Characterization of the protein import pathway in pea chloroplast

Dissertation der Fakultät für Biologie

der Ludwig-Maximilians-Universität München

vorgelegt von

WaiLing Chang

aus Ipoh, Malaysia

München

17. Juli 2014

Erstgutachter: Prof. Dr. Jürgen Soll

Zweigutachter: Prof. Dr. Jörg Nickelsen

Tag der Abgabe: 17. Juli 2014

Tag der mündlichen Prüfung: 27. Oktober 2014

Summary

In order to sustain their structure and metabolism, chloroplasts and other plastid types must import the majority of their proteins from the cytosol across the envelope membrane. Translocation of these precursor proteins across the double envelope membrane is achieved by two multimeric complexes - the so-called TOC and TIC complexes (**T**ranslocon at the **O**uter envelope of **C**hloroplast and **T**ranslocon at the **I**nner envelope of **C**hloroplast, respectively). N-terminal transit peptides essential for import of the precursor proteins are cleaved after their entry into the stroma. It was thus far believed that all of the different cytosolic precursor proteins would enter the chloroplast through the same, jointly acting TOC/TIC machineries. Recent evidence, however, suggests that multiple, regulated import pathways exist in plastids that involve different import machineries. Different combinations of TOC and TIC proteins were shown to establish different import sites in *Arabidopsis thaliana* with specificity for either photosynthetic proteins (the general import pathway) or non-photosynthetic „*housekeeping*" proteins. Moreover, numerous non-canonical import pathways such as the import of Tic32 and AtQORH mediated by the yet unknown novel import pathway and the import via the secretory pathway were shown to exist. Proteomics studies have revealed the presence of a large number of plastid proteins lacking predictable N-terminal transit sequences for import. The import mechanism for the majority of these proteins has not been determined yet. Examples of the transit sequenceless precursor proteins are the chloroplast envelope quinone oxidoreductase homologue, AtQORH and the chloroplast inner envelope protein 32, Tic32. Both proteins are imported into the inner plastid envelope membrane by a non-canonical pathway (Toc159- and Toc75-independent) and without any proteolytic cleavage. In the present study not only the import characteristic of nine tentative 'non-canonical' chloroplast precursor proteins but also the new interactions between these precursor proteins and the proteins at the organellar surfaces were analyzed. Moreover, a non-canonical precursor protein without the classical transit peptide, the iron superoxide dismutase (FSD1) could be identified. Biochemical crosslinking experiments revealed that FSD1 interacts with new members of the Toc159 family in pea, namely PsToc132 and PsToc120. Using deletion mutants as well as a peptide scanning approach, regions of the precursor protein, which are involved in

receptor binding could be defined. These are distributed across the entire sequence; surprisingly only the extreme N-terminus as well as a C-proximal domain turned out to be essential for targeting and import. En route into the plastid FSD1 engages components of the general import pathway, implying that in spite of the 'non-canonical' targeting information and recognition by a specific receptor, this precursor protein follows a similar way across the envelope as the majority of plastid precursor proteins.

Zusammenfassung

Um ihre Struktur und ihren Metabolismus aufrechtzuerhalten, müssen Plastiden den Hauptteil ihrer im Zytosol synthetisierten Proteine importieren, was deren Transfer über die Hüllmembranen erfordert. Importapparate in der äußeren und inneren Hüllmembran, genannt TOC (**T**ranslocon at the **O**uter envelope of **C**hloroplast) und TIC (**T**ranslocon at the **I**nner envelope of **C**hloroplast), wurden identifiziert, die den Import von diesen plastidären Proteinen vermitteln. N-terminale Transitpeptide, die für den Import dieser Präproteine/Vorstufenproteine unerlässlich sind, werden nach deren Import im Stroma abgespalten. Bisher wurde angenommen, dass alle verschiedenen im Cytosol gebildeten Vorstufenproteine über die gleiche TOC/TIC Maschinerie in den Chloroplasten transportiert werden. Neuere Analysen belegen jedoch die Existenz verschiedener, regulierter Importwege, die unterschiedlichen Importapparate involvieren. So konnte in der Modellpflanze *Arabidopsis thaliana* gezeigt werden, dass verschiedene Kombinationen von TOC und TIC Proteinen unterschiedliche Importwege bilden, die vorzugsweise entweder photosynthetisch aktive Proteine (der sogenannte ‚general import pathway') oder nicht-photosynthetisch aktive („*housekeeping*") Proteine importieren. Weiterhin wurden zahlreiche nicht-klassische Importwege beschrieben, wie zum Beispiel der Import von Tic32 und AtQORH sowie der Import über das endoplasmatische Retikulum und den Golgi-Apparat. Proteom-Analysen ergaben, dass zahlreiche in Plastiden lokalisierte Proteine keine prognostizierbaren N-terminalen Transitpeptide besitzen. Die Art und Weise ihres Imports ist bisher noch relativ unbekannt. Zwei Beispiele solcher Proteine sind ein in der plastidären Hüllmembran lokalisiertes quinone-oxidoreduktase-homolog, genannt AtQORH und eins der TIC-Komponenten, Tic32. Dessen Import in die innere Hüllmembran erfolgte unabhängig von Toc159 und Toc75; zwei Komponenten des Standardproteinimportapparates, sowie ohne jede proteolytische Spaltung. Die vorliegende Arbeit analysierte sowohl die molekulare Importeigenschaften der transitpeptidelosen plastidären Vorstufenproteine als auch deren Interaktion mit Proteinen an den Organellenoberflächen. Darüber hinaus wurde *„iron superoxide dismutase"* (FSD1) als eins der transitpeptidlosen plastidären lokalisierten Proteine identifiziert. Biochemische Crosslinking-Analysen zeigten, dass

FSD1 mit den neuen Toc159-Homologen in Erbsen, PsToc132 und PsToc120 interagiert. Diese Daten lassen stark vermuten, dass das Vorhandensein mehrerer Toc159-Homologe, welcher an den unterschiedlichen TOC-Komplexen in *Arabidopsis thaliana* beteiligt sind, in Erbsen als möglich erschien. Um die Beteiligung des PsToc120 Rezeptorproteins bei der Erkennung und Sortierung der Vorstufenproteine im Cytosol zu untersuchen, wurde eine Kombination aus Deletion und eines Peptid-Arrays des FSD1-Proteins angewendet. Die Bindedomänen zwischen dem PsToc120 Rezeptorprotein und dem Vorstufenprotein, FSD1, wurden bestimmt. Dies ist zufällig über die gesamte Sequenz verteilt. Erstaunlicherweise sind nur der extreme N-Terminus sowie die C-proximale Domäne von FSD1 nötig um die Zielsteuerung und den Import in den Chloroplasten zu gewährleisten. Außerdem zeigte eine systematische Charakterisierung der Importwege von FSD1, dass FSD1, während seines Transports in den Chloroplasten mit den Bestandteilen des Standardproteinimportapparates interagiert. Dies weist darauf hin, dass der Transport von FSD1 in den Chloroplasten, trotz seines ungewöhnlichen N-terminalen Transitpeptids und die Nutzung von speziellen Rezeptorkomponenten, auf die gleiche Weise wie die Mehrzahl der plastidären Proteine erfolgt.

Content

Abbreviations

2D	two dimensional
AMP-PNP	adenosine 5′-(β,γ-imido)triphosphate tetralithium salt hydrate
ATP	adenosine-5'-triphosphate
At	*Arabidopsis thaliana*
B-ME	β-mercaptoethanol
BLAST	basic local alignment search tool
BN-PAGE	blue-native polyacrylamide gel electrophoresis
bp	basepair
CBB	coomassie brilliant blue
cDNA	copy-DNA
C-terminus	carboxyl-terminus
cTP	chloroplast transit peptide
DMSO	dimethylsulfoxide
DoMa	*n*-dodecyl-β-D-maltoside
DTT	dithiothreitol
ECL	enhanced chemiluminescence
E.coli	*Escherichia coli*
ER	endoplasmic reticulum
Fd	ferredoxin
FNR	ferredoxin-NADP(H) oxidoreductase
g	gravity force
GFP	green fluorescence protein
GTPase	guanosine-5'-triphosphatase
IE	inner envelope / inner envelope membrane
IMP	inner membrane protein
IMS	intermembrane space
k, K	kilo, times 1000
kDa	kilo Dalton
Met	methionine
mRNA	messenger-RNA
MS	mass spectrometry
MW	molecular weight
N-terminus	amino-terminus
OD	optical density
OE	outer envelope / outer envelope membrane
OE23	oxygen evolving complex protein of 23 kDa
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG	polyethylene glycol

Ps	*Pisum sativum*
PVDF	polyvinylidene difluoride membrane
RACE	rapid amplification of cDNA ends
rpm	revolutions per minute
RT	room temperature
RuBisCO	ribulose-1,5-bisphosphate carboxylase/oxygenase
SD	standard deviation
SDS	sodium dodecyl sulphate
Sec	secretory
SP	signal peptide
SSU	ribulose-1,5-bisphosphate carboxylase/oxygenase, small subunit
TIC	translocon at the inner envelope of chloroplasts
TOC	translocon at the outer envelope of chloroplasts
TP	transit peptide
TX-100	Triton X-100
TM	transmembrane (domain)
w/v	weight per volume
WT	wild type

Chapter 1

Introduction

1.1. The imprint of endosymbiosis and the origin of plastids

Plastids represent a large set of organelles with distinct physiological functions and morphologies found within all plant cells (Constan et al., 2004; Lopez-Juez & Pyke, 2005). The best studied plastid is the chloroplast – a photosynthetic organelle in plant and green algae cells that is responsible for harvesting energy from sunlight and converting it into sugars and ATP. Genomic analysis of the chloroplast genome revealed that the origin of chloroplasts can be traced back to a cyanobacterial predecessor that was engulfed by a eukaryotic cell in an endosymbiotic event which took place approximately 1.5 – 2.0 billion years ago (Margulis, 1975). During evolution, the prokaryote was reduced to a double membrane-surrounded plastid and vertically transmitted to subsequent generations (Figure 1). An important characteristic of this evolutionary process is the genomic 'downsizing' of the cyanobacterial endosymbiont, i.e. the elimination of superfluous genes and the transfer of essentials ones to the nucleus of the host cell, rendering the endosymbiont a semi – autonomous organelle (Martin & Herrmann, 1998; Timmis et al., 2004). In higher plants, genes of cyanobacterial origin account for only a small percentage of proteins, mainly for those involved in translation and photosynthesis; while the majority of the chloroplastic proteins – more than 95% - are encoded in the nucleus of the host cell (Martin & Herrmann, 1998; Martin et al., 2002). Gene relocation from the chloroplast genome typically requires a return ticket for the gene product back to its place of function. A hallmark for this scenario is the development of protein trafficking systems and regulatory networks for the delivery of proteins translated in the cytoplasm back to the compartment of origin within the chloroplast where the proteins perform their function. This post-translational protein trafficking mechanism is mainly achieved two multimeric protein complexes (also known as the general import complexes) located at the outer (TOC – Translocon at the Outer envelope of Chloroplasts) and inner (TIC – Translocon at the Inner envelope of the Chloroplasts) envelopes of chloroplasts, respectively (Soll, 2002).

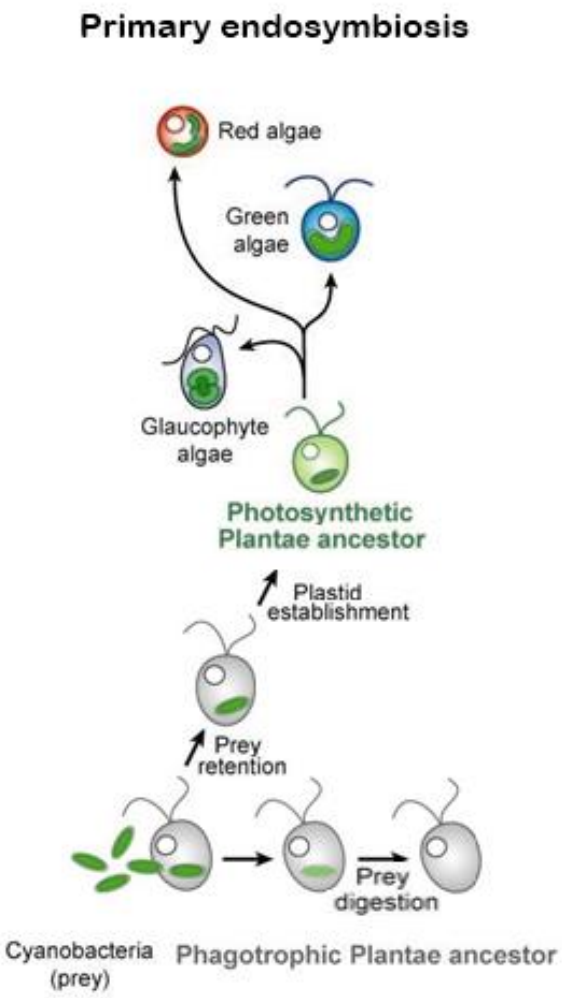


Figure 1 | Model of the primary endosymbiotic origin of the plastid. Usually, the phagotrophic Plantae ancestor digests the taken-up cyanobacterial preys. During the course of evolution, the event of phagotrophy led to the retention of the cyanobacterial prey and, subsequently, to massive gene lost in the endosymbiont and progressive transfer of the endosymbiotic genes to the host nucleus. Thereafter, this ancestral alga lost the ability for phagotrophy and diversified into the extant lineages of green, red, and glaucophyte algae.(Modified after Reyes-Prieto et al., 2007)

1.2. Organization and functions of the chloroplast

The chloroplasts serve not only as a platform for oxygenic photosynthesis, a process which is essential for all life on earth, but they also feature a large number of biosynthetic pathways. These include steps in carbon and nitrogen assimilation as well as biosynthesis of amino acids, lipids, vitamins, hormones and secondary metabolites (e.g. terpenoids and porphyrin) (Browse et al., 1986; Camara, 1984; Folkes, 1970; Gas et al., 2009; Gerrits et al., 2001; Leister, 2003; Neuhaus & Emes, 2000; Takahashi et al., 2005; Wang et al., 2010). Chloroplasts have a discoid structure, with an approximate diameter of 5 to 10 µm. Depending on the cell type and species, the number of

chloroplasts that are present in each cell may vary (between 1 to more than 100) (Block et al., 2007; Kirk, 1971; Lopez-Juez & Pyke, 2005; Rudowska et al., 2012). Chloroplasts can be subdivided into six distinct compartments (Figure 2): three different membrane systems (outer, inner and thylakoid membranes), an intermembrane space between the two envelope membranes, a soluble interior between the inner membranes and the thylakoid membranes called stroma, and an aqueous lumen within the thylakoids (Jarvis, 2008). Interestingly, the chloroplasts as well as mitochondria possess their own genomic DNA and show similarities, e. g. size and shape, with bacteria suggesting an endosymbiotic origin.

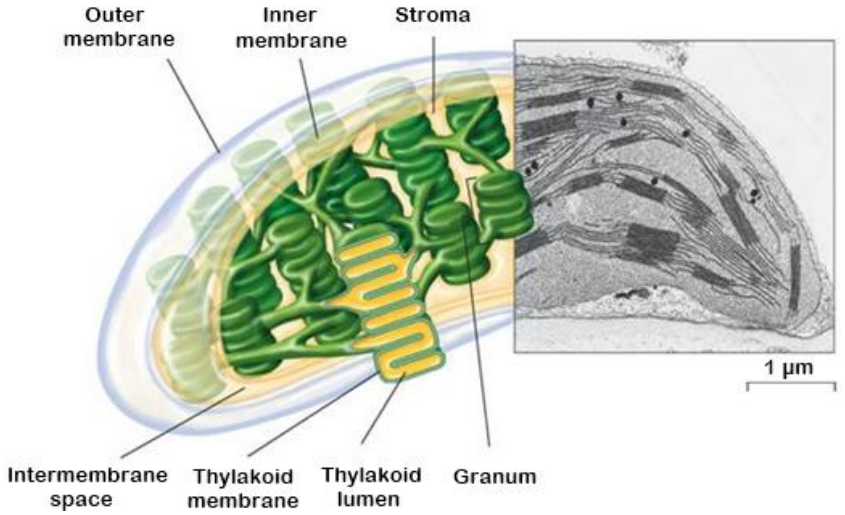


Figure 2 | The plant cell chloroplast. The sub-cellular organization of the chloroplast includes three different membrane systems: the outer and inner envelope membranes and the thylakoid membrane. Encased within these membrane systems are three additional compartments: the inter-membrane space (between the outer and the inner envelope membranes), the soluble stroma and the aqueous lumen within the thylakoid membranes. A granum is a stack of thylakoid disc (Adapted and modified from Thomson Higher Education 2007).

1.3. Translocation across the chloroplast envelopes

The process of genomic re–organizing following endosymbiosis caused obvious challenges for the cell: (1) the need to ensure correct targeting of proteins synthesized by cytosolic ribosomes to the chloroplast and / or other organelles, such as mitochondria or peroxisomes, (2) transport of these nascent proteins across a double membrane, and (3) re–routing of the transported proteins to their destination in the chloroplast, i.e. stroma, thylakoid and thylakoid lumen (Jarvis, 2008). While the mechanism of protein transport across the double membranes of chloroplast is a novel

process, intraplastidial sorting is thought to be a pre–existing ancient mechanism inherited from the cyanobacterial predecessors. Indeed, similar pathways between chloroplasts and their predecessors have been observed (Albiniak et al., 2012; Robinson et al., 2001).

The majority of chloroplast–destined proteins are synthesized as precursor proteins by cytosolic ribosomes with an N–terminal cleavable presequence (from here on referred to as chloroplast transit peptide, cTP), harbouring all the information necessary for the post-translational targeting and translocation into plastid. So far, the consensus features required for targeting are still poorly understood – owing largely to the heterogeneous nature of the cTP itself. They are remarkably divergent in both the primary amino acid composition as well as the structural organization. cTPs also vary substantially in length, ranging from 20 to >100 amino acid residues long (Zhang & Glaser, 2002). Despite the lack of primary sequence similarity, there are several shared features. These include the abundance of small non–polar residues as well as basic and hydroxylated amino acid residues (serine and threonine). By contrast acidic residues are almost absent in the cTPs (Bruce, 2000; Jarvis, 2008). Additionally, they appear not to form any specific secondary structures in solution, but adopt instead a random coil conformation. Recent analysis of the transit peptidome revealed the presence of multiple semi–conserved sub–domains with distinctive sequence motifs that seem to be involved at different stages of the targeting and translocation process (D. W. Lee et al., 2008; D. W. Lee et al., 2006; D. W. Lee et al., 2009b).

Such precursor proteins are recognized by receptors at the chloroplasts surface and are translocated into the organelle via the coordinated action of protein complexes of the general import machinery, composed of the TOC and TIC translocon at the outer and inner envelope membranes of chloroplasts, respectively. Cytosolic chaperones like Hsp90 and Hsp70, the latter forming a “guidance complex” with 14-3-3 proteins, support the targeting step by keeping the precursor proteins in an unfolded conformation, which is required for import (May & Soll, 2000; Qbadou et al., 2006). It is generally believed that the TOC and TIC translocons are able to interact physically in order to allow simultaneous translocation of the precursor proteins across the two chloroplast membrane (Schnell & Blobel, 1993) (Figure 3).

The translocation process was shown to be dependent on internal ATP (Theg et al., 1989) and has at least three distinguishable steps: (1) The 'docking' stage, where the precursor proteins attach reversibly to the chloroplast surface in an energy independent process (Kouranov & Schnell, 1997; Olsen et al., 1989); (2) The envelope associated precursor proteins form an 'early intermediate' with the import machinery spanning the chloroplast outer membrane en route to the stroma. This binding step is irreversible and is promoted by hydrolysis of low concentration of ATP (≤100 μM) (Schnell & Blobel, 1993). A low level of GTP additionally supports this step and enhances ATP – dependent docking, although GTP alone is unable to substitute for ATP (Olsen & Keegstra, 1992; Olsen, et al., 1989; Young et al., 1999); and, (3) The 'late intermediate stage' which involves simultaneous translocation of the precursor proteins across both envelope membranes into the chloroplast stroma followed by the removal of the transit peptide. This stage requires millimolar concentrations of ATP (≥1 mM), which possibly attribute for the activity of stromal chaperones (i.e. Hsp60, Hsp70 and Hsp93)(Inoue et al., 2013; Nielsen et al., 1997; Pain & Blobel, 1987; Shi & Theg, 2010). As the precursor proteins emerge into the stroma, they are processed to either their mature– or intermediate–sized forms by the stromal processing peptidase (SPP) (Inaba & Schnell, 2008; Kovacs-Bogdan et al., 2010; Li & Chiu, 2010; Soll & Schleiff, 2004) and are then assembled into its functional conformation with the assistance of molecular chaperones (Kessler & Blobel, 1996; Lubben et al., 1989). Alternatively, the proteins may be further directed to other sub – compartments within the chloroplast.

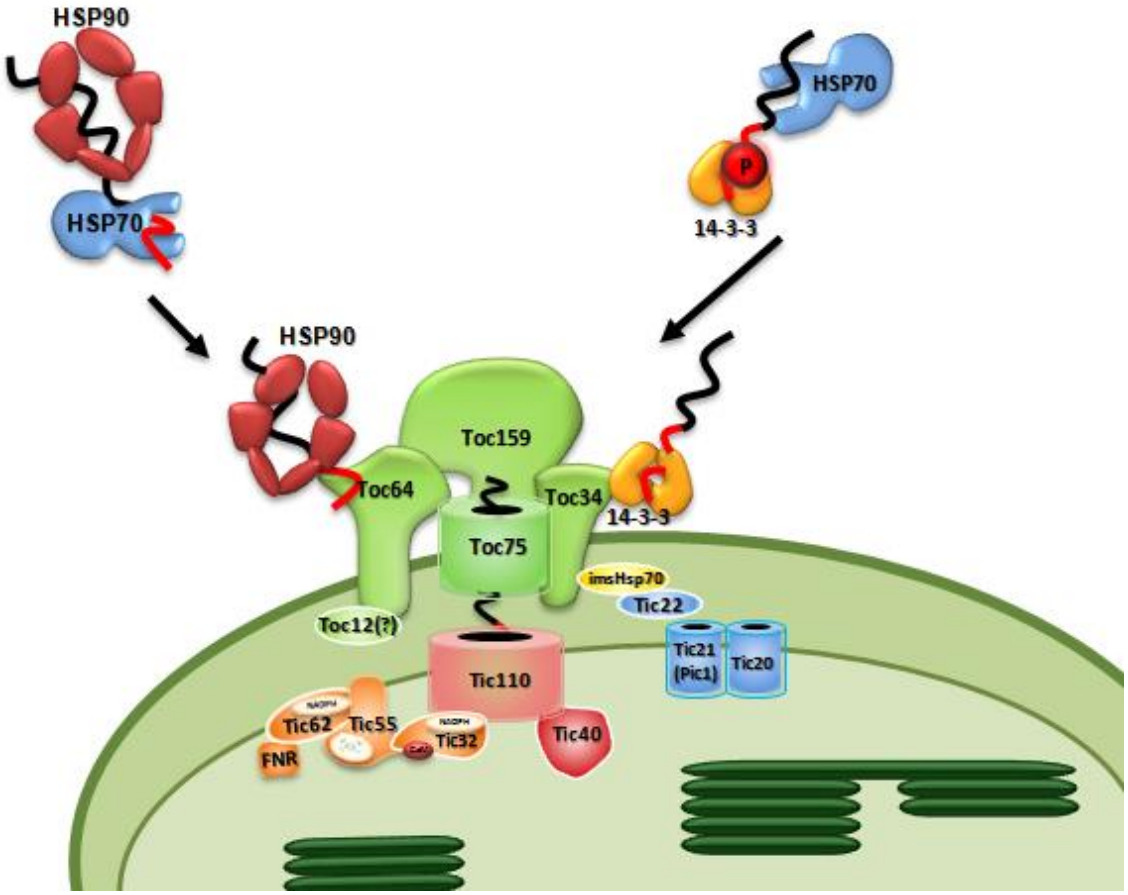


Figure 3 | Schematic illustration of the general TOC / TIC import machinery in pea. Most chloroplast proteins synthesized in the cytosol containing a removable N-terminal transit peptide for plastid targeting and translocation via the TOC / TIC complexes. These nascent proteins are transported to the chloroplast with the help of chaperone-assisted complexes (e-g-Hsp70/14-3-3 or Hsp90) and, are recognized by the receptor constituents of the TOC complex (Toc34, Toc159 and Toc64). Translocation across the outer envelope membrane and inner membrane space (IMS) is facilitated by the Toc75 channel and the IMS complex; (Toc12, imHsp70 and Tic22), respectively. Tic110 and Tic20 are proposed to form the Tic channel. Tic21 is also proposed as a putative protein-conducting channel. Tic40 functions as co-chaperone, whereas Tic32, Tic55 and Tic62 are three redox-sensing auxiliary elements which modulate protein import according to the metabolic redox state of the chloroplast.

1.4. The molecular architecture of the TOC complex

Toc75, Toc159 and Toc34 were among the first components of the chloroplast import machinery to be identified of pea chloroplasts (Schnell & Blobel, 1993; Waegemann & Soll, 1996). Toc75 is a β–barrel protein constituting the protein translocation channel across the outer envelope membrane (Keegstra & Cline, 1999; Schnell et al., 1994). Both receptor components, Toc159 and Toc34 associated with Toc75-are integral proteins at the outer membrane that function through a cycle of GTP hydrolysis (Hirsch & Soll, 1995; Kessler et al., 1994; Seedorf et al., 1995). They are unique to plastids and are responsible for recognition of nuclear–encoded precursor proteins at the outer envelope. Together, Toc159, Toc34 and Toc75 form a stable core

TOC complex capable and sufficient for precursor protein translocation in artificial lipid vesicles *in vitro* (Schleiff et al., 2003a). This core TOC complex was found to exist in an 'oligomeric form' with an approximate molecular mass between 550 kDa (Kikuchi et al., 2009; Schleiff et al., 2003b). This variation of molecular mass might be explained by the existence of multiple copies of the core TOC complex constituents, which is further supported by the reported Toc75:Toc34:Toc159 stoichiometry of 4:4:1 (Schleiff, et al., 2003b). 2D electron microscopy revealed that, the core TOC complex forms an approximately circular particle, enclosed by a dense outer ring with a central 'finger' domain that divides the central cavity into four apparent pores (Schleiff, et al., 2003b), correlating well with the observed stoichiometry.

Further components associated with the TOC core complex is an accessory receptor, Toc64. The Toc64 receptor was first reported as a 64 kDa protein, which co–purifies on sucrose density gradients with isolated TOC complex from pea chloroplasts (Sohrt & Soll, 2000). It possesses a short N–terminal hydrophobic transmembrane anchor, a central region with homology to amidases, and three tetratricopeptide repeats (TPRs) at its C–terminus, which are exposed to the cytosol. Further studies revealed that Toc64 only transiently associates with the TOC core complex (Schleiff, et al., 2003a) and functions in providing a docking site for Hsp90–affiliated preproteins via its TPR domain (Qbadou, et al., 2006), indicating a possible 'fine-tuning' function in post–translational protein translocation across the chloroplast outer envelope.

Although most of the components of the import machinery were originally identified from pea chloroplasts, homologues are reported in moss (*Physcomitrella* patens) as well as all seed plants analyzed. In some of them, several components (particularly components of the core TOC complex) are encoded by multigene families. For instance, the *Arabidopsis thaliana* genome encodes two paralogues of Toc34 (AtToc33 and AtToc34) (Gutensohn et al., 2000; Jarvis et al., 1998) and, four paralogues each of Toc159 (AtToc159, AtToc132, AtToc120 and AtToc90) (Bauer et al., 2000; Hiltbrunner et al., 2004; Ivanova et al., 2004; Kubis et al., 2004) and Toc75 (AtToc75-III, AtToc75-IV, AtToc75-I and AtToc75V/AtOep80) (Baldwin et al., 2005). The presence of homologues of the members of the core TOC complex might allow remodelling of the import

machinery in accordance to the biochemical requirements of the plastid dependent on the developmental stage.

1.5. The Toc159 and Toc34 GTPase receptor families

Precursor proteins recognition and translocation initiation at the TOC complex are mediated by two receptors; Toc159 and Toc34. Deriving from a common ancestor, Toc34 and its homologues were the first to branch off, and later, after the addendum of an extension at the N – terminus, Toc90 and the larger TOC receptors Toc120, Toc132 and Toc159 emerged (Hofmann & Theg, 2003). Interestingly, both of the receptors, Toc159 and Toc34 belong to a distinct plant family GTPases of eukaryotic origin (Reumann et al., 2005). Alignment – guided secondary structural analysis revealed that the core of the TOC receptors G – domain resembles the basic structure of other TRAFAC (translation factor related) family members, such as Ras GTPases (Aronsson & Jarvis, 2011). This analysis places the TOC receptors specifically in the septin – like superfamily within the TRAFAC class GTPases (Aronsson & Jarvis, 2011).

In general, members of the Toc159 / Toc34 superfamily share a typical domain structure organization (Figure 4). Toc34 is mostly composed of its GTPase binding domain (G–domain) and is anchored to the outer envelope of chloroplasts by a short hydrophobic patch at the C - terminus of the protein (Kessler, et al., 1994; Seedorf, et al., 1995). Toc159, on the other hand, can be subdivided into three functional domains: an acidic amino acid extension at the N–terminus (A–domain), a central GTP - binding domain (G–domain) and a C–terminal membrane anchoring domain of 52 kDa (M–domain) (Muckel & Soll, 1996). Both A - and G - domain are exposed to the cytosol whilst the M–domain substantiates the membrane anchoring (Hirsch et al., 1994). While the G–domain and M–domain between the different homologues exhibit a relatively high homology to each other, their A–domain, on the other hand, reveal a fairly low sequence conservation (Ivanova, et al., 2004). The N–terminal A–domain also represents the most variable region of the Toc159 receptor families both in length and primary structure. This structural remodelling (i.e. domain enlargement and negative charge introduction) suggests functional specialization and optimization, yet the exact functional relevance of the A–domain thus far remains largely in the shadows. However, a regulatory role of the for the A-domain during import of precursor proteins into chloroplasts has been

recently suggested (Inoue et al., 2010). A detailed description of the proposed functions of the A-domain of Toc159 homologues will be discussed in later chapter (see section 1.5.3 below).

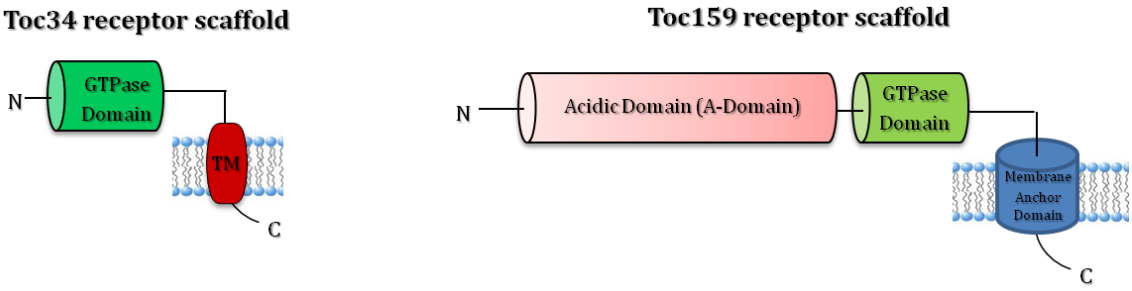


Figure 4 | General scheme of the molecular scaffold of Toc34 and Toc159 receptor families. The receptor constituents of the TOC complex fall into two subgroups; Toc159 homologues and Toc34 homologues, respectively. Toc159 receptors are tripartite proteins, which consist of the acidic domain (A, pink), the GTPase domain (G, shades of green) and membrane anchor domain (M, blue). Toc34 receptors consist mainly of the cytosolic GTPase domain (G, shades of dark green) and are anchored to the chloroplast outer envelopes via the COOH – terminal transmembrane helixes (TM, red) (Adapted from(Chang et al., 2012).

1.5.1. The Toc34 receptor family

The precursor of Toc34 receptor is synthesized without a cleavable transit peptide. It belongs to the class of TA (Tail-Anchored proteins) that require an AKR2A (Bae et al., 2008) for insertion into the outer membrane lipid bilayer of the chloroplast in its GTP-bound form (Qbadou et al., 2003). Similar to the homologues of Toc159, the two homologues of Toc34 in *Arabidopsis* (AtToc33 and AtToc34) also display different developmental expression profiles. *AtToc33* is expressed at very high levels in young, rapidly expanding photosynthetic tissues, whereas *AtToc34* is expressed at low levels throughout development (Gutensohn, et al., 2000; Jarvis, 2008; Kubis, et al., 2004). The knockout mutant of AtToc33 (*ppi*1: plastid protein import 1) does not demonstrate a strong phenotype as observed for the knockout mutant of AtToc159 (Jarvis, et al., 1998): it appears uniformly pale during the first two weeks of development, however these phenotypic defects were restored to that of wild-type in mature plants, strongly hinting a role in early chloroplast biogenesis, presumably during the expansion of cotyledons (Bauer et al., 2001; Gutensohn, et al., 2000). The knock-out mutant of AtToc34 (*ppi*3: plastid protein import 3), on the other hand, has no visible phenotype apart from delayed root growth, clearly suggesting a role of AtToc34 in root plastids biogenesis. However, both AtToc33 and AtToc34 are indispensable for the early development of

plastid biogenesis as the double mutant *ppi*1/*ppi*3 is embryo lethal (Constan, et al., 2004).

1.5.2. The Toc159 receptor family

Toc159 was initially identified as an 86 kDa fragment due to its high susceptibility to proteolysis (Bolter et al., 1998; Kessler, et al., 1994; Waegemann & Soll, 1995). Toc159 has been proposed to be involved in initial binding of the precursor proteins (Chen et al., 2000). A detailed understanding of the targeting and insertion of this important group of proteins is definitely on the horizon. Correct sub-cellular sorting and membrane anchoring of Toc159 relies on the vital information which is invariably found at the C–terminal segment of the protein (Muckel & Soll, 1996). For the initial docking and proper integration of Toc159 into the TOC complex, intrinsic Toc159 GTPase activities as well as the interaction of its M–domain with both the G–domain of Toc34 and Toc75 are essential (Bauer et al., 2002; Schleiff et al., 2002; Wallas et al., 2003). Characterisation of the T-DNA insertion mutant of *Arabidopsis* Toc159 (AtToc159), *ppi2* (plastid protein import 2) mutant, showed that the differentiation of proplastids into chloroplasts is arrested, resulting in an albino phenotype (Bauer, et al., 2000): in other words, the plant cannot develop photoautotrophically. The M-domain, the function of which is to anchor the protein in the outer membrane and to assemble the TOC core complex, was demonstrated to partially complement the preproteins import defect in *ppi2* mutant (K. H. Lee et al., 2003). The accumulation and expression level of photosynthesis-related proteins were drastically decreased. This however did not appear to be the case for non-photosynthetic plastid proteins. This observation led to the proposal that proteins of this class are imported by other members of the large TOC GTPases family, namely AtToc132 and AtToc120 (Bauer, et al., 2000; Kubis, et al., 2004). These different receptors were indeed assembled into different structurally distinct translocation complexes that comprise of either: AtToc159/33/75 or AtToc132/120/34/75, which functions are reflected by their individual receptor diversities (Bauer, et al., 2000; Ivanova, et al., 2004; Kubis, et al., 2004) (Figure 5).

