INVESTIGATING LIGHT-INDUCED *psbA* TRANSLATION IN CHLOROPLASTS

by

SONJA M. LJUNGDAHL

A THESIS

Presented to the Department of Biology
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Master of Science

September 2019

THESIS APPROVAL PAGE

Student: Sonja M. Ljungdahl

Title: Investigating Light-Induced psbA Translation in Chloroplasts

This thesis has been accepted and approved in partial fulfillment of the requirements for the Master of Science degree in the Department of Biology by:

Alice Barkan	Chairperson
Diane Hawley	Member
Jeff McKnight	Member

and

Janet Woodruff-Borden	Vice Provost and Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded September 2019

THESIS ABSTRACT

Sonja M. Ljungdahl

Master of Science

Department of Biology

September 2019

Title: Investigating Light-Induced psbA Translation in Chloroplasts

Light, while necessary for plants, can cause photo-oxidative damage. Adaptations to fluctuating light conditions optimize photosynthetic yield and minimize light-induced damage. Light-regulated synthesis of the chloroplast gene *psbA* and its protein product D1 is at the core of these responses. Shifting light intensity regulates D1 synthesis at the level of translation. This thesis investigates specific proteins we hypothesized mediate the effects of light on D1 synthesis. Our experiments on TPJ1, a maize mutant lacking one of these proteins, showed that TPJ1 does not influence *psbA* translation. Instead, our results show it is required for the translation of the chloroplast *psbJ* mRNA. In addition, I elucidated the biochemical interactions between two known *psbA* translational activators, OHP2 and HCF244, by identifying a short segment of OHP2 that is sufficient for its interaction with HCF244. This thesis includes published, co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Sonja M. Ljungdahl

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
Pacific University, Eugene

DEGREES AWARDED:

Master of Science, Biology, 2019, University of Oregon
Master of Arts, Teaching, 2010, Pacific University
Bachelor of Science, Biology, 2008, University of Oregon

AREAS OF SPECIAL INTEREST:

Molecular Biology

PROFESSIONAL EXPERIENCE:

High School Science Teacher, Springfield School District, August 2013-Present

Graduate Teaching Employee, University of Oregon, March 2017-June 2017

Teacher Leader, Stewardship Program in Aquatic Restoration and Conservation, September 2012-May2013

High School Science Teacher, McKenzie School District, August 2011-August 2013

GRANTS, AWARDS, AND HONORS:

M.J. Murdock Charitable Trust, Supplement to PIS-2016326, Clarifying the Connection Between Phenotype, Protein, and Genotype, Springfield High School, 2019.

Springfield Education Foundation Innovative Educator Grant, 21st Century Biology Skills, Springfield High School, 2019

M.J. Murdock Charitable Trust, Partners in Science Grant-2016326, University of Oregon, 2017.

Springfield Education Foundation Innovative Educator Grant, Science Mystery Video Project, Springfield High School, 2014

Oregon Department of Education STEAM on grant, Project-based science curriculum, Springfield High School, 2013.

Oregon Department of Education STEAM-on grant, After school science program with Vernier Probeware, McKenzie River High School, 2012.

PUBLICATIONS:

Williams-Carrier, R., Brewster, C., Belcher, S. E., Rojas, M., Chotewutmontri, P., **Ljungdahl, S**., & Barkan, A. (2019) The Arabidopsis pentatricopeptide repeat protein LPE1 and its maize ortholog are required for translation of the chloroplast psbJ RNA. *Plant Journal,* **99**, 56–66. https://doi.org/10.1111/tpj.14308

ACKNOWLEDGMENTS

I sincerely appreciate Professor Alice Barkan for her assistance in the preparation of this manuscript and support throughout my time in her lab. Margarita Rojas, Rosalind Carrier, and Kenneth Watkins were supportive, friendly, and shared their expertise with me. Carolyn Brewster was a punchy laboratory companion and collaborator, and I am grateful to have had her as an ally for the first year of these investigations. Non Chotewutmontri was particularly generous and patient in mentoring and aiding me in serial dilutions for quantitative analysis of yeast two hybrid colonies. I am grateful for the assistance of Professor Diane Hawley, who served as a member of my committee and tutored me in effective techniques for qualitatively analyzing yeast two hybrid systems, in particular x-gal filter lift assays to detect β-galactosidase activity. Appreciation to Assistant Professor Jeff McKnight for serving as a member on my committee. Thanks to my partner, David Levin, for consulting on grammar and style. This work was funded by the M.J.Murdock Charitable Trust, Partners in Science Grant-2016326, grants IOS-1339130 and MCB-1616016 to A.B. from the United States National Science Foundation, and by the University of Oregon Peter O'Day Fellowship to C.B. We wish to thank the UniformMu project and Maize Genetics Cooperative for providing the *tpj1* mutants, the ABRC for providing *lpe1, hcf107*, and *hcf173* mutants, and Dr. Hongbin Wang (Sun Yat-sen University) for providing important reagents (*lpe1-3* seed, LPE1 expression construct, a line expressing FLAG-tagged LPE1) and for helpful information about these materials.

To my past, present, and future students for inspiring me in all ways to be better.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

The chloroplast is the photosynthetic organelle present in plants and algae. It is believed to be the product of an endosymbiotic event, occurring over a billion years ago, in which a cyanobacterium was integrated into a single-celled eukaryotic host. Over time, and through the process of co-evolution, genes originally in the endosymbiont relocated into the nuclear genome, so that today's chloroplast consists of both nuclear and chloroplast-encoded proteins. The majority of a chloroplast's approximately 3,000 proteins are encoded in the nucleus and synthesized in the cytosol before they are shuttled into the chloroplast. Only ~100 chloroplast proteins are encoded by the chloroplast's genome; these proteins contribute to the chloroplast's photosynthetic apparatus and gene expression machinery and were likely present in the original cyanobacterium. Each major photosynthetic enzyme complex (Photosystem II, Cytochrome b_6f, Photosystem I, ATP synthase, and rubisco) consists of both chloroplast and nuclear gene products (Sun and Zerges, 2015; Zoschke and Bock, 2018).

While chloroplast genetic systems have similar features to those in bacteria, e.g., circular DNA, polycistronic transcription, and a population of ribosomes and ribonucleases similar to their bacterial ancestors, they also possess many derived traits. For example, chloroplast gene expression is predominantly orchestrated at the post-transcriptional level, distinct from the transcriptional controls observed in bacteria (Rochaix, 2006). Chloroplast translation is carried out by ribosomes and general translation factors derived from those in bacteria, but is regulated by nucleus-encoded

proteins that evolved post-endosymbiosis. Many of these belong to the pentatricopeptide repeat (PPR) family, a group that is theorized to have emerged during the coevolution of the ancient endosymbiont and its host (Barkan, 2011). PPR proteins are composed of tandem alpha-helical repeats that act as sequence-specific binding sites for particular chloroplast RNAs. There are approximately 400 PPR proteins in terrestrial plants today and they have been shown to be involved in every aspect of chloroplast gene expression: transcription, RNA stabilization and cleavage, splicing and editing of RNAs, and mediating translation (Barkan and Small, 2014). PPRs have been suggested to activate translation by binding to areas on mRNA which inhibit ribosome binding by folding into a hairpin. When the PPR is bound to the mRNA it is unable to conform to the hairpin shape, which makes the ribosome binding site accessible to the ribosome. This large protein family, while absent in the chloroplasts' cyanobacterial ancestors, is now integral to its function (Barkan and Small, 2014).

Photosynthesis harvests light energy from the sun through a series of enzyme-mediated reactions that take place in the thylakoid membrane and the stroma of the chloroplast. The presence of light not only drives photosynthesis, it also causes damage to the photosynthetic machinery. As environmental light conditions are fluctuating, plants must simultaneously minimize damage while optimizing the production of photosynthetic products. Plants must be able to respond rapidly to these fluctuations; therefore, many of these responses occur through the regulation of both nuclear and chloroplast genes (Rochaix, 2013).

Little is known about how light affects chloroplast translation; classical translational analysis techniques are limited in their ability to elucidate the translation

rates of proteins with low expression levels, or those with high turnover rates (Zoschke and Bock, 2018). In addition, many early experiments on chloroplast translation were performed on plants that had been de-etiolated, or used chloroplasts that had been isolated; these techniques obfuscate analysis of data by introducing secondary effects (Chotewutmontri and Barkan, 2018).

More recent data, using ribosome profiling techniques (Ingolia et al., 2009), has provided a genome-wide view of chloroplast translation (Chotewutmontri and Barkan 2016; Chotewutmontri and Barkan, 2018). Ribosome profiling reveals *in vivo* ribosome occupancy on mRNA transcripts at the time the organism is flash-frozen. Recent ribosome profiling experiments in *Zea mays* and *Arabidopsis* show that the ribosome occupancy on most chloroplast RNAs does not change significantly after shifting plants from light to dark or dark to light. However, there was one major exception: the *psbA* mRNA rapidly gains ribosomes after shifting plants to the light and loses them again after shifting to the dark (Chotewutmontri and Barkan, 2018, Figure 1.1).

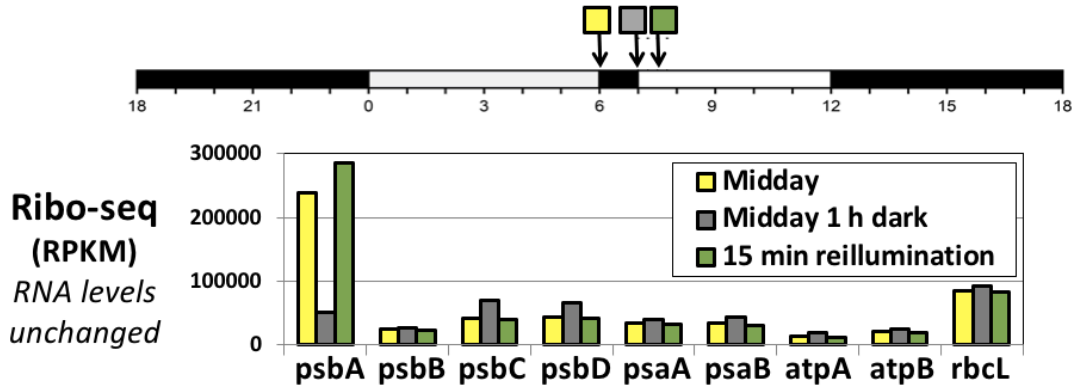


Figure 1.1 Dynamic changes in *psbA* ribosome occupancy following midday shifts from light-to-dark and vice versa. RPKM is the normalized abundance of ribosome footprints on the indicated gene. Only a subset of chloroplast genes is shown; ribosome occupancy on those not shown did not change substantively.

The psbA gene encodes the D1 protein, a reaction center subunit of Photosystem II (PSII). Photosystem II is a multi-subunit enzyme complex that is embedded in the thylakoid membrane and is responsible for harvesting energy from photons of light. Molecular processes at PSII result in the extraction of electrons from water molecules (releasing molecular oxygen) and transfer of those electrons down an electron transport chain, which produce a proton gradient that drives ATP synthesis and reduced NADPH that provides reducing power for carbon fixation. The D1 protein, a chlorophyll binding protein in the reaction center of the complex, suffers damage as a consequence of proximity to highly oxidizing species during these events. When D1 is damaged, photosynthesis is halted, thus this protein needs to be continuously replaced when illuminated (Sun and Zerges, 2015). This need for constant replacement during photosynthesis is concurrent with the increased translation of *psbA* RNA in the light, (Chotewutmontri and Barkan, 2018). The specific molecular actions that occur to drive D1's replacement are largely unrevealed. Zerges (2015) writes:

> "Nothing is known about how the *psbA* mRNA is specifically selected for translation to repair photodamaged PSII complexes, despite its importance to chloroplast biology and the productivity of agricultural plants."

This thesis makes progress towards elucidating the molecular mechanisms that underpin the rapid gain of ribosomes on *psbA* mRNA in response to light. Two central problems that this research addresses are: determining the proximal translational regulators of at the *psbA* RNA and understanding how those specific regulators are interacting with each other.

In Chapter 2, in an effort to further elucidate translational regulation of *psbA* RNA we performed ribosome profiling experiments on a maize mutant lacking the PPR protein LPE1, which had previously been conjectured to activate the translation of *psbA* RNA in *Arabidopsis.* The work in this chapter has been previously published with the co-authors Rosalind Williams-Carrier, Carolyn Brewster, Susan E. Belcher, Margarita Rojas, Prakitchai Chotewutmontri, and Alice Barkan.

Chapter 3 investigates the relationships between proteins required for the recruitment of ribosomes to psbA RNA at the thylakoid membrane. I use a yeast two-hybrid system to elucidate interactions among several proteins that form a complex required for PSII repair (HCF 244, OHP1, OHP2, HCF173, D1 stromal loop), several of which are required for *psbA* translation (HCF244, OHP1, OHP2, HCF173). These experiments were performed to help test a model in which the presence of D1 in this complex inhibits the activators, and that D1 degradation relieves this inhibition.

BRIDGE I

To elucidate mechanisms that mediate the effects of light on D1 synthesis, we examined the role of nucleus-encoded proteins that we hypothesized participated in this process. LPE1 came to our attention as a possible regulator of D1 translation based on the fact that it is known to localize to chloroplasts, it is a PPR protein and so was expected to bind RNA, and it exhibits an expression pattern that matches the expectation for genes involved in regulating D1 synthesis (Li et al. 2010; Majeran et al., 2010). In addition, the publication by Jin et al. (2018) concluded that LPE1 activates the translation of *psbA* RNA in *Arabidopsis*. In Chapter II we performed ribosome profiling experiments on mutants lacking LPE1, which revealed LPE1s role as a translational activator of two short chloroplast ORFs (*psbJ* and *psbN)*. Our findings warrant revision of LPE1's proposed function in *psbA* translation.

CHAPTER II

TITLE: THE ARABIDOPSIS PENTATRICOPEPTIDE REPEAT PROTEIN LPE1 AND ITS MAIZE ORTHOLOG ARE REQUIRED FOR TRANSLATION OF THE CHLOROPLAST PSBJ RNA

From Williams-Carrier, R., Brewster, C., Belcher, S. E., Rojas, M., Chotewutmontri, P., Ljungdahl, S., & Barkan, A. (2019) The Arabidopsis pentatricopeptide repeat protein LPE1 and its maize ortholog are required for translation of the chloroplast psbJ RNA. *Plant Journal,* 99, 56–66. https://doi.org/10.1111/tpj.14308

The experimental work in this chapter was performed either by me or by R. Williams-Carrier, C. Brewster, S.E. Belcher, M. Rojas, P.Chotewutmontri. Specifically, I contributed to the identification of maize mutants lacking *tpj1* (an orthologue to *lpe1)* and to the analysis of their mutant phenotypes. I used PCR techniques to identify mutant plants lacking *tpj1*. I performed initial TPJ1 immunoblot assays on mutant tissue to uncover deficiencies in the photosynthetic subunits suggesting its function in PSII biogenesis. I contributed to the ribo-seq analysis of maize *tpj1* mutants by tending and harvesting the plants, pooling the tissue, and participating in the ribosome profiling assays. Ribosome profiling experiments were performed under the tutelage of Roz Carrier. Carolyn Brewster performed Northern Blot assays, aided in ribo-seq analysis, tended plants, and performed final immunoblot assays.

