



Defining the core proteome of the chloroplast envelope membranes

Stefan Simm¹, Dimitrios G. Papisotiriou^{2†}, Mohamed Ibrahim^{1†}, Matthias S. Leisegang¹, Bernd Müller^{3†}, Tobias Schorge², Michael Karas^{2,4,5}, Oliver Mirus¹, Maik S. Sommer¹ and Enrico Schleiff^{1,4,5*}

¹ Institute of Molecular Cell Biology of Plants, Goethe University, Frankfurt, Germany

² Institute of Pharmaceutical Chemistry, Goethe University, Frankfurt, Germany

³ Department of Biology I, Ludwig-Maximilians-University, Munich, Germany

⁴ Center of Membrane Proteomics, Goethe University, Frankfurt, Germany

⁵ Cluster of Excellence 'Macromolecular Complexes', Goethe University, Frankfurt, Germany

Edited by:

Harvey Millar, The University of Western Australia, Australia

Reviewed by:

Lutz Andreas Eichacker, University of Stavanger, Norway

Sascha Rexroth, Ruhr-University Bochum, Germany

Markus Teige, University of Vienna, Austria

*Correspondence:

Enrico Schleiff, Center of Membrane Proteomics, Cluster of Excellence 'Macromolecular Complexes';

Institute of Molecular Cell Biology of Plants, Goethe University, Max-von-Laue Strasse 9, Frankfurt 60438, Germany.

e-mail: schleiff@bio.uni-frankfurt.de

†Current address:

Dimitrios G. Papisotiriou, Jealott's Hill International Research Centre,

Bracknell, Berkshire, UK;

Mohamed Ibrahim, Botany

Department, Faculty of Science, Ain

Shams University, Cairo, Egypt;

Bernd Müller, AB SCIEX Germany

GmbH, Darmstadt, Germany.

High-throughput protein localization studies require multiple strategies. Mass spectrometric analysis of defined cellular fractions is one of the complementary approaches to a diverse array of cell biological methods. In recent years, the protein content of different cellular (sub-)compartments was approached. Despite of all the efforts made, the analysis of membrane fractions remains difficult, in that the dissection of the proteomes of the envelope membranes of chloroplasts or mitochondria is often not reliable because sample purity is not always warranted. Moreover, proteomic studies are often restricted to single (model) species, and therefore limited in respect to differential individual evolution. In this study we analyzed the chloroplast envelope proteomes of different plant species, namely, the individual proteomes of inner and outer envelope (OE) membrane of *Pisum sativum* and the mixed envelope proteomes of *Arabidopsis thaliana* and *Medicago sativa*. The analysis of all three species yielded 341 identified proteins in total, 247 of them being unique. 39 proteins were genuine envelope proteins found in at least two species. Based on this and previous envelope studies we defined the core envelope proteome of chloroplasts. Comparing the general overlap of the available six independent studies (including ours) revealed only a number of 27 envelope proteins. Depending on the stringency of applied selection criteria we found 231 envelope proteins, while less stringent criteria increases this number to 649 putative envelope proteins. Based on the latter we provide a map of the outer and inner envelope core proteome, which includes many yet uncharacterized proteins predicted to be involved in transport, signaling, and response. Furthermore, a foundation for the functional characterization of yet unidentified functions of the inner and OE for further analyses is provided.

Keywords: membrane proteome, plant proteomics, chloroplast membrane proteins, mass spectrometry, envelope membrane proteome approach comparison

INTRODUCTION

The characterization of a single protein function is associated with an enumeration of different features. Some of these features are the subcellular localization of the protein, its interaction with other proteins, co- or post-translational modifications as well as its (enzymatic) activity. With the growing number of sequenced genomes, the "proteome," as sum of all proteins in an entire cell or cellular (sub-)compartment, becomes important for the understanding of cellular function (Wilkins et al., 1996; James, 1997). The mass spectrometric analysis of complete cellular proteomes still remains difficult, especially in the highly compartmentalized eukaryotic cells. Furthermore, proteomes are dynamic and change in response to different stimuli. They include different splice forms and post-translationally modified proteins in different abundances. Thus, different technical approaches have been developed to accommodate the complexity of a proteome (e.g., Karas and Hillenkamp, 1988; Aebersold and Mann, 2003),

especially to study the subcellular localization of membrane proteins as a complementary approach to the complete cell proteome analyses (van Wijk, 2000; Millar et al., 2009).

This complexity of eukaryotic cells leads us to focus on the proteome of chloroplast, which are organelles essential for different metabolic pathways like photosynthesis, and biosynthesis of fatty acids or amino acids. These organelles contain several thousand different proteins and the majority of which is cytosolically synthesized and has to be translocated across the envelope membranes (Leister, 2003; Schleiff and Becker, 2011). Thereby, the proteome of the organelle as such (Zabrouskov et al., 2003; Kleffmann et al., 2004) or of subfractions like the thylakoid lumen (Peltier et al., 2002), the thylakoid membranes (Eichacker et al., 2004; Friso et al., 2004), the stroma (Goulas et al., 2006; Peltier et al., 2006), plastoglobules (Ytterberg et al., 2006), or the envelope membranes (Schleiff et al., 2003b; Bräutigam and Weber, 2009) have been analyzed in the past. The current knowledge on the

proteomic content of chloroplasts has been deposited in several independent databases like PLPROT (Kleffmann et al., 2006) or AT_CHLORO (Ferro et al., 2010). However, especially the analysis of the envelope and more specifically the inner envelope (IE) and outer envelope (OE) membrane proteome is still a challenging task due to the hydrophobicity of membrane proteins (Eichacker et al., 2004). More specifically, the dissection of the IE and OE membrane proteome is still very poorly supported by direct proteomic studies (Ferro et al., 2003; Schleiff et al., 2003b).

The determination of a protein's localization is a very important tool for experimental guidance. In here, we aimed at the determination of a reliable proteome of the OE and IE membranes of chloroplasts. To this end, we comparatively analyzed the overall envelope proteomes of the model species *Arabidopsis thaliana* and *Medicago sativa*. To substantiate our findings, we individually analyzed IE and OE membranes of *Pisum sativum*, the only plant to date, for which the separation of both can be achieved (Ferro et al., 2003; Schleiff et al., 2003b). We chose the genetic model *A. thaliana* by its comprehensive genome and transcriptome data available (see, e.g., The Arabidopsis Information Resource, TAIR10; Lamesch et al., 2012). The legume *P. sativum* was chosen, as it is the model plant for biochemical analyses of chloroplast function (see, e.g., Franssen et al., 2012). Due to the paucity of data, the recently sequenced and closely related *M. sativa* was used to substantiate our findings for *P. sativum*.

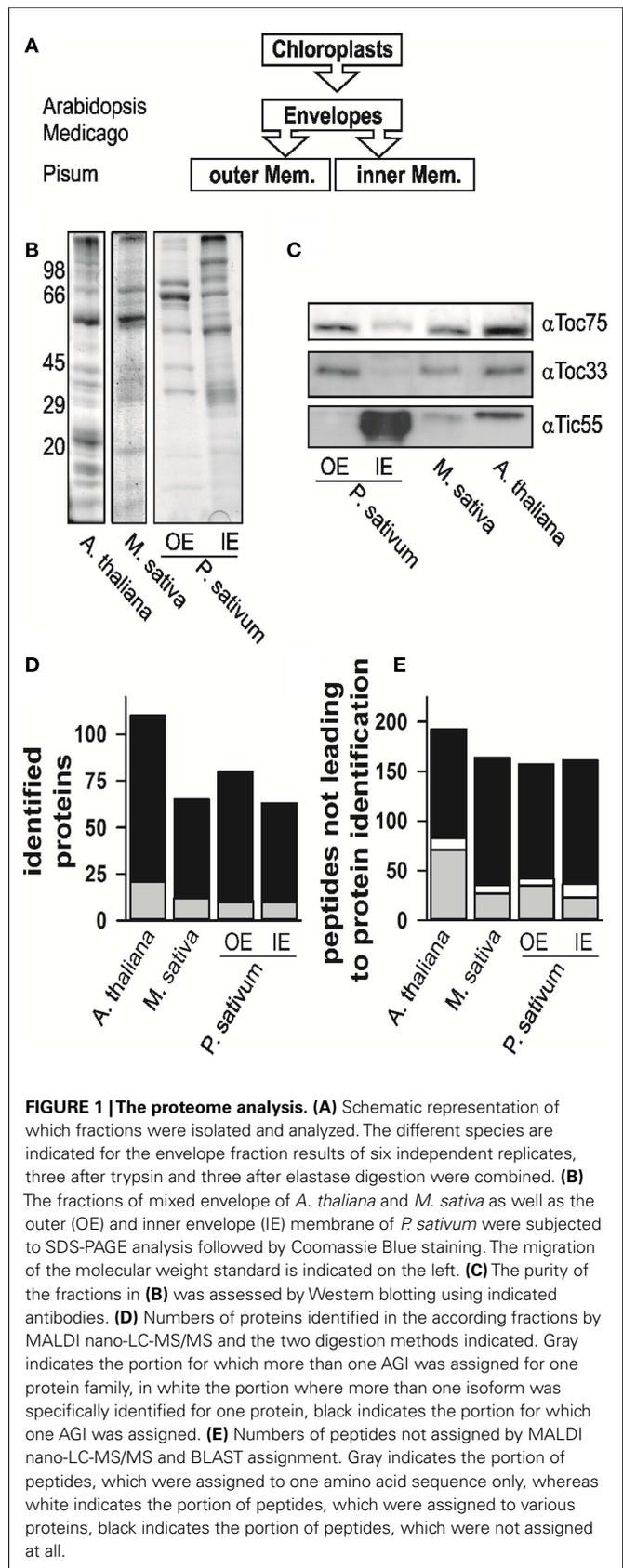
The identified proteins in these plant species were compared to each other and to the publicly available datasets of previous studies. We identified a total of 247 different proteins, of which – based on comparisons with other studies – 191 were assigned as putative envelope proteins. To our surprise, only 27 of these were found in all studies. Based on intersection and cross-contamination analysis of available previous studies, we were able to reliably assign 50/49 proteins as outer/inner membrane-localized, while at least 37 additional proteins in the mixed envelope fractions can be assigned as envelope proteins as well, but not reliably to a specific membrane.

RESULTS AND DISCUSSION

CHLOROPLAST PROTEOME ANALYSES

We analyzed the chloroplast proteomes with focus on the envelope membranes from three model plant species, namely *A. thaliana*, *P. sativum*, and *M. sativa*. We chose *A. thaliana* because of the availability of a comprehensive genome and many existing transcriptome data (e.g., The Arabidopsis Information Resource, TAIR10; Lamesch et al., 2012). Thus, the well annotated genome of *Arabidopsis* provides a solid base for the assignment of the identified inner and OE proteins. In turn, the legumes *P. sativum* and *M. sativa* are model plants for biochemical analyses of chloroplast function (e.g., Franssen et al., 2012), as well as crop plants. Using envelopes of different plant species allows the detection of proteins with different abundances. The varying achievable purity of the samples allows the detection of an additional different set of peptides.

We isolated and subfractionated chloroplasts to analyze the envelope proteomes (Figures 1A,B). The enrichment of the



obtained fractions was assessed by Western blotting using specific antibodies (Figure 1C). The analysis confirmed the enrichment of inner and outer membrane proteins in the mixed envelope fractions of *A. thaliana* and *M. sativa*, the mixed envelope fractions could not be further separated. In contrast, separation of envelope membranes in the IE and OE from *P. sativum* chloroplasts has been established previously (e.g., Schleiff et al., 2003a,b). Subsequently, the distinct fractions were analyzed by mass spectrometry.

The proteomes of all envelope membranes were analyzed by MALDI nLC-MS/MS (Table S14 in Supplementary Material) yielding in total 110 proteins in *A. thaliana* (Figure 1D, *Arabidopsis* EM, three independent isolations; Table S1 in Supplementary Material). In parallel, we identified 71 proteins in *M. sativa* (Figure 1D, *Medicago* EM, three independent isolations; Table S2 in Supplementary Material) and 124 different proteins in both membranes (87 IE; 73 OE) of *P. sativum* (Figure 1D, *Pisum* IE and *Pisum* OE, three independent isolations; Tables S3 and S4 in Supplementary Material).

Our peptide-based assignment relies on a stringent BLAST search, where an identity >95% and no mismatch or gap was allowed. Only a single amino acid substitution with a residue of similar properties or a single undefined amino acid position was accepted (for details see Experimental Section). The BLAST search was combined with a bidirectional best BLAST hit search to assign the homologous sequences in *A. thaliana* to the proteins identified in *P. sativum* or *M. sativa* to render the assignment from different species comparable. To confirm that the peptide-based assignment is consistent with the expected chloroplast localization, we analyzed the expression of the corresponding genes with respect to leaves and roots (e.g., Vojta et al., 2004). Indeed, almost all genes coding for the identified proteins including those identified by a single peptide only are highly expressed in leaf tissue (Figure A1 in Appendix). AT3G45360 is the only exception identified by more than one peptide with an expression value smaller than 10 in leaves. However, this gene is annotated as a transposable element. Furthermore, almost all genes are equally high or higher expressed in leaves in comparison to roots. The only gene with a significantly higher expression in roots than in leaves is AT3G09260 identified in *A. thaliana*. It encodes a β -glucosidase annotated as Pyk10, which was identified in ER-bodies (Matsushima et al., 2003). Although the protein most likely represents a contamination of the sample, its overall expression pattern supports the peptide-based protein assignment approach.

