the database in 2010, the content of AT_CHLORO was significantly improved. Indeed, since 2012, the database provides a graphical overview of sets of proteins recovered from specific searches. Since 2013, AT_CHLORO also stores information about calmodulin-binding proteins identified in chloroplast fractions using proteomic analyses [24]. Finally, since 2015, AT_CHLORO also provides information about specific enrichment of proteins in the main compartments (grana and stroma lamellae) of thylakoid membranes from *Arabidopsis thaliana* [16].

Altogether, the chloroplast envelope proteins represent only 1–2% of the chloroplast proteins. Within envelope membranes, a representative transporter like HMA1 for example [25], represent only 0.1% of the envelope proteins [4] and thus 1/250,000 of the plant cell proteins, thus far below the detection thresholds of earliest proteomic analyses. Consequently, due to their so low abundance, envelope proteins remained until recently the hidden part of the plastid proteome. The hunting season is still open since many known chloroplast envelope transporters deduced from our knowledge of the compartmentation of the cell metabolism remain to be identified [5]. Furthermore, proteins identified to date using subcellular proteomic approaches still represent a fraction of the complete plastid proteome. Indeed, some plastid proteins with highly particular functions are only present in specific plastid types

Uncorrected Proof

Editor's Proof

(Chapter 5) within plant tissues that are hardly accessible to subcellular fractionation approaches. On the other hand, subcellular proteomics have identified hundreds of plastid proteins whose function remains unknown.

The AT_CHLORO database thus aims at being updated with in-house data and curated information related to chloroplast proteins, with a more specific attention to envelope proteins (i.e., which makes the uniqueness of AT_CHLORO). In the near future, we plan to integrate data, from recently performed experiments, that will allow to (1) upgrade the composition of the chloroplast envelope proteome, (2) strengthen (or invalidate) the envelope localization of some proteins, and (3) provide updated and curated functional information about genuine envelope proteins, toward a comprehensive picture of the chloroplast envelope proteome.

4 Notes

1. Publication headings link to the PubMed entries of the corresponding manuscripts.
2. The first page to read, to get a global overview about information stored in the database, and tools that can be used to select specific groups of proteins (e.g., "transporters" or "proteases" using the search property "curated description" or "curated functions").
3. For example, PIP1C (AT1G01620), a Plasma membrane Intrinsic Protein or aquaporin, is listed in the envelope protein group list while most certainly localized in the plasma membrane of the cell. This information appears in column "Localization (curated)" of the Table, where PIP1C is classified as "other," meaning that is it most probably a nonplastid protein.
4. TAIR (https://www.arabidopsis.org/) is The Arabidopsis Information Resource, PPDB (http://ppdb.tc.cornell.edu) is a Plant Proteome Database, AtProteome (http://fgcz-atproteome.uzh.ch/) comprises the proteome data of the high-density, organ-specific proteome map for *Arabidopsis thaliana*, SUBA (http://suba3.plantenergy.uwa.edu.au/) is the SUBcellular localization database for Arabidopsis proteins, POGs (http://pogs.uoregon.edu/#/) is the Putative Orthologous Groups database containing genes in *Arabidopsis thaliana*, *Zea mays*, *Oryza sativa*, and/or *Populus trichocarpa*, and Aramemnon (http://aramemnon.uni-koeln.de/) is the Plant Membrane protein database,
5. Most chloroplast proteins are synthesized in the cytosol before their import into the chloroplast. During this import, most of them are cleaved with few established exceptions [3]. Since the precise cleavage site of many plastid proteins

Uncorrected Proof

Editor's Proof

remains unidentified, theoretical molecular weights (MW), isoelectric point (pI), and residue numbers (Length) are only provided for precursor forms of plastid proteins.

