Investigating the Outer Envelope Localization of OEP 18 in *A. thaliana* Chloroplasts

by

Tianlun Zhou

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Science
in
Biology

Waterloo, Ontario, Canada, 2021

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Plastids are plant organelles with specialized functions including photosynthesis. The specialized function of each plastid is informed by its distinct and dynamically regulated proteome. The majority of plastid proteins are synthesized in the cytosol and are imported into the plastid post-translationally. A variety of receptors and channels embedded within the plastid outer and inner envelope membranes regulate the import of plastid proteins, and thus control the plastid proteome composition. There are 117 known Outer Envelope Proteins (OEPs) that reside in the chloroplast outer membrane. Of these 117 OEPs, only a few have been extensively studied regarding their targeting. With the exception of Toc75, which uses the general pathway involving an N-terminal cleavable transit peptide (TP) for targeting, the OEPs that have been studied so far either use the signal-anchored, tail-anchored, or β-barrel pathway for targeting. However, previous studies showed that a key component of the translocon at the outer envelope of chloroplasts (Toc complex), Toc159, possesses a novel reverse TP-like sorting signal and an unconventional membrane anchor within the C-terminus for its own targeting and insertion into the outer membrane. Moreover, recent studies using bioinformatics tools predicted that eight additional OEPs also contain transit peptide-like sequences at their C-termini. Transient expression assays in *Arabidopsis* mesophyll protoplasts demonstrated that one of the candidates, OEP18, appears to be targeted to the chloroplast outer membrane using such a signal. My research was able to determine the precise subcellular location of OEP18 using protoplast transient expression assays, chloroplast fractionation, and Western blot analysis. Secondary structure and membrane topology prediction analyses for all 117 OEPs were also used for grouping according to their targeting pathways. My research also determined that the predicted reverse TP region of OEP18 is responsible for specific targeting to the chloroplast outer

membrane, but not its anchoring. These findings will allow for a better understanding of protein targeting to the outer membrane of chloroplasts.

Acknowledgement

I would like to thank Professor Chuong and Professor Smith for supervising and guiding this research project. I would also like to thank my committee members Professor Moffatt, Professor Marsden for guiding this project. Lastly, I would like to thank my colleagues Ben, Delaney, Nilan and Alyssa for their technical support and inspiration during my Master's degree.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

aa	Amino Acid
AKR2A	Ankyrin Repeat Protein 2A
CD	Circular Dichroism
CT	C-terminal
CT TP-like	C-terminal Transit Peptide-Like
IEM	Inner Envelope
IMS	Intermembrane space
NT	N-terminal
OEM	Outer Envelope
OEP	Outer Envelope Protein
RbcL	RuBisCO Large Subunit
RbcS	RuBisCO Small Subunit
SA	Signal Anchored
T159L	Toc159-like
TA	Tail-Anchored
TIC	Translocon of the Inner Membrane of Chloroplasts
TOC	Translocon of the Outer Membrane of Chloroplasts

Toc159	Translocase of chloroplast 159
TP	Transit Peptide

List of symbols

~	approximately
Δ	without

1. Introduction

1.1 The evolution and types of plastids

Plastids are double-membrane organelles of plant cells and are essential for plant metabolism and survival (Lopez-Juez, 2005). They originated from the endosymbiosis that occurred after a photosynthetic cyanobacterium was engulfed by an early eukaryote millions years ago (Bolter, 2018). The evolution of the chloroplast from the symbiotic cyanobacterium, a process otherwise known as organellogenesis, was the origin of species for Earth's land plants, green algae, red algae, and glaucophytes (Day & Theg, 2018; Patron & Waller, 2007). Through an endosymbiotic relationship, an engulfed bacterium contributes to the host's survival, and its majority of genes were transferred horizontally to the host cell's genome during evolution (Martin et al, 2002).

Organellogenesis of the symbiont required three essential steps (Bölter, 2018). First, the cyanobacterium's genetic material is transferred laterally to the host genome. The transfer of a few genes can be the trigger for the following rapid lateral transfer of major genetic materials (Lee & Hwang, 2018). Next, the host cell evolved methods to transcribe and translate the laterally transferred genes (Bölter, 2018). Finally, the host transfers the proteins encoded by those genes back to the symbiotic organelle, which is vital for completion of organellogenesis. In this symbiotic relationship, the host cell regulates plastid protein expression and import, increasing overall efficiency and productivity (Day & Theg, 2018).

Several types of plastids, are structurally and functionally distinct from the least differentiated precursor form, known as proplastids, with each type possessing a unique function. Although the different types of plastids serve different purposes and thus have a different proteomes and membrane structures, they are interchangeable in their response to different types of environmental stimuli and developmental stages, as shown in Figure 1 (Jarvis et al, 2013). For

example, amyloplasts are responsible for starch synthesis and storage; elaioplasts are responsible for lipid synthesis and storage, and also play an important role in pollen maturation; chromoplasts contain a variety of carotenoids that are thought to attract seed-dispersing vectors, such as animals and insects, by captivating them with bright colours; etioplasts remain arrested in the transition between proplastids and chloroplasts in the absence of light in some tissues, but serve as an essential component for metabolism in cabbage leaves (Wise, 2006). Gerontoplasts develop from chloroplasts during leaf senescence simultaneously with internal organelles such as thylakoid membranes and grana being disintegrated (Biswal and Mukesh, 2003). Proteinoplasts, also called proteoplasts develop specifically for protein storage are found in various seeds (Wise, 2006).

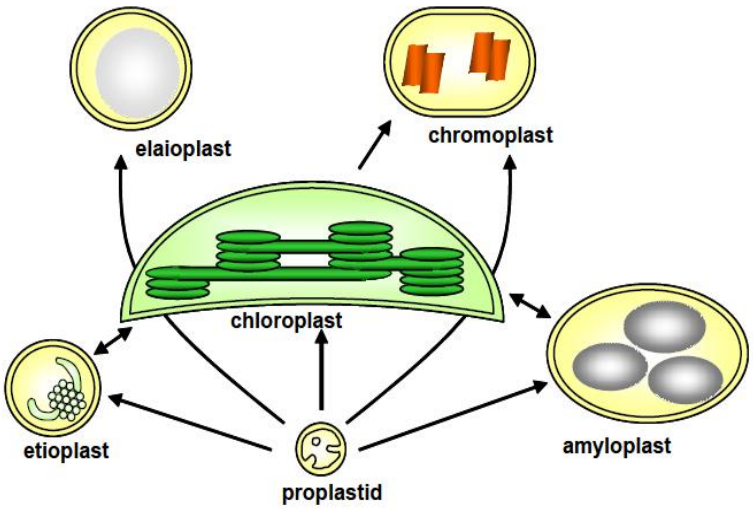

Figure 1.1 Diversity of plastid types and their interconversions (Lopez-Juez, 2005)

1.2 Chloroplast functions and structures

By far the most well-known type of plastid is the chloroplast, which is responsible for photosynthesis. Although all plastids are derived from proplastids, chloroplasts remain the dominant form among other plastids due to their vital role in plant survival. In addition to photosynthesis, chloroplasts also play a key role in photorespiration, chlorophyll synthesis, nitrogen assimilation, as well as lipid, fatty acid and amino acid synthesis (Wise, 2006). Each plastid type is interconvertible under different environmental stimuli, and in the presence of light chloroplasts can be derived from the differentiation of other plastids, such as etioplasts and amyloplasts (Figure 1; Thomas et al, 2009). Contrarily, this differentiation process can be reversed through a prolonged deprivation of light, allowing the chloroplast to revert back into an etioplast to carry out different functions (Thomas et al, 2009).

Chloroplasts are responsible for providing their host with carbohydrates, amino acids, specialized metabolites, lipids, and hormones. It also produces oxygen as a by-product from photosynthesis. The chloroplast is involved in ROS production and ion homeostasis, maintaining electron transport chains, and photosynthesis. The host cell in return protects the organelle from biotic and abiotic factors and maintains protein synthesis, regulation, and transport (Bölter, 2018; Lee & Hwang, 2018). These diverse functions are supported by the chloroplast proteome with approximately 3000 proteins. The majority of chloroplast proteins are encoded by nuclear genes, synthesized in the cytosol, and targeted post-translationally to the correct chloroplast subcellular compartment (Thomson et al, 2020).

The structure of chloroplasts consists of three membrane systems (the outer and inner and thylakoid membranes), and three internal subcompartments (the thylakoid lumen, stroma, and intermembrane space) as shown in Figure 1.2. The outer membrane or OEM contains several β-

barrel proteins with a substrate specificity similar to that of bacterial porins, hence its permeability allows the translocation of cytosolic preproteins (Bolter and Soll, 2001). The inner membrane or IEM serves as a physical barrier between the cell cytosol and the chloroplast's interior. Although both envelope membranes are composed largely of a lipid bilayer, they contain a high proportion of galactosyl diacylglycerides rather than the typical phospholipids, thereby suggesting a difference in the bilayer's overall function compared to the cellular membrane (Poincelot, 1976). The light-dependent reactions of photosynthesis occur within the thylakoid membrane (Jarvis, 2008). The thylakoid lumen, encased by thylakoid membrane, is the site of subsequent electron transfer chains during photosynthesis (Järvi, 2013). The stroma serves as a sub-organellar soluble space where the light-independent Calvin-Benson Cycle takes place (Jarvis, 2008). Lastly, the inter-membrane space or IMS serves as a buffer region of protein transport between the outer and inner membrane, and is where specific imported preproteins undergo redox reactions aided by the translocon complex located in the inner membrane (Kessler and Schnell, 2006). Both outer and inner membranes have a protein complex, Toc and Tic complex respectively, to allow the import of cytosolically translated proteins that are destined for the plastid (Schnell, 2019). Both development of the import pathways and chloroplast targeting signals of of host cell proteins are essential for proper import of chloroplast proteins, and thus maintaining the normal function of chloroplast (Day & Theg, 2018).

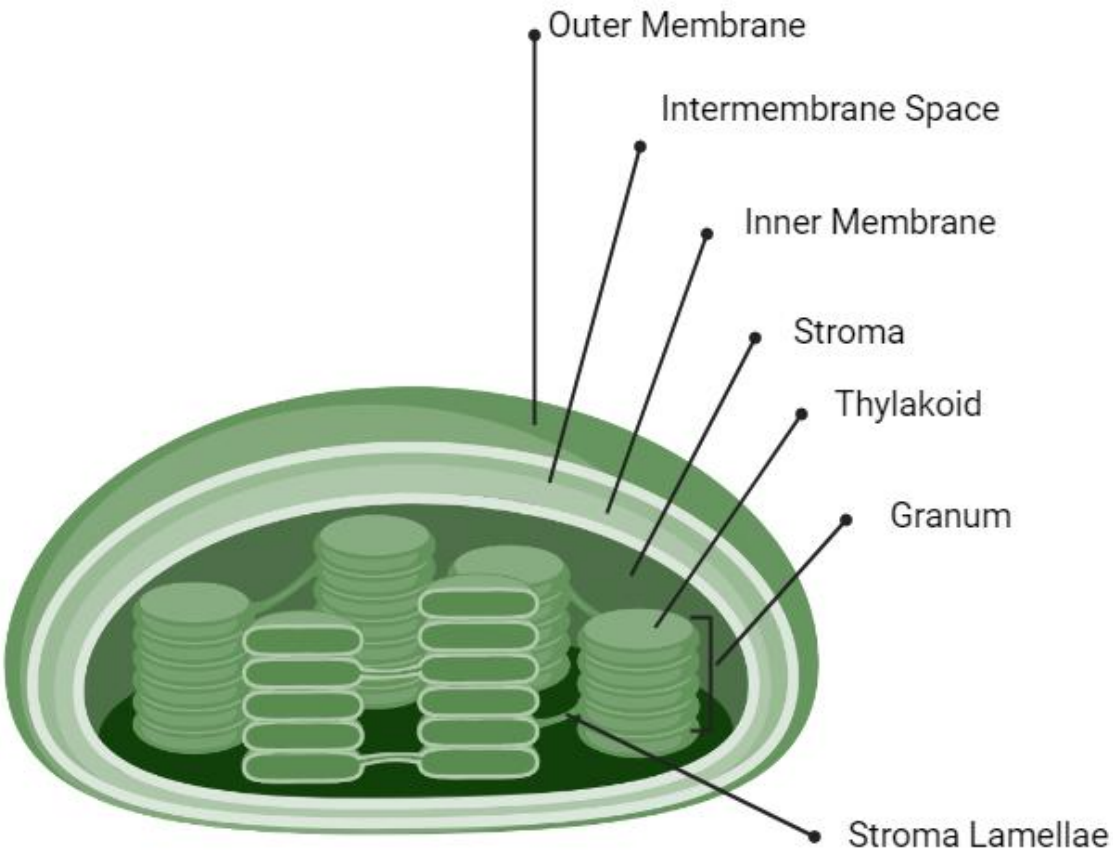

Figure 1.2 Structure of a chloroplast. The chloroplast is covered by the double membrane: the outer membrane, the inner membrane and the intermembrane space in-between. Thylakoids are membrane-bound compartments inside chloroplast, they are responsible for light-dependent reactions of photosynthesis. Stacks of thylakoids form granum, which are connected by stroma lamellae in-between. Created with BioRender.com

1.3 Chloroplast protein import

1.3.1 Transit Peptides

Due to lateral gene transfer, ~95% of plastid proteins are nuclear-encoded and cytosolically translated (Bölter, 2018; Lee & Hwang, 2018). Therefore, the nucleus-encoded proteins required for chloroplast biogenesis and maintenance of identity are translated in the cytosol as precursor proteins, targeted to and imported into the organelle (Jarvis, 2008). These proteins are

synthesized as precursor proteins (preproteins) in the cytosol and contain an N-terminal (NT) peptide extension, called a transit peptide (TP), that functions as a chloroplast localization-signal (Lee et al, 2013; Patron & Waller, 2007; Schnell, 2019). Similar peptide extensions used for protein import are seen widely among the endoplasmic reticulum, mitochondria, peroxisomes and chloroplasts (Bruce, 2000). This N-terminal TP is necessary and sufficient for import of preprotein from cytosol to chloroplast, yet different TP-containing preproteins can be destined to either outer or inner membranes instead of stroma during translocation through import channels, suggesting the TP not only grants import but also direct the destination of specific preproteins (Lee et al, 2017; Okawa et al, 2014).

Targeting peptides used in different organelles such as mitochondria are defined by a consensus sequence which can be easily predicted (Lee et al, 2014); however, chloroplast TPs have highly divergent sequences, making it difficult to predict them using the protein sequence alone (Lee & Hwang, 2018; Patron & Waller, 2007). TPs are highly variable in their length, amino acid composition and sequence, it is difficult to determine any consensus sequences for them (Chotewutmontri et al, 2012). Nevertheless, there are some common features in their primary sequences, as they generally contain a high concentration of serine (S) and threonine (T) as well as a low presence of acidic amino acids thus creating a net positive charge (Zhang and Glaser, 2002). Phosphorylation and dephosphorylation of S and T have been reported to regulate and increase the efficiency of import (Sjuts et al, 2017). During preprotein import, S/T residues are phosphorylated by STY kinases STY7, STY18 & STY46. Then, protein 14-3-3 binds phosphorylated S/T residues which increases guidance complex binding efficiency (Sjuts et al, 2017). Dephosphorylation of TPs at the OEM is required to maintain efficient import through the TOC complex. Failure to dephosphorylate preproteins results in slow import. This

phosphorylation cycle is not vital for import process; however, it is able to increase import efficiency in some cases (Bölter, 2018). This phosphorylation cycle can also be rapidly modified to adjust preprotein import specificity, allowing the cell to rapidly adjust the chloroplast proteome in response to environmental stressors (Sjuts et al, 2017). Presence of numerous prolines is also a common feature of chloroplast TPs as revealed by amino acid substitution and mutation studies (Lee et al, 2018). Interestingly, a mutated preprotein containing a Proline-less TP can be mis-targeted to chloroplast envelope membranes or cause protein aggregation in the stroma during chloroplast import. Additionally, proline residues may interact with stromal motor import proteins during late stages of import and increase the efficiency of import (Lee & Hwang, 2018). Amphiphilic α-helices are also commonly detected or predicted within TPs, as they form stable membrane-associations under mimetic conditions, yet relax into stable helices in hydrophilic environments (Lee et al, 2008), this becomes an important feature for TPs prediction.

Despite the high variance in the primary structure of TPs, some conserved motifs have been identified. Past research has shown that the numerous motifs within different TPs interact with different parts of the translocon complex during multiple import steps (Okawa et al, 2014), and an uncharged region of the TP and its hydrophobicity near the N-terminus is essential for the import of proteins into chloroplasts (Lee et al, 2002). TP sequences with moderate hydrophobicity can interact with the heat shock protein (Hsp) 70; as such, hydrophobicity enables efficient targeting to the OEM by the guidance complex (Lee & Hwang, 2018). Moreover, the hydrophobic segment shared among 77% of all transit peptides has been shown to be an Hsp70 binding site, in which Hsp70 acts as translocation motor and Hsp93 is likely involved in the initiation of preprotein import (Chotewutmontri and Bruce, 2015). Chaperone proteins in the cytosol recognize and bind particular TP motifs then carry preproteins to a

specific receptor in the outer membrane. Chaperone proteins belong to the Hsp 70 family and Hsp90 family, as well as protein 14-3-3, have been found to important for the transport of many preproteins to OEM-receptors. Hsp70 and 14-3-3 form a guidance complex which predominately targets preproteins to receptors in the TOC (translocon at the outer envelope of chloroplasts) complex (Bölter, 2018). Further studies also indicated that numerous motifs in different TPs can work as independent functional units, as unrelated proteins with synthetic TPs containing several critical motifs from the RbcS TP can be delivered to chloroplasts with moderate efficiency compared to original TP (Lee et al, 2015).

1.3.2 The general plastid import pathway

The import pathway to chloroplasts known as the "general import pathway" used by most proteins containing transit peptides is shown in Figure 1.3 (Kessler and Schnell, 2006). This pathway is mediated by a translocon at the outer envelope membrane of chloroplasts, known as Toc, and a translocon at the inner envelope membrane of chloroplasts, known as Tic (Jarvis, 2008). The TOC and TIC (translocon at the inner envelope of chloroplasts) complexes are physically linked by Tic236 and function together as a super-complex that transports proteins across the chloroplast double membrane (Chen et al, 2018; Schnell, 2019). This super complex assembly creates a membrane contact site which allows TPs to simultaneously interact with both the TOC and TIC complexes (Chen et al, 2018)

The TOC complex is assembled from GTPase receptors and a β-barrel channel. Toc159 and Toc34 family members function as the TOC complex protein receptors. These proteins are anchored to the OEM by C-terminal membrane domains, have a cytosolic GTPase domain, and cytosolic TP recognition sites (Schnell, 2019). They assemble with the β-barrel protein Toc75-III which is a protein-import channel and the core component of the TOC complex. Toc159 and

Toc34 mediate the initial interaction of the preprotein with the TOC complex in a selective and reversible manner; this interaction functions as check-point before preprotein import ensues (Schnell, 2019). During this initial reversible interaction, the disordered region of Toc159, called the acidic domain, binds the preprotein and the TP is partially inserted across the OEM. The mid-region of the TP interacts with Toc75-III and the N-terminus interacts with Tic20, which is a core component of the TIC complex at the inner membrane (Schnell, 2019). It was established that preproteins remain unfolded in the cytosol before import, thus allowing passage through the protein-import channels of both the OEM and IEM and are folded in the stroma after translocation is complete. However recent research shows that a 22 kDa preprotein is able to import through membranes in a folded conformation and the pore size of the channels are much larger than expected (Ganesan, 2018). Upon entering the intermembrane space, preproteins will bind to Hsp70 and enter the TIC complex with help from Tic22 (Kessler and Schnell, 2006). When a preprotein is selected for import, Toc receptors hydrolyze bound GTP to GDP and the energy released facilitates preprotein association with the stromal-chaperone import complex (Schnell, 2019). The stromal-chaperone import complex is tethered to the TIC complex by Tic110 and provides most of the energy required to facilitate movement through the TOC/TIC complex via ATP hydrolysis (Sjuts et al, 2017; Schnell, 2019). The complex is composed of import motor proteins cpHsp70, Hsp90c, and Hsp93 which pull the preprotein through the translocon super complex (Lee & Hwang, 2018).

Currently, TIC complex composition and the import motor complex remains a controversial topic, as new evidence on the functions of Tic100, Tic56 and Tic214 has come to light and challenges the traditional model (Nakai, 2015; Li et al, 2020), and it is also controversial that the core component of the TIC complex is Tic20 or Tic110 (Bölter, 2018). Thus, the classical model

will still be used to illustrate the pathway here. The transmembrane channel located in the inner membrane is composed of Tic20 and Tic110, in which Tic110 possesses a transit peptide-binding domain in its stromal region that acts as a docking site for preproteins, allowing them to bind to recruited Hsp93 (Kessler and Schnell, 2006). Additionally, Tic110 recruits Cpn60 which is involved in facilitating preprotein folding and mature protein processing. Cpn60 also recruits the enzyme stromal processing peptidase (SPP) that is used to cleave off the transit peptide in the stroma (Kessler and Schnell, 2006). The mature protein can assemble in the stroma or be sorted to the IEM, the TM, or the thylakoid through suborganellar targeting pathways (Sjuts et al, 2017). Unfolded preproteins are prone to forming toxic aggregates and must be degraded by proteosomes. Unfolded or misfolded protein with specific motifs are targeted for degradation by protein factors such as AtBAG1, Hsc70-4, and CHIP (Lee & Hwang, 2018; Schnell 2019; Thomson et al, 2020). These factors use a variety of mechanisms to target preproteins for degradation. For example, Hsc70-4 and CHIP form a complex that targets unfolded photosynthetic preproteins for degradation to prevent premature chloroplast biogenesis in etioplasts (Schnell, 2019).

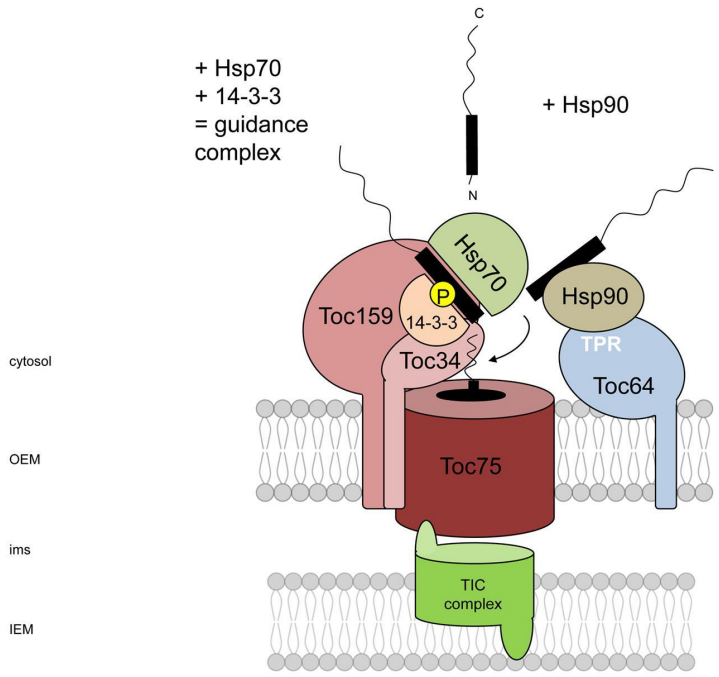

Figure 1.3 Chloroplast protein import pathway used by transit peptide-containing preproteins (Sjuts, et al 2017): Chloroplast protein import to the stroma is facilitated by the Toc complex in the outer membrane in conjunction with the Tic complex in the inner membrane. Briefly, the guidance complex formed by Hsp70 and 14-3-3 on Toc complex binds to specific regions of pre-proteins. the transit peptide on preproteins is recognized by GTPase receptors Toc159 and Toc34 and transported through the outer membrane by Toc75, a transmembrane channel. Once emerging in the stroma through the Tic complex, the transit peptide on preprotein is cleaved off by SPP, producing the mature protein that can be folded or targeted to a sub-chloroplast compartment.

1.4 Targeting of Chloroplast Outer Envelope Proteins

Outer Envelope Proteins (OEPs) are a family of proteins specifically localized to the outer membrane of chloroplasts. There are currently 117 known or predicted OEPs, although many of their functions and import machineries remain unknown (Inoue, 2015). Some of their known functions include ion/substrate transport, protein import/turnover, carbohydrate or lipid metabolism and intracellular signal transduction (Inoue, 2015). As their functions suggest, they play important roles in plastid identity and organelle metabolism.

Until recently, TOC75 was the only known OEP to be targeted to the outer membrane using the TP-mediated general import pathway (Tranel et al, 1995). Recent research also indicates that OEP80, a paralog of Toc75, also enters the chloroplast using a transit peptide but with a different sorting signal (Day et al, 2019). With the exception of Toc75 and OEP80, all preproteins with TPs are directed to the stroma (Gross et al, 2020). Thus, the vast majority of OEPs use pathways other than the general import pathway for their targeting. All OEPs are transcribed in the nucleus and translated on cytosolic 80S ribosomes (Kim et al, 2019). Some translated OEPs are bound by cytosolic factors that assist in OEP proteostasis and OEM-targeting (Kim et al, 2015). Cytosolic factors create a physico-chemical environment that maintains their import or targeting-competent state. Cytosolic factors also bind to hydrophobic regions of OEPs that interact unfavorably with the cytosol and prevent protein aggregation (Kim et al, 2019). Cytosolic factors can also aid in protein transport to the OEM. OEPs are thought to be bound and transported by cytosolic factors via diverse localization-signals within the OEP sequence and/or secondary structures (Lee et al, 2017). The localization mechanisms used by many OEPs are mostly unknown, due to limited knowledge of OEPs and the difficulties with transmembrane protein analysis. However newer research in proteomics and protein analysis techniques has led to more discovery of OEPs and detailed categorization of OEPs sequences (Bouchnak et al, 2019; Inoue et al, 2015).

OEPs can be divided into two structurally different groups: β-barrel proteins and α-helical transmembrane domain (TMD) proteins. Helical TMD proteins can be classified into four distinct structural subgroups. In two subgroups, the TMD is a single α-helix located at the N-terminus or C-terminus, named signal-anchored (SA) and tail-anchored (TA) proteins, respectively. A third subgroup includes proteins with multiple α-helical TMDs. Lastly, the TMD of proteins in the fourth subgroup contains both α-helices and β-sheets, and are named CT TP-

like proteins (Lee et al, 2014; Lung et al, 2014). Currently, there are four well established OEP import mechanisms: the N-terminal TP-mediated general import used by Toc75, β-barrel self-insertion, signal-anchor mediated insertion, and tail-anchor mediated insertion (Kim et al, 2019; Lee et al, 2017).