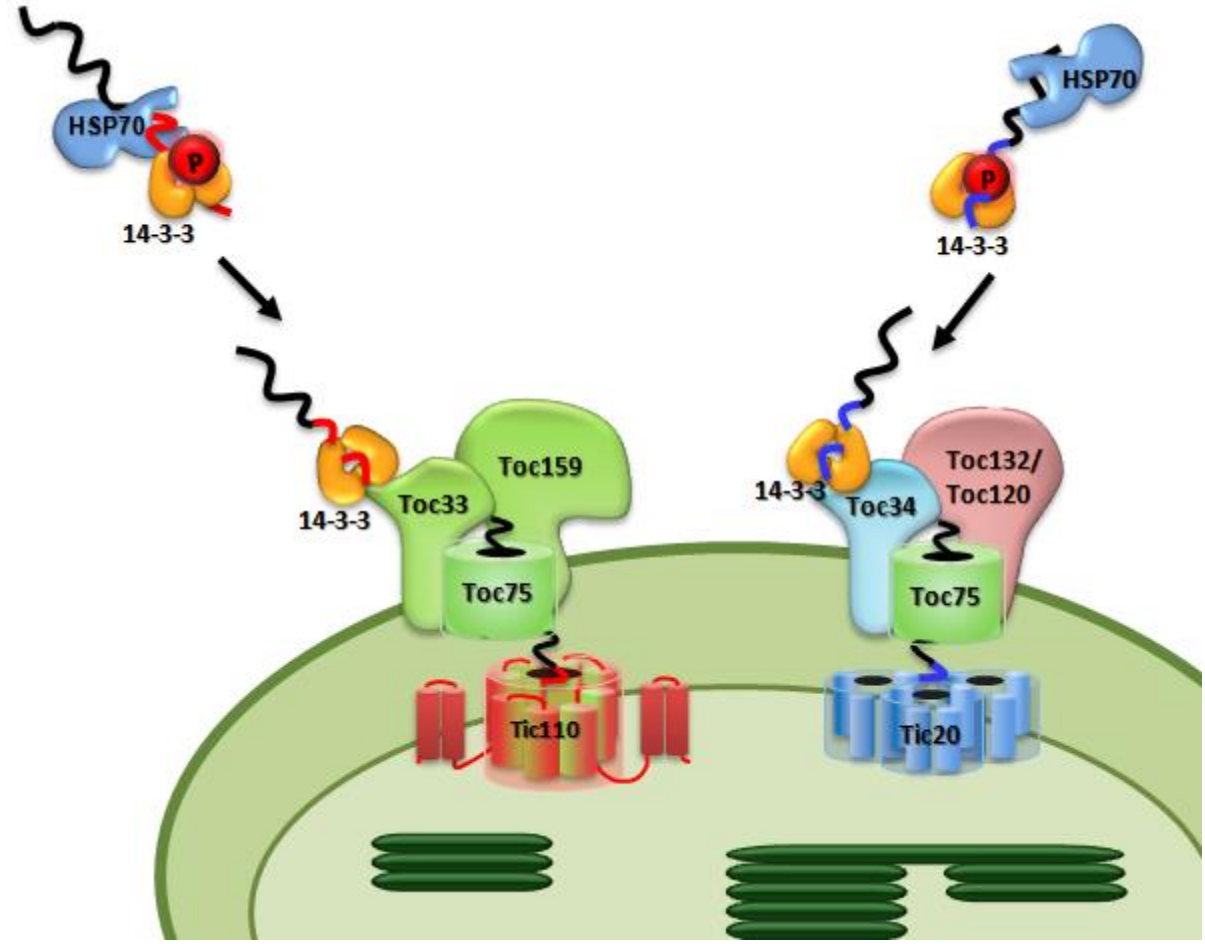


Figure 5 | Hypothetical model of two distinct core TOC complexes and two independent channels at the TIC translocons. The Toc159 and Toc34 family members assemble into distinct translocons with the Toc75 channel. The different TOC receptor isoforms mediate the recognition of distinct classes of nucleus-encoded preproteins (depicted in red and blue recognition signal) to maintain the proper levels of functional classes of proteins that are required for the biogenesis and homeostasis of the organelle (Inoue, et al., 2010; Ivanova, et al., 2004). Also depicted is the hypothetical model of Tic20 and Tic110 channels in the IE of chloroplasts. After tranlocating through OE, preproteins are imported either via Tic110 or Tic20 through the IE. Tic110 is thought to form a homodimer with a total of eight amphipathic transmembrane helices forming the translocation channel and four hydrophobic a-helices involved in the insertion into the membrane (Balsera, et al., 2009; Lubeck, Soll, Akita, Nielsen, & Keegstra, 1996). The proposed Tic20 channel is depicted as a homo-oligomer with a proposed molecular mass > 700 kD but only three molecules are drawn for simplicity. Due to the low overall abundance of Tic20 (Kovacs-Bogdan, et al., 2011), it might be responsible for the import of a smaller and distinct subset of precursor proteins

1.5.3. The ambiguous role of the A-domain of Toc159 receptor familiy

As aforementioned, the function of the A–domain of Toc159 is still under investigation owing largely to its dispensable function in chloroplast biogenesis (K. H. Lee, et al., 2003). It is highly improbable that the A–domain has evolved and was conserved throughout evolution without a functional significance. In all likelihood, the A–domain of the Toc159 homologues represent major determinants of distinct pathways for protein import into plastids (Ivanova, et al., 2004). Indeed the selectivity of

the different receptors (AtToc159, AtToc132 and AtToc132) towards precursor proteins were altered when their respective A–domain was swapped (Inoue, et al., 2010). Similarly, an atToc132GM overexpressor line (a constructs lacking the A–domain) was able to partially complement the import defects in *ppi2* mutant, but in a non–discriminating fashion towards the different classes of precursor proteins (Inoue, et al., 2010). These observations clearly suggest that the A–domain of Toc159 receptors families confers a certain degree of selectivity to the distinct TOC core complexes. Further, the isolated A–domain behaves as an intrinsically disordered protein (Richardson et al., 2009)(Figure 6). This places them in the category of natively unstructured proteins (Hernandez Torres et al., 2007). Many disordered regions are associated with protein–protein interaction and surprisingly implicated in an array of regulatory functions in eukaryotic cells (i.e. control of cell cycle and the regulation of transcription and translation) (Dyson & Wright, 2005). In agreement with the concept that reversible protein phosphorylation is central to the regulation of most aspects of cell function (Johnson, 2009), many disordered regions present in proteins are indeed regulated by phosphorylation (Dyson & Wright, 2005). Phosphoproteomic profiling of *Arabidopsis thaliana* proteins from several independent studies revealed that the A–domain as well as the full–length Toc159 are phosphorylated (de la Fuente van Bentem et al., 2008; Reiland et al., 2009). Similarly, cell fractionation followed by *in vitro* phospho–specific staining further demonstrated that full–length Toc159 and the free A–domain were indeed both phospho–proteins (Agne et al., 2010). As such, the regulation of the A–domain via phosphorylation is not surprising as it coincides nicely with the reported phospho–regulation of intrinsically disordered proteins (Dyson & Wright, 2005; Johnson, 2009).

Recent advances have demonstrated that sub-groups of transit peptides contain distinct motifs that could alter their import efficiency and receptor specificity (D. W. Lee, et al., 2008; D. W. Lee, et al., 2009b). Therefore, the finding that the Toc159 family A–domains are natively unstructured proteins is highly significant. Generally, many natively unstructured proteins possess a large surface area under physiological conditions (Dyson & Wright, 2005). The predominant unordered structure of the A–domain as well as its 50% coverage of the total length of the protein within the Toc159 family (with exception of AtToc90) makes them a perfect platform for interaction with several

binding partners simultaneously (Dyson & Wright, 2005). In addition, many natively unstructured proteins undergo transitions to a more stable secondary or tertiary structure upon binding to their target proteins (Dyson & Wright, 2005). Hence, the presence of distinct targeting motifs in the different classes of precursor proteins coupled with the subsequent induced subtle conformational changes of the A–domains may reflect the reported differential recognition between the different Toc159 receptors. Taken together, these findings strongly hint at a complex regulation of A-domain function that is important for the maintenance of the precursor protein selectivity at the TOC translocons.

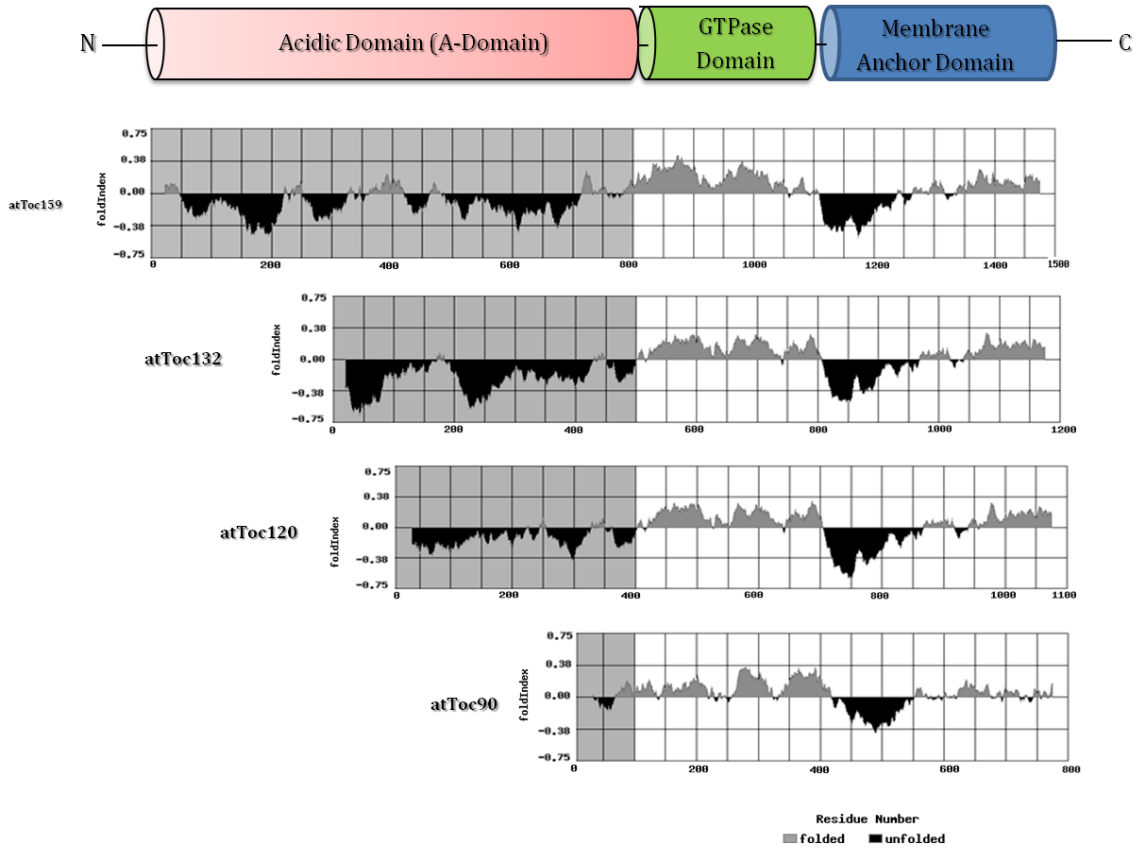


Figure 6 | The A – domains of the Toc159 receptors are predicted to be predominantly unstructured. FoldIndex (Prilusky et al., 2005) is used to predict the intrinsic disordered region of the Toc159 family A – domain. The regions predicted to be disordered are shaded in grey. (Adapted from Chang, et al., 2012).

1.6. The TIC complex

Translocation of nuclear–encoded precursor proteins into chloroplasts also requires the passage through the inner membrane (IE), a process which is facilitated by the TIC complex. In most cases, precursor proteins are thought to be translocated simultaneously through both TOC/TIC complexes. Some components of the TIC complex have been identified and extensively characterized throughout the years; however, the question regarding the precise nature of the TIC channel remains enigmatic. Three conserved membrane–spanning proteins; Tic110, Tic20 and Tic21, were proposed as candidates for inner membrane translocation channel. The latter has been proposed by Teng and co – workers (2006) as the third potential protein conducting channel at the inner membrane of chloroplasts. This notion however, has been controversially discussed in another study, where the reported Tic21 (Teng et al., 2006) most likely represents an ancient metal permease (PIC1), which regulates iron uptake and metal homeostasis in chloroplast and not a protein conducting channel (Duy et al., 2007). Despite the occurrence of several protein conducting channel candidates for the inner membrane of chloroplast, considerable lines of evidence clearly pinpoint Tic110 as the central subunit of the TIC complex, forming a high conductance cation selective channel (Balsera et al., 2009; Heins et al., 2002). Electrophysiology measurements indicate a pore size of 1.7 nm, similar to that of Toc75. Tic110 is encoded by a single – gene copy in *Arabidopsis thaliana* and is constitutively expressed in all tissues, indicating an indispensable role in plastid biogenesis (Inaba et al., 2005). Homozygous T–DNA insertion lines of Tic110 are embryo lethal, further establishing the role of Tic110 in plant viability. Additionally, Tic110 also contains binding sites for stromal Hsp93 and Cpn60 (Inaba et al., 2003; Jackson et al., 1998; Kessler & Blobel, 1996). Both chaperones function as part of the import motor, providing a driving force for translocation, as well as folding of the imported proteins in the stroma. This also accounts for the additional energy expenditure for the translocation of proteins across the IE. The evidence of more than one TIC channel constituents would only lead to a hypothesis that the TIC complex comprises of at least two translocation channels: Tic110 as the core translocation channel while Tic20 forms a distinct channel (Kovacs-Bogdan et al., 2011), independent of Tic110 (Figure 3). It has been proposed that both translocation channels might involved in translocation of different subsets of proteins, mirroring the translocation system at the inner membrane of mitochondria where Tim22 and Tim23 each forms a

distinct translocation channel, responsible for importing different sets of proteins (Mokranjac & Neupert, 2010).

1.7. Diversities of novel protein import pathways

1.7.1. Targeting to the chloroplast through 'non – canonical' transit peptides

The canonical transport of proteins possessing an N–terminal cleavable transit peptide through the TOC/TIC machinery is characteristic for the majority of chloroplast proteins. This is particularly true for stromal and thylakoid proteins. However, not all chloroplast proteins are synthesized with a cleavable transit peptide. Several outer envelope proteins, e.g. OEP7, OEP16, OEP21, OEP24 and OEP37, and most TOC proteins lack a cleavable transit peptide. They are instead directed to the outer envelope membrane by intrinsic targeting information located either within or adjacent to their hydrophobic transmembrane domains (TDMs) (Y. J. Lee et al., 2001). The assembly into the outer envelope membrane occurs spontaneously from the cytosolic side, independent of proteinaceous components (Keegstra & Cline, 1999; Qbadou, et al., 2003; Stengel et al., 2007). Similar features have been observed for two inner envelope proteins; the chloroplast envelope quinone oxidoreductase homologue (AtQORH) and Tic32. Both are targeted to the chloroplast with an intrinsic targeting sequence and the translocation process is not mediated by the standard TOC/TIC machinery (Miras et al., 2002; Miras et al., 2007; Nada & Soll, 2004) (Figure 6). However, in contrast with Tic32, the N-terminus of AtQORH is not required for targeting. Instead, ~40 central residues are crucial for this process (Miras, et al., 2007).

1.7.2. Interaction with the endomembrane system

Close association between the endoplasmic reticulum (ER) and the chloroplast envelope have been established several years ago (Franssen et al., 2011). An additional import pathway was suggested for glycoproteins and proteins with ER targeting signals that contain a signal peptide for the secretory pathway, but were nevertheless found to be localized in the chloroplast. While plastid protein transport through the ER is common in organisms with complex plastids containing more than two envelope membranes (Kovacs-Bogdan, et al., 2010), it was only recently shown to exist in angiosperms. These proteins (e.g. the carbonic anhydrase 1, CAH1 and nucleotide pyrophosphatase/ phosphodiesterase 1, NPP1) first seem to use their signal peptide to

enter the ER, before they are subsequently transported to the Golgi apparatus and finally to the chloroplast with the help of a vesicle transport system (Nanjo et al., 2006; Villarejo et al., 2005)(Figure 7). However, it remains elusive how these proteins are translocated into the chloroplast: it was hypothesized that substrates of this pathway are first released the IMS by vesicle fusion with the OE, from where they finally reach the stroma via the TIC complex or an unknown translocon. Alternatively, translocation may involve further vesicle formation at the IE membrane itself. However, there is no direct proof for any of these possibilities so far.

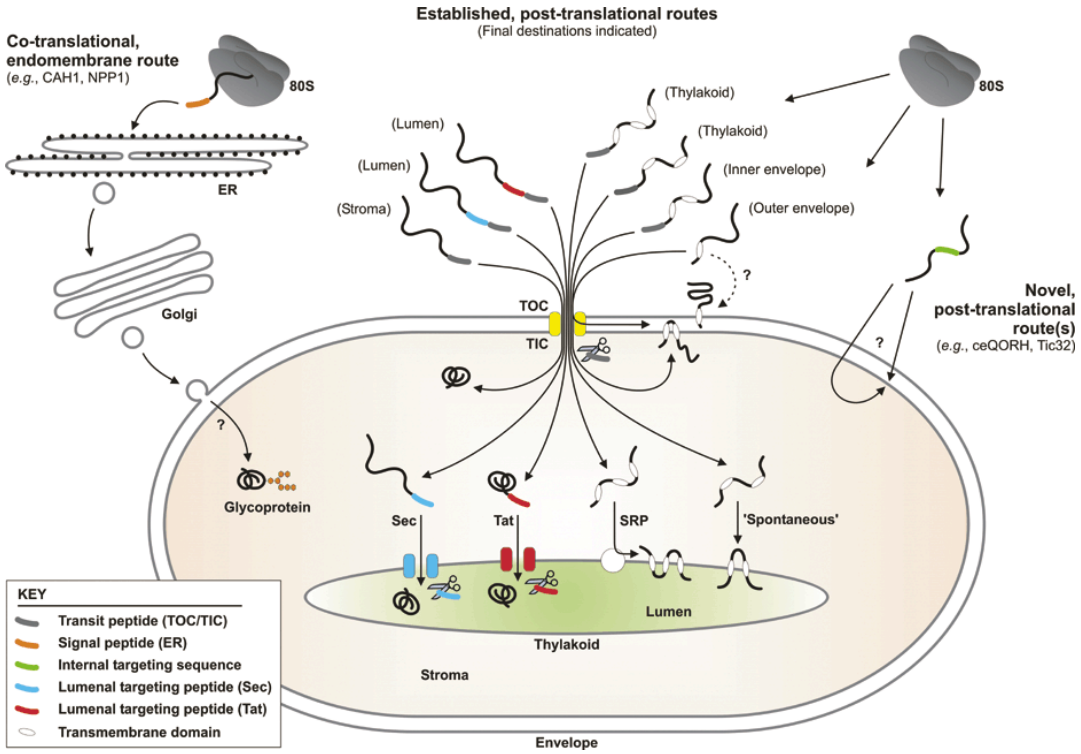


Figure 7 | An overview of protein targeting pathways to and within chloroplasts. At least five pathways for targeting nucleus-encoded proteins to chloroplasts have been described. The majority of the precursor proteins are targeted to and translocated into the chloroplast through the general TOC / TIC import machinery. Proteins destined to the stroma fold into their native conformation, otherwise they can be further targeted to the thylakoids (via the prokaryotic targeting pathways e.g. Sec, Tat, SRP or spontaneous insertion pathway). Many outer envelope proteins lack a transit peptide and follow the OM pathway(outer envelope membrane pathway, dotted arrow). Inner envelope proteins are either laterally released into the membrane from the TIC complex via their 'stop - transfer' signal, or they are re-inserted into the inner membrane after translocation into the stroma. TOC / TIC - independent pathways include a yet uncharacterized import pathways for some inner envelope proteins with intrinsic targeting signals and vesicle transfer to the chloroplast via the endoplasmic reticulum (ER)- and Golgi system. (Adapted from Jarvis, 2008).

1.8. Aims of this work

1.8.1 Novel import route for the 'non–canonical' chloroplast proteins

Successful translocation of nuclear-encoded chloroplast proteins across the chloroplast envelope membrane requires the well coordinated action of multiple proteinaceous components that comprise the TOC/TIC translocons. For many years, all proteins destined to the internal chloroplast compartments were believed to possess an N-terminal chloroplast targeting sequence (also known as the transit peptide), and to engage the TOC/TIC machinery. Recent studies of the *Arabidopsis thaliana* chloroplast proteome revealed, however, the existence of several 'non-canonical' chloroplast proteins, which enter the chloroplast internal compartments in an TOC/TIC-independent manner via their internal non-cleavable targeting sequences (Miras, et al., 2002; Miras, et al., 2007; Nada & Soll, 2004). While the knowledge about the TOC/TIC machinery is relatively detailed, much less is known about the components of the so-called 'non-canonical' translocation machinery. The finding that Tic32 did not compete with AtQORH for import (Miras, et al., 2007), implies that both proteins use different import pathways and that at least two additional yet unknown protein import pathways exist. A better understanding of the 'non-canonical' complex will provide essential insight into the complex nature of the mechanisms of protein trafficking into the chloroplast. The initial aim of the present work was, therefore, to elucidate the functional components that are involved in the import of this special class of 'non-canonical' chloroplast precursor proteins.

1.8.2 Multiple import pathways of the distinct TOC translocons

Multiple structural and functionally distinct TOC core complexes are mainly accounted by the functional selectivity of members of the Toc159 and Toc34 families in *Arabidopsis*. Indeed, both Toc132 and Toc120 were found to form a single TOC complex together, distinct from Toc159 (Ivanova, et al., 2004). While atToc33 co–immunoprecipitates predominantly with AtToc159, AtToc34 forms a complex together with AtToc132/AtToc120 (Ivanova, et al., 2004). This observation led to the notion that the core TOC complex in *Arabidopsis* comprises either: AtToc159/33/75 or AtToc132/120/34/75, whose functions are reflected by their individual receptor diversities (Ivanova, et al., 2004). In both complexes, only one functional Toc75

homologue (AtToc75-III) was detected (Ivanova, et al., 2004). These distinct import routes seem to converge at the TIC complex via AtTic110 (Ivanova, et al., 2004). Multiple structural and functionally distinct TOC core complexes have thus far been reported in *Arabidopsis*. Up to date, no experimental data have provided any direct indication for the presence of such TOC sub-complexes in *Pisum sativum* (pea). The identification of a 120–KDa/132–KDa protein, thus, revealed for the first time the existence of two putative homologues of the import receptor Toc159 in *Pisum sativum* (pea, PsToc159). All attempts to heterologously express and purify the full length or A-domain of Toc159 so far have failed. Especially the investigation of biochemical properties of the different TOC sub-complexes depends highly on the availability of antisera that could specifically distinguish the different TOC receptor homologues. Therefore, a second aim of this work concerned with the establishment of an expression system that is capable of producing adequate amounts of pure proteins of the PsToc159 homologues for these downstream applications.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Chemicals

All used chemicals were purchased in high purity from Sigma-Aldrich (Steinheim, Germany), Fluka (Buchs, CH), Roth (Karlsruhe, Germany), Roche (Penzberg, Germany), Merck (Darmstadt, Germany), AppliChem (Darmstadt, Germany) or Serva (Heidelberg, Germany). Radiolabeled amino acids (^{35}S – Methionine) were obtained from Perkin Elmer (Dreieich, Germany).

2.1.2 Molecular weight markers and DNA standards

PstI digested λ-Phage DNA (MBI Fermentas) was used as a molecular size marker for agarose gel electrophoresis. For SDS-PAGE the "MW-SDS-70L" marker from Sigma-Aldrich (Steinheim, Germany) was used.

2.1.3 Oligonucleotides

Name	Orientation	Sequence	Purpose
FSD1	F	CATGGAATTCATGGCTGCTTCAAGTGCT	cloning in pSP65
FSD1ΔN10	F	CATGGAATTCATGGTCCTCAAGCCACCTCCA	cloning in pSP65
FSD1ΔN20	F	CATGGAATTCATGGCTTTGGAGCCGCATATG	cloning in pSP65
FSD1ΔN30	F	CATGGAATTCATGCTGGAGTTTCACTGGGGA	cloning in pSP65
FSD1ΔC10	R	CATGGGATCCTTAGGCACTTACAGCTTCCCAAG	cloning in pSP65
FSD1ΔC20	R	CATGGGATCCTTAGGTCATGAATGTCTTTATGT AATC	cloning in pSP65
FSD1ΔC30	R	CATGGGATCCTTATCGGTTCTGGAAGTCAAGG	cloning in pSP65
FSD1	R	CATGGGATCCTTAAGCAGAAGCAGCCTTGGC	cloning in pSP65
FSD1.SpeI	R	CATGACTAGTATGGCTGCTTCAAGTGCTGTC	cloning in pOL
FSD1_.SalI	R	CATGGTCGACGAGCAGAAGCAGCCTTGGCGGC	cloning in pOL
FSD1(1-6).SpeI	F	CTAGTATGGCTGCTTCAAGTGCTCG	cloning in pOL
FSD1(1-6).SalI	R	TCGACGAGCACTTGAAGCAGCCATA	cloning in pOL
FSD1(1-10).SpeI	F	CTAGTATGGCTGCTTCAAGTGCTGTCACCGCAA ACCG	cloning in pOL
FSD1(1-10).SalI	R	TCGACGGTTTGCGGTGACAGCACTTGAAGCAGC CATA	cloning in pOL

FSD1 (1-20).SpeI	F	CTAGTATGGCTGCTTCAAGTGCTGTCACCGCAA ACTACGTCCTCAAGCCACCTCCATTCGCACTGCG	cloning in pOL
FSD1 (1-20).SalI	R	TCGACGCAGTGCGAATGGAGGTGGCTTGAGGAC GTAGTTTGCGGGACAGCACTTGAAGCAGCCATC	cloning in pOL
FSD1 (1-30).SpeI	F	CATGACTAGTATGGCTGCTTCAAGTGCTGTC	cloning in pOL
FSD1 (1-30).SalI	R	CATGGTCGACGTTGTTTGCTCATATGCG	cloning in pOL
FSD1.NcoI	F	CATGCCATGGATGGCTGCTTCAAG	cloning in pET21d
EcoRI_TP-pSSU	F	CAT GGAATTCATGGCTTCCTCTATGCTC	generation of chimeric constructs
TP-pSSU	R	GCTTGTGATGGAAGTAATGTCGTTGTTAGC	generation of chimeric constructs
3'TP-pSSU-FSD1	F	ACTTCCATCACAAGCATGGCTGCTTCAAGTG	generation of chimeric constructs
FSD1.BamHI	R	CATGGGATCCTTAAGCAGAAGCAGCC	generation of chimeric constructs
SPP	R	GCAGTTAACTCTTCCGCCGTTGCTTG	generation of chimeric constructs
SPP_FSD1	F	AAGAGTTAACTGCATGGCTGCTTCAAGTGC	generation of chimeric constructs
3'FSD1.mSSU	F	CAA GGCTGCTTCTGCTATGCAGGTGTGGCCTC	generation of chimeric constructs
3'FSD1.SPP	F	AGGCTGCTTCTGCTAACGGCGGAAGAG	generation of chimeric constructs
BamHI_FSD1	F	CATGGGATCCATGGCTGCTTCAAGTGCTG	generation of chimeric constructs
FSD1(1-50)	R	CTGTTTCTTGAGGTTGTCCACGTAAGCTC	generation of chimeric constructs
3'FSD1(1-50)_mSSU	F	ACCTCAAGAAACAGATGCAGGTGTGGCCTC	generation of chimeric constructs
SalI_mSSU	R	CATGGTCGACTTAACCGGTGAAGCTTG	generation of chimeric constructs
3'FSD1(1-50).SPP	F	ACC TCAAGAAACAGAACGGCGGAAGAG	generation of chimeric constructs
GSP1(TOC120)	R	GCTGCTGCACCACCCGGGCAGCCGGTTC	5'RACE PCR
NGSP1(TOC120)	R	CAAGAGGGGTGCTAGCAGCAACAGATGA	5'RACE PCR
GSP1(TOC132)	R	CTA CCCTGGGAGCAGGTTCCAACAAAGA	5'RACE PCR
NGSP1(TOC132)	R	AGGAGTGCTAGTAGCAACAGAATGTCCAGATGA G	5'RACE PCR
PsToc120A.EcoRI	F	CATGGAATTCATGGATAATGGTGGGTATGATGA G	cloning in pET21d
PsToc120A.HindIII	R	CATGAAGCTTCTGCTGCACCACCCGGGC	cloning in pET21d
PsToc132A.EcoRI	F	CATGGAATTCATGGTGGATGAGACCATTGACG	cloning in pET21d
PsToc132A.SalI	R	CATGGTCGACATCAAGAGGGGTGCTAGCAG	cloning in pET21d
PsToc159A.EcoRI	F	CATGGA ATTCATGGATTCCCAAACCCTATCTTC	cloning in pET21d
PsToc159A.SalI	R	CAT GGTCGACCTCGACAGAGAAAAGCCTAGATC	cloning in pET21d

2.1.4 Plasmids

Gene	Organism	Vector	Description	Source	Purpose
FSD1	*Arabidopsis*	pSP65	-	this work	*in vitro* transcription and translation
Lip2	*Arabidopsis*	pSP65	-	this work	*in vitro* transcription and translation
AtAnnAt1	*Arabidopsis*	pSP65	-	this work	*in vitro* transcription and translation
AtQORH	*Arabidopsis*	pSP65	-	this work	*in vitro* transcription and translation
PGR5	*Arabidopsis*	pSP65	-	this work	*in vitro* transcription and translation
PYR4	*Arabidopsis*	pSP65	-	this work	*in vitro* transcription and translation
TSP9	*Arabidopsis*	pSP65	-	this work	*in vitro* transcription and translation
Rap38	*Arabidopsis*	pSP65	-	this work	*in vitro* transcription and translation
Alket	*Arabidopsis*	pSP65	-	this work	*in vitro* transcription and translation
FSD1	*Arabidopsis*	pOL	C-terminal GFP	this work	localization
FSD1(1-6)	*Arabidopsis*	pOL	C-terminal GFP	this work	localization
FSD1(1-10)	*Arabidopsis*	pOL	C-terminal GFP	this work	localization
FSD1(1-20)	*Arabidopsis*	pOL	C-terminal GFP	this work	localization
FSD1(1-30)	*Arabidopsis*	pOL	C-terminal GFP	this work	localization
FSD1.ProtA	*Arabidopsis*	pET21d	C-terminal His - and ProtA – tag	this work	expression in *E.coli*
pSSU	*Arabidopsis*	pET21d	C-terminal His-tag	laboratory of Prof. Jürgen Soll	expression in *E.coli*
mSSU	*Arabidopsis*	pET21d	C-terminal His-tag	laboratory of Prof. Jürgen Soll	expression in *E.coli*
Toc34ΔTM	pea	pET21d	C-terminal His-tag	laboratory of Prof. Jürgen Soll	expression in *E.coli*
FSD1ΔN10	*Arabidopsis*	pSP65	-	this work	*in vitro* transcription and translation
FSD1ΔN20	*Arabidopsis*	pSP65	-	this work	*in vitro* transcription and translation
FSD1ΔN30	*Arabidopsis*	pSP65	-	this work	*in vitro* transcription and translation
FSD1ΔC10	*Arabidopsis*	pSP65	-	this work	*in vitro* transcription and translation

FSD1ΔC20	*Arabidopsis*	pSP65	-	this work	*in vitro* transcription and translation
FSD1ΔC30	*Arabidopsis*	pSP65	-	this work	*in vitro* transcription and translation
TP-pSSU-FSD1	*Arabidopsis*	pSP65	-	this work	*in vitro* transcription and translation
TP-pSSU.SPP-FSD1	*Arabidopsis*	pSP65	-	this work	*in vitro* transcription and translation
FSD1-mSSU	*Arabidopsis*	pSP65	-	this work	*in vitro* transcription and translation
FSD1-SPP-mSSU	*Arabidopsis*	pSP65	-	this work	*in vitro* transcription and translation
FSD1(1-50)-mSSU	*Arabidopsis*	pSP65	-	this work	*in vitro* transcription and translation
FSD1(1-50)-SPP-mSSU	*Arabidopsis*	pSP65	-	this work	*in vitro* transcription and translation
PsToc120A	pea	pET21d	C-terminal His-tag	this work	expression in *E.coli*
PsToc132A	pea	pET21d	C-terminal His-tag	this work	expression in *E.coli*

2.1.5 Kits

All the kits were utilized according to the manufacturer instructions.

Kit	Purpose	Source
QIAprep Spin Miniprep Kit	Plasmid DNA isolation	QIAGEN
QIAGEN Plasmid Midi Kit	Plasmid DNA isolation	QIAGEN
QIAGEN Plasmid Maxi Kit	Plasmid DNA isolation	QIAGEN
NucleoSpin Extract II	Purification of DNA	Macherey-Nagel
Rneasy Plant Mini Kit	RNA extraction from plant	QIAGEN
Wheat germ lysate translation kit	in vitro translation	Promega
Reticulocyte lysate translation kit	in vitro translation	Promega
BD SMART™ RACE cDNA Amplification Kit	5'RACE PCR	Clontech
CloneJET™ PCR Cloning Kit	Blunt – end cloning	Fermentas

2.1.6 Enzymes

The enzymes were utilized according to the manufacturer instructions.

Enzyme	Source
Restriction Endonucleases	Fermentas
T4 DNA ligase	Fermentas
Taq DNA polymerase	5 PRIME
Phusion DNA polymerase	New England Bio Labs
SP6 RNA polymerase	Fermentas
BD PowerScript™Reverse Transcriptase	Clontech
RNase-free DNase I	Amersham Biosciences
Lambda phopshatase	Sigma
Cellulase Onozuka R10	Serva
Macerozym R10	Yakult Honsha

2.1.7 Chromatography media

Beads	Purpose	Source
Ni-sepharose fast flow	His-tag purification	GE Healthcare
Protein A Sepharose CL-4B	Immunoprecipitation	GE Healthcare

2.1.8 Bacterial strains

Strain	Organism	Genotype
TOP10	*E.coli*	F-mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str^R) endA1 λ^-
BL21 (DE3) Star	*E.coli*	F- ompT lon $hsdS_B(r_B^- m_B^-)$ gal dcm rne131 (DE3)
BL21 (DE3) pRosetta	*E.coli*	F–ompT $hsdS_B(r_B^- m_B^-)$ gal dcm (DE3) pRARE (Cam^R)
BL21(DE3) pLysS	*E.coli*	F- ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ(DE3) pLysS (Cam^R)
BL21(DE3) pMICO	*E.coli*	F- ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ(DE3) pMICO (Cam^R)
BL21(DE3)	*E.coli*	F- ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ(DE3) [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]

2.1.9 *E.coli* media and plates

LB (Luria-Bertani) medium: 1% Trypton (Difco)

0.5% yeast extract (Difco)

1% NaCl

For agar-plates, 2% of agar was added.

2.1.10 Radioisotopes

^{35}S-Methionine/Cysteine mixture and ^{35}S -Cysteine with specific activity of 1000 Ci/mmol were provided from Amersham Biosciences (Freiburg, Germany).

2.1.11 Antibodies

Primary polyclonal antibodies (α-Toc86, α-Toc75 (III), α-Toc75 (V), α-Toc34, α-Toc64 (III) and α-Tic110) were generated in the laboratory of Prof. Jürgen Soll by injection of purified antigens into rabbit. Antibody α-psToc120A was produced for the purpose of this thesis (see Methods) by Pineda Antibody Service (Berlin, Germany). Secondary antibodies goat anti-rabbit alkaline phosphatase - / horseradish peroxidase conjugate) were obtained from Sigma.

2.1.12 Plant materials and growth conditions

Pisum sativum (sort "Arvica" were ordered from Bayerische Futtersaatbau [Ismaning, Germany]) was grown on vermiculite under 12 h day / 12 h night cycle in a climate chamber, at 20°C.

2.2 Methods

Molecular biology methods

2.2.1 General molecular biology methods

General molecular biological methods like culturing conditions of the bacteria, DNA precipitation, determination of DNA solutions and transformation were performed as described by Sambrook and Russell, 2001. The preparation of transformation competent cells was conducted according to the protocol published by Hanahan and co-worker (Hanahan et al., 1985). Restrictions, ligations and agarose gel electrophoresis were performed as described (Sambrook et al., 1989). Thereby, the reaction conditions were adjusted to the manufacturer's recommendations. *PstI* digested λ-phage – DNA was used as molecular weight standard for gel electrophoresis.

2.2.2 Cloning strategy

Several genes were cloned in the expression vector pET21d+ and transcription vector pSP65, respectively, during the course of this work. For the PCR amplification, the appropriate pairs of primers were used in order to amplify the desired fragments from a template DNA or from cDNA obtained either from *Arabidopsis thaliana* or *Pisum sativum* (see section 3.0.3). Different protocols for PCR reactions were utilized according to the size of the amplification product. In order to obtain compatible sticky ends, the amplified DNA and the destination vector were digested with the appropriate restriction endonucleases. After digestion, the DNA was loaded on an agarose gel and purified using the NucleoSpin Extract II (Macherey-Nagel). The ligation reaction between vector and insert was performed using the enzyme T4 DNA ligase (Fermentas) for at 16°C overnight. The ligation product was transformed in 50 µl of chemical competent *E.coli* TOP10 cells and plated on LB plates with the appropriate antibiotic. Single colonies were inoculated in liquid culture, let grown over night and the plasmid DNA was purified using the QIAprep Spin Miniprep Kit (QIAGEN). The DNA insert was fully sequenced.