While analyzing the data, a large number of the obtained peptides did not lead to an identification of a protein (Figure 1E, Tables S5–S7 in Supplementary Material). About 15–30% of these peptides mapped uniquely to a single sequence (in gray), while few peptides mapped to multiple sequences (in white). The large portion of peptides which remained unassigned (in black) might have three different reasons: (i) The choice of too stringent search parameters, (ii) contaminations of the samples, or (iii) the existence of natural variances of sequences in form of unknown splice variants or nucleotide polymorphisms of genes leading to alternative amino acid sequences. The analysis of this phenomenon, however, goes beyond the scope of this work.

COMPARISON TO OTHER ENVELOPE MEMBRANE PROTEOMIC APPROACHES

To establish a core envelope proteome we unified results of our and previous studies (Ferro et al., 2003, 2010; Froehlich et al., 2003; Bräutigam et al., 2008; Bräutigam and Weber, 2009). For that, we first assigned the *Arabidopsis* Genome Initiative (AGI) number of the closest homolog of *A. thaliana* to each of the proteins found in *M. sativa* and *P. sativum*. Combining our four data sets, we obtained 247 different proteins in total. The globally unified protein pool contains a total of 911 different proteins. Ferro et al. (2010) assigned their identified proteins according to the suborganellar (stroma, thylakoid, and envelope) localization, which we have used to assess the quality of our data (cross-contaminations from thylakoid and stroma). We defined four different categories (Table 1): Category I are proteins that were found in at least two studies but not in the stroma and thylakoid according to Ferro et al. (2010). Category II unites proteins, which were found in at least three studies but also in the stroma or thylakoid. Category III are proteins found in one study only, but exclusively in the envelope, and category IV are proteins found in less than three studies, but also in the stroma or thylakoid. The selection of three independent studies for category II as criterion takes into account that two studies each come from Bräutigam et al. (2008), Bräutigam and Weber (2009), and Ferro et al. (2003, 2010). For better visualization of the impact of our study we have marked the identified proteins of the categories as identified in here (a) or in previous studies (b).

From our point of view the list of proteins of category I is most reliable, because there are no cross-contaminations via thylakoid and stroma and they are supported by previous studies. Proteins of categories II and III have to be confirmed experimentally first and proteins of category IV are considered to be not reliably assigned.

We noticed that only 30 proteins were identified in all six studies (categories Ia, IIa, Table 1), of which three have been identified in the stroma or thylakoid as well. In total, we found 231 proteins of category I. Additionally, we found 346 proteins of category III according to Ferro et al. (2010), which are not cross-contaminations of the stroma or thylakoid (Ferro et al., 2010). Hence, they might represent envelope proteins as well. However, as stated above, this conclusion should be challenged by biochemical approaches. The latter holds true for particularly 72 proteins of category II, which have been identified in envelope and in stroma or thylakoid. However, 262 proteins have been assigned to category IV.

Based on the PPDB and SubaII databases, we next analyzed whether proteins have been previously assigned to the mitochondrion, peroxisome, nucleus, ER, golgi, plasma membrane or cytosol, and not to the plastid (Table 1; Heazlewood et al., 2007; Sun et al., 2009). Accordingly, 31/12 proteins of category I were assigned to other cellular localizations according to PPDB/SubaII, respectively. In category II we found 2/0 proteins and in category III 74/56 proteins, respectively, which have been identified in cellular compartments other than chloroplasts. Thus, about 10 and 20% of the proteins assigned to category I or category III are found in other cellular compartments than the chloroplast. The low abundance of mislocalized proteins in category II might reflect that the proteome of the stroma and thylakoid (Ferro et al.,

Table 1 | Categories for the classification of envelope membrane proteins.

Category	Mixed envelope fraction		Thylakoid or Stroma	Proteins identified in the given number of studies							
	This study	Others	Ferro et al. (2010)	6	5	4	3	2	1	SUM	OTH
Ia	+	At least one	–	27	12	25	17	18	–	99	11/8
Ib	–	At least two	–	–	5	11	41	75	–	132	20/4
IIa	+	At least two	+	3	14	11	16	–	–	44	0/0
IIb	–	At least three	+	–	2	9	17	–	–	28	2/0
IIIa	+	None	–	–	–	–	–	–	48	48	6/10
IIIb	–	One	–	–	–	–	–	–	298	298	68/46
IVa	+	Less than two	+	–	–	–	–	35	21	56	0/0
IVb	–	Less than three	+	–	–	–	–	53	153	206	16/0

Given is the category defined in the text (column 1), protein identification by us (+, column 2) or by any other proteomic study (column 3 defines the required number of identifications), identification in the thylakoid or stroma (+, Ferro et al., 2010; column 4) and the number of proteins identified in 6, 5, 4, 3, 2, or 1 study (column 5–10) as well as the number of proteins in each category (column 11), and the number of proteins, which have been identified in other cellular fractions than chloroplasts as well based on PPDB (Sun et al., 2009)/Suball (Heazlewood et al., 2007; column 12).

2010) has been established quite well. Nevertheless, the assignment of proteins in other organellar fractions does not necessarily mark them as false positive chloroplast proteins as (i) chloroplasts are the major organelles of plant cells and thus, contaminations of other fractions might exist and (ii) an increasing number of proteins are found to be dually localized (Carrie and Small, 2013).

COMPARISON OF THE IDENTIFIED ENVELOPE PROTEOMES OF THE DIFFERENT SPECIES

Next, we compared our envelope proteomes obtained for the different analyzed plants with focus on proteins assigned to categories I–III (Table S8 in Supplementary Material). The 191 of total proteins assigned included 48 proteins identified in *M. truncatula*, 68 proteins in *A. thaliana*, and 127 proteins in *P. sativum*. Thirty-nine proteins were identified in at least two plant species, 13 of which were found in all three (Figure 2A; Table 2). Dissecting the protein set of *P. sativum* into OE and IE localized, revealed a total of 46 OE and 60 IE proteins. Twenty-one proteins were found in both fractions. We compared the OE and IE proteins separately with the identified envelope proteins of the other two plants (Figure 2B). This analysis shows that all 13 proteins identified in all species were also found in the IE, while 7 of them are also found in the OE. Similarly, all proteins found in the overlap between *P. sativum* and *A. thaliana* are found in the IE fraction (11), while the overlap with the *M. sativa* envelope contains four proteins (AT2G01320, AT4G32250, Toc64-III, and Toc132) specifically found in the OE of *P. sativum*.

The set of proteins found in all three species include amongst others solute transporters like LptD and Iep37 and as part of the IE/OE preprotein translocases Toc75-III, Toc159, and Tic55-II. Remarkably, only a single protein with unknown function was identified in all envelope fractions, namely At5g08540. Additionally, seven proteins of category II are detected in all species including the photosynthesis proteins LHCB6, PSAD-2, and ATPB. Furthermore, three proteins involved in signaling and response (CA1, RCA, and FNR1) and SDX1 of the lipid biosynthesis are identified. Remarkably, we could identify only one protein of category I in the envelope fractions of *A. thaliana* and *M. sativa*,

which is the dually targeted (mitochondria and chloroplast) S-adenosylmethionine carrier 1 (SamC1; Palmieri et al., 2006). The category I proteins involved in transport (Oep16, NAP8), pre-protein import (CJD1, Tic110), and signaling (MDH) could be detected in the envelope fractions of *A. thaliana* and *P. sativum* (Table 2). It appears that subfractionation of IE and OE membranes in case of the samples from *P. sativum* lead to an increased detection of preprotein import (Toc120, Tic55-IV, and Tic40) and transporter (Oep37, NAP14, MEX1, KEA2, DiT1, and DiT2.1) proteins. For the envelope fractions of *A. thaliana* and *M. sativa* only the preprotein import protein Toc75-V (*M. sativa*) and the transport proteins KEA1, TIP1.1, PIP2A, and PCaP1 in *A. thaliana* and Oep16-2 in *M. sativa* could be identified.

THE OUTER AND INNER ENVELOPE MEMBRANE PROTEOME

Next, we inspected the individual proteomes of the OE and IE membrane of *P. sativum*, respectively. We only assigned proteins of categories I and III, which have been identified by at least two peptides. Due to the high uncertainty, proteins of category II were omitted (see above). Taking these criteria into account we could assign 30 proteins of known function to the OE (Table 3), and 34 proteins to the IE membrane (Table 4) and additional 22 proteins could not be clearly assigned (Table 5). In addition, we assigned 50 proteins of unknown function, 15 of them to the IE, 20 to the OE (fraction), and 15 to the envelope in general (Table 6). Thus, in total we were able to clearly assign 50 OE and 49 IE proteins (Figure 3) and will explain them in detail in the following sections.

Outer envelope proteins

We identified homologs to known OE proteins such as components of the TOC complex (Schleiff and Becker, 2011), like Toc75-III, Toc34, Toc159, Toc120 and Toc132, and Toc64-III which have been previously reported (Schleiff et al., 2003a; Ladig et al., 2011). The latter three were exclusively found in the OE membrane. Remarkably, we were not able to detect Toc75-V, except in the envelope fraction of *M. sativa* (Table 5). Further identified proteins with confirmed OE localization were Oep37, Oep21, and Oep16 (Schleiff et al., 2003a), SENSITIVE TO FREEZING 2

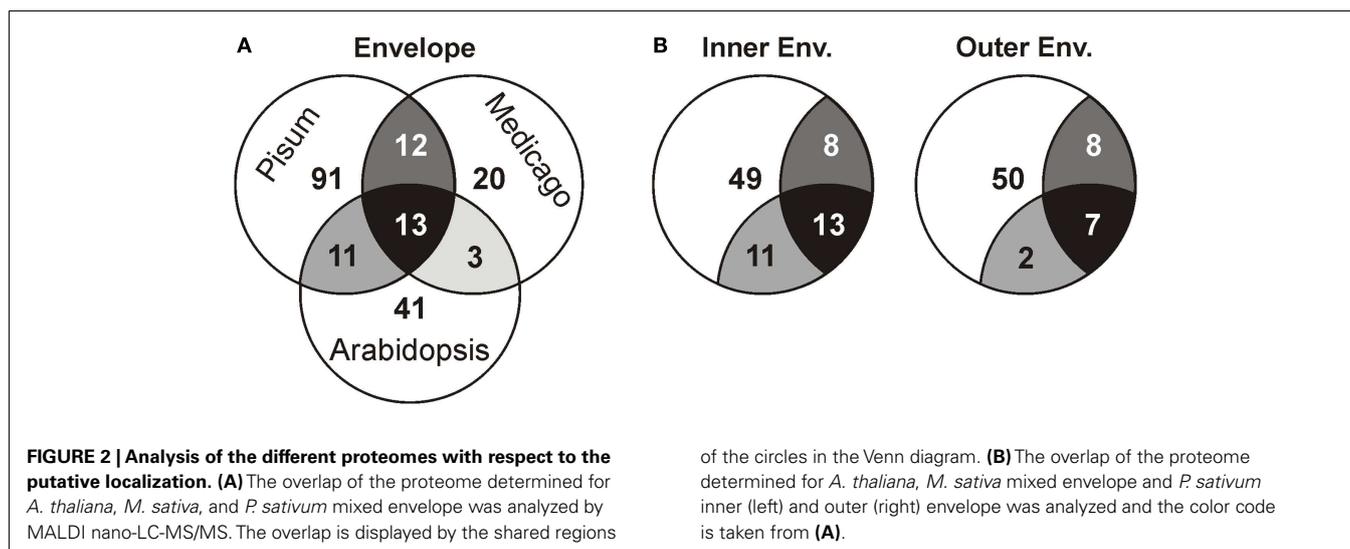


Table 2 | Proteins in the envelope fraction of at least two species.

