

**MASS SPECTROMETRY-BASED CHARACTERIZATION OF
CHLOROPLAST PROTEIN COMPLEXES AND CLP PROTEASE
FUNCTION**

A Dissertation

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

Paul Dominic Olinares

January 2011

© 2011 Paul Dominic Olinares

MASS SPECTROMETRY-BASED CHARACTERIZATION OF CHLOROPLAST PROTEIN COMPLEXES AND CLP PROTEASE FUNCTION

Paul Dominic Olinares, Ph.D.

Cornell University 2011

Mass spectrometry (MS) has enabled large-scale protein identification and quantification, yielding significant insights into relevant biological systems. MS has been extensively applied in this thesis to study chloroplast protein complexes and to quantify protein expression levels in the model plant *Arabidopsis thaliana*.

Plant plastids contain a 350 kDa Clp protease core complex consisting of two heptameric rings. The complex contains nine different proteins in one or more copies, namely five serine-type ClpP peptidases (ClpP1,3-6) and four non-proteolytic ClpR subunits (ClpR1-4). This core complex was purified from different transgenic lines harboring affinity-tagged Clp subunits. The absolute amount of each subunit and the corresponding stoichiometry within the heptameric rings was determined by MS analysis using stable isotope-labeled versions of peptides that uniquely represent each Clp protein, expressed from a synthetic gene. Results showed that the ClpP and ClpR proteins assemble into a single asymmetric complex, with the two component rings exhibiting differential proteolytic functionalities and adaxial surfaces; functional consequences are discussed.

To determine the consequences of the partial loss of Clp protease activity, the leaf proteomes of wild-type and a *CLPP3* null mutant were compared using MS-based spectral counting. 2116 proteins and protein families were quantified and their differential expression in the mutant was tested for significance. This showed a

general up-regulation of proteins involved in chloroplast proteome homeostasis and gene expression, but down-regulation of the photosynthetic machinery and specific responses of secondary metabolism. This demonstrates the essential contribution of ClpP3 in Clp core assembly and function, as well as the crucial role of the Clp core complex in chloroplast viability.

Large-scale, label-free quantification was used to characterize large (>800 kDa) soluble, chloroplast-localized protein and nucleoprotein assemblies in *Arabidopsis thaliana* which were separated by size exclusion chromatography. Hierarchical clustering using MS-derived spectral counts for each chromatography fraction effectively grouped the identified proteins into functional complexes. This combined experimental and bioinformatics analyses resolved chloroplast chromatin, numerous novel proteins, as well as chloroplast ribosomes in different assembly and functional states, with ribosome assembly factors and proteins involved in co-translational modifications, targeting and folding.

BIOGRAPHICAL SKETCH

Born and raised in Isabela, Philippines, **Dom Olinares** obtained his B.S. in Chemistry, *magna cum laude*, from the Ateneo de Manila University, where he graduated as Valedictorian of Class 2002. His undergraduate thesis involved the purification and refolding of a single-chain antibody fragment designed to be a potential breast cancer biomarker. Shortly after graduation, he joined the Jesuit Volunteer Philippines (JVP) and was assigned as a high school teacher in the remote coastal town of Hinunangan, Southern Leyte for a year.

He then went for graduate studies at the Department of Chemistry and Chemical Biology at Cornell University in 2005. He joined the lab of Dr. Klaas van Wijk in 2006 and worked on mass spectrometry-based characterization of protein complexes in plant chloroplasts.

Outside of the laboratory, Dom is involved with the Cornell Filipino Association, and is a principal banduria player and vocalist for 14Strings!, the Cornell Filipino Rondalla group. Dom also writes poems in the Filipino language.

*para kay Victoria, ang aking ina
at sa Pilipinas, ang aking Inang Bayan*

ACKNOWLEDGEMENTS

My deepest gratitude to Klaas van Wijk for the invaluable mentorship and for his unfaltering support throughout my research studies.

Many thanks to Professors Tom Brenna, Hening Lin and Tadhg Begley for guidance and for being part of my Special Committee.

My heartfelt thanks to past and present members of the van Wijk lab particularly Wojciech Majeran, Verenice Ramirez-Rodriguez, Boris Zybaïlov, Heidi Rutschow, Giulia Friso, Yukari Asakura and Anton Poliakov for the interdisciplinary training and for being patient with my questions and requests for help.

Thanks to Sharon Calhoun and Pat Hine (Chemistry Department) as well as Pat Welch and Tara Nihill (Plant Biology Department) for taking care of my interdepartmental status.

To the Filipino residents and Filipino graduate students in Ithaca—Tita Thess and Tito Ditto, Tita Jane and Tito Cliff, Tita Cynthia, Tita Mimi, Florio, Nesha, Rod, Apollo, Philipina and Edson for the warm welcome. *Maraming, maraming salamat.*

To my friends—Jay, Paulina, Ankush and Osman for being part of my journey.

To my family in the Philippines and in North America—Robert, Roy, Jobs, Tita Norma, Kuya Joel, Kuya Mike, Tita Precy, Manang Odessa. *Mahal ko kayong lahat.*

This work was supported by grants from the National Science Foundation, the United States Department of Agriculture and the New York State Office of Science, Technology and Research (NYSTAR) to Klaas Jan van Wijk.

TABLE OF CONTENTS

	page
Biographical Sketch	iii
Dedication	iv
Acknowledgements	v
Table of contents	vi
List of Figures	vii
List of Tables	ix
CHAPTER ONE. Introduction	1
CHAPTER TWO. Subunit composition, evolution and implications of an asymmetric plastid ClpPR protease	26
CHAPTER THREE. Mass spectrometry-based comparative proteomics of an Arabidopsis Clp mutant; consequences for assembly states of the ClpPR complex, proteome homeostasis and plastid-localized metabolism	64
CHAPTER FOUR. Megadalton complexes in the chloroplast stroma of <i>Arabidopsis thaliana</i> characterized by size exclusion chromatography, mass spectrometry and hierarchical clustering	117
APPENDIX	176

LIST OF FIGURES

CHAPTER ONE	page
Figure 1.1 A typical MS-based “bottom-up” proteomics workflow	2
Figure 1.2 The structure of ClpP from <i>E. coli</i>	7
Figure 1.3 The plastid-localized Clp protease family in <i>Arabidopsis</i>	11
Figure 1.4 Comparison of Clp sequences from <i>E. coli</i> and <i>Arabidopsis</i>	12
Figure 1.5 Growth and development of wildtype and Clp mutants	15
 CHAPTER TWO	
Figure 2.1 Workflow for determination of Clp core composition	29
Figure 2.2 Native PAGE analysis of affinity purified Clp assemblies	30
Figure 2.3 Optimization and accuracy of the QconCAT quantification	33
Figure 2.4 Evolutionary relationships among the Clp assemblies in prokaryotes and plant plastids	37
Supplementary Figure 2.1 Generation of transgenic plants	44
Supplementary Figure 2.2 Affinity purification and MS-based identification of Clp assemblies	45
Supplementary Figure 2.3 Representative extracted ion chromatograms of the Clp-QconCAT peptides	46
Supplementary Figure 2.4 Multiple sequence alignments of the plant plastid Clps (Clp P3-P6)	47
Supplementary Figure 2.5 Multiple sequence alignments of the plant plastid Clps (Clp P1, R1-R4)	48
Supplementary Figure 2.6 Phylogenetic analysis of the ClpP/R proteins	49

CHAPTER THREE

Figure 3.1 Development of the homozygous <i>clpp3-1</i> mutant	68
Figure 3.2 Large-scale SPC-based comparative proteomics of <i>clpp3-1</i>	69
Figure 3.3 Comparison of significantly expressed proteins in <i>clpp3-1</i>	74
Figure 3.4 Comparison of LTQ-Orbitrap data acquisition and and statistical analyses of the comparative proteomics dataset	81
Figure 3.5 Differential accumulation of photosynthetic proteins	85
Figure 3.6 Immunoblot analysis of proteins from wild-type and <i>clpp3-1</i>	86
Figure 3.7 Characterization of Clp assemblies in <i>clpp3-1</i>	95
Figure 3.8 Summary of the consequences of the loss of <i>CLPP3</i>	99

CHAPTER FOUR

Figure 4.1 Workflow for the experimental and bioinformatics analyses of the oligomeric state of chloroplast stroma	120
Figure 4.2 Size exclusion fractionation of chloroplast stroma	123
Figure 4.3 Overview of SEC migration profiles for MDa-sized complexes	133
Figure 4.4 Characterization of the 50S and 30S ribosomes	137

LIST OF TABLES

CHAPTER TWO	page
Table 2.1 Stoichiometry of the Clp subunits within each Clp ring	35
Supplementary Table 2.1 MS/MS analysis of the purified Clp assemblies	50
Supplementary Table 2.2 Properties of peptides comprising Clp-QconCAT	51
Supplementary Table 2.3 Linear MS instrument response for Clp-QconCAT	52
Supplementary Table 2.4 Quantification of Clp subunits in the Clp core	53
Supplementary Table 2.5 Quantification of Clp subunits in the Clp rings	54
Supplementary Table 2.6 Accessions for protein and cDNA Clp sequences	55
Supplementary Table 2.7 Primers used in the study	59
 CHAPTER THREE	
Table 3.1 Summary of subcellular location of identified proteins	71
Table 3.2 Chloroplast-localized proteins that are significantly upregulated or downregulated in <i>clp3-1</i> relative to wild-type	76
Table 3.3 Sample preparation and data acquisition parameters for the Clp proteomics studies	82
Table 3.4 Evaluation of peptide match rates for the different Clp proteomics studies	83
Table 3.5 Composition of Clp assemblies as determined by MS analysis	96
 CHAPTER FOUR	
Table 4.1 Elution profiles of proteins identified in MDA-sized fractions	127

CHAPTER ONE

INTRODUCTION*

Mass spectrometry-based quantitative proteomics

Proteomics allows the determination of qualitative and quantitative protein accumulation patterns across spatial, temporal and physiological cellular states, including protein or nucleoprotein complexes. Mass spectrometry (MS) using soft ionization techniques, combined with bioinformatics, has been the enabling tool in proteomics (1, 2). The information on absolute or relative amounts of proteins and the corresponding changes in abundance levels from MS-based quantification has facilitated the reconstruction of the dynamics of biological processes yielding remarkable functional insights (3).

A typical “bottom-up” MS-based proteomics workflow (3, 4) is described in Figure 1.1. In this workflow, the proteome sample of interest is extracted and further separated by affinity selection or biochemical fractionation such as a denaturing one-dimensional gel electrophoresis (SDS-PAGE). The fractionated proteins in solution or in excised gel bands are enzymatically digested by a protease, usually by trypsin, generating component peptides. To further reduce sample complexity, the resulting peptide mixture is typically separated by high-pressure liquid chromatography (LC). The eluted peptides are then analyzed by a mass spectrometer (MS) and survey (or precursor ion) scans (MS scans) are taken. The MS computer selects peptide ions (typically by abundance as measured by their peak intensities) for fragmentation and a

* Part of this introduction is adapted from a review that I co-authored: **Olinares, P.D., Kim, J., and van Wijk, K.J.** (2010). The Clp protease system; a central component of the chloroplast protease network. *Biochimica et Biophysica Acta (BBA)- Bioenergetics* special issue on “Regulation of electron transport in chloroplasts”. In press.

series of tandem or MS/MS scans of the resulting fragment ions are also recorded. The precursor mass information from the MS scan and the corresponding fragment ion masses from the MS/MS spectra are then matched against a protein sequence database to identify the peptides and infer the protein identities in the analyzed sample.

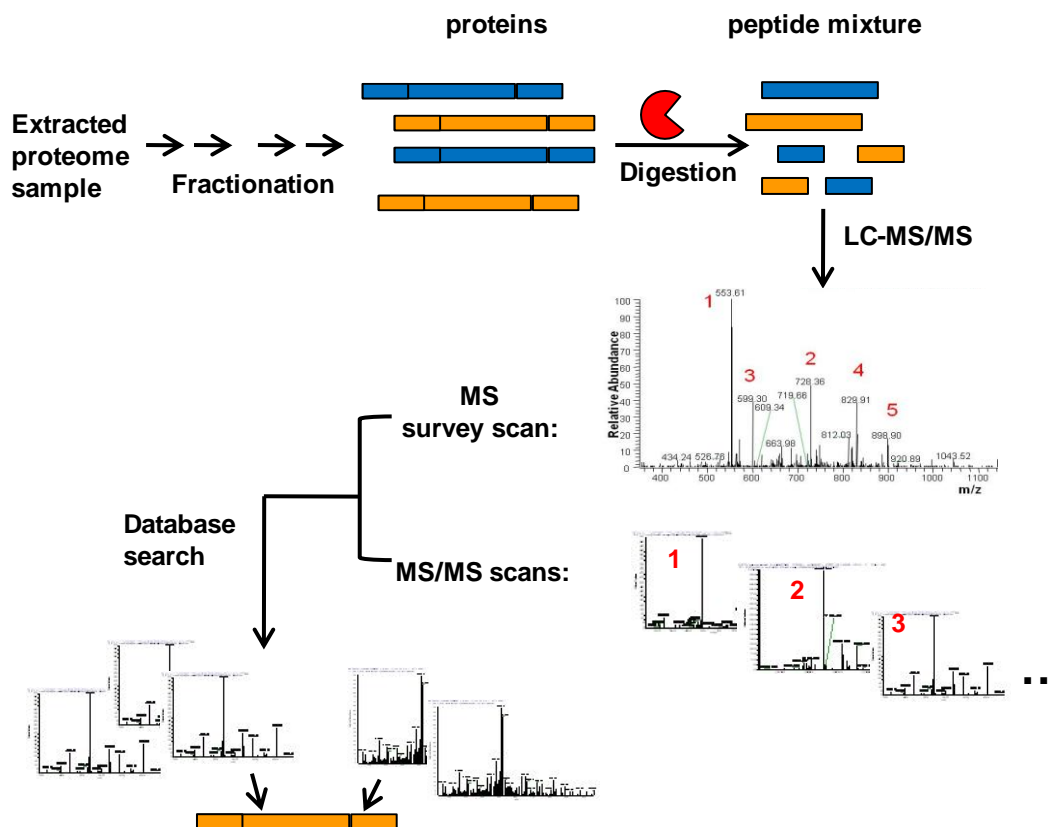


Figure 1.1. A typical MS-based “bottom-up” proteomics workflow. The proteome sample of interest is extracted and further separated by various protein fractionation strategies. The target proteins are then enzymatically digested by a protease to generate component peptides which are separated by liquid chromatography (LC) and analyzed by mass spectrometry (MS). Peptide ion masses in the MS survey scans and peptide fragment ion masses in the tandem or MS/MS scans are recorded. The MS and MS/MS spectra are then matched against a protein sequence database to identify the proteins in the analyzed sample.

For quantification, labeling of samples or spiking of labeled peptide standards can be performed in any of the steps prior to LC-MS analysis. The stable-isotope labeled peptides and their unlabeled versions exhibit similar physical and chemical

properties and thus co-elute during chromatography but can be distinguished by a diagnostic mass shift (typically several Da) during MS analysis. MS-derived peak intensities or peak areas for both labeled and endogenous peptides are directly compared to obtain the absolute amounts of the corresponding target peptides. For label-free quantification, the samples to be compared are run consecutively using the same LC-MS workflow. The corresponding peak intensities or peak areas from the MS scans or the number of matched MS/MS spectra that are matched to peptides from the database search can be used for differentiating protein abundance levels of the samples under comparison.

MS-based absolute quantification strategies involve the addition of known amounts of standards which include stable-isotope labeled synthetic peptides, peptide concatamers or proteins (5, 6). The use of individual heavy-coded peptides, popularly known as the AQUA (absolute quantification) strategy (7) requires chemical synthesis of reference peptides which can be expensive and time-consuming (8). Furthermore, difficulty in peptide resuspension as well as peptide degradation and modifications hinder accurate quantification (8). An alternative strategy involves the use of concatenated peptide standards, also known as QconCAT (quantification concatamer) (9) which are readily obtained by overexpression in a recombinant bacterial system and is thereby relatively inexpensive and sustainable (*i.e.*, one can always express, label and purify QconCAT proteins expressed from the initial QconCAT gene) (9, 10). Moreover, digestion of QconCAT releases equimolar amounts of constituent peptides making it an ideal reference for multiplexed absolute quantification of subunit composition and of stoichiometry in a protein complex or in biochemical pathways. Examples include the characterization of the stoichiometry of the elongation factor eIF2B-eIF2 in budding yeast (11) and the rod photoreceptor complex (12). The QconCAT technique was used in Chapter 2 to determine the subunit stoichiometry of

the Clp protease complex. Several variations in the QconCAT strategy have been made such as the inclusion of flanking sequences from the target peptides to equalize cleavage efficiency of peptides generated from QconCAT and from target proteins (11) and the fusion of a GFP construct to QconCAT for more accurate quantification of the protein standard prior to spiking (12). The third absolute quantification method uses labeled intact protein as reference, known as PSAQ (protein standard absolute quantification) (13). Since full-length proteins are spiked, this strategy provides the largest sequence coverage compared to the other two methods and can be added at earlier stages of sample preparation thereby minimizing experimental variation (13). However, quantification of multiple proteins requires synthesis of multiple recombinant protein standards and can thus be cost-prohibitive.

Comparative MS-based analysis of the relative abundances of proteins in different samples can involve the use of stable isotope labels or can be label-free (3, 4). Stable isotope labels can be incorporated by cultivation of cells in growth media supplemented with essential heavy isotope-labeled nutrients (*in vivo*) or by chemical derivatization (*in vitro*) (4). Examples of the former include metabolic labeling using [¹⁵N]-containing inorganic salts (14) and stable isotope labeling by amino acids in cell culture (SILAC) (15). *In vitro* approaches include the use of cysteine-reactive isotope-coded affinity tags (ICAT) (16) or amine-reactive isobaric mass tags such as iTRAQ (17). The disadvantages of *in vitro* labeling methods include the need for costly labeling kits, extra sample processing (labeling) steps and incompatibility with certain MS platforms (*e.g.*, iTRAQ on ion traps). On the other hand, *in vivo* labeling methods might not always be amenable for the biological system of interest. For example, plants are autotrophic and are capable of synthesizing their own essential amino acids regardless of any supplementation in growth media. Implementation of the SILAC methodology in *Arabidopsis* suspension cells achieved label incorporation of only 70-

80% (18), which is inadequate for accurate proteome analysis. However, metabolic labeling of plants with [¹⁵N]-containing inorganic salts have been demonstrated in plant cell cultures (19, 20) and in using intact plants grown hydroponically (21-23). Nevertheless, metabolic labeling in plant proteomics is still in its infancy mostly due to technical challenges.

Several label-free relative quantification strategies have emerged as practical alternatives to comparative label-based methods. One such strategy involves using peak intensities or peak areas of similar peptides identified in the samples being compared as abundance measures. However, this strategy requires challenging computational efforts particularly in peak alignment and background correction (4, 24). Another label-free technique relies on the observation that the frequency of detecting the component peptide for each protein, also known as spectral counts (SPCs), correlates well with the abundance of the corresponding protein (25-27). This abundance correlation extends over a linear dynamic range of at least two orders of magnitude for complex protein mixtures (25, 26, 28, 29). Since SPCs can be readily extracted from the MS database search result files, spectral counting can be relatively straightforward and is widely applicable for MS analysis of any biological sample. The disadvantage of this technique is that small fold changes (*i.e.*, less than ~2-fold) are harder to detect in particular for proteins of lower abundance. Nevertheless, spectral counting has yielded similar quantitative results as label-based methods including metabolic labeling (27), iTRAQ (30) and SILAC (31) with spectral counting generating greater protein coverage in some cases (30, 31). Given these advantages, spectral counting has been increasingly adopted in various comparative proteomics studies. This label-free SPC technique was applied in Chapters 3 and 4 and different tools were used to test for the significance of differential protein accumulation.

The Clp protease system

The Clp (caseinolytic protease) family has been found in almost all bacterial species and eukaryotic organelles except for archaea, mollicutes and some fungi (32). The Clp machinery has two oligomeric components, namely i) a barrel-shaped tetradecameric protease core with the catalytic sites sequestered inside the complex and ii) hexameric ring-like ATP-dependent chaperones. The chaperones recognize specific substrates with or without the aid of adaptors, unfold these substrates and translocate them into the proteolytic core for degradation (33, 34). Compartmentalization of the proteolytic sites within the core complex and coupling with chaperones and associated factors for substrate delivery enable targeted, adaptive and regulated protein degradation within the cell.

The Clp machinery is well-studied in the Gram negative bacterium *Escherichia coli*, where it was first isolated and characterized (35, 36). From X-ray crystal structure determination, the *E. coli* ClpP peptidase core is a homotetradecameric complex consisting of two stacked heptameric rings with the active sites enclosed within the equatorial cavity (37). The ClpP monomer structure resembles a hatchet with a wedge-shaped head consisting of six α/β repeats and a handle composed of a long β -strand and an α -helix (Figure 1.2A). The Ser-His-Asp catalytic triad is located in a cleft where the head and handle domains intersect. Each heptameric ring is assembled through extensive interactions among the head domains of the subunits (Figure 1.2B). To form the complex, two heptameric rings stack together through an intercalating network of handle domains from each subunit. The N-termini of the subunits line the axial pores while their C-termini protrude out of the complex (37).

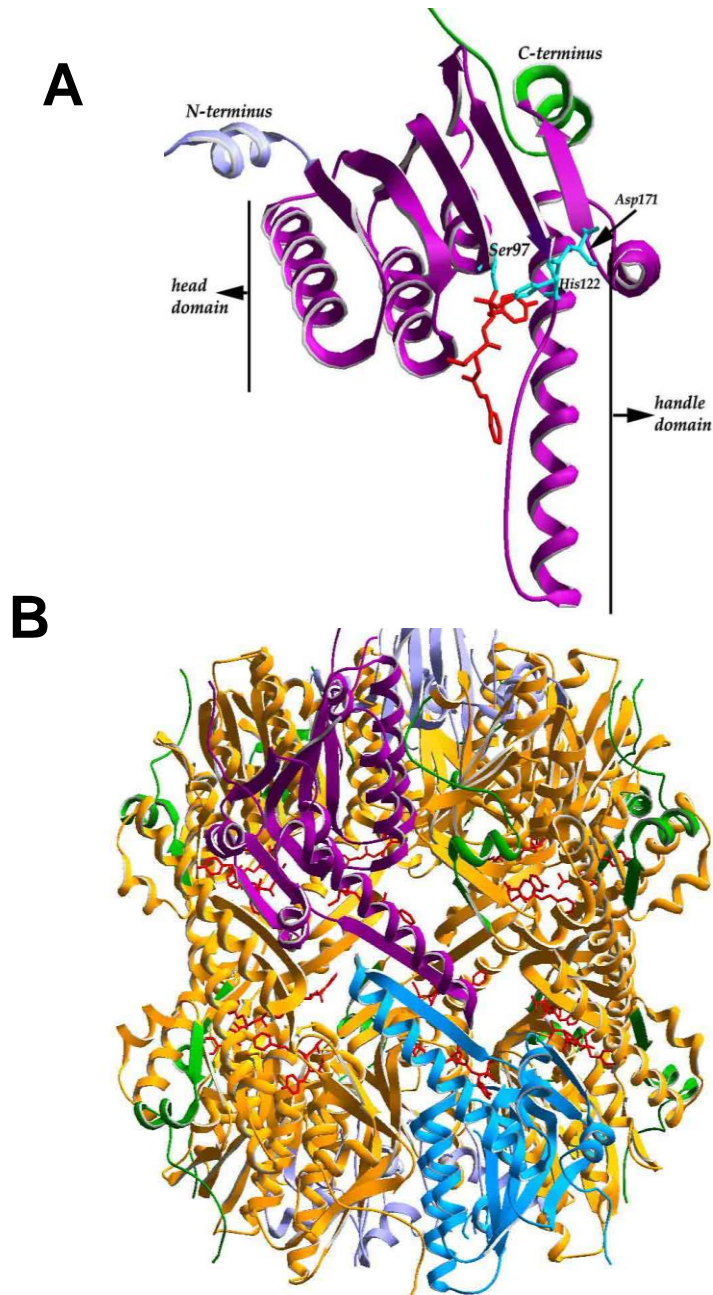


Figure 1.2 The structure of ClpP from *E. coli*. **A)** The structure of the *E. coli* ClpP monomer with a covalently bound tripeptide inhibitor shown in red. The N- and C-termini are highlighted in cyan and green, respectively. **B)** The *E. coli* Clp core is homotetradecameric. In this side-view, two subunits are colored differently (purple and blue) to highlight ring-to-ring contacts via their handle domains. The active sites are enclosed (see substrates in red) in the central cavity. Access to the complex is through the axial ends where the N-termini of the subunits (in cyan) protrude. The C-termini of the monomers (in green) jut out of the complex. The 3D structures were obtained from PDB: 2FZS (47, 48) and were rendered using SwissPDB Viewer (49, 50).

Access to the catalytic cavity of the Clp core is restricted by narrow axial entrance pores (37). The Clp protease core can slowly degrade small peptides by itself but requires the cooperation of ATP-driven chaperones to digest larger peptides or denatured proteins (38-40). These chaperones include ClpX and ClpA, which harbor one or two AAA+ (ATPase associated with cellular activities) domains, respectively. They form hexamers that mount coaxially on the ClpP core for substrate delivery (39). The interaction between the axial surfaces of the Clp core with the chaperone rings is highly dynamic, partly due to the symmetry mismatch between the hexameric chaperone and (double) heptameric protease core (41-43). Substrates are fed axially into ClpP (44, 45) and degraded processively into peptides of seven to eight residues (46). In addition, the association of a small adaptor protein, ClpS, confers additional substrate specificity to ClpAP complexes. ClpS changes the affinity of ClpA towards protein aggregates (51) and modulates the recognition and delivery of substrates with N-terminal degradation signals known as N-end rule substrates (52-54). ClpS binds to the N-domain of ClpA (55, 56). A molecule of ClpS can associate with high affinity to ClpA hexameric rings and is sufficient to induce ClpS-mediated substrate delivery and degradation by the ClpAP complex (57).

The Gram positive bacterium *Bacillus subtilis* is a free living bacterium and is much better adapted to higher temperatures than *E. coli*; these differences in temperature tolerance are likely to affect the protein homeostasis machineries, including proteases (58). *B. subtilis* has three Clp chaperones (ClpC, ClpE and ClpX) and one ClpP; surprisingly it lacks the ClpP-independent disaggregase ClpB which is present in *E. coli* and higher plant plastids. Unlike most other organisms, ClpP in *B. subtilis* is monomeric *in vivo* (and *in vitro*) but oligomerizes upon interaction with its adaptor-activated Clp chaperones, in particular MecA (59, 60). The ClpC chaperones depend on different adaptors such as MecA, YpbH, MscB, ClpS, whereas ClpX

interacts with the adaptor YjbH (58). Interestingly, non-photosynthetic plastids (apicoplasts) of the human malaria parasite *Plasmodium falciparum* contain both a ClpP and a ClpR protein, a ClpP-like protein without the catalytic triad (61). Both ClpP and ClpR proteins form mostly homoheptameric rings as observed by size-exclusion chromatography, analytical ultracentrifugation and electron microscopy. The X-ray structure of ClpP showed the protein as a compacted tetradecamer similar to that observed for *Streptococcus pneumoniae* and *Mycobacterium tuberculosis* ClpPs (61).

The Clp system in the photosynthetic bacterium *Synechococcus* sp. PCC 7942 consists of three *CLPP* genes (*P1,2,3*), one *CLPR*, *CLPX* and *CLPC* gene, as well as two *CLPS* genes (62). The ClpR protein is structurally similar to ClpP but lacks the catalytic residues for peptide bond hydrolysis. It is proteolytically inactive but its presence in the Clp core appeared not to limit the overall proteolytic activity (63). The cyanobacterial ClpC is a ClpA orthologue and exhibits protein refolding and protein disaggregation activities (64). Cyanobacteria possess two different functional Clp assemblies, namely a dispensable ClpP_{1,2} complex associating with ClpX and an essential ClpP₃/ClpR complex interacting with ClpC (62). In addition, the two adaptors ClpS1 and ClpS2 associate only with ClpC and not with ClpX (62). In another cyanobacterial species, *Nostoc* sp. PCC7120, ClpC has been shown to interact with NblA, a protein involved in the degradation of the light-harvesting phycobilisome complexes (65). This interaction is ATP/ADP-dependent and NblA is proposed to act as an adaptor for ClpC towards Clp-mediated phycobilisome clearance (65).

In the green alga *Chlamydomonas reinhardtii*, three ClpP genes (*CLPP1*, *CLPP4*, *CLPP5*) and five ClpR genes (*CLPR1-CLPR4*, *CLPR6*) encode for the chloroplast-localized heterooligomeric Clp protease core complex (66). Interestingly, plastid-encoded *clpP1* contains a large insertion sequence (IS1) and several ClpP1

protein variants have been observed within the Clp core, including the precursor ClpP_{1H} and several processed versions which lack various segments within the IS1 (66, 67). The processed forms do not arise from mRNA splicing (68), nor from protein intron self-splicing as in inteins, but instead from endoproteolytic cleavage within the IS1 region of ClpP_{1H}, mostly likely after the assembly of the ClpP_{1H}-containing protease complex (67).

The plastid-localized Clp proteolytic system further diversified in higher plants. The dicotyledon *Arabidopsis thaliana* has five serine-type Clp proteases (P1, P3-P6), four non-proteolytic ClpRs (R1-4), three Clp AAA+ chaperones (C1, C2, D) that are orthologues of the *E. coli* ClpA, the adaptor ClpS (homologous to the *E. coli* ClpS) and ClpT1, T2 with unknown functions but with similarity to the N-terminal domain of bacterial ClpA (69, 70) (see Figure 1.3). ClpB3, another plastid-localized AAA+ chaperone, lacks the I(L)GF motif combined with an upstream basic residue (69) which is implicated in interaction with the Clp protease core (71). Therefore, ClpB3 is unlikely to directly associate with the plastid Clp core complex. We identified all Clp proteins by mass spectrometry in chloroplasts of *Arabidopsis* and non-green plastids in *Brassica rapa* roots and *Brassica oleracea* petals (69, 72, 73). We note that mass spectrometry-based identification of the relatively small 12 kDa ClpS protein was challenging and was achieved only after size exclusion fractionation of stroma (73) (also, see the Plant Proteome Database, PPDB at <http://ppdb.tc.cornell.edu/>).

ClpR1, R3 and R4 have a 10-12 amino-acid insertion domain compared to the ClpP1-6 proteins, ClpR2 and *E. coli* ClpP (see Figure 1.4).(69). Our homology models suggested that this domain is protruding into the tunnel of the ClpP/R core, possibly affecting substrate presentation to the catalytic sites (69) Additional distinct features of chloroplast ClpR and ClpP proteins compared to the *E. coli* ClpP are the extended C-

termini, that might influence protein interactions to the adaxial site(s) of the ClpPR core, *i.e.* with hexameric rings of ClpC,D or with ClpT1,T2 (69) (Figure 1.4).

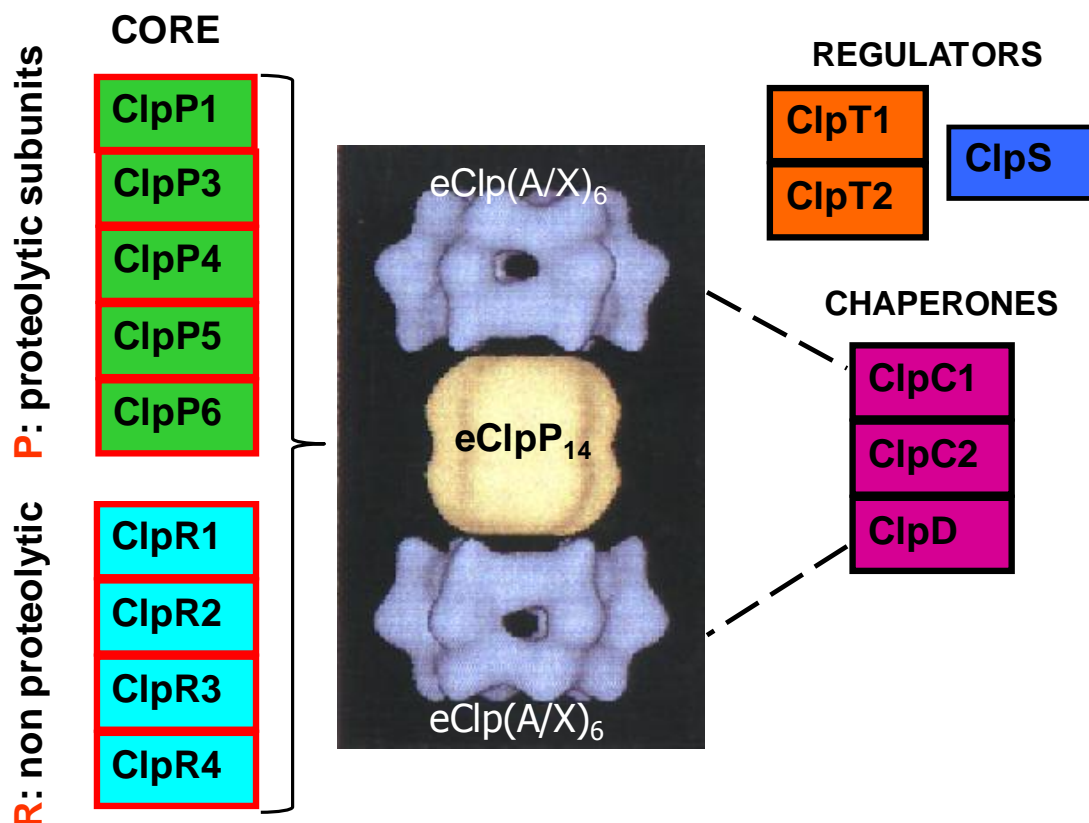


Figure 1.3. The plastid-localized Clp protease family in *A. thaliana*. The Clp core in *E. coli* is a homotetradecamer, designated here as eClpP₁₄ (yellow in cryoEM image of the eClpP-ClpA/X complex (41)). In contrast, *A. thaliana* has 9 different proteins that comprise the plastid-localized Clp protease core complex. ClpP1-6 harbor the conserved catalytic triad, while ClpR1-4 do not. *A. thaliana* plastids have three homologues for eClpA (ClpC1,C2, D) and none for eClpX. Two short novel proteins (ClpT1,2) similar to the N-terminal part of eClpA are also present in plant plastids. ClpS is homologous to the adaptor protein eClpS.

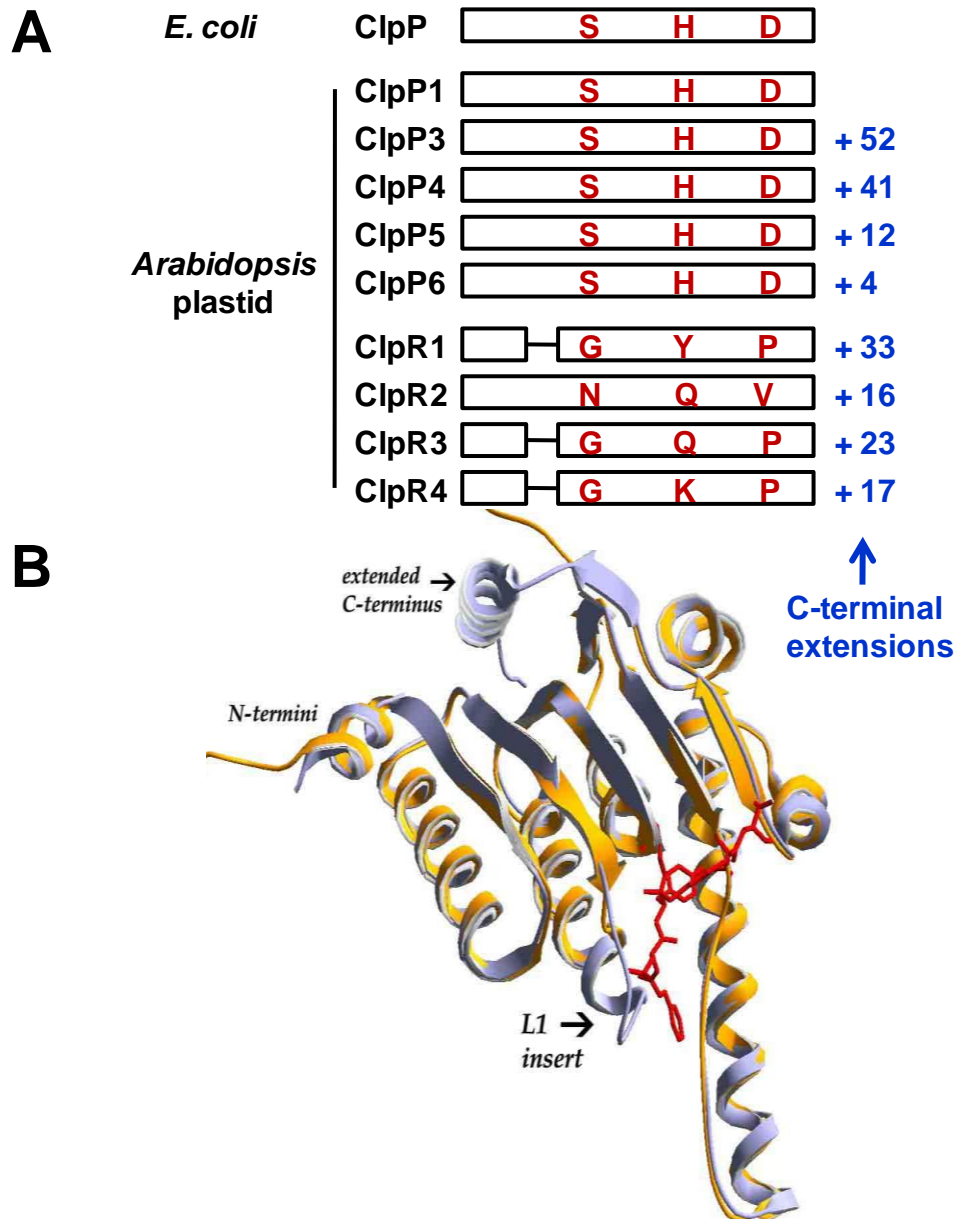


Figure 1.4. Comparison of the ClpP sequences from *E. coli* and the plastid-localized Clp proteins in *A. thaliana*. **A)** Sequence comparison of the *E. coli* ClpP and the plastid ClpPs and ClpRs. The serine-histidine-aspartic acid catalytic triad (S,H,D) (red) is replaced in the ClpRs. Three ClpR proteins harbor an extra insertion loop (10-12 residues). The C-terminal extensions of the plastid ClpP/R proteins relative to *E. coli* are also indicated in blue. Diagrams are not drawn to scale. **B)** Superimposed structures of *E. coli* ClpP (orange; with PDB ID: 2FZS (47, 48) and the homology model of ClpR4 (gray; with PDB ID: 1R99 (48, 69)). The covalently bound tripeptide inhibitor is shown in red. Compared to the *E. coli* ClpP, ClpR4 has an L1 insertion loop close to the substrate-binding region and has an extended C-terminus. The 3D structures were rendered using SwissPDB Viewer (49, 50).

Surprisingly, our analyses of the native soluble proteome of non-green plastids in roots and petals of *B. rapa* and *B. oleracea*, respectively, as well as chloroplasts of *A. thaliana*, have shown that they all contain a stromal 325-350 kDa Clp core protease complex consisting of all the nine Clp proteins (ClpP1, P3-P6, R1-R4) and ClpT1 and T2 (69). Furthermore, separation of these complexes by native isoelectric focusing showed a single complex in each plastid type (at pI of ~5) (69). Native gel electrophoresis of the ClpPR complex combined with Western blot analysis suggested the existence of two Clp core subcomplexes (heptameric rings): a ~180-200 kDa ring containing ClpPs (ClpP3-P6) and a ~230 kDa ring containing all the ClpRs (R1-R4) and ClpP1 (74). Furthermore, this western blot analysis also suggested that ClpT1 binds to the ClpP3-P6 ring (74). We point out that the individual rings have not yet been purified and that the precise stoichiometry within the rings is not clear, even if tentative quantification was obtained from gel band stain intensities for Clp proteins from stromal proteome analyses (75). In contrast to plastids, plant mitochondria contain a single homotetradecameric ClpP2 complex, presumably associated with ClpX chaperones (69, 76).

ClpT1,2 are unique to land plants and have not been found in algae or (cyano)bacteria (72). From our threading and structural modeling analysis, it was concluded that ClpT1 and T2 cannot fit within the Clp core ring structure but they dock well on the Clp core's aromatic pockets situated near the axial entrance (69). From these modeling studies, we proposed several possible roles for ClpT. For instance, ClpT may regulate the association of the hexameric ClpC1,C2 or D with the core by competing for docking sites on the core. Based on homology to the *E. coli* ClpA, ClpT1,2 have predicted binding sites for interactions with the adaptor ClpS (55, 69). As such, plastid ClpT1,2 might provide docking sites for ClpS to deliver substrates, most likely short peptides, into the core.

The Clp chaperones ClpC1 and C2 were found as dimers at ~200 kDa and after interaction with their substrates they are predicted to dock as hexameric rings on either axial side of the proteolytic core (69, 77, 78). However, the Clp core associated with the ClpC or D chaperone complex has not yet been isolated; this will be important to i) more firmly establish ClpC/D function, ii) to determine if ClpC/D can bind to both adaxial sites of the ClpPR core, and iii) to understand its relationship with ClpT1,2 (*i.e.*, whether ClpC/D can associate with a ClpPRT complex). Finally, little is known about plastid ClpS. It remains to be determined if ClpS does interact with the ClpC/D chaperones or ClpT1,2 and the adaptor function of ClpS has not been demonstrated in plants.

Lessons from Clp gene disruption studies in photosynthetic organisms

Arabidopsis mutants with reduced or complete loss of expression of nuclear-encoded Clp genes have been generated by anti-sense RNA, by T-DNA insertions and by EMS mutagenesis. Phenotypes for mutant lines with visible phenotypes (yellow-pale green and/or seedling lethal plants) in *Arabidopsis* are shown in Figure 1.5. So far, mutants for three of the four ClpR proteins and all four nuclear-encoded plastid-localized ClpP proteins (P3-6) have been obtained and analyzed in varying degrees. Down-regulation of the nuclear-encoded *CLPR2* gene by 80% delayed chloroplast biogenesis and affected protein homeostasis (79, 80). Similarly, partial down-regulation of *ClpP4* and of *ClpP6* by anti-sense RNA techniques reduced plant growth and development and resulted in pale-green plants (74, 81). Reduced levels of ClpR1 protein by EMS mutagenesis (*clpr1-1*) or T-DNA insertion (*clpr1-2*) to nearly undetectable levels resulted in virescent mutants, but *clpr1-1* plants did grow on soil and set viable seeds without problem (82, 83). In case of null mutants for *CLPR2* and *CLPR4*, the seeds did germinate and under heterotrophic conditions, pale green seedlings developed

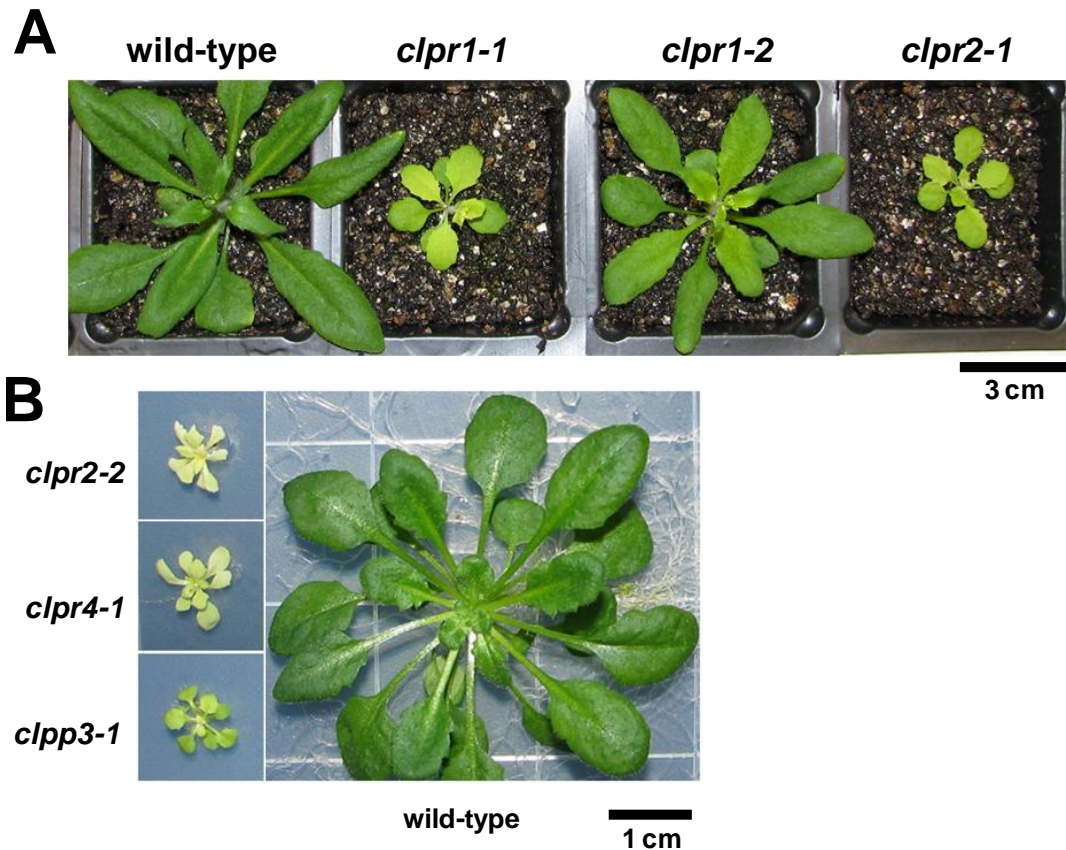


Figure 1.5. Growth and development of wild-type and various Clp mutants. **A)** Direct comparison of wild-type, null *clpr1-1*, knockdown *clpr1-2*, and knockdown *clpr2-1* mutants grown on soil for 28 days under a 16/8 hour light/dark cycle at $120 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$. Bar = 3 cm. **B)** Homozygous *clpr2-2*, *clpr4-1*, and *clpp3-1* null plants on MS agar plates with 2% sucrose grown for 2 months under a 10/14 hour light/dark cycle at $40 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$. Bar = 1 cm. Images taken by Jitae Kim.

slowly but produced no viable seeds (83). The relatively mild phenotype of the *CLPR1* mutants could be explained by a partial functional substitution by ClpR3 (83, 84). Null mutants in *CLPP4* and *CLPP5* were blocked in embryogenesis and never germinated, even when supplied with sugars (83). Collectively, analyses of Clp mutants revealed that the ClpPR core subunits, except for ClpR1, exhibit little functional redundancy and that the Clp protease core is essential for plastid development both during embryogenesis and in cotyledons and leaves.

Mutants for ClpC1, C2 and D, as well as ClpB3, have also been analyzed in fair detail. Loss of expression of the *ClpC1* chaperone resulted in reduced plant growth and chloroplast development, but homozygous plants remained autotrophic, producing viable seeds (85-87). Mutants lacking both ClpC1 and ClpC2 however were blocked in embryogenesis (88). Although ClpC1,2 accumulate predominantly in the stroma, ClpC1 is also associated with the chloroplast protein translocation machinery in the inner envelope, interacting in particular with Tic110 and Tic40 (89, 90). Consistently, loss of ClpC1 results into lower protein import rates into isolated chloroplasts (86, 87). Recently it was reported that the visible pale-green phenotype of a *CLPC1* mutant (*irm1*) harboring a point mutation (Gly773 into Arg773) can be complemented by providing extra iron to the plants, but not by extra zinc or manganese, suggesting altered chloroplast import of nuclear-encoded proteins involved in iron transport in *irm1* (91).

Whereas the most comprehensive genetic analysis of nuclear-encoded Clp genes was performed in *Arabidopsis*, downregulation of the plastid-encoded *CLPP1* gene was done in tobacco, which showed that the ClpP1 protein is essential for shoot development (68, 92). Downregulation of the *CLPP1* gene in the green algae *C. reinhardtii*, suggested that ClpP1 is involved in the degradation of the thylakoid proteins, in particular in genetic backgrounds that lead to misassembly of thylakoid complexes (93, 94). Furthermore, in the photosynthetic bacterium *Synechococcus* sp. PCC 7942, mutant analyses has revealed that *CLPP1* and *CLPP2* are dispensable but that *CLPP3* and *CLPR* are essential to cell viability (63, 95, 96).

MOTIVATIONS AND FOCUS OF THIS DISSERTATION

Chloroplasts are essential organelles of prokaryotic origin found in every plant cell. The chloroplast proteome is predicted to contain ~3000 different nuclear-encoded proteins and about 80 plastid-encoded proteins. The chloroplast also harbors a small genome and its own gene expression machinery. The imported and chloroplast-encoded proteins then interact and assemble into protein complexes to execute their function. As such, the determination of the composition, structure and dynamics of protein complexes is key to understanding biological processes within the chloroplast. This work involves the application of absolute and relative MS-based quantification strategies in characterizing the components and the function of chloroplast-localized protein complexes specifically, nucleoprotein assemblies involved in plastid gene expression and the Clp protease complex involved in protein turnover.

Numerous proteins involved in protein import, protein folding and proteolysis are crucial for chloroplast biogenesis and maintenance of chloroplast viability. The Clp family is the largest chloroplast-localized protease system in plants with nine different subunits (ClpP1, ClpP3-6, ClpR1-4) comprising the core protease complex and six others (ClpC1-2, ClpD, ClpT1-2, ClpS) involved in substrate recognition and delivery. Genetic and phenotypic analyses of various Clp mutants in plants showed that the Clp protease subunits have differential functional contributions and all of them, except ClpR1, are essential for embryo or seedling development. As such, knowing the stoichiometry and ring localization of the different ClpP/R proteins through MS-based absolute quantification is a prerequisite for understanding the functional contributions of these proteins. Comparative MS-based analysis of the leaf proteomes of a Clp mutant relative to wild-type also provided insight into the role of the Clp protease system in chloroplast viability and plant development.

