Plant-type RNA editing transferred into *Escherichia coli*: Expected and surprising new insights for editing factors and how they recognize their targets

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Yingying Yang

aus

Henan, China

Bonn, April, 2023

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. **Gutachter**: Prof. Dr. Volker Knoop
2. **Gutachter**: Prof. Dr. Peter Dormann

Tag der Promotion: 22.06.2023
Erscheinungsjahr: 2023

Table of Content

1. Introduction

In eukaryotic organisms, several post-transcriptional processes take place after a gene has been transcribed into pre-mRNA, before the mRNA is matured and ready to be translated into protein. These processes include decapping, tailing, and RNA splicing, which ensure that protein translation occurs correctly. One further important process is RNA editing, which will be the focus of this work.

1.1. RNA editing

RNA editing is a post-transcriptional process that is widely found in eukaryotes, prokaryotes, archaea, and viruses (Benne, 1994; Benne et al., 1986; Su and Randau, 2011). During this process, RNA molecules are modified in a variety of ways, including nucleotide insertion, deletion, and base substitutions (Knoop, 2011). In general, this evolutionarily conserved modification ensures that the mRNA is corrected before being translated into proteins. RNA editing has been found in mRNAs, tRNAs, rRNAs, and miRNAs, including untranslated regions, in the nucleus, mitochondria, and chloroplasts (Chateigner-Boutin and Small, 2010; Knoop, 2011; Su and Randau, 2011; Sun et al., 2016).

Base substitutions, such as adenosine to inosine deamination (e.g. in fungi (Teichert, 2018)) and pyrimidine changes, have been discovered not only to correct mRNA but also to generate different forms of proteins with and without editing, leading to downstream pathway discrepancies (Karcher and Bock, 2009; Su and Randau, 2011; Wagner et al., 1989). To state the most prominent example, in the human nuclear transcript of *apoB*, a C-to-U deamination exchange by RNA editing results in edited and unedited versions of the protein expressed in the liver and intestine, respectively. The corresponding key RNA editing factor, APOBEC-1, was later characterized as a zinc-dependent cytidine deaminase that recognizes nucleotides in cis upstream of the editing site (Mehta and Driscoll, 2002; Smith et al., 1997).

1.2. Plant-type RNA editing

The discovery of plant-type RNA editing was first published in 1989 by three independent research groups, reporting the RNA editing sites found in angiosperm mitochondrial transcripts (Covello and Gray, 1989; Gualberto et al., 1989; Hiesel et al., 1989). Two years later, the RNA editing site in maize chloroplast *rpl2* was reported, changing ACG (Thr) to a start codon (ATG, Met) (Hoch et al., 1991). To specify this type of RNA editing, which contains the single nucleotide substitution from C-to-U (forward editing) and U-to-C (reverse editing), the term plant-type RNA editing was defined (Figure 1). This is a site-specific cytidine deamination post-transcriptional process that is found only in the transcripts of mitochondria and chloroplasts (Blanc et al., 1995; Rajasekhar and Mulligan, 1993; Yu and Schuster, 1995).

Plant-type RNA editing can occur in various regions of RNA molecules, especially in coding sequences of messenger RNA (mRNA), but also in transfer RNAs (tRNA) (Binder et al., 1994; Kunzmann et al., 1998; Marchfelder et al., 1996; Marechal-Drouard et al., 1993), introns (Begu et al., 2011; Binder et al., 1992; Borner et al., 1995) and untranslated regions (Miyata et al., 2002). With the help of advanced bioinformatic tools and high-throughput RNA sequencing technologies, scientists can now identify and predict RNA editing sites more efficiently, thereby improving our understanding of the RNA-editing mechanism (Bentolila et al., 2013; Cheng et al., 2016; Lenz et al., 2018; Ruwe et al., 2013). Once the information encoded in the genome on the first and second position of a codon has been corrected, the resulting mature mRNA can be translated into a functional correct protein, maintaining its proper organelle function. The efficiency of RNA editing can vary depending on the editing site, with sites leading to no change in codon identity being often only partially edited, while sites that switch the codon to a different one are typically edited more efficiently (Edera et al., 2018; Miyata and Sugita, 2004; Small et al., 2020). Plant-type RNA editing has been shown to play important roles in a variety of biological processes in plants, including development, stress response, and energy metabolism (Li et al., 2021).

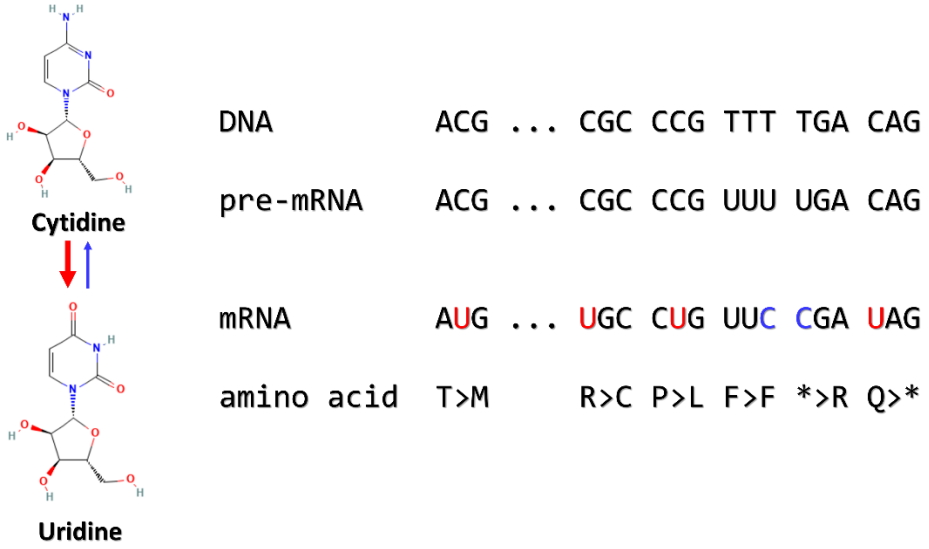


Figure 1. Plant-type RNA editing conversion between cytidine and uridine via deamination and amination during post-transcriptional processing. This type of editing, which can occur at the first and second codon positions, results in a switch in amino acid identity, while editing at the third codon position typically leads to a silent change. In organellar transcripts, both forward (dominate, red) and reverse (blue) RNA editing between cytidine and uridine are observed primarily during mRNA maturation. Additionally, this process can generate start codons (M) and stop codons (*). Chemical structure taken from PubChem https://pubchem.ncbi.nlm.nih.gov/.

1.2.1. Evolution of plant-type RNA editing

The C-to-U plant-type RNA editing mechanism has not been reported in algae. It is believed to have evolved in the common ancestor of land plants possibly as a defense against UV light exposure after transitioning from water to land (Fujii et al., 2011; Hollósy, 2002; Takahashi and Ohnishi, 2004). Forward RNA editing is found to be wildly distributed in all land plants except of the marchantiid liverworts (Figure 2) (Freyer, 1997; Groth-Malonek et al., 2007; O Malek, 1996; Rüdinger et al., 2012; Sper-Whitis et al., 1996; Steinhauser et al., 1999). The number of RNA editing sites can vary widely between species, ranging from tens to thousands. For example, the moss *Funaria hygrometria* has eight editing sites in mitochondria and two in chloroplast transcripts (Rüdinger et al., 2011a), while *Seleginella moellendorffii* contains 2152 editing sites in mitochondria (Hecht et al., 2011), and *Seleginella unicate* has 3415 sites in chloroplast transcripts (Oldenkott et al., 2014). Reverse

RNA editing, which converts uracil to cytidine, is found in hornworts, lycophytes (except Seleginellales), and monilophytes (Gerke et al., 2020; Grewe et al., 2011; Gutmann et al., 2020; Ichinose et al., 2022; Knie et al., 2016). According to the plant phylogeny model proposed by Qiu et al. (2006), reverse editing was gained in the common ancestor of hornworts and tracheophytes. However, recent research by Wickett et al. (2014), Lutzoni et al. (2018), Puttick et al. (2018) and Su et al. (2021) suggests that bryophytes form a monophyletic group in plant phylogeny. This indicates that both forward and reverse editing may have been gained at the origin of land plants (Knoop, 2022). Additionally, adenosine-to-inosine RNA editing in tRNA has been observed in plant organelles once (Delannoy et al., 2009; Karcher and Bock, 2002; Su and Randau, 2011). However, this work will focus solely on C-to-U RNA editing.

To clearly specify different RNA editing sites, Rüdinger et al. (2009) proposed a nomenclature system. The system labels editing sites with the gene name, the direction of editing (C-to-U: eU or U-to-C: eC), the position in the gene, and the resulting amino acid change. For example, the editing site nad4eU272SL represents a forward C-to-U editing site (eU) in the *NADH dehydrogenase subunit 4* (*nad4*) gene at nucleotide position 272 of the transcript, resulting in a change of codons from serine (S) to leucine (L).

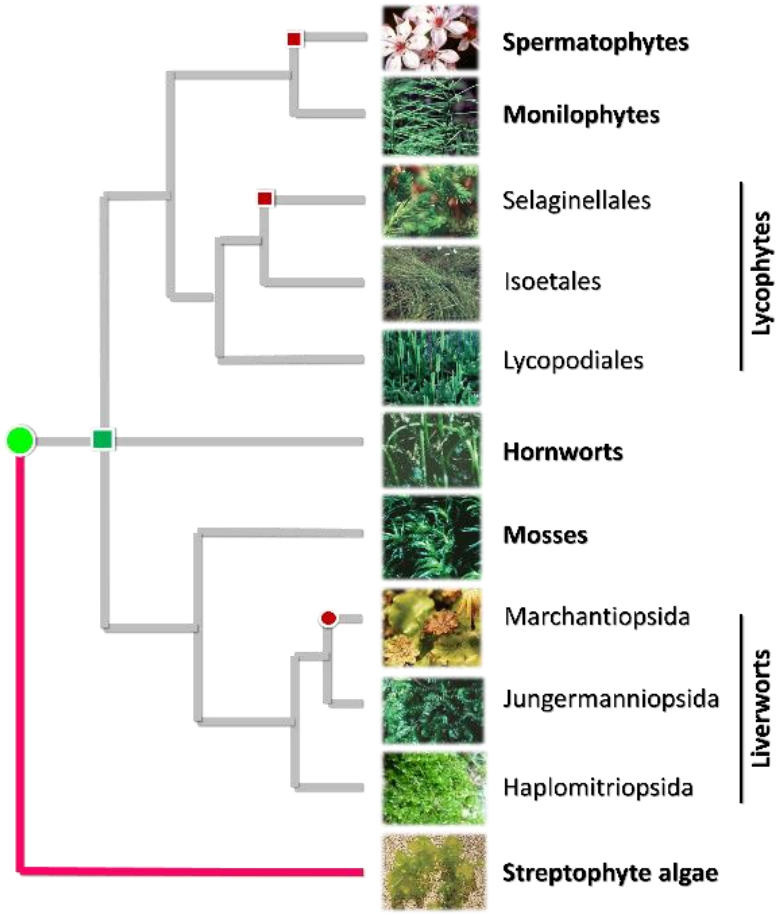


Figure 2. Cladogram depicts the gain and loss of plant-type RNA editing in land plants, based on a modified phylogeny from Knoop (2022). C-to-U RNA editing likely emerged (green dot) as a mechanism to defend against UV light exposure during the transition from water to land, and was lost (red dot) in Marchantiopsida. The phylogeny of bryophytes is still debated (Puttick et al., 2018; Su et al., 2021; Wickett et al., 2014), but U-to-C RNA editing appears to have arisen (green square) in the common ancestor of hornworts and tracheophytes, and was lost (red square) in Selaginellales and the common ancestor of seed plants. Plant names in bold indicate common clades, plant pictures from Simpson (2010).

