Supporting Information

Boex-Fontvieille et al. 10.1073/pnas.1507714112

SI Materials and Methods

Production of Transgenic Lines. Gateway technology was used for construction of transgenes used for the production of transgenic lines according to the manufacturer's instructions (Invitrogen). For the generation of transgenic *35S::AtWSCP* lines, a cDNA for the coding frame was amplified with the following primers: 5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGA-AGAA-TCCTTCAGTGATCTCTTTT-3′ and 5′-GGGGACC-ACTTTGTACAAGAAAGCTGGG-TCTCAACCCGGGAAG-TATAAGTTGCT-3′. The PCR product was introduced into pDONR221 and, after sequencing, introduced into pB7FWG2 (Plant System Biology, VIB-Ghent University). For construction of transgenic lines expressing *35S::AtWSCP-(His)$_6$*, a similar strategy was used, although the final vectors were modified to contain the coding region for the His tag. To produce *AtWSCP* promoter::*GUS* lines, the *AtWSCP* promoter sequence was cloned into pDONR221 with the primers 5′-GGGGACAAGTTTGTAC-AAA-AAAGCAGGCTTATGACAAATTACAAAAATGG-3′ and 5′-GGGGACCACTTTGTACA-AGAAAGCTG-GGTTGATTG-TTATGTGTGTTTTAGGG-3′, sequenced and then cloned into the pKGWFS7 destination vector (Plant System Biology, VIB-Ghent University). To obtain transgenic *35S::AtWSCP::GFP* lines, cDNA encoding the AtWSCP precursor protein without stop codon was introduced into pDONR221 and subsequently into pK7FWG2 with Gateway technology (1). Transgenic *35S::SP::GFP* lines were created similarly, using cDNAs encoding the signal peptide of AtWSCP and GFP. Stable transformation of *Arabidopsis* (Col-0 ecotype) was performed with *Agrobacterium tumefaciens* C58 pGV3121, using the floral dip technique described by Clough and Bent (2).

Protein Expression, Purification, and Antibody Production. For production of AtWSCP precursor protein, cDNA was amplified by PCR (2) with the primer pair 5′-GGGGACAAGTTTGTACAA-AAAAGCAGGCTTCAAGAATCCTTCA-GTGATCTCTTTT-3′ and 5′-GGGGACCACTTTGTACAAGAAAGCTGGGTCTC-AACC-CGGGAAGTATAAGTTGCT-3′. PCR products were cloned into pDONR221, sequenced, and introduced into the vector pDEST17 by using Gateway technology. The construct obtained was transformed into *Escherichia coli* strain BL21 AI. Expression of a NH_2-terminally His-tagged precursor was induced by adding arabinose (0.2% final concentration). After growth at 37 °C for 3 h, pellets obtained from 0.5 L of bacterial cultures were lysed in a buffer containing 20 mM NaH_2PO_4, 6 M urea, 20 mM imidazole, 500 mM NaCl, 0.5 mM PMSF, pH 7.5 and passed through a French Press (Thermo Electron; FA-078A). After centrifugation, the clear lysate was subjected to Ni-NTA agarose (Qiagen) and the purification was carried out according to the manufacturer's instructions. Two milligrams of pure protein were used to raise polyclonal antibodies in rabbits (Interchim).