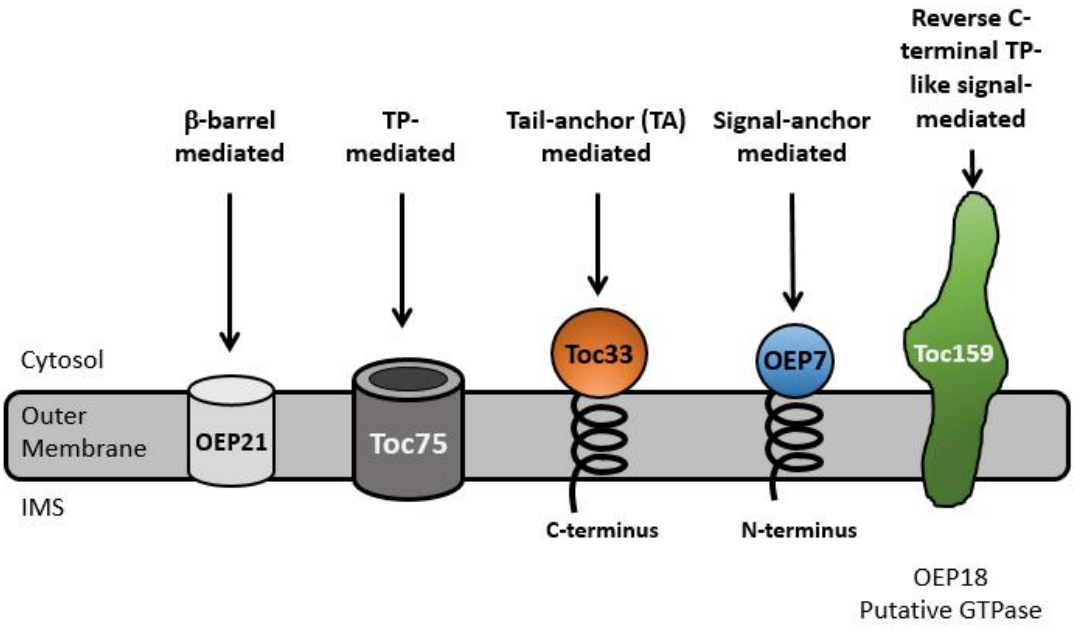

Figure 1.4 Four canonical chloroplast OEP targeting pathways and the novel Toc159-like pathway: Though the mechanisms for the targeting of most OEPs to the chloroplast outer membrane have not been defined, illustrated here are five distinct chloroplast outer membrane targeting mechanisms. OEP21 is an example of a β-barrel protein that appears to self-insert into the chloroplast outer membrane. Toc33 and OEP7 are representative proteins of types of targeting used by other known OEPs; Toc33 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 it is different from the tail-anchored proteins because it lacks a canonical α-helical transmembrane domain (TMD) (Lung et al, 2014).

1.4.1 β-barrel OEPs

β-barrel proteins are formed from 8-24 β-sheets which create a hydrophilic membrane pore (Tsaousis et al, 2017). The majority are transmembrane channels that recognize and transport specific substrates. β-barrel proteins are also responsible for cellular signaling, organelle interactions, apoptosis, and other cellular pathways (Jones & Rapaport, 2017). All of these channels share evolutionary ancestry and are exclusively found in the envelopes of chloroplasts & mitochondria and in the plasma membrane (PM) of gram-negative bacteria. Homology between β-barrel proteins across different organelles was proven by their highly similar secondary structures, function, and localization-signals (Jones & Rapaport, 2017). β-barrel proteins found in the plasma membrane of gram-negative bacteria and chloroplast outer and inner membranes can target the mitochondrial outer membrane (MOM) *in vivo* (Jones & Rapaport, 2017). Furthermore, mitochondrial β-barrel proteins can target the gram-negative bacteria PM, suggesting these channels have common targeting specificity (Jones & Rapaport, 2017). However mitochondrial β-barrel proteins cannot target the chloroplast OEM, suggesting chloroplast β-barrel OEPs have gained additional mechanisms that enable specific chloroplast localization and prevent localization to the mitochondria (Jones & Rapaport, 2017). It is possible that cytosolic factors assist in selective β-barrel localization to the OEM, but it is speculation (Jones & Rapaport, 2017). It is possible that conserved OEM targeting signals are embedded within their highly similar secondary structure instead of diverse primarily structure. Further evidence of a 20 hydrophobic C-terminal β-hairpin shown to be necessary and sufficient for β-barrel targeting to the mitochondria OM demonstrates the targeting-signal for OM β-barrel proteins lie within the secondary structure (Jones & Rapaport, 2017). Furthermore, fusing this hydrophobic b-hairpin to the chloroplast β-barrel proteins OEP37 and OEP24, resulted in

mistargeting to the MOM. Chloroplast β-barrel proteins also have a C-terminal b-hairpin motif that was proven not sufficient for import (Jones & Rapaport, 2017). Thus, the specific import mechanism used by chloroplast β-barrel protein appears to be using an altered mechanism and requires further study (Jones & Rapaport, 2017). Chloroplast β-barrel proteins are imported to the OEM post-translationally. A recent study suggests some chloroplast β-barrel proteins can facilitate their own import into the OEM *in vitro* (Gross et al, 2020). However, a small group of chloroplast β-barrel proteins were predicted to use N-terminal TPs for OEM import, including: Toc75-III, OEP24, OEP37, and OEP80/Toc75-V. OEP24 and OEP37 do not exhibit a change in size after import, suggesting an N-terminal TP signal was not utilized (Jones & Rapaport, 2017). Further investigation is required to determine the import mechanism of OEP24 and OEP37 (Jones & Rapaport, 2017; Kim et al, 2019). Toc75-III uses a bipartite TP 21 with two distinct elements, a classical N-terminal TP and a glycine rich (GR) region (Kim et al, 2019). OEP80/Toc75-V is found to be also using an N-terminal signal-peptide. However, these Toc75 signals are highly dissimilar and are thought to utilize distinct import pathways (Day et al, 2019; Gross et al, 2020). The classical TP found within the bipartite TP of Toc75-III utilizes some of the general import pathway apparatus. During import, the TP is pulled into the stroma which pull the GR region into the IMS where it is attached. Then SPP cleaves the TP in the stroma, and the GR region is cleaved by type I signal-peptidase in the IMS. Thus, this differentiates Toc75-III from the general import pathway and self-inserts in the OEM (Richardson et al, 2014). The import of Toc75-III can be competitively inhibited by preproteins. Moreover, OEM-receptors which recognize preprotein-chaperone complexes, Toc64, OEP61, and Toc33, have been implicated in Toc75-III import. This also supports the theory that Toc75-III binds cytosolic factors and uses some components of the general import pathway (Jones & Rapaport, 2017).

However, it is unknown if chaperone proteins are able to recruit Toc75-III to the general import apparatus (Kim et al, 2019). Recent research shows that Toc75-V forms a translocon which can insert Toc75-III and other β-barrel proteins into the OEM (Gross et al, 2020). Recently, a TP-signal was identified at the N-terminus of Toc75-V, however it is not bipartite and may not utilize the general import pathway. When Toc75-V was first analyzed for a probable TP, N-terminal peptide does not qualify, resulting in the dismissal of an N-terminal TP-signal (Gross et al, 2020). However, more recent studies provide evidence that Toc75-V uses an N-terminal TP which is necessary for targeting (Gross et al, 2020). During import, the Toc75-V TP is cleaved at a conserved cysteine residue followed by a consensus sequence. Unlike the processing of Toc75-III, the Toc75-V TP is cleaved after import is finished, suggesting they are diverted to the OEM by different import pathways. Thus, Toc75-III and Toc75-V use different cleavable N-terminal TP-signals and also OEM import-pathways (Day et al, 2019; Gross et al, 2020)

1.4.2 Signal-Anchored (SA) OEPs

Many SA proteins, such as Toc64 and OEP14 serve as membrane receptors (Lee et al, 2014). SA OEPs have a topology of the N-terminus embedded in the membrane and the C-terminus extending into the cytosol. SA-proteins specifically target the chloroplast OEM using a non-cleavable TP-signal which includes, an α-helical TMD anchor, and a C-terminal positively-charged flanking region (CPR). In 85% of chloroplast SA-proteins, the TMD-anchor has a hydrophobicity score of less than 0.4 on the Wimley White (WW) scale (Lee et al, 2011). This feature is distinct from the pathway of the ER localization as the TMD in 89% of ER SA-proteins have a hydrophobicity score of greater than 0.4 on the WW scale (Lee et al, 2011). The CPR consists of 3-5 lysine (K) and/or arginine (R) residues. Exchanging basic residues for inert

glycine residues causes Toc64 mistargeting to the plasma membrane. Thus, the hydrophobicity of the TMD and charge of the CPR are vital to maintain specific OEM targeting (Kim et al, 2019; Lee et al, 2014). SA-proteins require a cytosolic factor, ankyrin repeat protein 2 (AKR2) for recruitment and insertion into the OEM. AKR2 translationally targets SA-proteins to the OEM by binding the SA-protein N-terminus (Kim et al, 2019). AKR2 binds the hydrophobic regions of the TMD to prevent unfavourable interactions with the cytosol and the formation of protein aggregation. After AKR2 binds the SA-proteins, sHsp17.8 binds AKR2 and transfer them to the chloroplast (Kim et al, 2011). Then AKR2 uses an MGDG (monogalactosyldiacylglycerol) lipid and a PG (phosphatidylglycerol) lipid as a docking site to unload its cargo. Once AKR2 docked, SA-proteins are integrated into the OEM by Toc75, however, the exact mechanism requires further evidence (Kim et al, 2019).

1.4.3 Tail-Anchored (TA) OEPs

TA proteins maintain diverse and important functions in multiple membranes, including protein translocation, membrane fusion, vesicle-trafficking, electron transport, apoptosis, and protein quality control (Lee et al, 2014). TA OEPs have an opposite topology of SA OEPs as their C-terminus is inserted into the membrane leaving the N-terminus exposed to the cytosol. They have been shown to contain three sequentially ordered characteristic features, including: a positive C-terminal sequence (CTS), an α-helical TMD membrane-anchor, and a CT-tail with a maximum length of 50 aa (Zhuang et al, 2017).

The targeting signals of TA-proteins contain four physico-chemical features, each feature has a varying degree of importance depending on the identity of the TA-protein and the context provided by each feature (Kim et al, 2019). The four features of TA-protein targeting signals

include, a positive CTS, an α-helical TMD-anchor, moderate hydrophobicity within the TMD, and in some cases, a GTPase domain. The CTS is basic and either flanks the N-terminus of the TMD or flanks both sides of the TMD. The hydrophobicity of TA-protein TMDs is moderate; however, they exhibit more diverse hydrophobicity scores than SA-protein TMDs (Kim et al, 2019; Lee et al, 2014). The signal length, hydrophobicity, overall charge, CTS, as well as the spacing of features have been shown to contribute to the specific subcellular localization of TA-proteins and are especially important for ER TA protein integration (Teresinki et al, 2019). Chloroplast TA-proteins use a CTS with net positive charge, which is less important than the distribution of charge throughout the CTS (Lee et al, 2014). A subset of chloroplast TA-proteins contain an RK/ST motif within the CTS, which is important for selective plastid-targeting. The RK/ST motif is up to 9 aa long, contains at least 3 K or R residues and 3 S or T residues, and can be located anywhere in the CTS (Teresinski et al, 2019). Some RK/ST sequences are enriched in both positively and negatively charged residues, suggesting charge distribution is more important than the net charge. Interestingly, although these RK/ST motifs vary in sequence, they are interchangeable within those TA-proteins (Teresinski et al, 2019). Although the TA-proteins share similarities in structure and targeting features, these proteins use multiple localization pathways (Lee et al, 2017). In TA-proteins OEP9 and OEP7.2, the CTS and TMD are necessary and sufficient for targeting and their CTS regions contain a RK/ST motif (Lee et al, 2014). Additionally, a net positive charge and the distribution of charge in the CTS, as well as TMD length and hydrophobicity are essential features for selective targeting of OEP9 and OEP7.2 to plastids (Teresinski et al, 2019). On the other hand, TA-proteins Toc33 and Toc34 require a GTPase domain, a TMD, and a CTS for targeting (Kim et al, 2019). Although Toc34 and Toc33 have highly similar sequences, two RK/ST motifs are present in the CTS of Toc34, while the

CTS of Toc33 does not contain an RK/ST motif (Teresinki et al, 2019), suggesting the possibility that RK/ST motifs regulate OEM-targeting specificity. Moreover, RK/ST motifs may be used to regulate protein import-specificity in select tissue-types & cell-types, at various developmental stages, and in response to environmental cues (Teresinki et al, 2019). The C-terminal targeting signal of TA-proteins emerges from the ribosome exit tunnel when translation is terminated. TA-proteins are post-translationally targeted to plastids by cytosolic chaperone proteins (Kim et al, 2019; Lee et al, 2017). In a similar fashion to SA-proteins, AKR2, Hsp70, and Hsp90 transport chloroplast TA-proteins to the OEM. Hsp70 and Hsp90 increase the efficiency of TA-targeting but not fidelity, suggesting they aid AKR2 and other cytosolic factors in targeting and cannot act independently (Kim et al, 2019). Not all TA OEPs need cytosolic factors for OEM-targeting and transfer. Some TA-protein translocation is dependent on events which occur at the membrane and upon the OEM lipid composition (Lee et al, 2014). For example, Toc33 and Toc34 do not require cytosolic factors and self-insert into the OEM. Still, other TA-proteins, like OEP9.1, cannot self-insert into the OEM or use Toc receptors for import and instead rely on some unknown protein import factor (Kim et al, 2019; Teresinki et al, 2019).

1.4.4 Toc159 and CT TP-like OEPs

Although the canonical four import pathways have been established, there are still many OEPs that have yet to be identified that may not fall into the known groups. The most important one among them is Toc159, which as a core protein in Toc complex, plays a vital role in chloroplast protein import and chloroplast biogenesis. Toc159 was predicted to be a TA-protein due to its structural similarity with Toc34 and other TA-proteins. However, Toc159 is not anchored to the membrane using an α-helical TMD, the CTS region of Toc159 also has a net charge of 0 (Lung

& Chuong, 2012). Lastly, a reverse TP-like signal in the C-terminus was identified as a key targeting feature, thus, Toc159 does not meet the criteria of TA-protein classification (Teresinki et al, 2019). Toc159 does not fall into any canonical import pathways as the essential characteristics for specific targeting is missing, and its novel pathway is yet to be determined.

It has been shown that Toc159 is able to insert into the OEM when the G domain is in a GDP-bound conformation (Smith et al, 2002). The G domain alone can associate with the OEM. However, it requires the C-terminal membrane (M) domain for stable OEM insertion (Smith et al, 2002). Interestingly, Lung et al. (2014) demonstrated the complete C-terminal M domain of Toc159 is not essential for stable chloroplast association. To investigate M domain localization, they examined three distinct regions, named the M1, M2, and M3 sub-domains that were defined based on secondary structure predictions. Together, the M2 and M3 domains are necessary and sufficient for chloroplast-OEM targeting. The CT 56 aa of the M3 domain contains a localization signal which is necessary for OEM targeting and the upstream 44 aa of the M2 region is sufficient for membrane anchorage. A GTPase receptor of the Toc complex localized in the outer membrane, Toc159 is not predicted to be anchored to the outer membrane by an neither α-helical transmembrane domain or a β-strand motifs, while at the same time possessing a novel reverse transit peptide-like signal at its C-terminus (Lung and Chuong, 2012; Lung et al, 2014), which would potentially represent a fifth outer membrane targeting mechanism.

Preproteins can be predicted by examining the protein sequences for hydrophobicity, residue representation, phosphorylation sites, and conserved motifs and by analyzing the secondary structure for amphipathic α helices and potential β-sheet. The TP prediction software ChloroP can identify some targeting features within the N-terminus of protein sequences and predict the presence and cleavage site of TPs (Emanuelsson et al, 1999). ChloroP is the most robust

prediction software for chloroplast preproteins to date (Emanuelsson et al, 2007; Patron & Waller, 2007). The original publication describing ChloroP has over 1300 citations on PubMed and 78 of those citations occurred in 2020 (Emanuelsson et al, 1999). Thus, it is a widely accepted and reliable tool for studying preprotein import up to date (Emanuelsson et al, 2007; Bouchnak et al, 2020). The multi-selection multi-order (MM) model was proposed to understand the diversity of TP composition and specificity by Li & Teng (2013). The MM model suggests that TPs are assembled from numerous motifs that interact with distinct molecular factors during the import pathway. The arrangement of these motifs show little similarity in order and location. These predicted motifs can appear highly variable between different TPs (Lee & Hwang, 2018). Functional TP hybrids can also be created by fusing individual motifs from different TPs together, thus demonstrating that dissimilar motifs have interchangeable functions and can work as a whole. As TP motifs fuse to non-chloroplast proteins resulting in stromal-localization, it demonstrates that these motifs are vital for specific targeting (Lee & Hwang, 2018). The diversity in TP motifs makes it difficult to predict TPs using a chloroplast protein sequence. However, such diversity is the key to the complexity of preprotein import (Lee & Hwang, 2018; Li & Teng, 2013; Patron & Waller, 2007). ChloroP was used to identify the novel C-terminal TP-like targeting signal in Toc159. Rather than inputting the sequence from N-terminus to C-terminus, as is typical, the reverse amino acid sequence from C-terminus to N-terminus was also analyzed (Lung et al, 2014). The prediction of a reverse TP-like signal at the C-terminus further validates the speculated non-canonical import pathway. Further analysis by Lung et al. (2014) also suggested that unconventional anchoring of Toc159 to the chloroplast outer membrane via the C-terminal M-domain could be achieved via a lipophilic β-sheet-rich region in the predicted M2 sub-domain due to its high sequence homology with lipid binding domain LpxD, a left-

handed β-helical protein (Lung et al, 2014). The validation of the fifth non-conventional import pathway and enlightenment to its possible anchoring mechanism gave rise to the question whether other OEPs can also utilize this import pathway.

Although the general pathway of import through the Toc/Tic complex is largely understood, how OEPs that utilize this pathway localize within the outer membrane remain unknown. One prevailing hypothesis is the lateral translocation of preprotein during import through Toc complex. As mentioned previously, preproteins normally targeted to the stroma with mutated Proline-less TPs have been shown to be unable to reach the stroma, and to form cytosolic aggregates or to localize in the outer membrane (Lee et al, 2018). This indicates that Proline-less TPs may be able to deliver proteins to the chloroplast outer membrane, but unable to fully translocate into stroma. When this inefficient import happens, the chloroplast undergoes two pathways to ensure the subsequent cargos are translocated efficiently: either reverse the import process and export the preprotein back to cytosol, or translocate the preprotein to the membrane in the intermembrane space. This localizing pathway is not novel as endoplasmic reticulum membrane proteins have been shown to utilize lateral translocation from the import channels (Ismail et al, 2012). However, it is unclear whether altered TP bind less to the chaperone in stroma resulting in reduced pulling force thus stuck in the import channel; or if it binds to the Toc complex and triggers an alternative import/localization pathway.

1.5 T159L candidates and study of OEP18

Using ChloroP in a similar fashion as was done previously for Toc159 by Lung and Chuong (2012), Nick Grimberg (2016) generated a list of OEPs with potential forward and reverse transit peptides (Table 1). This analysis revealed that OEP18 was predicted to contain a reverse TP-like

signal at the C-terminus, and does not contain any predicted α-helical transmembrane domains at either its N- or C-terminus, thereby suggesting that this protein is a candidate for using the same non-canonical targeting pathway to the chloroplast outer membrane as Toc159. Further study of OEP18 also revealed that OEP18 most likely resides in the chloroplast outer membrane and utilizes its C-terminus as a transit peptide and/or membrane anchor (see Figure 1.5). GFP fused to the N-terminus of OEP18 likely disrupts the folding of OEP18, causing the membrane localization to be weaker compared to OEP18-GFP (as seen in the green ring-like structure around red chlorophyll signal in Figure 1.5). Without the C-terminus OEP18 lost its specific localization and aggregated within the stroma. The C-terminus of OEP18 alone likely still has a certain degree of specific localization but weaker than full-length constructs.

Table 1. The amino acid sequences of the eight T159L OEP candidates of *Arabidopsis thaliana* were analyzed using the bioinformatic tool ChloroP in the forward and reverse orientation to identify putative transit peptide and cleavage sites (Grimberg, 2016).

AGI #	Name	Amino Acid Length	Envelope	ChloroP Score (Forward)	cTP Length (Forward)	ChloroP Score (Reverse)	cTP Length (Reverse)
At4g16160	OEP16-2	178		0.433		0.516	33
At2g16640	Toc132	1206	Yes	0.428		0.514	34
At1g07930	E-Tu (protein synthesis)	449		0.432		0.516	25
At4g36650	pBRP	503		0.465		0.515	38

	(transcription)						
At3g63150	MIRO2 (mito)	643		0.495		0.549	59
At2g32240	DUF869	1333		0.427		0.549	82
At4g02482	Putative GTPase	134		0.473		0.548	44
At5g42070	OEP18	164	Yes	0.568	72	0.571	23

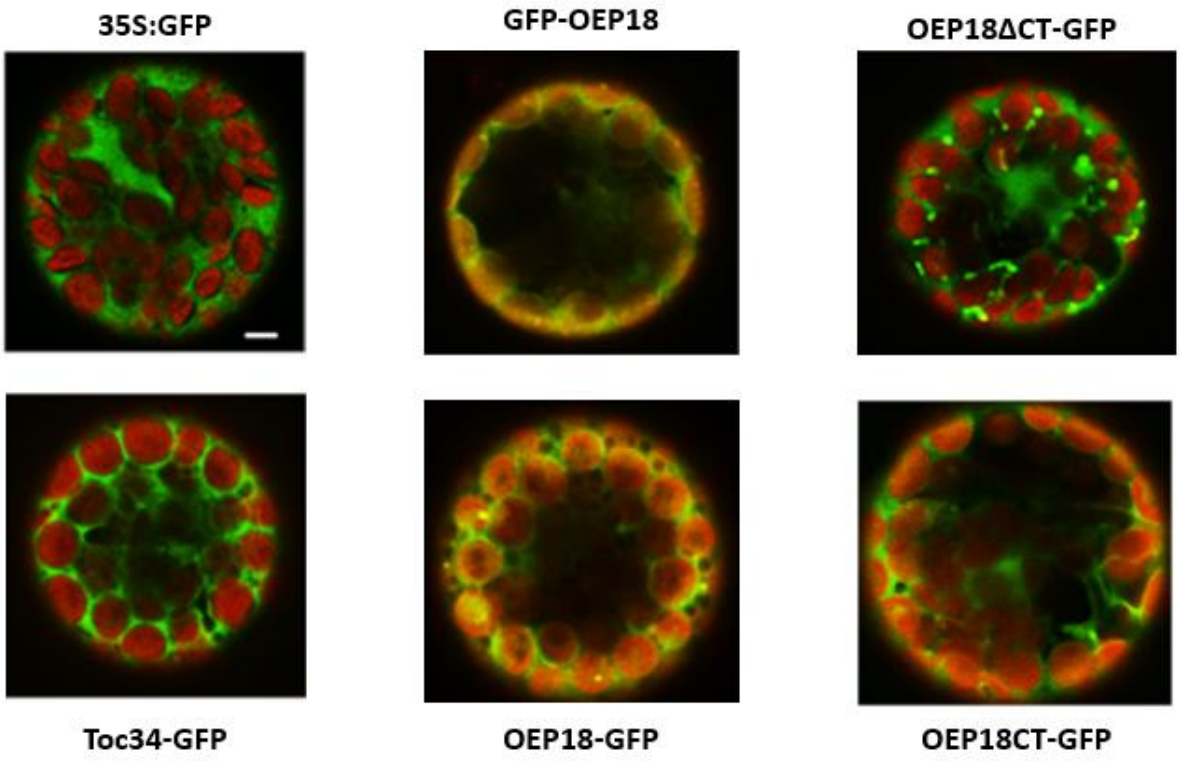

Figure 1.5 Confocal images of transfected protoplasts with OEP18 fusion constructs: Merged optical slices of protoplasts transformed with pSAT6-C1 35S:GFP as negative control, Toc34 as positive control, EGFP–OEP18FL, OEP18FL–EGFP, OEP18ΔCT-GFP and OEP18CT-GFP constructs (Grimberg, 2016)

1.6 Research Goals and hypotheses

Although four canonical import pathways are known to exist and the existence of the potentially novel T159-like import pathway (T159L) was discovered in recent years, the mechanism or pathway of targeting is still unknown for the majority of OEPs. It was validated that this novel T159L pathway utilizes the predicted reverse TP-like sequence near the CT for targeting. Eight potential T159L candidates among 117 OEPs have been identified using ChloroP prediction in the past research. The overall goal of the current research project was to further characterize the T159L targeting pathway by testing whether it is used by other OEPs, and to further validate the T159L candidates.

I hypothesize that OEP18 uses a C-terminal TP-like signal for targeting to chloroplasts and insertion into the outer envelope using the same pathway as Toc159. Three objectives were identified to test this hypothesis. The first goal of the current study was to investigate all 117 known OEPs in an attempt to classify them by predicting the likely targeting pathways. To accomplish this objective, various bioinformatic prediction tools including PSIPred, TMHMM and TOPCONS were used to predict a variety of physico-chemical properties of all OEPs including potential secondary structures and TMD regions from their amino acid sequences. The proteins were then categorized into appropriate predicted import pathways, a process that included comparison to OEPs with known or validated targeting pathways. The second goal was to further verify the subcellular localization of OEP full-length and truncated fusion constructs of OEP18 within *A. thaliana* protoplasts using a biochemical approach. Transfected protoplasts were fractionated into soluble (stroma) and insoluble (membrane and organelle), and the presence and relative amount of GFP-tagged full length or truncated OEP18 was determined using Western Blot analysis and ImageJ quantification. The third goal was to validate prediction

of OEP18 amino acid sequence and its secondary structure formation in the chloroplast outer membrane generated from the first part of my study, by testing for conformational change between hydrophobic and hydrophilic environments. His-tagged OEP18 produced in E. coli was purified using IMAC resins, validated by Western blotting with anti-His antibody. The secondary structure of purified and concentrated recombinant protein was measured using Circular Dichroism spectroscopy and analysis.

The detailed hypotheses are as following:

1. OEP18 and other T159L candidates possess distinctive characteristics that are not used by the four known pathways
2. The predicted C-terminal reverse transit peptide-like sequence is essential for targeting of OEP18 to the outer membrane of chloroplasts
3. The predicted C-terminal transit peptide possesses secondary structure information that is required to anchor OEP18 to the outer membrane

2. Material and methods

2.1 Chemical 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 blue prestained protein standard broad range (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 recombinant constructs were performed using the KOD DNA Polymerase (New England Biolabs, Pickering, ON, Canada; cat. no. F- 530S), whereas colony PCR were performed using Taq 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 117 known and predicted Arabidopsis chloroplasts OEPs (Inoue, 2015) were analyzed using TMHMM v2.0 program (http://www.cbs.dtu.dk/services/TMHMM/)

for possible transmembrane helix (TMH) domains, TOPCON (http://topcons.cbr.su.se/), which also predicts for TMH domains but collectively with 6 different algorithms, PSIpred (http://bioinf.cs.ucl.ac.uk/psipred/) for potential β-strand domains, α-helices, and coil domains, and PRED-TMBB (http://bioinformatics.biol.uoa.gr/PRED-TMBB/) for possible β-barrel formation. All nucleotide and amino acid sequences of 117 OEPs were obtained from the National Center of Biology Information (NCBI) database. Analyses using these tools were combined with ChloroP results from Grimberg (Nick Grimberg, 2016) in an effort to predict the import pathway used by of all 117 OEPs as a way of categorizing them. The T159L candidates previously predicted by Grimberg were also be ruled out if they are predicted and known to possess characteristics of other import pathways. The results were compared to other studies on OEP targeting pathways, including from Teresinki (2015).

2.3 Plant growth condition

Wild-type *Arabidopsis thaliana* (ecotype Columbia) seeds were incubated and stratified for at least 48h 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.

2.4 Construction of OEP18 fluorescent fusion constructs

The OEP18 full-length and truncated fluorescent fusion constructs used in this study were previously generated by Nick Grimberg (2016). Briefly, the entire OEP18 open reading frame (ORF) or DNA fragments encoding specific amino acid residues of OEP18 were cloned into the XhoI-BamHI sites at the N-terminus of EGFP of the pSAT6-N1 vector or at the C-terminus of EGFP of the pSAT6-C1 vector. Each DNA fragment was PCR-amplified using a different primer set with restriction enzyme sites (specifically XhoI and BamHI) to be in-frame with the coding sequence for EGFP of the corresponding vector.

