

**PROTEOLYTIC MATURATION AND PROTEIN DEGRADATION IN
ARABIDOPSIS THALIANA CHLOROPLASTS**

A Thesis

Presented to the Faculty of the Graduate School
of Cornell University

In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Elden E. Rowland

December 2017

© 2017 Elden E. Rowland

**PROTEOLYTIC MATURATION AND PROTEIN DEGRADATION IN
ARABIDOPSIS THALIANA CHLOROPLASTS**

Elden E. Rowland, PhD

Cornell University 2017

Proteolysis is crucial for the maturation, regulation and recycling of the chloroplast proteome. Although several dozen chloroplast proteases are known, information concerning their substrates and functions is limited. In particular, little is known about the structural features of substrates that trigger their proteolysis. Most chloroplast proteins are nuclear encoded and are targeted through an N-terminal chloroplast transit peptide (cTP) that is removed by stromal processing peptidase (SPP). To better understand proteolytic maturation, the soluble N-terminal proteome of the *Arabidopsis thaliana* chloroplast was characterized. A cTP cleavage motif was observed that suggests other peptidases, in addition to SPP, are involved in chloroplast protein maturation. There was a clear preference for small uncharged amino acids at the processed protein N-terminus suggesting the existence of a chloroplast specific 'N-end rule'. The soluble chloroplast peptidases PREP and OOP have been shown to degrade small polypeptides *in vitro* and are thought to be responsible for removal of cTP fragments and other degradation products. The CLP protease system can degrade intact protein substrates with the aid of ATP dependent (AAA+) CLPC chaperones that unfold and feed substrates into the CLP proteolytic core. An array of proteomic tools were used to compare *Arabidopsis* mutants deficient in the above peptidases with wild type.

Degradation products, including cTPs, were found to accumulate in peptidase mutants indicative, of rate-limited or blocked degradation pathways. Incomplete or altered N-terminal maturation for chloroplast proteins was dependent on the type and severity of the peptidase deficiency. These results provide molecular details to help explain dwarf, chlorotic mutant phenotypes and demonstrate the interplay between protein import, proteolytic processing and the downstream degradation of damaged or unwanted proteins in the chloroplast. Substrate and sequence cleavage specificity was determined for soluble chloroplast glutamyl-endoribonuclease (CGEP) and the plastoglobule localized metalloprotease PGM48. Structural models were used to predict peptidase substrate binding mechanisms.

BIOGRAPHICAL SKETCH

Born in England, Elden Rowland was raised in Bonavista, Newfoundland, Canada. He received a B.Sc. (Hons) in Chemistry from Dalhousie University in 2002. He managed the Proteomics Core Facility at Dalhousie's School of Medicine for eight years. During that time, he was a member of the Atlantic Research Center for Lipid Research and developed a widening interest biology. In 2011 he joined the lab of Klaas van Wijk at Cornell University, first as a technician, and then as a graduate student in the field of Plant Biology.

The force that through the green fuse drives the flower
Drives my green age; that blasts the roots of trees
Is my destroyer.
And I am dumb to tell the crooked rose
My youth is bent by the same wintery fever.

The force that drives the water through the rocks
Drives my red blood; that dries the mouthing streams
Turns mine to wax.
And I am dumb to mouth unto my veins
How at the mountain spring the same mouth sucks.

Dylan Thomas (1952)

ACKNOWLEDGMENTS

My deepest gratitude to Klaas van Wijk for the opportunity to study biology in his laboratory and for his unwavering support throughout.

Extra special thanks to Maureen Hanson, Tom Owens and Michelle Heck for serving on my committee and for providing insightful discussions and encouragement.

Thank you to all the van Wijk lab members past and present for your generosity, patience and for rich discussions. I would especially like to thank Jitae Kim and Nazmul Bhuiyan, for being inspirational mentors.

Much gratitude to the Department of Plant Biology for welcoming me to the plant world. I especially appreciate everyone's dedication to teaching. To my fellow graduate students, you are a constant source of inspiration and the best of friends to relax with.

Lots of love to Liz, Steve and Betony and my dearest love Laura.

This work was supported by grants from the National Science Foundation awarded to Klaas van Wijk and by Department of Plant Biology teaching assistantships.

TABLE OF CONTENTS

	Page
Biographical Sketch	iii
Dedication	iv
Acknowledgements	v
Table of Contents	vi
List of Figures	vii
List of Tables	xi
Preface	xii
CHAPTER ONE. Introduction	1
CHAPTER TWO. The Arabidopsis chloroplast stromal N-terminome: Complexities of amino-terminal protein maturation and stability	48
CHAPTER THREE. Genetic and functional interactions of the organellar peptidases CLP, PREP and OOP in <i>Arabidopsis thaliana</i>	104
CHAPTER FOUR. The plastoglobule-localized metallopeptidase PGM48 is a positive regulator of senescence in <i>Arabidopsis thaliana</i>	217
APPENDIX. Structural and functional characterization of stromal chloroplast glutamyl-endopeptidase (CGEP)	272

LIST OF FIGURES

CHAPTER ONE	page
Figure 1.1. Model of proteolytic activity and peptidases in chloroplasts.	04
Figure 1.2. N-terminal proteomic methodologies.	11
Figure 1.3. M16 peptidase phylogeny.	21
Figure 1.4. N-End rule, hierarchy, adaptors and proteases.	30
 CHAPTER TWO	
Figure 2.1. Conceptual illustration of Nt maturation of nuclear- and p-encoded proteins.	51
Figure 2.2. Nt amino acid frequency for stromal-exposed n-encoded chloroplast proteins.	59
Figure 2.3. Analysis of amino acid conservation around experimentally determined Nti for n-encoded stromal-exposed proteins and comparison to Nti generated by <i>in vitro</i> SPP cleavage assays reported in the literature.	62
Figure 2.4. Nt amino acid frequency for stromal-exposed p-encoded proteins and comparison to all known luminal-exposed Nti (both p- and c-encoded).	69
Figure 2.5. Working model for Nt maturation of n-encoded proteins and the classification of different types of Nti.	77
Supplemental Figure 2.1. TAILS experimental workflow, using a method adapted from Kleifeld et al., 2011.	100
Supplemental Figure 2.2. RBCS-4 Nt peptide MS spectrum labeled light and heavy to distinguish physiological mono-methylation status.	102
Supplemental Figure 2.3. TAILS identified Nti of intact stromal proteins and their degradation products.	103

CHAPTER THREE

Figure 3.1. Genetic interaction between PREP and OOP peptidases.	112
Figure 3.2. Quantitative comparison of total leaf proteomes of proteomics. <i>wt, prep1 prep2, oop</i> and the triple mutant.	118
Figure 3.3. Pathways to proteolytic protein maturation and degradation.	127
Figure 3.4. Comparison of <i>wt</i> and <i>prep1 prep2</i> peptidome.	138
Figure 3.5. Genetic interaction between the CLP core subunits and PREP peptidases.	145
Figure 3.6. Genetic interaction between the CLPT, CLPC and PREP peptidases.	150
Figure 3.7. Comparative quantitative leaf proteomics of <i>wt</i> and <i>prep1 prep2 clpt1 clpt2</i> .	155
Figure 3.8. Icelogo plots for differentially regulated Nt processing events in the <i>clpt1 clpt2</i> mutant.	168
Figure 3.9. Genetic interaction between the CLPR2 and OOP.	170
Figure 3.10. Model for proteolytic maturation and degradation.	176
Supplemental Figure 3.1. Characterization of OOP and PREP mutants.	199
Supplemental Figure 3.2. Characterization of <i>prep1 prep2 oop</i> at later stages of development; seed phenotype and germination rate.	200
Supplemental Figure 3.3. Quantitative comparison of chloroplast proteomes of proteomics <i>wt</i> and <i>prep1 prep2</i> .	202
Supplemental Figure 3.4. Quantitative comparison of total leaf proteomes of proteomics <i>wt, prep1 prep2, oop</i> and the triple mutant.	204
Supplemental Figure 3.5. Terminal amine isotopic labeling of substrates (TAILS) workflow for comparative Nt proteomics.	207
Supplemental Figure 3.6. Gel based N-terminal proteomics for <i>prep1 prep2 oop</i> stroma. Verification of relative Nt peptide abundance for select proteins.	209
Supplemental Figure 3.7. Peptidome experimental workflow.	211
Supplemental Figure 3.8. The functional interaction between PREP1,2	212

and CLPT.

CHAPTER FOUR

Figure 4.1. PG-M48 is localized in PG and has Zn-metallo-dependent endopeptidase activity cleaving upstream of hydrophobic residues.	223
Figure 4.2. M48 proteins are conserved in non-photosynthetic and photosynthetic organisms.	226
Figure 4.3. 3D protein structural model for PGM48 and predicted interface with the PG monolayer.	230
Figure 4.4. Overexpression and suppression of PGM48 in Arabidopsis and effect on mRNA levels during plant development and senescence.	234
Figure 4.5. Overexpression and suppression of PGM48 in Arabidopsis shows that PGM48 protein accelerates natural leaf senescence.	236
Figure 4.6. PG proteome composition in wt plants during natural leaf senescence and mRNA levels for PG genes	240
Figure 4.7. PG proteome composition in wt, OE and RNAi plants during natural leaf senescence determined by MS/MS-based label-free spectral counting of isolated PG.	243
Figure 4.8. Interaction of PGM48 with other PG proteins.	247
Figure 4.9. Functional model for PGM48 function in senescing chloroplasts and phenotype of the <i>CCD4</i> null mutant during dark-induced senescence.	253

APPENDIX

Figure 5.1. Proteolytic activity of recombinant CGEP.	278
Figure 5.2. N- and C-terminal processing of CGEP.	279
Figure 5.3. CGEP 3D structural model generated with iTASSER.	281
Figure 5.4. Icelogo plots for CGEP specific cleavage of peptide libraries; PICS Experiment 1	286

Figure 5.5. Icelogo plots for CGEP specific cleavage of peptide libraries; PICS Experiment 2.	289
Figure 5.6. Peptide library length distribution before and after reaction with CGEP in PICS experiments.	290
Figure 5.7. Proteins cleaved preferentially by CGEP in PICS type experiment with intact protein standards.	291
Figure 5.8. CGEP tDNA insertion lines and sub-cellular localization.	293
Figure 5.9. N-terminal peptides from CP12 identified in TAILS experiment comparing wt and <i>cgep-5</i> chloroplast stroma.	296

LIST OF TABLES

CHAPTER ONE	page
Supplemental Table 1.1. Proteins and gene identifiers used to construct M16 protease phylogeny	43
CHAPTER TWO	
Table 2.1. Examples of experimentally determined N-terminal peptides	57
CHAPTER THREE	
Table 3.1. Differentially regulated protein Nti identified in <i>prep1 prep2</i> and <i>prep1 prep2 oop</i> TAILS experiments	122
Table 3.2. Differentially regulated protein N-termini identified in <i>prep1 prep2 oop</i> low molecular weight TAILS experiments	132
Table 3.3. cTPs detected in peptidome	141
Table 3.4. Differentially accumulating chloroplast and mitochondrial proteins in <i>prep1 prep2 clpt1 clpt2</i> and <i>clpt1 clpt2</i> as compared to wt	159
Table 3.5. Differentially regulated protein N-termini identified in <i>clpt1 clpt2</i> and <i>prep1 prep2 clpt1 clpt2</i> TAILS experiments	163
Supplemental Table 3.2. Comparative total leaf proteome analysis of wt, <i>prep1 prep2</i> , <i>oop</i> and <i>prep1 prep2 oop</i> .	214
Supplemental Table 3.3. List of TAILS and peptidome experiments conducted.	215
Supplemental Table 3.6. Mitochondrial proteins processed by ICP55.	216

PREFACE

Like most cellular compartments, the chloroplast has a discrete, but dynamic, collection of proteins (referred to as the proteome) that are required for biological function. The chloroplast proteome is continually shaped by proteolysis. This thesis covers the life-cycle of chloroplast proteins from protein translocation and maturation, to their ultimate degradation into free amino acids. In order to study these processes I chose the model plant *Arabidopsis thaliana* due to its high quality genome sequence, its well annotated gene models, and the availability of extensive genetic resources. Proteomics and mass spectrometry techniques were used to track N-terminal protein maturation, identify protease substrates, as well as protein degradation products.

The INTRODUCTION (Chapter 1) explains the importance of proteolysis in the chloroplast and summarizes the relevant literature. Chapter 2 describes our study of the chloroplast N-terminal proteome, using a specialized labeling technique combined with mass spectrometry, which serves as a baseline for the rest of the thesis. This 2nd chapter characterizes the native population of mature N-terminal proteins sequences following import and processing of transit peptides and facilitates discussion of proteolytic degradation in later chapters.

The study described in Chapter 3 utilizes similar techniques as described in Chapter 2 to study various plastid proteases mutants and the consequences for the chloroplast proteome. This research revealed the connectivity of the plastid protease system or network. Chapter 4 describes the characterization of the specialized plastoglobular protease PGM48 that is senescence-induced and appears to play a role in degradation of the plastoglobular protein CAROTENOID CLEAVAGE DIOXYGENASE. The thesis also contains an appendix regarding the structure and

function of the soluble stromal protease cGEP both *in vitro* and *in vivo*. I suggest how this protease might fit into the above mentioned protease network.

CHAPTER ONE

INTRODUCTION

1.1 THE CHLOROPLAST

Chloroplasts are one of the defining features of plants and are the site of photosynthesis. They are part of a family of organelles called plastids that play a variety of roles in different tissues and at various developmental stages (Jarvis and Lopez-Juez, 2013). All plastids are derived from proplastids that are present in the shoot and root apical meristems. They cannot be synthesized *de novo* and therefore must be inherited from 'mother' cells via cell division. Proplastids may differentiate into amyloplasts in roots, etioplasts in the dark, chloroplasts in green tissue and chromoplasts in flowers and fruit (Liebers et al., 2017). Interestingly, most plastids can interconvert if the conditions are right. In addition to performing photosynthesis, plastids synthesize an array of essential metabolites such as purines, fatty acids, vitamins E and K, carotenoids, amino acids and hormones (Neuhaus and Emes, 2000; Rolland et al., 2012).

Chloroplasts are the result of an endosymbiotic relationship between a cyanobacterium and a eukaryotic heterotroph (Archibald, 2015). Throughout ~ one billion years of evolution most of the plastid progenitor's genes were lost or migrated to the host nucleus via a process termed horizontal gene transfer (Shinozaki et al., 1986; Zimorski et al., 2014). Less than 100 genes remain in the plastid genome but they play central roles in the light and dark reactions of photosynthesis, in transcription, proteolysis, and perhaps import (Nakai, 2015). Signaling between the nucleus and the

plastid has been demonstrated but the identities and integration of these signals is not well understood (Jarvis and Lopez-Juez, 2013).

The majority of the estimated ~3000 plastid proteins are encoded by nuclear genes and must be imported after their translation in the cytosol (van Wijk and Baginsky, 2011). This requires a robust system for protein trafficking and import to the plastid. The nature of transit peptides (TPs) that direct proteins to organelles continues to be an active area of research. Some 100 proteins are dually targeted to plastids and mitochondria and many function in organelle gene expression (e.g. tRNA-synthetases) and a few are proteases (Carrie and Small, 2013). The plant cell relies on the structural and electrostatic nature of TPs to direct them to the desired organelle (Garg and Gould, 2016).

1.2 PROTEOLYSIS AND PROTEOSTASIS

A highly dynamic proteome ensures that chloroplast biogenesis and homeostasis is maintained throughout development and in the face of (a)biotic stresses (Jarvis and Lopez-Juez, 2013). Protein homeostasis (proteostasis) refers to maintenance of a functional proteome (Hartl et al., 2011). More specifically, an optimum population of properly localized, folded and active proteins must be maintained. This requires control of transcription and translation, as well as a network of chaperones and other quality control systems including proteolysis (van Wijk, 2015). The chloroplast contains many different peptidases and together they help ensure proteostasis, by processing incoming immature proteins, and removing damaged or otherwise unwanted proteins, cleaved cTPs, and other protein fragments (Figure 1.1). Complete degradation of proteins yields

amino acids that can be used for protein synthesis or directed towards other metabolic activities (Rolland et al., 2012).

Proteolysis describes the enzymatic degradation of polypeptides - one protein degrades another. The central question in protease research is what structural features cause a protein substrate to be degraded. In other words, what prevents functional proteins from being degraded and allows unwanted proteins or protein fragments to be positively selected for degradation? Indeed, cellular peptidase systems incorporate many layers of complexity to regulate proteolysis (Sauer et al., 2004). The chloroplast protease network has many similarities to bacterial systems owing to its cyanobacterial progenitor. However, these proteases have evolved complexities that are absent in bacteria and harbor many unique features as compared to other organelles such as mitochondria (Nishimura and van Wijk, 2015).

1.2.1 Scope of this Thesis

A general objective of this work was to improve our understanding of proteolysis in the chloroplast. As will be demonstrated in the remainder of this Chapter, plastid proteases can be loosely broken down into two groups, namely those involved in maturation (trimming a protein to its final functional form) and those involved in degradation (deconstruction of proteins). These functional groups are discussed in detail below. Additionally, I discuss the M16 peptidase family (including extra-plastidic members) in *Arabidopsis thaliana* that includes peptidases involved in both maturation

and degradation. This thesis also provides many examples of the strengths and limitations of different experimental and *in silico* peptidase characterization strategies.

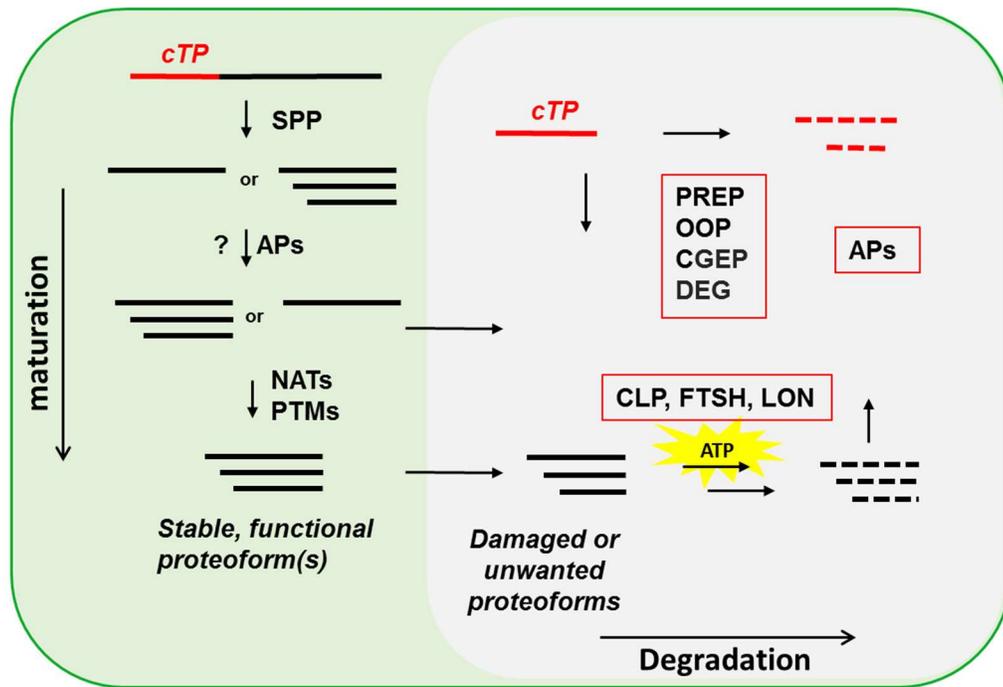


Figure 1.1. Model of proteolytic activity in the chloroplasts. Maturation and degradation pathways are shown. cTP, chloroplast transit peptide; SPP, stromal processing peptidase; APs, amino peptidases involved in degradation and potentially involved in processing of immature proteoforms; NATs, alpha-amino acetyltransferases; PTMs, post translational modifications; PREP, presequence protease; OOP, organellar oligopeptidase; CGEP, chloroplast glutamyl endopeptidase; DEG, CLP, FTSH and LON are the common peptidase names.

The focus in Chapter 2 is proteolytic maturation and characterization of the *Arabidopsis thaliana* (Col-0) chloroplast N-terminal (Nt) proteome is described. This project revealed the mature Nt start sites for several hundred chloroplast proteins and evaluated possible scenarios to arrive at these N-termini. In Chapter 3, the focus shifts

to degradation, but we discover that disruption of degradation machinery interferes with maturation; therefore these two processes are linked. The remainder of this thesis (Chapters 4 and the Appendix) describes the characterization of other chloroplast proteases unrelated to those discussed in the opening chapters. The long-term goal of this research is to define the chloroplast peptidase network and establish its contribution to plant growth and development.

1.3 THECHNOLOGY TO STUDY PROTEOLYSIS

Although chloroplast proteolysis has been intensively studied over the past several decades, many questions remain. Substrate-peptidase relationships are notoriously difficult to identify and elucidating the physiological significance of peptidases is challenging. Here I discuss various approaches that have been used to decipher protease structure and function.

1.3.1 Genetics approaches

Genetics is an invaluable tool and many important peptidases have been discovered by mutational analysis. Using a forward genetics strategies many cellular processes have been linked to proteases in mitochondria and chloroplasts (Rigas et al., 2009; Albrecht-Borth et al., 2013; Hong et al., 2016). Ling *et al* were searching for suppressors of outer chloroplast membrane TOC33 loss-of-function mutants and discovered an E3 ligase, SP1 that targets components of the chloroplast import apparatus (Ling et al., 2012). A number of chloroplast peptidases were identified in a

high-throughput screen for nuclear-encoded chloroplast targeted genes with potential roles in photosynthesis and amino acid metabolism (Savage et al., 2013).

By crossing different loss-of-function peptidase mutants in *Arabidopsis* we can determine if these genes act in similar pathways or processes (Chapter 3; (Kmiec et al., 2013)). As the chloroplast protease field has developed, a more complete list peptidases and their assisting chaperones and adaptor proteins has been established (van Wijk, 2015). Using this list of known and putative chloroplast peptidases a reverse genetics approach can be used to better resolve the functional roles and interconnectedness of peptidases in the chloroplast.

1.3.2 In vitro strategies

Once a peptidase has been identified, a series of logical questions arise and many of these can be answered using biochemical techniques *in vitro*. Is the protein a functional protease? Does it require cofactors, metals or ATP to operate? What is the protein structure? What is the cleavage specificity and substrate selectivity? What are the kinetics of proteolysis for a given peptidase?

If the peptidase can be expressed in *E. coli* (or other expression systems) and reasonable amounts of an active, recombinant protein can be purified, characterization can be relatively straightforward. If the peptidase is unstable or part of larger heteromeric multi-subunit complex, such as the plastid CLP peptidase system in higher plants, *in vitro* production of the enzyme is far more challenging. For simple peptidase monomers like presequence protease (PREP) and organellar oligopeptidase (OOP), studied extensively by the Glaser group (Teixeira and Glaser, 2013), or CGEP (Chapter

4) and PGM48 (Chapter 5), *in vitro* experiments can provide a wealth of knowledge. Determining a protein structure requires extensive skills and resources and is key to fully understanding how a peptidase interacts with its substrates. However, if the structure of a related peptidase is known, protein homology modeling can provide a rough idea of how the protease might interact with substrates and its surroundings (see Chapters 4 and 5 for examples).

Once purified, a peptidase can be incubated with synthetic peptides, recombinant proteins or cell extracts (Agard and Wells, 2009). The products of proteolysis can then be resolved by gel electrophoresis (SDS-PAGE), liquid chromatography (LC) and/or mass spectrometry (MS). If the substrate mixture is relatively simple, peptide products may be directly analyzed by MS and the identified peptides reveal the cleavage preference or specificity of the peptidase (as shown in Chapter 5). Alternatively, a substrate can be incubated with total or fractionated cell extracts containing known or unidentified peptidases (Abad et al., 1991; Teixeira et al., 2017). From these experiments, the necessary cofactors and energy requirements for optimal peptidase activity can be determined as well.

For a higher resolution picture of sequence cleavage specificity, the Proteomic Identification of Cleavage Sites (PICS) procedure was developed (Schilling et al., 2011; Biniossek et al., 2016). This N-terminal (Nt) proteomics technique involves incubation of a peptidase with a peptide library so that statistically significant cleavage site specificity can be determined. In a recent study, human cell lysates were incubated with either the peptidase CASPASE 2 or 6 before enrichment of cleaved proteins (Julien et al., 2016). In addition to the identification of many new candidate substrates, samples

were taken at different time points and analyzed by targeted MS (SRM) and immunoblotting, allowing the authors to determine the caspase specific rates of degradation for hundreds of substrates. This study highlights the fact that a peptidase may have multiple substrates, but some will be degraded much more efficiently than others.

Although these experiments are essential to fully characterize a peptidase, they often leave us in the dark as to the function of the protein in living cells. Therefore *in vivo* experiments that draw from the above genetic and *in vitro* results are key.

1.3.3 In vivo approaches

The essential function of a peptidase is to cleave single or multiple peptide bonds within a substrate. However, the hunt for the substrates often ends empty-handed or with inconclusive results. There are a variety of reasons for this but new approaches and technologies are continually developed to improved substrate identification.

Relative protein abundance levels can be determined using mass spectrometry (MS) analysis. Label-free quantitative MS is a straightforward method to determine the relative abundances of thousands of proteins in different genetic backgrounds (Friso et al., 2011). The van Wijk lab has characterized many *Arabidopsis thaliana* peptidase mutants with this technique, revealing dramatic remodeling of the chloroplast proteome and identifying putative substrates (Kim et al., 2013; Kim et al., 2015).

Since peptidases are involved in protein turnover, researchers have long sought to quantify protein turnover or lifetime. One strategy involves adding protein synthesis inhibitors to the tissue and then using immunoblotting to measure reduction in protein

abundance at different time points (Sjogren and Clarke, 2011). A valid criticism of this approach is that the translational inhibitors themselves drastically effect the proteome. Pulse-chase-type experiments involving stable isotope labeling in cell culture (SILAC) have been successfully applied to measure protein turnover in animals (Dhondt et al., 2017), but this approach does not work well in plants because they can synthesize their own amino acids. Recently, stable isotope labeling of plants was achieved by hydroponic growth on an N15 enriched nitrogen source. MS analysis of plants harvested at different time points following isotopic labeling allowed the determination of half-lives for over a thousand proteins (Li and Millar, 2017). Calculation of these rates was based on the assumption that protein synthesis rates are constant and a correction was made to account for new growth/biomass.

Untargeted or ‘shotgun’ proteomics techniques effectively sample abundant proteins. However, shorter lived, low abundance protein and peptide fragments are often below the limit of detection (Fortelny et al., 2015; Julien and Wells, 2017). Furthermore, these techniques cannot typically distinguish between different post-translationally modified proteoforms. Targeted proteomics surpass the standard limits of detection and quantification by scanning only over a small mass range specific to the protein or peptide of interest (Bereman et al., 2012), and this technique has been used to monitor putative peptidase substrates in different genetic backgrounds (Majovsky et al., 2014; Julien et al., 2016).

Quantitative *degradomics* methods including amino (N-) and carboxy (C-) terminal proteomics facilitate the detection of protease substrates *in vivo* (Demir et al., 2017). N-terminal (Nt) proteomics allows determination of protease cleavage sites (and

substrates) by identifying new (sometimes called *neo*-) N-termini generated by proteolysis (Agard and Wells, 2009; Lange and Overall, 2013). These methods have also been used to study alternative start sites for non-organellar proteins (Van Damme et al., 2014). This approach really began with the advent of Edman sequencing (Edman, 1950). For many years this was the sole method for determination of a protein's sequence, starting at the N-terminus. Advancements in MS eventually allowed the sequencing of peptides by tandem-MS and database searching against sequenced genomes. Due to poor ionization efficiency and limited MS based fragmentation, proteins are generally digested with a protease such as trypsin, to produce small peptides before MS analysis (Domon and Aebersold, 2006).

So-called 'top down' MS approaches where whole proteins without prior enzymatic digestion are analyzed by MS directly are gaining popularity (Fornelli et al., 2017). This approach has the potential to sample populations of proteins and their degradation fragments (Wildburger et al., 2017). However, separating intact (whole) proteins in solution remains challenging and MS detectors have a limited dynamic range that may preclude detection of less abundant proteolytic fragments.

The Terminal Amine Isotopic Labeling of Substrates (TAILS) approach, outlined in Figure 1.2, allows purification of protein N-termini. The method takes advantage of the high reactivity of protein α -amines (primary amines) at protein N-termini and on the side chains of Lys (Kleifeld et al., 2011). As with other N-terminomics techniques, internal peptides produced by digestion with trypsin are separated from N-terminal peptides using a separate labeling step, before and after

digestion. The TAILS approach is compatible with various quantitative labeling reagents such as ITRAQ, and SILAC (Prudova et al., 2014).

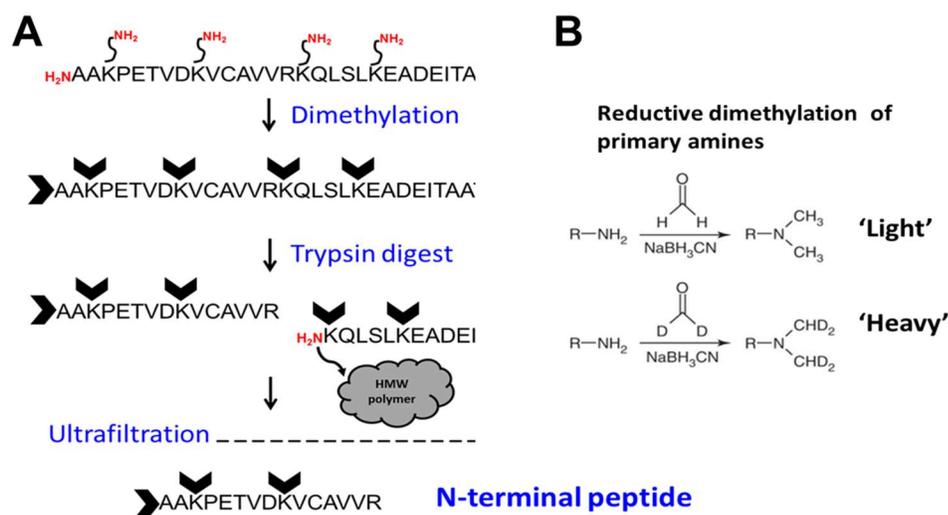


Figure 1.2. N-terminal proteomic methodologies. (A) Terminal amine isotopic labeling of substrates (TAILS) schematic. (B) Reductive dimethylation reaction for labeling the primary amines of proteins in solution at neutral pH.

Despite these technological advances, there are relatively few examples of these strategies revealing *bona fide* protease substrates in complex samples (*in vivo*) (Tsiatsiani et al., 2013; Carrie et al., 2015). In both these referenced cases, the specificity and localization of the peptidases was already known, greatly narrowing the proteome pool of potential substrates. N-terminal proteomics was applied to elucidate chloroplast N-terminal maturation and degradation in Chapters 2, 3 and in the Appendix.

1.4 PROTEOLYTIC MATURATION IN CHLOROPLASTS

1.4.1 *Protein import*

Most chloroplast (and mitochondrial) proteins are synthesized in the cytosol and must therefore be imported. The plastid import apparatus has been extensively studied (Demarsy et al., 2014; Nakai, 2015; Paila et al., 2015). Receptors on the outer membrane recognize proteins to be imported by a 20 to 100 residue N-terminal signal peptide called the chloroplast transit peptide (cTP). Import then proceeds in a GTP- and ATP-dependent manner. Most nuclear encoded proteins enter via the TOC/TIC channel, a pore complex that consists of at least a dozen protein subunits, many of which have one or more paralogs (Shi and Theg, 2013). The TOC/TIC channels within the envelope are not heterogeneous and there is evidence for at least two distinct complexes that preferentially import different classes of proteins. This is thought to facilitate the import of essential but low abundance ‘housekeeping’ proteins through one channel, while highly abundant photosynthetic proteins translocate through an alternate channel (Inoue et al., 2010). Furthermore, subunits of one such complex were shown to be the target of an E3 ligase (SP1), which triggers their degradation via the 26S proteasome, thus influencing protein import efficiency and selectivity (Ling et al., 2012).

1.4.2 *Stromal processing peptidase and cTP cleavage*

It is well established that the cTPs are cleaved by STROMAL PROCESSING PEPTIDASE (SPP) (Richter et al., 2005; Teixeira and Glaser, 2013). The cleaved cTPs are then rapidly degraded by SPP itself, and a combination of PREP, OOP,

aminopeptidases, and possibly others (Kmiec et al., 2014; Teixeira et al., 2017). See sections 1.5 - 1.6.

Complete loss of the SPP protein causes embryo lethality at the 16-cell stage, demonstrating that Nt maturation is crucial for plant development and cell viability (Trosch and Jarvis, 2011). The exact mechanism for this phenotype is unknown. However, loss-of function alleles for many plastid-localized proteins prevent development past the embryo stage (Savage et al., 2013). Most components of the import machinery in the inner envelope are nuclear encoded and must therefore be processed by SPP; unprocessed proteins likely have reduced function. Indeed, many of the plastid protein import mutants have similar phenotypes (Paila et al., 2014). Another factor to consider is that overall plastid protein stability will be hindered in SPP mutants, due to the long unstructured cTP extension that may trigger various degradation pathways (Bruce, 2000; Kmiec et al., 2014).

The ability to predict the N-terminus of a chloroplast protein is limited due to a lack of experimental information. The cleavage site specificity of recombinant SPP from pea has been probed with just a handful of substrates from a variety of species (Richter et al., 2005). Using this data and a machine learning approach, the ChloroP algorithm was developed (Emanuelsson, 1999). As the authors point out, the program is designed to predict the SPP cleavage site rather than the ultimate protein N-terminus. The software-derived consensus cleavage motif was VR↓AAAV. Implementing this motif produces an N-terminus one or two residues upstream of the observed mature protein N-terminus for most Arabidopsis chloroplast proteins as determined in Chapter 2 and in previous reports (Zybailov et al., 2008; Bienvenut et al., 2012). The presence

of numerous amino peptidases in the plastid, and the possibility that they contribute to N-terminal maturation has long been recognized, but no concrete evidence exists for their contribution to plastid protein maturation (Richter and Lamppa, 1998; Emanuelsson, 1999; Walling, 2006). Although the prospect of additional processing by a so far unidentified aminopeptidase is intriguing, it is possible that SPP does not so much recognize a sequence motif, but rather a structural shift from an extended beta strand to a more helical structure, as was pointed out in Gavel *et al* (1990). The authors note that the C-termini of the cTP are enriched in Ile, Val (both beta strand forming) and Arg, but lack Leu and Lys (strong alpha helix favoring) (Pace and Scholtz, 1998). Then the cTP leads into Ala/Cys/Met followed by Ala/Ser, having alpha helix promoting character (Von Heijne and Gavel, 1990; Pace and Scholtz, 1998). Therefore, SPP may cleave the cTP without a strictly defined sequence motif. We explore these possibilities as well as the necessity for chaperones to ensure SPP precision in chapter 3. Interestingly, Rudhe *et al* show that SPP is slightly more tolerant of mutations near the cleavage site of substrates than its mitochondrial counterpart β -MPP (Rudhe et al., 2004).

SPP is a soluble stromal protein and does not form a stable interaction with the chloroplast import apparatus (Schnell and Blobel, 1993; Schnell, 2014). Temporary existence of unprocessed pre-proteins in the stroma has been reported (Inoue et al., 2013). However, the general lack of observed cTPs in chloroplasts (both from *in vitro* chloroplast protein import experiments and analysis of extracted leaf chloroplast proteomes) indicates that cTP cleavage and degradation must be highly efficient.

1.4.3 Co-translational maturation of plastid encoded proteins

N-terminal methionine excision (NME) removes the first amino acid from newly synthesized proteins (Giglione et al., 2004). This reaction occurs co-translationally and is often coordinated with N- α acetylation (NAA) of the revealed amino acid while still attached to the ribosome (Starheim et al., 2012). NME is an essential process and the NME machinery is part of the minimal eubacterial genome (Hutchison et al., 1999). Indeed complete loss of cytosolic NME causes embryo lethality in Arabidopsis (Ross et al., 2005; Frottin et al., 2009). Both NME and NAA also occur for nuclear encoded, plastid targeted proteins in the cytosol (Zybailov et al., 2008; Bischof et al., 2011). Tight interactions with cytosolic chaperones usually ensure translocation of the intact protein to the TOC import receptors (Fellerer et al., 2011). However, if the import machinery is compromised, as is the case in the *ppil* TOC159 loss-of-function allele, then unprocessed NAA transit peptides accumulate in the cytosol (Bischof et al., 2011).

Plastid and bacterial encoded proteins are initiated with a formylated methionine, produced by a specialized formyl-Met-tRNA. Before NME can occur, the enzyme peptide deformylase must deformylate the N-terminal Met (there are two paralogues in Arabidopsis, PDF1A and PDF1B). These events and other cTP cleavages dictate the mature N-terminal sequence of plastid encoded proteins. Interestingly, not all proteins are deformylated and NME is limited by the penultimate amino acid. Generally NME only occurs for proteins with Ala, Ser, Val, Thr, Pro, Gly or Cys in the penultimate position. These small uncharged residues are thought to facilitate Met

cleavage; the same pattern is observed across the tree of life (Bonissone et al., 2013). Two notable exceptions detailed in Chapter 2 are 30S ribosomal protein S15 and ATP-synthase subunit CF β 1 for which the mature proteins are Ile and Arg terminated immediately downstream of the N-terminal initiating Met, respectively.

If plastid NME efficiency is reduced, as in the *pdf1b* loss-of function allele in Arabidopsis, dwarf, albino plants result (Giglione, 2003) again illustrating the importance of Nt processing. Protein destabilization is thought to be responsible for the phenotype and lower accumulation of the Photosystem II subunits D1 and D2 was observed in *pdf1b*. Furthermore, D2 turnover was shown in the green alga *Chlamydomonas reinhardtii* to be increased in response to treatment with the PDF inhibitor actinonin and also by mutating the penultimate residue to block removal of methionine (Giglione, 2003). An overall increase in proteolytic activity is suggested in response to NME inhibition, with CLP and FTSH proteases being involved (Adam et al., 2011).

1.4.4 Additional mitochondrial processing peptidases ICP55 and OCT1

The yeast mitochondrial Nt proteome was recently shown to be regulated by the peptidases ICP55 and OCT1, which begs the question, does such processing also occur in plastids? Vogtle *et al* noticed that many proteins accumulate with multiple Nt protein sequences that differ by a one or two amino acids; a similar phenomenon was observed in chloroplasts (see Charters 2 and 3). Systematic knock-out of yeast mitochondrial proteases led to the identification of intermediate cleavage peptidase (ICP55), which cleaves a single Leu, Phe or Tyr from the N-terminus of proteins following processing

by mitochondrial processing peptidase MPP (Vogtle et al., 2009). The processing peptidase OCT1 is able to cleave eight amino acids from the N-terminus of select mitochondrial proteins, again following MPP action (Vogtle et al., 2011). OCT1 also recognizes N-terminal sequences with Leu and Phe at their N-terminus but not Tyr. In most cases, both ICP55 and OCT1 generate N-termini with 'stabilizing' (small and uncharged) N-terminal residues. The authors went on to show that in the absence of processing by these enzymes, protein stability was greatly reduced *in organello* and *in vivo* (Vogtle et al., 2009; Vogtle et al., 2011). Mossmann *et al* demonstrate that loss of MPP β (MAS1) and PREP (CYM1) in yeast results in hindered or blocked processing of mitochondrial proteins (Mossmann et al., 2014). They observed widespread maturation defects when PREP was absent, such as unprocessed mitochondrial pre-proteins and incorrect N-terminal processing. We reveal similar feedback inhibition of SPP processing in the chloroplasts of Arabidopsis mutants deficient in PREP and OOP (Chapter 3).

1.5 M16 PEPTIDASES; PROTEIN STRUCTURES AND EVOLUTIONARY RELATIONSHIPS

Members of the M16 family of metallopeptidases are widely distributed across the tree of life and they have diverse functions and an interesting evolutionary history (Rawlings et al., 2012). They function in organelle protein processing and maturation (as is the case for SPP and MPP) and also in peptide degradation. This feature makes this family particularly interesting for important for organelle biology. Because this

thesis concerns several *Arabidopsis* M16 peptidases, their structural and evolutionary relationships are discussed here.

All M16s have an ‘inverted’ zinc binding motif (HxxEH) unique to peptidases that is required for their proteolytic activity (Aleshin et al., 2009). Typically, four structurally similar domains form two clam shell-like halves that encapsulate peptide substrates (Johnson et al., 2006; King et al., 2014). There are three M16 subfamilies (A, B and C): *E. coli* pitrilysin (PTR), human insulin degrading enzyme (IDE) and yeast STE23 fall into subfamily A (Alper et al., 2009; Taskin et al., 2017), whereas the eukaryotic MPP/SPP and PREP fall into subfamilies B and C respectively.

Crystal structures are available for members of each subfamily including: human IDE (Protein Data Bank code, 2G56) and *E. coli* pitrilysin PTR (1Q2L), yeast MPP (1HR6) as well as *Arabidopsis* PREP1 (2FGE), and human PREP (4L3T). Additionally, a structure was determined for *Bacillus halodurans* M16B categorized peptidase named BHP (3HDI) (Aleshin et al., 2009). For each peptidase, at least one crystal structure was determined with a bound substrate or inhibitor molecule, allowing detailed characterization of the substrate binding pocket.

E. coli PTR is an M16A family member localized to the periplasm. It was shown *in vitro* to cleave a variety of small peptides including insulin (Anastasi et al., 1993). PTR has 28% sequence identity to human IDE, the protease primarily responsible for insulin degradation in the blood (Rawlings et al., 2012).

Yeast MPP is a dimer made up of α - and β -subunits. Two structures were determined for the heterodimer, each with different mitochondrial transit peptides (mTPs) bound to the substrate cavity in an extended conformation (Taylor et al., 2001).

The normally unstructured mTPs were shown to hydrogen bond to β -strands in the proximity of the active site, with the N-terminus protruding into the cavity and the active site cleaving the substrate close to the C-terminus of the mTP, where the mature protein would normally start. There is no structure available for SPP but the existence of 4 domains found also in PREP and α/β -MPP suggests it forms a similar structure to MPP, perhaps with a larger cavity to accommodate lengthy cTPs. Interestingly SPP itself has an extremely long 142 residue cTP (Richter et al., 2005), which begs the question, can it cleave its own cTP or must it rely on an already mature and processed SPP?

MPP has a significant sequence similarity with the core 1 and 2 subunits of the mitochondrial cytochrome *bc1* complex (also named Complex III) (Gakh et al., 2002); in plants both MPP alpha and beta are integrated in the cytochrome *bc1* complex (Glaser and Dessi, 1999). It has been suggested that an M16-like protease was introduced into eukaryotes from the mitochondrial progenitor that eventually became associated with the *bc1* complex. In the case of some fungi such as *N. crassa*, a duplication allowed one of the subunits to become an autonomous soluble protease subunit while the other remained partially integrated into the *bc1* complex. In animals and yeast, two duplications of the proteases allowed the creation of two separate complexes, the non-proteolytic, membrane bound core 1 and 2 subunits (*bc1* complex) and a separate soluble α/β -MPP. Plants, for some reason, never separated these functions and MPP remains critical for both respiration and precursor processing.

The structures of Arabidopsis PREP1 at 2.1Å (Johnson et al., 2006) and human PREP at 2.0 Å (King et al., 2014) suggest that PREP acts as a monomer with a dynamically formed proteolytic chamber (10000-13000 Å³), opening and closing

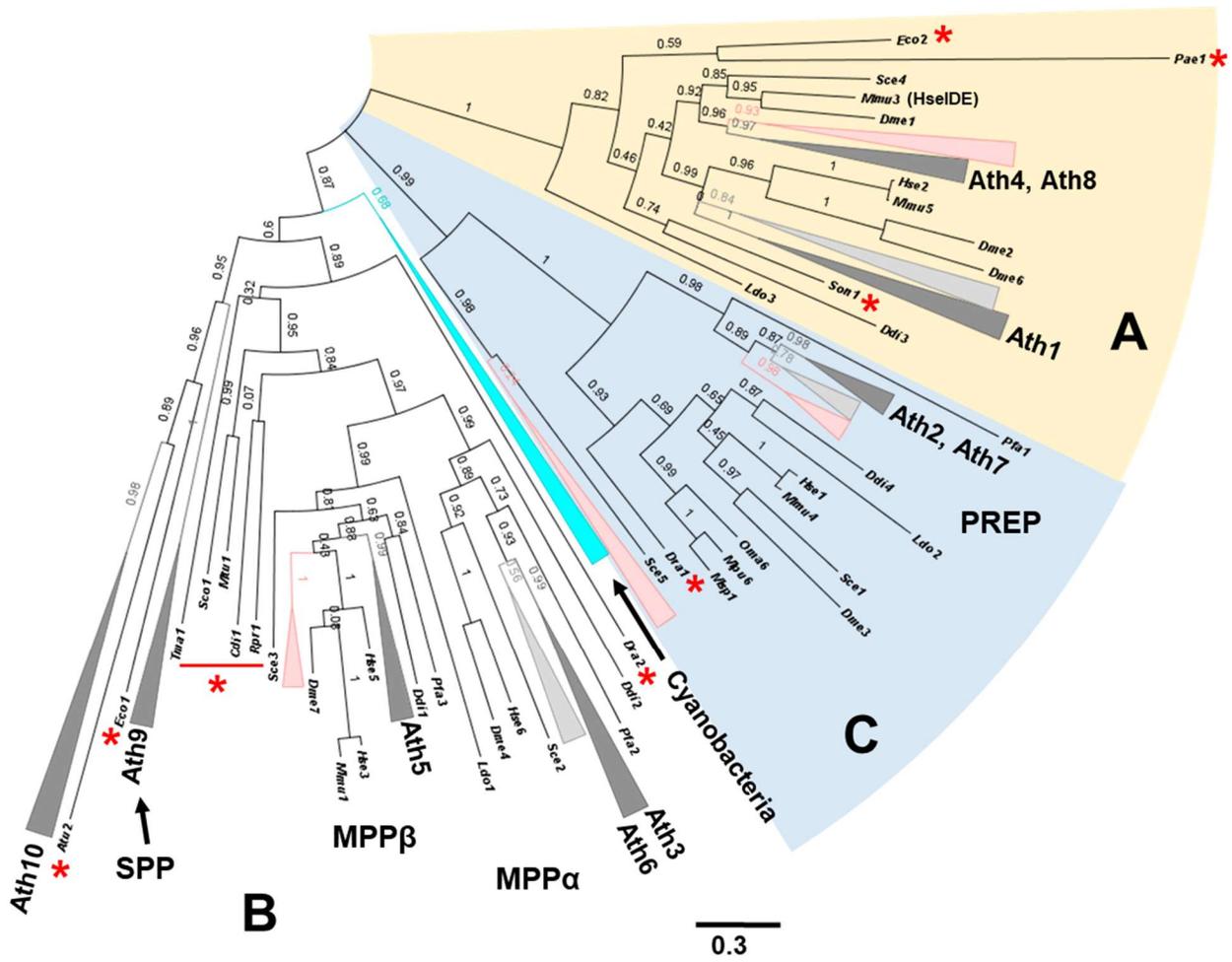
through a hinge mechanism. *In vitro* experiments indicated that the *Arabidopsis* PREP1,2 homologs can degrade peptides/proteins within a size range of 10 and 65 amino acids (Stahl et al., 2005), consistent with the size of the proteolytic chamber. They are suggested to have a preference for positively charged peptides including cTPs and mTPs due to the partially negatively charged binding pocket (Glaser and Alikhani, 2010; King et al., 2014). In addition to the canonical M16 active site residues (HXXEH), other conserved sites are required for PREP activity such as E94, E177, N109, and C-terminal R848 and Y854. The existence of catalytically important residues (not part of the M16 domain) in SPP may explain why a C-terminal truncated form of SPP is inactive *in vitro* (Richter et al., 2005; Johnson et al., 2006).

Several phylogenetic studies have been reported for M16 peptidases but they focused only on a single subfamily, or were limited in scope (Gakh et al., 2002; Stahl et al., 2005). I identified ten M16 peptidases in *Arabidopsis thaliana* containing the characteristic zinc binding motif. In order to better resolve their evolutionary origins, and to suggest functions of so far uncharacterized members, I generated a phylogenetic tree (Figure 1.3). Three major clades were observed corresponding to the A, B, C subfamilies. The *Arabidopsis* proteins fell into seven subclades across these three subfamilies. Two *Arabidopsis* M16s (phylogeny identifiers Ath4 and Ath8) are positioned in a clade with only plant and algal (green and red) proteins that is sister to a clade containing human IDE (clustered with Mmu3) and yeast STE23 (Sce4).

Figure 1.3. M16 peptidase phylogeny. Plant and algal clades are collapsed separately, except in cases where the algal sequences did not form a monophyletic clade and they were collapsed together with plants; plants, dark grey; green algae, light grey; red algae, pink. Prokaryotic proteins are marked with a red asterisk. A list of proteins and gene identifiers can be found in the APPENDIX of this chapter.

A: Ath1 -AT1G06900; Ath4 - AT2G41790; Ath8 - AT3G57470. **B:** Ath2 – PREP2 (AT1G49630); Ath7 – PREP1 (AT3G19170). **C:** Ath 3 - MPP-alpha; Ath5 - MPP-beta (AT3G02090); Ath6 - MPP-alpha (AT3G16480); Ath9 – SPP (AT5G42390); Ath10 - AT5G56730

BLAST searches were conducted for each Arabidopsis protein against plant sequences (in Phytozome), and the Landmark database (at NCBI) which includes sequences from 27 genomes spanning a wide taxonomic range. The compiled sequences were clustered to remove redundancies and aligned using MUSCLE or MAFFT and then trimmed with Trimal (Capella-Gutiérrez 2009). Trees were generated using the FastTree, Approximately-Maximum-Likelihood method (Price et al., 2010).



This clade is sister to one that includes AT1G06900 (Ath1) and human Nardilysin (Hse2). Two clades containing bacterial proteins are sister those above, including *E. coli* pitrilysin (Eco2), which is likely the ancestral member of the M16A peptidases. Ath1, 4 and 8 have no known functions but probably degrade small peptides based on their similarity to other characterized M16s, most likely in the cytosol since they lack predicted signal or transit peptides.

The PREP proteins form the second subfamily (M16C) and are primarily eukaryotic with the exception of a bacterial protein (*Deinococcus radiodurans*). All PREP proteins studied to date are targeted to the mitochondria, and also to plastids in photosynthetic organisms; hence this subfamily appears to be an organelle-specific invention. The *Plasmodium falciparum* (malaria parasite, species belonging the Apicomplexa) protein facilyisin (PFA1) is sister to the major green plant PREP clade and was shown to be targeted to the apicoplast (a non-photosynthetic plastid) with only minor amounts detected in mitochondria (Eggleston et al., 1999; Ponpuak et al., 2007).

The third subfamily (M16B) includes α -MPP (Ath3 and Ath6) and β -MPP (Ath5), each forming distinct subclades that are sister to a bacterial clade that includes *Rickettsia* (Rpr1), a suggested descendent of the mitochondria progenitor, and other bacteria. The *Bacillus halodurans* peptidase BHP mentioned above is a close homologue of the *Rickettsia* protein. Although there were reports of such proteins being able to act as monomers (Dabonne et al., 2005; Kitada et al., 2007), Aleshin *et al* showed that it forms a dimer, like α/β -MPP. However, the closed conformation bound to a peptide substrate suggests that functionally these MPP bacterial ancestors behave more like M16As and M16Cs (Aleshin et al., 2009). MPP (and SPP) are critical components

of their respective organelle proteomes (Yaffe and Schatz, 1984; Trosch and Jarvis, 2011) and is therefore likely that these peptidases played a key role in the persistence of the organelles after endosymbiosis.

Sister to MPP is the SPP containing clade that is made up of two subclades. One contains the confirmed Arabidopsis SPP (Ath9) and its plant algal homologues. The other clade contains the uncharacterized AT5G56730 (Ath10) and two bacterial proteins from *E. coli* (Eco1) and *Agrobacterium tumefaciens* (Atu2), perhaps representing ancient relatives. Surprisingly, all the cyanobacterial proteins form a separate clade that is equally distant from MPP and SPP clades. Homologues of SPP were previously reported in cyanobacteria (Richter et al., 2005); however, they are more similar to MPP and therefore there are no clear SPP homologues in cyanobacteria.

AT5G56730 (Ath10) has 25 % sequence identity with SPP over its N-terminal region but little similarity over the 600 C-terminal residues. Prediction of subcellular localization with TargetP (Emanuelsson et al., 2007) gives cTP and mTP scores of 0.03 and 0.316 respectively and we have detected its un-cleaved cTP by MS, suggesting that it is unlikely plastidic. However the protein was primarily detected in plastid samples in PPDB and it is part of the plastid reference genome (Huang et al., 2013).

1.6 PROTEOLYTIC DEGRADATION

So far we have discussed peptidases involved in protein processing and maturation, but the majority of chloroplast peptidases are likely involved in the degradation of whole proteins, protein aggregates and fragments. Therefore, they are said to function in proteome maintenance and quality control (Jarvis and Lopez-Juez,

2013; Nishimura et al., 2017). We still know very little about what triggers degradation of a substrate by these proteolytic machines. In the cytosol, ubiquitination is a major trigger for degradation by the 26S proteasome (Amm et al., 2014). Although this modification is absent in plastids, pre-proteins in the cytosol, destined for the plastid, have been shown to be ubiquitinated and degraded perhaps preventing the import apparatus from being overloaded (Lee et al., 2009).

In absence of ubiquitination, bacterial and organellar protease substrates must be marked for degradation by other post-translational modifications (PTM) and structural determinants (Baker and Sauer, 2006; Trentini et al., 2016). Caseinolytic peptidase or CLP is named after its ability to degrade the milk protein casein (Hwang et al., 1988; Squires and Squires, 1992). Casein is highly phosphorylated and intrinsically disordered in solution allowing it to be easily degraded (Naqvi et al., 2016). Even robust enzymes like trypsin cannot readily degrade globular proteins and require denaturation of substrates for complete degradation (Switzar et al., 2013). In order to degrade globular proteins *in vivo*, ATP dependent chaperones are required to unfold the substrate and present it to the peptidase (see section 1.6.2).

1.6.1 ATP independent peptidases of the chloroplast

Small or unstructured polypeptides do not require unfolding for them to be degraded and there are many peptidases that can potentially fulfill this role. It was recently demonstrated that in addition to PREP peptidase (section 1.4 -1.5), OOP and various aminopeptidases are required to completely degrade synthetic cTPs *in vitro* (Teixeira et al., 2017). OOP can degrade peptides in the 8 – 23 residue range in

agreement with the size of the substrate binding cavity based on the crystal structure (Kmicc et al., 2014). At least three metallo-aminopeptidases (M1 and M17 family) cooperate with PREP and OOP to generate free amino acids (Teixeira et al., 2017).

There are at least six DEG proteases that are localized to the plastid, either residing in the stroma or in the thylakoid lumen. Interestingly, their oligomeric state and peptidase activity is influenced by pH which fluctuates in the stroma (between 7 and 8) and lumen (between 3.5 and 7) depending on photosynthetic electron transport rate and chloroplast ATP synthesis (Kley et al., 2011; Schuhmann et al., 2012).

Chloroplast glutamyl endopeptidase (CGEP) was discovered in soluble protein extracts of leaves of pea and cucumber, and a homolog was identified in Arabidopsis. CGEP was partially purified from pea extracts and shown to cleave the recombinant N-terminal domain of LHCII at C-terminal of glutamate (Forsberg et al., 2005). We characterize and discuss Arabidopsis CGEP in Chapter 4.

Plastoglobuli are lipid monolayers particles that form from the outer membrane leaflet of thylakoid membranes during various abiotic stresses and during senescence (van Wijk and Kessler, 2017). A metallopeptidase unique to plants was identified in purified plastoglobuli (Bhuiyan and van Wijk, 2017); PGM48 is characterized and its *in vivo* function described in Chapter 5.

1.6.2 ATP dependent peptidases of the chloroplast

The largest group of plastid peptidases are the ATP-dependent proteases, FTSH, LON and CLP. Including the CLP chaperones and adaptors, there are ~20 proteins contributing to their combined functions in chloroplasts. There are a number recently

published and detailed reviews (Nishimura et al., 2015; van Wijk, 2015; Nishimura et al., 2017) on these peptidases, so I discuss them here only briefly.

Each of these protein degradation machines form oligomeric complexes with enclosed proteolytic cores. The LON protein subunits have both peptidase and chaperone domains, whereas CLP has a dedicated peptidase core formed by several different proteins, and separate associated chaperones and adaptor proteins. Both CLP and LON are soluble protein complexes localized to the stroma; FTSH is membrane bound and has either one or two transmembrane domains that determine the orientation of the peptidase domain. There are dedicated thylakoid FTSH proteins and others that are localized to the chloroplast envelope (van Wijk, 2015).

Our understanding of substrate recognition by ATP dependent peptidases in organelles is rapidly evolving. Perhaps the most well-known example in the chloroplast is the degradation of the photo-damaged D1 component of photosystem II by FTSH (and also by DEG) as part of the PSII repair cycle (Nath et al., 2013). Damaged, misfolded or aggregating proteins are generally thought to be targets for all of these ATP dependent systems (Baker and Sauer, 2006) and the chaperone components are thought to regulate (equilibrate and balance) the proteome to prevent excessive amounts of non-functional proteins, either through degradation, or re-folding and repair (Hartl et al., 2011). A recent study of CLP in *Bacillus Subtilis* revealed CLPC (the primary chaperone for the CLP system) dependent degradation of arginine phosphorylated substrates (Trentini et al., 2016). The kinase MCSB was previously shown to be required for degradation of various substrates in *B. Subtilis*, but this is the first time such a

phosphorylation-dependent mechanism for substrate recognition and delivery has been shown.

The chlorophyll biosynthetic enzyme glutamyl-tRNA reductase (GluTR) was shown to interact with the CLP adaptor proteins CLPS and CLPF in *Arabidopsis* (Nishimura et al., 2015). Degradation of GluTR was shown to be CLPS dependent in the night but not in the day (Apitz et al., 2016). The Nt domain of GluTR appears to be critical for this regulation and this may be one of first examples of an N-degron mediated degradation in chloroplasts (see section 1.6.3). CLPC was shown to interact with both the import apparatus via TIC110 and the CLP core (Flores-Perez et al., 2016). This suggests that CLP may also act as chloroplast gatekeeper of sorts, providing quality control as new proteins arrive and are processed.

1.6.3 N-end rule and its potential role in chloroplast substrate recognition

A major mechanism for the regulation of protein lifetime is the *N-end* rule, first described in 1989 by Varshavsky *et al* based on their observation that the half-life of a β -galactosidase fusion protein in yeast was greatly influenced by the N-terminal amino acid (Gonda et al., 1989). The *N-end* rule has been shown to be a key determinant of protein stability from eubacteria to eukaryotes (Dougan et al., 2012) and may help explain increased turnover of proteins in chloroplasts in response to developmental transitions, stress and incomplete protein maturation (Nishimura et al., 2013). The destabilizing N-terminal component that triggers degradation is referred to as an N-degron (Figure 1.4). There are primary N-degrons (Arg, Lys, His, Leu, Phe, Trp, and Tyr) that trigger degradation of substrates by the ClpAP protease system in bacteria

(Humbard et al., 2013) or by i-AAA and m-AAA proteases in yeast and mammalian mitochondria (Glynn, 2017). Secondary N-degrons (Asp, Glu) require the post-translational N-terminal addition of another primary N-degron amino acid, or N-alpha acetylation to be triggered. Lastly there are tertiary N-degrons (Cys, Asn, Gln) that must undergo some post translational modification before the secondary pathway above can be initiated - reviewed in (Graciet et al., 2010).

Most examples of N-end rule pathways in plants are extra-plastidic and are linked to ubiquitylation and the proteasome (Graciet et al., 2010). An exciting demonstration of the plant N-end rule is the sensing of molecular oxygen and the response to hypoxia (low oxygen). Multiple members of the group VII ERF transcription factors have been shown to activate hypoxia-response genes. The N-terminal cysteine of these proteins acts as a tertiary N-degron and under normal atmospheric conditions the cysteine is enzymatically oxidized by specific plant cysteine oxidases and the protein degraded. Upon hypoxia, (*e.g.* flooding) these proteins are no longer oxidized and are able to accumulate, generating the hypoxic response (Dissmeyer et al., 2017; Dong et al., 2017; White et al., 2017).

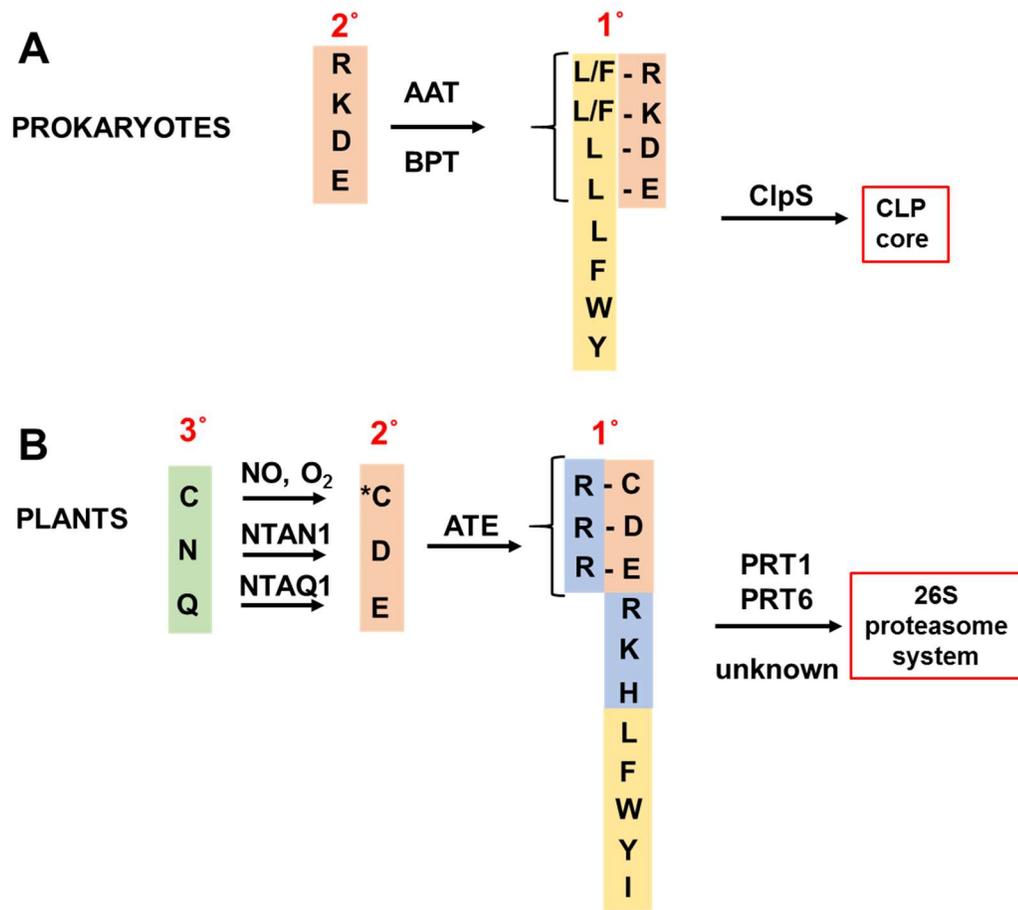


Figure 1.4. N-End rule, hierarchy, adaptors and proteases. (A) Prokaryotic system where certain protein N-terminally exposed amino acids act as primary N-degrons (leu, Phe, Trp or Tyr), that are recognized by ClpS and targeted to CLP protease for degradation. Secondary N-degrons are modified by the ligation of primary degrons by the aminotransferases AAT and BPT. (B) In plants, tertiary N-degrons are modified by oxidation of Cys, or deamidation of Gln and Asn to become secondary N-degrons that are modified by Arg-transferases (ATEs) to form primary N-degrons. These N-degrons are in turn ubiquitinated by the E3 ligases PRT1 and PRT6 targeting them to the proteasome. After Graciet and Wellmer (2010) Trends in Plant Science.

The nature of a plastid N-end rule remains unclear although a homolog of the bacterial N-end rule adaptor protein ClpS is present in plants; this ClpS homolog was named CLPS1 (Nishimura et al., 2013). Measuring the impact (stabilizing character) of the ultimate N-terminal residue has proven to be a complicated task. Apel and Bock *et al* generated transplastomic tobacco lines expressing a variable N-terminus-GFP fusion construct and then monitored protein half-life (Apel et al., 2010). A list of N-termini ranked based on their stability was generated for the GFP fusions in both *E. coli* and tobacco. However, incomplete NME in the case of certain N-termini (Glu and maybe Val) complicates these results. We cannot be sure whether these residues are actually stabilizing, or if the N-terminal methionine protects any potential N-degron; perhaps most importantly, the results of this publication mostly contradicts bacterial-like N-degrons (Dougan et al., 2012). Interestingly, I identified two high confidence cysteine N-terminated proteins in Arabidopsis that appear to accumulate under normal growth conditions (see Chapter 2).

LITERATURE CITED

- Abad MS, Oblong JE, Lamppa GK** (1991) Soluble Chloroplast Enzyme Cleaves preLHCP Made in *Escherichia coli* to a Mature Form Lacking a Basic N-Terminal Domain. *Plant Physiol* **96**: 1220-1227
- Adam Z, Frottin F, Espagne C, Meinnel T, Giglione C** (2011) Interplay between N-terminal methionine excision and FtsH protease is essential for normal chloroplast development and function in *Arabidopsis*. *Plant Cell* **23**: 3745-3760
- Agard NJ, Wells JA** (2009) Methods for the proteomic identification of protease substrates. *Curr Opin Chem Biol* **13**: 503-509
- Albrecht-Borth V, Kauss D, Fan D, Hu Y, Collinge D, Marri S, Liebers M, Apel K, Pfannschmidt T, Chow WS, Pogson BJ** (2013) A novel proteinase, SNOWY COTYLEDON4, is required for photosynthetic acclimation to higher light intensities in *Arabidopsis*. *Plant Physiol* **163**: 732-745
- Aleshin AE, Gramatikova S, Hura GL, Bobkov A, Strongin AY, Stec B, Tainer JA, Liddington RC, Smith JW** (2009) Crystal and solution structures of a prokaryotic M16B peptidase: an open and shut case. *Structure* **17**: 1465-1475
- Alper BJ, Rowse JW, Schmidt WK** (2009) Yeast Ste23p shares functional similarities with mammalian insulin-degrading enzymes. *Yeast* **26**: 595-610
- Amm I, Sommer T, Wolf DH** (2014) Protein quality control and elimination of protein waste: the role of the ubiquitin-proteasome system. *Biochim Biophys Acta* **1843**: 182-196
- Anastasi A, Knight CG, Barrett AJ** (1993) Characterization of the bacterial metalloendopeptidase pitrilysin by use of a continuous fluorescence assay. *Biochem J* **290 (Pt 2)**: 601-607
- Apel W, Schulze WX, Bock R** (2010) Identification of protein stability determinants in chloroplasts. *Plant J* **63**: 636-650
- Apitz J, Nishimura K, Schmied J, Wolf A, Hedtke B, van Wijk KJ, Grimm B** (2016) Posttranslational control of ALA synthesis includes GluTR degradation by Clp protease and stabilization by GluTR-binding protein. *Plant Physiol*
- Archibald JM** (2015) Endosymbiosis and Eukaryotic Cell Evolution. *Curr Biol* **25**: R911-921
- Baker TA, Sauer RT** (2006) ATP-dependent proteases of bacteria: recognition logic and operating principles. *Trends Biochem Sci* **31**: 647-653

- Bereman MS, MacLean B, Tomazela DM, Liebler DC, MacCoss MJ** (2012) The development of selected reaction monitoring methods for targeted proteomics via empirical refinement. *Proteomics* **12**: 1134-1141
- Bhuiyan NH, van Wijk KJ** (2017) Functions and substrates of plastoglobule-localized metallopeptidase PGM48. *Plant Signal Behav* **12**: e1331197
- Bienvenut WV, Sumpton D, Martinez A, Lilla S, Espagne C, Meinel T, Giglione C** (2012) Comparative large scale characterization of plant versus mammal proteins reveals similar and idiosyncratic N-alpha-acetylation features. *Mol Cell Proteomics* **11**: M111015131
- Binossek ML, Niemer M, Maksimchuk K, Mayer B, Fuchs J, Huesgen PF, McCafferty DG, Turk B, Fritz G, Mayer J, Haecker G, Mach L, Schilling O** (2016) Identification of Protease Specificity by Combining Proteome-Derived Peptide Libraries and Quantitative Proteomics. *Mol Cell Proteomics* **15**: 2515-2524
- Bischof S, Baerenfaller K, Wildhaber T, Troesch R, Vidi PA, Roschitzki B, Hirsch-Hoffmann M, Hennig L, Kessler F, Gruissem W, Baginsky S** (2011) Plastid proteome assembly without Toc159: photosynthetic protein import and accumulation of N-acetylated plastid precursor proteins. *Plant Cell* **23**: 3911-3928
- Bonissone S, Gupta N, Romine M, Bradshaw RA, Pevzner PA** (2013) N-terminal protein processing: a comparative proteogenomic analysis. *Mol Cell Proteomics* **12**: 14-28
- Bruce BD** (2000) Chloroplast transit peptides: structure, function and evolution. *Trends Cell Biol* **10**: 440-447
- Carrie C, Small I** (2013) A reevaluation of dual-targeting of proteins to mitochondria and chloroplasts. *Biochim Biophys Acta* **1833**: 253-259
- Carrie C, Venne AS, Zahedi RP, Soll J** (2015) Identification of cleavage sites and substrate proteins for two mitochondrial intermediate peptidases in *Arabidopsis thaliana*. *J Exp Bot* **66**: 2691-2708
- Dabonne S, Moallic C, Sine JP, Niamke S, Dion M, Colas B** (2005) Cloning, expression and characterization of a 46.5-kDa metallopeptidase from *Bacillus halodurans* H4 sharing properties with the pitrilysin family. *Biochim Biophys Acta* **1725**: 136-143
- Demarsy E, Lakshmanan AM, Kessler F** (2014) Border control: selectivity of chloroplast protein import and regulation at the TOC-complex. *Front Plant Sci* **5**: 483
- Demir F, Niedermaier S, Villamor JG, Huesgen PF** (2017) Quantitative proteomics in plant protease substrate identification. *New Phytol*
- Dhondt I, Petyuk VA, Bauer S, Brewer HM, Smith RD, Depuydt G, Braeckman BP** (2017) Changes of protein turnover in aging *Caenorhabditis elegans*. *Mol Cell Proteomics*

- Dissmeyer N, Rivas S, Graciet E** (2017) Life and death of proteins after protease cleavage: protein degradation by the N-end rule pathway. *New Phytol*
- Domon B, Aebersold R** (2006) Mass spectrometry and protein analysis. *Science* **312**: 212-217
- Dong H, Dumenil J, Lu FH, Na L, Vanhaeren H, Naumann C, Klecker M, Prior R, Smith C, McKenzie N, Saalbach G, Chen L, Xia T, Gonzalez N, Seguela M, Inze D, Dissmeyer N, Li Y, Bevan MW** (2017) Ubiquitylation activates a peptidase that promotes cleavage and destabilization of its activating E3 ligases and diverse growth regulatory proteins to limit cell proliferation in Arabidopsis. *Genes Dev* **31**: 197-208
- Dougan DA, Micevski D, Truscott KN** (2012) The N-end rule pathway: from recognition by N-recognins, to destruction by AAA+proteases. *Biochim Biophys Acta* **1823**: 83-91
- Edman P** (1950) Method for Determination of the Amino Acid Sequence in Peptides. *Acta Chemica Scandinavica* **4**: 283-293
- Eggleston KK, Duffin KL, Goldberg DE** (1999) Identification and characterization of falcilysin, a metallopeptidase involved in hemoglobin catabolism within the malaria parasite *Plasmodium falciparum*. *J Biol Chem* **274**: 32411-32417
- Emanuelsson O** (1999) ChloroP a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Science* **8**: 978-984
- Emanuelsson O, Brunak S, von Heijne G, Nielsen H** (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* **2**: 953-971
- Fellerer C, Schweiger R, Schongruber K, Soll J, Schwenkert S** (2011) Cytosolic HSP90 cochaperones HOP and FKBP interact with freshly synthesized chloroplast preproteins of Arabidopsis. *Mol Plant* **4**: 1133-1145
- Flores-Perez U, Bedard J, Tanabe N, Lymperopoulos P, Clarke AK, Jarvis P** (2016) Functional Analysis of the Hsp93/ClpC Chaperone at the Chloroplast Envelope. *Plant Physiol* **170**: 147-162
- Fornelli L, Durbin KR, Fellers RT, Early BP, Greer JB, LeDuc RD, Compton PD, Kelleher NL** (2017) Advancing Top-down Analysis of the Human Proteome Using a Benchtop Quadrupole-Orbitrap Mass Spectrometer. *J Proteome Res* **16**: 609-618
- Forsberg J, Strom J, Kieselbach T, Larsson H, Alexciev K, Engstrom A, Akerlund H-E** (2005) Protease activities in the chloroplast capable of cleaving an LHCII N-terminal peptide. *Physiologia Plantarum* **123**: 21-29
- Fortelny N, Pavlidis P, Overall CM** (2015) The path of no return--Truncated protein N-termini and current ignorance of their genesis. *Proteomics* **15**: 2547-2552

- Friso G, Olinares PD, van Wijk KJ** (2011) The workflow for quantitative proteome analysis of chloroplast development and differentiation, chloroplast mutants, and protein interactions by spectral counting. *Methods Mol Biol* **775**: 265-282
- Frottin F, Espagne C, Traverso JA, Mauve C, Valot B, Lelarge-Trouverie C, Zivy M, Noctor G, Meinnel T, Giglione C** (2009) Cotranslational proteolysis dominates glutathione homeostasis to support proper growth and development. *Plant Cell* **21**: 3296-3314
- Gakh O, Cavadini P, Isaya G** (2002) Mitochondrial processing peptidases. *Biochim Biophys Acta* **1592**: 63-77
- Garg SG, Gould SB** (2016) The Role of Charge in Protein Targeting Evolution. *Trends Cell Biol* **26**: 894-905
- Giglione C** (2003) Control of protein life-span by N-terminal methionine excision. *EMBO Life Sci* **61**: 1455-1474
- Giglione C, Boularot A, Meinnel T** (2004) Protein N-terminal methionine excision. *Cell Mol Life Sci* **61**: 1455-1474
- Glaser E, Alikhani N** (2010) The organellar peptidasome, PreP: a journey from Arabidopsis to Alzheimer's disease. *Biochim Biophys Acta* **1797**: 1076-1080
- Glaser E, Dessi P** (1999) Integration of the mitochondrial-processing peptidase into the cytochrome bc1 complex in plants. *J Bioenerg Biomembr* **31**: 259-274
- Glynn SE** (2017) Multifunctional Mitochondrial AAA Proteases. *Front Mol Biosci* **4**: 34
- Gonda DK, Bachmair A, Wunning I, Tobias JW, Lane WS, Varshavsky A** (1989) Universality and structure of the N-end rule. *J Biol Chem* **264**: 16700-16712
- Graciet E, Mesiti F, Wellmer F** (2010) Structure and evolutionary conservation of the plant N-end rule pathway. *Plant J* **61**: 741-751
- Hartl FU, Bracher A, Hayer-Hartl M** (2011) Molecular chaperones in protein folding and proteostasis. *Nature* **475**: 324-332
- Hong L, Dumond M, Tsugawa S, Sapala A, Routier-Kierzkowska AL, Zhou Y, Chen C, Kiss A, Zhu M, Hamant O, Smith RS, Komatsuzaki T, Li CB, Boudaoud A, Roeder AH** (2016) Variable Cell Growth Yields Reproducible OrganDevelopment through Spatiotemporal Averaging. *Dev Cell* **38**: 15-32
- Huang M, Friso G, Nishimura K, Qu X, Olinares PD, Majeran W, Sun Q, van Wijk KJ** (2013) Construction of plastid reference proteomes for maize and Arabidopsis and evaluation of their orthologous relationships; the concept of orthoproteomics. *J Proteome Res* **12**: 491-504

- Humbard MA, Surkov S, De Donatis GM, Jenkins LM, Maurizi MR** (2013) The N-degradome of *Escherichia coli*: limited proteolysis in vivo generates a large pool of proteins bearing N-degrons. *J Biol Chem* **288**: 28913-28924
- Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC** (1999) Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* **286**: 2165-2169
- Hwang BJ, Woo KM, Goldberg AL, Chung CH** (1988) Protease Ti, a new ATP-dependent protease in *Escherichia coli*, contains protein-activated ATPase and proteolytic functions in distinct subunits. *J Biol Chem* **263**: 8727-8734
- Inoue H, Li M, Schnell DJ** (2013) An essential role for chloroplast heat shock protein 90 (Hsp90C) in protein import into chloroplasts. *Proc Natl Acad Sci U S A* **110**: 3173-3178
- Inoue H, Rounds C, Schnell DJ** (2010) The molecular basis for distinct pathways for protein import into *Arabidopsis* chloroplasts. *Plant Cell* **22**: 1947-1960
- Jarvis P, Lopez-Juez E** (2013) Biogenesis and homeostasis of chloroplasts and other plastids. *Nat Rev Mol Cell Biol* **14**: 787-802
- Johnson KA, Bhushan S, Stahl A, Hallberg BM, Frohn A, Glaser E, Eneqvist T** (2006) The closed structure of presequence protease PreP forms a unique 10,000 Angstroms³ chamber for proteolysis. *EMBO J* **25**: 1977-1986
- Julien O, Wells JA** (2017) Caspases and their substrates. *Cell Death Differ* **24**: 1380-1389
- Julien O, Zhuang M, Wiita AP, O'Donoghue AJ, Knudsen GM, Craik CS, Wells JA** (2016) Quantitative MS-based enzymology of caspases reveals distinct protein substrate specificities, hierarchies, and cellular roles. *Proc Natl Acad Sci U S A* **113**: E2001-2010
- Kim J, Kimber MS, Nishimura K, Friso G, Schultz L, Ponnala L, van Wijk KJ** (2015) Structures, Functions, and Interactions of ClpT1 and ClpT2 in the Clp Protease System of *Arabidopsis* Chloroplasts. *Plant Cell* **27**: 1477-1496
- Kim J, Olinares PD, Oh SH, Ghisaura S, Poliakov A, Ponnala L, van Wijk KJ** (2013) Modified Clp protease complex in the ClpP3 null mutant and consequences for chloroplast development and function in *Arabidopsis*. *Plant Physiol* **162**: 157-179
- King JV, Liang WG, Scherpelz KP, Schilling AB, Meredith SC, Tang WJ** (2014) Molecular basis of substrate recognition and degradation by human presequence protease. *Structure* **22**: 996-1007
- Kitada S, Uchiyama T, Funatsu T, Kitada Y, Ogishima T, Ito A** (2007) A protein from a parasitic microorganism, *Rickettsia prowazekii*, can cleave the signal sequences of proteins targeting mitochondria. *J Bacteriol* **189**: 844-850

- Kleifeld O, Doucet A, Prudova A, auf dem Keller U, Gioia M, Kizhakkedathu JN, Overall CM** (2011) Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nat Protoc* **6**: 1578-1611
- Kley J, Schmidt B, Boyanov B, Stolt-Bergner PC, Kirk R, Ehrmann M, Knopf RR, Naveh L, Adam Z, Clausen T** (2011) Structural adaptation of the plant protease Deg1 to repair photosystem II during light exposure. *Nat Struct Mol Biol* **18**: 728-731
- Kmiec B, Teixeira PF, Berntsson RP, Murcha MW, Branca RM, Radomiljac JD, Regberg J, Svensson LM, Bakali A, Langel U, Lehtio J, Whelan J, Stenmark P, Glaser E** (2013) Organellar oligopeptidase (OOP) provides a complementary pathway for targeting peptide degradation in mitochondria and chloroplasts. *Proc Natl Acad Sci U S A* **110**: E3761-3769
- Kmiec B, Teixeira PF, Glaser E** (2014) Shredding the signal: targeting peptide degradation in mitochondria and chloroplasts. *Trends Plant Sci* **19**: 771-778
- Kmiec B, Teixeira PF, Glaser E** (2014) Shredding the signal: targeting peptide degradation in mitochondria and chloroplasts. *Trends Plant Sci*
- Lange PF, Overall CM** (2013) Protein TAILS: when termini tell tales of proteolysis and function. *Curr Opin Chem Biol* **17**: 73-82
- Lee S, Lee DW, Lee Y, Mayer U, Stierhof YD, Lee S, Jurgens G, Hwang I** (2009) Heat shock protein cognate 70-4 and an E3 ubiquitin ligase, CHIP, mediate plastid-destined precursor degradation through the ubiquitin-26S proteasome system in Arabidopsis. *Plant Cell* **21**: 3984-4001
- Li L, Millar AH** (2017) Protein Degradation Rate in Arabidopsis thaliana Leaf Growth and Development. *The Plant Cell Online*
- Liebers M, Grubler B, Chevalier F, Lerbs-Mache S, Merendino L, Blanvillain R, Pfannschmidt T** (2017) Regulatory Shifts in Plastid Transcription Play a Key Role in Morphological Conversions of Plastids during Plant Development. *Front Plant Sci* **8**: 23
- Ling Q, Huang W, Baldwin A, Jarvis P** (2012) Chloroplast biogenesis is regulated by direct action of the ubiquitin-proteasome system. *Science* **338**: 655-659
- Majovsky P, Naumann C, Lee CW, Lassowskat I, Trujillo M, Dissmeyer N, Hoehenwarter W** (2014) Targeted proteomics analysis of protein degradation in plant signaling on an LTQ-Orbitrap mass spectrometer. *J Proteome Res* **13**: 4246-4258
- Mossmann D, Vogtle FN, Taskin AA, Teixeira PF, Ring J, Burkhart JM, Burger N, Pinho CM, Tadic J, Loreth D, Graff C, Metzger F, Sickmann A, Kretz O, Wiedemann N, Zahedi RP, Madeo F, Glaser E, Meisinger C** (2014) Amyloid-beta peptide induces mitochondrial dysfunction by inhibition of preprotein maturation. *Cell Metab* **20**: 662-669

