Investigating Subcellular Localization of *Arabidopsis thaliana* Hydroperoxide Lyase I (At4g15440)

Author: Asher Kim

Student ID: 20933203

Supervisor: Dr. Simon D.X. Chuong

Second Reader: Dr. Andrew C. Doxey

Date of Submission: April 15, 2025

This report is submitted as a requirement of BIOL499 Senior Honours Project.

Acknowledgements

I am incredibly grateful to my supervisor, Dr. Simon D.X. Chuong, for providing me the opportunity to conduct this research and his invaluable support and mentorship throughout this project. I am thankful to the other members of the Chuong Lab, especially Dr. Barbara Moffatt and Charis Tam, for their advice and encouragement in the lab. I would also like to thank Dr. Andrew C. Doxey, for offering his time and expertise as my second reader.

Abstract

During the transition from an endosymbiont to a photosynthetic organelle, many of the cyanobacterial genes were transferred to the host's nuclear genome. As a result, most of the genes for chloroplast proteins evolved localization signals to target their destined region of the organelle. To localize to the chloroplast membrane, proteins may be signal anchored, tail anchored, β – barrel proteins, or contain a cleavable N-terminal transit peptide or reverse transit peptide-like C-terminal sequence. Hydroperoxide lyase (HPL, At4g15440) is an *Arabidopsis thaliana* chloroplast outer envelope protein involved in stress response by regulating stress response molecules called arabidopsides. While HPL is predicted to be a multi-pass alpha helical transmembrane protein with neither an N-transit peptide nor a C-transit peptide-like sequence, it is unknown which regions of HPL are responsible for its localization to the chloroplast outer membrane. Using bioinformatic software, three putative transmembrane domains were predicted between 75-100 amino acids (aa), 180-220 aa, and 350-370 aa in the 384 aa sequence. To verify regions involved in localization, the HPL open reading frame was ligated into plasmid vectors containing EGFP, and the vectors were transformed into onion epidermal cells via biolistic bombardment. EGFP-HPL fusion proteins appeared to be consistently localized to the nucleus, cytoplasm, and punctate-like structures. The punctate-like structures were identified to be plastids by co-bombarding with plastid marker ferredoxin-DsRed. Future studies will investigate deletion constructs of HPL to determine more specific regions responsible for HPL localization to chloroplasts.

Table of Contents

List of Tables

List of Figures

1. Introduction

Chloroplasts are well-known as the organelles that perform photosynthesis in plants, but they also carry out the biosynthesis of amino acids, lipids, vitamins, phytohormones, secondary metabolites, and other compounds (Song et al., 2021). Both chloroplasts and mitochondria originated from endosymbiotic bacteria, which gradually transformed into organelles (Lee et al., 2019). The bacteria gained a second membrane upon engulfment, and most of their genes were transferred to the nuclear genome (Kim et al., 2019; Stegemann et al., 2003). It is estimated that about 2000-3000 proteins are synthesized in the cytosol from the nuclear genome (Jin et al., 2022), and these proteins must be transported to the appropriate organelle membranes (Fish et al., 2022). Proteins destined for the chloroplast or mitochondria are typically produced with localization signals (Lee et al., 2019). In mitochondria, this signal is a presequence; in chloroplasts, this signal is a transit peptide (TP) (Lee et al., 2019). Both presequences and TPs contain short sequence motifs that direct protein import (Lee et al., 2019). There is a high amount of diversity among presequence and TP sequences, and neither have a consensus sequence (Lee et al., 2019). Interestingly, some *in vivo* and *in vitro* experiments have shown that some chloroplast proteins are imported to the mitochondria in the absence of plastids (Lee et al., 2019).

Additional cellular components help localize proteins. Molecular chaperones, including heat shock proteins 70 and 90, aid in the transportation of TP-containing preproteins to chloroplasts (Fish et al., 2022). Molecular chaperones may prevent preproteins from aggregating or becoming translocation-incompetent (Kim et al., 2019).

Generally, cytosolic ribosomes synthesize chloroplast preproteins, which are then transported to the chloroplast using their N-terminal TPs (Fish et al., 2022). The translocon at the outer membrane of the chloroplast (TOC) and the translocon at the inner membrane of the

chloroplast (TIC) identify and import the preproteins through the inner and outer chloroplast membranes (Fish et al., 2022). Mutations in TOC or TIC complexes have been observed to cause albinism or even embryonic lethality (Jin et al., 2022). There are three consensus components of TOC (Jin et al., 2022). The first, Toc75, is a β-barrel transmembrane protein that acts as the preprotein conducting channel of the complex (Jin et al., 2022). The other two components, Toc159 and Toc34, are both GTPases that act as TP import receptors (Jin et al., 2022; Lung et al., 2014). The classic model of TIC establishes Tic110 as the main component, accompanied by Tic20, Tic21, Tic22, Tic32, Tic40, Tic55, Tic62 (Jin et al., 2022). An alternative model proposes Tic20, Tic56, Tic100, and Tic214 as components of a protein translocation complex, as observed in *Arabidopsis* (Jin et al., 2022).

Proteins on the chloroplast outer and inner membranes have important roles including ion and metabolite exchange, protein import, and signal transduction (Kim et al., 2019). Chloroplast membrane proteins were traditionally categorized as signal anchored (SA), tail anchored (TA), or β-barrel proteins (Fish et al., 2022). SA proteins are anchored in the chloroplast membrane by an approximately 20 amino acid N-terminal α-helix (Fish et al., 2022). The targeting sequence consists of the transmembrane domain (TMD) and a C-terminal positively charged region (Lee et al., 2011). TA proteins are anchored in the chloroplast membrane by a C-terminal α-helix, which forms the targeting sequence with a C-terminal sequence (Moog, 2019). Unlike SA and TA proteins, β-barrel proteins do not possess α-helical TMDs (Fish et al., 2022). Instead, they are comprised of 8-12 β-strands containing 9-11 amino acids each, with the strands at a 45° angle from the membrane (Fish et al., 2022). Additionally, some chloroplast membrane proteins localize using a cleavable N-terminal TP or a reverse TP-like C-terminus sequence (Fish et al., 2022). For

example, Toc75 and OEP80 are β-barrel proteins with an N-terminal TP, and a TOC homolog in *Bienertia sinuspersici* contains a TP-like C-terminal sequence (Fish et al., 2022; Lung et al., 2014).

Presently, 138 chloroplast outer membrane proteins have been identified (Fish et al., 2022; Inoue, 2015). The functions of these proteins have been categorized as lipid metabolism, carbohydrate metabolism, solute and ion transport, protein turnover and modification, protein import, intracellular communication, and other metabolic and regulatory functions (Inoue, 2015).

One chloroplast outer membrane protein is hydroperoxide lyase 1 (HPL1), which was identified by Inoue in 2007 (Inoue, 2007). HPL1 is a member of the cytochrome P450 family and forms the CYP74 enzyme group along with allene oxide synthase and divinyl ether synthase (Hughes et al., 2009). The CYP74 enzymes are a unique group, specializing in hydroperoxide metabolism and not requiring oxygen nor NADPH-reductase for their enzyme mechanisms (Hughes et al., 2009). In *A. thaliana,* HPL1 is a key regulator of stress response molecules called arabidopsides, which are galactolipids with esters of 12-oxo-phytodienoic acid and dinor-12-oxo-phytodienoic acid (Nilsson et al., 2016).