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

Fusion protein	Vector	Oligonucleotide		
		name	Sequence (5′ to 3′)	Orientation
GFP-OEP18	C1	OEP18F1	CGC**CTCGAG**CTATGGCGAA TTCCATTTCATCA	Sense
		OEP18R1	CGC**GGATCC**TCACTTGTTTG AACTTTTGCT	Anti-sense
OEP18-GFP	N1	OEP18F2	CG**CCTCGAG**CATGGCGAAT TCCATTTCATCA	Sense
		OEP18R2	CGC**GGATCC**CCTTGTTTGAA CTTTTGCTAGA	Anti-sense
OEP18ΔCT-GFP	N1	OEP18F2	CGC**CCTCGAG**CATGGCGAA TTCCATTTCATCA	Sense
		OEP18R3	CGC**GGATCC**TCAAGTCACC ACGACCAAATGCAA	Anti-sense
OEP18CT-GFP	N1	OEP18F4	CGC**CCTCGAG**CTCTAAATCC TCCACTTCTGTA	Sense
		OEP18R2	CGC**GGATCC**CCTTGTTTGAA CTTTTGCTAGA	Anti-sense

2.5 Isolation and transfection of mesophyll protoplasts from *A. thaliana*

The initial procedure used for isolating mesophyll protoplasts from *A. thaliana* was modified from Lung et al. (2014) Briefly, 30-40 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 before 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 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 and the observation of round-shaped protoplasts under light microscopy. Cell wall digestion and protoplasts release was intermittently monitored by visualizing under a light microscope. Release of healthy protoplasts is indicated by the presence of spherical cells that are not clumped together (Lung and Chuong, 2012).

A new isolation method developed by Wu et al (2009) was also used for isolating protoplasts. Briefly, 20-30 leaves were placed on the sticky side of a piece of masking tape with the lower

epidermal layer facing upward, and then covered with a piece of 3M magic tape. Fingers or the round bottom of a falcon tube were used to gently press the magic tape on to the leaves to ensure the lower epidermal layer was in complete contact with the magic tape. The lower epidermal layers were then stripped away by removing the 3M magic tape. The enzyme solution was prepared in the same way as described previously, but with less enzyme [1% w/v cellulase and 0.25% w/v macerozyme]. Stripped leaves were incubated in enzyme solution for 1-1.5h at room temperature with gentle shaking while still adhered to the masking tape. The new "tape-sandwich" method uses less enzyme making it more cost-effective, and releases more healthy protoplasts per leaf, meaning fewer leaves are needed, and in a shorter amount of time (1 h vs 3 h), as compared to the previous method. The results from new method are consistent with the results from the older method previously for all fusion constructs, thus both samples are combined for analysis and the results from the older method make up 1/4 of the total sample size.

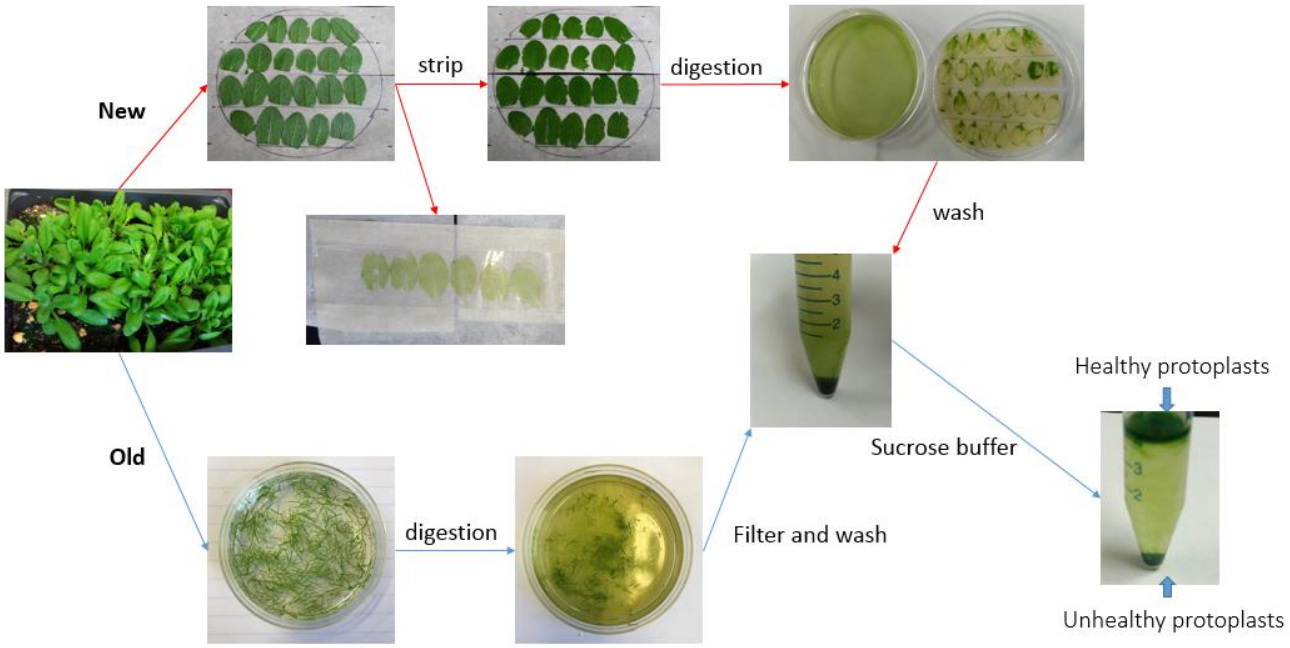

Figure 2.1 Comparison between "leaf strips" old cutting method and new tape-sandwich method: A flowchart comparing the cutting method and tape-sandwich method for isolating protoplasts starting from leaf processing to healthy protoplast isolation.

After incubation, the solution containing the released protoplasts was transferred using a Pasteur pipette from the Petri dish to a 15 mL falcon tube 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 when using the old cutting method; the new method did not require filtering because leaf tissue remained stuck to the masking tape. 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] to release any remaining released protoplasts was transferred using a Pasteur pipette into the Petri dish and the rinse solution was then transferred to the same 15 mL Falcon tube. The solution 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 isolate the healthy protoplasts which are found in the floating layer of this step. The internatant and pellet containing unhealthy and broken protoplasts were carefully removed without disturbing the floating layer using a Pasteur pipette. The healthy protoplasts were then diluted in 1 mL of W5 solution by a gentle swirling motion. 10 μL of protoplast solution was placed on a haemocytometer to estimate the number of isolated healthy protoplasts. The resuspended healthy protoplasts were incubated on ice for 30~60 min, during which the protoplasts settled to the bottom of the 15 mL tube.

After protoplasts were settled to the bottom, the supernatant was carefully removed, and protoplasts were resuspended in 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 test 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).

In a standard reaction of PEG-mediated transfection, 20,000 protoplasts were mixed with 5 µg of concentrated plasmid DNA (>1µg/µL) 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 8~10 times and 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 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 60,000 protoplasts with 15 µg of plasmid DNA, and the transfected protoplasts were cultured overnight for 16~18 hours with 3 mL of WI solution.

2.6 Fractionation of transfected protoplasts

After an overnight incubation, transfected protoplasts were visualized using an epifluorescence microscope to determine the rate of transfection. Protoplast would be used for the fractionation process if the estimated transfection rate was estimated to be at least 60%.

Total protein was extracted from ¼ of the transfected protoplasts from each reaction. Transfected protoplasts were pelleted at 200 g for 2 min and the supernatant was discarded. The protoplast pellet was vortexed for 2 minutes in 100 µL of lysis buffer (25 mM Tris-HCl (pH 6.8), 1% (v/v) Triton X-100 and 1 mM DTT) and centrifuged at 14,000 rpm for 10 min at room temperature. The supernatant containing protein was transferred to a fresh centrifuge tube, and total protein was concentrated by precipitating in five volumes of acetone at -20 ^{o}C for 2h. Proteins were collected by centrifugation at 14,000 rpm at 4 ^{o}C for 30 min. The supernatant was discarded, and the pellet was air dried for 10 min and stored at -20 ^{o}C until use.

The remaining ¾ of the remaining transfected protoplasts were fractionated using one of two separate methods. The first method was intended to only rupture protoplasts but not any organelles. Approximately 300,000 transfected protoplasts were pelleted at 100 g for 2 min and the supernatant was discarded. The protoplast pellet was then resuspended in 300 µL HS buffer [50 mM HEPES-KOH (pH 7.3), 0.330 M sorbitol and 1 mM PMSF] and flicked gently to mix. The protoplasts were lysed by gently pushing the solution through a rupture device consisting of 10µm nylon affixed to the end of a syringe. The flow through lysate solution was examined under a light microscope to ensure all protoplasts had been ruptured, but chloroplasts remained intact. The solution was centrifuged at 1,000 x g for 5 min in a swinging-bucket rotor. The supernatant containing soluble protein from the cytoplasm, stroma and vacuole was transferred to a new microfuge tube and concentrated in the same way as total protein extract with acetone

precipitation described above. The pellet contained protoplast membranes (e.g., plasma membranes and tonoplasts) and intact organelles including chloroplasts was stored at -20 ^{o}C until use.

The other method caused the complete rupturing of protoplasts and its organelles including chloroplasts. The transfected protoplasts after overnight incubation in WI buffer were pelleted and then incubated with 300 μL HS buffer but without sorbitol, and the solution was vortexed vigorously for 2 min. After centrifuging at 5,000 g for 10 min, the supernatant contained all soluble proteins from the cytosol chloroplast, stroma and all other ruptured organelles, and which were precipitated with acetone. The pellet contained total protoplast membranes including all ruptured organelle membranes. All total protein extracts, soluble and membrane fractions were stored at -20 ^{o}C prior to SDS-PAGE analysis.

2.7 Epifluorescence microscopy

Prior to fractionation of transfected protoplasts, 5 μL of transfected protoplasts in W5 solution was examined in flat-bottom chamber slides made of nail polish-premounted coverslip. 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 #4) under UV illumination, and the chlorophyll 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 three independent experiments.

2.8 SDS-PAGE, Western blot and ImageJ band intensity quantification

The methods for SDS-PAGE were adopted from Laemmli (1970). The 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 10 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 10 min and resolved by 12% SDS-PAGE. A 12% (w/v) resolving gel was prepared comprised of the following: 2.4 mL of 30% (w/v) acrylamide, 2.25 mL of 1 M Tris-HCl (pH 8.8), 1.26 mL of distilled water, 60 µL of 10% (w/v) SDS, 60 µL of 10% (w/v) APS, and 6 µL of TEMED. A 4.8% (w/v) stacking gel was prepared comprised of the following: 400 µL of 30% (w/v) acrylamide (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, 25 µL of 10% (w/v) APS, and 5 µL of TEMED. The commercial molecular weight ladder Precision Plus Protein Standards All Blue or All Blue prestained protein standard broad range (Bio-Rad, Mississauga, ON, Canada) was used. The control was also used as 100~200ng of purified recombinant EGFP in 10 µL 6x SDS-PAGE sample buffer. Gel electrophoresis was run at 90 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 140 V until the dye reach bottom of gel or ran off the gel.

The resolved proteins from SDS-PAGE were transferred onto a polyvinylidene difluoride membrane (PVDF) (Bio-Rad, Mississauga, ON, Canada) at 15 V for 40 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 10 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 5 min. The Ponceau-stained membrane was then rinsed in distilled water at least 4 times until protein bands became visible. The membrane was then 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 1h at room temperature with gentle shaking. The membrane was then incubated in blocking solution with a primary polyclonal antibody raised in rabbit against enhanced green fluorescent protein (EGFP; 1:5,000) overnight at 4 °C with shaking.

The membrane was washed 3 times for 10 min each in TBS-T and then incubated in blocking solution with an anti-rabbit secondary antibody conjugated to horseradish peroxidase (1:50,000) (Sigma- Aldrich, Oakville, ON, Canada) at room temperature for 1h with shaking. The same 3 x 10 min wash with TBS-T was used. 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 (Bio-Rad Laboratories Inc., Mississauga, ON, Canada). Captured images were processed using Adobe Photoshop CS (Adobe System Inc., Seattle, WA, USA). Protein band intensity of each lane from each membrane is then quantified using ImageJ program (https://imagej.nih.gov/ij/).

2.9 Construct design of OEP18 recombinant protein

For the purpose of producing recombinant OEP18 protein for structural and topology studies, the cDNA encoding the full-length OEP18 protein ordered from the Arabidopsis Biological Resource Centre (ABRC, OSU, Columbus, OH, USA) was cloned into the NdeI-XhoI sites of the pET28a expression vector. Briefly forward and reverse primers were used to introduce NdeI and XhoI restriction sites into the end of the cDNA using PCR such that the cDNA would be in frame with the translation start site and C-terminal 6x His tag coding sequence that is provided by the pET28a vector.

100 µL restriction digest reactions for purified OEP18 PCR products and pET28a vector were prepared by adding NEBuffer 3 (1x) (New England BioLabs category # B70003S), purified plasmid DNA (5 µg) or purified OEP18 (5 µg),NdeI (50U) 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 first enzyme. XhoI (50 U) was then added to the reaction tubes and placed in a water bath for 3-4h at 37°C. The tubes were then incubated at 65°C for 20 min to heat-inactivate XhoI. The double digested products were then examined on 1% agarose gel and gel-purified. The OEP18 inserts were ligated with the pET28a 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 heat shock transformed with chemically-competent DH5α *E. coli* cells and incubated on ice for 20 min. Transformed cells were selected by growth on LB agar plates containing ampicillin (100 µg/mL). Selected colonies were screened with colony PCR, plasmid DNA of each positive colonies was then amplified with Miniprep kit (Biobasic Inc., Markham, ON, Canada), and send to Sanger Sequencing Facility at The Centre for Applied Genomics (The Hospital for Sick Children, Toronto, ON, Canada) for sequencing

confirmation.

Table 2.2 Oligonucleotides used for construction of OEP18 recombinant protein

Name	Sequence (5′ to 3′)	Orientation
OEP18NdeF2N	CGC**CATATG**ATGGCGAATTCCATTTC	Sense
OEP18XhoR1	AAA**CTCGAG**TCACTTGTTTGAACTTT	Antisense

```
attttgttttgcgtttctgaatttgtggccattatcttctcacactctcttctcttagctc

acagaggaaagaaaaatggcgaattccatttcatcaatatctctgcctcgatgtttcatc
                 M  A  N  S  I  S  S  I  S  L  P  R  C  F  I
ttcaacaatggtagtcacaaatcaaggccatggccaagctcaagcagtttctttctcaac
 F  N  N  G  S  H  K  S  R  P  W  P  S  S  S  S  F  F  L  N
aaatcatcaaagcatcatcctcatccactactctctctttcttcttcgccgtcttccgtc
 K  S  S  K  H  H  P  H  P  L  L  S  L  S  S  S  P  S  S  V
gtagaaactgataatgacgacgacaacgacctcactttcagtggttgcagagcgtgtggg
 V  E  T  D  N  D  D  D  N  D  L  T  F  S  G  C  R  A  C  G
aaagaagagaaagagagtgggtgtaatggtgacggacggattcaaggtggcattgcaact
 K  E  E  K  E  S  G  C  N  G  D  G  R  I  Q  G  G  I  A  T
gtccccggtttcggttggtggccaattaaggcttacaggccttgtcccgcgtttgttgag
 V  P  G  F  G  W  W  P  I  K  A  Y  R  P  C  P  A  F  V  E
gccggaggtagataccgccgcatagggcagagcatggacgaggttgcatttggtcgtggt
 A  G  G  R  Y  R  R  I  G  Q  S  M  D  E  V  A  F  G  R  G
gactctaaatcctccacttctgtagacaccagtgattcactactacgccagacaaagcca
 D  S  K  S  S  T  S  V  D  T  S  D  S  L  L  R  Q  T  K  P
acaagttctagcaaaagttcaaacaagtgatatatagcccagtttttcaagtccaagaag
 T  S  S  S  K  S  S  N  K
aacttttttggaaatgtgtaatgatgattgggcctaaacatattatcggcccatttata

attttacggctcaatagttaagaaaccccaaaaaaaaaaaaaaaaa
```

Figure 2.2 DNA sequence of OEP18 and deduced amino acid sequence of OEP18 from *Arabidopsis* (Grimberg, 2016): The full-length nucleotide sequence (At5g42070) coding for the translational product of Arabidopsis OEP18.

2.10 Recombinant protein overexpression and purification with IMAC

The confirmed plasmid of pET28a vector containing in-frame OEP18 or pET28a-OEP18 was then heat-shock transformed within *E. coli* BL21 Codon+ cells. Transformed cells were selected by growth on LB agar plates containing kanamycin (34 μg/mL), chloramphenicol (50 μg/mL) and streptomycin (75 μg/mL). Selected colonies were verified with colony PCR with KOD polymerase. Positive colonies were incubated with 4 mL of LB broth containing 34 μg/mL kanamycin and 50 μg/mL chloramphenicol with shaking at 37°C overnight. 3 mL of overnight culture is further incubated with 300 mL of LB broth with the same antibiotics with shaking at 37°C for protein induction later. As OD600 reached 0.4~0.5 after 2~2.5h, 1mL of uninduced culture sample was transferred to a microfuge tube, centrifuged at 5,000 g for 2 min, and supernatant was removed. IPTG was added to the remaining uninduced culture to a final concentration of 1 mM, and the culture was then incubated at 37°C with shaking for 6h~12h till induction was complete. 250 μL of induced culture sample was transferred to a microfuge tube, centrifuged at 5,000 g for 2 min, and supernatant was removed. Both uninduced and induced samples with blue prestained protein standard broad range were resolved in 10% acrylamide SDS-PAGE gels, and stained with Coomassie stain G-250 with shaking for 1h. It was then destained with Coomassie destain solution [40% (v/v) methanol and 10% (v/v) acetic acid] for 2h. OEP18-His protein was confirmed based on its predicted molecular weight (19 kDa with addition of 6xHis tag).

The rest of the overnight culture was centrifuged at 5,000 g for 5 min, the supernatant was removed, and the cell pellet was stored at -20 °C. After overexpression was confirmed, the cell pellet was resuspended in10 mL lysis buffer containing 50 mM Tris-HCl (pH 7.5), 100 mM NaCl and 4 M urea. The cell pellet was completely dissolved by vortexing and rocking for 5 min.

Cells were completely ruptured using French Pressure cell press with 20,000 kPa setting. Solution was then centrifuged at 15,000 g for 30 min at 4°C. Insoluble fraction was stored in -20 °C, soluble fraction was transferred to a new Falcon tube and purified with Immobilized Metal Affinity Chromatography (IMAC).

0.6 mL of slurried Profinity™ IMAC Ni-Charged Resin (Bio-Rad Laboratories, Hercules, CA, USA, cat.no. 156-0135) was prepared and used according to manufacturer's instruction. Briefly, after resin in the column had been equilibrated with binding buffer [50 mM Tris-HCl (pH 7.5), 100 mM NaCl and 5 mM imidazole, 4 M urea], the soluble fraction was applied to the column and mixed thoroughly by pipetting and rocking. The column was then incubated at 4 °C with rocking for 30-60 min. The flow-through (FT) fraction was collected which represents all unbound proteins. The column was washed 5 times with 1 mL washing buffer [50 mM Tris-HCl (pH 7.5), 100 mM NaCl and 20 mM imidazole, 4 M urea] each time to remove unbound and weakly bound proteins, all 5 wash fractions were collected. The elution series starts with 1 mL of E1 buffer [50 mM Tris-HCl (pH 7.5), 100 mM NaCl and 100 mM imidazole, 4 M urea], then 1 mL E2 buffer with 200 mM imidazole and 1 mL E3 buffer with 500mM imidazole. All elution fractions were collected. All fractions collected during IMAC purification along with starting soluble fraction were analyzed on a 10% acrylamide gel SDS-PAGE stained with Coomassie. The E1 fraction was confirmed to contain purified target protein, and therefore the protein concentration was quantified using the Bradford (Bio-Rad, Mississauga, ON, Canada). Protein concentration was estimated against standard solutions of BSA from 20~500 μg/mL. The E1 fraction was then concentrated with Vivaspin 500 centrifugal concentrator to 20 μL. The concentrated sample was resolved using a 10% acrylamide gel SDS-PAGE gel and stained with Coomassie R-250. The target protein band was cut out and grinded to small pieces in a

microfuge tube. Gel pieces were then destained with traditional Coomassie destain solution [50% (v/v) acetonitrile and 50% 25 mM NH_4HCO_3]. Gel pieces were sent to Advanced Analysis Centre at University of Guelph for confirmation of the amino acid sequence.

2.11 Sample Dialysis for Circular Dichroism analysis

After IMAC purification and amino acid sequence of OEP18-His was confirmed, protein samples from IMAC elution fractions were combined and underwent sequential dialysis into compatible CD buffer (100 mM NaF, 20 mM Tris-HCl pH 7.5) without urea and imidazole. Dialysis was achieved in 6 steps in 1 L dialysis buffer over at least 6 hours, any protein precipitation was removed during buffer change between steps. After dialysis was complete, the concentration of the dialyzed protein sample was determined using the Bradford Assay, the sample was analyzed using SDS-PAGE to check protein integrity and then analyzed using Circular Dichroism (analysis performed by Michael Fish at Wilfrid Laurier University). Far-UV CD Spectroscopy was performed using an AVIV CD Spectrometer Model 215 at 25 degrees Celsius and analyzed using AVIV CDSD software. The sample was contained in a high precision quartz cuvette with a 1 mm path length. The deconvolution was performed with DichroWeb (dichroweb.cryst.bbk.ac.uk) using the K2D method. 3 different buffers were selected for CD: the CD compatible buffer (100 mM NaF and 50 mM Tris), 50% TFE and 50% liposome. TFE is the organic solvent which is able to induce structural conformations of protein which may not be able to form under aqueous condition. Liposomes were used to mimic the surrounding of membrane bilayer, structural change of protein will be observed if liposome-peptide binding occurs.

3. Results

3.1 Analysis and categorization of 117 OEPs of *Arabidopsis*

Inoue (2015) outlined the 117 proteins identified or predicted to be in the chloroplast outer membrane (COM) in *Arabidopsis*. Functions assigned to some of the OEPs include solute and ion transport, preprotein import, protein turnover, lipid and carbohydrate metabolism and intracellular communication (Inoue, 2015). Past research done by Nick Grimberg (2016) utilized ChloroP analysis to identify possible TP-like targeting sequences at the N- and C-termini of each protein. The sequences of all 117 COM proteins were analyzed in both forward and reverse orientation as was done previously by Lung and Chuong (2012) to identify a TP-like sequence in the reverse orientation at the C-terminus (CT) of Toc159 in *Bienertia sinuspersici*, which is able to target the protein to the outer membrane. The goal of my bioinformatic analysis was to further categorize the 117 OEPs into distinct COM targeting/insertion pathways by combining the previous ChloroP analysis with transmembrane α-helix (TMH) domain prediction using TMHMM 2.0 and TOPCONS, membrane topology prediction using PSIpred, and transmembrane β-strands and β-barrel prediction using PRED-TMBB. All amino acid sequences of 117 OEPs were obtained from the National Centre for Biotechnology Information (NCBI) database.

TMHMM predicts potential transmembrane helix domain (TMD) α-helices using Hidden Markov model, and TOPCONS is a collection of 6 other different algorithms predicting TMDs. PECT1 and Peptidase M16 are two representative proteins predicted to be SA and TA proteins respectively, as both proteins are predicted to have the highest possibility of containing a TMD α-helix at their NT and CT by both TMHMM and TOPCONS respectively (Figure 3.1). It is important to emphasize, however, that these prediction tools are not completely accurate. For

example, the TMD at the CT of Toc33, a well characterized TA protein localized to the COM, is only predicted with minimum probability by TMHMM, and 2 algorithms out of 6 by TOPCONS (Figure 3.2). Thus, proteins with TMD probabilities below threshold in TMHMM or detected by any algorithm in TOPCONS should still be considered for possible SA/TA pathway.

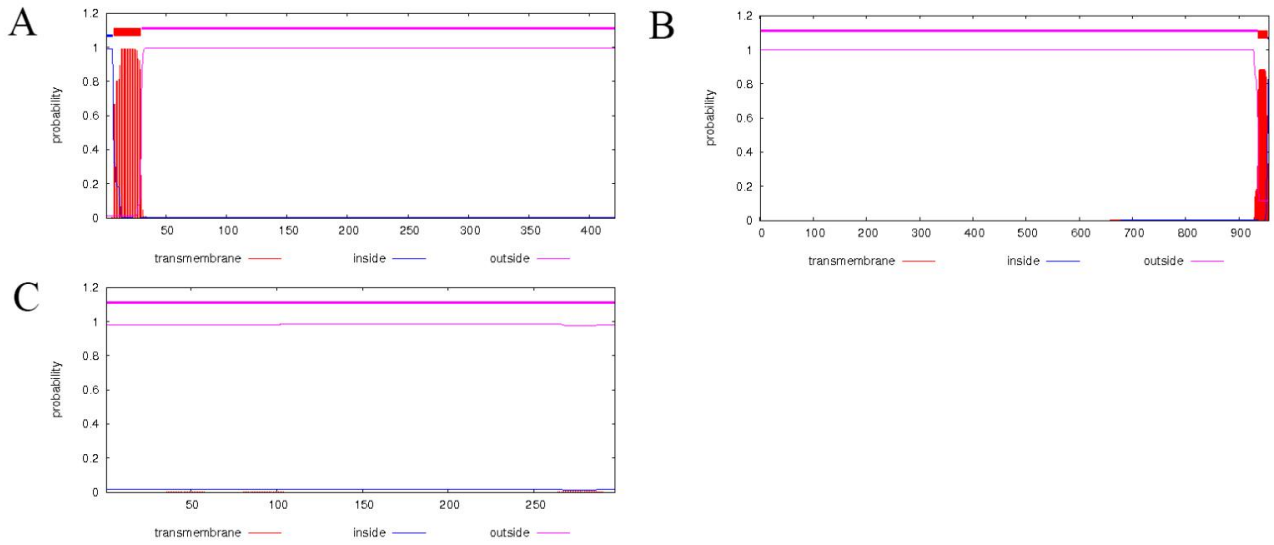

Figure 3.1 TMH prediction graphs by TMHMM of representative OEPs: the red bar on the graph indicating possibility of TMH at specific position within the target protein. A) PECT1, A known SA OEP with high probability of TMH at N-terminus; B) Peptidase M16, a known TA OEP with high probability of TMH at C-terminus; C) Toc33, a known TA OEP with a C-terminal transmembrane domain with minimal probability of TMH at C-terminus predicted by TMHMM

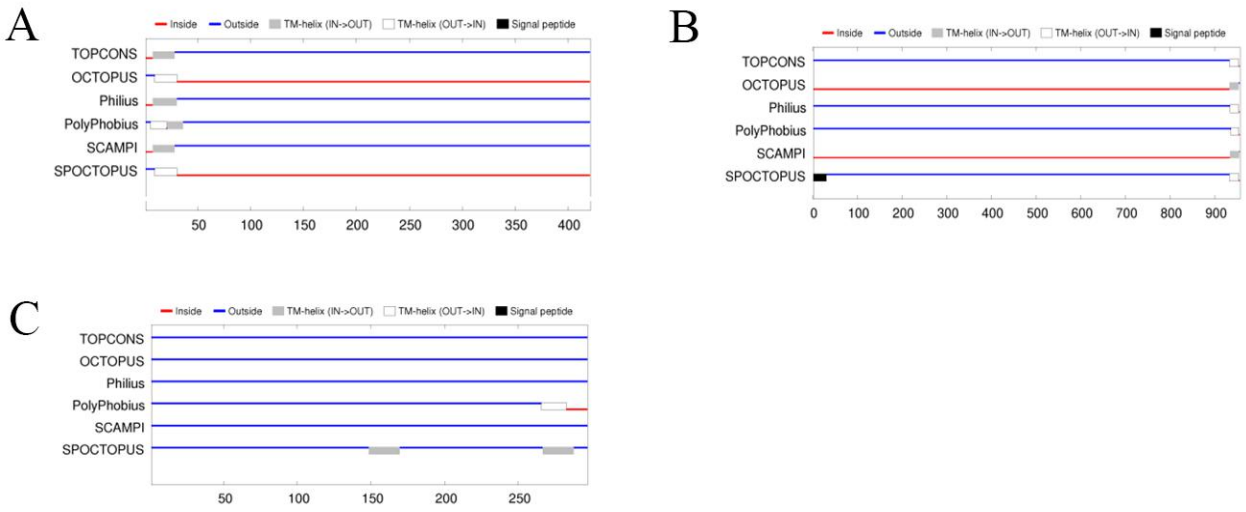

Figure 3.2 TMH prediction graphs by TOPCONS of representative OEPs: A) PECT1, A known SA OEP with predicted TMH at N-terminus by 6 algorithms; B) Peptidase M16, a known TA OEP with predicted TMH at C-terminus by 6 algorithms; C) Toc33, a known TA OEP with predicted TMH at C-terminus by only 2 algorithms. Boxes of grey and white on specific region of target protein indicates possible transmembrane helix. Red and blue line indicates predicted protein orientation if transmembrane helix is predicted to be present. Red line indicates the region of protein inside the membrane, and blue line indicates region of protein outside membrane.