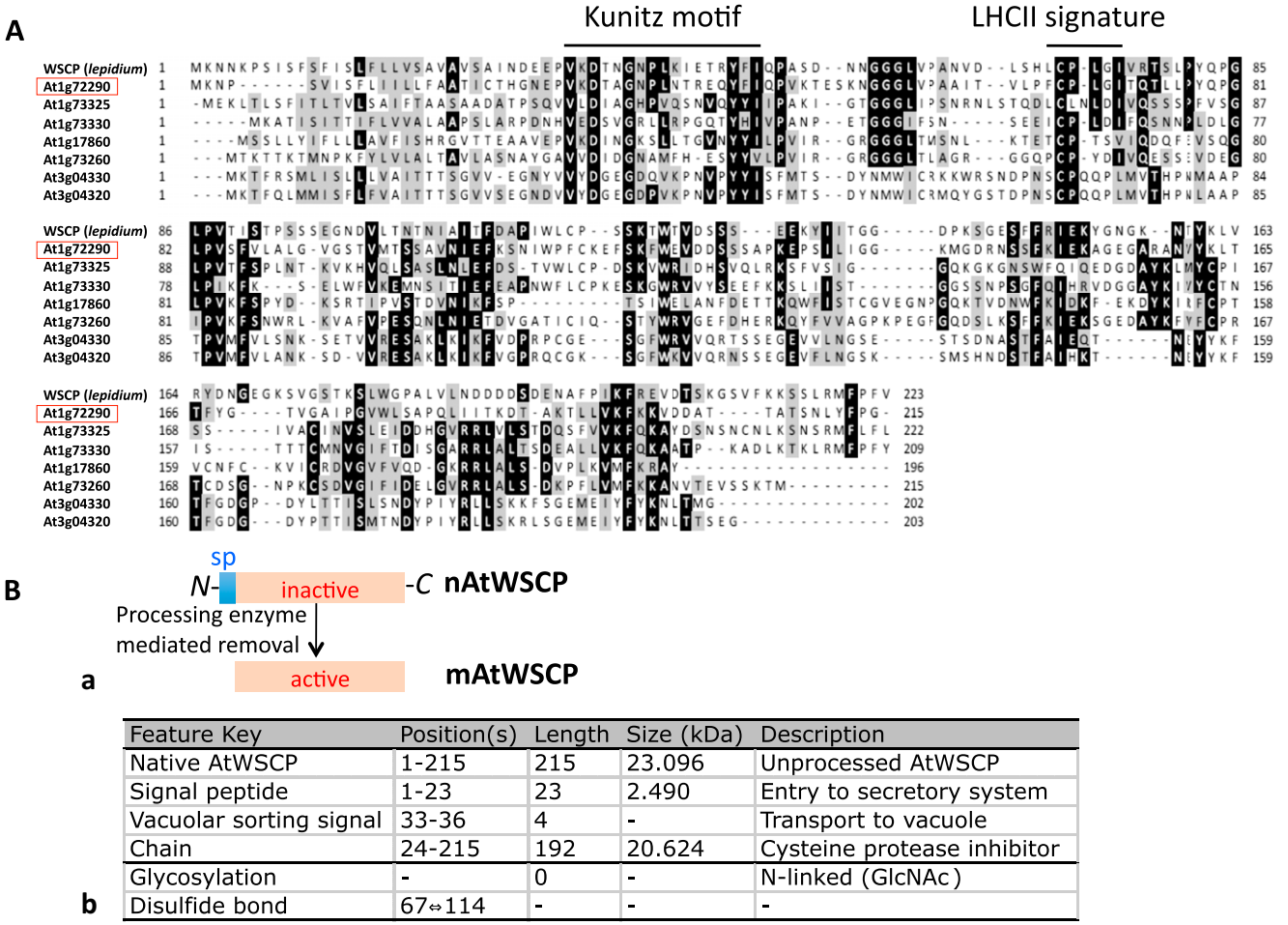
First-Strand cDNA Synthesis and Semiquantitative PCR. Total RNA was isolated from etiolated seedlings, and 2 μg were incubated in the presence of 300 pmol oligo-dT primers at 65 °C for 5 min. Subsequently, synthesis of the first cDNA strand was carried out with reverse transcriptase (SUPERSCRIPT II; Invitrogen) at 42 °C for 2 h under the conditions recommended by the supplier. The reaction was stopped by incubation at 72 °C for 10 min, and a 1-μL aliquot (1/30 of the total reaction volume) was subjected to semiquantitative PCR that had been optimized with respect to the number of cycles necessary to be in the exponential phase of amplification for each type of cDNA sequence. The primer pairs used are as follows: ACTfor 5′-CAGAATCAGATATCTAAA-AATCC-CGGAAA-3′; ACTrev 5′-TGGGATGACATGGAG-AAGAT-3′; AtWSCPfor 5′-TGGCATG-AGGAAAAAGCC-AAG-3′; AtWSCPrev 5′-TCAATGTTTTCTCAAGCTCAA-3′; LHCB2for 5′-CGGACCAGACCGTCCCAA-3′; LHCB2rev 5′-ATGCTTTGCGCGTGGATC-3′.

GUS Staining. Seedlings were infiltrated in staining solution (100 mM phosphate buffer, 10 mM EDTA, 1 mM X-Gluc, pH 7.0) for 10 min. Then, staining was pursued by incubating the plant samples at 37 °C overnight. Thereafter, destaining was done by successively incubating the tissues in 50%, 75%, and 96% ethanol. Finally, the tissues were transferred into a solution of ethanol and glycerol (50 and 30% each) and either kept at 4 °C or analyzed immediately with an Eclipse E-600 (Nikon) microscope or a SZX12 (Olympus) binocular. Photos were taken with an Olympus DP70 camera.

Molecular Modeling. The 3D modeling of AtWSCP (At1g72290) and RD21 (At1g47128) was performed, respectively, by using SWISS-MODEL and I-TASSER, online protein structure and function prediction tools (3, 4), and predictions on AtWSCP–RD21 interaction were made by using ClusPro (5).