1.3. Pentatricopeptide repeat (PPR) proteins

In the year 2000, the genome of the model plant *Arabidopsis thaliana* was fully sequenced, which marked a significant milestone in the discovery of pentatricopeptide repeat (PPR) proteins (Small and Peeters, 2000). The PPR protein family is one of the largest and most enigmatic gene families found in land plants,

characterized by tandem motifs (2-26) of 35-amino acid repeats. Unlike the tetratricopeptide repeat (TPR) motifs, which consist of 34 amino acids in each motif and are involved in protein-protein interactions, PPR proteins are considered key factors in some protein-RNA interactions (D'Andrea and Regan, 2003; Das et al., 1998; Karpenahalli et al., 2007; Small and Peeters, 2000). In the PPR protein, similar to TPR protein, two anti-parallel α-helices are linked by a loop forming the α-solenoid repeat. The tandem repeats of these motifs result in a superhelix encircling a tunnel with a hydrophilic central and positively charged bottom, providing the ability to interact with negatively charged RNA molecules (Small et al., 2023; Small and Peeters, 2000). PPR proteins are divided into different subgroups regarding their structure. While the P subclass PPR proteins consist only a stretch of 2-25 35 amino acid sequence motifs, participating in mitochondria or chloroplasts for organelle RNA processing, including RNA splicing, cleavage, RNA stability, transcription, and translation (Barkan and Small, 2014; Rugen et al., 2019; Small et al., 2023; Waltz et al., 2019), the PLS subclass PPR protein is built with canonic P motifs combined with longer (L, 35-36 amino acids) and shorter (S, 31 amino acids) variants (Barkan et al., 2012; Kindgren et al., 2015; Lurin et al., 2004). The PLS proteins usually show P, L, S repeats as triplets with the most C-terminal ones slightly differing in sequence conservation and therefore labeled as P2, L2, S2 (Cheng et al., 2016; Lurin et al., 2004). After the PLS stretch, most PLS proteins additionally a carboxy (C)-terminal domain, including the extension domains E1, E2, and the DYW domain, named after the three terminal amino acids aspartate (D), tyrosine (Y), and tryptophan (W) (Figure 3) (Cheng et al., 2016). The DYW-type PPR proteins are defined as the key factors of plant-type RNA editing, due to the zinc-binding cytidine deaminase function holds by the DYW domain (Cheng et al., 2016; Hayes and Santibanez, 2020; Oldenkott et al., 2019; Schallenberg-Rüdinger and Knoop, 2016; Small et al., 2020).

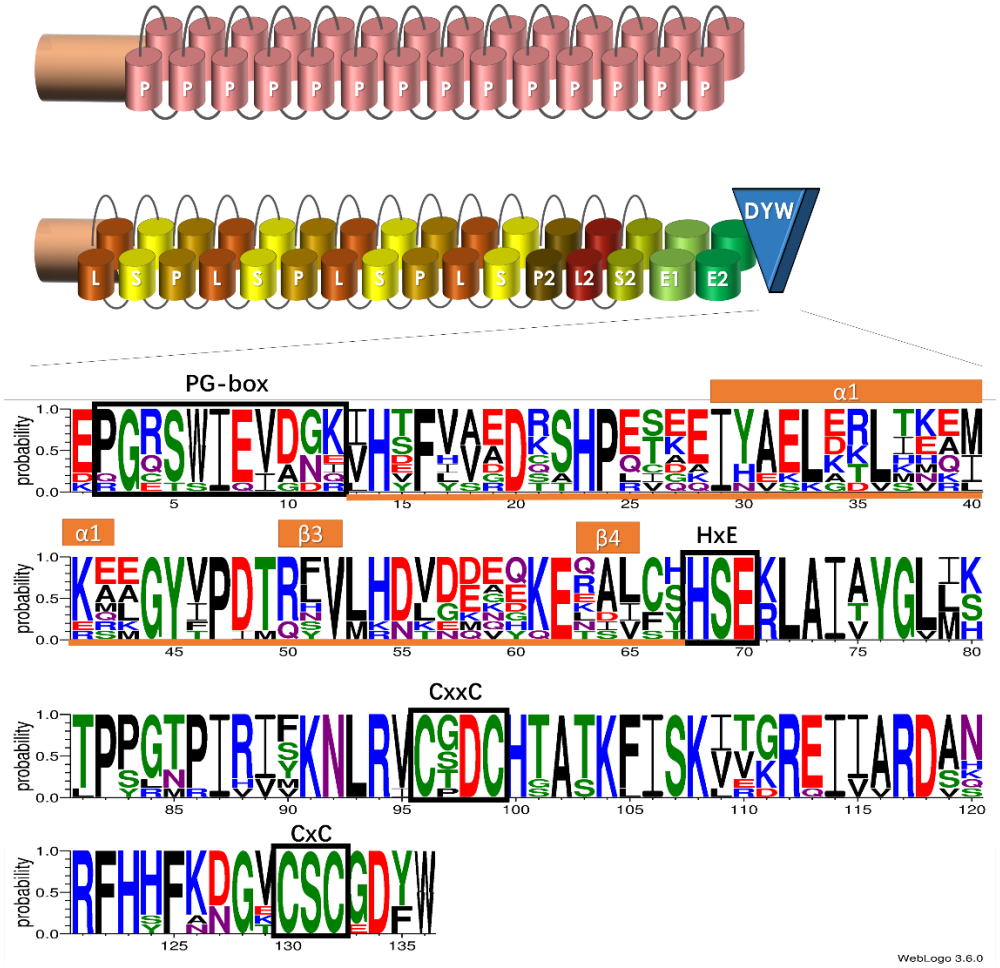


Figure 3. A model of a DYW-type PPR protein and its key motifs within the DYW domain. The DYW-type PLS proteins typically begin with an organellar signal peptide (light orange) in the N-terminal region, followed by several PLS repeats that terminate with P2, L2, and S2 motifs (dark brown, dark red, and dark yellow, respectively), with amino acid variations compared to P1, L1, and S1 motifs (brown, red, and yellow). The C-terminal E1 and E2 domains (green and dark green) are similar in structure to TPR repeats and consist of two alpha-helices structures each. The DYW domain is colored blue. Key motifs within the DYW domains are indicated in boxes. The PG-box is crucial for maintaining editing site conservation, while the HxEx$_n$CxxC domain binds to a zinc ion and is involved in cytidine deamination. The CxC domain captures the second zinc ion. Additionally, the orange lines denote the gating domain, which plays a pivotal role in opening the catalytic center for the precise positioning of the nucleotide to be edited. (Takenaka et al., 2021). Alpha-helixes and beta-sheets are indicated in orange boxes on top. Weblogo generated (https://weblogo.threeplusone.com/create.cgi) based on the DYW domains from the nine *P. patens* editing factors.

The C-terminal extensions E1 and E2 domains in the PLS subclasses of PPR proteins share a similar structure with TPR proteins. While TPR proteins are primarily involved in protein-protein interactions (Schallenberg-Rüdinger and Knoop, 2016), the exact role of E1 and E2 domains in PLS proteins is not well-defined. However, the RNA footprint of *A. thaliana* CRR2 suggests that these domains may also be involved in RNA recognition (Ruwe et al., 2018). CRR2 is a DYW-type PPR protein that does not participate in RNA editing. Rather, its function is to stabilize transcripts of chloroplast *rps7* and *ndhB*, thereby playing a crucial role in maintaining their integrity (Hashimoto et al., 2003). Despite this evidence, the precise function of the E1 and E2 domains in PLS proteins remains unclear. The function of the DYW domain, in contrast, is better known. The DYW domains stores the so-called "PG-box" motif, located at the beginning of the DYW domain, which is suggested to be required for editing site conservation (Hayes et al., 2013; Okuda et al., 2007). The $HxEx_nCxxC$ motif is the zinc-binding cytidine deaminase signature, and the CxC motif close to the DYW end captures a second zinc ion (Boussardon et al., 2014; Hayes et al., 2013; Schallenberg-Rüdinger and Knoop, 2016; Wagoner et al., 2015). These two zinc-binding sites were confirmed capturing Zn^{2+} ions after the DYW domain was successfully crystalized (Takenaka et al., 2021). With the crystallization structure, Takenake's group defined a so-called "gating domain", contains one α helix (α1) and two β sheets (β3 and β4) (Figure 3). The gating domain undergoes a conformational change upon binding to RNA, resulting in the opening of the catalytic center of the PPR protein, which in turn enables efficient cytidine deamination of a specific C nucleotide (Small et al., 2023; Takenaka et al., 2021).

1.3.1. PPR-RNA recognition code

Numerous knockout and knockdown studies of PPR protein genes in the model plants *Arabidopsis thaliana* and *Physcomitrium patens* have let to assign PPR proteins to their RNA targets (Fujii et al., 2013; Hammani et al., 2009; Rüdinger et al., 2011a; Schallenberg-Rüdinger et al., 2013a; Takenaka et al., 2019). From the variable loss of function studies, it was found that PPR proteins can serve more than one RNA target. By utilizing bioinformatic tools and statistical analysis in

combination with the PPR proteins assigned to their targets, researchers found that the fifth and the last amino acid of the PPR motifs, especially the P and S motifs, are important for RNA recognition, where the fifth amino acid distinguish between purine and pyrimidine, and the last amino acid assigned keto or amino group of ribonucleotieds (Barkan et al., 2012). The targets assigned by the same editing factor PPR proteins show high similarities (Glass et al., 2015; Hammani et al., 2009; Rüdinger et al., 2011a). Subsequently, a PPR-RNA recognition code was proposed (Barkan and Small, 2014; Barkan et al., 2012; Gerke et al., 2020), which follows the one-repeat-one-nucleotide pattern until the nucleotide -4 of the editing site (Figure 4). Besides the fifth and last amino acids in the PLS motif, other studies suggest that the second amino acid also helps recognizing RNA targets (Yagi et al., 2013). The crystal structure analyses of designed P subclass PPR proteins and RNA electrophoretic mobility shift assay (REMSA) studies confirmed and expanded the proposed binding code (Shen et al., 2016; Yan et al., 2019).