HPL1 is a multi-pass α-helical transmembrane protein and possesses neither a N-transit peptide nor C-transit peptide-like sequence (Fish et al., 2022). It is not known which regions of this key stress response protein are responsible for its insertion into the chloroplast outer membrane. Therefore, the goal of this experiment is to determine regions of HPL1 involved in localizing to the chloroplast outer membrane. By elucidating mechanisms of chloroplast protein transport, this study will contribute to the current understanding of chloroplast evolution and organelle biogenesis.

2. Materials and Methods

2.1 Bioinformatic Analysis

Bioinformatic tools were used in a preliminary analysis of the HPL1 protein sequence. PSIpred, TMHMM, Phobius, MEMSAT, PyMOL, and Alphafold were used to predict secondary structures, the protein's 3D structure and TMDs.

2.2 PCR Amplification of HPL1

Primers flanking the open reading frame HPL1 were purchased from Eurofins Genomics (Louisville, KY, USA; **Figure 1; Table 1)**. pUNI51 plasmids containing HPL1 open reading frame (ORF) cDNA were obtained from the Arabidopsis Biological Resource Centre, Columbus, OH, USA. PCR amplification of HPL1 was performed using these primers and Thermo Scientific Phusion high fidelity DNA polymerase, which produces blunt-ended DNA fragments. PCR products were stored in the freezer at -80°C degrees.

To verify the identities of the amplified products, the PCR products were run on a 0.8% 1xTAE agarose gel. The HPL ORF was recovered from the gel using a Thermo Scientific GeneJET Gel Extraction Kit, and concentration of extracted DNA was measured using a Thermo Scientific NanoDrop One.

2.3 Plasmid Ligation

The gel-purified amplified HPL1 PCR product (sequence) was ligated into pSAT6 plasmid vector systems. Both pSAT6-C1 and pSAT6-N1 were used, to produce proteins with EGFP at either of the N- and C-termini of HPL1. This was done to account for possible interactions between EGFP and the N- and C-termini of HPL1 (Snapp, 2005). The plasmid will be cut with SmaI, which produces blunt ends. Then, the plasmid vector and insert were incubated for overnight (1 day) at

4°C with 10X T4 ligation buffer, standard distilled water, and T4 ligase. The resulting plasmid contained an ampicillin resistance gene, gene regulatory elements, and HPL1 in frame with GFP (Snapp, 2005). Control plasmids with no insert were also be prepared.

2.4 Heat Shock Transformation of Competent *E. coli* cells

Heat shock transformation was used to transform competent DH5α *Escherichia coli* cells with the ligated plasmids. The competent *E. coli* cells had been stored at -80°C and were thawed for use in the transformation. Next, 60 µL of competent cells were added to each ligation mix and incubated on ice for 30 minutes. The mixture was heat shocked for 3 minutes at 37°C then kept on ice for 2 minutes. 800 µL of LB was added to each mixture, and each of the contents were transferred into 13 mL sterile tubes which were incubated in a shaker for 40 minutes at 37°C. The mixtures were centrifuged at 10 000 rpm for 1 minute to pellet cells, and about 800 µL of the supernatant was discarded. Finally, about 75 µL of the resuspended mixture was spread on LB agar ampicillin plates and incubated at 37°C for 1 day, after which they were stored at 4°C.

2.5 Colony Screening

From the ampicillin containing plate, 30 colonies were selected and added to about 1 mL of LB ampicillin in 13 mL sterile tubes. The tubes were incubated in a shaker for about 1 day at 37°C and 220 rpm. Then, plasmid DNA was collected using a crude plasmid DNA isolation method. This procedure is modified from the Qiagen Spin Miniprep kit so as to not require a spin column. The saturate cell culture was transferred to 1.5 mL Eppendorf tubes and centrifuged at 13,000 rpm for 2 minutes to pellet cells, and the supernatant was discarded. The pellet was resuspended in 100 µL of P1 by vortex. Next, 100 µL of P2 was added to each Eppendorf tube and they were gently inverted several times. 350 µL of P3 was added and the tubes were similarly inverted. The tubes were centrifuged at 13,000 rpm for 5 minutes which pelleted the debris. The

supernatant was carefully removed, added to a new 1.5 mL Eppendorf tube with 300 µL isopranol, and incubated on ice for about 15 minutes. The tubes were centrifuged for 15 minutes at 14,000 rpm at 4°C which pelleted the plasmid DNA. 300 µL of 70% ethanol was gently added to each tube and drained promptly. Once the tubes had finished air-drying, 30 µL of warm elution buffer was added to each.

Gel electrophoresis was used to predict which of the colonies had taken up the insert. About 5 µL of each sample of crude plasmid DNA was added to a drop (2-3 uL) of 6x loading dye, and samples were run alongside 5 µL of 1 Kb plus DNA Ladder RTU. The samples were run on a 0.8% agarose 1 X TAE gel at 80 V for a 30 mL gel or at 100 V for a 60 mL gel. Gels were imaged on a BioRad ChemiDoc MP Imaging System and using the ChemiDoc MP Nucleic Acid Gels GelRed imaging application, and the images were processed with Adobe Photoshop CS. Bands that had travelled more slowly and shifted towards the top of the gel were assumed to contain the insert.

2.6 Colony PCR Screening

Samples that appeared to contain the insert based on the results of the colony screening gel electrophoresis were further analyzed using PCR and gel electrophoresis. The crude plasmid DNA from these samples were used as templates in PCR reactions with KOD DNA polymerase. The PCR reactions used one insert-specific primer and one vector-specific primer in order to only amplify a product if the insert was positioned in the desired direction.

To analyze the PCR results, another agarose gel electrophoresis was run using the PCR products. 5 µL of 6x loading dye was added to the contents of the PCR tubes and run on the gel alongside 5 µL of 1 Kb plus DNA Ladder RTU. Positive controls were TRAF constructs that had reliably shown amplification with PCR. If a band was seen amplified at approximately the length

of the ORF (1155 bp), it was assumed that the colony contained the insert in the desired direction and the culture was incubated further. The colony's culture was inoculated into one or two 13 mL sterile tubes containing 4 mL of LB ampicillin. Then, they were incubated in a shaker at 37°C and 220 rpm for about 24 hours.

From the 4 mL cultures, the plasmids were isolated and purified using the Bio Basic Molecular Biology Kit for plasmid purification. The EZ-10 Spin Column Plasmid DNA Minipreps Kit BS413, BS414, BS614 protocol was followed on pages 4-7 of the EZ-10 Spin Column Handbook with a few exceptions (*EZ-10 Spin Column Handbook*, 2019). Before step 2, the pellet was resuspended by vortex, and in step 11, 30 µL rather than 50 µL of Elution Buffer was inserted into the column. After DNA extraction, the concentration of DNA was measured using a Thermo Scientific NanoDrop One.

2.7. Biolistic Bombardment of Onion Epidermal Cells

White onion bulbs were purchased from a local grocery store and inner layers were cut into approximately 1 cm x 1 cm rectangular sections. Onion samples were often used on the same day of purchase, and were rarely used past 2 days of purchasing. The onion epidermal cells were transformed via biolistic bombardment and DNA-coated tungsten balls. About 1 µg of the plasmid DNA extracted in section 2.6 was vortexed thoroughly with 8.0 µL of 60 mg/mL tungsten, 10 µL of 2.5 M $CaCl_2$, and 5 µL of 0.1 M spermidine. The mixtures were centrifuged for 1 minute to pellet the tungsten. Then the supernatant was discarded, and the tungsten was washed with 70% ethanol followed by 100% ethanol. The tungsten solutions were spread onto microcarrier discs and loaded into the top level of a BioRad Model PDS-1000/He Biolistic Particle Delivery System. Four onion slices were arranged on an open petri dish in a rectangular formation, inserted into the second lowest level of the Biolistic Particle Delivery System and bombarded with the DNA-coated

tungsten. The bombardment was done using Bio-Rad Laboratories Inc. 1,350 psi rupture disks at 11 cm under a vacuum of 26.5 in Hg. The bombarded onion samples were incubated in a dark drawer at room temperature for about 24 hours on a moist filter paper in a Petri dish. Controls of plasmids with no inserts were also prepared. About three replicated experiments of bombardments were completed for each construct.