Confocal Laser Scanning Microscopy. Confocal laser scanning microscopy was carried out by using a Leica TCS SP5 microscope and documented with Leica LAS software. Collection of GFP signals was performed after excitation with an argon laser (488 nm) in combination with a 510- to 525-nm emission filter. Pigment fluorescence was analyzed simultaneously (excitation with an argon laser at 488 nm; emission filter 650–750 nm). Counterstaining of cell walls was done with Calcofluor white (FLUKA), according to the manufacturer's instructions.

1. Karimi M, Inzé D, Depicker A (2002) GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci* 7(5):193–195.
2. Clough SJ, Bent AF (1998) Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16(6):735–743.
3. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: A unified platform for automated protein structure and function prediction. *Nat Protoc* 5(4):725–738.
4. Biasini M, et al. (2014) SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* 42(Web Server issue): W252–W258.
5. Comeau SR, Gatchell DW, Vajda S, Camacho CJ (2004) ClusPro: A fully automated algorithm for protein-protein docking. *Nucleic Acids Res* 32(Web Server issue): W96–W99.



Feature Key	Position(s)	Length	Size (kDa)	Description
Native AtWSCP	1-215	215	23.096	Unprocessed AtWSCP
Signal peptide	1-23	23	2.490	Entry to secretory system
Vacuolar sorting signal	33-36	4	-	Transport to vacuole
Chain	24-215	192	20.624	Cysteine protease inhibitor
Glycosylation	-	0	-	N-linked (GlcNAc)
Disulfide bond	67⇔114	-	-	-

b

Fig. S1. (*A*) The protein encoded by At1g72290, AtWSCP, belongs to a small family of Kunitz-type protease inhibitors in *A. thaliana*. Multiple sequence alignment of proteins was done with BioEdit program by using the ClustalW option. Dark boxes represent identical amino acids, and gray boxes represent conserved amino acids. The homologous *L. virginicum* WSCP (LvWSCP) protein was included in the sequence comparison. The positions of the Kunitz-type protease inhibitor motif, which is perfect for LvWSCP [LIVM]-x-D-x-[EDNTY]-[DG]-[RKHDENQ]-x-[LIVM] and LHCII-signature, which shows homology to the sequence motif [F/Y]DPLGL of the major light-harvesting chlorophyll *a*/b-binding protein of photosystem II and its related proteins, are delineated by horizontal lines. (*B*) Presence of a putative signal peptide and other localization and protein modification sites in AtWSCP. mAtWSCP, mature AtWSCP; nAtWSCP, native AtWSCP.

