Published in The Plant Cell, Vol. 23, Issue 11, 2011, p. 3911-3928
which should be used for any reference to this work

Plastid Proteome Assembly without Toc159: Photosynthetic Protein Import and Accumulation of *N*-Acetylated Plastid Precursor Proteins [C][W]

Sylvain Bischof,[a] Katja Baerenfaller,[a] Thomas Wildhaber,[a] Raphael Troesch,[a] Pierre-Alexandre Vidi,[b,1] Bernd Roschitzki,[c] Matthias Hirsch-Hoffmann,[a] Lars Hennig,[a,2] Felix Kessler,[b] Wilhelm Gruissem,[a,c] and Sacha Baginsky[a,3,4]

[a] Department of Biology, Eidgenössische Technische Hochschule Zurich, 8092 Zurich, Switzerland
[b] Laboratoire de Physiologie Végétale, 2007 Neuchâtel, Switzerland
[c] Functional Genomics Center Zurich, 8057 Zurich, Switzerland

Import of nuclear-encoded precursor proteins from the cytosol is an essential step in chloroplast biogenesis that is mediated by protein translocon complexes at the inner and outer envelope membrane (TOC). Toc159 is thought to be the main receptor for photosynthetic proteins, but lacking a large-scale systems approach, this hypothesis has only been tested for a handful of photosynthetic and nonphotosynthetic proteins. To assess Toc159 precursor specificity, we quantitatively analyzed the accumulation of plastid proteins in two mutant lines deficient in this receptor. Parallel genome-wide transcript profiling allowed us to discern the consequences of impaired protein import from systemic transcriptional responses that contribute to the loss of photosynthetic capacity. On this basis, we defined putative Toc159-independent and Toc159-dependent precursor proteins. Many photosynthetic proteins accumulate in Toc159-deficient plastids, and, surprisingly, several distinct metabolic pathways are negatively affected by Toc159 depletion. Lack of Toc159 furthermore affects several proteins that accumulate as unprocessed *N*-acetylated precursor proteins outside of plastids. Together, our data show an unexpected client protein promiscuity of Toc159 that requires a far more differentiated view of Toc159 receptor function and regulation of plastid protein import, in which cytosolic Met removal followed by N-terminal acetylation of precursors emerges as an additional regulatory step.

INTRODUCTION

Assembly of the chloroplast proteome requires transport complexes that recognize plastid precursor proteins in the cytosol and mediate their translocation into the chloroplast. Most of the nuclear-encoded chloroplast proteins contain a cleavable N-terminal transit peptide and are imported by the TOC/TIC (for translocon at the outer/inner membrane of the chloroplast) import machinery (Agne and Kessler 2009; Balsera et al., 2009). In *Arabidopsis thaliana*, recognition and selection of chloroplast-imported proteins are mediated by GTP binding proteins that belong to two small families: Toc34/33 and Toc159/132/120/90. Members of the Toc159 family share a GTP binding domain (G-domain) and a membrane anchoring domain (M-domain). They are distinguished by differently sized acidic domains (A-domain), with Toc90 entirely lacking the A-domain. The differently sized A-domains in Toc159, Toc132, and Toc120 are located at the N-terminal side of the G- and M-domains and largely account for the molecular mass differences between the different members of the Toc159 family (Agne and Kessler, 2009). The receptors are required for photosynthetic growth because their loss-of-function produces plants with a pale green or albino phenotype (Jarvis et al., 1998; Bauer et al., 2000). The analysis of Toc33 mutants (*ppi1* for plastid protein import 1) suggested that this receptor may be preferentially involved in the import of photosynthetic proteins (Kubis et al., 2003). Similarly, the loss of Toc159 function (*ppi2*) results in albino plants that do not grow beyond the cotyledon stage on soil because the accumulation of photosynthetic proteins is significantly reduced (Bauer et al., 2000).

Reverse genetic studies and precursor binding assays suggested two different classes of receptors with specialized functions:

[1] Current address: Basic Medical Sciences, Purdue University, West Lafayette, IN 47907.

[2] Current address: Department of Plant Biology and Forest Genetics, Uppsala BioCenter, Swedish University of Agricultural Sciences, Box 7080, 75007 Uppsala, Sweden.

[3] Current address: Martin-Luther-University Halle-Wittenberg, Institute of Biochemistry and Biotechnology, Plant Biochemistry, Weinbergweg 22 (Biozentrum), 06120 Halle (Saale), Germany.

[4] Address correspondence to sacha.baginsky@biochemtech.uni-halle.de. The authors responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) are: Wilhelm Gruissem (wgruissem@ethz.ch) and Sacha Baginsky (sacha.baginsky@biochemtech.uni-halle.de).

[C] Some figures in this article are displayed in color online but in black and white in the print edition.

[W] Online version contains Web-only data.

one class comprising Toc159/90 and the other class Toc132/120 (Hiltbrunner et al., 2004; Ivanova et al., 2004; Kubis et al., 2004; Smith et al., 2004; Infanger et al., 2011). The loss of either Toc132, Toc120, or Toc90 did not result in a visible phenotype, but *Toc132 Toc120* double mutants appeared pale green (Kubis et al., 2004) or were even embryo lethal (Ivanova et al., 2004), suggesting functional overlap between Toc132 and Toc120. Neither full-length Toc132 nor Toc120 could complement *ppi2* (Ivanova et al., 2004; Kubis et al., 2004). Based on these data, it was proposed that Toc132 and Toc120 are specific for the import of nonphotosynthetic proteins (Kubis et al., 2004). In contrast with Toc132 and Toc120, the loss of Toc90 function in the *ppi2* background resulted in a more severe phenotype, suggesting functional overlap between Toc159 and Toc90 (Hiltbrunner et al., 2004). Indeed, partial complementation of the *ppi2* phenotype was achieved by the overexpression of Toc90 (Infanger et al., 2011).

Although it is currently unclear how the different Toc receptors recognize their substrates, it is conceivable that amino acids in the N-terminal transit peptide of plastid precursor proteins are involved in the recognition and in establishing different import specificities. Canonical chloroplast targeting requires transit peptides that are typically 20 to 100 amino acids long and enriched in hydroxylated amino acids but low in acidic amino acids (Bruce, 2001). After import, most transit peptides are cleaved by a stromal processing peptidase (SPP; Richter and Lamppa, 1998, 1999). Processing of imported precursors is essential, and the lack of SPP results in embryo lethality in *Arabidopsis* (Trösch and Jarvis, 2011). Because the SPP cleavage site sequence is not highly conserved, it was proposed that recognition requires physicochemical properties of the transit peptide rather than a particular amino acid sequence (Emanuelsson et al., 2000; Zhang and Glaser, 2002; Rudhe et al., 2004). Following transit peptide removal, the newly generated protein N terminus may be acetylated (Ferro et al., 2003; Kleffmann et al., 2007; Zybailov et al., 2008). *N*-acetylation is catalyzed by N-terminal acetyltransferases in the cytosol and in the plastid (Sherman et al., 1985; Meinnel et al., 2006; Goetze et al., 2009). It is therefore possible that a chloroplast-targeted protein may be acetylated as a precursor in the cytosol and after import and processing as the mature protein. N-terminal acetylation is thus a useful modification to map the different N termini of proteins in different compartments. Interestingly, the lack of a cytoplasmic *N*-acetyltransferase affects plastid biogenesis suggesting a functional crosstalk between cytosolic *N*-acetylation and chloroplast function (Pesaresi et al., 2003).

The assembly and maintenance of the chloroplast proteome are highly regulated processes and involve not only the organellar import machinery but also multiple anterograde and retrograde signaling events that affect the transcription of nuclear- and chloroplast-encoded genes (Mochizuki et al., 2001, 2008; Strand et al., 2003; Larkin et al., 2003; Koussevitzky et al., 2007; Moulin et al., 2008; Kakizaki et al., 2009). Similar transcript patterns observed under different environmental conditions or genetic modifications identified key regulators that affect entire sets of plastid proteins and biochemical networks (Richly et al., 2003; Biehl et al., 2005). In order to understand chloroplast proteome assembly, it is therefore necessary to investigate the complexity and quantitative nature of the proteome together with the regulation of the corresponding genes.

Here, we report an integrated approach to investigate the qualitative and quantitative contribution of the Toc159 receptor to the assembly of the plastid proteome. We analyzed the proteome and the transcriptome of Toc159-deficient plants to discern the consequences of impaired protein import from systemic transcriptional responses resulting from the loss of photosynthetic capacity. The analysis of N-terminal protein acetylation showed that some chloroplast-targeted precursor proteins accumulate outside of plastids in Toc159-deficient plants. Interestingly, these precursors were not exclusively photosynthetic proteins but have diverse metabolic functions. Several of these proteins were acetylated at position 2, indicating that they entered the common modification pathway for cytosolic proteins. Our results show that Toc159 has broad specificity, which is different from the proposed role as a receptor for import of abundant photosynthetic proteins. Together, our combined proteome and transcriptome analysis of mutants lacking Toc159 provides new insights into nuclear–organelle interactions during chloroplast proteome assembly.

RESULTS

Quantitative Analysis of Protein Accumulation in Wild-Type and Toc159-Deficient Leaves

To investigate the contribution of Toc159 to the assembly of the plastid proteome, we measured the quantitative protein accumulation in two *Arabidopsis* mutant lines deficient in Toc159 in comparison to the wild type and the wild type grown on Suc (wtS) (Figures 1A to 1D). The Toc159 null mutant *ppi2* (Bauer et al., 2000) can grow beyond the cotyledon stage only on medium supplemented with Suc (Figure 1A). Elevated levels of soluble sugars interfere with developmental processes and affect gene expression (Bläsing et al., 2005; Gibson, 2005). To bypass the requirement of exogenous sugar supply, we generated a Toc159 cosuppression line (*Toc159cs*) that develops silencing of the *Toc159* gene only during later stages of development, resulting in normal appearing seedlings and older plants with chlorophyll-deficient white rosette leaves (Figure 1B). *Toc159cs* lines were originally constructed to overexpress the M- and G-domains of Toc159 in a wild-type background (see Supplemental Figure 1 online) but developed the cosuppression phenotype shown in Figure 1A. Several independent *Toc159cs* lines were selected containing only one single transgene insertion. Thermal asymmetric interlaced PCR was used to identify lines in which the transgene was inserted in a noncoding region (see Supplemental Figure 1D online). Immunoblotting analyses with Toc159 antibodies confirmed the complete loss of Toc159 in *ppi2* and in white leaves of *Toc159cs* (Figure 1E).

Using tandem mass spectrometry (MS/MS), we analyzed the leaf proteome composition of wild-type, wtS, *ppi2*, and *Toc159cs* plants in three biological replicates each. In total, 4582 different proteins were identified (wild type, 2760; wtS, 2882; *ppi2*, 3150; *Toc159cs*, 3119) (Figure 1F; see Supplemental Data Set 1 online). Quantitative information was obtained for all identified proteins using normalized spectral counting (nSpC) (Baerenfaller et al., 2008). Since our experiment was designed to

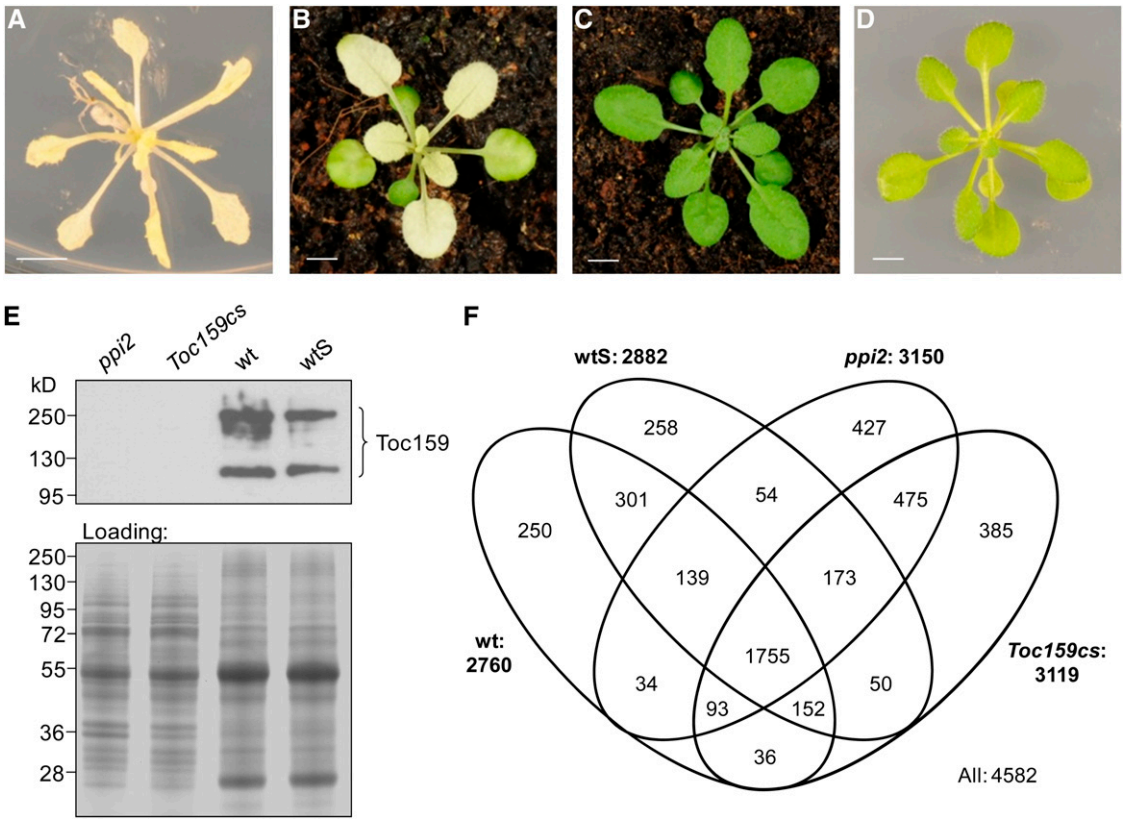


Figure 1. Quantitative Proteome Profiling of *ppi2*, *Toc159cs*, Wild-Type, and wtS Leaves.

(A) to **(D)** Phenotypes of 35-d-old plants used for large-scale proteome profiling grown in short-day conditions (8 h light/16 h dark) on soil or on half-strength Murashige and Skoog medium supplemented with 100 mM Suc for *ppi2* and wtS. *ppi2* **(A)**, *Toc159cs* **(B)**, wild type **(C)**, and wtS **(D)**. Bars = 5 mm.

(E) Protein immunoblotting using antibodies directed against Toc159 confirming the lack of Toc159 in *ppi2* and *Toc159cs* plants. Coomassie blue staining shows an equivalent protein loading for each protein extract separated by SDS-PAGE. wt, wild type.

(F) Proteins identified by MS/MS in *ppi2*, *Toc159cs*, wild-type, and wtS leaves. Total proteins were extracted from leaves, separated by gel electrophoresis, digested with trypsin, and analyzed by liquid chromatography MS/MS using a linear trap quadrupole Fourier transform–ion cyclotron resonance mass spectrometer.

analyze the consequences of Toc159 depletion for the assembly of the quantitative plastid proteome, we grouped the proteomics data of the wild type (wild type and wtS) and the Toc159-deficient plants (*ppi2* and *Toc159cs*) and compared the protein identification and accumulation within and between these two groups. In total, 2347 proteins were detected in the wild type and wtS and 2496 in *ppi2* and *Toc159cs*. The Spearman rank correlation coefficients obtained from the mean nSpC within these two groups are 0.85 for the wild type and wtS and 0.69 for *ppi2* and *Toc159cs*. Although a large number of proteins were identified in all four plant types (1755 proteins), their quantitative accumulation was considerably different between the wild-type and the Toc159-deficient plants with Spearman rank correlation coefficients between 0.43 and 0.47.