PSIpred using PSI-blast predicts potential β-strands within the target protein using artificial neural network learning methods. PRED-TMBB predicts potential β-strands using Hidden Markov model and also visualizes potential β-barrel (BB) formation in a 2D plane. Outer Membrane Proteins found in Gram-negative bacteria, mitochondria and chloroplasts contain 8~24 β-strands (Fairman, 2012), and the average length of strands range from 11 to 14 residues depending on type of porin and β-barrel (Tamm, 2004). OEP21 provides a representative example of a BB protein predictions: the average length of β-strands in OEP21 is predicted to be above 8 by both PSIpred and PRED-TMBB, and it is predicted to possess transmembrane properties and able to form a β-barrel (Figure 3.3). In early research, BB protein has been determined that in order to be physically capable of forming a β-barrel, it is essential to have at

least 8 transmembrane β-strands and an average length of 10 amino acids (Wimley, 2002). However different prediction tools for β-barrel proteins provide variable results in terms of the precise length and number of β-strands of the same OEP. Thus for the purpose of the current study, any OEPs predicted to have at least 6 β-strands and with an average length of at least 10 amino acids per strand are considered as potential β-barrel membrane proteins.

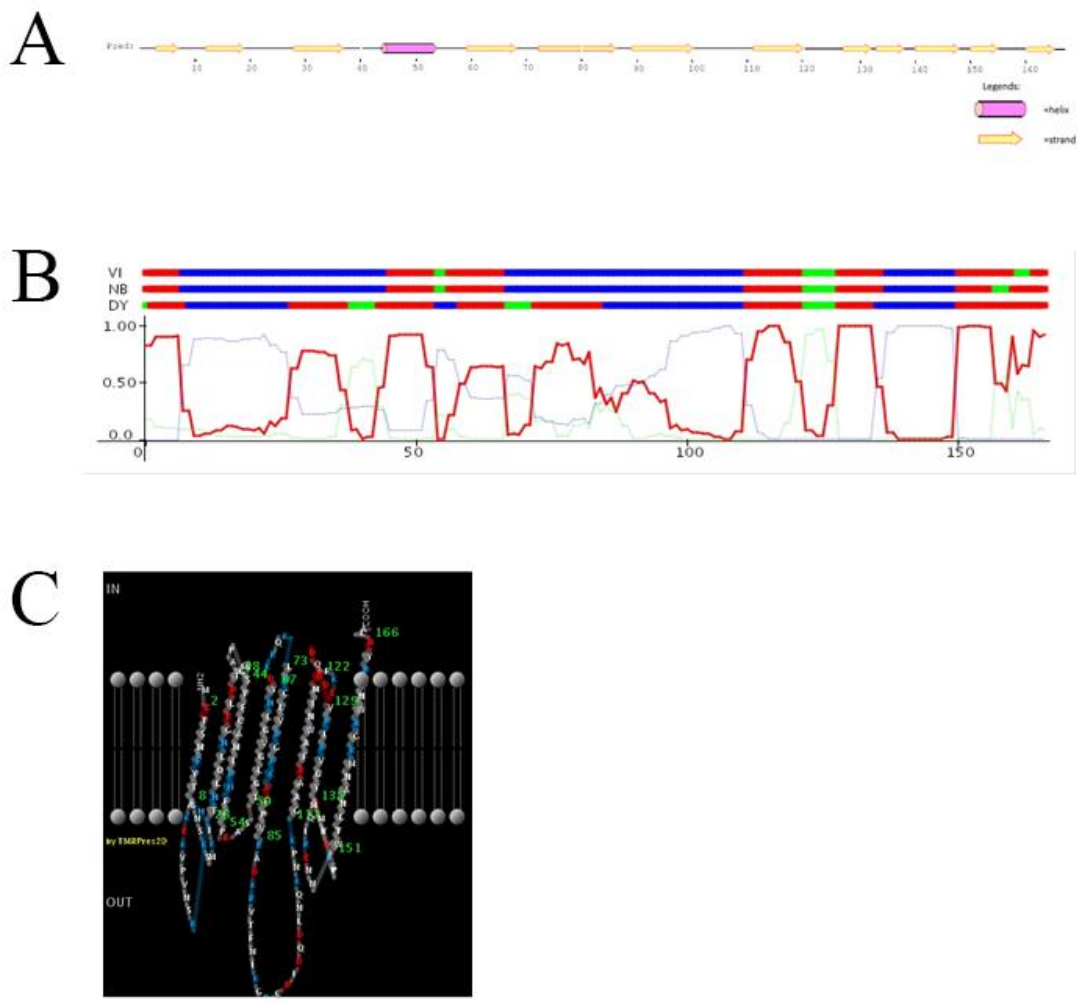

Figure 3.3 transmembrane β-strand and β-barrel prediction of OEP21: A) OEP21 is predicted to have 12 β-strands by PSIpred indicated by yellow arrow; B) OEP21 is predicted to have 9 β-strands by PRED-TMBB, the red region is the region of predicted β-strand; C) 2D visualization of potential β-barrel formation of OEP21 by PRED-TMBB, all eight predicted β-strands are folded in 2D to form a transmembrane β-barrel between the lipid bilayer.

Table 3.1 The amino acid sequences of 117 known or predicted chloroplast outer membrane proteins of *Arabidopsis* were analyzed using the bioinformatics tool TMHMM 2.0 and TOPCONS for transmembrane prediction, PSIpred for secondary structure prediction, and PRED-TMBB for transmembrane β-strand and β-barrel prediction. Proteins previously predicted to contain transit peptide-like sequences at their C-terminus in the reverse orientation (Grimberg, 2016) are highlighted in yellow.

AGI #	Name	A.A Length	ChloroP				TMHMM		TOPCONS		PSIpred	PRED TMBB	predicted pathway
			ChloroP Score (Forward)	cTP Length (Forward)	ChloroP Score (Reverse)	cTP Length (Reverse)	expected number of amino acid in TMH (location)	expected number of TMH	number of TMH (location) different from TMHMM	predicted by # of algorithm (out of 6)	number, average length of β-strands	possible β-barrel	
At1g20816	OEP21-1	167	0.446		0.432		0	0	0	0	12, 8	Y	BB
At1g45170	OEP24-1	213	0.538	31	0.44		0	0	0	0	16,9.2	Y	TP or BB
At1g76405	OEP21-2	167	0.448		0.431		0	0	0	0	11, 13.2	Y	BB
At2g01320	WBC7	728	0.451		0.426		98(12-34,492-514,526-548,553-575)	4	8	6	11, 6.2	N	x
At2g17695	OEP23/DUF1990	205	0.449		0.449		0	0	0	0	9, 6.7	Y	BB
At2g28900	OEP16-1	148	0.491		0.45		127-147	0	0	1	2, 5.5		TA
At2g43950	OEP37	343	0.576	73	0.438		0	0	0	0	17, 9	Y	TP or BB
At3g51870	PAPST1 homolog	381	0.51	26	0.447		54 (240-262,277-298)	2	0	3	3, 5.7	N	TP
At3g62880	OEP16-4	136	0.456		0.489		(55-78, 84-104, 105-130)	0	0	3	0		x
At4g16160	OEP16-2	178	0.433		0.516	33	157-177	0	0	1	3, 3.3		T159L or TA
At5g42960	OEP24-2	213	0.477		0.463		0	0	0	0	15, 9	Y	BB
At1g02280	Toc33	297	0.477		0.431		(265-290)	0	0	2	7, 6.7		TA
At2g16640	Toc132	1206	0.428		0.514	34	(1175-1195)	0	0	1	21, 6	N	T159L or TA
At2g17390	AKR2B	344	0.446		0.433		0	0	0	0	0		x
At3g16620	Toc120	1089	0.427		0.491		(1056-1076)	0	0	1	27, 7	N	TA

At3g17970	Toc64-III	589	0.443		0.435		18 (5-22)	1	2 (5-25, 398-418)	6	8, 5		SA
At3g44160	p39/OEP80tr1	362	0.478		0.492		(235-250)	0	0	0	23, 8	Y	BB
At3g46740	Toc75-III	818	0.585	79	0.447		(1-90)	0	0	0	33, 9	Y	TP or BB
At3g48620	p36/OEP80tr2	321	0.481		0.479		0	0	0	0	17, 7.5		x
At4g02510	Toc159	1503	0.488		0.478		(1460-1500)	0	0	2	25, 6		T159L
At4g09080	Toc75-IV	396	0.423		0.44		0	0	0	0	18, 10	Y	BB
At5g05000	Toc34	313	0.439		0.443		(260-313)	0	0	2	7, 5.7		TA
At5g19620	OEP80/Toc75-V	732	0.535	93	0.442		0	0	0	0	27, 8.2	Y	TP or BB
At5g20300	Toc90	793	0.471		0.434		(260-395)	0	0	1	22, 6	Y	TA or BB
At1g02560	ClpP5 (proteolysis)	298	0.568	62	0.427		(180-210)	0	0	2	4, 3.5		TP
At1g07930	E-Tu (protein synthesis)	449	0.432		0.516	25	0	0	0	0	18, 5.3		T159L
At1g09340	HIP1.3/RAP38/CSP41B (protein synthesis)	378	0.495		0.432		(1-70)	0	0	1	12, 3.8		x
At1g63900	SP1 (proteolysis)	347	0.451		0.442		(220-255)	0	2 (2-22, 223-243)	6	17, 6	N	x
At1g67690	M3 protease	710	0.458		0.451		22 (21-43)	1	0	1	6, 5		SA
At3g46780	pTAC16 (transcription)	510	0.512	19	0.469		0	0	0	0	12, 3.4		TP
At4g05050	UBQ11 (proteolysis)	229	0.428		0.431		0	0	0	0	15,5.5	Y	BB
At4g32250	Tyrosine Kinase	611	0.427		0.453		(550-580)	0	1 (552-572)	6	22, 3.2	N	TA
At4g36650	pBRP (transcription)	503	0.465		0.515	38	0	0	0	0	3,3		T159L
At5g16870	PTH2 family (protein synthesis)	169	0.444		0.433		(3-20)	0	1 (2-22)	5	4,6	N	SA

At5g35210	PTM (transcription)	1706	0.442		0.465		91 (1537-1554, 1567-1589, 1604-1626, 1633-1655, 1680-1702)	5	5	6	30, 4		x
At5g56730	Peptidase M16 family	956	0.439		0.466		17 (936-953)	1	1	6	21, 5		TA
At1g77590	LACS9	691	0.45		0.432		(1-15)	0	0	2	26, 5	N	SA
At2g11810	MGD3	465	0.445		0.432		0	0	0	0	13, 4.6		x
At2g27490	ATCOAE	232	0.475		0.445		(210-230)	0	1	5	5,4		TA
At2g38670	PECT1	421	0.439		0.433		21 (7-29)	1	1	6	14, 4		SA
At3g06510	SFR2/GGGT	656	0.491		0.44		31 (4-26, 482-504)	2	0	0	9,4		x
At3g06960	TGD4	479	0.439		0.435		(290-310)	0	0	0	21, 5		TA
At3g11670	DGD1	808	0.559	58	0.453		0	0	0	1	11, 5		TP
At3g26070	PAP/FBN3a	242	0.57	50	0.431		0	0	0	0	10, 6.4	Y	TP or BB
At3g63170	FAP1	279	0.577	73	0.427		26 (20-45)	1	0	0	7, 7.8	N	TP or SA
At4g00550	DGD2	473	0.452		0.447		(10-40)	0	0	1	11, 4		SA
At4g15440	HPL homolog	384	0.436		0.463		(180-220)	0	0	1	5, 4		x
At5g20410	MGD2	468	0.434		0.441		0	0	0	0	12, 4		x
At1g12230	Transaldolase	427	0.586	47	0.437		0	0	0	1	5, 3		TP
At1g13900	PAP2	656	0.513	19	0.487		24 (614-636)	1	1	6	34, 4.4		TP or TA
At2g19860	HXK2	502	0.472		0.426		21 (5-24)	1	0	2	11, 7	N	SA
At4g29130	HXK1	496	0.495		0.436		22 (5-24)	1	0	2	11, 6.5	N	SA
At1g34430	PDC E2	465	0.539	48	0.46		(390-420)	0	0	0	20, 5	N	TP
At1g44170	ALDH3H1	484	0.437		0.461		0	0	0	2	12, 5	N	TA

At2g34590	PDC E1beta	406	0.586		0.434		0	0	0	1	10, 5		TP
At2g47770	TSPO	196	0.44		0.472		82 (46-68, 88-110)	2	5	6	0		x
At3g01500	beta CA1	347	0.597	47	0.438		0	0	0	0	5, 6	Y	TP or BB
At3g16950	PDC E3	623	0.589	70	0.446		(610-630)	0	0	1	21, 5	N	TP
At3g25860	PDC E2	480	0.592	47	0.462		0	0	0	0	19, 4.5	N	TP
At3g27820	MDAR4	488	0.436		0.463		36 (459-481)	1	1	0	24, 5	Y	TA, SA, BB
At5g17770	CBR	281	0.455		0.44		22 (13-32)	1	1	6	13, 4.7		SA
At5g23190	CYP86B1	559	0.504	17	0.426		38 (36-54, 61-83)	2	2	0	8, 5		TP
At5g25900	KO1/GA3	509	0.545	28	0.435		23 (2-24)	1	2 (4-24, 452-472)	0	10, 3.5		TP, SA, TA
At2g16070	PDV2 (division)	307	0.427		0.438		(220-235)	0	0	2	3, 3		x
At2g20890	THF1/PSB 29 (plasma membrane)	300	0.579	67	0.433		0	0	0	0	1, 4		TP
At3g25690	CHUP1 (actin-dependent movement)	1004	0.466		0.44		0	0	0	1	2, 3.5		SA
At5g53280	PDV1 (division)	272	0.429		0.431		(210-220)	0	0	3	1, 4		TA
At5g58140	PHOT2 (actin-dependent movement)	915	0.455		0.422		(800-900)	0	0	0	15, 4.5		x
At1g27390	TOM20-2 (mito)	210	0.428		0.453		26 (183-200)	1	1	6	0		TA
At3g01280	VDAC1 (mito)	276	0.467		0.462		0	0	0	0	10, 8.5		BB
At3g12580	Hsc70-4 (cytosol)	650	0.43		0.432		0	0	0	0	19, 5	N	x
At3g21865	PEX22 (peroxisome)	283	0.462		0.452		17 (45-62)	1	0	0	8, 5		SA
At3g46030	Histone H2B (nucleus)	145	0.424		0.457		0	0	0	0	2, 4.5		x
At3g63150	MIRO2	643	0.495		0.549	59	22 (612-	1	1	6	14, 4.6		T159L or

	(mito)						634)						TA?
At4g14430	inoyl-CoA isomerase (peroxisome)	240	0.436		0.432		(100-120)	0	0	2	6, 4.3		x
At4g16450	Complex I subunit (mito)	106	0.463		0.463		(35-50, 65-75)	0	2	4	0		x
At4g31780	MGD1 (IEM)	533	0.575	33	0.437		0	0	0	0	12, 5		TP
At4g35000	APX3 (peroxisome)	287	0.431		0.453		22 (258-280)	1	1	6	1, 4		TA
At4g38920	Vacuolar ATPase sub	164	0.523	25	0.432		91 (12-34, 54-76, 93-115, 130-152)	4	4	6	0		TP
At5g02500	HSC70-1 (cytosol/nucleus)	651	0.43		0.428		0	0	0	0	20, 4.5		TP
At5g06290	Prx B (stroma)	273	0.598	90	0.431		(110-130)	0	0	1	9, 5		TP
At5g15090	VDAC3 (mito)	274	0.485		0.465		0	0	0	0	19, 8	Y	BB
At5g27540	EMB2473/ MIRO1 (mito)	648	0.477		0.465		22 (619-641)	1	1	6	11, 6.4	Y	TA
At5g35360	CAC2/BC (IEM)	555	0.571	70	0.428		(330-350)	0	0	0	18, 4.5		TP
At1g09920		192	0.426		0.49		20 (169-188)	1	1	6	1, 2		TA
At1g16000	OEP9	86	0.433		0.443		20 (36-55)	1	1	6	3, 8.3		SA or TA
At1g27300		200	0.427		0.462		18 (107-124)	1	1	6	0		x
At1g64850		162	0.438		0.482		(105-120, 140-155)	0	1	0	0		TA
At1g68680		75	0.46		0.441		37 (21-40, 50-67)	2	2	0	3, 4		SA or TA
At1g70480	DUF220	338	0.486		0.434		0	0	0	0	10, 8.5	Y	BB
At1g80890	OEP9.2	80	0.432		0.436		19 (30-47)	1	1	6	3, 5		SA or TA
At2g06010		188	0.429		0.432		0	0	0	0	13, 6.2	Y	BB

At2g24440		183	0.465		0.428		0	0	0	0	4, 5		x
At2g32240	DUF869	1333	0.427		0.549	82	21 (1308-1330)	1	1	6	2, 4		T159L or TA?
At2g32650	PTAC18 like	139	0.53	32	0.428		0	0	0	0	11, 5.7	Y	TP or BB
At2g44640		451	0.487		0.425		(295-315)	0	0	0	21, 6.4	Y	BB
At3g26740	CCL	141	0.58	41	0.416		0	0	0	0	4, 2.5		TP
At3g49350		539	0.596	55	0.43		0	0	0	0	3, 3		TP
At3g52230	OMP24 homolog	145	0.473		0.426		(90-115)	0	1 (90-110)	5	1, 3		TA
At3g52420	OEP7	64	0.442		0.445		21 (10-32)	1	1	6	0		SA or TA
At3g53560	TPR protein	340	0.575	75	0.429		0	0	1 (69-89)	5	0		TP
At3g63160	OEP6	69	0.442		0.493		20 (24-46)	1	1	6	1, 4		SA or TA
At4g02482	Putative GTPase (OEP15-1)	134	0.473		0.548	44	(1-370)	1	0	1	10, 6.5	N	T159L or TA
At4g15810	NTPase	918	0.471		0.461		(890-918)	0	0	1	17, 4		TA
At4g17170	RAB2	211	0.434		0.434		0	0	0	0	7, 6	Y	BB
At4g27680	NTPase	398	0.433		0.44		19 (13-31)	1	1	6	3, 5		SA
At4g27990	YGGT-B protein	218	0.566	83	0.439		(190-215)	0	2 (136-156, 197-217)	0	1, 2		TP
At5g11560		982	0.442		0.434		33 (949-971)	1	1	0	62, 6.5	N	SA or TA
At5g20520	WAV2	308	0.443		0.434		22 (7-26)	1	2 (5-25, 153-173)	0	9, 6.3	N	SA
At5g21920	YGGT-2	251	0.559	51	0.496		(120-150, 190-210)	0	2	6	1, 3		TP
At5g21990	OEP61-TPR	554	0.457		0.439		21 (531-553)	1	1	6	0		TA
At5g27330		628	0.524	27	0.474		15 (605-627)	1	1	5	0		TA
At5g42070	OEP18	164	0.568	72	0.571	23	(110-130)	0	0	0	6, 3.5		T159L
At5g43070	WPP1	155	0.45		0.434		0	0	0	0	1, 3		x

At5g51020	CRL	269	0.497		0.443		(20-40)	0	1 (19-39)	5	9, 7.6	N	SA
At5g59840	RAB8A-like	216	0.438		0.449		0	0	0	0	8, 6	Y	BB
At5g64816		130	0.429		0.454		(4-24)	0	1 (3-23)	4	2, 4		SA

SA=signal-anchored

TA=tail-anchored

TP=transit-peptide mediated

BB=β-barrel mediated

T159L=TOC159-like pathway

Predictions of transmembrane domains and membrane topology were combined with previous ChloroP analysis (Table 3.1) and collectively were used to categorize each OEP into one of the four canonical pathways and the new Toc-159-like pathway (Table 3.2). Signal-anchored and Tail-anchored proteins were defined, according to Inoue (2015), as those having a single transmembrane helix region at the NT or CT, respectively. Some of the predicted Tail-anchored proteins including Toc33, Toc34, TGD4, OEP9, OEP9.2, OEP6, and OEP6-TPR have been independently predicted to be TA-proteins by Teresinki (2015). The minimum length of amino acids per β-strand for β-barrel proteins was set as 6 residues based on solved and characterized β–barrel membrane proteins (Taylor et al, 2006). My predicted β-barrel OEPs have also been independently predicted in Inoue's list of known or predicted OEPs based on his past publications (http://www.plantsciences.ucdavis.edu/kinoue/OM.htm). The smallest potential β-barrel protein in this list contains 12 predicted β-strands with an average length of 8 amino acids, which meets the minimum requirement for being categorized as using the β-barrel mediated pathway.

There are 18 OEPs which can be categorized into multiple pathways. The two largest subgroups are TP or BB proteins, and SA or TA proteins (Table 3.3). TP or BB proteins contain both a predicted transit peptide at their NT and transmembrane β–strands capable of forming β–barrels. Toc75, the only known TP-mediated OEP up to date, possesses both a TP at its NT and β–strands forming a transmembrane β–barrel. The channel formed by Toc75 provides the preprotein pore of the Toc complex used by preproteins to cross the COM (Figure 3.4). Further experiments are required to determine whether OEPs in this subgroup use a pathway that depends solely on their innate β–barrel structure, their

predicted NT transit peptide or a combination of the two elements.

Table 3.2 Summary of predicted import pathway of 117 OEPs

Pathways	**# of predicted proteins**
Signal Anchor (SA) [a]	15
Tail Anchor (TA) [a]	20
β-barrel (BB) [b]	14
TP-mediated (TP) [C]	21
Toc159-like (T159L) [d]	8
multiple predictions (excluding T159L) [e]	18
Unknown [f]	21

[a] contains one and only transmembrane α-helix of at least 15 aa in length within 50 aa of either the N-terminus (SA) or the C-terminus (TA) predicted by at least one of two analysis programs (TMHMM and TOPCONS)

[b] contains predicted β-strands with at least 6 residues on average by PSIpred, and are predicted to be transmembrane β –strands and able to form transmembrane β-barrel by PRED-TMBB

[C] contains predicted N-terminal transit peptide-like sequence with ChloroP score > 0.5

[d] contains predicted C-terminal reverse transit peptide-like sequence with ChloroP score > 0.5

[e] contains multiple characteristics that can fit into at least two of the five pathways

[f] does not fit into any categories above.

Table 3.3 Breakdown of predicted T159L OEPs which can fit in multiple pathways

Pathways	**# of predicted proteins**
TP or BB	7
SA or TA	6
TP or SA	1
TA or BB	1
TP or TA	1
TA, SA, BB	1
TP, SA, TA	1

SA=signal-anchored
TA=tail-anchored
TP=transit-peptide mediated
BB=β-barrel mediated

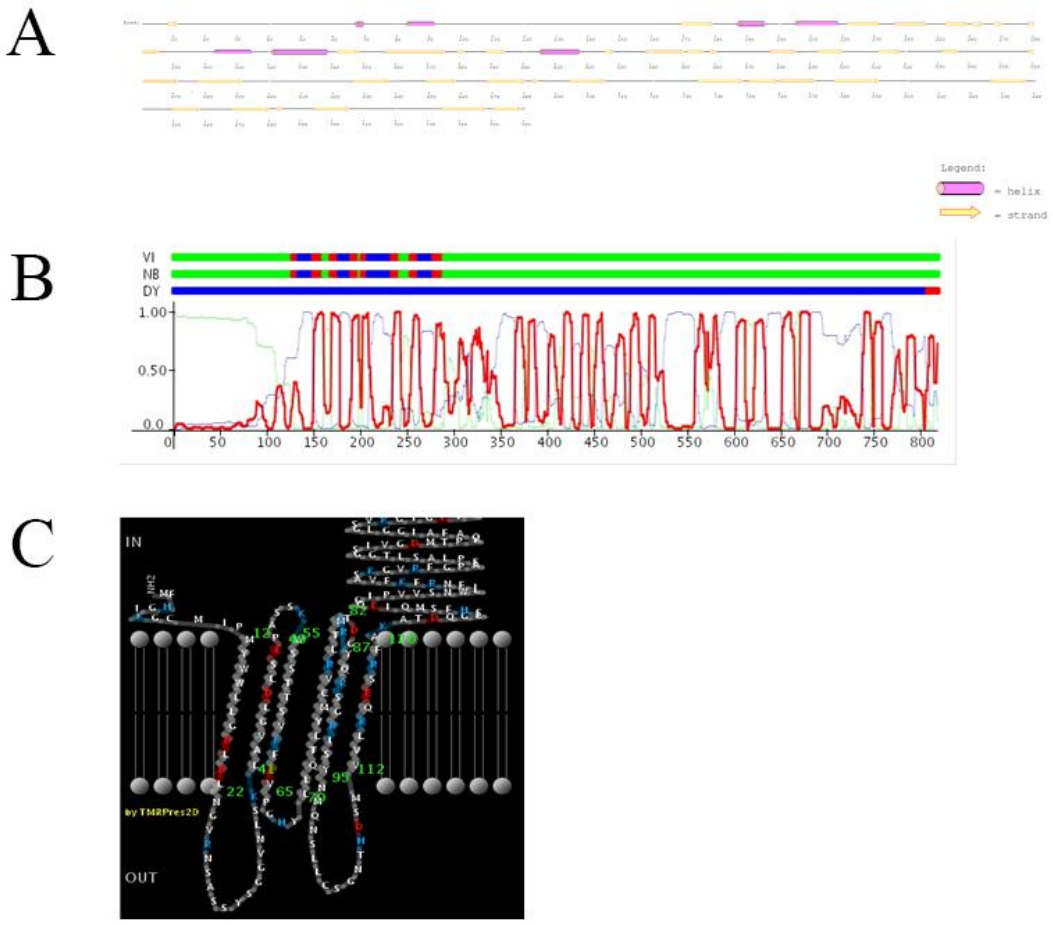

Figure 3.4 transmembrane β-strand prediction of Toc75: Toc75 is a known Toc complex subunit that forms a β-barrel and relies on a TP for its targeting to the COM. It is predicted to contain β-strands by A) PSIpred and B) PRED-TMBB. C) Potential β-barrel formation is visualized by PRED-TMBB. Only 6 β-strands are predicted positively to form the trasnmembrane β-barrel.