AGI	Abbr.	<i>A. thaliana</i>	<i>M. sativa</i>	<i>P. sativum</i>			AGI	Abbr.	<i>A. thaliana</i>	<i>M. sativa</i>	<i>P. sativum</i>		
				OE	IE	M					OE	IE	M
<i>At1g03130</i>	PSAD-2	X	X			X	<i>At3g26740</i>	CCL	X				X
<i>At1g06950</i>	Tic110	X				X	<i>At3g46740</i>	Toc75-III	X	X			X
<i>At1g08640</i>	CJD1	X				X	<i>At3g47520</i>	MDH	X			X	
<i>At1g15820</i>	LHCB6	X	X			X	<i>At3g63410</i>	Iep37	X	X			X
<i>At1g55670</i>	PSAG		X			X	<i>At4g02510</i>	Toc159	X	X			X
<i>At1g65260</i>	VIPP1		X			X	<i>At4g15440</i>	HPL1		X			X
<i>At1g67090</i>	RBCS1A		X			X	<i>At4g20360</i>	RAB8d	X				X
<i>At1g74470</i>	Unknown	X				X	<i>At4g25450</i>	NAP8	X				X
<i>At1g77590</i>	LACS9		X			X	<i>At4g32250</i>	Unknown		X		X	
<i>At2g01320</i>	Unknown		X		X		<i>At4g32260</i>	Unknown	X	X			
<i>At2g16640</i>	Toc132		X		X		<i>At4g32770</i>	SDX1	X	X			X
<i>At2g24820</i>	Tic55-II	X	X			X	<i>At4g33350</i>	Unknown		X			X
<i>At2g28900</i>	Oep16	X				X	<i>At4g39460</i>	SAMC1	X	X			
<i>At2g39730</i>	RCA	X	X			X	<i>At5g05000</i>	Toc34		X			X
<i>At2g44640</i>	LptD	X	X			X	<i>At5g08540</i>	Unknown	X	X			X
<i>At3g01500</i>	CA1	X	X			X	<i>At5g24650</i>	Unknown		X			X
<i>At3g04340</i>	emb2458	X				X	<i>At5g50920</i>	HSP93-V	X				X
<i>At3g16950</i>	LPD1	X				X	<i>At5g66190</i>	FNR1	X	X			X
<i>At3g17970</i>	Toc64-III		X		X		<i>AtCg00480</i>	ATPB, PB	X	X			X
<i>At3g23400</i>	FIB4	X	X										

Given is the *Arabidopsis* Genome Initiative (AGI) number (*italic* indicates category II), the short name and aliases, the identification in *A. thaliana* mixed envelope, *M. sativa* mixed envelope, *P. sativum* outer envelope, inner envelope, or mixed envelope. Identification of the protein is marked by an X in the column.

protein (Sfr2), a galactolipid-remodeling enzyme (Fourrier et al., 2008; Moellering et al., 2010) and CRUMPLED LEAF protein (Crl) and PDV2, which are both involved in plastid division (Asano et al., 2004; Glynn et al., 2008).

Additionally, we included proteins, for which significantly more peptides were found in the OE than in the IE fraction, albeit their exact localization is unclear. The long-chain acyl-CoA synthetase

Lacs9 (Schnurr et al., 2002), the ABC-type transporter WBC7 (Ferro et al., 2003), and the paralog of TGD4 (Xu et al., 2008) encoded by AT2G44640 (LptD; Haarmann et al., 2010) were shown to be localized in the envelope membranes, before (Ferro et al., 2003; Froehlich et al., 2003). Similarly, the kinase CoaE was identified in the chloroplast proteome, but experimental data on the localization does not exist (Zybailov et al., 2008).

Table 3 | Outer envelope proteins with known function.

	AGI	Abbr.	Name and function	TM fold	Other Loc.	Studies	Cat.
Preprotein import	At2g16640	Toc132	GTP-binding chloroplast preprotein receptor	Unknown	–	3	I
	At3g16620	Toc120	GTP-binding chloroplast preprotein receptor	Unknown	n.d.	2	I
	At3g17970	Toc64-III	Chloroplast preprotein receptor	α -Helical TM	–	3	I
	At3g46740	Toc75-III	Translocon channel	β -Barrel	–	6	I
	At4g02510	Toc159	GTP-binding chloroplast preprotein receptor	Unknown	–	6	I
	At5g05000	Toc34	GTP-binding chloroplast preprotein receptor	1 α -Helical TM	–	6	I
Lipid biosyn.	At1g77590	LACS9	Long-chain acyl-CoA synthetase	1 α -Helical TM	–	6	I
	At3g06510	SFR2	Beta-glucosidase	2 α -Helical TM	–	6	I
	At4g31780	MGD1	Type A monogalactosyldiacylglycerol synthase	None	–	5	I
Transport	At1g20816	Oep21	Outer envelope channel	None	n.d./–	3	I
	At1g45170	Oep24	Outer envelope channel	None	–/n.d.	1	III
	At2g01320	WBC7	Putative subfamily G ABC-type transporter	4 α -Helical TM	–	3	I
	At2g28900	Oep16-I	Outer envelope protein	2 α -Helical TM	–	6	I
	At2g43950	Oep37	Outer membrane ion channel	β -Barrel	–	6	I
	At2g44640	LptD	Lipopolysaccharide-assembly protein D	β -Barrel	–	6	I
Others	At2g16070	PDV2	Plastid division machinery	1 α -Helical TM	–	3	I
	At2g17390	AKR2B	AKR2-like protein	None	n.d./–	1	III
	At2g27490	COAE	Putative dephospho-CoA kinase	None	X	2	I
	At3g27820	MDAR4	Membrane-associated monodehydroascorbate reductase	2 α -Helical TM	X	2	I
	At4g05050	UBQ11	Polyubiquitin	None	n.d.	1	III
	At4g29130	HXK1	Glucose-responsive sensor hexokinase	1 α -Helical TM	X	4	I
	At5g17770	CBR1	NADH:cytochrome b5 reductase	1 α -Helical TM	n.d./X	2	I
	At5g51020	CRL	Affects pattern of plastid division	1 α -Helical TM	–	2	I
At5g58140	NPL1	Multifunctional blue-light-responsive photoreceptor	None	X	2	I	
Other Organelles	At1g27390	Tom20-2	Putative mitochondrial outer membrane translocase component	1 α -Helical TM	n.d./X	3	I
	At3g46030	HTB11	Putative H2B-type histone	None	X/n.d.	1	III
	At4g14430	ECH1b	Putative enoyl-CoA hydratase/isomerase	1 α -Helical TM	X	1	III
	At4g35000	APX3	Putative peroxisomal ascorbate peroxidase	1 α -Helical TM	X	5	I
	At4g38920	VHA-C3	c-Type subunit of vacuolar H(+)-ATPase membrane V0 subcomplex	4 α -Helical TM	n.d./X	1	III
	At5g43070	WPP1	Nuclear envelope-targeted protein involved in mitotic activity	None	n.d./X	2	I

Given is the functional pathway or the organellar compartment, the AGI number, the short name (Abbr.), the (putative) function, the transmembrane anchor architecture, other localization by the PPDB (Sun et al., 2009) and SUBAII (Heazlewood et al., 2007; n.d. not defined, –, no other localization, X, other localization), the number of studies where the protein was identified (our study; Ferro et al., 2003; Froehlich et al., 2003; Bräutigam et al., 2008; Bräutigam and Weber, 2009; Ferro et al., 2010), and the category from our study. Lipid biosynthesis (Lipid biosyn.).

The protein encoded by AT5G27330 is annotated as Prefold in chaperone subunit family protein and was predicted to be localized in the endoplasmic reticulum (Dunkley et al., 2006). Likewise, ascorbate peroxidase Apx3 (Narendra et al., 2006) was previously assigned to peroxisomal membranes, while Cbr1 (Fukuchi-Mizutani et al., 1999) was described as a protein of the microsomal electron-transfer system. Remarkably, both proteins were identified as substrates of the Akr2a-dependent transport (Shen et al., 2010), which is also involved in the transport of Oep7 to the chloroplast OE membrane (Bae et al., 2008). Furthermore, Apx3 was previously identified in the chloroplast proteome (Zybailov et al., 2008). Although unclear, these proteins are most likely dually localized to both, peroxisomes or ER and chloroplasts.

In contrast, we identified a couple of proteins, which indicate a slight impurity of the sample, namely Mdar4

(Lisenbee et al., 2005), which was clearly assigned to the peroxisomal membrane, Wpp1 (Patel et al., 2004) and Hxk1 (Moore et al., 2003), which are nuclear proteins, the mitochondrial proteins Tom20 and AT4G16450 (Lister et al., 2007; Klodmann et al., 2010) vacuolar protein AVA-P3, and IE protein MGD1, the MGDG synthase (Awai et al., 2001; Ladig et al., 2011). Further, LHCB1.4 is a thylakoid protein. AT4G27680, AT3G52230, AT2G32240, AT3G53560, AT2G24440, AT1G09920, AT3G49350, and AT1G68680 are unknown, while for the protein kinase encoded by AT4G32250 a stromal localization was proposed (Friso et al., 2004; Zybailov et al., 2009).

Inner envelope proteins

Analyzing the IE proteome of *P. sativum*, we realized that it was in contrast to the OE fraction heavily contaminated with proteins of

Table 4 | Inner envelope proteins with known function.

	AGI	Abbr.	Name and function	TM fold	Other Loc.	Studies	Cat
Preprotein import	At1g06950	Tic110	Inner envelope translocon component	1 α -Helical TM	–	6	I
	At1g08640	CJD1	DnaJ-like membrane protein	3 α -Helical TM	–	5	I
	At2g24820	Tic55-II	Inner envelope Rieske iron-sulfur protein	3 α -Helical TM	–	5	I
	At4g23420	Tic32-IVb	NAD- or NADP-dependent oxidoreductase	1 α -Helical TM	n.d.	2	I
	At4g25650	Tic55-IV	Inner envelope Rieske iron-sulfur protein	2 α -Helical TM	–	4	I
	At4g33350	Tic22-IV	Inner envelope translocon component	None	–	6	I
	At5g16620	Tic40	Inner envelope translocon component	1 α -Helical TM	–	6	I
Lipid biosyn.	At4g15440	HPL1	Membrane-associated hydroperoxide lyase	2 α -Helical TM	–	4	I
	At4g31500	SUR2	Cytochrome P450 monooxygenase	2 α -Helical TM	X	1	III
	At5g01220	SQD2	UDP-sulfoquinovose:DAG sulfoquinovosyltransferase	1 α -Helical TM	–	4	I
	At5g05580	FAD8	chloroplast omega-3 fatty acid desaturase	3 α -Helical TM	–	3	I
Transport	At1g80300	NTT1	Plastidic ATP/ADP antiporter	11 α -Helical TM	–	5	I
	At3g20320	TGD2	Putative subfamily I ABC protein	1 α -Helical TM	–	6	I
	At3g63410	IEP37	37 kDa chloroplast inner envelope protein	2 α -Helical TM	–	6	I
	At4g00630	KEA2	Putative potassium cation efflux antiporter	14 α -Helical TM	–	4	I
	At4g25450	NAP8	Putative subfamily B ABC-type transporter	5 α -Helical TM	–	6	I
	At5g12860	DiT1	Plastidic 2-oxoglutarate/malate-translocator	13 α -Helical TM	–	6	I
	At5g14100	NAP14	Putative subfamily I ABC protein	None	–	3	I
	At5g17520	MEX1	Putative maltose translocator	9 α -Helical TM	–	3	I
	At5g24650	PRAT2.2	Putative dual-targeted mitochondrial and plastidial membrane translocase	4 α -Helical TM	–	6	I
	At5g64290	DiT2.1	Plastidic glutamate/malate-translocator	11 α -Helical TM	–	6	I
SR	At1g32080	LRGB	LrgB-like membrane protein	12 α -Helical TM	–	5	I
	At3g47520	MDH	NAD-malate dehydrogenase	1 α -Helical TM	–	4	I
	At5g23040	CDF1	Cell growth defect factor	3 α -Helical TM	–	5	I
Proteases	At1g79560	FtsH12	ATP-dependent metalloprotease	2 α -Helical TM	–	4	I
	At5g53170	FtsH11	ATP-dependent metalloprotease	1 α -Helical TM	X	4	I
	At5g64580	FtsH14	Putative ATP-dependent metalloprotease	1 α -Helical TM	–	4	I
Embryon. develop.	At1g10510	Emb2004	Putative membrane protein	1 α -Helical TM	–	6	I
	At3g04340	Emb2458	Putative membrane protein	3 α -Helical TM	–	4	I
	At3g52590	UBQ1	Ubiquitin extension protein	None	X	1	III
	At5g22640	Emb1211	Putative membrane protein	None	–	4	I
	At5g53860	Emb2737	Putative membrane protein	None	–	3	I
Others	At1g65260	VIPP1	Membrane-associated vesicle-inducing prot.	None	–	6	I
	At2g37860	LCD1	Mutant lcd1-1 exhibits pale phenotype	2 α -Helical TM	–	4	I

Given is the functional pathway or the organellar compartment, the AGI number, the short name (Abbr.), the (putative) function, the transmembrane anchor architecture, other localization by the PPDB (Sun et al., 2009) and SUBAII (Heazlewood et al., 2007; n.d., not defined, –, no other localization, X, other localization), the number of studies where the protein was identified (our study; Ferro et al., 2003; Froehlich et al., 2003; Bräutigam et al., 2008; Bräutigam and Weber, 2009; Ferro et al., 2010), and the category from our study. Lipid biosynthesis (Lipid biosyn.); signaling and response (SR); embryonic development (Embryon. develop.).

the stroma and the OE. First, with clearly annotated OE proteins like Toc75-III, Toc159, Toc34, Lacs9, and Oep21. Second, with stromal proteins like the small subunit of ribulose biphosphate carboxylase Rbcs1A, Rbcl and the ATP-dependent RuBisCO activase (RCA), the malate dehydrogenase (MDH), and subunit PsG of photosystem I complex as prominent stromal contaminations. For Emb1211, PsA2-2, the beta-subunit of ATP synthase (ATPB), AT1G33810, and the geranyl reductase AT1G74470 a thylakoid localization was determined (Peltier et al., 2004).