Plastid Proteome Assembly in the Absence of Toc159

To investigate the changes in the abundance of plastid proteins, we extracted these from the leaf data sets using a high-confidence chloroplast proteome reference table comprising 1155 proteins (Baginsky and Gruissem, 2009; see Supplemental Data Set 2 online). We identified in total 926 plastid proteins in our data sets. In Toc159-deficient plants (the *ppi2* mutant and *Toc159cs*), we identified 767 plastid proteins and 569 were identified in all four profiled plant types (Figure 2A). The quantitative accumulation of plastid proteins was very similar between wild-type and wtS plastids and between *ppi2* and *Toc159cs*, with Spearman rank correlation coefficients of 0.945 and 0.813, respectively, and statistical tests support this conclusion. Out of 737 plastid proteins identified with more than 10 spectra, only 27 were significantly different between wild type and wtS and only 22 between *Toc159cs* and the *ppi2* (*t* test, P value < 0.05). Only two proteins are present in both sets of significantly changed proteins: ferritin (AT5G01600) and β-amylase (AT4G17090); thus, we have no indication for a systematic influence of Suc on quantitative plastid protein accumulation. These results indicate that effects of exogenous Suc supply on plastid proteome assembly are minor under our experimental conditions and that both growth

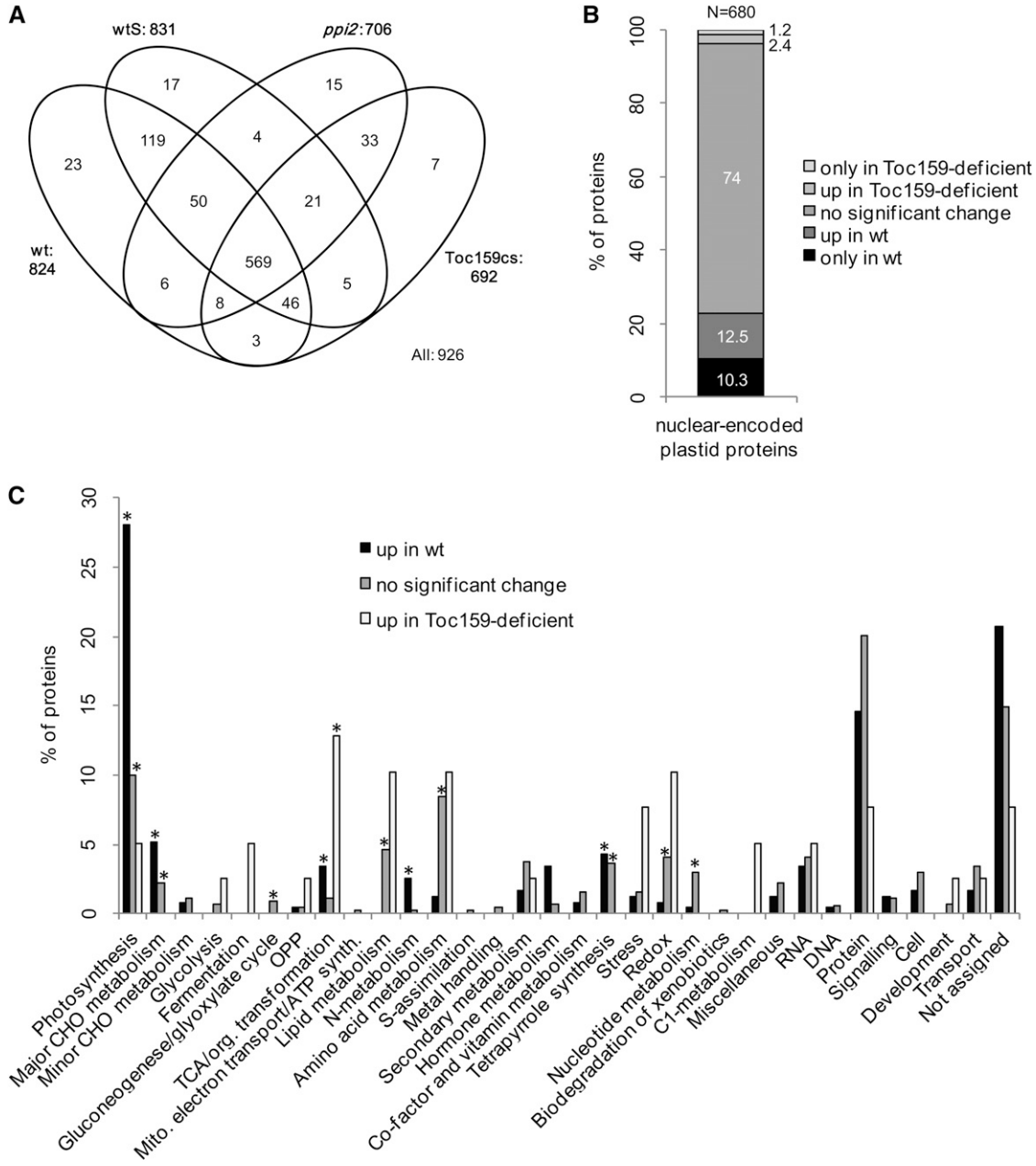


Figure 2. Quantitative Accumulation of Plastid Proteins Identified in Leaves.

(A) Venn diagram of the plastid proteins identified in leaves by MS/MS. Plastid localization is based on a chloroplast proteome reference table with 1155 proteins. wt, wild type.

(B) Differential accumulation of nuclear-encoded plastid proteins in wild-type and Toc159-deficient leaves. Numbers have been rounded to one decimal.

(C) Metabolic annotation of the nuclear-encoded genes destined to the plastid based on their distribution in MapMan bins. Categories indicated with an asterisk are significantly overrepresented, showing P values < 1e-05 in Fisher's exact test. CHO, carbohydrate; OPP, oxidative pentose phosphate.

conditions, with and without Suc, can be used for a quantitative comparison of the plastid proteomes in wild-type and Toc159-depleted plastids. Furthermore, statistical testing (*t* test, P value < 0.05) revealed that the abundances of the different components of the protein import machinery were not significantly affected by Suc, suggesting that similar constraints exist for the assembly of the plastid proteome in *ppi2* and *Toc159cs* (see Supplemental Figure 2 online).

Next, we used the quantitative data for 737 plastid proteins of which 680 are nuclear encoded that were identified with at least 10 spectra across all samples to identify groups of plastid proteins whose accumulation is significantly affected by the

absence of Toc159. For statistical analysis, we omitted proteins identified with <10 spectra because low-abundance proteins are not accurately quantified (Zybailov et al., 2009). We performed a *t* test and defined proteins that accumulated to higher levels in wild-type or in Toc159-deficient plants as those proteins that had a P value < 0.05 both in the comparisons of the wild type against *Toc159cs* and wtS against *ppi2*, and no significant difference in the comparisons of the wild type against wtS and *Toc159cs* against *ppi2* to exclude those proteins from the analysis whose accumulation is affected by Suc. In addition, we required a minimum fold change above 1.5 to exclude spurious small changes that are statistically significant from the interpretation. To these proteins we added those that had been identified in at least two replicates in each wild-type condition and not at all in both Toc159-deficient conditions and vice versa. Altogether, 184 plastid proteins accumulated to higher levels in wild-type and 24 in Toc159-deficient plants, of which 155 and 24, respectively, are nuclear encoded (Figure 2B; see Supplemental Data Set 3 online). The vast majority of the identified plastid proteins are not significantly affected by Toc159 depletion (74%). Two groups of plastid proteins ("no significant change" and "up in Toc159-deficient") include proteins whose accumulation is not significantly affected by the absence of Toc159. Functional annotation of these putative Toc159-independent proteins and Fisher's exact test revealed a significant enrichment of the categories *photosynthesis* ("no significant change") and tricarboxylic acid (TCA) cycle *TCA/org. transformation* ("up in Toc159-deficient") (Figure 2C). By contrast, the third group of proteins, "up in wild type," comprises 155 nuclear-encoded plastid proteins that accumulate to higher levels in wild-type plastids and thus comprises putative Toc159-dependent substrates (Figures 2B and 2C). Functional annotation showed that these 155 proteins belong to various functional categories (Figure 2C) with a significant enrichment in *photosynthesis*.

Proteins for Photosynthetic Functions Are Downregulated at the Transcriptional Level in *ppi2*

To distinguish proteins whose accumulation is directly affected by a loss of Toc159 function from those that are downregulated at the transcriptional level, we profiled the transcriptome of wild-type and *ppi2* plants using plants grown on Suc to account for a possible effect of Suc on gene expression. After RNA hybridization to AGRONOMICS1 genome tiling arrays (Rehrauer et al., 2010), signals above background in at least one of the samples were detected for 24,639 of the ~30,000 probe sets corresponding to 24,378 genes. Of these genes, 1998 (8.2%) were significantly upregulated in the wild type and 2268 (9.3%) in *ppi2* (Figure 3A; see Supplemental Data Set 4 online). Our transcriptomics data confirmed the trend that photosynthetic proteins are expressed at lower levels in the import mutant as previously reported in published microarray, SAGE and RT-PCR data (Kubis et al., 2003; Kakizaki et al., 2009; Lee et al., 2009b). Notably, 84% of the downregulated genes reported in the SAGE data set were also found downregulated under our experimental conditions (see Supplemental Data Set 4 online). Similarly, we could confirm the reported upregulation of several heat shock–related proteins (Kakizaki et al., 2009). Of the 1046 nuclear-encoded genes coding for plastid proteins for which we had array data, 381 (36.4%) had significantly reduced RNA levels in *ppi2*, with a significant overrepresentation of genes encoding proteins for photosynthetic functions (Figure 3B; see Supplemental Data Set 4 online). Thus, the decreased accumulation of photosynthetic proteins in *ppi2* is most likely a composite effect of reduced protein import and downregulation of genes for photosynthetic proteins, possibly as a systemic response to the loss of photosynthetic capacity.

We observed that 104 (expected in a random distribution are 59, P value in Fisher's exact test = 2.2e-07) of the set of 155 nuclear-encoded plastid proteins with decreased protein accumulation in Toc159-deficient plants for which we had transcript levels were also downregulated at the transcriptional level in *ppi2* (69.8%; Figures 3B and 3C; see Supplemental Data Set 3 online). For the 24 proteins that accumulated to higher levels in Toc159-deficient plants (Figure 3C), 75% showed no significant change in RNA expression level, 20.8% were upregulated in *ppi2* compared with the wild type, and only one was downregulated. From the 529 proteins (Figure 2B; i.e., the 74% plastid proteins from the category "no significant change"), more than 80% were either unchanged or downregulated at the transcriptional level in *ppi2*, thus indicating that they are efficiently imported into plastids despite the absence of Toc159. These Toc159-independent proteins also include proteins involved in photosynthesis (Figure 2C). In conclusion, the comparison between protein and transcript accumulation therefore confirms that Toc159 is not exclusively involved in the import of photosynthetic proteins and that many photosynthetic proteins enter the plastids in the absence of Toc159 (Figure 2C).

Peptide Detection Incidence Indicates Plastid Precursor Protein Accumulation outside of Plastids in Toc159-Deficient Plants

The accumulation of plastid proteins in Toc159-deficient leaves does not necessarily indicate their correct import into the chloroplast. We therefore analyzed the peptide detection incidence along the amino acid sequence of each protein to distinguish between mature plastid proteins and unprocessed precursors. For each identified protein, the most N-terminal detected peptide was considered and placed into the indicated amino acid bin corresponding to its location in the protein (Figure 4). After import, transit peptides of plastid proteins are usually cleaved off and degraded. Therefore, peptides from the transit peptide region should not be detectable from mature plastid proteins. The bin distribution of peptide detection reveals a detection gap at the N-terminal region of plastid proteins, thus suggesting that the transit peptides of most of the plastid proteins were cleaved (Figure 4A). The distribution was similar between the wild type, wtS, *ppi2*, and *Toc159cs*, suggesting that almost all precursor proteins were imported into plastids and processed in the absence of Toc159. As a control, we analyzed the peptide detection incidence for nonplastid proteins (Figure 4B). In this case, peptide detections from the N-terminal region were more prevalent, indicating that the low peptide detection in the N terminus of plastid proteins is most likely a consequence of N-terminal processing.

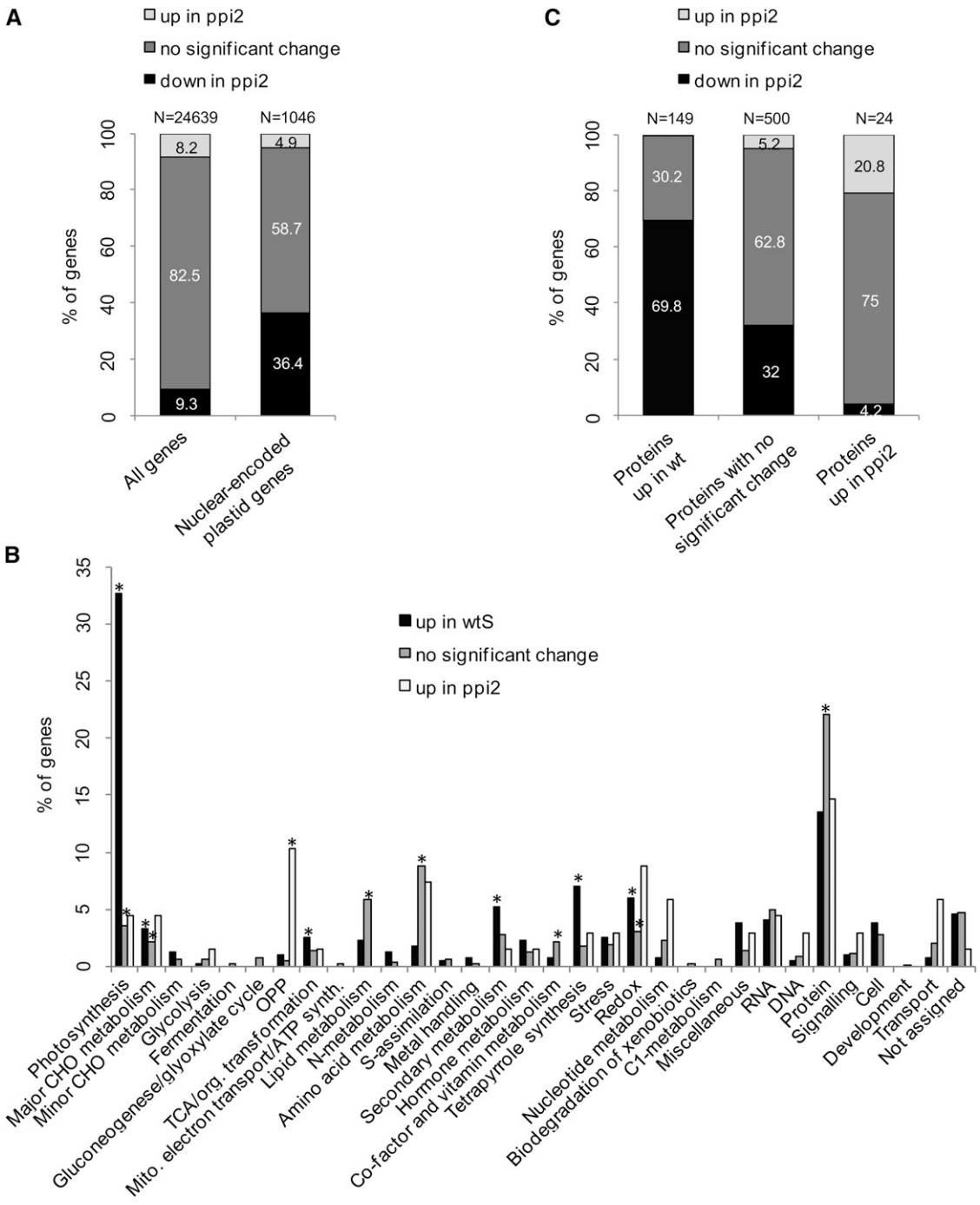


Figure 3. Transcriptional Response in *ppi2* Compared with wtS.