- Nakai M** (2015) The TIC complex uncovered: The alternative view on the molecular mechanism of protein translocation across the inner envelope membrane of chloroplasts. *Biochim Biophys Acta* **1847**: 957-967
- Naqvi MA, Irani KA, Katanishoostari M, Rousseau D** (2016) Disorder in Milk Proteins: Formation, Structure, Function, Isolation and Applications of Casein Phosphopeptides. *Curr Protein Pept Sci* **17**: 368-379
- Nath K, Jajoo A, Poudyal RS, Timilsina R, Park YS, Aro EM, Nam HG, Lee CH** (2013) Towards a critical understanding of the photosystem II repair mechanism and its regulation during stress conditions. *FEBS Lett* **587**: 3372-3381
- Neuhaus HE, Emes MJ** (2000) Nonphotosynthetic Metabolism in Plastids. *Annu Rev Plant Physiol Plant Mol Biol* **51**: 111-140
- Nishimura K, Apitz J, Friso G, Kim J, Ponnala L, Grimm B, van Wijk KJ** (2015) Discovery of a Unique Clp Component, ClpF, in Chloroplasts: A Proposed Binary ClpF-ClpS1 Adaptor Complex Functions in Substrate Recognition and Delivery. *Plant Cell* **27**: 2677-2691
- Nishimura K, Asakura Y, Friso G, Kim J, Oh SH, Rutschow H, Ponnala L, van Wijk KJ** (2013) ClpS1 is a conserved substrate selector for the chloroplast Clp protease system in Arabidopsis. *Plant Cell* **25**: 2276-2301
- Nishimura K, Kato Y, Sakamoto W** (2017) Essentials of Proteolytic Machineries in Chloroplasts. *Mol Plant* **10**: 4-19
- Nishimura K, van Wijk KJ** (2015) Organization, function and substrates of the essential Clp protease system in plastids. *Biochim Biophys Acta* **1847**: 915-930
- Pace CN, Scholtz JM** (1998) A helix propensity scale based on experimental studies of peptides and proteins. *Biophys J* **75**: 422-427
- Paila YD, Richardson LG, Schnell DJ** (2014) New Insights into the Mechanism of Chloroplast Protein Import and Its Integration with Protein Quality Control, Organelle Biogenesis and Development. *J Mol Biol*
- Paila YD, Richardson LG, Schnell DJ** (2015) New insights into the mechanism of chloroplast protein import and its integration with protein quality control, organelle biogenesis and development. *J Mol Biol* **427**: 1038-1060
- Ponpuak M, Klemba M, Park M, Gluzman IY, Lamppa GK, Goldberg DE** (2007) A role for falcilysin in transit peptide degradation in the Plasmodium falciparum apicoplast. *Mol Microbiol* **63**: 314-334

- Prudova A, Serrano K, Eckhard U, Fortelny N, Devine DV, Overall CM** (2014) TAILS N-terminomics of human platelets reveals pervasive metalloproteinase dependent proteolytic processing in storage. *Blood*
- Rawlings ND, Barrett AJ, Bateman A** (2012) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **40**: D343-350
- Richter S, Lamppa G** (1998) A chloroplast processing enzyme functions as the general stromal processing peptidase. *Proc Natl Acad Sci U S A* **95**: 7463 -7468
- Richter S, Zhong R, Lamppa G** (2005) Function of the stromal processing peptidase in the chloroplast import pathway. *Physiologia Plantarum* **123**: 362-368
- Rigas S, Daras G, Laxa M, Marathias N, Fasseas C, Sweetlove LJ, Hatzopoulos P** (2009) Role of Lon1 protease in post-germinative growth and maintenance of mitochondrial function in *Arabidopsis thaliana*. *New Phytol* **181**: 588-600
- Rolland N, Curien G, Finazzi G, Kuntz M, Marechal E, Matringe M, Ravanel S, Seigneurin-Berny D** (2012) The biosynthetic capacities of the plastids and integration between cytoplasmic and chloroplast processes. *Annu Rev Genet* **46**: 233-264
- Ross S, Giglione C, Pierre M, Espagne C, Meinnel T** (2005) Functional and developmental impact of cytosolic protein N-terminal methionine excision in *Arabidopsis*. *Plant Physiol* **137**: 623-637
- Rudhe C, Clifton R, Chew O, Zemam K, Richter S, Lamppa G, Whelan J, Glaser E** (2004) Processing of the dual targeted precursor protein of glutathione reductase in mitochondria and chloroplasts. *J Mol Biol* **343**: 639-647
- Sauer RT, Bolon DN, Burton BM, Burton RE, Flynn JM, Grant RA, Hersch GL, Joshi SA, Kenniston JA, Levchenko I, Neher SB, Oakes ES, Siddiqui SM, Wah DA, Baker TA** (2004) Sculpting the proteome with AAA(+) proteases and disassembly machines. *Cell* **119**: 9-18
- Savage LJ, Imre KM, Hall DA, Last RL** (2013) Analysis of essential *Arabidopsis* nuclear genes encoding plastid-targeted proteins. *PLoS One* **8**: e73291
- Schilling O, Huesgen PF, Barre O, Auf dem Keller U, Overall CM** (2011) Characterization of the prime and non-prime active site specificities of proteases by proteome-derived peptide libraries and tandem mass spectrometry. *Nat Protoc* **6**: 111-120
- Schnell DJ** (2014) Discussion with Dr. Schnell following research seminar. *In*, Cornell University
- Schnell DJ, Blobel G** (1993) Identification of intermediates in the pathway of protein import into chloroplasts and their localization to envelope contact sites. *J Cell Biol* **120**: 103-115

- Schuhmann H, Huesgen PF, Adamska I** (2012) The family of Deg/HtrA proteases in plants. *BMC Plant Biol* **12**: 52
- Shi LX, Theg SM** (2013) The chloroplast protein import system: from algae to trees. *Biochim Biophys Acta* **1833**: 314-331
- Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M** (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J* **5**: 2043-2049
- Sjogren LL, Clarke AK** (2011) Studying proteases and protein turnover in Arabidopsis chloroplasts. *Methods Mol Biol* **774**: 225-240
- Squires C, Squires CL** (1992) The Clp proteins: proteolysis regulators or molecular chaperones? *J Bacteriol* **174**: 1081-1085
- Stahl A, Nilsson S, Lundberg P, Bhushan S, Biverstahl H, Moberg P, Morisset M, Vener A, Maler L, Langel U, Glaser E** (2005) Two novel targeting peptide degrading proteases, PrePs, in mitochondria and chloroplasts, so similar and still different. *J Mol Biol* **349**: 847-860
- Starheim KK, Gevaert K, Arnesen T** (2012) Protein N-terminal acetyltransferases: when the start matters. *Trends Biochem Sci* **37**: 152-161
- Switzar L, Giera M, Niessen WM** (2013) Protein digestion: an overview of the available techniques and recent developments. *J Proteome Res* **12**: 1067-1077
- Taskin AA, Kucukkose C, Burger N, Mossmann D, Meisinger C, Vogtle FN** (2017) The novel mitochondrial matrix protease Ste23 is required for efficient presequence degradation and processing. *Mol Biol Cell* **28**: 997-1002
- Taylor AB, Smith BS, Kitada S, Kojima K, Miyaura H, Otwinowski Z, Ito A, Deisenhofer J** (2001) Crystal structures of mitochondrial processing peptidase reveal the mode for specific cleavage of import signal sequences. *Structure* **9**: 615-625
- Teixeira PF, Glaser E** (2013) Processing peptidases in mitochondria and chloroplasts. *Biochim Biophys Acta* **1833**: 360-370
- Teixeira PF, Kmiec B, Branca RM, Murcha MW, Byzia A, Ivanova A, Whelan J, Drag M, Lehtio J, Glaser E** (2017) A multi-step peptidolytic cascade for amino acid recovery in chloroplasts. *Nat Chem Biol* **13**: 15-17

- Trentini DB, Suskiewicz MJ, Heuck A, Kurzbauer R, Deszcz L, Mechtler K, Clausen T** (2016) Arginine phosphorylation marks proteins for degradation by a Clp protease. *Nature* **539**: 48-53
- Trosch R, Jarvis P** (2011) The stromal processing peptidase of chloroplasts is essential in Arabidopsis, with knockout mutations causing embryo arrest after the 16-cell stage. *PLoS One* **6**: e23039
- Tsiatsiani L, Timmerman E, De Bock PJ, Vercaemmen D, Stael S, van de Cotte B, Staes A, Goethals M, Beunens T, Van Damme P, Gevaert K, Van Breusegem F** (2013) The Arabidopsis metacaspase9 degradome. *Plant Cell* **25**: 2831-2847
- Van Damme P, Gawron D, Van Criekinge W, Menschaert G** (2014) N-terminal proteomics and ribosome profiling provide a comprehensive view of the alternative translation initiation landscape in mice and men. *Mol Cell Proteomics* **13**: 1245-1261
- van Wijk KJ** (2015) Protein maturation and proteolysis in plant plastids, mitochondria, and peroxisomes. *Annu Rev Plant Biol* **66**: 75-111
- van Wijk KJ, Baginsky S** (2011) Plastid Proteomics in Higher Plants: Current State and Future Goals. *Plant Physiol* **155**: 1578-1588
- van Wijk KJ, Kessler F** (2017) Plastoglobuli: Plastid Microcompartments with Integrated Functions in Metabolism, Plastid Developmental Transitions, and Environmental Adaptation. *Annu Rev Plant Biol* **68**: 253-289
- Vogle FN, Prinz C, Kellermann J, Lottspeich F, Pfanner N, Meisinger C** (2011) Mitochondrial protein turnover: role of the precursor intermediate peptidase Oct1 in protein stabilization. *Mol Biol Cell* **22**: 2135-2143
- Vogle FN, Wortelkamp S, Zahedi RP, Becker D, Leidhold C, Gevaert K, Kellermann J, Voos W, Sickmann A, Pfanner N, Meisinger C** (2009) Global analysis of the mitochondrial N-proteome identifies a processing peptidase critical for protein stability. *Cell* **139**: 428-439
- Von Heijne G, Gavel Y** (1990) A conserved cleavage-site motif in chloroplast transit peptides. *FEBS J* **261**: 4
- Walling LL** (2006) Recycling or regulation? The role of amino-terminal modifying enzymes. *Curr Opin Plant Biol* **9**: 227-233
- White MD, Klecker M, Hopkinson RJ, Weits DA, Mueller C, Naumann C, O'Neill R, Wickens J, Yang J, Brooks-Bartlett JC, Garman EF, Grossmann TN, Dissmeyer N, Flashman E** (2017) Plant cysteine oxidases are dioxygenases that directly enable arginyl transferase-catalysed arginylation of N-end rule targets. *Nat Commun* **8**: 14690

Wildburger NC, Esparza TJ, LeDuc RD, Fellers RT, Thomas PM, Cairns NJ, Kelleher NL, Bateman RJ, Brody DL (2017) Diversity of Amyloid-beta Proteoforms in the Alzheimer's Disease Brain. *Sci Rep* **7**: 9520

Yaffe MP, Schatz G (1984) Two nuclear mutations that block mitochondrial protein import in yeast. *Proc Natl Acad Sci U S A* **81**: 4819-4823

Zimorski V, Ku C, Martin WF, Gould SB (2014) Endosymbiotic theory for organelle origins. *Curr Opin Microbiol* **22**: 38-48

Zybilov B, Rutschow H, Friso G, Rudella A, Emanuelsson O, Sun Q, van Wijk KJ (2008) Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS One* **3**: e1994

APPENDIX

Supplemental Table 1.1. Proteins and gene identifiers used to construct M16 protease phylogeny (Figure 1.3)

Protein	Gene identifier	Species	Group
Acy2	WP_015214686.1	Anabaena cylindrica	cyanobacteria
Acy3	WP_015216675.1	Anabaena cylindrica	cyanobacteria
Ath1	AT1G06900	Arabidopsis thaliana	plants
Ath10	AT5G56730	Arabidopsis thaliana	plants
Ath2	AT1G49630	Arabidopsis thaliana	plants
Ath3	AT1G51980	Arabidopsis thaliana	plants
Ath4	AT2G41790	Arabidopsis thaliana	plants
Ath5	AT3G02090	Arabidopsis thaliana	plants
Ath6	AT3G16480	Arabidopsis thaliana	plants
Ath7	AT3G19170	Arabidopsis thaliana	plants
Ath9	AT5G42390	Arabidopsis thaliana	plants
Atu2	WP_042617173.1	Agrobacterium tumefaciens	bacteria
Bra1	Brara.D00283	Brassica rappa	plants
Bra2	Brara.E01598	Brassica rappa	plants
Bra3	Brara.E02505	Brassica rappa	plants
Bra4	Brara.F00346	Brassica rappa	plants
Bra5	Brara.I03967	Brassica rappa	plants
Bra6	Brara.J00475	Brassica rappa	plants
Bra7	Brara.J01099	Brassica rappa	plants
Ccr1	XP_005715652.1	Chondrus crispus	red algae
Ccr2	XP_005716261.1	Chondrus crispus	red algae
Ccr3	XP_005716826.1	Chondrus crispus	red algae
Ccr4	XP_005717973.1	Chondrus crispus	red algae
Cdi1	YP_001087816.1	Clostridioides difficile 630	bacteria
Cme1	XP_005537012.1	Cyanidioschyzon merolae 10D	red algae
Cme2	XP_005537047.1	Cyanidioschyzon merolae 10D	red algae
Cme3	XP_005537605.1	Cyanidioschyzon merolae 10D	red algae
Cme4	XP_005538043.1	Cyanidioschyzon merolae 10D	red algae
Cme5	XP_005538148.1	Cyanidioschyzon merolae 10D	red algae

Supplemental Table 1.1. (Continued)

Protein	Gene identifier	Species	Group
Cre1	Cre01.g020918	Chlamydomonas reinhardtii	green algae
Cre2	Cre09.g386113	Chlamydomonas reinhardtii	green algae
Cre4	Cre17.g723250	Chlamydomonas reinhardtii	green algae
Csu1	C169e_gw 1.4.49.1	Chlamydomonas reinhardtii	green algae
Csu2	20138	Chlamydomonas reinhardtii	green algae
Csu3	40264	Chlamydomonas reinhardtii	green algae
Csu4	150053	Chlamydomonas reinhardtii	green algae
Csu5	30279	Chlamydomonas reinhardtii	green algae
Csu6	pm.2_112	Chlamydomonas reinhardtii	green algae
Ddi1	XP_001134518.1	Dictyostelium discoideum AX4	amoebae
Ddi2	XP_001134603.1	Dictyostelium discoideum AX4	amoebae
Ddi3	XP_637100.1	Dictyostelium discoideum AX4	amoebae
Ddi4	XP_645544.1	Dictyostelium discoideum AX4	amoebae
Dme1	NP_524182.3	Drosophila melanogaster	metazoa
Dme2	NP_572757.2	Drosophila melanogaster	metazoa
Dme3	NP_610156.1	Drosophila melanogaster	metazoa
Dme4	NP_610333.1	Drosophila melanogaster	metazoa
Dme6	NP_649271.1	Drosophila melanogaster	metazoa
Dme7	NP_650401.1	Drosophila melanogaster	metazoa
Dra1	NP_294340.1	Deinococcus radiodurans R1	bacteria
Dra2	NP_296235.1	Deinococcus radiodurans R1	bacteria
Dsa1	Dusal.0069s00012	Dunaliella salina	green algae
Dsa2	Dusal.1182s00001	Dunaliella salina	green algae
Eco1	NP_416011.1	Escherichia coli str. K12 substr. MG1655	bacteria
Eco2	NP_417298.1	Escherichia coli str. K12 substr. MG1655	bacteria
Hse1	NP_001229236.1	Homo sapiens	metazoa
Hse2	NP_002516.2	Homo sapiens	metazoa
Hse3	NP_003356.2	Homo sapiens	metazoa
Hse5	NP_004270.2	Homo sapiens	metazoa
Hse6	NP_055975.1	Homo sapiens	metazoa
Ldo1	XP_003857880.1	Leishmania donovani	protozoan parasite
Ldo2	XP_003858467.1	Leishmania donovani	protozoan parasite
Ldo3	XP_003863405.1	Leishmania donovani	protozoan parasite
Mae1	WP_012265321.1	Microcystis aeruginosa	bacteria
Mmu1	NP_079683.2	Mus musculus	metazoa
Mmu3	NP_112419.3	Mus musculus	metazoa
Mmu4	NP_660113.1	Mus musculus	metazoa
Mmu5	NP_666262.2	Mus musculus	metazoa

Supplemental Table 1.1. (Continued)

Protein	Gene identifier	Species	Group
Mpo1	Mapoly0007s0115	Marchantia polymorpha	plants
Mpo2	Mapoly0009s0206	Marchantia polymorpha	plants
Mpo3	Mapoly0021s0007	Marchantia polymorpha	plants
Mpo4	Mapoly0022s0090	Marchantia polymorpha	plants
Mpo5	Mapoly0028s0022	Marchantia polymorpha	plants
Mpo6	Mapoly0029s0118	Marchantia polymorpha	plants
Mpo7	Mapoly0138s0034	Marchantia polymorpha	plants
Mpu1	70045	Micromonas pusilla CCMP1545 v3	green algae
Mpu2	t10048	Micromonas pusilla CCMP1545 v3	green algae
Mpu3	t30221	Micromonas pusilla CCMP1545 v3	green algae
Mpu4	gw 1.4.30.1	Micromonas pusilla CCMP1545 v3	green algae
Mpu5	31112	Micromonas pusilla CCMP1545 v3	green algae
Mpu6	11000029	Micromonas pusilla CCMP1545 v3	green algae
Msp1	gw 2.08.32.1	Micromonas sp. RCC299 v3	green algae
Msp2	50024	Micromonas sp. RCC299 v3	green algae
Msp3	100139	Micromonas sp. RCC299 v3	green algae
Msp4	130107	Micromonas sp. RCC299 v3	green algae
Msp5	140456	Micromonas sp. RCC299 v3	green algae
Msp6	200010329	Micromonas sp. RCC299 v3	green algae
Msp7	500010336	Micromonas sp. RCC299 v3	green algae
Mtr1	Medtr3g466820	Medicago truncatula	plants
Mtr2	Medtr5g004680	Medicago truncatula	plants
Mtr3	Medtr5g007380	Medicago truncatula	plants
Mtr4	Medtr5g095180	Medicago truncatula	plants
Mtr5	Medtr7g053330	Medicago truncatula	plants
Mtr6	Medtr7g091590	Medicago truncatula	plants
Mtr7	Medtr8g014660	Medicago truncatula	plants
Mtr8	Medtr8g101950	Medicago truncatula	plants
Mtr9	Medtr8g102230	Medicago truncatula	plants
Mtu1	NP_217298.1	Mycobacterium tuberculosis H37Rv	bacteria
Nsp1	WP_010999251.1	Nostoc sp. PCC 7120	cyanobacteria
Oma1	gw Euk.13.11.1	Ostreococcus lucimarinus	green algae
Oma2	190077	Ostreococcus lucimarinus	green algae
Oma3	210120	Ostreococcus lucimarinus	green algae
Oma4	40029	Ostreococcus lucimarinus	green algae
Oma5	500010332	Ostreococcus lucimarinus	green algae
Oma6	600010345	Ostreococcus lucimarinus	green algae
Oma7	2000010241	Ostreococcus lucimarinus	green algae

Supplemental Table 1.1. (Continued)

Protein	Gene identifier	Species	Group
Osa10	Os05g44916	Oryza sativa	plants
Osa11	Os07g38260	Oryza sativa	plants
Osa12	Os07g38270	Oryza sativa	plants
Osa13	Os07g38280	Oryza sativa	plants
Osa2	Os01g37825	Oryza sativa	plants
Osa3	Os01g51390	Oryza sativa	plants
Osa4	Os01g53700	Oryza sativa	plants
Osa5	Os01g57073	Oryza sativa	plants
Osa6	Os01g73550	Oryza sativa	plants
Osa7	Os02g52390	Oryza sativa	plants
Osa8	Os03g11410	Oryza sativa	plants
Osa9	Os03g21810	Oryza sativa	plants
Pae1	NP_250663.1	Pseudomonas aeruginosa PAO1	bacteria
Pfa1	XP_001350319.1	Plasmodium falciparum 3D7	protozoan parasite
Pfa2	XP_001351788.2	Plasmodium falciparum 3D7	protozoan parasite
Pfa3	XP_001352201.1	Plasmodium falciparum 3D7	protozoan parasite
Ppa1	Pp3c1_4220	Physcomitrella patens	plants
Ppa10	Pp3c3_11990	Physcomitrella patens	plants
Ppa11	Pp3c4_13680	Physcomitrella patens	plants
Ppa12	Pp3c5_3160	Physcomitrella patens	plants
Ppa13	Pp3c5_6610	Physcomitrella patens	plants
ppa14	Pp3c5_8950	Physcomitrella patens	plants
Ppa2	Pp3c10_5860	Physcomitrella patens	plants
Ppa3	Pp3c11_12620	Physcomitrella patens	plants
Ppa4	Pp3c13_21610	Physcomitrella patens	plants
Ppa5	Pp3c14_4243	Physcomitrella patens	plants
Ppa6	Pp3c16_17660	Physcomitrella patens	plants
Ppa7	Pp3c16_2100	Physcomitrella patens	plants
Ppa8	Pp3c16_8340	Physcomitrella patens	plants
Ppa9	Pp3c19_10200	Physcomitrella patens	plants
Ptr1	001G191100	Populus trichocarpa	plants
Ptr2	002G235700	Populus trichocarpa	plants
Ptr3	004G142300	Populus trichocarpa	plants
Ptr4	006G050700	Populus trichocarpa	plants
Ptr5	008G193200	Populus trichocarpa	plants
Ptr6	010G036700	Populus trichocarpa	plants
Ptr7	013G154600	Populus trichocarpa	plants
Ptr8	017G092400	Populus trichocarpa	plants
Rpr1	WP_004596018.1	Rickettsia prow azekii	bacteria
Sce1	YDR430C	Saccharomyces cerevisiae	metazoa
Sce2	YHR024C	Saccharomyces cerevisiae	metazoa
Sce3	YLR163C	Saccharomyces cerevisiae	metazoa
Sce4	YLR389C	Saccharomyces cerevisiae	metazoa

Supplemental Table 1.1. (Continued)

Protein	Gene identifier	Species	Group
Sce5	YOL098C	<i>Saccharomyces cerevisiae</i>	metazoa
Sco1	NP_629863.1	<i>Streptomyces coelicolor</i> A3(2)	bacteria
Sel1	WP_011378267.1	<i>Synechococcus</i>	cyanobacteria
Sel2	WP_011430985.1	<i>Synechococcus</i> sp. JA33Ab	cyanobacteria
Sel4	WP_017324394.1	<i>Synechococcus</i> sp. PCC 7336	cyanobacteria
Sly1	Solyc01g108600.2	<i>Solanum lycopersicum</i>	plants
Sly2	Solyc02g043860.2	<i>Solanum lycopersicum</i>	plants
Sly3	Solyc02g088700.2	<i>Solanum lycopersicum</i>	plants
Sly4	Solyc03g118430.2	<i>Solanum lycopersicum</i>	plants
Sly5	Solyc04g015680.2	<i>Solanum lycopersicum</i>	plants
Sly6	Solyc04g015690.2	<i>Solanum lycopersicum</i>	plants
Sly7	Solyc05g012480.2	<i>Solanum lycopersicum</i>	plants
Sly8	Solyc12g008630.1	<i>Solanum lycopersicum</i>	plants
Sly9	Solyc12g008710.1	<i>Solanum lycopersicum</i>	plants
Smo1	146300	<i>Selaginella moellendorffii</i>	plants
Smo2	152047	<i>Selaginella moellendorffii</i>	plants
Smo3	154839	<i>Selaginella moellendorffii</i>	plants
Smo4	181652	<i>Selaginella moellendorffii</i>	plants
Smo5	183257	<i>Selaginella moellendorffii</i>	plants
Smo6	440740	<i>Selaginella moellendorffii</i>	plants
Smo7	447027	<i>Selaginella moellendorffii</i>	plants
Smo8	448827	<i>Selaginella moellendorffii</i>	plants
Son1	NP_718646.1	<i>Shewanella oneidensis</i> MR1	bacteria
Ssp1	WP_010873020.1	<i>Synechocystis</i> sp. PCC 6803	cyanobacteria
Ssp2	WP_010873523.1	<i>Synechocystis</i> sp. PCC 6803	cyanobacteria
Tma1	NP_229147.1	<i>Thermotoga maritima</i> MSB8	bacteria
Vca1	Vocar.0006s0465	<i>Volvox carteri</i>	green algae
Vca2	Vocar.0024s0072	<i>Volvox carteri</i>	green algae
Vca3	Vocar.0024s0073	<i>Volvox carteri</i>	green algae
Vca4	Vocar.0032s0130	<i>Volvox carteri</i>	green algae
Vca5	Vocar.0045s0019	<i>Volvox carteri</i>	green algae
Vca6	Vocar.0069s0003	<i>Volvox carteri</i>	green algae
Vvi1	GSVIVG01013468001	<i>Vitis vinifera</i>	plants
Vvi2	GSVIVG01015250001	<i>Vitis vinifera</i>	plants
Vvi3	GSVIVG01015273001	<i>Vitis vinifera</i>	plants
Vvi4	GSVIVG01016955001	<i>Vitis vinifera</i>	plants
Vvi5	GSVIVG01024571001	<i>Vitis vinifera</i>	plants
Vvi6	GSVIVG01031766001	<i>Vitis vinifera</i>	plants
Vvi7	GSVIVG01033283001	<i>Vitis vinifera</i>	plants
Zma1	PH207Zm00008a000746	<i>Zea mays</i>	plants
Zma2	PH207Zm00008a023071	<i>Zea mays</i>	plants
Zma3	PH207Zm00008a026249	<i>Zea mays</i>	plants
Zma4	PH207Zm00008a029136	<i>Zea mays</i>	plants
Zma5	PH207Zm00008a032360	<i>Zea mays</i>	plants
Zma6	PH207Zm00008a032445	<i>Zea mays</i>	plants
Zma7	PH207Zm00008a034716	<i>Zea mays</i>	plants

CHAPTER 2
THE ARABIDOPSIS CHLOROPLAST STROMAL N-TERMINOME;
COMPLEXITIES OF N-TERMINAL PROTEIN MATURATION AND
STABILITY¹

2.1 ABSTRACT

Protein N-termini are prone to modifications and are major determinants of protein stability in bacteria, eukaryotes, and perhaps also in chloroplasts. Most chloroplast proteins undergo N-terminal maturation, but this is poorly understood due to insufficient experimental information. Consequently, N-termini of mature chloroplast proteins cannot be accurately predicted. This motivated an extensive characterization of chloroplast protein N-termini using terminal amine isotopic labeling of substrates (TAILS) and mass spectrometry, generating nearly 14,000 MS/MS spectra matching to protein N-termini. Many nuclear-encoded plastid proteins accumulated with two or three different N-termini; we evaluated the significance of these different proteoforms. Ala, Val, Thr (often in N- α acetylated form) and Ser were by far the most observed N-terminal residues, even after normalization for their frequency in the plastid proteome, while other residues were absent or highly under-represented. Plastid-encoded proteins

¹ Published in *Plant Physiology* (2015) **Elden Rowland, Jitae Kim, Nazmul Bhuiyan and Klaas J. van Wijk**

The contribution of the thesis' author to this work consisted of all mass spectrometry experiments and data analysis. JK and NB prepared chloroplast stroma. ER and KJVW wrote the manuscript.

showed a comparable distribution of N-terminal residues, but with a higher frequency of Met. Infrequent residues (*e.g.* Ile, Arg, Cys, Pro, Asp, Glu) were observed for several abundant proteins (*e.g.* HSP70/90, RBCL, Fd-GOGAT) likely reflecting functional regulation through their N-termini. In contrast, the thylakoid luminal proteome showed a wide diversity of N-terminal residues, including those typically associated with instability (Asp, Glu, Leu, Phe). We propose that after cleavage of the chloroplast transit peptide by stromal processing peptidase, additional processing by unidentified peptidases occurs to avoid unstable or otherwise unfavorable N-terminal residues. The possibility of a chloroplast N-end rule is discussed.

2.2 INTRODUCTION

Following synthesis, most proteins undergo various N-terminal protein modifications, including removal of the N-terminal (Nt) methionine and signal peptide, Nt α -acetylation (NAA), ubiquitination and acylation. These Nt modifications play an important role in the regulation of cellular functions. The N-termini of a proteins have been shown to be a major determinant of their stability in bacteria (Varshavsky, 2011), eukaryotes (Graciet et al., 2009), mitochondria, and perhaps in plastids/chloroplasts (Apel et al., 2010; Nishimura et al., 2013; van Wijk, 2015). The role of the N-terminus in protein stability is conceptualized in the ‘N-end rule’ which states that certain amino acids, when exposed at the N-terminus of a protein, act as trigger for degradation (Bachmair et al., 1986; Dougan et al., 2012; Tasaki et al., 2012; Gibbs et al., 2014).

Most of the ~3000 plastid proteins are nuclear-encoded (n-encoded) and are targeted to the plastid through an Nt chloroplast transit peptide (cTP). After import, the

cTP is cleaved by the stromal processing peptidase (SPP) (Richter and Lamppa, 1998; Trosch and Jarvis, 2011). The consensus site of cTP cleavage by SPP is only loosely defined, and rules, mechanisms and enzymes for possible subsequent processing, stabilization and other post-translational modifications (PTMs) are not well characterized – for discussion see (van Wijk, 2015). The exact N-terminus is unknown for many chloroplast proteins and cannot be accurately predicted because SPP specificity is not sufficiently understood (Emanuelsson et al., 2000; Zybailov et al., 2008), and probably also because additional Nt processing occurs for many chloroplast proteins (Figure 2.1A). The ~85 plastid-encoded (p-encoded) proteins typically undergo co-translational Nt deformylation, followed by methionine excision (NME) (Giglione et al., 2009) (Figure 2.1B); both of these PTMs are required for normal plastid/chloroplast development and protein stability (Dirk et al., 2001, 2002; Giglione et al., 2003; Meinnel et al., 2006). Both n-encoded and p-encoded proteins can undergo NAA inside the plastid (Zybailov et al., 2008) (Figure 2.1A,B). Postulated functions of NAA in eukaryotes include the mediation of protein location, assembly and stability (Jones and O'Connor, 2011; Starheim et al., 2012; Hoshiyasu et al., 2013; Xu et al., 2015), thereby affecting a variety of processes, including drought tolerance in *Arabidopsis* (Linster et al., 2015).

Typical proteomics workflows generally yield only partial coverage of protein sequences, and it is often difficult to know which peptides represent the true N- or C-termini. Systematic identification of N- or C-termini requires specific labeling and enrichment strategies, such as COmbined FRActional DIagonal Chromatography

(COFRADIC) developed by Gevaert and colleagues (Staes et al., 2011) and Terminal Amine Isotopic Labeling of Substrates (TAILS) developed by the group of Overall

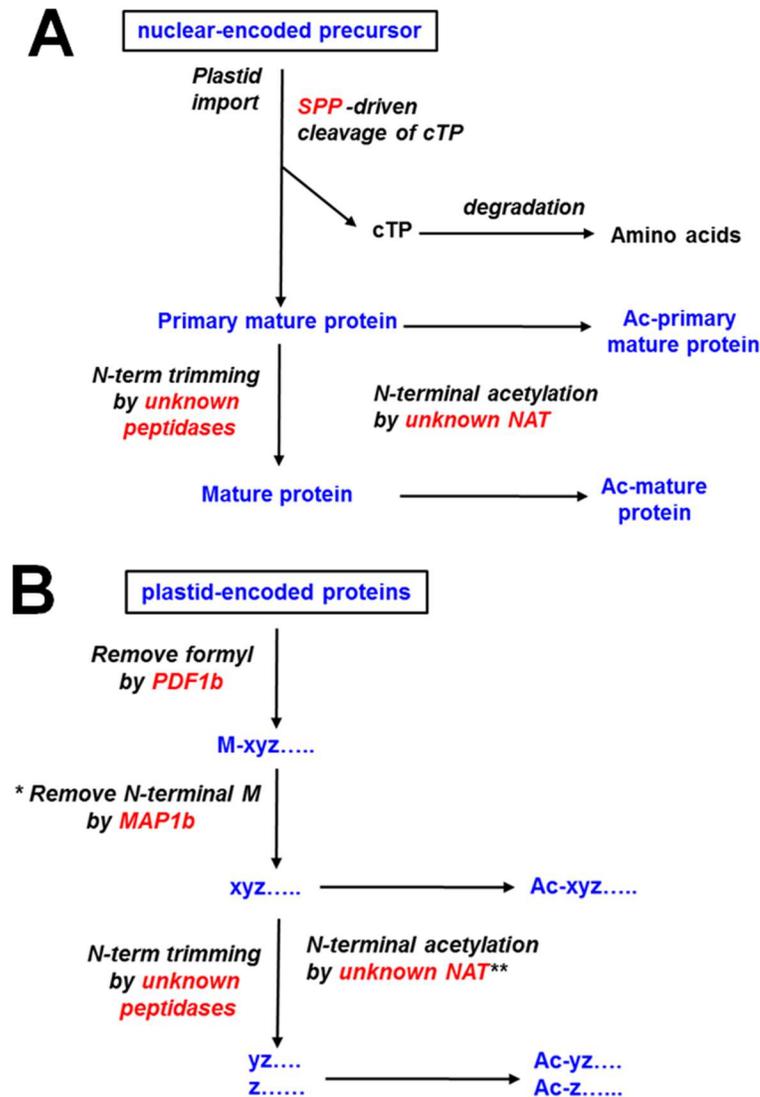


Figure 2.1. Conceptual illustration of Nt maturation of nuclear- and p-encoded proteins. Ac, acetylated; SPP stromal processing peptidase; cTP, chloroplast transit peptide; NAT, N-acetyl transferase; PDF, peptide deformylase; MAP, methionine amino peptidase. **(A)** Nt maturation of nuclear encoded plastid proteins including removal of cTP by SPP and potential subsequent Nt modifications. **(B)** Nt maturation of p-encoded proteins.

(Kleifeld et al., 2011; Lange and Overall, 2013). These strategies allow identification of different Nt proteoforms and were recently also applied to plants (Tsiatsiani et al., 2013; Carrie et al., 2015; Kohler et al., 2015; Zhang et al., 2015) and diatoms (Huesgen et al., 2013). We previously reported on N-termini (Nti) of chloroplast proteins based on MS/MS analysis, but because no Nt enrichment/labeling technique was used, only those that underwent NAA could be considered *bona fide* Nti (Zybailov et al., 2008). Nt Edman degradation sequencing was systematically carried out for thylakoid lumen proteins (Peltier et al., 2000; Peltier et al., 2002), but not for stromal proteins or chloroplast membrane proteins with their Nti exposed to the stroma. The Nti of thylakoid lumen proteins are mostly generated by luminal peptidase(s) (Hsu et al., 2011; Midorikawa et al., 2014) and the thylakoid lumen contains a different set of peptidases than the stroma; hence rules for Nt maturation and stability are likely different than for stroma-exposed proteins.

The objective of this study was to systematically determine the Nti of stroma-exposed chloroplast proteins of *Arabidopsis thaliana* (the N-terminome) and to provide a baseline for understanding Nt protein maturation and protein stability in the chloroplast stroma. To that end, we applied the TAILS technique and determined the Nti of ~250 chloroplast proteins by mass spectrometry (MS). We observed that many n-encoded plastid proteins accumulated with 2 or even 3 different Nt residues, in many cases both with or without NAA. The extent of accumulation of different N-terminal proteoforms is surprising and will be discussed. P-encoded proteins generally showed very similar Nt residues to those of n-encoded proteins, with the exception of Met. Our data show that small, apolar or hydroxylated residues (Ala, Val, Ser, Thr) are the most

frequent Nt residues of stromal proteins, whereas other residues are strictly avoided, or are only present for very specific proteins likely to aid in their function. Chloroplast protein degradation products were also detected, with enrichment for peptides generated by cleavage between Arg and Thr. We present testable hypotheses for understanding Nt processing and maturation, stability and a possible N-end rule in chloroplast stroma.

2.3 RESULTS

2.3.1 Systematic identification of protein N-termini

To systematically identify the Nti of chloroplast proteins, we employed the TAILS method for labeling and enrichment of chloroplast protein Nti, followed by MS/MS-based identification (Kleifeld et al., 2011). The TAILS workflow removes the internal non-Nt peptides, whereas both unmodified α -amino Nti and NAA-Nti are retained, therefore greatly simplifying the remaining proteome. For a general description of the TAILS method, we refer to excellent papers from Overall (Kleifeld et al., 2011; Lange and Overall, 2013), the materials and methods and Supplemental Figure 2.1A. In brief, the TAILS method involves first the dimethyl labeling of free N-terminal α -amines, as well ϵ -amines of Lys residues. Following digestion with protease (trypsin or GluC), the un-modified peptides are removed by crosslinking internal peptides to a soluble polymer, allowing the collection of N-terminal peptides. As starting material, we used developed leaf rosettes of soil-grown Arabidopsis plants, analyzing both soluble stromal protein extracts from isolated chloroplasts (4 independent preparations; 10 TAILS experiments), as well as total soluble leaf protein extracts (3 independent preparations; 9 TAILS experiments). Comparison of Nt

sequences from total leaf and chloroplast stromal extracts allowed us to consider processing of dual-targeted chloroplast proteins (*i.e.* targeted to other subcellular locations, including mitochondria) and to identify chloroplast pre-proteins (*i.e.* with their cTP still attached). Protein recovery across the labeling and enrichment steps was verified by SDS-PAGE followed by silver staining (Supplemental Figure 2.1B). Dimethyl labeling efficiency and proteolytic digestion was monitored by LC/MS/MS analysis of each sample prior to the negative selection step. This showed that >99% of lysines were dimethylated, indicating near quantitative labeling, which allowed a semi-quantitative comparison of different Nt proteoforms.

2.3.2 Assessment and filtering of Nt sequences

All MS/MS search results were pooled and filtered to identify only Nt labeled peptides. Of the complete set of acquired MS/MS spectra across all experiments, 13858 spectra matched to Nt peptides (Supplemental Table 2.1). We then pooled the Nt peptides with the same molecular mass and sequence (irrespective of charge state), resulting in 1037 non-redundant Nti matching to 577 proteins. Matched proteins were annotated for subcellular location to aid in identification of subcellular Nt maturation events (Supplemental Table 2.2). Peptides starting with the same Nt modification and amino acid sequence, but with different C-terminal ends or different modified side chains were merged into 894 Nti matching to 577 proteins (Supplemental Table 2.3). Importantly, these overlapping peptides strengthened Nt identifications. We did not condense peptide sequences with or without NAA, because these NAA sequences should be considered functionally distinct from their unmodified sequences. 544 of

these merged Nti matched to 250 plastid proteins and the remaining peptides matched to proteins located in other subcellular compartments or without assigned subcellular location (Supplemental Table 2.3).

The two main objectives of this study were to i) develop a working hypothesis for cTP cleavage specificity and subsequent maturation steps and ii) determine under- or over-representation of specific Nt amino acids for the steady state, stromal exposed proteome and deduce potential Nt stability rules. Chloroplast proteins with their Nti exposed to the plastid stroma, lumen, intra-envelope space or facing the cytoplasm should be considered separately because they undergo distinct processing steps in each location. Hence we carefully evaluated intra-plastid location for each identified protein. Those encoded by the plastid genome also represent a distinct set since they undergo specific co- and post-translational processing (Figure 2.1B). Sixteen of the detected n-encoded chloroplast proteins are known to be dual targeted to chloroplasts and mitochondria or cytosol (Carrie and Small, 2013) (Supplemental Table 2.3). For most of these dual-targeted proteins, we identified a single N-terminus which appeared to represent the chloroplast-localized form. This is not surprising because we used either protein extracts from photosynthetic leaves in which chloroplast proteins are far more abundant than mitochondrial proteins, or we used isolated chloroplasts. Three dual targeted plastid/cytosolic proteins (ADL1, GOX1 and GSTphi) were only identified in their cytosolic forms and were not further considered for chloroplast N-terminome analysis. In the remaining analysis we will focus on the n- or p-encoded proteins that have stromal exposed Nti. We note that even if these proteins have their Nti facing the stroma, they may actually be buried within the protein structure and thus only truly

exposed to the stroma during biogenesis or degradation.

2.3.3 *Nt AA frequency and acetylation state of n-encoded chloroplast proteins*

For 126 plastid proteins only a single N-terminus was identified; examples are shown in Table 2.1- scenario A. These proteins are interesting because other Nt proteoforms of these proteins must be quickly degraded, or the SPP cleaves the cTPs at only a single location or additional peptidases trim the Nti to a single proteoform. Multiple Nti were detected for ~ 100 proteins representing three different scenarios: i) Nt peptides well upstream of the predicted or previously documented mature N-terminus (for ~ a dozen proteins) (Table 2.1 - scenario B). These Nt peptides were generally found in the total leaf extracts, rather than stromal extracts. This suggests that these upstream Nti were from proteins not yet imported into the chloroplast. Table 2.1 (scenario B) shows the data for four such proteins where Nti of the unprocessed protein was only found in the total leaf extracts but the Nti of processed proteins only in the stroma or in both stroma and leaf extracts, ii) Proteins with multiple closely spaced Nti that each could represent the mature N-terminus of the respective protein (Table 2.1 – scenario C). In most cases a single N-terminus had both the highest number of spectral counts (SPC; these are matched MS/MS spectra) and the most Nt residue, thus representing the most likely candidate for the N-terminus of the steady state protein, iii) Nti of degradation products. 129 Nti (matching to 31 n-encoded and four p-encoded proteins) were likely degradation products (see section ‘*Accumulation of proteolytic products*’).

Table 2. 1 Examples of experimentally determined Nt peptides for selected nuclear-encoded proteins demonstrating three different physiological scenarios. A number of details are provided: Nt residue position for mature protein (a) (predicted/experimental), the residue immediately upstream of the observed Nt peptide (Prev. AA), the Nt modification (TAILS), number of matched MS/MS spectra (Spectral counts - SPC). Examples include cases where the predicted N-terminus is upstream, downstream, or identical to the observed N-terminus.

Scenario A (*)	Protein name	N-terminus (a)	Prev. AA	N-terminal peptide	TAILS	SPC
AT1G356880	50S ribosomal L21	84/66	F	AESVVEAEPETDIEAWWSDVSEVTEEKAKR	Dimethyl	46
AT4G096650	CF1d - atpD	49/48	M	SATAASSYAMALADVAKR	Dimethyl	51
AT3G27830	50S ribosomal L12-A	59/59	A	AVEAPEKEKEKGEISSLTLEEAR	Dimethyl	161
AT1G54630	ACP-3 plastid	52/53	C	AAKPEITVDKVCVVRR	Dimethyl	73
AT4G23100	GSH1	74/75	A	ASPTTEEAVVAITEPLTR	Dimethyl	58
AT5G04140	Fd.GOGAT 1(GS1)	63/106	A	CGVGFIAWLDNIPSHGVVKDALIALGCMHR	Dimethyl	6
Scenario B (**)	Protein name	N-terminus (a)	Prev. AA	N-terminal peptide	TAILS	SPC (stroma/leaf)
AT1G67090	RBCS-4	55/2 55/56	M C	ASSMLSSATMVASPAQATMVAPEFNGLKSSAAPPATR MQVWPPRGKKKFFELSYLPDLTDELAKEDVDYUR	Acetyl Dimethyl	0/3 9/3
AT5G38410	RBCS-3B	55/2 55/56	M C	ASSMLSSAAVVTSPAQATMVAPEFNGLKSSAAPPVTR MKVWPPRGKKKFFELSYLPDLSDVELAKEVDYLLR	Acetyl Dimethyl	0/3 5/0
AT4G38970	SFBA-2	47/2 47/47	M R	ASTLLKASPVLDKSEWVKGSVLFRR AASSYADELVKTAKTIASPRGR	Acetyl Dimethyl	0/1 0/26
		47/48	A	AASSYADELVKTAKTIASPRGR	Dimethyl	85/119
Scenario C (***)	Protein name	N-terminus (a)	Prev. AA	N-terminal peptide	TAILS	SPC
AT3G60750	Transketolase 1	66/66 66/67 66/67 66/68	R A A A	AAAVETVEPTDSSIVDKSVNSIR AAVETVEPTDSSIVDK AAVETVEPTDSSIVDKSVNSIR AVETVEPTDSSIVDKSVNSIR	Dimethyl Acetyl Dimethyl Dimethyl	37 4 215 74
AT4G24280	cpHSP70-1	93/75 93/78	R N	VNEKVVIGIDLGTNSAVVAAMEGGKPTVITNAEGQR EKVVGIDLGTNSAVVAAMEGGKPTVITNAEGQR	Dimethyl Dimethyl	16 71
AT4G24830	Argirinosuccinate synthase	74/74 74/75 74/75	R A A	AVLSGDGITALTDSKEAGLR VLSGDGITALTDSKEAGLR VLSGDGITALTDSKEAGLR	Dimethyl Acetyl Dimethyl	17 22 6

* **Scenario A.** Proteins for which only a single Nt peptide was identified with multiple MS/MS spectra

** **Scenario B.** Proteins for which unprocessed CTPs were detected starting with their penultimate residues (nuclear-encoded precursors), as well the Nt of the mature chloroplast-localized forms. The unprocessed forms were only identified in the total soluble extracts (leaf) and not in stromal extracts.

*** **Scenario C.** Proteins with variable Nt resulting from 'slippy' SPP cleavage specificity and/or from additional Nt maturation steps by amino peptidases following initial CTP cleavage by SPP. Both NAA and free N-alpha amino (dimethylated) residues were detected for Val and Ala N-terminated peptides.

The Nt amino acid frequency for all mature n-encoded chloroplast Nti was calculated (Figure 2.2A; Supplemental Table 2.4). This demonstrates that Ala and Ser are heavily favored as Nt residues, followed by Val and Thr, whereas the remaining residues are underrepresented (in particular Asp, Tyr and Trp each only once) or not observed at all (Pro and His) (see legend Figure 2.2A). The ratio between NAA and unmodified (but dimethylated in the TAILS procedure) Nti can be approximated based on matched MS/MS spectra, in particular if a relative high total number of MS/MS sequences (*e.g.* >50) are obtained. The NAA rates for the high frequency residues Val, Thr, Ala and Ser were respectively 54%, 47%, 21% and 19%. The few cases of Trp, Arg, Ile and Pro were mostly in NAA form, whereas NAA was not observed for Tyr, Leu, Phe, Asp, Cys.

Because many proteins were present as different Nt proteoforms, we ranked Nt peptides for each protein such that a single representative N-terminus for each protein could be assigned. This ranking was based on the number of observed spectra, the proximity to the predicted cTP cleavage site (ChloroP) and, if available, previously published N-terminal sequence data (See Supplemental Table 2.4 for N-terminal ranks and for a description of ranking process). Importantly, selecting a single ‘best ranked’ Nt for each protein hardly influenced the frequency distribution of the aa at the N-terminus (Figure 2.2B). Moreover, Ile, Leu, Try, Tyr and Asp were each found only once as best ranked Nt residue, whereas Pro, His and Phe were not observed at all. Interestingly, N-terminal Arg (3x) and Trp (1) were only found in their NAA form, perhaps suggesting that NAA is needed for stabilization.

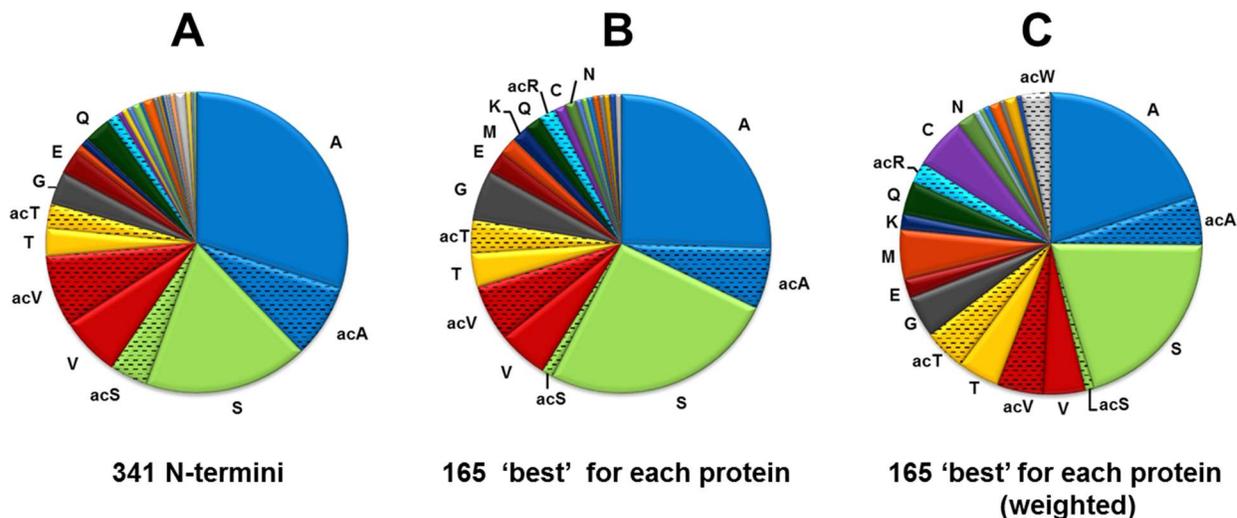


Figure 2.2. Nt amino acid frequency for stromal-exposed n-encoded chloroplast proteins.

(A) All detected stromal Nti (341), excluding unprocessed proteins and obvious breakdown products (see Supplemental Table 2.4). This shows that Ala and Ser are heavily favored as Nt residues, followed by Val and Thr, while 14 residues were underrepresented (Gly 14x, Gln 14x, Glu 10x, Ile 6x, Arg 5x, Lys 5x, Met 4x, Asn 3x, Leu 3x, Cys 2x, Phe 2x, Trp 1x, Tyr 1x, Asp 1x) or not observed (Pro and His). A significant portion of these highly favored residues were acetylated (Val 54%, Thr 47%, Ala 21%, Ser 19%), whereas the acetylation rate for other residues was either 0% (Tyr, Leu, Phe, Asp, Cys) or 100% (Trp) (acetylation is indicated as ‘ac’).

(B) Single highest ranked N-terminus per protein (165), excluding Nti with less than two SPC. Selecting a single ‘best’ or highest ranked N-terminus for each protein (see Methods) did hardly influence the Nt amino acid frequency, except that it slightly decreased the dominance of Ala, increased Ser and reduced acetylated Ser. Less frequent residues were Gly (9x), Glu (5x), Gln (4x), Lys (4x), Arg (3x), Met (3x), Asn (2x), Cys (2x), Leu (1x), Ile (1x), Trp (1x), Tyr (1x), Asp (1x), whereas Phe, Pro, His were not observed.

(C) Single highest ranked Nti as in panel (B) but normalized (weighted) to the frequency of each amino acid in the known (from PPDB; 1575 proteins) n-encoded plastid proteome with predicted cTPs removed.

Some amino acids are far more frequent in the known chloroplast proteome than others (Leu is the most frequent (~9.5%), followed by Ala, Ser and Val (each ~7.7%)), whereas His, Cys and Trp are the least abundant (1-2%), possibly biasing the Nt aa frequencies. Therefore, the frequencies of these 165 best ranked Nti were normalized to the natural frequency of each amino acid in the known n-encoded plastid proteome (1575 proteins – see Material and Methods) with predicted cTPs removed (Figure 2.2C). This showed again that Ala, Ser, and to a lesser extent Val and Thr (in NAA and free form), are still strongly favored, whereas Met and Cys are more prominent than before weighing (compare to Figure 2.2A) and Leu is clearly avoided.

2.3.4 *Physiological N-terminal methylation*

The four paralogs of RUBISCO small subunit (RBCS) were the only observed n-encoded mature proteins that started with an Nt Met residue. This Nt Met of RBCS has previously been shown to be methylated at its N-terminus through the activity of Rubisco methyltransferase (Houtz et al., 2008). The dimethylation reaction used in the TAILS method would mask this physiological (mono)methylation since it generates a dimethylated amino-terminus. To assess if *in vivo* N-terminal methylation occurs in chloroplasts for other proteins, TAILS experiments were also performed with deuteriated formaldehyde (CD₂O) instead of formaldehyde (CH₂O), which allowed us to differentiate between natural methylation and methylation by formaldehyde. Indeed, we observed that RBCS4 (Supplemental Figure 2.2) and RBCS1b (not shown) accumulated with *in vivo* monomethylated Nti Met. No other convincing cases for Nt-methylation were detected, which is perhaps not surprising because we observed so few

mature n-encoded proteins (only RBCS family members) that start with a Met residue. The lack of observed Nt Met of n-encoded stromal exposed proteome suggest very efficient NME. The lack of NME for just RBCS is likely due the presence of a bulky residue (Lys) immediately after the Met in case of RBCS. It should be noted that Lys methylation has been observed for several Arabidopsis chloroplast proteins downstream of their mature Nti (Zybailov et al., 2009; Alban et al., 2014). Lys-14 of RBCL has been shown in pea to be (tri)methylated (Houtz et al., 2008). However, we found no evidence for such a modification in Arabidopsis (the detected Nt peptide of RBCL is long enough to include this Lys - SPQTETKASVGFKAGVKEY), in agreement with a recent study indicating that Arabidopsis RBCL is not naturally (tri)methylated at this position (Mininno et al., 2012).

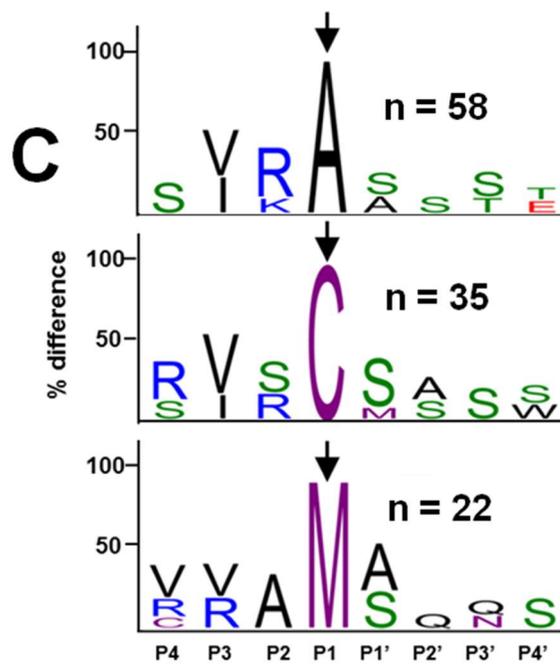
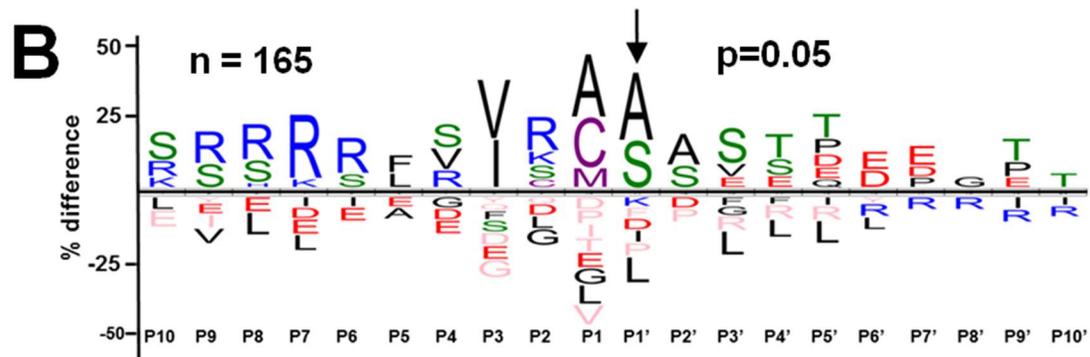
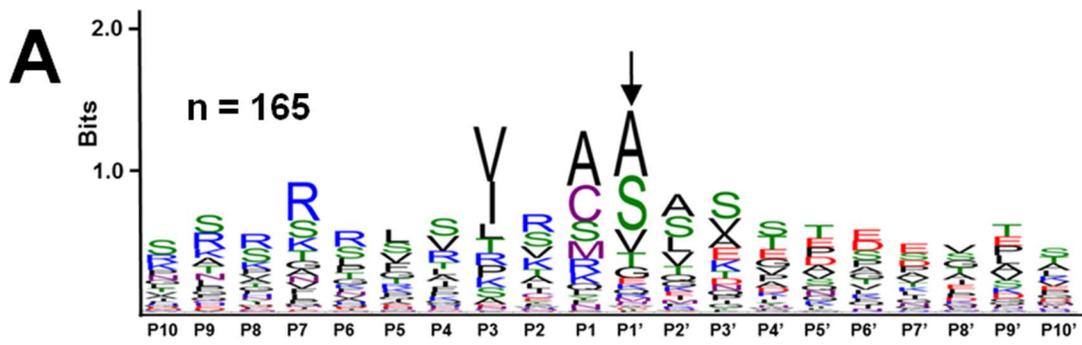
2.3.5 Conservation around the cTP cleavage site

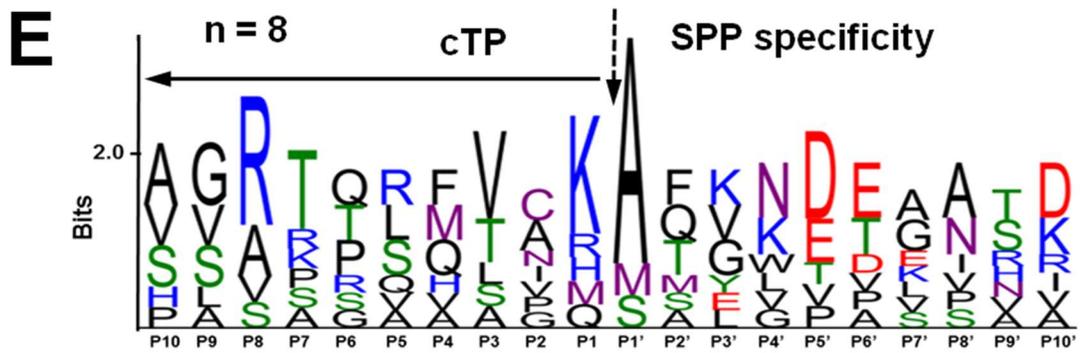
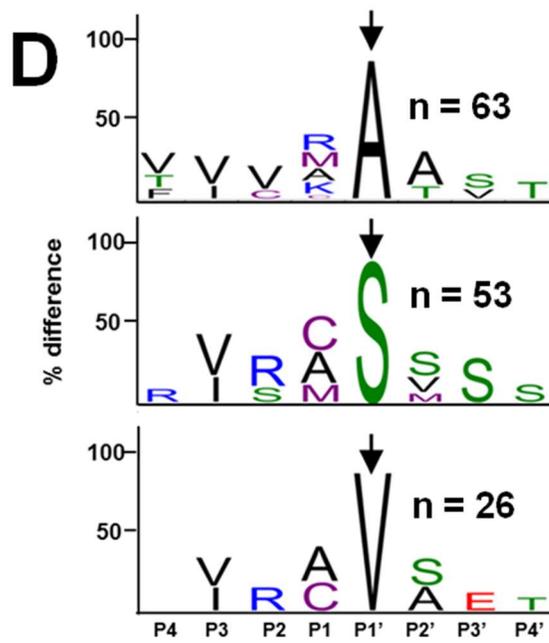
In an effort to obtain more insight into the relationship between cTP cleavage and the ultimate Nt residue/sequence, we generated a sequence logo of residues surrounding the observed mature Nti using the ‘best ranked’ N-terminus for each protein (as defined above) (Figure 2.3A). This data set is much larger than the previously published datasets of experimental chloroplast protein Nti. Furthermore these previous data sets were necessarily enriched for NAA Nti since only these could be confidently identified as *in vivo* Nti (in absence of N-terminal labeling) (Zybailov et al., 2008); the dimethyl labeling in the TAILS workflow allowed us to avoid this bias. The sequence logo shows only a weak consensus around the observed N-terminus (Figure 2.3A); however a (still weak) consensus motif was more clearly

Figure 2.3. Analysis of amino acid conservation around experimentally determined Nti for n-encoded stromal-exposed proteins and comparison to Nti generated by in vitro SPP cleavage assays reported in the literature. As per consensus, P1' is the observed Nt residue and P1 the residue immediate upstream of P1'. Solid arrows indicate the experimentally determined Nt residue. For plots A-D, the best ranked Nti of 165 plastid-proteins with n-encoded stromal-exposed Nti were used. In all plots, proteins were aligned round the experimentally determined Nt residue (P1'). Color coding for residues: blue – basic residues (R, K, H); red – acidic residues (D,E); black – apolar residues (A, V, L, I, P, F, W,G); purple - reactive residues (M, C); green – uncharged, polar residues (S, T, Y, Q,N).

(A) Sequence Logo of the 165 stromal-exposed proteins shows a weak motif around the mature Nt. The conservation level of aa in this sequence alignment is represented as vertical stacks of the aa symbols; the stack height reflects the level of conservation. **B,C,D,** IceLogo plots of the stromal-exposed proteins in which the aa frequency is normalized (weighted) against the total amino acid frequency of the n-encoded chloroplast proteome (from PPDB; 1575 proteins). AA residues significantly enriched are shown above the x-axis, whereas those underrepresented are shown below the x-axis. Residue below the x-axis colored in pink were entirely absent in this position in the experimental sequences. **(B)** IceLogo of the 165 n-encoded stromal-exposed proteins (at $p=0.05$). **(C)** IceLogo plots (at $p=0.01$) for n-encoded stromal-exposed proteins for which the residue immediately upstream of the experimentally determined Nti (P1) is an Ala (58 sequences), Cys (35 sequences), or Met (22 sequences). **(D)** IceLogo plots (at $p=0.01$) for n-encoded stromal-exposed proteins for which the experimentally determined Nti (P1') is an Ala (63 sequences), a Ser (53 sequences) or Val (26 sequences).

(E) Sequence logo for 8 sequences shown to be cleaved in vitro by SPP (7 using *P. sativum* SPP, 1 using *C. reinhardtii* SPP) with SPP either purified from chloroplasts or recombinant SPP expressed in *E. coli* and immobilized on beads via a Nt biotin tag. Substrates are from a range of organisms (wheat, tomato, spinach, pea, *C. reinhardtii*, *Arabidopsis*, *S. pratensis* and the fungus *N. crassa*).





visualized using an iceLogo (Colaert et al., 2009) (Figure 2.3B). The IceLogo involves weighing against the total amino acid frequency of the chloroplast proteome, thereby visualizing significantly over- and under-represented amino acids (Figure 2.3B). Cys was highly enriched in the P1 position, but not anywhere else. Furthermore, this showed that acidic residues were disfavored within the cTP, whereas basic residues (in particular Arg) were enriched in the cTP, but that Arg was avoided within the first 10 residues of the mature protein (Figure 2.3B). Small uncharged and often hydrophilic residues were favored within the first four residues of the observed proteins (P1' to P4') whereas Leu was under-represented in these positions. Cys, Met and Ala were strongly enriched immediately upstream of the N-terminus (P1 position). Both Cys and Met are easily oxidized and oxidized cysteine has been shown to act as a degradation signal outside of the plastid, leading to protein degradation by the proteasome (Graciet et al., 2010; Graciet and Wellmer, 2010).

The lack of a visible consensus cleavage site motif despite this large and high-quality dataset, suggests that SPP does not have a strict consensus cleavage motif for imported plastid proteins. Alternatively this lack of observed motif might indicate the activity of subsequent maturation steps by additional peptidases, thereby masking the SPP cleavage site motif. Indeed, chloroplasts do possess a significant number of mostly uncharacterized aminopeptidases (Walling, 2006; van Wijk, 2015). For instance, the observation that Cys, Met and Ala were strongly enriched immediately upstream of the N-terminus (P1 position) may be explained by the activity of amino-peptidases that specifically remove these unstable residues following SPP processing.

To try and distinguish between the various scenarios and possibly reveal hidden motifs, subsets of sequences with either highly conserved residues at P1 and P1' positions were analyzed separately by iceLogos (Figure 2.3C, D). Cys in the P1 position was preferentially flanked (in P2 and P1') by Ser and to a lesser extent R at P2, whereas Met in the P1 position was flanked by Ala (Figure 2.3C). Subsets of Nti with Ala, Ser or Val at the N-terminus (in P1') (Figure 2.3D) reveal that Ser and Val N-terminated proteins are mostly produced by cleavage after Cys or Ala, whereas Ala N-terminated proteins are preceded by Arg, Lys, Ala or Met. Furthermore, it can be observed that for both P1-Met (Figure 2.3C) and P1'-Ala (Figure 2.3D), Val/Ile conservation at P3 breaks down, which could be indicative of sequential processing. These comparisons suggests that a possible cTP cleavage motif is obscured by additional processing steps.

To better understand SPP cleavage and possible subsequent maturation by other peptidases, we collected all available direct evidence for SPP cleavage site specificity (Supplemental Table 2.6). Such specificity has been determined for recombinant proteins using either recombinant SPP from pea (Richter and Lamppa, 2002) or using semi-purified SPP from isolated chloroplasts of pea or *C. reinhardtii* (Supplemental Table 2.6) (Richter et al., 2005). It should be noted that these substrates are from 5 different plant species. Some of the substrates lack a cTP and seem less relevant to test the specificity of a processing peptidase (See Supplemental Table 2.9). Using only the eight *bona fide* intra-plastid proteins, we then generated a sequence logo of residues around the observed N-terminus (Figure 2.3E). This suggests cleavage primarily after basic residues (in particular Lys, but also Arg, His), and upstream of Ala (Figure 2.3E), which matches well with the top panel in Figure 2.3D. Determination of SPP cleavage

specificity using a wider variety of substrates from Arabidopsis, as well as analysis of putative chloroplast aminopeptidases, are needed to improve our understanding of plastid protein maturation.

2.3.6 The N-terminome of plastid-encoded proteins

The maturation process of p-encoded proteins (Figure 2.1B) is very different from n-encoded chloroplast proteins (Figure 2.1A). Moreover, the Nti of nascent p-encoded proteins are likely protected by proteins interacting with the 70S ribosome near the exit gate, such as trigger factor. Furthermore, N-terminal deformylation, NME and NAA are likely co-translational processes for p-encoded proteins (Giglione et al., 2009; Preissler and Deuerling, 2012; Sandikci et al., 2013; Giglione et al., 2014). Hence, the Nt sensitivity to proteolytic activity may differ between plastid- and n-encoded chloroplast proteins. P-encoded proteins are synthesized with an N-terminal methionine and a subset undergoes NME. In general, the penultimate position (P1') is the major determinant for NME and cleavage occurs if the side chain is small (Ala, Cys, Pro, Ser, Thr, Gly and Val) (Giglione et al., 2004). Whereas p-encoded proteins generally follow this rule, there are a few outliers and several other proteins undergo additional maturation steps (Zybailov et al., 2008; Zybailov et al., 2009; Bienvenut et al., 2012).

There are 88 proteins encoded by the plastid genome in Arabidopsis, 65 of these proteins have Nti in the stroma, whereas the other remaining proteins have their Nti exposed to the thylakoid lumen or their topology is currently not clear to us (Supplemental Table 2.7). Frequency analysis of the penultimate residues for Arabidopsis p-encoded proteins with stromal exposed Nti showed 16 possible residues

(absent are bulky His, Tyr, Trp, and Phe) (Figure 2.4A). Applying the general NME rule (Giglione et al., 2004) to these stromal exposed Nti results in a simpler amino acid distribution of chloroplast Nt residues with just 8 possible amino acids (Figure 2.4B).

We then combined our TAILS results with previous in-house MS/MS data for other Arabidopsis chloroplast proteome experiments in PPDB (*e.g.* (Zybailov et al., 2008; Zybailov et al., 2009; Kim et al., 2013; Lundquist et al., 2013; Nishimura et al., 2013), as well as information from (Giglione et al., 2004) which was mostly based on Nt Edman sequencing data from various plant species. The Edman sequencing method does not yield NAA state because these Nti prevent Edman chemistry ('blocked' Nti). The information from these other plant species was 'projected' onto Arabidopsis homologs if the Nti were identical. The distribution of Nt residues is summarized in Figure 2.4C and Supplemental Table 2.7. We then compared the pie chart in panel 4B (predicted after NME) with panel 4C (experimental observations). This shows the presence of experimental Nti starting with Ile and Arg, which must have been due to unusual NME activity, namely that Met was removed to expose Ile (PsaA and RPS15) or Arg (CF1 β); these are bulky residues that typically would prevent NME activity. It should be noted that in all three cases these Nt residues were acetylated, again suggesting that NAA is required for stabilization. NME did not occur for the three other observed proteins with Ile in the penultimate position (PetG, RPL14 and NDH-A), nor was Met removed for the only other observed case with Arg in the penultimate position (PsaJ). The Nti for p-encoded YCF1.2 (TIC214) (Kikuchi et al., 2013), RBCL

Figure 2.4. Nt amino acid frequency for stromal-exposed p-encoded proteins and comparison to all known luminal-exposed Nti (both p- and n-encoded).

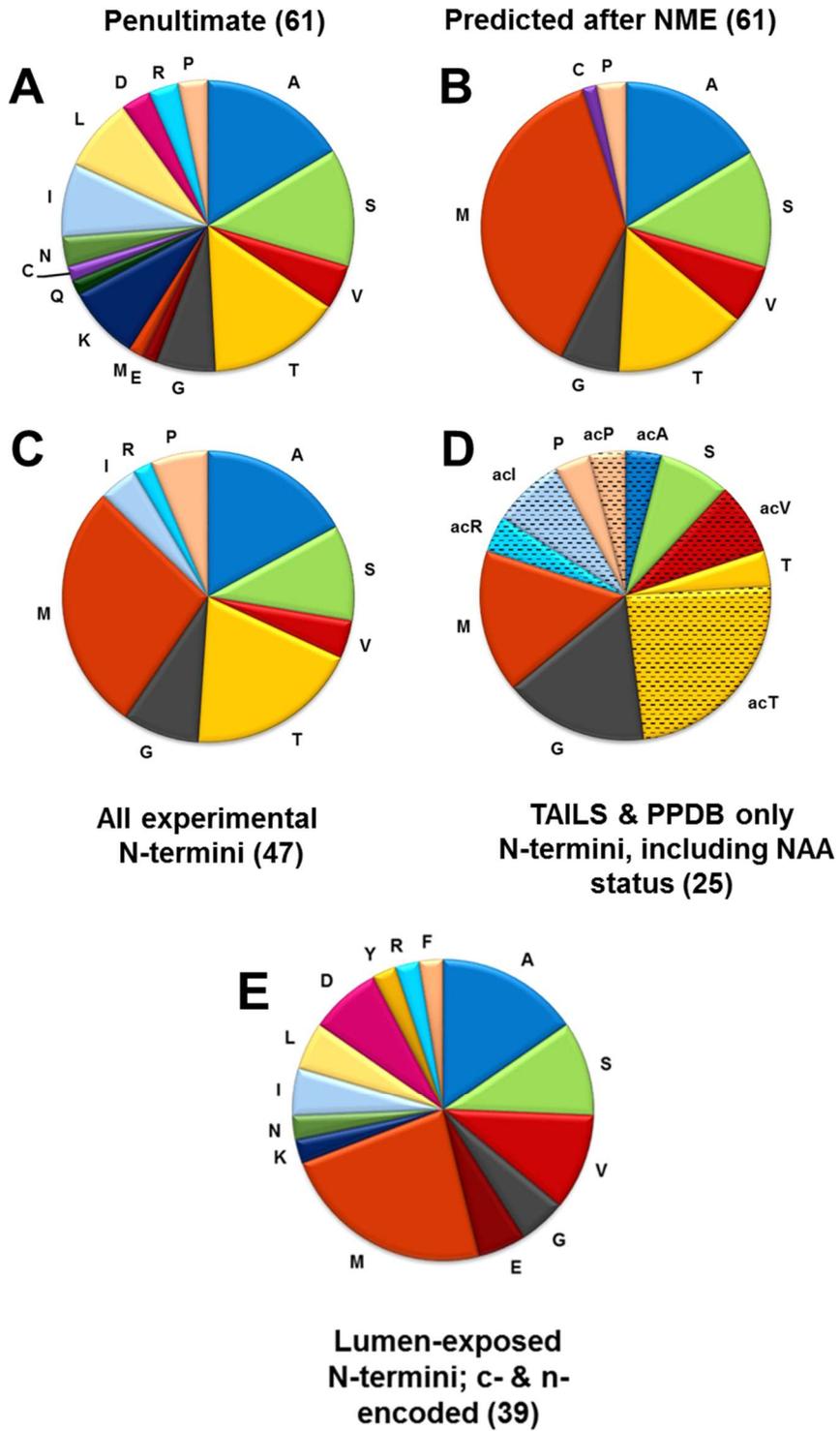
(A) The penultimate residues (i.e. residues immediately down-stream of the initiating Met) of 65 p-encoded proteins for which the N-terminus is facing the stroma. This sequence information is derived from the protein sequences listed in TAIR (<https://www.arabidopsis.org/>). Within this group there are 3 sets of identical homologues (ribosomal proteins S7A,B, ribosomal proteins S12A,B,C and a full length YCF1.2 protein and a truncated form – see Supplemental Table 2.7 for details). Rather than including each of these homologs, we counted each set only once, thus resulting into 61 Nti.

(B) The predicted Nt residues of mature proteins after application of the general NME rule for the p-encoded proteins in panel A.

(C) Experimentally determined Nt residues for p-encoded proteins for which the N-terminus is facing the stroma (total 47 proteins). Experimental evidence was obtained from the TAILS experiments described in this study, from semi-tryptic or NAA Nti detected in (Zybailov et al., 2008; Zybailov et al., 2009; Bienvenut et al., 2012) and additional data from in-house experiments in PPDB. Also included is information from (Gigliione et al., 2004) which was mostly based on Nt Edman sequencing data from various plant species. We note that Edman sequencing cannot sequence proteins for which the Nt is NAA; these modified Nti are ‘blocked’ preventing Edman chemistry). The experimental Nt information from these other plant species was ‘projected’ onto Arabidopsis homologs if the Nti were identical.

(D) Experimentally determined Nt residues for 25 p-encoded proteins for which the N-terminus is facing the stroma as determined by TAILS and in-house experiments in PPDB. This is a subset of the proteins in panel C.

(E) Experimentally determined Nt residues for 39 p- and n-encoded proteins for which the N-terminus is facing the thylakoid lumen. Experimental evidence was obtained from the TAILS experiments, previous publications (Zybailov et al., 2008; Zybailov et al., 2009) and additional data in PPDB (see Supplemental Table 2.5 for details).



and CP43 did not start with Met nor with the penultimate residue, indicating that these Nti must have been generated by additional peptidase activity; however the responsible peptidases are unknown. For RBCL, the N-terminus starts with the 3rd residue (Pro) (observed by 537 MSMS spectra) and it was always in NAA form; this is in agreement with previous observations (Zybailov et al., 2008). The unprocessed YCF1.2 protein is predicted to start with f-Met-Met, but both Met residues were removed, resulting into an Nt Val. In the case of CP43, 12 amino acids were removed, exposing an N-terminal Thr, which is known to undergo NAA and reversible phosphorylation (Vener et al., 2001; van Wijk et al., 2014). We did not observe this phosphorylated form because we did not take any precautions to prevent dephosphorylation (i.e. by addition of phosphatase inhibitors) and/or because we did not enrich for phosphopeptides, which is typically needed to observe the phosphorylated forms.

Finally, Figure 2.4D shows the extent of NAA for experimental observed Nti of the stromal-faced p-encoded proteins determined only by TAILS or from previous in-house experiments listed in PPDB (25 proteins in total). This shows that Arg, Ile, Ala and Val are always observed in their NAA form, but in case of RPS15, Ile was also observed unmodified.

2.3.7 The thylakoid-lumen exposed Nti show a wide distribution of amino acids

The thylakoid lumen has its own (limited) set of proteases. We assembled all available information for p- and n-encoded luminal exposed Nti (Figure 2.4E; Supplemental Table 2.5). In addition to the abundant Ala, Ser, Val and Met, this shows the presence of residues essentially absent at the Nti of stromal-exposed proteins.

Examples are Tyr (Y), Asp (D), Glu (E), Leu (L), indicating a far greater Nt flexibility, likely reflecting a lack of Nt-driven instability.

2.3.8 *Accumulation of proteolytic products*

129 Nti are likely breakdown products of chloroplast proteins (Supplemental Table 2.8). Interestingly, none are NAA, suggesting a short half-life and/or generation of these proteolytic products after stromal isolation and assuming that N- α acetylases are not very active at that point. About 60% of these Nti were from the very abundant RBCL, RBCS and RCA, which is not surprising given that these are among the most abundant proteins. However, only few degradation products were detected for several other highly abundant enzymes, such as transketolase, glutamate-ammonia ligase, CPN21. Perhaps this indicates that the RBCS/L and RCA have a shorter life-time than other abundant stromal proteins (see also (Recuenco-Munoz et al., 2015)). Importantly, we do note that the Nti of mature proteins are always far more frequent than the Nti of the fragments; an example is shown for abundant stromal proteins isomerase ROC4, transketolase 1 (TLK1), phosphoribulose kinase (PRK2) and Sedoheptulose Fructose biphosphatase (SFBA) (Supplemental Figure 2.3). Analysis of the breakdown products revealed a strong preference for cleavage after Arg and to a lesser extent before Thr.

2.3.9 *Correlation with other large scale N-terminome studies*

In the last few months, Arabidopsis studies were published that employed TAILS or (a variant of) COFRADIC to study protein N-termini in roots (Zhang et al., 2015), mitochondria (Carrie et al., 2015) and leaves of wild-type and a chloroplast import

mutant (Kohler et al., 2015). Additionally there was a large scale study of NAA leaf proteins (Bienvenut 2012) and an assessment of mitochondrial protein N-termini based on ‘classical’ proteomics (Huang 2009). None of these studies shared the objectives of the current study; nevertheless these studies are a good opportunity to probe the consistency with the data presented here. To that end we systematically cross-checked the observed Nti for stroma-exposed Nti, as well as Nti of mitochondrial proteins (Supplemental Table 2.9). Of the Nti of the 206 stromal-exposed mature proteins identified in our study, 104 matched exactly with those found by others. The observed start position for 16 other proteins observed in our dataset was within 5 residues of that found by others. Other chloroplast proteins in our dataset were either not observed by other studies or they were detected with an N-terminus too far down-stream to represent the *bona fide* mature N-terminus; examples are RBCS and related proteins RCA and CP12, the Calvin cycle enzymes GAP-A/B, and SFBA, and several enzymes in the methylerythritol 4-phosphate pathway.

We also detected 19 mitochondrial and 17 peroxisomal proteins (Supplemental Table 2.3). More than half of the mitochondrial Nti started with Ser and the Nti were typically preceded by Met, Ser, Leu, Phe, and Tyr, in good agreement with previously described mitochondrial presequence cleavage motifs (Huang et al., 2009; Carrie et al., 2015) (Supplemental Table 2.9). Peroxisomal proteins are targeted to the matrix by a non-cleavable tripeptide at the extreme C-terminus (PTS1) or a cleavable nanopeptide at the N-terminus (PTS2) (Hu et al., 2012). Of the 17 detected peroxisomal proteins all except four were NAA and started either at the initiating Met or Ala in the 2nd position, presumably because they are targeted through a PTS2 signal.

2.4 DISCUSSION

The objectives of this study were to determine the Nti of the stromal-exposed chloroplast proteome and develop a testable model for Nt processing, maturation and stability. Through systematic TAILS analysis of soluble proteins from total leaf extracts and isolated chloroplasts, we obtained nearly 14.000 MS/MS spectra matching to protein Nti. Following condensation and curation of this data set, as well as annotation of subcellular localization, we then obtained a comprehensive set of chloroplast Nti. Comparison of this dataset to previously published information for individually studied proteins and other N-terminome studies (see below) showed that our TAILS workflow provided reliable and physiologically relevant information. The parallel acquisition of N-terminomes of total leaf extract and stromal extracts from isolated chloroplasts was important for recognition of extra-plastidic proteins and chloroplast precursor proteins. This also confirmed that accumulation of unprocessed chloroplast proteins (or cleaved cTPs) within the chloroplast is exceedingly rare, indicating a high efficiency of cTP cleavage and subsequent degradation of cleaved cTPs within the chloroplast, in agreement with (Richter and Lamppa, 1999) and others.

2.4.1 *Working hypotheses for Nt maturation of n-encoded proteins*

Based on the analysis of Nt amino acid frequency, sequence logos and iceLogos, as well as published information (*e.g.* (Richter and Lamppa, 1998; Richter et al., 2005; Zybaïlov et al., 2008; Bienvenut et al., 2012)), we formulated a working model for Nt maturation of n-encoded proteins (Figure 2.5A) (for a broader discussion and many

cited references see (van Wijk, 2015)). Upon import into the chloroplast, the cTP is cleaved by SPP. This cleavage could either be very precise at a single position (a specific peptidyl bond) (Figure 2.5A – upper left portion), or less precise with cleavages occurring at closely spaced, multiple positions, depending on the residues neighboring the cleavage site. Additional peptidases will subsequently perhaps remove one, or in some cases two or three, residue(s) from the N-terminus; this likely depends on the Nt residue and the immediate down-stream sequence, as well as protein fold (accessibility of the N-terminus). Seven stromal amino peptidases (APs) were identified with high confidence and their relative abundance quantified in chloroplasts of Arabidopsis (Zybailov et al., 2008). These include the higher abundance Leu-AP, Glu-AP and Amino peptidase - P, as well as four lower abundance peptidases (Met-AP1B, Gly-AP, Pro-AP, Ser-AP). Whereas the substrate specificity of these peptidases has generally not been characterized, they are strong candidates for performing the proposed role in Nt maturation (Figure 2.5A). The combination of single and multiple SPP cleavages and activity of multiple amino-peptidases provides the most flexible scenario to arrive at the highly restricted N-terminome (*i.e.* high prevalence of Ala, Val, Thr, Ser) and best accommodates all observations. Finally, a limited number of proteins likely undergo an additional down-stream cleavage, as exemplified by p-encoded CP43 which accumulates with an NAA (and reversibly phosphorylated) Thr13; we suggest that a specific peptidase (as yet unidentified) generates the N-terminus of this abundant (and essential for photosynthesis) PSII core protein.