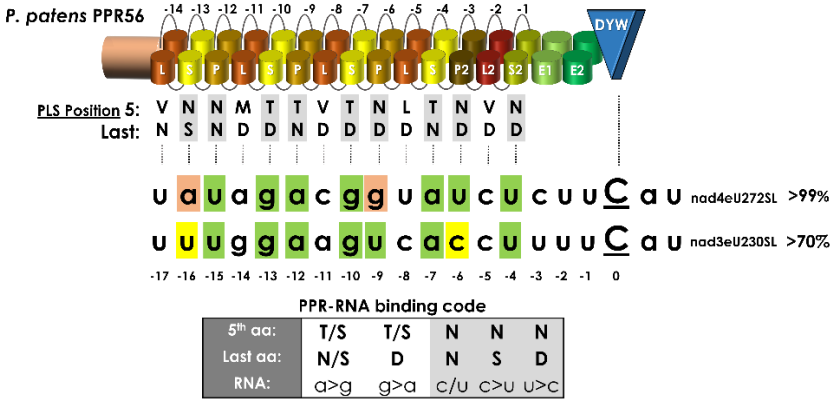


Figure 4. The PPR-RNA binding model of DYW-type *Physcomitrium patens* PPR56 is shown with its two targets nad4eU272SL and nad3eU230SL. Protein structure is labelled as in Figure 3. PPR56, which is a mitochondrial editing factor, contains 14 PLS repeats (numbering backward) and is responsible to edit nad4eU272SL (>99%) and nad3eU230SL (>70%). The letters on the PLS repeats indicate the fifth and last amino acids that are important for RNA recognition. The amino acid combinations from P- and S-repeats, which are thought to be important for target selection, are shaded. The function of L-repeats are still controversial. The PLS repeats bind to the RNA target in a one-repeat-one nucleotide pattern, with the editing cytidine marked in bold and underlined. The PPR-RNA binding code is based on Barkan et al. (2012). PPR model modified from Oldenkott (2020).

The P- and S-type PPR repeats in the PLS subclass PPR proteins generally follow the suggested PPR-RNA recognition code (Barkan and Small, 2014; Gerke et al., 2020). In contrast, the L-type PPR repeats show less conservation in the second, fifth, and last amino acids (Gutmann et al., 2020; Yagi et al., 2013). Although it was previously thought that the L-type PPR repeats do not play a role in PPR-RNA recognition, but rather act as spacers or provide conformational help when the P- and S-type PPR repeats bind to RNA targets (Yan et al., 2017b), *in vivo* studies suggest that at least some of the L motifs are still important for RNA recognition (Oldenkott et al., 2019).

1.3.2. Evolution of DYW-type PPR proteins

While it remains unclear why PLS-type PPR proteins are mainly restricted to the land plant clade, phylogenetic studies have revealed an expansion of the PLS-type PPR protein family in embryophytes (Fujii et al., 2011; Gutmann et al., 2020; O'Toole et al., 2008; Salone et al., 2007). Transcriptional data is now available for further studies through the OneKP project (Carpenter et al., 2019; Leebens-Mack et al., 2019). Given that PLS-type PPR proteins are involved in RNA editing, it is not surprising to see that the number of these proteins is linked to the frequency of RNA editing (Gutmann et al., 2020). Usually, loss of an RNA editing site (which involves replacing cytidine with thymidine on DNA level) leads to the loss of the corresponding PPR editing factor, unless this protein is able to serve additional sites or gets a neofunctionalization, such as PPR43 from *P. patens* (Hein and Knoop, 2018; Hein et al., 2016; Hein et al., 2019; Rüdinger et al., 2011a).

In the basal land plant linkage bryophyte, although E-type PPR proteins without DYW domain was found, plant-type RNA editing is mainly carried out by DYW-type PPR proteins that possess the complete C-terminal extension (E1, E2, DYW domains).. In the model moss *Physcomitrium patens*, at least 107 genes encoding for PPR proteins have been discovered, with only 10 of them possessing C-terminal extensions (Sugita, 2022). Although PLS-type PPR proteins lacking the C-terminal extensions have been identified, they have not been found to be involved in the editing mechanism directly. Nine of the DYW-type PPR proteins are assigned to one up to two of the in total 13 editing sites (Ichinose et al., 2014; Ichinose et al., 2013;

Ohtani et al., 2010; Rüdinger et al., 2011a; Schallenberg-Rüdinger et al., 2013a; Tasaki et al., 2010; Uchida et al., 2011) while the tenth one, PPR43, has a degenerated DYW domain and have an RNA splicing function (Ichinose et al., 2012; Rüdinger et al., 2011a). The remaining PLS-type PPR proteins are found to be involved in RNA splicing, RNA stabilization, RNA accumulation, and other post-transcriptional processes (Sugita, 2022). PPR proteins in *P. patens* are “intron-rich”, indicating that the PPR gene family underwent retrotransposition-mediated proliferation in the more ancient past (O'Toole et al., 2008). At least 18 pairs of paralogous PPR proteins have been discovered, which is likely the result of genome duplication activity during the expansion of the gene family (Rensing et al., 2008).

In angiosperms, DYW type PPR proteins have evolved into various ways (Hayes et al., 2013; Kotera et al., 2005; Okuda et al., 2009; Okuda et al., 2010; Zehrmann et al., 2011). While the some DYW-type PPR protein, such as MEF1, performs RNA editing by itself as seen for *P. patens*, storing a complete functional DYW domain (Zehrmann et al., 2009; Zehrmann et al., 2011), others utilize the "brock" mechanism, where truncated PPR proteins mainly consisting of PLS repeats are involved in target recognition, and DYW-like proteins provide deamination function in angiosperms (Small et al., 2023). Truncated PLS-type PPR proteins, such as CRR4, can complete the cytidine deamination function on ndhDeU2TM with the help of the DYW-like protein DYW1, a short protein with only 239 amino acids, that holds only an N-terminal truncated DYW domain (Boussardon et al., 2012). Truncation experiments with some DYW-type PPR proteins showed that after C-terminal truncation they were still able to edit their assigned sites, such as truncated MEF11, in the complementation study (Verbitskiy et al., 2010). However, PPR proteins such as RARE1 and QED1 were unable to edit their corresponding sites when the DYW domain was truncated (Wagoner et al., 2015). All these examples already reveal the diversity of DYW domain evolution in *A. thaliana*.

1.3.3. The more complicate editing complex: Editosome

In angiosperms, particularly in the extensively researched model plant maize and *A. thaliana*, various types of PLS PPR proteins are involved in RNA editing mechanisms.

It is frequently observed that truncated PLS proteins collaborate, such as CRR4 with DYW1, as mentioned above. MEF8 and MEF8S, with only five PLS repeats, are other examples that require helper proteins to complement the cytidine deamination function (Boussardon et al., 2012; Chateigner-Boutin et al., 2013; Diaz et al., 2017). Similarly, the DYW-type PPR protein DYW2, which lacks E1 and E2 domains, and the P-type PPR protein NUWA, which contains a coiled-coil domain on the C-terminal, are co-factors of the editing factor CLB19 (Andres-Colas et al., 2017; Guillaumot et al., 2017). Unlike DYW1, which only pairs with CRR4, DYW2 and MEF8 have been found to interact with several E-type PPR proteins. DYW2 has been shown to be dual-targeted in both mitochondria and chloroplasts, and is involved in over 100 RNA editing sites along with NUWA (Guillaumot et al., 2017). In the later complementation study, CWM1, an E-type PPR protein from *A. thaliana* responsible for editing 4 sites, was found to be a likely interaction partner of DYW2 as well (Oldenkott et al., 2020). Similarly, the DYW-like PPR protein PCW1 has been found to be responsible for editing 102 sites in maize mitochondria (Wang et al., 2022).

Apart from the cooperation between PPR proteins, other co-factors are involved in RNA editing in angiosperms (Figure 5) (Sun et al., 2016; Yan et al., 2017a). For example, cp31, a chloroplast ribonucleoprotein (cpRNP) containing RNA recognition motifs (RRMs), was found to be crucial for editing at two sites in *ndhB* and *psbL* (Hirose and Sugiura, 2001; Tillich et al., 2009). It has been suggested that this protein may indirectly influence RNA editing by dissolving RNA secondary structures (Schallenberg-Rüdinger and Knoop, 2016). Organelle RNA recognition motif proteins (ORRMs) are another group of co-factors found in both mitochondria and chloroplasts, influencing up to 44% (ORRM4) of editing sites (Shi et al., 2016; Sun et al., 2016). The multiple organellar RNA editing factors (MORFs) (Takenaka et al., 2012) or RNA editing factor interacting proteins (RIPs) (Bentolila et al., 2012) shows direct interaction with PPR proteins. Ten MORF/RIP proteins have been found in *A. thaliana*, and half of them were found as major editing factors (Sun et al., 2016). Mutations in these individual MORF/RIP proteins lead to a reduction or loss of editing at different sites (Bentolila et al., 2013; Bentolila et al., 2012; Takenaka et al., 2012). Yeast-two-hybrid experiments indicate direct interaction

between different MORF/RIP proteins and with PPR proteins (Glass et al., 2015; Härtel et al., 2013; Zehrmann et al., 2015). Organelle zinc-finger (OZ) protein OZ1 is another co-factor found in the editosome that interacts with ORRM1, PLS-type PPR proteins, P-type PPR proteins, and protoporphyrinogen IX oxidase 1 (PPO1) in chloroplasts is involved in up to 18 editing sites (Sun et al., 2016; Sun et al., 2015; Yan et al., 2017a; Zhang et al., 2014). In 2017, it was reported that the chloroplast RNA helicase ISE2 (increased size exclusion limit 2) is involved in RNA editing (Bobik et al., 2017). Given that the RNA targets may need to maintain a certain secondary structure for proper protein interaction (Schallenberg-Rüdinger et al., 2013a), this is not too surprising.

