PPR protein-related organellar RNA processing in *Arabidopsis thaliana*

Brody Makoto Hagino Frink

Table of Contents

Abstract

RNA processing changes the initially transcribed RNA into new and mature forms of RNA that are stable, translated correctly, and encode proper functional proteins. RNA processing comes in many forms. Modifications to the ends of RNA such as 5' capping and addition of poly-A tails, and splicing is commonly observed in nuclear-encode genes in eukaryotes. Despite their prokaryotic origin, plant mitochondrial and chloroplast transcripts are also regulated by various types of RNA processing. My research focuses on two types of RNA processing in plant organelles: RNA editing and RNA splicing.

RNA editing is a process ubiquitously found in living organisms that can insert, delete, or change specific nucleotides in RNA transcripts (Christofi & Zaravinos, 2019). Plants have an abundance of substitutional RNA editing events that take place in both the chloroplasts and mitochondria, numbering in the hundreds for flowering plant mitochondria and up to thousands in other vascular plants (Takenaka et al., 2013). Typically, these events occur through a deamination reaction from cytosine (C) to uracil (U), though the reverse U-to-C reaction is also common in hornworts, lycophytes, and ferns (Grewe et al., 2011; Knie et al., 2016; Kugita et al., 2003; Yoshinaga et al., 1996). In plants, the pentatricopeptide repeat (PPR) family of proteins is numerous, numbering from the hundreds to thousands in land plants, and functions in various forms of RNA processing including RNA editing (Barkan & Small, 2014). PPR proteins are named so due to their signature 35 amino acid repeats found clustered together at the N-terminus but have also been expanded to include additional motifs consisting of 31 and 36 amino acids. These repeats are necessary for proper target RNA recognition and binding as each repeat contains an amino acid pair which corresponds to a preferred nucleotide for that repeat (Barkan et al., 2012). In addition to these motifs, some PPR proteins also contain a DYW domain, named due to the highly

conserved triplet sequence at its C-terminus. For years it had been posited that the DYW domain was the catalytic domain of the deaminase reaction due to the conserved HXE and CXXC zinc-ion binding motif commonly present in cytidine deaminases, but it had not yet experimentally proven. Here, I show that a singular PPR protein, moss PPR56, can edit its target RNA in the absence of any other factors in a minimal buffer solution. PPR56 is a member of the DYW subgroup and contains the DYW domain. The purified PPR protein was incubated with its target RNA and editing was observed, proving that a single protein is sufficient for editing, as well as that the DYW domain must be the catalytic domain of the reaction. This and subsequent experiments further elucidated some of the preferred conditions of this PPR protein.

My second topic is on RNA splicing, a process necessary for the maturation of many nuclear-encoded RNAs which contain non-coding introns that must be removed before translation. Plant mitochondrial DNA also contains special types of introns called group II introns. Classic examples of these introns have self-catalytic activity and can splice themselves from RNA, but this function has degenerated evolutionarily and is lost in all introns located in *A. thaliana* mitochondria. Many of the intron-containing genes code for proteins related to oxidative phosphorylation in mitochondria and their proper maturation is necessary for energy production. Therefore, plants have developed numerous and diverse nuclear-encoded splicing factors that facilitate the removal of introns from pre-RNA. PPR proteins are one protein family that participate in splicing events. My research focused on a gene co-expressed with PPR proteins, including eight functionally annoated PPR proteins. This gene had not been characterized yet was conserved in angiosperms and embryonically lethal when knocked out, highlighting its importance for plant development. I found that it contributed primarily to three mitochondrial transcriptome splicing events in NADH dehydrogenase subunit RNAs. Due to its relation to PPR proteins, I named it

PPR **C**o-expressed **I**ntron **S**plicing 1 protein or "PCIS1". Together in a multinational collaboration with other laboratories, additional experiments investigated the characteristics of the protein and further solidified its role in mitochondrial splicing reactions.

List of Abbreviations

6xHis: Protein tag consisting of six histidine residues

ABI3: ABI family member 3

AMP: Adenosine Monophosphate

AMP-CP: ADP analog

AMP-PCP: ATP analog

ADP: Adenosine Diphosphate

BN-PAGE: Blue Native Polyacrylamide Gel Electrophoresis

CBB: Coomassie Brilliant Blue

CI: Mitochondrial Respiratory Complex I

Col-0: Columbia-0 ecotype plants

Comp: Complementation

DIG: Digoxigenin

DYW: Domain named based on amino acid motif of "DYW"

E894A: Non-functional PPR56 mutant with its glutamic acid (E) at position 894 changed to alanine (A)

GFP: Green Fluorescent Protein

GUS: β-Glucuronidase

HMW: High Molecular Weight

IPTG: Isopropyl β- d-1-thiogalactopyranoside

kDa: Kilo Dalton

L917A: PPR56 mutant unable to dimerize due to leucine (L) 917 changed to alanine (A)

mAU: Milli Absorbance Unit

MBP: Maltose Binding Protein

MS: Murashige and Skoog medium

NADH: Nicotinamide Adenine Dinucleotide (NAD) + Hydrogen (H)

NiNTA: Nickel-nitrilotriacetic Acid

NTP: Nucleotide Triphosphate (i.e., ATP, CTP, GTP, UTP)

OTP86: PPR56E-OTP86DYW

OXPHOS: Oxidative Phosphorylation

pABI3::PCIS1: *pcis1-1* homozyous mutant complemented with expression from the ABI3 promoter

PCIS1: PPR Co-expressed Intron Splicing 1

PPR: Pentatricopeptide repeat

SDS-PAGE: Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

THU: Tetrahydrouridine

Y2H: Yeast 2 Hybrid

Chapter 1

Introduction

RNA editing is a widely observed phenomenon in living organisms that can insert, delete, or change specific nucleotides in RNA transcripts (Christofi & Zaravinos, 2019). In plant organelles, numerous cytidines (C) are converted to uridine (U) by RNA editing. Many pentatricopeptide repeat (PPR) proteins are involved in C-to-U RNA editing in plant organelles (Hiesel et al., 1989; Takenaka et al., 2021). The PPR family of proteins found in many organisms, but extremely expanded in plants, are named as such due to the presence of their degenerated 35 amino acid repeats (Small & Peeters, 2000). Two subfamilies exist, the P subfamily which has generic P motifs and the PLS subfamily which contains P motifs along with long (L) and short (S) variants (Lurin et al., 2004; Manna, 2015; Schmitz-Linneweber & Small, 2008; X. Wang et al., 2021). RNA editing events in plant organelles are often required for changing the coding amino acid residues of mRNA, creation of start and stop codons, and removal of stop codons.

Specific site recognition for RNA editing events is achieved by the nucleotide affinity of each PPR motif, called the PPR code. Two amino acid residues at two locations in each repeat dictate preferred target nucleotide types depending on the combination. While no pair is 100% specific to any nucleotide, they strongly influence which nucleotide species preferentially bind. Much akin to the genetic code, the PPR code can be used to predict potential targets for the PPR proteins and can be changed artificially to allow for stronger affinity for certain RNA sequences or even changed to target different transcripts entirely (Barkan et al., 2012; Gully et al., 2015; Yagi et al., 2013). PLS subfamily PPR proteins are further split into subgroups with additional domains. One of these is the E subgroup which has E1 and E2 motifs whose function is still relatively vague but has been found to be important for RNA recognition and also for protein-protein interactions

(Bayer-Császár et al., 2017; Lurin et al., 2004). The E+ subgroup, in addition to the E1 and E2 motifs, has an N-terminal about 90 amino acids of DYW domain (Cheng et al., 2016). Another is the DYW subgroup which contains a DYW domain, named for its amino acid motif in the C-terminus. The domain has also been posited to be involved in RNA recognition (Ichinose & Sugita, 2018; Maeda et al., 2022) and has been a focus in PPR research as it has highly conserved zinc ion binding motifs also conserved in cytidine deaminases, which are able to convert cytosine residues into uracil. Due the conserved motifs, the DYW domain has been believed to be the enzymatic component necessary for the deamination of cytidine (Salone et al., 2007). E2 and E+ subgroups do not harbor a DYW domain and co-opt other proteins that contain N-terminally truncated DYW domains like DYW1, DYW2, and MEF8/8S (Boussardon et al., 2012; Diaz et al., 2017; Malbert et al., 2020). While truncation or mutations in the DYW domain of three different PPR proteins abolished RNA editing in planta (Ichinose & Sugita, 2018; Oldenkott et al., 2019), they could not definitively prove that the DYW domain is a cytidine deaminase. Expression of a single DYW subgroup RNA editing factor, PPR56 or PPR65 in *E. coli*, was sufficient to edit their targets. This strongly suggested a catalytic function for the DYW domain. However, the possibility remained that bacterial deaminase was recruited to the plant proteins (Oldenkott et al., 2019). Thus, in vitro editing with only a DYW-containing PPR protein would be necessary to close this open question.

The in vitro system would also help to understand the reaction mechanism of plant C-to-U editing. Although previous in vitro editing with mitochondrial and chloroplast lysates revealed several important aspects of the process, it has been impossible to exclude the undesired effects of other unrelated proteins and compounds in the lysates (Hirose & Sugiura, 2001; Takenaka & Brennicke, 2003).

Through my research here, I successfully established an in vitro editing system with purified proteins. The system uses PPR56, a PPR protein taken from the moss *Physcomitrium patens*. It targets the nad4eU272SL and nad3eU230SL sites, which lie within the *nad* genes in plant mitochondria and encode subunits of respiratory complex I. PPR56 contains an internal DYW domain and its expression induced C-to-U editing of target sites in *E. coli* (Oldenkott et al., 2019). I purified recombinant PPR56 protein expressed in bacteria and incubated it with in vitro transcribed target RNAs. Through adjustments in buffer conditions and reagent quantities, C-to-U conversion of the target site was successfully achieved using a single protein containing the DYW domain, ultimately confirming the catalytic activity of the DYW domain. Furthermore, the substitution of the DYW domain of PPR56 to that of another PPR protein, OTP86, successfully edited substrate RNA, indicating a common catalytic function for this domain.

Materials and Methods

Escherichia coli transformation for expression of PPR proteins

Recombinant protein expression constructs were made previously using a modified pETG41K vector (Oldenkott et al., 2019). *PPR56* coding sequence was taken from *Physcomitrium patens* and fused to an N-terminal MBP tag as well as a 6xHis tag (pETG41K::PPR56). In addition, the target sequence of the PPR56 protein was inserted downstream, allowing for RNA editing efficiency calculation in *E. coli*. PPR56^{E1E2}-OTP86^{DYW} fusion constructs were made by swapping PPR56's DYW domain with that of OTP86 (pETG41K:: PPR56^{E1E2}-OTP86$^{DYW)}$ (Hammani et al., 2009; Takenaka et al., 2021). DH5α was used for plasmid isolation, while BL21 strains of *E. coli* were used for induction of recombinant protein. The construct was transformed into *E. coli* through heat shock protocols. One µL of DNA construct was added to 15 µL of bacteria, left on ice for 30 minutes, heat shocked at 42°C for 30 seconds, and then grown at 37° for one hour before spreading on plates and leaving to grow overnight at 37°C.

Target RNA transcription

PCR fragments including one of the target sites of PPR56, nad4eU272SL, were amplified using the pETG41K::PPR56 as a template. For in vitro transcription of the target RNA including 33 bp upstream and 5 bp downstream of the nad4eU272SL site was amplified using primers that targeted the upstream and downstream regions of the *nad4* sequence in the pETG41K::PPR56 plasmid (T7-KS-attB2-F, SK-T7-Up-R). T7-KS-attB2-F primer also includes the sequence for the T7 promoter and KS. SK-T7-Up-R primer includes the sequence for SK (Table 1). PCR products were purified using Econospin II columns (Ajinomoto Bio-Pharma, Osaka, Japan), products were confirmed by

gel electrophoresis, and underwent DpnI digestion to remove residual vector. In vitro transcription was subsequently performed using ScriptMAX® Thermo T7 Transcription Kit (TOYOBO, Osaka, Japan) and RNA was then cleaned up using Direct-Zol RNA miniprep kits (Zymo Research, Irvine, CA, USA).

Protein isolation and purification

BL21 *E. coli* bacteria expressing the PPR56 or PPR56^{PPRE1E2}-OTP86^{DYW} construct were inoculated from LB plates into 5 mL of LB medium and 50 µg/mL Kanamycin. Pre-cultures were shaken overnight (16-20 hours) at 37°C, 180 rpm until saturated with bacteria. 50 mL of LB with Kanamycin was then inoculated with 500 µL of the pre-culture and shaken until OD600 values reached 0.4 – 0.6 (typically in 2 hours). Bacteria were set on ice for at least 5 minutes, then supplemented with 0.4 mM $ZnSO_4$ and 0.4 mM IPTG. Cultures were then shaken at 16°C at 180 rpm for 20 hours or overnight.

Bacteria in LB media were centrifuged at 5,000 rpm for 10 minutes at 4°C. Supernatant was discarded and the pellet was resuspended in lysis buffer (200 mM NaCl, 50 mM Tris pH 7.5, 2-mercaptoethanol 3.5 µL/10 mL) equaling 10% of the original culture (5 mL) on ice. 1 mM PMSF protease inhibitor was then added. The pellet was resuspended by vortexing and kept on ice during the next sonication steps. Bacteria were pulsed ten times for a total of six sets with 30-second breaks in between each set. These lysates were used in in vitro experiments but were further processed by first centrifuging at 15,000 prm for 10 minutes at 4°C. Amylose resin (New England Biolabs, Massachusetts, US) was equilibrated in lysis buffer and the lysate supernatant was added to the resin. The resin mixture was incubated at 4°C on a rotary for one hour before two wash steps

with lysis buffer and then elution with 10 mM maltose supplemented lysis buffer. This eluate was directly used in further experiments.

In vitro RNA editing assays

All experiments were conducted under the following conditions unless otherwise stated. In vitro assays with purified protein were performed in 10 or 20 µL volumes consisting of purified protein (1 µg / 10 µL), 10x deaminase buffer (100 mM Tris-HCl pH 7.5, 0.375 µL / 10 µL), purified RNA (2 ng / 10 µL or 3.7 nM, 168 bp, MW = 53,909.6 Da), RNase Out (2.5 mM, Invitrogen, Waltham, Massachusetts), and filled up with sterile distilled water. Bacterial lysate editing assays used 6.25 µL of lysate of *E. coli* expressed PPR protein instead of the purified protein. All reagents were kept on ice as much as possible to slow reactions before the intended time. Reagents were mixed by pipetting so as not to disturb protein activity. Reactions were incubated at 16°C for two and a half hours. Reactions were immediately returned back to ice and then cDNA was made by directly using 3 µL of the reaction with the ReverTra Ace qRT-PCR kit (TOYOBO, Osaka, Japan). The target sequence was then amplified using KS and SK primers with GoTaq Green Master Mix (Promega, Wisconsin, US) in 25 µL reactions. Amplified DNA was treated with 2 µL of ExoSAP (2 U ExoI, 10 U SAP) for one hour, and sequenced using the GENEWIZ service (Azenta Life Sciences, Japan). In vitro editing efficiencies were calculated by taking the ratio of the thymine trace peak vs the combined trace of both thymine and cytosine added together.

SDS-PAGE, Coomassie, and silver staining of the recombinant proteins

SDS-PAGE gels were made with 10% separating gels and 5% stacking gels. 10 µL of protein was added to 10 µL 2x SDS loading buffer (90 mM Tris-HCl pH 6.8, 20% glycerol, 2% SDS, 0.02% Bromophenol Blue) and the full 20 µL was run in each well after denaturing for 2 minutes at 95°C. 4 µL of P7008 protein marker (New England Biolabs, Massachusetts, US) was used to approximate band sizes. Gels were run at constant currents of 30 mA per gel for 60 minutes. Coomassie brilliant blue (CBB) stains were performed by first washing finished gels in water for 10 minutes before staining for one hour. Background stain was removed by washing again in 30 minutes water. Silver staining was performed using Sil-Best Stain One (Nacalai Tesque Inc., Kyoto, Japan) according to the kit instructions.

Western Blots

After SDS-PAGE was run, gels were equilibrated in transfer buffer for 10 minutes and then transferred onto nylon membranes at 20 V, 2.5 A for 40 minutes. Immobilon-P™ PVDF membranes (Merck Millipore, Tullagreen, Ireland) were blocked for 45 minutes to overnight at 4°C. Membranes were then washed in 1x TBT buffer (20 mM Tris-HCl pH 7.6, 150 mM NaCl, 1% v/v Tween-20) before 10,000x diluted MBP antibodies (New England Biolabs, Lot #0101501) were added and shaken for one hour at room temperature. Membranes were washed three times in TBT buffer then Amersham™ ECL anti-mouse antibodies (Cytiva, Massachusetts, USA) were added and shaken for 1.5 hours. After a final wash, membranes were left to react in Amersham™ CDP-Star detection reagent (Cytiva, Massachusetts, USA) for 5 minutes, covering thoroughly with

manual spreading. Final western blot images were visualized using an ImageQuant LAS 4000 (GE Healthcare, Chicago, Illinois, USA).

Size exclusion chromatography

Prior to the experiment, lysis buffer, sterile distilled water, and 20% ethanol were degassed using a vacuum chamber for 20 minutes. Protein isolation was completed as previously stated, however larger amounts of bacteria culture were used and thus falcon tubes were used in place of microcentrifuge tubes as 500 µL of protein is required for one round of chromatography. Size exclusion chromatography necessitated higher concentrations of proteins, thus Vivaspin concentration columns (Cytiva, Massachusetts, USA) were used to lower the total amount of protein volume while keeping the same amount of protein and filtering the mixture. Proteins were injected into an AKTApurifier (Cytiva) through a 500 µL injection loop and run with the degassed lysis buffer at 0.5 mL/min. The total volume was two times the total column volume or 48 mL. Samples were fractionated into 1 mL fractions and flash-frozen immediately with liquid nitrogen.

Protein crosslinking and tandem mass spectrometry

Crosslinking was performed using UV crosslinking in a UV Stratalinker 2400 (Funakoshi, Tokyo, Japan) at 254 nm. For mass spectrometry analysis, 420 µg of purified protein and 126 µg of target RNA or 6 mM GTP/THU was incubated for 10 minutes at room temperature. The mixture was transferred to the cap of the tube and underwent UV irradiation twice at 600 mJ/cm^2 each. Five-minute breaks were taken in between each to prevent overheating. Samples then underwent acetone precipitation to purify and concentrate products. Some products were immediately

visualized on SDS-PAGE gels and stained with silver staining. Protein/RNA mixtures destined for mass spectrometry continued as RNAse OUT was added (20 U/10 µL) as well as 20 mM DTT for one hour at 37°C. 40 mM of iodoacetamide was then added and left to incubate in the dark for 30 minutes. Finally, 14 µg of trypsin or chymotrypsin was added and incubated at 4 hours at 37°C. RNA was then isolated using RNeasy Mini Kits (Qiagen, Hilden, Germany). RNA was then digested in 50 µM of $LuCl_3$ overnight at 4°C. The mixture underwent one final purification step using GL-Tip SDB tips (GL Sciences, Tokyo, Japan) following manufacturer's instructions.

Further analysis including chromatography and tandem mass spectrometry was performed by Yuzo Watanabe at the Prof. Ishikawa Fuyuki laboratory at the Kyoto University Graduate School of Biostudies, Division of Integrative Life Sciences.

Results

RNA is edited to high degrees in *E. coli*

In bacteria, RNA editing with recombinant PPR56 is a robust system and has been used as a base for analyzing the catalysis of other DYW domains by swapping out those from other DYW domain-containing proteins (Maeda et al., 2022). To obtain active recombinant PPR protein available for in vitro editing, I confirmed highly efficient editing of the system in bacteria. I compared in *E. coli* editing efficiencies of the recombinant PPR56 protein and PPR56^{PPRE1E2}-OTP86^{DYW}, which uses the DYW domain from *Arabidopsis* OTP86 instead of the moss PPR56's intrinsic domain (Hammani et al., 2009). The DYW domain of OTP86 was of interest because its crystal structure had been examined (Takenaka et al., 2021). Thus, the establishment of an in vitro system with recombinant PPR56^{PPRE1E2}-OTP86^{DYW} would be advantageous to understand the detailed catalytic mechanism from a structural view. Both of these recombinant proteins with MBP and 6xHis tags were expressed under a *lac* operator system. A target RNA including the nad4eU272SL was inserted downstream of the PPR protein gene so that the translated editing factor could directly edit the target site in vivo (Figure 1). nad4eU272SL was chosen over another target of PPR56, nad3eU230SL, due to its higher observed editing in plants (100% to 70-100%) (Oldenkott et al., 2019). Recombinant PPR56 expression in *E. coli* can be induced by the addition of IPTG. RNA editing was detected at 13.8% in recombinant PPR56 transformants and 7.6% in recombinant PPR56^{PPRE1E2}-OTP86^{DYW} even without IPTG induction (Figure 2). This is probably due to some leaky expression without IPTG. After induction of the construct, editing efficiencies of both constructs rose to 100%, confirming that the recombinant PPR56 as well as PPR56^{PPRE1E2}-OTP86^{DYW} protein is very effective at eliciting editing in bacteria.

E. coli lysates expressing PPR56 edit target RNA

Mitochondrial lysates from plants have previously been reported to edit target sites in in vitro transcribed substrate RNA (Araya et al., 1992; Takenaka et al., 2007; Takenaka & Brennicke, 2003). As a stepping stone from an intrabacterial system to a purified protein-only system, I examined bacterial lysates expressing recombinant PPR56 for in vitro editing experiments. I tested two reaction temperatures, 16 and 28°C, and two incubations times of either 2.5 hours or 20 hours. Among the conditions tested, 2.5 hours at 16°C had the highest efficiency at 27.9% editing (Figure 3). Longer incubation times lowered editing efficiencies, and 28°C was more negative. Longer incubation times may lead to lower editing due to degradation of edited RNAs along with protection by bound protein or possibly the presence of the reverse reaction. These results may be because of activation at high temperatures of internal proteases and nucleases in the lysates that can target the recombinant PPR protein and target RNA. Therefore, 16°C and 2.5 hours incubation were applied for all further editing experiments.

Protein stabilizers do not stimulate lysate editing

Since lysates containing recombinant PPR56 are highly unstable (19.4% editing, 8.2% standard deviation over four measurements), the effects of four commonly used protein stabilizers were tested: betaine, mannitol, Triton X-100, and cOmplete protease inhibitor. All four stabilizers largely reduced editing efficiency (Figure 4). While these stabilizers can be useful for other proteins, they had only negative effects on recombinant PPR56, suggesting that the protein is quite selective for buffer conditions.

PPR56 protein is successfully purified

To purify the recombinant PPR56 and PPR56^{PPRE1E2}-OTP86^{DYW} proteins tagged with Hisx6 and MBP from the bacterial lysates used for the previous in vitro experiments, the solubility of the recombinant proteins was analyzed. A large portion of the 120 kDa proteins corresponding to the recombinant PPR56 protein was trapped within the insoluble pellet fraction though a prominent band could also be observed in the soluble supernatant fraction (Figure 5). I continued with purification using the soluble fractions and could isolate recombinant PPR56 protein using NiNTA resin which binds to the fused His-tag, though multiple additional bands were present in the SDS gel (Figure 5). Purification of the recombinant PPR proteins using the MBP tag reduced extra protein bands (Figure 6A). Western blotting with anti-His tag antibodies also exhibited strong signals for recombinant protein, although secondary products were still present (Figure 6B). Due to its overall purer results, MBP-based purification became the isolation method of choice for recombinant PPR proteins.

In vitro editing observed with only purified PPR protein

Using the purified recombinant proteins, I attempted to elicit RNA editing of transcribed RNA in vitro in a single reaction, using only Tris-HCl buffer to stabilize pH and RNase OUT to inhibit RNA degradation. The recombinant PPR56 was able to edit the target site in its target RNA substrate at 16.8% (Figure 7). Urea-treated recombinant protein isolated from the pellet fraction was lowered in activity. In vitro editing with recombinant PPR56^{PPRE1E2}-OTP86^{DYW} editing was also confirmed at 7.6%. This was the first time that a singular DYW domain containing PPR protein alone could edit the target site in vitro, thus proving that the DYW domain is indeed responsible for the deamination of cytidine.

Because lysates containing PPR56 are unstable, I analyzed the stability of the purified recombinant PPR56 during storage. One week of storage at 4°C reduced editing to 2.5%, which was not anticipated (Figure 8). This suggested that storage of the protein would not be feasible for very long and therefore freshly isolated protein were used for all the following in vitro editing assays. Editing efficiency is halved when the protein is vortexed, which also suggests that the activity of recombinant protein is highly unstable (Figure 9). Heat shocking the protein at 80°C for 5 minutes also eliminates protein activity.

Different incubation times were tested for $PPR56^{PPRE1E2}$-$OTP86^{DYW}$. Hour-long incubation did not improve editing efficiency, while 30 minutes of incubation exhibited 12.3% editing compared to 11.4% in 2.5 hours (Figure 10), suggesting that the editing reaction occurs rapidly after protein is exposed to substrate RNA. Although the reaction may only need 30 minutes, I continued to incubate for 2.5 hours as a standard. RNA concentration was also assessed and diluting RNA by 10 and 100 times (200 and 20 picograms per 10 µL, respectively) only resulted in minor reductions to recombinant PPR56 editing efficiency (32.1% base editing to 28.1% and 27.4% respectively, Figure 11). Hence PPR56 activity is largely RNA concentration independent.

A certain level of zinc may be necessary for efficient editing

DYW domains have similar motifs to known cytidine deaminases, HXE and CXXC which bind zinc ions needed for the actual deaminase reaction (Boussardon et al., 2014; Hayes et al., 2013; Takenaka et al., 2021). Since the presence of these ions is essential for the reaction to occur, I investigated how zinc concentration affects the editing efficiency of recombinant PPR56. During induction of the protein in *E. coli*, I commonly added 0.4 mM of $ZnSO_4$ as suggested by Oldenkott

et al. (Oldenkott et al., 2019). Luria-Bertani (LB) medium contains Yeast extract, which should contain trace amounts of zinc and therefore supplementation of zinc may not be necessary. I removed the extra $ZnSO_4$ in one trial and editing rose from 18.8% to 25.0% (Figure 12). There may be a preferred concentration of zinc and reactions can be further optimized. To clarify the direct effect of zinc ions on the editing reaction, 2 mM $ZnSO_4$ was added to the reaction. In two separate reactions, I used 5 mM 1, 10-phenanthroline, a zinc chelator, as well as its 1, 7- inactive form. All three chemicals reduced editing to 3.7, 3.3, and 3.0% editing respectively (from 18.8%), including the inactive zinc chelator. The crystal structures of DYW domains contain zinc ions, indicating that zinc is essential for catalysis (Takenaka et al., 2021). However, the negative effects of the three chemicals observed are likely unrelated to the recombinant protein and its binding state with zinc ions and due more to the change in the reaction environment and indirect effects on protein function, similar to the previously used protein stabilizers. .

NTPs stimulate PPR56 mediated RNA editing

Addition of NTPs increase the efficiency of in vitro editing using pea mitochondrial lysates, chloroplast lysates, and purified recombinant PPR65 (Hayes & Santibanez, 2020; Hirose & Sugiura, 2001; Takenaka & Brennicke, 2003). All four NTPs increased the efficiency of RNA editing of recombinant PPR56 when added to the reaction mix, with GTP showing the highest increases followed by ATP, UTP, and CTP (Figure 13). To find the optimal ATP concentration, 2, 6, and 15 mM were tested. 6 mM of ATP further promoted editing in $PPR56^{PPRE1E2}$-$OTP86^{DYW}$ but was reduced with PPR56. 15 mM showed further negative effects in both proteins (Figure 14). ATP's stimulatory effect was notably pronounced compared to Fig. 13, however this is due to further optimization of in vitro reactions. ADP at 2 mM barely affected editing while higher

concentrations along with any concentration of AMP also reduced editing. AMP-PCP (an ATP analog) heavily reduced editing while AMP-CP (ADP analog) had no effect (Figure 15).

GTP was a strong stimulant of the editing reaction overall with large increases to editing efficiencies in both PPR56 and PPR56^{PPRE1E2}-OTP86^{DYW}. Both induced nearly two-fold increases even at 2 mM concentration which was further amplified at 6 mM (Figure 16). Similar to ATP, positive effects of GTP lowered at 15 mM, but was still slightly higher than the effect at 2 mM. Effects of GTP were observed consistently over further experiments, solidifying its position as an enhancer of PPR protein mediated editing. The direct effect of NTPs on editing proteins is still unknown but they may interact allosterically. While ATP is most commonly found in the mitochondria with editing proteins in vivo, other NTPs share similar structures and in the case of GTP can enhance editing to a greater degree than ATP.

THU also enhances the editing reaction

Tetrahydrouridine (THU) had been used in editing experiments as cytidine deaminase inhibitor (Carlow & Wolfenden, 1998). The structure of THU is similar to cytidine making it a competitive inhibitor for proteins utilizing cytidine (Stoller et al., 1978). Thus, I expected that adding THU to the editing reaction should compete with the reaction and lower editing activity. However, this editing with recombinant PPR56 exposed to 2 mM of THU increased from 30.5% to 49.2% (Figure 17A). This increased further as the concentration was increased up to the highest tested concentration of 30 mM. Editing enhancement of PPR56^{PPRE1E2}-OTP86^{DYW} was more variable, with one of the tested preparations showing similar results to PPR56 (Figure 17B), while a few other tests showed small decreases (Figure 17A). THU also stimulates PPR56^{PPRE1E2}-OTP86^{DYW},

but the observed variation may be due to unstable reaction conditions. THU does not show the same level of inhibition at higher concentrations and its effect on PPR56^{PPRE1E2}-OTP86^{DYW} structure was observed in (Takenaka et al., 2021).

Recombinant PPR56 form multimers in the solution

Size exclusion chromatography of the recombinant PPR65, another PPR protein used in in vitro experiments, had been shown to exist in both monomeric and polymeric forms (Hayes & Santibanez, 2020). To investigate the status of the recombinant PPR56, I performed size exclusion chromatography with a molecular weight marker (Figure 18) and PPR56 (Figure 19). Recombinant PPR56 displayed two peaks, one at a volume of 8 mL and another at 16 mL. The 8 mL peak corresponding to over 700 kDa could be agglutination of the protein or an oligomer probably consisting of six monomers (Figure 19). The expected molecular size of the peak at 16 mL is too low to be the monomer and it is most likely MBP fragments approximately 42 kDa. Therefore, the monomer of PPR56 was very minor and was not detected in our size exclusion data. Aggregation of the PPR56 protein may be correlated with the decline in editing efficiency depending on the storage time of the protein. PPR56^{PPRE1E2}-OTP86^{DYW} and its point mutant at the putative dimerization interface in the DYW domain, L917A, showed no peaks for the monomeric form (Figure 20) (Takenaka et al., 2021). As L917A mutants should have impaired dimerization at the DYW domain, observed protein aggregation is more likely a recombinant protein-specific byproduct.

Search for nucleotides and target RNA binding sites in recombinant PPR56

Although crystal structures of DYW domains implied the pathway for substrate RNA on PPR proteins (Takenaka et al., 2021; Toma-Fukai et al., 2023), their exact protein-RNA interacting sites are unclear. In addition, binding sites of GTP and THU, which activate the reaction, within the PPR56 protein were also not identified, though they change the structure of the DYW domain. To examine this, I conducted UV crosslinking of PPR56 to its target RNA and nucleotides. Ultraviolet light irradiation stably binds amino acid residues and nucleotides in close proximity. Tandem mass spectrometry of the nucleotide crosslinked polypeptides allows the determination of RNA interaction sites within the PPR proteins.

At first, to optimize the intensity of UV light to induce protein-nucleotide crosslinking but not protein-protein crosslinking, multiple intensities of UV light were irradiated onto purified protein with or without substrate RNA present. It was evident that recombinant PPR56 protein began to crosslink with itself as the 120 kDa monomer band disappeared with increasing UV irradiation without the appearance of larger bands (Figure 21A). However, the monomer band did not disappear after UV irradiation when 6 mM GTP was added to the sample (Figure 21B). This also held true for all other NTPs and several of their analogs (Figure 22). NTPs seem to be able to stabilize the monomeric form of PPR56, which is one possible reason for how ATP and GTP may stimulate RNA editing. As analogs also stabilize the monomeric form, the chemical activity of the NTP may also be important to determine the extent of enhancement of editing. PPR proteins are known to associate with other editing factors to form an active editing complex, but in the case of a single PPR editing a RNA target, monomers appear to be active. Smaller polymers such as dimers have been observed in other factors, but as PPR56's polymerized form consists of upwards of six monomers, its likely that this is not coordinated and is due to protein clumping.

Two peptidases, trypsin and chymotrypsin, were tested for recombinant PPR56 protein to compare the polypeptide coverages by mass spectrometric analysis. Chymotrypsin is similar to trypsin but is less specific as to what it can cleave, cutting at the C-terminal side of tyrosine, tryptophan, and phenylalanine as opposed to lysine or arginine in trypsin. The peptide coverage rate by chymotrypsin cleavage was 92%, compared to 80% for trypsin cleavage (Figure 23). The addition of GTP or THU did not drastically affect the coverage rate (Figure 24). PPR56 recombinant protein was incubated in substrate RNA, GTP, or THU for ten minutes before UV crosslinking. This was followed by digestion with chymotrypsin, RNA purification, and digestion by $LuCl_3$, which was tested and found to digest total yeast RNA at 50 mM and did not digest PPR56 recombinant protein at the same concentration. (Figures 25 and 26) (Matsumura & Komiyama, 1997). Unfortunately, neither GTP nor THU was detected to be bound to any amino acid in tandem mass spectrometry. When substrate RNA incubated with recombinant protein, peptide coverage was too poor for analysis (Figure 27), so further experiments will be required in the future.

Discussion

First in vitro editing using PPR56 and DYW as a deaminase

The experiments detailed were the first instance of in vitro RNA editing using recombinant PPR56 protein and its RNA target, nad4eU272SL. Mitochondrial and chloroplast lysates taken from plants, as well as *E. coli*-based editing systems had previously been used to assess editing in vitro (Hirose & Sugiura, 2001; Oldenkott et al., 2019; Takenaka & Brennicke, 2003). Lysates contain unknown cofactors that may be required for RNA editing reactions, thus in vitro editing with a purified recombinant protein was necessary. *Physcomitrium patens* PPR56 contains an innate DYW domain. The DYW domain is important for target recognition and it was postulated that the domain was the active cytidine deaminase required for RNA editing from cytosine to uracil (Okuda et al., 2009; Salone et al., 2007). Studies using *E. coli*-based RNA editing systems showed that single nucleotide mutations in conserved regions of the DYW domain abolish target editing completely (Ichinose & Sugita, 2018; Nakamura & Sugita, 2008; Oldenkott et al., 2019), suggesting its primary role in cytidine deamination. A DYW domain-containing PPR protein had not been shown to edit without cofactors in lysates and thus I set out to show this. Both recombinant PPR56 and PPR56^{PPRE1E2}-OTP86^{DYW} proteins edited substrate RNA in minimal buffer conditions, proving that DYW-containing PPR proteins are self-sufficient and that the DYW domain has cytidine deaminase activity.

In vitro editing is highly sensitive to environmental conditions

Recombinant PPR56-mediated editing varied in basal editing rates but improved as reaction conditions and protocols were optimized. Protein stabilizers did not improve stability of editing.

Addition of certain reagents would abolish editing, emphasizing that the protein requires strict conditions for activity in buffer medium. As some of the reagents are well known protein stabilizers such as betaine and detergents like Triton-X 100 it is unclear as to how they affect recombinant PPR56 protein activity. A-to-I RNA editing by adenosine deaminase has been shown to be sensitive to pH conditions (Malik et al., 2021). Altered pH due to the addition of exogenous reagents could have negatively affected DYW domain-mediated editing efficiency, though dramatic pH changes in the reaction solution are unlikely as it contains 10 mM Tris-HCl buffer (pH 7.5). Further optimization of pH conditions may also yield higher editing of substrate RNAs. The RNA editing reaction by recombinant PPR56 is very quick, with maximal editing observed within thirty minutes. This may also further be optimized in the future as extended incubation times may be unneeded or detrimental..

DYW domains require the presence of zinc ions in order to stabilize its catalytic site and has been confirmed with X-ray crystallography in $OTP86^{DYW}$ (Boussardon et al., 2014; Hayes et al., 2013; Takenaka et al., 2021). Mutations in the zinc binding regions abolish editing. Thus, recombinant PPR56 and $PPR56^{PPRE1E2}$-$OTP86^{DYW}$ likely require the presence of zinc for the deamination reaction to occur. During induction, a zinc source in the form of $ZnSO_4$ is added. Removal of this ion source did not have any negative effect on editing efficiencies. This is probably because LB medium contains trace amounts of zinc already. When zinc was added in in vitro conditions, RNA editing efficiency was reduced. Addition of a zinc chelator and also an inactive analog reduced overall editing. The balance of zinc ions is important in determining the activity of the DYW domain and excess zinc may be equally as damaging as its absence. As the control inactive zinc chelator also reduced editing efficiencies however, these effects may be due to other changes in medium conditions such as pH, which needs to be examined in the future.

NTPs enhance RNA editing through unknown mechanisms

NTPs have stimulatory effects on RNA editing (Takenaka & Brennicke, 2003). Recombinant PPR65 editing in vitro increased in efficiency at 2 mM ATP, AMP-PCP, and AMP-CP (Hayes & Santibanez, 2020). Therefore, PPR56 was hypothesized to react similarly to the addition of NTPs. When tested, recombinant PPR56 increased its RNA editing efficiency at 2 mM ATP, but these effects were quickly lost as concentration increased to 6 mM and over. Recombinant $PPR56^{PPRE1E2}$-$OTP86^{DYW}$ was more durable with enhanced editing up to 6 mM but fell off at higher concentrations. ADP and ATP analogs, AMP-CP and AMP-PCP, also increased editing efficiencies in recombinant PPR65, but failed to do so with PPR56. The most effective cofactor was GTP, which drastically increased RNA editing capabilities of both recombinant PPR56 and $PPR56^{PPRE1E2}$-$OTP86^{DYW}$ at all tested concentrations, though 15 mM and onwards reduced the stimulatory effect. THU is a uridine analog and a known competitive inhibitor of cytidine deaminases (Carlow & Wolfenden, 1998). Addition of THU to the in vitro editing reactions increased the editing efficiencies of both recombinant proteins, contrary to its known function.

NTPs can greatly increase the amount of RNA editing that takes place. The degree of stimulation differs between each NTP and is also dependent on the acting protein. How NTPs mechanistically affect editing is not known. To elucidate a mechanism behind the stimulatory effect, I attempted to crosslink THU or GTP to recombinant editing proteins and then observe where binding may occur. This may have given us a hint as to what residues are important, but bound THU and GTP could not be found when analyzed with tandem mass spectrometry. I believe this is a problem with the protocol rather than the lack of interaction. Crosslinking and analysis protocols were based on those performed by Bae, et. al. 2020, which used hydrofluoric acid (HF)

to degrade substrate RNA to nucleotides before downstream analysis by tandem mass spectrometry (Bae et al., 2020). Degradation to single nucleotides rather than short polynucleotides is important because longer chains would muddle mass spectrometry's ability to distinguish peptides bound to RNA nucleotides. Due to HF's volatility, an alternative chemical, $LuCl_3$, was used based on papers detailing its similar activity (Matsumura & Komiyama, 1997). In my tests it did not seem to impact protein abundance in silver stained SDS-PAGE gels and size exclusion chromatography data. However, the use of this chemical may have further degraded peptides or nucleotides, which may be why the presence of bound nucleotides was not found when using substrate RNA.

NTPs are not intrinsically required for deamination events to take place (Takenaka et al., 2007; Takenaka & Brennicke, 2003). Recombinant PPR56 can edit in the absence of exogenous NTPs. The action of NTPs is not related to energy but possibly through protein stabilization. While crosslinking of protein to RNA was unsuccessful, NTPs could protect recombinant PPR56 protein from crosslinking with itself or agglutination under UV irradiation, as the monomeric band of recombinant PPR56 at 120 kDa was still visible when NTPs were added. This was true of all NTPs added as well as analogs. Since ATP and GTP are stimulatory reagents, it's likely that the monomeric form of recombinant PPR56 and $PPR56^{PPRE1E2}$-$OTP86^{DYW}$ are the enzymatically active forms of the protein. Recombinant protein examined by size exclusion chromatography used one day old protein and lacked a band corresponding to monomers (Figure 19). This may suggest that the monomeric fraction decreases as the protein ages and thus results in lowered editing efficiency. NTPs may stabilize the monomer form, keeping the active sites of PPR56 exposed and available for cytidine deaminase activity by DYW. UTP, CTP, and analogs may keep the protein in its monomeric state, but may inhibit activity directly through active site competition or allosteric

inhibition. THU is expected to act similarly but is instead stimulatory, suggesting that its effects may be due to another mechanism yet unknown.

As described in (Takenaka et al., 2021), OTP86's DYW domain contains a gating domain that is responsible regulation of enzymatic activity. The domain exists in two forms, an active and inactive form that was shown to change in response to addition of THU when observing crystal structures, but also ATP and GTP when examined using differential scanning fluorimetry. The three enhancers may have a two different effectson the DYW domain. One that maintains PPR56 in a monomeric form such that RNA interactions can freely occur and putative active sites are not blocked by steric inhibition by protein-protein dimers and/or oligomers. The second effect is to shift DYW gating domains from their inactive to active forms, such that active sites are available for enzymatic activity. Regulation of PPR proteins by THU and other NTPs is advantageous as many of their targets are located in the mitochondria where NTPs, namely ATP, production takes place. Suboptimal concentrations of ATP could stimulate the activity of RNA editing factors and thus increase production of oxidative phosphorylation-related complexes needed for ATP production to take place, but be reduced when ATP is in excess to avoid wasting energy and other negative effects. This is corroborated by recombinant protein in vitro, as editing proteins have an optimal concentration of ATP and GTP where editing activity is at its peak, but begins to fall off when concentrations increase. Expression and translation of editing proteins may also be controlled by ATP levels in the cell, but stability and abundance of the protein is unaffected as seen in NTP crosslink experiments.

PPR56 as a new system for the study of PPR proteins

The experiments performed here helped create and test a viable in vitro editing system that could study individual PPR proteins without the need and influence of other co-factors. Even with the advent of other PPR proteins that can edit in vitro, recombinant PPR56 has additional advantages (Hayes & Santibanez, 2020). PPR56 edits in in vitro lysate conditions at a higher rate compared to PPR65, 97% vs 76% in previous papers (Oldenkott et al., 2019). While off-targeting in PPR56 is more rampant compared to PPR65, PPR56's PLS domains maintain high editing even when more heavily conserved residues are changed. For the creation of designer editing systems, this offers more opportunities for unique sequences as targets. Recombinant PPR56 also edits in minimal buffer, reducing the need for more complex conditions. Probably one of the biggest advantages is that this system has been proven to work with other DYW domains. The DYW subgroup of PPR proteins are numerous and each have DYW domains with slight variations that may change preferred editing targets or efficiency. Use of OTP86's DYW domain in these experiments, though slightly reduced compared to recombinant PPR56, had respectable amounts of editing even when fused to $PPR56^{PPRE1E2}$ and experiments produced measurable changes in editing, which could help elucidate more about specific mechanisms of editing of each DYW domain. The efficacy of fusing other DYW domains to PPR56 is high and potentially even singular cytidines deaminases like DYW1 could be tested. PPR56 has even been utilized in human cells and could edit target sites (Lesch et al., 2022). Use of PPR56 and the expression systems developed will only help further our understanding of how these domains operate and of RNA editing as a whole.

