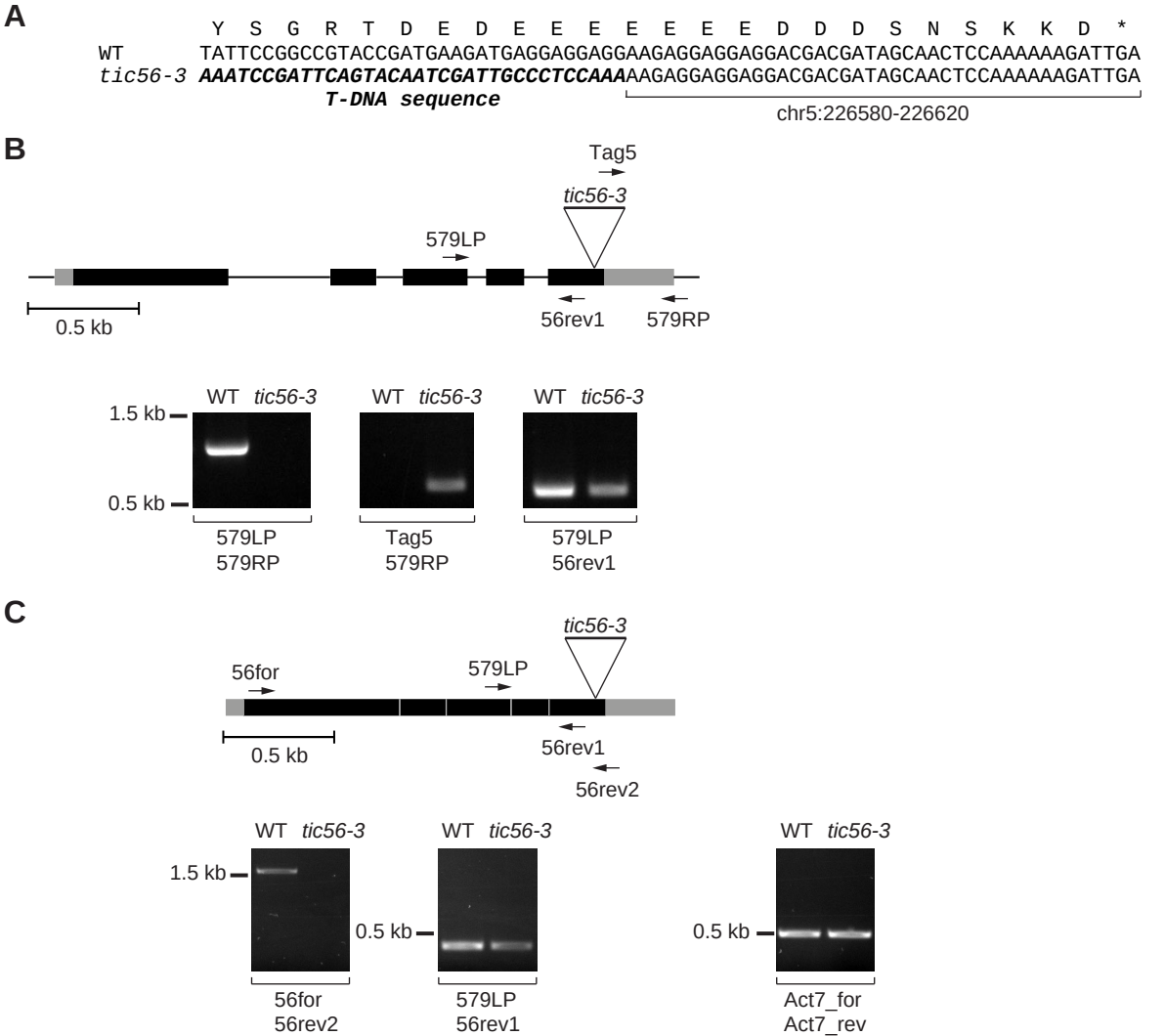
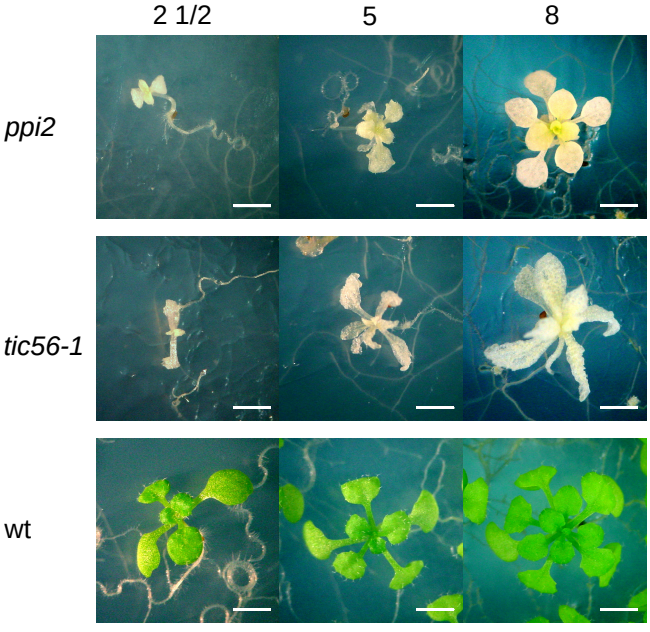
Supplemental Figure 1