2.4.2 *Classification of chloroplast stroma-exposed Nt residues and examples*

Figure 2.5B summarizes the observed frequencies of each amino acid in the stromal-exposed Nt position for the n- and p-encoded proteome. The most frequent, and perhaps the most stable Nti, start with the small polar (Ser, Thr) or apolar (Ala, Val, Gly) residues; together these represent ~75-80% of all Nti. Except for Gly, a significant percentage of these residues are NAA; however, the general function of NAA is poorly understood (Jones and O'Connor, 2011; Hollebeke et al., 2012; Starheim et al., 2012), but can influence protein stability as for example shown in Arabidopsis for a nod-like receptor (Xu et al., 2015). It was recently shown that reduced NAA rates trigger a drought response in Arabidopsis (Linster et al., 2015). In the case of p-encoded proteins, Met has a high frequency in the Nt position, dictated by the penultimate residue and the NME. In selected cases, such as the three PSII core proteins D1, D2 and CP43, the N-terminus plays an active regulatory role through reversible phosphorylation of the (stable) NAA Thr (Figure 2.5B) (see *e.g.* (Vener, 2007; Rokka et al., 2011)).

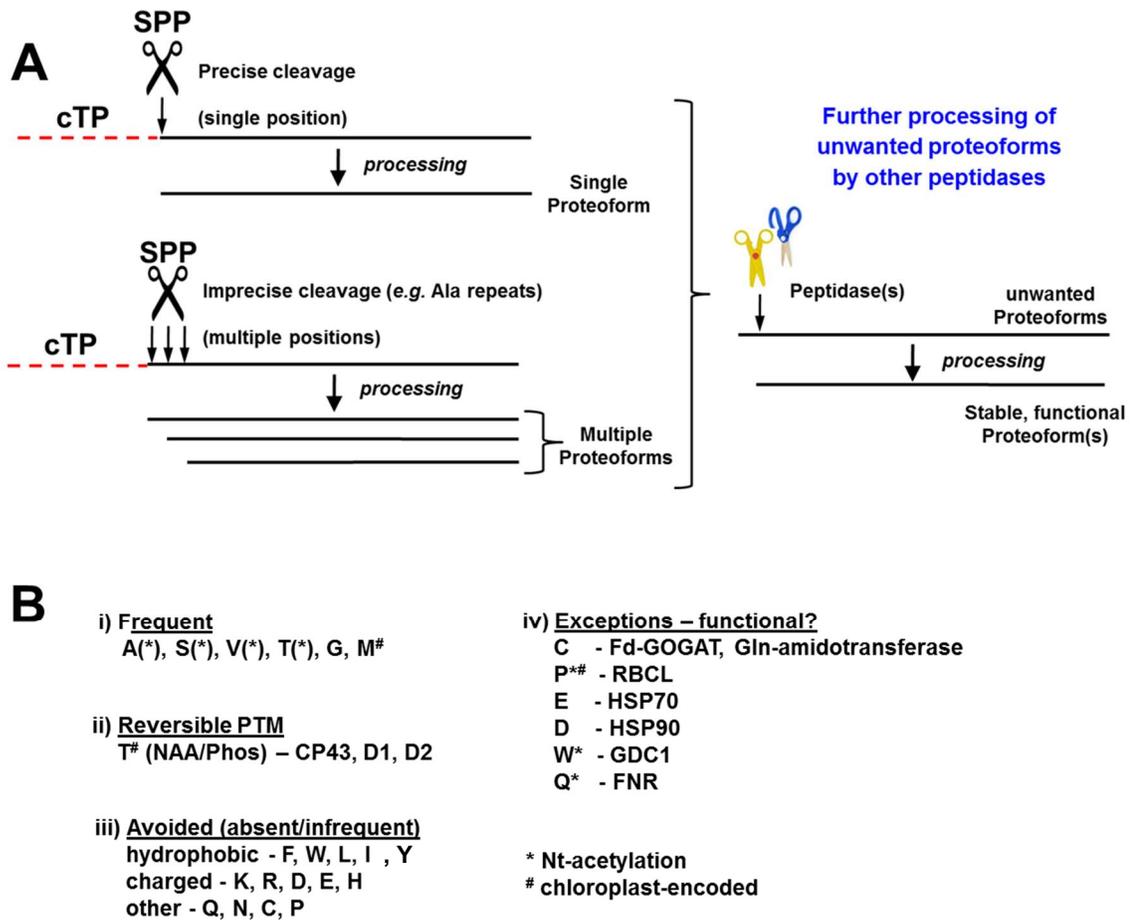


Figure 2.5. Working model for Nt maturation of n-encoded proteins and the classification of different types of Nti.

(A) Model for generation of mature and stable Nti of n-encoded chloroplast proteins. Upon chloroplast import, the cTP of precursor proteins are either cleaved at a specific single site or cleaved at closely spaced multiple positions. Proteins with unwanted and/or unstable Nti are further processed by one or more stromal amino peptidases to stabilize the proteins.

(B) Classification of different types of chloroplast stroma-exposed Nti and examples. We distinguish three types of amino acids: i) Amino acids that are very frequent in the Nt position and that are presumable very stable in the chloroplast stroma. ii) Nti with reversible PTM and that play a functional role, iii) Amino acids that are not or rarely observed, and likely result in destabilization of proteins in the chloroplast when these Nti are exposed to the stroma. Group iv shows examples of proteins that were observed with rare amino acids at the Nt position; these are discussed in the text

Whereas just six residues occupied most of the stromal-exposed Nti, other amino acids were never observed in the Nt position (His, Phe) or observed in just one or a few cases (and sometimes only in NAA form, *e.g.* Trp) (Figure 2.5B). We discuss a number of such cases:

Redox active Cys – Cysteines are redox active and the thiol often forms inter- or intramolecular disulfide bonds, participates in enzymatic reactions, and undergo many post-translational modifications. We observed two cases of Nt Cys residues; these are for Fd-GOGAT (AT5G04140) and GLN PHOSPHORIBOSYL PYROPHOSPHATE AMIDOTRANSFERASE 2 (AT4G34740). Surprisingly, both mature Nti start with ‘CGV’, as well as unusual acidic stretches upstream of the cTP cleavage site, even if these proteins are otherwise completely unrelated and have distinct functions. In both cases, these ‘CGV’ Nt sequences are conserved in plants, algae and even cyanobacteria, suggesting that these play a specific, but as yet unknown function. Furthermore, for both cases these proteins were only detected with these specific Cys Nt-proteofom further suggesting that these Nt Cys residues play a functional role.

Aromatic amino acids Tyr, Trp, His and Phe. These residues are destabilizing N-terminal residues in prokaryotes (Tyr, Trp and Phe) where they are likely targets of the ClpAPS protease system, and in the cytosol of eukaryotes (Tyr, Trp, His and Phe) where they are targets of the proteasome (Dougan et al., 2012; Tasaki et al., 2012; Gibbs et al., 2014). His and Phe are absent as Nt residues in chloroplast stroma, whereas Tyr and Trp were each only observed once. This Trp is the N-terminus of GDC1 (AT1G50900) and

was observed in 33 out of 35 times (always NAA); manual inspection of several of the underlying spectra confirmed the assignment. GDC1 (also named LTD) is a mostly stroma-localized protein involved in the sorting of members of the LHC protein family and interacts with the SRP particle in the stroma (Cui et al., 2011; Ouyang et al., 2011). The Trp is largely conserved across land plants and is typically preceded by a Cys residue. The significance of this NAA Trp remains to be determined. It is not known if these aromatic residues confer instability to proteins in the chloroplast.

Acidic residues Glu and Asp. The two abundant stromal chaperones cpHSP70-1 (AT4G24280) and HSP90 (AT2G04030) both start with acidic residues (Glu and Asp, respectively). The N-terminus of HSP90 is generated by cleavage after Cys. In case of HSP90 there was one other Nt-proteoform, starting with an Ala one residue downstream of the Asp; however, it was only observed twice compare to 29 times for the acidic Nt-proteoform. In case of HSP70, the acidic Nt-proteoform was observed 71 out of 87 times. These essential chaperones perhaps require these unusual acidic Nti to interact with their targets or partners. Another Nt Glu was found for ribosomal protein RPL13, also generated by cleavage after Cys. It seems quite logical that a stromal amino peptidase might exist that removes these Cys residues.

Proline We observed Nt Pro (an unreactive amino acid) with high number of MS/MS spectra for the abundant RBCL (the Pro was always NAA), in agreement with a previous study in spinach (Mulligan et al., 1988). Because Pro is the 3rd residue of this p-encoded protein after Met-Ser, it was likely generated by NME, followed by cleavage of the Ser

by a different peptidase. We predict that a specific, as yet unknown, peptidase evolved to carry out this specific cleavage of RBCL.

2.4.3 *Functional significance of Nt maturation and Nt proteoforms*

The N-terminome analysis presented in this study clearly established that many chloroplast proteins are represented by more than one Nt proteoform. Based on the number of matched MS/MS spectra presented in Supplemental Table 2.4, it is possible to calculate tentative abundance ratios between Nt proteoforms for these proteins. For example, in case of inorganic phosphatase (AT5G09650.1), observed by 185 MS/MS spectra, 5% started with Ser-Ala-Iso, 27% started with Ala-Iso and 68% started with the downstream Iso. In case of Enoyl-ACP reductase (AT2G05990.1) observed with 134 MS/MS spectra, two Nt-proteoforms were observed starting with Ala-Met-Ser (13%) or the downstream Ser (87%), but interestingly no Nt-proteoform starting with this Met was observed. Whereas the asymmetric distribution of Nt-proteoforms may relate to functional differences, very few studies exist that have looked at the significance of Nt proteoforms. One example of such studies is for thylakoid-associated FNR-1 (AT5G66190) and FNR-2 (AT1G20020) (Lehtimäki et al., 2014) that each have two Nt variants, with Nt sequences AQVT and AQIT being observed with and without the Ala for FNR1 and FNR2 respectively (Lehtimäki et al., 2014), in agreement with our TAILS data. However in this case, our TAILS data did not allow determination of the ratio between the two proteoforms for each FNR protein. Furthermore, NAA forms of each N-termini were observed and this modification appeared to be induced by light (Lehtimäki et al 2014). However, these Nt variations did not influence their association

with the thylakoid membrane and the exact physiological relevance remained unclear (Lehtimäki et al., 2014). A systematic analysis to investigate the functional significance of the phenomenon of multiple proteoforms is warranted.

2.4.4 Differences between plastid and n-encoded proteins

The maturation processes of nuclear- and p-encoded proteins differ from each other; it is strictly post-translational in case of n-encoded proteins but co-translational for many p-encoded proteins. Furthermore, following import and cTP cleavage, most n-encoded chloroplast proteins do not start with a Met, whereas p-encoded proteins do. Except for the much higher frequency of observed Nt Met (~30%), the p-encoded protein Nti are dominated by small, uncharged residues Ala, Ser, Thr, Val and Gly, similar as observed for n-encoded proteins. NAA rates appeared higher for p-encoded proteins than for n-encoded proteins perhaps because this occurs co-translationally, rather than post-translationally. Additionally, observation of a wide range of NAA amino acids (Ile, Pro, Arg, Trp, Gln) (but not Gly) suggest that more than one Nt acetylase (NAT) operates in the chloroplast (see (Starheim et al., 2012; Giglione et al., 2014)).

2.4.5 Comparison with protein maturation in mitochondria

Chloroplasts and mitochondria have many similarities with respect to protein biogenesis and more than 100 proteins are dually targeted to both organelles (Carrie and Small, 2013), including several proteases such as LON1 (Daras et al., 2014), PREP1 (Kmieć et al., 2014), FTSH11 (Urantowka et al., 2005) and OOP (Kmieć et al., 2013).

It is therefore likely that they also show similarities in protein maturation and in mechanisms of protease substrate recognition. Recent observations for plant mitochondrial N-terminal peptidase ICP55 (Carrie et al., 2015; Huang et al., 2015) and OCT1 (Carrie et al., 2015) support a similar model for maturation and stabilization of n-encoded as the proposed model for chloroplasts (Figure 2.5A). After cleavage of the N-terminal mTP by the general mitochondrial processing peptidase (MPP; the functional equivalent of chloroplast SPP), one (or sometimes 2) amino acid residue (in particular F, Y, L, I) is cleaved by ICP55 for high portion of mitochondrial proteins. The specificity of OCT1 was not very clear and it was suggested that OCT1 might act after assembly of proteins, rather than immediately following mTP cleavage (Carrie et al., 2015). Based on these observations for ICP55, it was suggested that removal of specific N-terminal residues (in particular F, Y and L) is needed to confer protein stability (Carrie et al., 2015; Huang et al., 2015). The Arabidopsis genome contains a homolog (AT4G29490) of mitochondrial ICP55, which is a candidate to play a similar function in chloroplasts.

2.4.6 *Nt residues, N-degron and the N-end rule*

The N-terminus of proteins has shown to be a major determinant of protein stability/half-life in many organisms. Early observations in yeast led to the formulation of the ‘N-end rule’ which states that certain amino acids, when exposed at the N-terminus of a protein, act as triggers for degradation (Bachmair et al., 1986). The N-end rule in prokaryotes is different than in eukaryotes in part because most prokaryotes do not have a proteasome and also lack ubiquitination (Tobias et al., 1991). In bacteria,

such N-end rule proteins are recognized by the adaptor ClpS, which delivers such proteins for unfolding and degradation to the Clp chaperone and protease system. Recent reviews summarize the history of these discoveries and the current understanding of the N-end rule pathway for prokaryotes and eukaryotes (Dougan et al., 2012; Tasaki et al., 2012; Gibbs et al., 2014). Whereas an N-end rule for chloroplasts/plastids in plants is not known, over-expression studies for plastid-encoded proteins have shown that the amino-acids at the N-terminus can greatly affect protein stability (Apel et al., 2010), and see (Gao et al., 2012) for discussion. Moreover, a plant homolog of ClpS was recently identified and characterized in chloroplasts of *Arabidopsis thaliana* (Nishimura et al., 2013). Recent data of the N-terminome of an *Arabidopsis* mitochondrial ICP55 null mutant indicated that ICP55 removes in particular the N-terminal residues Phe, Tyr and Leu. These residues are generally considered unstable residues and it was therefore suggested that plant mitochondria also utilize an N-end rule pathway (Carrie et al., 2015; Huang et al., 2015). Our study was in part motivated by the search for a possible N-end rule in the chloroplast stroma. From the observed amino acid frequencies in the stromal-exposed Nti, there appears indeed a strong overlap between residues avoided in chloroplast stromal-exposed Nti and the bacterial and mitochondrial primary N-end rules residues (Trp, Tyr, Phe, Leu). Secondary destabilizing residues Asp, Glu, Arg, Lys in prokaryotes also are among the low frequency amino acids in the Nt position. In contrast, Met, a secondary destabilizing residue in *E. coli* (but nevertheless quite frequent in prokaryotes (Bonissone et al., 2013)), is clearly a very frequent and likely stable residue for p-encoded proteins in chloroplast stroma. Secondary destabilizing residues only become destabilizing upon

transfer of an amino acid to the N-terminus (Dougan et al., 2012; Tasaki et al., 2012; Gibbs et al., 2014); such aminotransferase remains to be identified (or recognized) in chloroplasts. Measurements of chloroplast protein stability in dependence of their Nt residues in different peptidase, protease and protease adaptor (*e.g.* ClpS1) mutant backgrounds will be needed to determine to what extent chloroplast proteostasis is governed by an N-end rule.

2.5 MATERIAL AND METHODS

2.5.1 *Plant Growth and generation of protein samples*

A. thaliana (Col-0) was grown on soil in a temperature controlled chamber at 150 $\mu\text{mol photons.m}^{-2}.\text{s}$ in a 12 h light period and harvested at developmental stage 1-12. Total leaf was frozen in liquid nitrogen and ground to a powder in cooled mortar and pestle. Proteins were then extracted in 50mM HEPES-KOH, pH 8.0, 1mM EDTA, 1mM Pefabloc and 10 $\mu\text{g}/\mu\text{l}$ E64, using 1 ml volume/g fresh weight; cell debris was removed by spin columns (Friso et al., 2011). Protein concentrations were determined by the BCA protein assay (ThermoScientific). Chloroplast stroma was obtained from isolated chloroplasts as described in (Olinares et al., 2010).

2.5.2 *TAILS experiments*

A TAILS strategy was employed as described in (Kleifeld et al., 2011) and (Guryca et al., 2012) with minor modifications. Briefly, 100 to 200 μg protein in extraction buffer (above) was mixed 1:1 with 8M GuHCl in a single 1.6 or 2 ml tube. DTT was added at a final concentration of 5 mM and the solution incubated at 65°C for

1 h. Cysteines were alkylated with iodoacetamide, 15 mM final concentration, for 20 minutes in darkness at room temperature and the residual iodoacetamide quenched by addition of DTT, at a final concentration of 15mM. For dimethylation of amines, 2M formaldehyde (CH₂O) or CD₂O (heavy isotope) (made fresh in ddH₂O) and 1M NaCNBH₃ was added to give 40 mM and 20 mM respectively. The pH was lowered to 7.0 with 1M HCl and the solution incubated at 37°C for between 8 and 16 h. The reaction was quenched with 1M ammonium bicarbonate (NH₄HCO₃) at final concentration of 100mM for 2 h at 37°C. Proteins were precipitated with between 4 and 8 volumes of ice-cold acetone, 1 volume methanol and the solution kept at -80°C for 3 h. Proteins were pelleted at 14k xg for 20 min, the supernatant removed and the pellet washed twice with ice-cold methanol. The pellet was resuspended in 10 to 20 µl 8M GuHCl or DMSO followed by stepwise dilution with 50mM HEPES, pH 8.0 while vortexing to give a GuHCl concentration of < 0.8 M and a protein concentration of 1 mg/ml. Some precipitate typically remained after resuspension.

Sequencing grade trypsin (V5111 Promega) at a 1:100 or Glu-C (Promega V165A) at a ratio of 1:20, enzyme to protein ratio was added and the solution incubated overnight at 37°C. One aliquot (80 ug) of Glu-C digested sample was further digested (1:20, enzyme to protein ratio) with trypsin overnight (20 hr). Alternatively, 200 µg of labelled protein, precipitated protein was re-suspended in 1x Laemmli buffer and the resolved by SDS PAGE, 12% T Laemmli. The whole gel lane was then cut into 4 bands and in gel trypsin digest was performed as described in (Friso et al., 2011) except that no reduction and alkylation was performed and the digestion was performed in 50mM HEPES-KOH, pH 8.0. The resulting peptide extracts were dried and then resuspended

in 50mM HEPES-KOH, pH 8.0. Following each digest a 5% aliquot of the digested protein was reserved for testing labeling efficiency (before negative selection step). These aliquots were desalted using C-18 solid phase extraction spin columns (Pierce 89870) using the manufacturers guidelines and subjected to MS analysis. The remaining protein digest was added to 600 μ g (15 to 20 μ l aliquots) of dendritic high molecular weight polyglycerol-aldehyde polymer (HPG-ALD) (Flintbox Innovation Network) followed immediately by 1M NaCNBH₃ to give a 40mM final concentration. The pH was adjusted to 7.0 with 1M HCl and the sample incubated overnight at 37°C. The reaction was quenched as above with 100 mM NH₄HCO₃ and the solution filtered through a pre-washed (3 x 0.5 mL water, 2 x 100mM NH₄HCO₃) Amicon 30-kDa-molecular mass ultrafiltration device (Millipore). The filtrate was acidified with formic acid and the peptides desalted as described above. The peptides were then dried in a vacuum centrifuge and resuspended in 20 μ l of 2% formic acid. SDS PAGE was performed to ensure recovery of protein across the labeling and precipitation steps and to ensure digestion went to completion. Please note that the GuHCl concentration in the SDS sample buffer must be less than 0.04 M GuHCl to avoid precipitate and spoiled gel separation.

2.5.3 LC-MS/MS analysis

All samples were analyzed by nanoLC-MS/MS using a LTQ Orbitrap mass spectrometer (ThermoFisher) run at 100,000 resolution in MS (Orbitrap portion) at least once prior to the negative selection step, to confirm label efficiency, and at least twice following the negative selection step. The LC and MS tuning and acquisition conditions

were as previously described (Friso et al., 2011) with some minor variations. In some cases a reject list for the most abundant, persisting, highly concentrated RBCS and RBCL peptides was added.

2.5.4 Database search/peptide identification

Peak lists (mgf files) were generated from Thermo raw data files using DTA Supercharge. The peak lists were searched using MASCOT 2.4 (Matrix Science) against TAIR10, appended with all reverse sequences (Decoy) and common contaminants (71,149 sequences and 29,099,754 residues). Following an initial database search performed at 30 ppm MS tolerance, 0.8 Da MSMS tolerance, the peak list was then recalibrated as previously described (Friso et al., 2010). A semi-specific enzyme search was then conducted - semiArgC, semiGluC (V8), or semi(ArgC and GluC) - allowing for two missed cleavages, 4 ppm MS tolerance and 0.8 Da MSMS tolerance. Fixed modifications were carboxamidomethyl Cys and dimethyl Lys; variable modifications were oxidized Met, pyroGlu N-term Gln, acetyl N-term and dimethyl N-term (light, +28 Da or heavy, +32 Da). Another search including singly methylated N-term was conducted for select files in order to detect methylated Nt Pro. For samples labeled with heavy formaldehyde, a search was conducted with intermediate - light and heavy (+30 Da) dimethylation to account for proteins that underwent physiological (mono)methylated at the N-terminus.

2.5.5 Additional database searches

To test labeling efficiency, samples taken prior to negative selection were subjected to semitryptic or semiGluC searches with only fixed carboxamidomethyl Cys and variable dimethyl Lys, dimethyl Nt, acetyl Nt and oxidized Met. These search parameters enabled detection of unlabeled Lys side chains and semitryptic peptides that should not be present if dimethyl labelling was complete. To detect mono-methylated Pro, mono-methylated Nt was added to the list of variable modifications. To detect physiological Nt mono-methylation in samples labeled with heavy formaldehyde, a search was conducted with intermediate dimethylation – light/heavy (+30 Da). To monitor dimethyl labeling efficiency, pre-negative selection database search results were exported and the number of Lys per peptide calculated. The number of assigned dimethylated Lys was then compared to the Lys number for each peptide as well as the number of missed cleavages (which should equal the number of Lys unless they are followed by Pro).

2.5.6 Filtering of MS Data

Each MASCOT result (MS/MS ion search) was filtered, $P < 0.01$, minimum ion score 30. The spectra were then exported and sorted to remove any contaminant peaks such as trypsin and keratin and to ensure the number of reverse hits gave an FDR of $< 1\%$ with $FDR = 2 * FP / (FP + TP)$. Identified peptides from all db searches were combined and all peptides without an Nt label removed leaving only those with either dimethyl, acetyl or pyroGlu Q Nti. Peptides were then sorted by criteria in the following order: peptide sequence, modification 1, modification 2, and peptide score (highest to lowest). All identical peptide species were collapsed and the number summed to give

the number of spectral counts (SPC) for that peptide. In total 13858 spectra were matched using the above criteria (Supplemental Table 2.1), representing 1087 Nt peptides. All pyroGluQ Nti, downstream of K, R or E (resulting from artificial trypsin or GluC cleavage) were deemed artifacts and were removed (50 peptides). Certain groups of peptides were found that matched to the same N-terminus due to ragged ends at the C-terminus, or alternate enzyme cleavage sites (*e.g.* GluC) and SPC for these peptides were grouped such that the SPC for all peptides matching the same N-terminus were combined (peptide with most SPC is the parent of that group). If for example a missed cleavage led to the same NAA N-terminus being found twice, each with distinct a mass, both spectra are counted towards that N-terminus. Therefore, all redundant Nti that cannot be distinguished by Nt modification (148 peptides) were collapsed and their Nti added to the parent group leaving 894 unique Nt sequences (Supplemental Table 2.2) matching to 577 proteins.

2.5.7 Validation of Nti, terminology and subcellular localization

TAILS identified chloroplast Nti was compared with Nt information from the scattered Edman sequencing information available in the literature (*e.g.* (Richter and Lamppa, 1998; Peltier et al., 2002; Candat et al., 2013)) and with semi-tryptic peptides identified in PPDB. A unique Nt sequence refers to a single sequence that could be identified from different charge states of the same peptide or from separate overlapping peptides with the same modified/labeled amino-terminus. If the same sequence was identified with alternate modifications of the NAA amino group, these are considered unique Nti. As such, the same exact sequence can be found in three unique forms,

dimethylated (free N-term in original sample), NAA or pyroGlu. Proteins were annotated for subcellular location based on manually curated experimental information from PPDB - <http://ppdb.tc.cornell.edu/> (Sun et al., 2009).

2.5.8 *Generation of sequence Logos and iceLogos*

Sequence and iceLogos were generated with IceLogo v 1.2 (<http://www.proteomics.be>). The sequence logos generated are identical to those made with Weblogo (<http://weblogo.berkeley.edu/>). For nuclear-encoded proteins, we aligned sequences starting ten residues upstream of the N-terminus (P10 position) and ending ten residues downstream (P10'). Sets of sequences were loaded into iceLogo along with the 1663 known chloroplast proteins (PPDB), minus their predicted cTPs, as a reference (background) proteome. The iceLogo results were filtered to show only residues that were significantly different from the reference proteome with $P < 0.05$ for Figure 2.4B and $P < 0.01$ for Figure 2.4C and D. Amino acid colors were chosen to accentuate basic (blue), acidic (red), hydrophilic or polar (green) and sulfur containing (purple) residues. All other, generally hydrophobic, residues are in black.

2.5.9 *Annotation of protein subcellular localization*

Proteins were annotated for subcellular localization based on curated information as listed in PPDB (<http://ppdb.tc.cornell.edu/>) updated with the most recent data from the literature and other public resources. The subcellular localization in PPDB is based on available information from the literature, as well as information from specific data bases, such as those by the Rolland lab (AtChloro), but also SUBA (from

the Millar lab), TAIR and the new Arabidopsis Information Portal (AIP), etc. For annotation in PPDB, all this public information is considered together with extensive in-house information to then manually assign a subcellular localization, in particular for plastids/chloroplasts. Furthermore, it also considers information from orthologs in maize, as was recently described in (Huang et al., 2013). Annotation of the orientation of Nti of chloroplast proteins (facing the stroma, lumen or envelope intra-membrane space) was based on literature.

ACKNOWLEDGEMENTS This work was supported by grants from the National Science Foundation (NSF) MCB-1021963 and IOS-1127017 to KJVW.

SUPPORTING INFORMATION

Supplemental Figure 2.1. TAILS experimental workflow, using a method adapted from (Kleifeld et al., 2011).

Supplemental Figure 2.2. RBCS-4 Nt peptide MS spectrum labeled light and heavy to distinguish physiological mono-methylation status.

Supplemental Figure 2.3. TAILS identified Nti of intact stromal proteins and their degradation products.

Supplemental Table 2.1. 13858 matched MS/MS spectra that provide the basis for this N-terminome analysis matching to 606 non-redundant proteins.

Supplemental Table 2.2. Nt peptides after merging redundant Nt peptides (i.e. with the same mass), resulting in 1037 Nt sequences matching to 577 proteins and including degradation products.

Supplemental Table 2.3. Nt peptides after merging redundant and/or overlapping Nt sequences irrespective of PTMs (exception for NAA, treated as distinct) and C-terminal ends ('ragged ends') resulting in 894 Nt sequences, matching to 577 proteins.

Supplemental Table 2.4. 344 stromal exposed Nti of 205 n-encoded proteins used for figures 2 and 3 and associated analysis; degradation products are removed.

Supplemental Table 2.5. 39 Nti exposed to the thylakoid lumen.

Supplemental Table 2.6. Experimentally determined SPP cleavage sites described in the literature.

Supplemental Table 2.7. Nti of all 88 p-encoded proteins, their topology, predicted mature Nti based on the general NME rule and experimental Nt information from this

study, PPDB and published literature.

Supplemental Table 2.8. 129 degradation Nti of 498 stromal exposed Nti.

Supplemental Table 2.9 Comparison of identified chloroplast and mitochondrial protein Nti to other published TAILS and N-terminome studies.

LITERATURE CITED

- Alban C, Tardif M, Mininno M, Brugiére S, Gilgen A, Ma S, Mazzoleni M, Gigarel O, Martin-Laffon J, Ferro M, Ravanel S** (2014) Uncovering the protein lysine and arginine methylation network in Arabidopsis chloroplasts. *PLoS One* **9**: e95512
- Apel W, Schulze WX, Bock R** (2010) Identification of protein stability determinants in chloroplasts. *Plant J* **63**: 636-650
- Bachmair A, Finley D, Varshavsky A** (1986) In vivo half-life of a protein is a function of its amino-terminal residue. *Science* **234**: 179-186
- Bienvenut WV, Sumpton D, Martínez A, Lilla S, Espagne C, Meinel T, Giglione C** (2012) Comparative large scale characterization of plant versus mammal proteins reveals similar and idiosyncratic N-alpha-acetylation features. *Mol Cell Proteomics* **11**: M111015131
- Bonissone S, Gupta N, Romine M, Bradshaw RA, Pevzner PA** (2013) N-terminal protein processing: a comparative proteogenomic analysis. *Mol Cell Proteomics* **12**: 14-28
- Candat A, Poupart P, Andrieu JP, Chevrollier A, Reynier P, Rogniaux H, Avelange-Macherel MH, Macherel D** (2013) Experimental determination of organelle targeting-peptide cleavage sites using transient expression of green fluorescent protein translational fusions. *Anal Biochem* **434**: 44-51
- Carrie C, Small I** (2013) A reevaluation of dual-targeting of proteins to mitochondria and chloroplasts. *Biochim Biophys Acta* **1833**: 253-259
- Carrie C, Venne AS, Zahedi RP, Soll J** (2015) Identification of cleavage sites and substrate proteins for two mitochondrial intermediate peptidases in Arabidopsis thaliana. *J Exp Bot* **prepublication on-line**
- Colaert N, Helsen K, Martens L, Vandekerckhove J, Gevaert K** (2009) Improved visualization of protein consensus sequences by iceLogo. *Nat Methods* **6**: 786-787
- Cui YL, Jia QS, Yin QQ, Lin GN, Kong MM, Yang ZN** (2011) The GDC1 gene encodes a novel ankyrin domain-containing protein that is essential for grana formation in Arabidopsis. *Plant Physiol* **155**: 130-141
- Daras G, Rigas S, Tsitsekian D, Zur H, Tuller T, Hatzopoulos P** (2014) Alternative transcription initiation and the AUG context configuration control dual-organellar targeting and functional competence of Arabidopsis Lon1 protease. *Mol Plant* **7**: 989-1005
- Dirk LM, Williams MA, Houtz RL** (2001) Eukaryotic peptide deformylases. Nuclear-encoded and chloroplast-targeted enzymes in Arabidopsis. *Plant Physiol* **127**: 97-107
- Dirk LM, Williams MA, Houtz RL** (2002) Specificity of chloroplast-localized peptide deformylases as determined with peptide analogs of chloroplast-translated proteins. *Arch Biochem Biophys* **406**: 135-141

- Dougan DA, Micevski D, Truscott KN** (2012) The N-end rule pathway: from recognition by N-recognins, to destruction by AAA+proteases. *Biochim Biophys Acta* **1823**: 83-91
- 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
- Friso G, Majeran W, Huang M, Sun Q, van Wijk KJ** (2010) Reconstruction of metabolic pathways, protein expression, and homeostasis machineries across maize bundle sheath and mesophyll chloroplasts: large-scale quantitative proteomics using the first maize genome assembly. *Plant Physiol* **152**: 1219-1250
- Friso G, Olinares PDB, van Wijk KJ** (2011) The Workflow for Quantitative Proteome Analysis of Chloroplast Development and Differentiation, Chloroplast Mutants, and Protein Interactions by Spectral Counting. *In* RP Jarvis, ed, *Chloroplast Research in Arabidopsis*, Vol 775. Humana Press, New York, pp 265-282
- Gao M, Li Y, Xue X, Wang X, Long J** (2012) Stable plastid transformation for high-level recombinant protein expression: promises and challenges. *J Biomed Biotechnol* **2012**: 158232
- Gibbs DJ, Bacardit J, Bachmair A, Holdsworth MJ** (2014) The eukaryotic N-end rule pathway: conserved mechanisms and diverse functions. *Trends Cell Biol* **24**: 603-611
- Giglione C, Boularot A, Meinnel T** (2004) Protein N-terminal methionine excision. *Cell Mol Life Sci* **61**: 1455-1474
- Giglione C, Fieulaine S, Meinnel T** (2009) Cotranslational processing mechanisms: towards a dynamic 3D model. *Trends Biochem Sci* **34**: 417-426
- Giglione C, Fieulaine S, Meinnel T** (2014) N-terminal protein modifications: Bringing back into play the ribosome. *Biochimie* **prepublication on-line**: 134-146
- Giglione C, Vallon O, Meinnel T** (2003) Control of protein life-span by N-terminal methionine excision. *EMBO J* **22**: 13-23
- Graciet E, Mesiti F, Wellmer F** (2010) Structure and evolutionary conservation of the plant N-end rule pathway. *Plant J* **61**: 741-751
- Graciet E, Walter F, Maoileidigh DO, Pollmann S, Meyerowitz EM, Varshavsky A, Wellmer F** (2009) The N-end rule pathway controls multiple functions during Arabidopsis shoot and leaf development. *Proc Natl Acad Sci U S A* **106**: 13618-13623
- Graciet E, Wellmer F** (2010) The plant N-end rule pathway: structure and functions. *Trends Plant Sci* **15**: 447-453
- Guryca V, Lamerz J, Ducret A, Cutler P** (2012) Qualitative improvement and quantitative assessment of N-terminomics. *Proteomics* **12**: 1207-1216
- Hollebeke J, Van Damme P, Gevaert K** (2012) N-terminal acetylation and other functions of N-alpha-acetyltransferases. *Biol Chem* **393**: 291-298

- Hoshiyasu S, Kohzuma K, Yoshida K, Fujiwara M, Fukao Y, Yokota A, Akashi K** (2013) Potential involvement of N-terminal acetylation in the quantitative regulation of the epsilon subunit of chloroplast ATP synthase under drought stress. *Biosci Biotechnol Biochem* **77**: 998-1007
- Houtz RL, Magnani R, Nayak NR, Dirk LM** (2008) Co- and post-translational modifications in Rubisco: unanswered questions. *J Exp Bot* **59**: 1635-1645
- Hsu SC, Endow JK, Ruppel NJ, Roston RL, Baldwin AJ, Inoue K** (2011) Functional diversification of thylakoidal processing peptidases in *Arabidopsis thaliana*. *PLoS One* **6**: e27258
- Hu J, Baker A, Bartel B, Linka N, Mullen RT, Reumann S, Zolman BK** (2012) Plant peroxisomes: biogenesis and function. *Plant Cell* **24**: 2279-2303
- Huang M, Friso G, Nishimura K, Qu X, Olinares PD, Majeran W, Sun Q, van Wijk KJ** (2013) Construction of plastid reference proteomes for maize and *Arabidopsis* and evaluation of their orthologous relationships; the concept of orthoproteomics. *J Proteome Res* **12**: 491-504
- Huang S, Nelson CJ, Li L, Taylor NL, Stroher E, Petereit J, Millar AH** (2015) The intermediate cleavage peptidase ICP55 modifies enzyme N-termini and alters protein stability in *Arabidopsis* mitochondria. *Plant Physiol*
- Huang S, Taylor NL, Whelan J, Millar AH** (2009) Refining the definition of plant mitochondrial presequences through analysis of sorting signals, N-terminal modifications, and cleavage motifs. *Plant Physiol* **150**: 1272-1285
- Huesgen PF, Alami M, Lange PF, Foster LJ, Schroder WP, Overall CM, Green BR** (2013) Proteomic amino-termini profiling reveals targeting information for protein import into complex plastids. *PLoS One* **8**: e74483
- Jones JD, O'Connor CD** (2011) Protein acetylation in prokaryotes. *Proteomics* **11**: 3012-3022
- Kikuchi S, Bedard J, Hirano M, Hirabayashi Y, Oishi M, Imai M, Takase M, Ide T, Nakai M** (2013) Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science* **339**: 571-574
- Kim J, Olinares PD, Oh SH, Ghisaura S, Poliakov A, Ponnala L, van Wijk KJ** (2013) Modified Clp protease complex in the ClpP3 null mutant and consequences for chloroplast development and function in *Arabidopsis*. *Plant Physiol* **162**: 157-179
- Kleifeld O, Doucet A, Prudova A, auf dem Keller U, Gioia M, Kizhakkedathu JN, Overall CM** (2011) Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nat Protoc* **6**: 1578-1611
- Kmieciak B, Teixeira PF, Berntsson RP, Murcha MW, Branca RM, Radomiljac JD, Regberg J, Svensson LM, Bakali A, Langel U, Lehtio J, Whelan J, Stenmark P, Glaser E** (2013) Organellar oligopeptidase (OOP) provides a complementary pathway for targeting peptide degradation in mitochondria and chloroplasts. *Proc Natl Acad Sci U S A* **110**: E3761-3769

- Kmiec B, Teixeira PF, Glaser E** (2014) Phenotypical consequences of expressing the dually targeted Presequence Protease, AtPreP, exclusively in mitochondria. *Biochimie* **100**: 167-170
- Kohler D, Montandon C, Hause G, Majovsky P, Kessler F, Baginsky S, Agne B** (2015) Characterization of chloroplast protein import without Tic56, a component of the 1-MDa TIC translocon. *Plant Physiol* **prepublication online**
- Lange PF, Overall CM** (2013) Protein TAILS: when termini tell tales of proteolysis and function. *Curr Opin Chem Biol* **17**: 73-82
- Lehtimäki N, Koskela MM, Dahlstrom KM, Pakula E, Lintala M, Scholz M, Hippler M, Hanke GT, Rokka A, Battchikova N, Salminen TA, Mulo P** (2014) Posttranslational modifications of FERREDOXIN-NADP+ OXIDOREDUCTASE in Arabidopsis chloroplasts. *Plant Physiol* **166**: 1764-1776
- Linster E, Stephan I, Bienvenut WV, Maple-Grodem J, Myklebust LM, Huber M, Reichelt M, Sticht C, Geir Moller S, Meinnel T, Arnesen T, Giglione C, Hell R, Wirtz M** (2015) Downregulation of N-terminal acetylation triggers ABA-mediated drought responses in Arabidopsis. *Nat Commun* **6**: 7640
- Lundquist P, Poliakov A, Giacomelli L, Friso G, Appel M, McQuinn RP, Krasnoff SB, Rowland O, Ponnala L, Sun Q, van Wijk KJ** (2013) Loss of plastoglobule-localized kinases ABC1K1 and ABC1K3 leads to a conditional degreening phenotype, a modified prenyl-lipid composition and recruitment of JA biosynthesis. *Plant Cell* **25**: 1818-1839
- Meinnel T, Serero A, Giglione C** (2006) Impact of the N-terminal amino acid on targeted protein degradation. *Biol Chem* **387**: 839-851
- Midorikawa T, Endow JK, Dufour J, Zhu J, Inoue K** (2014) Plastidic type I signal peptidase 1 is a redox-dependent thylakoidal processing peptidase. *Plant J* **80**: 592-603
- Mininno M, Brugiére S, Pautre V, Gilgen A, Ma S, Ferro M, Tardif M, Alban C, Ravanel S** (2012) Characterization of chloroplastic fructose 1,6-bisphosphate aldolases as lysine-methylated proteins in plants. *J Biol Chem* **287**: 21034-21044
- Mulligan RM, Houtz RL, Tolbert NE** (1988) Reaction-intermediate analogue binding by ribulose biphosphate carboxylase/oxygenase causes specific changes in proteolytic sensitivity: the amino-terminal residue of the large subunit is acetylated proline. *Proc Natl Acad Sci U S A* **85**: 1513-1517
- Nishimura K, Asakura Y, Friso G, Kim J, Oh SH, Rutschow H, Ponnala L, van Wijk KJ** (2013) ClpS1 is a conserved substrate selector for the chloroplast Clp protease system in Arabidopsis. *Plant Cell* **25**: 2276-2301
- Olinares PD, Ponnala L, van Wijk KJ** (2010) Megadalton complexes in the chloroplast stroma of Arabidopsis thaliana characterized by size exclusion chromatography, mass spectrometry and hierarchical clustering. *Mol Cell Proteomics* **9.7**: 1594-1615
- Ouyang M, Li X, Ma J, Chi W, Xiao J, Zou M, Chen F, Lu C, Zhang L** (2011) LTD is a protein required for sorting light-harvesting chlorophyll-binding proteins to the chloroplast SRP pathway. *Nat Commun* **2**: 277

- Peltier JB, Emanuelsson O, Kalume DE, Ytterberg J, Friso G, Rudella A, Liberles DA, Soderberg L, Roepstorff P, von Heijne G, van Wijk KJ** (2002) Central Functions of the Lumenal and Peripheral Thylakoid Proteome of Arabidopsis Determined by Experimentation and Genome-Wide Prediction. *Plant Cell* **14**: 211-236.
- Peltier JB, Friso G, Kalume DE, Roepstorff P, Nilsson F, Adamska I, van Wijk KJ** (2000) Proteomics of the Chloroplast. Systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins. *Plant Cell* **12**: 319-342
- Preissler S, Deuerling E** (2012) Ribosome-associated chaperones as key players in proteostasis. *Trends Biochem Sci* **37**: 274-283
- Recuenco-Munoz L, Offre P, Valledor L, Lyon D, Weckwerth W, Wienkoop S** (2015) Targeted quantitative analysis of a diurnal RuBisCO subunit expression and translation profile in *Chlamydomonas reinhardtii* introducing a novel Mass Western approach. *J Proteomics* **113**: 143-153
- Richter S, Lamppa GK** (1998) A chloroplast processing enzyme functions as the general stromal processing peptidase. *Proc Natl Acad Sci U S A* **95**: 7463-7468
- Richter S, Lamppa GK** (1999) Stromal processing peptidase binds transit peptides and initiates their ATP-dependent turnover in chloroplasts. *J Cell Biol* **147**: 33-44
- Richter S, Lamppa GK** (2002) Determinants for removal and degradation of transit peptides of chloroplast precursor proteins. *J Biol Chem* **277**: 43888-43894
- Richter S, Zhong R, Lamppa G** (2005) Function of the stromal processing peptidase in the chloroplast import pathway. *Physiologia Plantarum* **123**: 362-368
- Rokka A, Aro EM, Vener AV** (2011) Thylakoid phosphoproteins: identification of phosphorylation sites. *Methods Mol Biol* **684**: 171-186
- Sandikci A, Gloge F, Martinez M, Mayer MP, Wade R, Bukau B, Kramer G** (2013) Dynamic enzyme docking to the ribosome coordinates N-terminal processing with polypeptide folding. *Nat Struct Mol Biol* **20**: 843-850
- Staes A, Impens F, Van Damme P, Ruttens B, Goethals M, Demol H, Timmerman E, Vandekerckhove J, Gevaert K** (2011) Selecting protein N-terminal peptides by combined fractional diagonal chromatography. *Nat Protoc* **6**: 1130-1141
- Starheim KK, Gevaert K, Arnesen T** (2012) Protein N-terminal acetyltransferases: when the start matters. *Trends Biochem Sci* **37**: 152-161
- Sun Q, Zybailov B, Majeran W, Friso G, Olinares PD, van Wijk KJ** (2009) PPDB, the Plant Proteomics Database at Cornell. *Nucleic Acids Res* **37**: D969-974
- Tasaki T, Sriram SM, Park KS, Kwon YT** (2012) The N-end rule pathway. *Annu Rev Biochem* **81**: 261-289
- Tobias JW, Shrader TE, Rocap G, Varshavsky A** (1991) The N-end rule in bacteria. *Science* **254**: 1374-1377

- Trosch R, Jarvis P** (2011) The stromal processing peptidase of chloroplasts is essential in Arabidopsis, with knockout mutations causing embryo arrest after the 16-cell stage. *PLoS One* **6**: e23039
- Tsiatsiani L, Timmerman E, De Bock PJ, Vercammen D, Stael S, van de Cotte B, Staes A, Goethals M, Beunens T, Van Damme P, Gevaert K, Van Breusegem F** (2013) The Arabidopsis metacaspase9 degradome. *Plant Cell* **25**: 2831-2847
- Urantowka A, Knorpp C, Olczak T, Kolodziejczak M, Janska H** (2005) Plant mitochondria contain at least two i-AAA-like complexes. *Plant Mol Biol* **59**: 239-252
- van Wijk KJ** (2015) Protein maturation and proteolysis in plant plastids, mitochondria, and peroxisomes. *Annu Rev Plant Biol* **66**: 75-111
- van Wijk KJ, Friso G, Walther D, Schulze WX** (2014) Meta-Analysis of Arabidopsis thaliana Phospho-Proteomics Data Reveals Compartmentalization of Phosphorylation Motifs. *Plant Cell* **26**: 2367-2389
- Varshavsky A** (2011) The N-end rule pathway and regulation by proteolysis. *Protein Sci* **20**: 1298-1345
- Vener AV** (2007) Environmentally modulated phosphorylation and dynamics of proteins in photosynthetic membranes. *Biochim Biophys Acta* **1767**: 449-457
- Vener AV, Harms A, Sussman MR, Vierstra RD** (2001) Mass spectrometric resolution of reversible protein phosphorylation in photosynthetic membranes of Arabidopsis thaliana. *J Biol Chem* **276**: 6959-6966.
- Walling LL** (2006) Recycling or regulation? The role of amino-terminal modifying enzymes. *Curr Opin Plant Biol* **9**: 227-233
- Xu F, Huang Y, Li L, Gannon P, Linster E, Huber M, Kapos P, Bienvenut W, Polevoda B, Meinel T, Hell R, Giglione C, Zhang Y, Wirtz M, Chen S, Li X** (2015) Two N-terminal acetyltransferases antagonistically regulate the stability of a nod-like receptor in Arabidopsis. *Plant Cell* **27**: 1547-1562
- Zhang H, Deery MJ, Gannon L, Powers SJ, Lilley KS, Theodoulou FL** (2015) Quantitative proteomics analysis of the Arg/N-end rule pathway of targeted degradation in Arabidopsis roots. *Proteomics* **prepublication online**
- Zybailov B, Rutschow H, Friso G, Rudella A, Emanuelsson O, Sun Q, van Wijk KJ** (2008) Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS One* **3**: e1994
- Zybailov B, Sun Q, van Wijk KJ** (2009) Workflow for large scale detection and validation of peptide modifications by RPLC-LTQ-Orbitrap: application to the Arabidopsis thaliana leaf proteome and an online modified peptide library. *Anal Chem* **81**: 8015-8024

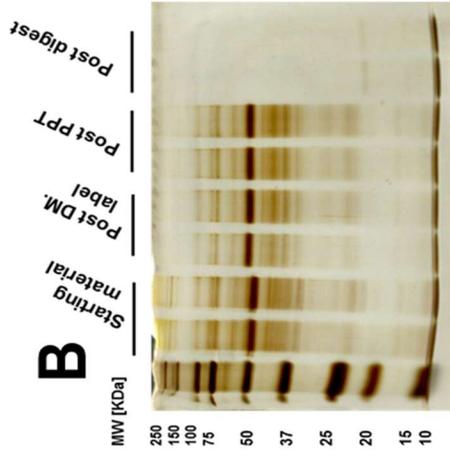
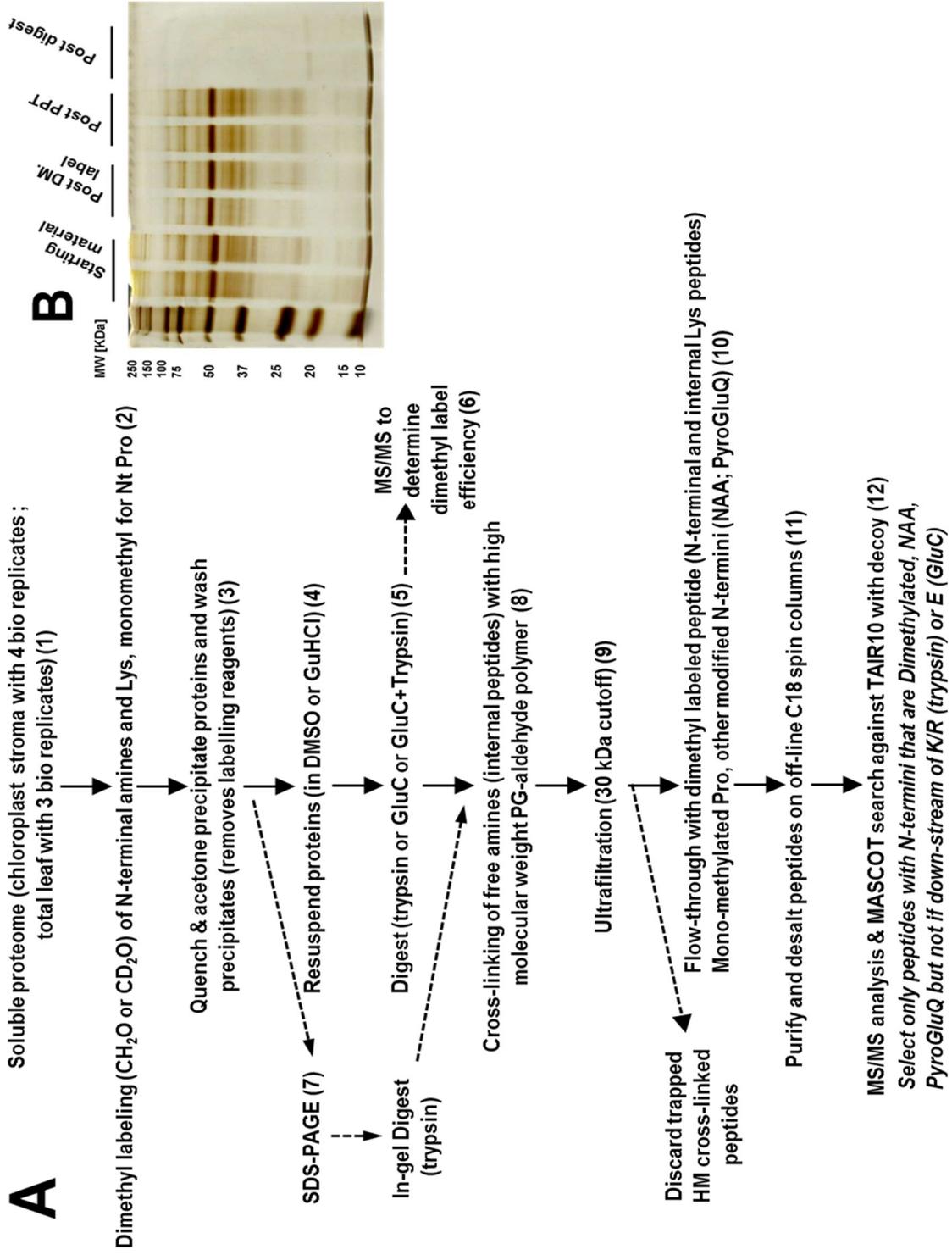
APPENDIX

Supplemental Figure 2.1. TAILS experimental workflow, using a method adapted from (Kleifeld, *et al.* 2011) and protein recovery.

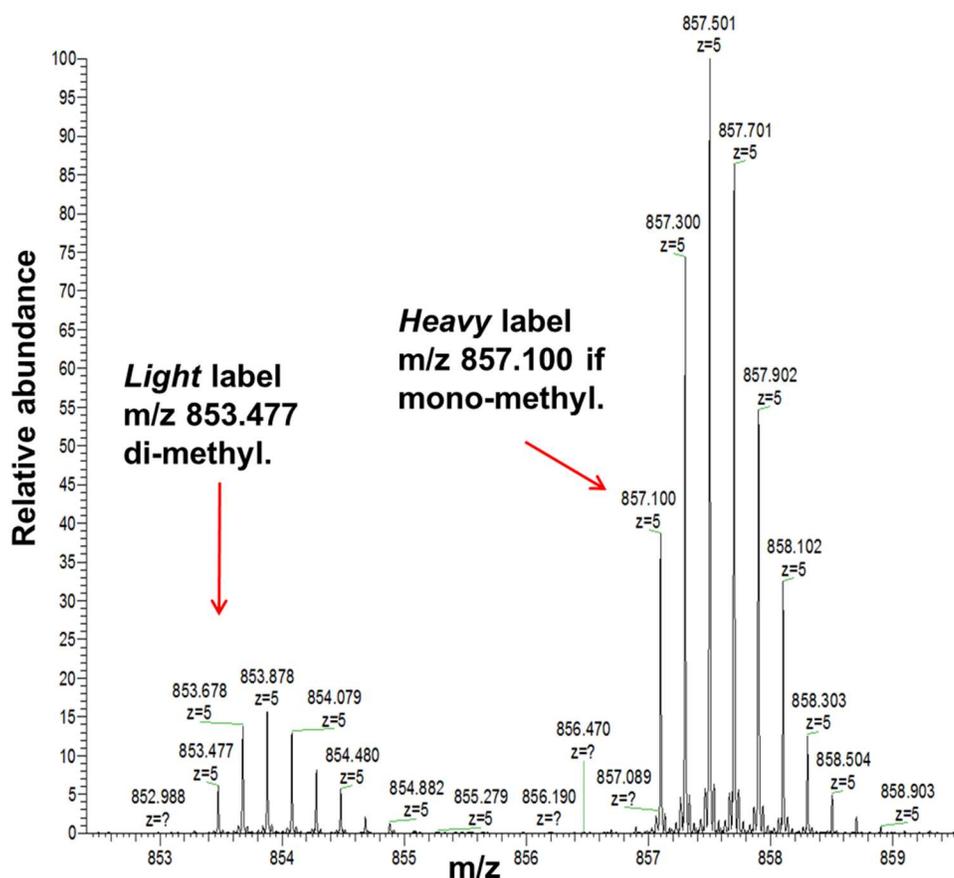
(A) 100-200 μg of proteins were labeled overnight in a single tube (2). Following quenching and acetone precipitation (3), proteins were resuspended (4) and subjected to overnight in-solution digestion (5) or resolved by SDS-PAGE and subjected to in-gel digest (7). Crosslinking of internal peptides was performed by adding the digested proteins directly to pre-aliquoted polyglycerol-aldehyde polymer along with NaCNBH_3 , pH adjusted to 7, and reacted overnight (8). The polymer and attached internal peptides were then removed by ultrafiltration (9) and the flow-through with dimethyl labeled peptide (N-terminal and internal Lys peptides), mono-methylated Pro, other modified N-termini (NAA; PyroGluQ) collected (10). Peptides were then desalted over C18 resin (11) followed by MS/MS analysis on a LTQ-Orbitrap and database searching (12).

(B) The silver-stained SDS-PAGE gel demonstrates the tracking of protein recovery throughout a TAILS experiment. This representative gels (total soluble leaf preparation) shows good recovery of protein post-labeling and post-precipitate; and confirmation of digestion with trypsin. Please note fine band around 200 kDa in starting material that is no longer visible following labeling and precipitation.

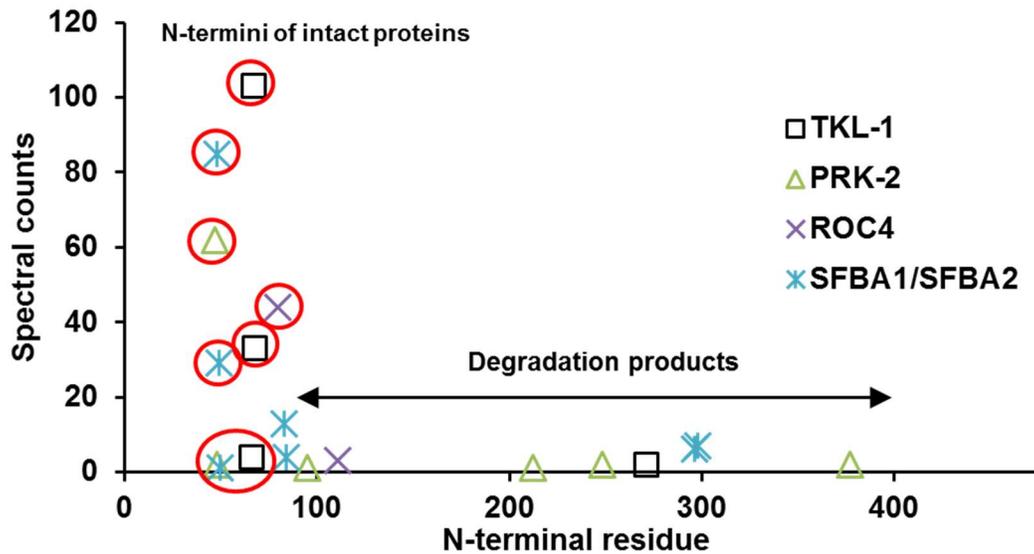
Abbreviations: DM, dimethyl; PPT, precipitate. In total, we analyzed 4 independent stromal preparations (10 TAILS experiments) and 3 independent soluble leaf protein extracts (9 TAILS experiments).



**RBCS-4 N-terminus:
MQVWPPIGKKKFETLSYLPDLTDSELAKEVDYLIR**



Supplemental Figure 2.2. RBCS-4 Nt peptide MS spectrum labeled light and heavy to distinguish physiological mono-methylation status. Identical stromal extracts were labeled either light (formaldehyde) or heavy (deuterated formaldehyde) and then mixed at 5:1 (not shown) or 1:5 ratio and then subjected to TAILS. Observed heavy m/z values indicate RBCS-4 is naturally mono-methylated, m/z within 4 ppm of the expected values for both the light and heavy forms of the peptide. For RBCS-4 the Nt peptide typically eluted between 58 and 60 minutes, ESI producing 4⁺ and 5⁺ charge states. The peptide contains 4 lysines, which, in addition to the N-terminus, should all be dimethylated by TAILS, the *light* labeled form having a m/z of 853.4736 (at 5⁺), *heavy* m/z of 857.4987 (at 5⁺). However, if the native N-terminus was already monomethylated, the heavy labeling should produce a product 2 amu lighter than predicted for full labeling. Indeed, the heavy labeled peptide has an m/z of 857.1047 (5⁺) corresponding to a naturally monomethylated form.



Supplemental Figure 2.3. TAILS identified Nti of intact stromal proteins and their degradation products. Distribution of detected Nti for abundant stromal proteins: isomerase ROC4, transketolase 1 (TKL1), phosphoribulose kinase (PRK2) and sedoheptulose fructose biphosphatase (SFBA). The x-axis shows the Nt residue position and the y-axis shows the number of matched spectra in stromal samples. The data point circled in red are the Nti of the intact, mature proteins. The other data points are the Nti of degradation products. The Nt peptide sequences of the degradation products are listed in Supplemental Table 2.8.

CHAPTER 3

GENETIC AND FUNCTIONAL INTERACTIONS OF THE ORGANELLAR PEPTIDASES CLP, PREP AND OOP IN *ARABIDOPSIS THALIANA*¹

3.1 ABSTRACT

Proteolysis is essential for proteome dynamics and homeostasis (proteostasis) in all living organisms. A network of different types of peptidases and chaperones ensure proteostasis within plant chloroplasts and mitochondria. This study describes the genetic and functional interactions between the chloroplast stromal CLP protease system and the PREP1, PREP2 and OOP peptidases dually localized in chloroplast stroma and mitochondria. A strong synergistic interaction was observed between mutants in CLPRT protease core components (*clpr1* and *clpr2*) and both PREP homologs but, surprisingly, with the most pronounced effect for the lower abundant PREP2. Mutants with very low CLP core capacity (*clpt1 clpt2*) in the PREP1,2 double mutant (*prep1 prep2*) background resulted in embryo lethality. Synergistic effects were also observed between a CLPC1 chaperone null mutant (*clpc1*) and *prep1 prep2*. Using label-free

¹Manuscript in preparation for submission to Plant Physiology (2017) **Jitae Kim, Elden Rowland, Giulia Friso, Anton Poliakov, Lalit Ponnala, Qi Sun, and Klaas J. van Wijk**

The contribution of the thesis' author to this work consisted of N-terminal proteomics and peptidomics experiments and label free spectral counting for *prep1 prep2* chloroplast experiments. JK performed all genetic analyses and characterization of peptidase mutants. GF and AP performed additional proteomics experiments. ER and KJVW wrote the manuscript.

mass spectrometry, terminal amine stable isotopic labeling of substrates (TAILS) and peptidomics for identification and quantification of protein N-termini and protein degradation products, we determined the proteome phenotypes of single and higher order mutants in these three proteolytic systems. We conclude that there is: i) interference with *in vivo* cTP processing in *prep* and *clp* mutant alleles, ii) accumulation of protein degradation products in *prep* and *prep oop* backgrounds, iii) pronounced protein folding stress in the CLP mutants, that is only minor in *prep1 prep2* and lacking in *oop*. Functional models are proposed that explain mechanisms of stromal N-terminal protein maturation disruption, and how these peptidases are part of a proteolysis network and degradation cascade.

3.2 INTRODUCTION

Plastids are essential organelles of prokaryotic origin present in all plant cells. Except for outer envelope proteins, most nuclear-encoded plastid proteins are imported using an N-terminal (Nt) chloroplast transit peptide (cTP) that is cleaved by STROMAL PROCESSING PEPTIDASE (SPP) recognizing the C-terminal region of the cTP (Richter and Lamppa, 1998, 1999; Trosch and Jarvis, 2011). SPP cleaves the cTP from the mature protein followed by its immediate release, whereas the cleaved cTP is released as two fragments through cleavage by SPP. These cTP fragments are then degraded by ATP- and metallo-dependent, non-serine type, stromal degradation activity (Richter and Lamppa, 1999, 2002). This activity, coined ‘transit peptide subfragment-degrading enzyme’ (TP-DE) could be inhibited by oligopeptides (25-65 aa) but not whole proteins, thus predicting the presence of a non-serine type oligopeptidase that

degrades cTP fragments, but that can also degrade other protein fragments (Richter and Lamppa, 2002). Cleaved cTPs and fragments thereof (in particular the C-terminal portion) interfere with SPP preprocessing activity *in vitro* by direct competition for the SPP substrate binding domain (reviewed in Richter et al., 2005). Loss of SPP negatively regulates protein import *in vivo* resulting in accumulation of precursor proteins in the cytosol (Zhong et al., 2003).

Different proteases, processing peptidases and aminopeptidases (from here on referred to as ‘peptidases’) are present in the plastid to remove and recycle cleaved cTPs and luminal transit peptides (ITPs), damaged, miss-folded proteins, or unwanted proteins, reviewed in (Kmiec et al., 2014; van Wijk, 2015; Nishimura et al., 2016). Collectively, these peptidases must form a peptidase network, with built-in redundancies, hierarchies and complementary activities. The challenge is to determine the contributions of each peptidase (system) to the network. The current study aims to understand the functional interactions and *in vivo* contributions of three soluble stromal proteolytic systems to this peptidase network in Arabidopsis, namely the CLP peptidase system (Nishimura and van Wijk, 2015) and the PRESEQUENCE PROTEASE 1 and 2 (PREP1,2) and the ORGANELLAR OLIGOPEPTIDASE (OOP) metallo-oligopeptidases (Teixeira and Glaser, 2013; Kmiec et al., 2014).

ATP-dependent CLP peptidase systems are present in bacteria, mitochondria and plastids, where they likely regulate a broad range of substrates (Sauer and Baker, 2011; Alexopoulos et al., 2012; Liu et al., 2014; Nishimura and van Wijk, 2015). CLP peptidase consist of ATP-dependent chaperones that form 600 kDa hexameric rings and associate with a barrel-shaped tetradameric peptidase core complex (~350 kDa).

Substrates are unfolded by these chaperones and delivered into the CLP core. Bacterial CLP peptidases use adaptor proteins such as CLPS and destabilizing signals in the N-, C-terminal or internal regions of proteins (known as degrons) for substrate recognition and delivery (Kirstein et al., 2009; Battesti and Gottesman, 2013). Plant chloroplasts harbor a highly diversified CLP system, consisting of a hetero-oligomeric peptidase core comprising five proteolytically-active subunits (CLPP1, CLPP3–6) and four proteolytically-inactive proteins (CLPR1–4), as well as two stabilizing/activating factors (CLPT1/2), three chaperones (CLPC1, CLPC2, and CLPD), the adaptor CLPS1 (Nishimura et al., 2013) and its co-adaptor CLPF (Nishimura and van Wijk, 2015). CLPS1 directs selected substrates to the CLPC chaperone complex for degradation and we recently identified a small set of likely candidate substrates (Nishimura et al., 2013), including GLUTR (Apitz et al., 2016). CLPF is proposed to be a co-adaptor of CLPS1 but this binary adaptor model needs further investigation (Nishimura et al., 2015). Interestingly, the mitochondrial CLP system in plants is more similar to bacterial systems, requiring fewer genes and having less heterogeneity (Nishimura and van Wijk, 2015). The CLP peptidases in bacteria typically generate short peptide products up to 25 residues in length, with the majority being ~5-15 amino acid fragments (Choi and Licht, 2005).

PREP1, PREP2 and OOP in Arabidopsis have been extensively studied by the Glaser lab, reviewed in (Teixeira and Glaser, 2013; Kmiec et al., 2014) and see (Teixeira et al., 2017). All three peptidases are dual-targeted to chloroplasts and mitochondria, and it was proposed that these three peptidases together degrade cleaved cTPs and mitochondrial transit peptides (mTPs) in the chloroplast stroma and the mitochondrial

matrix, respectively. PREP1,2 are members of the M16 family of metallo-peptidases, similar to the stromal and mitochondrial M16 processing peptidases, SPP and MPP, respectively. These peptidases have no ATP/GTP binding domains or known protein binding partners and therefore they cannot unfold proteins. Crystal structures of Arabidopsis PREP1 at 2.1Å (Johnson et al., 2006) and human PREP at 2.0 Å (King et al., 2014) suggested that PREP acts as a monomer with a dynamically formed proteolytic chamber (10000-13000 Å³) that opens and closes through a hinge mechanism. *In vitro* experiments indicated that the Arabidopsis PREP1,2 homologs can degrade peptides/proteins within a size range of 10 and 65 amino acids (Stahl et al., 2005), consistent with the size of the proteolytic chamber. They are suggested to have a preference for positively charged peptides because of a partially negatively charged binding pocket (Glaser and Alikhani, 2010; King et al., 2014). OOP is a member of the M3 metallo-peptidase family which also includes plant mitochondrial OCT1 that cleaves octapeptides from selected protein N-termini (Carrie et al., 2015). The crystal structure of OOP suggested a monomer with two major domains enclosing a smaller substrate binding cavity of 3000 Å³ (Kmieć et al., 2013). *In vitro* experiments with recombinant OOP indicates that it degrades only short peptides ranging between 8 and 23 amino acids (Kmieć et al., 2013). The upper substrate size limit of OOP and PREP distinguishes them from the CLP system, which can degrade larger and folded proteins due to the unfolding capacity of the CLP chaperones.

It was proposed that the primary functions of dual-localized PREP and OOP are to degrade cleaved cTPs and mTPs with OOP operating down-stream of PREP1,2 and/or CLP (Kmieć et al., 2013). Similarly, an earlier study of the OOP homolog in *E.*

coli (OPDA) suggested it to function down-stream of CLP and LON peptidases (Jain and Chan, 2007). Recently it was shown that PREP and OOP peptidases together can indeed degrade recombinant or synthetic cTPs and other peptides but that aminopeptidases, in particular M1, M17 and M18 metallo-exopeptidases, are required to completely degrade these fragments into individual amino acids (Teixeira et al., 2017).

Single and double *Arabidopsis* T-DNA insertion mutants for PREP1 (*prep1*) and PREP2 (*prep2*) (Cederholm et al., 2009) and single OOP (*oop-1* and *oop-2*) and triple PREP1,2 OOP mutants (*prep1 prep2 oop-2*) (Kmiec et al., 2013) have been published by the Glaser lab. The single *prep2* and *oop-1,2* mutants showed no obvious growth or developmental phenotypes, whereas *prep1* showed a slightly pale-green phenotype but without obvious growth penalty. *prep1 prep2* showed visibly reduced growth and slightly delayed development, which was more pronounced in *prep1 prep2 oop-2*. All mutants were viable and fertile on soil and appeared healthy and green without changes in leaf shape, but *prep1 prep2* and *prep1 prep2 oop2* showed delayed greening of the very young leaves. Complementation of *prep1-1 prep2-1* with a form of PREP1 that only targets to the mitochondria showed that the growth rosette phenotype was mostly due to loss of PREP1 activity in chloroplasts, rather than mitochondria (Kmiec et al., 2014).

The current study aims to understand the functional interactions and *in vivo* contributions of CLP, PREP and OOP, through the generation of higher order mutants and *in vivo* proteomics techniques that monitor chloroplast protein Nt maturation, accumulation of degradation products and protein abundance levels. This demonstrated

strong synergistic genetic interactions between the CLP and PREP peptidase systems resulting in embryo lethality, partial inhibition of Nt protein processing in absence of PREP and CLP activity, and differential accumulation of cTPs and protein degradation products. Strikingly, whereas loss of CLP capacity resulted in strong over-accumulation of all chloroplast stromal chaperone systems (*e.g.* CPN60/10/20, HSP90 and CLPB3) indicative of a significant folding and aggregation stress, the abundance of these chaperones systems were not or only marginally changed in single or higher order PREP1,2 and OOP mutants. We present models describing the positions of CLP, PREP and OOP peptidases within the chloroplast peptidase network, and the *in vivo* feedback inhibition of pre-protein maturation when peptide fragments accumulate and chaperones are limited.

3.3 RESULTS

3.3.1 Genetic interactions between PREP1, PREP2 and OOP

In an independent and long term effort towards unravelling the functional chloroplast peptidase network, we isolated four OOP alleles (*oop-1,2,3,4*) and one allele for each PREP1 (*prep1-1*; from hereon named *prep1*) and PREP2 (*prep2-1*; from hereon named *prep2*) (Figure 3.1A,B; Supplemental Figure 3.1A,B). RT-PCR showed that *oop-2* was the only true null allele at the transcript level (insertion near the 6th exon) (Supplemental Figure 3.1A), whereas *prep1* and *prep2* were essentially null alleles with < 1% of mRNA compared to wild type (wt) (Supplemental Figure 3.1B; Figure 3.1A,B). We generated double and triple mutants between *prep1*, *prep2* and *oop-2* (from hereon named *oop*). Consistent with previous observations by the Glaser lab, the *prep1 prep2*

double mutant showed a ~20-30% reduced size in plant size, but without any other obvious visible phenotypes, whereas *oop* did not show a visible phenotype (Figure 3.1A, Supplemental Figure 3.1A,B). RT-PCR analysis of *oop*, *prep1* and *prep2* did not show any changes in the level of the other peptidase genes (*e.g.* in *oop*, mRNA levels for *PREP1* and *PREP2* were unaffected, etc.). However, mRNA levels for OOP in the *prep1 prep2* double mutant increased by ~50% (Figure 3.1B) suggesting some feedback regulation and genetic interactions between these two peptidase systems. Double mutants *oop prep1* and *oop prep2* did not show visible phenotypes (Figure 3.1A). In contrast, the triple mutant showed delayed rosette development and flowering, and in particular the youngest, emerging leaves of the triple *prep1 prep2 oop* mutant were pale green and the mutant rosette was dramatically smaller than the *prep1 prep2* double mutant, indicative of a strong genetic interaction (synergistic effect) (Figure 3.1A). At later stages of development the triple mutant largely recovered especially after the transition to flowering (Supplemental Figure 3.2A,B). Mature plants were subjected to brighter light ($500 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$) for 11 days and the mutants showed no additional phenotypes when compared to wt (Supplemental Figure 3.2C). Seeds from the triple mutant had slightly collapsed (shriveled) structure, but the germination rate was normal (Supplemental Figure 3.2D,E).

Extensive tandem mass spectrometry (MS/MS) analysis of leaf or chloroplast proteomes of the single, double and triple mutants did not detect PREP1, PREP2 or

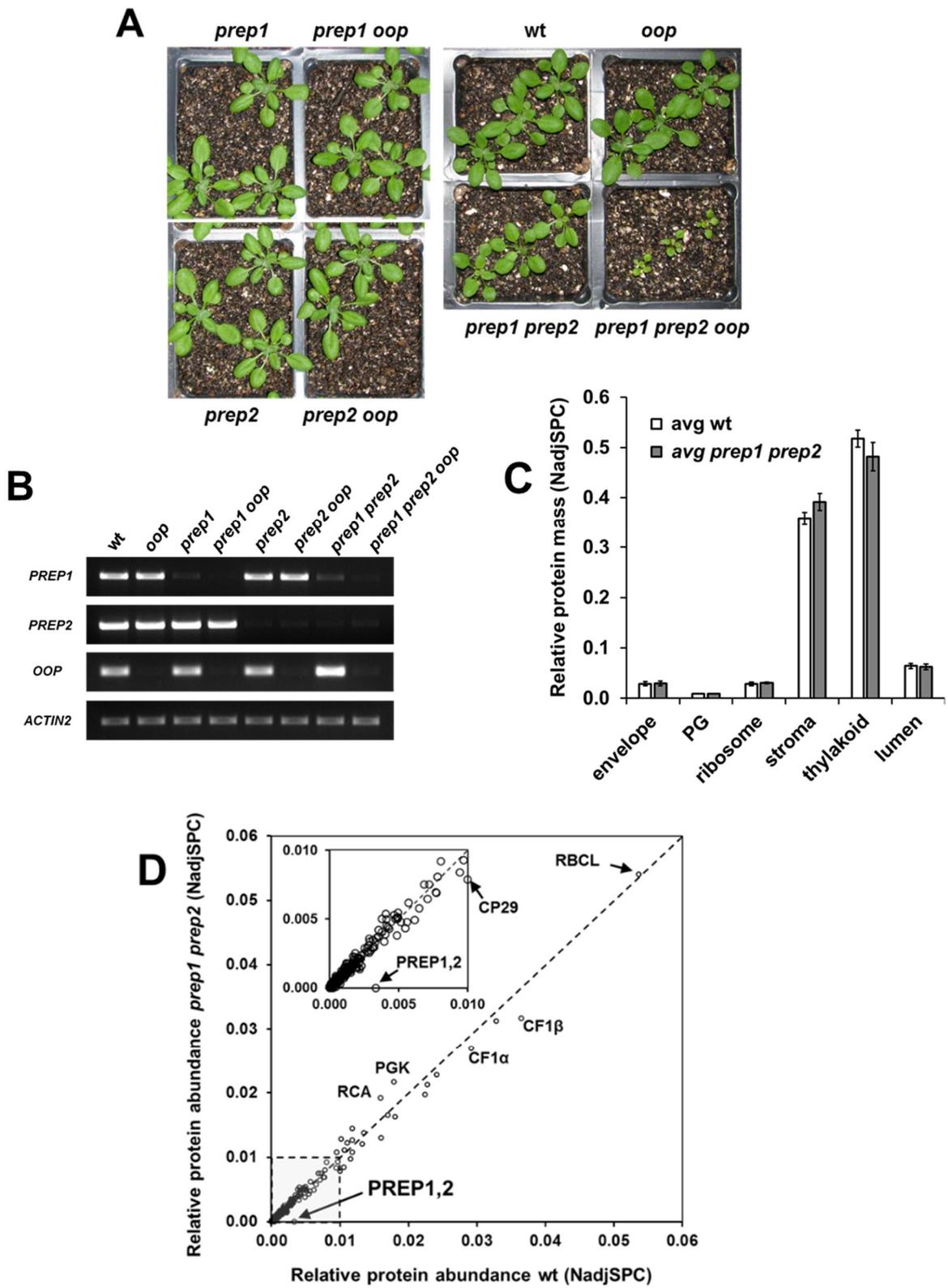
Figure 3.1. Genetic interaction between PREP and OOP peptidases.

(A) Comparison of wt, *oop*, *prep1 prep2* and *prep1 prep2 oop*. Plants were grown on soil for 20 days under a 10-h/14-h light/dark cycle at 120 $\mu\text{mol photons.m}^{-2}.\text{s}^{-1}$. Comparison of *prep1*, *prep1 oop*, *prep2* and *prep2 oop*. Plants were grown on soil for 23 days under a 10-h/14-h light/dark cycle at 120 $\mu\text{mol photons.m}^{-2}.\text{s}^{-1}$.

(B) Transcript accumulation levels in the leaves of wt, *oop*, *prep1*, *prep1 oop*, *prep2*, *prep2 oop*, *prep1 prep2* and *prep1 prep2 oop*. Transcript levels were determined by RT-PCR using gene-specific primer pairs; ACTIN2 was used as internal control.

(C) Relative protein mass in envelope, plastoglobules (PG), ribosomes, stroma, thylakoid and lumen in wt and *prep1 prep2* as determined by MS/MS. NadjSPC – normalized adjusted spectral count. Standard deviations are shown. N = 3.

(D) Comparison of relative chloroplast protein accumulation levels in wt and *prep1 prep2* averaged across three biological replicates. Several proteins are marked: PREP1,2, RUBISCO ACTIVATE (RCA), PHOSPHOGLYCERATE KINASE (PGK), ATP-SYNTHASE 1 (CF1) α,β , RUBISCO LARGE SUBUNIT (RBCL) and CHLOROPLAST PROTEIN 29 kDa (CP29).



OOP in the corresponding homozygous mutant allele backgrounds, but did identify them in wt and other mutants (see proteomics sections below). This genetic material was then used to systematically investigate *in vivo* relative protein abundance and the relative accumulation different Nt proteoforms and protein/peptide degradation fragments.

3.3.2 Molecular proteome phenotypes of *oop*, *prep1 prep2*, and *prep1 prep2 oop*

To determine the consequences of the loss of PREP and/or OOP peptidase capacity on the chloroplast and leaf proteomes that may explain the growth phenotypes, we carried out comparative proteome analyses of *oop*, *prep1 prep2*, the triple mutant and wt, organized in two separate experiments each in biological triplicate. For all experiments, wt and mutants were in the same developmental stage (1.11) as measured by the number of leaves per rosette (see for discussion on developmental effects on plant and chloroplast proteomes, (Rutschow et al., 2008)).

In the first experiment, we compared proteomes of isolated leaf chloroplasts of wt and the *prep1 prep2* to focus on chloroplast-specific proteome phenotypes. In the second experiment we compared total leaf proteomes of wt, *prep1 prep2*, *oop*, and the triple mutant. For all experiments, protein extracts were resolved by SDS-PAGE and complete gel lanes were sliced into sections (12 per gel lane) and proteins in-gel digested by trypsin and analyzed by LC-MS/MS. After database searching, processing and protein annotation, the data were analyzed statistically and through hierarchical clustering to determine genotypic effects. Across these experiments, 288 LC-MS/MS runs were carried out, resulting in half a million MS/MS spectra that could be

confidently matched to ~2500 non-redundant proteins after application of several stringent cut-offs and filters. Supplemental Figures 3.3 and 3.4 summarize the complete proteomics workflows, whereas Supplemental Table 3.1B and 3.2B summarize the identified proteins, their annotations, quantification and fold changes with the statistical significance levels (p-values and false-discovery rates) of differences between genotypes. Mass spectral data are available through ProteomeXchange (<http://www.proteomexchange.org/>) and matched peptides in PPDB. Protein annotations for name, function and subcellular location are from PPDB at <http://ppdb.tc.cornell.edu/> (updated May 2017).

3.3.3 The chloroplast proteome of *prep1 prep2*

We identified 913 proteins and sets of protein homologs (85 proteins assembled into 35 groups of homologs) of which 87% were confirmed plastid proteins representing 97% of the protein biomass in both wt and mutant (Supplemental Table 3.1A). Comparison of protein biomass per chloroplast sub-compartment did show a small (7%) reduction in thylakoid proteome and 9% increase in stromal proteome (Figure 3.1C). However, principle component analysis (PCA) based on adjusted spectral counts (adjSPC) or normalized adjusted spectral counts (NadjSPC) did not separate the wt from *prep1 prep2*, and pairwise correlation coefficients between the six samples were between 0.96 and 0.99 indicative of their general similarity (Supplemental Figure 3.3). The similarity between the chloroplast proteomes was also visible from an abundance cross-correlation for individual proteins in the two genotypes, with the exception of PREP1,2, absent in *prep1 prep2* (Figure 3.1D). When evaluated per process or pathway

(*e.g.* Calvin cycle), there were only relatively small differences (mostly within 10%). Differential genotypic accumulation of individual proteins (or groups of close homologs) was determined using the combined results of two different types of statistical analysis methods (QSPEC 5% FDR and GLEE $<p<0.01$); this identified four down-regulated proteins in the double mutant, namely PREP1,2, GRANULE BOUND STARCH SYNTHASE (GBSS; AT1G32900), PG-localized unknown protein 2 (AT3G43540), and Mg-protoporphyrin IX chelatase (AT5G13630; CHLH or GUN5) (Supplemental Table 3.1B). Applying only the QSPEC 5% FDR threshold, identified another 15 down-regulated thylakoid proteins (subunits of Photosystems I and II, the ATP-synthase and the cytochrome b6f complex, as well as FTSH2/8), six upregulated stromal proteins (within the Calvin cycle, a carbonic anhydrase, FD-GOGAT, and a subunit of the PEP RNA polymerase) and one down-regulated stromal protein (haloacid dehalogenase-like hydrolase-1) (Supplemental Table 3.1B); some of these are pointed out in Figure 3.1D. This is consistent with the general decrease of thylakoid- and increase of the stromal proteome mass.

The identified chloroplast peptidases and chaperones were evaluated for the potential impact of loss of PREP1,2 activity on the proteostasis network (Supplemental Table 3.1A). Surprisingly, none of the chloroplast peptidases or stromal chaperones (CPN10/20/60; HSP70, HSP90), the unfoldase CLPB3 nor the CLPC/D chaperones were significantly affected. This strongly contrasts our observations for several CLP core mutants in which all chloroplast chaperone systems increased many fold (Kim et al., 2009; Zybailov et al., 2009; Kim et al., 2013; Kim et al., 2015), reviewed in (Nishimura and van Wijk, 2015) (see below and DISCUSSION).