As expected, we identified proteins of the preprotein translocon of the inner membrane (TIC; Soll and Schleiff, 2004), namely Tic110, Tic55-II, Tic55-IV, Tic40, and Tic32-IVb as major components of the IE fraction. Although assigned to category II we identified IE membrane-associated cpHsp70 (two peptides; Su and Li, 2010) and CPN60 (two peptides; Stürzenbaum et al., 2005), which were two chaperones previously discussed to be involved in pre-protein import. We also detected two peptides for intermembrane space localized Tic22-IV. Remarkably, the chloroplast-targeted

Table 5 | Mixed envelope proteins with known function.

	AGI	Abbr.	Name and function	TM fold	Other Loc.	Studies	Cat.
Preprotein import	At5g19620	Toc75-V	Protein translocation channel at OEM	β -Barrel	–	5	I
Lipid biosyn.	AtCg00500	ACCD	Carboxyltransferase beta-subunit of acetyl-CoA carboxylase complex	None	–	1	III
Transport	At1g01790	KEA1	Putative potassium cation efflux antiporter	13 α -Helical TM	–	6	I
	At2g36830	TIP1.1	Putative tonoplast intrinsic protein	6 α -Helical TM	n.d./–	3	I
	At3g53420	PIP2A	Putative plasma membrane intrinsic protein 2a	6 α -Helical TM	–	2	I
	At4g16160	Oep16-2	Putative plastid outer envelope protein	β -Barrel	–	1	III
	At4g20260	PCaP1	Mediates hypocotyls cell elongation	None	–	2	I
	At4g39460	SamC1	S-Adenosylmethionine transporter	5 α -Helical TM	–	5	I
	At5g13450	ATP5	Mitochondrion MF1-ATP synthase subunit	None	X	3	I
SR	At1g15690	AVP-3	Type I proton-translocating pyrophosphatase	14 α -Helical TM	–	4	I
	At1g55020	LOX1	Lipoxygenase	1 α -Helical TM	n.d.	1	III
	At3g09260	PYK10	Beta-glucosidase	1 α -Helical TM	X	1	III
	At3g14210	ESM1	Putative GDSL-type lipase	1 α -Helical TM	n.d./–	2	I
Others	At1g55860	UPL1	Putative ubiquitin-protein ligase 1	1 α -Helical TM	n.d./X	1	III
	At1g80370	CYCA2_4	A-type cyclin	1 α -Helical TM	n.d./X	1	III
	At2g38040	CAC3	Alpha subunit of acetyl-CoA carboxylase complex	None	–	6	I
	At4g19170	CCD4	Putative carotenoid cleavage dioxygenase	None	–	1	III
	At4g22710	CYP706A2	Cytochrome P450 monooxygenase	1 α -Helical TM	n.d./–	1	III
	At5g25980	TGG2	Thioglucoside glucohydrolase	1 α -Helical TM	n.d./–	2	I
Other Organelle	At1g78900	VHA-A	A-type subunit of vacuolar H(+)-ATPase peripheral V1 subcomplex	None	n.d./–	4	I
	At2g38670	PECT1	Phosphoethanolamine cytidylyltransferase	1 α -helical TM	n.d./X	2	I
	At5g15920	SMC5	Putative SMC5-like component of chromosome metabolism	None	n.d.	1	III

Given is the functional pathway or the organellar compartment, the AGI number, the short name (Abbr.), the (putative) function, the transmembrane anchor architecture, other localization by the PPDB (Sun et al., 2009) and SUBAII (Heazlewood et al., 2007; n.d., not defined, –, no other localization, X, other localization), the number of studies where the protein was identified (our study; Ferro et al., 2003; Froehlich et al., 2003; Bräutigam et al., 2008; Bräutigam and Weber, 2009; Ferro et al., 2010), and the category from our study. Signaling and response (SR).

ferredoxin-NADP(+)-oxidoreductase FNR1 (Table 4), which was found to be associated with the IE via interaction with Tic62 before (Küchler et al., 2002), was clearly detected, whereas Tic62 was identified by only one peptide. Similarly, for Tic20 we found only a single peptide as well. The absence or the low coverage of the membrane-inserted TIC proteins might reflect the problems of analyzing membrane proteins in general (Eichacker et al., 2004).

Besides the TIC components, we identified Iep37, which is described as an IE protein involved in Polyquinone biosynthesis (Dreses-Werringloer et al., 1991). Similarly, the cell growth defect factor Cdf1 (Kawai-Yamada et al., 2005), which is able to induce apoptosis when expressed in yeast, was found to be localized in the IE of chloroplasts (Ladig et al., 2011). Sulfoquinovosyldiacylglycerol (SQDG) synthesis occurs in envelope membranes (Seifert and Heinz, 1992) and here identified SQDG synthase (SQD2; Yu et al., 2002) was localized in chloroplasts. Similarly, we detected the stromal FAD8 (Matsuda et al., 2005) involved in lipid desaturation, and TGD2 involved in transport of lipids from the ER to chloroplasts (three peptides; Awai et al., 2006).

Further, we detected the ATP/ADP antiporter of the IE (NTT1; Neuhaus et al., 1997), the preprotein and amino acid transporter family protein Prat2.2 (Murcha et al., 2007) and the potassium cation efflux antiporter KEA2 (Zybailov et al., 2008) with at least four peptides. In addition, one peptide each was found for the putative magnesium cation transporter MGT10 (Froehlich et al., 2003), for the triose-phosphate/phosphate translocator TPT (Schneider et al., 2002), for the mitoferrin-like carrier MFL1 (Tarantino et al., 2011), for the plastidial sodium-dependent pyruvate transporter BAT1 (Furumoto et al., 2011), and two peptides for the plastidic glutamate/malate-translocator (DIT2; Renné et al., 2003), the putative sugar transporter encoded by AT5G59250 (Froehlich et al., 2003), as well as three peptides for the plastidic 2-oxoglutarate/malate-translocator (DIT1; Weber et al., 1995), and for the maltose transporter Mex1 (Niittylä et al., 2004).

The beta-carbonic anhydrase (CA1; Fabre et al., 2007) of category II and the three metalloproteases (category I; FtsH4i, FtsH11, and FtsH12), detected in the IE fraction, were previously allocated to the stroma (Sakamoto et al., 2003), but they might be associated with the IE as well as suggested for Emb2458

Table 6 | Proteins from Category I and III with unknown function.

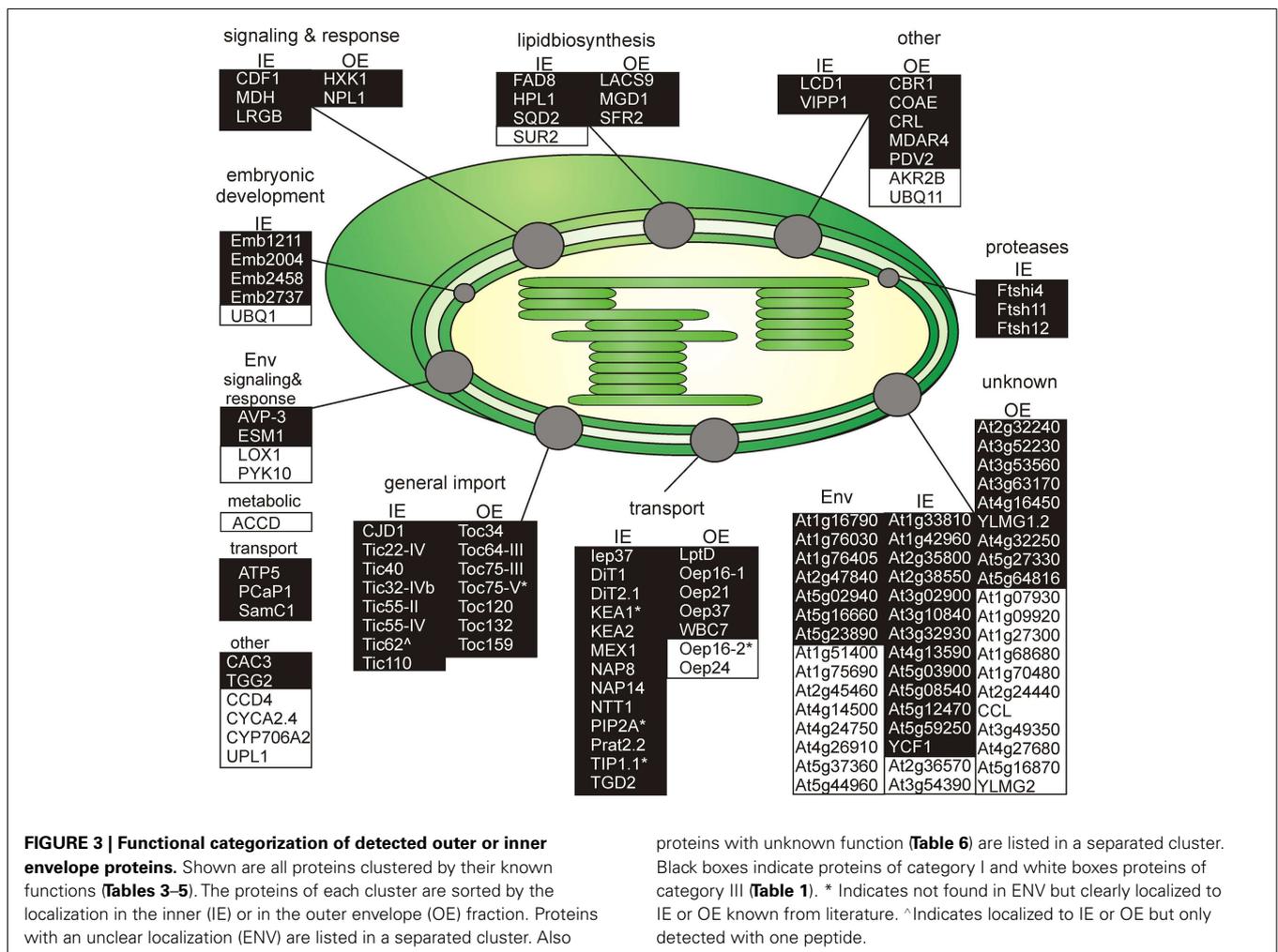
Loc.	AGI	Putative function ^A or Closest homologue ^B or GO annotation ^C	Putative TM fold	Studies	Cat.
Inner envelope membrane	At1g33810	TIM phosphate binding super family ^A	1 α -Helical TM	5	I
	At1g42960	Glutaredoxin2 C ^A	1 α -Helical TM	6	I
	At2g35800	Mitochondrial glutamate carrier (<i>M. truncatula</i>) ^B	2 α -Helical TM	4	I
	At2g36570	Leucine-rich repeat receptor-like protein kinase (Tyrosin kinase) ^{A,B}	1 α -Helical TM	1	III
	At2g38550	Transmembrane proteins 14C (<i>H. sapiens</i>) ^B	4 α -Helical TM	6	I
	At3g02900	YCF1.2 protein ^A	1 α -Helical TM	4	I
	At3g10840	Putative alpha/beta-fold-type hydrolase ^{A,B}	2 α -Helical TM	4	I
	At3g32930	Pterin 4 alpha carbinolamine dehydratase ^A	None	4	I
	At3g54390	Putative DNA-binding protein (SWI3, ADA2, N-CoR, and TFIIB) ^A	None	1	III
	At4g13590	PF27 (small family from bacteria and eukaryotes) belongs to the lysine exporter superfamily ^A	7 α -Helical TM	4	I
	At5g03900	Iron-sulfur cluster biosynthesis family protein ^A	2 α -Helical TM	4	I
	At5g08540	Oligonucleotide/oligosaccharide binding DNA ligase family ^A	1 α -Helical TM	6	I
	At5g12470	DUF3411 and Glycine rich proteien family ^A	4 α -Helical TM	6	I
	At5g59250	Putative sugar transporter ^A D-xylose-proton symporter-like protein ^B	11 α -Helical TM	6	I
	AtCg01130	Ycf1 protein ^A Ycf1 (<i>B. napus</i>) ^B	8 α -Helical TM	3	I
Outer envelope membrane	At1g07930	Putative elongation factor Tu GTP-binding protein ^{A,B}	None	1	III
	At1g09920	Putative PRLI-interacting factor K ^A	1 α -Helical TM	1	III
	At1g27300	Protein binding ^C	1 α -Helical TM	1	III
	At1g68680	Putative mitochondrial respiratory chain complex I ^C	2 α -Helical TM	1	III
	At1g70480	Putative steroidogenic acute regulatory protein; lipid transfer protein ^A	None	1	III
	At2g24440	Putative protein that binds to UDP-glucose:glycoprotein glucosyltransferase ^A /selT/selW/selH selenoprotein (<i>Z. mays</i>) ^B	None	1	III
	At2g32240	Putative SMC protein that bind DNA and act in organizing and segregating chromosomes for partition ^A	1 α -Helical TM	2	I
	At3g26740	CCL putative light regulated protein ^{A,B}	None	1	III
	At3g49350	Putative GTPase activator protein of Rab-like small GTPases ^{A,B}	None	1	III
	At3g52230	Chloroplast outer envelope 24 kD protein like (omp24) ^B	1 α -Helical TM	6	I
	At3g53560	Putative protein protein interaction function ^A	None	2	I
	At3g63170	Putative chalcone-flavonone isomerase ^{A,B}	1 α -Helical TM	4	I
	At4g16450	Putative NADH-ubiquinone oxidoreductase complex I ^{A,B}	1 α -Helical TM	3	I
	At4g27680	Putative ATPases associated ^A Spastin (<i>M. truncatula</i>) ^B	1 α -Helical TM	1	III
	At4g27990	YLMG1-2 (YGGT family protein) YGGT repeat found in conserved hypothetical integral membrane proteins ^A	3 α -Helical TM	5	I
	At4g32250	Putative Serine/Theronine protein kinase ^A G protein-coupled receptor kinase (<i>M. truncatula</i>) ^B	1 α -Helical TM	3	I
	At5g16870	Putative peptidyl-tRNA hydrolase (PTH2) ^{A,B}	1 α -Helical TM	1	III
	At5g21920	YLMG2 (YGGT family protein) YGGT repeat found in conserved hypothetical integral membrane proteins ^A	2 α -Helical TM	1	III
At5g27330	Putative SMC protein that bind DNA and act in organizing and segregating chromosomes for partition ^A	1 α -Helical TM	2	I	
At5g64816	Not determined	1 α -Helical TM	3	I	
Mixed envelope membrane	At1g16790	Ribosomal protein related ^C	None	5	I
	At1g51400	Photosystem II reaction center subunit T ^{A,B}	None	1	III
	At1g75690	Putative chaperone protein dnaJ-related ^A	1 α -Helical TM	1	III
	At1g76030	V-type (H +)-ATPase V1 like protein ^{A,B}	None	2	I
	At1g76405	Putative Cupin-like protein ^A /Outer envelope pore protein of 21 kDa (<i>P. sativum</i>)	None	3	I
	At2g45460	FHA Putative nuclear signaling domain (FHA) ^A	None	1	III
	At2g47840	Chloroplast protein import component Tic20/Ycf60 family protein ^{A,B}	3 α -Helical TM	6	I
	At4g14500	Putative steroidogenic acute regulatory lipid transfer protein ^{A,B}	3 α -Helical TM	1	III
At4g24750	Rhodanese homology domain ^A	1 α -Helical TM	1	III	