6. "Ch" means Chloroplast, "E" means Envelope, "Th" means Thylakoid, "S" means Stroma, "other" means other cell (non-plastid) compartment. "Ch/S" means Chloroplast Stroma. "Ch/Th" means Chloroplast Thylakoid. For Envelope localization, several options are available: "Ch/E/IM" for Inner Membrane of the Chloroplast Envelope, "Ch/E/OM" for Outer Membrane of the Chloroplast Envelope, and "Ch/E" for Chloroplast Envelope, when the accurate localization at the envelope scale is yet unidentified.
7. In this latter case, the multiple accession numbers should be written (or copied and pasted) as a column.
8. Note that the search "with specific features" option also allows to perform complex searches of specific properties, to search for an exact match (equal, not equal), a partial match (begins with, contains, ends with) to specify a portion of the term and find any words that contain that fragment, or even to search for missing data (is null).

Uncorrected Proof

Acknowledgments

D.S., S.B., L.M., I.B., M.K, C.B., and N.R acknowledge support from the ANR project ANR-15-IDEX-02. I.B. is supported by a joint PhD fellowship from the INRA Plant Biology and Breeding Division and from the Labex GRAL (ANR-10-LABX-49-01).

References

1. MacFadden GI (2014) Origin and evolution of plastids and photosynthesis in eukaryotes. Cold Spring Harb Perspect Biol 6(4):a016105
2. Martin W, Rujan T, Richly E et al (2002) Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc Natl Acad Sci U S A 99(19):12246–12251
3. Jarvis P, L pez-Juez E (2013) Biogenesis and homeostasis of chloroplasts and other plastids. Nat Rev Mol Cell Biol 14:787–802
4. Ferro M, Brugi re S, Salvi D, Seigneurin-Berny D et al (2010) AT_CHLORO, a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins. Mol Cell Proteomics 9(6):1063–1084
5. Rolland N, Curien G, Finazzi G et al (2012) The biosynthetic capacities of the plastids and integration between cytoplasmic and chloroplast processes. Annu Rev Genet 46:233–264
6. Emanuelsson O, Nielsen H, von Heijne G (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci 8(5):978–984
7. Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300:1005–1016
8. Agrawal GK, Bourguignon J, Rolland N et al (2011) Plant organelle proteomics: collaborating for optimal cell function. Mass Spectrom Rev 30(5):772–853

Editor's Proof

9. Kleffmann T, Hirsch-Hoffmann M, Gruissem W et al (2006) plprot: a comprehensive proteome database for different plastid types. Plant Cell Physiol 47(3):432–436
10. Sun Q, Zybailov B, Majeran W et al (2009) PPDB, the plant proteomics database at Cornell. Nucleic Acids Res 37(Database issue): D969–D974
11. Joshi HJ, Hirsch-Hoffmann M, Baerenfaller K et al (2011) MASCP gator: an aggregation portal for the visualization of Arabidopsis proteomics data. Plant Physiol 155(1):259–270
12. Heazlewood JL, Verboom RE, Tonti-Filippini J et al (2007) SUBA: the Arabidopsis subcellular database. Nucleic Acids Res 35(Database issue):D213–D218
13. Tanz SK, Castleden I, Hooper CM et al (2013) SUBA3: a database for integrating experimentation and prediction to define the SUBcellular location of proteins in Arabidopsis. Nucleic Acids Res 41:1185–1191
14. Hooper CM, Castleden IR, Tanz SK et al (2017) SUBA4: the interactive data analysis centre for Arabidopsis subcellular protein locations. Nucleic Acids Res 45(D1):D1064–D1074
15. Bruley C, Dupierris V, Salvi D et al (2012) AT_CHLORO: a chloroplast protein database dedicated to sub-plastidial localization. Front Plant Sci 3:205
16. Tomizioli M, Lazar C, Brugi re S et al (2014) Deciphering thylakoid sub-compartments using a mass spectrometry-based approach. Mol Cell Proteomics 13(8):2147–2167
17. Salvi D, Rolland N, Joyard J et al (2008) Purification and proteomic analysis of chloroplasts and their sub-organellar compartments. Methods Mol Biol 432:19–36
18. Seigneurin-Berny D, Salvi D, Dorne AJ et al (2008) Percoll-purified and photosynthetically active chloroplasts from *Arabidopsis thaliana* leaves. Plant Physiol Biochem 46(11):951–955
19. Salvi D, Moyet L, Seigneurin-Berny D et al (2011) Preparation of envelope membrane fractions from Arabidopsis chloroplasts for proteomic analysis and other studies. Methods Mol Biol 775:189–206
20. Moyet L, Salvi D, Tomizioli M et al (2018) Preparation of membrane fractions (envelope, thylakoids, grana and stroma lamellae) from Arabidopsis chloroplasts for quantitative proteomic investigations and other studies. Methods Mol Biol 1696:117–136
21. Joyard J, Ferro M, Masselon C et al (2009) Chloroplast proteomics and the compartmentation of plastidial isoprenoid biosynthetic pathways. Mol Plant 2:1154–1180
22. Joyard J, Ferro M, Masselon C et al (2010) Chloroplast proteomics highlights the subcellular compartmentation of lipid metabolism. Prog Lipid Res 49:128–158
23. Gloaguen P, Bournais S, Alban C et al (2017) ChloroKB: a Web application for the integration of knowledge related to chloroplast metabolic network. Plant Physiol 174(2):922–934
24. Dell'Aglio E, Giustini C, Salvi D et al (2013) Complementary biochemical approaches applied to the identification of plastidial calmodulin-binding proteins. Mol BioSyst 9(6):1234–1248
25. Seigneurin-Berny D, Gravot A, Auroy P et al (2006) HMA1, a new Cu-ATPase of the chloroplast envelope, is essential for growth under adverse light conditions. J Biol Chem 281:2882–2892

