FUNCTIONAL CHARACTERIZATION OF THE CLP PROTEASE SYSTEM IN *ARABIDOPSIS* CHLOROPLASTS THROUGH REVERSE GENETICS AND PROTEOMICS

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

Andrea Rudella

January 2007

FUNCTIONAL CHARACTERIZATION OF THE CLP PROTEASE SYSTEM IN *ARABIDOPSIS* CHLOROPLASTS THROUGH REVERSE GENETICS AND PROTEOMICS

Andrea Rudella, Ph.D.

Cornell University 2007

Proteases play an important role in regulating protein maturation, activity and life-time. The Clp protease system in *Arabidopsis thaliana* plastids accumulates at relatively high levels and consists of a proteolytic core and associated chaperones. The core is an assembly of five different catalytic ClpP subunits, four non-catalytic ClpR subunits, and two ClpS proteins with unknown function. ClpR,S are unique to photosynthetic organisms. Three ATP-dependent chaperones, ClpC1,C2,D, are expected to deliver substrates to the ClpPRS core. Control of Clp activity is not understood and Clp substrates are unknown.

Arabidopsis T-DNA insertion Clp mutants were isolated and genotyped. Null mutants for ClpP4,P5 are embryo-lethal under both auto- and heterotrophic conditions. Mutants of ClpP3,R4 did not form seedlings under autotrophic conditions but developed albino seedlings under heterotrophic conditions, displaying limited greening under low light. Null mutants for the chaperones ClpC1 and ClpD have pale-green and wild-type phenotypes, respectively. ClpP,R core subunits are likely essential, while there are redundancies in the ClpC,D subfamily. Two mutants with partial loss of gene expression for ClpR1 and ClpR2 (*clpr2-1*) exhibited pale-green phenotypes, with *clpr2-1* having a stronger phenotype.

ClpR2 protein accumulation in *clpr2-1* chloroplasts was 5-fold reduced, while the ClpPRS core was 3-fold downregulated, suggesting an induction of core composition

heterogeneity. Stromal chaperones were upregulated several fold and ClpC was recruited to the thylakoid membrane. Thylakoid protein homeostasis was unbalanced as deduced from increased accumulation of thylakoid proteases, plastoglobules, protein precursors and degradation products. *Clpr2-1* chloroplasts were smaller, with 30% less thylakoids than wild-type. Clearly, ClpR2 is not a redundant member of the Clp family and reduced *CLPR2* gene expression has adverse effects on plastid and plant development.

A comparative proteome analysis using differential stable isotope labeling of *clpr2-1* and wild-type stroma identified 298 proteins, and 113 were quantified. The Calvin cycle was down-regulated, explaining the slower development of *clpr2-1*. The most striking response was the high accumulation of the chloroplast protein translation machinery and chaperones. This suggests that the ClpPRS core complex may be involved in regulation of plastid gene expression, providing a first understanding of the functional role of the Clp family in plastids.

BIOGRAPHICAL SKETCH

Andrea graduated with a B.S. and M.S. degree in Agro-Industrial Biotechnologies from the University of Verona, Italy, in March 1999. Experimental research for his M.S. degree was carried out at the laboratory of Franco Dellaglio, Scientific and Technologic Department, University of Verona, Italy. The *M.S.* thesis is entitled "Eco-compatible cloning and expression of an alkaline cellulase in *Lactobacillus plantarum*".

After the mandatory Italian Military Service as Police Officer (2^{nd} Lieutenant), Andrea joined the laboratory of Klaas van Wijk, Biochemistry and Biophysics Department, Stockholm University, Sweden, in August 2000 as Research Assistant.

In January 2001 Andrea entered the Graduate School at Cornell University, Ithaca, NY, USA, for a *Ph.D.* in the field of Plant Biology. The research was carried on in the laboratory of Klaas van Wijk and the thesis presented in August 2006.

Andrea is now an MS Proteomics Solution Specialist, Europe, for Waters Co.

In addition to the thesis, Andrea also contributed to five publications concerning the analysis of plastid proteomes, but they are not part of this dissertation. Specifically, he applied expertise in protein identification by mass spectrometry (Peltier et al., 2002; Friso et al., 2004; Peltier et al., 2004b; Giacomelli et al., 2006; Peltier et al., 2006) as well as molecular biology and fluorescence confocal microscopy (Peltier et al., 2004b). Andrea also contributed to the development of the bioinformatics infrastructure for proteome analysis in the van Wijk lab.

to my wife Lisa

and my good, extended, Italian family

always loving and supporting

ACKNOWLEDGEMENTS

Needless to say, but many thanks to Klaas for offering me this opportunity and giving me all his support throughout the *Ph.D.* 'Tante grazie' to all the past and present lab members, a long list of names that I will always treasure. And Professors David Stern, Maureen Hanson, Jian Hua and Thomas Brutnell are acknowledged for serving in my Special Committee and steering my research efforts toward success.

This work was supported by the grants from the National Sciece Foundation (NSF, #MCB 0343444) and the US Department of Energy (DOE, DE-FG02-04ER15560) to Klaas Jan van Wijk

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1. page

CHAPTER 2.

CHAPTER 3.

CHAPTER 4.

LIST OF TABLES

CHAPTER ONE

INTRODUCTION[1]

The essential role of plastids in plant development and function

Plastids are organelles of prokaryotic origin and are essential for many aspects of plant cell development and function. Plastids are present in virtually every plant cell. (Zhang et al., 2003). Chloroplasts are the best known form of plastids and are present in all green tissues. Chloroplasts have thylakoid membranes that contain the photosynthetic electron transport chain which transforms light into chemical energy used for carbon fixation (photosynthesis). Chloroplasts and other plastid types are key sources of lipids, amino acids, vitamins, nucleotides, and secondary metabolites, such as hormones, alkaloids and isoprenoids. Plastids are also important in gravity sensing, synthesis of carotenoids, production of volatile compounds (e.g. terpenes). Plastids contain their own genome and protein synthesis machinery, but the majority of plastid proteins are encoded by nuclear genes and synthesized in the cytosol in the form of precursors which typically contain an N-terminal plastid/chloroplast targeting peptide (cTP). Thus formation of plastids involves coordinated expression of two different genomes. To facilitate so many different biochemical pathways, plastids must contain a large number of proteins and efforts were made to catalogue them by experimentation and prediction (reviewed in van Wijk, 2004). The subcellular localization program TargetP (http://www.cbs.dtu.dk/services/TargetP/) predicts that different plastid types collectively contain up to 4650 nuclear-encoded proteins in *Arabidopsis thaliana*. Given the reported sensitivity (0.85) and false positives rate (0.69) (Sun et al., 2004), we could expect a total plastid proteome of maximum (4650

[1] Part of this introduction is adapted from a recent review that we co-authored: **Adam, Z., Rudella, A., and van Wijk, K.J.** (2006). Recent advances in the study of Clp, FtsH and other proteases located in chloroplasts. Curr Opin Plant Biol **9,** 234-240.

x 0.69/0.85 =) 3774 nuclear-encoded and 87 plastid-encoded proteins. Considering that 26,751 protein-coding genes are currently annotated by "The Arabidopsis Information Resource" (TAIR, genomic annotation v6, www.arabidopsis.org), the plastid proteome may account for 14% of all plant proteins; this is a very significant investment by the plant into this organelle. Most plant cells contain many copies of plastids and chloroplast proteins account for more than half the total protein content in leaves.

Plastid biogenesis and protein homeostasis

Developing and maintaining the diverse array of plastid functions is a daunting task. Plastids are not formed *de novo* (Osteryoung and Nunnari, 2003; Pyke, 2006). Plastid biogenesis requires not only coordinated gene expression of the nuclear and plastid genomes, but also many post-transcriptional control steps. Consequently, many plastid proteins function in biogenesis, protein homeostasis and protection (i.e. from abiotic stress).

Chaperones and proteases are pivotal players in protein homeostasis and their actions are often intertwined. The importance of these proteins to plastid and plant cell function and development is underscored by the appearance of strong and often lethal phenotypes in mutants (Budziszewski et al., 2001). In *A. thaliana* for example, a null mutant for the alpha subunit of the chaperone 60 complex (Cpn60α) is seedling lethal (Apuya et al., 2001) and null mutants in the membrane bound metallo-protease FtsH family have a variegated leaf phenotype (Takechi et al., 2000; Sakamoto et al., 2002).

Plastid proteolysis

Proteases are involved in numerous aspects of biogenesis and maintenance of chloroplasts and plastids in general, including i) the removal and degradation of signal

sequences, ii) the degradation of partially assembled complexes or damaged proteins, iii) adaptation to changes in environmental conditions and iv) the controlled breakdown of chloroplasts in senescing leaves (for a review, see Sakamoto, 2006). Proteolysis is important for regulation of metabolic and signaling pathways, as well as protein homeostasis. Given the evolutionary origin of chloroplasts, it is not surprising that all plastid proteases identified so far are homologues of bacterial proteases (see i.e. Sokolenko et al., 2002). The Clp protease is the most abundant plastid protease and is probably composed of 14 gene products (reviewed in Adam et al., 2006; Sakamoto, 2006). Clp is the focus of this dissertation and will be extensively discussed further below.

As of May 2006, the MEROPS database (http://merops.sanger.ac.uk/) reports for *A. thaliana* 676 genes encoding for proteins with features of proteases. Based on sequence conservation analysis it is predicted that 528 are expected to be catalytically active (Rawlings et al., 2006). Proteases can be classified accordingly to their mechanism of action. They use most commonly hydroxyl or sulfhydryl groups as nucleophiles and less frequently activated water molecules. Serine, threonine and cysteine proteases belong to the first group, whereas aspartic, metallo and glutamic proteases belong to the second group (for a review, see van der Hoorn and Jones, 2004). Other common distinctions are the energy dependence of nucleotides (ATP/GTP) and the cleavage site in the substrate, which can be internal (endo-peptidases) or terminal (exo- or amino-peptidases).

Aminopeptidases in plastids are likely to have an important role in protein modification. Most proteins have stabilizing N-termini to avoid N-end rule degradation (Giglione and Meinnel, 2001). In *A. thaliana*, about half of the predicted 20 aminopeptidases might localize to plastids. Two peptide deformylases and three methionine aminopeptidases in plastids co-operate in the modification of N-termini and therefore half-life of plastid encoded proteins (Giglione et al., 2003, and for a review Giglione and Meinnel, 2001).

Here, I will briefly summarize our current knowledge of structure and function of the plastid proteases known to date (Fig. 1.1). There are at least 50 genes encoding plastid targeted proteases, many of which have homologues targeted to mitochondria, and a few proteases are dually targeted to plastids and mitochondria.

After import of precursor proteins into plastids (Fig. 1.1), their N-terminal chloroplast transit peptides (cTP) are cleaved off resulting in mature and functional proteins. An endoprotease from pea, termed stromal processing peptidase (SPP), was shown to be able to process several substrates with various efficiencies *in vitro* (reviewed in Richter and Lamppa, 2005). Some of these substrates were not processed to their final N-terminal residues. An additional requirement for yet unidentified aminopeptidases was then postulated (Richter and Lamppa, 1998). *Arabidopsis* and tobacco antisense SPP lines have various strong phenotypes and show reduced import rates of GFP fused to the cTP of ferredoxin (Zhong et al., 2003); this suggests that cTP processing is coupled to the chloroplast import. Further degradation of the cTPs to short peptides is likely facilitated by one or two Zn-metallo-proteases, assigned AtPreP1 and AtPreP2. Both are present in chloroplasts (Peltier et al., 2006) but are also targeted to mitochondria, where

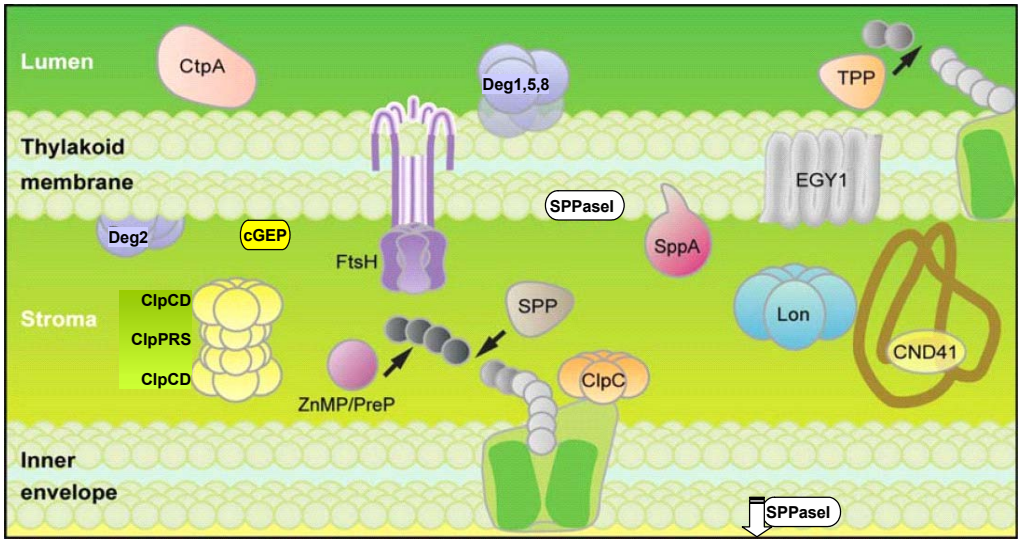


Figure 1.1. Schematic representation of known proteases in chloroplasts. Adapted from (Sakamoto, 2006).

they may perform similar functions (Bhushan et al., 2005; Stahl et al., 2005). However, the short peptide products of PreP might be further degraded to free amino acids by yet uncharacterized oligopeptidases, amino- and carboxypeptidases (Stahl et al., 2005). Proteins destined for the thylakoid lumen possess an additional N-terminal lumenal transit peptide (lTP) that is removed by one or more thylakoid processing peptidases (TPP) (Shackleton and Robinson, 1991). It is unclear what happens to lTPs once cleaved off, whether they are degraded directly in the lumen or they are transported back to the stroma. TPPs are type I signal peptidases. TPP-1 showed activity *in vitro* against a synthetic stromal intermediate of OEC23 (Chaal et al., 1998). TPP-2 (or SPase I) is located in the thylakoid membrane as determined by mass spectrometry (Kleffmann et al., 2004; Peltier et al., 2004a). TPP-2 has also been implicated in the processing of Toc75, an outer chloroplast envelope membrane protein with a cTP, followed by a cleavable outer membrane glycine-rich stop transfer sequence (Inoue et al., 2005). Thus TPP-2 is located in both the thylakoid membrane as well as in the chloroplast envelope, making it the first identified proteolytic enzyme that is found in the outer envelope (Inoue et al., 2005).

A well studied protein is the chloroplast-encoded 32 kDa D1 subunit of Photosystem II. The D1 protein has five trans-membrane domains. Upon integration into the thylakoid membrane, the C-terminus is processed at the lumenal side of the membrane by the C-terminal processing peptidase (CtpA) - its only substrate seems to be D1. The exposure of the new C-terminus is essential in coordination of Mn ions that function in the water splitting complex (Liao et al., 2000; Inagaki et al., 2001; Yamamoto et al., 2001; Roose and Pakrasi, 2004). The D1 protein is also highly sensitive to photodamage and has a very high turnover rate (for a review, see Zhang and Aro, 2002).

Proteolysis of integral membrane proteins poses a special challenge- since part of the protein is embedded in the lipid bilayer. Proteases can either cleave the (hydrophobic) transmembrane domains (TMD) within the bilayer or cleave the exposed (soluble) loops.

The cleaved fragments need to be 'pulled' out of the membrane bilayer. It has been proposed that members of the Deg protease family are responsible for cleaving the exposed loops of thylakoid proteins (Haussuhl et al., 2001). Deg proteins are ATP-independent serine endopeptidases with a trypsin-like catalytic domain. They contain PDZ domains involved in protein–protein interactions and Deg proteins are believed to form trimers that condense into hexamers. Deg1, 5 and 8 have been found in the thylakoid lumen whereas Deg2 is peripherally associated with the thylakoid membrane on the stromal side (Fig. 1.1) (for a review, see Huesgen et al., 2005). The cleavage of TMDs within the thylakoid lipid bilayer is believed to primarily occur by ATP-dependent FtsH Zn metallo-proteases. FtsH is a multigene family with 12 of its members targeted to plastids. Some of these FtsH proteins have 'lost' the catalytic domain. When single null mutants for chloroplast FtsH proteins were exposed to photoinhibitory conditions, the D1 protein was stabilized (Bailey et al., 2002), suggesting the direct involvement of FtsH in D1 degradation. *In vitro* studies seem to confirm this role as well (Ostersetzer and Adam, 1997; Lindahl et al., 2000, reviewed in Nixon et al., 2005). The predicted structure of FtsH contains two trans-membrane helices, followed by an ATPase domain, and a zinc-binding motif serving as the catalytic site (for recent reviews see Ito and Akiyama, 2005; Adam et al., 2006). FtsH forms single hexameric rings with the catalytic sites sequestered within (Bieniossek et al., 2006). In *A. thaliana* chloroplasts, FtsH2 and 5 are the most abundant components (Sinvany-Villalobo et al., 2004) and single null mutants in either of these subunits have a variegated phenotype (*var2* and *var1* respectively) and increased sensitivity to photoinhibition (reviewed in Aluru et al., 2006). Interestingly, the functionality of FtsH seems to have developed into pairs; FtsH2 mutants can be rescued by over-expressing FtsH8 but not FtsH1; on the other hand, FtsH5 mutants can be rescued by over-expressing FtsH1 but not FtsH8 (Yu et al., 2004b, 2005).

Another integral membrane thylakoid protease is the ATP-independent, metallo-protease ethylene-dependent gravitropism-deficient and yellow-green 1 (EGY1). Null

mutants for this gene show impairment in chloroplast biogenesis with reduced thylakoid accumulation as compared to wild-type plants (Chen et al., 2005). A protein similar to prokaryotic signal peptide peptidase A (SppA, also known as protease IV) is another ATP-independent serine-type protease that tightly interacts with the surface of the thylakoid membrane through an amphipatic helix and it forms homotetramers (Ichihara et al., 1986). SppA is strongly up-regulated by high light, but nothing is known about its substrates (Lensch et al., 2001). A glutamyl endopeptidase (cGEP) was identified in pea chloroplasts and was shown to localize at the thylakoid surface. In vitro, cGEP can cleave the N-terminal part of Lhcb1. A protein homolog of cGEP is present in *A. thaliana* and seems to have three conserved amino acid residues that are typical for the catalytic site of serine proteases (Forsberg et al., 2005).

Remobilization of free amino acids is very important especially during natural leaf senescence in flowering annuals when leaf chloroplasts are degraded and nitrogen moved to the fast developing flowers and seeds. In chloroplasts, only the aspartic protease CND41 from tobacco has been linked to this process. CND41 can degrade Rubisco, and its expression increases during senescence. Interestingly, CND41 binds to and can be inhibited by DNA. Lower DNA content in senescing chloroplasts has been invoked in explaining higher CND41 activity during this late developmental stage. Two genes in *A. thaliana* encode for CND41 homologues (Kato et al., 2004; Kato et al., 2005).

Two more multi-gene families encoding ATP-dependent proteases are present in plastids and are the Lon and Clp families. The Lon protease counts several members of unknown localization and is expected to form hexamers (Adam et al., 2001) – in this dissertation we identified a Lon protease in the chloroplast stroma. In *Escherichia coli*, Lon was shown to have an affinity for DNA (Suzuki et al., 1997) and it is interesting to know whether they retained this function in *A. thaliana*. The ATP-dependent Clp (caseinolytic protease) family consists of five serine-type ClpP (P1, P3-6) proteases, four ClpP-related ClpR (R1-4) proteins, three AAA^+ chaperones (C1, C2 and D; ATPases

Associated with diverse cellular Activities), and three members (ClpS1, S2, T) with unknown functions (Peltier et al., 2004b). ClpT localization is uncertain even though it contains a predicted cTP. Clp is not only the most diversified protease gene family to localize in plastids, but it is probably also the most abundant. Characterization of the plastid Clp protease is the focus of this dissertation. In the next sections, more details will be provided about the Clp proteolytic system in the better studied prokaryotes (section 1.4 below) and followed by what is known in *Arabidopsis* (section 1.5).

E. COLI CLP PROTEASE IS THE MODEL SYSTEM

Clp modules and assembly states

The Clp protease was first discovered in *E. coli* (Hwang et al., 1987; Katayama-Fujimura et al., 1987) and its structural and functional organization are best characterized in this organism. These general features are likely conserved among Clp homologs and provide therefore an excellent template for the study of the Clp plastid protease family. In *E. coli*, the Clp protease machinery can be divided into two main biochemical components: a proteolytic core complex, formed by the serine-type peptidase ClpP, and ATP-dependent chaperones, ClpA, ClpB and ClpX, members of the AAA^+ superfamily

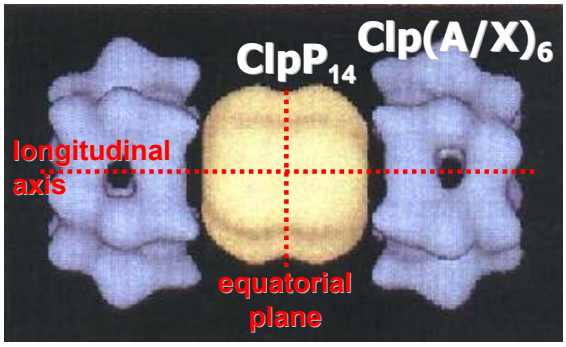


Figure 1.2. CryoEM picture depicting the Clp quaternary structure in *E. coli*. 14 subunits of identical proteolytic ClpP assemble into two juxtaposed heptameric rings ($ClpP_{14}$ in yellow), sequestering the catalytic sites within an axial channel. Homohexameric rings of ATPases ClpA or ClpX ($Clp(A/X)_6$ in blue) dock at either side of the ClpP core. In this picture, only $ClpX_6$ is represented. The main axes of symmetry are indicated. Adapted from (Beuron et al., 1998).

