2016

Characterizing an Alternative Chloroplast Outer Membrane Targeting Signal in Arabidopsis thaliana

Nicholas A. Grimberg
Wilfrid Laurier University, grim0770@mylaurier.ca

CHARACTERIZING AN ALTERNATIVE CHLOROPLAST OUTER MEMBRANE TARGETING SIGNAL IN *Arabidopsis thaliana*

by

Nicholas Grimberg

THESIS
Submitted to the Department of Biology
Faculty of Science
in partial fulfillment of the requirements for the
Master of Science in Integrative Biology
Wilfrid Laurier University

2016

Author's Declaration

I hereby declare that I am the sole author of this thesis. I authorize Wilfrid Laurier University to lend this thesis to other institutions or individuals for the purpose of scholarly research. I understand that my thesis may be made available electronically to the public.

Abstract

Chloroplasts are organelles that are unique to plant and algal cells and are the site of photosynthesis. Though chloroplasts contain their own genome, an estimated 95% of chloroplast proteins are encoded in the nucleus, and therefore rely on post-translational targeting to the organelle. The majority of known chloroplast proteins are targeted to the chloroplast interior by cleavable signals at the N-terminal end of preproteins known as transit peptides. The *t*ranslocon at the *o*uter envelope membrane of *c*hloroplasts (Toc) is a multimeric complex that recognizes and binds N-terminal transit peptides at the cytosolic surface of chloroplasts. Though transit peptides are necessary and sufficient for guiding nuclear-encoded preproteins into the chloroplast interior, the nature of sequence information of transit peptides is not fully understood due to their high divergence in length and composition. Over the last nine years, the number of proteins known or predicted to reside in the chloroplast outer envelope membrane of Arabidopsis has tripled to one hundred and seventeen. Although the functions for some of these outer envelope proteins (OEPs) have been characterized, the precise mechanism of their targeting to the chloroplast outer membrane has not been fully elucidated. Besides Toc75, the targeting mechanisms used by OEPs that have been characterized do not involve an N-terminal transit peptide. The bioinformatics tool ChloroP can be used to predict if amino acid sequences contain an N-terminal transit peptide. Recently, ChloroP analysis and protoplast transient expression assays were used to identify a novel chloroplast targeting signal in the C-terminus of the chloroplast preprotein receptor Toc159 in *Bienertia sinuspersici* (Lung and Chuong,

2012). Toc159 was also shown to lack a canonical transmembrane domain typically present in OEPs. While the unique C-terminal targeting sequence has been partially characterized in Toc159 (Lung et al., 2014), it left open the question of whether this type of signal is unique to Toc159, or if it is used by other OEPs as well. In the current study, to determine if other OEPs use this novel targeting pathway, ChloroP analysis identified eight potential candidates possessing the putative C-terminal targeting signal in Arabidopsis. Transient protoplast expression assays have been performed on OEP18, the protein predicted with the highest ChloroP score, to determine its subcellular localization and the sequences required for its targeting to chloroplasts. The primary purpose of the current study was to establish whether chloroplast outer membrane proteins other than Toc159 use a similar C-terminal targeting signal. Overall, the data in this thesis suggest that some OEPs other than Toc159, such as OEP18, may use this novel targeting pathway.

Acknowledgements

I would like to express deepest gratitude to my supervisors Dr. Matthew Smith and Dr. Simon Chuong. You each gave me the opportunity to achieve my graduate educational goals within your groups. This thesis would not have been possible without your expert guidance, inspiration, advice, and support. I also express my appreciation to Dr. Frédérique Guinel and Dr. Masoud Jelokhani-Niaraki for their insight and advice while serving on my thesis advisory committee.

I would also like to extend my gratitude to my colleagues and friends at Wilfrid Laurier University and the University of Waterloo: Max Pottier, Dustin Sigurdson, and Kelly Yeung, for their continuous support.

I would not have been able to complete this thesis without the love and support from my parents. I would also like to thank my sister, Kristina, for the artistic value that she added to this thesis. I would also like to thank my girlfriend, Stephanie, for proofreading this thesis and for putting up with my rants on the struggles I encountered in the lab throughout my two year MSc journey.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ABRC	Arabidopsis biological resource center
AKR2A	Ankyrin repeat domain-containing protein 2A
APS	Ammonium persulfate
BSA	Bovine serum albumin
cDNA	Complementary deoxyribonucleic acid
COM	Chloroplast outer membrane
CT	C-terminus
cTP	Chloroplast transit peptide
dd	Double distilled
DNA	Deoxyribonucleic acid
ECL	Enhanced chemiluminescence
EGFP	Enhanced green fluorescent protein
ER	Endoplasmic reticulum
FL	Full-length
GTP	Guanosine triphosphate
HSP	Heat shock protein
IDP	Intrinsically disordered protein
LB	Luria-Bertami
MGDG	Monogalactosyldiacyglycerol
NCBI	National Centre for Biotechnology Information
NT	N-terminus

OE	Oxygen-evolving
OEP	Outer envelope protein
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PEX	Peroxisomal biogenesis factor
PG	Phosphatidylglycerol
PTS	Peroxisome targeting signals
PVDF	Polyvinylidene fluoride
RbcS	Ribulose 1,5-bisphosphate carboxylase/oxygenase small-subunit
SA	Signal-anchored
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SPP	Stromal processing peptidase
SRP	Signal recognition particle
TA	Tail-anchored
TAIR	The Arabidopsis information resource
TBS	Tris-buffered saline
TEMED	Tetramethylethylenediamine
TFE	2,2,2-trifluoroethanol
TIC	Translocon at the inner envelope membrane of chloroplasts

TMH	Transmembrane helix
TOC	Translocon at the outer envelope membrane of chloroplasts
TP	Transit peptide
UV	Ultraviolet

1. Introduction

1.1 Plastids

Plastids are organelles surrounded by a double membrane envelope that play crucial roles in signaling and metabolic processes that are necessary for plant development and survival (Lopez-Juez and Pyke, 2005). It is currently accepted that plastids, regardless of the host in which they reside, evolved about 1.5 billion years ago from free-living cyanobacteria through the process of endosymbiosis (Lopez-Juez, 2007). Plastids have the unique ability to differentiate into several variants from an undifferentiated proplastid (Lopez-Juez, 2007) (Figure 1.1). These plastid types are inter-convertible during plant development and in response to different environmental conditions (Wise, 2006).

1.2 Chloroplasts structure and function

Chloroplasts develop from undifferentiated proplastids and are functionally unique from other types of plastids in that they are the site of photosynthesis. The differentiation of proplastids into chloroplasts occurs in tissues, such as leaves and stems, and is triggered by light-dependent pathways (Jarvis, 2008). The maintenance of chloroplast structure is also light-dependent, as prolonged exposure to darkness or inadequate illumination can cause redifferentiation to etioplasts in some cases (Thomas et al., 2009; Taiz and Zeiger, 2010) (Figure 1.1). Chloroplasts are composed of three independent membrane systems (inner and outer envelope membranes and thylakoid membranes) and three internal soluble subcompartments (stroma, thylakoid lumen, and envelope

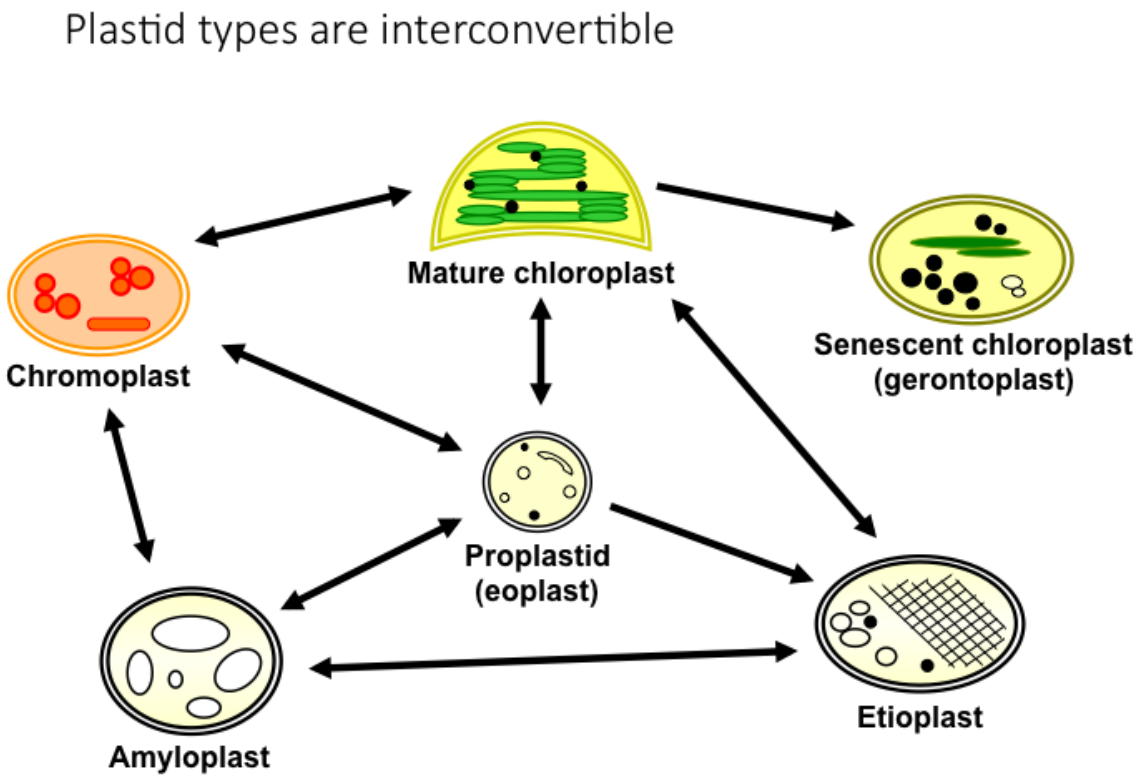


Figure 1.1. All plastid types are derived from proplastids. Many plastid types redifferentiate and interconvert from one type to another in a network of developmental transitions due to environmental changes and the tissues in which they reside.

intermembrane space) (Paila et al., 2015) (Figure 1.2). The outer and inner envelope membranes and thylakoid membranes are made up of a lipid bilayer with phospholipids and a high concentration of galactosyl diacylglycerides (Block et al., 1983; Poincelot, 1976). The outer envelope membrane contains a few β-barrel proteins that are similar to bacterial porins, and are used to translocate specific substrates across the outer envelope (Patel et al., 2008). Due to the permeability of the outer envelope membrane, the inner envelope membrane acts as the primary selectively-permeable barrier between the cytosol and the chloroplast interior (Figure 1.2). The intermembrane space has a buffering capability that is controlled by the permeability of the outer membrane, which is regulated in response to changes in metabolic needs of the organelle (Bölter and Soll, 2001). The thylakoid membranes are the site of the light-dependent reactions of oxygenic photosynthesis (Jarvis, 2008). The vast majority (~95%) of chloroplast proteins are encoded in the nucleus (Jarvis, 2008). Therefore, chloroplast biogenesis and maintenance is dependent on selective targeting and post-translational import of nuclear-encoded precursor proteins (preproteins) that are produced on cytosolic ribosomes (Cline and Henry, 1996).

1.3 Intracellular protein targeting

In a typical plant cell, a series of sophisticated intracellular protein trafficking pathways ensure that several thousand polypeptides are transported to the appropriate organelles and suborganellar compartments. Intracellular targeting of

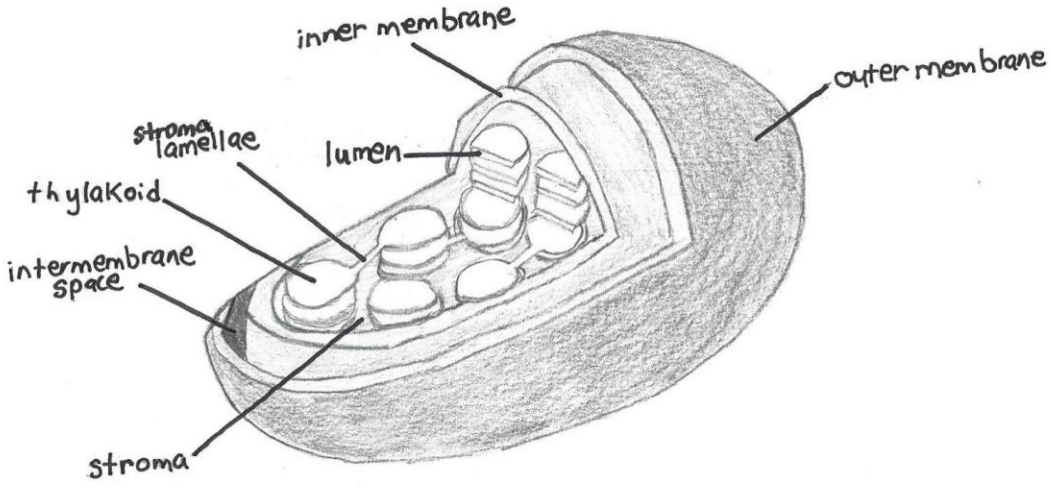


Figure 1.2. Chloroplasts contain a double-envelope membrane with outer and inner layers, between which is the intermembrane space. The thylakoid membrane, which is the site of photosynthesis, is extensively folded and characterized by the presence of thylakoids. Protons are pumped across the thylakoid membrane into the lumen during the light-dependent reaction of photosynthesis. The stroma surrounds the thylakoids and contains the chloroplast genome. Exterior to the outer membrane is the cytosol.

organellar proteins is a fundamental cellular process in eukaryotic cells. Intracellular proteins are delivered to their target organelle either by direct targeting from the cytoplasm or by vesicular trafficking from a donor to a receptor compartment (Lee et al., 2013). Peroxisomes, endoplasmic reticulum (ER), mitochondria, and chloroplasts are organelles that receive their proteins via direct targeting (Lee et al., 2013). Intracellular pathways are aided by targeting information embedded in amino acid sequences of the polypeptides being transported (Bruce, 2000). The names given to these sequences depend on what organelle the polypeptide is targeted to: "peroxisome targeting signals (PTS)" for the peroxisome, "signal peptides" for the endoplasmic reticulum, "nuclear localization signal (NLS)" for the nucleus, "presequences" for the mitochondria, and "transit peptides" for the chloroplasts (Bruce, 2000). In order for an amino acid sequence to be considered a targeting motif it must function in a specific position in the protein, confer residency in a specific organelle, and be disrupted through mutation (Teasdale and Jackson, 1996).

1.3.1 Protein targeting to peroxisomes

Peroxisomes are organelles found in most eukaryotic cells and are involved in various metabolic processes depending on cell type and environment (Johnson and Olsen, 2001). Many plant peroxisome types are involved in the reduction of reactive oxygen species (Corpas et al., 2001). Peroxisomes do not contain their own genome and therefore rely on post-translational import of proteins synthesized on cytosolic ribosomes (Olsen, 1998). Protein import into peroxisomes is necessary for

the regulation of peroxisomal responses to changes in environmental conditions (Goto-Yamada et al., 2015). Targeting sequences can exist on the N- or C-terminus of proteins destined for the peroxisome (Lee et al., 2013). Peroxisome targeting signal 1 (PTS1) is a C-terminal tripeptide sequence that is recognized by the cytosolic import receptor PEX5 and targeted to peroxisomes (Lee et al., 2013). PTS1 sequences comprise a family of sequences that generally conform to a similar pattern of amino acids: small side chain amino acid–basic amino acid–hydrophobic amino acid (Baker et al., 2016). Peroxisome targeting signal 2 (PTS2) is an N-terminal sequence that targets proteins to peroxisomes through recognition by the cytosolic import receptor PEX7 (Lee et al., 2013). In comparison to PTS1, a smaller group of peroxisome-destined proteins use the PTS2 sequence (Lanyon-Hogg et al., 2010). The PTS2 signal sequence is a conserved nonapeptide (Rachubinski and Subramani, 1995).

1.3.2 Protein targeting to the ER

Many eukaryotic proteins are synthesized by ribosomes studded on the cytosolic face of the endoplasmic reticulum (ER) membrane (Hebert and Molinari, 2007). The structure of the ER consists of a number of membrane-enclosed sacs called cisternae. Cisternae are formed by a single envelope membrane, which creates an inner lumen separated from the cytosol (Kim and Hwang, 2013). Resident proteins of the ER lumen, proteins of other compartments of the endomembrane system (e.g. Golgi), and proteins destined to be secreted, contain an N-terminal signal sequence known as a signal peptide. Signal peptides are typically

7-30 amino acids in length and divided into three regions: a positively charged N-terminus, a central hydrophobic domain, and a cleavable C-terminus (Kim and Hwang, 2013). Signal peptides are able to adopt an α-helical conformation to function as a signal sequence (McKnight et al., 1989). Protein translocation into the ER can occur either co- or post-translationally (Fewell and Brodsky, 2000). Cotranslational import is mediated by a molecular chaperone, a signal recognition particle (SRP), which recognizes and binds the signal sequence that emerges from a cytosolic ribosome (Keenan et al., 2001). The SRP guides the nascent polypeptide chain to the ER membrane. Upon arriving at the ER membrane, the SRP binds to an SRP receptor to dock the ribosome at the ER membrane, which allows the nascent polypeptide to be inserted into the ER membrane cotranslationally (Pool et al., 2002).

1.3.3 Protein targeting to mitochondria

Mitochondria are double-membrane organelles that play a crucial role in the metabolism of amino acids, cellular energy conversion, and regulation of apoptosis (Bolender et al., 2008). Though mitochondria contain their own genome, the majority of the over 1,000 resident proteins of mitochondria are nuclear-encoded and imported as precursor proteins from the cytosol (Murcha et al., 2014). Mitochondrial presequences are typically N-terminal signal sequences that target proteins from the cytosol to mitochondria (Mossmann et al., 2012). It was originally believed that the majority of proteins targeted to mitochondria had cleavable presequences (Glaser et al., 1998). The majority of mitochondrial proteins that

were first identified were located in the mitochondrial matrix. However, as more mitochondrial proteins in the outer membrane and inner membrane space have been identified, it is currently believed that possibly up to 50% of mitochondrial proteins do not contain cleavable presequences (Murcha et al., 2014). The majority of presequences of mitochondrial proteins in Arabidopsis range from 20–70 amino acids in length and no conserved targeting motifs have been characterized (Zhang and Glaser, 2002). The amino acid composition of presequences is similar to that of chloroplast transit peptides in that they both typically have an overrepresentation of serine and threonine residues (Zhang and Glaser, 2002). However, there are biochemical differences between presequences and transit peptides. About 84% of mitochondrial presequences in Arabidopsis are predicted to form an α-helix within the first 10 amino acids compared to about 30% of chloroplast transit peptides (Huang et al., 2009). Also, the first 10 amino acids of presequences are generally more hydrophilic than those of transit peptides (Huang et al., 2009).

1.3.4 Transit peptides

Chloroplast transit peptides (cTPs) are N-terminal sequences that target nuclear-encoded preproteins to chloroplasts with high specificity (von Heijne and Nishikawa, 1991). Some chloroplast proteins such as RbcS, OE23, and OE33 possess conserved transit peptide motifs for interaction with cytosolic factors thought to facilitate protein targeting to chloroplasts (Lee et al., 2013). However, the functions of these cytosolic factors have not been clearly demonstrated *in vivo* (Lee et al., 2013). Currently, no clear consensus sequence has been identified since cTPs are

highly divergent in length, composition, and organization (Bruce, 2000). Therefore, what precisely defines cTPs still remains to be fully understood. Though these N-terminal transit peptides are required for preprotein targeting to chloroplasts, some preproteins also require targeting information located in their C-terminal domains, which in some cases have been shown to influence the function of transit peptides (Ko and Ko, 1991). The length of transit peptides can range from 13 to 146 residues, while the majority are 30–80 amino acid residues, which are generally longer than plant mitochondrial presequences (Zhang and Glaser, 2002). Transit peptides are, however, similar to plant mitochondrial presequences in terms of amino acid content. They are both rich in hydrophobic, hydroxylated, and positively charged amino acid residues, and generally lack acidic amino acids (Zhang and Glaser, 2002). The most conserved residue in cTPs is an alanine immediately downstream of the N-terminal methionine (Emanuelsson et al., 2007). A limited number of investigations explains the relative paucity of information available concerning structural aspects of transit peptides. However, this remains an important aspect of cTPs to understand as part of their characterization. One of the few studies on cTP structure indicates that transit peptides are largely unstructured in aqueous environments (Bruce, 1998). This is in agreement with an earlier proposal by von Heijne and Nishikawa (1991) that cTPs are devoid of any regular secondary or tertiary structures and have evolved to form perfect random coils. However, when transit peptides are placed in membrane-mimicking environments such as 2,2,2-trifluoroethanol (TFE) and detergent micelles, analysis by circular dichroism spectrometry reveal that one or more regions of the transit peptide

become α-helical (Wienk et al., 1999). Interestingly, these properties are characteristic of intrinsically disordered proteins (IDPs) (Dyson and Wright, 2005). These data also provide evidence that the chloroplast envelope membrane may play a role in preprotein targeting (Bruce, 2000). Based on the galactosyl diacylglyceride content of the chloroplast envelope membrane, it has been proposed that two possible molecular interactions form the basis of the initial association between transit peptides and the chloroplast surface: an ionic interaction between the basic amino acids of the transit peptide and the anionic phospholipids, and hydrogen bonding between the hydroxylated amino acids of the transit peptide and the galactose headgroups of the glycolipids, particularly the chloroplast-specific monogalactosyldiacyglycerol (MGDG) (Pinnaduwage and Bruce, 1996; Bruce, 2000). However, Aronsson et al. (2008) demonstrated that Arabidopsis mutants with reduced levels of MGDG showed no significant decrease in protein targeting to chloroplasts and import efficiency of cTPs. Therefore, the type of interaction that constitutes the initial association between cTPs and the chloroplast surface may involve other lipid constituents or some unknown factors.

Upon entry into the chloroplast stroma, transit peptides are removed via proteolytic cleavage by the stromal processing peptidase (SPP) (Richter and Lamppa, 1998). The SPP has a broad range of specificity, which reflects the highly divergent nature of transit peptides (Richter and Lamppa, 1998). However, as the database of transit peptides has continued to increase over the past 30 years, so has the efficiency of motif-finding algorithms. These algorithms have led to the identification of a loosely conserved motif (VR↓AAAVXX, where the arrow head

indicates the cleavage site), which led to the development of the neural network-based predictor ChloroP, which is designed to discriminate N-terminal cTPs from other N-terminal sequences (Emanuelsson et al., 1999). ChloroP also predicts the SPP cleavage sites of a given amino acid sequence with 60% accuracy by using a scoring matrix generated by an automatic motif-finding algorithm (Emanuelsson et al., 1999).

1.4 Protein subcellular localization prediction programs

In order to understand the function of a protein, an important first step is to determine its subcellular localization. Besides ChloroP (Emanuelsson et al., 1999), other computational programs have been developed to predict the subcellular localization of proteins including iPSORT (Nakai and Horton, 1999), TargetP (Emanuelsson et al., 2000), TMHMM (Krogh et al., 2001), SignalP (Bendtsen et al., 2004), and Predotar (Small et al., 2004). In parallel, high-throughput experimental approaches have been developed in recent years to determine the subcellular localization of proteins *in vivo* (Emanuelsson et al., 2007). However, it is inevitable that these experiments produce some false-positives and false-negatives. Therefore, computational prediction tools can be used as starting points for identification of organelle-specific proteins and improve the quality of high-throughput data. Despite the accuracy of these computational tools, the results they produce should be taken as suggestions for further experimental analysis for they do have limitations. For example, ChloroP cannot always effectively discriminate between: 1) cTPs and mitochondrial targeting sequences due to the similarity of sorting

signals and 2) those proteins that exhibit dual-targeting to chloroplasts and mitochondria (Emanuelsson et al., 1999). Also, these tools do not provide insight into how transit peptides function in a common import pathway.