Chapter 2

Mitochondrial genomes (mtDNAs, mitogenomes) are separate entities from the genomic DNA in the nucleus, which have important aspects for the functioning of eukaryotic cells. In the endosymbiotic theory, mitochondria started as engulfed prokaryotes and contained their own genome and machinery necessary for energy production. As mitochondria changed and became more reliant on their host cell, they also lost much of their own genome needed to maintain their structure and function (Brennicke et al., 1993). The mitogenomes of land plants are large, complex in structure, and harbor more genes than their counterparts in animalia (Best, Mizrahi, et al., 2020; Bonen, 2018; Gualberto & Newton, 2017). The mitogenes encode proteins related to organellar gene expression (tRNAs, rRNAs and ribosomal proteins), as well as ATP biosynthesis and OXPHOS related functions. A notable feature of the mitogenomes in plants involves extensive RNA processing steps (e.g., RNA trimming, editing, or splicing) that are required for the maturation of primary organellar transcripts, before they can be translated, in a similar manner to the expression of nuclear genes found in eukaryotes (Zmudjak & Ostersetzer-Biran, 2018). Our work relates to the regulation of splicing of mitochondrial group II introns.

Group II introns consist of autocatalytic RNA elements and their cognate intron-encoded proteins (IEPs, or maturases) that assist in the splicing of their host pre-RNAs (Michel & Feral, 1995). This family of introns is highly prevalent in the mitogenomes of plants but these have degenerated over time, such as losing various elements that are considered to be essential for their self-splicing activity and also lacking the IEPs that function in their in vivo splicing (Bonen and Vogel, 2001). The intron degeneration and absence of IEPs was followed by the acquisition of many nuclear-encoded factors needed for the processing of mitochondrial group II intron splicing in plants (Brown et al., 2014). These include IEPs/maturases that have been translocated into the

host nuclear genomes (Mizrahi et al., 2022), as well as other RNA binding proteins that were recruited from the host genomes during their evolution to function in the splicing organellar group II intron in plants (Zmudjak & Ostersetzer-Biran, 2018). One family of proteins that is key for mitochondrial RNA metabolism and splicing in plants is the pentatricopeptide repeat (PPR) protein family (Small & Peeters, 2000b). PPR proteins gain their name from their canonical 35 amino acid repeats that confer the recognition of a specific RNA base (Barkan et al., 2012; Gully et al., 2015; Yagi et al., 2013). PPR proteins are classified into P and PLS classes based on the types of PPR motifs contained within each protein (Lurin et al., 2004; Schmitz-Linneweber & Small, 2008). PLS class PPR proteins consist of canonical P type, short S type, and long L type PPR motifs and are known for their function in C-to-U RNA editing within mitochondria and chloroplast transcriptomes. They utilize a DYW deaminase domain located within the protein or recruited separately to catalyze the reaction necessary for the conversion of cytidine to uridine (Boussardon et al., 2014; Hayes & Santibanez, 2020; Oldenkott et al., 2019; Takenaka et al., 2021). P-class PPR proteins, on the other hand, have been shown to be necessary for various post-transcriptional events including RNA stabilization, intron splicing, 5' and 3' processing, and translation. As most of them contain no additional domains except for PPR motifs, their main function likely involves binding to target RNA with high sequence specificity (Schmitz-Linneweber et al., 2006; Schmitz-Linneweber & Small, 2008).

Here, we describe the analyses of the PPR co-expressed intron splicing1 (*PCIS1* gene, At5g25500) factor, which was investigated due to its co-expression patterns with multiple PPR proteins that are postulated to function in mtRNA metabolism. As PCIS1 functions are essential during embryo development, the phenotypic and molecular characterization of its roles in *Arabidopsis* plants rely upon the expression of the gene in *pcis1* mutants expressing the gene under the *ABI3* promoter,

which is active only during embryogenesis. The functionally rescued p*ABI3:PCIS1* mutant lines displayed stunted growth and developmental phenotypes. Analysis of the RNA and protein profiles revealed impaired intron splicing of three different *nad* transcripts, as well as the loss of mitochondrial respiratory complex I. PCIS1 is a novel mitochondrial splicing cofactor, which likely cooperates with various PPR proteins (and presumably other factors) in the maturation and expression of NAD subunits.

Materials and Methods

Plant growth conditions and genotyping

Seeds of *Arabidopsis thaliana* wild type (Col-0) and T-DNA insertional mutant lines, i.e., *pcis1-1* (SALK_152244), and *pcis1-2* (SALK_094043C) (Figure 1A), were obtained from the Arabidopsis Biological Resource center (ABRC, Ohio, USA). The seeds of the mutant and wild type plants were surface sterilized with 70% ethanol for two minutes, and in 10% sodium hypochlorite solution for ten minutes, and then rinsed in MilliQ purified water before sowing onto soil or half-concentration of Murashige and Skoog agar plates (MS, Wako, Tokyo, Japan). The seeds were vernalized for 3 days at 4°C before being transferred into *Arabidopsis* growth chambers. Humidity was kept at 55% and temperatures at 23°C under long-day (16 hrs light: 8 hrs dark) conditions. T-DNA insertion lines were confirmed by PCR-based genotyping, using the LB1.3 and *PCIS1* specific primer sets At5g25500-5UTRF144 and At5g25500R742 for *pcis1-1* and At5g25500-3'UTR-R1547 for *pcis1-2*.

Transformation of Arabidopsis plants

The intronless *PCIS1* coding region was obtained by PCR using total *A. thaliana* DNA and primers designed with 15 bp adaptors to overlap with the end of the vectors. The modified pENTRA1 entry vector (Thermo Fischer Scientific, Massachusetts, US), pENTR41b in which two outsides of the gateway cassette sequence TTTAAAGG and TTCGCGGCCGCACTCGAGATATCTA, were replaced to GAATTCTGTACAGGCCTG and GCTTGCGGCCGCACTCGAG, respectively, to insert StuI and HindIII sites (Thermo Fischer Scientific, Massachusetts, US). This vector was linearized using StuI and HindIII and the amplified PCR fragment was inserted into the vector using NEBuilder (New England Biolabs, Massachusetts, US) or In-Fusion cloning kit (TAKARA, Japan). LR reactions with Gateway™ LR Clonase™ II Enzyme mix (Thermo Fischer Scientific, Massachusetts, US) were performed to transfer the insert fragments from the entry vector to destination vectors. A construct for expressing *PCIS1* embryonically were created by cloning of the full-length gene amplified with At5g25500-ATGiFF and At5g25500iFR into pENTR41b and transferred to the plant expression vector pH7WG containing the ABI3 promoter (Aryamanesh et al., 2017). A second complementation construct utilizing the *PCIS1* native promoter was created by incorporating the region 1,000 bp upstream of the *PCIS1* start codon along with the coding region. This was amplified with primers, At5g25500Up-8879943-F and At5g25500iFR. The construct was inserted into pENTR41b and then transferred into pGWB507 (Nakagawa et al., 2007). Finally, GFP fused with *PCIS1* was cloned into the pGWB5 destination vector for a third construct for GFP-based localization. This contains a 35S promoter for constitutive expression as well as a C-terminally located sGFP sequence. Primers used for the amplification step were At5g25500-ATGiFF and At5g25500nostopiFR. All constructs were transformed into *Agrobacterium tumefaciens* and used for transformation into *A. thaliana* plants through floral dip

protocols (Zhang et al., 2006), though *Agrobacterium* pre-cultures for floral dips were only grown overnight and bacteria were directly pipetted onto buds. Transformant plants were screened using half-concentration Murashige and Skoog medium supplemented 10 µg/mL Hygromycin B and 0.8% agarose. Transformants were grown in petri dishes, waiting 3-4 weeks before being transferred to soil and genotyped. All primers used in these experiments are listed with sequences in Supplemental Table 1.

GFP localization assay

Performed by Shohei Yamaoka at the Graduate School of Biostudies, Kyoto University, Kyoto, Japan.

Roots of 14-day-old seedlings of the transgenic plants expressing *PCIS1* fused with sGFP were stained with a mitochondrial marker, 1 µM MitoTracker Orange (Thermo Fischer Scientific), and examined by a confocal laser scanning microscope (FV1000, Evident (formerly Olympus), Japan). Fluorescent images of the root epidermal cells were acquired with the following excitation and emission wavelength: for sGFP, 488 nm and 510-530 nm; for MitoTracker Orange, 543 nm and 555-655 nm, respectively.

GUS Assays

The *PCIS1* gene (lacking its stop codon) with its native promoter 1,000 bp upstream were inserted into the pGWB533 vector to create a *PCIS1pro*::*PCIS1-GUS* fusion construct. Transformation into *A. thaliana* (Col-0 ecotype) mirrored the protocols mentioned above. Tissue fixation was performed by adding plant tissues to cold 90% acetone and incubating for 10 minutes at 4°C,

followed by two wash steps with water. GUS staining solution containing 1 mM X-Gluc, 100 mM $NaPO_4$ (pH 7.0), 10 mM EDTA, 5 mM potassium ferricyanide, 5 mM potassium ferrocyanide, and 0.1% Triton X-100 was added to the plant tissues and subjected to a vacuum for 15 minutes. Samples were incubated at 37°C overnight in the dark and then washed with 70% ethanol and 1:6 acetic acid to ethanol, before a final wash in 70% ethanol and observation.

Seed viability check

Col-0 plants and *pcis1-1* heterozygous plants were plated on ½ MS medium and grown for 1-1.5 weeks until germination. Seeds were arranged in a matrix to allow for easy viewing of growth. The total number of seeds, germinated and inviable, were counted and then compared against the total number of inviable seeds for both lines. The percent viability was calculated for each trial, averaged, and the standard deviation was calculated.

RNA isolation

Maxwell 16 LEV Plant RNA kits (Promega, Wisconsin, US) were used to isolate RNA from fresh plant tissue. Cut leaves from Col-0 and the *ABI3* promoter complemented *pcis1-1* plants (p*ABI3*::*PCIS1*) were frozen in liquid nitrogen prior to shaking. Whole plants were used for RNA-seq. Sample tubes contained a porcelain bead for grinding. Samples were then shaken in a Shake Master Neo shaker (BMS, Tokyo, Japan) for two 20-second cycles. Homogenization buffer was added and subsequent steps followed the included kit protocol.

RNA editing analysis

RNA was purified from leaves of 4 weeks old Col-0 and 6 weeks old partially complemented pABI3::PCIS1 plants aforementioned protocol. RNA editing in mitochondria was evaluated direct sequencing of the cDNA products as described previously (Takenaka, 2022). This book chapter also contains all primer information for reverse transcription, PCR reaction, and sequencing reaction. Sanger Sequencing data were obtained commercially and compared for analysis (Azenta Japan). Ratios between heights of C and T signals were calculated with the DNADynamo software (BlueTractorSoftware, Ltd., United Kingdom).

RNA-Seq analysis

RNA was isolated from 4 weeks old Col-0 and 6 weeks old pABI3 rescued *pcis1-1* plants used the aforementioned protocol. RNA-seq was then performed using GENEWIZ NGS RNA-seq. Mapping data was analyzed using the Integrated Genomics Viewer software (Broad Institute, California). All RNA-sequencing data are available at Sequence Read Archive under BioProject accession number PRJDB16342.

Blue native (BN) PAGE analysis

Performed by our collaborators at Oren Ostersetzer-Biran's laboratory at The Hebrew University of Jerusalem, Jerusalem, Israel.

Crude organellar membranous fractions were obtained from 200 mg MS-grown seedlings, as described previously (Pineau et al., 2008; Sultan et al., 2016). Blue native (BN)-PAGE of mitochondria preparations was performed as described previously (Pineau et al., 2008). The

organellar fractions (equivalent to 25 mg-FW of MS-grown seedlings) were solubilized with n-dodecyl-β-D-maltoside (DDM, 1.5% (w/v)) for 30 min on ice. The solubilized membranous fractions were pelleted (8 min at 22,000 g), resuspended in 0.2% Serva Blue G solution, and loaded onto a native gel. For non-denaturing-PAGE immunoblotting, the protein was transferred to a PVDF membrane (Bio-Rad, California, USA) in Cathode buffer (16 h at 4°C at 40 mA) and incubated with the primary antibodies. Detection of proteins and native organellar complexes was carried out by immunoblotting assays, following incubation with horseradish peroxidase (HRP)-conjugated secondary antibodies (Table 2).

Protein extraction and analysis

Performed by our collaborators at Oren Ostersetzer-Biran's laboratory at The Hebrew University of Jerusalem, Jerusalem, Israel.

Crude organellar membranous fractions were solubilized with 3x protein sample buffer and subjected to SDS-PAGE, essentially as described previously (Shevtsov-Tal et al., 2021). For immunoblotting, the proteins were transferred from the gel onto a PVDF membrane and blotted overnight at 4°C with specific antibodies (Table 2). Detection was carried out by chemiluminescence assays after incubation with an appropriate horseradish peroxidase (HRP)-conjugated secondary antibody.

RT-qPCR analysis

Performed by our collaborators at Oren Ostersetzer-Biran's laboratory at The Hebrew University of Jerusalem, Jerusalem, Israel.

Quantitative PCRs were performed as described previously (Shevtsov-Tal et al., 2021). The relative accumulation of various organellar transcripts was analyzed by a LightCycler 480 (Roche, Basel, Switzerland), using LightCycler 480 SYBR Green I Master mix in a final volume of 6 µL, with 3 µL (1.25 µM) of specific oligonucleotides (Table 3) designed to intron-exon regions (pre-mRNAs) or exon-exon (mRNAs) regions. Reactions were performed in triplicates in the following conditions: pre-heating at 95°C for 10 min, followed by 45 cycles of 10 sec at 95°C, 10 sec at 58°C and 10 sec at 72°C. The nuclear genes *ACTIN2* (At3g18780), *GAPDH* (At3g04120), *18S* rRNA (At3g41768) and the mitochondrial 26S rRNA (*rrn26*, At3g18780) were set as calibrators in the quantitative analyses.

Northern blot analysis

DNA probes containing DIG-11-dUTP (Jena Bioscience, Jena, Germany) were created from cDNA made from Col-0 total RNA. cDNA libraries were created using ReverTra Ace qPCR RT kits (Toyobo, Osaka, Japan). Primers pairs, whose products cover a majority of the spliced transcript, were made by directly adding 50% DIG-11-dUTP to the PCR reaction mix utilizing KOD FX Neo (TOYOBO, Japan) (Table 4). Targets for the probes included *nad2*, *nad4*, *nad5b*, *nad7*, *ccmFC*, *cox2*, *rpl2*, and *rps3*. A *nad2a* and a *nad5* probe only covering the last three exons was made due to difficulties in amplifying a little and no amounts of product were amplified with the primer sets for full-length probe, respectively. Specific hybridization of the probes was confirmed against a non-DIG containing control DNAs and purified using Econospin columns (Ajinomoto Bio Pharma Service, Osaka, Japan). Northern blots followed protocols described in previous papers (S. L. He & Green, 2013; Keren & Ostersetzer-Biran, 2011; Rio, 2014; Streit et al., 2009). Two µg of total RNA isolated from Col-0 and the *ABI3* promoter rescued transformant

with *pcis1-1* homozygous background (p*ABI3*::*PCIS1*) were run on 1.2% agarose, 1.1% formaldehyde gels for an hour and a half at 50 V. Gels were then transferred to nylon membranes overnight and then UV crosslinked at 254 nm for a total of 1200 mJ. DNA probes were then added to the membranes and allowed to hybridize overnight at 50°C. Visualization was performed using anti-digoxigenin-AP, Fab fragments (Roche, Basel, Switzerland) and CDP-Star (GE Healthcare, Illinois, US). An ImageQuant LAS 4000 machine (General Electric, Massachusetts, US) was used in the final step to detect the chemiluminescence.

RT-PCR and product pattern observation

Plant RNA from Col-0 and p*ABI3*:*PCIS1* lines were subjected to a reverse transcriptase reaction using ReverTra Ace qPCR RT kits (TOYOBO, Japan). The amount of RNA was normalized to 300 ng before the RT step was performed. cDNA was then amplified using the same primer sets used for northern blots for a round of PCR (Table 4). PCR product patterns were then visualized on a 1% agarose gel and compared.

Alexander staining

Stains were performed using stage 12 buds from Col-0 and *pcis1-1* heterozygous plants, with the same protocol used in (Peterson et al., 2010). The modified Alexander stain was dropped onto dried buds using a pipette and heated lightly with a lighter. Pollen was then viewed immediately under the microscope.

Yeast two hybrid assay

The coding sequence of *PCIS1* was cloned into both pGBKT7 and pGADT7 vectors, while full length coding sequences of ABO6, ABO8, CFM9, MatR, mCSF1, MISF74, MTL1, nMAT2, OZ2, PMH2, and SLO3 were cloned into pGADT7 vector using NEBuilder (New England Biolabs, Massachusetts, US). Accession numbers are available in Table 6. AH109 and Y187 strains of yeast were transformed with the assembled pGADT7 and pGBKT7 constructs, respectively. Yeast transformation was performed according to protocols written in (Gietz & Schiestl, 2007). Yeast mating and culturing followed protocols for Matchmaker Y2H systems (Clontech, California, USA). Mated yeast was pre-cultured and grown to an OD600 of 0.5 before plating on double and triple dropout media and incubation at 37˚C for four days.

PCIS1-GFP co-immunoprecipitation and RNA immunoprecipitation (RIP) assays

Col-0 and plants expressing *PCIS1-GFP* were grown for 6 weeks and GFP fusion protein were isolated using µMACS™ GFP isolation kits (Miltenyi Biotec, Germany). Three seedlings were frozen in liquid nitrogen and disrupted using a porcelain bead and shaken in a Shake Master Neo shaker (BMS, Tokyo, Japan) for two 20-second cycles. Samples were homogenized in RIP lysis buffer (20 mM HEPES-KOH pH 7.6, 100 mM KCl, 20 mM $MgCl_2$, 1 mM DTT, 1% Triton X-100, 1X cOmplete protease inhibitor EDTA-Free [Roche, Basel, Switzerland]) and incubated at 4°C for 30 minutes (10 RPM). Samples were centrifuged at 10,000 x g for 20 minutes at 4°C and then transferred to 50 µL of GFP beads and incubated again for 30 minutes at 4°C (10 RPM). Columns were equilibrated by running 200 µL of RIP lysis buffer. The plant samples were added to the column and washed three times with wash buffer (identical to RIP lysis buffer but with 0.1% Triton X-100). For co-immunoprecipitation assays, hot elution buffer was added to the column and the

protein products were collected and analyzed by mass spectrometry. RIP assays were performed by eluting the RNA with Trizol reagent instead of elution buffer, and purifying RNA using Direct-Zol RNA miniprep kits (Zymo Research, Irvine, CA, USA). cDNA libraries were created using ReverTra Ace qPCR RT kits (Toyobo, Osaka, Japan), and qPCR with primers targeting *nad* intron and exon sequences were used to assess transcript abundance.

Results:

PCIS1 is co-expressed with various PPR proteins

To identify novel organellar RNA processing factors, we focused on the *AT5G25500* gene (annotated here as PPR co-expressed intron splicing1, *PCIS1*), which is co-expressed with various PPR protein genes that are known or predicted to be involved in organellar RNA metabolism in *Arabidopsis thaliana* plants. The gene contains no introns and encodes a protein with 421 amino acid residues (Jones et al., 2014). The *in-silico* data indicated that *PCIS1* gene is well conserved in flowering plants, though there is no similarity with other known domains (Figures 28 and 29) (Sayers et al., 2022; Thomas et al., 2022). Some databases containing the InterPro site (www.ebi.ac.uk/interpro/) annotate this gene family as PTHR37763 "EXSOME COMPLEX EXONUCLEASE" (as of May 2024). However, we found no experimental or structural data supporting the annotation. These incorrect annotations are due to mis-annotation in the TAIR database (www.arabidopsis.org), which were displayed until recently. Prediction of transmembrane helices by TMHMM 2.0 did not show any significant transmembrane domains in PCIS1 (Krogh et al., 2001; Sonnhammer et al., 1998). The predicted structure of the protein is not intrinsically disordered but has no significant similarity with other domains (Figure 30) (Abramson et al., 2024; Erdős & Dosztányi, 2024). The functional prediction from the predicted protein structure by DeepFRI suggested a potential involvement in "protein interaction", albeit with a moderate score of 0.51 (Gligorijević et al., 2021).

Genes that were found to be notably co-expressed with *PCIS1*, according to the ATTEDII (z score>=8.0), included 29 PPR proteins (Obayashi et al., 2022)(Table 4). These include three PPR factors that are classified as E2 subclass, nine E+ subclass, and 14 DYW subclass PPR proteins. Among these, eight were previously reported to be involved in mitochondrial C-to-U

editing (i.e., MEF10, MEF13, OTP72, OTP90, SLO2, SLO4), or in the editing of plastidial RNAs (i.e., OTP81/QED1 and OTP82). Two additional genes include a P-class and a PLS subclass PPR proteins with no known or established functions yet. Its conservation among angiosperms, and its co-expression patterns with PPR proteins, may imply that *PCIS1* is involved in organellar RNA metabolism.

PCIS1 **knockout lines are embryonically lethal.**

Here we analyzed two T-DNA insertion lines found within the *PCIS1* gene-locus, i.e., *pcis1-1* (SALK_152244), and *pcis1-2* (SALK_094043C) (Figure 31). The *pcis1-1* mutant line has two tandem T-DNA insertions with opposite orientations found about 350 nucleotides downstream to the translation start (ATG) site, while the *pcis1-2* contains a T-DNA inserted about 450 nucleotides downstream to the initiation site (Figure 31). Genotyping of the *pcis1-1* offspring did not reveal any plantlets that were homozygous for the *pcis1-1* allele. We, therefore, investigated the characteristics of seeds found in the silique of heterozygous *pcis1-1* plants. Non-viable seeds may appear shrunken, white, aborted or completely missing from the silique. Likewise, the green mature siliques of heterozygous *pcis1-1* plants contained about 10 ~ 15 % white seeds, in which embryogenesis was delayed (Figure 32). Seeds found within the fully matured siliques were brown and showed no visible (abnormal) phenotype. However, after collection of dried seeds, about 23.7 % (36/152) of seeds appear smaller and shriveled (Figure 33). The seeds seem to have fragments “broken off” or exhibit large crevices in the outer shell, unlike the normal ovular shape. Some seeds also showed discoloration compared to the wild type seeds, being in a darker shade of the usual brown. Morphologically abnormal seeds from heterozygous plants were separated visually and grown on MS medium to access the relationship between appearance and viability.

Of the 36 abnormal seeds found, only eleven germinated compared to the 106 out of 116 seeds with a normal phenotype. Four of them sprout cotyledons, but these quickly die before progressing any further. The remaining seven seeds that germinated were genotyped and two wild type and five heterozygous plants were found. In total, the ratio of the inviable seeds is about 25.7%, coinciding with Mendelian genetics. Upon genotyping 24 separate, viable plants from heterozygous *pcis1-1* seeds (36 plants total), 17 plants were heterozygous and 7 were wild type for the T-DNA insertion. Similar to *pcis1-1*, about 25% of the seeds from the heterozygous *pcis1-2* mutants, failed to germinate, and no homozygous plants could be found among their progeny. These results indicate that *PCIS1* encodes a protein whose functions are essential during the embryogenesis phase (Figure 34).

To assess the possible impact of *PCIS1* mutations on microsporogenesis, pollen collected from heterozygous *pcis1-1* plants was analyzed by Alexander staining (Figure 35). This method showed that the viability of pollen obtained from the mutant plants is similar to that of the wild type plants, suggesting that the lethality is not related to pollen production and maturation.

Functional complementation lines successfully restored the growth defect phenotypes associated with the *pcis1-1* mutant plants

To confirm that the phenotypes seen in *pcis1-1* mutant result by disruption of the *PCIS1* gene, we used a functional complementation strategy. For this purpose, the *PCIS1* coding region was cloned into pGWB5 vector under the control of the CaMV 35S promoter, and then transformed into the heterozygous *pcis1-1* plants. Among the different hygromycin-resistant transformants, none were found to be homozygous for the T-DNA insertion of *pcis1-1*. We also genotyped the T2 offspring of the p35S::*PCIS1* with heterozygous backgrounds, but again no homozygous plants were found

among the progeny of the p35S::*PCIS1* plants (n=56). This is probably due to fine-tuned expression of *PCIS1* that cannot fully be replicated with constitutive expression under the 35S promoter. When we used the native promoter of *PCIS1* (including the region 1,000 bp upstream of the start codon) instead of the 35S promoter, five transformants with homozygous T-DNA insertion backgrounds were obtained (Figure 36). The functional complemented lines showed typical growth that was similar to that of the wild type plants (Figure 37). Therefore, we concluded that the mutation in the *PCIS1* gene-locus was indeed associated with the phenotypes observed in the T-DNA insertion lines.

Establishment of partially complemented *pcis1-1* plants that express the PCIS protein under the control of the ABI3 promoter

Mitochondria dysfunctions have been previously associated with low germination and stunted growth phenotypes and are often also associated with embryogenesis developmental defects (Colas Des Francs-Small & Small, 2014). Likewise, the analysis of *pcis1* mutants indicated that none of their progeny are homozygous for the *pcis1-1 or pcis1-2* alleles, due to embryo developmental defects. Here, we used the ABI3 promoter (Despres et al., 2001) to drive embryonic expression of *PCIS1* in homozygous *pcis1-1* mutants. The coding sequence of the *PCIS1* gene was cloned under the control of the *ABI3* promoter and transformed into heterozygous *pcis1-1* plants. This allowed us to obtain several p*ABI3*::*PCIS1* transformants that are also homozygous to the *pcis1-1* insertion. The partially complemented p*ABI3*::*PCIS1* plants (i.e., p*ABI3*::*PCIS1* #1, and p*ABI3*::*PCIS1* #2) showed typical development until the 4-5th leaf stage, but following this stage they showed notable retardation in development (Figure 37 and 38). These phenotypes were noticeable prior to the bolting stage (the vegetative to reproductive phase transition) and were more notable afterwards.

The majority of these plants failed to form a flowering stem, while the few plants that initiated bolting were delayed in growth by 2-4 months. Leaf phenotypes were also notably affected, with a dark green coloration and curly-leaf morphology (Figure 38). This phenotype is especially prominent in the late-developed leaves, which are also shrunk in size, as compared with the early-developed leaves. As many of the plants were unable to bolt, they continuously created new leaves resulting in a shrub-like morphology. Among the viable p*ABI3*::*PCIS1* transformants (homozygous for the *pcis1-1* allele), the line #2 plants were able to flower and set viable seeds, but with fewer seeds produced as compared to the wild type (Col-0) plants. The partially complemented seedlings were used for the molecular analysis of PCIS1 roles in mitochondria biogenesis and function.

The expression patterns of *PCIS1* gene

Public available gene expression databases indicate *PCIS1* is highly expressed during embryogenesis, the beginning of the germination, as well as in shoot apexes (Winter et al., 2007). To investigate the expression patterns of *PCIS1* gene in different tissues and growth stages by GUS activity staining, the *PCIS1pro*::*PCIS1* was fused with *GUS* gene and introduced into wild type plants. The *PCIS1* gene was fused with *GUS* as expression would be closer to natural conditions and is more useful in further experiments. Individually isolated *PCIS1pro::PCIS1-GUS* lines displayed GUS staining in actively dividing tissues in the plants, located in the apical meristems of the roots and shoots, early true leaves, as well as stigmas in flowering plants (Figure 39).

The PCIS1 protein is located within the mitochondria

Based on its deduced amino acid sequence (Hooper et al., 2017), the N-terminal region of PCIS1 is predicted to harbor a mitochondrial localization region, which facilitates its import into the mitochondria in plants. To analyze the subcellular location of PCIS1 protein in plants, a *PCIS1-GFP* fusion gene was introduced and expressed in Arabidopsis plants. *PCIS1-GFP* transformed plants grew faster and appeared healthier than Col-0 plants, however this was not experimentally examined. The PCIS1-GFP signal in the transgenic line fully overlapped with that of the mitochondrial maker, confirming that PCIS1 is indeed localized to the mitochondria, in vivo (Figure 40).

***PCIS1* facilitates the processing and maturation of several mitochondrial *nad* transcripts**

PCIS1 seemed to be co-expressed with several PPR proteins that function in mitochondrial RNA processing. It is therefore assumed that PCIS1 may function in mtRNA metabolism as well. As the majority of co-expressed PPR proteins function (or are predicted to be involved) in RNA editing (i.e., C>U base conversions), all the known RNA editing sites in the coding regions of mitochondria loci were analyzed in the partially-complemented p*ABI3*::*PCIS1* mutant lines by direct Sanger sequencing of cDNA derived from mature transcripts. Some editing sites show moderate differences in efficiencies between the wild type and mutant plants, but these cannot explain the severe phenotypes of the mutant lines (Table 7).

In addition to RNA editing, we also considered other RNA metabolism effects. RNA-Seq analysis of mature leaves of p*ABI3*::*PCIS1* lines and wild type plants revealed altered splicing activities of *nad2* intron 4 (*nad2* i4), *nad4* i3, and *nad7* i2, as indicated by the ratios of the number of mapped transcripts on the introns and their flanking exons, which showed about 2-to-20 times

higher reads in the intron regions than the wild type plants (Figure 41). The RNA-Seq data did not reveal any other significant changes in the accumulation or metabolism of other mitochondrial transcripts, including those which correspond to introns, nor in the processing of postulated 5' and 3' regions of the mitochondrial transcriptome (Best, Sultan, et al., 2020; Ruwe et al., 2016; Zoschke & Barkan, 2015). Likewise, the expression levels of intronless mitochondrial and plastid RNAs in the p*ABI3*::*PCIS1* plants were equivalent to those of the wild type plants. In addition, expression of *PCIS1* was undetectable in RNA-Seq data, supporting that the *ABI3* promoter no longer expressed the gene in mature plants.

In Arabidopsis mitochondria, the coding regions of *nad1*, *nad2*, *nad4*, *nad5*, *nad7*, cox2, *ccmFc*, *rpl2*, and *rps3* genes are interrupted by group II intron sequences (Sloan et al., 2018; Unseld et al., 1997). To examine whether the increased reads of intronic regions we see in the RNA-seq data of the *pcis1* mutants are associated with altered splicing activities, the relative mRNA to pre-mRNA ratios of each of the 23 introns in the mutants versus those of the wild type plants, were analyzed by quantitative RT-PCR (RT-qPCR) assays. mRNA can be detected with primers specific to exon sequences while pre-mRNAs are detected with an intron and exon specific primer pair. Introns with splicing affected in the mutant plants included *nad2* i4, *nad4* i3, and *nad7* i2 (Figure 42). Reduced splicing efficiencies (i.e., the ratios of pre-RNA to mRNA in mutant lines versus those of the wild type plants) were apparent in the cases of *nad2* intron 1, *nad4* introns 1 and 2, and *nad5* intron 1 (Figure 42). We also noticed increased steady-state levels (2x ~ 4x) in many other pre-mRNAs (Figure 43). However, splicing defects were considered only in cases where a higher accumulation of a specific pre-mRNA was correlated with the reduced level of its corresponding mature mRNA transcript (i.e., adjoining exons). As shown in Figure 43A, although the pre-mRNA to mRNA ratios in mutant plants vs Col-0 plants seemed altered for various

transcripts, notable reductions in mRNA levels where apparent only for spliced exon transcripts corresponding to *nad2* (exons 4-5), *nad4* (exons 3-4) and *nad7* (exons 2-3).

The gene of *PCIS1* seems to overlap with a 'natural antisense transcript' (i.e., At5g04005) that is annotated as antisense_lncRNA. T-DNA insertions of the two insertional lines, *pcis1-1* and *pcis1-2,* are both found outside the coding region of At5g04005, while RT-qPCR analysis confirmed that the expression of the *AT5G04005* gene product was not significantly affected in the mutant lines (Figure 43B).

In addition to the RNA-Seq and RT-qPCR experiments, I carried out RT-PCR assays with primers amplifying the entire coding region to assay whether mature mRNAs are present in *pcis1* mutants. Wild type controls only exhibited the PCR products of mature mRNA but not of immature unspliced transcripts, as did the native promoter complemented lines (Figure 44). *proABI3* complemented *pcis1-1* mutants however, differences in splicing patterns were present in multiple mitochondrial transcripts. In agreement with previous experiments, *nad2*, *nad4*, and *nad7* all had additional RT-PCR products that appeared to be derived from immature transcripts. *nad7* in particular had no PCR products corresponding to mature transcripts, suggest mutants cannot properly splice intron 2 in the absence of the PCIS1 protein. *nad2* and *nad4* contained a combination of the mature transcript and smaller fragments, showing that while splicing is impaired, it is not entirely lost (Figure 44).

Northern blots were also performed to directly view the presence of RNA transcripts using the coding region of each *nad* gene as a probe. Results were not as clear as RT-PCR gels, as non-specific binding to ribosomal RNAs and other transcripts were detected. The presence of additional transcripts was detected in both *nad4* and *nad7*, but no changes could be found in *nad2* (Figure 45). This could be due to the high background of ribosomal RNAs. *nad7*'s additional band was

found at around 2.7 knt which roughly coincides with the total length of the mature transcript (1,612 nt) and intron 2 (1,063 nt). *nad4* had an extra band of around 3.7 knt compared to the mature transcript (1,488 nt) and intron 3 (1,942 nt).

The biogenesis and function of Complex I is affected in *pcis1* mutants

The RNA metabolism defects we see in *pcis1* mutants are expected to affect the expression of Nad2, Nad4, and Nad7 proteins, and hence influence the biogenesis of the respiratory complex I (NADH dehydrogenase enzyme, CI). The accumulation of native respiratory complexes I, III, IV, and V in Col-0 and *pcis1-1* mutant plants was analyzed by Blue native PAGE (BN-PAGE), followed by *in-gel* activity assays (Figure 46) and immunoblot analyses (Figure 47). In-gel activity assays indicated that CI abundance or function are notably reduced (i.e., undetectable) in the mutant line (Figure 46A). Immunoblotting with CA2 antibodies further showed that the holo-CI is strongly reduced (i.e., below detectable levels) in the *pcis1-1* mutant plants. The anti-CA2 blot (corresponds to the membrane-arm domain of the holo-CI enzyme) indicated the presence of various CI assembly intermediates (Ligas et al., 2019) of about 100 to 550 kDa in *pcis1-1* mitochondria (Figure 47). Relative respiratory enzyme's levels were measured by densitometry of immunoblots, and quantified using ImageJ software (Chotewutmontri & Barkan, 2016). The signals of the chloroplast ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) enzyme (i.e., about 500 kDa in size, as they appear with antibodies to Rubisco large subunit, RbcL) (Figure 46B) was used as an internal loading control for the immunoblots. While CI was notably reduced in the mutants, the signals corresponding to the Cox2 of CIV and AtpA were higher (i.e., ~5.5 and 2.7-fold higher, respectively) in the mutant (Figure 47). The abundance of the RISP subunit of CIII was similar in the wild type and mutant plants.

The relative accumulation of various mitochondrial proteins was analyzed in 3-week-old MS-grown Col-0 seedlings and *pcis1-1* mutant plants, by SDS-PAGE followed by immunoblot assays (Figure 46C). The immunoblots indicated that the steady-state levels of the CI subunits, NAD9 (~23 kDa) (Lamattina et al., 1993) and carbonic anhydrase-like 2 (CA2, ~30 kDa) (Perales et al., 2005; Sunderhaus et al., 2006) subunits, are similar in the wild type and *pcis1-1* plants. Likewise, the cytochrome c-1 subunit (CytC, ~12 kDa), which is associated with complex III (Klodmann et al., 2011), the complex IV cytochrome c oxidase subunit 2 (COX2, ~30 kDa), the AtpA subunit (AtpA, ~55 kDa) of the ATP synthase enzyme (CV), and the outer mitochondrial membranous protein, voltage-dependent anion-selective channel protein 1 (VDAC1, ~27 kDa), accumulated to similar levels in the wild type and mutant plants. However, the levels of the twin cysteine protein, At12Cys (~12 kDa), which is known to be induced in mutants affected in complex I biogenesis (Y. Wang et al., 2016), were found to be notably higher in the *pcis1* mutants than in the wild type plants (Figure 46C).

In addition to the canonical respiratory enzymes, the OXPHOS in plants is further represented by the rotenone-insensitive type-II NAD(P)H dehydrogenases (NDs) and cyanide-resistant alternative oxidases (AOXs), which can bypass the four classical electron transport chain (Millar et al., 2011; Schertl & Braun, 2014). Similar to other mitochondrial mutants in plants, the defects in CI assembly were associated with the induction of various AOXs and NDs in the *pcis1* mutants (Figure 48). Immunoblot analyses further indicated that the steady-state levels of alternative oxidase 1a protein (AOX1a, ~35 kDa) was notably increased in the mutant plants (Figure 46C).

Together, the RNA and protein profiles strongly indicate a role for PCIS1 in the processing of mitochondrial group II introns, and that these functions are pivotal for the expression of various

complex I subunits. Mutants affected in PCIS1 show altered complex I function, which are strongly associated with embryo developmental defects (Ostersetzer-Biran, 2016).

PCIS1 does not interact with other splicing factors or relevant RNA transcripts

So far, eleven splicing factors (ABO6, ABO8, CFM9, mCSF1, MATR, MISF74, MTL1, nMAT2, OZ2, PMH2, and SLO3) have been shown to be required for one or few of the three PCIS1's target introns (Bentolila et al., 2021; Haïli et al., 2016; J. He et al., 2012; Hsieh et al., 2015; Keren et al., 2009; Köhler et al., 2010; Lee et al., 2019; Wahleithner et al., 1990; C. Wang et al., 2018; Yang et al., 2014; Zmudjak et al., 2013). The functional prediction analysis of PCIS1's amino acid sequence using the DeepFRI model shows forecasts protein binding as the only significant GO term for molecular function (Gligorijević et al., 2021). It is possible that PCIS1 interacts with some of these splicing factors and forms complexes needed for maturation of each intron. Protein-protein interactions between PCIS1 and the other known splicing factors involved in the three PCIS1 target introns were therefore examined through yeast-two-hybrid assays. However, PCIS1 showed no interaction with itself or any tested splicing factor (Figure 49). To investigate other potential interactions, protein co-immunoprecipitation and RNA immunoprecipitation assays using PCIS1-GFP transformed plants was analyzed using mass spectrometry and qRT-PCR, respectively. PCIS1-GFP proteins were detected along with many other proteins, but these were unrelated to splicing and very few were localized in the mitochondria (Table 7). Y2H of the top four potentially interacting proteins also showed no interaction (Figure 50). RNA interactions between PCIS1-GFP and total mRNA were analyzed by qRT-PCR. Primers specific for *nad* target RNAs including intron and exon-specific primers were used to compare abundance of unspliced transcripts. No

differences were found between the PCIS1-GFP co-IP RNAs and a Col-0 control. Overall RNA abundance was low and gave poor quality qPCR results, however. Additional experiments will be needed to further investigate protein and RNA interactions.

Discussion

PCIS1 is an angiosperm-specific gene that is co-expressed with various PPR proteins

Group II introns are large catalytic RNAs, which are related to spliceosomal introns, non-LTR retrotransposons, and telomers. Group II introns form interactions with various protein cofactors that are required for their splicing activities, in vivo (Brown et al., 2014; Zmudjak et al., 2017). Canonical group II introns encode proteins (known as maturases, or IEPs), which bind with high affinity and specificity to their host introns to facilitate their splicing (Mizrahi et al., 2022). The *Arabidopsis thaliana* mitochondrial genome contains 23 group II-type introns, of which 18 are *cis*-spliced and five are found in a *trans* configuration (i.e., encoded by separate gene loci) (Unseld et al., 1997). The plant organellar introns are highly degenerated, lacking various intronic regions that are required for their 'self-splicing' activity, and have also lost their cognate intron-encoded ORFs (Bonen, 2008; Brown et al., 2014; Zimmerly & Semper, 2015). Instead, the plant organellar introns have acquired various *trans*-acting splicing cofactors, most of which are encoded in the nuclear genomes. These include universal maturases that have been mobilized into the nuclear genomes (i.e., nMATs), as well as RNA-helicases, CRM proteins, and many PPR proteins (Zmudjak & Ostersetzer-Biran, 2018). Using *in silico* analyses, it was found that the *PCIS1* gene is co-expressed with various PPR proteins (Obayashi et al., 2022) that are known (or predicted) to be involved in the regulation of organellar gene expression (Barkan & Small, 2014). Several uncharacterized PPR genes belonging to the P-class are co-expressed with *PCIS1*. 22 PPR protein genes show a moderate co-expression pattern (7.0=<z-score<8.0) with *PCIS1*.

PCIS1 encodes a previously unidentified mitochondria-localized splicing cofactor (Figure 40), whose functions are essential during embryogenesis and germination and for normal growth and development (Figures 32 - 38). GUS fusion constructs of *PCIS1* show expression in actively

dividing tissues where mitochondria are highly active (Figure 39). Expression did not coincide completely with predictive algorithms, but the observed expression patterns were similar to other intron splicing factors (Lee et al., 2021; Su et al., 2017; Zhao et al., 2020). PCIS1 is found to be well conserved in angiosperms plants, including in those that are regarded as more 'basal' species, such as the *Amborella, Nymphaea, Magnolia,* and *Gingko* (PANTHER classification system and BLASTp) (Figures 28 and 29) (Sayers et al., 2022; Thomas et al., 2022). Complementation of the null phenotype was only successful with expression under the *PCIS1* native promoter, suggesting that expression is tightly regulated at low levels. As no homozygous plants could be obtained from the progeny of the null *PCIS1* mutants, we used a partial complementation strategy that is based on the expression of an essential gene under the embryo-specific ABI3 promoter (Despres et al., 2001). The p*ABI3*::*PCIS1* rescued mutants, which are homozygous for the *pcis1-1* allele, were notably affected in growth and development, but otherwise allowed us to obtain enough plant material to analyze their related RNA and protein profiles.

PCIS1 is necessary for splicing of three NADH dehydrogenase subunits

Analysis of the RNA profiles of *pcis1* mutants by RNA-seq and RT-qPCR analyses indicated notable defects in the maturation of *nad2*, *nad4, and nad7* pre-RNAs (Figure 42). As the coding regions of these three *nad* genes are interrupted by group II intron sequences, we analyzed the splicing efficiencies of the 23 introns in *Arabidopsis* mitochondria (Figures 42, 43A). The RT-qPCR assays indicated perturbation in the processing of *nad2* i4, *nad4* i3, and *nad7* i2. We also noticed increased levels of many other pre-RNAs in *pcis1*, but splicing defects can be concluded only in cases where an increase in a specific pre-mRNA is correlated with a reduction in its corresponding mRNA in the mutant plants (Figure 43A). An upregulation in many of organellar

primary transcripts in *pcis1* may occur as a result of the association of various splicing cofactors with the partially processed *nad2*, *nad4,* and *nad7* pre-mRNAs, which are unable to dissociate from the organellar spliceosomal nucleoprotein particles as long as the splicing reaction is incomplete in the absence of PCIS1. Accordingly, multiple splicing factors have been found to be involved in the processing of each of the three intron targets of PCIS1 in Arabidopsis plants (Colas Des Francs-Small & Small, 2014; Zmudjak & Ostersetzer-Biran, 2018). The *nad2* i4 requires at least six known splicing factors (i.e., ABO6, CFM9, MISF74, mCSF1, nMAT2, and PMH2). Likewise, the processing of *nad4* intron 3 involves ABO6, ABO8, CFM9, and PMH2, while the maturation of *nad7* intron 2 depends on eight different factors, i.e., CFM9, MatR, mCSF1, MTL1, nMAT2, OZ2, PMH2 and SLO3. Of the 11 factors, ABO8, MISF74, MTL1, and SLO3 are all PPR proteins. The participation of PCIS1 in the splicing of *nad2* i4, *nad4* i3, and *nad7* i2 may indicate that these pre-mRNAs share a common feature. However, we have not identified sequence motifs especially common among the three introns, which may be recognized by the PCIS1 protein. It is possible, therefore, that these introns share a common structural feature that is recognized by the protein or PCIS1 is recruited to a ribonucleoprotein complex by a yet unknown factor. Yeast 2 hybrid assays showed no interaction between PCIS1 and currently known splicing factors, but further discoveries may find other participating proteins.