(Fig. 1.2). In *E. coli* and many other bacteria, *CLPP* is dispensable but is particularly important under stress conditions (Maurizi et al., 1990b; Gottesman, 1996). In contrast, in cyanobacteria several members of the Clp family are essential (Schelin et al., 2002).

The *E. coli* ClpP protease subunits assemble into two juxtaposed homo-heptameric rings (hence a homo-tetradecamer) that sequester the catalytic sites in the inner chamber of this barrel-like $ClpP_{14}$ complex (Kessel et al., 1995; Beuron et al., 1998). This core complex shows a seven-fold symmetry around its axial pore and a dual symmetry at the equatorial plane (Fig. 1.2). ClpP is encoded as a precursor protein and upon assembly a 14 aa N-terminal pro-peptide is auto-catalytically cleaved off (Maurizi et al., 1990b; Maurizi et al., 1990a). The core alone is able to degrade only small peptides of up to six amino acids in an exoergonic reaction (Thompson et al., 1994). Proteolysis of larger peptides and proteins is facilitated by the association of hexameric rings of chaperone subunits to either side of the barrel-like structure (Fig. 1.2).

The chaperone ClpA/B/X proteins are characterized by a modular structure (Schirmer et al., 1996), with either one (ClpX) or two (ClpA/B) ATPase domains containing highly conserved Walker sequences (or Nucleotide Binding Domains, NBD), and a Sensor and Substrate Discrimination (SSD) C-terminal domain. ClpA/B have a weakly conserved 'middle' sequence between the two ATPase domains, and an N-terminal domain of variable size and conservation, important in substrate binding. However, only ClpA has a Zn binding domain (ZBD) within the N-terminal domain (Xia et al., 2004). At their C-termini, ClpA/X, but not ClpB, have a conserved I/LGF tri-peptide (preceded by a positively charged residue) that is implicated in interaction with the proteolytic core. The absence of this motif prevents ClpB from associating with ClpP. In fact it has been shown in *E. coli* that ClpB acts exclusively as an HSP/chaperone in combination with DnaJ/K to untangle protein aggregates (Mogk and Bukau, 2004; Schlieker et al., 2004), but that simple insertion of the I/LGF motif allows ClpB to functionally interact with the ClpP core complex (Weibezahn et al., 2004). ClpA and

ClpX can also act as chaperones to (re)fold (aggregated) proteins. However, when bound to $ClpP_{14}$, ClpA/X use their unfolding activity to feed the proteolytic chamber with substrates for degradation (Burton and Baker, 2005).

Clp mode of action

Many multimeric proteases form barrel-like structures that sequester the active sites away from the cellular proteome and maximize potential contacts with substrates within the chamber for optimum degradation. On top of the core are multimeric rings of ATPase/chaperone proteins. Such oligomeric structures are also observed for the archeabacterial and yeast proteasome (Lowe et al., 1995; Groll et al., 1997; Walz et al., 1998), Clp (Wang et al., 1997; Sousa et al., 2000) and yeast mitochondrial Lon (Stahlberg et al., 1999), as determined by X-ray crystallography or cryo-electron microscopy and single particle averaging. Proteolysis is an exoergonic process, but energy is usually required for regulation, substrate unfolding and translocation into the catalytic chamber.

High-resolution crystal structures were generated for *E. coli* ClpP and ClpA (Wang et al., 1997; Guo et al., 2002) and the ClpP(A/X) complexes were observed by cryo-electron microscopy (Kessel et al., 1995; Beuron et al., 1998). This has given a magnificent insight into the mode of action of the Clp machinery. Since ClpX homologues have not been found in *A. thaliana* plastids (see below), we will report mainly on ClpP-A interactions.

Substrates must be recognized by the chaperones before being presented to the core for degradation. Within the ClpA structure, the N-terminal domain has been held responsible for this function. Due to its mobility on top of the complex, the N-terminal domain has been compared to tentacles that capture substrates and bring them close to the axial pore (Xia et al., 2004). Using different model substrates, it was determined that ClpA with N-terminal truncations has weaker interactions with the ClpP core and rates of

degradation are reduced, probably because the N-terminal domain can induce allosteric changes that favour interaction with the ClpP core (Hinnerwisch et al., 2005). The N-terminus of ClpA is also the site of interaction with the specificity protein ClpS (Xia et al., 2004). ClpS is unrelated to the other members of the Clp family (ClpP/A/X/B). In *E. coli*, it has been shown that ClpS is responsible for directing ClpA specificity toward protein aggregates (Zeth et al., 2002a; Zeth et al., 2002c) and that ClpS is necessary for degradation of N-end rule substrates (Erbse et al., 2006). Interaction between ClpS and the N-terminal domain of ClpA involves the coordination of the Zn^{2+} present in this domain (see above and Xia et al., 2004). In *Bacillus subtilis* ClpC (the *E. coli* ClpA homolog) only functions in proteolysis after association of the adaptor protein MecA (Kirstein et al., 2006).

ClpA hexameric rings presumably bind only one substrate at the time and translocate it to the ClpP core (Piszczek et al., 2005). ATP is required for ring formation and substrate translocation into the proteolytic chamber (Maurizi et al., 1994; Hoskins et al., 2000; Singh et al., 2001), and just one subunit might be able to drive substrate translocation (Martin et al., 2005). ClpPA has been shown to degrade proteins to an average of 6-8 aa long peptides (Choi and Licht, 2005). Choi and Licht theorized also that ClpP and ClpA act with an allosteric mechanism in which substrate translocation activates proteolysis and proteolysis deactivates translocation. More mechanistic insights come from (Bewley et al., 2006), where the authors proposed a model (Fig. 1.3) in which six loops at the N-terminus of ClpP hold the chaperone ring in position creating a "pseudo-6-fold symmetric interaction" that is partially in contrast with the previously proposed ratcheting mechanism (Kessel et al., 1995). The axial pore in $ClpP_{14}$ is only 10 Å wide (Wang et al., 1997), but it can stretch to 20-25 Å (Burton et al., 2001), possibly by the 'extrusion' of 6 N-termini of ClpP thus providing increased access to the catalytic chamber. The remaining loop folds in so to provoke allosteric changes within the chamber: it might trigger loosening at the 'equator' resulting in openings for product

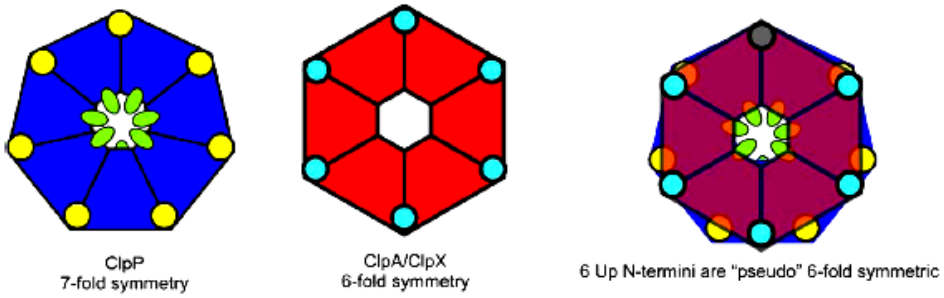


Figure 1.3. Schematic representation of the $ClpP_7/Clp(A/X)_6$ symmetry mismatch in *E. coli.*
$ClpP_7$ in blue; $Clp(A/X)_6$ in red; I/LGF residues in cyan, corresponding pockets in $ClpP_7$ in yellow; ClpP N-termini in green. From (Bewley et al., 2006).

release. Thus the 3D structure of the ClpP core seems to be very dynamic. Recently, substantial advances have been made in the understanding of the mechanism of product release for the Clp proteolytic machinery which might have evolved differently in different species. In *Homo sapiens* mitochondria, single ClpP heptameric rings can associate with ClpX hexameric rings to form an intermediate complex (with exposed catalytic sites). This complex combines with other heptameric rings to form the complete ClpP barrel. The dynamic nature or flexibility of the double ClpP rings (tetradecamer) allows release of proteolytic products from the catalytic chamber (Kang et al., 2005). In *E. coli*, where the ClpP tetradecamer assembles before interacting with the chaperones, pores have been identified between the so-called handles of ClpP - these are α-helices that lock into place the two juxtaposed heptameric rings at the 'equator' (Sprangers et al., 2005). Relaxation in this region has been hypothesised also for *Streptococcus pneumoniae* Clp (Gribun et al., 2005).

E. coli Clp substrates

Several Clp substrates have been identified in *E. coli.* A specific substrate for ClpAP is RepA, the plasmid P1 initiator protein. ClpA can convert dimers of RepA into monomers that bind DNA specifically or feed them to ClpP for degradation (Hoskins et

al., 2000). Another ClpAP substrate is RecN, a protein needed in DNA repair (Meddows et al., 2005), which is also substrate to ClpXP (Neher et al., 2006). ClpAP is involved in general removal of aggregated and missfolded proteins and nascent peptide chains stalled on the cytosolic ribosomes (for reviews see Chandu and Nandi, 2004; Sauer et al., 2004). Stalled nascent chains are selected for proteolysis by a unique tagging system (SsrA) that targets these polypeptides to proteolysis by ClpAP, ClpXP and FtsH proteases (Herman et al., 1998). Two proteins that can influence this process are SspB and ClpS. SspB enhances ClpXP-recognition and ClpS diverts ClpAP from tagged substrates toward aggregated proteins (Wah et al., 2002). Binding of ClpS to the N-terminal domain of ClpA blocks tag recognition while improving aggregates detection (Zeth et al., 2002b; Zeth et al., 2002c). As mentioned earlier, ClpAP degrades N-end rule substrates in a ClpS-dependent manner (Erbse et al., 2006). ClpXP substrates have been identified more in detail. Flynn and colleagues used the proteolytically incompetent epitope-tagged $ClpP^{trap}$ to collect trapped substrates and identify them by mass spectrometry. With this new set of substrates, the authors were able to determine the presence of at least five $ClpXP^{trap}$ recognition signals (Flynn et al., 2003). Data for trapped substrates in $ClpAP^{trap}$ are still unpublished. Recently, the same $ClpXP^{trap}$ strain has been used in substrate profiling after DNA damage (Neher et al., 2006).

THE ARABIDOPSIS CLP FAMILY

Sequence analysis

Whereas many *E. coli* proteases are represented by a single gene, they have evolved into multi-gene families in photosynthetic bacteria and plants (Sokolenko et al., 2002). Also, several specific protease components have evolved, likely to accommodate the photosynthetic apparatus and the role of the plastids in plant development. In *A. thaliana* the Clp family has expanded to some 25 nuclear genes and one plastid gene

(Peltier et al., 2004b). The plastid localized Clp proteolytic system is far more complex than in bacteria and it includes five serine-type ClpP (P1, P3-6) proteases, four ClpP-related ClpR (R1-4) proteins, three AAA^+ chaperones (C1, C2 and D) similar to ClpA/B from *E. coli* (see above), and three members (ClpS1, S2, T) with unknown functions (Peltier et al., 2004b) (Figure 1.4). ClpT localization is uncertain even though it contains a predicted cTP. We have shown by GFP-fusions that a single ClpP homolog (ClpP2) localizes to mitochondria where it is believed to interact with the three *Arabidopsis* ClpX gene products (X1-3) (Peltier et al., 2004b)(Figure 1.5). An additional Clp chaperone (ClpB3) is present in plastids. However, ClpB3, similarly to its *E. coli* homolog, does not contain the conserved I/LGF tripeptide and is unlikely to interact with the plastid Clp core complex. Table 1.1 summarizes the common characteristics between *E. coli* and *A. thaliana* Clp homologues with relevance to plastid proteolysis.

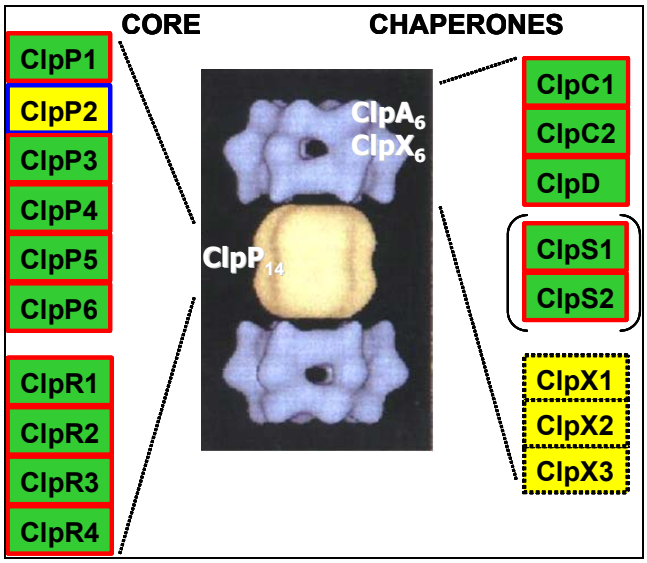


Figure 1.4. The *A. thaliana* Clp protease family.
E. coli has one gene encoding for the proteolytic subunit ClpP (yellow in the CryoEM picture in center, see Fig. 1.2 for more information); *A. thaliana* has 10 homologues. Six have conserved the catalytic triad (ClpP1-6) and 4 have not (ClpR1-4). In place of the 2 *CLPA/X* genes of *E. coli*, *A. thaliana* has 3 homologues for *CLPA* (*CLPC1,C2,D*) and 3 for *CLPX* (*CLPX1-3*). Two novel genes (*CLPS1,2*) similar to only the N-terminal part of ClpA have evolved in *A. thaliana*. *CLPP1* is plastid encoded; all the others are nuclear encoded. Green background indicates chloroplast localization, yellow mitochondrial localization (only predicted and not experimentally determined for ClpX1-3).

The ClpR proteins have significant homology to the ClpP proteins (24 to 38% primary sequence similarity for the predicted mature proteins), but they do not contain

Table 1.1. Comparison of characteristics between *Escherichia coli* and *Arabidopsis thaliana* Clp homologs with relevance to plastid proteolysis.

module	*Escherichia coli*		*Arabidopsis thaliana*	
	protein	**Characteristics**	**proteins**	**characteristics**
core protease	ClpP	2 juxtaposed homo-heptameric rings (cryo-electron microscopy)	ClpP1,3-6 ClpR1-4	2 juxtaposed hetero-heptameric rings (homology modeling)
		14 aa propeptide cleavage		none identified/suspected
		N-terminal loops involved in chaperone docking		??
		6-8 aa average products		??
		equator loosening for product release		putative polar equator openings for product release
chaperone	ClpA	homo-hexamers	ClpC1,2 ClpD	homo/hetero-hexamers?
		2 NBD, N-terminal domain		2 NBD, N-terminal domain
		Zn in N-terminal domain		??
		I/LGF and positively charged residue for interaction with ClpP		I/LGF and positively charged residue for interaction with ClpP
		involved in folding and proteolysis		involved in protein import, (folding and proteolysis??)
		adaptor proteins needed for function in proteolysis		??
adaptors	ClpS	binds N-terminal domain of ClpA through coordination of Zn	ClpT (unknown location)	??
		provides aggregated and N-term-rule proteins as substrates		??
	ssrA-tag	added to stalled ribosomal nascent chains to guide them to Clp-mediated degradation	??	??
	??	??	ClpS1,2	Tightly associated with the core. Unknown function

the conserved catalytic triad Ser-His-Asp. The chaperones ClpC1, C2, and D are believed to form hexameric rings and interact with the core. ClpS1,2 share homology with the N-terminal domains of ClpC1,2, but they do not contain any NBD and their functions are unknown. ClpS1,2 homologs do not appear in the genomes of prokaryotes including

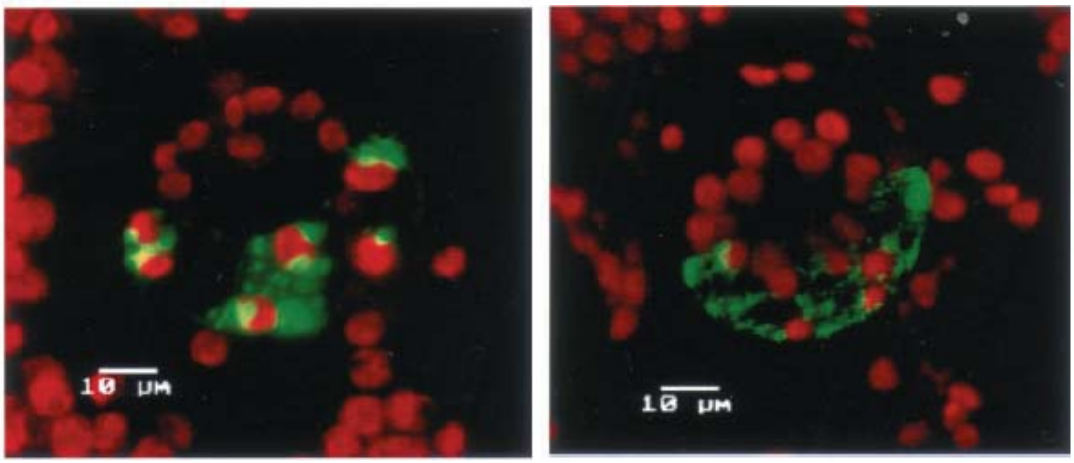


Figure 1.5. ClpP2 localizes to mitochondria.
Tobacco cells expressing ClpP2:GFP showing that ClpP2 (left) and the positive control coxIV:mGFP4 (right) are exclusively targeted to mitochondria. The figures show a projection of 9 transverse optical sections taken at 1 µm intervals along the optical axis. The pseudo colors applied are green for GFP and red for chlorophylls. From (Peltier et al., 2004b).

photosynthetic bacteria, but are present in the genome of the green alga *Chlamydomonas reinhardtii* (O. Vallon and W. Majeran - personal communication). The evolution and conservation of ClpS1,2 suggest that they may have acquired a function specific to higher plants and green algae (Peltier et al., 2004b).

Transcript analysis of most *CLPP/R* genes showed constitutive expression in roots and leaves of *A. thaliana.* There are contrasting reports about mRNA and protein levels for Clp subunits under various stresses or in different developmental stages (Nakabayashi et al., 1999; Weaver et al., 1999). It seems that expression of ClpP/C is not altered significantly after short-term stresses both at mRNA and protein levels (Zheng et al., 2002), although transcript levels increased after 2.5 hours of high light treatment (Sinvany-Villalobo et al., 2004). In contrast, *CLPD* expression is increased under drought conditions (ClpD is also knows as ERD1, early responsive to dehydration 1, Kiyosue et al., 1993) or during senescence (Nakashima et al., 1997) even though protein levels decrease (Weaver et al., 1999).

Oligomerization of Clp proteins

With the exception of ClpT, all proteins of the Clp proteolytic system have been identified by mass spectrometry analyses of the native soluble proteome of non-green plastids in *Brassica rapa* roots and *Brassica oleracea* petals and in chloroplasts of *A. thaliana* leaves (Peltier et al., 2004b). The chaperone protein ClpC1 and likely C2 (although no unique peptides for this homolog were found) were identified as dimers at ~200 kDa, not in association with the proteolytic core, while ClpD was not identified. The association of the chaperones to the core to activate proteolysis might be a transient interaction difficult to capture under standard purification conditions. Moreover, there is the possibility of adaptor proteins are needed to activate ClpC for proteolysis, as it is the case for the ortholog in *B. subtilis* (Kirstein et al., 2006), which would make isolation of a Clp chaperone/core complex more challenging. The plastid Clp core was isolated as a ~325-350 kDa and with a pI of ~5.0. This complex contained all the catalytically active ClpP1,3-6 and also the inactive ClpR1-4. In addition, ClpS1,2 were also identified in this complex, suggesting that they are tightly associated with the core. Therefore, the core complex has been designated ClpPRS. Taking into consideration protein spot patterns, amino acid compositions and staining properties, we determined that the *A. thaliana* Clp core protease complex consists of one to three copies of the five ClpP proteins, one to two copies of the four ClpR proteins and one copy of ClpS1 and S2 proteins (Peltier et al., 2004b). However, slightly different compositions cannot be excluded and it is also possible that Clp proteins can occupy different positions in the ring structure. For comparison, the stromal chloroplast Clp core complex from the green alga *C. reinhardtii* is also heteromeric with several ClpP/R proteins, but so far no ClpS proteins were detected (Majeran et al., 2005). In contrast to plastids, plant mitochondria contain a single homo-tetradecameric ClpP2 complex, that can presumably associate with ClpX chaperones (Peltier et al., 2004b).