1.5 Chloroplast protein import pathway

The highly uncharged N-terminus of transit peptides plays a crucial role in plastid protein import (Lee et al., 2002). About 77% of transit peptides contain a heat shock protein 70 (Hsp70) binding site (H70BS) in the N-terminal region (Chotewutmontri and Bruce, 2015). Hsp70 and Hsp93 are two chaperones that act as plastid translocation motors (Su and Li, 2010). When the expression of Hsp93 is knocked down, plastids can still import preproteins at 40-60% of the wild type level (Kovacheva et al., 2007). This suggests that Hsp70, or potentially another chaperone, functions as the translocation motor for the imported preprotein. Those transit peptides that lack H70BS (Chotewutmontri and Bruce, 2015) or have mutated H70BS (Rial et al., 2003) are still capable of being imported into plastids. Therefore, a dual involvement of Hsp93 and Hsp70 likely exists to bind transit peptides sequentially, simultaneously, or independently, and initiate the import process into chloroplasts (Chotewutmontri and Bruce, 2015). Transit peptides continue to mediate translocation across the chloroplast envelope once preproteins arrive at the chloroplast surface.

1.6 Toc complex

The *t*ranslocon at the *o*uter envelope membrane of *c*hloroplasts (Toc) is a multimeric complex that recognizes and binds N-terminal transit peptides of cytoplasmic preproteins (Chang et al., 2012). The Toc complex works in coordination with the *t*ranslocon at the *i*nner envelope membrane of *c*hloroplasts (Tic) to transport preproteins from the cytosol to the stroma (Li and Chiu, 2010).

The core complex of the Toc complex is made up of two GTP-dependent preprotein receptors, Toc34 and Toc159, and a membrane channel Toc75 (Figure 1.3) (Richardson et al., 2014). These core Toc components are integral membrane proteins (Schleiff et al., 2003). Toc34 and Toc159 each have a GTPase (G-) domain that is capable of binding the transit peptides of newly synthesized preproteins at the chloroplast surface (Richardson et al., 2014). Once bound to the G-domain, preproteins are transferred to Toc75 and translocated across the outer envelope membrane of the chloroplast through sophisticated intermolecular events (Kessler and Schnell, 2002). Toc34 and Toc159 are encoded by multi-gene families, such that different family members assemble in combination with Toc75 to form distinct core Toc complexes with different preprotein selectivities (Kessler and Schnell, 2009). In Arabidopsis, the Toc159 family is comprised of atToc159, atToc132, atToc120, and atToc90, while the Toc34 family contains atToc33 and atToc34 (Smith et al., 2004). It has been hypothesized that the receptors expressed in any given cell type depends on the metabolic needs of the organelle. For example, atToc159 would be expressed if the import of photosynthetic preproteins were required, since atToc159 has a high affinity for a subset of essential photosynthetic proteins.

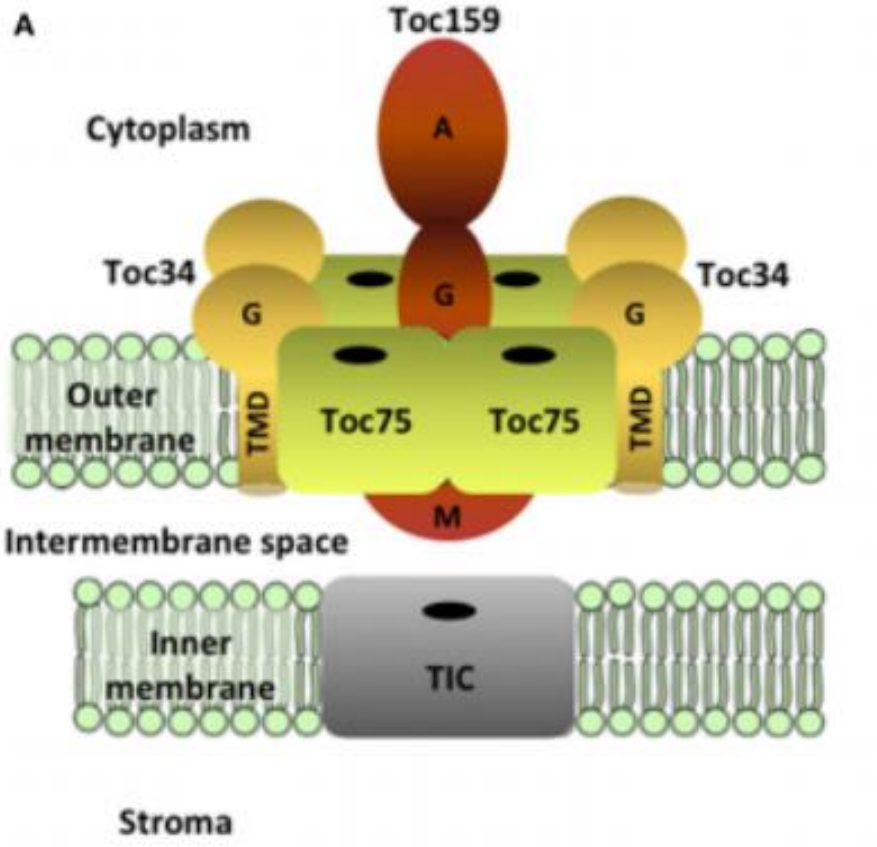


Figure 1.3. GTPase receptors Toc159 and Toc34 assemble with the β-barrel protein channel Toc75 to make up the Toc complex in the chloroplast outer membrane (Richardson et al., 2014).

atToc132, atToc120, and atToc90, on the other hand, have higher affinities for different subsets of proteins, which may include non-photosynthetic proteins (Smith et al., 2004; Dutta et al., 2014).

1.6.1 Models for studying the Toc complex

The majority of Toc components that mediate preprotein import were originally identified and characterized in pea (Smith, 2006) but have been most extensively studied in Arabidopsis (Kubis et al., 2004). More recently, with the completion of the tomato genome sequencing project, Yan et al. (2014) analyzed and cloned Toc GTPase cDNAs from tomato and identified four Toc159 homologues and two Toc34 homologues with high sequence similarity to those of Arabidopsis. This indicates that the tomato is a valid model for further study of preprotein import into chloroplasts (and potentially other plastid types), and furthermore suggests that investigation of the import apparatus may be sufficient, but does not need to be limited to Arabidopsis and pea. Investigation of the import apparatus may also be achieved using the Bienertia system of the Chenopodiaceae family due to its unique dimorphic chloroplasts within a single cell (Lung and Chuong, 2012). The single-cell C_4 species use a unique intracellular compartmentalization of two biochemically and morphologically distinct chloroplasts to accomplish the functions of the Kranz-type C_4 system (Edwards et al., 2004), The differentiation of the dimorphic chloroplasts must involve selective expression and/or accumulation of nuclear encoded proteins (Lung et al., 2012; Erhlinghaeuser et al., 2016). Finding multiple valid models for studying plastid preprotein import is significant, for different systems contain

varying levels of plastid-types. Therefore, new insights are gained on the expression patterns of chloroplast import components, leading towards a more thorough understanding of the chloroplast protein import apparatus functions.

1.6.2 N-terminus of Toc159

Although Toc159 and Toc34 contain homologous GTPase domains, Toc159 is a much larger protein than Toc34 due in part to its large N-terminal acidic (A-) domain. The physicochemical properties of the A-domain are characteristic of intrinsically disordered proteins (IDPs) (Richardson et al., 2009). The demonstration that the A-domain is an IDP led to the suggestion that this domain has a role in preprotein recognition. This proposal was supported by Dutta et al. (2014) who used a yeast two-hybrid approach to identify more preprotein substrates for the Toc159 receptor family. Results from this series of screens demonstrated how the A-domain may have a role in conferring specificity for preprotein subclasses, which interact with the G-domain of Toc159 through their cTPs (Dutta et al., 2014).

1.7 Outer Envelope Protein Targeting Pathways

Outer envelope proteins (OEPs) are a subset of chloroplast proteins that are nuclear-encoded and reside in the chloroplast outer membrane. OEPs are unique from other chloroplast proteins because, aside from Toc75 (Tranel et al., 1995), they do not target to chloroplasts using a transit peptide (Hofmann and Theg, 2005). The mechanisms of targeting for the vast majority of the one hundred and seventeen

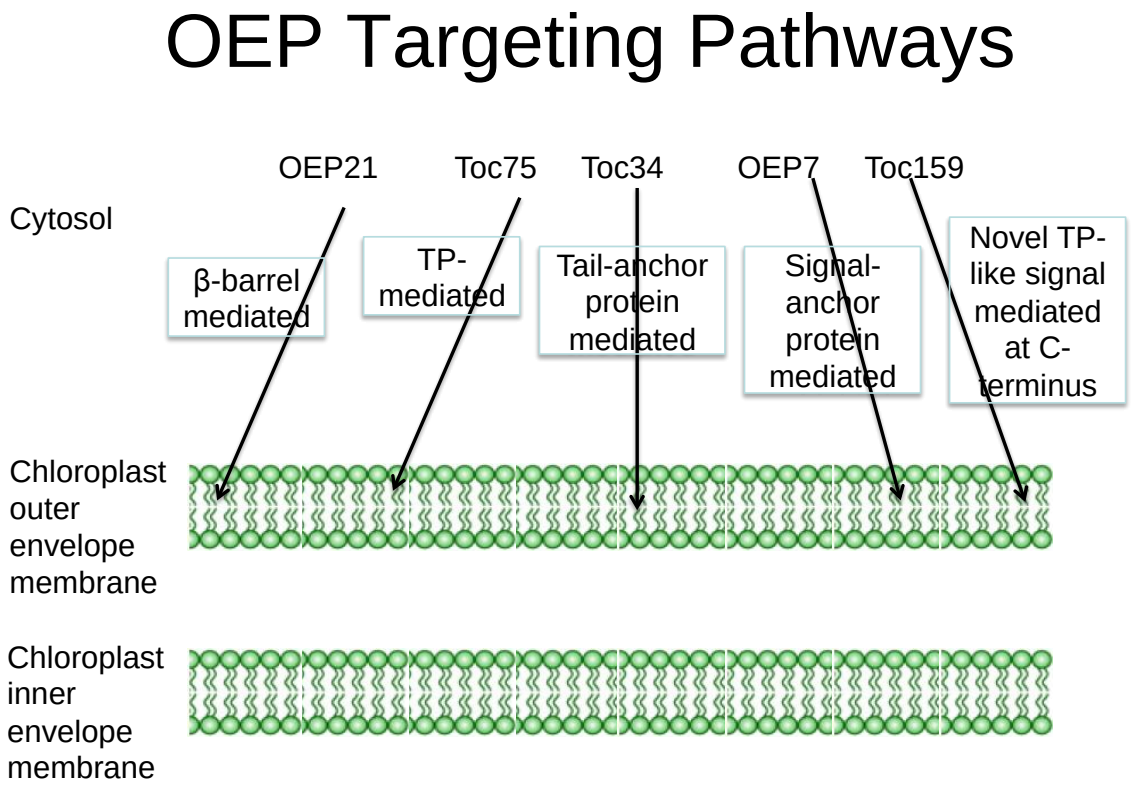


Figure 1.4. Though the mechanisms for the targeting of most OEPs to the chloroplast outer membrane have not been defined, illustrated here are five distinct chloroplast outer membrane targeting mechanisms. OEP21 is an example of a β-barrel that appears to self-insert into the chloroplast outer membrane. Toc34 and OEP7 are representative proteins of types of targeting used by other known OEPs; Toc34 is a tail-anchored protein and OEP7 is a representative of signal-anchored proteins. Toc75 is the only known OEP that uses an N-terminal transit peptide for targeting (Tranel et al., 1995). Toc159 was recently shown to contain targeting information in its C-terminus, but is different from the tail-anchored proteins because it lacks the canonical α-helical transmembrane domain (TMD) (Lung et al., 2014).

known or predicted OEPs of Arabidopsis (Inoue, 2015) have not been fully elucidated. However, multiple OEP targeting pathways have been characterized based on the location of the membrane-spanning domains in the amino acid sequence that, in many cases, constitute the protein sorting information (Figure 1.4) (Hofmann and Theg, 2005). Of the few OEPs whose targeting mechanisms have been studied, the majority contain a single α-helix that targets and anchors the protein to the hydrophobic chloroplast outer membrane (Bölter and Soll, 2011). OEP targeting pathways are broadly classified based on whether this single α-helical transmembrane domain (TMD) is located at the N- or C- terminus of the OEP, termed "signal-anchored" and "tail-anchored," respectively (Bölter and Soll, 2011). For example, Toc34 is a tail-anchored protein since it anchors to the chloroplast outer membrane using a short TMD located near its C-terminus whereas OEP7 is a signal-anchored protein which utilizes the N-terminal TMD to targets itself (Oreb, 2008). Toc75 is the only known OEP that uses a canonical N-terminal cleavable transit peptide for targeting (Tranel and Keegstra, 1996). A few integral β-barrel proteins, such as OEP21, appear to self-insert into the chloroplast outer membrane (Pohlmeyer et al., 1998). The cytosolic receptor AKR2A has been demonstrated to mediate targeting of select OEPs to the chloroplast outer membrane (Bae et al., 2008). AKR2A distinguishes OEPs from other substrates by the positively charged region of their transmembrane domains (Lee et al., 2011). The C-terminal domain of AKR2A mediates targeting to the chloroplast outer membrane (Bae et al., 2008) and the nature of the membrane association was recently shown to be between

AKR2A and the chloroplast outer membrane lipids MGDG and phosphatidylglycerol (PG) (Kim et al., 2014).

1.7.1 Toc159 targeting and anchoring to the chloroplast outer membrane

Toc159 contains a C-terminal membrane (M-) domain that anchors the receptor to the outer membrane of chloroplasts. Toc159 also depends on its G-domain to regulate targeting of Toc159 from the cytosol to the chloroplast outer membrane (Bauer et al., 2002). Specifically, targeting of Toc159 to chloroplasts is mediated by interaction between the homologous G-domains of Toc159 and Toc33 (Smith et al., 2002). The specific mechanism of membrane association/anchoring remains unknown due to lack of known homologues and the lack of a predicted transmembrane domain (Lung and Chuong, 2012). The 52-kDa C-terminal domain of Toc159 was given the name "membrane domain" due to its resistance to proteolysis in intact chloroplasts, which indicates it is protected from proteolysis by the membrane (Chen et al., 2000). However, it remains possible that the majority of the protein resides in the hydrophilic intermembrane space, or is resistant to proteolysis due to a poorly understood mechanism of association with the membrane. Bioinformatics tools have predicted the C-terminus of *Bienertia sinuspersici* Toc159 (BsToc159) to contain a cTP-like targeting signal, which also possesses similar physicochemical and structural properties to those of cTPs (Lung and Chuong, 2012). Structural prediction data also suggests that the C-terminus of BsToc159 forms amphipathic structures analogous to those of transit peptides (Lung and Chuong, 2012). Recently, a membrane association region was identified

within the 60-100 amino acid residues upstream from the end of the C-terminus of BsToc159 (Lung et al., 2014). The chloroplast outer membrane targeting signal for BsToc159 was shown to be within the final 56 residues of the C-terminus. Therefore, unlike other OEPs, the chloroplast outer membrane targeting sequence for BsToc159 did not constitute the membrane association region. Furthermore, the membrane association region was demonstrated to not contain the canonical transmembrane domain used by other OEPs. The membrane association region was predicted through bioinformatic analyses and sequence analyses to anchor via a non α-helical and non β-barrel dependent manner (Lung et al., 2014). It is speculated that the BsToc159 C-terminus might adopt a lipophilic β-helix for associating with the chloroplast outer membrane based on high sequence homology with the lipid-binding domain of LpxD, a left-handed β-helical protein (Lung et al., 2014). Therefore, the C-terminus of BsToc159 is an unconventional membrane anchor and may represent a new class of sorting signals to the chloroplast outer membrane. Due to the unconventional nature of BsToc159 targeting and similarities in characteristics with cTPs, it is possible this unique class of sorting signals is shared with other OEPs.

1.8 Overall Objectives and Hypothesis

The mechanism by which the vast majority of OEPs target to the chloroplast outer membrane is largely unknown. A few broadly classified families of OEP targeting pathways have been identified. However, the OEP Toc159 of *Bienertia sinuspersici* was recently demonstrated to target to the chloroplast outer membrane

using a novel transit peptide-like targeting signal in its C-terminal domain. The overall objective of this current study was to further characterize the novel chloroplast outer membrane targeting signal identified in BsToc159. To accomplish this objective, the goal was to find other OEPs that use this novel targeting mechanism. It was hypothesized that using a similar bioinformatic approach that Lung and Chuong (2012) used on BsToc159 would result in identifying other OEPs that use the unique C-terminal targeting signal. In the current study, all known and predicted OEPs of Arabidopsis chloroplasts were analyzed using a bioinformatic approach to first identify OEPs predicted to contain the putative C-terminal transit peptide-like targeting signal. The cDNA of the OEP with highest probability to contain the novel targeting signal was used to design EGFP fusion constructs. These constructs were expressed in onion epidermal cells and Arabidopsis protoplasts to determine the subcellular location of the protein. Furthermore, truncation mutants were designed by removing segments of the OEP and fusing them to EGFP to determine the sequence.

2. Materials and Methods

2.1 Chemicals and supplies

All chemicals were of analytical grade and purchased from Sigma-Aldrich (Oakville, ON, Canada), BioShop Canada Inc. (Burlington, ON, Canada) or Fisher Scientific (Ottawa, ON, Canada), unless otherwise specified. All equipment and supplies for agarose and polyacrylamide gel electrophoresis and transblotting were purchased from Bio-Rad (Mississauga, ON, Canada). Molecular weight standards for electrophoresis included Quick-Load 100 bp DNA Ladder (New England Biolabs, Pickering, ON, Canada), 1 kb DNA Ladder RTU (GeneDireX, Toronto, ON, Canada), and Precision Plus Protein Standards (Bio-Rad, Mississauga, ON, Canada). All restriction enzymes, T4 DNA ligase and other modifying enzymes were purchased from New England Biolabs (Pickering, ON, Canada). PCR reactions for the production of fluorescent fusion constructs were performed using the Phusion High-Fidelity DNA Polymerase (New England Biolabs, Pickering, ON, Canada; cat. no. F-530S), whereas colony PCR were performed using T*aq* DNA Polymerase (New England Biolabs, Pickering, ON, Canada; cat. no. M0267S). DNA sequencing service was provided by the Sanger Sequencing Facility at The Centre for Applied Genomics (The Hospital for Sick Children, Toronto, ON, Canada). Custom DNA oligonucleotides were synthesized by Eurofins Scientific (Huntsville, AL, USA). Purification of plasmid DNA was performed using the EZ-10 Spin Column Plasmid DNA Minipreps or Maxipreps Kits (Biobasic Inc., Markham, ON, Canada).

2.2 Bioinformatic Analysis

Amino acid sequences of the one hundred and seventeen known and predicted Arabidopsis chloroplast OEPs (Inoue, 2015) were analyzed using the ChloroP prediction program v1.1 (Emanuelsson et al., 1999; http://www.cbs.dtu.dk/services/ChloroP) in the forward and reverse orientation for the purpose of predicting the presence of a chloroplast Transit Peptide (cTP). Any score over 0.5 is considered an indication that the sequence likely contains a canonical N-terminal cTP. The same cut-off was used for the purpose of identifying which OEPs might contain a cTP-like sequence at their C-terminal end.

Further bioinformatic analysis was completed on the proteins predicted by ChloroP to have a putative transit peptide at their C-terminus. This was done to rule out the possibility of these proteins potentially containing a canonical tail-anchor or signal-anchor sequence to mediate targeting to the chloroplast outer membrane. The TMHMM v2.0 prediction program (Krogh et al., 2001; http://www.cbs.dtu.dk/services/TMHMM/) was used to predict the presence of transmembrane helices in the protein sequences predicted to contain C-terminal transit peptides.

The candidate proteins were further ruled out to follow a canonical OEP targeting pathway by cross referencing them to the known and predicted Arabidopsis tail-anchor proteins outlined in a recent publication by Marty et al. (2014).

The secondary structure of the OEP selected as the protein of study was predicted using the PSIPRED protein structure prediction server v3.0 (Jones, 1999; http://bioinf.cs.ucl.ac.uk/psipred/).

2.3 Construction of fluorescent protein fusion constructs

In the ChloroP analysis, OEP18 received the highest score for potentially containing a C-terminal transit peptide, and was chosen as the protein for further experiments. The full-length OEP18 cDNA was ordered from the Arabidopsis Biological Resource Centre (ABRC, OSU, Columbus, OH, USA). The construction of the EGFP fusion constructs for transient expression studies were produced by subcloning specific DNA fragments of interest at the 5'-end of EGFP sequence using pSAT6-35S:EGFP-N1 vector or the 3'-end of EGFP sequence using pSAT6-35S:EGFP-C1 vector (Appendix I), as described previously by Chung et al. (2005). The details of primers for the generation of each EGFP fusion construct can be found in Table 2.1. To generate the OEP-EGFP fusion construct, the entire OEP18-encoding sequence was PCR-amplified using a primer set with restriction enzyme site introduced in the forward primer in frame with the N-terminus of EGFP. Similarly, the EGFP-OEP18 fusion construct was generated by PCR-amplifying the entire OEP18-encoding sequence using a primer set with the restriction enzyme site incorporated in the reverse primer in frame with the C-terminus of EGFP. For the generation of the deletion OEP-EGFP fusion constructs, the selected region of OEP18 coding sequence was PCR-amplified using a primer set with restriction enzyme site incorporated in the reverse primer in frame with the N-terminus of EGFP. Primer recognition sites

such as 5 '-CTCGAG-3' inserted in the forward primer and 5 '-GGATCC-3' in the reverse primer sequences, were recognized by restriction endonucleases XhoI and BamHI, respectively.

Fifty µL restriction digest reactions for purified OEP18 PCR products and pSAT6-N1 and pSAT6-C1 vectors were prepared by adding NEBuffer 3 [1 x] (New England BioLabs category # B70003S), BSA [0.1 mg/mL], purified plasmid DNA [2 µg], BamHI [25 U], and ddH_2O. Initial digests were carried out in a water bath overnight at 37°C. Reaction tubes containing the initial digestion were then incubated at 65°C for 20 min to heat-inactivate BamHI. XhoI [25 U] was then added to the reaction tubes and placed in a water bath for 3- to 4-h at 37°C. The tubes were then incubated at 65°C for 20 min to heat-inactivate XhoI. The OEP18 inserts were ligated with the respected pSAT6 vectors with T4 DNA Ligase (New England BioLabs category # M0202S) using a 4:1 insert to vector ratio at 4°C, incubated overnight. Ligated plasmids were mixed with chemically-competent *E. coli* (XL-10 Gold) and incubated on ice for 20 min. Transformation of *E. coli* was achieved using the heat-shock method. Transformed cells were selected by growth on LB agar plates containing ampicillin (50 µg/mL).