A co-expression gene cluster including PCIS1 may be exclusively important for complex I maturation

Many of the proteins co-expressed with *PCIS1* were identified as RNA editing PPR cofactors, containing the E2, E+ or DYW subclasses (Table 5). Most of them target one or multiple editing sites in *nad* transcripts. MEF10 is required for *nad2*-842 editing. MEF13 targets the C>U

conversions of *nad2*-59, *nad4*-158, *nad5*-1665, *nad5*-1916, and *nad7*-213 sites, while OTP90 facilitates the editing of the C-residue in *nad1*-500. SLO2 edits *nad4L*-110 and *nad7*-739. SLO4 is essential for *nad4*-1033.

As PCIS1 and co-expressed RNA editing factors exclusively target nad genes, their co-expression pattern are valuable indicators for predicting the function of other uncharacterized genes in the cluster. The *PCIS1* co-expressed P-class PPR proteins may also be required for splicing or other maturation steps of *nad* transcripts, a prerequisite for the assembly of complex I in mitochondria (Table 2).

The biogenesis of respiratory complex I is strongly affected in *pcis1-1* mutants

The OXPHOS system in plants contains the 4 canonical electron transport chain (ETC) complexes (i.e., CI to CIV), and the ATP synthase (CV) enzyme, as well as various proteins of the alternative system (Møller et al., 2021). The NADH dehydrogenase (CI) is the largest ETC enzyme and also serves as the main site for electron transfer to the respiratory chain in most aerobic organisms. In *Arabidopsis*, CI consists of 47 subunits (Klusch et al., 2021; Soufari et al., 2020), with nine being encoded within the mitochondrial genome, i.e., Nad1, Nad2, Nad3, Nad4, Nad4L, Nad5, Nad6, Nad7, and Nad9 (Unseld et al., 1997). Five of the genes (*nad1*, *nad2*, *nad4*, *nad5*, and *nad7*) contain group II introns that must be removed post-transcriptionally from the coding regions they interrupt. These polypeptides and their nuclear-encoded counterparts are assembled into two main modules, known as the membrane arm (consisting of Nad1, Nad2, Nad3, Nad4, Nad4L, Nad5, Nad6) and the peripheral arm (containing Nad7 and Nad9) (Ligas et al., 2019).

The recruitment of functional Nad2, Nad4, and Nad7 proteins is pivotal for the assembly of a functional holo-complex I enzyme (Hsieh et al., 2015; C. Wang et al., 2018; Yang et al., 2014).

Accordingly, the splicing defects we see in *pcis1-1* mutants (Figures 42 and 43) were linked to altered CI biogenesis and activity, as indicated by BN-PAGE (Figures 46A, 46B, and 47) and SDS-PAGE (Figure 46C) followed by immunoblot analyses. Activity assays display a band in *pcis1-1* mutants that appears to correlate with a CI intermediate. This sub-complex may be partially active and further examination is required to elucidate its identity. The impairment in CI biogenesis and function was further associated with the upregulation of respiratory CIII, CIV, and the ATP synthase enzyme, as well as the induction of expression of genes encoding alternative respiratory chain components (i.e., AOXs and NDs) (Figures 46C and 48) and the At12Cys protein that is known to be upregulated (Wang *et al.* 2016). Previously it was shown that the At12Cys-2 transcript is induced in different 'mitochondrial mutants', whereas an upregulation in At12Cys protein accumulation is apparent solely for mutants affected in complex I biogenesis. The upregulation in the expression and accumulation of other OXPHOS enzymes is likely required to accommodate the alterations in cellular metabolism (e.g., ATP synthesis and other essential metabolic processes). It is further hypothesized that the embryo-arrest, growth, and developmental defect phenotypes seen in *pcis1* are tightly correlated with altered CI activities, as also observed in various other mutant plants affected in the assembly or function of CI (Best, Mizrahi, et al., 2020; Kühn et al., 2015; Ostersetzer-Biran, 2016).

Closing Remarks

RNA processing in some shape or form is essential to all life on this planet. Processing allows not only for the creation of mature transcripts needed for proper protein function, but also allows for control of expression patterns.

In chapter one, I focused on in vitro RNA editing of recombinant PPR56 protein. PPR56 specifically edits sites in *nad3* and *nad4* RNA transcripts, required for construction of the NADH dehydrogenase complex used in oxidative phosphorylation. Moreover, if and how PPR proteins edit their targets in the absence of other factors was not known. Through in vitro editing experiments using recombinant protein, I was able show that a single PPR protein was sufficient to edit an RNA target, and that its DYW domain holds cytidine deaminase activity. Recombinant PPR56 protein's activity was also enhanced by the presence of NTPs hinting at a control mechanism based upon physiological conditions within the mitochondria where they function. As PPR56 directly edits RNAs needed for energy production, this is important for maintaining a balance when energy stores are low or when overproduction would be disadvantageous. While many mechanistic characteristics of the editing reaction remain yet unknown, I also endeavored to investigate more about the binding between PPR proteins and their substrate RNAs. While the location of the binding site is still elusive, it would seem that the monomeric fraction of the protein is the active form and NTPs may help stabilize this state.

In chapter 2, I catalogued a novel protein I named PCIS1 based on its co-expression pattern with other PPR proteins. PCIS1 is required for group II intron splicing of three *nad* transcripts. Knockout mutants are embryonically lethal and required partial complementation using an embryo-specific promoter from the *ABI3* gene. Even under partial expression, mutant plants showed drastically stunted growth and difficulty forming stems necessary for reproduction. The

experiments performed with PCIS1 further reinforced the importance of RNA processing as impairment of splicing of just a few RNAs can lead ultimately to inviability. Specifically for group II intron splicing, it requires a finely tuned network of multiple splicing factors to operate (Li & Jiang, 2024), which adds another layer of control for pivotal process such as energy production. PCIS1's mechanism is still unknown, and yeast-two-hybrid as well as other preliminary RNA binding experiments failed to find any hints of interaction. Thus, it is highly likely that PCIS1 is part of an editing complex with proteins yet to be characterized.

In regards to both topics, there is still much that is not known and that is waiting to be discovered. The mechanisms behind specifically how PPR proteins edit is still not understood, but the prospects of using RNA editing as a tool is a great one. PPR56 presents the possibility of the creation of designer editing proteins that could edit targets of interest, similar to other systems in animals such as the use of ADAR and APOBEC. As PPRs have a pseudo-genetic code that dictates their preferred targets, this gives greater control over other methods which typically require the use of RNA guides. These tools could be extended even to the field of medicine where diseases caused by single nucleotide polymorphisms could potentially be cured through a weekly shot.

RNA splicing within the plant mitochondria is also still poorly understood. In the *A. thaliana* there exists only 23 group II introns, and yet each one requires the activity of multiple splicing factors in order to create the mature transcript. While some factors may influence splicing of multiple introns, some are also highly specific to one. How general factors and specific factors interface with another to form a splicing complex is largely unknown. In addition, there are many more factors that have yet to be discovered. While PCIS1 is one piece of the puzzle, the discovery of a novel splicing factor other than typical RNA-binding proteins that does not contain any known domains suggest a more complex network of splicing factors in plant organelles. Future studies

will need to examine and annotate other factors while a more comprehensive look at the interactions between the individual splicing factors will be needed.

References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., … Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*. https://doi.org/10.1038/s41586-024-07487-w

Araya, A., Domec, C., Begu, D., & Litvak, S. (1992). An in vitro system for the editing of ATP synthase subunit 9 mRNA using wheat mitochondrial extracts (plant mitochondrial genes/RNA editing). In *Biochemistry* (Vol. 89). https://www.pnas.org

Aryamanesh, N., Ruwe, H., Sanglard, L. V. P., Eshraghi, L., Bussell, J. D., Howell, K. A., Small, I., & des Francs-Small, C. C. (2017). The pentatricopeptide repeat protein EMB2654 is essential for trans-splicing of a chloroplast small ribosomal subunit transcript. *Plant Physiology*, *173*(2), 1164–1176. https://doi.org/10.1104/pp.16.01840

Bae, J. W., Kwon, S. C., Na, Y., Kim, V. N., & Kim, J. S. (2020). Chemical RNA digestion enables robust RNA-binding site mapping at single amino acid resolution. *Nature Structural and Molecular Biology*, *27*(7), 678–682. https://doi.org/10.1038/s41594-020-0436-2

Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y. S., Bond, C. S., & Small, I. (2012). A Combinatorial Amino Acid Code for RNA Recognition by Pentatricopeptide Repeat Proteins. *PLoS Genetics*, *8*(8). https://doi.org/10.1371/journal.pgen.1002910

Barkan, A., & Small, I. (2014). Pentatricopeptide repeat proteins in plants. In *Annual Review of Plant Biology* (Vol. 65, pp. 415–442). Annual Reviews Inc. https://doi.org/10.1146/annurev-arplant-050213-040159

Bayer-Császár, E., Haag, S., Jörg, A., Glass, F., Härtel, B., Obata, T., Meyer, E. H., Brennicke, A., & Takenaka, M. (2017). The conserved domain in MORF proteins has distinct affinities to the PPR and E elements in PPR RNA editing factors. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1860*(8), 813–828. https://doi.org/10.1016/j.bbagrm.2017.05.004

Bentolila, S., Gipson, A. B., Kehl, A. J., Hamm, L. N., Hayes, M. L., Mulligan, R. M., & Hanson, M. R. (2021). A RanBP2-type zinc finger protein functions in intron splicing in Arabidopsis mitochondria and is involved in the biogenesis of respiratory complex i. *Nucleic Acids Research*, *49*(6), 3490–3506. https://doi.org/10.1093/nar/gkab066

Best, C., Mizrahi, R., & Ostersetzer-Biran, O. (2020). Why so complex? The intricacy of genome structure and gene expression, associated with angiosperm mitochondria, may relate to the regulation of embryo quiescence or dormancy—intrinsic blocks to early plant life. *Plants*, *9*(5). https://doi.org/10.3390/plants9050598

Best, C., Sultan, L., Murik, O., & Ostersetzer-Biran, O. (2020). *Insights into the mitochondrial transcriptome landscapes of two Brassicales plant species, Arabidopsis thaliana (var. Col-0) and Brassica oleracea (var. botrytis)*. https://doi.org/10.1101/2020.10.22.346726

Bonen, L. (2008). Cis- and trans-splicing of group II introns in plant mitochondria. In *Mitochondrion* (Vol. 8, Issue 1, pp. 26–34). https://doi.org/10.1016/j.mito.2007.09.005

Bonen, L. (2018). Mitochondrial Genomes in Land Plants. In R. D. Wells, J. S. Bond, J. Klinman, & B. S. S. Masters (Eds.), *Molecular Life Sciences: An Encyclopedic Reference* (pp. 734–742). Springer New York.

Boussardon, C., Avon, A., Kindgren, P., Bond, C. S., Challenor, M., Lurin, C., & Small, I. (2014). The cytidine deaminase signature HxE(x)nCxxC of DYW1 binds zinc and is necessary for RNA editing of ndhD-1. *New Phytologist*, *203*(4), 1090–1095. https://doi.org/10.1111/nph.12928

Boussardon, C., Salone, V., Avon, A., Berthomé, R., Hammani, K., Okuda, K., Shikanai, T., Small, I., & Lurina, C. (2012). Two interacting proteins are necessary for the editing of the ndhD-1 site in Arabidopsis plastids. *Plant Cell*, *24*(9), 3684–3694. https://doi.org/10.1105/tpc.112.099507

Brennicke, A., Grohmann, L., Hiesel, R., Knoop, V., & Schuster, W. (1993). The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants. In *FEBS Letters* (Vol. 325, Issues 1–2, pp. 140–145). https://doi.org/10.1016/0014-5793(93)81430-8

Brown, G. G., des Francs-Small, C. C., & Ostersetzer-Biran, O. (2014). Group II intron splicing factors in plant mitochondria. In *Frontiers in Plant Science* (Vol. 5, Issue FEB). Frontiers Research Foundation. https://doi.org/10.3389/fpls.2014.00035

Carlow, D., & Wolfenden, R. (1998). Substrate Connectivity Effects in the Transition State for Cytidine Deaminase †. In *Biochemistry* (Vol. 37). UTC. https://pubs.acs.org/sharingguidelines

Cheng, S., Gutmann, B., Zhong, X., Ye, Y., Fisher, M. F., Bai, F., Castleden, I., Song, Y., Song, B., Huang, J., Liu, X., Xu, X., Lim, B. L., Bond, C. S., Yiu, S. M., & Small, I. (2016). Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. *Plant Journal*, *85*(4), 532–547. https://doi.org/10.1111/tpj.13121

Chotewutmontri, P., & Barkan, A. (2016). Dynamics of Chloroplast Translation during Chloroplast Differentiation in Maize. *PLoS Genetics*, *12*(7). https://doi.org/10.1371/journal.pgen.1006106

Christofi, T., & Zaravinos, A. (2019). RNA editing in the forefront of epitranscriptomics and human health. In *Journal of Translational Medicine* (Vol. 17, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s12967-019-2071-4

Colas Des Francs-Small, C., & Small, I. (2014). Surrogate mutants for studying mitochondrially encoded functions. In *Biochimie* (Vol. 100, Issue 1, pp. 234–242). Elsevier. https://doi.org/10.1016/j.biochi.2013.08.019

Despres, B., Delseny, M., & Devic, M. (2001). Partial complementation of embryo defective mutations: A general strategy to elucidate gene function. *Plant Journal*, *27*(2), 149–159. https://doi.org/10.1046/j.1365-313X.2001.01078.x

Diaz, M. F., Bentolila, S., Hayes, M. L., Hanson, M. R., & Mulligan, R. M. (2017). A protein with an unusually short PPR domain, MEF8, affects editing at over 60 Arabidopsis mitochondrial C targets of RNA editing. *Plant Journal*, *92*(4), 638–649. https://doi.org/10.1111/tpj.13709

Erdős, G., & Dosztányi, Z. (2024). AIUPred: combining energy estimation with deep learning for the enhanced prediction of protein disorder. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkae385

Gietz, R. D., & Schiestl, R. H. (2007). Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method. *Nature Protocols*, *2*(1), 1–4. https://doi.org/10.1038/nprot.2007.17

Gligorijević, V., Renfrew, P. D., Kosciolek, T., Leman, J. K., Berenberg, D., Vatanen, T., Chandler, C., Taylor, B. C., Fisk, I. M., Vlamakis, H., Xavier, R. J., Knight, R., Cho, K., & Bonneau, R. (2021). Structure-based protein function prediction using graph convolutional networks. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-23303-9

Grewe, F., Herres, S., Viehöver, P., Polsakiewicz, M., Weisshaar, B., & Knoop, V. (2011). A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte Isoetes engelmannii. *Nucleic Acids Research*, *39*(7), 2890–2902. https://doi.org/10.1093/nar/gkq1227

Gualberto, J. M., & Newton, K. J. (2017). *Plant Mitochondrial Genomes: Dynamics and Mechanisms of Mutation*. https://doi.org/10.1146/annurev-arplant-043015

Gully, B. S., Cowieson, N., Stanley, W. A., Shearston, K., Small, I. D., Barkan, A., & Bond, C. S. (2015). The solution structure of the pentatricopeptide repeat protein PPR10 upon binding atpH RNA. *Nucleic Acids Research*, *43*(3), 1918–1926. https://doi.org/10.1093/nar/gkv027

Haïli, N., Planchard, N., Arnal, N., Quadrado, M., Vrielynck, N., Dahan, J., des Francs-Small, C. C., & Mireau, H. (2016). The MTL1 pentatricopeptide repeat protein is required for both translation and splicing of the mitochondrial NADH DEHYDROGENASE SUBUNIT7 mRNA in arabidopsis1. *Plant Physiology*, *170*(1), 354–366. https://doi.org/10.1104/pp.15.01591

Hammani, K., Okuda, K., Tanz, S. K., Chateigner-Boutin, A. L., Shikanai, T., & Small, I. (2009). A study of new arabidopsis chloroplast rna editing mutants reveals general features of editing factors and their target sites. *Plant Cell*, *21*(11), 3686–3699. https://doi.org/10.1105/tpc.109.071472

Hayes, M. L., Giang, K., Berhane, B., & Mulligan, R. M. (2013). Identification of two pentatricopeptide repeat genes required for rna editing and zinc binding by c-terminal cytidine deaminase-like domains. *Journal of Biological Chemistry*, *288*(51), 36519–36529. https://doi.org/10.1074/jbc.M113.485755

Hayes, M. L., & Santibanez, P. I. (2020). A plant pentatricopeptide repeat protein with a DYW-deaminase domain is sufficient for catalyzing C-to-U RNA editing in vitro. *Journal of Biological Chemistry*, *295*(11), 3497–3505. https://doi.org/10.1074/jbc.RA119.011790

He, J., Duan, Y., Hua, D., Fan, G., Wang, L., Liu, Y., Chen, Z., Han, L., Qu, L. J., & Gonga, Z. (2012). DEXH box RNA Helicase-Mediated mitochondrial reactive oxygen species production in Arabidopsis mediates crosstalk between abscisic acid and auxin signaling. *Plant Cell*, *24*(5), 1815–1833. https://doi.org/10.1105/tpc.112.098707

He, S. L., & Green, R. (2013). Northern blotting. In *Methods in Enzymology* (Vol. 530, pp. 75–87). Academic Press Inc. https://doi.org/10.1016/B978-0-12-420037-1.00003-8

Hiesel, R., Wissinger, B., Schuster, W., & Brennicke, A. (1989). RNA editing in plant mitochondria. *Science*, *246*(4937), 1632–1634. https://doi.org/10.1126/science.2480644

Hirose, T., & Sugiura, M. (2001). Involvement of a site-specific trans-acting factor and a common RNA-binding protein in the editing of chloroplast mRNAs: Development of a chloroplast in vitro RNA editing system. *EMBO Journal*, *20*(5), 1144–1152. https://doi.org/10.1093/emboj/20.5.1144

Hooper, C. M., Castleden, I. R., Tanz, S. K., Aryamanesh, N., & Millar, A. H. (2017). SUBA4: The interactive data analysis centre for Arabidopsis subcellular protein locations. *Nucleic Acids Research*, *45*(D1), D1064–D1074. https://doi.org/10.1093/nar/gkw1041

Hsieh, W. Y., Liao, J. C., Chang, C. Y., Harrison, T., Boucher, C., & Hsieh, M. H. (2015). The SLOW GROWTH3 pentatricopeptide repeat protein is required for the splicing of mitochondrial NADH dehydrogenase subunit7 intron 2 in arabidopsis. *Plant Physiology*, *168*(2), 490–501. https://doi.org/10.1104/pp.15.00354

Ichinose, M., & Sugita, M. (2018). The DYW domains of pentatricopeptide repeat RNA editing factors contribute to discriminate target and non-target editing sites. *Plant and Cell Physiology*, *59*(8), 1652–1659. https://doi.org/10.1093/pcp/pcy086

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Keren, I., Bezawork-Geleta, A., Kolton, M., Maayan, I., Belausov, E., Levy, M., Mett, A., Gidoni, D., Shaya, F., & Ostersetzer-Biran, O. (2009). AtnMat2, a nuclear-encoded maturase required for splicing of group-II introns in Arabidopsis mitochondria. *RNA*, *15*(12), 2299–2311. https://doi.org/10.1261/rna.1776409

Keren, I., & Ostersetzer-Biran, O. (2011). *An optimized method for the analysis of plant mitochondria RNAs by Northern-blotting Epigenetic regulation of gene expression by Histones post translational modifications View project Gene Expression regulation of*

Chloroplasts and Mitochondria View project. https://www.researchgate.net/publication/263304583

Klodmann, J., Senkler, M., Rode, C., & Braun, H. P. (2011). Defining the protein complex proteome of plant mitochondria. *Plant Physiology*, *157*(2), 587–598. https://doi.org/10.1104/pp.111.182352

Klusch, N., Senkler, J., Yildiz, Ö., Kühlbrandt, W., & Braun, H. P. (2021). A ferredoxin bridge connects the two arms of plant mitochondrial complex I. *Plant Cell*, *33*(6), 2072–2091. https://doi.org/10.1093/plcell/koab092

Knie, N., Grewe, F., Fischer, S., & Knoop, V. (2016). Reverse U-to-C editing exceeds C-to-U RNA editing in some ferns - A monilophyte-wide comparison of chloroplast and mitochondrial RNA editing suggests independent evolution of the two processes in both organelles. *BMC Evolutionary Biology*, *16*(1). https://doi.org/10.1186/s12862-016-0707-z

Köhler, D., Schmidt-Gattung, S., & Binder, S. (2010). The DEAD-box protein PMH2 is required for efficient group II intron splicing in mitochondria of Arabidopsis thaliana. *Plant Molecular Biology*, *72*(4), 459–467. https://doi.org/10.1007/s11103-009-9584-9

Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology*, *305*(3), 567–580. https://doi.org/10.1006/jmbi.2000.4315

Kugita, M., Yamamoto, Y., Fujikawa, T., Matsumoto, T., & Yoshinaga, K. (2003). RNA editing in hornwort chloroplasts makes more than half the genes functional. *Nucleic Acids Research*, *31*(9), 2417–2423. https://doi.org/10.1093/nar/gkg327

Kühn, K., Obata, T., Feher, K., Bock, R., Fernie, A. R., & Meyer, E. H. (2015). Complete mitochondrial complex I deficiency induces an up-regulation of respiratory fluxes that is abolished by traces of functional complex I. *Plant Physiology*, *168*(4), 1537–1549. https://doi.org/10.1104/pp.15.00589

Lamattina, L., Gonzales, D., Gualberto, J., & Grienenberger, J. -M. (1993). Higher plant mitochondria encode an homologue of the nuclear-encoded 30-kDa subunit of bovine mitochondrial complex I. *European Journal of Biochemistry*, *217*(3), 831–838. https://doi.org/10.1111/j.1432-1033.1993.tb18311.x

Lee, K., Leister, D., & Kleine, T. (2021). Arabidopsis mitochondrial transcription termination factor mterf2 promotes splicing of group iib introns. *Cells*, *10*(2), 1–29. https://doi.org/10.3390/cells10020315

Lee, K., Park, S. J., Park, Y. Il, & Kang, H. (2019). CFM9, a Mitochondrial CRM Protein, Is Crucial for Mitochondrial Intron Splicing, Mitochondria Function and Arabidopsis Growth and Stress Responses. *Plant and Cell Physiology*, *60*(11), 2538–2548. https://doi.org/10.1093/pcp/pcz147

Lesch, E., Schilling, M. T., Brenner, S., Yang, Y., Gruss, O. J., Knoop, V., & Schallenberg-Rüdinger, M. (2022). Plant mitochondrial RNA editing factors can perform targeted C-to-U editing of nuclear transcripts in human cells. *Nucleic Acids Research*, *50*(17), 9966–9983. https://doi.org/10.1093/nar/gkac752

Li, X., & Jiang, Y. (2024). Research Progress of Group II Intron Splicing Factors in Land Plant Mitochondria. In *Genes* (Vol. 15, Issue 2). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/genes15020176

Ligas, J., Pineau, E., Bock, R., Huynen, M. A., & Meyer, E. H. (2019). The assembly pathway of complex I in Arabidopsis thaliana. *Plant Journal*, *97*(3), 447–459. https://doi.org/10.1111/tpj.14133

Lurin, C., Andrés, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyère, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M. L., Mireau, H., Peeters, N., Renou, J. P., Szurek, B., Taconnat, L., & Small, I. (2004). Genome-wide analysis of arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell*, *16*(8), 2089–2103. https://doi.org/10.1105/tpc.104.022236

Maeda, A., Takenaka, S., Wang, T., Frink, B., Shikanai, T., & Takenaka, M. (2022). DYW deaminase domain has a distinct preference for neighboring nucleotides of the target RNA editing sites. *Plant Journal*, *111*(3), 756–767. https://doi.org/10.1111/tpj.15850

Malbert, B., Burger, M., Lopez-Obando, M., Baudry, K., Launay-Avon, A., Härtel, B., Verbitskiy, D., Jörg, A., Berthomé, R., Lurin, C., Takenaka, M., & Delannoy, E. (2020). The analysis of the editing defects in the dyw2 mutant provides new clues for the prediction of rna targets of arabidopsis e+-class ppr proteins. *Plants*, *9*(2). https://doi.org/10.3390/plants9020280

Malik, T. N., Doherty, E. E., Gaded, V. M., Hill, T. M., Beal, P. A., & Emeson, R. B. (2021). Regulation of RNA editing by intracellular acidification. *Nucleic Acids Research*, *49*(7), 4020–4036. https://doi.org/10.1093/nar/gkab157

Manna, S. (2015). An overview of pentatricopeptide repeat proteins and their applications. In *Biochimie* (Vol. 113, pp. 93–99). Elsevier B.V. https://doi.org/10.1016/j.biochi.2015.04.004

Matsumura, K., & Komiyama, M. (1997). Enormously Fast RNA Hydrolysis by Lanthanide(III) Ions under Physiological Conditions : Eminent Candidates for Novel Tools of Biotechnology). In *J. Biochem* (Vol. 122).

Michel, F., & Feral, J.-L. (1995). STRUCTURE AND ACTIVITIES OF GROUP II INTRONS. *Annual Reviews of Biochemistry*, *64*, 435–461. www.annualreviews.org

Millar, A. H., Whelan, J., Soole, K. L., & Day, D. A. (2011). Organization and regulation of mitochondrial respiration in plants. *Annual Review of Plant Biology*, *62*, 79–104. https://doi.org/10.1146/annurev-arplant-042110-103857

Mizrahi, R., Shevtsov-Tal, S., & Ostersetzer-Biran, O. (2022). Group II Intron-Encoded Proteins (IEPs/Maturases) as Key Regulators of Nad1 Expression and Complex I Biogenesis in Land

Plant Mitochondria. In *Genes* (Vol. 13, Issue 7). MDPI. https://doi.org/10.3390/genes13071137

Møller, I. M., Rasmusson, A. G., & Van Aken, O. (2021). Plant mitochondria – past, present and future. In *Plant Journal* (Vol. 108, Issue 4, pp. 912–959). John Wiley and Sons Inc. https://doi.org/10.1111/tpj.15495

Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., Toyooka, K., Matsuoka, K., Jinbo, T., & Kimura, T. (2007). Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *Journal of Bioscience and Bioengineering*, *104*(1), 34–41. https://doi.org/10.1263/jbb.104.34

Nakamura, T., & Sugita, M. (2008). A conserved DYW domain of the pentatricopeptide repeat protein possesses a novel endoribonuclease activity. *FEBS Letters*, *582*(30), 4163–4168. https://doi.org/10.1016/j.febslet.2008.11.017

Obayashi, T., Hibara, H., Kagaya, Y., Aoki, Y., & Kinoshita, K. (2022). ATTED-II v11: A Plant Gene Coexpression Database Using a Sample Balancing Technique by Subagging of Principal Components. *Plant and Cell Physiology*, *63*(6), 869–881. https://doi.org/10.1093/pcp/pcac041

Okuda, K., Chateigner-Boutin, A. L., Nakamura, T., Delannoy, E., Sugita, M., Myouga, F., Motohashi, R., Shinozaki, K., Small, I., & Shikanai, T. (2009). Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in Arabidopsis chloroplasts. *Plant Cell*, *21*(1), 146–156. https://doi.org/10.1105/tpc.108.064667

Oldenkott, B., Yang, Y., Lesch, E., Knoop, V., & Schallenberg-Rüdinger, M. (2019). Plant-type pentatricopeptide repeat proteins with a DYW domain drive C-to-U RNA editing in Escherichia coli. *Communications Biology*, *2*(1). https://doi.org/10.1038/s42003-019-0328-3

Ostersetzer-Biran, O. (2016). Respiratory complex i and embryo development. *Journal of Experimental Botany*, *67*(5), 1205–1207. https://doi.org/10.1093/jxb/erw051

Perales, M., Eubel, H., Heinemeyer, J., Colaneri, A., Zabaleta, E., & Braun, H. P. (2005). Disruption of a nuclear gene encoding a mitochondrial gamma carbonic anhydrase reduces complex I and supercomplex I+III2 levels and alters mitochondrial physiology in Arabidopsis. *Journal of Molecular Biology*, *350*(2), 263–277. https://doi.org/10.1016/j.jmb.2005.04.062

Peterson, R., Slovin, J. P., & Chen, C. (2010). A simplified method for differential staining of aborted and non-aborted pollen grains. *International Journal of Plant Biology*, *1*(2), 66–69. https://doi.org/10.4081/pb.2010.e13

Pineau, B., Layoune, O., Danon, A., & De Paepe, R. (2008). L-galactono-1,4-lactone dehydrogenase is required for the accumulation of plant respiratory complex I. *Journal of Biological Chemistry*, *283*(47), 32500–32505. https://doi.org/10.1074/jbc.M805320200

Rio, D. C. (2014). Northern blots for small RNAs and microRNAs. *Cold Spring Harbor Protocols*, *2014*(7), 793–797. https://doi.org/10.1101/pdb.prot080838

Ruwe, H., Wang, G., Gusewski, S., & Schmitz-Linneweber, C. (2016). Systematic analysis of plant mitochondrial and chloroplast small RNAs suggests organelle-specific mRNA stabilization mechanisms. In *Nucleic Acids Research* (Vol. 44, Issue 15, pp. 7406–7417). Oxford University Press. https://doi.org/10.1093/nar/gkw466

Salone, V., Rüdinger, M., Polsakiewicz, M., Hoffmann, B., Groth-Malonek, M., Szurek, B., Small, I., Knoop, V., & Lurin, C. (2007). A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Letters*, *581*(22), 4132–4138. https://doi.org/10.1016/j.febslet.2007.07.075

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., … Sherry, S. T. (2022). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, *50*(D1), D20–D26. https://doi.org/10.1093/nar/gkab1112

Schertl, P., & Braun, H. P. (2014). Respiratory electron transfer pathways in plant mitochondria. In *Frontiers in Plant Science* (Vol. 5, Issue APR). Frontiers Research Foundation. https://doi.org/10.3389/fpls.2014.00163

Schmitz-Linneweber, C., & Small, I. (2008). Pentatricopeptide repeat proteins: a socket set for organelle gene expression. In *Trends in Plant Science* (Vol. 13, Issue 12, pp. 663–670). https://doi.org/10.1016/j.tplants.2008.10.001

Schmitz-Linneweber, C., Williams-Carrier, R. E., Williams-Voelker, P. M., Kroeger, T. S., Vichas, A., & Barkan, A. (2006). A pentatricopeptide repeat protein facilitates the trans-splicing of the maize chloroplast rps12 pre-mRNA. *Plant Cell*, *18*(10), 2650–2663. https://doi.org/10.1105/tpc.106.046110

Shevtsov-Tal, S., Best, C., Matan, R., Chandran, S. A., Brown, G. G., & Ostersetzer-Biran, O. (2021). nMAT3 is an essential maturase splicing factor required for holo-complex I biogenesis and embryo development in Arabidopsis thaliana plants. *Plant Journal*, *106*(4), 1128–1147. https://doi.org/10.1111/tpj.15225

Sloan, D. B., Wu, Z., & Sharbrough, J. (2018). Correction of persistent errors in arabidopsis reference mitochondrial genomes. In *Plant Cell* (Vol. 30, Issue 3, pp. 525–527). American Society of Plant Biologists. https://doi.org/10.1105/tpc.18.00024

Small, I. D., & Peeters, N. (2000a). The PPR motif – a TPR-related motif prevalent in plant organellar proteins. *Trends in Biochemical Sciences*, *25*(2), 46–47. https://doi.org/https://doi-org.kyoto-u.idm.oclc.org/10.1016/S0968-0004(99)01520-0

Small, I. D., & Peeters, N. (2000b). The PPR motif – a TPR-related motif prevalent in plant organellar proteins. *Trends in Biochemical Sciences*, *25*(2), 45–47. http://dodo.cpmc.

Sonnhammer, E. L. L., Von Heijne, G., & Krogh, A. (1998). *A hidden Markov model for predicting transmembrane helices in protein sequences* (Issue 1). AAAI Press. www.aaai.org

Soufari, H., Parrot, C., Kuhn, L., Waltz, F., & Hashem, Y. (2020). Specific features and assembly of the plant mitochondrial complex I revealed by cryo-EM. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-18814-w

Stoller, R. G., Myers, C. E., & Chabnert, B. A. (1978). ANALYSIS OF CYTIDINE DEAMINASE AND TETRAHYDROURIDINE INTERACTION BY USE OF LIGAND TECHNIQUES. In *Bwhrmlcal Pharmacology* (Vol. 27). Pergamon Press. IY7X Prlntcd m Grrat Bnlaln.

Streit, S., Michalski, C. W., Erkan, M., Kleeff, J., & Friess, H. (2009). Northern blot analysis for detection and quantification of RNA in pancreatic cancer cells and tissues. *Nature Protocols*, *4*(1), 37–43. https://doi.org/10.1038/nprot.2008.216

Su, C., Zhao, H., Zhao, Y., Ji, H., Wang, Y., Zhi, L., & Li, X. (2017). RUG3 and ATM synergistically regulate the alternative splicing of mitochondrial nad2 and the DNA damage response in Arabidopsis thaliana. *Scientific Reports*, *7*. https://doi.org/10.1038/srep43897

Sultan, L. D., Mileshina, D., Grewe, F., Rolle, K., Abudraham, S., Głodowicz, P., Niazi, A. K., Keren, I., Shevtsov, S., Klipcan, L., Barciszewski, J., Mower, J. P., Dietrich, A., & Ostersetzer-Biran, O. (2016). The reverse transcriptase/RNA maturase protein MatR is required for the splicing of various group ii introns in brassicaceae mitochondria. *Plant Cell*, *28*(11), 2805–2829. https://doi.org/10.1105/tpc.16.00398

Sunderhaus, S., Dudkina, N. V., Jänsch, L., Klodmann, J., Heinemeyer, J., Perales, M., Zabaleta, E., Boekema, E. J., & Braun, H. P. (2006). Carbonic anhydrase subunits form a matrix-exposed domain attached to the membrane arm of mitochondrial complex I in plants. *Journal of Biological Chemistry*, *281*(10), 6482–6488. https://doi.org/10.1074/jbc.M511542200

Takenaka, M. (2022). Quantification of Mitochondrial RNA Editing Efficiency Using Sanger Sequencing Data. In O. Van Aken & A. G. Rasmusson (Eds.), *Plant Mitochondria. Methods in Molecular Biology* (Humana, Vol. 2363, pp. 263–278). Springer US. https://doi.org/10.1007/978-1-0716-1653-6

Takenaka, M., & Brennicke, A. (2003). In Vitro RNA Editing in Pea Mitochondria Requires NTP or dNTP, Suggesting Involvement of an RNA Helicase. *Journal of Biological Chemistry*, *278*(48), 47526–47533. https://doi.org/10.1074/jbc.M305341200

Takenaka, M., Takenaka, S., Barthel, T., Frink, B., Haag, S., Verbitskiy, D., Oldenkott, B., Schallenberg-Rüdinger, M., Feiler, C. G., Weiss, M. S., Palm, G. J., & Weber, G. (2021). DYW domain structures imply an unusual regulation principle in plant organellar RNA editing catalysis. *Nature Catalysis*, *4*(6), 510–522. https://doi.org/10.1038/s41929-021-00633-x

Takenaka, M., Verbitskiy, D., van der Merwe, J. A., Zehrmann, A., Plessmann, U., Urlaub, H., & Brennicke, A. (2007). In vitro RNA editing in plant mitochondria does not require added energy. *FEBS Letters*, *581*(14), 2743–2747. https://doi.org/10.1016/j.febslet.2007.05.025

Takenaka, M., Zehrmann, A., Verbitskiy, D., Härtel, B., & Brennicke, A. (2013). RNA editing in plants and its evolution. In *Annual Review of Genetics* (Vol. 47, pp. 335–352). Annual Reviews Inc. https://doi.org/10.1146/annurev-genet-111212-133519

Thomas, P. D., Ebert, D., Muruganujan, A., Mushayahama, T., Albou, L. P., & Mi, H. (2022). PANTHER: Making genome-scale phylogenetics accessible to all. In *Protein Science* (Vol. 31, Issue 1, pp. 8–22). John Wiley and Sons Inc. https://doi.org/10.1002/pro.4218

Toma-Fukai, S., Sawada, Y., Maeda, A., Shimizu, H., Shikanai, T., Takenaka, M., & Shimizu, T. (2023). Structural insight into the activation of an Arabidopsis organellar C-to-U RNA editing enzyme by active site complementation. *Plant Cell*, *35*(6), 1888–1900. https://doi.org/10.1093/plcell/koac318

Unseld, M., Marienfeld, J. R., Brandt, P., & Brennicke Axel. (1997). The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides. *Nature Genetics*, *15*, 57–61. http://www.nature.com/naturegenetics

Wahleithner, J. A., Macfarlane, J. L., & Wolstenholme, D. R. (1990). A sequence encoding a maturase-related protein in a group II intron of a plant mitochondrial nadl gene (broad bean/reverse transcriptase/evolutionary conservation/sequence rearrangements). In *Proc. Nail. Acad. Sci. USA* (Vol. 87).

Wang, C., Aubé, F., Quadrado, M., Dargel-Graffin, C., & Mireau, H. (2018). Three new pentatricopeptide repeat proteins facilitate the splicing of mitochondrial transcripts and complex i biogenesis in Arabidopsis. *Journal of Experimental Botany*, *69*(21), 5131–5140. https://doi.org/10.1093/jxb/ery275

Wang, X., An, Y., Xu, P., & Xiao, J. (2021). Functioning of PPR Proteins in Organelle RNA Metabolism and Chloroplast Biogenesis. In *Frontiers in Plant Science* (Vol. 12). Frontiers Media S.A. https://doi.org/10.3389/fpls.2021.627501

Wang, Y., Lyu, W., Berkowitz, O., Radomiljac, J. D., Law, S. R., Murcha, M. W., Carrie, C., Teixeira, P. F., Kmiec, B., Duncan, O., Van Aken, O., Narsai, R., Glaser, E., Huang, S., Roessner, U., Millar, A. H., & Whelan, J. (2016). Inactivation of Mitochondrial Complex i Induces the Expression of a Twin Cysteine Protein that Targets and Affects Cytosolic, Chloroplastidic and Mitochondrial Function. *Molecular Plant*, *9*(5), 696–710. https://doi.org/10.1016/j.molp.2016.01.009

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G. V., & Provart, N. J. (2007). An "electronic fluorescent pictograph" Browser for exploring and analyzing large-scale biological data sets. *PLoS ONE*, *2*(8). https://doi.org/10.1371/journal.pone.0000718

Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T., & Nakamura, T. (2013). Elucidation of the RNA Recognition Code for Pentatricopeptide Repeat Proteins Involved in Organelle RNA Editing in Plants. *PLoS ONE*, *8*(3). https://doi.org/10.1371/journal.pone.0057286

Yang, L., Zhang, J., He, J., Qin, Y., Hua, D., Duan, Y., Chen, Z., & Gong, Z. (2014). ABA-Mediated ROS in Mitochondria Regulate Root Meristem Activity by Controlling

PLETHORA Expression in Arabidopsis. *PLoS Genetics*, *10*(12). https://doi.org/10.1371/journal.pgen.1004791

Yoshinaga, K., Iinuma, H., Masuzawa, T., & Uedal, K. (1996). Extensive RNA editing of U to C in addition to C to U substitution in the rbcL transcripts of hornwort chloroplasts and the origin of RNA editing in green plants. In *Nucleic Acids Research* (Vol. 24, Issue 6). Oxford University Press.

Zhang, X., Henriques, R., Lin, S. S., Niu, Q. W., & Chua, N. H. (2006). Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. *Nature Protocols*, *1*(2), 641–646. https://doi.org/10.1038/nprot.2006.97

Zhao, P., Wang, F., Li, N., Shi, D. Q., & Yang, W. C. (2020). Pentatricopeptide repeat protein MID1 modulates nad2 intron 1 splicing and Arabidopsis development. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-58495-5

Zimmerly, S., & Semper, C. (2015). Evolution of group II introns. In *Mobile DNA* (Vol. 6, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s13100-015-0037-5

Zmudjak, M., Colas des Francs-Small, C., Keren, I., Shaya, F., Belausov, E., Small, I., & Ostersetzer-Biran, O. (2013). mCSF1, a nucleus-encoded CRM protein required for the processing of many mitochondrial introns, is involved in the biogenesis of respiratory complexes I and IV in Arabidopsis. *New Phytologist*, *199*(2), 379–394. https://doi.org/10.1111/nph.12282

Zmudjak, M., & Ostersetzer-Biran, O. (2018). RNA Metabolism and Transcript Regulation. In *Annual Plant Reviews online* (pp. 143–184). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119312994.apr0548

Zmudjak, M., Shevtsov, S., Sultan, L. D., Keren, I., & Ostersetzer-Biran, O. (2017). Analysis of the roles of the arabidopsis nMAT2 and PMH2 proteins provided with new insights into the regulation of group II intron splicing in land-plant mitochondria. *International Journal of Molecular Sciences*, *18*(11). https://doi.org/10.3390/ijms18112428

Zoschke, R., & Barkan, A. (2015). Genome-wide analysis of thylakoid-bound ribosomes in maize reveals principles of cotranslational targeting to the thylakoid membrane. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(13), E1678–E1687. https://doi.org/10.1073/pnas.1424655112

Acknowledgements

All of this would not be possible without the support of those around me.

Of course, I would like to thank Dr. Mizuki Takenaka first for his leadership and positive attitude. He has been a constant driving force throughout my studies here and has supported me both during my daily life and in each experiment as well.

My appreciation extends to his wife, Sachi, and his two children who have all treated me so well since I have been here and have been wonderful to work with.

This continues to all of the other professors in our lab: Dr. Toshiharu Shikanai, Dr. Ryuji Tsugeki, and Dr. Yoshiki Nishimura. All four of our laboratory's professors have given me plenty of advice and support. And all have been excellent discussion partners, no matter which facet of life the topic extended to.

Dr. Tomoo Shimada and Dr. Yoshito Oka contributed many reagents and kind conversations during my research period and I am grateful for their constant support.

I want to mention Dr. Oren Ostersetzer-Biran and his students in Israel who contributed their time and data to produce a major part of my research.

I'd also like to thank Shohei Yamaoka for expertise with confocal microscopy and his support and kindness.

My experiences in our laboratory would be nothing without the rest of the people that inhabited it and I'd like to quickly name off each and every one of them: Naoto Doi, Sho Fujii, Takashi Hamaji, Kota Ishibashi, Jingxiu Ji, Yuto Kakiuchi, Haruki Kanazawa, Haruka Higashi, Kirie Kaneko, Nanako Katake, Aine Kawashima, Ayako Maeda, Takehiro Matsumoto, Issei Ohara, Sota Okui, , Chao Huang, Yu Ogawa, Nayu Otsuki, Ruchika, Mari Takusagawa, Masumi Taniguchi, Yukari Tokutsu, Tenghua Wang, Ziling Weng, Hiroshi Yamamoto, Hana Yano, Hato Yonamine, and Qi Zhou.

I’d also like to thank Leonardo Basso, Ryouhei Kobayashi, Kazuya Nomura, and Deborah Schatz-Daas who have been supporting me inside and outside our laboratory.