REFERENCES

1. Aebersold, R. & Goodlett, D. R. (2001) *Chem Rev* 101, 269-95.
2. Aebersold, R. & Mann, M. (2003) *Nature* 422, 198-207.
3. Domon, B. & Aebersold, R. (2010) *Nat Biotechnol* 28, 710-21.
4. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. (2007) *Anal Bioanal Chem* 389, 1017-31.
5. Brun, V., Masselon, C., Garin, J. & Dupuis, A. (2009) *J Proteomics* 72, 740-9.
6. Kito, K. & Ito, T. (2008) *Curr Genomics* 9, 263-74.
7. Gerber, S. A., Kettenbach, A. N., Rush, J. & Gygi, S. P. (2007) *Methods Mol Biol* 359, 71-86.
8. Mirzaei, H., McBee, J. K., Watts, J. & Aebersold, R. (2008) *Mol Cell Proteomics* 7, 813-23.
9. Beynon, R. J., Doherty, M. K., Pratt, J. M. & Gaskell, S. J. (2005) *Nat Methods* 2, 587-9.
10. Pratt, J. M., Simpson, D. M., Doherty, M. K., Rivers, J., Gaskell, S. J. & Beynon, R. J. (2006) *Nat Protoc* 1, 1029-43.
11. Kito, K., Ota, K., Fujita, T. & Ito, T. (2007) *J Proteome Res* 6, 792-800.
12. Nanavati, D., Gucek, M., Milne, J. L., Subramaniam, S. & Markey, S. P. (2008) *Mol Cell Proteomics* 7, 442-7.
13. Dupuis, A., Hennekinne, J. A., Garin, J. & Brun, V. (2008) *Proteomics* 8, 4633-6.
14. Oda, Y., Huang, K., Cross, F. R., Cowburn, D. & Chait, B. T. (1999) *Proc Natl Acad Sci U S A* 96, 6591-6.
15. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A. & Mann, M. (2002) *Mol Cell Proteomics* 1, 376-86.

16. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H. & Aebersold, R. (1999) *Nat Biotechnol* 17, 994-9.
17. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A. & Pappin, D. J. (2004) *Mol Cell Proteomics* 3, 1154-69.
18. Gruhler, A., Schulze, W. X., Matthiesen, R., Mann, M. & Jensen, O. N. (2005) *Mol Cell Proteomics* 4, 1697-709.
19. Engelsberger, W. R., Erban, A., Kopka, J. & Schulze, W. X. (2006) *Plant Methods* 2, 14.
20. Lanquar, V., Kuhn, L., Lelievre, F., Khafif, M., Espagne, C., Bruley, C., Barbier-Brygoo, H., Garin, J. & Thomine, S. (2007) *Proteomics* 7, 750-4.
21. Bindschedler, L. V., Palmblad, M. & Cramer, R. (2008) *Phytochemistry* 69, 1962-72.
22. Palmblad, M., Bindschedler, L. V. & Cramer, R. (2007) *Proteomics* 7, 3462-9.
23. Ippel, J. H., Pouvreau, L., Kroef, T., Gruppen, H., Versteeg, G., van den Putten, P., Struik, P. C. & van Mierlo, C. P. (2004) *Proteomics* 4, 226-34.
24. Stevenson, S. E., Chu, Y., Ozias-Akins, P. & Thelen, J. J. (2009) *J Proteomics* 72, 555-66.
25. Liu, H., Sadygov, R. G. & Yates, J. R., 3rd (2004) *Anal Chem* 76, 4193-201.
26. Old, W. M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K. G., Mendoza, A., Sevinsky, J. R., Resing, K. A. & Ahn, N. G. (2005) *Mol Cell Proteomics* 4, 1487-502.
27. Zybaylov, B., Coleman, M. K., Florens, L. & Washburn, M. P. (2005) *Anal Chem* 77, 6218-24.
28. Zhang, G., Fenyo, D. & Neubert, T. A. (2009) *J Proteome Res* 8, 1285-92.

29. Cooper, B., Feng, J. & Garrett, W. M. (2010) *J Am Soc Mass Spectrom* 21, 1534-46.
30. Patel, V. J., Thalassinos, K., Slade, S. E., Connolly, J. B., Crombie, A., Murrell, J. C. & Scrivens, J. H. (2009) *J Proteome Res* 8, 3752-9.
31. Collier, T. S., Sarkar, P., Franck, W. L., Rao, B. M., Dean, R. A. & Muddiman, D. C. (2010) *Anal Chem*.
32. Yu, A. Y. & Houry, W. A. (2007) *FEBS Lett* 581, 3749-57.
33. Baker, T. A. & Sauer, R. T. (2006) *Trends Biochem Sci* 31, 647-53.
34. Striebel, F., Kress, W. & Weber-Ban, E. (2009) *Curr Opin Struct Biol* 19, 209-17.
35. Hwang, B. J., Park, W. J., Chung, C. H. & Goldberg, A. L. (1987) *Proc Natl Acad Sci U S A* 84, 5550-4.
36. Katayama-Fujimura, Y., Gottesman, S. & Maurizi, M. R. (1987) *J Biol Chem* 262, 4477-85.
37. Wang, J., Hartling, J. A. & Flanagan, J. M. (1997) *Cell* 91, 447-56.
38. Thompson, M. W., Singh, S. K. & Maurizi, M. R. (1994) *J Biol Chem* 269, 18209-15.
39. Grimaud, R., Kessel, M., Beuron, F., Steven, A. C. & Maurizi, M. R. (1998) *J Biol Chem* 273, 12476-81.
40. Weber-Ban, E. U., Reid, B. G., Miranker, A. D. & Horwich, A. L. (1999) *Nature* 401, 90-3.
41. Beuron, F., Maurizi, M. R., Belnap, D. M., Kocsis, E., Booy, F. P., Kessel, M. & Steven, A. C. (1998) *J Struct Biol* 123, 248-59.
42. Gribun, A., Kimber, M. S., Ching, R., Sprangers, R., Fiebig, K. M. & Houry, W. A. (2005) *J Biol Chem* 280, 16185-96.

43. Bewley, M. C., Graziano, V., Griffin, K. & Flanagan, J. M. (2006) *J Struct Biol* 153, 113-28.
44. Singh, S. K., Grimaud, R., Hoskins, J. R., Wickner, S. & Maurizi, M. R. (2000) *Proc Natl Acad Sci U S A* 97, 8898-903.
45. Ishikawa, T., Beuron, F., Kessel, M., Wickner, S., Maurizi, M. R. & Steven, A. C. (2001) *Proc Natl Acad Sci U S A* 98, 4328-33.
46. Choi, K. H. & Licht, S. (2005) *Biochemistry* 44, 13921-31.
47. Szyk, A. & Maurizi, M. R. (2006) *J Struct Biol* 156, 165-74.
48. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000) *Nucleic Acids Res* 28, 235-42.
49. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. (2003) *Nucleic Acids Res* 31, 3381-5.
50. Guex, N. & Peitsch, M. C. (1997) *Electrophoresis* 18, 2714-23.
51. Dougan, D. A., Reid, B. G., Horwich, A. L. & Bukau, B. (2002) *Mol Cell* 9, 673-83.
52. Erbse, A., Schmidt, R., Bornemann, T., Schneider-Mergener, J., Mogk, A., Zahn, R., Dougan, D. A. & Bukau, B. (2006) *Nature* 439, 753-6.
53. Wang, K. H., Sauer, R. T. & Baker, T. A. (2007) *Genes Dev* 21, 403-8.
54. Dougan, D. A., Truscott, K. N. & Zeth, K. (2010) *Mol Microbiol* 76, 545-58.
55. Guo, F., Esser, L., Singh, S. K., Maurizi, M. R. & Xia, D. (2002) *J Biol Chem* 277, 46753-62.
56. Zeth, K., Ravelli, R. B., Paal, K., Cusack, S., Bukau, B. & Dougan, D. A. (2002) *Nat Struct Biol* 9, 906-11.
57. De Donatis, G. M., Singh, S. K., Viswanathan, S. & Maurizi, M. R. (2010) *J Biol Chem* 285, 8771-81.

58. Kirstein, J., Moliere, N., Dougan, D. A. & Turgay, K. (2009) *Nat Rev Microbiol* 7, 589-99.
59. Kirstein, J., Schlothauer, T., Dougan, D. A., Lilie, H., Tischendorf, G., Mogk, A., Bukau, B. & Turgay, K. (2006) *Embo J* 25, 1481-91.
60. Mei, Z., Wang, F., Qi, Y., Zhou, Z., Hu, Q., Li, H., Wu, J. & Shi, Y. (2009) *J Biol Chem* 284, 34366-75.
61. El Bakkouri, M., Pow, A., Mulichak, A., Cheung, K. L., Artz, J. D., Amani, M., Fell, S., de Koning-Ward, T. F., Goodman, C. D., McFadden, G. I., Ortega, J., Hui, R. & Houry, W. A. (2010) *J Mol Biol* 404, 456-77.
62. Stanne, T. M., Pojidaeva, E., Andersson, F. I. & Clarke, A. K. (2007) *J Biol Chem* 282, 14394-402.
63. Andersson, F. I., Tryggvesson, A., Sharon, M., Diemand, A. V., Classen, M., Best, C., Schmidt, R., Schelin, J., Stanne, T. M., Bukau, B., Robinson, C. V., Witt, S., Mogk, A. & Clarke, A. K. (2009) *J Biol Chem* 284, 13519-32.
64. Andersson, F. I., Blakytyn, R., Kirstein, J., Turgay, K., Bukau, B., Mogk, A. & Clarke, A. K. (2006) *J Biol Chem* 281, 5468-75.
65. Karradt, A., Sobanski, J., Mattow, J., Lockau, W. & Baier, K. (2008) *J Biol Chem* 283, 32394-403.
66. Majeran, W., Friso, G., van Wijk, K. J. & Vallon, O. (2005) *Febs J* 272, 5558-71.
67. Derrien, B., Majeran, W., Wollman, F. A. & Vallon, O. (2009) *J Biol Chem* 284, 15408-15.
68. Huang, C., Wang, S., Chen, L., Lemieux, C., Otis, C., Turmel, M. and Liu, X.-Q. (1994) *Mol. Gen. Genetics* 244, 151-159.

69. Peltier, J. B., Ripoll, D. R., Friso, G., Rudella, A., Cai, Y., Ytterberg, J., Giacomelli, L., Pillardy, J. & Van Wijk, K. J. (2004) *J Biol Chem* 279, 4768-81.
70. Adam, Z., Rudella, A. & van Wijk, K. J. (2006) *Curr Opin Plant Biol* 9, 234-40.
71. Kim, Y. I., Levchenko, I., Fraczkowska, K., Woodruff, R. V., Sauer, R. T. & Baker, T. A. (2001) *Nat Struct Biol* 8, 230-3.
72. Peltier, J. B., Ytterberg, J., Liberles, D. A., Roepstorff, P. & van Wijk, K. J. (2001) *J Biol Chem* 276, 16318-27.
73. Olinares, P. D., Ponnola, L. & van Wijk, K. J. (2010) *Mol Cell Proteomics* 9.7, 1594-1615.
74. Sjogren, L. L., Stanne, T. M., Zheng, B., Sutinen, S. & Clarke, A. K. (2006) *Plant Cell* 18, 2635-49.
75. Peltier, J. B., Ytterberg, A. J., Sun, Q. & van Wijk, K. J. (2004) *J Biol Chem* 279, 49367-83.
76. Halperin, T., Zheng, B., Itzhaki, H., Clarke, A. K. & Adam, Z. (2001) *Plant Mol Biol* 45, 461-8.
77. Sokolenko, A., Lerbs-Mache, S., Altschmied, L. & Herrmann, R. G. (1998) *Planta* 207, 286-95.
78. Peltier, J. B., Cai, Y., Sun, Q., Zabrouskov, V., Giacomelli, L., Rudella, A., Ytterberg, A. J., Rutschow, H. & van Wijk, K. J. (2006) *Mol Cell Proteomics* 5, 114-133.
79. Rudella, A., Friso, G., Alonso, J. M., Ecker, J. R. & van Wijk, K. J. (2006) *Plant Cell* 18, 1704-21.
80. Zybaïlov, B., Friso, G., Kim, J., Rudella, A., Rodriguez, V. R., Asakura, Y., Sun, Q. & van Wijk, K. J. (2009) *Mol Cell Proteomics* 8, 1789-1810.

81. Zheng, B., MacDonald, T. M., Sutinen, S., Hurry, V. & Clarke, A. K. (2006) *Planta* 224, 1103-15.
82. Koussevitzky, S., Stanne, T. M., Peto, C. A., Giap, T., Sjogren, L. L., Zhao, Y., Clarke, A. K. & Chory, J. (2007) *Plant Mol Biol* 63, 85-96.
83. Kim, J., Rudella, A., Ramirez Rodriguez, V., Zybailov, B., Olinares, P. D. & van Wijk, K. J. (2009) *Plant Cell* 21, 1669-92.
84. Stanne, T. M., Sjogren, L. L., Koussevitzky, S. & Clarke, A. K. (2009) *Biochem J* 417, 257-68.
85. Sjogren, L. L., MacDonald, T. M., Sutinen, S. & Clarke, A. K. (2004) *Plant Physiol* 136, 4114-26.
86. Constan, D., Froehlich, J. E., Rangarajan, S. & Keegstra, K. (2004) *Plant Physiol* 136, 3605-15.
87. Kovacheva, S., Bedard, J., Patel, R., Dudley, P., Twell, D., Rios, G., Koncz, C. & Jarvis, P. (2005) *Plant J* 41, 412-28.
88. Kovacheva, S., Bedard, J., Wardle, A., Patel, R. & Jarvis, P. (2007) *Plant J* 50, 364-79.
89. Akita, M., Nielsen, E. & Keegstra, K. (1997) *J Cell Biol* 136, 983-94.
90. Nielsen, E., Akita, M., DavilaAponte, J. & Keegstra, K. (1997) *EMBO J* 16, 935-946.
91. Wu, H., Ji, Y., Du, J., Kong, D., Liang, H. & Ling, H. Q. (2010) *Ann Bot* 105, 823-33.
92. Shikanai, T., Shimizu, K., Ueda, K., Nishimura, Y., Kuroiwa, T. & Hashimoto, T. (2001) *Plant Cell Physiol* 42, 264-73.
93. Majeran, W., Olive, J., Drapier, D., Vallon, O. & Wollman, F. A. (2001) *Plant Physiol* 126, 421-33.
94. Majeran, W., Wollman, F. A. & Vallon, O. (2000) *Plant Cell* 12, 137-50.

95. Schelin, J., Lindmark, F. & Clarke, A. K. (2002) *Microbiology* 148, 2255-65.
96. Barker-Astrom, K., Schelin, J., Gustafsson, P., Clarke, A. K. & Campbell, D. A. (2005) *Arch Microbiol* 183, 66-9.
97. Sokolenko, A., Pojidaeva, E., Zinchenko, V., Panichkin, V., Glaser, V. M., Herrmann, R. G. & Shestakov, S. V. (2002) *Curr Genet* 41, 291-310.

CHAPTER TWO

Subunit composition, evolution and implications of an asymmetric plastid ClpPR protease*

INTRODUCTION

Clp proteases are oligomeric self-compartmentalized serine-type proteases found in eubacteria and in mitochondria and plastids of eukaryotes (1). The Clp proteolytic core in *E. coli* consists of 14 identical proteolytic ClpP subunits, organized as a barrel-like structure with two stacked heptameric rings (2). The Clp protease system greatly expanded and diversified in plastids of higher plants with five ClpP (ClpP1, ClpP3-6) and four non-catalytic ClpR proteins (ClpR1-4), assembled in a single ~350 kDa complex (3) with unknown subunit stoichiometry and ring composition. Except for the plastid-encoded ClpP1, all Clp proteins in plants are nucleus-encoded (3, 4). Genetic and phenotypic analyses of various Clp mutants in plants showed that the proteolytic and non-proteolytic subunits have differential functional contributions and all of them, except ClpR1, are essential for embryo or seedling development. ClpP1 in tobacco is essential for shoot development (5, 6), whereas complete loss of *CLPP5* is embryo lethal and complete loss of *CLPR2*, *CLPR4* results in seedling lethality (7). A knockdown mutant in *CLPR2* (*clpr2-1*) and antisense lines against *CLPP4* and *CLPP6* in *Arabidopsis* exhibited delayed chloroplast and plant development and a virescent or variegated phenotype (8-10). The *Arabidopsis* EMS mutant *clpr1-1* with a premature stop codon showed a virescent phenotype (11), but overexpression of *CLPR3* in this

* **Olinares, P.D.B., Kim, J., Davis, J.I., and van Wijk, K.J.** (2010). Subunit composition, evolution and implications of an asymmetric plastid ClpPR protease. Manuscript submitted. PDBO performed most experiments; JK generated the *clpp3-1* mutant and the ClpP3-StrepII tagged complemented lines; JID provided support for the phylogenetic analysis; PDBO and KJW planned the experiments and wrote the paper.

background led to complementation, suggesting that ClpR1 is redundant to ClpR3 (7). Knowing the stoichiometry and ring localization of the different ClpPR proteins is a prerequisite for understanding the functional contribution of the different ClpP and ClpR protein to the protease complex. Moreover, Clp protease substrates are delivered via the adaxial sides of the protease core and this adaxial surface is determined by Clp ring composition and organization. ClpP3-5 and ClpR1-4 have extended C-termini (upto 52 residues beyond the *E. coli* ClpP C-terminus). Previously, we predicted that these extended C-termini can form α -helical fragments that can fold over the adaxial surfaces, affecting docking of ClpC/D chaperones and ClpT1,2 (3). However, it is not known if these C-termini are indeed stable when the ClpPR proteins are assembled in the core complex. It is known for *E. coli* that the N-terminus of ClpP is cleaved autocatalytically (12).

In this study, the chloroplast Clp protease complex was affinity-purified from *clpr4* and *clpp3* *Arabidopsis* null mutants complemented with C-terminal StrepII-tagged versions of *CLPR4* and *CLPP3*, respectively. The subunit stoichiometry was determined by mass spectrometry-based absolute quantification using stable isotope-labeled proteotypic peptides generated from a synthetic gene. We also show that the C-terminal extensions of the plastid ClpP/R subunits are conserved across plants and are not proteolytically removed. Thus, the chloroplast Clp core is comprised of diverse Clp subunits that assemble into a single asymmetric complex, unlike any bacterial Clp core, with two rings exhibiting differential proteolytic functionalities and adaxial surface features for chaperone interaction.

RESULTS AND DISCUSSION

To understand the functional significance of the diversification of Clp proteins in plastids, we determined the ring composition and absolute subunit stoichiometry of the plastid-localized Clp protease complex in *Arabidopsis* by affinity purification coupled with mass spectrometry-based quantification of the subunits (Figure 2.1 for an overview of the workflow). Native gel separations of the chloroplast stroma proteome indicated that the Clp core complex migrates at 350 kDa (3, 13) and its component rings at 180-200 kDa (10). To facilitate purification of the ClpPR core, as well as its individual rings, we generated two different transgenic plant lines expressing tagged ClpR4 or ClpP3 subunits by complementing homozygous *CLPR4* and *CLPP3* null mutants (*clpr4-1* and *clpp3-1*) with respectively a 1x35S-driven *CLPR4* or *CLPP3* cDNA fused to C-terminal StrepII tags (Supplemental Figure 2.1). The eight-residue StrepII tag was attached to the C-terminus rather than the N-terminus to prevent interference with the N-terminal chloroplast targeting peptide. A C-terminal tag also likely minimizes interference with the Clp core function since the N-terminal domains of ClpP modulate chaperone interactions and substrate delivery in the *E. coli* system (14). The complemented mutants did grow on soil and exhibited a wild-type phenotype (Supplemental Figure 2.1A). Moreover, complementation was achieved at the protein level, with chloroplast proteins accumulating at wild-type levels (Supplemental Figure 2.1B). We also employed other C-terminal tagging systems, including His₆ and a tandem affinity tag; whereas partial or complete phenotypic complementation was achieved, the isolated Clp complex was never sufficiently pure as judged by MS/MS analysis (data not shown); these efforts will therefore not be further discussed.

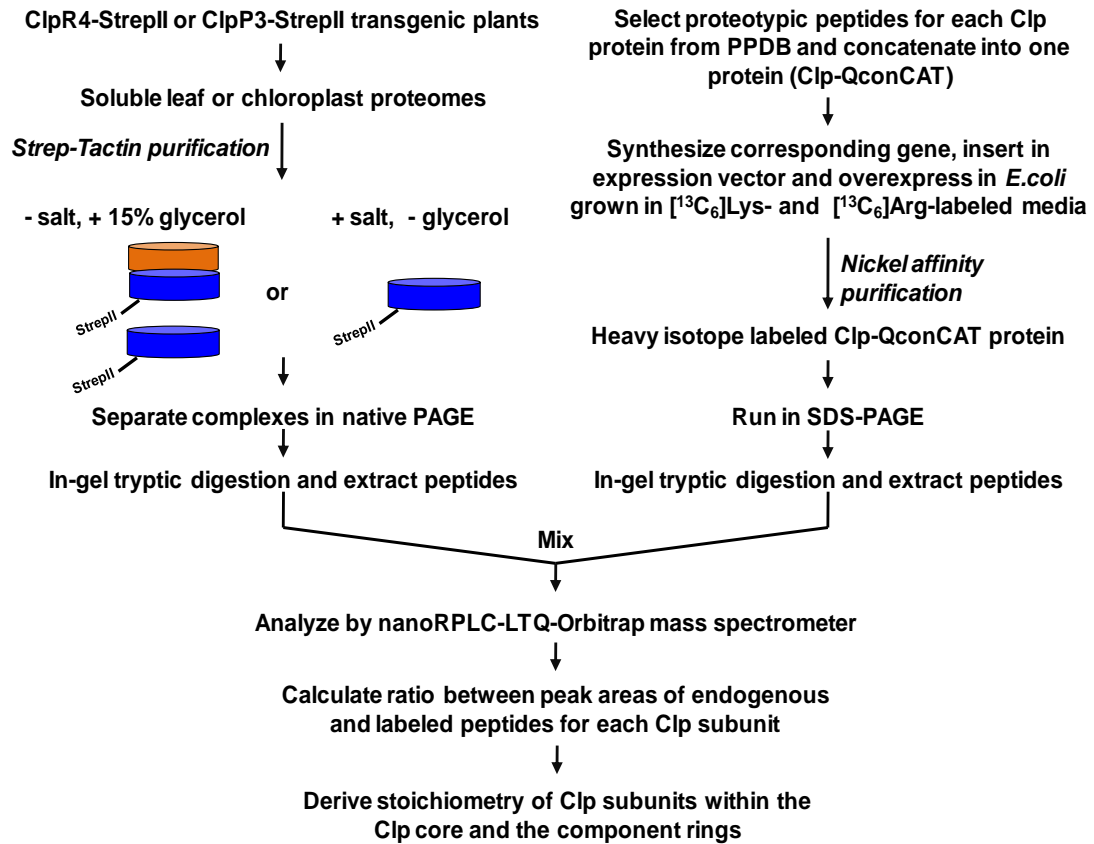


Figure 2.1. Workflow for determination of the ring composition and absolute subunit stoichiometry of the tetradecameric plastid-localized ClpPR protease complex in *Arabidopsis* using mass-spectrometry based absolute quantification with stable isotope labeled proteotypic peptides generated in *E. coli*. PPDB is the Plant Proteome Database (<http://ppdb.tc.cornell.edu/>).

After optimization of purification conditions (not shown), immunoblot analyses with anti-StrepII serum showed that we obtained a tagged ~350 kDa Clp core complex and a tagged ~200 kDa Clp ring from both the ClpR4 and ClpP3 StrepII-tagged lines (Figure 2.2A). In the case of ClpR4, some monomeric StrepII-R4 accumulated when using total leaf extracts as starting material for purification, rather than isolated chloroplast stroma, indicating some destabilization or accumulation outside of the plastid (Figure 2.2A). High resolution and high accuracy tandem mass

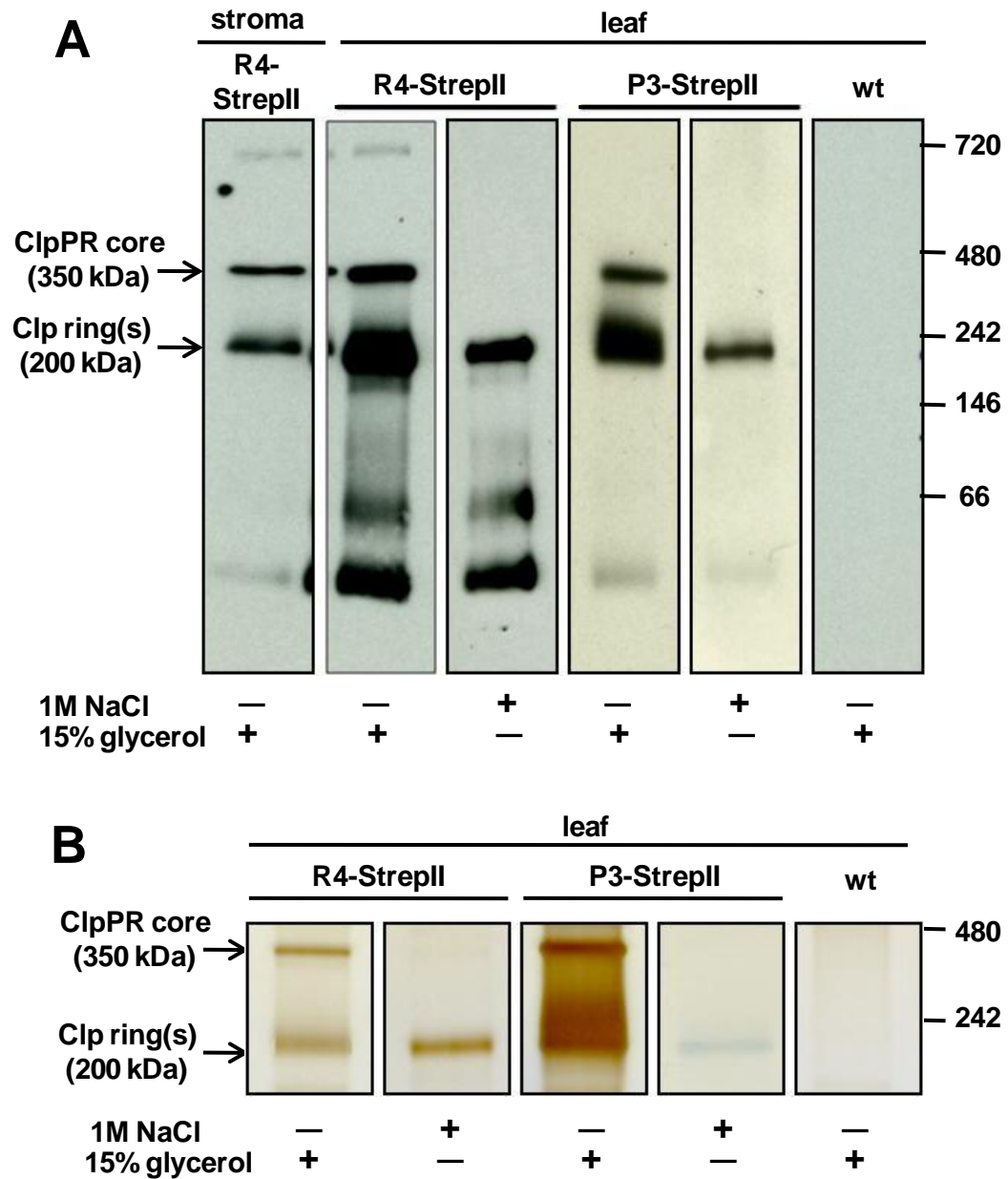


Figure 2.2. Native PAGE analysis of affinity purified StrepII-tagged Clp assemblies (350 and 200 kDa) from soluble chloroplast (stroma) or leaf extracts of transgenic plants expressing R4-StrepII or P3-StrepII. **A,B** On-column salt incubation in the absence of glycerol dissociated the individual rings from the core complex facilitating the purification of only the StrepII-tagged Clp ring. **A**) Immunoblot analyses with an anti-StrepII antibody. The 30 kDa bands observed in both transgenic lines correspond to unprocessed forms of R4-StrepII and P3-StrepII and the 66 kDa band corresponds to dimers. A faint band at 700 kDa indicates cross-interaction of the antibody to the Cpn60 complex. **B**) Silver stain of a gel separation with similar samples shown in panel **A**. The complete gel is shown in Supplemental Figure 2.2A.

spectrometry (MS/MS) of the ~350 kDa and ~200 kDa gel bands (Figure 2.2B, Supplemental Figure 2.2 and Supplemental Table 2.1), identified all nine ClpP and ClpR proteins in both. This suggested that the ~200 kDa band contained a mixture of heptameric rings resulting from destabilization of the Clp core. To obtain highly purified individual heptameric rings, we incubated the column-bound tagged 350 kDa Clp complexes in 1M NaCl without glycerol. After subsequent column washing, elution with buffer containing 2.5 mM desthiobiotin and 15% glycerol released a single StrepII-tagged Clp ring at ~200 kDa for both R4 and P3 tagged lines (Figure 2.2). MS/MS analyses showed that the purified 200 kDa Clp ring from the StrepII-tagged ClpR4 plants contained only ClpP1 and ClpR1-4 whereas the purified ring from StrepII-tagged ClpP3 plants contained only ClpP3, P4, P5, and P6 (Supplemental Table 2.1). Clearly, the on-column salt incubation separated the tagged ring-containing tetradecameric Clp core into two individual rings, similar as for the homotetradecamer in *E. coli* (15, 16), and that subsequent washing removed the untagged ring. Importantly, this showed that each ClpP and ClpR subunit is only present in one of the rings.

The identical ClpP rings in *E. coli* reassociate when glycerol is added after initial incubation in high salt conditions (even at >2 M KCl) (16). In contrast, no 350 kDa complex was observed upon addition of glycerol to the purified R4-tagged or P3-tagged rings from the salt incubation experiments (not shown), suggesting that identical plastid Clp rings do not form stable double-ring complexes. This is consistent with the essential nature of most ClpPR subunits (7, 10). Overall, these results establish that the plastid-localized ClpP and ClpR subunits assemble into one core complex which is comprised of a ClpP1/R1/R2/R3/R4 and a ClpP3/P4/P5/P6 ring. Furthermore, inter-ring associations within the plastid Clp core are primarily ionic and can be perturbed by high salt concentrations, similar as in the bacterial Clp complex.

The plant specific ClpT1,2 proteins (with high similarity to the N-terminus of the ClpC chaperones (3)) were consistently identified with the 350 kDa core, but the on-column salt incubations removed these subunits (Supplemental Table 2.1) supporting the peripheral nature of their interaction with the ClpPR subunits, as previously suggested based on homology modeling (3).

To quantify the amount of each Clp subunit within the purified Clp assemblies and to determine the stoichiometry among the subunits, we implemented an innovative method (assigned quantification concatamer or QconCAT) based on spiking the sample with an equimolar set of stable isotope-labeled signature proteotypic tryptic peptides generated from a synthetic gene expressed in *E. coli* (17) (Figure 2.1). Proteotypic peptides are peptides that uniquely match to a protein of interest and that possess physico-chemical characteristics that are favorable for frequent detection in specific mass spectrometry platforms (18) (in our case a nanoLC-ESI-LTQ-Orbitrap instrument). First, we selected the two most optimal proteotypic peptides for each of the 16 members of the Clp family from extensive mass spectrometry analysis of *Arabidopsis* chloroplasts available at our Plant Proteome Database (<http://ppdb.tc.cornell.edu/>). Peptide sequences containing residues prone to post-translational modifications (eg methionine and cysteine) were avoided, except for one P3 peptide that contained a methionine residue (see Figure 2.3A). A gene (designated here as Clp-QconCAT gene) encoding for 29 Clp-derived peptides, plus a number of additional peptides matching to cytosolic or plastidic household proteins as controls, and a C-terminal His₆ tag for affinity purification (Supplemental Table 2.2) was then synthesized. The resulting 1,506 bp gene was expressed from a plasmid in *E. coli* cells grown in media containing either [¹³C₆]-lysine and [¹³C₆]-arginine (for stable isotope labeled Clp-QconCAT protein) or unlabeled lysine and arginine (for unlabeled Clp QconCAT protein) (in addition to all other unlabeled amino acids). The Clp-QconCat

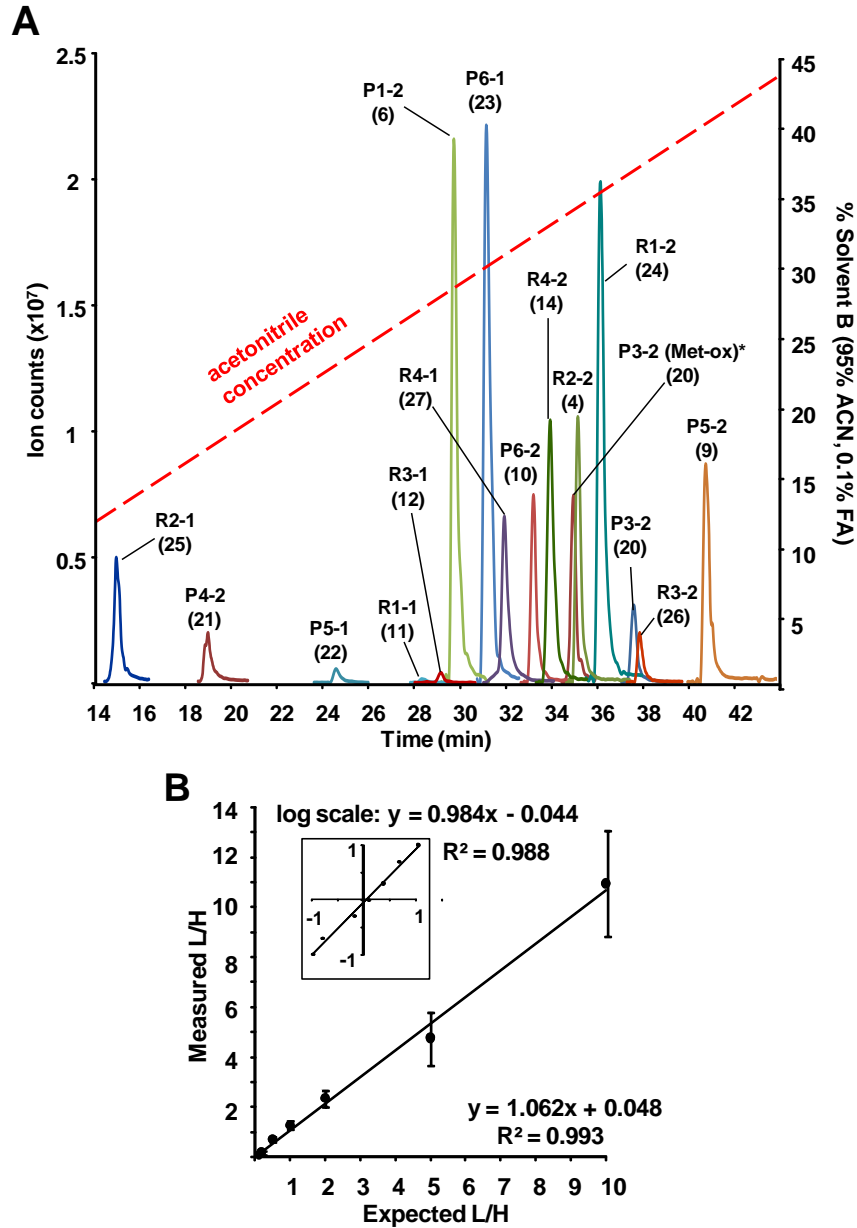


Figure 2.3. Optimization and accuracy of the QconCAT quantification method. **A)** Representative extracted ion chromatograms (XICs) of the signature peptides for Clp subunits derived from 100 fmol Clp-QconCAT protein. Four out of the nine Clp core subunits can be quantified with two peptides and the rest with one peptide. An XIC plot of all quantifiable Clp-QconCAT peptides is shown in Supplemental Figure 2.3. The numbers in parentheses designate the order in the Clp-QconCAT construct (Supplemental Table 2.3). All quantifiable Clp-QconCAT peptides eluted within the first 45 min. **B)** The average linear MS instrument response for the Clp-QconCAT protein in varying concentrations (10 fmol to 1 pmol of light Clp-QconCAT mixed with constant 100 fmol of heavy Clp-QconCAT). $n=32$ peptides quantified, error bars = S.D. The inset panel shows the same plot on a logarithmic scale. For detailed linear response for each peptide, see Supplemental Table 2.3.

protein was then isolated by Ni²⁺-NTA affinity purification. In-gel tryptic digestion and MS/MS analysis of the isolated protein confirmed its purity and identity (not shown). Quantitative LC-MS analysis was optimized using heavy (H, labeled) and light (L, unlabeled) affinity-purified *E. coli* Clp-QconCAT protein. L/H ratios for each peptide were calculated from peak areas of extracted ion chromatograms (XICs). Peptide retention times (Figure 2.3A) were reproducible (median standard deviation was 0.2 min). Using 100 fmol heavy ClpQconCAT protein and variable amounts of its light version, a linear MS instrument response was observed within two orders of magnitude (0.1 to 10 L/H ratios) with high precision at the protein level (Figure 2.3B) and for each of the identified Clp-QconCAT peptides (Supplemental Table 2.3). Overall, 32 out of the 37 QconCAT component peptides could be detected and quantified.

We then used the ¹³C₆-labeled QconCAT protein to quantify proteins in the purified 350 kDa core, as well as the mixed and purified individual 200 kDa Clp rings. These gel-separated complexes were each in-gel digested and spiked with in-gel digested ¹³C₆-labeled ClpQconCAT protein. From the measured sample-to-standard (L/H) peak area ratios, we then derived the stoichiometry of the Clp subunits for each ring (Table 2.1). We chose ClpR4 and ClpP6 as the reference for the ClpP1/R and ClpP3-6 rings, respectively, since these proteins were reliably quantified with two peptides at high precision. The resulting stoichiometry for the two rings were ClpP1:R1:R2:R3:R4 = 3:1:1:1:1 and ClpP3:P4:P5:P6 = 1:2:3:1. The ratio of the total moles of each ring within the complex is 1 confirming that the component rings indeed assemble in a 1:1 configuration (Supplemental Table 2.4).

Our previous homology models of the plastid Clp core complex indicated that the extended C-termini of ClpP3-6 and ClpR1-3 (as compared to *E. coli* ClpP) may form two to four α -helical fragments that can fold over the adaxial side of the core,

Table 2.1. Stoichiometry of the Clp subunits within each Clp ring^a

Clp	Peptide ^e	Clp core ^b		Clp ring (no salt added) ^c		Clp ring (with salt incubation) ^d		Summary
		AVE	SD	AVE	SD	AVE	SD	
ClpP1/R ring								
P1	Q6	2.9	0.4	3.3	0.2	3.0	0.4	3
R2	Q25	1.3	0.2	1.6	0.3	1.3	0.1	1
	Q4	0.8	0.1	0.8	0.1	0.8	0.1	
R3	Q26	1.0	0.1	1.0	0.1	1.0	0.4	1
R4	Q27	0.9	0.2	0.9	0.1	0.9	0.02	1
	Q14	0.9	0.1	1.1	0.1	1.1	0.02	
		n=4		n=6		n=3		
ClpP3-6 ring								
P3	Q20	1.2	0.2	0.9	0.2	1.0	0.2	1
P4	Q8	2.2	0.1	2.5	0.2	2.5	0.2	2
P5	Q22	2.4	0.7	2.3	0.4	2.3	0.4	3
	Q9	3.1	0.8	3.1	0.9	3.3	0.8	
P6	Q23	1.4	0.3	1.1	0.1	1.1	0.05	1
	Q10	1.1	0.3	1.0	0.4	0.9	0.05	
		n=4		n=5		n=4		

^aObtained from normalization against the average peak ratios of ClpR4 or ClpP6 peptides. See Supplemental Tables 2.4 and 2.5 for details on peak ratios.

^bFrom affinity-purified 350-kDa Clp core complex from both R4-StrepII and ClpP3-StrepII transgenic lines.

^cFrom the 200-kDa Clp ring from the affinity purifications without salt incubation.

^dThe ClpP1/R1-4 ring was purified from the ClpR4-StrepII line and the ClpP3-6 ring was obtained from the ClpP3-StrepII line.

^eOrder by which these proteotypic peptide appeared starting from the N-terminus of the Clp-QconCAT construct. See Supplemental Table 2.2 for details.

thereby potentially interfering with docking of the ClpC/D chaperones and/or ClpT1,2 (3) (in particular the longer extensions of ClpP3 (52 aa) and P4 (41 aa) (Supplemental Figures 2.4 and 2.5). Sequence alignments to other fully sequenced plant species showed that these extensions are conserved in higher plants (Supplemental Figures 2.4 and 2.5). To determine that these extended C-termini were indeed still present in the assembled complexes and not proteolytically cleaved, we mapped all MS/MS

identified peptides in the Clp rings and core to their protein sequences. Results showed that the extended C-termini are indeed part of the assembled ClpPR complex (Supplemental Figures 2.4 and 2.5). Multiplying the subunit stoichiometries with the length of each C-terminal extension, shows a 2-fold higher amount of total sequence extension (174 versus 89 aa) in the ClpP3-6 ring than in the ClpP1/R ring; this suggests potentially stronger effects on adaxial docking for the ClpP3-6 ring.

Non-photosynthetic bacteria do not possess ClpR genes and have either one to three ClpP genes (1). In contrast, (photosynthetic) cyanobacteria harbor three ClpP and one ClpR gene. In *Synechococcus elongatus* sp. PCC 7942, these Clp gene products are assembled in two different Clp core complexes; one complex is not essential and consists of a mixture of ClpP1 and ClpP2 in unknown stoichiometry and the other complex is essential and consists of ClpP3 and ClpR in a 3:4 ratio in both heptameric rings (19, 20). Homology searches against the recently sequenced plant genomes (maize, rice, poplar, the spikemoss *Selaginella* and the moss *Physcomitrella*) and the green alga *Chlamydomonas reinhardtii* revealed that all plastid ClpPR core subunits in *Arabidopsis* have orthologs in plants and in green alga (except for *CLPP4* in the case of *C. reinhardtii*) (Supplemental Figure 2.6). This indicates that the expansion of the Clp family occurred early in the green lineage and is maintained throughout the evolution of land plants. Unlike the essential ClpP3/R core in *Synechococcus*, the two heptameric rings of the essential plastid ClpPR core complex in *Arabidopsis* are composed of completely different subunits, and consequently the adaxial sides (providing docking for Clp chaperones and ClpT1,2) are different. Careful phylogenetic analysis shows that the subunits of the ClpP1,R1-4 ring originate from the essential ClpP3-ClpR genes (Figure 2.4; Supplemental Figure 2.6) (21). In particular, the plant ClpP1 and ClpR2 genes originate from an ancestral cyanobacterial ClpP3, whereas ClpR1, ClpR3, and ClpR4 arose from duplications of the ancestral

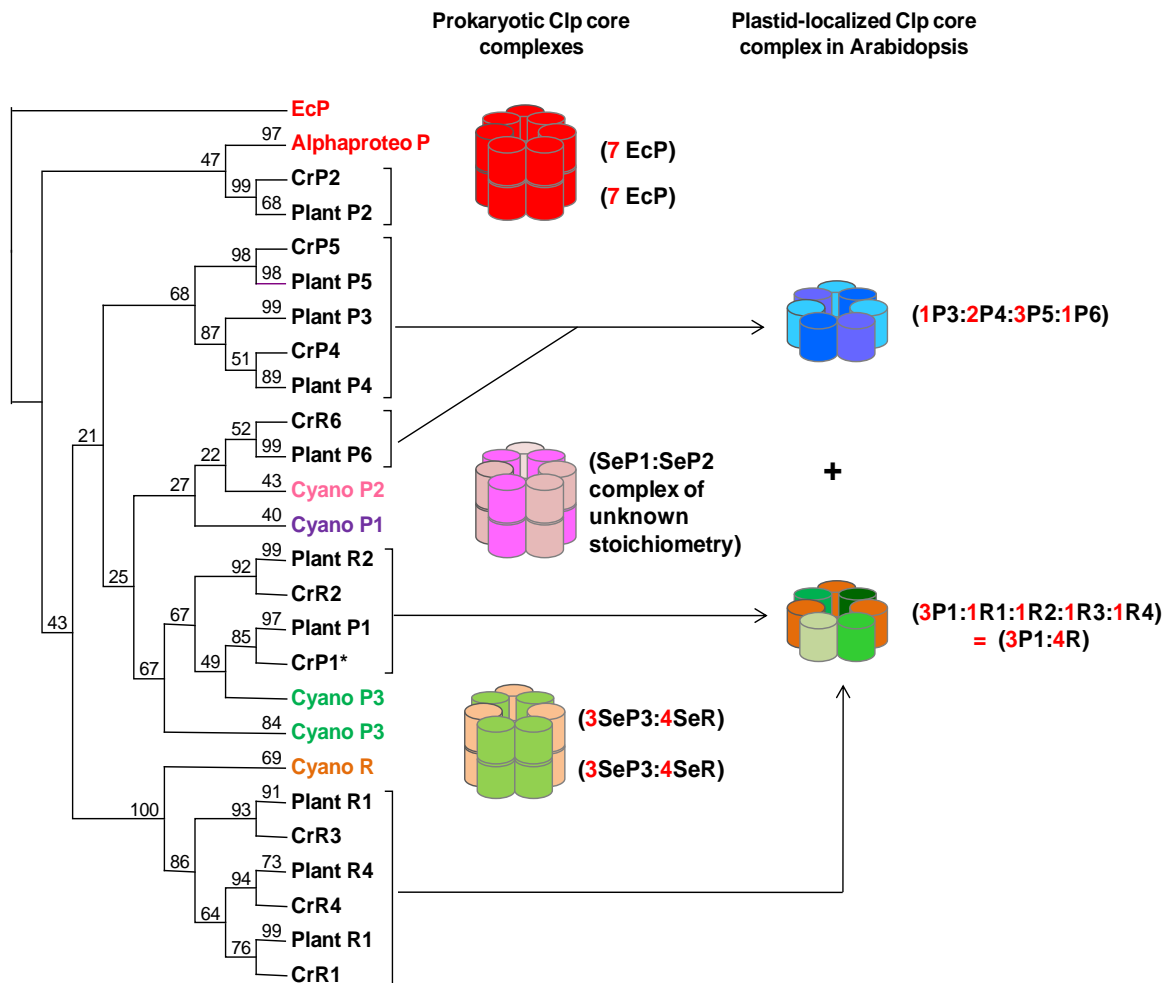


Figure 2.4. Evolutionary relationships among the Clp protease assemblies in prokaryotes and plant plastids. Phylogenetic analysis of the ClpP/R proteins from proteobacteria, photosynthetic bacteria and plastids employing the Maximum Likelihood (ML) approach (23) with the General Time Reversible (GTR) model for nucleotide substitution (24) and a gamma model of substitution rate heterogeneity (25). 118 Clp cDNA sequences from bacteria (*E. coli* and 2 alphaproteobacteria species), cyanobacteria (five species), green alga (*Chlamydomonas reinhardtii*, Cr) and plants (seven species) were aligned and analyzed. The Cr ClpP1* is the plastid-encoded ClpP1 in green alga with the large insertion (IS1) sequence removed. Support values are shown from 500 non-parametric bootstrap replicates. The *E. coli* ClpP was designated as the outgroup. Bacterial Clp subunits are highlighted and their corresponding homo- or heterooligomeric assembly states are represented: *E. coli* and proteobacterial ClpP (red) and cyanobacterial ClpP1 (purple), ClpP2 (fuchsia), ClpP3 (green) and ClpR (orange). The oligomeric state, composition and stoichiometry of the plastid-localized Clp complex are then linked to the Clp assemblies from its prokaryotic progenitors.

cyanobacterial ClpR (Figure 2.4; Supplemental Figure 2.6). Moreover, the duplication event that yielded *CLPP3* and *CLPP4* in land plants occurred early in the terrestrial adaptation as both genes are already found in moss (Supplemental Figure 2.6). It is striking that the chloroplast P1/R ring has an overall 3:4 ClpP:R ratio similar to that in the cyanobacterial ClpP3/R core complex. Intriguingly, the 3:4 active:inactive composition is also observed in heptameric component rings of the eukaryotic proteasome (22), suggesting a functional restraint on the composition of these various proteolytic assemblies. The origins of the subunits of the plastid ClpP3-6 ring is unclear; phylogenetic analyses suggest that ClpP6 is related to cyanobacterial ClpP2 (Figure 2.4).

From its component ring stoichiometries, the plastid Clp core complex is expected to exhibit different proteolytic capacities (one heptameric ring has seven, whereas the other has three catalytic subunits). The lower concentration of catalytic sites in the ClpP1/R1-4 ring is likely to exhibit slower rates of proteolysis which may or may not be the rate limiting step in the sequence of substrate selection, delivery and unfolding. In cyanobacteria, the inclusion of the catalytically inactive ClpR in the ClpP3/R complex appeared not to be rate-limiting, since restoration of the catalytic sites and replacement of the whole internal domain of ClpR with that of ClpP3 did not enhance the proteolytic activity of the resulting ClpP3/chimeric ClpR complex relative to wildtype (19).

In all studied organisms, substrates for the Clp core complex are delivered through ATP-dependent ClpA- or ClpX-type chaperones; these chaperones form hexameric rings that dock onto the adaxial sides of the Clp core complex (26). This interaction needs to accommodate the different oligomeric states (heptamer and hexamer) and, in *E. coli* this interaction has been shown to be dynamic and reversible (14, 27). The essential ClpP3/R complex in cyanobacteria only interacts with ClpC, a

homologue of *E. coli* ClpA, whereas the non-essential ClpP1P2 complex only interacts with ClpX (20). Higher plant plastids do not contain any of the three predicted ClpX proteins (instead these ClpX proteins are confined to the plant mitochondria (3, 28)) but they do contain three ClpA homologues, namely ClpC1,C2,D. We therefore speculate that the chloroplast ClpP1/R ring provides the docking surface for ClpC/D interaction. In contrast, the ClpP3-6 ring with unclear origin, is most likely the preferred interaction partner for ClpT1,2 since these peripheral subunits are only observed in higher plants. We speculate that the ClpP3-6 ring co-evolved with ClpPT1,2 and together represent a specific adaptation to the plastid proteome (for a discussion on the possible roles of ClpT1,2 - see (3)). The extended C-terminal sequences of in particular ClpP3 and ClpP4 may play a significant role in regulating the interaction with ClpT1,2 and ClpC1,2D, and thereby contributing to substrate selection and delivery.