Table 2.1 List of oligonucleotides used for the construction of the EGFP fusion constructs

Fusion protein	Vector	Oligonucleotide		
		Name	Sequence (5' to 3')	Orientation
EGFP-OEP18FL	C1	OEP18F1	CGCctcgagCT***ATG***GCGAATTCCATTTCATCA	Sense
		OEP18R1	CGCggatcc***TCA***CTTGTTTGAACTTTTGCT	Anti-sense
OEP18FL-EGFP	N1	OEP18F2	CGCctcgagC***ATG***GCGAATTCCATTTCATCA	Sense
		OEP18R2	CGCggatccCCTTGTTTGAACTTTTGCTAGA	Anti-sense
OEP18ΔC□-EGFP	N1	OEP18F2	CGCctcgagC***ATG***GCGAATTCCATTTCATCA	Sense
		OEP18R3	CGCggatccTCAAGTCACCACGACCAAATGCAA	Anti-sense
OEP18CT-EGFP	N1	OEP18F4	CGCctcgagCTCTAAATCCTCCACTTCTGTA	Sense
		OEP18R2	CGCggatccCCTTGTTTGAACTTTTGCTAGA	Anti-sense
OEP18ΔNT-EGFP	N1	OEP18F3	CGCctcgagCTGCGTGTGGGAAAGAAGAGAAAGA	Sense
		OEP18R2	CGCggatccCCTTGTTTGAACTTTTGCTAGA	Anti-sense
OEP18NT-EGFP	N1	OEP18F1	CGCctcgagCT***ATG***GCGAATTCCATTTCATCA	Sense
		OEP18R4	CGCggatccCTCTGCAACCACTGAAAGT	Anti-sense

2.4 Plant Propagation and Growth Conditions

Wild-type *Arabidopsis thaliana* (ecotype Columbia) seeds were incubated and stratified for at least 24 h in the dark at 4°C in 0.5% (w/v) agar solution. The cold-stratified seeds were then sown in 18x13x6 cm cell packs containing a 1:1 soil mixture of Sunshine LC1 mix and Sunshine LG3 germination mix (SunGro Horticulture Inc., Bellevue, WA, USA). Plants were grown at 22°C under a 16 h:8 h, light : dark cycle in an environment-controlled growth chamber (Conviron Ltd., Winnipeg, MB, Canada) with a light intensity of approximately 150 μmol m^{-2} s^{-1}. Seeds were covered with a plastic dome during germination for the first week in the growth chamber. The seedlings were watered and fertilized regularly with 20:20:20 (N:P:K) fertilizer (Plant Products Co. Ltd., Brampton, ON, Canada). Leaves from 3- to 4-week-old plants were used for protoplast preparation.

2.5 Isolation of mesophyll protoplasts from *A. thaliana*

The procedures for isolation and transfection of mesophyll protoplasts from Arabidopsis were modified from Yoo et al. (2007), and Lung et al. (2014). Briefly, 50 healthy leaves from 3- to 4- week old plants were harvested and cut into 0.5- to 1-mm strips using a sharp double-edge stainless steel razor blade (Electron Microscopy Sciences, Hatfield, PA, USA; cat. no. 72000) on a glass plate. Each razor blade was replaced after cutting approximately 10 leaves to ensure leaf cuttings were made as clean as possible, without tissue tearing or crushing. Leaf strips were immediately transferred using flat-tip forceps into a Petri plate containing 10 mL of enzyme solution (Figure 2.2). Enzyme solution was freshly prepared by heating: CS-

mannitol buffer [0.4 M mannitol, 20 mM MES-KOH (pH 5.7), 20 mM KCl] to 70°C for 10 min and then cooling it to 55°C, and adding cellulase R-10 and macerozyme R-10 (Yakult Pharmaceutical, Tokyo, Japan) to final concentrations of 1.5% (w/v) and 0.4% (w/v), respectively. The enzyme solution was then cooled to room temperature, followed by the addition of BSA (Sigma-Aldrich, Oakville, ON, Canada) and $CaCl_2$ to final concentrations of 0.1% (w/v) and 10 mM, respectively. Leaf strips in the enzyme solution (Figure 2.1A) were vacuum infiltrated for 15 min in the dark using a desiccator. Leaf strips were then incubated in the enzyme solution for 3.5 h in the dark at room temperature without shaking until the cell wall digestion was completed as indicated by the green color of the solution (Figure 2.1B) and the observation of round-shaped protoplasts under light microscopy. Cell wall digestion and protoplasts release was monitored by visualizing under a light microscope. Release of healthy protoplasts is indicated by the presence of spherical cells that are not clumped together.

The solution containing the released protoplasts was transferred using a Pasteur pipette from the Petri dish onto a piece of 75 μm nylon mesh (Sefar America Inc., Kansas City, MO, USA) to filter the protoplasts into a 15 mL falcon tube. The remaining digested leaf strips in the Petri dish were rinsed with 2 mL of W5 solution [2 mM MES (pH 5.7), 154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl]. The 2 mL W5 solution rinse containing any remaining released protoplasts was transferred using a Pasteur pipette from the Petri dish onto the same piece of 75 μm nylon mesh to filter the released protoplasts into the same 15 mL falcon tube. The 15 mL falcon

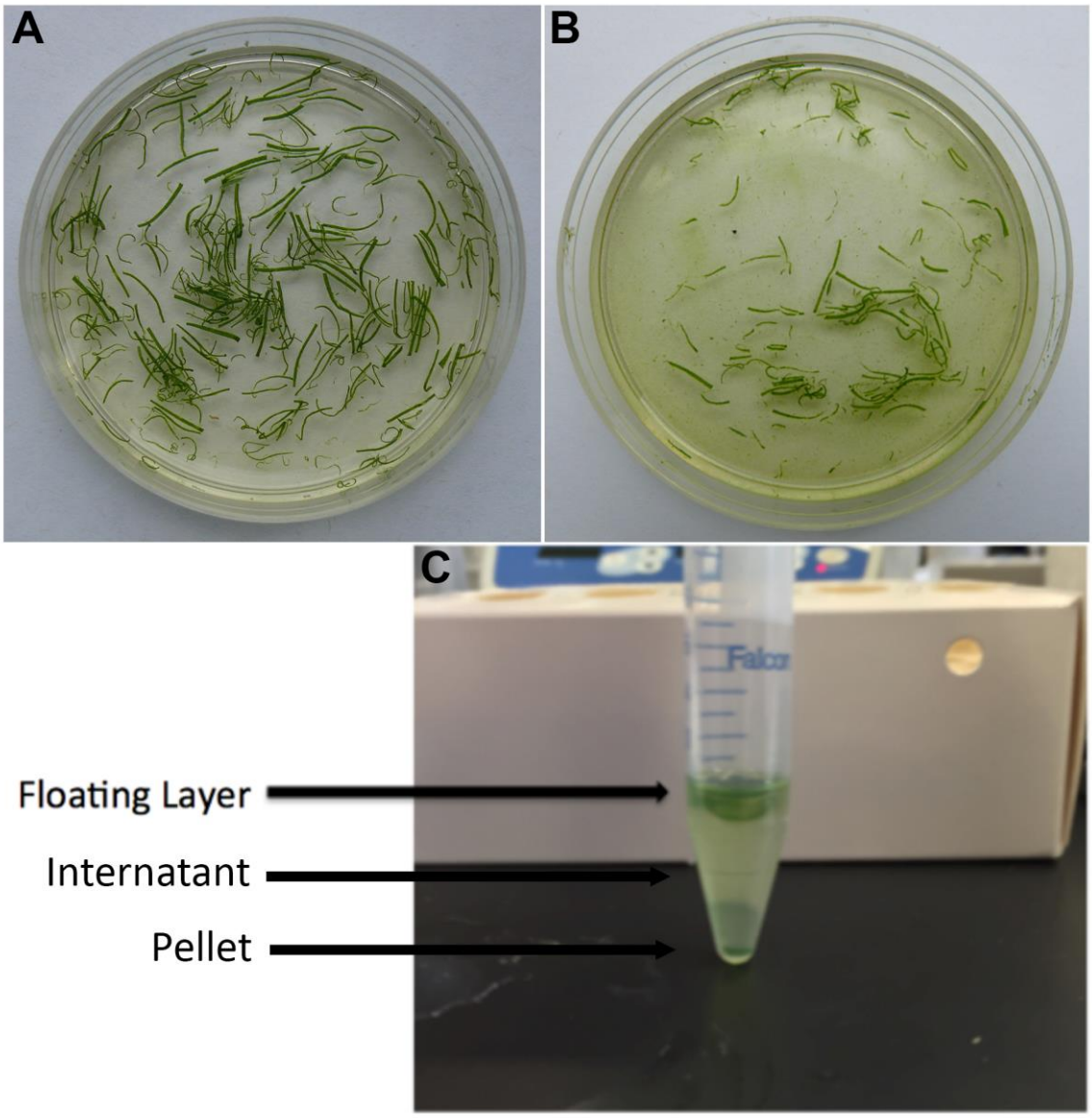


Figure 2.1. Isolation of mesophyll protoplasts from Arabidopsis.

(A) Leaf strips from 3-week old Arabidopsis plants in enzyme solution at the beginning of digestion. (B) Leaf sections after 3.5 h of digestion showed protoplasts released into the solution. (C) Healthy protoplasts floated to the top in the CS-Sucrose buffer after a 2.5 min centrifugation in the Silencer H-20 swinging bucket rotor. The pellet and internatant comprised unhealthy protoplasts and cell debris that were carefully removed using a Pasteur pipette.

tube was centrifuged at 100 g for 2 min in a swinging-bucket rotor (Silencer H-20) to pellet the protoplasts. The supernatant was carefully removed and discarded using a Pasteur pipette. The protoplast pellet containing healthy, unhealthy, and broken protoplasts was resuspended in 2 mL of CS-sucrose buffer [0.4 M sucrose, 20 mM MES-KOH (pH 5.7), 20 mM KCl] and centrifuged at 100 *g* for 2.5 min in the swinging-bucket rotor (Silencer H-20) to float the healthy protoplasts (Figure 2.1C). The internatant and pellet containing unhealthy and broken protoplasts were carefully removed without disturbing the floating layer using a Pasteur pipette. The green floating layer of healthy protoplasts was then diluted in 1 mL of W5 solution using a gentle swirling motion. Ten µL of solution was placed on a haemocytometer to estimate the number of isolated healthy protoplasts. The resuspended healthy protoplasts were incubated on ice for at least 30 min, during which the protoplasts settled to the bottom of the 15 mL tube.

The supernatant was removed, and settled protoplasts were resuspended in a volume of Mg-Man buffer [0.4 M mannitol, 4 mM MES (pH 5.7), 15 mM $MgCl_2$] to a final concentration of 200,000 protoplasts per mL. The protoplast viability was performed by incubating 100 µL of isolated protoplasts in CS-sucrose buffer with 4 µL of 0.2% (w/v) fluorescein diacetate (Sigma-Aldrich, Oakville, ON, Canada; cat. no. F5502) in acetone for 15 min at room temperature, washed twice by centrifugation at 100 *g* for 2 min and resuspension in 100 µL of CS-sucrose buffer. The stained protoplasts were examined under a Zeiss AxioImager D1 epifluorescence microscope (Carl Zeiss Canada Inc., Toronto, ON, Canada).

2.6 Transfection of mesophyll protoplasts from *A. thaliana*

In each standard reaction of PEG-mediated transfection, approximately 20,000 protoplasts were mixed with 5-10 μg of plasmid DNA and 110 μL of PEG solution containing 40% (w/v) PEG4000 (Sigma-Aldrich, Oakville, ON, Canada; cat. no. 81240), 0.4 M sucrose and 100 mM $CaCl_2$. The tube was mixed gently by inverting it 4-6 times and then was incubated in the dark at room temperature for 15 min. The transfected protoplasts were mixed with 440 μL of W5 solution to stop the reaction and centrifuged at 100 g for 2 min. The protoplast pellet was resuspended in 1 mL of WI solution containing 0.5 M mannitol, 4 mM MES-KOH (pH 6.5) and 20 mM KCl and transferred to a 35 mm petri plate and incubated in a growth chamber (Environmental Growth Chambers, OH, USA) at 23 ^{o}C with a photon flux density of approximately 30 μmol m^{-2} s^{-1} overnight. To allow sufficient protoplasts for subsequent Western blot analysis, the standard procedures were scaled up by transfecting 80,000-100,000 protoplasts with 40-50 μg of plasmid DNA, and the transfected protoplasts were cultured overnight in a 50 mm petri plate with 2 mL of WI solution.

2.7 Biolistic Bombardment of Onion Epidermal Cells

Tungsten microcarriers were coated with plasmid DNA of the various fusion constructs essentially as previously described (Sanford et al., 1993). Briefly, 30 mg of tungsten M-17 particles (~1.1 μm in diameter; Bio-Rad, Mississauga, ON, Canada) were washed in 70% (v/v) ethanol by vortexing vigorously for 3 min, and then soaked for 15 min. After a brief centrifugation, the particles were rinsed three times

by vortexing in sterile distilled water. The washed tungsten particles were pelleted by centrifugation, resuspended in 500 μL of 50% (v/v) glycerol and stored at -20 ^{o}C. One milligram of tungsten particles was coated with 5 μg of plasmid DNA in a suspension containing 16 mM spermidine and 0.1 M $CaCl_2$ by vortexing vigorously for 2 min followed by a 5-min incubation step. The DNA-coated tungsten microcarriers were collected by a brief centrifugation step, washed in 70% (v/v) and 100% (v/v) ethanol, and loaded onto the macrocarrier discs (Bio-Rad, Mississauga, ON, Canada). The dried DNA-coated tungsten particles were bombarded into the adaxial epidermis of three 1.5 x 2.5 cm sections of onion bulb from a distance of 10 cm at a helium pressure of 1,350 p.s.i using the Biolistic PDS-1000/He particle-delivery system (Bio-Rad, Mississauga, ON, Canada) according to the manufacturer's instructions. The bombarded samples were incubated on moist filter paper in petri plates at room temperature in the dark for 12-16 h and observed under a Zeiss AxioImager D1 epifluorescence microscope (Carl Zeiss Canada Inc., Toronto, ON, Canada).

2.8 Total Protein Collected from Transfected Protoplasts

After an overnight incubation, transfected protoplasts were visualized using an epifluorescence microscope to determine the rate of transfection. Total protein was extracted if the estimated transfection rate was over 80%. Total proteins from transfected protoplasts were extracted as follows: transfected protoplasts were pelleted at 200 g for 2 min and the supernatant was discarded. The protoplast pellet was vortexed vigorously in 100 μL of lysis buffer [25 mM Tris-HCl (pH 6.8), 1%

(v/v) Triton X-100, 1 mM DTT) for 3 min and centrifuged at 14,000 rpm for 5 min at 4 ^{o}C. The protein extracts from total protoplasts were quantified by the Bradford assays (Bradford, 1976) using the Protein Assay Dye Reagent Concentrate (Bio-Rad, Mississauga, ON, Canada), according to the manufacturer's instructions. The protein concentrations were estimated against standard solutions of BSA from 0.5 to 10 mg mL^{-1}. Protein samples were then concentrated by precipitating the extract in five volumes of acetone at -20 ^{o}C for 1 h. Proteins were collected by centrifugation at 14,000 rpm at 4 ^{o}C for 20 min. The supernatant was discarded and the pellet was air dried for 15 min.

2.9 SDS-PAGE

The total protein fractions isolated from transfected protoplast were resolved by SDS-polyacrylamide gel electrophoresis (PAGE) using a Mini-Protean Electrophoresis Cell (Bio-Rad, Mississauga, ON, Canada). Briefly, the acetone-precipitated protein pellets were resuspended in 20 μL of 6x SDS-PAGE sample buffer [72 mM Tris-HCl (pH 6.8), 30% (v/v) glycerol, 2% (w/v) SDS, 0.12% (w/v) bromophenol blue, and 6% (v/v) β-mercaptoethanol], boiled at 95 ^{o}C for 5 min and resolved by 12% polyacrylamide gel electrophoresis (PAGE) using the Mini-PROTEAN® III Electrophoresis Cell (Bio-Rad, Mississauga, ON, Canada). A 12% (w/v) resolving gel was prepared comprised of the following: 2.4 mL of 30% (w/v) acrylamide/bisacrylamide, 2.25 mL of 1 M Tris-HCl (pH 8.8), 1.26 mL of distilled water, 60 μL of 10% (w/v) SDS, 30 μL of 10% (w/v) APS, and 3 μL of TEMED. A 4.8% (w/v) stacking gel was prepared comprised of the following: 400 μL of 30%

(w/v) acrylamide/bisacrylamide (37.5:1), 312 μL of 1 M Tris-HCl (pH 6.8), 1.75 mL of distilled water, 25 μL of 10% (w/v) SDS, 12.5 μL of 10% (w/v) APS, and 2.5 μL of TEMED. The commercial molecular weight ladder Precision Plus Protein Standards All Blue (Bio-Rad, Mississauga, ON, Canada) was used. Gel electrophoresis was run at 75 V until samples entered the stacking gel in running buffer [24.8 mM Tris, 0.192 M glycine, and 0.1% (w/v) SDS] and then at 120 V until the dye ran off the gel.

2.10 Western Blot

Resolved proteins from SDS-PAGE were transferred onto a polyvinylidene difluoride membrane (PVDF) (Bio-Rad, Mississauga, ON, Canada) at 15 V for 45 min in transfer buffer [48 mM Tris, 39 mM glycine, 20% (v/v) methanol, and 0.0375% (w/v) SDS] at room temperature using the Trans-Blot SD Semi-Dry Electrophoretic Transfer Cell (Bio-Rad, Mississauga, ON, Canada). Briefly, SDS PAGE gel and the methanol-prewetted PVDF membrane were washed in transfer buffer for 15 min before placing in the Trans-Blot SD Semi-Dry Electrophoretic Transfer Cell. To visualize the transferred proteins, the PVDF membrane was placed in 0.1% (w/v) Ponceau stain in 5% (v/v) acetic acid for 15 min. The membrane was then rinsed in distilled water until protein bands became visible. The membrane was incubated in blocking solution containing 5% (w/v) skim milk powder in TBS-T buffer [25 mM Tris-HCl (pH 7.4), 137 mM NaCl, 2.7 mM KCl, and 0.05% (v/v) Tween-20] for 1 h at room temperature, with gentle shaking. The membrane was then incubated again in blocking solution with a primary polyclonal antibody raised in rabbit against enhanced green fluorescent protein (EGFP; 1:4,000) overnight at 4 °C, with shaking.

The membrane was then washed 3 times for 10 min each in TBS-T containing 2% (w/v) skim milk and then incubated in blocking solution with an anti-rabbit secondary antibody conjugated to horseradish peroxidase (1:100,000) (Sigma-Aldrich, Oakville, ON, Canada) at room temperature for 2 h with shaking. The same 3 x 10 min wash was repeated on the membrane. The membrane was then incubated in a 1:1 mix of solution A and solution B from Amersham ECL-Advance Solution (GE Healthcare, Baie d'Urfe, QC, Canada) in the dark for 5 min to enable detection of chemiluminescent signals. Excess ECL mix was removed from the membrane by tilting. Imaging was achieved using a BioRad ChemiDoc MP Imaging System. The signal accumulation mode was set to capture 15 images between 5- and 155- seconds of exposure. The best image out of the multiple exposure times was selected and saved on the computer. The membrane was imaged for a second time under the setting "colourmetric" with an exposure time set to 0.5 seconds to visualize the Precision Plus All Blue ladder. The colourmetric image representing the ladder was merged with the chemi blot displaying the protein bands. Captured images were processed using Adobe Photoshop CS (Adobe System Inc., Seatle, WA, USA).

2.11 Epifluorescence Microscopy

The bombarded adaxial epidermis of the onion scales were peeled and mounted in water onto a glass slide prior to microscopic observation. Stained or transfected protoplasts were examined in flat-bottom chamber slides made of nail polish-premounted coverslips. Epifluorescence and bright field micrographs were

acquired using a Zeiss AxioImager epifluorescence microscope equipped with the AxioVision Imaging software (Carl Zeiss Canada Inc., Toronto, ON, Canada). Fluorescein diacetate and EGFP signals were detected using the EGFP/FITC filter set (turret #2) under UV illumination whereas the DsRed signal was detected using the DsRed/rhodamine filter set (turret #5). All images were processed using Adobe Photoshop CS (Adobe Systems Inc., Seattle, WA, USA). Representative images were presented after similar results were obtained from at least 3 independent experiments. Colocalization of the EGFP and DsRed signals were obtained from scatterplots and Pearson's coefficients generated from the Fiji colocalization threshold plug-in of ImageJ (National Institutes of Health, USA).

2.12 Confocal Microscopy

The subcellular localization of OEP18 fusion constructs in Arabidopsis protoplasts were also visualized using an Olympus FV1000 confocal laser-scanning microscope. Transfected protoplasts were transferred to 8-chamber Lab-Tek II (Nalgene Nunc, Denmark) cover glass slides. Samples were first scanned under the bright field setting. Two excitation wavelengths, 488 nm and 594 nm, were used to detect EGFP and chlorophyll autofluorescence, respectively. Serial Z-stack images were taken at 1 μm intervals using a 40x objective at 1024x1024 pixel resolution. All images were further processed using Adobe Photoshop CS (Adobe System Inc., Seattle, WA, USA). Multiple independent experiments were performed for each construct with similar results.

3. Results – Bioinformatic Analyses

3.1 ChloroP Analysis of the One-Hundred and Seventeen Chloroplast Outer Membrane Proteins of Arabidopsis

The recently published paper by Inoue (2015) outlines the one hundred and seventeen proteins identified or predicted to be in the chloroplast outer membrane (COM) in Arabidopsis. The functions of these proteins involve solute and ion transport, preprotein import, protein turnover, lipid and carbohydrate metabolism, and intracellular communication (Inoue, 2015). Each of the proteins was analyzed using ChloroP to identify potential TP-like targeting information in their sequences (Table 3.1). Specifically, the amino acid sequences of each protein were input in the forward, as well as the reverse orientation, as an alternative approach to potentially identify additional OEPs that might contain similar targeting information as that present at the C-terminus of BsToc159 (Table 3.1). The idea of running sequences through in the reverse orientation was inspired by the research performed by Lung and Chuong (2012) on the C-terminus (CT) of Toc159 in *Bienertia sinuspersici*, which demonstrated how the CT targets and anchors the receptor to the chloroplast outer membrane through an unconventional targeting sequence that resembles transit peptides.

The ChloroP analysis identified 8 proteins with scores that suggested they could potentially contain transit peptides in the reverse orientation at their C-terminus: OEP16-2, Toc132, E-Tu, pBRP, MIRO2, DUF869, putative GTPase, and At5g42070, with predicted transit peptide lengths of 33, 34, 25, 38, 59, 82, 44, and 23 amino acids, respectively. Out of the 8 candidate proteins, ChloroP predicted

Table 3.1. The amino acid sequences of one-hundred and seventeen known or predicted chloroplast outer membrane proteins of Arabidopsis were analyzed using the bioinformatics tool ChloroP in the forward and reverse orientation to identify putative transit peptides and cleavage sites. Proteins with predicted transit peptides at the C-terminus in the reverse orientation are highlighted in yellow.