And finally, I’d like to thank my friends and family for their support through the many years I’ve been here. There are countless others I would like to mention, but I truly appreciate everyone and made me who I am.

Lastly, thank *you* for reading this dissertation!

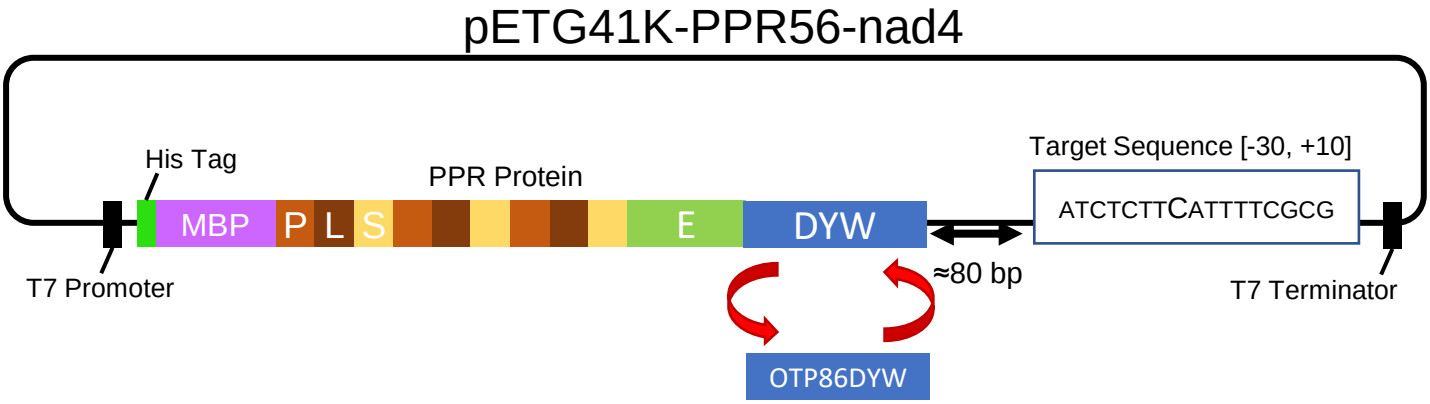


Figure 1: Overall design of the PPR56 expression system
The pETG41K vector was utilized to express PPR56 and PPR56^{PPRE1E2}-OTP86^{DYW} within *E. coli* bacteria. Gene expression was controlled using a T7 promoter and terminator. The PPR protein included two tags at the N-terminus, a 6xHis tag and maltose binding protein tag (purple). The DYW domain in blue is swappable with domains from other proteins and has been swapped with that from OTP86 in these experiments. Also included downstream of the PPR protein is the nad4eU272SL target RNA sequence which is used not only for assessing in vivo editing but also for creation of purified substrate RNA for in vitro experiments.

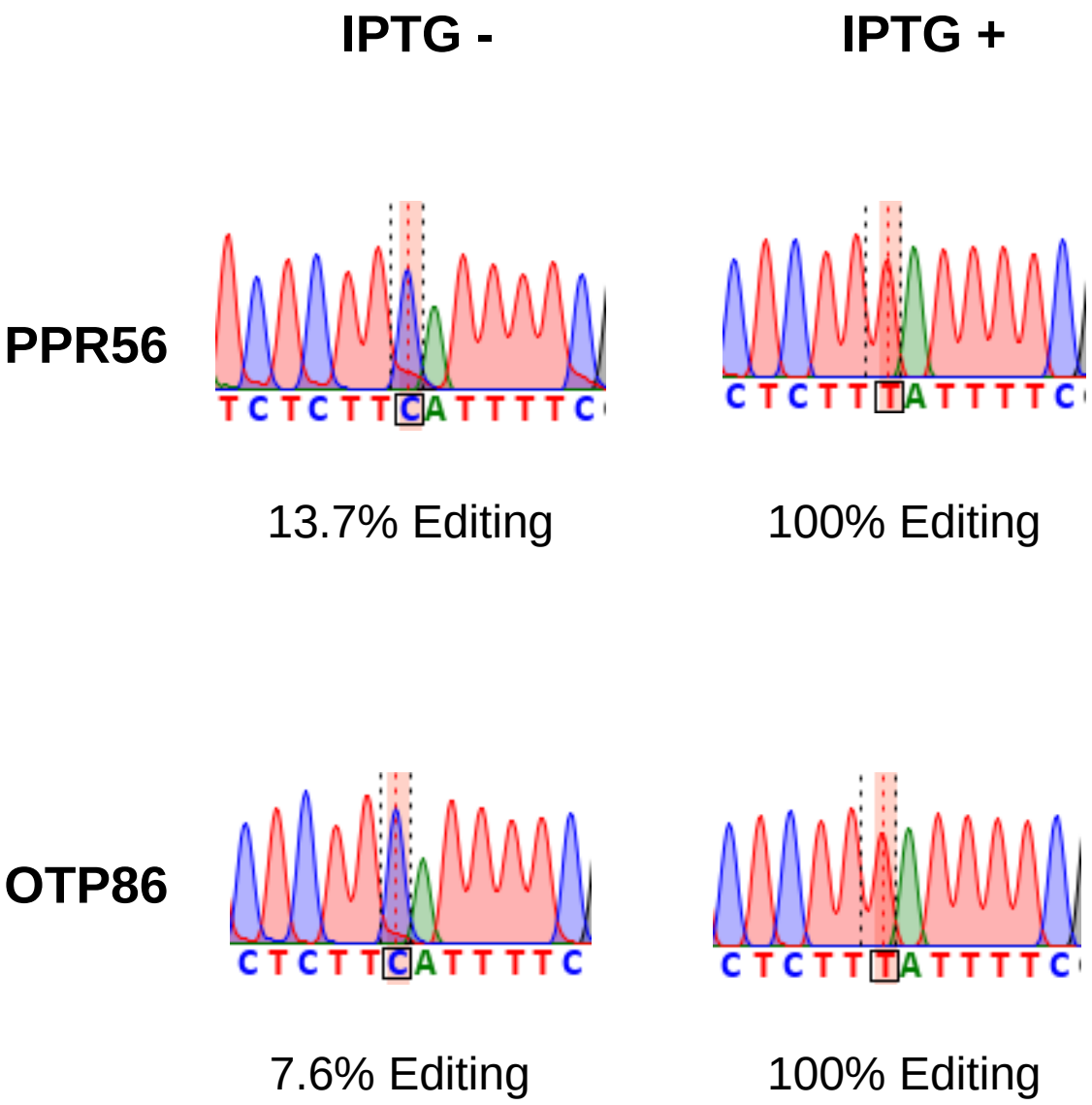


Figure 2: RNA editing of nad4eU272SL under in vivo conditions
E. coli was transformed with the PPR56 and PPR56^{PPRE1E2}-OTP86^{DYW} (OTP86) constructs and protein expression was induced with IPTG. In vivo RNA editing was examined before and after induction. Pre-induced RNAs had only minor editing possibly due to leaky expression, but properly induced RNA was 100% edited in both PPR56 and OTP86 transformed bacteria.

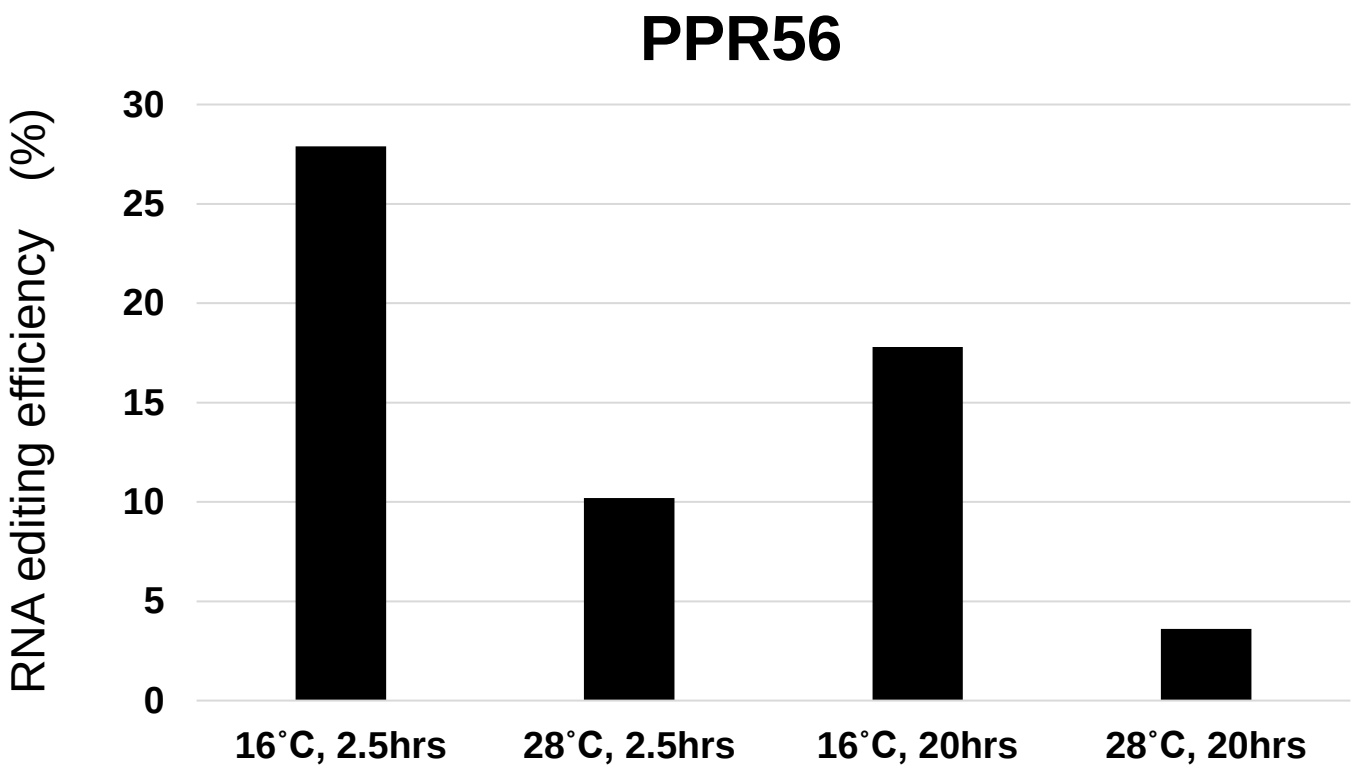


Figure 3: In vitro lysate editing is temperature and time sensitive
Lysates taken from recombinant PPR56-expressing bacteria were mixed with substrate RNA and left to react for 2.5 and 20 hours at both 16 and 28°C. Editing efficiencies were higher at 16°C and also at the shorter duration of 2.5 hours.

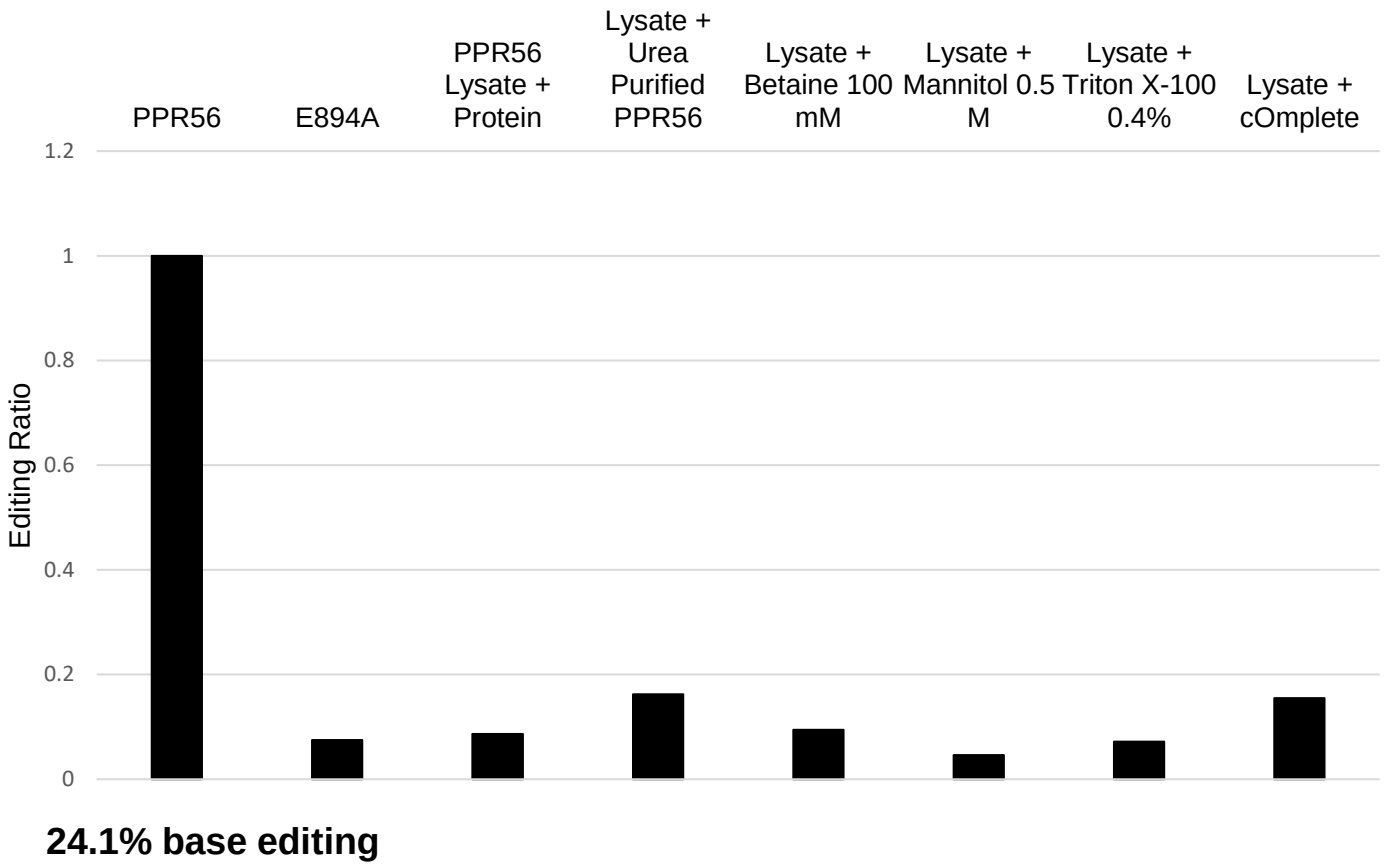


Figure 4: Addition of protein stabilizing reagents fail to increase lysate editing

Lysate in vitro editing experiments were performed using PPR56 expressing bacterial lysates. Before reaction incubation, addition reagents were added to attempt to stimulate editing. This included additional purified protein, protein purified in urea-containing denaturing conditions, betaine, mannitol, Triton X-100 and cOmplete protease inhibitor. All substantially reduced editing efficiencies compared to the base trial with 24.1% editing. All changes are represented on a log base 2 scale.

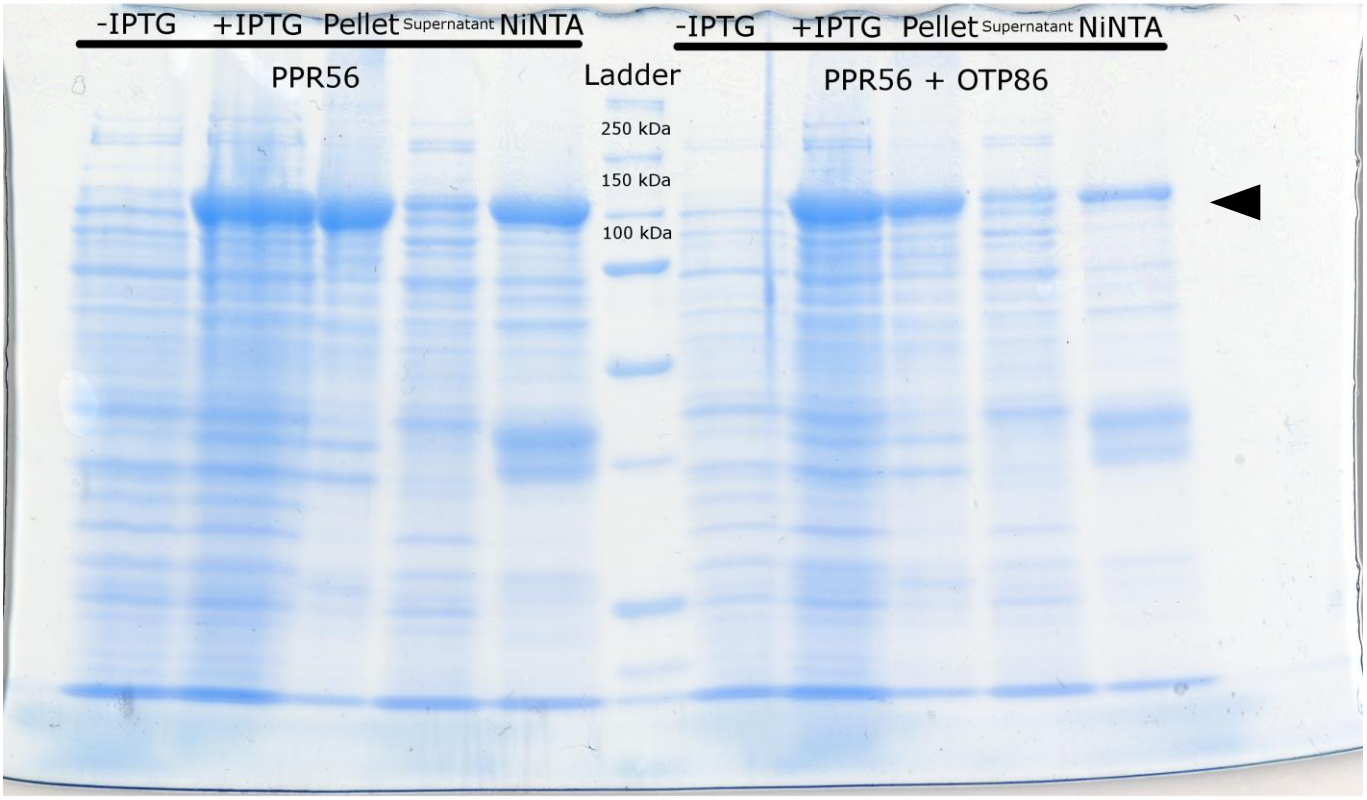


Figure 5: First purification using NiNTA targeting 6xHis tags

Coomassie brilliant blue stains of the initial purified proteins were performed in PPR56 and PPR56^{PPRE1E2}-OTP86^{DYW} (OTP86) using the 6xHis tag as a target. Bacterial lysates before and after induction are represented as –IPTG and +IPTG. The centrifuged lysate pellet contains insoluble proteins while the lysate supernatant is the soluble fraction. Supernatant was used for further purification with NiNTA resin as it retains activity, and a sufficient amount of protein was obtained.

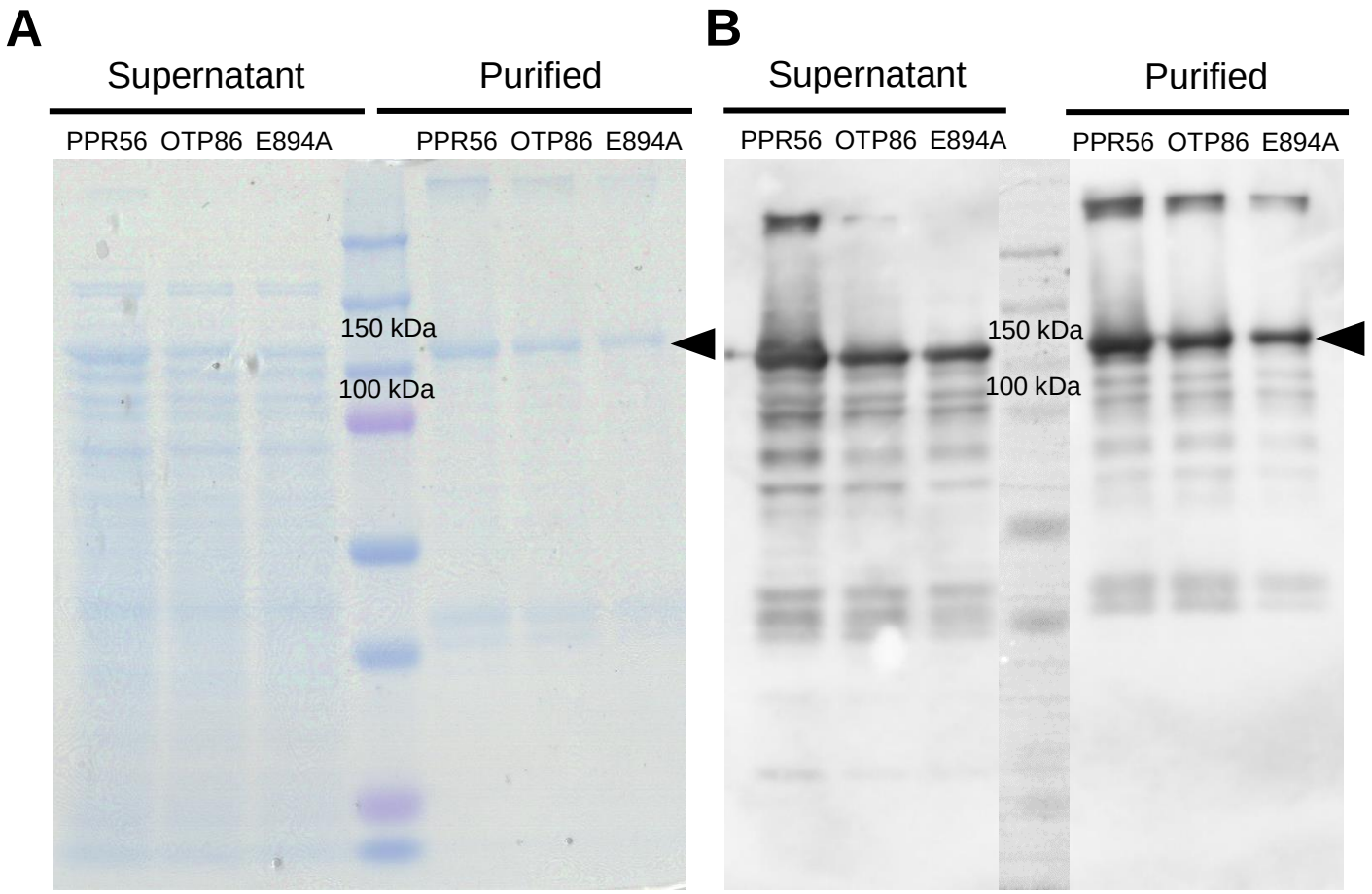
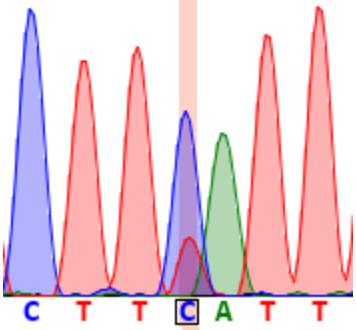


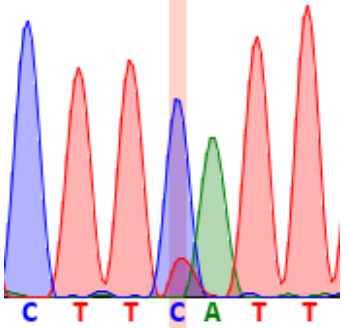
Figure 6: MBP-isolated PPR56 has less secondary bands and strong signal in western blots

(A) Coomassie brilliant blue stains of MBP purified recombinant PPR56, PPR56^{PPRE1E2}-OTP86^{DYW} (OTP86), and PPR56-E894A showed significantly less secondary banding compared to the NiNTA resin counterparts. (B) Western blot using Anti-His antibodies of MBP purified recombinant PPR56

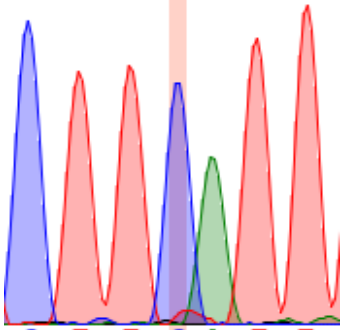
PPR56 Bacterial Lysate
24.1% Editing



PPR56 Purified Protein
16.8% Editing



PPR56 Purified Protein
(Urea Purified)
6.0% Editing



PPR56^{PPRE1E2}-OTP86^{DYW}
Purified Protein
7.6% Editing

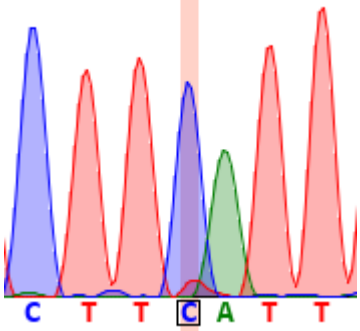


Figure 7: Purified PPR56 protein can edit substrate RNA in the absence of lysate or other reagents

In vitro RNA editing experiments were performed with PPR56 bacterial lysate as well as purified PPR56 alone. The MBP purified protein was able to edit its target RNA at 15% editing with no lysate or additional reagents added, confirming that a single protein is sufficient for RNA editing. Protein purified using denaturing conditions in urea was unable to edit to a significant degree.

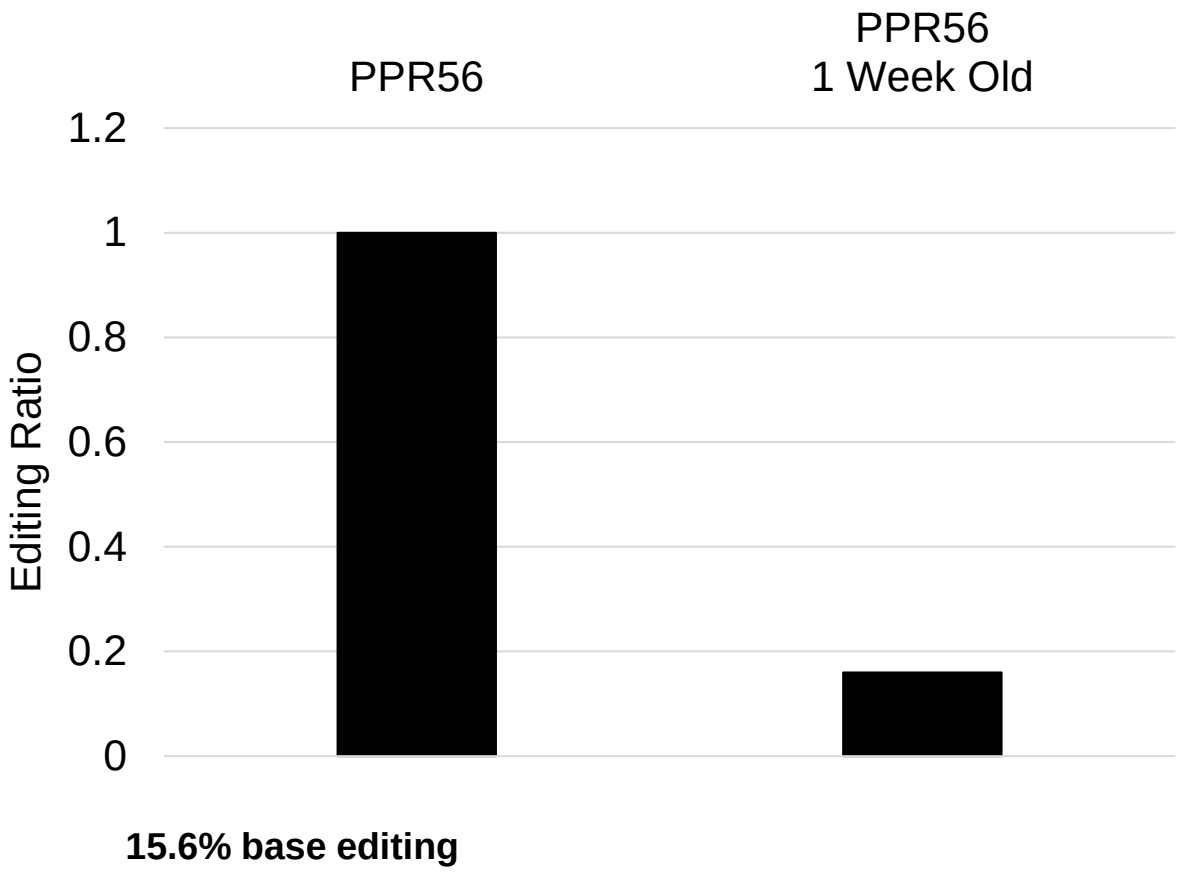


Figure 8: Recombinant PPR56 protein activity is sensitive to heat and time

Recombinant PPR56 protein was tested for long term viability by both heat shocking protein at 85°C for 15 minutes and by using older protein stored at 4°C for one week. Both saw reductions to editing efficiencies which were compounded when both were combined. PPR56^{PPRE1E2}-OTP86^{DYW} (OTP86) protein similarly was affected by heat shock.

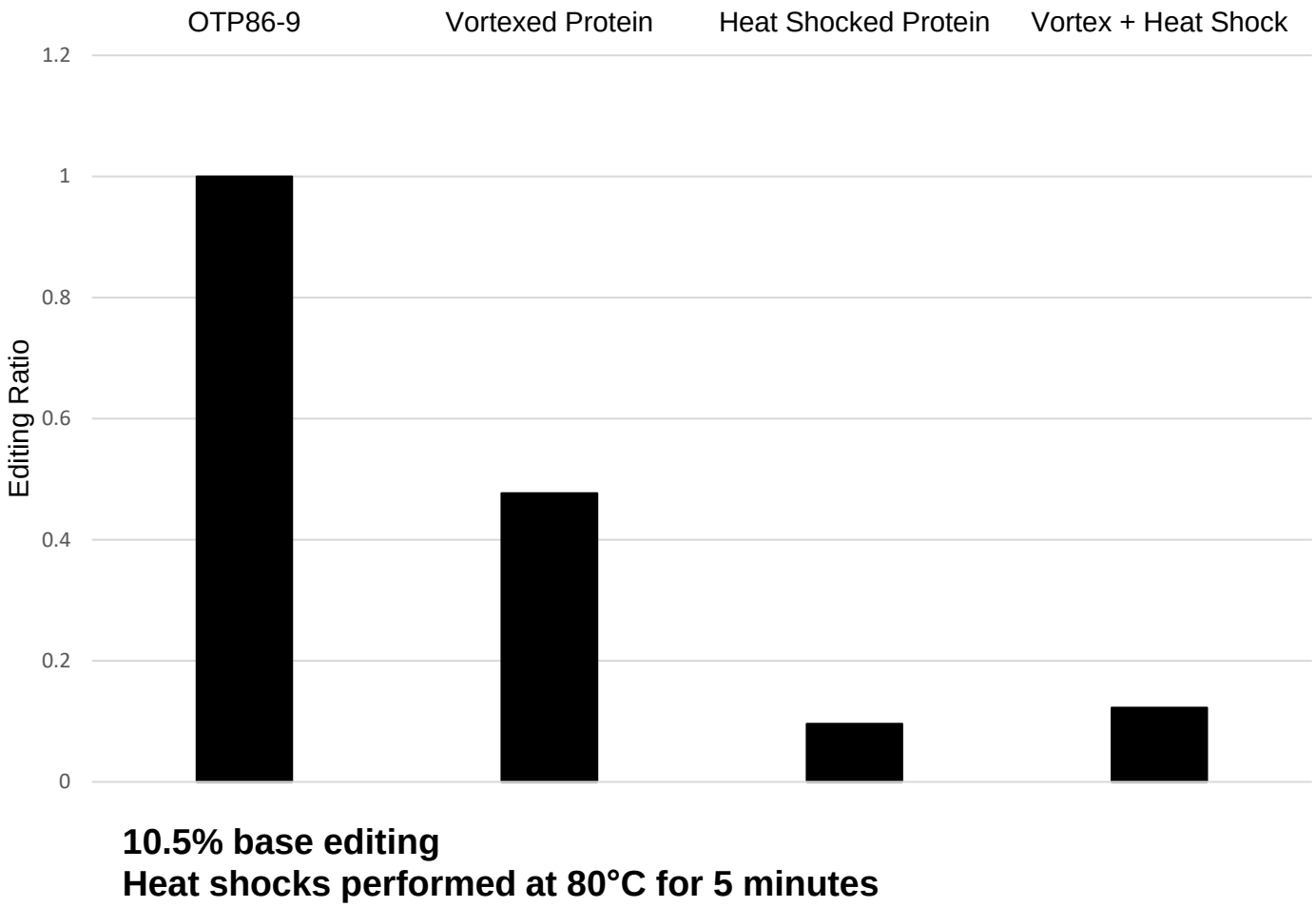


Figure 9: PPR56^{PPRE1E2}-OTP86^{DYW} is sensitive to heat and vibration by vortex

PPR56^{PPRE1E2}-OTP86^{DYW} (OTP86) was vortexed for about 10 seconds and its editing efficiency was found to decrease by over a half. OTP86 protein was exposed to 80°C heat for just 5 minutes and saw large decreases in editing similar in scale to previous heat shock trials.

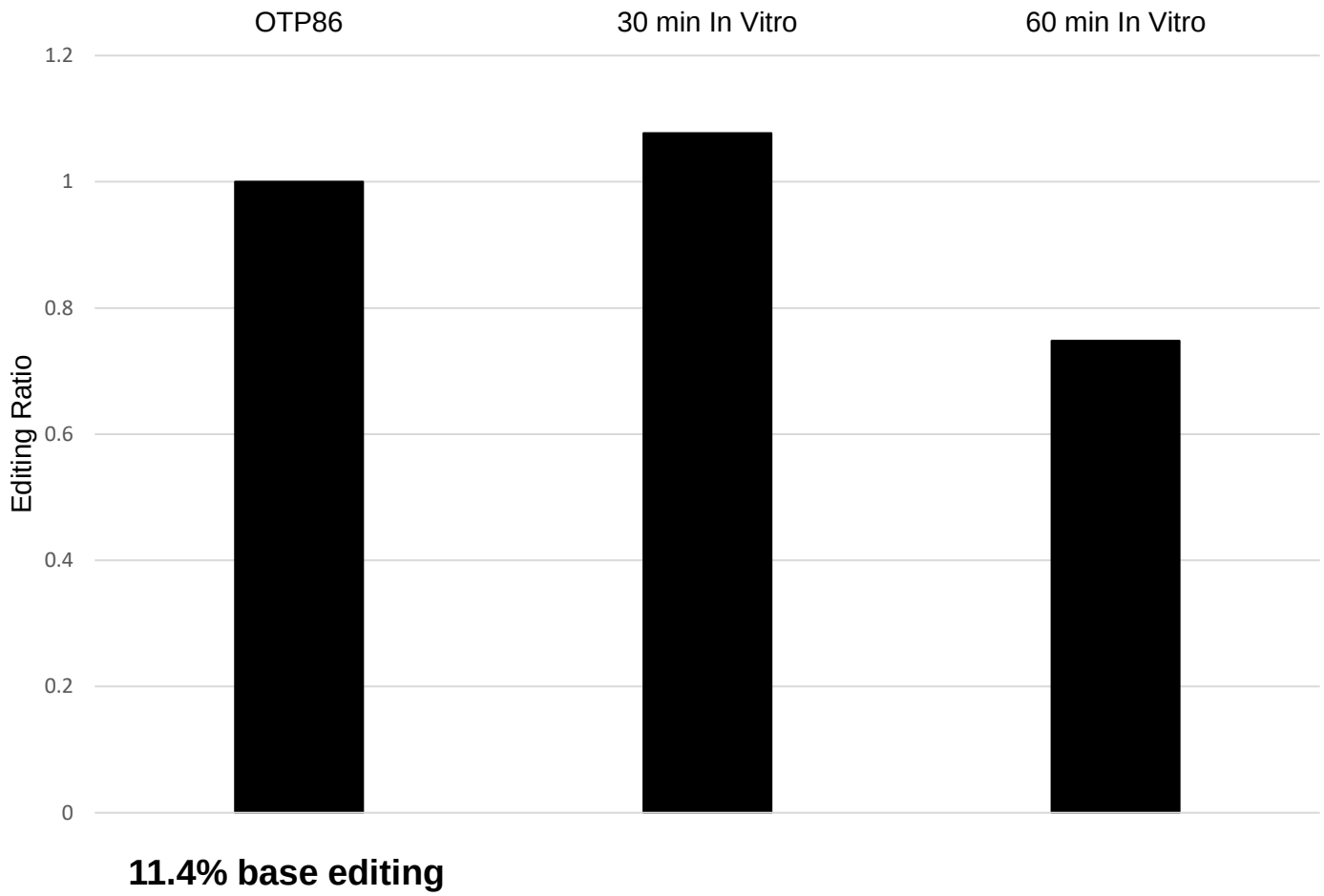
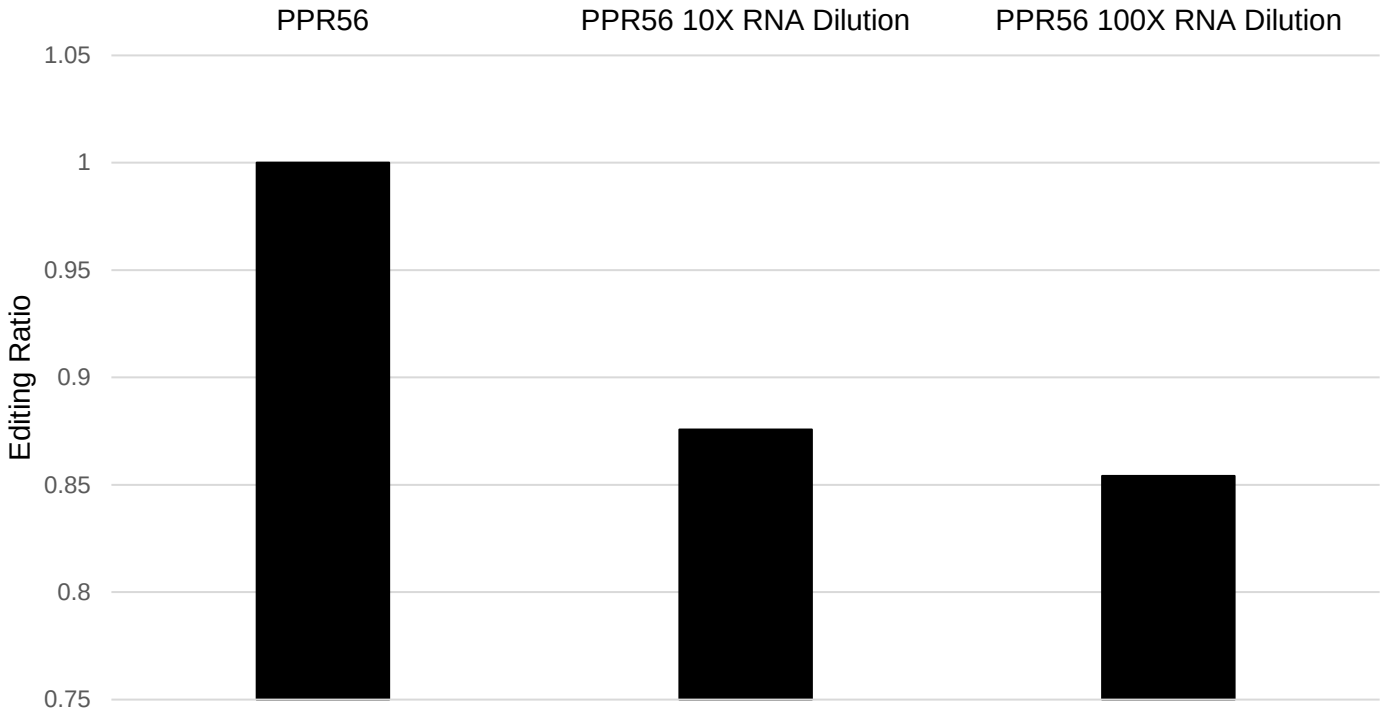


Figure 10: PPR56^{PPRE1E2}-OTP86^{DYW} mediated RNA editing occurs rapidly

PPR56^{PPRE1E2}-OTP86^{DYW} (OTP86) was used in in vitro editing experiments where incubation times with target RNA was reduced to 30 and 60 minutes compared to the usual two and a half hours. Editing to similar extents as the control were observed with only 30 minutes incubation time showing that editing reaches plateau rapidly. Reductions in editing over longer stretches of time may be due to RNA degradation.



32.1% base editing

Figure 11: Editing efficiency is not affected by RNA substrate concentration

nad4eU272SL target RNA usually had a concentration of around 0.2 ng/µL in the final reaction mix used in in vitro editing experiments. This was diluted by ten times to 0.02 ng/µL and to 0.002 ng/µL. Editing efficiency was slightly decreased as concentration was lowered, but overall editing was still about 85% of controls, thus RNA concentration does not greatly affect the reaction.