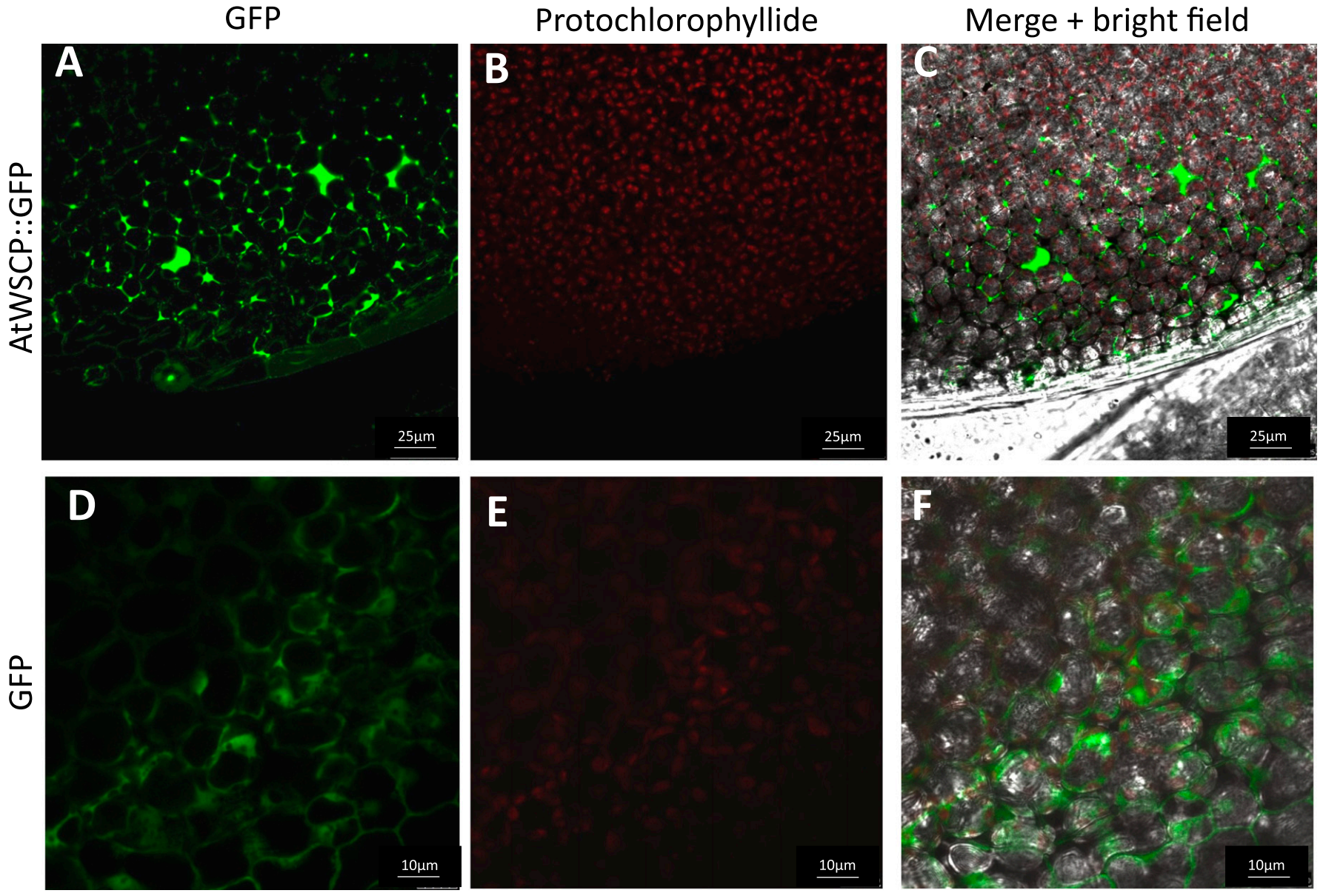


Fig. S2. Transgenic lines expressing the fusion protein consisting of AtWSCP plus GFP (AtWSCP::GFP; *A*–*C*) or GFP (*D* and *E*) were analyzed by confocal laser scanning microscopy. Green fluorescence signals (shown in *A* and *D*) and red fluorescence signals of protochlorophyllide (seen in *B* and *E*) in the cotyledons of etiolated seedlings. *C* and *F* show merges of both fluorescence signals viewed under bright field.