AGI #	Name	A.A Length	Function	Envelope	ChloroP Score (Forward)	cTP Length (Forward)	ChloroP Score (Reverse)	cTP Length (Reverse)
At1g20816	OEP21-1	167	Solute/Ion Transport	Yes	0.446		0.432	
At1g45170	OEP24-1	213	Solute/Ion Transport		0.538	31	0.44	
At1g76405	OEP21-2	167	Solute/Ion Transport	Yes	0.448		0.431	
At2g01320	WBC7	728	Solute/Ion Transport	Yes	0.451		0.426	
At2g17695	OEP23/DUF1990	205	Solute/Ion Transport	Yes	0.449		0.449	
At2g28900	OEP16-1	148	Solute/Ion Transport	Yes	0.491		0.45	
At2g43950	OEP37	343	Solute/Ion Transport	Yes	0.576	73	0.438	
At3g51870	PAPST1 homolog	381	Solute/Ion Transport	Yes	0.51	26	0.447	
At3g62880	OEP16-4	136	Solute/Ion Transport		0.456		0.489	
At4g16160	OEP16-2	178	Solute/Ion Transport		0.433		0.516	33
At5g42960	OEP24-2	213	Solute/Ion Transport	Yes	0.477		0.463	
At1g02280	Toc33	297	Protein Import Components and their homologs	Yes	0.477		0.431	
At2g16640	Toc132	1206	Protein Import Components and their homologs	Yes	0.428		0.514	34
At2g17390	AKR2B	344	Protein Import Components and their homologs		0.446		0.433	
At3g16620	Toc120	1089	Protein Import Components and their homologs		0.427		0.491	
At3g17970	Toc64-III	589	Protein Import Components and their homologs	Yes	0.443		0.435	
At3g44160	p39/OEP80tr1	362	Protein Import Components and their homologs		0.478		0.492	
At3g46740	Toc75-III	818	Protein Import Components and their homologs	Yes	0.585	79	0.447	
At3g48620	p36/OEP80tr2	321	Protein Import Components and their homologs		0.481		0.479	

AGI #	Name	A.A Length	Function	Envelope	ChloroP Score (Forward)	cTP Length (Forward)	ChloroP Score (Reverse)	cTP Length (Reverse)
At4g02510	Toc159	1503	Protein Import Components and their homologs	Yes	0.488		0.478	
At4g09080	Toc75-IV	396	Protein Import Components and their homologs		0.423		0.44	
At5g05000	Toc34	313	Protein Import Components and their homologs	Yes	0.439		0.443	
At5g19620	OEP80/Toc75-V	732	Protein Import Components and their homologs	Yes	0.535	93	0.442	
At5g20300	Toc90	793	Protein Import Components and their homologs		0.471		0.434	
At1g02560	ClpP5 (proteolysis)	298	Protein Turnover	Yes	0.568	62	0.427	
At1g07930	E-Tu (protein synthesis)	449	Protein Turnover		0.432		0.516	25
At1g09340	HIP1.3/RAP38/CSP41B (protein synthesis)	378	Protein Turnover	Yes	0.495		0.432	
At1g63900	SP1 (proteolysis)	347	Protein Turnover		0.451		0.442	
At1g67690	M3 protease	710	Protein Turnover		0.458		0.451	
At3g46780	pTAC16 (transcription)	510	Protein Turnover	Yes	0.512	19	0.469	
At4g05050	UBQ11 (proteolysis)	229	Protein Turnover		0.428		0.431	
At4g32250	Tyrosine Kinase	611	Protein Turnover	Yes	0.427		0.453	
At4g36650	pBRP (transcription)	503	Protein Turnover		0.465		0.515	38
At5g16870	PTH2 family (protein synthesis)	169	Protein Turnover		0.444		0.433	
At5g35210	PTM (transcription)	1706	Protein Turnover	Yes	0.442		0.465	
At5g56730	Peptidase M16 family	956	Protein Turnover	Yes	0.439		0.466	
At1g77590	LACS9	691	Lipid Metabolism	Yes	0.45		0.432	
At2g11810	MGD3	465	Lipid Metabolism		0.445		0.432	
At2g27490	ATCOAE	232	Lipid Metabolism	Yes	0.475		0.445	

AGI #	Name	A.A Length	Function	Envelope	ChloroP Score (Forward)	cTP Length (Forward)	ChloroP Score (Reverse)	cTP Length (Reverse)
At2g38670	PECT1	421	Lipid Metabolism		0.439		0.433	
At3g06510	SFR2/GGGT	656	Lipid Metabolism	Yes	0.491		0.44	
At3g06960	TGD4	479	Lipid Metabolism	Yes	0.439		0.435	
At3g11670	DGD1	808	Lipid Metabolism		0.559	58	0.453	
At3g26070	PAP/FBN3a	242	Lipid Metabolism	Yes	0.57	50	0.431	
At3g63170	FAP1	279	Lipid Metabolism	Yes	0.577	73	0.427	
At4g00550	DGD2	473	Lipid Metabolism		0.452		0.447	
At4g15440	HPL homolog	384	Lipid Metabolism	Yes	0.436		0.463	
At5g20410	MGD2	468	Lipid Metabolism		0.434		0.441	
At1g12230	Transaldolase	427	Carbohydrate Metabolism and Regulation	Yes	0.586	47	0.437	
At1g13900	PAP2	656	Carbohydrate Metabolism and Regulation		0.513	19	0.487	
At2g19860	HXK2	502	Carbohydrate Metabolism and Regulation		0.472		0.426	
At4g29130	HXK1	496	Carbohydrate Metabolism and Regulation	Yes	0.495		0.436	
At1g34430	PDC E2	465	Other Metabolism and Regulation	Yes	0.539	48	0.46	
At1g44170	ALDH3H1	484	Other Metabolism and Regulation		0.437		0.461	
At2g34590	PDC E1beta	406	Other Metabolism and Regulation	Yes	0.586		0.434	
At2g47770	TSPO	196	Other Metabolism and Regulation		0.44		0.472	
At3g01500	beta CA1	347	Other Metabolism and Regulation	Yes	0.597	47	0.438	

AGI#	Name	A.A Length	Function	Envelope	ChloroP Score (Forward)	cTP Length (Forward)	ChloroP Score (Reverse)	cTP Length (Reverse)
At3g16950	PDCE3	623	Other Metabolism and Regulation	Yes	0.589	70	0.446	
At3g25860	PDCE2	480	Other Metabolism and Regulation	Yes	0.592	47	0.462	
At3g27820	MDAR4	488	Other Metabolism and Regulation	Yes	0.436		0.463	
At5g17770	CBR	281	Other Metabolism and Regulation		0.455		0.44	
At5g23190	CYP86B1	559	Other Metabolism and Regulation		0.504	17	0.426	
At5g25900	KO1/GA3	509	Other Metabolism and Regulation		0.545	28	0.435	
At2g16070	PDV2 (division)	307	Intracellular Communication	Yes	0.427		0.438	
At2g20890	THF1/PSB29 (plasma membrane)	300	Intracellular Communication	Yes	0.579	67	0.433	
At3g25690	CHUP1 (actin-dependent movement)	1004	Intracellular Communication	Yes	0.466		0.44	
At5g53280	PDV1 (division)	272	Intracellular Communication		0.429		0.431	
At5g58140	PHOT2 (actin-dependent movement)	915	Intracellular Communication	Yes	0.455		0.422	
At1g27390	TOM20-2 (mito)	210	Functions/locations defined in compartments other than the chloroplast OM		0.428		0.453	
At3g01280	VDAC1 (mito)	276	Functions/locations defined in compartments other than the chloroplast OM	Yes	0.467		0.462	

AGI #	Name	A.A Length	Function	Envelope	ChloroP Score (Forward)	cTP Length (Forward)	ChloroP Score (Reverse)	cTP Length (Reverse)
At3g12580	Hsc70-4 (cytosol)	650	Functions/locations defined in compartments other than the chloroplast DM		0.43		0.432	
At3g21865	PEX22 (peroxisome)	283	Functions/locations defined in compartments other than the chloroplast DM		0.462		0.452	
At3g46030	Histone H2B (nucleus)	145	Functions/locations defined in compartments other than the chloroplast DM		0.424		0.457	
At3g63150	MIRO2 (mito)	643	Functions/locations defined in compartments other than the chloroplast DM		0.495		0.549	59
At4g14430	inoyl-CoA isomerase (peroxisome)	240	Functions/locations defined in compartments other than the chloroplast DM		0.436		0.432	
At4g16450	Complex I subunit (mito)	106	Functions/locations defined in compartments other than the chloroplast DM		0.463		0.463	
At4g31780	MGD1 (IEM)	533	Functions/locations defined in compartments other than the chloroplast DM	Yes	0.575	33	0.437	
At4g35000	APX3 (peroxisome)	287	Functions/locations defined in compartments other than the chloroplast DM	Yes	0.431		0.453	
At4g38920	Vacuolar ATPase sub	164	Functions/locations defined in compartments other than the chloroplast DM		0.523	25	0.432	

AGI #	Name	A.A Length	Function	Envelope	ChloroP Score (Forward)	cTP Length (Forward)	ChloroP Score (Reverse)	cTP Length (Reverse)
At5g02500	HSC70-1 (cytosol/nucleus)	651	Functions/locations defined in compartments other than the chloroplast OM	Yes	0.43		0.428	
At5g06290	Prx IIB (stroma)	273	Functions/locations defined in compartments other than the chloroplast OM	Yes	0.598	90	0.431	
At5g15090	VDAC3 (mito)	274	Functions/locations defined in compartments other than the chloroplast OM	Yes	0.485		0.465	
At5g27540	EMB2473/MIRO1 (mito)	648	Functions/locations defined in compartments other than the chloroplast OM		0.477		0.465	
At5g35360	CAC2/BCA (IEM)	555	Functions/locations defined in compartments other than the chloroplast OM	Yes	0.571	70	0.428	
At1g09920		192	Function unknown/unclear		0.426		0.49	
At1g16000	OEP9	86	Function unknown/unclear		0.433		0.443	
At1g27300		200	Function unknown/unclear		0.427		0.462	
At1g64850		162	Function unknown/unclear	Yes	0.438		0.482	
At1g68680		75	Function unknown/unclear	Yes	0.46		0.441	
At1g70480	DUF220	338	Function unknown/unclear		0.486		0.434	
At1g80890	OEP9.2	80	Function unknown/unclear		0.432		0.436	
At2g06010		188	Function unknown/unclear		0.429		0.432	
At2g24440		183	Function unknown/unclear		0.465		0.428	
At2g32240	DUF869	1333	Function unknown/unclear		0.427		0.549	82
At2g32650	PTAC18 like	139	Function unknown/unclear		0.53	32	0.428	

AGI	Name	A.A Length	Function	Envelope	ChloroP Score (Forward)	cTP Length (Forward)	ChloroP Score (Reverse)	cTP Length (Reverse)
At2g44640		451	Function unknown/unclear	Yes	0.487		0.425	
At3g26740	CCL	141	Function unknown/unclear		0.58	41	0.416	
At3g49350		539	Function unknown/unclear		0.596	55	0.43	
At3g52230	OMP24 homolog	145	Function unknown/unclear	Yes	0.473		0.426	
At3g52420	OEP7	64	Function unknown/unclear		0.442		0.445	
At3g53560	TPR protein	340	Function unknown/unclear	Yes	0.575	75	0.429	
At3g63160	OEP6	69	Function unknown/unclear	Yes	0.442		0.493	
At4g02482	Putative GTPase	134	Function unknown/unclear		0.473		0.548	44
At4g15810	NTPase	918	Function unknown/unclear		0.471		0.461	
At4g17170	RAB2	211	Function unknown/unclear	Yes	0.434		0.434	
At4g27680	NTPase	398	Function unknown/unclear		0.433		0.44	
At4g27990	YGGT-B protein	218	Function unknown/unclear	Yes	0.566	83	0.439	
At5g11560		982	Function unknown/unclear		0.442		0.434	
At5g20520	WAV2	308	Function unknown/unclear		0.443		0.434	
At5g21920	YGGT-2	251	Function unknown/unclear		0.559	51	0.496	
At5g21990	OEP61-TPR	554	Function unknown/unclear		0.457		0.439	
At5g27330		628	Function unknown/unclear		0.524	27	0.474	
At5g42070		164	Function unknown/unclear	Yes	0.568	72	0.571	23
At5g43070	WPP1	155	Function unknown/unclear		0.45		0.434	
At5g51020	CRL	269	Function unknown/unclear	Yes	0.497		0.443	
At5g59840	RAB8A-like	216	Function unknown/unclear		0.438		0.449	
At5g64816		130	Function unknown/unclear	Yes	0.429		0.454	

At5g42070 as having the highest probability of containing a transit peptide at its C-terminus in the reverse orientation. At5g42070 was the only protein out of the 8 candidates predicted by ChloroP to contain a transit peptide at its N-terminus as well. For simplicity, "At5g42070" was referred to as "OEP18" here, since the protein is confirmed to be an *o*uter *e*nvelope *p*rotein of Arabidopsis chloroplasts (Inoue, 2015) and its open reading frame encodes a protein product of 17.7 kDa in size.

3.2 Further Bioinformatic Analysis on the 8 Candidate Proteins

Further bioinformatic analysis was performed on the 8 proteins predicted by ChloroP to have a transit peptide at the C-terminus (Table 3.2). This was done to rule out the possibility of these proteins potentially containing a canonical tail-anchored or signal-anchored sequence that could mediate targeting to the chloroplast outer membrane. Since signal-anchored and tail-anchored proteins are anchored to the chloroplast outer membrane by a single transmembrane domain (Inoue, 2015), the 8 candidate amino acid sequences were input into the bioinformatics program TMHMM2.0 which predicts the presence of transmembrane domains. Of the 8 candidate proteins, MIRO2 and DUF869 were predicted by TMHMM2.0 to have a transmembrane domain, thereby making them unlikely to target to the chloroplast outer membrane using the unique, non-canonical transit peptide-like signal.

The 8 candidate proteins were also compared to the known and predicted Arabidopsis tail-anchored proteins published by Marty et al. (2014) to further rule out the proteins that follow the canonical tail-anchored mediated targeting pathway

(Table 3.2). Of the 8 proteins, MIRO2 and DUF869 were known and predicted, respectively, to belong to the subset of tail-anchored proteins. This evidence strengthened the likelihood that MIRO2 and DUF869 do not target to the chloroplast outer membrane using the unique, non-canonical transit peptide-like signal.

3.3. OEP18 Nucleotide Sequence

The nucleotide sequence for the OEP18 cDNA was obtained through the National Centre for Biotechnology Information (NCBI) database. The nucleotide sequence of the full-length At5g42070 cDNA clone (707 nucleotides) revealed 565 nucleotides in the entire mRNA coding region, 76 nucleotides in the 5' untranslated region, 111 nucleotides in the 3' untranslated region and a poly(A) tail of 20 nucleotides. The open reading frame (ORF) encodes a putative polypeptide of 164 amino acids with a calculated molecular weight of 17.7 kDa (Figure 3.1).

3.4. Predicted transit peptide properties of the OEP18 carboxyl terminus (CT)

In an attempt to determine the chloroplast envelope-targeting signal predicted by ChloroP, the amino acid sequences of OEP18 CT and NT were analyzed. It was observed that the CT of OEP18 exhibits characteristics similar to those of cTPs (Table 3.3). For example, the hydroxylated residues (i.e. serine and threonine) are overrepresented in both termini with 27.8% in the NT and 53.8% in the CT and acidic residues (i.e. aspartic acid and glutamic acid) are underrepresented with similar scores: 8.3% in the NT and 7.1% in the CT. However, the calculated isoelectric point of the NT is intermediate whereas that of the CT is basic (pI =10).

These properties are in agreement with the characteristic features of cTPs (von Heijne et al., 1989; Patron and Waller, 2007; Lung, 2012).

3.5. Secondary Structure Prediction for OEP18

The bioinformatic tool PSIPRED was used to predict the secondary structure of OEP18 (Figure 3.2). PSIPRED predicted the presence of a putative β-strand within the N-terminus of OEP18. The protein was also predicted to contain a putative α-helix within the C-terminus of OEP18. Three short β-strands and three α-helices were predicted to be in the middle of the protein.

Table 3.2. Further bioinformatic analysis on the 8 proteins predicted to contain targeting information at the C-terminus in the reverse orientation. Amino acid sequences were analyzed using the online tool TMHMM2.0 to predict the presence and length of transmembrane helices. Known or predicted tail-anchor proteins of Arabidopsis outlined in Marty et al. (2014) was also used to see if any of the 8 candidate proteins follow the canonical tail-anchor mediated pathway to the chloroplast outer membrane. OEP7 was used as a control to test the efficiency of TMHMM2.0, as it is known to insert into the chloroplast outer membrane using a transmembrane helix.

AGI #	Name	Amino Acid Length	Function	Envelope	ChloroP Score (Forward)	cTP Length (Forward)	ChloroP Score (Reverse)	cTP Length (Reverse)	Known or Predicted to possess TA orientation	Number of predicted transmembrane helices	Expected number of amino acids in TMH
At4g16160	OEP16-2	178	Solute/Ion Transport		0.433		0.516	33	No	0	0
At2g16640	Toc132	1206	Protein Import Components and their homologs	Yes	0.428		0.514	34	No	0	0
At1g07930	E-Tu (protein synthesis)	449	Protein Turnover		0.432		0.516	25	No	0	0
At4g36650	pBRP (transcription)	503	Protein Turnover		0.465		0.515	38	No	0	0
At3g63150	MIRO2 (mito)	643	Functions/locations defined in compartments other than the chloroplast OM		0.495		0.549	59	Yes (Known)	1	22
At2g32240	DUF869	1333	Function unknown/unclear		0.427		0.549	82	Yes (Predicted)	1	21
At4g02482	Putative GTPase	134	Function unknown/unclear		0.473		0.548	44	No	0	0
At5g42070	OEP18	164	Function unknown/unclear	Yes	0.568	72	0.571	23	No	0	0
Known transmembrane helices containing protein											
At3g52420	OEP7	64	Function unknown/unclear		0.442		0.445		No	1	21

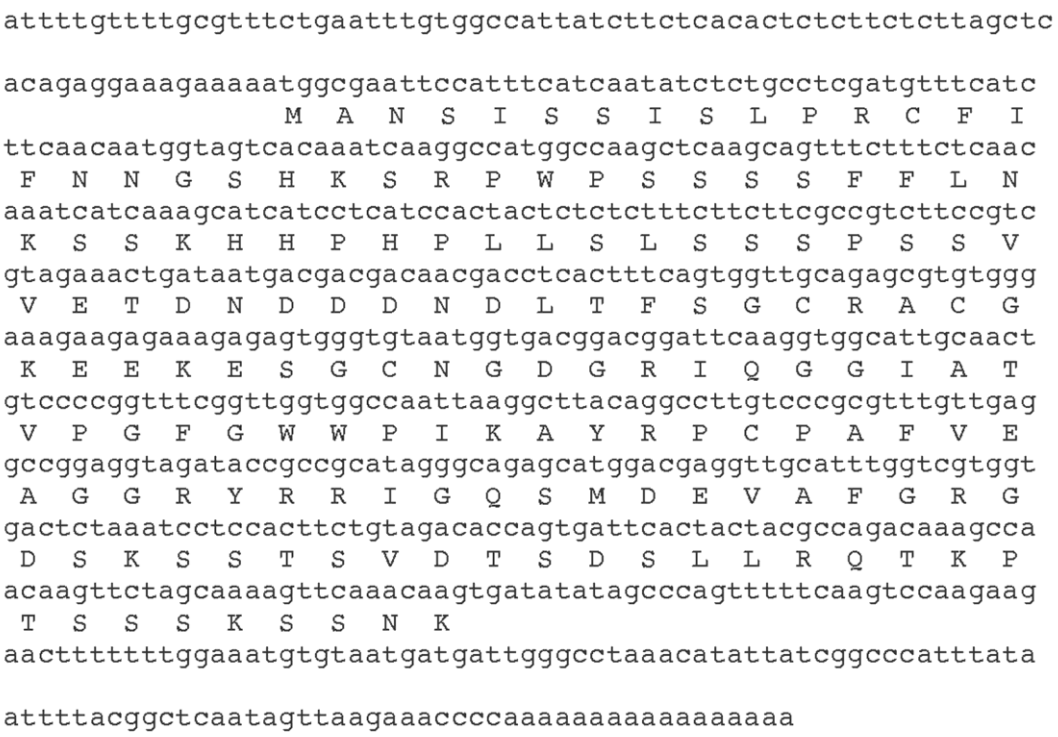


Figure 3.1. The primary sequence of OEP18 mRNA and the deduced amino acids derived from the mRNA

The full-length nucleotide sequence (At5g42070) coding for the translational product of Arabidopsis OEP18.

Table 3.3 Physiochemical properties of OEP18 amino- and carboxyl-termini

Peptide	Terminus	Residues	Occurrence (%) D+E[1]	Occurrence (%) S+T[2]	Calculated pI
OEP18	Amino	72	8.3	27.8	6.89
OEP18	Carboxyl	28	7.1	53.6	10.00
RbcS	Amino	80	0.3	25.0	9.36
Fd	Amino	52	0	28.8	12.60
BsToc159[3]	Carboxyl	69	1.4	20.3	10.21

1 The occurrence of acidic residues including aspartic (D) and glutamic (E) acids.

[2] The occurrence of hydroxylated residues including serine (S) and threonine (T).

[3] The values were previously reported by Lung (2012).

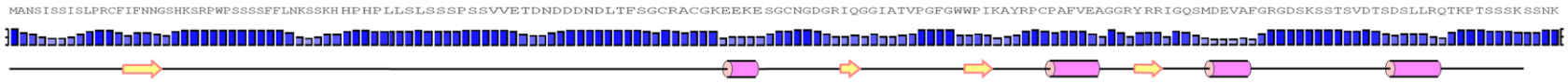


Figure 3.2. Secondary structure prediction of OEP18 using the bioinformatics tool PSIPRED (Jones, 1999).

The height of the blue bars represents the confidence level of the prediction for each residue. The purple cylinders represent putative α-helices. The yellow arrows represent putative β-strands.

4. Results – Construct Design and Cellular Expression

4.1 Construction of the OEP18 Full Length Fusion Constructs

To examine the subcellular localizations of OEP18, two transient expression fusion constructs were made by fusing the entire coding sequence for OEP18 to either the amino or carboxy terminus of EGFP forming OEP18FL-EGFP or EGFP-OEP18FL, respectively (Figure 4.1). The OEP18FL fusion constructs were used to transfect onion epidermal cells.

4.2 Onion Epidermal Cells Bombarded with the Full-Length OEP18 Fusion Constructs

Onion epidermal cells were bombarded with the two full-length OEP18 EGFP-fusion constructs to examine their expression before transfecting them into Arabidopsis protoplasts (Figure 4.2). Information on the subcellular localization of the constructs can be readily acquired from the bombarded onion epidermal cells. The EGFP null construct was used as a control. In the absence of OEP18, the EGFP control construct showed mostly nuclear and cytosolic localization (Figure 4.2A). In cells that were transfected with OEP18 fused to the C-terminus of EGFP (EGFP–OEP18FL), the EGFP signal was mostly detected in the nucleus and cytoplasm (Figure 4.2B). Occasionally, the fusion protein appeared as irregular-shaped punctate structures. However, cells that were transformed with OEP18 fused at the N-terminus of EGFP (OEP18FL–EGFP) showed little cytosolic EGFP signal (Figure 4.2C). Instead, the EGFP signal in most cells appeared as punctate structures with elongated extensions that resembled stromules (Figure 4.2C inset).

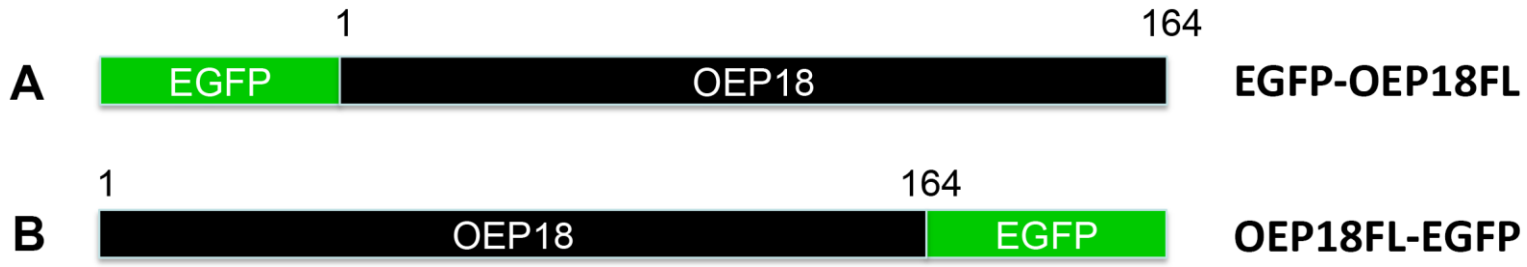


Figure 4.1. Schematic maps of full-length OEP18 fusion constructs.

OEP18 is 164 amino acids in length. **(A)** OEP18 is fused to the C-terminus of EGFP (EGFP–OEP18FL) of the pSAT6-C1 vector. **(B)** OEP18 is fused to the N-terminus of EGFP (OEP18FL–EGFP) of the pSAT6-N1 vector. "FL" denotes it is the "full length" of the protein.