Homology modeling suggests functions and a novel mechanism for release of proteolytic products

3-D homology modeling and threading was applied to the *A. thaliana* Clp family to try to understand why such complexity has arisen. This showed that the ClpP/R proteins fit well together in a tetradecameric complex, whereas ClpS1 and S2 are not part of these rings but fit on the axial sites of the ClpPR core (Peltier et al., 2004b). The predicted 3D structures for ClpR1 (Figure 1.6), R3 and R4, show the presence of the L1 insertion loop which is oriented into the catalytic chamber and could influence substrate presentation to the catalytic sites of the ClpPs. An additional feature for all of ClpR and the ClpP proteins, except ClpP1 and P6, is an extended C-terminus as compared to the *E. coli* ClpP. The role of these extensions is not clear, but might influence interaction with the chaperone rings or ClpS1,2. The function of ClpS1 and S2 is unknown and their observed tight association with the core complex is puzzling (Peltier et al., 2004b). Binding of ClpS to an axial side of the core is incompatible with binding by ClpC or D to that side. Several hypotheses for ClpS function seem plausible: i) regulation of ClpC or D chaperone interaction with the core through competition for docking sites; ii) an adaptor role for interaction with other proteins; and iii) interaction with substrates, in particular short, unstructured peptides, since ClpS1 and S2 do not contain any ATPase domain (required for unfoldase activity). This last hypothesis is intriguing in the light of an observation in *E. coli*: a slight change in the model of ClpA and ClpP allosteric interaction (see above) can potentially produce very small degradation products (Choi and Licht, 2005). If ClpS1,2 interaction with the ClpPR core is enough to produce "always active" catalytic sites, then Clp could be responsible for removal of peptides resulting from primary proteolytic events.

Sequence alignments and homology modeling revealed the presence of conserved residues among all ClpP/R members that likely constitute lateral pores next to the catalytic sites (Peltier et al., 2004b). These pores could be used for release of small

proteolytic fragments instead of the postulated release through the axial channels (Figure 1.7). Some evidence for the general hypothesis of lateral exit as compared to axial exit has now been provided for the *E. coli* Clp complex (see above).

Lessons from Clp gene disruption studies

Relatively little has been published on the functional importance of the Clp family in plants. The plastid-encoded ClpP1 subunit was shown to be essential for shoot development in tobacco (Shikanai et al., 2001; Kuroda and Maliga, 2003), whereas it does not appear to affect root development. The *CLPP1* gene was 'lost' in maize BMS cell cultures grown under heterotrophic conditions (Cahoon et al., 2003) - this may be specific for maize or may indicate that *CLPP1* is not needed in non-green plastids during heterotrophic growth in cell cultures. However, the parasitic plant *Epifagus virginiana* maintains a *CLPP* gene in its minimal plastid genome supporting an important role of the

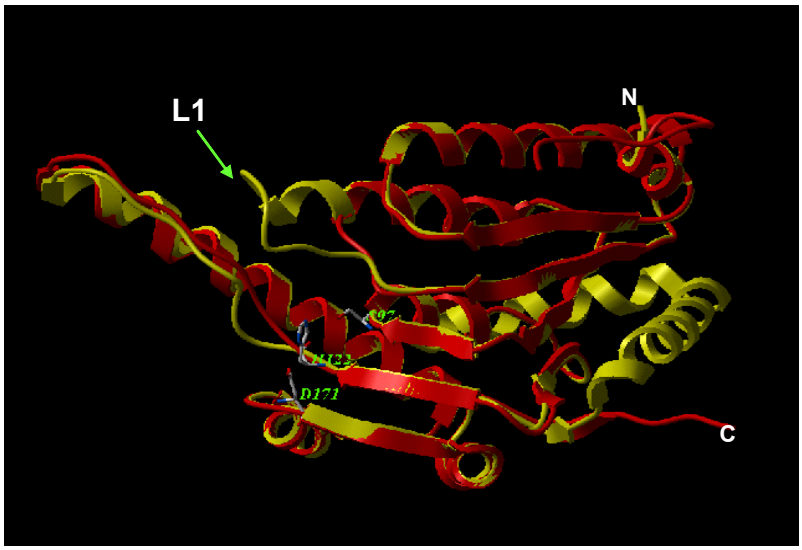


Figure 1.6. Homology modeling of ClpP/R.
A ClpR1 monomer (in yellow) is superimposed to *E. coli* ClpP to highlight the structural conservation and the differences among the two homologues. Catalytic residues in green. Insertion loop (L1), N- and C-termini (N and C respectively) are indicated.
From (Peltier et al., 2004b).

plastid Clp complex for plants in a natural environment is that (Cui et al., 2006). Plastid *CLPP1* transcription is likely to be tightly regulated in tobacco, since over-expression of the *CLPP1* promoter to drive expression of (exogenous) transgenes genes in plastids had adverse effects on plant development (Kuroda and Maliga, 2002). Down-regulation of the plastid-encoded *CLPP1* gene in *C. reinhardtii* suggested that ClpP1 is involved in the degradation of the thylakoid bound subunits of the cytochrome b_6f and PSII complexes (Majeran et al., 2000; Majeran et al., 2001), although PSII components are readily degraded also in a Clp-independent manner (Giglione et al., 2003). In the photosynthetic bacterium *Synechococcus* sp. PCC 7942, *CLPP1* and *P2* are dispensable, while *P3* and *R* are essential (Schelin et al., 2002; Barker-Astrom et al., 2005). In a recent review, A. Clarke speculates that *A. thaliana* KO mutants for *CLPP* genes might be embryo lethal. Interestingly, he reports that antisense lines for *CLPP4* have a yellow-heart, variegated phenotype whereas *CLPP5* and *P6* have a pale green phenotype. *CLPR1* KO was reported to have a variegated phenotype similar to *CLPP4* antisense. Clarke reports also of KO lines for *CLPT* and *CLPD* that show no phenotype, but no evidence was provided (Clarke et al., 2005).

In *A. thaliana*, loss of expression of the ClpC1 chaperone resulted in reduced plant

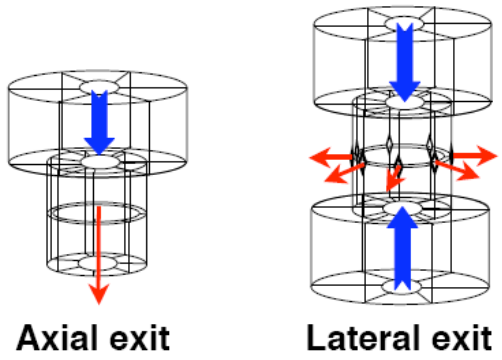


Figure 1.7. Models for proteolytic products exit.
Hexameric ClpC chaperone rings feed substrates (blue arrows) into the catalytic chamber of the ClpPR core. Cleaved products exit (red arrows) through the axial channel (left) or via putative polar lateral openings (right).
From (Adam et al., 2006).

growth and chloroplast development, but homozygous plants are autotrophic and seeds are viable (Constan et al., 2004; Sjogren et al., 2004; Kovacheva et al., 2005). This partial redundancy of ClpC1 is likely due to the expression of the ClpC2 homolog (Kovacheva et al., 2005). Although ClpC1,2 accumulate predominantly in the stroma, ClpC1 is also associated with the chloroplast protein import machinery in the inner envelope (Akita et al., 1997; Nielsen et al., 1997). Consistently, loss of ClpC1 results in lower protein import rates into isolated chloroplasts (Constan et al., 2004; Kovacheva et al., 2005). A mutant with reduced expression of ClpC2 was isolated in a screen for suppressors of the variegated phenotype of FtsH2 (*var2*) (Park and Rodermel, 2004). The isolated ClpC2 mutant line did not have any visible phenotype, probably because of the minor expression levels as compared to the homolog ClpC1. The explanation provided by the Rodermel lab for the phenotypic suppression was that ClpC2 is a suppressor of thylakoid biogenesis and maintenance and that it might act by accelerating photooxidative stress. We rather favor the interpretation that ClpC2 is needed for thylakoid biogenesis. When ClpC2 levels are low, thylakoid biogenesis is slowed down, removing the need for a threshold concentration of FtsH2. Whatever the correct explanation, the suppression data clearly show an interesting connection between different plastid localized proteolytic systems that warrants further investigation.

PREMISES AND MOTIVATION FOR THIS THESIS

A biochemical approach has allowed the determination of the composition of the Clp complex in chloroplasts of *A. thaliana* (Peltier et al., 2004b). The 11 *CLPP/R/S* gene products of the core always co-purify, but a mixture of complexes cannot be excluded, and it is not clear if there is functional redundancy between the different ClpP/R/S proteins. A genetic approach is needed to determine if any of the ClpP/R core subunits are essential and such an approach was the main thrust of this thesis. Particularly

intriguing is the role of the ClpR proteins, not catalytically active but yet present in every Clp core complex analyzed so far. Therefore, many of the efforts described in this thesis focus of the function of the ClpR proteins.

CHAPTER TWO

Reverse genetics of the plastid Clp protease gene family of *Arabidopsis thaliana*; ClpP and R subunits are essential, whereas ClpC and D chaperones are partially redundant

INTRODUCTION

The Clp machinery in *A. thaliana* plastids is present in non-photosynthetic plastids in roots and petals as well as in chloroplasts. It consists of five serine-type ClpP (P1,P3-6) proteases, four ClpR (R1-4) proteins, three Clp AAA^+ chaperones (C1,C2 and D), and two small ClpS (S1,S2) proteins with unknown functions but similar to the N-terminal domain of ClpC (Adam and Clarke, 2002; Peltier et al., 2004b). ClpP1 is encoded by the plastid genome, while all others are encoded by the nuclear genome. The ClpR proteins lack the three conserved catalytic amino acid residues found in ClpP family members (Adam and Clarke, 2002). These 14 gene products form a proteolytic system consisting of i) a 325-350 kDa ClpPRS protease core complex (Peltier et al., 2004b), ii) 200 kDa dimers of ClpC chaperones, and iii) the ClpD chaperone with an unknown assembly state (Weaver et al., 1999; Zheng et al., 2002; Peltier et al., 2004b). The core complex can be divided into a tetradecameric catalytic ClpPR complex and associated ClpS1,2 proteins. The chaperones ClpC,D are believed to form hexameric rings that interact with the axial sides of the ClpPR complex to recognize and feed substrates. Because ClpS proteins also dock on the axial sides of the ClpPR complex, a regulatory role for the ClpS proteins was hypothesized (Peltier et al., 2004b). For comparison, mitochondrial, *Escherichia coli* and human Clp are also tetradecameric proteases, but contain 14 copies of the same ClpP gene product and no ClpR or ClpS homologues (see i.e. Kang et al., 2002; Sauer et al., 2004). Moreover, the presence of ClpR and ClpS proteins is unique to photosynthetic organisms. Why the Clp machinery has evolved to such a complexity in plastids is currently unknown.

Relatively little is known of the function of the Clp protease in plants. Some insights are emerging from the isolation and characterization of mutant alleles for members of the Clp family. Analysis of null mutants for *CLPC1* and *C2* have recently been described (Constan et al., 2004; Park and Rodermel, 2004; Sjogren et al., 2004; Kovacheva et al., 2005) showing partial redundancy for ClpC1 and ClpC2 (Kovacheva et al., 2005). It was suggested that a null mutant for *CLPD* has no phenotype, but no experimental evidence was provided (Clarke et al., 2005). The need for at least one Clp chaperone might be deduced by the retention of a copy of ClpC by the (non-photosynthetic) malaria parasite *Plasmodium falciparum* in its reduced plastid genome (Wilson et al., 1996).

Transcript analysis of most ClpP/R genes showed constitutive expression in roots and leaves of *A. thaliana* with only minor changes in gene expression under specific stress conditions or during senescence (Shanklin et al., 1995; Nakabayashi et al., 1999; Zheng et al., 2002).

A genetic approach is needed to determine if any of the nuclear-encoded ClpP/R core subunits are essential. Particularly intriguing is the role of the ClpR proteins, most likely not catalytically active but yet present in every Clp core complex in photosynthetic organisms analyzed so far. It is tempting to suggest that the Clp protease system plays a central role in plastid proteolysis, in many aspects paralleling the role of the 26S proteasome in the cytosol and nucleus (Vierstra, 1996; Moon et al., 2004). Identification of Clp-specific substrates and substrate recognition mechanisms will be a major challenge.

In this chapter, we will describe the isolation of *Arabidopsis* T-DNA insertion Clp mutants. Taking advantage of the T-DNA insertion mutant collection at the Salk Institute (Alonso et al., 2003), we have isolated a number of mutants in the different classes of plastid Clp proteins displaying an interesting array of phenotypes. Null mutants for ClpP4,P5 are embryo-lethal under both auto- and heterotrophic conditions. ClpP3 and

ClpR4 mutants with less than ~1% mRNA accumulation are seedling lethal, but can develop only under heterotrophic conditions and very low light fluencies, generating white leaves with some greening and sterile flowers. Accordingly, chloroplast ultrastructure is affected with the strongest phenotype at higher light intensities. Null mutants for the chaperones ClpC1 and ClpD have pale green and wild type phenotypes, respectively, and they have recently been described in the literature. Two mutants with partial loss (more than 2-fold reduction) of gene expression for ClpR1 (*clpr1-1*) and ClpR2 (*clpr2-1*) exhibited pale green phenotypes. However, they are clearly not variegated or yellow-heart as the anti-sense lines in (Clarke et al., 2005). The phenotype of *clpr2-1* is more severe than that of *clpr1-1*, but both can be maintained as a homozygous population under autotrophic conditions (on soil). *clpr2-1* was selected for biochemical characterization, as will be described extensively in chapters 3 and 4. It is most likely that all ClpP,R core subunits are essential, while there are redundancies within the ClpC,D subfamily.

RESULTS

Collection of mutants for plastid-localized Clp proteins

The expanded plastid localized Clp protein family is puzzling and suggests that the original bacterial Clp genes have duplicated and diverged to acquire specific functions needed for plastid development and protein homeostasis. To better understand the relationship between the different Clp members, protein alignments were generated using the processed form of the Clp proteins (i.e. without the TargetP predicted chloroplast transit peptide, cTP, http://www.cbs.dtu.dk/services/TargetP/) and the structural domains of *E. coli* ClpP (1TYF in PDB) and ClpA N-terminal domain (1K6K in PDB) as consensus sequences. The relatively high sequence conservation within the ClpP/R and

ClpC/D/S protein subfamilies, expressed as percentage of identities, is summarized in Table 2.1.

Table 2.1. Identity matrix table in % for the plastid localized Clp polypeptides. Where applicable, predicted mature proteins are used (i.e. mP5). Proteins are diveded for structural similarity: ClpP/R (upper right corner) and ClpC,D,S (lower left corner). Each group is compared to the homolog sequence from *E. coli* deposited in PDB (1TYF for ClpP and 1K6K for ClpA).

mP5	mP3	mP4	P1	mR2	mR1	mR3	mR4	mP6	
59	51	47	43	38	37	34	44	39	1TYF
	48	45	35	33	28	29	32	37	mP5
		42	36	32	27	30	32	40	mP3
mC1	20		34	33	29	32	36	38	mP4
mC2	20	91		38	26	24	29	31	P1
mD	12	49	49		28	30	32	32	mR2
mS1	16	25	26	17		35	35	24	mR1
mS2	19	23	23	17	55		38	30	mR3
	1K6K	mC1	mC2	mD	mS1			28	mR4

Arabidopsis ClpP proteins have an identity score between them ranging from 31% (comparing P1 and P6) to 48% (comparing P3 and P5); ClpR proteins range from 28% (R1 compared to R2) to 38% identity (R3 compared to R4); ClpR compared to ClpP proteins range from 24% (R3-P1) to 38% (R2-P1). It should be noted that R2 is diverse from ClpR1,3,4 because it lacks the L1 insertion loop and the extended C-terminus (Peltier et al., 2004b). Within the ClpR family, R2 is most similar to ClpP proteins.

ClpC1 and C2 share 91% identity, explaining the difficulties in correct discrimination between the two by tandem mass spectrometry (Peltier et al., 2004b). ClpD is only 49% identical to ClpC and protein expression levels are much lower than ClpC. ClpS1 and S2 share 55% of their residues and respectively ~25% and 17% with ClpC and D.

To better understand the functional significance of the chloroplast targeted Clp gene family, we screened the T-DNA insertion collection (in Col-0 ecotype) available at that time (2001-2002) established at the Salk Institute (http://signal.salk.edu/cgi-

bin/tdnaexpress) (Alonso et al., 2003). Blast searches identified putative insertions in the genomic regions belonging to *CLPP3,4,5, R1,2,4, S1, C1* and *D* (Table 2.2). After retrieving T2 seeds for the putative lines, seedlings were grown on MS agar plates and the T-DNA insertions were confirmed by PCR followed by DNA sequencing. Antibiotic resistance markers are frequently silenced in transgenic plants (Schubert et al., 2004) and therefore these markers were not used for genotyping. Confirmed heterozygous plants were selected and seeds collected. PCR-based segregation analysis of the selved T2 was carried out for seedlings germinating on agar plates with and without addition of sucrose. The characterization of the isolated mutant lines is detailed below and is summarized in Table 2.2.

Table 2.2. List of SALK mutant lines isolated in this study

Gene	Mutant	Salk hit	Hit location[(a)]	Gene expression	Phenotype	KmR [(b)]
At1g66670	*clpp3-1*	SALK_065330	2/4 Intron	KD/KO*	lethal	yes
At5g45390	*clpp4-1*	SALK_000913	5' UTR -16	KO	lethal	yes
At1g02560	*clpp5-1*	SALK_007708	3/9 Exon	KO	lethal	no
At1g02560	*clpp5-2*	SALK_052763	5' UTR	wt-like	wt-like	yes
At1g49970	*clpr1-1*	SALK_088407	1/8 Intron	KD	pale	n.d.
At1g49970	*clpr1-2*	SALK_021405	5' UTR	wt-like	wt-like	no
At1g12410	*clpr2-1*	SALK_046378	5' UTR -7	KD	pale	no
At4g17040	*clpr4-1*	JP7_7H07L	2/6 Intron	KD/KO*	lethal	yes
At4g25370	*clps1-1*	SALK_052772	3/5 Intron	wt-like	wt-like	yes
At5g50920	*clpc1-1*	SALK_014058	4/9 Exon	KO	pale	no
At5g51070	*clpd-1*	JP7_7A07L	7/12 Exon	KO	wt-like	yes
a	The location of T-DNA insertion is indicated as # of exon or intron over the total # in the gene or as in 5' untranslated region (UTR) followed eventually by the number of base pairs before the start codon.					
b	Selectable marker resistance to kanamycin . n.d. - not determined					
*	severe Knock Down (KD) or leaky KO					

Null mutants for CLPP4 and CLPP5 are embryo-lethal

The line SALK_000913 has a T-DNA insertion in *CLPP4* 16 bp upstream of the start codon and was designated *clpp4-1* (Figure 2.1A). For *CLPP5*, the line SALK_007708

has an insertion in the 3rd exon and was designated *clpp5-1*, whereas the line SALK_052763 has an insertion upstream of the 5' untranslated region (UTR) and was designated *clpp5-2* (Figure 2.1A). Homozygous seedlings were identified only for *clpp5-2*, both under autotrophic and heterotrophic conditions. The *clpp5-2* homozygotes grew as wild type (not shown) and RT-PCR analysis showed that *CLPP5* gene expression was not affected (Figure 2.1A), likely because the T-DNA insertion is too far upstream of the *CLPP5* promoter. This line was not further characterized.

For *clpp4-1* and *clpp5-1* no homozygous seedlings could be found, with close to 25% of the seeds not germinating both under autotrophic and heterotrophic conditions, consistent with a single embryo lethal mutation. Different germination and growth conditions were tested without any improvement in germination. These included lower light fluencies and richer media containing up to 2% sucrose and supplementation with Gamborg's vitamin mix and mevalonic acid (MVA) (for an explanation, see discussion).

Heterozygous *clpp4-1* and *clpp5-1* mutants had wild type phenotypes, indicative of a recessive mutation without a gene dosage effect. Siliques of heterozygous plants contained developing seeds devoid of pigmentation, indicative of abnormal chloroplast development in embryos (Figure 2.1B). These seeds, when dry, appeared as empty sacs and no embryo could be recovered under a dissecting microscope. The conclusion is that *clpp4-1* and *clpp5-1* are both null mutants for their respective genes and that embryo development is blocked resulting in a lethal phenotype.