To introduce single point mutations, site directed mutagenesis was performed. The whole plasmids were amplified by PCR using the proper pairs of primers carrying the mutation and the protocols for PCR reactions were modified according to the size of the

amplification product. Chimeric constructs were generated by overlap PCR, using appropriate oligonucleotides fusing the desired gene fragments.

2.2.3 Extraction of plant RNA

Five to seven days old plantlets were ground in liquid nitrogen (N_2), and total RNA was isolated as recommended in the manual supplied with the RNA Easy Isolation kit (Qiagen, Hilden, Germany). cDNA was prepared from 1 µg samples of DNase-treated RNA using the BD SMART™ RACE cDNA Amplification Kit (Clontech, Germany). cDNA was diluted 10-fold and the diluted cDNA aliquot can be stored at -20°C up to 3 months.

2.2.4 RT-PCR and 5'-RACE PCR

RT-PCR and 5'-RACE PCR were performed to the manufacturer's instructions (BD SMART™ RACE cDNA Amplification Kit, Clontech, Germany). For amplification purposes, the RACE products were cloned into pJET1.2 / blunt cloning vector (Thermo Fisher Scientific, Germany) and, subsequently, into the expression vector pET21d+ vector with a C-terminal fused hexahistidine-tag.

Biochemical methods

2.2.5 Heterologous protein expression in *E. coli* and purification via Ni2+-NTA matrix

All recombinant proteins used in this work were expressed in *E.coli* BL21 (pLysS) or BL21 (DE3) cells. Cells were grown at 37°C in LB medium in the presence of 100 µg/ml to an OD_{600}=0.6. Expression was induced by addition of 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG), and cells were grown for either for 3 h at 37°C or at 12°C overnight. All proteins were purified via their C-terminal His-tag using Ni^{2+}-NTA Sepharose (GE Healthcare, Munich, Germany) under native conditions and eluted increasing the imidazole concentration up to 250mM imidazole. The proteins were always used fresh, concentrated and buffer exchanged for 50mM Tris/HCl, pH 8.0, 150mM NaCl prior to usage.

For protein purification in inclusion bodies, cells were lysed in lysis buffer (50 mM Tris/HCl pH 8.0, 150 mM NaCl, 5 mM β – mercaptoethanol) and centrifuged for 30 min

at 14000 rpm. The resulting pellet was resuspended in detergent buffer (20 mM Tris/HCl pH 7.5, 1% Deoxycholic acid, 1% Nonidet P40, 200 mM NaCl, and 10 mM β – mercaptoethanol) and centrifuged for 10 minutes at 10000 rpm. The pellet obtained was washed twice with Triton buffer (20 mM Tris/HCl pH 7.5, 0.5% Triton X-100, and 5 mM β – mercaptoethanol) and 2 times in Tris buffer (50 mM Tris/HCl pH 8.0, 10 mM DTT). The inclusion bodies were finally incubated in buffer A (8 M UREA, 50 mM Tris/HCl pH 8.0, 100 mM NaCl, 2 mM β – mercaptoethanol) at RT for 1 h, centrifuged for 10 min at 10000 rpm and the supernatant was incubated with 350 µl Ni-sepharose fast flow (GE Healthcare) for 1 h at RT. The Sepharose was washed twice respectively with washing buffer B (8 M UREA, 50 mM Tris/HCl pH 8.0, 100 mM NaCl, 40mM imidazole, 2 mM β – mercaptoethanol) and buffer C (8 M UREA, 50 mM Tris/HCl pH 8.0, 1M NaCl, 40mM imidazole, 2 mM β – mercaptoethanol). The recombinant proteins were eluted by increasing the imidazole concentration up to 400 mM.

For immunization of rabbits, the protein was dialysed (two times 1 L each) against 10 mM Tris/HCl (pH 7.0), 150 mM NaCl using a dialysis membrane with a MWCO of 14 kDa (Spectra/Por, Spectrum Laboratories, Rancho Dominguez, CA, USA). For CD spectroscopy, see section 3.1.9.

2.2.6 Electroelution of proteins from polyacrylamide gels

The recombinant PsToc120A protein remains inseparable from the crude lysate even after Ni^{2+} affinity chromatography, gel filtration and anion exchange affinity purification. Hence after Ni^{2+}-affinity purification, the eluted proteins were solubilised in Laemmli buffer, boiled for 5 min at 95°C and separated on 10% SDS-PAGE. The overexpressed protein was excised and eluted out from the gel from a self assembled electro-eluter. In brief, the dialysis membranes were equilibrated briefly in protein elution buffer for 1 minute. The gel slice was loaded into the electro-eluter and elution was performed at 25 mA/ dialysis membrane constant current overnight. The sample in the dialysis membrane was transferred into a microfuge tube and dialysed against 50 mM Tris/HCl (pH 7.0), 150 mM NaCl to remove SDS bound to the protein. The precipitated protein, during dialysis, was separated from the soluble fraction after a brief centrifugation at 10,000 g for 10 min. The soluble protein fraction was used for immunization.

2.2.7 Total protein extractions from *Pisum sativum* (pea) leaf

To obtain soluble and total membrane-attached proteins, 50-150 mg of pea fresh weight was harvested was harvested and flash-frozen in liquid N_2. The frozen leaves were thoroughly ground in liquid N_2 by using a pre-cooled mortar and pestle. To obtain soluble and membrane-attached proteins the powder was mixed with 400 µl of urea-buffer (50 M Tris (pH 8.0), 0.2 M EDTA, 6 M urea), incubated for 10 min at RT and centrifuged at maximum speed in a microfuge (10 min, 13,000 x g) and the supernatant was saved. To extract the membrane proteins the pellet was resuspended in 200 µl of SDS-buffer (50 mM Tris-HCl (pH 8.0), 0.2 mM EDTA, 1% SDS), incubated for 15 min at RT, centrifuged again at maximum speed and the supernatant containing total membrane proteins was also saved. For the extraction of soluble and membrane proteins in one step, ground samples are extracted by addition of 750 µl SDS buffer (50 mM Tris (pH 8,0), 0.2 mM EDTA, 1% SDS), followed by incubation at RT and centrifugation as described above. The protein concentration of the supernatants was determined with the Bradford assay.

2.2.8 SDS polyacrylamide gel electrophoresis and western blot

The electrophoretical separation of proteins in denaturing polyacrylamide gels was carried out according to the method of Laemmli, 1970. Separating gels with polyacrylamide concentrations ranging from 10-15% were used. Before being applied to the gel, proteins were solubilized in sample buffer (Laemmli-buffer) and incubated for 20 min at RT. Gels were stained either by Coomassie Brilliant Blue-R250 (Sambrook et al., 1989) or silver-stained using a protocol according to Blum et al., 1987 with modifications.

For immunodetection, proteins were transferred onto either Nitrocellulose or PVDF membranes by semi-dry-blotting in a Trans Blot Cell (BioRad, München, Germany) in blotting buffer (25 mM Tris-HCl (pH 8.2-8.4), 192 mM glycine, 10-20% MeOH, 0.025% SDS) for 1h at 400 mA as described previously (Towbin et al., 1979). Proteins of the size markers were either stained with Ponceau S solution (nitrocellulose) or amido black solution (PVDF). Membranes with bound proteins were first incubated for 30 min in blocking buffer containing skimmed milk powder (2-5% milk powder, 1x TBS, 0.05% Tween 20). Incubation with the primary antibody (diluted in blocking buffer 1:250-

1:4000, depending on the antibody) was done for 2-3h at RT or overnight at 4°C. Non-bound antibody was removed from the membrane by washing for 3x10 min in TBS-T (1x TBS with 0.1% Tween 20). The secondary antibody was selected according to the desired method of visualization (see section 3.1.6).

2.2.9 BN-PAGE with isolated chloroplast

Blue native gel electrophoresis was carried out as described in previously (Schägger and von Jagow, 1991 and Wittig et al., 2006) with the following modifications: Chloroplasts (equivalent to 4-50 µg of chlorophyll (Chl) were solubilized in 50 mM Bis-Tris/HCl, (pH 7.0), 750 mM 6-aminocaproic acid, 1% *n*-dodecyl β-D-maltoside (DOMA). After incubation on ice for 15 min, samples were centrifuged at 80,000 rpm for 10 min at 4°C. The supernatant was supplemented with 0.1 vol. of a Coomassie Blue solution (5% Coomassie Brilliant Blue G-250, 750 mM 6-aminocaproic acid) and loaded on a 5-12% (w/v) polyacrylamide gradient gel. Electrophoresis was carried out at increasing voltage (stacking gel: 100 V maximum; separating gel: 15 mA/400 V maximum for a 12 x 14 cm gel, 8 mA maximum for a 6 x 8 cm gel) at 4°C. The cathode buffer contained 0.02% dye and was replaced by buffer lacking dye after approximately one-third of the electrophoresis run.

2.2.10 Immunoblot development

For colorimetric reaction with the alkaline phosphatase (AP) system, the secondary antibody (goat anti-rabbit IgG (whole molecule)-AP conjugated; Sigma-Aldrich Chemie GmbH, Taufkirchen) was applied to the membrane for 1 h. After removal of unbound secondary antibody by washing (3x10 min in wash buffer) and a final wash in western developer (105.7 mM Tris-HCl (pH 9.5), 100 mM NaCl, 50 mM MgCl2,) the immune – reaction was visualized by incubation with western developer containing 66 µl/10ml of NBT (4-Nitro blue tetrazolium chloride, 50 mg/ml in 100% dimethylformamid) and 66 µl/10ml of BCIP (5- Bromo-4-chloro-3-indolyl-phosphate in 100% dimethylformamid). The reaction was stopped by washing in ~5mM EDTA.

For the chemiluminescent method of protein detection (ECL), HRP-conjugated goat anti – rabbit antibody as used as the secondary antibody. Proteins were visualized either with Pierce ECL Western Blotting Substrate (Thermo Fisher Scientific Inc., Rockford,

USA) or according to the following protocol: Solution 1 (100 mM Tris-HCl (pH 8.5), 1% (w/v) luminol, 0.44% (w/v) coomaric acid) and solution 2 (100 mM Tris-HCl (pH 8.5), 0.018% (v/v) H_2O_2) were mixed in a 1:1 ratio and added to the blot membrane (~700 µl per small gel). After incubation for 1 min at RT (in the dark) the solution was removed and the luminescence detected with a film (Super RX Fuji Medical X – ray film; FUJIFILM Corporation, Germany).

2.2.11 Co-immunoprecipitation

Outer envelope vesicles were solubilized by incubation with 1.5% DeMa, 25 mM HEPES/KOH pH 7.6, 150 mM NaCl for 5 min at 20°C. Undissolved particles were removed (100,000 × g, 10 min, 4°C) and the supernatant was diluted 10 times in IP buffer (25 mM HEPES/KOH pH 7.6, 150 mM NaCl, 0.2% DeMa). After the addition of 15µl antiserum to the mixture, the sample was incubated for 1 h at 4°C, followed by incubation with 50 µl in IP buffer pre equilibrated protein-A sepharose (Amersham Bioscience, Freiburg, Germany) for 1 h at 4°C. After washing with IP buffer, the bound protein was eluted by cooking in SDS – PAGE loading buffer. The eluted fractions were separated by SDS-PAGE and subsequently transferred to a nitrocellulose membrane and immune – decorated with the indicated antisera.

2.2.12 Chemical cross-linking and mass spectrometric (MS) analysis

After import, chloroplasts were re-isolated on a Percoll cushion, washed and chemical cross-linking was performed by incubation of chloroplasts with 0.5 mM dithiobis succinimdylproprionate (DSP) in 330 mM sorbit, 50 mM HEPES/KOH pH 7.6, 0.5 mM CaCl2, for 15 minutes at 4ºC. The reaction was stopped by the addition of 125 mM glycine and further incubation at 4ºC for 15 minutes. Chloroplasts were washed twice in 330 mM sorbit, 50 mM HEPES/KOH pH 7.6, 0.5 mM CaCl2 and finally lysed in hypotonic buffer (20 mM HEPES/KOH pH 7.6, 5 mM EDTA) for 30 minutes on ice. A total membrane fraction was recovered by centrifugation at 256,000xg for 30 minutes. Membranes were solubilized in 1% SDS (w/v), 25 mM HEPES/KOH pH 7.6, 150 mM NaCl, diluted tenfold in the above buffer in the absence of SDS, centrifuged for 2 minutes at 20,000xg and the supernatant was used for immunoprecipitation with the antisera against Toc75(III), Toc75(V), Toc34, Tic110 and OEP16. Antisera for the previously indicated proteins were incubated with membranes and 0.5% egg albumin, rotating for

1 hour at RT, followed by purification on Protein A – Sepharose. The affinity matrix was washed 3 times with 10 bead-volumes of the mentioned buffer before the elution with Laemmli sample buffer in the presence of b-mercaptoethanol to split the crosslink products.

Coomassie- or silver-stained protein spots were cut from SDS-PAGE gels and send for identification to the "Zentrallabor für Proteinanalytik" (ZfP, Adolf-Butenandt-Institut, LMU München). There, tryptic peptides were detected either by Peptide Mass Fingerprint (MALDI, **M**atrix **A**ssisted **L**aser **D**esorption/**I**onisation) or LC-MS/MS (**L**iquid **C**hromatography with **MS**) runs. Protein identification was then accomplished with a Mascot software assisted database search. Only hits displaying a threshold score of >=60 were analyzed further.

2.2.13 Peptide array affinity assay

Customized FSD1 peptide arrays were ordered from JPT Peptide technologies. Peptides were synthesized at 5 nmol/spot with acetylated N-termini and covalently bound by C-termini with a polyethelene glycol linker to the cellulose membrane. The recombinant A-domain of PsToc120 (PsToc120A) was analyzed in the affinity arrays. The peptide array was blocked with 0.3% skim milk in 1X TBS buffer for 1h and subsequently incubated with 5µg/ml PsToc120A overnight at 4°C. The binding of the proteins to the peptides was detected after 3h incubation with rabbit anti-PsToc120A (1:500, 0.3% skim milk in 1X TBS) primary antibody and 1h incubation with HRP-conjugated anti-rabbit (1:20000, 0.3% skim milk in 1X TBS) secondary antibody. The membrane was washed three times for 10 min with 0.3% skim milk in 1X TBS after primary and secondary incubation. A negative control was performed by excluding PsToc120A from the incubation protocol. The detection was performed with ECL Plus detection reagents. The intensities of the spots were analyzed with ImageQuant TL 8.1 software (GE healthcare, München)

2.1.14 Circular dichroism for the analysis of protein secondary structure

Far-UV CD spectra were measured on an Aviv 215 spectropolarimeter (Aviv Biomedical). Measurements were performed using rectangular quartz cells with 0.1 cm path length. psToc120A-His was typically measured at a concentration of 0.1mg/ml, in CD buffer (10 mM Tris-HCl pH 8.0), with added components where indicated, unless

stated otherwise within figures. Samples were equilibrated at the indicated temperature for 10 min prior to measurements. Spectra of protein samples and buffer alone were measured with a 0.5 nm/s scanning speed at 0.5 nm intervals, and are an average of four scans. Averaged buffer spectra were subtracted from protein sample spectra and the resultant corrected spectra were smoothed and subsequently converted to mean residue ellipticity using Aviv CDSD software (Aviv Biomedical). Spectra were deconvoluted on the Dichroweb website (Whitmore and Wallace 2004) using the K2D method (Andrade et al. 1993).

Methods in cell biology

2.1.15 *In vitro* transcription and translation

Transcription of linearized plasmids was carried out as previously described (Firlej-Kwoka, Strittmatter, Soll, & Bolter, 2008). Translation was carried out using the Flexi Rabbit Reticulocyte Lysate System from Promega (Madison, USA) following the manufacturers protocol in presence of ^{35}S-methionine for radioactive labelling.

2.1.16 Isolation of intact chloroplast from *Pisum sativum* (pea)

For isolation of intact chloroplasts (Keegstra and Youssif, 1986; Waegemann and Soll, 1995) pea seedlings grown for 9-11 days on vermiculite, under 12/12 hours dark/light cycle were used. All procedures were carried out at 4°C. About 200 g of pea leaves were grinded in a kitchen blender in approximately 300 ml isolation medium (330 mM sorbit, 20 mM MOPS, 13 mM Tris, 3 mM $MgCl_2$, 0.1% (w/v) BSA) and filtered through four layers of mull and one layer of gauze (30 µm pore size). The filtrate was centrifuged for 1 minute at 3200 rpm and the pellet was gently resuspended in about 1ml wash medium (330 mM sorbit, 50 mM HEPES/KOH, pH 7.6, 3 mM $MgCl_2$). Intact chloroplasts were reisolated via a discontinuous Percoll gradient of 40% and 80% (in 330 mM sorbit, 50 mM HEPES/KOH, pH 7.6) and centrifuged for 5 minutes at 7000 rpm in a swing out rotor. After centrifugation two green bands; intact chloroplasts are found at the bottom band. This band was taken and washed two times, and finally resuspended in a suitable volume of wash medium. Samples of chloroplasts (5µl) were dissolved in 5 ml of 80% acetone and chlorophyll concentration was estimated by measuring the optical density at three wavelengths against the solvent (Arnon, 1949). Chloroplasts were then used for further import experiments.

2.1.17 Treatment of chloroplast and translation products before *in vitro* import assay

ATP depletion from chloroplast and translation product

Before chloroplasts isolation, the peas were left over night in the dark. After the isolation procedure, intact chloroplasts were left on ice in the dark for 30 minutes in order to deplete ATP and therefore allow subsequent import experiments to be carried out with only exogenously added ATP as energy source. To deplete endogenous ATP from *in vitro*

translation product, Micro Bio-Spin Chromatography Columns (Bio-Rad, Hercules, CA, USA) were used according to manufacturer´s recommendations.

Protease pre – treatment of isolated intact chloroplast

Protease treatment of chloroplasts before insertion of radioactively labelled protein was carried out using chloroplasts corresponding to 1 mg chlorophyll, 1 mg thermolysin and 0.5 mM $CaCl_2$. Wash medium (330 mM sorbitol, 50 mM HEPES/KOH, pH 7,6, 3 mM $MgCl_2$) was added to the final volume of 1 ml and the sample was incubated for 30 minutes on ice. To stop the protease reaction, 5 mM EDTA was added. Intact chloroplasts were re-isolated via a discontinuous Percoll gradient containing 5 mM EDTA and washed twice as described before.

2.1.18 *In vitro* import experiments and chloroplast post-treatment

Import of radioactively labelled proteins into intact chloroplast

^{35}S-labelled precursor proteins (translation products) in the maximal amount of 10% (v/v) in the reaction were mixed with freshly prepared intact pea chloroplasts (equivalent to 5 – 10 µg chlorophyll) in import buffer (330 mM sorbit, 50 mM HEPES/KOH pH 7.6, 3 mM $MgCl_2$, 10 mM methionine, 10 mM cysteine, 20 mM potassium gluconate, 10 mM NaHCO3, 2% BSA (w/v)) and up to 3mM ATP in a final volume of 100 µl (Waegemann and Soll, 1995). The import mix was incubated at 25°C for up to 20 minutes, depending on experimental requirements. Chloroplasts were reisolated over a 40% Percoll cushion, washed, and samples were separated by SDS-PAGE. Resulting gels were fluorographed (Bonner and Laskey, 1974) if needed, dried and laid on x-ray sensitive films over night.

Chloroplast post – treatment with thermolysin

To control the efficiency of protein import across the outer envelope of chloroplasts the intact organelles were treated with the protease thermolysin. After import chloroplasts were pelleted at 3200 rpm for 1 minute at 4°C and resuspended in 100 ml digestion buffer (330 mM sorbitol, 50 mM HEPES/KOH pH 7.6, 0.5 mM $CaCl_2$). The addition of thermolysin (2 µg per 10µg of chlorophyll) started the digestion which was performed for 20 minutes on ice. The reaction was stopped by addition of 5 mM EDTA. Chloroplasts

were pelleted and washed in the washing buffer (330 mM sorbitol, 50 mM HEPES/KOH pH 7.6, 5 mM EDTA).

ATP concentration scale

To characterize ATP requirements of proteins imported into isolated intact chloroplasts, radioactively labelled, ATP-depleted translation product was incubated with chloroplasts corresponding to 10 – 15µg chlorophyll in the import mixture (see section 4.4.1) without or with different concentration of ATP: 100, 1000 and 3000 µM. Samples were incubated at 25°C for 8 minutes. Chloroplasts were re-isolated over 40% Percoll cushions and subsequently samples were analyzed by SDS-PAGE. All steps were performed in the dark to minimize the generation of internal ATP via photophosphorylation.

Competition with heterologously expressed proteins

Up to 10 µM of purified competitor protein pSSU, as well as its mature form mSSU (see section 3.1.1) were added to the import mixture. The import experiment was performed as described in section 3.2.4.1. Maximum 10 µg of chlorophyll per reaction was used and the import reaction lasted 5 (pSSU) to 10 or 15 minutes (all other tested transcription constructs) at 25°C.

Inhibition of import with spermine

Chloroplasts corresponding to 10 – 15µg chlorophyll were pre-treated with spermine, a known inhibitor of the import channel Toc75 for 10 minutes and washed twice with wash medium prior to the import assay. The import experiment was then performed as described in section 3.2.4.1.

2.1.19 Stromal processing assay

Intact chloroplasts were isolated as described in the section 3.3.2. Chloroplasts corresponding to 800 mg chlorophyll were pelleted at 1500xg for 1 minute at 4°C and lysed in 1 ml of 5 mM ice-cold HEPES/KOH pH 8.0 for 15 minutes on ice. Samples were centrifuged for 10 minutes at 16,000xg at 4°C and the supernatant was centrifuged again for 30 minutes at 137,000xg, 4°C. In the processing assay the supernatant containing an active stromal processing peptidase was used. Samples containing 15 µl of supernatant,

20 mM HEPES/KOH pH 8.0, and 4-8 µl radioactively labelled translation product were mixed in a total volume of 25 µl and incubated for 90 minutes at 26°C. The reaction was stopped by addition of Laemmli buffer and samples were analyzed by SDS-PAGE.

2.1.20 Isolation and transient transformation into *Arabidopsis thaliana* protoplasts

Mesophyll protoplasts were isolated from leaves of three to four-week-old *Arabidopsis* plants grown on soil. Leaves were cut in small pieces and incubated in the 10 ml enzymes-buffer (1 % Cellulase R10, 0.3 % Macerozyme R10, 40 mM Mannitol, 20 mM KCl, 20 mM MES pH 5.7, 10 mM $CaCl_2$, 0.1 % BSA) in the dark for 90 min at 40 rpm. Protoplasts were released by shaking 1 min at 80 rpm, filtered with a 100 µm Nylon-membrane and centrifuged 2 min at 100 x g. Protoplasts were resuspended in 500 µl MMg buffer (400 mM Mannitol, 15 mM $MgCl_2$, 4 mM MES pH 5.7, Osm 540), separated on a gradient made by 9 ml MSC buffer (10 mM MES, 20 mM $MgCl_2$, 1.2 % sucrose, pH 5.8, Osm 550) and 2 ml MMg buffer via centrifugation 10 min at 75 x g. Intact protoplasts were washed once in W5 buffer (150 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, 2mM MES pH 5.7, Osm 550-580) and resuspended in MMg buffer. 100 µl protoplasts (about 4 x 10^4 protoplasts) were mixed with 10-50 µg DNA (GFP-fusion constructs) and with 110 µl PEG buffer (40 % PEG 4000 (Fluka), 200 mM Mannitol, 100 mM $Ca(NO_3)_2$) and incubated 15 min in dark. Protoplasts were diluted with 500 µl W5 buffer and collected by centrifugation 2 min at 100 x g. Protoplasts were resuspended in 1 ml W5 buffer and incubate at 25°C overnight in dark. GFP fluorescence was observed with a TCS-SP5 confocal laser scanning microscope (Leica, Wetzlar, Germany).

2.1.21 Software, databases and algorithms used in the present study

Table 1 | List of used software tools (freeware)

Name	Version	Author/Reference	URL
Chromas lite	2.0.1	Technelysium Pty Ltd.	http://www.technelysium.com.au/chromas_lite.html
GeneDoc	2.6.002	Nicholas and Nicholas, 1997	http://www.nrbsc.org/gfx/genedoc/
AnnHyb	4.938	Olivier Friard	http://bioinformatics.org/annhyb
LAS AF lite	3.0	Leica Microsystem	http://www.csc.mrc.ac.uk/microscopy/links.tml/

Table 2 | List of used software tools (licensed)

Name	Version	Publisher/Licensor
VectorNTI	9.1.0	Invitrogen
AIDA (Advanced Image Data Analyzer)	4.938	Raytest Isotopenmeßgeräte GmbH

Table 3 | List of used databases and algorithms (available online)

Name	Version/ release	Author/Reference	URL
BLAST		Altschul et al., 1990; Altschul et al., 1997	http://www.ncbi.nlm.nih.gov/BLAST
Dichroweb		Whitmore and Wallace,2004; Whitmore and Wallace, 2008	http://dichroweb.cryst.bbk.ac.uk/ html/home.shtml
ChloroP	1.1	Emanuelsson, O. et al. 1999; Emanuelsson ,O. et al. 2007	http://www.cbs.dtu.dk/services/ChloroP
TargetP	1.1	Emanuelsson, O. et al. 2000	http://www.cbs.dtu.dk/services/TargetP
Predotar	1.03	Small, I. et al. 2004	http://urgi.versailles.inra.fr/predotar/ predotar.html
WoLFPSORT		Horton, P. et al. 2007	http://wolfpsort.seq.cbrc.jp/
iPSORT		Bannai, H. et al. 2002	http://ipsort.hgc.jp/
PSORT		Nakai and Horton, 1999	http://psort.hgc.jp/form.html
PCLR	0.9	Andrew, L. et al. 2001	http://www.andrewschein.com/cgi-bin/pclr/pclr.cgi
ProteinProwler	1.2	Hawkins and Boden, 2006	http://bioinf.scmb.uq.edu.au/pprowler _webapp_1-2/
BaCelLo		Pierleoni, P. et al. 2006	http://gpcr.biocomp.unibo.it/bacello/ pred.htm
Jalview	2.8.1	Waterhouse, A.M., et al. (2009)	http://www.jalview.org/
ClustalW2	2.0	Larkin, M.A. et al. (2007)	http://www.ebi.ac.uk/Tools/msa/ clustalw2

Chapter 3

Results

3.1 Chloroplast proteins without cleavable transit peptides

Earlier systematic studies on the chloroplast proteome have suggested that a relatively large number of proteins that reside in the chloroplasts do not possess canonical targeting information (such proteins lack transit peptides for engagement of the general import pathway) (Kleffmann et al., 2004; Zybailov et al., 2008), and this has led to the elucidation of novel and unusual pathways for chloroplast protein trafficking (Miras et al., 2002; Miras et al., 2007; Nada & Soll, 2004). Whether these novel non-canonical chloroplast proteins that were identified in the previous proteomic studies are targeted to the chloroplast by the TOC/TIC-independent translocation mechanism is still a matter of debate. An intriguing question would be how these proteins cross the outer and inner chloroplast envelopes and which membrane proteins are involved in this complex process. In contrast to comparable studies of the general import machinery, no translocation intermediates have been obtained thus far for the TOC/TIC-independent translocation event. Import intermediates have played a crucial role for the identification and molecular characterization of the general translocation machinery (Li et al., 2012; Schnell & Blobel, 1993; Schnell, Kessler, & Blobel, 1994). Generated by a simple chemical cross-linking assay which immobilized translocating precursor proteins at the chloroplast membranes; these membrane-bound intermediates will remain in close association with components of the chloroplast translocation machinery and stable even after detergent solubilization of the chloroplast membranes. Therefore, this assay could be employed as a 'fishing rod' to bait for the putative proteinaceous components of the TOC/TIC-independent transport machinery.

3.2 *In silico* sub-cellular analysis of the putative 'non-canonical' chloroplast proteins

As a first approach to characterize the molecular identity of the of the TOC/TIC-independent transport pathway, a subset of 9 tentative non-canonical chloroplast proteins were selected as baits to 'fish' for the potential candidates of the TOC/TIC-independent transport machinery. The main criterion for the selection was based on their robust prediction for the lack of a chloroplast transit peptide (cTP). In order to increase the maximum accuracy of the cTP prediction, nine different prediction algorithms were employed. Only proteins that were predicted to lack a chloroplast transit peptide by at least six out of the nine prediction algorithms were selected (Table 4).

Table 4 | Non-canonical chloroplast proteins identified in the proteomic studies and selected for experimental determination of their sub-cellular localization

AGI Code	Description	Lab Annotation	Non-cTP prediction**
At4g25100	Iron-superoxide dismutase 1	FSD1	7
At3g47070	Thylakoid phosphoprotein	TSP9	9
At2g05620	Proton gradient regulator, essential for photoprotection	PGR5	7
At1g09340	Protein of controversially discussed function	Rap38	6
At4g13010	Quinone – oxidoreductase	AtQORH	7
At4g20010	Plastid transcriptionally active 3	PTAC3	8
At5g53580	Putative aldo/keto reductase family	AldKet	6
At4g22930	Dihydroorotase	PYR4	7
At4g31050	Putative lipoate-protein ligase B	Lip2	8
At1g35720	Stress-responsive calcium-dependent membrane-associated annexin	AtAnnAt1	9

** Number of algorithms that predict a non-cp location, namely absence of a cTP. In all, nine different predictors were employed: TargetP, Predotar, ChloroP, Wolfpsort, iPSORT, PSORT; PCLR, BaCelLo, ProteinProwler.

3.3 Several non-canonical chloroplast proteins can be directly demonstrated to be localized in the chloroplasts

Although all the proteins of the test set have a reported chloroplast localization as well as lacked a transit peptide in the previous proteomic studies (Ferro et al., 2003; Kleffmann, et al., 2004; Zybailov, et al., 2008), the sub-cellular localization and processing of these proteins were re-assessed. To experimentally determine the

chloroplast localization and processing of the members of the 9-proteins test set, an *in vitro* import assay was employed. Since the assay is best established in *Pisum sativum* (pea), chloroplasts were isolated from pea plants for all experiments. Briefly, the proteins were radiolabeled by coupled *in vitro* transcription and translation from their respective cDNA clones, and incubated with isolated chloroplasts. The chloroplasts were then recovered by centrifugation through Percoll cushion, and incubated with the protease thermolysin to remove non-imported proteins. Samples taken before and after thermolysin treatment, as well as aliquots of the translation products were then subjected to SDS-PAGE and autoradiography. In the *in vitro* import assay, detection of the protein prior to thermolysin treatment indicates that the protein has attached to the organelle; if the signal persists after thermolysin treatment, one can infer that import has occurred. Moreover, the presence of an additional smaller band is characteristic for post-import cleavage of the cTP. As control of the TOC/TIC and TOC/TIC-independent pathways, the small subunit precursor of ribulose-1,5-bisphosphate carboxylase / oxygenase (pSSU) and inner chloroplast envelope quinone-oxidoreductase homologue (AtQORH) were used.

As Figure 8A and Table 5 illustrate, most of the proteins in the tested set can be successfully imported into chloroplasts. Only AtAnnAt1, PGR5 and Lip2 showed no evidence of chloroplast localization *in vitro*, as there were no protease protected products detectable. Moreover, AtAnnAt1 and PGR5 did not even attach to chloroplasts (Figure 8B; Table 5). Precursors of the imported proteins were found to be processed to a smaller mature protein that was protease resistant. In most cases, the import behaviour was comparable to that of the positive control for the general import pathway, pSSU. Thus, in deviation to the computational prognosis these proteins do feature a cleavable transit peptide, indicating that the applied algorithms have still to be optimized. The exception being FSD1; similar to AtQORH, FSD1 did not undergo a proteolytic maturation during incubation. At a first glance this result suggested that FSD1 was not imported. When chloroplasts were re-isolated after import and treated with thermolysin, however, FSD1 was easily detectable in a protease-resistant, plastid-bound form (Figure 8A). In addition, FSD1 was degraded by the protease in the absence of chloroplasts, validating its protease sensitivity (Figure 8C). These observations revealed that FSD1 had been imported but without detectable proteolytic cleavage, a

result that is consistent with the reported result in the previous proteomic studies. Since FSD1 was the only protein in the representative set that was not processed after import (an indication of a potential novel translocation mechanism), the following experiments were thus be carried out using this protein.

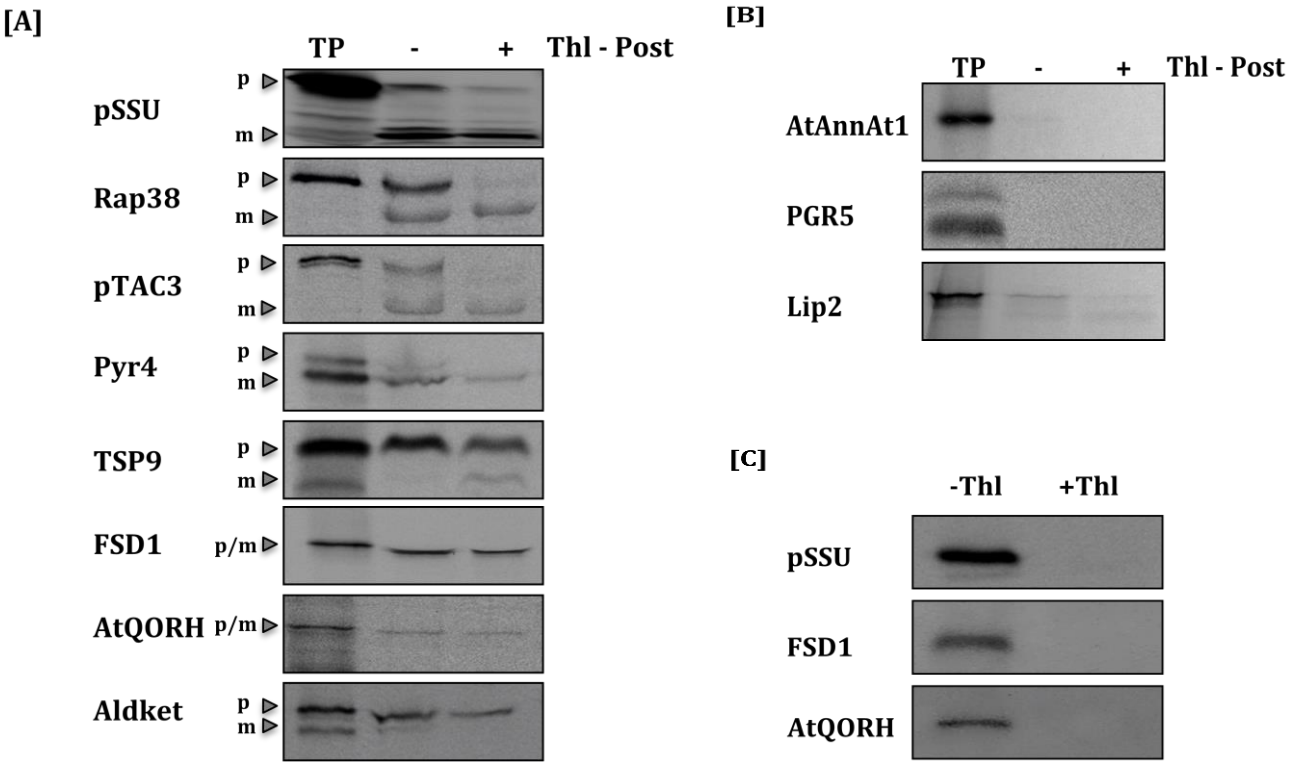


Figure 8 | Sub-cellular localization of the non-canonical chloroplast proteins by means of *in vitro* import assays. The data shown are for proteins with confirmed [A] or non – confirmed [B] chloroplast localization. For each protein *in vitro* import assays are shown. Import assays into intact pea chloroplasts were performed by incubating *in vitro* 35^S-synthesized radiolabeled precursor proteins with chloroplasts corresponding to 10 µg chlorophyll at 25°C in a standard import reaction. Parallel incubations were carried out using the precursors of the small subunit of RuBisCo, pSSU and the quinone – oxidoreductase, AtQORH as a control for the general import pathway and TOC/TIC-independent pathway, respectively. Gels were loaded with 10% of the radiolabeled translation product added to the import reaction mixture ('TP'), and the radiolabeled protein recovered from chloroplasts that had ('+Thl-Post') or had not ('–Thl-Post') been treated with thermolysin after termination of the import reaction; p, precursor; m, mature [C] Degradation of radiolabeled pSSU, FSD1 and AtQORH incubated with thermolysin in the absence of chloroplasts.