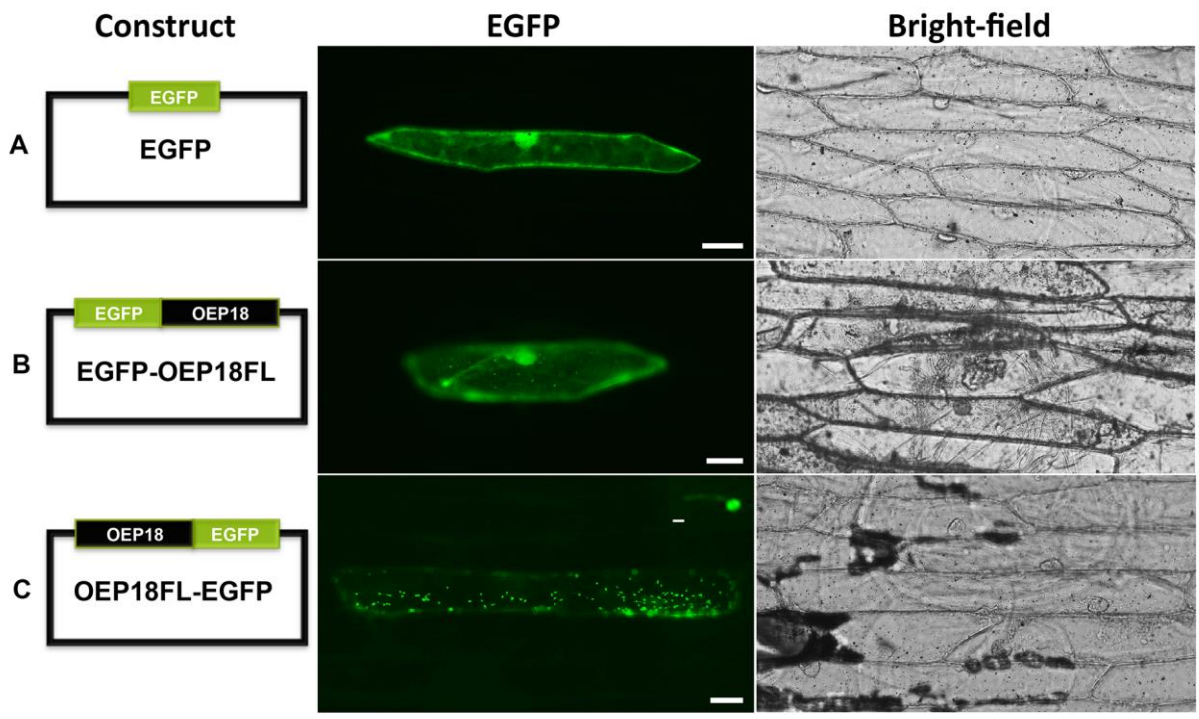


Figure 4.2. Transient expression of EGFP fusion proteins with full-length sequences of OEP18

Onion epidermal cells were bombarded with the two full-length OEP18 fusion constructs to examine their subcellular localizations. (A) A representative epidermal cell expressing the null vector containing only EGFP. (B) A cell expressing the EGFP–OEP18FL construct with OEP18 fused to the C-terminus of EGFP. (C) A cell expressing the OEP18FL–EGFP construct with OEP18 fused to the N-terminus of EGFP. Each is a representative cell from at least three independent experiments. The inset shows a magnified view of a punctate structure with an elongated tail resembling a stromule. Scale bar = 50 μm

4.3 Verifying the Identity of the Punctate Structures through Colocalization Analysis of Full Length Fusion Proteins with DsRed in Onion Epidermal Cells

To further evaluate the efficiency of the full-length OEP18 constructs to target to plastids, onion epidermal cells were co-bombarded with the EGFP fusion constructs and a second construct encoding DsRed fused to the transit peptide of ferredoxin (Figure 4.3). Ferredoxin is a known protein of plastids, and thus would direct DsRed to plastids. Overlapping signals between the EGFP and DsRed verifies that the elongated punctate structures are in fact stromules from etioplasts. Green punctate signals that do not colocalize with DsRed are most likely insoluble aggregates that represent protein misfolding (Lung et al., 2014) or proteins that are targeted to non-plastid organelles. The EGFP signal in the control cells showed no colocalization with DsRed, therefore indicating the protein is not targeted to plastids. Cells expressing the EGFP–OEP18FL fusion protein showed little-to-no overlap with the DsRed-decorated plastids, while the OEP18FL–EGFP signals colocalized with DsRed signals, as shown by the yellow punctate signals from the merge of the two channels. This is also supported by the distribution of the two signals clustering along the diagonal line in the scatter plot (Figure 4.3). Finally, Pearson's correlation coefficients (R_r) were higher for OEP18FL–EGFP/DsRed than for EGFP–OEP18FL/DsRed (0.7338 ± 0.0527 vs 0.3020 ± 0.0681; Table 4.1).

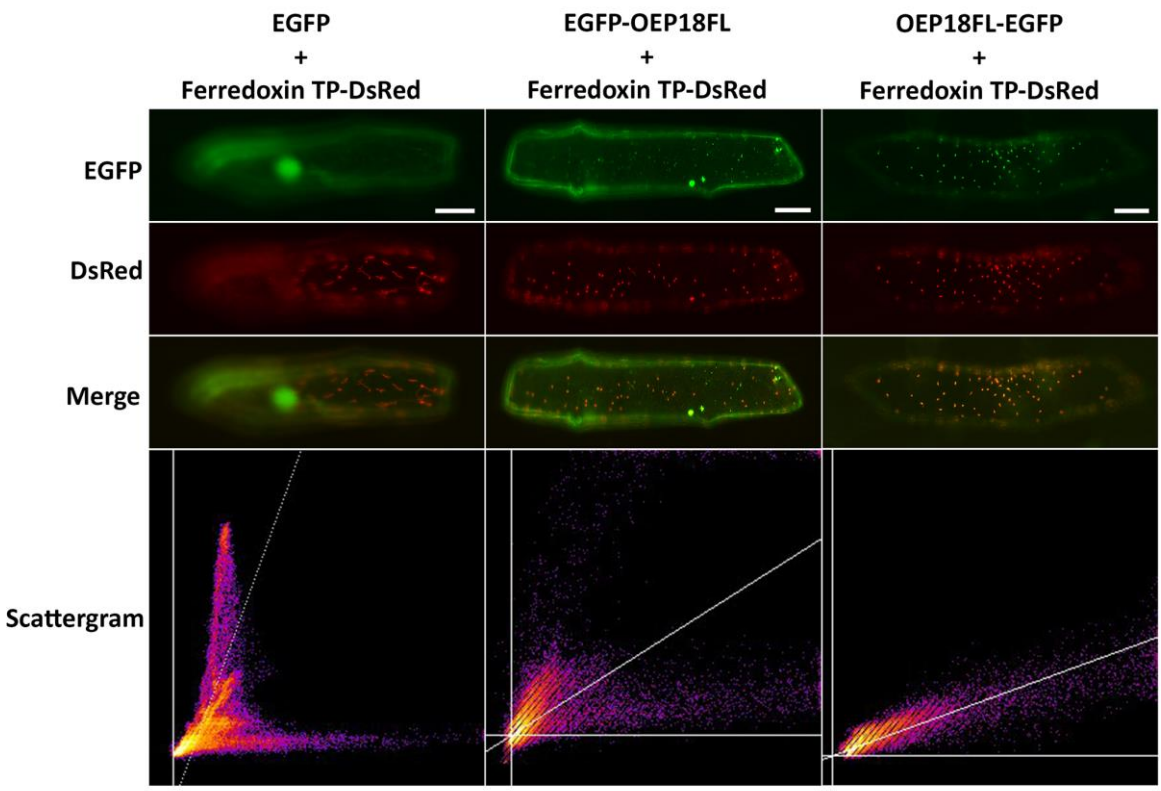


Figure 4.3. Colocalization analysis of EGFP fusion proteins with OEP18FL in onion epidermal cells

Onion epidermal cells were co-bombarded with full-length OEP18 fusion constructs and the ferredoxin transit peptide fused to DsRed. Representative images from multiple independent experiments of EGFP (green signal), DsRed (red signal), and a merge of the two channels are shown for the control and two full-length OEP18 constructs. A scatterplot was generated using the Fiji "Colocalization threshold" plug-in of ImageJ to demonstrate the colocalization between EGFP and DsRed signals for each construct. Scale bars = 50 µm

Table 4.1. Pearson's correlation coefficients (R_r) of the two fluorescent channels for the EGFP, EGFP–OEP18FL, and OEP18FL–EGFP constructs co-bombarded with DsRed tagged to the transit peptide of ferredoxin in onion epidermal cells

Construct	EGFP	EGFP-OEP18FL	OEP18FL-EGFP
R_r	0.1035 ± 0.0682	0.3020 ± 0.0681	0.7338 ± 0.0527

Values represent the mean from four replicates (N=4) of each construct (± SD) from multiple independent experiments was calculated. The maximum theoretical R_r score is 1.

4.4 Epifluorescence Imaging of Full-Length OEP18 Fusion Constructs in Transfected Arabidopsis Protoplasts

Chloroplast targeting efficiency of each full-length construct was evaluated by transiently expressing them in Arabidopsis protoplasts (Figure 4.4). Arabidopsis protoplasts were isolated and transfected with the full-length OEP18 constructs using methods as described previously in Section 2.5, and visualized 12-16 h after transfection using epifluorescence microscopy. In protoplasts that were transformed with the control vector containing EGFP, the EGFP signals were mostly found in the nucleus and cytoplasm (Figure 4.4A). Protoplasts transfected with the EGFP-OEP18FL construct showed some green fluorescent signals forming ring-like structures, suggesting some chloroplast targeting (Figure 4.4B). In comparison, protoplasts expressing the OEP18FL–EGFP construct displayed some nuclear expression, but also thinner green fluorescent ring-like structures, more strongly suggesting chloroplast targeting (Figure 4.4C).

4.5 Subcellular Localization of Full Length Fusion Constructs in Transfected Arabidopsis Protoplasts using a High Resolution Confocal Microscope

To further differentiate the subcellular localization of the two OEP18 constructs, transfected Arabidopsis protoplasts were examined using a confocal microscope (Figure 4.5). Z-stack (3-dimensional) reconstructed images were captured for each construct, as were optical slices representing the view of a transfected protoplast from a single optical plane.

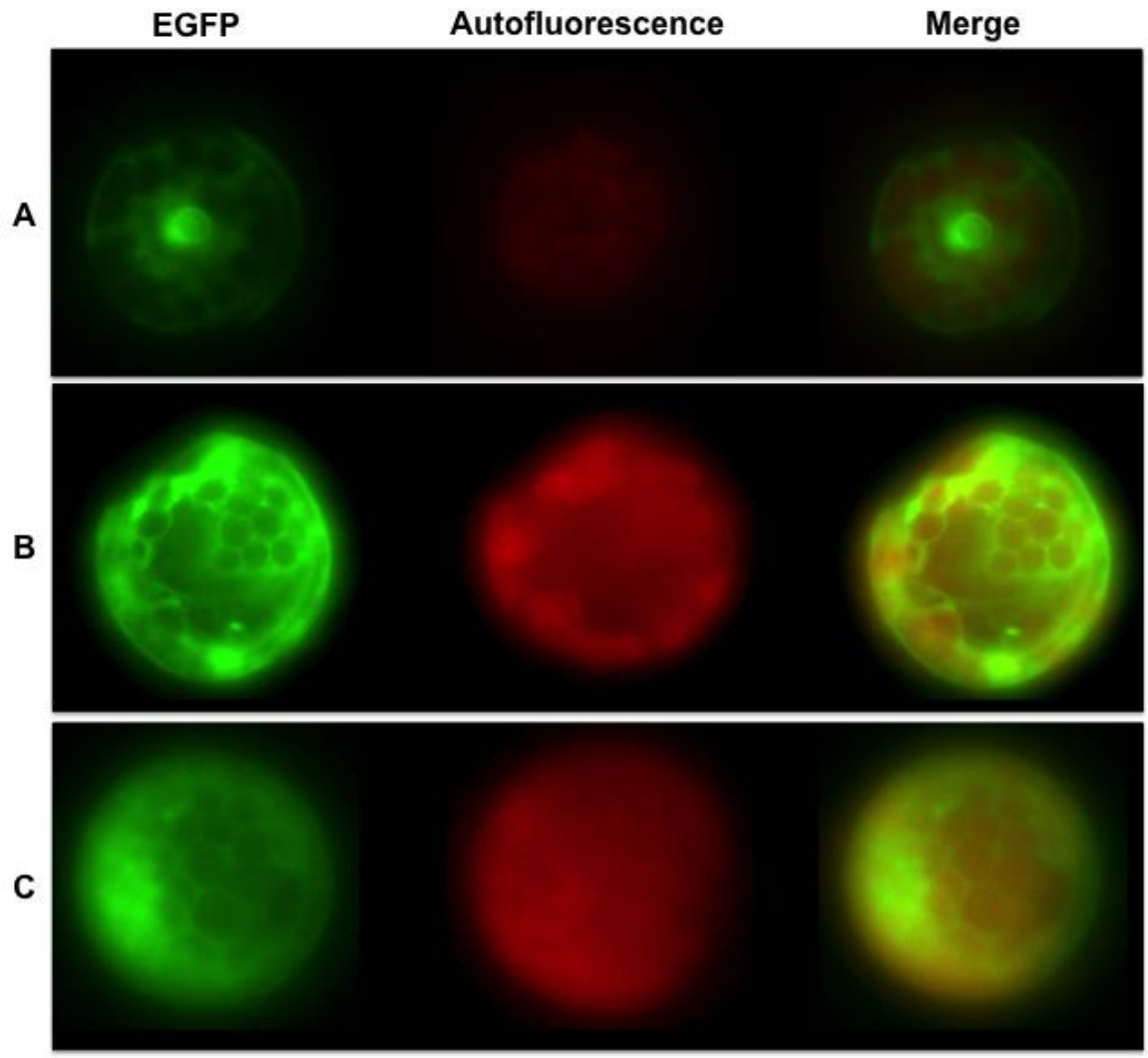


Figure 4.4. Transient expression of EGFP fusion proteins with full-length sequences of OEP18 in Arabidopsis protoplasts

Full-length OEP18 fusion constructs transiently expressed in Arabidopsis protoplasts and visualized using epifluorescence microscopy. Representative images from multiple independent experiments of protoplasts transfected with **(A)** pSAT6-N1 control, **(B)** EGFP-OEP18FL, and **(C)** OEP18FL-EGFP are presented. For each construct, representative images of EGFP (green), autofluorescence (red), and a merge of the two channels are displayed.

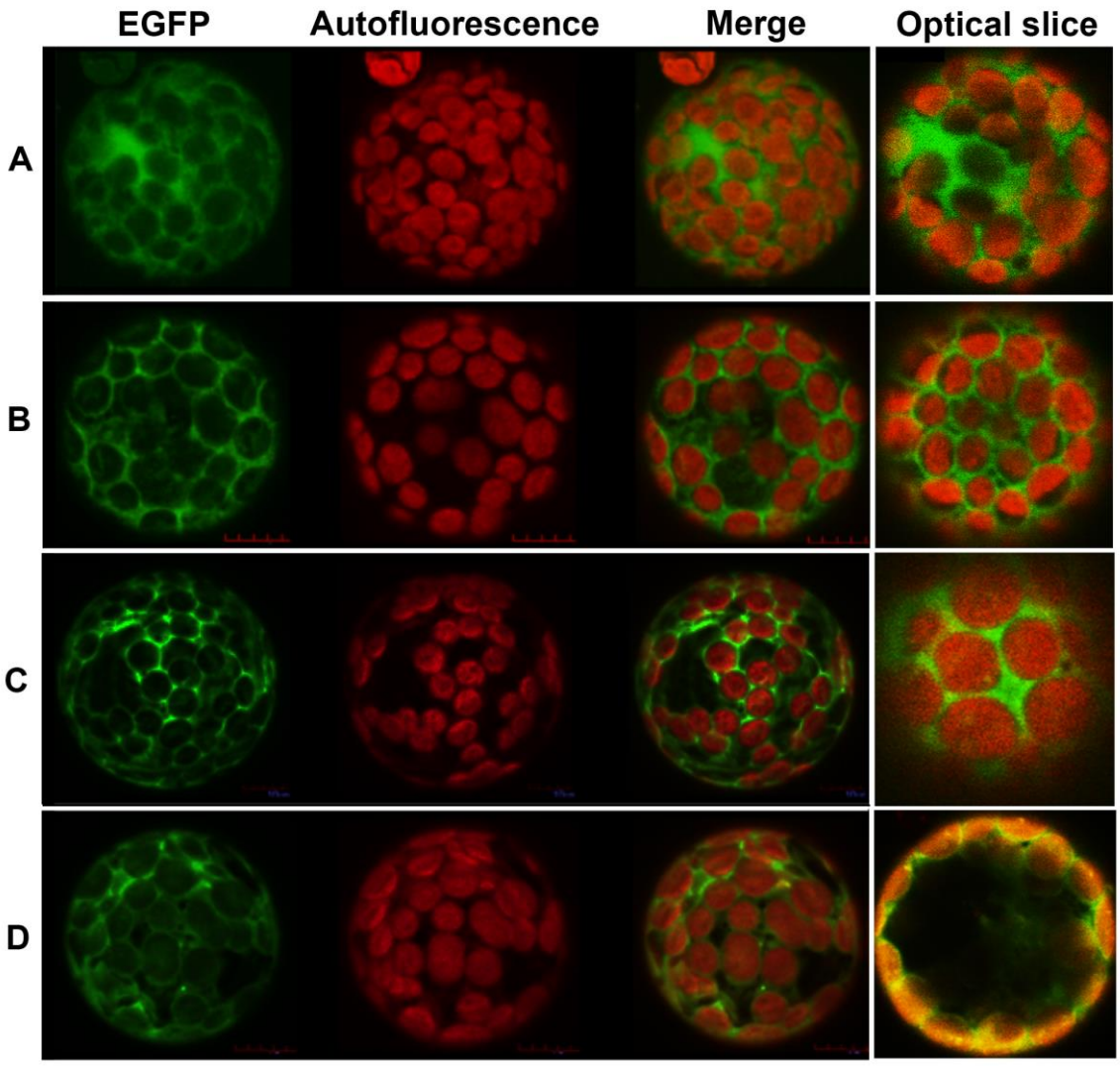


Figure 4.5. Full-length OEP18 fusion constructs transiently expressed in Arabidopsis protoplasts and visualized using confocal microscopy

3-D projections and merged optical slices of protoplasts transformed with **(A)** pSAT6-C1 negative control, **(B)** Toc34 positive control, **(C)** EGFP–OEP18FL, and **(D)** OEP18FL–EGFP constructs. Representative images were taken from multiple independent experiments. For each construct, representative images of EGFP (green), chlorophyll autofluorescence (red), and a merge of the two channels are displayed.

The empty pSAT6-C1 vector containing EGFP was used as a negative control. The large spaces between the chloroplasts filled with green fluorescence indicate pSAT6-C1 was expressed mostly in the cytoplasm (Figure 4.5A). Toc34, a known OEP (Bauer et al., 2002), was used as a positive control. Protoplasts expressing this construct showed a thin ring-like appearance of green fluorescent signals surrounding the red autofluorescence, indicating chloroplast outer membrane expression (Figure 4.5B). Cells transfected with the EGFP–OEP18FL construct displayed some thin ring-like patterns in the z-stack projection, but the optical slices displayed thicker ring-like structures and green fluorescent signals filling the spaces between red autofluorescence, indicating cytosolic expression (Figure 4.5C). The z-stack projection of cells expressing the OEP18FL–EGFP construct displayed a distinct ring-like appearance of green fluorescent signals surrounding the red autofluorescence, indicating chloroplast outer membrane expression. The optical slice also showed relatively little green fluorescent signals between the red autofluorescence of the chloroplasts, indicating far less cytosolic expression in cells expressing the OEP18FL–EGFP construct (Figure 4.5D).

4.6 Detection of the Full-Length OEP18 fusion proteins *in planta* using Western Blot analysis

The transient expression of the full-length OEP18 fusion proteins *in planta* was verified using western blot analysis using an anti-EGFP antibody (Figure 4.6). Arabidopsis protoplasts were transfected with the full-length EGFP constructs and

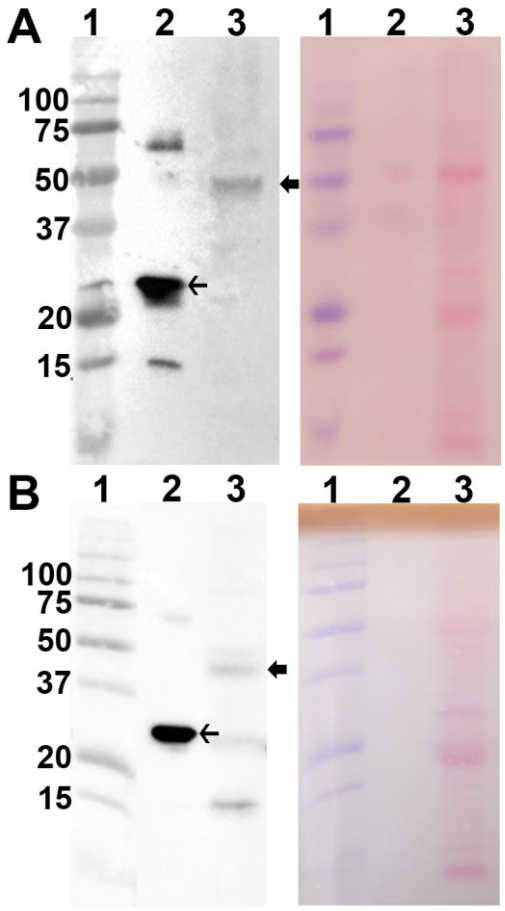


Figure 4.6. Detection of full-length OEP18 fusion constructs in transfected protoplasts using western blot analysis

(A) Total protein extracts from protoplasts transfected with the EGFP-OEP18FL or (B) OEP18FL-EGFP construct were separated by SDS-PAGE, followed by immunoblotting with an anti-EGFP polyclonal antibody. Recombinant EGFP (lane 2; skinny arrow) was used as a control and the protein ladder is labeled (lane 1). Full-length OEP18 fusion constructs are represented in lane 3 (thick arrow). The right panels show the Ponceau stained PVDF membranes to demonstrate a successful transfer of protein from the SDS PAGE gel. Numbers to the left indicate the position of the marker proteins in kilodaltons.

transfection rate was estimated using epifluorescence microscopy. Protoplasts with transfection rates of at least 60% were used in Western blot analysis. The expected molecular mass of the fusion protein (EGFP, 27 kDa and OEP18,18 kDa) was 45 kDa. EGFP–OEP18FL (Figure 4.6A) and OEP18FL–EGFP (Figure 4.6B) were detected at approximately 45 kDa, indicating their presence of the full-length fusion proteins in the protoplast transient expression assays.

4.7 Construct Design for OEP18 Truncation Constructs in a pSAT6-N1 Vector

In order to identify the sorting signal used by OEP18 to target itself to chloroplasts, a number of transient expression constructs were created by fusing various regions of the protein to the N-terminus of EGFP (Figure 4.7). The size of each truncation was based on the ChloroP prediction of transit peptide lengths at the N- and C- terminus of OEP18 (Table 3.1, Table 3.2). Transit peptides were predicted to be 23 and 72 amino acids in length at the C-terminus and N-terminus, respectively.