MATERIALS AND METHODS

Plant growth conditions Plants were grown on agar plates with 0.5x Murashige and Skoog medium with 2% sucrose under a 10/14-h light/dark cycle at 40 $\mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$. Plants were grown on soil under a 10/14-h light/dark cycle at 100-120 $\mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$.

Complementation of *clpr4-1* with *CLPR4-StrepII* and *clpp3-1* with *CLPP3-StrepII*

The primers used in this study are shown in Supplemental Table 2.7. *CLPR4* and *CLPP3* cDNAs were PCR amplified and the C-terminal StrepII sequence was attached using Platinum Pfx polymerase (Invitrogen) or GOTaq Polymerase (Promega). The PCR products *CLPR4-StrepII* and *CLPP3-StrepII* were then subcloned into pENTR-D-TOPO (Invitrogen) or pCR8-TOPO (Invitrogen), respectively. The resulting

pENTR-D-TOPO construct was then digested with *Mlu*I (New England Biolabs) and the *CLPR4-StrepII*-containing fragment was gel-purified and subcloned into the Gateway destination vector pEARLEYGATE100 (29) using LR Clonase™ Enzyme Mix (Invitrogen). The same subcloning procedure was employed with the *CLPP3-StrepII* sequence in pCR8-TOPO without any restriction digestion step. Competent cells of *Agrobacterium tumefaciens* strain GV3101 were transformed with pEarleyGate100-*CLPR4-StrepII* or pEarleyGate 100-*CLPP3-StrepII*. Heterozygous *clpr4-1* or *clpp3-1* plants were used for *Agrobacterium*-mediated plant transformation by the floral dip method (30). Transformants were screened using 10 µg/mL DL-phosphinothricin (BioWorld). Complemented plants that are homozygous to the T-DNA insertion were selected and verified by PCR genotyping.

Affinity purification of StrepII-tagged Clp assemblies Total leaf material from transgenic plants were ground in liquid nitrogen and solubilized in 50 mM HEPES-KOH, pH 8.0, 15% glycerol and 10 mM MgCl₂ (extraction buffer, EB) with protease inhibitor cocktail. The suspension was then filtered in miracloth and spun at 100,000xg. The supernatant was concentrated, loaded on a Strep-Tactin column (IBA, Germany), and washed with EB for isolation of the Clp core or washed with EB without glycerol for purification of Clp rings. For isolating Clp rings, the column-bound Clp core was incubated in EB with 1 M NaCl for one hour and then subsequently washed with EB without glycerol. Elution was performed using EB with 2.5 mM desthiobiotin.

QconCAT design, expression and purification The selected ClpQconCAT peptides (Supplemental Table 2.2) were concatenated and was flanked by a leader N-terminal sequence (MAGKVIR-) and a C-terminal sequence (AGKVICSAEGSK-) as in (31).

The corresponding DNA sequence for the Clp-QconCAT peptides was synthesized (Entelechon GmbH, Germany). Cloning, protein expression in stable isotope-labeled or unlabeled growth media and protein purification was performed as previously described (17). L-[¹³C₆]Lys and L-[¹³C₆]Arg (97-99% enrichment, CIL Inc.) were used for labeling.

Protein extractions, gel analysis, immunoblots and in-gel trypsin digestion Prior to SDS-PAGE separation, guanidine was removed from the purified Clp-QconCAT as in (32). The purified Clp-QconCAT pellet was then resuspended with sonication in 50 mM Tris-HCl pH 6.8, 1% SDS prior to protein analysis.

For total leaf extractions (with SDS), leaf material was ground in liquid nitrogen, resuspended in 50 mM Tris-HCl pH 6.8, 2% SDS and protease inhibitor cocktail and filtered through a frit column (ThermoFisher Scientific) by a quick centrifugation. The filtrate was collected for protein analysis.

Protein concentrations were determined using the BCA Protein Assay Kit (ThermoScientific). Light blue native PAGE was performed for separation of affinity-purified Clp assemblies using the NativePAGE Novex gel system (Invitrogen) with pre-cast 4-16% acrylamide Bis-Tris gels (Invitrogen). For denaturing protein separations (*e.g.*, Clp-QconCAT and total leaf samples), precast 10.5-14% gradient acrylamide Laemmli gels (Biorad) were used. For immunoblots, proteins were blotted onto PVDF membranes and probed with antibodies using chemiluminescence for detection, following standard procedures. The following antibodies were used: anti-StrepII (GenScript), anti-RH3 and anti-OEC23 (from Dr. Alice Barkan) and anti-PsaF (from Dr. Hendrik Scheller). For MS analysis, visible protein bands were in-gel digested with trypsin as previously described (33).

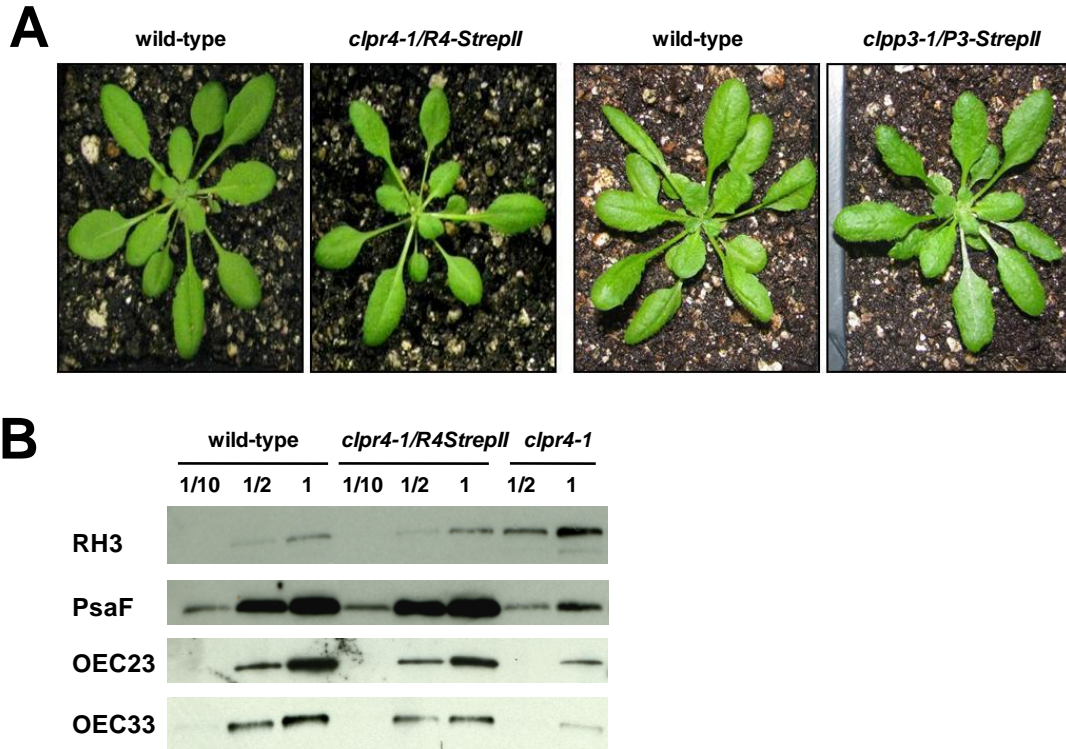
LC-MS analysis and data processing After in-gel protein tryptic digestion, the extracted peptides were resuspended in 2% formic acid and an aliquot was automatically loaded on a guard column (LC Packings MGU-30-C18PM) via an autosampler followed by separation on a PepMap C₁₈ reverse-phase nanocolumn (LC Packings nan75-15-03-C18PM) using 85-min gradients with 95% water, 5% ACN, 0.1% FA (solvent A) and 95% ACN, 5% water, 0.1% FA (solvent B) at a flow rate of 200 nl/min. The gradient proceeds as follows for solvent B: 0 to 6 min – 10%, 40 min – 45%, 48 to 50 min - 95%, 51 to 57 min - 10%, 65 to 67 min - 95%, 68 min to 85 min – 10%. Two blanks were run after every sample. Each sample was analyzed in triplicate. The MS acquisition cycle consisted of a survey scan in the Orbitrap with a set mass range (250 to 1800 *m/z*) at resolution 60,000 followed by five data-dependent MS/MS scans acquired in the LTQ. Dynamic exclusion was set with exclusion size, 500; repeat count, 2; repeat duration, 30 s; exclusion time, 90 s; exclusion window, ± 0.6 *m/z*. Target values were set at 5×10^5 and 10^4 for the survey and tandem MS scans, respectively.

The extracted ion chromatograms (XICs) were calculated using the XISsee software (34) with a retention time tolerance of ± 2.5 min and *m/z* tolerance of ± 10 ppm. Peak area quantification was done on SILAC mode with [¹³C₆]Lys and [¹³C₆]Arg. Peak smoothing was enabled and default parameters for peak detection were used. Peak areas for peptides identified with multiple charge states or modifications were combined prior to calculating the light-to-heavy ratios. To generate the XICs for all Clp-QconCAT peptides in one plot (Figure 2.3 and Supplemental Figure 2.3) the peak profiles were obtained from an MS raw file using MASIC (35) (<http://www.pnl.gov/>), exported in Microsoft Excel and then reconstructed.

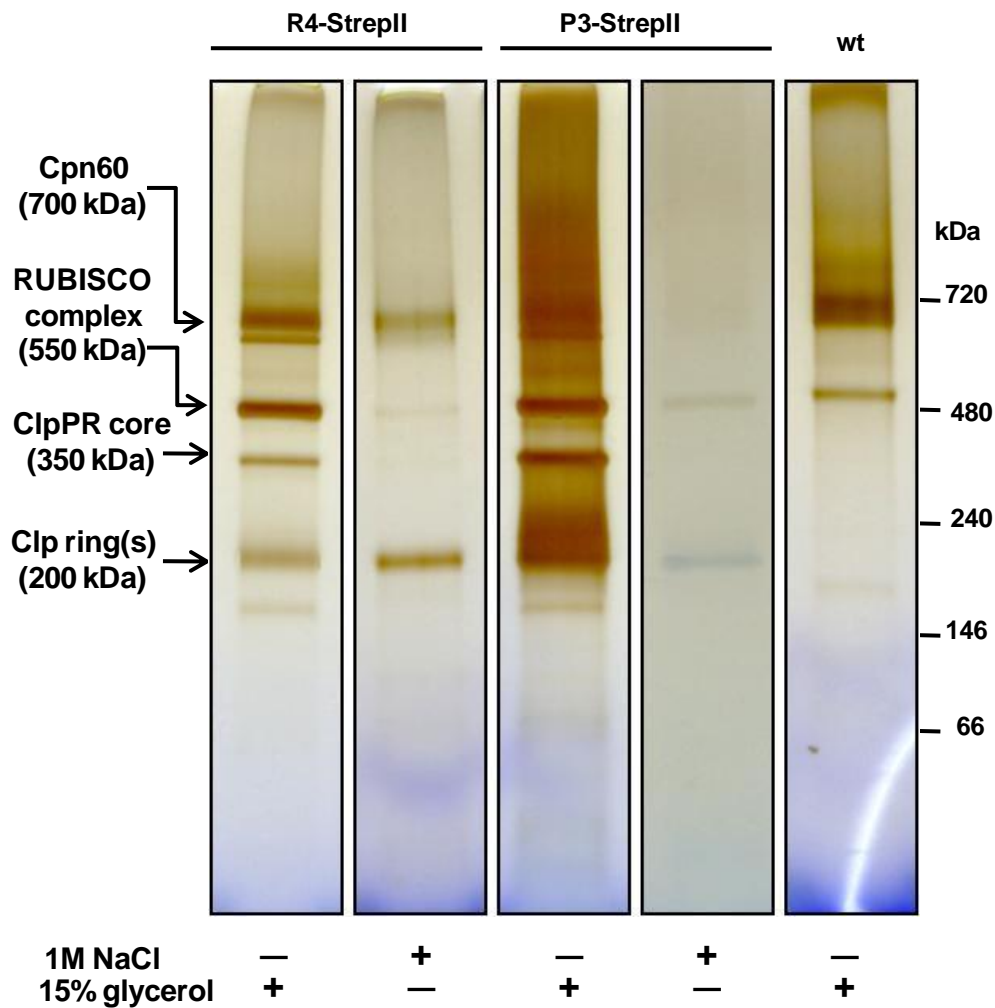
For peptide identifications, MS raw files were processed using our workflow (36). For samples with spiked stable isotope-labeled QconCAT, [$^{13}\text{C}_6$]Lys- and [$^{13}\text{C}_6$]Arg-labeling were included in the database searches as variable modifications. All filtered results were uploaded into the Plant Proteomics DataBase, PPDB (<http://ppdb.tc.cornell.edu/>) (37).

Phylogenetic tree construction The accession numbers for the sequences used for phylogenetic analyses are shown in Supplemental Table 2.6. The cDNA and protein sequences for the prokaryotic Clp proteins and for Chlamydomonas, poplar and moss Clps were obtained from KEGG (<http://www.genome.jp/kegg/>), Uniprot database (<http://www.uniprot.org/>), NCBI (<http://www.ncbi.nlm.nih.gov/>) and EMBL (<http://www.ebi.ac.uk>). The genome sequencing databases for Arabidopsis (<http://www.arabidopsis.org/>), rice (<http://rice.plantbiology.msu.edu/>) maize (<http://www.maizesequence.org/index.html>) and Selaginella (<http://genome.jgi-psf.org>) were used to obtain the respective plant Clp sequences. Protein sequences were initially aligned using TCOFFEE (http://toolkit.tuebingen.mpg.de/t_coffee) employing Mlalign_id_pair and mslow_pair alignments. The resulting protein alignment was used as a template to align the corresponding cDNA sequences. The phylogenetic tree was generated using the RAxML (Randomized Axelerated Maximum Likelihood) software version 7.0.3 (38). The Maximum Likelihood (ML) approach (23) was employed with the General Time Reversible (GTR) model for nucleotide substitution (24) and a gamma model for substitution rate heterogeneity (25). The likelihood trees were generated with 100 replicates using the GTRGAMMA model. Multiple non-parametric bootstrapping was performed with 500 replicates. The resulting trees were visualized and edited using MEGA v.4 or v.5 (39).

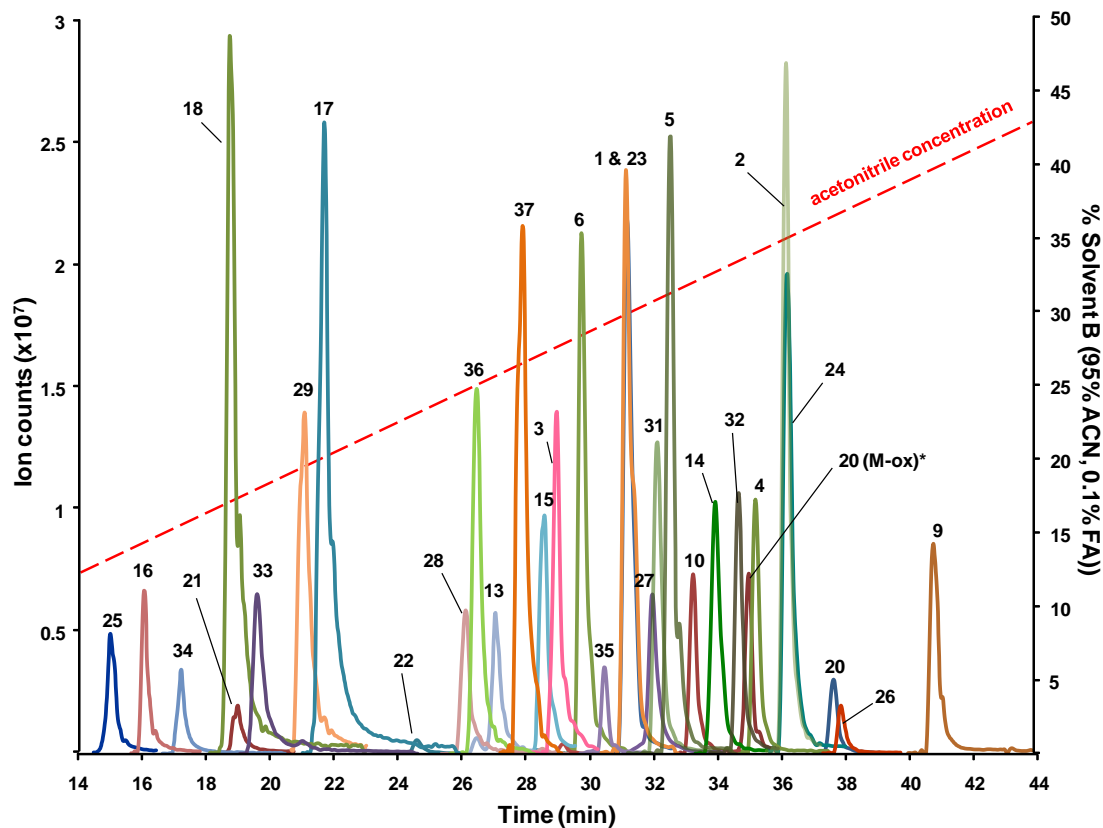
APPENDIX



Supplemental Figure 2.1. Generation of transgenic plants expressing a StrepII-tagged Clp subunit. **A)** Comparative analysis of wild-type, *clpr4-1* complemented with *CLPR4-StrepII* and *clpp3-1* complemented with *CLPP3-StrepII* grown on soil for about five weeks under a 10/14-h light/dark cycle at $100 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$. The complemented lines exhibit wild-type phenotype and are autotrophic. **B)** Protein gel blot analysis of titrations (indicated as 1/10, 1/2 and 1) of total leaf protein extracts from wild-type, *clpr4-1* complemented with *CLPR4StrepII* and homozygous *clpr4-1* plants grown for five weeks on agar plates with Murashige and Skoog medium with 2% sucrose under a 10/14-h light/dark cycle at $40 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$. Membranes were probed with antibodies generated against different chloroplast proteins namely the stromal RNA helicase 3 (RH3); PsaF, a small peripheral subunit of photosystem I complex; OEC23 and OEC33, both part of the oxygen evolving complex of the photosystem II assembly. RH3 was upregulated whereas the thylakoid-bound PsaF, OEC23 and OEC33 were downregulated in *clpr4-1*. Accumulation of these proteins in the complemented plants was restored to wild-type levels indicating successful complementation.



Supplemental Figure 2.2. Affinity purification and MS-based protein component identification of StrepII-tagged Clp assemblies. Native PAGE analysis of affinity-purified StrepII-tagged Clp assemblies (350 and 200kDa) from total leaf extractions showing the gel lanes at full-length. On-column salt incubation dissociated the individual rings from the core revealing the Strep-tagged Clp ring. Additional bands at 720 and 500 kDa were observed but were also found in mock purifications (with wild-type). MS/MS analyses of these bands indicate the co-purification of the Chaperone 60 complex (700 kDa) and the highly abundant Ribulose-1,5-bisphosphate carboxylase oxygenase (RUBISCO) complex (~550 kDa).



Supplemental Figure 2.3. Representative extracted ion chromatograms of the peptides derived from 100 fmol Clp-QconCAT protein. The numbers designate the position and identity of the peptide (see Supplemental Table 2.2). 32 out of the 37 Clp-QconCAT peptides can be detected and quantified. All the quantifiable Clp-QconCAT peptides eluted within the first 45 minutes of the run. The % acetonitrile concentration (red dashed line) was estimated based on the mobile phase gradient used with correction for dead volume during the chromatographic run.

ClpP3-P6 ring subunits

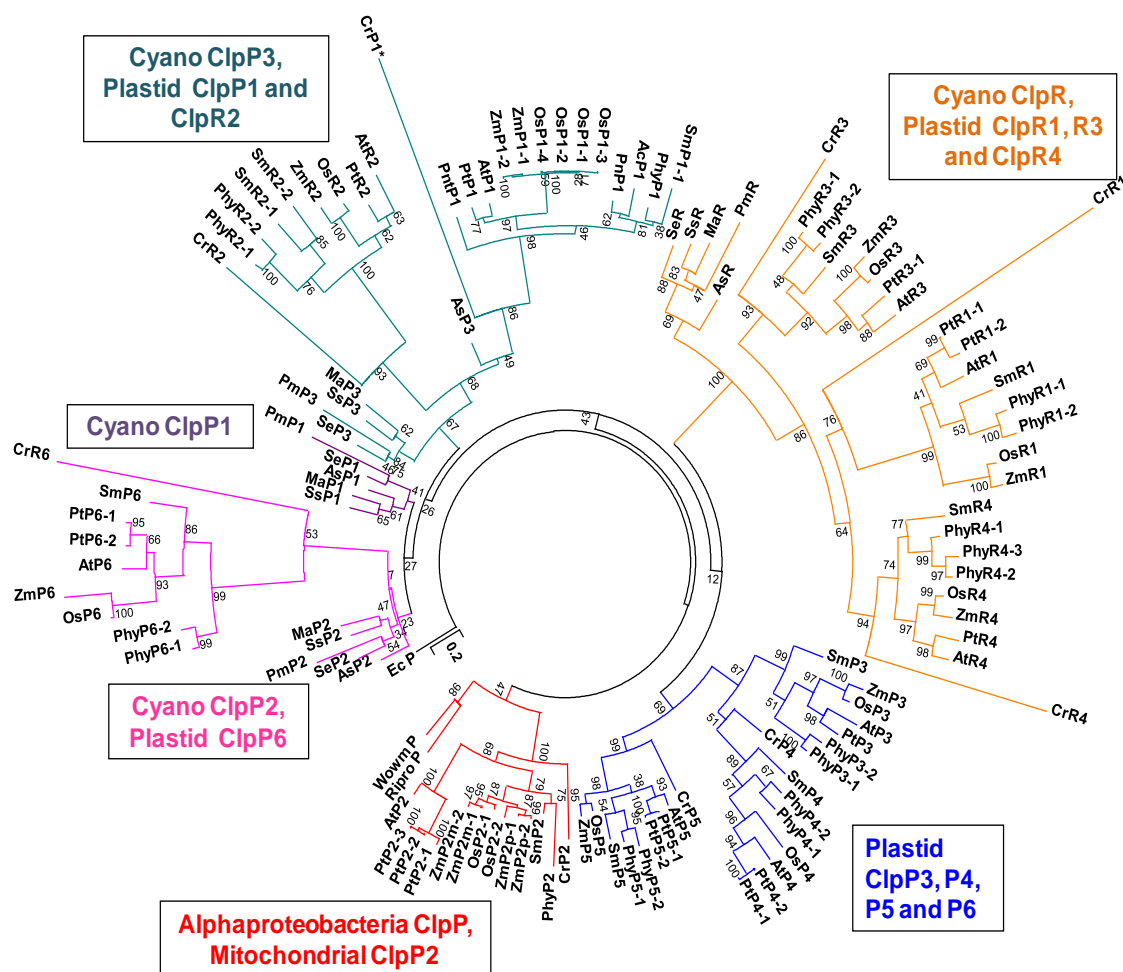
EcP	GLVDSILTHR N	
PhypP3-1	GLVDGVIG---	DGNPDIVAPIGSPIEPPKPKINSNWTIK-----
PhypP3-2	GLVDGVIG---	DGNPDLVAPIGSPIEPPKPKINSNWTIKE -FRE-RRYMPSEDAARQI----- E-APKS
SmP3	GLVDGVID---	DGKPGLVAPI GEAKEPPKTRVSDFTIKD -GK--- KNRPSED-----TQ
AtP3	GLIDAVID---	DGKPGLIAPIGDGTTPPKTKVWDLWK VEG -TKKDNTNLPSEK SMTQNGY-----AAAE
PtP3	GLVDEVID---	DGKPGLVAPLTDASPPPKTRVWDLWKIEG -SKKAKNNLPSEHKMLQNGYVGGGDGDRGVREKETPSPV
OsP3	GIVDSVID---	EGKPGLVAPLAGAVPPKSRVWYLNWNASGPTRK IMKNLPSEEKLI QNGN -GSASGDDGKF--KE- I STA
ZmP3	GIVDNIID---	EGKPGLVAPLAGSVPPP KSRVWYLNWNASGPTRK IMKHL PSEEKLI HNGN -GSATGDDGKL --KE -ATAT
PhypP4-1	GIIDGVID---	KDAIISINGLEVSDR -VKPRKENLAALEDPRKFLTPEIPDDEIY
PhypP4-2	GMIDGVID---	KDATIAINELPQVPER -VKPRQENLAALEDPRKFLTPEIPDDEIY
SmP4	GLIDGVID---	QENIIPV --- PAMPEKKIERRRDIKDAEADPMKFLKPQIPDDEIF
AtP4	GLIDGVID---	GDSIIPL --- EPVPDR -VKPR VNYEEISKDPMKFLTPEIPDDEIY
PtP4-1	GIIDGVID---	RDSIIPL --- APVPER -VTPTLNYEDMRKDPMKFLNPDVDPDEIY
PtP4-2	GIIDGVID---	RDSIIPL --- APVPER -VTPTLNYEDMRKDPMKFLNPDVDPDEIY
OsP4	GLIDSVID---	GDSIIPL --- EPVPER -VKPKYNYEELYKDPQKFLTDPVDPDEIY
PhypP5-1	GLIDAVIS---	NPLKALRPLPSANGSADSSPTPT
PhypP5-2	GLIDAVIS---	NPLKALRSPPAANGSADPTPTLA
SmP5	GLIDGVIS---	NPLKALSPLA-----
AtP5	GLIDGVIM---	NPLKALQPLAAA-----
PtP5-1	GLIDGVIL---	NPLKVLQPLAAAA -----DQQ
PtP5-2	GLIDGVIL---	NPLKVLQPLAAAA -----DQQ
OsP5	GLIDGVIM---	NPLKALQPLPAS ----- S
ZmP5	GLIDGVIM---	NPLKALQPLPAS ----- S
PhypP6-1	GLIDGLL---	ETEF
PhypP6-2	-----	E---
SmP6	GLIDGLL---	ETEY
AtP6	LIDGLL---	ETEF
PtP6-1	GLIDGIL---	ETEY
PtP6-2	GLIDGIL---	ETEY
OsP6	GLVDALL---	ETRY
ZmP6	-----	LI-----

Supplemental Figure 2.4. Multiple sequence alignment of the C-terminal extensions of the ClpP3-6 subunits comprising the ClpP ring of the 350 kDa plastid-localized Clp protease core complex in plants relative to the *E.coli* Clp sequence (outlined in orange box). The peptide sequences of these C-terminal extensions that were identified by MS/MS analyses (see PPDB: Plant Proteome Database at (<http://ppdb.tc.cornell.edu/>)) are highlighted in green.

ClpP1/R ring subunits

EcP	GLVDSILTHRN	
PhypP1	GIVDLVAIEN-----T	
PnP1	GIVDLIALEN-----D	
AcP1	GVVDLVAVEN-----VSR	
SmP1	GIIDLVMMDT-----SP---PRPIVSS	
PnthP1	GIVDVVAEG-----	
AtP1	GIVDLVAVQ-----	
PtP1	GIVDLVAVA-----	
OsP1-1	GLVDIVGDEMLDEHCDTDPVWFPEMFKDW	C-terminal extensions for ClpP1 only in monocots
OsP1-2	GLVDIVGDEMLDEHCDTDPVWFPEMFKDW	
OsP1-3	GLVDIVGDEMLDEHCDTDPVWFPEMFKDW	
OsP1-4	GLVDIVGDEMLDEHCDTDPVWFPEMFKDW	
ZmP1-1	GLVDIVGDEMLDEHCDTDPVWFPEMFKDW	
ZmP1-2	GLVDIVGDEMLDEHCDTDPVWLEMEFKDW	
PhypR1-1	GLADKIIIEPRGIAMEKRNIDEMLAQSKAARASYTRPGAAA---SAGAGGR	
PhypR1-2	GLADKIIIEPRGIAMERRNIDEMLAQSKAARASYARPGAAA---PAGAGGR	
SmR1	GLADKVIIEEP-----D	
AtR1	GIADKIADSQDSSFEKRDYDGTLAGRAMRPGGGSP-----AAPAGLR	
PtR1	GIADKLLSSNDDAFEKRDYDALLAQTKAMKAQAAGPR-----AAPSGSR	
OsR1	GLADTILHSLDGSFKPKDLTAQLAKAQEMRQSGKRPAAGAGRWSTPSVPR	
ZmR1	GLADTILHSLDGSFKPKDLTAQLAKAQAMRQSGKRAAAGAGRWSTPTAPR	
PhypR2-1	GIIDKVVPRPKRIKPDA---RRQESVG--VGLG	
PhypR2-2	GIIDKVVPRPKRIKPDA---RRQESIG--VGLG	
SmR2-1	GLIDKIIIRPPTVKPDA---SSKAEAG--RGIG	
SmR2-2	GLIDKIVRPPKYASSS---LSIGGVGPYNASF	
AtR2	GLIDKIVRPPRIKEDA---PRQD-ES--AGLG	
PtR2	GLIDRIIRPPRIDDDV---GPSD-AS--AGLG	
OsR2	GIIDRIIRPSRIKKEGSTAQKKDLRN--LGLG	
ZmR2	GIIDRIVRPSRIKKEGSTAQRRDMRN--LGLG	
PhypR3-1	GVADKILWRGQEA-MAETMKTEDWDKGAGIRVVERSASNGAGDSGLG	
PhypR3-2	GVADKILWRGQEA-MAETMKAEDWDKGAGIRVVERPASYGAGDTGLG	
SmR3-1	GVVDKILWRGQEA-MGETLSPEQWDKRAGIRAVERPMTYGAGTSGLG	
SmR3-2	GVVDKILWRGQEA-MGETLSPEQWDKRAGIRAVERPMTYGAGTSGLG	
AtR3	GVIDRILWRGQEKTIADVVPSEEFDKNAGIKS-----VV	
PtR3-1	GVIDKILWQGEKIMADVLPPEWDKSAKIKVA-----DPF	
OsR3	GVIDKILWRGQEKYMADMLSPPEWDKVAGVRRP-----DIM	
ZmR3	GVIDKILWRGQEKYMADMLSPDEWDKVAGVRHP-----DLM	
PhypR4-1	GLIDKVLVYAEGL-DEKRSIASDLKKAQ-LI	
PhypR4-2	GIIDKVLVYAEGL-DEKRSIASDLKKAQ-LI	
PhypR4-3	GIIDKVLVYVEGL-DEKRSIASDLKKAQ-LI	
SmR4	GLIDKVLRTKGEKSKGVVDQLRQAQLLK	
AtR4	GIIDKVVYNERGSQDRGVVSDLKKAQ-LI	
PtR4	GIIDKVIYNERTTEDRGVVSDLKKAQ-LI	
OsR4	GIIDKVLVYNEKSQEDGGVVSELKRSN-LI	
ZmR4	-----W-KN	

Supplemental Figure 2.5. Multiple sequence alignment of the C-terminal extensions of the ClpP1 and ClpR1-4 subunits comprising the ClpP1-R ring of the 350 kDa plastid-localized Clp protease core complex in plants relative to the *E. coli* Clp sequence (outlined in orange box). The peptide sequences of these C-terminal extensions that were identified by MS/MS analyses (see PPDB: Plant Proteome Database at <http://ppdb.tc.cornell.edu/>) are highlighted in green.



Supplemental Figure 2.6. Phylogenetic analysis of the ClpP/R proteins from proteobacteria, photosynthetic bacteria, green alga and plants. The phylogenetic tree inferred for the 118 Clp sequences from various prokaryotes and photosynthetic eukaryotes employing the Maximum Likelihood (ML) approach¹¹ with the General Time Reversible (GTR) model for nucleotide substitution¹² and a gamma model of substitution rate heterogeneity¹³. Support values are shown from 500 non-parametric bootstrap inferences. This includes the *Escherichia coli* ClpP; the ClpP from alphaproteobacteria presumed to be the ancestor of mitochondrial organelles: *Rickettsia prowazekii* (Rp) and *Wolbachia wmel* (Ww); the cyanobacteria which are the ancestor of plastids: *Anabaena sp. (strain PCC 7120)* (As), *Synechocystis sp. (strain PCC 6803)* (Ss), *Synechococcus elongatus (strain PCC 7942)* (Se), *Microcystis aeruginosa* (Ma) and *Prochlorococcus marinus MED4* (Pm); the green alga *Chlamydomonas reinhardtii* (Cr) and representative plant species: the bryophyte *Physcomitrella patens* (PhyP), the lycopod *Selaginella moellendorffii* (Sm), the dicots *Arabidopsis thaliana* (At) and *Populus trichocarpa* (Pt) as well as dicots *Zea mays* (Zm) and *Oryza sativa* (Os). The plastid ClpP1 homologues for the ferns *Psilotum nodum* (Pn) and *Adiantum capillus renesis* (Ac) and the gymnosperm *Pinus thunbergii* (Pnt) were also collected from recent chloroplast genome sequencing projects for these species. The Cr ClpP1* is the plastid-encoded ClpP1 in green alga with the large insertion (IS1) sequence removed. The tree was rooted against Ec ClpP.

Supplemental Table 2.1. MS/MS analyses of the affinity-purified Clp assemblies^a

Accession	Clp	R4-StrepII			P3-StrepII			
		no salt		+ salt	no salt			+ salt
		core (350 kDa)	ring(s) (200 kDa)	ring (200 kDa)	core (350 kDa)	ring(s) (200 kDa) ^b	ring(s) (180 kDa) ^b	ring (200 kDa)
ATCG00670.1	P1	5	2	6	43	11	40	
AT1G49970.1	R1	8	1	19	19	10	28	
AT1G12410.1	R2	6	4	30	25	20	60	
AT1G09130.1	R3	11	2	33	35	34	39	
AT4G17040.1	R4	10	4	43	57	26	77	
AT1G66670.1	P3	3	1		35	50	70	18
AT5G45390.1	P4	9	6	3		122	127	72
AT1G02560.1	P5	8	4	2	66	86	109	41
AT1G11750.1	P6	10	6		43	86	76	40
AT4G25370.1	T1	2			38	36	26	
AT4G12060.1	T2	3			6	2		

^a MS/MS analyses was performed from peptides extracted from the in-gel trypsin-digested protein bands from native gel separations of the affinity-purified Clp complexes (see Fig.2 for the corresponding gel images). Shown here are the total adjusted spectral counts per protein.

^b Two bands were observed at slightly different masses and were excised separately for in-gel digestion and MS/MS analysis.

Supplemental Table 2.2 The peptides comprising Clp-QconCAT and their corresponding properties.

Accession	Peptide	Pos. ^a	Peptide	Charge States	a.a ^b	[M+2H] ²⁺	[M+3H] ³⁺	Q? ^c
Clp core subunits								
ATCG00670.1	P1-1	19	IAPPHAR	2,3	7	406.2323,	271.1573	N
	P1-2	6	SPGEGDTSWVDIYNR	2	15	848.3841		Y
AT1G49970.1	R1-1	11	YLQAQAAIDYGIADK	2	15	820.4199		Y
	R1-2	24	TAPPDLPSLLLDAR	2	14	739.9143		Y
AT1G12410.1	R2-1	25	IALQSPAGAAR	2	11	527.8038		Y
	R2-2	4	FNAEEAIEYGLIDK	2	14	806.3987		Y
AT1G09130.1	R3-1	12	EPIYIYINSTGTTR	2	14	814.4199		N
	R3-2	26	DILVELLSK	2	9	515.3132		Y
AT4G17040.1	R4-1	27	GSAHEQPPDLASYLFK	2, 3	17	928.9625	619.6441	Y
	R4-2	14	YFSPTEAVEYGIIDK	2	15	866.4274		Y
AT1G66670.1	P3-1	7	DNTNLPSEK	2	9	523.2491		N
	P3-2	20	LPSFEELDTTNMLLR	2	15	889.9533		Y
AT5G45390.1	P3-2 (ox)	20	LPSFEELDTTNM(ox)LLR	2	15	897.9508		Y
	P4-1	8	SFEQVLK	2	7	425.7371		Y
AT1G02560.1	P4-2	21	ADVSTIALGIAASTASILGAGTK	3	24	1101.579		N
	P5-1	22	ANLNGYLAYHTGQSLEK	2	17	939.9709		Y
AT1G11750.1	P5-2	9	FQSISQLFQYR	2	12	765.4092		Y
	P6-1	23	IIFIGQPINAQVAQR	2	15	834.4832		Y
	P6-2	10	VISQLVTLASIDDK	2	14	751.4272		Y
Clp chaperones/adaptors								
AT4G25370.1	T1-1	13	DETLSSLGK	2	9	488.2715		Y
	T1-2	28	AIAWAIDEK	2	9	508.7742		Y
AT4G12060.1	T2-1	29	ALDSALDQNLK	2	11	594.317		Y
	T2-2	15	ILATLGFTDEK	2	11	604.3321		Y
AT1G68660.1	S-1	30	VILHNDNFNK	2, 3	10	607.3198	405.2157	N
	S-2	16	GGGVLDKPIIEK	2, 3	12	613.3612	409.2432	Y
AT5G50920.1	C1-1	31	VPEPTVDETIQILK	2	14	791.4403		Y
AT5G50920.1,	C1/2-1	17	VLENLGADPSNIR	2	13	699.3728		Y
AT3G48870.1	C1/2-2	32	GSGFVAVEIPFTPR	2	14	738.8959		Y
AT5G51070.1	D1-1	18	VVGQDEAVAAISR	2	13	657.8542		Y
	D1-2	33	VFEAAVEYSR	2	10	585.7931		Y
Localization markers^d								
AT1G04410.1	MD-1	1	VLVVANPANTNALILK	2	16	825.5011		Y
	MD-2	34	LSVPVSDVK	2	9	472.2766		Y
AT5G20290.1	RPS8-1	3	VLDVVYNASNELVR	2	15	852.9494		Y
	RPS8-2	36	SAIVQVDAAPFK	2	12	623.3455		Y
AT3G04790.1	PRI-2	35	SLGIPLVGLDTHPR	2, 3	14	737.9225	492.2841	Y
	PRI-1	2	LLSSGELYDIVGIPTSK	2	17	896.4906		Y
AT5G51820.1	PGM1-1	5	VAEIPDIDLSQVGVTK	2	16	842.4618		Y
	PGM1-2	37	IYGNTLSISEIK	2	12	669.3692		Y

^a Order by which these proteotypic peptide appeared starting from the N-terminus of the Clp-QconCAT construct

^b Number of amino acid residues per peptide

^c Whether it can be detected and quantified (Y) or not (N)

^d These proteins serve as reference for organellar location. Malate dehydrogenase (MD) and the 40S ribosomal subunit S8 (RPS8) are both cytosolic whereas ribose 5-phosphate isomerase (PRI) and phosphoglucomutase (PGM) are both plastid-localized.

Supplemental Table 2.3. Linear instrument response for individual Clp-QconCAT peptides.*

Peptide	Position	slope	y-intercept	Regression coefficient
ClpP1/R ring				
P1-2	6	1.40	0.05	0.995
R1-1	11	1.09	0.04	0.971
R1-2	24	1.20	0.04	0.990
R2-1	25	1.05	0.14	0.999
R2-2	4	1.12	0.04	0.987
R3-2	26	0.76	0.13	0.968
R4-1	27	1.13	0.06	0.996
R4-2	14	1.32	0.05	0.991
ClpP3-P6 ring				
P3-2	20	0.89	0.26	0.987
P4-2	8	1.04	0.05	0.999
P5-1	22	1.57	0.09	0.999
P5-2	9	0.57	0.42	0.956
P6-1	23	1.21	0.05	0.994
P6-2	10	1.22	0.33	0.975
Clp chaperones/adaptors				
T1-1	13	1.05	0.04	0.998
T1-2	28	1.10	0.07	0.998
T2-1	29	1.00	0.08	0.998
T2-2	15	0.72	0.19	0.987
S-2	16	0.86	0.12	0.998
C1/2-1	17	1.07	0.07	0.997
C1/2-2	32	0.70	0.30	0.971
C1-1	31	0.92	0.02	0.986
D1-1	18	1.08	0.09	0.999
D1-2	33	1.31	0.004	0.993
Localization markers				
MD-1	1	1.11	0.09	0.992
MD-2	34	1.03	0.00	0.998
PGM1-1	5	1.09	0.09	0.988
PGM1-2	37	1.30	0.06	0.996
PRI-1	2	0.87	0.13	0.981
PRI-2	35	1.14	0.03	0.996
RPS8-1	3	1.26	0.04	0.992
RPS8-2	36	0.85	0.15	0.989

* For 100 fmol of heavy Clp-Qconcat peptides mixed with varying amounts of their light versions spanning two orders of magnitude (10 fmol to 1 pmol).

Supplemental Table 2.4. Quantification of Clp subunits and determination of subunit stoichiometry in the Clp core complex.^a

Clp	Pep	Moles (x100 fmol)				Stoichiometry					
		R4-1	R4-2	P3-1	P3-2	R4-1	R4-2	P3-1	P3-2	AVE	SD
P1	Q6	1.6	0.6	3.3	4.4	3.4	3.2	2.5	2.7	2.9	0.4
R1	Q11	0.6	0.3	1.3	1.7	1.2	1.5	1.0	1.0	1.2	0.2
R2	Q25	0.7	0.3	1.4	1.8	1.5	1.5	1.1	1.1	1.3	0.2
	Q4	0.3	0.2	0.9	1.3	0.7	0.8	0.7	0.8	0.8	0.1
R3	Q26	0.5	0.2	1.1	1.7	1.0	1.0	0.8	1.1	1.0	0.1
R4	Q27	0.5	0.2	0.9	1.2	1.0	0.9	0.7	0.8	0.9	0.2
	Q14	0.5	0.2	1.1	1.4	1.0	1.1	0.8	0.9	0.9	0.1
TOTAL^b		4.7	1.9	10.1	13.6						

Clp	Pep	Moles (x100 fmol)				Stoichiometry					
		R4-1	R4-2	P3-1	P3-2	R4-1	R4-2	P3-1	P3-2	AVE	SD
P3	Q20	0.5	0.3	1.7	1.8	1.0	1.4	1.3	1.1	1.2	0.2
P4	Q8			2.8	3.5			2.1	2.2	2.2	0.1
P5	Q22	1.6	0.5	2.2	3.4	3.2	2.6	1.6	2.1	2.4	0.7
	Q9	1.1	0.5	4.9	6.3	2.2	2.7	3.7	3.9	3.1	0.8
P6	Q23	0.7	0.3	1.5	1.9	1.5	1.7	1.2	1.2	1.4	0.3
	Q10	0.7	0.2	1.1	1.4	1.4	1.3	0.8	0.8	1.1	0.3
TOTAL^b		4.5	1.8	14.1	18.2						

^a Based on affinity purification and native gel separation of various Clp assemblies from R4-StrepII (n=2) or P3-StrepII (n=2) transgenic lines. Molar amounts are derived from peak area ratios between the endogenous Clp peptides and 100 fmol spiked Clp-QconCAT peptides). Stoichiometry was calculated by normalization against the molar amounts of ClpR4 or ClpP6 proteins.

^b Sum of the molar amounts of the constituent Clp subunits per ring. The average ratio between the total moles of ClpP1/R and ClpP3-6 within the Clp core is 1.2.

Supplemental Table 2.5. Quantification of Clp subunits and determination of subunit stoichiometry in the Clp rings.^a

		Clp ring with no salt incubations (200 kDa)														purified Clp ring from on-column salt incubation (200 kDa)									
Clp	Pep	Moles (x100 fmol)						Stoichiometry								Moles (x100 fmol)			Stoichiometry						
		R4-4	R4-5	P3-3	P3-4	P3-5	P3-6	R4-4	R4-5	P3-3	P3-4	P3-5	P3-6	AVE	SD	R4-6	R4-7	R4-8	R4-6	R4-7	R4-8	AVE	SD		
P1	Q6	4.6	0.9	1.2	1.7	6.5	6.9	3.2	3.1	3.3	3.3	3.5	3.5	3.3	0.2	1.2	2.7	1.1	2.8	3.4	2.7	3.0	0.4		
R1	Q11	1.5	0.4	0.6	0.7	2.2	2.3	1.1	1.4	1.6	1.3	1.2	1.2	1.3	0.2	0.5	1.0	0.6	1.1	1.2	1.3	1.2	0.1		
R2	Q25	1.8	0.5	0.7	0.8	2.3	3.0	1.3	1.8	1.9	1.5	1.2	1.6	1.6	0.3	0.6	1.0	0.5	1.4	1.3	1.2	1.3	0.1		
	Q4	0.9	0.2	0.3	0.5	1.7	1.8	0.7	0.9	0.7	0.9	0.9	0.9	0.8	0.1	0.3	0.7	0.4	0.7	0.8	0.9	0.8	0.1		
R3	Q26	1.3	0.3	0.4	0.5	1.8	2.5	0.9	1.1	1.1	1.0	0.9	1.3	1.0	0.1	0.3	0.8	0.6	0.6	1.0	1.3	1.0	0.4		
R4	Q27	1.4	0.2	0.4	0.5	1.7	1.8	1.0	0.8	1.0	0.9	0.9	0.9	0.9	0.1	0.4	0.7	0.4	0.9	0.9	0.9	0.9	0.016		
	Q14	1.4	0.3	0.3	0.6	2.0	2.1	1.0	1.2	1.0	1.1	1.1	1.1	1.1	0.1	0.5	0.9	0.5	1.1	1.1	1.1	1.1	0.016		

Clp	Pep	Moles (x100 fmol)					Stoichiometry								Moles (x100 fmol)				Stoichiometry						
		R4-5	P3-3	P3-4	P3-5	P3-6	R4-5	P3-3	P3-4	P3-5	P3-6	AVE	SD	P3-7	P3-8	P3-9	P3-10	P3-7	P3-8	P3-9	P3-10	AVE	SD		
P3	Q20	0.2	1.7	2.2	2.2	3.6	0.7	1.2	1.1	0.9	1.0	1.0	0.2	0.6	1.5	0.6	2.1	1.0	1.0	0.9	1.1	1.0	0.1		
P4	Q8		3.8	5.2	5.7	9.2		2.6	2.7	2.2	2.7	2.5	0.2	1.4	3.2	1.5	4.9	2.4	2.0	2.1	2.6	2.3	0.3		
P5	Q22	0.8	2.2	5.1	5.8	8.2	2.5	1.5	2.6	2.2	2.4	2.3	0.4	0.9	3.3	1.5	4.2	1.6	2.0	2.1	2.2	2.0	0.3		
	Q9	0.7	4.3	7.5	10.2	12.9	2.0	3.0	3.9	3.9	3.7	3.3	0.8	1.4	5.0	2.1	7.7	2.5	3.1	3.0	4.1	3.2	0.7		
P6	Q23	0.3	1.7	2.3	3.0	3.8	1.1	1.2	1.2	1.2	1.1	1.1	0.1	0.7	1.8	0.8	2.1	1.1	1.1	1.1	1.1	1.1	0.02		
	Q10	0.3	1.2	1.6	2.2	3.2	0.9	0.8	0.8	0.8	0.9	0.9	0.1	0.5	1.4	0.6	1.6	0.9	0.9	0.9	0.9	0.9	0.02		

^a Based on affinity purification and native gel separation of various Clp assemblies from R4-StrepII (N=8) or P3-StrepII (N=10) transgenic lines. Molar amounts are derived from peak area ratios between the endogenous Clp peptides and 100 fmol spiked Clp-QconCAT peptides). Stoichiometry was calculated by normalization against the molar amounts of ClpR4 or ClpP6 proteins.

Supplemental Table 2.6. Accessions for protein and cDNA sequences of the Clp proteins from prokaryotes, green alga and plants.