Table 5 | Non-canonical chloroplast proteins identified in the proteomic studies and selected for experimental determination of their sub-cellular localization

Protein	In Vitro Import	Cleavable TP
FSD1	●	No
TSP9	●●(?)	Yes(?)
PGR5	-	Localization unclear
Rap38	●●	Yes
AtQORH	●	No
pTAC3	●●	Yes
AldKet	●●	Yes
Pyr4	●●	Yes
Lip2	(●)	Found attached to chloroplast
AtAnnAt1	-	Localization unclear

In vitro import assay:
(●) found attached to the chloroplast; ● imported into the chloroplast; ●● imported and processed; - unclear localization.

3.4 FSD1 is present in the chloroplast stroma

As a complementary approach to the *in vitro* uptake assay, the sub-cellular localization of FSD1 was further analyzed by transient transformation of mesophyll protoplasts from *Arabidopsis thaliana* with a GFP-fusion construct. The GFP- and autofluorescence of transformed protoplasts was monitored with a confocal laser scanning microscope and fluorescence signals subsequently merged. Since chlorophyll emits a red autofluorescence when excited by the confocal laser beam, this signal was used as an indicator for chloroplasts (Figure 9A, second panel). In the following experiments, the localization of the C-terminal GFP fusions of the full-length FSD1 construct was monitored. The FSD1-GFP signal was visible exclusively within the stromal compartment of the chloroplasts, clearly overlapping with the red autofluorescence emitted from the chloroplast (Figure 9A). The GFP-signal from the FSD1 construct is clearly comparable to the pattern obtained with control constructs that are targeted into the stroma of chloroplasts (Figure 9A; pSSU-GFP). In addition to transient transformation of *Arabidopsis thaliana* mesophyll protoplasts, biochemical fractionation was also employed to analyze the localization of FSD1. For this purpose, sub-fractions from pea chloroplasts were analyzed by immunoblotting. FSD1 indeed displayed localization in the stroma (Figure 9B).

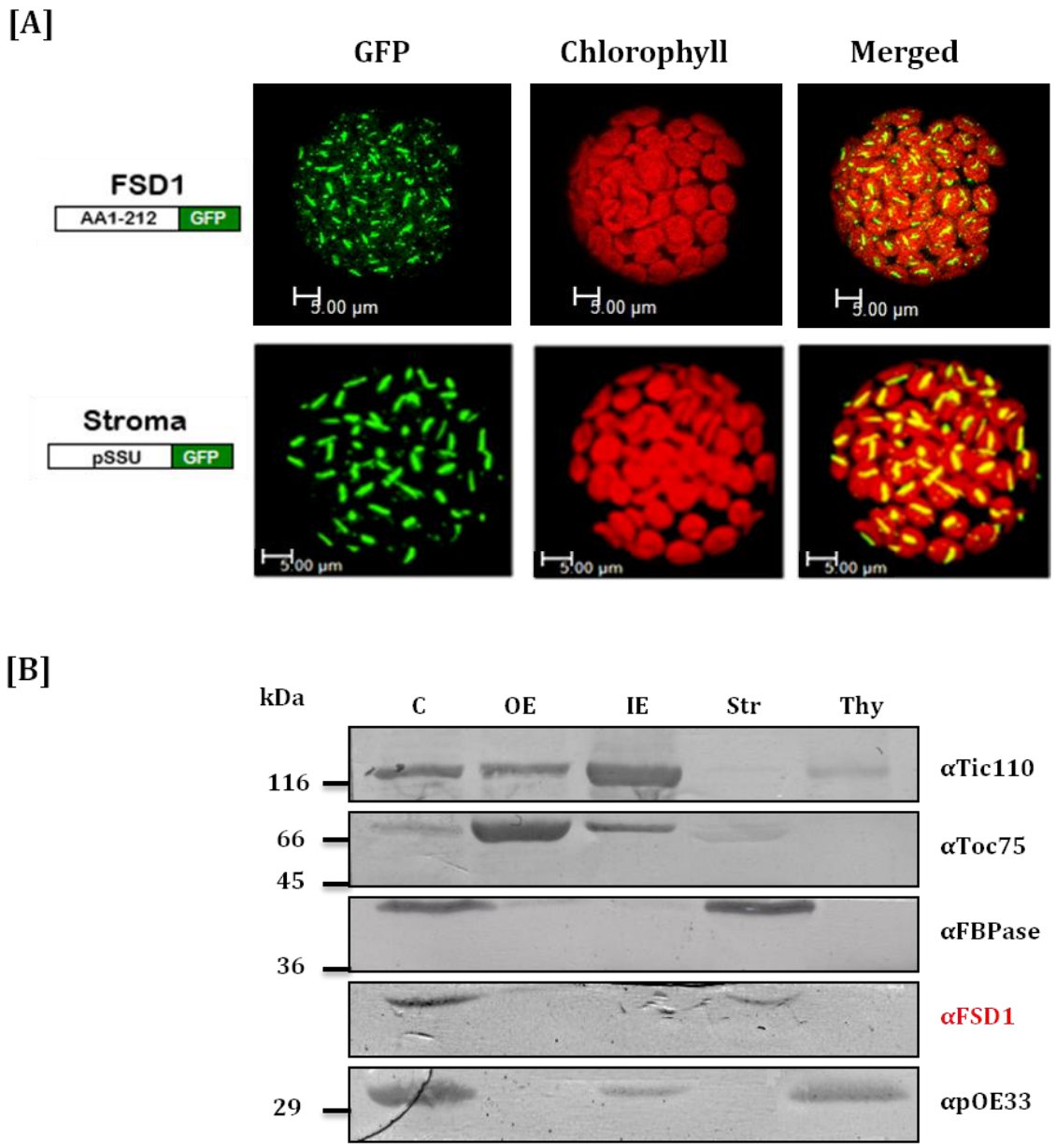


Figure 9 | FSD1 is targeted to the chloroplast *in vivo* and is located in the stroma. [A] *Arabidopsis thaliana* mesophyll protoplasts were transiently transformed with FSD1-GFP fusion constructs and pSSU as stroma control (GFP constructs). The resulting localization of the constructs was analyzed using a confocal microscope. One representative transformed protoplast is depicted for each construct. The first column shows the GFP signal, the second column the chlorophyll autofluorescence and the third column the merge picture. All signals were detected exclusively in chloroplasts. FSD1 was found to localize to the stroma. The used construct is depicted on the left. AA, amino acids; bar, 5 µm [B] Immunoblot localization of FSD1 in pea chloroplast sub-fractions; total chloroplast (C), outer envelope (OE), inner envelope (IE), stroma (Str) and thylakoid (Thy) using antibodies against FSD1 (Agrisera AB, Sweden), FBPase (stroma marker), pOE33 (thylakoid marker), Toc75 and Tic110 (outer and inner envelope marker respectively).

3.5 *In vitro* characterization of FSD1 import properties

3.5.1 Energy dependence of FSD1 import

Successive translocation of precursor proteins across the chloroplast membranes is an energy-dependent process involving the hydrolysis of nucleoside triphosphates at the outer envelope, in the intermembrane space and in the stroma (Kouranov & Schnell, 1997; Pain & Blobel, 1987; Young, Keegstra, & Froehlich, 1999). Generally, a concentration of nucleotide triphosphates above 100 µM is required for complete translocation of a standard precursor protein across the outer envelope and inner envelope and into the stroma (Theg, Bauerle, Olsen, Selman, & Keegstra, 1989). To screen for the energetic requirement of FSD1 translocation across the chloroplast envelope membranes, the endogenous nucleotide triphosphates were first depleted from the 35^S-radiolabeled FSD1 translation products via gel filtration prior to the *in vitro* import assay. Likewise, to minimize the production of endogenous nucleotide triphosphates, the chloroplasts and the subsequent import reaction were kept in the dark. Consequently, only the influence of externally added nucleotide triphosphates on the import characteristic was investigated. As a control for the energy requirement of stromal-localized proteins during the import reaction, the import of 35^S-radiolabeled pSSU was also monitored.

As depicted in Figure 10, import of 35^S-radiolabeled FSD1 is largely diminished in the absence of ATP (lane 2), while the addition of exogenous ATP resulted in a significant increase of the import yield of the precursor protein (lane 3 – 6). In general, a 4- to 5-fold stimulation in the presence of ATP can be observed for the import of FSD1. A similar import characteristic was also observed for pSSU. It should be noted that the import of FSD1 seems slightly decreased in the presence of a high ATP concentration (Figure 10A and 10B). The import of the inner envelope AtQORH, on the other hand, is not dependent on the presence of nucleoside triphosphates (Figure 10A). Thus, it could be concluded that the import of FSD1 is dependent on ATP but the ATP concentration required is lower compared to the standard stroma localized pSSU (Figure 10).

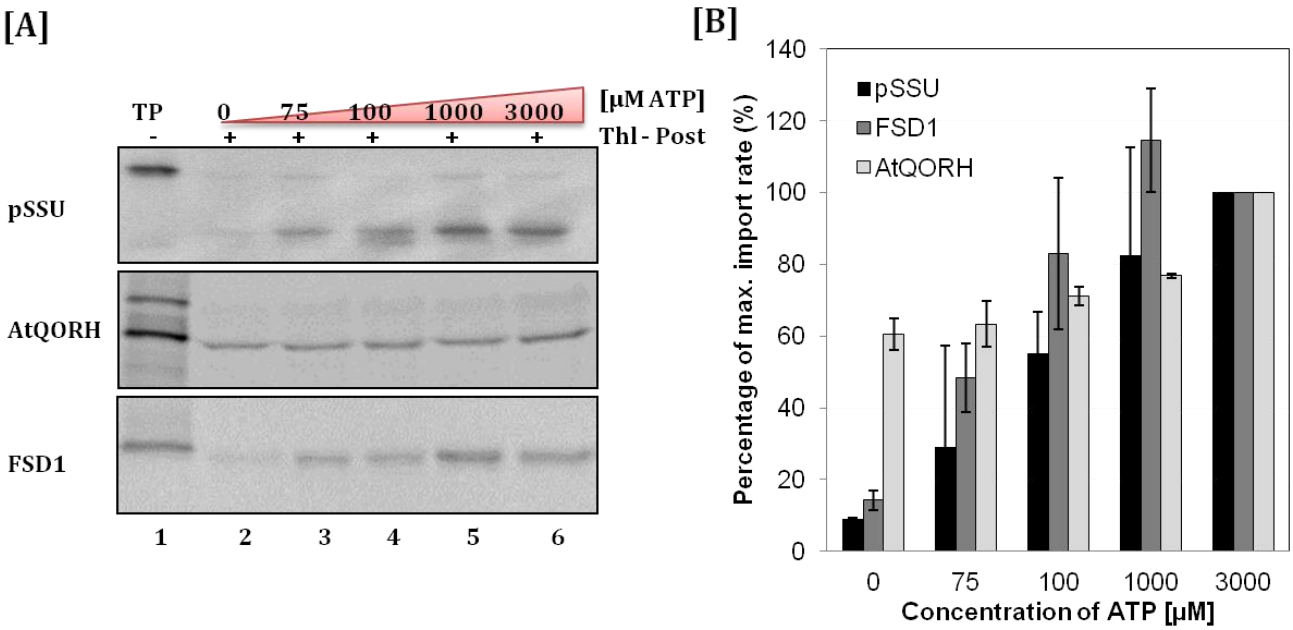


Figure 10 | ATP-dependence of FSD1 import. Import into intact pea chloroplasts was performed by incubating *in vitro* synthesized ^{35}S-FSD1 with chloroplasts corresponding to 10 µg chlorophyll at 25°C in a standard import reaction as described in Methods (see 3.2.4.1). Results were analyzed by a 12.5% SDS-PAGE. [A] Internal ATP was depleted from both chloroplasts and translation products. Different ATP concentrations were externally added (0, 10, 75, 1000 and 3000 µM). TP represent 10% of the translation products used in the import experiments. All samples were treated with thermolysin after the import reaction (Lane 2-6). [B] Quantitative analysis of the data from replicate experiments (n=3) including those presented in [A] with standard deviation bars.

3.5.2 FSD1 depends on proteinaceous components on the chloroplast surface for import

Import of precursor proteins into chloroplasts often requires protease-sensitive components at the outer envelope membrane. To investigate whether the import site of FSD1 may be of proteinaceous nature, import experiments were performed with purified pea chloroplast that were pre-treated with thermolysin. Under appropriate conditions thermolysin degrades surface-exposed epitopes of outer envelope proteins as well as receptors on the chloroplast surface, but leaves inner envelope and deeply embedded outer envelope proteins intact (Cline, Werner-Washburne, Andrews, & Keegstra, 1984; Joyard et al., 1983). Import of precursor proteins that require intact proteinaceous components should be inhibited after protease treatment. The thermolysin pre-treatment was assessed by immunoblotting and showing that surface exposed domains of the receptor proteins Toc159 and Toc34 were removed while the inner envelope protein Tic110 remained intact (Figure 11A). Chloroplasts from the same batch as used for the immunoblot analysis were used for the import assays. Briefly, intact chloroplasts were treated with 1 mg thermolysin per 1 mg chlorophyll for 20

minutes on ice in the dark. After treatment, intact chloroplasts were re-isolated on a Percoll cushion in the presence of 5 mM EDTA and used for import reactions. In parallel, non-treated chloroplasts were used in control reactions. The 35^S-radiolabeled FSD1 and the control protein 35^S-radiolabeled AtQORH and 35^S-radiolabeled pSSU, which import is known to be dependent on thermolysin sensitive components on the chloroplast surface (Miras, et al., 2007; Schnell, Blobel, & Pain, 1991), were incubated with chloroplasts corresponding to 15 µg chlorophyll at 25°C for 10-12 minutes for 35^S-radiolabeled FSD1 and 35^S-radiolabeled AtQORH, and 5 minutes for 35^S-radiolabeled pSSU (Figure 11B). Protease pre-treatment resulted in a significant decrease in binding and import of pSSU. Intriguingly, the translocation of AtQORH was also diminished, though it had been shown that it does not require the main TOC receptor proteins (Miras, et al., 2002). It seems that although AtQORH is still imported into protease treated chloroplasts, a protease sensitive component enhances import efficiency. It should be noted that an approximate 30% bypass imports could still be observed for both pSSU and AtQORH (Figure 11B). These residual imports might be caused by the partial or incomplete degradation of the surface-bound receptors by thermolysin as seen by the presence of residual Toc159 and Toc34 receptors on the pre-treated chloroplast in the immunoblot (Figure 11A). FSD1 import was, on the other hand, completely abolished, demonstrating that FSD1 depends on protease sensitive receptors on the chloroplasts.

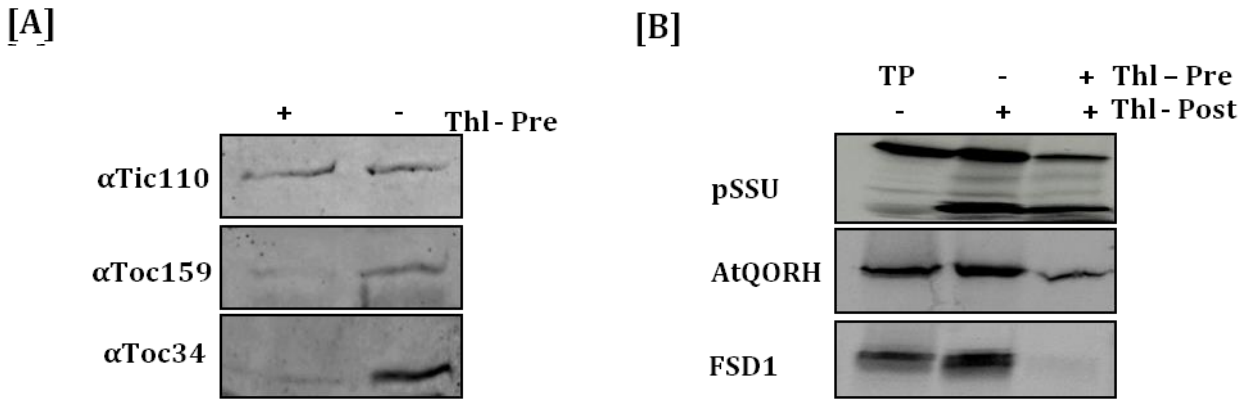


Figure 11. | Import of FSD1 into chloroplasts is dependent of protease-sensitive receptors. (A) Intact purified chloroplasts were either treated or not treated with the protease thermolysin for 20 minutes at 4°C. Chloroplasts were re-purified on Percoll gradients prior to further use. Efficiency of the proteolytic treatment was controlled by immunoblotting using antisera against the outer envelope localized translocon subunits Toc159 and Toc34 and the inner envelope translocon subunit Tic110. [B] Chloroplasts from the identical batch as tested in [A] were used for a standard import reaction. All other conditions are as indicated on top of the gel.

3.5.3 FDS1 uses a distinct pathway that engages some of the components of the common TOC/TIC – machinery across the chloroplast envelope membranes

Having established that FSD1 is targeted to the stroma via an unconventional targeting signal, the question arose whether the translocation across the outer envelope occurs through the TOC core, TOC159/75/34 translocon. In a first set of experiments, *in vitro* import experiments were carried out in the presence or absence of the bacterially expressed and purified, unlabelled pSSU (Figure 12). Heterologously expressed proteins containing a typical transit sequence, like pSSU, normally compete for translocation with other precursor proteins that use the general translocation pathway via the TOC159/75/34 (Perry, Buvinger, Bennett, & Keegstra, 1991; Schnell, et al., 1991). As internal control for the general- and TOC/TIC-independent pathway, imports of 35^S-radiolabeled pSSU and 35^S-radiolabeled AtQORH were tested in the presence or absence of the unlabeled competitor respectively. The data presented in Figure 12A clearly illustrate that the amount of the 35^S-radiolabeled pSSU that was imported into the chloroplast decreased as the concentration of the unlabelled pSSU competitor increased. Import was inhibited approximately by 80% in the presence of 2 μM of the competitor; only an approximately 5% residual import remained after the application of 10 μM the competitor (Figure 12B). The import of 35^S-radiolabeled AtQORH was unaffected even in presence of excess pSSU competitors (5 and 10 μM of competitor; Figure 12A, lane 5-6), when translocation of pSSU itself was already completely abolished (Figure 12A). These observations revealed that AtQORH did not engage the TOC159/75/34 complex for translocation into chloroplasts, a result that is consistent with previous works (Miras, et al., 2002; Miras, et al., 2007). FSD1, on the other hand, confers a clear sensitivity to the presence of the pSSU. At 1μM concentration of competitor, import of FSD1 is greatly diminished, but in contrast to pSSU the translocation of FSD1 is never completely blocked, even at the highest competitor concentration as judged by the appearance of the mature FSD1 after protease treatment (Figure 12A). Translocation across the outer envelope occurs through the import channel Toc75 (Hinnah, Wagner, Sveshnikova, Harrer, & Soll, 2002). In order to investigate whether Toc75 is involved in FSD1 translocation, isolated pea chloroplasts were pre-incubated with a known inhibitor of the import channel Toc75. The positively charged spermine binds to Toc75 and blocks the import channel, thus inhibiting the function of the TOC complex (Hinnah, et al., 2002). In the presence of spermine, the import of pSSU was completely abolished while

it had little or no influence on the translocation of FSD1 into the chloroplasts. A similar result was observed for AtQORH after spermine treatment (Figure 12C, lane 3). Judging from these results, it appears that FSD1 may engage in a novel operational pathway that is distinct from pSSU across the outer envelope membrane of chloroplast.

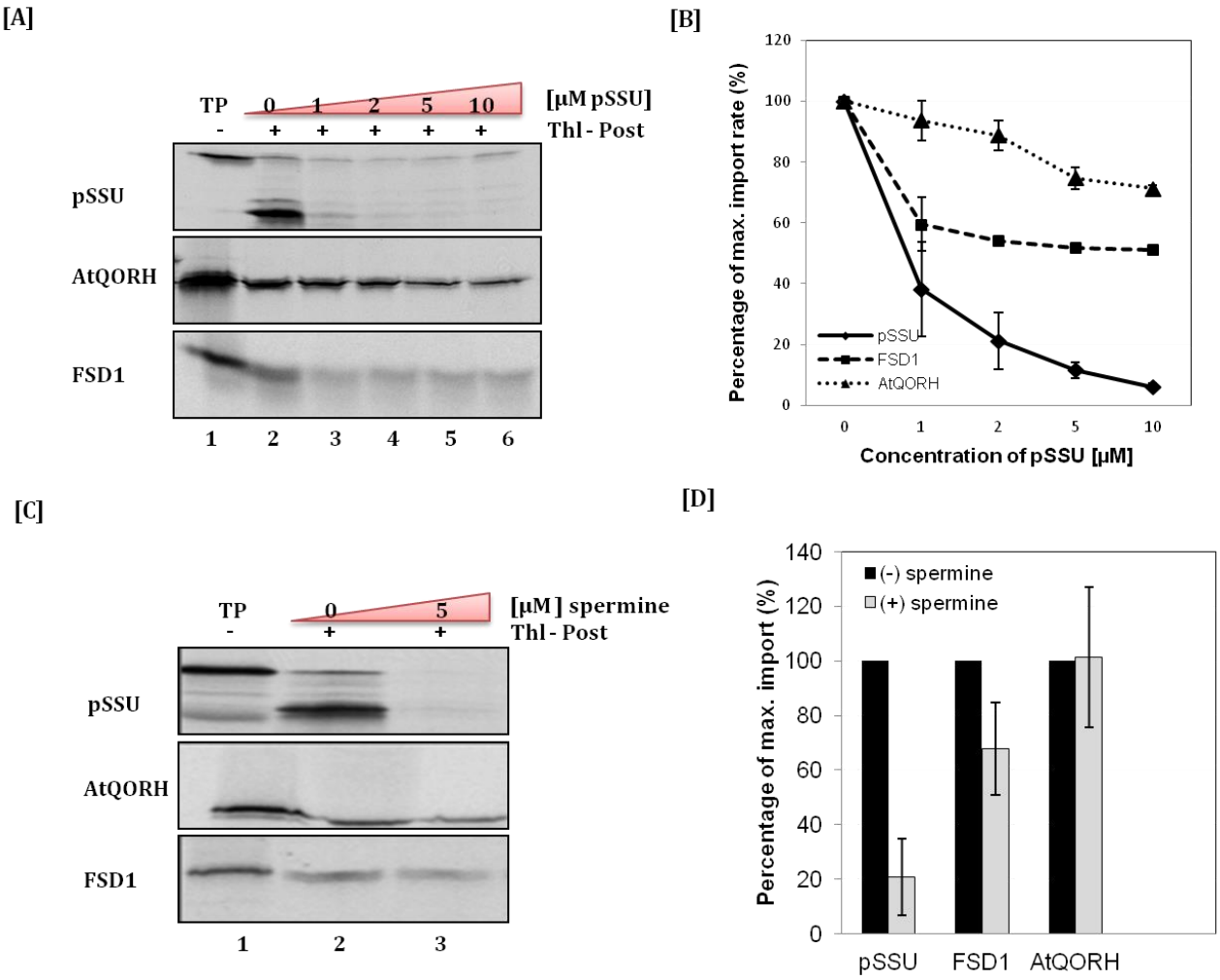


Figure 12 | FSD1 import into chloroplasts is not affected in the presence of different TOC translocon inhibitors. [A] Import of FSD1 and pSSU were assayed in the presence of increasing pSSU competitive concentrations as indicated. [B and D] Quantitative analysis of the data from replicate experiments (n=3) including those presented in [A and C] with standard deviation bars. (C) Import of FSD1, AtQORH and pSSU was conducted in the absence or presence of 5 mM spermine.

3.5.4 FSD1, AtQORH and Tic32 do not use the same import pathway into the chloroplast

Evidence for the operation of distinct import pathways was previously reported for the chloroplast import of another inner envelope protein without canonical transit-sequence, Tic32 (Nada & Soll, 2004). Competition experiments were used to address the question of whether FSD1, AtQORH and Tic32 share components of the same import

pathway (Figure 13). For this purpose, bacterially expressed and purified FSD1 was added to import reactions containing 35^S-radiolabeled AtQORH and 35^S-radiolabeled Tic32. Three different competitor concentrations were used: 1 µM, 2 µM, and 5 µM. As depicted in Figure 13A, strong competition only occurred in the case of FSD1 itself. The fact that the import of AtQORH and Tic32 still occurred in the presence of excess FSD1, it can be concluded that FSD1 translocation does not engage components involved in either Tic32 or AtQORH translocation. The classical precursor protein pSSU shows a slight import inhibition in the presence of recombinant FSD1, which has similar properties as observed in the reciprocal experiment (Figure 12A): addition of 1µM of FSD1 results in an inhibitory effect that is not intensified by higher amounts of competitor. Taken together, these results imply that FSD1 might share some translocation components with pSSU on their way into chloroplasts but demonstrated nonetheless distinct import patterns to those of the general import pathways.

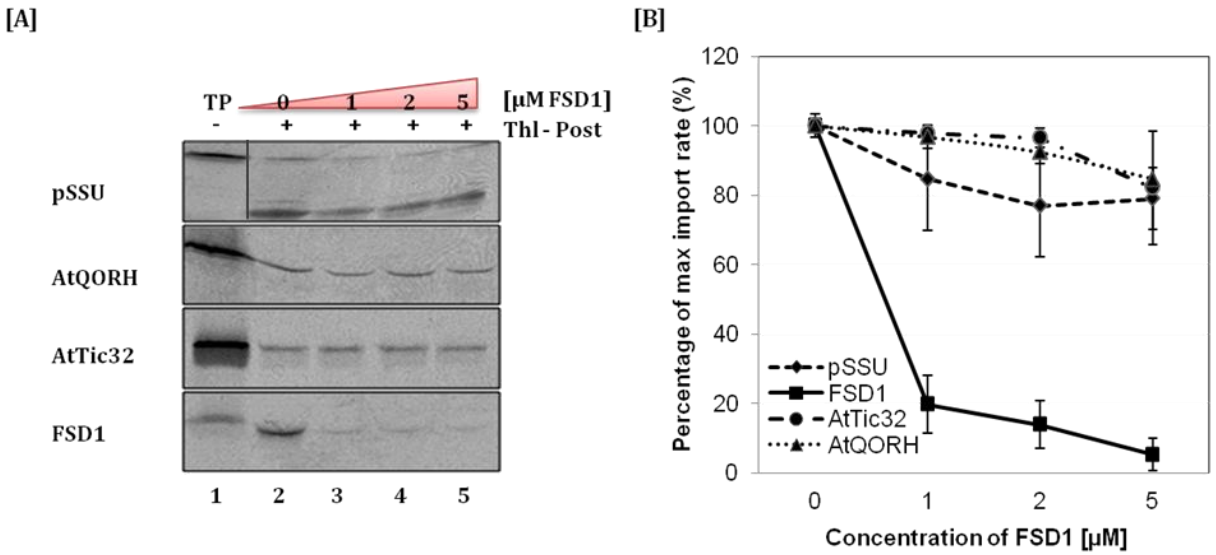


Figure 13 | FSD1, AtQORH and AtTic32 do not use the same import site into chloroplasts. [A] 35^S-FSD1, 35^S-AtQORH and 35^S-AtTic32 were synthesized *in vitro* and added to incubation mixtures containing 3 mM ATP and the indicated 1 µM, 2 µM, and 5 µM concentrations of heterologously expressed unlabeled FSD1, used as competitors. Mock incubations lacked competitor. After 15 min the reactions were stopped on ice. Plastids were re-purified on Percoll and supplemented with thermolysin (Thl) to degrade non-imported protein, as indicated. Proteins were detected by autoradiography following SDS-PAGE. [B] Quantitative analysis of the data from replicate experiments (n=3) including those presented in [A] with standard deviation bars.

3.6 The mechanism of FSD1 targeting to chloroplasts

3.6.1 The N-terminus of FSD1 is essential for recognition and targeting of the protein to chloroplasts

Primary amino acid sequence comparison between the FSD1 proteins and its prokaryotic homologues revealed that the N-terminal region is found exclusively in the plant FSD1 proteins (Figure 14). The acquisition of an N-terminal extension is concurrent with plastid evolution (McFadden, 1999). Hence, this observation strongly suggests that these extensions of the plant FSD1 proteins are directly required for the specific chloroplast targeting.

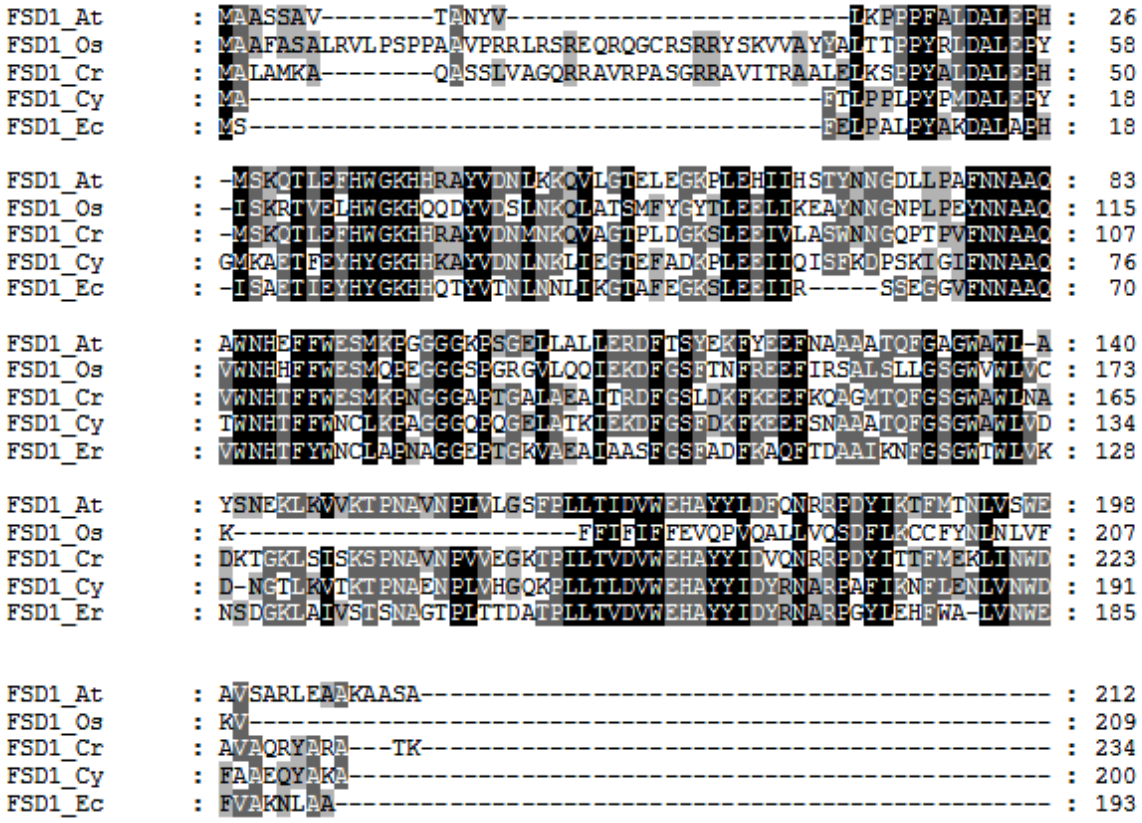
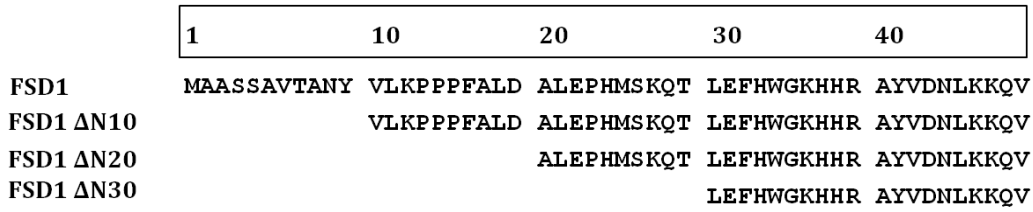


Figure 14 | Alignment of *Arabidopsis thaliana* FSD1 protein with several plant FSD1 proteins, cyanobacterial and bacterial closest homologues. Residues conserved in the four proteins are *black* coloured. Similar residues (according to the following groupings: ASPTG, ILMV, KRH, NQ, DE, YWF and C) conserved in the primary sequence of FSD1 and in any other primary sequence are *light gray* coloured. FSD1_At (this work), FSD1_Os (AAX09664.1), FSD1_Cr (EDP05850.1), FSD1_Cy [cyanobacterium PCC 7702] (WP_017323315.1), FSD1_Ec (YP_006091930.1).

For most internal chloroplast proteins the N-terminal sequence contains essential targeting information: classical precursor proteins comprise cleavable N-terminal transit peptides, but also Tic32 has its non-cleavable targeting signal in the N-terminus. The only known exception is AtQORH, which is guided by internal sequence information (Miras, et al., 2002). Since FSD1 clearly exhibited some import characteristics similar to canonical precursor proteins, the targeting properties of FSD1 extreme N-terminus were examined. For this purpose, several truncated versions of FSD1, in which 10-30 amino acids were progressively removed from the N-terminus of FSD1, respectively, were generated (Figure 15A). Full-length FSD1 and the deletion mutants were subsequently synthesized *in vitro* and used for protein import studies. The truncated proteins showed a strong reduction in the extent of chloroplast binding (Figure 15B). With the exception of the full-length FSD1 all mutant proteins remained protease accessible, demonstrating that no productive interaction with chloroplasts had occurred. Deletion of the first ten amino acids already abolished import (Figure 15B; FSD1 ΔN10). The same observation was made for FSD1ΔN20 and FSD1ΔN3 (Figure 15B). Deletions at the C-terminus, on the contrary, did not appear to have any significant impact on the targeting and import efficiency of FSD1 (Figure 15B; FSD1 ΔC10 – ΔC30), indicating that the N-terminus of FSD1 constitutes an indispensable part in signal recognition and targeting to chloroplasts.

[A]



[B]

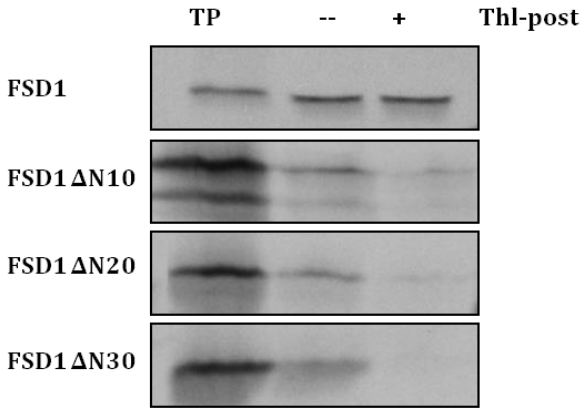


Figure 15 | The N-terminus of the stromal FSD1 is essential for targeting to chloroplasts. [A] The N-proximal 50 amino acids of FSD1 (GenBank accession no. AAG40062.1) are shown in addition to the respective deletion mutants. [B] Progressive N-terminal deletions of FSD1 were synthesized *in vitro* as radiolabeled preproteins, respectively. Import and analysis conditions were as outlined in Figure 6.