Proteins that make up the subgroup of proteins identified as either SA or TA-targeted proteins are less than 100 amino acids in length and their predicted TMH region is located toward the middle of the sequence, which is less than 50 amino acid distance from both ends, thus rendering prediction solely based on TMH location impossible. Both prediction tools also provide transmembrane protein orientation by indicating regions of inside and outside the membrane. OEP9, a known TA protein is among this subgroup. TA proteins would have their TMH anchored to the membrane, thus exposing their NT to the outside (stroma) and CT to the inside (inter membrane space). However, both TMHMM and TOPCONS predict the opposite orientation for OEP9 (Figure 3.5), providing evidence that TMHMM and TOPCONS are in some degree unreliable in distinguishing between SA and TA proteins based on bioinformatic analysis alone for small proteins with less than 100 residues.

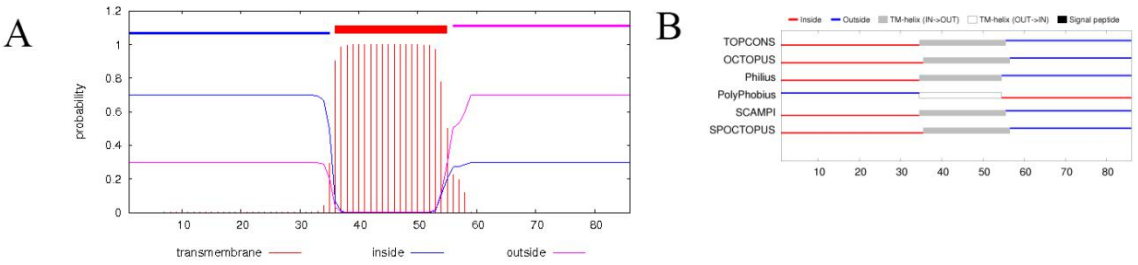

Figure 3.5 TMH prediction graph of OEP9: OEP9, a known TA OEP is predicted to be SA OEP by both A) TMHMM and B) TOPCONS based on orientation around TMH. It is unreliable to distinguish whether OEP9 is signal-anchored or tail-anchored based on the two prediction tools.

3.2 Further analysis of previously identified eight T159L candidates

Secondary structure predictions of the eight OEP candidates that were predicted to use

the T159L pathway by PSIpred are shown in Figure 3.6. The eight candidates are highly variable in size and secondary structure characteristics. However, seven out of eight highlighted candidates possess a predicted α-helix of variable length within the reverse TP-like sequence at the CT (P>0.5) with the exception of pBRP (Figure 3.7). The presence of a predicted common structural feature supports the possibility that this predicted α-helix plays a role in targeting to the OEP and possibly contributed in association between the OEP and chloroplast outer membrane. Among the candidates, MIRO2 and DUF869 were predicted with high possibility of TMH at their C-terminus (Figure 3.8). OEP15-1, OEP16-2, Toc132 and OEP18 were predicted with none or very minimal possibility of TMH on either the N- or C-terminus (Figure 3.9 and 3.10).

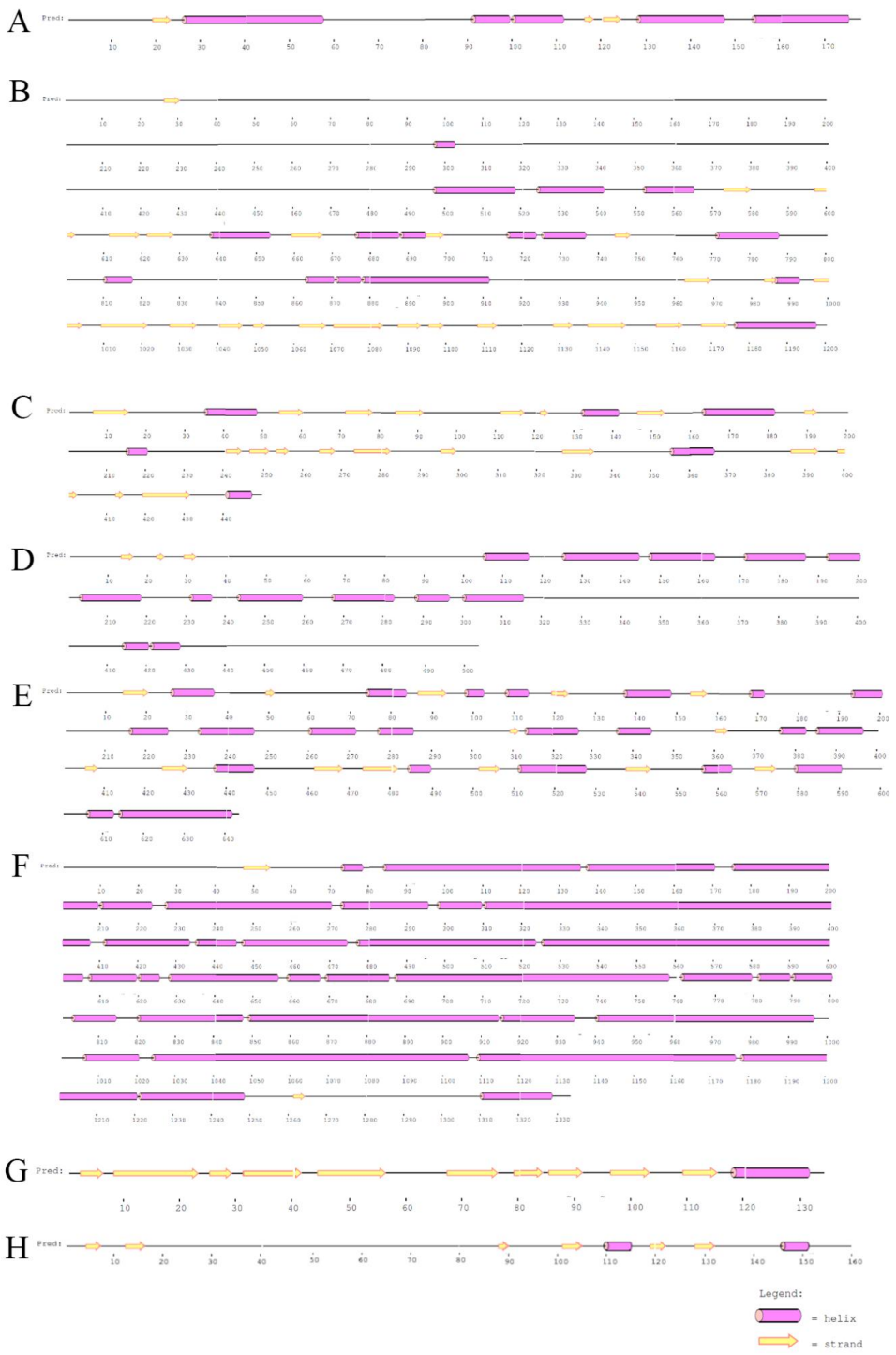

Figure 3.6 Secondary structure prediction of 8 candidate proteins predicted to be targeted to the COM using the T159L pathway. The secondary structures of eight candidate proteins were predicted using PSIpred, A) OEP16-2, B) Toc132, C) E-Tu, D) pBRP, E) MIRO2, F) DUF869, G) Putative GTPase (OEP15-1) and H) OEP18.

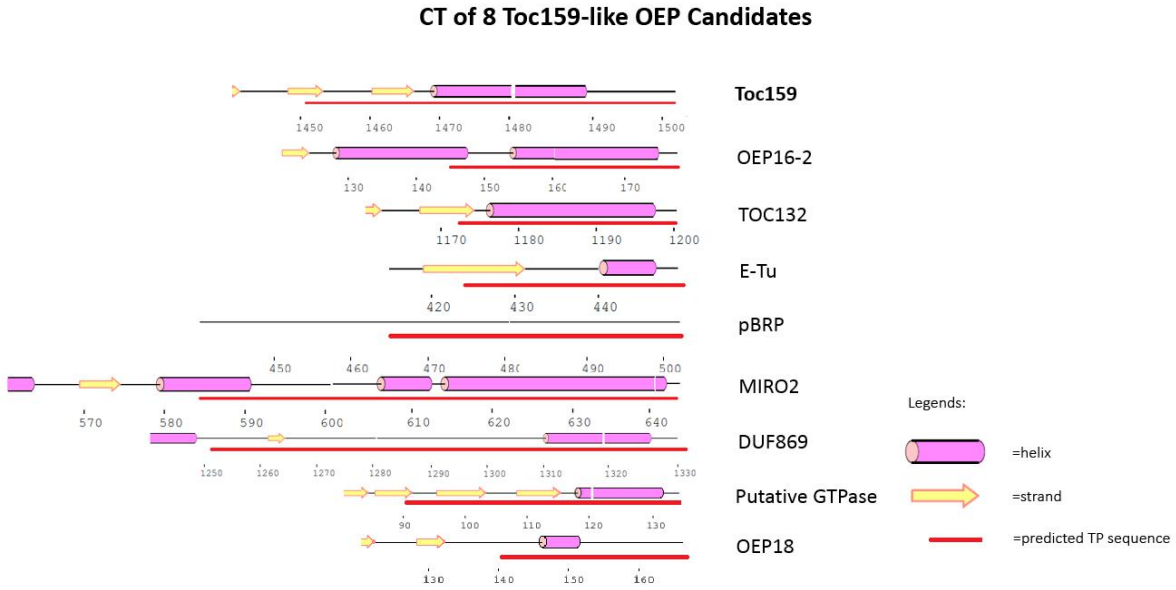

Figure 3.7 Comparison of the C-terminal secondary structure predictions of 8 candidates. The secondary structural predictions generated by PSIpred (Figure 3.6) are compared. The predicted TP-like domain predicted by TOPCONS is indicated by a red line. Putative GTPase has been renamed as OEP15-1 by my colleague Alyssa Overton, OEP15-1 will be used instead in this thesis (personal communication)

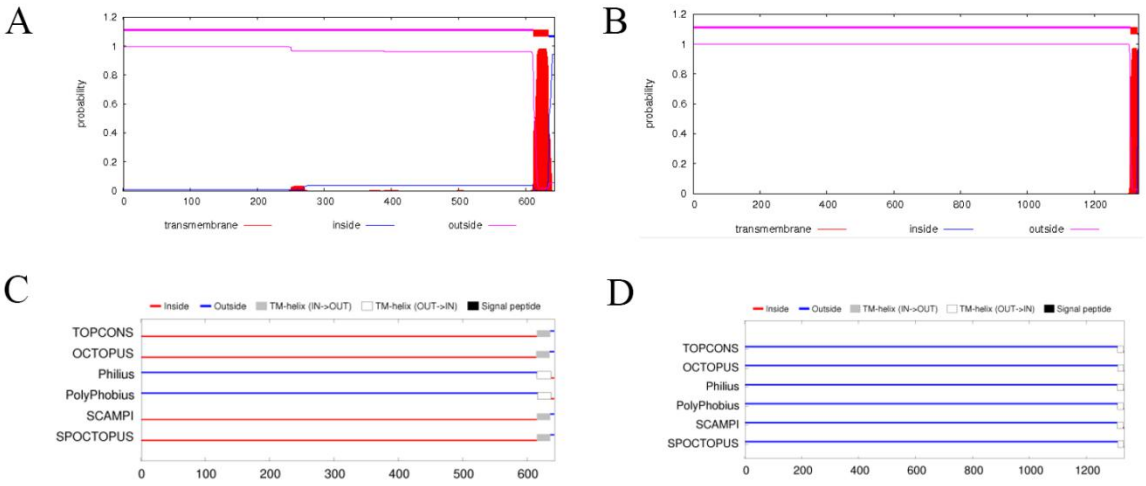

Figure 3.8 TMH prediction of MIRO2 and DUF869 by TMHMM and TOPCONS. TMHMM predicts high possibility of TMH at C-terminus of A) MIRO2 and B) DUF869. All 6 algorithms of TOPCONS predict the same TMH region at the C-terminus of C) MIRO2 and D) DUF869.

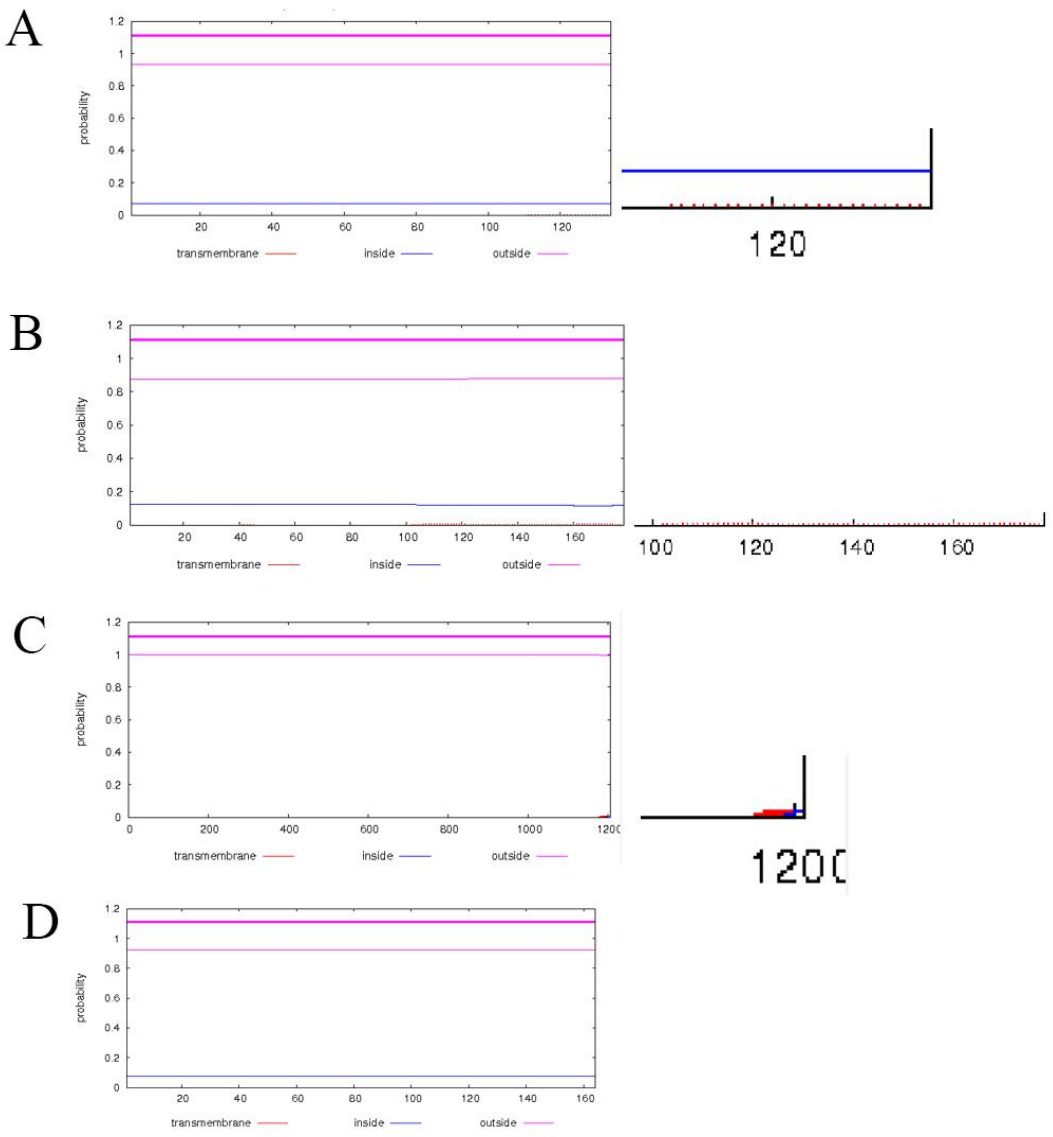

Figure 3.9 TMH predictions for four T159L OEP candidates by TMHMM. TMHMM was used to identify possible transmembrane α-helices in the four proteins predicted to use the Toc159-like pathway for targeting to the COM. TMHMM predicts very minimal possibility of a TMH at the C-terminus of A) OEP15-1, B) OEP16-2, C) Toc132. D) OEP18 shows no potential of TMH at C-terminus at all.

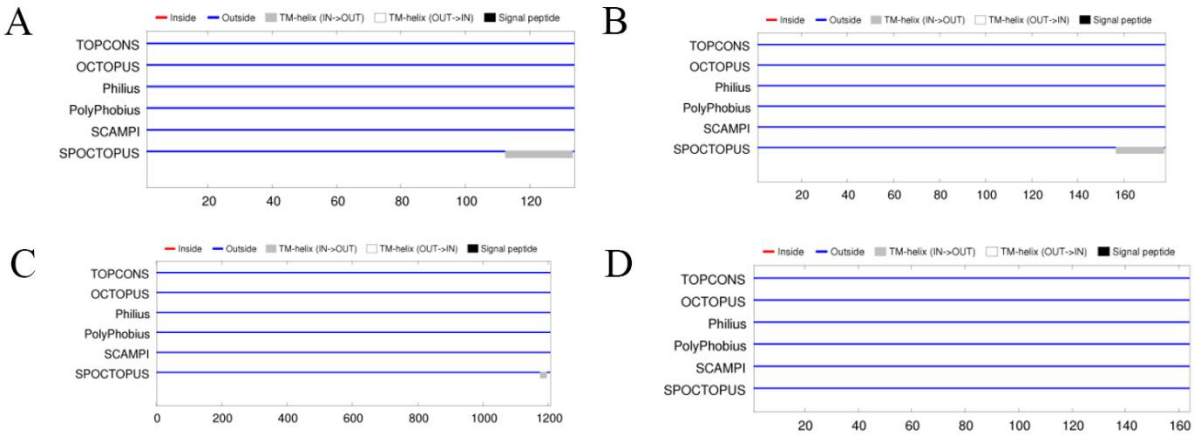

Figure 3.10 TMH prediction graph of 4 T159L OEP candidates by TOPCONS: 1 out of 6 algorithms in TOPCONS predicts possible TMH at C-terminus of A) OEP15-1, B) OEP16-2, C) Toc132. D) OEP18 shows no potential of TMH at C-terminus at all.

3.3 Microscopic observation of *A. thalianas* protoplasts transfected with OEP18 fusion constructs using Epifluorescence microscopy

To examine subcellular localization of transiently expressed OEP18 in *A. thaliana* protoplasts. Four constructs were made previously by Nick Grimberg (2016). Briefly, coding sequences of full-length or truncated OEP18 were fused to the N-terminus or C-terminus of EGFP in the pSAT6C1 or pSAT6N1 vectors (Figure 3.11).

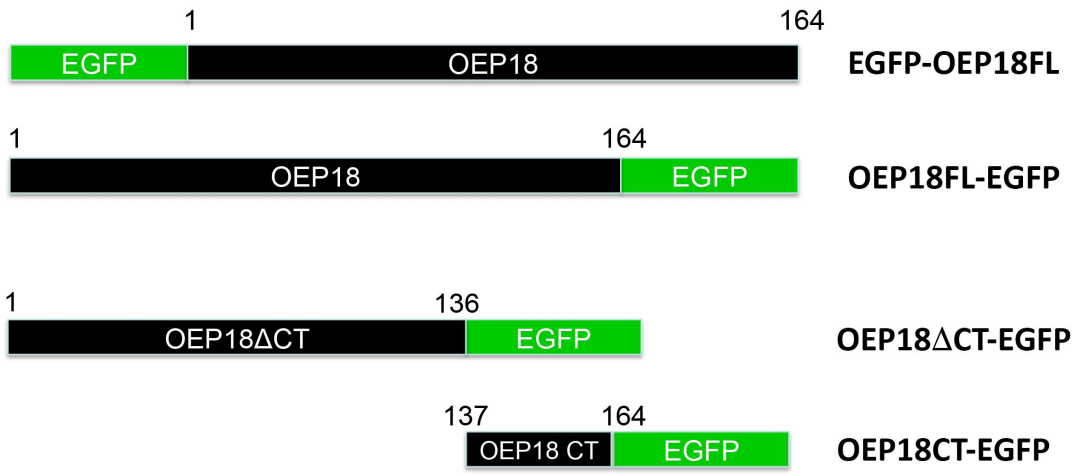

Figure 3.11 schematic maps of full-length and truncated OEP18 fusion constructs: Cleavage site between truncated constructs were determined by the predicted reverse TP sequence by ChloroP. The full-length constructs EGFP-OEP18FL and OEP18FL-EGFP have EGFP attached to the N- and C-terminus of full-length OEP18 respectively. OEP18ΔCT-GFP has GFP attached to the C-terminus of truncated OEP18 from aa 1 to aa 136, with C-terminus containing predicted reverse TP cleaved off. OEP18CT-GFP has GFP attached the C-terminus of truncated OEP18 from aa 137 to 164 with predicted reverse TP.

Arabidopsis protoplasts were isolated and transfected with each fusion construct as previously described in section 2.5. Transfected protoplasts were examined with epifluorescence microscopy 16~18h after transfection (Figure 3.12). Protoplasts transfected with the control vector expressing only GFP showed signal mostly in the nucleus and cytoplasm (Figure 3.12A). Protoplasts transfected with the OEP18-GFP construct shows some degree of chloroplast membrane targeting as observed as forming a thin ring-like structure around chloroplasts (Figure 3.12B). Protoplasts transfected with GFP-OEP18 also shows similar appearance, but with less signal strength and a thinner ring-like structure as compared to OEP18-GFP, indicating a slightly weaker membrane targeting (Figure 3.12C). However, the difference between the control vector and two full-length OEP18 fusion constructs are very minimal and subjective. Thus, their

transient expression requires a more objective and accurate comparison by using Western Blot and Image-J quantification. In comparison, OEP18ΔCT-GFP showed clumping fluorescence signal in punctate-like structures in the cytoplasm, suggesting protein aggregation and impaired specific membrane targeting (Figure 3.12D). On the other hand, the truncated OEP18 construct with only C-terminus shows similar pattern as control vector and full-length constructs, with nuclear and cytoplasmic expression and thin ring-like structure (Figure 3.12E).

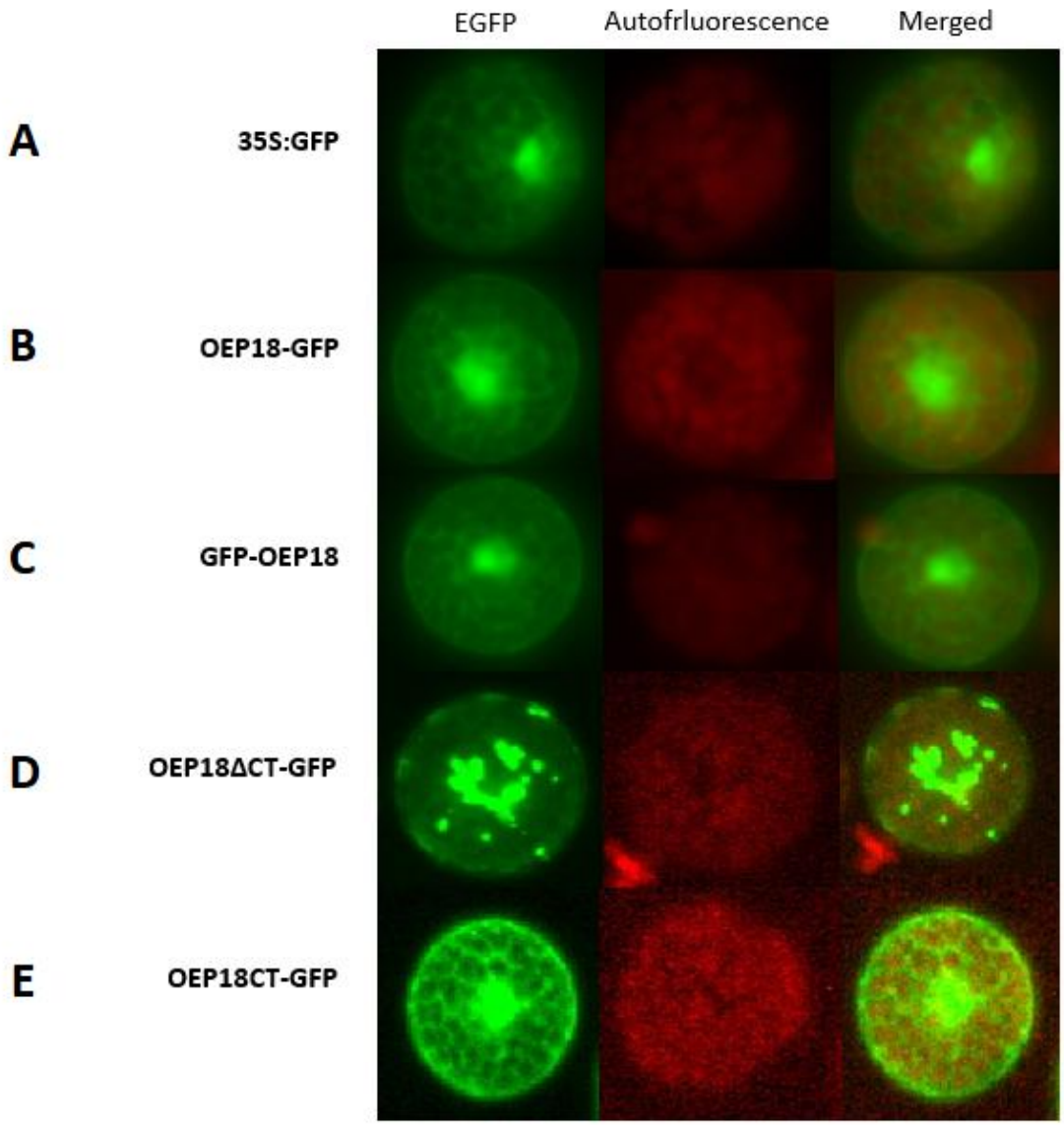

Figure 3.12 transient expression of EGFP fusion proteins with full-length or truncated OEP18 in *Arabidopsis* protoplasts: For each construct representative images of green EGFP, red autofluorescence and merged image of two channels are displayed. OEP18-GFP, GFP-OEP18 and OEP18CT-GFP all have green fluorescent web-like structure around chloroplasts. OEP18ΔCT-GFP has strong punctuate-like structure within cytoplasm, indicating impaired targeting to chloroplast, and protein aggregation caused by misfolding. Panel order: A) 35S:GFP; B) OEP18-GFP; C) GFP-OEP18; D) OEP18ΔCT-GFP; E) OEP18CT-GFP.

3.4 Detection and quantification of OEP18 fusion proteins in fractionated protoplasts using Western Blot analysis

To further determine the subcellular localization of the four OE18 fusion constructs, protoplasts were fractionated into total soluble and total insoluble (membrane) fractions containing stroma and chloroplast membranes, respectively, and protein signals were then examined and semi-quantified using Western blot analysis and ImageJ. Transfected protoplasts were fractionated using methods described in section 2.6. The transfection rate was estimated via epifluorescence microscopy, protoplasts with transfection rate of at least 60% were used in Western blot analysis and signal quantification. The predicted/expected molecular masses for OEP18-GFP, GFP-OEP18, OEP18ΔCT-GFP and OEP18CT-GFP were 45kDa, 45kDa, 37kDa, and 30kDa, respectively. All corresponding bands were detected on each membrane, indicating intact fusion proteins were expressed in every assay for both protoplast-only rupture and total rupture method (Figure 3.13). Two signal quantification approaches were used for western blot images collected from both rupture methods: the first approach used all the band signal in one lane including both intact protein and degraded protein and degraded signals (Figure 3.14 and 3.16); the second approach used only the signal at corresponding molecular weight of the intact target protein (Figure 3.15 and 3.17). Paired t-test was used to test whether the two full-length constructs shows significant difference in signal distribution between different rupture methods and quantification methods (Table 3.4).