(A) Expression data for 24,639 genes of *ppi2* compared with wtS grown on Suc (all genes) and for the subset of 1046 nuclear-encoded plastid proteins. The percentage of downregulated genes in *ppi2* highly increases when only nuclear-encoded genes destined to the plastid are taken into account.
(B) Metabolic annotation of the nuclear-encoded genes destined to the plastid based on their distribution in MapMan bins. Categories indicated with an asterisk are significantly overrepresented, showing P values < 1e-05 in Fisher's exact test.
(C) Expression data for the genes for which the proteins were found significantly upregulated in wild type (wt)/wtS, unchanged, or significantly upregulated in toc159cs/wtS in the proteomics data. CHO, carbohydrate; OPP, oxidative pentose phosphate.

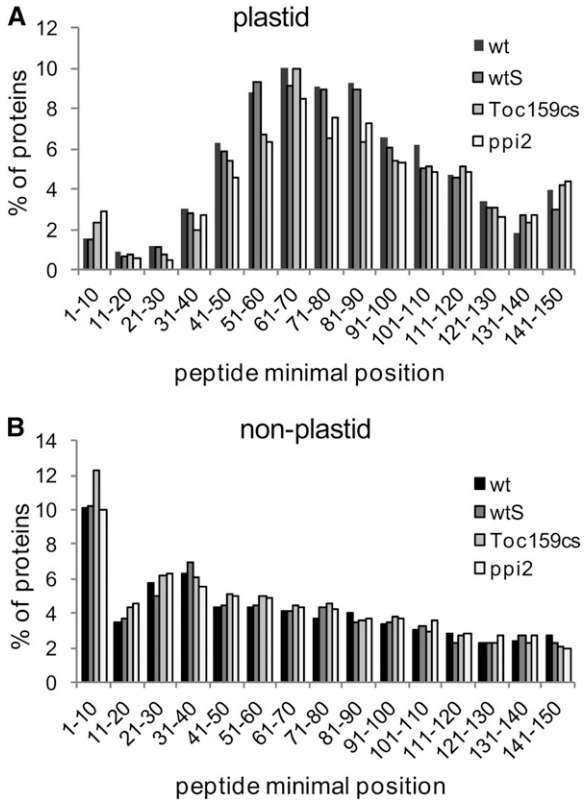


Figure 4. Peptide Detection Incidence and Transit Peptide Cleavage in Leaves.

(A) Peptide minimal position of plastid proteins identified in leaves. The positions of the most N-terminal detected peptide of each identified protein were grouped into bins. The reduced detection of peptides at the N terminus of plastid proteins supports the removal of transit peptides after import in the wild type and in Toc159-deficient plants. The number of proteins for each blot (wild type, wtS, *Toc159cs*, and *ppi2*) is 763, 771, 645, and 659.

(B) Peptide minimal position of nonplastid proteins identified in leaves indicates that the N terminus of most nonplastid proteins is not cleaved and readily detectable by MS. The number of proteins for each blot (wild type, wtS, *Toc159cs*, and *ppi2*) is 1936, 2051, 2427, and 2444.

We observed small peaks at positions 1 to 10 for plastid proteins in *ppi2* and *Toc159cs* that were not present in the wild type (Figure 4A). This detection bin contains 22 proteins, 14 of which have plastid transit peptides that TargetP predicts to be processed and removed after protein import. We retrieved information about the localization of these proteins from the Plant Proteomics Database (PPDB) (Sun et al., 2009) and marked all proteins whose localization is ambiguous (see Supplemental Data Set 5 online). The remaining proteins include established chloroplast proteins, such as a photosystem II oxygen-evolving complex protein (AT5G66570), RNA binding protein 29 (AT2G37220), or phosphoglycerate kinase (AT3G12780), which have predicted transit sequences of 29, 47, and 75 amino acids, respectively. Peptides within the cleaved transit peptide sequences were not found in the wild type or in the *Arabidopsis* proteome map (Baerenfaller et al., 2008), indicating that normally the transit sequences are efficiently cleaved and degraded. Their detection in *ppi2* and *Toc159cs* therefore provides a first hint that precursor proteins accumulate in these two genotypes (Table 1). To gain further insights into processing of chloroplast-targeted proteins in the wild type and in the mutants, we systematically analyzed N-terminal protein acetylation.

Plastid Precursor Proteins Accumulate outside of Plastids in Toc159-Deficient Plants Where They Are Modified by N-Terminal Met Excision and N-Terminal Acetylation

The detection of peptides within the N-terminal transit peptide sequence of several plastid proteins suggests their accumulation as unprocessed precursors in Toc159-deficient plants. However, the exact size of the transit peptide is difficult to predict and therefore could be erroneous. For further information about precursor processing, we therefore mapped the in vivo N termini of proteins by analyzing N-terminal protein acetylation (*N*-acetylation) (see Supplemental Data Set 6 online). Since *N*-acetylation occurs in the cytosol and in plastids, it allows us to distinguish between plastid precursor and mature proteins (Figure 5A). Cytosolic proteins are predominantly acetylated at position 2, which corresponds to the detection of N-acetylated peptides in the 1-10 amino acid bin (Figure 5B). This distribution was similar among the wild type, wtS, *ppi2*, and *Toc159cs*, indicating that cytosolic N-terminal Met excision (NME) and subsequent *N*-acetylation is functional in all plant lines (Meinnel et al., 2006).

By contrast, differences were found in the distribution of *N*-acetylation sites in plastid proteins. In the wild type, peaks of *N*-acetylated peptides were detected between amino acid positions 30 and 80, with a minor peak in the 1-10 amino acid bin (Figure 5A). *N*-acetylation between positions 30 and 80 corresponds to the position of the mature N terminus after removal of transit sequences and is therefore characteristic for imported and processed proteins. The minor peak at positions 1 to 10 comprises four nuclear-encoded proteins, a dynamin-like protein (AT5G42080), a protein of unknown function (AT5G01750), a malate dehydrogenase (AT5G11670), the Fe-superoxide dismutase 1 (AT4G25100), and three plastid-encoded proteins (see Supplemental Data Set 7 online). The nuclear-encoded proteins do not have predicted transit sequences, which is supported by our data.

For *ppi2* and *Toc159cs* plastid proteins, two *N*-acetylation peaks were visible: a major peak at positions 1 to 10 and a second, minor peak at positions 30 to 80 (Figure 5A). As discussed above, the peak at positions 30 to 80 is indicative for precursor processing after import in *ppi2* and *Toc159cs* plastids. The major peak at positions 1 to 10 in Toc159-deficient plants comprises three plastid-encoded proteins and 16 nuclear-encoded proteins (see Supplemental Data Set 7 online). Ten of these nuclear-encoded plastid proteins have a canonical transit sequence of a predicted length between 29 and 75 amino acids (Table 1). The accumulated proteins are not restricted to abundant photosynthetic proteins but also include minor constituents of the chloroplast proteome, such as a 29-kD ribonucleoprotein, the chaperonin 20, and a thiazole-requiring enzyme. Notably, the sequence context around the identified *N*-acetylation sites is consistent with established

Table 1. Unprocessed Precursor Proteins Accumulate outside Plastids in Toc159-Deficient Plants

Accession	Description	*N*-Acetylation Position (Leaves)	Minimal Peptide Position in Leaves			TP Length	Metabolic Function
		toc159 (Wild Type)	*toc159*	Wild Type	AtProteome		
N-acetylated plastid precursor proteins in Toc159-deficient plants							
AT1G10960	ATFD1 (FERREDOXIN1)	2	2	NF	59	69	Photosystem
AT1G60950	FED A (FERREDOXIN2)	2	2	130	59	52	Photosystem
AT2G37220	29-kD ribonucleoprotein, chloroplast	2	2	63	68	47	RNA met.
AT2G39730	RUBISCO ACTIVASE	2	73	73	73	58	Photosystem
AT3G12780	PGK1	2	2	80	80	75	Photosystem
AT1G61520	LHCA3; chlorophyll binding	2	14	14	49	48	Photosystem
AT3G54050	Fructose-1,6-bisphosphatase, putative	2 (60)	2	60	60	57	Photosystem
AT5G20720	CPN20 (CHAPERONIN20)	2	2	59	59	50	Protein
AT5G54770	THI1 (THIAZOLE REQUIRING)	2	2	47	46	45	Not assigned
AT5G66570	PSBO-1	2	2	107	107	29	Photosystem
Putative plastid precursor proteins identified in Toc159-deficient plants							
AT1G62640	3-Ketoacyl-acyl carrier protein synthase III	NF	44	92	72	74	Lipid met.
AT1G80030	DNAJ heat shock protein, putative	NF	72	133	118	92	Stress
AT2G04030	CR88 (embryo defective 1956)	NF	8	69	62	60	Abiotic stress
AT2G34460	Flavin reductase-related	NF	35	36	48	51	Unknown
AT3G63140	CSP41A	NF	50	102	80	72	RNA met.
AT4G13430	Isopropyl malate isomerase	NF	48	56	56	76	TCA/org.
AT5G06290	2-Cys peroxiredoxin B	NF	20	104	104	90	Redox
AT5G08280	Hydroxymethylbilane synthase	NF	70	81	81	86	Tetrapyrrole
AT5G16390	Acetyl coenzyme A carboxylase 1	NF	64	67	78	82	Lipid met.
AT5G20250	DIN10 (DARK INDUCIBLE10)	NF	9	20	20	–	Minor CHO met.
AT5G26742	EMB1138	NF	27	61	61	60	RNA met.
AT5G49910	cpHSC70-2	NF	45	152	152	92	Abiotic stress
AT5G62790	1-Deoxy-D-xylulose 5-phosphate reductoisomerase	NF	14	109	54	86	Secondary met.

Cytosolic location of plastid proteins was assessed by detection of *N*-acetylation within the predicted transit peptide (TP). Provided is also the minimal position of detected peptides in wild-type leaves and in AtProteome (Baerenfaller et al., 2008). Data "Wild Type" encompass data from wild type and wtS, and "toc159" data from *ppi2* and *Toc159cs*, respectively. CHO, carbohydrate; NF, not found; met., metabolism; org., organic acid; -, no transit peptide predicted.

requirements for the cytosolic NME/acetylation pathway (i.e., small, uncharged residues, such as Ala following the initiator Met) (Figures 5C and 5D) (Sherman et al., 1985; Meinnel et al., 2006).

The detection of unprocessed plastid precursor proteins in the import-deficient plants could result either from a cleavage and processing defect in the mutant plastids or from precursor protein accumulation outside of plastids, most likely in the cytosol. The data presented in Figure 6 show that the amino acid context of the identified *N*-acetylation sites of processed plastid proteins was similar in the wild type and mutants. Although it is possible that a defect in the plastid protease network could potentially affect precursor processing in plastids (Zybailov et al., 2009), our data show that the accumulation of proteins involved in the posttranslational protein homeostasis network is not negatively affected by a defect in Toc159 (see Supplemental Data Set 8 online). Most of the *N*-acetylated mature plastid proteins that we identified here start with Ala or Val, and Ile and Val are overrepresented at the position three amino acids N-terminal to the *N*-acetylation site in the precursor protein. This is true for the mature proteins in all genotypes tested and confirms previous data (Zybailov et al., 2008). Together, these data argue against a cleavage and processing defect in *ppi2* and *Toc159cs* lines. We therefore tested whether the detection of unprocessed precursors results from their accumulation outside of plastids using isolated plastids.

Proteome Analysis with Isolated Plastids Confirms the Accumulation of Precursor Proteins outside Plastids in Toc159-Deficient Mutants

To confirm the accumulation of *N*-acetylated plastid precursors outside of plastids in Toc159-deficient plants, we isolated plastids from wild-type and *ppi2* plants because unprocessed cytosolic precursors should be removed during plastid isolation. We analyzed quantitative protein accumulation in isolated wild-type and *ppi2* plastids using MS/MS in three biological replicates and identified 3231 unique proteins (see Supplemental Figure 3 and Supplemental Data Set 9 online). We focused our analysis on 856 proteins that were present in the chloroplast proteome reference table (Baginsky and Gruissem, 2009). Of these proteins, 708 were identified in the wild type, 737 in *ppi2*, and 589 in both

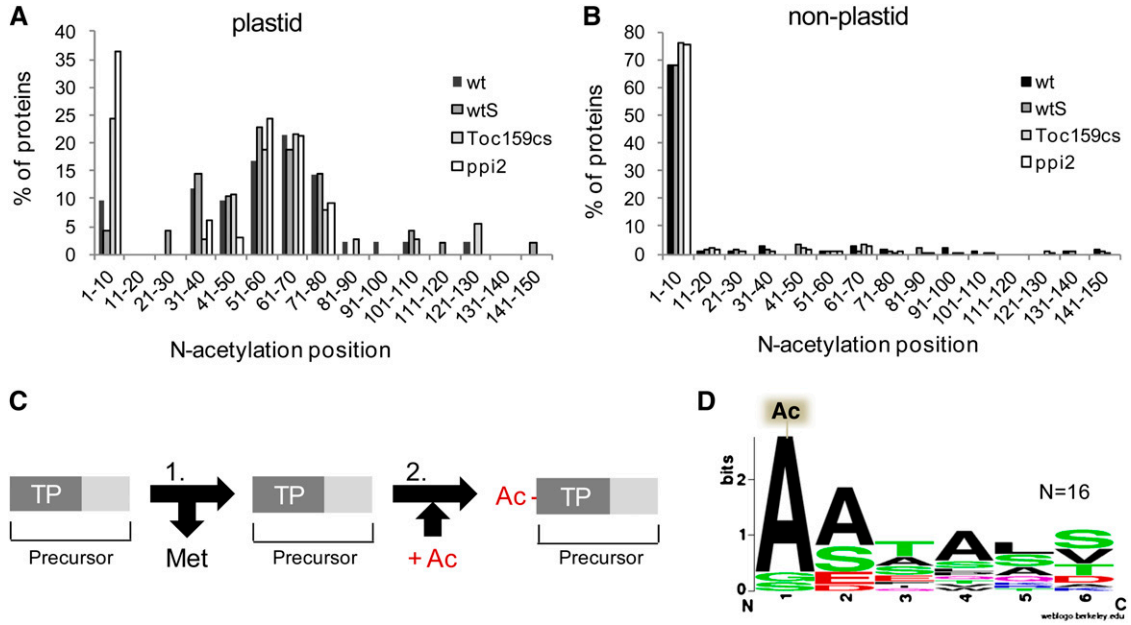


Figure 5. Met Removal and N-Terminal Acetylation in Leaves.

(A) *N*-acetylation sites of plastid proteins identified by MS/MS. *N*-acetylated proteins were grouped according to the position of their *N*-acetylation site. *N*-acetylation sites of plastid proteins identified at positions 30 to 80 reflect the correct cleavage and processing of transit peptides after import in wild-type and in Toc159-deficient plants. The increased percentage of *N*-acetylated plastid proteins at positions 1 to 10 in Toc159-deficient leaves suggests their accumulation outside plastids as a consequence of partially impaired import. The number of proteins for each blot (wild type, wtS, *Toc159cs*, and *ppi2*) is 42, 48, 37, and 33.