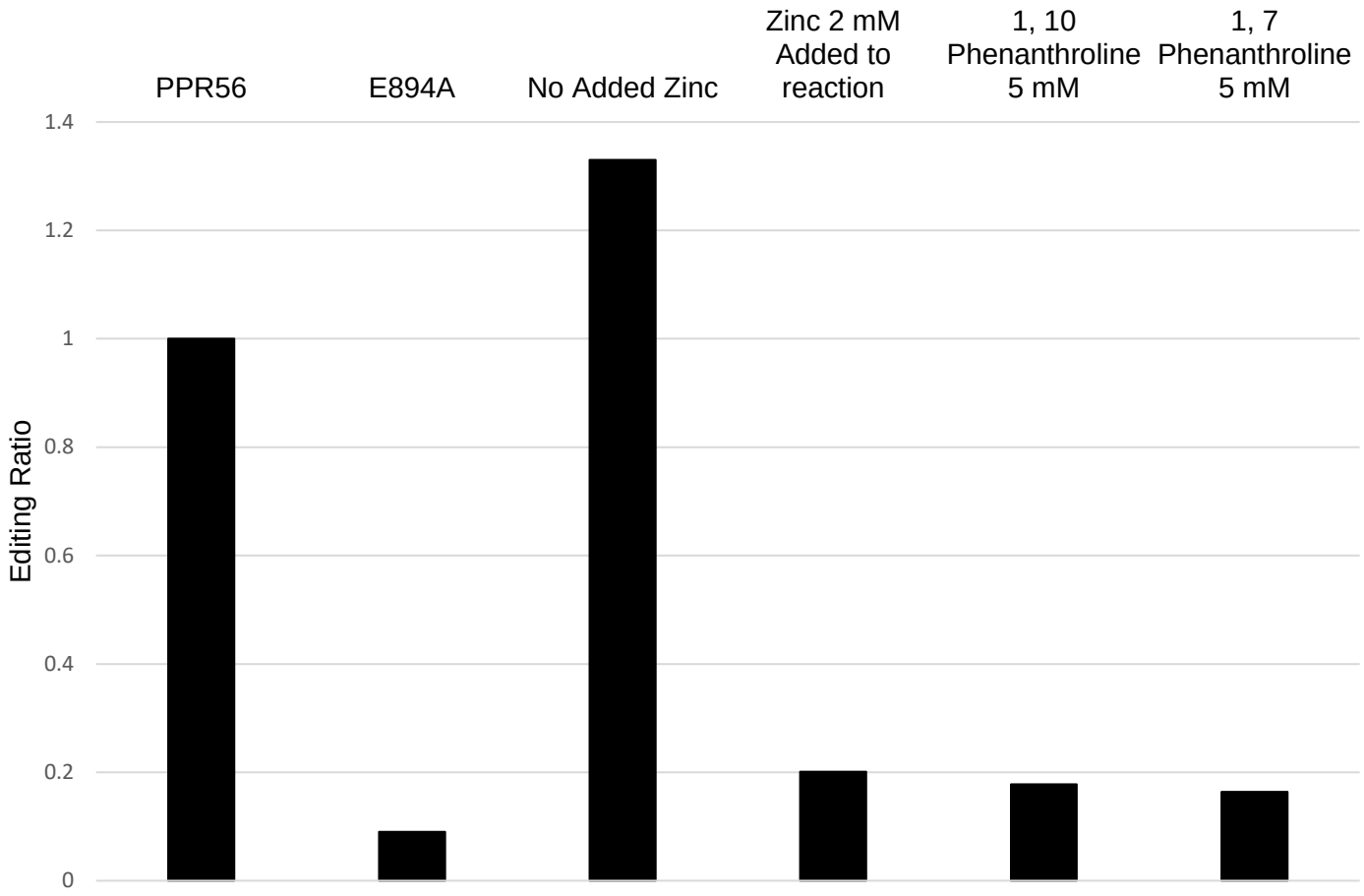
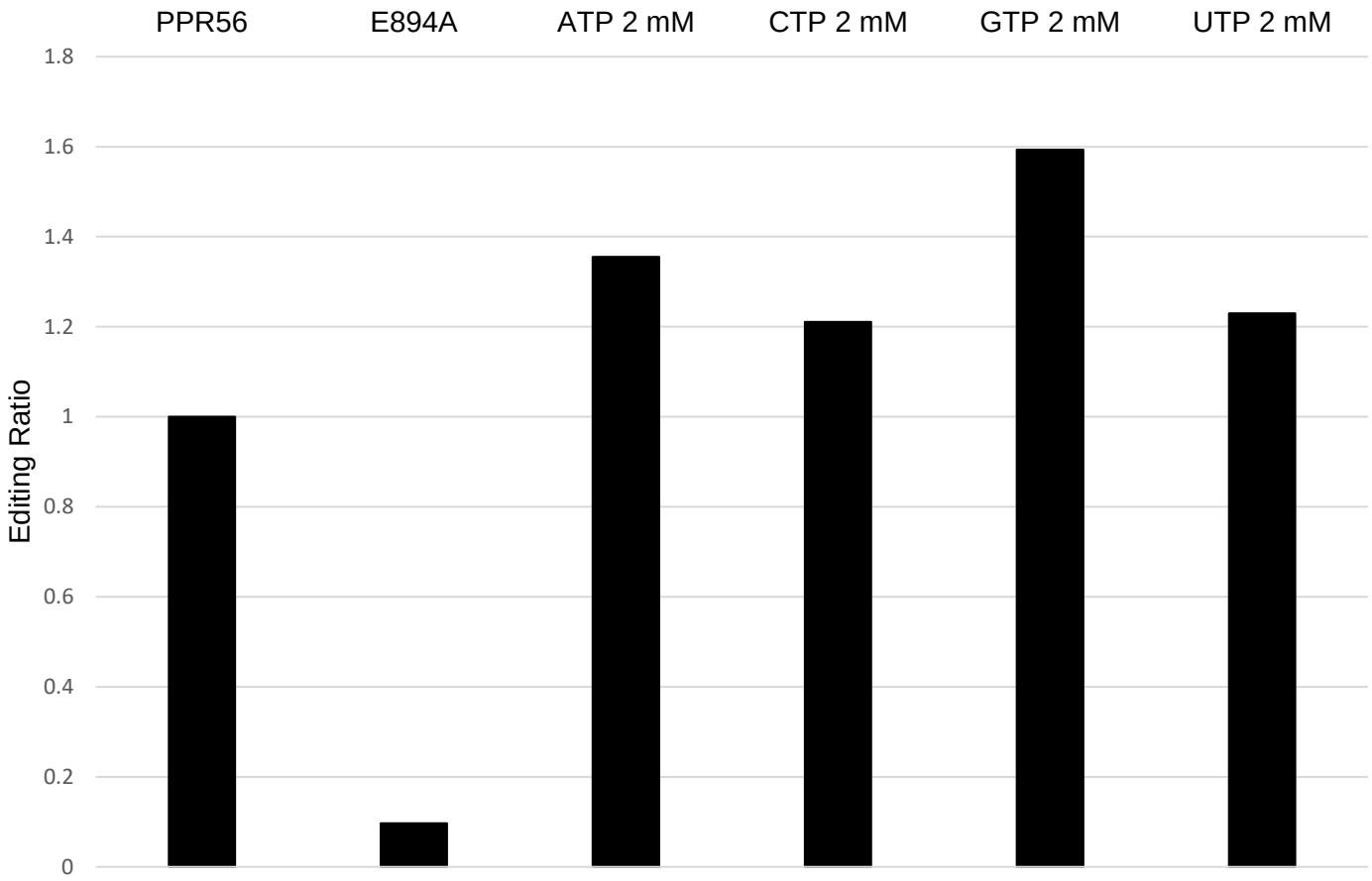


Figure 12: PPR56 activity may have a preferred zinc concentration

PPR56 has two zinc coordination domains which may be influenced by zinc concentrations. During bacterial growth, 0.4 mM $ZnSO_4$ is added to stimulate growth. This was removed in the “no added zinc” trial with no adverse effects. 2 mM of $ZnSO_4$ supplemented directly into the in vitro reaction and reduced editing efficiency. Zinc chelator 1, 10 phenanthroline and its inactive form 1, 7 phenanthroline. Both lowered editing equally, suggesting the effect was due to the addition of the chemical and not due to the removal of zinc ions.



16.3% base editing

Figure 13: NTPs stimulate the editing reaction

PPR56 mediated in vitro editing were performed in the presence of ATP, CTP, GTP, and UTP at at 2 mM. All four NTPs increased the amount of editing to various degrees, with GTP and ATP seeing the largest increases.

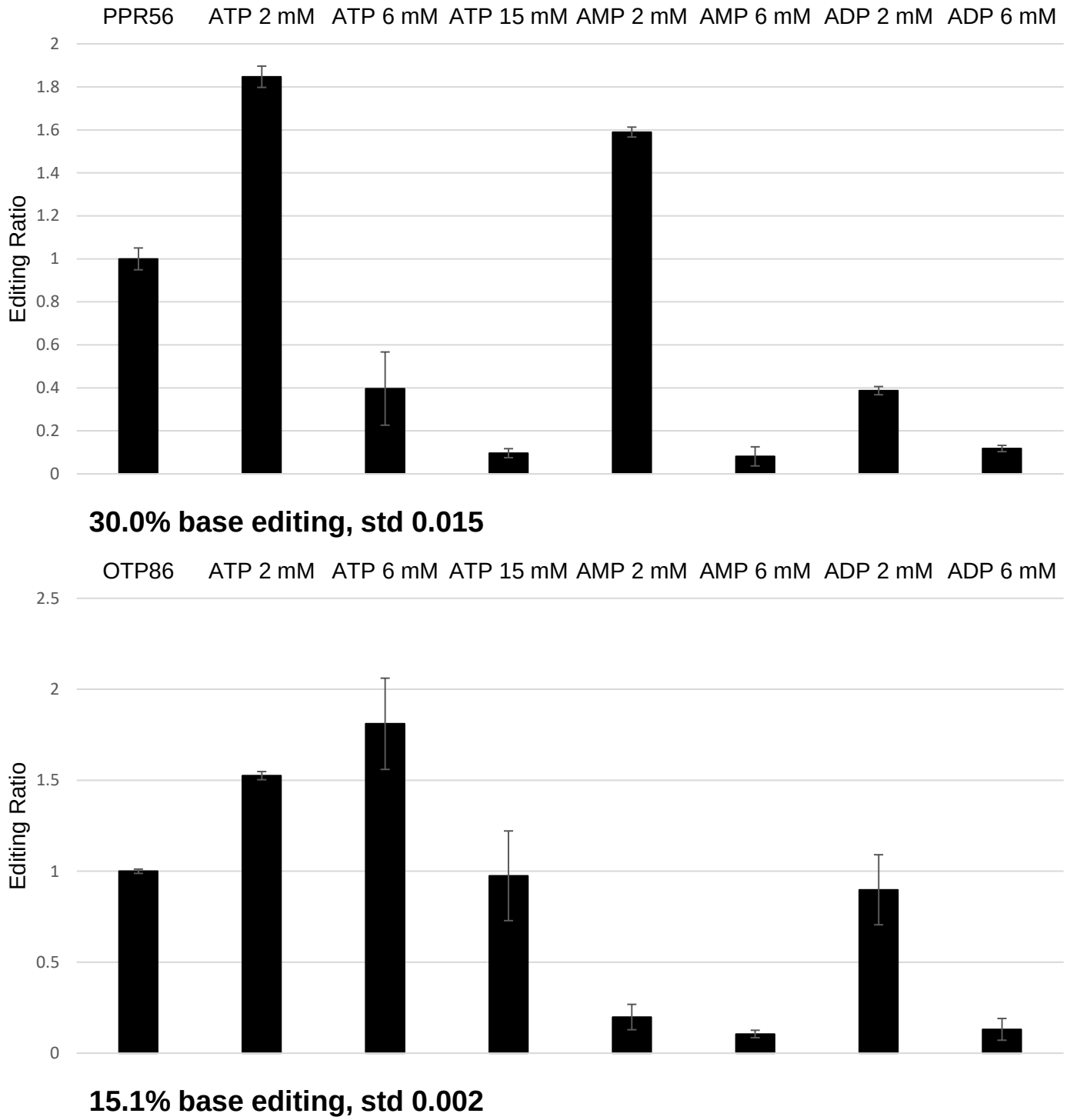


Figure 14:ATP's effects are protein-dependent

PPR56 and PPR56^{PPRE1E2}-OTP86^{DYW} (OTP86) mediated in vitro editing were performed in the presence of ATP, AMP, and ADP at multiple concentrations. ATP at 2 mM stimulated the editing reaction in both proteins but suppressed editing in PPR56 as the concentration was increased. OTP86 did not experience the same decreases however. AMP and ADP minorly affected editing at 2 mM but all reduced editing at 6 mM.

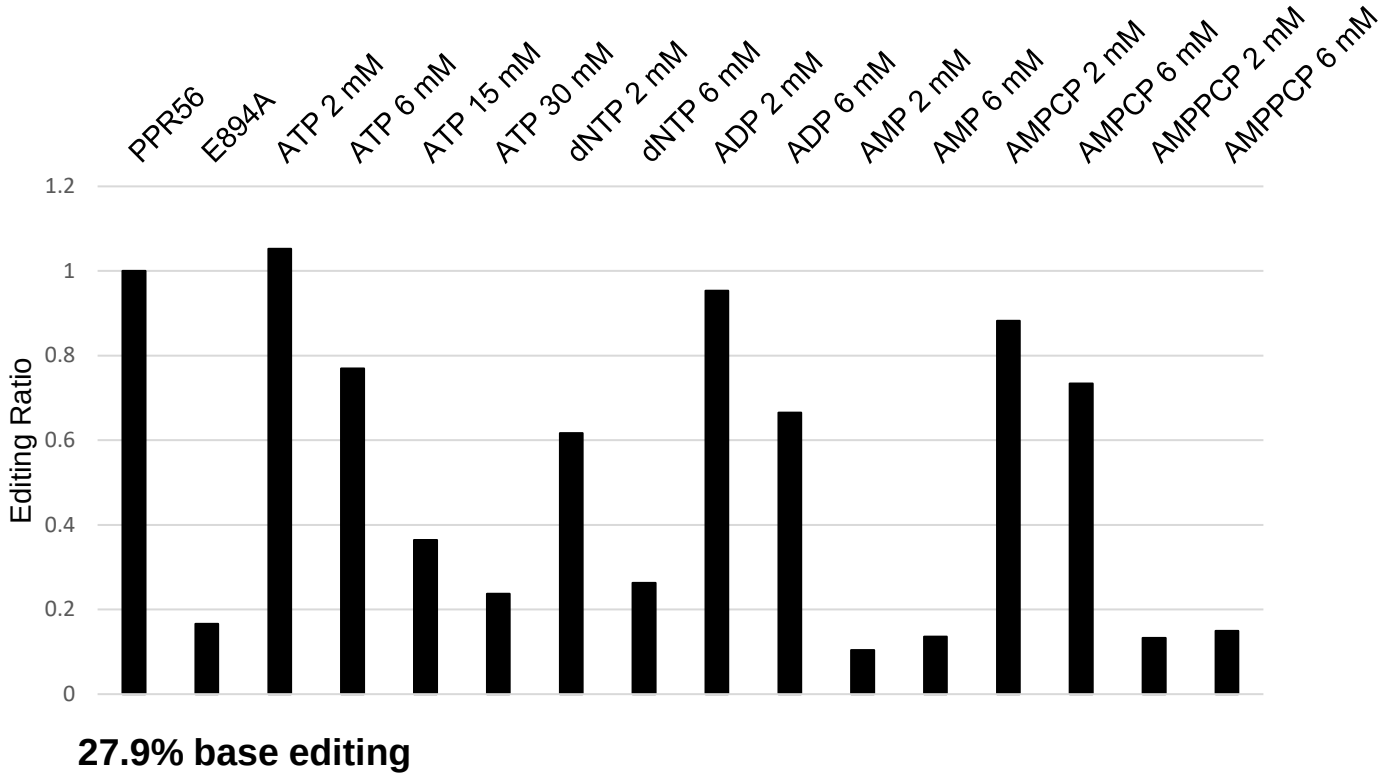
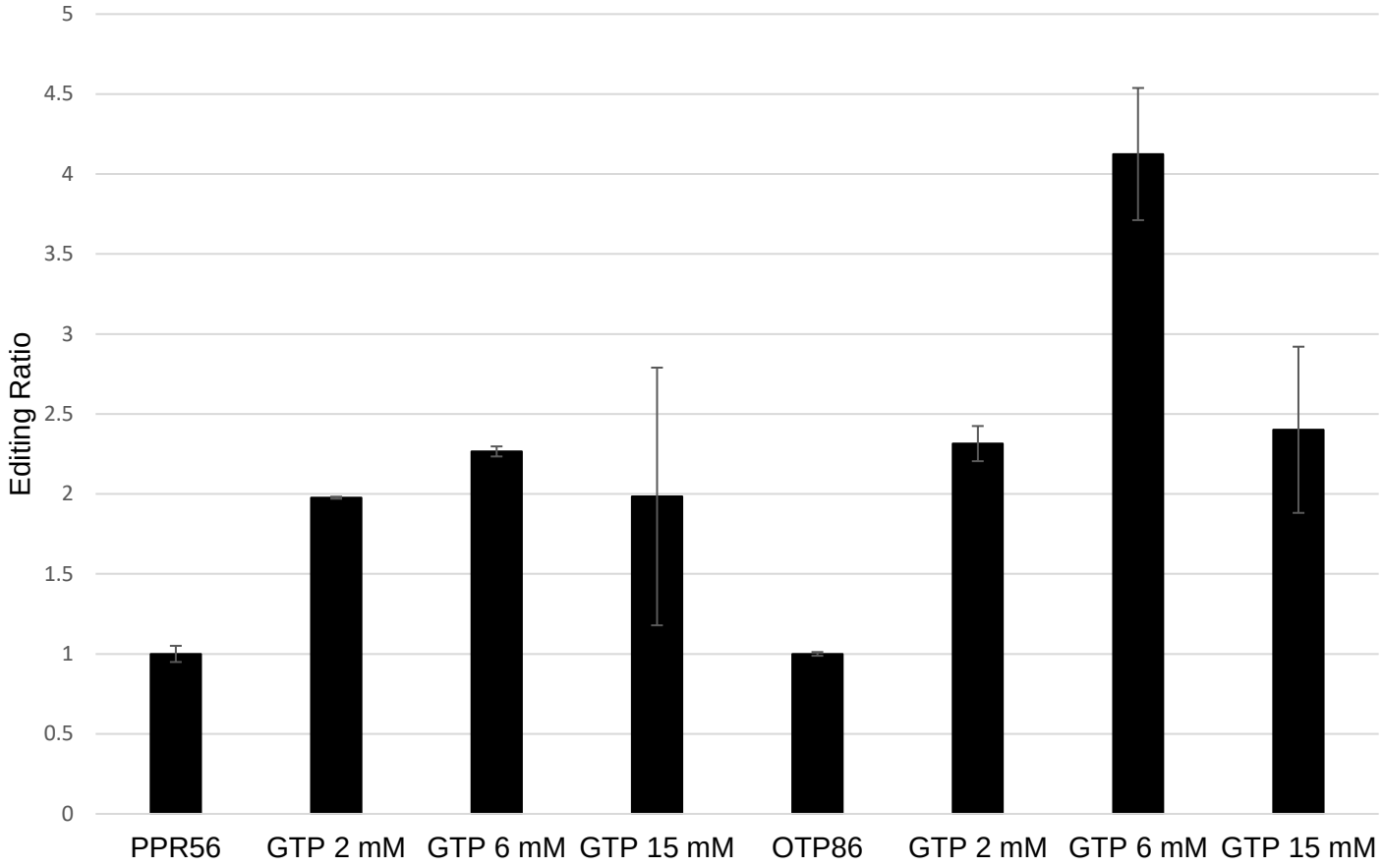


Figure 15: ATP analogs show little positive effect

PPR56 mediated in vitro editing was performed in the presence of ATP, dNTP mixes, ADP, AMP, AMP-CP (ADP analog), and AMP-PCP (ATP analog). ATP only slightly increased editing at 2 mM but had negative effects at higher concentrations. dNTP mixes, ADP and AMP-CP were similar but had no positive effects at any concentration. AMP and AMP-PCP drastically reduced editing even at low concentrations.

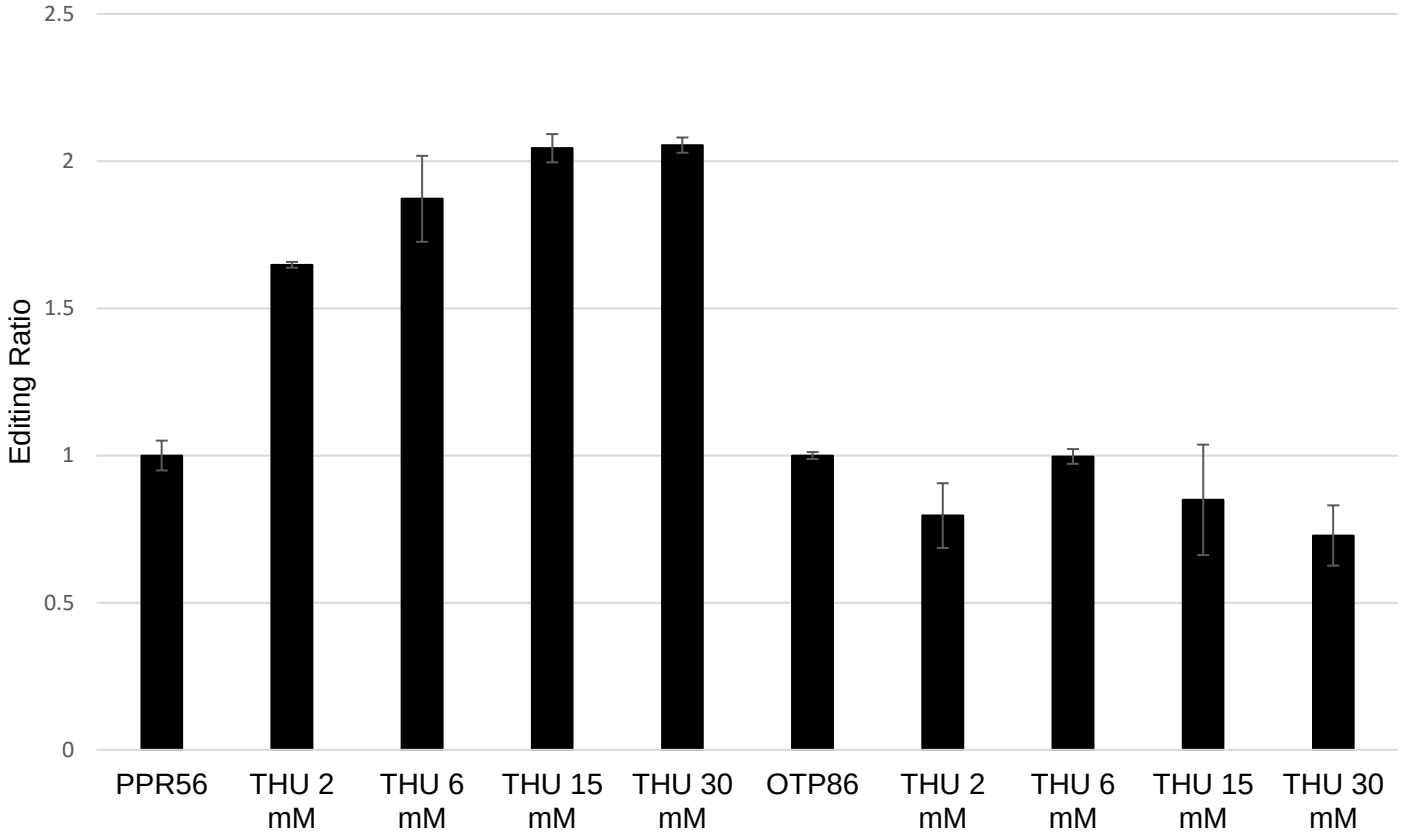


PPR56 - 30.0% base editing, std 0.015

OTP86 - 15.1% base editing, std 0.002

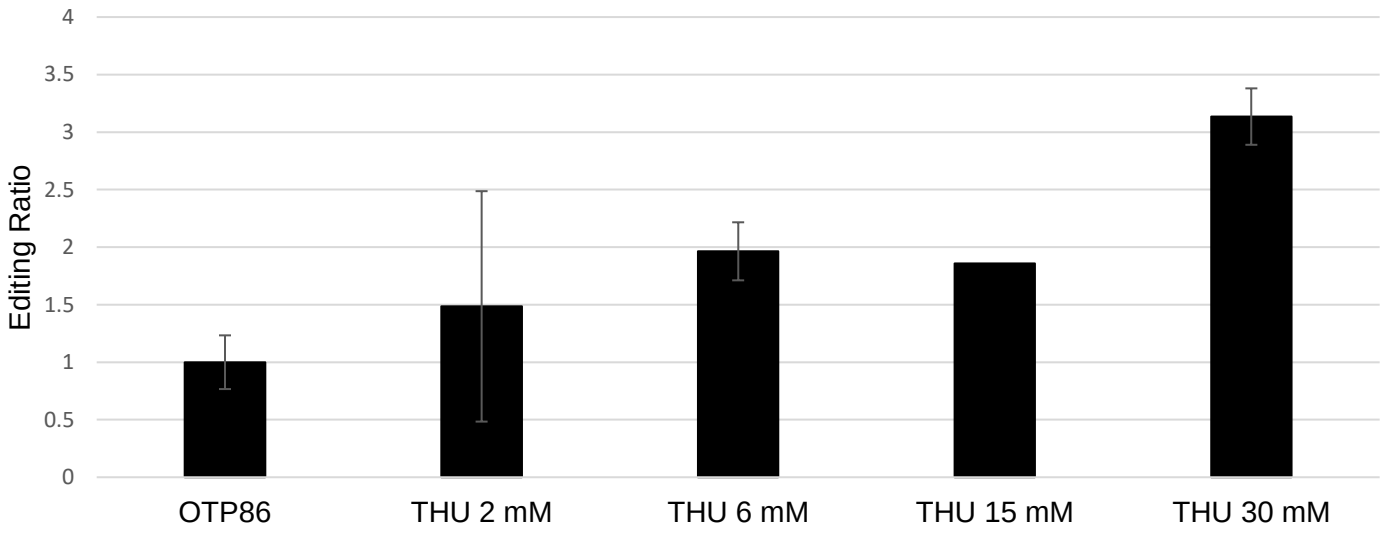
Figure 16: GTP greatly increases in vitro RNA editing efficiencies

PPR56 and PPR56^{PPRE1E2}-OTP86^{DYW} (OTP86) were exposed to varying levels of GTP during the in vitro reaction. Large increases in editing efficiencies were found, peaking at 6 mM. Editing percent was over doubled in PPR56 and quadrupled in OTP86. Higher concentrations (15 mM) are still positive but lowered comparatively.



PPR56 - 30.0% base editing, std 0.015

OTP86 - 15.1% base editing, std 0.002



OTP86 – 11.6% base editing, std 0.027

Figure 17: Tetrahydrouridine is a selective enhancer of RNA editing
A known cytidine deaminase inhibitor, tetrahydrouridine (THU), was added to in vitro reactions of PPR56 and PPR56^{PPRE1E2}-OTP86^{DYW} (OTP86). THU greatly increased editing efficiency in PPR56 with continued increases as concentration increased. THU was not as stable in OTP86 with some trials displaying no change in efficiency while some showed similar increases. THU may preferentially affect certain PPR proteins more

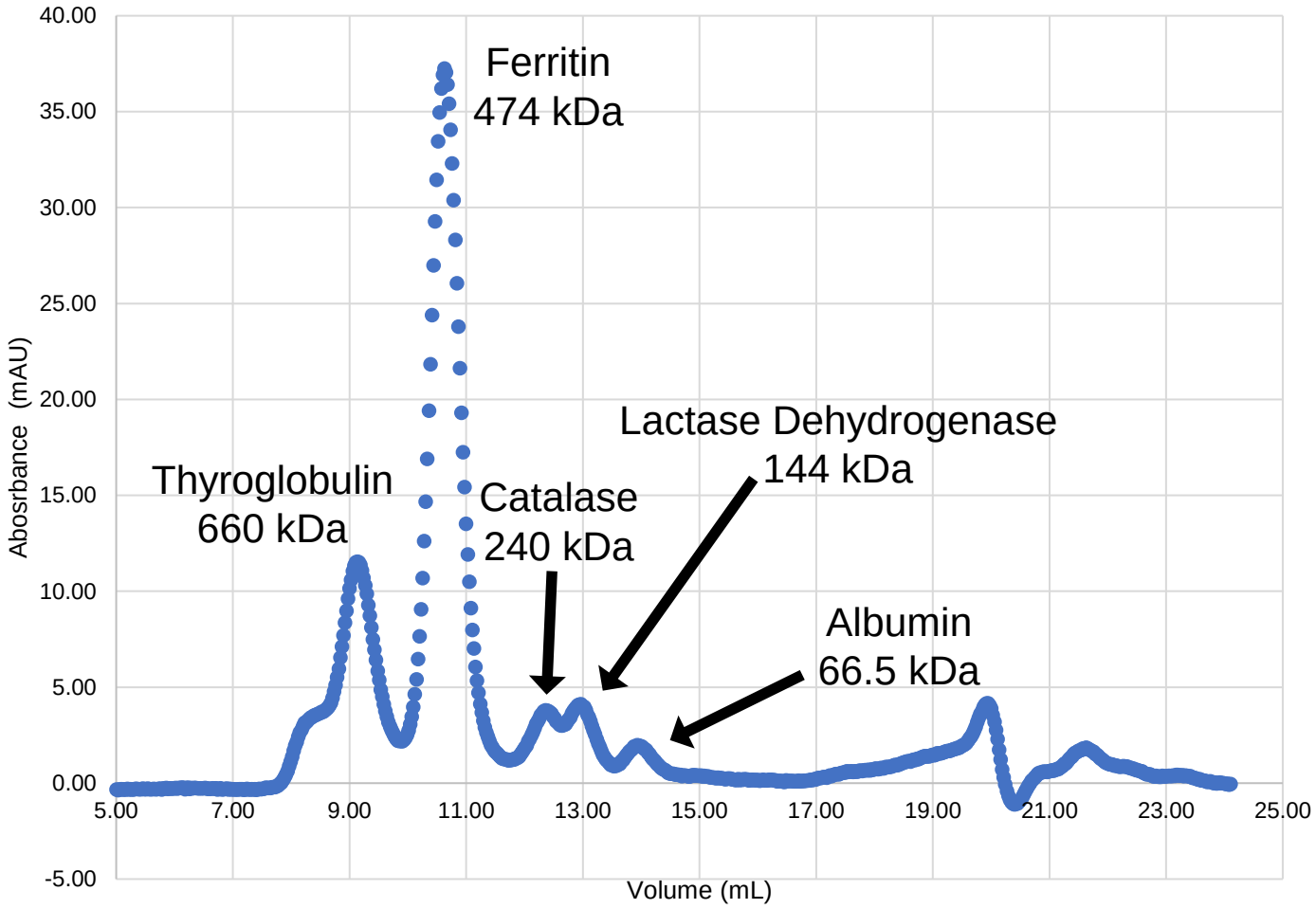


Figure 18: Size exclusion chromatography of Amersham HMW Marker

In order for the relative masses of each peak to be estimated, a protein marker was run under the same conditions as other size exclusion experiments. The marker contains a mixture of five proteins corresponding to the listed masses in kDa. Proteins that peak at those respective volumes should roughly have the same mass.

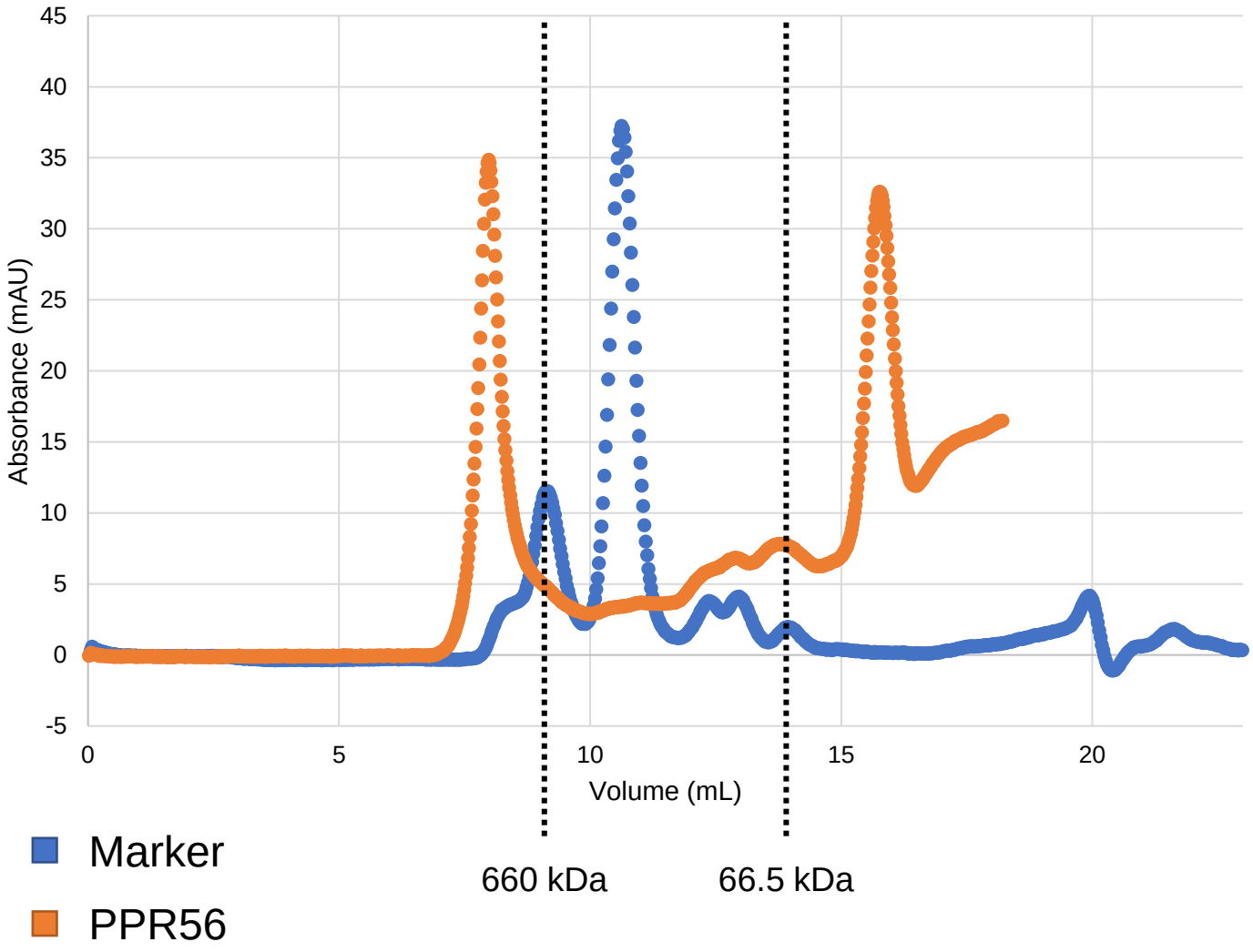


Figure 19: Initial size exclusion chromatography of PPR56
500 µL of purified PPR56 in 24 mL of lysis buffer was subjected to size exclusion chromatography at 0.5 mL/min. As proteins flow through the column, they are separated by size and these are at different times represented by the total volume eluted on the x-axis. Protein abundance is measured by absorbance at 280 nm in milli-absorbance units. PPR56 segregated into two peaks around 8 and 15.7 mL.

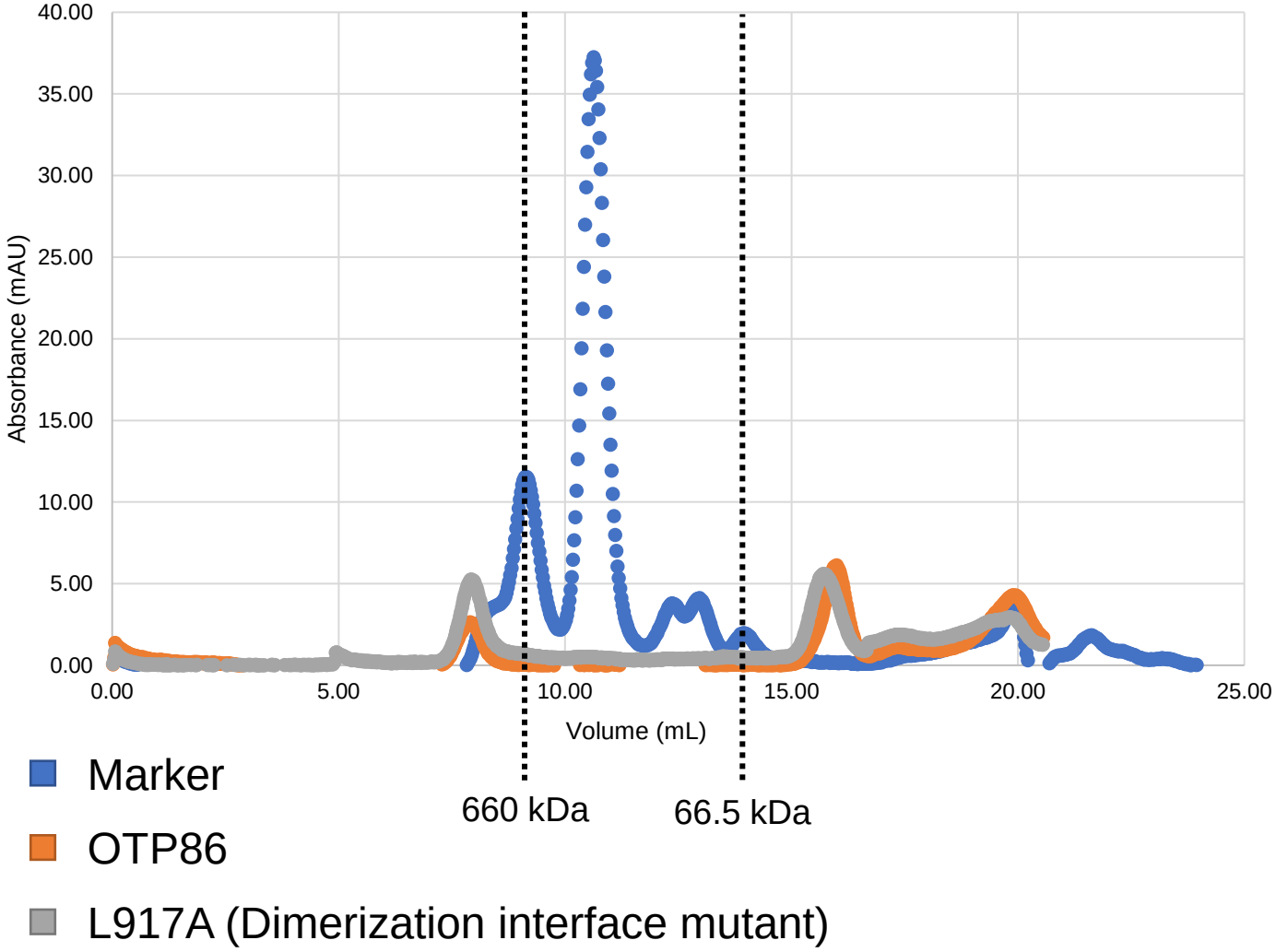


Figure 20: PPR56^{PPRE1E2}-OTP86^{DYW} and similar proteins oligomerize and lack monomeric fraction

PPR56^{PPRE1E2}-OTP86^{DYW} (OTP86) and L917A were analyzed next using size exclusion chromatography. L917A contains a mutation in the dimerization interface which should prevent the accumulation of oligomers. Samples separated into the same peaks as PPR56 with two peaks at 8 and 16 mL, representing a large oligomer and fragments of maltose binding protein, respectively. No peak close to the expected 120 kDa for the monomer was found.

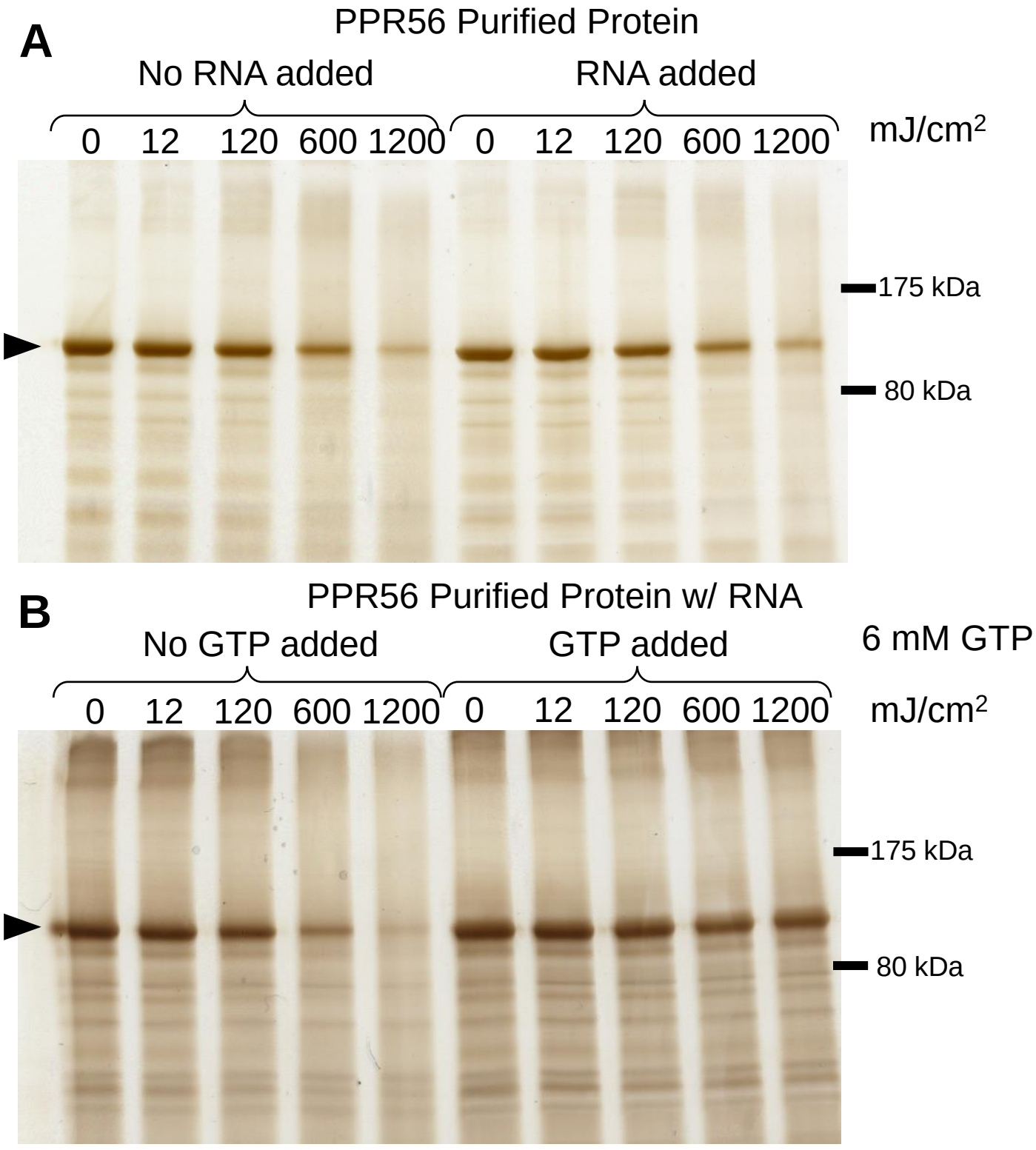
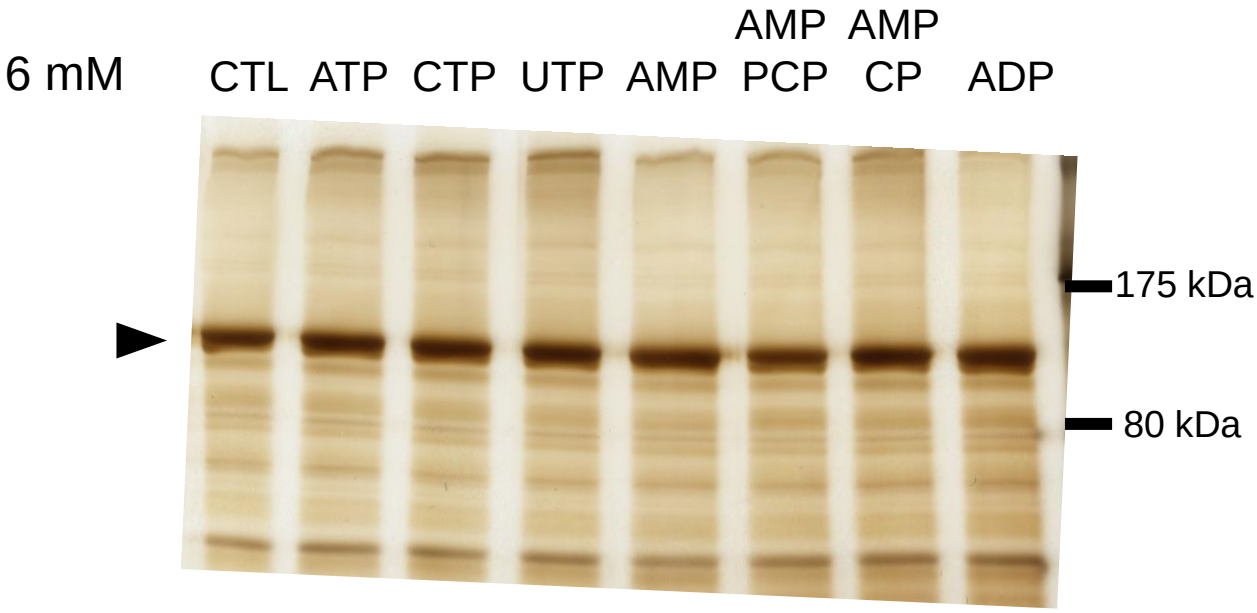


Figure 21: GTP stabilizes the PPR56 monomer when exposed to crosslinking UV radiation

PPR56 purified protein, with or without its corresponding substrate RNA, was incubated for 30 minutes and then exposed to 254 nm UV light for the listed exposures ranging from zero to 1200 mJ/cm^2. Crosslinked proteins were then run on SDS-PAGE gels and stained with silver stain. (A) PPR56 monomer band at 120 kDa slowly disappeared as exposure was increased. Addition of target RNA showed displayed no change. (B) When 6 mM GTP was added during incubation, monomer bands were still visible even at the highest amount of irradiation.