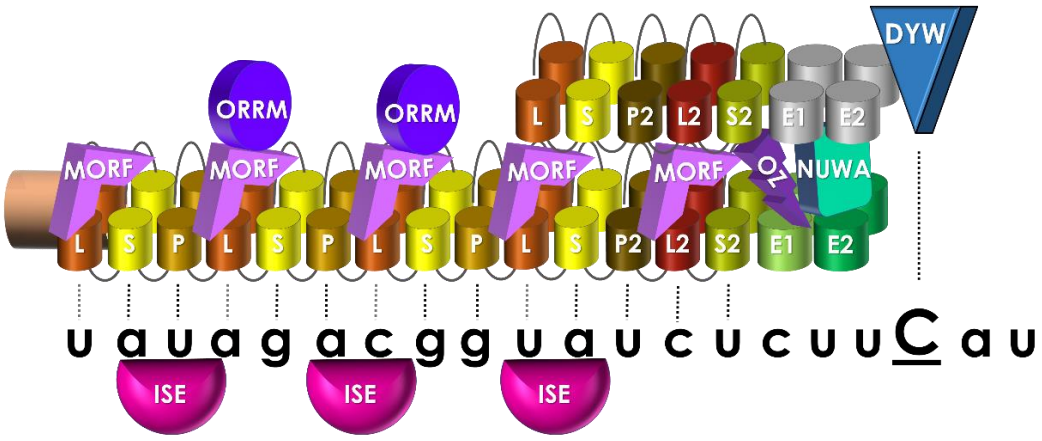


Figure 5. A hypothetical model of "editosome" in many angiosperm RNA editing mechanism. Unlike in the basal plant *P. patens*, RNA editing in angiosperms has been found to be more complex. DYW-type PPR proteins can be "broken" into two pieces to co-work on one editing site (such as CRR4 and DYW1), or the protein with a DYW domain may be involved in editing multiple sites (e.g. DYW2). Other proteins, such as MORF/RIP, ORRM, OZ, ISE, and P-type PPR protein NUWA have been found to interact with PLS-type PPR proteins and assemble into a different and dynamically transient protein complex called the "editosome" for RNA editing, as describe in text. Note that ORRM only interacts with MORF/RIP, but not with the PLS-type PPR proteins. Further investigation is needed to fully understand the molecular interactions involved.

Since the E1 and E2 domains in the C-terminal extensions exhibit high similarity with the TPR repeat structure, it is plausible that they may participate in protein-protein interactions similar to TPR proteins (Cheng et al., 2016; Schallenberg-Rüdinger and Knoop, 2016). Additionally, the presence of the conserved "PG-box"

at the beginning of the DYW domains suggests that this region may also play a crucial role in protein-protein interactions (Hayes et al., 2013).

Although the term "editosome" has been used to describe the editing complex made up of different proteins (Figure 5), it has not yet been successfully purified in the lab (Bentolila et al., 2012; Huang et al., 2019; Sandoval et al., 2019). Since the same or related proteins could also be involved in other RNA maturation processes, it seems that the "editosome" is likely a weak and transient interaction for editing events, rather than a stable and well-defined complex (Small et al., 2023).

1.3.4. DYW-type PPR proteins outside of land-plants

Although plant-type RNA editing was once thought to be restricted to land plants, DYW-type PPR proteins have been discovered in aquatic plants. *Nitella hyaline*, a streptophytes algae, harbors several PPR proteins and one DYW domain. In addition, the chlorophyte algae *Tetraselmis cordiformis*, a flagellated microalga, possesses DYW-like motifs and a PPR-like protein with a DYW domain (Gutmann et al., 2020). DYW-like sequences containing the key cytidine deaminase signature $HxEx_nCxxC$ was found in algae Pyraminoadales, Dolichomastigales, Prasinococcales and Chlorodendrales (Gutmann et al., 2020). DYW-type PPR proteins are also present in a variety of other organisms, including the euglenozoa *Diplonema* (Moreira et al., 2016), the slime mold *Physarum polycephalum* (Schaap et al., 2015; Schallenberg-Rüdinger et al., 2013b), and the heterolobosean protists *Acrasis kona* (Fu et al., 2014) and *Naegleria gruberi* (Rüdinger et al., 2011b). Considering that these organisms have close contact with land plants and are capable of incorporating foreign DNA, horizontal gene transfer may be a plausible explanation (Cazalet et al., 2010; Choudhary et al., 2007; Fu et al., 2014; Knoop and Rüdinger, 2010).

In *Physarum polycephalum*, 4 C-to-U editing events were found together with other type of editing, including nucleotide insertion, deletion, and conversion of C-to-G and U-to-G (Bundschuh et al., 2011). In the protists *Acrasis kona* and *Naegleria gruberi*, C-to-U RNA editing sites (atp6eU722SL and cobeU409HY in *A. kona*, cox1eU1120HY and cox3eU787RW in *N. gruberi*) were identified in their mitochondrial transcriptomes (Fu et al., 2014; Rüdinger et al., 2011b). In *N. gruberi*,

ten DYW-type PPR proteins were found encoded in the genome. Ten DYW-type PPR proteins were also identified in the *N. gruberi* genome, and although not all of them possess mitochondrial signal peptides, some of them stores a long PPR repeat regions, and the DYW domains of most of them show high conservation in key motifs, including the PG box, the cytidine deaminase signature HSEK, CxxC and CxC, which could provide clues to assign the editing sites to the corresponding PPR protein (Knoop and Rüdinger, 2010; Rüdinger et al., 2011b).

1.4. The model organisms *Physcomitrium patens* and *Escherichia coli*

1.4.1. Physcomitrium patens

In the basal branch of land plants (Figure 2), the moss *Physcomitrium patens*, from the family Funariaceae in the order Funariales, has a significant advantage as other bryophytes (liverworts, mosses, hornworts) in that its haploid gametophyte dominates its life cycle, making phenotypic studies more manageable (Beike et al., 2015; Beike et al., 2014; McDaniel et al., 2010; Medina et al., 2019; Rensing et al., 2020). Its nuclear genome and its two organellar genomes were sequenced and made available from 2008 (Lang et al., 2018; Rüdinger et al., 2008; Sugiura et al., 2003; Terasawa et al., 2007). Stable (Schaefer and Zrÿd, 1997) and transient (Bezanilla et al., 2003) transformation methods have been available since 1997, including traditional partial bombardment (Cho et al., 1999), PEG-mediated protoplast transformation (Schaefer and Zrÿd, 1997) and *Agrobacterium tumefaciens* T-DNA insertion (Cove et al., 2009). The nuclear genome is amenable to mutational, knockout, knockdown, and overexpression studies due to its high rates of homologous recombination (Kamisugi et al., 2005; Sugita, 2022). Other molecular tools such as PREPACT have been established in the *P. patens* as well (Bezanilla et al., 2003; Collonnier et al., 2017). Additionally, gene targeting in the chloroplast genome was feasible as early as 2004 (Sugiura and Sugita, 2004). With these various technologies, *P. patens* has become a suitable non-seed model plant for studying abiotic stress and molecular-based research (Rensing et al., 2020).

Out of the 107 PPR proteins found in *P. patens*, only 16 belong to the PLS subclass, and ten of those contain a DYW domain. The five PLS-type PPR proteins (PPR9, PPR25, PPR31, PPR34 and PPR69) lacking complete C-terminal extensions have not been found to be involved in RNA editing (Sugita, 2022). PPR105 which storing a E-domain was not found to be involved in editing function as well. Instead, PPR9 has been suggested as a co-factor for PPR43 in intron splicing of *cox1* (Ichinose et al., 2020), while PPR31 is responsible for intron splicing of *atp9* and *nad5* (Ichinose et al., 2020). PPR69 has been identified as a homolog of PDM1/SEL1, a RNA splicing and accessing factor in *A. thaliana* (Sugita, 2022; Zhang et al., 2015). The remaining PLS-type PPR proteins are not yet well characterized (Sugita, 2022). Of the ten DYW-type PPR proteins in *P. patens*, PPR43 is the only one not assigned to RNA editing function. Its degenerate DYW domain is instead implicated in intron 3 splicing of *cox1* (Ichinose et al., 2012).

After combining the genome and transcriptome data, 13 C-to-U RNA editing sites were found in *P. patens* (Figure 6), 11 in mitochondrial and 2 in chloroplast transcripts (Ichinose et al., 2014; Ichinose et al., 2013; Ohtani et al., 2010; Rüdinger et al., 2011a; Schallenberg-Rüdinger et al., 2013a; Tasaki et al., 2010). Knockout studies were conducted to investigate the role of DYW-type PPR proteins in *P. patens*. The results revealed that the lines ΔPPR45 and ΔPPR98 could not survive, therefore knockdown studies were performed, showing that PPR45 edits two sites in *rps14* and PPR98 is responsible for atp9eU92SL (Ichinose et al., 2014; Ichinose et al., 2013; Schallenberg-Rüdinger et al., 2013a). KO PPR65 (Schallenberg-Rüdinger et al., 2013a), KO PPR71 (Tasaki and Sugita, 2010; Tasaki et al., 2010), KO PPR77 and KO PPR91 (Ohtani et al., 2010) showed a strong phenotype characterized by very small colony size. Further analysis showed that PPR65 needs to edit the ccmFC103PS site before PPR71 can edit the ccmFC112SF site (Schallenberg-Rüdinger et al., 2013a). In contrast, KO PPR56 (Ohtani et al., 2010), KO PPR78 (Rüdinger et al., 2011a), and KO PPR79 (Uchida et al., 2011) did not exhibit a strong phenotype. All of these 13 sites were assigned to 9 DYW-type PPR proteins, with one or two targets each (Figure 6). In contrast to the complicated RNA editing mechanisms found in angiosperms, *P. patens* shows an ancient RNA editing system,

where only DYW-type PPR proteins are sufficient for the cytidine deamination activity, without any helper proteins (Oldenkott et al., 2019; Sugita, 2022). As the first organism in which all of its editing sites and the factors involved have been fully characterized, *P. patens* has become a favorite plant for studying RNA editing mechanisms.

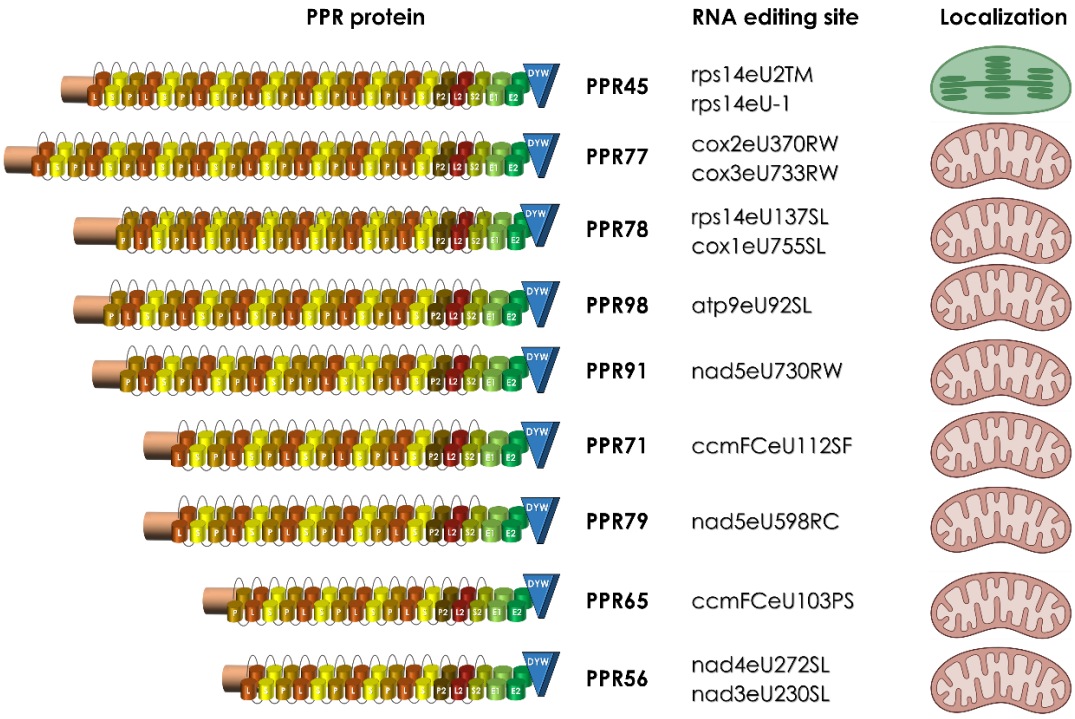


Figure 6. Nine DYW-type PPR proteins as RNA editing factors assigned to their corresponding editing sites in *P. patens*. Eight of the PPR proteins edit their targets in mitochondrial transcripts, while PPR45 is assigned to chloroplast targets in *rps14*. Mitochondria and chloroplast figure from Bio-render (https://www.biorender.com/).