A null mutant for CLPP3 is seedling-lethal

A T-DNA insertion was identified in the line SALK_065330 in the 2nd intron of *CLPP3* and was designated *clpp3-1* (Figure 2.2A). Close to a quarter of the seeds either did not germinate or germinated but became arrested in seedling development under autotrophic conditions (Figure 2.2B). PCR-based genotyping confirmed that these seeds or arrested seedlings were homozygous for *clpp3-1*. Seedlings developed further under heterotrophic

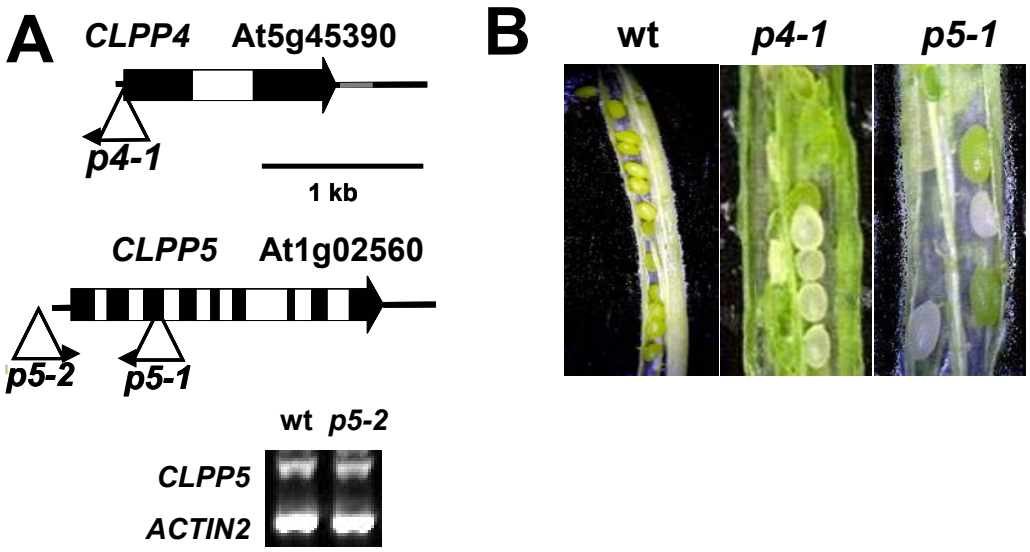


Figure 2.1. T-DNA insertion mutants for *CLPP4* and *CLPP5*.
(A) Gene models for *CLPP4* and *CLPP5*. Exons as black boxes, introns as 'empty' boxes and untranslated regions as a black line, or a grey line if an intron. TAIR accession numbers and location of the inserted T-DNA are indicated, with the left borders displayed as arrow-heads. *clpp4-1* and *clpp5-1* are null mutants, whereas *clpp5-2* has no effect on *CLPP5* expression. RT-PCR for *CLPP5* in *clpp5-2* shows wild type mRNA levels while *ACTIN2* serves as a control.
(B) A maturing silique from a heterozygote plant for *clpp5-1* and *clpp4-1* segregates white and green embryos. A wild type silique with all green embryos is shown for comparison.

conditions (1% sucrose), displaying delayed development, with almost completely white cotyledons and true leaves, and often developing three cotyledons (instead of two) (Figure 2.2B). Surprisingly, no visible phenotype was observed in the roots (not shown). RT-PCR from true leaves of *clpp3-1* could not detect any *CLPP3* transcript, and we estimated it to be <1% of wt (see discussion), but more sensitive determinations are underway in the van Wijk lab (Figure 2.2C). The *clpp3-1* homozygous plantlets could not be obtained on soil. If maintained long enough, *clpp3-1* plantlets bolted but flowers were sterile. *clpp3-1* is therefore seedling-lethal and has to be maintained as a heterozygous population.

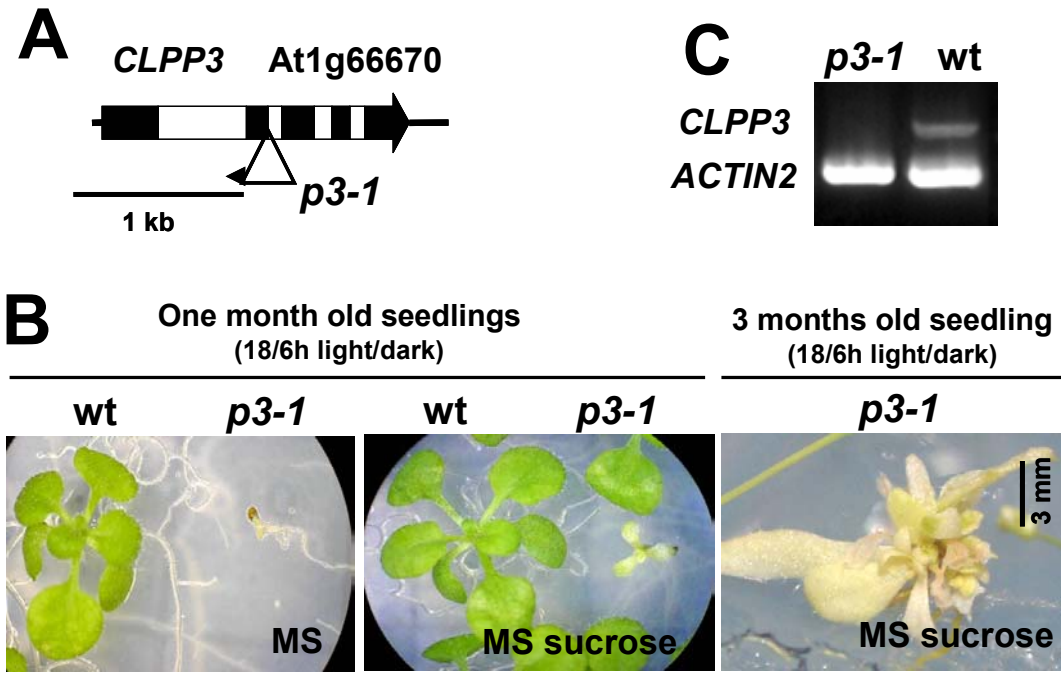


Figure 2.2. Isolation of a T-DNA insertion mutant for *CLPP3*.
(A) Gene model for *CLPP3*. Exons as black boxes, intron empty boxes, untranslated regions as a black line. TAIR accession number is indicated. The T-DNA is inserted in intron 2 of *CLPP3* (left border indicated as an arrow-head).
(B) Development of *clpr3-1* on agar plates under an 18h/6 h light/dark cycle at 100 µmol photons $.m^{-2}.s^{-1}$. *clpp3-1* seedlings are arrested at the cotyledon stage on MS medium without supplemental carbon source. When 1% sucrose is supplemented, *clpp3-1* seedlings develop albino cotyledons (often with three rather than two) and true leaves very slowly as compared to wild type. After 3 months (about 3 wild type generations) *clpp3-1* is still albino, stunted and not flowering.
(C) Semi-quantitative RT-PCR of *CLPP3* transcript in leaves of wild type and *clpr3-1*. Amplification of *ACTIN2* serves as a control.

clpp3-1 is impaired in embryo and chloroplast development

In siliques of heterozygous *clpp3-1* plants, white seeds segregated in a similar fashion as seen for *clpp4-1* and *clpp5-1* (not shown). However, homozygous *clpp3-1* embryos could be excised under a dissecting microscope from dry seeds and were analyzed under a laser scanning confocal microscope. Wild type embryos fill the entire seed, have a well developed and separated pair of cotyledons, and accumulate chlorophyll both in cotyledons and hypocotyl (Figure 2.3A). In contrast, homozygous *clpp3-1* embryos are about half the size of wild-type and don't fill the seed which appears empty and wrinkled.

Homozygous *clpp3-1* embryos are nonetheless visible under a dissecting microscope and can be isolated when using great care. *clpp3-1* embryos appear underdeveloped under the confocal microscope, with stunted cotyledons often in triplicate while the radicle appears asymmetric as compared to wild type (Figure 2.3B). Chlorophyll auto-fluorescence is virtually absent but could be detected by increasing the excitation energy at least a 25% as compared to wild type (Figure 2.3B-C).

To further investigate if the defect in chlorophyll accumulation in *clpp3-1* embryos resulted also in impaired leaf chloroplast development, leaf chloroplast ultrastructure was investigated through transmitted electron microscopy (TEM) by the Cornell microscopy imaging facility (CIMC). Young/emerging leaves from wild type were used as comparison to approximate the delayed development of *clpp3-1* seedlings (Figure 2.3D). Photos of thin sections from *clpp3-1* white leaves revealed small plastids as compared to wild type, with pro-thylakoid membranes and, most likely, phytoferritin crystals (Figure 2.3E). Clearly, lack of *CLPP3* gene expression adversely affects chloroplast and plant development as well as embryo development in the silique.

Reduced expression of CLPR1 and CLPR2 results in pale-green plants

Two T-DNA insertions in the genomic region of *CLPR1* were isolated. In the line SALK_088407, *CLPR1* is disrupted in the 1st intron and the line was renamed *clpr1-1*. For SALK_021405 (renamed *clpr1-2*) the T-DNA was located upstream of the 5' UTR (Figure 2.4A). Both mutant lines could be germinated on soil with 100% efficiency. While *clpr1-2* grew as wild type (not shown), *clpr1-1* segregated as seedlings with yellow cotyledons that rapidly greened (Figure 2.4B). *clpr1-1* plants grow slower and leaves are paler, rounder and flatter than wild type (Figure 2.4B). RT-PCR analysis from leaves showed that *CLPR1* mRNA accumulation was unaffected in *clpr1-2*, whereas expression was reduced below 50% in *clpr1-1* as compared to wild type (Figure 2.4C).

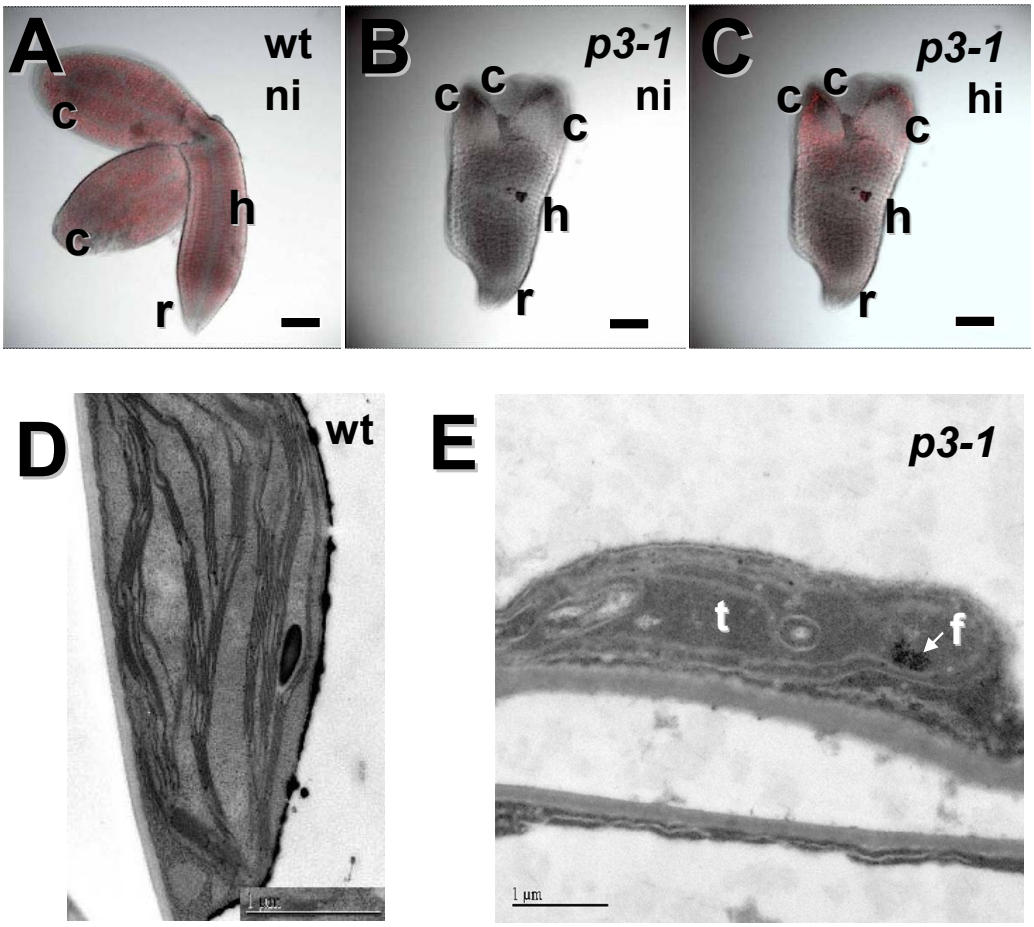


Figure 2.3. *clpp3-1* embryos and plastids develop abnormally.
(A-C) Overlays of Nomarski bright field and chlorophyll red fluorescence confocal microscopy images for wild type and *clpp3-1* embryos excised from imbibed dry seeds. The wild type embryo in **(A)** shows clear cotyledons (c), hypocotyl (h) and radicle (r) formation as well as accumulation of chloroplasts. The mutant embryo in **(B,C)** develops a normal root tip, but shorter and thicker hypocotyl, stunted cotyledons, often present in three-fold. Chlorophyll fluorescence in the mutant is only visible after increasing excitation energy from a normal intensity (ni) in (A,B) to saturating values for wild type (hi) in **(C)**. Scale bars are 80 µm.
(D-E) Transmitted electron microscopy micrograph of thin sections of mesophyll cells showing chloroplasts from wild type **(D)** and *clpp3-1* **(E)**. In chloroplasts of *clpp3-1*, pro-thylakoids membranes (t) and a phytoferritin deposit (f) are visible. Scale bars are 1.0 µm.

SALK_046378 contains a T-DNA insertion 7 bp upstream of the start codon for *CLPR2* and the mutant was named *clpr2-1*. *clpr2-1* plants germinate on soil and develop yellow/pale green leaves (Figure 2.4A,B). *CLPR2* mRNA accumulation is reduced more

than 2-fold in *clpr2-1* as compared to wild type levels (chapter 3). Although *clpr1-1* and *clpr2-1* were not null mutants, they resulted in strong phenotypes, indicating non-redundant functions for these non-catalytic ClpR1 and ClpR2 subunits. Both *clpr1-1* and *clpr2-1* lines were complemented by introducing a copy of the cDNA or genomic DNA, respectively, by floral dipping. The complemented *clpr1-1/R1:XH102*[1] displays green leaves as wild type and grows with a similar pace (Figure 2.4B). *clpr2-1* was complemented and selected for an extensive characterization, as detailed in chapters 3 and 4, because it has a stronger phenotype than *clpr1-1*.

Loss of CLPR4 gene expression in clpr4-1 is seedling-lethal under autotrophic conditions

T2 seeds for the Salk line JP_7H07L were germinated and a T-DNA insertion in the 2nd intron of *CLPR4* was confirmed by PCR amplification and DNA sequencing (Figure 2.5A). This line was renamed *clpr4-1*. Homozygous seedlings were identified by PCR-based segregation analysis of the selved T2 for seedlings germinating on agar. However, development was arrested at the cotyledon stage under autotrophic conditions, similar to *clpp3-1* (Figure 2.5B). Heterozygous mutants had a wild type phenotype, indicative of a recessive mutation. In presence of sucrose, homozygous seedlings developed beyond the cotyledon stage (see below).

CLPR4 transcript was not detected by RT-PCR from the homozygous *clpr4-1* seedlings grown under heterotrophic conditions and we estimated it to be <1% of wt (see discussion), but more sensitive determinations are underway in the van Wijk lab (Figure 2.5C). *clpr4-1* is a near-null mutant with seedling-lethal phenotype.

[1] Complementation of *clpr1-1* was achieved by Verenice Ramirez-Rodriguez and Jitae Kim in the van Wijk lab.

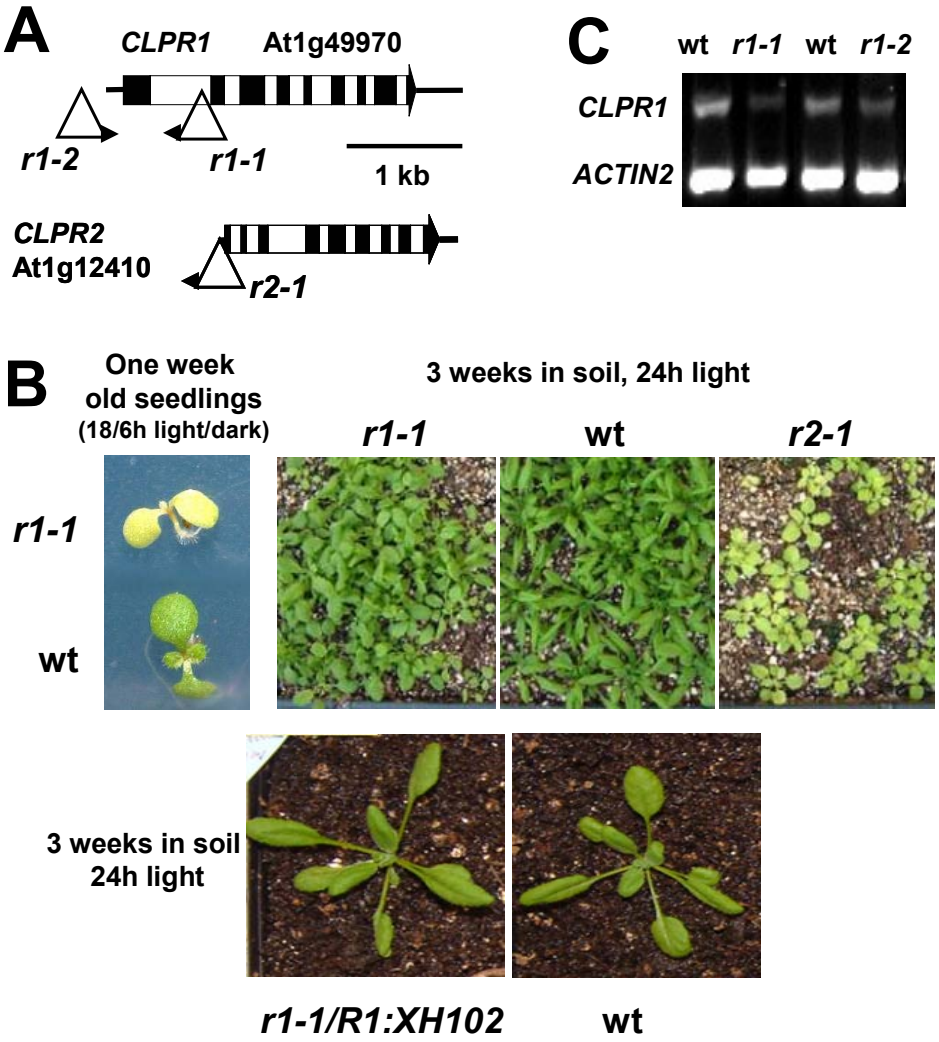


Figure 2.4. Isolation of T-DNA insertion mutants for *CLPR1* and *CLPR2*.
(A) Gene model for *CLPR1* and *CLPR2*. Exons are indicated as black boxes, intron 'empty' boxes, untranslated regions (5' and 3' UTRs) as a black line. TAIR accession number is indicated. Two T-DNA insertion mutants in *CLPR1* (left border indicated as arrow-heads) were analyzed with the insertion in *clpr1-1* located in intron 1 and the insertion in *clpr1-2* upstream of the 5' UTR. The T-DNA insertion in *clpr2-1* is in the 5' UTR.
(B) *clpr1-1* develops as a yellow seedling that rapidly greens, but grows slower and remains paler as compared to wild type. One week old seedlings for *clpr1-1* and wild type are shown (left panel) grown on agar plates with MS medium without supplemented carbon source grown under an 18h/6 h light/dark cycle at 100 µmol photons. $m^{-2}.s^{-1}$. The right hand panels show 3 week old plants for *clpr1-1* and wild type as well as *clpr2-1* germinated in soil under continuous light (100 µmol. $m^{-2}.s^{-1}$). At the bottom, the complemented *clpr1-1/R1:XH102* is shown next to wild type germinated in soil under continuous light (100 µmol photons. $m^{-2}.s^{-1}$).
(C) RT-PCR of *CLPR1* transcript in leaves of wild-type, *clpr1-1* and *clpr1-2* reveals reduced expression in *clpr1-1* but not in *clpr1-2*, after normalization for *ACTIN2* amplification.

Embryo development is impaired in clpr4-1

A phenotype for homozygous *clpr4-1* was observed in siliques of heterozygous plants as segregating white seeds, whereas wild type and heterozygous seeds were green. Mature, dry seeds of homozygous *clpr4-1* were smaller and appeared as almost empty wrinkled sacs (Figure 2.6A). To investigate the phenotype of these homozygous seeds, embryos were gently excised from 24h imbibed mature seeds and analyzed for structure and chlorophyll accumulation by confocal microscopy (Figure 2.6B-E). Embryos developed stunted cotyledons, often in triplicate, a short thick hypocotyl, and an asymmetric radicle

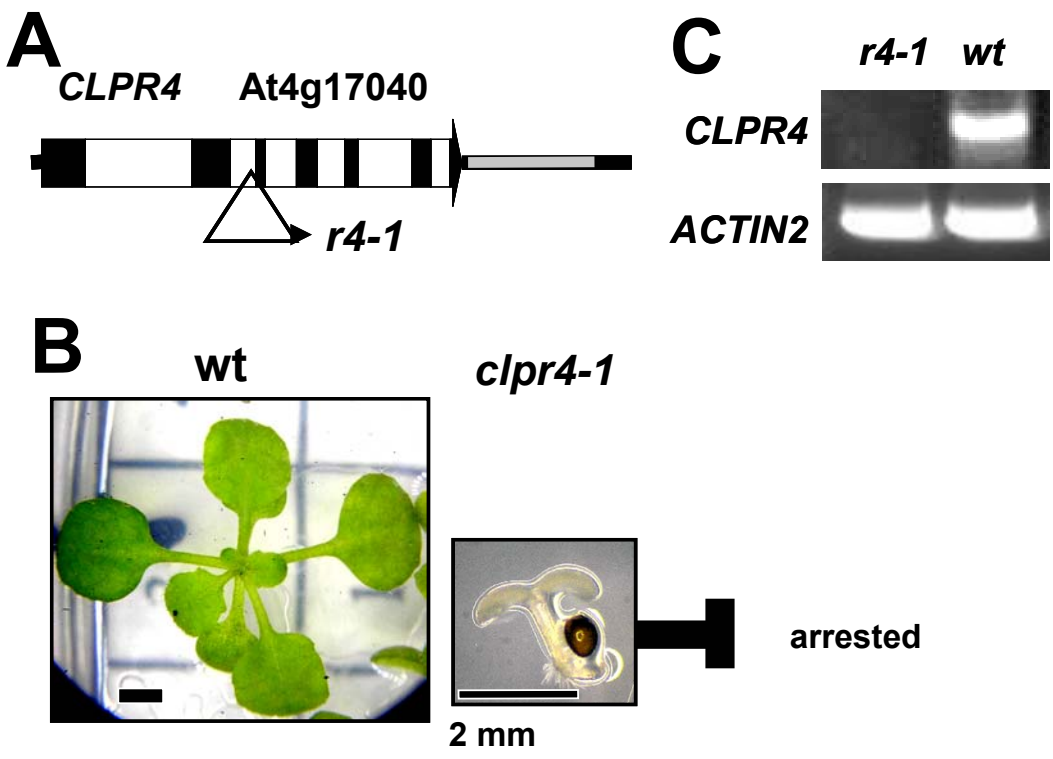


Figure 2.5. Isolation of a T-DNA insertional mutant for *CLPR4*.
(A) The T-DNA insertion site in *CLPR4* (At4g17040) is located between exon 2 and 3 in the coding region (left border indicated as an arrow-head). The exons are represented by black boxes.
(B) Development of *clpr4-1* on agar plates with MS medium without supplemental carbon grown under an 18h/6h light/dark cycle at 100 µmol photons $.m^{-2}.s^{-1}$ (NL). The seedlings are arrested in the cotyledon stage.
(C) Semi-quantitative RT-PCR of *CLPR4* transcript in leaves of wild type and *clpr4-1*. Amplification of *ACTIN2* serves as a control.