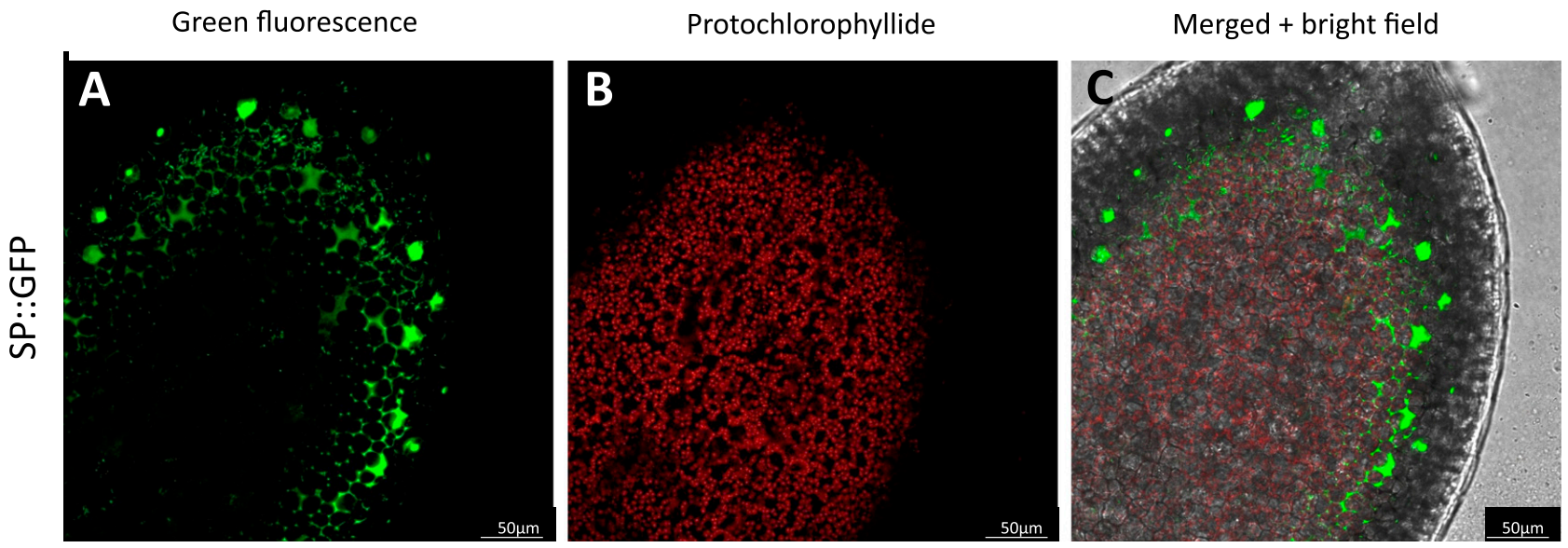


Fig. S3. Localization of a fusion protein consisting of the signal peptide of AtWSCP and GFP (SP::GFP) in the cotyledons of etiolated seedlings. Transgenic lines were generated expressing the SP::GFP coding sequence under the control of *35S* promoter. Fluorescence signals of GFP (*A*) and protochlorophyllide (*B*) were collected by confocal laser scanning microscopy. *C* shows the superposition of both fluorescence images under bright field.

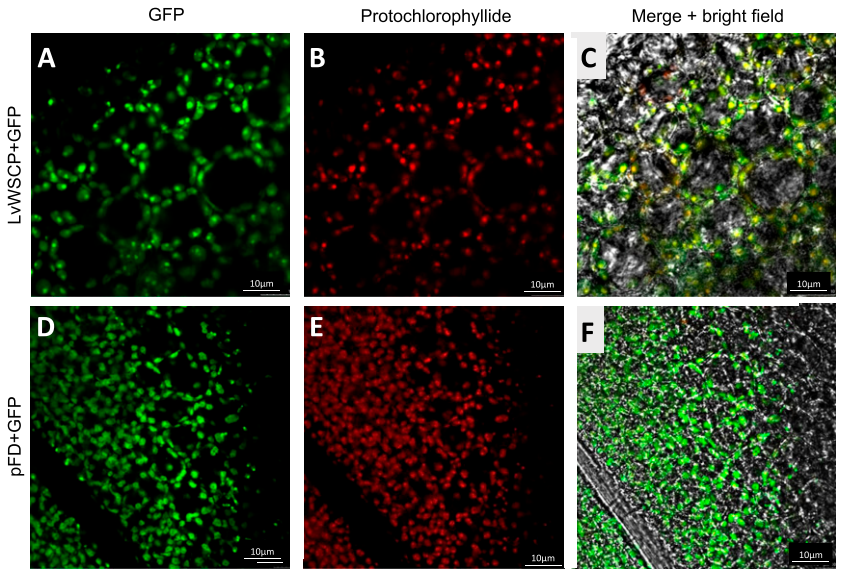


Fig. S4. LvWSCP::GFP localization in plastids of etiolated seedlings obtained from seeds of the T_3 generation of transgenic lines expressing WSCP from *L. virginicum* fused to GFP at its COOH terminus under the control of *35S* promotor (*A*–*C*). Etiolated transgenic seedlings expressing a fusion protein consisting of the ferredoxin precursor and GFP were analyzed in parallel (*D*–*F*). Fluorescence signals of GFP (*A* and *D*) and protochlorophyllide (*B* and *E*) were collected simultaneously by confocal laser scanning microscopy. Overlap of green GFP and red chlorophyll fluorescence (*C* and *F*) shows clearly that both fusion proteins are imported into etioplasts.