2.8 Imaging of GFP Constructs Using Fluorescent Microscopy

The expression and localization of the HPL1 GFP fusion construct as observed using fluorescent microscopy. There were about 10 biological replicates of expressing onion cells observed per bombardment. The bombarded onion cells showing expression were observed using a Zeiss Axio Imager D1 and images were taken using AxioVision software. To image the fluorescence of GFP, the Axio Imager D1 was set to the 38 HE GFP filter set, which used an excitation wavelength of 470 nm and an emission wavelength of 525 nm. Images were processed using Adobe Photoshop CS.

2.9 Co-bombardment with Ferredoxin TP – DsRed Fusion Protein

After several bombardments using pSAT6::HPL, onion epidermal cells were co-bombarded with both pSAT6::HPL full length constructs and ferredoxin transit peptide – DsRed (FdTP::DsRed) fusion proteins. The ferredoxin transit peptide was used to target plastids, as ferredoxin is a protein in the electron transport chain in chloroplasts (van't Hof et al., 1993), while DsRed was used as a fluorescent marker. To image the fluorescence of DsRed, the Axio Imager D1 was set to the 43 HE CY3 filter set, which uses an excitation wavelength of 550 nm and an emission wavelength of 570 nm.

2.10 Deletion Constructs

The HPL1 ORF was subdivided into two regions, each about 500 bp long, by PCR amplification using forward 1 reverse 3 (F1R3) primers and forward 2 reverse 4 (F2R4) primers, shown in **Figure 1**. The regions were inserted into pSAT6 plasmids, cultivated in competent DH5-α *E. coli* cells, and screened for the presence of the insert as described in 2.1-2.6.

3. Results

3.1 Bioinformatic Analysis

PSIPRED 4.0 was used to predict the secondary structures of HPL1 (Jones, 1999). As seen in **Figure 1**, several helix regions and a few short strand regions were predicted with high confidence in some segments. Roughly half of the sequence was predicted to constitute a helix, and two groups of several strands were predicted towards the C-terminus. No regions of the protein were predicted to contain a signal peptide, transmembrane helix, or membrane interaction region.

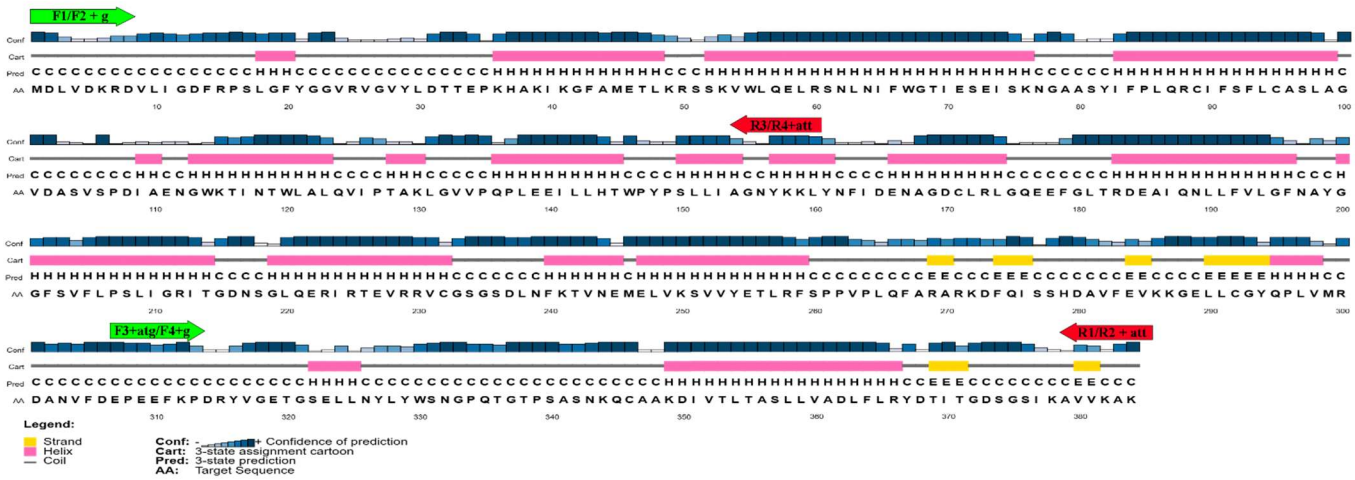

Figure 1: PSIPRED chart of secondary structure predictions of HPL1 amino acid sequence, including primers and confidences of predictions (Jones, 1999). Pairs of primers (F1 and F2,

R3 and R4, etc.) span the regions indicated by the arrows, and differ by a few nucleotides, as shown in the text in the arrows.

Phobius was used to predict the transmembrane domains and signal peptide regions of HPL1 (Käll et al., 2004). As seen in **Figure 2**, three putative TMDs were predicted between approximately 80-105 aa, 180-215 aa, and 345-370 aa. The second TMD was predicted with a probability of 0.6, and the first and third TMDs were predicted with lower probabilities of <0.1. The remainder of the protein was predicted to be mostly non-cytoplasmic, and none of the protein was predicted to be a signal peptide, transmembrane helix, or membrane interaction region.

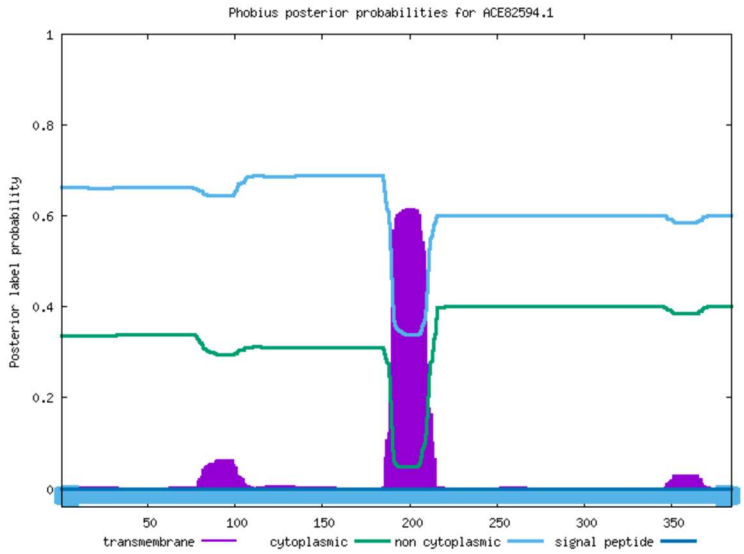

Figure 2: Phobius plot of predicted transmembrane, cytoplasmic, non-cytoplasmic, and signal peptide regions in HPL1 (Käll et al., 2004).

Alphafold was used to predict the 3D structure of HPL1 (Jumper et al., 2021; Varadi et al., 2024). Shown in **Figure 3**, Alphafold predicted a roughly globular protein with many helices and a few sheets, with most of the structure predicted with a very high per-residue model confidence

score, and couple regions of high or low confidence. The sheets were arranged together in two regions, and the helices were clustered together angled in several distinct directions.