(B) Distribution of *N*-acetylation sites of nonplastid proteins indicates that most of these undergo Met removal and *N*-acetylation in the cytosol. The number of proteins for each blot (wild type, wtS, *Toc159cs*, and *ppi2*) is 121, 140, 191, and 168.

(C) Schematic representation of the cytosolic two-step Met removal/acetylation pathway. (1) N-terminal start Met is usually removed from newly synthesized precursor proteins containing a transit peptide (TP) if the second residue is small and uncharged. (2) Addition of *N*-acetylation at the second residue by *N*-acetyltransferases.

(D) Sequence alignment of *N*-acetylation sites (Ac) of all plastid proteins *N*-acetylated at position 2 in wild-type and in Toc159-deficient plants.

genotypes (see Supplemental Figures 3A and 3B online). The high number of proteins common to both plastid types provides additional evidence that most plastid proteins accumulate in both wild-type and *ppi2* plastids despite the lack of Toc159. The N-terminal peptide detection incidence of isolated wild-type and *ppi2* plastids was very similar, indicating that the transit sequence of most plastid proteins was cleaved after import (see Supplemental Figure 3C online). Also, the amino acid context around the identified *N*-acetylation sites of wild-type and *ppi2* plastid proteins was in agreement with the previous observation for all leaf data sets, confirming that processing of imported plastid proteins is functional in the wild type and in *ppi2* plastids (Figure 6; see Supplemental Data Set 9 and Supplemental Figure 3D online).

Analysis of acetylated N termini supports the import and correct processing of plastid precursor proteins in isolated wild-type and *ppi2* plastids (Figure 7A) because peptide detection peaked between positions 30 and 80 (cf. Figure 5A). Comparison of the *N*-acetylation site distribution in plastid proteins between leaves and isolated plastids from wild-type plants confirmed that precursor proteins do not accumulate in wild-type plants (Figure 7B; see Supplemental Data Sets 9 and 10 online). By contrast, the distribution of *N*-acetylation sites was different for leaves and isolated plastids of *ppi2* (Figure 7C). The major peak in bin 1-10 in leaves was absent in isolated plastids, suggesting that cytosolic precursor proteins were removed during plastid isolation (see Supplemental Data Set 10 online). Since peptide detection in data-dependent acquisition may result in false-negative results, we searched for the peptide masses of N-terminal peptides from precursor proteins in the leaf data set and in isolated plastids. The absence of detectable peptide masses in MS spectra provides a strong argument against its presence in the sample because the mass spectrometric measurement is extremely sensitive. Most peptides of a complex sample will give rise to a data point (i.e., peak) in the spectrum, even though only a very small fraction of these are identified in data-dependent acquisition experiments (Beck et al., 2011). While the precursor peptide masses were clearly identified in leaf extracts of Toc159-deficient plants, no masses matching the precursor peptide were identified in isolated plastids (exemplified for At3g12780 and At1g61520 in Figure 8 and Supplemental Figure 4 online). If nuclear-encoded plastid precursor proteins were present in plastids, the isolation procedure should have enriched these compared with the leaf extract and, thus, they should be more readily detectable. Based on these data, we therefore conclude that the unprocessed precursor proteins in *ppi2* and *Toc159cs* accumulate outside of

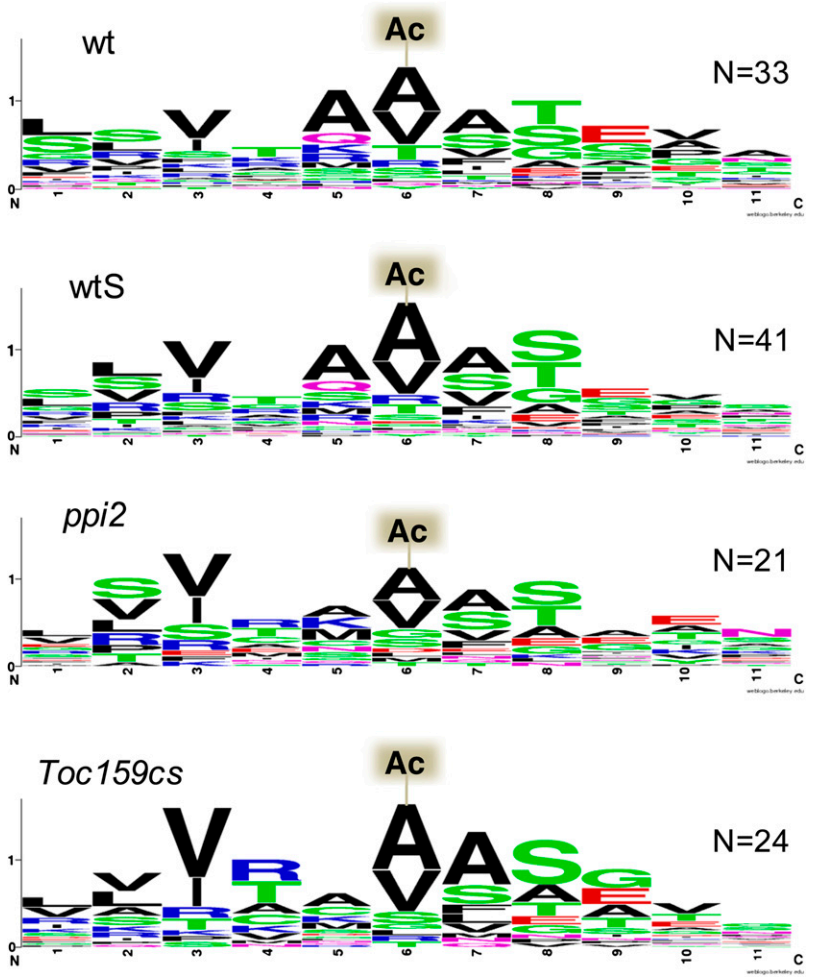


Figure 6. Sequence Context around *N*-Acetylation Sites of Plastid Proteins Identified in Leaves.

Most plastid proteins are *N*-acetylated (Ac) on Val or Ala. The similar sequence context in wild-type and in Toc159-deficient plants indicates that precursor processing of imported proteins is also functional in the mutant lines. *N*-acetylation sites ranging between positions 25 and 90 were used for the alignment.

plastids. The typical processing pattern of cytosolic proteins that is also observed for the detected precursor proteins (i.e., NME and N-terminal acetylation) (Figure 5) furthermore provides a strong argument for their cytosolic localization.

Identification of Toc159-Dependent and Toc159-Independent Plastid Proteins

Using the quantitative protein and transcript data presented above, we defined substrates that depend on Toc159 for import into the chloroplast and searched for features that distinguish such Toc159-dependent and Toc159-independent plastid proteins. We defined as Toc159-dependent those proteins that accumulated to significantly higher levels in the wild type and wtS compared with *ppi2* and *Toc159cs* leaves as described above (see Supplemental Data Set 3 online). To exclude those proteins whose decreased abundance in *ppi2* could be attributed to lower transcript levels, we additionally excluded those from the list of Toc159-dependent proteins that were significantly downregulated in *ppi2* compared with wtS. As a final criterion, plastid proteins annotated by PPDB and AT_CHLORO to be localized in the outer envelope were removed (Sun et al., 2009; Ferro et al., 2010). Together, 44 proteins fulfilled these requirements and constitute the set of Toc159-dependent proteins, whose accumulation is affected at the protein and not at the transcript level (see Supplemental Data Set 11 online). By contrast, Toc159-independent proteins are those proteins that have increased or equal protein abundance in *ppi2* or *Toc159cs* compared with the wild type and wtS, while their expression levels are similar in *ppi2* and wtS. Together, 308 Toc159-independent proteins matched these criteria (see Supplemental Data Sets 1 and 3 online). Most of these were also identified in isolated *ppi2* plastids, further supporting their Toc159-independent import (see Supplemental Data Set 9 online).

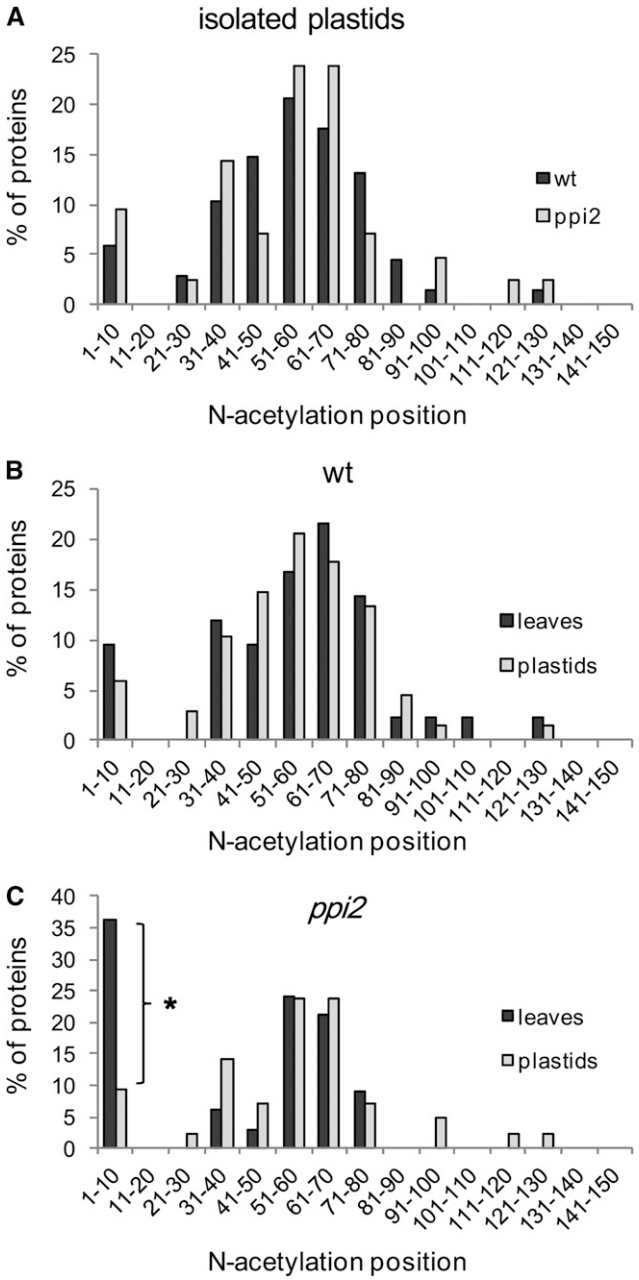


Figure 7. Comparison of N-acetylation Sites Identified in Isolated Plastids and Leaves.

(A) Similar distribution patterns of N-acetylation sites identified in isolated the wild-type and *ppi2* plastids indicate that plastid proteins were imported and processed correctly. The number of proteins for each blot (the wild type and *ppi2*) is 68 and 42.

(B) Similar distribution patterns of N-acetylation sites between isolated wild-type chloroplasts and wild-type leaves. The number of proteins for each blot (leaves and plastids) is 42 and 68.

(c) Comparison of the N-acetylation sites between isolated *ppi2* chloroplasts and *ppi2* leaves. A major difference (*) is observed at positions 1-10 suggesting the accumulation outside plastids of unprocessed precursor proteins when import is partially impaired in *ppi2*. The number of proteins for each blot (leaves and plastids) is 33 and 42.

Among the Toc159-independent proteins was plastoglobulin 35 (PGL35) (At4g04020) that we selected as an additional control for Toc159-independent protein import. The PGL35 protein can be considered representative for this group of proteins because it accumulates to similar levels in wild-type and albino leaves (average nSpC values are as follows: wild type, 1.35; wtS, 1.16; *ppi2*, 1.22; *Toc159cs*, 1.46) (see Supplemental Data Set 1 online). We fused PGL35 to green fluorescent protein (PLG35:GFP) (Vidi et al., 2006) and bombarded wild-type and white *Toc159cs* leaves with the fusion construct attached to gold particles (Figure 9). We observed the accumulation of PLG35:GFP in several wild-type plastids and in smaller plastids of white *Toc159cs* leaves (Figure 9A). By contrast, the precursor protein of the small subunit of ribulose-1,5-bis-phosphate carboxylase/oxygenase fused to GFP only accumulated in wild-type plastids (Figure 9B). These data confirm our proteomics strategy for the quantification and selection of Toc159-independent proteins.

To extract information about sequence features that might govern import specificity, we aligned the sequences of the 100 first amino acids of the putative Toc159-dependent and Toc159-independent substrates (see Supplemental Figure 5 online). The enriched amino acids comprise mostly Ser residues, and the pattern did not reveal any specific sequence or domain structure that would allow distinguishing proteins in the two data sets. Other attempts to define relevant information in the amino acid sequences by multiple sequence alignment using ClustalW were not successful, and none of the approaches allowed us to distinguish the proteins in the two data sets by sequence information.

DISCUSSION

The Chloroplast Proteome in the Absence of Toc159

We used quantitative proteomics and transcriptomics to investigate the contribution of the import receptor Toc159 to the assembly of the chloroplast proteome. In addition to the loss-of-function mutant *ppi2*, we used a Toc159 cosuppression line that allowed us to study the chloroplast proteome composition without exogenous sugar supply (Figure 1). We expected that the natural distribution of source/sink relationships in the *Toc159cs* plants better reflects the in vivo situation compared with the growth of *ppi2* on Suc. Interestingly, the qualitative and quantitative accumulation of plastid proteins in leaves was very similar in *Toc159cs* and *ppi2*, and also in the wild type and wtS, suggesting that exogenous Suc supply has only a minor effect on the assembly of the plastid proteome.

Based on the above data, *ppi2* and white leaves of *Toc159cs* are equivalent with respect to plastid proteome assembly in the absence of Toc159. Both genotypes accumulate photosynthetic proteins (Figures 2, 4, and 6), and proteins from several other metabolic pathways in the plastids of Toc159-deficient leaves are reduced (Figure 2). Thus, the substrate specificity of Toc159 is not restricted to photosynthetic proteins. Most of the proteins that had a lower abundance in *ppi2* and *Toc159cs* were also downregulated at the gene expression level, suggesting that their decreased accumulation is a result of two different

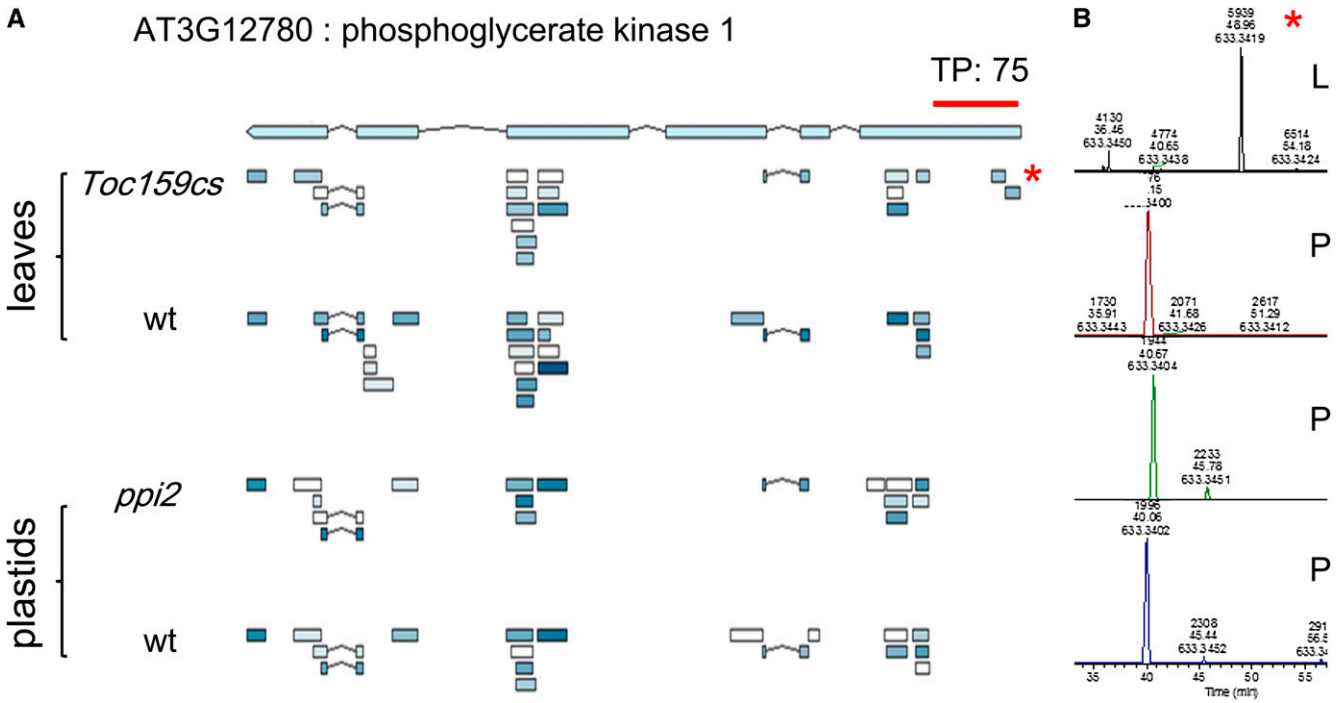


Figure 8. Transit Peptide Visualization and Extracted Ion Chromatogram Quantification for Phosphoglycerate Kinase1 Peptides Identified in Leaves and Isolated Plastids.