AUTHORS:

Sonja M. Ljungdahl, R. Williams-Carrier, C. Brewster, S.E. Belcher, M. Rojas, P. Chotewutmontri, Alice Barkan.

INTRODUCTION

Expression of the chloroplast genome requires hundreds of nucleus-encoded proteins, the majority of which act at the post-transcriptional level to promote the processing, stability, or translation of specific RNAs (Barkan 2011). Most of these belong to protein families whose members function exclusively (or nearly so) in the expression of organellar genes. The pentatricopeptide repeat (PPR) protein family is a prominent example of this phenomenon (Barkan and Small 2014). PPR proteins are defined by degenerate tandem repeats of approximately 35 amino acids that adopt a helix-turn-helix fold (Small and Peeters 2000). Consecutive repeats stack to form an elongated superhelix whose surface binds single-stranded RNA (Yin *et al.* 2013). PPR proteins are found in all eukaryotes but the size of the PPR family is highly variable among species. They comprise one of the largest protein families in plants, where they have diversified into several subfamilies (Cheng *et al.* 2016). Proteins that consist primarily of canonical PPR motifs, referred to as P-type proteins, generally promote the splicing, translation or stability of specific organellar RNAs, whereas proteins with variant repeat tracts denoted PLS generally specify sites of organellar RNA editing. In either case, RNA recognition by PPR tracts involves modular 1-repeat, 1-nucleotide interactions whose specificity is influenced by the identities of amino acids at several positions in each repeat (Barkan *et al.* 2012; Takenaka *et al.* 2013; Yagi *et al.* 2013).

We have been analyzing PPR proteins that localize to chloroplasts in order to elucidate the basis for the sequence-selectivity of PPR tracts and the functional consequences of PPR-RNA partnerships. Our attention was drawn to the P-type PPR protein encoded by maize gene GRMZM2G056116 due to its unusual expression pattern:

whereas mRNA levels for the majority of chloroplast PPR proteins peak in a "chloroplast biogenesis" zone toward the base of the seedling leaf blade, mRNA from this gene peaks at the leaf tip - the site of mature chloroplasts (see http://bar.utoronto.ca/efp_maize/cgi-bin/efpWeb.cgi) (Li *et al.* 2010; Majeran *et al.* 2010). This expression pattern suggested a role in maintaining expression of one or more of the handful of chloroplast genes whose expression peaks at the leaf tip, most of which encode subunits of Photosystem II (PSII) (Chotewutmontri and Barkan 2016). In accord with this prediction, the Arabidopsis ortholog (AT3G46610), denoted LPE1, was recently reported to be necessary for PSII accumulation (Jin *et al.* 2018).

The PSII deficiency in *lpe1* mutants was attributed to a defect in translation of the chloroplast *psbA* mRNA (Jin *et al.* 2018). This conclusion was based on results from polysome and pulse-labeling assays. However, polysome assays have limitations when assessing translation of short open reading frames (ORFs) on polycistronic RNAs, and pulse-labeling assays have limitations when assessing synthesis of small proteins (including several PSII subunits) and proteins that are rapidly degraded when their assembly is disrupted (such as the *psbA* gene product). A more recently developed method called ribosome profiling provides a genome-wide, high-resolution, and quantitative snapshot of ribosome positions on mRNA, and can detect gene expression defects that are missed by polysome and pulse-labeling assays (Rojas *et al.* 2018; Zoschke *et al.* 2013). Ribosome profiling uses deep-sequencing or high-resolution microarrays to quantitatively map ribosome footprints: short mRNA segments that are protected by ribosomes from ribonucleases. We revisited the function of LPE1 by using ribosome profiling-by-sequencing (Ribo-seq) to analyze chloroplast gene expression in

lpe1 mutants in maize and Arabidopsis. Our results show that the maize LPE1 ortholog strongly stimulates translation of the chloroplast *psbJ* ORF, and that this is the sole basis for the PSII defect in the maize mutant. Because the abbreviation LPE1 is already in use for a different gene in maize, we named the maize ortholog Translation of *psbJ* 1 (TPJ1) to reflect this function. In addition, our data reveal previously undetected roles for Arabidopsis LPE1 in translation of the chloroplast *psbJ* and *psbN* ORFs, and provide evidence that the modest decrease in ribosome occupancy on the *psbA* mRNA in *lpe1* mutants is a secondary effect of their PSII deficiency. Taken together with *in vitro* RNA binding data, these results provide strong evidence that the loss of PSII in Arabidopsis *lpe1* mutants results from defects in the expression of *psbJ* and *psbN*.

RESULTS

Maize TPJ1 functions in PSII biogenesis

LPE1 (AT3G46610) and TPJ1 (GRMZM2G056116) consist of 14 PPR motifs preceded by a predicted chloroplast transit peptide (Appendix A). LPE1 has been shown to localize to chloroplasts (Jin *et al.* 2018) and TPJ1 was detected in the chloroplast nucleoid proteome (Majeran *et al.* 2012). To analyze TPJ1 function in maize, we analyzed the phenotypes of two *Mu* transposon insertion alleles. The *tpj1-1* insertion maps to the 5'-untranslated region (UTR) and is expected to be a hypomorphic allele, whereas the insertion in *tpj1-2* interrupts the protein coding region (Fig. 2.1a). Plants that are homozygous for *tpj1-2* exhibit a mild elevation in chlorophyll fluorescence (Fig. 2.1b), suggesting a photosynthetic defect. The mutants survive past the seedling stage when grown in soil, albeit with reduced vigor (Fig. 2.1b right). Arabidopsis *lpe1* null mutants

likewise survive past the seedling stage in soil (Jin *et al.* 2018). Therefore, TPJ1 and LPE1 are not essential for photosynthesis.

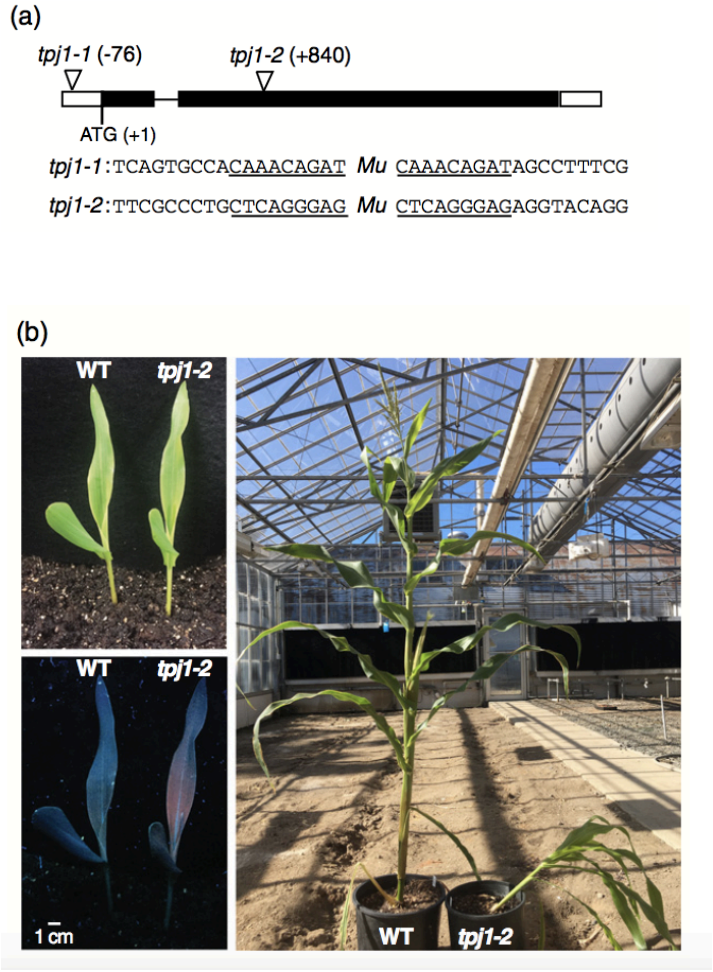


Figure 2.1. Overview of maize *tpj1* mutant alleles.
(a) Insertion sites. TPJ1 is encoded by maize locus GRMZM2G056116 (B73 RefGen_v3) or Zm00001d036489 (B73 RefGen_v4). The gene contains one intron (thin line). The distance in nucleotides of each *Mu* insertion from the start codon is indicated. The sequences of the insertion sites are shown below, with the 9-bp target site duplication underlined.
(b) Mutant phenotypes. The seedlings to the left were grown for 10 days in soil. The same seedlings were illuminated either with white light (top) or UV light (bottom). The phenotype of mature plants (roughly 8 weeks old) is shown to the right.

Immunoblot analysis showed that *tpj1-2* mutants have reduced levels of core PSII subunits (Fig. 2.2). The magnitude of this deficiency differed among subunits: PsbC was strongly reduced (~10% of wild-type levels) whereas PsbA, PsbB, and PsbD were

reduced roughly 4-fold. Subunits of the ATP synthase (AtpA), Photosystem I (PsaD), and the cytochrome b_6f complex (PetD) accumulated to approximately 50% of normal levels whereas the large subunit of Rubisco (RbcL) was unaffected. Heteroallelic progeny of a complementation test cross (*tpj1-1/tpj1-2)* showed a weaker phenotype, as expected due to the fact that the *tpj1-1* insertion maps a considerable distance upstream from the start codon (Fig. 2.1a). Nonetheless, the PsbC and PsbD subunits of PSII were clearly reduced. The results for *tpj1-2* are similar to those reported for Arabidopsis *lpe1* except that PsbA was more strongly reduced than PsbC in *lpe1* mutants (Jin *et al.* 2018).

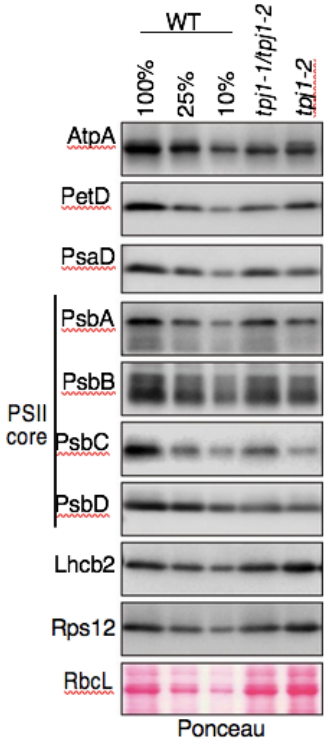


Figure 2.2 Immunoblot analysis of components of the photosynthetic apparatus in maize *tpj1* mutants. Seedling leaf extracts (5 µg protein or the indicated dilutions) from plants of the indicated genotype were fractionated by SDS-PAGE and analyzed by immunoblotting. Replicate blots were probed with antibodies specific for the indicated proteins. An excerpt of one of the replicate blots stained with Ponceau S (bottom) shows the abundance of the large subunit of Rubisco (RbcL). AtpA is a subunit of the thylakoid ATP synthase, PetD is a subunit of the cytochrome b_6f complex, PsaD is a subunit of PSI, Lhcb2 is a subunit of the peripheral light harvesting complex of PSII, and Rps12 is found in the chloroplast 30S ribosomal subunit.

TPJ1 is required for *psbJ* translation

Given that PPR proteins generally function in organellar gene expression, it seemed likely that TPJ1 promotes PSII accumulation by stimulating expression of one or more chloroplast gene encoding a PSII subunit. To detect chloroplast genes whose expression relies on TPJ1, we analyzed *tpj1-2* mutant seedlings by ribosome profiling. The absence of ribosome footprints from the *tpj1* gene in the *tpj1-2* mutant confirmed that *tpj1-2* is a null allele Appendix B). In three replicate experiments, we observed a dramatic loss of *psbJ* ribosome footprints in the *tpj1-2* mutant (Fig. 2.3a, Appendix C, Appendix F). The *psbJ* gene resides in the *psbE-psbF-psbL-psbJ* transcription unit. A sequence coverage plot (Fig. 2.3b) shows that ribosome density along the *psbJ* ORF is drastically reduced in the *tpj1* mutant, in comparison with the three ORFs upstream. Data from one replicate suggested a defect in translating the third exon of *rps12* (Appendix C), but this was not reproducible. This spurious result was likely due to sampling error, as this exon encodes only nine amino acids and is represented by a small number of Ribo-seq reads. The abundance of ribosome footprints mapping to all other chloroplast genes was similar between the *tpj1-2* and wild-type samples.

To determine whether the loss of ribosome footprints from *psbJ* results from an mRNA defect, we analyzed *psbJ* transcripts by RNA gel blot hybridization (Fig. 2.3c). A *psbJ*-specific probe detected only the tetracistronic *psbE-psbF-psbL-psbJ* transcript, and this RNA was found at normal levels in the *tpj1-2* mutant. Therefore, the loss of ribosome footprints from *psbJ* results from a defect in *psbJ* translation. Motivated by the spurious *rps12* result in the Ribo-seq data (see above), we probed a replicate blot to detect the second and third exons of *rps12* mRNA (Fig. 2.3c). The results show that a major transcript at approximately 1200 nucleotides is reduced in the mutant whereas two

smaller transcripts (~1000 and 450 nucleotides) were detected only in the mutant. The nature of the RNA defect is unclear, although the transcript pattern is inconsistent with a splicing defect. In any case, RPS12 protein accumulates normally in *tpj1-2* mutants (Fig. 2.2). As the effects of TPJ1 on *rps12* mRNA have no apparent impact on Rps12 abundance or function, we did not investigate this further.

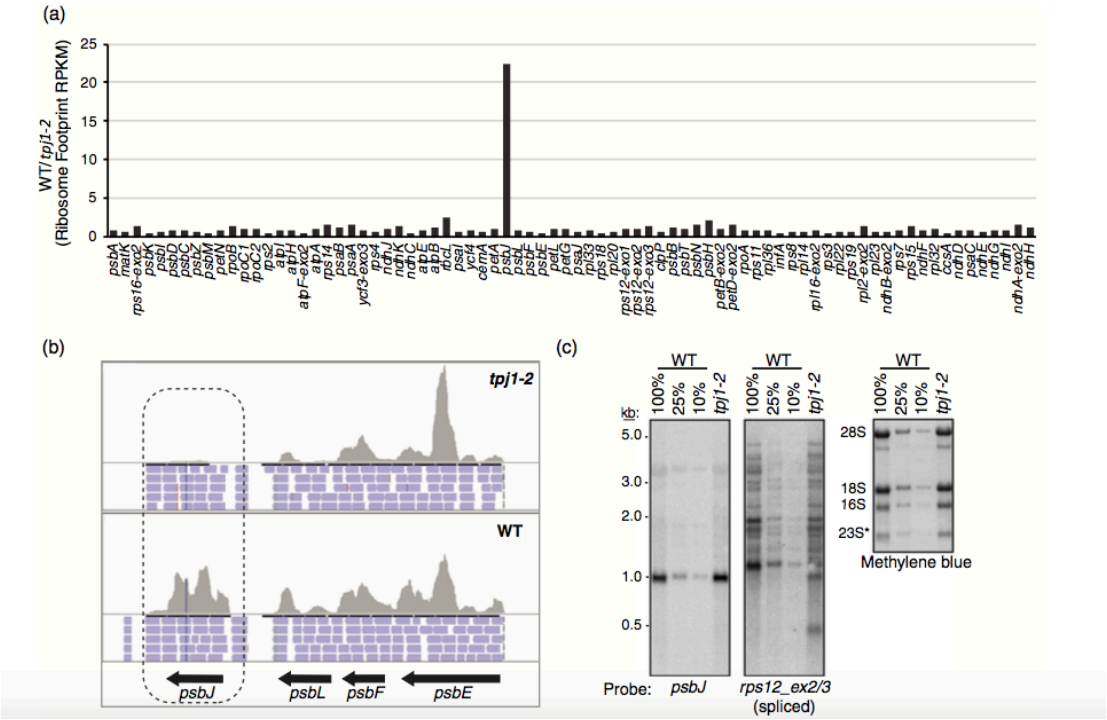


Figure 2.3. Ribo-seq analysis of chloroplast gene expression in maize tpj1-2 mutants.
(a) Ratio of normalized ribosome footprint abundance in the wild-type relative to the tpj1-2 mutant for each chloroplast gene. Values were normalized to the number of reads mapping to chloroplast ORFs. Two replicate experiments are shown in Figure S3. The read count and RPKM values for all chloroplast genes are shown in Table S2.
(b) Screen capture from the Integrated Genome Viewer (IGV) showing the absence of ribosome footprints on the psbJ ORF in the tpj1-2 mutant. The psbE– psbF–psbL–psbJ transcription unit is shown.
(c) RNA gel blot hybridization analysis of psbJ and rps12 transcripts in tpj1-2 mutants. Replicate panels from the same gel were hybridized with probes specific for psbJ or exons 2 and 3 of rps12. One blot was stained with methylene blue to illustrate the abundance of rRNAs as a loading control. The bands marked 28S and 18S are cytosolic rRNAs, and the bands marked 16S and 23S* are chloroplast rRNAs.