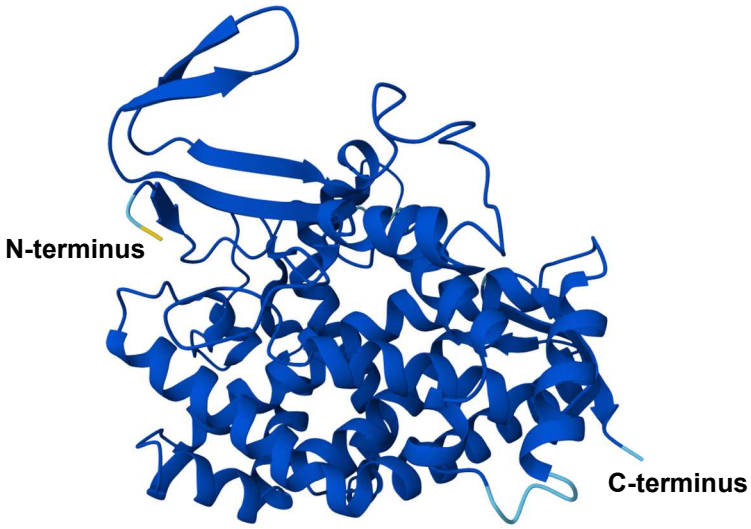

Figure 3: 3D structure prediction made by Alphafold (AFDB Accession Number AF-B3LF83-F1-v4). Dark blue colouring represents a very high model confidence score, light blue represents high confidence, and yellow represents low confidence (Jumper et al., 2021; Varadi et al., 2024). The N-terminus is coloured yellow, while the C-terminus is light blue.

To visualize the locations of the putative TMDs within the protein's 3D structure, the Alphafold predicted HPL protein structure was loaded into PyMOL. This visualization, shown in **Figure 4**, shows that Phobius' three predicted TMDs coincided neatly with three of Alphafold's predicted α-helices. Further, the three helices were clustered together towards the middle of the globular protein at slight angles to one another.

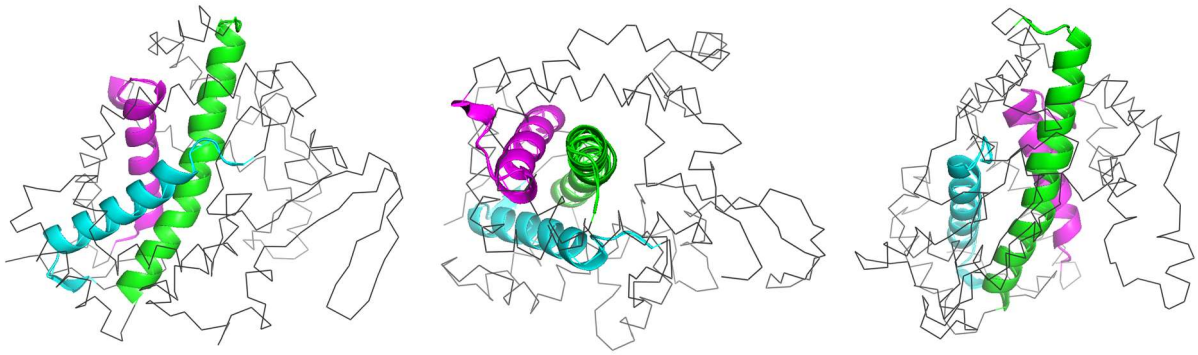

Figure 4: Several angles of the PyMOL 3D representation of HPL1 highlighting transmembrane domains predicted by Phobius (*The PyMOL Molecular Graphics System*, n.d.). The first transmembrane domain from 80-105 aa is coloured pink, the second transmembrane domain from 180-215 aa is coloured green, and the third transmembrane domain from 345-370 aa is coloured teal. The remainder of the protein is represented as a ribbon representation so the TMDs can be more easily viewed.

Finally, Deep TMHMM was used to further predict the transmembrane topology of HPL1. As seen in **Figure 5**, the entire protein sequence was predicted to be internal with a very high probability of 1.0.

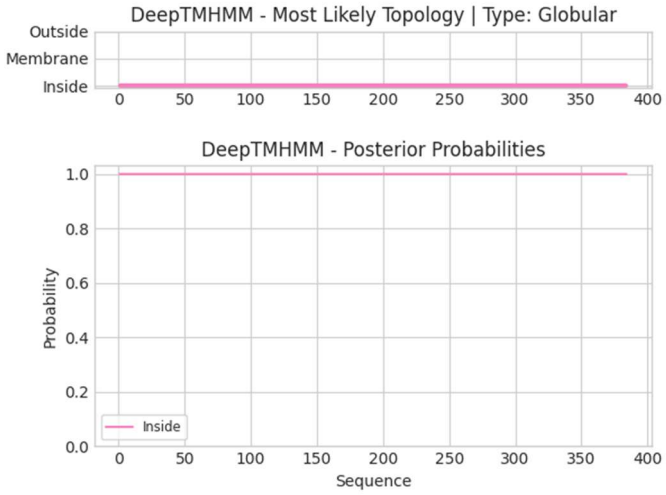

Figure 5: Deep TMHMM plot of HPL1 predicted transmembrane topology (Hallgren et al., 2022).

MEMSAT-SVM was used to predict extracellular, cytoplasmic, and transmembrane domains, as seen in **Figure 6**. This software also predicted three transmembrane domains, but from 118-146 aa, 207-192 aa, and 351-388 aa.

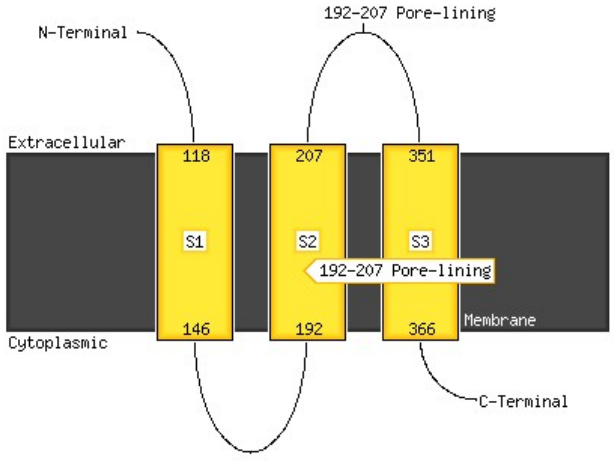

Figure 6: MEMSAT-SVM Schematic of TMDs (Nugent & Jones, 2009).

3.2 PCR of HPL Full Length and Deletion Sequences

Figure 7 shows the results of the initial PCR of the full length HPL1 ORF. Both the forward 1 reverse 1 (F1R1) primers and forward 2 reverse 2 (F2R2) primers resulted in an amplified band slightly larger than 1000 bp, which was very similar to the expected length of 1155 bp. The two sets of primers both amplify the full length HPL1 ORF, and differ only in the presence or absence of a start codon, stop codon, or additional base pair to preserve the reading frame, which are shown in **Figure 1** and **Table 1**.