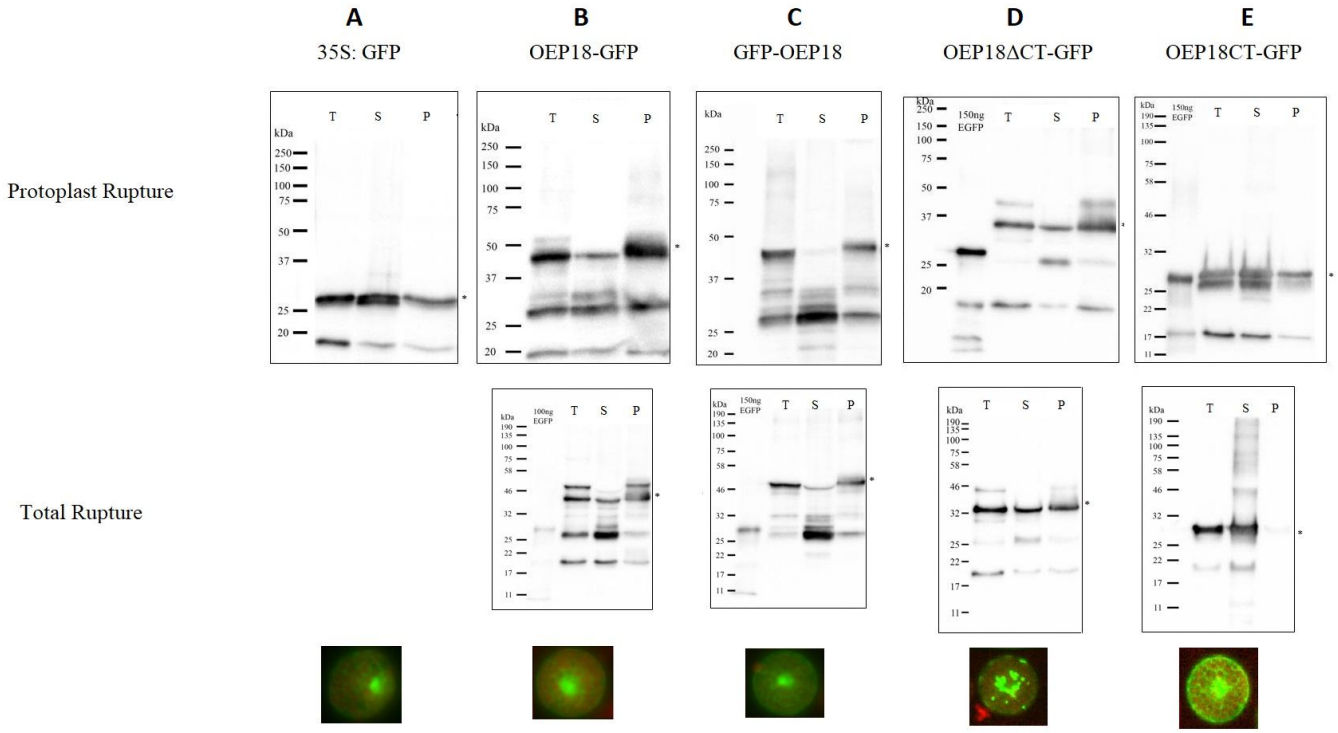

Figure 3.13 Representative Western blot images of transfected protoplasts obtained from protoplast rupture and total rupture methods. 35S:GFP control vector and 4 full-length or truncated OEP18+GFP fusion constructs expressed in *Arabidopsis* protoplasts fractionated with protoplast rupture or total rupture method. Each fraction was separated by SDS-PAGE, followed by immunoblotting with an anti-EGFP polyclonal antibody. Numbers to the left indicate the position of the marker protein ladder in kilodaltons. Recombinant EGFP was used as a control on each immunoblot. T represents ¼ of total protein from each replicate. S and P represent soluble fraction and insoluble pellet fractions respectively. Each replicate has 60,000 protoplasts with at least 60% transfection rate. * Marks the location of intact target protein. The bottom images are the representative merged microscopy images of each construct transiently expressed in protoplasts from Figure 3.12. Panel order: A) 35S:GFP; B) OEP18-GFP; C) GFP-OEP18; D) OEP18ΔCT-GFP; E) OEP18CT-GFP.

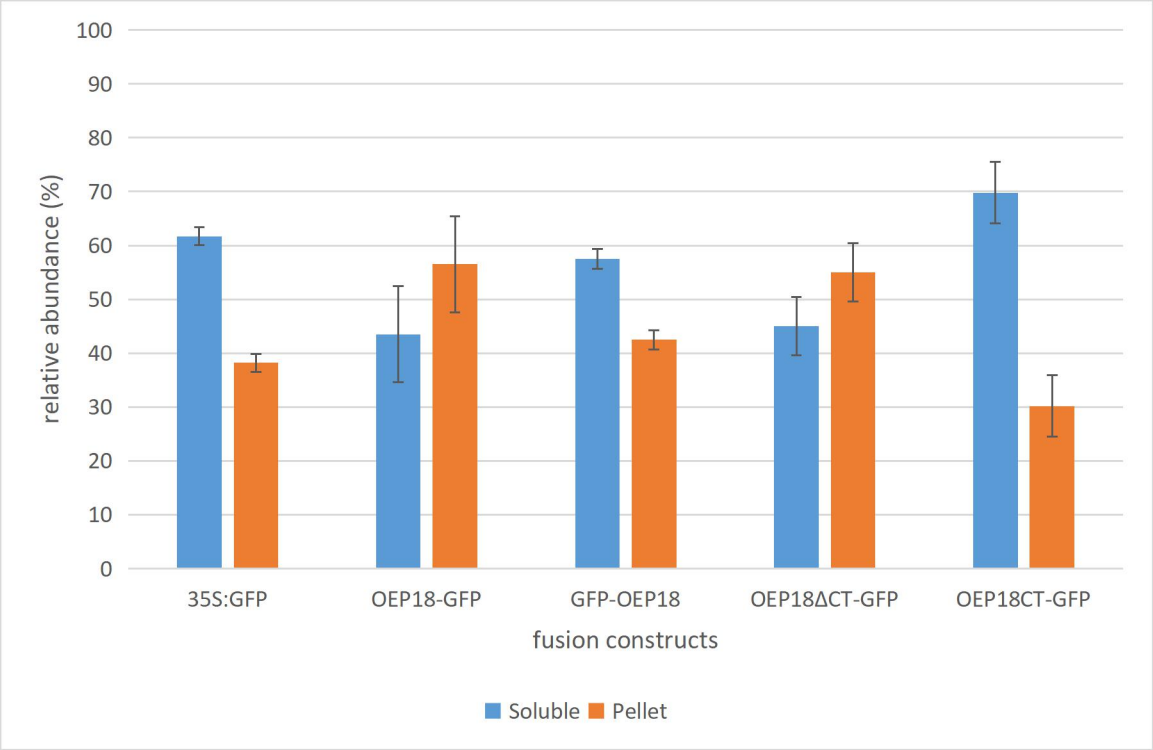

Figure 3.14 relative abundance of all signals in soluble and insoluble fractions for the four OEP18 fusion constructs obtained from protoplast rupture method: 35S:GFP control vector and 4 full-length or truncated OEP18+GFP fusion constructs expressed in *Arabidopsis* protoplasts fractionated with only protoplasts ruptured. The western blot images of each construct were selected for clear images and bands. The sample size was set to 8 minimum. All signals including intact protein and degraded protein for both soluble (blue) and insoluble membrane (orange) fractions in each Western Blot images was collected and quantified using ImageJ. The average relative abundance for each fraction = total signal strength of each fraction / total signal strength of both fractions combined. The error bar on each column is the standard deviation determined from each sample group.

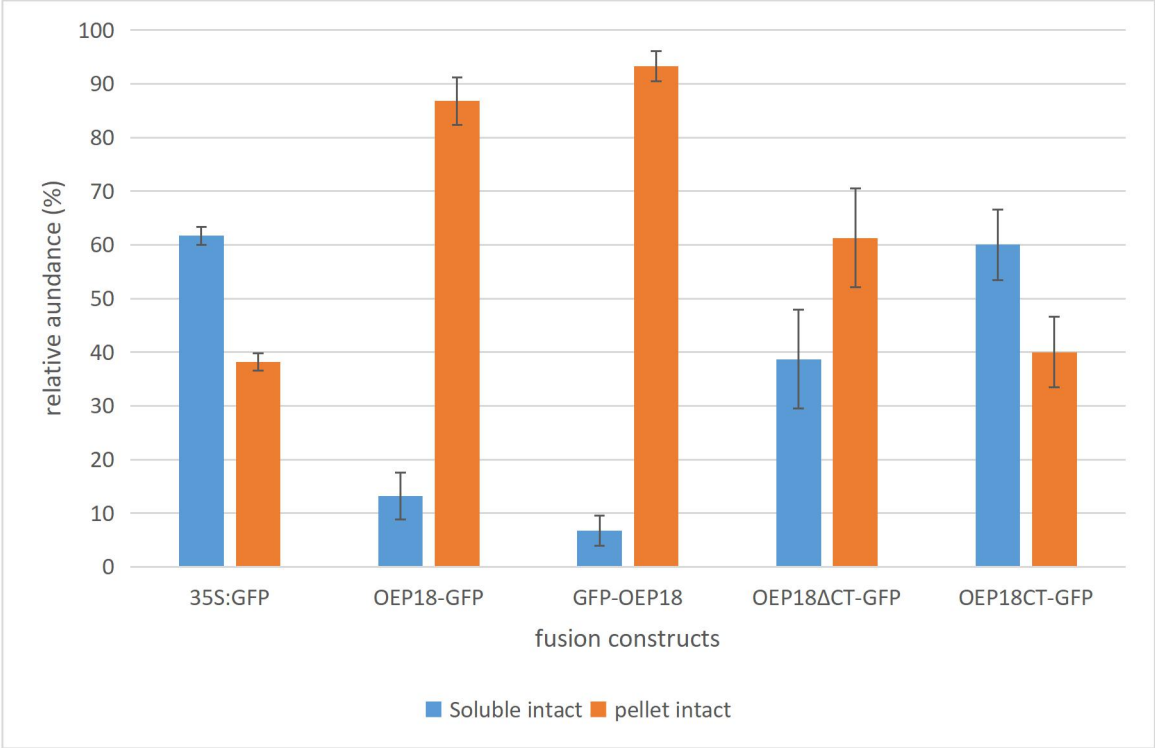

Figure 3.15 relative abundance of intact protein signal in soluble and insoluble fractions for the four OEP18 fusion constructs obtained from protoplast rupture method: 35S:GFP control vector and 4 full-length or truncated OEP18+GFP fusion constructs expressed in *Arabidopsis* protoplasts fractionated with only protoplasts ruptured. The western blot images of each construct were selected for clear images and bands. The sample size was set to 8 minimum. Only intact protein signal for both soluble (blue) and insoluble membrane (orange) fractions in each Western Blot images were collected and quantified using ImageJ. The average relative abundance for each fraction = intact protein signal strength of each fraction / total intact protein signal strength of both fractions combined. The error bar on each column is the standard deviation determined from each sample group.

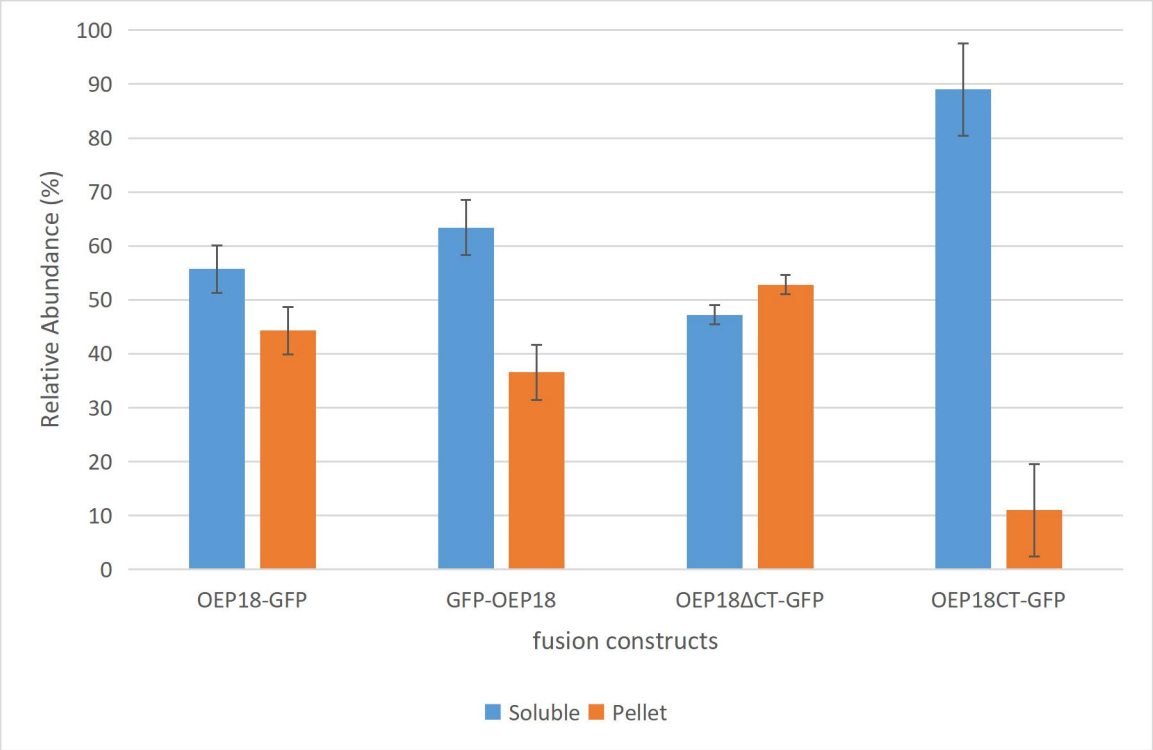

Figure 3.16 relative abundance of all signal for the four OEP18 fusion constructs obtained from total rupture method: The 4 full-length or truncated OEP18+GFP fusion constructs expressed in *Arabidopsis* protoplasts fractionated with the total rupture method. The western blot images of each construct were selected for clear images and bands. The sample size was set to 8 minimum. All signals including intact protein and degraded protein for both soluble (blue) and insoluble membrane (orange) fractions in each Western Blot images was collected and quantified using ImageJ. The average relative abundance for each fraction = total signal strength of each fraction / total signal strength of both fractions combined. The error bar on each column is the standard deviation determined from each sample group.

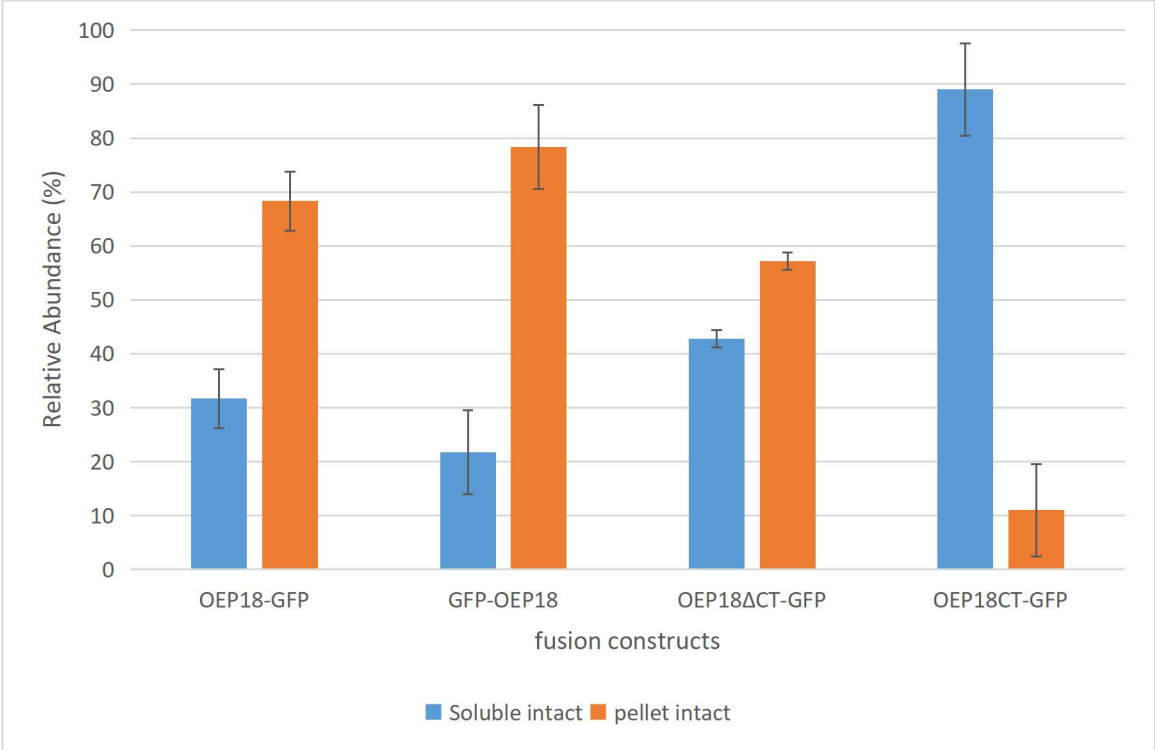

Figure 3.17 relative abundance of intact protein signal for the four OEP18 fusion constructs obtained from total rupture method: The 4 full-length or truncated OEP18+GFP fusion constructs expressed in *Arabidopsis* protoplasts fractionated with the total rupture method. The western blot images of each construct were selected for clear images and bands. The sample size was set to 8 minimum. Only intact protein signal for both soluble (blue) and insoluble membrane (orange) fractions in each Western Blot images were collected and quantified using ImageJ. The average relative abundance for each fraction = intact protein signal strength of each fraction / total intact protein signal signal strength of both fractions combined. The error bar on each column is the standard deviation determined from each sample group.

Table 3.4 Paired t-test between two full length constructs under different rupture methods and signal quantification methods

rupture method	quantification method	constructs	Pellet signal abundance (%)	standard deviation	t-value	p-value (n=8,a=0.05)	Significant difference (Y/N)
protoplast rupture	total signal	OEP18-GFP	56.5	8.9	2.75	1.895	Y
		GFP-OEP18	42.5	1.8			
	intact signal only	OEP18-GFP	86.8	4.4	0.81		N
		GFP-OEP18	93.3	2.81			
Total rupture	total signal	OEP18-GFP	44.3	4.38	1.94		Y
		GFP-OEP18	36.6	5.11			
	intact signal only	OEP18-GFP	68.3	5.45	-1.33		N
		GFP-OEP18	78.3	7.8			

3.5 OEP18 recombinant protein design

In order to study the secondary structures of OEP18, a recombinant construct was made by incorporating a cDNA encoding full length OEP18 (Figure 2.2) into the pET28a expression vector such that a recombinant protein would be produced with a 6X His-tag fused to the C-terminus. The molecular weight of the corresponding recombinant protein is predicted to be 22 kDa. The construct has been made and validated as explained in section 2.9.

3.6 Protein induction profile

The sequence-verified construct was heat-shock transformed into BL21 Codon+ *E. coli* cells. Overexpression of recombinant protein was achieved using 1mM IPTG in 300 mL uninduced culture. After 16-hour induction, a prominent band of about 22 kDa was visible (tON) compared to the uninduced sample (t0) (Figure 3.18). The majority of the overexpressed ~22 kDa protein is located in the soluble fraction (S), as compared to the insoluble fraction (P). Based on the molecular weight of the prominent band, attempts were made to purify the recombinant protein using IMAC with 4 M urea after ruptured with French Press.

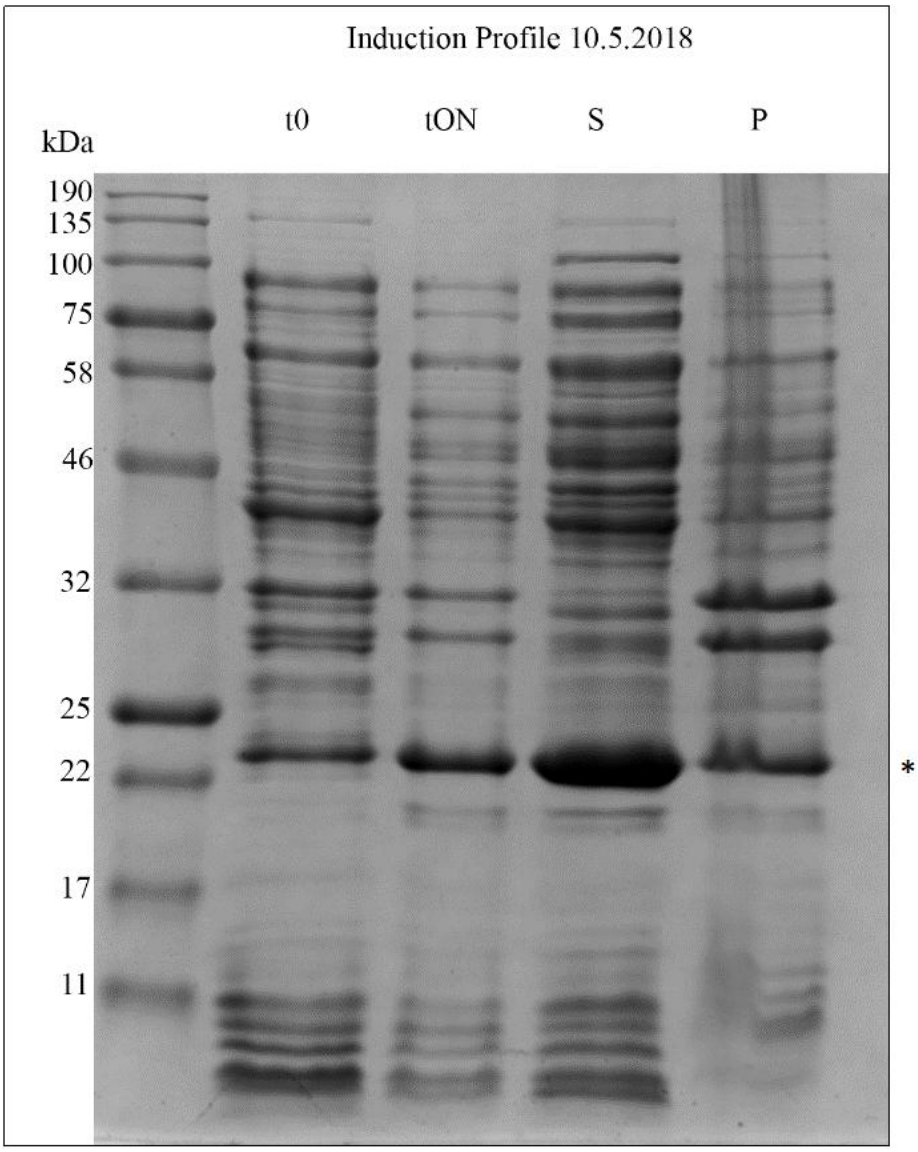

Figure 3.18 Protein profile of overexpressed recombinant OEP18 Samples collected from uninduced culture (t0), induced culture (tON), soluble fraction (S) and pellet fraction (P) after cell rupture by French Press are separated by SDS-PAGE and stained with Coomassie. Each lane contains approximately same cell mass. Numbers to the left indicate the position of the marker protein ladder in kilodaltons. * Marks the location of intact target protein.

3.7 IMAC purification and OEP18 verification by Western Blot

The urea-solubilized recombinant OEP18 after cell rupture was purified from the soluble extract using Immobilized metal affinity chromatography (IMAC). The purification process is explained in section 2.10. OEP18 with 6x His tag protein or OEP18-His was purified with Ni-charged resin and be eluted with 100 mM imidazole in the elution buffer (Figure 3.19). The presence of recombinant OEP18-His within elution fraction with 100 mM imidazole (E1) was verified by Western blot analysis, probed with a primary polyclonal antibody raised in mouse against 6xHis protein (1:3000) and anti-mouse anti-igG secondary antibody (1:10,000) (Figure 3.20).

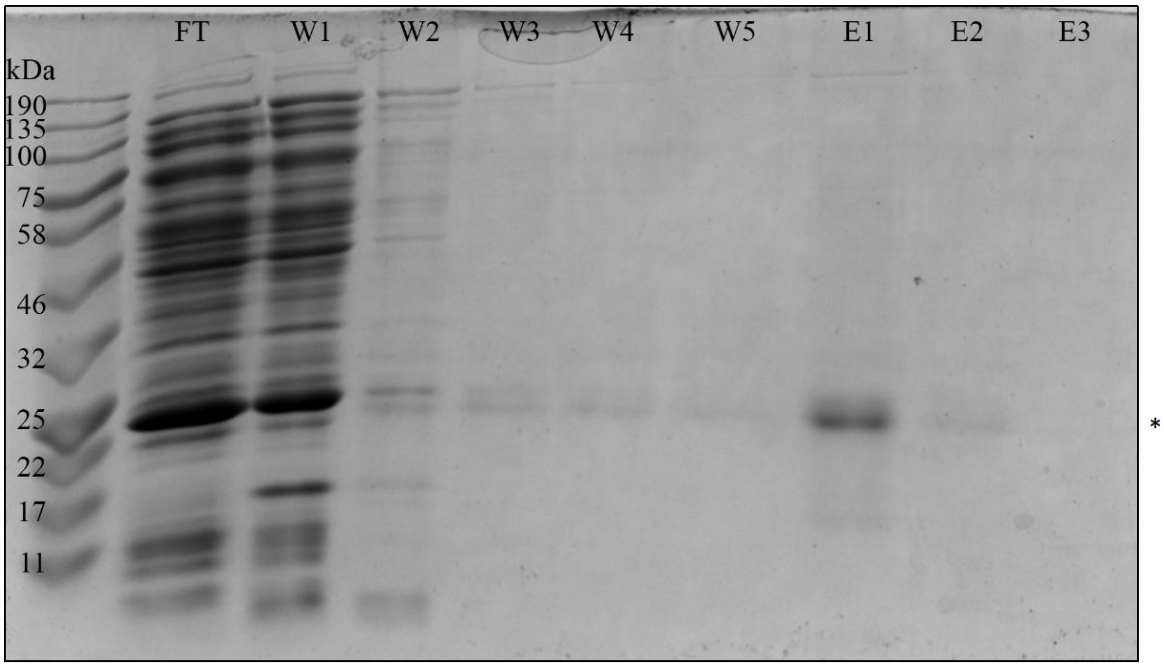

Figure 3.19 Protein profile of each fraction in each step of IMAC purification
Recombinant OEP18-His was purified from Soluble fraction after cell rupture by IMAC. After 5 washing steps (W1-W5) with 1 mL washing buffer each, OEP18-His with 21 kDa was eluted with elution solution with 100, 200 and 500 mM imidazole each (E1-E3). Samples collected from flowthrough (FT), W1-W5 and E1-E3 were separated on SDS-PAGE, the gel was stained with Coomassie. Numbers to the left indicate the position of the marker protein ladder in kilodaltons. * Marks the location of intact target protein.