(A) Visualization of peptide detection along the protein. Several spectra matched to the transit peptide region (indicated as TP; 75 amino acids predicted length) in Toc159-deficient (toc159) leaves. By contrast, we could not detect any spectra matching the predicted TP region in wild-type leaves or in isolated *ppi2* or wild-type plastids, although the coverage of Phosphoglycerate Kinase1 (PGK1) detection is similar between the different samples (cf. peptide color scheme and peptide localization between the different samples; the darker the shading, the more spectra matching to the indicated peptide were observed).

(B) Extracted exact mass ion chromatogram for the spectra matching to the transit peptide region from leaves (top panel; marked with an asterisk) and corresponding samples from isolated plastids. In three samples from isolated plastids, an ion matching the exact mass of the PGK1 transit peptide identification (asterisk) was detected at a shifted retention time of ~8 min (bottom panels). MS/MS information on these ions revealed that they are clearly different from the peptide eluting at 48 min (top panel, precursor peptide). The upper peptide represents the peptide matching to the PGK1 transit peptide, the three ions below produced a poor-quality MS/MS spectrum that could not be assigned to any peptide (it was assigned to format dehydrogenase with a score below the threshold). L, leaves; P, plastids.

[See online article for color version of this figure.]

responses, the transcriptional response as well as the import defect (Figure 3). This is especially true for photosynthetic proteins. It could be argued that import deficiency directly affects gene expression, but such a scenario would require separate signaling circuits that control the expression of individual nuclear genes encoding plastid proteins, which seems unlikely. A more likely explanation is that nuclear-encoded plastid proteins are regulated in modules, which is supported by the proposed master switch model (Richly et al., 2003; Biehl et al., 2005).

Many plastid proteins accumulate in Toc159-deficient plants, some of them to similar or even above wild-type levels. Apparently, the absence of Toc159 is partially compensated by one or more of its homologs: Toc132, Toc120, or Toc90. We could not detect these receptors by MS, either in leaves or in isolated plastids, but transcripts of Toc132 and Toc120 were slightly elevated in *ppi2* compared with the wild type (see Supplemental Figure 1 and Supplemental Data Set 4 online). A recent study showed that the overexpression of Toc90 can partially restore the accumulation of Toc159 client proteins in *ppi2* (Infanger et al., 2011). However, the low abundance of the Toc90 protein and the low expression level of its transcript in *ppi2* make it unlikely that the residual protein import observed here is exclusively based on Toc90 activity (Bauer et al., 2000; Infanger et al., 2011). It is more plausible that Toc132 and Toc120 sustain protein import in Toc159-deficient plants because their expression is induced in *ppi2*. Elevated receptor abundance was previously shown to restore an albino phenotype to a pale-green phenotype, suggesting that receptor abundance is limiting for efficient protein import (Hiltbrunner et al., 2004; Ivanova et al., 2004; Kubis et al., 2004). Together with these observations, our data call for a model in which the import capacity is controlled not only by substrate specificity but also by the abundance of the different receptors and their ability to associate into functional translocon complexes.

A recent report suggested that the A-domain of the Toc159 receptor family mediates precursor selectivity (Inoue et al.,

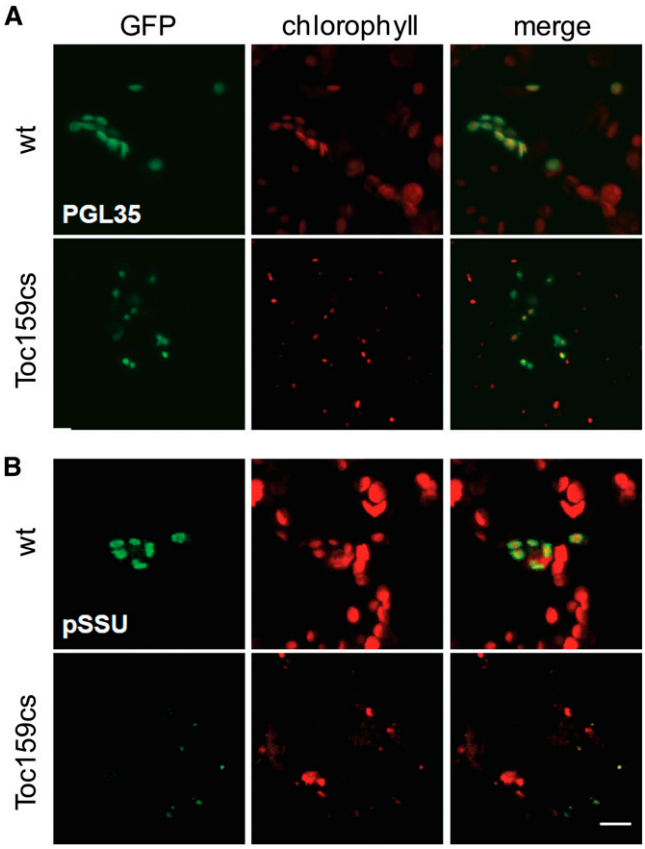


Figure 9. Toc159-Independent Import of 35S:PGL35 in White *Toc159cs* Leaves.

(A) Epidermal cells of wild-type and white *Toc159cs* leaves transformed by biolistic transformation. GFP fused to the transit peptide of PGL35 is imported successfully into wild-type and *Toc159cs* plastids. Maximal projections of confocal images are shown.
(B) Import into plastids in wild-type and white *Toc159cs* leaves of GFP fused to the transit peptide of ribulose-1,5-bis-phosphate carboxylase/oxygenase small subunit (pSSU).
Bars = 10 μm.

2010), which explains why overexpression of full-length Toc132 or Toc120 failed to complement *ppi2* (Ivanova et al., 2004; Kubis et al., 2004), while constructs containing only the G- and M-domains of Toc132 were able to do so (Agne et al., 2009; Richardson et al., 2009; Inoue et al., 2010). Loss of the A-domain resulted in import receptors with less selective precursor protein recognition and import capacities. It seems that the A-domain confers precursor selectivity to a constitutive import process, whose efficiency depends mainly on the abundance of the different Toc receptors. Initial results suggest that the A-domain may be cleaved upon assembly of the Toc complex (Agne et al., 2010). The conclusion that the A-domain mediates import specificity could explain why our data did not identify sequence specificity determinants of Toc159-mediated protein import (see Supplemental Figure 5 online). Although we identified information-rich regions within the first 20 amino acids up to position 55 as reported for Toc159-mediated RbcS import (Smith et al., 2004; Lee et al., 2008; Lee et al., 2009a), we could not identify sequence determinants that distinguish Toc159-dependent from Toc159-independent proteins (see Supplemental Figure 5 online). Additional components in the cytosol, such as 14-3-3 proteins, might be involved in determining specific transit peptide recognition by different TOC receptors, but further research is necessary to establish the exact role of cytosolic guidance complexes in this process (May and Soll, 2000).

N-Terminal Met Excision and Acetylation of Plastid Precursor Proteins in the Cytosol

We found several unprocessed precursor proteins that were *N*-acetylated at position 2 most likely in the cytosol of Toc159-deficient plants (Figure 5). The detection of the unprocessed precursors raises the question of which specific features allow their accumulation in Toc159-deficient plants. The identified proteins are not simply abundant photosynthetic proteins that may saturate the impaired plastid import machinery in *ppi2* and *Toc159cs* and thus accumulate because of their high abundance. The accumulation of most high-abundance photosynthetic proteins seems to be prevented by regulatory signaling circuits that result in reduced expression of their genes (Figures 2 and 3). Interestingly, some of the unprocessed precursors of nonphotosynthetic proteins were also downregulated at the transcriptional level, but their decrease was not as pronounced as that of photosynthetic proteins (Table 1; see Supplemental Data Set 4 online). This suggests that these nonphotosynthetic proteins are integrated in a signaling circuit that is different from the signaling circuit that controls the expression of photosynthetic proteins. Together, our data support the view that the accumulation of nuclear-encoded plastid precursor proteins triggers an unimported precursor protein response (Kakizaki et al., 2009; Lee et al., 2009b). This signaling response could act through additional unknown signaling pathways that do not depend on the functional state of the plastids.

Our data show that plastid precursor proteins were modified by NME and subsequent *N*-acetylation outside the plastid. This modification mechanism is used for 60 to 90% of all proteins in eukaryotes (Goetze et al., 2009), and it is conceivable that it also modifies plastid precursor proteins under normal wild-type conditions during their cytosolic synthesis and transition. The sequence context around the identified *N*-acetylation sites is consistent with described requirements for this type of modification (i.e., small, uncharged residues like Ala follow the start Met) (Sherman et al., 1985; Meinnel et al., 2006). Our observation may therefore indicate a general modification pathway for plastid proteins in the cytosol. *N*-acetylation serves as a degradation signal for the ubiquitin ligase-proteasome system in yeast, and it was proposed to be involved in protein folding quality control (Hwang et al., 2010). Support for a function of plastid precursor protein *N*-acetylation in vivo came from the characterization of a mutant for a cytoplasmic *N*-acetyltransferase. This mutant is affected in chlorophyll accumulation and photosynthetic capacity, suggesting crosstalk between cytosolic *N*-acetylation and chloroplast function (Pesaresi et al., 2003).

The fact that plastid precursor stability is tightly controlled in the cytosol was recently substantiated by the detection of a

ubiquitin-dependent degradation pathway for nonimported Lhcb4 precursors that is mediated by the heat shock cognate protein 70-4 and ubiquitination by an E3 ligase (Lee et al., 2009b). Inhibition of the proteasome resulted in the enhanced accumulation of ubiquitinated Lhcb4 precursors in *ppi2*, thus indicating the existence of a degradation system for plastid precursors. Based on these observations, it is conceivable that cytosolic *N*-acetylation may serve as an additional degradation signal for nonimported plastid proteins to avoid accumulation of unprocessed precursor proteins in the cytosol. A short half-life of nonimported *N*-acetylated plastid precursor proteins could be the reason why these have not been identified in wild-type plants. Future work is needed to establish the role of *N*-acetylation in the assembly of the plastid proteome and its contribution to cellular function.

In summary, we used quantitative proteomics in combination with transcriptional profiling to establish a definition of Toc159-dependent and Toc159-independent plastid protein import. Based on our results, we could not identify any amino acid sequence specificity determinants that allow distinguishing Toc159-dependent and Toc159-independent proteins. The numerous photosynthetic proteins that accumulate as correctly processed mature proteins in *ppi2* and *Toc159cs* and the distinct functions affected in Toc159-deficient plastids furthermore indicate an unexpected client protein promiscuity of Toc159 and suggest that the current model for the import of photosynthetic and housekeeping nonphotosynthetic proteins must be revised. The identification of *N*-acetylated plastid precursor proteins outside the plastid in Toc159-deficient plants could point to a control mechanism for plastid precursor proteins that may prevent precursor accumulation under normal wild-type conditions. Altogether, we present evidence for complex regulation of plastid proteome assembly in import-deficient plastids that encompasses import preference, transcriptional regulation, and potentially differential degradation of plastid precursor proteins by *N*-acetylation–triggered processes. The quantitative contribution of each individual pathway must be assessed under different conditions, and more information on their dynamic regulation is necessary. This information in combination with increasingly sensitive analyses of precursor protein accumulation via detection of unprocessed N termini, possibly by combined fractional diagonal chromatography (Gevaert et al., 2007), in the future could provide a more comprehensive understanding of chloroplast proteome assembly.

METHODS

Plant Material

All experiments used *Arabidopsis thaliana* plants of accession Wassilewskija. Plants were grown at 20°C in photoperiods of 8 h light and 16 h darkness with a light intensity of 150 μmol m^{-2} s^{-1}. For protein extraction, leaves (leaf 6 from 5 to 10 plants per sample) were harvested after 35 d, 1 h before the end of the photoperiod and frozen in liquid nitrogen. The wild-type and *Toc159cs* plants were grown on soil, whereas wtS and *ppi2* were grown on half-strength Murashige and Skoog medium supplemented with 100 mM Suc. For each genotype, three independent biological replicates were performed (i.e., plants were grown three times under exactly the same conditions), and leaf 6 was harvested as described above. The genotypic characterization of *ppi2* and *Toc159cs* was described by Bauer et al. (2000) and in Supplemental Figure 1 online, respectively.

Generation of Toc159 Cosuppression Plants

A fragment from *Toc159* corresponding to the GTPase and membrane domains was amplified by PCR using 5′-CGGGATCCAAAATGGCTCAGGATCACCACCACCACCACCACGGCACGAAGCTTTTCTCTATGGAT-3′ (includes *Bam*HI) and 5′-GCTCTAGATTAGTACATGCTGTACTTGTCGTTCGTC-3′ (includes *Xba*I) primers, digested with *Bam*HI and *Xba*I, and inserted into the corresponding *Nco*I and *Xba*I sites from the pCHF7 binary vector (a gift from C. Fankhauser, University of Lausanne, Switzerland). Plants were transformed using the floral dip method as described (Clough and Bent, 1998). Transformants were selected on soil by spraying phosphinothricin (30 mg/L BASTA; Duchefa), and DNA prepared from leaf material was used for diagnostic PCR. Analysis of segregation was performed by germinating seeds on half-strength Murashige and Skoog medium and 30 μg/mL phosphinothricin. Phosphinothricin-sensitive plants displayed yellow cotyledons and were easily distinguishable from Toc159 cosuppressing plants with white leaves and cotyledons. Lines with a segregation indicative of a single T-DNA insertion and with high incidence of Toc159 cosuppression were selected. T-DNA insertions were mapped in these lines by three rounds of thermal asymmetric interlaced PCR using pCHF7_LB_a (5′-CGAACATCGGTCTCAATGCAAAAG-3′), pCHF7_LB_b (5′-CTACCTCGGCTCTGCGAAG-3′), pCHF7_LB_c (5′-CGGATACTTACGTCACGTCTTGC-3′), and random primers.

Protein Extraction and Immunoblotting

Protein extracts were prepared from leaves by grinding shock-frozen tissue. Subsequently, extraction buffer was added (40 mM Tris, pH 6.8, 40 mM DTT, 4% SDS, and 2× protease inhibitor cocktail [Roche]). Homogenates were centrifuged twice at 20,000 rcf for 20 min at room temperature. Protein concentration was determined using a BCA protein assay kit (Thermo Scientific) before adding DTT. Twenty micrograms of protein from leaves were subjected to SDS-PAGE on 12% gels. Immunoblotting using α-Toc159, α-PGL35, and α-CAB antibodies was performed as described previously (Ivanova et al., 2004). For each immunoblotting experiment, a second gel was prepared in parallel as loading control, which was stained with Coomassie Brilliant Blue according to standard procedures.

