

Investigating Protein Targeting to the Outer Membrane of Plastids

by

Delaney Nash

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

©Delaney Nash 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, such as photosynthesis. The specialized function of each plastid is informed by its distinct and dynamically regulated proteome. The vast 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 regulate the import of plastid proteins, thus, control the plastid proteome composition.

Proteins embedded within the outer envelope membrane of plastids have been generally categorized into four groups which include, β -barrel proteins, tail-anchored proteins, signal-anchored proteins, and CT TP-like proteins. Each group is defined by distinct structural and plastid-targeting characteristics. β -barrel proteins are composed of β -sheets and their plastid-targeting signal and mechanism is not well understood. Tail-anchored and signal-anchored proteins are tethered to the plastid outer envelope membrane by a single transmembrane alpha-helix located at the proteins C-terminus or N-terminus, respectively, and use a variety of physiochemical features for plastid-targeting. Lastly, the only currently defined CT TP-like protein, Toc159, utilizes a C-terminal plastid-targeting signal with transit peptide-like features.

In this study, the structure and plastid-targeting signal of the plastid protein Outer Envelope Protein 16-2 (OEP16-2) was investigated. Computational structural analysis showed that OEP16-2 is embedded within the outer envelope membrane by four alpha-helical transmembrane domains and does not share structural similarity with defined categories of outer envelope proteins. Furthermore, three internal transmembrane alpha helical domains

were sufficient for plastid-targeting. These internal targeting domains cannot be characterized by currently defined outer envelope protein targeting strategies. Thus, OEP16-2 was classified in a fifth outer envelope protein category, defined by multiple transmembrane alpha helices and internal targeting domains. Future experiments will examine the structure and plastid-targeting signal of other outer envelope proteins with multiple transmembrane helices.

Acknowledgements

I would like to thank Professor Chuong for supporting this research project and encouraging my professional growth. I would like to thank my committee members Professor Moffatt, Professor Lolle, and Professor Smith for guiding this project and providing valuable feedback. I would also like to thank my colleagues Alyssa Overton and Nilanth Yogadasan for their friendship, support, and advice. Lastly, I would like to thank my partner Max and my family for their endless support and encouragement during my Master's degree.

Table of contents

<i>Title Page</i>	<i>i</i>
<i>Declaration</i>	<i>ii</i>
<i>Abstract</i>	<i>iii</i>
<i>Acknowledgements</i>	<i>v</i>
<i>List of Figures</i>	<i>x</i>
<i>List of Tables</i>	<i>xii</i>
<i>List of Abbreviations</i>	<i>xiii</i>
<i>List of Symbols</i>	<i>xiv</i>
Chapter 1. Introduction	1
1.1. <i>The Evolution, Structure, and Function of Chloroplasts</i>	1
1.2. <i>Plastid-types and Plastid-Proteomes</i>	3
1.3. <i>Plastid-localized Protein Transcription, Translation, and Translocation</i>	5
1.4. <i>Features of N-terminal Transit Peptides</i>	9
1.5. <i>The Diverse Function and Structure of TOC Complexes and their Receptors</i>	12
1.6. <i>Chaperone-Receptor Interactions & Toc Receptor Regulation</i>	14
1.7. <i>UPS regulation of TOC complexes and Preproteins</i>	15
1.8. <i>Targeting Mechanisms Utilized by Plastid Outer Envelope Proteins</i>	17
1.9. <i>Recruitment and Integration of Beta-Barrel Proteins at the Plastid Outer Envelope</i>	19
1.10. <i>Recruitment and Integration of Signal-Anchored Proteins at the Plastid Outer Envelope</i>	22
1.11. <i>Recruitment and Integration of Tail-Anchored Proteins at the Plastid Outer Envelope</i>	23
1.12. <i>The Role and Function of Chaperone Protein ARK2</i>	27
1.13. <i>A Potentially Novel OEP Targeting Mechanism</i>	30
1.14. <i>Predicting N-terminal Transit Peptides using Computational Tools</i>	32
1.15. <i>Function, Localization, Evolution, and Expression of OEP16-2</i>	32
1.16. <i>Structural Prediction of OEP16-2</i>	35
1.17. <i>Hypothesis and Research Objectives</i>	36
Chapter 2. Materials and Methods	39
2.1. <i>Computational analysis of the C-Terminal OEP16-2 sequence</i>	39
2.2. <i>Constructing OEP16-2 Fusion Constructs</i>	40
2.3. <i>Onion Cell Bombardment using the Biolistic Particle Delivery System</i>	43
2.4. <i>Computational Analysis of Secondary Structure and Domain Prediction</i>	45
2.5. <i>Protoplast Preparation and Transfection</i>	45
2.6. <i>Protoplast Subfractionation to Obtain Soluble, Insoluble, and Total Protein Fractions</i>	47

2.7.	<i>Protoplast Protein Separation by SDS-PAGE and Detection by Western Blot Analysis</i>	48
2.8.	<i>Computational Investigation of the Predicted OEP16-2 H2 domain</i>	50
2.9.	<i>Computation Prediction of OEP Targeting Structures and Structural Classification</i>	50
Chapter 3. Computational Assessment of the Predicted OEP16-2 Chloroplast Targeting Signal		52
3.1.	<i>Overview</i>	52
3.2.	<i>Computational Investigation of the OEP16-2 Using ChloroP</i>	52
	Figure 3.1. <i>TP analysis of OEP16-2 using the ChloroP server</i>	53
3.3.	<i>Computational Investigation of the OEP16-2 C-terminus for Transit Peptide Features</i>	53
	Table 3.1. <i>Comparing the Amino Acid Composition of the OEP16-2 Full-Length & C-terminal Sequences using ProtParam</i>	55
3.4.	<i>Computational Comparison of OEP16-2 Sequences Using an MSA</i>	55
	Figure 3.2. <i>Multiple sequence alignment of 29 OEP16-2 sequences from a variety of higher plant species</i>	57
3.5.	<i>Secondary Structure Prediction of OEP16-2 Using PSI-Pred</i>	58
	Figure 3.3. <i>Psi-Pred secondary structure prediction of AtOEP16-2</i>	59
3.6.	<i>Computational Investigation of OEP16-2 Using Phobius</i>	60
	Figure 3.4. <i>Phobius Analysis of the OEP16-2 Protein Sequence</i>	61
3.7.	<i>Computational Investigation of the OEP16-2 C-terminus Using HeliQuest Analysis</i>	61
	Figure 3.5. <i>Helical Wheel Projection of the OEP16-2 C-Terminus</i>	62
3.8.	<i>Summation of OEP16-2 Computational Predictions</i>	63
Chapter 4. Identifying the chloroplast targeting signal within the OEP16-2 Sequence by Epifluorescence		64
4.1.	<i>Overview</i>	64
4.2.	<i>Localization of Six Original Fluorescent Constructs after Onion Cell Bombardment</i>	64
	Figure 4.1. <i>Design of Six OEP16-2:EGFP Fluorescent Fusion Constructs</i>	65
	Figure 4.2. <i>Onion Cell Bombardment with Original Six OEP16-2 Fluorescent Fusion Constructs</i>	67
	Figure 4.3. <i>Onion Cell Bombardment Co-localization of Fusion Constructs and a Plastid Marker</i>	68
	Table 4.1. <i>Expected vs. Observed Localization of Six Initial Fusion Constructs Transiently Expressed in Onion Epidermal Cells</i>	69
4.3.	<i>Design and Transient Expression of Domain Truncation Constructs in Onion Epidermal Cells</i>	69
	Figure 4.4. <i>Overlaid NMR OEP16-1 structure onto an MSA created by Drea et al (2006)</i>	71
	Figure 4.5. <i>AtOEP16-2 Domain Prediction</i>	72
	Figure 4.6. <i>Design of Domain Truncation Fusion Constructs</i>	74

Figure 4.7.	<i>Onion cell bombardments using Domain Truncation Fusion Constructs</i>	75
Figure 4.8.	<i>Onion Epidermal Cell Co-bombardment of Domain Truncation Fusion Constructs with a Plastid Marker</i>	77
Table 4.2.	<i>Localization of Domain Truncation Constructs in Onion Epidermal Cells</i>	78
4.4.	<i>Transient Expression of H2-Domain Constructs in Onion Epidermal Cells</i>	80
Figure 4.9.	<i>Design of H2 Domain EGFP Fusion Constructs</i>	80
Figure 4.10.	<i>Onion Cell Co-bombardment of H2 Domain Fusion Constructs with a Plastid Marker</i>	80
Table 4.3.	<i>Localization of H2 Truncation Constructs in Onion Epidermal Cells</i>	81
4.5.	<i>Transient Expression of the S-Domain EGFP fusion Construct in Onion Epidermal Cells</i>	81
Figure 4.11.	<i>Design of OEP16-2-SD:EGFP Fusion Construct</i>	81
Figure 4.12.	<i>Onion Cell Co-bombardment with OEP16-2-SD:EGFP and a Plastid Marker</i>	82
Chapter 5. Immunodetection of OEP16-2 Subcellular Localization in Arabidopsis Protoplasts		83
5.1.	<i>Overview</i>	83
5.2.	<i>Quality of Subfractionation and Immunoblot Analysis</i>	83
5.3.	<i>Immunoblot analysis of Protoplast Plastid and Cytosolic Protein Fractions</i>	84
Figure 5.1.	<i>Ponceau Stain and Immunodetection of Fractionated Protoplasts Transfected with EGFP</i>	85
Figure 5.2.	<i>Ponceau Stain and Immunodetection of Fractionated Protoplasts Transfected with OPE16-2-H2:EGFP</i>	86
Figure 5.3.	<i>Ponceau stain and Immunodetection of Fractionated Protoplasts transfected with OEP16-2-H1:EGFP</i>	87
Chapter 6. Computational Assessment of the H2 Domains Physiochemical Properties		88
6.1.	<i>Overview</i>	88
6.2.	<i>OEP16-2 Protein Sequence Conservation</i>	89
Figure 6.1.	<i>OEP16-2 Multiple Sequence Alignment from Various Plant Species</i>	90
6.3.	<i>Properties and Patterns within the Predicted H2-Domain Sequence</i>	
Figure 6.2.	<i>The Amino Acid Trends within the Predicted H2-Domain</i>	93
Table 6.1.	<i>Amino acid Composition Analysis of the Predicted H2-Domain</i>	93
Figure 6.3.	<i>Phobius prediction of the predicted H2-domain</i>	94
Figure 6.4.	<i>Heli-Quest Predictions of the H2-Domain Protein Sequence</i>	95
Figure 6.5	<i>Construct Design of Future Experiments</i>	96
6.4.	<i>A Potential Pathway for OEP16-2 OEM-Translocation</i>	96
Figure 6.6.	<i>SWISS-MODEL Protein Sequence Alignment of AKR2A and OEP16-2</i>	98
Figure 6.7.	<i>Structural Model of OEP16-2 and AKR2A</i>	99

Chapter 7. Structural Classifications of Known OEPs	100
7.1. Overview	100
Figure 7.1. OEP Structural Categorization Pipeline	100
7.2. Identification & Categorization of OEPs by Structural Class for OEP Candidate Selection	101
Table 7.1. Compilation of OEPs and their Predicted Structural Classification	102
Figure 7.2. The Percentage of OEPs Grouped into Four Structural Classes	106
7.3. Investigating Predicted Multi-Pass Alpha Helical Proteins for Potential Targeting Features	107
Figure 7.3. MSA of OEP16 isoforms with HP20, HP30, HP30-2, and TIM22-3	109
Table 7.2. Predicted Multi-pass Alpha Helical Proteins Sorted in Functional Groups	110
8.0 Concluding Remarks	112
9.0 References	114
10.0 Appendix	120
A1. The primary protein sequence of AtOEP16-2 (At4G16160) retrieved from NCBI	120
A2. Protein Accession Numbers of OEP16-2 Isoforms from 29 plant species used to create the MSA in Figure 3.2.3.	120
A3. Sequences of Subcloned OEP16-2 EGFP Fusion Constructs	121
A4. OEP16-2 sequences from 30 species & the protein accession number retrieved from NCBI by pBLAST used to generate an MSA	123
A5. Solutions prepared for protocols in methods and materials	124
A6. List of Bioinformatic Servers and URLs	125

List of Figures

- Figure 1.1 The General Chloroplast Structure*
- Figure 1.2 Plastid-type Transition Network*
- Figure 1.3 Import of Preproteins across the Chloroplast Double Membrane*
- Figure 1.4 Outer Envelope Protein Targeting Mechanisms*
- Figure 1.5 Structure of ARK2A Lipid-Binding Pockets L₁ and L₂*
- Figure 1.6 MSA of OEP16-1 & OEP16-2 with Predicted Secondary Structure*
- Figure 1.7 OEP16-1 structure based on CD and NMR analysis*
- Figure 3.1 TP analysis of OEP16-2 using the ChloroP server*
- Figure 3.2 Multiple sequence alignment of 29 OEP16-2 sequences from a variety of higher plant species*
- Figure 3.3 Psi-Pred secondary structure prediction of AtOEP16-2*
- Figure 3.4 Phobius Analysis of the OEP16-2 Protein Sequence*
- Figure 3.5 Helical Wheel Projection of the OEP16-2 C-Terminus*
- Figure 4.1 Design of Six OEP16-2:EGFP Fluorescent Fusion Constructs*
- Figure 4.2 Onion Cell Bombardment with Original Six OEP16-2 Fluorescent Fusion Constructs*
- Figure 4.3 Onion Cell Bombardment Co-localization of Fusion Constructs and a Plastid Marker*
- Figure 4.4 Overlaid NMR OEP16-1 structure onto an MSA created by Drea et al (2006)*
- Figure 4.5 AtOEP16-2 Domain Prediction*
- Figure 4.6 Design of Domain Truncation Fusion Constructs*
- Figure 4.7 Onion cell bombardments using Domain Truncation Fusion Constructs*
- Figure 4.8 Onion Epidermal Cell Co-bombardment of Domain Truncation Fusion Constructs with a Plastid Marker*
- Figure 4.9 Design of H2 Domain EGFP Fusion Constructs*
- Figure 4.10 Onion Cell Co-bombardment of H2 Domain Fusion Constructs with a Plastid Marker*
- Figure 4.11 Design of OEP16-2-SD:EGFP Fusion Construct*
- Figure 4.12 Onion Cell Co-bombardment with OEP16-2-SD:EGFP and a Plastid Marker*
- Figure 5.1 Ponceau Stain and Immunodetection of Fractionated Protoplasts transfected with EGFP*
- Figure 5.2 Ponceau Stain and Immunodetection of Fractionated Protoplasts Transfected with OPE16-2-H2:EGFP*
- Figure 5.3 Ponceau stain and Immunodetection of Fractionated Protoplasts transfected with OEP16-2-H1:EGFP*
- Figure 6.1 OEP16-2 Multiple Sequence Alignment from Various Plant Species*
- Figure 6.2 The Amino Acid Trends within the Predicted H2-Domain*
- Figure 6.3 Phobius prediction of the predicted H2-domain*
- Figure 6.4 Heli-Quest Predictions of the H2-Domain Protein Sequence*
- Figure 6.5 Construct Design of Future Experiments*
- Figure 6.6 SWISS-MODEL Protein Sequence Alignment of AKRA2A and OEP16-2*

- Figure 6.7* *Structural modelling of OEP16-2 and comparison to AKRA2A*
Figure 7.1 *OEP Structural Categorization Pipeline*
Figure 7.1 *The Percentage of OEPs Grouped into Four Structural Classes*
Figure 7.2 *MSA of OEP16 isoforms with HP20, HP30, HP30-2, and TIM22-3*

List of Tables

<i>Table 2.1</i>	<i>URLs of Computational Tool Servers</i>
<i>Table 2.2</i>	<i>Primer Pairs and Vectors used to Subclone Specific OEP16-2 Sequences</i>
<i>Table 3.1</i>	<i>Comparing the Amino Acid Composition of the OEP16-2 Full-Length & C-terminal Sequences using ProtParam</i>
<i>Table 4.1</i>	<i>Expected vs. Observed Localization of Six Initial Fusion Constructs Transiently Expressed in Onion Epidermal Cells</i>
<i>Table 4.2</i>	<i>Localization of Domain Truncation Constructs in Onion Epidermal Cells</i>
<i>Table 4.3</i>	<i>Localization of H2 Truncation Constructs in Onion Epidermal Cell</i>
<i>Table 4.4</i>	<i>Localization of Truncation Fusions Constructs in Onion Epidermal Cells</i>
<i>Table 6.1</i>	<i>Amino acid Composition Analysis of the Predicted H2-Domain</i>
<i>Table 7.1</i>	<i>Compilation of OEPs and their Predicted Structural Classification</i>
<i>Table 7.2</i>	<i>Compilation of Predicted Multi-pass Alpha Helical OEPs and their Predicted Structural Classification</i>
<i>Table 7.3</i>	<i>Predicted Multi-pass Alpha Helical Proteins Sorted in Functional Groups</i>

List of Abbreviations

aa	<i>Amino Acid</i>
AKR2A	<i>Ankyrin Repeat Protein 2A</i>
ARD	<i>Ankyrin Repeat Domain</i>
CT	<i>C-terminal</i>
CD	<i>Circular Dichroism</i>
CT TP-like	<i>C-terminal Transit Peptide-Like</i>
H1	<i>Alpha Helix 1</i>
H2	<i>Alpha Helix 2</i>
H3	<i>Alpha Helix 3</i>
H4	<i>Alpha Helix 4</i>
H5	<i>Alpha Helix 5</i>
HWP	<i>Helical Wheel Projection</i>
IEM	<i>Inner Envelope</i>
IMS	<i>Intermembrane space</i>
LHC	Light Harvesting Complex
MSA	<i>Multiple Sequence Alignment</i>
NMR	<i>Nuclear Magnetic Resonance</i>
NT	<i>N-terminal</i>
OEM	<i>Outer Envelope</i>
OEP	<i>Outer Envelope Protein</i>

OEP16-1	<i>Outer Envelope Protein 16-1</i>
OEP16-2	<i>Outer Envelope Protein 16-2</i>
OEP16-3	<i>Outer Envelope Protein 16-3</i>
OEP16-4	<i>Outer Envelope Protein 16-4</i>
PEG	<i>Polyethylene Glycol</i>
ppi	<i>Plastid Protein Import Mutant</i>
PRAT	<i>Preprotein Amino Acid Transporter</i>
RuBisCO	<i>Ribulose 1,5-Bisphosphate Carboxylase Oxygenase</i>
RbcL	RuBisCO Large Subunit
RbcS	RuBisCO Small Subunit
SA	<i>Signal Anchored</i>
TL	<i>Thylakoid Lumen</i>
TM	<i>Thylakoid Membrane</i>
TIC	<i>Translocon of the Inner Membrane of Chloroplasts</i>
TOC	<i>Translocon of the Outer Membrane of Chloroplasts</i>
TP	<i>Transit Peptide</i>
TA	<i>Tail-Anchored</i>

List of symbols

~ *approximately*

Δ *without*

Chapter 1. Introduction

1.1. The Evolution, Structure, and Function of Chloroplasts

Plastids are plant-specific organelles that have a variety functional types. The most well studied plastid type is the chloroplast due to its role in photosynthesis. The chloroplast evolved through an endosymbiotic event that occurred millions of years ago (Bölter, 2018).

Endosymbiosis occurred when a heterotrophic eukaryote containing mitochondria engulfed a photosynthetic cyanobacterium via phagocytosis (Lee & Hwang, 2018). The evolution of the chloroplast from the symbiotic cyanobacterium, termed organellogenesis, was the origin of species for the land plants, green algae, red algae, and glaucophytes (Day & Theg, 2018; Patron & Waller, 2007). This endosymbiotic event is evident from the many similarities found between the structure and function of chloroplasts and cyanobacterium, including in protein sequence conservation, in the assembly and function of their photosynthetic machinery, in their genome structure and content, and in their membrane and proteome composition (Lee and Hwang, 2018).

Generally, the double membranes that enclose both cyanobacterium and plastids share similar protein and lipid compositions (Day & Theg, 2018). The chloroplast is composed of an outer envelope membrane (OEM) and inner envelope membrane (IEM) separated by an intermembrane space (IMS). The IEM encloses the stroma and the thylakoid membrane (TM); the TM contains the thylakoid lumen (TL; Figure 1.1; Lee et al., 2013).

It is speculated that complete organellogenesis of the symbiont required three major steps (Bölter, 2018). First, the lateral gene transfer of cyanobacterium genetic material to the host genome occurred. Next, the host cell evolved methods to transcribe and translate the

laterally transferred genes (Bölter, 2018). Finally, the host developed mechanisms to retarget the previously symbiont-encoded proteins back to the symbiotic organelle. The final step was likely pivotal in completing organellogenesis (Bölter, 2018). Additionally, the lateral transfer of a few genes may have initiated a rapid lateral transfer of genetic material to the host genome (Lee & Hwang, 2018). Through lateral gene transfer, the chloroplast genome has been reduced to ~100 genes, and as a result, ~95-98% of chloroplast-proteins are nuclear-encoded and cytosolically translated (Bölter, 2018; Lee & Hwang, 2018). Lateral gene transfer gave the host cell regulatory control over chloroplast-protein expression and import; ultimately enabling a harmonious and productive relationship between the symbiont and the host (Day & Theg, 2018). Cytosolic factors as well as complex protein machinery in the chloroplast outer and inner envelope membranes have evolved to selectively and specifically regulate protein import with fidelity for diverse substrates (Schnell, 2019). Additionally, the development of chloroplast-protein import pathways allowed host cell proteins to develop chloroplast targeting signals, as well as, novel functions and pathways (Day & Theg, 2018).

The chloroplast evolved mechanisms to provide the host with oxygen, carbohydrates, amino acids, specialized metabolites, lipids, and hormones. Moreover, the chloroplast plays a major role in ROS production and ion homeostasis, maintains an electron transport chain, and photosynthesizes. In exchange, the host cell protects the organelle from biotic and abiotic factors and maintains protein synthesis, regulation, and transport (Bölter, 2018; Lee & Hwang, 2018). Approximately 3000 proteins in the chloroplast enable these diverse functions. The majority chloroplast proteins are encoded by nuclear genes then synthesized in the cytosol and

targeted post-translationally to the correct chloroplast subcellular compartment (Thomson et al., 2020).

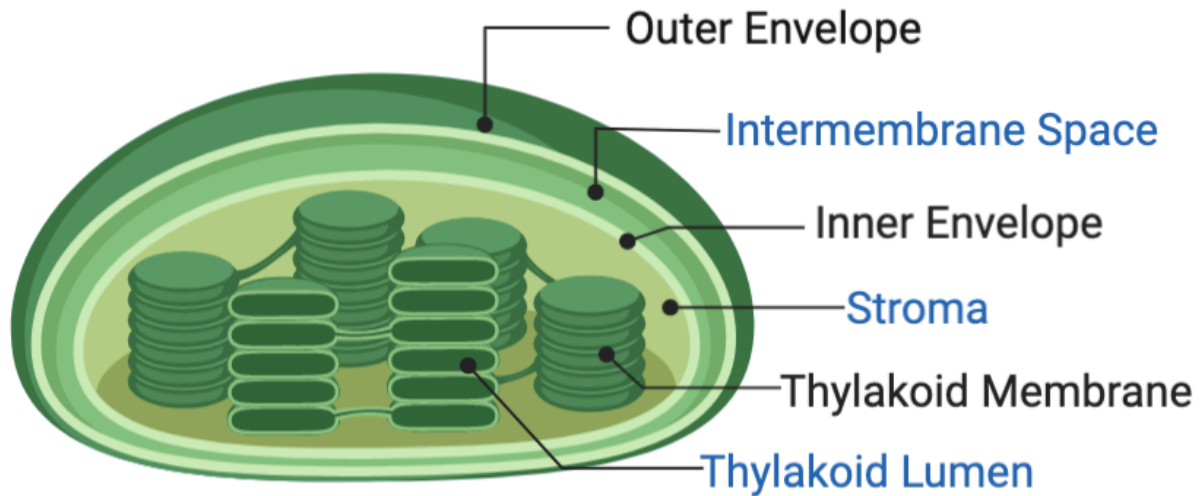


Figure 1.1. The General Chloroplast Structure. The chloroplast, composed of three lipid bilayers and three compartments, is enclosed by an outer envelope and inner envelope separated by the intermembrane space. The IEM contains the stroma and thylakoid membrane, and the TM contains the thylakoid lumen. Created with BioRender.com.

1.2. Plastid-types and Plastid-Proteomes

Many types of plastids have evolved, each has a different function, and most are tissue specific. Every plastid begins as an undifferentiated precursor plastid called a proplastid, which can transition to and between specific plastid-types when triggered by developmental or environmental cues (Figure 1.2; Chu & Li, 2018; Jarvis et al., 2013). Each plastid-type has a different proteome and internal membrane structure (Jarvis et al., 2013). Moreover, specialized tissues contain specific plastid-types with unique proteomes and pathways to meet

the tissues' individual needs. For example, the bright pigmentation of carotenoids, synthesized in the chromoplasts of fruits, entice vectors for seed dispersal (Ling et al., 2012). Proteins are selectively imported into different plastid-types, which ultimately dictates the plastid proteome. Protein-selective import is facilitated by OEM-receptors which differentially transport specific protein sets. Moreover, OEM-receptors are differentially expressed in each plastid-type and regulate specific changes in protein import that are essential for age and tissue-specific function and development (Chu & Li, 2018).

Dynamic protein-import regulation is particularly important during the biogenesis of plastids in germination and early developmental stages (Thomson et al., 2020). Tight control of protein-import is also essential to rapidly shift plastid proteomes in response to environmental fluctuations, such as sudden light exposure. Additionally, plastids in young and dividing tissues have higher protein demands and requirements than plastids in adult and non-dividing tissues. These many dynamic changes require plastids to constantly acclimate and alter their proteome throughout their life-cycle (Chu & Li, 2018; Sjuts et al., 2017; Yang et al., 2019).

Many regulatory mechanisms of plastid-protein import exist, including regulation via peptide-receptor interactions at the OEM. Protein receptors and channels embedded in the OEM recognize and/or aid in transport of specific plastid-localizing protein groups, thus, ultimately regulating the plastid-proteome by dictating protein translocation (Ling et al., 2012; Schnell, 2019). These OEM receptors evolved following lateral gene transfer from the endosymbiont to enable subcellular plastid-protein targeting (Sjuts et al., 2017).

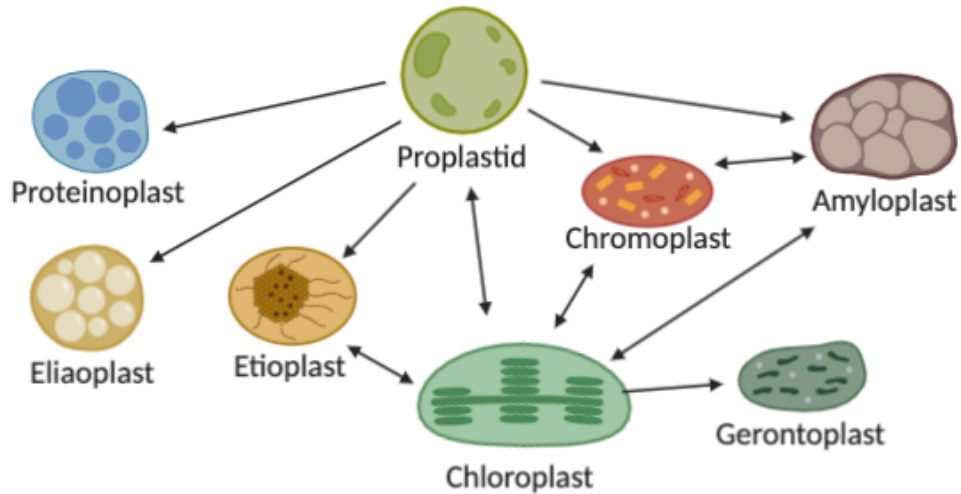


Figure 1.2. Plastid-type Transition Network. Arrowheads represent a possible plastid-type transition. All plastids begin as a proplastid and can transition between types based on environmental and developmental cues. Chloroplasts perform photosynthesis; etioplasts rapidly transition to chloroplasts upon light exposure; gerontoplasts are senescent chloroplasts; chromoplasts contain carotenoid pigments; elaioplasts, proteinoplasts, and amyloplast are storage organelles which store lipids, proteins, and starch, respectively (Jarvis & López-Juez, 2013). Created with BioRender.com.

1.3. Plastid-localized Protein Transcription, Translation, and Translocation

Following lateral gene transfer of chloroplast genetic material to the nucleus, the host cell evolved mechanisms to transcribe these genes and translate the encoded proteins in the cytosol (Day & Theg, 2018). Upon translation, proteins need to be efficiently translocated to plastids to prevent the accumulation of these proteins in the cytosol. Moreover, the soluble cytosolic environment can cause membrane associated proteins to mis-fold and aggregate (Kim et al., 2019). In order to maintain proteostasis during translation and transport the host cell evolved complex methods of plastid-protein translocation. Proteostasis is maintained by cytosolic factors that create a physiochemical environment which kinetically favours protein-

translocation (Kim et al., 2019). Cytosolic factors play a key role in the specific and efficient import of most plastid-proteins by recognizing and sorting them to the correct chloroplast-targeting pathway (Sjuts et al., 2017).

The majority of plastid proteins are post-translationally translocated to the chloroplast using the general chloroplast-import pathway. 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). Chaperone proteins in the cytosol recognize and bind particular TP motifs then carry preproteins to a specific OEM-receptor. Chaperone proteins in the heat shock protein (Hsp) 70 family and Hsp90 family, as well as protein 14-3-3, transport 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). The TOC and TIC (translocon at the inner envelope of chloroplasts) complexes function together as a super-complex that shuttles proteins across the chloroplast double membrane (Schnell, 2019).

The TOC complex is assembled from protein receptors and a beta-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 voltage-gated 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 the core component of the TIC complex (Schnell, 2019). The TOC and TIC complexes are physically linked by Tic236 which forms a super complex (Chen et al., 2018). 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). 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).

Different models for the import motor complex assembly have been proposed and the role and importance of each component is highly debated (Li et al., 2020). Proteins of the import motor complex also behave as chaperone proteins which fold and maintain the integrity of preproteins during import. Once preproteins are imported to the stroma, the TP is cleaved by stromal processing peptidase (SPP). The cTP is defined as the region that will be cleaved by SPP in the stroma, however more processing and regulatory information may lie elsewhere in the mature protein (Sjuts et al, 2017). Additional stromal factors, such as Cpn60, are responsible for preprotein folding and processing to form the mature protein. The mature

protein can assemble in the stroma or be sorted to the IEM, the TM, or the TL through suborganellar targeting pathways (Figure 1.3; Sjuts et al, 2017). These additional pathways include the cpSRP pathway, cpSec1 pathway, and the twin arginine translocase pathway (Day & Theg, 2018).

Some details within this description of the general import pathway are not completely accepted. For instance, it is highly debated whether the core component of the TIC complex is Tic20 or Tic110 (Bölter, 2016). Nevertheless, most steps in import are generally agreed upon and this account is sufficient for our purposes.

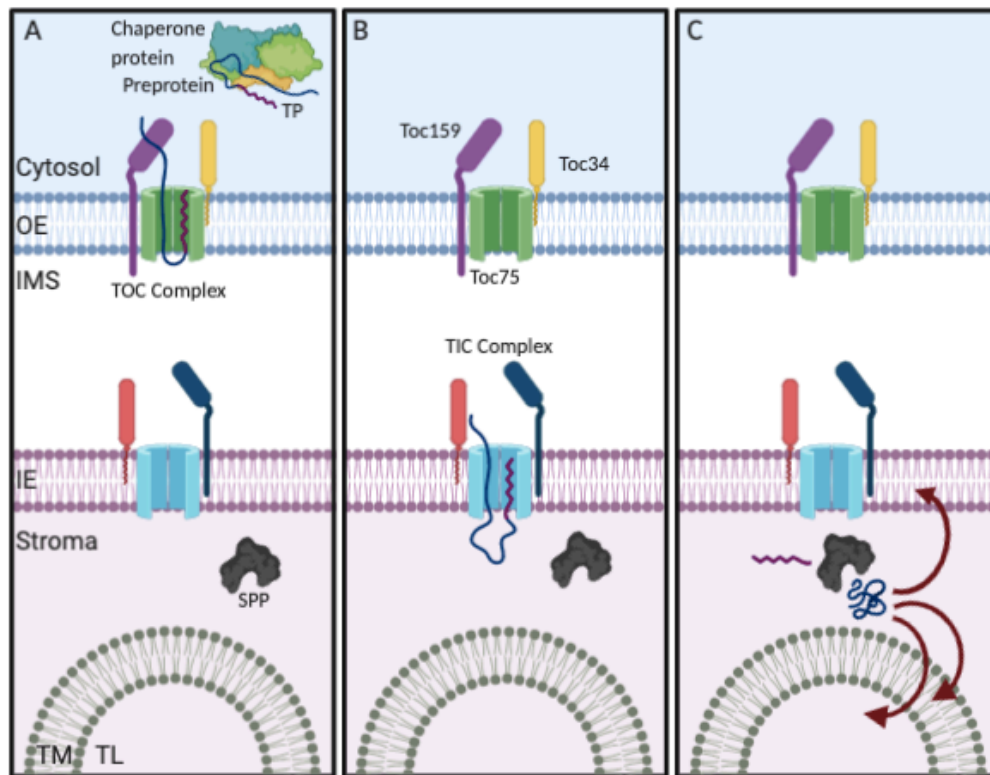


Figure 1.3. Import of Preproteins across the Chloroplast Double Membrane. At the OEM, preproteins are bound by Toc159 and Toc33 then transported through the beta-barrel protein Toc75 (A). Preproteins are then passed through the TIC complex (B). Once in the stroma, the TP is cleaved by SPP and the mature protein is formed. The mature protein can remain in the stroma or be transported to the IEM, TM or the TL (C) (Sjuts, et al 2017).

1.4. Features of N-terminal Transit Peptides

Peptide sequences are used as subcellular localization signals to target proteins to specific subcellular compartments, some examples include the nuclear localization sequence and endoplasmic reticulum retention signal. Many of these localization signal peptides are defined by a consensus sequence making them easy to predict, however, this is not the case for chloroplast TP sequences. Chloroplast TPs have highly divergent sequences which makes it difficult to predict them using the protein sequence alone (Lee & Hwang, 2018; Patron & Waller, 2007). Subgroups of TP-types have been identified which share some common features; however, no single TP has every feature (Lee & Hwang, 2018).

The number of currently known TP features continues to grow and includes: an abundance of K/R, S/T, and P residues, a lack of acidic residues, moderate hydrophobicity, amphipathic alpha helices under mimetic conditions, and conserved sequence motifs (Bruce, 2001; Lee et al., 2008; Lee & Hwang, 2018; Patron & Waller, 2007). The presence of basic residues, lysine (K) and arginine (R), and a lack of acidic residues creates an overall net positive charge. When basic residues are substituted for inert alanine residues, preprotein import efficiency decreases. An abundance of proline (P) residues is thought to create rigid structures in the TP that function as a targeting-signal. Additionally, P residues may interact with stromal motor import proteins during late stages of import and increase the efficiency of import (Lee & Hwang, 2018). TP sequences with moderate hydrophobicity can interact with Hsp70, as such, hydrophobicity enables efficient OEM-recruitment by the guidance complex (Lee & Hwang, 2018). The structural of TP amphiphilic alpha helices is fluid, these alpha helices form stable membrane-associations under mimetic conditions, yet, relax into stable coils in hydrophilic

environments (Lee et al., 2008). Subgroups of conserved sequence motifs are common within specific classes of preproteins. These common motifs likely coordinate the import of proteins that require simultaneous expression (Bruce, 2001; Lee et al., 2008). Serine (S) and threonine (T) residues are found across many TP subgroups, their phosphorylation and dephosphorylation regulates and increases 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). Subsequent dephosphorylation of TPs at the OEM is required to maintain efficient import through the TOC complex. Failure to dephosphorylate preproteins results in extremely slow import. This S/T phosphorylation cycle is not essential for the import process however; it has been reported to increase import efficiency by several fold in some cases (Bölter, 2018). Moreover, this phosphorylation cycle can be rapidly modified to quickly adjust preprotein import, allowing the cell to rapidly acclimate the chloroplast proteome in response to environmental stressors (Sjuts et al, 2017).

Preproteins can be predicted by examining the protein sequence hydrophobicity, residue representation, phosphorylation sites, and conserved motifs, as well as, by analyzing the secondary structure for amphipathic alpha helices (discussed in section 1.14). Additionally, preprotein TPs can be identified by probing for the conserved site where the enzyme, Signal Processing Peptidase (SPP), cleaves the TP from the preprotein. The SPP cleavage site is not always considered part of the TP as it does not function in import. Yet, it remains an important feature for correct processing and is therefore a useful predictive feature for identifying TPs (Lee et al., 2008). The transit peptide prediction software ChloroP can identify some targeting

features within the N-terminus of protein sequence and predict the presence or absence of a TP (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 remains a relevant and important tool for understanding preprotein import (Emanuelsson et al., 2007; Bouchnak et al., 2020).