3.3.4 *Direct comparison of total leaf proteomes of wt, prep1 prep2, oop, and prep1 prep2 oop*

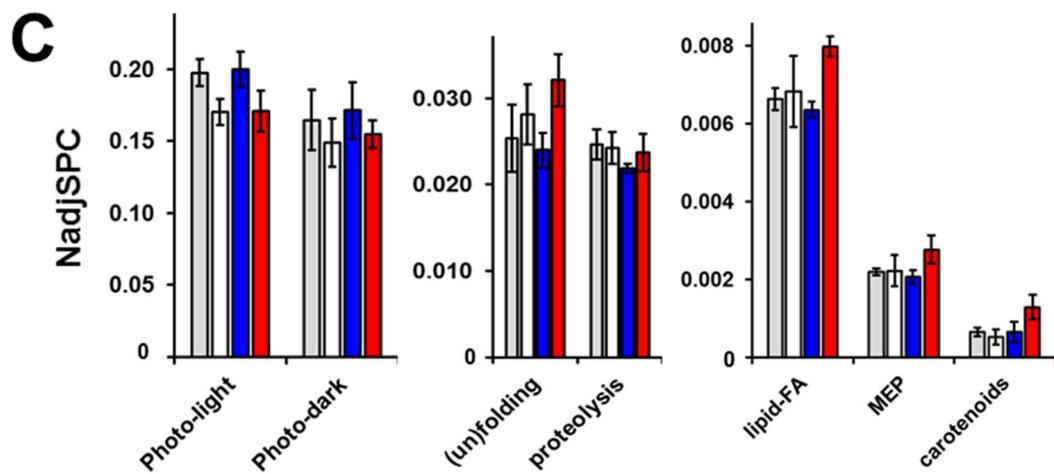
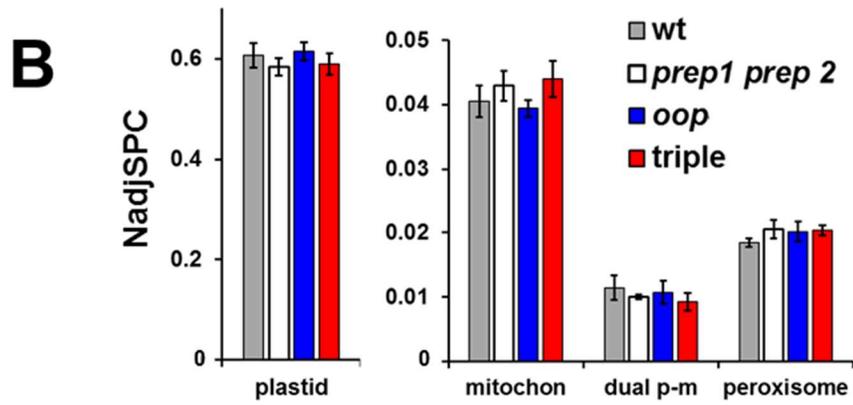
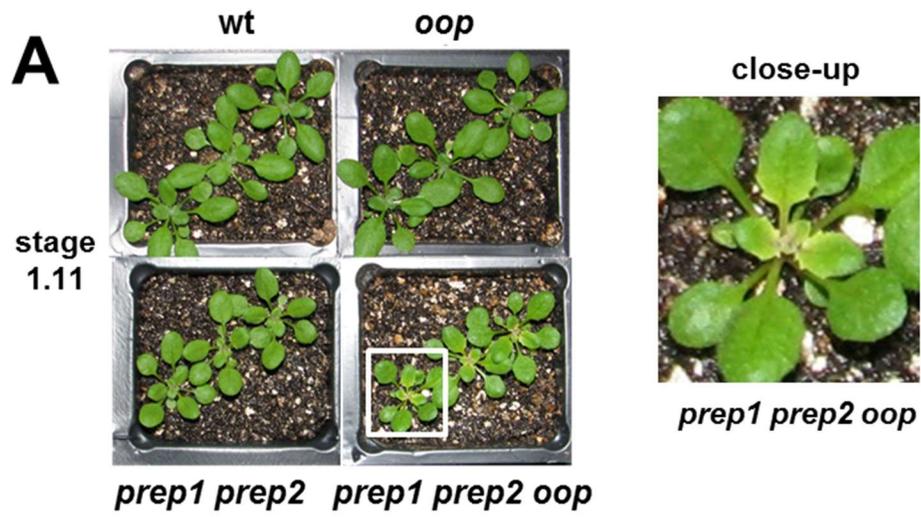
All four genotypes were in developmental stage 1.11 and were grown for 20 days (wt and *oop*), 22 days (*prep1 prep2*) or 28 days (*prep1 prep2 oop*) (Figure 3.2A) and total leaf proteomes analyzed by MS/MS using the workflow as summarized in Supplemental Figure 3.3A. In total 2548 proteins and protein groups were identified matching to nearly 330 thousand MS/MS spectra (Supplemental Table 3.2A). The total quantitative proteomes (based on NadjSPC) of the four genotypes were weakly distinguishable by PCA, and pairwise correlation coefficients between the six samples were between 0.91 and 0.97 (Supplemental Figure 3.4B,C). We identified more than 900 plastid proteins, 130 mitochondrial proteins and 50 peroxisomal proteins, as well as 37 proteins dual-localized in plastids and mitochondria, many of which were tRNA-synthases. Given that OOP and PREP1,2 are dually targeted to chloroplasts and mitochondria, we first evaluated the relative protein mass distribution (from NadjSPC) across plastids and mitochondria across the genotypes. Chloroplasts, mitochondria and peroxisomes represented 61%, 4.1% and 1.8% of the total protein mass, respectively, whereas dual-targeted plastid-mitochondrial proteins represented 1.1% (Figure 3.2B). Plastid proteomes in the *prep1 prep2* and the triple mutant were reduced by 3-4%, mostly due to reduced thylakoid proteins (lumen, integral and peripheral), whereas mitochondrial proteomes were increased by 6-9%; in contrast *oop* was similar to wt (Figure 3.2B). We then evaluated various plastid and mitochondrial functions (Figure 3.2C; Supplemental Figure 3.4). Investments in the photosynthetic electron transport

Figure 3.2. Quantitative comparison of total leaf proteomes of proteomics wt, *prep1 prep2*, *oop* and the triple mutant in developmental stage 1.11.

(A) Representative 1.11 stage plants of the four genotypes selected for the comparative leaf proteome analysis. The close-up shows the pale-green phenotype of the youngest emerging leaves of the triple mutant.

(B) Relative protein mass distribution of chloroplasts, mitochondria and peroxisomes, as well as dual-targeted plastid-mitochondrial proteins across the four genotypes. Standard deviations are indicated.

(C) Differential investments in specific chloroplast functions across the four genotypes as determined by relative protein abundance. Color coding as in panel B and standard deviations are indicated.



chain and the Calvin-Benson cycle were reduced in the *prep1prep2* and triple mutants by 14% and 6-10%, respectively. Several metabolic pathways, including for fatty acids/lipids, the MEP pathway and carotenoids (Figure 3.2C), relatively increased in the triple mutant, by 20%, 26% and 97%, respectively, but were essentially unaffected in *prep1 prep2* and *oop*. Evaluation of different plastid protein biogenesis and proteostasis functions showed no significant changes on ribosome content, translation initiation/elongation (not shown) or proteolysis (Figure 3.2C), but showed a 27% increase in stromal chaperone and protein isomerase protein mass in the triple mutant (Figure 3.2C). Most of these responses were also visible in the *prep1 prep2* mutant but were less pronounced than in the triple mutant.

Finally, we tested pair-wise significance for each protein or protein group between all genotypes. Fifty proteins showed one or more pairwise genotypic effects and these included PREP (significant loss in *prep1 prep2* and triple) and OOP (significant loss in *oop* and triple) (Supplemental Table 3.2B). The highest number of differentially accumulated proteins was between triple/wt and triple/*oop*, consistent with the visible leaf phenotypes. Nineteen were plastid proteins, two (PREP1/2 and OOP) were dual-localized across plastid and mitochondria, one was mitochondrial and the remainder were mostly cytosolic and/or nuclear-localized. Among the plastid proteins were six members of photosynthetic thylakoid complexes, reflecting their down-regulation in particular in *prep1prep2* and the triple mutant. The unfoldase CLPB3 was significantly upregulated in the triple mutant when compare to *oop*. Among the non-plastid/mitochondrial proteins, REDUCED CHLOROPLAST COVERAGE 1 (REC1) and REC2 are of particular interest in that they are involved in chloroplast biogenesis

and control of total chloroplast volume per cell (Larkin et al., 2016); REC1,2 were down-regulated in the triple mutant.

3.3.5 Comparison of Nt proteomes of wt, prep1 prep2 and prep1 prep2 oop using TAILS

To specifically recognize the ultimate N-termini (Nti) of mature chloroplast proteins or protein degradation products, we employed the TAILS technique in which protein Nti are chemically labeled by reductive dimethylation, either *light* or *heavy* (using different stable isotopes) to distinguish the sample from the control. Following trypsin digestion of proteins, Nti are enriched by removal of internal (trypsin generated) non-labeled peptides (Kleifeld et al., 2010). We compared the soluble Nt proteomes of total leaf extracts or isolated chloroplasts from wt with *prep1 prep2* or *prep1 prep2 oop* (the ‘triple mutant’). Each TAILS experiment consisted of three biological replicates and a stable isotope label switch; see Supplemental Figure 4.5 and Supplemental Table 3.3A for details. Average mutant to wt peptide ratios (log₂ scale) for each TAILS experiment are compiled in Supplemental Table 3.4 and 3.5. A summary of proteins with differentially regulated Nti is presented in Table 3.1. In addition to the relative peptide abundance (mutant/wt) ratios, we have included the total number of spectral counts matching to each peptide for each experimental comparison to provide a rough estimation of the total proteoform mass represented by each Nt peptide (Table 3.1). We did not include a total protein abundance ratio (determined in previous section) here since overall protein levels were hardly changed in the respective mutants.

Table 3.1. Differentially regulated protein Nti identified in *prep1 prep2* and *prep1 prep2 oop* TAILS experiments. Peptides were quantified by MS1 filtering using Skyline software (MacLean et al., 2010). Average ratios for three replicates are displayed with standard deviations. Mutant to wild type ratios for three independent experiments (see Supplemental Table 3.5) are displayed: *prep1 prep2* (stroma), *prep1 prep2 oop* (stroma) and *prep1 prep2 oop* (total leaf). (a) Predicted Nt residue position after cTP removal (ChloroP), (b) experimental observed Nt position, (c) acetylated Nt, (d) P1 residue immediately upstream protease cleavage site based on Schechter and Berger nomenclature, (e) mutant/wt ratio, Log 2 scale (f) total spectral counts matching to this peptide for both wt and mutant.

Table 3.1

Accession	Protein Description	Pred. Exp. Nt (a)	Exp. Nt (b)	Ac Nt (c)	P1 (d)	N-Terminal Peptide	prep1_prep2			prep1_prep2_oop			prep1_prep2_oop		
							log2 (Mu/WT) (e)	STDEV (f)	total SPC (f)	log2 (Mu/WT)	STDEV	total SPC	log2 (Mu/WT)	STDEV	total SPC
Proteins with cTPs and alternate Nt.proteoforms detected															
AT4G38970	SFBA-2	47	2	1	M	ASTSLLKASPVLDKSEWVKQSVLFR AASSYADELWKTAKTIASPGR ASSYADELWKTAKTIASPGR SSYADELWKTAKTIASPGR SYADELWKTAKTIASPGR	2.9 -0.1 -0.1 -0.1	(0.3) (0.3) (0.4) (0.3)	60 385 45 60	2.0 -0.1 0.1	(0.1) (0.2) (0.2)	64 363 49	3.3 2.0 -0.1	(0.3) (0.1) (0.1)	55 100 1036
AT5G38410	RBCS-3B	55	2	1	M	ASSMLSSAAVWTPAQATM/APFTGLKSSAAFPVTR TM/APFTGLKSSAAFPVTR	-0.1	(0.3)	60	0.1	(0.2)	49	-0.2 2.1 3.2	(0.2) (0.2) (0.9)	17 14
AT1G67090	RBCS-4	55	2	1	M	MKWVPIGKIKFETLSYLPDLSDELAKEDVYLLR ASSMLSSATM/SPAQATM/APFNGLKSSAAFPATR VASPAQATM/APFNGLKSSAAFPATR	-0.3 3.4	(0.1) (0.4)	94	-0.3 3.4	(0.1) (0.4)	94	-0.6 5.7 3.8	(0.1) (0.6) (0.3)	182 8 12
AT5G54770	TH11	46	2	1	M	MQVWPPIGKIKFETLSYLPDLTDELAKEDVYLLR TSLYLPDLTDELAKEDVYLLR SYLPDLTDELAKEDVYLLR TDELAKEDVYLLR SELAKEDVYLLR TDSAQVLKEVEECKKEYPNAFR AAIASTLSLSSSTKPKR ATTAGYDLNAFTFDPIKESIVSR ATTAGYDLNAFTFDPIKE TTAGYDLNAFTFDPIKESIVSR GYDLNAFTFDPIKESIVSR	-0.1 -0.2 -0.4 -0.2 -0.1 -0.2 3.0 -0.1 -0.3	(0.1) (0.1) (0.0) (0.1) (0.2) (0.0) (0.6) (0.3) (0.1)	88 106 82 71 32 64 6 50 42	-0.1 -0.2 -0.4 -0.2 -0.1 -0.2 3.4 -0.3 -0.2	(0.1) (0.1) (0.0) (0.1) (0.2) (0.0) (0.1) (0.1) (0.1) (0.1) (0.1) (0.1) (0.1) (0.1)	42 142 70	2.8 2.7 -0.5 -0.4	(0.2) (0.2) (0.1) (0.1) (0.1) (0.1) (0.1) (0.1) (0.1)	44 56 174 136
Proteins with alternate Nt.proteoforms detected															
AT4G18440	adenylosuccinate lyase	65	47		C	SVSTTTDTPKLVSTKVTAMDGVSSR	0.1	(0.6)	28	0.2	(0.1)	28	-1.3	(0.5)	15
AT4G24830	argininosuccinate synthase	74	74		R	VSTTTDTPKLVSTKVTAMDGVSSR AVLSGDGTALTTDSKEAGLR VLSGDGTALTTDSKEAGLR VLSGDGTALTTDSKEAGLR	1.2 5.0 0.0	(0.6) (0.7) (0.4)	12 10 30	1.5 4.4 0.1	(0.2) (0.6) (0.1)	26 36 82 46	1.2 4.5 -0.5	(0.7) (0.1) (0.1)	51 114
AT1G03475	coproporphyrinogen (III) oxidase	49	48		C	SVSIEKEVPETERPFTFLR SVSIEKEVPE VSIEKEVPETERPFTFLR	-0.1 -0.2 2.0	(0.3) (0.4) (0.3)	96 62 7	-0.2 -0.2 2.7	(0.1) (0.1) (0.0)	127 18	-0.3 2.4	(0.0) (0.2)	89 9
AT3G23940	dihydroxyacid dehydratase	36	35		C	SAGSVTADPSPPTDITNKLKLYSSR SAGSVTADPSPPTDITNKLKLYSSR AGSVTADPSPPTDITNKLKLYSSR	-0.1 0.3	(0.3) (0.1)	95 44	-0.2 1.4	(0.2) (0.2)	101 22	-0.2 2.0	(0.1) (0.1)	144 10

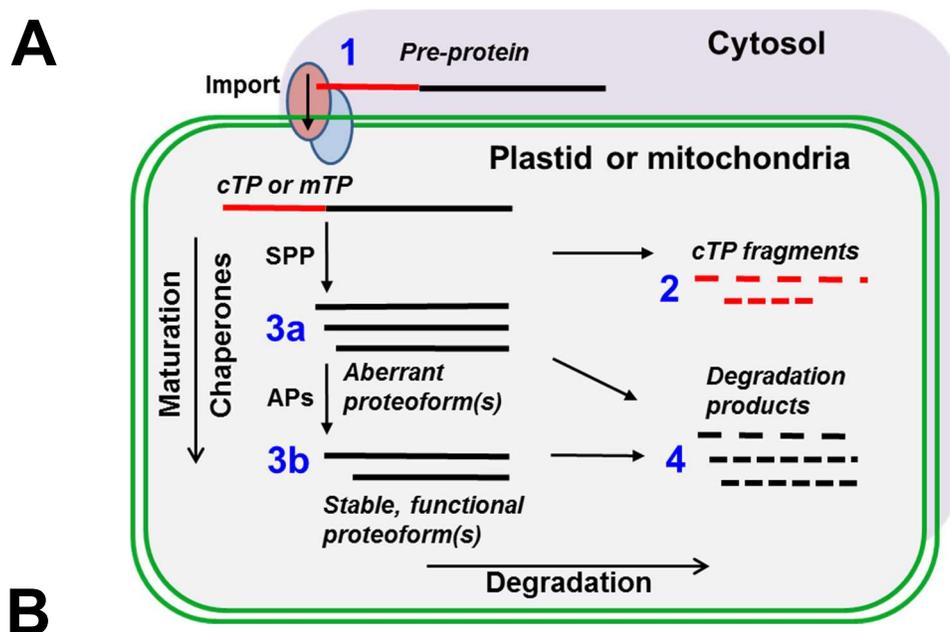
Table 3.1 (continued)

Accession	Protein Description	Pred. Exp. Mt (a)	Nt (b)	Ac Nt (c)	P1 (d)	N-Terminal Peptide	prep1_prep2			prep1_prep2_ooop			prep1_prep2_ooop		
							log 2 (Mu/WT) (e)	STDEV	total SPC (f)	log 2 (Mu/WT)	STDEV	total SPC	log 2 (Mu/WT)	STDEV	total SPC
AT1G24360	Ketoacyl-ACP Reductase	58	58		K	AQATATEQSPGEVWQKVESPVVMTGASR						2.3	(0.2)	2	
AT1G16080	unknown protein	42	42	1	M	ATATEQSPGEVWQKVESPVVMTGASR AIIAAASAATAKKLAPAMVGGGR AAASAATAKKLAPAMVGGGR AAASAATAKKLAPAMVGGGR AAASAATAKKLAPAMVGGGR	-0.6	(0.2)	30	-0.5	(0.0)	106	(0.0)	106	
AT5G51110	unknown protein	51	51		R	SNLAQDFLGDFFGAR	0.0	(0.3)	12	2.5	(0.3)	12	1.0	(0.1)	109
AT1G63970	MDS	53	49		S	ASVSAASSAVDWIESVTSEKPTKTLPR SYSAASSAVDWIESVTSEKPTKTLPR AASSAVDWIESVTSEKPTKTLPR ASSAVDWIESVTSEKPTKTLPR SSAVDWIESVTSEKPTKTLPR SAVDWIESVTSEKPTKTLPR AVDWIESVTSEKPTKTLPR	0.4	(0.5)	48	0.0	(0.2)	13	-0.6	(0.2)	23
AT2G37660	3-beta-hydroxy-delta5-steroid dehydrogenase	70	70		S	AAATTEPLTLVLTGAGGR	0.2	(0.4)	6	2.7	(0.5)	20	-0.1	(0.0)	68
AT4G17300	Asp-tRNA synthetase	64	60		C	TAVSESLGSGDGNKVESYEKR AVSESLGSGDGNKVESYEKR VSESLGSGDGNKVESYEKR	0.2	(0.4)	6	-0.1	(0.2)	66	2.9	(0.1)	56
AT1G51100	CRR41	39	40		C	SSNPESKQDFNLTPEAEKASEMLQPIR	0.3	(0.3)	20	0.0	(0.1)	62	-0.8	(0.1)	63
AT1G48860	EP SP synthase	40	72		A	SVSTAEKASEMLQPIR	0.2	(0.4)	8	2.0	(0.5)	6	-1.0	(0.1)	30
AT4G21210	protein kinase	87	19		S	NLNPNSKPKAGSDSVSLINASEPGSER	0.8	(0.4)	40	-0.6	(0.3)	18	-0.7	(0.1)	70
AT3G55600	SBP ase	60	75		C	EIGQSLEEFLAQATPKDKGLR	0.8	(0.4)	40	-0.4	(0.3)	2	-0.6	(0.1)	106
AT1G09340	HIP1.3	50	44		K	GALYVSASSEKKILMGTR	-0.1	(0.3)	58	-0.7	(0.0)	42	1.6	(0.0)	74
AT4G05180	psbQ OEC16	49	83		A	EAPIKVGGPPLPSGGLPGTDNSDQAR GGPPLPSGGLPGTDNSDQAR	1.0	(0.1)	26	0.9	(0.3)	74	1.6	(0.0)	74
AT1G05385	Psb27	65	68		V	SEEKLDGEGVWGAFAKTLDFPNER	0.5	(0.3)	14	-0.3	(0.0)	153	-0.4	(0.1)	150
AT3G27925	DEG1 protease	45	106		A	FWSTPKKLOTDELATVR	3.4	(0.2)	8	3.4	(0.2)	8	-0.1	(0.2)	108
AT1G17490	unknown protein (not plastid)	n/a	2	1	M	AEPKQSSLSEVKTNVGSRR	0.5	(0.2)	150	0.5	(0.2)	150	-0.1	(0.2)	108
AT3G22230	60S RPL27 (not plastid)	n/a	2		M	VKFLKQNKAVILLOGR	1.7	(0.3)	52	1.7	(0.3)	52	1.1	(0.1)	64
							1.2	(0.1)	8	1.2	(0.1)	8	1.1	(0.1)	64
							1.6	(1.7)	47	1.6	(1.7)	47	0.0	(0.2)	34
													0.6	(0.3)	149

To illustrate the different biological scenarios leading to various Nt sequences, and for the purpose of discussion, we have classified protein N-termini (Nti) as either (1) unprocessed pre-proteins with cTP attached, (2) cleaved and or fragmented cTPs, (3a) intermediate or aberrant proteoforms, (3b) primary, mature proteoforms, or (4) far downstream degradation products (Figure 3.3A). Experiments, tables and figures where these phenomena were observed are summarized in Figure 3.3B.

3.3.6 Unprocessed pre-proteins are detected in triple mutant total leaf extracts but not in stromal extracts

Nt peptides matching to the cTPs of the abundant stromal proteins fructose-bisphosphase aldolase 2 (SFBA-2), rubisco small subunit 3 and 4 (RBCS-3, 4) and thiazole biosynthetic enzyme (THI1) were clearly detected in total leaf proteomes of the triple mutant but were at or below the limit of detection in wt (Table 3.1; Supplemental Table 3.5). The other major Nt peptides matching to these proteins are displayed for comparison; the peptide matching to the primary Nt proteoform for each protein can be distinguished by the high number of spectral counts (SPC). cTPs were not readily detected in chloroplast stromal extracts (only the cTPs from RBCS-4 detected in the triple mutant) indicating that these cTPs were part of pre-proteins *en route* to the plastid (Class 1 Nti – Figure 3.3) and that cTPs in the stroma continue to be efficiently degraded by peptidases other than PREP1,2 and OOP (see DISCUSSION). We performed a number of follow-up experiments to determine if cTPs do accumulate inside the plastid (see below sections 3.3.9 - 3.3.12). The mature N-terminus (with cTP removed) of RBCS3 and 4 was slightly decreased in triple mutant stroma and total leaf experiments



Nt peptide class	Genotype	Method	Table	Figure
1* cTP as part of preprotein	<i>prep1 prep; triple</i>	TAILS, LMW-TAILS, label free SPC	1,2,	S2C
2 cTP fragments	<i>prep1 prep2; quadruple</i>	peptidomics	4	4
3a intermediate or aberrant proteoform	<i>all genotypes</i>	TAILS, LMW-TAILS, label free SPC	1,2,5	S2C
3b mature proteoform	<i>all genotypes</i>	TAILS, LMW-TAILS	1,2,5	S2C
4 degradation products	<i>prep1 prep2; triple</i>	LMW-TAILS, peptidomics	2,4	4

* Only detected in total leaf extracts and not in chloroplast (stromal) extracts

Figure 3.3. Pathways to proteolytic protein maturation and degradation. Classification of N-terminal (Nt) peptides detected in TAILS and or peptidome experiments. 1, cTP as part of pre-protein; 2, cTP fragments cleaved from protein; 3a, intermediate or aberrant Nt proteoforms; 3b, fully mature stable Nt proteoforms; 4, degradation products downstream of mature Nt. SPP, stromal processing peptidase; APs, amino peptidase potentially involved in processing of immature proteoforms.

relative to wt, in line with little to no change observed in our comparative proteomics experiments above.

A number degradation products (Class 4 Nti – Figure 3.3) for RBCS-4 were also observed and their relative abundances were similar to that of the mature N-terminus for this protein, indicating that these fragments do not accumulate in the triple mutant background.

3.3.7 Altered maturation or processing of n-encoded plastid proteins in prep1 prep2 and the triple mutant

The most striking feature of the mutant Nt proteomes was the occurrence of altered protein processing or maturation (Peptide Class 3a, Figure 3.3). Alternate Nti (different from the primary Nt proteoform) with exposed Nt Ala made up the majority of over-accumulating peptides in the mutants followed closely by Nt Val (Table 3.1). For the protein THI1, unchanged at the protein level, the most abundant and presumably native N-terminus (TTAGY....) was reduced in the mutant to 70% of wt level, whereas the form with an additional alanine (ATTAGY...) was at least 8-fold up regulated. A very similar scenario was observed for chloroplast glutamine synthetase (GS2; AT5G35630), inorganic phosphatase like (AT5G09650) and arginosuccinate synthase (AT4G24830) (Table 3.1). Interestingly, the residues exposed when Ala is removed are different for these three proteins (Leu, Ile and Val) indicating that it is indeed the Ala that is disfavored at the N-terminus of these proteins in wt. However, we know that many highly abundant stromal proteins start with Ala, for example SFBA-2, phosphoglycerate kinase-1 (PGK-1) and transketolase-1 (TKL-1) (Table 3.1). Indeed

Ala is the most common start residue for chloroplast proteins (Rowland et al., 2015), so the removal, or not, of Ala from the N-terminus of proteins is context dependent (see DISCUSSION). In the case of chloroplast heat shock protein 70-1 (cpHSP70-1), a proteoform with an additional three residues (VVNEKV...) was 4-fold and 16-fold up in the double and triple mutants respectively, although the normal mature N-terminus (EKVVGI...) was unchanged (Table 3.1). Glyceraldehyde-3-phosphate dehydrogenase B (GAPB) had an N-terminus with an additional 14 residues upstream of the mature N-terminus (KLKVAI...) (Rowland et al., 2015), that was elevated in the *prep1 prep2* mutant. Please note that these aberrant Nti generally represent only a fraction of the overall proteoforms accumulating in these mutants. Only 3 and 7% of the total spectral counts in stroma for cpHSP70-1 Nti were from the aberrant form in *prep1 prep2* and the triple mutant respectively (Table 3.1).

Proteins with truncated N-termini (shorter than the major form) were also elevated in the mutant (Peptide Class 3a, Figure 3.3). The major Nt form of coporphrynogen III oxidase (AT1G03475) starts with the residues SVSIEK... and was reduced by 20% in the mutant, whereas the proteoform lacking the Nt Ser (VSIEK...) was at least four fold elevated in both mutants. A similar scenario was observed for dihydroxyacid dehydratase (AT3G23940) and AT4G17300. 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MDS) had seven unique Nti observed in similar proportions (based on SPC across this experiment) forming a ladder of alternate proteoforms with both longer and shorter forms being elevated (Table 3.1) in *prep1 prep2* stroma.

3.3.8 Mitochondrial proteins processed by ICP55 are not influenced by lack of PREP and OOP

The aminopeptidase ICP55 was previously shown in the mitochondria of yeast and plants to aid in Nt maturation of proteins following removal of their mTP by MPP (Vogtle et al., 2009; Carrie et al., 2015). ICP55 primarily cleaves the peptidyl bond C-terminal of Tyr, Phe, Leu or Ile (and sometimes two residues in a row). We wondered whether the absence of PREP and OOP in mitochondria may influence these processing events, and therefore evaluated the total leaf TAILS dataset for mitochondrial peptides. We identified Nt peptides for 11 out of the ~70 previously observed ICP55 substrates. Two were dually localized to plastids and mitochondria; AT4G32915 had an Nt two residue upstream of the mature mitochondrial processed form (SYSSD... where cleavage C-terminal of Y normally occurs by ICP55); Phe-tRNA synthetase (AT3G58140) had a distinct plastid processed Nt that is different from the mitochondrial form. The other mitochondrial proteins all had Nti matching the expected mature, ICP55 processed form. Additionally, we found the partial immature Nt of CLPP2 with an additional Tyr (YSLPM...). The relative accumulation for eight Nt peptides from these proteins was not significantly changed in the triple mutant, including the immature form of CLPP2 (Supplemental Table 3.6).

3.3.9 Accumulation of protein degradation products in prep1prep2 and triple mutants. Low molecular weight TAILS reveals degradation products accumulate in the triple mutant

There was little evidence for differential accumulation of degradation products or cleaved cTPs in the above TAILS experiments (see Figure 3.3B), perhaps due to inadequate sensitivity and low dynamic range in the presence of highly abundant RBCL (53 kDa) and other large proteins. Therefore, we removed proteins larger than 10 kDa or 30 kDa using filtration devices, followed by Nt peptide enrichment (this section) or direct MS analysis (section 3.3.11). Additionally, for the triple mutant, we labeled protein Nti, resolved the low molecular weight proteome by SDS-PAGE and performed in gel digestion and MS analysis (section 3.3.10). We discuss these experiments and their main findings briefly with more detailed information in Supplemental Tables 3.5, 3.7, 3.8 and summarized in Tables 3.2 and 3.3.

Total soluble leaf proteomes from the triple mutant or wt were compared using a low molecular weight (LMW)-TAILS variant (<30 kDa) (Supplemental Figure 3.5A). The Nt fragments corresponding to cTPs identified in Table 3.1 were again found to accumulate in the triple mutant relative to wt (Table 3.2). The unprocessed molecular weights for these proteins are 20, 37 and 43 kDa for RBCS, THI1 and SFBA-2 respectively, so it is feasible that at least the first three and perhaps some of the fourth of these highly abundant proteins passed through the filtration device. Therefore, we suggest these are again Peptide Class 1 (Figure 3.3). No other cTP fragments were detected.

Importantly, a number of identified Nt peptide fragments corresponding to degradation fragments (Peptide Class 4, Figure 3.3) were strongly upregulated in the triple mutant. Fragments of GUN5 (AT5G13630), RH3 (AT5G26742), ketol-acid reductoisomerase (AT3G58610) and various ribosomal subunits accumulated in the

Table 3.2. Differentially regulated protein N-termini identified in *prep1 prep2 oop* low molecular weight TAILS experiments. Peptides were quantified by MS1 filtering in Skyline (MacLean et al., 2010). Average ratios, log₂ (mutant/wt), for three replicates are displayed with standard deviations and total spectral counts for that peptide (sum of wt and mutant). (a) localization: s, stroma; t, thylakoid (b) Predicted Nt residue position after cTP removal (ChloroP), (c) experimentally observed Nt position, (d) number of residues from Nt of peptide to C-terminus of protein (e) acetylated Nt, (f) P1 residue immediately upstream protease cleavage site based on Schechter and Berger nomenclature, (#) degradation product with Nt position far downstream of mature protein N-terminus.

Table 3.2

Accession	Protein Description	Loc. (a)	Pred. Nt (b)	Exp. Nt (c)	Res. from C-term. (d)	Ac Nt (e)	P1 (f)	N-Terminal Peptide	<i>prep1 prep2 oop</i> total leaf		
									log 2 (Mu/WT)	STDEV	total SPC
AT4G38970	SFBA-2	s	47	2	396	1	M	ASTSLLKASPVLDKSEWVKQSVLFR	2.9	(1.3)	9
				47	351	R	AASSYADELVKTAKTIASPGR	2.1	(0.2)	30	
AT5G38410	RBCS-3B	s	55	2	179	1	M	ASSMLSSAAVWTPAQATMVAPFTGLKSSAAFPVTR	4.3	(2.6)	2
				19	162	A	TMVAPFTGLKSSAAFPVTR	3.4	(0.9)	22	
AT1G67090	RBCS-4	s	55	2	178	1	S	ASSMLSSATMVASPAQATMVAPFNGLKSSAAFPATR	3.2	(2.5)	10
AT5G38430	RBCS-1b	s	55	56	125		C	MKWVPPIGKKKFETLSYLPDLTDVELAKEVDYLLR	0.2	(0.6)	34
AT5G54770	THI1	s	46	2	347	1	M	AAIASTLSLSSTKPQR	1.4	(0.7)	2
				46	303	R	ATTAGYDLNAFTFDPIKESIVSR	3.0	(0.1)	20	
ATCG00490	RBCL			65	414		G	TWTTWWDGLTSLDR	-0.5	(0.1)	44
				232	247	E	TGEIKGHYLNATAGTCEEMIKR	0.2	(1.0)	14	
				406	73	G	TLGHPWGNAPGAVANR	-0.5	(0.2)	18	
				9	34	445	D	TDILAAFR	-0.4	(0.1)	29
AT3G60750	TKL-1	s	66	66	675		R	AAAVETVEPTDSSIVDKSVNSIR	4.0	(0.3)	24
				67	674	A	AAVETVEPTDSSIVDKSVNSIR	-0.2	(0.1)	38	
				68	673	A	AVETVEPTDSSIVDKSVNSIR	0.9	(0.2)	38	
AT5G51110	unknown protein	s	51	52	168		S	NLAQDFLGDGFGAR	-0.5	(0.1)	32
AT1G54580	ACP	s	52	53	83		C	AAKPETVDKVC AVR	-0.6	(0.1)	24
AT5G26742	DEAD box RNA helicase (RH3)	s	61	426	322		Y	ELPNDPETFVHR #	5.8	(0.7)	16
AT4G04020	fibrillin 1a (FBN1a)	s	56	56	262		R	ATDIDDEWQQDGVER	0.8	(0.1)	22
AT5G13630	CHLH; GUN5	s	87	1111	270		E	LDEPVEQNFVR #	4.2	(0.7)	18
				1112	269	L	DEPVEQNFVR #	2.3	(1.3)	7	
				1227	154	D	SDPTNLVQSLR #	1.8	(1.4)	15	
AT4G10300	DUF861		38	39	95		A	ESTEKLGITIEKNPPESKLTQLGVR	-0.6	(0.1)	64
AT3G13120	30S RPS10	s	79	57	134		F	PDTLDPTPEILDEPASEVPSSSSISVDADKMAPKQKIR	0.6	(0.1)	3
ATCG01120	30S RPS15	s	18	2	86		M	IKNMFEEQKEESR	0.2	(0.3)	81
AT3G27160	30S RPS21	s	48	82	101		F	SSGYNVQVFVEDNESEER #	0.9	(0.2)	22
ATCG00770	30S RPS8	s	16	2	132		M	GKDTIADIITSIR	0.9	(0.4)	16
AT2G43030	50S RPL3	s	50	97	174		A	TDGYDAVQIGYRR #	4.4	(1.2)	9
				97	174	A	TDGYDAVQIGYR #	5.4	(0.9)	28	
AT1G05190	50S RPL6	s	59	179	44		G	YDKSEIGQFAATVR #	2.9	(0.8)	10
AT4G24830	argininosuccinate synthase	s	74	74	420		R	AVLSGDGTALTTDSKEAGLR	2.7	(1.1)	10

Table 3.2 (continued)

Accession	Protein Description	Loc. (a)	Pred. Nt (b)	Exp. Nt (c)	Res. from C-term. (d)	Ac Nt (e)	P1 (f)	N-Terminal Peptide	<i>prep1 prep2 oop</i> total leaf		
									log ₂ (Mu/WT)	STDEV	total SPC
AT5G45390	CipP4	s	61	244	48		Y	GLIDGVIDGDSIIPLEVPDR #	6.1	(2.3)	14
AT1G03475	coproporphyrinogen III oxidase	s	49	49	337		S	VSIEKEVPETERPFTFLR	2.4	(0.1)	7
AT4G24280	cpHSP70-1	s	93	75	643		R	VWNEKWGIDLGTNSAVAAMEGGKPTVTNAEGQR	3.6	(1.0)	6
				78	640		N	EKWVGDIDLGTNSAVAAMEGGKPTVTNAEGQR	0.2	(0.1)	84
AT2G04030	cpHSP90	s	61	62	718		C	DAVAEKETTEEGSGEKFEYQAEVSR	n/a	(1.2)	n/a
AT5G20720	Cpn21	s	51	51	202		K	AASVWAPKYTSIKPLGDR	4.6	(0.2)	19
AT4G20360	EF-Tu-1	s	68	100	376		L	TMALASIGSSVAKKYDEIDAAPEER #	1.1	(0.6)	37
AT5G35630	GS2	s	46	51	379		L	ALQSDNSTVNRVETLLNLDTKPYSDR	3.4	(0.3)	17
				152	278		D	TWTPAGEPIPTNKR	3.3	(1.1)	14
AT4G02520	glutathione transferase ATGSTF2	s	34	2	210		M	AGIKVFGHPASIATR	0.8	(0.5)	8
AT1G42970	GAPB	s	46	427	20		S	GDPLEDFCKTNPADDEECKVYD #	1.2	(0.8)	8
				432	15		E	DFCKTNPADDEECKVYD #	1.6	(0.4)	12
AT5G09650	Inorganic phosphatase like	s	67	59	241		S	AYNPQVKVQEEGPAESLDYR	3.4	(0.3)	8
				60	240		A	MNPQVKVQEEGPAESLDYR	-0.5	(0.1)	66
AT3G58610	ketol-acid reductoisomerase	s	71	566	25		A	QLRPTVDISVPADADFVRPELR #	3.6	(1.5)	14
AT1G11430	MORF9; DAG-related	s	59	182	50		N	GEIIPCTYPTYQPKQR #	3.2	(0.9)	15
				183	49		G	EIIPCTYPTYQPKQR #	4.0	(0.6)	11
AT4G21210	phosphatase/ protein kinase	s	87	19	384		S	NLNPNSKPAGSDSVSLNASEPGSER	2.4	(0.3)	8
AT1G68590	PSRP-3A	s	57	59	107		E	TVTGDTSNDTPQQTIKWKPKDEKSR #	5.0	(0.8)	9
AT2G39730	Rubisco activase	s	59	71	403		R	WRGLAYDTSDDQQDITR	1.9	(0.3)	22
AT2G39730	Rubisco activase			190	284		M	SAGELESGNAGEPAKLIR	1.2	(0.4)	22
ATCG00480	CF1b - atpB	t	14	113	385		N	VLGEPVDNLGPVDTR #	6.2	(0.9)	30
				115	383		L	GEPVDNLGPVDTR #	4.8	(1.2)	8
AT5G23120	HCF136 Tat ltp	t	61	79	324		A	DEQLSEWER	0.3	(0.2)	32
AT4G31560	HCF153 - biogenesis cytb6	t	49	113	24		E	GDISLDDVIQEPVLQR #	4.3	(2.3)	12
AT3G61470	LHCI-2.1	t	45	189	68		W	FDPLGWGSGSPAKLKELR #	4.5	(1.0)	20
AT5G01530	LHCI-4.1-CP29	t	41	92	198		Q	FDIDSLDQNLAKNLAGDVIGTR #	5.5	(1.0)	24
		t	41	94	196		D	IDSLDQNLAKNLAGDVIGTR #	3.2	(2.8)	18
ATCG00350	psaA - subunit Ia	t	22	2	748		M	IIRSPPEVKILVDRDP	3.5	(1.1)	7
				51	699		H	ADAHDFDSDSHTDLEEISR #	3.1	(0.6)	20
AT4G12800	psaL - subunit XI	t	51	161	58		E	GEPSIAPSLTLTGR #	3.4	(1.0)	6
ATCG00270	psbD D2	t	83	332	21		A	AQDQPHENLIFPEEVLPR #	2.2	(1.0)	10
AT5G66570	psbO OEC33	t	30	248	84		R	GGSTGYDVALPAGGRGDE #	6.0	(1.6)	10
AT4G05180	psbQ OEC16-like Tat ITP	t	49	84	146		E	AIPIKVGPPPLPSGGLPGTDNSDQAR	4.4	(0.7)	12
				90	140		V	GGPPLPSGGLPGTDNSDQAR #	2.1	(0.3)	30
AT4G28660	psbW -like	t	65	84	114		M	MVKPALQFIQGTDEMTPDKLTR	-0.4	(0.1)	98

mutant samples but were absent or below the limit of detection for wt (Table 3.2; Supplemental Table 3.6). This is quite different from the pattern observed in the whole proteome TAILS experiments above where Class 4 peptides were not generally found to be upregulated; see RBCS-4 and TKL-1 examples in Table 3.1. It should be noted that Nti representing degradation products detected for these proteins were many (10 to 100) times less abundant than the major mature proteoforms, explaining why these went undetected in the TAILS experiments without enrichment of LMW species. Thus, removal of primary Nt proteoforms by filtration, greatly enhanced our ability to detect and quantify degradation products. Interestingly, fragments of RBCL were not elevated in the mutant (Table 3.2) suggesting that not all degradation products are sensitive to the absence of PREP and OOP.

Eleven of the upregulated Nt were less than 70 residues from the C-terminus putting them in the range of substrate lengths accessible to PREP (Teixeira et al., 2017). Exact substrate/protein fragment lengths in the original sample cannot be determined by Nt proteomics because proteins are digested with trypsin. We discuss this in relation to the wider chloroplast peptidase network below (see DISCUSSION).

In summary, the triple mutant does over-accumulate a variety of protein degradation products *in vivo* as compared to wt and many of these peptides may be direct substrates of PREP (or OOP).

3.3.10 Low molecular weight tricine gels reveal subtle changes in accumulation of degradation products and confirms maturation defects

In a separate effort to detect protein fragments accumulating in triple mutant plants, a gel-based approach was taken, whereby the dimethyl labeled low molecular weight stromal proteome was resolved by tris-tricine SDS-PAGE, followed by trypsin in-gel digestion and MS/MS analysis. 66 unique Nt peptides were identified, primarily in the upper gel bands (Supplemental Table 3.7). Any peptides present in the lower bands, near the dye front of the gel were likely washed away during the in-gel digest procedure. (A second gel comparing wt and mutant stroma extracts was fixed with 5% glutaraldehyde before silver staining but this did not visualize any obvious LMW fragments (Supplemental Figure 3.5A). One dimethyl labeled Nt fragment from RBCS1 and RBCS-2 (YLPDLTDSSELAKEVDYLIR and YLPDLSDELAKEVDYLLR) was 2-fold elevated in the triple mutant while other fragments from these proteins were unchanged (Supplemental Figure 3.5B). If RBCS was cleaved once before Tyr 72, that would leave a 109 residue C-terminal fragment. Perhaps PREP normally contributes to the removal of this fragment. We also confirmed that the Ala extended Nt of CPN10 was elevated in the mutant (Supplemental Figure 3.5C).

3.3.11 Peptidomics demonstrate that cTPs and degradation products accumulate for the prep1 prep2 double mutant

Having shown that degradation products accumulate in the triple mutant and because much of the molecular phenotype seems to be due to PREP, we focused on the *prep1 prep2* double mutant. We used a peptidome approach to purify endogenous peptide fragments from total soluble leaf or chloroplast extracts (Fricker et al., 2012; Wu et al., 2015). In brief, we heated and or acidified protein extracts resulting in protein

aggregation, followed by filtration through 10 or 30 kDa filtration devices to collect peptides that were subjected to LC-MS/MS analysis without digestion (see MATERIALS and METHODS and Supplemental Methods). We compiled results from four independent experiments, identifying 1677 unique peptides corresponding to 467 proteins (Supplemental Table 3.8). This experiment is technically challenging because of many other small molecules that are co-purified with small peptides, limiting chromatographic separations, consuming much of the ion current and limiting the time spent fragmenting peptides.

Two thirds of the matched spectra were from chloroplast proteins. Peptide distributions between wt and mutant tissues are shown in Figure 3.4A. On average, a greater number of peptides were identified in mutant tissue than in wt. Interestingly, 25% of matched spectra were from thylakoid luminal proteins, especially plastocyanin (PC-1 and 2). This may be due to higher stability of peptides derived from these proteins in the acidic extraction buffer. There were similar numbers of spectra matching to luminal proteins from wt and mutant plants (1364 and 1228 respectively), indicating that these peptide fragments are not genotype specific. We compared the number of spectra matching to proteins from different sub-chloroplast compartments and protein functional classes (Figure 3.4 B). While there were similar counts for peptides from thylakoid proteins exposed to the lumen and stroma, integral thylakoid proteins were far more abundant in mutant tissue, primarily due to elevated LHC fragments (also enriched in the mutant for the LMW TAILS experiment above). Proteins localized to the stroma and ribosome were increased by 40% in the mutant

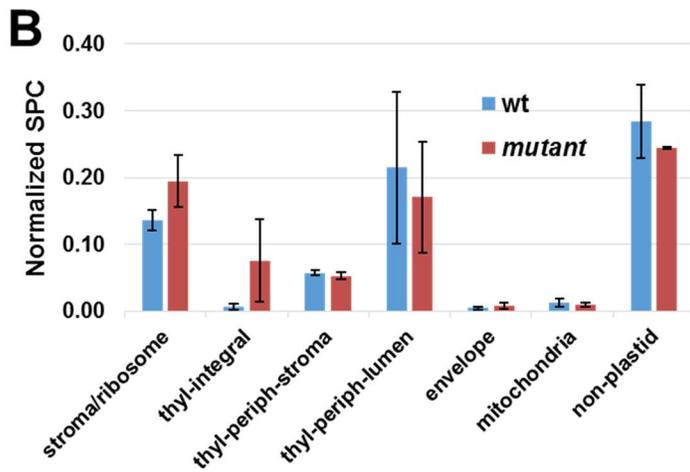
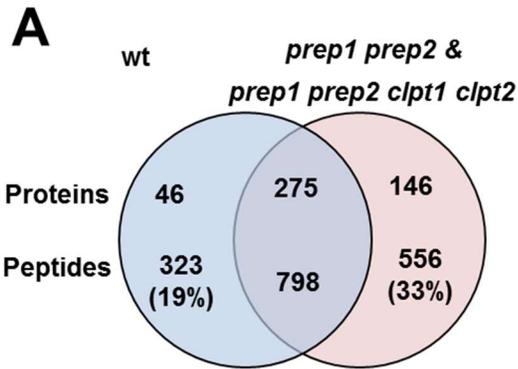
Figure 3.4. Comparison of wt and *prep1 prep2* peptidome.

(A) Peptide and protein genotype distributions across pooled peptidome experiments; Supplemental Table 3.8. Sum of all observed unique peptides and proteins across all peptidome experiments. Counts from single quadrupole mutant experiment pooled with those from *prep1 prep2* mutant.

(B) Normalized average spectral counts (SPC) for across different cellular compartments for wt and *prep1 prep2* from three independent experiments utilizing total soluble leaf extracts. Experiments 1, 2 and 3 in Supplemental Table 3.8.

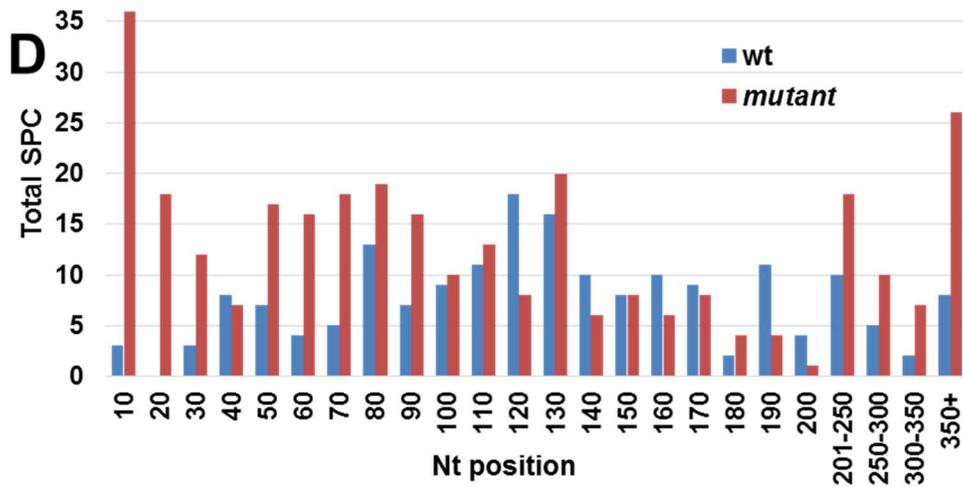
(C) Sum SPC for select plastid proteins and protein groups across all peptidome experiments.

(D) Nt position of identified peptides, binned in groups of 10 or 50, vs. summed SPC in wt and PREP mutants.



C

Protein or Group	wt	prep1 prep2
D1	0	42
tetrapyrroles	0	28
THI1	0	11
RBCS	9	75
RBCL	70	155
ribosomes	62	132
HSP	68	140
LHCs	13	95
psba/b	927	1071
atp synthase	46	90



relative to the wt. These results can be explained in part by the presence of protein fragments from a few abundant proteins. Peptides matching to D1, THI1, CHLI-1 and CHLH were all found exclusively in mutant samples (Figure 3.4C, Supplemental Table 3.8). These proteins were recently shown to have high turnover rates (Li and Millar 2017) and our results suggest that PREP1,2 are involved in their turnover.

Peptides from LHCs were primarily detected in mutant samples, as were fragments of RBCS and these fragments generally match to the cTP region of these proteins (Table 3.3). Additionally we detected cTPs from THI1, SFBA-2 and PC-1 and RCA. Because the majority of experiments utilized total leaf tissue we cannot be certain that these peptides are inside the chloroplast or if they were part of pre-proteins *en route* to the plastid (as discussed above). However for certain cases such as for peptides from RBCS (Table 3.3), the peptides all have C-termini immediately upstream of the known cTP cleavage site, indicating that these peptides are accumulating inside the plastid (Class 2 peptides, Figure 3.3). The starting position of each chloroplast targeted peptide was mapped across each protein and the counts for both mutant and wt plant was plotted against the starting position (Figure 3.4D). A clear preference was observed for peptide fragments falling within the first 30 residues of plastid proteins in *prep1 prep2*.

In order to confirm that these peptide fragments are indeed inside the plastid and are directly related to absence of PREP and or OOP, we analyzed whole chloroplast extracts (see MATERIALS and METHODS and Supplemental Methods). These extracts yielded low numbers of identified peptides, presumably because of extremely low free peptide concentrations in chloroplasts. We were able to identify

cTP fragments from LHCII-2.2, LHCII-3 and PC1 exclusively in the mutant chloroplast extracts, although with few spectral counts (Table 3.3). Because we analyzed these samples directly with any chemical modification or digestion, this demonstrates that peptide fragments accumulate in *prep1 prep2* mutants *in vivo*.

Table 3.3. cTPs detected in peptidome in *prep1 prep2* and *prep1 prep2 clpt1 clpt2 (aabbccDd)* total leaf and chloroplast extracts. See Supplemental Table 3.3B for experimental details and Supplemental Table 3.8 for complete peptidome data set.

Table 3.3. cTPs detected in peptidome

Accession	Protein Description	Exp. Nt ^b	Ac Nt ^c	Nt P1 ^d	Peptide	Ct P1 ^e	Tryp. Exp. ^f	wt ^a				prep1 prep2 ^a						
								1	2	3	4	1	2	3a	3b #	4		
AT3G56940	CHL27	2	1	M	AAEMALVKPISK	F							1					
AT1G60950	Fd2-leaf	2	1	M	ASTALSSAIVGTSFIR	R	1									2		
AT3G54050	FBPA, highCEF1	2	1	M	AATAATTTSSHLLSSSR	H	1										3	
AT4G38970	SFBA-2	2	1	M	ASTSLLKASPVLDKSEWVKGQ	S											1	
AT4G25100	FSD1	2	1	M	AASSAVTANYVLKPPPFALDALEPHMSK	Q	1		1									
AT2G05100	LHCII-2.1	2	1	M	ATSAIQQSSFAGQTALKPSN	E							2					
		2	1	M	ATSAIQQSSFAGQTALKPSNELLR	K	1										1	
		2	1	M	ATSAIQQSSFAGQTALKPSNELLRK	V	1										1	
AT2G05070	LHCII-2.2	2	1	M	ATSAIQQSSFAGQTALKPSSDLIQKVGVLGGGRVTM	R											3	
		2	1	M	ATSAIQQSSFAGQTALKPSSDLIQK	V	1										1	
AT3G27690	LHCII-2.3	2	1	M	ATSAIQHSSFAGQTTLKPSN	D							3					
AT5G54270	LHCII-3	2	1	M	ASTFTSSSSVLTPTTFLGQTKASSFNPLRDVSLGSPKYTM	G											2	
AT5G01530	LHCII-4.1-CP29	2	1	M	AATSAAAAAASSIMGTRVAPGIHPGSGRFTA	V							2	5	5			
		2	1	M	AATSAAAAAASSIMGTR	V	1								2			
AT3G08940	LHCII-4.2-CP29	2	1	M	AATSTAAAASSIMGTR	V	1										1	
AT1G15820	LHCII-6-CP24	2	1	M	AMAVSGAVLSGLGSSFLTGGKRGATA	L							1					
AT1G20340	PC-1	2	1	M	ASVTSATVAIPSFTGLK	A	1								1			
		45		S	SLKNFGVAAVAAAAASIALAGNAM	A											2	
AT4G27440	PORB	2	1	M	ALQAASLVSSAFSVR	K	1										1	
AT4G02770	psaD-2	2	1	M	ATQAAGIFNSAITTAATSGVK	K	1		2						1	4		
AT4G28750	psaE-1	2	1	M	AMTTASTVFLPANVTSVAGASSSR	S	1										1	
AT1G31330	psaF	2	1	M	SLTIPANLVLNPRSNKSLTQSVPKS	S							2					
		2	1	M	SLTIPANLVLNPRSNKSLTQ	S											1	
AT5G66570	psbO OEC33	2	1	M	AASLQSTATFLQSAK	I	1									1	1	
AT5G38420	RBCS-2b	32		A	SFPVTRKANNDITSITSNGGRVSC	M							6	4				
		40		A	NNDITSITSNGGRVSC	M											9	
		41		N	NDITSITSNGGRVSC	M											4	
AT5G38410	RBCS-3b	2	1	M	ASSMLSSAAVWSPAQATMVAPFTGLK	S	1								2	3		
AT1G67090	RBCS-4	2	1	M	ASSMLSSATMVASPAQATMVAPFNGLK	S	1									1	1	
		2	1	M	ASSMLSSATMVASPA	Q												1
		26		N	GLKSSAAFPAATTRKANNDITSITSNGGRVNC	M											1	
		29		K	SSAAFPAATTRKANNDITSITSNGGRVNC	M											2	
		30		S	SAAFPAATTRKANNDITSITSNGGRVNC	M											2	
		32		A	AFPATTRKANNDITSITSNGGRVNC	M									6	1		
		40		A	NNDITSITSNGGRVNC	M											2	
		41		N	NDITSITSNGGRVNC	M											5	
AT5G54770	THI1	2	1	M	AAIATLSLSSTKPQR	L	1										6	

Table 3.3 (Continued)

Accession	Protein Description	Exp. Nt ^b	Ac Nt ^c	Nt P1 ^d	Peptide	Ct P1 ^e	Tryp. Exp. ^f	wt ^a				<i>prep1 prep2</i> ^a			
								1	2	3	4	1	2	3a	3b #
AT2G39730	RCA	10			AINRAPLSLNGSGSGAVSAPASTFLGKKVTVS	R						4	2		
		13			RAPLSLNGSGSGAVSAPASTFLGKKVTVS	R						5			
		19			NGSGSGAVSAPASTFLGKKVTVS	R						1	3		
		66			TDGDRWRGLAYDTSDDQQDITRG	K						3	6	2	
		73			GLAYDTSDDQQDITRG	K	1							5	
		76			YDTSDDQQDITRGKGMVDS	V								4	
		190			SAGELESGNAGEPAKLIRQ	R								2	
		227			INDLDAGAGRMGGTTQYVNN	Q								3	

^a Sum SPC for independent experiments, each having multiple replicates.

^b Experimentally observed position of peptide Nt within parent protein

^c Acetylated Nt

^{d,e} P1 residue upstream of N-terminal protease cleavage or P1' residue downstream of C-terminal cleavage site, based on Schechter and Berger nomenclature

^f Peptide identified in experiment with trypsin digest

[#] Single experiment with *prep1 prep2 clpt1 clpt2 (aabbccDd)* total leaf tissue.

3.3.13 Mining spectral counting datasets for peptide phenotypes

Our initial proteome analysis of *prep1 prep2*, *oop* and the triple mutant focused on changes at the protein level (section 3.3.4). We wondered if any peptide level observations above could be further confirmed in this dataset. For each peptide, we compared the total number of observations per genotype. Peptides were also evaluated for their relative position to the start of the protein and to the predicted or experimentally determined mature N-terminus (ChloroP; Emanuelsson 2007). We then looked for peptides from chloroplast proteins that fell upstream of the cTP cleavage site. Peptides within the cTPs of RBCS, THI1, RPL12-A and PrxII E were detected only in *prep1 prep2* and triple mutants (Supplemental Table 3.2C). Interestingly, three of those peptides bridged the normal cTP cleavage site, again showing that pre-proteins

accumulate in PREP mutants. Two peptides upstream of the major mature N-terminus of GAPB, perhaps representing an immature proteoform were also enriched in the mutants. Peptides from LHCI-1-1, LHCII-1.4 and LHCII-1.1 close to or within the cTP were found exclusively in the double and triple mutants. Finally, a number of cases of Nt maturation defects as observed in our TAILS experiments were again shown to over-accumulate in the mutants. As such, we were able to confirm the results from our Nt proteomics experiments in this label free spectral counting dataset.

3.3.14 Synergistic interactions between the CLP peptidase system and PREP peptidases

To test the genetic and functional interactions between the chloroplast CLP peptidase core and the PREP peptidases, we crossed the *clpr1-2* null mutant (with ~30-50% CLP core capacity) (Kim et al., 2009), the *clpr2-1* mutant (with ~ 20% of CLPPR capacity) (Rudella et al., 2006) and the *clpt1 clpt2* double mutant with ~10-15% CLPR capacity (Kim et al., 2015) to *prep1* and *prep2* single mutants and the *prep1 prep2* double mutant (Figures 3.5 and 3.6). Additionally, we crossed the *clpc1-1* chaperone null mutant (Constan et al., 2004; Nishimura et al., 2013) to these single and double *prep* mutants (Figure 3.6D). *clpc1-1* has ~20-30% total CLPC chaperone capacity with the remainder of CLP chaperone capacity contributed by the lower abundant homolog CLPC2 and perhaps CLPD (it is not clear to what extent CLPC1,2 have an overlapping substrate pool with CLPD). For simplicity we left out the allele numbers for these mutants in the figures and the remainder of the text.

Figure 3.5. Genetic interaction between the CLP core subunits and PREP peptidases. Plants were either directly grown on soil under a 10-h/14-h light/dark cycle at $120 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$, or first grown on agar plates with half-strength Murashige and Skoog medium and 2% sucrose, as indicated. Transcript accumulation levels in the leaves were determined by RT-PCR using gene-specific primer pairs and ACTIN2 as internal control.

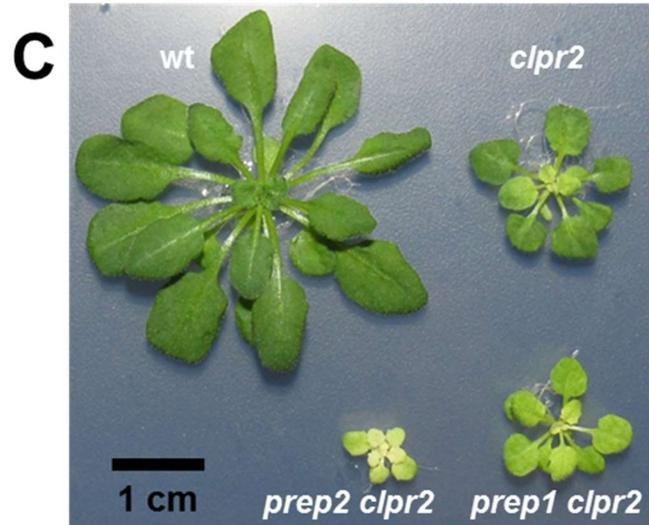
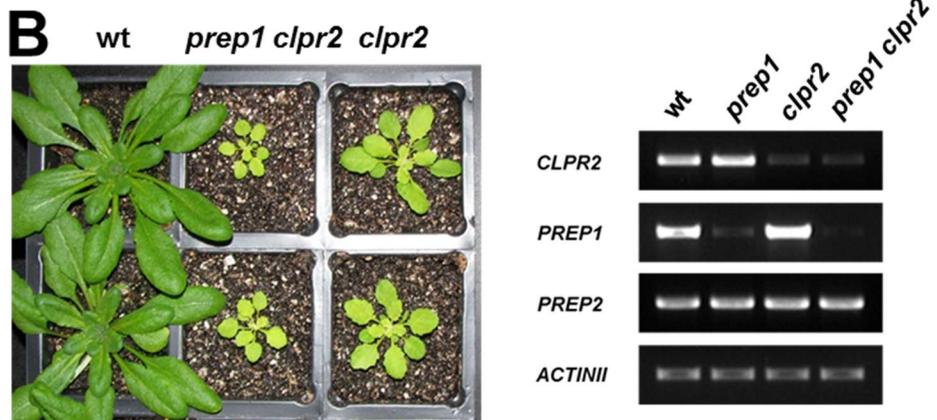
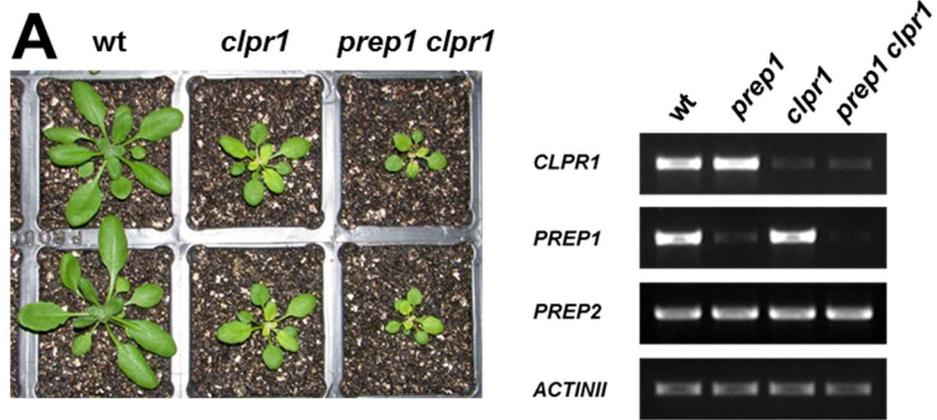
(A) Comparison of plant growth and transcripts of *wt*, *clpr1* and *prep1 clpr1* of plants grown on soil for 26 days.

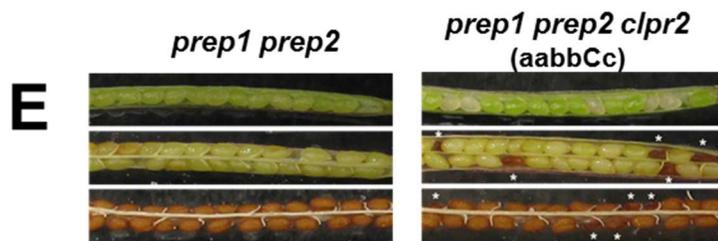
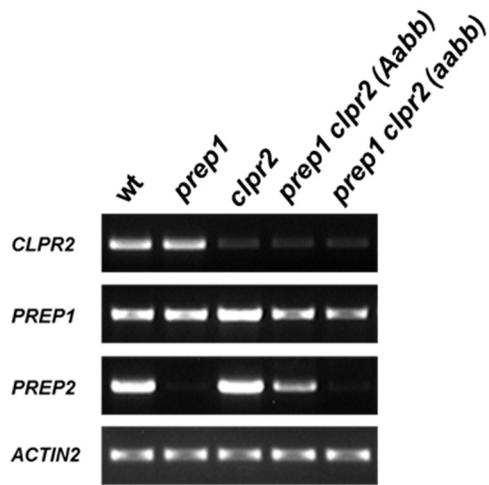
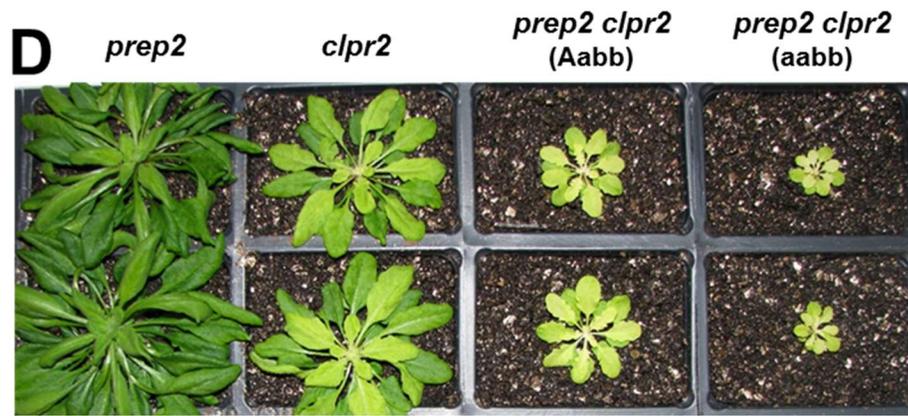
(B) Comparison of plant growth and transcripts of *wt*, *clpr2* and *prep1 clpr2* of plant grown on soil for 33 days.

(D) Comparison of plant growth and transcripts of *wt*, *clpr2*, *prep1 clpr2*, and *prep2 clpr2*. All plants were first grown on agar plates for 45 days under a 10/14 h light/dark cycle at $80 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$.

Comparison of plant growth and transcripts of *prep2*, *clpr2*, *prep2 clpr2* (*Aabb*) and *prep2 clpr2* (*aabb*). All plants were first grown on agar plates under a 10/14 h light/dark cycle at $40 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ for 12 weeks and then transferred to soil and grown for 10 days.

(E) Opened developing (top images), less mature (middle images) and mature (bottom images) siliques of *prep1 prep2* and *prep1 prep2 clpr2* (+/-). Seeds in developing siliques of the *prep1 prep2* are all green, while the siliques of *prep1 prep2 clpr2* (+/-) allele show green and white seeds on average in a 3:1 ratio. In the mature silique stage, triple homozygous mutant seeds are recognizable as smaller, very dark-brown and wrinkled seeds as indicated by the asterisk.





Both the *clpr1 prep1* (Figure 3.5A) and *clpr2 prep1* (Figure 3.5B) rosette plants are >50% smaller than the parents, indicating a strong negative synergistic interaction. However both double mutants are viable, produce viable seed, and can grow on soil despite their very small stature and virescent phenotype. Transcript levels of the *PREP2* homolog were unchanged in the single and double mutants, indicating lack of transcriptional compensation for loss of *PREP2* or *CLP* peptidase capacity (Figures 3.5A,B).

To our surprise, the negative synergistic effect between *clpr2* and *prep2* was much stronger than between *clpr2* and *prep1* (Figures 3.5C). The stronger growth retardation effect for the *PREP2* allele is surprising given that *PREP2* is far less abundant than *PREP1*, and suggests at least partially specific functions for each *PREP* homolog, in agreement with suggested differences in cleavage specificities between the homologs (Stahl et al., 2005). Comparison between homozygous *clpr2-1 prep2* and a *clpr2 prep2* line heterozygous for the *PREP2* T-DNA insertion showed a clear gene dosage effect on the growth phenotype (Figure 3.5D). The *prep1 prep2 clpr2* triple homozygous mutant is embryo lethal and consistently the ratio of green: white seeds in the developing silique is 2.9:1 ($\chi^2 = 0.118$), with the white seed being triple homozygous (Figure 3.5E). These white seeds turn brown when the seeds of other segregating genotypes are still pale green, and then turn into dark brown, wrinkled seeds in mature siliques (Figure 3.5E). This synergistic genetic interaction between *PREP1,2* and *CLP1,2* suggests that *PREP* and the *CLP* peptidase systems have a functional overlap (see DISCUSSION).

Previous studies indicated that CLPT1 and CLPT2 are largely redundant since the *clpt1 clpt2* double null mutant has a dwarf phenotype, and the single mutants have no obvious visible or molecular phenotypes (Kim et al., 2015). The double mutant heterozygous for one of the alleles is more robust (albeit with reduced growth) than the dwarf double mutant, and it is therefore more suitable for the generation of higher order mutants (Figure 3.6A). We crossed this *clpt1clpt2* double mutant (*aaBb*) with homozygous *prep1 prep2*, to generate triple and quadruple mutants (Figure 3.6B). *prep1 prep2 clpt1* is smaller and paler green than *prep1 prep2* indicative of a negative synergistic effect. In contrast, *prep1 prep2 clpt2-1* is similar to *prep1 prep2*. The quadrupole *prep1 prep2 clpt1 clpt2* heterozygous for the *clpt2* allele (*aabbccDd*) is smaller and paler green than *prep1 prep2 clpt1* (Figure 3.6B), whereas the *prep1 prep2 clpt1 clpt2* quadruple homozygous mutant is embryo lethal (Figure 3.6C). The progeny in the developing silique of *prep1 prep2 clpt1 clpt2 (aabbccDd)* shows a ratio of green:white seeds of 3.1:1 ($\chi^2 = 0.028$), with the white seeds turning into dark brown wrinkled seeds when the silique is mature (Figure 3.6C). These white seeds and wrinkled seeds are the quadrupole homozygous mutant. Considering the involvement of CLPT in CLPPR core stabilization and activity (Kim et al., 2015), the strong genetic interaction between CLPT1,2 and PREP peptidase is consistent with the observed genetic interaction between CLPR1,2 and PREP. Analysis of the CLPPR core assembly state by native gels and immunoblotting (using antibodies for the P and R rings) showed that the CLPPR core assembly state in *prep1 prep2* is similar to wt, and the assembly state in *prep1 prep2 clpt1 clpt2 (aabbccDd)* is comparable to the assembly state in homozygous *clpt1 clpt2* (Supplemental Figure 3.8A).

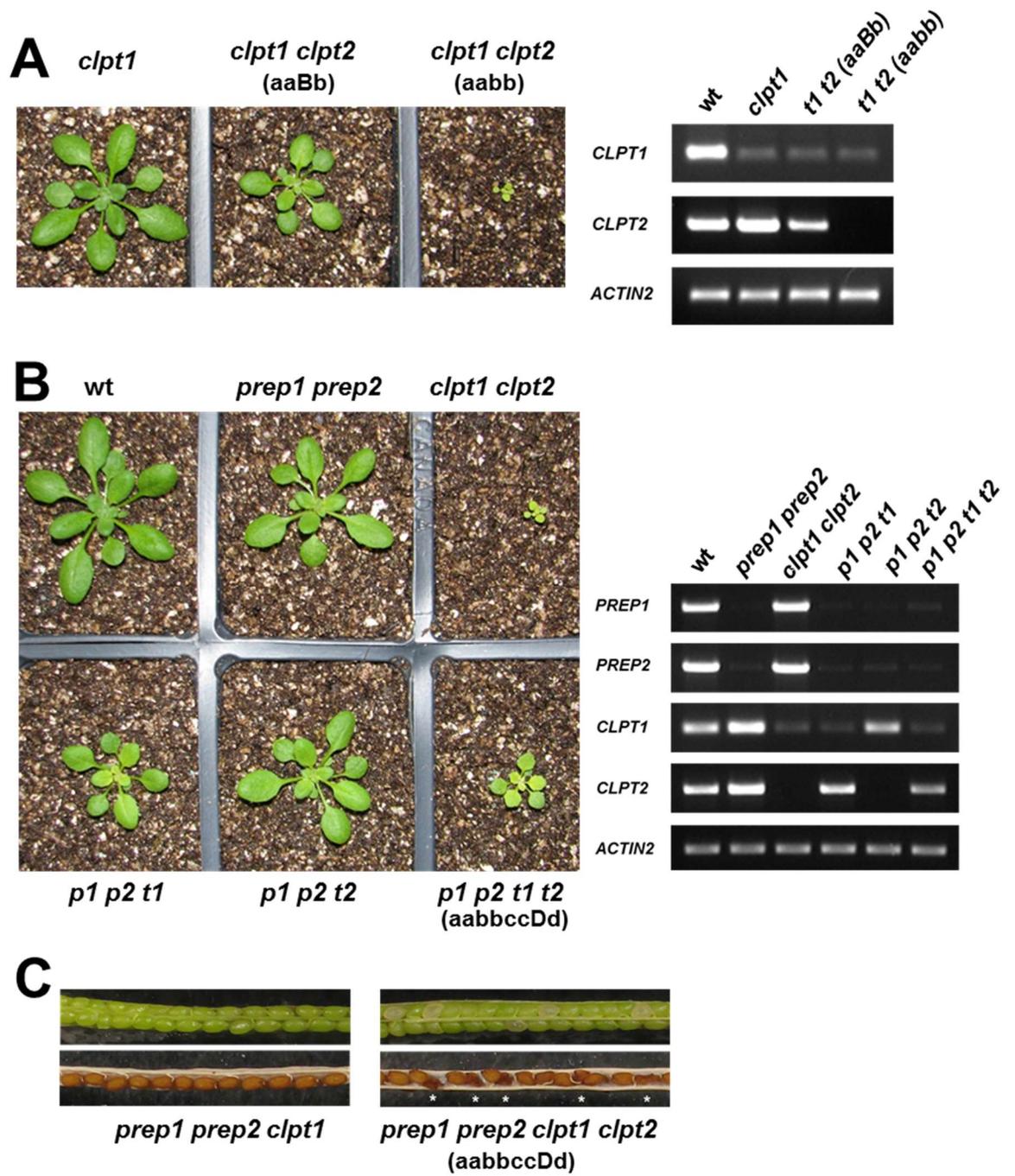
Figure 3.6. Genetic interaction between CLPT, CLPC and PREP peptidases. Plants were either directly grown on soil under a 10-h/14-h light/dark cycle at 120 $\mu\text{mol photons.m}^{-2}.\text{s}^{-1}$ for 21 days (panel A,B) or 28 days (panel D). Transcript accumulation levels in the leaves were determined by RT-PCR using gene-specific primer pairs and ACTIN2 as internal control.

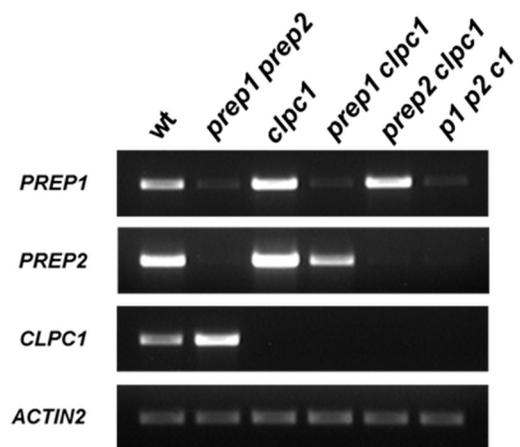
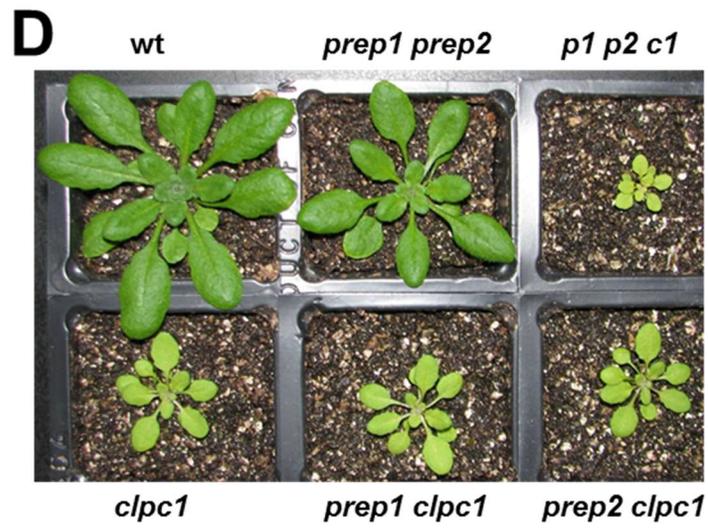
(A) Comparison of *clpt1*, *clpt1 clpt2* (aaBb), and *clpt1 clpt2*.

(B) Comparison of wt, *prep1 prep2*, *clpt1 clpt2*, *prep1 prep2 clpt1*, *prep1 prep2 clpt2*, and *prep1 prep2 clpt1 clpt2* (aabbccDd).

(C) Opened developing (top images) and mature (bottom images) siliques of *prep1 prep2 clpt1* and *prep1 prep2 clpt1 clpt2* (aabbccDd). Seeds in developing siliques of the *prep1 prep2 clpt1-2* are all green, while the siliques of *prep1 prep2 clpt1 clpt2* (+/-) allele show green and white seeds on average in a 3:1 ratio. In the mature silique stage, quadruple homozygous mutant seeds are recognizable as smaller, dark-brown and wrinkled seeds as indicated by the asterisks

(D) Comparison of wt, *prep1 prep2*, *prep1 prep2 clpc1*, *clpc1*, *prep1 clpc1*, and *prep2 clpc1*.





The double mutants *prep1 clpc1* and *prep2 clpc1* are very similar to the *clpc1*, whereas the triple homozygous *prep1 prep2 clpc1* is much smaller than either *prep1 prep2* or *clpc1* (Figure 3.6D). However, the triple mutant can be grown on soil and maintained as homozygous line. Similar to the crosses between *prep1 prep2* and *clpr1*, *clpr2* and *clpt1 clpt2*, this confirms a strong genetic interaction between the CLP and PREP system. Some of this genetic material was then used to investigate the molecular proteomes of these mutants, as described in the next sections.

3.3.15 The molecular proteomics phenotype of the prep1 prep2 clpt1 clpt2 quadruple mutant

Previously we studied the proteome phenotype of the *clpt1-2 clpt2-1* homozygous mutant as compared to wt (Kim et al., 2015). This showed a proteostasis phenotype similar as other CPPR core mutants, with 51 significantly changed proteins that were over-represented by chloroplast-localized proteins. To determine the molecular proteome phenotype of the combined loss of both PREP1,2 and strong loss of CLP peptidase core capacity, we compared total leaf proteomes of wt and the quadruple mutant of *prep1-2 prep2-1 clpt1-2 clpt2-1** both at developmental stage 1.11 (Figure 3.7A for an example of the plants); (* heterozygous for the *clpt2* allele to avoid seedling lethality). We use a similar workflow (SDS-PAGE gels, 12 gel bands/lane, etc; see Supplemental Figure 3.8B) as for proteome analysis of the *PREP* and *OOP* mutants discussed above and as our previous *clpt1 clpt2* proteome analysis (Kim et al., 2015). The workflow identified 1653 proteins and 80 protein groups (193 proteins) matched to 120 thousand non-redundant MS/MS spectra (Supplemental

Table 3.9A). This identified 664 plastid proteins, 87 mitochondrial proteins, as well as 34 proteins dual-localized in plastids and mitochondria. There were no significant overall differences in chloroplast and mitochondrial protein mass (respectively ~60% and 5%) in the mutant compared to wt, but the plastoglobular (PG) proteome increased by 90% (135% in *clpt1 clpt2* (Kim et al., 2015)), the thylakoid proteome decreased by 21% (23% in *clpt1 clpt2* (Kim et al., 2015)), and the stromal proteome increased by 20% (Figure 3.7B).

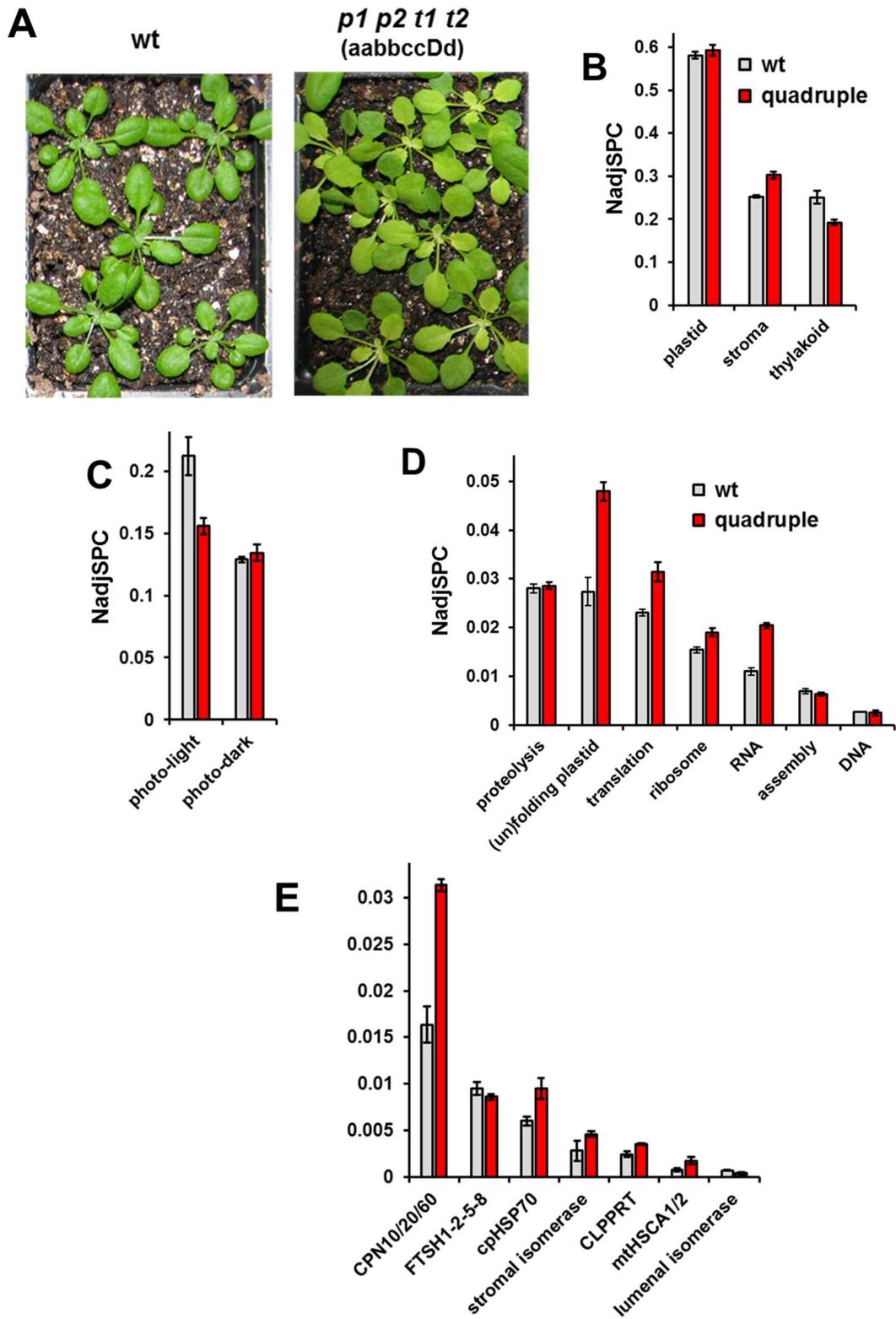
The total quantitative proteomes (based on NadjSPC) of wt and the quadruple mutant were highly distinguishable by PCA, with the biological replicates for each genotype tightly clustered and the two genotypes well separated (Supplemental Figure 3.8C). In total 55 (39 plastid, 3 mitochondrial and 1 dual-localized to plastid/mitochondria) and 10 proteins (4 plastid and 1 dual-localized to plastid/mitochondria) were respectively significantly up- and down-regulated in the quadruple mutant (Supplemental Table 3.9B). Hence plastid proteins were strongly overrepresented in this proteome phenotype. Table 3.4 shows the significantly changed plastid and mitochondrial proteins and includes the results for these proteins for the *clpt1 clpt2* mutant and wt from (Kim et al., 2015) for comparison.