Uncorrected Proof

PUBLICATIONS

Bouchnak I., Brugière S., Moyet L., Le Gall S., Salvi D., Kuntz M., Tardif M. and Rolland N. Unravelling hidden components of the chloroplast envelope proteome: opportunities and limits of better MS sensitivity (*submitted to MCP*)

Moyet L., Salvi D., **Bouchnak I**., Miras S., Perrot L., Seigneurin-Berny D., Kuntz M. and Rolland N. Calmodulin 5, a sensor relay for environmental control of vesicular trafficking to the chloroplast (*submitted to Plant Cell)*

Bouchnak I., Moyet L., Salvi D., Kuntz M. and Rolland N. Preparation of chloroplast subcompartments from Arabidopsis for the analysis of protein localization by immunoblotting or proteomics and other studies. *J Vis Exp In Press*.

CHAPITRES D'OUVRAGE

Rolland N., **Bouchnak I**., Moyet L., Salvi D. and Kuntz M. The main functions of plastids. Methods Mol Biol, 2018. **1829**: p. 73-85.

Salvi D., Bournais S., Moyet L., **Bouchnak I**., Kuntz M., Bruley C. and Rolland N. The first step when looking for information about subplastidial localization of proteins. Methods Mol Biol, 2018. **1829**: p. 395-406.

COMMUNICATIONS ORALES

Bouchnak I., Salvi D., Moyet L., Kuntz M., Brugière S., Tardif M. and Rolland N. Chloroplast biogenesis: alternative protein targeting pathways. Séminaire jeunes chercheurs BIG/CEA, Grenoble (France), 2018, January 12

Bouchnak I., Salvi D., Moyet L., Kuntz M., Brugière S., Tardif M., Ferro M. and Rolland N. Chloroplast biogenesis: towards the role of alternative protein targeting pathways in Arabidopsis. *Chloroplast Metabolism and Photosynthesis. University of Neuchâtel*, Neuchâtel (Switzerland), 2017, June 26-28

Bouchnak I., Salvi D., Moyet L., Kuntz M., Brugière S., Tardif M., Ferro M. and Rolland N. Chloroplast biogenesis: towards the role of alternative protein targeting pathways in Arabidopsis. *9thEuropean Plant Science Retreat. INSA Toulouse*, Toulouse (France), 2017, June 20-23

Bouchnak I., Salvi D., Moyet L., Kuntz M., Brugière S., Tardif M., Ferro M. and Rolland N. Chloroplast biogenesis: towards the role of alternative protein targeting pathways in Arabidopsis. *Journées des jeunes chercheurs du département BAP de l'INRA.* , ISVV Bordeaux (France), 2017, April 20-21

POSTERS

- GRAL 48h. Autrans (France), 2018, March 26-27
- **Bouchnak I**., Salvi D., Moyet L., Kuntz M., Brugière S., Tardif M. and Rolland N. Chloroplast biogenesis: towards the role of alternative protein targeting pathways.