Ribo-seq analysis of Arabidopsis *lpe1* mutants revealed defects in *psbJ* and *psbN* expression.

The report that the TPJ1 ortholog in Arabidopsis (LPE1) is required specifically for *psbA* translation (Jin *et al.* 2018) contrasts with our findings for TPJ1, indicating either that the function is not conserved or that the defect in *psbJ* translation was missed in the Arabidopsis study. To resolve this issue, we used Ribo-seq to assay chloroplast gene expression in two *lpe1* exon insertion alleles: the *lpe1-3* allele described previously (Jin *et al.* 2018) and a new allele, *lpe1-4* (Appendix D). Both alleles condition a severe loss of the D1 and D2 reaction center subunits of PSII (Appendix D), consistent with the prior

report. In addition, we observed a several-fold decrease in the abundance of the PsaD subunit of photosystem I (PSI). An effect on PSI was not detected previously (Jin *et al.* 2018). The basis for this difference is unclear but may result from differences in growth conditions or developmental stage.

In an initial Ribo-seq experiment, we compared *lpe1-4* mutants to siblings of normal phenotype (Fig. 2.4); the mutant and normal plants were the same age but the mutant plants were less developed due to their photosynthetic defect (Appendix D). The ratio of normalized ribosome footprint abundance for each chloroplast gene in the wild-type relative to the mutant revealed *psbJ* as having the largest expression defect (Fig. 2.4a); the 10-fold difference between the wild-type and mutant is similar to what we observed for *psbJ* in the maize *tpj1* mutant. The data show, in addition, a moderate reduction in ribosome footprints from the *psbA* and *psbN* ORFs in the *lpe1-4* mutant.

We then used Ribo-seq to analyze chloroplast gene expression in the *lpe1-3* allele described previously (Jin *et al.* 2018). Because the *lpe1-3* plants varied in size, we were able to select mutant and Col-0 plants at a similar developmental stage (Appendix D). The results showed an approximately 12-fold loss of ribosome footprints from *psbJ* (Fig. 2.4b), validating the importance of LPE1 for *psbJ* expression. Interestingly, the loss of ribosome footprints from *psbN* was much more severe in the *lpe1-3* experiment than in the *lpe1-4* analysis, whereas the *psbA* defect was milder. Sequence coverage plots show a strong decrease in ribosome footprints along the entire *psbJ* and *psbN* ORFs in *lpe1-3* mutants (Fig. 2.4c). RNA gel blot hybridizations showed that the *psbJ, psbN,* and *psbA* mRNAs were found at near normal levels in *lpe1* mutants (Fig. 2.4d). Taken together, these results show that LPE1, like its maize ortholog, strongly stimulates translation of *psbJ.* In addition, the results suggest that LPE1 activates *psbN* translation. However, the large difference in the magnitude of the *psbN* defect in *lpe1-3* and *lpe1-4* alleles is puzzling. It may be that the *lpe1-4* allele is hypomorphic, as its insertion maps near the 3'-end of the *LPE1* gene within sequences encoding the final PPR motif (Appendix A).

Ribo-seq analysis of *hcf107* mutants suggests that the modest decrease in *psbA* ribosome occupancy in *lpe1* mutants is a secondary effect of their PSII deficiency. In past work, we observed a small decrease in *psbA* ribosome occupancy in ribosome profiling analyses of various non-photosynthetic mutants (e.g. Shen *et al.* 2017; Zoschke *et al.* 2013; Zoschke *et al.* 2016). Therefore, we considered the possibility that the small reduction in *psbA* translation in *lpe1* mutants is a secondary effect of their PSII defect. To address this possibility, we used Ribo-seq to analyze Arabidopsis *hcf107* mutants, which

lack PSII due to defects in the expression of the chloroplast *psbH* and *psbB* genes (Felder *et al.* 2001). We used an *hcf107* insertion mutant (see Appendix D), and grew them in parallel with the *lpe1-3* plants to which they were compared.

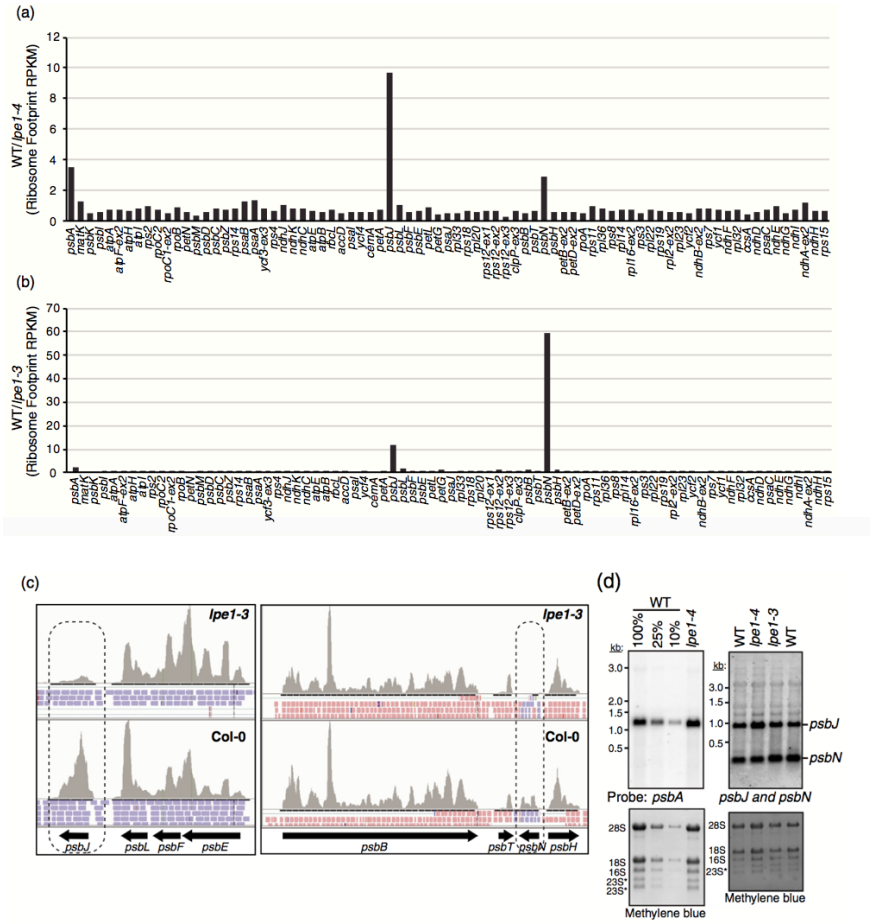


Figure 2.4. Ribo-seq analysis of chloroplast gene expression in Arabidopsis lpe1 mutants. The read count and RPKM values for all chloroplast genes in these experiments are shown in Table S2.
(a) Ratio of normalized ribosome footprint abundance in the wild-type relative to the lpe1-4 mutant for each chloroplast gene. Values were normalized to the number of reads mapping to chloroplast ORFs. Mutants were compared with siblings of normal phenotype. Tissue was harvested 18 days post-vernalization; mutant seedlings were at an earlier developmen- tal stage due to their photosynthetic defect (see Figure S4a).
(b) Ratio of normalized ribosome footprint abundance in the wild-type relative to the lpe1-3 mutant for each chloroplast gene. Mutants were compared to Col-0 grown in parallel and harvested 20 days post-vernalization. Mutant and Col-0 plants selected for this analysis were at a similar developmental stage (see Figure S4a).
(c) Screen captures from the Integrated Genome Viewer (IGV) showing the loss of ribosome footprints from the psbJ and psbN ORFs in the lpe1-3 mutant.
(**d) RNA gel blot hybridization analysis of psbJ, psbA, and psbN RNA in lpe1 mutants.** The psbJ and psbN data come from consecutive probings of the same blot; each probe detects just the one transcript that is marked. Blots were stained with methylene blue to illustrate the abundance of rRNAs as a loading

control (bottom). The bands marked 28S and 18S are cytosolic rRNAs, and the bands marked 16S and 23S* are chloroplast rRNAs.

A scatter plot of RPKM values for all chloroplast genes in *hcf107* mutants relative to their normal siblings (Fig. 2.5a) revealed strongly reduced expression of *psbH* and a moderate reduction in *psbB* expression, validating prior conclusions about HCF107 (Felder *et al.* 2001). In addition, values for *psbA* were slightly lower in *hcf107* mutants than in the wild-type. An analogous display of the *lpe1-3* data shows a similar effect on *psbA* (Fig. 2.5b). When the *hcf107* and *lpe1-3* data are compared directly (Fig. 2.5c), the loss of *psbN* and *psbJ* expression in *lpe1-3,* and the loss of *psbB* and *psbH* expression in *hcf107* are unambiguous, but the difference at *psbA* is negligible. Furthermore, the distribution of Ribo-seq reads along the *psbA* ORF in *hcf107* and *lpe1-3* mutants is very similar (Fig. 2.5d). These results strongly suggest that the small reduction in ribosomes on *psbA* mRNA in *lpe1* mutants is a secondary effect of their PSII deficiency.

Ribo-seq analysis shows that HCF173 is required specifically to recruit ribosomes to *psbA* mRNA.

LPE1 was proposed to activate *psbA* translation by recruiting HCF173 to *psbA* mRNA (Jin *et al.* 2018). HCF173 has been shown to activate *psbA* translation (Schult *et al.* 2007), but the assays that demonstrated the *psbA* translation defect in *hcf173* mutants left unclear the magnitude of the defect or whether expression of other PSII subunits was also compromised. To further evaluate the proposed collaboration between LPE1 and HCF173, we used Ribo-seq to analyze chloroplast translation in an Arabidopsis *hcf173* mutant (Fig. 2.6). The results demonstrated a severe loss of ribosomes from *psbA* mRNA

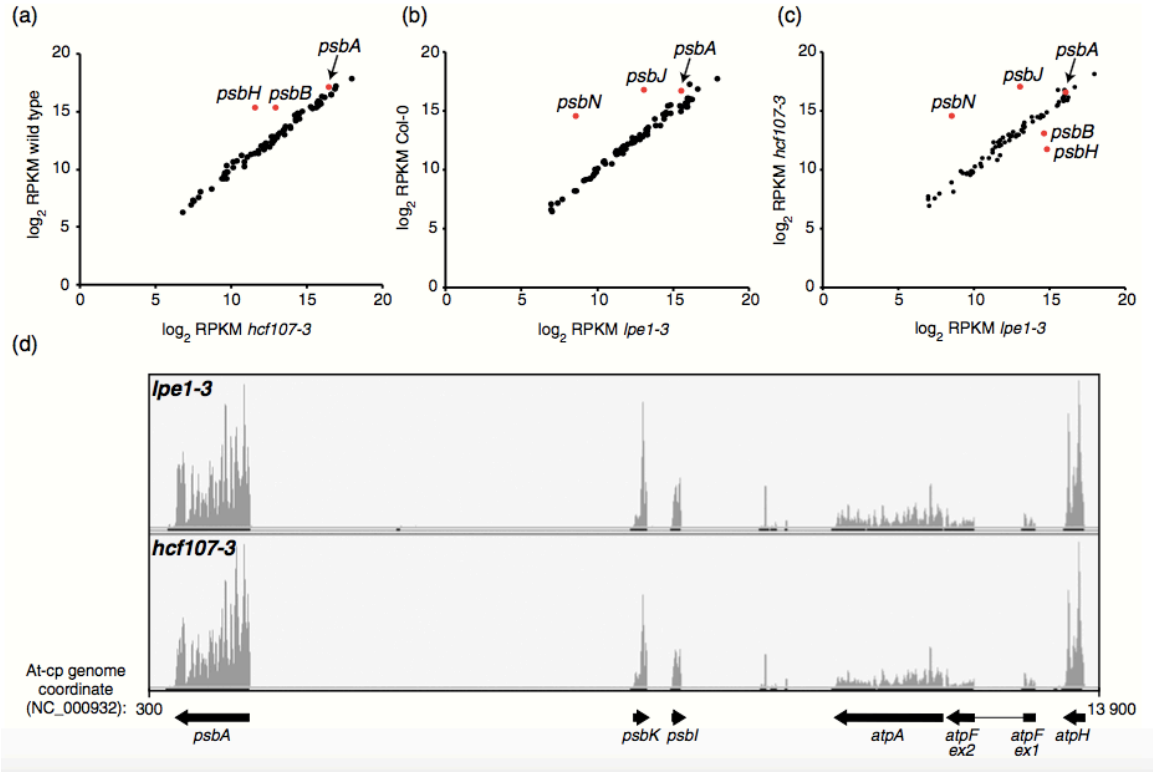


Figure 2.5. Comparison of Ribo-seq data for *hcf107* and *lpe1-3* mutants. The read count and RPKM values for all chloroplast genes are shown in Table S2.
(a) Scatter plot showing the normalized abundance of ribosome footprints mapping to each chloroplast gene in the *hcf107* mutant in comparison to normal siblings. The results validate the previous conclusion that HCF107 is required for *psbH* expression and stimulates *psbB* expression (Felder *et al.* 2001).
(b) Scatter blot of the normalized abundance of ribosome footprints mapping to each chloroplast gene in the *lpe1-3* mutant in comparison to Col-0 plants grown in parallel. The underlying data are the same as those shown in Fig. 2.4b.
(c) Scatter plot comparison of chloroplast Ribo-seq data for *hcf107* and *lpe1-3* mutants that had been grown and analyzed in parallel.
(d) Screen capture comparing Ribo-seq read coverage at *psbA* and nearby genes in *lpe1-3* and *hcf107* mutants.

in *hcf173* mutants (~17-fold), whereas the expression of all other chloroplast genes was unaffected. RNA gel blot hybridizations show that the loss of *psbA* ribosome footprints in *hcf173* mutants is due, in part to decreased *psbA* mRNA levels, as reported previously (Schult *et al.* 2007), but *psbA* mRNA remains plentiful in the mutant (Fig. 2.6 inset). These results confirm that HCF173 is essential for *psbA* translation and show, in addition, that it has no substantive effect on the expression of other chloroplast genes.