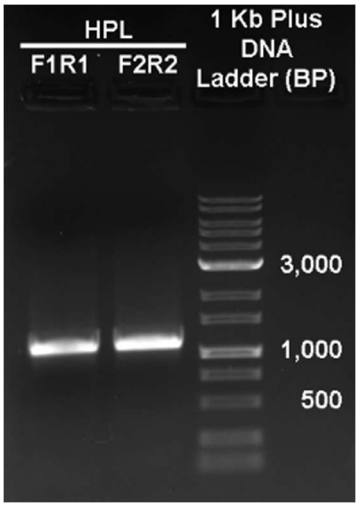

Figure 7: Phusion PCR of HPL full length sequence. Phusion PCR of the HPL ORF was performed using two sets of primers, the F1R1 primers and F2R2 primers. Both sets of primers resulted in a band near the expected size of 1155 bp.

Figure 8 shows Phusion PCR result of the HPL deletion sequences, where the F1R3 and F2R4 primers amplified a region of about 500 bp. The primers and their amplified sequence are shown in **Figure 1**.

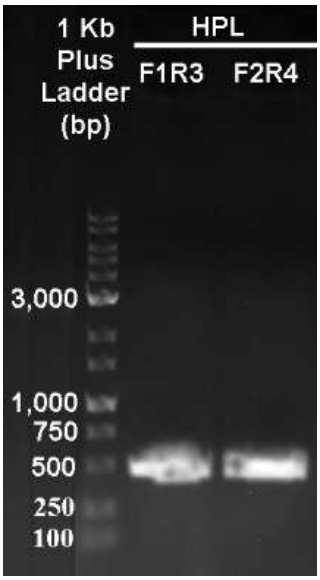

Figure 8: Phusion PCR of HPL deletion sequences. Phusion PCR of the HPL deletion sequences was performed using two sets of primers, the F1R3 primers and F2R4 primers.

3.3 Colony Screening for Presence of Insert

Figure 9 and **Figure 10** show the colony PCR results for pSAT6N1 and pSAT6C1 respectively. In **Figure 9**, N1-HPL colonies 6, 14, 23, 27, 28 and 30 travelled more slowly and were shifted backwards relative to the other colonies, so they were assumed to contain the insert and were further analyzed by PCR and gel electrophoresis. In **Figure 10**, similar observations and analyses were made on the C1-HPL 3, 5, 6, 9, 14, 16 and 18 colonies.

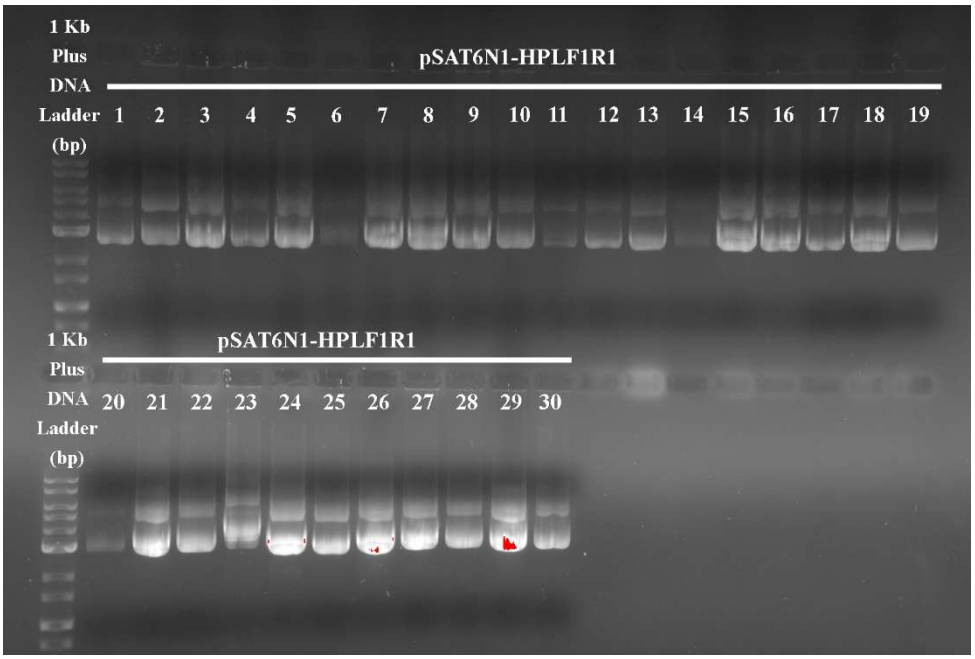

Figure 9: pSAT6N1-HPL colony screening for the presence of the HPL insert.

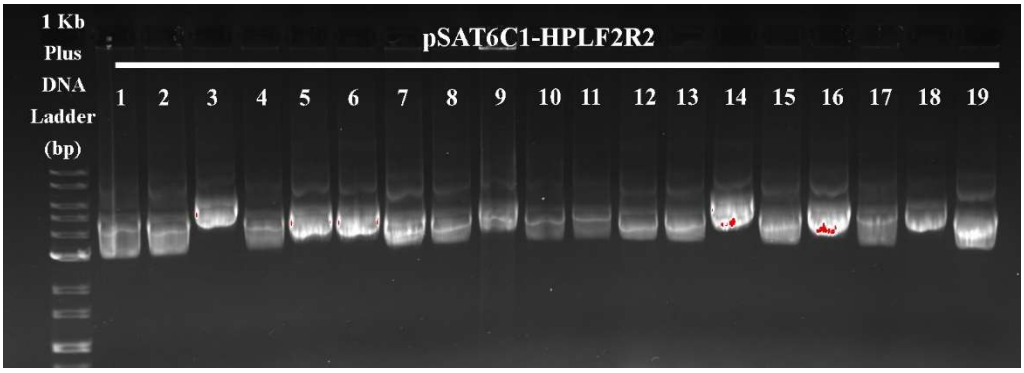

Figure 10: pSAT6C1-HPL colony screening for the presence of the HPL insert.

3.4 Colony PCR Screening for Direction of Insert

PCR was performed on crude plasmid DNA that appeared to contain an insert. As seen in **Figure 11**, the twenty-third N1-HPL colony (N1-HPL 23) showed amplification between 1200-1300 bp, which is slightly larger than the HPL insert length of 1155 bp. Similarly, **Figure 12** shows the third and fourteenth C1-HPL colonies (C1-HPL 3 and C1-HPL 14) with amplified bands also between 1200-1300 bp. These band sizes were expected, since the amplified sequence contained

the HPL ORF and the multiple cloning site. Primers were selected so that the HPL ORF and MCS would be amplified only if the insert was in the correct orientation. Therefore, it was assumed that the inserts were in the correct orientation in the plasmids for N1-HPL 23, C1-HPL 3, and C1-HPL 14.

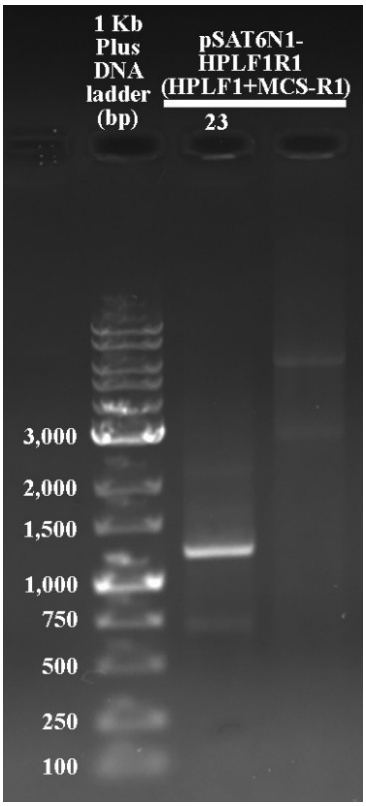

Figure 11: Colony PCR results of pSAT-6 N1-HPL. The PCR amplification of the plasmid DNA from N1-HPL 23 screened showed a band at the expected size between 1200-1300 bp.