(Continued)

Table 6 | Continued

Loc.	AGI	Putative function ^A or Closest homologue ^B or GO annotation ^C	Putative TM fold	Studies	Cat.
	At4g26910	Putative 2-oxoglutarate dehydrogenase E2 component ^{A,B}	None	1	III
	At5g02940	DUF1012 ^A /Os03g163100 (<i>O. sativa</i>) ^B	3 α -Helical TM	4	I
	At5g16660	Apolipoprotein domain ^A /Os12g0583400 (<i>O. sativa</i>) ^B	1 α -Helical TM	4	I
	At5g23890	Os03g0862100 (<i>O. sativa</i>) ^B	1 α -Helical TM	6	I
	At5g37360	Ammonium transporter AES61175.1 (<i>M. truncatula</i>) ^B	2 α -Helical TM	1	III
	At5g44960	Putative F-box like protein associated in nuclear processes ^A	None	1	III

Given is the fraction the protein was localized in (outer/inner/mixed envelope membrane), the AGI number, putative function or functional domain indicated by ^A or closest homolog indicated by ^B or function annotated in GeneOntology (GO) ^C, the transmembrane architecture, the number of studies where the protein was identified with (our study; Ferro et al., 2003; Froehlich et al., 2003; Bräutigam et al., 2008; Bräutigam and Weber, 2009; Ferro et al., 2010), and the category.



(Froehlich et al., 2003). The same holds true for the DnaJ-like membrane protein of unknown function (CJD1; Zybailov et al., 2008). The tocopherol cyclase SXD1 (category II; Provencher et al., 2001) is chloroplast-localized and is involved in tocopherol synthesis, which takes place in the IE membrane (Lichtenthaler et al., 1981). Thus, it is most likely that these six proteins are membrane-associated and correctly assigned to the IE membrane. The proteins encoded by AT1G33810, AT1G42960, AT2G35800,

AT2G38550, AT3G02900, AT3G10840, AT3G32930, AT4G13590, AT5G03900, AT5G08540, and AT5G12470 are assigned as (inner) envelope proteins (Ferro et al., 2003, 2010; Froehlich et al., 2003; Bräutigam et al., 2008; Bräutigam and Weber, 2009), but their function remains to be explored. We further confirmed the IE localization of the plastid-encoded Ycf1.2 (Ladig et al., 2011). The latter might be inserted by the recently identified Sec translocon (Skalitzky et al., 2011).

Non-assignable and unknown proteins

Next to the proteins with known functions that could be clearly assigned to the OE/IE membrane in *P. sativum*, we identified two additional classes of proteins. The first are proteins that have a known function but could not clearly be allocated to either of the membranes (Table 5), because these proteins were found only in the mixed envelope of *A. thaliana* and/or *M. sativa*. Most of these proteins function as transporters like KEA1, TIP1.1, PIP2A, Oep16-2, PCaP1, and SamC1 or signaling and response (AVP-3, LOX1, PYK10, and ESM1). Toc75-V (Schleiff et al., 2003a) was the only preprotein import protein, which could be identified in the mixed envelope fraction but not in the OE or IE membrane of *P. sativum*.

The second are proteins of which neither function nor localizations are known yet (Table 6). These proteins were assigned concerning their identification in OE or IE membrane of *P. sativum*. Two of 15 IE-assigned proteins of unknown function (At2g36570, At3g54390) were only detected in our study, whereas ~50% of OE-assigned unknown proteins are of category III. To characterize the unknown proteins of the two groups and support them as potential new IE/OE envelope proteins we used TOPCONS single (Figure A2 in Appendix, Hennerdal and Elofsson, 2011) and Aramemnon (Schwacke et al., 2003) for secondary structure prediction. Eighty-five percent of the unknown IE proteins possess at least one predicted transmembrane helix (Table 6) and might therefore be anchored or embedded into the IE membrane. None of the unknown OE proteins are found to be β -barrel structures, which would have been an argument for an OE localization (Schleiff et al., 2003a). However, it has to be taken into account, the prediction of eukaryotic β -barrel proteins is not as reliable as of helical proteins (Mirus and Schleiff, 2005). Also, the putative function via Pfam (Finn et al., 2010) and CDD (Marchler-Bauer and Bryant, 2004) and the closest homolog via reciprocal best BLAST hit search were predicted to allocate the proteins correctly (Table 6). Interestingly, most of the unknown proteins assigned to the IE are localized via PPDB and SUBAII to the plastid except of At2g36570 (other localization) and At3g54390 (not determined), whereas most of the OE-assigned proteins are not determined at least by one database and only six proteins are localized in the plastid (At3g26740, At3g52230, At3g53560, At3g63170, At4g27990, and At4g32250).

EXPERIMENTAL SECTION

ISOLATION AND FRACTIONATION OF CHLOROPLASTS

Arabidopsis thaliana

Chloroplasts were isolated from 20-day-old *A. thaliana* plants (Col-0 ecotype Columbia; 8 h light/16 h dark photoperiod of $120 \mu\text{mol m}^{-2} \text{s}^{-1}$; 25°C). Plants were harvested before light onset and all procedures were carried out at 4°C. Leaves were cut and homogenized in 450 mM Sorbitol, 20 mM Tricin-KOH pH 8.4, 10 mM EDTA, 5 mM NaHCO₃, 1 mM PMSF, using a waring blender (four pulses: low speed 3 s; medium speed 3 s; high speed 2 s; low speed 4 s). The homogenate was filtered through four layers of cheesecloth and one layer of miracloth and centrifuged for 5 min at $1,500 \times g$ and 4°C. The pellet was resuspended using a paintbrush in 300 mM Sorbitol, 20 mM Tricin-KOH pH 7.6, 5 mM MgCl₂, 2.5 mM EDTA, 1 mM PMSF (resuspension buffer), placed on top of percoll gradients by underlying 12 ml of 45%

(v/v) Percoll™ with 8 ml of 85% (v/v) Percoll™, and centrifuged for 10 min at $10,000 \times g$. Intact chloroplasts between 40 and 80% (v/v) Percoll™ were collected after removal of broken chloroplasts by water jet pump. Intact chloroplasts were washed twice by centrifugation for 5 min at $1,500 \times g$ in resuspension buffer and collected.

Chloroplasts were lysed by resuspension in 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 mM PMSF (TE buffer) to a final concentration of 2 mg chlorophyll/ml. The suspension was placed on top of a sucrose step-gradient (2.4 ml 1.2 M; 4 ml 1.0 M; 4 ml 0.45 M sucrose in TE buffer) and centrifuged for 2 h at $125,000 \times g$ and 4°C. Chloroplast fractions were recovered by Pasteur pipettes, diluted 1:3 in TE buffer, centrifuged, pooled, and immediately frozen in liquid nitrogen and stored in -80°C .

Pisum sativum

Chloroplast isolation was adapted from Schleiff et al. (2003a,b). Pea (*P. sativum* cv. Arvika) plants were grown for 8 days in a greenhouse (8 h dark/16 h light, $70 \mu\text{mol m}^{-2} \text{s}^{-1}$; 25°C). Pea leaves were harvested and homogenized in the 330 mM Sorbitol, 13 mM Tris, 20 mM MOPS, 0.1 mM MgCl₂, 0.02% (w/v) BSA, 1 mM β -ME, 0.3 mM PMSF using a waring blender (five pulses, low/medium/high/low/medium, all 2 s). The suspension was filtered through four layers of cheesecloth and one layer of miracloth and centrifuged for 5 min at $1,500 \times g$ and 4°C. The pellet was resuspended in the remaining buffer, transferred with cut 5 ml-pipette tip on top of Percoll gradients prepared by underlying 13 ml of 40% (v/v) Percoll™ with 8 ml of 80% (v/v) Percoll™, centrifuged for 10 min at $10,000 \times g$ and 4°C. Intact chloroplasts were collected from the phase between 40 and 80% Percoll™ and washed twice in 330 mM Sorbitol, 1 mM β -ME, and 0.3 mM PMSF.

Chloroplasts were osmotically shocked by adding 2.4 M sucrose solution to a final concentration of 0.6 M sucrose and incubation for 10 min in dark, followed by mechanical disruption with 50 strokes in a dounce homogenizer. Solution was mixed with 2.4 M sucrose solution to a final concentration of 1.35 M, overlaid with 10 ml 1.1 M, 10 ml 1.0 M, and 8 ml 0.45 M sucrose solutions, respectively. Chloroplast sub-compartments were recovered after centrifugation for 18 h at $125,000 \times g$ and 4°C, resuspended in 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 mM PMSF, and stored in -80°C .

Medicago sativa

Chloroplast isolation, and subsequent fractionation, from Alfalfa seedlings was performed as described for pea chloroplast with the following modifications. Seedlings were grown for 20 days and leaves were homogenized in a waring blender (2×3 pulses at low speed for 3 s; at medium speed for 3 s at high speed for 2 s). Further, Percoll™ gradients were prepared by underlying 13 ml of 42% (v/v) Percoll™ with 8 ml of 82% (v/v) Percoll™.

PROTEOME ANALYSIS BY MALDI NANO-LC-MS/MS

Preparation for enzymatic digestion

An amount of 120 μg membranes were washed using 25 mM NH₄HCO₃ pH 8.0 and carbamidomethylated prior to digestion. After 2 min of centrifugation at $12,000 \times g$ the supernatant was removed, and the pellet was gently resuspended in 100% (v/v) methanol. Sample reduction with DTT was performed at 56°C for 45 min and 10 μl of a 500 mM iodoacetamide in 25 mM

NH_4HCO_3 solution was used for sulfhydryl alkylation. Following a 10 min period of sonication, the methanol was diluted to 60% (v/v) using 25 mM NH_4HCO_3 buffer. The proteolytic digestion was performed by adding either 2 μg of trypsin (three biological replicates each organism and envelope fraction) or 10 μg of elastase (three biological replicates each organism and envelope fraction) for 16 h at 36°C. Prior to storing at -20°C the peptide-containing sample was centrifuged at $12,000 \times g$ for 2 min in order to remove all undigested membranes and finally the supernatant was concentrated to 15 μl .