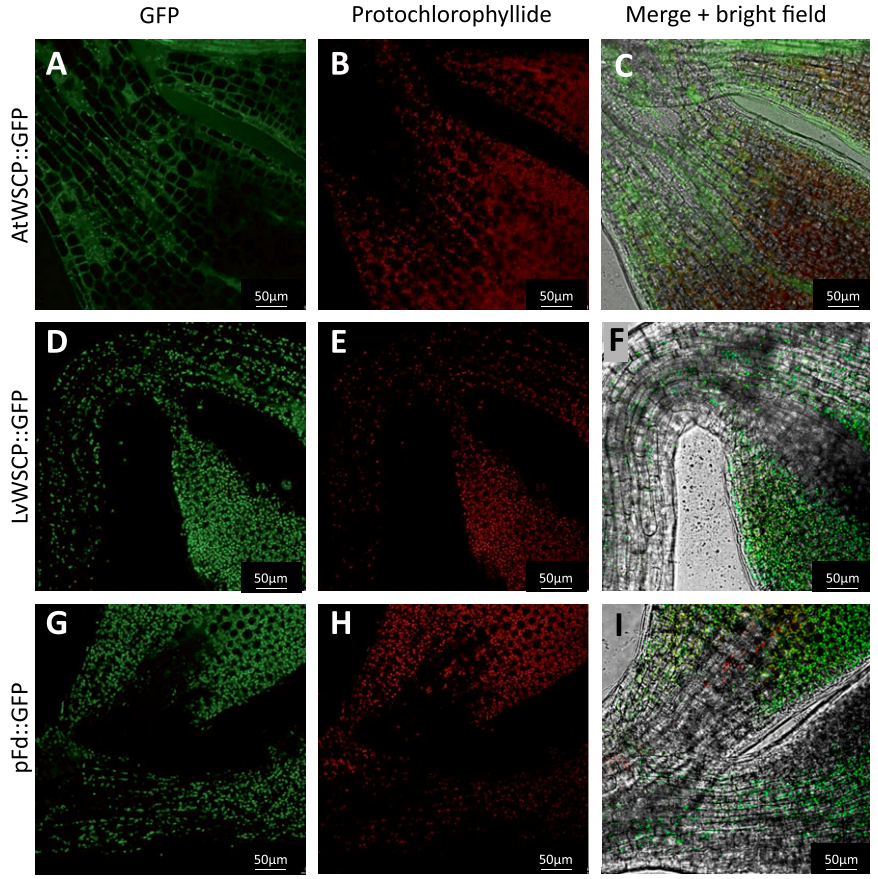
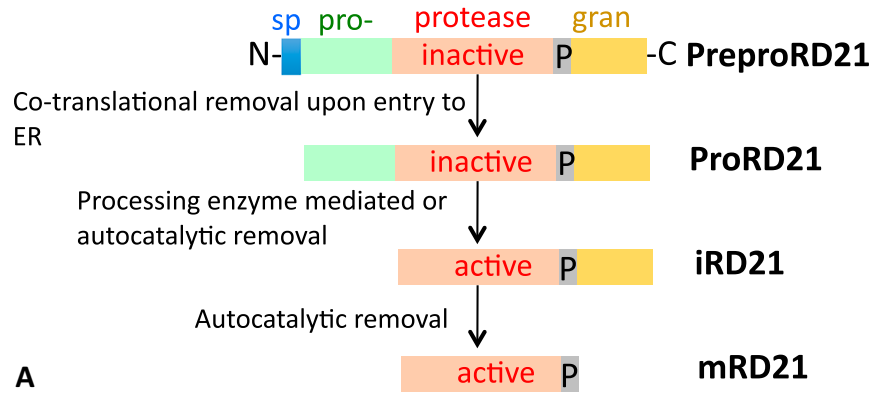


Fig. S5. Localization of AtWSCP::GFP (*A*–*C*), LvWSCP::GFP (*D*–*F*), and pFd::GFP (*G*–*I*) in the apical hook of etiolated seedlings. Transgenic lines expressing AtWSCP::GFP, LvWSCP::GFP, and pFd::GFP under the control of *35S* promotor were generated and etiolated seedlings of the T_3 generation analyzed by confocal laser scanning microscopy. The images show fluorescence signals of GFP (*A*, *D*, and *G*) and protochlorophyllide (*B*, *E*, and *H*) and respective merges viewed under bright field (*C*, *F*, and *I*). Note the obviously overlapping GFP and pigment fluorescence distributions that are indicative for a plastid localization of LvWSCP::GFP and pFd::GFP and the different localization of AtWSCP::GFP in cell walls/extracellular spaces.



A

Feature Key	Position(s)	Length	Description
Signal peptide	1-21	21	Entry to secretory system
Propeptide	22-136	115	Self-inactivation
Chain	137-356	220	Cysteine endoprotease
Propeptide	357-462	106	Unknown function
Glycosylation	90-90	1	N-linked (GlcNAc)
Disulfide bond	158⇔200	-	-
Disulfide bond	192⇔233	-	-
Disulfide bond	291⇔342	-	-
Disulfide bond	375⇔387	-	-
Disulfide bond	381⇔402	-	-
Glycosylation	414-414	1	N-linked (GlcNAc)

B

Fig. S6. Maturation steps of preproRD21 and its structural features. (*A*) RD21 is encoded as a preproprotein with a NH_2-terminal signal peptide, and two propeptides. The signal peptide is required for the entry into the secretory system, and the NH_2-terminal propeptide is required for inactivation of enzyme under undesired conditions. The actual function of the COOH-terminal propeptide is yet unknown. gran, COOH-terminal granulin domain; iRD21, intermediate RD21; mRD21, mature RD21; P, proline-rich domain; PreproRD21, precursor of RD21; Pro-, NH_2-terminal prodomain; proRD21, proprotein precursor of RD21; Protease, protease domain; Sp, Signal peptide. (*B*) List of structural features and putative/validated functions of different domains with the list of predicted secondary modifications and disulfide bridges.