To explain the numerous underlying mechanisms which enable this large diversity in TP identity, Li & Teng (2013) proposed the multi-selection multi-order (MM) model. The MM model suggests that TPs are assembled from numerous motifs that interact with distinct molecular factors during the import pathway. Also, TP motifs are used for preprotein quality control and import regulation (Lee & Hwang, 2018). These numerous motifs are arranged in long N-terminal peptide extensions and show little preference for order and location. In many cases the sequence of a single motif is not well-conserved and can appear highly variable between different TPs (Lee & Hwang, 2018). Intriguingly, functional TP hybrids can be synthesized by fusing individual motifs from different TPs together, demonstrating that seemingly dissimilar motifs have interchangeable functions. Additionally, TP motifs are necessary and sufficient for protein targeting, as their fusion to non-chloroplast proteins results in stromal-localization (Lee & Hwang, 2018). The large diversity in TP features makes it difficult to predict TPs using a chloroplast protein sequence. However, this diversity is the underlying mechanism that regulates complex and dynamic preprotein import (Lee & Hwang, 2018; Li & Teng, 2013; Patron & Waller, 2007).

1.5. The Diverse Function and Structure of TOC Complexes and their Receptors

Preproteins with highly diverse N-terminal TPs are transported to the chloroplast in a tightly regulated manner. OEM receptors must recognize this wide range of substrates and selectively import preproteins into the chloroplast. Toc receptors provide the TOC translocon with the fidelity, specificity, and selectivity required for complex and dynamic preprotein import (Schnell, 2019). Different Toc receptor isoforms assemble in structurally and functionally diverse translocons, each assembly has selectivity for specific NT TPs. These Toc complexes recognize distinct classes of preproteins and differentially regulate import in response to developmental cues and environmental stressors (Chu & Li, 2018). Toc receptors use their TP recognition sites and GTPase activity to regulate preprotein access to the translocon machinery (Schnell, 2019).

The general Toc complex assembly includes a Toc159 family member, a Toc34 family member, and Toc75-III, in a reported stoichiometry of 1:4:4, respectively (Sjuts et al., 2017). Toc159 family members include: Toc159, Toc132, Toc120, and Toc90, while Toc34 family members include, Toc33 and Toc34 (Bölter, 2018). Toc159 isoforms have an acidic (A) domain, a GTPase (G) domain, and a membrane (M) domain. The A-domain is highly variable between different isoforms and aids in the selective and specific binding of preproteins (Thomson et al., 2020). The M-domain anchors Toc159 to the OEM membrane and possibly extends into the IMS. The portion of the M-domain located within the IMS is in close proximity to TPs during the initial reversible stage of preprotein import which has led some to speculate that the M-domain also plays a role in import (Kouranov & Schnell, 1997). Toc159 family members prefer a monomeric conformation *in vivo*, at resting state Toc159 binds to GTP. On the other hand,

Toc34 isoforms prefer a homodimerized state, creating a nucleotide-binding pocket at the dimer interface; at rest this pocket is GDP-bound (Schnell, 2019). In their active states, Toc159 and Toc34 simultaneously bind the preprotein N-terminus and C-terminus, respectively (Thomson et al., 2020). Preprotein binding causes the Toc34 homodimer to dissociate, which in turn, stimulates the exchange of GDP for GTP. Then, Toc34 and Toc159 hydrolyze bound GTP using their GTPase domains. GTP-hydrolysis prompts preprotein association with stromal chaperone proteins and induces a conformational change in receptors which releases the bound preprotein (Schnell, 2019). This recent model of receptor-preprotein binding conformation is debated and some propose a dimer-interactions occur between Toc159 proteins and/or between Toc159 and Toc34 proteins (Chang et al., 2017). Further research is required to resolve the import function and assembly of the TOC complex.

It is thought that various Toc159 and Toc34 receptor isoforms assemble in distinct translocons that selectively regulate preprotein access to the translocon machinery. For example, a translocon assembled from Toc159, Toc33, and Toc75-III will transport preproteins that are directly or indirectly used in photosynthesis. Conversely, the Toc132 or Toc120 receptor assembles with Toc34 and Toc75-III to transport preproteins encoded by housekeeping genes. Thus, the cell can regulate chloroplast biogenesis and general preprotein import by regulating the presence/absence of specific translocon assemblies (Sjuts et al., 2017). This diverse translocon theory was formulated around several observations, the most prominent being the phenotype observed in the Toc159 knockout mutant, *ppi2*. *ppi2* mutants are albino, seedling lethal, and cannot be rescued by Toc132 or Toc120 but can be partially rescued by Toc90. This suggested Toc159 is fundamental for photosynthetic preprotein import

and supported the theory that diverse translocon assemblies recognize specific classes of preproteins. However, transcript analysis of *ppi2* mutants grown in sucrose revealed a significant decrease in the expression of photosynthetic genes; suggesting protein deficiency is partially a result of downregulated gene expression and not defective import (Bischof et al., 2011). Moreover, studies have shown an equal number of photosynthetic and non-photosynthetic preprotein will bind to both Toc159 and Toc132 (Sjuts et al., 2017). Therefore, the model of diverse translocon assembly may be oversimplified and must be revisited to resolve these discrepancies. It is suggested that these diverse Toc assemblies selectively and differentially import preproteins in response to developmental cues plastid-age specific cues, tissues-specific cues, and environmental stressors (Chu & Li, 2018; Sjuts et al., 2017; Richardson et al., 2014).

1.6. Chaperone-Receptor Interactions & Toc Receptor Regulation

Prior to preprotein import, Toc receptors behave as OEM docking sites for chaperone proteins bound with preproteins. Toc receptors are the predominant docking site for the guidance complex, however, there are other chaperone proteins and receptors that are also involved in preprotein recruitment to the OEM (Bölter, 2018). For example, Hsp90 can complex with Hsp70 to bind specific preproteins and shuttle them to the OEM-receptor Toc64. Toc64 then passes the preprotein to either Toc33 or Toc34 for subsequent import through the TOC/TIC complex. Toc64 overlaps in function with Toc33 and increases import efficiency. However, Toc64 is not essential for import, implying the TOC complex contains sufficient components for preprotein import across the OEM (Bölter, 2018).

Import of preproteins can also be regulated by phosphorylating specific Toc receptors. Toc159 isoforms are phosphorylated within their A-domain by KOC1 (Yang et al., 2019). Each Toc159 isoform has a distinct phosphorylation pattern due to the divergence between the A-domain sequences. Toc33 can also be phosphorylated, however, Toc34 is only phosphorylated in certain species, such as in *Pisum sativum* (Sjuts et al, 2017). KOC1 null mutants have impaired preprotein import, demonstrating that Toc159 phosphorylation plays an important role in import (Schnell, 2019). Conversely, other studies have shown that Toc33 phosphorylation impairs preprotein import (Sjuts et al, 2017). Phosphorylation is also speculated to play a role in Toc receptor degradation, however, this hypothesis remains unverified. Thus, the importance phosphorylation plays in preprotein import regulation is complex and requires further investigation (Sjuts et al., 2017).

1.7. UPS regulation of TOC complexes and Preproteins

Another method of regulating TOC expression level is via the ubiquitin-proteasome system (UPS). The UPS uses ubiquitination and UPS protein factors to selectively target translocon assemblies for degradation by proteosomes (Thomson et al., 2020). The UPS not only regulates the degradation of specific translocon assemblies, it also targets unimported preproteins for degradation (Yang et al., 2019). Ubiquitylation is the ligation of a ubiquitin protein to a lysine residue on a receiving protein. The 8.5 kDa ubiquitin tag will target Toc-receptors for degradation by the 26S proteasome which works in two different UPS pathways, the CHLORAD and DELLA/GA pathways (Thomson et al., 2020).

The chloroplast-associated degradation (CHLORAD) UPS pathway was first discovered in leaves and is crucial for chloroplast development in seedlings and stress responses (Schnell, 2019). The CHLORAD system involves 4 proteins, ubiquitin, SP1 (suppressor of *ppi1* locus 1 E3 ligase), SP2 (an outer membrane OMP85 family protein), and Cdc48 (cell division control protein 48). SP1 is an E3 ubiquitin ligase protein which marks Toc receptors for degradation. Ubiquitinated receptors are released from the membrane by the β -barrel retrotranslocon SP2 using motor energy from the ATPase, Cdc48. Once in the cytosol, ubiquitinated Toc receptors are targeted to the 26S proteasome for degradation (Yang et al., 2019; Thomson et al., 2020). SP1 co-immunoprecipitates with all Toc receptor isoforms, thus, the CHLORAD pathway can potentially regulate the expression level of all TOC components (Thomson et al., 2020).

The UPS DELLA/GA pathway also regulates TOC complex degradation and is crucial during germination. During chloroplast biogenesis, the import of photocomplexes is tightly regulated to avoid imbalances that generate phototoxic aggregates and harm the plant in early life (Schnell, 2019; Thomson et al., 2020). Cytosolic regulator proteins in the DELLA family, bind and target Toc159 for degradation, ultimately preventing the formation of unwanted photosynthetic complexes. The production of gibberellic acid (GA) decreases levels of DELLA proteins enabling Toc159 to accumulate and assemble with Toc75-III and Toc33. Consequently, the assembly of the Toc159/Toc33 translocon allows the import of photosynthetic preproteins to ensue. On the other hand, housekeeping proteins are not regulated by the GA/DELLA system; thus, it seems unlikely that other Toc receptors are widely regulated this way (Schnell, 2019; Thomson et al., 2020).

Cytosolic UPS pathways also target preproteins for degradation. Unfolded preproteins are highly prone to forming toxic aggregates and therefore must be degraded by proteosomes. When the preprotein is unfolded, specific motifs are targeted for degradation by various 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 unimported photosynthetic preproteins for degradation to prevent premature chloroplast biogenesis in etioplasts (Schnell, 2019).

Diverse interactions between Toc receptors and cytosolic chaperones, Toc receptor phosphorylation patterns, and UPS pathways are just some regulatory mechanisms which add layers of complexity to an already multifaceted import system. It is this complexity that allows for tight, dynamic, and highly specific control of preprotein import (Thomson et al., 2020).

1.8. Targeting Mechanisms Used by Plastid Outer Envelope Proteins

With the exception of Toc75-III and Toc75-V, preproteins with TPs are directed to the stroma and cannot be diverted to the OEM (Gross et al., 2020). Thus, outer envelope proteins (OEPs) cannot use the general import pathway and require a different localization mechanism. All OEPs are transcribed in the nucleus and translated on cytosolic 80S ribosomes (Kim et al., 2019). In some cases, translated OEPs are bound by cytosolic factors that assist in OEP proteostasis and OEM-targeting. Cytosolic factors create a physiochemical environment that maintains the preproteins capacity for import. Cytosolic factors maintain import competence by binding hydrophobic regions of the OEP that interact unfavourably with the cytosol and

prevent protein aggregates from forming (Kim et al., 2019). Cytosolic factors can also aid in protein transport to the OEM. Generally, it is thought that OEPs are bound and transported by cytosolic factors via diverse localization-signals embedded within OEP primary and secondary structures (Lee et al., 2017).

The localization mechanisms used by many OEPs are uncharacterised, due in large part to the limited number of known OEPs and the difficulties associated with transmembrane protein analysis. Due to this gap in knowledge, we are limited in our ability to engineer plastid proteomes and manipulate protein import to the chloroplast (Anderson et al., 2019). However, advances in proteomics and protein analysis techniques has led to the identification of an increasing number of OEPs (Bouchnak et al., 2019; Inoue et al., 2015). OEPs can be divided into two structurally diverse groups which include: β -barrel proteins and helical transmembrane domain (TMD) proteins. Helical TMD proteins can be classified in four distinct structural subgroups. In two subgroups, the TMD is a single alpha helix located at either the N-terminus or C-terminus, named signal-anchored and tail-anchored proteins, respectively. A third subgroup includes proteins with multiple alpha helical TMDs. Lastly, the TMD of proteins in fourth subgroup contains both alpha helices and β -sheets, named CT TP-like proteins (Lee et al., 2014; Lung et al., 2014). Each group and subgroup of OEPs utilize different mechanisms for OEM recruitment and integration. Currently, there are four well established OEP-localization strategies including: an N-terminal TP used by Toc75-III and Toc75-V, β -barrel self-insertion, signal-anchor mediated insertion, and tail-anchor mediated insertion (Figure 1.4; Kim et al., 2019; Lee et al., 2017).

1.9. Recruitment and Integration of β -Barrel Proteins at the Plastid Outer Envelope

β -barrel proteins are formed from 8-24 β -sheets which create a hydrophilic membrane pore (Tsaousis et al., 2017). Many are transporter channels that recognize and translocate specific substrates, including, small ions, molecules, peptides, nucleic acids, and proteins. β -barrel proteins are also involved in cellular signalling, organelle interactions, apoptosis, and many other important 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 is made evident by their primary and 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 (OM) *in vivo* (Jones & Rapaport, 2017). Furthermore, mitochondrial β -barrel proteins can target the gram-negative PM, suggesting these channels have some conserved targeting function (Jones & Rapaport, 2017). Intriguingly, 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, however, evidence of this has not been experimentally verified (Jones & Rapaport, 2017). Generally, β -barrel proteins are highly diverse in their primary sequence, yet, highly conserved in their secondary structures. Therefore, it is more likely that a conserved OEM targeting-signal is found within the β -barrel secondary structure and not the primary sequence. Moreover, a

hydrophobic C-terminal β -hairpin was shown to be necessary and sufficient for β -barrel targeting to the mitochondria OM, demonstrating the targeting-signal for OM β -barrel proteins lies within the secondary structure (Jones & Rapaport, 2017). Furthermore, fusing this hydrophobic β -hairpin to the chloroplast β -barrel proteins OEP37 and OEP24, resulted in mistargeting to the OM. Chloroplast β -barrel proteins also have a C-terminal β -hairpin motif; however, it is not sufficient for import. Thus, the specific import mechanism used by chloroplastic β -barrel protein appears to be more complex and requires further study (Jones & Rapaport, 2017).

Chloroplast β -barrel proteins are imported to the OEM post-translationally. The hydrophobic β -barrel proteins interact unfavourably with the cytosol which causes the formation of toxic protein aggregates. As such, maintaining proteostasis in the cytosol is crucial for protein import. It is speculated that β -barrel proteins use cytosolic chaperone proteins to maintain proteostasis during OEM import (Kim et al., 2019). However, there is no experimental evidence to support this notion. Recent advances suggest chloroplast β -barrel proteins use distinct targeting signals and import pathways (Gross et al., 2020).

In vitro experiments have shown some chloroplast β -barrel proteins can facilitate their own insertion into the OEM (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 following import, suggesting an N-terminal TP signal is either not used or not cleaved. Further investigation is required to determine the import mechanism of OEP24 and OEP37 (Jones & Rapaport, 2017; Kim et al., 2019). It has long been established that Toc75-III uses a bipartite TP

containing two distinct elements, a classical N-terminal TP & a glycine rich (GR) region (Kim et al., 2019). More recently, it was established that OEP80/Toc75-V (from now referred to as Toc75-V) also uses an N-terminal signal-peptide. However, these Toc75 signals are highly dissimilar and 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 drags the GR-region into the IMS where it becomes detained. SPP proceeds to cleave the TP in the stroma, then, the GR-region is cleaved by type I signal-peptidase in the IMS. This diverts Toc75-III from the general import pathway and inserts it in the OEM (Richardson et al., 2014). The import of Toc75-III can be competitively inhibited by preproteins which implicates the general pathway in its import. 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 machinery (Jones & Rapaport, 2017). However, it is unknown if chaperone proteins like Hsp70, Hsp90, or 14-3-3 are capable of recruiting Toc75-III to the general import apparatus (Kim et al., 2019). Additionally, evidence suggests that Toc75-V forms a translocon which can integrate 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 a bipartite TP and thus, may not utilize the general import pathway. When Toc75-V was initially analyzed for a TP the 52 most N-terminal residues appeared to be dispensable for targeting, resulting in the dismissal of an N-terminal TP-signal (Gross et al., 2020). Yet, more recent studies provide evidence that Toc75-V uses an N-terminal TP which is necessary and sufficient

for targeting (Gross et al., 2020). Following 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 distinct cleavable N-terminal localization-signals and OEM import-pathways (Day et al, 2019; Gross et al., 2020).

1.10. Recruitment and Integration of Signal-Anchored Proteins at the Plastid Outer Envelope

OEPs using signal-anchor (SA) mediated insertion have an N-terminal TMD that anchors the protein to the OEM leaving the C-terminus exposed to the cytosol (Inoue, 2015). Many SA-proteins are protein-receptors, such as Toc64 and OEP14, and are generally found in eukaryotic cellular membranes (Lee et al., 2014). SA-proteins specifically target the chloroplast OEM using a non-cleavable TP-signal which includes, an alpha-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 may act as a deterrent for ER mis-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. In the case of Toc64, exchanging basic residues for inert glycine residues results in mistargeting to the plasma membrane. Therefore, the hydrophobicity of the TMD and charge of the CPR are essential to maintain specific OEM targeting (Kim et al., 2019; Lee et al., 2014).

SA-proteins require a cytosolic factor called the ankyrin repeat-containing protein 2 (AKR2) for recruitment and integration into the OEM. AKR2 translationally targets SA-proteins

to the OEM by binding the SA-protein N-terminus as it emerges from the ribosome exit tunnel (Kim et al., 2019). There are two AKR2 isoforms, AKR2A and AKR2B both transport SA-proteins to the chloroplast and maintain proteostasis. AKR2 binds the hydrophobic regions of the TMD to prevent unfavourable interactions with the cytosol and the formation of non-specific aggregations. After AKR2 binds its cargo, dimerized heat shock protein sHsp17.8 binds AKR2 and facilitates targeting to the chloroplast (Kim et al., 2011). Once at the OEM, AKR2 uses an MGDG (monogalactosyldiacylglycerol) lipid and a PG (phosphatidylglycerol) lipid as a docking site to unload its cargo. After successful OEM docking, SA-proteins bound to AKR2 are integrated into the OEM by Toc75, however, the exact mechanism which facilitates this integration is unclear (Kim et al., 2019).

1.11. Recruitment and Integration of Tail-Anchored Proteins at the Plastid Outer Envelope

Chloroplast OEM tail-anchored proteins have a $N_{out}-C_{in}$ topology. They contain three sequentially ordered features, including: a positive C-terminal sequence (CTS), an alpha-helical TMD membrane-anchor, and a CT-tail with a maximum length of 50aa (Zhuang et al., 2017). TA proteins are common to all eukaryotes and some prokaryotes. They are found in the majority of cellular membranes and maintain diverse and important functions, including but not limited to, protein translocation, membrane fusion, vesicle-trafficking, electron transport, apoptosis, and protein quality control. Moreover, important OEM receptors, such as Toc33 and Toc34 are tail-anchored proteins (Teresinki et al., 2019). Therefore, deducing the targeting features which facilitate chloroplast TA-protein targeting will elucidate mechanisms that direct receptors to the TOC complex assembly (Kim et al., 2019).

The targeting signals of TA-proteins contain four physiochemical features; however, 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 series of basic residues called a CTS, an alpha-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 a wider range of 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 will contribute to the specific subcellular localization of many TA-proteins and are especially important for ER TA-protein integration (Teresinki et al., 2019). Chloroplast TA-proteins use a CTS with a net positive charge, however, the net charge matters less 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 9aa long, contains at least 3 K or R residues and 3 S or T residues, and can be located anywhere in the CTS. 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 amongst TA-proteins which harbour them. Therefore, the distribution of charges in the RK/ST is likely more important than the overall charge, which is commonly seen in CTS regions (Teresinki et al., 2019).

Despite the TA-protein 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 (Teresinki et al., 2019). On the other hand, TA-proteins Toc33 and Toc34 require a GTPase domain, a TMD, and a CTS for sufficient 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). It is likely 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, in 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 then post-translationally targeted to plastids by cytosolic chaperone proteins (Kim et al., 2019; Lee et al., 2017). Chaperone proteins AKR2, Hsp70, and Hsp90 transport chloroplast TA-proteins to the OEM. Hsp70 and Hsp90 increase the efficiency of TA-targeting but not fidelity of targeting, suggesting they aid AKR2 and other cytosolic factors in targeting and cannot act independently (Kim et al., 2019). The presence of multiple TA-protein localization pathways has led to speculation that multiple chaperone proteins recognize distinct TA-protein targeting signals. ER-localized TA-proteins are targeted and integrated using the GET pathway and GET proteins. Recently, GET homologs have been identified in plant and algal groups (Zhuang et al., 2017). Notably, the GET homolog ArsA1 in

Chlamydomonas reinhardtii has been implicated in targeting TA-proteins to the endosymbiotic organelles. However, this relationship remains to be resolved in model organisms for terrestrial plants. Coordination between ArsA1 and AKR2 may occur but remains to be determined (Lee et al., 2017). Not all TA OEPs need cytosolic factors for OEM-targeting and translocation. In some cases, 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).

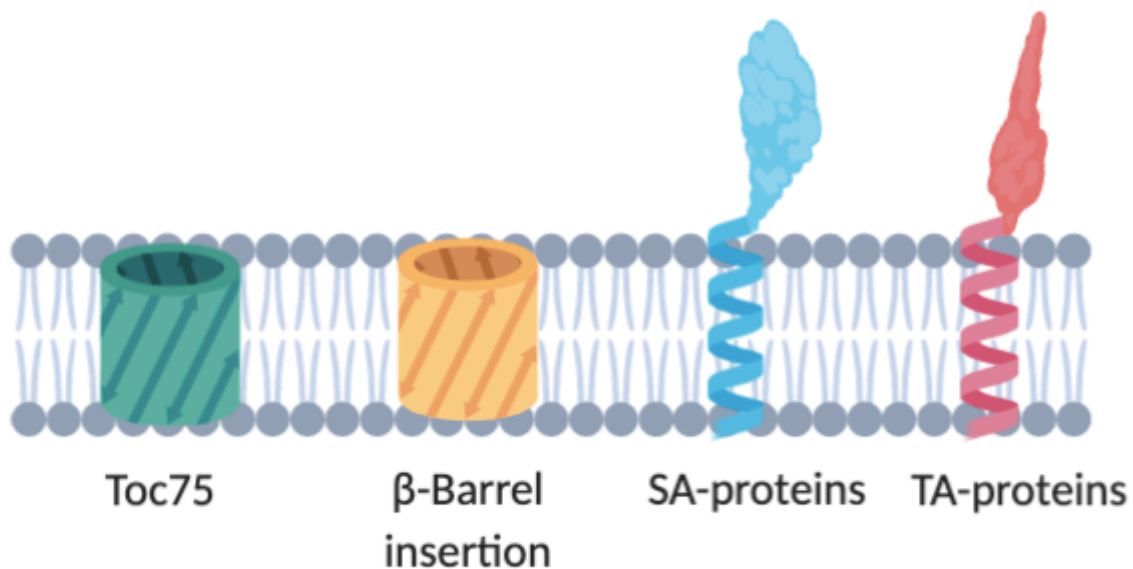


Figure 1.4. Outer Envelope Protein Targeting Mechanisms. The four OEM targeting mechanisms include a modified N-terminal TP used by Toc75-III and Tov75-V, β -barrel self-insertion, Signal-anchored mediated insertion (SA-signal) and tail-anchored mediated insertion (TA-signal). This image was generated using BioRender.

1.12. The Role and Function of Chaperone Protein ARK2

The chaperone protein AKR2 has two isoforms, AKR2A and AKR2B, both isoforms can bind chloroplast TA-proteins and SA-proteins. However, AKR2A has been the focus of most study. AKR2A is a cytosolic chaperone protein which maintains proteostasis of its cargo by binding hydrophobic and charged regions in the TMD and CTS/CPR, ultimately preventing protein aggregation. The import efficiency of AKR2A increases in the presence of the cofactor sHsp17.8, which can directly bind both AKR2A and OEM lipids (Kim et al., 2011; Lee et al., 2014).

AKR2A will dock at the RPL23A ribosomal site when SA- and TA-proteins are translated by the 80S ribosome in the cytosol. As the TA-protein targeting-signal emerges from the ribosome exit tunnel, the N-terminal ankyrin repeat domains (ARD) of AKR2A will recognize and bind the unfolded protein (Kim et al., 2019). After AKR2A binds its cargo it is recruited to the chloroplast through some unknown mechanism. Once AKR2A reaches the OEM, an MGDG-lipid and a PG-lipid function as a protein-docking station via interactions with two lipid binding-pockets formed by the C-terminal ARD domains of AKR2A. MGDG lipids are specific to the chloroplast membrane, thereby, functioning as an organelle-specific marker and providing a mechanism for chloroplast targeting fidelity. The AKR2A PG-lipid interaction is a prerequisite for MGDG-lipid binding and tightens AKR2A's interaction with the OEM. The synergistic and coincidental binding of MGDG and PG lipid heads by AKR2A facilitates a tight and specific interaction with the chloroplast OEM. The two AKR2A lipid-binding pockets are formed by several ARDs located within the C-terminus and this structure has been determined by X-ray crystallography (Kim et al., 2014). Computational and mutation studies identified the residues which directly interact with lipid-heads in the pockets. Residues E246 and H223 inhabit the L₁

pocket and specifically bind an MGDG lipid head. Residues Y294 and R296 are situated in the L₂ pocket and specifically bind a PG lipid head (Kim et al., 2014). Additionally, there are a large number of aromatic residues on the protein face that surrounds the L pockets, which is a common characteristic of a protein surface that interacts with a membrane (Kim et al., 2014). The mechanism by which AKR2A off-loads its proteins at the OEM for integration is unknown. Once proteins are unloaded by AKR2A, they can be assimilated into the OEM by a number of mechanisms, depending on the protein identity and the context provided by the targeting signal. Toc75 can play a role in assimilating some AKR2A cargo proteins, however, other proteins such as Toc33 and Toc34 are capable of self-insertion, still others like OEP9.1, require unknown protein factors for import (Lee et al., 2013; Teresinski et al., 2019). After AKR2A unloads its cargo it must be released from the membrane to continuously target proteins. However, the mechanism of AKR2A membrane release remains to be determined. It is speculated that the lipid-head binding-pocket interaction is disrupted, destabilizing the AKR2 membrane-association and allowing its release (Kim et al., 2019).

During AKR2 evolution, domain functions were acquired from the host cell and endosymbiont; creating a protein which is functional in eukaryotic plant cells, yet, has plastid targeting fidelity. The N-terminal ARDs which bind cargo-proteins evolved from the eukaryotic host cell and include, the PEST, C1, and C2 domains. C1 and C2 domains directly bind cargo-proteins while the PEST sequence stabilizes the interaction (Kim et al, 2014). The C-terminal ARD of AKR2A originated from the host but was adapted to specifically target plastid lipids. ARDs are common protein-protein interacting domains found in ~6% of all eukaryotic protein sequences (Kim et al, 2014). ARDs are used in fundamental cellular processes, including,

cytoskeletal organization, cell signaling, transcriptional regulation, the inflammatory response, cell cycle regulation, and cell development/differentiation (Kim et al, 2014). During AKR2A evolution the ARD evolved from a protein-protein interacting domain to a protein-lipid interacting domain using the context provided by endosymbiont lipids (Kim et al, 2014). The ARDs evolved binding-capacity for the plastid-specific MGDG-lipid and the PG-lipid, providing plastid targeting context (Figure 1.5). This adaptation most likely occurred during organellogenesis of the endosymbiont (Kim et al, 2014).

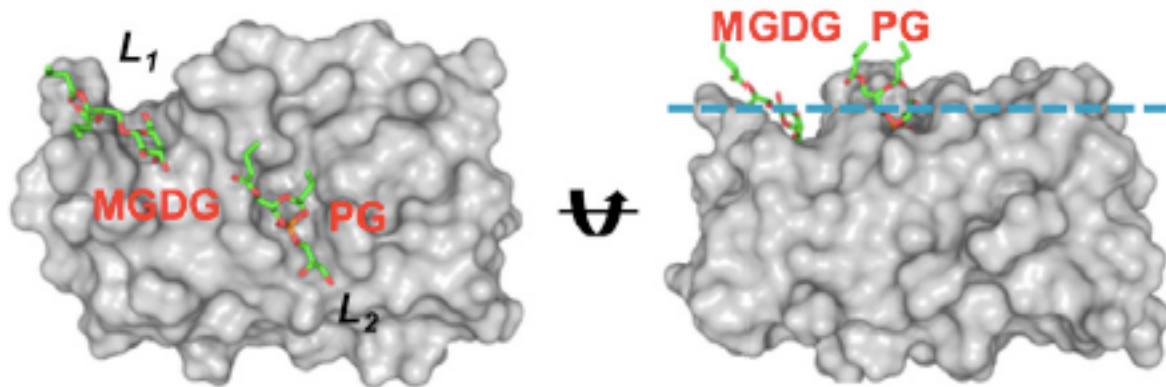


Figure 1.5. Structure of ARK2A Lipid-Binding Pockets L_1 and L_2 . This figure was adapted from Kim et al. (2014). ARD structure as determined by X-ray crystallography. Two three-dimensional surface representations of ARK2A depicts lipid-binding pockets denoted by L_1 and L_2 and their respective lipid interactions with either MGDG or PG. Each model illustrates a different surface perspective. The blue dashed line divides the lipid interacting surface (upper) from the remainder of the protein (lower). The arrow denotes the rotation required to achieve the model on the right from the model on the left.

Figure 1.5. Structure of AKR2A Lipid-binding Pockets L_1 and L_2 . Reprinted from An Ankyrin Repeat Domain of AKR2 Drives Chloroplast Targeting through Coincident Binding of Two Chloroplast Lipids by 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. Retrieved from Cell and Developmental Biology. Copyright 2014 by Elsevier Inc.

1.13. A Potentially Novel OEP Targeting Mechanism

While some mechanisms for targeting β -barrel, SA-, and TA-proteins to the chloroplast OEM have been established, there are non-classical OEPs which fall outside of these defined structural groups. Minimal research has attempted to characterize the localization mechanisms used by these non-classical OEPs. The most notable non-classical OEP is the essential Toc receptor, Toc159, which was originally considered a TA-protein due to its structural similarity with Toc34 and other TA-proteins. However, unlike TA-protein insertion, Toc159 is not anchored to the membrane using an alpha-helical TMD. Furthermore, the CTS region of Toc159 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 (Teresinski et al., 2019). Non-classical OEPs like Toc159 cannot use the same translocation mechanisms as β -barrel, SA-, and TA-proteins because they lack the necessary targeting-features. Therefore, the translocation pathway used by many OEPs has yet to be characterised (Lee et al., 2014).

Elucidating the localization mechanism of Toc159 is of great interest because it plays an essential role in preprotein import and chloroplast biogenesis. Toc159 is translocated into the OEM when the G domain is in a GDP-bound conformation (Smith et al., 2002). The G domain 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 its three structurally distinct segments, named the M1, M2, and M3 regions. The CT 56aa of the M3 domain contains a localization signal which is necessary for OEM targeting and the upstream 44aa of the M2 region is sufficient for membrane anchorage. Together, the M2 and M3 domains are necessary and sufficient for chloroplast-OEM targeting. Upon closer examination of the M3 signal-containing region, Lung et al. (2014) identified features that are reminiscent of classical chloroplast TPs. The reversed TP-like signal has an abundance of S/T residues and forms an amphipathic alpha helix. Moreover, the AtToc159 sequence was reversed and analyzed using the TP prediction software ChloroP. A TP was successfully identified, indicating the presence of a reverse C-terminal TP-like (CT TP-like) signal (Lung et al., 2014). From this evidence, they concluded a novel OEP targeting mechanism is used by Toc159 in *Arabidopsis thaliana* and *Bienertia sinuspersici* to target to the chloroplast OEM (Lung & Chuong, 2012). Furthermore, they hypothesized a novel C-terminal TP-like targeting signal may be used by a select subclass of OEPs for OEM translocation (Lung et al., 2014). Subsequently, ChloroP was used to analyze the reverse sequence of 117-known chloroplast OEPs in *A. thaliana*, which were compiled by Inoue (2015). Of 117 OEPs, 8 returned scores over the ChloroP threshold value, one of which included Outer Envelope Protein 16-2 (OEP16-2; At4G16160; Grimberg, 2016).

1.14. Predicting N-terminal Transit Peptides using Computational Tools

A myriad of computational tools can be used to probe a protein sequence for TP features. The tool ChloroP analyzes the NT of protein sequences to identify select TP features; it predicts the size of the TP and the SPP cleavage site. The localization-threshold score for

ChloroP is 0.5; any value above this threshold is indicative of a chloroplast-localization signal (Emanuelsson et al., 1999). Other tools can be used in tandem with ChloroP to investigate individual TP features. ProtParam analyzes amino acid (aa) composition revealing aa biases, such as an increased percentage of S&T residues (Gasteiger et al., 2005). The tool PSI-Pred predicts secondary structures such as alpha helices by analyzing the primary structure of an inputted protein (Jones, 1999). Phobius analyzes proteins for transmembrane domains which can indicate an amphipathic alpha helix (Käll et al., 2007). A helical wheel projection (HWP) predicts the hydrophobicity of alpha helical faces. An HWP generates alpha helices using frames of 18aa and analyzes the resulting side-chain projections by calculating the hydrophobicity score of the projections using the Kyte-Doolittle hydrophobicity scale. Using these hydrophobicity scores the program identifies any hydrophilic faces on the alpha helix which can be indicative of an amphipathic helix (Gautier et al., 2007). Additionally, protein sequences can be probed for characteristic TP motif subgroups to identify likely targeting signals (Lee et al., 2013).

1.15. Function, Localization, Evolution, and Expression of OEP16-2

OEP16-2 is a member of the PRAT (Preprotein Amino Acid Transporter) protein family and has 2 isoforms, OEP16-1 (At2G28900) and OEP16-4 (At3G62880). The PRAT family contains six subgroups of proteins which target the mitochondria and/or chloroplast membranes. These subgroups include families, HP20, HP30, TIM17, TIM22, TIM23, and OEP16 (Rossig et al., 2014). PRAT proteins form transmembrane protein pores and function as amino acid and/or small peptide transports (Pohlmeyer et al., 1997). However, further research is needed to

characterise the specific function of each isoform (Rossig et al., 2014). Evolutionary analysis of OEP16 protein suggests OEP16-4 diverged first, while OEP16-1 and OEP16-2 share a more recent common ancestor (Pudelski, et al. 2010). This is made evident by the high degree of sequence conservation between OEP16-1 and OEP16-2 (Figure 1.6; Drea, et al. 2006).

OEP16-1 likely functions a voltage-gated homodimer with selectivity for amino acids and pPORA (NADPH:protochlorophyllide oxidoreductase A precursor; Pohlmeier et al., 1997; Samol et al., 2011). Some studies have suggested OEP16-1 also selectively imports OEP16-1 and OEP16-2 have different protein expression profiles. OEP16-2 is expressed in desiccant tissues, such as seeds and pollen grains, and is controlled by an ABA-inducible promoter. OEP16-1 is expressed primarily in leaf tissue at moderate levels throughout most development phases and is upregulated by low-temperature stress (Drea et al., 2006). Localization assays demonstrate that OEP16-1 and OEP16-2 target the chloroplast OEM, however, it remains unclear which envelope layer OEP16-4 is targeted to. However, the chloroplast targeting-signal and pathway used by each OEP16 isoform is currently unknown (Pudelski, et al. 2010).