PPR56 Protein, No RNA Added
No Crosslink



PPR56 Protein, No RNA Added
1200 mJ Crosslink

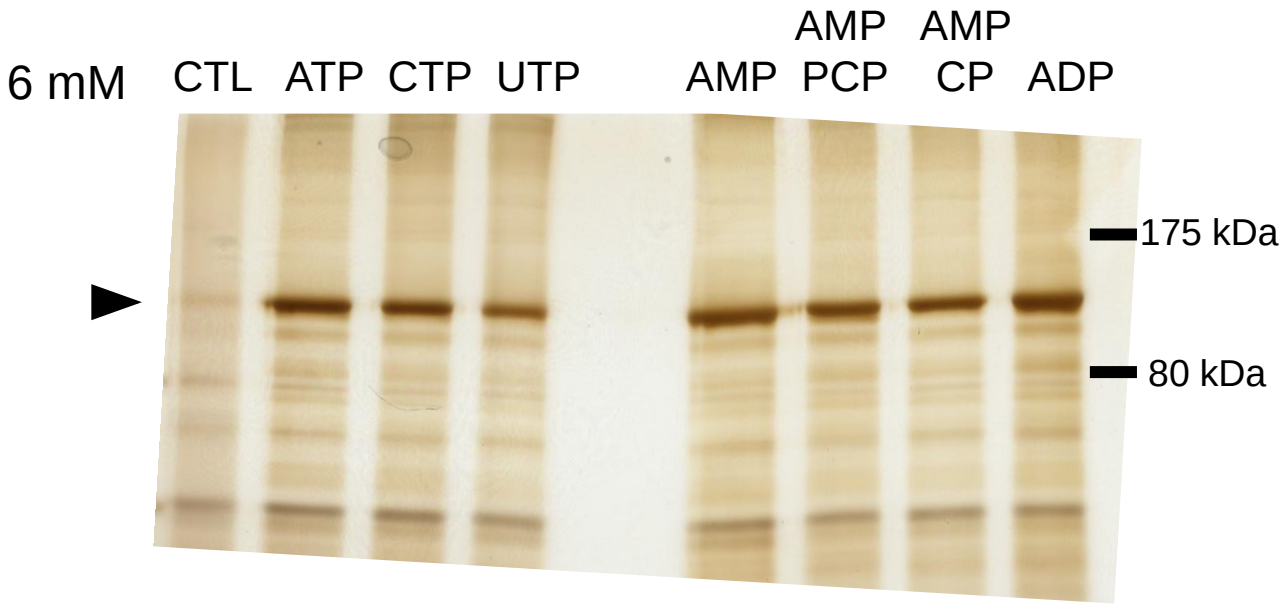


Figure 22: All NTPs are capable of stabilizing PPR56 monomer under UV crosslinking exposure

PPR56 protein was exposed to UV light for a set 1200 mJ. Before irradiation, the protein was incubated with select NTPs for 30 minutes. Silver stained gels with no crosslink showed the normal monomeric band at 120 kDa. When crosslinked, this disappeared similar to previous experiments, but the addition of 6 mM of any NTP as well as analogs were sufficient to protect the protein from crosslinking.

Purified PPR56 protein digested with:

Trypsin

Protein sequence coverage: 80%

Matched peptides shown in ***bold red***.

```
   1 MKHHHHHHPM KIEEGKLVIW INGDKGYNGL AEVGKKFEKD TGIKVTVEHP
  51 DKLEEKFPQV AATGDGPDII FWAHDRFGGY AQSGLLAEIT PDKAFQDKLY
 101 PFTWDAVRYN GKLIAYPIAV EALSLIYNKD LLPNPPKTWE EIPALDKELK
 151 AKGKSALMFN LQEPYFTWPL IAADGGYAFK YENGKYDIKD VGVDNAGAKA
 201 GLTFLVDLIK NKHMNADTDY SIAEAAFNKG ETAMTINGPW AWSNIDTSKV
 251 NYGVTVLPTF KGQPSKPFVG VLSAGINAAS PNKELAKEFL ENYLLTDEGL
 301 EAVNKDKPLG AVALKSYEEE LAKDPRIAAT MENAQKGEIM PNIPQMSAFW
 351 YAVRTAVINA ASGRQTVDEA LKDAQTPGSG TMTSLYKKAG LENLYFQGVV
 401 LMNRLQRGLI TDSFMYVEVL KRCLKQKDLM AAKQVHDCII KSRMEQNAHV
 451 MNNLLHVYIE CGRLQEARCV FDALVKKSGA SWNAMIAGYV EHKHAEDAMR
 501 LFREMCHEGV QPNAGTYMII LKACASLSAL KWGKEVHACI RHGGLESDVR
 551 VGTALLRMYG KCGSINEARR IFDNLMNHDI ISWTVMIGAY AQSGNGKEAY
 601 RLMLQMEQEG FKPNAITYVS ILNACASEGA LKWVKRVHRH ALDAGLELDV
 651 RVGTALVQMY AKSGSIDDAR VVFDRMKVRD VVSWNVMIGA FAEHGRGHEA
 701 YDLFLQMQTE GCKPDAIMFL SILNACASAG ALEWVKKIHR HALDSGLEVD
 751 VRVGTALVHM YSKSGSIDDA RVVFDRMKVR NVVSWNAMIS GLAQHGLGQD
 801 ALEVFRRMTA HGVKPDRVTF VAVLSACSHA GLVDEGRSQY LAMTQVYGIE
 851 PDVSHCNCMV DLLGRAGRLM EAKLFIDNMA VDPDEATWGA LLGSCRTYGN
 901 VELGELVAKE RLKLDPKNAA TYVLLSNIYA EAGKWDMVSW VRTMMRERGI
 951 RKEPGRSWIE VDNKIHDFLV ADSSHPECKE INESKDKVIE KIKAEGYIPD
1001 TRLVLKNKNM KDKELDICSH SEKLAIVYGL MHTPPGNPIR VFKNLRVCTD
1051 CHGATKLISK VEGREIIVRD ANRFHHFKDG VCSCGDYW
```

Chymotrypsin

Protein sequence coverage: 92%

Matched peptides shown in ***bold red***.

```
   1 MKHHHHHHPM KIEEGKLVIW INGDKGYNGL AEVGKKFEKD TGIKVTVEHP
  51 DKLEEKFPQV AATGDGPDII FWAHDRFGGY AQSGLLAEIT PDKAFQDKLY
 101 PFTWDAVRYN GKLIAYPIAV EALSLIYNKD LLPNPPKTWE EIPALDKELK
 151 AKGKSALMFN LQEPYFTWPL IAADGGYAFK YENGKYDIKD VGVDNAGAKA
 201 GLTFLVDLIK NKHMNADTDY SIAEAAFNKG ETAMTINGPW AWSNIDTSKV
 251 NYGVTVLPTF KGQPSKPFVG VLSAGINAAS PNKELAKEFL ENYLLTDEGL
 301 EAVNKDKPLG AVALKSYEEE LAKDPRIAAT MENAQKGEIM PNIPQMSAFW
 351 YAVRTAVINA ASGRQTVDEA LKDAQTPGSG TMTSLYKKAG LENLYFQGVV
 401 LMNRLQRGLI TDSFMYVEVL KRCLKQKDLM AAKQVHDCII KSRMEQNAHV
 451 MNNLLHVYIE CGRLQEARCV FDALVKKSGA SWNAMIAGYV EHKHAEDAMR
 501 LFREMCHEGV QPNAGTYMII LKACASLSAL KWGKEVHACI RHGGLESDVR
 551 VGTALLRMYG KCGSINEARR IFDNLMNHDI ISWTVMIGAY AQSGNGKEAY
 601 RLMLQMEQEG FKPNAITYVS ILNACASEGA LKWVKRVHRH ALDAGLELDV
 651 RVGTALVQMY AKSGSIDDAR VVFDRMKVRD VVSWNVMIGA FAEHGRGHEA
 701 YDLFLQMQTE GCKPDAIMFL SILNACASAG ALEWVKKIHR HALDSGLEVD
 751 VRVGTALVHM YSKSGSIDDA RVVFDRMKVR NVVSWNAMIS GLAQHGLGQD
 801 ALEVFRRMTA HGVKPDRVTF VAVLSACSHA GLVDEGRSQY LAMTQVYGIE
 851 PDVSHCNCMV DLLGRAGRLM EAKLFIDNMA VDPDEATWGA LLGSCRTYGN
 901 VELGELVAKE RLKLDPKNAA TYVLLSNIYA EAGKWDMVSW VRTMMRERGI
 951 RKEPGRSWIE VDNKIHDFLV ADSSHPECKE INESKDKVIE KIKAEGYIPD
1001 TRLVLKNKNM KDKELDICSH SEKLAIVYGL MHTPPGNPIR VFKNLRVCTD
1051 CHGATKLISK VEGREIIVRD ANRFHHFKDG VCSCGDYW
```

Figure 23: Mass spectrometry of trypsin and chymotrypsin-digested PPR56

In preparation for crosslink experiments, PPR56 protein was analyzed using mass spectrometry with two different proteases, trypsin and chymotrypsin. The observed coverage after mass spectrometry for each protease is listed and detected amino acids are in red with undetected amino acids in black. Chymotrypsin cleaves peptides more readily and also displayed higher sequence coverage at 92%.

GTP Trypsin

Protein sequence coverage: 80%

Matched peptides shown in ***bold red***.

```
   1 MKHHHHHHPM KIEEGKLVIW INGDKGYNGL AEVGKKFEKD TGIKVTVEHP
  51 DKLEEKFPQV AATGDGPDII FWAHDRFGGY AQSGLLAEIT PDKAFQDKLY
 101 PFTWDAVRYN GKLIAYPIAV EALSLIYNKD LLPNPPKTWE EIPALDKELK
 151 AKGKSALMFN LQEPYFTWPL IAADGGYAFK YENGKYDIKD VGVDNAGAKA
 201 GLTFLVDLIK NKHMNADTDY SIAEAAFNKG ETAMTINGPW AWSNIDTSKV
 251 NYGVTVLPTF KGQPSKPFVG VLSAGINAAS PNKELAKEFL ENYLLTDEGL
 301 EAVNKDKPLG AVALKSYEEE LAKDPRIAAT MENAQKGEIM PNIPQMSAFW
 351 YAVRTAVINA ASGRQTVDEA LKDAQTPGSG TMTSLYKKAG LENLYFQGVV
 401 LMNRLQRGLI TDSFMYVEVL KRCLKQKDLM AAKQVHDCII KSRMEQNAHV
 451 MNNLLHVYIE CGRLQEARCV FDALVKKSGA SWNAMIAGYV EHKHAEDAMR
 501 LFREMCHEGV QPNAGTYMII LKACASLSAL KWGKEVHACI RHGGLESDVR
 551 VGTALLRMYG KCGSINEARR IFDNLMNHDI ISWTVMIGAY AQSGNGKEAY
 601 RLMLQMEQEG FKPNAITYVS ILNACASEGA LKWVKRVHRH ALDAGLELDV
 651 RVGTALVQMY AKSGSIDDAR VVFDRMKVRD VVSWNVMIGA FAEHGRGHEA
 701 YDLFLQMQTE GCKPDAIMFL SILNACASAG ALEWVKKIHR HALDSGLEVD
 751 VRVGTALVHM YSKSGSIDDA RVVFDRMKVR NVVSWNAMIS GLAQHGLGQD
 801 ALEVFRRMTA HGVKPDRVTF VAVLSACSHA GLVDEGRSQY LAMTQVYGIE
 851 PDVSHCNCMV DLLGRAGRLM EAKLFIDNMA VDPDEATWGA LLGSCRTYGN
 901 VELGELVAKE RLKLDPKNAA TYVLLSNIYA EAGKWDMVSW VRTMMRERGI
 951 RKEPGRSWIE VDNKIHDFLV ADSSHPECKE INESKDKVIE KIKAEGYIPD
1001 TRLVLKNKNM KDKELDICSH SEKLAIVYGL MHTPPGNPIR VFKNLRVCTD
1051 CHGATKLISK VEGREIIVRD ANRFHHFKDG VCSCGDYW
```

THU Trypsin

Protein sequence coverage: 75%

Matched peptides shown in ***bold red***.

```
   1 MKHHHHHHPM KIEEGKLVIW INGDKGYNGL AEVGKKFEKD TGIKVTVEHP
  51 DKLEEKFPQV AATGDGPDII FWAHDRFGGY AQSGLLAEIT PDKAFQDKLY
 101 PFTWDAVRYN GKLIAYPIAV EALSLIYNKD LLPNPPKTWE EIPALDKELK
 151 AKGKSALMFN LQEPYFTWPL IAADGGYAFK YENGKYDIKD VGVDNAGAKA
 201 GLTFLVDLIK NKHMNADTDY SIAEAAFNKG ETAMTINGPW AWSNIDTSKV
 251 NYGVTVLPTF KGQPSKPFVG VLSAGINAAS PNKELAKEFL ENYLLTDEGL
 301 EAVNKDKPLG AVALKSYEEE LAKDPRIAAT MENAQKGEIM PNIPQMSAFW
 351 YAVRTAVINA ASGRQTVDEA LKDAQTPGSG TMTSLYKKAG LENLYFQGVV
 401 LMNRLQRGLI TDSFMYVEVL KRCLKQKDLM AAKQVHDCII KSRMEQNAHV
 451 MNNLLHVYIE CGRLQEARCV FDALVKKSGA SWNAMIAGYV EHKHAEDAMR
 501 LFREMCHEGV QPNAGTYMII LKACASLSAL KWGKEVHACI RHGGLESDVR
 551 VGTALLRMYG KCGSINEARR IFDNLMNHDI ISWTVMIGAY AQSGNGKEAY
 601 RLMLQMEQEG FKPNAITYVS ILNACASEGA LKWVKRVHRH ALDAGLELDV
 651 RVGTALVQMY AKSGSIDDAR VVFDRMKVRD VVSWNVMIGA FAEHGRGHEA
 701 YDLFLQMQTE GCKPDAIMFL SILNACASAG ALEWVKKIHR HALDSGLEVD
 751 VRVGTALVHM YSKSGSIDDA RVVFDRMKVR NVVSWNAMIS GLAQHGLGQD
 801 ALEVFRRMTA HGVKPDRVTF VAVLSACSHA GLVDEGRSQY LAMTQVYGIE
 851 PDVSHCNCMV DLLGRAGRLM EAKLFIDNMA VDPDEATWGA LLGSCRTYGN
 901 VELGELVAKE RLKLDPKNAA TYVLLSNIYA EAGKWDMVSW VRTMMRERGI
 951 RKEPGRSWIE VDNKIHDFLV ADSSHPECKE INESKDKVIE KIKAEGYIPD
1001 TRLVLKNKNM KDKELDICSH SEKLAIVYGL MHTPPGNPIR VFKNLRVCTD
1051 CHGATKLISK VEGREIIVRD ANRFHHFKDG VCSCGDYW
```

GTP Chymotrypsin

Protein sequence coverage: 89%

Matched peptides shown in ***bold red***.

```
   1 MKHHHHHHPM KIEEGKLVIW INGDKGYNGL AEVGKKFEKD TGIKVTVEHP
  51 DKLEEKFPQV AATGDGPDII FWAHDRFGGY AQSGLLAEIT PDKAFQDKLY
 101 PFTWDAVRYN GKLIAYPIAV EALSLIYNKD LLPNPPKTWE EIPALDKELK
 151 AKGKSALMFN LQEPYFTWPL IAADGGYAFK YENGKYDIKD VGVDNAGAKA
 201 GLTFLVDLIK NKHMNADTDY SIAEAAFNKG ETAMTINGPW AWSNIDTSKV
 251 NYGVTVLPTF KGQPSKPFVG VLSAGINAAS PNKELAKEFL ENYLLTDEGL
 301 EAVNKDKPLG AVALKSYEEE LAKDPRIAAT MENAQKGEIM PNIPQMSAFW
 351 YAVRTAVINA ASGRQTVDEA LKDAQTPGSG TMTSLYKKAG LENLYFQGVV
 401 LMNRLQRGLI TDSFMYVEVL KRCLKQKDLM AAKQVHDCII KSRMEQNAHV
 451 MNNLLHVYIE CGRLQEARCV FDALVKKSGA SWNAMIAGYV EHKHAEDAMR
 501 LFREMCHEGV QPNAGTYMII LKACASLSAL KWGKEVHACI RHGGLESDVR
 551 VGTALLRMYG KCGSINEARR IFDNLMNHDI ISWTVMIGAY AQSGNGKEAY
 601 RLMLQMEQEG FKPNAITYVS ILNACASEGA LKWVKRVHRH ALDAGLELDV
 651 RVGTALVQMY AKSGSIDDAR VVFDRMKVRD VVSWNVMIGA FAEHGRGHEA
 701 YDLFLQMQTE GCKPDAIMFL SILNACASAG ALEWVKKIHR HALDSGLEVD
 751 VRVGTALVHM YSKSGSIDDA RVVFDRMKVR NVVSWNAMIS GLAQHGLGQD
 801 ALEVFRRMTA HGVKPDRVTF VAVLSACSHA GLVDEGRSQY LAMTQVYGIE
 851 PDVSHCNCMV DLLGRAGRLM EAKLFIDNMA VDPDEATWGA LLGSCRTYGN
 901 VELGELVAKE RLKLDPKNAA TYVLLSNIYA EAGKWDMVSW VRTMMRERGI
 951 RKEPGRSWIE VDNKIHDFLV ADSSHPECKE INESKDKVIE KIKAEGYIPD
1001 TRLVLKNKNM KDKELDICSH SEKLAIVYGL MHTPPGNPIR VFKNLRVCTD
1051 CHGATKLISK VEGREIIVRD ANRFHHFKDG VCSCGDYW
```

THU Chymotrypsin

Protein sequence coverage: 90%

Matched peptides shown in ***bold red***.

```
   1 MKHHHHHHPM KIEEGKLVIW INGDKGYNGL AEVGKKFEKD TGIKVTVEHP
  51 DKLEEKFPQV AATGDGPDII FWAHDRFGGY AQSGLLAEIT PDKAFQDKLY
 101 PFTWDAVRYN GKLIAYPIAV EALSLIYNKD LLPNPPKTWE EIPALDKELK
 151 AKGKSALMFN LQEPYFTWPL IAADGGYAFK YENGKYDIKD VGVDNAGAKA
 201 GLTFLVDLIK NKHMNADTDY SIAEAAFNKG ETAMTINGPW AWSNIDTSKV
 251 NYGVTVLPTF KGQPSKPFVG VLSAGINAAS PNKELAKEFL ENYLLTDEGL
 301 EAVNKDKPLG AVALKSYEEE LAKDPRIAAT MENAQKGEIM PNIPQMSAFW
 351 YAVRTAVINA ASGRQTVDEA LKDAQTPGSG TMTSLYKKAG LENLYFQGVV
 401 LMNRLQRGLI TDSFMYVEVL KRCLKQKDLM AAKQVHDCII KSRMEQNAHV
 451 MNNLLHVYIE CGRLQEARCV FDALVKKSGA SWNAMIAGYV EHKHAEDAMR
 501 LFREMCHEGV QPNAGTYMII LKACASLSAL KWGKEVHACI RHGGLESDVR
 551 VGTALLRMYG KCGSINEARR IFDNLMNHDI ISWTVMIGAY AQSGNGKEAY
 601 RLMLQMEQEG FKPNAITYVS ILNACASEGA LKWVKRVHRH ALDAGLELDV
 651 RVGTALVQMY AKSGSIDDAR VVFDRMKVRD VVSWNVMIGA FAEHGRGHEA
 701 YDLFLQMQTE GCKPDAIMFL SILNACASAG ALEWVKKIHR HALDSGLEVD
 751 VRVGTALVHM YSKSGSIDDA RVVFDRMKVR NVVSWNAMIS GLAQHGLGQD
 801 ALEVFRRMTA HGVKPDRVTF VAVLSACSHA GLVDEGRSQY LAMTQVYGIE
 851 PDVSHCNCMV DLLGRAGRLM EAKLFIDNMA VDPDEATWGA LLGSCRTYGN
 901 VELGELVAKE RLKLDPKNAA TYVLLSNIYA EAGKWDMVSW VRTMMRERGI
 951 RKEPGRSWIE VDNKIHDFLV ADSSHPECKE INESKDKVIE KIKAEGYIPD
1001 TRLVLKNKNM KDKELDICSH SEKLAIVYGL MHTPPGNPIR VFKNLRVCTD
1051 CHGATKLISK VEGREIIVRD ANRFHHFKDG VCSCGDYW
```

Figure 24: Addition of GTP and THU does not change sequence coverage

PPR56 protein was incubated in 6 mM GTP and THU before UV crosslink and MS preparation. Proteins from both trials were then digested with trypsin and chymotrypsin and subjected to mass spectrometry. Sequence coverage was largely unchanged from controls with chymotrypsin again offering higher sequence coverage in general.

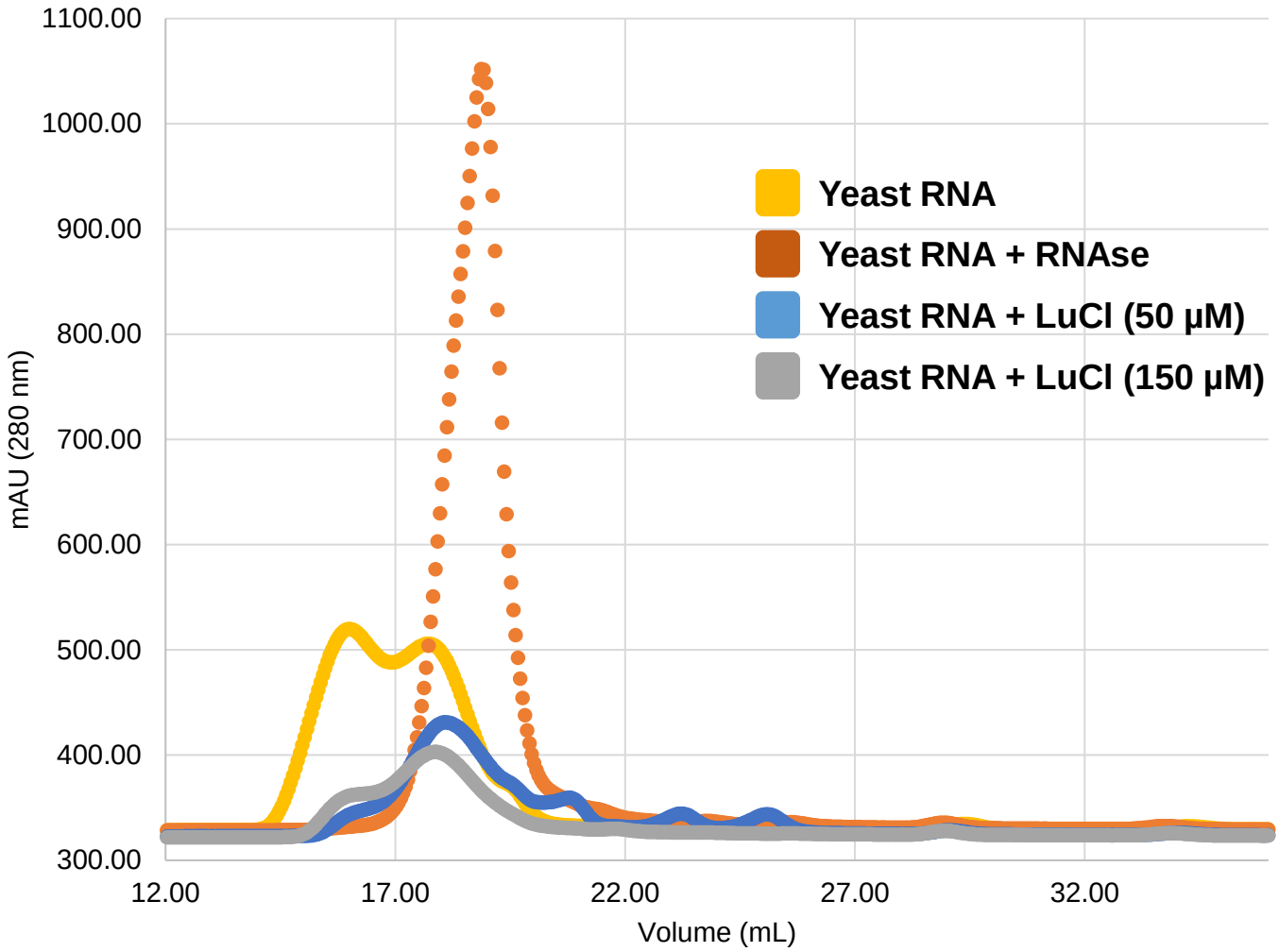


Figure 25: Lutetium (III) chloride completely digests control yeast RNA

Size exclusion chromatography of standard yeast total RNA (yellow) was observed along with yeast RNA digested by both RNAse (red), 50 µM LuCl (Blue), and 150 µM LuCl (gray) overnight at 4°C. All three reagents could degrade yeast RNA into nucleotides.

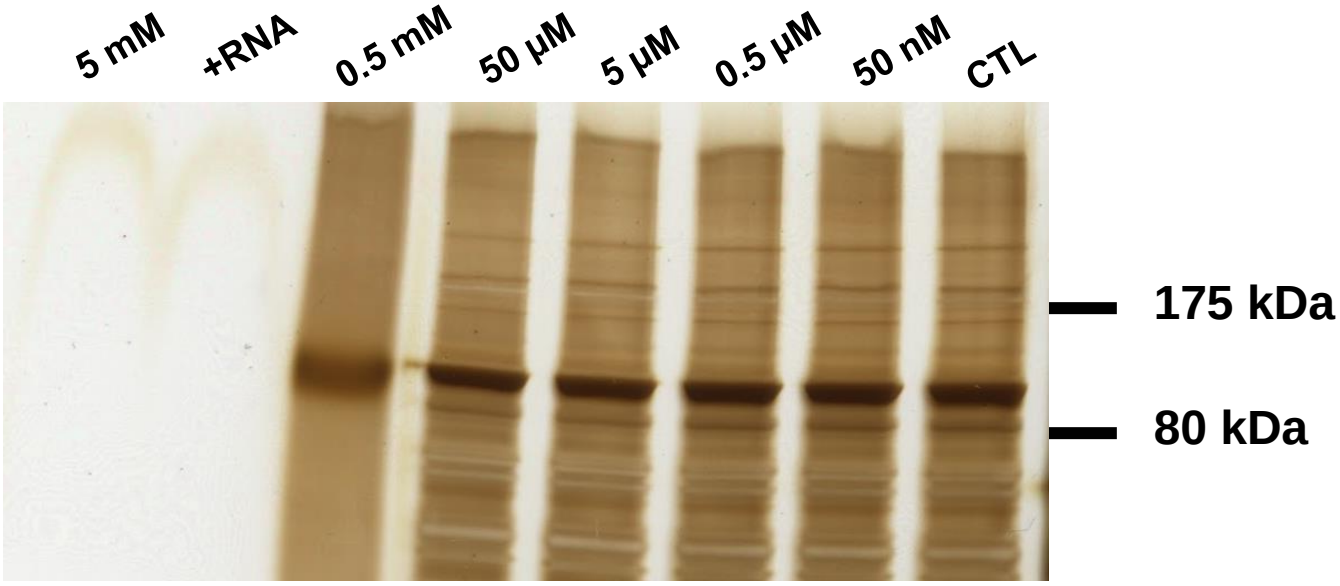


Figure 26: Lutetium (III) chloride does not degrade PPR56 recombinant protein at 50 µM and lower concentrations

Silver staining of SDS gels run with 1 µg of PPR56 recombinant protein. Concentrations indicate the concentration of LuCl used in overnight 4°C digestions. 120 ng of RNA was added to one trial at 5 mM LuCl. Recombinant protein is degraded until LuCl is diluted to 50 µM.

Protein sequence coverage: 66%

Matched peptides shown in ***bold red***.

```
   1 MKHHHHHHPM KIEEGKLVIW INGDKGYNGL AEVGKKFEKD TGIKVTVEHP
  51 DKLEEKFPQV AATGDGPDII FWAHDRFGGY AQSGLLAEIT PDKAFQDKLY
 101 PFTWDAVRYN GKLIAYPIAV EALSLIYNKD LLPNPPKTWE EIPALDKELK
 151 AKGKSALMFN LQEPYFTWPL IAADGGYAFK YENGKYDIKD VGVDNAGAKA
 201 GLTFLVDLIK NKHMNADTDY SIAEAAFNKG ETAMTINGPW AWSNIDTSKV
 251 NYGVTVLPTF KGQPSKPFVG VLSAGINAAS PNKELAKEFL ENYLLTDEGL
 301 EAVNKDKPLG AVALKSYEEE LAKDPRIAAT MENAQKGEIM PNIPQMSAFW
 351 YAVRTAVINA ASGRQTVDEA LKDAQTPGSG TMTSLYKKAG LENLYFQGVV
 401 LMNRLQRGLI TDSFMYVEVL KRCLKQKDLM AAKQVHDCII KSRMEQNAHV
 451 MNNLLHVYIE CGRLQEARCV FDALVKKSGA SWNAMIAGYV EHKHAEDAMR
 501 LFREMCHEGV QPNAGTYMII LKACASLSAL KWGKEVHACI RHGGLESDVR
 551 VGTALLRMYG KCGSINEARR IFDNLMNHDI ISWTVMIGAY AQSGNGKEAY
 601 RLMLQMEQEG FKPNAITYVS ILNACASEGA LKWVKRVHRH ALDAGLELDV
 651 RVGTALVQMY AKSGSIDDAR VVFDRMKVRD VVSWNVMIGA FAEHGRGHEA
 701 YDLFLQMQTE GCKPDAIMFL SILNACASAG ALEWVKKIHR HALDSGLEVD
 751 VRVGTALVHM YSKSGSIDDA RVVFDRMKVR NVVSWNAMIS GLAQHGLGQD
 801 ALEVFRRMTA HGVKPDRVTF VAVLSACSHA GLVDEGRSQY LAMTQVYGIE
 851 PDVSHCNCMV DLLGRAGRLM EAKLFIDNMA VDPDEATWGA LLGSCRTYGN
 901 VELGELVAKE RLKLDPKNAA TYVLLSNIYA EAGKWDMVSW VRTMMRERGI
 951 RKEPGRSWIE VDNKIHDFLV ADSSHPECKE INESKDKVIE KIKAEGYIPD
1001 TRLVLKNKNM KDKELDICSH SEKLAIVYGL MHTPPGNPIR VFKNLRVCTD
1051 CHGATKLISK VEGREIIVRD ANRFHHFKDG VCSCGDYW
```

Figure 27: Peptide coverage is diminished when RNA is used as a substrate

Recombinant PPR56 protein incubated with target RNA, crosslinked, and digestion with chymotrypsin does not yield the same peptide coverage as THU and GTP experiments.

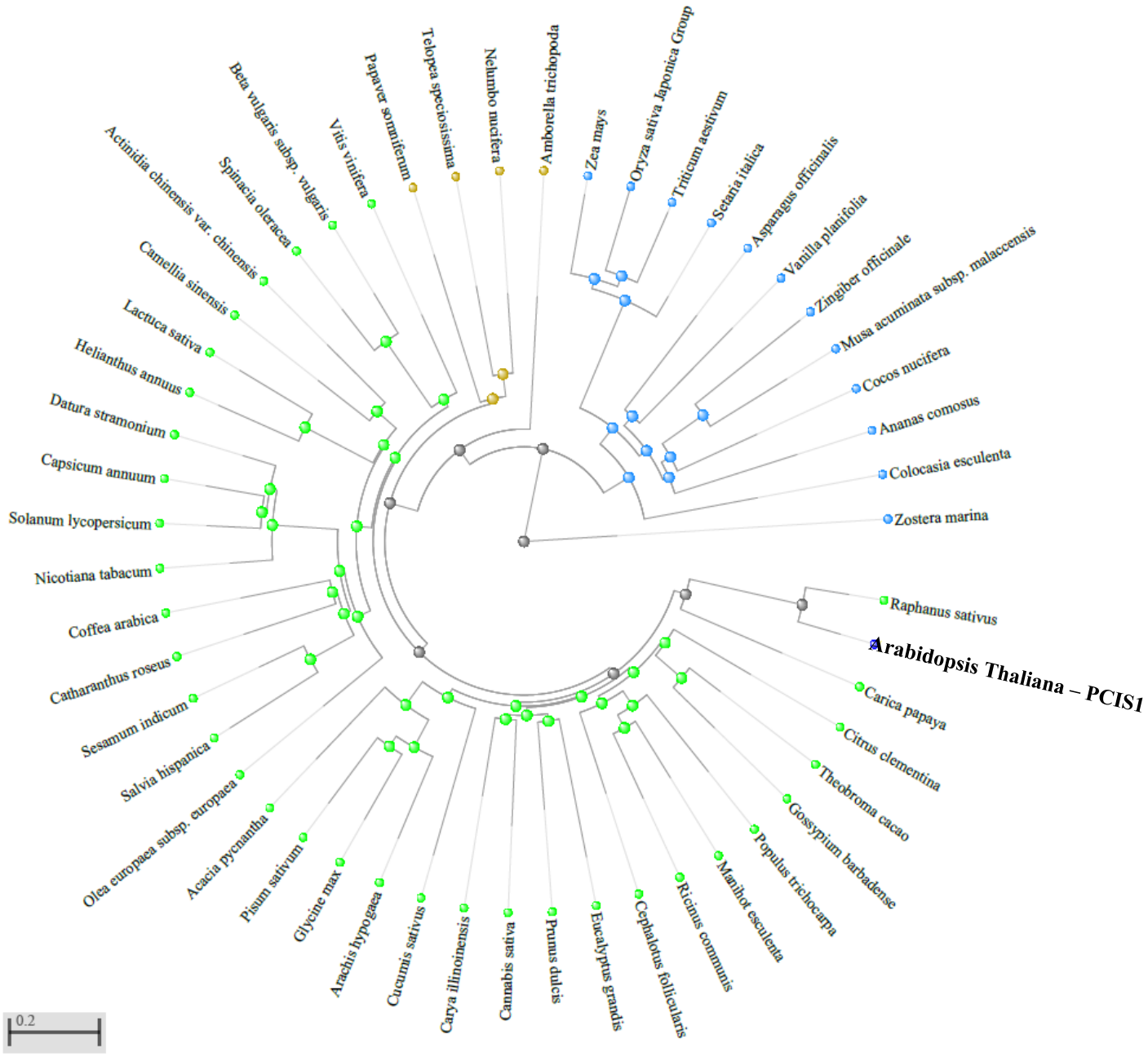


Figure 28: Phylogenic tree of PCIS1 homologs in planta

PCIS1 amino acid sequence was subjected to a protein BLAST and similar proteins were found in over 500 other plants. Representative organisms were chosen based on phylogenetic significance and plotted. Green dots represent eudicots, monocots are in blue, and other flowering plants are in brown.

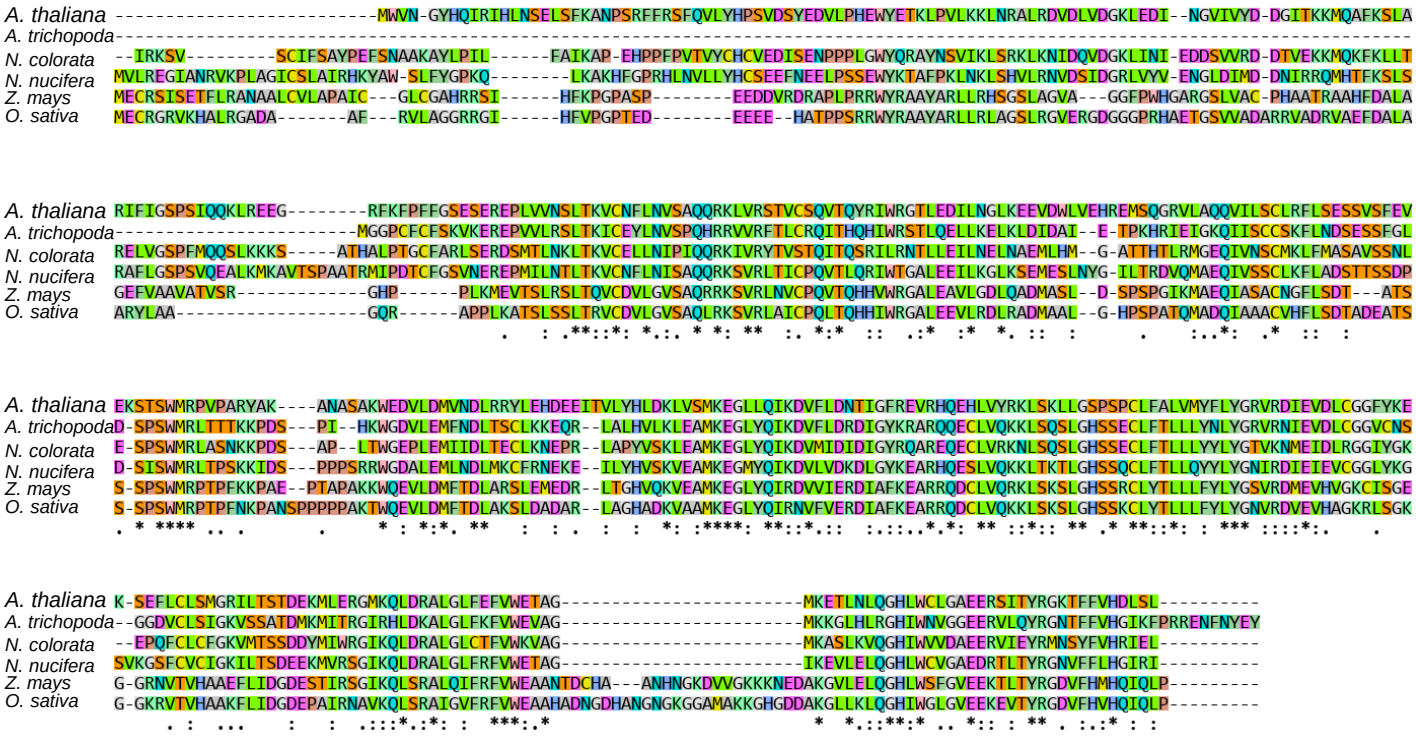


Figure 29: PCIS1 is conserved in basal flowering plants and other significant angiosperms

The PCIS1 amino acid sequence was aligned with three evolutionarily significant angiosperms, *A. trichopoda*, *N. colorata*, and *N. nucifera*. *Z. mays* and *O. sativa* were also compared and all plants shared high similarity in the middle and ends of the sequence. The beginning of all the sequences were the least conserved between the six flowering plants.

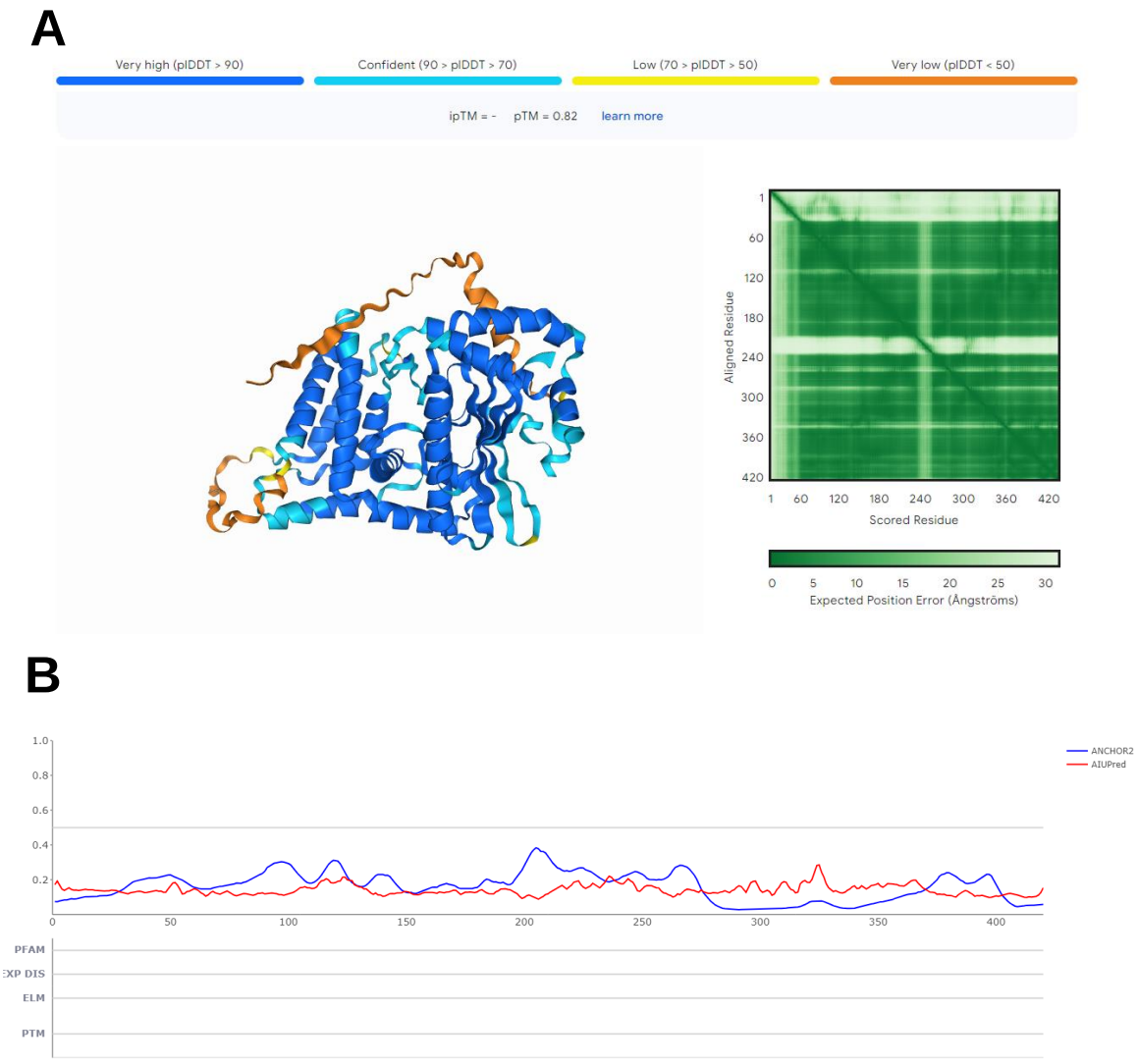


Figure 30: Structural prediction and disorder of the PCIS1 protein

A. PCIS1 protein structural prediction using AlphaFold 3. **B.** The amino acid sequence of PCIS1 was submitted to AIUPred (https://iupred.elte.hu/). AIUPred (red) and ANCHOR2 (blue) algorithms aim to predict intrinsically disordered protein regions and are signified by scores higher than 0.5 (y-axis). Predicted proteins families (PFAM), experimentally verified disordered regions (EXP DIS), eukaryotic linear motifs (ELM), and post-transcriptional modifications (PTM) are also displayed.