- Doctoral School of Chemistry and Life Sciences Annual Meeting of 2017 (JAD2017). UGA Grenoble (France)**,** 2017, June 15
- **Bouchnak I**., Salvi D., Moyet L., Kuntz M., Brugière S., Tardif M., Ferro M. and Rolland N. Chloroplast biogenesis: towards the role of alternative protein targeting pathways.

- Exploring the hidden depths of the proteome BePAc 2016. ***Ghent (Belgium)*****,** 2016, November 17-18
- **Bouchnak I**., Salvi D., Moyet L., Kuntz M., Brugière S., Tardif M., Ferro M. and Rolland N. Chloroplast biogenesis: towards the role of alternative protein targeting pathways.

- Journées des jeunes chercheurs du département BAP de l'INRA. ***ENS Lyon (France)*****,** 2016, April 11-12
- **Bouchnak I**., Salvi D., Moyet L., Kuntz M., Brugière S., Tardif M., Ferro M. and Rolland N. Chloroplast biogenesis: Role of alternative protein targeting pathways?

ENSEIGNEMENT

Label RES : 128 heures d'enseignement- DCE 2016-2018

Affectation :

Etablissement d'enseignement : Grenoble INP
Composante d'enseignement : PHELMA d'enseignement
CNU d'enseignement : 64 - Biochimie et biologie moléculaire

Année universitaire 2017-2018:

- UE Biologie (S3), Section Biomedical Engineering (BIOMED, niveau M1):
  - 8h de CM en Biologie Moléculaire
  - 4h de TD en Bioinformatique
  - 40h de TP en Biochimie et Biologie Moléculaire

Année universitaire 2016-2017:

- UE Biologie (S3), Section Biomedical Engineering (BIOMED, niveau M1):

- 6h de TD en Bioinformatique

- 62h de TP en Biochimie et Biologie Moléculaire

- UE Physiologie (S4), Section Biomedical Engineering (BIOMED, niveau M1):

- 8h de TP en Biologie Cellulaire

Toutes les disciplines ont été enseignées en anglais.