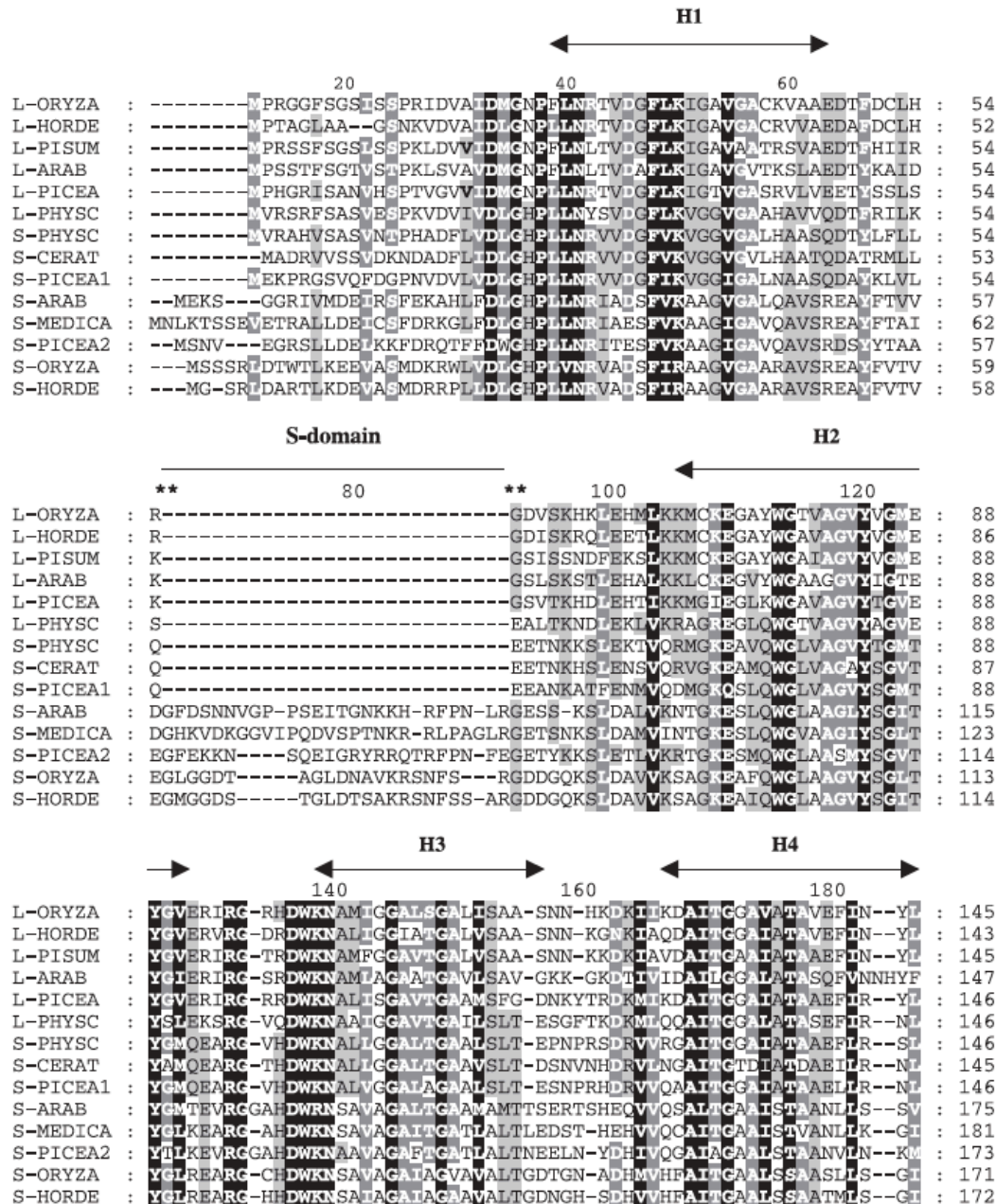


Figure 1.6. MSA of OEP16-1 & OEP16-2 with Predicted Secondary Structure by from Drea et al. (2006). The MSA aligns six OEP16-1 sequences, formerly called OEP16-L, and eight OEP16-2 sequences, formerly called OEP16-S. Symbols H1, H2, H3 and H4 denote alpha helical domains predicted by CD analysis of OEP16-1 from *Pisum sativum* (Linke, et al., 2004). The region of the S-domain in OEP16-2 is also indicated. A high degree of sequence conservation can be seen in the predicted alpha helical regions.

Figure 1.6. MSA of OEP16-1 & OEP16-2 with Predicted Secondary Structure by from Drea et al. (2006). Reprinted from Gene duplication, exon gain and neofunctionalization of OEP16-related genes in land plants by Drea, S. C., Lao, N. T., Wolfe, K. H. & Kavanaugh, T. A., 2006. Retrieved from The Plant Journal. Copyright 1999-2020 by John Wiley & Sons Inc.

1.16. Structural Prediction of OEP16-1

Of all OEP16 isoforms, OEP16-1 has been studied in the most detail. OEP16-1 is composed of four alpha helices denoted H1, H2, H3 & H4. Its structure has been validated by circular dichroism (CD) analysis and nuclear magnetic resonance (NMR), the results of which are in good agreement (Figure 7; Linke et al., 2004; Zook et al., 2013). The structure of OEP16-2 has not yet been characterised. Nonetheless, the high degree of conservation between OEP16-1 and OEP16-2 enables the inference of secondary structures in OEP16-2 using known OEP16-1 structures. Drea et al. (2006) inferred the location of secondary structures and domains in OEP16-2 by creating a multiple sequence alignment (MSA) between OEP16-1 and OEP16-2 then overlaying the secondary structure of PsOEP16-1 determined via CD analysis (Figure 1.6). A notable difference between the isoforms is the presence of the S-domain in OEP16-2. The S-domain is a disordered region with low conservation located between the H1 and H2 domain. Its function is currently unknown, and it is not a universal feature of all OEP16-2 sequences (Figure 6; Drea et al. 2006).

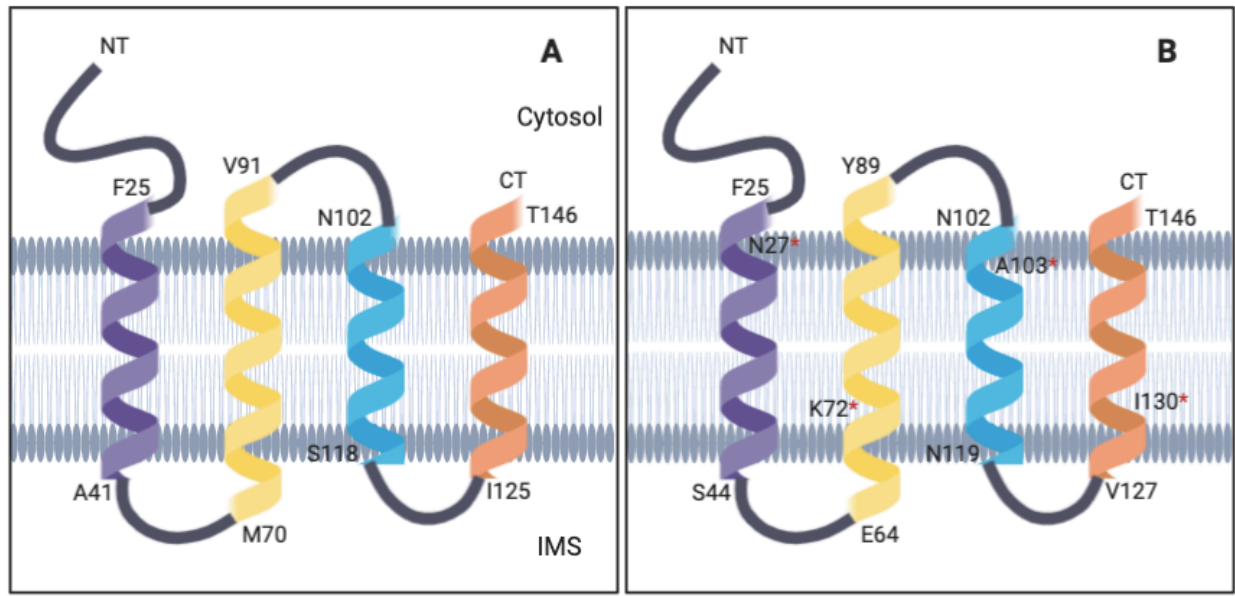


Figure 1.7. OEP16-1 structure based on CD and NMR analysis. Secondary structural depiction of OEP16-1 by CD (A) and NMR (B) studies. OEP16-1 contains four alpha helices, H1 purple, H2 yellow, H3 blue, H4 orange. The N-terminus (NT) and C-terminus (CT) are oriented in the cytosol. The aa position and identity are denoted for the first and last residue in each helix. Asterisk labeled residues () are kinks predicted by TALSO+ (Linke et al. 2004; Zook et al. 2013).*

1.17. Hypothesis and Overall Objectives

The targeting mechanism and pathway used by many OEPs are uncharacterised. The OEP Toc159 has been shown to use a novel CT TP-like signal to target the plastid outer envelope (Lung et al., 2014). To further this research, a set of common characteristics belonging to CT TP-like signals are going to be identified. First, OEPs containing potential CT TP-like signals were identified by Grimberg, 2016. After experimentally verifying CT TP-like activity in OEPs, these signals will be compared to identify common features that function as an OEM targeting-signal. The objective of this study was to investigate the targeting mechanism used by a CT TP-like signal candidate OEP16-2 (Grimberg, 2016). I hypothesized that OEP16-2 uses a

CT TP-like signal to target the plastid outer envelope. Three objectives were set to investigate this hypothesis, first, the subcellular localization pattern of OEP16-2 was examined; then, the region containing the TP-signal within the OEP16-2 sequence was determined; finally, features which may facilitate OEM localization were identified.

1. *Examining the subcellular localization of OEP16-2*

Fluorescent fusion constructs were made using the OEP16-2 sequence and enhanced green fluorescent protein (EGFP). Constructs were designed with EGFP fused to either the N- or C-terminus of OEP16-2 to test the effect of EGFP orientation on OEP16-2 localization. Constructs were transiently expressed in two live-cell systems, onion epidermal cells and *A. thaliana* mesophyll cell protoplasts, then observed using epifluorescent microscopy. Immunodetection of protein extracts from *A. thaliana* mesophyll cell protoplasts confirmed the observed subcellular localization pattern.

2. *Determine the region containing the TP-signal within the OEP16-2 sequence*

EGFP fusion constructs were made using various truncations of the OEP16-2 sequence wherein each truncation contained a different set of domains. Together these constructs assessed the targeting function of each individual domain and various domain combinations. Fusion constructs were transiently expressed in the previously mentioned live-cell systems. The subcellular localization of fusion constructs transiently expressed in protoplasts was confirmed via immunodetection.

3. *Identify TP-signal features which facilitate OEM localization*

Once the targeting region of OEP16-2 was determined, computational analysis of the targeting region was performed. The OEP16-2 targeting region was probed for features common to other plastid targeting-signals, such as biases in residue composition. Lastly, a list of known OEPs was generated through literature searches. OEPs without known targeting-signals and with similar structure to OEP16-2 were compared to the identified OEP16-2 targeting-signal to uncover common features which could function in OEM-targeting.

Chapter 2. Materials and Methods

2.1. Computational analysis of the C-Terminal OEP16-2 sequence

The protein sequence of AtOEP16-2 was retrieved from NCBI (At4G16160; A1). The protein sequence was inputted into online servers, ChloroP, ProtParam, PSI-Pred 4.0, Phobius, & HeliQuest Analysis (Table 2.1; Figure 3.1-3.5; Table 3.1). A multiple sequence alignment (MSA) was generated using OEP16-2 sequences from 29 different plant species (Figure 3.2; A2). Sequences were retrieved by performing a pBLAST analysis using the full length AtOEP16-2 sequence. Each sequence selected was an annotated OEP16-2 protein, hypothetical and uncharacterised proteins were not used. Additionally, protein sequences from a variety of different genus were selected to provide diversity and reduce evolutionary bias within the alignment. Sequences were compiled in a Fasta file and imported into the alignment tool seaview. The MUSCLE algorithm was used to generate an MSA and seaview was used to create a consensus60 sequence from the MSA (Figure 3.2).

Table 2.1. URLs of Computational Tool Servers.

Online Server	Link
ChloroP	http://www.cbs.dtu.dk/services/ChloroP/
ProtParam	http://protparam.net/index.html
PSI-Pred	http://bioinf.cs.ucl.ac.uk/psipred/
Phobius	http://phobius.sbc.su.se/
HeliQuest Analysis	https://heliquest.ipmc.cnrs.fr/cgi-bin/ComputParams.py

Each online server used to analyze the primary protein sequence of OEP16-2, these results are found and discussed in chapter 3.

2.2. Constructing OEP16-2 Fusion Constructs

The AtOEP16-2 (At4G16160) sequence was retrieved from NCBI and the full coding sequence was synthesized by Bio Basics (Bio Basic Canada) and subcloned into the pBluescript vector. Primers were designed to amplify specific regions of the OEP16-2 for subcloning into the XhoI and/or BamHI cut sites in the MCS of either the pSAT6C1 or pSAT6N1 vector (Table 2.2). PCR products underwent gel purification, depending on the product size, fragments were loaded onto a 1.0-1.8% agarose gel and run between 80-110 volts for 25-45 minutes. Following electrophoresis, products of the correct size were excised using a razor blade and recovered using a Biobasics miniprep spin column (BS354). The concentration of recovered PCR product was estimated by nanodrop nucleic acid analysis. Recovered PCR products were digested in a 37°C water bath for 1-4 hours using the enzymes XhoI and/or BamHI purchased from New England BioLabs (R0146S & R3136S). Following digestion, products were either gel purified in the previously described manner or purified directly from the Biobasics restriction enzyme digest DNA purification kit (BS354). The recovered DNA concentration was measured by nanodrop nucleic acid analysis. The pSAT6C1 and pSAT6N1 vectors were digested using two successive digestions in a 37°C water bath. First BamHI digested the vector for 2-16 hours, then was inactivated by placement on an 80°C heat block for 20 minutes, then, vectors were digested with XhoI for 2-4 hours and gel purified. Products were then purified on a 1.0% agarose gel, run between 100-110 volts for 25-45 minutes, then excised using a razor blade and cleaned up using a Biobasics mini-prep kit (BS354). The concentration of recovered plasmid DNA was estimated by nanodrop nucleic acid analysis.

Digested PCR products (insert) and plasmid DNA (vector) were ligated using commercial T4 DNA ligase (M0202S). The insert and vector were ligated in either a 3:1, 5:1, 7:1 or 10:1 ratio and incubated overnight (16-20 hours) at 4°C or incubated at room temperature for 2-4 hours before being transferred to a 4°C fridge for overnight incubation. Following overnight incubation, each ligation reaction was mixed with 50-150 µL of competent *E. coli* DH5α cells and left on ice for 30 minutes. Cells were heat shocked in a water bath at 37°C for 3 minutes or 42°C for 90-120 seconds then placed on ice for 2-5 minutes. Cells were allowed to recover for 40-50 minutes in 800 µL of LB broth placed in a climate controlled shaker set at 37°C and 250 rpms. Recovered cells were spun in a centrifuge for 2 minutes at 10000 rpms, then resuspended in 200 µL of LB broth and plated on selective LB media which contained 100 µM ampicillin. Inoculated plates were placed in a 37°C chamber for 16-20 hours, plates which grew colonies were stored at 4°C and screened for positive transformants.

Plates with colonies were initially screened for positive transformants using colony PCR, or, in the case of small colonies, were immediately cultured overnight for plasmid isolation using the Biobasics miniprep kit (BS614). Colony PCR and plasmid PCR utilized an insert flanking primer and a primer that annealed to the vector to verify insert presence and orientation. PCR products were visualized on an agarose gel run at 80-100 volts for 25-60 minutes. Colonies that contained a DNA product of the correct size were then cultured overnight in LB broth containing 100 µM ampicillin for 16-20 hours in a 37°C incubator shaking at 250 rpms and streaked on LB plates that contained ampicillin and stored at 4°C. Plasmid DNA from cultures was extracted using a Biobasics mini-prep kit (BS614) and the concentration was estimated by nanodrop nucleic acid analysis. Plasmid DNA underwent a single and/or double digestion

reaction and products were visualized using an agarose gel to verify the fragment size of digested subcloned plasmids. The sequence of positive colonies was verified by sequencing performed at Sick Kids Hospital (A3).

Table 2.2. Primer Pairs and Vectors used to Create OEP16-2 Constructs.

Construct Name	Vector	Forward Primer	Reverse Primer
EGFP:OEP16-2-FL	C1	OEP16-2XhoF1 - 5' AAA <u>CTC</u> <u>GAG</u> GA ATG GAG AAG AGT GGA 3'	OEP16-2BamR3 - 5' TAC <u>GGA TCC</u> CTA GAA AAC GCT AGA AAG GAG 3'
EGFP:OEP16-2Δ33CT	C1	OEP16-2XhoF1 - 5' AAA <u>CTC</u> <u>GAG</u> GA ATG GAG AAG AGT GGA 3'	OEP16-2BamR4 - 5' AGA <u>GGA TCC</u> CTA AGC CAT TGC CGC TCC 3'
EGFP:OEP16-2-33CT	C1	OEP16-2XhoF3 - 5' AAA <u>CTC</u> <u>GAG</u> GA ATG ACG ACG TCT GAG 3'	OEP16-2BamR3 - 5' TAC <u>GGA TCC</u> CTA GAA AAC GCT AGA AAG GAG 3'
OEP16-2-FL:EGFP	N1	OEP16-2XhoF2 - 5' AAA <u>CTC</u> <u>GAG</u> ATG GAG AAG AGT GGA GG 3'	OEP16-2BamR1N - 5' CGC <u>GGA TCC</u> G GAA AAC GCT AGA AAG 3'
OEP16-2Δ33CT:EGFP	N1	OEP16-2XhoF2 - 5' AAA <u>CTC</u> <u>GAG</u> ATG GAG AAG AGT GGA GG 3'	OEP16-2BamR2N - 5' AAA <u>GGA TCC</u> G AGC CAT TGC CGC TCC TGT 3'
OEP16-2-33CT:EGFP	N1	OEP16-2XhoF4 - 5' AAA <u>CTC</u> <u>GAG</u> ATG ACG ACG TCT GAG 3'	OEP16-2BamR1N - 5' CGC <u>GGA TCC</u> G GAA AAC GCT AGA AAG 3'
OEP16-2Δ53CT:EGFP	N1	OEP16-2XhoF2 - 5' AAA <u>CTC</u> <u>GAG</u> ATG GAG AAG AGT GGA GG 3'	OEP16-2BamR5 - 5' TTT <u>GGA</u> <u>TCC</u> C TCC ACG AAC CTC TGT 3'
OEP16-2-53CT:EGFP	N1	OEP16-2XhoF5 - 5' AAA <u>CTC</u> <u>GAG</u> ATG GGA GCT CAT GAT TGG 3'	OEP16-2BamR1N - 5' CGC <u>GGA TCC</u> G GAA AAC GCT AGA AAG 3'
OEP16-2Δ96CT:EGFP	N1	OEP16-2XhoF2 - 5' AAA <u>CTC</u> <u>GAG</u> ATG GAG AAG AGT GGA GG 3'	OEP16-2BamR6 - 5' AAA <u>GGA TCC</u> C AGG GAA CCT ATG TTT 3'
OEP16-2-96CT:EGFP	N1	OEP16-2XhoF6 - 5' AAT <u>CTC</u> <u>GAG</u> ATG GGG GAA AGC AGC AAA TCT3'	OEP16-2BamR1N - 5' CGC <u>GGA TCC</u> G GAA AAC GCT AGA AAG 3'
OEP16-2Δ121CT:EGFP	N1	OEP16-2XhoF2 - 5' AAA <u>CTC</u> <u>GAG</u> ATG GAG AAG AGT GGA GG 3'	OEP16-2BamR6 - 5' AAA <u>GGA TCC</u> C AGG GAA CCT ATG TTT 3'
OEP16-2-121CT:EGFP	N1	OEP16-2XhoF7 - 5' AAA <u>CTC</u> <u>GAG</u> ATG GAC GGG GCA GGT T 3'	OEP16-2BamR1N - 5' CGC <u>GGA TCC</u> G GAA AAC GCT AGA AAG 3'

OEP16-2-H2:EGFP	N1	OEP16-2XhoF6 - 5' AAT <u>CTC</u> <u>GAG</u> ATG GGG GAA AGC AGC AAA TCT3'	OEP16-2BamR5 - 5' TTT <u>GGA</u> <u>TCC</u> C TCC ACG AAC CTC TGT 3'
EGFP:OEP16-2-H2	C1	OEP16-2 H2_XhoF1 - GTG TTC AAC TCG AGG CGG GGA AA	OEP16-2 H2 BamR1 - CA AAT CAT GGA TCC CTA TCC ACG AAC CTC
OEP16-2-SD	N1	OEP16-2XhoF7 - 5' AAA <u>CTC</u> <u>GAG</u> ATG GAC GGG GCA GGT T 3'	OEP16-2BamR6 – AAA <u>GGA</u> <u>TCC</u> CAG GGA ACC TAT GTT T
OEP16-2ΔH1-H2	OEP16-2-SD	OEP16-2BamF3 – 5' ATC TCG <u>GGA TCC</u> TGG GAG CTC ATG AT	OEP16-2BamR1N - 5' CGC <u>GGA TCC</u> G GAA AAC GCT AGA AAG 3'

The listed forward and reverse primer pairs were used in PCR reactions to amplify the corresponding OEP16-2 for subcloning into the specific vector. Restriction enzyme cut sites are underlined, start codons are in black, bolded and italicized, stop codons are in red, bolded, and italicized.

2.3. Onion Cell Bombardment using the Biolistic Particle Delivery System

Biolistic bombardment was performed to transiently express OEP16-2 fluorescent fusion constructs in onion epidermal cells. Tungsten particles (microcarriers) were prepared following methods described by Sanford et al, (1993). The prepared microcarriers (8 μL of 60 mg/mL) were coated with 1400-1600 ng of plasmid DNA, 10 μL of 2.5 M CaCl₂, and 5 μL of 0.1 M spermidine by vigorous pipetting, then vortexed at high speed for three minutes. In the case of co-bombardment assays, 1400-1600 ng of plasmid DNA containing the Fd-TP:DsRed sequence was included when coating tungsten balls, in addition to 1400-1600 ng of OEP16-2 fusion construct plasmids. The plasmid coated microcarriers were washed twice, first with 100 μL of 70% ethanol then with 100 μL of 100% ethanol. Coated microcarriers were resuspended in 12 μL of 100% ethanol, spread on a mounted macrocarrier holder, and left to dry. The Bio-Rad Biolistic PDS-1000/He particle delivery system (Bio-Rad Canada) was set by opening the helium

tank, setting the regulator pressure 200 psi above the rupture disk psi, turning on the evacuation chamber and the pump. Three sliced onion peels approximated 2 cm by 2 cm were placed in the center of a petri dish on the second lowest level within the evacuation chamber. A clean screen was placed inside the fixed nest, the macrocarrier was inverted above and the cover lid was firmly tightened. A 1350 psi rupture disk was dipped in 100% isopropanol, placed into the retaining cap, and screwed into place with a torque wrench. Finally, the macrocarrier platform was placed directly below the retaining cap. The chamber was evacuated until a pressure of approximately 27 mmHg was reached, then held. Helium pressure accumulated inside gas acceleration tube by pressing fire until the rupture disk burst and propelled the tungsten balls toward the onion sample. The chamber pressure was released, and the onion samples were placed on moist filter paper in dark drawer at room temperature overnight (16-20 hours). Following overnight incubation, the onion epidermal layer was peeled and mounted in a drop of water on a glass slide and a cover slip was placed on top. Peels were viewed under a Zeiss AxioImager D1 Epifluorescent microscope using a brightfield lens and fluorescent filters. The excitation and emission wavelengths for the EGFP and DsRed signals were 470 nm & 525 nm and 550 nm & 570 nm, respectively. During co-bombardment assays, EGFP and DsRed signals captured and merged using AxioImaging software to assess signal co-localization. Each construct bombardment and/or co-bombardment was performed three to six times to validate a consistent expression pattern for each construct.

2.4. Computational Analysis of Secondary Structure and Domain Prediction

The location of domains within the OEP16-2 sequence were predicted in order to construct OEP16-2 constructs that evaluate the targeting capacity of individual domains and domain combinations. Four predictions from different sources were superimposed onto the OEP16-2 sequence to generate a final prediction (Figure 4.5). The computational prediction tool PSI-Pred 4.0 predicted five alpha helices while the tool Phobius predicted three weak transmembrane domains (Figure 3.3 & Figure 3.4). Additionally, an MSA generated by Drea et al. (2006) was used to infer the domain positions in OEP16-2 by overlaying CD analysis of OEP16-1 domains onto the OEP16 MSA (Figure 1.6, Drea, et al., 2006). More recently, NMR analysis was used to resolve the structure of OEP16-1 and the domain positions were identified (Zook et al., 2013). These domain positions were overlaid onto the OEP16 MSA to infer the position of domains in AtOEP16-2 (Figure 4.4). All four predictions were superimposed onto the OEP16-2 sequence which revealed four distinct alpha helical regions and the S-domain (Figure 4.5). The first and last residue predicted in each distinct alpha helical region was used to assign domain positions and generate the overall prediction.

2.5. Protoplast Preparation and Transfection

Protoplasts were prepared and transfected with OEP16-2 fusion constructs for subsequent subfractionation and immunoblotting. Protoplast were prepared using the tape method and buffer treatments as outlined by Wu, et al (2009). First, leaves from a 15-25-day old Arabidopsis plants were placed onto masking tape with the upper epidermis facing down. Then, strips of 3M scotch tape were gently pressed onto the exposed lower epidermis using the

bottom of a 15 mL falcon tube. Scotch tape was gently peeled, removing the lower epidermis. The exposed mesophyll cells were laid face down in 10ml of enzyme buffer (0.4M mannitol, 20 mM MES-KOH, 20 mM KCl, 1% (w/v) cellulase R-10, 0.25% (w/v) macerozyme R-10, 0.1% (w/v) BSA, 10 mM CaCl₂) and incubated at RT for 30-90 minutes while rotating at approximately 40-60 rpm. The protoplast solution was pipetted into a 15 ml tube using a wide-bore pipette tip. W5 buffer (2 mM MES, 154 mM NaCl, 125 mM CaCl₂, 5 mM KCl) was used to wash the tape and dish to collect missed protoplasts and were added to the 15ml tube. Protoplasts were pelleted by centrifugation for 3 minutes at 100 g in a swing-bucket centrifuge. The supernatant was discarded, and protoplasts were washed by resuspension in 10mL of chilled W5 buffer. A second wash was repeated in the same manner, then protoplasts were resuspended in 3 mL of CS-Sucrose buffer (0.4 M sucrose, 20 mM MES-KOH, 20 mM KCl). Protoplasts were spun at 100 g for 3 minutes in the swing-bucket centrifuge which caused healthy protoplasts to form a floating layer. The supernatant and pellet were removed using a glass pasture pipette and the floating layer was resuspended in W5 buffer to achieve a final volume of 1ml, then, protoplasts were incubated on ice for 30 minutes to pellet the protoplasts. Prior to incubation, 10 µl of protoplast solution was loaded onto a haemocytometer and placed under a light microscope to approximate the number of recovered protoplasts. After incubation, the supernatant was removed, and the protoplast pellet was resuspended in Mg-Man buffer (0.4 M mannitol, 4 mM MES-KOH, 15 mM MgCl₂) to achieve a concentration of 20000 cells per 100 µl. The required Mg-Man volume in millilitres was calculated by averaging the cell count in the four corner squares of the haemocytometer then dividing by 20. Prepared protoplasts were incubated at RT for 15 minutes with plasmid DNA and PEG-solution using a ratio of 20000 cells per 5-10 µg of

DNA and 110 μl of PEG-solution (40% (w/v) PEG-4000, 0.2 M mannitol, 0.1 M CaCl_2). Following incubation, 440 μl of W5 buffer was added for every 20000 cells in solution and gently mixed to stop chemical transfection. Protoplasts were pelleted using swing-bucket centrifugation at 100 g for 2 minutes and the pellet was resuspended in 1-2 mL of WI buffer (0.5 M mannitol, 4 mM MES-KOH, 20 mM KCl). Protoplasts were then placed under mesh cloth in a growth chamber set at approximately 23°C, emitting 30 $\mu\text{mol m}^{-2} \text{s}^{-1}$ of light, and left to recover for up to 16 hours. Following overnight recovery, the transformed protoplast solution was mixed, 10 μL was loaded onto a depression slide and viewed under the Zeiss AxioImaging epifluorescent microscope. EGFP signal was observed using an excitation wavelength of 470 nm and emission wavelength of 525 nm. The transfection rate was estimated by comparing the number of fluorescent cells to the total number of protoplasts. Protoplast preparations with transfection rates higher than 70% were used for subsequent fraction and immunoblot analysis.

2.6. Protoplast Subfractionation to Obtain Soluble, Insoluble, and Total Protein Fractions

Protoplasts were prepared, transformed, and incubated overnight as previously described using batches of 350 000 - 500 000 cells. Following overnight recovery, protoplasts solutions were gently homogenized, 50 000 – 75 000 cells were placed in a 1.5mL tube and used as the total protein fraction. The total fraction was prepared by pelleting cells by centrifugation at 100g for 2 minutes and resuspension in solubilization buffer (100mM Tris-HCl pH8, 100mM NaCl, 1% (v/v) SDS, 1% (v/v) Triton X-100). Protoplasts were vortexed at high speed on an angle for 2 minutes then, the solution was centrifuged at 15 000 rpms for 15 minutes. The supernatant was moved to a fresh 1.5mL tube and 4-5 volumes of ice-cold

acetone was added. The solution was incubated at -20°C for 1-2 hours, then spun for 15 minutes at 14 000 rpms in a centrifuge cooled to 4°C. The resulting protein pellet was resuspended in 15µL of 6x SDS loading dye (375 mM Tris-HCl, 9% w/v SDS, 50% v/v Glycerol, 0.03% w/v Bromophenol Blue). The remaining 300 000 – 425 000 protoplasts were fractioned into a cytosolic soluble fraction and a chloroplast insoluble fraction. First, protoplasts were pelleted by swing-bucket centrifugation at 100g for 2 minutes. The pellet was resuspended in 300µL of lysis HS buffer (50mM HEPES-KOH pH 7.3, 330mM sorbitol, 1mM PMSF) per 75 000 cells and gently mixed. The protoplasts were lysed by pushing the solution through a 10µm mesh fixed to the end of a syringe. The insoluble fraction was pelleted by centrifugation at 5000g for 5 minutes and the soluble supernatant fraction was placed into a new tube. The pellet was resuspended in 10µl of chloroplast lysis buffer (25mM Tris-HCl pH 6.8, 1% (v/v) Triton X-100, 1mM DTT) and 10µl of 6xSDS solution then mixed using a pipette until fully resuspended. Protein in the soluble fraction was precipitated by adding 1 volume of acetone/TCA (50% acetone, 10% TCA) and incubated on ice for 5-10 minutes. The solution was spun at 15000g for 15 minutes at 4°C and the supernatant was discarded. The pellet was washed in 100% acetone and spun at 15000g for 2 minutes at 4°C, then washed in 80% acetone and spun in the same manner. The pellet air dried and was resuspended in 20µL of 6x SDS loading dye.

2.7. Protoplast Protein Separation by SDS-PAGE and Detection by Western Blot Analysis

The total, soluble, and insoluble fraction were boiled for 10 minutes on a 95°C heat block. Samples were loaded onto an SDS-PAGE gel composed of 10% separating layer and a 4.8%

stacking layer and run alongside the Biobasics Two Colour Prestained Protein Ladder (BZ0010R). Samples in the gel were run in a 4°C fridge at 80 volts through the stacking layer, then, at 110 volts through the separating layer. The gel and Whatman paper were incubated for 15 minutes in 1x transfer buffer (38.4mM Tris, 31.2mM glycine, 1.04mM SDS, 20% (v/v) methanol) and a PVDF membrane was soaked in 100% methanol. The gel was assembled on a chemiblot semi-dry transfer apparatus in a sandwich with prepared Whatman paper and the PVDF membrane. The sandwich consisted of a piece of Whatman paper, followed by the PDVF membrane, the gel, and a second piece of Whatman paper, a glass test tube was used to roll out any air bubbles. The transfer apparatus was run at 18 volts for 30 minutes and the PVDF paper was soaked in ponceau stain for 10 minutes. The PVDF paper was then rinsed with dH₂O several times, dried on a kim wipe, and photographed. The PVDF membrane was incubated 20ml of blocking buffer (5% skim powder, 100mM Tris-HCl pH 7.5, 150mM NaCl, 0.3% Tween-20) and left shaking at RT for one hour. The PVDF paper was then incubated in 20 ml of 1 primary antibody-solution (1:5000 rabbit anti-EGFP sera dilution, 5% skim milk, 100mM Tris-HCl pH 7.5, 150mM NaCl, 0.3% Tween-20) by shaking at RT for 2 hours or overnight at 4°C. The PVDF paper was washed 3 times for 10 minutes each wash in 1xTBS-T buffer (100mM Tris-HCl pH 7.5, 150mM NaCl, 0.3% Tween-20). The washed PVDF paper was incubated in 20ml of 2 secondary antibody solution (1:25000 anti-rabbit conjugated to horse radish peroxidase dilution, 5% skim milk, 100mM Tris-HCl pH 7.5, 150mM NaCl, 0.3% Tween-20) by shaking at RT for 2 hours. The PVDF paper was then washed in 1x TBS-T solution for 10 minutes 3 times. Then, incubated in 2-4ml of freshly prepared Immun-Star AP Chemiluminescence Kit (170-5061) for 5 minutes in the dark. ChemiDoc MP system Hi-Sensitivity and colometric filters were used to capture the

antibody signal and pre-stained ladder, respectively. Replicates of each blot were successfully performed 2-3 times.

2.8. Computational Investigation of the Predicted OEP16-2 H2 domain

The primary protein sequence of the predicted OEP16-2 H2 domain was analyzed by the tools, ProtParam, PSI-PRED, Phobius, & Heli-Quest Analysis (Gasteiger et al., 2005; Jones, 1999; Käll et al., 2007; Gautier et al., 2008). The OEP16-2 primary protein sequence retrieved from NCBI and the AKR2A structure retrieved from the Protein database were input into the SWISS-MODEL program (Waterhouse et al., 2018). The OEP16-2 structure with the highest degree of confidence was selected for analysis. Chimera was used to examine the predicted OEP16-2 structure, a surface model was generated that highlighted the hydrophobic, hydrophilic, and neutral faces of the protein surface.

2.9. Computation Prediction of OEP Targeting Structures and Structural Classification

OEP protein accessions were obtained from four sources, Inoue (2015), Plant Protein Database (PPDB), AT_CHLORO database, and Bouchnak et al (2019). Each source was cross-referenced to identify proteins common and unique to each reference. A total of 137 unique OEP sequences were identified from these four sources (Table 7.1). UniProt was used to retrieve amino acid sequences and experimentally verified structural annotations. Amino acid sequences were input into PSI-PRED 4.0 to predict secondary structures (Jones, 1999). A distinct characteristic of β -barrel proteins is that they contain 8-24 beta-sheets of 6-22 aa in length (Tsaousis et al., 2017). Proteins predicted by PSI-Pred to contain the correct number and

length of β -sheets were further investigated for β -barrel features using HHomp and PRED-TMBB (Zimmermann et al., 2018; Bagos & Liakopoulous, 2004). Proteins identified by HHomp as having any degree of β -barrel homology were classified as β -barrel proteins. Additionally, sequences predicted by PRED-TMBB to have a discrimination threshold of less than 0.995 and characterised as a β -barrel protein were categorized as a β -barrel protein. All proteins identified as a β -barrel by PRED-TMBB were also identified by HHomp, however, not all β -barrel proteins identified by HHomp were identified by PRED-TMBB. Any protein sequence that was not classified as a β -barrel protein and contained at least one alpha helix predicted by PSI-PRED was then analyzed for alpha helical transmembrane protein properties.

Proteins predicted to contain an alpha helix or helices by PSI-PRED were investigated using the computational tools, Phobius, TMHMM, TM-Pred, and MEMSAT-SMV (Käll et al., 2007; Sonnhammer et al., 1998; Hofmann & Stoffel, 1993; Nugent & Jones, 2009). The resulting predictions and experimental annotations available on UniProt were evaluated on a case-by-case basis and the results from all sources were collectively considered to determine the most appropriate structural classification. If the majority of helical detection tools predicted multiple alpha helices the protein was classified as an alpha helical multi-pass protein. However, if the majority of helical detection tools predicted a single alpha helix in a consistent region within the protein sequence the protein was classified as a single-pass alpha helix protein. Proteins that were not predicted to contain any transmembrane helices or conform to a β -barrel structure were categorized as other.

Chapter 3. Computational Assessment of the Predicted OEP16-2 Chloroplast Targeting Signal

3.1 Overview

The CT 33 amino acids of OEP16-2 were computationally assessed for TP-like features that could function as a transit peptide-like signal. NT transit peptides contain a variety of targeting features, including an overrepresentation of basic K&R, polar S&T, and hydrophobic proline residues and an underrepresentation of acidic D&E residues. Additionally, TPs are typically moderately hydrophobic. Moreover, some TPs form stable alpha helices in mimetic environments, yet, relax into stable coils under hydrophilic conditions, this type of structure is known as amphipathic (Lee et al., 2008; Lee & Hwang, 2018). The tools ChloroP, PSI-PRED Phobius, ProtParam, HeliQuest Analysis, as well as, MSAs were used to examine the CT of OEP16-2 for these features to assess potential for a CT TP-like signal (Emanuelsson et al., 1999; Jones, 1999; Käll et al., 2007; Gasteiger et al., 2005; Gautier et al., 2005).

3.2. Computational Investigation of the OEP16-2 Using ChloroP

Grimberg (2016) used ChloroP to predict a TP-like signal in the reverse C-terminus of OEP16-2. It is unlikely that the reversal of TP-like features negatively impacts their targeting capacity because of the fluidity in TP structure and organization as described by the TP MM model. The MM model states the order of targeting features and motifs is irrelevant to the TP overall function (Li & Teng, 2013). Thus, reversed TP features are likely functional and can be predicted by established computational tools. The ChloroP TP analysis tool predicted a 33 amino acid TP in the C-terminus of OEP16-2 (Figure 3.1) and established a working TP region for future analysis. However, ChloroP tends to underestimate the length of TPs (Emanuelsson et

al., 1999). Moreover, the reversal of amino acid sequences may impact the prediction accuracy. In fact, the CT TP-like signal in BsToc159 was underestimated by 5 amino acids (Lung & Chuong, 2012). Thus, the 33 aa of the CT in OEP16-2 may also be an underestimation in length. As such, this was a useful working model, however, is subject to revision.