Protein Accessions (Prokaryotic Clps)					
	Proteo-bacteria P	Cyano P1	Cyano P2	Cyano P3	Cyano R
<i>E. coli</i> (Ec)	POA6G7				
<i>Rickettsia prowazekii</i> (Rp)	RP520				
<i>Wolbachia wmel</i> (Ww)	WD0319				
<i>Anabaena</i> sp. (strain PCC 7120) (As)		Q8YXH5	Q8YQX8	Q8YP43	Q8YP42
<i>Synechocystis</i> sp. (strain PCC 6803) (Ss)		P54416	Q59993	P74467	P74466
<i>Synechococcus elongatus</i> (strain PCC 7942) (Se)		P54415	O34125	Q9L4P3	Q9L4P4
<i>Microcystis aeruginosa</i> (Ma)		MAE_11870	MAE_62720	MAE_57180	MAE_57190
<i>Prochlorococcus marinus</i> MED4 (Pm)		PMM0742	PMM1656	PMM1314	PMM1313
cDNA Accessions (prokaryotic Clps)					
	Proteo-bacteria P	Cyano P1	Cyano P2	Cyano P3	Cyano R
<i>E. coli</i> (Ec)	ENA AAA23588				
<i>Rickettsia prowazekii</i> (Rp)	RP520				
<i>Wolbachia wmel</i> (Ww)	WD0319				
<i>Anabaena</i> sp. (strain PCC 7120) (As)		gi 17227497: 1470240-1470854	gi 17227497: 4443275-4443973	gi 17227497: 5219511-5220104	gi 17227497: 5220184-5220846
<i>Synechocystis</i> sp. (strain PCC 6803) (Ss)		gi 16329170: 3212598-3213194	gi 16329170: 3245421-3246101	gi 16329170: 2206551-2207159	gi 16329170: 2205774-2206451
<i>Synechococcus elongatus</i> (strain PCC 7942) (Se)		ENA AAC67306	ENA AAB68677	ENA AAB68677	ENA CAB81780
<i>Microcystis aeruginosa</i> (Ma)		MAE_11870	MAE_62720	MAE_57180	MAE_57190
<i>Prochlorococcus marinus</i> MED4 (Pm)		PMM0742	PMM1656	PMM1314	PMM1313

Supplemental Table 2.6 (continued)

Protein Accessions (Plastid Clps)					
	P1	P3	P4	P5	P6/R6
Chlamydomonas reinhardtii (Cr)	P42380		A8II60	A8IL21	A8INX1
Physcomitrella patens (PhyP)	Q6YXM7	P3-1: A9RYV9, P3-2: A9T6II	P4-1: A9TZR6, P4-2: A9RE44	P5-1: A9S5E8, P5-2: A9TE25	P6-1: A9TH43, P6-2: A9SWP4
Selaginella moellendorffii (Sm)	C7B2H1	XP_002960346.1	XP_002981945.1	XP_002963649.1 , XP_002983780.1	XP_002968942.1
Arabidopsis thaliana (At)	ATCG00670.1	AT1G66670.1	AT5G45390.1	AT1G02560.1	AT1G11750.1
Populus trichocarpa (Pt)	A4GYT6	B9H362	P4-1: B9GZW8, P4-2: A9PA38	P5-1: B9GST6, P5-2: B9I9I1	P6-1: A9PDP3, P6-2: B9HRQ3
Zea mays (Zm) ("GRM" removed from Accessions)	ZM2G448161_P01, ZM2G427444_P06	ZM2G001755_P01		ZM2G121456_P01	ZM2G092632_P01
Oryza sativa (Os)	Os08g15270, Os10g21300, Osp1g005990, Os12g10590	Os01g32350.1	Os10g43050.1	Os03g19510.1	Os03g29810.1
Psilotum nodum (Pn)	Q8WHZ7				
Adiantum capillus renesis (Ac)	Q85FJ8				
Pinus thunbergii (Pnt)	P41609				

Supplemental Table 2.6 (continued)

cDNA accessions (for plastid ClpPs)					
	P1	P3	P4	P5	P6/R6
Chlamydomonas reinhardtii (Cr)	ENA ACJ50097		gi 159465134: 119-1156	gi 159465360: 108-878	gi 159466487: 159-1010
Physcomitrella patens (PhyP)	gi 34501376:11632-11702, 12338-12632, 13074-13307	P3-1: gi 168013185, P3-2: gi 168043510	P4-1: gi 168064315, P4-2: gi 167999543	P5-1: gi 168018158: 395-1306 , P5-2: gi 168048811	P6-1: gi 168050942, P6-2: gi 168036512
Selaginella moellendorffii (Sm)	gi 255961289: c>68286-68216, c67451-66902	XM_002960300.1	XM_002981899.1	XM_002963603.1 XM_002983734.1	XM_002968896.1
Arabidopsis thaliana (At)	ATCG00670.1	gi 145337232: 57-986	gi 145358888	gi 145334998	gi 145335409
Populus trichocarpa (Pt)	ENA ABO36731	gi 224079412	P4-1: gi 224074499:77-739 , P4-2: EF145046	P5-1: gi 224068557: 123-1025 , P5-2: gi 224128301: 37-543	P6-1: gi 224080314: 72-869 , P6-2: gi 224103436:1-741
Zea mays (Zm) ("GRMZM" removed from Accessions)	P1-1: 2G448161_T01 , P1-2: 2G427444_T06	2G001755_P01	2G001755_T01	2G121456_T03	2G121456_T03
Oryza sativa (Os)	P1-1: 13108.m09245, P1-2: 13110.m07757, P1-3: gi 11466763: 67638-68288 , P1-4: 13112.m01122	13101.m03260	13110.m04073	13103.m02334	13103.m03417
Psilotum nodum (Pn)	ENA BAB84241				
Adiantum capillus renesis (Ac)	ENA AAP29415				
Pinus thurnbergii (Pnt)	gi 7524593: 29611-30201				

Supplemental Table 2.6 (continued)

Protein accessions (plastid ClpR and mitochondrial ClpP2)					
	R1	R2	R3	R4	mito P2
Chlamydomonas reinhardtii (Cr)	A8INX1	jgi Chlre4 183767	A8I547	A8IH07	A8IX06
Physcomitrella patens (PhyP)	P6-1: A9TH43, P6-2: A9SWP4	R1-1:A9TVY1, R1-2: A9SI89	R2-1: A9SWI4, R2-2: A9T3W3	R3-1: A9SG20, R3-2: A9RZ41	R4-1: A9RQL9, R4-2: A9TF07, R4-3: A9SSY7
Selaginella moellendorffii (Sm)	XP_002968942.1	XP_002970082.1	XP_002968834.1	XP_002990376.1	XP_002982594.1
Arabidopsis thaliana (At)	AT1G11750.1	AT1G49970.1	AT1G12410.1	AT1G09130.1	AT4G17040.1
Populus trichocarpa (Pt)	P6-1: A9PDP3, P6-2: B9HRQ3	R1-1: B9HMY7, R1-2: B9NHA6	B9N3L0	A9PFT0	A9PAS6
Zea mays (Zm)	ZM2G092632_P01	ZM2G099529_P01	ZM2G148106_P01	ZM2G030072_P02	AC207652.3_FGP003
Oryza sativa (Os)	Os03g29810.1	Os05g51450.1	Os06g04530.1	Os03g22430.1	Os01g16530.1
cDNA accessions (Plastid ClpR and mitochondrial ClpP2)					
	R1	R2	R3	R4	mito P2
Chlamydomonas reinhardtii (Cr)	jgi Chlre4 183767 estExt_fgensch2_kg.C_10275	gi 159467460:144-992	gi 159464778:1-1248	gi 159470090:140-970	gi 159477908
Physcomitrella patens (PhyP)	R1-1: gi 168061570, R1-2: gi 168026988	R2-1: gi 168036369, R2-2: gi 168041626	R3-1: gi 168025571, R3-2: gi 168013354	R4-1: gi 168007319, R4-2: gi 168049468, R4-3: gi 168033938:1-714	gi 168031791
Selaginella moellendorffii (Sm)	XM_002970036.1	XM_002968788.1	XM_002990330.1,	XM_002982548.1	XM_002983734.1
Arabidopsis thaliana (At)	gi 145336568	gi 145335431	gi 145335309	gi 30683910	gi 186524962:74-799
Populus trichocarpa (Pt)	R1-1:gi 224103734:80-1255, R1-2: gi 224097119	gi 224125071:126-965	gi 224125177:74-1099	gi 224126446:25-957	P2-1: gi 224081362:28-762 , P2-2: gi 224094112:142-876 , P2-3: gi 224094112:142-876
Zea mays (Zm) ("GRMZM" removed from Accessions)	2G099529_T01	2G148106_T01	2G030072_T02	AC207652.3_FGT003	P2m-1: 2G320135_T02, P2m-2: 2G474883_T01, P2p-1: 2G053236_T01, P2p-2: 2G111756_T01
Oryza sativa (Os)	13105.m05548	13106.m00412	13103.m02699	13101.m01848	P2-1: 13104.m04426, P2-2: 13102.m04675

Supplemental Table 2.7. Primers used in this study

Construct	Primers (5' to 3')
R4cDNASTrepII Step1, for PCR amplification	1:CCATGGAGGTAGCAGCAGCGA 2: AGGATGAGACCAAATGAGTTGTGCC
R4cDNASTrepII Step2, for PCR amplification	1:CACCATGGAGGTAGCAGCAGCGAC 2: TCACTTCTCGAATTGAGGATGAGACCA
R4StrepII in pEARLEYGATE100, for genotyping	1:GGGAATGTCTCTCGTACCTTCAGTT 2: GGCGCTCTATCATAGATGTCGCTATAAACC
R4StrepII in pEARLEYGATE100, for genotyping	1:CCATGGAGGTAGCAGCAGCGA 2: GGCGCTCTATCATAGATGTCGCTATAAACC
R4 gene specific, for genotyping	1;GGGAATGTCTCTCGTACCTTCAGTT 2: ACAACTGGACACTGTTGCATAATGA
P3 cDNA Strep II, for PCR amplification	1:ATGGAGATGAGTTTGCGTCTCGCTTC 2:CTACTTCTCGAATTGAGGATGAGACCATTC AAT GGCGGCATAAACCATTCTGTGTC
P3StrepII in pEARLEYGATE100, for genotyping	1:ATGGAGATGAGTTTGCGTCTCGCTTC 2: GGCGCTCTATCATAGATGTCGCTATAAACC
TDNA insertion, for genotyping	1:GGCAATCAGCTGTTGCCCGTCTCACTGGTG 2: ACAACTGGACACTGTTGCATAATGA
ClpQconCAT in pET21, for confirmation	1:TTATGCTAGTTATTGCTCAGCGGTG 2: CCATTTCTCGTATCGCATTTCAC
ClpQconCAT in pET21, for confirmation	1:TCCGGCGTAGAGGATCGAGATC 2: TGGAATGCGATACGAGAAATGG

ACKNOWLEDGEMENTS

This work was supported by grants from the National Science Foundation (MCB-1021963 and MCB-0718897) to KJVW. We would like to thank the members of the van Wijk laboratory for helpful discussions.

REFERENCES

1. Yu, A. Y. & Houry, W. A. (2007) *FEBS Lett* **581**, 3749-57.
2. Wang, J., Hartling, J. A. & Flanagan, J. M. (1997) *Cell* **91**, 447-56.
3. Peltier, J. B., Ripoll, D. R., Friso, G., Rudella, A., Cai, Y., Ytterberg, J., Giacomelli, L., Pillardy, J. & Van Wijk, K. J. (2004) *J Biol Chem* **279**, 4768-81.
4. Sokolenko, A., Pojidaeva, E., Zinchenko, V., Panichkin, V., Glaser, V. M., Herrmann, R. G. & Shestakov, S. V. (2002) *Curr Genet* **41**, 291-310.
5. Shikanai, T., Shimizu, K., Ueda, K., Nishimura, Y., Kuroiwa, T. & Hashimoto, T. (2001) *Plant Cell Physiol* **42**, 264-73.
6. Kuroda, H. & Maliga, P. (2003) *Nature* **425**, 86-9.
7. Kim, J., Rudella, A., Ramirez Rodriguez, V., Zybaïlov, B., Olinares, P. D. & van Wijk, K. J. (2009) *Plant Cell* **21**, 1669-92.
8. Rudella, A., Friso, G., Alonso, J. M., Ecker, J. R. & van Wijk, K. J. (2006) *Plant Cell* **18**, 1704-21.
9. Zheng, B., MacDonald, T. M., Sutinen, S., Hurry, V. & Clarke, A. K. (2006) *Planta* **224**, 1103-15.
10. Sjogren, L. L., Stanne, T. M., Zheng, B., Sutinen, S. & Clarke, A. K. (2006) *Plant Cell* **18**, 2635-49.
11. Koussevitzky, S., Stanne, T. M., Peto, C. A., Giap, T., Sjogren, L. L., Zhao, Y., Clarke, A. K. & Chory, J. (2007) *Plant Mol Biol* **63**, 85-96.
12. Maurizi, M. R., Clark, W. P., Kim, S. H. & Gottesman, S. (1990) *J Biol Chem* **265**, 12546-52.
13. Peltier, J. B., Ytterberg, J., Liberles, D. A., Roepstorff, P. & van Wijk, K. J. (2001) *J Biol Chem* **276**, 16318-27.

14. Gribun, A., Kimber, M. S., Ching, R., Sprangers, R., Fiebig, K. M. & Houry, W. A. (2005) *J Biol Chem* **280**, 16185-96.
15. Maurizi, M. R., Singh, S. K., Thompson, M. W., Kessel, M. & Ginsburg, A. (1998) *Biochemistry* **37**, 7778-86.
16. Maglica, Z., Kolygo, K. & Weber-Ban, E. (2009) *Structure* **17**, 508-16.
17. Pratt, J. M., Simpson, D. M., Doherty, M. K., Rivers, J., Gaskell, S. J. & Beynon, R. J. (2006) *Nat Protoc* **1**, 1029-43.
18. Kuster, B., Schirle, M., Mallick, P. & Aebersold, R. (2005) *Nat Rev Mol Cell Biol* **6**, 577-83.
19. Andersson, F. I., Tryggvesson, A., Sharon, M., Diemand, A. V., Classen, M., Best, C., Schmidt, R., Schelin, J., Stanne, T. M., Bukau, B., Robinson, C. V., Witt, S., Mogk, A. & Clarke, A. K. (2009) *J Biol Chem* **284**, 13519-32.
20. Stanne, T. M., Pojidaeva, E., Andersson, F. I. & Clarke, A. K. (2007) *J Biol Chem* **282**, 14394-402.
21. Majeran, W., Friso, G., van Wijk, K. J. & Vallon, O. (2005) *Febs J* **272**, 5558-71.
22. Groll, M., Ditzel, L., Lowe, J., Stock, D., Bochtler, M., Bartunik, H. D. & Huber, R. (1997) *Nature* **386**, 463-71.
23. Felsenstein, J. (1981) *J Mol Evol* **17**, 368-76.
24. Tavar, S. (1986) in *Some mathematical questions in biology-DNA sequence analysis.*, ed. Miura, R. M., Providence, R.I.), Vol. 17, pp. 57-86.
25. Yang, Z. (1996) *Trends Ecol. Evol.* **11**, 367-372.
26. Thompson, M. W., Singh, S. K. & Maurizi, M. R. (1994) *J Biol Chem* **269**, 18209-15.
27. Beuron, F., Maurizi, M. R., Belnap, D. M., Kocsis, E., Booy, F. P., Kessel, M. & Steven, A. C. (1998) *J Struct Biol* **123**, 248-59.

28. Halperin, T., Zheng, B., Itzhaki, H., Clarke, A. K. & Adam, Z. (2001) *Plant Mol Biol* **45**, 461-8.
29. Earley, K. W., Haag, J. R., Pontes, O., Opper, K., Juehne, T., Song, K. & Pikaard, C. S. (2006) *Plant J* **45**, 616-29.
30. Zhang, X., Henriques, R., Lin, S. S., Niu, Q. W. & Chua, N. H. (2006) *Nat Protoc* **1**, 641-6.
31. Beynon, R. J., Doherty, M. K., Pratt, J. M. & Gaskell, S. J. (2005) *Nat Methods* **2**, 587-9.
32. Mirzaei, H., McBee, J. K., Watts, J. & Aebersold, R. (2008) *Mol Cell Proteomics* **7**, 813-23.
33. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. (1996) *Anal Chem* **68**, 850-8.
34. Fusaro, V. A., Mani, D. R., Mesirov, J. P. & Carr, S. A. (2009) *Nat Biotechnol* **27**, 190-8.
35. Monroe, M. E., Shaw, J. L., Daly, D. S., Adkins, J. N. & Smith, R. D. (2008) *Comput Biol Chem* **32**, 215-7.
36. Zybaylov, B., Sun, Q. & van Wijk, K. J. (2009) *Anal Chem* **81**, 8015-24.
37. Sun, Q., Zybaylov, B., Majeran, W., Friso, G., Olinares, P. D. & van Wijk, K. J. (2009) *Nucleic Acids Res* **37**, D969-74.
38. Stamatakis, A. (2006) *Bioinformatics* **22**, 2688-90.
39. Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007) *Mol Biol Evol* **24**, 1596-9.

CHAPTER THREE

Mass spectrometry-based comparative proteomics of an Arabidopsis Clp mutant; consequences for assembly states of the ClpPR complex, proteome homeostasis and plastid-localized metabolism*

INTRODUCTION

Intracellular proteolysis is essential for proteome homeostasis, regulation of metabolic and signaling pathways and ultimately in the maintenance of organellar and cellular viability. Chloroplasts contain multiple soluble and membrane-bound proteases and processing peptidases (1, 2), that operate in parallel as well as in series, presumably with partially overlapping substrates. The Clp protease system is the most abundant and complex soluble protease family in the plastid. It consists of five serine-type Clp proteases (P1, P3-P6) and four non-proteolytic ClpRs (R1-4) which comprise the Clp protease core, three Clp AAA+ chaperones (C1, C2, D) similar to the *E. coli* ClpA, the adaptor ClpS (homologous to the *E. coli* ClpS) and ClpT1, T2 with unknown functions but with similarity to the N-terminal domain of bacterial ClpA (3, 4). Genetic and phenotypic analyses of various Clp mutants in plants showed that the proteolytic and non-proteolytic subunits have differential functional contributions, but all of them are essential for embryo or seedling development. The plastid-encoded ClpP1 in tobacco is essential for shoot development (5, 6), whereas full inactivation of *CLPP5* gene expression is embryo lethal. The complete loss of *CLPR2* or *CLPR4* is seedling lethal (7). A knockdown mutant in *CLPR2* (*clpr2-1*) and antisense lines against *CLPP4* and *CLPP6* in Arabidopsis exhibited delayed chloroplast and plant

* This work is included in the paper: **Kim, J., Olinares, P.D.B., and van Wijk, K.J.** (2010). Differential contributions of the ClpP subunits in the plastid-localized Clp protease complex in Arabidopsis. In preparation..

development and a virescent or variegated phenotype (8-10). The *Arabidopsis* EMS mutant *clpr1-1* also showed a virescent phenotype (11), but overexpression of *CLPR3* in this background led to full complementation, suggesting that ClpR1 is partially redundant to ClpR3 (12).

Mass spectrometry-based proteomics has enabled large-scale identification and differential profiling of complex proteomes yielding significant insights into relevant biological systems (13). This approach typically involves a liquid chromatography tandem mass spectrometry (LC-MS/MS) setup and employs hybrid mass spectrometers with fast duty cycles for intensity-based sampling of peptide ions. Among the information generated from LC-MS/MS experiments is the number of tandem mass spectral observations for all peptides of a given protein—designated as spectral counts (SPCs)—which have been shown to correlate well with the abundance of the corresponding protein (14-16). This abundance correlation extends over a linear dynamic range of at least two orders of magnitude for complex protein mixtures (14, 15, 17, 18). Since SPCs can be readily obtained within the workflow of identifying proteins from database searches of proteome datasets, spectral counting is a straightforward technique for estimating protein abundances within a sample and for comparing protein amounts across related samples. It also offers a practical alternative to label-based quantification methods which have been limited by expensive reagents for stable isotope labeling or incompatibilities with the biological sample under study and available MS platforms (19).

Maximizing the potential of spectral counting as a quantitative method has involved optimizations throughout the typical proteomics analysis workflow from sample preparation and fractionation, instrument setup, data processing and statistical analysis. Intensity-based peptide sampling in LC-MS/MS is semi-random and largely depends on sample complexity, chromatographic separation and instrument

parameters (14). Considerations on the impact of several of these factors to increase sampling depth have been assessed (17, 20). To more reliably reflect proteome abundances, appropriate transformations of raw SPCs have accounted for peptide length and total SPC within the sample (21) or probability of peptide detection (22). Various statistical programs for significance analysis of spectral counting studies have emerged based mainly on modeling the behavior of SPC datasets (20, 23-27).

We have previously employed the mass spectrometry-based spectral counting methodology in characterizing the unfractionated leaf proteomes of two Clp mutants: one with reduced ClpR2 expression (*clpr2-1*) (28) and a ClpR4 null mutant (*clpr4-1*) (7). Large-scale proteome studies of these mutants identified 2800 proteins and revealed that the strongest effects occurred within the chloroplast consistent with their yellow/pale-green phenotype and delayed growth. Reduced accumulation of the *CLPR2* protein led to the upregulation of chloroplast chaperones and several proteins involved in plastid gene expression (28). A near complete loss of the photosynthetic machinery was observed in the leaf proteome of the *CLPR4* null mutant consistent with its severe albino phenotype and its dependence on exogenous sugar supply for survival (7).

In this study, we characterized a null mutant in *CLPP3* (*clpp3-1*) that exhibits an intermediate phenotype between the *clpr2-1* knockdown and *clpr4-1* null mutant. The *clpp3-1* mutant is seedling lethal but can break this arrest by growth on sucrose as in *clpr4-1*. However, after developing several leaves, the *clpp3-1* can be transferred on soil and can survive autotrophically. To gain insight on the consequences of the loss of *CLPP3*, we performed large-scale comparative analysis by spectral counting of the *clpp3-1* null mutant and wild-type after several weeks of growth on soil. Furthermore, characterization of the oligomeric states of the remaining ClpPR subunits revealed the presence of assembled ClpPR rings and a Clp core complex of higher molecular

weight relative to wild-type with yet undetermined composition. Comparison of the *clpp3-1* proteomes with that of *clpr2-1* and *clpr4-1* showed similar results with downregulation of proteins involved in photosynthesis and upregulation of proteins involved in protein folding and in plastid gene expression.

In addition, we tested several modifications in our spectral counting methodology relative to our previous comparative proteomics studies on Clp mutants. We increased the sampling cycle for each ion that is selected for MS/MS analysis to obtain better quality fragment ion spectra yielding increased spectrum-to-peptide matches and higher SPCs. Moreover, we evaluated two statistical frameworks—one involving conventional signal-to-noise ratio corrections (GLEE) and another employing Bayesian modeling (QSpec) for significance analyses of SPC datasets. Both methods produced comparable results. Overall, an improved SPC acquisition for robust statistical assessment enabled a more comprehensive and deeper proteome analysis of a Clp mutant.

RESULTS

Comparative total leaf proteome analysis of the *clpp3-1* null mutant

The homozygous *clpp3-1* plants are not embryo lethal, but are seedling lethal on soil; however this developmental arrest could be overcome by growth in heterotrophic conditions (with sucrose) (Figure 3.1). After nine weeks of heterotrophic development, the mutants can be transferred on soil, grow autotrophically thereafter, flower and yield viable seeds.

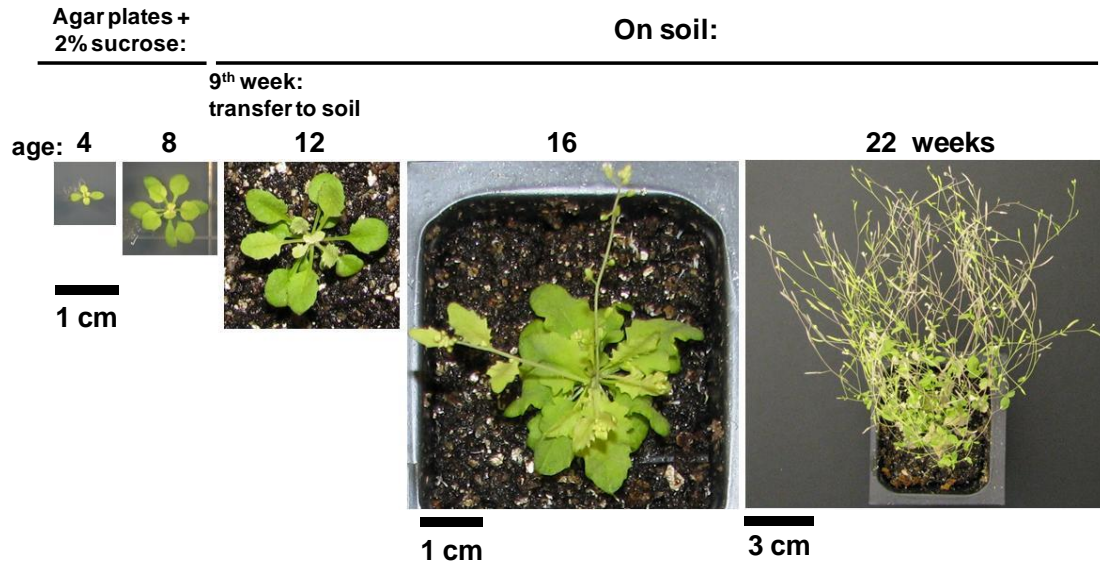


Figure 3.1. Development of the homozygous *clpp3-1* mutant. Plants were first grown on agar plates with 0.5xMurashige-Skoog medium and 2% sucrose under 10/14 hour light/dark cycle at $40 \mu\text{mol photons.m}^{-2}.\text{s}^{-1}$. The plants were then transferred to soil after nine weeks and grown under 10/14 hour light/dark cycle at $100 \mu\text{mol photons.m}^{-2}.\text{s}^{-1}$. Images taken by Jitae Kim.

To gain insight in the consequences of the loss of *CLPP3*, we compared the total leaf proteome of pale-green *clpp3-1* plants initially grown on heterotrophic conditions (agar plates + 2% sucrose) and then transferred to soil for autotrophic growth and soil-grown wild-type plants (Figure 3.2A). Leaf material was harvested from both genotypes at the same developmental stage (both had 20 leaves) (see (29) for discussion on correcting for developmental delay in proteome experiments). Total leaf proteomes were extracted with SDS and 50 μg of each proteome was separated by SDS-PAGE and visualized by Coomassie staining (Figure 3.2B). Each gel lane was excised in 20 slices, followed by in-gel trypsin digestion and protein identification by on-line nano-liquid chromatography tandem mass spectrometry (nanoLC-MS/MS) with an LTQ-Orbitrap, as described in (30) (see Figure 3.2C). Three biological replicates per genotype were analyzed resulting in 120 MS/MS runs. After database search, 22% of the 800,000 MS/MS total acquired spectra were matched to tryptic

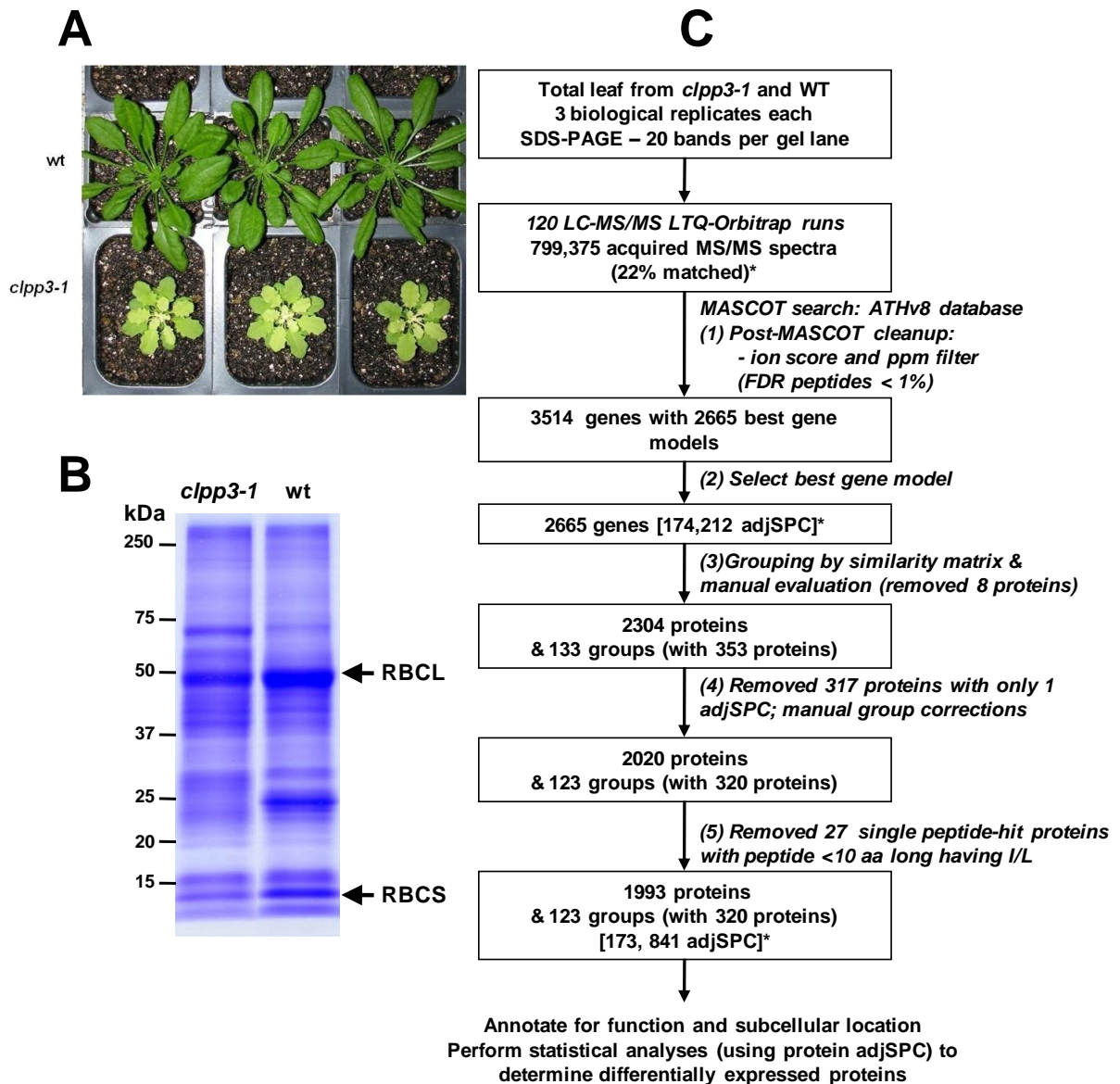


Figure 3.2. Large-scale SPC-based comparative proteomics of the *clpp3-1* mutant. **A)** The *clpp3-1* and wt plants used for proteome analysis. Wild type (Col-0) plants were grown on soil for 30 days under a short day cycle (10h/14h of light/dark) at $100 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$. Homozygous *clpp3-1* plants were initially grown on agar plates with 0.5x Murashige-Skoog medium and 2% sucrose under short day conditions at $40 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$, and then transferred to soil and grown under short day cycle at $100 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ for 70 days. **B)** 1D-SDS PAGE gel separation (biological replicate 1) of the total leaf proteomes from *clpp3-1* and wt. The highly abundant Ribulose-1,5-bisphosphate carboxylase/oxygenase subunits (RBCL and RBCS) are indicated. **C)** MS analysis and bioinformatics workflow for the comparative proteome analysis. adjSPC, adjusted spectral counts.

peptides. This match rate increased to 25% when results from semi-tryptic searches were included (data not shown). The search results were further processed (Figure 3.2C) to reduce false positive identifications (steps 2, 5 and 6), select the best gene model for each gene (step 3) and group proteins that shared more than 80% of their MS/MS spectra to avoid overidentification of proteins belonging to protein families (step 4) using the workflow as described in (31). Protein annotations for subcellular location, function and other relevant physico-chemical features were obtained from PPDB at (<http://ppdb.tc.cornell.edu/>) (32). The MapMan system (33) was used for functional classification. This workflow resulted in the identification and quantification of 1993 proteins and 123 protein groups. 37% of the total identified proteins were chloroplast-localized (Table 3.1).

We then derived unique SPCs (MS/MS spectra that uniquely matched) and adjusted SPCs (adjSPCs) for each protein. The latter is the sum of unique SPCs and SPCs from shared peptides for proteins belonging to a protein family distributed in proportion to the unique SPCs of the proteins in the group, if applicable. Using protein adjSPCs as abundance measures, the chloroplast-to-total leaf proteome mass was reduced in *clpp3-1* relative to wild-type levels (61% in mutant vs. 68% in wild-type) consistent with the mutant's yellow/pale-green leaf phenotype, the requirement for exogenous sugar during the early phase of seedling development and delayed growth. This decrease was also observed in proteome analyses of a knockdown (*clpr2-1*) and a knockout (*clpr4-1*) Clp mutant (7, 8, 28).

Significance analysis for determining differentially expressed proteins

The average pairwise correlation coefficients among the three biological replicates within the wild-type and *clpp3-1* datasets were 0.990 and 0.983, respectively, indicating high reproducibility between the independent replicates analyzed for each

genotype. To determine which of the 2116 proteins (including protein families) were differentially expressed in *clpp3-1* relative to wild-type, two statistical tools for significance analysis of large-scale spectral count datasets were employed.

Table 3.1. Summary of the subcellular location of quantified proteins (N= 2116) in the total leaf comparative proteomics of *clpp3-1* and wild-type.

Predicted Location (using TargetP) ^a	Proteins
cTP	832 (39%)
mTP	221 (10%)
sTP	234 (11%)
None	829 (39%)
Curated subcellular location (from PPDB) ^b	
Chloroplast	779 (37%)
stroma (including nucleoid)	374 (18%)
thylakoid (including lumen and plastoglobuli)	186 (9%)
envelope (inner and outer)	69 (3%)
Mitochondria	62 (3%)
Cytosol	142 (7%)
Peroxisome	24 (1%)
Plasma membrane	31 (1%)
Others (nucleus, ER, golgi, vacuole)	43 (2%)
None	1034 (49%)

^a Subcellular location was predicted by TargetP (34). cTP, chloroplast transit peptide; mTP, mitochondrial transit peptide; sTP, secretory pathway signal peptide.

^b Curated locations are from the Plant Proteome Database (PPDB) at (<http://ppdb.tc.cornell.edu/>).

Due to the stochastic nature of data-dependent peptide selection and fragmentation, proteins with fewer SPCs exhibit greater variations thereby yielding less reliable quantitative data and confounding the detection of biologically relevant differences between samples (20). One approach is to model the relationship between average signal as measured by adjSPCs and the corresponding variation for better assessment of signal-to-noise such that abundance differences between two samples can be confidently judged. This mean-variance dependence was modeled using a linear log-equation (also known as power law) in the PLGEM (power law global error

estimation) software (20). However, this assumption of linearity was deemed inadequate for variation adjustments of high and low abundant proteins. As such, a modification of PLGEM, known as GLEE (global logarithm error estimation), which employs a third-order polynomial fitting of mean-variance behavior was developed to account for variations particularly along the low and high adjSPC ranges[†].

In contrast to empirical mean-variance fitting for signal-to-noise correction in GLEE, QSpec employs a probabilistic approach through a hierarchical Bayesian strategy which treats SPCs as observations following a Poisson distribution (23). Bayesian modeling was implemented through Markov chain Monte Carlo simulations of the appropriate distributions. To increase statistical power, QSpec pools statistical information across all quantified proteins for analysis (23). The main QSpec statistic is a Bayes factor (BF) for each protein which is the ratio of the likelihood of the SPC distribution fitting a regression model that includes a differential parameter versus the null model. Thus, a higher BF signifies higher probabilistic evidence that the model with a differential parameter better fits the data and that the corresponding protein is indeed differentially expressed (23).

The adjSPC of a protein normalized to the total adjSPCs in the sample (NadjSPC) provides an estimate of its amount (mass) in the sample. As an extension, various corrections and normalization have been tested to derive protein concentration within each sample including adjustment by protein length (21) or by number of observable tryptic peptides (22, 30) yielding spectral abundance factors (SAFs). The PLGEM software (the predecessor of GLEE) and QSpec recommends the use of SAFs instead of adjSPCs as input for statistical analyses. However, evaluation of the results

[†] **Poliakov, A., Ponnala, L., Olinares, P.D., Asakura, Y., and van Wijk, K.** (2010). A statistical solution for pair-wise comparative proteome analysis using large scale label-free spectral counting. In preparation.

using adjSPC and SAF values as input revealed that the abundances of longer or shorter proteins are under- or overestimated for SAF-based quantifications, respectively leading to flawed significance assessment. Ultimately, the component peptides of the proteins were sampled and sequenced in the mass spectrometer and not the whole protein itself. As such, the variation in the raw data (SPCs) is independent of protein length. The bias against long or short proteins was particularly more pronounced with PLGEM and GLEE where SAF adjustments led to erroneous variance estimations due to arbitrary signal adjustments by protein length (not shown). QSpec was more tolerant of SAF adjustments as they were considered auxiliary components of the regression models used (23). Given these considerations, our current study used protein adjSPCs and not SAFs for GLEE and QSpec analysis.

Figure 3.3A compares the results of the significance analyses of the total leaf datasets using QSpec (BF > 10) and GLEE (p-value <0.01) at 5% false discovery rate (FDR). Both methods detected a similar number of differentially expressed proteins. 65 out of the 264 proteins that were considered significant in both tests had unknown curated location. 164 out of the 199 proteins with known localizations (85%) were chloroplast-targeted suggesting that the chloroplast was strongly affected by the inactivation of ClpP3 expression. Examination of the 30 differentially expressed proteins that were localized as nonplastid did not show any functional trends suggesting no major effects outside the plastid in the mutant. Most of the proteins that were only found significant in either test had zero values in one or more replicates in either genotype. We note that GLEE replaces zero values with one SPC (the least number of SPC) whereas QSpec considers zero values as missing data points along the Poisson-modeled SPC distribution.

The chloroplast proteome comprised more than 50% of the total leaf proteome by mass in both wild-type and *clpp3-1*. Moreover, the chloroplast proteome was reduced

in the *clpp3-1* mutant relative to wild-type (61% in mutant vs 68% in wild-type). To better assess the impaired state of the *clpp3-1* chloroplasts, we also performed significance analyses using QSpec and GLEE on the chloroplast proteome subset as shown in Figure 3.3B. As with the leaf analysis, both tests showed similar results and proteins unique to either test also involved zero values. Table 3.2 lists the 161 chloroplast-localized proteins/protein families that were significantly up- or downregulated in the *clpp3-1* background from both QSpec and GLEE tests. In addition, accumulation ratios between the *clpp3-1* and wild-type (fold-changes) for chloroplast proteins were normalized to their respective total chloroplast proteome amounts.

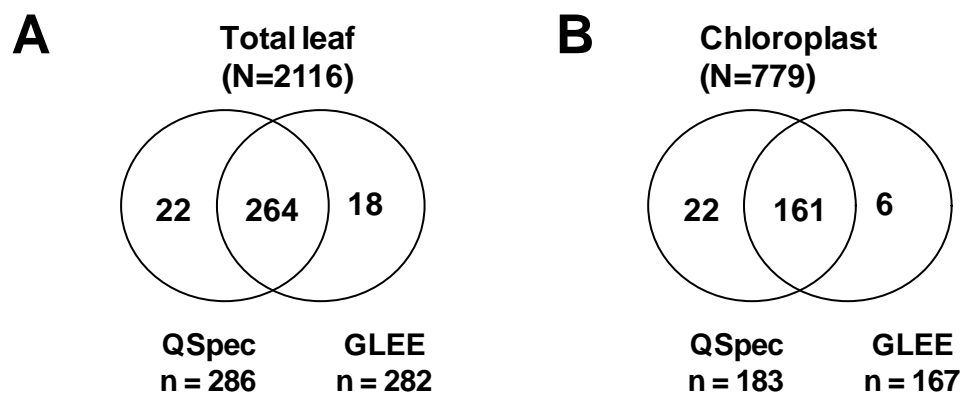


Figure 3.3. Comparison of proteins determined to be significantly expressed in the *clpp3* background relative to wild-type in (A) leaf and (B) chloroplast datasets. This includes proteins that have BF>10 for QSpec or p-value <0.01 for GLEE both at 5% FDR level.

Comparison with previous comparative proteomics studies of Clp mutants

We have previously characterized two Clp protease mutants in Arabidopsis by large-scale SPC-based proteomics using a nanoLC-LTQ-Orbitrap platform (7, 28). We employed the same LC-MS setup in this study but we loaded less protein sample (four-fold less) with a corresponding smaller gel electrophoresis format (3x less protein per gel volume). The two proteomics studies and the current study all involved

excision of 20 gel slices and in-gel tryptic digestion was performed in a similar manner. In addition, we also modified our MS/MS acquisition parameters (Table 3.3). As shown in Figure 3.4A, we generated our MS precursor survey scan and MS/MS fragment scan readouts from the Orbitrap and LTQ analyzers, respectively. We set our survey scan at maximum allowable resolution ($R=100,000$ at 400 m/z) in the Orbitrap to enable high accuracy precursor mass data for unambiguous peptide matches. This also enabled confident identification of biologically relevant post-translational modifications and RNA editing sites as we have previously shown (35). Parallel to the MS survey scan acquisition in the Orbitrap is the selection of the top five most abundant ions for isolation, fragmentation by collision induced dissociation (CID) and fragment ion spectrum acquisition in the linear ion trap (LTQ). A snapshot MS spectrum in the Orbitrap ($R=7,500$ at 400 m/z , acquire at ~ 250 ms) was used for ion selection in the LTQ. With this setting and with one MS/MS microscan for each ion selected for fragmentation (about 250 ms each), the median duty cycle in our two previous studies was 1.8s (Table 3.3). This slightly varies due to variable times in filling both traps to reach the ion population required for detection and sampling (target value). Nevertheless, trap fill times are typically low and do not appreciably contribute to overall duty cycle times (36). In our current workflow, we intended to increase our matched rates by increasing the frequency of sampling ions selected for fragmentation. In this case, we required an average of three microscans for each MS/MS acquisition (~ 700 ms per MS/MS scan). This led to a two-fold increase in the overall duty cycle relative to previous studies (Figure 3.4A).

Table 3.2. Chloroplast-localized proteins that are significantly upregulated or downregulated in *clpp3-1* plants relative to wildtype ^a

Accession	Protein	Loc ^b	Function ^c	<i>clpp3/</i> <i>wt</i> ^d	Direction ^e		
					<i>p3</i>	<i>r2</i>	<i>r4</i>
AT1G08380.1	psaO	TI	PSI	0.32*	D		
AT1G52230.1	psaH-2 - subunit VI	TI	PSI	0.43	D		D
AT4G12800.1	psaL - subunit XI	TI	PSI	0.48	D		D
AT4G28750.1	psaE-1 subunit IV	TS	PSI	0.52	D		
AT1G31330.1	psaF- subunit III	TI	PSI	0.53	D	D	D
AT5G64040.1	psaN - TAT LTP	TL	PSI	0.62	D		D
ATCG00340.1	psaB - subunit Ib	TI	PSI	0.70	D		
AT1G61520.1	LHCI-3 CAB4	TI	PSI	0.73	D	D	
AT1G45474.1	LHCI-5	TI	PSI-NDH assembly	2.69	U		
ATCG00520.1	YCF4	S	PSI assembly	3.41*	U		
AT4G05180.1	psbQ OEC16-like Tat ITP	TL	PSII	0.16	D	D	D
AT2G34420.1	LHCII-1.5	TI	PSII	0.23	D		
AT3G55330.1	OEC23-like-4 Tat ITP TL25.6	TL	PSII	0.25	D		
AT5G66570.1	psbO OEC33	TL	PSII	0.50	D	D	D
AT3G50820.1	psbO OEC33-like	TL	PSII	0.52	D	D	D
ATCG00680.1	psbB CP47	TI	PSII	0.54	D	D	D
AT1G06680.1	PsbP-1 OEC23 Tat ITP	TL	PSII	0.56	D	D	
AT4G21280.1	psbQ OEC16 Tat ltp	TL	PSII	0.59	D		D
ATCG00270.1	psbD D2	TI	PSII	0.60	D	D	
AT3G08940.2	LHCII-4.2	TI	PSII	0.61	D		
ATCG00020.1	psbA D1	TI	PSII	0.62	D		
ATCG00280.1	psbC CP43	TI	PSII	0.68	D	D	D
AT4G10340.1	LHCII-5 CP26	TI	PSII	0.71	D	D	
AT3G27690.1							
AT2G05070.1	LHCII-2.3, 2.2, 2.1	TI	PSII	0.71	D	D	
AT2G05100.1							
AT3G47070.1	thylakoid phosphoprotein (TSP9)	TS	PS state transtion	0.21*	D		
ATCG00720.1	petB - Cytochrome b6	TI	Cytochrome b6/f	0.39	D		D
ATCG00730.1	petD - subIV	TI	Cytochrome b6/f	0.45*	D		
ATCG00540.1	petA - cytochrome f	TI	Cytochrome b6/f	0.50	D		D
AT4G03280.1	petC - Rieske Fe-S protein	TL	Cytochrome b6/f	0.60	D		D
ATCG00470.1	CF1e - atpE	TS	ATP synthase	0.69	D		D
AT4G04640.1	CF1y - atpC	TS	ATP synthase	0.69	D		
AT2G05620.1	PGR5	TS	Chlororespiration	0.31*	D		
ATCG01100.1	NDH A (NDH-1)	TI	NDH complex	0.21	D		
ATCG01110.1	NDH H (NDH-7)	T	NDH complex	0.36	D		
AT2G39470.1	NDH subunit PPL2	TL	NDH complex	0.46	D		
AT2G47910.1	CRR6	S	NDH complex assembly	3.89*	U		
AT5G38410.1	Rubisco small subunit 3b and 2b						
AT5G38420.1	(RBCS-3B, 2B)	S	Calvin cycle	0.45	D	D	D
AT5G38430.1	Rubisco small subunit 1b (RBCS-1b)	S	Calvin cycle	0.51	D	D	D
ATCG00490.1	Rubisco large subunit (RBCL)	S	Calvin cycle	0.56	D	D	D
AT1G67090.1	Rubisco small subunit-4 (RBCS-4)	S	Calvin cycle	0.60	D		D
AT2G46910.1	Fibrillin (FIB8)	PG	PG coat	3.35*	U		

Table 3.2 (continued)

AT4G04020.1	Fibrillin (FIB1a)	PG	PG coat	3.74	U	U	
AT4G22240.1	Fibrillin (FIB1b)	PG	PG coat	6.06	U	U	U
AT2G21330.1	Fructose-bisphosphate aldolase-1 (SFBA-1)	PG + S	PG metabolism	1.48	U	U	
AT5G42650.1	allene oxide synthase (AOS)	PG	PG metabolism	2.14	U		D
AT5G08740.1	NADH dehydrogenase (NDC1)	PG	PG metabolism	3.78*	U		
AT4G19170.1	9-cis-epoxycarotenoid dioxygenase (CCD4)	PG	PG metabolism	4.35*	U		
AT1G79600.1	ABC1 kinase 2	PG	PG metabolism	12.43*	U		U
AT4G31390.1	ABC1 kinase 3	PG	PG metabolism	3.25*	U		
AT3G04870.1	ζ-carotene desaturase (ZDS)	S	PG metabolism & carotenoid synthesis	0.49	D		
AT1G06950.1	Tic110	IE	Protein targeting	2.09	U	U	
AT4G01800.1	cpSecA	TS	Protein targeting	2.28	U		
AT2G18710.1	cpSecY	TI	Protein targeting	2.73*	U		
AT4G14870.1	cpSecE	TI	Protein targeting	5.11*	U		
AT5G13410.1	Isomerases TAT ITP	TL	Protein folding	0.42	D		
AT3G01480.1	CYCLOPHILIN 38 (CYP38)	TL	Protein folding	0.58	D		
AT3G62030.1	Peptidylprolyl isomerase ROC4	S	Protein folding	1.57	U	U	
AT5G55220.1	Trigger Factor	S	Protein folding	1.71	U		
AT2G28000.1	Cpn60-alpha-1	S	Protein folding	2.10	U	U	
AT3G13470.1	Cpn60-beta-1 & 2	S	Protein folding	2.33	U	U	
AT5G56500.2	Cpn60-beta-3	S	Protein folding	9.46	U	U	
AT5G20720.1	Cpn21 (also Cpn20)	S	Protein folding	1.54	U		
AT2G44650.1	Cpn10-1	S	Protein folding	2.25	U		
AT2G04030.1	cpHSP90	S	Protein folding	2.40	U	U	U
AT4G24280.1	cpHSP70-1, 2 (DnaK homologue)	S	Protein folding	2.53	U	U	U
AT5G49910.1							
AT5G15450.1	ClpB3	S	Protein folding	3.88	U	U	U
AT5G42390.1	Stromal processing peptidase (SPP)	S	Protease	3.55	U		
AT4G30920.1	Leucyl aminopeptidase (LAP2)	S	Protease	2.19	U		
AT3G19170.1	AtPrep1	S	Protease	2.44	U	U	U
AT1G73990.1	SppA	TI	Protease	2.97	U		
AT5G05740.1	EGY2	T	Protease	3.38*	U		
AT3G04260.1	SAP domain-containing protein (pTAC3)	S	DNA binding	2.85*	U		
AT1G80480.1	PRLI-interacting factor L (pTAC17)	S	DNA binding	3.24*	U		
AT5G35970.1	Deadbox DNA helicase-related	S	DNA binding	4.25	U		
AT3G10690.1	DNA gyrase subunit A (TAC)	S	DNA repair	3.67	U		
AT1G09750.1	CND41	S	Transcription regulation	0.39*	D		
AT5G14260.1	SET domain-containing protein	S	Transcription regulation	1.97	U		
AT1G14410.1	Why1 (pTAC1)	S	Transcription regulation	2.49*	U		U
AT2G02740.1	Why3 (pTAC11)	N	Transcription regulation	3.26*	U		
AT4G16390.1	RNA-binding protein P67	S	RNA binding	2.67	U		
AT4G09040.1	RNA recognition motif (RRM) protein	S	RNA binding	4.5*	U		
AT3G03710.1	3'-5' exoribonuclease (RIF10)	S	RNA processing	1.67	U		
AT5G26742.1	RNA helicase 3 (RH3)	S	RNA processing	2.18	U	U	U

Table 3.2 (continued)

AT5G63420.1	RNase J homologue	S	RNA processing	3.68*	U		
AT1G70070.1	DEAD/DEAH box helicase	S	RNA processing	4.74*	U		
ATCG00830.1	50S ribosomal protein L2A & L2B	S	Protein synthesis	0.50	D		
ATCG01310.1							
AT3G44890.1	50S ribosomal protein L9	S	Protein synthesis	0.62	D		
AT3G20230.1	50S ribosomal protein L18	S	Protein synthesis	4.57*	U		
AT5G40950.1	50S ribosomal protein L27	S	Protein synthesis	0.47	D		
AT4G29060.1	pETs (fusion of EF-Ts and PSRP-7)	S	Protein synthesis	1.43	U	U	U
AT4G20360.1	Elongation factor Tu (EF-Tu-1)	S	Protein synthesis	1.72	U	U	U
AT1G62750.1	Elongation factor Tu-G (EF-G) (sco1)	S	Protein synthesis	2.37	U	U	
AT5G13650.1	TypeA/bipA like	S	Protein synthesis	3.43	U	U	U
AT3G62910.1	Peptide chain release factor, AtPRF1	S	Protein synthesis	3.22*	U		
AT1G16720.1	HCF173	S	Protein synthesis	2.14	U		
AT3G48110.1	Gly t-RNA synthetase (GlyRS-2) (EDD1)	S	Protein synthesis	2.18	U		
AT5G52520.1	tRNA synthetase class II	S	Protein synthesis	2.35*	U		
AT5G49030.1	Ile-tRNA synthetase class II (OVA2)	S	Protein synthesis	2.55	U		
AT4G25130.1	Methionine sulfoxide reductase type A 4	S	Protein synthesis	2.68*	U		
AT4G33760.1	tRNA synthetase class II (D, K and N)	S	Protein synthesis	3.01	U		
AT5G16715.1	Val-tRNA synthetase (ValRS-2)	S	Protein synthesis	3.21*	U		
AT2G36250.1	FtsZ2.1	S	Plastid division	2.02	U		U
AT3G52750.1	FtsZ2.2	S	Plastid division	2.75*	U		
AT2G38040.1	Carboxyltransferase (CT) alpha subunit	IES	FA synthesis	0.55	D		
AT3G22960.1	pyruvate kinase-1	S	FA synthesis	1.79	U		
AT5G35360.1	biotin carboxylase (BC)	IES	FA synthesis	1.99	U		
AT2G30200.1	malonyl-CoA:Acyl carrier protein transacylase	S	FA synthesis	2.00	U		
AT4G14070.1	Plastidial Long-Chain Acyl-CoA Synthetase	E	FA synthesis	2.20	U		
AT4G16155.1	E3 - dihydrolipoamide dehydrogenase 2 (ptlpd2)	S	FA synthesis	2.96*	U		
AT5G62790.1	1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR)	S	Isoprenoid synthesis (non-MVA)	1.80	U		
AT5G60600.1	4-hydroxy-3-methylbutyl diphosphate synthase (HDS)	TS	Isoprenoid synthesis (non-MVA)	3.27	U	U	U
AT4G27440.1	PORB	TS	Tetrapyrrole synthesis	0.44	D		
AT5G08280.1	Hydroxymethylbilane synthase (HEMC)	S	Tetrapyrrole synthesis	0.60	D		
AT1G69740.1	δ -aminolevulinic acid dehydratase-1 (ALAD-1)	S	Tetrapyrrole synthesis	1.66	U		U
AT2G43750.1	Cysteine synthase	S	Amino acid metabolism	0.55	D		
AT3G58610.1	Ketol-acid reductoisomerase	S	Amino acid metabolism	1.70	U		U
AT1G18500.1	2-isopropylmalate synthase (IMS1)	S	Amino acid metabolism	2.24	U		
AT1G74040.1							
AT4G29840.1	Threonine synthase (MTO2)	S	Amino acid metabolism	2.25	U		

Table 3.2 (continued)

AT4G13430.1	Isopropylmalate isomerase large subunit	S	Amino acid metabolism	2.27	U	
AT4G31990.1	Aspartate aminotransferase (AAT1/Asp5)	S	Amino acid metabolism	2.78	U	U
AT5G48960.1	5'-nucleotidase	S	Nucleotide metabolism	2.94*	U	
AT3G18680.1	Asp/Glu-uridylylate kinase	S	Nucleotide metabolism	4.1*	U	
AT5G04140.1	Fd-GOGAT 1	S	N-metabolism	1.45	U	U
AT5G53460.1	NADH-GOGAT (GLT1)	S	N-metabolism	10.78	U	U
AT4G04770.1	ATNAP1	S	S-assimilation	3.06*	U	
AT1G32500.1	SufD (AtNap6)	S	S-assimilation	3.71*	U	
AT1G10760.1	Water dikinase (Sex1)	S	Starch degradation	1.57	U	U
AT5G26570.1	Phosphoglucan water dikinase (PWD)	S	Starch degradation	2.22	U	
AT1G69830.1	α -amylase (AtAMY3)	S	Starch degradation	2.71	U	U
AT5G24300.1	Starch synthase 1 (SS1)	S	Starch synthesis	1.98	U	
AT5G19220.1	ADP-glucose pyrophosphorylase (ADG2)	S	Starch synthesis	2.08	U	
AT1G32900.1	Starch synthase	S	Starch synthesis	2.19	U	
AT1G04420.1	Aldo/keto reductase family protein	S	Sugar metabolism	2.19	U	
AT1G09830.1	Phosphoribosylamine-glycine ligase (PUR2)	S	Purine synthesis	3.77*	U	
AT1G29900.1	Carbamoylphosphate synthetase	S	Pyrimidine synthesis	1.92	U	U
AT3G26060.1	Peroxiredoxin Q (Prx Q)	S	ROS defense	0.32	D	
AT3G54660.1	Glutathione reductase,	S	ROS defense	2.67	U	
AT4G23100.1	γ -glutamylcysteine synthetase (GSH1)	S	ROS defense	3.05	U	
AT3G06730.1	Thioredoxin family protein	S	ROS defense	3.05*	U	
AT5G01600.1	Ferritin-1	S	Fe storage	4.83*	U	U
AT5G12470.1	MEP3	E	Metabolite transporter	0.34*	D	
AT5G23890.1	MEP2	E	Metabolite transporter	2.22	U	U
AT4G00630.1	K ⁺ efflux antiporter	E	Metabolite transporter	2.70*	U	
AT3G49560.1	Inner membrane translocase subunit (Tim17)	E	Metabolite transporter	3.13*	U	
AT5G24650.1	Inner membrane translocase subunit (Tim17/22)	E	Metabolite transporter	3.18	U	
AT1G80300.1	ATP/ADP translocator 1 (AtNTT1)	IE	Metabolite transporter	6.49*	U	
AT1G15500.1	ATP/ADP translocator 2 (AtNTT2)	IE	Metabolite transporter	11.15*	U	U
AT1G17650.1	Glyoxylate reductase 2 (GR2)	S	OPP	2.09	U	
AT1G64190.1	6-phosphogluconate dehydrogenase-1	S	OPP	2.87*	U	
AT1G74730.1	unknown protein	T	Unknown	0.19*	D	
AT5G08050.1	unknown protein	T	Unknown	0.34*	D	
AT2G26340.1	unknown protein	T	Unknown	0.36*	D	
AT2G42220.1	Rhodanese-like protein	T	Unknown	0.41	D	
AT1G71500.1	Rieske [2Fe-2S] domain	T	Unknown	0.42	D	
AT5G02940.1	unknown protein	E	Unknown	2.67*	U	

Table 3.2 (continued)

AT2G31890.1	unknown protein	S	Unknown	2.78	U
AT3G63170.1	unknown protein	S	Unknown	2.84*	U
AT3G04550.1	unknown protein	S	Unknown	3.14*	U
AT2G44640.1	unknown protein	E	Unknown	4.35*	U
AT2G39670.1	Radical SAM-containing protein	S	Unknown	4.67*	U
AT5G51110.1	unknown protein	S	Unknown	7.62*	U

^a Only proteins that passed the significance analyses by both QSpec (Bayes Factor > 10) and GLEE (p-value < 0.01) at 5% FDR threshold are shown. Proteins were normalized to total chloroplast protein abundance in the respective samples.