3.6.2 Additional information is required for plastid localization of FSD1 in vivo

To investigate whether the first ten amino acids at the N-terminus of FSD1 is sufficient to drive the translocation of FSD1 to the plastid *in vivo*, the localization of partial or full-length FSD1 constructs were monitored via transient transformation of *Arabidopsis thaliana* mesophyll protoplasts. Intriguingly, while the ten amino acids at the N-terminus of FSD1 are important for plastid recognition (Figure 15B; FSD1 ΔN10), they were not sufficient to drive import of a GFP construct into plastids (Figure 16; FSD1_N6- to N30-GFP). For all chimeric constructs except the full length FSD1-GFP cytosolic localization of the fluorescent signal was observed. This demonstrated that the N-terminal part of FSD1 (30 first residues) is not sufficient for plastid localization of a reporter protein. To establish if the C-terminus also contains important targeting information, we generated C-terminal deletion constructs where 10, 20 or 30 amino acids were missing, FSD1ΔC10, FSD1ΔC20 and FSD1ΔC30 (Figure 17). None of these proteins were affected in translocation efficiency, indicating that the C-terminus of FSD1 is not necessary for targeting or import into chloroplasts. Please note that upon import of the C-terminal deletion constructs two protease resistant bands

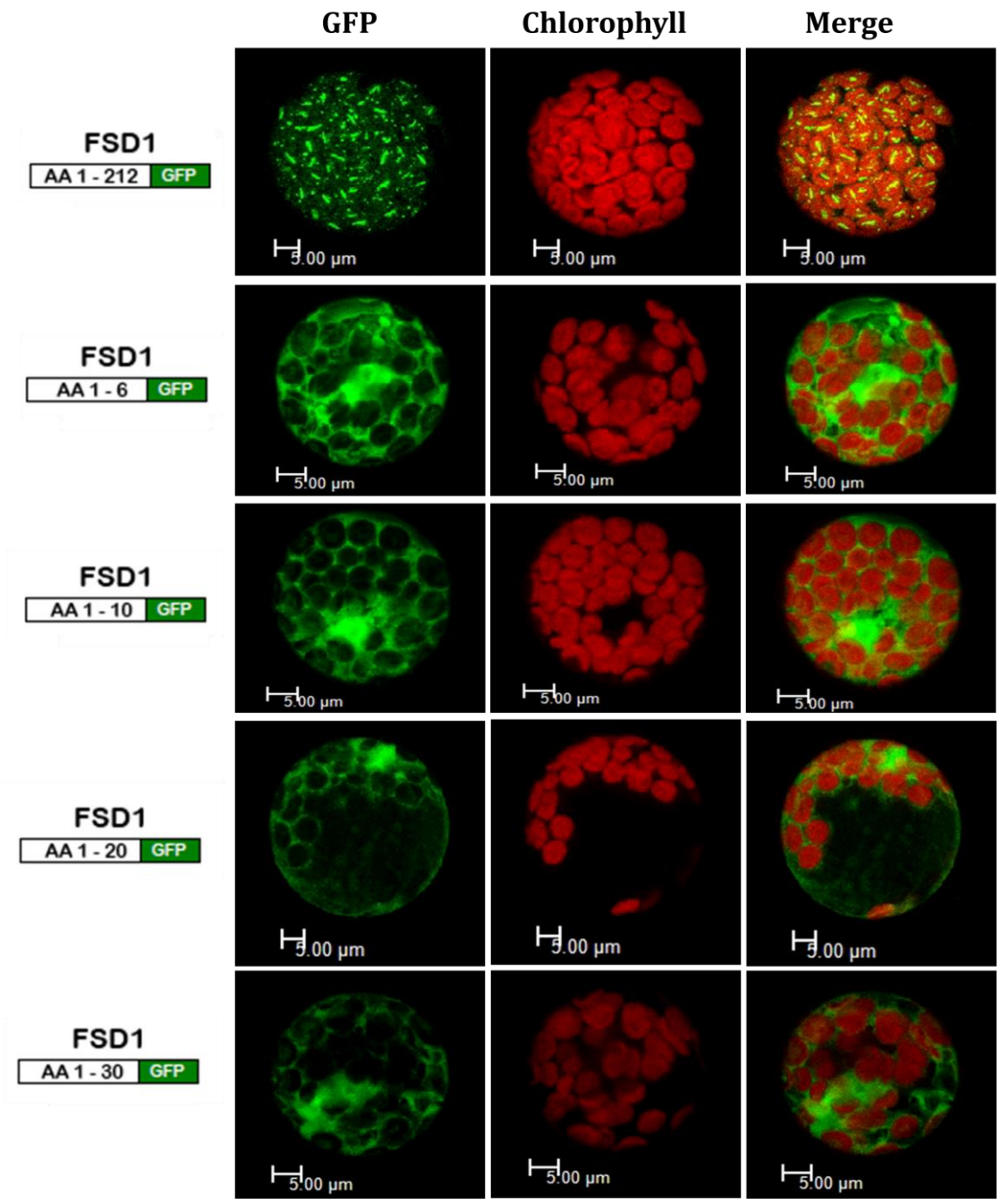


Figure 16 | The interplay between the N–terminal region of FSD1 and the downstream sequences are essential for targeting and localization of the protein to the chloroplast. *Arabidopsis thaliana* mesophyll protoplasts were transiently transformed with GFP constructs of either full-length FSD1 or partial N-terminus constructs of FSD1. Maximum intensity signals from confocal images are shown for GFP-fluorescence (GFP), chlorophyll-autofluorescence (chlorophyll), and an overlay of both (merged). The used constructs are depicted on the left. AA, amino acids; bar, 5 µm.

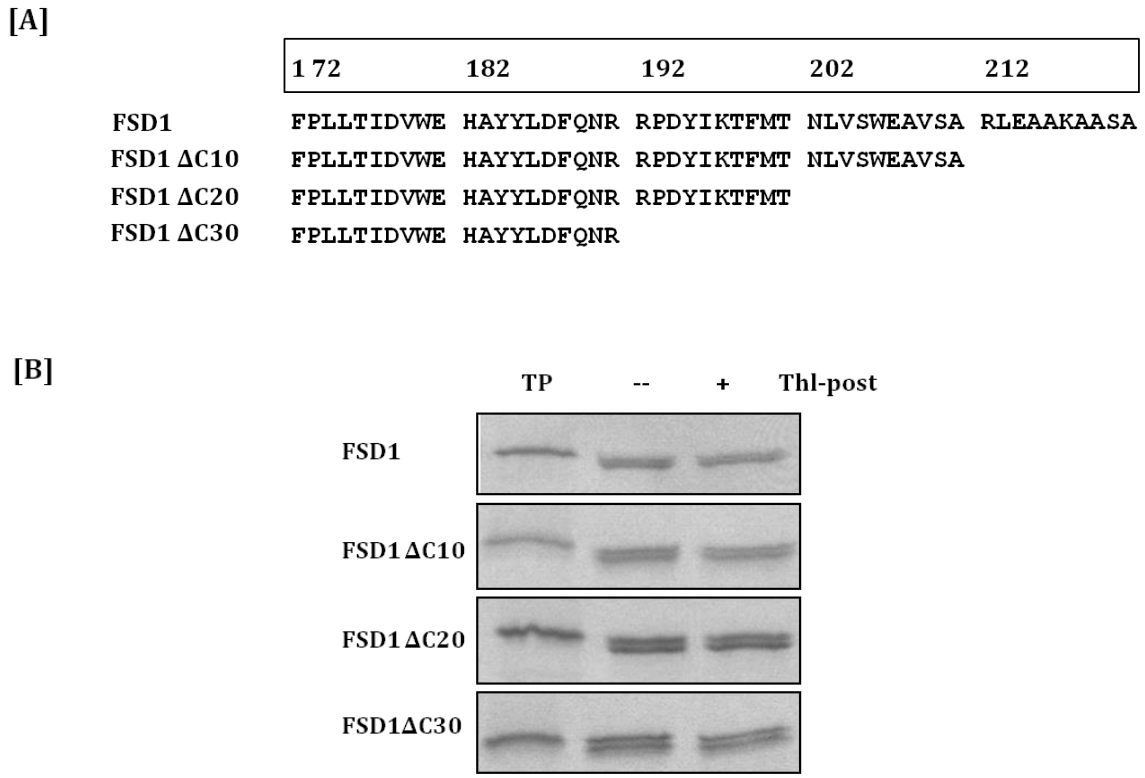


Figure 17 | The extreme C-terminus of the stromal FSD1 is not necessary for import. [A] The C-proximal 50 amino acids of FSD1 (GenBank accession no. AAG40062.1) are shown in addition to the respective deletion mutants. [B] Progressive C-terminal deletions of FSD1 were synthesized *in vitro* as radiolabeled preproteins, respectively. Import and analysis conditions were as outlined in Figure 6

3.7 Cross-linking of the 'non-canonical' FSD1 precursor to chloroplast envelope components

3.7.1 The N-terminus of FSD1 specifically interacts with large outer envelope proteins

A strategy to further map specific interactions between FSD1 and envelope-based translocation components in chloroplasts, precursor binding and chemical cross-linking assays were employed. The chimeric FSD1 precursor protein was used as cross-linking substrate for these studies. The FSD1-protA hybrid protein consisted of FSD1 fused at its C-terminus to the IgG binding domains of staphylococcal protein A. This hybrid was chosen as it embodied both the import characteristics of FSD1 and the high affinity IgG binding sites of protein A. The latter feature provided a simple means of purifying the precursor using IgG-Sepharose. The fusion of the IgG binding domains to FSD1 had a negligible effect on the binding and import characteristics of the FSD1 precursor protein in the *in vitro* assays using intact chloroplasts (data not shown). The hybrid FSD1-protA protein was cross-linked to isolated chloroplasts at the late stage of the import process

in the presence of 2 mM ATP and 0.1 mM GTP (Schnell & Blobel, 1993) using the membrane-permeable cross-linker, DSP (Ji, 1983). The late stage of import had been chosen as it contains the potential late import intermediate that is simultaneously inserted across the outer and inner membranes (Schnell & Blobel, 1993). Hence, this late intermediate at this import stage provided the potential to identify components of both at the outer and the inner membrane translocons.

Isolated intact pea chloroplasts were incubated in at 25°C in the presence of 2 mM ATP and 0.1 mM GTP together with the FSD1-protA hybrid protein for 3 minutes. The precursor-bound chloroplasts were re-isolated and treated with 5 mM of the cleavable cross-linker DSP for 30 minutes on ice to induce cross-linking. After cross-linking, the reaction was quenched with 50 mM glycine and intact chloroplasts were re-isolated, lysed and fractionated to yield a total envelope membrane fraction. The membrane fractions were solubilized with 1% dodecylmaltoside (DOMA), clarified by centrifugation, and used for immunoprecipitation using IgG sepharose. In order to keep the target protein covalently attached to the components in the envelope membranes, the immunoprecipitates were resolved under non-reducing conditions and analyzed by a 4-15% gradient SDS-PAGE. A control experiment was performed in a similar manner except for the absence of the cross-linking substrate, FSD1-protA.

When the cross-linked chloroplasts containing bound FSD1-protA hybrid proteins were immunoprecipitated with IgG sepharose, two complexes that migrated with estimated molecular masses of 600 and 300 KDa, T1 and T2 were observed. Both bands were only present in the immunoprecipitates of the FSD1-protA preparation used for cross-linking and not in the control reaction (Figure 18). These two protein bands were then excised and the individual proteins that constitute the cross-linked complexes were identified by mass spectrometric (MS). Results from the MS analyses revealed, however, only major association to the thylakoidal ATP synthase complex and the stromal RuBisCo with the hybrid FSD1-protA precursor protein, indicating that the large majority of the precursor-bound envelope membranes were recovered in the thylakoidal fraction when chloroplasts were treated with cross-linkers. Similar results were observed in other replicates regardless of whether the chloroplasts were lysed hypotonically or hypertonically (data not shown). These observations are most probably due to the

abundance and the lower sedimentation value of the thylakoid membranes. Hence, it explains the lower yield of the purified, cross-linked FSD1-protA / envelope complexes that was recovered.

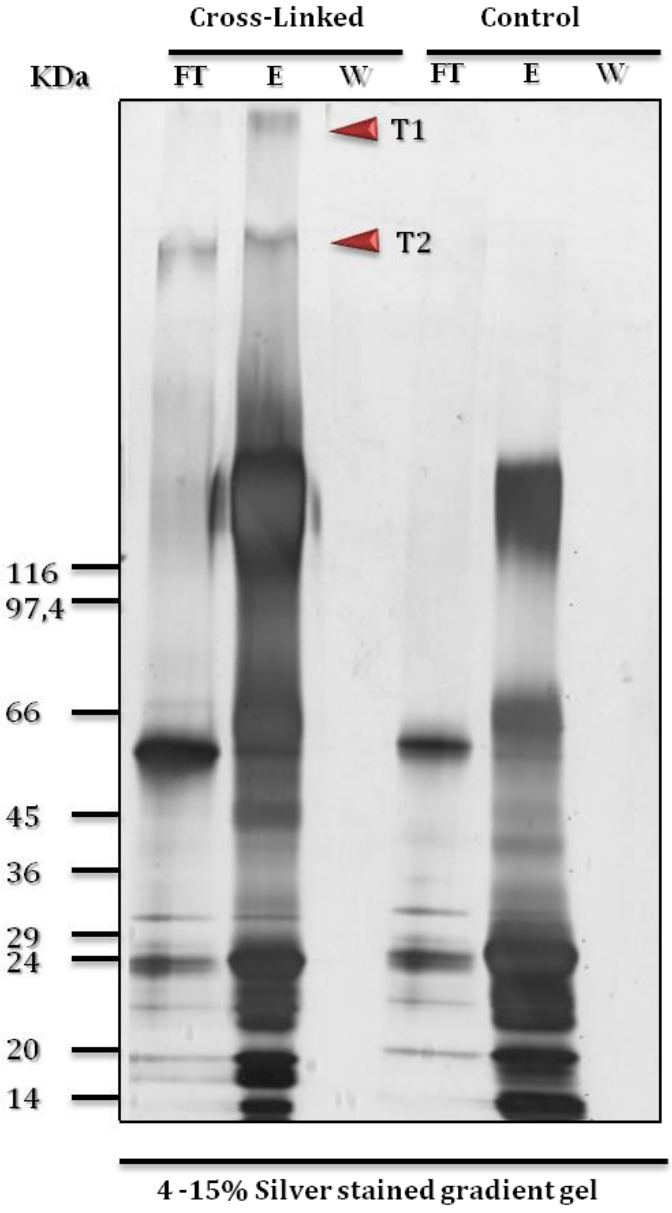


Figure 18 | Chemical cross-linking and immunoprecipitation analyses of the hybrid FSD1-protA-bound chloroplasts. Pea chloroplasts were incubated with the hybrid FSD1-protA precursor protein under the mentioned import conditions for 3 minutes, re-isolated, lysed and cross-linked with DSP. Supernatants of solubilized total membranes were immunoprecipitated with IgG sepharose beads. Immunoprecipitates were subsequently analyzed by a non-reducing 4-15% gradient SDS-PAGE followed by Silver staining. Molecular masses of markers in kilodaltons; KDa. Two gel bands were excised and sent for mass spectrometric analysis.

To overcome this issue, isolated and purified pea outer envelope membrane vesicles were used for the ensuing precursor binding and cross-linking analyses. The purified outer envelope membranes were isolated as right-side-out vesicles, in other words, the cytosolic side of the vesicles is exposed at the surface of the vesicles (Waegemann, Eichacker & Soll, 1992). Therefore, it provides the opportunity for the characterization

of the individual components in the cross-linked, precursor-bound / envelope complexes without the contamination from stromal and thylakoidal protein complexes.
For most internal chloroplast proteins the N-terminal sequence contains essential targeting information: classical precursor proteins comprise cleavable N-terminal transit peptides, but also Tic32 has its non-cleavable targeting signal in the N-terminus. The only known exception is AtQORH, which is guided by internal sequence information (Miras, et al., 2002). Since FSD1 clearly exhibited some import characteristics similar to canonical precursor proteins, the targeting properties of FSD1 extreme N-terminus were addressed.

As cross-linking substrate, a shorter hybrid peptide which consists of the N-terminal region of FSD1 (1 – 25 amino acids) was used. This hybrid protein was modified with a biotin molecule and an additional cysteine residue at the C-terminus. Here, the N-terminal FSD1-biotin hybrid protein represents the ideal substrate as it encompasses both the targeting specificity of FSD1 (as demonstrated in Figure 15) and biotin's femtomolar association constant with streptavidin / or avidin. The biotinylation of the N-terminal FSD1 hybrid protein is valuable for purifying the hybrid protein using streptavidin-sepharose as well as for the specific immunolocalization of the biotinylated hybrid protein using the VECTASTAIN® ABC system. Since only outer envelope membranes were used, the interactions between the N-terminal FSD1 hybrid protein and the outer membrane translocons were probed under binding conditions. The modified N-terminal FSD1-biotin precursor protein was used in the binding experiments in the presence of 0.1 mM ATP and incubated at 4°C for 5 minutes. After recovery, the precursor-docked envelope membranes were treated with 5 mM of the cross-linker N-(α-maleimidoacetoxy) succinimide ester (AMAS) for 30 minutes on ice to initiate cross-linking. Cross-linked products, containing the precursor-bound outer membrane complexes, were solubilized with 1% DOMA and prior to incubation with streptavidin-sepharose matrix. The eluates were separated by a 12.5% SDS-PAGE followed by immunodection with an α-biotin antibody. As depicted in Figure 18, various cross-linked products were observed when the precursor-bound membranes were treated with AMAS. No cross-reactivity was observed with an empty streptavidin-sepharose matrix as shown in the control reactions (Figure 18, control). Five gel pieces were excised from the SDS-PAGE gel lane of the cross-linked fraction and sent for analysis via automated

nano-spray LC-MS/MS (Figure 18A, band 1–5). Comparative analyses of the extracted peptide masses with pea sequence contigs unambiguously revealed close associations of the hybrid N-terminal FSD1-biotin protein with most components of the TOC core complex, particularly the receptor constituent Toc159. A summary of the peptide masses that matched to Toc159, To132, Toc120 and Toc75 is shown in Table S3 in Appendix. Amidst the highly abundant TOC core components, two proteins with molecular masses of 132- and 120 KDa were also identified in the MS data. Homology-based protein identification disclosed high similarity of the 132- and 120 KDa products with Toc132 and Toc120 translocons in *Arabidopsis thaliana.* The identified peptides matched partial sequences in the pea database which corresponded to proteins of 120 kDa and 132 kDa, respectively (see Table 6). Homology search against the Arabidopsis proteome revealed strong homology to AtToc132 and AtToc120, respectively, indicating that FSD1 was cross-linked to yet unidentified paralogues of Toc159 in pea.

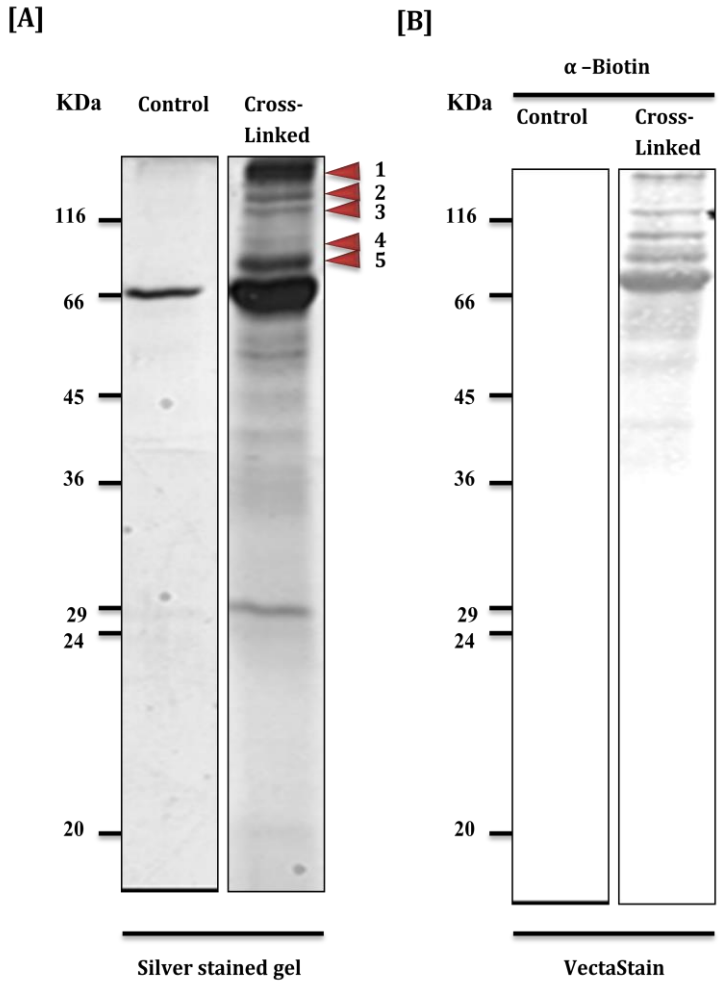


Figure 19 | Precursor binding and cross-linking analyses of the N-terminal FSD1-biotin hybrid precursor protein with isolated and purified pea outer envelope vesicles. Binding of N-terminal FSD1-biotin hybrid precursor protei was carried out at 0.1 mM ATP, 10 mM $MgCl_2$ in binding buffer for 5 minutes on ice using 25 µg of envelope proteins. Membranes were re-isolated, washed and cross-linked with AMAS [N-(α-maleimidoacetoxy)], a succinimide ester with a spacer length of 4.4 Å. Supernatant of solubilized outer membranes were purified via streptavidin-sepharose matrix followed by immunodetection by an antibody against biotin (VectaStain, biotin/avidin detection system). [A] The cross-linked products were separated by SDS-PAGE (12.5%) and were [B] identified by VectaStain. Arrows indicated bands excised for mass spectrometry sequencing. Molecular masses of markers in kilodaltons; KDa.

Table 6 | Identification of PsToc132 and PsToc120 by LC-MS/MS after combined tryptic digestion

Matched peptides of PsToc120 [a]	Observed mass [b]	Mr (expt)	Mr (calc)	ppm	Score
R.VNYTVSDTQPR.K	640.3153	1278.616	1278.616	-3.43	29.2
R.PAGLGSAAPLLEPAAR.V	745.9187	1489.8228	1489.8228	-1.61	33
R.KTEDSSIGEADEYDETR.R	648.947	1943.8226	1943.8226	-2.07	72

Matched peptides of PsToc132 [a]	Observed mass [b]	Mr (expt)	Mr (calc)	ppm	Score
R.LFVK.E	310.2127	681.4104	681.4104	0.6	31
K.FCNFR.R	372.1681	742.3216	742.3216	-0.59	18
K.DLAYTLR.S	426.2348	850.455	850.455	0.23	32
K.LQMIRVKFLRLANRL.G	710.3957	1418.7768	1418.7768	-1.5	48
K.ATSLGFDMQTVGK.D	685.8334	1369.6255	1369.6255	-1.85	50
K.EKIPVSFSGQVTK.D	581.8261	1161.6376	1161.6376	-0.25	27

a The amino acid residues appearing before and after the dot correspond to residues proceeding and following the peptide in the protein sequence.

b Average peptide mass

3.8 Toc132 and Toc120 represent two novel components of the TOC translocon

3.8.1 Identification of the novel Toc159 homologues of pea outer envelope

Toc159 represents one of the two families of GTPases that mediate the initial targeting of precursor proteins to the chloroplasts. Encoded by four different genes in *Arabidopsis*, these different homologues of Toc159 assemble into distinct TOC complexes that confer differential import characteristics of photosynthetic and non–photosynthetic plastidic proteins, respectively (see Introduction, section 5.1 and section 7.3). Such complexity of the TOC components is well characterized in *Arabidopsis* and rice (*Oryza sativa*) (Ivanova et al., 2004; Kubis et al., 2004; Voigt et al., 2005). No biochemical evidence, however, has been brought forward thus far demonstrating the existence of such gene families in pea. Therefore, the identification of Toc132 and Toc120 peptides while analysing the MS data from the precursor binding and cross-linking analyses raised the possibility that similar multigene families could also be present in pea.

To test the hypothesis, generation of antisera that could specifically recognize and distinguish these related but structurally distinct components of the TOC complexes in pea, in particular the Toc159 homologues, Toc132 (PsToc132) and Toc120 (PsToc120), was therefore prerequisite to confirm their localization. Previous studies on the *Arabidopsis* Toc159 homologues revealed a common tripartite domain structure consisting of the conserved C-terminal membrane anchor domain (M-domain), the central GTPase domain (G-domain), and the highly variable N-terminal acidic domain (A-domain) (Bauer et al., 2000; Ivanova, et al., 2004; Kubis, et al., 2004). Primary amino acid sequence analysis using ClustalW indeed showed a strong sequence identity between the G- and M-domain of the *Arabidopsis* Toc159 homologues. Despite of the high homology (~47.7%) between the *Arabidopsis* Toc132 (AtToc132) and Toc120 (AtToc120), the A-domain is clearly the most divergent region in the protein (Table 7). This renders the A-domain an ideal template for antibody production.

Table 7 | Comparison of amino acid sequence identity between the three members of the Toc159 family in *Arabidopsis* using ClustalW.

	Full length		A-Domain		G-Domain		M-Domain	
	AtToc132	AtToc120	AtToc132	AtToc120	AtToc132	AtToc120	AtToc132	AtToc120
AtToc159	36,9	40,1	15,5	21,2	51,2	50,9	50,1	48,1
AtToc132	-	75,9	-	47,7	-	92,8	-	87,0

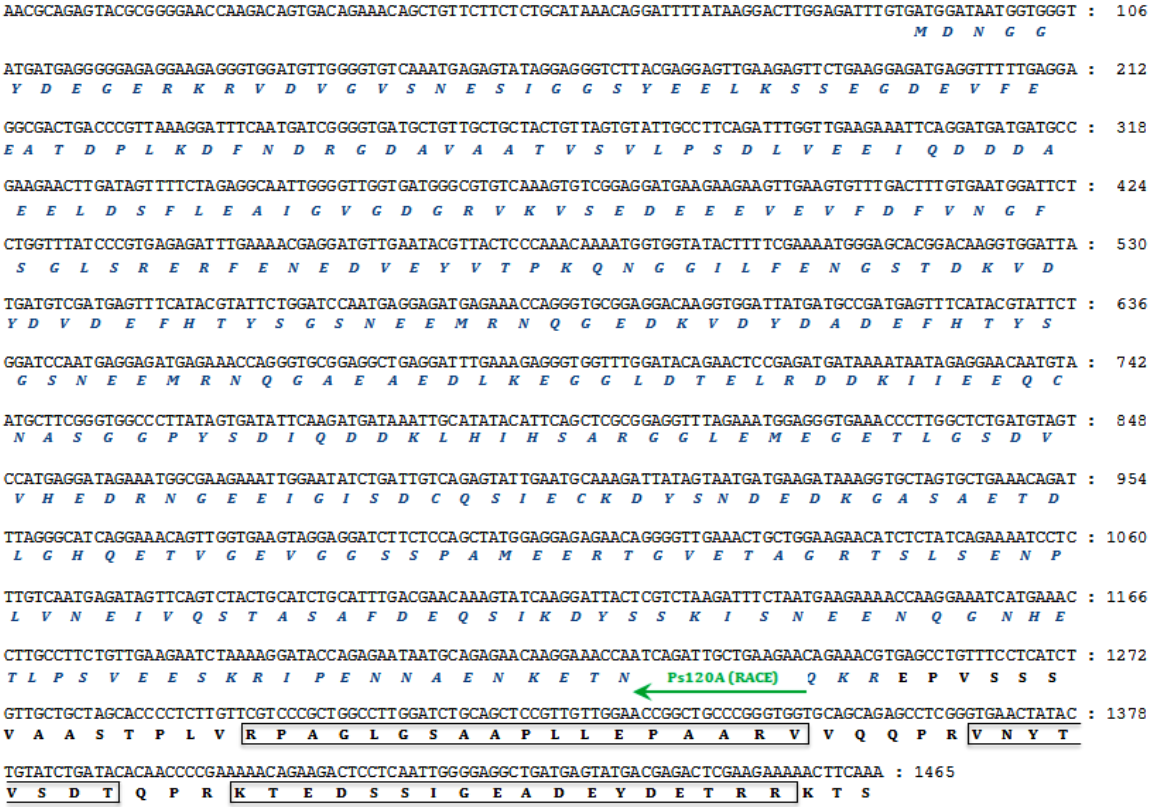
* The table presents the percentage of identity between the full length protein and the three domains of *Arabidopsis* Toc159 homologues.

3.8.2. Isolation and characterization of pea cDNA encoding the Toc159 homologues, PsToc132 and PsToc120

In order to resolve the question as to where the N-terminal A-domain of both PsToc132 and PsToc120 starts respective to that of their *Arabidopsis* homologues, a set of degenerative oligonucleotide primers (see 'Materials and Method', section 2.1.3) was used to isolate cDNA from pea encoding the PsToc132 and PsToc120. The primers were synthesized based on the sequence of the respective PsToc132 and PsToc120 peptides obtained from the MS analysis. The 5'-RACE PCR amplification produced two products with the size of 1465 bp and 1437 bp, respectively. The first 1465 bp cDNA clone encodes for part of the putative translocase of 120 kDa in pea, PsToc120. It contained an open reading frame of 1157 bp which encodes for the 391 amino acids long, putative A-domain of PsToc120 with a calculated molecular mass of 43.4 KDa (Figure 20A). BLAST analysis of the deduced amino acid sequence showed 35.5% sequence similarity to AtToc120 (Figure 20B) as well as to other plastidial Toc120 translocases (*Cicer arietinnum*, 57%, *Glycine max*, 35%, *Medicago truncatula*, 43%, *Cucumis melo,* 83%, *Malus domestica,* 27%, *Theobroma cacao* 41%, *Cucumis sativus* 54%, and *Solanum tuberosum* 69% see Table S1 in Appendix). On the other hand, the nucleotide sequence of the second 1437 bp cDNA clone showed a 69% homology to AtToc132, suggesting that part of this clone encodes for part of the PsToc132 translocase. This clone, however, lacked a true start methionine, therefore was most likely not N-terminally complete (Figure 21A and Figure 21B). Further amino acid sequence analysis indicated that the putative A-domain of Toc132 in pea is composed of at least 400 amino acids with a calculated molecular mass 44.4 KDa. Homology search demonstrated a 38.6% sequence similarity to the A-domain of AtToc132 (Figure 21B). Likewise, the deduced amino acid sequence of the PsToc132

A-domain also demonstrated homology to other plastidial Toc132 translocases (*Cicer arietinnum*, 81%, *Glycine max*, 72%, *Medicago truncatula*, 53%, *Cucumis melo,* 66%, *Malus domestica,* 52%, *Theobroma cacao* 59%, *Cucumis sativus* 66%, and *Solanum tuberosum* 64% see Table S2 in Appendix). In addition, the abundance of acidic amino acids (glutamic acid and aspartic acid) as well as the hydroxyl-containing serine and threonine that has been proposed as characteristic features of the A-domain in *Arabidopsis* Toc159 homologues was also found both putative A-domain sequences of PsToc132 and PsToc120 (Agne et al., 2010; Agne & Kessler, 2010; Chang et al., 2012). Therefore, both extension sequences represent genuine A-domains of PsToc132 and PsToc120 in pea, respectively.

[A]



[B]

Figure 20 | Toc120, a novel Toc159 homologue at the outer envelope of chloroplasts in pea. [A] The putative amino acid sequence of the pea A–domain of Toc120 (PsToc120A) is shown in italics blue. Position of the gene specific primer used in the 5' – RACE id indicated with arrow. The peptides identified by mass spectrometry sequencing are framed.[B] Primary amino acid sequence comparison between PsToc120A and AtToc120A (NP_188284.1) aligned with ClustalW 1.7 Gaps are introduced to maximize identical sequences. Amino acids identical in at least two of the sequences are shaded in black; conserved substitutions are shaded in grey. The GTPase region (G-domain) is underlined in red.

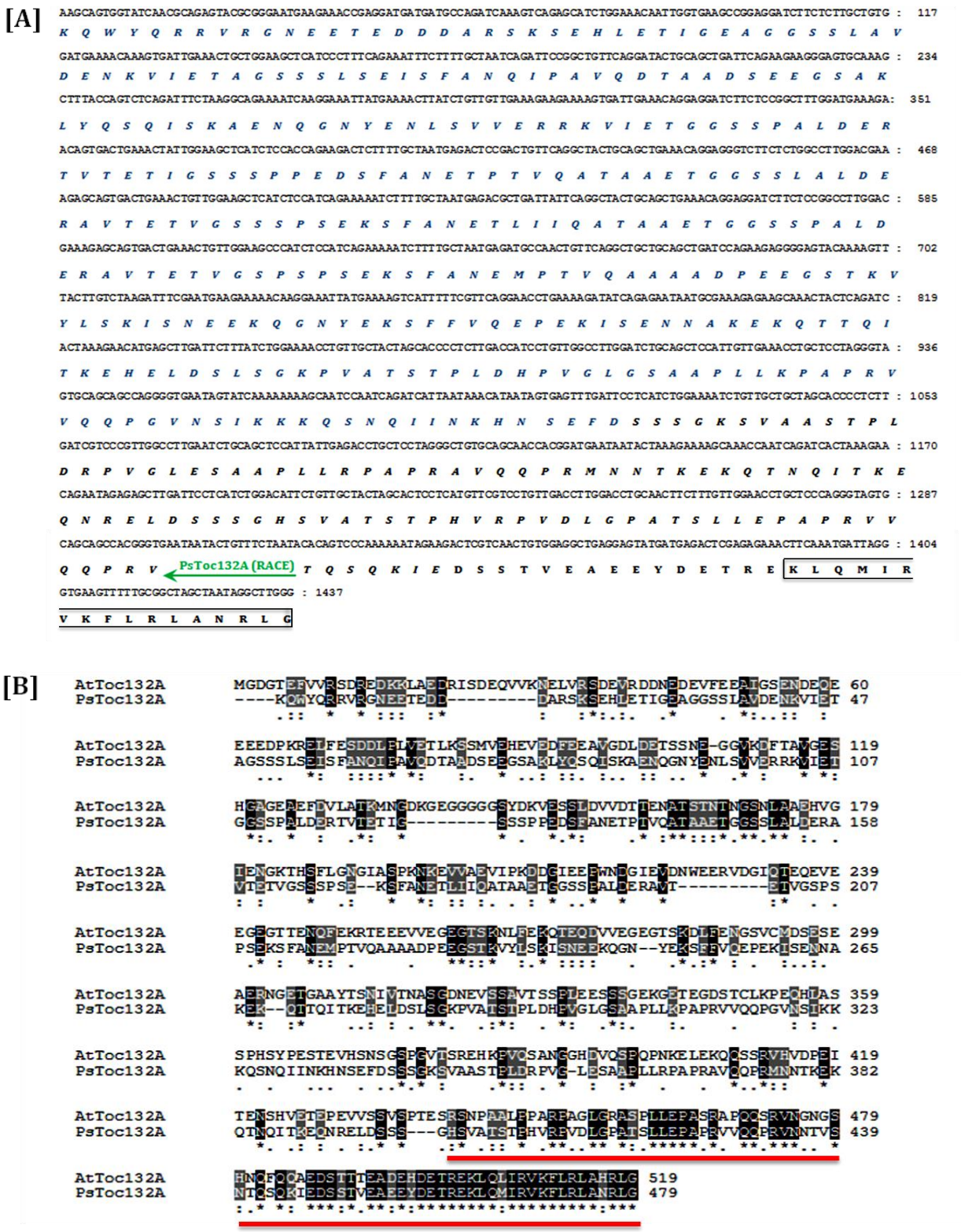


Figure 21 | Toc132, a novel Toc159 homologue at the outer envelope of chloroplasts in pea. [A] The putative amino acid sequence of the pea A–domain of Toc132 (PsToc132A) is shown in italics blue. Amino acid sequence of the G-domain is shown in black and the putative sequence amplified via 5'-RACE is shown in italics black. Position of the gene specific primer used in the 5' – RACE id indicated with arrow. The peptides identified by mass spectrometry sequencing are framed.[B] Primary amino acid sequence comparison between PsToc132A and AtToc132A (NP_179255.1) aligned with ClustalW 1.7. Gaps are introduced to maximize identical sequences. Amino acids identical in at least two of the sequences are shaded in black; conserved substitutions are shaded in grey. The GTPase region (G-domain) is underlined in red.

3.9 Expression and purification of the A-domains of PsToc132 and PsToc120

For characterization of PsToc132 and PsToc120, antisera against PsToc132 and PsToc120 were raised against their relatively divergent A-domains. These N-terminal extensions, encompassing amino acid residues 1-400 and 1-391 of PsToc132 and PsToc120, respectively ($Ps132A_{His}$ and $Ps120A_{His}$; Figure 20A and Figure 21A), were cloned in-frame into pET21(d)+ vector (Novagen, Madison, WI) and heterologously expressed as fusions to C-terminal hexahistidine-tag in *Escherichia coli* (*E. coli*). To this end, several different *E. coli* strains were tested for the optimal expression yield and stability of the respective recombinant A-domains of PsToc132 and PsToc120. The properties of the different *E. coli* strains and helper plasmids are summarized in Table 8.