Name	Length	Score	cTP	CS-score	cTP-length
OEP16-2-Forward	178	0.433	-	8.300	38
OEP16-2-Reverse	178	0.516	Y*	8.583	33

Figure 3.1. TP analysis of OEP16-2 using the ChloroP server. The sequence of OEP16-2 was analyzed by ChloroP in both the forward and reverse direction (Grimberg, 2016). ChloroP provides the length of the protein (Length), a localization threshold score (Score), the predicted presence or absence of a TP (cTP) denoted by Y or -, respectively, MEME matrix score for the predicted SPP cleavage site (CS-score) and the TP-length (cTP-length). The red asterisk emphasizes the predicted TP in the reverse OEP16-2 protein sequence.

3.3. Computational Investigation of the OEP16-2 Using ProtParam

Classical NT TPs exhibit a number of characteristic features and trends (discussed in 1.4) which are present in different combinations within select TPs, as described by the MM model (Lee & Hwang, 2018; Li & Teng, 2013). Trends in aa composition are commonly seen in NT TPs, including, an underrepresentation of acidic residues D&E, and an overrepresentation of basic K&R residues, polar S&T residues, and hydrophobic P residues (Lee & Hwang, 2018). These

trends were investigated in OEP16-2 using ProtParam which compared the 33 aa C-terminal OEP16-2 protein sequence to the full-length protein (Table 3.1). The 33 aa of the C-terminus had 17.04% more polar S&T residues and 3.49% less acidic D&E residues than the full-length sequence, consistent with TP-trends. However, the CT also had a 6.52% decrease in basic K&R residues and has a complete absence of proline residues, which is inconsistent with TP-trends. Previous groups have shown TPs contain an average of 19% serine residues versus 6% found in the mature protein sequence and almost a complete absence of acidic residues (von Heijne et al., 1988). Thus, the overrepresentation of S/T residues and the underrepresentation of D/E residues in the CT may function as a plastid-targeting feature.

ProtParam also calculates the grand average of hydropathicity (GRAVY) for a given sequence input. Values greater than 0 indicate hydrophobicity while values less than 0 indicate hydrophilicity (Kyte & Doolittle, 1982). Moderate to weak hydrophobicity is a common trend seen in many classical TPs (Lee & Hwang, 2018). The full-length OEP16-2 sequence has a GRAVY of -0.125, indicating moderate hydrophilicity. Conversely, the CT OEP16-2 sequence has a GRAVY of 0.252 which indicates moderate hydrophobicity and supports the CT TP-like prediction in OEP16-2 (Chotewutmontri & Bruce, 2015).

Taken together, the moderate hydrophobicity of the CT, the overrepresentation of S&T residues, and the underrepresentation of D&E residues supports the prediction of a CT TP-like signal in OEP16-2. Yet, the lack of basic residues and proline residues is inconsistent with TP aa trends and does not support the conclusion of a CT TP-like signal in OEP16-2. However, due to the high degree of variability in TP aa trends, we do not expect every aa trend to be present. Thus, ProtParam analysis supports the possibility of an CT TP-like signal in OEP16-2.

Table 3.1. Comparing the Amino Acid Composition of the OEP16-2 Full-Length & C-terminal Sequences using ProtParam

	Full Length OEP16-2 Sequence	C-terminal OEP16-2 Sequence	% ^{CT} - % ^{FL}
<i>Total # of aa residues</i>	178	33	
<i>Grand Average of Hydropathicity (GRAVY)</i>	-0.125	0.252	
<i># and (%) of D residues</i>	7 (3.93%)	0 (0%)	-3.93%
<i># and (%) of E residues</i>	10 (5.26%)	2 (6.06%)	+0.8
<i># and (%) of acidic residues (D&E)</i>	17 (9.55%)	2 (6.06%)	-3.49%
<i># and (%) of K residues</i>	8 (4.49%)	0 (0%)	-4.49%
<i># and (%) of R residues</i>	9 (5.06%)	1 (3.03%)	-2.03%
<i># and (%) of basic residues (K&R)</i>	17 (9.55%)	1 (3.03%)	-6.52%
<i># and (%) of P residues</i>	4 (2.25%)	0 (0)	-2.25%
<i># and (%) of S residues</i>	18 (10.11%)	6 (18.18%)	+8.07%
<i># and (%) of T residues</i>	11 (6.18%)	5 (15.15%)	+8.97%
<i># and (%) of Polar S&T residues</i>	29 (16.29%)	11 (33.33%)	+17.04%

The amino acid composition, net charge, hydrophobicity, and polarity is compared between the full length OEP16-2 sequence and the predicted 33aa C-terminal TP. The number and % of each residue are listed for each sequence input. Differences between the % residue composition of the CT and FL sequence are calculated.

3.4. Computational Comparison of OEP16-2 Sequences Using an MSA

An MSA was generated using OEP16-2 sequences from 29 different plant varieties to assess sequence conservation and the consistency of CT aa trends (Figure 3.2; A2). A high degree of sequence conservation is evident throughout the entire alignment, particularly in regions containing predicted alpha helices. Conserved residues within the MSA were identified by generating a consensus60 sequence (Figure 3.2). Consensus60 sequences include residues that are conserved in 60% or more of aligned positions. Residues that are conserved in at least

60% of aligned positions are highly conserved throughout evolution. Within the 33 aa of the AtOPE16-2 C-terminus, polar and acidic residues sites are conserved, however, the residue identity is not always consistent. Of the 11 S&T residues in the CT AtOEP16-2, 5 of these sites are conserved in the consensus60, in both residue identity and position, while 1 site is flexible to either an S or T residue. Comparatively, 12 S&T residue sites are conserved throughout the upstream consensus60 sequence. The CT consensus60 sequence contains 15.15% S&T residues while the upstream consensus60 sequence contains 6.74% of residues. Therefore, the trend of S&T overrepresentation is consistent in the OEP16-2 MSA. Interestingly, the other polar residues N&Q are almost completely conserved in identity and position at 4 sites in the CT. Of the 2 D&E residues sites, 1 is conserved in residue identity and 1 is flexible to a D or E residue. The single basic R residue is not conserved. The overrepresentation of S&T residues and underrepresentation of D&E residues is consistent within the MSA, supported their predicted role as transit peptide. Overall, the identity and position of the 33 CT amino acids in OEP16-2 is highly conserved throughout plant species as 28 out of 33 residue positions are conserved within the C-terminal consensus60 sequence.

1

Arabidopsis	-----MEK	SGGRIVMDEI	RSFEKAHL-F	DLGHPLLNR	ADSFVKAAGV	GALQAVSREA	YFTVVDGAGF
Capsella	-----MEK	SGGRVMDEI	RSFEKAHL-F	DLGHPLLNR	ADSFVKAAGV	GALQAVSREA	YFTVVDGAGF
Eutrema	-----ME	KSGRNVMEI	RSFEKASL-F	DLGHPLLNR	ADSFVKAAGV	GALQAVSREA	YFTVVDGAGF
Brassica	-----MEK	SGGRKVMDEI	RSFEKASL-F	DLGHPLLNRV	ADSFVKAAGV	GALQAVSREA	YFTVVDGAGF
Quercus	---MSSSSSNL	E-KRSLDEL	SSFEGKGF-F	DLGHPLLNR	AESFVKAAGI	GAIQAVSREA	YFTAVESAGL
Ziziphus	MSLGGGSGNL	E-TRTLLDEL	RSFDKGGF-F	DLGHPLLNR	AESFVKAAGI	GAVQAVSREA	YFTAFEGFDS
Prunus	---MNTSSSNL	E-TRPSLQEL	RSFEKGGF-F	DFGHPLLNR	AESFLKAAGI	GAIQAVSREA	YFTATEGLDS
Carica	---MSSNF	E-NRSLDEL	RSFDKGGF-F	DLGHPLLNR	AESFVKAAGI	GAIQAVSREA	YFTAVESAGV
Arachis	---MNSKSNL	E-TRSLDEL	SSFNOGGL-F	DFGHPLLNR	AESFVKAAGI	GAVQAVSREA	YFTAIESSGL
Medicago	---MNSNSNL	E-TRTLLDEL	SDFNKGGF-F	DFGHPLLNR	AESFVKAAGI	GAVQAVSREA	YFTVIEGTGI
Malus	MLNTTSSSTSL	ETKPSLLPEL	RSFDKSGF-F	DLGHPLLNR	AESFVKAAGI	GAIQAVSREA	YFIAIEGFDS
Manihot	-----MSSLDEV	RRFEKECF-F	DLGHPLLNR	AESFVKAAGV	GAIQAVSREA	YFTAIEGSGL	
Populus	---MSHNL	E-TRSLMGEI	RRFDKCCF-F	DFGHPLLNR	AESFVKAAGI	GAIQAVSREA	YFTAIEGSGF
Abrus	---MNLNTSSNL	E-TRSLNEI	CNFDKGGF-F	DLGHPLLNR	AESFVKAAGI	GAVQAVSREA	YFTAIESTGV
Ricinus	---MSNKL	Q-TRTFMDEL	RGFEKGM-F	DLGHPLLNR	AESFVKAAGI	GAIQAVSREA	YFTAIEGSGL
Mucuna	---MNLNSSNL	E-TRSLDEL	CSFDKGGF-F	DLGHPLLNR	AETFVKAAGI	GAVQAVSREA	YFTAVEGTGA
Spatholobus	---MNHNTSSNL	E-TRSLDEL	CNFDKGGF-F	DLGHPLLNR	AESFVKAAGI	GAVQAVSREA	YFTATEGTGT
Citrus	---MTSNM	E-NRSLFHEL	PGFEKGGF-F	DLGHPLLNR	TESFVKAAGI	GAIQAVTREA	YFTAVEGSGF
Pistacia	---MSNL	E-TRSLDEF	RSFDKGGF-F	DLGHPLLNR	VESFVKAAGI	GAIQAVSRDA	YFTAIEGSLG
Glycine	---MNLNTSSNL	E-TRSLDEL	CNFDKGGF-F	DLGHPLLNR	LETFVKAAGI	GAVQAVSREA	YFTAVEVSGT
Nymphaea	---MNG	NGNRSLLDEL	RSFDKGGF-F	DLGHPLLNR	AESFVKAAGI	GAVQAVSREA	YLTLVGECAS
Ananas	---MSSNGGKL	E-TRTFLDEI	RSMEKRWL-F	DLGHPLLNR	AESFVKAAGI	GAIQAVSREA	YFTAVEGVVA
Syzygium	---M	EARTHSLNE	LNFDKGGF-F	DLGHPLLNR	AESFIKAAGI	GAVQAVSREA	YFTAVEVSLG
Cajanus	---MNLNTSSNL	E-TRSLDEL	CNFDKGGF-F	DLGHPLLNRV	VESFVKAAGV	GAVQAVSREA	YFTAVEGTGA
Jatropha	---MSSNL	E-TRSLMDEF	RNLEKGSW-F	DLGHPLLNR	AESFVKAAGI	GAIQAVSREA	YLTASQGNGL
Solanum	---MSSNM	EC---RSL	HCFDKGGF-F	DLGHPLLNR	SESFVKAAGI	GAVQAVSREA	YFTAESESTGG
Vigna	---MNLNTSSNL	E-TRSLDEL	CSFDKGGF-F	DLGHPLLNR	LESFVKAAGI	GAVQAVSREA	FFSAIEGNGT
Dendrobium	---MSEVNL	E-GRSLDEL	KSFDKGSL-F	DFGHPLLNRV	AESFVKAAGI	GALQAVSREV	YFSAVDGSSI
Momordica	---MGSRNL	E-NRSLVDEI	RSFDSGGFLY	DLGHPLLNRV	AESFIKAAGI	GALQAVSREA	YFTAAESLDS
Consensus60	MMXXXXXNL	EXXRSLDEX	XXFXKXGLF	DLGHPLLNR	AESFVKAAGI	GAXQAVSREA	YFTAXEGXGX

71

Arabidopsis	DSNNVGPPE	ITG---	NKK	HRFPNLRGES	-SKSLDALVK	NTGKESLOWG	LAAGLYSGIT	YGMTEVRGGA
Capsella	DSNNVGPPE	-----	SKK	HRFPNLRGES	-SKSLDALVK	NTGKESLOWG	LAAGLYSGIT	YGMKEVRGGA
Eutrema	DSSNLGPPE	NNG---	SKK	HRFPNLRGES	-SKSLDALVK	NTGKESLOWG	LAAGLYSGIA	YGMKEARGGA
Brassica	DSSMGPPE	DTC---	SKK	HRFPNLRGEN	NSKSLEALVK	NTGKESLOWG	LAAGLYSGIT	YGMKEARGGA
Quercus	DSGGDVSSSE	LSC---	AKK	RRFPDLRGET	NRKSLEAMVK	NTGKESLOWG	LAAGVYSGLT	YGLKEARG-S
Ziziphus	N-GSSVPE	ISG---	AKK	HRFPDLRGET	NRKSLEAMVK	STGKESLOWG	LAAGVYSGLT	YGLKEARG-A
Prunus	SGGGIPPEIS	---G---	NKK	HRFPDLRGEN	NRKSLEAMVK	HTGKESLOWG	LAAGVYSGLT	YGLTEARG-A
Carica	D-SSNLPE	LSC---	AKK	KRFPDLKGES	NRKSLEAMVK	HTGKESLOWG	LAAGIYSGLT	YGLKEARG-S
Arachis	DN-TGGMPPE	VSG---	AKK	HRFPDLRGET	NRKSLEAMVK	HTGKESLOWG	LAAGIYSGLT	YGLKEARG-A
Medicago	DN-AGGMPPE	ISG---	AKK	NRFPDLRGET	SSKSIEAMVK	NTGKESLOWG	LAAGLYSGIT	YGMKEARG-T
Malus	STNGIPPEIS	VPG---	NKR	QRFPDLRGEN	NRKSLEAMVK	NTGKESLOWG	LAAGVYSGLT	YGLREARG-A
Manihot	DSN-GVPE	LSASSDKKR		HRFPDLKGET	NRKSLEALVK	STGKESLOWG	LAAGMYSGLT	YGLREARG-A
Populus	ESS-CGVPE	ISVDG---	KKR	HRAPDLRGET	NRKSLEALVR	NTGKESLOWG	LAAGVYSGLT	YGLSEARG-V
Abrus	D-NSGGLPTE	ISG---	AKK	NCLPLRGET	SSKSVEALVK	KTGKESLOWG	VAAGIYSGLT	YGLKEARG-A
Ricinus	DSS-SSVPE	LSPAGAAKKR		NRFPDLRGET	NRKSLEALVK	STGKESLOWG	LAAGVYSGLT	YGLREARG-A
Mucuna	DNNSGGLPTE	ISG---	AKK	NQLPLRGET	NNKSLEAMVK	STGKESLOWG	VAAGIYSGLT	YGLKEARG-A
Spatholobus	E-NSGGLPSE	ISG---	AKK	NHLPDLRGET	SSKSLEAMVK	STGKESLOWG	VAAGLYSGLT	YGLKEARG-A
Citrus	DS-SNNVSD	MGG---	AKK	HQFPNLKGET	NRKSLEAMVK	NTGKESLOWG	VVAGIYSGLT	YGLREARG-A
Pistacia	DSS---DMSE	LGG---	SKK	HSFSNIKGET	NRKSLEAMVK	NTGKESLOWG	LAAGVYSGLT	YGLKEARG-V
Glycine	D-NSGGLPPE	ISS---	AKK	NRLPSLKGET	NNKSLEAMVK	NTGKESLOWG	VAAGLYAGLT	YGLKEARG-A
Nymphaea	DAGA--DPDL	IKT---	RRH	PRFPNFKGET	CAKSLEAMVK	DTGKESLOWG	IAAGVYSGLT	YGLKEARG-A
Ananas	DSDG---VPD	IAT--ASSKR		NKFPDVRGEN	GKKSLEVVK	NTSKESLOWG	MAAGVYSGLT	YGLREVRG-T
Syzygium	DSGN--GAPE	TPG---	AKK	HRFPDLRGET	NRKSLEAMVK	STGKESLOWG	LAAGVYSGLT	YGLKETRG-A
Cajanus	DK-SGALPPD	LSS---	AKK	NHLPDLRGET	SNKSLEAMVK	NTGKEALOWG	VAAGIYSGLT	YGLKEARG-A
Jatropha	DSS-SGLPPE	LSCA---	AKK	RRFPNLKGET	SVKSLEALVK	RTGKESLOWG	LAAGVYSGLT	YGLREARG-A
Solanum	DTN-SIPPE	ITG---	PKK	NRFPDLRGET	NRKSVEALVK	STGKESVOWG	LAAGMYSGLT	YGLKEARG-V
Vigna	---NRGGLPAE	VSN---	TKK	NQLPLRGET	SNKSLEAMVK	STGKESLOWG	LAAGLYSGLT	YGLKEARG-A
Dendrobium	DSG---TVPE	LPG---	ARK	RRFPDLRGET	NRKSLEALVR	YTGRESFOWG	LTAGIYSGLT	YGLREARG-T
Momordica	NSG---VPPE	LST---	AKK	HRFPDLRGET	NRKSLEAMVK	NTGKESLOWG	LAAGVYSGLT	YGLREARG-A
Consensus60	DSXXXGXEXE	XXGXGASXXX		XRFPDLRGET	NNKSLEAVVK	XTGKESLOWG	LAAGXYSGLT	YGLKEARGGA

	141		*				
Arabidopsis	HDWK	NSAVA	GALTGAAMAM	TTSERTSHEQ	VVQSALTGAA	ISTAANLLSS	VF-
Capsella	HDWK	NSAVA	GALTGAAMAM	TTSDRTSHEQ	VVQSALTGAA	ISTAANLLSS	VF-
Eutrema	HDWR	NSAVA	GALTGAAMAM	TTSERTNHEQ	VVQSALTGAA	ISTAANLLSN	VF-
Brassica	HDWR	NSAVA	GALTGAAMAM	TTSERTSHEQ	VVQSALTGAA	ISTAANLLSS	VF-
Quercus	HDWK	NSAVA	GAITGMALAL	TS-EGSSHEQ	IVQCAITGAA	ISTAANLLTG	IF-
Ziziphus	HDWK	NSAVA	GAITGVALAL	TT-EDYSHEQ	IVQCAITGAA	ISTAANLLTG	IF-
Prunus	HDWK	NSAVA	GAITGVALAL	TS-EGSSHEQ	IVQCAITGAA	ISTAANLLSG	IF-
Carica	HDWK	NSAVA	GAITGVAVAL	TS-EDYSHEQ	VVQCAITGAA	ISTAANLLSG	IF-
Arachis	HDWK	NSVVA	CAITGATLAL	TS-EDTSHEQ	IVQCAITGAA	ISTAANLLTG	IF-
Medicago	HDWK	NSAVA	GAITGAALAL	TS-DNTSHEQ	IAQCAITGAA	ISTAANLLTG	IF-
Malus	HDWK	NSAVA	GAITGVALAL	TS-EGSSHEQ	IVQCAITGAA	ISTAANLLSG	IF-
Manihot	HDWK	NSAVA	CAITGMALAL	TT-DDVSHEQ	VVQCAITGAA	ISTAANLLTA	GIF-
Populus	HDWK	NTAVA	GAITGVALAL	TT-ADISHEQ	IVQCAITGAA	ISTAANLLTG	IF-
Abrus	HDWK	NSAVA	GAITGATLAL	TL-ENSTHEQ	IVQCAITGAA	ISTAANLLTG	IF-
Ricinus	HDWK	NSAVA	CAVTGMALAL	TA-DDVSHEQ	IVQCAITGAA	ISTAANLLTG	-IF-
Mucuna	HDWK	NSAVA	GAITGATLAL	TL-EDSTHEQ	IVQCAITGAA	ISTAANLLTG	IF-
Spatholobus	HDWK	NSAVA	GAITGATLAL	TL-EDSNHEQ	IVQCAITGAA	ISTAANLLTG	IF-
Citrus	HDWK	NSAVA	GAITGVALAL	TT-DDSSHEQ	VVQCAITGAA	ISTAANLLTG	IF-
Pistacia	HDWK	NSAMA	GAITGVALAL	TS-EDPSHEQ	IVQCAITGAA	ISSAANLLTG	IF-
Glycine	HDWK	NSAVA	GAITGATLAL	TL-EDSTHEQ	IVQCAITGAA	ISTAANLLTG	IF-
Nymphaea	HDWK	NSAIA	GALTGATLAL	TT-DDTSHER	IVQCAITGAA	LSTAANLLNG	IF-
Ananas	HDWK	NSAVA	GAITGAAVAL	TS-DNASHEQ	IVQCAITGAA	LSTAANLLSG	IF-
Syzygium	HDWK	NSAVA	GAITGVALAL	TS-DDTSHEQ	IVQCAITGAA	ISTAANLLSG	IF-
Cajanus	HDWK	NSAVA	GAITGATLAL	TL-DDSTHEQ	IVQCAITGAA	ISTAANLLTG	IF-
Jatropha	HDWK	NSAVA	GALTGLALGL	TT-DDVSHEQ	LVQCAITGAA	ISAAANLLTG	I-F-
Solanum	HDWK	NSALA	GAITGAALAL	TL-EERSHEQ	VVQCAITGAA	ISTAANLLTG	IF-
Vigna	HDWK	NSAVA	GAITGATLAL	TL-DDSTHEH	IVQCAITGAA	ISTAANLLTG	IF-
Dendrobium	HDWK	NSVVA	CAVTGAALAL	TL-QDTTHEQ	LVQCAITGAA	LSTAANLLRG	IF-
Momordica	HDWK	NSAIA	GAITGVALAL	TS-DESSHEQ	IVQCAITGAA	VSTAANIFAG	VF-
Consensus60	HDWK	NSAVA	GAITGXALAL	TS-XXXSHEQ	IVQCAITGAA	ISTAANLLXC	IF-

Figure 3.2. Multiple sequence alignment of 29 OEP16-2 sequences from a variety of higher plant species. The AtOEP16-2 sequence was aligned with 28 OEP16-2 sequences from various species, accession numbers, and Latin binomials are listed in A2. The genus is listed beside each sequence line, numbers indicate the MSA position. Red asterisk labels the 1st aa of the predicted CT TP-like signal.

3.5. Secondary Structure Prediction of OEP16-2 Using PSI-PRED

PSI-PRED predicts secondary structures from an inputted amino acid sequence. The tool predicted five alpha helices, H1, H2, H3, H4, & H5, in the full-length OEP16-2 (Figure 3.3). The positions of H2-H5 are in good agreement with the secondary structural prediction by Drea et al. (2006; Figure 1.5). Additionally, PSI-PRED estimates high confidence in the H2-H5 helical predictions (Figure 3.3). The PSI-PRED predicted H2, H3, H4, and H5 regions were used as a parameter for preliminary investigation of the OEP16-2 targeting signal, hereafter referred to as H1, H2, H3, & H4 domains, respectively. The 33 CT amino acids included the predicted H4

domain and the end of the H3 domain. Moderate hydrophobicity in the OEP16-2 CT, identified using ProtParam, indicated the CT may form moderately hydrophobic alpha helix (Lee et al., 2008). Within the 33 aa of the CT, an alpha helix was predicted from positions 9-31. The transmembrane and amphipathic nature of the predicted CT alpha helix was assessed using Phobius and HeliQuest Analysis, respectively.

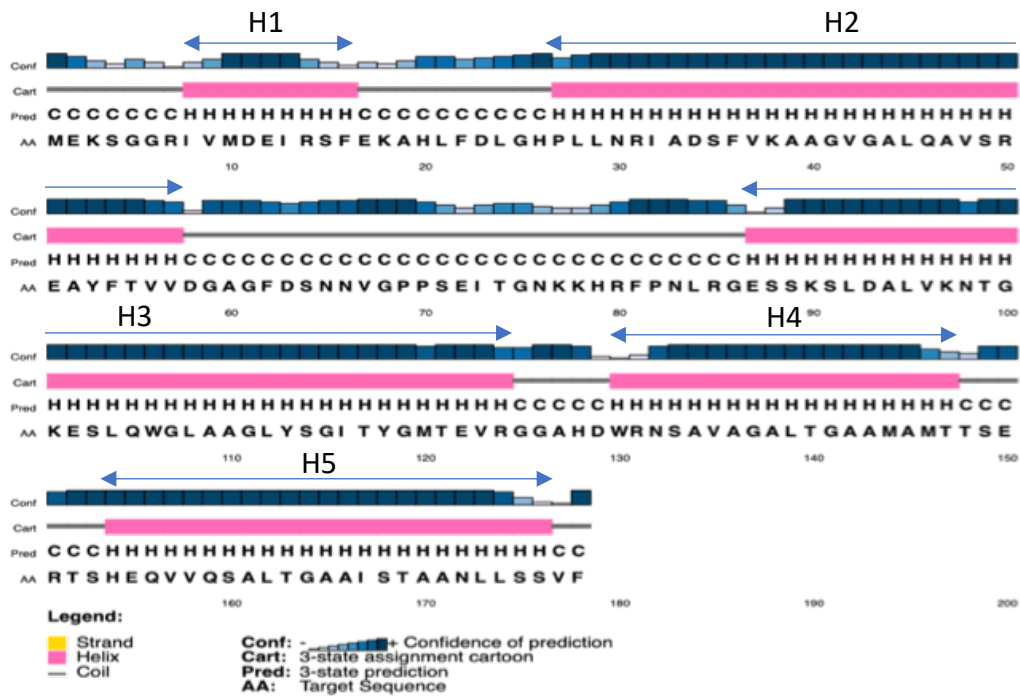


Figure 3.3. Psi-Pred secondary structure prediction of AtOEP16-2. PSI-Pred predicted 5 alpha helices (pink). Conf denotes the prediction confidence at each position, darker taller bars indicate higher confidence. Cart illustrates the predicted structure, pink denoted an alpha helix, grey denotes a coil. Pred provides an alphabetical designation for the structure at each site, H denotes a helix, C denotes a coil. Numbers indicate the amino acid position.

3.6. Computational Investigation of OEP16-2 Using Phobius

Phobius assessed the inputted protein sequence for transmembrane domains, cytoplasmic localization, non-cytoplasmic localization, and signal peptides using a hidden Markov model algorithm (Käll et al., 2007). A probability plot is used to predict these distinct regions and generate an overall prediction. Phobius predicted an overall non-cytoplasmic localization for the full length OEP16-2 sequence (Figure 3.4). However, analysis of the Phobius probability plot revealed three weak transmembrane predictions which corresponded to the predicted H3 & H4 domains, as well as, the C-terminal half of the H2 domain. The strongest of these three weak transmembrane predictions was found in the H4 domain from residues 155-178. Therefore, the 23 CT amino acids may form a moderately hydrophobic transmembrane targeting feature, similar to the moderately hydrophobic alpha helices found in some classical TPs (Lee et al., 2008; Lee & Hwang, 2018).

```

ID Arabidopsis_thaliana
FT TOPO_DOM 1 178 NON CYTOPLASMIC.
//

```

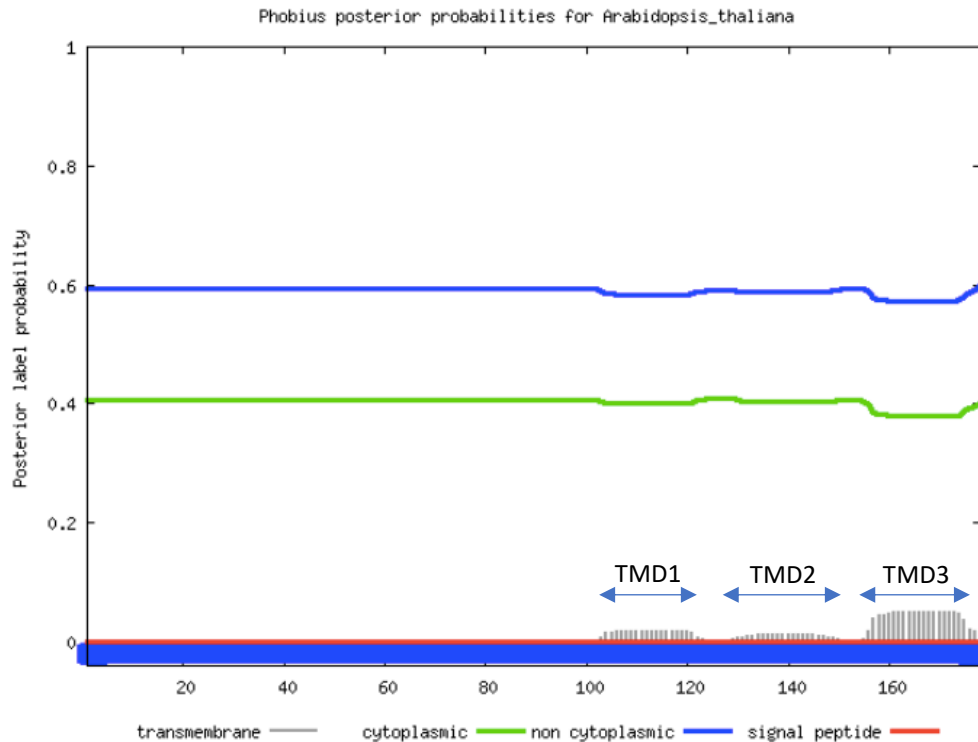


Figure 3.4. Phobius Analysis of the OEP16-2 Protein Sequence. The Phobius plot assessed the inputted amino acid sequences at each position for four features, transmembrane regions (grey bars), cytoplasmic localization (green line), non-cytoplasmic localization (blue line), and signal peptides (red line). AA position is denoted on the X-axis and the probability of each prediction is indicated by the y-axis. The overall prediction for each distinct region (FT) is provided above the plot. Three weak to moderately hydrophobic transmembrane domains were predicted, denoted by TMD1, TMD2, and TMD3.

3.7. Computational Investigation of the OEP16-2 C-terminus Using HeliQuest Analysis

HeliQuest Analysis generated alpha helices from an inputted protein sequence and characterised physiochemical properties of the predicted helices. Helices were generated in frames of 18aas, starting from amino acids 1-18, each subsequent frame was shifted by one aa until the entire sequence was assessed. The 33 OEP16-2 CT amino acids were inputted into

HeliQuest Analysis (Figure 3.5; Gautier et al., 2008). A hydrophobic face containing 5 amino acids was predicted on an alpha helix formed by residues 15-33. Moreover, the side of the helix that opposed the hydrophobic face was saturated with S&T residues and lacked acidic residues. These features are characteristic of membrane-associating amphipathic helices (Lee et al., 2008). Thus, the CT of OEP16-2 may contain an amphipathic alpha helix from residues 159-178 with moderate hydrophobicity. This structure may function as a targeting feature in the predicted CT TP-like targeting signal.

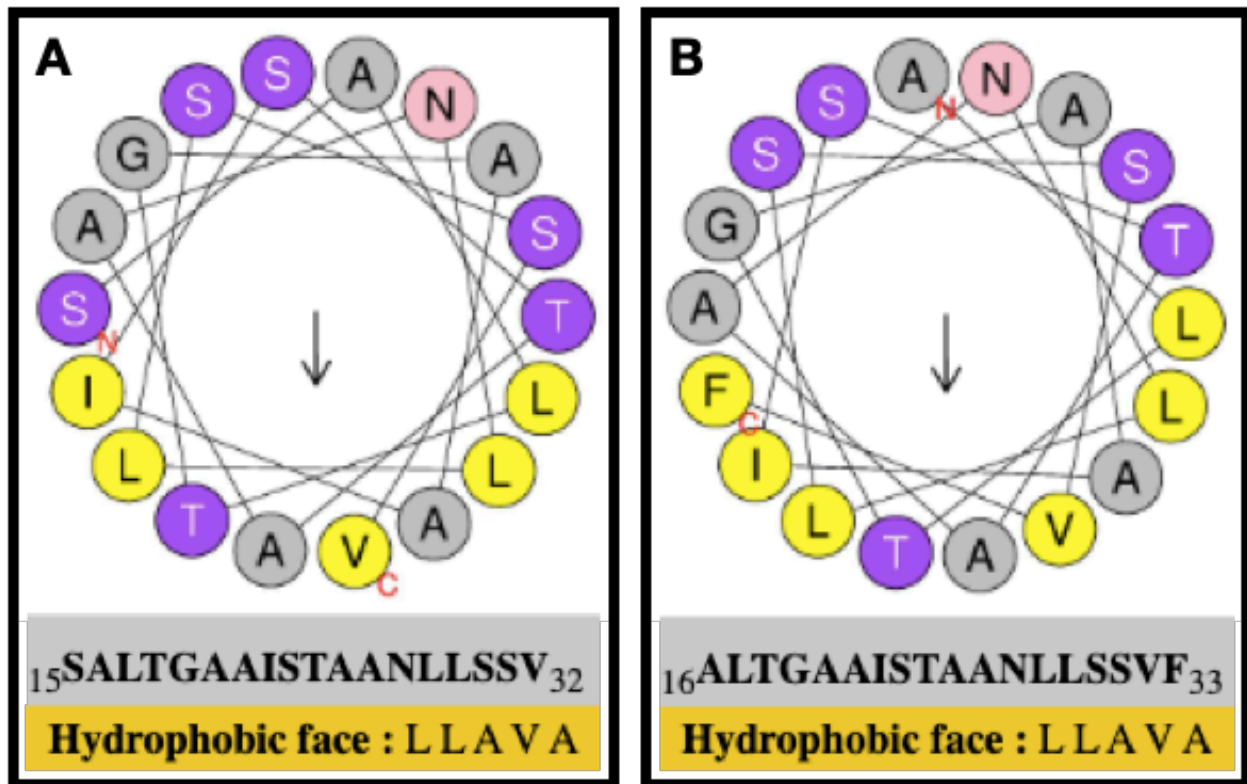


Figure 3.5. Helical Wheel Projection of the OEP16-2 C-terminal 33 Amino Acids. The 33 aa residues of the OEP16-2 CT was analyzed in frames of 18 aa for hydrophobic faces. A hydrophobic face was identified in two frames, A includes residues 15-32, B includes residues 16-33. N and C denote the first and last aa in-frame. Amino acids composing the hydrophobic face include: L L A V A.

3.8. Summation of OEP16-2 Computational Predictions

In summary, multiple computational analysis tools supported the prediction of a 33 aa long CT TP-like signal in OEP16-2 that was made by ChloroP. ProtParam analysis revealed aa trends similar to trends found in NT TPs and an MSA of OEP16-2 sequence demonstrated the conservation of these trends. Additionally, a moderately hydrophobic alpha helix with amphipathic properties was predicted by Phobius, PSI-Pred, and HeliQuest Analysis. Some classical TPs are disordered under cytosolic conditions, however, form a stable membrane embedded alpha helix under membrane mimetic conditions. The predicted OEP16-2 CT amphipathic alpha helix could potentially function in a similar manner. Collectively, these features made OEP16-2 a suitable candidate for further localization studies, wherein the CT targeting function was assessed.

Chapter 4. Identifying the chloroplast targeting signal within the OEP16-2 Sequence by Epifluorescence

4.1. Overview

Fluorescent fusion constructs containing various truncations of the OEP16-2 sequence were transiently expressed in onion epidermal cells. The resulting expression patterns were observed using epifluorescent microscopy. Constructs were co-bombarded with a plastid marker, DsRed fused to the TP of ferredoxin, to visually assess co-localization to plastids. EGFP and DsRed signals were captured and overlaid to assess the degree of co-localization for each construct. The plastid-targeting regions within the OEP16-2 sequence were identified.

4.2. Localization of Six Original Fluorescent Constructs after Onion Cell Bombardment

Initially, six fluorescent OEP16-2 fusion constructs were constructed to examine targeting function of the OEP16-2 33 CT amino acids and to assess the effect of EGFP orientation on protein localization (Figure 4.1). OEP16-2 sequences were subcloned into the two vectors pSAT6N1 and pSAT6C1, which have an MCS flanking the NT or CT of EGFP, respectively. Each EGFP fusion construct included either the full-length sequence (OEP16-2-FL), a truncated sequence lacking the CT (OEP16-2 Δ 33CT), or a truncated sequence containing the 33 CT amino acids (OEP16-2-33CT).