1.4.2. Escherichia coli

Escherichia coli, a gram-negative bacterium, is a widely-used model organism in the field of biology. Since the first successful plasmid transformation in *E. coli*, numerous biotechnologies have been developed for this fast-growing organism, making it a valuable tool for many different applications (Becker and Wittmann, 2016; Cai et al., 2016; Chen et al., 2013; Lee et al., 2013; Rodriguez et al., 2014; Sauer et al., 2008; Zhao et al., 2016). While *Physcomitrium patens* displays a

relatively simple RNA editing mechanism, a stable mitochondrial transformation method has been established for this plant (Takenaka et al., 2013). This limitation has made *E. coli* a promising alternative for RNA editing studies. While *E. coli* does not perform pyrimidine-type RNA editing in general, no PLS-type PPR proteins were found in *E. coli*. A-to-I RNA editing has been found in the ACG anticodon of tRNA, and the corresponding factor, tadA, belongs to the adenosine deaminase family and is not linked to the PPR protein family (Wolf et al., 2002). This makes it possible to establish a plant-type RNA editing system in *E. coli* without interference from plant-specific factors.

PPR65 and PPR56 from *P. patens*, along with their corresponding targets, were successfully transferred into *E. coli* (Figure 7), achieving comparable editing as *in planta* (Oldenkott et al., 2019). This study demonstrated that a single DYW-type PPR protein is sufficient to complete C-to-U RNA editing, with the DYW domain acting as the cytidine deaminase. The first target switch could be successfully achieved in the plant-type RNA editing mechanism. Taking advantage of the heterologous system, the *E. coli* transcriptome provided a large number of off-targets of PPR65 (6) and PPR56 (79) besides the targets we provided (on-target), highlighting key positions for PPR protein recognition. The PPR-RNA code was largely confirmed as well. Based on this study, a chimera of PPR56 with the DYW domains of OTP86 from *A. thaliana* can efficiently edit the provided *nad4* targets. Key motifs within the DYW domain of OTP86 were also tested and found to behave similarly to the DYW domain of moss (Takenaka et al., 2021). With a similar system, synthetic PLS-type PPR proteins with complete C-terminal extensions were shown to edit their targets with the help of MORF2 (Royan et al., 2021) in bacteria as well; the active PPR65 protein from *P. patens* was purified and demonstrated to retain its cytidine deamination function *in vivo* (Hayes and Santibanez, 2020).

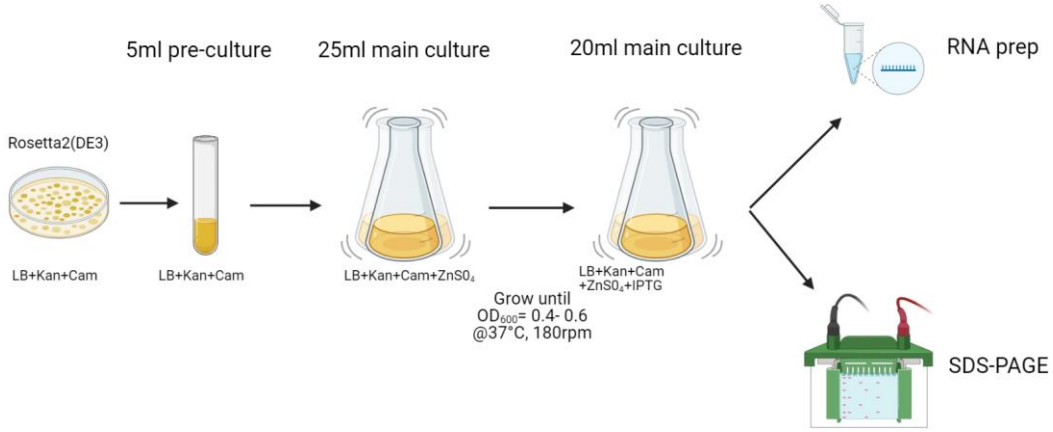


Figure 7. Expression of RNA editing factors and their targets in *E. coli*. The pET41Kmod vector containing the coding sequence for the protein, along with the provided target sequence, was transformed into *E. coli* strain Rosetta2(DE3) on LB agar plates supplemented with kanamycin (Kan, 50 µg/ml) and chloramphenicol (Cam, 5 µg/ml). After selecting a suitable clone, a 5 ml pre-culture was prepared and grown overnight at 37°C and 180 rpm. Subsequently, 250 µl of the pre-culture was transferred to a 25 ml main culture supplemented with additional 0.4 mM $ZnSO_4$. Once the OD600 reached 0.4-0.6, the main culture was adjusted to a volume of 20 ml and induced with 0.4 mM of IPTG to activate the T7 promoter. The culture was then grown for 20 hours at 16°C and 180 rpm. The bacteria were harvested for RNA preparation, followed by cDNA synthesis to assess editing efficiency, while protein expression was checked by SDS-PAGE.. Figure based on Oldenkott et al. (2019) and created by using bio-render.

This study is aimed to investigate the plant-type RNA editing system by focusing mainly on PPR56. Target and protein modifications at various positions were tested, and several chimeras were examined to explore the role of different motifs. Additional data sets of off-targets in the *E. coli* system were also analyzed to evaluate the plant-type RNA editing system as a potential molecular tool.

2. Additional materials and methods

All materials and methods used for production of the results displayed in this thesis are described in the mentioned publications (Yang et al., 2023a; Yang et al., 2023b; Yang et al., 2023c). From the cloning of the constructs until the sequencing of the results for editing efficiency, it was followed the same procedure as described in Yang et al. (2023a). Additional materials and methods are described below.

2.1. Protein solubility test

To confirm that the lack of editing in the constructs was not due to insoluble proteins caused by protein mutations, soluble proteins were extracted from a 20-hour *E. coli* culture (2 mL) based on the method described by Takenaka et al. (2021). For western blot analysis, a 1:500 dilution of 6x-His Tag Monoclonal Antibody (HIS.H8, Invitrogen) and a 1:5000 dilution of Rabbit anti-Mouse IgG (H+L) Secondary Antibody, AP (Invitrogen) were used. Signals were detected using SIGMAFAST™ Fast Red TR/Naphthol AS-MX Tablets (Sigma-Aldrich).

2.2. Additional oligonucleotides

Additional oligonucleotides are listed in the Supplementary Table 1.

3. Results

3.1. Establishing protein solubility test for PPR proteins expressed in *E. coli* system

In Oldenkott et al. (2019), PPR65 and PPR56 from *P. patens* was successfully expressed and edit their targets in *E. coli* system (Figure 7). The approval of how many soluble protein was still questing. In Takenaka et al. (2021), the amount of soluble chimera PpPPR56 with DYW domain of AtOTP86 was tested by western blot. It was then wondered how many soluble protein of PPR65 and PPR56 shown in Oldenkott et al. (2019) was expressed. The SDS-PAGE gel shows that PPR65 was always highly over-expressed comparing with PPR56 in *E. coli* system, although PPR56 edit its target more efficiently, even with much more off-targets (133) in *E. coli* transcriptome comparing with 6 off-targets for PPR65 (Yang et al., 2023a).

Soluble total proteins were extracted based on Takenaka et al. (2021). The proteins from $4.5*10^8$ *E. coli* cells were then applied to western blot to fish out PPR proteins fused with MBP (see section 2.1). With different dilution of the first antibody, the 1:1500 dilution shows the clearest soluble PPR proteins for both PPR56 and PPR65. Surprisingly, the amount of soluble PPR proteins are correlated with the total over-expressed protein on SDS-PAGE gel (**Figure 8**).

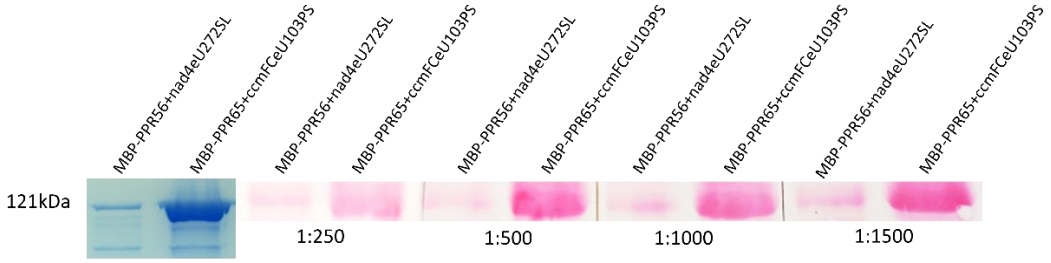


Figure 8. A western blot analysis to test the amount of soluble protein in PPR65 and PPR56 expressed in the *E. coli* system. The left side (blue bands) shows the SDS-PAGE gel with blue bands representing PPR56 and PPR65 proteins. The right side (red bands) shows the soluble proteins with different dilutions (1:250, 1:500, 1:1000, 1:1500) of anti-His antibody.

3.2. Yang et al 2023a (submitted)

Yingying Yang, Kira Ritzen hofen, Jessica Otrzonsek, Mareike Schallenberg-Rüdinger and Volker Knoop (2023)

Beyond a PPR-RNA recognition code : Many aspects matter for the multi-targeting properties of RA editing factor PPR56

Plant C-to-U RNA editing is a cytidine deamination process which occurs in transcripts of both mitochondria and chloroplasts allowing plants to correct the information encoded in their organellar DNAs before it is translated into protein. RNA-binding pentatricopeptide repeat (PPR) proteins are most prominent as specific RNA editing factors and form the largest nuclear encoded gene family in plants. Adapting plant-type RNA editing could offer a useful new tool for targeted transcript alteration in diverse systems in the future. Towards that end, it is important to understand its mechanism in detail. RNA editing factors of *Physcomitrella patens* have successfully been transferred into *Escherichia coli*, including PPR56 with its two native targets nad3eU230SL and nad4eU272SL. To elucidate the important parameters for efficient C-to-U editing, we tested how (i) protein mutations (ii) target mutations, (iii) diverse target sequence extensions, (iv) secondary structures, as well as (v) diverse placements of the two targets in different combinations in PPR56 would affect editing efficiency. Accompanying *E. coli* transcriptome analyses provides additional insights from off-target identification.