^b Subplastidial location: E - envelope; IE - inner envelope membrane; IES - inner envelope membrane-associated, stroma side; S - stroma; PG - plastoglobule; TS - thylakoid membrane-associated, stroma side; TI - thylakoid membrane, integral-bound; TL - thylakoid membrane-associated, lumenal side; L - lumen

^c PS, Photosystem; NDH, NADH dehydrogenase; FA, fatty acid; MVA, mevalonate pathway; ROS, reactive oxygen species; OPP, oxidative pentose phosphate pathway

^d Protein accumulation ratio between *clpp3-1* mutant and wildtype (three biological replicates each) obtained from QSpec calculations.

^e Based on the *clpp3-1*/wt accumulation ratios, the protein can be upregulated (U) or downregulated (D). These were derived from this study (*p3* dataset), (28) (*r2* dataset) and (7) (*r4* dataset).

* had zero value(s) in one or more replicates. As such, the fold-change involved an imputation for zero values.

As shown in Table 3.4, we acquired a comparable or slightly lower number of MS/MS spectra in this study relative to the previous studies as a consequence of longer duty cycles at similar overall LC-MS analysis time. However, more spectra were matched to peptides in this study both in absolute (average adjSPCs) and relative (peptide match rate) terms. This gain in adjSPCs (22% in this study vs. 14% in previous studies) is most likely due to the acquisition of higher quality MS/MS spectra. In terms of proteins identified based on these matched peptides per replicate (one replicate involves one gel lane with 20 excised gel bands that were processed for MS analysis), there was a slightly lower number of proteins identified per replicate in *clpp3-1* relative to *clpr4-1*. However, the number of chloroplast proteins quantified was higher in *clpp3-1* suggesting that a considerable number of identified proteins in *clpr4-1* had low adjSPCs (<2 adjSPCs). In the two previous studies, quantification was only performed for chloroplast proteins with total adjSPC > 10 summed from both

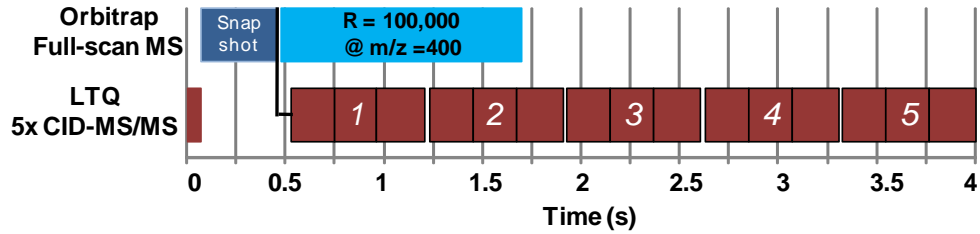
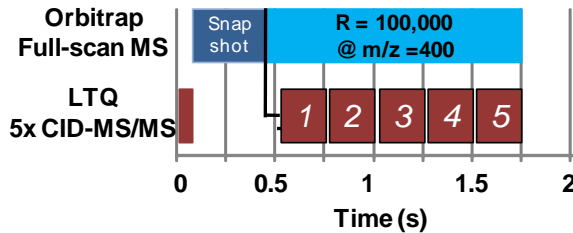
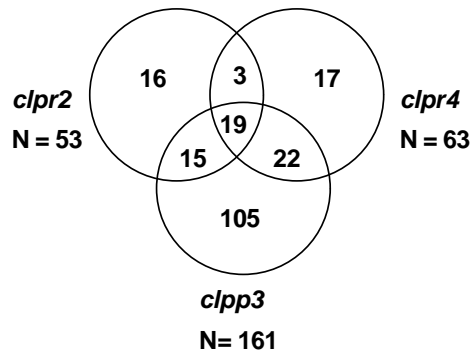
A*clpp3* proteomics setup:*clpr2* and *clpr4* proteomics setup:**B**

Figure 3.4. Comparison of LTQ-Orbitrap data acquisition cycle and statistical analysis of SPC-based comparative proteomics of a *clpr2* (Zybailov et al, 2009), *clpr4* (Kim et al, 2009) and *clpp3* (this study) mutants. A) Acquisition schemes for the three proteomics studies with full MS scan in the Orbitrap and CID-MS/MS scans in the LTQ. MS survey scan acquisition in the Orbitrap and the selection of the top five most abundant ions for isolation, fragmentation by collision induced dissociation (CID) and fragment ion spectrum acquisition in the linear ion trap (LTQ) were performed in parallel. A snapshot MS spectrum in the Orbitrap ($R=7,500$ at 400 m/z , acquisition time ~ 250 ms) was used for ion selection in the LTQ. B) Comparison of the number of chloroplast-localized proteins that passed the significance thresholds for differential expression in the three comparative proteomics studies. Pairwise G-test was used for statistical analysis of *clpr2* and *clpr4* with a 95% confidence threshold and correction for multiple hypothesis testing. GLEE (p -value < 0.01) and QSpec ($BF > 10$) was employed for significance analysis of the *clpp3-1* dataset. The differentially expressed proteins found in *clpr2* or *clpr4* that were also detected in *clpp3* are indicated in Table 3.2.

genotypes. Of the 779 quantified chloroplast proteins in *clpp3-1*, 163 had a total adjSPC below 10. Even with this correction, still more proteins were quantified in this study. Furthermore, as a result of more adjSPC per protein (higher signal) and detection of low abundant proteins with sufficient adjSPC for statistical analysis, more proteins passed the significance thresholds imposed by two independent statistical tools in the current study (Figure 3.4B) assuming that the underlying biology is the same for all the Clp mutants being compared.

Table 3.3 Sample preparation and MS data acquisition parameters used among the Clp comparative proteomic studies.^a

Study	Leaf stage ^b	Growth	Sample loaded (ug)	Protein/gel volume (ug/mm ³)	Target Ions (x 10 ⁴) ^c		Microscans per scan		Median duty cycle (s)
					MS (Orbi)	MS ² (LTQ)	MS (Orbi)	MS ² (LTQ)	
<i>r2</i>	1.07, 1.14	soil	200	0.3	50	5	1	1	1.8
<i>r4</i>	1.14	sucrose	200	0.3	20	10	1	1	1.7
<i>p3</i>	1.20	sucrose then on soil	50	0.1	50	10	1	3	4

^a*r2*: from (28), with three biological replicates and two technical replicates each, *r4*: from (7), with two biological replicate and one technical replicate, *p3*: this study with three biological replicates and one technical replicate each. 20 gel slices were excised and analyzed per biological replicate.

^bCorresponds to the number of leaves in the harvested plants. Wildtype and Clp mutant pairs were harvested and analyzed at similar developmental stages.

^c Ion population sampled per microscan in each analyzer.

Table 3.4 Evaluation of peptide matched rates and number of proteins identified among the Clp comparative proteomic studies.^a

Study	MS/MS Spectra	MS/MS spectra matched (adjSPCs) ^b	Peptide match rate (%) ^b	Proteins matched	Plastid proteins quantified ^c	Statistics used
<i>r2</i>	115,749 ± 51,000	15044 ± 4477	14 ± 1	1310 ± 285	384	G-test
<i>r4</i>	190,800 ± 5824	25943 ± 7288	14 ± 7	2124 ± 642	298	G-test
<i>p3</i>	133, 229 ± 3616	29035 ± 851	22 ± 4	1817 ± 130	779 (616)	Qspec & GLEE

^a Shown are average values across all replicates irrespective of genotypes per study with S.D. for *r2* (n=6, only wildtype datasets were *considered*); from (28), *r4* (n=4): from (7), *p3* (n=6): current study.

^b Only for MS/MS spectra matched to tryptic peptides.

^c For the *r2* and *r4* datasets, only proteins with total adjSPCs >10 for across wild-type and mutant replicates. For the *p3* dataset, total adjSPCs>2 across all replicates. Also indicated in parenthesis are proteins with total adjSPCs>10.

In the next sections, functional analysis was performed with focus on chloroplast-localized proteins whose abundance levels in the *clpp3-1* mutant were considered significantly changed relative to wild-type by both QSpec and GLEE.

Photosynthesis

116 proteins involved in photosynthesis and associated processes (cyclic electron flow and light stress) were quantified. Most of the proteins involved in photosynthesis were two-fold downregulated (Table 3.2 and Figure 3.5) except for several enzymes that are shared among carbon fixation, glycolysis and oxidative pentose phosphate pathways whose accumulation levels remained unaltered (marked as other Calvin cycle enzymes in Figure 3.5). With regards to the linear electron transport pathway, the accumulation of the two photosystem complexes (PSI and PSII) and cytochrome b6/f complex was diminished (about two-fold) and the light-harvesting antennae (LHCI and LHCII), ferredoxin reductase complex and ATP synthase complex were all slightly decreased (*clpp3-1/wt* ~ 0.9) in *clpp3-1*. Immunoblot analyses of two luminal proteins namely OEC23 and PsaF confirmed this trend (Figure 3.6). Furthermore, the unprocessed

form of Psaf was also found to accumulate in *clpp3-1* indicating problems in protein processing (Figure 3.6). Proteins involved in cyclic electron flow (PGR5, PGRL, PIFI and the NDH complex) were similarly downregulated (*clpp3-1/wt* ~0.6). The amounts of state transition proteins (Stt7, TSP9, CS) were unaltered (Figure 3.5).

The energy and reducing equivalents generated from light energy capture and linear electron flow are used to fix carbon dioxide into reduced carbohydrates within the Calvin cycle. Several enzymes in this pathway perform reversible reactions and are also shared in glycolysis and oxidative pentose pathways involved in sugar breakdown for energy production. The three protein complexes/enzymes that are unique to carbon fixation include the Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco complex), sedoheptulose-1,7-bisphosphatase (SBPase) and phosphoribulokinase (PRK1). The Rubisco complex is a heterooligomer composed of eight subunits of the plastid-encoded large subunit (RBCL) and eight small subunits (RBCs) with four RBC isoforms found in Arabidopsis (reviewed in (37)). In the *clpp3-1* mutant, only the Rubisco subunits were significantly downregulated (two-fold lower, see Table 3.2 and Figure 3.5) among all the enzymes involved in Calvin cycle. The *clpp3-1/wt* abundance ratio was 0.68 for SBPase and this decrease was deemed significant by QSpec but not by GLEE. PRK1 and PRK2 levels were similar to wild-type. The levels of Rubisco activase, an ATP-independent enzyme that facilitates conformational changes within the RUBISCO complex during its carboxylase activity (reviewed in (37)) did not change.

Interestingly, three proteins involved in the assembly of photosynthetic complexes were significantly upregulated in the mutant including Ycf4 (*clpp3-1/wt* ~ 3.4), LHCI-5 (*clpp3-1/wt* ~ 2.7) and CRR6 (*clpp3-1/wt* ~ 3.9) (see Table 3.4). The Ycf4 orthologue in *Chlamydomonas reinhardtii* is essential for PSI complex assembly (38) and was observed to form a 1.5 MDa-complex containing newly synthesized PSI

components suggesting that it might act as a scaffold during initial PSI biogenesis (39). LHCI-5 is a low abundant protein shown to loosely associate with PSI (40) and interacts with the light harvesting complex I at the Lhca2/Lhca3 site (41). The NDH-PSI supercomplex was absent in knockout lines of LHCI-5 indicating a potential role for LHCI-5 in mediating NDH and PSI association (42). The mutants of CRR6 (*chlororespiratory reduction 6*) exhibited impaired NDH complex accumulation (43). Co-purification of CRR6 with NdhH, a subunit of the NDH-A subcomplex, indicated that CRR6 is involved in NDH subcomplex-A assembly (42).

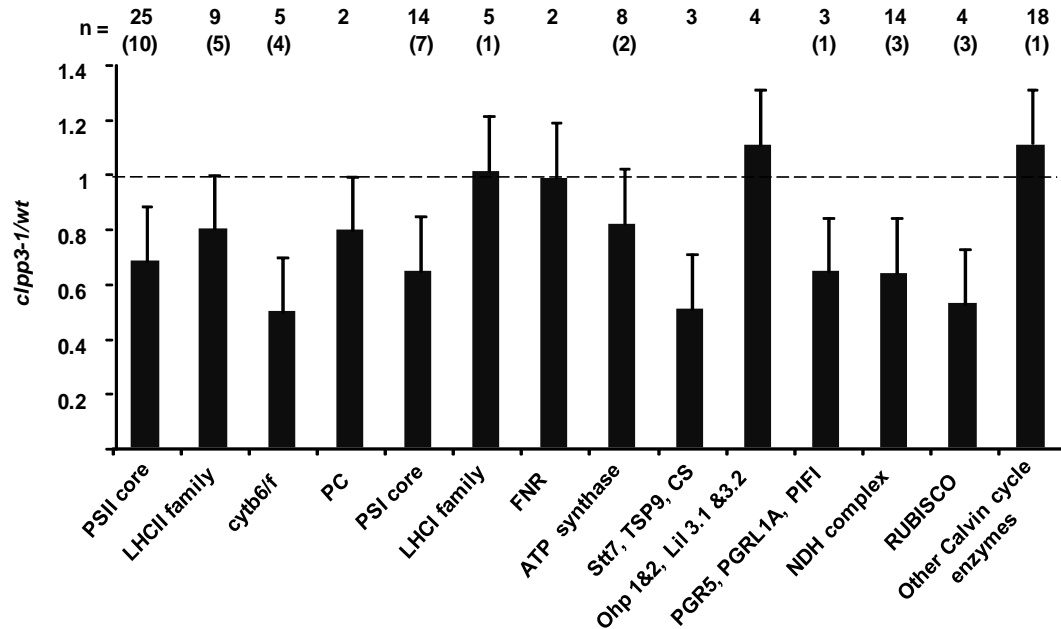


Figure 3.5. Differential accumulation of photosynthetic proteins grouped as complexes and/or by function in *clpp3-1* leaves compared with wild type leaves, determined by spectral counting, based on the *clpp3-1*/wt ratios for all quantified proteins. The number of proteins quantified for each group (n) is indicated and the protein subset that passed the significance test by GLEE and QSpec per group are enclosed in parenthesis. PS, photosystem; LHC, light-harvesting complex; PC, plastocyanin; FNR, ferredoxin reductase; Stt7, State transition thylakoid 7; CS, calcium-sensing receptor; Ohp, one-helix protein; Lil, light harvesting-like protein; PGR5, proton gradient regulation 5; PGRL1A, PGR-like protein 1A; PIFI, post-illumination chlorophyll fluorescent increase); NDH, NADH dehydrogenase; Rubisco, ribulose-1,5-bisphosphate carboxylase/oxygenase. Stt7, TSP9 and CS belong to the state transition protein group. Ohps and Lils are light stress proteins. PGR5, PGRL1A and PIFI are involved in cyclic electron flow (chlororespiration).

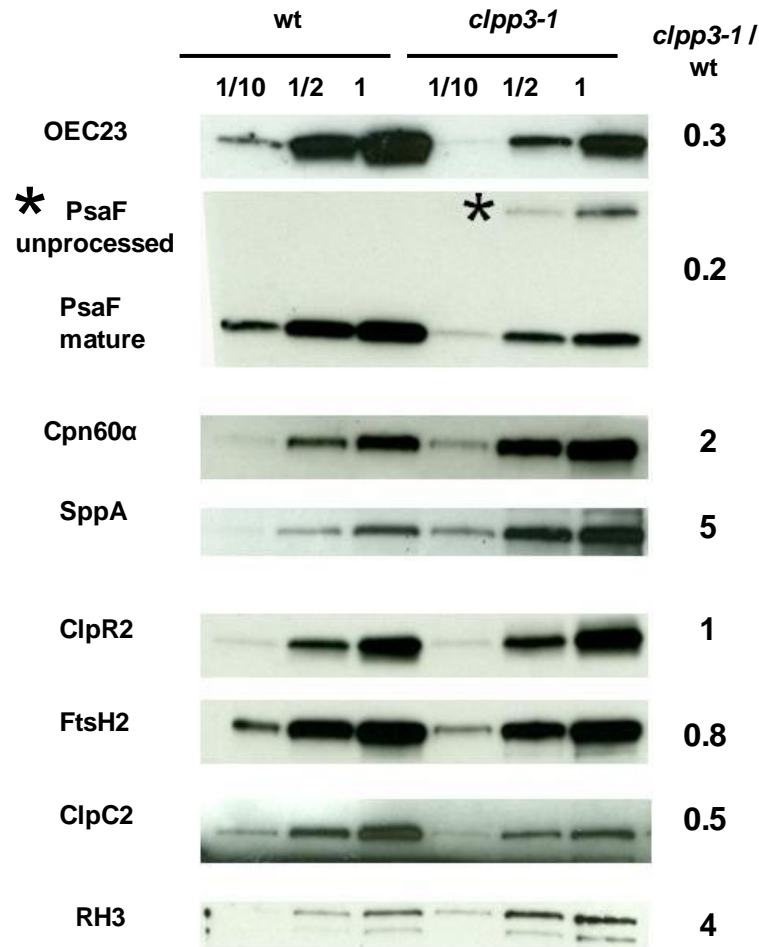


Figure 3.6. Immunoblot analysis of titrations (indicated as 1/10x, 1/2x, 1x) of total leaf protein extracts from wild type and *clpp3-1* grown as in Figure 3.2A. Membranes were probed with antibodies generated against different proteins of the photosystem I and II (PsaF: a small peripheral subunit of PSI; OEC23: oxygen evolving complex component of PSII); chaperones and proteases (ClpR2: subunit of the Clp protease complex; Clp C2; stromal chaperone; Hsp60alpha; SppA: ATP-independent light induced serine-type thylakoid protease; FtsH2: thylakoid protease of the Zn-metallo-protease family; SppA) and the stromal RNA helicase RH3. 1x = 20ug.

Plastoglobules

Plastoglobules (PGs) are thylakoid-associated lipoprotein particles that serve as compartments for synthesis, storage and degradation of quinones, tocopherols and carotenoids and are also involved in stress defense and chlorophyll and thylakoid membrane turnover (44). Microscopy analyses of chloroplasts from knockdown

(*clpr2-1*) and null (*clpr4-1*) Clp mutants (7, 8) showed an increase in PG size and quantity. Proteomic studies of the mutants also revealed a strong increase in specific PG proteins, particularly with some of the fibrillins (7, 28). As structural proteins, fibrillins maintain the PG coat and likely control PG size. Out of the seven PG-localized fibrillins, three (FIB8, FIB1a and FIB1b) were found to be three- to six-fold upregulated in *clpp3-1*. PG-localized metabolic enzymes including fructose-bisphosphate aldolase-1 (SFBA-1), allene oxide synthase (AOS), NADH dehydrogenase (NDC1), 9-cis-epoxycarotenoid dioxygenase (CCD4) and two ABC1 kinases (ABC1K2 and ABC1K3) were also upregulated. ABC1K4, another PG-localized ABC1 kinase and tocopherol cyclase were only detected in the *clpp3-1* mutant and were considered significantly changed by QSpec but not by GLEE.

Protein Import and Targeting

In this study, 19 proteins involved in protein import and targeting to various chloroplast compartments were quantified. Only one subunit of an envelope protein translocase and three members of a thylakoid (45) import pathway were found to differentially accumulate. Tic 110, an inner envelope membrane protein translocase, is the most abundant component of the Tic translocon (45-47). Knockout lines of Tic110 are embryolethal indicating its crucial role in chloroplast biogenesis (48). Protein targeting to the thylakoid lumen is mediated by two pathways, namely the cpSec pathway and cpTAT pathway (reviewed in (49)). A knockout line of cpSecA in Arabidopsis is seedling lethal but can be rescued if grown heterotrophically on sucrose at very low light intensities (50). In maize, disruption of the *cpSecY* gene is seedling lethal (51). Inactivation of the *cpSecE* gene in maize resulted in a less severe, pale-green seedling phenotype (52). In the *clpp3-1* mutant, the components of the ATP-

dependent Sec pathway (cpSecA, cpSecE and cpSecY) were upregulated (two- to five-fold) suggesting impediments in the translocation of lumen-localized proteins.

Protein Quality Control

24 proteins involved in protein folding and 34 proteases were quantified, 10 of which were found significantly upregulated (mostly by more than two-fold). These include the chloroplast chaperones CPN60 α,β (GroEL homologues), cpHSP70-1,2 (DnaK homologues) and cpHSP90 which play crucial roles in protein folding (53, 54). ClpB3, the chloroplast homologue of the bacterial ClpB protein (55) was four-fold upregulated. ClpB is involved in unfolding aggregated proteins together with the DnaK chaperone system (homologous to cpHSP70) (56, 57). The differential accumulation of these chaperone systems was also observed in other Clp mutants ((7, 8, 28), see Table 3.4). Consistent with these studies, ROC4, an abundant stromal peptidylprolyl isomerase (30, 58) with *in vitro* rotamase activity (59) and the Zn²⁺-protease AtPreP1 which were suggested to be involved in degradation of cleaved chloroplast transit peptides (60-62) were observed to significantly accumulate in the *clpp3-1* mutant.

Additional upregulated proteins (two- to three-fold) involved in protein quality control include the co-chaperonins Cpn10 and Cpn20, the peptidases SPP and LAP2 as well as the proteases EGY2 and SppA. The co-chaperonins Cpn10 and Cpn20 (GroES homologues) assist the Cpn60 complex (GroEL homologue) for protein folding (63, 64). The stromal processing peptidase (SPP) is involved in transit peptide removal of most nuclear-encoded, chloroplast-targeted proteins (65, 66). Leucyl aminopeptidase 2 (LAP2) belongs to the family of soluble aminopeptidases (67). EGY2 is a paralogue of EGY1, an ATP-independent thylakoid-bound metalloprotease that is crucial for thylakoid development and accumulation of chlorophyll-containing

proteins in chloroplast membranes (68). SppA is a light stress-induced, thylakoid-bound ATP-independent serine-type protease with unknown substrates (69).

The plastid-localized Clp protease family was also surveyed. As expected, ClpP3 was not detected in the homozygous *clpp3-1* null background. Despite the complete loss of *CLPP3*, the accumulation levels of the other Clp protease subunits as well as the Clp chaperones T1,T2 were not deemed significantly changed relative to wild-type. This was also consistent with previous proteomics studies of Clp mutants (7, 8, 28) where Clp accumulation levels were unaltered in spite of the reduction or loss of a Clp subunit. In particular, the transcript and protein levels for all Clp protease subunits (except for ClpR4) did not change in the albino *clpr4-1* null mutant (7). The downregulation (*clpp3/wt* of 0.68) of ClpC1/C2 was considered significant by QSpec, but not by GLEE.

Estimation of relative protein abundance from immunoblot analyses of several chloroplast protein chaperones and proteases (Figure 3.6) confirmed the fold-changes measured from the MS-based proteomic experiment (Table 3.2). Cpn60-alpha subunit and SppA were two-fold and five-fold upregulated, respectively. The abundance levels of ClpR2, a subunit of the Clp complex and FtsH2, a subunit of the FtsH complex were unchanged. ClpC2 was two-fold reduced in *clpp3-1* relative to wild-type.

Consistent with the overall reduction of the thylakoid proteome, two lumen-localized isomerases were observed as significantly two-fold downregulated namely CYP38 (or TL40) and the TAT isomerase. Cyclophilin 28 (CYP38) is a multidomain protein with peptidyl prolyl *cis-trans* isomerase and phosphatase inhibitor domains (70) necessary for the assembly and maintenance of PSII (71) particularly in the proper folding of the D1 protein (and CP43) and successful assembly of the OEC complex (72).

Plastid gene expression

The chloroplast retains a small genome and its own functional gene expression machineries. 50 chloroplast-encoded proteins were quantified and comprised 34% of the total chloroplast proteome mass in wild-type but decreased to 23% in the *clpp3-1* mutant. 13 of these chloroplast-encoded proteins were found significantly downregulated by two-fold (Table 3.2). This indicates a bottleneck in one or several processes involved in chloroplast gene expression from transcription, RNA processing, translation and co-translational processing and folding.

Transcription

The plastid-encoded RNA polymerase (PEP) complex is composed of four subunits namely RpoA, RpoB, RpoC1 and RpoC2 (73) and is the predominant transcription complex in mature chloroplasts (74). These subunits were detected with low adjSPCs and their accumulation levels in the *clpp3-1* and wild-type were similar.

Analysis of purified transcriptionally active chromosomes (TACs) from *Arabidopsis* chloroplasts identified the PEP complex and 18 proteins (pTACs 1 to 18) which contain RNA/DNA binding domains (75). Several pTAC proteins were found two- to four-fold upregulated in *clpp3-1* including pTAC3, pTAC17, pTAC1, pTAC11 and DNA gyrase A. In addition, a DNA helicase (At5g35970.1) which was found in DNA-enriched fractions in high molecular weight fractionation of stroma (76) and a SET domain-containing protein (At5g14260.1) predicted to be involved in transcription regulation were also significantly accumulating in the mutant.

CND41 was isolated from the chloroplast nucleoid (protein-DNA complexes) in tobacco cell culture (77) and exhibits aspartic protease activity as well as DNA-binding activity (78) consistent with its association with the nucleoid. Interestingly, CND41 was found significantly downregulated (three-fold) in *clpp3-1* (Table 3.2).

RNA processing and maturation

Proteins that were shown or were predicted to be involved in RNA processing were significantly upregulated (two- to five-fold) in *clpp3-1* including two RNA-binding proteins (P67 and an RNA Recognition motif-containing protein), two ribonucleases (PNPase and Rnase J) and two RNA helicases. Polynucleotide phosphorylase (PNPase) is an exoribonuclease that is indispensable for 3'-end maturation of 23S rRNA transcripts and the efficiency of the 3'-end processing and polyadenylation of mRNAs, as well as RNA degradation (79-82). The endonuclease RNase J has been implicated in 16S rRNA maturation in *Bacillus subtilis* and associates with both assembled 70S ribosomes and 30S particles suggesting a role in ribosome assembly (83). The *Arabidopsis* orthologue of Rnase J was found to be essential for plant embryogenesis (EMB2746) (84), <http://www.seedgenes.org>. RNA helicase 3 (RH3), which belongs to a subgroup of DEAD-box proteins containing a Gly/Arg/Ser-rich C-terminal extension (85) has not been well-studied in plants, but members of this helicase family are often involved in ribosome maturation and/or rRNA processing and stability (86, 87). Consistent with their function in RNA processing, PNPase, Rnase J and RH3 were observed in ribonucleoprotein-rich stromal fractions of around 1-3 Megadalton (76).

Protein synthesis

In this study, 21 and 30 proteins comprising the 30S and 50S ribosomal particles were quantified, respectively. The average *clpp3-1*/wt ratio of the 30S ribosome subunits was 1.2 ± 0.3 and none were detected as significantly expressed. In contrast, the average accumulation level of 50S ribosome subunits was slightly lower (0.8 ± 0.2 , excluding RPL18) with three subunits (L2, L27 and L9) showing significant two-fold reduction and one subunit (L18) increasing 4.5-fold. Examination of the RPL18

profile showed that it had zero values for the wild-type and the fold-change calculation thus involved an imputation and might have been overestimated.

In addition, most elongation factors in protein synthesis namely EF-Tu, EF-G, EF-Ts and TypA/bipA were upregulated by two to three-fold in the mutant (Table 3.2). These proteins have been shown to co-migrate with the 70S ribosomes suggesting their participation in translation (76). During elongation, EF-Tu introduces new aminoacyl tRNAs into the peptidyl transfer center in the 50S ribosome, EF-G translocates the peptidyl-tRNA after which spontaneous peptidyl transfer occurs and EF-Ts facilitates GTP hydrolysis which releases EF-Tu for another round of peptide elongation(88, 89). Note that in Arabidopsis, EF-Ts is synthesized as fused to a 30S ribosome-interacting protein PSRP-7 which might be post-translationally processed to render various fused or independent proteins (90). The bacterial TypA/bipA binds to the 70S ribosomes at the same site as EF-G (91-93). Both bacterial and plant TypA/BipA appear to be particularly important under stress conditions (94-98). Moreover, five out of the 22 amino acid synthetases quantified were found two- to three-fold upregulated (Table 3.2).

The *Arabidopsis* orthologue of a peptide release factor (RF1) and trigger factor (TF) were also upregulated. RF1 recognizes the stop codon and binds to the ribosome to terminate translation (88, 89). AtRF1 is essential for chloroplast development (99). The *E. coli* TF binds to the 70S exit tunnel and prevents misfolding and aggregation of emerging nascent proteins (100).

Isoprenoid and tetrapyrrole metabolism

Isoprenoids are central in plant development and are derived from acetyl-CoA in the cytosol through the mevalonate pathway, or from pyruvate and glyceraldehyde-3-phosphate in the plastid through the methylerythritol phosphate (MEP) pathway (101).

We quantified 16 enzymes in the plastid isoprenoid pathway and two of them namely 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR) and 4-hydroxy-3-methylbutyl diphosphate synthase (HDS, also named GcpE or CLB4) were upregulated by 1.8-fold and 3-fold, respectively. HDS is thylakoid-associated and can accept electrons directly from the photosynthetic machinery for its catalytic activity (102).

We quantified 23 proteins involved in the tetrapyrrole biosynthetic pathway with coverage of all known enzymes involved in chlorophyll synthesis. Two enzymes involved in the initial pyrrole synthesis were differentially expressed. alpha-aminolevulinic acid dehydratase (ALAD) condenses two molecules of ALA to form porphobilinogen (PBG) and four PBG molecules are polymerized by PBG deaminase to form a linear tetrapyrrole, 1-hydroxymethylbillane (HMB) (103). Interestingly, ALAD was upregulated (*clpp3/wt* ~1.7) whereas PBG deaminase was downregulated (*clpp3/wt* ~0.7). An isoform of NADPH:protochlorophyllide reductase (PORB) which catalyzes the photoreduction of protochlorophyllide to chlorophyllide (103) was downregulated (*clpp3/wt* ~0.4).

Starch metabolism

Out of the 18 quantified proteins involved in starch metabolism, six were observed as significantly upregulated by about two- to three-fold (Table 3.2). Two starch synthases and an ADP-glucose phosphorylase (ADG2) are involved in starch synthesis. The other three were starch degradation enzymes including two phosphoglucan dikinases (Sex1 and PWD1) and an alpha-amylase (AtAMY3). The dikinases are involved in control of phosphoryltic starch degradation during the night (104).

Metabolite transporters

Several metabolite transporters were upregulated in *clpp3-1*. KEA2 is a putative K⁺ efflux antiporter belonging to the monovalent cation:proton antiporter family (105). In gram-negative bacteria, K⁺/H⁺ antiporters are controlled by reduced and oxidized forms of glutathione and are used to acidify the cytoplasm against toxic electrophiles (106). In chloroplasts, envelope membrane-bound nucleoside triphosphate transporters (NTTs) are involved in replenishing stromal ATP pools in the night by sequestering cytosolic ATP from glycolytic starch breakdown ((107). Two chloroplast NTT isoforms were both upregulated six- to ten-fold in the mutant.

Clp assemblies in the *clpp3-1* mutant

To determine the consequences of the loss of ClpP3 in the assembly state of the ClpPR complex, total soluble leaf proteomes of wt and *clpp3-1* as well as the knockdown mutant *clpr2-1* were extracted under non-denaturing conditions and proteins were separated by one-dimensional native gel electrophoresis. Immunoblot analyses using anti-ClpR2 serum detected the 350 kDa and 200 kDa bands in wild-type corresponding to the Clp core complex and the ClpP1/R1-4 ring, respectively (Figure 3.7A and see discussion in Chapter 2). These migration profiles were also observed in the *clpr2-1* knockdown mutant. Moreover, the band intensities of these Clp assemblies in *clpr2-1* corresponded to a five-fold reduction as compared to wild-type consistent with reduced ClpR2 expression (20% relative to wild-type levels as observed in (8)). The 200 kDa band was also detected in the *clpp3-1* mutant suggesting an assembled ClpR2-containing ring (most likely ClpP1/R1-4). Interestingly, a band of slightly higher molecular weight (~400 kDa) instead of the 350 kDa band for the Clp core was detected in *clpp3-1*.

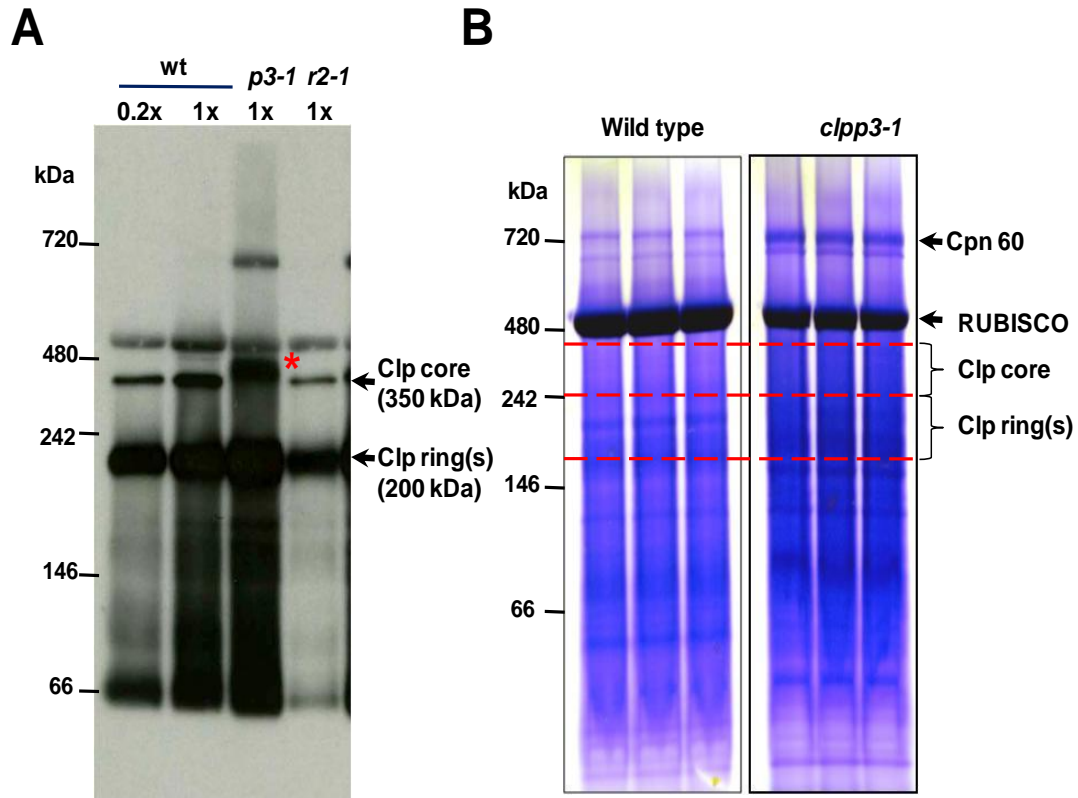


Figure 3.7. Characterization of the Clp oligomeric states in the *clpp3-1* mutant. **A)** Immunoblot analyses of the Clp assemblies in wild-type, *clpp3-1* and *clpr2-1* mutants probed by anti-ClpR2 antiserum. 1x corresponds to 50 ug of soluble total leaf. The Clp core and Clp ring(s) are indicated. The red star marks the protein band for an alternative Clp core complex with a higher molecular weight (~400 kDa) than the wild-type (350 kDa). **B)** Native gel separation of total leaf proteomes from wild-type and *clpp3-1* for MS characterization. 100ug soluble total leaf proteome was loaded per gel lane. Three replicates per genotype were separated. The corresponding molecular weight ranges that were excised, in-gel digested with trypsin and analyzed by MS are marked in red dashed lines. Additional complexes namely Cpn 60 (~700 kDa) and Rubisco (ribulose-1,5-bisphosphate carboxylase/oxygenase complex) (~550 kDa) are indicated.

To further characterize the Clp assemblies in *clpp3-1*, gel bands at 400-350 kDa and 250-150 kDa were excised, digested with trypsin and subjected to mass spectrometry analysis (Figure 3.7B and Table 3.5). The components of the ClpP3-P6 ring were detected in both Clp core and Clp ring in wild-type consistent with previous studies (see Chapter 2). ClpP3 was not detected but ClpP4, P5 and P6 were observed in *clpp3-1* in both mass ranges. The relative abundances of these subunits in *clpp3-1* were comparable to wild-type levels in the 300-400 kDa mass range but were about

three-fold up in the 250-150 kDa range indicating assembly of these subunits as individual rings, rather than as a double-ring complex. The components of the ClpP1/R ring (except for ClpR1) could only be observed in the 150-250 kDa range and accumulated at similar levels in both genotypes. Thus, a 400 kDa Clp assembly containing ClpP4-P6 and ClpP1/R3/R2/R4 of unknown stoichiometry was observed in the *clpP3-1* null mutant.

Table 3.5. Composition of Clp assemblies determined by MS analysis^a

Subunit	Clp core			Clp ring(s)		
	WT	<i>p3</i>	<i>p3/wt</i>	WT	<i>p3</i>	<i>p3/wt</i>
P3	2			2		
P4	27	29	0.9	18	62	3.0
P5	12	12	0.8	7	30	3.8
P6	4	5	1.0	3	9	2.7
SUM^b	45	46	0.9	30	101	3.0
P1	7	14	1.7	12	13	1.0
R1						
R2				4	2	0.4
R3		4		5	10	1.8
R4	2	2	0.8	3	5	1.5
SUM^b	9	20	1.9	24	30	1.3

^a Number of matched MS/MS spectra (SPC) are shown to estimate protein abundance. The analyzed gel bands are shown in Fig. 2b. The *p3/wt* ratios were derived from the protein SPCs for each genotype normalized to the total SPC per gel band.

^b Total SPC from all the subunits per Clp ring.

DISCUSSION

Optimizations in spectral counting-based comparative proteomics

The MS-based SPC methodology involves counting MS/MS spectra that matched to peptides and using this as an abundance measure of the protein comprised by the detected peptides. During the database search, the masses of fragment ions corresponding to the fragmentation of the peptide are matched to those from *in silico*-

derived peptides of target proteins with a corresponding confidence score. As such, the MS/MS spectra obtained must be of sufficient quality and this means high signal intensities of the peptide fragment ions for higher confidence in database matching. In our current workflow, we determined the consequence of increased sampling of ions selected for fragmentation by requiring an average of three microscans for each MS/MS acquisition instead of one microscan. This was equivalent to obtaining three replicates of the same peptide sequencing event that could potentially yield MS/MS spectra with higher intensities for fragment ions. On the other hand, increasing the sampling time per peptide extended the duty cycle two-fold and reduced the total MS/MS spectra acquired across the chromatographic timescale. Nevertheless, assessment of peptides matched revealed that MS/MS sampling with three microscans enabled more confident matches with peptides (higher peptide ion score) thereby increasing the number of SPCs per protein. This is particularly advantageous for detection and quantification of proteins with low or medium abundance levels. As expected, the total MS/MS spectra acquired was reduced but the total number of detected proteins was only marginally lower than in previous studies and these proteins would only have low adjSPCs (typically < 2 adjSPCs). Moreover, this difference might be due to reduced protein sample loading as the proteome analyzed in this study is about four-fold less than in previous studies. A more systematic study is needed to fully assess the impact of MS/MS acquisition frequency and time on SPC distribution across proteins of varying abundance levels; however, our study clearly shows the benefits of an increased sampling frequency for relative quantification. Overall, the experimental and bioinformatic workflow in this study is well-applicable for large-scale SPC-based differential proteomics of any biological sample with reduced sample requirements. Furthermore, the introduction of a new LTQ instrument (LTQ Velos) (36) with faster MS/MS acquisition speed than the LTQ instrument used

in the current study might enable the adoption of the MS/MS acquisition scheme described here for gaining better SPC datasets.

We also employed and evaluated two statistical methods for significance analyses of SPC datasets which model the variability in protein SPCs in different ways. GLEE estimates variation from fitting SPCs to calculated variation using third polynomial order equations whereas QSpec implements a probabilistic strategy. Overall, both statistical methods yielded comparable results except for proteins with zero values which likely represent low abundant proteins that are harder to detect and quantify. SPC data are discrete and substitution of missing values particularly for low SPCs (*e.g.*, from low abundant proteins) is not straightforward.

Comparative proteomics of the *clpp3-1* mutant

We analyzed the leaf proteome of the pale-green *clpp3-1* null mutant after heterotrophic growth and transfer on soil. We quantified 2116 proteins (including protein groups) with the majority (>60%) of the leaf proteome mass being comprised by chloroplast proteins. Furthermore, the abundance levels of 161 out of the 779 chloroplast-localized proteins were judged to be significantly affected in the mutant from QSpec and GLEE analysis at 5% FDR (Table 3.2). Functional analysis of these proteins was performed with the perspective that plastid biogenesis and protein homeostasis are functionally intertwined with primary and secondary metabolism (Figure 3.8). Consistent with its pale-green phenotype and delayed growth, the *clpp3-1* mutant exhibited reduced accumulation levels of thylakoid-bound photosynthetic complexes and the Rubisco complex. In contrast, the thylakoid-associated plastoglobules involved in stress response and membrane turnover were strongly upregulated that might indicate a thylakoid membrane homeostasis problem. We also observed upregulation of several plastid enzymes and envelope transporters involved in metabolism. The response of these metabolic pathways can in part be explained by

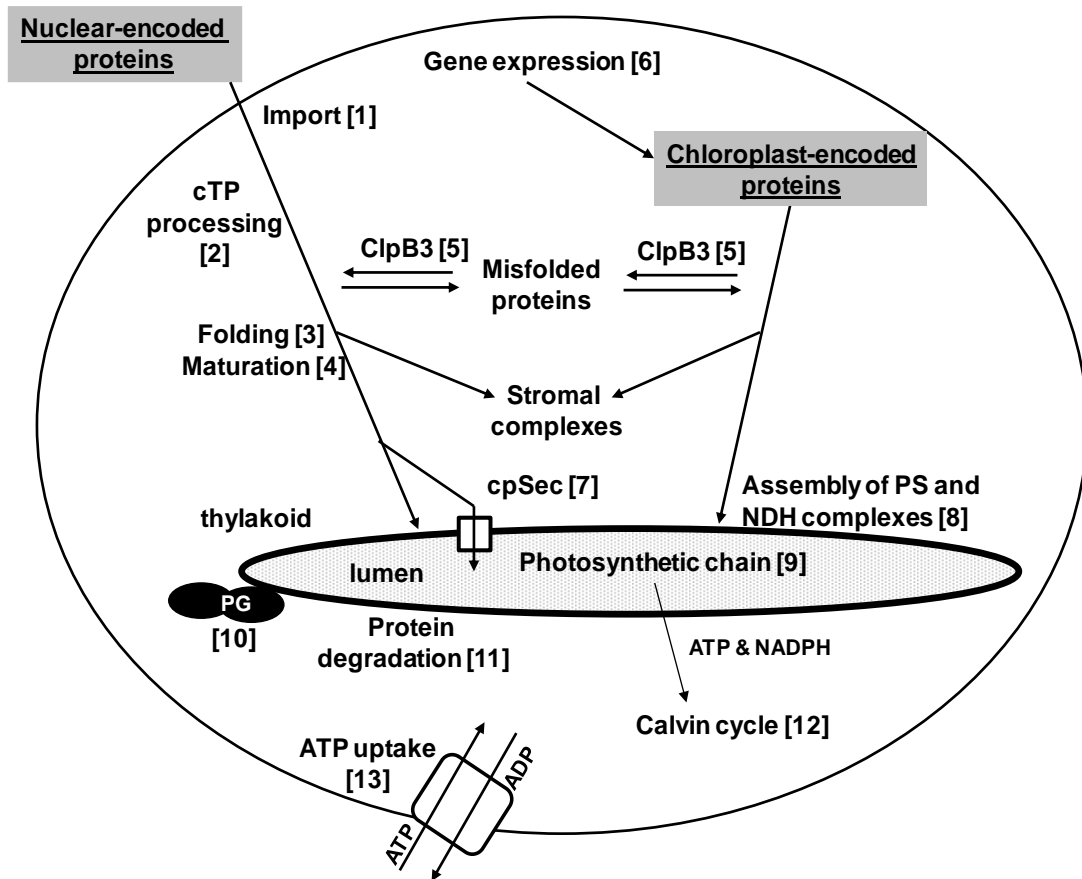


Figure 3.8. Summary of the consequences of the loss of ClpP3 in the chloroplast. Nucleus-encoded proteins are imported into the chloroplast through the Toc/Tic translocon followed by removal of the N-terminal chloroplast transit peptide (cTP) by the stromal processing peptidase (SPP) (108). The cleaved transit peptides are then processed by the protease Prep1 (60-62). Proteins destined for the thylakoid membrane system, including the thylakoid lumen, are targeted through different pathways including the ATP-dependent cpSecA pathway (49). The chloroplast-encoded proteins are synthesized on the 70S ribosomes either in the stroma or at the thylakoid surface (109). Several of the upregulated proteins/protein complexes in *clp3-1* are: Tic110 [1]; SPP and Prep1 [2]; cpHSP70, Cpn60, Cpn 20 and Cpn10 [3]; cpHSP90 [4]; ClpB3 [5]; several transcription-related proteins, elongation factors, 50S RPL subunits, and RNA processing proteins [6]; cpSecA/Y/E [7]; assembly factors for PSI and NDH complexes (Ycf4, LHCI-5, CRR6) [8]; plastoglobule components [10]; EGY2 and SppA [11]; ATP/ADP translocase (NTT1/2) [13]. Proteins that were significantly downregulated are components of the complexes involved in light reaction [9] and Rubisco involved in Calvin cycle [12]. PS, photosystem; NDH, NADH dehydrogenase. (Figure adapted from (28) with modifications).

the loss of photosynthetic capacity in Clp mutants. A clear example is the multi-fold upregulation of the inner envelope ATP/ADP translocators (NTTs) which import cytosolic ATP into the chloroplast (107) confirming the reduced ATP-generating

capacity in the chloroplast. Other effects, such as the increase of several enzymes in the MEP and tetrapyrrole pathways can be explained by a combination of developmental effects on gene expression and the reduced electron transport affecting the availability of electrons for reduction steps in the MEP pathway.