Mass spectrometry

Extracted peptides were subjected to MALDI nLC-MS/MS. Specifically, extracted peptides were injected into an Easy-nLC from Proxeon Systems (Thermo Fisher Scientific, Dreieich, Germany) using solvent A [8% (v/v) acetonitrile, 0.1% (v/v) trifluoroic acid]. Separation was performed on a thermostatic (40°C) custom made C_{18} column (X-Bridge™ BEH 180 C_{18} 300 Å 3.5 μm , 75 $\mu\text{m} \times 150$ mm) at a flow rate of 300 nl/min with increasing acetonitrile concentrations. The linear-gradient profile was used for tryptic peptide digests started with 8–90% solvent B [95% (v/v) acetonitrile, 0.1% (v/v) trifluoroic acid] in 75 min, a stagnation at this level for 8 min, followed by a quick decline to 8% in 5 min and finally, an additional 2 min at 8% for column equilibration. In the case of elastase generated peptide mixtures, the linear-gradient profile duration was increased to 105 min. The separated peptides were then mixed on a tee (Upchurch Scientific) with matrix solution supplied by an auxiliary pump (flow rate, 1.0 $\mu\text{l}/\text{min}$). This solution contained 3 mg/ml α -Cyano-4-hydroxycinnamic acid (α -CHCA; Bruker Daltonics, Germany) dissolved in 70% (v/v) acetonitrile, 30% (v/v) H_2O , and 0.1% (v/v) trifluoroic acid. The final mixture was directly spotted every 20 s on a blank 123 mm \times 81 mm Opti-TOF™ LC/MALDI insert metal target. Subsequent MALDI-TOF/TOF measurements were carried out using the 4800 TOF/TOF Analyzer (Applied Biosystems, Germany). All peptides used for calibration were taken from the Sequazyme™ Peptide Mass Standards kit (Applied Biosystems, Germany). Spectra were acquired in the positive reflector mode between 700 and 4000 m/z with fixed laser intensity. A total of 750 laser shots per spot were accumulated. The precursor selection for MS/MS was carried out via the software of the instrument to avoid unnecessary multiple selections of identical precursor peptides. Up to 10 precursors per spot were selected for fragmentation each requiring a minimum signal-to-noise ratio of 30. The fragmentation of the selected precursors was performed at collision energy of 1 kV using air as collision gas at a pressure of 1×10^{-6} torr. Depending on the spectral quality, 1250–2500 laser shots were recorded. Potential matrix cluster signals were removed from precursor selection by excluding all masses in the range from 700 to 1400 m/z having values of 0.030 ± 0.1 m/z as well as the internal calibrant mass.

DATA ANALYSIS AND PRESENTATION

Format parsing

Mascot generic format (mgf) files were retrieved from each nLC-MALDI MS/MS run (three biological replicates each organism and envelope fraction; Table S14 in Supplementary Material) using

the built-in Peaks2Mascot feature, exporting up to 65 peaks per MS/MS spectrum, each requiring a minimum signal-to-noise of 5. MS/MS queries were processed using the Mascot database search engine v2.2.03 (Matrix Science Ltd.). Data were analyzed using the following settings: below 60 ppm MS precursor mass tolerance (except for the OE of *P. sativum* in combination with trypsin which was 90 ppm due to a technical problem with the instrument that day) and below 0.5 Da MS/MS mass tolerance for MALDI-TOF/TOF. For all database searches, the post-translational modifications carbamidomethylation of cysteins and oxidation of methionines were both selected as variable. When tryptic searches were performed, up to three missed cleavages were taken into consideration in combination with a specific cleavage after K and R and not before P. In all elastase searches, the number of missed cleavages was set to the maximum value of 9 and enzyme specificity was set to A, V, L, I, S, and T, but not before P. For all samples, a custom Viridiplantae database was generated from UniProtKB containing 887,260 entries as of March 02, 2011. Additionally, for *P. sativum* and *M. sativa* samples, customized databases containing 79,106 and 47,532 sequences were provided by the EST-library (Franssen et al., 2012) and MT3.0 of the IMGAG¹, respectively. False discovery rates (FDR, Table S14 in Supplementary Material) given are those originating from the internal Mascot decoy database search function. For each nLC-MALDI-MS/MS run and each sample, the ions score cut off was calculated individually as $-\log(p)$ with $p = 0.05$ (95% confidence level; Table S14 in Supplementary Material). The Mascot analyses were described in the paper of Rietschel et al. (2009). For multiple fragmentations of identical precursors, due to the reappearance in repetitions, only data from the highest scoring peptide were kept. Significant proteins present in all three triplicates were taken and summarized in one table for each type of experiment. Afterward, these tables of elastase and trypsin treatments, containing non-identical hits and peptides, were fused.

Peptide assignment

Depending on the source the peptides identified by Mascot or Sequest were afterward aligned either to the protein database of TAIR9 (*A. thaliana*²), the protein database of MIPS (*M. truncatula*³), or the data file of contigs und singlets (*P. sativum*, data file from Franssen et al. (2012) using a standalone version of Blast from NCBI (substitution matrix BLOSUM62 with linear gap penalty). Following criteria were applied: peptides were only assigned to proteins in the database, if (i) they were aligned with an identity of $>95\%$ (determined via blastp), (ii) they had no gaps or mismatches except for (iii) a single substitution with amino acid residues with similar qualities (defined by the substitution matrix) or a single undefined amino acid position (declared by X). Short peptides (<11 aa), which were already covered by assigned peptides, were not subject to the previously mentioned criteria. Those short peptides were assigned to the protein, although they were

¹ ftp://ftpmips.gsf.de/plants/medicago/MT_3_0/, International Medicago Genome Annotation Group

² <ftp://ftp.arabidopsis.org/home/tair/>

³ ftp://ftpmips.gsf.de/plants/medicago/MT_3_0/

not aligned with BLAST, which is insufficiently accurate regarding the assignment of peptides shorter than 11 amino acids. This method was used to reduce redundancy and as a more stringent criterion for the detection of proteins via the predicted peptides of Mascot. Also, we used a single method to assign the different species and databases in the same way under the same parameter settings of BLAST. Additionally, we searched in parallel for the closest homolog of *A. thaliana* in the other species.

The peptides allocated to *P. sativum* or *M. truncatula* are also allocated to the possible orthologs in *A. thaliana*. On the basis of the *A. thaliana* gene identifiers and their allocated peptides, the splice variants of the proteins were merged to a single gene identifier. The next step to reduce the abundance was connecting all gene identifiers with exactly the same allocated peptides. These gene identifiers were summed up and given the name of the gene identifier with the most allocated peptides or the shortest amino acid sequence by identity. In the end gene identifiers with an overlap of allocated peptides were also combined to one gene identifier. The name of the gene identifier was chosen on the basis of the number of uniquely allocated peptides or the length of the amino acid sequence. All proteins with only one allocated peptide were handled as not significant and are listed in Tables S5–S7 in Supplementary Material.

Prediction of outer/inner envelope membrane proteins

All gene identifiers including splice variants and proteins, which could be identified with the allocated peptides were used to predict the envelope membrane proteins. Two different experimental approaches were applied for *P. sativum*. The first approach for mass spectrometry analyses contained purified OE proteins. The other approach contained purified IE proteins. The peptides detected by MS were blasted against a database of contigs and singlets of *P. sativum*. For classification of the detected contigs and singlets to outer or IE proteins, we first had to find orthologs in *A. thaliana*. The contigs of the *P. sativum* database were blasted against the *A. thaliana* protein database and subsequently the best hit was reblasted against the *P. sativum* contigs database to verify the *A. thaliana* protein. The dedicated *A. thaliana* gene identifiers were used for the prediction of the OE and IE membrane proteins. All gene identifiers with at least four assigned peptides were used for the analysis of the membrane protein prediction.

Also the identified gene identifiers were allocated to the sub-compartments in the chloroplasts. For this the Plant Proteome Database (Sun et al., 2009) was used, which includes the experimentally annotated localizations of the *A. thaliana* gene identifiers. In the end, the amino acid sequences of the identified proteins in the envelope pools were used to predict transmembrane α -helices via TOPCONS single⁴ (Hennerdal and Elofsson, 2011).

Database comparison

The proteins of the three different organisms detected in our envelope studies were compared to previous envelope studies including proteomic data for the membrane envelope of plastids by Bräutigam et al. (2008), Bräutigam and Weber (2009), Ferro et al. (2003, 2010), and Froehlich et al. (2003). Also the detected proteins

are categorized concerning their occurrence in the different studies and stroma or thylakoid in this study or the study of Ferro et al. (2010).

Domain and homolog searches, structural predictions

First, the function and the name of the protein represented by the gene identifiers of Tables 3–6 were looked up in Aramemnon rel. 7.0⁵ (Schwacke et al., 2003). Afterwards, the predicted transmembrane fold was annotated. If Aramemnon predicts transmembrane β -barrel structures the sequences of the gene identifiers were used to build 3D models of respective amino acid sequence with the help of alignments to known protein structures via the protein fold recognition server Phyre2 (Kelley and Sternberg, 2009). For the gene identifiers of unknown function, the putative domains were searched using the Protein families database (Pfam; Finn et al., 2010) and the Conserved Domain Database (CDD; Mitra et al., 2007).

CONCLUSION

The determination of subcellular and suborganellar proteomes or alterations thereof (due to, e.g., environmental changes) by mass spectrometry is still limited in respect to protein abundance and sample purity (Figure 1), but most likely not by bioinformatic methods used for protein assignment (Figure A1 in Appendix). The assignment of peptides depends in general on their length and the false positive rate can be regulated by mapping criteria. Unassigned peptides usually observed in such studies can in parts be explained by the stringency of the mapping criteria, but point also toward natural variances at the protein level.

In the study at hand, we performed proteomic analyses of chloroplast envelope membranes from three different plant species. The necessity to sustain proteomic studies on the analyses of different species was formerly shown by the unexpected high diversity of soluble chloroplast proteomes, when comparing data from *A. thaliana* and *P. sativum* (Bayer et al., 2011). The comparison of envelope fractions from different plant species in our study increased the number of detected proteins but did not result in a large intersection of these envelope proteins (Figure 2; Table 2).

Furthermore, when comparing our findings with previous proteomic envelope approaches, we were able to refine the available proteome data and assign a reliable, comprehensive core proteome. Contrary to expectations, intersection of proteins identified in these studies was rather small (Table 1). Altogether, we identified 191 potential envelope proteins (categories I–III). After detecting putative cross-contaminations of stromal and thylakoid proteins the remaining 136 envelope proteins were clustered according to their predicted/confirmed localization and cellular function (Figure 3). To this end 35 IE, 24 OE, and 19 known non-assignable envelope proteins were identified. Amongst these UBQ1 and SUR2 as well as AKR2B, UBQ11, Oep16-2, and Oep24 were newly assigned to IE and OE, respectively.

Moreover, we identified 21 new potential envelope proteins of category III of unknown function which might give rise to further analyses. Finally, we observed differences concerning the predicted localizations in the independent studies which point

⁴<http://single.topcons.net/>

⁵<http://aramemnon.uni-koeln.de/index.ep>

toward a possible membrane-association or a possible dual or multi-sublocalization inside the chloroplast or cell.

ACKNOWLEDGMENTS

We are grateful to Markus T. Bohnsack for support. The work was supported by grants from the Deutsche Forschungsgemeinschaft SFB807-P17 and from the Volkswagenstiftung to Enrico Schleiff.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://www.frontiersin.org/Plant_Proteomics/10.3389/fpls.2013.00011/abstract

Table S1 | Proteins identified in the *A. thaliana* envelope membrane fraction. The first column gives the AGI number, the second the number of identified splice variants, the third the AGI of the splice variants, the fourth the AGI code of similar proteins detected, the fifth column the number of peptides assigned to the protein only, the sixth column the number of peptides additionally assigned to other proteins, and the seventh column a short description of the protein. In the second sheet the AGI number and all identified peptides are listed. Every peptide is identified by MS/MS.

Table S2 | The proteins identified in the *M. sativa* envelope membrane fraction. The first column gives the AGI number, the second column the *Medicago* specific ID, the third the number of identified splice variants, the fourth the AGI of the splice variants, the fifth the AGI code of similar proteins detected, the sixth column the number of peptides assigned to the protein only, the seventh column the number of peptides additionally assigned to other proteins, and the eighth column a short description of the protein. In the second sheet the *Medicago* ID number and all identified peptides are listed. Every peptide is identified by MS/MS.

Table S3 | The proteins identified in the *P. sativum* outer envelope membrane fraction. The first column gives the AGI number, the second column the *Pisum* specific ID, the third the number of identified splice variants, the fourth the AGI of the splice variants, the fifth the AGI code of similar proteins detected, the sixth column the number of peptides assigned to the protein only, the seventh column the number of peptides additionally assigned to other proteins, and the eighth column a short description of the protein. In the second sheet the *Pisum* ID number and all identified peptides are listed. Every peptide is identified by MS/MS.

Table S4 | The proteins identified in the *P. sativum* inner envelope membrane fraction. The first column gives the AGI number, the second column the *Pisum* specific ID, the third the number of identified splice variants, the fourth the AGI of the splice variants, the fifth the AGI code of similar proteins detected, the sixth column the number of peptides assigned to the protein only, the seventh column the number of peptides additionally assigned to other proteins, and the eighth column a short description of the protein. In the second sheet the *Pisum* ID number and all identified peptides are listed. Every peptide is identified by MS/MS.

Table S5 | Peptides identified by analysis of *A. thaliana* fractions not assigned to a protein. The peptide, the type of digestion yielding the peptide and the fraction(s) the peptide was identified in is given in sheet one. In sheet two the *Arabidopsis* ID, the peptide, the type of digestion yielding the peptide, the fraction(s) the peptide was identified in, and the short description of the protein is given for all proteins identified by a single peptide only. In sheet three the *Arabidopsis* IDs, the peptide, the type of digestion yielding the peptide, and the fraction(s) the peptide was identified in is given for all peptides leading to the identification of multiple proteins.