(Figure 2.6C-E), very similar to what observed for *clpp3-1* (Figure 2.3B-C). Wild type embryos showed strong auto-fluorescence from chlorophyll, whereas the *clpr4-1* embryos had only patchy chlorophyll accumulation and with lower intensity than wild type (Figure 2.6B,C) that could be detected by increasing the excitation energy a 25% as compared to wild type (not shown). Thus loss of *CLPR4* expression leads to shoot defects during embryo development and embryo plastids do not accumulate chlorophyll at significant levels.

Seedling development of clpr4-1 is light dependent

Homozygous *clpr4-1* seedlings were blocked at the cotyledon stage when grown under autotrophic conditions on agar and did not develop on soil. The seedlings were then tested for rescue by growth under heterotrophic conditions on agar supplemented with 2% sucrose. Under a 18 h light/6 h dark period and 100 μmol photons.m^{-2}.s^{-1} (further referred to as "normal light" or NL), seedlings initially accumulated high levels of anthocyanins in hypocotyl and cotyledons, and subsequently developed white leaves, without any greening (Figure 2.7A). Growth rates were extremely reduced as compared to wild type plants. However, at lower light and shorter day conditions (LL, 10h light/14 h dark and 40 μmol.m^{-2}.s^{-1}), leaves were able to accumulate some chlorophylls and, with several months of delay, also set flowers (Figure 2.7B). These flowers were sterile, never setting any seed. The leaf shapes of *clpr4-1* were highly serrated and curled and shape depended on the light regime. For plants developing some greening in their leaves, their photosystem II efficiency was measured as the ratio between the variable chlorophyll fluorescence (F_v) to maximum chlorophyll fluorescence (F_m) (F_v/F_m) (Krause, 1991). *clpr4-1* grown at LL conditions had an F_v/F_m = 0.51 ±0.15 (n=4), while wild type leaves grown at NL conditions had F_v/F_m values of 0.84 ±0.00 (n=8). Clearly Photosystem II efficiency is 40% reduced in light green *clpr4-1* leaves (and even more in white leaves).

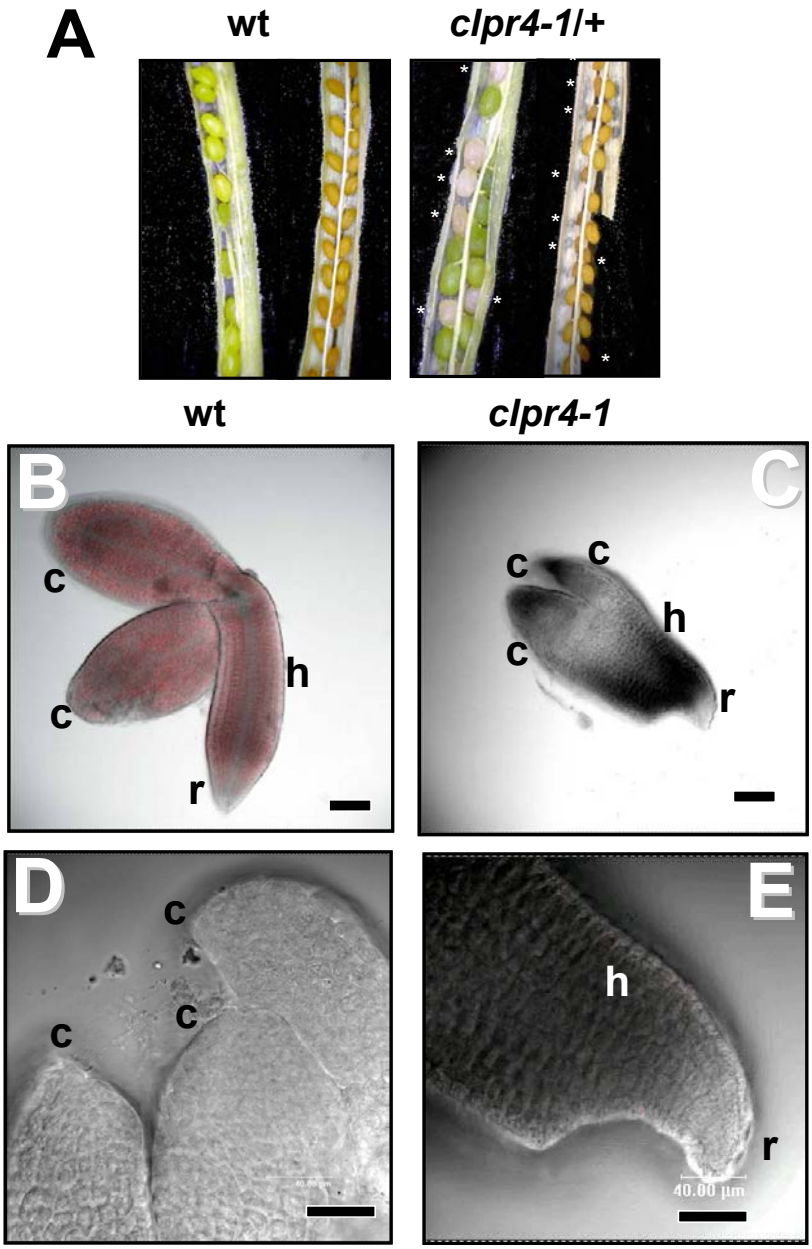


Figure 2.6. *clpr4-1* embryos develop abnormally

(A) Photos of opened siliques at different ripening stages. Heterozygous plants for *clpr4-1 (clpr4-1*/+) develop white seeds in their siliques (asterisks). At a later stage, mutant seeds are recognizable as smaller, wrinkled seeds (asterisks).

(B-E) Overlays of Nomarski bright field and chlorophyll red fluorescence confocal microscopy images for wild type and *clpr4-1* embryos extracted from imbibed dry seeds. The wild type embryo in **(B)** shows clear cotyledons (c), hypocotyl (h) and radicle (r) formation with accumulation of chloroplasts. The mutant embryo in **(C)** develops a normal root tip, but shorter and thicker hypocotyl **(E)** and stunted cotyledons, often present in three-fold **(D)**. Scale bars are 80 µm in (B,C) and 40 µm in (D,E).

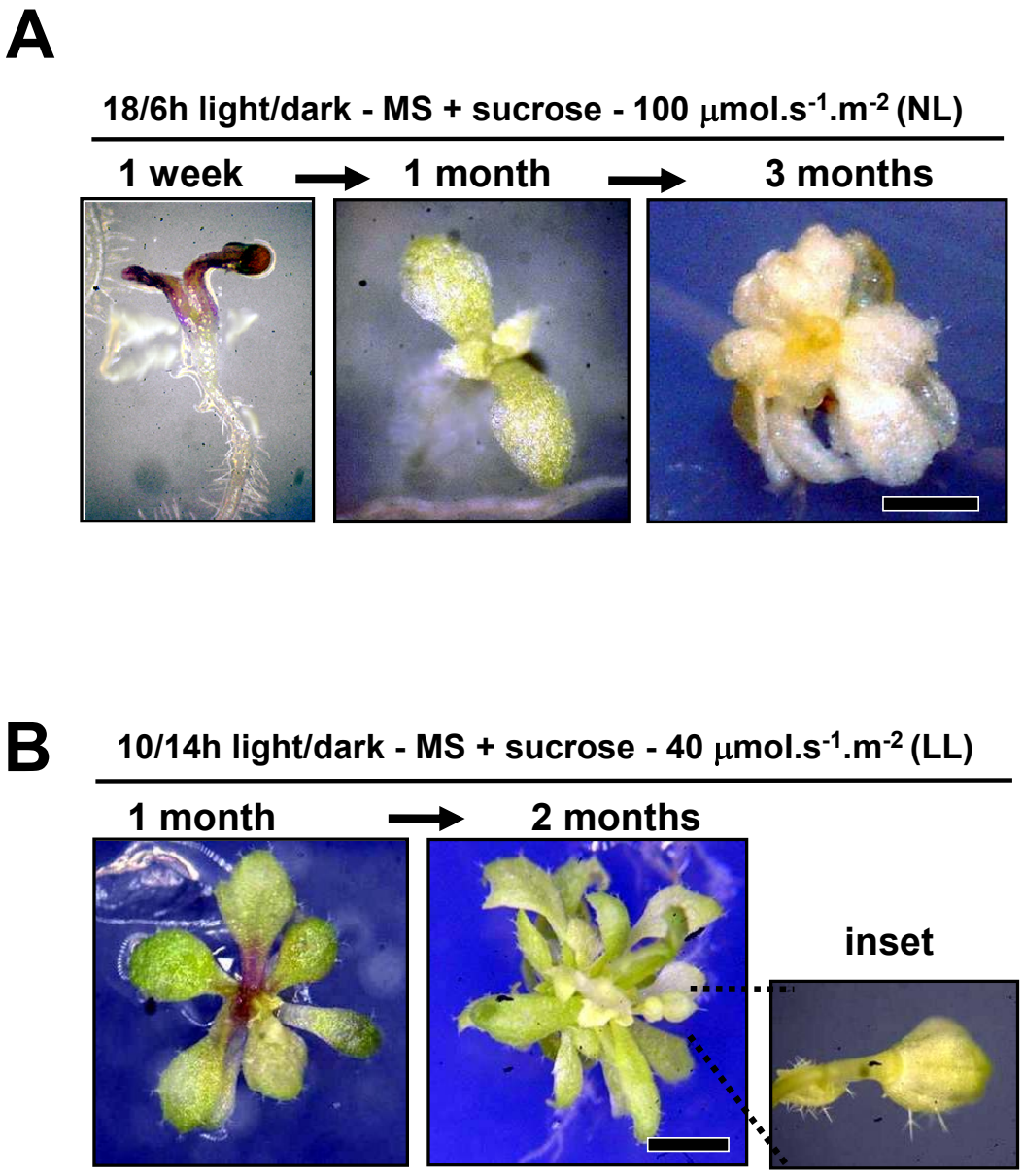


Figure 2.7. Seedling development rescue of homozygous *clpr4-1* under different growth conditions.
(A) Development of *clpr4-1* on agar plates with MS medium, supplemented with 2% sucrose grown under an 18h/6h light/dark cycle at 100 μmol photons.m^{-2}.s^{-1} (NL). The seedlings were initially red from anthocyanin accumulation, and subsequently developed into white plantlets with a rosette diameter of about 10 mm after 3 months. These seedlings never bolted and eventually died off.
(B) Development of *clpr4-1* on agar plates with MS medium, supplemented with 2% sucrose grown under a 10/14h light/dark cycle at 40 μmol photons.m^{-2}.s^{-1} (LL). The seedlings developed into light-green rosettes that formed sterile flowers after about 2 months. The rosette diameter was maximally about 15 mm.

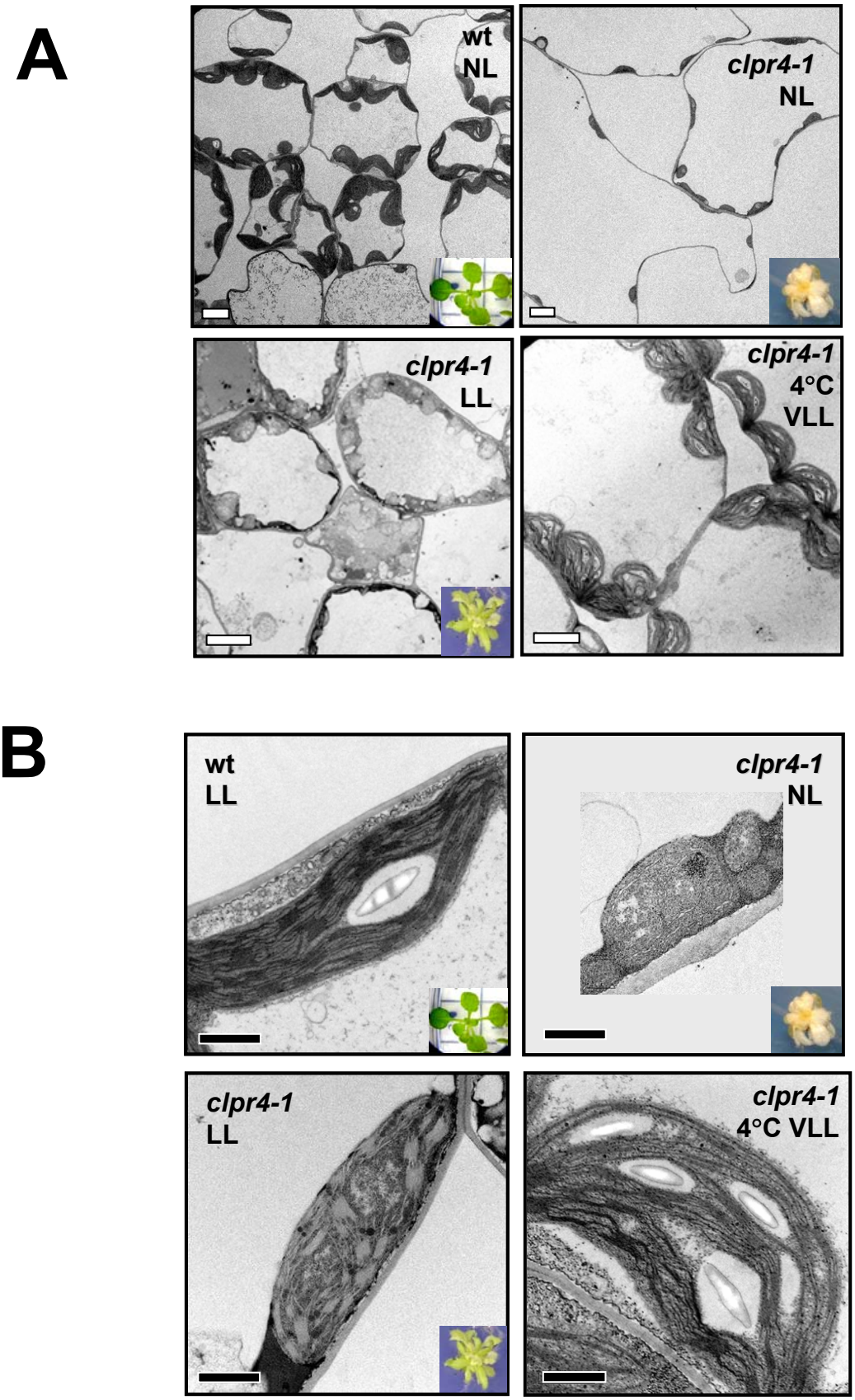
Attempts to rescue the seedlings by supplementing the growth medium with Gamborg's vitamin mix, mevalonic acid, and isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP) were not successful (for an explanation, see discussion). Thus (near) loss of *CLPR4* expression severely impacts seedling development and seedlings can only develop under low light, short day-length and heterotrophic conditions. However, plants are not fertile and, even after prolonged growth (3 months), remain very small.

Light microscopy and TEM analysis of clpr4-1 mesophyll cells and chloroplasts

To understand the effects on cell and chloroplast development in *clpr4-1*, the anatomy of leaf cross-sections and the chloroplast ultra-structure were investigated by respectively light- and transmission electron microscopy (TEM) (Figure 2.8A,B). White leaves (NL grown) of *clpr4-1* contained much larger and vacuolated, but fewer cells as compared to young emerging wild type leaves (NL grown) (Figure 2.8A). The cell size of light green (LL grown) *clpr4-1* was similar to wild type (NL). In addition, chloroplasts size and number were strongly reduced in white leaves, and to a lesser extent also in the light green leaves (Figures 2.8A). Interestingly, when light green plants (LL) were shifted to 4°C at ~10 μmol photons $.m^{-2}.s^{-1}$ (VLL) for 3 weeks, their leaves darkened to almost wild type-levels. These *clpr4-1* plants were also subjected to microscopic analysis and surprisingly chloroplast size and morphology were like wild type (Figure 2.8A).

TEM at higher magnification of *clpr4-1* white leaves showed that thylakoids were nearly absent, present only in a rudimentary form. The chloroplasts from light green *clpr4-1* leaves showed accumulation of stacked but disorganized thylakoid membranes and accumulation of a higher number of plastoglobules, visible as electron dense particles (Figure 2.8B). Surprisingly, thylakoid structure became like wild type in leaves kept under VLL (Figure 2.8B).

Figure 2.8. Light microscopy and TEM of *clpr4-1* leaves at different light fluencies.
(A) A wild type thick section of mesophyll cells from young leaves is compared to corresponding tissue for *clpr4-1*, grown under normal light conditions (100 μmol photons.m^{-2}.s^{-1}, NL), low light (40 μmol photons.m^{-2}.s^{-1}, LL), or shifted to very low light (<10 μmol photons.m^{-2}.s^{-1}, VLL) at 4°C for 3 weeks. Scale bars are 5 μm.
(B) A wild type thin section of plastids from mesophyll cells from young leaves is compared to *clpr4-1* plastids, grown under normal light conditions (NL), low light (LL), and or shifted to very low light (<10 μmol photons.m^{-2}.s^{-1}, VLL) at 4°C for 3 weeks. Scale bars are 1 μm.


A
wt
NL
clpr4-1
NL
clpr4-1
LL
clpr4-1
4°C
VLL
B
wt
LL
clpr4-1
NL
clpr4-1
LL
clpr4-1
4°C VLL

clpr4-1 and clpp3-1 do not respond to phototropic cues during greening

Chloroplast development in *clpr4-1* is highly affected by growth light intensity. To study whether *clpr4-1* shows any phototropic defects, seeds from heterozygote *clpr4-1* and wild type were germinated in the dark and subsequently exposed to unidirectional white light (Figure 2.9). Wild type and wild type-like seedlings rapidly greened and re-orientated toward the source of the light. In contrast, homozygous *clpr4-1* did not respond to the directional light cue and, although they de-etiolated and displayed some greening under these low light conditions, they kept growing in the original direction (Figure 2.9). The same experiment was repeated for *clpp3-1* and gave similar results (not shown). Apparently in *clpr4-1* and *clpp3-1*, disruption of chloroplast development has also adverse effects on phototropic responses.

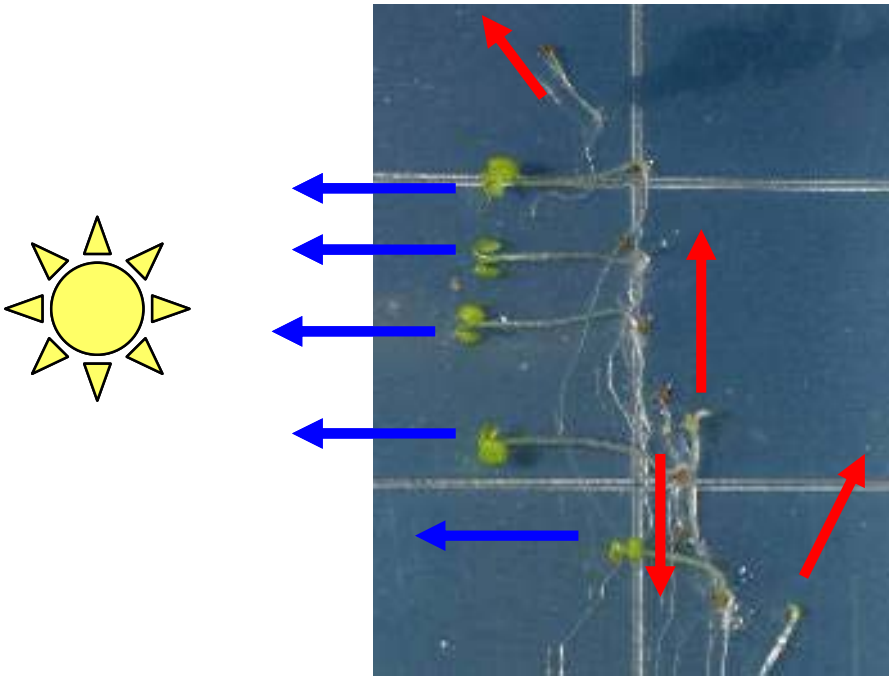


Figure 2.9. Phototropic response of *clpr4-1* seedlings.
Segregating seeds from heterozygous *clpr4-1* plants were germinated in the dark on agar plates supplemented with sucrose. Subsequently, the plates were exposed to white fluorescent light perpendicular to the direction of growth (horizontal vs vertical). Re-orientation of deetiolated seedlings is marked by arrows in blue for wild type-like and in red for *clpr4-1*.