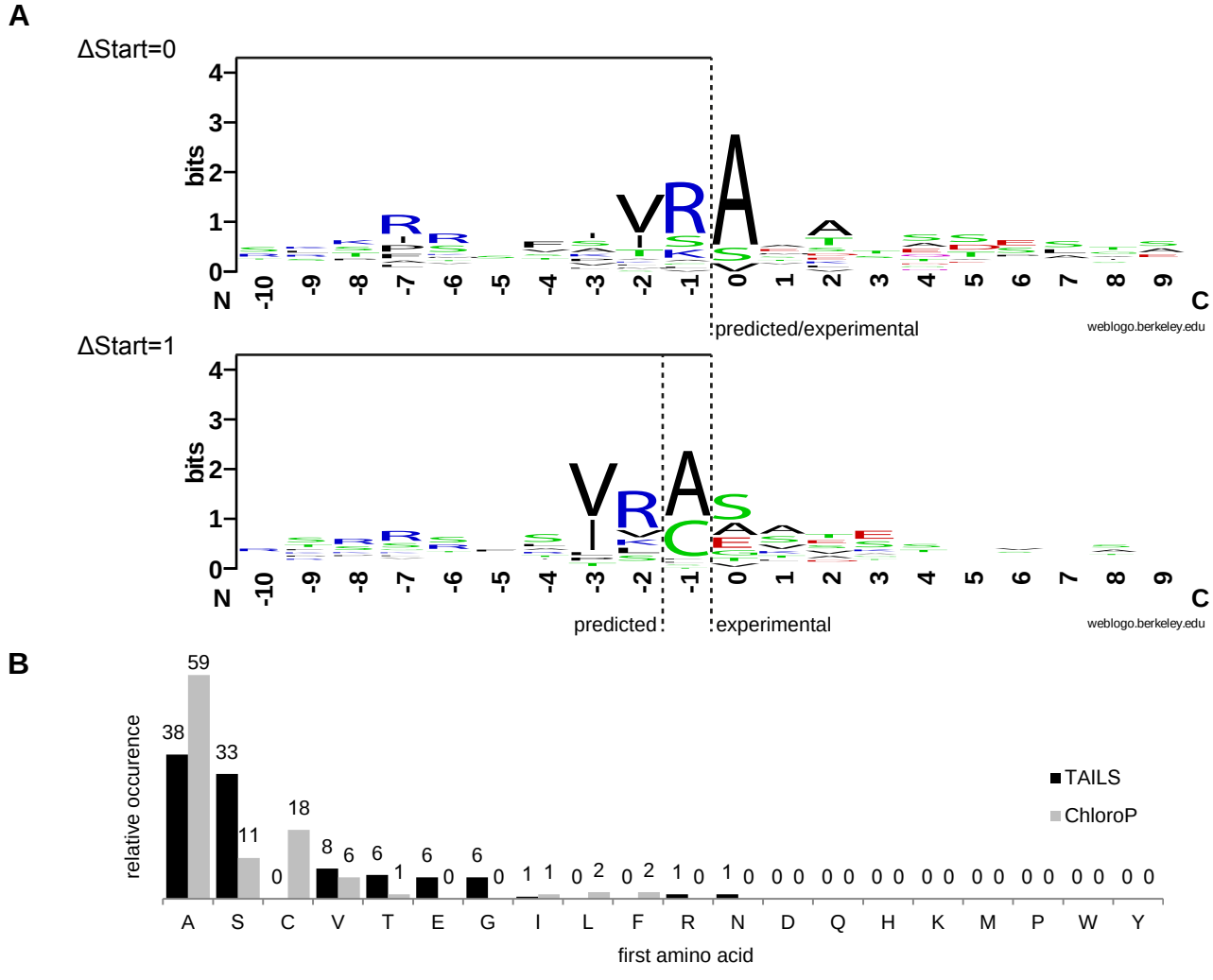
Supplemental Figure 1 Genotyping and gene expression analysis of the *tic56-3* mutant (A) Border sequence of the *tic56-3* T-DNA insertion toward the 3' end of the *TIC56* gene. (B) Genotype analysis by PCR. The schematic representation of the *TIC56* gene structure shows the position of the *tic56-3* T-DNA insertion and the binding sites of primers used for PCR (below). Black boxes, exons; lines, introns; grey boxes, untranslated regions. (C) RT-PCR analysis of *TIC56* expression in wild-type and *tic56-3*. The schematic representation of the *TIC56* transcript shows the position of the *tic56-3* T-DNA insertion and the binding sites of primers used for RT-PCR. The data presented confirms previous data on the location of the T-DNA insertion in *tic56-3* (Kikuchi et al., 2013) and shows the occurrence of a *TIC56* transcript that doesn´t encode the full-length protein. We tried left- and right-border primer as well as a TAIL-PCR approach with *TIC56*-specific primer to obtain the sequence of the other T-DNA junction but failed. This could be due to a T-DNA truncation or rearrangement at the right border frequently occurring in T-DNA insertion mutants.