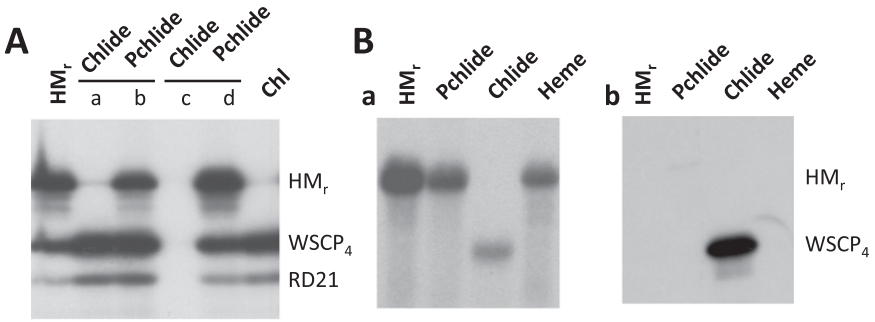
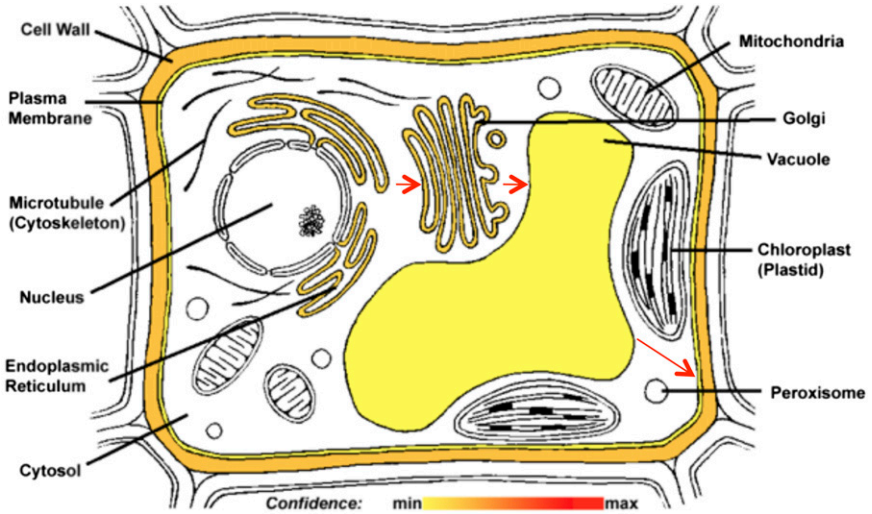
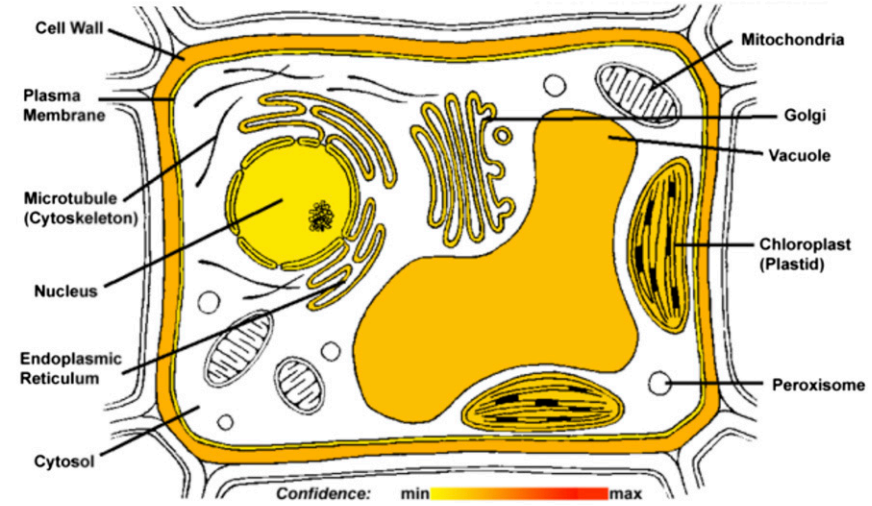


Fig. S7. In vitro reconstitution of AtWSCP-$(His)_6$ into larger complexes with RD21. (*A*) Presence of higher molecular mass AtWSCP::RD21 complexes (HM_r) and respective subunits and dissociation of the reconstituted HM_r complexes by added Chlide and Chl. Lanes a and b show incubations carried out for 30 min, whereas lanes c and d show incubations carried out for 24 h after the addition of pigment. Note the autolytic degradation of the released subunits by virtue of RD21's protease activity in the Chlide-containing assays and the lack of a comparable effect by added Pchlide. (*B*) Chlide-dependent dissociation of reconstituted AtWSCP::RD21 complexes by Chlide, but neither Pchlide or heme used for comparison. *a* shows a Western blot that was probed with the WSCP antibody. *b* shows a replicate analysis in which Chlide-containing protein complexes were traced by their red light-induced pigment autofluorescence.