The most up-regulated plastid proteins were the ATP/ADP translocase NTT2 involved in ATP import into the chloroplast and indicative of shortage of ATP generating activity within the chloroplasts (15x up), DPT1 involved in *psaA/B* transcript accumulation (6x up), THI1 involved in thiamin biosynthesis (5x up), metallo-chaperone CCS (4.5x), two DEADbox RNA helicases, including RH3 (4x up), PG-localized FIBRILLINS 1a,b (~3x up) and also the general stromal processing

Figure 3.7. Comparative quantitative leaf proteomics of wt and *prep1 prep2 clpt1 clpt2 (aabbccDd)*

(A) wt and mutants at growth stage 1.11, grown for 24 days for wt and 29 days for the mutant on soil under a 10-h/14-h light/dark cycle at 100 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$

(B-E) Protein mass investments in subcellular compartments, functions and sets of proteins as determined by relative protein abundance. Standard deviations across the three biological replicates are indicated.



peptidase SPP (2.9x up). Other significantly upregulated plastid proteins include several elongation factors, stromal chaperones (CPN60, 70 and 90 family members). Most of these showed very similar upregulation in other CLP core mutants, including the *clpt1* *clpt2* (see Table 1). The two most upregulated mitochondrial proteins were HSCA-2/HSC70-2 (5x up) and HSCA-1/HSC70-1 (1.7 up) both involved in Fe-S sulfur cluster assembly (Leaden et al., 2014). The significantly down-regulated plastid/mitochondrial proteins (Table 3.4) were of course PREP1,2 (undetectable in the mutant), carboxyltransferase α subunit - part of ACCase complex (5x down), inner plastid envelop transporter MEP3/RER4 with unassigned function (4x down), thylakoid NDH subunit A (4x down) and luminal psbQ OEC16-like (2.5x down).

Among the 17 non-organellar significantly upregulated proteins, are the cytosolic chaperones HSP90 and HSP70 members, two S-adenosyl-L-homocysteine hydrolases whereas the ubiquitin-protein ligase (BIG; DARK OVER-EXPRESSION OF CAB 1) was 2x decreased in the quadruple mutant.

The significant effects on plastid proteins are also reflected in general plastid functions (Figure 3.7C-E); for instance investment thylakoid-bound photosynthetic electron transport is decreased by 27% (Figure 3.7C), whereas total plastid stromal (un)folding protein mass is increased by 75% (Figure 3.7D). Within mitochondria, investment in the electron transport chain/OXPHOS pathway and TCA cycle proteome are unchanged, but mitochondrial (un)folding investments are up by 31%, mostly attributed by the statistical significant increase in HSCA701,2 (Figure 3.7E). The total plastid peptidase investment did not change in the triple mutant (Figure 3.7D) consistent with the lack of statistically significant changes individual peptidases; this is similar to

our observations for other CLP core mutants, including *clpt1 clpt2* (Kim et al., 2015). Figure 3.7E also shows the two major plastid peptidases, the thylakoid-bound FTSH1,2,5,8 complex and the stromal CLPRT core. In contrast to investment in plastid proteolysis, investments in stromal (un)folding capacity strongly increased (65%), mostly due to the strong increase in both the CPN60 and HSP70 systems (Figure 3.7E), again in agreement with the significance analysis of the individual proteins (CPN60, cpHSP70 and cpHSP90 are up in the quadruple mutant) (Table 3.4). Finally, total protein investments in chloroplast translation factors and ribosomes, as well as proteins involved in RNA metabolism are increased in the quadruple mutant (Figure 3.7E).

In conclusion, it appears that the quadruple mutant has a very similar phenotype as the *clpt1 clpt2* double mutant, suggesting that the proteome phenotypes of the quadruple mutant is mostly driven by the loss of Clp core capacity. This is also consistent with the very weak *prep1 prep2* proteome phenotype, as shown above. Hence the strong genetic interaction between the CLP and PREP system is not easily explained by the general proteome profiles. Possible explanation for this are, i) changes to (undetected) lower abundant proteins, ii) the accumulation of peptidase degradation products, or iii) feedback inhibition, blocking plastid proteins import and maturation that interfere with growth and development.

Table 3.4. Differentially accumulating chloroplast and mitochondrial proteins in *prep1 prep2 clpt1 clpt2* and *clpt1 clpt2* as compared to wt. Statistical significance was based on combined results of two statistical analysis methods (QSPEC 5% FDR and GLEE $p < 0.01$).

Accession	Protein Description	location	simple functional bin	Total adjSPC	quadruple /wt (NadjSPC) ^a	<i>clpt1</i> <i>clpt2</i> /wt ^b	significant in <i>clpt1 clpt2</i> ^b
AT5G38410.1	Rubisco small subunit 3b (RBCS-3B)	p-stroma	Photosynthesis-dark	232	2.2	0.5	x
AT5G61410.1	ribulose-5-phosphate-3-epimerase (RPE)	p-stroma	Photosynthesis-dark	142	2.4	1.5	
ATCG01100.1	NDH A (NDH-1)	p-thy	Photosynthesis-light	55	0.3	0.9	
AT4G05180.1	OEC16-like	p-thy-lumen	Photosynthesis-light	113	0.4	0.5	x
AT3G01500.3	beta-carbonic anhydrase-1 (beta CA1)	p-stroma	carbonicanhydrases	583	1.5	0.9	
AT1G53240.1	mitochondrial malate dehydrogenase [NAD] mMDH1	mito	respiration	187	1.5	1.1	
AT2G38040.1	α -carboxyltransferase (ACCCase complex)	p-env-inner	lipid-FA	50	0.2	0.5	
AT1G36180.1	acetyl-CoA carboxylase - ACC2	p-str/env	lipid-FA	29	absent in wt	nd	
AT3G22960.1	pyruvate kinase-1 (typically homotetramer)	p-stroma	lipid-FA	88	1.9	1.1	
AT5G53460.1	NADH-GOGAT or NADH-glutamate synthase (GLT1)	p-stroma	N-metabolism	293	1.5	5.6	x
AT3G58610.1	ketol-acid reductoisomerase	p-stroma	AA-metabolism	327	1.6	1.3	
AT3G48560.1	acetolactate synthase	p-stroma	AA-metabolism	35	3.9	2.2	
AT1G12520.1	Cu-metallo chaperone (AtCCS)	p-stroma	Metals	32	4.5	nd	
AT5G60600.1	4-hydroxy-3-methylbutyl diphosphate synthase (HDS)	p-stroma	isoprenoids	258	1.6	3.2	x
AT3G45140.1	lipoxygenase LOX2	p-stroma	JA	404	1.3	1.0	
AT5G54770.1	THI1 (ARA6)	dual	cofactor&vitamin	403	4.8	1.4	
AT5G13630.1	Mg-protoporphyrin IX chelatase (CHLH) (GUN5)	p-stroma	tetrapyrroles	566	1.5	1.0	
AT3G11630.1	2-Cys Peroxiredoxin A (Prx A or BAS1)	p-stroma	redox	128	1.8	2.0	x
AT1G76080.1	thioredoxin (CDSP32)	p-stroma	redox	62	3.1	3.7	x
AT4G04020.1	fibrillin 1a (FBN1a)	p-PG	miscellaneous	136	2.8	2.6	x
AT4G22240.1	fibrillin 1b (FBN1b)	p-PG	miscellaneous	76	3.5	5.3	x
AT5G26742.2	DEAD box RNA helicase (RH3)	p-nucleoid	RNA	333	3.5	2.7	x
AT1G70070.1	DEAD/DEAH box helicase	p-stroma	RNA	52	3.9	28.5	x
AT3G53460.1	RNA binding protein CP29 A' (RNP29A')	p-stroma	RNA	128	1.8	1.9	
AT3G18680.1	Defect in pasA/B transcript accumulation (DPT1)	p-stroma	RNA	33	6.4	2.9	
AT3G03710.1	3'-5' exoribonuclease (RIF10)	p-stroma	RNA	110	1.7	3.0	
AT3G62030.1	peptidylprolyl isomerase ROC4 (CYP20-3)	p-stroma	(un)folding	363	1.7	1.4	x

Table 3.4 (Continued)

Accession	Protein Description	location	simple functional bin	Total adjSPC	quadruple		significant in clpt1 clpt2 ^b
					/wt (NadjSPC) ^a	clpt1/clpt2/wt ^b	
AT2G28000.1	Cpn60-alpha-1	p-stroma	(un)folding	1268	1.8	1.8	x
AT2G04030.1	cpHSP90 (Hsp90-5)	p-stroma	(un)folding	444	2.0	1.9	x
AT5G49910.1	cpHSP70-2	p-stroma	(un)folding	194	1.6	2.0	x
AT3G13470.1							
AT1G55490.1	Cpn60-beta-1/2/3	p-stroma	(un)folding	1442	2.1	1.9	
AT5G56500.1							
AT5G09590.1	mtHSCA-2 (mtHSC70-2)	mito	assembly	53	5.0	35.3	x
AT4G37910.1	mtHscA1 - Fe-S assembly (mtHSC70-1)	mito	assembly	102	1.7	0.8	
AT3G19170.1							
AT1G49630.1	PREP1 and PREP2	dual	proteolysis	192	absent in mutant	1.6	x
AT5G42390.1	stromal processing peptidase (SPP)	p-stroma	proteolysis	153	2.9	2.4	x
AT5G45390.1	ClpP4 (P-ring)	p-stroma	proteolysis	146	1.8	1.5	
AT2G43030.1	50S ribosomal protein L3	p-ribosome	ribosome	135	1.6	0.9	
AT2G33800.1	30S ribosomal protein S5	p-ribosome	ribosome	49	3.3	1.0	
ATCG00800.1	30S ribosomal protein S3	p-ribosome	ribosome	245	1.6	1.7	
AT1G06950.1	Tic110	p-env-inner	sorting	639	1.3	1.6	x
AT4G01800.1	cpSecA	p-stroma	sorting	245	1.4	1.5	
AT4G29060.1	PSRP-7	p-ribosome	translation	501	1.5	1.3	
AT1G62750.1	elongation factor Tu-G (EF-G) (sco1)	p-stroma	translation	479	1.3	1.4	x
AT4G20360.1	elongation factor Tu (EF-Tu-1)	p-stroma	translation	1071	1.3	1.7	x
AT5G13650.1	elongation factor protein, typeA/bipA like (SVR3)	p-stroma	translation	191	2.2	3.1	x
AT5G12470.1	transporter MEP3	p-env-inner	transport	47	0.2	0.6	
AT1G15500.1	transporter NTT2 - adenine nucleotide translocase	p-env-inner	transport	107	14.9	11.4	
AT2G37660.1	3-beta-hydroxy-delta5-steroid dehydrogenase	p-stroma	unknown	98	1.9	1.5	

^a All values are significant (P = 0.01)

^b Relative protein abundance based on NadjSPC (Kim et al 2015)

3.3.16 Comparative Nt proteomics for *clpt1 clpt2* and *prep1 prep2 clpt1 clpt2* reveals extensive an extensive maturation deficit.

To better understand the synergistic genetic interaction between the PREP and CLP system, we compared the stromal Nt proteomes of *clpt1 clpt2*, *prep1 prep2 clpt1 clpt2* (*aabbccDd*) and wt in two independent TAILS experiments. This identified many N-terminal differences between genotypes, see summary Table 3.5. Relative protein accumulation levels are included for the double mutant. Additional details on identification and quantification of these peptides can be found in Supplemental Tables 3.4 and 3.5.

The vast majority of detected Nt peptides start within a few residues of the previously identified Nt start site (Rowland 2015) or the cTP cleavage sites predicted by ChloroP (Emanuelsson et al., 2007). Many of the alternate proteoforms (Class 3a peptides, Figure 3.3) accumulating in *prep1 prep2 oop* were also observed in these CLP mutants (Table 3.5). In contrast, Nt peptides falling within the cTP of nuclear encoded proteins (Class 1 or 2 peptides, Figure 3.3) were not detected. Some obvious breakdown products were detected (Class 4 peptides, Figure 3.3); however, these peptides were not generally elevated or reduced in these CLP mutants (Table 3.5, Supplemental Table 3.4, 3.5).

The maturation defects observed in *clpt1 clpt2* were far more dramatic than those observed in other genotypes with the quadrupole mutant being intermediate between *clpt1 clpt2* and *prep1 prep2 oop*. The same aberrant Nti were identified, but the fraction of total protein Nti represented by these alternate forms was much greater in *clpt1 clpt2*. That is, aberrant proteoforms that are up in *PREP* higher order mutants represent only

a small fraction of the total Nti detected for those proteins. In *clpt1 clpt2* on the other hand, these Nti often represent up to 50% of the total Nti for those proteins (See TH11, GS2, Inorganic phosphatase like, and cpHSP70-1, Table 3.5). Accordingly, the major Nti, normally observed in wt plants was often reduced in *clpt1 clpt2*. Alanine terminated peptides again were very prominent in the list of Nt peptides accumulating in CLP peptidase mutants. Long protein extensions were observed for GAPB, NDPK2 and sedoheptulose-bisphosphatase (SBPase) indicating again that protein maturation is hindered when proteolysis is limited in the chloroplast.

PEP carboxylase family protein (AT4G10750.1) is an interesting case where a Tyr extended N-terminus accumulates eight fold in *clpt1 clpt2* and four fold in the quadrupole mutant (Table 3.5). Tyrosine has been shown in bacteria to act as an N-degron, triggering ClpS dependent degradation (Varshavsky, 2011). Furthermore, Nt Tyr was shown in yeast and Arabidopsis mitochondria to be removed by ICP55 and to be a destabilizing residue (Vogtle et al., 2011; Carrie et al., 2015; Huang et al., 2015). This protein is therefore a good candidate to test for CLPS dependent degradation in chloroplasts via the N-end rule.

Table 3.5. Differentially regulated protein N-termini identified by TAILS in *clpt1 clpt2* and *prep1 prep2 clpt1 clpt2*. Peptides were quantified by MS1 filtering using Skyline software (MacLean et al., 2010). Average ratios for three replicates are displayed with standard deviations. Ratios for two independent experiments (see Supplemental Table 3.5) are shown for comparison: *clpt1 clpt2* (stroma) and *prep1 prep2 clpt1clpt2* (stroma). (a) Localization: s, stroma; t, thylakoid. (b) Predicted Nt residue position after cTP removal (ChloroP). (c) Experimental observed Nt position. (d) Acetylated Nt. (e) P1 residue immediately upstream protease cleavage site based on Schechter and Berger nomenclature. (f) Relative protein abundance based on NadjSPC, Kim *et al* 2015.

Table 3.5

Accession	Protein Description	loc. (a)	Pred. Nt (b)	Exp. Nt (c)	Ac Nt (d)	P1 (e)	N-Terminal Peptide	c1pt1 c1pt2			prep1 prep2 c1pt1 c1pt2			c1pt1 c1pt2 (f)	
								log ₂ (Mu/WT)	STDEV	total SPC	log ₂ (Mu/WT)	STDEV	total SPC	total protein log ₂ (Mu/WT)	total adjSPC
A14G04020	fibrillin 1a (FBN1a)	s	56	56		R	ATDIDDEWGDGQGVVER	4.4	(0.4)	84	0.6	(0.5)	64	1.4	123
				57		A	TDIDDEWGDGQGVVER	4.9	(1.0)	10					
				76		S	STVSVADKAIESVEETER	4.6	(1.2)	46					
A14G24280	cpHSF70-1	s	93	75		R	VVNEKVVGDIDLTTNSAVAAAMEGGKPTIVTNAEGQR	7.8	(0.1)	209	5.6	(0.2)	62	0.9	213
				78		N	EKVVGDIDLTTNSAVAAAMEGGKPTIVTNAEGQR	-0.9	(0.3)	233	0.7	(0.1)	280		
				87		G	TTNSAVAAAMEGGKPTIVTNAEGQR	0.7	(0.3)	6	0.6	(0.6)	38		
				133		R	QAVVNPENITFSVKR				0.5	(0.2)	34		
A14G38970	SFBA-2	s	47	47		R	AASSYADELVKTAKTIASPGR	9.0	(0.1)	995	4.9	(0.2)	132	0.0	377
				48		A	ASSYADELVKTAKTIASPGR	-2.0	(0.2)	853	0.1	(0.2)	1238		
				50		S	SYADELVKTAKTIASPGR	-1.1	(0.2)	42	0.3	(0.3)	83		
				195		N	GPSALAVKEAAWGLAR	0.3	(0.9)	6	0.1	(0.2)	33		
				344		N	TOLKTWGGRPENVNAQTLLAR	-1.4	(1.1)	2	0.1	(0.1)	32		
A15G09650	Inorganic phosphatase like	s	67	58		C	SAIYNPQVKVQEEGPAESLDYR				-0.1	(0.3)	38	-0.7	41
				59		S	AIYNPQVKVQEEGPAESLDYR	6.1	(0.3)	121	4.5	(0.0)	96		
				60		A	IYNPQVKVQEEGPAESLDYR	-2.7	(0.2)	185	-0.1	(0.1)	148		
A15G20720	Cpn21	s	51	51		K	AASVVAPKYTSIKPLGDR	8.7	(0.6)	137	6.0	(0.1)	66		
				52		A	ASVVAPKYTSIKPLGDR	-0.4	(0.2)	164	0.2	(0.2)	174		
A15G35630	GS2	s	46	51		L	ALQSDNSTVNRVETLLNLDTKPYSDR	8.7	(0.1)	135	4.4	(0.4)	54		
				52		A	LQSDNSTVNRVETLLNLDTKPYSDR	-1.4	(0.2)	162	-0.1	(0.4)	97		
				351		L	TGKHETASIDQFSWGVANR				0.0	(0.6)	30		
A15G54770	TH11	s	46	46		R	ATTAGYDLNAFTFDPIKESVSR	8.0	(0.2)	119	4.4	(0.5)	58	0.5	189
						A	TTAGYDLNAFTFDPIKESVSR	-1.0	(0.2)	180	0.5	(0.1)	176		
						A	GYDLNAFTFDPIKESVSR	-0.2	(0.1)	104	0.5	(0.1)	152		
A15G63310	NDPK2-stromal	s	63	63		R	ASSAESGIFLPHVASMEDVEETYIMVKPDGIQR	2.1	(0.7)	26				0.8	44
						A	SSSAESGIFLPHVASMEDVEETYIMVKPDGIQR	2.8	(0.4)	6					
						S	SAESGIFLPHVASMEDVEETYIMVKPDGIQR	2.4	(0.5)	18					
						A	SMEDVEETYIMVKPDGIQR	-2.2	(0.5)	171	-0.4	(0.1)	252		

Table 3.5 (Continued)

Accession	Protein Description	loc. (a)	Pred. Nt (b)	Exp. Nt (c)	Ac Nt (d)	P1 (e)	N-Terminal Peptide	c1p1 c1p2			prep1 prep2 c1p1 c1p2			c1p1 c1p2 (f)	
								log ₂ (Mu/WT)	STDEV	total SPC	log ₂ (Mu/WT)	STDEV	total SPC	total protein log ₂ (Mu/WT)	total adjSPC
AT1G03475	coproporphyrinogen III oxidase	s	49	48		C	SVSIEKEVPETERPFTFLR	-1.4	(0.2)	92	0.5	(0.1)	164	0.2	42
				49		S	VSIEKEVPETERPFTFLR	5.3	(0.2)	59	3.0	(0.0)	11		
AT1G16080	unknown protein	s	42	42	1	M	AMAAASAATAKALAPAVIVGGGR	4.5	(0.5)	47	2.7	(0.5)	28	0.7	49
				44	1	M	AAASAATAKALAPAVIVGGGR	-2.1	(0.1)	40	-0.8	(0.1)	68		
AT1G29900	carbamoylphosphate synthetase	s	63	72	1	A	AASAATAKALAPAVIVGGGR	-1.2	(0.2)	122	-0.1	(0.2)	120	-0.2	164
				78		H	VLPVSEADTTTKPFSPFVGVGR	1.4	(0.3)	45	0.7	(0.3)	33		
AT1G42970	GAPB	s	46	47		S	ELADTTTKPFSPFVGVGR	-1.0	(0.1)	77	-0.2	(0.2)	118	0.1	308
				67		M	SSIGGEASFFDAVAQIIPKAVTTSIPVR	5.1	(0.9)	28	5.8	(0.1)	32		
				77		K	AVTTSIPVRGETVAKLKVAINFGFR	5.8	(0.9)	53	-0.1	(0.4)	14		
AT1G48860	EPSP synthase	s	40	71		R	ETVAKLKVAINFGFR	4.4	(0.4)	48				2.7	7
				72		A	ASVSTAEKASEIVLOPIR	-0.5	(0.1)	84	0.2	(0.2)	202		
AT1G51100	CRR41 - NDH assembly factor	s	39	40		C	SSNPESKQQFINLTPAPESINTTSAEKFPKIEKR	-1.4	(0.2)	29	-0.3	(0.2)	110		
				41		S	SNPESKQQFINLTPAPESINTTSAEKFPKIEKR	2.1	(0.4)	2	2.5	(0.1)	4		
AT1G65970	MDS	s	53	49		S	ASVSAASSAVDVNESVTSEKPTKTLPPFR	2.8	(0.2)	32	0.7	(0.2)	6	0.6	8
				50		A	SVSAASSAVDVNESVTSEKPTKTLPPFR	0.0	(0.3)	2	0.4	(0.3)	8		
				53		S	AASSAVDVNESVTSEKPTKTLPPFR	2.4	(0.4)	26	1.0	(0.4)	56		
				54		A	ASSAVDVNESVTSEKPTKTLPPFR	-0.1	(0.4)	38	0.2	(0.1)	110		
				55		A	SSAVDVNESVTSEKPTKTLPPFR	1.3	(0.2)	47	1.0	(0.3)	102		
				56		S	SAVDVNESVTSEKPTKTLPPFR	0.0	(0.5)	2	1.0	(0.1)	45		
				57		S	AVDVNESVTSEKPTKTLPPFR	0.4	(1.0)	14	2.1	(0.2)	27		
AT1G80600	acetylornithine transaminase	s	44	42	1	S	VLTNAGDOAVSVKASVSQKVEEAAKVIIVGYAR	-1.3	(0.1)	10				-0.1	44
				55		K	ASVSQKVEEAAKVIIVGYAR	4.4	(0.5)	38					
				56		A	SVSQKVEEAAKVIIVGYAR	-1.3	(0.1)	16					
AT2G01140	SFBA-3	s	41	41		R	AGAYSDELVTAKSIASPSGR	5.6	(0.5)	88				-3.0	8
				42		A	GAYSDELVTAKSIASPSGR	-1.1	(0.2)	180	0.4	(0.1)	128		
AT2G04030	cpHSP90	s	61	62		C	DAAVAEEKITTEEGSGEKFYQAEVSR	0.2	(0.1)	16	0.9	(0.1)	109	1.0	232
				63		D	AAVAEEKITTEEGSGEKFYQAEVSR	2.2	(0.3)	18	1.2	(0.1)	73		
AT2G04400	IGPS	s	66	53		R	AQOSDLKESLAVSSSSVEDKGNVLR	5.2	(0.2)	83	1.8	(0.3)	48		
				54		A	QOSDLKESLAVSSSSVEDKGNVLR	0.4	(0.4)	88	0.3	(0.5)	214		

Table 3.5 (Continued)

Accession	Protein Description	loc. (a)	Pred. Nf (b)	Exp. Nf (c)	Ac Nt (d)	P1 (e)	c1p1 c1p2			prep1 prep2 c1p1 c1p2			c1p1 c1p2 (f)	
							log ₂ (Mu/WT)	STDEV V	total SPC	log ₂ (Mu/WT)	STDEV	total SPC	total protein log ₂ (Mu/WT)	total adfSPC
AT2G05990	Enoyl-ACP reductase	s	75	75		R	5.0	(0.8)	64	0.6	(0.5)	30	0.0	52
							AMSESENKAPSGLPIDLR							
						M	-2.3	(0.3)	177	0.3	(0.1)	160		
AT2G44650	Cpn10-1	s	40	42		K	6.7	(1.7)	73	5.2	(0.9)	72	0.4	60
						A	-1.4	(0.2)	46	0.2	(0.1)	114		
AT2G45300	EPSP synthase 2	s	74	70		M	3.0	(0.8)	46	1.8	(0.3)	55	2.7	7
AT3G03710	RIF-10	s	49	64		R	4.4	(0.4)	29	4.2	(0.6)	36	1.6	43
						A	-0.6	(0.2)	16	1.0	(0.4)	50		
AT3G12780	PGK-1	s	76	78		M	-0.3	(0.1)	165	0.1	(0.1)	186	-0.1	1011
AT3G22890	ATP sulfurylase	s	48	48		R	5.6	(0.8)	67	3.2	(0.1)	24	0.8	23
						A	-0.8	(0.1)	92	-0.8	(0.2)	145		
AT3G23940	dihydroxyacid dehydratase	s	36	35		C	-1.2	(0.1)	60	0.1	(0.0)	268	1.0	30
						S	1.9	(0.1)	28	2.3	(0.1)	103		
						Q				0.5	(0.0)	33		
AT3G48420	haloacid dehalogenase-like hydrolase-2	s	66	67		C	-1.3	(0.1)	21	-0.1	(0.2)	91	0.6	16
						S	2.3	(0.3)	77	2.0	(0.2)	2		
						A	-0.2	(0.1)	5					
AT3G55800	SBPase	s	60	60		K	0.8	(1.6)	7				-0.3	186
						K	1.9	(2.0)	11					
						C	-2.4	(0.1)	394	-0.4	(0.2)	965		
AT3G60750	TKL-1	s	66	66		R	7.3	(0.3)	153	4.2	(0.1)	118	-0.2	699
						A	-1.3	(0.1)	1213	0.1	(0.2)	928		
						A	1.1	(0.2)	366	1.5	(0.0)	158		
						E				-0.3	(0.7)	14		
						M				-0.2	(0.6)	2		
						D	-0.2	(0.6)	34	0.1	(0.4)	24		
						N				0.3	(0.3)	48		
AT1G20020	FNR-2	t	56	56		K	5.6	(0.1)	19	4.6	(0.2)	52	-0.1	164
						A	-1.9	(0.1)	70	-0.4	(0.3)	247		
						Q	-2.5	(0.3)	10	0.3	(0.4)	32		

Because *clpt1 clpt2* has the most pronounced maturation defects, we compiled all Nt peptides that were elevated or reduced in this genetic background, and aligned the protein sequences four residues up and down stream of the cTP cleavage site that generated each N-terminus (Figure 3.8; Supplemental Table 3.10). We then generated Icelogos (Colaert et al., 2009) to display over and underrepresented amino acids. This showed that Nt peptides accumulating in the mutant do not have a distinctive Nt motif that is significantly different from that observed for all mature chloroplast proteins, Figure 3.8C (Rowland et al., 2015). Rather it seems that Nti elevated in the mutant tend to result from cleavage following Arg/Lys/Ser (in the P1 position) whereas Nti that are down or unchanged in the mutant are generally formed by cleavage after Ala/Cys/Met, as is the case for the majority of plastid targeted proteins. Therefore, proteoforms that accumulate in peptidase mutants are generally formed by the cleavage motif, V/I - R/S/K ↓ A whereas proteoforms reduced are generated by V/I - X - A/C/M ↓ S/A (Figure 3.8). Furthermore, the relative populations of each proteoform is influenced by the severity of the peptidase deficiency.

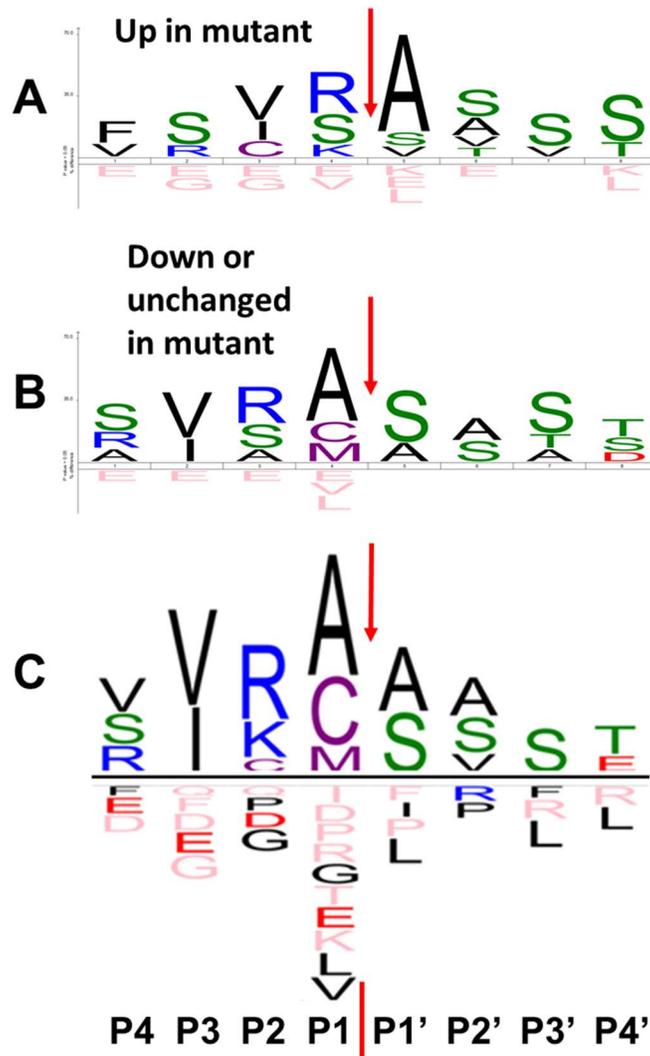


Figure 3.8. Icelogo plots for differentially regulated Nt processing events in the *clpt1 clpt2* mutant. The protein sequences surrounding the ultimate Nt cleavage were compared, peptide nomenclature positions P4 through P4' were plotted using Icelogo; chloroplast proteome with cTPs removed used a reference, $P = 0.05$. **(A)** Nt proteoforms at least two fold up in *clpt1 clpt2* ($n = 62$); **(B)** Nt proteoforms unchanged or down in *clpt1 clpt2* ($n = 50$); **(C)** Arabidopsis N-terminome consensus cTP cleavage site for soluble nuclear encoded chloroplast proteins ($n = 165$) (Rowland *et al* 2015).

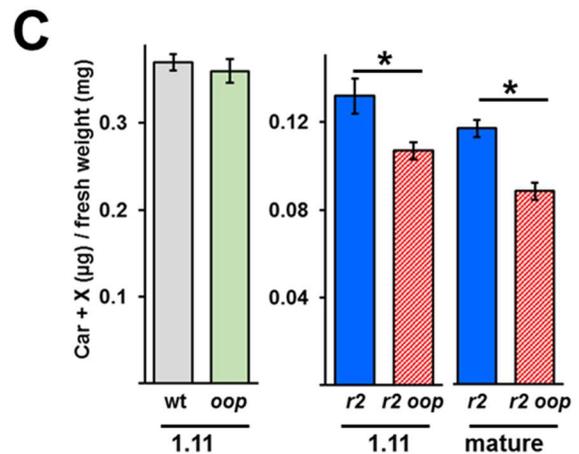
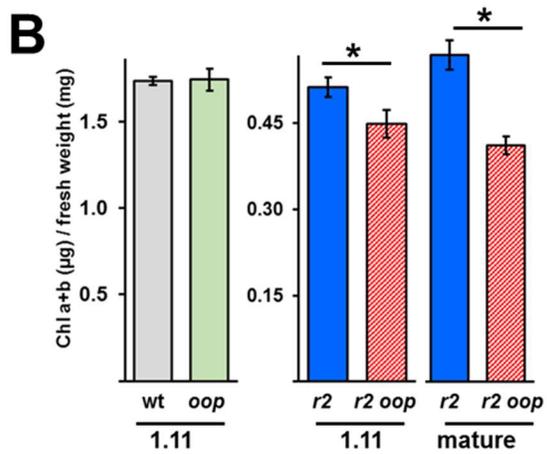
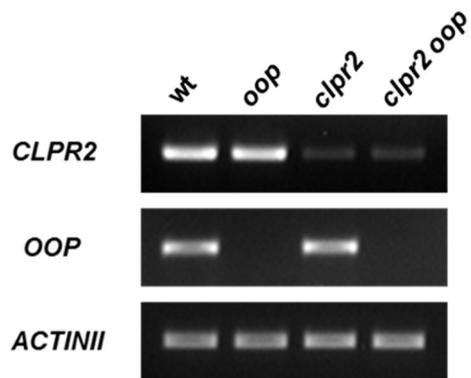
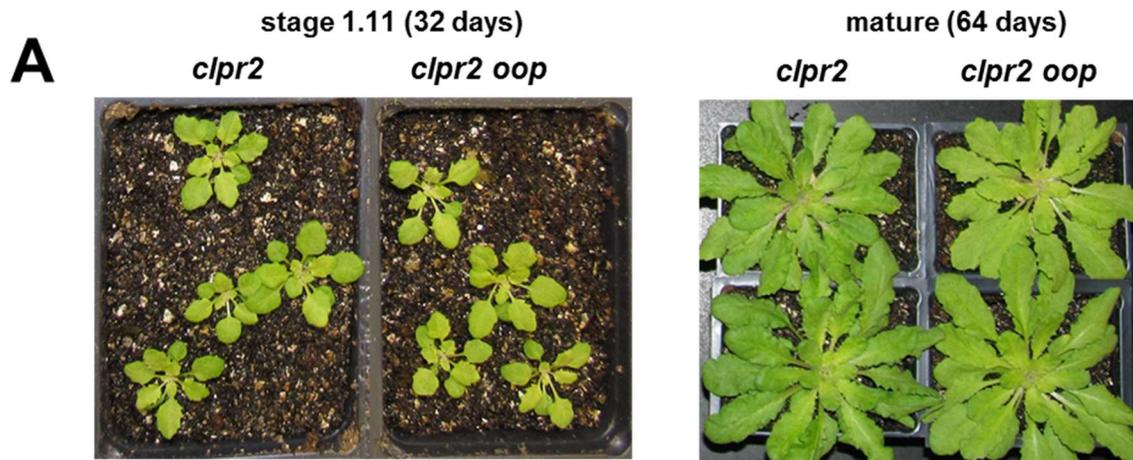
3.3.17 Genetic interactions between CLPR2 and OOP

To investigate the genetic interactions between the OOP and the CLP system, we crossed virescent *clpr2-1* with *oop* (*oop* has no obvious phenotype – see Figure 3.1A). The double mutant was more virescent than *clpr2-1*, and chlorophyll and carotenoid extraction and analysis showed significantly ($P < 0.01$) reduced chlorophyll and carotenoid content on a fresh weight basis (Figure 3.9). In contrast, we did not observe measurable and significant differences in pigment content between wt and *oop* (Figure 3.9B,C). The observed weak synergistic interaction suggests that OOP and CLP core have some functional overlap. We did not analyze the molecular proteome phenotype of the double mutant as this is likely to be dominated by the strong *clpr2-1* phenotype, which we analyzed in detail previously (Rudella et al., 2006; Zybailov et al., 2009). The *oop* proteome phenotype was discussed in previous sections.

Figure 3.9. Genetic interaction between the CLPR2 and OOP.

(A) Comparison of *clpr2* and *clpr2 oop*. Plants were grown on soil for 32 days under a 10-h/14-h light/dark cycle at $120 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$. Comparison of *clpr2* and *clpr2 oop*. Plants were grown on soil for 64 days under a 10-h/14-h light/dark cycle at $120 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$. Transcript accumulation levels in the leaves of wt, *oop*, *clpr2* and *clpr2 oop*. Transcript levels were determined by RT-PCR using gene-specific primer pairs; ACTIN2 was used as internal control.

(B, C). Pigment levels in wt, *oop*, *clpr2* and *clpr2 oop* mutants. Chlorophyll **(B)** and Carotenoid **(C)** contents were measured after 20 days (wt and *oop*, 1.11 stage), 32 days (*clpr2* and *clpr2 oop*, 1.11 stage), and 64 days (*clpr2* and *clpr2 oop*) of sowing. Standard deviations are indicated. N=5 for 1.11 stage plants and N=3 for 64 days old plants.



3.4 DISCUSSION

2.4.1 Proteolysis in the chloroplast

Cellular peptidase systems incorporate many layers of complexity to control when proteins get degraded. These enzymes not only degrade damaged or unwanted proteins, they also play crucial roles in the processing and maturation of protein (Poveda-Huertes et al., 2017). The chloroplast protease network has many similarities to bacterial systems owing to its cyanobacterial origin. However, proteases of the chloroplast have evolved complexities that are absent in bacteria or even in other organelles such as mitochondria (van Wijk, 2015).

Here we have combined a rigorous genetic analysis of the interactions between three chloroplast peptidase systems with high resolution MS techniques that probe the *in vivo* molecular proteome phenotypes (proteolytic signature) of an array of peptidase mutants. We set out to determine the effects of loss or reduction of PREP, OOP and CLP peptidases on the proteome and their relative contributions to the peptidase network in chloroplasts. We evaluated whether peptidase substrates, including cleaved cTPs and peptide fragments, accumulate in peptidase mutants. Secondly, we determined if and how other cellular processes, including protein maturation and proteostasis, were affected by a reduction of peptidase activity.

2.4.2 Genetic and functional interactions between three stromal peptidase systems

In this study we probed the genetic and functional interactions between the abundant multi-component CLP protease system and the simpler dual-localized mitochondrial/chloroplast soluble PREP and OOP peptidase systems. Specifically, we

crossed mutants with decreased CLP protease core activity or CLP chaperone capacity with PREP and OOP null mutant. The combined loss of CLP protease or chaperone capacity with complete loss of PREP capacity resulted in strong leaf and rosette growth phenotypes that correlated with the reduction in CLP core capacity. CLP core capacity of less than 10-20% (*clpt1 clpt2*) resulted in embryo lethality when combined with loss of PREP activity. Comparative, quantitative proteomics of the quadruple *prep1 prep2 clpt1 clpt2* (aabbccDd) showed a strong proteome phenotype that strongly resembled the *clpt1 clpt2* proteome phenotype (Kim et al., 2015). The OOP and CLP core mutants also showed a weak synergism indicative of functional interactions.

From these genetic interactions, a picture emerges of the relationships and hierarchical positions of PREP, OOP and CLP in the peptidase network: There is clearly significant overlap in the functions of PREP and CLP and to a lesser extent between OOP and CLP. This implies that PREP and OOP do not act strictly downstream of CLP but rather they act somewhat in parallel, and can compensate for loss of the other as was previously shown for PREP and OOP (Kmiec et al., 2013).

3.4.3 The differential response of the chloroplast and mitochondrial proteostasis machineries and networks to loss of protease capacity

When proteolysis is compromised, proteostasis is unbalanced (Jarvis and Lopez-Juez, 2013). Accumulation of partially degraded proteins and/or signal peptides may destabilize membranes (Kmiec et al., 2014) and interfere with enzymes normally involved in protein processing – e.g. other proteases (Poveda-Huertes et al., 2017). Additionally, accumulation of inactive enzymes may clog metabolic pathways or the

formation of protein aggregates may physically interfere with general cellular processes. A robust proteolytic machinery normally removes these potentially toxic products (Jarvis and Lopez-Juez, 2013). It is well established that unbalanced proteostasis leads to over-accumulation of chaperones and other proteases (Hartl et al., 2011; Kim et al., 2015). Our comparative proteomics analysis revealed that in the absence of both PREP and OOP, chaperone levels only modestly increased abundance and proteostasis is largely stable. In sharp contrast, the CLP peptidase mutants show strong proteostasis defects that persists throughout the life of the plant.

3.4.4 Maturation is compromised in PREP, PREP/OOP and CLP peptidase mutants

Nt proteomics (TAILS) allowed us to measure the relative abundance of different Nt peptides between wt and the various peptidase mutants. These Nti are proportional to differently processed isoforms of each protein (proteoforms). We previously characterized the Nt proteome of the Arabidopsis chloroplast. For a given protein, there is generally a primary proteoform that corresponds to the mature function protein, see Rowland *et al* for a discussion (Rowland et al., 2015). Therefore, we refer to minor, atypical proteoforms as aberrant or as being improperly processed.

Processing defects (Class 3a peptides, Figure 3.3) were clearly observed in soluble protein extracts from total leaf and chloroplast stroma for both *prep1 prep2* and *prep1 prep2 oop* mutants. Extended or truncated proteoforms with an exposed Nt Ala or Val made up the majority of mutant accumulating peptides. Please note that many of these aberrant proteoforms are present in wt tissue, but typically represent only ~5 – 10% of the total protein mass for a given gene. The same selection of Nt peptides

accumulate in PREP mutants but these aberrant forms are many times more abundant than in wt (see spectral count numbers and relative abundance ratios in Table 3.1). Interestingly, the relative over- or under-accumulation of Nt proteoforms was almost identical in the double and triple mutants demonstrating that loss of PREP (and not loss of OOP) is the primary cause of this phenotype.

Maturation defects were also observed in *clpt1 clpt2* and *prep1 prep2 clpt1 clpt2*. The magnitude of up or down regulated Nt proteoforms was far greater than those observed for the PREP/OOP mutants, and the aberrant forms represented a much larger portion of the overall protein – up to 50% of the total Nti for that protein. The quadrupole mutant, that is heterozygous for *clpt2*, had defects intermediate between *prep1 prep2 oop* and *clpt1 clpt2*. Interestingly four of the substantially altered proteins were chloroplast chaperones (HSP70-1, cpHSP90, CPN21 and CPN10-1) and it is possible that some of these aberrant proteoforms have reduced chaperone efficiency compared to the properly processed forms, perhaps contributing to the imbalance in proteostasis. Both chloroplast HSP70 and cpHSP90 normally start with the relatively rare (for plastid localized proteins) amino acids, Asp and Glu, indicating that their exposure may be important for functional integration. In the aberrant forms of these proteins these residues are masked or removed, perhaps contributing to the severe growth phenotype of *clpt1 clpt2*.

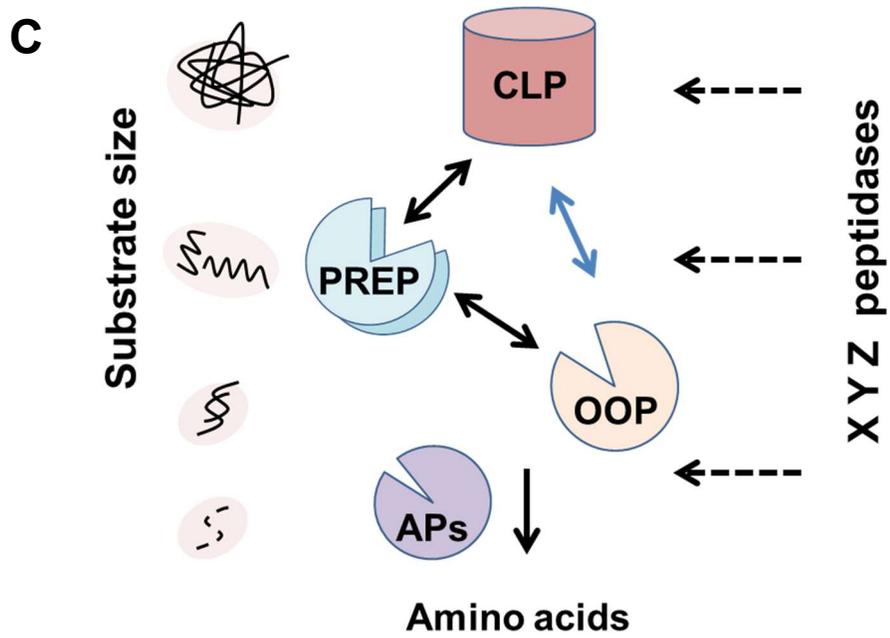
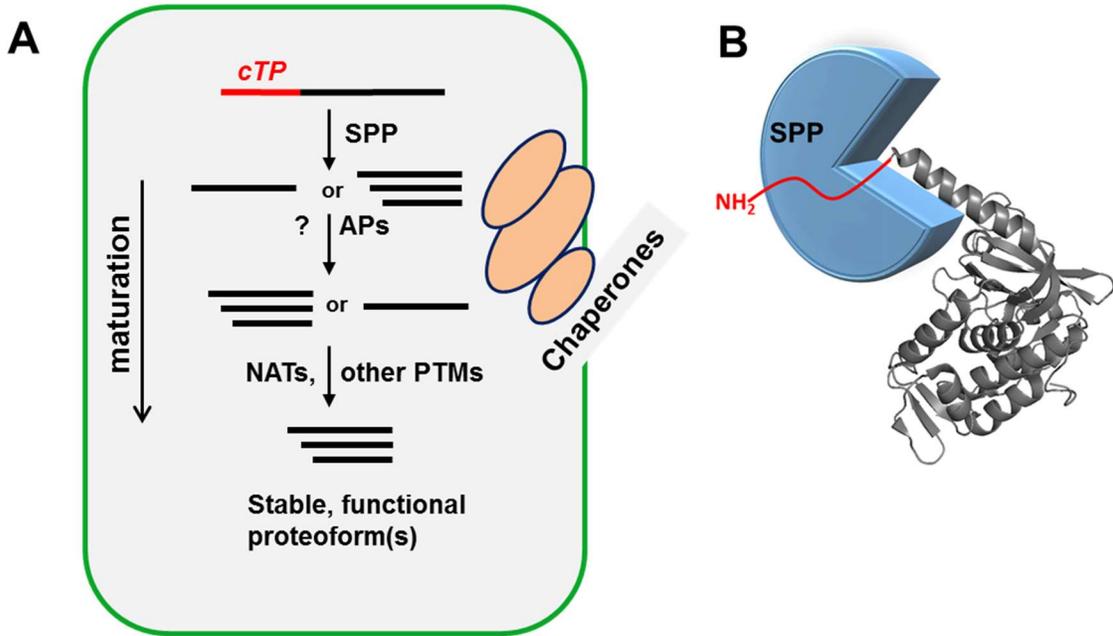
The significance of different N-terminal amino acid extensions is unknown. We present various scenarios/models for their formation (Figure 3.10A,B). It is clear that Ala (and also Val) N-terminated proteins accumulate in all of the peptidase mutants analyzed,

Figure 3.10. Model for proteolytic maturation and degradation.

(A) Nuclear encoded undergo a series for proteolytic processing steps and other modifications after import into the chloroplast. Charerones contribute to these steps and protein folding. cTP, chloroplast transit peptide; SPP, stromal processing peptidase; APs, amino peptidase potentially involved in processing of immature proteoforms; NATs, alpha amino acetyltransferases; PTMs, post translational modifications.

(B) SPP normally cleaves the unstructured cTP from the folded or semi-folded pre-protein (crystal structure of Arabidopsis THI1 shown; PDB: 1RP0). Under conditions of folding stress and limited chaperone availability, the precision of cTP cleavage is reduce; the line between unstructured cTP and structured protein is blurred.

(C) Model of chloroplast peptidase network and hierarchy of CLP, PREP and OOP. Different substrate classes are selected by CLP, PREP and OOP, or by combinations thereof. Additional known and unknown peptidases cleave substrates and produce proteolytic products that enter into this network laterally.



either representing a shorter or longer form than the primary protein N-terminus. Furthermore, tri- or di-Ala N-terminated proteins tend to accumulate more so than proteins that start with a single Ala. Except in the case of TKL-1 where the major proteoform sequence begins with di-Ala, and the mono- and tri-Ala forms accumulate at much lower levels (Tables 3.1, 3.2 and 3.5). This suggests that processing is context dependent, *i.e.* different for each protein undergoing maturation. Three possible explanations for these maturation defects are as follows: (i) Aminopeptidase mediated trimming of protein N-termini by other aminopeptidases is hindered in certain protease mutants (*e.g.* ICP55-like activity (Carrie et al., 2015)). This seems unlikely because the Nt residues removed and exposed by these cleavage events are highly variable whereas aminopeptidases tend to have some amino acid specificity. As previously mentioned, many proteins with and without Nt Ala were observed, perhaps suggesting an Ala-aminopeptidase. However, the most common mature N-terminus for nuclear encoded plastid proteins is Ala, thus ruling out any universal Nt Ala cleavage event. (ii) Improperly processed Nt proteoforms are normally removed by the protease network, in a quality control-like mechanism, and this pathway is blocked in peptidase mutants. This seems feasible, however none of the aberrant proteoforms appear to contain residues or motifs like those of bacterial, mitochondrial or proposed plastid N-degrons (Apel et al., 2010; Carrie et al., 2015; Rampello and Glynn, 2017). (iii) Limited precision of SPP under conditions of imbalanced proteostasis, folding stress and peptide fragment accumulation causing interference between substrates and SPP (Figure 3.10 A,B). We favor this explanation for accumulation of aberrant proteoforms. Feedback inhibition has been clearly demonstrated in yeast mitochondria in PREP (Cym1)

mutants and also in human cells overexpressing a β -amyloid peptide (Mossmann et al., 2014). Mossmann *et al* present a model whereby accumulating peptides, caused by lack of PREP or flooded capacity, bind to and inhibit cleavage or pre-proteins by MPP. We do not observe intact pre-proteins in stroma but we do find evidence for proteins with extended Nti up 16 residues longer (Table 3.1) in TAILS experiments.

Therefore, we suggest that SPP cannot properly cleave cTPs because of accumulating peptide substrates and because chaperone capacity is limited due to increased folding stress. In the *clpt1 clpt2* and quadrupole mutants, SPP abundance is increased at least two-fold, suggesting a bottleneck in SPP capacity likely due to decreased SPP cleavage efficiency. More than 20 years ago Gavel and von Heijne proposed that SPP does not recognize a strict sequence motif during cTPs cleavage, but rather it reads a transition from an unstructured to a structured region surrounding the cTP cleavage site (von Heijne 1990). This theory fits well with our results that show that cTP cleavage is somewhat flexible and is context dependent.

3.4.5 Accumulation of protein degradation products in absence of PREP and OOP oligopeptides

PREP and OOP can degrade small to medium sized peptides *in vitro* (see INTRODUCTION) and are proposed to have a preference for cTPs and mTPs because of their intrinsically disordered nature and positive charge (Johnson et al., 2006; Teixeira and Glaser, 2013; Teixeira et al., 2017). We conducted a series of experiments designed to monitor endogenous peptides in leaf tissue and chloroplasts extracts. Our LMW TAILS experiment revealed that peptide degradation products indeed accumulate

in the *prep1 prep2 oop* mutant (Table 3.2). Fragments from various stromal proteins including GUN5, RH3, CLPP4, GAPB and various ribosomal subunits were found to accumulate. Additionally, a number of fragments from thylakoid proteins were highly enriched in the mutant. We do not know the physiological significance of these degradation fragments. It is reasonable to assume that these fragments contributed to the maturation defects described above, this phenomenon also being observed in yeast mitochondria (Poveda-Huertes et al., 2017).

Our peptidome experiments on *prep1 prep2*, as well as *prep1 prep2 clpt1 clpt2* mutants, demonstrate that cTP fragments and degradation products (Class 2 and 4 Peptides, Fig.3; Fig. 4; Table 3) accumulate in PREP mutants, and that OOP perhaps plays only a minor role in cTP removal (Figure 3.10 C). Importantly, a number of the cTPs fragments observed were produced by cleavage at the known cTP cleavage site (Table 3.3, see C-terminal end of detected peptides). For RBCS-2b and RBCS-4, we identified nine unique peptides, ranging in size from 15 to 30 residues, all ending with a C-terminal Cys, which is the well-established end of the cTP (Houtz et al., 2008; Atkinson et al., 2017). Also, the cTPs of LHCII 2.1, LHCII 3 and PC1, exclusively found in *prep1 prep2* chloroplast extracts, all end at the known cTP cleavage site, suggesting again that cTP fragments did indeed accumulate. These cTP fragments were likely directly produced and release by SPP and not efficiently degraded when there was no PREP1,2 activity. Furthermore, a number of fragments from proteins known to have short half-lives in chloroplasts (Li and Millar, 2017) were found to accumulate in PREP mutants, including D1, THI1 and several tetrapyrroles biosynthetic enzymes(Figure 3.4C and Table 3.2).

3.4.6 Models for functional interactions between CLP, PREP and OOP peptidase systems

Based on genetic interactions, our proteome analysis, and results from previously published studies on the upper substrate size limitations for PREP and OOP, we propose a functional peptidase network model within the chloroplast (Figure 3.10C). Different substrate classes are accepted by CLP, PREP, OOP (Nishimura and van Wijk, 2015; Teixeira et al., 2017). However, the strong genetic interactions between CLP and PREP, and PREP and OOP (black arrows) suggest that they have substrates in common (Figure 3.10C). The weak interaction between CLP and OOP (blue arrow) suggests only minor overlap of substrates.

Given the substrate size limitation of OPP and PREP, and the ATP-driven unfolding capacity of the CLP system, CLP must have its own pool of substrates (Nishimura and van Wijk, 2015) that are inaccessible to PREP and OOP (Kmiec et al., 2014). In contrast, PREP is tailored to degrade cTPs and mTPs, in line with its evolutionary relationship to SPP and MPP (Mach et al., 2013). Degradation products generated by either PREP or the CLP systems could potentially be further degraded by OOP. We showed that cTPs and unprocessed pre-proteins do accumulate in PREP mutants *in vivo* where CLP is fully functional, but not in mutants where only the Clp system is (partially) inactivated (Figure 3.3). These results support previous evidence suggesting that PREP is a dedicated cTP degrader. However, the fact that CLP interacts (genetically) with PREP (especially PREP2) and OOP suggests some interdependence and that they have partial overlap of substrates. It is also possible that the pool of

recycled amino acids is compromised in these CLP-PREP-OOP higher order mutants due to blocked proteolytic pathways. This may explain the reduced vigor of the triple and quadrupole mutants, especially in the young emerging leaves (Figure 3.2 and 3.7). Although not discussed in detail here, aminopeptidases are certainly part of this peptidase network (Teixeira et al., 2017).

Protein substrates for CLP are generally thought to be recognized and unfolded by the CLPC chaperone (Sauer and Baker, 2011). Certain substrates may be specifically selected by the CLPS1-F adaptors (Nishimura et al 2015), perhaps through recognition of specific N-degrons (Dissmeyer et al., 2017). The products of CLP proteolysis were shown in *E. coli* to be 5 - 12 residue peptides (Choi and Licht, 2005), placing some of these in the PREP and OOP substrate range. Other substrate classes likely enter this peptidase network laterally, following cleavage by known and so far uncharacterized peptidase systems; (van Wijk, 2015) provides a comprehensive list of candidate peptidases.

In conclusion, we provide details on a major proteolytic pathway in chloroplasts. The results presented here offers insight into the relative contributions of PREP, OOP and CLP to maintaining proteostasis.

3.5 MATERIALS AND METHODS

3.5.1 *Plant Growth, Mutant Isolation and RT-PCR analysis*

T-DNA insertion lines in Columbia-0 for *PREP1* (AT3g19170), *PREP2* (AT1g49630) and *OOP* (AT5g65620) are SALK_048944 (*prep1-1*), SALK_133220 (*prep2-1*), SALK_058439 (*oop-1*), SALK_061339 (*oop*), SAIL_150_D06 (*oop-3*), and

SAIL_1238_D07 (*oop-4*), respectively. The location of the T-DNA insertions was confirmed by DNA sequencing. Other mutants from the van Wijk lab have been described previously, *i.e.*, *clpr1-2* and *clpr2-1* (Kim et al., 2009), *clpt1-2* and *clpt2-1* (Kim et al., 2015), and *clpc1-1* (Olinares et al., 2011). Genotyping and RNA extraction were carried out as described previously (Rudella et al., 2006). Various growth conditions are detailed in the Figure legends. For RT-PCR, total RNA was isolated with an RNeasy plant mini kit (Qiagen). First strand was synthesized from equal amounts of total RNA with Superscript III Reverse Transcriptase (Invitrogen).

3.5.2 Chloroplast stroma and total leaf proteome isolation for analysis of CLP assembly states

For chloroplast stroma isolations, leaves of the wt (Col-0) and various mutant alleles were briefly homogenized in grinding medium (50 mM HEPES-KOH, pH 8.0, 330 mM sorbitol, 2mM EDTA, 5mM ascorbic acid, 5mM cysteine and 0.03% BSA) and filtered through a nylon mesh. The crude chloroplasts were then collected by a 2-min spin at $1100 \times g$ and further purified on 35-85% Percoll cushions (Percoll in 0.6% Ficoll, 1.8% polyethylene glycol) by a 10-min spin at $3750 \times g$ and one additional wash in the grinding medium without ascorbic acid, cysteine and BSA. Chloroplasts were subsequently lysed in 10 mM HEPES-KOH, pH 8.0, 5 mM $MgCl_2$, and 15% glycerol with a mixture of protease inhibitors under mild mechanical disruption. The lysate was then subjected to ultracentrifugation ($100,000 \times g$) to collect membranes as pellet and the stroma as supernatant. For total leaf proteome isolation under non-denaturing conditions, total leaf material was ground in liquid nitrogen and solubilized in 50 mM

HEPES-KOH, pH 8.0, 15% glycerol and 10 mM MgCl₂ with protease inhibitor cocktail. The suspension was then filtered in miracloth and spun at 100,000xg. The protease cocktail contained 1mM Pefabloc-SC, 50 µg/mL antipain, 40 µg/mL bestatin, 20 µg/mL chymostatin, 10 µg/mL E64, 10 µg/mL phosphoramidon, 1 µg/mL aprotinin.

3.5.3 Pigment, Native PAGE, Immunoblot and Protein concentrations

Chlorophyll and carotenoid contents on a fresh weight basis were measured in 80% acetone as described (Lichtenthaler, 1987). Light blue native PAGE was performed for separation of stromal and total leaf extracts under non-denaturing conditions using the Native PAGE Novex gel system (Invitrogen) with pre-cast 4-16% acrylamide Bis-Tris gels (Invitrogen). For immunoblots, proteins were blotted onto nitrocellulose or PVDF membranes and probed with antibodies using chemiluminescence for detection, following standard procedures. Antisera were generated in rabbits as described previously (Kim et al., 2013; Nishimura et al., 2013). Protein concentrations were determined using the BCA Protein Assay Kit (ThermoFisher).

3.5.4 Comparative Proteomics and Mass spectrometry

Three biological replicates per genotype were analyzed within each of the three sets of comparative proteomics experiments.

Experiment one: wt and the *prep1-1 prep2-1* double were grown for 20 days (wt) and 29 days (mutants) under a 10-h/14-h light/dark cycle at 100 µmol photons.m⁻².s⁻¹.

Experiment two: wt and quadrupole *prep1-1 prep2-1 clpt1-2 clpt2-1 (aabbccDd)* plants were grown for 21 days (wt) and 29 days (mutants) under a 10-h/14-h light/dark cycle at 100 $\mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$.

Experiment three: wt and homozygous *oop*, *prep1-1 prep2-1* and *prep1-1 prep2-1 oop-2* plants were grown for 20 days (wt and *oop-2*), 22 days (*prep1-1 prep2-1*), and 28 days (*prep1-1 prep2-1 oop-2*) under a 10-h/14-h light/dark cycle at 120 $\mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$.

Intact chloroplasts were isolated and total chloroplast proteomes used for analysis, whereas for experiments 2 and 3, total leaf proteomes were extracted in presence of SDS after grinding 250-350 mg fresh leaves in liquid N₂ into a fine powder, followed by removal of cell debris on frit spin columns (pore size 30 μm), as in (Friso et al., 2011). For all experiments, 50 μg total leaf protein of mutants and wt samples were each run out on a Biorad Criterion Tris-HCl precast gels (10.5-14% acrylamide gradient). Each gel lanes were cut into 12 bands followed by reduction, alkylation, and in-gel digestion with trypsin as described in (Friso et al., 2011).

3.5.5 Mass spectrometry, data processing and significance testing

The re-suspended peptide extracts were analyzed by data-dependent MS/MS using an on-line LC-LTQ-Orbitrap (ThermoFisher) as described in (Kim et al., 2015). Mass spectrometry data processing, data searching against TAIR10 using Mascot v2.4 and subsequent filtering and quantification based on normalized and adjusted spectral counts was carried out as in (Nishimura et al., 2013). Mass spectrometry-derived information, as well as annotation of protein name, location and function for the identified proteins can be found in the PPDB (<http://ppdb.tc.cornell.edu/>). Protein

functions were assigned using the MapMan bin system (Thimm et al., 2004) that we further curated and incorporated into the Plant Proteome Database (PPDB) at <http://ppdb.tc.cornell.edu>.

Pairwise significance analyses for genotypic differences within each experiment was carried out based on the combined outcome of two statistical packages, QSPEC and GLEE, specifically developed for spectral counting analysis, as described in (Kim et al., 2015). Proteins were deemed significantly different between genotypes at $p < 0.01$ using GLEE (using NadjSPC) and with $< 5\%$ False Discovery Rate using QSPEC (using AdjSPC).

The MS and proteomics data was submitted to the ProteomeXchange Consortium (Vizcaino et al., 2014) via the PRIDE partner repository (<http://www.ebi.ac.uk/pride>).

3.5.6 Plant growth for and protein extraction for TAILS experiments

Total leaf extracts for TAILS experiments were identical to those used for comparative proteomics (Experiment 3 above). Chloroplast stroma was prepared from Arabidopsis grown for between 6 and 9 weeks such that the plants were at a similar developmental stage but not yet flowering under a 10-h/14-h light/dark cycle at 100 or 150 $\mu\text{mol photons.m}^{-2}.\text{s}^{-1}$. The light conditions were identical for each experiment comparing wt and mutant Arabidopsis tissue. Chloroplast stroma for TAILS comparison of wt and *prep1 prep2* was extracted from the same plants used for comparative proteomics above.

3.5.7 TAILS methodology

Each TAILS experiment was performed in biological triplicate and for each sample, two technical replicates were performed with the isotopic dimethyl label switched (e.g. wt 'light' vs. mutant 'heavy' and wt 'heavy' vs. mutant 'light'). Each sample was injected onto the LC/MS system twice (Supplemental Figure 3.5). Exceptions were made for *prep1 prep2* TAILS and *prep1 prep2 oop* LMW-TAILS experiments where three biological replicates were used but no label switch (technical replicate) was performed. See Supplemental Table 3.3A for details of samples analyzed.

Total soluble leaf extracts and chloroplast stroma were prepared as previously described (Friso 2011; Olinares, 2010). The extraction buffer with protease inhibitors was as follows, 50 mM HEPES, pH 8, 1mM Pefabloc-SC, 50 µg/mL antipain, 40 µg/mL bestatin, 20 µg/mL chymostatin, 10 µg/mL E64, 10 µg/mL phosphoramidon, 1 µg/mL aprotinin. We followed the method TAILS method (Kleifeld, 2011) described previously with minor modifications (Rowland 2015). The dimethyl labeling reaction was quenched with 0.1 M glycine. All incubations were performed at 37°C in an Eppendorf ThermoMixer with light shaking. Between 600 and 800 µg of HPG-ALD resin was used for each 100 µg of digested peptides. Following overnight reaction of peptides with HPG-ALD resin and quenching with 0.1 M ammonium bicarbonate, samples were acidified with formic acid prior to removal of polymer bound internal peptides by ultrafiltration using Amicon 30-kD molecular mass filters (Millipore). The flow through was then immediately desalted using C18 SPE spin columns (ThermoFisher 89870).

3.5.8 TAILS database searching, filtering and curation. Quantification with Proteome Discoverer 2.1 and Skyline

Analysis of mass spectrometry data from TAILS experiments including peptide identification and relative peptide quantification, was performed with Proteome Discoverer™ 2.1 (ThermoFisher). Two database searches were conducted with MASCOT v2.4 (Matrix Science) and Sequest HT (ThermoFisher). Two separate database searches (MASCOT v2.4) were performed for each sample to identify light and heavy labeled peptides using parameters outlined for trypsin digested samples in (Rowland et al., 2015). The results of the database search were pooled and filtered using Percolator; the max Delta Cn was 0.05, target decoy false discovery rate was 0.01, validation was based on q-value. The peptide quantification mode was set to dimethylation duplex using the light and medium labels. Exported peptide group lists were filtered to remove unlabeled (not Nt peptides) and peak areas for wt and mutant peaks were used to calculate mutant to wt ratios. The average ratio from two LC-MS injections was used for each sample. Average values and standard deviations were calculated for each triplicate experiment for mutant (light) versus wt (heavy) and for the reciprocal labeling experiment (Supplemental Table 3.4). Differentially accumulating peptides (at least two fold up or down in the mutant) were selected for verification along with other peptides matching to those proteins, regardless of their wt to mutant ratio. These peptide lists were imported in Skyline (MacLean et al., 2010) and the MS1 filtering feature was used to (i) manually confirm that the peptide identification was correct and (ii) to generate a second set of peak area ratios and inspect for reproducibility across replicates (Supplemental Table 3.5).

Low molecular weight TAILS experiments were conducted as above except 1 mg dimethyl labeled protein per replicate was filtered through 30-kDa molecular mass filters and the filtrate carried forward before the negative selection step to remove large and intact proteins.

3.5.9 Gel based Nti determination. Nt profiling without enrichment Nt peptides.

Proteins were reduced and alkylated and dimethyl labeled in 4M GuHCl, 50mM HEPES, pH 8 as for TAILS experiments. Wt (light) and *prep1 prep2 oop* (heavy) samples were then mixed and acetone precipitated. The pellet was suspended in SDS sample buffer and the proteins resolved by Tris-Tricine SDS PAGE on 16.5 % acrylamide precast gels (Biorad, Criterion). Gel bands were cut and digested as described previously (Friso et al., 2011), with water washes in place of DTT and iodoacetamide incubations. Peptide extracts were analyzed by data-dependent LC-MS/MS (see section 3.5.5). Two separate database searches (MASCOT v2.4) were performed for each sample to identify light and heavy labeled peptides using parameters outlined for trypsin digested samples in (Rowland et al., 2015). Database search results were filtered in MS Excel to include only N-terminally labeled peptides. Un-adjusted spectral counts (SPC) for wt and mutant were summed for unique peptides from each replicate.

3.5.10 Peptidome analysis

Two different strategies were used to purify either total leaf or chloroplast peptidomes. The first technique was loosely based on that described by (Lyons and

Fricker, 2011). Soluble protein extracts containing 1 - 2 mg total protein were rapidly heated to 90°C, incubated for 3 min then cooled on ice before acidification with 0.01 M HCl (final concentration). The protein precipitate was pelleted by centrifuge at 14,000 x g, 4°C, for 10 min and the supernatant transferred to an Amicon, 30-kD ultrafiltration device (Millipore). Half of each peptide extract was neutralized with ammonium bicarbonate (200 mM final concentration) and digested with 0.5 µg trypsin for 16 hr. The second technique loosely followed the protocol described by (Xu et al., 2015). Total leaf tissue (~1 g fresh weight) was ground under liquid nitrogen, or purified chloroplasts (see section 3.5.2) were extracted with 0.5% formic acid with protease inhibitors (0.5 mM EDTA, 0.5 mM pefabloc SC, 5 µM E64) on ice. Chloroplasts were lysed in a glass potter in extraction solution. Extracts were then clarified by centrifuge at 20k x g at 4°C for and filtered using either 10- or 30-kD ultrafiltration devices. The above filtrates were desalted by C-18 SPE spin columns (Pierce 89870) using the manufacturer's guidelines. Extracts were brought to dryness in a Speed Vac (ThermoFisher) and then re-suspended in 2% formic acid, 3% acetonitrile for LC-MS/MS analysis. Mass spectrometry was performed on an LTQ Orbitrap as previously described with the following adjustments (section 3.5.5). The LC gradients were 120 or 140 min in length and the normalized CE was set to 35 to facilitate fragmentation of large peptides. See Supplemental Table 3.3B for a list of peptidome experiments conducted. *No enzyme* data base searches were conducted with MASCOT v2.4 with the following variable modifications: Acetyl (N-term), Deamidated (N,Q), Formyl (N-term) and oxidized (M). The precursor mass tolerance was 6 ppm; fragment ion mass tolerance 0.8 Da, Two missed cleavages were allowed. For samples digested with trypsin, a *semitryptic* data base search was conducted

with variable modifications: Acetyl (N-term) and oxidized (M); precursor mass tolerance 20 ppm; fragment ion mass tolerance 0.8 Da.

LITERATURE CITED

- Alexopoulos JA, Guarne A, Ortega J** (2012) ClpP: a structurally dynamic protease regulated by AAA+ proteins. *J Struct Biol* **179**: 202-210
- Apel W, Schulze WX, Bock R** (2010) Identification of protein stability determinants in chloroplasts. *Plant J* **63**: 636-650
- Apitz J, Nishimura K, Schmied J, Wolf A, Hedtke B, van Wijk KJ, Grimm B** (2016) Posttranslational control of ALA synthesis includes GluTR degradation by Clp protease and stabilization by GluTR-binding protein. *Plant Physiol*
- Atkinson N, Leitao N, Orr DJ, Meyer MT, Carmo-Silva E, Griffiths H, Smith AM, McCormick AJ** (2017) Rubisco small subunits from the unicellular green alga *Chlamydomonas* complement Rubisco-deficient mutants of *Arabidopsis*. *New Phytol* **214**: 655-667
- Battesti A, Gottesman S** (2013) Roles of adaptor proteins in regulation of bacterial proteolysis. *Curr Opin Microbiol* **16**: 140-147
- Carrie C, Venne AS, Zahedi RP, Soll J** (2015) Identification of cleavage sites and substrate proteins for two mitochondrial intermediate peptidases in *Arabidopsis thaliana*. *J Exp Bot* **66**: 2691-2708
- Cederholm SN, Backman HG, Pesaresi P, Leister D, Glaser E** (2009) Deletion of an organellar peptidosome PreP affects early development in *Arabidopsis thaliana*. *Plant Mol Biol* **71**: 497-508
- Choi KH, Licht S** (2005) Control of peptide product sizes by the energy-dependent protease ClpAP. *Biochemistry* **44**: 13921-13931
- Colaert N, Helsens K, Martens L, Vandekerckhove J, Gevaert K** (2009) Improved visualization of protein consensus sequences by iceLogo. *Nat Methods* **6**: 786-787
- Constan D, Froehlich JE, Rangarajan S, Keegstra K** (2004) A stromal Hsp100 protein is required for normal chloroplast development and function in *Arabidopsis*. *Plant Physiol* **136**: 3605-3615
- Demir F, Niedermaier S, Villamor JG, Huesgen PF** (2017) Quantitative proteomics in plant protease substrate identification. *New Phytol*
- Dissmeyer N, Rivas S, Graciet E** (2017) Life and death of proteins after protease cleavage: protein degradation by the N-end rule pathway. *New Phytol*

- Emanuelsson O, Brunak S, von Heijne G, Nielsen H** (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* **2**: 953-971
- Fricker LD, Gelman JS, Castro LM, Gozzo FC, Ferro ES** (2012) Peptidomic analysis of HEK293T cells: effect of the proteasome inhibitor epoxomicin on intracellular peptides. *J Proteome Res* **11**: 1981-1990
- Friso G, Olinares PD, van Wijk KJ** (2011) The workflow for quantitative proteome analysis of chloroplast development and differentiation, chloroplast mutants, and protein interactions by spectral counting. *Methods Mol Biol* **775**: 265-282
- Friso G, Olinares PDB, van Wijk KJ** (2011) The Workflow for Quantitative Proteome Analysis of Chloroplast Development and Differentiation, Chloroplast Mutants, and Protein Interactions by Spectral Counting. *In* RP Jarvis, ed, *Chloroplast Research in Arabidopsis*, Vol 775. Humana Press, New York, pp 265-282
- Glaser E, Alikhani N** (2010) The organellar peptidosome, PreP: a journey from Arabidopsis to Alzheimer's disease. *Biochim Biophys Acta* **1797**: 1076-1080
- Hartl FU, Bracher A, Hayer-Hartl M** (2011) Molecular chaperones in protein folding and proteostasis. *Nature* **475**: 324-332
- Houtz RL, Magnani R, Nayak NR, Dirk LM** (2008) Co- and post-translational modifications in Rubisco: unanswered questions. *J Exp Bot* **59**: 1635-1645
- Huang S, Nelson CJ, Li L, Taylor NL, Stroher E, Peteriet J, Millar AH** (2015) INTERMEDIATE CLEAVAGE PEPTIDASE55 Modifies Enzyme Amino Termini and Alters Protein Stability in Arabidopsis Mitochondria. *Plant Physiol* **168**: 415-427
- Huesgen PF, Alami M, Lange PF, Foster LJ, Schroder WP, Overall CM, Green BR** (2013) Proteomic amino-termini profiling reveals targeting information for protein import into complex plastids. *PLoS One* **8**: e74483
- Jarvis P, Lopez-Juez E** (2013) Biogenesis and homeostasis of chloroplasts and other plastids. *Nat Rev Mol Cell Biol* **14**: 787-802
- Johnson KA, Bhushan S, Stahl A, Hallberg BM, Frohn A, Glaser E, Eneqvist T** (2006) The closed structure of presequence protease PreP forms a unique 10,000 Angstroms³ chamber for proteolysis. *EMBO J* **25**: 1977-1986
- Julien O, Wells JA** (2017) Caspases and their substrates. *Cell Death Differ* **24**: 1380-1389

- Kim J, Kimber MS, Nishimura K, Friso G, Schultz L, Ponnala L, van Wijk KJ** (2015) Structures, Functions, and Interactions of ClpT1 and ClpT2 in the Clp Protease System of Arabidopsis Chloroplasts. *Plant Cell* **27**: 1477-1496
- Kim J, Olinares PD, Oh SH, Ghisaura S, Poliakov A, Ponnala L, van Wijk KJ** (2013) Modified Clp protease complex in the ClpP3 null mutant and consequences for chloroplast development and function in Arabidopsis. *Plant Physiol* **162**: 157-179
- Kim J, Rudella A, Ramirez Rodriguez V, Zybaïlov B, Olinares PD, van Wijk KJ** (2009) Subunits of the plastid ClpPR protease complex have differential contributions to embryogenesis, plastid biogenesis, and plant development in Arabidopsis. *Plant Cell* **21**: 1669-1692
- King JV, Liang WG, Scherpelz KP, Schilling AB, Meredith SC, Tang WJ** (2014) Molecular basis of substrate recognition and degradation by human presequence protease. *Structure* **22**: 996-1007
- Kirstein J, Moliere N, Dougan DA, Turgay K** (2009) Adapting the machine: adaptor proteins for Hsp100/Clp and AAA+ proteases. *Nat Rev Microbiol* **7**: 589-599
- Kleifeld O, Doucet A, auf dem Keller U, Prudova A, Schilling O, Kainthan RK, Starr AE, Foster LJ, Kizhakkedathu JN, Overall CM** (2010) Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat Biotechnol* **28**: 281-288
- Kmiec B, Teixeira PF, Berntsson RP, Murcha MW, Branca RM, Radomiljac JD, Regberg J, Svensson LM, Bakali A, Langel U, Lehtio J, Whelan J, Stenmark P, Glaser E** (2013) Organellar oligopeptidase (OOP) provides a complementary pathway for targeting peptide degradation in mitochondria and chloroplasts. *Proc Natl Acad Sci U S A* **110**: E3761-3769
- Kmiec B, Teixeira PF, Glaser E** (2014) Phenotypical consequences of expressing the dually targeted Presequence Protease, AtPreP, exclusively in mitochondria. *Biochimie* **100**: 167-170
- Kmiec B, Teixeira PF, Glaser E** (2014) Shredding the signal: targeting peptide degradation in mitochondria and chloroplasts. *Trends Plant Sci* **19**: 771-778
- Larkin RM, Stefano G, Ruckle ME, Stavoe AK, Sinkler CA, Brandizzi F, Malmstrom CM, Osteryoung KW** (2016) REDUCED CHLOROPLAST COVERAGE genes from Arabidopsis thaliana help to establish the size of the chloroplast compartment. *Proc Natl Acad Sci U S A* **113**: E1116-1125

- Leaden L, Busi MV, Gomez-Casati DF** (2014) The mitochondrial proteins AtHscB and AtIsc1 involved in Fe-S cluster assembly interact with the Hsp70-type chaperon AtHscA2 and modulate its catalytic activity. *Mitochondrion* **19 Pt B**: 375-381
- Li L, Millar AH** (2017) Protein Degradation Rate in Arabidopsis thaliana Leaf Growth and Development. *The Plant Cell Online*
- Lichtenthaler HK** (1987) Chlorophylls and Carotenoids: Pigments of photosynthetic Biomembranes. *Methods in Enzymology* **148**: 350-382
- Liu K, Ologbenla A, Houry WA** (2014) Dynamics of the ClpP serine protease: A model for self-compartmentalized proteases. *Crit Rev Biochem Mol Biol* **49**: 400-412
- Lyons PJ, Fricker LD** (2011) Peptidomic approaches to study proteolytic activity. *Curr Protoc Protein Sci* **Chapter 18**: Unit18 13
- Mach J, Poliak P, Matuskova A, Zarsky V, Janata J, Lukes J, Tachezy J** (2013) An Advanced System of the Mitochondrial Processing Peptidase and Core Protein Family in Trypanosoma brucei and Multiple Origins of the Core I Subunit in Eukaryotes. *Genome Biol Evol* **5**: 860-875
- MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ** (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**: 966-968
- Mossmann D, Vogtle FN, Taskin AA, Teixeira PF, Ring J, Burkhart JM, Burger N, Pinho CM, Tadic J, Loreth D, Graff C, Metzger F, Sickmann A, Kretz O, Wiedemann N, Zahedi RP, Madeo F, Glaser E, Meisinger C** (2014) Amyloid-beta peptide induces mitochondrial dysfunction by inhibition of preprotein maturation. *Cell Metab* **20**: 662-669
- Nishimura K, Apitz J, Friso G, Kim J, Ponnala L, Grimm B, van Wijk KJ** (2015) Discovery of a Unique Clp Component, ClpF, in Chloroplasts: A Proposed Binary ClpF-ClpS1 Adaptor Complex Functions in Substrate Recognition and Delivery. *Plant Cell*
- Nishimura K, Asakura Y, Friso G, Kim J, Oh SH, Rutschow H, Ponnala L, van Wijk KJ** (2013) ClpS1 is a conserved substrate selector for the chloroplast Clp protease system in Arabidopsis. *Plant Cell* **25**: 2276-2301

- Nishimura K, Kato Y, Sakamoto W** (2016) Chloroplast Proteases: Updates on Proteolysis within and across Suborganellar Compartments. *Plant Physiol* **171**: 2280-2293
- Nishimura K, Kato Y, Sakamoto W** (2017) Essentials of Proteolytic Machineries in Chloroplasts. *Mol Plant* **10**: 4-19
- Nishimura K, van Wijk KJ** (2015) Organization, function and substrates of the essential Clp protease system in plastids. *Biochim Biophys Acta* **1847**: 915-930
- Olinares PD, Kim J, Davis JI, van Wijk KJ** (2011) Subunit stoichiometry, evolution, and functional implications of an asymmetric plant plastid ClpP/R protease complex in Arabidopsis. *Plant Cell* **23**: 2348-2361
- Poveda-Huertes D, Mulica P, Vogtle FN** (2017) The versatility of the mitochondrial presequence processing machinery: cleavage, quality control and turnover. *Cell Tissue Res* **367**: 73-81
- Rampello AJ, Glynn SE** (2017) Identification of a Degradation Signal Sequence within Substrates of the Mitochondrial i-AAA Protease. *J Mol Biol* **429**: 873-885
- Richter S, Lamppa GK** (1998) A chloroplast processing enzyme functions as the general stromal processing peptidase. *Proc Natl Acad Sci U S A* **95**: 7463-7468
- Richter S, Lamppa GK** (1999) Stromal processing peptidase binds transit peptides and initiates their ATP-dependent turnover in chloroplasts. *J Cell Biol* **147**: 33-44
- Richter S, Lamppa GK** (2002) Determinants for removal and degradation of transit peptides of chloroplast precursor proteins. *J Biol Chem* **277**: 43888-43894
- Rowland E, Kim J, Bhuiyan NH, van Wijk KJ** (2015) The Arabidopsis Chloroplast Stromal N-Terminome: Complexities of Amino-Terminal Protein Maturation and Stability. *Plant Physiol* **169**: 1881-1896
- Rudella A, Friso G, Alonso JM, Ecker JR, van Wijk KJ** (2006) Downregulation of ClpR2 Leads to Reduced Accumulation of the ClpPRS Protease Complex and Defects in Chloroplast Biogenesis in Arabidopsis. *Plant Cell* **18**: 1704-1721
- Rutschow H, Ytterberg AJ, Friso G, Nilsson R, van Wijk KJ** (2008) Quantitative proteomics of a chloroplast SRP54 sorting mutant and its genetic interactions with CLPC1 in Arabidopsis. *Plant Physiol* **148**: 156-175

- Sauer RT, Baker TA** (2011) AAA+ proteases: ATP-fueled machines of protein destruction. *Annu Rev Biochem* **80**: 587-612
- Stahl A, Nilsson S, Lundberg P, Bhushan S, Biverstahl H, Moberg P, Morisset M, Vener A, Maler L, Langel U, Glaser E** (2005) Two novel targeting peptide degrading proteases, PrePs, in mitochondria and chloroplasts, so similar and still different. *J Mol Biol* **349**: 847-860
- Teixeira PF, Glaser E** (2013) Processing peptidases in mitochondria and chloroplasts. *Biochim Biophys Acta* **1833**: 360-370
- Teixeira PF, Kmiec B, Branca RM, Murcha MW, Byzia A, Ivanova A, Whelan J, Drag M, Lehtio J, Glaser E** (2017) A multi-step peptidolytic cascade for amino acid recovery in chloroplasts. *Nat Chem Biol* **13**: 15-17
- Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Kruger P, Selbig J, Muller LA, Rhee SY, Stitt M** (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J* **37**: 914-939
- Trosch R, Jarvis P** (2011) The stromal processing peptidase of chloroplasts is essential in Arabidopsis, with knockout mutations causing embryo arrest after the 16-cell stage. *PLoS One* **6**: e23039
- van Wijk KJ** (2015) Protein maturation and proteolysis in plant plastids, mitochondria, and peroxisomes. *Annu Rev Plant Biol* **66**: 75-111
- Varshavsky A** (2011) The N-end rule pathway and regulation by proteolysis. *Protein Sci* **20**: 1298-1345
- Vizcaino JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Rios D, Dianas JA, Sun Z, Farrah T, Bandeira N, Binz PA, Xenarios I, Eisenacher M, Mayer G, Gatto L, Campos A, Chalkley RJ, Kraus HJ, Albar JP, Martinez-Bartolome S, Apweiler R, Omenn GS, Martens L, Jones AR, Hermjakob H** (2014) ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol* **32**: 223-226
- Vogtle FN, Prinz C, Kellermann J, Lottspeich F, Pfanner N, Meisinger C** (2011) Mitochondrial protein turnover: role of the precursor intermediate peptidase Oct1 in protein stabilization. *Mol Biol Cell* **22**: 2135-2143
- Vogtle FN, Wortelkamp S, Zahedi RP, Becker D, Leidhold C, Gevaert K, Kellermann J, Voos W, Sickmann A, Pfanner N, Meisinger C** (2009) Global analysis of the mitochondrial N-proteome identifies a processing peptidase critical for protein stability. *Cell* **139**: 428-439

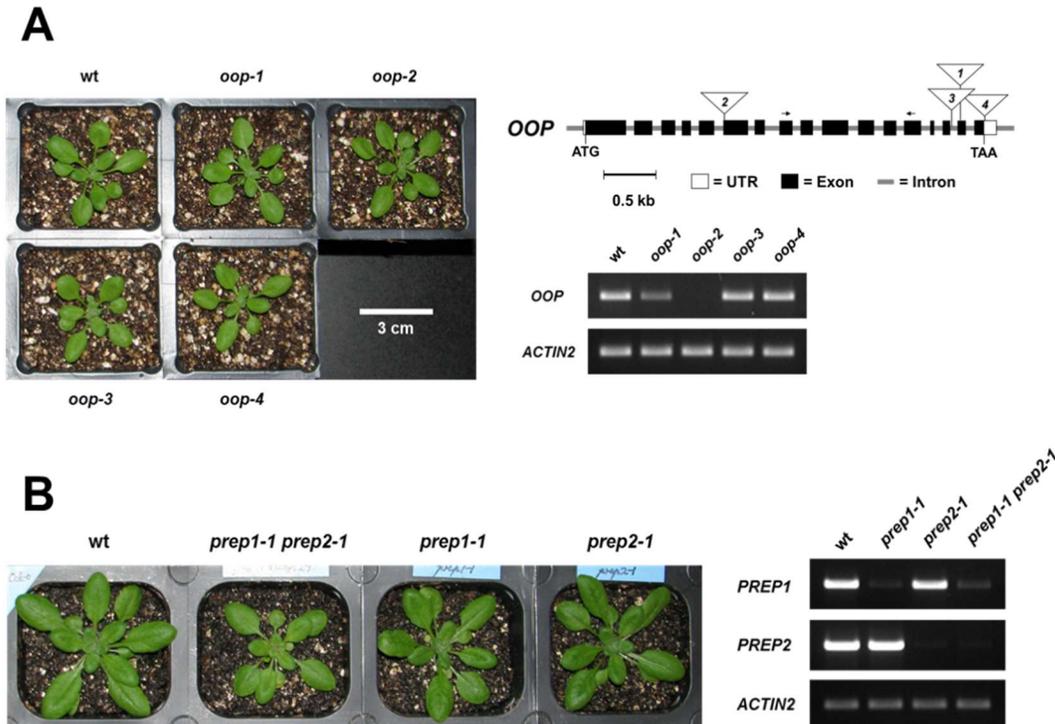
- Wu C, Monroe ME, Xu Z, Slyszy GW, Payne SH, Rodland KD, Liu T, Smith RD** (2015) An Optimized Informatics Pipeline for Mass Spectrometry-Based Peptidomics. *J Am Soc Mass Spectrom* **26**: 2002-2008
- Xu Z, Wu C, Xie F, Slyszy GW, Tolic N, Monroe ME, Petyuk VA, Payne SH, Fujimoto GM, Moore RJ, Fillmore TL, Schepmoes AA, Levine DA, Townsend RR, Davies SR, Li S, Ellis M, Boja E, Rivers R, Rodriguez H, Rodland KD, Liu T, Smith RD** (2015) Comprehensive quantitative analysis of ovarian and breast cancer tumor peptidomes. *J Proteome Res* **14**: 422-433
- Zhong R, Wan J, Jin R, Lamppa G** (2003) A pea antisense gene for the chloroplast stromal processing peptidase yields seedling lethals in Arabidopsis: survivors show defective GFP import in vivo. *Plant J* **34**: 802-812
- Zybailov B, Friso G, Kim J, Rudella A, Rodriguez VR, Asakura Y, Sun Q, van Wijk KJ** (2009) Large scale comparative proteomics of a chloroplast Clp protease mutant reveals folding stress, altered protein homeostasis, and feedback regulation of metabolism. *Mol Cell Proteomics* **8**: 1789-1810

APPENDIX

Supplemental Figure 3.1. Characterization of OOP and PREP mutants

(A) Comparison of wt, *oop-1*, *oop-2*, *oop-3* and *oop-4*. Plants were grown on soil for 23 days under a 10-h/14-h light/dark cycle at $120 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$. Gene model structures and position of T-DNA inserts in the *OOP* mutants used in this study. Exons (black boxes for coding sequence; open boxes for 5' and 3' untranslated regions [UTRs]) T-DNA insertions (triangles), and RT-PCR primers (arrows) are indicated. Transcript accumulation levels in the leaves of wt, *oop-1*, *oop-2*, *oop-3* and *oop-4*. Transcript levels were determined by RT-PCR using gene-specific primer pairs; ACTIN2 was used as internal control.