Table 8 | *Escherichia coli (E. coli)* expression strains tested for the ability to express the acidic domain (A-domain) of PsToc132 and PsToc120

Strains	Plasmid	Properties
BL21 (DE3)	None	Routinely used *E. coli* expression strain
BL21 (DE3) Star	None	Strain with increased mRNA stability
BL21 (DE3) / pLys S	pLys S	Contain a T7 phage lysozyme which represses expression from a T7 promoter until IPTG induction
BL21 (DE3) / pRosetta	pRosetta	Correct for codon bias
BL21 (DE3 / pMICO	pMICO	Correct for codon bias and tighten induction control

All attempts to propagate $Ps132A_{His}$ in the *E. coli* system were unsuccessful despite several different combinations of parameters used; bacterial strains, helper plasmids, culture medium, expression temperature and expression duration (data not shown). The very low level of expression of $Ps132A_{His}$ is probably due to its susceptibility to degradation and aggregation in the cultured medium, as well as its toxicity to *E. coli.* For this reason and due to time limitation, the expression $Ps132A_{His}$ was therefore not further analyzed in this study. Similarly, the expression of $Ps120A_{His}$ was equally challenging. Most of the strains tested (BL21(DE3), BL21(DE3) Star, BL21(DE3)/pRosetta and BL21(DE3)/pMICO) were completely unable to express the recombinant $Ps120A_{His}$. The remaining BL21(DE3)/pLys S strain was able to express

Ps120A$_{His}$ at 12°C overnight as a soluble protein but the amount of overexpressed Ps120A$_{His}$ was barely detectable on a coomassie-stained gel (Figure 22).

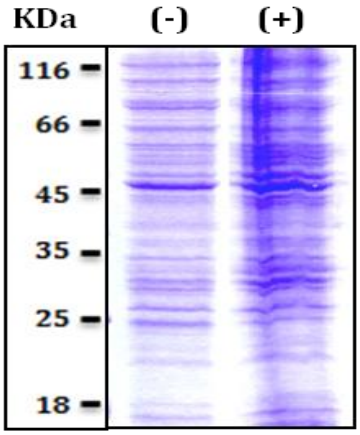


Figure 22. | Expression of Ps120A$_{His}$ protein in *E. coli* BL21(DE3) pLysS cells. Total bacterial lysate from before induction of expression (-) is compared to a sample taken after induction from an overnight incubation at 12°C (+).

Even though the expression of Ps120A$_{His}$ was hardly visible on the coomassie-stained gel, the crude bacterial lysate containing the recombinant protein from a 1L cultured medium were subjected to Ni^{2+}-NTA affinity chromatography for purification nonetheless (see 'Materials and Methods', section 2.2.5). As depicted in Figure 20 (lane 1), the recombinant Ps120A$_{His}$ still could not be purified to homogeneity despite stringent binding and washing conditions during Ni^{2+}-NTA affinity chromatography as well as other combinations of affinity purification (gel filtration, anion exchange, and etc.; data not shown). In order to gain the desired level of purity, the Ni^{2+}-purified Ps120A$_{His}$ was subjected to a further purification step by electroeluting the protein directly from the SDS-polyacrylamide gel, as described in 'Material and Methods', section 2.2.6 (Figure 23, lane 2-6). The resulting protein after electroelution was sufficient in quantity and quality for (I) antibody production and (II) the direct interactions studies of the A-domain of PsToc120 with the precursor protein, FSD1.

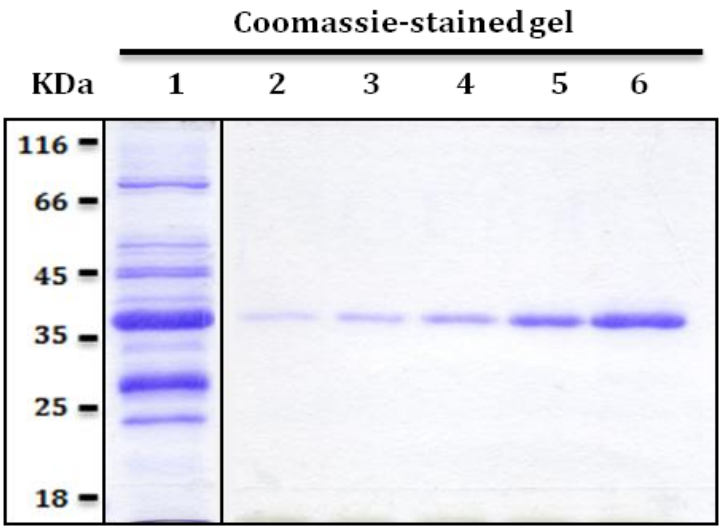


Figure 23 | Purification of the Ps120A$_{His}$ protein. Proteins were fixed and stained with Coomassie Brilliant Blue R-250. Lane 1: crude lysate from the bacterial cultures after Ni^{2+}-affinity purification. Lane 2-6: Purified electroeluted Ps120A$_{His}$ protein.

The theoretical molecular weight of Ps120A$_{His}$ is 43.3 KDa; however, as depicted in Figure 23, the recombinant protein migrated at an estimated molecular weight of 39 KDa on the SDS-PAGE, which is an approximate 5 KDa smaller than expected. The instability of the A-domain and its susceptibility to proteolytic cleavage are the likeliest explanations for the observed aberrant electrophorectic mobility of the recombinant Ps120A$_{His}$. Indeed, the propensity of the A-domain to proteases has been previously reported (Bolter et al., 1998; Chen et al., 2000). Hence, prior to immunization and any further interactions studies, the identity of Ps120A$_{His}$ from the electroeluted sample was confirmed by MS sequencing. The peptide masses revealed a 32% sequence coverage to the recombinant Ps120A$_{His}$, with seven peptides matching solely to the putative A-domain of PsToc120 (Figure 24, sequences marked in red).

MDNGGYDEGE RKRVDVGVSN ESIGGSYEEL KSSEGDEVFE EATDPLKDFN
DRGDAVAATV SVLPSDLVEE IQDDDAEELD SFLEAIGVGD GRVKVSEDEE
EVEVFDFVNG FSGLSRERFE NEDVEYVTPK QNGGILFENG STDKVDYDVD
EFHTYSGSNE EMRNQGAEDK VDYDADEFHT YSGSNEEMRN QGAEAEDLKE
GGLDTELRDD KIIEEQCNAS GGPYSDIQDD KLHIHSARGG LEMEGETLGS
DVVHEDRNGE EIGISDCQSI ECKDYSNDED KGASAETDLG HQETVGEVGG
SSPAMEERTG VETAGRTSLS ENPLVNEIVQ STASAFDEQS IKDYSSKISN
EENQGNHETL PSVEESKRIP ENNAENKETN QIAEEQKREP VSSSVAASTP
LVRPAGLGSA APLLEPAARV VQQSLG

Figure 24 | Identified peptides of Ps120A$_{His}$ from the electroeluted sample. Depicted is the deduced sequence of the A-domain of PsToc120 in pea. Peptides identified in the MS analysis are marked in red.

3.10 Antibody production against the A-domain of PsToc120

After the identity of Ps120A$_{His}$ was confirmed, the heterologously expressed Ps120A$_{His}$ from 3.2.1 was rebuffered and sent to Pineda Antikörper-Service (Berlin) for antibody generation. The antiserum (used as 1:500 dilutions in 1% skimmed milk and 0.05% Tween 20, 0.1% BSA in TBS (100 mM Tris/HCl, pH7.5, 150 mM NaCl) received after the first bleeding (60 days) already showed specific reaction with the heterologously expressed Ps120A$_{His}$ (Figure 25; lane 1-2). No signal is detected in a parallel experiment with the pre-immune serum (Figure 25; lane 3-4). However, a mild cross-reaction was observed with a protein at 66 KDa (Figure 25, marked by asterisk), which presumably is the PsToc75-like protein import channel, PsToc75-V (Eckart et al., 2002). This cross-reactivity of the Ps120A$_{His}$ antiserum might due the fact that the A-domain of PsToc120 and the PsToc75-V could be sharing similar antigenic epitopes.

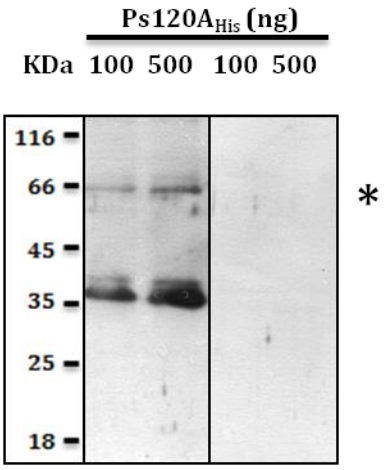


Figure 25 | Ps120A$_{His}$ antiserum specificity test. Antiserum generated by immunization of a rabbit with purified Ps120A$_{His}$ (first two lanes) was compared to the corresponding pre-immune serum (last two lanes) in an immunoblot titration with the increasing amounts of PsToc120 A-domain antigen that were used to generate Ps120A$_{His}$ antiserum. Signals were detected by incubation first with antiserum (first bleeding; 1:500 in TBS-T) followed by a horseradish peroxidase coupled secondary antibody (see Materials and Methods; section 2.2.8). TBST: 1 x Tris buffered saline + 0.05% Tween-20 + 0.1% BSA. *Asterisk: unspecific reaction (see text)*

3.11 PsToc120 is located at the outer envelope of pea chloroplasts

To confirm the localization of PsToc120, sub-fractions from pea chloroplasts were tested for the presence of PsToc120. An envelope fraction, containing the isolated outer and inner envelope of chloroplasts (Figure 26; lane 1 and 2, respectively), a stromal fraction (Figure 26; lane 3), a thylakoid fraction (Figure 26; lane 4), as well as purified pea mitochondria (Figure 26; lane 5) were analyzed by immunoblotting using antiserum that is directed against the A-domain of PsToc120 receptor. The antiserum recognizes specifically a protein of ~ 170 KDa at the outer envelope of chloroplasts (Figure 26; lane 1). Although cross-reactivity could be observed in all the tested fractions, which is presumably caused by the same reason mentioned in the previous section (see 3.10), however, no immune reactive proteins that run at the same molecular mass could be detected in any of the other sub-compartments of chloroplasts or in the mitochondria (Figure 26). The PsToc120 receptor migrates to an apparent molecular weight of an approximate 50 KDa larger than its actual molecular weight (Figure 26; OE fraction). Such aberrant electrophorectic mobility is not unusual for acidic proteins. In fact, the full length *Arabidopsis* Toc159 receptor also demonstrated a similar electrophorectic pattern while the Toc159 receptor devoid of the A-domain (Toc86) migrates as expected in previous studies (Bolter, et al., 1998; Chen, et al., 2000). This unusual migration pattern of the A-domain is thought to be caused by the repulsion of the negatively charged SDS and the acidic residues of the A-domain (Graceffa et al., 1992). Taken together, these results represent the first evidence that of the existence of a multigene family of Toc159 receptor in pea.

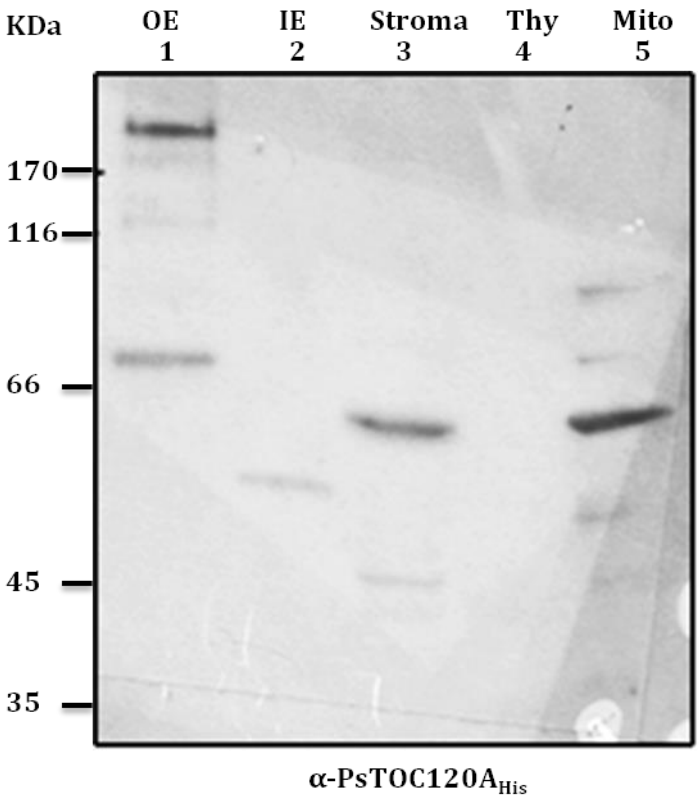


Figure 26 | Localization of the novel Toc159 homologues, Toc120 in pea. Different chloroplast subfractions, 25 µg protein of each outer envelope (OE), inner envelope (IE), stroma, thylakoids (Thy) as well as purified pea mitochondria (Mito) were separated by SDS-PAGE followed by immunodecoration using PsToc120A antisera

3.12 PsToc120 forms distinct TOC complexes the outer envelope of chloroplasts

Reverse genetic analyses demonstrated that the Toc159 gene family encodes a set of selective protein import receptors which assembles into different structurally and functionally unique TOC complexes that are responsible for the distinct protein targeting pathway to the chloroplasts (Bauer, et al., 2000; Ivanova, et al., 2004). These observations raise the question of whether also PsToc120 assembles into diverse TOC complexes in pea. To examine the association of the PsToc120 receptor with the other TOC components, detergent solubilized pea outer envelope membranes were subjected to co-immunoprecipitations using antiserum against PsToc120. As depicted in Figure 24A, both the putative translocation channel, PsToc75, and the small TOC receptor GTPase, PsToc34, could specifically be co-immunoprecipitated in the presence of PsToc120A$_{His}$ antiserum, indirectly demonstrating the close association of the PsToc120 receptor with the two TOC core components in the outer membrane of chloroplasts. Remarkably, PsToc75–like channel protein, PsToc75-V, was found to co-

immunoprecipitate with PsToc120 (Figure 27A). This pea Toc75 paralogue has been previously shown to migrate at the molecular weight of 66 KDa on a SDS-PAGE (Eckart, et al., 2002). Given the non-specific immune reaction that was observed between the PsToc120A$_{His}$ antibody and a 66 KDa protein in Figure 25 (marked by asterisk) and several cases of reported cross-reactivities with PsToc75-V protein (Eckart, et al., 2002), no conclusion could be drawn at the present moment whether the demonstrated interaction between the two proteins was authentic. As expected, the TPR-domain containing Toc64 receptor was not present in the α-PsToc120A$_{His}$ immunoprecipitates (Figure 27A). This observation correlates nicely with the reported dynamic association of the TPR-receptor with the TOC core complex (Becker et al., 2004b; Schleiff et al., 2003b). Due to the lack of antiserum that could specifically differentiate between PsToc120 and PsToc159 at the present moment, the co-immunoprecipitation with PsToc159 was excluded in this experiment. The sensitivity and specificity of the co-immunoprecipitation assay was demonstrated by a parallel control assay with the pre-immune serum (data not shown).

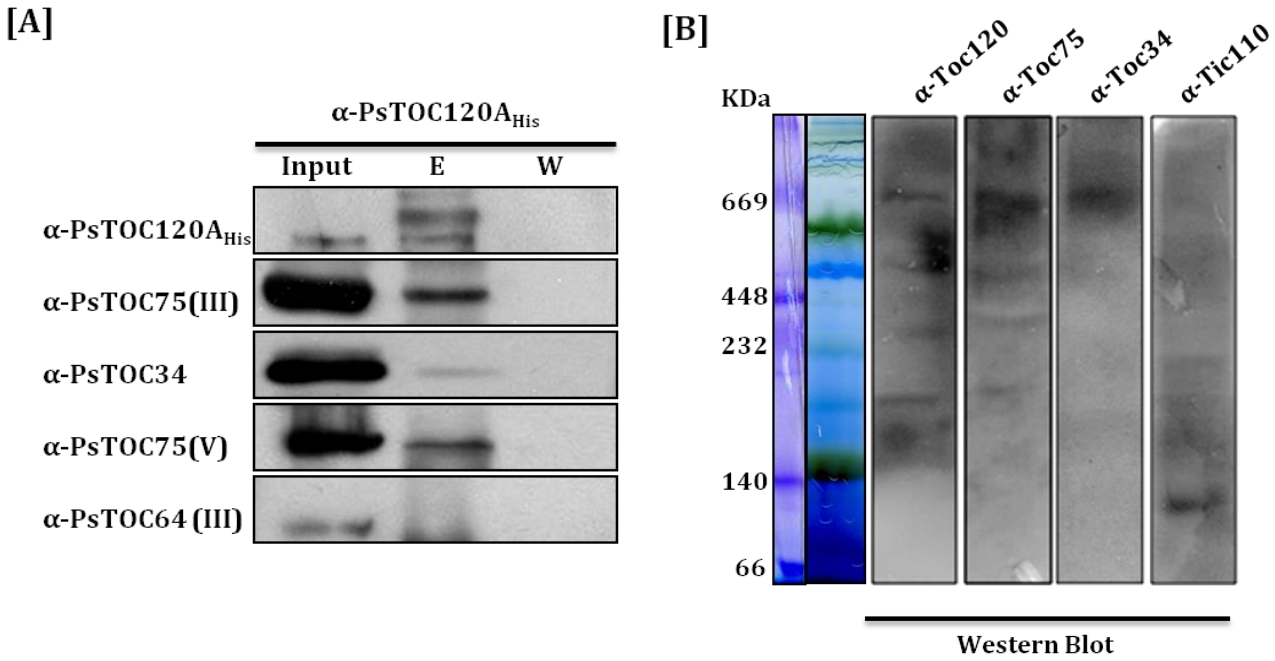


Figure 27 | PsToc120 is an outer envelope protein of chloroplasts and associated with the TOC core translocon. [A] Solubilized outer envelope vesicles were solubilized and incubated with PsToc120A antiserum. The input (5%), wash (5%) and eluted fractions (100%) were separated by a 12.5% SDS-PAGE, transferred to a PVDF membrane and immunodecorated with the indicated antisera.[B] BN-PAGE (5–13.5%) of chloroplasts isolated from pea leaves (20 µg Chlorophyll each) in the presence of protease inhibitor cocktail. An unstained gel lane indicating the major thylakoidal complexes and immunoblots with α-Toc120, α-Toc75, α-Toc34, and α-Tic110. Molecular mass standard markers are tyroglobulin 696, ferritin 448, catalase 232, lactate dehydrogenase 232, and albumin 66 in KDa, respectively.

In addition to *in vitro* co-immunoprecipitation assay, the complex formation between PsToc120 and the other TOC components was also assessed under native conditions. For this purpose, detergent solubilized chloroplast extracts were used and were resolved using a 5-13% non-denaturing BN-PAGE in the presence of protease inhibitor cocktails. The individual outer envelope proteins were identified using their respective antisera. Figure 27B shows an immunoblot analysis of the first dimension BN-PAGE analysis of the chloroplast extracts. All three proteins, PsToc120, PsToc34 and PsToc75, migrated to identical positions on the BN-PAGE in the approximate range of 700-800 KDa. These observed interactions between PsToc120 and the two TOC core components are in good agreement with the co-immunoprecipitation results depicted in Figure 24A. Additionally, a minor fraction of the component of the inner envelope membranes, PsTic110 was observed in the 100 KDa region that most probably corresponds to the monomer of PsTic110 (Figure 27B, the very right panel). This mobility differences suggest that the TOC and TIC complexes are not in direct association under the applied conditions and that the interactions observed in Figure 27B are specific to the outer envelope components. Taken together, the data from both *in vitro* and *in* vivo experiments suggest that the novel receptor, PsToc120, indeed forms a distinctive TOC core complex with an approximate molecular mass of ~800 KDa under native conditions together with PsToc75 and PsToc34.

3.13 PsToc120 interacts specifically with the precursor protein FSD1

After characterizing the localization of PsToc120 and establishing the fact about the close association of PsToc120 with the TOC core complex, the receptor function of PsToc120 was addressed. In order to determine a receptor function, increasing amounts of recombinant PsToc120A_{His} were added to an *in vitro* import assay containing the ^{35}S radiolabeled FSD1 and isolated chloroplasts. A PsToc120 concentration dependent inhibition of the precursor protein, FSD1 translocation was demonstrated (Figure 28A). Surprisingly, a similar inhibitory effect was observed by the addition of 35^{S}-radiolabeled AtQORH (Figure 28A). These data, therefore, indirectly suggest that PsToc120 might act as the common receptor for both tested precursor proteins. The inhibition of PsToc120 is specific towards FSD1 and AtQORH since an increment of the recombinant receptor does not reduce the import of 35^{S}-radiolabeled pSSU (Figure 28A). The N-terminal region of FSD1 is indispensable for targeting but not sufficient to drive the translocation

of the precursor protein into the chloroplasts (Figure 15 and Figure 16), indicating that the C-terminal region of FSD1 may carry additional sequence motifs that are vital for receptor recognition and/or for the process of protein import itself. To confirm this hypothesis, the precursor protein FSD1 (1-50).mSSU was constructed. FSD1(1-50).mSSU is a chimeric protein composed of the N-terminal 50 amino acids of FSD1 fused to the mature sequence of the small subunit of RuBisCo, and the inhibitory effect of the receptor PsToc120 towards the resulting construct was examined. Astoundingly, an excess of PsToc120 has little or no inhibitory effect on the import of 35^{S}-radiolabeled hybrid FSD1(1-50).mSSU precursor protein while the translocation of the full length FSD1 was completely abolished in the excess of recombinant receptor (Figure 28B). These observations suggest that the C-terminus of FSD1 is not only crucial for the translocation process but it also contains critical sequence information that confers PsToc120-dependence in protein import into chloroplasts.

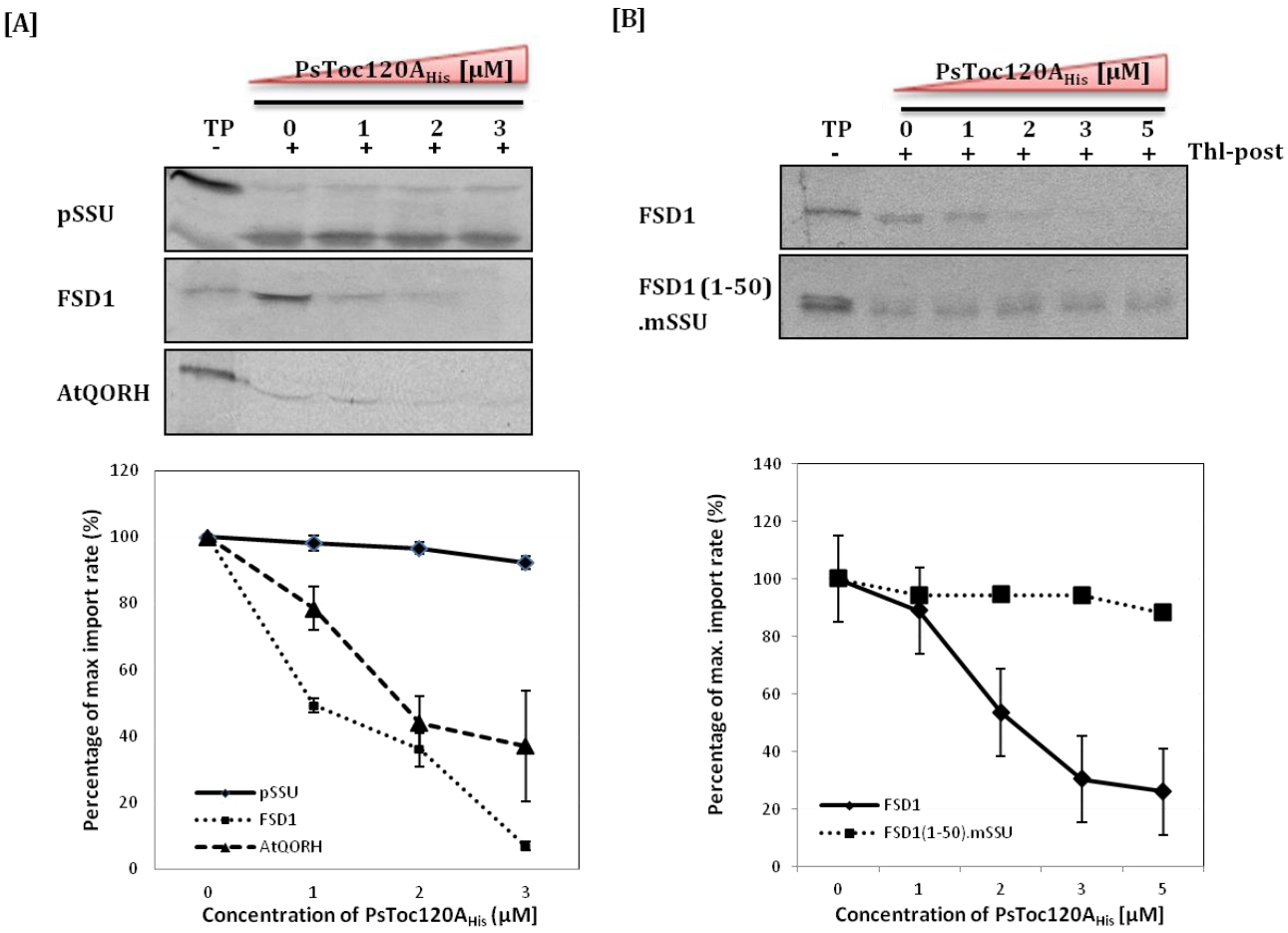


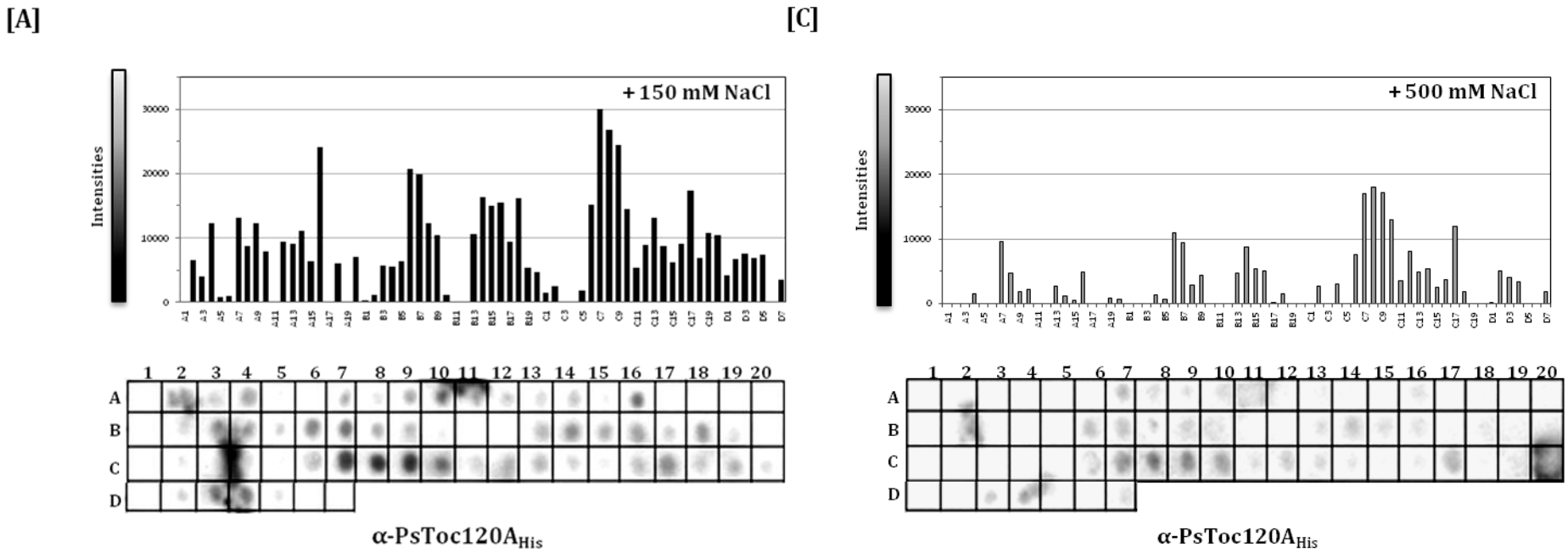
Figure 28 | PsToc120 binds specifically to the C -terminal of FSD1. [A] Radiolabeled 35^S -pSSU, 35^S-AtQORH and 35^S –FSD1 were imported into isolated chloroplasts in the absence or presence of different concentration of recombinant PsToc12A. Increasing amounts of PsToc120A was used for import inhibition. Mock incubations lacked competitor. Quantitative analysis of the data from replicate experiments (n=3) including those presented in [A and B] with standard deviation bars are shown in the lower panel of their respective autoradiograph. [B] Same experiment as in [A], however a chimeric FSD1 which devoid of the C–terminal was used.

3.14 FSD1 contains multiple sequence motifs that contribute to the protein binding specificity to the PsToc120 receptor

The results indicated that the N-terminal region of FSD1 is indispensable for targeting, but not sufficient to drive translocation of the protein into chloroplasts (Figure 15 and 16), while the extreme C-terminus is not necessary for import (Figure 17). This implies that other regions of FSD1 carry additional sequence information that is vital for receptor recognition and/or for the process of protein import. In order to clarify the substrate binding specificity of PsToc120, the precise binding sites between the receptor and the precursor protein were screed using immobilized peptide array, representing the primary sequences of FSD1, for recognition by PsToc120. The peptides were 15 amino acids in length, with each subsequent peptide on the array moved by 3 amino acids in the sequence towards the C-terminal end. This produced an array of 67 peptides, each with a 15 amino acids overlap with the preceding peptides. The FSD1 peptide array was incubated with recombinant PsToc120A_{His} and the binding specificity of the receptor was detected with a specific primary and HRP-conjugated secondary antibodies. As revealed in the peptide scan analysis, PsToc120 demonstrated a high relative binding specificity across the array. A negative control that was performed with only the α-PsToc120A_{His} primary antibody and the HRP-conjugated secondary antibody revealed no binding to the peptides (data not shown). The PsToc120 receptor interacted with several consecutive peptides stretches that are located at both the N-and C-terminus of FSD1 (Figure 29A). Additionally, the binding of the PsToc120 receptor to FSD1 is predominantly mediated by a salt-sensitive electrostatic interaction. The presence of 500 mM NaCl in the binding buffer significantly lowers the binding affinity of the receptor to an approximate 1-2 fold, and in some positions the interactions between PsToc120 and FSD1 were completely abolished (Figure 29C). The amino acid sequences of the PsToc120-binding peptides and the minimal binding motif present in each of the peptides are depicted in Figure 29B. Notably, the identified binding motif of

10-NYVLKPPPFALDALE-24 matches the sequence of the N-terminal FSD1-biotin hybrid protein used in the cross-linking assay (see above). This short peptide sequence is most probably indispensable in the initial targeting and recognition of the precursor FSD1 by the PsToc120 receptor at the chloroplast surface while the binding motifs at the C-terminus might be involved in conferring specificity to FSD1 targeting and its subsequent translocation into the chloroplast.

To further confirm the results from the peptide array analysis several truncations of FSD1 for *in vitro* import experiments were generated. These were lacking only recognition motif six (amino acids 1-178 were still present) or motifs five and six (amino acids 1-131 present), respectively (Figure 30A). Import assays of these proteins revealed that deletion of motif six does not diminish import, which is well in line with the results from the C-terminal deletions displayed in Figure 17. In contrast, the additional loss of motif five resulted in almost complete import inhibition (Figure 30B). This corresponds perfectly to the intensity observed in the peptide array analysis where binding was strongest to the region defined as motif five. These observations together with the previous data from the *in vitro* interactions studies with the recombinant PsToc120A receptor strongly point to a *bona fide* interaction between PsToc120 and FSD1 in specific regions of the substrate protein.



[B]

```
1    MAASSAVTANYVLKPPPFALDALEPHMSKQTLEFHWGKHHRAYVDNLKKQ  50
50   VLGTELEGKPLEHIIHSTYNNGDLLPAFNNAAQAWNHEFFWESMKPGGGG  100
101  KPSGELLALLERDFTSYEKFYEEFNAAAATQFGAGWAWLAYSNEKLKVVK  150
151  TPNAVNPLVLGSFPLLTIDVWEHAYYLDFQNRRPDYIKTFMTNLVSWEAV  200
201  SARLEAAKAASA                                        212
```

Figure. 29 | Interaction of PsToc120A with the substrate FSD1 using peptide spot arrays. [A] the interaction of PsToc120A$_{His}$ and FSD1 was analyzed using a peptide scan approach. Recombinant PsToc120A$_{His}$ was incubated in a final concentration of 5 µg/ml with a peptide library comprising the FSD1 amino acids 1–212. The peptide library contained 67 15-mer peptides, overlapping by 13 residues. Bound protein was detected by immunoblotting using an antibody against PsToc120A$_{His}$. [B] Prominently bound peptides are marked in grey in the FSD1 sequence and the deducted binding regions are numbered form one to six in red. [C] Similar analysis as in [A] except that the peptide scan analysis was carried out under high salt conditions. The bar charts demonstrate the quantified intensity of each spot of the peptide array.

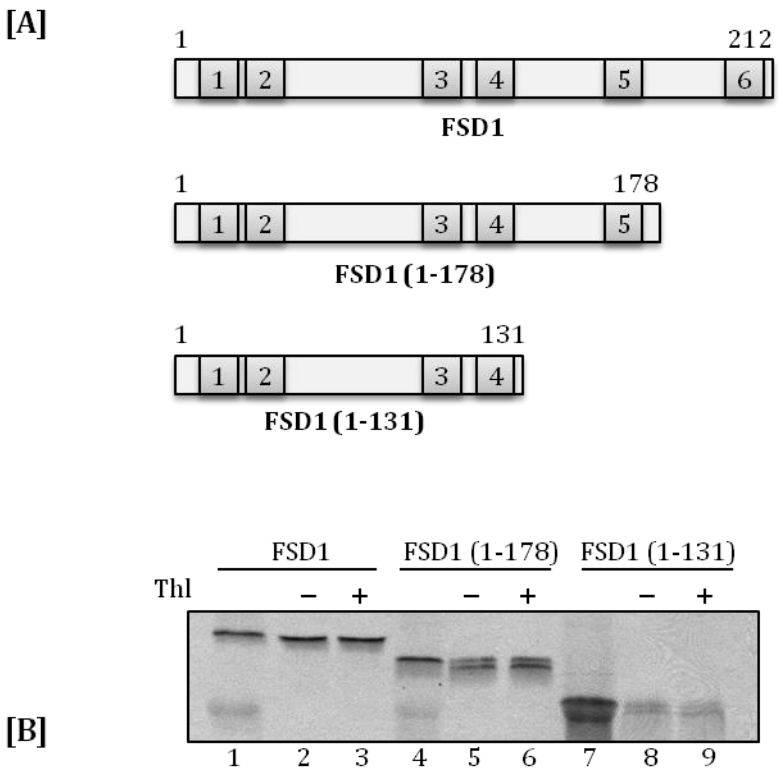


Figure 30 | The fifth recognition motif in the C-proximal end of FSD1 import is essential for binding and import. (A) Schematic representation of the constructs used for import in [B]. [B] *In vitro* import assays of C-terminal truncations are depicted. Lane 1 shows 10% of translation product. Thl (+) indicates thermolysin digestion after the import reaction (lanes 3, 6, 9).

Chapter 4
Discussion

The import of the majority of nucleus-encoded chloroplast proteins is ensured by them common TOC/TIC pathway (Li & Chiu, 2010). These proteins share the feature of possessing N-terminal extensions collectively referred to as the chloroplast transit peptide (cTP) that mediate their delivery to the chloroplast surface and their subsequent translocation across the envelope membranes. It was initially thought that all of the different chloroplast targeted precursor proteins enter the organelle through the TOC/TIC machinery. During the recent years, however, several classes of chloroplast resident proteins showed evidence of divergence in their import pathways *en route* to the chloroplast. The first class of such 'non-canonical' proteins is synthesized in the cytosol as precursor proteins with cleavable signal peptide. These proteins are transferred to the chloroplast via the endomembrane system that complements the TOC/TIC pathway (Nanjo et al., 2006; Radhamony & Theg, 2006; Villarejo et al., 2005). The import of the second class of proteins, on the other hand, is sustained by multiple, differentially regulated TOC and TIC core complexes (Bauer et al., 2000; Ivanova, Smith, Chen, & Schnell, 2004; Kovacs-Bogdan, Benz, Soll, & Bolter, 2011; Kubis et al., 2004). The actual role and mechanism of action of these TOC core complexes will be discussed in section 4.4. The third class of proteins is synthesized in the cytosol at their mature size that carries a non-cleavable cTP. This group includes (i) most of the chloroplast outer membrane proteins, which import seems not to involved the some of the TOC components (for review, see Keegstra & Froehlich, 1999; Schleiff & Klosgen, 2001), and (ii) a minor fraction of non-canonical chloroplast resident proteins identified in the recent proteomic analyses (Ferro et al., 2003; Kleffmann et al., 2004). Two of such proteins are AtQORH (Miras et al., 2002) and Tic32 (Nada & Soll, 2004). Both proteins are located at the inner envelope membrane of the chloroplast. Despite their common structural feature of possessing a non-cleavable cTP, both proteins have very distinctive import patterns (Miras et al., 2002; Miras et al., 2007; Nada & Soll, 2004). These observations thus suggest the operation of more than one novel import pathways

4.1 *In vitro* characterization of chloroplast proteins without cleavable targeting sequence

Although the novel import pathways that mediate post-translational delivery of proteins to the chloroplast via a non-cleavable cTP has already been described to some extend (Miras, et al., 2002; Miras et al., 2007; Nada & Soll, 2004), there are, however, still major questions looming that arose from these observations have yet to be answered. In particular the identity of the proteinaceous import component as well as the precise information that is required for substrate targeting and specificity by these yet uncharacterized translocation machineries, remained to be determined. In the present study, detailed analyses were performed to address these open questions. To search for the ideal bait to 'fish' for components of the novel import pathways, nine tentative non-canonical proteins with confirmed stromal localization and a strong prediction against the presence of a classical cTP from the earlier proteomic analyses of chloroplasts were examined *in vitro*. Six of the proteins in the tested set showed an unambiguous localization in the chloroplast (Table 5). Surprisingly, out of this sextet, three proteins proved to be processed upon translocation into chloroplasts, which was in stark contrast to the computational predictions. Though the import characteristics of these precursor proteins were not investigated any further, it seems quite likely that they employ the general TOC/TIC import pathway. There was, however, one candidate, FSD1, which was not cleaved upon import, as indicated by the lack of post-import processing in the *in vitro* import assay (Figure 8A). FSD1 thus constituted a promising substrate for the present endeavor. The import characteristics of FSD1 was therefore further appraised with regard to targeting information contained within the mature sequence, energy dependence and the engagement of known translocation components (see section 4.2-4.6).