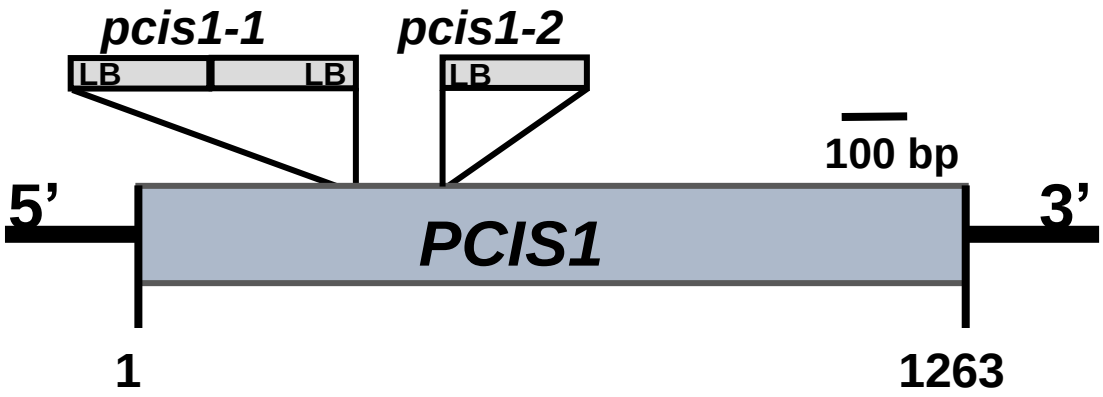


Figure 31: PCIS1 gene diagram with insertion lines
PCIS1 is a smaller gene with two T-DNA insertion lines available for use, *pcis1-1* and *pcis1-2*. *pcis1-1* was found to have two insertions located with left borders facing outwards and bases 313 to 335 deleted. The *pcis1-2* lines contains a single insertion further downstream.

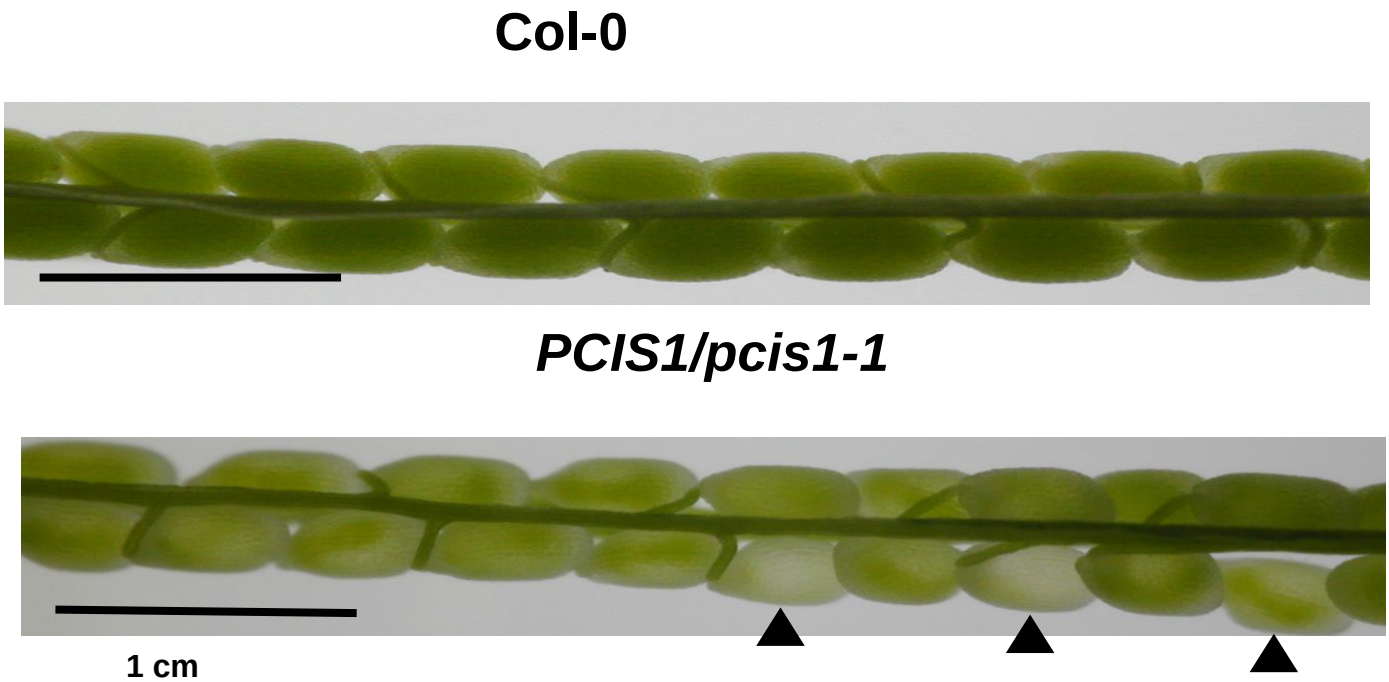


Figure 32: *pcis1-1* heterozygous plants show aborted seeds in siliques

pcis1-1 heterozygous mutant plants were regrown and siliques were examined under light microscope after 68 days of growing. Col-0 control plants show darkened, green seeds with embryos visible. Under inspection of the heterozygous *pcis1-1* seeds, some seeds would be transparent with no green color visible and no evident embryo.

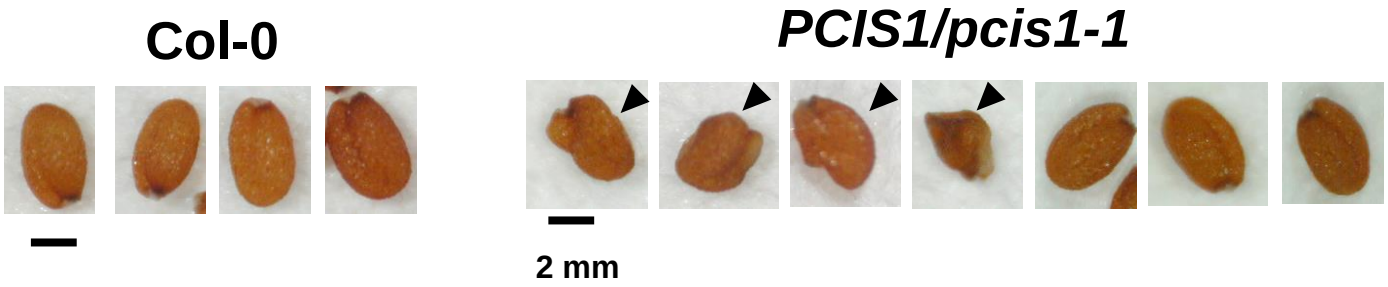


Figure 33: *pcis1-1* heterozygous seeds have visible deformities after drying

Seeds from heterozygous *pcis1-1* plants were dried for at least 3 weeks and examined under a light microscope. Col-0 seeds show typical morphology with an oval shape and uniform brown color. 23.7% of *pcis1-1* heterozygous seeds showed altered phenotypes with darkened color and shrinking of the outer seed coat.

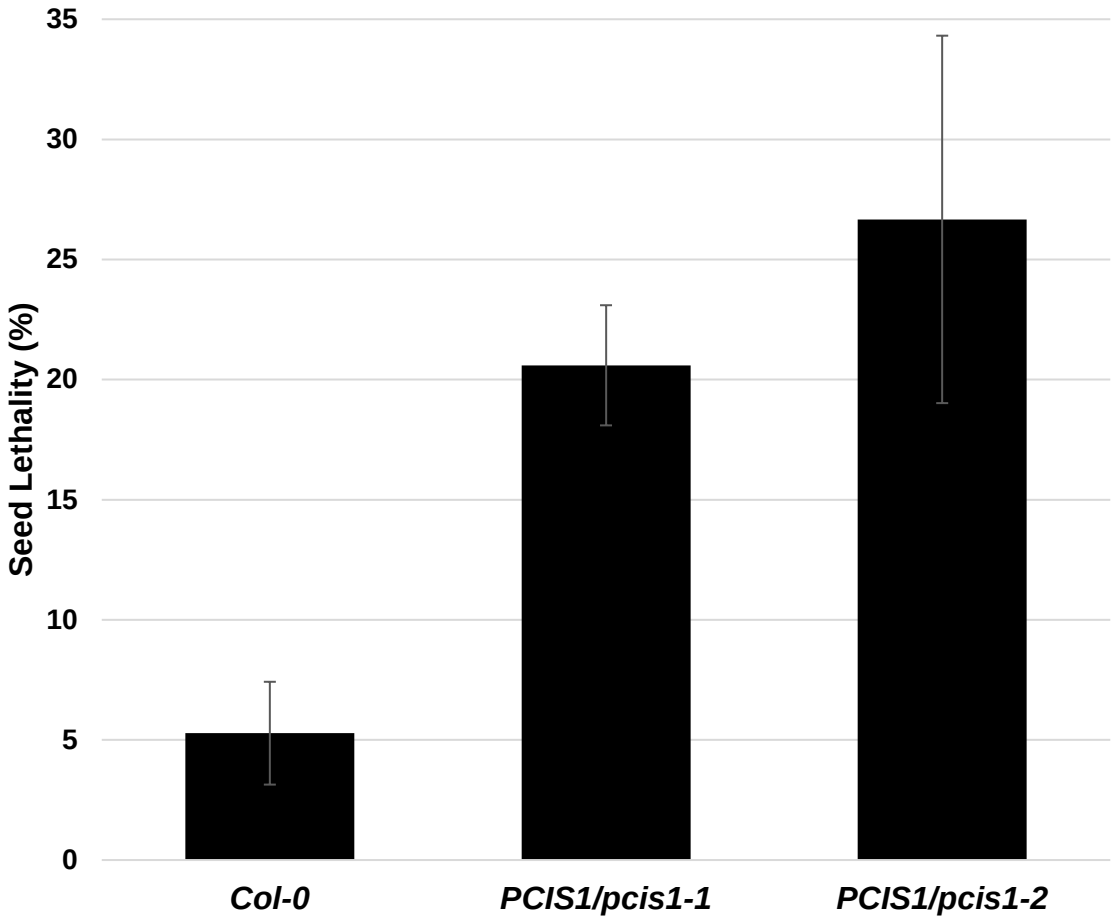


Figure 34: Seed lethality between Col-0 and *pcis1-1* heterozygotes
Seeds taken from both 656 Col-0,332 *PCIS1/pcis1-1*, and 233 *PCIS1/pcis1-2* heterozygous plants were regrown in half Murashige and Skoog medium. Plants were grown for 3 weeks and viability of seeds was assessed by examining germination rates. Seeds from heterozygous lines had significantly higher rates of embryonic lethality roughly four times greater than Col-0.

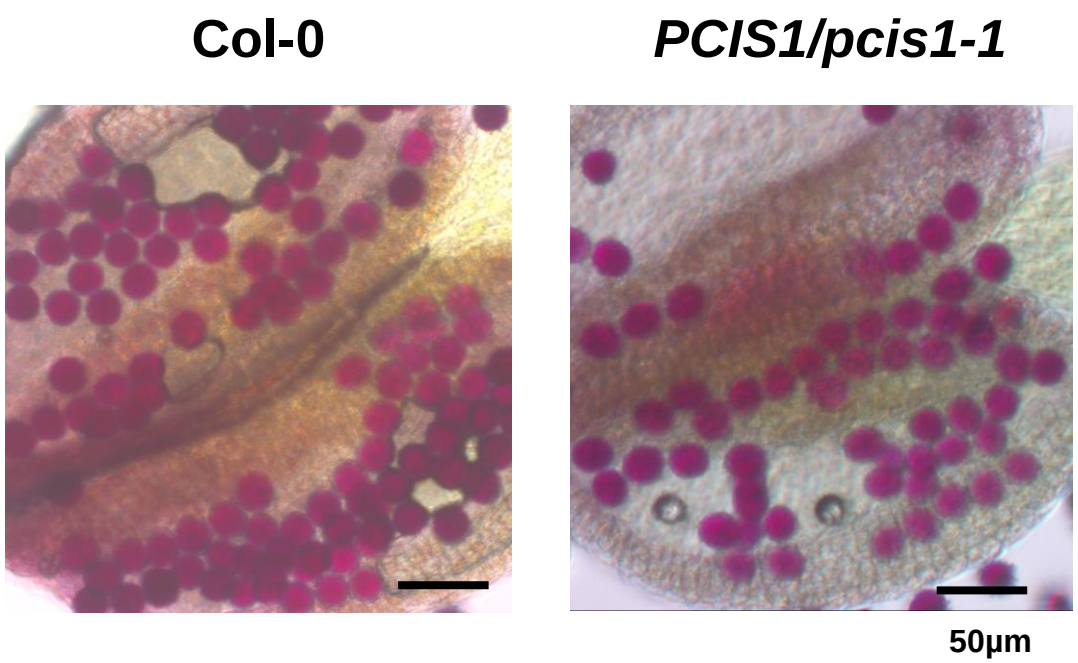


Figure 35: Alexander staining of pollen granules show no defects in pollen formation and development

Pollen from Col-0 and *pcis1-1* heterozygous plants were compared by first staining with Alexander stain and then observation under light microscope. Viable pollen granules have their pollen coat stained a greenish-blue with a clear magenta color throughout the cytoplasm. Inviable pollen granules lack the magenta coloring. Both Col-0 and heterozygous pollen showed no inviable pollen.

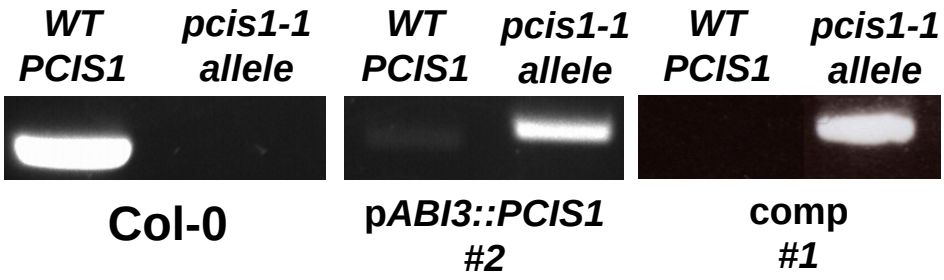


Figure 36: Genotyping of complementation lines

DNA collected from Col-0, ABI3 promoter *pcis1-1* complementation lines, and native promoter complementation lines was amplified using primers specific to the endogenous gene and for the T-DNA insertion. Amplified DNA was then analyzed on 1% agarose gels. Both complementation lines lacked the wild-type allele but contained the T-DNA insertion.

Figure 37: Growth phenotype of complementation lines

Growth phenotypes of ABI3 promoter *pcis1-1* complementation line and of the native promoter complementation line were taken one and a half months after sowing. Compared to Col-0, ABI3 complementation lines were smaller and did not bolt. Restoring normal expression using the native promoter returned growth to wild type levels.

Col-0

1 cm

p*ABI3::PCIS1 #2*

Figure 38: Leaf phenotype of ABI3 complementation line

Leaves from plants grown for one and a half months from Col-0 and ABI3 promoter *pcis1-1* complementation lines were removed and lined up in order of age. Col-0 showed vibrant, light green leaves with rounded morphology. ABI3 *pcis1-1* mutants leaves appear similar for the first few leaves, but as newer leaves grow they become darker and shriveled.

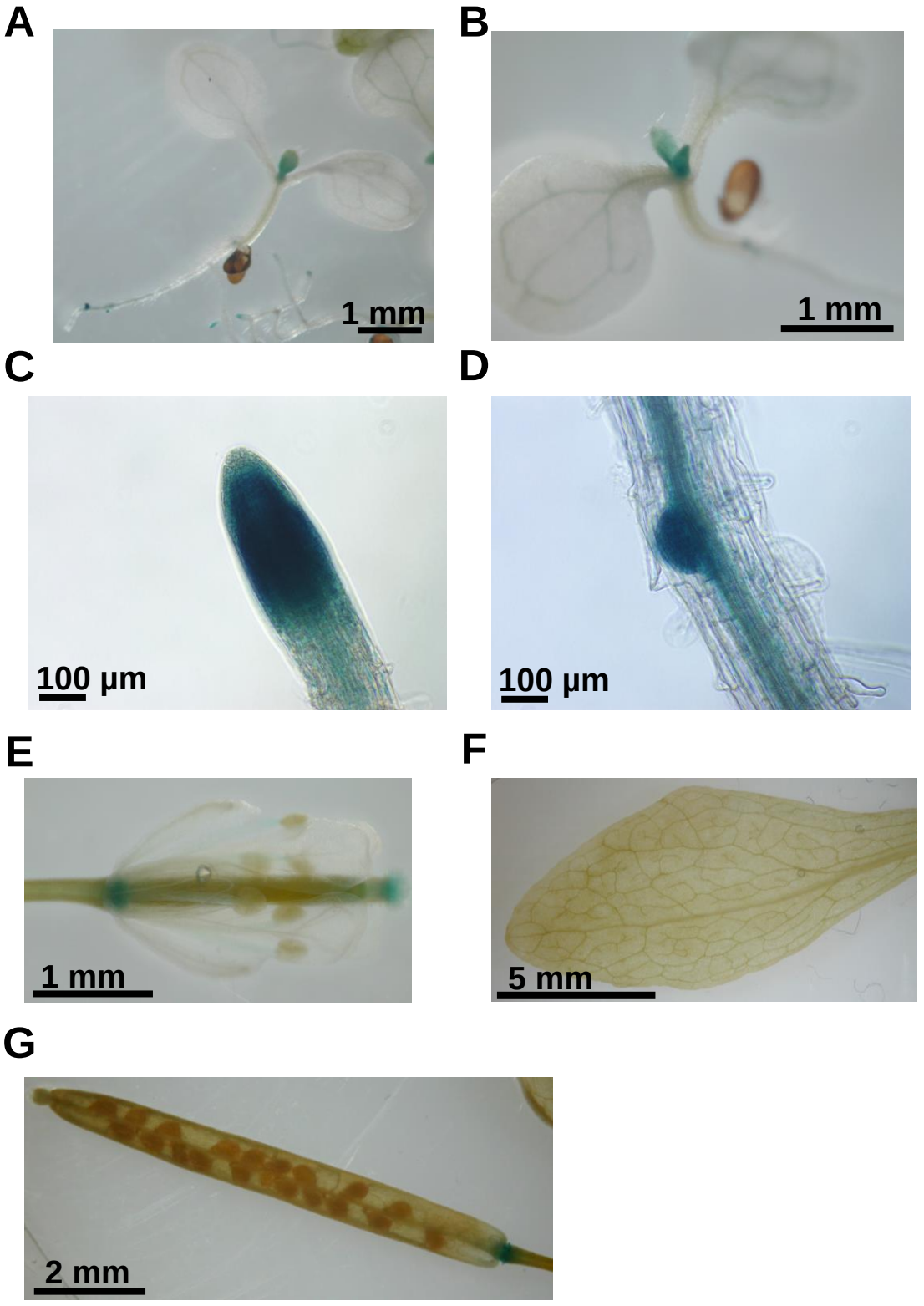


Figure 39: GUS staining of *PCISpro::PCIS1-GUS* fusion constructs

PCIS1pro::PCIS1-GUS expression in Col-0 plants in the: **A.** 6 day old seedlings **B.** magnified true leaves in 6 day old seedlings **C.** primary root apical meristem (6 days) **D.** secondary root apical meristem (6 days) **E.** flower (40 days) **F.** mature leaf (40 days) **G.** silique (40 days). Samples were stained in X-Gluc and blue coloring indicates GUS activity.

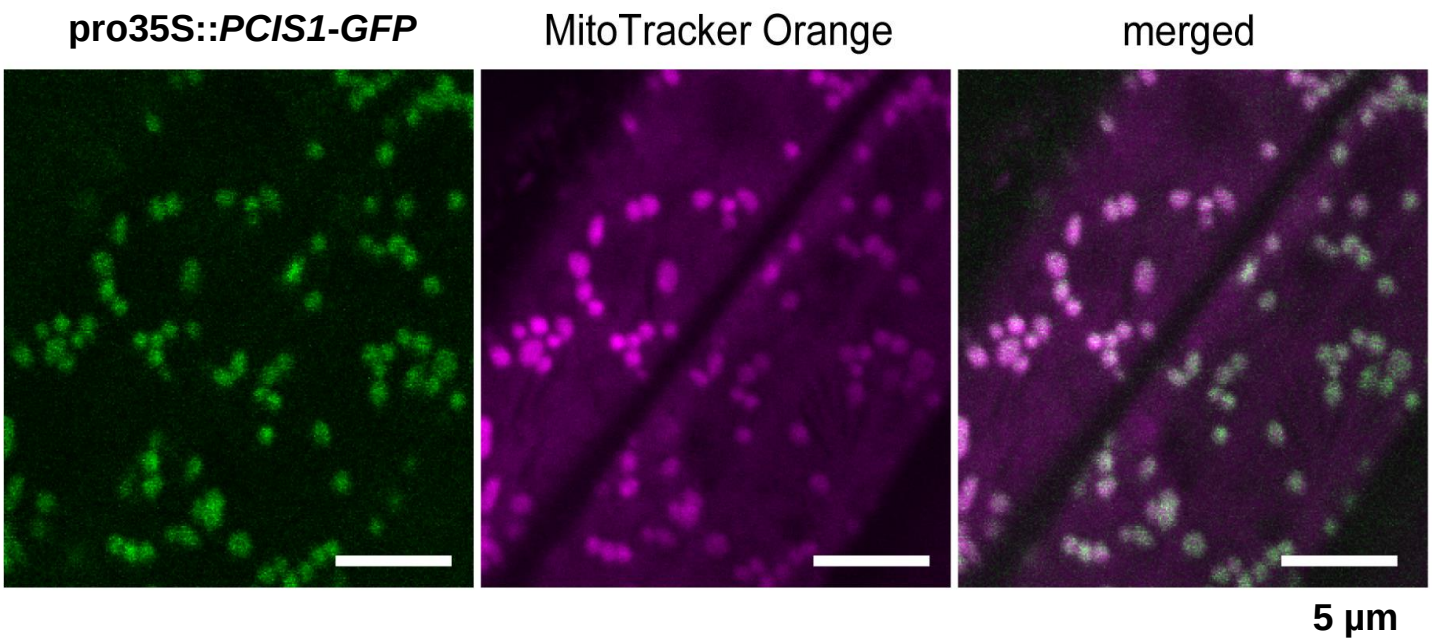


Figure 40: PCIS1 is localized in the mitochondria.
Confocal images of root epidermal cells of two-week old transgenic plants expressing pro35S::*PCIS1-EGFP* stained with MitoTracker Orange. Scale bars, 5 µm.

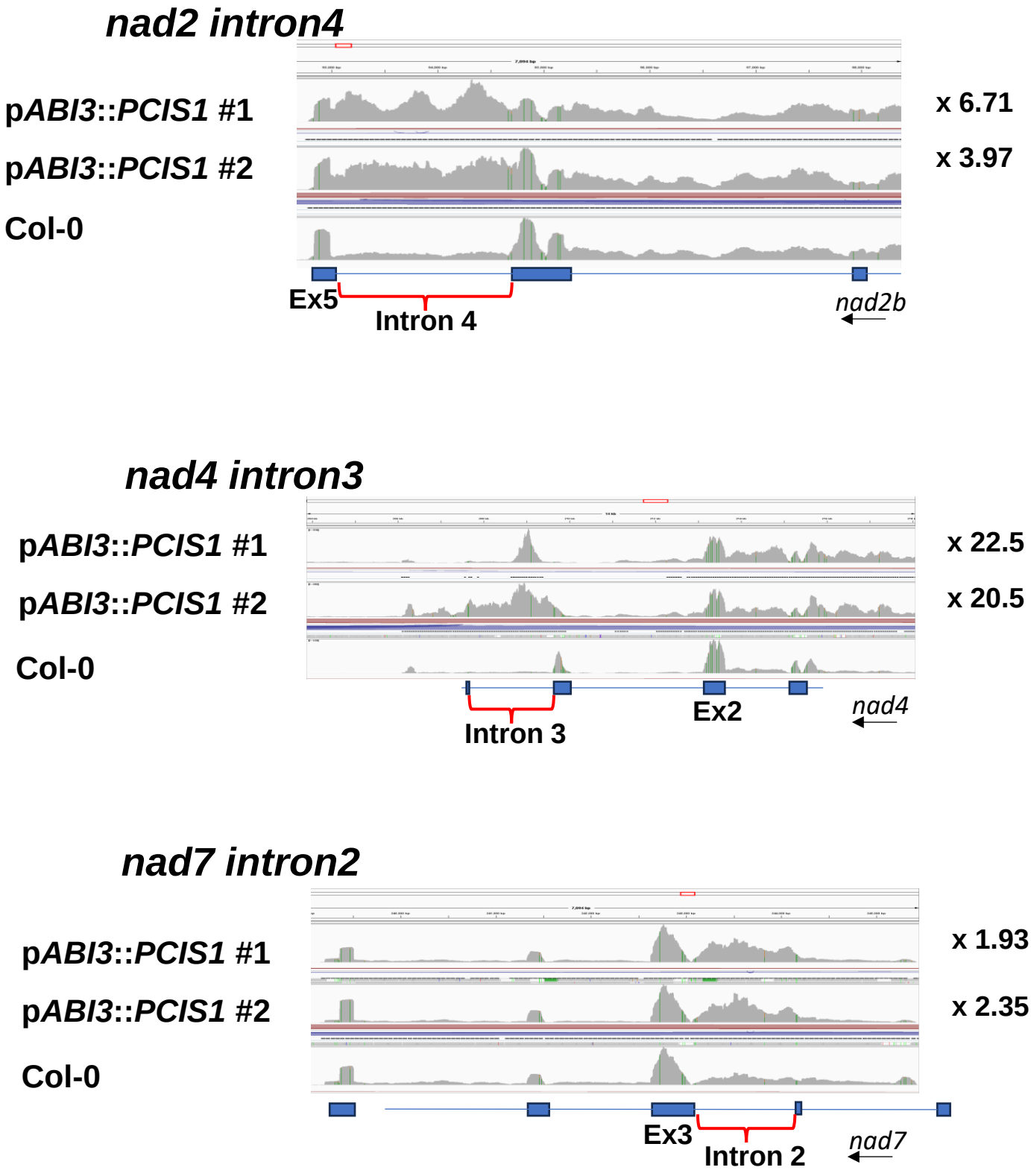


Figure 41: RNA-seq data shows accumulation of intron transcripts
mRNA taken from Col-0 and two separate ABI3 promoter *pcis1-1* complementation lines underwent RNA-seq. Reads were visualized using Integrated Genomics Viewer software and reads for introns in *nad2*, *nad4*, and *nad7* were found to be higher in abundance in ABI3 lines. Ratios of the relative abundance of each was calculated by taking the highest peak in the complementation lines and comparing them to Col-0 peaks.

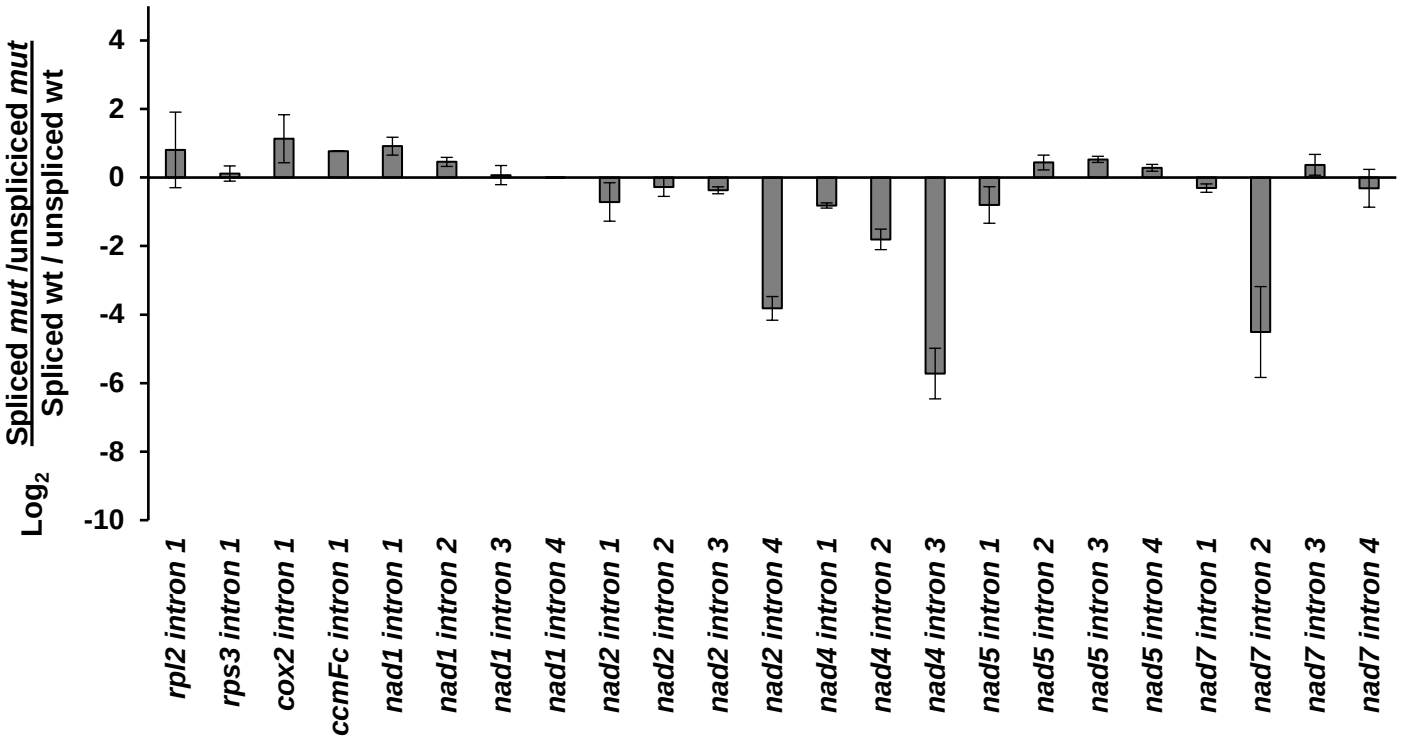


Figure 42: ABI3 promoter rescued *pcis1-1* have reduced splicing in *nad* genes

Reverse-transcription quantitative real-time PCR (RT-qPCR) analysis of the splicing efficiencies of the 23 mitochondrial group II introns in *Arabidopsis thaliana* plants. RNA extracted from whole plant tissue of wild type (Col-0) and *pcis1-1* seedlings was analyzed by RT-qPCR, after normalization to *GAPDH* (AT1G13440), *ACTIN2* (At3g18780), *18S rRNA* (At3g41768), *rrn26* (i.e., mitochondrial 26S-rRNA, Atmg00020). The histogram shows the log2 of pre-RNAs to mRNAs ratios in *pcis1-1* lines to wild type plants. The values are means of four biological replicates (error bars indicate one standard deviation).

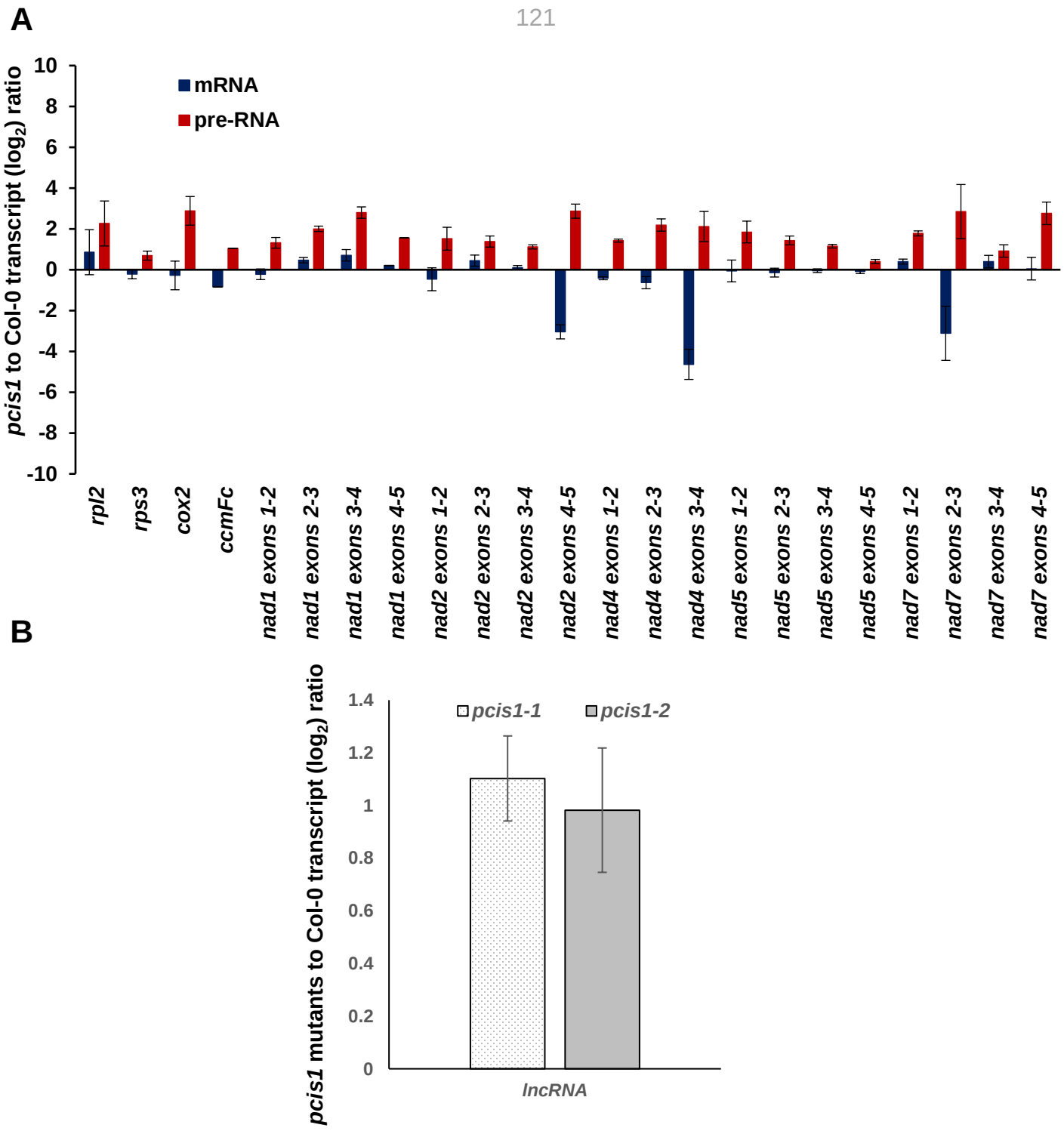


Figure 43: Abundance of mitochondrial transcripts in *pcis1-1* mutant plants.
Reverse-transcription quantitative real-time PCR (RT-qPCR) analysis of the expression of (A) the mRNAs (i.e., spliced exons) and pre-RNAs corresponding to the 23 introns in *Arabidopsis thaliana* mitochondria or (B) the accumulation of the *lncRNA* (At5g04005) transcripts. RNA extracted from whole plant tissue of wild type and mutant seedlings was analyzed by RT-qPCR, after normalization to *GAPDH* (AT1G13440), *ACTIN2* (At3g18780), *18S rRNA* (At3g41768), *rrn26* (i.e., mitochondrial 26S-rRNA, Atmg00020). The histogram shows the log2 ratio of mRNA or pre-RNA transcripts abundance in *pcis1-1* lines compared with those of wild type plants. The values are means of four biological replicates (error bars indicate one standard deviation).

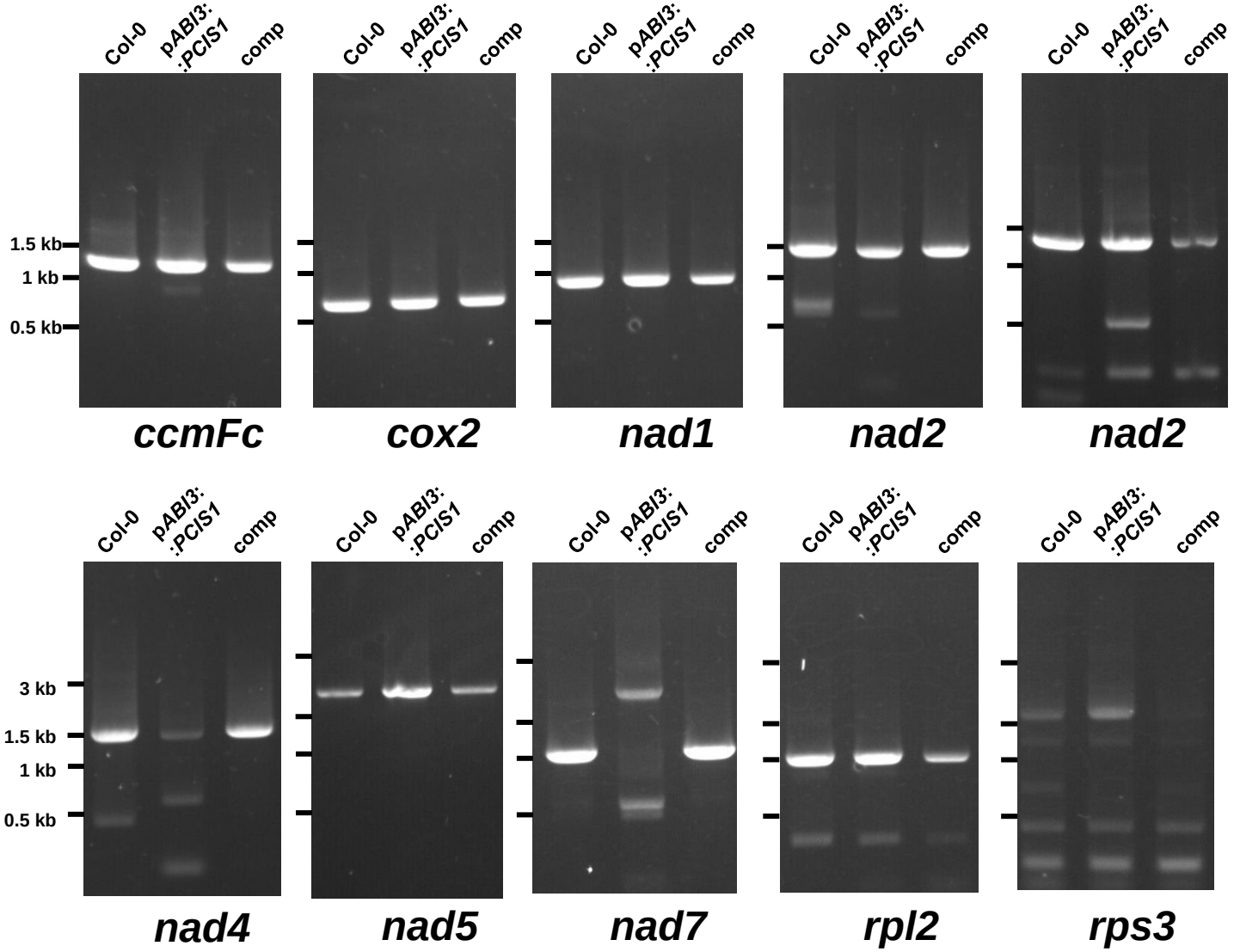


Figure 44: Detection of mature mRNAs in *pcis1-1* (p*ABI3*) mutants.
To evaluate mature mRNA for mitochondrial genes including introns, primer sets amplifying the whole or nearly the entire CDS of mitochondrial genes are used for RT-PCR of Col-0, *ABI3* promoter complemented *pcis1-1* line (*pcis1-1*(p*ABI3*)), and native promoter complemented *pcis1-1* line (comp). Size markers are denoted in lines on the left of each picture.

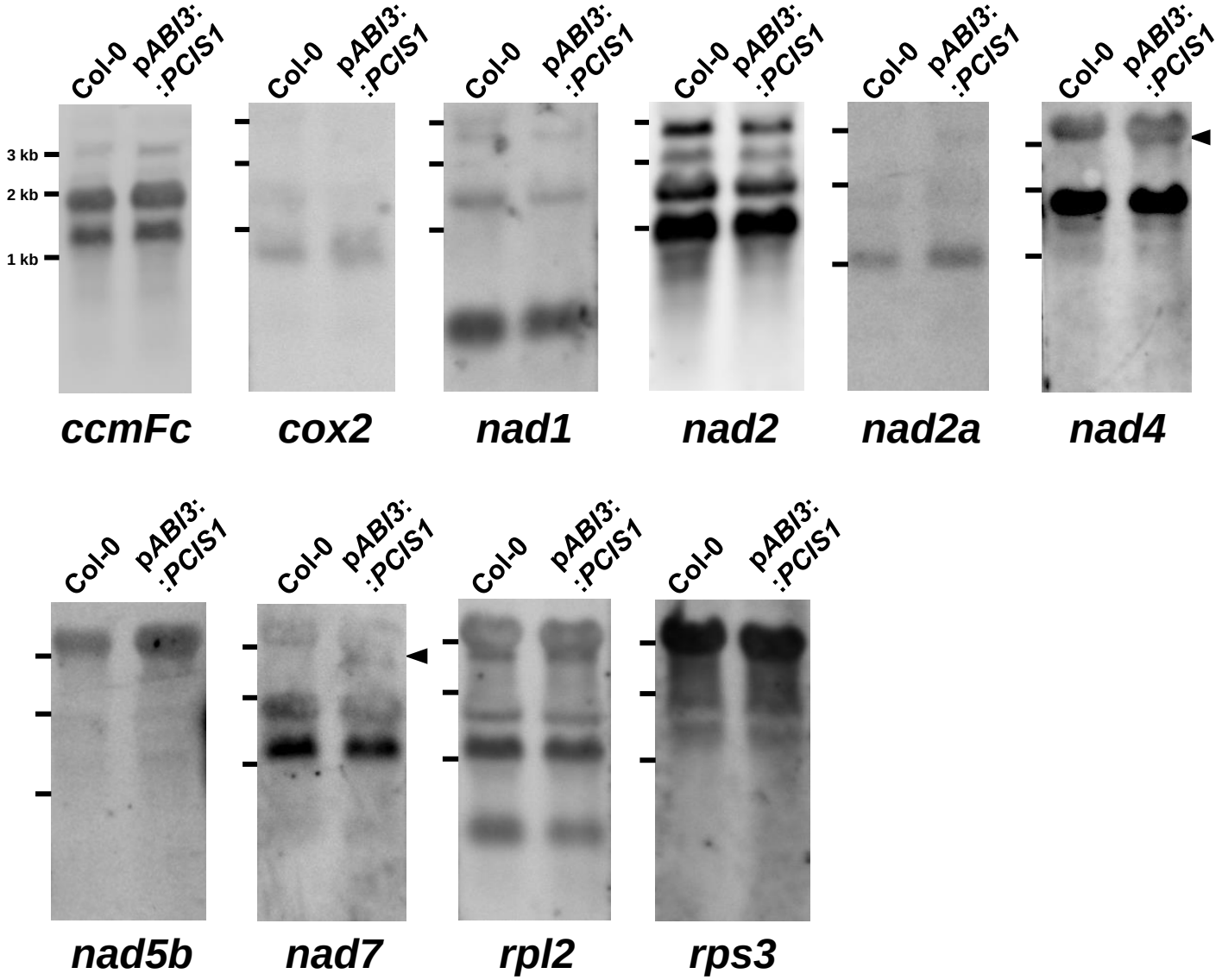


Figure 45: Northern blots show impaired splicing in *nad4* and *nad7* transcripts.

Two µg of total RNA from Col-0 and *ABI3* promoter complemented *pcis1-1* line (*pcis1-1*(p*ABI3*)) were analyzed using northern blot hybridization with DIG-11-dUTP labeled PCR products covering nearly the entire length of the mature CDS of each mitochondrial gene. Lines denoting size markers are included on the left of each image.