Table S6 | Peptides identified by analysis of *P. sativum* fractions not assigned to a protein. The peptide, the type of digestion yielding the peptide, and the fraction(s) the peptide was identified in is given in sheet one. In sheet

two the *Arabidopsis* ID, the *Pisum* ID, the peptide, the type of digestion yielding the peptide, the fraction(s) the peptide was identified in, and the short description of the protein is given for all proteins identified by a single peptide only. In sheet three the *Arabidopsis* IDs, the *Pisum* IDs, the peptide, the type of digestion yielding the peptide, and the fraction(s) the peptide was identified in is given for all peptides leading to the identification of multiple proteins.

Table S7 | Peptides identified by analysis of *M. sativa* fractions not assigned to a protein. The peptide, the type of digestion yielding the peptide, and the fraction(s) the peptide was identified in is given in sheet one. In sheet two the *Arabidopsis* ID, the *Medicago* ID, the peptide, the type of digestion yielding the peptide, the fraction(s) the peptide was identified in, and the short description of the protein is given for all proteins identified by a single peptide only. In sheet three the *Arabidopsis* IDs, the *Medicago* IDs, the peptide, the type of digestion yielding the peptide, and the fraction(s) the peptide was identified in is given for all peptides leading to the identification of multiple proteins.

Table S8 | List of all identified proteins. The *Arabidopsis* IDs of all proteins identified in this study including those with only one peptide matching are listed. The first column gives the ID, the second column the predicted compartment the protein is supposed to be localized in, the column 3 the *Arabidopsis* fraction, columns 6 and 7 the two *Pisum* fractions, and column 8 the *Medicago* fraction; the last column indicates whether the protein is identified in at least one fraction by more than one peptide (norm) or whether identification occurred by one peptide match only (onehit). The fraction the protein was identified in is marked by X.

Table S9–S12 | List of all proteins in category I. The first column is the AGI identifier, the second column the name and aliases of the protein, and the third column the number of studies, where the protein was identified. Category Ia are proteins found in our study and at least one other study and category Ib are proteins identified not in our study but at least two other studies. Category IIa are proteins found in our study and at least two other studies but also in the stromal or thylakoid fraction. Category IIb are proteins found in three other studies and also in the stromal or thylakoid fraction. Category IIIa are proteins only identified in our study and category IIIb are proteins found only in one study excluding our study. Category IVa and IVb contains proteins identified in the stromal or thylakoid fraction and only in our and less than two other studies (IVa) or in less than three other studies (IVb).

Table S13 | List of overlapping and not overlapping proteins in the Venn diagram. The first column gives the AGI identifier, the second column the name and aliases of the protein, the third column the number of studies where the protein was identified, the fourth column the category of the protein, the columns 5–9 show in which envelope fractions and plant species the proteins could be identified. X, identified; –, not identified.

Table S14 | List of the ions score cutoff and FDR for nLC-MALDI MS/MS. The first column gives the used MS method, the second column the organism and fraction, the third column the restriction enzyme, the fourth column the used database for searching, the fifth column the number of repetition, the sixth column the ions score cutoff in $-10\log(p)$ by $p = 0.05$, and the seventh column the false discovery rate (FDR). The used databases are the UniProtKB, the MT3.0 from IMGAG for *Medicago truncatula*, and the EST-library by Franssen et al. (2012) for *Pisum sativum*.

Tables S15–S57 | Raw data measured by nLC-MALDI MS/MS. Each excel sheet is grouped in two levels. The first level contains information for each identified accession ID. The first column gives the accession (UniProtKB, IMGAG, or EST-library by Franssen et al., 2012), the second column the coverage, the third column the number of peptide spectrum matches (#PSMs), the fourth column the number of peptides, the fifth column the number of amino acids (#AAs), the sixth column the molecular weight (MW in kDa), the seventh column the isoelectric point (pI), the eighth column the score, and the ninth column the description. The second level contains all peptide information for each accession ID. The second column gives the confidence icon (Low; Medium; High), the third column the peptide sequence, the fourth column the protein accessions, the fifth column the number of proteins, the

sixth column the number of protein groups, the seventh column the activation type (Collision Induced Dissociation, CID), the eighth column the modifications, the ninth column the ion score, the 10th column the expectation value (exp. value), the 11th column the delta score (Δ score), the 12th column the rank, the 13th column the identity High, the 14th column the homology

threshold, the 15th column the charge, the 16th column the mass to charge ratio in daltons (m/z), the 18th column the delta mass (ΔM , difference between the theoretical mass of the peptide and the experimental mass of the precursor ion), the 19th column the matched ions, and the 20th column the spectrum file.

REFERENCES

- Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. *Nature* 422, 198–207.
- Asano, T., Yoshioka, Y., Kurei, S., Sakamoto, W., and Machida, Y. (2004). A mutation of the CRUMPLED LEAF gene that encodes a protein localized in the outer envelope membrane of plastids affects the pattern of cell division, cell differentiation, and plastid division in *Arabidopsis*. *Plant J.* 38, 448–459.
- Awai, K., Maréchal, E., Block, M. A., Brun, D., Masuda, T., Shimada, H., et al. (2001). Two types of MGDG synthase genes, found widely in both 16:3 and 18:3 plants, differentially mediate galactolipid syntheses in photosynthetic and nonphotosynthetic tissues in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 98, 10960–10965.
- Awai, K., Xu, C., Tamot, B., and Benning, C. (2006). A phosphatidic acid-binding protein of the chloroplast inner envelope membrane involved in lipid trafficking. *Proc. Natl. Acad. Sci. U.S.A.* 103, 10817–10822.
- Bae, W., Lee, Y. J., Kim, D. H., Lee, J., Kim, S., Sohn, E. J., et al. (2008). AKR2A-mediated import of chloroplast outer membrane proteins is essential for chloroplast biogenesis. *Nat. Cell Biol.* 10, 220–227.
- Bayer, R. G., Stael, S., Csaszar, E., and Teige, M. (2011). Mining the soluble chloroplast proteome by affinity chromatography. *Proteomics* 11, 1287–1299.
- Bräutigam, A., Shrestha, R. P., Whitten, D., Wilkerson, C. G., Carr, K. M., Froehlich, J. E., et al. (2008). Low-coverage massively parallel pyrosequencing of cDNAs enables proteomics in non-model species: comparison of a species-specific database generated by pyrosequencing with databases from related species for proteome analysis of pea chloroplast envelopes. *J. Biotechnol.* 136, 44–53.
- Bräutigam, A., and Weber, A. P. M. (2009). Proteomic analysis of the proplastid envelope membrane provides novel insights into small molecule and protein transport across proplastid membranes. *Mol. Plant* 2, 1247–1261.
- Carrie, C., and Small, I. (2013). A reevaluation of dual-targeting of proteins to mitochondria and chloroplasts. *Biochim. Biophys. Acta* 1833, 253–259.
- Drees-Werringloer, U., Fischer, K., Wachter, E., Link, T. A., and Flügge, U. I. (1991). cDNA sequence and deduced amino acid sequence of the precursor of the 37-kDa inner envelope membrane polypeptide from spinach chloroplasts. Its transit peptide contains an amphiphilic alpha-helix as the only detectable structural element. *Eur. J. Biochem.* 195, 361–368.
- Dunkley, T. P. J., Hester, S., Shadforth, I. P., Munions, J., Weimar, T., Han-ton, S. L., et al. (2006). Mapping the *Arabidopsis* organelle proteome. *Proc. Natl. Acad. Sci. U.S.A.* 103, 6518–6523.
- Eichacker, L. A., Granvogel, B., Mirus, O., Müller, B. C., Miess, C., and Schleiff, E. (2004). Hiding behind hydrophobicity. Transmembrane segments in mass spectrometry. *J. Biol. Chem.* 279, 50915–50922.
- Fabre, N., Reiter, I. M., Becuwe-Linka, N., Genty, B., and Rumeau, D. (2007). Characterization and expression analysis of genes encoding alpha and beta carbonic anhydrases in *Arabidopsis*. *Plant Cell Environ.* 30, 617–629.
- Ferro, M., Brugiére, S., Salvi, D., Seigneurin-Berny, D., Court, M., Moyet, L., 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., Brugiére, S., Miras, S., Kowalski, S., Louwagie, M., et al. (2003). Proteomics of the chloroplast envelope membranes from *Arabidopsis thaliana*. *Mol. Cell. Proteomics* 2, 325–345.
- Finn, R. D., Mistry, J., Tate, J., Cog-gill, P., Heger, A., Pollington, J. E., et al. (2010). The Pfam protein families database. *Nucleic Acids Res.* 38, D211.
- Fourrier, N., Bédard, J., Lopez-Juez, E., Barbrook, A., Bowyer, J., Jarvis, P., et al. (2008). A role for SENSITIVE TO FREEZING2 in protecting chloroplasts against freeze-induced damage in *Arabidopsis*. *Plant J.* 55, 734–745.
- Franssen, S. U., Shrestha, R. P., Bräutigam, A., Bornberg-Bauer, E., and Weber, A. P. M. (2012). Comprehensive transcriptome analysis of the highly complex *Pisum sativum* genome using next generation sequencing. *BMC Genomics* 12:227. doi:10.1186/1471-2164-12-227
- Friso, G., Giacomelli, L., Ytterberg, A. J., Peltier, J.-B., Rudella, A., Sun, Q., et al. (2004). In-depth analysis of the thylakoid membrane proteome of *Arabidopsis thaliana* chloroplasts: new proteins, new functions, and a plastid proteome database. *Plant Cell* 16, 478–499.
- Froehlich, J. E., Wilkerson, C. G., Ray, W. K., McAndrew, R. S., Osteryoung, K. W., Gage, D. A., et al. (2003). Proteomic study of the *Arabidopsis thaliana* chloroplast envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis. *J. Proteome Res.* 2, 413–425.
- Fukuchi-Mizutani, M., Mizutani, M., Tanaka, Y., Kusumi, T., and Ohta, D. (1999). Microsomal electron transfer in higher plants: cloning and heterologous expression of NADH-cytochrome b5 reductase from *Arabidopsis*. *Plant Physiol.* 119, 353–362.
- Furumoto, T., Yamaguchi, T., Ohshima-Ichii, Y., Nakamura, M., Tsuchida-Iwata, Y., Shimamura, M., et al. (2011). A plastidial sodium-dependent pyruvate transporter. *Nature* 476, 472–475.
- Glynn, J. M., Froehlich, J. E., and Osteryoung, K. W. (2008). *Arabidopsis* ARC6 coordinates the division machineries of the inner and outer chloroplast membranes through interaction with PDV2 in the intermembrane space. *Plant Cell* 20, 2460–2470.
- Goulas, E., Schubert, M., Kieselbach, T., Kleczkowski, L. A., Gardeström, P., Schröder, W., et al. (2006). The chloroplast lumen and stromal proteomes of *Arabidopsis thaliana* show differential sensitivity to short- and long-term exposure to low temperature. *Plant J.* 47, 720–734.
- Haarmann, R., Ibrahim, M., Stevanovic, M., Bredemeier, R., and Schleiff, E. (2010). The properties of the outer membrane localized lipid A transporter LptD. *J. Phys. Condens. Matter* 22, 454124.
- Heazlewood, J. L., Verboom, R. E., Tonti-Filippini, J., Small, I., and Millar, A. H. (2007). SUBA: the *Arabidopsis* subcellular database. *Nucleic Acids Res.* 35, D213–D218.
- Hennerdal, A., and Elofsson, A. (2011). Rapid membrane protein topology prediction. *Bioinformatics* 27, 1322–1323.
- James, P. (1997). Protein identification in the post-genome era: the rapid rise of proteomics. *Q. Rev. Biophys.* 30, 279–331.
- Karas, M., and Hillenkamp, F. (1988). Laser desorption/ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* 60, 2299–2301.
- Kawai-Yamada, M., Saito, Y., Jin, L., Ogawa, T., Kim, K.-M., Yu, L.-H., et al. (2005). A novel *Arabidopsis* gene causes Bax-like lethality in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 280, 39468–39473.
- Kelley, L. A., and Sternberg, M. J. E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* 4, 363–371.
- Kleffmann, T., Hirsch-Hoffmann, M., Gruissem, W., and Baginsky, S. (2006). plprot: a comprehensive proteome database for different plastid types. *Plant Cell Physiol.* 47, 432–436.
- Kleffmann, T., Russenberger, D., von Zychlinski, A., Christopher, W., Sjölander, K., Gruissem, W., et al. (2004). The *Arabidopsis thaliana* chloroplast proteome reveals pathway abundance and novel protein functions. *Curr. Biol.* 14, 354–362.
- Klodmann, J., Sunderhaus, S., Nimtz, M., Jansch, L., and Braun, H. P. (2010). Internal architecture of mitochondrial complex I from *Arabidopsis thaliana*. *Plant Cell* 22, 797–810.
- Küchler, M., Decker, S., Hörmann, F., Soll, J., and Heins, L. (2002). Protein import into chloroplasts involves redox-regulated proteins. *EMBO J.* 21, 6136–6145.
- Ladig, R., Sommer, M. S., Hahn, A., Leisegang, M. S., Papanotiropoulos, D. G., and Ibrahim, M., et al. (2011). A high-definition native polyacrylamide gel electrophoresis system for the analysis of membrane complexes. *Plant J.* 67, 181–194.

- Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40, D1202–D1210.
- Leister, D. (2003). Chloroplast research in the genomic age. *Trends Genet.* 19, 47–56.
- Lichtenthaler, H. K., Prenzel, U., Douce, R., and Joyard, J. (1981). Localization of prenylquinones in the envelope of spinach chloroplasts. *Biochim. Biophys. Acta* 641, 99–105.
- Lisenbee, C. S., Lingard, M. J., and Trelease, R. N. (2005). Arabidopsis peroxisomes possess functionally redundant membrane and matrix isoforms of monodehydroascorbate reductase. *Plant J.* 43, 900–914.
- Lister, R., Carrie, C., Duncan, O., Ho, L. H. M., Howell, K. A., Murcha, M. W., et al. (2007). Functional definition of outer membrane proteins involved in preprotein import into mitochondria. *Plant Cell* 19, 3739–3759.
- Marchler-Bauer, A., and Bryant, S. H. (2004). CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.* 32, W327–331.
- Matsuda, O., Sakamoto, H., Hashimoto, T., and Iba, K. (2005). A temperature-sensitive mechanism that regulates post-translational stability of a plastidial omega-3 fatty acid desaturase (FAD8) in Arabidopsis leaf tissues. *J. Biol. Chem.* 280, 3597–3604.
- Matsushima, R., Kondo, M., Nishimura, M., and Hara-Nishimura, I. (2003). A novel ER-derived compartment, the ER body, selectively accumulates a β -glucosidase with an ER-retention signal in Arabidopsis. *Plant J.* 33, 493–502.
- Millar, A. H., Carrie, C., Pogson, B., and Whelan, J. (2009). Exploring the function-location nexus: using multiple lines of evidence in defining the subcellular location of plant proteins. *Plant Cell* 21, 1625–1631.
- Mirus, O., and Schleiff, E. (2005). Prediction of β -barrel membrane proteins by searching for restricted domains. *BMC Bioinformatics* 6:254. doi:10.1186/1471-2105-6-254
- Mitra, S. K., Gantt, J. A., Ruby, J. F., Clouse, S. D., and Goshe, M. B. (2007). Membrane proteomic analysis of Arabidopsis thaliana using alternative solubilization techniques. *J. Proteome Res.* 6, 1933–1950.
- Moellering, E. R., Muthan, B., and Benning, C. (2010). Freezing tolerance in plants requires lipid remodeling at the outer chloroplast membrane. *Science* 330, 226–228.
- Moore, B., Zhou, L., Rolland, F., Hall, Q., Cheng, W.-H., Liu, Y.-X., et al. (2003). Role of the Arabidopsis glucose sensor HXK1 in nutrient, light, and hormonal signaling. *Science* 300, 332–336.
- Murcha, M. W., Elhafez, D., Lister, R., Tonti-Filippini, J., Baumgartner, M., Philippar, K., et al. (2007). Characterization of the preprotein and amino acid transporter gene family in Arabidopsis. *Plant Physiol.* 143, 199–212.
- Narendra, S., Venkataramani, S., Shen, G., Wang, J., Pasapula, V., Lin, Y., et al. (2006). The Arabidopsis ascorbate peroxidase 3 is a peroxisomal membrane-bound antioxidant enzyme and is dispensable for Arabidopsis growth and development. *J. Exp. Bot.* 57, 3033–3042.
- Neuhaus, H. E., Thom, E., Mohlmann, T., Steup, M., and Kampfenkel, K. (1997). Characterization of a novel eukaryotic ATP/ADP translocator located in the plastid envelope of Arabidopsis thaliana L. *Plant J.* 11, 73–82.
- Niittylä, T., Messerli, G., Trevisan, M., Chen, J., Smith, A. M., and Zee-man, S. C. (2004). A previously unknown maltose transporter essential for starch degradation in leaves. *Science* 303, 87–89.
- Palmieri, L., Arrigoni, R., Blanco, E., Carrari, F., Zanon, M. I., Studart-Guimaraes, C., et al. (2006). Molecular identification of an Arabidopsis S-adenosylmethionine transporter. Analysis of organ distribution, bacterial expression, reconstitution into liposomes, and functional characterization. *Plant Physiol.* 142, 855–865.
- Patel, S., Rose, A., Meulia, T., Dixit, R., Cyr, R. J., and Meier, I. (2004). Arabidopsis WPP-domain proteins are developmentally associated with the nuclear envelope and promote cell division. *Plant Cell* 16, 3260–3273.
- Peltier, J.-B., Cai, Y., Sun, Q., Zabrouskov, V., Giacomelli, L., Rudella, A., et al. (2006). The oligomeric stromal proteome of Arabidopsis thaliana chloroplasts. *Mol. Cell Proteomics* 5, 114–133.
- Peltier, J.-B., Emanuelsson, O., Kalume, D. E., Ytterberg, J., Friso, G., Rudella, A., et al. (2002). Central functions of the luminal and peripheral thylakoid proteome of Arabidopsis determined by experimentation and genome-wide prediction. *Plant Cell* 14, 211–236.
- Peltier, J.-B., Ytterberg, A. J., Sun, Q., and van Wijk, K. J. (2004). New functions of the thylakoid membrane proteome of Arabidopsis thaliana revealed by a simple, fast, and versatile fractionation strategy. *J. Biol. Chem.* 279, 49367–49383.
- Provencher, L. M., Miao, L., Sinha, N., and Lucas, W. J. (2001). Sucrose export defective1 encodes a novel protein implicated in chloroplast-to-nucleus signaling. *Plant Cell* 13, 1127–1141.
- Renné, P., Dreßen, U., Hebbeker, U., Hille, D., Flüggé, U.-I., Westhoff, P., et al. (2003). The Arabidopsis mutant dct is deficient in the plastidic glutamate/malate translocator DiT2. *Plant J.* 35, 316–331.
- Rietschel, B., Arrey, T. N., Meyer, B., Bornemann, S., Schuerken, M., Karas, M., et al. (2009). Elastase digests: new ammunition for shotgun membrane proteomics. *Mol. Cell. Proteomics* 8, 1029–1043.
- Sakamoto, W., Zaltsman, A., Adam, Z., and Takahashi, Y. (2003). Coordinated regulation and complex formation of yellow variegated1 and yellow variegated2, chloroplastic FtsH metalloproteases involved in the repair cycle of photosystem II in Arabidopsis thylakoid membranes. *Plant Cell* 15, 2843–2855.
- Schleiff, E., and Becker, T. (2011). Common ground for protein translocation: access control for mitochondria and chloroplasts. *Nat. Rev. Mol. Cell Biol.* 12, 48–59.
- Schleiff, E., Eichacker, L. A., Eckart, K., Becker, T., Mirus, O., Stahl, T., et al. (2003a). Prediction of the plant beta-barrel proteome: a case study of the chloroplast outer envelope. *Protein Sci.* 12, 748–759.
- Schleiff, E., Soll, J., Küchler, M., Kühlbrandt, W., and Harrer, R. (2003b). Characterization of the translocon of the outer envelope of chloroplasts. *J. Cell Biol.* 160, 541–551.
- Schneider, A., Hausler, R. E., Kolukisaoglu, U., Kunze, R., van der Graaff, E., Schwacke, R., et al. (2002). An Arabidopsis thaliana knock-out mutant of the chloroplast triose phosphate/phosphate translocator is severely compromised only when starch synthesis, but not starch mobilisation is abolished. *Plant J.* 32, 685–699.
- Schnurr, J. A., Shockey, J. M., de Boer, G.-J., and Browse, J. A. (2002). Fatty acid export from the chloroplast. Molecular characterization of a major plastidial acyl-coenzyme A synthetase from Arabidopsis. *Plant Physiol.* 129, 1700–1709.
- Schwacke, R., Schneider, A., van der Graaff, E., Fischer, K., Catoni, E., Desimone, M., et al. (2003). ARAMEMNON, a novel database for Arabidopsis integral membrane proteins. *Plant Physiol.* 131, 16–26.
- Seifert, U., and Heinz, E. (1992). Enzymatic characteristics of UDP-sulfoquinovose-diacylglycerol-sulfoquinovosyl transferase from chloroplast envelopes. *Botan. Acta* 285, 197–205.
- Shen, G., Kuppu, S., Venkataramani, S., Wang, J., Yan, J., Qiu, X., et al. (2010). ANKYRIN REPEAT-CONTAINING PROTEIN 2A is an essential molecular chaperone for peroxisomal membrane-bound ASCORBATE PEROXIDASE3 in Arabidopsis. *Plant Cell* 22, 811–831.
- Skalitzky, C. A., Martin, J. R., Harwood, J. H., Beirne, J. J., Adamczyk, B. J., Heck, G. R., et al. (2011). Plastids contain a second Sec translocase system with essential functions. *Plant Physiol.* 155, 354–369.
- Soll, J., and Schleiff, E. (2004). Plant cell biology: protein import into chloroplasts. *Nat. Rev. Mol. Cell Biol.* 5, 198–208.
- Stürzenbaum, S. R., Arts, M. S. J., and Kammenga, J. E. (2005). Molecular cloning and characterization of Cpn60 in the free-living nematode *Plectus acuminatus*. *Cell Stress* 10, 79.
- Su, P.-H., and Li, H.-M. (2010). Stromal Hsp70 is important for protein translocation into pea and Arabidopsis chloroplasts. *Plant Cell* 22, 1516–1531.
- Sun, Q., Zybailov, B., Majeran, W., Friso, G., Olinares, P. D. B., and van Wijk, K. J. (2009). PPDB, the Plant Proteomics Database at Cornell. *Nucleic Acids Res.* 37, D969.
- Tarantino, D., Morandini, P., Ramirez, L., Soave, C., and Murgia, I. (2011). Identification of an Arabidopsis mitoferrin like carrier protein involved in Fe metabolism. *Plant Physiol. Biochem.* 49, 520–529.
- van Wijk, K. J. (2000). Proteomics of the chloroplast: experimentation and prediction. *Trends Plant Sci.* 5, 420–425.
- Vojta, A., Alavi, M., Becker, T., Hörmann, F., Küchler, M., Soll, J., et al. (2004). The protein translocon of the plastid envelopes. *J. Biol. Chem.* 279, 21401–21405.
- Weber, A., Menzlaff, E., Arbinger, B., Gutensohn, M., Eckerskorn, C., and Fluegge, U.-I. (1995). The 2-oxoglutarate/malate translocator of chloroplast envelope membranes: molecular cloning of a transporter containing a 12-helix motif and expression of the functional protein in yeast cells. *Biochemistry* 34, 2621–2627.

- Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J.-C., et al. (1996). From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino Acid Analysis. *Nat. Biotechnol.* 14, 61–65.
- Xu, C., Fan, J., Cornish, A. J., and Benning, C. (2008). Lipid trafficking between the endoplasmic reticulum and the plastid in Arabidopsis requires the extraplastidic TGD4 protein. *Plant Cell* 20, 2190–2204.
- Ytterberg, A. J., Peltier, J.-B., and van Wijk, K. J. (2006). Protein profiling of plastoglobules in chloroplasts and chromoplasts. A surprising site for differential accumulation of metabolic enzymes. *Plant Physiol.* 140, 984–997.
- Yu, B., Xu, C., and Benning, C. (2002). Arabidopsis disrupted in SQD2 encoding sulfolipid synthase is impaired in phosphate-limited growth. *Proc. Natl. Acad. Sci. U.S.A.* 99, 5732–5737.
- Zabrouskov, V., Giacomelli, L., van Wijk, K. J., and McLafferty, F. W. (2003). A new approach for plant proteomics: characterization of chloroplast proteins of Arabidopsis thaliana by top-down mass spectrometry. *Mol. Cell. Proteomics* 2, 1253–1260.
- Zybilov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O., Sun, Q., et al. (2008). Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS ONE* 3:e1994. doi:10.1371/journal.pone.0001994
- Zybilov, B., Sun, Q., and van Wijk, K. J. (2009). Workflow for large scale detection and validation of peptide modifications by RPLC-LTQ-orbitrap: application to the Arabidopsis thaliana leaf proteome and an online modified peptide library. *Anal. Chem.* 81, 8015–8024.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 09 October 2012; accepted: 15 January 2013; published online: 06 February 2013.

Citation: Simm S, Papatirou DG, Ibrahim M, Leisegang MS, Müller B, Schorge T, Karas M, Mirus O, Sommer MS and Schleiff E (2013) Defining the core proteome of the chloroplast envelope membranes. *Front. Plant Sci.* 4:11. doi: 10.3389/fpls.2013.00011

This article was submitted to *Frontiers in Plant Proteomics*, a specialty of *Frontiers in Plant Science*.

Copyright © 2013 Simm, Papatirou, Ibrahim, Leisegang, Müller, Schorge, Karas, Mirus, Sommer and Schleiff. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.

APPENDIX