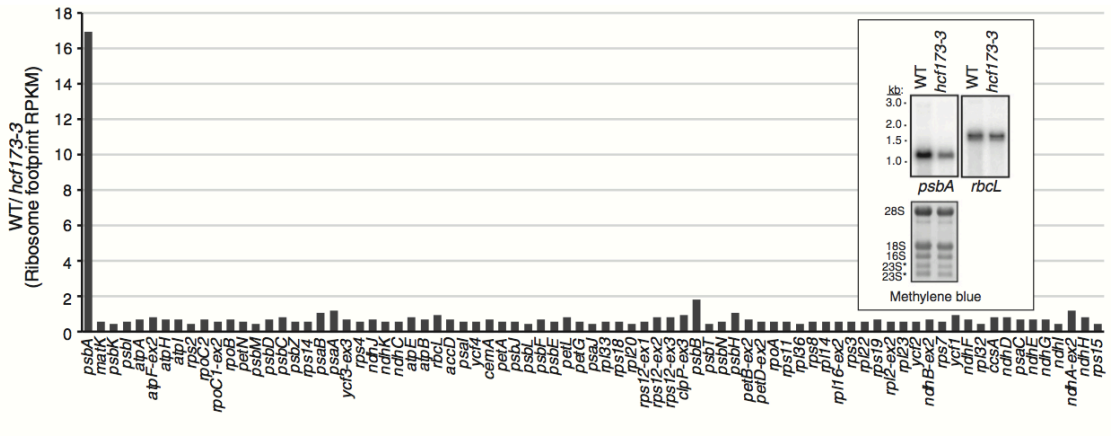


Figure 2.6. Ribo-seq analysis of chloroplast gene expression in Arabidopsis *hcf173* mutants. Ratio of normalized ribosome footprint abundance in the wild-type relative to the *hcf173* mutant for each chloroplast gene. Values were normalized to the number of reads mapping to chloroplast ORFs. The read count and RPKM values for all chloroplast genes are shown in Table S2. The inset shows RNA gel blot hybridizations of RNA from the same homogenates used for ribosome profiling. The blot was hybridized sequentially with probes specific for *psbA* and *rbcL*. The methylene blue-stained blot is shown below.

Recombinant LPE1 binds with higher affinity to the *psbJ* 5' UTR than to the *psbA* 5' UTR.

It was reported that recombinant LPE1 binds the *psbA* 5'UTR *in vitro* (Jin *et al.* 2018). However, the significance of that interaction is called into question by the results above suggesting that LPE1 has no direct impact on *psbA* expression. In fact, no evidence had been presented that LPE1's interaction with the *psbA* 5' UTR was sequence-specific. We revisited this issue by comparing binding of recombinant LPE1 to sequences in the *psbJ* and *psbA* 5' UTRs (Fig. 2.7). We tested binding to a 38-nucleotide RNA sequence mapping between 20 and 58 nucleotides upstream of the *psbJ* start codon; this region seemed a good candidate for harboring an LPE1 binding site because it includes a highly conserved block (Fig. 2.7) and its position with respect to the start codon is similar to that of the few well-defined binding sites for chloroplast translational activators (Fujii *et al.* 2013; Hammani *et al.* 2012; Prikryl *et al.* 2011). Furthermore, binding nearer than this to

the start codon would encroach on the footprint of the initiating ribosome (Chotewutmontri and Barkan 2016), and would likely inhibit rather than activate translation. For *psbA,* we tested binding to the complete 5' UTR, but divided it into three overlapping segments (Fig. 2.7, bottom). The results show that LPE1 binds with much higher affinity to the *psbJ* sequence than to the *psbA* sequences. The very weak binding to *psbA* RNAs appears to be non-specific, as it is at the level that is often observed for promiscuous RNA-binding by PPR proteins *in vitro* (e.g. Hammani *et al.* 2012; Sun *et al.* 2013; Williams-Carrier *et al.* 2008). Taken together with the Ribo-seq data above, these results strongly suggest that LPE1 directly activates *psbJ* translation by binding the *psbJ* 5' UTR and that it does not directly activate *psbA* translation.

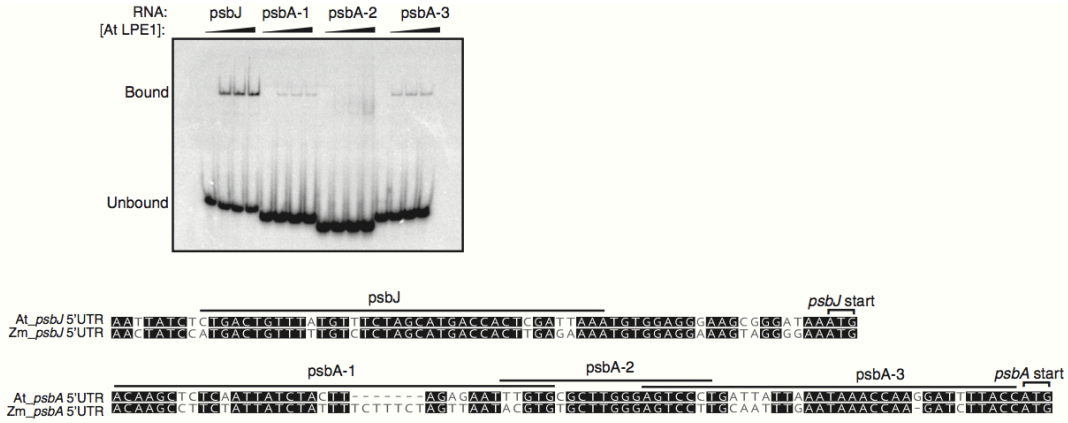


Figure 2.7. Gel mobility shift assay demonstrating relative affinities of recombinant LPE1 for sequences in the *psbJ* and *psbA* 5' UTR.
The RNA sequences used for binding assays are marked on the alignments below. Binding reactions included RNA at 15 pM and recombinant LPE1 at 0, nM, 125 nM, 250 nM, or 500 nM. The experiment was performed five times using heparin concentrations ranging from 0-50 µg/ml, with similar results.

DISCUSSION

Our results show that the orthologous proteins TPJ1 and LPE1 in maize and Arabidopsis, respectively, share a conserved function as translational activators of the chloroplast *psbJ* ORF. Our data show that *psbN* translation is also compromised in *lpe1* mutants. Given that PsbN is an assembly factor for PSII (Torabi *et al.* 2014), the *psbN* expression defect likely contributes to their loss of PSII. Although *psbA* ribosome occupancy was reduced slightly in *lpe1* mutants, *hcf107* mutants have a similar reduction; this suggests that the small decrease in *psbA* translation in *lpe1* mutants is a secondary effect of their PSII defect. By contrast, the *psbJ* and *psbN* expression defects in *lpe1* mutants cannot be explained in this way because *hcf173* and *hcf107* mutants express *psbJ* and *psbN* normally.

These findings show that the loss of PsbJ is the sole cause of the PSII deficiency in maize *tpj1* mutants and strongly suggest that it is a major contributor in Arabidopsis *lpe1* mutants. Analysis of *psbJ* knockout mutants in tobacco showed that PsbJ is important for the accumulation of the PsbP subunit of the oxygen enhancing complex and for the association of LHCII with the PSII core (Suorsa *et al.* 2004). Tobacco *psbJ* mutants have a light-sensitive phenotype that is more severe in mature than in young leaves (Hager *et al.* 2002; Suorsa *et al.* 2004; Swiatek *et al.* 2003). These mutants also exhibit a decrease in PSI under some conditions (Hager *et al.* 2002; Swiatek *et al.* 2003). We observed a decrease of PSI in *lpe1* mutants (Appendix D) whereas the prior study did not (Jin *et al.* 2018), suggesting that PsbJ in Arabidopsis likewise impacts PSI stability under specific conditions. The maize *tpj1* mutant now provides an opportunity to investigate PsbJ function in the context of a C4 monocot.

It seems likely that TPJ1 and LPE1 exert their effect on *psbJ* translation by binding its 5'-UTR. Although an association between LPE1 and the *psbJ* RNA was not detected by RNA coimmunoprecipitation (Jin *et al.* 2018), false-negatives are common in experiments of that type. Furthermore, the tagged LPE1 used for that experiment was not shown to be functional. It is not possible to use the "PPR code" to infer the sequence specificities of LPE1 and TPJ1 because their PPR motifs are irregular. That said, we showed that LPE1 binds selectively to a conserved region of the *psbJ* 5'-UTR *in vitro* (Fig. w.7). This region is predicted in both maize and Arabidopsis to contribute to stable RNA conformations that would occlude the *psbJ* translation initiation region (Appendix E). We hypothesize that TPJ1 and LPE1 binding prevents the formation of these inhibitory RNA structures, analogous to the mechanism suggested for various other chloroplast translational activators (reviewed in Zoschke and Bock 2018).

Our data call into question the functions proposed previously for LPE1 (Jin *et al.* 2018). The PSII deficiency in the *lpe1-3* mutant had been attributed to a defect in *psbA* translation, but the much stronger reduction in *psbJ* and *psbN* expression shown here accounts for their PSII deficiency. Furthermore, it was observed previously that a D1 degradation product is found at elevated levels in *lpe1-3* mutants [see Fig. S10 in (Jin *et al.* 2018)]; this seems inconsistent with the view that D1 is synthesized at reduced rates and suggests instead that the loss of D1 is due to accelerated degradation. It was also proposed that LPE1 activates *psbA* translation by binding directly to *psbA* RNA and recruiting HCF173 (Jin *et al.* 2018). A key piece of evidence for this model was the binding of LPE1 to the *psbA* 5'UTR *in vitro* (Jin *et al.* 2018). We show here, however, that recombinant LPE1 binds with much higher affinity to the *psbJ* 5'UTR; the very weak

binding to *psbA* sequences is at a similar level to the non-specific interactions that are often observed with PPR proteins.

Our experiments identify two plausible explanations for the apparent reduction in *psbA* translation reported previously for *lpe1* mutants: (i) the absence of PSII causes secondary effects on *psbA* translation (as demonstrated by our *hcf107* data); (ii) the comparison of stunted mutant plants to wild-type plants at a more advanced developmental stage may have contributed to the apparent *psbA* translation deficit due to the unique developmental dynamics of *psbA* translation. The *psbA* gene is one of a handful of chloroplast genes whose translational efficiency increases through leaf development in maize, and it stands out as the gene for which this increase is greatest (Chotewutmontri and Barkan 2016). It seems likely that an analogous scenario holds true in Arabidopsis. Because non-photosynthetic Arabidopsis mutants grow more slowly than the wild-type, it is common to compare mutant plants to normal seedlings of the same age but at a more advanced developmental stage. Indeed, the prior report on LPE1 (Jin *et al.* 2018) and our initial Ribo-seq analysis of *lpe1-4* mutants (Fig. 2.4a) took this approach. However, when we selected mutant and normal plants at a similar developmental stage, their difference in *psbA* ribosome occupancy was minor (compare Fig. 2.4a to Fig. 2.4b) and this residual effect can be accounted for by the PSII defect in *lpe1* mutants.

Defects in PsbJ and PsbN synthesis would be very difficult to detect with pulse-labeling assays due to their short length and their lack of sulfur-containing amino acids. Furthermore, polysome assays are problematic for *psbJ* due to the fact that the *psbJ* ORF constitutes only a small fraction of the polycistronic RNA on which it resides. In fact, these methods are poorly suited for assessing translation rates of the majority of

chloroplast ORFs. Therefore, incorporation of ribosome profiling into the standard toolkit for the analysis of mutants with impaired photosynthesis can be expected to reveal numerous additional nucleus-encoded proteins that activate the translation of specific chloroplast ORFs.

EXPERIMENTAL PROCEDURES

Plant material

Maize TPJ1 is encoded by GRMZM2G056116 (B73 RefGen_v3) or Zm00001d036489 (B73 RefGen_v4). Evidence for its orthology with Arabidopsis LPE1 (AT3G46610) can be found at http://cas-pogs.uoregon.edu/#/pog/12707. The *tpj1* insertion lines were generated by the UniformMu project (McCarty *et al.* 2013) and provided by the Maize Genetics Cooperative. The *tpj1-1* allele is cluster mu1039450 and seed stock UFMU-03095. The *tpj1-2* allele is cluster mu1042597 and seed stock UFMu-05008. Ribo-seq data for the *tpj1* gene showed a virtual absence of expression *tpj1* expression in *tpj1-2* mutants, indicating that *tpj1-2* is a null allele (Appendix B). Seed for Arabidopsis mutants was obtained from the Salk collection (Alonso *et al.* 2003). All T-DNA insertion sites were confirmed by DNA sequencing. *lpe1-4* is SALK_064817 and has an insertion 1798 nucleotides downstream of the start codon. Seed for *lpe1-3* was generously provided by Hongbin Wang (Sun Yat Sen University). The *hcf107* allele used for Ribo-seq is SALK-079285C and has a T-DNA insertion ~20 bp upstream of the start codon. The *hcf173* allele used for Ribo-seq is SALK_035984, and has a T-DNA insertion in the first exon, 65 nucleotides downstream from the start codon.

Maize was sown in soil and grown in diurnal cycles (16 h light, 8 h dark) at 28˚C and 26°C for the light and dark periods, respectively. Plants were illuminated using a light intensity of approximately 300 μmol m^{-2} s^{-1}. The second and third leaves to emerge were harvested 9 or 10 days after planting and flash frozen in liquid nitrogen prior to processing for Ribo-seq or RNA extraction. Arabidopsis seeds were sterilized by incubating the seeds in a 1% bleach and 0.1% SDS solution for 10 minutes followed by a 70% ethanol wash, and three washes with sterile water. Seeds were sown on sterile MS medium [4.33 g/L Murashige and Skoog Basal Medium (Sigma), 2% sucrose, 0.3% Phytagel (Sigma), pH 5.7]. After vernalization, plants were grown in a growth chamber at 22ºC in diurnal cycles (10 h light, 14 h dark) at a light intensity of 100 μmol m^{-2} s^{-1} (*lpe1-4* experiment) or 75 μmol m^{-2} s^{-1} (*hcf107* and *lpe1-3* experiments), and harvested at midday on the day indicated in the legends showing each experiment. Plants (aerial portion only) were flash frozen in liquid nitrogen and stored at -80 ˚C until use.

RNA gel blot and immunoblot analysis

Immunoblots and RNA gel blots were performed as described previously (Barkan 1998). Proteins for immunoblots in maize were extracted from the apical 2-cm of the second leaf of 10-day seedlings. RNA gel blot hybridizations for maize used RNA extracted from the same homogenate used for Ribo-seq analysis. RNA gel blot hybridizations for Arabidopsis used RNA extracted from separate aliquots of the flash-frozen leaf tissue used for Ribo-seq analysis. The hybridization probes are described in Table S1 (Appendix F). Antibodies to PsbB, PsbC, PsbD, PsaA, Lhcb2, and Rps12 were purchased

from Agrisera. Antibodies to PsbA, PetD, PsaD, AtpB were generated by our group and were described previously (Pfalz *et al.* 2009).