Light microscopy and TEM analysis of primary root cell layers in clpr4-1

It is known that the ClpP/R proteins accumulate in roots as a 325-350 kDa complex (Peltier et al., 2004b). Therefore, the primary roots of homozygous *clpr4-1* plants were analyzed by colleagues from the Cornell Integrated Microscopy Center (CIMC) for plastid phenotypes. Figure 2.10A,B shows TEM photos of mitochondria and plastids in the root cortex and in root vascular cells of wild type and *clpr4-1* (NL). The size and shape of the root plastids did not seem significantly affected. Prothylakoid accumulation in plastids of root cortex cells was quite comparable between wild type and *clpr4-1*, possibly with a modest increase in plastoglobule accumulation in the *clpr4-1* plastids. Mitochondria ultra-structure was also unaffected in *clpr4-1*

Clp genes expression analysis in clpr4-1

To understand if loss of expression of *CLPR4* affects the expression of other members of the Clp family, RT-PCR was performed from transcripts extracted from leaves of *clpr4-1*. Amplification could readily be detected for all the Clp genes tested (*CLPP3,4,5*; *CLPR1,2*; and *CLPS1,2*) (not shown).

The plastid Clp core proteins have so far always been found together in a single complex in plastids of different tissues (Peltier et al., 2004b). However, the presence of different complexes in different cell types within one tissue type (e.g. root, leaf, petal) cannot be excluded. A way to address this question is the use of endogenous promoters fused to reporter genes, such as β-glucoronidase (GUS). Therefore, stable transgenic plants were created harboring *promoter:GUS* fusions for *CLPR2* (*pR2:GUS*) and *CLPR4* (*pR4:GUS*). Preliminary histochemical analysis of GUS activity revealed that the expression patterns of these two constructs at the young seedling stage is rather similar (Figure 2.11). Expression of *CLPR2* and *CLPR4* appears constitutive throughout the aerial part of the plantlets as compared to wild type background levels. Expression in the roots was concentrated in the main vasculature, possibly with higher expression levels of

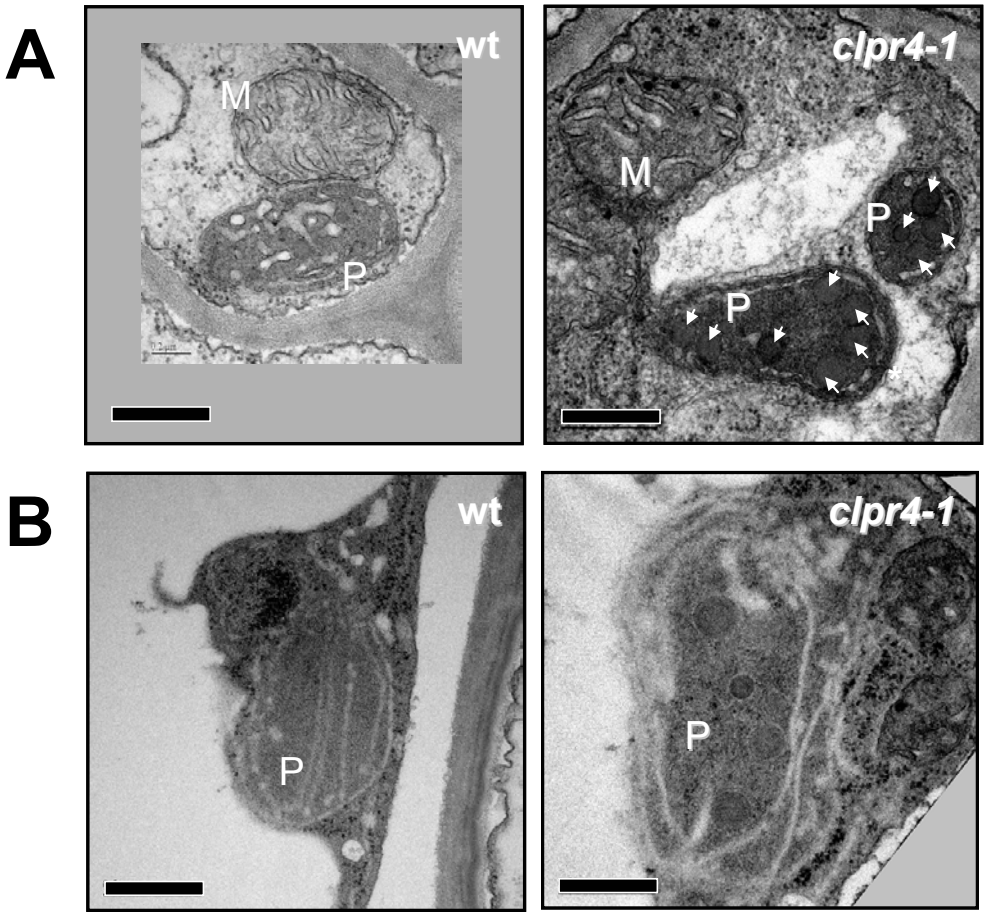


Figure 2.10. TEM of *clpr4-1* roots.
(A) Thin sections of primary root vasculature showing plastids (P) and mitochondria (M) for wild type and *clpr4-1* tissues grown under a 10h/14h light/dark cycle at 40 μmol photons.m^{-2}.s^{-1} (LL). In plastids of *clpr4-1*, many electron dense particles are visible, most likely representing plastoglobuli (arrowheads).
(B) Thin sections of primary root cortex showing plastids (P) for wild type and *clpr4-1* tissues. Scale bars are 0.5 μm.

pR2:GUS in the root tip. No expression could be detected for *pR4:GUS* in the inflorescence branch leading to the flower buds. Clearly there might be situations and/or cell types in which not all the *CLPP/R* genes are expressed at the same time. A more comprehensive and detailed analysis is needed. To this end, undergraduate Kieren Patel - in the van Wijk lab under my supervision - created stable transgenic lines harboring *promoter:GUS* fusions for all the plastid *CLPP/R/S/C/D/T* genes (unpublished results).

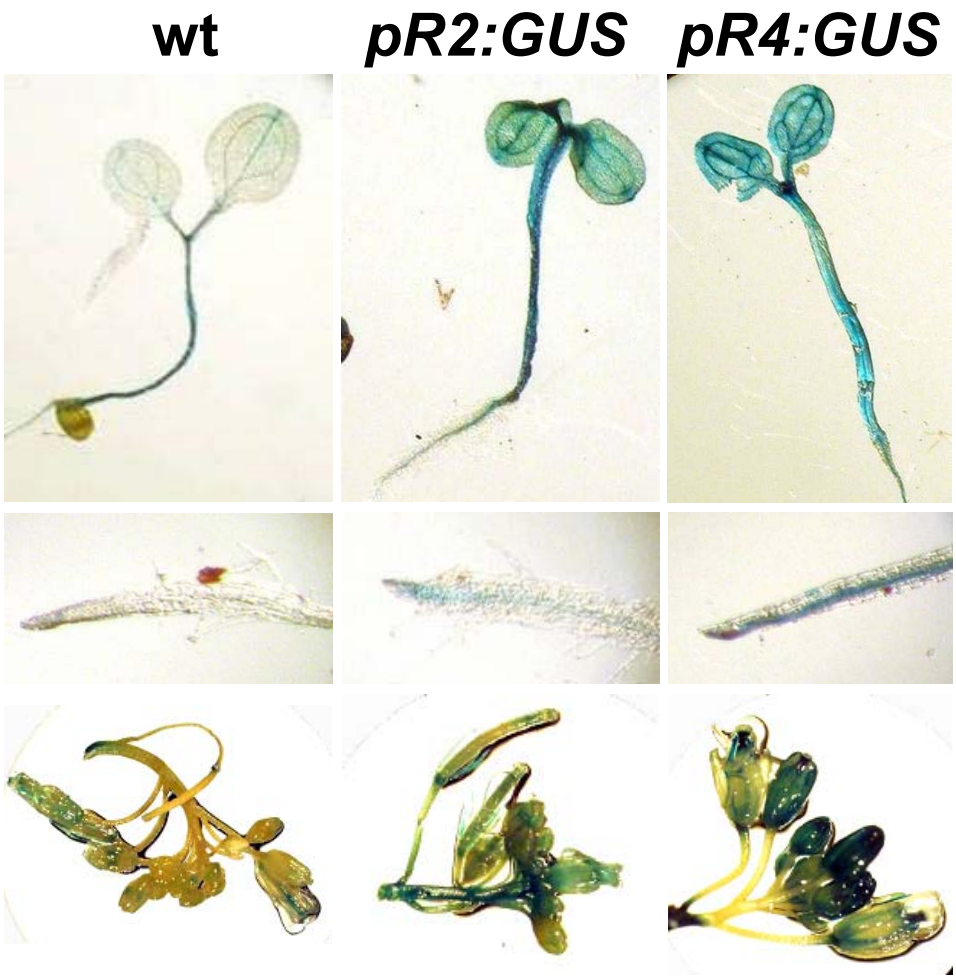


Figure 2.11. *CLPR2* and *CLPR4* expression analysis by *promoter:GUS* fusions
Histochemical GUS staining in transgenic seedlings carrying a *CLPR2* or *CLPR4 promoter:GUS* fusion (*pR2:GUS* and *pR4:GUS* respectively) as compared to wild type plants. Middle panels show a close up of root tips and bottom panels staining of inflorescences.

clpr4-1 leaves are specifically depleted in chloroplast proteins

The microscopy of white and green *clpr4-1* leaves clearly showed that chloroplast size and number per cell were strongly reduced in *clpr4-1*, particularly in the white leaves. To determine if and how this was reflected in the protein composition of the leaf, total leaf proteome (soluble and membrane proteins) was extracted from wild-type and white *clpr4-1* leaves (NL), all grown on agar plates. A titration of leaf proteomes was loaded on a 1-D SDS-PAGE gel and stained with the fluorescent dye Sypro Ruby, for maximal

linearity. This showed that the accumulation levels of Rubisco and LHCII, the two most abundant chloroplast and leaf proteins, are strongly reduced in *clpr4-1* (Figure 2.12A).

To better visualize the difference in proteomes between wild type and *clpr4-1*, the total soluble leaf proteome extracts from wild type and *clpr4-1* white leaves (NL) were compared using native 2D gels (Figure 2.12B-C). The Rubisco complex and many other abundant metabolic enzymes are virtually absent in the mutant. Only the spot containing Chaperone 60 (Cpn60) could be assigned as a plastid protein complex in the *clpr4-1*, with accumulation levels surprisingly similar to wild type. The 26S proteasome, mitochondrial ATP synthase complex, cytosolic methionine synthase (MS1,2) and vacuolar thioglucoside glucohydrolase 1 (TGG1) - as indicators of the rest of the cellular proteome – all accumulated to normal levels. This showed that mainly the chloroplast proteome was affected in *clpr4-1*, in agreement with microscopic analysis described above.

'Shotgun' proteomics of total leaf extracts from green and white clpr4-1 plants

To obtain a more in-depth understanding of the leaf proteome in *clpr4-1* and to determine if cellular functions other than those in chloroplasts were affected, we compared the total proteomes of wild type (NL), light green (LL) and white (NL) *clpr4-1* leaves. Total leaf proteome of wild type and the white and light green plantlets were extracted with SDS and digested in DMSO solution with trypsin. After digestion and sample clean-up, the three peptide extracts were analyzed by nanoLC-ESI-MS/MS, injecting equal amounts of protein extracts. In total 80 proteins were identified (111 when including ambiguous hits to various homologues) and they are reported in Table 2.3. Protein identities were assigned to different functional categories using the MapMan Bin classification system (http://gabi.rzpd.de/projects/MapMan/, Thimm et al., 2004) and if possible we also classified the subcellular location based on experimental information from published data, TAIR annotation or based on subcellular prediction with TargetP and Predotar (http://urgi.infobiogen.fr/predotar/predotar.html).

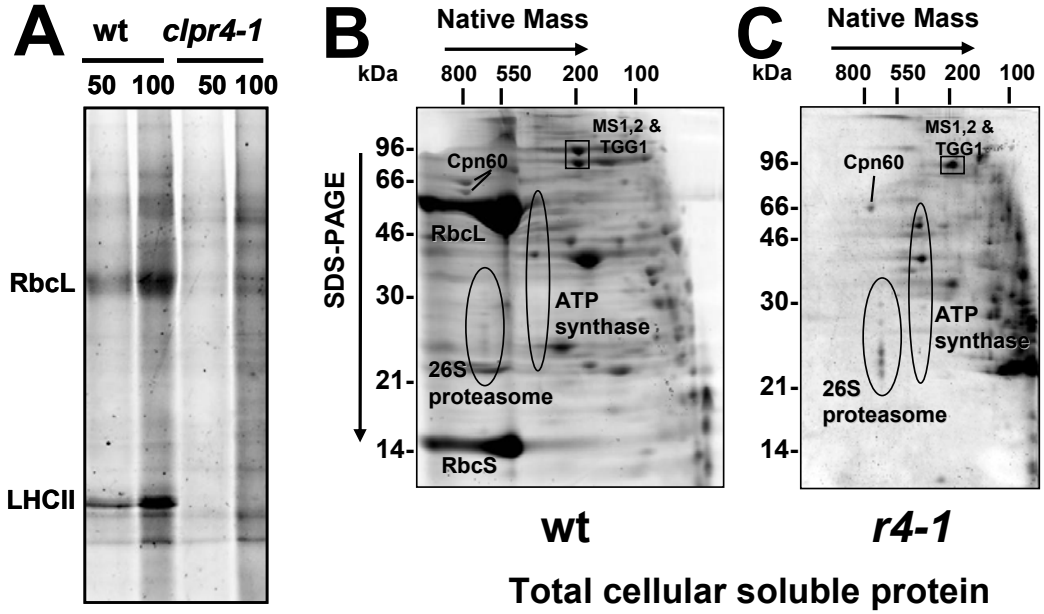


Figure 2.12. Protein profiles of *clpr4-1* leaves
(A) Sypro stained 1-D SDS-PAGE of total leaf proteome extracted from wild type and *clpr4-1* seedlings grown on MS medium in agar supplemented with 2% sucrose grown under an 18h/6h light/dark cycle at 100 μmol photons.m^{-2}.s^{-1} (NL). 50 or 100 μgrams of proteins loaded as indicated.
(B-C) CN-PAGE and Sypro Ruby staining of 2 mg of total soluble proteins from leaves of wild type **(B)** and *clpr4-1* **(C)** seedlings grown as in (A, NL). In wild type, the predominant chloroplast stromal proteins are the Rubisco large and small subunits. Chaperone 60 and the extra-chloroplastic 26S proteasome complex and mitochondrial ATP synthase complex (ovals) and cytosolic methionine synthase 1,2 (MS1- At5g17920; MS2 - At3g03780) and vacuolar thioglucoside glucohydrolase 1 (TGG1; At5g20980) are indicated.

The overlap among the different proteomes was surprisingly small (Figure 2.13A). Only 5 proteins were shared and are chloroplast Rubisco activase, transketolase 1, mitochondrial elongation factor 1-alpha, ATP synthase beta chain and subunit 1. Glutathione S-transferase Phi (AtGSTF2) and glyceraldehyde-3-phosphate dehydrogenase C-1,2 (GapC-1,2) were unique to the *clpr4-1* genotype; mitochondrial malate dehydrogenase [NAD], a tumor protein homologue and HSP81 were shared among NL-grown plants (wild type and white *clpr4-1*) and 11 chloroplastic proteins (LHCII-1, Large and Small subunits of Rubisco, Fructose-bisphosphate aldolase-1 and -2 (SFBA-1,2), Phosphoglycerate kinase-1 (PGK-1), Elongation factor Tu-1 (EF-Tu-1),

oxigen evolving complex 33 kD subunit (OEC33), ATP synthase coupling factor subunit α (CF1α), and photosystem II core subunit D2) plus cytosolic dehydroascorbate reductase-1,2 (DHAR1,2) were shared among green plants (wild type and light green *clpr4-1*).

Table 2.3. Shot-gun proteome analysis of leaves from wild-type plants and from green and white *clpr4-1* leaves. Wild-type plants and white *clpr4-1* were grown under normal light conditions (NL) at 100 μmol/m^2.sec at long day length (16 hrs), while green *clpr4-1* were grown at low light (LL) and short day lenght (40 μmol/m^2.sec; 10 hrs). Total leaf proteome was in-solution digested with trypsin and peptides were analyzed by nano-LC-MS/MS followed by ATH1v6 search using Mascot.

Gene[a]	WT	*clpr4-1*		Location[b]	Annotation[c]	MapManBin[d]	Ambiguity[e]
		LL	NL				
At1g15820	X			P	LHCII-6 - CP24	1.1.1.1 PS.lightreaction.photosystem II.LHC-II	
At5g01530	X			P	LHCII-4.1-CP29	1.1.1.1 PS.lightreaction.photosystem II.LHC-II	
At3g08940	X			P	LHCII-4.2 - CP29	1.1.1.1 PS.lightreaction.photosystem II.LHC-II	
At5g66570	X	X		P	psbO OEC33	1.1.1.2 PS.lightreaction.photosystem II.PSII polypeptide subunits	At3g50820
AtCg00020	X			P	psbA D1	1.1.1.2 PS.lightreaction.photosystem II.PSII polypeptide subunits	
At1g06680	X			P	psbP OEC23	1.1.1.2 PS.lightreaction.photosystem II.PSII polypeptide subunits	
At4g05180	X			P	psbQ OEC16-like	1.1.1.2 PS.lightreaction.photosystem II.PSII polypeptide subunits	
At1g29930	X	**X**[(*)]		P	LHCII-1.3,2,1	1.1.1.2 PS.lightreaction.photosystem II.PSII polypeptide subunits	At1g29910 At1g29920
AtCg00280	X			P	psbC CP43	1.1.1.2 PS.lightreaction.photosystem II.PSII polypeptide subunits	
At4g21280	X			P	psbQ OEC16	1.1.1.2 PS.lightreaction.photosystem II.PSII polypeptide subunits	
AtCg00270	X	X		P	psbD D2	1.1.1.2 PS.lightreaction.photosystem II.PSII polypeptide subunits	
AtCg00680	X			P	psbB CP47	1.1.1.2 PS.lightreaction.photosystem II.PSII polypeptide subunits	
At1g61520	X			P	LHCI-3 - LHCI-680A CAB4	1.1.2.1 PS.lightreaction.photosystem I.LHC-I	
At1g03130	X			P	psaD-1,2 subunit II	1.1.2.2 PS.lightreaction.photosystem I.PSI polypeptide subunits	At4g02770
At3g16140	X			P	psaH-1,2 - subunit VI	1.1.2.2 PS.lightreaction.photosystem I.PSI polypeptide subunits	At1g52230
AtCg00480	X			P	CF1b - atpB	1.1.4 PS.lightreaction.AtP synthase	
AtCg00120	X	X		P	CF1a - atpA	1.1.4 PS.lightreaction.AtP synthase	
At5g66190		X		P	FNR-1	1.1.5.3 PS.lightreaction.other electron carrier (ox/red).ferredoxin reductase	

Table 2.3 (continued)

AtCg00490	X	X		P	Rubisco large subunit (RBCL)	1.3.1 PS.calvin cyle.rubisco large subunit	
At2g39730	X	X	X	P	Rubisco activase	1.3.13 PS.calvin cyle.rubisco interacting	
At1g67090	X	X		P	Rubisco small subunit-4 (RBCS-4)	1.3.2 PS.calvin cyle.rubisco small subunit	
At5g38420	X	X		P	Rubisco small subunit 2,3b (RBCS-2,3b)	1.3.2 PS.calvin cyle.rubisco small subunit	At5g38410
At3g12780	X	X		P	phosphoglycerate kinase-1 (PGK-1)	1.3.3 PS.calvin cyle.phosphoglycerate kinase; 4.10 glycolysis.phosphoglycerate kinase; 4.10 glycolysis.phosphoglycerate kinase	
At3g26650	X			P	glyceraldehyde 3-phosphate dehydrogenase A-1,2 (GAPA-1,2)	1.3.4 PS.calvin cyle.GAP	At1g12900
At1g42970		X		P	glyceraldehyde-3-phosphate dehydrogenase B (GAPB)	1.3.4 PS.calvin cyle.GAP	
At4g38970	X	X		P	fructose-bisphosphate aldolase-2 (SFBA-2)	1.3.6 PS.calvin cyle.aldolase; 4.07 glycolysis.aldolase	
At2g21330	X	X		P	fructose-bisphosphate aldolase-1 (SFBA-1)	1.3.6 PS.calvin cyle.aldolase; 4.07 glycolysis.aldolase	At2g21330.3
At3g60750	X	X	X	P	transketolase-1 (tKL-1)	1.3.8 PS.calvin cyle.transketolase; 7.2 OPP.non-reductive PP	
At5g06860			X	O	polygalacturonase inhibiting protein 1 (PGIP1)	10.6.3 cell wall.degradation.pectate lyases and polygalacturonases	
At5g07440			X	M	NAD(H)glutamate dehydrogenase 2 (GDH1)	12.3.1 N-metabolism.N-degradation.glutamate dehydrogenase	
At3g22200			X	M	gamma-aminobutyrate transaminase subunit (POP2)	13.1.1.1 amino acid metabolism.synthesis.central amino acid metabolism.GABA	
At1g17290		X		M	alanine aminotransferase, putative	13.1.1.3 amino acid metabolism.synthesis.central amino acid metabolism.alanine; 13.2.4.4 degradation.branched-chain group.leucine; 13.2.4.5 degradation.branched chain group.isoleucine	
At3g58610			X	P	ketol-acid reductoisomerase	13.1.4.1 amino acid metabolism.synthesis.branched chain group.common	