(B) Comparison of wt, *prep1-1 prep2-1*, *prep1-1* and *prep2-1*. Plants were grown on soil for 25 days under a 10-h/14-h light/dark cycle at $120 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$. Transcript accumulation levels in the leaves of wt, *prep1-1 prep2-1*, *prep1-1* and *prep2-1*. Transcript levels were determined by RT-PCR using gene-specific primer pairs; ACTIN2 was used as internal control.



Supplemental Figure 3.2. Characterization of *prep1 prep2 oop* at later stages of development; seed phenotype and germination rate.

(A) *prep1 prep2 oop* and wt plants grown on soil under a 10-h/14-h light/dark cycle at $150 \mu\text{mol photons.m}^{-2}.\text{s}^{-1}$ for 58 days then transferred to a 10-h/14-h light/dark cycle at $500 \mu\text{mol photons.m}^{-2}.\text{s}^{-1}$ for 11 days.

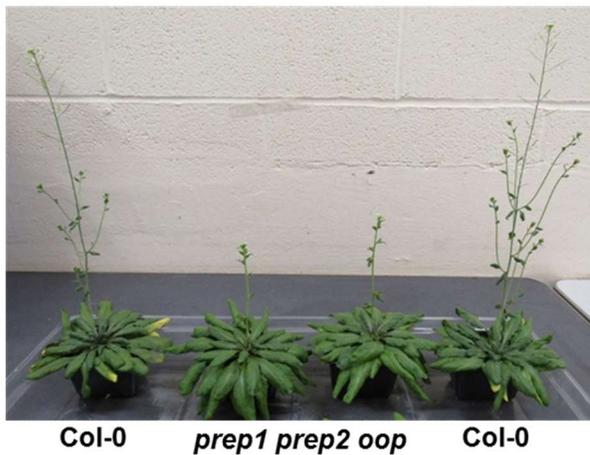
(B) *prep1 prep2 oop* and wt plants grown on soil under a 10-h/14-h light/dark cycle at $150 \mu\text{mol photons.m}^{-2}.\text{s}^{-1}$ for 67 days

(C) *prep1 prep2 oop* and wt plants grown on soil under a 10-h/14-h light/dark cycle at $150 \mu\text{mol photons.m}^{-2}.\text{s}^{-1}$ for 78 days

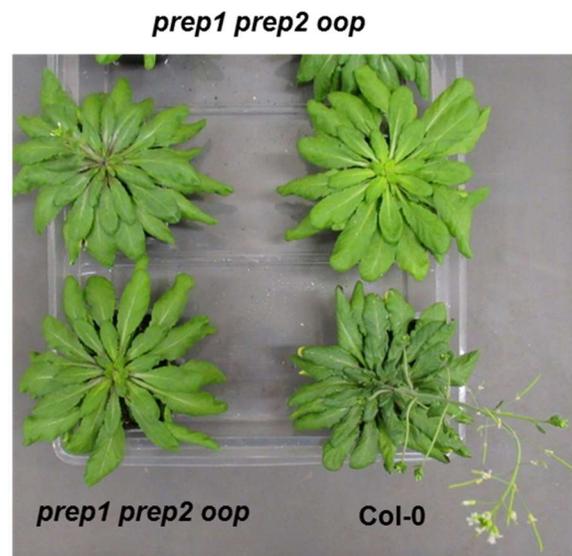
(D) Seeds harvested from above *prep1 prep2 oop* and wt plants

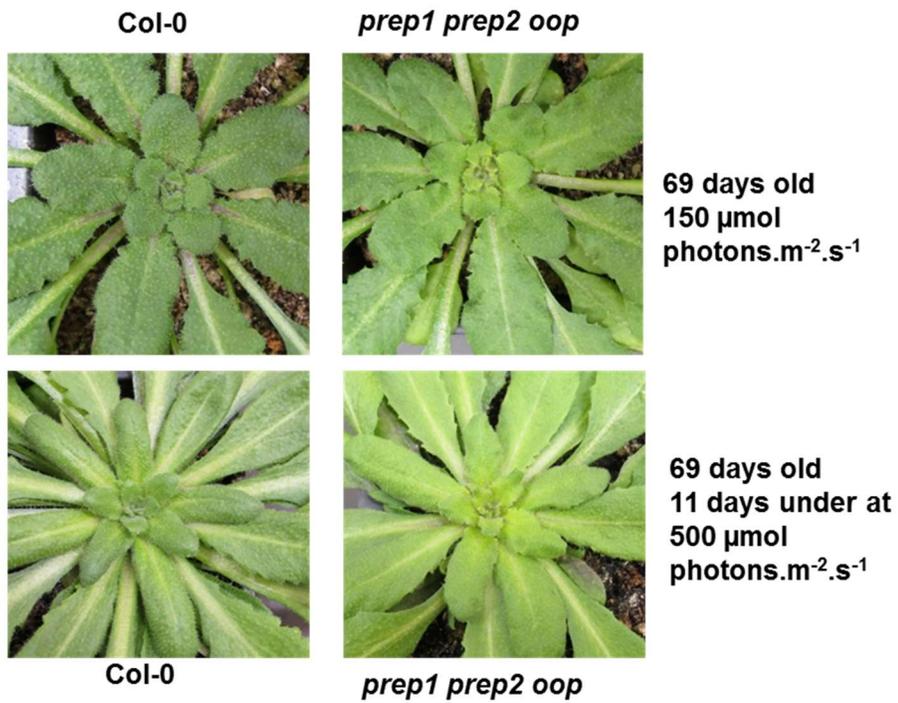
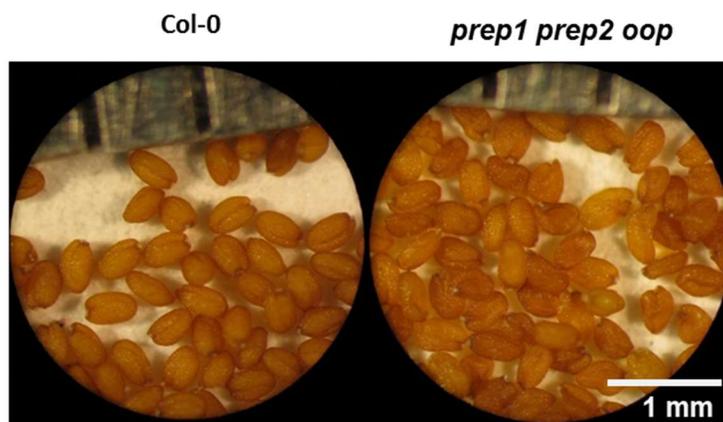
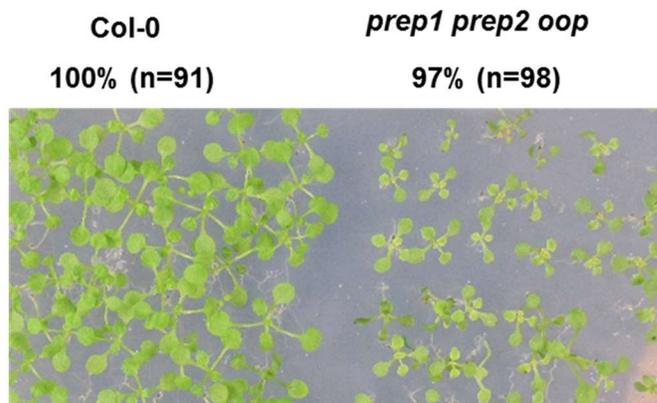
(E) *prep1 prep2 oop* and wt plants grown on $\frac{1}{2}$ MS medium, 2% sucrose under a 10-h/14-h light/dark cycle at $150 \mu\text{mol photons.m}^{-2}.\text{s}^{-1}$ for 16 days for determination of germination rate.

A



B



C**D****E**

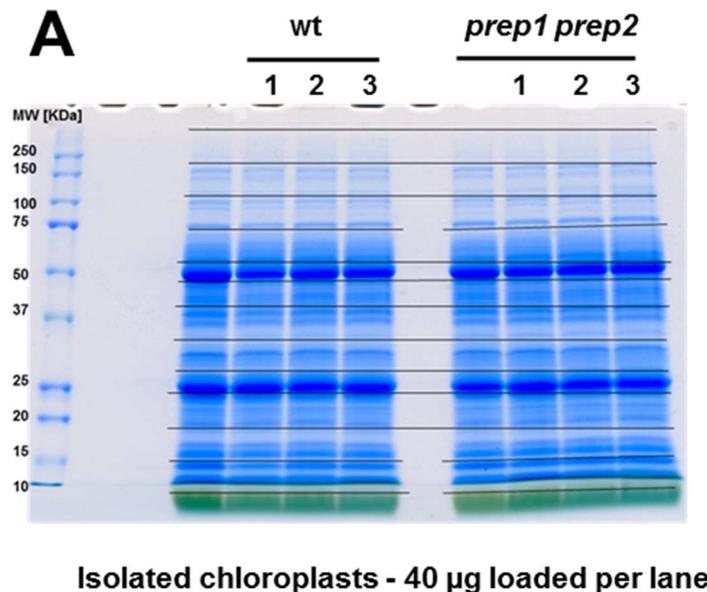
Supplemental Figure 3.3. Quantitative comparison of chloroplast proteomes of proteomics wt and *prep1 prep2* with triple biological replicates of soil grown plants in developmental stage 1.11 using label-free spectral counting. The *prep1-1 prep2-1* and wt plants used for chloroplast proteome analysis. Plants were grown for 6 weeks on soil under a 10-h/14-h light/dark cycle at $120 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$.

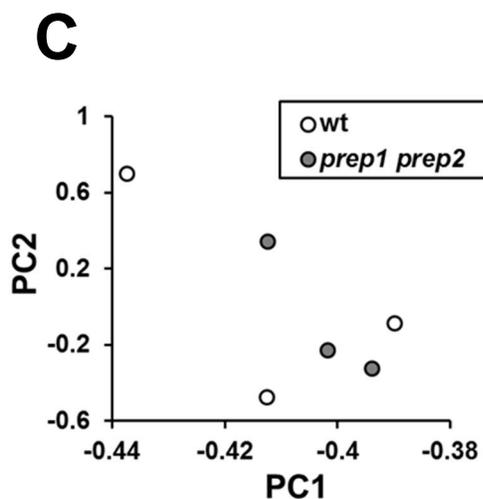
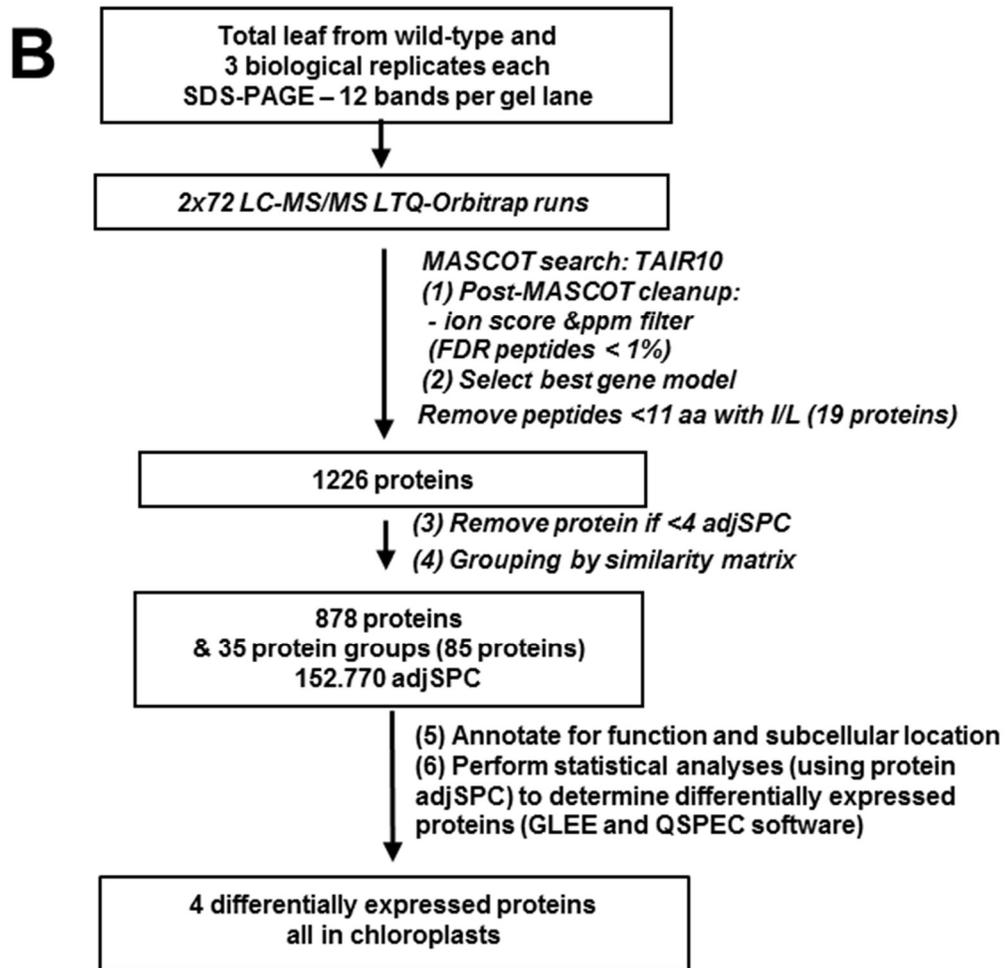
(A) Coomassie blue-stained SDS-PAGE gel with four biological replicates of stromal proteomes for each genotype. Three replicates were used for the MSMS analysis. Complete gel lanes were sliced into 12 sections per gel lane.

(B) Complete experimental and bioinformatics proteomics workflow. Proteins were in-gel digested by trypsin and analyzed by LC-MS/MS. After database searching, processing and protein annotation, the data were analyzed statistically to determine genotypic effects. Supplemental Table S1A,B summarizes the identified proteins, their annotations, quantification and fold changes with the statistical significance levels (p-values and false-discovery rates) of differences between genotypes.

(C) Principle component analysis (PCA) based on NadSPC of both genotypes.

(D) Correlation coefficients between genotypes and replicates based on NadjSPC.





D

NadjSPC	wt-rep1	wt-rep2	wt-rep3	p1xp2-rep1	p1xp2-rep2	p1xp2-rep3
wt-rep1	1.000					
wt-rep2	0.976	1.000				
wt-rep3	0.963	0.982	1.000			
p1xp2-rep1	0.982	0.978	0.969	1.000		
p1xp2-rep2	0.973	0.991	0.989	0.981	1.000	
p1xp2-rep3	0.964	0.981	0.985	0.979	0.987	1.000

Supplemental Figure 3.4. Quantitative comparison of total leaf proteomes of proteomics wt, *prep1 prep2*, *oop* and the triple mutant. Three biological replicates of soil grown plants under a 10-h/14-h light/dark cycle at $120 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ and leaves harvested at developmental stage 1.11. Plants were grown for 20 days for *oop-2* and the wt, 22 days for *prep1 prep2* and 28 days for *prep1 prep2 oop-2*.

(A) Protein extracts were run on a SDS-PAGE gel, stained by Coomassie blue, and complete gel lanes were sliced into sections (12 per gel lane) for protein in-gel digestion by trypsin and analysis by LC-MS/MS.

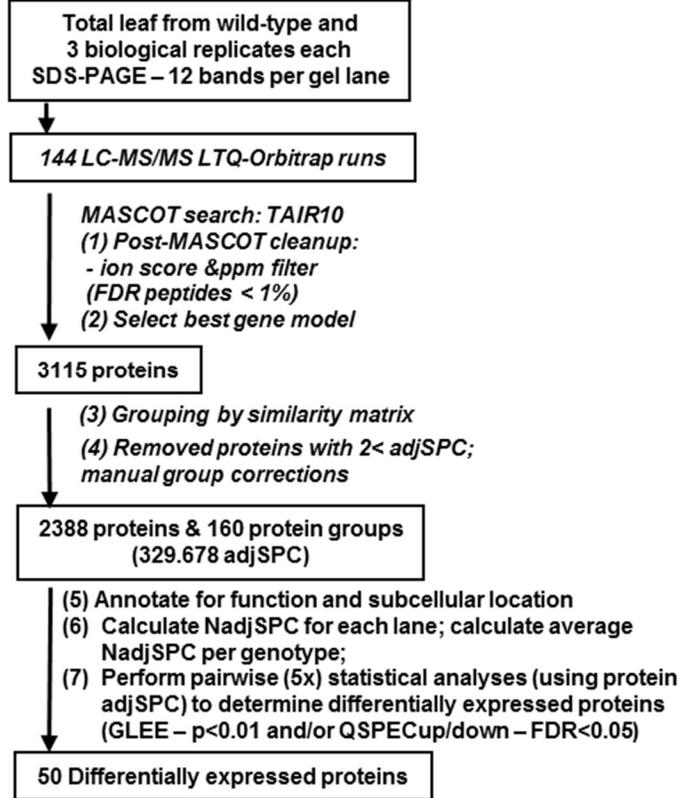
(B) Complete experimental and bioinformatics proteomics workflow. Supplemental Table S2A,B summarizes the identified proteins, their annotations, quantification and fold changes with the statistical significance levels (p-values and false-discovery rates) of differences between genotypes.

(C) Principle component analysis (PCA) based on NadSPC of both genotypes.

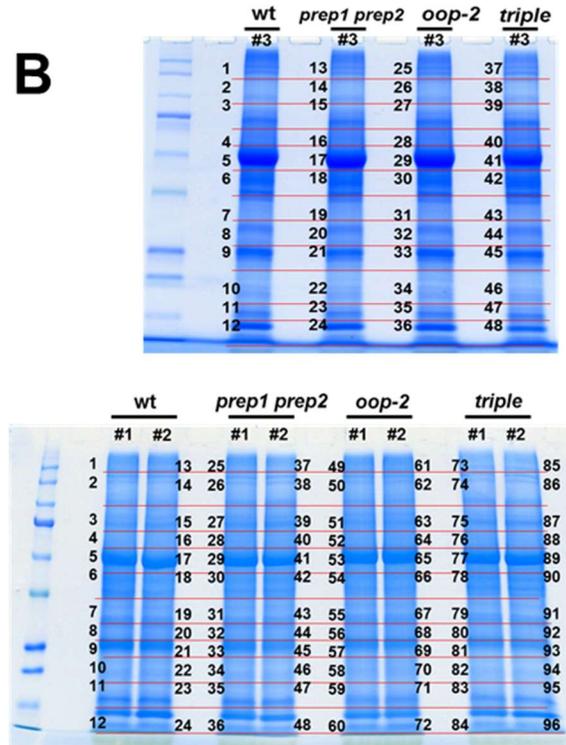
(D) Correlation coefficients between genotypes and replicates based on NadjSPC.

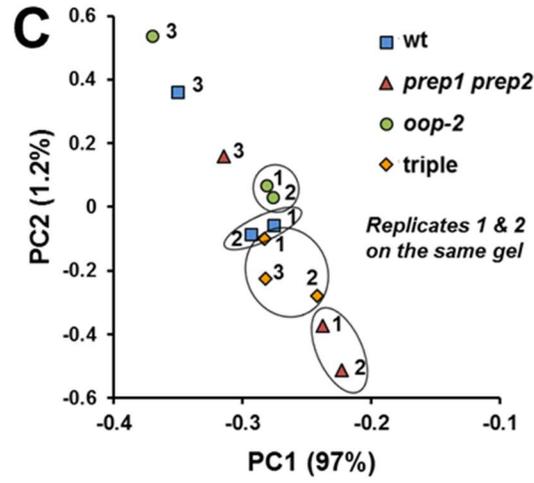
(E) Differential investments in specific chloroplast functions across the four genotypes as determined by relative protein abundance. Standard deviations are indicated.

A



B

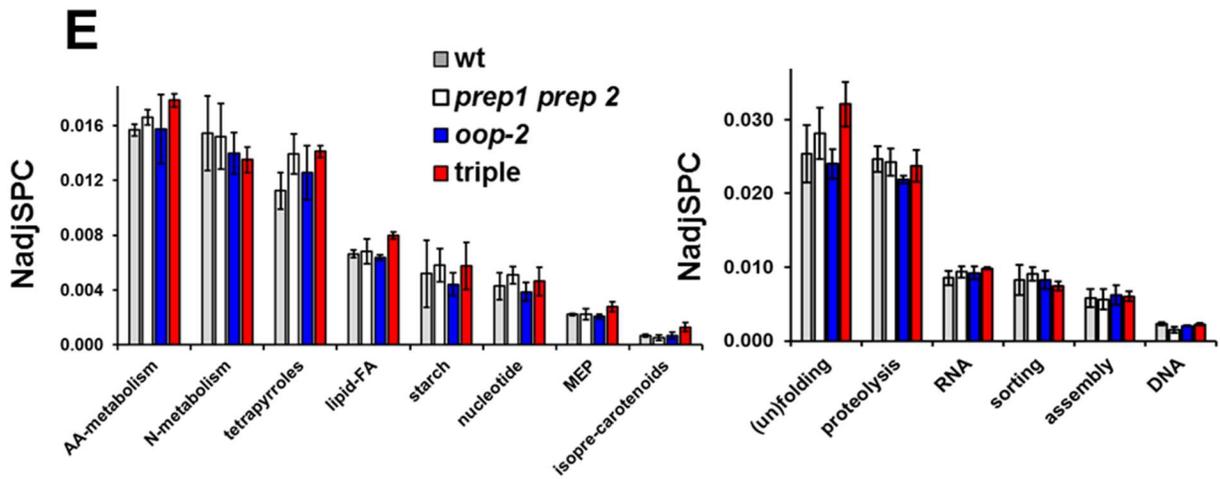




D

NadjSPC

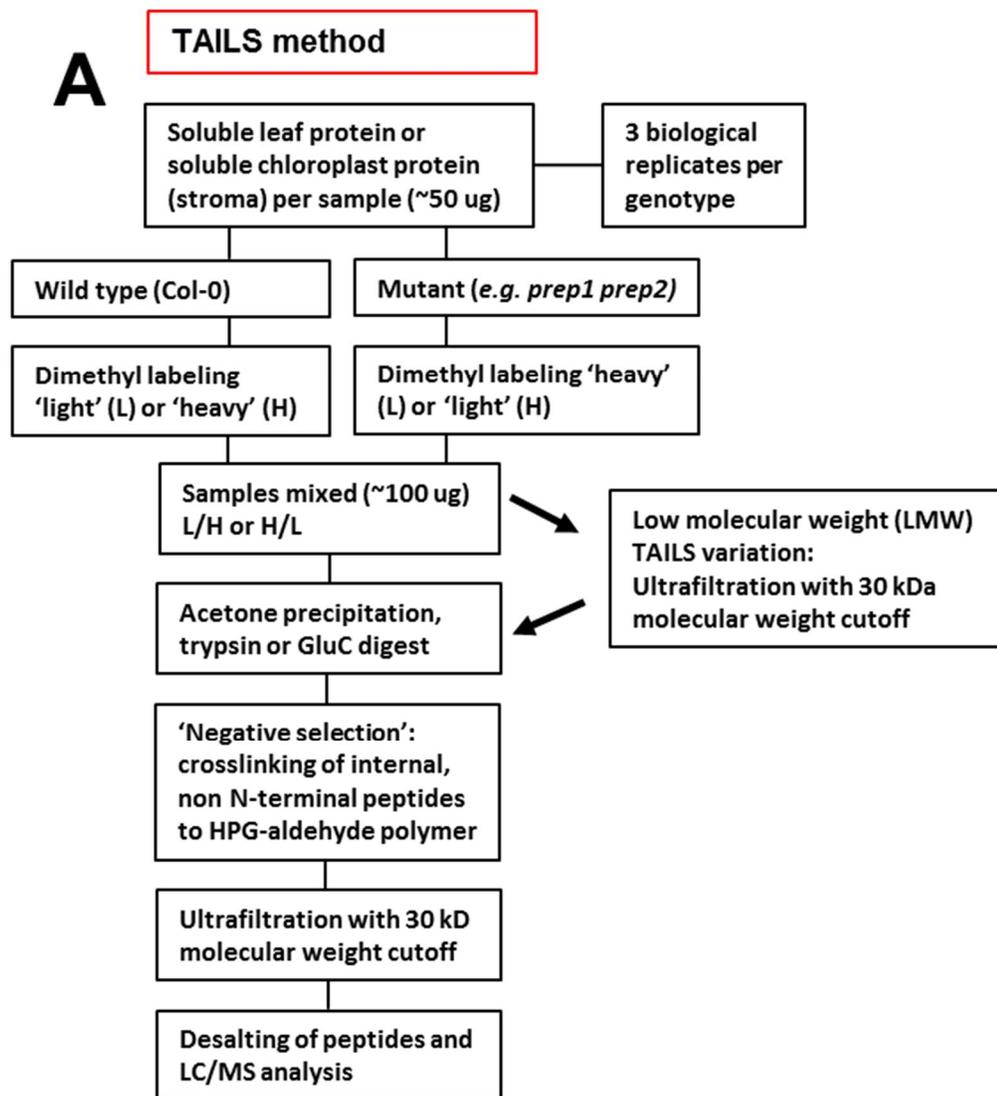
	rep1-wt	rep1-prep1 prep2	rep1-oop-2	rep1-triple	rep2-wt	rep2-prep1 prep2	rep2-oop-2	rep2-triple	rep3-wt	rep3-prep1 prep2	rep3-oop-2	rep3-triple
rep1-wt	1.000											
rep1-prep1 prep2	0.967	1.000										
rep1-oop-2	0.980	0.970	1.000									
rep1-triple	0.978	0.967	0.979	1.000								
rep2-wt	0.978	0.966	0.981	0.966	1.000							
rep2-prep1 prep2	0.946	0.973	0.934	0.944	0.956	1.000						
rep2-oop-2	0.973	0.963	0.986	0.971	0.984	0.940	1.000					
rep2-triple	0.966	0.971	0.960	0.973	0.969	0.968	0.963	1.000				
rep3-wt	0.967	0.941	0.971	0.967	0.970	0.921	0.978	0.961	1.000			
rep3-prep1 prep2	0.972	0.953	0.972	0.975	0.974	0.938	0.978	0.970	0.991	1.000		
rep3-oop-2	0.967	0.933	0.973	0.965	0.968	0.905	0.975	0.944	0.989	0.984	1.000	
rep3-triple	0.964	0.955	0.953	0.973	0.951	0.952	0.958	0.980	0.964	0.973	0.947	1.000



Supplemental Figure 3.5. Terminal amine isotopic labeling of substrates (TAILS) workflow for comparative Nt proteomics.

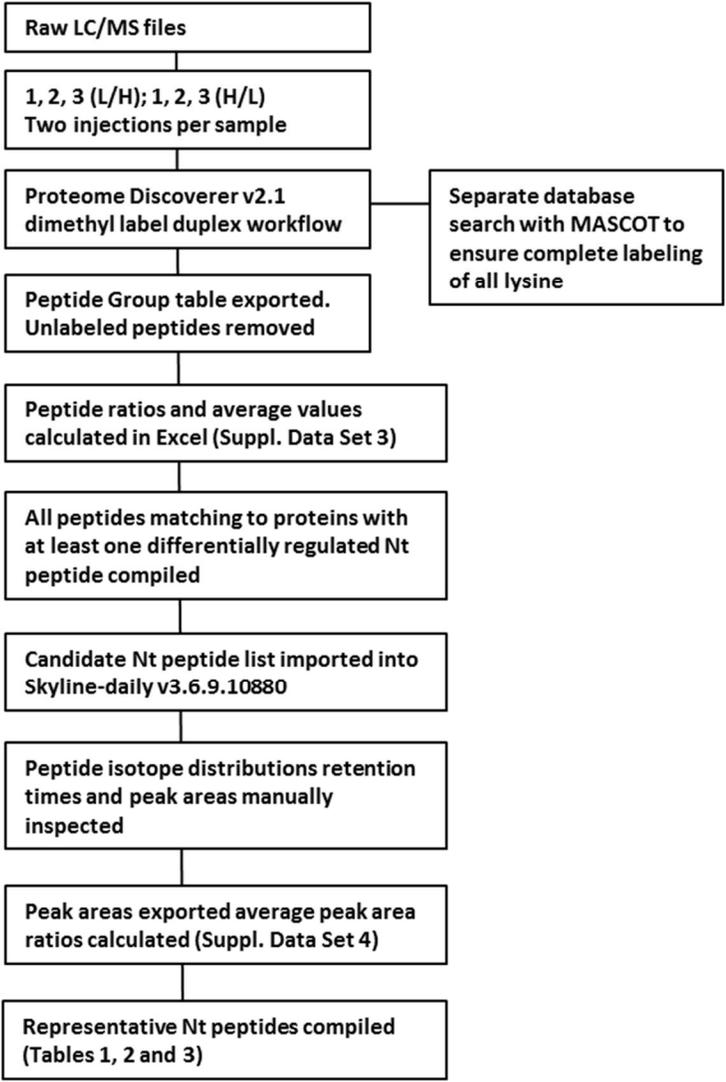
(A) Flow chart detailing TAILS method and variation thereof for enrichment of N-terminal protein sequences.

(B) Data analysis workflow for TAILS experiments.



B

TAILS Data Analysis

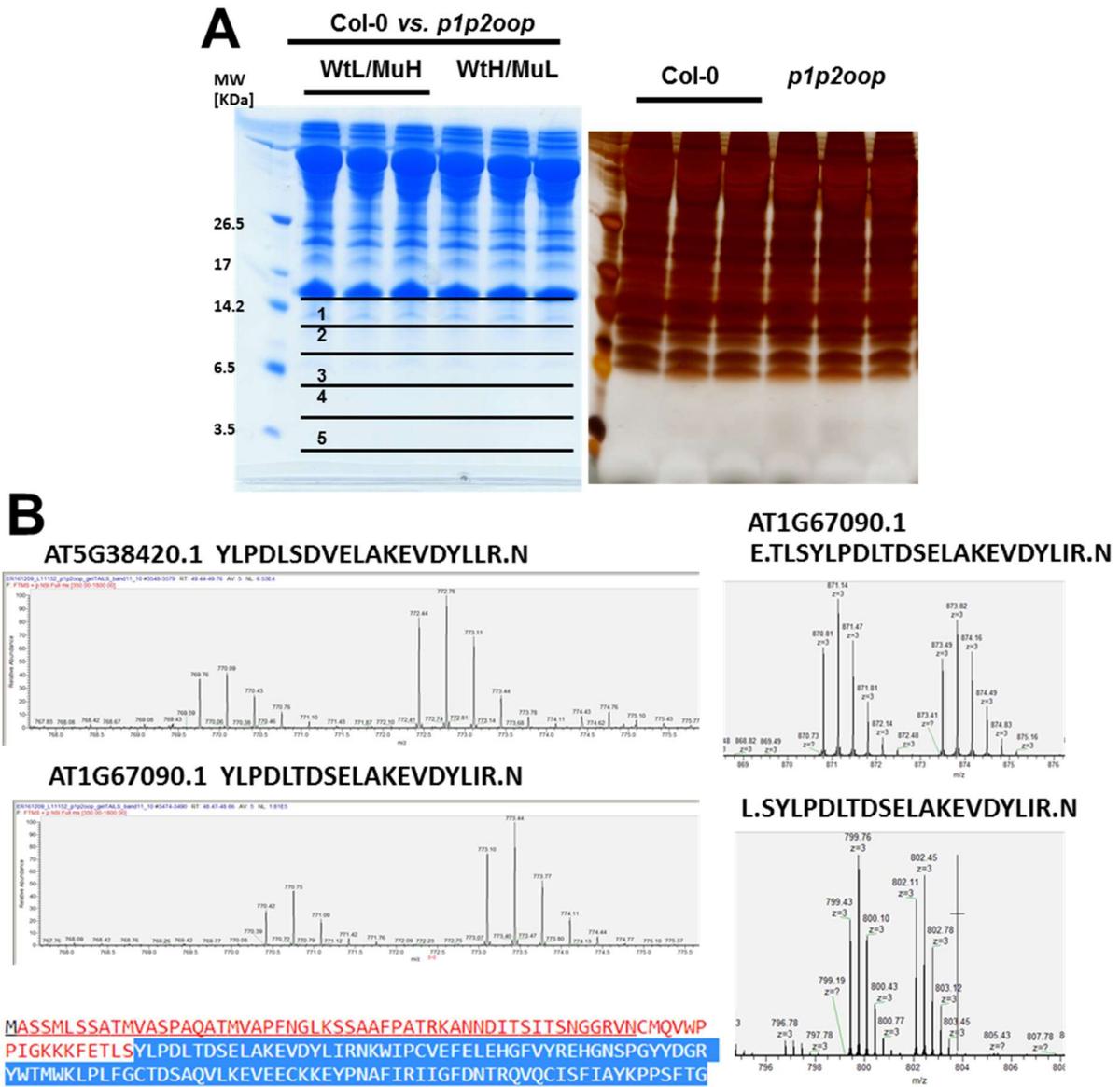


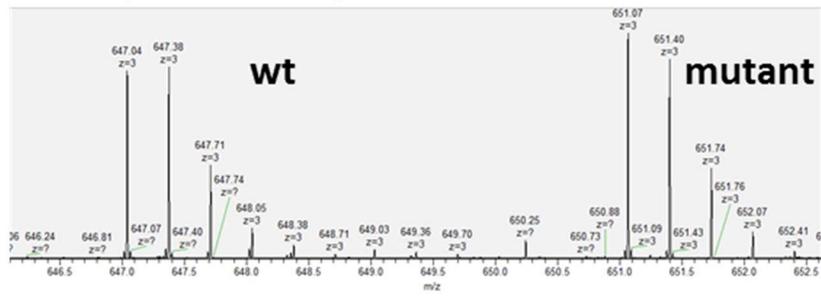
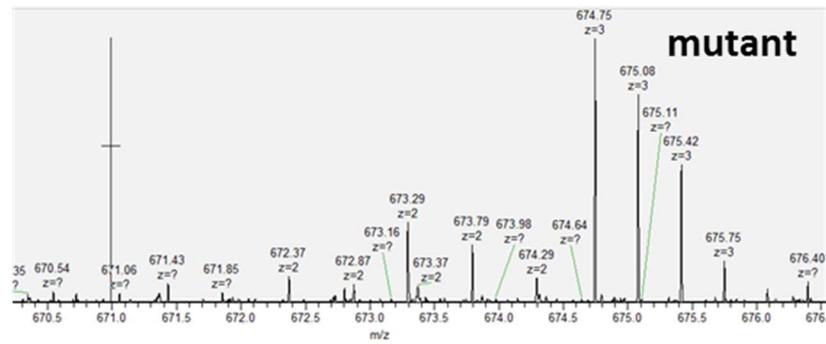
Supplemental Figure 3.6. Gel based N-terminal proteomics for *prep1 prep2 oop* stroma. Verification of relative Nt peptide abundance for select proteins.

(A) Stromal extracts dimethylated in 8M GuHCl followed by acetone precipitation, then resolved by SDS tris-tricine (16.5%). 90 ug protein per lane (mix of both genotypes labeled 'light' or 'heavy. Gel slices cut, in gel digested and analyzed by LC-MS. Second gel fixed with glutaraldehyde and stained with silver nitrate. 40 ug protein per lanes Col-0 and *p1p2oop* (lanes 1 – 6).

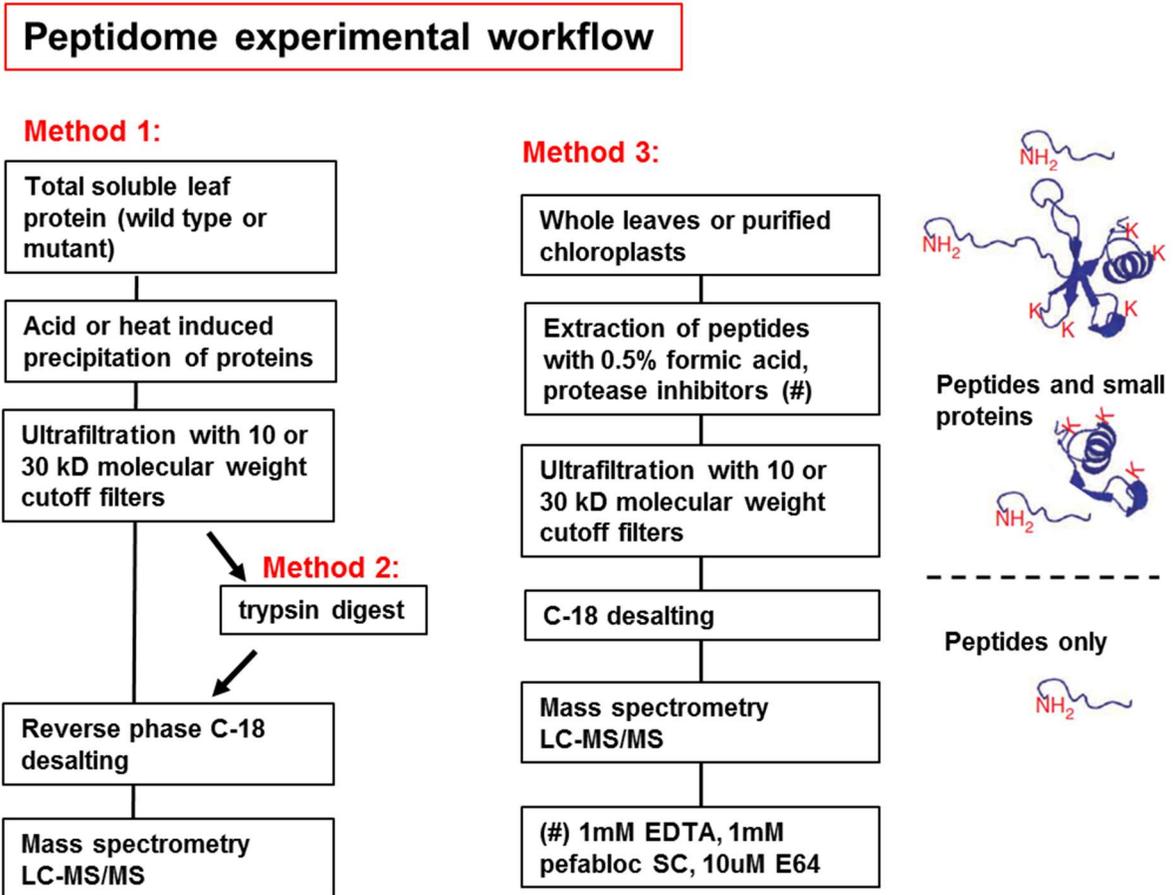
(B) Representative mass spectra for N-terminal peptides matching RBCS-4.

(C) Representative mass spectra for N-terminal peptides matching CPN10.



C**CPN10 (AT2G44650.1)****ISTKWEPTKVVVPQAR****CPN10 (AT2G44650.1)****AISTKWEPTKVVVPQAR**

Supplemental Figure 3.7. Peptidome experimental workflow. Flow chart detailing peptidome methods used to analyze *prep1 prep2* and *prep1 prep2 clpt1 clpt2* (*aabbccDd*) total leaf and chloroplast extracts.



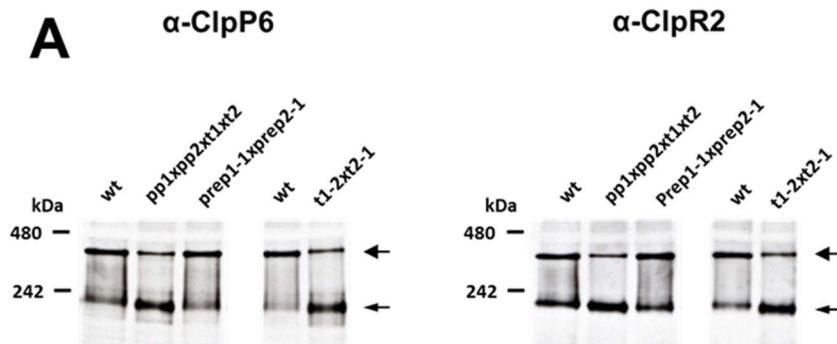
Supplemental Figure 3.8. The functional interaction between PREP1,2 and CLPT.

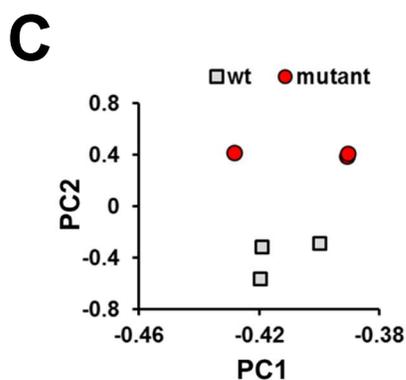
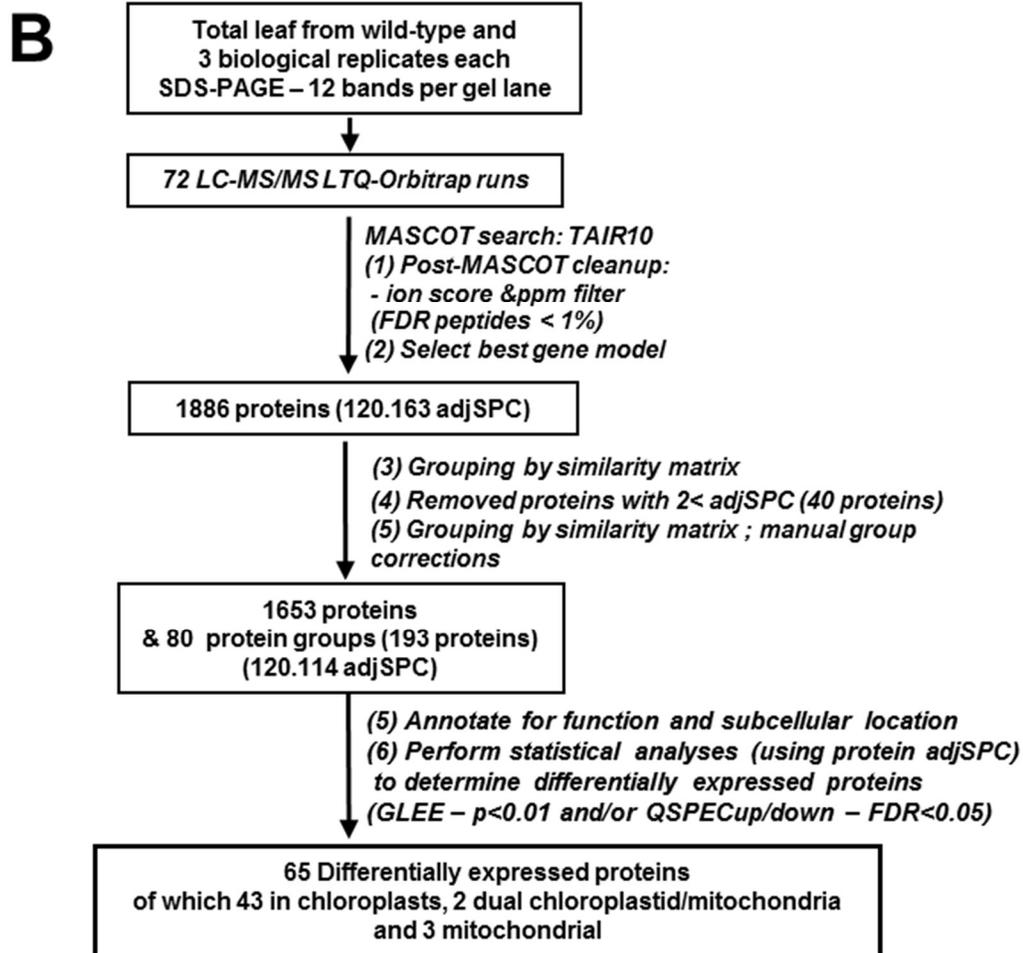
(A) Assembly state of the chloroplast CLPPR core in *wt*, *prep* and *clpt* mutants. Assembly state of the CLPPRT proteins determined by immunoblot analysis after native gel electrophoresis. Stromal proteins obtained from isolated chloroplasts from the *wt*, *prep1 prep2 clpt1 clpt2 (aabbccDd)*, *prep1 prep2*, and *clpt1 x clpt2* were separated on native gels, transferred to blots, and probed with antisera against CLPP6 and CLPR2. The larger arrows indicate CLPPRT core complexes (350–400 kD), while the smaller arrows indicate CLP complexes between 180 and 240 kD, corresponding to heptameric CLPPR rings. Thirty micrograms of stromal protein was loaded in each lane.

(B) Complete experimental and bioinformatics proteomics workflow for *prep1 prep2 clpt1 clpt2 (aabbccDd)* and *wt* plants. Total leaves were harvested at growth stage 1.11, 24 days for the *wt* and 29 days for the mutant. Plants were grown on soil under a 10-h/14-h light/dark cycle at $100 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$. Supplemental Table 3.9A,B summarizes the identified proteins, their annotations, quantification and fold changes with the statistical significance levels (p-values and false-discovery rates) of differences between genotypes.

(C) Principle component analysis (PCA) based on NadSPC of both genotypes.

(D) Correlation coefficients between genotypes and replicates based on NadjSPC





D

NadjSPC	wt-1	wt-2	wt-3	mut-1	mut-2	mut-3
wt-1	1.000	0.992	0.977	0.963	0.962	0.961
wt-2	0.992	1.000	0.969	0.959	0.957	0.957
wt-3	0.977	0.969	1.000	0.940	0.939	0.935
mut-1	0.963	0.959	0.940	1.000	0.991	0.990
mut-2	0.962	0.957	0.939	0.991	1.000	0.989
mut-3	0.961	0.957	0.935	0.990	0.989	1.000

Supplemental Table 3.2 Comparative total leaf proteome analysis of wt, *prep1 prep2*, *oop* and *prep1 prep2 oop*.

(A) Comparative proteomics of wt and mutants. All identified proteins, their annotation, spectral count data and significance analysis.

(B) Significant differentially regulated proteins between wt and mutants.

(C) Identified peptides unique to *prep prep2* and the triple mutant that verify altered N-terminal processing observed in TAILS experiments.

Supplemental Figure 3.2 C.

Accession	Protein Description	Pred . Nt	Exp. Nt	P1	Peptide	SUM SPC			Peptide Class, (Fig.3.3)	
						wt	<i>oop</i>	<i>prep</i> 1 <i>prep</i> <i>triple</i>		
AT3G27830.1	RPL12-A	59	51	R	ATHLRPIAAVEAPEKIEK			6	6	1
AT1G55490.1	Cpn60-beta-2	54	426	K	LSGGVAVIQVGAQTETELK	2		3	24	4
AT1G42970.1	GAPB	46	46	R	MSSIGGEASFFDAVAAQIIPK			1	1	3
			47	M	SSIGGEASFFDAVAAQIIPK			2	5	3
AT3G45140.1	lipoygenase LOX2	57	55	A	SRANIEQEGNTVKEPIQNIK			4	1	3
AT3G62030.1	ROC4	78	78	K	SM(ox)AAEEEEVIEPQAK	1		8	6	3
			78	K	SMAEEEEVIEPQAK			4	4	3
AT3G52960.1	PrxII E	71	63	R	SFATTPVTASISVGDKLPDSTLSYLDPSTGDVK			8	8	1
AT4G21210.1	phosphatase/kinase	87	19	S	NLNPNKSKPAGSDSVSLNASEPGSER			2	5	1 or 2
			26	K	PAGSDSVSLNASEPGSER				3	1 or 2
AT5G38430.1	RBCS-1b, RBCS-2b	55	39	K	ANNDITSITSNGGR			1	2	1
AT5G38410.1	RBCS-3B	55	42	K	DITSIASNGGR			1	1	1
AT1G67090.1	RBCS-4	55	29	K	SSAAFPATR				2	1
			53	R	VNCM(ox)QVWPPIGK			4	7	1
			53	R	VNCMQVWPPIGK				1	1
AT3G60750.1	TKL-1	66	66	R	AAAVETVEPTDSSIVDK	6	6	22	15	3
			66	R	AAAVETVEPTDSSIVDKSVNSIR	1		3	2	3
AT5G54770.1	THI1	46	46	R	ATTAGYDLNAFTDPIKESIVSR			13	14	3
AT3G54890.1	LHCI-1-1 - LHCI-730	36	26	K	FVSAGVPLPNAGNVGR				2	3
AT1G29910.1	LHCII-1.1	24	17	K	AVNLSPAASEVLGSGR			11	7	3
			20	N	LSPAASEVLGSGR			8	7	3
AT2G34430.1	LHCII-1.4	24	20	K	LSPAASEVFGTGR			4	5	1
AT2G05100.1	LHCII-2.1	31	219	R	LAMFSMFGFFVQAVTGK			4	9	4
AT5G54270.1	LHCII-3	23	23	K	ASSFNPLRDVVSLGSPK				3	3
			40	K	YTM(ox)GNDLWYGPDR			6	7	3
AT1G15820.1	LHCII-6 - CP24	47	43	K	TLIVAAAAQPK			2		3
AT5G02120.1	Ohp1 or Lil2/Hlip/Scp	41	42	A	AKLPEGVIVPK				2	3
AT1G79040.1	psbR	41	7	M	LSSVTLKPAGFTVEK				1	1 or 2
AT4G04640.1	CF1y - atpC	43	10	M	WVSSKPSLSADSSSLFR			2	1	1 or 2
AT1G20020.1	FNR-2	56	56	K	AQITTEDTPTPAK			5	3	3
AT3G04790.1	PRI	38	38	K	AQSVALSQDDLK			2	3	3
			38	K	AQSVALSQDDLK			2	1	3

Supplemental Table 3.3 List of TAILS and peptidome experiments conducted.
Table display genotypes compared to wt, type and age of tissue at harvest and method variations. (A) TAILS. (B) Peptidome

Supplemental Table 3.3 A

TAILS Experiment	Genotype compared to wild type	No. biological replicates	Method variation	Tissue used	Plant age at harvest for wild type and mutant respectively
1	<i>prep1 prep2 oop-2</i>	3		Total soluble leaf	20 and 28 days
2	<i>prep1 prep2 oop-2</i>	3*	Low molecular weight TAILS variant	Total soluble leaf	20 and 28 days
3A	<i>prep1 prep2 oop-2</i>	3		Stroma	6, 7 and 8 weeks
3B	<i>prep1 prep2 oop-2</i>	1		Stroma	4 and 6 weeks
4	<i>prep1 prep2</i>	3		Stroma	7 weeks
5A	<i>prep1 prep2</i>	3*	High molecular weight TAILS variant digested with trypsin	Stroma	7 weeks
5B	<i>prep1 prep2</i>	3*	High molecular weight TAILS variant digested with GluC	Stroma	7 weeks
6	<i>prep1 prep2 clpt1 clpt2 (aabbccDd)</i>	3		Stroma	7 and 9 weeks
7	<i>clpt1 clpt2</i>	3		Stroma	7 and 8 weeks
8	<i>prep1 prep2 oop-2</i>	3*	Low molecular weight gel, whole protein dimethyl labeling	Stroma	6, 7 and 8 weeks

* no label swap

Supplemental Table 3.3 B

Peptideome Experiment	Sample	No. biological replicates	Plant age at harvest	Peptidome method
1	wt	3	26 Days	3
	<i>prep1 prep2</i>	3	26 Days	3
2	wt	2	26 Days	3
	<i>prep1 prep2</i>	2	26 Days	3
3a	wt	3	3 weeks	1, 2
3a	<i>prep1 prep2</i>	1	3 weeks	1, 2
3b	<i>prep1 prep2 clpt1 clpt2 (aabbccDd)</i>	3	4 weeks	1, 2
4*	wt	2	7 weeks	3
	<i>prep1 prep2</i>	2	7 weeks	3

* Experiment using purified chloroplast extracts. (1) Based on Lyons and Fricker, 30 kDa molecular weight (MW) cutoff filter. (2) As per method 1 except with trypsin digest. (3) Based on Xu *et al.* 10 or 30 kDa MW cutoff filters.

Supplemental Table 3.6. Mitochondrial protein N-termini, processed by ICP55.
 N-terminal sequences identified in wt and prep1 prep2 oop total leaf extracts,
 relative abundances determined by TAILS. Peptide peak areas were calculated in
 Skyline.

Supplemental Table 3.6

Acession	N-terminal peptide	AVG 6 reps (MU/WT)	STDEV
AT4G32915.1	SYSSDS DSSVLQPPDVAR	0.8	0.1
AT1G48030.1	ASSGSDENDVVIIGGGPGGYVAAIKASQLGLKTTCKIEKR	0.8	0.1
AT3G06050.1	SKLAEGTDITSAAPGVSLQKAR	1.1	0.1
AT3G13930.1	SSSDLPPHQEIGMPSLSPMTTEGNIAR	1.1	0.3
AT5G08670.1	ATSSPASSAAPSSAPAKDEGKKTYDYGGKGAIGR	1.6	0.6
AT5G55070.1	SSDSGDVVEAVVPHMGESITDGTAAFLKPGDR	1.0	0.1
AT5G23140.1	YSLIPMVEIHSSR	0.9	0.1
AT5G23140.1	SLIPMVEIHSSR	1.0	0.4

CHAPTER 4

PLASTOGLOBULE LOCALIZED METALLOPEPTIDASE PGM48 IS A POSITIVE REGULATOR OF SENESCENCE IN *ARABIDOPSIS THALIANA*¹

4.1 ABSTRACT

Plastoglobuli (PG) are thylakoid-associated monolayer lipid particles with a specific proteome of ~30 PG core proteins and isoprenoid- and neutral lipids. During senescence, PG increase in size, reflecting their role in dismantling the thylakoid membrane. Here we show that the only PG-localized peptidase, PGM48, acts as a positive regulator of leaf senescence. We discovered that PGM48 is a member of the M48 peptidase family with PGM48 homologs forming a novel clade (M48D) only found in photosynthetic organisms. Unlike the M48A, B, C clades, members of M48D have no transmembrane domains, consistent with the unique subcellular location, the PG. *In vitro* assays showed Zn-dependent proteolytic activity and substrate cleavage upstream of hydrophobic residues. Overexpression of PGM48 accelerated natural leaf senescence, whereas suppression delayed senescence. Quantitative proteomics of PG from senescing rosettes of PGM48 overexpression lines showed a dramatically reduced level of CAROTENOID CLEAVAGE ENZYME 4 (CCD4), and significantly increased levels of senescence-induced ABC1 kinase 7 (ABC1K7) and PHYTYL ESTER SYNTHASE 1 (PES1). Yeast-2-hybrid experiments identified PG core proteins ABC1 KINASE 3, PES1 and CCD4 as PGM48 interactors, whereas several other PG-localized proteins and chlorophyll degradation enzymes did not interact. We

¹ **AUTHOR CONTRIBUTIONS** NHB, GF, ER designed and performed the experimental analysis, KM carried out co-expression analyses. GF and ER carried out the mass spectrometry analyses. KJVW provided oversight over the project, helped design the various and wrote the paper together with NHB.

discuss mechanisms through which PGM48 could possibly accelerate the senescence process.

4.2 INTRODUCTION

Plastoglobuli (PG) are lipoprotein particles found in plastids of most plant tissues. In chloroplasts, PG are contiguous with the outer lipid leaflet of the thylakoid membrane which is compatible with channeling of hydrophobic metabolites between the thylakoid membrane and the PG (Austin et al, 2006). Chloroplast-localized PG size typically ranges from 45 to 60 nm, but increase dramatically in size in senescing or stressed chloroplasts (*e.g.* light stress, nitrogen deprivation) of wild-type (wt) plants and in various chloroplast mutants (Austin et al., 2006; Lundquist et al., 2013). The dynamic response of PG suggest that they play a functional role in chloroplast biogenesis, senescence and stress response (Rottet et al., 2015).

PG contain various prenyl-lipids, in particular tocopherol, various quinones (plastoquinone 8, phylloquinone, plastoquinone), carotenoids and fatty-acid phytylesters, as well as triacylglycerols (TAGs) (Gaude et al., 2007; Zbierzak et al., 2009; Eugeni-Piller et al., 2011; Lippold et al., 2012). Mass spectrometry (MS) analysis of isolated PG proteins and quantitative comparison to the thylakoid, stromal and total leaf proteomes showed that PG contain a small but specific proteome of a ~30 proteins, assigned the PG core proteome, (Vidi et al., 2006; Ytterberg et al., 2006; Lundquist et al., 2012) and several additional proteins that are recruited to PG under stress conditions and in mutant backgrounds (Lundquist et al., 2013). The most abundant PG proteins are several members of the plastid-specific FIBRILLIN family (FBN) (Singh and McNellis, 2011) and members of the ACTIVITY OF *BC₁* COMPLEX KINASE (ABC1K) family (Lundquist et al., 2012). Other PG core proteins include a well-studied tocopherol (vitamin E) cyclase (VTE1) (Porfirova et al., 2002), a key enzyme in tocopherol and plastochromanol biosynthesis, and PHYTYL ESTER SYNTHASE 1 and 2 (PES1,2) involved in the formation of phytylesters following cleavage of chlorophyll and thylakoid lipids (Lippold et al.,

2012). PG-localized NADP(H) dehydrogenase C1 (NDC1) is involved in vitamin K biosynthesis (phyllo-quinone) (Fatihi et al., 2015) and in reduction of oxidized plastochromanols within PG (Eugeni-Piller et al., 2011). PG core protein CCD4 and its homologs in various plant species (Martinez et al., 2008) are involved in carotenoid cleavage, in particular 9,10 and 9', 10' positions, resulting in one or more apocarotenoids, especially β -ionone (Rubio et al., 2008; Huang et al., 2009). An *Arabidopsis thaliana* (further referred to as *Arabidopsis*) genome-wide association study identified CCD4 as a major negative regulator of seed carotenoid content (Gonzalez-Jorge et al., 2013); *ccd4* loss-of-function mutants exhibited increased β -carotene content upon seed desiccation and much higher carotenoid levels than the wild-type after dark-induced leaf senescence. *Arabidopsis* CCD4 is also implicated in an apocarotenoid signaling cascade leading to inhibition of chloroplast and leaf development in *Arabidopsis* (Avendano-Vazquez et al., 2014; Hou et al., 2016). White color petals of *Chrysanthemum* were converted into yellow color petals by RNAi-mediated suppression of the CCD4 homolog *CCD4a*, indicating its cleavage of carotenoids into colorless compounds (Zhu et al., 2010). In mandarin fruit, a CCD4 homolog was reported to be involved in cleavage of β -cryptoxanthin and zeaxanthin to yield a red pigment, β -citraurin (Rubio et al., 2008). Collectively this shows that the CCD4 clade cleaves a variety of carotenoids into various apocarotenoids and that CCD4 in leaves plays a role in leaf development and retrograde signaling; however the signaling molecules and pathway are not known (Hou et al., 2016). Other PG core proteins have various predicted functional domains, such as two UBI-E methyltransferases (UBIE-MT1,2) (related to *E. coli* UbiE involved in methylation of reactions in both ubiquinone and menaquinone biosynthesis (Lee et al., 1997)), but their functions have not yet been studied. We also identified a low abundant putative peptidase of the M48 family, here assigned PGM48. The functions and targets of PGM48 are unknown and PGM48 is the focus of this paper.

Based on publicly available *Arabidopsis* genome-wide mRNA data, we previously generated a co-expression network using the PG core proteins as nodes (Lundquist et al., 2012). This suggested four major co-expression modules, each

showing functional enrichment, including senescence, chloroplast proteolysis, carotenoid metabolism, chloroplast redox regulation, the Calvin cycle and chloroplast biogenesis. The senescence module was formed around PG-localized ABC1K7, PES1, PGM48, PG-ASSOCIATED SENESCENCE GENE (PGSAG; AT1G73750) and included the co-expressors PHEOPHYTENASE (PPH) and PHEIDE a OXYGENASE (PaO) both involved in chlorophyll degradation. Leaf senescence is a highly regulated process, involving expression of thousands of genes, and resulting in chlorophyll degradation, loss of photosynthetic activity and remobilization of the chloroplast constituents for seed development or other tissues (Breeze et al., 2011; Schippers, 2015; Schippers et al., 2015). During senescence, internal structures within chloroplasts are dismantled and the PG increase in size during this process, suggesting their direct involvement in senescence. Comparative, quantitative metabolite analysis of leaf PG isolated from four stages of natural senescence in beach leaves (*Fagus sylvaticus*) showed that in particular, prenylquinones and free fatty acids, but not glycolipids or proteins, accumulated in PG during the senescence process (Tevini and Steinmuller, 1985). Whereas only very small amounts of chlorophylls and carotenoids or lipids were observed in PG prior to senescence, during natural senescence, the TAG content dramatically declined in PG, whereas carotenoids (mostly in esterified form) and free FA levels strongly increased (Tevini and Steinmuller, 1985). However, the removal of protein complexes must also involve protein degradation, but despite major efforts by labs around the world, no intra-plastid peptidase has been shown to be involved in leaf senescence (Martinez et al., 2008; van Wijk, 2015)).

Here we studied the function of PGM48 both *in vitro* and *in vivo* in *Arabidopsis*. This shows that PGM48 is a M48-type metallopeptidase with a conserved HExxH metal binding site. Phylogeny indicates that PGM48 forms a specific clade of peptidases only found in photosynthetic organisms; we named this subfamily M48D. We discuss the likely evolutionary origin of this subfamily and compare it to M48A, M48C, M48B subfamilies. Importantly, we discovered that overexpression of PGM48 accelerated natural leaf senescence whereas suppression of PGM48 by RNAi delayed senescence. Quantitative proteomics of PG from senescing

rosettes of overexpression lines showed a dramatically reduced level of carotenoid cleavage enzyme 4 (CCD4), and significantly increased levels of ABC1 kinase 7 (ABC1K7) and PHYTYL ESTER SYNTHASE 1 (PES1). Furthermore, we show that PGM48 physically interacts with PES1, CCD4 and ABC1K3 but not with chlorophyll catabolic enzymes. We discuss how PG-localized PGM48 can possibly contribute to leaf senescence. This is the first intra-plastid protease directly linked to the leaf senescence process.

4.3 RESULTS

4.3.1 *PGM48 is a low abundant PG-localized protein in Arabidopsis*

PGM48 was identified by tandem mass spectrometry (MS/MS) in purified PG from leaves using a LTQ-Orbitrap mass spectrometer in *Arabidopsis* (Lundquist et al., 2012; Lundquist et al., 2013), as well as maize (Huang et al., 2013). PGM48 was never observed in any other samples than purified PG, despite hundreds of MS/MS analyses of *Arabidopsis* or maize leaf, chloroplast or thylakoid fractions carried out in our lab over the last decade (see PPDB) or by others for isolated chloroplast envelope membranes of *Arabidopsis* (Ferro et al., 2010) or other plant species such as *Pisum sativum*, *Medicago sativa* or *Zea mays* (Manandhar-Shrestha et al., 2013; Simm et al., 2013; Gutierrez-Carbonell et al., 2014). Furthermore, MS/MS analysis of many PG preparations (Lundquist et al., 2012; Lundquist et al., 2013) and this paper, resulted in high sequence coverage of the mature portion of PGM48. No peptides were detected matching against the predicted chloroplast transit peptide (cTP) (Supplemental Figure 4.1), which is consistent with a plastid location of PGM48; the most N-terminal residue was R61, immediately down-stream of a cysteine, as frequently observed for chloroplast proteins (Rowland et al., 2015). To further verify localization across chloroplast compartments, we purified thylakoid, stroma and PG proteomes from isolated chloroplasts and probed distribution with specific antisera for stroma (CPN60), thylakoids (LHCB1) and PG core proteins ABC1K1 and ABC1K3.

Consistent with our proteomics data, PGM48 signal was highly enriched in PG similar as ABC1K1, ABC1K3 and VTE1 and unlike CPN60 or LHCB1 (Figure 2.1A). We conclude that PGM48 is highly enriched in PG and likely carries out its peptidase function in this specialized compartment.

4.3.2 *PGM48 is a functional metallo Zn-endopeptidase*

To test if PGM48 is a functional peptidase, we expressed mature (*i.e.* starting from R61; excluding the cTP) Arabidopsis PGM48 with a histidine tag in *E. coli* and the recombinant protein was affinity purified. PGM48 was incubated with β -casein as substrate and degradation was visualized by Coomassie staining after separation by SDS-PAGE (Figure 2.1B). β -casein was degraded by PGM48 and this activity was completely inhibited by addition of the metal chelator EDTA (Figure 2.1B). Other peptidase inhibitors including the cysteine peptidase inhibitor E64 and metallo-aminopeptidase inhibitor bestatin partially inhibited degradation, whereas chymostatin (an inhibitor of various types of peptidases) did not inhibit PGM48 activity. Through site-directed mutagenesis of the conserved metal binding HEXXH in PGM48 (see Figure 2.1C - lower panel with the sequence logo based on angiosperm PGM48 homologs), we replaced H191 or E192 into A191 or D192, respectively. Recombinant mutated PGM48 (PGM48-H191A and PGM48-E192D) completely lacked peptidase activity (Figure 2.1C – upper panel), clearly showing that PGM48 is a functional metallo-endopeptidase.

To determine cleavage site specificity, recombinant active PGM48 and inactive PGM48-H191A were each incubated with β -casein, and following desalting, peptide products were analyzed by high resolution MS through direct infusion. Figure 1D shows an MS spectrum from extracts of the active PGM48. The dominant peaks matched to eight peptides as determined by MS/MS ranging from 1.7 to 5.0 kDa (Figure 2.1D); these peptides were absent in PGM48-H191A (Supplemental Figure 2.2). The methionine containing peptides were also observed with oxidized methionine residues and/or in different charge states (due to the ionization process in the MS source) allowing us to account for almost all ions in the mass spectrum (see

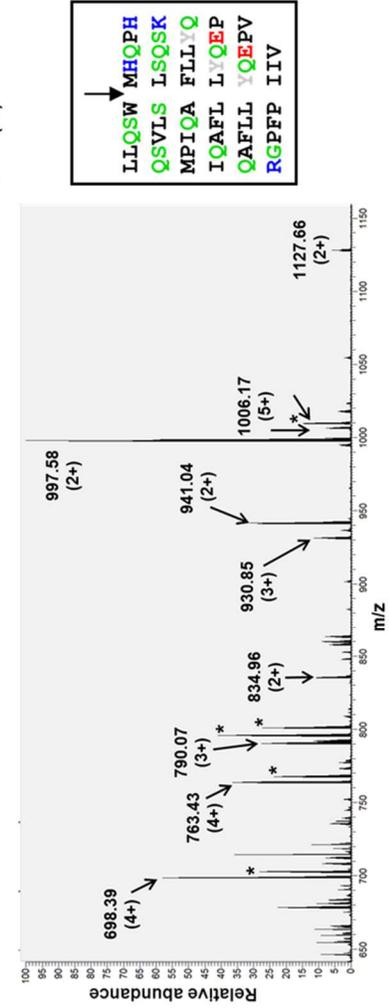
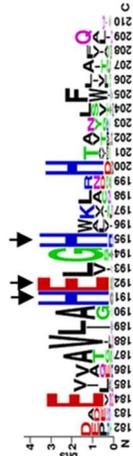
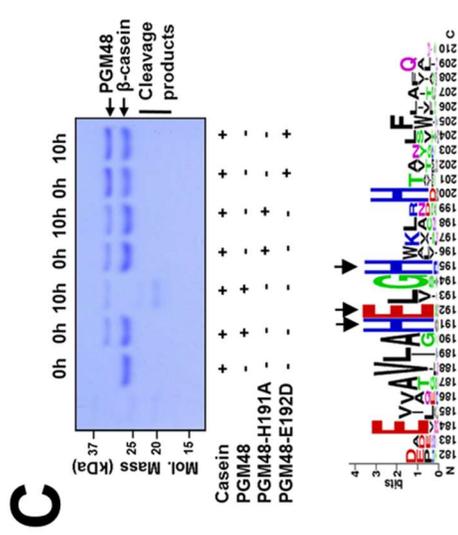
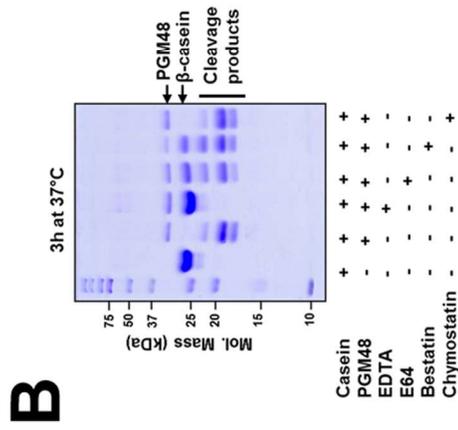
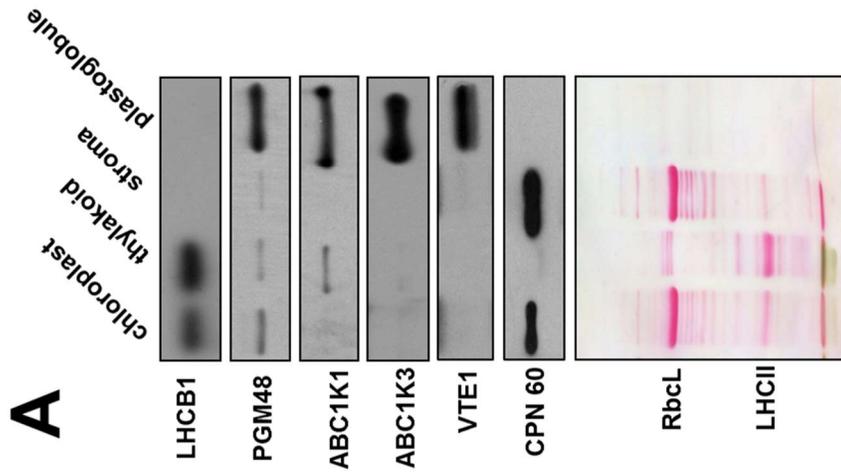
Figure 4.1. PG-M48 is localized in PG and has Zn-metallo-dependent endopeptidase activity cleaving upstream of hydrophobic residues.

(A) Proteomes from isolated chloroplasts, thylakoid membranes, stroma and PG were separated by SDS-PAGE, transferred to membranes and blotted with specific antisera against PG core proteins PGM48, ABC1K1, ABC1K3, VTE1, as well as LHCB1 (marker for thylakoid membrane) and CPN60 (marker for stroma). 15 μ gram protein was loaded for chloroplast, stroma and thylakoid fractions and \sim 5 μ gram protein was loaded for PG – due to the extremely high lipid/protein content, protein concentrations are difficult to measure accurately. The Ponceau stained membrane (lower panel) serves as loaded control. RBCL – Rubisco Large Subunit (also marker for stromal proteome); LHCII - family of major LHCII proteins.

(B) Recombinant PGM48 (1 μ g) was incubated for 3 hrs at 37°C with β -casein (4 μ g) to determine the peptidase activity and the effects of peptidase inhibitors E64 (1 mM), Bestatin (1 mM), Chymostatin (1 mM) and the metal chelator EDTA (5 mM). EDTA inhibited PGM48 peptidase activity completely.

(C) Metal dependency of PGM48. Upper panel shows in vitro proteolytic activity of recombinant PGM48 and PGM48-H191A and PGM48-E192D mutated in their predicted metal binding site (HExxH). PGM48, but not mutated PGM48, shows peptidase activity. The lower panel shows a sequence logo for the conserved metal binding motif of eleven PGM48 homologs in angiosperms (for a complete list of homologs see Supplemental Table 1). Color coding: Green: T, G, S, Y, C – uncharged, polar; Pink: N, Q – uncharged polar amine; Red: D, E – Acidic; Blue: K, R, H – basic; Black: A, V, L, I, P, F, W, M – apolar.

(D) PGM48 cleavage sites of β -casein as determined by mass spectrometry. The lower panel shows a high resolution MS spectrum (from the Orbitrap portion of the LTQ-Orbitrap) of a casein digest by PGM48. Positive charged peptides matching to casein peptide fragments are marked; m/z values and charge states are indicated. The upper portion shows the C-terminal portion of β -casein and the identified peptides as indicated by horizontal lines with the respective m/z value and associated charge states. The boxed inset shows the inferred cleavage sites with the β -casein sequence. Collectively, this shows that PGM48 prefers cleavage upstream of hydrophobic residues (methionine (M), phenylalanine (F), leucine (L), isoleucine (I) (in black) or Tyrosine (Y) (in grey). Other colors: Green: Q, S – uncharged, polar; D, E – acidic in red; K, R, H – basic in blue; remaining residues in black.



Supplemental Table 2.1). The detected fragments correspond to cleavage sites depicted in the sequence of PGM48, and displayed in the Figure 2.1D insert, with a clear preference for cleavage immediately upstream of hydrophobic residues (L, I, M, F or Y) and downstream of hydrophobic (L, W, P) or polar residues (A,S).

4.3.3 PGM48 homologs form a new clade in the M48 family

PGM48 belongs to the M48 family of ATP-independent metallo (Zn)-endopeptidases as defined in the MEROPS peptidase database (Rawlings et al., 2016). According to MEROPS, the M48 family has 3 clades, M48A, M48B and M48C, each represented by a prototypic enzyme named STE24 or CAAX peptidase, bacterial HTPX and mitochondrial OMA1, respectively (see INTRODUCTION for references). However, sequence analysis of M48 proteins in 42 species across the tree-of-life, suggests that there are four clades (Figure 4.2; Supplemental Table 4.2; Supplemental Dataset 4.1 for alignments), with PGM48 and its homologs forming a separate clade that we assigned M48D. We briefly discuss the distribution of these M48 clades across phylogenetic groups and comment on known functions, followed by a hypothesis for the origin of PGM48 proteins in plants.

Clade M48A consists of well-studied peptidases localized to the endoplasmic reticulum (ER) or inner nuclear membrane in photosynthetic and non-photosynthetic eukaryotes, and STE24 homologs in non-photosynthetic bacteria, but absent in Archaea and photosynthetic bacteria. Eukaryotic STE24 are membrane proteins (multiple transmembrane domains) with a large hollow barrel-shaped chamber enclosing the catalytic site. They function to cleave C-termini of proteins with a CAAX motif carrying a cysteine linked prenyl group. Recently, STE24 has also been shown to clear clogged SEC protein translocons, in addition to its function in CAAX processing (Ast et al., 2016). Angiosperms, gymnosperms, lycopod, moss and most red/green algae each have a single STE24 homolog (Figure 4.2, Supplemental Table 4.2). Arabidopsis STE24 is ER-localized and cleaves prenylated CAAX proteins (Bracha et al., 2002; Bracha-Drori et al., 2008).

Figure 4.2. M48 proteins are conserved in non-photosynthetic and photosynthetic organisms.

Phylogenetic tree of the M48 family based on alignment of amino acid sequences of 92 M48 homologs from diverse archaea, prokaryotes and eukaryotes. Four clades are assigned and are M48A, M48B, M48C and the new clade M48D and bootstrap values are indicated at the nodes of the tree. The sequence alignment is shown in the Supplemental Data File and Supplemental Table 1 provides a complete listing of species and number of M48 members. Abbreviations: *Vitis vinifera* (Vvi); *Populus trichocarpa* (Ptr); *Medicago truncatula* (Mtr); *Brassica rapa* (Bra); *Arabidopsis thaliana* (Ath); *Glycine max* (Gma); *Zea mays* (Zma); *Brachypodium distachyon* (Bdi); *Hordeum vulgare* (Hvu); *Oryza sativa* (Osa); *Sorghum bicolor* (Sbi); *Picea sitchensis* (Psi); *Selaginella moellendorffii* (Smo); *Physcomitrella patens* (Ppa); *Cyanidioschyzon merolae* (Cme); *Chondrus crispus* (Ccr); *Volvox carteri* (Vca); *Chlamydomonas reinhardtii* (Cre); *Chlorella variabilis* (Cva); *Micromonas pusilla* (Mpu); *Synechocystis* sp. PCC6803 (Syne); *Anabaena cylindrica* (Acy); *Nostoc* sp. PCC7120 (Nos); *Synechococcus* sp. PCC 7942 (Syn); *Homo sapiens* (Hsa); *Penicillium marneffei* (Pma); *Saccharomyces cerevisiae* (Sce); *Agrobacterium tumefaciens* (Atu); *Nitrosomonas europaea* (Neu); *Escherichia coli* (Eco); *Sulfuricurvum kujiens* (Sku); *Streptomyces griseus* (Sgr); *Bacillus subtilis* (Bsu); *Leptospira interrogans* (Lin); *Leptospira biflexa* (Lbi); *Sulfolobus islandicus* (Sis); *Metallosphaera*

Clade M48B is represented by bacterial HTPX Zn-endopeptidase. Members of this clade are present in Archaea and in non-photosynthetic and photosynthetic bacteria, but not in eukaryotes (Figure 4.2). Most prokaryotes have one HTPX homolog (Supplemental Table 4.2). HTPX homologs have been studied in *e.g.* *B. subtilis* and *E. coli* and have shown to be integral plasma membrane metallopeptidases with their active side facing the cytosol. HTPX in *E. coli* complements the ATP-dependent FTSH peptidase in quality control of the plasma membrane proteome (Akiyama, 2009).