All amino acid modifications within the DYW domain related to the 2nd zinc binding site (PPR56|DYW:H123A, H123Y, C130A, and C132A) (Figure 3) completely abolished editing in both nad4 and nad3 targets. Editing was also abolished in the 1st zinc binding amino acid mutation (K91A), but K71R after the "HSE" motif remained edited in both targets. PPR56|DYW:G3A had less of an effect on editing efficiency than P2A, similar to the conserved motif "SHP" in moss with mutations of H23A and P24A. The key motif suggested to be important for DYW chimera compatibility has been proven to be important, as MH79IS mutations abolished

editing in both targets, but H80K did not significantly affect editing.

Mutations on the PLS motifs aim to redirect the PPR protein to new targets according to the PPR-RNA binding code. PPR56|S-13NS>ND, P-12NN>NS, S-7TD>TN, P2-3ND>NS, L-2VD>ND, and S2-1ND>NS could not complement editing with mutations on the corresponding target sites. Meanwhile, PPR56|S-10TD>TN, P-9TN>TD, P-6ND>TD, S-4TN>TD, P2-3ND>NN, and S2-1ND>TD showed higher editing in at least one target when using the mutated target in the corresponding site. Notably, although S-7TD>TN could not redirect to a new target, the double mutation S-7TD>TN|S-4TN>TD could edit both targets with mutations on target -10 and -7 sites. Given that the N-terminal part of the PLS stretch usually contributes less to target recognition and binding, truncations on the N-terminal part (first PLS triplet) surprisingly greatly reduced editing on both targets.

Different single nucleotide modifications cover all nucleotides corresponding to the PLS motifs (together with the results of Oldenkott et al. 2019). In most cases, the *nad4* target is more tolerant of modifications than the *nad3* target. This holds true for multi-nucleotide modifications as well. Although L motifs do not contribute to target binding, it is surprising to see the reduction of editing in target mutations corresponding to L-2 and L-8. Interestingly, switching nucleotides corresponding to the L motif in triple mutations of *nad3* (g-14a|a-11c|c-8u) results in higher editing than the opposite switching in the *nad4* target (76% vs. 26%).

Nucleotides around editing sites (position -1 to -3, +1 to +2) has been modified as well. Start codon and stop codons could be generated with high editing of *nad4* target. Furthermore, the sequences upstream of the PPR binding site were also mutated. Stepwise shortening progressively reduced RNA editing efficiencies, even though the expected core PPR-binding region of the target remained unaffected. Although RNA secondary structure could be one of the explanations, the current structure prediction tools could not explain this yet. To investigate further, artificial strong hairpin structures were tested on the *nad4* target, and indeed, they abolished editing.

E. coli, as a heterologous system, provides the opportunity to check for off-targets

in its transcriptome data. Off-targets of native PPR56 have been investigated in Oldenkott et al. 2019, where 79 off-targets were found. With more data accumulated and updates from the analyzing tools, now 133 off-targets have been found. The off-targets highlighted the key nucleotides that would be important for target selection of PPR56: S-10TD:g, P-9TN:a, S-7TD:g, S-4TN:a, P2-3ND:u, and S2-1ND:u. Interestingly, with one amino acid mutation, PPR56|S-4TN>TD has 449 off-targets identified, while the S-10TD>TN mutants only have 16. Although a strong shift in preference of nucleotides at the corresponding positions has been observed, it is surprising that a single amino acid can influence the protein's flexibility so strongly.

Knowing the upstream sequence of the PPR binding part could influence editing. A series of tandem targets have been tested in the *E. coli* system, and the upstream target can significantly enhance the editing of the downstream target. This enhancement does not solely rely on the editing of the upstream target. This holds true when cloning off-target sequences in the vector system as well. Additionally, targets have been placed upstream of the PPR56 coding sequence (5') and downstream (3') for comparison. Interestingly, the target in the 3' position can enhance editing at the 5' site in most cases. Using this target enhancement capability, a potential target, cox3eU290SF, has been found for PPR56. This site is pre-edited with a thymine in DNA already and is conserved in its pre-edited state in all available moss mitochondrial DNA. However, it has been confirmed as an editing site in the lycophytes *Isoetes engelmannii*, *Selaginella moellendorffii*, and in the fern *Haplopteris ensiformis*.

Part of the study was included in my master's thesis (Yang, 2019), while other parts were included in the bachelor's theses of Kira Ritzenhofen (Ritzenhofen, 2021) and Jessica Otrzonsek (Otrzonsek, 2020). I cloned most of the constructs and analyzed their editing efficiencies. I also analyzed the RNA-seq data for PPR56 and its mutants and mainly participated in the further downstream analysis of the off-target candidate sites.

Beyond a PPR-RNA recognition code: Many aspects matter for the multi-targeting properties of RNA editing factor PPR56

Yingying Yang, Kira Ritzenhofen, Jessica Otrzonsek, Mareike Schallenberg-Rüdinger* and Volker Knoop*

IZMB – Institut für Zelluläre und Molekulare Botanik, Abteilung Molekulare Evolution, Universität Bonn, Kirschallee 1, D-53115 Bonn, Germany

*Corresponding authors
mareike.ruedinger@uni-bonn.de, Phone: +49 228 73-6464
volker.knoop@uni-bonn.de, Phone: +49 228 73-6466
Fax: +49 228 73-6467

ORCID-IDs
Yingying Yang: 0000-0002-5518-0307
Kira Ritzenhofen: 0009-0008-1608-3871
Jessica Otrzonsek: 0009-0003-5218-8105
Mareike Schallenberg-Rüdinger: 0000-0002-6874-4722
Volker Knoop: 0000-0002-8485-9423

Key words: *Physcomitrium patens,* mitochondrial RNA editing factor, PLS-type pentatricopeptide repeats, DYW-type cytidine deaminase, PPR-RNA recognition code, *Escherichia coli*

Abstract

The mitochondrial C-to-U RNA editing factor PPR56 of the moss *Physcomitrium patens* is an RNA-binding pentatricopeptide repeat protein equipped with a terminal DYW-type cytidine deaminase domain. Transferred into *Escherichia coli,* PPR56 works faithfully on its two native RNA editing targets, nad3eU230SL and nad4eU272SL, and also converts cytidines into uridines at over 100 off-targets in the bacterial transcriptome. Accordingly, PPR56 is attractive for detailed mechanistic studies in the heterologous bacterial setup, allowing for scoring differential RNA editing activities of many target and protein variants in reasonable time. Here, we report (i) on the effects of numerous individual and combined PPR56 protein and target modifications, (ii) on the spectrum of off-target C-to-U editing in the bacterial background transcriptome for PPR56 and two variants engineered for target re-direction and (iii) on combinations of targets in tandem or separately at the 5'- and 3'-ends of large mRNAs. The latter experimentation finds enhancement of RNA editing at weak targets in many cases, including cox3eU290SF as a new candidate mitogenome target. We conclude that C-to-U RNA editing can be much enhanced by transcript features also outside the region ultimately targeted by PPRs of a plant editing factor, possibly facilitated by its enrichment or scanning along transcripts.

Introduction

The recent years have seen much progress towards understanding the molecular machinery behind cytidine-to-uridine RNA editing in plant chloroplasts and mitochondria [1–4]. The research on RNA editing and other processes of RNA maturation in the two endosymbiotic organelles of plant cells has clearly profited from parallel approaches taken not only with model flowering plants like *Arabidopsis*, maize or rice but also with bryophyte model organisms [5]. Flowering plants (angiosperms) feature complex RNA editosomes variably composed of numerous and diversely interacting proteins to target specific sites for C-to-U conversion in the organelle transcriptomes [3,6–8]. In contrast, a much simpler scenario has emerged for C-to-U RNA editing in "early-branching" land plants among which the moss *Physcomitrium patens* holds a key role as a model organism [4,9,10]. All characterized RNA editing factors in *Physcomitrium* combine a stretch of pentatricopeptide repeats (PPRs) responsible for sequence-specific RNA recognition with a terminal DYW-type cytidine deaminase carrying out the site-specific C-to-U conversion.

To a large part, the complex editosomes of angiosperms seem to be the result of frequent separation of RNA target recognition and the catalytic DYW domain, now relying on protein-protein interaction including various helper proteins interacting *in trans* [11–19]. This evolutionary pathway is exemplified with the recently investigated case of angiosperm RNA editing factor CWM1 that is C-terminally truncated in *Arabidopsis* and relies on helper proteins but features an orthologue with a terminal DYW domain in the early-branching flowering plant *Macadamia* that was able to complement an RNA editing KO in *Physcomitrium* [20]. Single editing factors retaining those functionalities in just one polypeptide, as in the case of the here investigated PPR56, mainly exist in early-arising plant lineages like the mosses [3].

Physcomitrium patens has a prominent role with its only 13 C-to-U RNA editing sites assigned to nine site-specific RNA editing factors. However, *Physcomitrium* is in no way representative for other bryophytes, which feature the full spectrum of RNA editing being entirely absent in the marchantiid liverworts, with massive C-to-U RNA editing in the early-branching moss *Takakia lepidozioides* [21] or

with abundant "reverse" U-to-C RNA editing co-existing with C-to-U editing in hornworts like *Anthoceros agrestis* [22]. Among altogether more than 100 pentatricopeptide repeat proteins in *Physcomitrium* only nine are RNA editing factors and all of them, including PPR56 investigated here, are characterized by a PLS-type PPR array linked to a terminal DYW cytidine deaminase domain via the E1 and E2 domains [9]. It is likely no surprise that the simple one-protein RNA editing setup of *Physcomitrium* could be functionally transferred into heterologous systems like the bacterium *Escherichia coli* [23] and, more recently, also into human cell lines [24]. The bacterial setup in particular offers an easy access to exploring the interaction of an RNA editing factor and its targets by allowing the investigation of numerous protein and target variants in short time.

The mitochondrial RNA editing factor PPR56 of *Physcomitrium patens* has been functionally characterized some years ago [25] and appeared particularly suited for further investigations for several reasons. Firstly, it has two native mitochondrial target sites that are converted with different efficiencies by specific cytidine deamination in the moss (Fig. 1A). Editing target nad4eU272SL is converted to more than 99% in the steady state mitochondrial transcriptome of *Physcomitrium*. Editing efficiency at its second target, nad3eU230SL, is more variable and may depend on environmental conditions but is generally above 70% *in planta* [25,26]. The RNA editing target site labels follow a nomenclature proposal that indicates the respective genetic locus (here *nad* subunits of respiratory chain complex 1, the NADH ubiquinone oxidoreductase), the RNA editing event towards uridine (eU), the transcript position counting from the first nucleotide of the AUG start codon and the resulting codon change, here serine to leucine in both cases [26,27].