Table 2.3 (continued)

At1g11860		X		M	aminomethyltransferase-related precursor protein	13.2.5.2 amino acid metabolism.degradation.serine-glycine-cysteine group.glycine	
At5g01600		X		P	Ferritin-1	15.2 metal handling.binding, chelation and storage	
At3g07390		X		O	auxin-responsive protein / auxin-induced protein (AIR12)	17.2.3 hormone metabolism.auxin.induced-regulated-responsive-activated	
At3g45140	X			P	lipoxygenase AtLOX2	17.7.1.2 hormone metabolism.jasmonate.synthesis-degradation.lipoxygenase	
At1g75040		X		O	pathogenesis-related protein 5 (PR-5)	20.1 stress.biotic	
At5g26000		X		O	tGG1, tHIOGLUCOSIDE GLUCOHYDROLASE 1 (myrosinase)	20.1 stress.biotic	
At3g12500			X	O	basic endochitinase	20.1 stress.biotic	
At5g56010	X		X	O	heat shock proteins -HSP81/ Hsp90	20.2.1 stress.abiotic.heat	At5g56030 At5g56000 At5g52640
At5g20630	X			O	germin-like protein (GER3)	20.2.99 stress.abiotic.unspecified	
At5g42980			X	O	thioredoxin H-type 3 (tRX-H-3)	21.1 redox.thioredoxin	
At1g19570	X	X		O	dehydroascorbate reductase-1,2 (DHAR1,2)	21.2 redox.ascorbate and glutathione	At1g19550
At5g64120		X		O	peroxidase, putative	26.12 misc.peroxidases	
At2g38380			X	O	peroxidase 22 (PER22) (P22) (PRXEA) / basic peroxidase E	26.12 misc.peroxidases	
At3g32980			X	O	peroxidase 32 (PER32) (P32) (PRXR3)	26.12 misc.peroxidases	
At3g49110			X	O	peroxidase 33 (PER33) (P33) (PRXCA) / neutral peroxidase C (PERC)	26.12 misc.peroxidases	
At3g49120			X	O	peroxidase, putative	26.12 misc.peroxidases	
At4g27520			X	O	plastocyanin-like domain-containing protein	26.19 misc.plastocyanin-like	
At3g18070		X		O	glycosyl hydrolase family 1 protein	26.3 misc.gluco-, galacto- and mannosidases	At3g18080

Table 2.3 (continued)

At4g22240			X	P	fibrillin (FIB1b)	26.31* misc. fibrillins	
At3g44300			X	O	nitrilase 2 (NIt2)	26.8 misc.nitrilases, *nitrile lyases, berberine bridge enzymes, reticuline oxidases, troponine reductases	
At4g02520		X	X	O	glutathione transferase Phi (AtGStF2)	26.9 misc.glutathione S transferases	
At1g78380			X	O	putative glutathione transferase	26.9 misc.glutathione S transferases	
At3g54400		X		O	aspartyl protease family protein	27.3.99 RNA.regulation of transcription.unclassified	
At1g22530		X		O	SEC14 cytosolic factor family protein / phosphoglyceride transfer family protein	28.99 DNA.unspecified; 34.99 transport misc	At1g72150
At3g27830		X		P	50S ribosomal protein L12-A,C	29.2.1 protein.synthesis.chloroplast - plastid ribosomal protein	At3g27850
At1g13930		X		O	60S ribosomal protein	29.2.2 protein.synthesis.misc ribososomal protein	
At2g20450			X	O	60S ribosomal protein L14 (RPL14A,B)	29.2.2 protein.synthesis.misc ribososomal protein	At4g27090
At2g37190			X	O	60S ribosomal protein L12 (RPL12A,B,C)	29.2.2 protein.synthesis.misc ribososomal protein	At3g53430 At5g60670
At1g07920	X	X	X	M	elongation factor 1-alpha / EF-1-alpha	29.2.4 protein.synthesis.elongation	At1g07940 At5g60390 At1g07930
At4g20360	X	X		P	elongation factor tu (EF-tu-1)	29.2.4 protein.synthesis.elongation	
At5g20720		X		P	Cpn21 (also Cpn20)	29.6* protein.(un)folding (includes isomerases and chaperones - if not heat induced)	
At2g01210	X			O	leucine-rich repeat transmembrane protein kinase, putative	30.2.3 signalling.receptor kinases.leucine rich repeat III	
At1g04820	X			O	tubulin alpha	31.1 cell.organisation	At1g50010 At5g19780 At5g19770 At4g14960
At1g78900			X	O	AtPase 70 kDa	34.1 transport.p- and v-AtPases	
At5g15090		X		M	voltage-dependent anion-selective channel protein hsr2	34.18 transport.unspecified anions	
At1g45201		X		O	similar to lipase class 3 family protein	35.1 not assigned.no ontology	

Table 2.3 (continued)

At3g14990			X	O	4-methyl-5(b-hydroxyethyl)-thiazole monophosphate biosynthesis protein, putative	35.1 not assigned.no ontology	
At4g20260			X	O	DREPP plasma membrane polypeptide family	35.1 not assigned.no ontology	
At3g16640	X		X	O	tCtP homolog - tumor protein homologue	35.2 not assigned.unknown	
At2g46820	X			P	expressed protein	35.2 not assigned.unknown	
At5g53560			X	O	cytochrome b5	35.2 not assigned.unknown	
At5g03690.2		X		O	fructose-bisphosphate aldolase	4.7 glycolysis.aldolase	At3g52930 At2g36460
At3g04120		X	X	M	glyceraldehyde-3-phosphate dehydrogenase C-1,2 (GapC-1,2)	4.9 glycolysis.glyceraldehyde 3-phosphate dehydrogenase	At1g13440
At1g53240	X		X	M	malate dehydrogenase [NAD]	8.1.9 tCA / org. transformation.tCA.malate DH	At3g15020
At5g08690	X	X	X	M	H+-transporting AtP synthase beta chain	9.9 mitochondrial electron transport / AtP synthesis.F1-AtPase	At5g08670 At5g08680
AtMg01190	X	X	X	M	atp1 AtPase subunit 1	9.9 mitochondrial electron transport / AtP synthesis.F1-AtPase	At2g07698
At2g34420		**X**(**)			LHCII-1.5		

a	Accession number from TAIR v6 (www.arabidopsis.org). (Putative) chloroplast localized entries in bold
b	Localization curated by the van Wijk laboratory, based on experimental evidence and primary literature: P - plastid, M - mitochondria, O – other
c	Annotation curated by the van Wijk laboratory, based on primary literature, TAIR, and functional domain predictors
d	Bins curated by the van Wijk laboratory, based on MapMan data, primary literature, and TAIR
e	Accession numbers of proteins truly ambiguous with the reported first identification

*cTP petide identified. In this case rules out ambiguity. **cTP identified, but Mowse score of 20 is too low for identification confidence level $p<0.05$

The distribution of the identified proteins over the subcellular localizations within each of the three proteomes showed strong differences, with identified plastid proteins representing only 13% in white plants, 49% in the light green plants, and 74% in wild type (Figure 2.13B). Clearly, the relative abundance of chloroplast proteins is much lower in *clpr4-1* leaves than in wild type leaves. Consequently, the most abundant

functions were very different between the three leaf samples: 44% of the proteins identified in wild type were thylakoid membrane proteins involved in photosynthetic electron transport, and 23% were involved in the stromal Calvin cycle. In contrast, in the light green *clpr-4-1* plants, only 14% were part of the thylakoid electron transport chain, and 24% involved in the Calvin cycle. Others were involved in stress-defense (ascorbate and glutathione based defense), a lipase and others. LHCII-1.3 accumulated as a precursor in the mutant while this and other precursor forms are undetectable in wild type, as well as LHCII-1.5, even though the peptide score was too low to count toward a positive identification. In the white leaves, Rubisco activase was the only photosynthesis-related protein. The other chloroplast proteins were transketolase-1 (TKL-1), a ketol-acid reductoisomerase and a fibrillin (FIB1b).

It can be concluded that chloroplast development in *clpr4-1* is impaired, but that photosynthetic proteins can accumulate under low light fluencies.

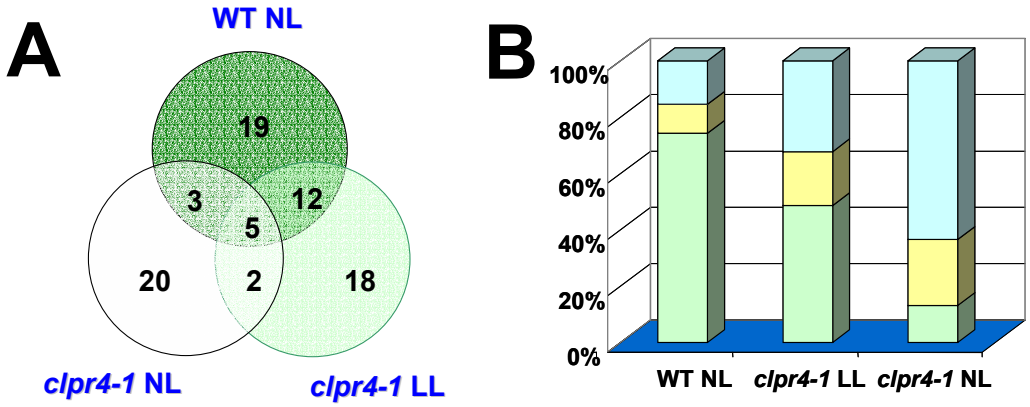


Figure 2.13. 'Shotgun' proteomics analysis of wild type and *clpr4-1* leaves
(A) Venn diagram representation for proteins identified by nano-LC ESI-MS/MS after in-solution digestions for wild type and *clpr4-1* (NL, 18h/6h light/dark cycle at 100 µmol photons.m^{-2}.s^{-1}), and *clpr4-1* (LL, 10h/14h light/dark cycle at 40 µmol photons.m^{-2}.s^{-1}).
(B) Localization of proteins identified in (A). Green: plastid; yellow: mitochondria; cyan: other localization.

Western blot analysis of green clpr4-1 leaves

To investigate the accumulation of chloroplast and thylakoid proteins, the light green *clpr4-1* plants (LL) were analyzed by western blot. To ensure a quantitative analysis, a titration of 3 different concentrations of proteins was used. The data presented in Figure 2.14 shows that accumulation of PSII and PSI core subunits was reduced in *clpr4-1* by at least a factor 10 (on a total leaf protein basis). In contrast, PsbS, an unusual thylakoid membrane protein with pigment binding domains and a central role in non-photochemical quenching (Li et al., 2000), accumulated at near wild type levels. PsbS has been shown to accumulate in absence of PSII complexes, in etioplast membranes and in mutants devoid of pigment accumulation (Funk et al., 1995). The thylakoid protease SppA was 2.5 fold up-regulated on a total leaf protein basis. The Zn-metallo proteases FsH5 (Var1) and FtsH2 (Var2) and the stromal protease Deg2 were also detected in *clpr4-1*, but they were down-regulated by a factor two on a total leaf protein basis. Importantly, Cpn60 was detected at wild type levels (Figure 2.14), confirming the observations from 2D gels (see above). Considering that chloroplast protein content is considerably lower in *clpr4-1* leaves as compared to wild type (see i.e. Figure 2.8), it is clear that Cpn60 must be several fold up-regulated on a chloroplast protein basis. Taking these observations together, loss of *CLPR4* expression also causes an imbalance in chloroplast protein homeostasis.

Null mutants for CLPC1 and CLPD show different phenotypes

T-DNA insertion lines were obtained for members of the chaperone and chaperone-like Clp subfamily. SALK_052772 was confirmed to contain an insertion in the 3rd intron of *CLPS1* and was renamed *clps1-1*. RT-PCR amplification for the *CLPS1* transcript showed wild type-levels (not shown) and led to the conclusion that the T-DNA in *clps1-1* is efficiently spliced out. This results in a 100% germination rate of wild type looking plants. *clps1-1* was therefore not further characterized. SALK_014058 with a confirmed

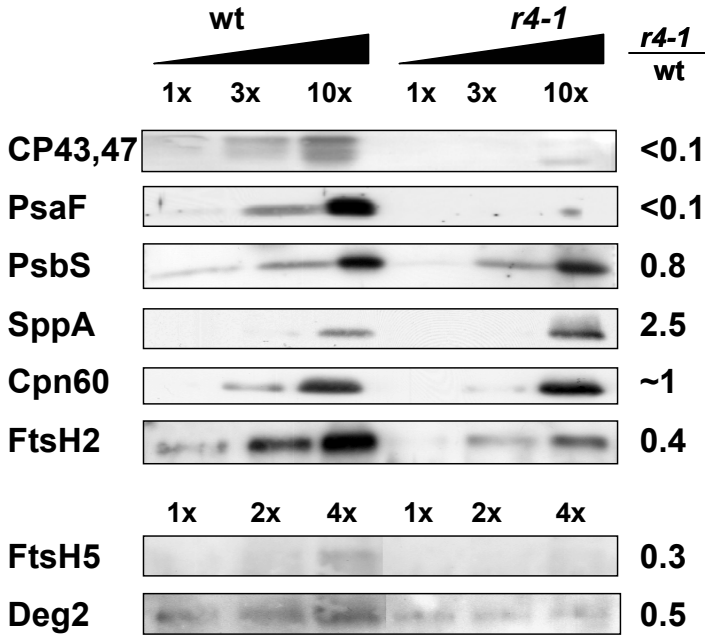


Figure 2.14. Immunoblot analyses of plastid protein populations
To evaluate the effect of reduced ClpR4 expression on the thylakoid protein complexes, as well as proteases and chaperones in stroma and thylakoid, a western blot analysis was carried out. Titrations (indicated as 1x, 2x, 3x, 4x, 10x) of total leaf protein extracts from wild type and *clpr4-1* grown under a 10h/14h light/dark cycle at 40 μmol photons.m^{-2}.s^{-1} (LL) were separated by SDS-PAGE and blotted onto PVDF membranes. Protein ratios in *clpr4-1*/wild type are indicated. Membranes were probed with antibodies generated against different proteins of the photosystem I and II (CP43,47: Core Protein 43 and 47 of PSII; PsaF: a small peripheral subunit of PSI; PsbS a unique antenna protein in PSII), as well as chaperones and proteases (stromal chaperone Cpn60; stromal protease Deg2; thylakoid proteases FtsH2 and 5 - members of the FtsH Zn-metallo-protease family; ATP-independent light induced serine-type thylakoid protease SppA).

T-DNA in the 4th exon of *CLPC1* was renamed *clpc1-1*, whereas JP7_7A07L with confirmed T-DNA in the 7th exon of *CLPD* was renamed *clpd-1* (Figure 2.15A). Both lines germinated on soil, but while *clpd* was wild type-like, *clpc1-1* grew as a pale green plant (Figure 2.15B). RT-PCR analysis showed that both lines are null mutants for their respective transcripts (Figure 2.15C). In contrast to what we have seen for the *CLPP/R* genes (above), it appears that a certain degree of functional redundancy is present for the Clp chaperone subunits, with ClpC1 being more important than ClpD.

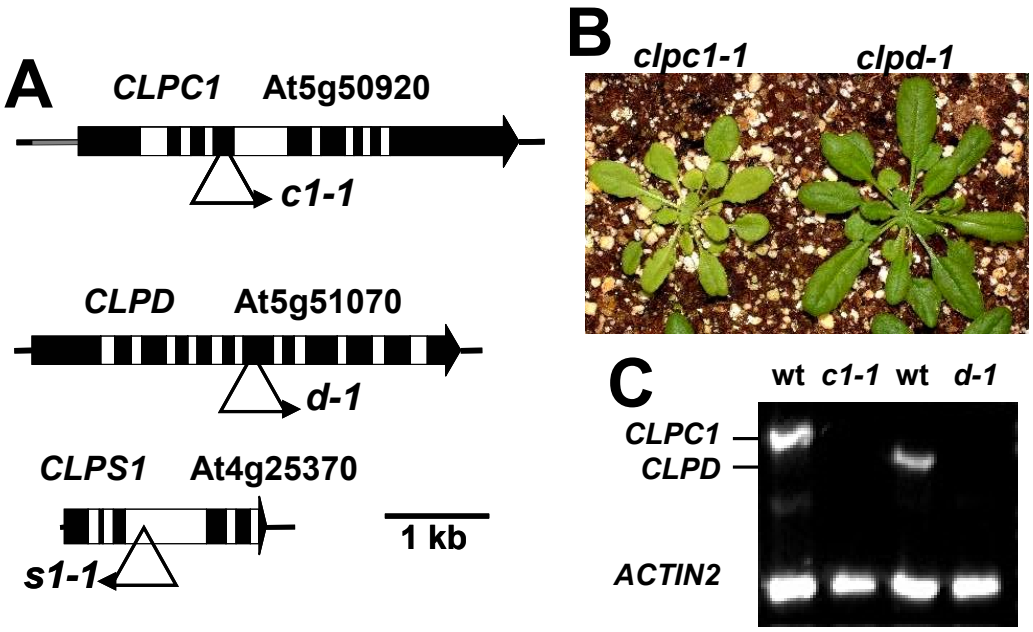


Figure 2.15. Isolation of T-DNA insertion mutants for *CLPC1*, *CLPD* and *CLPS1*.
(A) Gene models for *CLPC1*, *CLPD* and *CLPS1*. Exons as black boxes, intron empty boxes, untranslated regions as a black line or grey line if intron. TAIR accession numbers are indicated. T-DNA insertions (left border indicated as arrow-heads) are in 4th exon of CLPC1, 7th exon of CLPD and 3rd intron of CLPS1. *clps1-1* does not affect mRNA accumulation and does not have a phenotype (not shown).
(B) *clpc1-1* germinates on soil and develops as a pale-green plant as compared to wild type. *clpd-1* germinates on soil and develops as a wild type-like plant. Plants were grown for 3 weeks under continuous light at 100 μmol photons.m^{-2}.s^{-1}.
(C) RT-PCRs of mRNA isolated from leaves of wild-type, *clpc1-1* and *clpd-1* reveal null mutations in *CLPC1* and *CLPD,* respectively.

DISCUSSION

The Clp machinery in *A. thaliana* plastids is present in non-photosynthetic plastids in roots and petals, as well as in chloroplasts. It consists of a tetradecameric 350 kDa complex with two heptameric rings together containing gene products of five serine-type *CLPP* (*1,3-6*) proteases and four *CLPR* (1-4) genes. Associated with the complex are two small ClpS (S1,S2) proteins that likely have a regulatory function. The chaperones ClpC1,2 and D accumulate in the chloroplast and likely form hexameric rings before

docking onto the catalytic core to deliver substrates (Peltier et al., 2004b), as observed in *E. coli* (Sauer et al., 2004). ClpP1 is encoded by the plastid genome, while all others are encoded by the nuclear genome. The ClpR proteins lack the three conserved catalytic amino acid residues (Adam and Clarke, 2002). Particularly intriguing are the roles of the ClpR and ClpS proteins, unique to photosynthetic organisms and present in every Clp core complex analyzed so far. Why the Clp machinery has evolved to such a complexity in plastids is currently unknown.

The experiments presented in this chapter aim at addressing the functional roles of the nuclear-encoded plastid localized Clp proteins and to determine if there are redundancies within these ClpP/R and Clp chaperone subfamilies.

CLPP and CLPR genes do not show redundancy

Taking advantage of the T-DNA insertion collection established at the Salk Institute, we isolated, confirmed and characterized mutants with various degrees of reduced expression for *CLPP3,4,5* and *R1,2,4*. Full complementation was achieved for *clpr1-1* and *r2-1*, while segregation analysis for others (*clpp3-1, p4-1, p5-1* and *r4-1*) suggested linkages between genotype and phenotype. Complementation of the mutant lines *clpp3-1, p4-1, p5-1* and *r4-1* and isolation of additional, independent alleles are in progress by the van Wijk lab, with the objective to more firmly confirm linkages between gene disruption and phenotype. All heterozygous lines have a wild type phenotype.