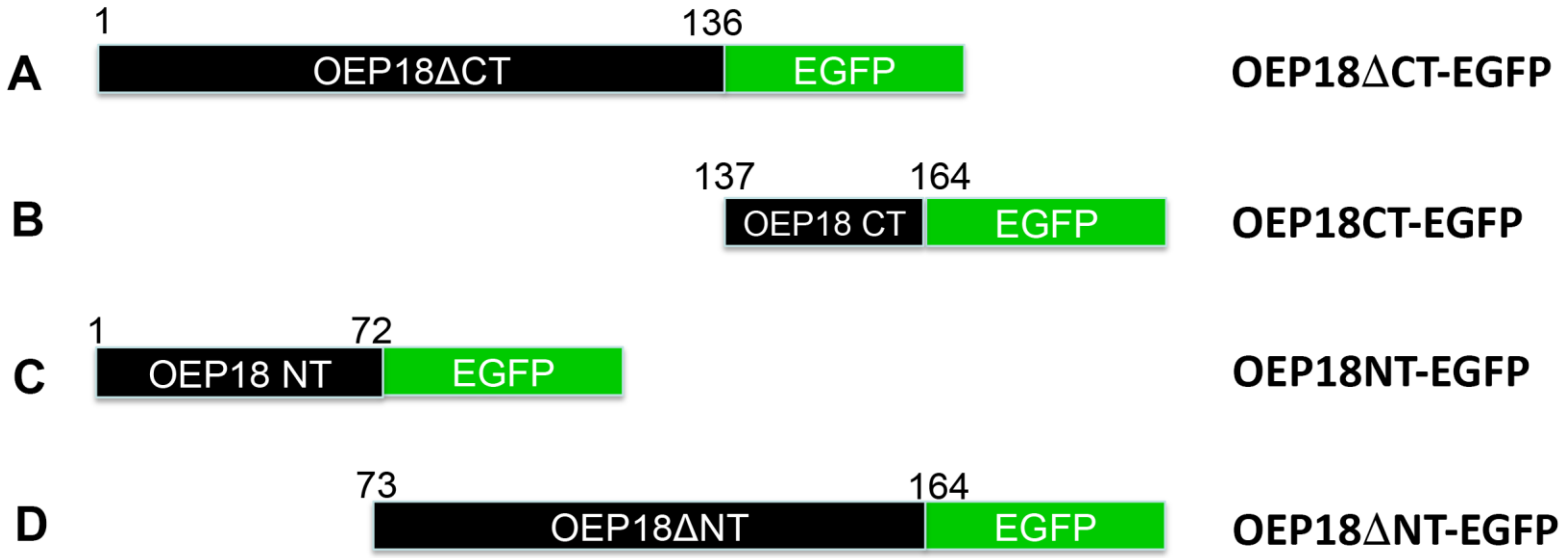


Figure 4.7. Schematic map of OEP18 deletion fusion constructs

(A) The 27 amino acids at the C-terminus were deleted and the rest of the protein was fused to EGFP (i.e. OEP18ΔCT–EGFP). **(B)** The 27 amino acids at the C-terminus of OEP18 were fused to EGFP (i.e. OEP18CT–EGFP). **(C)** The 72 amino acids at the N-terminus of OEP18 were fused to EGFP (i.e. OEP18NT–EGFP). **(D)** The 72 amino acids at the N-terminus were deleted and the rest of the protein was fused to EGFP (i.e. OEP18ΔNT–EGFP). Truncation designs were based on the transit peptides lengths predicted by ChloroP.

4.8 Onion Epidermal Cells Bombarded with OEP18 Truncated Constructs

The OEP18 truncated constructs were bombarded in onion epidermal cells to verify their expression and acquire subcellular targeting information (Figure 4.8). Cells transformed with OEP18ΔCT–EGFP and OEP18NT–EGFP constructs showed EGFP expression in the nucleus and cytoplasm (Figure 4.8A and C). EGFP signals from OEP18CT–EGFP and OEP18ΔNT–EGFP formed mostly punctate structures and showed much less nuclear and cytoplasmic expression (Figure 4.8B and D).

4.9 Verifying the Identity of the Punctate Structures through Colocalization Analysis between Truncation Constructs and DsRed in Onion Epidermal Cells

Onion epidermal cells were co-bombarded with the OEP18 truncation constructs and DsRed to further acquire subcellular targeting information and to determine the identity of the punctate structures (Figure 4.9). In cells co-transfected with the OEP18CT–EGFP or OEP18ΔNT–EGFP and the ferredoxin TP-DsRed constructs, the EGFP signals colocalized with DsRed signals, as shown by the yellow punctate signals from the merge of the two channels. This is also supported by the distribution of the two signals clustering along the diagonal line in the scatter plots produced using the Fiji "colocalization threshold" plug-in ImageJ (National Institutes of Health, USA). Furthermore, Pearson's correlation coefficients (R_r) for OEP18CT–EGFP and OEP18ΔNT–EGFP were high (0.6723 ± 0.0871 and 0.7395 ± 0.0432, respectively; Table 4.2).

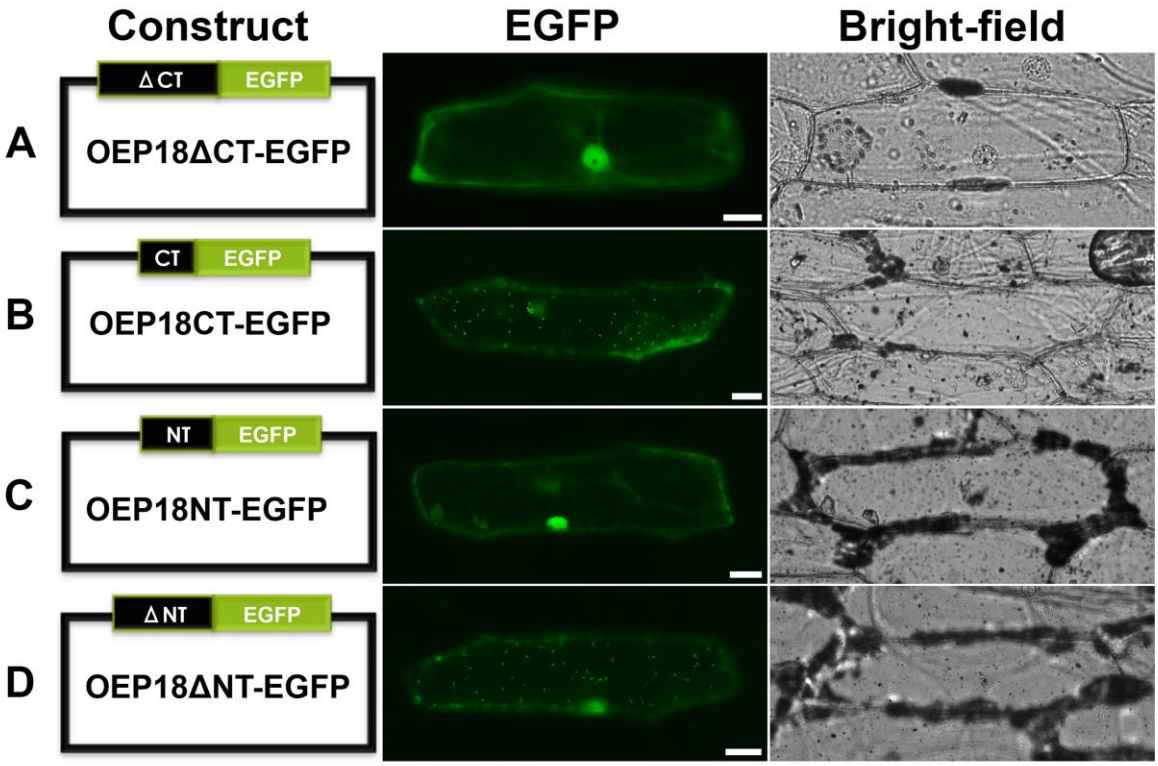


Figure 4.8. Transient expression of EGFP fusion proteins with partial OEP18 sequences in onion epidermal cells

Onion epidermal cells bombarded with OEP18 truncation constructs to examine their subcellular localizations. **(A)** OEP18ΔCT–EGFP, **(B)** OEP18CT–EGFP, **(C)** OEP18NT–EGFP, and **(D)** OEP18ΔNT–EGFP are representative images from multiple independent experiments. Scale bars = 50 μm

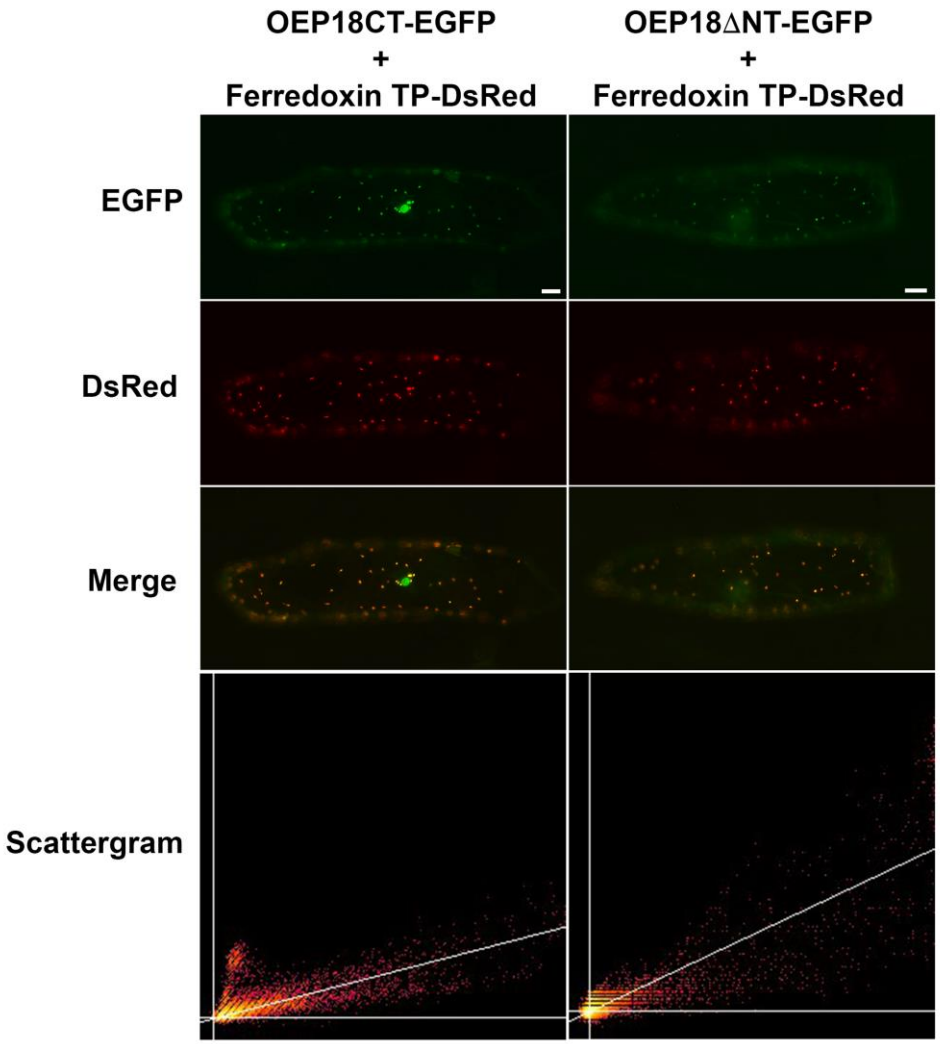


Figure 4.9. Colocalization analysis of EGFP fusion proteins with partial OEP18 sequences in onion epidermal cells

Onion epidermal cells were co-bombarded with OEP18 truncation constructs and the ferredoxin transit peptide fused to DsRed. Representative images from multiple independent experiments of EGFP (green signal), DsRed (red signal), and a merge of the two channels are shown for OEP18CT–EGFP and OEP18ΔNT–EGFP. Scatterplots were generated using ImageJ to demonstrate the colocalization between EGFP and DsRed signals for both truncated constructs. Scale bars = 50 μm

Table 4.2. Pearson's correlation coefficients (R_r) of the two fluorescent channels for the OEP18CT–EGFP and OEP18ΔNT–EGFP constructs in co-bombarded onion epidermal cells with DsRed tagged to the transit peptide of ferredoxin.

Construct	OEP18CT-EGFP	OEP18ΔNT-EGFP
R_r	0.6723 ± 0.0871	0.7395 ± 0.0432

Values represent the mean from four replicates (N=4) of each construct (± SD) from multiple independent experiments was calculated. The maximum theoretical R_r score is 1.

4.10 Visualizing Arabidopsis Protoplasts Transfected with OEP18 Truncation Constructs by Epifluorescence Microscopy

Arabidopsis protoplasts were isolated and transfected with OEP18 truncation constructs and subcellular targeting was visualized using epifluorescence microscopy (Figure 4.10). In protoplasts transformed with the OEP18ΔCT–EGFP construct, the EGFP signals were found in the nucleus and cytoplasm as irregular-shaped punctate structures that likely represent protein misfolding or proteins targeted to non-plastid organelles (Figure 4.10A). Protoplasts transfected with the OEP18CT–EGFP construct showed some green fluorescent signals forming ring-like structures indicating chloroplast targeting (Figure 4.10B). Protoplasts transfected with the OEP18NT–EGFP construct showed EGFP signals in the nucleus and cytoplasm (Figure 4.10C). Protoplasts transfected with the OEP18ΔNT–EGFP construct displayed EGFP signals in the nucleus and also some ring-like structures indicating chloroplast targeting (Figure 4.10D).

4.11 Subcellular Localization of OEP18 Truncation Constructs in Transfected Arabidopsis Protoplasts using a High Resolution Confocal Microscope

A high resolution confocal microscope was used to further differentiate the subcellular localization of the truncation constructs transfected in Arabidopsis protoplasts (Figure 4.11). Z-stack (3-dimensional) reconstructed images were captured for each construct, as well as optical slices representing the view of a transfected protoplast from a single optical plane. Protoplasts expressing the

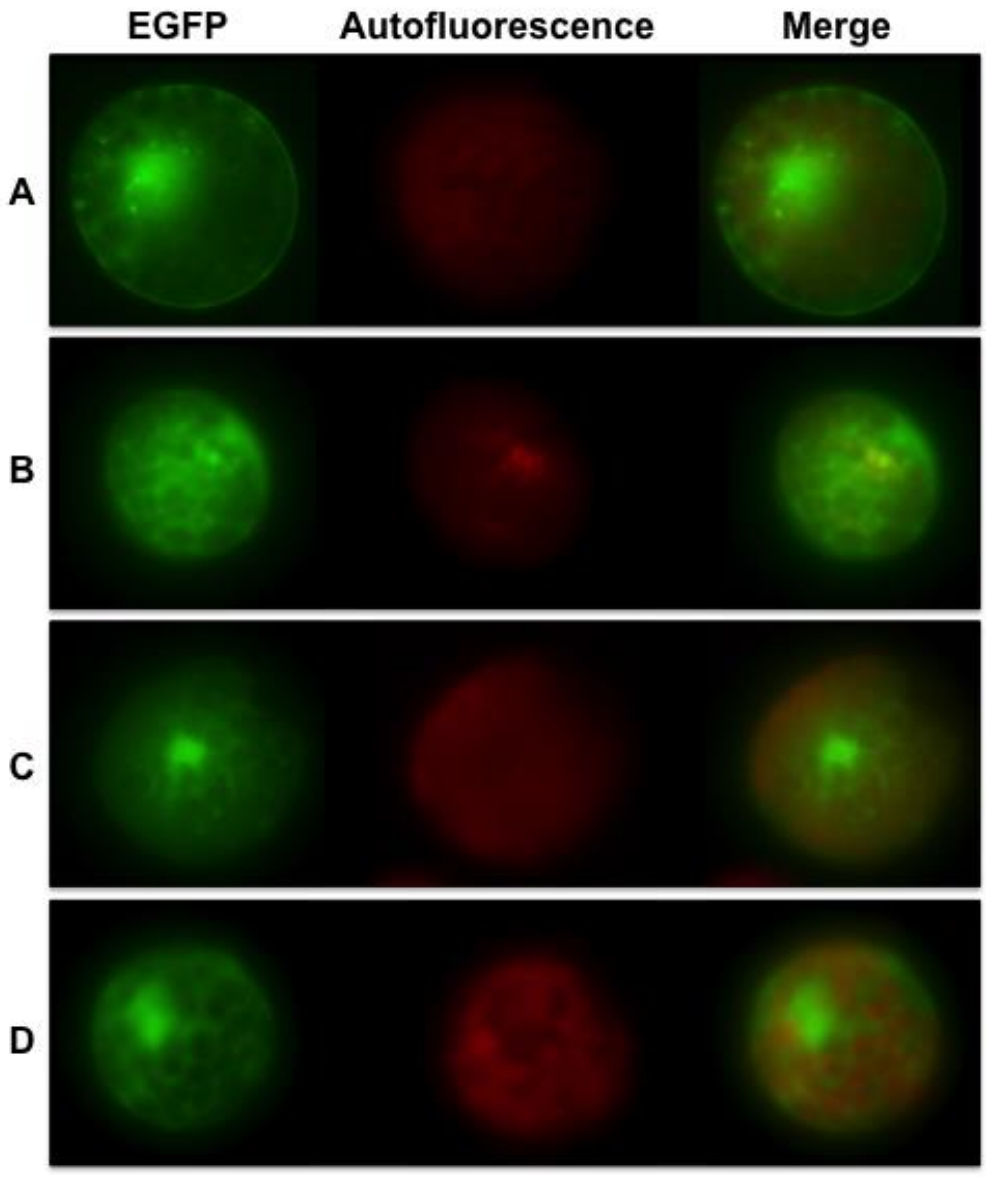


Figure 4.10. OEP18 truncation constructs transiently expressed in Arabidopsis protoplasts and visualized by epifluorescence microscopy.

(A) OEP18ΔCT–EGFP, **(B)** OEP18CT–EGFP, **(C)** OEP18NT–EGFP, and **(D)** OEP18ΔNT–EGFP are representative images from multiple independent experiments. For each construct, representative images of EGFP (green), autofluorescence (red), and a merge of the two channels are displayed.

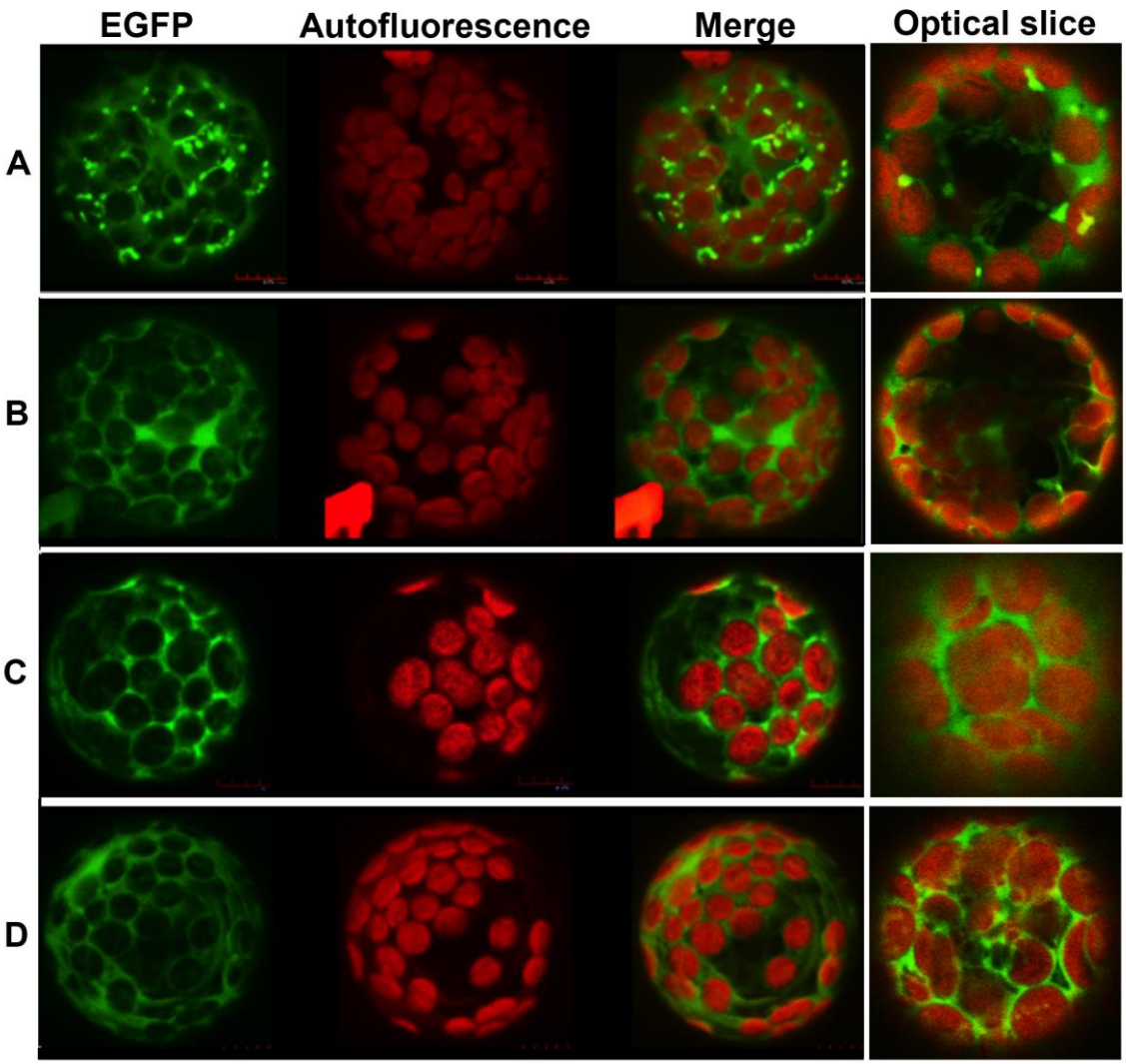


Figure 4.11. OEP18 truncation constructs transiently expressed in Arabidopsis protoplasts and visualized by laser scanning confocal microscopy.

3-D projections and merged optical slices of protoplasts transformed with **(A)** OEP18ΔCT–EGFP, **(B)** OEP18CT–EGFP, **(C)** OEP18NT–EGFP, and **(D)** OEP18ΔNT–EGFP constructs. Representative images are displayed from multiple independent experiments. For each construct, representative images of EGFP (green), chlorophyll autofluorescence (red), and a merge of the two channels are displayed.

OEP18ΔCT–EGFP construct displayed mostly irregular-shaped punctate structures in the cytoplasm indicating protein misfolding or proteins targeted to non-plastid structures (Figure 4.11A). In protoplasts transformed with the OEP18CT–EGFP construct, some EGFP signals were found in the nucleus, while some green fluorescent ring-like structures surrounded the red autofluorescence, indicating chloroplast outer membrane targeting. The optical slice for the OEP18CT–EGFP construct shows minimal green fluorescent signal accumulation between the red autofluorescence, which is also an indication of chloroplast outer membrane targeting (Figure 4.11B). The z-stack projection of protoplasts transfected with the OEP18NT–EGFP construct showed some thin ring-like structures indicating chloroplast outer membrane targeting, but the optical slices displayed green fluorescent signal accumulation between the red autofluorescence, indicating mostly expression in the cytoplasm (Figure 4.11C). The z-stack projection and optical slices of protoplasts transformed with the OEP18ΔNT–EGFP construct displayed mostly thin green fluorescent ring-like structures surrounding the red autofluorescence, indicating chloroplast outer membrane targeting (Figure 4.11D).

5. Discussion

The majority of chloroplast-destined proteins are encoded in the nucleus and targeted to chloroplasts after being synthesized by cytosolic ribosomes (Jarvis, 2008). Most chloroplast proteins are imported via N-terminal transit peptides (TPs) (Keegstra and Cline, 1999). However, with the exception of Toc75 (Tranel et al., 1995), outer envelope proteins (OEPs) of chloroplasts are targeted via different mechanisms (Hofmann and Theg, 2005). Multiple OEP targeting pathways are known, but with the recent spike in the identification of new OEPs, the mechanisms for their targeting have not been completely elucidated. The current study aimed to determine if a recently-identified novel OEP targeting mechanism characterized in BsToc159 is used by any other OEPs.