Biogenèse du chloroplaste: voies d'import alternatives

Le chloroplaste est un composant majeur de la cellule végétale. Cet organite est le fruit d'une endosymbiose, survenue entre une cellule eucaryote et une cyanobactérie. Ainsi, 95% des gènes codant pour les protéines plastidiales ont été transférés vers le génome nucléaire au cours de l'évolution. En conséquence, la plupart des protéines chloroplastiques sont aujourd'hui codées par le noyau, synthétisées dans le cytosol sous forme de précurseurs dotés d'une une extension *N*-terminale clivable (le "peptide de transit") et ensuite importées sans les chloroplastes via le système TOC/TIC (Translocons localisés au niveau des membranes externe et interne de l'enveloppe des chloroplastes). Jusqu'à récemment, toutes les protéines destinées aux compartiments chloroplastiques internes étaient censées posséder une séquence d'adressage *N*-terminale clivable et engager la machinerie d'import général TOC/TIC. Cependant, des études récentes reposant sur des approches protéomiques ont révélé l'existence de plusieurs protéines chloroplastiques dépourvues de la séquence additionnelle clivable. La première évidence de telles protéines dites non canoniques a été fournie par notre équipe, étudiant le protéome de l'enveloppe du chloroplaste d'Arabidopsis, qui a conduit à l'identification d'une protéine quinone oxidoréductase homologue nommée « ceQORH ». Bien que dépourvues de peptide de transit clivable, il s'est avéré que ces protéines sont capables de rejoindre les compartiments chloroplastiques internes. D'autre part, il a été également montré que l'import de ces protéines dans le chloroplaste n'est pas médiée par la machinerie de translocation générale TOC/TIC. De plus, il s'est avéré que ces protéines ont la particularité d'être multilocalisées dans les cellules de différents tissus de la feuille. Cependant, les mécanismes moléculaires qui contrôlent la localisation sub-cellulaire de telles protéines chloroplastiques non canoniques demeurent encore inconnus. Pour mieux caractériser fonctionnellement les composantes des systèmes d'import alternatifs de protéines chloroplastiques non canoniques, nous avons adopté une approche directe qui reposait sur des techniques biochimiques combinant le pontage chimique, la purification par affinité et la spectrométrie de masse. Cette stratégie nous a permis d'identifier un partenaire, impliqué dans le contrôle de l'adressage de la protéine ceQORH dans le chloroplaste. Alternativement, nous avons réalisé une bio-analyse du protéome de l'enveloppe du chloroplaste et qui nous a permis de revisiter la composition du protéome de l'enveloppe du chloroplaste. Afin d'expliquer la localisation sub-cellulaire variable de la protéine ceQORH, les membres de l'équipe ont émis l'hypothèse d'une interaction probable de cette protéine avec un partenaire cytosolique. Dans la dernière partie de cette étude, nous avons validé l'interaction, *in planta*, entre ceQORH et son partenaire par une approche génétique qui portait sur l'analyse de l'impact de l'absence de ce dernier sur la régulation de la localisation sub-cellulaire de la protéine ceQORH.

Mots-clés: Chloroplaste, Arabidopsis, enveloppe, protéome, alternative, adressage, import, non-canonique, protéines

Chloroplast biogenesis: Alternative protein targeting pathways

Chloroplasts are a major component of plant cells. Their origin traces back to a cyanobacterial ancestor that was engulfed by an ancient eukaryotic cell and eventually integrated as an organelle during evolution. As a result, more than 95% of the ancestral cyanobacterial genes were transferred to the host cell nucleus. Proteins encoded by these relocated genes need to return to internal chloroplast compartments. This import is mainly achieved by the general TOC/TIC machinery located at the chloroplast surface. Until recently, all proteins destined to chloroplast were believed to possess an *N*-terminal and cleavable chloroplast targeting peptide, and to engage the TOC/TIC machinery. However, recent studies have revealed the existence of several non-canonical preproteins, lacking cleavable transit peptides. The first evidence for such 'non-canonical' chloroplast proteins was provided by our team studying the Arabidopsis chloroplast envelope proteome, leading to the identification of a quinone oxidoreductase homologue termed « ceQORH ». Furthermore, a few such proteins were demonstrated to use alternative targeting pathways, independent of the TOC/TIC machinery. To better characterize components of such alternative targeting machineries, a targeted study combining affinity purification and mass spectrometry aiming to identify alternative receptors at the chloroplast surface has been performed. This study allowed us to identify new "partner" involved in the control of chloroplast targeting of ceQORH protein. Alternatively, we also revisited the chloroplast envelope proteome composition and initiated a gene candidate approach. In addition, some non-canonical proteins are shared by plastids and other cell compartments. However, molecular mechanisms controlling subcellular localization of these non-canonical plastid proteins remain unknown. In order to explain the variable subcellular localization of ceQORH protein, our team hypothesized a probable interaction of ceQORH with a cytosolic partner. In the last part of this study, we validated the interaction between ceQORH and its partner *in planta* by a genetic approach analyzing the impact of the absence of the cytosolic partner on the regulation of the sub-cellular localization of ceQORH protein.

Keywords: Chloroplast, Arabidopsis, envelope, proteome, alternative, targeting, pathways, non-canonical, proteins

Laboratoire de Physiologie Cellulaire & Végétale

BIG/CEA-Grenoble

17, rue des martyrs, 38054 Grenoble Cedex 9