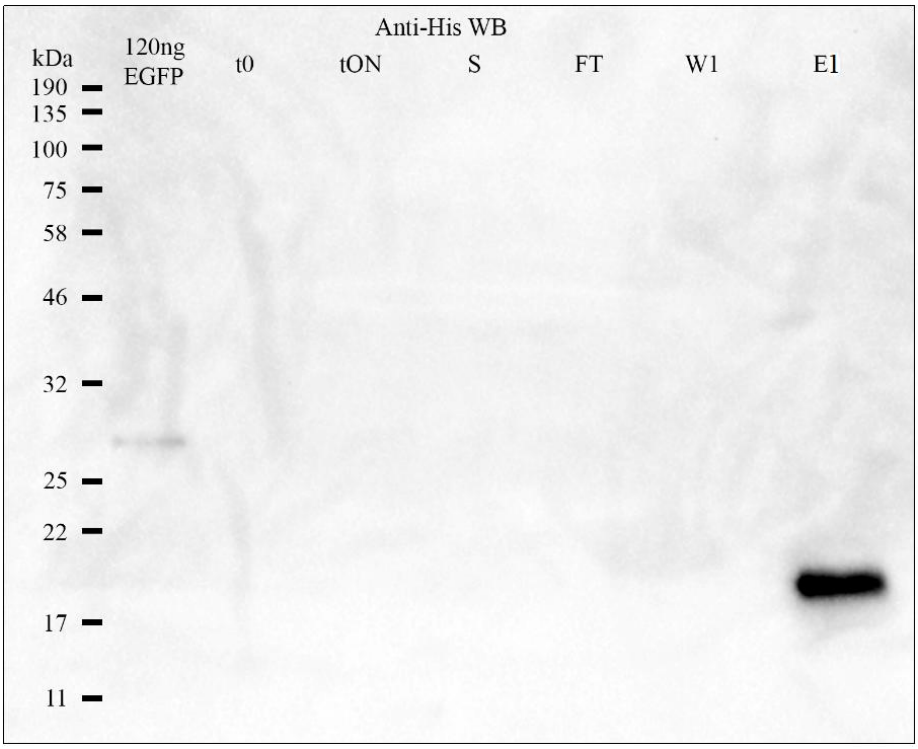

Figure 3.20 Western blot image of protein profiles in IMAC purification and in protein induction with Anti-His antibody. Several samples were selected to run together with SDS-PAGE in order to validate OEP18-His is purified in E1 fraction in IMAC purification: uninduced culture (t0), induced culture (tON) and soluble fraction (S) from protein induction; and flowthrough (FT), wash fraction (W1) and elution fraction (E1) from IMAC purification. Strong anti-His signal is only present in E1 fraction, which was recognized by anti-His primary antibody and visualized by Anti-mouse anti-IgG secondary antibody. EGFP lane was used as control. Numbers to the left indicate the position of the marker protein ladder in kilodaltons.

3.8 Circular Dichroism analysis of OEP18

IMAC elution fractions which showed strong signal of OEP18 were combined and dialyzed to exchange the protein from elution buffer [50 mM Tris-HCl (pH 7.5), 100 mM NaCl and 100 mM imidazole, 4 M urea] to CD compatible buffer [100 mM NaF, 50 mM Tris-HCl (pH 7.5)]. This is necessary to completely remove urea and imidazole, which would interfere with CD, and was accomplished using 6 steps of dialysis to gradually reduce the Urea concentration in an attempt to avoid protein precipitation. After dialysis,

protein concentration was tested using a Bradford Assay and was determined to be 3.5 M (77 ug/mL). The sample was also examined by SDS-PAGE to confirm that the protein remained soluble after dialysis (Figure 3.21).

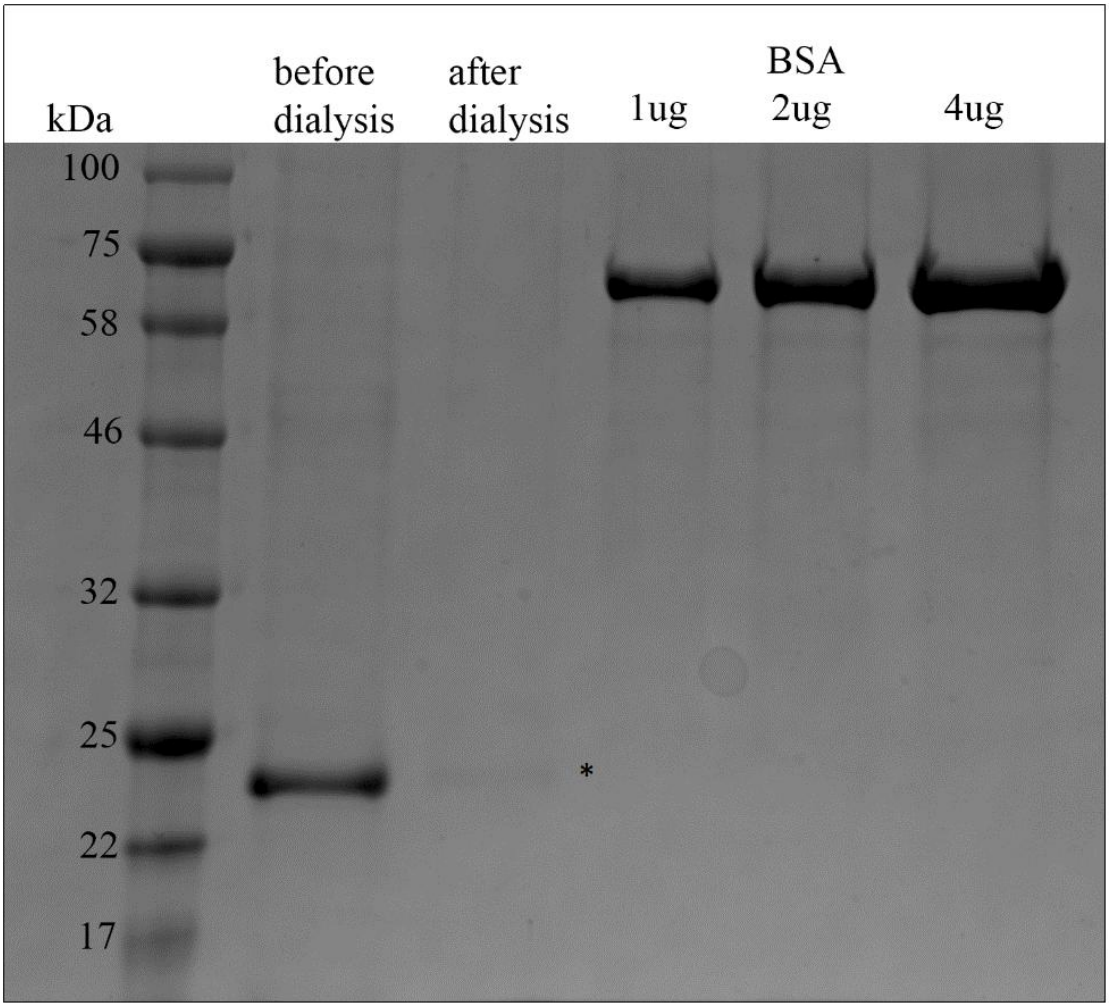

Figure 3.21 Comparison of protein profile of OEP18-His before and after dialysis. 5 uL of sample after dialysis was run with 10 uL of elution sample before dialysis with SDS-PAGE. The gel is stained by Coomassie. 1 ug, 2 ug and 4 ug of BSA were also used for comparing band size and approximate protein concentration. Numbers to the left indicate the position of the marker protein ladder in kilodaltons. * marks the location of intact target protein. The protein signal in after-dialysis lane is faintly present.

The secondary structure content of the recombinant OEP18 protein was examined using Circular Dichroism spectroscopy immediately (i.e. without freezing) following dialysis

and determination of protein presence and concentration. All aspect of Circular Dichroism, including buffer preparation, experiments and analysis was performed by Michael Fish at Wilfrid Laurier University. Data generated from CD was analyzed using DichroWeb program (dichroweb.cryst.bbk.ac.uk). In the presence of 100 mM NaF, liposome, or 50% TFE, OEP18 adopted conformation with some degree of alpha helicity as indicated by the minima around 208 and 222 nm (Figure 3.22). The extent of helical structure induced by NaF and liposomes were less than that induced by 50% TFE. Four independent protein samples were prepared starting from protein induction for consistency and reliability of protein samples. The average structural content of the protein in each buffer is shown in Table 3.5. With original CD buffer, OEP18-His showed 10% α-helix and 36% β-strand. OEP180-His shows approximately the same secondary structure composition in liposome buffer. In 50% TFE buffer, OEP18-His shows increased percentage of α-helix.

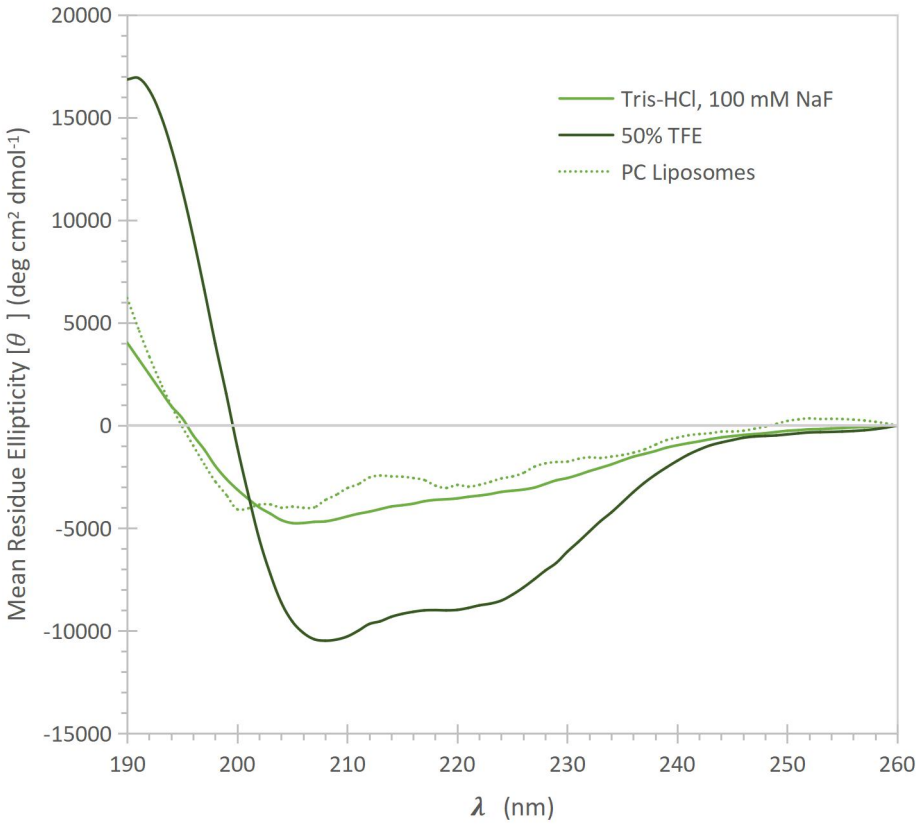

Figure 3.22 Representative Circular Dichroism Curve of OEP18. Protein sample with different buffer content was ran against each blank. Three buffers include: original CD buffer (light green), 50% TFE (dark green) and liposome (green dash). Credit: Michael Fish, Wilfrid Laurier University.

Table 3.5 structural analysis of CD curve

DichroWeb Analysis (K2D Method)			
	Tris-HCl NaF	TFE	PC Lipo
Alpha Helix	10%	29%	11%
Beta Sheet	36%	13%	36%
Random Coil	55%	59%	54%

Credit: Michael Fish, Wilfrid Laurier University

4. Discussion

4.1 Prediction of import pathways of all 117 OEPs using several bioinformatic tools

Bioinformatic analysis of all 117 known or predicted *Arabidopsis* chloroplast OEPs provided important insight for this study as a large number of OEPs currently remain uncharacterized. ChloroP is designed to predict chloroplast transit peptides at the N-terminus of proteins, and in the current study it has been used in a novel way - as first demonstrated by Lung and Chuong (2012) for Toc159 - to predict the presence of reverse transit peptide-like sequences in OEP18 (Lung et al, 2014). When applied to all 117 OEPs, 8 candidate OEPs were identified that might also contain TP-like sequences at their C-terminus that could be involved in their targeting (Nick Grimberg, 2016). One goal of this study was to use additional bioinformatic tools including TMHMM, TOPCONS, PSIpred and PRED-TMBB, in order to predict potential import pathways used by as many OEPs as possible based on their predicted transmembrane domain(s), possible secondary structures and potential β-barrel formation. I also aimed to rule out any T159L candidates which are known or strongly predicted to use other import pathways. As a result of the analysis, 70 OEPs were found to possess characteristics of only one import pathway, 18 OEPs were found to possess characteristics of multiple known import pathways, and 21 OEPs remain without a predicted import mechanism (Table 3.2).

The specificity of targeting and localization can be determined by different domains of a protein. To ensure reliability of bioinformatic analysis of this study, multiple bioinformatic tools were used. To determine import pathways of those proteins predicted to possess characteristics of multiple targeting pathways, it is essential to study their

import mechanisms experimentally. That there were 21 OEPs that could not be clearly assigned to one of the known import pathways may be a result of the unreliability of the bioinformatic tools or analysis that was done, or might indicate there are still novel import pathways used by some proteins that possibly utilize unconventional signals, membrane topologies or involve protein-protein interactions with other unknown receptors.

DUF869 and MIRO2 were two of the 8 previously identified T159L OEP candidates that were strongly predicted to be TA proteins in the current study, based on having strongly predicted transmembrane C-terminal helix domains. DUF869 and MIRO2 are strongly predicted to have a transmembrane α-helix (P=1) near their CT by TMHMM and all 6 algorithms of TOPCONS (Figure 3.7), which overlaps in both cases with the predicted α-helix within the TP-like sequence. Such structural features are consistent with TA-proteins. Indeed, the likelihood that DUF869 and MIRO2 are Tail-anchored proteins is further strengthened by Marty et al. (2014) who predicted those two proteins to be among the TA-proteins of COM. MIRO2 was further identified as dual-localized in mitochondria as a TA protein (Costello, 2017). Thus, DUF869 and MIRO2 can be eliminated as candidates using the Toc159-like pathway, and included in the list of proteins that are likely belonging to the TA-protein category.

The OEP16-2, OEP132 and OEP15-1 also possess very minimal probability by TMHMM (Figure 3.9) and only detected by one of six algorithms of TOPCONS (Figure 3.10). The amino acid sequence of OEP15-1 was analyzed by Teresinki (2015) due to detection by TOPCONS. The amino acid composition of its predicted TMH domain does not contain any positively charged amino acid including R and K residues, which is distinct from

typical known TA-proteins consisting of ~18% of positively charged amino acid. More importantly, this predicted transmembrane α-helix by TOPCONS overlaps with the α-helix in predicted reverse TP-sequence by PSIpred and it contains 11.1% of S and T residues. As high percentage of S and T residues is one of the few common features among TP sequences (Zhang and Glaser, 2002), this supports that the CT of OEP15-1 could still utilize the T159L over TA pathway; a possibility being examined by Alyssa Overton, a member of the Chuong and Smith labs. Thus, it is possible that α-helix at CT of OEP16-2, OEP132 and OEP15-1 could possess transmembrane properties, but it is highly unlikely these three proteins utilize the TA pathway judging from the TMH predictions. The target protein of the current study, OEP18, does not contain a predicted TM α-helix at either the NT or CT, and only contains 3.5 average residues per β-strand. The bioinformatic analysis shows that OEP18 does not have any predicted TMH region nor does it meet the minimum criteria for β-barrel formation (Figure 3.9). With the highest ChloroP score for reverse TP prediction at the CT, OEP18 has the highest possibility of utilizing the Toc159-like pathway among the T159L OEP candidates.

4.2 Limitation of predictions of the collective bioinformatic tools

Several limitations of the bioinformatic analysis were identified in the current study. First, prediction algorithms are not completely accurate. For example, the TMH region at the CT of Toc33, a well characterized TA protein, is only predicted with minimum probability by TMHMM, and is only recognized as a TMD by algorithms for 2 out of 6 prediction programs used by TOPCONS (Figure 3.1C). Although two prediction tools and seven algorithms were used in total for predicting transmembrane helix domains, it is

difficult to set the optimal threshold and decide which prediction to accept when algorithms provide contradicting predictions. Another example is the prediction of Toc75 by PRED-TMBB (Figure 3.4C). Though Toc75 is predicted correctly to be a β-barrel protein, its predicted number of β-strands forming the β-barrel is 6 instead of 16 which was well established in previous research (Hinnah et al, 2002). This raises concerns of the reliability of topologies predicted by PRED-TMBB for the OEPs. Secondly, it is difficult to predict whether an OEP is an SA or TA protein when the protein contains less than 100 amino acids and the transmembrane helix domain is within 50 amino acids of both the N- and C-terminus. One example is OEP9, which is a known TA OEP having the transmembrane helix region in the middle of its amino acid sequence (Figure 3.5). It is currently impossible to determine if such an OEP is TA or SA using a bioinformatic approach. Third, it is difficult to predict the more likely import pathway used by a protein that is predicted to possess characteristics of multiple pathways. One example is the subgroup of OEPs predicted to potentially use the chloroplast β-barrel self-insertion pathway and the N-terminal TP pathway. This subgroup of OEPs is the largest in the subgroups that are predicted for multiple import pathways, including: Toc75-III, OEP24, OEP37, PAP/FBN3a, beta CA1, PTAC18 like, OEP16-1 and OEP80/Toc75-V. In vitro experiments with OEP24 and OEP37 do not reveal a change in size following import, suggesting an N-terminal TP signal is either not used or is not cleaved, suggesting OEP24 and OEP37 utilize the β-barrel self-insertion pathway (Jones & Rapaport, 2017; Kim et al, 2019). There are also false negative within the categorized unknown group, several OEPs with known import pathways were overlooked by the bioinformatic tools. For example, PDV2, a known TA protein responsible for cell division (Glynn et al, 2008), was

predicted to have possible TMH near its C-terminus further than 100 aa by both TMHMM and TOPCONS. Further investigation is required for determining the import mechanism used by the OEPs predicted to potentially use multiple import pathways.

4.3 OEP18 is targeted to plastids and has higher targeting efficiency when GFP is fused to the C-terminus

The subcellular localization of full-length and truncated fusion constructs of OEP18 within *A. thaliana* protoplasts were examined using a biochemical approach. Transfected protoplasts were fractionated into soluble (cytosol and stroma) and insoluble (membrane and organelle) fractions, and the presence and relative amount of GFP-tagged full length or truncated OEP18 was determined using Western Blot analysis and ImageJ quantification. The protoplasts transfected with EGFP fusion constructs were also examined using confocal microscopy.

The purpose of using two different full-length fusion constructs (OEP18-GFP and GFP-OEP18) was to determine whether fusing GFP to the N- or C-terminus of OEP18 affected its chloroplast targeting efficiency. From the microscopy images of recombinant protein transiently expressed in protoplast, Both full length constructs shows the fluorescent ring-like structure around the chloroplast, and OEP18-GFP had a slightly thicker ring (Figure 3.12B&C). The pellet fraction for OEP18-GFP also shows a thicker band than GFP-OEP18 at target molecular weight (Figure 3.13B&C). Based on observation made from microscopy images and representative Western Blot images, it is likely that GFP tagged to the N-terminus of OEP18 has slightly less targeting efficiency than tagged to the C-terminus. Thus, for the truncated fusion constructs in this study, GFP was chosen to fuse

to the C-terminus of truncated OEP18.

Both full-length constructs showed a higher signal abundance in the insoluble membrane fraction than the negative control 35S: GFP (Figure 3.14), indicating OEP18 can target to chloroplast efficiently regardless of tagging GFP on either N- or C-terminus. OEP18-GFP showed more signal abundance than GFP-OEP18 in the membrane fraction for both protoplast rupture and total-rupture method (Figure 3.14, 3.16). However, the GFP signal of the protein at the predicted MW of the GFP-OEP18 protein is higher than the signal for the other full-length protein, OEP18-GFP (Figure 3.15, 3.17). Both fusion constructs also show very high signal abundance (>80%) in membrane fraction than cytosol when examining only intact protein signal (Figure 3.15, 3.17), indicating that the large portion of full-length fusion protein that remained in the cytoplasm is possibly degraded. Based on paired t-test on pellet signal abundance of full-length constructs in both two rupture methods and two quantification methods, it is validated that OEP18-GFP indeed has significantly more signal abundance in the membrane fraction than GFP-OEP18 as all protein signal is involved; however it is false that OEP18-GFP is more susceptible to degradation as no significant difference of signal abundance between two constructs exist in samples if only examining the intact protein signal (Table 3.4). These results further reinforce the conclusion drawn from the microscopy images and Western Blot images.

4.4 The C-terminal reverse TP-like sequence of OEP18 is essential for targeting to chloroplast outer membrane, but cannot anchor to membrane alone

The purpose of using the two truncated OEP18 fusion constructs was to determine if the predicted C-terminal reverse TP-like sequence of OEP18 is required for targeting to the

chloroplast outer membrane. The two truncated OEP18 fusion constructs used in this study were OEP18ΔCT-GFP and OEP18CT-GFP.

The epifluorescent images of OEP18ΔCT-GFP show strong punctate-like structures in the cytoplasm unlike the web-like structure produced by the full-length fusion constructs around chloroplasts (Figure 3.12). The failure of OEP18ΔCT-GFP to target to chloroplasts indicates that the C-terminus, containing the reverse TP-like sequence, is needed for targeting. It appears that the lack of the C-terminus also results in protein aggregation (punctates) caused by protein misfolding. OEP18ΔCT-GFP has a stronger signal in the insoluble membrane fraction than the negative control 35S:GFP (Figure 3.14, 3.16), which may suggest that OEP18ΔCT-GFP can target to the chloroplast. However, it is likely caused by that the misfolded and aggregated OEP18ΔCT-GFP in the cytoplasm remains insoluble after both rupture methods and is thus included in the insoluble membrane fraction, which was expected based on previous microscopy images.

The epifluorescent images of OEP18CT-GFP show similar web-like structure around chloroplasts as OEP18-GFP and GFP-OEP18 but with stronger signal (Figure 3.12). This suggests that the C-terminus of OEP18 alone is able to target GFP to chloroplasts as efficiently as the full length OEP18. However, the quantified signal distribution contradicts with the observation from the epifluorescent images. The signal distribution of OEP18CT-GFP was approximately the same as 35S:GFP when only protoplast rupture was used to isolate fractions (Figures 3.14, 3.16). Moreover, when using the total rupture method, the signal distribution of OEP18CT-GFP in soluble fraction is significantly higher than any other constructs including the full-length fusion constructs (Figures 3.15, 3.17). The low signal abundance of OEP18-CT in the membrane fraction contradicted

with the observation made from the epifluorescence images. However, if combined with the results of OEP18ΔCT-GFP in the same experiment, one plausible conclusion is that although the C-terminus of OEP18 containing the reverse TP-like sequence is essential for targeting to chloroplast outer membrane, the CT of OEP18 alone is insufficient to anchor or interact with the outer membrane. The low signal abundance of OEP18CT-GFP in the membrane fraction is likely that the result of its inability to anchor to membrane. Although the truncated fusion protein has been imported to the chloroplast outer membrane, it is easily separated from the membrane during both rupture methods, thereby becoming part of the soluble fraction. It is also possible that OEP18CT-GFP is imported to chloroplasts but remains within the inter-membrane space due to the lack of information normally provided by the rest of OEP18. This aligns with the observation from confocal images and also explains the drastic increase of signal abundance of OEP18CT-GFP in soluble fraction from protoplast rupture to total rupture, as the soluble fraction of total rupture method include chloroplast inter-membrane space in addition to cytoplasm (Figures 3.15, 3.17). Another possibility is that the CT of OEP18 is essential for specific targeting, but it is not sufficient alone. However, these are pure speculation and requires further investigation.

4.5 OEP18 secondary structure is richer than predicted with Bioinformatic tools

The next step is to further validate prediction of OEP18 amino acid sequence and its secondary structure formation in the chloroplast outer membrane generated from the first part of the study, by testing for confrontational change between hydrophobic and hydrophilic environments. It was also to shed light how OEP18 interact with the outer

membrane. A recombinant construct OEP18-His was made by incorporating a cDNA encoding full length OEP18 (Figure 2.2) into the pET28a expression vector such that a recombinant protein would be produced with a 6X His-tag fused to the C-terminus and purified with IMAC resins, validated by Western blotting with anti-His antibody (Figures 3.19, 3.20). OEP18-His was further dialyzed from urea solvent to CD-compatible buffer. The secondary structure of purified and concentrated protein and its confrontational change under different buffer condition was measured using Circular Dichroism spectroscopy and analysis.

Previously OEP18 was only predicted to contain 2 α-helices (20 aa in total) and 6 small β-strands (21 aa in total) by TMHMM and PSI-Pred respectively, the structural content was predicted to be 6% α-helix and 13% β-strand. However, with original CD buffer content, it was 10% α-helix and 36% β-strand, indicating that some regions of OEP18 below the prediction threshold was able to form secondary structures especially β-strands. With addition of 50% TFE, structural content changed to 29% α-helix and 13% β-strand. Since CD analysis is unable to examine the structure changes of individual regions of the protein sequence, it is unclear which part of OEP18 changed to α-helix as solvent become more organic. Regions of β-strands decrease drastically under organic conditions and are also very close to the prediction. It is important to point out that the structural content remained unchanged from the Tris-HCl NaF buffer after addition of 50% liposome which mimic the surrounding of membrane bilayers (Table 3.4). This indicates that no liposome-peptide binding was present, OEP18 was inserted into the outer membrane without physically binding to the bilayers. The structure of OEP18 also most likely remain unchanged after its localization. However, there are also other possible causes for

the unchanged secondary structure of OEP18 between original CD buffer and liposome buffer: the liposome used for this study is not a perfect mimic of the chloroplast outer membrane; a possible receptor protein is required for OEP18 targeting and anchoring to the chloroplast outer membrane, which is not included in the liposome buffer; there can be other interactions between OEP18 and liposome that do not result in a change in the secondary structure of OEP18.

During this thesis, Alphafold, a new bioinformatic tool for predicting protein 3D topology was published. This program utilizes deep machine learning and the large database from Protein Data Bank (PDB) and is proven to be the most accurate 3D protein folding up to date (Yang et al, 2020). The six T159L OEP candidates were searched in Alphafold database for predicting its secondary structure and topology, all candidates were found except E-Tu. OEP18, Toc159, Toc132, OEP16-2 and OEP15-1 are all predicted to have hydrophobic α-helix near their C-Terminus (Figure 4.1). Interestingly pBRP was also predicted to have the same feature but with low confidence, which align with the PSIpred predictions. The secondary structure position predicted by Alphafold also fits the secondary structure mapped by PSIpred for the six OEPs, suggesting the predictions made by PSIpred are reliable. Although the majority of OEP18's structure is predicted with low confidence, the hairpin structure formed by two anti-parallel β-strands was predicted with high confidence (Figure 4.1A). The two anti-parallel β-strands are also predicted to be the same length and position by PSIpred (Figure 3.6H). The loop between the two β-strands which is also predicted with high confidence, is confirmed with only strong hydrophobic amino acid, this feature is also present in pBRP around the T159L candidates (FIgure 4.1A&D). A hydrophobic loop is one of the most common

types of non-covalent interactions between membrane protein and the lipid layers, it creates a weak interaction by embedding part of the hydrophobic loop within the membrane (Lomize et al, 2007). Combined with the CD result that OEP18 does not have any liposome-peptide binding in organic buffer, the predicted hydrophobic loop is a plausible explanation of how OEP18 is anchored to the outer membrane after the C-terminus containing the reverse TP-like sequence guides its import of OEPs into chloroplast. This speculation needs further evidence from *in vitro* experiments.