5.1 Identification of Candidate OEPs of Study using a Unique Bioinformatic Approach

Bioinformatic analysis on the one hundred and seventeen known or predicted Arabidopsis chloroplast OEPs provided an essential foundation for the current study. ChloroP predicts the presence of a potential chloroplast transit peptide (cTP) in an amino acid sequence submitted by the user. Since transit peptides are used to target proteins to the chloroplast stroma and not to the COM (with the exception of Toc75), analyzing OEPs using ChloroP is seemingly unorthodox. Furthermore, using ChloroP to predict the presence of TP-like sequences at the C-terminus by analyzing amino acid sequences in the reverse order is even more unconventional. However, this method of analysis was recently shown

to be successful in identifying a novel chloroplast targeting signal in the OEP Toc159 of *Bienertia sinuspersici* (Lung et al., 2014). ChloroP was originally used to predict a TP-like sorting signal at the C-terminus of BsToc159 (Lung and Chuong, 2012). Eight of the 117 known or predicted Arabidopsis chloroplast OEPs were predicted by ChloroP to share a similar TP-like targeting signal at their C-terminus in the current study (Table 3.1). Further bioinformatic analyses indicated that two of the eight candidate proteins contained a transmembrane domain, which suggests that these proteins may use one of the established OEP targeting pathways and were thus eliminated as candidates in the current study (Table 3.2). ChloroP predicted OEP18 (At5G42070) with the highest probability to contain a TP-like sorting signal at its C-terminus, and therefore was the protein of focus for the rest of the study.

5.2 OEP18 can Target to Plastids when Expressed as an EGFP Fusion protein using a pSAT6-N1 Vector

The entire coding sequence for OEP18 was fused to the amino and carboxy ends of EGFP using pSAT6-N1 and pSAT6-C1 vectors to generate the OEP18FL–EGFP and EGFP–OEP18FL fusion proteins, respectively (Figure 4.1). The purpose of generating the two different full-length fusion proteins was to determine if there were any differences in chloroplast targeting when EGFP is fused to the N- or C-terminus of OEP18. Onion epidermal cells were bombarded with the full-length fusion constructs to confirm protein expression in a plant cell system and acquire protein targeting information. Results from the onion epidermal cells show largely nuclear and cytosolic expression for the EGFP–OEP18FL construct (Figure 4.2B). In

contrast, there were numerous punctate structures present in cells expressing the OEP18FL–EGFP construct (Figure 4.2C). Under high magnification, long extensions resembling stromules, which are indicative of etioplasts (Lung et al., 2014), can be seen emanating from the punctate structures present in cells bombarded with the OEP18FL–EGFP construct (Figure 4.2C inset). In comparison, under high magnification, the green punctate signals produced in the onion epidermal cells bombarded with the EGFP–OEP18FL construct were irregular-shaped and did not contain long extensions. Moreover, co-bombardment experiments with DsRed fused to the ferredoxin TP confirmed that the punctate structures were etioplasts in cells expressing the OEP18FL–EGFP construct. The signals from the OEP18FL–EGFP construct and the signals from DsRed colocalized as indicated by the yellow signals produced in the merge of the two channels (Figure 4.3). Furthermore, the two fluorescent signals clustering along the diagonal line in the scatter plot and the high Pearson's correlation coefficient confirmed colocalization between OEP18FL–EGFP and DsRed signals (Figure 4.3; Table 4.1). Co-bombardment of cells with the constructs encoding EGFP–OEP18FL and DsRed fused to the ferredoxin TP, on the other hand, produced comparatively less overlap between the fluorescent signals (Figure 4.3). The Pearson's correlation coefficient for the cells expressing EGFP–OEP18FL and DsRed fused to the ferredoxin TP was slightly higher than the control (Table 4.1). However, this is likely due to only partial targeting of fusion proteins to plastids and because the irregular-shaped EGFP punctate structures overlap with or reside in close proximity to some of the DsRed-localized plastids. It was concluded

that OEP18 retains best its ability to target to plastids when EGFP is fused to its C-terminus (OEP18FL–EGFP).

Both OEP18 full-length EGFP fusion constructs were also expressed in Arabidopsis protoplasts to determine if their targeting behavior is consistent in chloroplast-containing cells. Epifluorescence imaging indicated the presence of ring-like structures surrounding the red autofluorescence from chlorophyll representing chloroplasts (Figure 4.4). Chloroplast targeting was apparent for OEP18FL–EGFP, highlighted by the green fluorescent signals that produced fine ring-like structures (Figure 4.4C) in comparison to the slightly more diffuse ring-like structures present in the protoplasts transformed with the EGFP–OEP18FL construct (Figure 4.4B). To confirm this observation at a higher resolution, confocal laser scanning microscopy was used to examine the subcellular distribution for the same full-length constructs. The z-stack 3-dimensional projections from the confocal microscopy showed fine ring-like structures in the protoplasts transformed with the OEP18FL–EGFP construct, indicating targeting to the chloroplast outer envelope (Figure 4.5D). Individual optical slices were also examined to rule out the possibility of observing a false ring in the 3-dimensional projections. The optical slices also reveal distinct, ring-like structures surrounding chloroplasts in cells expressing the OEP18FL–EGFP construct (Figure 4.5D), which closely resembled the ring-structures present in the cells expressing the positive control, Toc34 (Figure 4.5B). The EGFP signals produced diffuse ring-like structure for the cells expressing the EGFP–OEP18FL construct (Figure 4.5C), which resembled the diffuse green fluorescent signal expression present in the cells expressing the negative control,

pSAT6-C1 construct (Figure 4.5A). Overall, these data from the onion epidermal cells and transformed protoplasts indicate that the OEP18FL–EGFP construct is likely targeted to the chloroplast outer membrane. The data from the bombarded onion epidermal cells and transformed protoplasts also indicate that the EGFP–OEP18FL construct is expressed mostly in the nucleus and cytoplasm and is not efficiently targeted to chloroplasts. Therefore, chloroplast targeting is achieved when EGFP is fused to the C-terminus of OEP18 (OEP18FL–EGFP) and not targeted effectively when EGFP is fused to the N-terminus of OEP18 (EGFP–OEP18FL). Therefore, fusion proteins with OEP18 on the N-terminus of EGFP were used for the remainder of the study.

5.3 The C-terminus of OEP18 contains chloroplast outer membrane targeting information

Once it was established that OEP18 must be expressed in front of EGFP to achieve efficient targeting of the fusion protein to the chloroplast outer membrane, OEP18 truncation constructs were designed (Figure 4.7) and expressed in onion epidermal cells and Arabidopsis protoplasts to determine what domain(s) of the OEP18 protein are required for chloroplast targeting. Onion epidermal cells showed punctate structures in the cells bombarded with the OEP18CT–EGFP and OEP18ΔNT–EGFP constructs indicating plastid targeting (Figure 4.8 B and D). In contrast, the onion epidermal cells bombarded with the OEP18ΔCT–EGFP and OEP18NT–EGFP constructs showed mostly nuclear and cytoplasmic expression (Figure 4.8 A and C). The punctate structures were confirmed to be plastids using

co-bombardment with DsRed fused to the ferredoxin TP. Colocalization between the OEP18CT–EGFP or OEP18ΔNT–EGFP constructs and the ferredoxin TP fused to DsRed was demonstrated by the yellow signals produced in the merge of the EGFP and DsRed signals (Figure 4.9). The scatter plots that showed clustering of signals around the diagonal line (Figure 4.9) and the high Pearson's correlation coefficients (Table 4.2) further indicated a high degree of colocalization between the OEP18CT–EGFP or OEP18ΔNT–EGFP constructs and the ferredoxin TP fused to DsRed. These data indicate that the C-terminus of OEP18 is necessary for targeting the protein to plastids.

Epifluorescence imaging of OEP18 truncation constructs expressed in Arabidopsis protoplasts show that when the C-terminus of OEP18 is present, EGFP signals form fine ring-like structures around the chloroplasts, as compared to the more diffuse ring-like structures and misfolded protein aggregates when the C-terminus of OEP18 is absent (Figure 4.10). Z-stack projections were generated using a high-resolution confocal microscope, which showed EGFP signals producing fine ring-like structures surrounding chloroplasts in the protoplasts transfected with constructs containing the OEP18 C-terminus (Figure 4.11 B and D). In comparison, EGFP signals producing diffuse ring-like structures were observed in protoplast transfected with the construct without the C-terminus of OEP18 (Figure 4.11 A and C). The possibility of observing a false ring from the z-stack projections was ruled out by examining individual optical slices for each construct. The individual optical slices revealed that when the C-terminus of OEP18 is expressed in protoplasts, fine ring-like structures produced from EGFP signals formed around the

red autofluorescence, indicating chloroplast outer membrane targeting (Figure 4.11 B+D). In comparison, when protoplasts were expressing OEP18 constructs with their C-terminus deleted, individual optical slices revealed diffuse green signal accumulation between the chloroplasts, indicating protein expression in the cytoplasm (Figure 4.11 A+C). The results from the transfected Arabidopsis protoplasts support the results from onion epidermal cell bombardments and it can be concluded that the C-terminus of OEP18 is necessary to target the protein to the chloroplast outer membrane.

Results from the bioinformatic analyses strongly predicted that chloroplast targeting information would be present in the C-terminus of OEP18. Out of all 117 known or predicted OEPs of Arabidopsis analyzed by ChloroP, OEP18 received the highest score for potentially containing a TP-like targeting signal at its C-terminus (Table 3.1). Furthermore, the secondary structure prediction of OEP18, using the bioinformatics tool PSIPRED, predicted the presence of an α-helix within the first 10 amino acids residues that make up the OEP18 C-terminus (Figure 3.2). Similarly, an α-helix was also predicted in the predicted 51-amino acid TP-like chloroplast targeting signal of BsToc159 (Lung et al., 2014). About 30% of transit peptides contain an α-helix within the first 10 amino acids (Huang et al., 2009). This structure may serve as a targeting signal for OEP18 to the chloroplast outer membrane.

The amino acid composition (Figure 3.1) of OEP18 strongly suggests that chloroplast targeting information is present in the C-terminus, and not the N-terminus. Transit peptides typically have an overrepresentation of serine and

threonine residues (Zhang and Glaser, 2002). Fifteen out of the twenty-eight amino acids (54%) that make up the C-terminal end of OEP18 are serine and threonine residues (Table 3.3). This TP-like property is consistent with the C-terminus of OEP18 containing a TP-like targeting signal. In comparison, only 20 out of the 72 amino acids (28%) that make up the N-terminus of OEP18 are serine and threonine residues (Table 3.3). This low representation of hydroxylated amino acid residues indicates that the N-terminus of OEP18 likely does not contain a TP-like targeting signal. Furthermore, only 2 out of 28 amino acids at the C-terminal end of OEP18 are aspartic acid and glutamic acid (Table 3.3). Transit peptides generally lack acidic amino acids (Zhang and Glaser, 2002). This further indicates that the C-terminus of OEP18 is likely to contain a TP-like targeting signal.

Collectively, these data from the onion epidermal cell bombardments, protoplast transient expression assays, bioinformatic analyses, and amino acid composition analyses lead to the conclusion that the C-terminus of OEP18 appears to contain a TP-like targeting signal that is necessary and sufficient for targeting the protein to chloroplasts. Although ChloroP predicted the presence of a transit peptide in the N-terminus of OEP18, data from the onion epidermal cell bombardments and Arabidopsis protoplast transformations (Figure 4.11C) combined with amino acid composition analysis indicate that the N-terminus of OEP18 does not likely contain a TP-like signal and it cannot efficiently target OEP18 to chloroplasts.

5.4 OEP18 cannot be categorized into any of the broadly classified OEP targeting families

The canonical chloroplast targeting mechanisms for many OEPs are characterized based on their transmembrane domains (Lee et al., 2013). Besides Toc75, the other identified integral β-barrel proteins (OEP21, OEP24, OEP37) appear to self-insert into the chloroplast outer membrane via membrane-spanning domains that also contain the proteins' targeting information (Pohlmeyer et al., 1998; Bölter et al., 1999, Goetze et al., 2006). In the current study, not only did the secondary structure prediction analysis indicate no β-strands in the C-terminus of OEP18, but also there are not nearly enough β-strands present within the total OEP18 sequence for it to even possibly form a β-barrel (Figure 3.2). Furthermore, of the few β-strands predicted in the N-terminus and central region of OEP18, they are too short (mean = 3.25 residues per strand) to constitute a membrane-spanning region (6–25 residues) (Taylor et al., 2006). Therefore, OEP18 is not a β-barrel protein that self-inserts into the chloroplast outer membrane.

There are two broadly classified OEP families characterized by the location of α-helical transmembrane domains within the protein (Hofmann and Theg, 2015). OEPs that contain a single transmembrane domain at their N-terminus or at their C-terminus are referred to as "signal-anchored" and "tail-anchored" proteins, respectively (Bölter and Soll, 2011; Dhanoa et al., 2010). This transmembrane domain functions as a membrane anchor, as well as a targeting signal (Bölter and Soll, 2011). Results from the current study demonstrate that the C-terminus of OEP18 is necessary and sufficient to target the protein to plastids, and specifically to

the chloroplast outer membrane. The N-terminus alone was not sufficient to target OEP18 to the chloroplast outer membrane (Figure 4.11 A+C). Therefore, OEP18 cannot be classified as a signal-anchored protein, since the protein does not have chloroplast targeting information in its N-terminus. To be classified as a tail-anchored protein, the OEP must (I) have the majority of the protein exposed to the cytosolic side, (II) contain a transmembrane domain at or near the C-terminus, and (III) the C-terminal tail must protrude into the organelle interior (Kutay et al., 1993; Abell and Mullen, 2011). The data from the current study show that the C-terminus of OEP18 contains the chloroplast outer membrane targeting information. However, much like what was recently identified in BsToc159 (Lung and Chuong, 2012), bioinformatic analysis using TMHMM2.0 indicates OEP18 does not contain a transmembrane domain (Table 3.2). Therefore, despite having targeting information in its C-terminus, if OEP18 does not contain a transmembrane domain, it cannot be classified as a tail-anchored protein. Overall, OEP18 does not appear to fall into any of the broadly classified families of OEP targeting mechanisms.

5.5 The targeting mechanism used by OEP18 resembles the new class of sorting signal recently identified in Toc159

The lack of a canonical transmembrane domain and the presence of a TP-like sorting signal at the OEP18 C-terminus suggest that OEP18 may belong to the newly identified class of sorting signal found in BsToc159 (Lung et al., 2014). Future experiments need to be conducted in order to conclude that OEP18 shares this novel chloroplast outer membrane sorting signal. More truncation mutants can be made

for OEP18 to further dissect and define the specific sequence within the C-terminus required for chloroplast outer membrane targeting and association. The C-terminal truncation used in the current study was designed based on the ChloroP prediction of a transit peptide 23 amino acids in length (Table 3.1), and was shown to target to the chloroplast outer membrane. However, the majority of chloroplast TPs are 30 to 80 amino acid residues long (Zhang and Glaser, 2002). ChloroP analysis on BsToc159 by Lung and Chuong (2012) predicted a 51- residue length TP-like sorting signal at the C-terminus. The 51 C-terminal residues were capable of targeting BsToc159 to the chloroplast outer membrane. However, it was later demonstrated that including an additional five amino acid residues upstream from the 51-residue C-terminal truncation (for a total of 56 C-terminal residues) targeted BsToc159 to the chloroplast outer membrane with higher efficiency (Lung et al., 2014). This difference in efficiency was visualized in Arabidopsis protoplasts and was also determined using chloroplast fractionation and western blot analysis. Therefore, designing more truncation mutants and determining chloroplast targeting efficiency will better define the targeting sequence within the OEP18 C-terminus. Furthermore, the design of more C-terminal truncations in BsToc159 helped identify the novel membrane association region 60–100 residues upstream of the C-terminus (Lung et al., 2014). OEP18, like BsToc159, is predicted to contain no canonical transmembrane domain (Table 3.2) and thus, likely associates with the chloroplast outer membrane using a non-canonical method. Designing truncation constructs using parts of the OEP18 C-terminus will help identify the sequence

required for chloroplast outer membrane association. It is possible that OEP18 is a peripheral membrane protein.

5.6 Other potential TP-like OEP targeting candidates

Results from the ChloroP analysis combined with TMHMM2.0 analysis indicated that five other known or predicted OEPs of Arabidopsis chloroplast may contain a TP-like sorting signal at their C-terminus (Table 3.1; Table 3.2). The same methods from the current study could be applied to these other OEPs to determine if they also potentially belong to the same novel class of chloroplast outer membrane sorting signal.

5.7 Integrating bioinformatic approaches leads to characterizing novel targeting mechanisms in other organelles beyond chloroplasts

In recent years, unique bioinformatic approaches and machine learning techniques have led to characterizing novel intracellular targeting mechanisms in organelles besides chloroplasts (Angermueller et al., 2016). Furthermore, these approaches are helping to discover new proteins putatively residing in the organelle of interest. For instance, tail-anchored proteins have always been recognized as a class of proteins integral to all cellular membranes and are defined by a single transmembrane domain near the C-terminus (Kutay et al., 1993). Tail-anchored proteins are very complex, as they must differentiate between targeting to the ER, mitochondria, and in the case of plant cells, plastids (Abell and Mullen, 2011). Marty et al. (2014) recently demonstrated that a dibasic targeting signal motif, originally

identified in the electron carrier cytochrome b_5 (Hwang et al., 2004), is present in many known tail-anchored proteins residing in the outer membrane of mitochondria. Marty et al. (2014) combined predictions from a variety of bioinformatic tools, which led to mutational analysis of the dibasic motif, further showing that the targeting motif is far more divergent than previously defined. The newly expanded targeting motif led to the novel identification of forty-three candidate tail-anchored proteins containing the putative mitochondrial outer membrane dibasic targeting signal motif in Arabidopsis (Marty et al., 2014).

A subset of peroxisome-destined proteins is targeted to peroxisomes using a PTS1 signal sequence. PTS1 sequences have been recognized to conform to a similar pattern of amino acids: small side chain amino acid–basic amino acid–hydrophobic amino acid (Baker et al., 2016). However, a recently developed computational prediction tool and *in vivo* subcellular targeting analyses demonstrated that the plant PTS1 motif is more diverse than previously known, including more non-canonical sequences and amino acid residues (Chowdhary et al., 2012). Also, the newly identified proteins containing a non-canonical PTS1 were found to have four to five amino acid residues in front of the targeting tripeptide that enhanced targeting to peroxisomes (Chowdhary et al., 2012). The newly developed PTS1 prediction tool by Chowdhary et al. (2012) led to *in vivo* targeting analysis that validated twenty-three new PTS1 tripeptides in Arabidopsis. This novel computational approach has opened the door to potentially identify more plant peroxisomal proteins using non-canonical PTS1s, and suggests that similar

approaches such as those used in the this study could reveal new targeting sequences for other organelles in plants as well.

5.8 Integrating multiple fields of study and the larger context

Molecular biology can sometimes lose its direction without putting the goals and objectives of a study into perspective. Taking a look at this field of study from a big picture point of view, plants are essential to the worldwide ecosystem. Photosynthetic plants are vital for maintaining atmospheric oxygen levels and providing the primary source of energy that drives metabolic processes in all living organisms. The underlying functions of plants are carried out by a series of sophisticated intracellular molecular processes, and studying these processes leads to an understanding of plant growth and development. The scientific tools used to study these processes are derived from a diverse body of research including biochemical, molecular, evolutionary, ecological, genetic, biophysical, chemical, and computational approaches. In the current study, computational approaches were used to generate scatterplots to determine plastid colocalization of OEP18, predict the secondary structure elements of OEP18, and predict which OEPs might contain the putative C-terminal TP-like sorting signal. Molecular and cellular approaches were integrated to design the OEP18-EGFP fusion constructs and examined their in vivo expression. Biochemical approaches were used to confirm the presence of the OEP-EGFP fusion proteins using western blot analysis. Microscopy was used to visualize the subcellular localization of various OEP18-EGFP fusion constructs in onion epidermal cells and Arabidopsis protoplasts. Biophysical methods, such as

structural analysis, can be used in the future to more accurately determine the secondary structural elements of OEP18. Other molecular approaches such as *in vitro* chloroplast targeting assays can be used as an alternative approach to fluorescence microscopy to determine OEP18 targeting information. The application of a variety of scientific approaches is crucial to maximize our understanding of complex biological phenomenon.

5.9 Conclusions

OEP18 was predicted using bioinformatics, in the same manner as BsToc159, to contain a TP-like sorting signal at its C-terminus. Since BsToc159 was shown to contain targeting information at its C-terminus (Lung et al., 2014), it was hypothesized that the C-terminus of OEP18 would also be responsible for targeting the protein to the chloroplast outer membrane. The current study has demonstrated that C-terminus of OEP18 is necessary and sufficient for targeting the protein to chloroplasts. When the C-terminus of OEP18 is removed, the protein is mostly expressed in the nucleus and cytoplasm, or forms punctate structures that are indicative of protein misfolding or mistargeted. Bioinformatic analyses indicated that OEP18 is not an integral β-barrel protein and does not contain a transmembrane domain. Therefore, OEP18 likely does not belong to one of the broadly classified families of OEP targeting mechanisms and furthermore, may potentially associate with the chloroplast outer membrane as a peripheral membrane protein. In the future, more OEP18 truncation constructs should be designed to define the exact sequence within the C-terminus required for targeting

and membrane association. In addition, structural analysis can be used to acquire more information on the type of association between OEP18 and the chloroplast outer membrane. Overall, the results from the current study indicate that the C-terminus of OEP18 contains targeting information to the chloroplast outer membrane. Furthermore, OEP18 may share the novel chloroplast targeting mechanism first characterized in Toc159.

6. References

Abell BM. Mullen RT. 2011. Tail-anchored membrane proteins: exploring the complex diversity of tail-anchored-protein targeting in plant cells. Plant Cell Rep. 30: 137–151

Angermueller C, Pärnamaa T, Parts L, Stegle O. 2016. Deep learning for computational biology. Mol. Syst. Biol. 12: 878

Aronsson H, Schöttler MA, Kelly AA, Sundqvist C, Dörmann P, Karim S, Jarvis P. 2008. Monogalactosyldiacylglycerol deficiency in Arabidopsis affects pigment composition in the prolamellar body and impairs thylakoid membrane energization and photoprotection in leaves. Plant Physiol. 148: 580–592

Bae W, Lee YJ, Kim DH, Lee J, Kim S, Sohn EJ, Hwang I. 2008. AKR2A-mediated import of chloroplast outer membrane proteins is essential for chloroplast biogenesis. Nat. Cell Biol. 10: 220–227

Baker A, Hogg TL, Warriner SL. 2016. Peroxisome protein import: a complex journey. Biochem. Soc. Trans. 44: 783–789

Bauer J, Hiltbrunner A, Weibel P, Vidi PA, Alvarez-Huerta M, Smith MD, Schnell DJ, Kessler F. 2002. Essential role of the G-domain in targeting of the protein import receptor AtToc159 to the chloroplast outer membrane. J. Cell. Biol. 159: 845–854

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340: 783–795

Bionda T, Tillmann B, Simm S, Beilstein K, Ruprecht M, Schleiff E. 2010. Chloroplast import signals: the length requirement for translocation in vitro and in vivo. J. Mol. Biol. 402: 510–523

Block MA, Dorne AJ, Joyard J, Douce R. 1983. Preparation and characterization of membrane fractions enriched in outer and inner envelope membranes from spinach chloroplasts. J. Biol. Chem. 258: 13281–13286

Bradford MM. 1976. A rapid and sensitive method for quantification of microgram quantities of protein utilizing the principle of protein-dye-binding. Anal. Biochem. 72: 248–254

Bruce BD. 2000. Chloroplast transit peptides: structure, function, and evolution. Trends Cell Biol. 10: 440–447

Bruce BD. 1998. The role of lipids in plastid protein transport. Plant Mol. Biol. 38: 223–246

Bolender N, Sickmann A, Wagner R, Meisinger C, Pfanner N. 2008. Multiple pathways for sorting mitochondrial precursor proteins. EMBO Rep. 9: 42–49

Bölter B, Soll J. 2011. Protein import into chloroplasts: dealing with the (membrane) integration problem. Chembiochem. 12: 1655–1661

Bölter B, Soll J. 2001. Ion channels in the outer membranes of chloroplasts and mitochondria: open doors or regulated gates? EMBO J. 20: 935–940

Bölter B, Soll J, Hill K, Hemmler R, Wagner R. 1999. A rectifying ATP-regulated solute channel in the chloroplastic outer envelope from pea. EMBO. J. 18: 5505–5516

Chang WL, Soll J, Bölter B. 2012. The gateway to chloroplast: re-defining the function of chloroplast receptor proteins. Biol. Chem. 393: 1263–1277

Chen K, Chen X, Schnell DJ. 2000. Initial binding of preproteins involving the Toc159 receptor can be bypassed during protein import into chloroplasts. Plant Physiol. 122: 813–822

Chotewutmontri P, Bruce BD. 2015. Non-native N-terminal Hsp70 molecular motor recognition elements in transit peptides support plastid protein translocation. J Biol. Chem. 290: 7602–7621

Chowdhary G, Kataya ARA, Lingner T, Reumann S. 2012. Non-canonical peroxisome targeting signals: identification of novel PTS1 tripeptides and characterization of enhance elements by computational permutation analysis. BMC Plant Biol. 12:142

Chung SM, Frankman EL, Tzfira T. 2005. A versatile vector system for multiple gene expression in plants. Trends Plant Sci. 10: 357–361

Cline K, Henry R. 1996. Import and routing of nucleus-encoded chloroplast proteins. Annu. Rev. Cell Dev. Biol. 12: 1–26.