Supplemental Figure 2



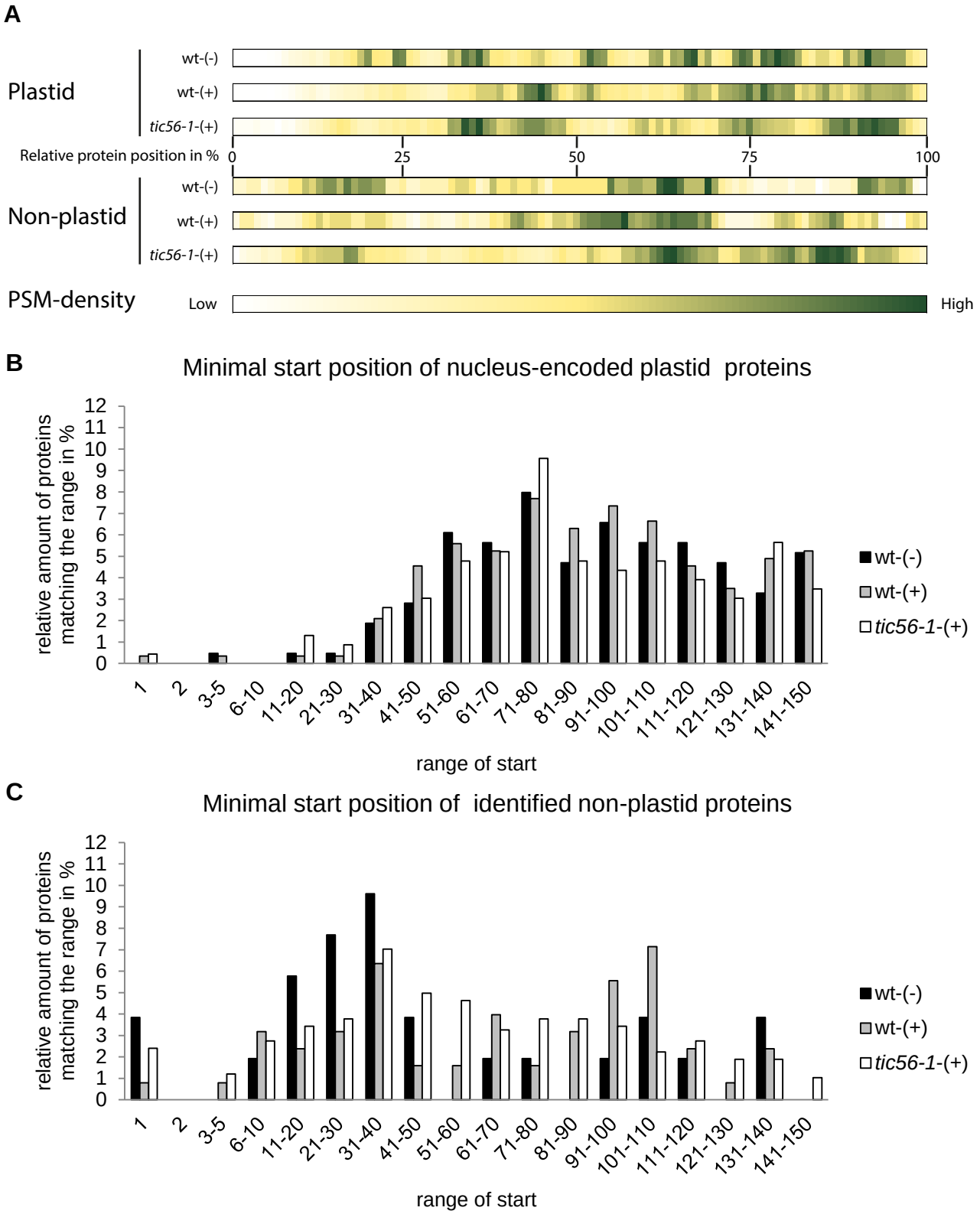
Supplemental Figure 2 Phenotypes of the mutants *ppi2* and *tic56-1* compared to wild-type. Observation of *ppi2*, *tic56-1* and wild-type (wt) seedlings grown for 2.5, 5 and 8 weeks under short day conditions on MS agar supplemented with 0.8% sucrose. (Scale bars: 2 mm).