The three remaining candidates in the test set (AtANnAt1, PGR5 and Lip2) revealed no evidence of plastid localization in the *in vitro* import assay (Figure 8B). A localization of PGR5 in the chloroplast is highly feasible since several independent studies have demonstrated the role of PGR5 (Proton gradient regulator, essential for photoprotection) in the regulation of photosystem I cyclic electron flow and photoprotection in *Arabidopsis* (Munekage et al., 2002; Okegawa et al., 2007; Suorsa et al., 2012). It is possible that the import of PGR5 into the chloroplast is hampered by

some additional factors that are missing in the *in vitro* system. Both Lip2 (putative lipoate-protein ligase B) and AtAnnAt1 (stress-responsive calcium-dependent membrane-associated annexin) were only found attached to the chloroplasts. A close association with the golgi apparatus has been reported for AtAnnAt1 (Clark, Lee, Dauwalder, & Roux, 2005). Extraplastidial contamination is often the common demise in the analyses of chloroplast proteome, particularly those comprising of total chloroplast / and total chloroplast envelopes fraction (Ferro et al., 2010; Ferro, et al., 2003). This observation along with the non-plastidial localization in the present study might explain the identification of both proteins (Lip2 and AtAnnAt1) in the course of chloroplast proteomic studies.

The fact that only one protein from the initial nine tentative non-canonical chloroplast proteins that does carry a non-cleavable cTP, clearly hints that the fractions of chloroplast proteins that possess a non-cleavable cTP are not as large as previously anticipated by Kleffmann and co-workers. These findings also highlights the insufficient sensitivity of the existing algorithms applied for transit peptide prediction and need to be optimized, which will be a challenging task for the future. Although the shortcoming of cTP prediction could be compensated by a combination of the available chloroplast localization methods, the fractions of novel non-canonical chloroplast proteins identified in the proteomic studies should be re-accessed with caution as the specificity of cTP prediction is also largely governed by the ambiguity between chloroplast and mitochondrial targeting sequence (Chew & Whelan, 2004) as well as the complexity and the divergence of the chloroplast targeting sequence themselves (Bruce, 2000). Further analyses would be necessary to clarify whether all of the novel non-cTP chloroplast proteins will be targeted to the chloroplasts in a similar fashion as AtQORH and Tic32.

4.2 Import characteristics of FSD1 indicate that it uses some general components but shows distinct properties

In accordance to the literature, chloroplast proteins, especially those destined for the inner compartments of chloroplast generally carry a cleavable cTP, which will be proteolytically removed upon import (Soll & Schleiff, 2004). Chloroplast sub-fractionation as well as the localization of a GFP-tagged FSD1 construct (Figure 9A and 9b) revealed a distinct localization of the protein in the stroma. Therefore, one might

expect that FSD1 is also targeted to the chloroplast in a similar fashion as the other stromal counterparts. The initial *in vitro* characterization of FSD1, revealed, however, no evidence of post-import processing (Figure 8A). Collectively, these findings demonstrate that FSD1 may be the first described protein to be targeted to the stroma of the chloroplast without a classical N-terminal cleavage of the cTP. In order to examine the possibility of FSD1 as substrate for the novel import pathway, the import characteristics of FSD1 was further appraised. Generally, substrates of the novel pathways could be preliminarily distinguished from those of the classical TOC/TIC-mediated import pathway on the basis of the following import characteristics: (i) the presence of a cleavable cTP; (ii) competition by a TOC/TIC-dependent substrate; (iii) engagement with a different TOC core complex; and (iv) requirement of a protease-sensitive surface component.

Evaluation of the import characteristics of FSD1 demonstrated the unique features of FSD1 import. Some of which are reminiscent to the substrates of the general import pathway, such as the ATP-dependence of translocation (Figure 10), the need for protease-sensitive receptors at the plastid surface (Figure 11), and the necessity of the proximal N-terminus for targeting and import (Figure 15 and 16). The energy requirement of chloroplast protein is often closely attribute to the final subplastidial localization of the import substrate as well as the implication of molecular chaperones in the cytosol. Therefore, the larger energy requirement of 3 mM ATP of FSD1 corroborate nicely to the protein final subplastidial localization in the stroma (Theg, Bauerle, Olsen, Selman, & Keegstra, 1989). This observation also hints to a potential involvement of molecular chaperones in the targeting and translocation of FSD1 – a proposition that inadvertently substantiate the necessity of import receptor in the translocation of FSD1. Furthermore, competition with recombinant expressed pSSU, which travels via TOC/TIC-machinery into chloroplasts, resulted in a decreased in the import of FSD1. However, in stark contrast to pSSU, the translocation of FSD1 is never completely inhibited, even at the highest competitor concentration used (Figure 12A). All these infer that FSD1 might engage at least some components of the general import pathway.

On the other hand, FSD1 also demonstrated import patterns that are distinguishable from substrate of the common TOC/TIC – pathway. For instance, the presence of

spermine has just a marginal effect on the import of FSD1 as well as the control protein, AtQORH while the translocation of pSSU is again completely inhibited (Figure 12C). Spermine is a small positively charged aliphatic polyamine that carries 4 positive charges at physiological pH of pH 7.0 (Figure 30) (Dudley, Rosenheim, & Starling, 1926). This overall positive charge of the surface of spermine is somewhat characteristic of a 'typical' chloroplastic transit peptide as described by (von Heijne & Nishikawa, 1991). Indeed, it has been previously shown that spermine is able to induce voltage-dependent block within the Toc75 pore region (Bolter, Soll, Schulz, Hinnah, & Wagner, 1998) in a step that is after recognition but before translocation (Hinnah, Wagner, Sveshnikova, Harrer, & Soll, 2002). Therefore, the presented *in vitro* data hint to a Toc75-independent translocation of FSD1. Assuming that the blockage of spermine is not 100% and the fact that Toc75 is one of the most abundant proteins at the outer envelope of chloroplast, the author will not exclude the implication of the Toc75 translocation channel in FSD1 import. Furthermore, it occurs that FSD1 and Tic32, another non-canonical chloroplast protein which imports mechanism is still elusive, both engaged in independent course *en route* to the chloroplast, since an excess of FSD1 did not hamper Tic32 translocation (Figure 13). The control protein AtQORH that was used for a partially characterized non-canonical import pathway behaved differently from FSD1 in the same assay. Taken together, these findings strongly infer that FSD1 represents a third class of substrate protein that uses a different operational pathway to those of pSSU, AtQORH and Tic32.

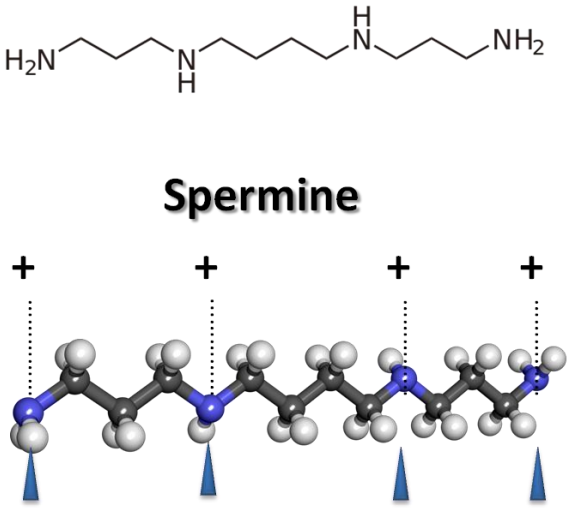


Figure 30|The chemical structure as well as a ball-and-stick model of spermine. Spermine is a biogenic polyamine ($C_{10}H_{26}N_4$) that is formed from spermidine. It is found in a wide variety of organisms and tissues and is essential growth factor in some bacteria (Guirard & Snell, 1964) .

4.3 The proximal N-terminal region of FSD1 interacts specifically with large outer envelope proteins and the identification of novel *Pisum sativum* Toc159 (PsToc159) homologues

Multiple, regulated versions of TOC and TIC core translocons that confer varying specificity towards different class of chloroplast proteins have been reported recently (Bauer, et al., 2000; Ivanova, et al., 2004; Kovacs-Bogdan, et al., 2011; Kubis, et al., 2004), out of which some might be involved in FSD1 import. For instance, one could imagine that FSD1 is translocated into the chloroplast via a TOC core complex containing a Toc120/Toc132 or Toc90 instead of Toc159. Indeed, chemical cross-linking studies using a synthetic peptide that comprises the first 25 amino acid from the extreme N-terminus of FSD1 as well as the subsequent MS analysis indicated that FSD1 did not engage to the Toc159 receptor of the classical TOC core complex but to a 120-/132 KDa component of the chloroplast outer envelope membranes (Figure 19). Thorough scrutinizing of the resulting peptide masses revealed that the FSD1 peptide bound to two proteins with similarities to AtToc132 and AtToc120, respective (Table 6).

Further bioinformatics analysis of the isolated, putative A-domain of PsToc132 and PsToc120 disclosed characteristics that are typical to that of the A-domain of AtToc159 homologues (Figure 31). Notably, the presence of a unusually high number of charge / acidic amino acid residues (Figure 31, see sequence logo residues marked in red) as well as a highly repetitive motif that is comparable to the reported G(D/E)XVV(D/E)X(V/I) consensus sequence in the A-domain of PsToc159 (Figure 31) (Chen, Chen, & Schnell, 2000) could be detected throughout the sequence. Additionally, both isolated A – domains behave as an intrinsically disordered protein (Figure 32). Many disordered regions are associated with protein - protein interaction and surprisingly implicated in an array of regulatory functions in eukaryotic cells (i.e. control of cell cycle and the regulation of transcription and translation) (Dyson & Wright, 2005). In agreement with the concept that reversible protein phosphorylation is central to the regulation of most aspects of cell function (Johnson, 2009), many disordered regions present in proteins are indeed regulated by phosphorylation (Dyson & Wright, 2005).Indeed, both putative A-domains are also enriched in phosphorylation sites, as indicated by the abundance of serine and threonine residues (Figure 31, see sequence logo residues marked in green) – an observation that is consistent with the demonstrated phosphorylated form of both

the A-domain well as the full–length AtToc159 in the recent phosphoproteomic profiling of the *Arabidopsis* proteins (de la Fuente van Bentem & Hirt, 2009; Reiland et al., 2009). Furthermore, immunoblotting using an antiserum against the *E. coli* expressed A-domain of PsToc120 specifically recognizes one protein at the outer envelope membranes of pea chloroplasts but not in the other chloroplast subfractions or pea mitochondria (Figure 26). Although the *in vivo* localization of PsToc132 could not be verified, due to difficulties in propagating the putative A-domain of PsToc132 in *E.coli*, the existence and localization of the PsToc132 receptor in the chloroplast seems likely.

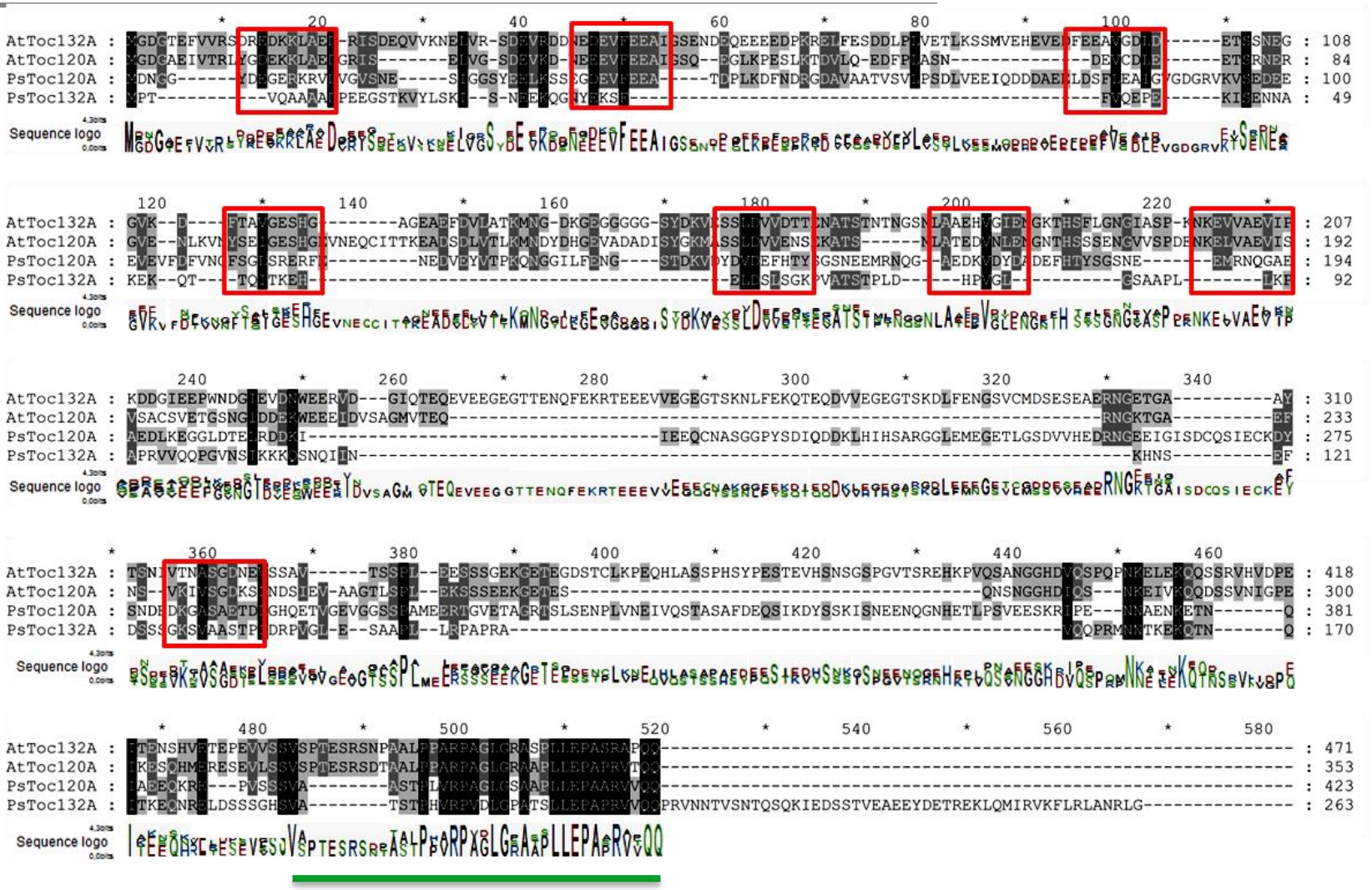
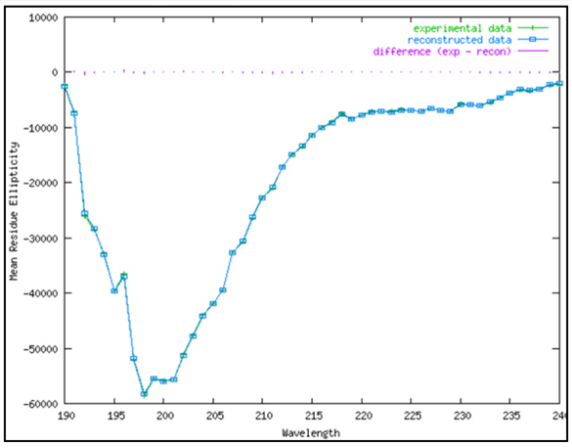


Figure 31| PsToc132 and PsToc120 as novel Toc159 homologue at the outer envelope of chloroplasts in pea. Primary amino acid sequence comparison between PsToc132 and PsToc120 and their respective *Arabidopsis* homologues aligned with ClustalW 1.7. Gaps are introduced to maximize identical sequences. Amino acids identical in at least two of the sequences are shaded in black; conserved substitutions are shaded in grey. The GTPase region (G-domain) is underlined in green. The G(D/E)XVV(D/E)X(V/I) consensus sequence are marked in red box. LOGOs were generated from the sequences of the A-domains of the TOC GTPase receptors.

[A]

Helix1	Helix2	Strand1	Strand2	Turns	Unordered	Total
0.05	0.13	0.13	0.08	0.24	0.38	1.01



[B]

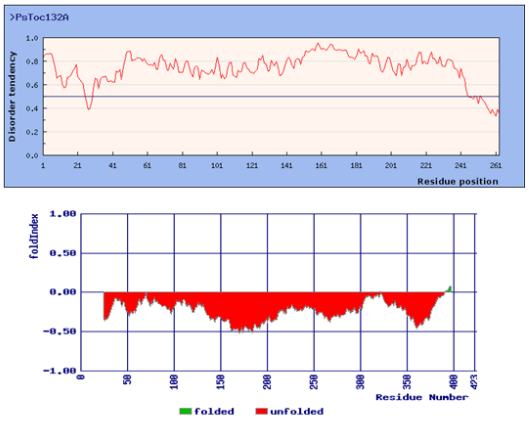


Figure 32| Secondary structure analysis of PsToc132A and PsToc120A. [A]CD spectra of PsToc120A (3 µM) at 25 °C, pH 8.0**.** Spectra demonstrated that PsToc120A are mainly ainly random coil at 25°C and physiological pH, with some residual secondary structure. [B]Disorder prediction of PsToc132A was predicted using IUPred (top) and FoldIndex (bottom).

4.4 Initial characterization of the assembly between the novel PsToc120 and the TOC core components at the chloroplast envelope

The presence of multiple structurally and functionally distinct TOC core complexes in *Arabidopsis* is mainly attributed to the diversities of the AtToc159 and AtToc34 receptor families (Ivanova, et al., 2004; Jelic, Soll, & Schleiff, 2003; Kubis, et al., 2004). Most often, such complexity of the TOC components is hitched as adaptation strategies towards the diverse gene–expression profiles during plastid differentiation in higher plants (Inaba et al., 2005). Therefore, the identification of the PsToc159 receptor homologues and its close association with the precursor protein, FSD1, in the present study supports the prevailing notions of the existence of such similar complexity of TOC components in pea.

Indeed, the PsToc120 receptor was found to form a single complex together with PsToc34 and PsToc75 in the *in vitro* co-immunoprecipitation assay (Figure 27A). Co-migration analysis in of the PsToc12o/34/75 complex in BN-PAGE revealed a molecular mass of ~700-800 KDa and (Figure 24B). The data corroborate with the earlier reported molecular architecture and organization of the TOC core complex (Kikuchi, Hirohashi, &

Nakai, 2006; Schleiff, Soll, Kuchler, Kuhlbrandt, & Harrer, 2003b). Moreover, the demonstrated interaction between PsToc120, PsToc75 and PsToc34 occurs specifically at the outer envelope membranes of the chloroplast, since no co-migration behaviour could be detected between PsToc120 and the inner envelope protein, PsTic110. With the current limitation (antisera that could specifically distinguish the different PsToc159 receptors are still lacking), the current data preclude any firm conclusions to be drawn at the moment. However, the possibility that the observed PsToc120/75/34 complex may represent a complex that is structurally and functionally distinct to those containing PsToc159 could not be excluded.

4.5 The novel PsToc120 receptor interacts specifically with FSD1 via its acidic domain (A-domain)

The current hypothesis predicts that the members of the AtToc159 receptor family are required for the import of a set of different precursor proteins that are required at different stages of plastid development (Bauer, et al., 2000). Essentially, AtToc159 has been proposed to bind specifically with the highly abundant, photosynthetic proteins (Bauer, et al., 2000) while the AtToc132/AtToc120 are more involved in accepting proteins fulfilling house-keeping functions(Kubis et al., 2003). Upon sequence alignment of the respective A-domains of the AtToc159 homologues, it turned out that the highest sequence variability between the different AtToc159 receptors lies within these acidic regions, whereas the G- and M-domains are quite conserved (Ivanova, et al., 2004). Swapping of the respective A-domains between the different AtToc159 isoforms *in planta* altered their selectivity in precursor protein binding (Inoue, Rounds, & Schnell, 2010). This hypothesis was, however, questioned by a recent proteomic study which analyzed the proteome of the *ppi2* mutant plants lacking AtToc159 (Bischof et al., 2011). Many proteins involved in photosynthesis have been found to be present in the mutant plastids, clearly implying that import of these precursor proteins does not exclusively rely on AtToc159. At least, the absence of AtToc159 in the mutant plants, *ppi2,* could be partially compensated by one or the other homologues of the AtToc159 receptor.

FSD1 clearly represents a protein with photosynthesis related function, since the scavenging of reactive oxygen species is highly relevant during active photosynthesis in

the light. Thus, one might have expected to find it prominently bound to PsToc159. However, competition with the recombinant A-domain of PsToc120 resulted in a concentration dependent inhibition of FSD1 translocation (Figure 28A), indicating that the A-domain of PsToc120 receptor interacts specifically to FSD1 and thus blocking the proteins binding to the intrinsic receptor at the chloroplast surface. In contrast, pSSU import remained unaffected. Intriguingly, a similar inhibitory effect was observed for AtQORH. These data suggest that PsToc120 might act as a common receptor for both FSD1 and AtQORH. Although the earlier *in vitro* import assay uncovered the non-proteinaceous nature import of AtQORH (Figure 11), the sensitivity of AtQORH to the excess PsToc120A could be due to fact that AtQORH, having further components at the chloroplast surface that is insensitive to protease treatment. Conversely, it could also be that the precursor protein having a much higher affinity to its import channel, so that it could bypass the receptor.

The fact that FSD1 is associated with the newly identified orthologue of AtToc120, PsToc120, implies that the substrate specificity of Toc132/Toc120 is not restricted to house-keeping proteins as previously anticipated. Rather, these findings call for a model of which the import pathway that is engaged by a protein might rather depend on intrinsic sequence properties than its final function within plastids. Additional factors in the cytosol, such as Hsp70 and Hsp90 that have been previously been implicated in the import process (Jackson-Constan, Akita, & Keegstra, 2001; Qbadou et al., 2006; Zhang & Glaser, 2002), might be involved in determining the specific recognition of the cTP by the different Toc159 receptors. Concerning the composition of the translocon responsible for FSD1 import other than those containing Toc120, one can only speculate The fact that peptides from PsToc75 were also detected in the MS data argues for FSD1 using the PsToc75 import channel. In *Arabidopsis* AtToc120 and AtToc132 associate with AtToc75 and AtToc33/34; this results in the existence of several distinct complexes with the one common element being the channel AtToc75. Thus, one could hypothesize that FSD1 is specifically recognized by PsToc120 (and maybe PsToc132) and then engages PsToc75. This is exemplarily represented in the model in Figure 33. But at this point it is just a hypothesis which awaits confirmation. Another scenario that could be envisioned is that distinct TOC complexes exist in pea - as has been shown in *Arabidopsis* – that consist of different combinations of PsToc159, −132, and −120 with Toc34. These

distinct TOC complexes with different TOC GTPase receptors could have different, but overlapping, substrate specificities accounting for the partial competition of FSD1 for pSSU import. This would be in line with the hypothesized situation in other systems that have already been shown to have multiple Toc159 isoforms.

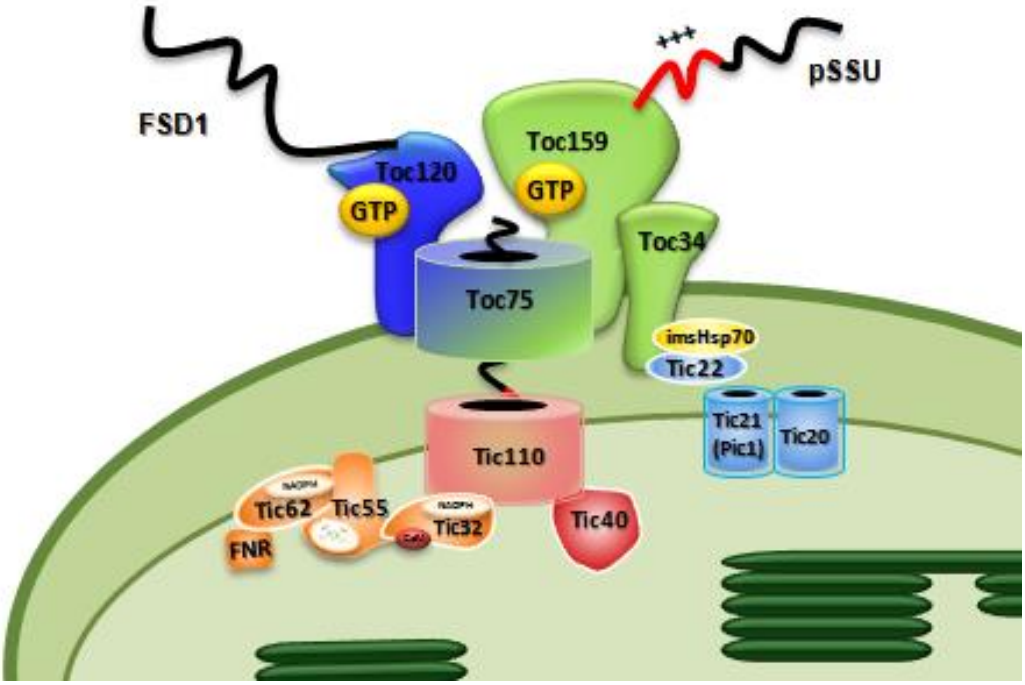


Figure 33| Hypothetical model for dynamic TOC complexes. The general import pathway comprises Toc159, Toc34, and Toc75 as core components. The hypothetical translocon responsible for FSD1 import consists of Toc120, most likely Toc75 and unknown component(s). Green color indicates the pathway taken by canonical substrates, whereas blue signifies an alternative translocon. A mix of both colors indicates participation in both translocation machineries. Thus, Toc75 as the common channel can form a complex with both Toc120 and Toc159, but not necessarily at the same time.

4.6 Multiple sequence motifs in the FSD1 are required for efficient PsToc120 recognition

The classical cleavable N-terminal chloroplast transit peptide contains all information that is necessary and sufficient (in most cases) for receptor recognition as well as translocation across the chloroplast membranes. Despite the divergent nature in their primary sequence, distinct 'homology blocks' throughout the chloroplast transit peptide have been identified (Bruce, 2001). The emerging concept suggests that these multiple 'homology blocks' carry sequence information that is distinct and complementary for targeting to plastids as well as recognition of components of the translocon system at the outer and inner envelope membranes of chloroplasts. Indeed, recent biochemical analyses revealed that specificity of the Toc159-dependent and Toc132/Toc120-dependent pathways in *Arabidopsis* is conferred by multiple sequence elements that are spread across the transit peptide of chloroplasts (Lee et al., 2006; Lee, Lee, Oh, & Hwang, 2009b).

Generally, the transit peptide consist of the following common domain architecture: (i) a membrane-interacting domain at either the N- and / or C-terminal extremities, which is implicated in lipid-mediated binding of the precursor proteins with the chloroplast envelope lipids (Pilon et al., 1995; Pinnaduwage & Bruce, 1996; van't Hof et al., 1993), and (ii) a central region that is involved in the recognition of the import machineries at the respective chloroplast outer and inner envelope membranes (Pilon, et al., 1995). At first glance, the functional organization of the FSD1 sequence seems reminiscent to that of the classical chloroplast transit peptide, with the exception that it is non-cleavable. Indeed, the FSD1 sequence motifs have evolved into several distinct sub-domains to facilitate its proper targeting into the chloroplast as well as recognition by the TOC receptors. Determination of the sequence motifs in FSD1 that confer PsToc120-dependent binding via peptide array analysis revealed specific areas of the protein that are more strongly bound to the receptor, PsToc120 than others (Figure 29A). The reliability of the array could be judged by the detection of the N-terminal peptide that was used for cross-linking as among the most strongly bound regions. From that array, six regions within FSD1 which appear important for binding to the A-domain could be defined. They have an apparent distribution across the protein, including the N- and C-termini. In order to confirm these data, C-terminal truncations of FSD1, which in fact

corroborated the regions essential for binding and import of the protein, were constructed and applied in import assays (Figure 30). While the extreme C-terminus itself is not important (Figure 17), the C-proximal region five which is most strongly labeled in the peptide array proved to be indispensable. The presented data is in line with the previous findings reported by Lee and co-workers (Lee, et al., 2009b) that even in the classical canonical precursor protein, pSSU, multiple sequence elements within the mature part of the protein are required for efficient translocation. These concurrent interactions between the Toc159 receptors and the multiple motifs within transit peptides / precursor proteins is made possible by the natively disordered structure of the A-domains of the Toc159 receptors (Figure 32). Many natively unstructured proteins, in general, have a large surface area under physiological conditions, making them a perfect platform for interaction with several binding partners simultaneously (Dyson & Wright, 2005). The prevalent unordered structure of the A – domain as well as its 50% coverage of the total length of the protein within the Toc159 receptor family (with exception of Toc90) deposits it in a nice position to facilitate interactions with multiple motifs within transit peptides / precursor proteins.

Taken together, the results suggest that the sequence information that is layout across FSD1 contribute collectively to specific interaction with PsToc120 as well as efficient translocation of the precursor protein into the chloroplast. While the C-terminal region of the protein is dispensable for the import process, it is required in addition to the N-terminal region for proper initiation of the PsToc120-dependent pathway as this specificity was abolished when the C-terminal domain was swapped (Figure 25B). The extreme N-proximal region of FSD1, on the other hand, is essential for correct plastid targeting – an observation that substantiate the acquisition of the additional N-terminal extension of plant FSD1 during the course of evolution (Figure 12).

Bibliography

Agne, B., Andres, C., Montandon, C., Christ, B., Ertan, A., Jung, F., et al. (2010). The acidic A-domain of Arabidopsis TOC159 occurs as a hyperphosphorylated protein. *Plant Physiol, 153*(3), 1016-1030.

Albiniak, A. M., Baglieri, J., & Robinson, C. (2012). Targeting of lumenal proteins across the thylakoid membrane. [Research Support, Non-U.S. Gov'tReview]. *J Exp Bot, 63*(4), 1689-1698.

Aronsson, H., & Jarvis, P. (2011). Dimerization of TOC receptor GTPases and its implementation for the control of protein import into chloroplasts. *Biochem J, 436*(2), e1-2.

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*(2), 220-227.

Baldwin, A., Wardle, A., Patel, R., Dudley, P., Park, S. K., Twell, D., et al. (2005). A molecular-genetic study of the Arabidopsis Toc75 gene family. *Plant Physiol, 138*(2), 715-733.

Balsera, M., Goetze, T. A., Kovacs-Bogdan, E., Schurmann, P., Wagner, R., Buchanan, B. B., et al. (2009). Characterization of Tic110, a channel-forming protein at the inner envelope membrane of chloroplasts, unveils a response to Ca(2+) and a stromal regulatory disulfide bridge. *J Biol Chem, 284*(5), 2603-2616.

Bauer, J., Chen, K., Hiltbunner, A., Wehrli, E., Eugster, M., Schnell, D., et al. (2000). The major protein import receptor of plastids is essential for chloroplast biogenesis. *Nature, 403*(6766), 203-207.

Bauer, J., Hiltbrunner, A., & Kessler, F. (2001). Molecular biology of chloroplast biogenesis: gene expression, protein import and intraorganellar sorting. *Cell Mol Life Sci, 58*(3), 420-433.

Bauer, J., Hiltbrunner, A., Weibel, P., Vidi, P. A., Alvarez-Huerta, M., Smith, M. D., et al. (2002). Essential role of the G-domain in targeting of the protein import receptor atToc159 to the chloroplast outer membrane. *J Cell Biol, 159*(5), 845-854.

Bischof, S., Baerenfaller, K., Wildhaber, T., Troesch, R., Vidi, P. A., Roschitzki, B., et al. (2011). Plastid proteome assembly without Toc159: photosynthetic protein import and accumulation of N-acetylated plastid precursor proteins. *Plant Cell, 23*(11), 3911-3928.

Block, M. A., Douce, R., Joyard, J., & Rolland, N. (2007). Chloroplast envelope membranes: a dynamic interface between plastids and the cytosol. *Photosynth Res, 92*(2), 225-244.

Bolter, B., May, T., & Soll, J. (1998). A protein import receptor in pea chloroplasts, Toc86, is only a proteolytic fragment of a larger polypeptide. *FEBS Lett, 441*(1), 59-62.

Browse, J., Warwick, N., Somerville, C. R., & Slack, C. R. (1986). Fluxes through the prokaryotic and eukaryotic pathways of lipid synthesis in the '16:3' plant Arabidopsis thaliana. *Biochem J, 235*(1), 25-31.

Bruce, B. D. (2000). Chloroplast transit peptides: structure, function and evolution. *Trends Cell Biol, 10*(10), 440-447.

Buren, S., Ortega-Villasante, C., Blanco-Rivero, A., Martinez-Bernardini, A., Shutova, T., Shevela, D., et al. (2011). Importance of post-translational modifications for functionality of a chloroplast-localized carbonic anhydrase (CAH1) in Arabidopsis thaliana. *PLoS One, 6*(6), e21021.

Camara, B. (1984). Terpenoid metabolism in plastids : sites of phytoene synthetase activity and synthesis in plant cells. *Plant Physiol, 74*(1), 112-116.

Chang, W. L., Soll, J., & Bolter, B. (2012). The gateway to chloroplast: re-defining the function of chloroplast receptor proteins. *Biol Chem, 393*(11), 1263-1277.

Chen, K., Chen, X., & Schnell, D. J. (2000). Initial binding of preproteins involving the Toc159 receptor can be bypassed during protein import into chloroplasts. *Plant Physiol, 122*(3), 813-822.

Constan, D., Patel, R., Keegstra, K., & Jarvis, P. (2004). An outer envelope membrane component of the plastid protein import apparatus plays an essential role in Arabidopsis. *Plant J, 38*(1), 93-106.

de la Fuente van Bentem, S., Anrather, D., Dohnal, I., Roitinger, E., Csaszar, E., Joore, J., et al. (2008). Site-specific phosphorylation profiling of Arabidopsis proteins by mass spectrometry and peptide chip analysis. [Research Support, Non-U.S. Gov't]. *J Proteome Res, 7*(6), 2458-2470.

Duy, D., Wanner, G., Meda, A. R., von Wiren, N., Soll, J., & Philippar, K. (2007). PIC1, an ancient permease in Arabidopsis chloroplasts, mediates iron transport. *Plant Cell, 19*(3), 986-1006.

Dyson, H. J., & Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol, 6*(3), 197-208.

Firlej-Kwoka, E., Strittmatter, P., Soll, J., & Bolter, B. (2008). Import of preproteins into the chloroplast inner envelope membrane. [Research Support, Non-U.S. Gov't]. *Plant Mol Biol, 68*(4-5), 505-519.

Folkes, B. F. (1970). The physiology of the synthesis of amino acids and their movement into the seed proteins of plants. *Proc Nutr Soc, 29*(1), 12-20.

Franssen, F. M., Sauerwein, H. P., Ackermans, M. T., Rutten, E. P., Wouters, E. F., & Schols, A. M. (2011). Increased postabsorptive and exercise-induced whole-body glucose production in patients with chronic obstructive pulmonary disease. *Metabolism, 60*(7), 957-964.

Gas, E., Flores-Perez, U., Sauret-Gueto, S., & Rodriguez-Concepcion, M. (2009). Hunting for plant nitric oxide synthase provides new evidence of a central role for plastids in nitric oxide metabolism. *Plant Cell, 21*(1), 18-23.