The phenotypes of these six insertion lines showed strong positive correlation between the reduction in mRNA levels and reduction of viability, irrespective of the gene - this is summarized in Figure 12.16. The phenotypes range from embryo lethal (for null mutants) to light sensitive, sterile, obligate heterotrophic, white seedlings (mRNA accumulation below 1%) and to pale green autotrophic plants (mRNA accumulation between 1% and 50%). Our data are consistent with observations for reduced expression of the plastid-encoded *CLPP1* gene in tobacco, where *CLPP1* was essential for shoot

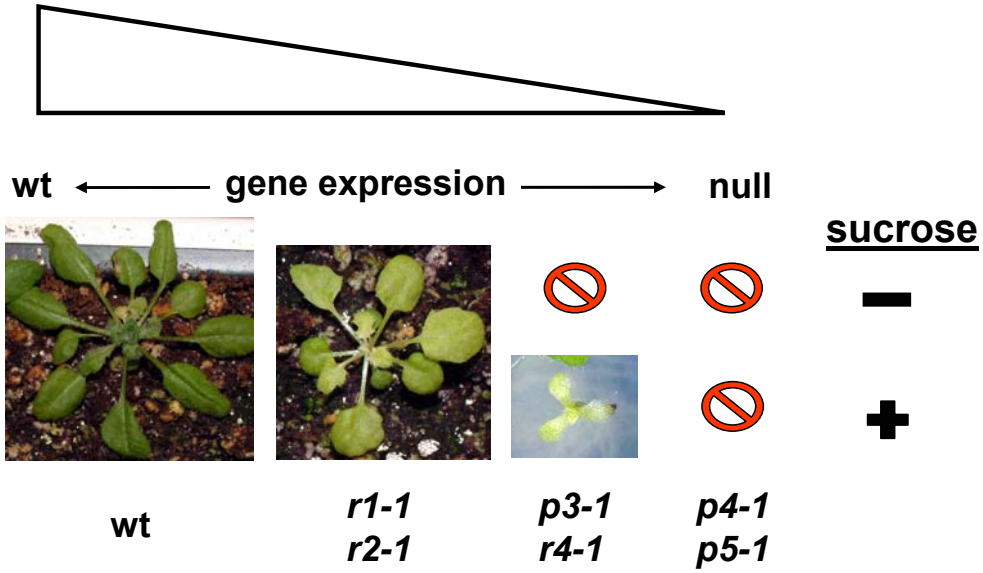


Figure 2.16. Schematic overview of the range of *clpp/r* phenotypes observed in this study.
The phenotypes of the six ClpP/R lines showed strong positive correlation between the reduction in mRNA levels and reduction of viability, irrespective of the gene. The phenotypes range from embryo lethal (for null mutants), to white light sensitive and sterile obligate heterotrophic (mRNA accumulation below 1%) and to pale green autotrophic plants (mRNA accumulation between 1% and 50%). A wt phenotype appears in heterozygous plants. The influence of addition of 2% sucrose on development and phenotype is indicated. This suggests that each ClpP/R protein makes a unique contribution to assembly, structure or function of the ClpPR core, without any obvious redundancies. 'No' symbol indicates that the line is embryo lethal.

development in tobacco while it did not appear to affect root development (Shikanai et al., 2001; Kuroda and Maliga, 2003). This suggests that each of the tested ClpP/R protein makes a unique contribution to assembly, structure or function of the ClpPR core, without any obvious redundancies. This absence of redundancies is consistent with the finding of one ClpPRS complex, without detectable variation in composition in all plastid types analyzed so far (Peltier et al., 2004b).

The nine ClpP/R proteins are quite similar and homology modeling showed several possible structural contributions that may provide an explanation for the absence of redundancy. ClpR1,3,4 have a 9-10 aa insertion loop predicted to face into the central chamber of the ClpPR core - this loop may affect the presentation of substrate to the

ClpPs' catalytic sites within the catalytic chamber. Another feature is the different length of the C-termini that might influence interaction with ClpS or ClpC,D. Last, but not least, the presence of putative openings for proteolytic products release might necessitate specific subunit/subunit interaction (Peltier et al., 2004b). There is additional evidence for a possible substrate release mechanism at the equatorial plane of the Clp core complex: flexibility at the equatorial plane of *Streptococcus pneumoniae* ClpP is important for efficient product release (Gribun et al., 2005; Sprangers et al., 2005). The same authors demonstrated also how loops at the N-termini of ClpP are important in the docking of the chaperone rings (Gribun et al., 2005). Clearly, Clp core subunits are proteins with a high degree of structural functionality.

Embryo lethal lines (clpp4-1, p5-1) and CLPP1-deletion in tobacco

The function of the ClpPR core is proteolysis but the set of substrates is unknown. The embryo lethal phenotype of the ClpP/R null mutants is most likely due to an important function of the Clp complex during embryogenesis. Studies on *CLPP1* gene expression in tobacco plastids strengthen this view; they have either failed to obtain homoplasmic plants in a *CLPP1* deletion mutant or they excised the *CLPP1* gene only after germination using a cre-lox system (Shikanai et al., 2001; Kuroda and Maliga, 2003). It is known that photosynthetic mutants are not embryo lethal - examples are *psad1-1* and *psad2-1* mutants lacking the D-subunit of Photosysthem I (Ihnatowicz et al., 2004), or the *dpa1* mutant lacking one of two ATP-γ subunits (Bosco et al., 2004). This suggests that ClpPR substrates must include proteins with functions other than photosynthesis. Embryo lethal mutants have been observed for null mutants in various categories of plastid proteins, such as Tic32 of the chloroplast import machinery and in lysophosphatidic acid acyltransferase - in both cases the transition from heart to torpedo stage was blocked (Hormann et al., 2004; Yu et al., 2004a). Furthermore, several mutants that are blocked in expression of pentatricopeptide (PPR) proteins are embryo lethal (Cushing et al., 2005).

The embryo lethal phenotype of the Clp mutants suggests that proteolysis by the ClpPR is important for maintaining plastid function beyond photosynthesis.

Light sensitive, sterile, obligate heterotrophic lines (clpr4-1, p3-1)

clpp3-1 and *r4-1* growth could be partially rescued by supplementation of sucrose to the media, but no beneficial effects were observed by supplementation of vitamins and precursors in the isoprenoid biosynthetic pathway. These precursors were tested since a striking similarity in phenotype between *clpp3-1* and *r4-1* and mutants for enzymes in this key secondary metabolism pathway (see for example Gutierrez-Nava Mde et al., 2004 and for reviews on plastid isoprenoids Lange and Ghassemian, 2003; Eisenreich et al., 2004). Without sucrose, seedling development was stalled in the cotyledon stage and true leaves never formed, while under heterotrophic growth conditions and very low light intensities, homozygous seeds developed into very small plants with sterile flowers. The partial rescue of *clpp3-1* and *r4-1* allowed us to analyze these mutants by microscopy as well as at the protein level. We focused our characterization on *clpr4-1* but we anticipate that most findings extend to *clpp3-1*. The growth analysis under various light intensities and the microscopy and protein analysis (western and proteomics) showed that the chloroplasts were extremely light sensitive - it appears that light fluencies might set the threshold limits for minimal ClpPR accumulation to allow thylakoid formation and (some) greening. Shotgun proteome analysis of total white and green *clpr4-1* leaf proteomes indicated a strong reduction in photosynthetic and plastid localized carbon primary metabolism. No Lhcs were detected in white leaves, whereas greener *clpr4-1* leaves accumulated Lhcs precursors, similar as in *clpr2-1*, as will be shown in more detail in chapters 3 and 4.

The appearance of triplicate cotyledons –and the observed lack of phototropic response- in *clpp3-1* and *r4-1* mutants may suggest a perturbation in polar auxin transport or even increased amounts of ABA. These two hormones have been implicated in the

establishment of leaf primordia (for a recent review see Lumba and McCourt, 2005). It is then conceivable that in *clpp3-1* and *r4-1* embryo's shoot apical meristem, pro-cotyledons start forming without the correct distribution of auxin and ABA allowing the emergence of a third cotyledon. ABA is synthesized in the chloroplast via the isoprenoid pathway (Lange and Ghassemian, 2003) and auxin in the cytosol from plastid precursors (reviewed in Grubb and Abel, 2006). It is likely that the appearance of triplicate cotyledons and lack of phototropic response are a pleiotropic phenotype since no direct link with Clp activity can be envisioned.

Pale green autotrophic plants (clpr1-1, r2-1)

Lines c*lpr1-1 and r2-1* have only a partially reduced accumulation of mRNA, which resulted in pale green autotrophic plants with delayed development and reduced rates of biomass accumulation. *clpr2-1* will be extensively analyzed in chapters 3 and 4 and we refer therefore to these chapters. Since it appears that all ClpP/R proteins form a single ClpPR core without redundancy between the different subunits, we expect that lessons learned from the detailed phenotypic analysis on *clpr2-1* are equally relevant for the other *CLPP/R* genes.

The chaperones ClpC and ClpD are functionally redundant

The plastid Clp chaperone subfamily is much simpler than the ClpP/R subfamily and consists in plastids of ClpC1,2 and D. Most likely all three chaperones function to deliver substrates to the ClpPRS core, although formal proof is still lacking. ClpC1,2 proteins also interact with Tic110 and Tic40 of the chloroplast protein translocation machinery in the inner envelope (Akita et al., 1997; Nielsen et al., 1997) and loss of ClpC1 resulted in lower protein translocation rates into isolated chloroplasts (Constan et al., 2004; Kovacheva et al., 2005). ClpD expression was reported to be induced by cold/drought conditions or senescence, but its precise function under such conditions has not been

determined (ClpD is also knows as ERD1, early responsive to dehydration 1)(Kiyosue et al., 1993; Nakashima et al., 1997; Zheng et al., 2002).

The null ClpC1 and ClpD mutants isolated in the present study have surprisingly mild phenotypes. *clpd-1* plants are indistinguishable from wild type plants under normal growth conditions, suggesting redundancy and/or a marginal role for ClpD under optimal growth conditions. It remains possible that *clpd-1* has a phenotype under stress conditions or maybe a mild phenotype during natural or induced senescence. A *CLPD* null mutant without phenotype was reported, but no experimental evidence was provided (Clarke et al., 2005). *clpc1-1* developed slower than wild type and has a pale green phenotype under normal growth conditions. The appearance of a phenotype for ClpC1 is not surprising since it is the most abundant Clp chaperone member and since it likely has at least a function in proteolysis and chloroplast protein import. Null mutants for *CLPC1* have recently been described and displayed similar phenotypes as *clpc1-1* (Constan et al., 2004; Sjogren et al., 2004; Kovacheva et al., 2005). The partial redundancy of ClpC1 is likely due to the expression of the ClpC2 and ClpD homologues (Kovacheva et al., 2005). A mutant with splice site variation in *CLPC2* resulted in sharply reduced accumulation of the ClpC2 protein, but without visible phenotype, likely due to a functional redundancy with the homologues ClpC1 and ClpD (Park and Rodermel, 2004).

We can conclude that, at least under normal growth conditions, ClpC1,2 and ClpD are functionally redundant with ClpC1 being the most prominent member of the Clp chaperone subfamily. It remains to be determined to what extent the three chaperones deliver substrates to the ClpPR core and to what extent they function in import and 'regular' folding and unfolding processes.

Intron splicing can remove T-DNA insertions

The T-DNA used for generating the Salk *Arabidopsis* mutant collection is about 4.5kb long. It inserts in the plant genome as single or multiple copies in tandem and by doing so

it disrupts gene expression (http://signal.salk.edu/tdna_FAQs.html). However, T-DNA insertion in the coding sequence (exons and introns) of a gene does not ensure disruption of gene expression. T-DNA insertions in introns (and possibly also exons) can in principle be spliced out. *clpp5-1*, *clpc1-1* and *clpd-1* have insertions in relatively small exons, and gene expression was completely abolished in each of these lines. *clpp3-1* and *clpr4-1* have insertions in small introns and no transcript was amplified using a single round of amplification - this shows that mRNA accumulation was reduced to less than ~1%. However, it is possible that smaller amounts of messenger RNA are spliced out to provide the plastid with enough Clp protein to grow and develop (but remaining sterile) under heterotrophic conditions at very low light intensities. This is the most plausible explanation for the difference between the small plantlets of *clpp3-1* and *r4-1* and the embryo lethal phenotypes of *clpp4-1* and *p5-1*. We also described *clpr1-1* and *clps1-1* with insertions in introns in the coding region of ~1 kb for *clpr1-1* or over 1kb for *clps1-1*. *clpr1-1* has a reduced gene expression (between 1% and 50% of wt) and the plants are pale green. *clps1-1* has wild type levels of ClpS1 mRNA accumulation and no visible phenotype under normal growth conditions. Thus, there appears to be a correlation between intron size and effects of a T-DNA insertion in such an intron: the larger the intron, the smaller the T-DNA effect on mRNA accumulation. Our data suggest that T-DNA insertion in introns larger than 1 kb can relatively easily be spliced out. However, an analysis of a large population of T-DNA insertion lines is needed to confirm these observations.

CONCLUSIONS

Our data, and information from published literature, suggest that each of the tested ClpP/R protein makes a unique contribution to assembly, structure or function of the ClpPR core, without any obvious redundancies. This absence of redundancies is consistent with the finding of one ClpPRS complex, without detectable variation in

composition in all plastid types analyzed so far (Peltier et al., 2004b). It remains to be determined why each of these ClpP/R proteins is not redundant. In contrast, the ClpC,D chaperone functions are not essential and their relative contributions to proteolysis, import and (un)folding remains to be determined. The observed range of phenotypes for the ClpP/R mutants and the strong impact on embryo development and viability show that the substrates of the ClpPR core must include proteins that contribute to functions other than photosynthesis.

METHODS

Plant growth, mutant isolation and genotyping

A. thaliana (Col 0) wild type and mutant seeds were surface sterilized, vernalized and germinated on MS medium (Sigma) under various light fluencies (as indicated in the result section) at 23°C in controlled growth chambers (Percival). Media when needed were supplemented with 2% sucrose or 50 mg/L kanamycin. Genomic DNA was extracted by grinding leaf tissue in extraction buffer (0.2M Tris-HCl pH 9.0; 0.4M LiCl, 25mM EDTA, 1% (wild type/vol) SDS) followed by isopropanol precipitation 1:1 (v/v) (adapted from (Krysan et al., 1996).

PCR was performed using Platinum Taq DNA polymerase (Invitrogen) and primers used were: *CLPP3* agatctggagatgagtttgcgtctc and gctagcttcaatggcggcataaccat; *CLPP4* gggatgtactagtcgatcgtttgag and cgcaaagccaataccagtaacc; *CLPP5* gggattgtgtttgcaggaagacg and ccagctgtaactgatccacc; *CLPR1* ggcttatggagctcatctctctctt and aattttacgcattgccttacaacct; *CLPR2* cttcattgagatcgtgagattggtt and ggcctgtattcttggctagttcatt; *CLPR4* ggtcctaagagagcaacagaggaaa and gggaatgtctctcgtaccttcagtt; *CLPS1* tggtcttggagctcataaggagaga and cctacccagcatccatatgaagaca; left border of T-DNA, ggcaatcagctgttgcccgtctcactggtg, caccagtgagacgggcaacagctgattgcc, and

cccaggctttacactttatgcttccggctc (Integrated DNA Technologies). Sequence verification and DNA analyses were performed using VectorNTI Suite 9.0 (Invitrogen).

Total RNA was isolated with an RNeasy Plant Mini Kit (Qiagen). First strand was synthesized from equal amounts of total RNA with Superscript III RT-RNaseH (Invitrogen) and primers used were: *CLPR4* ccatggaggtagcagcagcga and gctagcaatgagttgtgccttttaa; *ACTIN2* gcaactgggatgatatggaaaaga *and caaacgagggctggaacaagact.; CLPP3* same as DNA primers; *CLPP4 ccacctctcagaacaacttcaccat and gctagcctcatcatcaggtatctcgg; CLPP5 agctgtatattccggcaatctctgga and gctagctgctgcaagtggctgga; CLPR1 tttgttcaggaaagcatctcgtcta and ctcgagtcatcttagtccggctggagctgc; CLPR2 agatctggcggtctcgtttaataca and gctagcccctgcgctttcgtcttggc* (nested primers *catatgtctgggactgggaaagcgagcagg and gaattcctagcctagccctgcgctttcgtc); CLPS1 cggagaatggttcttcggataagat and gaattcctattgaccttgtttcttgaagctc; CLPS2 (At4g12060) catatggctctacaaactaatcgcccaagaag and gctagcttcatctaaaaagccagattc.*

Complementation of clpr1-1

The cDNA for *CLPR1* was subcloned into pENTR (Invitrogen). The DNA was introduced into a pXH102 Gateway destination/plant-binary vector (a kind gift of Dr. D. Kumar) using LR clonase (Invitrogen) and sequenced. *clpr1-1* homozygous plants were transformed by floral dipping with *Agrobacterium tumefaciens* (Clough and Bent, 1998).

Microscopy

For transmission electron microscopy, the fixation protocol was modified from (Gluaert, 1975). Briefly, the leaf materials were excised and placed in 2.5% glutaraldehyde in 0.1M sodium cacodylate pH 6.8 for 0.5 hour at room temperature, 10 minutes of which was under vacuum. The samples were then transferred to 4°C for 1.5 hours. Following fixation, the samples were washed with 0.1M sodium cacodylate buffer pH 6.8 for three

10 minute incubations at 4°C. After washing, the samples were placed at 4°C overnight in 1% osmium tetroxide. After three buffer washes, leaves were *en bloc* stained in 2% uranyl acetate for one hour at room temperature in dark. Roots were fixed in 1.25 % glutaraldehyde in 0.05 M sodium cacodylate pH 6.8 for one half hour at room temperature and then transferred to 2.5 % glutaraldehyde in 0.05 M sodium cacodylate pH 6.8 for one hour at 4°C. The samples were dehydrated through a graded series of ethanol. After dehydration the samples were infiltrated through a graduated series of ethanol and Spurr and then embedded in 100% Spurr with accelerator curing at 70°C overnight. Images were taken on a FEI TECNAI 12 BioTwin (Philips).

For confocal microscopy, seeds were imbibed in water for 24 hrs at 4°C. Embryos were separated from their seed coat and mounted immediately in 50% glycerol. Imaging was performed on a laser scanning confocal Leica TCS SP2 microscope equipped with Nomarski optics.

Promoter:GUS fusions and assay

The promoter region for *CLPR2* was amplified using primers (BamHI) ggatcctagatcacatcacatcaca and ccatggtcgtcgagcgtgtt (NcoI), sub-cloned into pGEM-Teasy (Promega), and the encoded restriction sites used to swap the CaMV35S promoter region in pCAMBIA 1305.2 (www.cambia.org). The promoter region (1040 bp) for *CLPR4* was amplified using primers cacccgtgtgaaaccaggcaagt and ccatttcttctttgttgcagag, sub-cloned into pCR8 (Invitrogen), and then LR-cloned (Invitrogen) into pMDC162 (Curtis and Grossniklaus, 2003).

The resulting vectors were electroporated into *Agrobacterium tumefaciens* strain GV3101 (a kind gift of Dr. J. Hua) and used for floral dipping according to (Clough, 2005). Transgenic plants were selected based on hypocotyl length after dark germination under selective pressure for hygromycin resistance (20 μg/L), adapted from (Lahser, 2003).

Histochemical GUS assay was performed according to Kirsten Bomblies' protocol in ((Weigel and Glazebrook, 2002), pg 243-245) without embedding.

Protein analysis

Total leaf proteins were extracted by grinding in 100mM Tris-HCl pH 8.8, 1% SDS, 5 mM EDTA. Equal amounts were loaded on duplicate 12% Laemli PAGE gels. One gel was stained with Sypro Ruby. The other gels were blotted onto PVDF membranes and probed with specific antibodies using chemiluminescence for detection, following standard procedures.

Native total cellular soluble proteomes of wild type and *clpr4-1* were collected by grinding leaves in a mortar pre-cooled at 4°C in 10mM Hepes pH 8.0, 5mM $MgCl_2$ and a protease inhibitor cocktail. Membranes and cellular debris were separated by centrifugation. About 2 mg protein was separated by Colorless Native Gels (CN-PAGE) as described originally in (Schägger et al., 1994) and (Peltier et al., 2001; Peltier et al., 2004b), followed by SDS-PAGE. The resulting 2-dimensional gels were stained with Sypro Ruby (Invitrogen) and gels were scanned using a CCD camera (FluorS, Biorad).

For shotgun proteomics analysis, leaves were ground in 0.1M Tris-HCl pH 8.8, 4% SDS. SDS removal and in solution protein digestion with trypsin in 10% DMSO was done according to (Peltier et al., 2004a).

Peptide sequencing was performed by on-line nanoRP-LC-MS/MS on a Q-TOF 1.5 (Micromass, Waters). Protein identification by Mascot 2.1, searching a local copy of the TAIR v6.0 *Arabidopsis* database (www.arabidospsis.org), was validated by software developed in-house and manual inspection of the spectra. Minimum requirements for identification relied on p-value <0.05, a clear Y-ion series and peptide MOWSE score cut-off value of 20. Protein information was obtained form TAIR, PPDB (ppdb.tc.cornell.edu/), MapMan (gabi.rzpd.de/projects/MapMan/), TargetP (www.cbs.dtu.dk/services/TargetP/) and Predotar (urgi.infobiogen.fr/predotar/).

Proteins were quantified using the Bicinchoninic acid assay (Smith et al., 1985).

Chlorophyll fluorescence measurements

As a measure of the efficiency of Photosystem II, F_v/F_m from 10 min dark adapted leaves was measured on wild-type and *clpr4-1* using a chlorophyll fluorescence instrument (FMS2; HansaTech, UK).

ACKNOWLEDGEMENTS

This work was supported by the grants from the National Science Foundation (#MCB 0343444) and the US department of Energy (DE-FG02-04ER15560) to KJVW. We thank the members of the van Wijk laboratory, and especially Verenice Rodriguez-Ramirez for embryo excisions and Verenice Rodriguez-Ramirez and Jitae Kim for complementation of *clpr1-1*. Drs Jian Hua and Tom Owens are acknowledged for help and suggestions, Dr. Dinesh Kumar for providing the vector pXH102, and Mandayam V Parthasarathy and Anita Aluisio of the microscopy facility at Cornell University (http://www.cimc.cornell.edu/) for the TEM analysis.