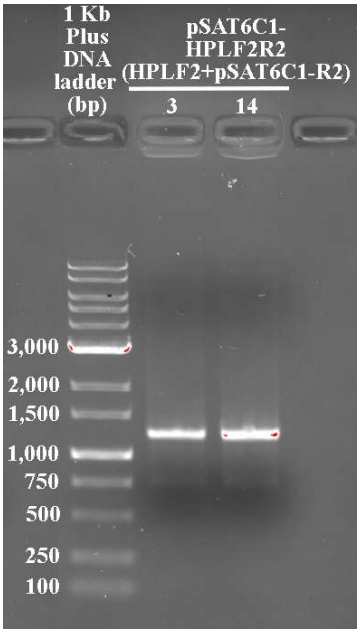

Figure 12: Colony PCR results of pSAT-6 C1-HPL. The PCR amplification of the plasmid DNA from C1-HPL 3 and C1-HPL 14 screened had a band at the expected size between 1200-1300 bp.

3.5 Biolistic Bombardment of Onion Cell Epidermis

There was some variation in the expression patterns observed in the onion epidermal cells, but the onion cells shown in **Figure 13** and **Figure 14** are representative of the 3 replicated experiments of bombardments per construct, each containing 4 samples of onion slices and about 10 biological replicates of expressing cells.

Figure 13 shows onion epidermal cells expressing EGFP. Negative controls of empty pSAT6 vectors are shown in **(A)** and **(C)**, and both vectors had nuclear and cytosolic EGFP signals in expressing cells. The two plasmid constructs differ only in the placement of EGFP relative to the multiple cloning site, where the pSAT6-C1 vector **(A)** has the MCS downstream of EGFP, and the pSAT6-N1 vector **(C)** has the MCS upstream of EGFP.

(B) and **(D)** in **Figure 13** show representative onion epidermal cells expressing the full length HPL1::EGFP fusion protein. The pSAT6-N1::HPL construct **(D)** showed localization to some punctate structures pressed against the cell wall, while pSAT6-C1::HPL **(B)** showed localization to a few punctate structures among wispy trails and clouds of EGFP expression.

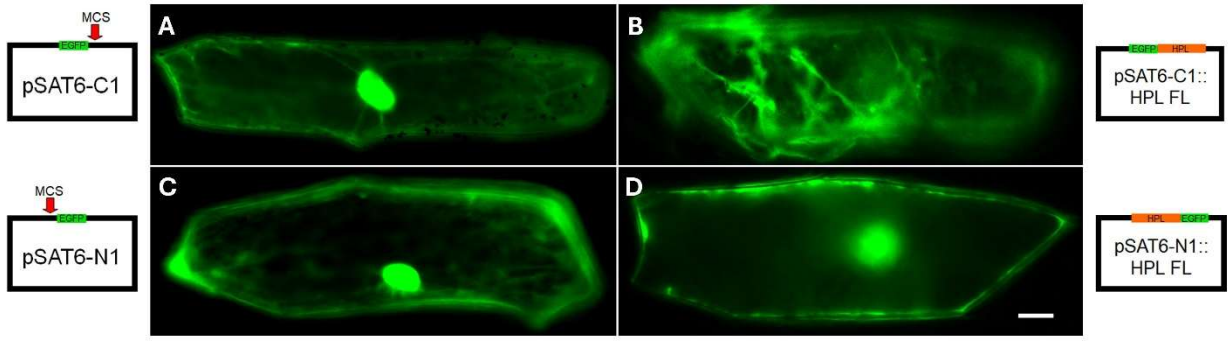

Figure 13: Onion epidermal cells expressing EGFP fusion constructs. (A, C) Empty pSAT6-C1 and –N1 vectors were bombarded into onion epidermal cells as negative controls. **(B, D)** Onion epidermal cells expressing the full length HPL1 construct. Scale bar represents 30 μm.

The results of the co-bombardment of onion epidermal cells with EGFP fusion constructs and ferredoxin transit peptide – DsRed fusion constructs are shown in **Figure 14**. In the first column of **Figure 14**, pSAT6-C1::HPL full length constructs targeted numerous punctate structures uniform in size, often with "tails". As shown in the second column, ferredoxin transit peptide and Ds-Red constructs targeted plastids, which aligned with the targeting of EGFP. As can be seen in the Merged images, overlap of EGFP and DsRed appears yellow. pSAT6-N1::HPL full length constructs also expressed localization to punctate structures, often with lower intensity than that of the pSAT6-C1::HPL constructs. The punctate structures also neatly overlapped with the ferredoxin transit peptide and Ds-Red constructs which identified plastids.

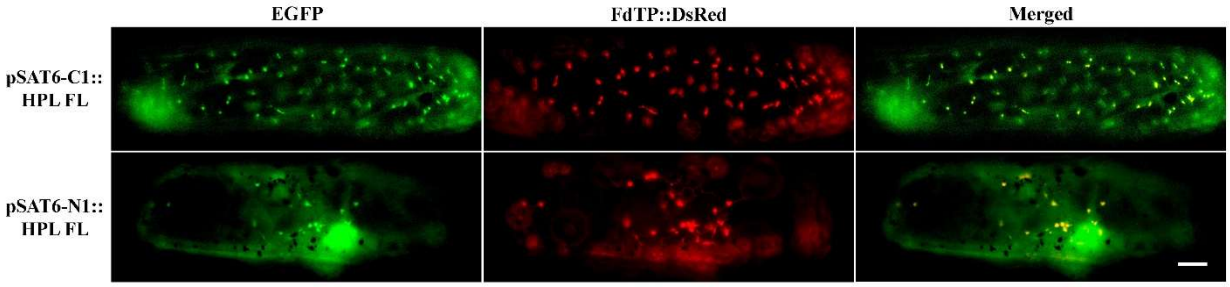

Figure 14: Co-bombardment of onion epidermal cells with EGFP fusion constructs and FdTP::Ds-Red constructs. Scale bar represents 30 μm.

4. Discussion

4.1 HPL1 Secondary Structure Predictions

The results of the bioinformatics analysis had several agreements and disagreements between the different software used. First, PSIPRED predicted 20 α-helices and 6 β-sheet strands (**Figure 1**) while Alphafold predicted about 14 helices and 6-8 strands (**Figure 3**). Deep TMHMM predicts that the entire protein is located inside the cell with a very high probability of 1.0, although Phobius predicts that the protein is nearly entirely non-cytoplasmic with a probability of about 0.65. On another hand, the TMDs predicted by Phobius correlated well with Alphfold's predicted helices, which are highlighted in **Figure 4**. These putative TMDs also roughly correspond with α-helices predicted by PSIPRED, although the second TMD is broken into two helices by PSIPRED, as seen in **Figure 1** around 200 aa. It is possible that these 3-4 helices are the major transmembrane regions of HPL. The first and third TMDs predicted by Phobius had low confidences of under 0.1, so those two regions may have other roles. Additional experimental results are needed to verify the role of the three putative transmembrane domains and other regions of HPL.

HPL1 may also be a hybrid of transmembrane helices and a partial β-barrel. Similar hybrid proteins have been observed, such as TOC75 which possesses both a cleavable bipartite N-terminal chloroplast TP and a β-barrel (Fish et al., 2022).