Supplemental Figure 3



Supplemental Figure 3 Comparison of experimental and theoretical processing sites and resulting N-terminal amino acids. **(A)** Sequence logos were created using the sequences of all proteins of all three plant lines without duplicates, whose start position matched with the prediction (upper panel) or was shifted by one amino acid (lower panel). For these, ten amino acids upstream and downstream of the experimentally observed starting position were included. **(B)** All nucleus-encoded plastid proteins determined as correct processed mature proteins of the whole TAILS experiment were combined and the frequency of occurrence of a certain amino acid at the N-terminus was determined. The experimental frequency is shown in black, and compared to the theoretical frequency with the same set of proteins, shown in grey.

Supplemental Figure 4



Supplemental Figure 4 Plastid proteins in/of *tic56-1* are protected during thermolysin treatment. A crude plastid pellet of wild-type and *tic56-1* was treated (+) or not (-) with thermolysin followed by proteome analysis. (A) The vertical bars/columns represent the protein length of nucleus-encoded plastid proteins (top) or non-plastid proteins (bottom) identified by mass spectrometry. Here, the protein length from N- to C-terminus is displayed in percent. The PSM density illustrates the distribution and amount of matched peptides within the proteins. The very low PSM density in the N-terminal part of plastid proteins can be interpreted as a sign of transit peptide removal. The high PSM density distributed over the other parts of plastid proteins, even in the thermolysin treated samples, hints to protection of the proteins against proteolysis. In **(B)** and **(C)** the distribution of the minimal starting positions of the proteins is shown. In all three samples the bulk of plastid proteins appear to lack their N-terminal targeting sequences **(B)**. In contrast to the (contaminant) non-plastid proteins the distribution of the starting ranges is unchanged for plastid proteins by thermolysin treatment, indicating protection of plastid proteins against proteolysis as well as efficacy of the enzyme treatment.

Supplemental Methods

Genotyping and gene expression analysis of the *tic56-3* mutant.