Ribosome profiling

Ribosome footprint preparation, library construction, and data analyses were performed as described previously (Chotewutmontri and Barkan 2018). Ribo-seq analysis of maize *tpj1* mutants used the apical half of the second leaf to emerge in 9-day old plants, pooling tissue from two mutant seedlings and two phenotypically-normal siblings. Three biological replicates were performed for *tpj1* mutants. Ribosome profiling of Arabidopsis mutants used the aerial portion of plants harvested 16 days *(hcf173*), 18 days (*lpe1-4*), or 20 days (*hcf107* and *lpe1-3)* after vernalization, pooling between two and four seedlings for each sample. Phenotypically normal siblings served as the wild-type controls for *hcf173*, *hcf107,* and *lpe1-4.* Col-0 was used as the wild-type control for *lpe1-3* because the mutant line was homozygous.

Gel mobility shift assays

Recombinant LPE1 was expressed as a fusion to thioredoxin and 6xhistidine tag from a pET-32a vector in *E. coli* strain CD41 (Lucigen), using the same expression construct used in the original report on LPE1 (Jin *et al.* 2018). The plasmid and expression protocol were generously provided by Hongbin Wang (Sun Yat-sen University). The protein was enriched by nickel affinity chromatography in a buffer containing 20mM sodium phosphate (pH 7.5), 500 mM NaCl, 20 mM imidazol, and 5 mM β-mercaptoethanol. The column was washed with the same buffer supplemented with 200 mM imidazole and the

protein was eluted with 800 mM imidazole. The recovered protein was further purified by gel filtration chromatography on a Superdex 200 column as described previously (Barkan *et al.* 2012), dialyzed into 30 mM Tris–HCl pH 8.0, 400 mM NaCl, 50% glycerol, 10 mM β-mercaptoethanol and stored at −20°C. Gel mobility shift assays were performed as previously described (Barkan *et al.* 2012) using radiolabeled synthetic RNA oligonucleotides (IDT). Binding reactions contained 15 pM RNA, 45 mM Tris–HCl (pH 8.0), 180 mM NaCl, 10% glycerol, 4 mM DTT, 10 U RNasin, 0.1 mg/mL BSA, 25 μg/mL heparin and protein at 500 nM, 250 nM, or 125 nM. Binding reactions were incubated for 45 min at 25°C and resolved on a non-denaturing 5% polyacrylamide gel at 4°C. The data were visualized with a phosphorimager. The experiment was performed five times using heparin concentrations ranging from 0-50 μg/ml, with similar results.

Accession *numbers*

tpj1 **-** GRMZM2G056116 (B73 RefGen_v3) or Zm00001d036489 (B73 RefGen_v4)

LPE1- AT3G46610

HCF107- AT3G17040

HCF173- AT1G16720

BRIDGE II

While we showed in Chapter II that LPE1 does not function in *psbA* translation, I continue to focus on regulators of *psbA* translation in Chapter III. There I describe a series of Yeast-two hybrid experiments which probe interactions among proteins hypothesized to be proximal translational regulators of *psbA* RNA.

CHAPTER III

TITLE: INTERACTIONS AMONG PROTEINS REQUIRED FOR THE SYNTHESIS OF D1.

AUTHORS:

Sonja M. Ljungdahl

INTRODUCTION

Previous research supports the hypothesis that the plant proteins HCF244 and HCF173 are required for initiation of *psbA* translation (Link et al., 2012). Studies in the cyanobacterium *Synechocystsis* revealed the presence of a protein-pigment complex consisting of Ycf39 (the ortholog of HCF244), an assembly factor called Ycf48, two high-light inducible proteins (Hlips), HliC and HliD, as well as three core subunits of PSII: D1, D2, and cytochrome b559. This complex is hypothesized to represent a very early stage of PSII assembly, and to insert chlorophyll recycled from degraded D1 into the nascent D1 during PSII repair (Knoppova et al., 2014). Ycf39 knock-out experiments revealed that the D1 DE loop, extending out into the stroma between the 4th and 5th transmembrane segments, is required for the recruitment of Ycf39 to the Ycf39-Hlip complex. This suggests that Ycf39 may bind to a site on the D1 DE loop. Taken together, the data from cyanobacteria suggest that Yc39 plays a role in D1 assembly during PSII repair at the thylakoid membrane's cytoplasmic surface (Knoppova et al., 2014).

Myouga et al. (2018) found a related complex in thylakoid membranes of the flowering plant Arabidopsis (*Arabidopsis thaliana)* that includes HCF244 (the ortholog of Ycf39), HCF136 (ortholog of Ycf48), and OHP1 and OHP2 (orthologs of Hlips) (Myouga et al., 2018). OHP1 is required for PSII accumulation (Myouga et al., 2018), and the loss of OHP2 results in the total reduction of both HCF244 and OHP1 (Hey and Grimm, 2018). These results suggest that this PSII assembly/repair complex has been conserved between cyanobacteria and chloroplasts. Furthermore, unpublished work from the Barkan lab showed that OHP1 and OHP2 are required for the recruitment of ribosomes to *psbA* mRNA, as had been shown previously for HCF173 and HCF244.

These findings and unpublished experiments from the Barkan lab suggest a conserved protein complex that couples D1 synthesis to PSII repair at the thylakoid membrane (Figure 3.1). To discover how the proteins in this complex regulate ribosome recruitment to the *psbA* mRNA, i.e., which proteins are directly interacting with each other, I performed a series of Yeast 2 hybrid experiments. Contemporaneously, a paper by Hey and Grimm was published (2018), which confirmed an interaction between HCF244 and OHP2. Their research includes a series of BiFC or Bimolecular Florescence complementation experiments, showing that a region of the OHP2 protein (*the middle region*) is necessary for the interaction of OHP2 and HC244. In the present study, I demonstrate that the OHP2 middle region is sufficient for an interaction with HCF244.

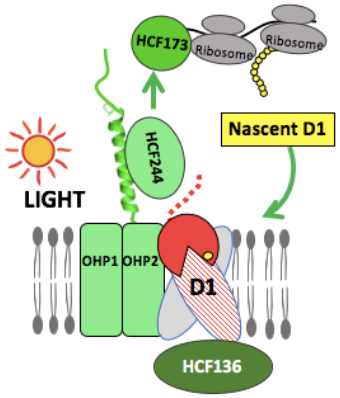


Figure 3.1 Working model that connects light-induced D1 degradation to the activation of *psbA*-ribosome recruitment in the context of a PSII assembly/repair complex.

MATERIALS AND METHODS

Plasmid Construction

Yeast-two hybrid assays were performed using pBD-GAL4 Cam as bait and pAD-GAL4-2.1 as the prey vector from the HybriZAP-2.1 Two-Hybrid system kit (Agilent technologies). I linearized the vectors using New England Biolab's (NEBs) 10x high fidelity (HF) restriction enzymes and set up a 40 microliter reaction: 4 µL of 10x NEB Cut Smart Buffer, 4 µL of plasmid (4 µg), 0.5 µL (10 units) of Sal1 restriction enzyme, 0.5 µL (5 units) of EcoR1 restriction enzyme, and 30 µL of sterile water. The solutions were incubated at 37 °C for one hour and then run out on a 1% agarose gel in 1XTAE Buffer. Plasmids were gel purified using Thermo Fisher Scientific's GeneJET Gel Extraction Kit according to the manufacturer's instructions.

Gene blocks for all sequences of interest in this investigation were synthesized by Integrated DNA technologies; sequences are provided in Appendix G. Plasmid vectors were assembled by Gibson Assembly (New England Biolabs) on all constructs and

cloned in frame between the EcoR1 and Sal1 sites of pBD-GAL4 Cam and pAD-GAL4-2.1 vectors (Figure 3.2). For the construction of pADHCF244, gene block At-HCF244 pAD gene block n-term (496 bp) and At-HCF244 pAD/pBD gene block c-term (567 bp) were cloned into pAD-GAL4-2.1. For the construction of pADOHP2_middle, gene block, pAD_At-OHP2_Basic Gene Block (200bp) was cloned into pAD-GAL4-2.1. For the construction of pADOHP2_pro_middle, gene block, pAD_At-OHP2_ProRS_Basic Gene Block (329 bp) was cloned into pAD-GAL4-2.1. For the construction of pBDHCF173 gene block, At-HCF173 pBD gene block n-term (664 bp) and At-HCF173 pAD/pBD gene block c-term (913 bp) were cloned into pBD-GAL4 Cam. For the construction of pBDHCF244 gene block, At-HCF244 pBD gene block n-term (495 bp) and At-HCF244 pAD/pBD gene block c-term (567 bp) were cloned into pBD-GAL4 Cam. All plasmid inserts were sequenced at (Sequetech) and examined for accurate insertion alignment and nucleotide sequence (See Appendix G for nucleotide sequences of cDNA fragments).

Yeast two-hybrid assay

Plasmid constructs were transformed into *E. coli* strain XL1 Blue cells (Agilent technologies) and grown at 37°C on LB carbenicillin (50 µg/ml final concentration) plates. Competent yeast *S. cerevisiae* YRG-2 strain (Agilent technologies) were transformed with the two hybrid plasmid constructs, pBD-GAL4 Cam (DNA binding domain) and pAD-GAL4-2.1 (DNA activation domain), according to the manufacturer's instructions. They were plated on selection plates lacking leucine and histidine (pAD-GAL4-2.1) or plates lacking tryptophan and histidine (pBD-GAL4 Cam) and incubated in 30°C room for 2-4 days. Each experiment was performed four times. I used the control

plasmids provided in the HybriZAP kit pBD-WT and pAD-WT as the strong interaction controls, pBD-MUT and pAD-MUT as the weak interaction controls, and pLamin C and pAD-MUT as the negative interaction controls. Screening for *HIS3* reporter gene expression was performed by growing constructs overnight in liquid drop out media without leucine, tryptophan or both. I pelleted and resuspended all cultures to give an OD_{800} of 0.5. A 10 μL volume of the suspensions as well as 10x, 100x, and 1000x dilutions were dropped on plates containing solid SD media lacking either leucine and tryptophan (growth control), or leucine, tryptophan, and histidine (reporter gene assay). These plates were grown at 30°C for 2-4 days. Filter lift assays were performed on plates according to HybriZAP-2.1 Manual (see section on Screening-Filter Lift Assay) to screen for the expression of lacZ reporter gene. As per the recommendation of professor Diane Hawley, I used Schleicher and Schuell No. 589 Blue Ribbon Filter paper. In addition, I diverged from Agilent technology's protocol by holding the filter paper in liquid nitrogen for thirty seconds one time instead of three times for 10 seconds.

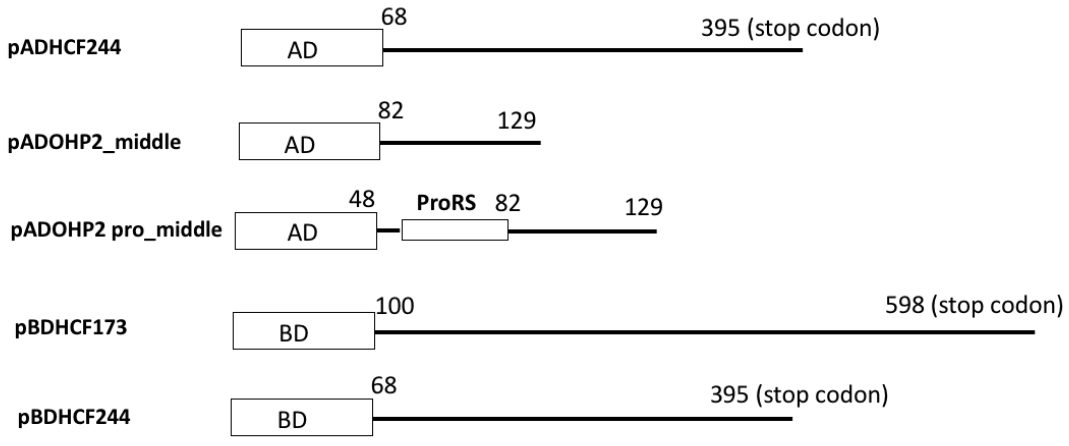


Figure 3.2. Gene constructs used in Y2H experiments. PCR products of gene block fragments were cloned into yeast-two hybrid vectors pAD-GAL4-2.1 or pBD-GAL4 Cam.

RESULTS

Detection of the interaction between OHP2 and HCF244

I performed a series of Yeast Two-Hybrid experiments to identify interactions among *Arabidopsis thaliana* proteins required for ribosome recruitment to *psbA* RNA and thus, D1 synthesis (Link et al., 2012, Knoppova et al., 2014, Myouga et al., 2018, Hey and Grimm, 2018). I screened for interactions between HCF244, HCF173 and two segments of OHP2 (Figure 3.2): OHP2-pro-middle represents the entire region that extends out of the membrane into the stroma whereas OHP2_middle lacks a proline-rich region that is at the N-terminus of the "stromal tail". In addition, I tested individual transformants to confirm there was no autoactivation from any single construct (Figure 3.3).

The screen yielded one interacting pair, OHP2 and HCF244; both the shorter (OHP2_middle) and longer (OHP2 pro_middle) OHP2 constructs showed a positive interaction with HCF244. The strength of the interaction between OHP2 and HCF244 was comparable with strong interaction controls pADWT + pBD WT for positive histidine expression (Figure 3.4).

This result shows that the OHP2 middle region alone is sufficient for an interaction with HCF244. No interaction was detected between pADHCF244+pBDHCF173, pADOHP2 pro_middle + pBDHCF173, pADOHP2_middle + pBDHCF173, or in the negative control reaction (pLaminC + pADmut) (Figure 3.4).

To validate the interaction between OHP2 and HCF244, I performed a filter lift assay to detect the expression of the second reporter gene, *lacZ,* on colonies that showed positive histidine expression. The result confirmed the interaction between pADOHP2_middle and pBDHCF244, but the colonies that included the longer

pADOHP2 pro_middle did not show expression of *lacZ* (Figure 3.5). It is likely that the interaction between HCF 244 and the longer protein segment is weaker than the interaction between HCF 244 and the middle region alone.

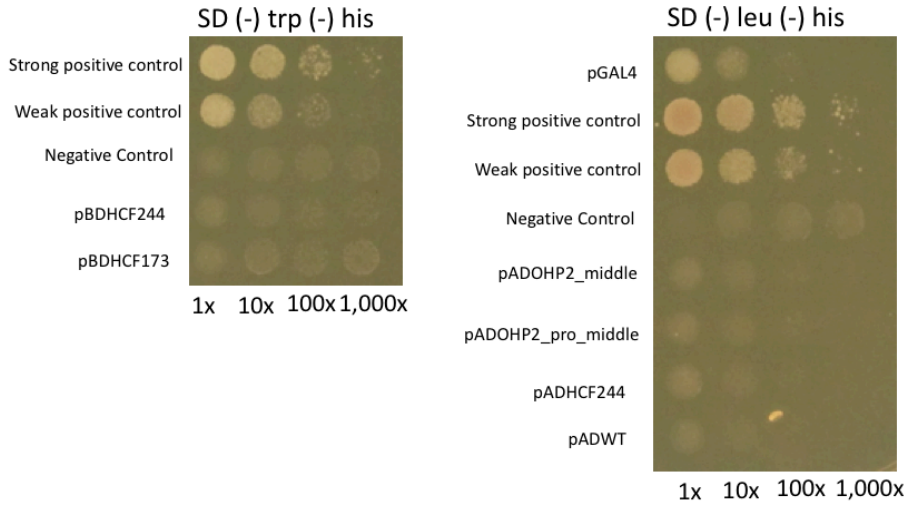


Figure 3.3: Constructs do not activate HIS3 reporter gene on their own. These plates show growth on media lacking tryptophan and histidine (left) or lacking leucine and histidine (right). Lack of growth confirms there is no autoactivation from individual constructs.