Sample Preparation for Large-Scale Proteomics and MS

Protein extraction was performed as described above, and 600 μg of protein from leaves and 200 μg from isolated chloroplasts were subjected to SDS-PAGE on 12% gels. After electrophoresis, gels were stained with Coomassie Brilliant Blue according to standard procedures and cut into 12 sections for leaves and 20 sections for isolated chloroplasts respectively. Each gel slice was diced into small pieces. In-gel digestion was performed according to Shevchenko et al. (1996). After digestion, dried peptides were resuspended in 3% acetonitrile and 0.2% trifluoretic acid and desalted using Sepak cartridges (Waters).

Dried peptides were resuspended in 3% acetonitrile and 0.2% formic acid and analyzed with a linear trap quadrupole Fourier transform–ion cyclotron resonance mass spectrometer (ThermoFischer Scientific) coupled to an Eksigent nano liquid chromatography system (Eksigent Technologies). Peptide mixtures were loaded onto laboratory-made capillary columns (BGB Analytik) of 75-μm inner diameter, 8-cm length, packed with 3 μm, 200 Å Magic C18 AQ beads (Michrom BioResources). Peptides were eluted from the column by an increased acetonitrile gradient from 3% acetonitrile, 0.2% formic acid to 36% acetonitrile, 0.2% formic acid over 55 min, and from 36% acetonitrile, 0.2% formic

acid to 80% acetonitrile, 0.2% formic acid over 5 min, followed by a 10-min wash step with 80% acetonitrile and 0.2% formic acid. Peptide ions were detected in a survey scan from mass-to-charge ratio 300 to 1600 at 100,000 full width half maximum nominal resolution followed by three data-dependent MS/MS scans (isolation width 2 atomic mass units, relative collision energy 35%, dynamic exclusion enabled, repeat count 1, followed by peak exclusion for 2 min).

Data Mining and Protein Quantification

MS/MS spectra were searched either with Sequest and PeptideProphet (Yates et al., 1995; Keller et al., 2002) using the Trans-Proteomic Pipeline (TPP v2.9) or with Mascot version 2.2.04 (Matrix Science) against the *Arabidopsis* TAIR9 protein database (downloaded on June 29, 2009) with a concatenated decoy database supplemented with contaminants (67,079 entries). For Sequest, the search parameters were as follows: requirement for tryptic ends, one missed cleavage allowed, mass tolerance = ± 5 ppm, variable modification of Met (M, PSI-MOD name: oxidation, ModAccession: MOD: 00412, mono Δ = 15.99491) and static modification of Cys (C, PSI-MOD name: iodoacetamide derivative, ModAccession: MOD: 00397, mono Δ = 57.0214). For PeptideProphet, the cutoff was set to a minimum probability of 0.9. For Mascot, the search parameters were as follows: requirement for semitryptic ends, one missed cleavage allowed, mass tolerance = ±5 ppm. Besides carbamidomethylation of Cys residues as fixed modification, oxidation of Met, N-terminal protein acetylation, and N-terminal acetylation were included as variable modifications. N-terminal acetylated peptides were accepted with a Mascot ion score higher than 24 and a Mascot expectation value smaller than 0.05. All peptide assignments from Sequest and Mascot except those of contaminants were filtered for ambiguity, and the peptides (N-terminal acetylated peptides included) matching to more than one protein were excluded from further analysis. This does not apply to different splice variants of the same protein or to different loci sharing exactly the same sequence. Different *N*-acetylated peptides matching to the same protein were also excluded from further analysis. Furthermore, spectrum assignments to decoy database peptides were excluded. All remaining spectrum assignments were entered into the pep2pro database and are available from the pep2pro website at www.pep2pro.ethz.ch (data set Plastid Proteome Assembly in Absence of Toc159). From the final data, PRIDE 2.1 XML files were created and exported to the PRIDE database (accession numbers from 14834 to 14839) (Vizcaíno et al., 2010). The spectrum false discovery rate was calculated by dividing the number of decoy database spectrum assignments by the number of spectrum assignments in the final data set. The false positive rate was between 0.34 and 0.74% for all measured biological replicates of leaves and of isolated chloroplasts.

Protein quantification with nSpC was done according to Baerenfaller et al. (2008). Briefly, the expected contribution of each individual protein to the samples total peptide pool was calculated correcting the values with a normalization factor, which balances for the theoretical number of tryptic peptides per protein and sample depth according to the following formula:

$$nSpC_K = Spectra_K \times \left(\frac{TTP_K \times MS}{MP} \right)^{-1}$$

where $nSpC_K$ is the normalized spectral count for protein K, TTP_K is the theoretical tryptic peptides of protein K, MS is the total number of measured spectra in the data set, And MP is the total number of theoretical tryptic peptides of the identified proteins in the data set.

For the determination of the number of TTP_K, the whole protein database was digested in silico. If Arg or Lys was followed by Pro (KP/RP site), the site was both cut and not cut (resulting in three tryptic peptides). If several of these sequence pairs followed each other, we only considered cutting one KP/RP site per time. The resulting peptides were labeled as theoretical tryptic peptides of 400 to 6000 D and at least six amino acids.

Statistical Analysis of the Data

In the statistical analysis of the data, only proteins identified with at least 10 spectra were taken into account. Four separate *t* tests (two-sided, Welch test) were done comparing the wild type and wtS, the wild type and *Toc159cs*, wtS and *ppi2*, and *Toc159cs* and *ppi2*. Proteins considered significantly more abundant in the wild type or the Toc159-deficient samples had to fulfill the following criteria: P $value_{wild\ type\_Toc159cs} < 0.05$, P $value_{wtS\_ppi} < 0.05$, P $value_{wild\ type\_wtS} \geq 0.05$ or not applicable, and P $value_{Toc159cs\_ppi2} \geq 0.05$ or not applicable. In addition to the significance test, the proteins had to show an at least 1.5-fold change in the comparisons wild-type/Toc159cs and wtS/ppi2. To the proteins fulfilling these criteria, we added those that were not identified at all in wild-type conditions but in at least two of the three replicates both for *ppi2* and *Toc159cs* (category "Only in *Toc159cs*/*ppi2*"), and those that were not at all identified in Toc159-deficient samples, but in at least two of the three replicates both for wild-type and wtS (category "Only in wild type/wtS"). All analyses were done with R (R Development Core Team, 2010).

Microarray Hybridization and Evaluation

RNA was extracted from 35-d-old plants. The experiment was performed with three biological replicates, and Affymetrix *Arabidopsis* AGRONOMICS1 microarrays were used. Labeling of samples, hybridizations, and measurements were performed as described (Hennig et al., 2004; Rehrauer et al., 2010). Signal values were derived using the robust multiarray analysis algorithm implemented in the statistical language R (R Development Core Team, 2010) using probe sets comprising exonic probes based on the TAIR9 genome annotation. For details of probe set definition and low-level data analysis, see Rehrauer et al. (2010). Differentially expressed genes were selected using LIMMA (Smyth, 2004) followed by multiple testing correction according to Storey and Tibshirani (2003). Genes were considered as differentially expressed if P value < 0.05. This resulted in a list of 15,882 genes. Downstream analysis was performed on a subset of the differentially expressed genes that had a fold change of more than 2.Significance of overlaps of gene sets was calculated with a Fisher's exact test. The microarray data have been submitted to ArrayExpress with the experiment number E-TABM-951.

Chloroplast Localization, Sequence Alignment, and Functional Annotation

Chloroplast localization annotation was based on a chloroplast protein reference table that comprises 1155 proteins (reviewed in Baginsky and Gruissem, 2009). Plastid localization of these 1155 proteins were validated in different proteomics studies by software tools that integrate targeting prediction with coexpression data and several other criteria that were combined and weighted with a naïve Bayesian classifier. Amino acid sequence alignment and visualization were performed with Weblogo (weblogo.berkeley.edu) (Crooks et al., 2004). In silico localization prediction was done using TargetP (Emanuelsson et al., 2000). Functional annotation was based on MapMan (mapman.mpimp-golm.mpg.de) (Usadel et al., 2009), and the overrepresentation of functional categories was assessed with Fisher's exact test.

Chloroplast Isolation

Intact chloroplasts from 35-d-old wild-type or *ppi2* plants grown on soil or on half-strength Murashige and Skoog medium supplemented with 100 mM Suc medium were isolated according to Fitzpatrick and Keegstra (2001) with the following modifications. Cellulase Onozuka R-10 and macerozyme R-10 (Serva; catalog numbers 16419 and 28302) were diluted to 1 and 0.25% (w/v), respectively. L-ascorbate (0.1% [w/v]) and 0.05% (v/v) protease inhibitor cocktail for plant cell extracts (Sigma-Aldrich;

catalog number P9599) was added to the breakage buffer. Intact wild-type chloroplasts were harvested at the interphase of a Percoll step gradient (40% [v/v] and 85% [v/v]) and *ppi2* chloroplasts on a linear gradient of 10 to 60%. Different density gradient centrifugation protocols were used for wild-type and *ppi2* plastids because they differ in their density and the linear gradients resulted in improved purity. To minimize contamination from other cellular compartments, isolated chloroplasts were subjected to three successive Percoll gradients, including three washings steps.

Expression of GFP Fusions in Wild-Type and *Toc159cs* Leaves

Expression in wild-type and *Toc159cs* leaves of a GFP fusion to the transit peptide of ribulose-1,5-bis-phosphate carboxylase/oxygenase small subunit (pSSU-GFP) or of a GFP fusion to an N-terminal fragment of PLG35 (residues 1 to 74) was achieved using biolistic transformation (Vidi et al., 2006). Gold particles (1.0 μm; Bio-Rad) were sterilized in ethanol and resuspended in water to a final concentration of 60 μg/μL. Plasmid DNA (5 μg) was precipitated on the gold carrier (50 μL) by adding 50 μL of 2.5 M $CaCl_2$ and 20 μL of 0.1 M spermidine under continuous vortexing. After washing with ethanol, DNA-coated gold particles were suspended in ethanol, spread on macrocarrier discs (Bio-Rad), and used for biolistic transformation. Leaves of mature *Arabidopsis* plants grown on soil were placed upside down on half-strength Murashige and Skoog medium and bombarded with a PDS-1000/He Biolistic R Particle Delivery System (Bio-Rad) operating at 1100 p.s.i. He pressure. Forty-eight hours after transformation, leaves were placed on glass slides for confocal microscopy imaging (Leica TCS 4D). GFP fluorescence was monitored using the fluorescein isothiocyanate (488 nm) laser line, and plastid autofluorescence was monitored using tetramethyl rhodamine isothiocyanate (568 nm) excitation. Images were processed using ImageJ (http://rsbweb.nih.gov/ij/).

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Characterization of Toc159 Cosuppression Lines.

Supplemental Figure 2. Protein and Transcript Abundances of the Components of the TIC/TOC Import Machinery.

Supplemental Figure 3. Proteome Profiling of Isolated Wild-Type and *ppi2* Chloroplasts.

Supplemental Figure 4. Extracted Ion Chromatogram Quantification for N-Terminal Peptides in Leaves and Isolated Plastids from the Light-Harvesting Complex Protein LHCA3 (AT1G61520).

Supplemental Figure 5. Toc159-Independent and Toc159-Dependent Substrates.

Supplemental Data Set 1. Proteins Identified by MS/MS in Wild-Type, wtS, *Toc159cs*, and *ppi2* Leaves.

Supplemental Data Set 2. Reference List of 1155 High-Confidence Plastid Proteins.

Supplemental Data Set 3. Statistical Analysis of Differential Protein Accumulation in Wild-Type and Toc159-Deficient Plants Compared with Their Transcriptional Regulation.

Supplemental Data Set 4. Transcriptional Profiling Data for wtS and *ppi2* Plants.

Supplemental Data Set 5. Nuclear-Encoded Proteins with a Minimal Peptide Position between 1 and 10.

Supplemental Data Set 6. *N*-Acetylated Peptides Identified by MS/MS in Wild-Type, wtS, *Toc159cs*, and *ppi2* leaves.

Supplemental Data Set 7. Plastid Proteins Identified by MS/MS in Wild-Type, wtS, *Toc159cs*, and *ppi2* Leaves *N*-Acetylated between Positions 1 and 10.

Supplemental Data Set 8. Stromal Proteins Involved in the Plastid Posttranslational Protein Homeostasis Network (Zybailov et al., 2008).

Supplemental Data Set 9. Proteins Identified in Isolated *ppi2* and Wild-Type Plastids.

Supplemental Data Set 10. *N*-Acetylated Peptides Identified by MS/MS in Isolated Wild-Type and *ppi2* Plastids.

Supplemental Data Set 11. Toc159-Independent and Toc159-Dependent Proteins.

ACKNOWLEDGMENTS

We thank members of the Functional Genomics Center Zurich, Dorothea Rutishauser, Peter Gehrig, Claudia Fortes, Paolo Nanni, Jonas Grossmann, Christian Panse, and Simon Barkow-Oesterreicher for their support in MS and bioinformatics. This work was supported by Eidgenössische Technische Hochschule Zurich, by Schweizerischer Nationalfonds Grant 31003A_127380/1 to F.K., and Schweizerischer Nationalfonds Grant 3100A0-112565 to S. Baginsky.

AUTHOR CONTRIBUTIONS

S. Bischof, K.B, W.G., and S. Baginsky designed the research. S. Bischof, K.B., T.W., R.T., P.-A.V., B.R., M.H.-H., and S. Baginsky performed the research. S. Bischof, K.B., L.H., F.K., and S. Baginsky analyzed data. S. Bischof, W.G., and S. Baginsky wrote the article.

REFERENCES

Agne, B., Andrès, C., Montandon, C., Christ, B., Ertan, A., Jung, F., Infanger, S., Bischof, S., Baginsky, S., and Kessler, F. (2010). The acidic A-domain of Arabidopsis TOC159 occurs as a hyperphosphorylated protein. Plant Physiol. **153:** 1016–1030.

Agne, B., Infanger, S., Wang, F., Hofstetter, V., Rahim, G., Martin, M., Lee, D.W., Hwang, I., Schnell, D.J., and Kessler, F. (2009). A toc159 import receptor mutant, defective in hydrolysis of GTP, supports preprotein import into chloroplasts. J. Biol. Chem. **284:** 8670–8679.

Agne, B., and Kessler, F. (2009). Protein transport in organelles: The Toc complex way of preprotein import. FEBS J. **276:** 1156–1165.

Baerenfaller, K., Grossmann, J., Grobei, M.A., Hull, R., Hirsch-Hoffmann, M., Yalovsky, S., Zimmermann, P., Grossniklaus, U., Gruissem, W., and Baginsky, S. (2008). Genome-scale proteomics reveals *Arabidopsis thaliana* gene models and proteome dynamics. Science **320:** 938–941.

Baginsky, S., and Gruissem, W. (2009). The chloroplast kinase network: new insights from large-scale phosphoproteome profiling. Mol. Plant **2:** 1141–1153.

Balsera, M., Soll, J., and Bölter, B. (2009). Protein import machineries in endosymbiotic organelles. Cell. Mol. Life Sci. **66:** 1903–1923.

Bauer, J., Chen, K., Hiltbunner, A., Wehrli, E., Eugster, M., Schnell, D., and Kessler, F. (2000). The major protein import receptor of plastids is essential for chloroplast biogenesis. Nature **403:** 203–207.