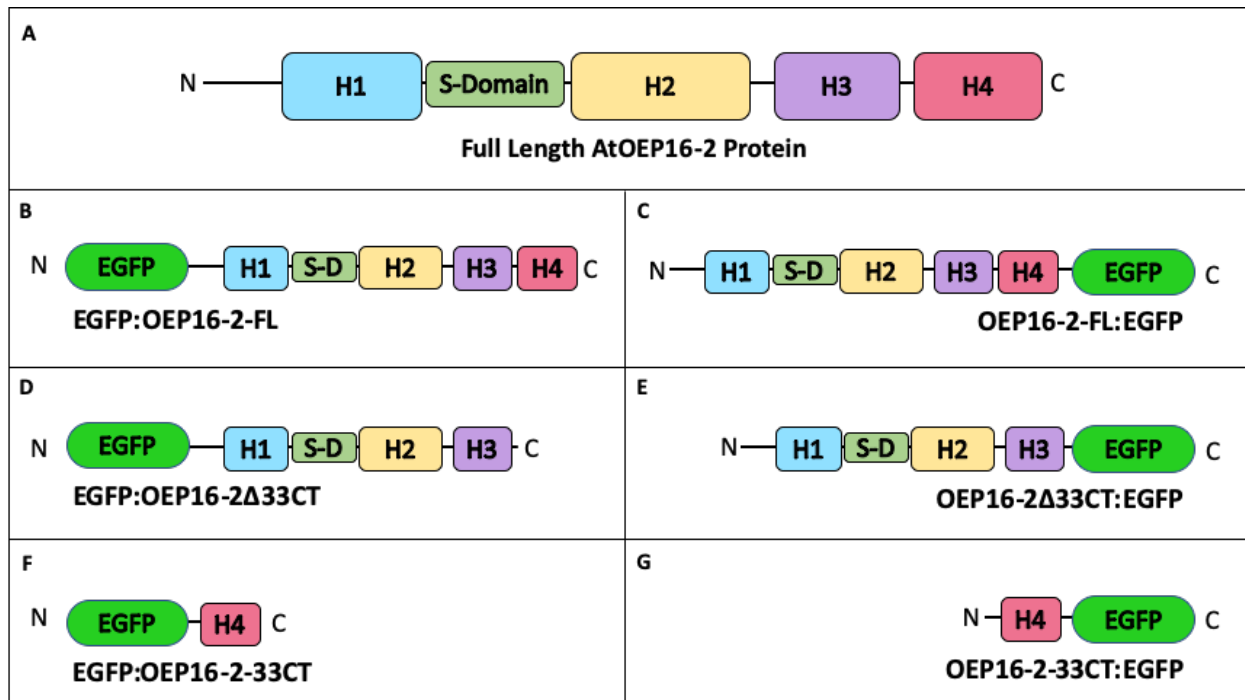


Figure 4.1. Design of Six OEP16-2:EGFP Fluorescent Fusion Constructs. The Full Length AtOEP16-2 Protein illustration (A) depicts the full length OEP16-2 sequence and its predicted domains, H1, H2, H3, H4, & the S-domain. Six fluorescent fusion constructs were designed to assess the targeting capacity of the C-terminal domain H4 (B-G). The pSAT6C1 vector was used to fuse full-length and truncated OEP16-2 sequences to the C-terminus of EGFP (B,D&F) while the pSAT6N1 vector was used to fuse full-length and truncated OEP16-2 sequences to the N-terminus of EGFP (C,E&G). OEP16-2-FL constructs contain the full-length sequence (B&C), OEP16-2Δ33CT constructs lack the 33 CT amino acids (D&E) effectively removing the H4 domain, while OEP16-2-33CT constructs solely contain the 33 CT amino acids (F&G).

We hypothesized that OEP16-2 contained a CT TP-like targeting signal. Thus, we expected constructs containing the OEP16-2-FL & OEP16-2-33CT sequence would target plastids, as they contain the pertinent targeting information based upon the current analysis. On the other hand, a cytosolic localization pattern was expected for OEP16-2Δ33CT constructs, as the necessary targeting information was removed. Lastly, we reasoned that OEP16-2 fusion

to the N-terminus of EGFP may inhibit plastid targeting by physically obstructing the targeting features predicted at the CT of OEP16-2. Thus, it was expected that constructs subcloned into the pSAT6N1 vector would localize to the cytoplasm, regardless of CT presence. Intriguingly, the opposite trends were observed when constructs were transiently expressed in onion epidermal cells (Table 3.1). OEP16-2-FL:EGFP and OEP16 Δ 33CT:EGFP localized to plastid-like punctate structures (Figure 4.2) and co-localized with the plastid marker (Figure 4.3). Moreover, OEP16-2-33CT:EGFP, as well as, all of the constructs subcloned into the pSAT6C1 vector did not exhibit plastid-targeting (Figure 4.2 & Figure 4.3).

Therefore, the CT sequence of OEP16-2 was not deemed necessary or sufficient for targeting. Instead, the data indicate that a region within residues 1-145 of the OEP16-2 functions in plastid-targeting. Additionally, OEP16-2 fusion to the C-terminus of EGFP inhibited plastid localization, indicating plastid-targeting is inhibited by physical obstruction of the OEP16-2 at the NT, not at the CT as expected. As such, additional experiments were designed to uncover the specific plastid-targeting region within the OEP16-2 sequence. Additional constructs contained various OEP16-2 domain truncations and were subcloned into pSAT6N1 vector. The pSAT6C1 vector was not used because the orientation of EGFP fusion to OEP16-2 inhibited targeting.

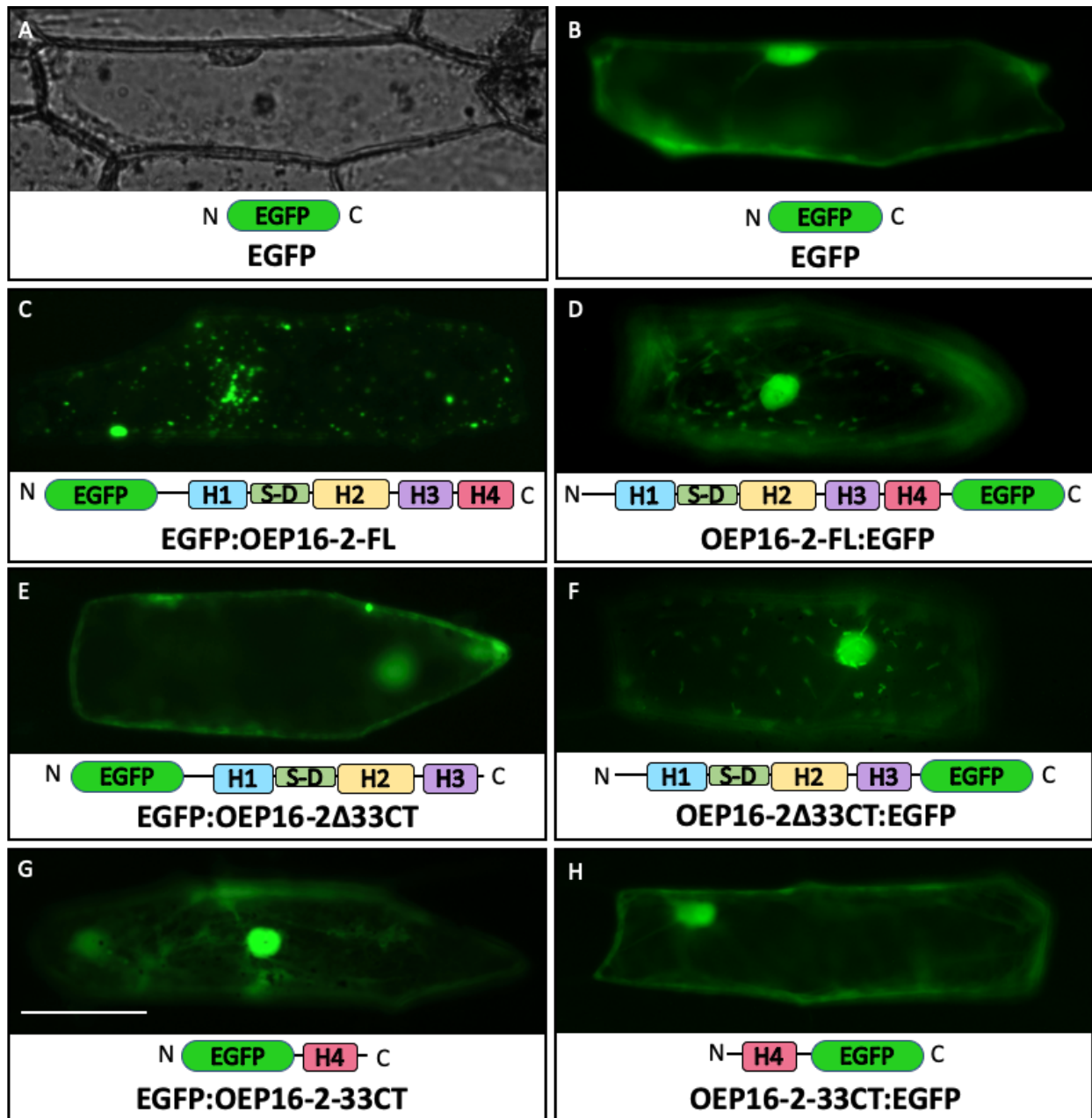


Figure 4.2. Onion Cell Bombardment with Original Six OEP16-2 Fluorescent Fusion Constructs. Onion epidermal cells were visualized using epifluorescent microscopy following biolistic bombardment with a fluorescent construct. (B-H) EGFP expression pattern, (A) bright-field image onion cell in (B). (B,E,G&H) EGFP, EGFP:OEP16-2 Δ 33CT, EGFP:OEP16-2-33CT, and OEP16-2-33CT:EGFP localized to cytosol and nucleus. (C) EGFP:OEP16-2-FL localized to structures varying in shape and size. (D&F) OEP16-2:EGFP-FL & OEP16-2 Δ 33CT:EGFP localized to plastid-like punctate structures. Scale bar = 0.1 mm.

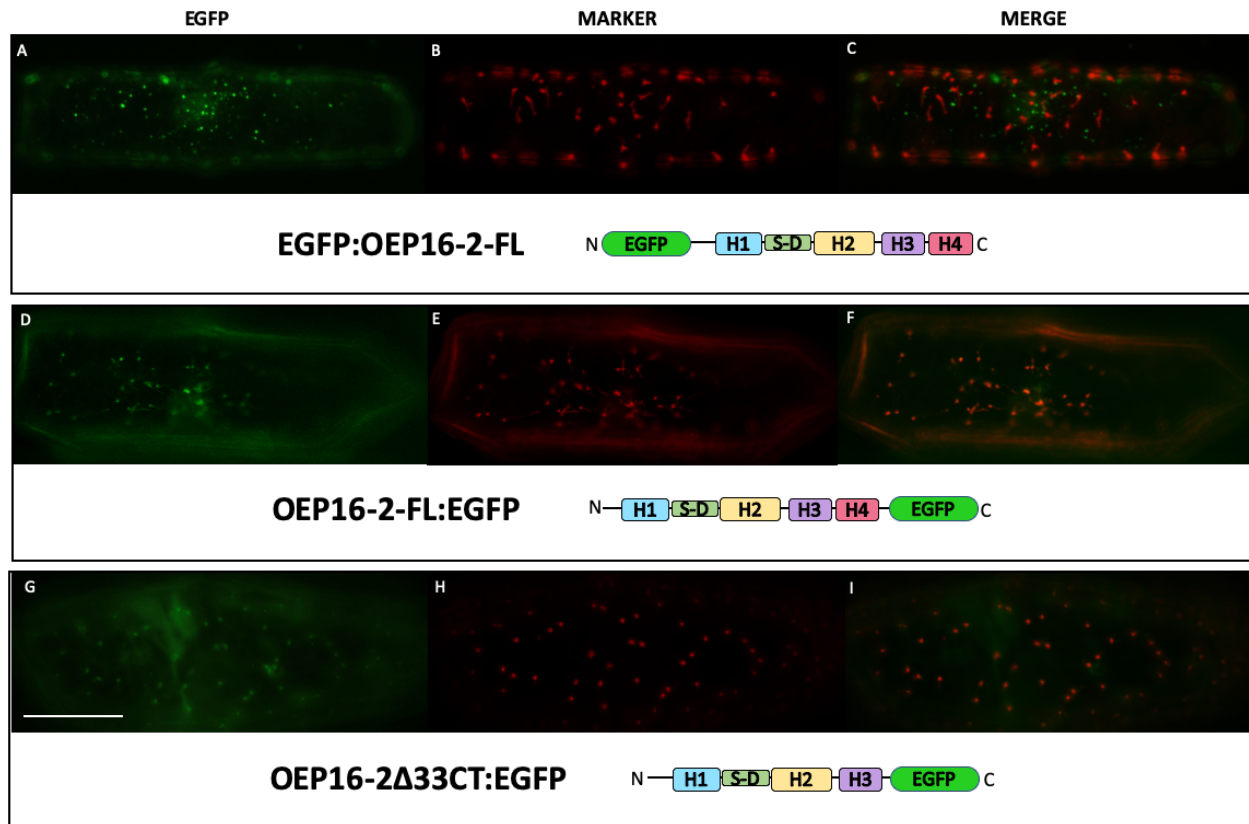









Figure 4.3. Onion Cell Bombardment Co-localization of Fusion Constructs and a Plastid Marker. Constructs localizing to punctate structures were co-bombarded with a plastid marker, the TP of ferredoxin fused to DsRed. (A,D&G) EGFP fluorescence, (B,E&H) DsRed fluorescence. (C,F&I) merged EGFP and DsRed signals. EGFP:OEP16-2-FL (A) and the plastid marker (B) did not co-localize (C). OEP16-2-FL:EGFP (D) and the plastid marker (E) co-localized (F). Additionally, OEP16-2 Δ 33CT:EGFP (G) and the plastid marker (H) colocalized (I). Scale bar = 0.1 mm.

Table 4.1. Expected vs. Observed Localization of Six Initial Fusion Constructs Transiently Expressed in Onion Epidermal Cells.

Construct Name	Construct Illustration	Expected Localization	Observed Localization
EGFP		Cytoplasm	Cytoplasm
EGFP:OEP16-2-FL		Plastid	Cytoplasm/ Undefined
EGFP:OEP16-2Δ33CT		Cytoplasm	Cytoplasm
EGFP:OEP16-2-33CT		Plastid	Cytoplasm
OEP16-2-FL:EGFP		Plastid or Cytoplasm	Plastid
OEP16-2Δ33CT:EGFP		Cytoplasm	Plastid
OEP16-2-33CT:EGFP		Plastid or Cytoplasm	Cytoplasm

A summary of the expected versus observed localization pattern for six fluorescent fusion constructs. The name and an illustration of each construct correspond to constructs outlined in Figure 3.2.1. Localization is denoted as Plastid, Cytoplasm, Plastid and/or Cytoplasm, or Undefined. All pSAT6C1 fusion constructs localized to the cytoplasm which included: EGFP:OEP16-2-FL, EGFP:OEP16-2Δ33CT, & EGFP:OEP16-2-33CT. Of the pSAT6N1 constructs, OEP16-2-FL:EGFP and OEP16-2Δ33CT:EGFP localized to plastids, while OEP16-2-33CT:EGFP localized to the cytoplasm.

4.3. Design and Transient Expression of Domain Truncation Constructs

OEP16-2 domain truncation constructs were designed to assess the targeting capacity of both individual domains and domain combinations (Figure 4.6). The position of domains in the primary structure of OEP16-2 has not been experimentally validated, thus, domain positions were predicted using computational secondary structural analysis and sequence alignments. The structural analysis tool PSI-PRED predicted five alpha helices in the OEP16-2 sequence

(Figure 3.3). The tool Phobius predicted three transmembrane domains that overlapped with three of the helices predicted by PSI-PRED (Figure 3.4; Figure 4.5). An MSA generated by Drea et al. (2006) was used to infer OEP16-2 domain positions through alignments with OEP16-1 isoforms (Figure 4.4). The MSA was generated from OEP16-1 and OEP16-2 sequences, the position of alpha helices predicted by CD analysis of PsOEP16-1 was overlain on the MSA to infer domain positions (Drea et al., 2006). The more recently published NMR structure of PsOEP16-1 was overlain on this MSA and strongly correlated with the previous CD prediction (Zook et al., 2013; Figure 4.4). When the MSA inferred domains and computational predictions were compared, four alpha helical regions are consistently identified (Figure 4.5). The predicted amino acid position of these four helices, denoted H1-H4, and the S-domain (S-D) was used to design the subsequent fusion constructs. However, future CD analysis of OEP16-2 will be necessary to determine the correct position of these domains.

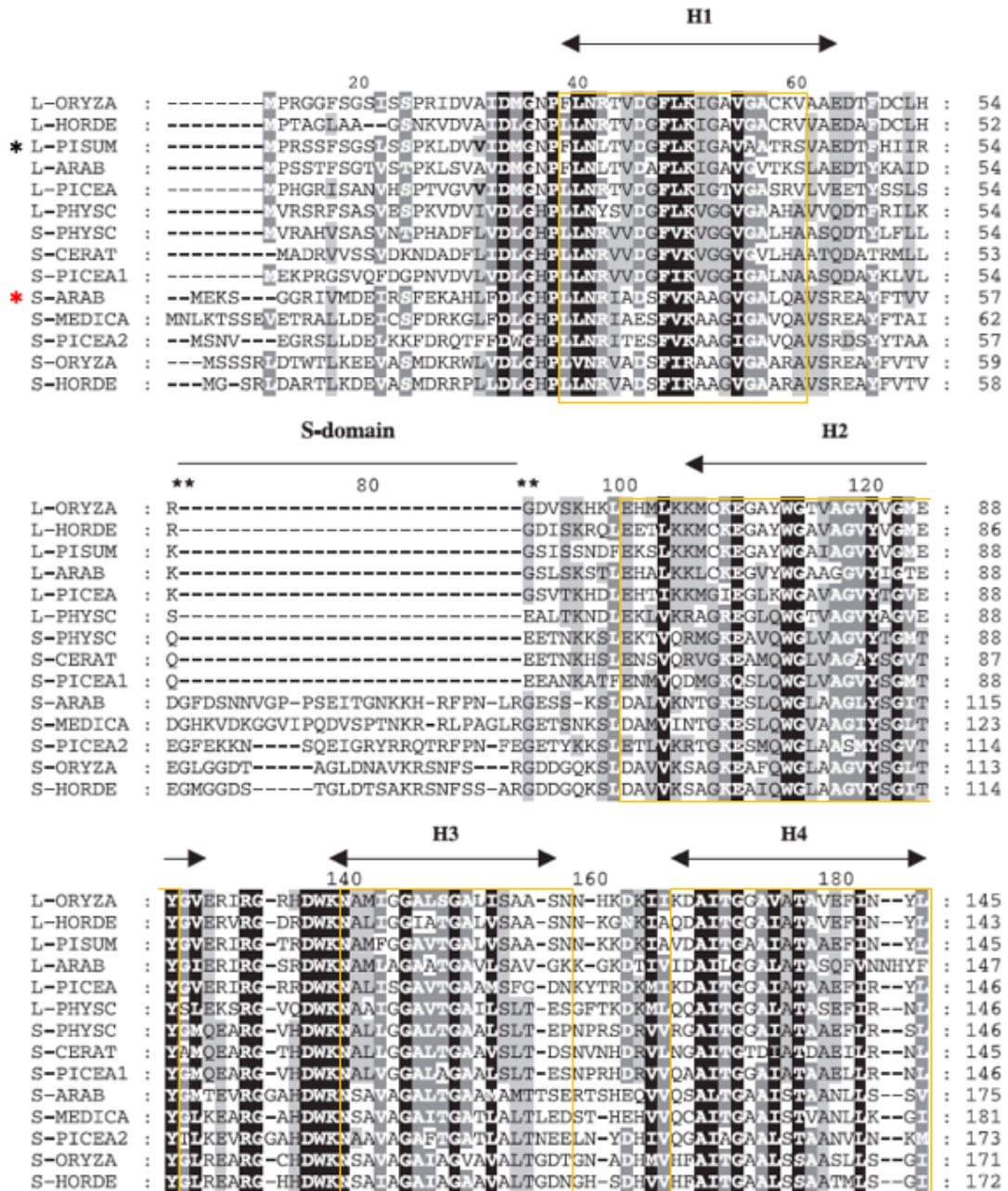


Figure 4.4. Overlaid NMR OEP16-1 structure onto an MSA created by Drea et al (2006). Drea et al. (2006) generated a multiple sequence alignment from OEP16-1 (L) and OEP16-2 (S) sequences. Then, overlaid the amino acid position of secondary structures predicted by CD analysis of PsOEP16-1, represented by the arrowheads under H1, H2, H3, H4, and the S-Domain (Linke et al., 2004). The amino acid position of secondary structures predicted by NMR analysis of PsOEP16-1 are indicated by yellow boxes (Zook et al., 2013). The CD and NMR prediction are

in good agreement. Asterisks indicate the PsOEP16-2 sequence (black) and the AtOEP16-2 sequence (red).

Figure 4.4. Overlaid NMR OEP16-1 structure onto an MSA created by Drea et al (2006). Adapted from Gene duplication, exon gain and neofunctionalization of OEP16-related genes in land plants by Drea, S. C., Lao, N. T., Wolfe, K. H. & Kavanaugh, T. A., 2006. Retrieved from The Plant Journal. Copyright 1999-2020 by John Wiley & Sons Inc.

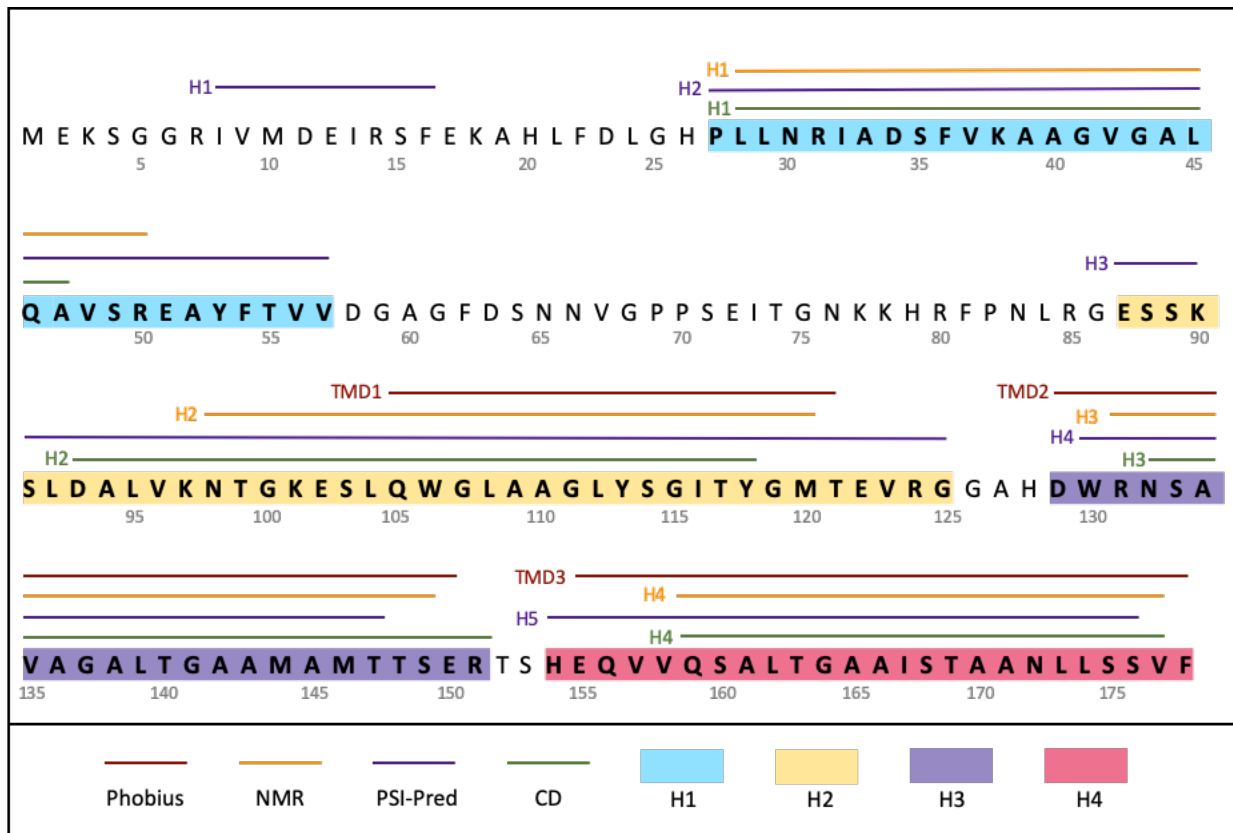


Figure 4.5. AtOEP16-2 Domain Prediction. The predicted position of domains H1 (blue), H2 (yellow), H3 (purple), and H4 (pink) are highlighted within the AtOEP16-2 sequence. The position of structures predicted by Phobius and PSI-Pred, as well as structures inferred by the alignment of OEP16-1 CD and NMR analysis (Figure 3.2.4), are represented as coloured lines above the sequence. All predictions are in good agreement. The position of each domain encompasses all predicted amino acids from each analysis, excluding amino acids of the H1

prediction made by PSI-Pred. Domains are positioned as follows, H1: 27-57aa, H2: 87-125aa, H3: 129-151aa, and H4: 154-178aa.

Domain truncation constructs were designed in pairs, each pair contained an opposite set of domains and assessed a different combination of domains (Figure 4.6). Construct OEP16-2 Δ 33CT contained the H1, S-D, H2 & H3 domains, while OEP16-2-33CT contained the H4 domain. OEP16-2 Δ 53CT and OEP16-2-53CT included the H1, H2, & S-domain, and the H3&H4 domain, respectively. OEP16-2 Δ 96CT and OEP16-2-96CT contained the H1&S-domain, and the H2, H3, & H4 domain, respectively. Lastly, OEP16-2 Δ 121CT and OEP16-2-121 included the H1 domain, and the S-domain, H2, H3, & H4 domain, respectively (Figure 4.6). All plastid-targeting constructs were expected share one or more similar domains while all constructs lacking one or more similar domains would localize to the cytoplasm.

Constructs containing either the S domain, H2 domain, or H3 domain localized to plastid-like punctate structures, including, OEP16-2 Δ 33CT, OEP16-2 Δ 53CT, OEP16-2-53CT, OEP16-2 Δ 96CT, OEP16-2-96CT, and OEP16-2-121 (Figure 4.7; Table 4.2). Constructs lacking the S, H2, or H3 domains localized to the cytoplasm, including, OEP16-2-33CT and OEP16-2 Δ 121CT (Figure 4.7; Table 4.2). S, H2, and H3 domain containing constructs were co-bombarded with a plastid marker and demonstrated co-localization to plastids (Figure 4.8). This indicated that the predicted H2 and H3 domains likely contain information that is necessary and sufficient for plastid targeting (Table 4.2). To further investigate the targeting capacity of the S and H2 domains a set of S domain and H2 domain truncation constructs were created.

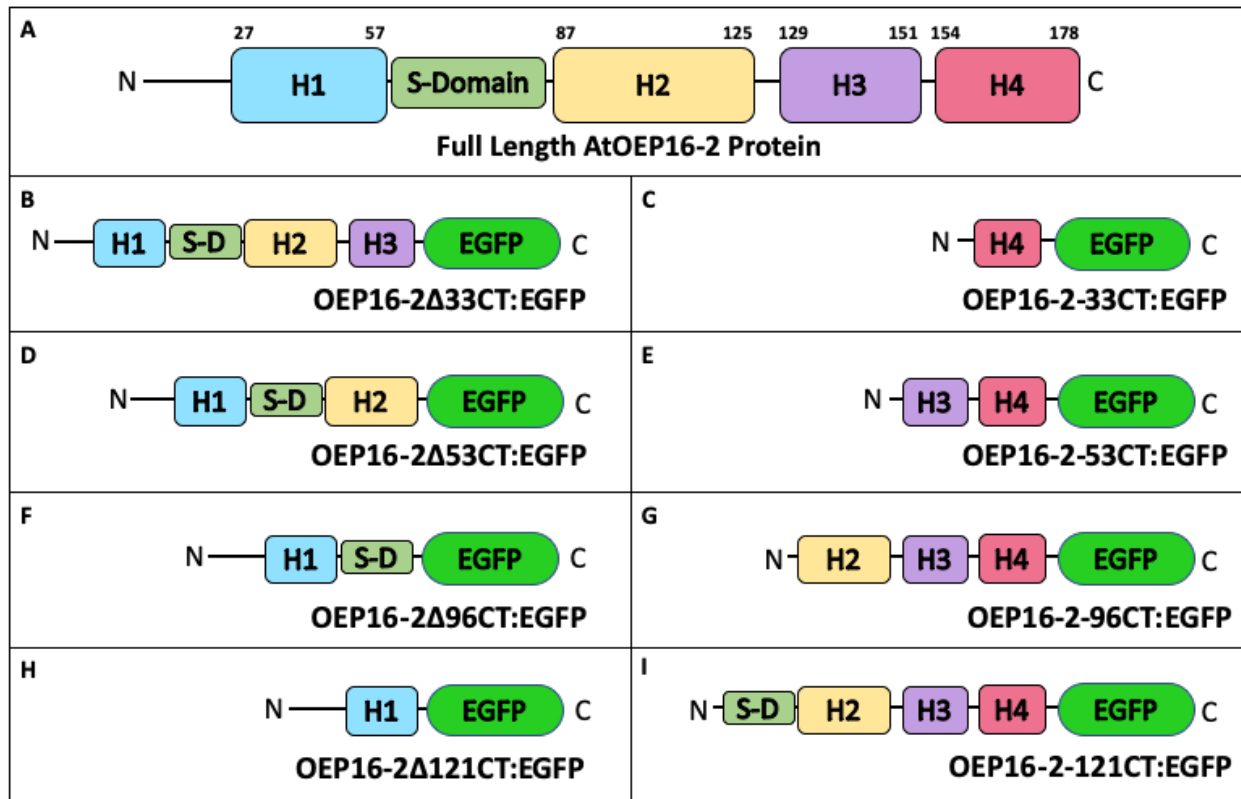


Figure 4.6. Design of Domain Truncation Fusion Constructs. The Full Length AtOEP16-2 Protein illustration (A) depicts the predicted aa positions of domains H1, H2, H3, H4, and the S-domain. (B-I) Eight fluorescent fusion constructs were designed to assess the targeting capacity of individual domains and domain combinations. (B-I) The pSAT6N1 vector was used to fuse truncated OEP16-2 sequences to the N-terminus of EGFP. (B,D,F&H) Constructs are donated as Δ XCT, where X indicates the number of removed C-terminal amino acids. (C,E,G&I) Constructs XCT, where X indicates the number of CT amino acids in the construct. Constructs are designed in pairs (B&C, D&E, F&G, H&I) to assess the targeting capacity of constructs with or without each individual domain. Constructs B&C were previously designed as part of the original six fusion constructs (Figure 3.2.1).

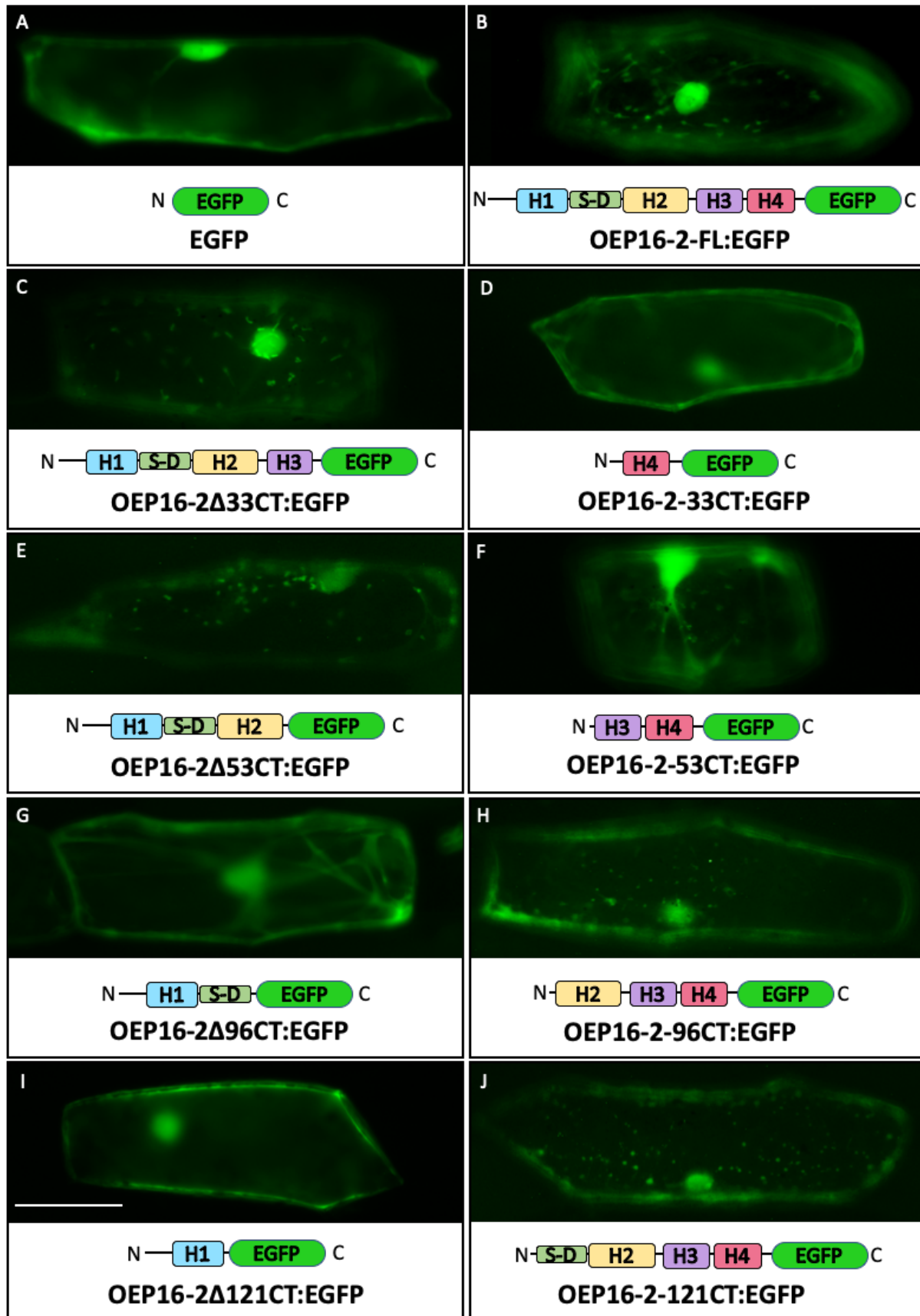


Figure 4.7. Onion cell bombardments of Domain Truncation Fusion Constructs. Onion epidermal cells were visualized using epifluorescent microscopy following biolistic bombardment with a fluorescent construct. (B,C,E,F,H&J) Plastid-like punctate structures were observed in cells transiently expressing constructs that contained the H2 and/or H3 domain. (A,D,G&I) Cytoplasmic and nuclear localization patterns were observed in cells transiently expressing constructs which lacked the H2 and H3-domains. Scale bar = 0.1 mm.

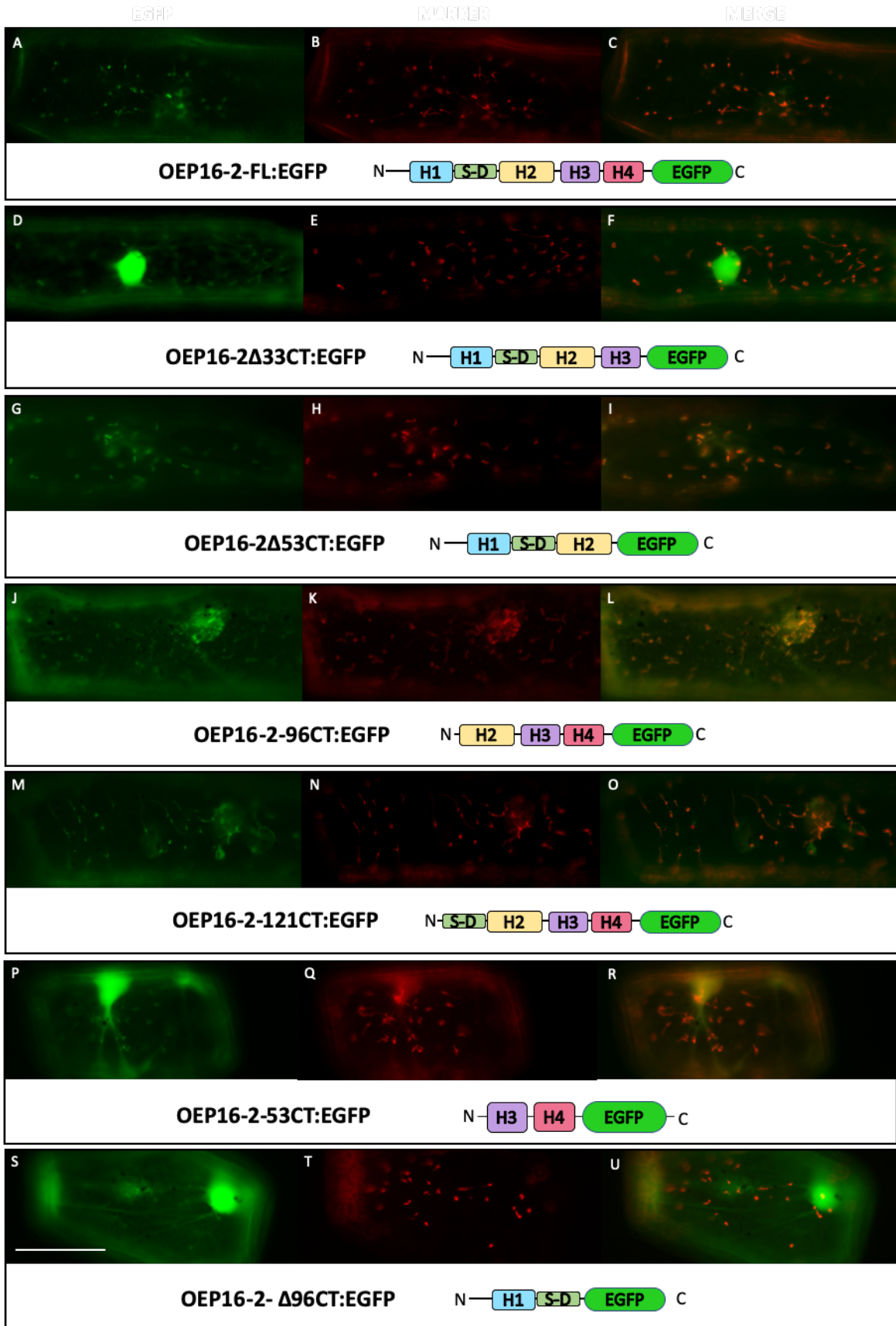









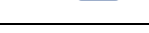


Figure 4.8. Onion Epidermal Cell Co-bombardment of Domain Truncation Fusion Constructs with a Plastid Marker. Constructs localizing to punctate structures were co-bombarded with a plastid marker, the TP of ferredoxin fused to DsRed. (A,D,G,J,M&P) EGFP fluorescence, (B,E,H,K,N&Q) DsRed fluorescence. (C,F,I,L,O&R) EGFP and DsRed signals were merged to assess co-localization. (C,F,J,M,P&S) Each EGFP fusion construct co-localized with the co-bombarded plastid marker. Scale bar = 0.1 mm.