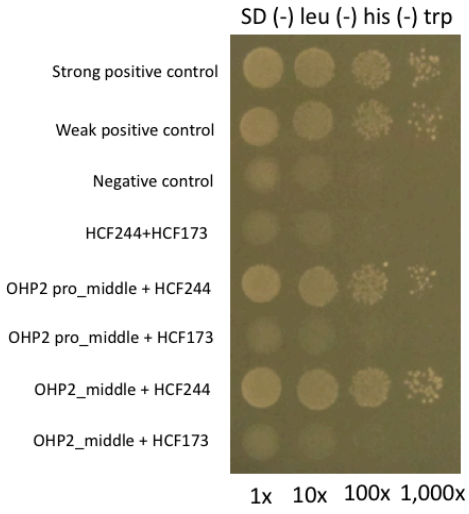


Figure 3.4. OHP2 middle region is sufficient for an interaction with HCF244 in yeast. This plate shows growth on media lacking leucine, tryptophan and histidine. Growth without histidine indicates an interaction between proteins. An interaction is evident in both strong and weak positive controls as well as with HCF244 and both versions of OHP2 proteins: the middle plus proline-rich region and the middle region alone.

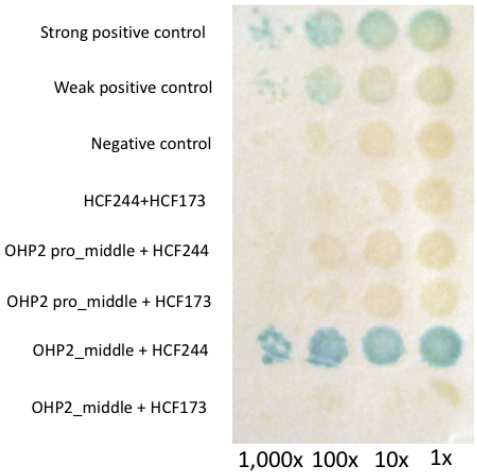


Figure 3.5. OHP2 middle region + HCF244 activate the β-galactosidase reporter. Transcription of *lacZ* reporter gene (β-galactosidase activity) was assayed for each co-transformant. In all assays, pADWT + pBD WT was used as strong positive control, pBD-MUT and pAD-MUT as the weak interaction controls, and pLamin C and pAD-MUT as the negative interaction controls. An interaction is evident in both strong and weak positive controls as well as with OHP2 middle region and HCF244.

Comparison of OHP2 middle interaction region with orthologs in other species reveals conserved amino acid sequence.

The OHP2 middle interaction region is the C-terminal portion of the protein that extends out into the stroma. Multiple sequence alignment in this region between *Arabidopsis* and Maize, Poplar, and rice reveals a highly conserved region at the interaction site (Figure 3.6).

Figure 3.6: Sequence alignment of OHP2 in rice *Oryza sativa* (1), Maize *Zea mays* (2, 3), *Arabidopsis thaliana* (4), Aspen *Populus tremula* (5,6). Approximate region (Amino acids 81-130 (Arabidopsis) sufficient for interaction with HCF244 underlined in red.

Helical Wheel Projection in the middle region of the OHP2 stromal tail Shows Basic, Acidic and Polar Clustered Regions

We used I-TASSER to predict the structure of the middle region of OHP2. The structure shows an alpha helix followed by 11 unstructured residues. A helical wheel projection shows clustering of acidic residues in two regions, basic residues in two regions, and polar residues in two regions. In addition, the software predicts that the helix binds a peptide, or possibly Calcium (Figure 3.7).

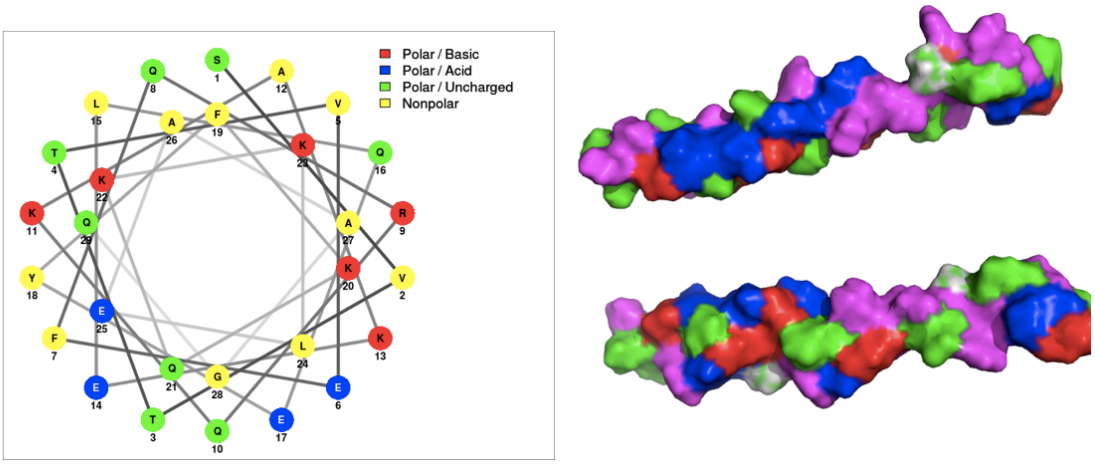


Figure 3.7: Helical Wheel Projection in the middle region of the OHP2 stromal tail Shows Basic, Acidic and Polar Clustered Regions

DISCUSSION

In this study I confirmed an interaction between OHP2 and HCF244, specifically that the interaction domain is in the section of OHP2 that projects out into the stroma between amino acids 82-129. This result shows that the OHP2 middle region alone is sufficient for an interaction with HCF244. This complements the conclusion reached by Grimm and Hey (2018), who showed through BiFC-bimolecular florescence complementation experiments that the OHP2 middle region is required for an interaction with HCF244. By nature of BiFC experiments, Grimm and Hey showed an interaction between the two proteins, while my Y2H reveals this interaction is direct and not mediated by other proteins present in chloroplasts.

This study supports the conjecture that a conserved thylakoid membrane complex regulates translation of *psbA* RNA. This complex includes HCF244 and HCF173, which have been previously shown to be required for *psbA* ribosome association (Link et al., 2012; Schult et al., 2007), and additionally have no impact on the translation of other

mRNAs (Williams-Carrier et al., 2019, and unpublished results from the Barkan lab). The complex also includes OHP1 and OHP2, which are also required for *psbA* translation (unpublished results from the Barkan lab). Fractionation data from Link et al. (2012) show that HCF244 is tightly associated with the stromal face of the thylakoid membrane, and is remarkably resistant to proteases. We conjecture that HCF244 is tethered to the membrane and protected from proteases by its tight interaction with OHP2. It is possible that these proximal regulators function in the activation and/or repression of *psbA* translation by modulating the accessibility of ribosomes to *psbA's* ribosome binding site. Future experiments should be aimed at exploring this hypothesis. In addition, more work is required to understand how the light-induced signal triggers these changes in proximal regulator activity. For example, predictions with I-TASSER (Figure 3.7) predicted that OHP2's helical projection in the stroma binds calcium, which is known to fluctuate in response to changing light conditions (Hochmal et al., 2015; Vainonen et al., 2008). These results suggest that calcium may be involved in the signal transduction pathway that triggers the recruitment of ribosomes to *psbA* mRNA in response to light.

CHAPTER IV

CONCLUSION

In this study we investigated the role of proteins hypothesized to mediate the effects of light on *psbA translation*, and consequently D1 synthesis. We demonstrated that the maize protein TPJ1 does not influence *psbA* translation. Instead, our results show it is required for the translation of the chloroplast *psbJ* mRNA, which encodes a different subunit of PSII. Additionally, I performed Yeast two-hybrid assays to illuminate protein-protein interactions, which we hypothesized are involved in the synthesis of D1. My research shows a specific amino acid sequence on the OHP2 protein is sufficient for interaction with the protein HCF 244. It is intriguing that this region of OHP2 is predicted to bind Ca++ by the I-TASSER protein structure prediction algorithm. Future experiments can address whether Ca++ in fact binds OHP2, and whether this influences its interactions with HCF244 and ability to activate *psbA* translation. This work contributes to the literature specifying the molecular actions driving D1 replacement.

APPENDIX A

SUPPLEMENTAL FIGURE 1

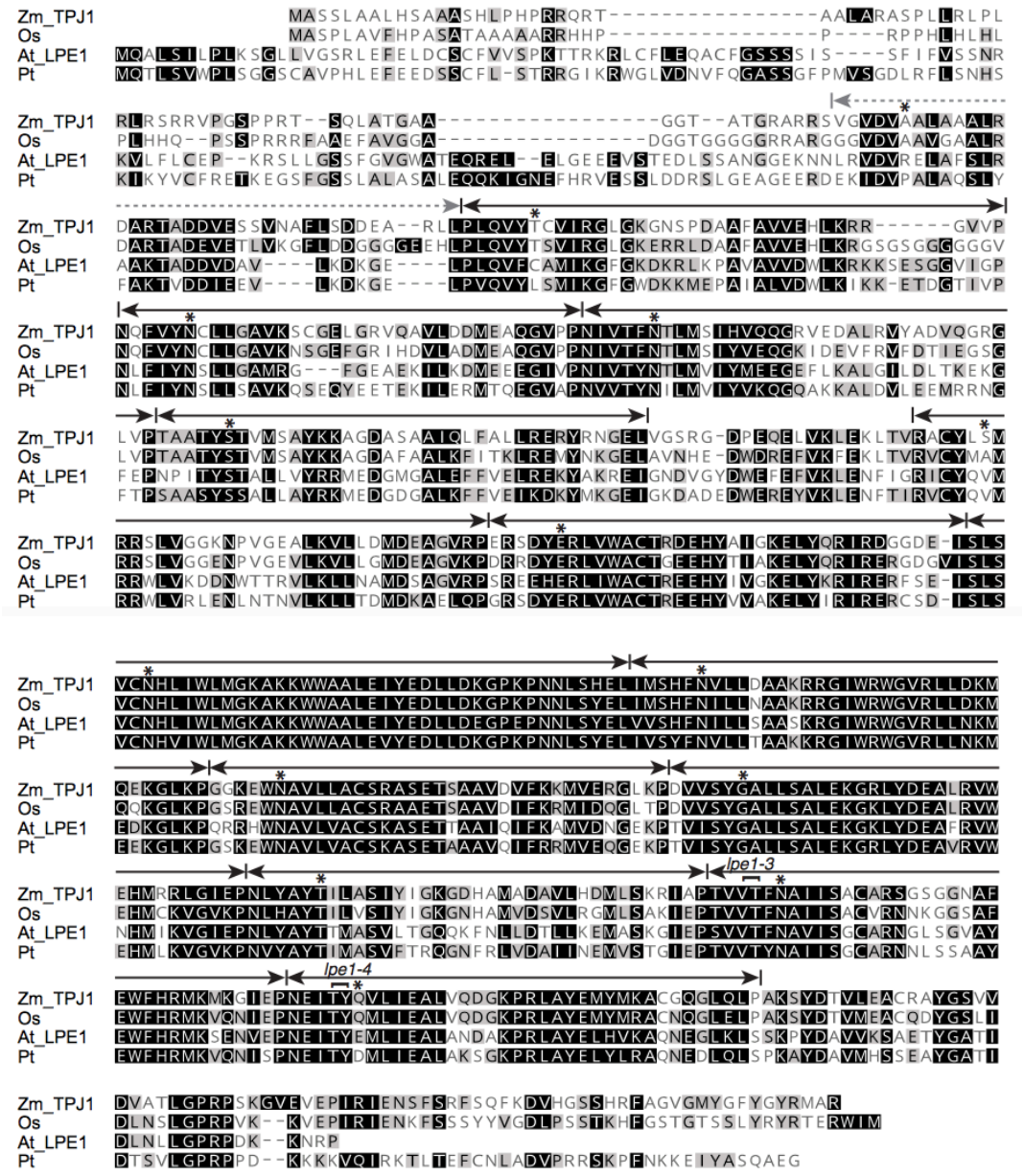


Figure S1. Multiple sequence alignment of LPE1 orthologs in *Arabidopsis thaliana* (At), *Zea mays* (Zm), *Oryza sativa* (Os), and *Populus trichocarpa* (Pt). These correspond to AT3G46610, GRMZM2G056116, Os12g18640, and POPTR_0017s10700, respectively. PPR motifs are marked above with arrows. The first motif is highly degenerate and is marked with a dashed line. The asterisks mark the amino acid in each repeat that discriminates pyrimidines from purines according to the PPR code (Barkan *et al.* 2012). The Predotar (Small *et al.* 2004) and Target P (Emanuelsson and Heijne 2001) algorithms predict that the rice and maize orthologs have chloroplast transit peptides (see http://cas-pogs.uoregon.edu/#/pog/12707). The positions of the T-DNA insertions in the Arabidopsis *lpe1* alleles used here are marked.

APPENDIX B

SUPPLEMENTAL FIGURE 2

Figure S2. Screen captures of Ribo-seq data from the Integrated Genome Viewer (IGV) showing absence of *tpj1* expression from the *tpj1-2* allele. The ubiquitin 2 gene is displayed to provide an internal standard. The number of read counts for each gene is indicated in parentheses.

APPENDIX C

SUPPLEMENTAL FIGURE 3

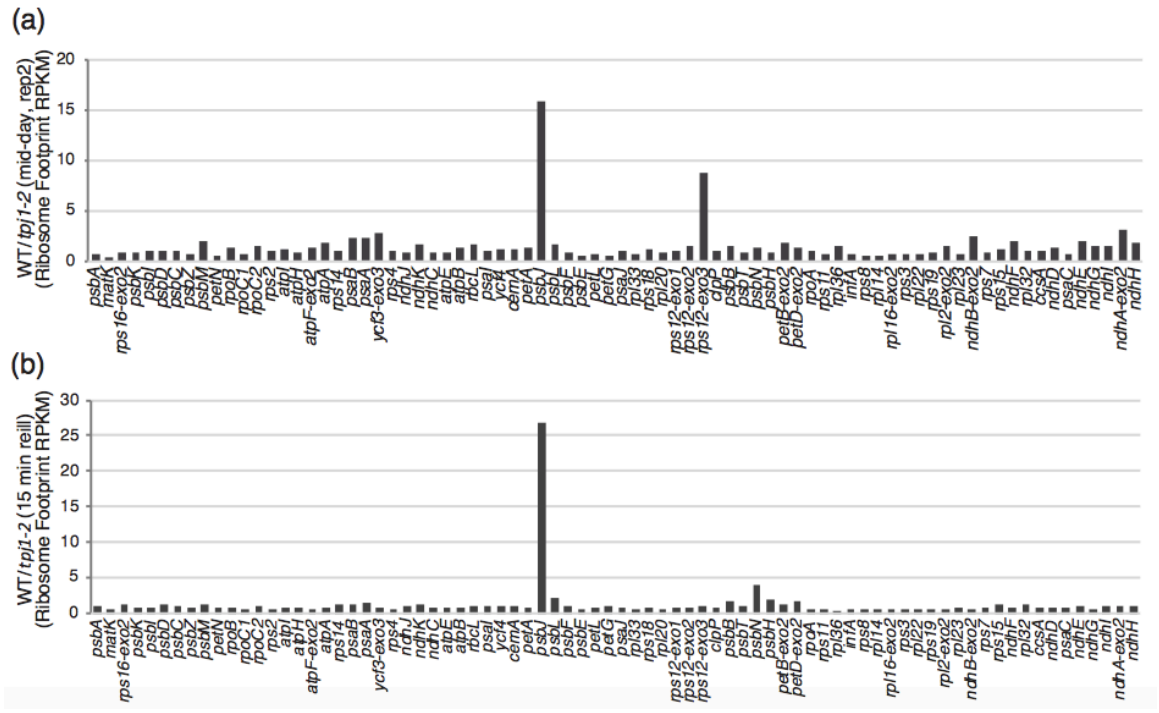


Figure S3. Replicate Ribo-seq assays of maize *tpj1-2* mutants. These experiments came from separate plantings, and used plants harvested at midday (top) or plants that were exposed to one hour dark and then reilluminated for 15 minutes (bottom). Graphs show the ratio of normalized reads mapping to each chloroplast gene in the wild-type siblings relative to the mutants. The apparent defect in ribosome coverage on exon 3 of *rps12* in the first replicate (panel a) is likely a result of sampling error due to the small number of reads.