The role of Clp in protein quality control

The *clpp3-1* mutant displayed an increased need for protein import, folding and unfolding capacity. Consistent upregulation of proteins involved in import (Tic110, SecA system), folding and maturation (Cpn60/Cpn20/Cpn10/cpHSP70/cpHSP90 and ROC4) and unfolding (ClpB3) was observed (Figure 3.8). The accumulation of ClpB3 suggests protein aggregation in the chloroplast and protein folding stress. Inactivation of *CLPB3* expression in the *clpr2-1* background led to seedling lethality (28) confirming the central role of ClpB3 in dealing with perturbed proteome homeostasis in the Clp mutants. Our working hypothesis is that reduced Clp protease activity leads to the buildup of unwanted/damaged proteins that accumulate as protein aggregates. ClpB3, together with cpHSP70, are upregulated to unfold and reactivate these aggregates. The engagement of cpHsp70 in refolding activity of aggregated plastid proteins likely reduces the capacity for folding of newly imported nuclear-encoded proteins or newly synthesized chloroplast-encoded proteins and thus further contributes to destabilized proteome homeostasis. The upregulation of both Cpn60/20/10 (GroEL/GroES homologues) and cpHSP90 suggests an extra demand for ATP-dependent folding activity, possibly due to less favorable/efficient folding conditions (*e.g.*, limited ATP supply). The increase in Tic110 suggests a bottleneck in protein import into the chloroplast; we speculate that this is due to delayed release from the import channel and from Tic110 are engaged in unproductive stabilization of proteins, rather than aiding in completion of the import process.

The upregulation of SPP and AtPrep1 in the mutant further implies problems in processing of the chloroplast transit peptide (cTP) sequences (Figure 3.8). During and after import of nuclear-encoded proteins, chaperones are also needed for presentation of imported precursor proteins to SPP. However, this is not met in the Clp mutant due to entrapment of the chaperones in unproductive refolding aggregates. As such, a subfraction of the imported thylakoid-localized proteins remains unprocessed as observed with the thylakoid-bound photosystem protein PsaF (Figure 3.6). Protein targeting into the lumen is also impaired as observed in the accumulation of the ATP-dependent cpSec pathway which might be partly due to limited ATP supply as photosynthetic capacity is reduced. The accumulation of unprocessed proteins in the thylakoid and reduction of import of luminal proteins would then result to incomplete and impaired assembly of photosystem complexes as indicated by the increased accumulation of assembly factors (Ycf4, LHCI-5 and CRR6). Unassembled subunits accumulate in the thylakoid membrane thereby requiring protein turnover by thylakoid-bound proteases such as EGY2 and SppA.

The role of Clp in chloroplast gene expression

The abundance levels of chloroplast-encoded proteins were decreased in *clpp3-1* despite an increase in components of the plastid nucleoid (plastid chromosome with DNA-binding proteins) involved in DNA transcription and mRNA processing. Accumulation levels of most chloroplast ribosomal subunits remained unaltered relative to wild-type suggesting the presence of assembled ribosomes which may or may not be functional. Chloroplast protein synthesis was not blocked but was most likely impaired. In particular, elongation factors, termination factors and co-translational proteins were upregulated suggesting either delayed 70S ribosome maturation, stalled (nonfunctional but assembled) ribosomes or both. It has been shown that *CLPR1* and *CLPR2* mutants have delayed processing of the dicistronic

23S-4.5 S chloroplast rRNA (11, 28, 110). Since rRNA processing is tightly coupled to the assembly process, the delayed rRNA processing could either be the cause or the consequence of delayed ribosome maturation.

In several bacterial species such as in *E. coli*, the Clp system is involved in recycling stalled ribosomes which involves an attachment of an *ssrA* tag of several amino acid residues to the C-terminus of the incompletely synthesized proteins for targeted degradation (reviewed in (111, 112)). In *E. coli*, the ClpX chaperone and the ClpP core (ClpXP) is responsible for more than 90% of the degradation of *ssrA*-tagged proteins and ClpAP and Lon in degrading the rest (5% and 2%, respectively) (113). In photosynthetic organisms, *ssrA* sequences have been found in genomes of cyanobacteria (114, 115) and in plastid genomes from the red lineage (116, 117), but not in higher plant plastid genomes. Reactivation of stalled ribosomes and its potential coupling with proteolysis remains uncharacterized in plant plastids.

An alternative Clp core complex in clpp3-1

The 350 kDa chloroplast-localized ClpPR complex is a single asymmetric complex of ClpP1/R1/R2/R3/R4 and ClpP3/P4/P5/P6 rings with a 3:1:1:1:1 and 1:2:3:1 stoichiometry, respectively (see chapter 2). As such, each Clp core complex contains one copy of ClpP3. Interestingly, native gel separations followed by immunoblot and MS characterizations of total leaf proteome from *clpp3-1* indicated the presence of a 400 kDa Clp assembly which contains most of the Clp subunits except ClpP3 suggesting an alternative Clp core where ClpP3 was replaced by another Clp subunit. In addition, the slightly higher molecular mass of this Clp core suggests the presence of additional copies of Clp chaperones ClpT1/T2 (molecular weight of 20 kDa each).

The presence of a Clp core has been observed in a null ClpR1 mutant (*clpr1-1*) suggesting that ClpR1 is partially redundant (7). Indeed, overexpression of ClpR3 in the *clpr1-1* background can complement for the loss of ClpR1 (7). However, *clpr1-1*

can grow on soil (11), whereas *clpp3-1* is seedling lethal; this suggests a different scenario for the complete loss of ClpP3. Overexpression of *CLPP4*, *CLPP5* and *CLPP6* in the *clpp3-1* background did not rescue the phenotype (Kim and van Wijk, unpublished).

Among the Clp subunits, ClpP3 has the longest C-terminal extension (+52 amino acids) and is not proteolytically removed upon Clp core assembly (Chapter 2). We predicted that these extended C-termini can form α -helical fragments that can fold over the adaxial surfaces, affecting docking of ClpC/D chaperones and ClpT1,2 (3). As such, even if ClpP3 can be substituted by another Clp subunit in an alternative Clp complex, association (or prevention of association) of Clp chaperones is possibly impaired. The subunit that would have replaced the Clp subunit might also serve as the binding site for ClpT1/T2 interaction that might explain the extra mass (~50 kDa) observed in the 400 kDa Clp assembly in *clpp3-1*. Furthermore, the interaction between the modified ClpP4-P6 and the unmodified ClpP1/R rings appears to be less stable as indicated by the three-fold accumulation of ClpP4-P6 subunits in the 200 kDa mass range (Figure 3.7 and Table 3.5). Moreover, the functional state of this alternative Clp core complex is unknown. The asymmetry of the Clp core complex indicates differential Clp chaperone binding and consequently, differential substrate pools. The seedling lethality of *clpp3-1* implies that the alternative Clp core complex is not fully functional particularly during embryogenesis. The determination of the composition and stoichiometry of the alternative Clp core complex in *clpp3-1* will be important in understanding the compensatory response in the loss of ClpP3 in Clp core complex assembly and function.

MATERIALS AND METHODS

Plant growth and total leaf proteome extraction for MS analysis Wild type (Col-0) plants were grown on soil for 30 days under a short day cycle (10h/14h of light/dark) at 100 $\mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$. Homozygous *clpp3-1* plants were first grown on agar plates with 0.5x Murashige-Skoog medium and 2% sucrose under short day conditions at 40 $\mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$, and then transferred to soil and grown under a short day cycle at 100 $\mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ for 70 days. Total leaf proteins were extracted by grinding 180 mg fresh leaves in liquid N₂ into a fine powder. 1 ml extraction buffer (1% SDS, 125 mM Tris-HCl (pH 8.8), 5 mM EDTA, 5 mM tributylphosphine, 2.5 mg/ml of protease inhibitor Pefablok) was added and a pestle was used to solubilize the material. Unsolubilized materials were removed by centrifugation and proteins in the resulting supernatant were precipitated in 75% acetone at -80°C. Proteins were collected as pellets by centrifugation, followed by two additional acetone washes to remove lipids. The resulting protein pellet was solubilized in 1% SDS, 50 mM Tris-HCl (pH 8.25) and protein concentrations were determined using the BCA Protein Assay Kit (ThermoFisher). 50 μg total leaf protein of *clpr4-1* and wt samples were each run out on a Biorad Criterion Tris-HCl precast gels (10.5-14% acrylamide gradient). Each of the gel lanes were cut into 20 bands followed by reduction, alkylation, and in-gel digestion with trypsin as described in (118).

nanoLC-LTQ Orbitrap analysis and data processing The resuspended peptide extracts were analyzed by data-dependent MS/MS using an on-line LC-LTQ-Orbitrap (Thermo Electron Corp.). Peptide samples were automatically loaded on a guard column (LC Packings MGU-30-C18PM) via an autosampler followed by separation on a PepMap C₁₈ reverse-phase nanocolumn (LC Packings nan75-15-03-C18PM) using 90-min gradients with 95% water, 5% ACN, 0.1% FA (solvent A) and 95% ACN, 5%

water, 0.1% FA (solvent B) at a flow rate of 200 nl/min. Two blanks were run after every sample (see Zybaylov et. al., 2008 for the gradient and sample injection scheme). The acquisition cycle consisted of a survey MS scan in the Orbitrap with a set mass range from 350 to 1800 m/z at the highest resolving power (100,000) followed by five data-dependent MS/MS scans acquired in the LTQ. Dynamic exclusion was used with the following parameters: exclusion size, 500; repeat count, 2; repeat duration, 30 s; exclusion time, 180 s; exclusion window, ± 6 ppm or ± 100 ppm. Target values were set at 5×10^5 and 10^4 for the survey and tandem MS scans, respectively. MS survey scan in the Orbitrap was acquired in one microscan. Fragment ion spectra were acquired in the LTQ as an average of three microscans. Mass spectrometry data processing, data searching against ATH v8 using Mascot and subsequent filtering and quantification based on normalized and adjusted spectral counts was carried out as in (35) and as outlined in Figure 3.2C. Mass spectrometry-derived information, as well as various types of annotated properties of all identified proteins can be found in the PPDB (<http://ppdb.tc.cornell.edu/>). The MapMan Bin system (33) was used for functional assignment. The software MASIC(119) (<http://www.pnl.gov/>) was used to extract MS and MS/MS relevant statistics such as duty cycle from Thermo raw files.

Significance analysis The GLEE software was developed in MATLAB version 7 (MathWorks Inc.) and a standalone executable version of the software code using the MATLAB Compiler was created (Poliakov, Ponnala, Olinares, Asakura and van Wijk, unpublished). GLEE was run in a Windows platform with a cubic polynomial equation fitting and 1,000 iterations for estimation of variation.

QSpec analysis was performed in LINUX platform using the software provided from (23). 5,000 Markov chain Monte Carlo simulations were performed with 20,000 iterations to ensure convergence of the algorithm. No normalization by protein length or peptide length was included.

Isolation of leaf proteome in denaturing or non-denaturing conditions for immunoblot analysis Total leaf material was ground in liquid nitrogen and solubilized in 50 mM HEPES-KOH, pH 8.0, 15% glycerol and 10 mM MgCl₂ (extraction buffer, EB) with protease inhibitor cocktail. The suspension was then filtered in miracloth and spun at 100,000xg. For total leaf extractions (with SDS), leaf material was ground in liquid nitrogen, resuspended in 50 mM Tris-HCl pH 6.8, 2% SDS and protease inhibitor cocktail and filtered through a frit column (ThermoFisher Scientific) by a quick centrifugation. The filtrate was collected for protein analysis. Protein concentrations were determined using the BCA Protein Assay Kit (ThermoScientific).

Light blue native PAGE was performed for separation of total leaf extracts under non-denaturing conditions using the NativePAGE Novex gel system (Invitrogen) with pre-cast 4-16% acrylamide Bis-Tris gels (Invitrogen). For denaturing protein separations, precast 10.5-14% gradient acrylamide Laemmli gels (Biorad) were used. For immunoblots, proteins were blotted onto PVDF membranes and probed with antibodies using chemiluminescence for detection, following standard procedures. The anti-sera used were generous gifts from various colleagues: anti-RH3 and anti-OEC23 (from Dr. Alice Barkan), anti-PsaF (from Dr. Hendrik Scheller), anti-Cpn60alpha (from Dr. Masato Nakai), anti-ClpC2 (from Dr. Steve Rodermel), anti-SppA (Dr. Anna Sokolenko) and anti-FtsH5 (Dr. Wataru Sakamoto).

ACKNOWLEDGEMENTS

This work was supported by grants from the National Science Foundation (MCB-1021963 and MCB-0718897) to KJVW. Part of this work was carried out by using the resources of the Computational Biology Service Unit of Cornell University which is partially funded by the Microsoft Corporation. We thank Jitae Kim for the images of

the *clpp3-1* plants, Anton Poliakov and Lalit Ponalla for the use of the GLEE software and computational support and Yukari Asakura for the anti-ClpR2 antibody. We thank Dr Qi Sun for his valuable support. We also acknowledge Drs. Alice Barkan, Hendrik Scheller, Masato Nakai, Steve Rodermel, Anna Sokolenko and Wataru Sakamoto for providing antisera for our immunoblot analyses.

REFERENCES

1. Sakamoto, W. (2006) *Annu Rev Plant Biol.* **57**, 599-621.
2. Kato, Y. & Sakamoto, W. (2010) *Int Rev Cell Mol Biol* **280**, 185-218.
3. Peltier, J. B., Ripoll, D. R., Friso, G., Rudella, A., Cai, Y., Ytterberg, J., Giacomelli, L., Pillardy, J. & Van Wijk, K. J. (2004) *J Biol Chem* **279**, 4768-81.
4. Adam, Z., Rudella, A. & van Wijk, K. J. (2006) *Curr Opin Plant Biol* **9**, 234-40.
5. Shikanai, T., Shimizu, K., Ueda, K., Nishimura, Y., Kuroiwa, T. & Hashimoto, T. (2001) *Plant Cell Physiol* **42**, 264-73.
6. Kuroda, H. & Maliga, P. (2003) *Nature* **425**, 86-9.
7. Kim, J., Rudella, A., Ramirez Rodriguez, V., Zybailov, B., Olinares, P. D. & van Wijk, K. J. (2009) *Plant Cell* **21**, 1669-92.
8. Rudella, A., Friso, G., Alonso, J. M., Ecker, J. R. & van Wijk, K. J. (2006) *Plant Cell* **18**, 1704-21.
9. Zheng, B., MacDonald, T. M., Sutinen, S., Hurry, V. & Clarke, A. K. (2006) *Planta* **224**, 1103-15.
10. Sjogren, L. L., Stanne, T. M., Zheng, B., Sutinen, S. & Clarke, A. K. (2006) *Plant Cell* **18**, 2635-49.
11. Koussevitzky, S., Stanne, T. M., Peto, C. A., Giap, T., Sjogren, L. L., Zhao, Y., Clarke, A. K. & Chory, J. (2007) *Plant Mol Biol* **63**, 85-96.
12. Kim, J., Olinares, P.B.D, van Wijk, K.J. (2010) *unpublished*.
13. Domon, B. & Aebersold, R. (2010) *Nat Biotechnol* **28**, 710-21.
14. Liu, H., Sadygov, R. G. & Yates, J. R., 3rd (2004) *Anal Chem* **76**, 4193-201.

15. Old, W. M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K. G., Mendoza, A., Sevinsky, J. R., Resing, K. A. & Ahn, N. G. (2005) *Mol Cell Proteomics* **4**, 1487-502.
16. Zybaylov, B., Coleman, M. K., Florens, L. & Washburn, M. P. (2005) *Anal Chem* **77**, 6218-24.
17. Zhang, Y., Wen, Z., Washburn, M. P. & Florens, L. (2009) *Anal Chem*.
18. Cooper, B., Feng, J. & Garrett, W. M. (2010) *J Am Soc Mass Spectrom* **21**, 1534-46.
19. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. (2007) *Anal Bioanal Chem* **389**, 1017-31.
20. Pavelka, N., Fournier, M. L., Swanson, S. K., Pelizzola, M., Ricciardi-Castagnoli, P., Florens, L. & Washburn, M. P. (2007) *Mol Cell Proteomics*.
21. Zybaylov, B., Mosley, A. L., Sardi, M. E., Coleman, M. K., Florens, L. & Washburn, M. P. (2006) *J Proteome Res* **5**, 2339-47.
22. Lu, P., Vogel, C., Wang, R., Yao, X. & Marcotte, E. M. (2007) *Nat Biotechnol* **25**, 117-24.
23. Choi, H., Fermin, D. & Nesvizhskii, A. I. (2008) *Mol Cell Proteomics* **7**, 2373-85.
24. Carvalho, P. C., Hewel, J., Barbosa, V. C. & Yates, J. R., 3rd (2008) *Genet Mol Res* **7**, 342-56.
25. Heinecke, N. L., Pratt, B. S., Vaisar, T. & Becker, L. (2010) *Bioinformatics* **26**, 1574-5.
26. Pham, T. V., Piersma, S. R., Warmoes, M. & Jimenez, C. R. (2010) *Bioinformatics* **26**, 363-9.
27. Li, M., Gray, W., Zhang, H., Chung, C. H., Billheimer, D., Yarbrough, W. G., Liebler, D. C., Shyr, Y. & Slebos, R. J. (2010) *J Proteome Res* **9**, 4295-305.

28. Zybaylov, B., Friso, G., Kim, J., Rudella, A., Rodriguez, V. R., Asakura, Y., Sun, Q. & van Wijk, K. J. (2009) *Mol Cell Proteomics* **8**, 1789-1810.
29. Rutschow, H., Ytterberg, A. J., Friso, G., Nilsson, R. & van Wijk, K. J. (2008) *Plant Physiol* **148**, 156-75.
30. Zybaylov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O., Sun, Q. & van Wijk, K. J. (2008) *PLoS ONE* **3**, e1994.
31. Friso, G., Majeran, W., Huang, M., Sun, Q. & van Wijk, K. J. (2010) *Plant Physiol* **152**, 1219-50.
32. Sun, Q., Zybaylov, B., Majeran, W., Friso, G., Olinares, P. D. & van Wijk, K. J. (2009) *Nucleic Acids Res* **37**, D969-74.
33. Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., Selbig, J., Muller, L. A., Rhee, S. Y. & Stitt, M. (2004) *Plant J* **37**, 914-39.
34. Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. (2000) *J Mol Biol* **300**, 1005-16.
35. Zybaylov, B., Sun, Q. & van Wijk, K. J. (2009) *Anal Chem* **81**, 8015-24.
36. Olsen, J. V., Schwartz, J. C., Griep-Raming, J., Nielsen, M. L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., Wouters, E. R., Senko, M., Makarov, A., Mann, M. & Horning, S. (2009) *Mol Cell Proteomics* **8**, 2759-69.
37. Spreitzer, R. J. & Salvucci, M. E. (2002) *Annu Rev Plant Biol* **53**, 449-75.
38. Boudreau, E., Takahashi, Y., Lemieux, C., Turmel, M. & Rochaix, J. D. (1997) *Embo J* **16**, 6095-104.
39. Ozawa, S., Nield, J., Terao, A., Stauber, E. J., Hippler, M., Koike, H., Rochaix, J. D. & Takahashi, Y. (2009) *Plant Cell* **21**, 2424-42.
40. Ganeteg, U., Klimmek, F. & Jansson, S. (2004) *Plant Mol Biol* **54**, 641-51.

41. Lucinski, R., Schmid, V. H., Jansson, S. & Klimmek, F. (2006) *FEBS Lett* **580**, 6485-8.
42. Peng, L., Cai, W. & Shikanai, T. (2010) *Plant J.*
43. Munshi, M. K., Kobayashi, Y. & Shikanai, T. (2006) *Plant Physiol* **141**, 737-44.
44. Brehelin, C., Kessler, F. & van Wijk, K. J. (2007) *Trends Plant Sci* **12**, 260-6.
45. Block, M. A., Dorne, A. J., Joyard, J. & Douce, R. (1983) *J Biol Chem* **258**, 13281-6.
46. Schnell, D. J., Kessler, F. & Blobel, G. (1994) *Science* **266**, 1007-12.
47. Lubeck, J., Soll, J., Akita, M., Nielsen, E. & Keegstra, K. (1996) *Embo J* **15**, 4230-8.
48. Inaba, T., Alvarez-Huerta, M., Li, M., Bauer, J., Ewers, C., Kessler, F. & Schnell, D. J. (2005) *Plant Cell* **17**, 1482-96.
49. Jarvis, P. & Robinson, C. (2004) *Curr Biol* **14**, R1064-77.
50. Liu, D., Gong, Q., Ma, Y., Li, P., Li, J., Yang, S., Yuan, L., Yu, Y., Pan, D., Xu, F. & Wang, N. N. (2010) *J Exp Bot* **61**, 1655-69.
51. Roy, L. M. & Barkan, A. (1998) *J Cell Biol* **141**, 385-95.
52. Williams-Carrier, R., Stiffler, N., Belcher, S., Kroeger, T., Stern, D. B., Monde, R. A., Coalter, R. & Barkan, A. (2010) *Plant J* **63**, 167-77.
53. Boston, R. S., Viitanen, P. V. & Vierling, E. (1996) *Plant Mol Biol* **32**, 191-222.
54. Wandinger, S. K., Richter, K. & Buchner, J. (2008) *J Biol Chem* **283**, 18473-7.
55. Myouga, F., Motohashi, R., Kuromori, T., Nagata, N. & Shinozaki, K. (2006) *Plant J* **48**, 249-60.
56. Goloubinoff, P., Mogk, A., Zvi, A. P., Tomoyasu, T. & Bukau, B. (1999) *Proc Natl Acad Sci U S A* **96**, 13732-7.

57. Haslberger, T., Zdanowicz, A., Brand, I., Kirstein, J., Turgay, K., Mogk, A. & Bukau, B. (2008) *Nat Struct Mol Biol* **15**, 641-50.
58. Peltier, J. B., Cai, Y., Sun, Q., Zabrouskov, V., Giacomelli, L., Rudella, A., Ytterberg, A. J., Rutschow, H. & van Wijk, K. J. (2006) *Mol Cell Proteomics* **5**, 114-133.
59. Lippuner, V., Chou, I. T., Scott, S. V., Ettinger, W. F., Theg, S. M. & Gasser, C. S. (1994) *J Biol Chem* **269**, 7863-8.
60. Bhushan, S., Stahl, A., Nilsson, S., Lefebvre, B., Seki, M., Roth, C., McWilliam, D., Wright, S. J., Liberles, D. A., Shinozaki, K., Bruce, B. D., Boutry, M. & Glaser, E. (2005) *Plant Cell Physiol* **46**, 985-96.
61. Stahl, A., Nilsson, S., Lundberg, P., Bhushan, S., Biverstahl, H., Moberg, P., Morisset, M., Vener, A., Maler, L., Langel, U. & Glaser, E. (2005) *J Mol Biol* **349**, 847-60.
62. Glaser, E., Nilsson, S. & Bhushan, S. (2006) *Biol Chem* **387**, 1441-7.
63. Bertsch, U., Soll, J., Seetharam, R. & Viitanen, P. V. (1992) *Proc Natl Acad Sci U S A* **89**, 8696-700.
64. Sharkia, R., Bonshtien, A. L., Mizrahi, I., Weiss, C., Niv, A., Lustig, A., Viitanen, P. V. & Azem, A. (2003) *Biochim Biophys Acta* **1651**, 76-84.
65. Richter, S. & Lamppa, G. K. (1998) *Proc Natl Acad Sci U S A* **95**, 7463-8.
66. Richter, S. & Lamppa, G. K. (1999) *J Cell Biol* **147**, 33-44.
67. Walling, L. L. (2006) *Curr Opin Plant Biol* **9**, 227-33.
68. Chen, G., Bi, Y. R. & Li, N. (2005) *Plant J* **41**, 364-75.
69. Lensch, M., Herrmann, R. G. & Sokolenko, A. (2001) *J Biol Chem* **276**, 33645-51.
70. Fulgosi, H., Vener, A. V., Altschmied, L., Herrmann, R. G. & Andersson, B. (1998) *Embo J* **17**, 1577-87.

71. Fu, A., He, Z., Cho, H. S., Lima, A., Buchanan, B. B. & Luan, S. (2007) *Proc Natl Acad Sci U S A* **104**, 15947-52.
72. Sirpio, S., Khrouchtchova, A., Allahverdiyeva, Y., Hansson, M., Fristedt, R., Vener, A. V., Scheller, H. V., Jensen, P. E., Haldrup, A. & Aro, E. M. (2008) *Plant J* **55**, 639-51.
73. Burnell, J. N. & Hatch, M. D. (1988) *Plant Physiol* **86**, 1252-1256.
74. Liere, K. & Börner, T. (2007) in *Cell and Molecular Biology of Plastids*, pp. 121-174.
75. Pfalz, J., Liere, K., Kandlbinder, A., Dietz, K. J. & Oelmüller, R. (2006) *Plant Cell* **18**, 176-97.
76. Olinares, P. D., Ponnola, L. & van Wijk, K. J. (2010) *Mol Cell Proteomics* **9.7**, 1594-1615.
77. Nakano, T., Murakami, S., Shoji, T., Yoshida, S., Yamada, Y. & Sato, F. (1997) *Plant Cell* **9**, 1673-82.
78. Murakami, S., Kondo, Y., Nakano, T. & Sato, F. (2000) *FEBS Lett* **468**, 15-8.
79. Walter, M., Kilian, J. & Kudla, J. (2002) *Embo J* **21**, 6905-14.
80. Yehudai-Resheff, S., Hirsh, M. & Schuster, G. (2001) *Mol Cell Biol* **21**, 5408-16.
81. Yehudai-Resheff, S., Portnoy, V., Yogeve, S., Adir, N. & Schuster, G. (2003) *Plant Cell* **15**, 2003-19.
82. Slomovic, S., Portnoy, V., Yehudai-Resheff, S., Bronshtein, E. & Schuster, G. (2008) *Biochim Biophys Acta* **1779**, 247-55.
83. Britton, R. A., Wen, T. Y., Schaefer, L., Pellegrini, O., Uicker, W. C., Mathy, N., Tobin, C., Daou, R., Szyk, J. & Condon, C. (2007) *Molecular Microbiology* **63**, 127-138.
84. Meinke, D., Sweeney, C. & Muralla, R. (2009) *PLoS One* **4**, e7386.

85. Aubourg, S., Kreis, M. & Lecharny, A. (1999) *Nucleic Acids Research* **27**, 628-636.
86. Linder, P. (2006) *Nucleic Acids Res* **34**, 4168-80.
87. Iost, I. & Dreyfus, M. (2006) *Nucleic Acids Res* **34**, 4189-97.
88. Ramakrishnan, V. (2002) *Cell* **108**, 557-572.
89. Liljas, A. (2004) *Structural aspects of protein synthesis* (World Scientific, Singapore; Hackensack, N.J.).
90. Beligni, M. V., Yamaguchi, K. & Mayfield, S. P. (2004) *Plant Cell* **16**, 3357-69.
91. Caldon, C. E., Yoong, P. & March, P. E. (2001) *Molecular Microbiology* **41**, 289-297.
92. Caldon, C. E. & March, P. E. (2003) *Current Opinion in Microbiology* **6**, 135-139.
93. Owens, R. M., Pritchard, G., Skipp, P., Hodey, M., Connell, S. R., Nierhaus, K. H. & O'Connor, C. D. (2004) *Embo J* **23**, 3375-85.
94. Pfennig, P. L. & Flower, A. M. (2001) *Molecular Genetics and Genomics* **266**, 313-317.
95. Beckering, C. L., Steil, L., Weber, M. H. W., Volker, U. & Marahiel, M. A. (2002) *J. Bacteriol.* **184**, 6395-6402.
96. Eymann, C., Homuth, G., Scharf, C. & Hecker, M. (2002) *J. Bacteriol.* **184**, 2500-2520.
97. Krishnan, K. & Flower, A. M. (2008) *Journal of Bacteriology* **190**, 7675-7683.
98. Wang, F., Zhong, N. Q., Gao, P., Wang, G. L., Wang, H. Y. & Xia, G. X. (2008) *Plant Cell Environ* **31**, 982-94.

99. Motohashi, R., Yamazaki, T., Myouga, F., Ito, T., Ito, K., Satou, M., Kobayashi, M., Nagata, N., Yoshida, S., Nagashima, A., Tanaka, K., Takahashi, S. & Shinozaki, K. (2007) *Plant Mol Biol* **64**, 481-97.
100. Ferbitz, L., Maier, T., Patzelt, H., Bukau, B., Deuerling, E. & Ban, N. (2004) *Nature* **431**, 590-6.
101. Phillips, M. A., D'Auria, J. C., Gershenzon, J. & Pichersky, E. (2008) *Plant Cell* **20**, 677-96.
102. Seemann, M., Tse Sum Bui, B., Wolff, M., Miginiac-Maslow, M. & Rohmer, M. (2006) *FEBS Lett* **580**, 1547-52.
103. Moulin, M. & Smith, A. G. (2005) *Biochem Soc Trans* **33**, 737-42.
104. Zeeman, S. C., Smith, S. M. & Smith, A. M. (2007) *Biochem J* **401**, 13-28.
105. Maser, P., Thomine, S., Schroeder, J. I., Ward, J. M., Hirschi, K., Sze, H., Talke, I. N., Amtmann, A., Maathuis, F. J., Sanders, D., Harper, J. F., Tchieu, J., Gribskov, M., Persans, M. W., Salt, D. E., Kim, S. A. & Guerinot, M. L. (2001) *Plant Physiol* **126**, 1646-67.
106. Munro, A. W., Ritchie, G. Y., Lamb, A. J., Douglas, R. M. & Booth, I. R. (1991) *Molecular Microbiology* **5**, 607-616.
107. Reinhold, T., Alawady, A., Grimm, B., Beran, K. C., Jahns, P., Conrath, U., Bauer, J., Reiser, J., Melzer, M., Jeblick, W. & Neuhaus, H. E. (2007) *Plant J* **50**, 293-304.
108. Jarvis, P. (2008) *New Phytol.*
109. Schunemann, D. (2007) *Biol Chem* **388**, 907-15.
110. Yu, F., Liu, X., Alsheikh, M., Park, S. & Rodermel, S. (2008) *Plant Cell* **20**, 1786-804.
111. Dulebohn, D., Choy, J., Sundermeier, T., Okan, N. & Karzai, A. W. (2007) *Biochemistry* **46**, 4681-93.

112. Karzai, A. W., Roche, E. D. & Sauer, R. T. (2000) *Nat Struct Biol* **7**, 449-55.
113. Lies, M. & Maurizi, M. R. (2008) *J Biol Chem* **283**, 22918-29.
114. Keiler, K. C., Shapiro, L. & Williams, K. P. (2000) *Proc Natl Acad Sci U S A* **97**, 7778-83.
115. Gaudin, C., Zhou, X., Williams, K. P. & Felden, B. (2002) *Nucleic Acids Res* **30**, 2018-24.
116. Williams, K. P. & Bartel, D. P. (1998) *Nucleic Acids Res* **26**, 163-5.
117. Gueneau de Novoa, P. & Williams, K. P. (2004) *Nucleic Acids Res* **32**, D104-8.
118. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. & Mann, M. (2006) *Nat Protoc* **1**, 2856-60.
119. Monroe, M. E., Shaw, J. L., Daly, D. S., Adkins, J. N. & Smith, R. D. (2008) *Comput Biol Chem* **32**, 215-7.

CHAPTER FOUR

Megadalton complexes in the chloroplast stroma of *Arabidopsis thaliana* characterized by size exclusion chromatography, mass spectrometry and hierarchical clustering*

INTRODUCTION

Chloroplasts are essential plant organelles of prokaryotic origin that perform a variety of metabolic and signaling functions. Best known for their role in photosynthesis, they also carry out the biosynthesis of many primary and secondary metabolites like lipids, amino acids, vitamins, nucleotides, tetrapyrroles, and hormones (1). Subcellular localization prediction by TargetP (2), combined with a correction for false positive and false negative rates, suggested that all non-green plastid types and chloroplasts together contain some 3500 proteins in *Arabidopsis thaliana* (3). More than 95% of the chloroplast proteins are nuclear-encoded and post-translationally imported into the chloroplast (4-6). Over the last decade, several studies were published that aimed to identify (subfractions of) the *Arabidopsis* chloroplast proteome, *e.g.* (7-10). The precise number of *bona fide* chloroplast proteins from these proteomics studies is probably somewhere around 1000-1300; comparing this number with the predicted chloroplast proteome indicates that ~50% of the proteome has still not been observed. Recently, we concluded that when compared to the predicted *Arabidopsis* chloroplast proteome, the chloroplast proteome identified to date is particularly underrepresented (40-70%) for proteins involved in signaling, stress, development, unassigned function

* This work was originally published in *Molecular and Cellular Proteomics*. **Olinares, P.D., Ponnala, L. and van Wijk, K.J.** (2010). Megadalton Complexes in the Chloroplast Stroma of *Arabidopsis thaliana* characterized by Size Exclusion Chromatography, Mass Spectrometry and Hierarchical Clustering. *Molecular and Cellular Proteomics* 9: 1594-1615 © Copyright American Society for Biochemistry and Molecular Biology. P.D.O performed all the experiments. L.P. wrote the script for the hierarchical clustering using Matlab. P.D.O. and K.J.V.W. wrote the manuscript.

and DNA/RNA metabolism (9). To probe deeper into the chloroplast proteome, enrichment for low-abundant proteins prior to MS analysis is required.

Many biochemical functions are executed by protein assemblies. Several studies have catalogued the assembly states of chloroplast proteins in plants. Separation of the oligomeric *Arabidopsis* stromal proteome by two-dimensional native gel electrophoresis (CN-PAGE) profiled 240 non-redundant proteins and captured information for 124 complexes (11). However, native gel electrophoresis has a practical size limit and only protein complexes below ~1000 kDa can be effectively separated, thereby missing megadalton-sized complexes. Several megadalton-sized complexes in plants have been characterized by targeted purification schemes, including the spinach 30S and 50S ribosomal particles (12-14), cytosolic ribosomes (15, 16), the tobacco plastid-encoded RNA polymerase (PEP) complex (17), maize mitochondrial pyruvate dehydrogenase complex (PDC) (18) and pea chloroplast acetyl-CoA carboxylase (ACCase) complex (19). Proteome characterization of a membrane-depleted, Triton-insoluble and high density pellet from pea plastids was highly enriched for the chloroplast PDC as well as proteins involved in plastid gene expression and carbon fixation (20). However, because no subsequent fractionation was performed, specific protein associations could not be resolved.

To extend chloroplast proteome coverage and characterize MDa-sized macromolecular assemblies to complement the previous CN-PAGE analysis of complexes up to 0.8 MDa, we fractionated the soluble chloroplast stroma by size exclusion chromatography (SEC) with a particular focus on complexes greater than 0.8 MDa. Proteins were identified by mass spectrometry analysis using an LTQ Orbitrap, a high accuracy and high sensitivity hybrid instrument (21, 22). SEC migration profiles for identified proteins were generated from matched spectral counts. Hierarchical clustering and protein heat maps of the SEC migration profiles revealed

that the identified protein complexes include 30S, 50S and 70S ribosomal particles, PDC, PEP and ACCase, indicating successful MDa-size fractionation. In addition, many ‘new’ proteins were detected and they were enriched for functions in plastid gene expression, in particular putative ribosomal biogenesis factors. Finally, protein annotations and identification data are available via the Plant Proteomics Database (PPDB at <http://ppdb.tc.cornell.edu/>), and mass spectrometry data with their metadata, are deposited in PRIDE (<http://www.ebi.ac.uk/pride/>).

The concept of using chromatography (or other continuous fractionation techniques) of protein complexes (or other types of cellular protein fractions) with mass spectrometry-based quantification to determine co-localization has been applied using stable isotope labeling (23, 24) or label-free techniques (25, 26). When combined with cluster analysis (this study and (24)), principle component analysis (23), or correlation of normalized elution profiles (this study and (25, 26)), this strategy is clearly a powerful tool and is widely applicable to other subcellular proteomes.

RESULTS

Isolation, size exclusion fractionation and MS analysis of the Arabidopsis stromal proteome

This study aims to extend the chloroplast proteome coverage and examine MDa-sized assemblies. A summary of the complete workflow of the experimental and computational analysis is shown in Figure 4.1. Intact chloroplasts were isolated from mature *Arabidopsis* leaf rosettes and were lysed under non-denaturing conditions. The lysate was subjected to ultracentrifugation to remove chloroplast membrane-bound proteomes (pellet) from the soluble fraction (stroma). Size separation of the non-denatured stromal proteome was then performed by size exclusion chromatography

(SEC), using a column that resolved up to ~5 MDa complexes, followed by SDS-PAGE.

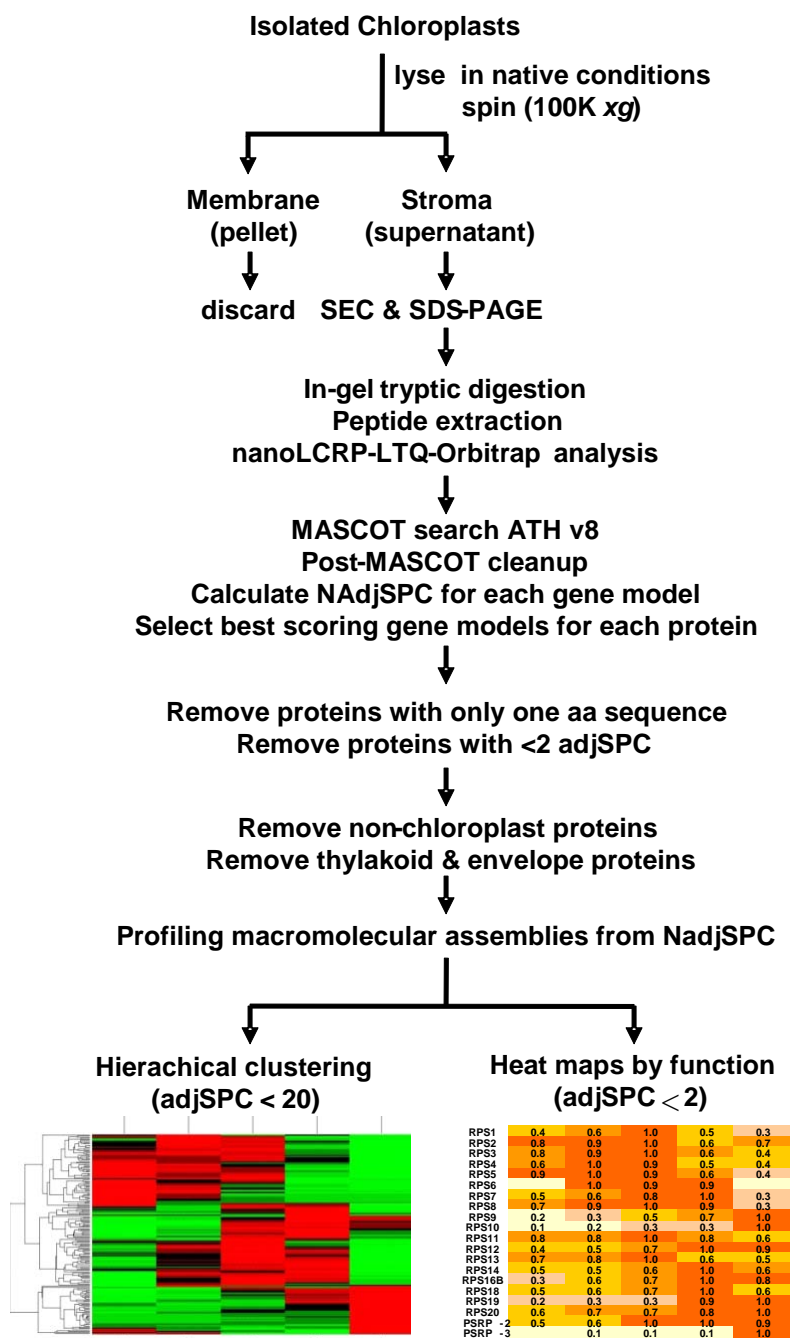
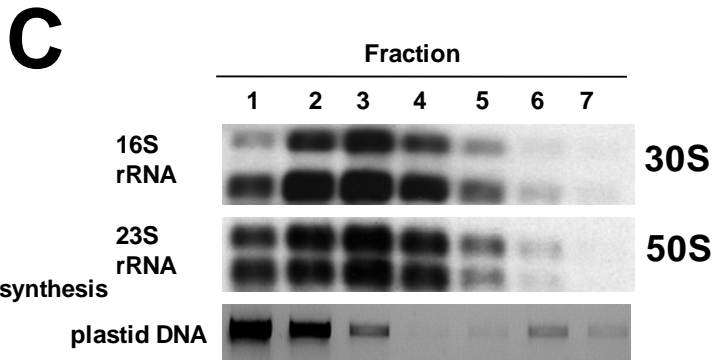
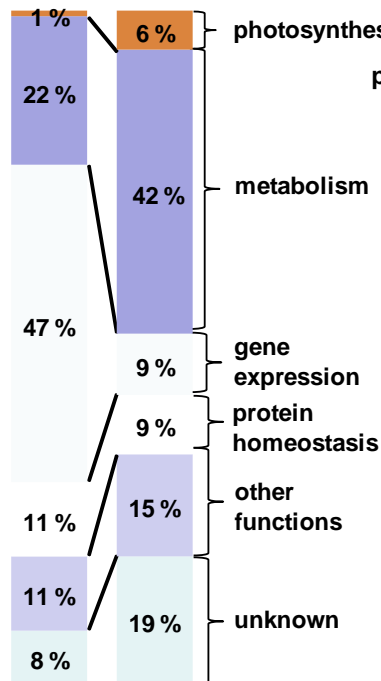
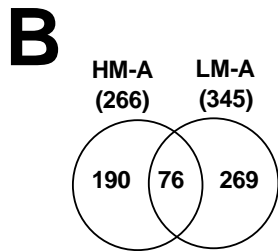
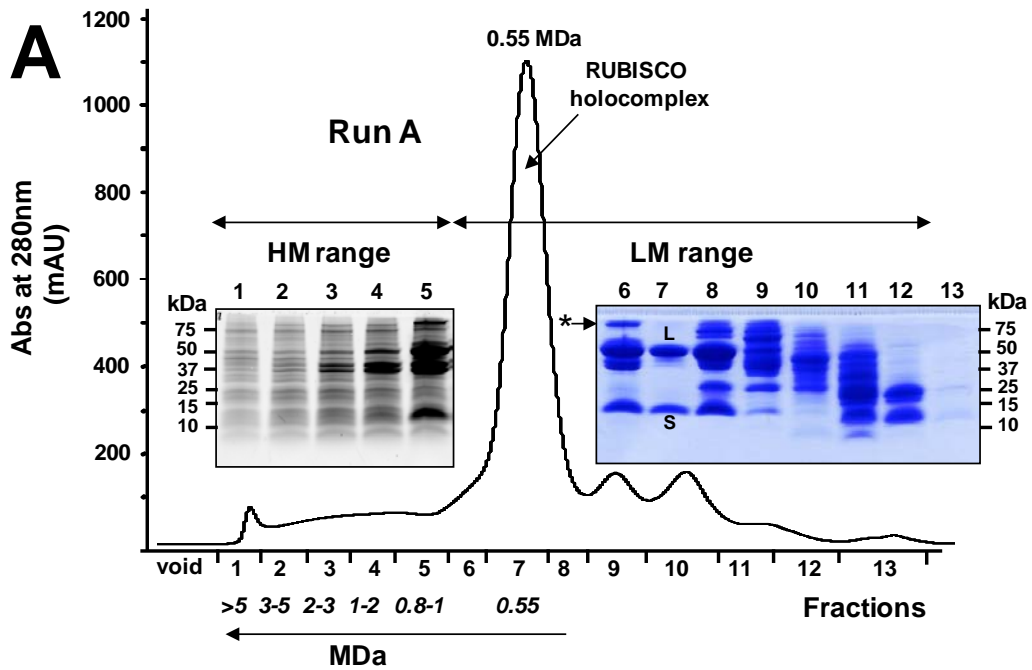


Figure 4.1. Workflow for the experimental and bioinformatics analyses of the *Arabidopsis* chloroplast stroma proteome.

Several lysis media and SEC elution conditions with varying MgCl_2 concentration, NaCl concentration and flow rates were tested for optimization of resolution and complex stability. Removal of MgCl_2 in the lysis and elution buffers led to the destabilization of high molecular weight complexes such as 70S ribosomes (data not shown). Chromatograms for runs at various NaCl concentrations (50mM, 100mM and 150 mM NaCl) revealed a slight shift to later elution times with increasing ionic strength, suggesting partial destabilization of protein interactions (not shown). However, similar overall peak profiles were observed, suggesting that the core complexes remained intact, but that transient and weak protein-protein or protein-RNA associations were destabilized at higher salt concentrations. Based on these optimizations, SEC fractionation was performed at a flow rate of 0.25 mL/min using either buffer A (25 mM HEPES pH 8.0, 10 mM NaCl , 10 mM MgCl_2) or buffer B (50 mM HEPES, 50 mM NaCl , pH 8.0, 5 mM MgCl_2).

A typical SEC chromatogram of the stromal proteome under native conditions is shown in Figure 4.2A. The highest peak at fraction 7 (~13 mL) corresponds to the ribulose-1,5-bisphosphate carboxylase oxygenase (Rubisco) holocomplex at ~550 kDa, which comprises ~58% of the total stromal mass (11) (Figure 4.2A). Five fractions for the SEC run with buffer A covering the mass range of 0.8 to >5 MDa were designated as HM-A (fractions 1-5) and the remainder of the fractions were assigned as LM-A (fractions 6-13), spanning a mass range of upto 0.8 MDa. The LM fraction covers the same native mass range as the CN-PAGE stromal proteome analysis (11), while the HM is complementary to this CN-PAGE analysis. For the SEC run with buffer B, we only analyzed and discussed the HM fraction (HM-B).

Figure 4.2. Size exclusion fractionation of Arabidopsis stroma. **A)** Chromatogram of SEC-fractionated stroma (~3 mg proteins) eluted with buffer A (25mM HEPES pH 8.0, 10mM NaCl, 10mM MgCl₂). The SEC separation with buffer A is further divided into HM-A (fractions 1-5) and LM_A (fractions 6-13) with their corresponding SDS-PAGE gels. Protein bands in the HM_A gel were visualized with the fluorescent Sypro Ruby whereas those for LM-A were stained with Coomassie Blue. 100% of fractions 1-5 were loaded on the gel, while approximately equal amounts of proteins (estimated from the integrated peak areas for each fraction) were loaded for the LM range (fraction 6-13). The Rubisco large and small subunits are indicated by L and S, respectively. The asterisk indicates the CPN60 complex of 800 kDa. **B)** Comparison of the HM-A and LM-A datasets based on the functional distribution of the non-overlapping datasets shown as stacked bar graphs. **C)** Distribution of chloroplast rRNA of 30S and 50S ribosomal particles, as well as the plastid chromosome. The upper panels show Northern blots of extracted RNA for 16S and 23S rRNA using antibodies targeting DIG-labeled rRNAs. The lower panel shows chloroplast DNA detected by PCR amplification of the plastid gene encoding for the chloroplast 16S rRNA.



The proteins in the SEC fractions were separated by SDS-PAGE (Figure 4.2A). Each gel lane was excised into slices, followed by in-gel trypsin digestion and nLC-LTQ-Orbitrap analysis of the extracted peptides using data-dependent acquisition and dynamic exclusion in a workflow previously optimized for identification and quantification by spectral counting (9). 110 MS runs were analyzed and the search results and relevant associated information (e.g. ion scores, mass errors, matched sequences, etc) were uploaded in the PPDB. The PPDB provides an integrated platform for comparing our protein identifications with other in-house MS-based proteomics experiments, annotated properties, as well as published information (29). All mass spectral data (the mgf files reformatted as PRIDE XML files) were made available via the Proteomics Identifications database (PRIDE) at <http://www.ebi.ac.uk/pride/>.

It has been shown for LC-MS-based analyses of proteomes that the number of matching MS/MS spectra, here assigned as 'spectral counts' (SPCs) correlates well with protein abundance if there is a sufficient number of SPCs obtained per protein (32-35). To correct for SPCs derived from peptides shared between proteins, adjusted SPCs (adjSPC) were calculated as the sum of unique SPC and a proportional distribution of shared SPCs (see Experimental Procedures). In addition, the relative concentration for each identified protein in the HM range, was calculated as the normalized Spectral Abundance Factor (NSAF), derived from adjSPC weighted for the number of theoretical tryptic peptides with a relevant length ('observable peptides') (36). To increase the confidence and significance of the datasets in this study, proteins that were matched to only one amino acid sequence (irrespective of charge state, post-translational modification, or number of SPCs) were excluded from further analysis.

Comparison with other chloroplast proteomics studies demonstrates enrichment of plastid gene expression components.

In total we identified 1081 proteins in the HM-A, LM-A and HM-B fractions (not shown, but see PPDB) of which 347 were not identified in our comprehensive total chloroplast proteome analysis (membranes and stroma) in which 1325 proteins were identified (9). For the purpose of the current study, we imposed an additional stringency demanding at least 2 different matched amino acid sequences per protein and at least 2 adjSPC, reducing the new dataset to 664 proteins. The reason for this extra stringency is that low abundance proteins are the focus in this study (upto more than 100,000-fold lower in abundance than Rubisco) and this extra filter reduced identification of non-chloroplast proteins and false positives, even if several ‘novel’ chloroplast proteins were removed (including PPR proteins). Following the grouping of closely related homologues and removal of 54 non-chloroplast proteins and 47 chloroplast membrane(-associated) proteins, we identified 542 stromal proteins (and protein groups) of which 86% had a TargetP predicted cTP (see PPDB). 60 of these 542 stromal proteins were not observed in our previous analysis (9); 37 of those 60 were in the HM set and were dominated by chloroplast ribosomal proteins, various splice factors and proteins with unknown function, such as PPR domain proteins and DNA binding proteins. Thus, the analysis of MDa-sized assemblies revealed low abundance proteins mostly involved in plastid gene expression.