The locations of positively charged amino acids can also suggest potential TMDs, as positive amino acids tend to be located at the edges of transmembrane helices (Baker et al., 2017). There is only one noticeable region rich in positively charged residues in HPL1. At around 230 aa, there are four arginines clustered near each other. This region also happens to be at the C-terminal end of the second putative TMD. The other predicted TMDs do not have similar clusters of lysine or arginine, but they may still be involved in localization.

4.2 HPL1 Full Length Localization

The biolistic bombardment of onion epidermis with full length HPL::EGFP fusion constructs was performed to verify its localization to plastids, establish a baseline for the construct's organelle targeting, and investigate whether there might be any interaction between EGFP and potential targeting mechanisms at HPL1's C1 or N1 termini. As seen in **Figure 13** and **Figure 14**, the full length HPL1::EGFP fusion constructs can be observed at the nucleus, punctate structures, and wispy, trailing areas. The connective, wispy areas seen in **Figure 13 (B)** and other cells expressing HPL1::EGFP may be stromules, which can traffick proteins between plastids (Hanson & Sattarzadeh, 2013). To identify the punctate structures, a co-bombardment with FdTP::DsRed was performed. The FdTP::DsRed fusion protein is expected to localize to plastids (van't Hof et al., 1993), and there was a high amount of overlap between the punctate structures targeted by HPL1::EGFP and FdTP::DsRed, which is shown in the Merged column of **Figure 14**. Therefore, the punctate structures targeted by HPL1::EGFP are likely to be plastids.

Although most onion epidermal cells expressing HPL1::EGFP showed targeting to punctate structures, there seemed to be more noticeable and consistent targeting with the pSAT6-C1::HPL construct. The ability of the pSAT6-N1::HPL construct to target plastids may have been diminished slightly by the presence of EGFP at the C1 terminus. Both Phobius and MEMSAT proposed putative TMDs near the C1 terminus of HPL1. If there is a TMD or other localizing structure at the C1 terminus, EGFP, which is a 27 kDa protein (Prasher et al., 1992), may interfere with or obscure it.

It was expected that the full length HPL1 would localize to chloroplasts, based on previous proteomics studies (Froehlich et al., 2003; Simm et al., 2013), and also because the chloroplast membrane is heavily involved in lipid metabolism (Cook et al., 2021). It is possible that HPL1 also localizes to other organelles in *A. thaliana*. In other species, HPL1 has been observed localizing to a wide variety of targets, such as the endomembrane and lipid bodies in almond plants (Mita et al., 2005), the cytosol, lipid droplets, and plastids in *Medicago truncatula* (De Domenico et al., 2007), and thylakoids in potato plants (Farmaki et al., 2007).

4.3 Difficulties and Possible Improvements

Some difficulties were encountered during colony screening and onion bombardment. First, there were some difficulties with screening the transformed C1-HPL colonies. It took more attempts than expected to obtain a C1-HPL colony containing an insert in the correct orientation. Although shifting in bands was seen in several rounds of colony screening, implying that the insert was present, none of the colonies had amplified bands in colony PCR screening. Positive controls were consistently amplified in the colony PCR screening, so it was unlikely an issue with PCR. The final round of C1-HPL colony screening showed very large shifting in the bands, as seen in **Figure 10**. All previous rounds of C1-HPL colony screening had smaller amounts of band shifting,

similar to what is seen in **Figure 9**. It is possible that all of the colonies screened after PCR contained the insert in the wrong direction. Alternatively, it is possible that the cells did not take up the insert at all in those previous heat shock transformations and colony screenings, since the shifting was much less noticeable than in the successful colony screening shown in **Figure 10**.

Additionally, two of the N1-HPL colonies, N1-HPL 17 and N1-HPL 18, appeared to contain the insert in the correct orientation based on colony screening and colony PCR screening, but they showed no EGFP expression after several biolistic bombardments. This may have been due to damage or another issue with the HPL DNA that occurred throughout the experiment, because a later heat shock transformation yielded the N1-HPL 23 colony which showed EGFP expression upon its first bombardment.

Some of the onions showed fewer expressing cells than expected, even with the positive control vector which had no insert. It was expected that each onion sample would yield about 20 expressing cells, but some onion slices showed only 5-6. Onions purchased on the same day as the bombardment tended to have improved bombardment results, so the lower yield may have been due to the freshness of the onions. Additionally, this experiment took place during the winter months, so grocery store produce may have had variable or reduced quality.

4.4 Next Steps and Further Research

This project will be continued by investigating the localization of deletion constructs of HPL1. Some deletion constructs have already been made as described in 2.10, but colonies containing the inserts in the correct directions have not yet been found. Future research will complete screening for successful transformants, and bombard the isolated plasmid DNA into onion epidermal cells as described in 2.5-2.9. Additional deletion constructs will also be created,

containing or excluding each of the putative TMDs. These constructs will also be bombarded into onion epidermal cells to investigate how the presence and absence of these regions affect the localization HPL1.

Additionally, the localization of HPL1 may be studied in *A. thaliana* protoplasts. Previous studies (Lee et al., 2019; Lung et al., 2014) have also investigated chloroplast protein translocation using *A. thaliana* protoplasts. Similar experiments may be conducted using HPL1.

Literature Cited

Baker, J. A., Wong, W.-C., Eisenhaber, B., Warwicker, J., & Eisenhaber, F. (2017). Charged residues next to transmembrane regions revisited: "Positive-inside rule" is complemented by the "negative inside depletion/outside enrichment rule." *BMC Biology*, *15*(1), 66. https://doi.org/10.1186/s12915-017-0404-4

Cook, R., Lupette, J., & Benning, C. (2021). The Role of Chloroplast Membrane Lipid Metabolism in Plant Environmental Responses. *Cells*, *10*(3), 706. https://doi.org/10.3390/cells10030706

De Domenico, S., Tsesmetzis, N., Di Sansebastiano, G. P., Hughes, R. K., Casey, R., & Santino, A. (2007). Subcellular localisation of Medicago truncatula 9/13-hydroperoxide lyase reveals a new localisation pattern and activation mechanism for CYP74C enzymes. *BMC Plant Biology*, *7*, 58. https://doi.org/10.1186/1471-2229-7-58

EZ-10 Spin Column Handbook. (2019). Bio Basic Inc. https://www.biobasic.com/us/amfilerating/file/download/file_id/25476/

Farmaki, T., Sanmartín, M., Jiménez, P., Paneque, M., Sanz, C., Vancanneyt, G., León, J., & Sánchez-Serrano, J. J. (2007). Differential distribution of the lipoxygenase pathway enzymes within potato chloroplasts. *Journal of Experimental Botany*, *58*(3), 555–568. https://doi.org/10.1093/jxb/erl230

Fish, M., Nash, D., German, A., Overton, A., Jelokhani-Niaraki, M., Chuong, S. D. X., & Smith, M. D. (2022). New Insights into the Chloroplast Outer Membrane Proteome and Associated Targeting Pathways. *International Journal of Molecular Sciences*, *23*(3), Article 3. https://doi.org/10.3390/ijms23031571

Froehlich, J. E., Wilkerson, C. G., Ray, W. K., McAndrew, R. S., Osteryoung, K. W., Gage, D. A., & Phinney, B. S. (2003). Proteomic Study of the Arabidopsis thaliana Chloroplastic Envelope Membrane Utilizing Alternatives to Traditional Two-Dimensional Electrophoresis. *Journal of Proteome Research*, *2*(4), 413–425. https://doi.org/10.1021/pr034025j