Beck, M., Claassen, M., and Aebersold, R. (2011). Comprehensive proteomics. Curr. Opin. Biotechnol. **22:** 3–8.

Biehl, A., Richly, E., Noutsos, C., Salamini, F., and Leister, D. (2005). Analysis of 101 nuclear transcriptomes reveals 23 distinct regulons and their relationship to metabolism, chromosomal gene distribution and co-ordination of nuclear and plastid gene expression. Gene **344:** 33–41.

Bläsing, OE., Gibon, Y., Günther, M., Höhne, M., Morcuende, R., Osuna, D., Thimm, O., Usadel, B., Scheible, WR., and Stitt, M. (2005). Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in *Arabidopsis*. Plant Cell **17:** 3257–3281.

Bruce, B.D. (2001). The paradox of plastid transit peptides: Conservation of function despite divergence in primary structure. Biochim. Biophys. Acta **1541:** 2–21.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. Plant J. **16:** 735–743.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: A sequence logo generator. Genome Res. **14:** 1188–1190.

Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. **300:** 1005–1016.

Ferro, M., et al. (2010). AT_CHLORO, a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins. Mol. Cell. Proteomics **9:** 1063–1084.

Ferro, M., Salvi, D., Brugière, S., Miras, S., Kowalski, S., Louwagie, M., Garin, J., Joyard, J., and Rolland, N. (2003). Proteomics of the chloroplast envelope membranes from *Arabidopsis thaliana*. Mol. Cell. Proteomics **2:** 325–345.

Fitzpatrick, L.M., and Keegstra, K. (2001). A method for isolating a high yield of Arabidopsis chloroplasts capable of efficient import of precursor proteins. Plant J. **27:** 59–65.

Gevaert, K., Impens, F., Van Damme, P., Ghesquière, B., Hanoulle, X., and Vandekerckhove, J. (2007). Applications of diagonal chromatography for proteome-wide characterization of protein modifications and activity-based analyses. FEBS J. **274:** 6277–6289.

Gibson, S.I. (2005). Control of plant development and gene expression by sugar signaling. Curr. Opin. Plant Biol. **8:** 93–102.

Goetze, S., et al. (2009). Identification and functional characterization of N-terminally acetylated proteins in *Drosophila melanogaster*. PLoS Biol. **7:** e1000236.

Hennig, L., Gruissem, W., Grossniklaus, U., and Köhler, C. (2004). Transcriptional programs of early reproductive stages in Arabidopsis. Plant Physiol. **135:** 1765–1775.

Hiltbrunner, A., Grünig, K., Alvarez-Huerta, M., Infanger, S., Bauer, J., and Kessler, F. (2004). AtToc90, a new GTP-binding component of the Arabidopsis chloroplast protein import machinery. Plant Mol. Biol. **54:** 427–440.

Hwang, C.S., Shemorry, A., and Varshavsky, A. (2010). N-terminal acetylation of cellular proteins creates specific degradation signals. Science **327:** 973–977.

Infanger, S., Bischof, S., Hiltbrunner, A., Agne, B., Baginsky, S., and Kessler, F. (2011). The chloroplast import receptor Toc90 partially restores the accumulation of Toc159 client proteins in the *Arabidopsis thaliana* ppi2 mutant. Mol. Plant **4:** 252–263.

Inoue, H., Rounds, C., and Schnell, D.J. (2010). The molecular basis for distinct pathways for protein import into *Arabidopsis* chloroplasts. Plant Cell **22:** 1947–1960.

Ivanova, Y., Smith, M.D., Chen, K., and Schnell, D.J. (2004). Members of the Toc159 import receptor family represent distinct pathways for protein targeting to plastids. Mol. Biol. Cell **15:** 3379–3392.

Jarvis, P., Chen, L.J., Li, H., Peto, C.A., Fankhauser, C., and Chory, J. (1998). An Arabidopsis mutant defective in the plastid general protein import apparatus. Science **282:** 100–103.

Kakizaki, T., Matsumura, H., Nakayama, K., Che, F.S., Terauchi, R., and Inaba, T. (2009). Coordination of plastid protein import and nuclear gene expression by plastid-to-nucleus retrograde signaling. Plant Physiol. **151:** 1339–1353.

Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal. Chem. **74:** 5383–5392.

Kleffmann, T., von Zychlinski, A., Russenberger, D., Hirsch-Hoffmann, M., Gehrig, P., Gruissem, W., and Baginsky, S. (2007). Proteome dynamics during plastid differentiation in rice. Plant Physiol. **143:** 912–923.

Koussevitzky, S., Nott, A., Mockler, T.C., Hong, F., Sachetto-Martins, G., Surpin, M., Lim, J., Mittler, R., and Chory, J. (2007). Signals from chloroplasts converge to regulate nuclear gene expression. Science **316:** 715–719.

Kubis, S., Baldwin, A., Patel, R., Razzaq, A., Dupree, P., Lilley, K., Kurth, J., Leister, D., and Jarvis, P. (2003). The *Arabidopsis* ppi1 mutant is specifically defective in the expression, chloroplast import, and accumulation of photosynthetic proteins. Plant Cell **15:** 1859–1871.

Kubis, S., Patel, R., Combe, J., Bédard, J., Kovacheva, S., Lilley, K., Biehl, A., Leister, D., Ríos, G., Koncz, C., and Jarvis, P. (2004). Functional specialization amongst the *Arabidopsis* Toc159 family of chloroplast protein import receptors. Plant Cell **16:** 2059–2077.

Larkin, R.M., Alonso, J.M., Ecker, J.R., and Chory, J. (2003). GUN4, a regulator of chlorophyll synthesis and intracellular signaling. Science **299:** 902–906.

Lee, D.W., Kim, J.K., Lee, S., Choi, S., Kim, S., and Hwang, I. (2008). *Arabidopsis* nuclear-encoded plastid transit peptides contain multiple sequence subgroups with distinctive chloroplast-targeting sequence motifs. Plant Cell **20:** 1603–1622.

Lee, D.W., Lee, S., Oh, Y.J., and Hwang, I. (2009a). Multiple sequence motifs in the rubisco small subunit transit peptide independently contribute to Toc159-dependent import of proteins into chloroplasts. Plant Physiol. **151:** 129–141.

Lee, S., Lee, D.W., Lee, Y., Mayer, U., Stierhof, Y.D., Lee, S., Jürgens, G., and Hwang, I. (2009b). Heat shock protein cognate 70-4 and an E3 ubiquitin ligase, CHIP, mediate plastid-destined precursor degradation through the ubiquitin-26S proteasome system in *Arabidopsis*. Plant Cell **21:** 3984–4001.

May, T., and Soll, J. (2000). 14-3-3 Proteins form a guidance complex with chloroplast precursor proteins in plants. Plant Cell **12:** 53–64.

Meinnel, T., Serero, A., and Giglione, C. (2006). Impact of the N-terminal amino acid on targeted protein degradation. Biol. Chem. **387:** 839–851.

Mochizuki, N., Brusslan, J.A., Larkin, R., Nagatani, A., and Chory, J. (2001). Arabidopsis genomes uncoupled 5 (GUN5) mutant reveals the involvement of Mg-chelatase H subunit in plastid-to-nucleus signal transduction. Proc. Natl. Acad. Sci. USA **98:** 2053–2058.

Mochizuki, N., Tanaka, R., Tanaka, A., Masuda, T., and Nagatani, A. (2008). The steady-state level of Mg-protoporphyrin IX is not a determinant of plastid-to-nucleus signaling in Arabidopsis. Proc. Natl. Acad. Sci. USA **105:** 15184–15189.

Moulin, M., McCormac, A.C., Terry, M.J., and Smith, A.G. (2008). Tetrapyrrole profiling in Arabidopsis seedlings reveals that retrograde plastid nuclear signaling is not due to Mg-protoporphyrin IX accumulation. Proc. Natl. Acad. Sci. USA **105:** 15178–15183.

Pesaresi, P., Gardner, N.A., Masiero, S., Dietzmann, A., Eichacker, L., Wickner, R., Salamini, F., and Leister, D. (2003). Cytoplasmic N-terminal protein acetylation is required for efficient photosynthesis in *Arabidopsis*. Plant Cell **15:** 1817–1832.

R Development Core Team (2010). R: A Language and Environment for Statistical Computing. (Vienna, Austria: R Foundation for Statistical Computing).

Rehrauer, H., et al. (2010). AGRONOMICS1: A new resource for Arabidopsis transcriptome profiling. Plant Physiol. **152:** 487–499.

Richardson, L.G., Jelokhani-Niaraki, M., and Smith, M.D. (2009). The acidic domains of the Toc159 chloroplast preprotein receptor family are intrinsically disordered protein domains. BMC Biochem. **10:** 35.

Richly, E., Dietzmann, A., Biehl, A., Kurth, J., Laloi, C., Apel, K., Salamini, F., and Leister, D. (2003). Covariations in the nuclear chloroplast transcriptome reveal a regulatory master-switch. EMBO Rep. **4:** 491–498.

Richter, S., and Lamppa, G.K. (1998). A chloroplast processing enzyme functions as the general stromal processing peptidase. Proc. Natl. Acad. Sci. USA **95:** 7463–7468.

Richter, S., and Lamppa, G.K. (1999). Stromal processing peptidase binds transit peptides and initiates their ATP-dependent turnover in chloroplasts. J. Cell Biol. **147:** 33–44.

Rudhe, C., Clifton, R., Chew, O., Zemam, K., Richter, S., Lamppa, G., Whelan, J., and Glaser, E. (2004). Processing of the dual targeted precursor protein of glutathione reductase in mitochondria and chloroplasts. J. Mol. Biol. **343:** 639–647.

Sherman, F., Stewart, J.W., and Tsunasawa, S. (1985). Methionine or not methionine at the beginning of a protein. Bioessays **3:** 27–31.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal. Chem. **68:** 850–858.

Smith, M.D., Rounds, C.M., Wang, F., Chen, K., Afitlhile, M., and Schnell, D.J. (2004). atToc159 is a selective transit peptide receptor for the import of nucleus-encoded chloroplast proteins. J. Cell Biol. **165:** 323–334.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. **3:** e3.

Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. USA **100:** 9440–9445.

Strand, A., Asami, T., Alonso, J., Ecker, J.R., and Chory, J. (2003). Chloroplast to nucleus communication triggered by accumulation of Mg-protoporphyrinIX. Nature **421:** 79–83.

Sun, Q., Zybailov, B., Majeran, W., Friso, G., Olinares, P.D., and van Wijk, K.J. (2009). PPDB, the Plant Proteomics Database at Cornell. Nucleic Acids Res. **37**(Database issue): D969–D974.

Trösch, R., and Jarvis, P. (2011). The stromal processing peptidase of chloroplasts is essential in *Arabidopsis*, with knockout mutations causing embryo arrest after the 16-cell stage. PLoS ONE **6:** e23039.

Usadel, B., Poree, F., Nagel, A., Lohse, M., Czedik-Eysenberg, A., and Stitt, M. (2009). A guide to using MapMan to visualize and compare Omics data in plants: A case study in the crop species, Maize. Plant Cell Environ. **32:** 1211–1229.

Vidi, P.A., Kanwischer, M., Baginsky, S., Austin, J.R., Csucs, G., Dörmann, P., Kessler, F., and Bréhélin, C. (2006). Tocopherol cyclase (VTE1) localization and vitamin E accumulation in chloroplast plastoglobule lipoprotein particles. J. Biol. Chem. **281:** 11225–11234.

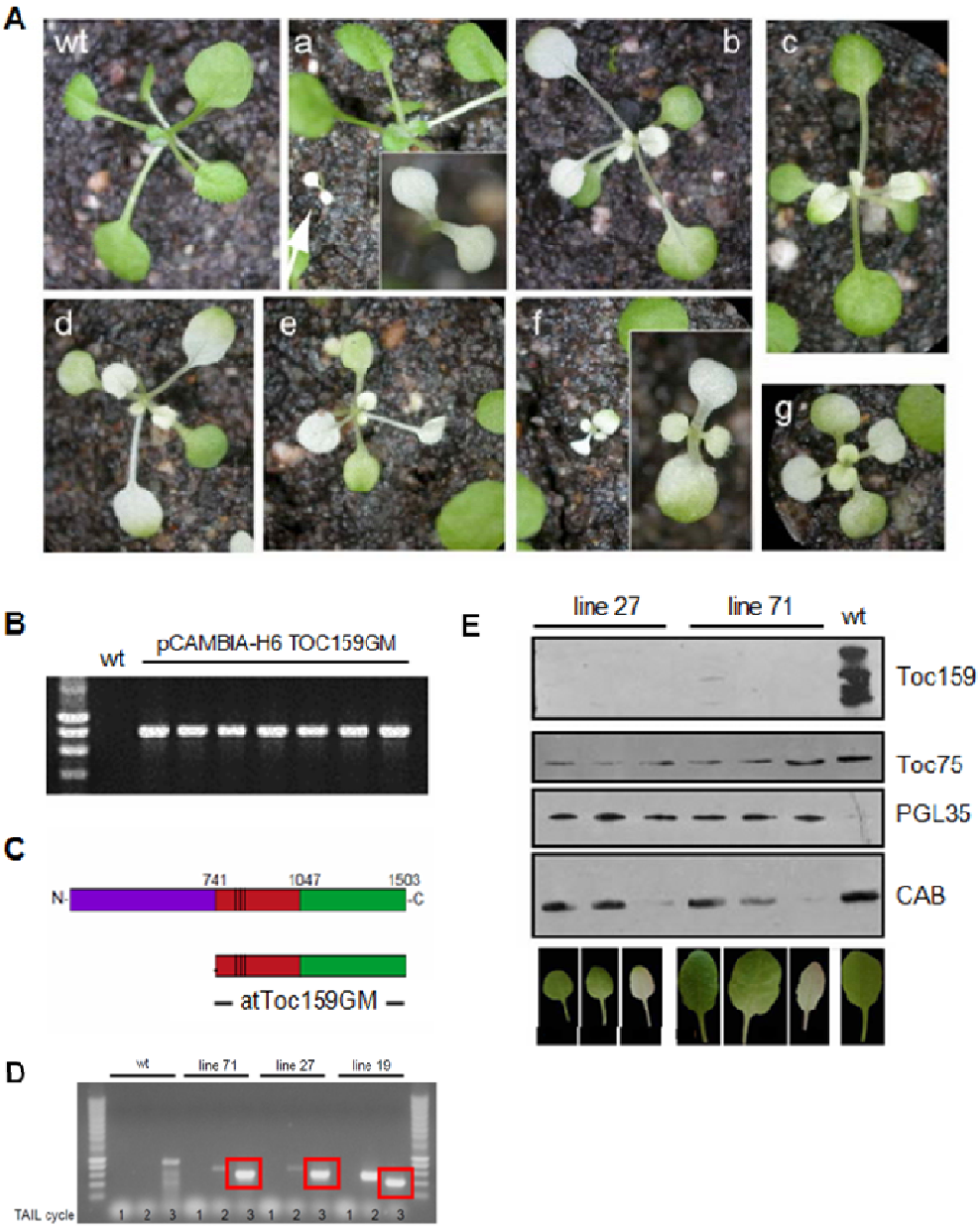
Vizcaíno, J.A., Reisinger, F., Côté, R., and Martens, L. (2010). PRIDE: Data submission and analysis. Curr. Protoc. Protein Sci. **25:** 4.

Yates III, J.R., Eng, J.K., McCormack, A.L., and Schieltz, D. (1995). Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. Anal. Chem. **67:** 1426–1436.

Zhang, X.P., and Glaser, E. (2002). Interaction of plant mitochondrial and chloroplast signal peptides with the Hsp70 molecular chaperone. Trends Plant Sci. **7:** 14–21.