Comparison of HM and LM datasets reveals effective SEC fractionation and enrichment for plastid gene expression components

Figure 4.2B compares the identified stromal proteins in HM-A and LM-A fractions. Importantly, there was only an 11% overlap (76 proteins) between these two datasets, showing that SEC effectively preserved and separated specific protein complexes

above and below 0.8 MDa. This overlap consisted predominantly of highly abundant proteins involved in primary carbon metabolism, such as Rubisco components. Comparison of the distribution of functions of the non-overlapping protein fractions shows a >5 fold enrichment in plastid gene expression in the HM fractions (47% all proteins), at the expense of all other functions, except for protein homeostasis components (*e.g.* chaperones).

Table 4.1 lists the annotated functions and relative abundance of proteins (NSAF) in the HM sets. Moreover, the heat maps (from NadjSPC) show the distribution for each protein across the SEC mass range. Excluding the 77 proteins that were also found in the LM dataset, 27 out of the 30 most abundant proteins are ribosomal subunits, further confirming that the HM fractions are highly enriched with proteins involved in plastid gene expression. This enrichment is consistent with the presence of plastid rRNAs and plastid DNA in the HM fraction (Figure 4.2C). The average relative abundance of ribosomal proteins across the HM fractions is 0.01 (Table 4.1) and this will be used to assess the accumulation levels of ribosome-associated proteins (see further below).

Profiling of macromolecular assemblies in the stroma

The SEC elution profiles of proteins, derived from NadjSPC values across the chromatogram, reflect the size range(s) of complexes in which they participate, as well as their quantitative distribution over these complexes. To obtain a global view of SEC migration protein profiles and to facilitate the identification of putative interacting proteins, hierarchical clustering was employed to group proteins that exhibit similar SEC elution trends (Figure 4.1). Several dendograms from clustering with different minimum thresholds for total adjSPCs were analyzed. Based on these tests, we chose a

Table 4.1. Elution profiles of stromal proteins identified in megadalton-sized SEC fractions

Protein	Accession	MWt (kDa) ^a	unique pep.	% Cov ^b	MAX ^c	SEC Fraction ^d					Rel. Abun. (x10 ⁻³)
						1	2	3	4	5	
30S Ribosomes											
RPS1	At5g30510.1	41	12	56	82						5.7
RPS2	Atcg00160.1	27	10	57	50						11.7
RPS3	Atcg00800.1	25	15	65	109						14.8
RPS4	Atcg00380.1	23	12	57	66						15.0
RPS5	At2g33800.1	27	6	25	44						3.5
RPS6	At1g64510.1	18	2	17	5						1.5
RPS7	Atcg01240.1	17	8	40	23						7.1
	Atcg00900.1										
RPS8	Atcg00770.1	15	8	54	88						20.5
RPS9	At1g74970.1	17	9	56	189						21.5
RPS10	At3g13120.1	13	3	45	16						1.6
RPS11	Atcg00750.1	6	6	43	43						22.7
RPS12	Atcg00065.1	7	2	29	2						11.0
	Atcg01230.1										
	Atcg00905.1										
RPS13	At5g14320.1	14	13	70	70						15.7
RPS14	Atcg00330.1	12	2	23	12						17.5
RPS15	Atcg01120.1	11	6	51	54						18.9
RPS16-1	At4g34620.1	13	5	60	27						10.4
RPS17	At1g79850.1	11	3	29	15						9.1
RPS18	Atcg00650.1	12	2	19	20						7.7
RPS19	Atcg00820.1	11	5	61	32						17.5
RPS20	At3g15190.1	14	7	29	52						15.3
PSRP-2	At3g52150.1	21	7	53	43						4.3
PSRP-3	At1g68590.1	12	2	35	15						1.3
PSRP-4	At2g38140.1	7	3	30	3						7.4
PSRP-7 (as part of PETs)	At4g29060.1	95	42	65	117						1.6
50S Ribosome											
RPL1	At3g63490.1	30	15	51	151						14.2
RPL2	Atcg00830.1	25	13	44	27						11.8
	Atcg01310.1										
RPL3	At2g43030.1	24	12	56	106						9.9
RPL4	At1g07320.1	27	12	43	113						11.0
RPL5	At4g01310.1	24	16	66	76						8.8
RPL6	At1g05190.1	18	13	62	64						11.6
RPL9	At3g44890.1	18	9	56	101						14.7
RPL10	At5g13510.1	20	9	55	55						9.3
RPL11	At1g32990.1	17	9	64	71						10.7

Table 4.1 (continued)

RPL12	At3g27830.1 At3g27850.1	14	5	63	129		26.6
RPL13	At1g78630.1	21	11	55	30		7.0
RPL14	Atcg00780.1	14	6	41	62		19.8
RPL15	At3g25920.1	23	10	50	51		6.0
RPL16	Atcg00790.1	15	5	42	50		9.8
RPL17	At3g54210.1	18	4	20	9		3.6
RPL18A	At1g48350.1	14	5	41	15		6.0
RPL18B	At3g20230.1	14	4	36	6		0.5
RPL19A	At4g17560.1	18	6	51	31		3.6
RPL19B	At5g47190.1	18	8	63	19		8.2
RPL20	Atcg00660.1	14	6	45	29		7.5
RPL21	At1g35680.1	15	11	62	40		13.1
RPL22	Atcg00810.1	19	7	43	22		8.1
RPL23	Atcg00840.1 Atcg01300.1	11	4	32	5		4.9
RPL24	At5g54600.1	17	5	34	23		5.8
RPL27	At5g40950.1	20	5	25	37		7.6
RPL28	At2g33450.1	5	5	57	15		20.8
RPL29	At5g65220.1	14	6	41	95		28.5
RPL31	At1g75350.1	13	8	59	40		9.8
RPL32	Atcg01020.1	6	1	15	4		7.4
RPL33	Atcg00640.1	8	2	30	5		2.7
RPL35	At2g24090.1	10	1	16	11		4.2
RPL36	Atcg00760.1	4	1	24	2		4.5
PSRP-5	At3g56910.1	10	3	29	7		3.9
Translation							
Initiation factor (IF-1)	At4g11175.1	16	2	34	4		2.9
Initiation factor 2 (IF-2)	At1g17220.1	103	26	38	82		2.9
Initiation factor 3 (IF-3)	At4g30690.1	30	3	14	4		0.3
Elongation factor G (EF-G)	At1g62750.1	78	4	8	3		0.1
Elongation factor Tu (EF-Tu-1)	At4g20360.1	45	18	61	208		17.8
TypA/bipA	At5g13650.1	68	10	29	31		0.9
LepA	At5g08650.1	70	12	29	36		1.0
PSRP-1	At5g24490.1	32	5	32	16		2.0
Co-translation							
cpSRP54	At5g03940.1	53	13	38	38		1.4
Trigger Factor	At5g55220.1	59	7	17	13		0.3
Trigger factor	At2g30695.1	16	2	16	1		0.1
Peptide deformylase 1B (PDF1B)	At5g14660.1	25	4	19	6		0.7

Table 4.1 (continued)

Methionine aminopeptidase 1B (MAP1B)	At1g13270.1	35	3	15	3		0.1
cpHSP70-1	At4g24280.1	67	20	211	54		2.3
cpHSP70-2	At5g49910.1	67	20	138	32		2.0
GrpE-1	At5g17710.1	29	3	89	30		1.4
Ribosome Biogenesis^f							
AtObgC (30S, 50S)	At5g18570.1				5		0.1
Hflx (50S, 70S)	At5g57960.1	55	2	7	2		0.1
AtNOA1 (30S)	At3g47450.1	62	2	5	3		0.3
YqeH (30S)	At3g57180.1	66	3	7	9		0.3
Era (30S)	At5g66470.1	44	4	10	2		0.1
EngA (30S, 50S, 70S)	At3g12080.1	66	5	9	9		0.5
RimM (30S)	At5g46420.1	68	3	11	4		0.4
RbfA (30S)	At4g34730.1	18	2	22	2		0.1
RsmD	At3g28460.1	30	2	12	5		0.2
YrdC family protein	At3g01920.1	34	2	10	5		0.2
SpoU family protein	At2g19870.1	66	3	11	5		0.1
Iojax-related	At3g12930.1	19	3	30	7		2.9
DAL (DAG-like protein)	At2g33430.1	25	3	30	7		0.4
RNA degradation and maturation							
CSP41B	At1g09340.1	43	18	73	91		5.1
PNPase (RIF10)	At3g03710.1	94	17	34	76		2.4
RNAse J	At5g63420.1	93	30	42	120		3.7
RNA helicases (RH)							
RH3	At5g26742.1	74	30	57	102		5.5
RH26	At5g08610.1	86	6	10	7		0.2
DEAD/DEAH box helicase	At1g70070.1	126	18	25	49		1.1
DEAD/DEAH box helicase	At4g09730.1	69	7	21	4		0.2
RH50	At3g06980.1	84	3	6	5		0.1
PEP complex							
RNA polymerase A (rpoA)*	Atcg00740.1	38	3	9	3		0.5
RNA polymerase B (rpoB)*	Atcg00190.1	121	5	8	8		0.3
RNA polymerase C1 (rpoC1)*	Atcg00180.1	79	4	7	8		0.4
RNA polymerase C2 (rpoC2)*	Atcg00170.1	156	7	7	5		0.1
Transcription and DNA Binding							
DNA gyrase subunit A *	At3g10690.1	97	9	14	16		0.3
pTAC2*	At1g74850.1	89	2	3	4		0.2

Table 4.1 (continued)




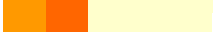
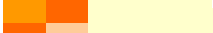
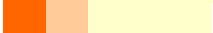



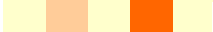
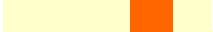






















pTAC16*	At3g46780.1	52	4	11	7		0.4
pTAC17*	At1g80480.1	42	4	13	11		0.4
Thioredoxin*	At3g06730.1	12	2	20	3		1.1
pfkB-type carbohydrate kinase*	At1g69200.1				2		0.0
DNA mismatch repair MutS	At1g65070.1	91	2	4	5		0.1
DNA repair protein recA	At1g79050.1	42	5	23	12		0.4
Deadbox DNA helicase	At5g35970.1	100	26	40	123		4.2
DNA topoisomerase	At4g31210.1	137	5	7	9		0.1
mTERF family protein ARCA; nucleotide binding WD-40	At1g18080.1	36	3	13	4		0.5
DnaJ domain-containing protein	At2g22360.1	39	7	11	4		0.2
DnaJ domain-containing protein	At4g39960.1	48	7	50	22		0.9
RNA splicing							
CAF1	At2g20020.1	75	2	4	2		0.1
Protein with one CRM domain	At4g39040.1	26	2	10	2		0.2
Protein with three CRM domains	At3g18390.1	90	4	9	3		0.2
RNC1	At4g37510.1	57	2	5	3		0.1
WTF1	At4g01037.1	53	4	11	7		0.3
matK maturase	Atcg00040.1	63	2	7	2		0.1
PPR proteins							
PPR4	At5g04810.1	100	5	9	6		0.1
CRP1	At5g42310.1	80	6	13	9		0.3
P67	At4g16390.1	76	6	12	12		0.4
PPR	At3g53700.1	77	3	8	2		0.1
PPR	At5g46580.1	76	5	9	2		0.1
PPR	At1g19720.1	101	3	4	5		0.1
PPR	At3g04760.1	59	3	4	6		0.2
PPR	At3g49240.1	71	6	14	13		0.3
RNA-binding proteins							
CP29 B'	At2g37220.1	26	3	34	16		1.6
CP33	At1g01080.1	25	6	41	19		2.3
RNA binding protein	At2g35410.1	26	8	52	30		3.0
RNA recognition motif (RRM)	At4g09040.1	25	5	40	40		4.0
S1 RNA-binding protein	At3g23700.1	33	5	29	15		1.6
S1 RNA-binding protein	At1g12800.1	80	2	5	1		0.1
S1 RNA-binding protein	At1g71720.1	48	4	13	2		0.1
HCF173	At1g16720.1	57	24	57	77		3.7

Table 4.1 (continued)

Pyruvate Dehydrogenase (PDC)							
E1 alpha subunit	At1g01090.1	41	13	47	65		2.0
E1 beta subunit	At2g34590.1	37	2	8	4		0.5
E1 beta subunit	At1g30120.1	40	4	19	14		4.6
E2 (LTA2)	At3g25860.1	45	13	49	121		2.6
E2	At1g34430.1	43	11	40	111		0.4
E3 (ptlpd1)	At3g16950.1	53	11	41	20		1.8
E3 (ptlpd2)	At4g16155.1	58	13	43	58		
Acetyl Carboxylase (ACCase)							
Biotin carboxylase (CAC2)	At5g35360.1	51	22	66	190		6.1
Carboxyltransferase (CT) alpha SU	At2g38040.1	80	30	51	110		3.3
SUubunit	Atcg00500.1	56	11	30	53		3.1
Biotin carboxyl carrier protein (BCCP-1)	At5g16390.1	21	7	61	70		8.4
Biotin carboxyl carrier protein (BCCP)	At3g15690.2	22	9	77	31		7.1
Biotin carboxyl carrier protein (BCCP)	At1g52670.1	24	6	48	38		5.9
Biotin carboxyl carrier protein (BCCP)	At3g56130.1	24	4	35	35		3.8
Ketoacyl-ACP Synthase I (KAS1)	At5g46290.1	45	3	12	5		0.2
Acetyl-CoA synthetase	At5g36880.1	73	2	5	3		0.1
Acyl carrier protein 4 (ACP4)	At4g25050.1	9	2	51	30		5.6

^{a,b} The predicted chloroplast transit peptide (cTP) was removed for nuclear-encoded proteins.

^c The maximum adjSPC (MAX) for each protein across the five fractions.

^d Based on the adjSPC normalized to MAX per fraction (NadjSPC). Map is based on color code below.

^e Based on Normalized Spectral Abundance Factor (NSAF) for the HM dataset

^f Indicated are ribosomal particle(s) by which they interact based on bacterial studies (see text for references).

* Proteins that were found in transcriptionally active chromosomes (pTACs), see (117).

Color Code :

0.80-1.0
0.60-0.80
0.40-0.60
0.20-0.40
0-0.20

minimum threshold of 20 total adjSPC per protein; this threshold minimized noise that might skew the linkages among correlated proteins. The resulting dendrogram (Figure 4.3) displayed three main clusters namely cluster I (essentially proteins confined to complexes >2 MDa), cluster II (1-3 MDa) and cluster III (< 1 MDa). The first two clusters are linked and they are both dominated by ribosomal proteins and associated factors, as will be discussed in the following sections. Cluster III is subdivided into cluster III-1 containing several ribosomal proteins, and cluster III-2 consisting mainly of proteins that are starting to elute below 1 MDa. Cluster III-2 included components of the Rubisco complex (~550 kDa), Chaperone 60 (~800 kDa), Fd-glutamine:2-oxoglutarate amidotransferase (~700kDa) and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) (600 kDa), which were well resolved by CN-PAGE (11).

Many of the proteins that were grouped by hierarchical clustering coincided with known assemblies, such as plastid PDC, 30S, 50S, and 70S ribosomal particles and ACCase (Figure 4.3). For instance, the maize mitochondrial PDC was found to have an estimated mass of about 8-9 MDa (18, 37). Similarly as for the mitochondria PDC, the *Arabidopsis* plastid-localized PDC components cluster in the mass range greater than 3 MDa (Figure 4.3B). As we will show in more detail below, the clustering further confirmed that SEC effectively separated stromal complexes. The remainder of the Results section will discuss the composition, organization and function of the observed MDa-sized macromolecular assemblies.

70S Ribosomes

The spinach plastid 70S ribosome consists of 58 ribosomal proteins with 25 and 33 components comprising the 30S and 50S complexes, respectively (12-14). The majority of these proteins are homologous to bacterial 70S ribosome components, but

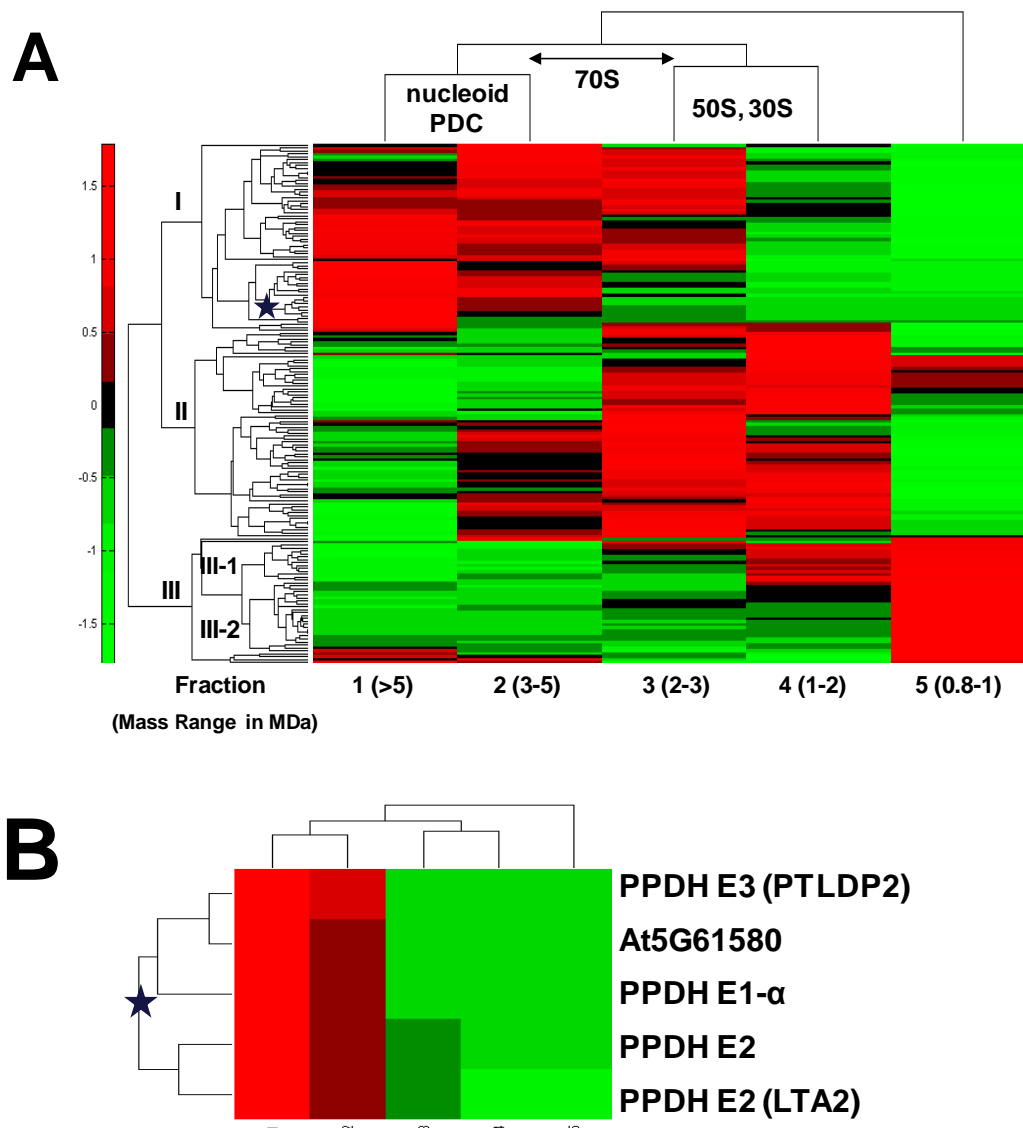


Figure 4.3. Overview of SEC migration profiles of MDa-sized complexes. **A)** Dendrogram of SEC migration profiles obtained by hierarchical clustering of HM-A proteins (total adjSPC cutoff of 20). The color code representing the relative abundance of the protein per SEC fraction is shown. The three main clusters (I, II and III) and subclusters of III (III-1 and III-2) are also noted. **B)** The subcluster (marked as a star in A) of the pyruvate dehydrogenase complex (composed of E1, two E2 proteins and E3 subunits) with SEC elution profile peaking at fraction 1 (mass range >5MDa).

several are unique to the chloroplast (plastid specific ribosomal proteins or PSRPs) and they are proposed to perform plastid-specialized ribosomal functions (12, 14, 38).

PSRP-2,-3,-4 associate with the 30S subunit, whereas PSRP-5 and PSRP-6 associate with the 50S subunit (12-14). PSRP-7 interacts with the 30S particle in *C. reinhardtii*, *Arabidopsis* and rice but is missing in spinach (39). PSRP-7 is synthesized as a polyprotein consisting of the mature PSRP-7 and the elongation factor EF-Ts (PETs) which might be post-translationally processed to render various fused or independent proteins (39). *Arabidopsis* orthologues for all but three ribosomal proteins were identified in the HM fractions, accounting for 57 ribosomal proteins (Table 4.1). About 50% and 75% of the 30S and 50S ribosomal components in *Arabidopsis* are nuclear-encoded, respectively, consistent with that observed in spinach (12-14). Interestingly, RPL23 is nuclear-encoded in spinach, but has two chloroplast-encoded genes in *Arabidopsis* (Atcg00840.1 and Atcg01300.1) suggesting evolutionary divergence.

Several *Arabidopsis* chloroplast ribosomal proteins that were identified are encoded by multiple genes. Five of these, namely RPS7, RPS12, RPL2, RPL12 and RPL23 are comprised of two or three identical gene products. Four of these protein groups are chloroplast-encoded, suggesting a possible regulatory role or adaptation to specific conditions in plastid gene expression. In addition, two paralogues each were assigned to each RPL18 (23% identity) and to RPL19 (74% identity, variable N-terminal regions) and these related proteins were distinguishable by MS analysis. Interestingly, RPL18A was found to be six times more abundant than RPL18B. In addition, one chloroplast and two nuclear genes encode for RPS16 in *Arabidopsis* (40, 41), but only one paralogue (At4g34620.1) was detected in this study.

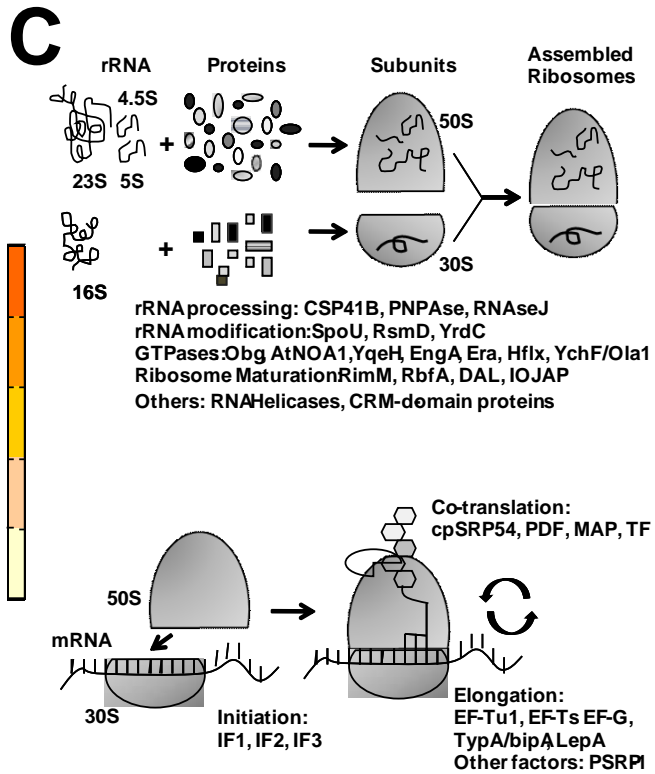
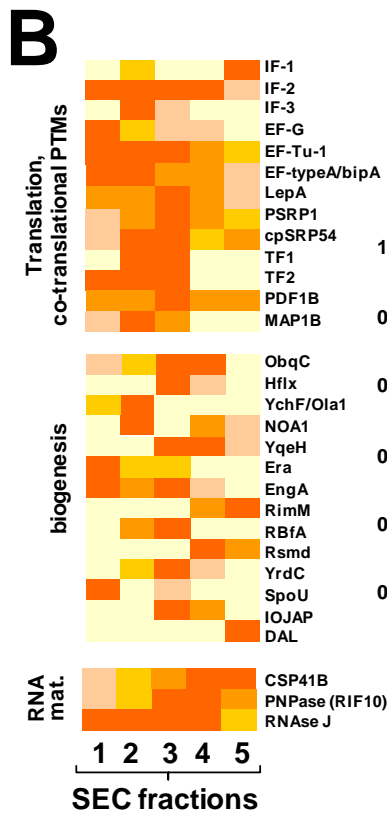
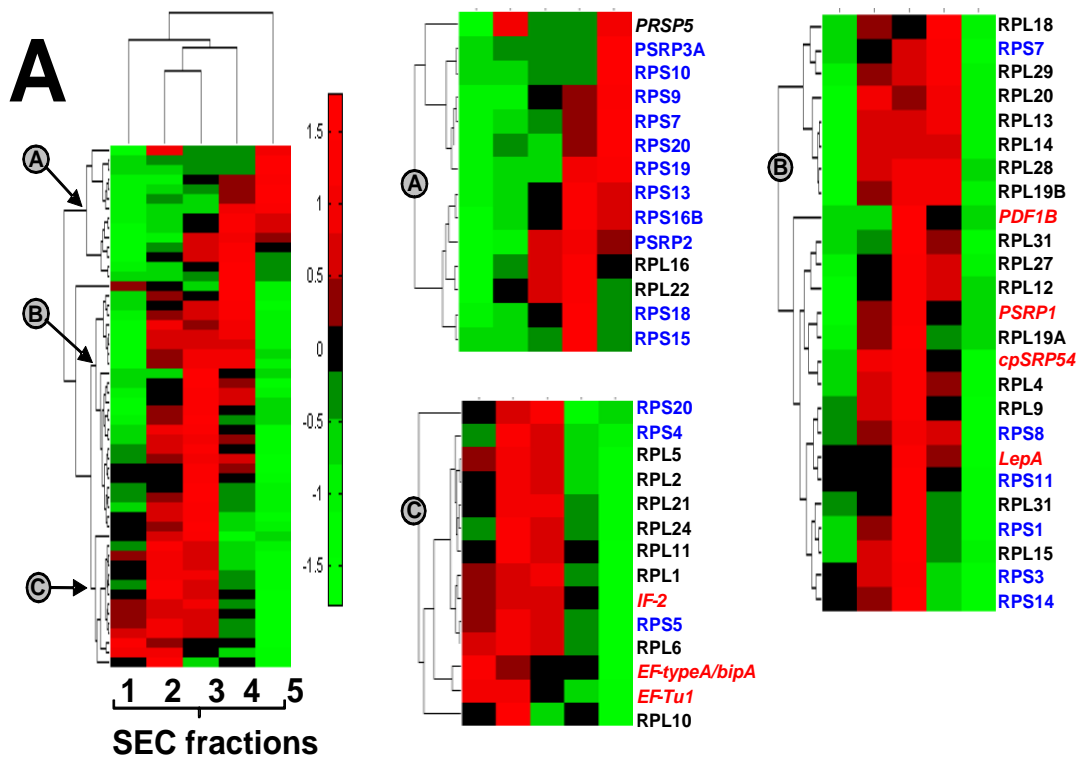
The ribosomal components RPS21, RPL34 and PSRP-6 were neither detected in this study, nor in previous published chloroplast proteomics studies on *Arabidopsis*. One reason might be that they are short, lysine- and arginine-rich proteins, yielding small tryptic peptides that were not amenable for LC-MS analysis. Another possibility

is the location and function of these proteins in the ribosome where they could easily be detached. RPS21 is found comprising the top of the head of the 30S region in prokaryotic ribosomes (42). PSRP-6 has been found to be loosely associated with the ribosome (38) and is detected at lower amounts compared to other spinach ribosomal proteins (12).

To explore the assembly state of the ribosome and to co-localize ribosome-associated factors, we carried out a separate hierarchical clustering of all 53 assigned ribosomal proteins, (co-)translational factors and ribosome biogenesis factors with 20 or more adjSPC (Figure 4.4A). The left hand panel shows the complete dendrogram in which three clusters (A,B,C) were distinguished, with cluster A representing the 30S particle, cluster B the 50S particle and cluster C the translating 70S ribosome. Close-ups of these three clusters with protein names are shown in the three other panels of Figure 4.4A. Protein components and factors known/expected to specifically associate with the 30S particle are in blue, and those part of the 50S particle are in black. Proteins that have a function in ribosome biogenesis or translation, but that are not integral part of the ribosome, are listed in red and italics. The 30S particle peaked in the 1-2 MDa range (fraction 4) and the 50S particle peaked in the 2-3 MDa range (fraction 3), whereas the 70S ribosomes peaked in fraction 2 (Figure 4.4A). Northern blot analyses of rRNAs for 30S and 50S extracted from these fractions were consistent with these profiles (Figure 4.2C). Overall, the protein and RNA profiles of these ribonucleoparticles correspond to 30S and 50S subunits, possibly in different stages of maturation, 70S ribosomes and polysomes.

We note that the ribosome-associated factors are sub-stoichiometric to the ribosomal proteins and many were quantified with less than 20 total adjSPC and they were therefore excluded from the clustering in Figure 4.4A. However, to show

Figure 4.4. Characterization of the 30S and 50S ribosomal particles and associated proteins by hierarchical clustering, annotation and heat maps. **A)** Hierarchical clustering of all 53 assigned ribosomal proteins, (co-)translational factors and ribosome biogenesis factors with 20 or more adjSPC. The left hand panel shows the complete dendrogram; three clusters were distinguished, with cluster A representing the 30S particle, cluster B the 50S particle and cluster C the translating 70S ribosome. Close-ups of these three clusters with protein names are shown in three other panels. Protein components and factors specifically associated with the 30S particle are in blue, and those part of the 50S particle are in black. Proteins that have a function in ribosome biogenesis or translation, but that are not integral part of the ribosome, are listed in red and in italics. **B)** Schematic, summarizing overview of the ribosome assembly process (upper panel) and protein translation process (lower panel) with identified factors indicated. **C)** Heat map of annotated proteins (based on NadjSPC) identified with 2 or more adjSPC and are known or expected to associate with ribosomes and pre-ribosomal particles. The proteins are grouped according to function as indicated (translation and co-translational modifications, ribosome biogenesis, RNA maturation). The scale is indicated with NadjSPC values for each fraction normalized to the maximum value (see Table 4.1).



the distribution of these proteins across the HM fractions, we show a heat map of all 30 proteins in Figure 4.4B. In the next sections we will discuss these ribosome-associated factors in more detail.

Protein translation

The translation machinery requires the participation of ribosome-associating initiation, elongation, and termination factors, as well as proteins aiding in ribosome recycling (42, 43). Translation initiation proceeds with the binding of the mRNA transcript to a free 30S subunit, followed by association of the initiator tRNA (fMet-tRNA) and the initiation factors IF-1, IF-2 and IF-3, together forming a pre-initiation complex (42, 43) (Figure 4.4C). We observed all three IFs, with IF1 and IF2 being ~10-fold more abundant than IF3.

Elongation ensues after binding of the mature 50S subunit to the pre-initiation complex and subsequent release of the three IFs (Figure 4.4C). Several rounds of elongation of the nascent peptide proceed by alternating actions of EF-Tu, which introduces new aminoacyl tRNAs into the peptidyl transfer center in the 50S ribosome, and EF-G which translocates the peptidyl-tRNA after spontaneous peptidyl transfer occurs. During each round, EF-Tu dissociates from the peptidyl transfer center upon GTP hydrolysis with the help of EF-Ts, its nucleotide exchange factor (42, 43). A point mutation in the chloroplast-localized EF-G protein leads to impairment of chloroplast development within cotyledons (which become white) but not in true leaves (44). Aside from its function in translation, the EF-Tu orthologue in maize plastids was suggested to also serve as a chaperone during heat stress (45, 46). Additional translation elongation factors include the GTPases LepA and TypA/bipA proteins; in bacteria both have been shown to bind to the 70S ribosomes at the same site as EF-G (47-49). In *E. coli*, LepA is proposed to recognize ribosomes with miss-

translocated tRNAs and induce back-translocation for corrective retranslocation (50). The bacterial and plant TypA/BipA appear to be particularly important under stress conditions (51-55).

The elongation factors detected in the HM fractions are EF-Tu, EF-G, EF-Ts, LepA and TypA/BipA (Figure 4.4B). These elongation factors have different SEC elution profiles (Figure 4.4A and 4.4B), but they were mostly found in the 2-5 MDa range and were most likely associated with 70S ribosomes or polysomes. The primary elongation factor EF-Tu was an abundant protein (same abundance range as the ribosomes) and was seen as a broad peak across all the five fractions (Figure 4.4B and 4.4C). Similarly, CN-PAGE analysis of stroma showed EF-Tu migrating at multiple native masses (11). TypeA/bipA, LepA and PSRP-1 exhibited similar accumulation levels at about 20% of the ribosome. The PETs which harbors the EF-Ts domain fused to PSRP-7 peaked in the 0.8-1 MDa range (Table 4.1). Cryo-EM structure analysis, as well as genetic and biochemical studies, have shown that PSRP-1 is a translation factor rather than an integral ribosomal protein (38, 56). PSRP-1 contacts the space between the 30S and 50S subunit thereby stabilizing the 70S ribosome and is proposed to be involved in translation regulation during stress (56). PSRP-1 clustered with the 50S particle and 70S ribosomes, which supports its role as a translation factor.

Translation termination occurs after release factors (RFs), which recognize the stop codon, bind to the ribosome. Ribosome recycling then occurs wherein a ribosome recycling factor (RRF) binds to post-termination complexes and, in coordination with EF-G, splits the 70S ribosomes for the next round of translation (42, 43). No RFs and RRF were observed in the HM fraction, but *Arabidopsis* orthologues of RF1 (At3g62910.1) and RRF (At3g63190.1) were found in LM-A fractions.

Co-translational Protein Processing and Folding

During translation, the nascent polypeptide chain that extends out of the peptidyl exit tunnel of the 70S ribosome is subjected to N-terminal modifications, as well as protein folding (57). Proteins involved in such co-translational activities were indeed found in fractions with 70S ribosomes (Table 4.1, Figure 4.4). These include the enzymes peptide deformylase (PDF) and methionine aminopeptidase (MAP), the chloroplast signal recognition particle (cpSRP54) and trigger factor (TF).

PDF and MAP are hydrolytic enzymes that together perform co-translational N-terminal methionine removal. PDF removes all *N*-formyl groups, exposing the amino group of the first methionine, a prerequisite for the subsequent action of MAP (57). Two PDFs (PDF1A and PDF1B) have been found in *Arabidopsis* but only PDF1B (At5g14660.1), shown to be dually targeted to the chloroplast and mitochondria in *Arabidopsis* (58-60), was observed in this study. Several MAP proteins (MAP1B, MAP1C and MAP1D) exist in plants and the chloroplast-localized MAP1B (At1g13270.1) (59) was found in this study.

In co-translational protein targeting, nascent polypeptides with hydrophobic domains are recognized by the signal recognition particle (SRP) and they are then targeted as a ribosome nascent chain complex to membranes (57). The chloroplast sorting component cpSRP54 has both a role in post-translational targeting of nuclear-encoded thylakoid proteins, and has also been implicated in co-translational targeting (61-64) and, consistently, was detected in this study peaking with ribosomes at the 2-5 MDa range (Figure 4.4). In addition, polypeptides are being folded as they emerge from the peptide exit tunnel. The *E. coli* TF binds to the 70S exit tunnel and prevents miss-folding and aggregation of emerging nascent proteins (65). A 55 kDa TF (At5g55220.1) with conserved domains, as well as a 19 kDa truncated form of TF

(At2g30695.1) were identified with the 70S ribosomal subunits. The 55 kDa full length TF protein was 3-fold more abundant than the truncated form (Table 4.1).

Chloroplasts contain two dominant protein chaperones systems, namely the CPN60/CPN21/10 (66, 67) and the HSP70/GrpE system (68). The *Arabidopsis* HSP70-1 (At4g24280.1) and HSP70-2 (At5g49910.1) are abundant proteins with a broad substrate pool and their abundance was constant across all the HM-A fractions, consistent with its wide range of substrates. The nucleotide exchange factor of HSP-70, GrpE (At5g17710.1) showed a similar distribution. Based on information about *E. coli* homologues, it is quite likely that HSP70 also functions in co-translational folding, thereby assisting the TF protein (57).

Ribosome biogenesis

The formation of a functional ribosome from more than 50 proteins and four rRNA molecules entails a complex series of coordinated processes, including processing and modification of ribosomal components, assembly and maturation (69, 70). Several *Arabidopsis* orthologues of bacterial ribosome assembly factors were identified in the HM-A fractions (Figure 4.4B and Table 4.1). Most of these proteins were initially annotated with unknown function, but careful domain analysis, observed homology to bacterial ribosome biogenesis factors (from PSI-BLAST searches;(71), and their detection in the ribosomal fractions provided support for their involvement in ribosome biogenesis.

Several GTPases have been implicated in ribosome assembly in bacteria (47, 48, 72). The bacterial Obg (ObgE) co-fractionates with the 50S subunit (73-75) and the 30S subunit, but not with the 70S ribosomes (75). It associates with 16S and 23S rRNAs *in vitro* and with an RNA helicase (75). A chloroplast-localized *Arabidopsis* Obg homologue AtObgC (At5g18570.1) has been identified and its GTPase activity

has been demonstrated (76). Knockout mutants of AtObgC are embryo lethal showing its crucial role in chloroplast development (76, 77) and studies on knockdown mutants have shown that AtObgC is essential for plastid rRNA processing (76). HflX, a GTPase that is related to ObgE, associates with the 50S particle (78, 79), the 70S ribosomes (79) and both 16S and 23S rRNA (78). In this study, both AtObgC and the HflX homologues (At5g7960.1) were observed mainly at ~2 MDa (the 50S ribosomal fraction) consistent with the results from bacterial studies. YchF/OLA1 is another subclass of the Obg family and exhibits altered nucleotide specificity by hydrolyzing ATP more efficiently than GTP (80). The *Arabidopsis* YchF/OLA1 protein (At1g56050.1) migrated in the 3-5 MDa range (Figure 4.4B).

YqeH, a circularly permuted GTPase, also co-sediments with the 30S particle and is essential for 16S rRNA maturation (81). Two YqeH orthologues in *Arabidopsis* were identified in the HM fractions and they behave similarly (found in the 30S and 70S fractions). One of these orthologues, the AtNOA1 protein (At3g47450.1), exhibits GTPase activity *in vitro* (82). Moreover, the bacterial YqeH complements AtNOA1-deficient mutants providing further evidence that AtNOA1 functions as a chloroplast-localized YqeH in plants (83, 84).

Era is another *E. coli* GTPase that binds 16S rRNA and is involved in 30S subunit maturation (85). We observed that the Era orthologue in *Arabidopsis* (At5g66470.1) was present in the 70S fractions rather than the 30S fraction. The *E. coli* EngA (YphC in *B. subtilis*), has two contiguous GTPase domains whose nucleotide occupancy modulates its binding to either 30S alone or to 30S, 50S and 70S ribosomes (86). It co-sediments with 16S and 23S rRNAs (86) and is essential for bacterial growth (87). The *Arabidopsis* orthologue of EngA (At5g6050.1) migrates at > 2 MDa range (50S, 70S and polysomes fractions). Two mutant alleles of the *Arabidopsis* EngA are arrested in embryogenesis at the globular stage (88),

<http://www.seedgenes.org>) consistent with an essential function in chloroplast development.

Additional ribosomal biogenesis factors involved in 16S rRNA processing and 30S ribosome maturation include RimM and the ribosome binding factor A protein (RbfA) (89-92). The *Arabidopsis* homologues of RimM (At5g46420.1) peaked at the 30S fraction (0.8-2 MDa) and RbfA peaked with the 70S ribosome (Figure 4.4B; Table 4.1).

rRNA modifications, particularly, base methylations are common to all organisms. Among the bacterial rRNA/tRNA methyltransferases are RsmD, YrdC and SpoU (93-96). The *Arabidopsis* orthologues of these rRNA modifying enzymes migrated at different masses with RsmD (At3g28460.1) at 0.8-2 MDa, YrdC (At3g01920.1) at 2-3 MDa and SpoU (At2g19870.1) at >5MDa (Figure 4.4B, Table 4.1). Other maturation factors include the DAG-like protein DAL and IOJAP. DAL (At2g33430.1) eluted at the 0.8 -1 MDa range and is involved in the maturation of rRNAs and ribosome assembly (97). The IOJAP protein (At3g12930.1) eluted at the 2-3 MDa fraction. In maize, IOJAP-deficient leaf tissues are ribosome-less and they lack 16S and 23S rRNAs and protein synthesis activity (98).

RNA maturation and degradation.

Three ribonucleases were observed in the HM fractions (Figure 4.4B and Table 4.1) and various lines of evidence suggest their involvement in rRNA maturation in various ribosomal assembly stages. CSP41A (At3g63140.1) and CSP41B (At1g09340.1) are two related endoribonucleases with multiple roles in chloroplast gene expression (99, 100). In *Arabidopsis*, CSP41B co-purified with pre-ribosomal particles but not with mature ribosomes nor polysomes (99, 100). CN-PAGE analysis of stroma showed that both CSP41A and CSP41B migrated at >1MDa range (in the stacking gel) (11).

Consistently, we observed only CSP41B mostly in fractions with 30S particles. In contrast, CSP41A was identified in LM-A fractions together with the remainder of CSP41B, in agreement with previous observations (11, 100).

Polynucleotide phosphorylase (At3g03710.1) or PNPase is an exoribonuclease that is indispensable for 3'-end maturation of 23S rRNA transcripts and the efficiency of the 3'-end processing and polyadenylation of mRNAs, as well as RNA degradation (101-104). In *E. coli*, PNPase is part of a “degradosome” complex along with the endoribonuclease RNase E, a DEAD-BOX RNA helicase and the glycolytic enzyme enolase (105). However, the chloroplast PNPase has been observed as a 600-kDa homooligomer in spinach (106) and as a 410-kDa tetramer in Arabidopsis (11). Interestingly, PNPase in the SEC fraction was observed to peak in ~1-3MDa (50S fractions) (Figure 4.4B) suggesting interactions with RNA-containing complexes.

RNase J is another endonuclease that is implicated in 16S rRNA maturation and associates with both assembled 70S ribosomes and 30S particles suggesting a role in ribosome assembly (107). The *Arabidopsis* orthologue (At5g63420.1) was found to be essential for embryogenesis (EMB2746) (88), <http://www.seedgenes.org>. In this study, it was found eluting at a wide range of masses (> 1MDa) at relatively abundant levels suggesting interactions with RNA bound to various ribonucleoprotein complexes (polysomes, 70S and 30S particles).

RNA helicases

DEAD-box proteins possess the characteristic Asp-Glu-Ala-Asp sequence, an RNA-binding motif and an ATP-hydrolyzing domain, and are mainly involved in ATP-dependent rearrangement of inter- and intra-molecular RNA structures or remodeling of ribonucleoprotein complexes (108, 109). There are 58 predicted genes for DEAD-

box proteins in *Arabidopsis* (110, 111); nine of these are predicted by TargetP (2) to be plastid-localized.

Several DEAD-box RNA helicases (RHs) were identified in this study and appeared in almost all fractions >1 MDa, suggesting associations with a variety of RNA-containing protein complexes. RH3 (At5g26742.1), which belongs to a subgroup of DEAD-box proteins containing a Gly/Arg/Ser-rich C-terminal extension (112) was found to be the most abundant (10x higher) of these RHs. RH3 was also identified in fractions above 1 MDa in the CN-PAGE fractionation of stroma (11). An *Arabidopsis* RH3 knockout mutant is embryo-lethal (EMB1138) ((88), <http://www.seedgenes.org>) indicating a crucial role in plant development. RH26 (At5g0860.1) has a long N-terminal extension containing seven internal Arg-Ser-Asp repeats (112). In addition, RH26-deficient plants display a pale-green leaf phenotype suggesting impairment in chloroplast development (113) (<http://rarge.psc.riken.jp/a/chloroplast/>). RH50 (At3g06980.1) is the *Arabidopsis* orthologue of the rice OsBIRH1, which exhibits RNA helicase activities *in vitro* and is involved in conferring plant resistance against various stresses (114).

Transcription and DNA-Binding

Proteins involved in transcription and other functions involving DNA-association were observed in the HM fractions, although at relatively low concentrations (1/10th of the ribosomal abundance) (Table 4.1). The plastid-encoded RNA polymerase (PEP) complex is composed of four subunits namely RpoA, RpoB, RpoC1 and RpoC2 (115) and is the predominant transcription complex in mature chloroplasts (116). The tobacco PEP complex in tobacco was affinity purified and was observed in a native gradient gel migrating at >900 kDa (17). In the current study, the PEP complex was observed to migrate at masses >2 MDa suggesting association with HM complexes,

most likely plastid DNA as part of the nucleoid (DNA-protein assembly). Indeed, PCR analysis confirmed the presence of plastid DNA, particularly in fractions with masses >3 MDa (Figure 4.2C).

Analysis of purified transcriptionally active chromosomes (TACs) from *Arabidopsis* chloroplasts revealed 35 proteins involved in plastid gene expression including the PEP complex and 18 proteins (pTACs 1 to 18) which contain RNA/DNA binding domains (117). In the current analysis, we found pTACs 2, 16 and 17 eluting with the PEP fraction. Moreover, a DNA gyrase (At3g10690.1), a thioredoxin protein (At3g06730.1) and a pfk-B type carbohydrate kinase (At1g69200.1), that were found in the TAC preparations, were also seen in the current study. In addition, several DNA-binding proteins that were not observed in the TAC analysis (117), clearly co-eluted with the PEP complex and pTAC proteins in the current study. These include two proteins involved in DNA damage repair, namely DNA mismatch repair MutS (At1g65070.1) and DNA repair protein RecA (At1g79050.1). A DNA topoisomerase (At4g31210.1) and a relatively high abundant DNA helicase (At45g35790.1) were observed peaking at 1-2 MDa. Finally, several proteins with DnaJ domains (At2g22360.1 and At4g39960.1) were found in fractions >2 MDa and the protein heat map suggested association with the nucleoid (Table 4.1).

Post-transcriptional events: RNA intron splicing, processing and editing

Proteins harboring single or multiple CRS1-YhbY domains, also called the chloroplast RNA splicing and ribosome maturation (CRM) domain, participate in the assembly of catalytic ribonucleoprotein complexes, namely group II intron particles and the 50S ribosomal subunit (118). In *Arabidopsis*, 16 proteins are predicted to have single or multiple CRM domains (118). CAF1 (At2g20020.1) has two CRM domains and is involved in group II intron RNA splicing (40). Aside from CAF1, two other CRM-

domain proteins were identified in this study. These include At4g39040.1 and At3g18390.1 which have one and three CRM domains, respectively. Two mutant alleles for the latter protein are embryo-lethal (EMB1865) with embryo development arrested at the globular stage, suggesting its importance in plastid development (88), (<http://www.seedgenes.org>).

In maize chloroplasts, WTF1 is found in group II intron ribonucleoprotein complexes (600–800 kDa) and cooperates with RNC1 in promoting group II intron splicing (119). At4g01037.1 and At4g37510.1, the respective *Arabidopsis* orthologues of maize WTF1 and RNC1, did not exhibit identical elution profiles in our SEC analysis but still indicated interactions with ribonucleoprotein complexes, with WTF1 eluting at 1-2MDa and RNC1 at 3-5 MDa (Table 4.1). In addition, a chloroplast-encoded maturase K (Atcg00040.1) was observed at the 1-2 MDa region in this study (Table 4.1). The RNA-binding properties and function of maturase K in other plants have been studied (120, 121).

About 450 pentatricopeptide repeat (PPR) proteins are predicted in *Arabidopsis* and they are recognized as major mediators of organelle gene expression although their specific functions remain poorly characterized (122). Eight PPRs were observed in the HM fractions (Table 4.1). In maize, the PPR proteins PPR4 and CRP1 are required for the translation of several chloroplast-encoded transcripts (123-126). The *Arabidopsis* orthologues of PPR4 (At5g04810.1) and CRP1 (At5g432310.1) eluted at 3-5 MDa and 1-3 MDa, respectively. In addition, the RNA-binding P67 (At4g16390.1) (127) was found at fractions >2 MDa (Table 4.1).

Three proteins with S1 RNA-binding domains (At1g17200.1, At3g23700.1 and At1g1280.1) were found eluting at fractions >1MDa, with At3g23700.1 being 10-fold more abundant than the other two. Plants deficient in At1g7200.1 expression exhibit pale-green cotyledons and leaves (113), <http://rarge.psc.riken.jp/a/chloroplast/>)

indicating a role in chloroplast development. Another RNA binding protein, HCF173 (At1g16720.1) was found in the 1-3 MDa fractions consistent with its detection in mRNA-containing HM complexes (128).

Chloroplast ribonucleoproteins (cpRNPs) are highly abundant proteins that associate with various RNA species for RNA processing and/or stabilization (129). In tobacco, cp29A, cp29B and cp33 were mostly found as non-ribosome-bound ribonucleoprotein complexes but were also detected in fractions >600 kDa (129). In this study, cp29B (At2g37220.1) and two cp33 proteins (At1g01080.1 and At2g35410.1) were observed in the 0.8-2 MDa range with similar elution profiles (Table 4.1). Several cpRNPs are involved in RNA editing (130, 131) although this has not yet been demonstrated for the cpRNPs identified in this study.

Lipid Metabolism

Aside from ribosomes and ribosome-associated proteins, the other function that is enriched in the HM dataset is fatty acid synthesis. Two hetero-oligomeric complexes, namely PDC and ACCase, were observed (see Figure 4.3; Table 4.1).

The PDC is composed of multiple copies of three enzymes. The E1 component is a pyruvate dehydrogenase (consisting of α - and β - subunits), E2 is a dihydrolipoamide acetyltransferase and E3 is a dihydrolipoamide dehydrogenase. PDCs form large complexes composed of a core complex of eight trimers (cube) or 20 trimers (pentagonal dodecahedron) of E2 with E1, with E3 promoting substrate channeling across the three enzyme components (132). The maize mitochondrial PDC was found to have an estimated mass of about 8-9 MDa (18) due in part to the 2.7 MDa E2 core (37). The plastid-localized PDC from pea chloroplasts dissociates rapidly *in vitro* making the estimation of its organization and composition difficult (133).