Hallgren, J., Tsirigos, K. D., Pedersen, M. D., Armenteros, J. J. A., Marcatili, P., Nielsen, H., Krogh, A., & Winther, O. (2022). *DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks* (p. 2022.04.08.487609). bioRxiv. https://doi.org/10.1101/2022.04.08.487609

Hanson, M. R., & Sattarzadeh, A. (2013). Trafficking of Proteins through Plastid Stromules[W]. *The Plant Cell*, *25*(8), 2774–2782. https://doi.org/10.1105/tpc.113.112870

Hughes, R. K., De Domenico, S., & Santino, A. (2009). Plant Cytochrome CYP74 Family: Biochemical Features, Endocellular Localisation, Activation Mechanism in Plant Defence and Improvements for Industrial Applications. *ChemBioChem*, *10*(7), 1122–1133. https://doi.org/10.1002/cbic.200800633

Inoue, K. (2007). The Chloroplast Outer Envelope Membrane: The Edge of Light and Excitement. *Journal of Integrative Plant Biology*, *49*(8), 1100. https://doi.org/10.1111/j.1672-9072.2007.00543.x

Inoue, K. (2015). Emerging knowledge of the organelle outer membranes – research snapshots and an updated list of the chloroplast outer envelope proteins. *Frontiers in Plant Science*, *6*. https://doi.org/10.3389/fpls.2015.00278

Jin, Z., Wan, L., Zhang, Y., Li, X., Cao, Y., Liu, H., Fan, S., Cao, D., Wang, Z., Li, X., Pan, J., Dong, M.-Q., Wu, J., & Yan, Z. (2022). Structure of a TOC-TIC supercomplex spanning

two chloroplast envelope membranes. *Cell*, *185*(25), 4788-4800.e13. https://doi.org/10.1016/j.cell.2022.10.030

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices1. *Journal of Molecular Biology*, *292*(2), 195–202. https://doi.org/10.1006/jmbi.1999.3091

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Käll, L., Krogh, A., & Sonnhammer, E. L. L. (2004). A Combined Transmembrane Topology and Signal Peptide Prediction Method. *Journal of Molecular Biology*, *338*(5), 1027–1036. https://doi.org/10.1016/j.jmb.2004.03.016

Kim, J., Na, Y. J., Park, S. J., Baek, S.-H., & Kim, D. H. (2019). Biogenesis of chloroplast outer envelope membrane proteins. *Plant Cell Reports*, *38*(7), 783–792. https://doi.org/10.1007/s00299-019-02381-6

Lee, D. W., Lee, S., Lee, J., Woo, S., Razzak, M. A., Vitale, A., & Hwang, I. (2019). Molecular Mechanism of the Specificity of Protein Import into Chloroplasts and Mitochondria in Plant Cells. *Molecular Plant*, *12*(7), 951–966. https://doi.org/10.1016/j.molp.2019.03.003

Lee, J., 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. https://doi.org/10.1105/tpc.110.082230

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. *Frontiers in Plant Science*, *5*. https://doi.org/10.3389/fpls.2014.00722

Mita, G., Quarta, A., Fasano, P., De Paolis, A., Di Sansebastiano, G. P., Perrotta, C., Iannacone, R., Belfield, E., Hughes, R., Tsesmetzis, N., Casey, R., & Santino, A. (2005). Molecular cloning and characterization of an almond 9-hydroperoxide lyase, a new CYP74 targeted to lipid bodies*. *Journal of Experimental Botany*, *56*(419), 2321–2333. https://doi.org/10.1093/jxb/eri225

Moog, D. (2019). Higher Complexity Requires Higher Accuracy: Tail-Anchored Protein Targeting to the Outer Envelope Membrane of Plant Plastids via a Specific C-Terminal Motif. *Plant and Cell Physiology*, *60*(3), 489–491. https://doi.org/10.1093/pcp/pcz021

Nilsson, A. K., Fahlberg, P., Johansson, O. N., Hamberg, M., Andersson, M. X., & Ellerström, M. (2016). The activity of HYDROPEROXIDE LYASE 1 regulates accumulation of galactolipids containing 12-oxo-phytodienoic acid in Arabidopsis. *Journal of Experimental Botany*, *67*(17), 5133–5144. https://doi.org/10.1093/jxb/erw278

Nugent, T., & Jones, D. T. (2009). Transmembrane protein topology prediction using support vector machines. *BMC Bioinformatics*, *10*, 159. https://doi.org/10.1186/1471-2105-10-159

Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G., & Cormier, M. J. (1992). Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene*, *111*(2), 229–233. https://doi.org/10.1016/0378-1119(92)90691-H

Simm, S., Papasotiriou, D. G., Ibrahim, M., Leisegang, M. S., Müller, B., Schorge, T., Karas, M., Mirus, O., Sommer, M. S., & Schleiff, E. (2013). Defining the Core Proteome of the Chloroplast Envelope Membranes. *Frontiers in Plant Science*, *4*, 11. https://doi.org/10.3389/fpls.2013.00011

Snapp, E. (2005). Design and Use of Fluorescent Fusion Proteins in Cell Biology. *Current Protocols in Cell Biology / Editorial Board, Juan S. Bonifacino ... [et Al.]*, *CHAPTER*, Unit-21.4. https://doi.org/10.1002/0471143030.cb2104s27

Song, Y., Feng, L., Alyafei, M. A. M., Jaleel, A., & Ren, M. (2021). Function of Chloroplasts in Plant Stress Responses. *International Journal of Molecular Sciences*, *22*(24), Article 24. https://doi.org/10.3390/ijms222413464

Stegemann, S., Hartmann, S., Ruf, S., & Bock, R. (2003). High-frequency gene transfer from the chloroplast genome to the nucleus. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(15), 8828–8833. https://doi.org/10.1073/pnas.1430924100

The PyMOL Molecular Graphics System (Version 3.1.0). (n.d.). [Computer software]. Schrödinger, LLC.

van't Hof, R., van Klompenburg, W., Pilon, M., Kozubek, A., de Korte-Kool, G., Demel, R. A., Weisbeek, P. J., & de Kruijff, B. (1993). The transit sequence mediates the specific interaction of the precursor of ferredoxin with chloroplast envelope membrane lipids.

Journal of Biological Chemistry, *268*(6), 4037–4042. https://doi.org/10.1016/S0021-9258(18)53576-6

Varadi, M., Bertoni, D., Magana, P., Paramval, U., Pidruchna, I., Radhakrishnan, M., Tsenkov, M., Nair, S., Mirdita, M., Yeo, J., Kovalevskiy, O., Tunyasuvunakool, K., Laydon, A., Žídek, A., Tomlinson, H., Hariharan, D., Abrahamson, J., Green, T., Jumper, J., … Velankar, S. (2024). AlphaFold Protein Structure Database in 2024: Providing structure coverage for over 214 million protein sequences. *Nucleic Acids Research*, *52*(D1), D368–D375. https://doi.org/10.1093/nar/gkad1011

Appendix 1: Primer sequences

Table 1: List of primers and sequences used in HPL1::EGFP constructs

Primer	**Sequence (5' to 3')**
HPLF1	ATGGATCTAGTTGATAAAAGAG
HPLF2	GATGGATCTAGTTGATAAAAGAG
HPLF3	ATGGAACCGGAGGAATTTAAAC
HPLF4	GGAACCGGAGGAATTTAAACC
HPLR1	TTTAGCTTTAACAACAGC
HPLR2	TTATTTAGCTTTAACAACAGC
HPLR3	AAGCTTTTTGTAATTTCCGGC
HPLR4	TTAAAGCTTTTTGTAATTTCCGG