APPENDIX D

SUPPLEMENTAL FIGURE 4

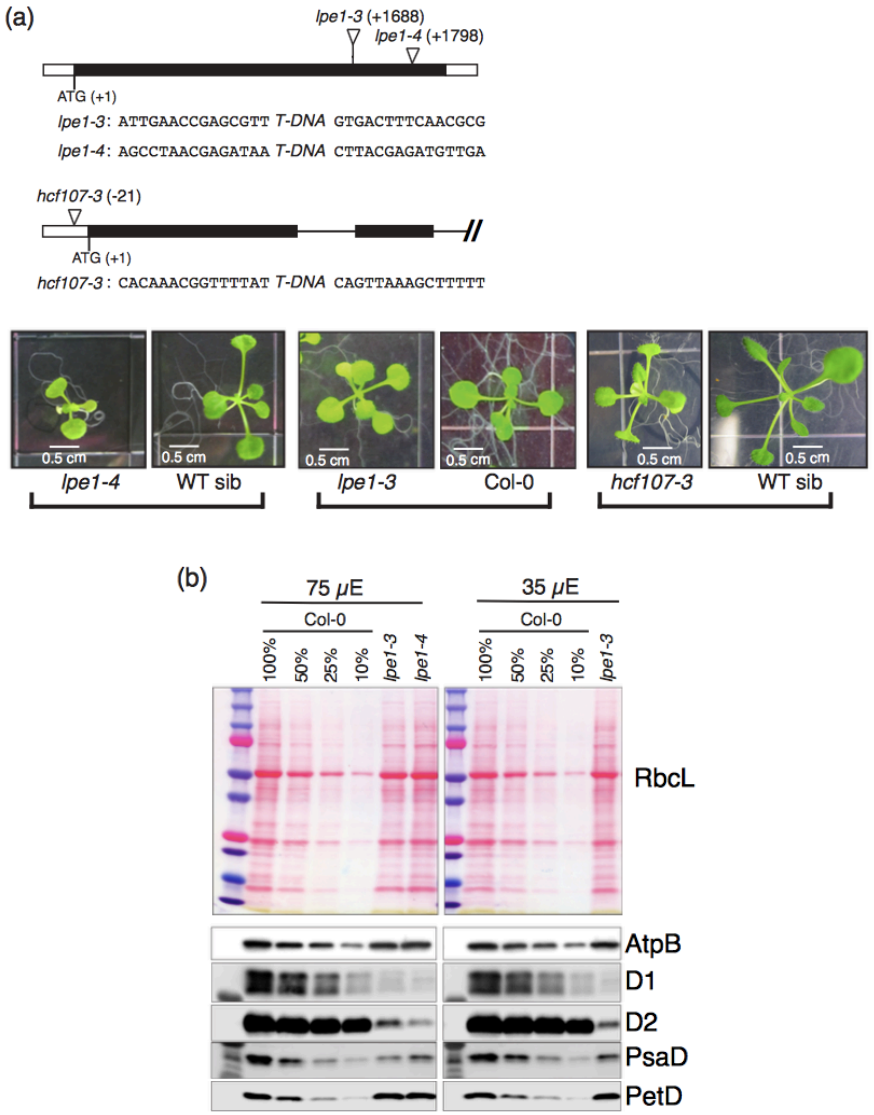
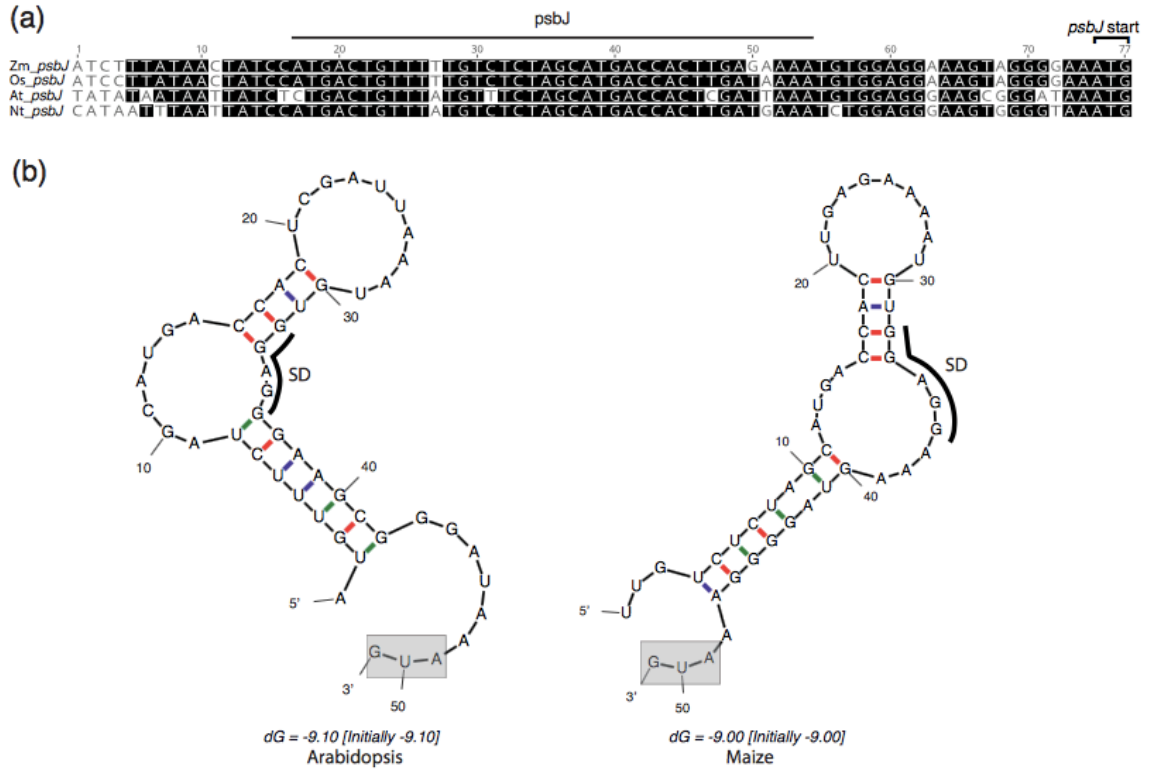


Figure S4. Arabidopsis *lpe1* and *hcf107* alleles used in this study.
(a) The insertion sites and phenotypes of plants used for Ribo-seq. The plants were grown for 18 days (*lpe1-4)* or 20 days (*lpe1-3* and *hcf107*) on MS medium in diurnal cycles as described in Methods. The *lpe1-3* line was homozygous for the insertion, so Col-0 was used as the control.
(b) Seedling leaf extracts from plants harvested 18 days after transfer to growth conditions were fractionated by SDS-PAGE and analyzed by immunoblotting. Plants were grown under different light intensities (75 or 35 µE as indicated). Each blot was probed consecutively with antibodies specific for the indicated proteins. An image of each Ponceau S-stained blot is included to illustrate sample loading and the abundance of the large subunit of Rubisco (RbcL). The *lpe1-3* mutants grown at 75 µE were used for ribosome profiling.

APPENDIX E

SUPPLEMENTAL FIGURE 5



APPENDIX F

SUPPLEMENTAL TABLE 1

Table S1. Probes used for RNA gel blot hybridizations

Arabidopsis	**Sequence**	**Source**
psbJ	ACAAGAATACCAGCTACAGTACCTATTACCCAAAGAGGAATCCTTCCAGTAGTATC AGCC	Oligonucleotide (IDT)
psbA for analysis of *lpe1*	Mixture of two oligonucleotides: CTTTCGCTTTCGCGTCTCTCTAAATTGCAGTCAT GGCTTATATTGCTCGTTTTTTACTAAACTAGATCTAGACTAACACTAACG	Oligonucleotide (IDT)
psbA for analysis of *hcf173*	PCR product from maize; Zm CP coordinates 295-1074	
rbcL	PCR product from maize; Zm CP coordinates 57036-57607	
psbN	CCGTGTTCCTCGAATGGATCTCTTAGTTGTTGAGAGGGTTGCCCAAAGGCAGTAT ATAGAGCATACCCAGT	Oligonucleotide (IDT)
Maize		
psbJ	Maize cp genome coordinates 63299...63515	PCR
rps12 exons 2-3	Maize cp genome coordinates 129659...130428, spliced	PCR from cDNA

Primers for genotyping insertion alleles in maize and Arabidopsis. Primers were ordered from IDT.

SB-16GRM0561165' (*tpj1* forward primer)	GGTGGCGGCTCCTCAGCAAT
SB1344GRM0561163' (*tpj1* reverse primer)	ATCAGCTCGTGGGACAGGTT
EOMumix3 (Mu TIR primer mix)	GCCTCCATTTCGTCGAATCCC and GCCTCTATTTCGTCGAATCCG
At3g17040-F (hcf107 forward primer)	CCAAATCACCCTTAACGTAACA
At3g17040-R (hcf107 reverse primer)	ACGATTGACCGGAGATTCCA
AT3G46610-F (lpe1-4 forward primer)	GGTGAAAAGCCAACGGTCAT
AT3G46610-R (lpe1-4 reverse primer)	CCCCACGGTAACCTTGTACA
At-hcf173-f1 (hcf173-3 forward primer)	AGTAACATGGCTGCGACTGA
At-hcf173-r2 (hcf173-3 reverse primer)	GTAGCCACGGAGCATGAGTT
Lba1 (T-DNA left border)	TGGTTCACGTAGTGGGCCATCG

Table S1. Primers and probes used in this study.

APPENDIX G

SUPPLEMENTAL TABLE 2

At-HCF244 pBD gene block n-term
CAAAGACAGTTGACTGTATCGCCGGCGGTGAACTTGGCGCCAGGGACACCGGTGAGACCCACGAGCATTCTGGTAGTAGGAGCAACAGG
AACTCTTGGAAGGCAGATAGTACGTAGGGCTTTAGATGAGGGATATGATGTACGTTGTCTGGTCCGTCCGAGACCGGCTCCCGCAGATTT
TTTGAGGGACTGGGGGGCGACCGTCGTCAATGCGGATTTAAGCAAACCTGAAACGATACCCGCTACCCTAGTCGGGATTCACACAGTAAT
AGACTGTGCCACCGGTCGTCCCGAAGAGCCTATAAAAACAGTAGACTGGGAAGGGAAGGTAGCGCTTATCCAGTGCGCAAAGGCCATGG
GCATTCAAAAGTATGTCTTCTACAGCATCCATAACTGTGATAAACATCCGGAAGTACCTCTAATGGAAATAAAATACTGTACAGAGAAGT
TCCTGCAAGAATCTGGATTAAACCATATTACAATCAGGCTGTGCGGG

At-HCF244 pAD gene block n-term
CGAATTAGGATCCTCTGCTAGCAGAGCGGTGAACTTGGCGCCAGGGACACCGGTGAGACCCACGAGCATTCTGGTAGTAGGAGCAACAG
GAACTCTTGGAAGGCAGATAGTACGTAGGGCTTTAGATGAGGGATATGATGTACGTTGTCTGGTCCGTCCGAGACCGGCTCCCGCAGATT
TTTGAGGGACTGGGGGGCGACCGTCGTCAATGCGGATTTAAGCAAACCTGAAACGATACCCGCTACCCTAGTCGGGATTCACACAGTAA
TAGACTGTGCCACCGGTCGTCCCGAAGAGCCTATAAAAACAGTAGACTGGGAAGGGAAGGTAGCGCTTATCCAGTGCGCAAAGGCCATG
GGCATTCAAAAGTATGTCTTCTACAGCATCCATAACTGTGATAAACATCCGGAAGTACCTCTAATGGAAATAAAATACTGTACAGAGAAG
TTCCTGCAAGAATCTGGATTAAACCATATTACAATCAGGCTGTGCGGG

At-HCF244 pAD/pBD gene block c-term
ACCATATTACAATCAGGCTGTGCGGGTTTATGCAGGGCTTGATTGGACAATATGCCGTGCCGATACTAGAAGAAAAGTCAGTCTGGGGCA
CTGATGCCCCGACTAGAGTCGCTTACATGGATACGCAAGATATAGCCAGACTAACACTTATCGCGTTAAGGAATGAAAAGATCAACGGT
AAACTATTGACTTTTGCTGGGCCGCGTGCTTGGACGACGCAAGAGGTTATTACCCTTTGCGAGCGTTTGGCCGGGCAAGATGCAAACGTG
ACCACCGTTCCTGTGAGTGTATTACGTGTAACCAGACAGTTAACGAGATTCTTCCAGTGGACGAACGATGTCGCAGACAGACTAGCTTTC
TCTGAGGTCTTGTCCTCAGACACAGTATTCTCAGCGCCGATGACGGAAACTAACAGCCTTCTAGGTGTAGATCAGAAGGATATGGTTACA
CTAGAAAAATACTTACAAGATTACTTCTCCAATATATTAAAGAAGCTGAAAGATCTTAAAGCCCAAAGTAAACAAAGCGACATCTATTTC
TAATCGACTCTAGAGCCCTATAGTGAGT

pAD_OHP2_basic gene block
CGAATTAGGATCCTCTGCTAGCAGAGCACAAAAGGCTGTGGCTGTCGATGGCAAGTCTGTCACGACCGTAGAATTTCAGAGGCAGAAGG
CAAAGGAATTACAAGAGTACTTCAAACAGAAGAAATTGGAAGCGGCAGGTCAAGGACCGTTCTTCGGGTTCCAGCCAAAAAATTGATCG
ACTCTAGAGCCCTATAGTGAGT

pAD_At-OHP2_ProRS_Basic Gene Block
CGAATTAGGATCCTCTGCTAGCAGAGTATTTATCATAAGGTGTTCTCAGACGGAAGGCCCTCTTAGGAGGCCCTCAGCACCCCCCACACT
GAGGGAACCTCAAAAACCCGTTCCGCCAAGTCAACCTAGCAGTAGTCCCCCACCGAGTCCGCCCCCACAAAAGGCTGTGGCTGTCGATG
GCAAGTCTGTCACGACCGTAGAATTTCAGAGGCAGAAGGCAAAGGAATTACAAGAGTACTTCAAACAGAAGAAATTGGAAGCGGCAGG
TCAAGGACCGTTCTTCGGGTTCCAGCCAAAAAATTGATCGACTCTAGAGCCCTATAGTGAGT