The PDC components eluted at a mass range >5 MDa (Figure 4.3 and Table 4.1). All of the subunits of the plastid-localized PDC are nuclear-encoded and all, except the E1 α component (At1g01090.1), are encoded by more than one gene. An E1 β -subunit (At1g3020.1) has been previously characterized (134) and was observed in this experiment with unique peptides. LTA2 (At3g25860.1) is an E2 component and has been shown to exhibit dihydrolipoamide acetyltransferase activity (135). The T-DNA insertion mutant for LTA2 is embryo lethal (136). Another E2 subunit (At1g34330.1) was found in the genome database (132) and a T-DNA insertion mutant (EMB3003) for this gene is also embryo lethal (88), <http://www.seedgenes.org>). Interestingly, two elution peaks were observed for E2—one at >5MDa and another at 1-2MDa (see Table 4.1) suggesting different oligomeric states for the E2 core with or without bound E1 and E3. The two plastidic E3 isoforms (At3g16950.1 or ptlpd1 and At4g16155.1 or ptlpd2) are 85% identical and were previously characterized (137, 138). The presence of different paralogues for the three central components suggests that the plastid PDC population is heterogeneous; further experimentation is needed to determine the biological significance.

The plastid-localized heteromeric ACCase catalyzes the first committed step in *de novo* fatty acid synthesis which occurs solely in plastids. ACCase is composed of biotin carboxylase (BC), biotin containing carboxylase protein (BCCP), alpha-carboxyltransferase (alpha-CT) and beta-carboxyltransferase (beta-CT) subunits. The pea chloroplast ACCase was found to elute at about 650-700 kDa in gel filtration analysis (19). In this study, the *Arabidopsis* ACCase subunits were all found in the 1-2 MDa range (Table 4.1). These include BC (At5g35360.1) (139), alpha-CT (At2g38040.1) and the chloroplast-encoded beta-CT (Atcg00500.1) as well as four BCCPs that were identified with distinct peptides. BCCP1 (At5g16390.1) has been well-characterized (140, 141), The other three BCCPs namely At1g52670.1,

At3g56130.1 and At3g15690.1 have similar molecular weights (~25 kDa), but domain analysis revealed that they are missing the critical lysine residue for biotin attachment (data not shown). Nevertheless, the observation that these BCCPs elute at the same size range as the ACCase suggests that they associate with the ACCase. This is further supported by the isolation of these BCCPs together with BCCP1 from PII-affinity chromatography where PII is a signaling protein that modulates ACCase activity (142). Nevertheless, the functional roles of these BCCPs remain to be characterized.

Additional proteins involved in the fatty acid biosynthesis pathway co-eluted with ACCase. KAS1 (At5g46290.1), an essential enzyme involved in the construction of unsaturated fatty acid carbon skeletons and acetyl-CoA synthetase (acetate-CoA ligase) (At5g36880.1) were also found at the HM fractions peaking at 3-5MDa and 1-2MDa, respectively (Table 4.1). Acyl carrier proteins (ACPs) carry the acyl chains during the synthesis of 16- and 18-carbon fatty acids. Several ACP isoforms are found in *Arabidopsis* and are expressed in a tissue-specific manner (143). ACP4 (At4g25050.1), the most abundant and most leaf-specific isoform (143), was the one observed in this study, and peaked at 0.8-1 MDa.

DISCUSSION

High molecular mass protein-protein and protein-nucleotide complexes and expanded proteome coverage

Protein-protein and protein-nucleotide interactions play a crucial role in orchestrating biological processes. Using gel filtration-based size fractionation, this study provides an overview of soluble chloroplast-localized assemblies between 0.8 and ~5 MDa, the 'high mass' (HM) range, representing about 10-13% of the mass of the stromal proteome. This analysis complemented our previous, gel-based analysis that resolved stromal complexes upto ~800 kDa (11). When excluding the abundant Calvin cycle

components, the HM range was dominated in biomass by 30S-50S-70S ribosome particles and associated factors, as well as the multi-functional enzymes complexes, PDC and ACCase involved in fatty acid metabolism. Furthermore, the plastid chromosome with interacting proteins separated from the bulk of ribosomes, and was associated with a specific set of DNA-binding proteins, dominated by the heteromeric plastid DNA polymerase, PEP.

The distribution of proteins across the mass range could be quantified based on the number of adjusted SPC, using appropriate normalizations. Hierarchical clustering of the dataset effectively grouped the proteins into biologically related functions and complexes, indicating that the complexes were stable during the gel filtration runs, and further benchmarking against known protein assemblies demonstrated that the clustering yielded biologically meaningful associations. Therefore, we could assign lesser or unknown proteins to various complexes, even if targeted validation by *e.g.* co-immunoprecipitations will ultimately be needed.

A second objective was to improve coverage of the chloroplast proteome and find proteins in underrepresented functional classes. Indeed, when comparing to previous chloroplast proteome analyses, we identified several low abundance proteins involved in RNA metabolism and ribosome assembly, mostly in the HM fractions. This indicates that many of the proteins involved in plastid gene expression are associated with large RNA containing assemblies (ribosome, RNA splice complexes) and the plastid chromosome. Further targeted analysis of these nucleotide-protein complexes is likely to reveal additional proteins. Affinity purifications that target specific binding domains (*e.g.* metals, ATP and other cofactors; see *e.g.* (144-146) using highly purified chloroplast preparations, will be needed to further improve proteome coverage.

Metabolic channeling in fatty acid metabolism

Metabolic functions were strongly underrepresented in the HM fraction, but were otherwise dominated by fatty acid metabolism in terms of protein biomass, in particular the ≥ 5 MDa PDC and the 0.8-1 MDa heteromeric ACCase. We suggest that this bias towards fatty acid metabolism, relates to the mixed hydrophobic and hydrophilic nature of the substrates and complexity of the co-factors and reactions. This complexity requires enclosure of the different intermediates within these MDa complexes, also designated as metabolic channeling (147, 148).

DNA binding – the chloroplast chromatin

Chloroplast DNA is packaged into nucleoids (organellar nuclei) consisting of multiple copies of the plastid genome complexed with proteins that are minimally characterized, even if a number of them were identified from purified nucleoids or TACs (117, 149, 150). The majority of the nucleoid is tightly anchored to plastid membranes and requires detergent treatment, differential centrifugation and possibly size exclusion column chromatography or co-immunoprecipitations for purification (117, 151, 152). In contrast, our current study did not involve detergent treatments and the extracted stromal proteome had negligible membrane contamination. Nevertheless, plastid DNA was clearly present in fractions >2 MDa and we identified the four subunits of the PEP complex and a dozen DNA binding proteins (e.g. involved in DNA repair, DNA organization), some of which were observed previously (117). However, we clearly did not observe the relatively abundant membrane-associated MFP1 protein (At3g16000) (easily observed in thylakoid proteome analysis with high protein MOWSE scores; e.g. (153); <http://ppdb.tc.cornell.edu/>) that anchors the plastid DNA to the membrane (154). The DNA and associated proteins peaked in higher mass assemblies than the 70S ribosomes, although there was some overlap. A more

extensive proteome analysis of the membrane-associated nucleoid and follow-up functional analysis of the associated proteins, that also addresses to what extent transcription and translation are coupled, is overdue.

The RNA processing, splice and degradation machinery

Most of the plastid-encoded genes in higher plants are organized as operons, which are generally transcribed as polycistronic transcriptional units (155, 156). The resulting primary transcripts are modified to generate functional RNAs by RNA cleavage of pre-existing RNAs, RNA stabilization and degradation, intron splicing and RNA editing. In addition to known factors involved in RNA metabolism, we discovered several new proteins in the HM range that, based on their functional domains (eg CRM, PPR) are likely to be involved in RNA metabolism. Group II intron ribonucleoprotein complexes were found enriched at size ranges of about 500-800 kDa (119, 157), even if they can also be found in higher mass ranges. We identified *Arabidopsis* homologues for most known group II intron splicing proteins and most of them migrate at 1-2 MDa, but were also found at higher mass ranges, suggesting a possible coupling between transcription and RNA processing. Furthermore these RNA interacting proteins also form associations with RNA-containing high molecular weight assemblies such as ribosomes and PEP or other RNA processing complexes. More detailed biochemical analysis, including protein-RNA interactions including using RIP-Chip analysis (158), as well as targeted affinity purifications combined with protein mass spectrometry, will be required to fully understand the organization of RNA processing and how it interfaces with transcription, as well as protein translation and protein assembly.

Ribosomes and Ribosome biogenesis

The most abundant proteins in the HM dataset were ribosomal proteins and indeed, the mass spectrometry analysis identified nearly all known or predicted ribosomal proteins, as well as most known and many potential ribosome-associated proteins. Interestingly, several of the ribosomal proteins were represented by multiple gene products; in some cases one gene was plastid-encoded whereas the other was nuclear-encoded, in other cases both homologues were plastid-encoded but differed greatly in protein abundance. The resulting ribosome heterogeneity is likely to have functional consequences, and may represent adaptation to particular developmental states, cell types or stresses; these observations warrant further investigation. Functional assignments of ribosome-associated proteins include translation, co-translational modifications and ribosome biogenesis, indicating that SEC fractionated ribosomes in various assembly and functional states (Figure 4.4).

The identification of 12 *Arabidopsis* orthologues of bacterial ribosome assembly factors in the HM-A fractions suggests that ribosome assembly in chloroplasts resembles that of the prokaryotic system. Most of these factors exhibited similar low accumulation levels (50 to 100-fold lower than the average ribosome abundance). So far, only two chloroplast-localized RA-GTPases in *Arabidopsis* were characterized namely AtObgC and the YqeH orthologue AtNOA. Analysis of their SEC elution profiles in this study established that AtObgC associates with 30S and 50S and that AtNOA1 interacts with 30S and 70S ribosomes consistent with bacterial studies (Figure 4.4). Most bacterial RA-GTPases are essential to bacterial viability (47, 48, 72). Similarly, functional analysis of AtObgC and AtNOA1 mutants has revealed that both proteins are essential for plastome synthesis and chloroplast development (76, 77, 82, 83). Targeted biochemical and genetic analyses of other *Arabidopsis* orthologues of bacterial ribosome biogenesis proteins including RA-

GTPases (EngA, Era, Hflx, YqeH) and other factors (RimM, RbfA, RsmD, YrdC, SpoU), as well as plant-specific maturation factors (DAL and IOJAP) will provide insights on their specific roles in chloroplast ribosome biogenesis and the consequences for plastid protein homeostasis. The proteome analysis in this study has opened up this challenging topic for further investigation in *Arabidopsis*.

Proteins with CRM domains have been shown to participate in the assembly of catalytic ribonucleoprotein complexes, namely group II intron particles and the 50S ribosomal subunit (118). So far, all the characterized CRM domain proteins in plants (CRS1, CAF1 and CAF2) associate with RNA *in vivo* and are involved in group II intron splicing (159-162). Here we also identified a CRM domain protein (At4g39040.1) migrating at 1-2 MDa (50S ribosomes) and is potentially a ribosome maturation factor, because its *E. coli* homologue Yhby was found to tightly associate with pre-50S ribosomes harboring immature 23S rRNAs (118).

The *Arabidopsis* CSP41B and PNPase have multiple functions in plastid gene expression including 23S rRNA maturation (99-104). Their SEC elution profiles show that they elute mainly in the 1-2 MDa fraction suggesting interactions with the 50S particle although this does not rule out associations with other RNA-containing complexes. Another endonuclease, RNase J, is found to be crucial for 16S rRNA processing in *B. subtilis* (107). The elution profile of its *Arabidopsis* orthologue (>1 MDa) indicates interactions with various ribonucleoprotein complexes including 30S and 70S particles consistent with bacterial studies (107). Overall, these findings support the role of CSP41B, PNPase and RNase J in ribosome assembly and maturation. RNA helicases have been implicated in various RNA processing functions including rRNA maturation during ribosome biogenesis (109). Four *Arabidopsis* DEAD-BOX RNA helicases were seen in the fractions at mass ranges > 1MDa and should be considered potential candidates for ribosome biogenesis. Interestingly, RH3

is specifically up-regulated in *Arabidopsis clpr2-1* mutants with reduced plastid-localized ClpPR protease levels and exhibits a delay in ribosome assembly and/or defects in RNA metabolism (30).

Translation and Co-translational modifications

The isolation of ribosomes, together with translational factors and proteins involved in post-translational modifications, suggests that translating ribosomes were captured. Our analysis clearly provided support for ribosome association of various N-terminal modifying enzymes (MAP, PDF), as well as targeting/folding proteins (TF and cpSR54). The *E. coli* TF has been shown to provide a folding cavity for the nascent protein emerging from the ribosome tunnel (163-165). We did not find any obvious growth phenotype for a TF null mutant in *Arabidopsis* (Rutschow and van Wijk, unpublished), which is perhaps not surprising since a clear phenotype in *E. coli* is only seen in combination with deletion of the DnaK chaperone (166). It has been shown for *E. coli* that during protein synthesis, in particular the RPL23 protein serves as a platform for the association of enzymes, targeting factors and chaperones that act upon the nascent polypeptide that emerges from the exit tunnel (57). Interestingly, *Arabidopsis* has two identical chloroplast-encoded L23 proteins and future studies should determine how these two paralogues contribute to protein synthesis and homeostasis.

MATERIALS AND METHODS

Plant growth, chloroplast stroma proteome isolation and size fractionation *A. thaliana* (Col 0) was grown under 10-h light/14-h dark cycles at 25/17 °C in controlled growth chambers (Conviron) for about 55 days and leaves were collected from mature rosettes, about one week prior to bolting. Leaves 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 plastids were then collected by a 2-min spin at $1100 \times g$ and further purified on 40-85% 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 $MgCl_2$ with a mixture of protease inhibitors under mild mechanical disruption. The lysate was then subjected to ultracentrifugation (100,000 $\times g$) to pellet the membrane components. The supernatant (stroma) was then collected and concentrated using Amicon 4, 10 kDa MWCO (Millipore). Protein amounts were determined using the Bradford reagent (Bio-rad) or the BCA protein assay kit (ThermoScientific).

One to three mg of stroma was loaded in a Superose 6 10/300 GL column (GE Healthcare Biosciences) using an AKTA FPLC system (Amersham Biosciences) with an Hg lamp as a detector. Absorbance was measured at 280 nm. Elution was performed with buffer A (25 mM HEPES pH 8.0, 10 mM NaCl, 10 mM $MgCl_2$) or buffer B (50 mM HEPES, 50 mM NaCl, pH 8.0, 5 mM $MgCl_2$) at an optimal flow rate of 0.25 mL/min. 300- μ L subfractions were initially collected. Subfractions were pooled as follows: three for fractions 1-6, four for fractions 7-12 and six for fraction 13 (see Figure 4.2 for the SEC chromatogram and the fraction designations). Proteins from pooled fractions were either concentrated using Amicon Microcon YM-10 (Millipore) or precipitated with 80% acetone, and were then separated further using SDS-PAGE on 12%T Laemmli or 12%T Tricine mini-gels. Protein bands were visualized using the fluorescent Sypro Ruby for fractions 1-5 of SEC-separated sample from buffer A and the rest of the gels were stained with Coomassie Blue. Each gel lane was excised into four or five bands followed by reduction, alkylation, in-gel

digestion with trypsin and peptide extraction, as described in (27). Peptide extracts were dried down and resuspended in 15–20 μ l 5% formic acid for MS/MS analysis.

nLC-LTQ Orbitrap analysis and data processing The resuspended peptide extracts were analyzed by data-dependent MS/MS using an on-line LC-LTQ-Orbitrap (Thermo Electron Corp.). Peptide samples were automatically loaded on a guard column (LC Packings MGU-30-C18PM) via an autosampler followed by separation on a PepMap C₁₈ reverse-phase nanocolumn (LC Packings nan75-15-03-C18PM) using 90-min gradients with 95% water, 5% ACN, 0.1% FA (solvent A) and 95% ACN, 5% water, 0.1% FA (solvent B) at a flow rate of 200 nl/min. Two blanks were run after every sample (see Zybaylov et. al., 2008 for the gradient and sample injection scheme). The acquisition cycle consisted of a survey MS scan in the Orbitrap with a set mass range from 350 to 1800 m/z at the highest resolving power (100,000) followed by five data-dependent MS/MS scans acquired in the LTQ. Dynamic exclusion was used with the following parameters: exclusion size, 500; repeat count, 2; repeat duration, 30 s; exclusion time, 180 s; exclusion window, ± 6 ppm or ± 100 ppm. Target values were set at 5×10^5 and 10^4 for the survey and tandem MS scans, respectively. Regular scans were used both for the precursor and tandem MS with no averaging.

Peak lists (.mgf format) were generated using DTA supercharge (v1.19) software (<http://msquant.sourceforge.net/>) and searched with Mascot v2.2 (Matrix Science). For off-line calibration, first a preliminary search was conducted with the precursor tolerance window set at ± 30 ppm. Peptides with the ion scores above 33 were chosen as benchmarks to determine the offset for each LC-MS/MS run. This particular ion score value (33) was chosen in accordance with the results of the search against the target-decoy database – see further below and (28). This offset was then applied to adjust precursor masses in the peak lists of the respective .mgf file for

recalibration using a Perl script (unpublished B. Zybailov). The recalibrated peak lists were searched against the TAIR ATH database v8, including sequences for known contaminants (*e.g.* keratin, trypsin) (total 33013 entries) with or without a concatenated with decoy database where all the sequences were in reverse orientation. Each of the peak lists were searched using Mascot v2.2 (maximum p-value of 0.01) for full tryptic peptides using a precursor ion tolerance window set at ± 6 ppm, variable methionine oxidation and fixed cysteine carbamido-methylation, and a minimal ion score threshold of 33 and mass range of 700-3500 Da for precursor ions. To reduce the false identification rate of proteins identified by one peptide, the Mascot search results were further filtered as follows: ion score threshold was increased to 35, and mass accuracy on the precursor ion was required to be within ± 3 ppm. Overall, this yielded a peptide false discovery rate (FDR) of 1.5%, with peptide FPR calculated as: $2 * (\text{decoy_hits}) / \text{total_hits}$ derived from searches against the target-decoy database. The false protein identification rate of protein identified with 2 or more peptides was zero. All filtered results were uploaded into the Plant Proteomics DataBase, PPDB (<http://ppdb.tc.cornell.edu/>) (29). All mass spectral data (the mgf files reformatted as PRIDE XML files) were made available via the Proteomics Identifications database (PRIDE) at <http://www.ebi.ac.uk/pride/>.

Determination of protein SEC elution profiles and relative protein abundance

To determine the relative protein abundance distribution by spectral counting, the number of matched MS/MS spectra or spectral count (SPC) for each protein was obtained. This was further classified as total SPC, unique SPC (uniquely matching to an accession), and adjusted SPC (adjSPC). The latter is the sum of unique SPCs and SPCs from shared peptides across accessions with SPC distributed in proportion to their unique SPCs, if applicable. Proteins that shared more than 80% of their matched

peptides with other proteins across the complete data set were grouped into families. For many *Arabidopsis* genes, more than one protein model is predicted. In this study, protein models with the highest total adjSPC across all experiments were used; if the protein models did not differ in total adjSPC, protein model 1 was selected. To increase the robustness and significance of the data set, we removed all proteins that were identified with only one amino acid sequence, irrespective of charge state, post-translational modifications or number of SPCs. Proteins that were quantified with 2 or less adjSPCs were also removed. To generate protein elution heat maps for the HM dataset, the AdjSPC for each protein per SEC fraction was normalized to the highest AdjSPC across five HM fractions (NAdjSPC).

To calculate the relative abundance for each protein across all HM fractions, the total adjSPC was divided by the number of observable tryptic peptides within the mass range 700-3500 Da (with the predicted transit peptide removed) yielding the spectral abundance factor (SAF). The SAF values were then normalized to the total SAF in the whole dataset yielding normalized SAFs (NSAFs).

Hierarchical clustering analysis To group proteins with similar elution profiles in the HM dataset, hierarchical clustering was employed using the Statistics toolbox of MATLAB version 7 (Mathworks, Inc.). The linear correlation (ρ) between every pair of proteins with NAdjSPC distribution across the SEC HM fractionation range: X_1, \dots, X_n and Y_1, \dots, Y_n where $n=5$

$$\rho_{XY} = \frac{\sum_{j=1}^n (X_j - \bar{X})(Y_j - \bar{Y})}{\sqrt{\sum_{j=1}^n (X_j - \bar{X})^2 \sum_{j=1}^n (Y_j - \bar{Y})^2}}$$

was derived. This was then converted into a distance measure $\Delta_{XY} = 1 - \rho_{XY}$. Protein pairs with similar elution profiles have higher correlations and in turn, have smaller distance values. A linkage map based on the average distance among protein pairs was then constructed to yield a hierarchical cluster tree (dendogram).

Nucleic acid extraction and subsequent detection of DNA and ribosomal RNA from SEC fractions DNA and RNA were isolated from SEC fractions 1-7 by phenol/chloroform extraction followed by ethanol precipitation. Briefly, an aliquot from each fraction was combined with phenol/chloroform/isoamyl alcohol in a 50:50 mixture together with 0.3% SDS, 1.5 mM EDTA and 20 ng/uL Glycoblue (Ambion). SDS and EDTA were added to dissociate proteins from nucleoprotein complexes. Glycoblue enhances nucleic acid recovery and increases visibility of the sample pellet. The aqueous phase was extracted after phase separation. To increase nucleic acid yield, the organic phase was then back-extracted with TESS buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 100 mM NaCl, 0.2% SDS). The collected aqueous phase was then combined with the initial extraction. After adjustment of salt concentration with 3 M sodium acetate, nucleic acid was precipitated with 70% ethanol, pelleted, dried and resuspended in TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA).

An aliquot of the extracted DNA/RNA was treated with Ribonuclease H (Invitrogen) to degrade RNA from RNA:DNA hybrids and plastid DNA was probed by PCR amplification of the gene for 16S rRNA (see (30) for primers used). PCR samples were then separated on agarose gels and visualized by ethidium bromide staining. The presence of 16S and 23S rRNA in the extracted DNA/RNA were determined by Northern blot analysis through hybridization with digoxigenin (DIG)-11-dUTP-labeled rRNA probes and subsequent detection with anti-DIG antibodies, as previously described (30).

The Plant Proteomics Database Mass spectrometry-based information of all identified proteins was extracted from the Mascot search pages and filtered for significance (*e.g.* minimum ion scores, etc.), ambiguities, and shared spectra as previously described. This information includes MOWSE scores, number of matching peptides, total SPCs, unique SPCs, adjSPCs, highest peptide score, highest peptide error (in ppm), lowest absolute error (ppm), sequence coverage, and tryptic peptide sequences (29). All these are available in the PPDB by using the search function "Proteome Experiments" and by selecting the desired output parameters. Alternatively information for specific accessions (either individually or a group) can be extracted using the search function "Accessions," and if desired, this search can be limited to specific experiments. Finally information for a particular accession can also be found on each "protein report page." The MapMan bin system (31) was used for functional assignment, and proteins were reassigned to other bins if needed.

ACKNOWLEDGEMENTS

This work was supported by the grants from the National Science Foundation (MCB 0343444), the United States Department of Agriculture (Biochemistry program; #2003-02409) and the New York State Office of Science, Technology and Research (NYSTAR) to KJVW. Part of this work was carried out by using the resources of the Computational Biology Service Unit of Cornell University which is partially funded by the Microsoft Corporation. We thank Dr Qi Sun for his valuable support. We thank several other members of the van Wijk lab, Boris Zybaïlov, Giulia Friso, Wojciech Majeran, Yuka Asakura and Heidi Rutschow, for support and discussions.

REFERENCES

1. Buchanan, B., Gruissem, W. & Jones, R. L. (2000) (American Society of Plant Physiologists, Rockville).
2. Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. (2000) *J Mol Biol* 300, 1005-16.
3. van Wijk, K. J. (2004) *Plant Physiol Biochem* 42, 963-77.
4. Jarvis, P. & Robinson, C. (2004) *Curr Biol* 14, R1064-77.
5. Soll, J. & Schleiff, E. (2004) *Nat Rev Mol Cell Biol* 5, 198-208.
6. Kessler, F. & Schnell, D. J. (2006) *Traffic* 7, 248-57.
7. Ferro, M., Salvi, D., Brugiere, S., Miras, S., Kowalski, S., Louwagie, M., Garin, J., Joyard, J. & Rolland, N. (2003) *Mol Cell Proteomics* 28, 28.
8. Kieselbach, T. & Schroder, W. P. (2003) *Photosynth Res* 78, 249-64.
9. Zybaïlov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O., Sun, Q. & van Wijk, K. J. (2008) *PLoS ONE* 3, e1994.
10. Baginsky, S., Kleffmann, T., von Zychlinski, A. & Gruissem, W. (2005) *J Proteome Res* 4, 637-40.
11. Peltier, J. B., Cai, Y., Sun, Q., Zabrouskov, V., Giacomelli, L., Rudella, A., Ytterberg, A. J., Rutschow, H. & van Wijk, K. J. (2006) *Mol Cell Proteomics* 5, 114-133.
12. Yamaguchi, K., von Knoblauch, K. & Subramanian, A. R. (2000) *J Biol Chem* 275, 28455-65.
13. Yamaguchi, K. & Subramanian, A. R. (2000) *J Biol Chem* 275, 28466-82.
14. Yamaguchi, K. & Subramanian, A. R. (2003) *Eur J Biochem* 270, 190-205.
15. Zanetti, M. E., Chang, I. F., Gong, F., Galbraith, D. W. & Bailey-Serres, J. (2005) *Plant Physiol* 138, 624-35.

16. Carroll, A. J., Heazlewood, J. L., Ito, J. & Millar, A. H. (2008) *Mol Cell Proteomics* 7, 347-69.
17. Suzuki, J. Y., Jimmy Ytterberg, A., Beardslee, T. A., Allison, L. A., Wijk, K. J. & Maliga, P. (2004) *Plant J* 40, 164-72.
18. Thelen, J. J., Miernyk, J. A. & Randall, D. D. (1998) *Plant Physiol* 116, 1443-50.
19. Sasaki, Y., Sekiguchi, K., Nagano, Y. & Matsuno, R. (1993) *FEBS Lett* 316, 93-8.
20. Phinney, B. S. & Thelen, J. J. (2005) *J Proteome Res* 4, 497-506.
21. Hu, Q., Noll, R. J., Li, H., Makarov, A., Hardman, M. & Graham Cooks, R. (2005) *J Mass Spectrom* 40, 430-43.
22. Scigelova, M. & Makarov, A. (2006) *Proteomics* 6 Suppl 2, 16-21.
23. Dunkley, T. P., Watson, R., Griffin, J. L., Dupree, P. & Lilley, K. S. (2004) *Mol Cell Proteomics* 3, 1128-34.
24. Dong, M., Yang, L. L., Williams, K., Fisher, S. J., Hall, S. C., Biggin, M. D., Jin, J. & Witkowska, H. E. (2008) *J Proteome Res* 7, 1836-49.
25. Andersen, J. S., Wilkinson, C. J., Mayor, T., Mortensen, P., Nigg, E. A. & Mann, M. (2003) *Nature* 426, 570-4.
26. Majeran, W., Zybilov, B., Ytterberg, A. J., Dunsmore, J., Sun, Q. & van Wijk, K. J. (2008) *Mol Cell Proteomics* 7, 1609-38.
27. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. & Mann, M. (2006) *Nat Protoc* 1, 2856-60.
28. Zybilov, B., Sun, Q. & van Wijk, K. J. (2009) *Anal Chem* 81, 8015-24.
29. Sun, Q., Zybilov, B., Majeran, W., Friso, G., Olinares, P. D. & van Wijk, K. J. (2009) *Nucleic Acids Res* 37, D969-74.

30. Zybaylov, B., Friso, G., Kim, J., Rudella, A., Rodriguez, V. R., Asakura, Y., Sun, Q. & van Wijk, K. J. (2009) *Mol Cell Proteomics* 8, 1789-1810.
31. Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., Selbig, J., Muller, L. A., Rhee, S. Y. & Stitt, M. (2004) *Plant J* 37, 914-39.
32. Liu, H., Sadygov, R. G. & Yates, J. R., 3rd (2004) *Anal Chem* 76, 4193-201.
33. Zybaylov, B., Coleman, M. K., Florens, L. & Washburn, M. P. (2005) *Anal Chem* 77, 6218-24.
34. Old, W. M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K. G., Mendoza, A., Sevinsky, J. R., Resing, K. A. & Ahn, N. G. (2005) *Mol Cell Proteomics* 4, 1487-502.
35. Lu, P., Vogel, C., Wang, R., Yao, X. & Marcotte, E. M. (2007) *Nat Biotechnol* 25, 117-24.
36. Zybaylov, B., Mosley, A. L., Sardi, M. E., Coleman, M. K., Florens, L. & Washburn, M. P. (2006) *J Proteome Res* 5, 2339-47.
37. Thelen, J. J., Muszynski, M. G., David, N. R., Luethy, M. H., Elthon, T. E., Miernyk, J. A. & Randall, D. D. (1999) *J Biol Chem* 274, 21769-75.
38. Sharma, M. R., Wilson, D. N., Datta, P. P., Barat, C., Schluenzen, F., Fucini, P. & Agrawal, R. K. (2007) *Proc Natl Acad Sci U S A* 104, 19315-20.
39. Baligni, M. V., Yamaguchi, K. & Mayfield, S. P. (2004) *Plant Cell* 16, 3357-69.
40. Asakura, Y. & Barkan, A. (2006) *Plant Physiol* 142, 1656-63.
41. Ueda, T., Matsuda, N., Anai, T., Tsukaya, H., Uchimiya, H. & Nakano, A. (1996) *Plant Cell* 8, 2079-2091.
42. Liljas, A. (2004) *Structural aspects of protein synthesis* (World Scientific, Singapore; Hackensack, N.J.).
43. Ramakrishnan, V. (2002) *Cell* 108, 557-572.

44. Albrecht, V., Ingenfeld, A. & Apel, K. (2006) *Plant Mol Biol* 60, 507-18.
45. Bhadula, S. K., Elthon, T. E., Habben, J. E., Helentjaris, T. G., Jiao, S. & Ristic, Z. (2001) *Planta* 212, 359-66.
46. Rao, D., Momcilovic, I., Kobayashi, S., Callegari, E. & Ristic, Z. (2004) *Eur J Biochem* 271, 3684-92.
47. Caldon, C. E., Yoong, P. & March, P. E. (2001) *Molecular Microbiology* 41, 289-297.
48. Caldon, C. E. & March, P. E. (2003) *Current Opinion in Microbiology* 6, 135-139.
49. Owens, R. M., Pritchard, G., Skipp, P., Hodey, M., Connell, S. R., Nierhaus, K. H. & O'Connor, C. D. (2004) *Embo J* 23, 3375-85.
50. Qin, Y., Polacek, N., Vesper, O., Staub, E., Einfeldt, E., Wilson, D. N. & Nierhaus, K. H. (2006) *Cell* 127, 721-733.
51. Pfennig, P. L. & Flower, A. M. (2001) *Molecular Genetics and Genomics* 266, 313-317.
52. Beckering, C. L., Steil, L., Weber, M. H. W., Volker, U. & Marahiel, M. A. (2002) *J. Bacteriol.* 184, 6395-6402.
53. Eymann, C., Homuth, G., Scharf, C. & Hecker, M. (2002) *J. Bacteriol.* 184, 2500-2520.
54. Krishnan, K. & Flower, A. M. (2008) *Journal of Bacteriology* 190, 7675-7683.
55. Wang, F., Zhong, N. Q., Gao, P., Wang, G. L., Wang, H. Y. & Xia, G. X. (2008) *Plant Cell Environ* 31, 982-94.
56. Sharma, M. R., Dönhöfer, A., Barat, C., Marquez, V., Datta, P. P., Fucini, P., Wilson, D. N. & Agrawal, R. K. (2009) *Journal of Biological Chemistry* 285, 4006-14.

57. Kramer, G., Boehringer, D., Ban, N. & Bukau, B. (2009) *Nat Struct Mol Biol* 16, 589-97.
58. Serero, A., Giglione, C. & Meinnel, T. (2001) *J Mol Biol* 314, 695-708.
59. Giglione, C., Serero, A., Pierre, M., Boisson, B. & Meinnel, T. (2000) *Embo J* 19, 5916-29.
60. Dirk, L. M., Williams, M. A. & Houtz, R. L. (2001) *Plant Physiol* 127, 97-107.
61. Franklin, A. E. & Hoffman, N. E. (1993) *J Biol Chem* 268, 22175-80.
62. Li, X., Henry, R., Yuan, J., Cline, K. & Hoffman, N. E. (1995) *Proc Natl Acad Sci U S A* 92, 3789-93.
63. Nilsson, R., Brunner, J., Hoffman, N. E. & van Wijk, K. J. (1999) *Embo J* 18, 733-42.
64. Rutschow, H., Ytterberg, A. J., Friso, G., Nilsson, R. & van Wijk, K. J. (2008) *Plant Physiol* 148, 156-75.
65. Ferbitz, L., Maier, T., Patzelt, H., Bukau, B., Deuerling, E. & Ban, N. (2004) *Nature* 431, 590-6.
66. Nishio, K., Hirohashi, T. & Nakai, M. (1999) *Biochem Biophys Res Commun* 266, 584-7.
67. Sharkia, R., Bonshtien, A. L., Mizrahi, I., Weiss, C., Niv, A., Lustig, A., Viitanen, P. V. & Azem, A. (2003) *Biochim Biophys Acta* 1651, 76-84.
68. Schroda, M., Vallon, O., Whitelegge, J. P., Beck, C. F. & Wollman, F. A. (2001) *Plant Cell* 13, 2823-39.
69. Kaczanowska, M. & Ryden-Aulin, M. (2007) *Microbiology and Molecular Biology Reviews* 71, 477-+.
70. Connolly, K. & Culver, G. (2009) *Trends in Biochemical Sciences* 34, 256-263.

71. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res* 25, 3389-402.
72. Britton, R. A. (2009) *Annual Review of Microbiology* 63, 155-176.
73. Scott, J. M. & Haldenwang, W. G. (1999) *Journal of Bacteriology* 181, 4653-4660.
74. Lin, B., Thayer, D. A. & Maddock, J. R. (2004) *Journal of Bacteriology* 186, 481-489.
75. Sato, A., Kobayashi, G., Hayashi, H., Yoshida, H., Wada, A., Maeda, M., Hiraga, S., Takeyasu, K. & Wada, C. (2005) *Genes to Cells* 10, 393-408.
76. Bang, W. Y., Hata, A., Jeong, I. S., Umeda, T., Masuda, T., Chen, J., Yoko, I., Suwastika, I. N., Kim, D. W., Im, C. H., Lee, B. H., Lee, Y., Lee, K. W., Shiina, T. & Bahk, J. D. (2009) *Plant Molecular Biology* 71, 379-390.
77. Chigri, F., Sippel, C., Kolb, M. & Vothknecht, U. C. (2009) *Molecular Plant* 2, 1373-1383.
78. Jain, N., Dhimole, N., Khan, A. R., De, D., Tomar, S. K., Sajish, M., Dutta, D., Parrack, P. & Prakash, B. (2009) *Biochemical and Biophysical Research Communications* 379, 201-205.
79. Shields, M. J., Fischer, J. J. & Wieden, H.-J. (2009) *Biochemistry* 48, 10793-10802.
80. Koller-Eichhorn, R., Marquardt, T., Gail, R., Wittinghofer, A., Kostrewa, D., Kutay, U. & Kambach, C. (2007) *J Biol Chem* 282, 19928-37.
81. Anand, B., Surana, P., Bhogaraju, S., Pahari, S. & Prakash, B. (2009) *Biochemical and Biophysical Research Communications* 386, 602-606.
82. Moreau, M., Lee, G. I., Wang, Y., Crane, B. R. & Klessig, D. F. (2008) *Journal of Biological Chemistry* 283, 32957-32967.

83. Flores-Perez, U., Sauret-Gueto, S., Gas, E., Jarvis, P. & Rodriguez-Concepcion, M. (2008) *Plant Cell* 20, 1303-15.
84. Sudhamsu, J., Lee, G. I., Klessig, D. F. & Crane, B. R. (2008) *J Biol Chem* 283, 32968-76.
85. Tu, C., Zhou, X. M., Tropea, J. E., Austin, B. P., Waugh, D. S., Court, D. L. & Ji, X. H. (2009) *Proceedings of the National Academy of Sciences of the United States of America* 106, 14843-14848.
86. Tomar, S. K., Dhimole, N., Chatterjee, M. & Prakash, B. (2009) *Nucleic Acids Research* 37, 2359-2370.
87. Hwang, J. & Inouye, M. (2006) *Molecular Microbiology* 61, 1660-1672.
88. Meinke, D., Sweeney, C. & Muralla, R. (2009) *PLoS One* 4, e7386.
89. Bylund, G. O., Wipemo, L. C., Lundberg, L. A. C. & Wikstrom, P. M. (1998) *Journal of Bacteriology* 180, 73-82.
90. Lovgren, J. M., Bylund, G. O., Srivastava, M. K., Lundberg, L. A. C., Persson, O. P., Wingsle, G. & Wikstrom, P. M. (2004) *Rna-a Publication of the Rna Society* 10, 1798-1812.
91. Datta, P. P., Wilson, D. N., Kawazoe, M., Swami, N. K., Kaminishi, T., Sharma, M. R., Booth, T. M., Takemoto, C., Fucini, P., Yokoyama, S. & Agrawal, R. K. (2007) *Molecular Cell* 28, 434-445.
92. Datta, P. P., Wilson, D. N., Swami, N. K., Sharma, M. R., Booth, T., Kawazoe, M., Shirouzu, M., Takeruoto, C., Yokoyama, S., Fucini, P. & Agrawl, R. K. (2007) *Biophysical Journal*, 9a.
93. Lesnyak, D. V., Osipiuk, J., Skarina, T., Sergiev, P. V., Bogdanov, A. A., Edwards, A., Savchenko, A., Joachimiak, A. & Dontsova, O. A. (2007) *Journal of Biological Chemistry* 282, 5880-5887.

94. Kaczanowska, M. & Ryden-Aulin, M. (2005) *Biochimica Et Biophysica Acta- Gene Structure and Expression* 1727, 87-96.
95. El Yacoubi, B., Lyons, B., Cruz, Y., Reddy, R., Nordin, B., Agnelli, F., Williamson, J. R., Schimmel, P., Swairjo, M. A. & de Crecy-Lagard, V. (2009) *Nucl. Acids Res.* 37, 2894-2909.
96. Tkaczuk, K., Dunin-Horkawicz, S., Purta, E. & Bujnicki, J. (2007) *BMC Bioinformatics* 8, 73.
97. Bisanz, C., Begot, L., Carol, P., Perez, P., Bligny, M., Pesey, H., Gallois, J. L., Lerbs-Mache, S. & Mache, R. (2003) *Plant Mol Biol* 51, 651-63.
98. Walbot, V. & Coe, E. H. (1979) *Proceedings of the National Academy of Sciences of the United States of America* 76, 2760-2764.
99. Beligni, M. V. & Mayfield, S. P. (2008) *Plant Mol Biol* 67, 389-401.
100. Bollenbach, T. J., Sharwood, R. E., Gutierrez, R., Lerbs-Mache, S. & Stern, D. B. (2009) *Plant Mol Biol* 69, 541-52.
101. Walter, M., Kilian, J. & Kudla, J. (2002) *Embo J* 21, 6905-14.
102. Yehudai-Resheff, S., Hirsh, M. & Schuster, G. (2001) *Mol Cell Biol* 21, 5408-16.
103. Yehudai-Resheff, S., Portnoy, V., Yogev, S., Adir, N. & Schuster, G. (2003) *Plant Cell* 15, 2003-19.
104. Slomovic, S., Portnoy, V., Yehudai-Resheff, S., Bronshtein, E. & Schuster, G. (2008) *Biochim Biophys Acta* 1779, 247-55.
105. Carpousis, A. J. (2002) *Biochem Soc Trans* 30, 150-5.
106. Baginsky, S., Shteiman-Kotler, A., Liveanu, V., Yehudai-Resheff, S., Bellaoui, M., Settlage, R. E., Shabanowitz, J., Hunt, D. F., Schuster, G. & Gruissem, W. (2001) *Rna* 7, 1464-75.

107. Britton, R. A., Wen, T. Y., Schaefer, L., Pellegrini, O., Uicker, W. C., Mathy, N., Tobin, C., Daou, R., Szyk, J. & Condon, C. (2007) *Molecular Microbiology* 63, 127-138.
108. Rocak, S. & Linder, P. (2004) *Nature Reviews Molecular Cell Biology* 5, 232-241.
109. Linder, P. (2006) *Nucleic Acids Res* 34, 4168-80.
110. Boudet, N., Aubourg, S., Toffano-Nioche, C., Kreis, M. & Lecharny, A. (2001) *Genome Research* 11, 2101-2114.
111. Mingam, A., Toffano-Nioche, C., Brunaud, V., Boudet, N., Kreis, M. & Lecharny, A. (2004) *Plant Biotechnology Journal* 2, 401-415.
112. Aubourg, S., Kreis, M. & Lecharny, A. (1999) *Nucleic Acids Research* 27, 628-636.
113. Myouga, F., Akiyama, K., Motohashi, R., Kuromori, T., Ito, T., Iizumi, H., Ryusui, R., Sakurai, T. & Shinozaki, K. (2009) *Plant J* epub- preview.
114. Li, D. Y., Liu, H. Z., Zhang, H. J., Wang, X. E. & Song, F. M. (2008) *Journal of Experimental Botany* 59, 2133-2146.
115. Little, M. C. & Hallick, R. B. (1988) *J Biol Chem* 263, 14302-7.
116. Liere, K. & Börner, T. (2007) in *Cell and Molecular Biology of Plastids*, pp. 121-174.
117. Pfalz, J., Liere, K., Kandlbinder, A., Dietz, K. J. & Oelmüller, R. (2006) *Plant Cell* 18, 176-97.
118. Barkan, A., Klipcan, L., Ostersetzer, O., Kawamura, T., Asakura, Y. & Watkins, K. P. (2007) *Rna* 13, 55-64.
119. Kroeger, T. S., Watkins, K. P., Friso, G., van Wijk, K. J. & Barkan, A. (2009) *Proc Natl Acad Sci U S A* 106, 4537-42.
120. Liere, K. & Link, G. (1995) *Nucleic Acids Res* 23, 917-21.

121. Vogel, J., Hubschmann, T., Borner, T. & Hess, W. R. (1997) *J Mol Biol* 270, 179-87.
122. Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M. L., Mireau, H., Peeters, N., Renou, J. P., Szurek, B., Taconnat, L. & Small, I. (2004) *Plant Cell* 16, 2089-103.
123. Schmitz-Linneweber, C., Williams-Carrier, R. E., Williams-Voelker, P. M., Kroeger, T. S., Vichas, A. & Barkan, A. (2006) *Plant Cell* 18, 2650-63.
124. Schmitz-Linneweber, C., Williams-Carrier, R. & Barkan, A. (2005) *Plant Cell* 17, 2791-804.
125. Barkan, A., Walker, M., Nolasco, M. & Johnson, D. (1994) *Embo J* 13, 3170-81.
126. Fisk, D. G., Walker, M. B. & Barkan, A. (1999) *Embo J* 18, 2621-30.
127. Lahmy, S., Barnèche, F., Derancourt, J., Filipowicz, W., Delseny, M. & Echeverria, M. (2000) *FEBS Letters* 480, 255-260.
128. Schult, K., Meierhoff, K., Paradies, S., Toller, T., Wolff, P. & Westhoff, P. (2007) *Plant Cell* 19, 1329-46.
129. Nakamura, T., Ohta, M., Sugiura, M. & Sugita, M. (2001) *J Biol Chem* 276, 147-152.
130. Hirose, T. & Sugiura, M. (2001) *Embo J* 20, 1144-52.
131. Tillich, M., Hardel, S. L., Kupsch, C., Armbruster, U., Delannoy, E., Gualberto, J. M., Lehwark, P., Leister, D., Small, I. D. & Schmitz-Linneweber, C. (2009) *Proc Natl Acad Sci U S A* 106, 6002-7.
132. Mooney, B. P., Miernyk, J. A. & Randall, D. D. (2002) *Annu Rev Plant Biol* 53, 357-75.
133. Camp, P. J. & Randall, D. D. (1985) *Plant Physiol* 77, 571-577.

134. Johnston, M. L., Luethy, M. H., Miernyk, J. A. & Randall, D. D. (1997) *Biochim Biophys Acta* 1321, 200-6.
135. Mooney, B. P., Miernyk, J. A. & Randall, D. D. (1999) *Plant Physiol* 120, 443-52.
136. Lin, M., Behal, R. & Oliver, D. J. (2003) *Plant Mol Biol* 52, 865-72.
137. Lutziger, I. & Oliver, D. J. (2000) *FEBS Lett* 484, 12-6.
138. Drea, S. C., Mould, R. M., Hibberd, J. M., Gray, J. C. & Kavanagh, T. A. (2001) *Plant Mol Biol* 46, 705-15.
139. Sun, J., Ke, J., Johnson, J. L., Nikolau, B. J. & Wurtele, E. S. (1997) *Plant Physiol* 115, 1371-83.
140. Ke, J., Choi, J. K., Smith, M., Horner, H. T., Nikolau, B. J. & Wurtele, E. S. (1997) *Plant Physiol* 113, 357-65.
141. Ke, J., Wen, T. N., Nikolau, B. J. & Wurtele, E. S. (2000) *Plant Physiol* 122, 1057-71.
142. Feria Bourrellier, A. B., Ferrario-Mery, S., Vidal, J. & Hodges, M. (2009) *Biochem Biophys Res Commun* 387, 700-4.
143. Bonaventure, G. & Ohlrogge, J. B. (2002) *Plant Physiol* 128, 223-35.
144. Ito, J., Heazlewood, J. L. & Millar, A. H. (2006) *J Proteome Res* 5, 3459-69.
145. Schoonheim, P. J., Veiga, H., Pereira Dda, C., Friso, G., van Wijk, K. J. & de Boer, A. H. (2007) *Plant Physiol* 143, 670-83.
146. Fuller, B., Stevens, S. M., Jr., Sehnke, P. C. & Ferl, R. J. (2006) *Proteomics* 6, 3050-9.
147. Winkel, B. S. (2004) *Annu Rev Plant Biol* 55, 85-107.
148. Jorgensen, K., Rasmussen, A. V., Morant, M., Nielsen, A. H., Bjarnholt, N., Zagrobelny, M., Bak, S. & Moller, B. L. (2005) *Curr Opin Plant Biol* 8, 280-91.

149. Sato, N., Terasawa, K., Miyajima, K. & Kabeya, Y. (2003) *Int Rev Cytol* 232, 217-62.
150. Sakai, A., Takano, H. & Kuroiwa, T. (2004) *Int Rev Cytol* 238, 59-118.
151. Cannon, G. C., Ward, L. N., Case, C. I. & Heinhorst, S. (1999) *Plant Mol Biol* 39, 835-45.
152. Krause, K., Maier, R. M., Kofer, W., Krupinska, K. & Herrmann, R. G. (2000) *Mol Gen Genet* 263, 1022-30.
153. Friso, G., Giacomelli, L., Ytterberg, A. J., Peltier, J. B., Rudella, A., Sun, Q. & Wijk, K. J. (2004) *Plant Cell* 16, 478-99.
154. Jeong, S. Y., Rose, A. & Meier, I. (2003) *Nucleic Acids Res* 31, 5175-85.
155. Barkan, A. (1988) *Embo J* 7, 2637-44.
156. Westhoff, P. & Herrmann, R. G. (1988) *Eur J Biochem* 171, 551-64.
157. Jenkins, B. D. & Barkan, A. (2001) *Embo J* 20, 872-9.
158. Barkan, A. (2009) *Methods Mol Biol* 553, 13-37.
159. Till, B., Schmitz-Linneweber, C., Williams-Carrier, R. & Barkan, A. (2001) *Rna* 7, 1227-38.
160. Ostheimer, G. J., Williams-Carrier, R., Belcher, S., Osborne, E., Gierke, J. & Barkan, A. (2003) *Embo J* 22, 3919-3929.
161. Asakura, Y. & Barkan, A. (2007) *Plant Cell* 19, 3864-75.
162. Asakura, Y., Bayraktar, O. A. & Barkan, A. (2008) *Rna* 14, 2319-2332.
163. Ullers, R. S., Houben, E. N., Brunner, J., Oudega, B., Harms, N. & Luirink, J. (2006) *J Biol Chem* 281, 13999-4005.
164. Rutkowska, A., Mayer, M. P., Hoffmann, A., Merz, F., Zachmann-Brand, B., Schaffitzel, C., Ban, N., Deuerling, E. & Bukau, B. (2008) *J Biol Chem* 283, 4124-32.

165. Maier, T., Ferbitz, L., Deuerling, E. & Ban, N. (2005) *Curr Opin Struct Biol* 15, 204-12.
166. Kramer, G., Rauch, T., Rist, W., Vorderwulbecke, S., Patzelt, H., Schulze-Specking, A., Ban, N., Deuerling, E. & Bukau, B. (2002) *Nature* 419, 171-4.

APPENDIX

Additional publications/manuscripts not included in this dissertation:

Poliakov, A., Ponnala, L., Olinares, P.D., Asakura, Y., and van Wijk, K.J. (2010).

A statistical solution for pair-wise comparative proteome analysis using large scale label-free spectral counting. In preparation.

Booth, J.G., Eilertsen, K., Olinares, P.D. and Yu, H. (2010). A Bayesian mixture model for comparative spectral count data analyses. In preparation.

Friso, G., Olinares, P.D. and van Wijk, K.J. (2010). The workflow for quantitative proteome analysis of chloroplast development and differentiation, chloroplast mutants and protein interactions by spectral counting. *Methods in Molecular Biology*. Issue on: Chloroplast Research in Arabidopsis: Methods and Protocols. In press.

Kim, J., Rudella, A., Ramirez-Rodriguez, V., Zybaylov, B., Olinares, P.D., van Wijk, K.J. (2009). Subunits of the plastid ClpPR protease complex have differential contributions to embryogenesis, plastid biogenesis, and plant development in Arabidopsis. *Plant Cell* 21: 1–24.

Sun, Q., Zybaylov, B., Majeran, W., Friso, G., Olinares, P.D., van Wijk, K.J. (2008). PPDB, the Plant Proteomics Database at Cornell. *Nucleic Acids Res.* 37:D969-D974.