A AtWSCP



B RD21

Fig. S8. Subcellular localization of AtWSCP (*A*) and RD21 (*B*) using Cell eFP Browser. A putative translocation route of AtWSCP was marked by arrows.

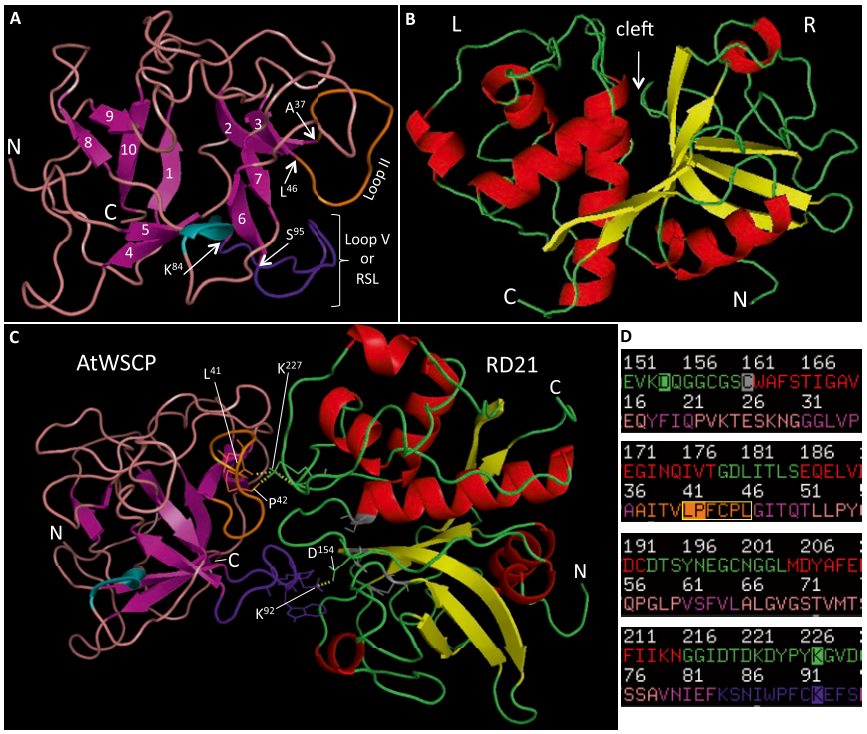


Fig. S9. Amino acid sequences and structures of AtWSCP, RD21, and the RD21–AtWSCP complex; these structures were predicted by using SWISS-MODEL, I-TASSER, and ClusPro, respectively. (*A*) Ribbon diagram displaying structure of AtWSCP. The reactive site loop (RSL) and the second loop are marked by arrows. These two loops facilitate the interaction between RD21 and AtWSCP, where the RSL intrudes into the active site of RD21 and residues in the second loop as well as RSL form hydrogen bonds with the amino acids in RD21 and stabilize their interaction. (*B*) Ribbon diagram showing the structure of the mature RD21 (mRD21; consult Fig. S6). Left and right lobs are designated as L and R, and active site cleft is marked by an arrow. (*C*) Ribbon diagram of the RD21–AtWSCP complex. Selected amino acid residues that form hydrogen bonds (displayed by dashes yellow line) in the partnering protein are shown by lines. In *A*–*C*, the β-strands of AtWSCP are shown as magenta, and RD21 as yellow arrows. The α-turn of AtWSCP is shown as cyan, whereas the α-helices of RD21 are shown as red ribbons. The loops connecting β-sheet and α-helices of RD21 are shown as green, and AtWSCP as deep salmon, except for the second loop between second and third β-strands, which is depicted by orange and the RSL between fifth and sixth β-strands, which is shown in purple. (*D*) Partial amino acid sequences of RD21 and AtWSCP are shown to display the interacting amino acids (highlighted) and different structural elements (shown in different colors). In RD21 sequence, the α-helices are shown in red, β-sheet in yellow, connecting loops in green, and interacting amino acids and catalytic Cys161 are highlighted. Similarly, in AtWSCP sequence, the β-sheets are shown in magenta, the connecting loops in deep salmon except for the second loop and the RSL that are, respectively, shown in orange and purple, the interacting amino acids are highlighted, and the LHCII signature sequence is marked by yellow rectangle. Amino acid locations in the respective proteins were marked at the top of each sequence.