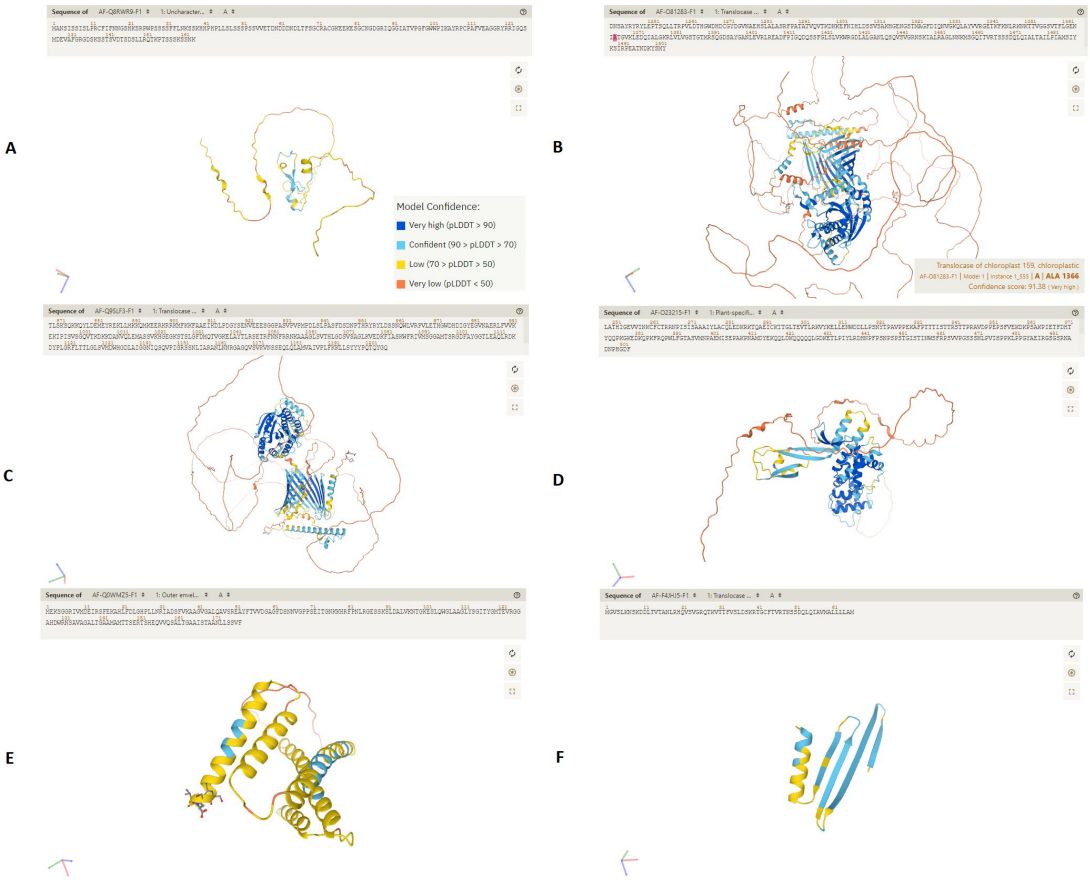

Figure 4.1 3D topology prediction of 6 T159L OEP candidates by Alphafold v2.0. Alphafold v2.0 predicts the possible secondary structure and topology of target protein. Flat arrow indicates predicted β-strand, spiral indicates predicted α-helix. Panel order: A) OEP18; B) Toc159; C) Toc132; D) pBRP; E) OEP16-2; F) OEP15-1. OEP18 shows a predicted hydrophobic loop formed by two anti-parallel β-strands. Hydrophobic loop are only predicted in pBRP and OEP18.

4.6 Future directions

In this study the data appear to support the hypothesis that the C-terminus of OEP18 containing the reverse TP-like sequence is essential for specific targeting to the chloroplast outer membrane. As far as this study investigates, it is determined that the C-terminus of OEP18 cannot anchor to the membrane on its own, but whether the C-terminus of OEP18 plays a role in anchoring/ interacting with the membrane still remains unclear. To further investigate how OEP18 interacts with the outer membrane, more truncated or substitution fusion constructs with GFP need to be examined to define the region within OEP18 that is required for membrane association and its proper folding, the predicted hydrophobic loop by Alphafold can be an excellent point of interest to start. The next truncated constructs to investigate should be two GFP-tagged constructs that are either truncated at the C-terminus or the predicted hydrophobic loop. A possible receptor can also be required for proper targeting/ anchoring of OEP18 to the chloroplast outer membrane. To further investigate this possibility, Chloroplast *in vitro* import assay is a suitable tool to use. Briefly, the chloroplast are prepared to be competent for in vitro import of recombinant preproteins synthesized using *in vitro* translation system derived from wheat germ lysates. The assays are mainly used for determining suborganellar location of imported preproteins, and allow closer study of import mechanism (Ling & Jarvis, 2016). Another future direction of this research can be distinguishing the difference in the import pathway between Outer Envelope Protein and Inner Envelope Protein. The protein import pathway to the chloroplast inner membrane and also thylakoid membrane still largely remains unknown, it can be a promising avenue of research.

The various functions of plants are carried out by a series of sophisticated intracellular molecular processes, and studying these processes leads to a better understanding of plant growth and development. The application of a variety of scientific approaches is crucial to maximize our understanding of complex biological development. In this study multiple approaches were used: bioinformatic approach was used to determine possible secondary structure and characteristics of all 117 OEPs and categorize them into predicted import pathways; cellular approach was used to examine the *in vivo* expression of designed fusion constructs in protoplasts; biochemical approach was used to accurately quantify the presence of expressed fusion protein using Western Blot analysis. It is important to investigate the hypothesis in different approaches as complex biological phenomenon cannot be understood fully from a single angle. An advancement in one approach often broadens the research direction in the other approaches, thus it is vital for scientific research to bring in new tools or methods to increase our knowledge of current field and set further goals.

5. Conclusion

The three hypotheses have been investigated in this study using different approaches. The first hypothesis is that OEP18 and other T159L candidates possess distinctive characteristics that are not used by the four known pathways, and it was proven partially true. All 117 predicted OEP (Inoue, 2015) proteins are examined by various bioinformatics to contain any distinctive characteristics used by four canonical pathways or the TP-like sorting signal at the C-terminus as seen in Toc159 for its novel T159L pathway. The eight predicted T159L candidates including OEP18 (Grimberg, 2016) were examined closely for their putative assignment to the T159L pathway. Six of the eight candidates, including OEP18, were strongly predicted to use the T159L following the bioinformatic analyses and predictions that were used. DUF869 and MIRO2 were eliminated from the candidates based on their strong predictions of transmembrane α-helix in the C-terminus overlapping within the TP-like region, indicating their possibility of utilizing the TA-anchor pathway rather than T159L pathway. The remaining six T159L candidates including OEP18 are strongly predicted to possess all required features of the T159L pathway by various bioinformatics.

The second and third hypothesis are that predicted CT reverse-TP like sequence is required for targeting of OEP18 to the outer chloroplast membrane; and if so, it possesses secondary structure information required for anchoring of OEP18 to the outer membrane. OEP18 was further tested by expressing EGFP fusion constructs in live cells and observing using a combination of fluorescence microscopy and signal quantification. When GFP is attached to the N-terminus of OEP18 instead of the C-terminus, targeting to the chloroplast outer membrane is interfered with. When the C-terminus of OEP18 is

removed, the protein is not targeted to chloroplasts and is retained in the cytoplasm. The punctate structures observed under epifluorescence microscope indicates formation of protein aggregation and misfolding. It is demonstrated that the C-terminus region is required for OEP18 proper folding and specific import to chloroplast outer membrane, however the C-terminus alone does not provide attachment to the membrane in the later steps, membrane association is likely provided by specific structure formation formed by OEP18 as whole.

The secondary structure composition and possible topology change of OEP18 in different environment were further analyzed with Circular Dichroism. It was determined that no liposome-peptide binding is present under similar environment of lipid bilayers, suggesting OEP18 is not physically bounded to the outer membrane during interaction, rather the physical structure of OEP18 facilitates the anchoring to the membrane. It is also possible that a receptor is required for specific targeting and anchoring of OEP18 to the chloroplast outer membrane. Based on the recently available Alphafold 3D structure prediction, the two short parallel β-strands and the hydrophobic loop in-between are predicted to be the key feature for membrane association as a peripheral membrane protein. In the future, more reliable tools need to be included in the bioinformatic analysis for more accurate predictions. More truncated or substitution fusion constructs with GFP need to be designed to define the exact region within OEP18 required for membrane association and its proper folding. In addition, more structural analysis can be used to determine the OEP18 topology during membrane interaction. A new hypothesis for future study is that both the predicted hydrophobic loop and the CT reverse TP-like sequence are required for proper anchoring of OEP18 to the chloroplast outer membrane.

In conclusion, the results from this study indicate that the C-terminus of OEP18 contains essential information for its specific import from the cytosol to the chloroplast outer membrane, however the C-terminus alone does not have sufficient structural information for association with the outer membrane. Furthermore, OEP18 is validated to share the same novel chloroplast targeting pathway as is used by Toc159, which can help to further validate the other T159L candidates.

6. Reference

Biswal, U, Mukesh, R (2003). Transformation of Chloroplast to Gerontoplast. *Chloroplast Biogenesis*. pp. 155–242

Bölter, B (2018). En route into chloroplasts: preproteins' way home. *Photosynthesis Research*, 138: 263-275.

Bolter, B., & Soll, J (2001). Ion Channels in the Outer Membranes of Chloroplasts and Mitochondria: Open Doors or Regulated Gates? *The EMBO Journal* 20 (5): 935–40.

Bouchnak, I., Brugiere, S., Moyet, L., Gall, S. L., Salvi, D., Kuntz, M., Tardif, M. & Rolland, N (2019). Unraveling Hidden Components of the Chloroplast Envelope Proteome: Opportunities and Limits of Better MS Sensitivity. *Molecular and Cellular Proteomics, 18:* 1285-1306.

Bruce BD (2000). Chloroplast transit peptides: structure, function and evolution. *Trends Cell Biol*. 2000 Oct;10(10):440-7.

Sjuts, I., Soll, J. & Bölter, B (2017). Import of Soluble Proteins into Chloroplasts and Potential Regulatory Mechanisms. *Front. Plant Sci*. 8(168).

Chang J, Chen L, Yeh Y, Hsiao C, and Li H. (2017). Chloroplast Preproteins Bind to the Dimer Interface of the. *Molecular and Cell Biology* 173: 2148–62. doi:10.1104/pp.16.01952.

Chen, Y. L., Chen, L. J., Chu, C. C., Huang, P. K., Wen, J. R. & Li, H. M. (2018). Tic236 links the outer and inner membrane translocons of the chloroplast. *Nature, 7734:* 125-129.

Chotewutmontri, P. & Bruce, B. (2015). Non-native, N-terminal Hsp70 Molecular Motor

Recognition Elements in Transit Peptides Support Plastid Protein Translocation. *J Biol Chem.290(12):* 7602-7621.

Chotewutmontri P, and Barkan A. (2016). Dynamics of Chloroplast Translation during Chloroplast Differentiation in Maize. *PLOS Genetics*, 1–28. doi:10.1371/journal.pgen.1006106.

Chotewutmontri P, and Bruce BD. (2015). Non-Native, N-Terminal Hsp70 Molecular Motor Recognition Elements in Transit Peptides Support Plastid Protein. *The Journal of Biological Chemistry* 290 (12): 7602–21. doi:10.1074/jbc.M114.633586.

Chotewutmontri P, Reddick LE, Mcwilliams DR, Campbell IM, and Bruce BD. (2012). Differential Transit Peptide Recognition during Preprotein Binding and Translocation into Flowering Plant Plastids. *The Plant Cells* 24: 3040–59. doi:10.1105/tpc.112.098327.

Cooper GM. (2000). The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates; Chloroplasts and Other Plastids.

Costello JL, Castro IG, Camões F, Schrader TA, Mcneall D, Yang J, Wilmanns M, Jedd G, Islinger M, and Schrader M. (2017). Predicting the Targeting of Tail-Anchored Proteins to Subcellular Compartments in Mammalian Cells. *Journal of Cell Science*, 130: 1675–87. doi:10.1242/jcs.200204.

Day PM, Inoue K, Theg SM. (2019). Chloroplast Outer Membrane β-barrel Protein Use Components of the general import Apparatus. *The Plant Cell*, 31 (8) 1845-1855; DOI: 10.1105/tpc.19.00001

Day, M. P. & Theg, S. M. (2018). Evolution of protein transport to the chloroplast

envelope membranes. *Photosynthesis Research*, 138: 315-326.

Dhanoa PK, Richardson LGL, Smith MD, Gidda SK, Matthew PA, Andrews DW, and Mullen RT. (2010). Distinct Pathways Mediate the Sorting of Tail-Anchored Proteins to the Plastid Outer Envelope. *PLoS ONE* 5 (4). doi:10.1371/journal.pone.0010098.

Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols, 2:* 953-971.

Emanuelsson, O., Nielsen, H. & von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Science, 8(5):* 978-984.

Fairman, James W, Nicholas Noinaj, and Susan K Buchanan. (2012). The Structural Biology of β-Barrel Membrane Proteins: A Summary of Recent Reports. *Curr Opin Struct Biol 21* (4): 523–31. doi:10.1016/j.sbi.2011.05.005.

Ganesan I, Shi LX, Labs M, Theg SM. (2018) Evaluating the Functional Pore Size of Chloroplast TOC and TIC Protein Translocons: Import of Folded Proteins. *The Plant Cell*, Vol. 30: 2161–2173

Glynn, J. M., Froehlich, J. E., & Osteryoung, K. W. (2008). Arabidopsis ARC6 coordinates the division machineries of the inner and outer chloroplast membranes through interaction with PDV2 in the intermembrane space. *The Plant cell*, 20(9), 2460–2470. https://doi.org/10.1105/tpc.108.061440

Grimberg N. (2016). Characterizing an alternative chloroplast outer membrane targeting signal in *Arabidopsis thaliana*. *Wilfred Laurier University, Department of Biology,*

Master's Thesis.

Gross, L. E., Spies, N., Simm, S. & Schleiff, E. (2020). Toc75V/OEP80 is processed during translocation into chloroplasts, and the membrane-embedded form exposes its POTRA domain to the intermembrane space. *FEBS Open Bio 10:* 444-454.

Hinnah SC, Wagner R, Sveshnikova N, Harrer R, Soll J. (2002). The Chloroplast Protein Import Channel Toc75: Pore Properties and Interaction with Transit Peptides. *Biophysical Journal*. 93(2): 899-911

Huang S, Taylor NL, Narsai R, Eubel H, Whelan J, and Millar AH. (2009). Experimental Analysis of the Rice Mitochondrial. *Plant Physiology* 149: 719–34. doi:10.1104/pp.108.131300.

Inoue K, and Gilroy S. (2015). Emerging Knowledge of the Organelle Outer Membranes – Research Snapshots and an Updated List of the Chloroplast Outer Envelope Proteins. *Frontiers in Plant Science* 6 (April): 1–5. doi:10.3389/fpls.2015.00278.

Ismail N, Hedman R, Schiller N, von Heijne G. (2012) A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane integration. *Nat Struct Mol Biol* 19: 1018–1022

Janes RW. (2010). Synchrotron Radiation Circular Dichroism (SRCD) Protein Conformations and Protein Interactions. *Biochem. Soc. Trans.* 38: 861–73. doi:10.1042/BST0380861.

Järvi S, Gollan PJ, and Aro E. (2013). Understanding the Roles of the Thylakoid Lumen in Photosynthesis Regulation. *Frontiers in Plant Science* 4: 1–14.

doi:10.3389/fpls.2013.00434.

Jarvis, P. & López-Juez, E. (2013). Biogenesis and homeostasis of chloroplasts and other plastids. *Nat. Reviews Mol Cell Bio*, 14: 787-802.

Jarvis P. (2008). Targeting of Nucleus-Encoded Proteins to Chloroplasts in Plants. *New Phytologist* 179: 257–78. doi:10.1111/j.1469-8137.2008.02452.x.

Jones, T. & Rapaport, D. (2017). Early stages in the biogenesis of eukaryotic β-Barrel proteins. *FEBS Letters, 591:* 2671-2681.

Keeling, P. J. (2004). Diversity and Evolutionary History of Plastids and Their Hosts."*American Journal of Botany* 91(10): 1481-1493.

Kessler F, Schnell DJ. (2006). The function and diversity of plastid protein import pathways: a multilane GTPase highway into plastids. *Traffic*.7(3):248-57.

Kessler F, and Schnell DJ. (2007). The Function and Diversity of Plastid Protein Import Pathways: A Multilane GTPase Highway into Plastids. *Traffic* 7: 248–57. doi:10.1111/j.1600-0854.2005.00382.x.

Kim TS, Kim HD, Park YJ, Kong E, Yang HW, Jung Y, Kim Y, & Kim J. (2019). JNK activation induced by ribotoxic stress is initiated from 80S monosomes but not polysomes. *BMB Rep*. 52(8):502-507.

Kim, D. H., Park, M. J. Gwon, G. H., Silkov, A., Xu, Z. Y., Yang, E. C., Song, S., Song, K., Kim, Y., Yoon, H. S., Honig, B., Cho, W., Cho, Y. & Hwang, I. (2014). An Ankyrin Repeat Domain of AKR2 117 Drives Chloroplast Targeting through Coincident Binding of Two Chloroplast Lipids. *Dev. Cell*, 30: 598-609.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680–685.

Lee, D. W. & Hwang, I. (2018). Evolution and Design Principles of the Diverse Chloroplast Transit Peptides. *Mol. Cells*, 41(3): 161-167.

Lee, D. W., Kim, J. K., Lee, S., Choi, S., Kim, S. & Hwang, I. (2008). *Arabidopsis* Nuclear-Encoded Plastid Transit Peptides Contain Multiple Sequence Subgroups with Distinctive Chloroplast-Targeting Sequence Motifs. *The Plant Cell, 20:* 1603-1622.

Lee, D. W., Woo, S., Geem, K. R., & Hwang, I. (2015). Sequence Motifs in Transit Peptides Act as Independent Functional Units and Can Be Transferred to New Sequence Contexts. *Plant physiology*, *169*(1), 471–484. doi:10.1104/pp.15.00842

Lee, D.W., Lee, J., and Hwang, I. (2017). Sorting of nuclear-encoded chloroplast membrane proteins. *Curr. Opin. Plant Biol*. 40, 1-7.

Lee, D.W., Yoo, Y.J., Razzak, M.A., and Hwang, I. (2018). Prolines in transit peptides Are crucial for efficient preprotein translocation into chloroplasts. *Plant Physiol*. 176, 663-677.

Lee DW, Jung C, and Hwang I. (2013). Biochimica et Biophysica Acta Cytosolic Events Involved in Chloroplast Protein Targeting. *BBA - Molecular Cell Research* 1833 (2). Elsevier B.V.: 245–52. doi:10.1016/j.bbamcr.2012.03.006.

Lee, J., Kim, D. H. & Hwang, I. (2014). Specific targeting of proteins to outer envelope membranes of endosymbiotic organelles, chloroplast, and mitochondria. *Front. Plant Sci.5(173).*

Lee J, Kim DH, Hwang I, Abell BM, and Hallam S. (2014). Specific Targeting of Proteins to Outer Envelope Membranes of Endosymbiotic Organelles, Chloroplasts, and Mitochondria. *Frontiers in Plant Science* 5: 1–11. doi:10.3389/fpls.2014.00173.

Lee, KH, Kim, DH., Lee, SW., Zhoo Hyeon 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 (3): 388–97.

Lee, L., Lee, H., Kim, J., Lee, S., Kim, D. H., 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. *The Plant Cell, 23(4):* 1588-1607.

Li, H. M. & Teng, Y. S. (2013). Transit peptide design and plastid import regulation. *Trends in Plant Sci. 18(7):* 360-366.

Li, H.M., Schnell, D. & Theg, S. M. (2020). Protein Import Motors in Chloroplasts: On the Role of Chaperones. *Plant Cell, 32:* 536-542.

Ling, Q., & Jarvis, P. (2016). Analysis of Protein Import into Chloroplasts Isolated from Stressed Plants. *Journal of visualized experiments*: JoVE, (117), 54717. https://doi.org/10.3791/54717

Lomize AL, Pogozheva ID, Lomize MA, Mosberg HI. (2007). The role of hydrophobic interactions in positioning of peripheral proteins in membranes. *BMC Struct Biol*. 29;7:44.

Lopez-juez E, and Pyke KA. (2005). Plastids Unleashed: Their Development and Their Integration in Plant Development. *Int. J. Dev. Biol.* 49: 557–77. doi:10.1387/ijdb.051997el.

Lung SC, and Chuong S. (2012). A Transit Peptide – Like Sorting Signal at the C Terminus Directs the *Bienertia sinuspersici* Preprotein Receptor Toc159 to the Chloroplast Outer Membrane. *The Plant Cell.* 24: 1560–78. doi:10.1105/tpc.112.096248.

Lung SC, Smith MD, Weston JK, Gwynne W, Secord N, and Chuong S. (2014). The C-Terminus of *Bienertia sinuspersici* Toc159 Contains Essential Elements for Its Targeting and Anchorage to the Chloroplast Outer Membrane. *Frontiers in Plant Science* 5: 1–17. doi:10.3389/fpls.2014.00722.

Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M. and Penny, D. (2002). Evolutionary analysis of *Arabidopsis*, cyanobacterial and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus." *Proc. Natl. Acad. Sci. USA* 99: 12246-12251.

Marty NJ, Teresinski HJ, Hwang YT, Clendening EA, Gidda SK, Sliwinska E, Zhang D, et al. (2014). New Insights into the Targeting of a Subset of Tail-Anchored Proteins to the Outer Mitochondrial Membrane. *Frontiers in Plant Science* 5: 1–20. doi:10.3389/fpls.2014.00426.

Mcfadden GI. (2014). Origin and Evolution of Plastids and Photosynthesis in Eukaryotes. *Cold Spring Harbor Perspectives in Biology*, 1–10.

Nakai M. (2015). The TIC complex uncovered: The alternative view on the molecular mechanism of protein translocation across the inner envelope membrane of chloroplasts.

Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1847 (9): 957-967.

Okawa, K., Inoue, H., Adachi, F., Nakayama, K., Ito-Inaba, Y., Schnell, D.J., Uehara, S., and Inaba, T. (2014). Targeting of a polytopic membrane protein to the inner envelope membrane of chloroplasts in vivo involves multiple transmembrane segments. *J. Exp. Bot*. 65, 5257-5265.

Patron, N. J. & Waller, R. F. (2007). Transit peptide diversity and divergence: A global analysis of plastid targeting signals. *BioEssays* 29: 1048-1058.

Poincelot RP. (1976). Lipid and Fatty Acid Composition of Chloroplast Envelope Membranes from Species with Differing Net Photosynthesis. *Plant Physiology* 58: 595–98.

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 *Frontiers in Plant Science* 5: 1–14. doi:10.3389/fpls.2014.00269.

Schein, AI, Kissinger, JC, and Ungar, LH. (2001). Chloroplast Transit Peptide Prediction: A Peek inside the Black Box. *Oxford University Press* 29 (16): 4–9.

Schnell, D. J. (2019). The TOC GTPase Receptors: Regulators of the Fidelity, Specificity and Substrate Profiles of the General Protein Import Machinery of Chloroplasts. *The Protein Journal*, 38: 343-350.

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

Tamm LK, Hong H, Liang B. (2004). Folding and Assembly of H -Barrel Membrane Proteins. *Biochimica et Biophysica Acta* 1666: 250–63. doi:10.1016/j.bbamem.2004.06.011.

Taylor, PD, Toseland CP, Attwood TK, Flower DR. (2006). Beta Barrel Trans-Membrane Proteins: Enhanced Prediction Using a Bayesian Approach." *BIoinformation* 1 (6): 231–33.

Teresinski, H. J., Gidda, S. K., Nguyen, T., Howard, N., Porter, B. K., Grimberg, N., Smith, M. D., Andrews, D. W., Dyer, J. M. & Mullen, R. T. (2019). An RK/ST C-Terminal Motif is Required for Targeting of OEP7.2 and a Subset of Other Arabidopsis Tail-Anchored Proteins to the Plastid Outer Envelope Membrane. *Plant Cell Physiol. 60(3):* 516-537.

Teresinski, H. J., Gidda, S. K., Nguyen, T., Howard, N., Porter, B. K., Grimberg, N., Smith, M. D., Andrews, D. W., Dyer, J. M. & Mullen, R. T. (2019). An RK/ST C-Terminal Motif is Required for Targeting of OEP7.2 and a Subset of Other Arabidopsis Tail-Anchored Proteins to the Plastid Outer Envelope Membrane. *Plant Cell Physiol*. 60(3): 516-537.

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

Thomson, S. M., Pulido, P. & Jarvis, P. R. (2020). Protein import into chloroplasts and its regulation by the ubiquitin-proteasome system. *Biochemical Society Transactions*, 48: 71-82.

Tranel PJ, Froehlich J, Goyal A, and Keegstral K. (1995). A Component of the Chloroplastic Protein Import Apparatus Is Targeted to the Outer Envelope Membrane via a Novel Pathway. *The EMBO Journal* 14 (11): 2436–46.

Tsaousis, G. N., Hamodrakas, S. J. & Bagos, P. G. (2017). Predicting Beta Barrel Transmembrane Proteins Using HMMs. *Methods Mol Biol. 1552:* 43-61.

Wise, Robert R. (2006). Chapter 1 The Diversity of Plastid Form and Function. *The Structure and Function of Plastids*, 3–26.

Wu FH, Shen SC, Lee LY, Lee SH, Chan MT, Lin CS. (2009). Tape-Arabidopsis Sandwich - a simpler Arabidopsis protoplast isolation method. *Plant Methods*, 24:5:16.

Yang J, Anishchenko I, Park H, Peng Z, Ovchinnikov S, Baker D. (2020). Improved protein structure prediction using predicted inter-residue orientations. *bioRxiv*. doi: https://doi.org/10.1101/846279

Zhang, XP, and Elzbieta G. (2002). Interaction of Plant Mitochondrial and Chloroplast Signal Peptides with the Hsp70 Molecular Chaperone. *TRENDS in Plant Science* 7 (1): 19–21.

Zhuang, X., Chung, K. P., & Jiang, L. (2017). Targeting tail-anchored proteins into plant organelles. *PNAS 114(8):* 1762-1764.

7. Appendix

7.1 Sequencing confirmation of OEP18 EGFP fusion constructs

OEP18 full-length in pSAT6-C1 vector

MADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEF
VTAAGITLGMDELYKSGLRSRAMANSISSISLPRCFIFNNGSHKSRPWPSSSSFFLNKSSKHHPHPLLSLS
SSPSSVVETDNDDDNDLTFSGCRACGKEEKESGCNGDGRIQGGIATVPGFGWWPIKAYRPCPAFVEAGGRY
RRIGQSMDEVAFGRGDSKSSTSVDTSDSLLRQTKPTSSSKSSNK

OEP18 full-length in pSAT6-N1 vector

HLIISFKQKQFSENFHHLRTIAMVRTQISSMANSISSISLPRCFIFNNGSHKSRPWPSSSSFFLNKSSKHH
PHPLLSLSSSPSSVVETDNDDDNDLTFSGCRACGKEEKESGCNGDGRIQGGIATVPGFGWWPIKAYRPCPA
FVEAGGRYRRIGQSMDEVAFGRGDSKSSTSVDTSDSLLRQTKPTSSSKSSNKGILMVSKGEELFTGVVPIL
VELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKS
AMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYTATTSISWPTAE
ERHQGELQIRHNIEDGSVQLATTTADTPIATPXLLRHTXLNTQSALNKNPNRXE

OEP18w/oCT in pSAT6-N1 vector

QQFKSFLLKQKQFSENFHHLRTIAMVRTQISSMANSISSISLPRCFIFNNGSHKSRPWPSSSSFFLNKSSK
HHPHPLLSLSSSPSSVVETDNDDDNDLTFSGCRACGKEEKESGCNGDGRIQGGIATVPGFGWWPIKAYRPC
PAFVEAGGRYRRIGQSMDEVAFGRGDLRILMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYG
KLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAE
VKFEGDTLVNRIELKGIDFKEDGNILGHKLEY

OEP18CT in pSAT6-N1 vector

LQQFKSFLLKQKQFSENFHHLRTIAMVRTQISSSKSSTSVDTSDSLLRQTKPTSSSKSSNKGILMVSKGEE
LFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDH
MKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSH
NVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMV
LLEFVTAAGITLGMDELYK