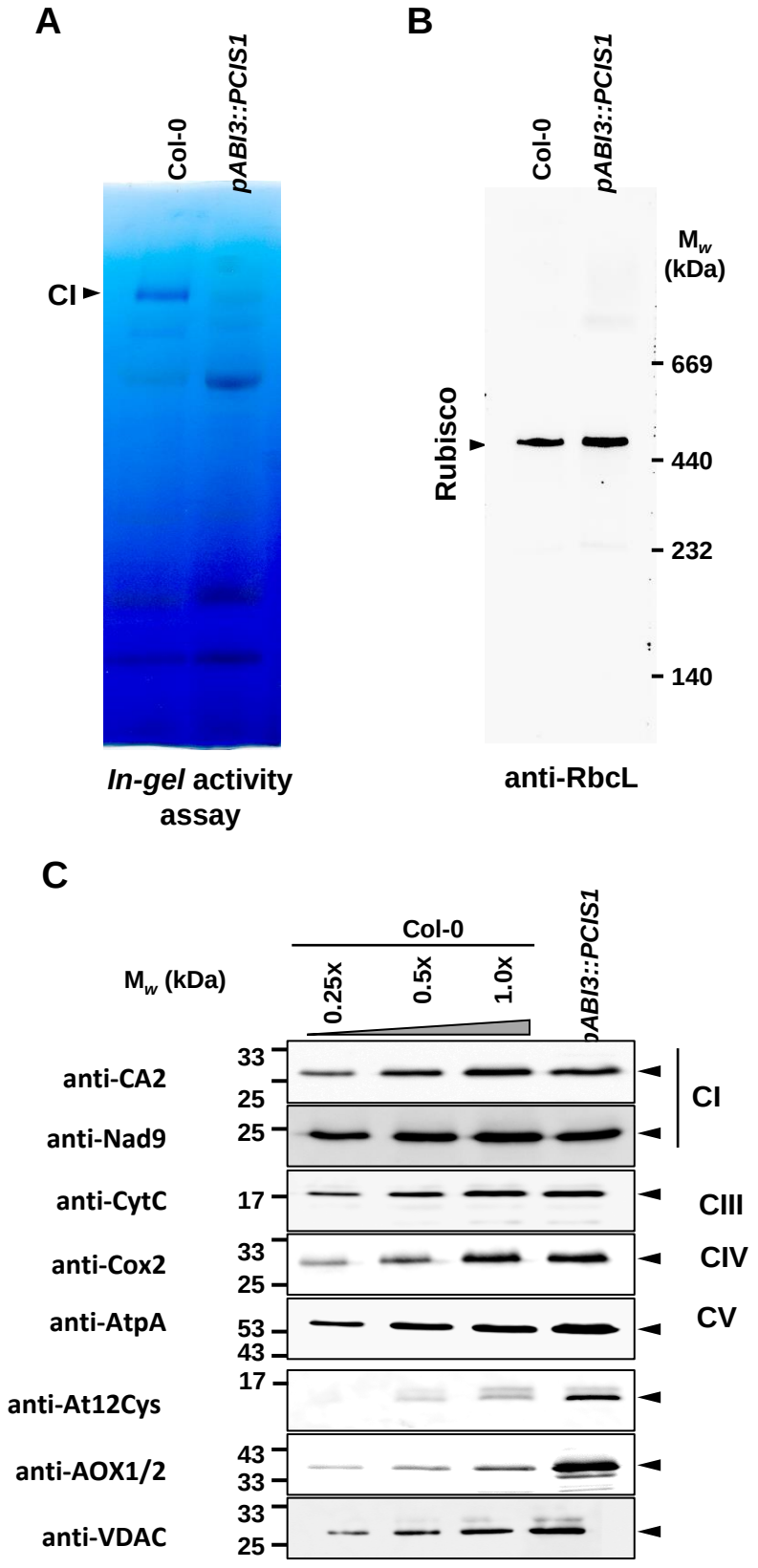


Figure 46: Mitochondrial complex I in-gel activity assay. Crude organellar preparations, obtained from three-weeks old Arabidopsis wild type (Col-0) and *ABI3* promoter complemented *pcis1-1* line (p*ABI3*::*PCIS1*) seedlings, were solubilized with DDM (1.5% (w/v)) and the organellar complexes were resolved by BN-PAGE. **A.** In-gel CI-activity assays were performed essentially as described previously (Eubel et al., 2005). **B.** For immunodetections, the proteins were transferred onto a PVDF membrane and probed with specific antibodies (Table S2) to Rubisco large subunit (RbsL), as indicated below the blot. Arrows indicate to the native Rubisco enzyme. Arrows indicate the native holo-CI (c. 1,000 kDa) in panel A and the Rubisco enzyme (~500 kDa) in panel B. **C**. Immunoblots with total membranous proteins extracted from 3-week-old MS-grown wild type (Col-0) plants and hemi-complemented *pABI3::PCIS1* mutants. The blots were probed with antibodies (mono- or polyclonal) raised to different organellar proteins, as indicated in each blot (see Table S2).

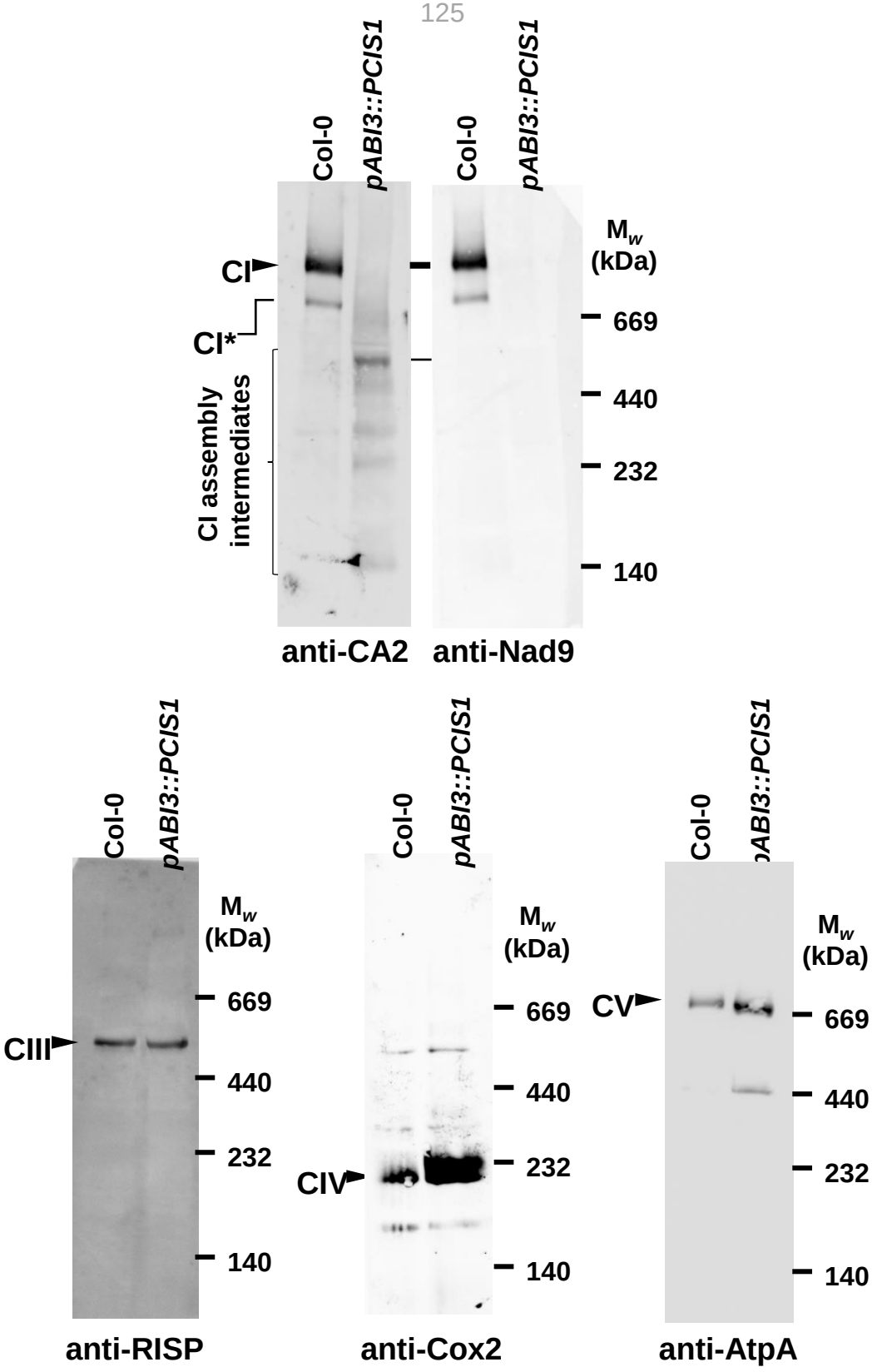


Figure 47: Mitochondrial complex I is notably reduced in Blue Native PAGE gels. Blue native PAGE gels of Col-0 and *ABI3* promoter complemented *pcis1-1* line (p*ABI3*::*PCIS1*) with antibodies targeting carbonic anhydrase 2 (CA2) and Nad9 subunits of CI, Rieske iron-sulfur protein (RISP) of CIII, the Cox2 subunit of cytochrome C oxidase (CIV), and AtpA subunit of the ATP synthase (CV) enzymes were used to assay the accumulation of the corresponding respiratory complexes. Arrows point toward native complexes I (~1000 kDa), CIII dimer (III2, ~500 kDa), CIV (about 220 kDa), and CV (~660 kDa). CI* indicates the last assembly intermediate (~850 kDa band) in complex I assembly pathway. Various bands of partial (sub-) CI assembly intermediates are indicated. An in-gel activity assay also displays the activity of complex I denoted with a black arrow.

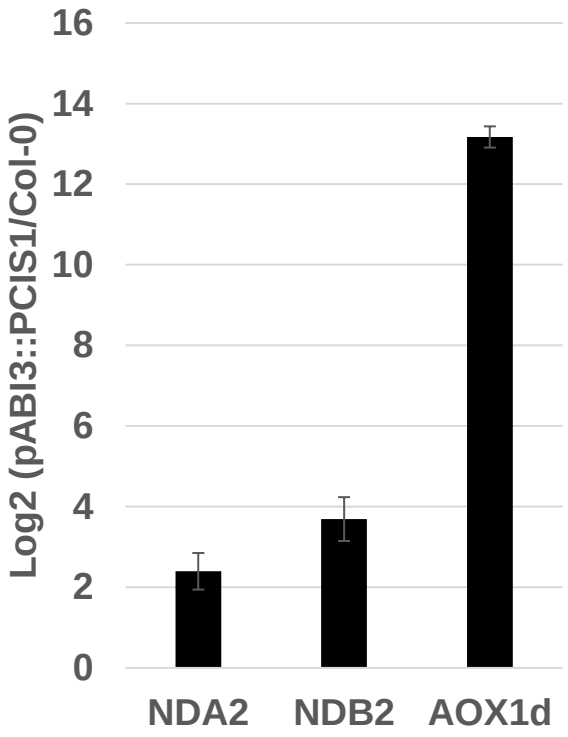


Figure 48: Comparison of expression of multiple alternative pathway-related genes in oxidative phosphorylation.
RNA-seq data displayed that two alternative NADH dehydrogenases (NDA2, NDB2) and alternative oxidase 1D expression was upregulated when ABI3 promoter complemented *pcis1-1* lines were compared to Col-0. All genes had false discovery rates near or lower than 0.03 and their respective fold changes are represented on a log base 2 scale.

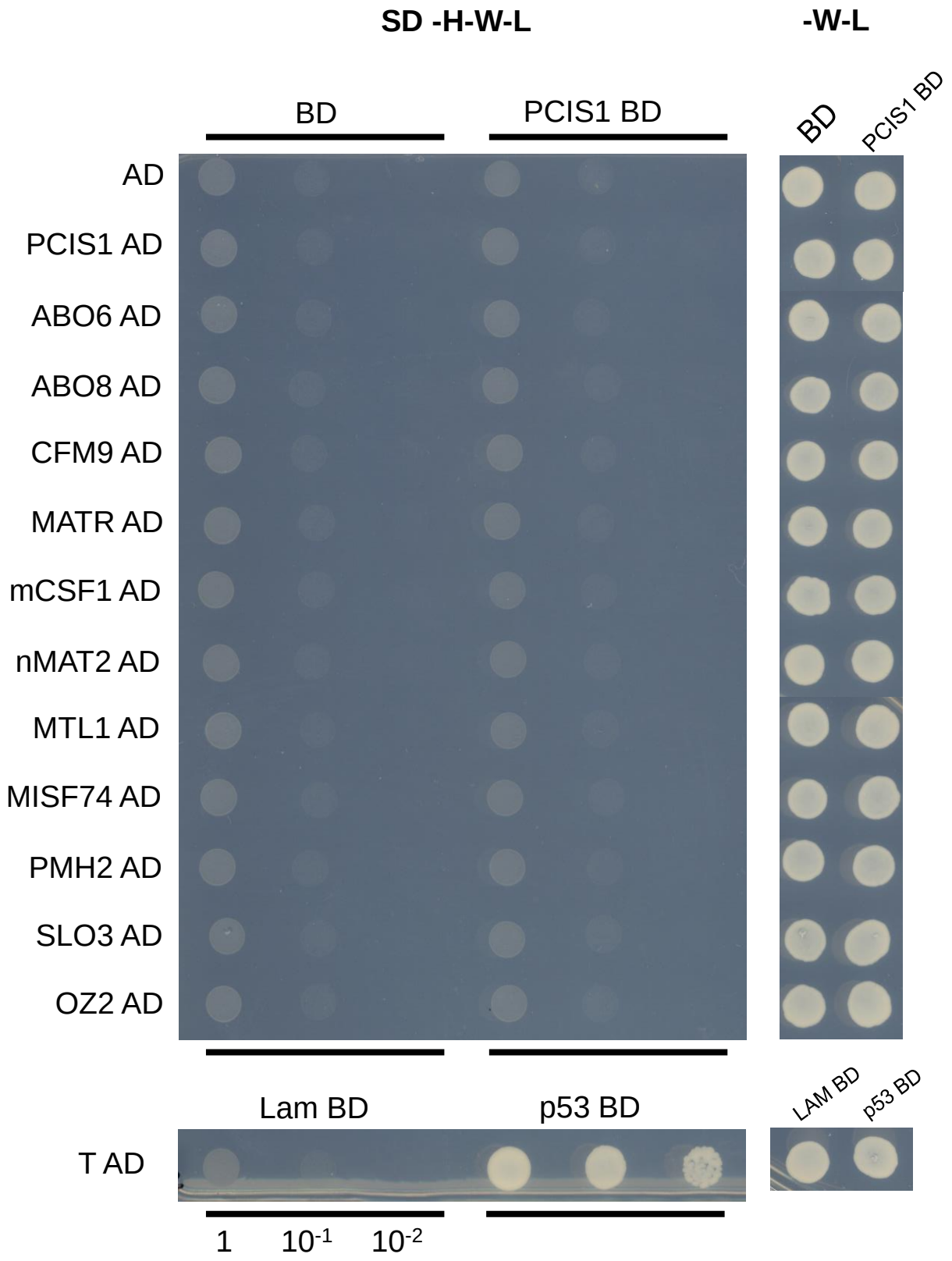


Figure 49: Yeast 2 hybrid assays show no interaction between PCIS1 and 11 other mitochondrial splicing factors. Mated yeast was plated on selective triple dropout medium (-H-W-L) on the left and double dropout medium (-W-L) on the right. AD and BD represent empty activating domain and binding domain vector-transformed yeast, respectively. Matchmaker negative (Lam) and positive (p53) controls are included. Dilutions of 1, 10^{-1}, and 10^{-2} on triple dropout medium were plated and incubated for four days.

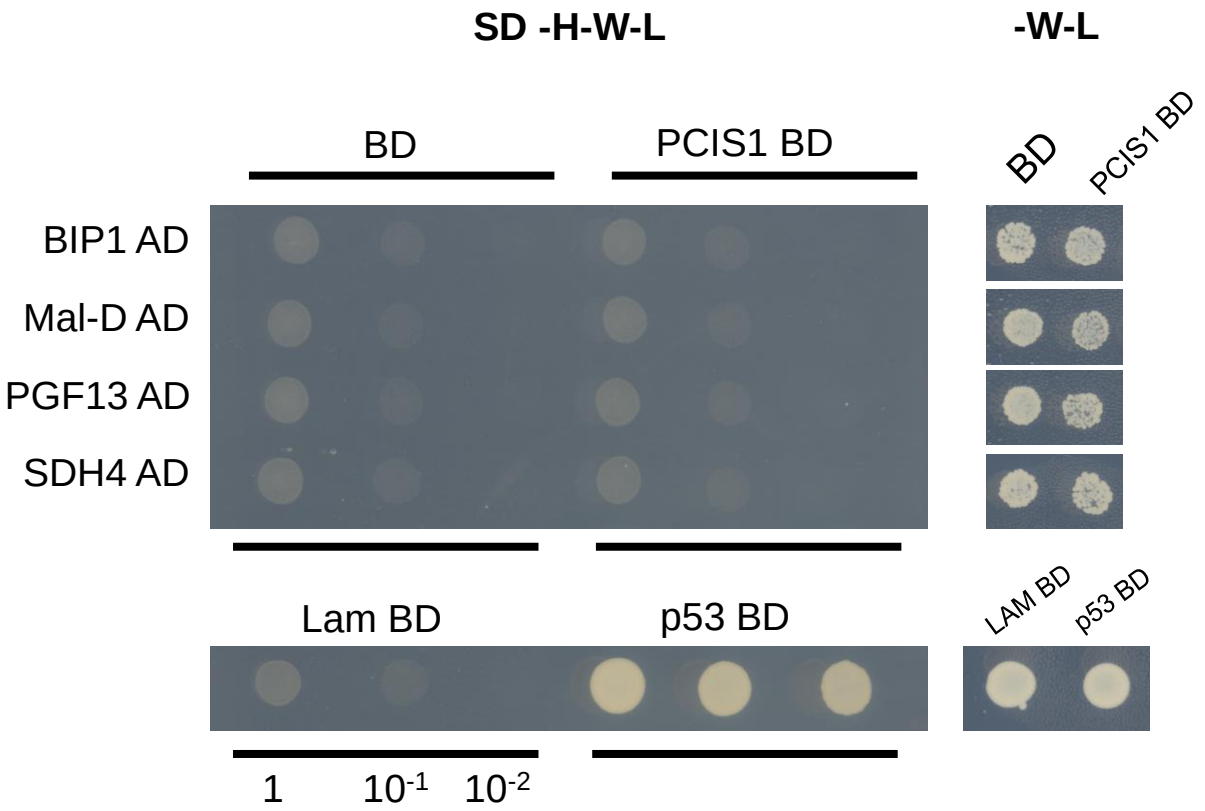


Figure 50: Yeast 2 hybrid assays show no interaction between PCIS1 the top four detected interacting proteins from GFP fusion Co-IP. Mated yeast was plated on selective triple dropout medium (-H-W-L) on the left and double dropout medium (-W-L) on the right. AD and BD represent empty activating domain and binding domain vector-transformed yeast, respectively. Matchmaker negative (Lam) and positive (p53) controls are included. Dilutions of 1, 10^{-1}, and 10^{-2} on triple dropout medium were plated and incubated for four days. Interactions were tested between BIP1, Mal d 1-associated protein (Mal-D), PGF13, and SDH4

Table 1. List of oligonucleotides used in this work.

Primer Name	Function	Sequence
T7-KS-attB2-F	Substrate RNA Transcription	TAATACGACTCACTATAGGGCTC GAGGTCGACGGTATCGTGTACAA AGTGGTGACTCGA
SK-T7up-R	Substrate RNA Transcription	CGCTCTAGAACTAGTGGATCCGG GCTTTGTTAGCAGCCGGATCTC
KS	Editing Efficiency Sequencing	CCCTCGAGGTCGACGGTATCGAT A
SK	Editing Efficiency Sequencing	GCCGCTCTAGAACTAGTGGATCC C
At5g25500-5UTRF144	Genotyping (Forward)	CCCAATTTCGCCGTTTCTTCA
At5g25500R742	Genotyping (Reverse)	GCGTTACCTGTGAACACACA
At5g25500-3'UTR-R1547	Genotyping (Reverse) / qPCR	ACAGGTAGTTGCTTGTACCTCA
At5g25500 ATGiFF	InFusion Cloning	CGAATTCTGTACAGGCATGTGGGTT AATGGGTATCATC
At5g25500nostopiFR	InFusion Cloning	GTGCGGCCGCAAGCTTGAGACTAA GGTCATGCACG
At5g25500iFR	InFusion Cloning	GTGCGGCCGCAAGCTTTCAGAGAC TAAGGTCATGCACG
At5g25500Up-8879943-F	Native Promoter Complementation	CGAATTCTGTACAGGCGGTGACAC GTTAACGACACT
At5g25500-R85-Seq	Sequencing	AACGGAAGAAGCGACTTGGA
At5g25500-F1438	Sequencing / qPCR	CATTTGTGGTGCCTTGGAGC
LB1.3	Genotyping (tDNA)	ATTTTGCCGATTTCGGAAC
pGBKT7NdeI-iFF	Yeast 2 Hybrid	AGGAGGACCTGCATATCGAATTCT GTACAGGC
pGBKT7BamHI-iFR	Yeast 2 Hybrid	GCAGGTCGACGGATCGTGCGGCCG CAAGCTT
pGADT7NdeI-iFF	Yeast 2 Hybrid	CAGATTACGCTCATATCGAATTCTG TACAGGC
pGADT7XhoI-iFR	Yeast 2 Hybrid	TCATCTGCAGCTCGAGTGCGGCCGC AAGCTT
pBT up	Yeast 2 Hybrid	CAGTGGAGACTGATATGCCTC
pBT rev	Yeast 2 Hybrid	CCTACAGGAAAGAGTTACTCAAG
pACT F	Yeast 2 Hybrid	CTATCTATTCGATGATGAAG
3AD	Yeast 2 Hybrid	AGATGGTGCACGATGCACAG
AT5G04895ATGiFF	Yeast 2 Hybrid	CGAATTCTGTACAGGCATGCGTTTC ACAAAAAGAATAAGCC

AT5G04895iFRStop	Yeast 2 Hybrid	GTGCGGCCGCAAGCTTTTACTTGCC CTTGGATCGCC
AT5G04895-UpF28	Yeast 2 Hybrid	TGCCTCTTCGCTCATTTCGT
AT5G04895-DownR140	Yeast 2 Hybrid	TCTGCCCCAAAAACGACACT
AT5G04895-F681	Yeast 2 Hybrid	ATAAGCGACCACAACGGGAG
AT5G04895-F846	Yeast 2 Hybrid	TGGGGAGACTGGATGTGGTA
AT5G04895-F2336	Yeast 2 Hybrid	GCTCCAGAGTCTTTGGCTGT
AT5G04895-R2644	Yeast 2 Hybrid	TGACAGCACGAACCAGAGAC
AT4G11690ATGiFF	Yeast 2 Hybrid	CGAATTCTGTACAGGCATGGCGGC AAAATCCCAG
AT4G11690iFRStop	Yeast 2 Hybrid	GTGCGGCCGCAAGCTTTTATTTTGA GCTTACATGTGAATCG
AT3G27550ATGiFF	Yeast 2 Hybrid	CGAATTCTGTACAGGCATGTTTGTC TCTAGGAGTCT
AT3G27550iFRStop	Yeast 2 Hybrid	GTGCGGCCGCAAGCTTCTAATCACT ATCCCAATCCT
ATMG00520ATGiFF	Yeast 2 Hybrid	CGAATTCTGTACAGGCATGAAAGA GGCGATCAGAATGGT
ATMG00520iFRStop	Yeast 2 Hybrid	GTGCGGCCGCAAGCTTCTAGACTA GACTAGTAGTTGAGTGT
ATMG00520-F734	Yeast 2 Hybrid	CCTTCCTCATAGAAGCCGCC
AT4G31010ATGiFF	Yeast 2 Hybrid	CGAATTCTGTACAGGCATGTTCTTG ATTCGTCTCTCCCG
AT4G31010iFRStop	Yeast 2 Hybrid	GTGCGGCCGCAAGCTTCTAGGTTGT CTCGTCAGGAG
AT5G46920ATGiFF	Yeast 2 Hybrid	CGAATTCTGTACAGGCATGCGTAG AAGCTTCTCTGT
AT5G46920iFRStop	Yeast 2 Hybrid	GTGCGGCCGCAAGCTTTTACATGCG TGCAATGCGAAGC
AT5G46920-F820	Yeast 2 Hybrid	GCGCTTGTTACACCTGTTGT
AT5G64320ATGiFF	Yeast 2 Hybrid	CGAATTCTGTACAGGCATGGTTATG CTAGCGAGATCA
AT5G64320iFRStop	Yeast 2 Hybrid	GTGCGGCCGCAAGCTTTCAAAAAG CTGCATTATAAAACCTT
AT5G64320-F843	Yeast 2 Hybrid	GGGATGTGTTCCTGATGCTGA
AT3G22330ATGiFF	Yeast 2 Hybrid	CGAATTCTGTACAGGCATGATCACT ACAGTGCTACGACG
AT3G22330iFRStop	Yeast 2 Hybrid	GTGCGGCCGCAAGCTTTCAGTAAG ATCTTTTCCCATCATTT

AT3G22330-F637	Yeast 2 Hybrid	GGTACACCGATTGGACAGCA
AT3G61360ATGiFF	Yeast 2 Hybrid	CGAATTCTGTACAGGCATGCTTCAG AAGATCTCCTCCG
AT3G61360iFRStop	Yeast 2 Hybrid	GTGCGGCCGCAAGCTTCTACATTAG CTGTATTTCAGGAGG
AT4G01400ATGiFF	Yeast 2 Hybrid	CGAATTCTGTACAGGCATGATTCGC CGGCCGATCTACG
AT4G01400iFRStop	Yeast 2 Hybrid	GTGCGGCCGCAAGCTTCTATAAATG TCTCCTCCTCT
AT1G55040ATGiFF	Yeast 2 Hybrid	CGAATTCTGTACAGGCATGGCTGCT TCAATCTCTCT
AT1G55040iFRStop	Yeast 2 Hybrid	GTGCGGCCGCAAGCTTCTATCTCTC GATAACTCTTC
AT1G55040-UpF21	Yeast 2 Hybrid	GGCTGCTTCTGCAACCTAAA
AT1G55040-DownR178	Yeast 2 Hybrid	CGAGGCATATGAGCTCACGA
AT1G55040-F386	Yeast 2 Hybrid	CTTGTGTTGTGCTTGAGCCG
AT1G55040-F693	Yeast 2 Hybrid	CCTAGTCGGTCATGGATGTCC
AT1G55040-F1377	Yeast 2 Hybrid	AGACAGGGTTGTGAAACGGG
AT2G35900ATGiFF	Yeast 2 Hybrid	CGAATTCTGTACAGGCATGGGTTGG ATCTGGATTGA
AT2G35900iFRStop	Yeast 2 Hybrid	GTGCGGCCGCAAGCTTTCAAACGTC CTTGGCCAAGC
AT5G28540ATGiFF	Yeast 2 Hybrid	CGAATTCTGTACAGGCATGGCTCGC TCGTTTGGAGC
AT5G28540iFRStop	Yeast 2 Hybrid	GTGCGGCCGCAAGCTTCTAGAGCTC ATCGTGAGACT
AT5G28540-F655	Yeast 2 Hybrid	AAGAAGGGTGGCGAGAAGAA
AT2G46505ATGiFF	Yeast 2 Hybrid	CGAATTCTGTACAGGCATGTCTCTC CGCCGCACTAT
AT2G46505iFRStop	Yeast 2 Hybrid	GTGCGGCCGCAAGCTTTCAGAGAA GAAACAAGATAA
AT4G23820ATGiFF	Yeast 2 Hybrid	CGAATTCTGTACAGGCATGTGGAG ACTCTCTGTGTC
AT4G23820iFRStop	Yeast 2 Hybrid	GTGCGGCCGCAAGCTTTTAGAATGT GGAAGAACACA

Table 2. List of antibodies used for the analysis of wild type and mutant plants.

Antibody	**Protein I.D.**	**origin**	**serum**	**dilution**	**Reference / source**
At12Cys		*Arabidopsis thaliana*	Rabbit (polyclonal)	1/1,000	(Y. Wang et al., 2016)
AtpA	Mitochondrial ATP-synthase subunit 1 (A)	*Nicotiana tabacum*	Mouse (monoclonal)	1/500	PM014 (Thomas Elton collection)
CA2	γ-carbonic anhydrase-like subunit 2	*Arabidopsis thaliana*	Rabbit (polyclonal)	1/1,000	(Perales et al., 2005; Sunderhaus et al., 2006)
Cox2	Cytochrome C oxidase subunit-2	*Arabidopsis thaliana*	Rabbit (polyclonal)	1/5,000	Agrisera antibodies, AS04 053A
Nad9	NADH-dehydrogenase complex subunit-9	*Triticum spp.*	Rabbit (polyclonal)	1/50,000	(Lamattina et al., 1993)
RISP	Rieske iron-sulfur protein	*Arabidopsis thaliana*	Rabbit (polyclonal)	1/5,000	Gift of Prof. Ian Small, UWA
RbcL	Rubisco large subunit	*Pisum sativum*	Rabbit (polyclonal)	1/5,000	Gift of Prof. Michal Shapira

Table 3. List of oligonucleotides used for the analysis of the splicing efficiencies of mutant plants by RT-qPCR.

Gene	Forward primer	Reverse primer
rpl2	CCGAAGACGGATCAAGGTAA	CGCAATTCATCACCATTTTG
rpl2 intron exon2	TTAGGAAGAGCCGTACGAGG	CGCAATTCATCACCATTTTG
rps3	AGCCGAAGGTGAGTCTCGTA	CCGATTTCGGTAAGACTTGG
rps3 intron1 exon2	AGCCGAAGGTGAGTCTCGTA	TCTACGGCGGGGTCACTAT
cox2	TGGGGGATTAATTGATTGGA	TGATGCTGTACCTGGTCGTT
cox2 intron1 exon2	TGGGGGATTAATTGATTGGA	AGCAGTACGAGCTGAAAGGC
ccmFc	GTGGGTCCATGTAAATGATCG	CACATGGAGGAGTGTGCATC
ccmFc intron1 exon1	CCCGGATCGAATCAGAGTT	CACATGGAGGAGTGTGCATC
nad1 exon1-2	GACCAATAGATACTTCATAAGAGACCA	TTGCCATATCTTCGCTAGGTG
nad1 intron1 exon2	GACCAATAGATACTTCATAAGAGACCA	CGTGCTCGTACGGTTCATAG
nad1 exon2-3	ATTCAGCTTCCGCTTCTGG	TCTGCAGCTCAAATGGTCTC
nad1 intron2 exon2	GGTTGGGTTAGGGGAACATC	TCTGCAGCTCAAATGGTCTC
nad1 exon3-4	AAAAGAGCAGACCCCATTGA	TCCGTTTGATCTCCCAGAAG
nad1 intron3 exon4	AAAAGAGCAGACCCCATTGA	GGGAGCTGTATGAGCGGTAA
nad1 exon4-5	AGCCCGGGATCTTCTTGA	TCTTCAATGGGGTCTGCTC
nad1 intron4 exon5	AGCCCGGGATCTTCTTGA	ACGGAGCTGCATCCCTACT
nad2 exon1-2	GCGAGCAGAAGCAAGGTTAT	GGATCCTCCCACACATGTTC
nad2 intron1 exon2	GCGAGCAGAAGCAAGGTTAT	CCCATTCCTAACCAGTGGAG
nad2 exon2-3	AAAGGAACTGCAGTGATCTTGA	AATATTTGATCTTAGGTGCATTTTC
nad2 intron2 exon2	CCCGATCCGATAGTTTACAA	AATATTTGATCTTAGGTGCATTTTC
nad2 exon3-4	GCGCAATAGAAAGGAATGCT	CTATGGGTCTACTGGAGCTACCC
nad2 intron3 exon4	GCGCAATAGAAAGGAATGCT	GGCGAATTTCAAACTTGTGG
nad2 exon4-5	CAAAGGAGAGGGGTATAGCAA	TATTTGTTCTTCGCCGCTTT
nad2 intron4exon4	CTTATTCGTGGCAACCTTCC	TATTTGTTCTTCGCCGCTTT
nad4 exon1-2	ATTCTATGTTTTTCCCGAAAGC	GAAAAACTGATATGCTGCCTTG
nad4 intron1 exon2	CCGTATGATGCGGAAGTCTC	GAAAAACTGATATGCTGCCTTG
nad4 exon2-3	AATACCCATGTTTCCCGAAG	TGCTACCTCCAATTCCCTGT
nad4 intron2 exon3	GCGGAACGACCAGAAAAATA	TGCTACCTCCAATTCCCTGT
nad4 exon3-4	TTCCTCCATAAATTCTCCGATT	TGAAATTTGCCATGTTGCAC
nad4 intron3 exon4	TCTAGCTTGGTTCGGAGAGC	TGAAATTTGCCATGTTGCAC
nad5 exon1-2	TGGACCAAGCTACTTATGGATG	CCATGGATCTCATCGGAAAT
nad5 intron1 exon2	TGGACCAAGCTACTTATGGATG	TTCGCAAATAGGTCCGACT
nad5 exon2-3	TACCTAAACCAATCATCATATC	CTGGCTCTCGGGAGTCTCTT
nad5 intron2-exon2	GTACGATCGTGTCGGGTGA	CTGGCTCTCGGGAGTCTCTT
nad5 exon3-4	AACTCGGATTCGGCAAGAA	GATATGATGATTGGTTTAGGTA
nad5 intron3-exon4	AACTCGGATTCGGCAAGAA	GCCGTGTAATAGGCGACCA
nad5 exon4-5	AACATTGCAAAGGCATAATGA	GTTCCTGCGTTTCGGATATG
nad5 intron4 exon5	AACATTGCAAAGGCATAATGA	CCTGTAAACCCCCATGATGT

nad7 exon1-2	ACCTCAACATCCTGCTGCTC	AAGGTAAAGCTTGAAGATAAGTTTTGT
nad7 intron1 exon2	ACGGTTTTTAGGGGGATCTG	AAGGTAAAGCTTGAAGATAAGTTTTGT
nad7 exon2-3	GAGGGACTGAGAAATTAATAGAGTACA	TGGTACCTCGCAATTCAAAA
nad7 intron2 exon3	AGTGGGAGAGCCGTGTTATG	TGGTACCTCGCAATTCAAAA
nad7 exon3-4	ACTGTCACTGCACAGCAAGC	CATTGCACAATGATCCGAAG
nad7 intron3 exon4	TAAAGTGAAGTGGTGGGCCT	CATTGCACAATGATCCGAAG
nad7 exon4-5	GATCAAAGCCGATGATCGTAA	AGGTGCTTCAACTGCGGTAT
nad7 intron4 exon5	CGGCCAAATGACTACAGGAT	AGGTGCTTCAACTGCGGTAT

Table 4. Primers used for Northern blots, RT-PCR assays, and RNA immunoprecipitation assays

Primer Name	**Sequence**
ccmFc-ex1-F1	ATGGTCCAACTACATAACTTTTTCTT
ccmFc-ex2-R2315	TGCAATTATGAACTCCACGGA
cox2-ex1-F6	TGTTCTAAAATGGTTATTCCTCACAA
cox2-ex2-R2126	TTAAGCTTCCCCGGTTTGG
nad1-ex1-F11	CTGTTCCAGCTGAAATACTTGGA
nad1-rev	TTAAGGAAGCCATTGAAAGG
nad2a_qF	GGATCCTCCCACACATGTTC
nad2b_qR	CAAAGGAGAGGGGTATAGCAA
nad2a-ex2-R1545	TTCCAGAGGAAAATGCACCT
nad2b-ex3-F9	TGATCTATGGGTCTACTGGAGCT
nad4-ex1-F	AGAACATTTCTGTGAATGC
nad4_ex3_4R	TGAAATTTGCCATGTTGCAC
nad5-ex1-F2	TGTATCTACTTATCGTATTTTTGCCCC
nad5-3UTR-R18	GCTCCTCCAGTTCGATTATTATTCT
nad5_ex2_3R	CTGGCTCTCGGGAGTCTCTT
nad5_ex3_4R	AACTCGGATTCGGCAAGAA
nad7_ex1_2F	ACCTCAACATCCTGCTGCTC
nad7-ex5-R	TCCACCTCTCCAAACACAAT
rpl2-ex2-F1	TCACACAGTGAATAAGGGCTT
rpl2-ex1-R2601	TGAGACCAGGGAGAGCAAGA
rps3-ex1-F1	ATGGCACGAAAAGGAAATCCG
rps3-rev2	CGTTTCGGATATAGCACGTCTCC
nad2_noNGS_7F	CAGGAATACCCCCGTTAGCC
nad2_ex4_562R	ACGCTAGTCACTACTCCCACT
nad2b_qF	TATTTGTTCTTCGCCGCTTT
nad2_int4_ex4R	CTTATTCGTGGCAACCTTCC
nad2_int4_1518F	GAACCTGCCCGGAGTGAG
nad2_int4_1635R	CCCGAAACCAAGGGGATACT
nad2_ex5_noNGS_7R	AGAATCCATGTCCTAGGTGTATCA

nad2_noNGS_7F2	TGATACACCTAGGACATGGATTCT
nad2_ex5_143R	AGGAGAGGGGTATAGCAAGGA
nad2b-ex3-F9	TGATCTATGGGTCTACTGGAGCT
nad2_ex2_3F	AAAGGAACTGCAGTGATCTTGA
nad4_ex3_139F	GCACCATGCCGAATCTCTCT
nad4_ex3_289R	ATCCCAAGCGCTGCTAATGT
nad4_ex3_4F	TTCCTCCATAAATTCTCCGATT
nad4_int3_132R	GGGACTTATGCAACGGGCTA
nad4_int3_6F	GACCGTCCGTTGAACTACCA
nad4_int3_ex4F	TCTAGCTTGGTTCGGAGAGC
nad4_ex3_4R	TGAAATTTGCCATGTTGCAC
nad4_ex4_1F	CTCGTTCGGATGGGTGTTCA
nad7exon2F	GAGGGACTGAGAAATTAATAGAGTACA
nad7exon2_R	TCTGAACGATCAGAATAAGGTA
nad7_int2_55R	GGTCAGCACCTCCGAAAGAA
nad7_int2_ex3F	AGTGGGAGAGCCGTGTTATG
nad7_int_1028R	CGACAGCTTTTTCGTACACGT
nad7_int2_ex3F	AGTGGGAGAGCCGTGTTATG
nad7_ex2_3R	TGGTACCTCGCAATTCAAAA
nad7_ex3_170F	TTAACTCCGTTCCTGTGGGC
nad7_ex3_286R	ATCTTGTGCCACTCCACCTG
matR_qF	AATTTTTGCGAGAGCTGGAA
matR_qR	TTGAACCCCGTCCTGTAGAC
ACTIN(ACT1)_F	CGACAATGGAACTGGAATGGTTA
ACTIN(ACT1)_R	GTGCCTCGGTAAGTAGAATAGGA

Table 5. List of genes that are co-expressed with *PCIS1*.

A co-expression z-score indicates the significance of the co-expression patterns, where 0 is a random level, and 3 is a false positive rate of 0.1%. Scores, z>=8.0 and 8.0>z>=7.0 were adopted as co-expressed and moderately co-expressed genes, respectively.

	Gene ID	Gene annotation	Function	PPR protein class	ath-u.3 / AT5G25500 (z-score)
1	AT3G02330	MEF13 (ccmFc-415 cox3-314 nad2-59 nad4-158 nad5-1665 nad5-1916 nad7-213)	PPR protein	E+	10.8
2	AT3G58590		PPR protein	PLS	10.7
3	AT1G08070	OTP82 (ndhB-836 ndhG-50)	PPR protein	DYW	10.2
4	AT3G13880	OTP72 (rpl16-440)	PPR protein	E+	10.2
5	AT4G35130		PPR protein	DYW	10
6	AT3G58520		Ubiquitin carboxyl-terminal hydrolase family protein		10
7	AT5G61370		PPR protein	P	9.5
8	AT3G08820		PPR protein	DYW	9.5
9	AT1G13410		PPR protein	E+	9.3
10	AT4G38010	SLO4 (cob-982 nad4-1033)	PPR protein	E2	9.3
11	AT2G01510		PPR protein	DYW	9.3
12	AT3G25970		PPR protein	E+	9.3
13	AT5G52850		PPR protein	DYW	9.2
14	AT5G39350		PPR protein	E2	9.1
15	AT2G13600	SLO2 (nad4L-110 nad7-739 mttB-144 mttB-145)	PPR protein	E+	9

16	AT4G39530		PPR protein	E+	8.8
17	AT5G66500		PPR protein	E2	8.6
18	AT2G37310		PPR protein	E+	8.6
19	AT3G26782		PPR protein	DYW	8.6
20	AT1G20230		PPR protein	DYW	8.6
21	AT5G27110		PPR protein	E+	8.5
22	AT4G37170		PPR protein	DYW	8.4
23	AT3G11460	MEF10 (nad2-842)	PPR protein	DYW	8.3
24	AT1G25360	OTP90 (ccmB-80 ccmC-184 ccmF-1246 nad1-500 mttB-97)	PPR protein	DYW	8.3
25	AT2G29760	OTP81/QED1 (matK-706 ndhB-872 rpoB-2432 rps12-69553 accD-58642)	PPR protein	DYW	8.3
26	AT1G56690		PPR protein	DYW	8.2
27	AT4G33170		PPR protein	DYW	8.2
28	AT5G56310		PPR protein	E+	8.1
29	AT3G15140		Polynucleotidyl transferase: ribonuclease H-like superfamily protein		8.1
30	AT5G08310		PPR protein	P	8.1
31	AT4G02750		PPR protein	DYW	8
32	AT2G36730		PPR protein	E2	7.9
33	AT1G32415		PPR protein	E+	7.9
34	AT3G03580		PPR protein	DYW	7.8
35	AT3G49142		PPR protein	DYW	7.7
36	AT3G62890		PPR protein	DYW	7.7
37	AT4G18840		PPR protein	E2	7.7
38	AT5G49410		thiamine-phosphate synthase		7.7
39	AT4G21065		PPR protein	DYW	7.7
40	AT1G09680		PPR protein	P	7.7

41	AT2G40720		PPR protein	E+	7.7
42	AT2G29260		NAD(P)-binding Rossmann-fold superfamily protein		7.6
43	AT3G53360		PPR protein	E+	7.6
44	AT2G35030	COD1 (cox2-253 cox2-698 nad4-1129)	PPR protein	E+	7.6
45	AT3G18970	MEF20 (rps4-226)	PPR protein	E2	7.6
46	AT1G68930		PPR protein	DYW	7.5
47	AT5G63290		Radical SAM superfamily protein		7.5
48	AT3G53220		Thioredoxin superfamily protein		7.4
49	AT1G09190		PPR protein	E2	7.4
50	AT1G17630	CWM1 (ccmB-428 ccmC-463 nad5-598)	PPR protein	E+	7.3
51	AT1G33350		PPR protein	E+	7.2
52	AT2G33680		PPR protein	E+	7.2
53	AT1G06440		Ubiquitin carboxyl-terminal hydrolase family protein		7.2
54	AT4G14050	MEF35 (nad4-1373 rpl16-209)	PPR protein	E+	7.2
55	AT5G16420		PPR protein	P	7.2
56	AT5G62990	EMB1692	Ubiquitin carboxyl-terminal hydrolase family protein		7.1
57	AT4G04190		transmembrane protein		7.1
58	AT5G39680	EMB2744	PPR protein	P	7.1
59	AT5G09270		transmembrane protein		7.1
60	AT2G22070		PPR protein	DYW	7
61	AT4G14820		PPR protein	DYW	7

Table 6. List of genes and associated accession numbers

PCIS1	AT5G25500
ABO6	AT5G04895
ABO8	AT4G11690
CFM9	AT3G27550
MatR	ATM00520
mCSF1	AT4G31010
MISF74	AT4G01400
MTL1	AT5G64320
nMAT2	AT5G46920
OZ2	AT1G55040
PMH2	AT3G22320
SLO3	AT3G61360
Mal d 1-associated protein	AT2G35900
BIP1	AT5G28540
SDH4	AT2G46505
PGF13	AT4G23820