Gerrits, N., Turk, S. C., van Dun, K. P., Hulleman, S. H., Visser, R. G., Weisbeek, P. J., et al. (2001). Sucrose metabolism in plastids. *Plant Physiol, 125*(2), 926-934.

Gutensohn, M., Schulz, B., Nicolay, P., & Flugge, U. I. (2000). Functional analysis of the two Arabidopsis homologues of Toc34, a component of the chloroplast protein import apparatus. *Plant J, 23*(6), 771-783.

Heins, L., Mehrle, A., Hemmler, R., Wagner, R., Kuchler, M., Hormann, F., et al. (2002). The preprotein conducting channel at the inner envelope membrane of plastids. [Research Support, Non-U.S. Gov't]. *Embo J, 21*(11), 2616-2625.

Hernandez Torres, J., Maldonado, M. A., & Chomilier, J. (2007). Tandem duplications of a degenerated GTP-binding domain at the origin of GTPase receptors Toc159 and thylakoidal SRP. *Biochem Biophys Res Commun, 364*(2), 325-331.

Hiltbrunner, A., Grunig, K., Alvarez-Huerta, M., Infanger, S., Bauer, J., & Kessler, F. (2004). AtToc90, a new GTP-binding component of the Arabidopsis chloroplast protein import machinery. *Plant Mol Biol, 54*(3), 427-440.

Hirsch, S., Muckel, E., Heemeyer, F., von Heijne, G., & Soll, J. (1994). A receptor component of the chloroplast protein translocation machinery. *Science, 266*(5193), 1989-1992.

Hirsch, S., & Soll, J. (1995). Import of a new chloroplast inner envelope protein is greatly stimulated by potassium phosphate. *Plant Mol Biol, 27*(6), 1173-1181.

Hofmann, N. R., & Theg, S. M. (2003). Physcomitrella patens as a model for the study of chloroplast protein transport: conserved machineries between vascular and non-vascular plants. [Research Support, U.S. Gov't, Non-P.H.S.]. *Plant Mol Biol, 53*(5), 621-632.

Inaba, T., Alvarez-Huerta, M., Li, M., Bauer, J., Ewers, C., Kessler, F., et al. (2005). Arabidopsis tic110 is essential for the assembly and function of the protein import machinery of plastids. *Plant Cell, 17*(5), 1482-1496.

Inaba, T., Li, M., Alvarez-Huerta, M., Kessler, F., & Schnell, D. J. (2003). atTic110 functions as a scaffold for coordinating the stromal events of protein import into chloroplasts. *J Biol Chem, 278*(40), 38617-38627.

Inaba, T., & Schnell, D. J. (2008). Protein trafficking to plastids: one theme, many variations. *Biochem J, 413*(1), 15-28.

Inoue, H., Li, M., & Schnell, D. J. (2013). An essential role for chloroplast heat shock protein 90 (Hsp90C) in protein import into chloroplasts. *Proc Natl Acad Sci U S A, 110*(8), 3173-3178.

Inoue, H., Rounds, C., & Schnell, D. J. (2010). The molecular basis for distinct pathways for protein import into Arabidopsis chloroplasts. *Plant Cell, 22*(6), 1947-1960.

Ivanova, Y., Smith, M. D., Chen, K., & Schnell, D. J. (2004). Members of the Toc159 import receptor family represent distinct pathways for protein targeting to plastids. *Mol Biol Cell, 15*(7), 3379-3392.

Jackson, D. T., Froehlich, J. E., & Keegstra, K. (1998). The hydrophilic domain of Tic110, an inner envelope membrane component of the chloroplastic protein translocation apparatus, faces the stromal compartment. *J Biol Chem, 273*(26), 16583-16588.

Jarvis, P. (2008). Targeting of nucleus-encoded proteins to chloroplasts in plants. *New Phytol, 179*(2), 257-285.

Jarvis, P., Chen, L. J., Li, H., Peto, C. A., Fankhauser, C., & Chory, J. (1998). An Arabidopsis mutant defective in the plastid general protein import apparatus. *Science, 282*(5386), 100-103.

Johnson, L. N. (2009). The regulation of protein phosphorylation. *Biochem Soc Trans, 37*(Pt 4), 627-641.

Keegstra, K., & Cline, K. (1999). Protein import and routing systems of chloroplasts. *Plant Cell, 11*(4), 557-570.

Kessler, F., & Blobel, G. (1996). Interaction of the protein import and folding machineries of the chloroplast. *Proc Natl Acad Sci U S A, 93*(15), 7684-7689.

Kessler, F., Blobel, G., Patel, H. A., & Schnell, D. J. (1994). Identification of two GTP-binding proteins in the chloroplast protein import machinery. *Science, 266*(5187), 1035-1039.

Kikuchi, S., Oishi, M., Hirabayashi, Y., Lee, D. W., Hwang, I., & Nakai, M. (2009). A 1-megadalton translocation complex containing Tic20 and Tic21 mediates chloroplast protein import at the inner envelope membrane. *Plant Cell, 21*(6), 1781-1797.

Kirk, J. T. (1971). Chloroplast structure and biogenesis. *Annu Rev Biochem, 40*, 161-196.

Kouranov, A., & Schnell, D. J. (1997). Analysis of the interactions of preproteins with the import machinery over the course of protein import into chloroplasts. *J Cell Biol, 139*(7), 1677-1685.

Kovacs-Bogdan, E., Benz, J. P., Soll, J., & Bolter, B. (2011). Tic20 forms a channel independent of Tic110 in chloroplasts. *BMC Plant Biol, 11*, 133.

Kovacs-Bogdan, E., Soll, J., & Bolter, B. (2010). Protein import into chloroplasts: the Tic complex and its regulation. *Biochim Biophys Acta, 1803*(6), 740-747.

Kubis, S., Patel, R., Combe, J., Bedard, J., Kovacheva, S., Lilley, K., et al. (2004). Functional specialization amongst the Arabidopsis Toc159 family of chloroplast protein import receptors. *Plant Cell, 16*(8), 2059-2077.

Lee, D. W., Kim, J. K., Lee, S., Choi, S., Kim, S., & Hwang, I. (2008). Arabidopsis nuclear-encoded plastid transit peptides contain multiple sequence subgroups with distinctive chloroplast-targeting sequence motifs. *Plant Cell, 20*(6), 1603-1622.

Lee, D. W., Lee, S., Lee, G. J., Lee, K. H., Kim, S., Cheong, G. W., et al. (2006). Functional characterization of sequence motifs in the transit peptide of Arabidopsis small subunit of rubisco. *Plant Physiol, 140*(2), 466-483.

Lee, D. W., Lee, S., Oh, Y. J., & Hwang, I. (2009b). Multiple sequence motifs in the rubisco small subunit transit peptide independently contribute to Toc159-dependent import of proteins into chloroplasts. *Plant Physiol, 151*(1), 129-141.

Lee, K. H., Kim, S. J., Lee, Y. J., Jin, J. B., & Hwang, I. (2003). The M domain of atToc159 plays an essential role in the import of proteins into chloroplasts and chloroplast biogenesis. *J Biol Chem, 278*(38), 36794-36805.

Leister, D. (2003). Chloroplast research in the genomic age. *Trends Genet, 19*(1), 47-56.

Li, H. M., & Chiu, C. C. (2010). Protein transport into chloroplasts. *Annu Rev Plant Biol, 61*, 157-180.

Lopez-Juez, E., & Pyke, K. A. (2005). Plastids unleashed: their development and their integration in plant development. *Int J Dev Biol, 49*(5-6), 557-577.

Lubben, T. H., Donaldson, G. K., Viitanen, P. V., & Gatenby, A. A. (1989). Several proteins imported into chloroplasts form stable complexes with the GroEL-related chloroplast molecular chaperone. *Plant Cell, 1*(12), 1223-1230.

Margulis, L. (1975). Symbiotic theory of the origin of eukaryotic organelles; criteria for proof. *Symp Soc Exp Biol*(29), 21-38.

Martin, W., & Herrmann, R. G. (1998). Gene transfer from organelles to the nucleus: how much, what happens, and Why? *Plant Physiol, 118*(1), 9-17.

Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., et al. (2002). Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc Natl Acad Sci U S A, 99*(19), 12246-12251.

May, T., & Soll, J. (2000). 14-3-3 proteins form a guidance complex with chloroplast precursor proteins in plants. [Comparative Study
In Vitro
Research Support, Non-U.S. Gov't]. *Plant Cell, 12*(1), 53-64.

Miras, S., Salvi, D., Ferro, M., Grunwald, D., Garin, J., Joyard, J., et al. (2002). Non-canonical transit peptide for import into the chloroplast. *J Biol Chem, 277*(49), 47770-47778.

Miras, S., Salvi, D., Piette, L., Seigneurin-Berny, D., Grunwald, D., Reinbothe, C., et al. (2007). Toc159- and Toc75-independent import of a transit sequence-less precursor into the inner envelope of chloroplasts. *J Biol Chem, 282*(40), 29482-29492.

Mokranjac, D., & Neupert, W. (2010). The many faces of the mitochondrial TIM23 complex. [Research Support, Non-U.S. Gov't
Review]. *Biochim Biophys Acta, 1797*(6-7), 1045-1054.

Muckel, E., & Soll, J. (1996). A protein import receptor of chloroplasts is inserted into the outer envelope membrane by a novel pathway. *J Biol Chem, 271*(39), 23846-23852.

Nada, A., & Soll, J. (2004). Inner envelope protein 32 is imported into chloroplasts by a novel pathway. *J Cell Sci, 117*(Pt 17), 3975-3982.

Nanjo, Y., Oka, H., Ikarashi, N., Kaneko, K., Kitajima, A., Mitsui, T., et al. (2006). Rice plastidial N-glycosylated nucleotide pyrophosphatase/phosphodiesterase is transported from the ER-golgi to the chloroplast through the secretory pathway. [Research Support, Non-U.S. Gov't]. *Plant Cell, 18*(10), 2582-2592.

Neuhaus, H. E., & Emes, M. J. (2000). Nonphotosynthetic Metabolism in Plastids. *Annu Rev Plant Physiol Plant Mol Biol, 51*, 111-140.

Nielsen, E., Akita, M., Davila-Aponte, J., & Keegstra, K. (1997). Stable association of chloroplastic precursors with protein translocation complexes that contain proteins from both envelope membranes and a stromal Hsp100 molecular chaperone. *EMBO J, 16*(5), 935-946.

Olsen, L. J., & Keegstra, K. (1992). The binding of precursor proteins to chloroplasts requires nucleoside triphosphates in the intermembrane space. *J Biol Chem, 267*(1), 433-439.

Olsen, L. J., Theg, S. M., Selman, B. R., & Keegstra, K. (1989). ATP is required for the binding of precursor proteins to chloroplasts. *J Biol Chem, 264*(12), 6724-6729.

Pain, D., & Blobel, G. (1987). Protein import into chloroplasts requires a chloroplast ATPase. *Proc Natl Acad Sci U S A, 84*(10), 3288-3292.

Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., et al. (2005). FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. [Research Support, Non-U.S. Gov't]. *Bioinformatics, 21*(16), 3435-3438.

Qbadou, S., Becker, T., Mirus, O., Tews, I., Soll, J., & Schleiff, E. (2006). The molecular chaperone Hsp90 delivers precursor proteins to the chloroplast import receptor Toc64. *EMBO J, 25*(9), 1836-1847.

Qbadou, S., Tien, R., Soll, J., & Schleiff, E. (2003). Membrane insertion of the chloroplast outer envelope protein, Toc34: constrains for insertion and topology. *J Cell Sci, 116*(Pt 5), 837-846.

Reiland, S., Messerli, G., Baerenfaller, K., Gerrits, B., Endler, A., Grossmann, J., et al. (2009). Large-scale Arabidopsis phosphoproteome profiling reveals novel chloroplast kinase substrates and phosphorylation networks. *Plant Physiol, 150*(2), 889-903.

Reumann, S., Inoue, K., & Keegstra, K. (2005). Evolution of the general protein import pathway of plastids (review). *Mol Membr Biol, 22*(1-2), 73-86.

Reyes-Prieto, A., Weber, A. P., & Bhattacharya, D. (2007). The origin and establishment of the plastid in algae and plants. *Annu Rev Genet, 41*, 147-168.

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

Robinson, C., Thompson, S. J., & Woolhead, C. (2001). Multiple pathways used for the targeting of thylakoid proteins in chloroplasts. [Review]. *Traffic, 2*(4), 245-251.

Rudowska, L., Gieczewska, K., Mazur, R., Garstka, M., & Mostowska, A. (2012). Chloroplast biogenesis - correlation between structure and function. *Biochim Biophys Acta, 1817*(8), 1380-1387.

Schleiff, E., Jelic, M., & Soll, J. (2003a). A GTP-driven motor moves proteins across the outer envelope of chloroplasts. *Proc Natl Acad Sci U S A, 100*(8), 4604-4609.

Schleiff, E., Soll, J., Kuchler, M., Kuhlbrandt, W., & Harrer, R. (2003b). Characterization of the translocon of the outer envelope of chloroplasts. *J Cell Biol, 160*(4), 541-551.

Schleiff, E., Soll, J., Sveshnikova, N., Tien, R., Wright, S., Dabney-Smith, C., et al. (2002). Structural and guanosine triphosphate/diphosphate requirements for transit peptide recognition by the cytosolic domain of the chloroplast outer envelope receptor, Toc34. *Biochemistry, 41*(6), 1934-1946.

Schnell, D. J., & Blobel, G. (1993). Identification of intermediates in the pathway of protein import into chloroplasts and their localization to envelope contact sites. *J Cell Biol, 120*(1), 103-115.

Schnell, D. J., Kessler, F., & Blobel, G. (1994). Isolation of components of the chloroplast protein import machinery. *Science, 266*(5187), 1007-1012.

Seedorf, M., Waegemann, K., & Soll, J. (1995). A constituent of the chloroplast import complex represents a new type of GTP-binding protein. *Plant J, 7*(3), 401-411.

Shi, L. X., & Theg, S. M. (2010). A stromal heat shock protein 70 system functions in protein import into chloroplasts in the moss Physcomitrella patens. *Plant Cell, 22*(1), 205-220.

Sohrt, K., & Soll, J. (2000). Toc64, a new component of the protein translocon of chloroplasts. *J Cell Biol, 148*(6), 1213-1221.

Soll, J. (2002). Protein import into chloroplasts. [Research Support, Non-U.S. Gov't Review]. *Curr Opin Plant Biol, 5*(6), 529-535.

Soll, J., & Schleiff, E. (2004). Protein import into chloroplasts. *Nat Rev Mol Cell Biol, 5*(3), 198-208.

Stengel, A., Soll, J., & Bolter, B. (2007). Protein import into chloroplasts: new aspects of a well-known topic. *Biol Chem, 388*(8), 765-772.

Takahashi, M., Konaka, D., Sakamoto, A., & Morikawa, H. (2005). Nocturnal uptake and assimilation of nitrogen dioxide by C3 and CAM plants. *Z Naturforsch C, 60*(3-4), 279-284.

Teng, Y. S., Su, Y. S., Chen, L. J., Lee, Y. J., Hwang, I., & Li, H. M. (2006). Tic21 is an essential translocon component for protein translocation across the chloroplast inner envelope membrane. *Plant Cell, 18*(9), 2247-2257.

Theg, S. M., Bauerle, C., Olsen, L. J., Selman, B. R., & Keegstra, K. (1989). Internal ATP is the only energy requirement for the translocation of precursor proteins across chloroplastic membranes. *J Biol Chem, 264*(12), 6730-6736.

Timmis, J. N., Ayliffe, M. A., Huang, C. Y., & Martin, W. (2004). Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet, 5*(2), 123-135.

Villarejo, A., Buren, S., Larsson, S., Dejardin, A., Monne, M., Rudhe, C., et al. (2005). Evidence for a protein transported through the secretory pathway en route to the higher plant chloroplast. *Nat Cell Biol, 7*(12), 1224-1231.

Waegemann, K., & Soll, J. (1995). Characterization and isolation of the chloroplast protein import machinery. *Methods Cell Biol, 50*, 255-267.

Waegemann, K., & Soll, J. (1996). Phosphorylation of the transit sequence of chloroplast precursor proteins. *J Biol Chem, 271*(11), 6545-6554.

Wallas, T. R., Smith, M. D., Sanchez-Nieto, S., & Schnell, D. J. (2003). The roles of toc34 and toc75 in targeting the toc159 preprotein receptor to chloroplasts. *J Biol Chem, 278*(45), 44289-44297.

Wang, X., Xue, L., Sun, J., & Zuo, J. (2010). The Arabidopsis BE1 gene, encoding a putative glycoside hydrolase localized in plastids, plays crucial roles during embryogenesis and carbohydrate metabolism. *J Integr Plant Biol, 52*(3), 273-288.

Young, M. E., Keegstra, K., & Froehlich, J. E. (1999). GTP promotes the formation of early-import intermediates but is not required during the translocation step of protein import into chloroplasts. *Plant Physiol, 121*(1), 237-244.

Zhang, X. P., & Glaser, E. (2002). Interaction of plant mitochondrial and chloroplast signal peptides with the Hsp70 molecular chaperone. *Trends Plant Sci, 7*(1), 14-21.

Appendix

Table S1| Comparison of sequence identity of A-domain of PsToc120 with selected plastidial Toc132/120 receptor.

Accesions	Description	Max score	Total score	Query cover	E value	Ident*
NP_188284.1	Translocase of chloroplast 120 [*Arabidopsis thaliana*]	47.8	93.9	21%	2e-10	69%
XP_004500736.1	Translocase of chloroplast 132/120 [*Cicer arietinum*]	321	427	98%	2e-94	57%
XP_003594564.1	Translocase of chloroplast [*Medicago truncatula*]	215	215	86%	2e-67	43%
XP_008447970.1	Translocase of chloroplast 120[*Cucumis melo*]	48.1	77.8	15%	3e-10	83%
XP_003540651.2	Translocase of chloroplast 132/120 [*Glycine max*]	129	162	74%	2e-37	35%
XP_008375043.1	Translocase of chloroplast 120 [*Malus domestica*]	60.5	85.5	92%	4e-14	27%
XP_007041900.1	Multimeric translocon complex at the OE membrane, 132 [*Theobroma cacao*]	55.5	93.2	42%	2e-12	41%
XP_004163662.1	Translocase of chloroplast 132, [*Cucumis sativus*]	49.7	93.9	20%	1e-10	54%
XP_006362716.1	Translocase of chloroplast 120, [*Solanum tuberosum*]	46.2	78.9	29%	5e-10	69%

Table S2| Comparison of sequence identity of A-domain of PsToc132 with selected plastidial Toc132/120 receptor.

Accesions	Description	Max score	Total score	Query cover	E value	Ident*
NP_179255.1	Translocase of chloroplast 132 [*Arabidopsis thaliana*]	137	23%	23%	1e-23	68%
XP_004500736.1	Translocase of chloroplast 132/120 [*Cicer arietinum*]	263	45%	45%	1e-48	81%
XP_003594564.1	Translocase of chloroplast [*Medicago truncatula*]	523	99%	99%	3e-97	53%
XP_008447970.1	Translocase of chloroplast 120[*Cucumis melo*]	137	40%	40%	8e-28	66%
XP_003540651.2	Translocase of chloroplast 132/120 [*Glycine max*]	116	17%	17%	2e-32	72%
XP_008375043.1	Translocase of chloroplast 120 [*Malus domestica*]	116	27%	27%	5e-32	52%
XP_007041900.1	Multimeric translocon complex at the OE membrane, 132 [*Theobroma cacao*]	228	61%	61%	3e-30	59%
XP_004163662.1	Translocase of chloroplast 132, [*Cucumis sativus*]	227	59%	59%	9e-28	66%
XP_006362716.1	Translocase of chloroplast 120, [*Solanum tuberosum*]	152	29%	29%	1e-28	64%

* % of amino acid identity was determined using BLASTP (Basic Local Alignment Search Tool), using the following settings: comparison matrix BLOSUM62;, Gap penalties, -11,-1; End-gap penalties, -5,-1 (http://blast.st-va.ncbi.nlm.nih.gov/Blast.cgi)

Figure S1| Bioinformatic analysis of PsToc120 receptor protein in *Pisum sativum* and selected relatives from other plant species. Amino acid sequence alignment of the A-domain of Toc120 receptor from PsToc120 (*Pisum sativum*, deduced), AtToc120 (NP_188284.1*)*, CaToc132/120 (XP_004500736.1), MtToc132/120 (XP_003594564.1), GmToc132/120 (XP_003540651.2), CmToc120 (XP_008447970.1), MdToc120A (XP_008375043.1), TcToc132/120A (XP_007041900.), CsToc132/120A (XP_004163662.1) and StToc120A (XP_006362716.1) Identical and conserved amino acid residues were labeled in various colors, respective.ly. Dashes indicated gaps introduced to optimize the alignment. Sequences were aligned using ClustalW2 software (http://www.ebi.ac.uk/Tools/msa/clustalw2/). The GTPase region is underlined in red. Abbreviations: *Ps, Pisum sativum, At, Arabidopsis thaliana, Ca, Cicer arietinum, Mt, Medicago truncatula, Cm, Cucumis melo, Gm, Glycine max, Md, Malus domestica, Tc, Theobroma cacao, Cs, Cucumis sativus, St, Solanum tuberosum,*

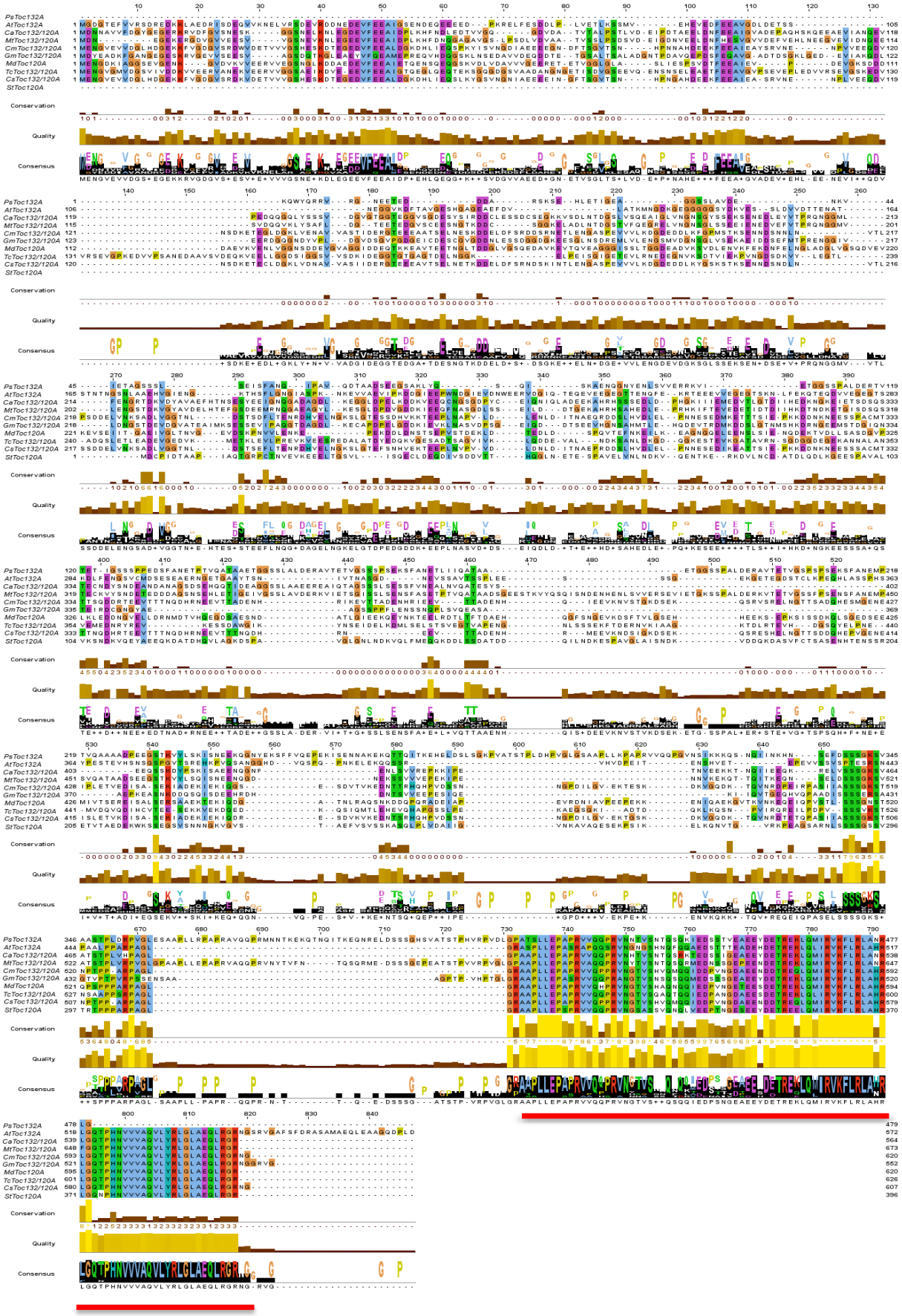


Figure S2| Bioinformatic analysis of PsToc132 receptor protein in *Pisum sativum* and selected relatives from other plant species. Amino acid sequence alignment of the A-domain of Toc120 receptor from PsToc132 (*Pisum sativum*, deduced), AtToc132 (NP_179255.1*),* CaToc132/120 (XP_004500736.1), MtToc132/120 (XP_003594564.1), GmToc132/120 (XP_003540651.2), CmToc120 (XP_008447970.1), MdToc120A (XP_008375043.1), TcToc132/120A (XP_007041900.), CsToc132/120A (XP_004163662.1) and StToc120A (XP_006362716.1) Identical and conserved amino acid residues were labeled in various colors, respective.ly. Dashes indicated gaps introduced to optimize the alignment. Sequences were aligned using ClustalW2 software (http://www.ebi.ac.uk/Tools/msa/clustalw2/) The GTPase region is underlined in red. Abbreviations: *Ps, Pisum sativum, At, Arabidopsis thaliana, Ca, Cicer arietinum, Mt, Medicago truncatula, Cm, Cucumis melo, Gm, Glycine max, Md, Malus domestica, Tc, Theobroma cacao, Cs, Cucumis sativus, St, Solanum tuberosum,*

Table S3 | Peptides masses identified in the chemical-crosslinking analysis of FSD1 with the chloroplast of *Pisum sativum*

Arabidopsis ID	Species_ID	Peptide	Trypsin	Fraction	Description
At4g02510	Ps_contig_mira-and-tgicl-ass_31701	R.LFGFR.S	X	OE	Translocon at the outer envelope membrane of chloroplast 159, GTP-binding, transmembrane receptor.
		K.FKNFK.R	X	OE	
		R.LDLQTR.D	X	OE	
		K.QLAYIVR.R	X	OE	
		R.VFDTPGLK.S	X	OE	
		K.LEDQIALGK.R	X	OE	
		K.ILSEATNISK.T	X	OE	
		R.DMNDLPMLR.S	X	OE	
		K.SSAFEQSYNR.S	X	OE	
		K.AYLEEYDYR.V	X	OE	
		K.SATINSIFGETK.T	X	OE	
		R.LTLVAGR.Q	X	OE	
		R.KVLSTVK.K	X	OE	
		K.LSGQINVR.T	X	OE	
		K.SPPDIVLYVKR.L	X	OE	
		R.SVTSALGPTIWR.N	X	OE	
		K.SSAFEQSYNRK.V	X	OE	
		R.FLEPNSQLLTR.P	X	OE	
		R.SQNDSAYGANVEVR:-	X	OE	
		K.KSPPDIVLYVDR.L	X	OE	
		R.SHIVQQAIGAVGDLR.L	X	OE	
		R.AGTVVSDTDLSEEDKK.K	X	OE	
		R.AGTVVSDTDLSEEDKKK.L	X	OE	
		R.LFSVERPAGLGPSLQTGK.P	X	OE	
		R.VVEVEDESHVGNTVEGEAR.S	X	OE	
		K.AASGAGGEDGGGITLTAQDGRS.L	X	OE	
		R.QSIDILENK.V	X	OE	
		R.LVLVGSTGTVR.S	X	OE	
		K.QWREELKR.M	X	OE	
		K.SSAFEQSYNR.K	X	OE	
		K.MTPILIMLLR.R	X	OE	
		R.QIGEMFSLDAAK.E	X	OE	
At2g**1**6640	Ps_contig_mira-and-tgicl-ass_32415	R.LFVLK.E	X	OE	Multimeric translocon complex in the outer envelope membrane 132, transmembrane receptor,GTP-binding.
		K.FCNFR.R	X	OE	
		K.DLAYTLR.S	X	OE	
		K.IPVSFSGQVTL.D	X	OE	
		K.ATSLGFDMQTVGK.D	X	OE	
		K.EKIPVSFSGQVTK.D	X	OE	
		K.VEDKLIANK.Q	X	OE	
		K.DVNLQMEMASSVK.Y	X	OE	

At2g16640	Ps_contig_mira-and-tgicl-ass_7321	R.LGLAEQLR.G	X	OE	Multimeric translocon complex in the outer envelope membrane 132, transmembrane receptor,GTP-binding
At3g16620	*Ps_contig_mira-and-tgicl-ass_37108*	R.GAGQISIR.L R.YSNLVAR.A K.VVGYSQQLQFGQ.- K.QTACFTNYFCSR.I	X X X X	OE OE OE OE	Multimeric translocon complex in the outer envelope membrane 120, transmembrane receptor,GTP-binding.
	Ps_contig_mira-and-tgicl-ass_5413	K.AYLDEVEYR.E	X	OE	
	Ps_contig_mira-and-tgicl-ass_6441	R.LFVLK.D K.IPFSFSGQVSK.E	X X	OE OE	
	Ps_contig_mira-and-tgicl-ass_7910	R.VNYTVSDTQPR.K R.PAGLGSAAPLLEPAAR.V R.KTEDSSIGEADEYDETR.R	X X X	OE OE OE	
At3g46740	Ps_contig_mira-and-tgicl-ass_36583	R.FGERF.- K.IEFFR.R R.NLQGLNR.S R.HQLTVTK.F K.GNPTVVYR.R K.EKIEFFR.R K.LSIQYLDK.L R.FVNGTIVGSR.N K.ANITENFSR.Q K.ISDILFFDR.N K.GYNMGEIGAAR.N R.NILELAAEIR.I R.MGQGSSYGAGMK.L R.DESNHICSNGQR.V R.EVVCEVVEGDITK.L K.ELESLATCGMFEK.V K.SAEVSTEWSIVPGR.G K.QLLPGHTFNIEAGK.Q K.LGNVVEGNTEGPVVQR.E K.MEYAHPYLDGVDNPR.N R.AEYAVDHNSGTGAVFFR.F R.GGRPTLASLQPGGTITFEHR.N R.VLPNGAISADGPPTTLSGTGIDR.M	X X X X X X X X X X X X X X X X X X X X X X X X	OE OE OE OE OE OE OE OE OE OE OE OE OE OE OE OE OE OE OE OE OE OE OE	Translocon at the outer envelope membrane of chloroplast 75-III, P-P-bond-hydrolysis-driven protein translocation channel

Acknowledgement

First, I would like to thank my PhD director and mentor, Prof. Dr. Jürgen Soll, for giving me the opportunity to be a member of his group in the laboratory of Plant Biochemistry and Physiology at the Faculty of Biology of the Ludwig-Maximilians-University. His unique style of motivation and training has transformed me into a young scientist – I have learned more in the past two years than I thought possible and I am very fortunate indeed to have had the opportunity to be a member of his laboratory. I would also like to thank my supervisor PD Dr. Bettina Bölter for her advice and for sharing her vast knowledge of all things molecular biology-related.

Thanks to Dr. Anna Stengel, Dr. Philipp Benz and Dr. Erika Kovács-Bogdán for all the fruitful discussion in the hallways and for and sharing their passion for science and to other past and present members of the of the Soll's laboratory – Sabine Grahl, Andreas Jungfer, Natalie Schuck – for the great chats and company, and making the lab a great place to work.

I am forever indebted to my family for their constant encouragement and seemingly unending patience. A special thank to my mother for their support and encouragement, both through her wise words and her actions. Witnessing her own personal struggles during mine as a graduate student has been humbling. Finally, a great thanks to Andreas Anselm, Yoovin Kim, Araceli Rivera Ordaz, and all my friends from Munich for reminding me that life outside of the lab *does* exist.

Curriculum Vitae

Personal details

First name:	WaiLing
Family name:	Chang
Date of birth:	3rd of May 1982
Place of birth:	Ipoh, Malaysia
Nationality:	Malaysian

Education

10/2009 – Present **PhD in Molecular Biology and Biochemistry**
Ludwig – Maximilians – University (LMU), Munich, Germany.

10/2007– 09/2009 **Master in Molecular and Cellular Biology**
Ruprecht – Karls – Universität, Heidelberg
Final mark: 1.8 (Good)

06/2002 –08/2005 **Bachelor in Biochemistry**
University of Malaya, Kuala Lumpur, Malaysia
Completed with First Class Honours (CPGA 3.85 of 4.00)

Research Experience

10/2009 – Present **Protein import into Chloroplast, PhD thesis.** Work carried out in the lab of Prof. Jürgen Soll under the supervision of Dr. Bettina Bölter, Department of Biology I (Botany), Biocenter of LMU, Munich, Germany.

12/2008– 09/2009 **"Characterization degradation kinetics of the circadian CLK protein in *Drosophila melanogaster*", Master thesis.** Work carried out in the lab of Dr. Frank Weber, Department of Biochemistry (BZH), Ruprecht – Karls – Universität Heidelberg, Germany.

06/2002 –08/2005 **"Evaluation of *in vitro* antioxidant activity in the traditional medicinal shrubs of Malaysia". Bachelor thesis.** Work carried out in the lab of Assoc. Prof. Dr. Habsah Abdullah, Department of Biochemistry,

University of Malaya, Kuala Lumpur, Malaysia

Publications

Chang WL, Bölter B, Soll J. *A new member of the PsToc159 family contributes to distinct protein targeting pathways in pea chloroplasts.* Front Plant Sci. 2014 May 28;5:239. doi: 10.3389/fpls.2014.00239.

Chang WL, Bölter B, Soll J. *The gateway to chloroplast: re – defining the function of chloroplast protein import receptors.* [Review] Biol. Chem. (2012) 393(11): 1263 - 1277

Zorn D, Hung HC, Maurer C, **Chang WL**, Rademacher C, Young MW, Weber F. *CYCLE is a post translational regulator of CLOCK, controlling stability and nucleo – cytoplasmic shuttling of the Drosophila circadian activator.* (2012) [Submitted for publication]

Hung HC, Maurer C, Zorn D, **Chang WL**, Weber F. Sequential and compartment-specific phosphorylation controls the life cycle of the circadian CLOCK protein. J. Biol. Chem. (2009) 284: 23734-23742.

Presentation at international meetings

April 2012, SFB 594 Mitgliedversammlung - Molekulare Maschinen in Proteinfaltung und Proteintransport, Martinsried, Germany. *„Receptor diversities in Plants"*. Talk

April 2011, EMBO conference series in Protein Transport System: System, mechanisms and medical aspects, Santa Margherita di Pula, Sardinia, Italy.*" Alternative import route into chloroplast"* Poster

March 2009, 60. Mosbacher Kolloquium: Molecular and Cellular Mechanism of Memory, Mosbach, Germany. *"Characterization of the localization – dependent degradation kinetics of Drosophila melanogaster CLK protein."* Poster

Awards and scholarships

2009 – Present LSM Munich student membership, providing support in terms of travel grants and access to scientific/soft skill workshops

2007 – 2009	Molecular and Cellular Biology (MCB) fellowship awarded by the University of Heidelberg.
2005	Amersham Bioscience prize awarded by The Malaysian Society for Biochemistry and Molecular Biology for the best student in Biochemistry thesis on the basis of the final semester examination results

Eidesstattliche Versicherung

Ich versichere hiermit eidesstattlich, dass die vorgelegte Dissertation von mir selbständig und nur unter Verwendung der angegebenen Hilfsmittel und Quellen angefertigt wurde.

München, den 17.07.2014 WaiLing Chang

Erklärung

Hiermit erkläre ich, dass ich zuvor nicht versucht habe, anderweitig eine Dissertation einzureichen oder mich einer Doktorprüfung zu unterziehen. Die vorliegende Dissertation wurde keiner weiteren Prüfungskommission weder in Teilen noch als Ganzes vorgelegt.

München, den 17.07.2014 WaiLing Chang