Corpas FJ, Barroso JB, del Rio LA. 2001. Peroxisomes as a source of reactive oxygen species and nitric oxide signal molecules in plant cells. Trends Plant Sci. 6: 145–150

Dhanoa PK, Richardson LG, Smith MD, Gidda SK, Henderson MP, Andrews DW, Mullen RT. 2010. Distinct pathways mediate the sorting of tail-anchored proteins to the plastid outer envelope. PLoS One. 5: e10098

Dutta S, Teresinski H, Smith MD. 2014. A split-ubiquitin yeast two-hybrid screen to examine the substrate specificity of atToc159 and atToc132 two Arabidopsis chloroplast preproteins import receptors. PLoS One. 9: 95026.

Dyson HJ, Wright PE. 2005. Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell. Biol. 6: 197–208

Edwards GE, Franceschi VR, Voznesenskaya EV. 2004. Single-cell C_4 photosynthesis versus the dual-cell (Kranz) paradigm. Annu. Rev. Plant Biol. 55: 173-196

Emanuelsson O, Brunak S, von Heijne G, Nielson H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. 2: 953–971

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

Emanuelsson O, Nielsen H, von Heijne G. 1999. ChloroP a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci 8: 978–984

Encyclopædia Britannica. 2015. Chloroplast. Encyclopædia Britannica Inc.

Erlinghaeuser M, Hagenau L, Wimmer D, Offermann S. 2016. Development, subcellular positioning and selective protein accumulation in the dimorphic chloroplasts of single-cell C_4 species. Curr. Opin. Plant Biol. 31: 76-82

Fewell, SW, Brodsky JL. 2000. Entry into the endoplasmic reticulum: protein translocation folding and quality control. Landes Bioscience.

Gautier R, Douguet D, Antonny B, Drin G. 2008. HeliQuest: a web server to screen sequences with specific α-helical properties. Bioinformatics. 24: 2101–2102

Glaser E, Sjoling S, Tanudji M, Whelan J. 1998. Mitochondrial protein import in plants: signals, sorting, targeting, processing and regulation. Plant Mol. Biol. 38: 311–338

Goetze TA, Philippar K, Ilkavets I, Soll J, Wagner R. 2006. OEP37 is a new member of the chloroplast outer membrane ion channels. J. Biol. Chem. 281: 17989–17998

Goto–Yamada GS, Mano S, Yamada K, Oikawa K, Hosokawa Y, Nishimura IH, Nishimura M. 2015. Dynamics of the light-dependent transition of plant peroxisomes. Plant Cell Physiol. 56: 1264–1271

Hebert DN, Molinari M. 2007. In and out of the ER: protein folding, quality control, degradation, and related human diseases. Physiol. Rev. 87: 1377–1408

von Heijne G, Nishikawa K. 1991. Chloroplast transit peptides: The perfect random coil? FEBS Let. 278: 1–3

Hofmann NR, Theg SM. 2005. Protein- and energy-mediated targeting of chloroplast outer envelope membrane proteins. Plant J.. 44: 917–927

Hofmann NR, Theg SM. 2005. Chloroplast outer membrane protein targeting and insertion. Trends Plant Sci. 10: 450–457

Huang S, Taylor NL, Narsai R, Eubel H, Whelan J, Millar AH. 2009. Experimental analysis of the rice mitochondrial proteome, its biogenesis, and heterogeneity. Plant Physiol. 149: 719–734

Hwang YT, Pelitire SM, Henderson MP, Andrews DW, Dyer JM, Mullen RT. 2004. Novel targeting signals mediate the sorting of different isoforms of the tail-anchored membrane protein cytochrome b5 to either endoplasmic reticulum or mitochondria. Plant Cell. 16: 3002–3019

Inoue K. 2015. Emerging knowledge of the organelle outer membranes-research snapshots and an updated list of the chloroplast outer envelope proteins. Front. Plant. Sci. 6: 278

Jarvis P. 2008. Targeting of nucleus-encoded proteins to chloroplasts in plants. New Phytol. 179: 257–285.

Johnson TL, Olsen LJ. 2001. Building new models for peroxisome biogenesis. Plant Physiol. 127: 731–739

Jones DT. 1999. Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292: 195–202

Keegstra K, Cline K. 1999. Protein import and routing systems of chloroplasts. Plant Cell. 11: 557–570

Keenan RJ, Freymann DM, Stroud RM, Walter P. 2001. The signal recognition particle. Annu. Rev. Biochem. 70: 755–775

Kessler F, Schnell D. 2009. Chloroplast biogenesis: diversity and regulation of the protein import apparatus. Curr. Opin. Cell Biol. 21: 494–500.

Kessler F, Schnell DJ. 2002. A GTPase gate for protein import into chloroplasts. Nat. Struct. Biol. 9: 81–83

Kim DH, Hwang I. 2013. Direct targeting of proteins from the cytosol to organelles: the ER versus endosymbiotic organelles. Traffic 14: 613–621

Kim DH, Park MJ, Gwon GH, Silkov A, Xu ZY, Yang EC, Song S, Song K, Kim Y, Yoon HS, Honig B, Cho W, Cho Y, Hwang I. 2014. An ankyrin repeat domain of AKR2 drives

chloroplast targeting through coincident binding of two chloroplast lipids. Dev. Cell 30: 598–609

Ko K, Ko ZW. 1991. Carboxyl-terminal sequences can influence the *in vitro* import and intraorganellar targeting of chloroplast protein precursors. Biol. Chem. 267: 13910–13916

Kovacheva S, Bedard J, Wardle A, Patel R, Jarvis P. 2007. Further *in vivo* studies on the role of the molecular chaperone Hsp93 in plastid protein import. Plant J. 50: 364–379

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305: 567–580

Kubis S, Patel R, Combe J, Be'dard J, Kovacheva S, et al. 2004. Function specialization amongst the Arabidopsis Toc159 family of chloroplast protein import receptors. Plant Cell 16: 2059–2077

Kutay U, Hartmann E. Rapoport TA. 1999. A class of membrane proteins with a C-terminal anchor. Trends Cell Biol. 3: 72–75

Lanyon-Hogg T, Warriner SL, Baker A. 2010. Getting a camel through the eye of a needle: the import of folded proteins by peroxisomes. Biol. Cell. 102: 245–263

Lee DW, Jung C, Hwang I. 2013. Cytosolic events involved in chloroplast protein targeting. Biochim. Biophys. Acta. 1833: 245–252

Lee KH, Kim DH, Lee SW, Kim ZH, Hwang I. 2002. *In vivo* import experiments in protoplasts reveal the importance of the overall context but not specific amino acid residues of the transit peptide during import into chloroplasts. Mol. Cells 14: 388–397

Lee J, Lee H, Kim J, Lee S, Kim DH, Kim S, Hwang I. 2011. Both the hydrophobicity and a positively charged region flanking the C-terminal region of the transmembrane domain of signal-anchored proteins play critical roles in determining their targeting specificity to the endoplasmic reticulum or endosymbiotic organelles in Arabidopsis cells. Plant Cell 23: 1588–1607

Li HM, Chiu CC. 2010. Protein transport into chloroplasts. Annu. Rev. Plant Biol. 61: 157–180

Lopez-Juez E. 2007. Plastid biogenesis, between light and shadows. J. Exp. Bot. 58: 11–26

Lopez-Juez E, Pyke KA. 2005. Plastids unleashed: their development and their integration in plant development. Int. J. Dev. Biol. 49: 557–577

Lung SC. 2012. Characterization of the chloroplast protein import receptors in the single-cell C_4 species *Bienertia sinuspersici*. Retrieved from UWspace: University of Waterloo Library

Lung SC, Yanagisawa M, Chuong SDX. 2012. Recent development and progress in single-cell C_4 photosynthesis. Front. Biol. 7: 539-547

Lung SC, Chuong SD. 2012. A transit peptide-like sorting signal at the C terminus directs the *Bienertia sinuspersici* preproteins receptor Toc159 to the chloroplast outer membrane. Plant Cell. 24: 1560–1578

Lung SC, Smith MD, Weston JK, Gwynne W, Secord N, Chuong SDX. 2014. The C-terminus of *Bienertia sinuspersici* Toc159 contains essential elements for its targeting and anchorage to the chloroplast outer membrane. Front. Plant Sci. 5: 722

Marty NJ, Teresinski HJ, Hwang YT, Clendening EA, Gidda SK, Sliwinska E, Zhang D, Miernyk JA, Brito GC, Andrews DW, Dyer JM, Mullen RT. 2014. New insights into the targeting of a subset of tail-anchored proteins to the outer mitochondrial membrane. Front. Plant Sci. 5: 426

McKnight CJ, Briggs MS, Gierasch LM. 1989. Functional and nonfunctional LamB signal sequences can be distinguished by their biophysical properties. J. Biol. Chem. 264: 17293–17297

Mossmann D, Meisinger C, Vögtle FN. 2012. Processing of mitochondrial presequences. Biochim. Biophys. Acta. 1819(9-10): 1098–1106

Murcha MW, Kmiec B, Kubiszewki-Jakubiak S, Teixeira PF, Glaser E, Whelan J. 2014. Protein import into plant mitochondria: signals, machinery, processing, and regulation. J. Exp. Bot. 65(22): 6301–6335

Nakai K, Horton P. 1999. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem. Sci. 24: 34–36

Olsen LJ. 1998. The surprising complexity of peroxisome biogenesis. Plant Mol. Biol. 38: 163–189

Oreb M, Hofle A, Mirus O, Schleiff E. 2008. Phosphorylation regulates the assembly of chloroplast import machinery. J. Exp Bot. 59: 2309–2316

Paila YD, Richardson LGL, Schnell DJ. 2015. New insights into the mechanism of chloroplast protein import and its integration with protein quality control, organelle biogenesis and development. J. Mol. Biol. 427: 1038–1060

Patel R, Hsu SC, Bédard J, Inoue K, Jarvis P. 2008. The Omp85-related chloroplast outer envelope protein OEP80 is essential for viability in Arabidopsis. Plant Physiol. 148: 235–245

Patron NJ, Waller RF. 2007. Transit peptide diversity and divergence: a global analysis of plastid targeting signal. BioEssays 29: 1048-1058

Pinnaduwage P, Bruce BD. 1996. In vitro interaction between a chloroplast transit peptide and chloroplast outer envelope lipids is sequence-specific and lipid class-dependent. J. Biol. Chem. 271: 32907–32915

Pohlmeyer K, Soll J, Grimm R, Hill K, Wagner R. 1998. A high-conductance solute channel in the chloroplastic outer envelope from pea. Plant Cell. 10: 1207–1216

Poincelot RP. 1976. Lipid and fatty acid composition of chloroplast envelope membranes from species with differing net photosynthesis. Plant Physiol. 58: 595–598

Pool RM, Stumm J, Fulga TA, Sinning I, Dobberstein B. 2002. Distinct modes of signal recognition particle interaction with the ribosome. Science. 297: 1345–1348

Rachubinski RA, Subramani S. 1995. How proteins penetrate peroxisomes. Cell. 83: 525–528

Rasband WS. ImageJ. U.S. National Institutes of Health, Bethesda, Maryland, USA.

Rial DV, Ottado J, Ceccarelli EA. 2003. Precursors with altered affinity for Hsp70 in their transit peptides are efficiently imported into chloroplasts. J. Biol. Chem. 278: 46473–46481

Richardson LGL, Paila YD, Siman SR, Chen Y, Smith MD, Schnell DJ. 2014. Targeting and assembly of components of the TOC protein import complex at the chloroplast outer envelope membrane. Front. Plant. Sci. 5: 269.

Richardson LGL, Jelokhani-Niaraki M, Smith MD. 2009. The acidic domains of the Toc159 chloroplast preprotein receptor family are intrinsically disordered protein domains. BMC Biochem. 10: 35.

Richter S, Lamppa GK. 1998. A chloroplast processing enzyme functions as the general stromal processing peptidase. Proc. Natl. Acad. Sci. USA. 95: 7463–7468

Sanford JC, Smith FD, Russell JA. 1993. Optimizing the biolistic process for different biological applications. Methods Enzymol. 217: 482–509

Schleiff E, Soll J, Kuchler M, Kuhlbrandt W, Harrer R. 2003. Characterization of the translocon of the outer envelope of chloroplasts. J. Cell. Biol. 160: 541–551

Small I, Peeters N, Legeai F, Lurin C. 2004. Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics 4: 1581–1590

Su PH, Li HM. 2010. Stromal Hsp70 is important for protein translocation into pea and Arabidopsis chloroplasts. Plant Cell 22: 1516–1531

Smith MD. 2006. Protein import into chloroplasts: an ever-evolving story. Can. J. Bot. 84: 531–542

Smith MD, Fitzpatrick LM, Keegstra K, Schnell DJ. 2002. *In vitro* analysis of chloroplast protein import. Yamada KM (ed) Current Protocols in Cell Biology. John Wiley & Sons, New York, pp. 11.16.11–11.16.21

Smith MD, Hiltbrunner A, Kessler F, Schnell DJ. 2002. The targeting of the atToc159 preprotein receptor to the chloroplast outer membrane is mediated by its GTPase domain and is regulated by GTP. J. Cell. Biol. 159: 833–843

Smith MD, Rounds CM, Wang F, Chen K, Afitlhile M, Schnell DJ. 2004. atToc159 is a selective transit peptide receptor for the import of nucleus-encoded chloroplast proteins. J. Cell. Biol.165: 323–333.

Taiz L, Zeiger E. 2010. Plant Physiology, Fifth edition. Sinauer Associates, Sunderland, MA, USA

Taylor PD, Toseland CP, Attwood TK, Flowers DR. 2006. Beta barrel trans-membrane proteins: enhanced prediction using a Bayesian approach. Bioinformatics. 1: 231–233

Teasdale RD, Jackson MR. 1996. Signal-mediated sorting of membrane proteins between the endoplasmic reticulum and the golgi apparatus. Annu. Rev. Cell Dev. Biol. 12:27–54

Thomas H, Huang L, Young M, Ougham H. 2009. Evolution of plant senescence. BMC Evol. Biol. 9: 163

Tranel PJ, Froehlich J, Goyal A, Keegstra K. 1995. A component of the chloroplastic protein import apparatus is targeted to the chloroplast outer envelope membrane via a novel pathway. EMBO J. 14: 2436–2446

Tranel PJ, Keegstra K. 1996. A novel, bipartite transit peptide targets OEP75 to the outer membrane of the chloroplastic envelope. Plant Cell 8: 2093–2104

von Heijne G, Steppuhn J, Herman SG. 1989. Domain structure of mitochondrial and

chloroplast targeting peptides. Eur. J. Biochem. 180: 535-545

Wienk HLJ, Czisch M, de Kruijff B. 1999. The structural flexibility of the ferredoxin transit peptide. FEBS Lett. 453: 318–326

Wise R. 2006. The diversity of plastid form and function. Springer. 23: 3–26

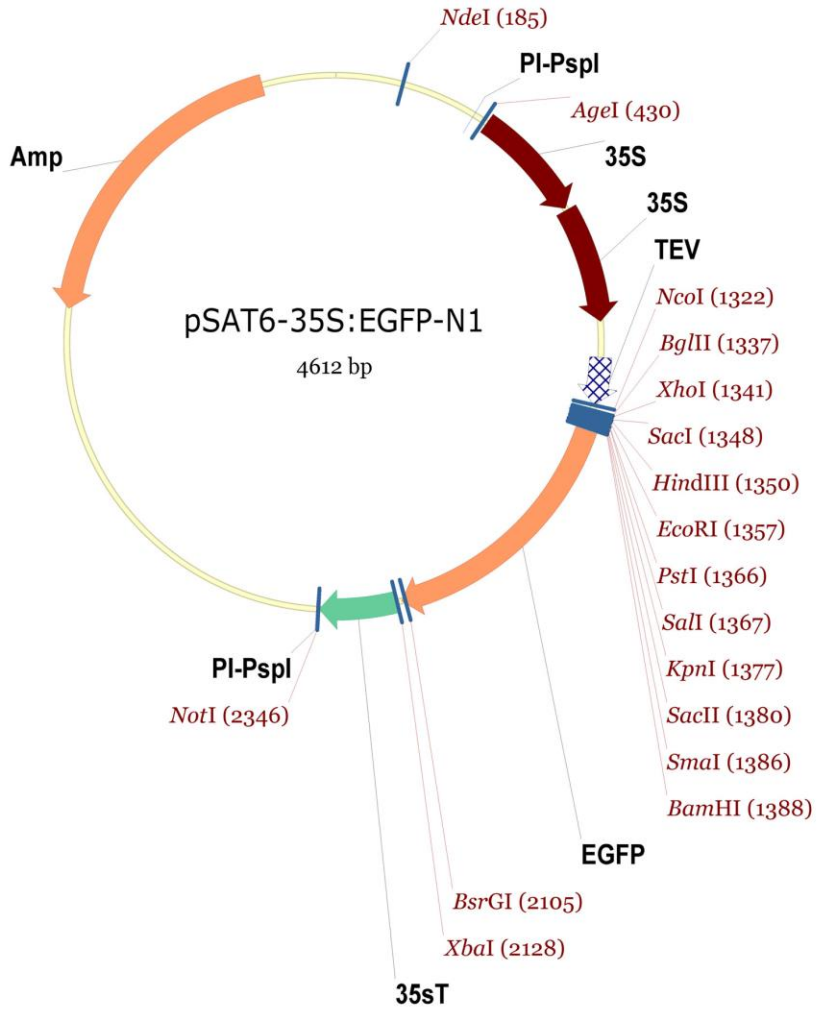
Yan J, Campbell JH, Glich BR, Smith MD, Liang Y. 2014. Molecular characterization and expression analysis of chloroplast protein import components in tomato *Solanum lycopersicum*. PLoS One 9: e95088

Yoo SD, Cho YH, Sheen J. 2007. *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. Nat. Protoc. 2: 1565–1572

Zhang XP and Glaser E. 2002. Interaction of plant mitochondrial and chloroplast signal peptides with the Hsp70 molecular chaperone. Trends Plant Sci. 7: 14–21

7. Appendix

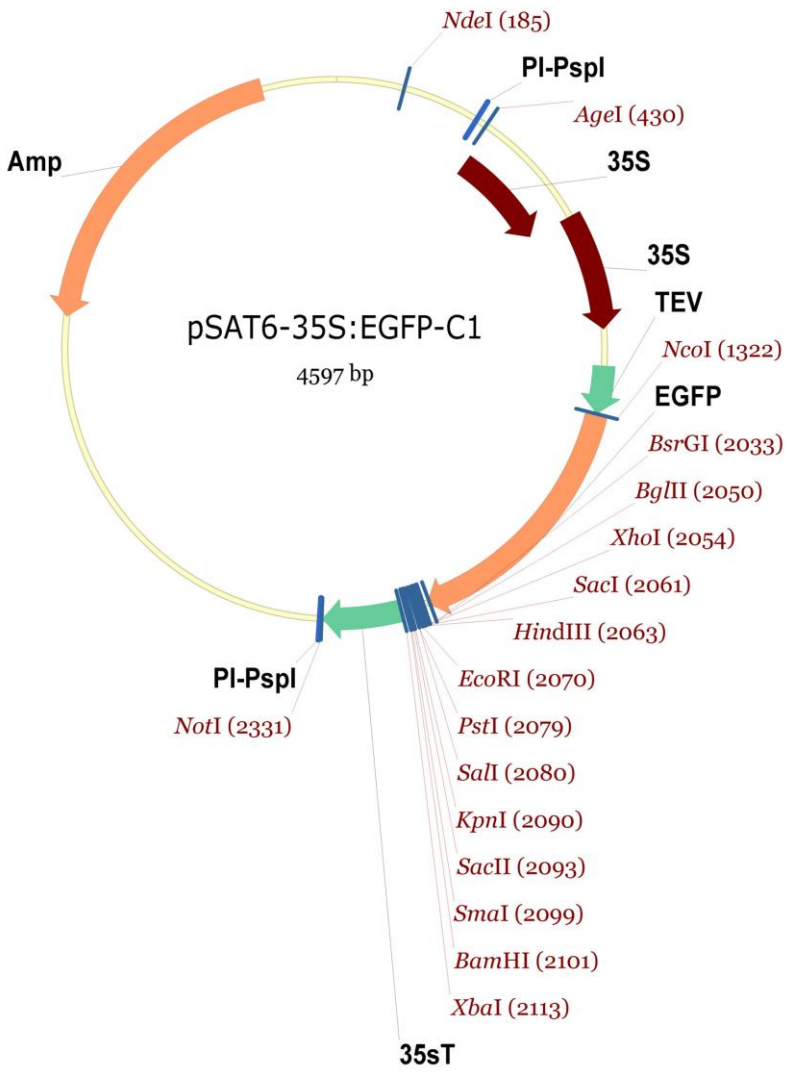
Appendix I – Vector Maps



pSAT6-35S:EGFP-N1 sequence landmarks:

- Cauliflower mosaic virus 35S promoters (35S)	441 - 1,089
- Translational enhancer from tobacco etch virus (TEV)	1,190 - 1,320
- Enhanced green fluorescent protein (EGFP)	1,395 - 2,114
- Cauliflower mosaic virus 35S terminator (35sT)	2,133 – 2,343
- Ampicillin selection marker (Amp)	3,552 – 4,412

Lung SC. 2012. Characterization of the chloroplast protein import receptors in the single-cell C_4 species *Bienertia sinuspersici*. Retrieved from UWspace: University of Waterloo Library



pSAT6-35S:EGFP-C1 sequence landmarks:

- Cauliflower mosaic virus 35S promoters (35S)	441 - 1,089
- Translational enhancer from tobacco etch virus (TEV)	1,190 - 1,320
- Enhanced green fluorescent protein (EGFP)	1,323 - 2,039
- Cauliflower mosaic virus 35S terminator (35sT)	2,118 – 2,328
- Ampicillin selection marker (Amp)	3,537 – 4,397

Lung SC. 2012. Characterization of the chloroplast protein import receptors in the single-cell C_4 species *Bienertia sinuspersici*. Retrieved from UWspace: University of Waterloo Library