Table 4.2. Localization of Domain Truncation Constructs in Onion Epidermal Cells.

Construct Name	Construct Illustration	Construct Localization
EGFP		Cytoplasm
OEP16-2-FL		Plastid
OEP16-2-33CT		Cytoplasm
OEP16-2-53CT		Plastid
OEP16-2-96CT		Plastid
OEP16-2-121CT		Plastid
OEP16-2-Δ33CT		Plastid
OEP16-2-Δ53CT		Plastid
OEP16-2-Δ96CT		Plastid
OEP16-2-Δ121CT		Cytoplasm

Each construct was transiently expressed in onion epidermal cells via biolistic bombardment. The construct name and illustration correspond to constructs outlined in Figure 3.2.6. Constructs OEP16-2-FL, -53CT, -96CT, -121CT, Δ33CT, & Δ53CT all localized to plastids and contain the H2 domain and/or the H3 domain. Conversely, EGFP as well as constructs OEP16-2-33CT, Δ96CT, & Δ121CT all localized to the cytoplasm and do not contain the H2 domain or the H3 domain.

4.4. Transient Expression of H2-Domain Constructs in Onion Epidermal Cells

Two H2 domain containing constructs were synthesized including, OEP16-2-H2:EGFP, and EGFP:OEP16-2-H2 (Figure 4.9). OEP16-2-H2 EGFP fusion constructs contained the H2 domain and one preceding glycine residue, residue positions 86-125 in the full-length OEP16-2 sequence. EGFP fusion to the OEP16-2 NT inhibited plastid targeting in previous experiments (section 4.2). Yet, the H2 domain is located internally within the OEP16-2 sequence and N-terminally flanked by approximately 86 aa, begging the question why the NT addition of EGFP inhibits an internal targeting signal. EGFP is significantly larger than the NT flanking 86aa in OPE16-2 and therefore it may be size of EGFP that inhibits targeting. Thus, EGFP:OEP16-2-H2 was designed and transiently expressed to further explore this observation. EGFP:OEP16-2-H2 localized to the cytoplasm when transiently expressed in onion cells, and did not co-localize with the plastid marker, indicating the EGFP to the NT of the H2 does indeed inhibit the targeting signal (Figure 4.10). On the other hand, OEP16-2-H2:EGFP co-localized with a plastid marker to plastid-like punctate structures, demonstrating the H2 domain is sufficient for plastid targeting and the CT fusion of EGFP does not inhibit targeting (Figure 4.10; Table 4.3). The orientation of EGFP at the NT of OEP16-2 may hinder the binding of chaperone proteins, thus inhibiting localization to the plastid OEM. Alternatively, the orientation of EGFP at the OEP16-2 NT may hinder the H2 domain from interacting directly with the OEM or receptors of the OEM, thus inhibiting plastid localization. In the future, the translocation pathway used by OEP16-2 can be investigated using protein-protein and protein-lipid interaction assays.

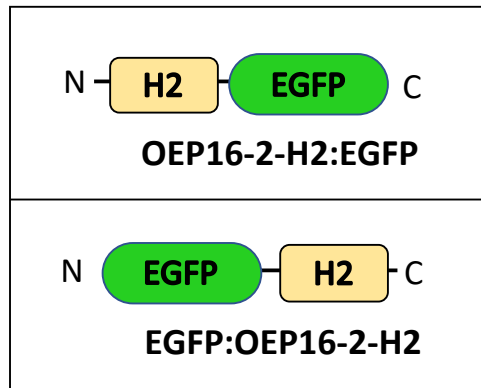


Figure 4.9. Design of H2 Domain EGFP Fusion Constructs. Two additional OEP16-2 fusion constructs were designed to assess the targeting function of H2. OEP16-2-H2 constructs contained the H2 domain fused to either the NT or CT of EGFP and assessed the effect of EGFP orientation on plastid-targeting.

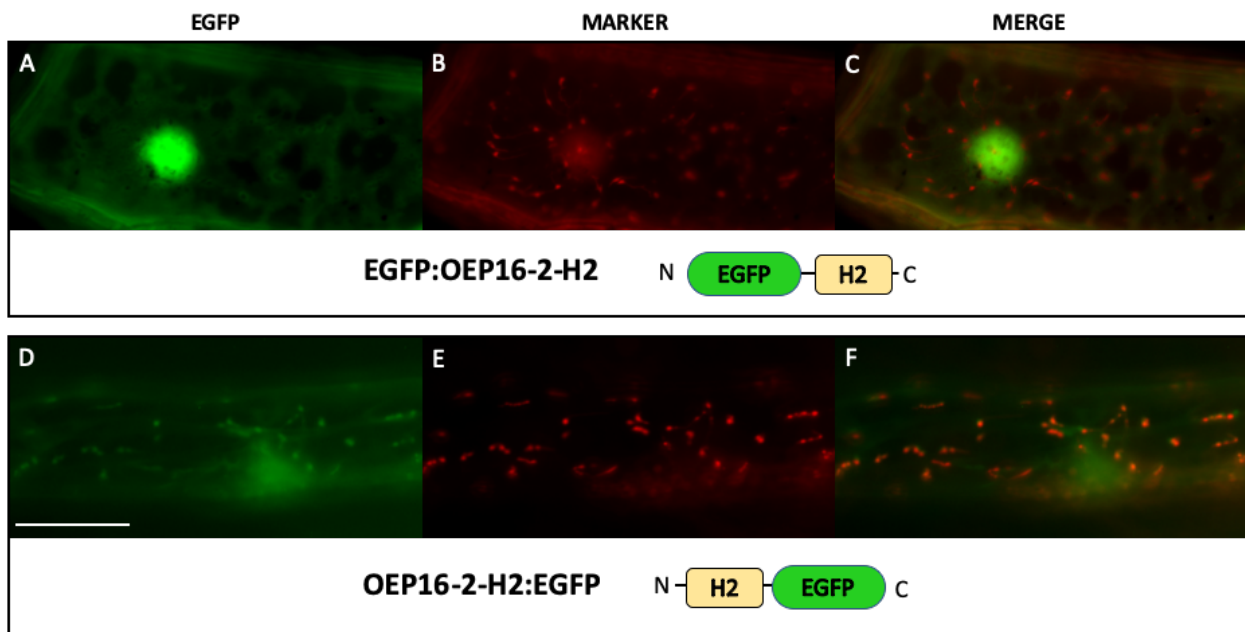






Figure 4.10. Onion Cell Co-bombardment of H2 Domain Fusion Constructs with a Plastid Marker. H2 domain fusion constructs were co-bombarded with a plastid marker, the TP of ferredoxin fused to DsRed. (A&D) EGFP fluorescence, (B&E) DsRed fluorescence. (C&F) EGFP and DsRed signals were merged to assess co-localization. OEP16-2 Δ H2:EGFP (A) and the plastid marker (B) did not co-localize (C). OEP16-2-H2:EGFP (D) and the plastid marker (E) co-localized (F). Scale bar = 0.1mm.

Table 4.3. Localization of H2 Domain Fusion Constructs in Onion Epidermal Cells.

Construct Name	Construct Illustration	Localization
EGFP		Cytoplasm
OEP16-2-FL:EGFP		Plastid
OEP16-2-H2:EGFP		Plastid
EGFP:OEP16-2-H2		Cytoplasm

Each construct was transiently expressed in onion epidermal cells by biolistic bombardment. EGFP & EGFP:OEP16-2-H2 localized to the cytoplasm while OEP16-2-H2:EGFP construct & OEP16-2-FL:EGFP localized to plastids.

4.5. Transient Expression of the S-Domain EGFP fusion Construct in Onion Epidermal Cells

An EGFP fusion construct containing the OEP16-2 S domain was constructed (OEP16-2-SD:EGFP; Figure 4.11). OEP16-2-SD:EGFP and a plastid marker (DsRed fused to the TP of ferredoxin) were co-bombarded in onion epidermal cells using biolistic bombardment (Figure 4.12). The OEP16-2-SD:EGFP signal co-localized to plastid-like structures with the plastid marker. This demonstrates that the OEP16-2 S domain likely contains a sufficient plastid targeting signal. However, the subcellular localization of OEP16-2-SD:EGFP must be validated by immunoblot analysis to confirm this localization pattern.

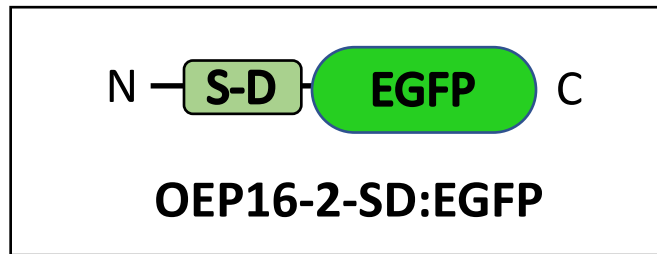


Figure 4.11. Design of OEP16-2-SD:EGFP Fusion Construct. The predicted S domain of OEP16-2 was subcloned into the MCS of the pSAT6N1 vector to create the construct OEP16-2-SD:EGFP. The S domain sequence was subcloned in-frame with a downstream EGFP sequence using three linker amino acids, GIL.

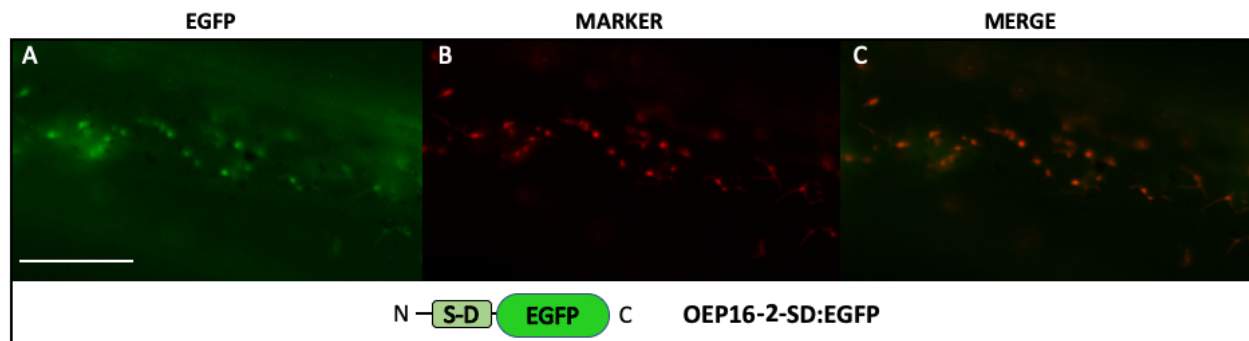


Figure 4.12. Onion Cell Co-bombardment with OEP16-2-SD:EGFP and a Plastid Marker. Onion epidermal cells were co-bombarded with OEP16-2-SD:EGFP (A) and a plastid marker (B). The EGFP and marker signals localized to the same plastid-like structures and these signals overlap when merged (C).

Chapter 5. Immunoblot EGFP Detection of Protoplast Subcellular Fractions

5.1 Overview

Subcellular fractionation and immunodetection were used to validate the expression patterns of fluorescent fusion proteins observed in bombarded onion epidermal cells.

Protoplasts were prepared and chemically-transfected with various OEP16-2 constructs.

Transfected protoplasts were fractionated into a total fraction, a soluble cytosolic fraction, and chloroplast insoluble fractions. Proteins were separated using SDS-PAGE, then transferred to and detected on a PVDF membranes using EGFP-specific antibodies.

5.2. Quality of Subfractionation and Immunoblot Analysis

Immunodetection of transfected protoplast fractions was used to validate the plastid subcellular localization of the OEP16-2 H2-domain. Ponceau stain was used to demonstrate the location of the RuBisCO large subunit (RbcL), the RuBisCO small subunit (RbcS) and the light harvesting complex (LHC) within the total and intact chloroplast fractions, which have expected sizes of 55kDa, 15kDa, and 26kDa, respectively (Pitzschke & Persak, 2012; Tiller et al., 2012).

RuBisCO leaks into the soluble fraction during protoplast lysis and is therefore not always a reliable plastid marker. However, the LHC of approximately 26kDa does not leak into the soluble fraction due to its association with the thylakoid membrane and functions as a plastid marker. The presence of the LHC in the insoluble fraction and absence in the soluble fraction demonstrates successful fractionation of plastid membranes. The two stained coloured ladder from Bio Basic Canada did not accurately align with the expected sizes of the RbcL, RbcS, and LHC. This trend was consistently seen throughout replicates. Therefore, the protein size

inferred by the ladder markers was slightly inaccurate and made some proteins appear smaller or larger than their true size. The recovery of plastid marker proteins would likely improve by using a second ladder for comparison or a greater number of cells for protoplast transfection and subfractionation. This methodology could also be improved by using an stable *Arabidopsis thaliana* line transformed with a plastid-protein fused to EGFP, such as EGFP:Toc159, which would function as a plastid-specific marker.

5.3. Immunodetection of EGFP and Fusion Constructs in Protoplast Fractions

Several constructs were transfected into protoplasts, then protoplasts were fractioned, and fractions were probed with EGFP antibodies. EGFP by itself localized to the cytosolic fraction and total protoplast fraction, demonstrating it does not contain plastid-targeting information (Figure 5.1). The OEP16-2-H2:EGFP construct localized to the intact chloroplast fraction and the total protoplast fraction (Figure 5.2). Thus, EGFP immunodetection demonstrates that the H2 domain contains sufficient plastid-targeting information and validates the subcellular localization pattern observed in bombarded onion epidermal cells. In contrast, the OEP16-2-H1:EGFP construct localized to the cytosolic fraction and total protoplast fraction, demonstrating that the H1 domain does not contain a sufficient plastid-targeting signal and the OEP16-2 targeting-signal is specific to the H2 domain (Figure 5.3). Thus, OEP16-2 plastid-targeting information is specific to the H2-domain and not present in the H1-domain.

In future experiments, an EGFP construct containing the OEP16-2 H3 domain will be transiently expressed in protoplasts for subsequent fractionation and western blot analysis. OEP16-2-H3:EGFP localization to plastids will demonstrate this domain is sufficient for plastid

targeting. Additionally, the sub-organelle location of various OEP16-2 constructs must be verified using density gradients to separate the chloroplast membranes and compartments. The presence of OEP16-2 in the OEM density layer will validate that OEP16-2 specifically localizes to the plastid OEM. A percoll gradient will be used to separate intact chloroplasts and sucrose gradients use density principles to separate the chloroplast membranes and compartments. It is expected that OEP16-2:EGFP constructs containing the S, H2, and H3 domains will localize to the OEM layer with the gradient.

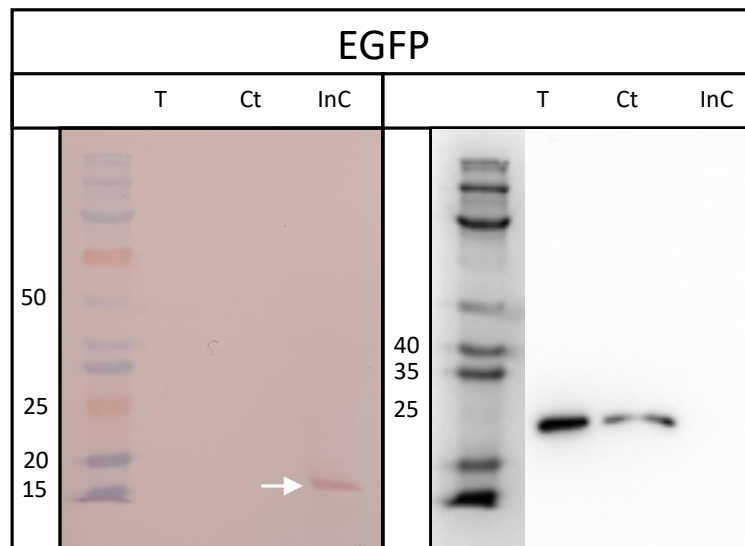


Figure 5.1. Ponceau Stain and Immunodetection of Fractionated Protoplasts Transfected with EGFP. Protoplasts fractionated into total protoplast (T), cytosolic (Ct), and intact chloroplast (InC) protein fractions. In the Ponceau stained blot, the RuBisCO small subunit can be seen at approximately 17kDa in the insoluble fraction but had an expected size of 15kDa (white arrow). In the immunoblot, EGFP is present in the total fraction and soluble fraction at approximately 25kDa. The expected size of EGFP was 26.9kDa.

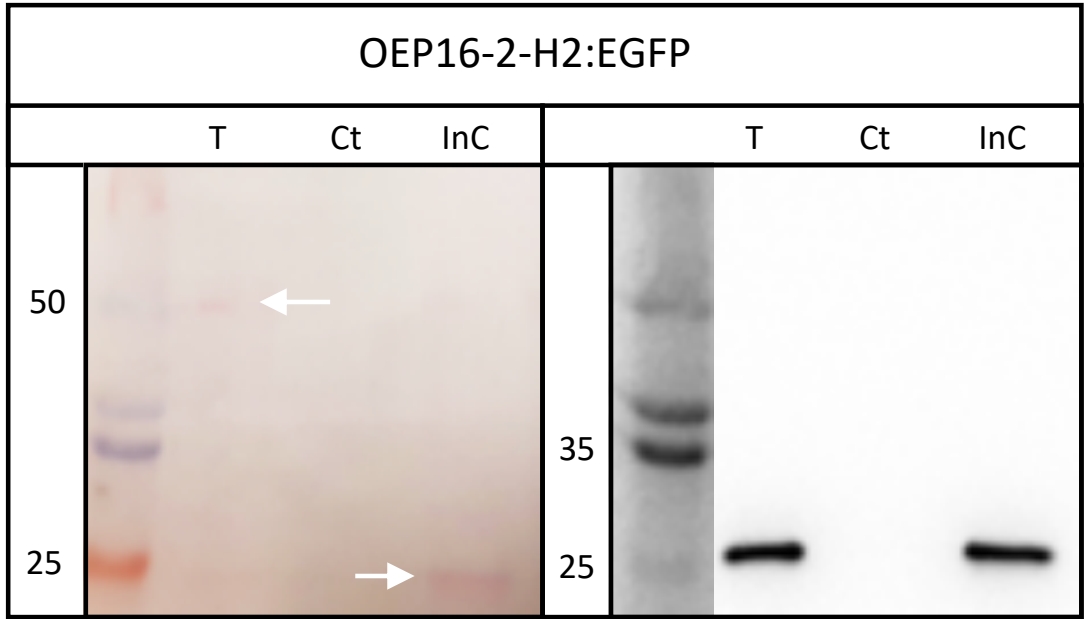


Figure 5.2. Ponceau Stain and Immunodetection of Fractionated Protoplasts Transfected with OPE16-2-H2. Protoplasts fractionated into total protoplast (T), cytosolic (Ct), and intact chloroplast (InC) protein fractions. In the Ponceau stained blot, RuBisCO can be faintly seen in the total protoplast fraction at approximately 50kDa, however was expected to appear at 55kDa (white arrow). Additionally, the light harvesting complex can be faintly seen at approximately 24kDa in the intact chloroplast fraction, however was expected at 26kDa (white arrow). In the immunoblot, the OEP16-2-H2 protein can be seen in the total fraction and soluble fraction just above the 25kDa marker, the expected size of OEP16-2-H2:EGFP was 31.5kDa.

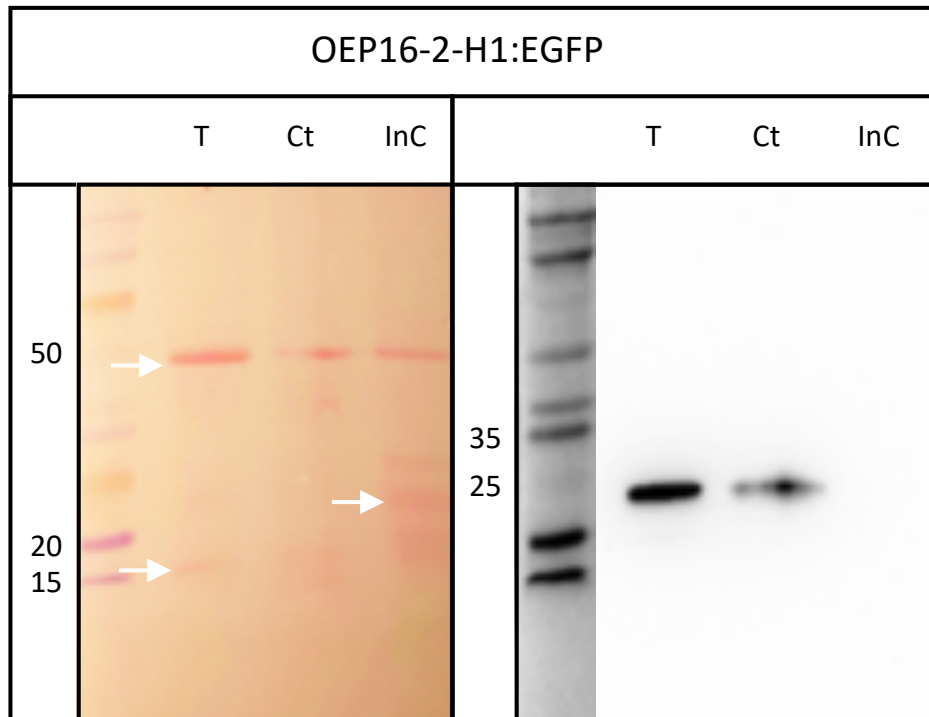


Figure 5.3. Ponceau stain and Immunodetection of Fractionated Protoplasts transfected with OEP16-2-H1. Protoplasts fractionated into total protoplast (T), cytosolic (Ct), and insoluble chloroplast (InC) protein fractions. In the Ponceau stained blot, RbcL can be seen in the at approximately 50kDa in all fractions, but had an expected size of 55kDa, RbcS can be faintly seen in the total protoplast fraction at approximately 17kDa, however had an expected size of 15kDa, and the light harvesting complex can be faintly seen at approximately 24kDa in the intact chloroplast fraction but had an expected size of 26kDa (white arrows). In the immunoblot, the OEP16-2-H1:EGFP protein can be seen in the total fraction (T) and cytosolic fraction (Ct) at approximately 25kDa but had an expected size of 30kDa.

Chapter 6. Computational Assessment of the H2 Domains Physicochemical Properties

6.1 Overview

To further elucidate the targeting function of the OEP16-2 H2 domain, physicochemical properties of this region were assessed to identify potential mechanisms which facilitate targeting. Plastid-targeting signals used by preproteins, SA-proteins, TA-proteins, and CT TP-like proteins share some common characteristics, including, negative charge, an abundance of S/T residues, and moderate hydrophobicity (Kim et al., 2019; Lee & Hwang, 2018; Teresinki et al., 2019). Therefore, it is probable that similar features are also present within the OEP16-2 targeting signal and facilitate plastid-targeting. Computational analysis of the OEP16-2 H2 domain was used to probe for similar features and identify possible targeting mechanisms. Tools including MSAs, ProtParam, Heli-Quest, Phobius, and PSI-PRED were used and future experiments are proposed based on these findings.

6.2. OEP16-2 Protein Sequence Conservation

OEP16-2 can be found in the genomes of many vascular plants and has a well-conserved protein sequence (Drea et al., 2006). An MSA of thirty OEP16-2 sequences from different species demonstrated that the predicted helical regions are well conserved throughout vascular plants (Figure 6.1; A4). The alignment contained species from a wide variety of clades including the OEP16-2 sequence from *Amborella trichopodea*, which is thought to be the most primitive living flowering plant (*Amborella* Genome Project, 2013). This indicates that OEP16-2 is broadly found throughout all flowering plant species, and that all five domains were present and conserved early in evolution (Figure 6.1). Interestingly, the S-domain is not present within a number of earlier evolved species, including the moss species *Physcomitrella patens*, the fern species *Ceratopteris richardii*, and the conifer species *Picea glauca* (Drea et al., 2006; Figure 1.6). Thus, the function of the S-domain is likely specific to the needs of flowering plants.

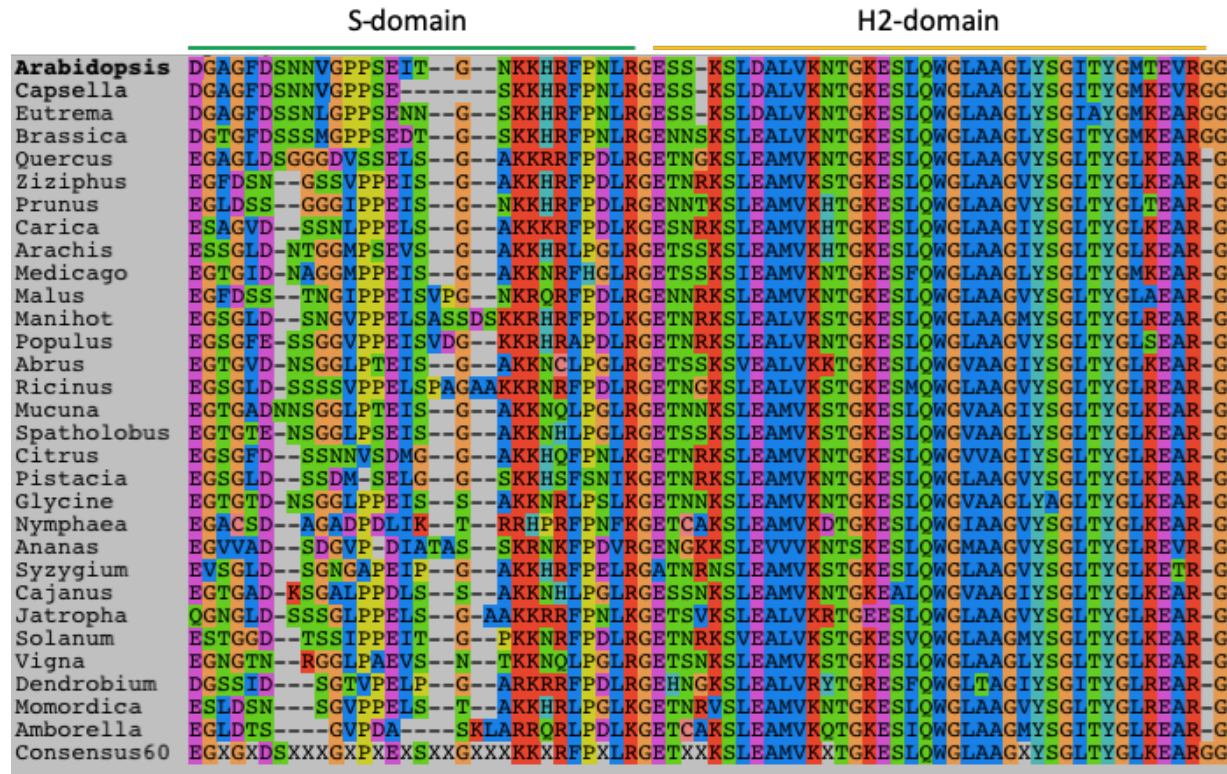


Figure 6.1. OEP16-2 Multiple Sequence Alignment from Various Plant Species. The OEP16-2 protein sequence from 30 plant species were aligned using the MUSCLE local alignment algorithm. Positions of the predicted S-domain and H2-domain are indicated above the alignment as a green and yellow bar, respectively. A consensus60 sequence was generated from the aligned residues, X represents a position with less than 60% aa conservation.

6.3. Properties and Patterns within the Predicted H2-Domain Sequence

The predicted H2 domain, made up of approximately 39 residues, contains two regions that have distinct compositions of amino acids (Figure 6.2). The N-terminal half of the H2 domain is composed of 30% charged residues, 35% polar residues, & 35% hydrophobic residues while the C-terminal half contains 10% charged residues, 25% polar residues, & 65% hydrophobic residues. Additionally, the grand average of hydrophobicity (GRAVY) score for the

NT region is -0.825 and for the CT region is 0.305, indicating a hydrophilic and hydrophobic region, respectively (Table 6.1). Collectively, this indicates the NT region is charged and hydrophilic, while the CT region is inert and hydrophobic. It is possible that the combined properties and structure of these distinct NT and CT regions function as a plastid targeting signal. However, it is also possible that only one of the two regions function as the plastid targeting signal. To test this theory, two constructs which include either the NT or CT terminal H2 domain region fused to the NT of EGFP will be constructed (Figure 6.5). If neither construct can target plastids when transiently expressed in onion cells, then both regions are essential for plastid-targeting. If one construct can target plastids, then the region contained within the plastid-targeting construct is essential for translocation. Lastly, if both constructs are capable of plastid-targeting, then alone each region is sufficient for plastid-targeting.

The distinct properties in the NT and CT regions of the H2 domain may not play any role in targeting. Instead a conserved sequence motif may function as the plastid-targeting signal. A notable motif in the CT region contains a series of four glycine residues spaced equally by three amino acids, G X₃ G X₃ G X₃ G. To test for sequence motifs, a series of alanine substitution constructs can be generated for future experiments. Each construct will have ten alanine substitutions which will overlap in five positions with another construct (Figure 6.5). The substituted regions within constructs that are unable to target plastids will indicate the essential targeting residues, and these residues should be conserved throughout OEP16-2 protein sequences.

It is also possible that the inherent secondary or tertiary structure of the H2-domain may function as a sufficient plastid-targeting feature. To explore this idea, structural properties

of the H2-domain were computationally analyzed. PSI-PRED predicted that the entire H2-domain, from residues 87-124, form an alpha helix (Figure 3.3). However, Phobius indicated that a transmembrane domain is present only in amino acids 104-120 (Figure 6.3). Additionally, Heli-quest predicted a hydrophobic face from amino acid 105-123 (Figure 6.4). This indicates that the H2-domain forms an amphipathic alpha helix, wherein the NT of the helix is hydrophilic and contains charged/polar residues while the CT end is a hydrophobic and transmembrane region. This amphipathic alpha helical structure may function as a plastid-targeting signal. To test this, amino acid substitution constructs can be designed which either relax the alpha helical structure, reduce the NT charge, or decrease the CT hydrophobicity. If these constructs are unable to target plastids it can be concluded that they function as a plastid targeting signal.

This series of H2-domain constructs (Figure 6.5) will provide insight into the properties that function as sufficient targeting features. Moreover, other OEPs can be investigated for these targeting features and additional protein candidates can be identified for localization assays. Once targeting features are identified, the mechanism and pathway used by these features to translocate to the OEM can be studied.

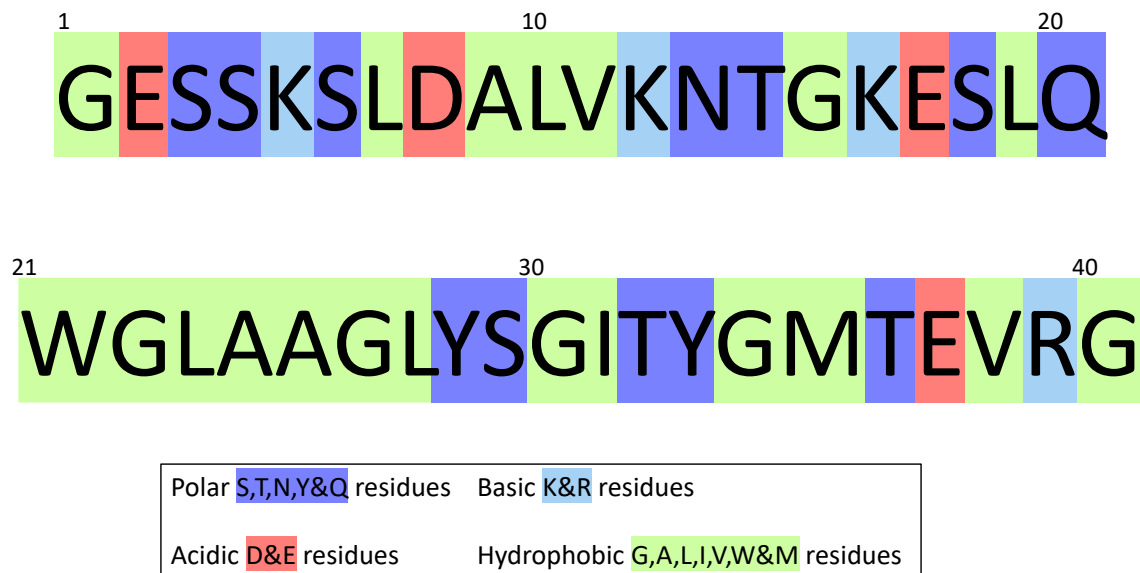


Figure 6.2. Amino Acid Trends within the Predicted H2-Domain. The amino acid sequence of the predicted H2-domain was colour coded to indicate residues that are polar (purple), non-polar (grey), basic (blue), acidic (red), and hydrophobic (green).

Table 6.1. Amino acid Composition Analysis of the Predicted H2 Domain

Property	Count & % Total	
	Amino Acid 1-20 (86-105)	Amino Acid 21-40 (106-125)
Negatively Charged Residues	3 – 15%	1 – 5%
Positively Charged Residues	3 – 15%	1 – 5%
Polar Residues	7 – 35%	5 – 25%
Hydrophobic Residues	7 – 35%	13 – 65%
GRAVY	-0.825	0.305

The total number and percentage of charged, polar, and hydrophobic residues within the first and second half of the predicted H2-Domain and the GRAVY score of each region.

FT TOPO_DOM 1 40 NON CYTOPLASMIC.
//

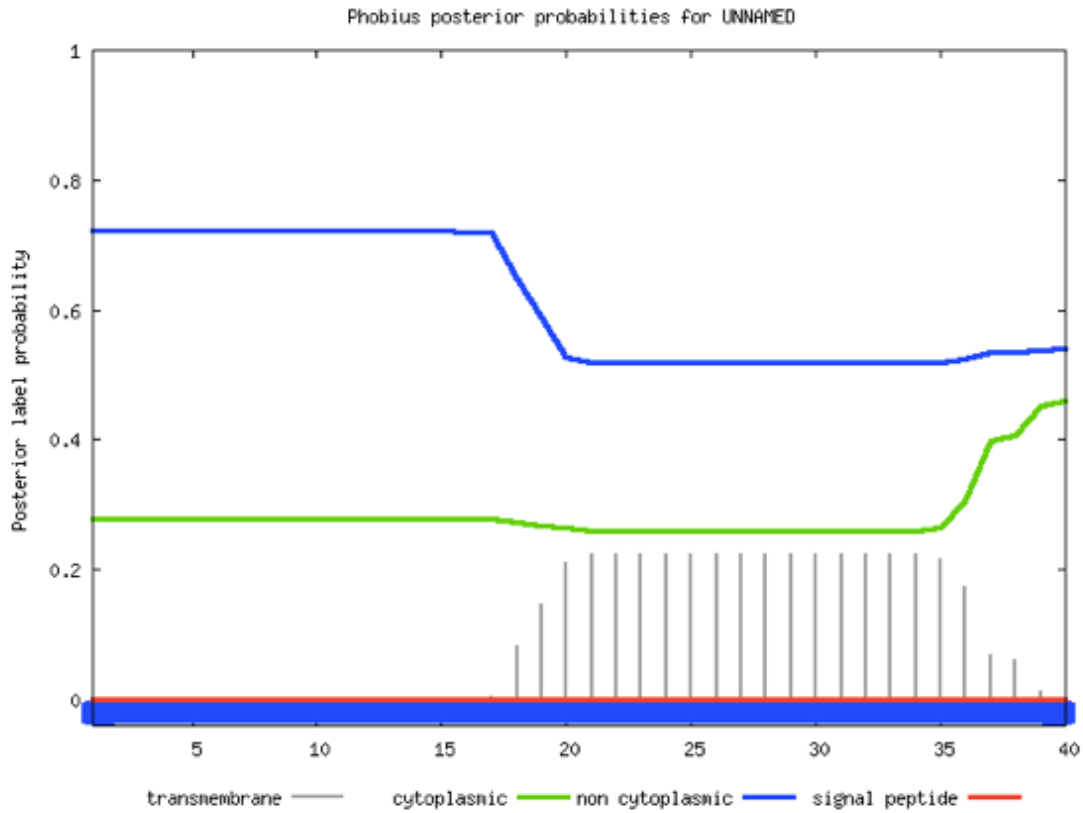


Figure 6.3. Phobius Prediction of the Predicted H2 Domain. The predicted H2 domain sequence contains a weak transmembrane region from approximately residues 20-35, which corresponds to residues 105-120 in the full length OEP16-2 sequence, as indicated by grey bars.