Zybailov, B., Friso, G., Kim, J., Rudella, A., Rodríguez, V.R., Asakura, Y., Sun, Q., and van Wijk, K.J. (2009). Large scale comparative proteomics of a chloroplast Clp protease mutant reveals folding stress, altered protein homeostasis, and feedback regulation of metabolism. Mol. Cell. Proteomics **8:** 1789–1810.

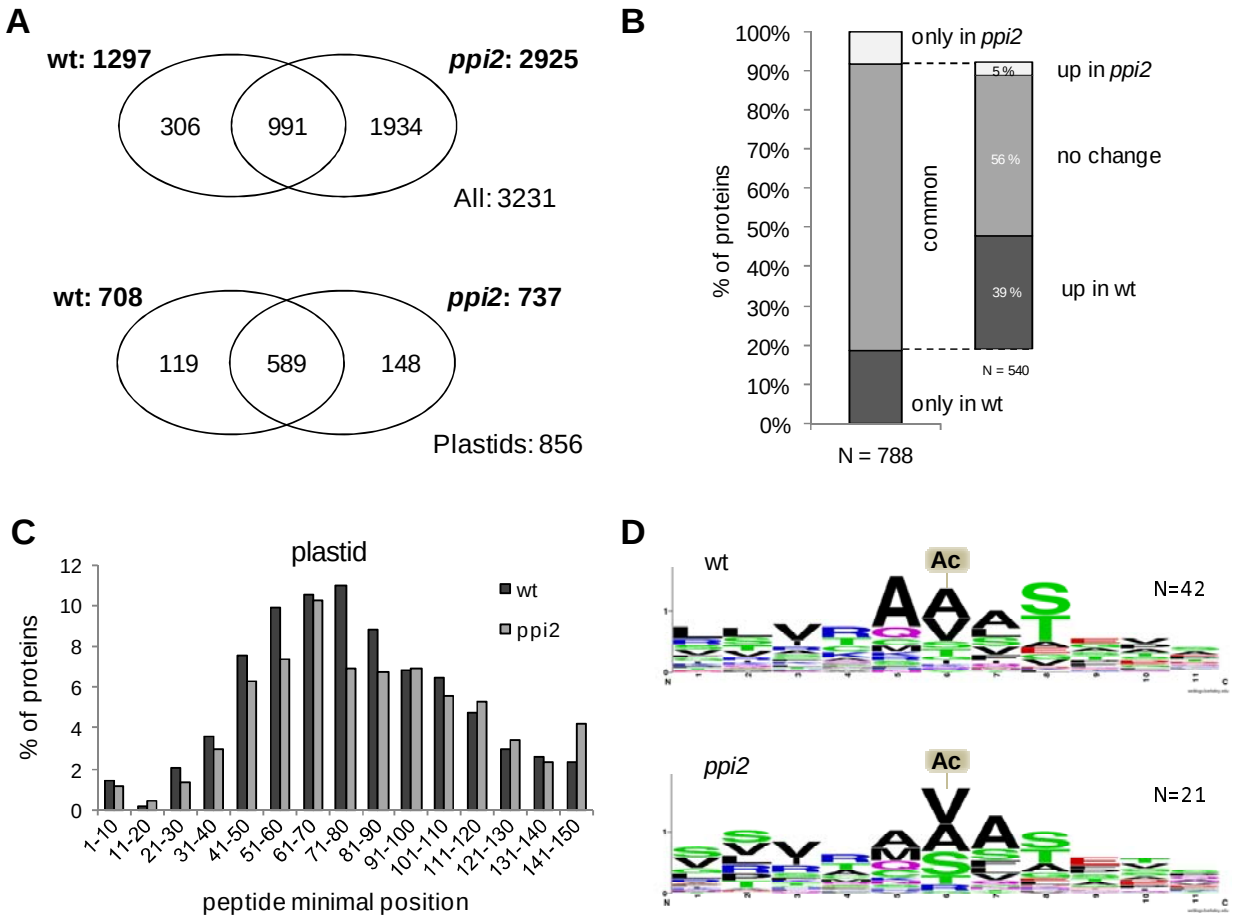
Zybailov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O., Sun, Q., and van Wijk, K.J. (2008). Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. PLoS One **3:**e1994.



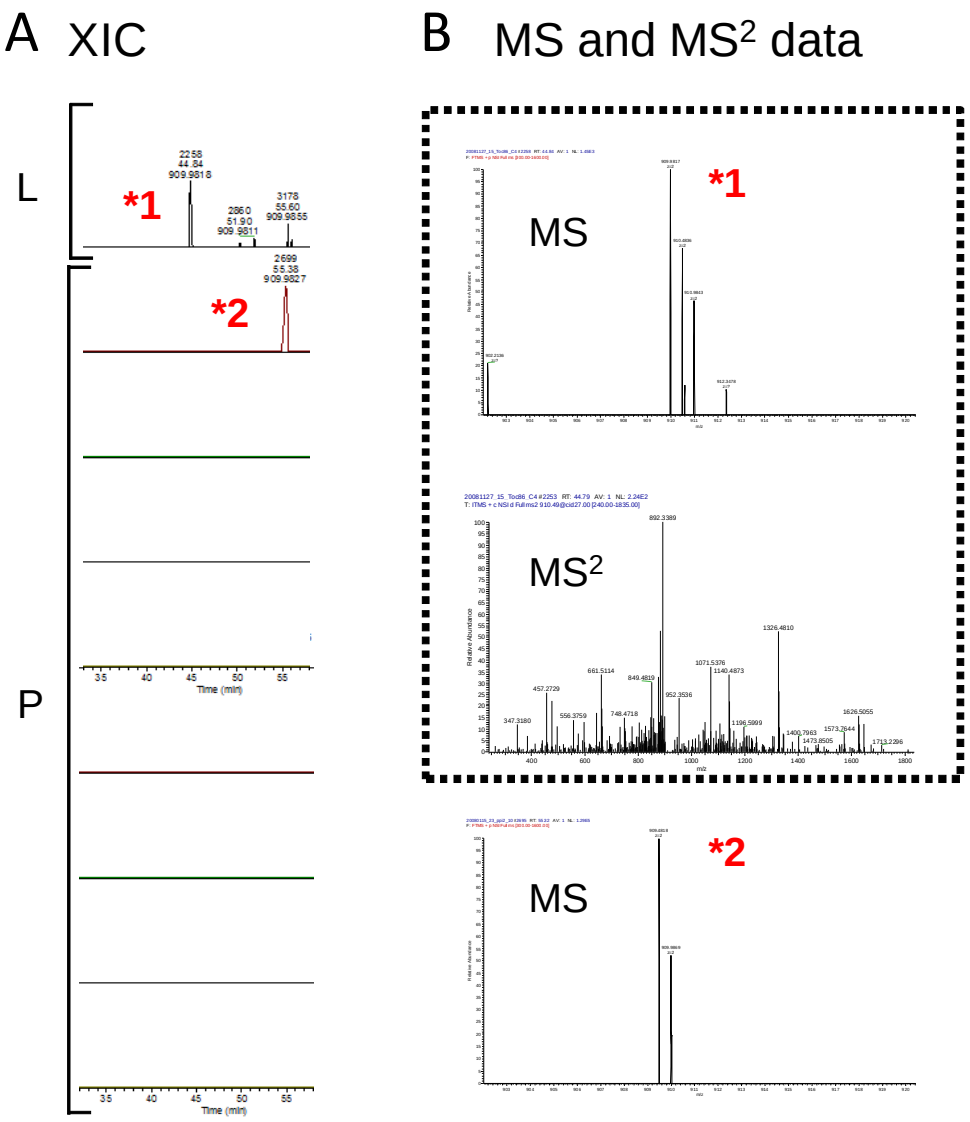
Supplemental Figure 1: Characterization of Toc159 co-suppression lines
(A) Phenotype of *Arabidopsis* plants co-suppressing atToc159. Images from wild-type (wt) and from plants transformed with the pCAMBIA-atTOC159GM vector encoding an truncated form of atToc159 under the control of the 35S promoter (a-g). Transformants showed various extent of leaf depigmentation. (B) PCR analysis showing the presence of the atToc86 transgene in the plants a-g imaged in A. (C) Domain structure of atToc159. AtToc159 GM corresponds to the GTPase (red) and membrane (green) domains of atToc159. Amino acid positions are indicated. (D) TAIL-PCR characterization of T-DNA LB flanking sequences with lines whose segragation pattern suggested single insertions. Insertions in lines 27 and 71 were not in a predcited coding seqeunce. (E) Western blot analysis of atToc86 transformant lines 27 and 71. Total leaf protein extracts (25 µg) were probed using antibodies specific to atToc159, atToc75 and atCAB as indicated. The leaves analyzed (images shown at the bottom) showed depigmentation to different degrees.

TOC components		Protein abundance				Transcript
		wt	wtS	*Toc159cs*	*ppi2*	*ppi2* / wt
AT4G02510	TOC159	1.35	1.01	0	0	-7.15
AT1G02280	TOC33	0.39	0.33	0.95	0.19	2.25
AT5G05000	TOC34	0.35	0.2	0.08	0.09	1.94
AT3G46740	TOC75-III	1.99	1.95	0.79	0.48	1.39
AT5G19620	OEP80	0.05	0.14	0.17	0.11	-1.14
AT3G17970	TOC64-III	0.16	0.17	0.44	0.13	1.33
TIC components						
AT2G47840	TIC20-II	0.38	0.9	0.26	0.27	1.08
AT5G55710	TIC20-V	0.07	0.83	0	0.15	-2.21
AT2G15290	TIC21	0.17	0.25	0.14	0.44	-1.35
AT4G33350	TIC22	0.33	0.47	0.4	0.22	-1.59
AT3G23710	TIC22	0.04	0.04	0.52	0.38	1.65
AT4G23430	TIC32-IV	0.03	0	0.08	0	-2.73
AT5G16620	TIC40	0.65	0.57	1.25	0.74	1.46
AT2G24820	TIC55	0.25	0.69	0.54	0.56	-3.07
AT3G18890	TIC62	2.77	2.23	0	0.27	-4.00
AT1G06950	TIC110	5.34	4.52	2.72	4.38	2.32
Heat shock proteins and stromal processing peptidase						
AT3G48870	HSP93-III	0.65	0.48	1.25	1.45	2.77
AT5G50920	HSP93-V	4.58	3.81	4.72	3.85	-1.77
AT5G42390	SPP	0.61	0.42	0.21	0.1	-1.78
TOC and TIC components not identified by MS/MS						
AT2G16640	TOC132	0	0	0	0	1.17
AT3G16620	TOC120	0	0	0	0	1.72
AT5G20300	TOC90	0	0	0	0	1.00
AT1G35860	TOC75-I	0	0	0	0	-1.23
AT4G09080	TOC75-IV	0	0	0	0	-1.06
AT5G09420	TOC64-V	0	0	0	0	3.45
AT1G08980	TOC64-I	0	0	0	0	-1.40
AT1G80920	TOC12	0	0	0	0	-1.35
AT4G11410	TIC32-IVb	0	0	0	0	-1.37
AT1G04940	TIC20-I	0	0	0	0	-1.13
AT4G03320	TIC20-IV	0	0	0	0	7.87

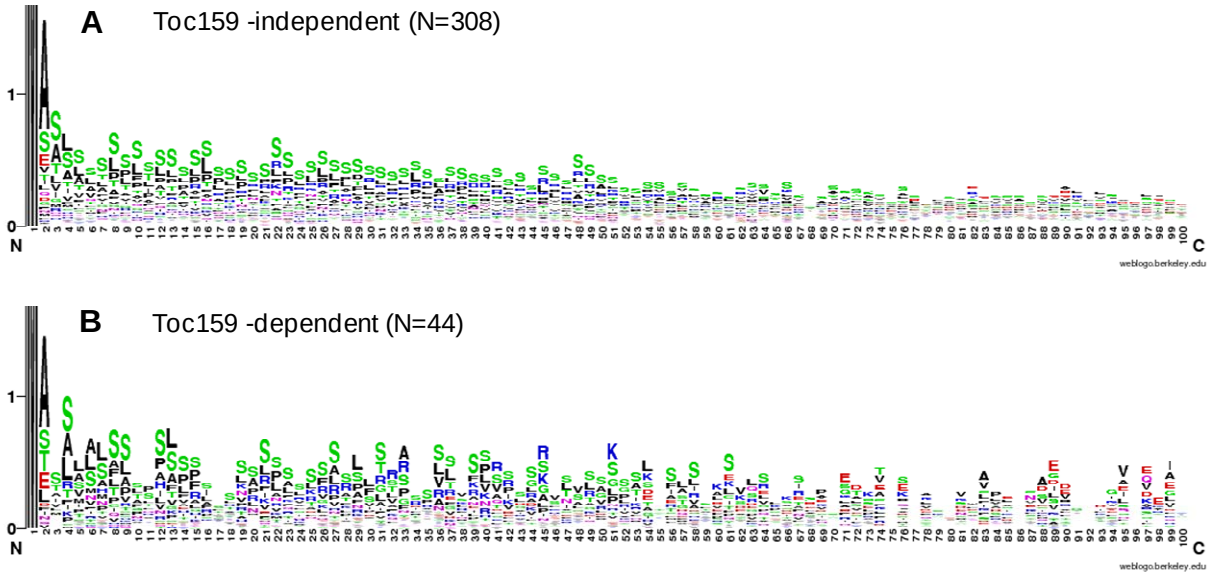
Supplemental Figure 2: Quantification of components of the plastid import machinery.
Provided are the APEX proteome quantification data and the transcript ratios between wt and *ppi2* (see Materials and Methods).



Supplemental Figure 3: Proteome profiling of isolated wt and *ppi2* chloroplasts.
(A) Proteins identified by mass spectrometry in isolated wt and *ppi2* plastids. Plastid localization is based on a chloroplast proteome reference table of 1153 proteins.
(B) Comparison between nuclear-encoded plastid proteins identified in wt and *ppi2* plastids. Down- and up-regulated plastid proteins have a two times lower or higher mean abundance in *ppi2* compared to wt.
(C) Peptide minimal position of nuclear-encoded plastid proteins. The position of the most N-terminal detected peptides of each identified protein were grouped into bins. The reduced detection of peptides at the N-terminus of plastid proteins supports the removal of transit peptides after import in wt and in *ppi2* plants.
(D) Sequence context around N-acetylation sites of plastid proteins. The similar sequence context in wt and in *ppi2* plants indicates that precursor processing of imported proteins is also functional in *ppi2*. N-acetylation sites ranging between positions 25 and 90 were taken for alignment.



Supplemental Figure 4: Extracted ion chromatogram quantification of the peptide identified in the transit peptide region of LHCA3 (AT1G61520).
(A) Extracted exact mass ion chromatogram for the spectra matching to the transit peptide region from leaves (upper panel, marked with an *1) and the corresponding samples from isolated plastids (marked with *2). In one of the samples from isolated plastids (three wt and three *ppi2*), an ion matching the exact mass of the LHCA3 transit peptide (*2) was detected at a shifted retention time of approx. 11 minutes (panels below). L: Leaves, P: Plastids.
(B) MS/MS (MS^2) information on this ion compared to the transit peptide peak (*1) revealed their difference. The peptide identified from isolated plastids represents the A+1 isotope of an unrelated parent ion, as opposed to the monoisotopic (A peak) mass of the transit peptide peak (*1). XIC: extracted ion chromatogram.



Supplemental Figure 5: Toc159-independent and Toc159-dependent substrates.
(A) Transit peptide alignment of the 100 first amino acids of 308 Toc159-independent proteins.
(B) Transit peptide alignment of the 100 first amino acids of 40 Toc159-dependent proteins.