Clade M48C is represented by inner membrane mitochondrial OMA1. This clade is found in both photosynthetic and non-photosynthetic eukaryotes and non-photosynthetic prokaryotes, but not in photosynthetic bacteria (Figure 4.2). OMA1 homologs in eukaryotes are (most likely) localized in mitochondria; they have been studied in mammals, zebrafish and yeast, in particular in proteolysis of OPTIC ATROPHY 1 (OPA1/MGM1), a key conserved inner membrane dynamin-like GTPase involved on mitochondrial fusion (Bohovyck et al., 2015; Korwitz et al., 2016). OMA1 is present in angiosperms, most green and red algae, lycophytes, moss and angiosperms, but not in the gymnosperm *Picea sitchensis*, either due to independent gene loss or perhaps incomplete genome sequence annotation (Figure 4.2; Supplemental Table 4.2). The OMA1 homolog in Arabidopsis has been observed in several mitochondrial proteomics studies (Finkemeier et al., 2011; Klodmann et al., 2011) and homologs in maize, rice, tobacco and other plant species have predicted mitochondrial transit peptides. However, it appears that OMA1 function has not been studied in any photosynthetic eukaryotes, nor in prokaryotes.

Based on the cladogram (Figure 4.2), it is clear that PG-localized M48 homologs form a separate clade; we assigned this as a new clade M48D. Proteins in this clade are only found in photosynthetic organisms, including cyanobacteria, green algae, red algae, gymnosperms (conifers) and angiosperms (Figure 4.2 and Supplemental Figure 4.3 for the M48D clade alone), but absent in the moss *Physcomitrella patens* and the lycophyte *Selaginella moellendorffii*. None of the members of this new M48 clade have been studied so far, but in addition to

Arabidopsis PGM48, we did identify the maize PGM48 protein (GRMZM2G111200) in isolated PG from leaves (Huang et al., 2013). We also identified the rice PGM48 (Os01g73910) in isolated chloroplasts with good sequence coverage (in three independent replicates) but not in total leaf extracts (see PPDB). PGM48 in maize, rice and most other angiosperms have a TargetP predicted cTP (Supplemental Figure 4.1).

It should be noted that *Arabidopsis* and other photosynthetic organism, have another peptidase family of CAAX proteases, named M79, that typically cleave a C-terminal tripeptide from an isoprenylated protein; there are five Arabidopsis proteins in this M79 family. One M79 family member in Arabidopsis is plastid-localized SNOWY COTYLEDON 4 (SCO4; AT5G60750), but there is no evidence that it has CAAX activity (Albrecht-Borth et al., 2013).

4.3.4 Structural model of PG48 and interaction with the monolayer PG

It is unknown how PGM48 interacts with PG, how substrates are recognized and interact with PGM48, or how substrates obtain access to the catalytic site. Fortunately, high resolution X-ray crystallography-based structures for M48 homologs in clade M48A from human (PDB 4AW6) (Quigley et al., 2013) and *Saccharomyces mikatae* (PDB 4IL3) (Pryor et al., 2013) (37% identity between these two STE24 homologs), clade M48B for *Vibrio parahaemolyticus* (PDB 3CQB) and from clade M48C for *Geobacter sulfurreducens* (PDB 3C37) are available in the Protein Data Bank (<http://www.rcsb.org/pdb/>). Based on these available structures we built a homology model for mature Arabidopsis PGM48 (Figure 4.3) with the objective to begin answering the key questions stated above. The various scoring parameters of the top scoring i-TASSER model for PGM48 (see legend Figure 4.3) suggested a meaningful model that is suitable and sufficient to address general folding, formation of the active site, internal cavity and interaction with the PG. Lower scoring models were not further considered. The PGM48 model was most similar to the 3.8Å structure of human STE24. Structural analysis showed that STE24 has a seven transmembrane (TM) α -helical barrel structure surrounding a large, water-filled chamber, capped by the Zn metallopeptidase domain with the catalytic site facing into the chamber

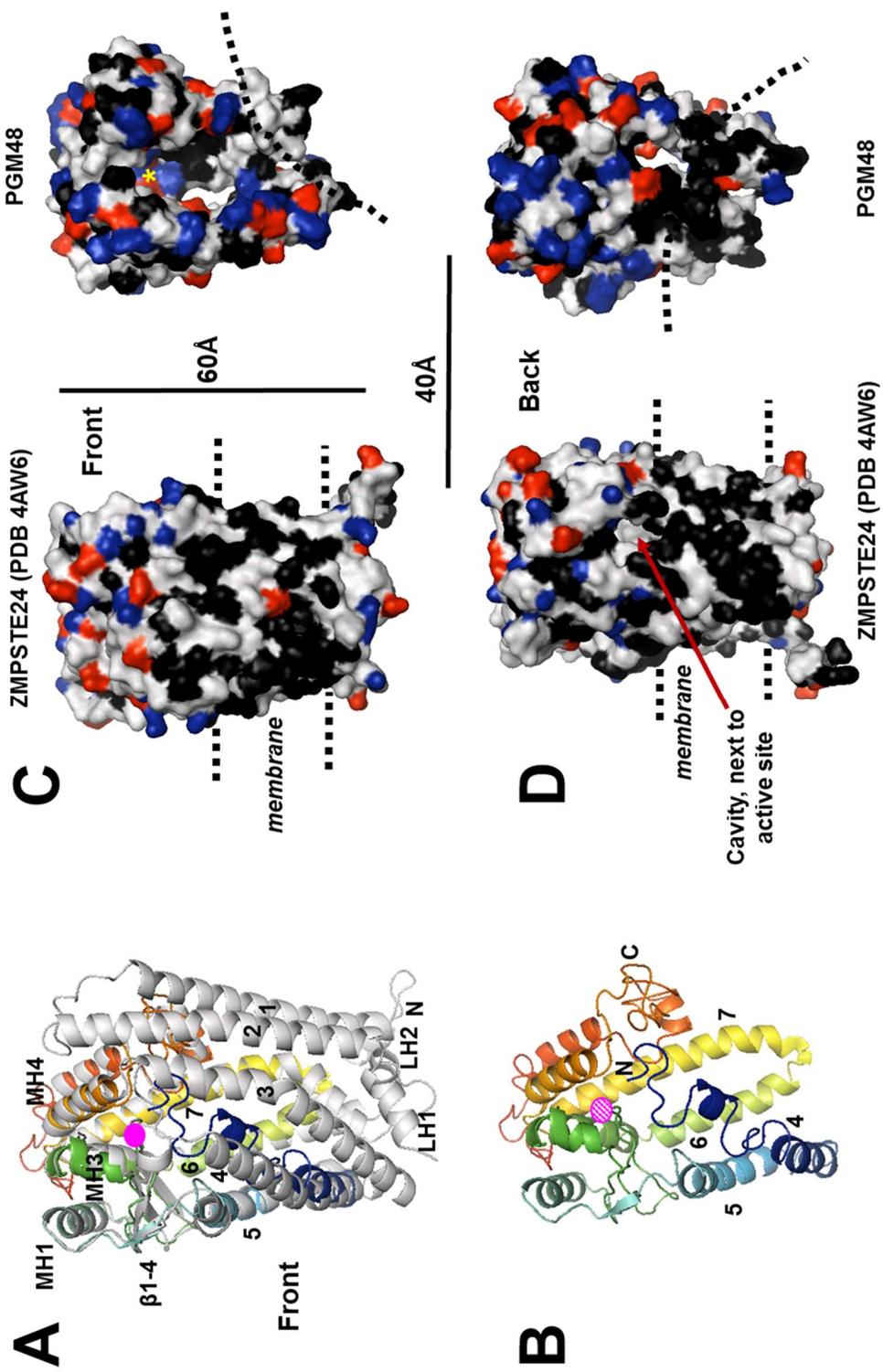
Figure 4.3. 3D protein structural model for PGM48 and predicted interface with the PG monolayer.

The 3D protein structural model of PGM48 was generated with i-TASSER (Yang, 2015) using the predicted mature PGM48 sequence (residues 48 - 344). The top scoring i-TASSER model for PGM48 had a C-score of -0.98 (C-scores range from -5 (poorest) to 2(best)), and estimated TM-score of 0.59 ± 0.14 . Analysis of PGM48 protein model using ProSA (Wiederstein, 2007) gave a Z-score of -3.73, which is well within the range of scores found for protein structures of this molecular weight generated by crystallography and NMR. The protein model matched most closely to the 3.8Å structure PDB 4AW6 of the M48 protein STE24 from humans (Quigley, 2013). Images were generated with PyMol Version 1.7.4 software (Schrödinger, LLC).

(A) Front view of STE24 structure (4AW6) in grey overlaid with the colored modelled PGM48. PGM48 is colored in rainbow from blue N-terminus to red C-terminus. The Zn metal ion is marked in pink (not to scale). The STE24 helices 1-7, luminal helices (L1-3) and cytoplasmic helices (MH1, 3, 4) are marked using the naming from (Quigley, 2013) Helices 1, 2, 3, 7A, LH1 and LH2 are absent in the PGM48 model.

(B) Front view of the PGM48 model. Helices are numbered according to 4AW6 and the N- and C-termini are indicated; the position of the Zn ion is indicated in dashed pink.

(C, D) Side-by-side comparison of the structure 4AW6 of STE24 and the PGM48 model using a surface representation. Amino acid side chains are colored in black for hydrophobic residues (Leu, Ile, Val, Phe, Trp, Ala), red for acidic residues (Glu, Asp) and blue for basic residues (Lys, Arg, His). Dimensions (in Å) and approximate location of the ER/nuclear membrane lipid bilayer (STE24) and postulated PG lipid surface (PGM48) (dashed lines) are indicated



(Quigley et al., 2013). Figure 4.3A shows this structure overlaid with the PGM48 model, Figure 4.3B shows PGM48 model alone, and Figures 4.3C, D show it as a space filling model side-by-side with PGM48. The model quality is highest in the central portion of PGM48, which contains the catalytic site (residues 180 through 265). PGM48 completely lacks the sequence corresponding to TM1, 2, 3 present in STE24 (Figures 4.3A, B). The PGM48 model shows no transmembrane domains, but it has shorter hydrophobic helices that partially overlay to TM3-7 (Figure 4.3A, B); this is consistent with PGM48 being associated with a monolayer particle. The PGM48 active site residues (H191, E192, H195 and E240) are located in the short α -helices 7 and MH3, in close proximity to a $6 \times 10\text{\AA}$ diameter cavity, similar to that observed for STE24. Substrates are proposed to enter the STE24 cavity through a pore located close to the lipid surface. These substrates are prenylated at their C-termini (through the CAAX motif) and this (hydrophobic) prenyl group helps guide the substrate along the surface of the lipid bilayer into the central pore (Quigley et al., 2013). After two cleavages, the C-terminally truncated substrate is released. Interestingly, the STE24 homolog in yeast was recently shown to clear the ER Sec protein translocon from proteins clogging the translocation pore (Ast et al., 2016); these substrates are probably not prenylated. Similarly, despite older reports of protein prenylation (*e.g.* palmitoylation) in chloroplasts (Mattoo and Edelman, 1987; Parmryd et al., 1997) but see (Parmryd et al., 1999), no genetic or further biochemical support for intra-plastid prenylation have been found. Hence PGM48 is unlikely to have CAAX peptidase activity and it is most likely that PGM48 can cleave proteins without (C-terminal) lipid modifications. The model for PGM48 shows a pore situated at the postulated lipid interface (dashed lines) (Figure 4.3C, D) that could provide substrate access to the central cavity with the active site (Figure 4.3C, yellow asterisks).

4.3.5 Generation of under- and over-expression lines of PGM48

To determine the physiological function and (candidate) substrates of PGM48, we obtained, confirmed and characterized *Arabidopsis* T-DNA insertion lines for

PGM48 that we named *pgm48-1* (insertion in exon) and *pgm48-2* (insertion in intron) (Supplemental Figure 4.4A). As expected, RT-PCR was not able to amplify full length mRNA for either line due to the relatively large size of the T-DNA, and RT-PCR analysis showed >2 fold reduced mRNA accumulation of the region upstream of the T-DNA insertion for both lines (Supplemental Figure 4.4B). MS/MS proteome analysis was carried out on isolated PG of *pgm48-1*, which identified PGM48 at 2-5 fold reduced levels compared to wt (Supplemental Figure 4.4C). We concluded that neither line is a true loss-of function (null) mutant. No visible phenotype was observed in developing rosettes under a variety of abiotic conditions, but natural senescence was delayed in both alleles by about 1 week (Supplemental Figure 4.4D).

Because both T-DNA alleles were leaky, likely weakening the phenotypes, we generated more complete loss-of-function mutants using silencing *PGM48* by RNAi (driven by a 2x35S promoter). Additionally, we generated transgenic plants overexpressing full length PGM48 with a C-terminal StrepII tag driven by the constitutive 35S promoter. RT-PCR for several independent transformants (18 day old plants a few days prior to bolting) showed that *PGM48* mRNA was several fold increased in overexpression (OE) lines, and reduced to nearly undetectable levels in RNAi plants (even when using a higher cycle number) compared to wt plants (Supplemental Figure 4.5A). Immunoblotting with anti-strepII antiserum confirmed accumulation of intact PGM48-StrepII in the OE lines (Figure 4.4A). Immunoblotting of purified PG from wt, OE and RNAi lines showed that PGM48 was 2-3 fold higher in PG from the OE-1 line but virtually undetectable in the RNAi-1 line (Figure 4.4B). Semi-quantitative RT-PCR mRNA analysis (low cycle number) for rosettes of 21-28-35-42 day old plants (grown under long day length – 18 h light) showed that *PGM48* mRNA levels were many fold higher in the OE-1 and 2 lines than wt plants, and that *PGM48* mRNA was undetectable in the RNAi-1 and 2 lines throughout these developmental stages (Figure 4.4C; Supplemental Figure 4.5A). Extensive MS/MS protein analysis of isolated PG of wt and transgenic lines further confirmed the lack of PGM48 accumulation in the RNAi lines (RNAi-1 and 2) and increased levels in the

OE line (OE-1 and 2) (see below). Thus we successfully generated Arabidopsis lines with high constitutive levels of PGM48 or very low (undetectable) levels of PGM48.

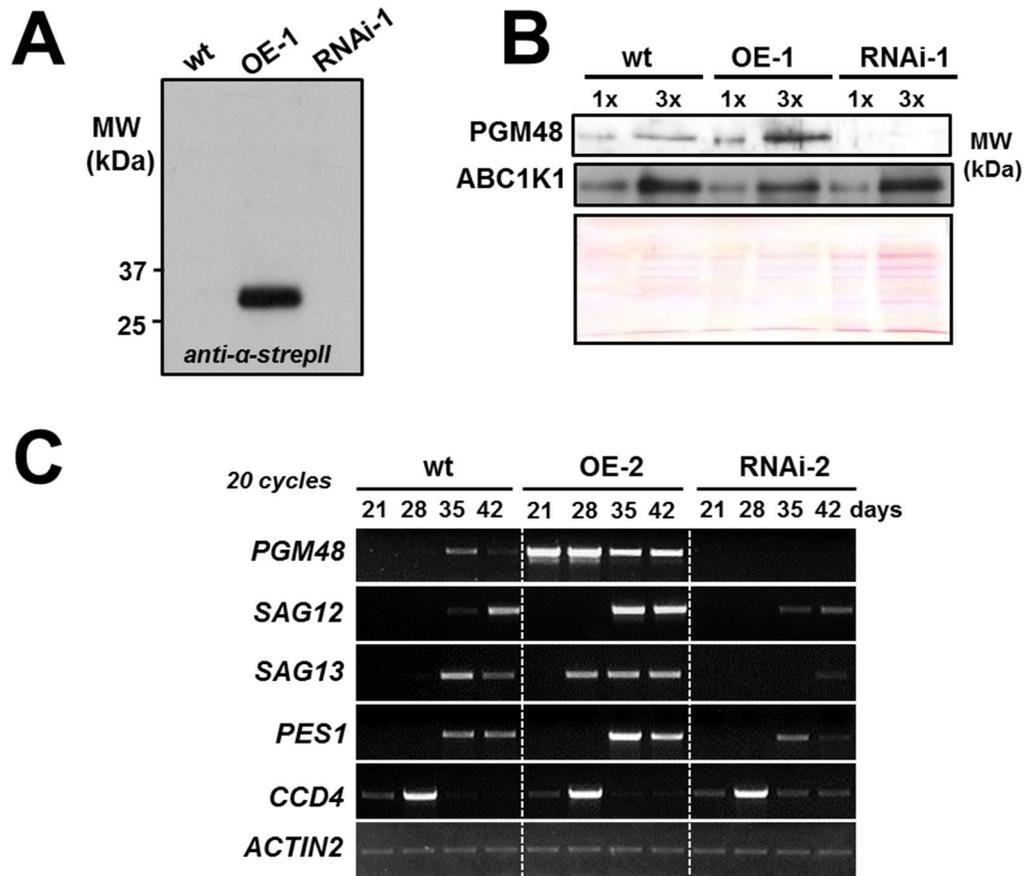


Figure 4.4. Overexpression and suppression of PGM48 in Arabidopsis and effect on mRNA levels during plant development and senescence. **(A)** Accumulation of transgenic StrepII-tagged PGM48 proteins in PG isolated from wt, OE and RNAi rosettes as determined by immunoblotting using specific anti-StrepII serum. **(B)** Accumulation level of endogenous and transgenic PGM48 and ABC1K1 proteins in PG isolated from wt, OE-1 and RNAi-1 rosettes as determined by immunoblotting. The ponceau stained blot is shown for loading control. 1x and 3x correspond to ~ 4 and 12 μ gram of loaded isolated PG proteins. **(C)** mRNA levels of PGM48, SAG12, SAG13, PES1, CCD4, in rosettes of wt, OE-2 and RNAi-2 lines after 21, 28, 35 and 42 days growth under 18h light/6h dark, 130 μ mol photons $m^{-2} s^{-1}$ light intensity

4.3.6 *PGM48 activates leaf senescence*

During the vegetative stages of seedling development (*i.e.* prior to bolting), the under- and over-expression lines did not show a visible phenotype (not shown). Bolting (the switch to reproductive phase and first appearance of inflorescence) started between 18-21 days in all three genotypes and was completed by day 21 (Supplemental Figure 4.5B) with the first flowers to appear at 22-25 days (under long day – 18 h light period). There was little change in bolting time between the three genotypes (Supplemental Figure 4.5B). *Arabidopsis* is a monocarpic plant and in absence of external (stress) factors, the onset of leaf senescence is driven by the developmental switch from vegetative to reproductive (Schippers et al., 2015). During these early flowering stages, the first visible signs of leaf senescence can be observed in the oldest leaves of the OE line, as well as wt plants (at 21-22 days), whereas the onset of senescence in RNAi lines was delayed by 2 days (Figure 4.5A; Supplemental Figure 4.5C). These differences became more pronounced in older rosettes (Figure 4.5A; Supplemental Figure 4.5D for additional transgenic lines). Quantification of senescence (categorized as green, green-yellow, yellow and brown/necrotic) for the individual leaves shows the progression of senescence from 28 to 35 days in all three genotypes, clearly demonstrating the accelerated senescence in the OE line and the delay in the RNAi line (Figure 4.5B). Consistently, leaf chlorophyll concentrations were significantly lower in the OE lines and higher in the RNAi lines as compared to wt at 28 days (Figure 4.5C) and 35 days (Supplemental Figure 4.5E). Interestingly, the total carotenoid to chlorophyll ratio was significantly higher in the OE line compared to wt and the RNAi line (Figure 4.5C), but there were no significant differences in leaf 5 and 6 at 35 days (Supplemental Figure 4.5E), likely because these leaves were now in a very advanced stage of senescence (Figure 4.5A,B; Supplemental Figure 4.5D).

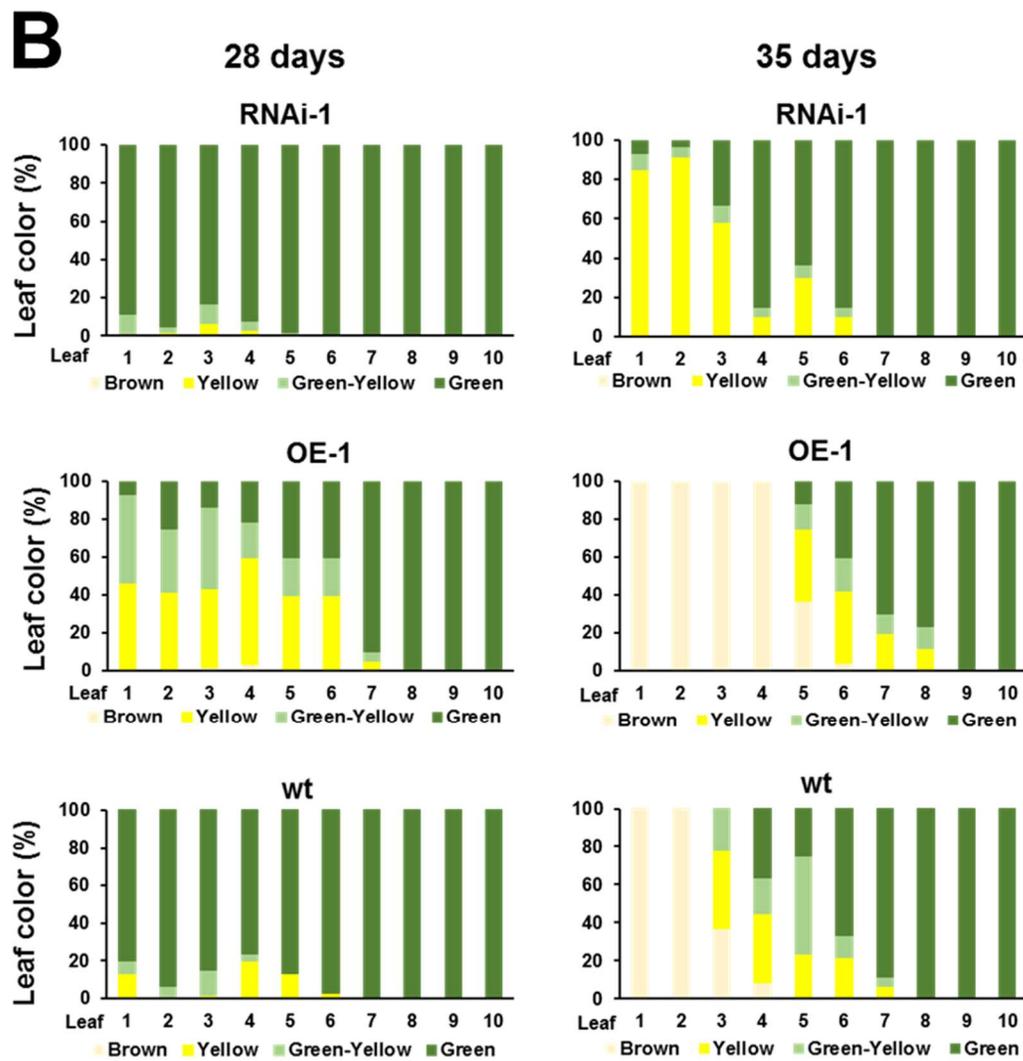
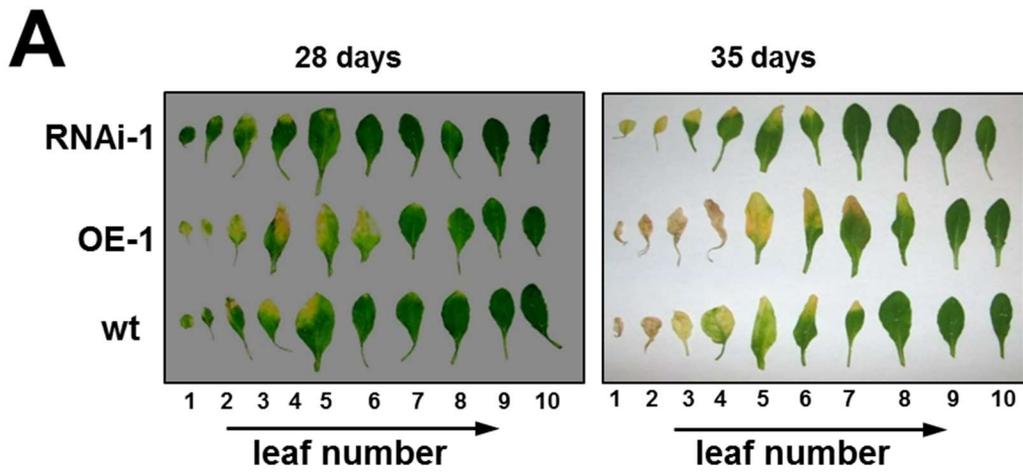
The acceleration and delay in senescence in respectively the OE and RNAi lines was also reflected at the mRNA level of the well-known senescence markers *SAG12* (encoding for the papain-like cysteine peptidase located in senescence-associated vacuoles (SAV) (Otegui et al., 2005; Carrion et al., 2013) and *SAG13*, a dehydrogenase with unknown function (Miller et al., 1999) (Figure 4.4C; see

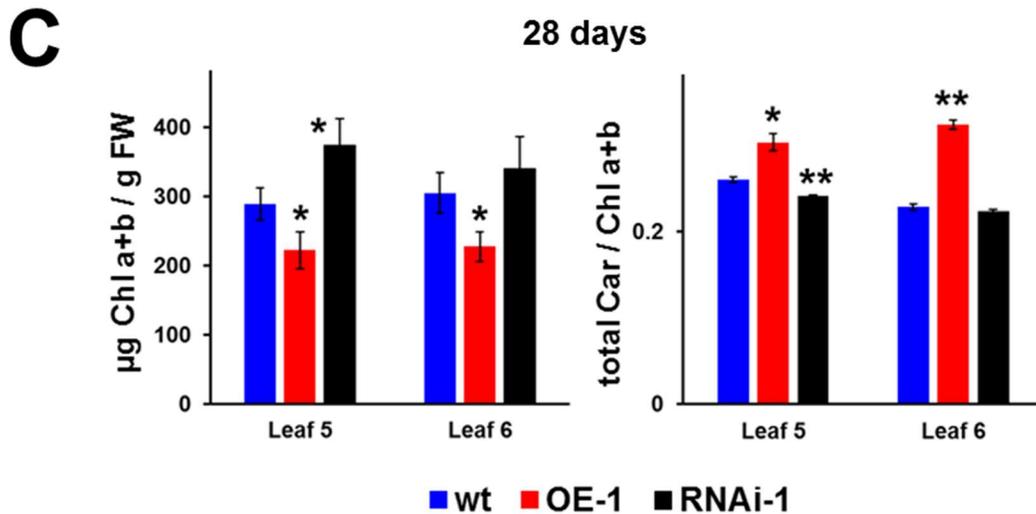
Figure 4.5. Overexpression and suppression of PGM48 in Arabidopsis shows that PGM48 protein accelerates natural leaf senescence.

(A) Examples of progression of natural leaf senescence of wt, OE and RNAi lines under 18h light / 6h dark, 130 $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$ light intensity. Ten individual leaves, numbered from old to young, were separated from a plant rosette for each of the three genotypes at day 28 and 35.

(B) Percentage of leaf color measured from leaf 1 to 10 for each genotype; the average values of for three rosette plants per genotype are presented. Dark green – healthy looking green tissue; light green - light green tissue undergoing degreening; yellow – yellow leaf section mostly devoid of visible chlorophyll; brown color - dead leaf.

(C) Chlorophyll a + b content on fresh weight bases and total carotenoid/total chlorophyll ratios for leaf 5 and 6 in the 3 genotypes at 28 days, for plants as shown in panels A, B. Significance levels are indicated - * $p < 0.05$; ** $p < 0.01$.





Supplemental Figure 4.5F for additional replicate). *SAG12* expression was visible in 35 days in the OE lines, but only detected at 42 days in wt and, at a much lower level in the RNAi-1 line. Similarly, *SAG13* was visible at 28 days in the OE lines, but at 35 days in wt, and at very low level in RNAi lines (Figure 4.4C; Supplemental Figure 4.5F). Thus mRNA accumulation patterns confirm accelerated senescence in the PGM48 OE lines and delayed senescence in the RNAi lines. Thus PGM48 is an activator of leaf senescence.

4.3.7 Effects of natural leaf senescence on the PG transcriptome and proteome

It is well established that the amount and size of PG increases during leaf senescence in many plant species (Besagni and Kessler, 2013). However, it is unknown if the proteome changes. To determine if the PG proteome in wt plants is affected by natural leaf senescence, we isolated PG from whole rosettes of plants in the bolting and advanced natural senescence stages. The PG proteomes were identified and quantified by label-free spectral counting using MS/MS after SDS-PAGE and tryptic digestion (Figure 4.6A; Supplemental Dataset 4.2). Whereas most PG core proteins did not change in relative abundance within the isolated PG, the relative abundance of PGM48 increased four-fold (Figure 4.6A). Furthermore, CCD4 stood

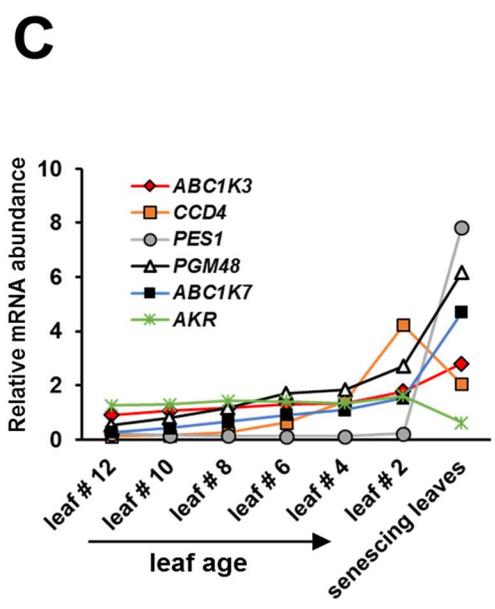
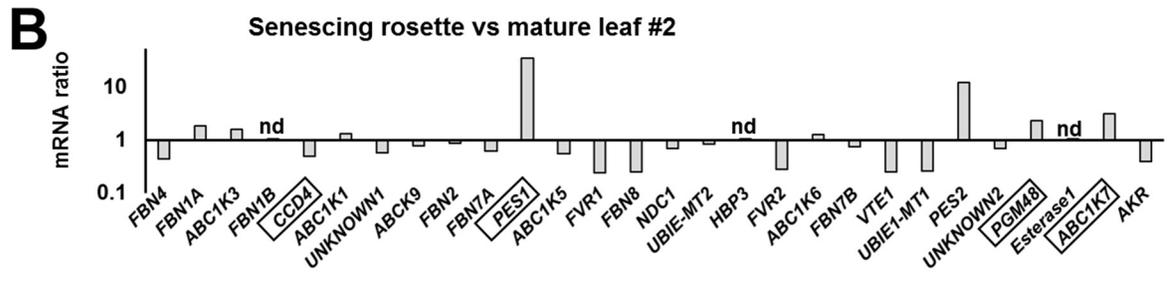
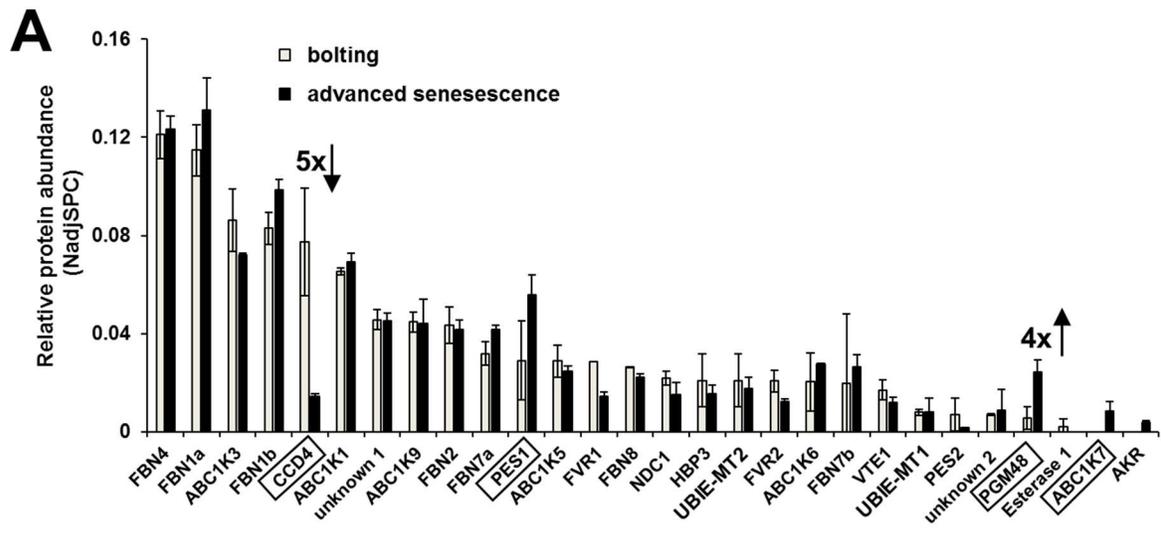
out for its strong, five-fold decrease during senescence. *ABC1K7* (part of the senescence mRNA expression module, together with *PGM48* and *PES1*) was not detected in the bolting stage, but was clearly detected in the senescence stage (with 18 MS/MS spectra) (Figure 4.6A). To understand if the proteome patterns are also observed at the mRNA level, we extracted the most relevant leaf mRNA expression data from the public domain for all genes encoding for PG proteins. A developmental leaf series of ten individual leaves from 17 day old rosette plants (continuous light), with leaf 2 being the oldest and leaf 12 the youngest, and senescent rosettes is most relevant here. Figure 4.6B, compares mRNA levels between senescing leaves and mature leaf #2 (best representing the bolting stage) for all available PG genes. This shows that in particular *PES1*, *PGM48* and *ABC1K7* are induced during senescence, whereas several others, in particular *CCD4*, *FLAVIN REDUCTASE-RELATED 1,2 (FVRI,2)*, *FBN8*, *VTE1* and *UBIE-MT1* decreased. Figure 4.6C shows mRNA abundance levels for *PGM48*, *CCD4*, *ABC1K7* and *PES1* in individual leaves from rosettes of 17 day old (continuous light), with leaf 2 being the oldest and leaf 12 the youngest. mRNA for *ALDO/KETO REDUCTASE (AKR)* is shown for comparison as an example of a gene not induced by senescence. *CCD4* mRNA very prominently increases with leaf age (upto ~ the bolting stage), but is then clearly declining in senescent leaves. This contrasts to *PGM48* which also increases with increasing leaf age but continues to increase in senescing leaves. Indeed, plotting the leaf mRNA expression level for all PG genes shows that *CCD4* and *PGM48* really stand out in their upregulation during leaf aging (Supplemental Figure 4.6A).

The micro-array data are consistent with the mRNA expression patterns observed experimentally by RT-PCR (Figure 4.4C), in that *CCD4* mRNA level of the rosette peaks around 28 days (when only a few of the oldest leaves on the rosette show some senescence – Figure 4.5A,B), but is dramatically decreased in advanced senescing plants (35 days) (Figure 4.4C; see also Supplemental Figure 4.5F for independent series). In contrast, mRNA levels of *PGM48*, *PES1* and *ABC1K7* peak in later stages of senescence (Figure 4.4D; Supplemental Figure 4.5F).

Figure 4.6. PG proteome composition in wt plants during natural leaf senescence and mRNA levels for PG genes

(A) PGM48 relative protein accumulation levels determined by label-free spectral counting quantitative proteomics of isolated PG at bolting stage and advanced senescence stage in wt. Results from two replicates of independent PG preparations were analyzed; relative abundance was normalized to the total amount of PG core protein. Plants were grown under short day conditions (10 h light/14h dark). For the complete abundance data of all PG core proteins, see Supplemental Dataset 2. CCD4 and PGM48 showed >4 to 5 fold difference between the bolting and advanced senescence stage.

(B, C) mRNA expression data from publicly available microarray data for leaves from wt (col-0); data downloaded from <http://bar.utoronto.ca/>. Panel B shows the relative abundance ratio between mRNA from mature leaf 2 and senescent leaves for all genes encoding for the PG core proteome, except for FBN1B, HBP3 and ESTERASE1 for which no data are available. Panel C shows mRNA abundance levels of individual leaves from rosettes of 17 day old (continuous light), with leaf 2 being the oldest and leaf 12 the youngest. This genes selected are PGM48 and CCD4, and showing ABC1K7 and PES1 as senescence-induced genes for comparison and AKR as a gene not induced by senescence. CCD4 mRNA very prominently increases with leaf age, but is clearly declining in senescent leaves; This contrasts to PGM48 which also increases with increasing leaf age but continues to increase in senescing leaves.



4.3.8 PG proteome from OE, RNAi and wt lines during senescence

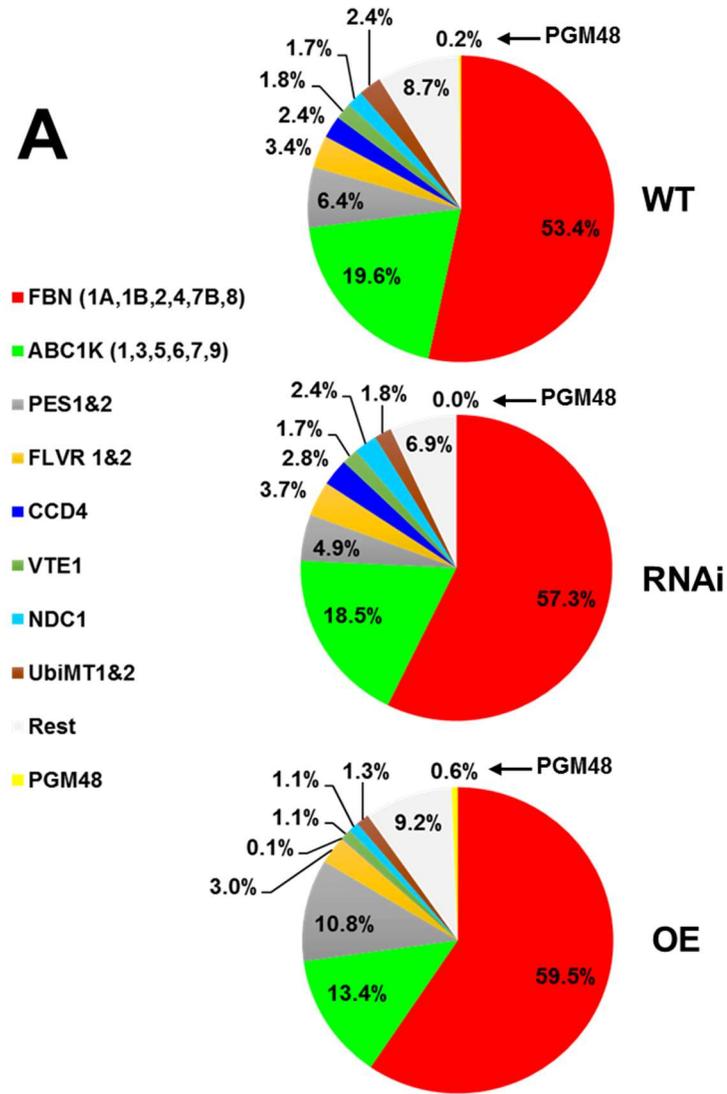
We then determined and compared the proteomes of PG isolated from wt, OE and RNAi lines from 35 days old senescing rosettes (as in Figure 5 and Supplemental Figure 4.5) with three biological replicates (using OE-1, 2 and RNAi-1, 2) to provide insight in possible targets of PGM48 peptidase activity. This identified 29 proteins out of the 30 known PG core proteins (low abundant At1G73750 with unknown function was missing) (Supplemental Dataset 4.3). Generally, the PG proteomes of the three genotypes were similar with 53-59% of protein mass invested in the FBN family and 13-20% to the ABC1K family (Figure 4.7A). However four PG core proteins (PGM48, CCD4, ABC1K7 and PES1) significantly ($p < 0.05$ or < 0.01) under- or over accumulated in the OE or RNAi lines (Figure 4.7B and Supplemental Dataset 4.3). PGM48 was 2.5 fold higher compared to wt and undetected in the RNAi lines, consistent with mRNA levels (Figure 4.7B). CCD4 decreased ~20 fold in the OE line (undetected in two replicates and just a few MS/MS spectra in the 3rd replicate) but was unchanged in the RNAi line compared to wt. ABC1K7 was 3-fold lower in the RNAi line compared to wt, whereas PES1 increased two-fold in the OE line and decreased 25% in the RNAi line.

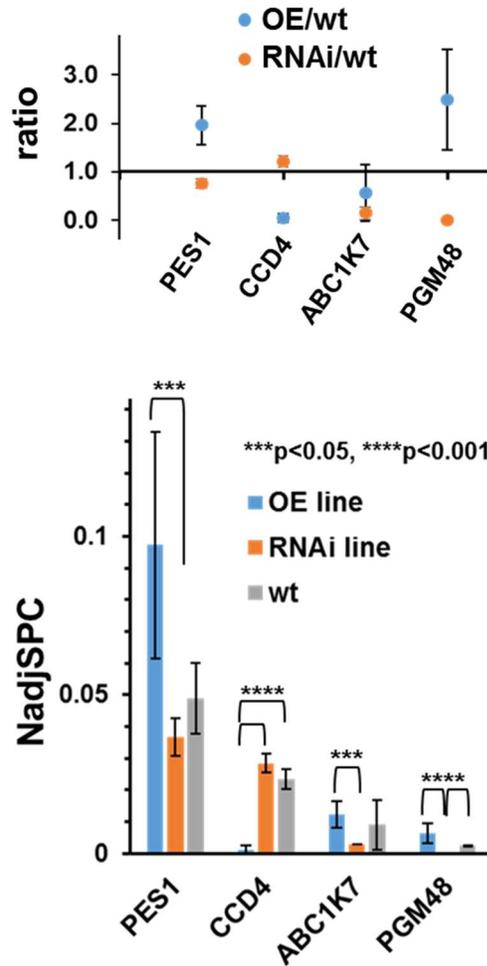
Figure 4.7. PG proteome composition in wt, OE and RNAi plants during natural leaf senescence determined by MS/MS-based label-free spectral counting of isolated PG.

(A) Relative mass abundance of PG core proteins and protein families normalized to the abundance of the sum of all observed PG proteins. 28 PG core proteins were observed. Proteomics data are listed in Supplemental Dataset 3 and spectral data are available at ProteomeXchange (see METHODS). PGs were isolated from 35 days old rosettes of wt, OE-1,2 and RNAi-1,2 (plants were grown under long day conditions at 18h light/6h dark) and analyzed by MS. PG proteins were determined by label-free quantitative comparative proteomics. PGM48 was not identified in the RNAi lines. The analysis was carried out with three sets of independent biological replicates with two replicates for OE-1 and RNAi-1 and one replicate for each OE-2 and RNAi-2 (see Supplemental Dataset 3).

(B) Statistically significant pair-wise differences of normalized abundance levels (using NadjSPC) for PG core proteins across the three genotypes was determined by student t-test. PES1, CCD4, ABC1K7 and PGM48 showed significant pair-wise differences ($p < 0.05$ or $p < 0.01$) as indicated in the bar diagram. Protein abundances were quantified based on the number of matched adjusted MS/MS spectra (NadjSPC). Data shown here are average of three independent replicate and bars indicate standard deviations ($n=3$). Abundance ratios between OE/wt and RNAi/wt of PES1, CCD4, ABC1K7 and PGM48 are shown in the upper panel.

A



B

4.3.9 *PGM48* interacts with *PES1*, *CCD4* and *ABC1K3* *in vitro*

Since *PGM48* shows a role in senescence, we tested if *PGM48* can interact with chloroplast chlorophyll catabolic enzymes *NYC1*, *RCCR*, *HCAR*, *PAO* and *PPH* proteins (Sakuraba et al., 2015) in mating-based split ubiquitin yeast two hybrid assays. This did not identify interactions between these proteins and *PGM48* (Figure

4.8A). However, testing interaction of PGM48 with nine PG core proteins showed that PGM48 interacted with PES1, CCD4, ABC1K3 and a weak interaction was also observed with UBIE-MET, but no interaction was found for ABC1K1, PES2, ABC1K7, ABC1K9 or PGSAG (Figure 4.8A). Finally, thylakoid peptidase FTSH2, chloroplast membrane SAG protein (AT2G20920), and senescence-induced stay-green protein 1 (NYE1/SGR1) did not interact with PGM48 either (Figure 4.8A).

We then tested PGM48 interactions with PES1 and CCD4 using *in vitro* pull-down assays with purified recombinant proteins expressed in *E. coli*. Recombinant PGM48 was generated with a His₆ tag at the C terminus, whereas recombinant CCD4 and PES1 were N-terminally fused with GST. Recombinant PGM48-His₆ was incubated with recombinant GST-CCD4, GST-PES1, or GST (as negative control), followed by affinity purification using Ni-NTA resin. After extensive washing, bait protein (PGM48-His) eluted with GST-CCD4 and GST-PES1, but not with GST (Figure 4.8B), confirming the observation by the yeast 2H system. Subsequently, we probed for interactions for the same set of proteins, but this time using the GST-fusions as bait, employing glutathione resin to capture GST-CCD4, GST-PES1, GST (as negative control) and testing if PGM48 is pulled down with these baits (Figure 4.8C). We included GST-PGSAG as an additional bait. As shown in Figure 4.8C, PGM48 eluted with GST-CCD4 and GST-PES1 indicating a positive interaction with PGM48, whereas PGM48 did not elute with GST-PGSAG nor GST alone (Figure 4.8C). Thus these reciprocal pull-down experiments further supported interactions of PGM48 with CCD4 and PES1. Co-immunoprecipitations of endogenous PGM48 in PG isolated from naturally senescing leaves using our anti-PGM48 antiserum followed by MS/MS, did successfully enrich for PGM48, but did not detect specific protein interactors. Whereas we did apply cross-linkers prior to these co-IP, it is likely that PGM48 interactions to target substrates are highly transient and therefore difficult to capture.

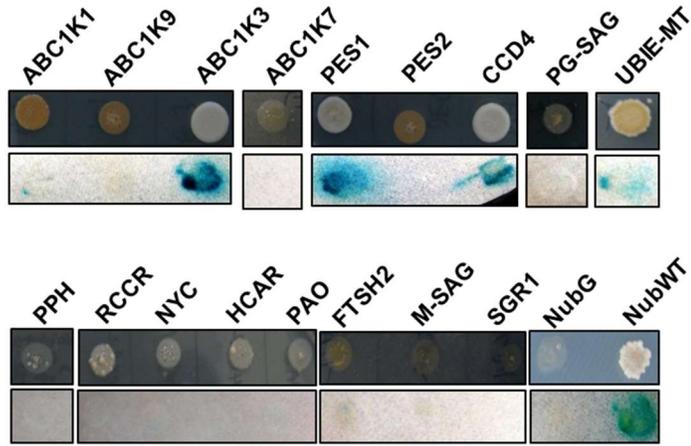
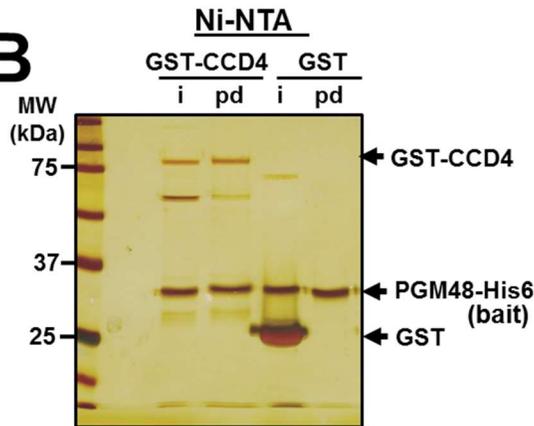
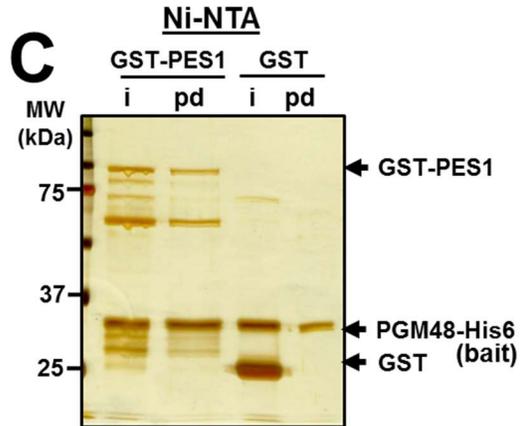
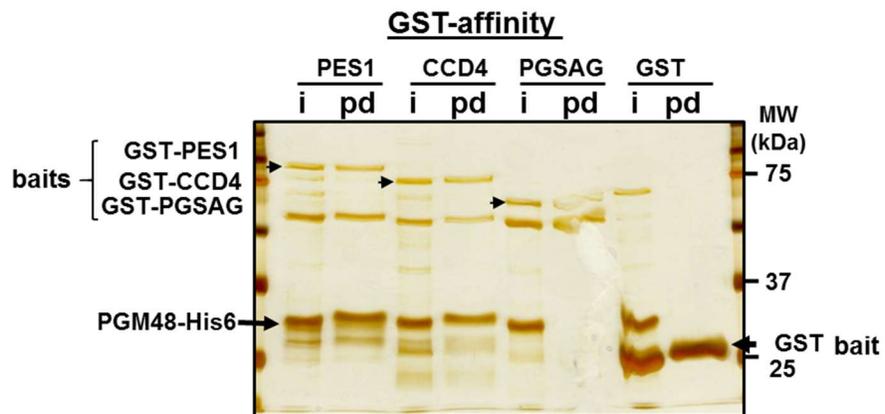
Figure 4.8. Interaction of PGM48 with other PG proteins.

(A) Yeast two-hybrid assays for interactions between PGM48 and selected proteins. PGM48 was used as a bait and selected candidate proteins were used as prey. Bait plasmid contains Cub-PLV and prey plasmid contains NubG. NubG moiety was fused to the N terminus of prey proteins. The resulting plasmids were transferred into the different yeast strains for bait and prey. The transformed yeasts harboring bait and prey constructs were mated and resulting transformants were analyzed on selective medium lacking Ade, His, Try, Leu, Ura and Met (upper lane) and for β -galactosidase (β -Gal) activity (lower lane). Soluble NubG and Nub-WT were used as negative and positive control, respectively. UBIE-MT1 – UbiE-methyltransferase related 1 (AT1G78140).

(B) Ni-NTA resin mediated pull down of PGM48-His with GST-CCD4. Purified GST alone (10 μ g) or GST fused CCD4 (2 μ g) was incubated with purified PGM48 fused with His6 (2.0 μ g), together with Ni-NTA resin. Input (50% of the reaction) and eluted samples (80%) were loaded into a SDS-PAGE gel and analyzed by silver staining. i = input; pd = pull-down eluate.

(C) Ni-NTA resin mediated pull down of PGM48-His with GST-PES1. Purified GST alone (10 μ g) or GST fused PES1 (3 μ g) was incubated with purified PGM48 fused with His6 (2.0 μ g), together with Ni-NTA resin. Input (50%) and eluted samples (80%) were loaded into a SDS-PAGE gel and analyzed by silver staining. i = input; pd = pull-down eluate.

(D) Glutathione resin mediated pull down of GST-PES1 and GST-CCD4 with PGM48-His. Purified GST-PES1 (2 μ g), GST-CCD4 (2 μ g), GST-PGSAG (2 μ g) and GST alone (10 μ g) was incubated with purified PGM48-His (3 μ g) recombinant protein, together with glutathione resin. Input (50% of the reaction) and eluted samples (80%) were loaded into a SDS-PAGE gel and analyzed by silver staining. i = input; pd = pull-down eluate.

A**Y2H - PGM48-bait****B****C****D**

4.4 DISCUSSION

4.4.1 PGM48 represent a specific adaptation of photosynthetic organisms

PG are plastid micro-compartments with multiple functions integrated in plastid metabolism, developmental transitions and environmental adaptation (van Wijk, 2016). The challenge is to unravel how PG and its molecular components contribute to plastid homeostasis. This study concerns the role of the PG-localized metallo-peptidase PGM48 in leaf chloroplasts. Through *in vitro* assays, we showed that PGM48 is a relatively low abundant Zn-dependent metallopeptidase and belong to the peptidase clan MA containing a variety of metallopeptidases. The families in clan MA are united by the presence of an HEXXH motif in which the two His residues are Zn ligands and the Glu has a catalytic function (Rawlings et al., 2016). Importantly, phylogenetics showed that PGM48 is part of a previously unknown clade within the M48 family, here assigned M48D, with proteins only present in photosynthetic organisms, including cyanobacteria, algae and higher plants. Moreover, unlike M48 homologs in the other three clades, members of the M48D clade have no predicted transmembrane domain, further indicating that this subfamily represents a specific adaptation (invention) of photosynthetic organisms. The lack of TMDs is consistent with the complete lack of TMDs in other PG localized proteins (Lundquist et al., 2012) and is logic given that PG are monolayer particles. In addition to PGM48 in Arabidopsis PG, we also identified the maize PGM48 homolog in isolated maize PG (Huang et al., 2013); we predict that PGM48 homologs in other plant species, as well as algae, are also located in PG. The phylogeny suggests that plastid PGM48 homologs originate from the cyanobacterial endosymbiont. Most cyanobacteria have both PGM48 (clade M48D) and HTPX homologs (clade M48B), suggesting that the PGM48 members originated from gene duplication of HTPX and obtained a specialized function in photosynthetic organisms. The primitive lycopod *Selaginella moellendorffii* and moss *Physcomitrella patens* lack PGM48, but each do possess a member of the M48A and C clades, indicating independent gene loss in these species, either because these species don't have PG or because they developed alternative

strategies to carry out PGM48 functions. *P. patens* chloroplasts can accumulate lipid-droplets under stress conditions, but it is not clear if these are indeed PG (Wang et al., 2009). There are several reports of PG in cyanobacteria (often named lipid particles or lipid droplets in the cyanobacterial literature) (van de Meene et al., 2006; Peramuna and Summers, 2014), but their proteome composition is not known. High-resolution three-dimensional reconstruction of the cyanobacterium *Synechocystis sp. PCC 6803* showed that lipid particles were abundant and that their distribution was restricted to thylakoids (van de Meene et al., 2006), similar as in *Synechococcus sp. PCC 7002* (Nierzwicki-Bauer et al., 1983). Their intracellular location suggests a role in thylakoid maintenance or thylakoid biogenesis, similar as for plastid PG. The PGM48 homolog in *Synechocystis sp. PCC 6803* (sll1280) was identified by mass spectrometry (Wegener et al., 2010; Liberton et al., 2016). Experimental studies on PGM48 in cyanobacteria may help further determine why the M48D clade evolved in photosynthetic organisms.

4.4.2 PGM48 substrate selection, interaction and cleavage

The PGM48 homology model suggests that the protein in its native state has a hydrophobic surface suitable to interact with the PG monolayer. The pore providing access to the active site in the cavity is facing the aqueous phase but is not far from the presumed lipid surface. This would make proteins that are part of the PG the most likely substrate candidates. It is so far unclear how a substrate(s) is recognized by PGM48, similar as it is unknown how substrates for members of the other M48 clades are selected (see below). In case of the *in vitro* substrate β -casein, cleavage by PGM48 occurred at (or from) the C-terminal end. Beta casein has very little native structure (it is classified as an intrinsically disorganized protein) and is highly sensitive to peptidase digestion. (Oldfield and Dunker, 2014). The observation that *in vitro* cleavage of beta-casein by recombinant PGM48 occurred at/from the C-terminal end suggest that this is the preferred confirmation of substrate presentation and cleavage. PG48 and its homologs in the M48D clade likely have unique properties and

substrates compared to the other clades, and its location in PG seems to exclude a general role in chloroplast proteostasis.

4.4.3 PGM48 is associated with senescence and control of senescence and stay-green mutants

Senescence induction depends on the developmental window. Following cell proliferation and expansion (Phase I), the leaf matures and become competent for external signals that induce senescence (Phase II). That competence increases with leaf age and in phase III, this will result in the initiation of developmental senescence, independent of external factors (Schippers, 2015; Schippers et al., 2015). A transcriptional network drives senescence in stage III and also involves several plant hormones as positive regulators (ethylene, abscisic acid - ABA, jasmonic acid -JA, salicylic acid - SA), or negative regulators (brassinosteroid – BR and cytokinin - CK). The impact of these hormones on senescence has been demonstrated though manipulation of TFs, signaling/receptor components or hormone biosynthetic pathways. For example overexpression of the TF EIN3 causes early senescence (Chen and Bleecker, 1995), whereas ethylene insensitive mutants *ein1* and *ein2* show delayed senescence (Li et al., 2013). Similarly, manipulation of CK biosynthesis can delay the onset of senescence (Gan and Amasino, 1995) because CK increases sink strength of the tissue (Schippers et al., 2015). During leaf senescence, thylakoid membranes and their protein complexes together with associated co-factors are dismantled in a controlled fashion (Hortensteiner, 2009; Hortensteiner and Krautler, 2011; Besagni and Kessler, 2013; Kusaba et al., 2013; Avila-Ospina et al., 2014; Ishida et al., 2014). Transmission and scanning electron microscopy, documenting dramatic increase of PG volume during this natural senescence process (Lichtenthaler, 1968; Tuquet and Newman, 1980; Tevini and Steinmuller, 1985; Ghosh et al., 1994; Guiamet et al., 1999; van Wijk, 2016). The experimental data presented in this study, clearly show that over- or under-expression of PGM48 result in respectively an acceleration or delay of natural leaf senescence. One can therefore postulate that PGM48 i) removes a negative regulator (*e.g.* associated with BR or CK pathways), or ii) results in activation (or

increased levels) of positive senescence regulators or signals (*e.g.* associated with ethylene, ABA, JA or SA). However, PGM48 is exclusively accumulating in PG, which narrows the immediate substrate pool to the PG proteome or proteins that transiently interact with the PG, and eliminates direct effects on nuclear TF or extra-plastidic enzymes.

4.4.4 Hypothetical functional models for PGM48 action and role in senescence

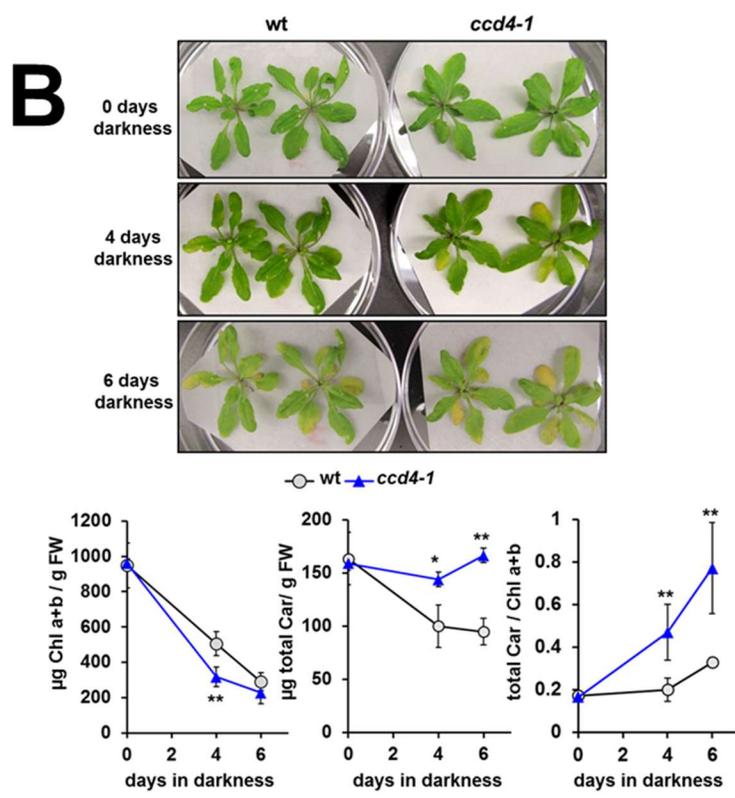
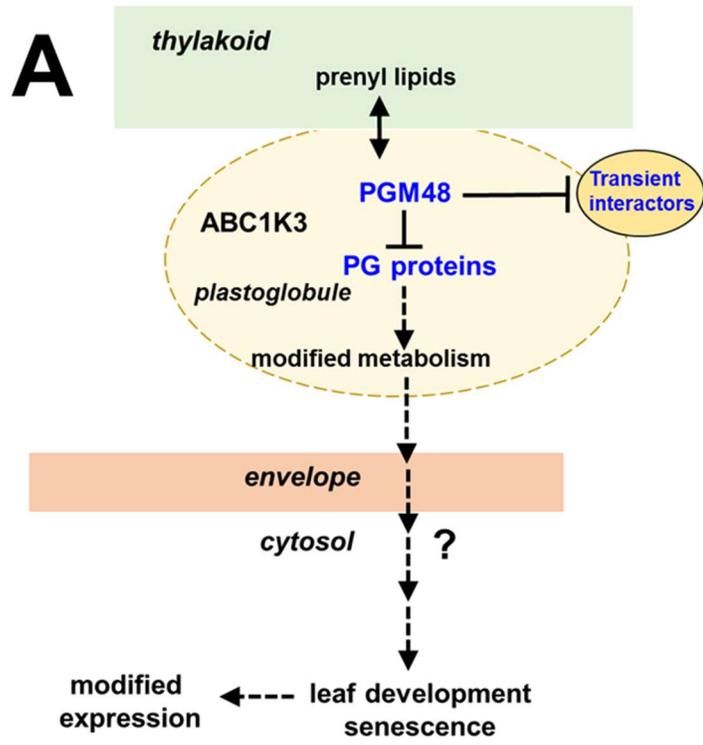
The early PGM48 over-expression and late under-expression senescence phenotypes, indicates that M48 degrades or inactivates a protein that is either an (indirect) repressor of senescence or activates a protein (indirectly) required for senescence by partial cleavage. The metabolite composition of leaf PG isolated from four stages of natural senescence in beach leaves (*Fagus sylvaticus*) were quantitatively (on weight basis) compared (Tevini and Steinmuller, 1985). In particular, prenylquinones and free fatty acids, but not glycolipids or proteins, accumulated in PG during the senescence process ((Tevini and Steinmuller, 1985) and references therein). Interestingly, the PG content of carotenoids and carotenoid esters was very low in green leaves, but increased in early stages of senescence (while total leaf carotenoid content decreases), and then decreased to low levels in more advanced stages of senescence (Tevini and Steinmuller, 1985). Thus the PG participates in the controlled removal of carotenoids from the thylakoid membrane during the senescence process. The mechanism by which the carotenoid content decreases in PG is unknown.

Based on our current observations and information from published literature, we propose a functional model that explain how PGM48 affects leaf senescence (Figure 4.9A). The model proposes that PGM48 degrades or partially cleaves one or more PG-localized or transiently associated PG proteins, resulting in modified metabolism and/or metabolite content, thereby accelerating leaf senescence through an unknown retrograde signaling pathway. Possible signals for induction or acceleration of senescence are ABA or JA. Currently, there is no direct evidence for PGM48 substrates, but based on the data presented in the current study and insight from the literature, we suggest that CCD4 is a good candidate.

Figure 4.9. Functional models for PGM48 function in senescing chloroplasts and phenotype of the CCD4 null mutant during dark-induced senescence.

(A) This model proposes that PGM48 degrades or partially cleaves one of more PG-localized proteins, resulting in modified metabolism and/or metabolite content, thereby accelerating leaf senescence through an unknown retrograde signaling pathway. Possible signals for induction or acceleration of senescence are ABA or JA. So far, no direct substrates for PGM48 have not been identified but we suggest CCD4 as a strong candidate (see DISCUSSION).

(B) The Arabidopsis *ccd4* T-DNA insertion null mutant (SALK_097984) shows accelerated dark-induced leaf senescence after 4 days darkness as compared to wt as evidenced by accelerated chlorophyll degradation, whereas degradation of carotenoids is delayed (lower panels). Throughout dark-induced senescence, but not prior to darkness, carotenoid to chlorophyll ratios are higher in the *ccd4* mutant, consistent with its role in carotenoid degradation (Gonzalez-Jorge et al., 2013). Student t-test (* $p < 0.1$, ** $p < 0.05$).



Degradation of CCD4 by PGM48 is supported by the fact that i) both PGM48 and CCD4 locate exclusively to PG, ii) yeast-2-hybrid and *in vitro* protein interaction data suggest physical interaction between PGM48 and CCD4. The comparison of mRNA accumulation patterns for all genes encoding for PG proteins shows a strikingly similar high induction of *CCD4* and *PGM48*, in particular as compared to all other PG genes (Figure 4.6 and Supplemental Figure 4.6); whereas the latter observation constitutes no direct proof that PGM48 degrades CCD4, it does fit the pattern. The very clear observations of the loss of CCD4 protein within the PG during senescence will require proteolytic activity and PGM48 is an excellent candidate, also because it is the only protease in PG. Recombinant PGM48 did not degrade recombinant GST-CCD4 *in vitro* (data not shown) but this does not exclude *in vivo* degradation, perhaps because degradation would require a monolayer lipid interface and/or ABC1K3 dependent activation of PGM48 or phosphorylation of CCD4.

CCD4 as a target for PGM48 is consistent with the observed phenotype for the PGM48 OE and RNAi lines. The Arabidopsis *ccd4* T-DNA insertion null mutant (SALK_097984) shows accelerated dark-induced leaf senescence after 4 days darkness as compared to wt, as evidenced by accelerated chlorophyll degradation and delayed degradation of carotenoids (Figure 4.9B). Throughout dark-induced senescence, but not prior to darkness, carotenoid to chlorophyll ratios are higher in the *ccd4* mutant, consistent with its role in carotenoid degradation (Gonzalez-Jorge et al., 2013). Consistently, carotenoid to chlorophyll ratios are also increased in the PGM48 overexpression line during senescence compatible with reduced CCD4 activity. CCD4 is known to function in degradation of carotenoids resulting in production of various volatile apocarotenoids and not only reducing carotenoid content (Gonzalez-Jorge et al., 2013) but also affecting leaf development though as yet uncharacterized signal transduction pathways (Avendano-Vazquez et al., 2014; Hou et al., 2016). PGM48 driven degradation of CCD4 would reduce production of these apocarotenoid signals, and may also result in increased flux into the downstream ABA biosynthetic pathway; an increase in ABA levels would stimulate leaf senescence. It remains to be determined if PGM48 driven degradation of CCD4 is indeed the primary cause of the

senescence phenotypes observed. Future research is needed to obtain direct proof for PGM48 substrates, but identifying such direct substrate-protease relationships is notoriously challenging for many proteases.

4.5 METHODS

4.5.1 *Phylogenetic analysis and mRNA expression analysis*

To generate a phylogenetic cladogram, 88 M48 proteins from 39 species across the tree-of-life were aligned by using MUSCLE (<http://www.ebi.ac.uk/Tools/msa/muscle/>). The alignment is available as Supplemental Data text 4.1. The aligned sequences were exported in Clustal format and viewed in Jalview (www.jalview.org/). Sequences were then converted in PHYLIP format, and phylogenetic trees were generated (1000 iterations) using the CIPRES Web portal (<http://www.phylo.org/>) selecting the tool “RAxMLHPC blackbox”. The resulting phylogenetic tree was annotated in FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>). RAxML bootstrap support values are shown at the nodes of the tree. *In silico* mRNA expression Arabidopsis tissue specific expression profile was derived from the e-FP browser <http://bar.utoronto.ca/ExpressionAngler/>.

4.5.2 *Generation of anti-PGM48 anti-serum*

The nucleotide sequence encoding amino acids 72 to 325 of PGM48 were amplified by PCR. The resulting DNA fragment was ligated into restriction sites (*Bam*HI and *Xho*I) of the C-terminal His affinity tag of the pET21a expression vector. BL21 *E. coli* cells were transformed by pET21a vector harboring this truncated *PGM48* gene and cells were harvested from liquid culture after addition of 1mM IPTG for 3 h incubation at 22⁰C. The over-expressed proteins were solubilized in 200 mM NaCl, 50 mM Tris, and 8M Urea at pH 8 and purified on a nickel-nitrotri-acetic acid agarose (NTA) resin matrix. A polyclonal antibody against this truncated PGM48

protein was raised in rabbits by injecting purified antigen (Alfa Diagnostic International, Texas). Antisera were affinity purified against the same antigen coupled to a HiTrap N-hydroxysuccinimide (*NHS*) ester-activated column (GE Healthcare Life Science).

4.5.3 T-DNA insertion mutants, genotyping and RT-PCR

T-DNA insertion lines *pgm48-1* (SALK_082409) and *pgm48-2* (GABI_324A06) were obtained from the Arabidopsis Biological Resource Center (ABRC) and European Arabidopsis stock center (NASC), respectively. TDNA inserted plant was identified by genotyping and insertion was confirmed by DNA sequencing. Details on genotyping primers were listed in a table (Supplemental Table 4.3). RNA was collected from homozygous plants and performed RT-PCR by using two different primers set. For transcript analysis, total RNA was extracted from Arabidopsis leaves using the RNeasy plant mini kit (QIAGEN). RNA was reverse transcribed with random hexamer primers by using Superscript III reverse transcriptase from Invitrogen. mRNA was normalized by *ACTIN2* and the PCR condition was 28 cycles at 94°C for 2 min, 55°C for 30S and 72°C for 1 min. A complete list of primers can be found in Supplemental Data (Supplemental Table 3).

4.5.4 Generation of constitutive overexpression and RNAi lines

To generate transgenic plants overexpressing PGM48 (OE lines), full length of PGM48 cDNA was cloned by using forward and reverse primers. For a complete list of primers see Supplemental Table 4.3. A nucleotide sequence encoding the StrepII tag was added in the reverse primer before the stop codon. PCR products were cloned into pCR8 topo vector (in vitrogen) and verified by DNA sequencing. This clone was ligated into gateway pEARLYGATE100 vector (ABRC stock center) by using LR enzymes (in vitrogen). To suppress PGM48 gene expression through RNAi (RNAi lines), a partial cDNA (Nucleotide 1 to 304) was cloned into pCR8 topo vector, verified by DNA sequencing, and cloned into the pRNAi-GG vector containing the 2X35S promoter (from ABRC stock center-Yan et al, 2012). These vectors were

transformed into *Agrobacterium* and subsequently transformed into wt *Arabidopsis* by floral dip method.

4.5.5 Plant Material, growth, light regimes, antibiotic selection, genotyping and RT-PCR

Transgenic plants were selected by using BASTA (for pEARLYGATE100 vector) and kanamycin (pRNAi-GG vector) containing plates. Plants surviving on selective medium were genotyped and confirmed presence of the transgene and were transferred to soil for seed production. Plants were grown in a growth chamber with long day conditions (18h light/6h dark) and temperature at 22°C with 130 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ light intensity. For transcript analysis, total RNA was extracted from *Arabidopsis* leaves using the RNeasy plant mini kit (QIAGEN). RNA was reverse transcribed with random hexamer primers by using Superscript III reverse transcriptase from Invitrogen. mRNA was normalized by *ACTIN2* and the PCR condition was 20 or 22 cycles at 94°C for 2 min, 55°C for 30S and 72°C for 1 min. A complete list of primers can be found in Supplemental Table 3.

4.5.6 Pigment concentrations

Chlorophyll and carotenoid concentrations were determined by absorbance spectrometry after extraction in 80% acetone (Porra, 1989). For chlorophyll and carotenoid determination, we collected specific leaves (leaves 5 and 6) from OE, RNAi and wt plants. In case of *ccd4-1* mutant we have collected the leaves which turn to yellow after 4 and 6 day dark treatment and same leaf number was also collected from wt.

4.5.7 PG isolation

The PG isolation method was from (Ytterberg et al., 2006; Lundquist et al., 2012). For PG preparations from senescing leaves, plants were grown on soil under 120 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ with 18 hour photoperiod for 35 days. Rosettes were harvested and homogenized in grinding buffer (50 mM HEPES-KOH pH 8.0, 5 mM

MgCl₂, 100 mM sorbitol, 5 mM ascorbic acid, 5 mM reduced cysteine, 0.05 % (w/v) BSA). Homogenate was filtered through four layers of 20- μ m miracloth and thylakoid membranes were pelleted by centrifugation for 6 minutes at 1800xg. Thylakoid pellets were washed once in 4 volumes of grinding buffer and resuspended in Medium R (50 mM Hepes-KOH pH 8.0, 5 mM MgCl₂, and cocktail of peptidase inhibitors) containing 0.2 M sucrose. Resuspended thylakoids were sonicated 4x 5s at output power 23 Watts (Fischer Scientific, sonic dismembrator Model 100), returning the samples to ice between each sonication event. Sonicated samples were centrifuged for 30 minutes at 150,000 xg and PG released from the thylakoid floated to the surface of the solution. PG were removed and combined with Medium R with 0.7M sucrose to achieve a sucrose concentration of 0.5M, which was then overlaid with Medium R with 0.2M sucrose and Medium R with no sucrose. The gradient was centrifuged 90 minutes, 150,000 xg. The resulting floating pad of PG was removed, flash frozen in liquid N₂, and stored at -80° C.

4.5.8 *Yeast two hybrid assay*

Y2H was carried out as previously described (Bhuiyan et al., 2015). The coding sequence of mature PGM48 (amino acid 48 to 344) was cloned in pMetYC-DEST vector (Grefen et al., 2007) and used as bait for interaction studies. Different PG localized proteins, chlorophyll catabolite enzymes, and other candidate proteins were cloned into pXN22-DEST or pNX32-DEST (Grefen et al., 2007) vector and used as prey for the interaction with bait proteins. Haploid yeast strains THY.AP4 and THY.AP5 were transformed by bait constructs and prey constructs, respectively. Bait plasmid contains Cub-PLV and prey plasmid contains NubG. NubG moiety was fused to the N terminus of prey proteins. Yeast strains were purchased from the ABRC stock center. After mating THY.AP4 and THY.AP5, diploid cells were selected growing on synthetic medium lacking Try and Leu. Positive colonies were subjected for testing interaction. Interactions between bait and prey proteins were performed according to the protocol described (Grefen et al., 2007). Interaction were verified by growing yeast colonies on synthetic minimal medium lacking Ade, His, Try, Leu, Ura and Met and

also by β -galactosidase (β -Gal) activity. Soluble NubG and Nub-WT was used as negative and positive control, respectively.

4.5.9 *PGM48 site-directed mutagenesis and in vitro proteolytic activity assays*

Mature PGM48 (without cTP) was cloned by using forward (PGM48-M-FW) and reverse (PGM48-M-RV) primers. The forward primer contains *Bam*HI and the reverse primer contains *Sal*I sites. The resulting PCR fragment was ligated into pCR8 topo vector and confirmed by DNA sequencing. pCR8 vector harboring *PGM48* gene was digested by *Bam*HI and *Sal*I restriction enzymes. The resulting DNA fragment was ligated into restriction sites (*Bam*HI and *Sal*I) of the C-terminal His affinity tag of the pET21a expression vector.

Two mutants H191A and E192D of PGM48 were constructed by using a PCR method. For the mutant H191A, the C terminal part of mature protein was amplified from pCR8 plasmid harboring *PGM48* gene by using forward primer (PGM48H191A-FW) and reverse primers (PGM48-M-RV). Forward primer M48H191AFW contains the mutation site CAT (His) to GCT (Ala). N terminal part of mature protein was amplified by using forward (PGM48-M-FW-) and reverse (PGM48H191A-RV) primers. Reverse primer PGM48H191A-RV contains the introduced site CAT (His) to GCT (Ala). The amplified two fragments were gel purified and mixed and used as template (1:1) for second round PCR to amplify mature protein by using the forward and reverse primer sets, PGM48-M-FW and PGM48-M-RV, respectively. Second mutant E192D was amplified the same way as H191A except for the different primer sets to introduce the mutation from GAA (Glu) to GAT (Asp). The PCR fragments were ligated into pCR8 topo vector and the mutations were confirmed by DNA sequencing. BL21 *E. coli* cells were transformed by pET21a vector harboring the various PGM48 constructs and cells were harvested from liquid culture after addition of 1mM IPTG for 3 h incubation at 22^oC.

Overexpressed wt and mutants PGM48 were solubilized in 300 mM NaCl, 50 mM Tris, 10% glycerol, at pH 8 and purified on a nickel-nitrotri-acetic acid agarose (Ni-NTA) resin matrix. The purified protein was dialyzed by using dialysis cassette

(slide-A-Lyzer, Thermo Scientific) against buffer 100 mM NaCl and 50 mM Tris and 10% glycerol. After dialysis the protein was concentrated by using microcon centrifugal filter units (Millipore). *In vitro* proteolytic activity was performed by incubating recombinant proteins with β -casein at 37⁰C for 3 hrs with or without various peptidase inhibitors. The reaction was stopped by adding 3% SDS and then followed by separation of the protein products by SDS-PAGE followed by staining with Coomassie or silver nitrate.

To determine cleavage site specificity, recombinant active PGM48 and inactive PGM48-H191A were incubated with β -casein for 3 hrs at 37⁰ C, followed by desalting using C4 Ziptip (Millipore) using the manufacturer's guidelines. The peptide mixture was directly infused into a LTQ-Orbitrap MS instrument at a flow rate of 0.3 μ L per minute through a 15 μ m fused silica emitter tip (New Objective). Samples were first analyzed using the orbitrap detector (FT-MS scan) at 7.5k or 15k resolution over a 600 to 2000 m/z range. The AGC target was set to 1 e 6, tube lens and capillary voltages were 200 and 60 Volts respectively. The needle voltage was 2000 Volts. MS/MS spectra were acquired either manually in Tune mode or by running a Data Dependent Acquisition (DDA) method identical to that used for proteomics (see below) except without LC separation. As such, MS/MS spectra were acquired for peptides visible in the FT-MS spectra. These MS/MS spectra were searched against the Swiss-Prot proteome (Jan, 2014 version, 542258 sequences) using MASCOT v 2.4.1 (Matrix Science). For the database search the MS and MS/MS tolerance was 3ppm and 0.5 Da respectively, no enzyme was specified and oxidized Met and phosphorylated Ser/Thr were chosen as variable modifications.

4.5.10 In vitro protein-protein interactions

Mature PES1 and CCD4 were cloned into pCR8 topo vector and ligated into the pGEX5-1 expression vector using *EcoRI* and *XhoI* sites. Recombinant proteins were expressed in *E. coli* BL21 cells by incubation at 22⁰C with IPTG for 5 hours; solubilized proteins were purified by using glutathione resin at 4⁰C. The purified proteins were dialyzed by using dialysis cassette (slide-A-Lyzer, Thermo scientific)

against 100 mM NaCl and 50 mM Tris-HCl (pH8) and 10% glycerol. After dialysis, proteins were concentrated using microcon centrifugal filter units (Millipore). To analyze interactions between PGM48 and PES1 or CCD4, bait and prey proteins were incubated at room temperature for 2-3 hours and then resin was added to the mixture and incubated for another 1 hour. Resin was washed by washing buffer for 5 times and finally resin bound protein/s was eluted by elution buffer (1.5% SDS, 50 mM Tris-HCl (pH 8), glycerol 15%) by heating at 75⁰C for 5 min. Interaction was visualized by SDS-PAGE gel and staining with silver nitrate.

4.5.11 Proteomics, MS and display in PPDB and submission to public data repository

For protein identification and quantification, each gel lane was cut in consecutive gel slices, followed by in-gel digestion using trypsin and subsequent peptide extraction as described previously (Friso et al., 2011). Peptide extracts for each gel band were then analyzed by on-line nanoLC-MS/MS using an LTQ-Orbitrap (Thermo). Resulting spectral data were searched against the predicted Arabidopsis proteome (TAIR10), including a small set of typical contaminants and the decoy, as described in (Nishimura et al., 2013). Only proteins with 2 or more matched spectra were considered. Protein abundances were quantified according to the number of matched adjusted MS/MS spectra (AdjSPC) as explained in (Friso et al., 2011). MS-derived information, as well as annotation of protein name, location and function for the identified proteins, can be found in the Plant Proteome Database (<http://ppdb.tc.cornell.edu>). MS-derived information, as well as annotation of protein name, location and function for the identified proteins can be found in the PPDB (<http://ppdb.tc.cornell.edu/>). The MS proteomics data have been deposited to the ProteomeXchange Consortium (Vizcaino et al., 2014) via the PRIDE partner repository (<http://www.ebi.ac.uk/pride>) with the dataset identifier PXD003684 and 10.6019/PXD003684.

ACKNOWLEDGMENTS This research was supported by a grant from the National Science Foundation MCB1021963 to KJVW

SUPPLEMENTAL INFORMATION

Supplemental Dataset 4.1. Sequence alignment used for the phylogenetic analysis displayed in Figure 2 (Text file)

Supplemental Dataset 4.2. The relative accumulation levels for all observed PG core proteins of wt after 6 (bolting stage) and 12 (advanced senescence) weeks growth under short day conditions (10 h light/14h dark) (Excell file).

Supplemental Dataset 4.3. PG proteomes of wt, RNAi and OE line. Plants were grown under long day conditions (18h light / 6h dark) for 35 days and PG were purified after extraction from isolated thylakoid membranes (Excell File)

Supplemental Table 4.1. Mass spectrometry analysis of β -casein digests with recombinant wt PGM48.

Supplemental Table 4.2. M48 homologs in the four clades in species across the tree-of-life

Supplemental Table 4.3. Primers used in this study.

Supplemental Figure 4.1. N-terminus of Arabidopsis PGM48 and cTP predictions for Arabidopsis and homologs in other dicots and monocots.

Supplemental Figure 4.2. MS spectra of digests by active PGM48 and mutant inactive PGM48 (PGM48-H191A).

Supplemental Figure 4.3. Phylogenetic tree of clade M48D with PGM48 homologs.

Supplemental Figure 4.4. Characterization of T-DNA insertion mutants in *PGM48* and PGM48 is a senescence induced protein in Arabidopsis

Supplemental Figure 4.5. Additional phenotypic analysis of RNAi and OE lines

Supplemental Figure 4.6. mRNA expression data from publicly available microarray data for leaves from wt (col-0) for all genes encoding for the PG core proteome and for ZEP.