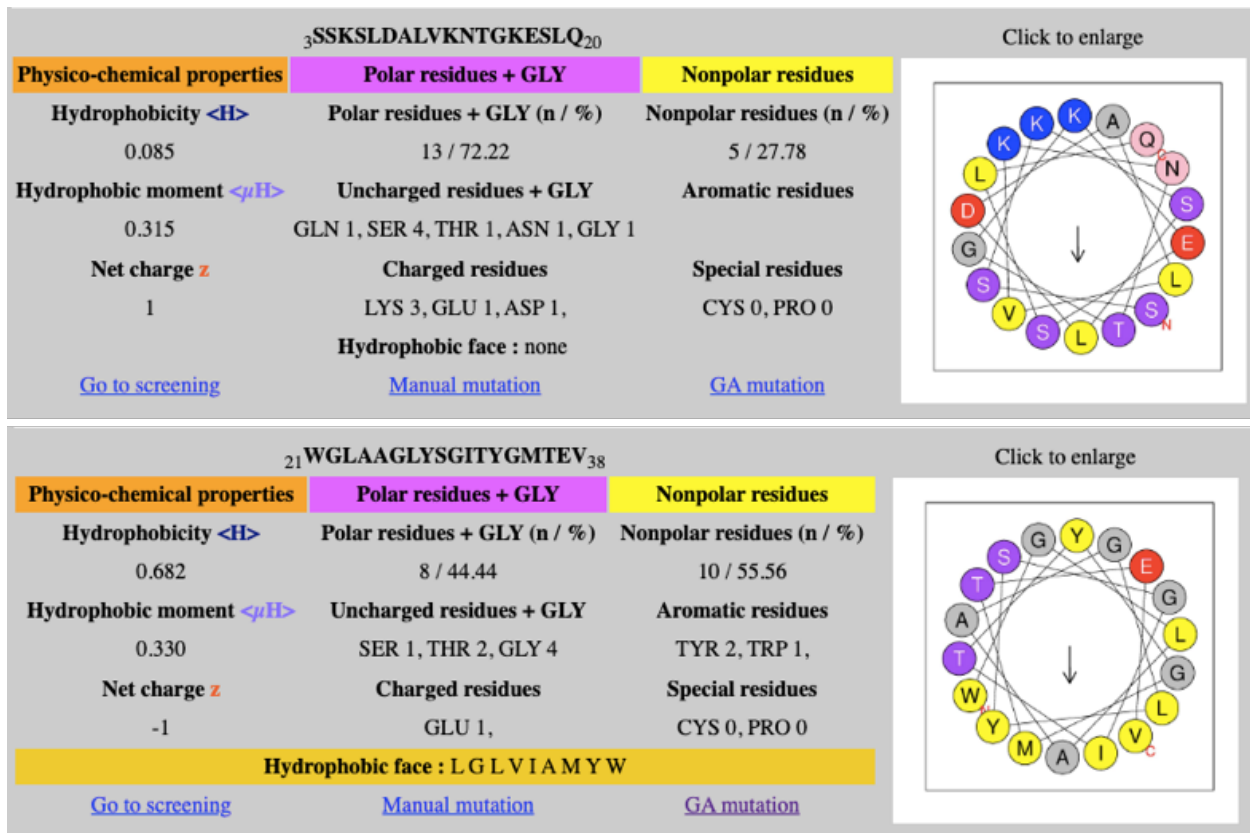


Figure 6.4. Heli-Quest Predictions of the H2-Domain Protein Sequence. The 40 predicted residues of the predicted H2-domain were analyzed using Heli-Quest. The helix generated by residues 3-20 contained many polar and charged residues and did not contain a hydrophobic face. The helix generated by residues 21-38 contained many hydrophobic residues and a hydrophobic face which included residues LGLVIAMYW.

	NT Construct	CT Construct
Experiment 1	GESSKSLDALVKNTGKESLQ	WGLAAGLYSGITYGMTEVRG
Experiment 2	1 AAAAAAAAAAVKNTGKESLQWGLAAGLYSGITYGMTEVRG	
	2 GESSKAAAAAAAAAAKESLQWGLAAGLYSGITYGMTEVRG	
	3 GESSKSLDALAAAAAAAAAAWGLAAGLYSGITYGMTEVRG	
	4 GESSKSLDALVKNTGAAAAAAAAAAGLYSGITYGMTEVRG	
	5 GESSKSLDALVKNTGKESLQAAAAAAAAAAITYGMTEVRG	
	6 GESSKSLDALVKNTGKESLQWGLAAAAAAAAAAATEVRG	
	7 GESSKSLDALVKNTGKESLQWGLAAGLYSGAAAAAAAAA	

Figure 6.5. Construct Design for Future Experiments. Constructs for experiment one will include either the first or second half of the predicted H2 domain, named the NT construct or the CT construct. Constructs in experiment two will include the H2-domain sequence wherein ten alanine substitutions have been made and two constructs overlap by five residue positions.

6.4. A Potential Pathway for OEP16-2 OEM-Translocation

The H2 domain sequence was examined for similarities to known plastid-targeting mechanisms. Similarities in residue composition were found between the CT ARD of AKR2A and the H2 domain of OEP16-2. The CT ARD of AKR2A forms a lipid-head binding pocket that allows AKR2A to associate with the OEM. The OEP16-2 protein sequence and the structural model of AKR2A were input into SWISS-MODEL. AKR2A is a soluble chaperone protein while OEP16-2 is a transmembrane pore. Thus, the accuracy of the model was expected to be low and was only generated to assess the ability of the H2 domain to form a lipid-head binding pocket. SWISS-MODEL aligned the OEP16-2 H2 domain and portions of the H3 domain to the

CT ARD of AKR2A with some confidence (Figure 6.6). However, the remainder of the OEP16-2 protein sequence did not align with AKR2A and therefore the overall confidence in this model is low, as expected. Thus, the only portion of this model worth considering is structure of the H2 and H3 domains. The AKR2A L1 and L2 binding pockets aligned well with the H2 and H3 domains, suggesting these domains could form a lipid-binding pocket that facilitates OEP16-2 targeting to the OEM. The OEP16-2 surface model generated by SWISS-MODEL contains two pockets within the H2 and H3 regions (Figure 6.7). If these pockets are present within the native structure of OEP16-2 they could potentially function as lipid-binding pockets which facilitate OEM localization. Future analysis can examine this potential interaction using lipid-protein binding assays. The protein-lipid binding activity of OEP16-2 could be assessed by treating chloroplasts with trypsin. Trypsin degrades proteins in the chloroplast outer envelope membrane and some proteins in the intermembrane space (Kim et al., 2014). If OEP16-2 localization requires OEP receptors, then OEP16-2 will not be imported into trypsin-treated chloroplasts. On the other hand, if OEP16-2 translocation is dependent on a protein-lipid interaction, then OEP16-2 will be imported and detected within trypsin-treated chloroplasts (Kim et al., 2014). This methodology is useful because it yields informative results even if the protein-lipid binding hypothesis is incorrect. If protein-lipid binding activity is observed, a second experiment to test the binding activity of OEP16-2 to specific lipid-types can be performed. The chemical duramycin causes PE- and MGDG-lipids to clump, disrupting the OEM lipid-organization. Treating chloroplasts with duramycin can disrupt protein interactions with PE- and MGDG-lipids. Thus, if OEP16-2 is not translocated into duramycin treated chloroplasts, then OEP16-2 likely binds to PE- and/or MGDG-lipids during translocation (Kim et al., 2014).

This lipid-protein interaction would then need to be validated using a quantitative lipid-protein assay, such as a protein lipid overlay (PLO) assay (Dowler et al., 2002). However, treatment of chloroplasts with trypsin may show that OEP16-2 relies on protein factors for import into the OEM. To investigate this, OEP16-2 can be transiently expressed in *ppi* mutant protoplasts. Mutants *ppi1* and *ppi2* are *Toc33* and *Toc159* knock-out mutants, respectively (Sjuts et al., 2017). Thus, if OEP16-2 is not imported into *ppi1* and/or *ppi2* protoplasts, it is likely that OEP16-2 translocation is dependent on components of the TOC complex and general import pathway.

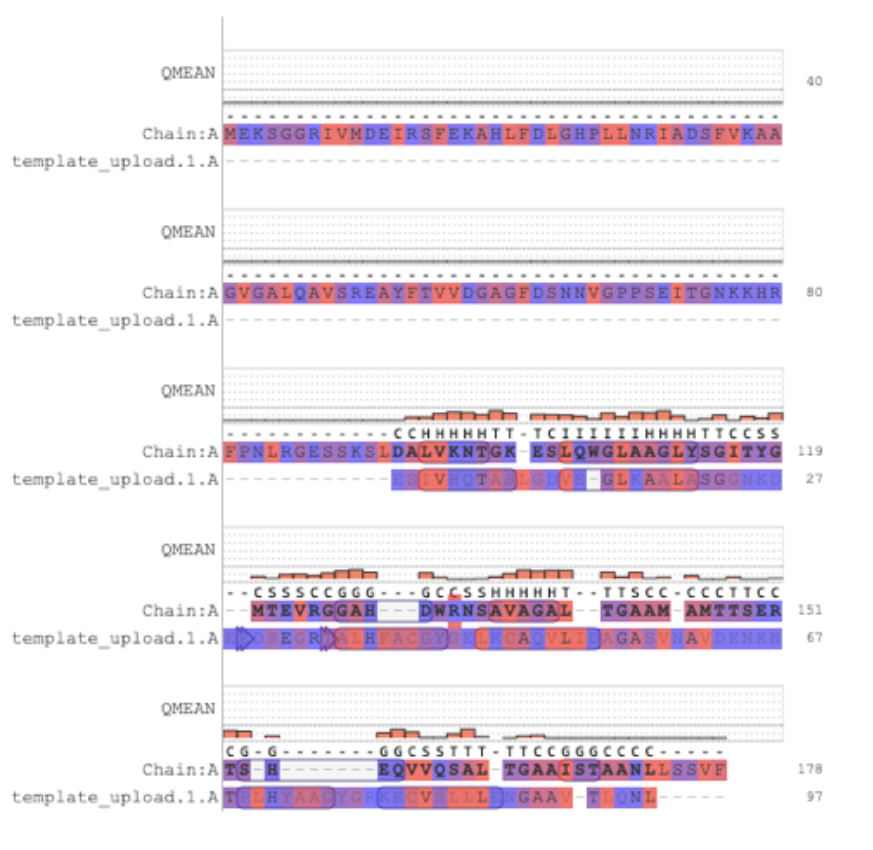


Figure 6.6. SWISS-MODEL Protein Sequence Alignment of AKR2A and OEP16-2. The SWISS-MODEL program identified homology between portions of the OEP16-2 sequence and the CT ARD domains of AKR2A. A QMEAN confidence score was generated for each aligned pair of residues, represented by a bar over each position, taller bars indicated higher confidence.

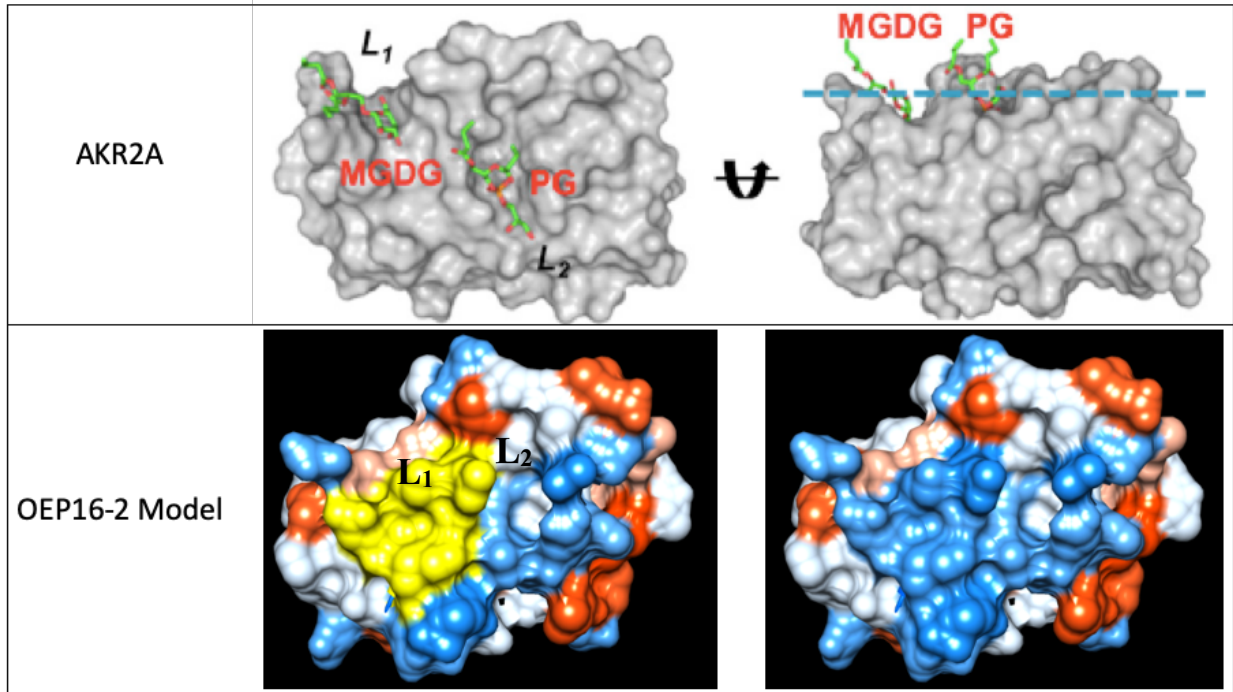


Figure 6.7. Structural Model of OEP16-2 and AKR2A. A surface model of AKR2A compared to the surface model of OEP16-2 which was generated by SWISS-MODEL using homology to AKR2A. The MGDG lipid-head binding pocket L₁ and PG lipid-head binding pocket L₂ in AKR2A are shown. The OEP16-2 model contains two similar pockets denoted by L₁ and L₂ above the respective pocket.

Figure 6.7. Structural modelling of OEP16-2 and comparison to AKR2A. Adapted from An Ankyrin Repeat Domain of AKR2 Drives Chloroplast Targeting through Coincident Binding of Two Chloroplast Lipids by 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. Retrieved from Cell and Developmental Biology. Copyright 2014 by Elsevier Inc.

7.0 Identifying OEP Candidates for Future Localization Assays

7.1. Overview

The internal plastid-targeting signal used by OEP16-2 could also be used by other multi-pass alpha helical OEPs. Identifying other multi-pass proteins and performing targeting assays with them could uncover targeting features similar to the OEP16-2 signal and expand our knowledge of how these signals function. Protein candidates for future targeting assays were identified by structurally categorizing known OEPs and computationally analyzing multi-pass proteins. First, a master list of OEPs was generated and structurally classified using the pipeline outlined in section 2.9 (Figure 7.1). Then, identified multi-pass proteins were further categorized by function and individual proteins were assessed to identify candidates appropriate for future targeting assays.

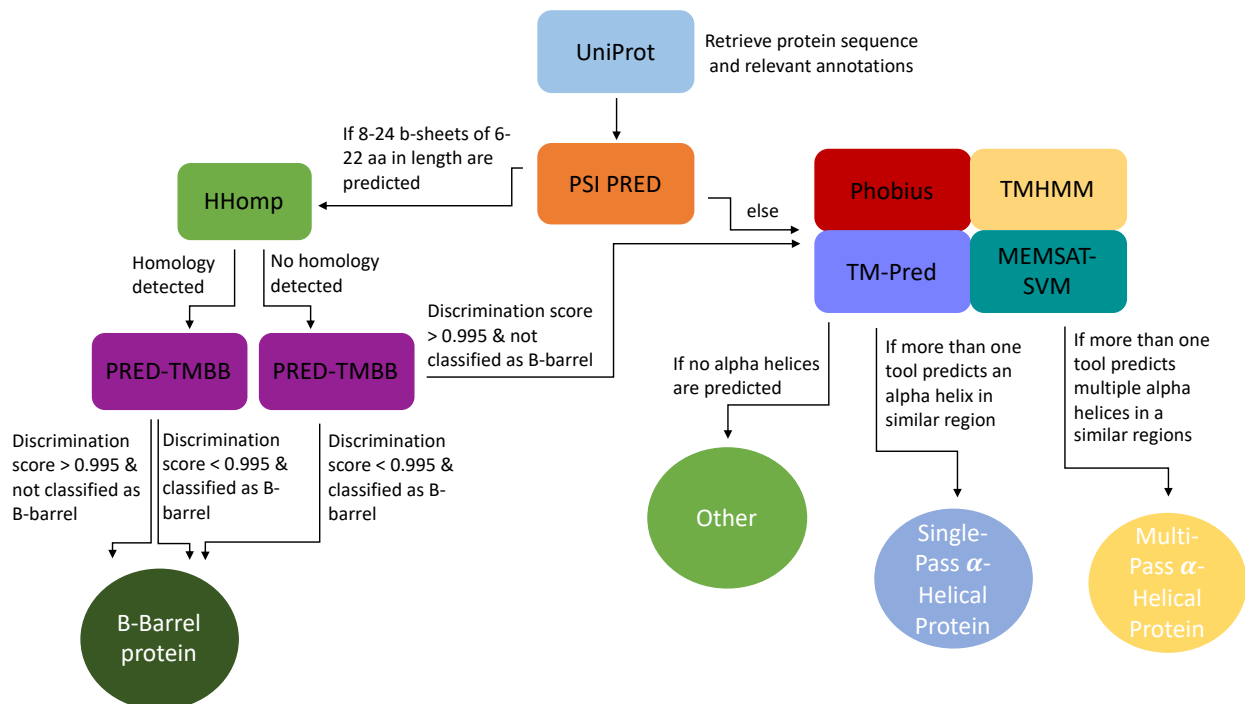


Figure 7.1. OEP Structural Categorization Pipeline. The pipeline designed to computationally assess and categorize predicted outer envelope proteins. The amino acid sequence of OEPs

were retrieved from the UniProt database. Each sequence was analyzed using PSI-PRED, if 8-24 b-sheets of 6-22 aa were predicted the protein was further analyzed with HHomp and PRED-TMBB. If HHomp detects homology to B-barrel proteins and/or if PRED-TMBB produced a discrimination score of < 0.995 and predicted a B-barrel protein, the OEP was classified as a B-barrel protein. Otherwise, the sequence was analyzed using Phobius, TMHMM, TM-Pred, and MEMSAT-SVM. If multiple tools predicted helices in overlapping locations the multi-pass alpha helical protein designation was assigned. If multiple tools predicted one helix in an overlapping location the single-pass alpha helical protein designation was assigned. Proteins unsuitable for any category were assigned the designation of other.

7.2. Classification of OEPs by Structural Class

The known OEP targeting mechanisms, including β -barrel insertion, SA-/TA-mediated insertion, and CT TP-like targeting are not suitable methods for the translocation of multi-pass alpha helical transmembrane proteins due to their inherent structure. Multi-pass alpha helical proteins must insert several helices across the membrane as opposed to inserting a single helix or a β -barrel. Therefore, they likely require a different translocation mechanism than the ones previously described. Moreover, we experimentally verified that the multi-pass protein OEP16-2 does not utilize a CT TP-like signal. Thus, multi-pass proteins likely use a different targeting mechanism than those previously described. This idea is supported by the identification of the OEP16-2 internal targeting signal, which is quite different than previously identified plastid-targeting signals. To further characterize this targeting mechanism the targeting signals of other multi-pass alpha-helical proteins must be identified for comparison.

A list of all known OEPs was generated through literature and database searches. A list of 137 proteins was compiled using the Plant Protein Database (PPDB), the AT_CHLORO

database, and lists generated by Inoue (2015) & Bouchnak et al., (2019). OEPs were structurally categorized as either a β -barrel protein, single-pass protein, multi-pass alpha helical protein, or other (Table 7.2). Of the 137 proteins examined, 27% of them were identified as multi-pass alpha helical proteins because several alpha helices were predicted within the amino acid sequence (Figure 7.2; Figure 7.1). This suggests there are many other OEPs that could share a common OEM-translocation pathway with OEP16-2.

Table 7.1 Compilation of OEPs and their Predicted Structural Classification.

Gene Accession	Protein Accession	Protein Name	Reference	Structural Classification	Structural Classification Support
At1g02280	O23680	Translocase of chloroplast 33	(i, ii, iii & iv)	single-pass	a, d, & g
At1g07930	Q8W4H7	Elongation factor 1-alpha 2	(i)	other	n/a
At1g09340	Q9SA52	Chloroplast stem-loop binding protein of 41 kDa (CSP41B)	(i)	single-pass	d, e, f, & g
At1g09920	Q8L7A5	Expressed Protein	(i)	single-pass	a, d, e, f, & g
At1g12230	F4IC59	Aldolase superfamily protein	(i)	multi-pass	e, f, & g
At1g13900	Q9LMG7	Probable inactive purple acid phosphatase 2	(i & ii)	single-pass	a, d, e, & g
At1g16000	P93048	GAG1At protein	(i)	single-pass	a, d, e, f, & g
At1g20816	Q6ID99	Outer envelope pore protein 21A	(i, iii & iv)	B-barrel	a & b
At1g26340	Q9FDW8	Cytochrome b5 isoform A	(iii)	single-pass	a, d, e, & g
At1g27300	Q9FZK5	F17L21.9	(i)	single-pass	a, d, e, f, & g
At1g27390	P82873	Mitochondrial import receptor subunit TOM20-2	(i)	single-pass	a, d, e & g
At1g34430	Q9C8P0	EMB3003	(i)	multi-pass	d, e, f
At1g44170	Q70DU8	Aldehyde dehydrogenase family 3 member H1	(i)	multi-pass	d, e, & f
At1g45170	Q1H5C9	Outer envelope pore protein 24A	(i, ii & iii)	B-barrel	a, b & c
At1g54150	Q9SYH3	E3 ubiquitin-protein ligase SPL2	(ii)	multi-pass	a, d, e, f & g
At1g59560	Q94HV7	E3 ubiquitin-protein ligase SPL1	(ii)	multi-pass	a, e, f & g
At1g63900	Q8L7N4	E3 ubiquitin-protein ligase SP1	(i, ii & iii)	multi-pass	a, e, f & g
At1g64850	Q9XIR0	At1g64850	(i)	multi-pass	a, d, e, f, & g
At1g67690	F4HTQ1	Probable thimet oligopeptidase	(i)	single-pass	d, e & g
At1g68680	Q8L9R6	At1g68680	(i)	multi-pass	a, d, e, f & g
At1g70480	F4I5G3	OBP32pep	(i & iii)	single-pass	d, e, f & g
At1g76405	Q9FPG2	OEP21B	(i, ii, iii & iv)	B-barrel	a & b

At1g77590	Q9CAP8	Long chain acyl-CoA synthetase 9	(i, ii & iii)	multi-pass	d, e, f & g
At1g80890	Q9SAH1	At1g80890	(i)	single-pass	a, d, e, f, & g
At2g01320	Q9ZU35	ABC transporter G family member 7	(i & iii)	multi-pass	a, d, e, f, & g
At2g06010	Q8VY85	OBP3-responsive protein 4	(i)	other	-
At2g11810	Q9SI93	Monogalactosyldiacylglycerol synthase 3	(i & ii)	multi-pass	d, e, f, & g
At2g16070	Q9XII1	Plastid division protein PDV2	(i, ii, iii & iv)	single-pass	a, d, e, f, & g
At2g16640	Q9SLF3	Translocase of chloroplast 132	(i, ii, iii & iv)	single-pass	a, e, & g
At2g17390	Q29Q26	Ankyrin repeat domain-containing protein 2B	(i)	other	a
At2g17695	Q8GXB1	UPF0548 protein At2g17695	(i & iii)	single-pass	d, e, & g
At2g19860	P93834	Hexokinase-2	(i)	single-pass	a, e, & g
At2g20890	Q9SKT0	Thylakoid formation 1	(i)	single-pass	a, f, & g
At2g43950	O80565	Outer envelope pore protein 37	(i, ii, iii & iv)	B-barrel	a, b & c
At2g24440	Q9ZQ24	Expressed Protein	(i)	other	-
At2g25660	F4ISL7	TIC236	(iii)	single-pass	a, d, & e
At2g27490	Q9ZQH0	Dephospho-CoA kinase	(i & iii)	single-pass	d, f, & g
At2g28900	Q9ZV24	OEP16-1	(i, ii & iii)	multi-pass	a, d, e, & f
At2g32240	F4ISU2	Early endosome antigen	(i)	single-pass	d, e, f, & g
At2g32290	Q8L762	Beta-amylase 6	(iii)	other	-
At2g32650	O48852	At2g32650	(i)	other	-
At2g34585	Q8S8R9	At2g34585	(iii)	single-pass	d, e, f, & g
At2g34590	O64688	Pyruvate dehydrogenase E1 component subunit beta-3	(i)	multi-pass	d, e, & f
At2g38670	Q9ZVI9	Ethanolamine-phosphate cytidyltransferase	(i)	single-pass	a, d, e, & g
At2g40690	Q949Q0	Glycerol-3-phosphate dehydrogenase 2	(iii)	multi-pass	f
At2g44640	O80503	Expressed Protein	(i & iii)	multi-pass	d, e, & f
At2g47770	O82245	Translocator protein homolog	(i)	multi-pass	a, d, e, f, & g
At3g01280	Q9SRH5	Mitochondrial outer membrane protein porin 1	(i)	single-pass	a & g
At3g01500	P27140	Beta carbonic anhydrase 1	(i)	other	a
At3g03870	Q9SRW4	F20H23.8 protein	(iii)	single-pass	a, d, e, f, & g
At3g06510	Q93Y07	Galactolipid galactosyltransferase SFR2	(i, ii, iii & iv)	single-pass	a
At3g06960	Q9M903	Trigalactosyldiacylglycerol 4	(i, ii, iii & iv)	single-pass	a, f & g
At3g07430	Q9SRS3	YlmG homolog protein 1-1	(iii)	multi-pass	a, d, e, f, & g
At3g11670	Q9S7D1	Digalactosyldiacylglycerol synthase 1	(i & ii)	other	-
At3g12580	Q9LHA8	Heat shock protein 70 kDa protein 4	(i)	other	a
At3g16620	Q9LUS2	Translocase of chloroplast 120	(i & iii)	single-pass	a, f, & g
At3g16950	A8MS68	Dihydroliopoyl dehydrogenase 1	(i)	other	a
At3g17970	Q9LVH5	Outer envelope protein 64	(i, ii, iii & iv)	multi-pass	a, d, e, & f
At3g19720	Q84N64	Dynammin-like protein ARC5	(i, ii & iii)	other	a
At3g21865	Q9LSX7	Peroxisome biogenesis protein 22	(i)	single-pass	a, e, f, & g

At3g25690	Q9LI74	CHUP1	(i, ii & iii)	other	a
At3g25860	Q9SQI8	Dihydrolipoyllysine-residue acetyltransferase	(i)	other	a
At3g26070	Q9LU85	Probable plasmid-lipid-associated protein	(i)	other	a
At3g26740	Q96500	Light-regulated protein 1	(i)	other	a
At3g27820	Q9LK94	Monodehydroascorbate reductase 4	(i)	multi-pass	a, d, e, & f
At3g44160	Q5PP51	Outer envelope protein 39	(i & ii)	B-barrel	a & b
At3g46030	Q9LZT0	Histone H2B.7	(i)	other	a
At3g46740	Q9STE8	TOC75-3	(i, ii, iii & iv)	B-barrel	a & b
At3g46780	Q9STF2	Plastid Transcriptionally active 16	(i)	other	a
At3g48620	F4JF35	Outer envelope protein 36	(i & ii)	B-barrel	a & b
At3g49350	Q4V3B4	At3g49350	(i)	other	-
At3g49560	Q9SCK3	Chloroplastic import inner membrane translocase subunit HP30-1	(iii & iv)	multi-pass	a, d, e, f, & g
At3g51870	O65023	Probable envelope ADP/ATP carrier	(i)	multi-pass	a, d, e, f, & g
At3g52230	Q9SUY2	AT3g52230/F4F15_340	(i, ii, iii & iv)	single-pass	d, e, f, & g
At3g52420	Q9SVC4	Outer envelope membrane protein 7	(i & iii)	single-pass	a, d, e, f, & g
At3g53560	Q8L606	Tetratricopeptide repeat (TPR)-like superfamily protein	(i)	single-pass	d, e, f, & g
At3g57090	Q9M1J1	Mitochondrial fission 1 protein A	(iii & iv)	single-pass	a, d, e, f, & g
At3g62880	Q9LZH8	Outer envelope pore protein 16-4	(i)	multi-pass	a, d, e, & f
At3g63150	F4J0W4	Mitochondrial Rho GTPase 2	(i)	single-pass	a, d, e, & g
At3g63160	Q9M1X3	AT3g63160/F16M2_10	(i, ii & iii)	single-pass	a, d, e, f, & g
At3g63170	Q9M1X2	Fatty-acid-binding protein 1	(i & iii)	other	a
At3g63520	O65572	Carotenoid 9,10(9',10')-cleavage dioxygenase 1	(iii)	other	a
At4g00550	Q8Q1S1	Digalactosyldiacylglycerol synthase 2	(i & ii)	other	-
At4g02482	F4JHJ5	Translocase of chloroplast-like protein	(i)	single-pass	a, d, e, & g
At4g02510	O81283	Translocase of chloroplast 159	(i, ii, iii & iv)	single-pass	a, d, e, & f
At4g05050	P0CH33	Polyubiquitin 11	(i)	other	a
At4g09080	Q5IZC8	TOC75-4	(i & ii)	B-barrel	a, b, & c
At4g12470	Q9SU35	pEARL11-like lipid transfer protein 1	(ii)	single-pass	f & g
At4g13550	F4JT30	Putative triglyceride lipase	(iii)	single-pass	d, f, & g
At4g14430	O23299	Enoyl-CoA delta isomerase 2	(i)	single-pass	f & g
At4g15440	B3LF83	Probable inactive linoleate hydroperoxide lyase	(i, ii & iii)	multi-pass	d, e, f, & g
At4g15810	F4JKW7	P-loop containing nucleoside triphosphate hydrolases superfamily protein	(i)	multi-pass	d, e, & f
At4g16160	Q0WMZ5	Outer envelope pore protein 16-2	(i)	multi-pass	a, d, e, & f
At4g16450	Q84W12	At4g16450	(i)	multi-pass	a, d, e, f, & g
At4g17170	P92963	Ras-related protein RABB1c	(i)	other	a
At4g26670	Q94EH2	Chloroplastic import inner membrane translocase subunit TIM22-2	(ii, iii, iv)	multi-pass	a, d, e, f, & g
At4g27680	Q9T090	26S proteasome regulatory particle chain RPT6-like protein	(i & iii)	single-pass	a, d, e, f, & g
At4g27990	Q9SUE0	YlmG homolog protein 1-2	(i & iii)	multi-pass	a, d, e, f, & g

At4g29130	Q42525	Hexokinase-1	(i, iii & iv)	single-pass	a, d, & e
At4g31780	O81770	Monogalactosyldiacylglycerol synthase 1	(i)	multi-pass	g
At4g32250	Q8RWX4	AT4g32250/F10M6_110	(i & iii)	single-pass	d, e, f, & g
At4g35000	Q42564	L-ascorbate peroxie 3	(i)	single-pass	a, d, e, f, & g
At4g36650	O23215	Plant-specific TFIIIB-related protein 1	(i)	other	a
At4g38920	P0DH93	V-type proton ATPase subunit c3	(i)	multi-pass	a, d, e, f, & g
At5g02500	P22953	Heat shock 70 kDa protein 1	(i)	other	a
At5g05000	Q38906	Translocase of Chloroplast 34	(i, ii, iii & iv)	single-pass	a, d, & g
At5g06290	Q9C5R8	2-Cys peroxiredoxin BAS1-like	(i)	multi-pass	d, e, & f
At5g11560	F4JXW9	Catalytics/EMC1_C domain containing protein	(i)	single-pass	a, d, & e
At5g13530	Q9FY48	E3 Ubiquitin-protein ligase KEG	(ii)	multi-pass	f & g
At5g15090	Q9SMX3	Mitochondrial outer membrane protein porin 3	(i)	B-barrel	a, b, & c
At5g16010	Q9LFS3	3-oxo-5-alpha-steroid 4-dehydrogenase family protein	(iii)	multi-pass	a, d, e, f, & g
At5g16870	Q1H5E3	At5g16870	(i)	single-pass	e, f, & g
At5g17770	Q9ZNT1	NADH--cytochrome b5 reductase 1	(i)	single-pass	a, d, e, & g
At5g19620	Q9C5J8	Outer envelope protein 80	(i, ii, iii & iv)	B-barrel	a, b & c
At5g20300	Q6S5G3	Translocase of chloroplast 90	(i & iii)	single-pass	a, f, & g
At5g20410	O82730	Monogalactosyldiacylglycerol synthase 2	(i & ii)	other	-
At5g20520	Q8RXP6	Alpha/beta hydrolase domain-containing protein WAV2	(i)	single-pass	a & g
At5g21920	Q9C595	YlmG homolog protein 2	(i)	multi-pass	a, d, e, f, & g
At5g21990	B7ZWR6	Outer envelope protein 61	(i & ii)	single-pass	a, d, e, & g
At5g23190	Q9FMY1	Cytochrome P450 86B1	(i)	single-pass	a & g
At5g24650	Q9FLT9	Chloroplastic import inner membrane translocase subunit HP30-2	(iii & iv)	multi-pass	a, d, e, f, & g
At5g02580	Q84TGO	Argininosuccinate lyase	(ii)	single-pass	e & g
At5g25900	Q93ZB2	Ent-kaurene oxidase	(i & ii)	single-pass	a,
At5g27330	F4K498	Prefoldin chaperone subunit family protein	(i)	single-pass	a, d, e, f, & g
At5g27540	Q8RXF8	Mitochondrial Rho GTPase 1	(i)	single-pass	a, d, e, & g
At5g35210	F4JYC8	DDT domain-containing protein PTM	(i & ii)	multi-pass	a, d, e, f, & g
At5g35360	O04983	Biotin carboxylase	(i)	single-pass	d, e, f, & g
At5g42070	Q8RWR9	Uncharacterized protein At5g42070	(i)	other	-
At5g42960	Q8HOY1	Outer envelope pore protein 24B	(i, ii, iii & iv)	B-barrel	a, b, & c
At5g43070	Q9FMH6	WPP domain-containing protein	(i)	other	a
At5g51020	Q9FI46	Chromophore lyase CRL	(i, ii, iii & iv)	single-pass	a, d, e, f, & g
At5g53280	Q9FK13	Plastid division protein PDV1	(i & ii)	single-pass	a, d, e, & g
At5g55510	Q6NKU9	Mitochondrial import inner membrane translocase subunit TIM22-3	(ii, iii, & iv)	multi-pass	a, d, e, f, & g
At5g56730	Q9FJT9	Zinc protease PQQL-like	(i)	single-pass	d, e, & g
At5g58140	P93025	Phototropin-2	(i)	other	a
At5g59840	Q9FJF1	Putative GTP-binding protein ara-3	(i)	single-pass	f & g
At5g64816	Q8L8Q8	Uncharacterized protein At5g64816	(i & iii)	single-pass	d, e, f, & g

A list of currently identified OEPs was compiled. The source(s) they are from are indicated by symbols, (i) Inoue (2015), (ii) PPDB, (iii) Bouchnak (2019), (iv) AT_CHLORO. The gene & protein accession number, the gene name, and the predicted structural category (B-barrel, single-pass, multi-pass, other) of each protein are listed. Database annotations and computational analysis that supported each structural classification are provided, (a) UniProt annotation, (b) HHomp, (c) PRED-TMBB, (d) Phobius, (e) TMHMM, (f) TM-Pred, and (g) MEMSAT-SMV.