RNA or genomic DNA was extracted from seedlings or leaves frozen in liquid nitrogen according to the protocols by Onate-Sanchez and Vicente-Carbajosa (2008) and Edwards et al. (1991). For RNA extraction plants were grown on half-strength Murashige and Skoog (MS) agar supplemented with 0.8% (w/v) sucrose for 28 days under short-day conditions. For first strand cDNA synthesis 1 µg of RNA, the RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Scientific) and oligo(dT)18 primer was used. Sequences of primers used for genotyping and RT-PCR: 579LP 5' ACTGGAATCTGATCACATGCC 3', 56rev1 5' CTTCAGGTCCTCTTCTCTCAGC 3', 56rev2 5' CTATGGATCCCCATCTTTTT TGGAGTTGC 3', 579RP 5' TATCGCCACTTAACATTTCGG 3', 56for 5' GTATAAGATC TCAACAATGTCGTCGATGAACTTCAATCC 3' and Tag5 5' CTACAAATTGCCTTTTC TTATCGAC 3'.

Thermolysin treatment and proteome analysis of *tic56-1* plastids

Protoplasts from 5 weeks old wild-type and *tic56-1* plants grown on ½ MS-medium containing 3% (w/v) sucrose were isolated as described (Material and Methods: Transient expression of eGFP fusion proteins in Arabidopsis protoplasts). The breakage of protoplasts was done according to (Fitzpatrick and Keegstra, 2001) omitting BSA in the breakage buffer. Plastids were enriched by centrifugation at 2.000xg for 5 min. The crude plastid samples were washed with HS buffer (50 mM Hepes KOH pH 8.0, 330 mM sorbitol). For the protease protection assay 200 µg protein of each wild-type sample and 27 µg protein of *tic56-1* sample was used. The crude plastid samples were treated with 100 µg/ml thermolysin in a final volume of 200 µl HS buffer for 30 min at 4°C according to (Froehlich, 2011). As a control a wild-type sample was treated in parallel without protease added. After quenching plastids were collected by centrifugation at 3.000xg for 5min and washed twice with HS buffer. Proteins were extracted and acetone-methanol-precipitated (Doucet et al., 2011). The protein-pellets were subjected to an in-solution digest with trypsin using RapiGest (Waters) according to the manufacturer's instructions. Peptide samples were measured on the LTQ Orbitrap Velos (Thermo Scientific) as described (Material and Methods: Terminal amine isotopic labeling of substrates) but with a modified chromatography method (0 150 min 5 40% B, 150 155 min 40 80 % B, 155 160 min 80 % B, A= water with 0.1 % formic acid, B=acetonitrile with 0.1 % formic acid). The RAW-files were analysed using the Proteome Discoverer 1.2 (Thermo Scientific), the search engine SEQUEST and the TAIR10 database. As variable

modifications N-terminal acetylation and as fixed modifications carbamidomethylation of cysteine were allowed. The precursor mass tolerance was set to 7 ppm and the fragment mass tolerance to 0.8 Da. For further analysis only peptides with a maximal FDR of 5% were used. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (Vizcaino et al., 2013) with the dataset identifier PXD001207.

Literature cited (Supplement)

Doucet A, Kleifeld O, Kizhakkedathu JN, Overall CM (2011) Identification of proteolytic products and natural protein N-termini by Terminal Amine Isotopic Labeling of Substrates (TAILS). Methods Mol Biol **753**: 273-287

Edwards K, Johnstone C, Thompson C (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res **19**: 1349

Fitzpatrick LM, 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

Froehlich J (2011) Studying Arabidopsis envelope protein localization and topology using thermolysin and trypsin proteases. Methods Mol Biol **774**: 351-367

Kikuchi S, Bedard J, Hirano M, Hirabayashi Y, Oishi M, Imai M, Takase M, Ide T, Nakai M (2013) Uncovering the protein translocon at the chloroplast inner envelope membrane. Science **339**: 571-574

Onate-Sanchez L, Vicente-Carbajosa J (2008) DNA-free RNA isolation protocols for Arabidopsis thaliana, including seeds and siliques. BMC Res Notes **1**: 93

Vizcaino JA, Cote RG, Csordas A, Dianes JA, Fabregat A, Foster JM, Griss J, Alpi E, Birim M, Contell J, O'Kelly G, Schoenegger A, Ovelleiro D, Perez-Riverol Y, Reisinger F, Rios D, Wang R, Hermjakob H (2013) The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res **41**: D1063-1069