At-HCF173 pAD/pBD gene block c-term
GACACTACGGCCAGCAAATACGACGGAGGAATGGATGCAAAGTTCGAATTTACGGAGACCGAGAGAGCGGAGTTCTCAGGCTACGTCTT
CACAAGAGGCGGATATGTAGAACTTAGCAAAAAACTGAGTTTGCCCCTGGGAACGACTCTGGATAGATATGAAGGACTTGTTTTGTCTGT
AGGCGGTAATGGACGTTCCTATGTGGTGATTTTGGAGGCTGGCCCTAGTTCCGACATGAGCCAGAGCAAGCAGTACTTTGCGAGAATTTC
CACCAAAGCCGGTTTCTGTAGGGTCAGAGTTCCTTTCAGCGCGTTCAGGCCAGTAAACCCGGAAGATCCGCCACTTGACCCCTTCCTAGT
GCACACTCTGACAATCAGGTTTGAGCCAAAAAGACAAAGGCCGGTCGATGGGCTTGCCGGAGCGCAACAAGATTTAAGATCCTTTTCCCT
GGTATTCGAGTATATAAAAGCCCTTCCAGCTGGACAAGAGACGGACTTTATTCTGGTGTCATGCACGGGATCTGGGGTCGAGGCCAACAG
GAGAGAGCAAGTCCTAAAGGCAAAGAGAGCTGGCGAAGACTCCTTGCGTCGTTCCGGCTTAGGGTACACCATTATTCGTCCTGGCCCCTT
GAAAGAAGAGCCTGGTGGCCAGAGAGCATTGATCTTTGACCAGGGCAACAGAATCTCTCAAGGGATTAGCTGCGCCGATGTTGCTGATA
TATGTGTAAAAGCTCTTCATGATTCTACAGCAAGAAATAAATCCTTTGACGTTTGCCACGAATACGTAGCCGAGCAAGGAATTGAGCTGT
ACGAACTAGTGGCGCACCTGCCGGATAAAGCTAATAATTATCTTACGCCTGCTCTATCAGTGTTAGAAAAGAACACGTAGTCGACTCTAG
AGCCCTATAGTGAGT

At-HCF173 pBD gene block n-term
AAAGACAGTTGACTGTATCGCCGGCCAAATTGGATGATGTAAATCCTGTCGGGCTTGGTCGTAGGTCTCGTCAAATTTTCGATGAAGTTT
GGAGAAAATTCTCAGGCTTAGGACAAATGAGTAGGACTACAAGGCCCGACGAGCAAGAAACTTTGGACTCCCTGTTGATCAGAGAGGGA
CCAATGTGTGAATTTGCGGTCCCAGGAGCCCAAAACGTGACAGTATTGGTCGTAGGAGCAACTTCCAGGATTGGTCGTATTGTCGTTAGA
AAGTTAATGCTTAGAGGATACACCGTCAAGGCACTTGTACGTAAACAAGATGAAGAAGTGATGTCTATGTTGCCAAGGAGTGTGGATATT
GTAGTGGGTGATGTGGGGGAACCATCCACTCTAAAGTCAGCCGTTGAAAGCTGCTCTAAGATAATCTATTGTGCCACCGCAAGGTCCACA
ATAACGGCTGATTTGACCCGTGTAGACCATTTAGGTGTTTACAATCTTACCAAAGCTTTCCAGGACTACAACAACCGTTTAGCCCAGTTGA
GAGCAGGCAAATCCAGTAAGTCAAAGCTACTTCTAGCGAAGTTTAAAAGTGCCGAGAGCTTGGATGGATGGGAGATCAGACAAGGAACA
TATTTTCAGGACACTACGGCCAGCAAATACGACGG

At-HCF173 pAD gene block n-term
CGAATTAGGATCCTCTGCTAGCAGAGCAAAATTGGATGATGTAAATCCTGTCGGGCTTGGTCGTAGGTCTCGTCAAATTTTCGATGAAGT
TTGGAGAAAATTCTCAGGCTTAGGACAAATGAGTAGGACTACAAGGCCCGACGAGCAAGAAACTTTGGACTCCCTGTTGATCAGAGAGG
GACCAATGTGTGAATTTGCGGTCCCAGGAGCCCAAAACGTGACAGTATTGGTCGTAGGAGCAACTTCCAGGATTGGTCGTATTGTCGTTA
GAAAGTTAATGCTTAGAGGATACACCGTCAAGGCACTTGTACGTAAACAAGATGAAGAAGTGATGTCTATGTTGCCAAGGAGTGTGGAT
ATTGTAGTGGGTGATGTGGGGGAACCATCCACTCTAAAGTCAGCCGTTGAAAGCTGCTCTAAGATAATCTATTGTGCCACCGCAAGGTCC
ACAATAACGGCTGATTTGACCCGTGTAGACCATTTAGGTGTTTACAATCTTACCAAAGCTTTCCAGGACTACAACAACCGTTTAGCCCAGT
TGAGAGCAGGCAAATCCAGTAAGTCAAAGCTACTTCTAGCGAAGTTTAAAAGTGCCGAGAGCTTGGATGGATGGGAGATCAGACAAGGA
ACATATTTTCAGGACACTACGGCCAGCAAATACGACGG

Table 1. Nucleotide sequences of cDNA fragments encoding proteins required for D1 synthesis.

REFERENCES CITED

Alonso, J.M. *et al.* (2003) Genome-wide insertional mutagenesis of Arabidopsis thaliana. *Science*, **301**, 653-657.

Barkan, A. (1998) Approaches to investigating nuclear genes that function in chloroplast biogenesis in land plants. *Methods Enzymol*, **297**, 38-57.

Barkan, A. (2011) Expression of plastid genes: organelle-specific elaborations on a prokaryotic scaffold. *Plant Physiol*, **155**, 1520-1532.

Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y.S., Bond, C.S. and Small, I. (2012) A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet*, **8**, e1002910.

Barkan, A. and Small, I. (2014) Pentatricopeptide Repeat Proteins in Plants. *Annu Rev Plant Biol*, **65**, 415-442.

Cheng, S. *et al.* (2016) Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. *Plant J*, **85**, 532-547.

Chotewutmontri, P. and Barkan, A. (2016) Dynamics of Chloroplast Translation during Chloroplast Differentiation in Maize. *PLoS Genet*, **12**, e1006106.

Chotewutmontri, P. and Barkan, A. (2018) Multilevel effects of light on ribosome dynamics in chloroplasts program genome-wide and psbA-specific changes in translation. *PLoS Genet*, **14**, e1007555.

Felder, S. *et al.* (2001) The nucleus-encoded HCF107 gene of Arabidopsis provides a link between intercistronic RNA processing and the accumulation of translation-competent psbH transcripts in chloroplasts. *Plant Cell*, **13**, 2127-2141.

Fujii, S., Sato, N. and Shikanai, T. (2013) Mutagenesis of individual pentatricopeptide repeat motifs affects RNA binding activity and reveals functional partitioning of Arabidopsis PROTON gradient regulation3. *Plant Cell*, **25**, 3079-3088.

Hager, M., Hermann, M., Biehler, K., Krieger-Liszkay, A. and Bock, R. (2002) Lack of the small plastid-encoded PsbJ polypeptide results in a defective water-splitting apparatus of photosystem II, reduced photosystem I levels, and hypersensitivity to light. *J Biol Chem*, **277**, 14031-14039.

Hammani, K., Cook, W. and Barkan, A. (2012) RNA binding and RNA remodeling activities of the Half-a-Tetratricopeptide (HAT) protein HCF107 underlie its effects on gene expression. *PNAS*, **109**, 5651-5656.

Hochmal, A. K., Schulze, S., Trompelt, K., & Hippler, M. (2015). Calcium-dependent regulation of photosynthesis ☆. *BBA - Bioenergetics*, **1847**(9), 993–1003.

Ingolia, N., Ghaemmaghami, S., Newmam, J., Weissman, J. (2009) Genome-side analysis in vivo of translation with nucleotid resolution using ribosome profiling. *Science* **324**:218-23

Hey, D., Grimm, B., Berlin, H., Fakultät, L., Biologie, I., & Pflanzenphysiologie, A. G. (2018) ONE-HELIX PROTEIN2 (OHP2) Is Required for the Stability of OHP1 and Assembly Factor HCF244 and Is Functionally Linked to PSII Biogenesis. *Plant Physiology,* **177**, 1453–1472.

Jin, H. *et al.* (2018) LOW PHOTOSYNTHETIC EFFICIENCY 1 is required for light-regulated photosystem II biogenesis in Arabidopsis. *Proc Natl Acad Sci U S A*, **115**, E6075-E6084.

Knoppová, J., Sobotka, R., Tichy, M., Yu, J., Konik, P., Halada, P., Nixon, P. J., Komenda, J. (2014). Discovery of a chlorophyll binding protein complex involved in the early steps of photosystem II assembly in Synechocystis. *The Plant cell*, ***26***(3), 1200-12.

Li, P. *et al.* (2010) The developmental dynamics of the maize leaf transcriptome. *Nat Genet*, **42**, 1060-1067.

Link, S., Engelmann, K., Meierhoff, K., & Westhoff, P. (2012) The Atypical Short-Chain Dehydrogenases HCF173 and HCF244 Are Jointly Involved in Translational Initiation of the psbA mRNA of Arabidopsis. *Plant Physiology*, **160**, 2202–2218.

Majeran, W., Friso, G., Asakura, Y., Qu, X., Huang, M., Ponnala, L., Watkins, K.P., Barkan, A. and van Wijk, K.J. (2012) Nucleoid-enriched proteomes in developing plastids and chloroplasts from maize leaves: a new conceptual framework for nucleoid functions. *Plant Physiol*, **158**, 156-189.

Majeran, W. *et al.* (2010) Structural and metabolic transitions of C4 leaf development and differentiation defined by microscopy and quantitative proteomics in maize. *Plant Cell*, **22**, 3509-3542.

McCarty, D.R., Suzuki, M., Hunter, C., Collins, J., Avigne, W.T. and Koch, K.E. (2013) Genetic and molecular analyses of UniformMu transposon insertion lines. *Methods Mol Biol*, **1057**, 157-166.

Myouga, F., Takahashi, K., Tanaka, R., Nagata, N., Kiss, A., Funk, C., Nomura, Y., Nakagami, H., Jansson, S., Shinozaki, K., Stable Accumulation of Photosystem II Requires ONE-HELIX PROTEIN1 (OHP1) of the Light Harvesting-Like Family. *Plant Physiology*, **176** (3) 2277-2291.

Pfalz, J., Bayraktar, O., Prikryl, J. and Barkan, A. (2009) Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts. *EMBO J*, **28**, 2042-2052.

Prikryl, J., Rojas, M., Schuster, G. and Barkan, A. (2011) Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proc. Natl. Acad. Sci. USA*, **108**, 415-420.

Rojas, M., Ruwe, H., Miranda, R., Zoschke, R., Hase, N., Schmitz-Linneweber, C. and Barkan, A. (2018) Unexpected functional versatility of the pentatricopeptide repeat proteins PGR3, PPR5, and PPR10. *Nucleic Acids Res*, **46,** 10448-10459.

Rochaix, J. (2006). The Role of Nucleus- and Chloroplast-Encoded Factors in the Synthesis of the Photosynthetic Apparatus. The Structure and Function of Plastids. 145-165, New York, New York, Springer, Dordrecht.

Rochaix, J. (2013). Redox Regulation of Thylakoid Protein Kinases, *Antioxidants & Redox Signaling,* **18**(16), 2184–2201.

Schult, K., Meierhoff, K., Paradies, S., Toller, T., Wolff, P. and Westhoff, P. (2007) The nuclear-encoded factor HCF173 is involved in the initiation of translation of the *psbA* mRNA in Arabidopsis thaliana. *Plant Cell*, **19**, 1329-1346.

Shen, J., Williams-Carrier, R. and Barkan, A. (2017) PSA3, a Protein on the Stromal Face of the Thylakoid Membrane, Promotes Photosystem I Accumulation in Cooperation with the Assembly Factor PYG7. *Plant Physiol*, **174**, 1850-1862.

Small, I. and Peeters, N. (2000) The PPR motif - a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem Sci*, **25**, 46-47.

Sun, T., Germain, A., Hammani, K., Barkan, A., Hanson, M. and Bentolila, S. (2013) An RNA recognition motif-containing protein is required for plastid RNA editing in Arabidopsis and maize. *Proc. Natl. Acad. Sci. USA*, **110**, E1169-1178.

Sun, Y., & Zerges, W. (2015) Biochimica et Biophysica Acta Translational regulation in chloroplasts for development and homeostasis. *Biochimica et Biophysica Acta*, ***1847***(9), 809–820.

Suorsa, M., Regel, R.E., Paakkarinen, V., Battchikova, N., Herrmann, R.G. and Aro, E.M. (2004) Protein assembly of photosystem II and accumulation of subcomplexes in the absence of low molecular mass subunits PsbL and PsbJ. *European J Biochem*, **271**, 96-107.

Swiatek, M., Regel, R.E., Meurer, J., Wanner, G., Pakrasi, H.B., Ohad, I. and Herrmann, R.G. (2003) Effects of selective inactivation of individual genes for low-molecular-mass subunits on the assembly of photosystem II, as revealed by chloroplast transformation: the psbEFLJoperon in Nicotiana tabacum. *Mol Genet Genomics*, **268**, 699-710.

Takenaka, M., Zehrmann, A., Brennicke, A. and Graichen, K. (2013) Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. *PloS One*, **8**, e65343.

Torabi, S. *et al.* (2014) PsbN is required for assembly of the photosystem II reaction center in Nicotiana tabacum. *Plant Cell*, **26**, 1183-1199.

Vainonen, J. P., Sakuragi, Y., Stael, S., Tikkanen, M., Allahverdiyeva, Y., Paakkarinen, V., Aro, E., Suorsa M., Scheller H.V., Vener A.V., Aro E.M. (2008). Light regulation of CaS , a novel phosphoprotein in the thylakoid membrane of Arabidopsis thaliana. *FEBS J*. **275,** 1767–1777.

Williams-Carrier, R., Kroeger, T. and Barkan, A. (2008) Sequence-specific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand. *RNA*, **14**, 1930-1941.

Williams-carrier, R., Brewster, C., Belcher, S. E., Rojas, M., Chotewutmontri, P., Ljungdahl, S., & Barkan, A. (2019) The Arabidopsis pentatricopeptide repeat protein LPE1 and its maize ortholog are required for translation of the chloroplast psbJ RNA. *Plant Journal,* **99**, 56–66. https://doi.org/10.1111/tpj.14308

Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T. and Nakamura, T. (2013) Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PloS One*, **8**, e57286.

Yin, P. *et al.* (2013) Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature*, **504**, 168-171.

Zoschke, R. and Bock, R. (2018) Chloroplast Translation: Structural and Functional Organization, Operational Control and Regulation. *Plant Cell*, **30**, 745-770.

Zoschke, R., Watkins, K. and Barkan, A. (2013) A rapid microarray-based ribosome profiling method elucidates chloroplast ribosome behavior in vivo *Plant Cell*, **25**, 2265-2275.

Zoschke, R., Watkins, K.P., Miranda, R.G. and Barkan, A. (2016) The PPR-SMR protein PPR53 enhances the stability and translation of specific chloroplast RNAs in maize. *Plant J*, **85**, 594-606.