Percentage of OEPs in Four Structural Classes

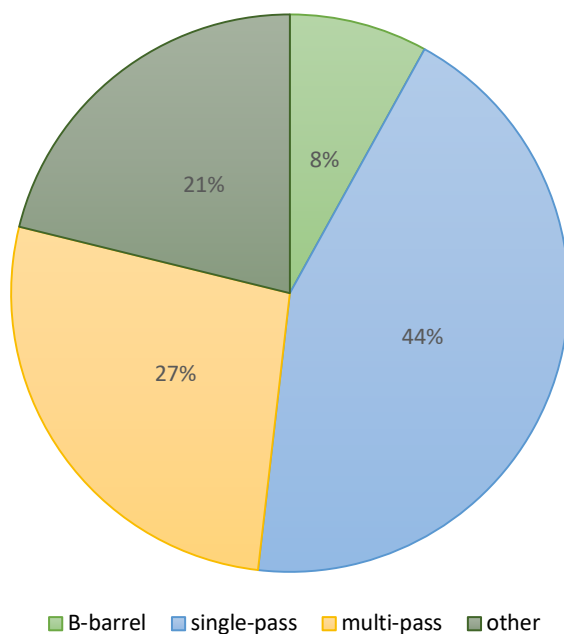


Figure 7.2. The Percentage of OEPs Grouped into Four Structural Classes. A list of 137 OEPs was categorized into four structural classes, including β -barrel proteins, single-pass alpha helical proteins, multi-pass alpha-helical proteins, and other proteins. Of these proteins, 44% were classified as single-pass alpha helical proteins, 8% were β -barrel proteins, 27% multi-pass alpha helical proteins, and other proteins.

7.3. Investigating Predicted Multi-Pass Alpha Helical Proteins for Potential Targeting-Features

The identified multi-pass proteins (Table 7.2) were examined for similarities to OEP16-2 in order to find candidates best suited for future targeting studies. Several of these proteins were members of the Tim17/22/23 superfamily and contain four transmembrane alpha helices. Thus, they share evolutionary and structural similarities with OEP16-2 and could also share similar targeting signal, making these proteins suitable candidates for further study. The identified Tim17/22/23 family members included, HP20 (hypothetical protein 20), HP22 (hypothetical protein 22), HP30 (hypothetical protein 30), & HP30-2 (hypothetical protein 30-2; Rossig et al., 2013). When these proteins were aligned with OEP16 isoforms, regions of sequence conservation were identified in the areas that aligned with the H1, H2, and H3 domains of OEP16-2. Therefore, these areas should be the starting point for future targeting studies. Additionally, the G X₃ G X₃ G X₃ G motif found in the CT half of the H2 domain of OEP16-2 is also found within the HP20, HP30, and HP30-2 sequence and could function as a targeting-signal (Figure 7.2). HP20, HP30, and HP30-2 function as a protein import site for IEM proteins (Rossig et al., 2013). Previous studies have shown that HP20 is located within the OEM while HP30 and HP30-2 are located within in IEM (Rossig et al., 2013). Therefore, HP20 is the best candidate for future OEM-targeting studies.

The 37 identified alpha helical multi-pass proteins were grouped by function (Table 7.2). Interestingly, proteins in the E3-ubiquitin family, which function in Toc-receptor turnover, were identified (Thomson et al., 2020). These proteins contain two alpha-helical transmembrane domains and may also be interesting candidates for targeting-signal studies. Other functional groups identified included solute and ion transporters, proteins involved in carbohydrate, lipid,

and other types of metabolism, and intracellular communication. Investigating any of these proteins may yield useful information about chloroplast OEP targeting mechanisms.

As protein analysis methodologies have advanced an increasing number of OEPs have been identified (Inoue, 2015). Moreover, these OEPs are involved in highly-specific and complex cellular processes. This suggests that the OEM is a dynamic and important regulatory structure and not a non-specific passive barrier as previous research has suggested (Day & Theg, 2018).

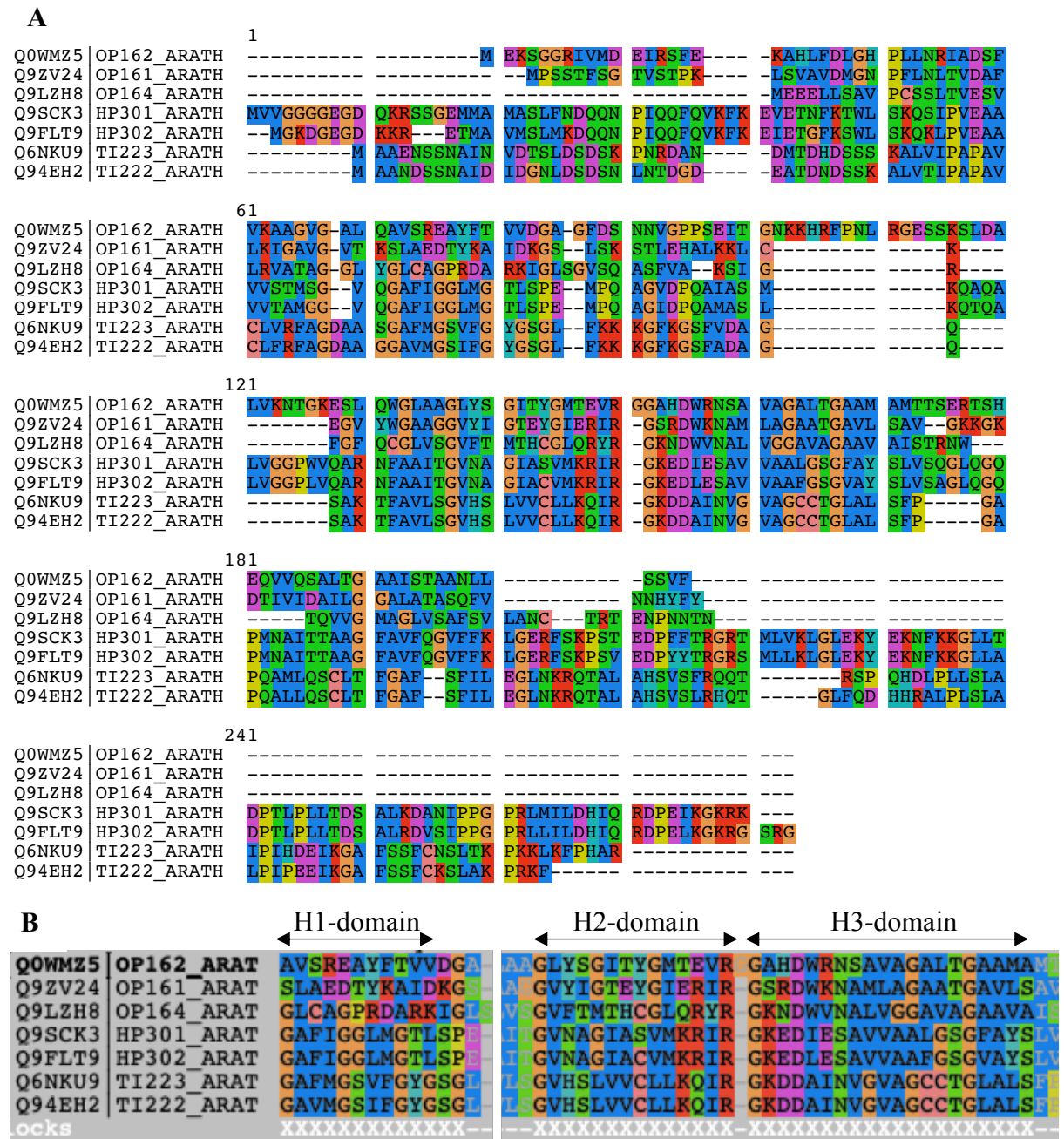


Figure 7.3. MSA of OEP16 isoforms with HP20, HP30, HP30-2, and TIM22-3. The full MSA alignment of OEP16-1, OEP16-2, OEP16-4, HP30, HP30-2, HP20, and TIM22-3 generated using the MUSCLE algorithm (A). Regions of conservation denoted by highly stringent G-blocks (X) align with portions of the H1-domain, H2 domain, and H3 domain (B).

Table 7.2. Predicted Multi-pass Alpha Helical Proteins Sorted in Functional Groups.

Function	Gene Accession	Protein Accession	Protein Name	Reference	Structural Classification	Structural Classification Support
Protein Turnover and modification	At1g63900	Q8L7N4	E3 ubiquitin-protein ligase SP1	(i, ii & iii)	Multi-pass	a, e, f & g
	At1g54150	Q9SYH3	E3 ubiquitin-protein ligase SPL2	(ii)	Multi-pass	a, d, e, f & g
	At1g59560	Q94HV7	E3 ubiquitin-protein ligase SPL1	(ii)	Multi-pass	a, e, f & g
	At5g13530	Q9FY48	E3 Ubiquitin-protein ligase KEG	(ii)	Multi-pass	f & g
Solute/Ion transporters	At2g28900	Q9ZV24	OEP16-1	(i, ii & iii)	Multi-pass	a, d, e, & f
	At3g62880	Q9LZH8	Outer envelope pore protein 16-4	(i)	Multi-pass	a, d, e, & f
	At4g16160	Q0WMZ5	Outer envelope pore protein 16-2	(i)	Multi-pass	a, d, e, & f
	At2g01320	Q9ZU35	ABC transporter G family member 7	(i & iii)	Multi-pass	a, d, e, f, & g
	At3g51870	O65023	Probable envelope ADP/ATP carrier	(i)	Multi-pass	a, d, e, f, & g
	At4g38920	P0DH93	V-type proton ATPase subunit c3	(i)	Multi-pass	a, d, e, f, & g
Protein Import Components	At4g26670	Q94EH2	Chloroplastic import inner membrane translocase subunit TIM22-2 (HP20)	(ii, iii, iv)	Multi-pass	a, d, e, f, & g
	At5g55510	Q6NKU9	Mitochondrial import inner membrane translocase subunit TIM22-3	(ii, iii, & iv)	Multi-pass	a, d, e, f, & g
	At3g17970	Q9LVH5	Outer envelope protein 64	(i, ii, iii & iv)	Multi-pass	a, d, e, & f
	At5g24650	Q9FLT9	Chloroplastic import inner membrane translocase subunit HP30-2	(iii & iv)	Multi-pass	a, d, e, f, & g
	At3g49560	Q9SCK3	Chloroplastic import inner membrane translocase subunit HP30-1	(iii & iv)	Multi-pass	a, d, e, f, & g
Intracellular Communication	At5g35210	F4JYC8	DDT domain-containing protein PTM	(i & ii)	Multi-pass	a, d, e, f, & g
Carbohydrate Metabolism	At1g12230	F4IC59	Aldolase superfamily protein	(i)	Multi-pass	e, f, & g
Lipid Metabolism	At1g77590	Q9CAP8	Long chain acyl-CoA synthetase 9	(i, ii & iii)	Multi-pass	d, e, f & g
	At4g15440	B3LF83	Probable inactive linoleate hydroperoxide lyase	(i, ii & iii)	Multi-pass	d, e, f, & g
	At2g11810	Q9SI93	Monogalactosyldiacylglycerol synthase 3	(i & ii)	Multi-pass	d, e, f, & g
	At5g16010	Q9LFS3	3-oxo-5-alpha-steroid 4-dehydrogenase family protein	(iii)	Multi-pass	a, d, e, f, & g
	At4g31780	O81770	Monogalactosyldiacylglycerol synthase 1	(i)	Multi-pass	g
	At2g40690	Q949Q0	Glycerol-3-phosphate dehydrogenase 2	(iii)	Multi-pass	f
Other Metabolism	At2g47770	O82245	Translocator protein homolog	(i)	Multi-pass	a, d, e, f, & g
	At3g27820	Q9LK94	Monodehydroascorbate reductase 4	(i)	Multi-pass	a, d, e, & f
	At1g34430	Q9C8P0	EMB3003	(i)	Multi-pass	d, e, f

	At1g44170	Q70DU8	Aldehyde dehydrogenase family 3 member H1 Pyruvate dehydrogenase E1 component subunit beta-3	(i)	Multi-pass	d, e, & f
	At2g34590	O64688		(i)	Multi-pass	d, e, & f
	At5g06290	Q9C5R8	2-Cys peroxiredoxin BAS1-like	(i)	Multi-pass	d, e, & f
Unknown	At4g27990	Q9SUE0	YlmG homolog protein 1-2	(i & iii)	Multi-pass	a, d, e, f, & g
	At3g07430	Q9SRS3	YlmG homolog protein 1-1	(iii)	Multi-pass	a, d, e, f, & g
	At2g44640	O80503	Expressed Protein	(i & iii)	Multi-pass	d, e, & f
	At1g64850	Q9XIR0	At1g64850	(i)	Multi-pass	a, d, e, f, & g
	At4g16450	Q84W12	At4g16450	(i)	Multi-pass	a, d, e, f, & g
	At5g21920	Q9C595	YlmG homolog protein 2	(i)	Multi-pass	a, d, e, f, & g
	At1g68680	Q8L9R6	At1g68680 P-loop containing nucleoside triphosphate hydrolases superfamily protein	(i)	Multi-pass	a, d, e, f & g
	At4g15810	F4JKW7		(i)	Multi-pass	d, e, & f

The OEPs predicted to be multi-pass alpha-helical proteins and the source(s) they are from, (i) Inoue (2015), (ii) PPDB, (iii) Bouchnak (2019), (iv) AT_CHLORO. The gene and protein accession numbers, the gene name, and the predicted structural class of each protein are listed. The database annotations and computational analysis that supported each structural classification are listed, (a) UniProt annotation, (b) HHomp, (c) PRED-TMBB, (d) Phobius, (e) TMHMM, (f) TM-Pred, and (g) MEMSAT-SMV. The 37 predicted multi-pass alpha helical proteins were grouped by function. These functional groups included, protein turnover and modification, solute and ion transporters, intracellular communication, protein import components, carbohydrate metabolism, lipid metabolism, other metabolism, and unknown.

8.0 Concluding Remarks

I originally hypothesized that OEP16-2 uses a CT TP-like signal to target the plastid outer envelope membrane. Epifluorescent analysis of onion cells transfected with various OEP16-2:EGFP fusion constructs revealed the C-terminal 33 amino acids are not a sufficient targeting signal. Further analysis showed the S, H2, and H3 domains in OEP16-2 contain a sufficient plastid targeting signal.

A number of possible features within these domains could function as a plastid targeting signal. The OEP16-2 H2 domain was computationally analyzed for features, such as residue biases and lipid-interacting pockets, that could function as a plastid targeting signal. Experiments were recommended to further investigate the targeting function of these features.

OEPs can be categorized into groups, each group is defined by a distinct set of characteristics and proteins within each group utilize a similar OEM localization pathway. These groups include B-barrel proteins, signal-anchored proteins, tail-anchored proteins, and CT TP-like proteins. The targeting domains and structure of OEP16-2 does not meet the criteria of any currently defined OEP groups. Therefore, OEP16-2 likely utilizes a different OEM targeting pathway.

A global OEP analysis was performed to classify proteins by their structure and localization pathway. A list of 137 known and predicted OEPs were compiled from various sources and categorized by targeting strategy. The categories included, B-barrel proteins, single-pass proteins which encompassed SA, TA, and CT TP-like proteins, multi-pass alpha helical protein, or other. A total of 27 multi-pass alpha helical OEPs were identified and categorized by function. Of the identified multi-pass alpha helical OEPs, 7 belonged to the PRAT

(preprotein amino acid transporter) protein family. Conservation was identified in the amino acid sequences of these 7 PRAT proteins. These areas of conservation aligned with the H2 and H3 domains of OEP16-2. This suggested the targeting signal within the H2 and H3 domains may be conserved throughout OEM targeting PRAT proteins. Lastly, the PRAT family protein HP20 was recommended as a candidate for future targeting assays because its localization to the outer envelope has been experimentally verified and it functions as an IEM protein transporter.

9.0 References

- Amborella Genome Project (2013). The Amborella Genome and the Evolution of Flowering Plants. *Science*, 324: 1467-1477.
- Anderson, S. A., Singhal, R. & Fernandez, D. E. (2019). Membrane-Specific Targeting of Tail-Anchored Proteins SECE1 and SECE2 Within Chloroplasts. *Front. Plant Sci.* 10(1401).
- Bagos, P. G. & Liakopoulos, T. D. (2004). PRED-TMBB: a web server for predicting the topology of beta-barrel outer membrane proteins. *Nucleic Acid Res.* 32: W400-404.
- Bischof, S., Barerenfaller, K., Wildhaber, T., Troesch, R., Vidhi, P. A., Roschitzki, B., Hirsch-Hoffmann, M., Hennig, L., Kessler, F., Gruissem, W. & Baginsky, S. (2011). Plastid Proteome Assembly without Toc159: Photosynthesis Protein Import and Accumulation of *N*-Acetylated Plastid Precursor Proteins. *Plant Cell*, 23: 3911-3928.
- Bölter, B. (2018). En route into chloroplasts: preproteins' way home. *Photosynthesis Research*, 138: 263-275.
- Bouchnak, I., Brugiére, 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.
- Chang, J. S., Chen, L. J., Yeh, Y. H., Hsiao, C. D. & Li, H. M. (2017). Chloroplast Preproteins Bind to the Dimer Interface of the Toc159 Receptor during Import. *Plant Phys.* 173: 2148-2162.
- 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.
- Chu, C. C. & Li, H. M. (2018). Developmental regulation of protein import into plastids. *Photosynthesis Research*, 138: 327-334.
- 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.

- Day, M. P. & Theg, S. M. (2018). Evolution of protein transport to the chloroplast envelope membranes. *Photosynthesis Research*, 138: 315-326.
- Dowler, S., Kular, G. & Alessi, D. R. (2002). Protein lipid overlay assay. *Sci. STKE*, 129: pl6.
- Drea, S. C., Lao, N. T., Wolfe, K. H. & Kavanaugh, T. A. (2006). Gene duplication, exon gain and neofunctionalization of OEP16-related genes in land plants. *The Plant Journal*, 46, 723-735.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. *Proteomics Protocols Handbook*, 52: 571-607.
- Gautier R., Douguet, D., Antony, B. & Drin, G. (2008). HELIQUEST: a web server to screen sequences with specific α -helical properties. *Bioinformatics*, 24(18): 2101-2102.
- 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.
- Hofmann, K. & Stoffel, W. (1993). TMbase – A database of membrane spanning proteins segments. *Biol Chem.*, 374: 166.
- Inoue, K. (2015). Emerging knowledge of the organelle outer membranes – research snapshots and an updated list of the chloroplast outer envelope proteins. *Front. Plant Sci.* 6:278.
- Jarvis, P. & López-Juez, E. (2013). Biogenesis and homeostasis of chloroplasts and other plastids. *Nat. Reviews Mol Cell Bio*, 14: 787-802.
- Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292: 195-202.
- Jones, T. & Rapaport, D. (2017). Early stages in the biogenesis of eukaryotic B-Barrel proteins. *FEBS Letters*, 591: 2671-2681.
- 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., Jung, C. & Hwang, I. (2013). Cytosolic events involved in chloroplast protein targeting. *Biochimica et Biophysica Acta*, 1833: 245-252.

- 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, 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., Lee, J. & Hwang, I. (2017). Sorting of nuclear-encoded chloroplast membrane proteins. *Current Opinion in Plant Biology*, 40: 1-7.
- 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., Schnell, D. & Theg, S. M. (2020). Protein Import Motors in Chloroplasts: On the Role of Chaperones. *Plant Cell*, 32: 536-542.
- Li, H. M. & Teng, Y. S. (2013). Transit peptide design and plastid import regulation. *Trends in Plant Sci.* 18(7): 360-366.
- Ling, Q., Huang, W., Baldwin, A. & Jarvis, P. (2012). Chloroplast Biogenesis Is Regulated by Direct Action of the Ubiquitin-Proteasome System. *Science*, 338: 655-659.
- Linke, D., Frank, J., Pope, M. S., Soll, J., Ilkavets, I., Fromme, P., Burstein, E. A., Reshetnyak, Y. K. & Emelyanenko, V. I. (2004). Folding Kinetics and Structure of OEP16. *Biophys. Journal*, 86: 1479-1497.
- Lung, S. C. & Chuong, S. D. X. (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: 1560-1578.
- Lung, S. C., Smith, M. D., Weston, J. K., Gwynne, W., Secord, N. & Chuong, S. D. X. (2014). The C-terminus of *Bienertia sinuspersici* Toc159 contains essential elements for its targeting and anchorage to the chloroplast outer membrane. *Front. Plant Sci.* 5: e722.
- Käll L., Krogh, A. & Sonnhammer, E. L. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic Acids Res.*, 35: W429-432.
- 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

Drives Chloroplast Targeting through Coincident Binding of Two Chloroplast Lipids. *Dev. Cell*, 30: 598-609.

Kim, D.H., Xu, Z.Y., Na, Y.J., Yoo, Y.J., Lee, J., Sohn, E.J. & Hwang, I. (2011). Small heat shock protein Hsp17.8 functions as an AKR2A cofactor in the targeting of chloroplast outer membrane proteins in Arabidopsis, *Plant Physiol.*, 157, 132–146.

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

Nugent, T. & Jones, D. T. (2009). Transmembrane protein topology prediction using support vector machines. *BMC Bioinformatics*, 10(159).

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

Pitzschke, A. & Persak, H., (2012). Poinsettia protoplasts—A simple, robust, and efficient system for transient gene expression studies. *Plant Methods*, 8(1): 14.

Pohlmeyer, K., Soll, J., Steinkamp, T., Hinnah, S. & Wagner, R. (1997). Isolation and characterization of an amino acid-selective channel protein present in the chloroplastic outer envelope membrane. *PNAS*, 94(17): 9504-9509.

Pudelski, B., Kraus, S., Soll, J. & Philippar, K. (2010). The plant PRAT proteins – preprotein and amino acid transport in mitochondria and chloroplasts. *Plant Biology*, 12, 42-55.

Richardson, L. G., Paila, Y. D., Siman, S. R., Chen, Y., Smith, M. D. & Schnell, D. J. (2014). Targeting and assembly of components of the TOC protein import complex at the chloroplast outer envelope membrane. *Front. Plant Sci.* 5(269).

Rossig, C., Reinbothe, C., Gray, J., Valdes, O., von Wettstein, D. & Reinbothe, S. (2014). New functions of the chloroplast Preprotein and Amino acid Transporter (PRAT) family members in protein import. *Plant Signaling & Behaviour*, 9: e27693.

Samol, I., Rossig, C., Buhr, F., Springer, A., Pollmann, S., Lahroussi, A., von Wettstein, D., Reinbothe, C. & Reinbothe, S. (2011). The Outer Chloroplast Envelope Protein OEP16-1 for Plastid Import of NADPH:Protochlorophyllide Oxidoreductase A in *Arabidopsis thaliana*. *Plant Cell Physiol.*, 52(1): 96-111.

Sanford, J. C., Smith, F. D. & Russel, J. A. (1993). Optimizing the Biolistic Process for Different Biological Applications. *Methods in Enzymology*, 217: 483-509.

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.

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

Sonnhammer, E. L. L., von Heijne, G. & Krogh, A. S., (1998). A hidden Markov model for prediction transmembrane helices in protein sequences. *Proceeding of the Sixth International Conference on Intelligent Systems for Molecular Biology.* 175-180.

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.

Tiller, N., Weingartner, M., Thiele, W., Maximova, E., Schöttler, M. A. & Bock, R. (2012). The plastid-specific ribosomal proteins of *Arabidopsis thaliana* can be divided into non-essential proteins and genuine ribosomal proteins. *The Plant Journal*, 69: 302-316.

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.

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

Von Heijne, G., Steppuhn, J. & Herrmann, R. G. (1988). Domain structure of mitochondrial and chloroplast targeting peptides. *Eur. J. Biochem.* 180: 535-545.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, D., Tauriello, G., Gummienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R. & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46:W296-303.

Wu, F., Shen, S., Lee, L., Lee, S., Chan, M. & Lin C. (2009). Tape-Arabidopsis Sandwich – a simpler Arabidopsis protoplast isolation method. *Plant Methods*, 5(16).

Yang, X., Li, Y., Qi, M., Liu, Y. & Li, T. (2019). Targeted Control of Chloroplast Quality Improve Plant Acclimation: From Protein Import to Degradation. *Front. Plant Sci.* 10(958).

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

Zimmermann, L., Stephens, A., Nam, S. Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A. N. & Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol.* *S0022-2836(17)*: 30587-30589.

Zook, J.D., Molugu, T. R., Jacobsen, N. E., Lin, G., Soll, J., Cherry, B. R., Brown, M. F. & Fromme, P. (2013). High-Resolution NMR Reveals Secondary Structure and Folding of Amino Acid Transporter from Outer Chloroplast Membrane. *PLoS ONE*, *8(10)*, e78116.

10.0 Appendix

A1. The primary protein sequence of AtOEP16-2 (At4G16160) retrieved from NCBI.

MEKSGGRIVMDEIRSFKAHLFDLGHPLLNRADSFVKAAGVVGALQAVSREAYFTVVDGAGFDSNNVGGPSE
ITGNKKHRFPNLRGESSKSLDALVKNTGKESLQWGLAAGLYSGITYGMTEVRGGAHDWNRNSAVAGALTGA
AMAMTTSETSHEQVVQSALTGAAISTAANLLSSVF

A2. Protein Accession Numbers of OEP16-2 Isoforms from 29 plant species used to create the MSA in Figure 3.2.3.

Genus species	Protein Accession Number
<i>Arabidopsis thaliana</i>	NP_849394.1
<i>Capsella rubella</i>	XP_023635268.1
<i>Eutrema salsugineum</i>	XP_024005643.1
<i>Brassica napus</i>	XP_013699139.1
<i>Quercus lobata</i>	XP_030974967.1
<i>Ziziphus jujuba</i>	XP_015899996.1
<i>Prunus persica</i>	XP_007227528.2
<i>Carica papaya</i>	XP_021906866.1
<i>Arachis hypogaea</i>	XP_025630202.1
<i>Medicago truncatula</i>	XP_003609756.2
<i>Malus domestica</i>	XP_008390546.2
<i>Manihot esculenta</i>	XP_021594264.1
<i>Populus trichocarpa</i>	XP_002312339.1
<i>Abrus precatorius</i>	XP_027339297.1
<i>Ricinus communis</i>	XP_002519203.1
<i>Mucuna pruriens</i>	RDX90919.1
<i>Spatholobus suberectus</i>	TKY72522.1
<i>Citrus clementina</i>	XP_006419410.1
<i>Pistacia vera</i>	XP_031264535.1
<i>Glycine max</i>	XP_003533195.1
<i>Nymphaea colorata</i>	XP_031498698.1
<i>Ananas comosus</i>	XP_020107689.1
<i>Syzygium oleosum</i>	XP_030456230.1
<i>Cajanus cajan</i>	XP_029126119.1
<i>Jatropha curcas</i>	XP_012084054.1
<i>Solanum pennellii</i>	XP_015087098.1
<i>Vigna radiata</i>	XP_014499162.1

Dendrobium catenatum

XP_020697588.1

Momordica charantia

XP_022138821.1

A3. Sequences of Subcloned OEP16-2 EGFP Fusion Constructs. Start codons are underlined, stop codons are denoted by asterisks, EGFP (dark green), H1 domain (blue), S-domain (light green), H2 domain (yellow), H3 domain (purple), H4 domain (pink).

OEP16-2 Sequence

MEKSGGRIVMDEIRSFKAHLFDLGHPLLNRIADSFVKAAGVGALQAVSREAYFTVVDGAGFDSNNVGGPPSE
ITGNKKHRFPNLRGESSKSLDALVKNTGKESLQWGLAAGLYSXITYGMTEVRGGAHDWRNSAVAGALTGAA
MAMTTSERTSHEQVVQSALTGAAISTAANLL

EGFP:OEP16-2-FL

VTAAGITLGMDELYKSGLRSRGMEKSGGRIVMDEIRSFKAHLFDLGHPLLNRIADSFVKAAGVGALQAVSRE
AYFTVVDGAGFDSNNVGGPPSEITGNKKHRFPNLRGESSKSLDALVKNTGKESLQWGLAAGLYSXITYGMTEV
RGGAHDWNRNSAVAGALTGAAMAMTTSERTSHEQVVQSALTGAAISTAANLL

EGFP:OEP16-2Δ33CT

VTAAGITLGMDELYKSGLRSRGMEKSGGRIVMDEIRSFKAHLFDLGHPLXNRIADSFVKAAGVGALQAVSR
EAYFTVVDGAGFDSNNVGGPPSEITGNKKHRFPNXRGESSKSLDALVKNTGKESLQWGLAAGLYSGITYGMTE
VRGGAHDGGTAR*

EGFP:OEP16-2-33CT

VTAAGITLGMDELYKSGLXSRGMTTSERTSHEQVVQSALTGAAISTAANLLSSVF*GST

OEP16-2-FL:EGFP

MMEKSGGRIVMDEIRSFKAHLFDLGHPLLNRIADSFVKAAGVGALQAVSREAYFTVVDGAGFDSNNVGGPPSE
ITGNKKHRFPNLRGESSKSLDALVKNTGKESLQWGLAAGLYSGITYGMTEVRGGAHDWRNSAVAGALTGA
AMAMTTSERTSHEQVVQSALTGAAISTAANLLSSVFRILMVSKGEELFTGVVPILV

OEP16-2Δ33CT:EGFP

MMEKSGGRIVMDEIRSFKAHLFDLGHPLLNRIADSFVKAAGVGALQAVSREAYFTVVDGAGFDSNNVGGPPSE
ITGNKKHRFPNLRGESSKSLDALVKNTGKESLQWGLAAGLYSGITYGMTEVRGGAHDWRNSAVAGALTGA
AMARILMVSKGEELFTGVVPILV

OEP16-2-33CT:EGFP

MTTSERTSHEQVVQSALTGAAISTAANLLSSVFRILMVSKGEELFTGVVPILV

OEP16-2-Δ53CT:EGFP

MEKSGGRIVMDEIRSFKAHLFDLGHPLLNRADSFVKAAGV GALQAVSREAYFTVV
DGAGFDSNNV GPPSEITGNKKHRFPNLRGESSKSLDALVKNTGKESLQWGLAAGLYSGITYGMTEVRRGIL
MVSKGEELFTGVVPILV

OEP16-2-53CT:EGFP

MGAHDWRNSAVAGALTGAAMAMTTTSERTSHEQVVQSALTGAAISTAANLLSSVFRILMVSKGEELFTGVV
PILV

OEP16-2-Δ96CT:EGFP

MEKSGGRIVMDEIRSFKAHLFDLGHPLLNRADSFVKAAGV GALQAVSREAYFTVV DGAGFDSNNV GPPSE
ITGNKKHRFP GILMVSKGEELFTGVVPILV

OEP16-2-96CT:EGFP

MGESSKSLDALVKNTGKESLQWGLAAGLYSGITYGMTEVRRGGAHDWRNSAVAGALTGAAMAMTTTSERTS
HEQVVQSALTGAAISTAANLLSSVFRILMVSKGEELFTGVVPILV

OEP16-2-Δ121CT:EGFP

MEKSGGRIVMDEIRSFKAHLFDLGHPLLNRADSFVKAAGV GALQAVSREAYFTVV GILMVSKGEELFTGVV
PILV

OEP16-2-121CT:EGFP

MDGAGFDSNNV GPPSEITGNKKHRFPNLRGESSKSLDALVKNTGKESLQWGLAAGLYSGITYGMTEVRRGGA
HDWRNSAVAGALTGAAMAMTTTSERTSHEQVVQSALTGAAISTAANLLSSVFRILMVSKGEELFTGVVPILV

OEP16-2-H2:EGFP

MGESSKSLDALVKNTGKESLQWGLAAGLYSGITYGMTEVRRG GILMVSKGEELFTGVVPILV

EGFP:OEP16-2-H2

VTAAGITLGMDELYKSGLR SRGESSKSLDALVKNTGKESLQWGLAAGLYSGITYGMTEVRRG*GST

OEP16-2ΔH1/H2:EGFP

MDGAGFDSNNV GPPSEITGNKKHRFP GILGAHDWRNSAVAGALTGAAMAMTTTSERTSHEQVVQSALTGA
AISTAANLLSSVFRILMVSKGEELFTGVVPILV

OEP16-2-SD:EGFP

MDGAGFDSNNVGPPEITGNKKHRFPGILMVSKGEELFTGVVPILV

A4. OEP16-2 sequences from 30 species & the protein accession number retrieved from NCBI by pBLAST used to generate an MSA (Figure 6.2.1).

Genus species	Protein Accession Number
<i>Arabidopsis thaliana</i>	NP_849394.1
<i>Capsella rubella</i>	XP_023635268.1
<i>Eutrema salsugineum</i>	XP_024005643.1
<i>Brassica napus</i>	XP_013699139.1
<i>Quercus lobata</i>	XP_030974967.1
<i>Ziziphus jujuba</i>	XP_015899996.1
<i>Prunus persica</i>	XP_007227528.2
<i>Carica papaya</i>	XP_021906866.1
<i>Arachis hypogaea</i>	XP_025630202.1
<i>Medicago truncatula</i>	XP_003609756.2
<i>Malus domestica</i>	XP_008390546.2
<i>Manihot esculenta</i>	XP_021594264.1
<i>Populus trichocarpa</i>	XP_002312339.1
<i>Abrus precatorius</i>	XP_027339297.1
<i>Ricinus communis</i>	XP_002519203.1
<i>Mucuna pruriens</i>	RDX90919.1
<i>Spatholobus suberectus</i>	TKY72522.1
<i>Citrus clementina</i>	XP_006419410.1
<i>Pistacia vera</i>	XP_031264535.1
<i>Glycine max</i>	XP_003533195.1
<i>Nymphaea colorata</i>	XP_031498698.1
<i>Ananas comosus</i>	XP_020107689.1
<i>Syzygium oleosum</i>	XP_030456230.1
<i>Cajanus cajan</i>	XP_029126119.1
<i>Jatropha curcas</i>	XP_012084054.1
<i>Solanum pennellii</i>	XP_015087098.1
<i>Vigna radiata</i>	XP_014499162.1
<i>Dendrobium catenatum</i>	XP_020697588.1
<i>Momordica charantia</i>	XP_022138821.1
<i>Amborella Trichopodea</i>	XP_020520323.1

A5. Solutions prepared for protocols in methods and materials

LB Broth

LB broth components were mixed in DI H₂O to obtain a final concentration of 1% (w/v) NaCl, 1% (w/v) tryptone, and 0.5% (w/v) yeast extract. Then, autoclaved in a 30-minute liquid cycle.

Solid LB

Components were combined in DI H₂O to obtain a final concentration of 1% (w/v) NaCl, 1% (w/v) tryptone, 0.5% (w/v) yeast extract, and 2% (w/v) agar. Then autoclaved in a 30-minute liquid cycle.

1x TAE Buffer

A 50x stock of TAE buffer was prepared by mixing components in DI H₂O to achieve a final concentration of 2M Tris base, 5.71% (v/v) acetic acid, and 0.05M EDTA (pH 8.0). A 1x dilution was mixed to obtain a final concentration of 2% (v/v) 50x TAE and 98% (v/v) DI H₂O.

6x DNA Loading Buffer

6x DNA loading buffer was prepared by combining components in DI H₂O to obtain a final concentration of 30% (v/v) glycerol, 0.25% (w/v) bromophenol blue, and 0.25% (w/v) xylene cyanol FF.

6x Laemmli SDS PAGE Sample Loading Buffer

Constituents were combined in DI H₂O to achieve a final concentration of 375mM Tris-HCl, 9% w/v SDS, 50% v/v Glycerol, and 0.03% w/v Bromophenol Blue.

1x SDS-PAGE Running Buffer

A 10X stock was made, 30g of Tris, 1440g of glycine, and 10g of SDS were dissolved in 800ml of DI H₂O and then brought up to 1L using DI H₂O. 100ml of the 10x stock was mixed with 900ml of DI H₂O to achieve a 1x dilution.

SDS-PAGE gel Recipe to Separate OEP16-2 fusion constructs

Separating Gel:

Components were mixed in DI H₂O to achieve a final concentration of 375mM Tris-HCl (pH 8.8), 10% acrylamide, 0.1% (v/v) SDS, 0.05% (v/v) APS, and 0.05% (v/v) TEMED.

Stacking Gel:

Components were mixed in DI H₂O to obtain a final concentration of 125mM Tris-HCl (pH 6.8), 4.8% (v/v) acrylamide, 0.1% (v/v) SDS, 0.05% APS, 0.1% (v/v) TEMED.

1x Transfer Buffer

Transfer buffer components were mixed in DI H₂O to obtain a 10x stock concentration of 250mM Tris-HCl (pH 7.6) and 1.92M glycine. A 1x dilution was achieved by mixing solutions to a final concentration of 80% (v/v) DI H₂O, 10% (v/v) 10x stock, and 10% (v/v) methanol.

Ponceau Stain

Ponceau stain was prepared by mixing components in dH₂O to achieve a final concentration of 1% (v/v) acetic acid and 0.5% (w/v) Ponceau S.

A6. List of Bioinformatic Servers and URLs

ChloroP - <http://www.cbs.dtu.dk/services/ChloroP/>

HeliQuest Analysis - <https://heliquest.ipmc.cnrs.fr/cgi-bin/ComputParams.py>

HHomp - <https://toolkit.tuebingen.mpg.de/tools/hhomp>

Phobius - <https://phobius.sbc.su.se/index.html>

PRED-TMBB - <http://bioinformatics.biol.uoa.gr/PRED-TMBB/>

ProtParam - <http://protparam.net/index.html>

PSI-PRED & MEMSAT-SVM - <http://bioinf.cs.ucl.ac.uk/psipred/>

SWISS-MODEL - <https://swissmodel.expasy.org/interactive>

TMHMM - <http://www.cbs.dtu.dk/services/TMHMM/>

TM-Pred - https://embnet.vital-it.ch/software/TMPRED_form.html