Defining a PPR-RNA recognition code has been a tremendous step forward in understanding the operation of pentatricopeptide repeat proteins [28–31]. At the core of this code, the identities of the 5th and the last (L) amino acid within the two antiparallel α-helices constituting an individual PPR are key to recognizing individual ribonucleotides with position '5' distinguishing purines (adenosines or guanosines) from pyrimidines (cytidines or uridines) and position 'L' defining preferences for amino (A or C) or keto nucleobases (G or U). However, the situation is notably more complex for PPR

proteins acting as RNA editing factors, which not only feature canonical 'P-type' PPRs of 35 amino acids but also variants with different consensus profiles and slightly variable lengths. Most widely distributed are the variants 'L' (long, 35-36 aa) and 'S' (short, 31-32 aa) contributing to PLS-type PPR arrays in most plant RNA editing factors. Yet more PPR variants such as 'SS' and 'LL' have recently been identified in the growing amount of genomic data for the huge PPR gene families in land plants, now also including hornworts, lycophytes and ferns [32].

The PPR-RNA code outlined above can be applied only to P- and S-type but not to L-type PPRs and the functional role of the latter remained mysterious. Notably, despite a conceptually slightly better overall fit of the nad3eU230SL target to the P- and S-type PPRs of PPR56 (Fig. 1A), the nad4eU272SL target is edited more efficiently not only in the native moss background but also in the recently established heterologous *E. coli* RNA editing assay system [23]. Hence, additional parameters beyond the conceptual matches of an array of PPRs to its targets evidently contribute to RNA editing efficiencies.

Here, we explored the impact of PPR56 protein mutations and of modified, extended, combined and differently placed RNA targets in the easily amenable bacterial system to identify the relevant elements contributing to efficient RNA editing. Most importantly, we found that sequences further upstream of the region ultimately bound by the PPR array contribute to high RNA editing efficiency and that tandem combinations of target sequences can significantly enhance RNA editing at previously less efficiently edited downstream targets. The latter include both selected off-targets in the *E. coli* transcriptome as well as cox3eU290SF as a predicted further candidate plant mitogenomic target of PPR56.

Moreover, we observed that placing the otherwise moderately edited nad3eU230SL target of PPR56 in the 5'- vs. the 3'-UTR of a long mRNA can enhance RNA editing even above the level observed in its native plant mitochondrial environment. Hence, the wider environment of the core RNA target sequence as defined by the PPR array contributes notably to the observed RNA editing efficiencies. Altogether, we conclude that the operation of PLS-type RNA editing factors like PPR56

112 relies not only on the defined code for P- and S-type PPRs but also on the hitherto enigmatic L-type
113 PPRs and on the wider transcript environment possibly favoring its enrichment in the neighborhood
114 of its ultimate target or even suggesting a 5'-to-3'-scanning mechanism towards the cytidine finally
115 targeted for deamination.

Results

PPR56, mutant nomenclature and the vector assay systems

PPR56 is a typical "complete", and likely evolutionarily ancestral, plant C-to-U RNA editing factor equipped with a highly conserved carboxyterminal DYW-type cytidine deaminase domain linked to an upstream PLS-type PPR array via the E1 and E2 extension motifs (Fig. 1A). For clarity, we here introduce nomenclature standards to label mutations on the protein or on the target side, respectively, that have been introduced for studying RNA editing functionality. For mutations on the protein side, we use a protein domain label behind a pipe symbol, followed by a colon and the position and amino acid identities in single-letter annotation before and after changes, e.g. PPR56|DYW:G3A for the mutation converting the glycine of the conserved PG box (Fig. 1B and suppl. fig. 1) into alanine. As a shorthand notation for mutations targeting the crucial positions '5' and 'L' of a given PPR, we simply indicate the introduced identities without numbering, e.g. PPR56|P-6ND>TD for the mutation converting the native ND combination in PPR P-6 for a conceptually better match to the guanidine that is naturally present in position -9 upstream of the nad4eU272SL editing site (Fig. 1A).

For mutations on the RNA target side, we will use small letters to label nucleotide changes and indicate positions relative to the editing site, which are added behind the respective RNA editing site labels after pipe symbols. For example, nad4eU272SL|u-4g will indicate the U-to-G exchange introduced four nucleotides upstream of the RNA editing site, which is assumed to be juxtaposed with the terminal S2-type PPR of PPR56 (Fig. 1A).

We mainly used the previously established heterologous expression system in *Escherichia coli* based on vector pET41Kmod [23]. The coding sequence of PPR56 is cloned in fusion with an upstream His_6-MBP tag behind an IPTG-inducible T7 promoter controlled by the lac operator and the respective target sequences are inserted in the 3'-UTR followed by a T7 terminator sequence. For further experimentation allowing to place target sequences alternatively also in the 5'-UTR, we

equipped pET41Kmod with an additional MCS upstream of the protein coding sequence, giving rise to pET41Kmod2 (Suppl. Fig. 2).

Mutating the DYW domain

Mutations had previously been introduced into the DYW domain of PPR65, another *Physcomitrium patens* RNA editing factor, to confirm the crucial role of conserved amino acids residues, including the ligands of a Zn^{2+} ion in the catalytic center of the cytidine deaminase [23]. Here, we have focused on other evolutionarily conserved positions in the DYW cytidine deaminase domain of PPR56 (Suppl. Fig. 1). Introducing mutations into the DYW domain of PPR56 (Fig. 1B) has the advantage that effects can be tested on its two native targets in parallel as opposed to only one target in the case of PPR65. The new set of mutants now also addresses a second Zn-binding site at the C-terminus of the DYW domain suggested to play a structural role outside of the catalytic center [33–35]. All mutations eliminating the relevant histidine or cysteine residues for coordination of the second zinc (PPR56|DYW:H123A, H123Y, C130A and C132A) indeed fully abolished detectable RNA editing on both targets (Fig. 1B).

Other mutations further upstream in the DYW domain, however, had surprisingly differential effects on the two targets of PPR56 with a generally much stronger impact on the less efficiently edited *nad3* target, which turned out to be generally more sensitive also upon other alterations (see below). Replacing proline with alanine in the eponymous PG box at the N-terminus of the DYW domain (PPR56|DYW:P2A) has a much stronger effect than the corresponding replacement of the following glycine residue (G3A), despite 100% conservation of the latter in all nine *Physcomitrium* RNA editing factors (Suppl. Fig. 1). Similarly, despite universal conservation of a downstream HP dipeptide motif in all *Physcomitrum* RNA editing factors (Suppl. Fig. 1), the corresponding mutations PPR56|DYW:H23A and P24A show significant remaining RNA editing activity with the exception of H23A on the *nad3* target (Fig. 1B). The position directly following the glutamate E70 in the catalytic center is conserved as either lysine or arginine in the DYW domains of RNA editing factors (Suppl. Fig. 1). However, exchanging lysine against arginine in that position (PPR56|DYW:K71R) results in

167 significantly reduced RNA editing of 79% at the *nad4* and of only 19% at the *nad3* target, respectively
168 (Fig. 1B). Notably, the reverse exchange (PPR65|DYW:R71K) had similarly led to reduced editing
169 efficiency for PPR65 [23], indicating that the respective identity of the basic amino acid in this
170 position is more important than could be expected.

171 We also addressed a variable region in the DYW domain that was previously postulated to
172 confer compatibility for creation of editing factor chimeras [36]. Exchanging the MH dipeptide to IS
173 (MH79IS) abolished editing activity completely whereas the single amino acid exchange (H80K) had
174 no negative, but even a slightly enhancing effect on the nad3eU230SL target (Fig. 1B). The
175 crystallization study of the DYW domain of OTP86, a chloroplast RNA editing factor of *Arabidopsis*
176 *thaliana,* suggested a regulation mechanism for DYW-type cytidine deaminases and defined a "gating
177 domain" blocking the catalytic site in an inactive state [33]. We tested the function of the corresponding
178 region in PPR56 by changing a conserved hydrophobic residue in its center into a positively charged lysine
179 (V36K), which abolished editing of the *nad3* target completely and reduced editing of the *nad4* target to
180 58% (Fig. 1B). The lysine in position 91 was found to mediate the accessibility of the catalytically
181 important E70 of the OTP86 DYW cytidine deaminase and exchanging the K in this position in PPR56 to A
182 (K91A) abolishes editing activity on both targets altogether (Fig. 1B).

183 ## Mutations in target positions juxtaposed with P- and S-type PPRs

184 To explore the different efficiencies of RNA editing at the two native targets of PPR56, we first
185 extended the set of mutations in target positions juxtaposed with the P- and S-type PPRs that are
186 assumed to follow the known PPR-RNA code rules (Fig. 2). Only one target mutation had previously
187 been found to enhance RNA editing at the *nad3* target: nad3eU230SL|c-6u, which improves the
188 conceptual fit to PPR P-3ND, hence fitting expectations. In the majority of mutants, we observe that
189 effects are much stronger for the nad3eU230SL than for the nad4eU272SL target (Fig. 2). Examples
190 are nad4eU272SL|u-4c (63%) vs. nad3eU230SL|u-4c (0%), nad4eU272SL|a-7g (20%) vs.
191 nad3eU230SL|a-7g (0%), nad4eU272SL|g-10a (27%) vs. nad3eU230SL|g-13a (0%), nad4eU272SL|g-
192 13a (35%) vs. nad3eU230SL|g-13a (0%) and, most dramatically for nad4eU272SL|u-15c (>99%) vs.

193 nad3eU230SL|u-15c (0%). The latter case is particularly surprising given that (i) N-terminal PPRs
194 generally play minor roles, (ii) PPR P-12NN is not expected to discriminate between U and C and (iii)
195 both natural targets have a uridine in that position. Exchanging conceptually perfect matches to PPRs
196 P-9TN and S2-1ND through mutations a-12g or u-4g abolishes RNA editing at both targets alike, again
197 fitting expectations (Fig. 2). Combining deleterious mutations g-13a and g-10a target abolishes
198 editing not only at the *nad3* target but also at the *nad4* target completely, indicating an additive
199 effect (Fig. 2). Changing the positions where the two targets differ opposite of P- or S-type PPRs to
200 the respective other nucleotide identities reduced RNA editing in both cases, to 54% for
201 nad4eU272SL|a-16u|g-9u|u-6c and to 49% for nad3eU230SL|u-16a|u-9g|c-6u, respectively.