LITERATURE CITED

- Akiyama Y** (2009) Quality control of cytoplasmic membrane proteins in *Escherichia coli*. *J Biochem* 146: 449-454
- Albrecht-Borth V, Kauss D, Fan D, Hu Y, Collinge D, Marri S, Liebers M, Apel K, Pfannschmidt T, Chow WS, Pogson BJ** (2013) A novel proteinase, SNOWY COTYLEDON4, is required for photosynthetic acclimation to higher light intensities in *Arabidopsis*. *Plant Physiol* 163: 732-745
- Ast T, Michaelis S, Schuldiner M** (2016) The Protease Ste24 Clears Clogged Translocons. *Cell* 164: 103-114
- Austin JR, 2nd, Frost E, Vidi PA, Kessler F, Staehelin LA** (2006) Plastoglobules are lipoprotein subcompartments of the chloroplast that are permanently coupled to thylakoid membranes and contain biosynthetic enzymes. *Plant Cell* 18: 1693-1703
- Avendano-Vazquez AO, Cordoba E, Llamas E, San Roman C, Nisar N, De la Torre S, Ramos-Vega M, Gutierrez-Nava MD, Cazzonelli CI, Pogson BJ, Leon P** (2014) An Uncharacterized Apocarotenoid-Derived Signal Generated in zeta-Carotene Desaturase Mutants Regulates Leaf Development and the Expression of Chloroplast and Nuclear Genes in *Arabidopsis*. *Plant Cell*
- Avila-Ospina L, Moison M, Yoshimoto K, Masclaux-Daubresse C** (2014) Autophagy, plant senescence, and nutrient recycling. *J Exp Bot* 65: 3799-3811
- Besagni C, Kessler F** (2013) A mechanism implicating plastoglobules in thylakoid disassembly during senescence and nitrogen starvation. *Planta* 237: 463-470
- Bhuiyan NH, Friso G, Poliakov A, Ponnala L, van Wijk KJ** (2015) MET1 is a thylakoid-associated TPR protein involved in photosystem II supercomplex formation and repair in *Arabidopsis*. *Plant Cell* 27: 262-285
- Bohovych I, Fernandez MR, Rahn JJ, Stackley KD, Bestman JE, Anandhan A, Franco R, Claypool SM, Lewis RE, Chan SS, Khalimonchuk O** (2015) Metalloprotease OMA1 Fine-tunes Mitochondrial Bioenergetic Function and Respiratory Supercomplex Stability. *Sci Rep* 5: 13989
- Bracha-Drori K, Shichrur K, Lubetzky TC, Yalovsky S** (2008) Functional analysis of *Arabidopsis* postprenylation CaaX processing enzymes and their function in subcellular protein targeting. *Plant Physiol* 148: 119-131

- Bracha K, Lavy M, Yalovsky S** (2002) The Arabidopsis AtSTE24 Is a CAAAX Protease with Broad Substrate Specificity. *Journal of Biological Chemistry* 277: 29856-29864
- Breeze E, Harrison E, McHattie S, Hughes L, Hickman R, Hill C, Kiddle S, Kim YS, Penfold CA, Jenkins D, Zhang C, Morris K, Jenner C, Jackson S, Thomas B, Tabrett A, Legaie R, Moore JD, Wild DL, Ott S, Rand D, Beynon J, Denby K, Mead A, Buchanan-Wollaston V** (2011) High-resolution temporal profiling of transcripts during Arabidopsis leaf senescence reveals a distinct chronology of processes and regulation. *Plant Cell* 23: 873-894
- Carrion CA, Costa ML, Martinez DE, Mohr C, Humbeck K, Guamet JJ** (2013) In vivo inhibition of cysteine proteases provides evidence for the involvement of 'senescence-associated vacuoles' in chloroplast protein degradation during dark-induced senescence of tobacco leaves. *J Exp Bot* 64: 4967-4980
- Chen QG, Bleecker AB** (1995) Analysis of ethylene signal-transduction kinetics associated with seedling-growth response and chitinase induction in wild-type and mutant arabidopsis. *Plant Physiol* 108: 597-607
- Eugeni-Piller L, Besagni C, Ksas B, Rumeau D, Brehelin C, Glauser G, Kessler F, Havaux M** (2011) Chloroplast lipid droplet type II NAD(P)H quinone oxidoreductase is essential for prenylquinone metabolism and vitamin K1 accumulation. *Proc Natl Acad Sci U S A* 108: 14354-14359
- Fatihi A, Latimer S, Schmollinger S, Block A, Dussault PH, Vermaas WF, Merchant SS, Basset GJ** (2015) A Dedicated Type II NADPH Dehydrogenase Performs the Penultimate Step in the Biosynthesis of Vitamin K1 in Synechocystis and Arabidopsis. *Plant Cell* 27: 1730-1741
- 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, Rolland N** (2010) AT_CHLORO, a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins. *Mol Cell Proteomics* 9: 1063-1084
- Finkemeier I, Laxa M, Miguet L, Howden AJ, Sweetlove LJ** (2011) Proteins of diverse function and subcellular location are lysine acetylated in Arabidopsis. *Plant Physiol* 155: 1779-1790
- Friso G, Olinas PDB, van Wijk KJ** (2011) The Workflow for Quantitative Proteome Analysis of Chloroplast Development and Differentiation, Chloroplast Mutants, and Protein Interactions by Spectral Counting. *In RP*

Jarvis, ed, Chloroplast Research in Arabidopsis, Vol 775. Humana Press, New York, pp 265-282

- Gan S, Amasino RM** (1995) Inhibition of leaf senescence by autoregulated production of cytokinin. *Science* 270: 1986-1988
- Gaude N, Brehelin C, Tischendorf G, Kessler F, Dormann P** (2007) Nitrogen deficiency in Arabidopsis affects galactolipid composition and gene expression and results in accumulation of fatty acid phytyl esters. *Plant J* 49: 729-739
- Ghosh S, Hudak KA, Dumbroff EB, Thompson JE** (1994) Release of Photosynthetic Protein Catabolites by Blebbing from Thylakoids. *Plant Physiol* 106: 1547-1553
- Gonzalez-Jorge S, Ha SH, Magallanes-Lundback M, Gilliland LU, Zhou A, Lipka AE, Nguyen YN, Angelovici R, Lin H, Cepela J, Little H, Buell CR, Gore MA, Dellapenna D** (2013) CAROTENOID CLEAVAGE DIOXYGENASE4 Is a Negative Regulator of beta-Carotene Content in Arabidopsis Seeds. *Plant Cell* 25: 4812-4826
- Grefen C, Lalonde S, Obrdlik P** (2007) Split-ubiquitin system for identifying protein-protein interactions in membrane and full-length proteins. *Curr Protoc Neurosci* Chapter 5: Unit 5 27
- Guamet JJ, Pichersky E, Nooden LD** (1999) Mass exodus from senescing soybean chloroplasts. *Plant Cell Physiol* 40: 986-992
- Gutierrez-Carbonell E, Takahashi D, Lattanzio G, Rodriguez-Celma J, Kehr J, Soll J, Philippar K, Uemura M, Abadia J, Lopez-Millan AF** (2014) The distinct functional roles of the inner and outer chloroplast envelope of Pea (*Pisum sativum*) as revealed by proteomic approaches. *J Proteome Res* 13: 2941-2953
- Hortensteiner S** (2009) Stay-green regulates chlorophyll and chlorophyll-binding protein degradation during senescence. *Trends Plant Sci* 14: 155-162
- Hortensteiner S, Krautler B** (2011) Chlorophyll breakdown in higher plants. *Biochim Biophys Acta* 1807: 977-988
- Hou X, Rivers J, Leon P, McQuinn RP, Pogson BJ** (2016) Synthesis and Function of Apocarotenoid Signals in Plants. *Trends Plant Sci* 21: 792-803
- Huang FC, Molnar P, Schwab W** (2009) Cloning and functional characterization of carotenoid cleavage dioxygenase 4 genes. *J Exp Bot* 60: 3011-3022

- Huang M, Friso G, Nishimura K, Qu X, Olinares PD, Majeran W, Sun Q, van Wijk KJ** (2013) Construction of plastid reference proteomes for maize and Arabidopsis and evaluation of their orthologous relationships; the concept of orthoproteomics. *J Proteome Res* 12: 491-504
- Ishida H, Izumi M, Wada S, Makino A** (2014) Roles of autophagy in chloroplast recycling. *Biochim Biophys Acta* 1837: 512-521
- Klodmann J, Senkler M, Rode C, Braun HP** (2011) Defining the protein complex proteome of plant mitochondria. *Plant Physiol* 157: 587-598
- Korwitz A, Merkwirth C, Richter-Dennerlein R, Troder SE, Sprenger HG, Quiros PM, Lopez-Otin C, Rugarli EI, Langer T** (2016) Loss of OMA1 delays neurodegeneration by preventing stress-induced OPA1 processing in mitochondria. *J Cell Biol* 212: 157-166
- Kusaba M, Tanaka A, Tanaka R** (2013) Stay-green plants: what do they tell us about the molecular mechanism of leaf senescence. *Photosynth Res* 117: 221-234
- Lee PT, Hsu AY, Ha HT, Clarke CF** (1997) A C-methyltransferase involved in both ubiquinone and menaquinone biosynthesis: isolation and identification of the Escherichia coli ubiE gene. *J Bacteriol* 179: 1748-1754
- Li Z, Peng J, Wen X, Guo H** (2013) Ethylene-insensitive3 is a senescence-associated gene that accelerates age-dependent leaf senescence by directly repressing miR164 transcription in Arabidopsis. *Plant Cell* 25: 3311-3328
- Liberton M, Saha R, Jacobs JM, Nguyen AY, Gritsenko MA, Smith RD, Koppelaar DW, Pakrasi HB** (2016) Global Proteomic Analysis Reveals an Exclusive Role of Thylakoid Membranes in Bioenergetics of a Model Cyanobacterium. *Mol Cell Proteomics* 15: 2021-2032
- Lichtenthaler HK** (1968) Plastoglobuli and the fine structure of plastids. *Endeavor* 27: 144-149
- Lippold F, vom Dorp K, Abraham M, Hapzl G, Wewer V, Yilmaz JL, Lager I, Montandon C, Besagni CI, Kessler F, Stymne S, Daqurmman P** (2012) Fatty Acid Phytol Ester Synthesis in Chloroplasts of Arabidopsis. *The Plant Cell* 24: 2001-2014
- Lundquist P, Poliakov A, Bhuiyan NH, Zybaïlov B, Sun Q, van Wijk KJ** (2012) The functional network of the Arabidopsis thaliana plastoglobule proteome based on quantitative proteomics and genome-wide co-expression analysis. *Plant Physiol* 58: 1172-1192

- Lundquist P, Poliakov A, Giacomelli L, Friso G, Appel M, McQuinn RP, Krasnoff SB, Rowland O, Ponnala L, Sun Q, van Wijk KJ** (2013) Loss of plastoglobule-localized kinases ABC1K1 and ABC1K3 leads to a conditional degreening phenotype, a modified prenyl-lipid composition and recruitment of JA biosynthesis. *Plant Cell* 25: 1818-1839
- Lundquist PK, Davis JL, van Wijk KJ** (2012) ABC1K atypical kinases in plants: filling the organellar kinase void. *Trends Plant Sci* 17: 546-555
- Manandhar-Shrestha K, Tamot B, Pratt EP, Saitie S, Brautigam A, Weber AP, Hoffmann-Benning S** (2013) Comparative proteomics of chloroplasts envelopes from bundle sheath and mesophyll chloroplasts reveals novel membrane proteins with a possible role in c4-related metabolite fluxes and development. *Front Plant Sci* 4: 65
- Martinez DE, Costa ML, Guamet JJ** (2008) Senescence-associated degradation of chloroplast proteins inside and outside the organelle. *Plant Biol (Stuttg)* 10 Suppl 1: 15-22
- Mattoo AK, Edelman M** (1987) Intramembrane translocation and posttranslational palmitoylation of the chloroplast 32-kDa herbicide-binding protein. *Proc Natl Acad Sci U S A* 84: 1497-1501
- Miller JD, Arteca RN, Pell EJ** (1999) Senescence-associated gene expression during ozone-induced leaf senescence in Arabidopsis. *Plant Physiol* 120: 1015-1024
- Nierzwicki-Bauer SA, Balkwill DL, Stevens SE, Jr.** (1983) Three-dimensional ultrastructure of a unicellular cyanobacterium. *J Cell Biol* 97: 713-722
- Nishimura K, Asakura Y, Friso G, Kim J, Oh SH, Rutschow H, Ponnala L, van Wijk KJ** (2013) ClpS1 is a conserved substrate selector for the chloroplast Clp protease system in Arabidopsis. *Plant Cell* 25: 2276-2301
- Oldfield CJ, Dunker AK** (2014) Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem* 83: 553-584
- Otegui MS, Noh YS, Martinez DE, Vila Petroff MG, Staehelin LA, Amasino RM, Guamet JJ** (2005) Senescence-associated vacuoles with intense proteolytic activity develop in leaves of Arabidopsis and soybean. *Plant J* 41: 831-844
- Parmryd I, Andersson B, Dallner G** (1999) Protein prennylation in spinach chloroplasts. *Proc Natl Acad Sci U S A* 96: 10074-10079

- Parmryd I, Shipton CA, Swiezewska E, Dallner G, Andersson B** (1997) Chloroplastic prenylated proteins. *FEBS Lett* 414: 527-531
- Peramuna A, Summers ML** (2014) Composition and occurrence of lipid droplets in the cyanobacterium *Nostoc punctiforme*. *Arch Microbiol* 196: 881-890
- Porfirova S, Bergmuller E, Tropsch S, Lemke R, Dormann P** (2002) Isolation of an *Arabidopsis* mutant lacking vitamin E and identification of a cyclase essential for all tocopherol biosynthesis. *PNAS* 99: 12495-12500
- Porra RJ, Thompson, W.A., and Kriedemann, P.E.** (1989) Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls *a* and *b* extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochim. Biophys. Acta* 975: 384-394
- Pryor EE, Jr., Horanyi PS, Clark KM, Fedoriw N, Connelly SM, Koszelak-Rosenblum M, Zhu G, Malkowski MG, Wiener MC, Dumont ME** (2013) Structure of the integral membrane protein CAAX protease Ste24p. *Science* 339: 1600-1604
- Quigley A, Dong YY, Pike AC, Dong L, Shrestha L, Berridge G, Stansfeld PJ, Sansom MS, Edwards AM, Bountra C, von Delft F, Bullock AN, Burgess-Brown NA, Carpenter EP** (2013) The structural basis of ZMPSTE24-dependent laminopathies. *Science* 339: 1604-1607
- Rawlings ND, Barrett AJ, Finn R** (2016) Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* 44: D343-350
- Rottet S, Besagni C, Kessler F** (2015) The role of plastoglobules in thylakoid lipid remodeling during plant development. *Biochim Biophys Acta* 1847: 889-899
- Rowland E, Kim J, Bhuiyan NH, van Wijk KJ** (2015) The *Arabidopsis* Chloroplast Stromal N-Terminome: Complexities of Amino-Terminal Protein Maturation and Stability. *Plant Physiol* 169: 1881-1896
- Rubio A, Rambla JL, Santaella M, Gomez MD, Orzaez D, Granell A, Gomez-Gomez L** (2008) Cytosolic and plastoglobule-targeted carotenoid dioxygenases from *Crocus sativus* are both involved in beta-ionone release. *J Biol Chem* 283: 24816-24825
- Sakuraba Y, Park SY, Paek NC** (2015) The Divergent Roles of STAYGREEN (SGR) Homologs in Chlorophyll Degradation. *Mol Cells* 38: 390-395

- Schippers JH** (2015) Transcriptional networks in leaf senescence. *Curr Opin Plant Biol* 27: 77-83
- Schippers JH, Schmidt R, Wagstaff C, Jing HC** (2015) Living to Die and Dying to Live: The Survival Strategy behind Leaf Senescence. *Plant Physiol* 169: 914-930
- Simm S, Papatiriu DG, Ibrahim M, Leisegang MS, Muller B, Schorge T, Karas M, Mirus O, Sommer MS, Schleiff E** (2013) Defining the core proteome of the chloroplast envelope membranes. *Front Plant Sci* 4: 11
- Singh DK, McNellis TW** (2011) Fibrillin protein function: the tip of the iceberg? *Trends in Plant Science* 16: 432-441
- Tevini M, Steinmuller D** (1985) Composition and Function of Plastoglobuli II. Lipid-Composition of Leaves and Plastoglobuli During Beech Leaf Senescence. *Planta* 163: 91-96
- Tuquet C, Newman DW** (1980) Aging and regreening in soybean cotyledons. I Ultrastructural changes in plastids and plastoglobuli. *Cytobios* 29: 43-59
- van de Meene AM, Hohmann-Marriott MF, Vermaas WF, Roberson RW** (2006) The three-dimensional structure of the cyanobacterium *Synechocystis* sp. PCC 6803. *Arch Microbiol* 184: 259-270
- van Wijk KJ** (2015) Protein maturation and proteolysis in plant plastids, mitochondria, and peroxisomes. *Annu Rev Plant Biol* 66: 75-111
- van Wijk KJK, F.** (2016) Plastoglobuli: Plastid Microcompartments with Integrated Functions in Metabolism, Plastid Developmental Transitions, and Environmental Adaptation. *Annual Review Plant Biology* 68
- Vidi PA, Kanwischer M, Baginsky S, Austin JR, Csucs G, Dormann P, Kessler F, Brehelin C** (2006) Tocopherol cyclase (VTE1) localization and vitamin E accumulation in chloroplast plastoglobule lipoprotein particles. *J Biol Chem* 281: 11225-11234
- Vizcaino JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Rios D, Dienes JA, Sun Z, Farrah T, Bandeira N, Binz PA, Xenarios I, Eisenacher M, Mayer G, Gatto L, Campos A, Chalkley RJ, Kraus HJ, Albar JP, Martinez-Bartolome S, Apweiler R, Omenn GS, Martens L, Jones AR, Hermjakob H** (2014) ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol* 32: 223-226

- Wang XQ, Yang PF, Liu Z, Liu WZ, Hu Y, Chen H, Kuang TY, Pei ZM, Shen SH, He YK** (2009) Exploring the mechanism of *Physcomitrella patens* desiccation tolerance through a proteomic strategy. *Plant Physiol* 149: 1739-1750
- Wegener KM, Singh AK, Jacobs JM, Elvitigala T, Welsh EA, Keren N, Gritsenko MA, Ghosh BK, Camp DG, 2nd, Smith RD, Pakrasi HB** (2010) Global proteomics reveal an atypical strategy for carbon/nitrogen assimilation by a cyanobacterium under diverse environmental perturbations. *Mol Cell Proteomics* 9: 2678-2689
- Wiederstein M, Sippl MJ** (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35: W407-410
- Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y** (2015) The I-TASSER Suite: protein structure and function prediction. *Nat Methods* 12: 7-8
- Ytterberg AJ, Peltier JB, van Wijk KJ** (2006) Protein profiling of plastoglobules in chloroplasts and chromoplasts; a surprising site for differential accumulation of metabolic enzymes. *Plant Physiol* 140: 984-997
- Zbierzak AM, Kanwischer M, Wille C, Vidi PA, Giavalisco P, Lohmann A, Briesen I, Porfirova S, Brehelin C, Kessler F, Dormann P** (2009) Intersection of the tocopherol and plastoquinol metabolic pathways at the plastoglobule. *Biochemical Journal* 425: 389-399
- Zhu C, Bai C, Sanahuja G, Yuan D, Farre G, Naqvi S, Shi L, Capell T, Christou P** (2010) The regulation of carotenoid pigmentation in flowers. *Arch Biochem Biophys* 504: 132-141

APPENDIX

ARABIDOPSIS CHLOROPLAST LOCALIZED GLUTAMYL- ENDOPEPTIDASE CGEP¹

5.1 ABSTRACT

Chloroplast localized glutamyl-endorpeptidase (CGEP) is an abundant plant protease of unknown function localized to the chloroplast stroma. CGEP has been purified from a variety of angiosperms and was shown to cleave small peptides C-terminal to glutamate (Glu) *in vitro*. Here I report the characterization of *Arabidopsis thaliana* CGEP with an emphasis on *in vivo* function. Recombinant CGEP was shown to be an active serine endo-protease, degrading the protein β -casein. Additionally, CGEP has autocatalytic activity that allows a 15 residue C-terminal peptide to be cleaved. Three dimensional structural modeling allowed us to make predictions about substrate recognition and showed that the unprocessed CGEP C-terminus is in close proximity of the active site. We compared the soluble chloroplast proteomes of wild type (wt) with that of a CGEP loss-of-function mutant (*cgep-5*) to determine the molecular phenotype and identify putative substrates. Ribosomal subunits strongly over-

¹Manuscript in preparation for submission to Plant Physiology (2017) **Nazmul Bhuiyan, Elden Rowland, Giulia Friso and Klaas J. van Wijk**

The contribution of the thesis' author to this work consisted of proteomic identification of cleavage sites (PICS), 3D structural modeling and N-terminal proteomics. Label free spectral counting was performed by GF. All other experiments and analysis by NB. ER wrote this first draft of the Manuscript.

accumulated in the mutant. N-terminal proteomics identified the C-terminal domain of TIC110 and a CP12 degradation product as potential substrates for CGEP.

5.2 INTRODUCTION

The alpha/beta-hydrolase fold is present in a large group (Clan SC in MEROPS protease database) of enzymes with diverse biological functions. These enzymes possess an Asp-His-Ser catalytic triad that catalyzes the hydrolysis of esters, lipids and proteins (Rea and Fulop, 2006). Within this group is a family of serine proteases named prolyl oligopeptidases (POPs) that perform a range of proteolytic functions. The POP family (S9 in MEROPS) is widely distributed across the tree of life (Venalainen et al., 2004) and its members are distinct from other well-known serine class proteases such as trypsin. POPs are generally ~100 kDa in molecular mass and are limited to the cleavage of small peptides and peptide hormones up to about ~30 residues in length (Rea and Fulop, 2011; Van Elzen and Lambeir, 2011). Interestingly, its members include exopeptidases limited to the cleavage of amino, di or tri-peptides from the ends of peptides, and endopeptidase that are able to cleave internal polypeptide bonds of proteins. Furthermore, they have broad substrate cleavage specificity being able to cleave between acidic, basic or hydrophobic amino acids including proline, as the name suggests (Rea and Fulop, 2006).

Crystal structures have been solved for a number of POPs (S9 peptidase family members). Typically they contain an α/β -hydrolase domain, that contains the catalytic triad and an adjacent β -propeller domain. The two domains form bowl-like structures that close together to form a large central cavity, restricting access to the catalytic site

(Rea and Fulop, 2006). Little is known about the dynamics of substrate binding such as how the peptidase opens to receive peptide substrates, or if closing is a requirement for peptidase activity. *Porcine* POP has an internal cavity of $\sim 8500 \text{ \AA}^3$ suggesting an upper peptide size limit of ~ 30 residues. Open and closed structures were solved for two bacterial POPs (Shan et al., 2005) indicating that the peptidase may open to receive substrates. Hydrogen-deuterium exchange mass spectrometry (MS) was used to probe structural dynamics of a human S9A. The authors suggest various modes of substrate gating. Interestingly, they point out that there is 4 \AA pore going through the center of the beta propeller domain, said to be too small for substrates without a big shift in structure. Their results support a loop side opening model (Tsirigotaki et al., 2017).

Plants and algae contain a unique subfamily of POPs, classified in the MEROPS as S9D, termed chloroplast glutamyl-endopeptidase (CGEP) (Rawlings et al., 2012). Bacterial homologues suggest this subfamily is of bacterial origin. Bacterial POP analysis shows some sequences with an active site motif identical to that of *Arabidopsis* CGEP (Kaushik and Sowdhamini, 2014). CGEP proteases have been purified from the leaves of spinach, cucumber and pea by consecutive rounds of orthogonal chromatography and each purified peptidase was shown to cleave synthetic peptides or recombinant protein substrates C-terminal to Glu (P1 position as per (Schechter and Berger, 1967)) (Laing and Christeller, 1997; Yamauchi et al., 2001; Forsberg et al., 2005). Short stretches of the protein sequences from cucumber and pea were identified by MS that had high similarity with the *Arabidopsis* protein, At2G46390. Laing *et al* showed the protease purified from spinach chloroplasts,

efficiently cleaved the synthetic peptide Z-Leu-Leu-Glu-βNA, but none of the other nine synthetic peptides tested containing a range of other amino acids including Asp, that has similar acidic character to Glu. Oxidized insulin β chain was digested after overnight incubation and the fragments corresponded to cleavage after Glu (P1 position). Neither RUBICO, RNase nor the protein casein was digested. The ~100 kDa protease had an optimum pH of 8.0, was only inhibited by serine peptidase inhibitors and formed a homo-tetramer in solution (Laing and Christeller, 1997). Yamauchi *et al* showed that cucumber CGEP had the same cleavage specificity (using synthetic peptides) and oligomeric state. Additionally, they suggest that the peptidase is normally inhibited *in vivo* because they could not detect the activity in crude extracts. Three unidentified low molecular weight proteins were purified and shown to be specific inhibitors to CGEP, not inhibiting other commercial peptidases (Yamauchi *et al.*, 2001). The pea CGEP studied by Forsberg *et al* showed similar biochemical properties except it was reported as a 130 kDa monomer in solution. CGEP's ability to degrade the recombinant N-terminal domain (69 residues) of pea LHCII 1.1 was tested. Six proteolytic products were identified by MS and Edman sequencing ranging in length from 13 to 38 residues; for only one of the five Glu residues was no cleavage observed², likely because the fragment was too small for detection by MS (Forsberg *et al.*, 2005).

² N-terminal domain of pea LHCII 1.1. All but the last Glu was cleaved *in vitro* (Forsberg *et al.*, 2005).

KSATTKKVASSGSPWYGPDRVKYLGPFSG**E**SPSYLTG**E**FPGDYGWDTAGLSA
DP**E**TFSKN**R**E**L**EVIHSR

Our lab identified a homologue of Arabidopsis CGEP in maize that was highly enriched in bundle-sheath cells (surrounding the vasculature) as compared to mesophyll cells (Majeran and van Wijk, 2009). Quantitative label-free MS analysis showed that CGEP is abundant in the chloroplasts of rice, maize and Arabidopsis (van Wijk, 2015; Nishimura et al., 2017). Here we report a detailed structural and functional analysis of Arabidopsis CGEP with emphasis on its *in vivo* activities. The substrate cleavage specificity was determined at high resolution using proteomic identification of cleavage sites (PICS) (Schilling et al., 2011) with a peptide library derived from the soluble Arabidopsis proteome. We revealed that CGEP has autocatalytic activity *in vitro* and *in vivo* that allows a 15 residue C-terminal extension to be removed. Three dimensional structural modeling facilitated predictions about substrate recognition and showed that the protein C-terminus is in close proximity of the active site. We compared the soluble chloroplast proteomes of wild type (wt) with that of a CGEP loss-of-function mutant (*cgep-5*) to determine the molecular phenotype and identify putative substrates. N-terminal proteomics for the same material identified the C-terminal domain of TIC110 and a CP12 degradation product as potential substrates for CGEP.

5.3 RESULTS

5.3.1 *Arabidopsis CGEP is an active peptidase localized to the chloroplast stroma*

Recombinant CGEP with an N-terminal GST tag (GST-CGEP) was expressed in *E. coli* and purified over glutathione-agarose beads. The protease was shown to almost completely digest the 25 kDa protein β -casein in contrast to previous

suggestions that CGEP could not degrade whole proteins (Figure 5.1 A). Inhibitor studies confirmed that CGEP is a serine protease. Mutation of the catalytic Ser to Arg (S781R, GST-SDM1) abolished its proteolytic activity towards β -casein (Figure 5.1 B). Interestingly there was a slight increase in the molecular weight of CGEP upon mutation of the catalytic site that seemed too large to be caused by a single point mutation. Therefore we performed in-gel digestion and MS analysis of wt GST-CGEP and GST-SDM1. This revealed a C-terminal truncation in wt that was absent in SDM-1 (Figure 5.2 A).

CGEP is an abundant chloroplast protein and therefore has been detected many times in our comparative proteomic studies of the Arabidopsis chloroplast, compiled in the Plant Proteome Database (PPDB). When proteins are digested with trypsin, as part of the proteomic workflow, information concerning *in vivo* N- and C-terminal protein processing is often obscured (See Chapters 1 and 2). However, semitryptic database searches can identify non-tryptic cleavage events that result from endogenous protease activity. As such we were able to identify the mature N-terminus that results from cleavage of the chloroplast transit peptide (cTP) and a C-terminal peptide that corresponds exactly to the C-terminal processing we observed in *in vitro* (Figure 5.2A,B).

Mutation of C-terminal Glu residues in CGEP (SDM2 and SDM3) showed that CGEP was only sensitive to the E946; AE949A; E951A mutations, but not E928A and D931A, suggesting that only these latter glutamate residues are accessible to the active site (Figure 5.2A). Addition of DTT (5mM) and different salt concentrations (100-500mM) did not affect CGEP activity (data not shown).

A

Sample	C-terminal region of recombinant CGEP (peptides detected in red)	C-term. StrepII tag <i>in vivo</i>
Wild type	...LWETDRWLQKYCVPNTSDADTSPDQSKEGSDSADKVSTGTGGGNPEFGEHEVHSKLRRSLL	
SDM1	...LWETDRWLQKYCVPNTSDADTSPDQSKEGSDSADKVSTGTGGGNPEFGEHEVHSKLRRSLL	yes
SDM2	...LWETDRWLQKYCVPNTSDADTSPDQSKAGSASADKVSTGTGGGNPEFGEHEVHSKLRRSLL	
SDM3	...LWETDRWLQKYCVPNTSDADTSPDQSKEGSDSADKVSTGTGGGNPAFGAHAVHSKLRRSLL	yes

B

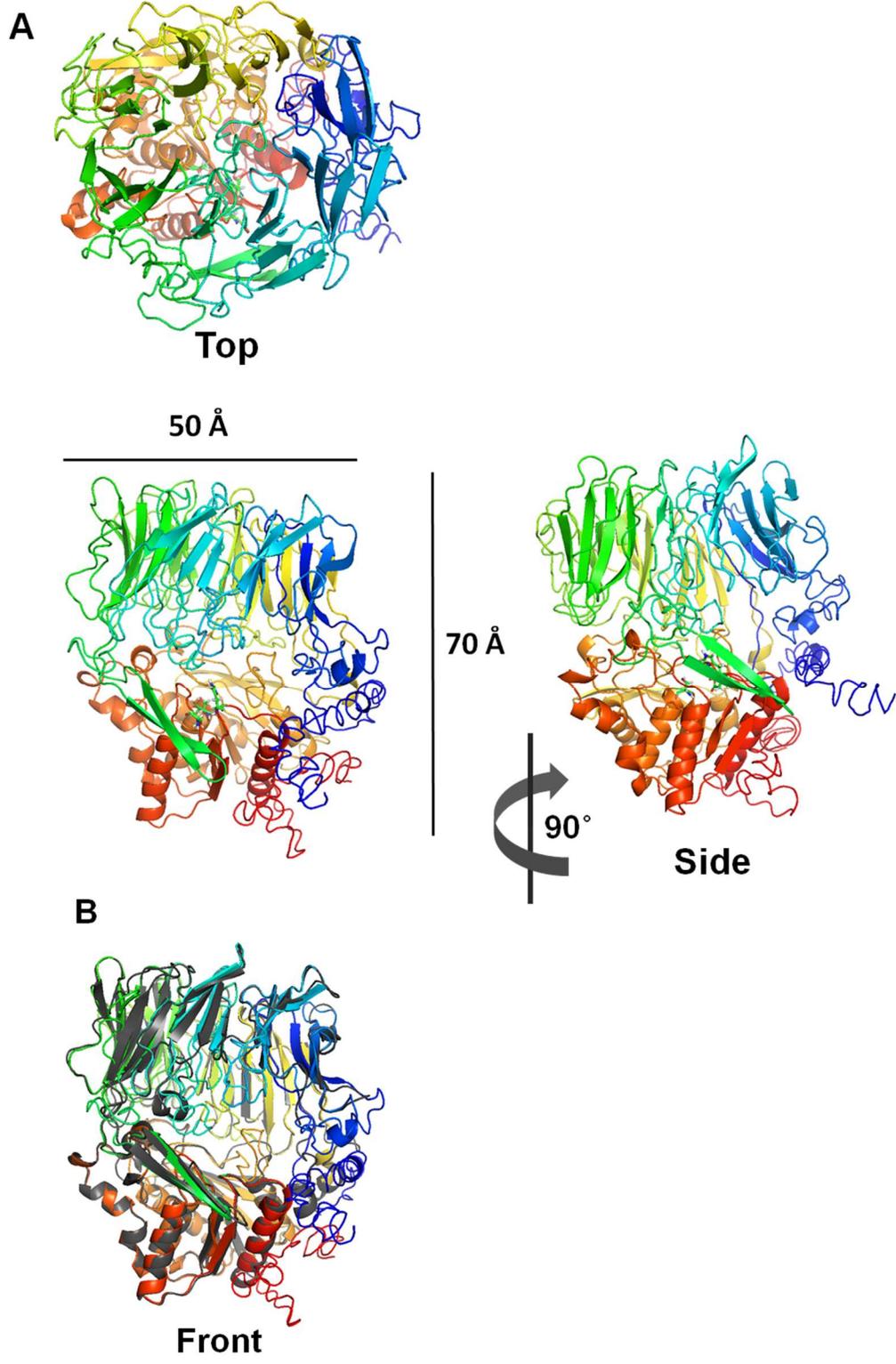
MMRFHKACHRFSLSPCHLSPPSPSPASSLLLLPKLSGFSTLSTRRCVRRRFSENPLTT
 VMASRSASRLRSLASACSGGAEDGGGTSNGLSASATATEDDELAIGTYRLPPEIRDI
 VDAPPVPALESFSPHRDKILFLKRRALPPLADLARPEEKLAGVRIDGYCNTRSRMSFYTGL
 GIHQLLPDGTLSPKEITGIPDGGKINFVTSNDGKHLAFSIRVDENGNSSKPVVWVADV
 ETGVARPLFNSQDIFLNAIFESFVWIDNSTLLVSTIPSSRGEPPKPLVPSGPKTLSNET
 KTVVQVRTFQDLLKDEYDADLFDYYASSQLVLAASLDGTVKEVGVPAVYTSLDPSTDHKYL
 LVSSLHRPYSFIVPCGRFPKKEVWTTDGRFVRQLCDLPLAEDIPIASNSVRKGMRSINW
 RADKPSTLYWAETQDGGDAKMEVSPRDIVYMQSAEPLAGEEPEVLHKLDLRYGGISWCDD
 TLALVYESWYKTRRTRTWVISPGSNDVSPRILFDRSSEDVYSDPGSTMLRRTDAGTYVIA
 KIKKENDEGTYYLLNGSGATPQGNVPFLDLFDINTGNKERIWESDKEKYFETVVALMSDQ
 KEGDLKMEELKILTSKESKTENTQYSLQLWPDRKVQQITNFFHPYPQLASLQKEMIRYQR
 KDGVLQATLYLPPGYDPSKDGPLPCLFWSYPGEFKSKDAAGQVRGSPNEFAGIGSTSAL
 LWLARRFAILSGPTIPIIGEGDEEANDRYVEQLVASAEAAVEEVVRRGVADRSKIAVGGH
 SYGAFMTANLLAHAPHLFACGIARSGAYNRTLTPFGFQNEEDRTLWEATNVYVEMSPFMSA
 NKIKKPILLIHGEEDNPNGLTQMQRFFNALKGHGALCRLVVLPHESHGYSARESIMHV
 LWETDRWLQKYCVPNTSDADTSPDQSKEGSDSADKVSTGTGGGNPEFGEHEVHSKLRRSLL

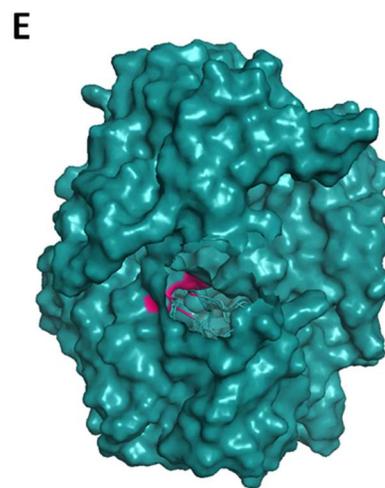
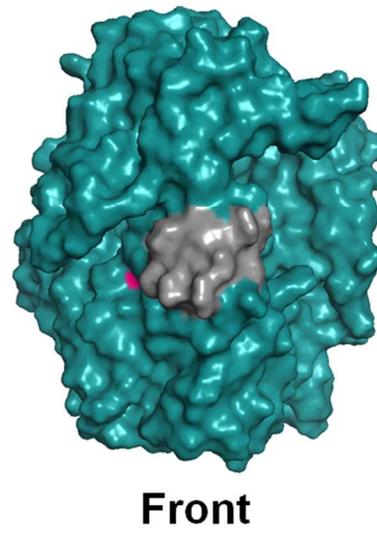
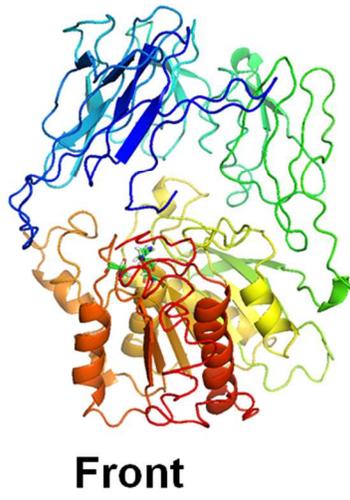
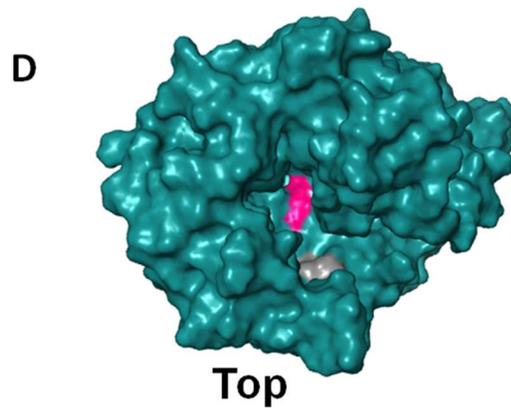
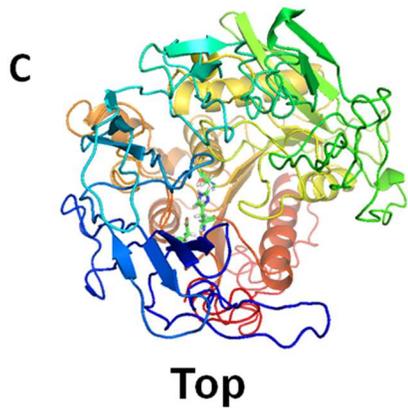
Figure 5.2. N- and C-terminal processing of CGEP. (A) Wild type and mutant forms of recombinant CGEP were digested and analyzed by MS confirming autocatalytic C-terminal processing. C-terminal cleavage/processing was not observed in SDM1 and SDM2 mutants. When CGEP null mutants (Figure 5.7) plants were transformed with the various mutant forms of rCGEP, the C-terminal StrepII tag was cleaved in wild type and SDM2, but not in SDM1 and SDM3 that lack the active site serine and the protease recognized glutamate respectively. Peptides identified by MS/MS are colored red. **(B)** Complete coding sequence for Arabidopsis CGEP. Tryptic peptides (red) and semitryptic peptides (blue) identified in PPDB demonstrate *in vivo* cTP cleavage (mature N-terminus) and C-terminal proteolytic processing.

5.3.2 3D structural model of CGEP

In order to better understand CGEP peptidase substrate interactions, a 3D structural model was constructed using the Iterative Threading Assembly Refinement server (I-TASSER) (Zhang, 2008). The model was based on the mature GCEP primary sequence residues 63 – 961 (predicted cTP removed). The best scoring threading template was the crystal structure of *Stenotrophomonas maltophilia* Dipeptidyl Aminopeptidase IV (PDB 2ECFA), which is member of POP subfamily S9B (See section 4.1). The model shows a typical β -propeller domain (upper) and α/β -hydrolase domain (lower) containing the catalytic triad (S781, D855 and H889) displayed in ball and stick representation (Figure 5.3 A). The active site is partially accessible through a shallow cavity ~ 15 Å wide by at most 10 Å high, visible from the front view. A very narrow cavity (< 4 Å) can also be seen from the top view, extending through the center of the β propeller domain (Figure 5.3A). This feature was noted in other POPs and is thought to be too narrow for a substrate to pass through, without substantial rearrangement of tertiary structure (Tsirigotaki et al., 2017). The structures of CGEP (rainbow from blue, N-terminus to red, C-terminus) and Dipeptidyl Aminopeptidase IV (gray) were overlaid for comparison (Figure 5.3 B) showing close alignment of alpha helices and beta sheets throughout but little to no alignment around the mouth of the central cavity at the N- and C-terminal regions. The C-score of -2.29, associated with this model is relatively low, over a range of -5 [poorest] to 2 [best], largely due to uncertainty in the model over the N-terminal region (Supplemental Figure 5.1) resulting in a large proportion of random coils in the

Figure 5.3. CGEP 3D structural model generated with iTASSER. (A) Cartoon representation of CGEP with active site residues S718, D792, H826 displayed in ball and stick representation, Colored rainbow from blue, N-terminus to red, C-terminus. C-score=-2.29; estimated TM-score = 0.45 ± 0.14 ; estimated RMSD = $14.4 \pm 3.7 \text{ \AA}$. **(B)** CGEP model colored rainbow, overlaid (PDB code: 2ECFA) crystal structure of Dipeptidyl Aminopeptidase IV *Stenotrophomonas maltophilia* (grey). **(C)** Front and top view of model for CGEP lacking N-terminal domain (residues 378 - 962), colored rainbow from blue, N-terminus to red, C-terminus. C-score=-0.93; estimated TM-score = 0.60 ± 0.14 ; estimated RMSD = $9.8 \pm 4.6 \text{ \AA}$. **(D)** Space filling model of structure in C with active site residues in pink and cleaved C-terminal extension in grey. **(E)** Space filling model of structure in D with 15 residue C-terminal extension cut, exposing active site residues.





model (Figure 5.3 A). Therefore, we generated a second model for CGEP with an N-terminal truncation (residues 387 to 961) that provided a better C-score of -0.93, and left most of the structure intact including the proteolytic domain (Figure 5.3 C). This structure for truncated CGEP showed that the C-terminal stretch of 15 residues that is auto-catalytically removed is in close proximity to the active site and its removal may increase accessibility to the active site (Figure 5.3 E).

5.3.3 *CGEP substrate and sequence cleavage specificity*

Protease cleavage nomenclature designates the residues up and down stream of the substrate cleavage site as P_n ... P₂, P₁, P₁', P₂' ... P_n' (Schechter and Berger, 1967). Based on previous reports and the C-terminal auto-catalysis event we were confident that cGEP was capable of cleaving substrates with Glu in the P₁ position. However, we wished to know if Arabidopsis CGEP could cleave after other residues or if there were additional cleavage determinants, beyond the P₁ position, sometimes referred to as sub-site cooperativity (Ng et al., 2009). We also wanted to better resolve the upper and lower substrate size limits because the results of our casein cleavage assay conflict with the low maximum substrate size threshold in a previous report (Laing and Christeller, 1997). We applied the PICS method or variations thereof (Schilling et al., 2011; Biniossek et al., 2016) utilizing both peptide and protein libraries. For each experiment we reacted the fusion proteins GST-CGEP and GST-SDM1 as a control, with the libraries and compared the fragments by MS analysis. Peptide libraries were made by digesting either total soluble leaf extracts or chloroplast stroma with trypsin, LysC or GluC (V8) such that the generated peptides

contained the full range of amino acids throughout. Aliquots of these libraries were analyzed by LC-MS to determine the extent of protein digestion and the relative peptide lengths and sequences to be reacted with recombinant CGEP (Figure 5.6). The two different methods conducted are described below.

In the first experiment peptide libraries were first dimethylated to block all primary α -amino groups (lysine side chains and amino termini) and these libraries were reacted with CGEP (or SDM1 as a negative, inactive protease control). Following digestion, the newly generated amino termini (cleaved by CGEP) were reacted with a cleavable biotin crosslinker that was used to enrich these peptides for analysis by LC-MS (See MATERIALS and METHODS). For the second experiment we reacted unmodified peptide libraries with CGEP or SDM1 and then dimethyl labeled all peptides either with *light* or *heavy* stable isotope reagents (to distinguish the sample from the control) before mixing in equal proportions. Samples were analyzed by nano-LC-MS/MS and peptides identified by MS/MS ion searching with MASCOT (Matrix Science). Across each sample the number of matched spectral counts (SPC) per unique peptide was summed and the list collapsed to identify cleavage events that were unique to the CGEP treated samples (Supplemental Table 5.1 A - E).

We observed 351, 248 and 93 CGEP specific cleavage events (observed by at least two SPC and absent in control) in libraries made with trypsin, LysC and GluC respectively (Experiment 1). 90% of these peptides resulted from cleavage after Glu (Glu in P1 position). The protein sequences 10 residues up- and 10 residues downstream of the cleavage sites were plotted as iceLogos (Colaert et al., 2009) in order to visualize the substrate cleavage preference of CGEP (Figure 5.4 A-C, right panels).

Figure 5.4. Icelogo plots for CGEP specific cleavage of peptide libraries from total soluble leaf extracts reacted with rCGEP or rSDM1; PICS Experiment 1. Peptides absent in control (SDM1 – active site abolished) with at least 2 spectral counts in CGEP samples (left panels) and peptides enriched in control (right panels), used to generate IceLogo. P10 through P10', mature plastid proteome (cTPs removed) used as background proteome, % difference shown, P=0.01 **(A)** trypsin generated peptide library (351 and peptides 199 for CGEP and control respectively). **(B)** LysC generated peptide library (248 and 177 peptides). **(C)** GluC generated peptide library (93 and 73 peptides). Please note that the GluC generated library contained many missed cleavages, un-cleaved Glu residues that were subsequently cleaved by CGEP and not by SDM1.

Beyond the strong enrichment for Glu in the P1 position there was very little sequence preference or avoidance found in either sample apart from a 10% enrichment ($P=0.01$) for Gly in the P1' position (Figure 5.4 A, B). We also generated iceLogos for the peptides unique to the inactive SDM1 control and observed a trypsin-like activity that was also found for peptides shared between the sample and control. Only Arg is observed in the P1 position because the dimethylation of Lys blocks cleavage with P1 Lys. This likely results from some proteolytic activity carried over from the *E. coli* used to express the recombinant proteins. We blocked trypsin activity used to generate the library by addition of the irreversible inhibitor Pefabloc SC, and then purified the peptides by solid phase extraction (SPE) so we doubt this is the source of trypsin-like activity. This activity is also observed in controls that did not use trypsin. The second PICS experiment produced very similar results producing 102 and 59 CGEP specific peptides (present in two replicates) in trypsin and GluC libraries respectively (Figure 5.5), showing a slight preference Thr and Gly in the P1' position.

We investigated whether there were any residues that prevented cleavage after Glu, *e.g.* Pro in the P1' position that prevents cleavage by trypsin and GluC. All amino acids were observed at the P1' position except for Pro (Experiment 1). However, proline does not favorably react with the crosslinking reagent used in this method, which may explain its absence. The second experiment did identify two high scoring peptides (with multiple SPC) with Pro in the P1' position demonstrating that *Arabidopsis* GluC can cleave between Glu and any other amino acid (Figure 5.5).

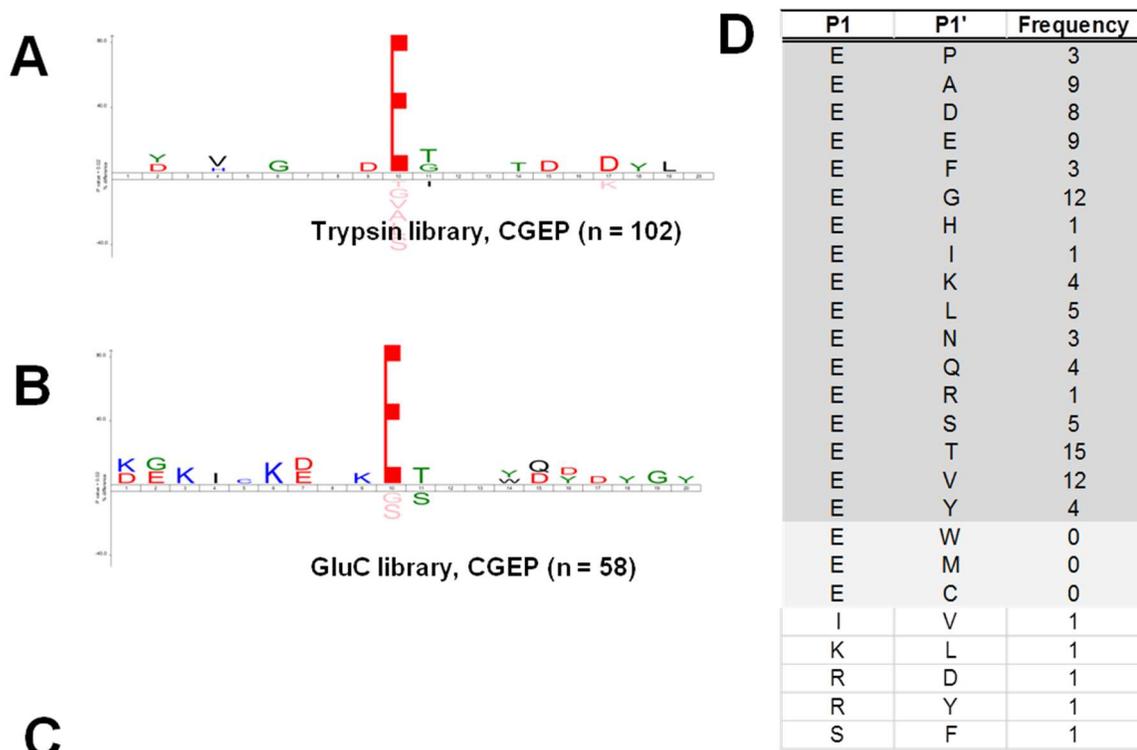


Figure 5.5. Icelogo plots for CGEP specific cleavage of peptide libraries; PICS Experiment 2. Peptides absent in control (SDM1 – active site abolished) and present in both 5 ug and 10 ug CGEP samples used to generate IceLogo. P10 through P10', mature plastid proteome (cTPs removed) used as background proteome, % difference shown, P=0.02 **(A)** trypsin generated peptide library (102 peptide). **(B)** GluC generated peptide library (58 peptide). Please note that the GluC generated library contained many missed cleavages, uncleaved Glu residues that were subsequently cleaved by CGEP and not by SDM1. **(C)** Peptides identified with N-terminal Pro identified in separate database search for mono-methylated peptides absent in primary database search. **(D)** Amino acid frequency in the P1 and P1' positions) for unique, CGEP specific, peptidyl cleavages. Schechter and Berger, 1967 protease substrate nomenclature

To assess whether CGEP has a maximum or minimum size limit we compared the lengths of peptides in the library before and after digestion with CGEP. The largest peptide generated by CGEP was 29 residues in length and corresponded to a 41 residue substrate that was identified in the library before digestion. The most common peptide lengths after digestion with CGEP were 12, 13 and 14 residues for libraries generated with trypsin, LysC and GluC respectively (Figure 5.6).

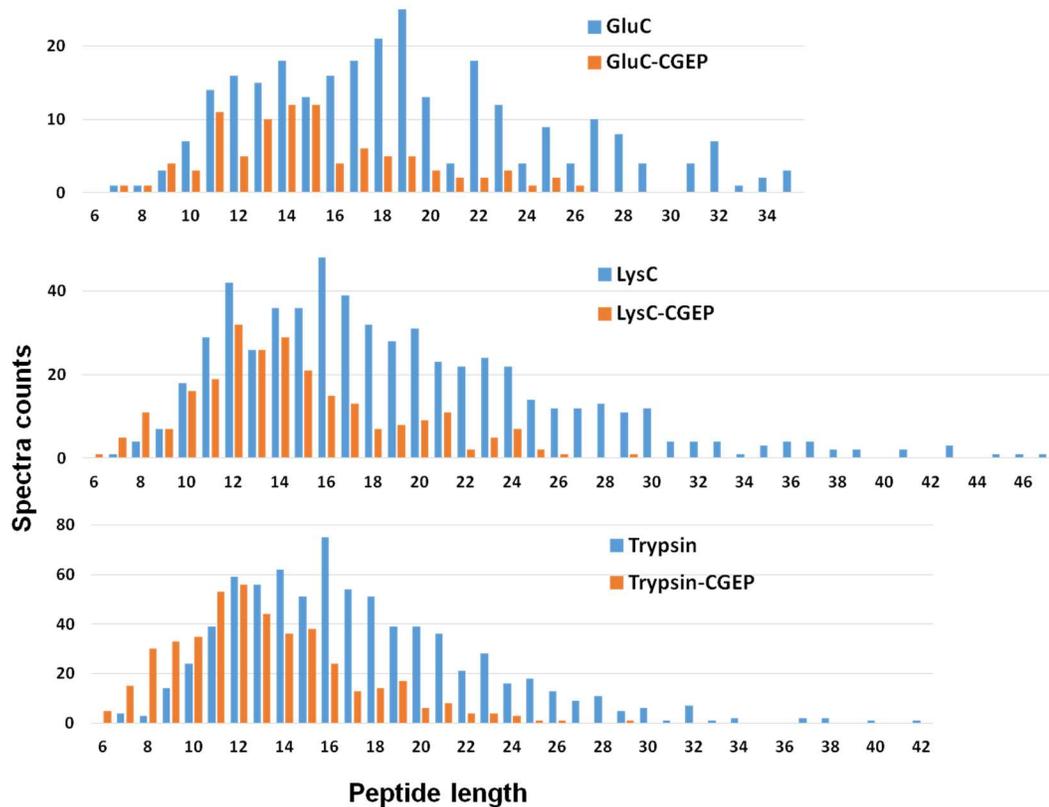


Figure 5.6. Peptide library length distribution before and after reaction with cGEP in PICS experiments. Peptide lengths for all peptides identified in PICS Experiment 1 (Figure 5.4) plotted against total number of spectral counts for GluC, LysC and trypsin generated libraries. See MATERIALS AND METHODS for experimental details.

Additionally we performed a similar experiment using whole protein standards instead of peptides. Just a few protein fragments were unique to the cGEP treated sample including two from myoglobin that both resulted from the same cleavage site on the periphery of the protein (Figure 5.7). This again suggests that CGEP can cleave midsize proteins if the cleavage site is accessible (*i.e.* on the surface).

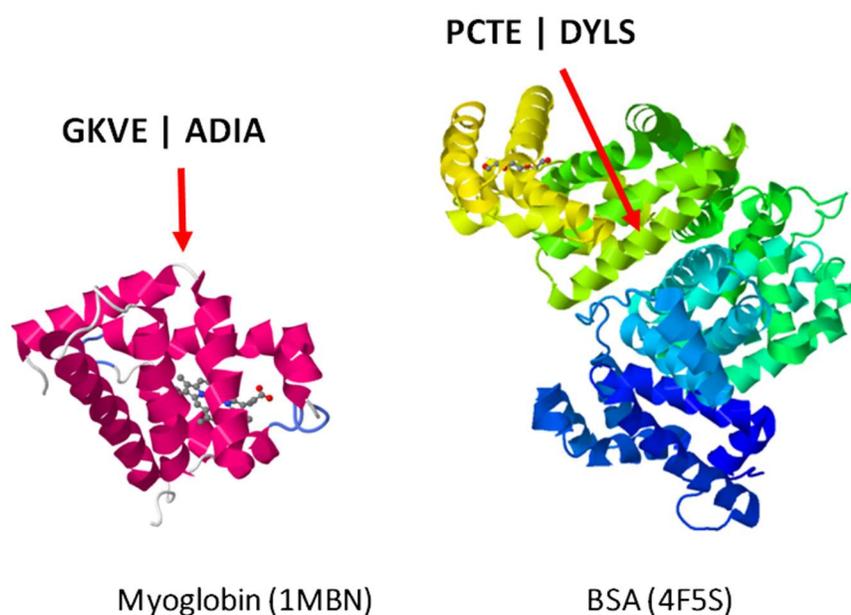


Figure 5.7. Proteins cleaved preferentially by CGEP in PICS type experiment with intact protein standards. Cleavage events unique to CGEP treated proteins (absent in SDM1 control) and location of cleaved sequence in 3D protein structures available in the Protein Data Bank.

5.3.5 CGEP null mutants have no visible phenotype

Although CGEP is an abundant stromal protease, tDNA insertion lines that completely prevent accumulation of CGEP transcript or CGEP protein had no visible phenotype (Figure 5.8 A-D). Immunoblotting of membrane and soluble chloroplast fractions demonstrated that CGEP is present in the stroma and is not associated with thylakoids (Figure 5.8 E). The mutants were grown under various abiotic stress regimes including high-light, drought, cold and heat stress and again, no visible phenotype was observed (data not shown).

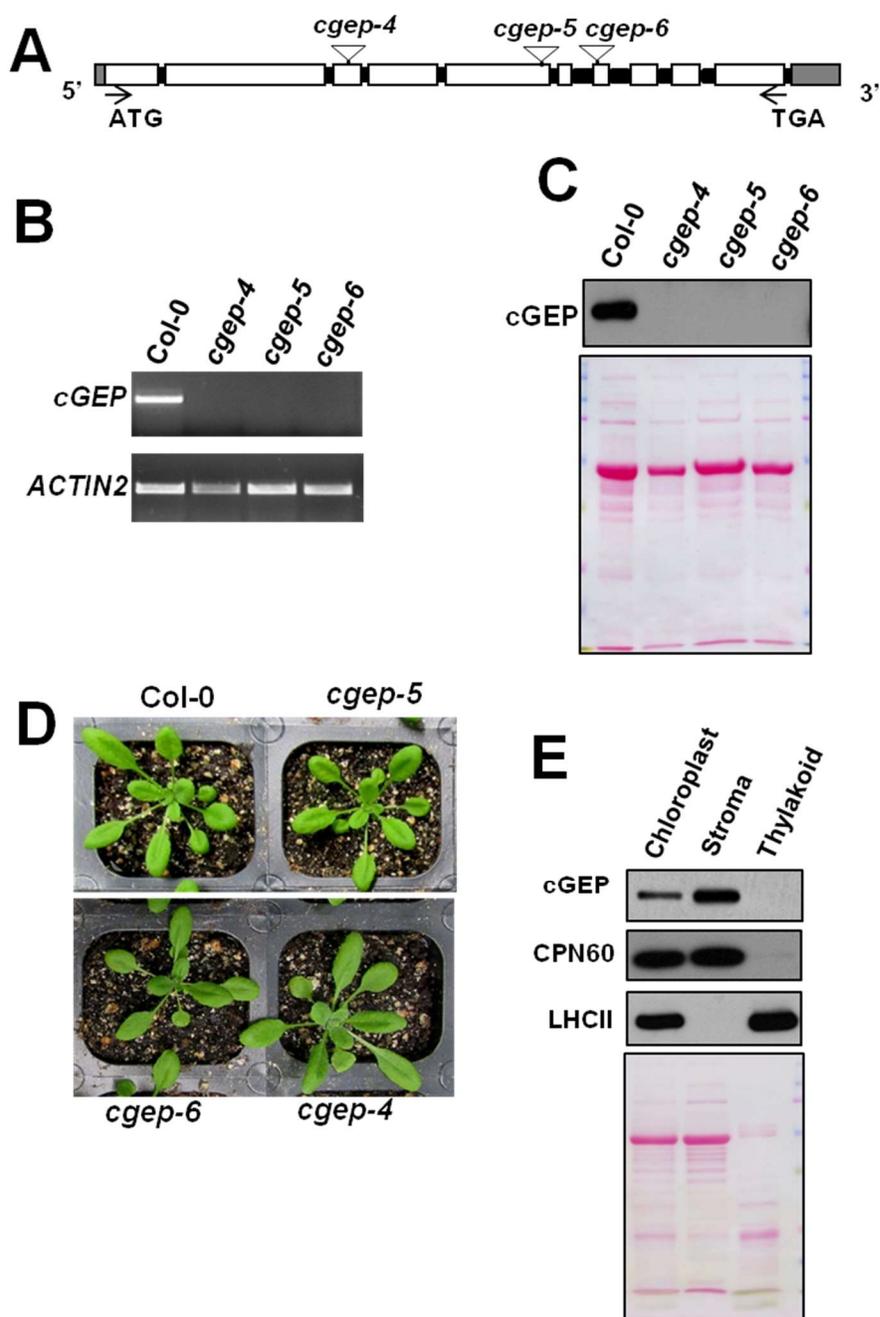


Figure 5.8. CGEP T-DNA insertion lines and sub-cellular localization. (A) Location of T-DNA insertions for established loss-of-function mutant lines. **(B)** RT-PCR with CGEP specific primers for *cgep* mutant lines. **(C)** Immunoblot demonstrating that CGEP protein does not accumulate in *cgep* mutant lines. **(D)** Representative images of wt and *cgep* mutant plants. **(E)** Immunoblot for CGEP in whole chloroplasts, purified stroma and thylakoid fractions. Ponceau stained blot is shown as a loading control.

5.3.6 *N-terminal proteomics*

In order to detect the proteolytic signature resulting from a loss of CGEP function we conducted N-terminal (Nt) proteomic analysis using the TAILS method (Kleifeld et al., 2011). Chloroplast stroma from *cgep-5* and wild type (wt) was compared with five biological replicates. Two chloroplast protein fragments from (TIC110 and CP12) were highly enriched in *cgep-5* suggesting that these fragments are normally degraded in the presence of CGEP, or that loss of CGEP causes up-regulation of other proteases that act upon these protein substrates. Previously we have observed that the loss of chloroplast proteases results in significant restriction of proteolytic maturation for imported nuclear encoded proteins (Chapter 3). Interestingly, we observed no such effect in *cgep-5* suggesting that CGEP forms a separate branch of the chloroplast protease network that is distinct from the central degrading peptidases (*e.g.* CLP, PREP and OOP).

An Nt peptide matching to Translocon at the inner envelope membrane of chloroplasts 110 (TIC110; AT1G06950) over-accumulated in *cgep-5* (in all five replicates). The 27 aa, N-terminal peptide KAPEEDPVQEKEEDEDEEWGSLESLR starts 660 residues downstream of the translated protein start site. There is some debate over the structure of the C-terminal domain of TIC110. It was suggested to contain four transmembrane domains spanning the inner envelope (Balsera et al., 2009). However, the crystal structure of the C-terminal domain of TIC10 from the red alga *Cyanidioschyzon merolae* indicates that C-terminus forms a rod-shaped helix repeat that is exposed to the stroma (Tsai et al., 2013). The observed 27 residue peptide contains nine Glu so it would be a prime target for CGEP if interactions allow.

The other N-terminal peptide accumulating in *cgep-5* was a fragment of the Calvin cycle regulatory protein CP12 (AT2G47400). This protein is a redox regulated, *intrinsically disordered* protein that under low light conditions (when the protein is oxidized) forms a complex with GAPDH and PRK, lowering their activity and reducing the flux through the Calvin cycle (Lopez-Calcagno et al., 2014). When CP12 is reduced, under conditions of high photosynthetic activity, the complex dissociates allowing the Calvin cycle to accelerate. A variety of Nt peptides from CP12 were detected, most of them unchanged in the mutant. However the Nt peptide ADDPVS GECVAAWDEVEELSAAASHAR, just downstream of the first redox regulated cysteine, strongly over-accumulated in the mutant stroma. Again this peptide contains multiple Glu residues that GCEP may cleave, lowering the concentration of this peptide in wild type. Interestingly, the overall protein abundance of CP12 (as well as its homologue CP12-like AT3G62410), was ~5 fold reduced in the mutant. It remains to be determined why the CP12 abundance is reduced and through which mechanisms this occurs.

Peptide	Modification	Mu	Wt	Lab Annot.
IKAAPEGGISDVVEKSIKEAQETCAGDPVSGECVAAWDEVEELSAASHAR	Dimethyl (N-term)	0	1	CP12 domain-containing protein
AAPEGGISDVVEKSIKEAQETCAGDPVSGECVAAWDEVEELSAASHAR	Acetyl (N-term)	1	6	CP12 domain-containing protein
AAPEGGISDVVEKSIKEAQETCAGDPVSGECVAAWDEVEELSAASHAR	Dimethyl (N-term)	9	10	CP12 domain-containing protein
ATSEGEISEKVEKSIQEAKETCADDPVSGECVAAWDEVEELSAASHAR	Dimethyl (N-term)	6	8	CP12 protein
TSEGEISEKVEKSIQEAKETCADDPVSGECVAAWDEVEELSAASHAR	Dimethyl (N-term)	5	7	CP12 protein
EGEISEKVEKSIQEAKETCADDPVSGECVAAWDEVEELSAASHAR	Dimethyl (N-term)	8	8	CP12 protein
GEISEKVEKSIQEAKETCADDPVSGECVAAWDEVEELSAASHAR	Dimethyl (N-term)	8	10	CP12 protein
SEKVEKSIQEAKETCADDPVSGECVAAWDEVEELSAASHAR	Dimethyl (N-term)	1	0	CP12 protein
QEAKETCADDPVSGECVAAWDEVEELSAASHAR	Dimethyl (N-term)	1	0	CP12 protein
ADDPVSGECVAAWDEVEELSAASHAR	Dimethyl (N-term)	12	0	CP12 protein
PVSGECVAAWDEVEELSAASHAR	none	12	12	CP12 protein
DKKKAGGSDPLEEYCNDNPETDECRTYDN	none	1	0	CP12 protein
EKVVGIDLGTTNSAVAAMEGGKPTIVTNAEGQR	Dimethyl (N-term)	24	24	cpHSP70-1 (DnaK homologue)
DAVAEKETTEEGSGEKFEYQAEVSR	Dimethyl (N-term)	12	12	cpHSP90 (Hsp90-5)

Figure 5.9. N-terminal peptides from CP12 identified in TAILS experiment comparing wt and *cgep-5* chloroplast stroma. Spectral counts are approximately equally distributed between wt and mutant stroma for most CP12 N-terminal peptides and for HSP70 and HSP90 (normalization control) except for the peptide ADDPVS... (high-lighted in green) found exclusively in mutant samples with 12 matched MS/MS spectra.

5.5 MATERIALS AND METHODS

5.5.1 CGEP site-directed mutagenesis and in vitro proteolytic activity assays

Mature CGEP (without cTP) was cloned by using forward (AtCGEP-M-FW-BamHI) and reverse (AtCGEP-M-RV-XhoI) primers. The forward primer contains BamHI and the reverse primer contains XhoI sites. The resulting PCR fragment was ligated into the pCR8 topo vector and confirmed by DNA sequencing. The pCR8 vector harboring CGEP gene was digested by BamHI and XhoI restriction enzymes. The resulting DNA fragment was ligated into restriction sites (BamHI and XhoI) of pGEX vector which has an N terminal GST tag. Three mutants SDM1 (CGEP S781R), SDM2 (CGEP E928A D931A) and SDM3 (CGEP E946A E949A E951A) of

AtCGEP were constructed by using a PCR method as described by Bhuiyan et al., 2016. For the mutant SDM1, the C terminal part of mature protein was amplified from pCR8 plasmid harboring CGEP gene by using forward primer (AtCGEPS781R-FW) and reverse primers (AtcGEP-M-RV-XhoI). Forward primer AtCGEPS781R-FW contains the mutation site TCC (Ser) to CGC (Arg). The N-terminal part of mature protein was amplified by using forward (AtCGEP-M-FW-BamHI) and reverse (AtCGEP S781R-RV) primers. Reverse primer AtCGEPS781R-RV contains the introduced site TCC (Ser) to CGC (Arg). The amplified two fragments were gel purified, mixed and used as template (1:1) for second round PCR to amplify mature protein by using the forward and reverse primer sets, AtCGEP-M-FW and AtCGEP-M-RV, respectively. SDM2 was amplified the same way as SDM1 except for the different primer sets to introduce two mutations from GAA (Glu) to GCA (Ala) and GAT (Asp) to GCT (Ala). SDM3 was amplified by using AtCGEP-M-FW-BamHI as a forward primer and -XhoI-E946A E949A E951A-RV as reverse primer. This reverse primer contains three mutation sites E946A (AGT to ACT), E949A (AGC to ACC) and E951A (AAG to ACG). The PCR fragments were ligated into pCR8 topo vector and the mutations were confirmed by DNA sequencing. pCR8 vectors harboring different CGEP mutants were digested by BamHI and XhoI sites and resulting fragments were ligated into the same sites of pGEX-5 vector to fuse with GST at the N term of genes of interest. BL21 *E. coli* cells were transformed by pGEX vectors harboring various CGEP constructs and cells were harvested from liquid culture after addition of 1mM IPTG for 3 h incubation at 22°C. Over-expressed wt and mutants CGEP in *E.coli* were solubilized in 500 mM NaCl, 50 mM Tris, 10% glycerol, at pH 8

and purified on a glutathione resin matrix. The purified protein was dialyzed using a dialysis cassette (slide-A-Lyzer, Thermo Scientific) against buffer 100 mM NaCl and 50 mM Tris and 10% glycerol. After dialysis the protein was concentrated by using microcon centrifugal filter units (Millipore). *In vitro* proteolytic activity was assayed by incubating recombinant CGEP and inactive CGEP-S781R (SDM1) with β -casein for 6 hrs at 37°C. The reaction was stopped by addition of 3% SDS, followed by separation of the protein products by SDS-PAGE and visualization with Coomassie staining.

T-DNA insertion mutants, genotyping and RT-PCR T-DNA insertions lines cgep-4 (SAIL_574_D03), cgep-5 (SALK_066117) and cgep-6 (SAIL_589_G08) were obtained from the Arabidopsis Biological Resource Center (ABRC) and European Arabidopsis stock center (NASC), respectively. Plants containing T-DNA insertions were identified by genotyping, and insertion was confirmed by DNA sequencing. For transcript analysis, total RNA was extracted from Arabidopsis leaves using the RNeasy plant mini kit (QIAGEN). RNA was reverse transcribed with random hexamer primers by using Superscript III reverse transcriptase from Invitrogen. mRNA was normalized by ACTIN2 and the PCR condition was 25 cycles at 94°C for 2 min, 55°C for 30 s and 72°C for 1 min.

5.5.2 Structural model

3D protein structural models for CGEP were generated with i-TASSER (Yang and Zhang, 2015) using the mature CGEP sequence (residues 63 to 961) or CGEP with a truncated N-terminal domain (residues 387-961). The top scoring i-TASSER

model for CGEP (mature) had a C-score of -2.29 (C-scores range from -5 [poorest] to 2 [best]); estimated TM-score = 0.45 ± 0.14 ; estimated RMSD = $14.4 \pm 3.7 \text{ \AA}$. The top scoring i-TASSER model for CGEP (Nt truncation) had a C-score of -0.93; estimated TM-score = 0.60 ± 0.14 ; estimated RMSD = $9.8 \pm 4.6 \text{ \AA}$. Images were generated with PyMol Version 1.7.4 software (Schrödinger).

5.5.3 PICS for determination of protease cleavage specificity

The PICS procedure was based on the methods described by Biniossek and Schilling (Schilling et al., 2011; Biniossek et al., 2016). To generate peptide libraries, 1 mg protein extract (in 50 mM HEPES, 40 $\mu\text{g/ml}$ bestatin, 10 $\mu\text{g/ml}$ phosphoramidon) was mixed with an equal volume of 8M GuHCl to denature proteins. DTT was added to a final concentration 5 mM and the sample incubated at 65°C for 1 hr. After cooling to room temperature, cysteines were alkylated by addition of 15 mM iodoacetamide and incubation for 20 min in darkness. Excess iodoacetamide was quenched with 10mM DTT and the sample was gradually 8-fold diluted with 200 mM HEPES, pH8. Samples were digested with 20 μg trypsin or GluC or 15 μg LysC per one mg protein at 37°C for 16 hrs in 1M GuHCl, 200 mM HEPES, pH8. Any precipitate was removed by centrifugation and an aliquot of the sample (1 mg) resolved by SDS-PAGE and silver staining to ensure the protein digestion was complete. Pefabloc-SC was added to a final concentration of 5 mM to inactivate the digestion protease. The peptide library was then acidified with formic acid and desalted using 1mL Resprep C18 columns (Restek Corp.). The acetonitrile in the elution buffer was removed by Speed Vac and the peptides suspended in 50 mM HEPES, 100 mM NaCl, pH 8 for PICS experiment

2. Alternatively, peptides were dimethylated prior to desalting and carried forward for PICS Experiment 1 (see below).

PICS experiment 1: After digestion of the proteome (with trypsin or GluC or LysC), peptides were dimethylated with CD₂O as per PICS experiment 2 and then desalted with 1mL Resprep C18 columns as described above. Purified dimethylated peptide libraries (120 to 170 µg) were reacted with 6.5 µg of either wild type CGEP or the inactive S781R mutant and incubated for 16 hrs at 37°C. CGEP activity was abolished by heating at 70°C for 10 min. To remove small molecules containing primary amines, samples were again desalted with 1mL Resprep C18 columns as described above and each sample was suspended in 100 µl 200mM HEPES, pH 8. Samples/peptides were then reacted with 5 µl, 10 mM Sulfo-NHS-SS-Biotin (Pierce/Thermo), 0.5 mM final concentration for 2 hrs at 25 C. 1 ml of Streptactin resin was washed 5 times with 50 mM HEPES, 150 mM NaCl, pH 8 and the resin split between six tubes. Above samples were then added to the resin and incubated 2 hrs at 25°C with shaking. Each sample was then transferred to a 0.5 ml spin filter (Pierce). Resin was washed 10 x with 500 µl of wash buffer with a light spin in desktop centrifuge to avoid drying of resin. 300 µl of 50 mM HEPES, 20 mM DTT, pH 8 was added and incubated for 10 minutes at 25°C. Peptides were eluted by centrifugation into a clean tube followed by an additional 200 µl of the above buffer. Samples were desalted using Resprep C18 columns as described above and suspended in 30µl, 2% acetonitrile, 2% formic acid for LC/MS analysis.

PICS experiment 2: 50 µl of peptide library (1 µg/µl in 50 mM HEPES, 100 mM NaCl, pH 8) was mixed with 5 or 10 µl of recombinant CGEP (1 µg/µl in 50mM

TrisHCl, 100 mM NaCl, pH 8, 30% glycerol) or the inactive S781R mutant and incubated for 15 hr at 37°C. Following reaction with CGEP, peptides were dimethylated with either light (control –S781R) or heavy (sample - CGEP) formaldehyde. 2M CH₂O ('light' formaldehyde) or CD₂O ('heavy' formaldehyde) was added to give a final concentration of 40 mM, followed immediately by addition of 1 M NaCNBH₃ to give final concentration of 30 mM. The samples were incubate for 2 hr in thermocycler at 25°C and then a second aliquot of CH₂O and NaCNBH₃ was added, as above, to give 80 and 60 mM final concentrations respectively and the samples were incubated overnight at 25°C. The dimethylation reaction was quenched with 0.1 M glycine, final concentration. The sample (CGEP digested - Heavy label) and control (S781R digested - light label) reactions were then mixed in a fresh tube. A 5 µg aliquot was desalted by C18 ziptip (Millipore) using the manufacturer's guidelines and the peptide eluant brought to dryness by Speed Vac. Samples were suspended in 20µl, 2% acetonitrile, 2% formic acid for LC/MS analysis.

For LC-MS/MS analysis 6.4 µl of each sample was loaded onto a C18 trapping column and then eluted onto a 15 cm x 75 µm I.D. C18 PepMap column (ThermoFisher) interfaced to an LTQ Orbitrap (ThermoFisher). A 90 min linear gradient from 3% to 40% solvent B was used to separate the peptides. A data dependent acquisition method was used whereby MS spectra were acquired in the Orbitrap at 100K resolution followed by 5 data dependent MS/MS scans were conducted in the ion trap. Peak lists (mgf files) for database searching were generated from Thermo raw data files using DTA Supercharge. The peak lists were searched using MASCOT 2.4 (Matrix Science) against TAIR10, appended with all reverse

sequences (Decoy) and common contaminants (71,149 sequences and 29,099,754 residues). Following an initial database search performed at 30 ppm MS tolerance, 0.8 Da MSMS tolerance, the peak lists were recalibrated as previously described (Friso et al., 2010), a semi-specific enzyme search was then conducted - semiArgC, semiGluC (V8) - allowing for 3 missed cleavages, 6 ppm MS tolerance and 0.8 Da MSMS tolerance. For PICS Experiment 1, fixed modifications were carboxamidomethyl Cys and dimethyl Lys (heavy, +32 Da), variable modifications were oxidized Met, pyroGlu N-term Gln, dimethyl N-term (heavy, +32 Da) and Thioacyl N-term. For PICS experiment 2, fixed modifications were carboxamidomethyl Cys and dimethyl Lys, variable modifications were oxidized Met, pyroGlu N-term Gln, acetyl N-term and dimethyl N-term (light, +28 Da or heavy, +32 Da). Another search including singly methylated N-term was conducted for select files in order to detect methylated Nt Pro. The database search results were parsed and sorted in Excel.

Sequence logo and Icelogo plots were generated with Icelogo version 1.2 (<http://www.proteomics.be>). The mature plastid proteome (cTPs removed) was used as a background to normalize for natural amino acid abundance in the library (PPDB).

LITERATURE CITED

- Balsera M, Goetze TA, Kovacs-Bogdan E, Schurmann P, Wagner R, Buchanan BB, Soll J, Bolter B** (2009) Characterization of Tic110, a channel-forming protein at the inner envelope membrane of chloroplasts, unveils a response to Ca(2+) and a stromal regulatory disulfide bridge. *J Biol Chem* **284**: 2603-2616
- Binossek ML, Niemer M, Maksimchuk K, Mayer B, Fuchs J, Huesgen PF, McCafferty DG, Turk B, Fritz G, Mayer J, Haecker G, Mach L, Schilling O** (2016) Identification of Protease Specificity by Combining Proteome-Derived Peptide Libraries and Quantitative Proteomics. *Mol Cell Proteomics* **15**: 2515-2524
- Colaert N, Helsens K, Martens L, Vandekerckhove J, Gevaert K** (2009) Improved visualization of protein consensus sequences by iceLogo. *Nat Methods* **6**: 786-787
- Forsberg J, Strom J, Kieselbach T, Larsson H, Alexciev K, Engstrom A, Akerlund H-E** (2005) Protease activities in the chloroplast capable of cleaving an LHCII N-terminal peptide. *Physiologia Plantarum* **123**: 21-29
- Kaushik S, Sowdhamini R** (2014) Distribution, classification, domain architectures and evolution of prolyl oligopeptidases in prokaryotic lineages. *BMC Genomics* **15**: 985
- Kleifeld O, Doucet A, Prudova A, auf dem Keller U, Gioia M, Kizhakkedathu JN, Overall CM** (2011) Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nat Protoc* **6**: 1578-1611
- Laing WA, Christeller JT** (1997) A Plant Chloroplast Glutamyl Proteinase. *Plant Physiol* **114**: 715-722
- Lopez-Calcagno PE, Howard TP, Raines CA** (2014) The CP12 protein family: a thioredoxin-mediated metabolic switch? *Front Plant Sci* **5**: 9
- Majeran W, van Wijk KJ** (2009) Cell-type-specific differentiation of chloroplasts in C4 plants. *Trends Plant Sci* **14**: 100-109
- Ng NM, Pike RN, Boyd SE** (2009) Subsite cooperativity in protease specificity. *Biol Chem* **390**: 401-407

- Nishimura K, Kato Y, Sakamoto W** (2017) Essentials of Proteolytic Machineries in Chloroplasts. *Mol Plant* **10**: 4-19
- Rawlings ND, Barrett AJ, Bateman A** (2012) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **40**: D343-350
- Rea D, Fulop V** (2006) Structure-function properties of prolyl oligopeptidase family enzymes. *Cell Biochem Biophys* **44**: 349-365
- Rea D, Fulop V** (2011) Prolyl oligopeptidase structure and dynamics. *CNS Neurol Disord Drug Targets* **10**: 306-310
- Schechter I, Berger A** (1967) On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* **27**: 157-162
- Schilling O, auf dem Keller U, Overall CM** (2011) Protease specificity profiling by tandem mass spectrometry using proteome-derived peptide libraries. *Methods Mol Biol* **753**: 257-272
- Schilling O, Huesgen PF, Barre O, Auf dem Keller U, Overall CM** (2011) Characterization of the prime and non-prime active site specificities of proteases by proteome-derived peptide libraries and tandem mass spectrometry. *Nat Protoc* **6**: 111-120
- Shan L, Mathews, II, Khosla C** (2005) Structural and mechanistic analysis of two prolyl endopeptidases: role of interdomain dynamics in catalysis and specificity. *Proc Natl Acad Sci U S A* **102**: 3599-3604
- Tsai JY, Chu CC, Yeh YH, Chen LJ, Li HM, Hsiao CD** (2013) Structural characterizations of the chloroplast translocon protein Tic110. *Plant J* **75**: 847-857
- Tsirigotaki A, Elzen RV, Veken PV, Lambeir AM, Economou A** (2017) Dynamics and ligand-induced conformational changes in human prolyl oligopeptidase analyzed by hydrogen/deuterium exchange mass spectrometry. *Sci Rep* **7**: 2456
- Van Elzen R, Lambeir AM** (2011) Structure and function relationship in prolyl oligopeptidase. *CNS Neurol Disord Drug Targets* **10**: 297-305
- van Wijk KJ** (2015) Protein maturation and proteolysis in plant plastids, mitochondria, and peroxisomes. *Annu Rev Plant Biol* **66**: 75-111
- Venalainen JI, Juvonen RO, Mannisto PT** (2004) Evolutionary relationships of the prolyl oligopeptidase family enzymes. *Eur J Biochem* **271**: 2705-2715

Yamauchi Y, Ejiri Y, Sugimoto T, Sueyoshi K, Oji Y, Tanaka K (2001) A high molecular weight glutamyl endopeptidase and its endogenous inhibitors from cucumber leaves. *J Biochem* **130**: 257-261

Yang J, Zhang Y (2015) I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res* **43**: W174-181

Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**: 40