202 ## Mutants in the PPR array

203 We tested whether target sequence mutations could be compensated by protein mutations in the
204 corresponding PPRs (Fig. 3). This was not the case for nad4eU272SL|u-4c, edited to 63% by
205 unmodified PPR56 (Fig. 2), but to only 30% by the conceptually adapted version PPR56|S2-1ND>NS
206 (Fig. 3A). Moreover, target variant nad3eU230SL|u-4c was neither edited by PPR56 (Fig. 2) nor by
207 PPR56|S2-1ND>NS (Fig. 3A). Unmodified targets nad4eU272SL and nad3eU230SL were still edited to
208 78% and 27% by the modified PPR56, respectively. Notably, canonical positions 5 and L in the
209 terminal S2-1 PPR matching with the corresponding position -4 as in PPR56 are more of an exception
210 than the rule for plant RNA editing factors.

211 For five other mutations in specific PPRs (S-13NS>ND, P-12 NN>NS, S-7TD>TN, P2-3ND>NS and
212 L2-2VD>ND, respectively), we found that RNA editing of the native targets was likewise significantly
213 decreased (with the exception of PPR56|P-12NN>NS on the nad4 target) and could not be rescued by
214 corresponding mutations in either target (Fig. 3A). This is most prominently seen for S-7TD>TN
215 abolishing RNA editing altogether and which could not rescue the corresponding mutation g-10a (Fig.
216 3A). Other mutations in the P2-L2-S2 triplet, again, had generally stronger effects on the *nad3* target.
217 Adapting P2-3 for a conceptually better match to cytidine by a ND>NS change did not improve
218 editing of any target (Fig. 3A). The changes introduced in the C-terminal P2-L2-S2 PPR triplet also

included L2-2VD>ND leading to a drastic drop in RNA editing through this single amino exchange in an L-type PPR, which would be expected to have increased preference for pyrimidines in P- and S-type PPRs (Fig. 3A). Most surprising, however, was the outcome of mutating the most N-terminal S-type PPR S-13NS>ND, which abolished RNA editing completely at both targets despite the mismatching adenosine in that position in the *nad4* target. Introducing the conceptually fitting uridine in position -16 did not restore editing (Fig. 3A).

Several other mutations in P- and S-type PPRs (S-10TD>TN, P-9TN>TD, P-6ND>TD, S-4TN>TD, P2-3ND>NN and S2-1ND>TD) had moderate consequences or could be rescued to a significant amount by corresponding changes in the targets (Fig. 3B). The S-10TD>TN and the corresponding target mutant g-13a fits the general insight of an overall more resilient *nad4* target with reduced editing of the original target (31%) and higher editing of the adapted one (g-13a, 63%), while editing of the original *nad3* target and in the nad3eU230SL|g-13a mutant is abolished completely. The inverse mutation in the directly neighboring PPR P-9TN>TD again has only moderate effects on the *nad4* target (Fig. 3B). However, and very surprisingly, this mutant can only be rescued by the corresponding a-12g mutation in the *nad3* but not in the *nad4* target. Somewhat similar is the outcome for the PPR P-6ND>TD mutant.

Given the striking outcome of completely abolished RNA editing for the S-7TD>TN mutant that could not even be partially rescued by the corresponding g>a exchanges in the two targets (Fig. 3A), we combined this mutation with the successful inverted exchange in S-4TN>TD (Fig. 3B) in a double mutant (Fig. 3C). Very surprisingly, this double mutant PPR56|S-7TD>TN|S-4TN>TD was able to edit both correspondingly adapted targets nad3eU230SL|g-10a|a-7g to 15% and nad4272SL|g-10a|a-7g to even 72%, indicating that the S-7TD>TN mutation does not cause a principally dysfunctional PPR56.

Overall, RNA editing factors characteristically show less conservation at the 5'-end of their PLS-type PPR arrays. However, the single amino acid mutation in PPR S-13NS>ND surprisingly abolished

RNA editing and could not be rescued on the target side (Fig. 3A). Effects were more moderate for mutating PPR P-12NN>NS. However, the original targets were still edited with higher efficiencies than the conceptually adapted ones with cytidines instead of uridines opposite to P-12NN>NS (Fig. 3A). To further address this, we created two progressive N-terminal truncations of PPR56 (Fig. 3D), either deleting PPR L-14 and the conceptually mismatching PPR S-13NS alone or a truncation including the following PPR P-12NN. For the shorter truncation RNA editing was abolished completely for the *nad3* target but only reduced to 89% for the generally more robust *nad4* target (Fig. 3D). This result may be explained by the moderately better fit of S-13NS to the cytidine in the *nad3* vs. the adenine in the *nad4* target. The further truncation including PPR P-12 further reduced RNA editing strongly at the *nad4* target (Fig. 3D).

The role of L-type PPRs

L-type PPRs only rarely feature amino acids in positions 5 and L that follow the PPR-RNA code rules. Notably, the two targets of PPR56 differ in the nucleotide identities opposite of its three central L-type PPRs L-11MD (a vs. g) , L-8VD (c vs. a) and L-5LD (u vs. c). Hence, we mutated these positions to check whether they could contribute to the different RNA editing efficiencies observed for nad4eU272SL and nad3eU230SL (Fig. 4). In a series of mutations adapting nucleotide identities to the respective other target, we find that changes in positions -14 (g<>a) and -8 (c<>u) do not significantly affect RNA editing in either target. Changes in position -11 (c<>a) decrease editing more significantly, however, and this is also the case after introducing a guanosine nucleotide in that position, eradicating editing for the *nad3* target altogether. Similar observations can be made for position -5 where the two native targets share a cytidine and the *nad3* target again proves to me more sensitive to changes. Notably, the corresponding triple-mutations converting positions -14, -11 and -8 to the identities in the respective other target decrease editing at the *nad4* target significantly to 26% and slightly improve editing at the *nad3* target to 76% (Fig. 4).

The immediate environment of the editing sites

The general avoidance of a guanosine in position -1 immediately upstream of a cytidine to be edited has been recognized since long and is unequivocally supported by large editome data sets [37]. Moreover, there is increasing evidence that the E1, E2 and the DYW domains downstream of the PPR arrays can contribute to target recognition selectivity [36,38]. Accordingly, we also targeted positions in the immediate environment of the respective RNA editing sites for mutations (Fig. 5). Exchanging the uridines in position -1 against guanosine indeed abolishes RNA editing altogether at both native targets of PPR56 (Fig. 5). For other positions, the *nad3* target is again more affected, even by identical nucleotide exchanges in the same positions as in the *nad4* target. For example, this is clearly seen for target mutations both immediately downstream of the respective edits, i.e. nad4eU272SL|a+1u (>99%) vs. nad3eU230SL|a+1u (49%) and nad4eU272SL|u+2g (>99%) vs. nad3eU230SL|u+2g (61%) as well as upstream of the respective edits: nad4eU272SL|c-3u (>99%) vs. nad3eU230SL|u-3c (22%) or nad4eU272SL|u-2g (31%) vs. nad3eU230SL|u-2g (0%).

We tested for the possibility to artificially create stop or start codons through C-to-U editing, focusing on the *nad4* target that had proven to be significantly more tolerant against variations. Indeed, all three possible stop codons (UAA, UAG, UGA) could be efficiently created by editing after mutations in positions +1 and/or +2 with >99% editing efficiencies (Fig. 5). Moreover, a combined nucleotide exchange in positions -1 and +1 (nad4eU272SL|u-1a|a+1g) also allows for artificial creation of a start codon by C-to-U editing quite efficiently (82%).

RNA secondary structures inhibit, but native sequences further upstream enhance RNA editing

The binding of an RNA editing factor can certainly be expected to compete with RNA secondary structure formation by base pairing. Target point mutations were routinely tested for potential secondary structure formations to exclude this as a potential cause for observed editing deficiencies [23]. We now intentionally created artificial secondary structures embedding the unchanged nad4eU272SL sequence targeted by PPR56 with upstream or with downstream

sequences creating base-pairings with the core PPR target region (Suppl. Fig. 3). An artificially added sequence upstream of the nad4eU272SL editing site potentially creating eight base pairs with positions -8 to -1 upstream of the cytidine editing left RNA editing efficiency unaffected whereas an extended regions creating 13 base pairs reduced RNA editing activity to only 19% (Suppl. Fig. 3). In contrast, RNA editing was abolished completely when artificial sequences were added behind position +5 relative to the cytidine editing target when creating potential base pairings with positions -10 to +1 or even only -8 to +1, respectively (Suppl. Fig. 3).

Establishing the RNA editing setup in *E. coli*, the PPR56 targets were cloned to include 17 additional nucleotides of the native sequence further upstream of the sequence that is ultimately expected to be targeted by the PPR array [23]. We now tested whether these additional 5'-sequences had an effect on RNA editing efficiencies and found significant effects, indeed (Fig. 6). Stepwise shortening the native target sequences at their 5'-ends progressively reduced RNA editing efficiencies considerably even though this would leave the expected core PPR-binding region of the target unaffected. Replacing the AU-rich region upstream of position -20 by a GC-rich sequence even abolished RNA editing at the nad3eU230SL target altogether (Fig. 6). These results suggested that native sequences beyond the target ultimately bound by the PPR array may contribute to enrich PPR proteins in the neighborhood of the target or possibly even a 5'-to-3' sliding of the protein on the mRNA towards its ultimate binding position for C-to-U conversion.

C-to-U RNA editing off-targets in the *E. coli* transcriptome

An initial screening of the *E. coli* transcriptome upon expression of PPR56 had identified 79 C-to-U RNA editing off-targets using strict criteria and confirmation from initially two independent RNA-seq replicates [23]. However, further candidates for C-to-U editing off-targets existed in the independent data sets that remained unconfirmed by the respective other replicate. We now created and analyzed four further RNA-seq data sets to screen for off-targets upon expression of PPR56 in constructs without or with different co-provided target combinations (Suppl. Data 2). Including the further replicates now resulted in the identification of altogether 133 off-targets (detected in a

minimum of two independent data sets) for the wild-type PPR56 (Fig. 7). The conservation profile for the 133 off-targets of wild-type PPR56 excellently confirms strong preferences for nucleotide positions opposite of P- and S-type PPRs as predicted from the PPR code in six cases: S-10TD:g, P-9TN:a, S-7TD:g, S-4TN:a, P2-3ND:u and S2-1ND:u. As generally known, we see a higher discrimination for the identities of purine than of pyrimidines. However, instead of an expected selectivity for uridine in position -9 opposite of PPR P-6ND we find a slightly stronger preference for guanidine. Notably, a guanosine is also unexpectedly present in the more efficiently edited native *nad4* target of PPR56. Additionally, there is strong selectivity for pyrimidines not only in positions -3 to -1 (mostly as UCU) but also in position -5 opposite of PPR L2-VD (Fig. 7). Moreover, L-type PPR L-8VD appears to select against guanosine whereas no selectivity for pyrimidines is found in positions -16 and -15 opposite of PPRs S-13NS and P-12NN.

Additionally, we included RNA-seq analyses for three datasets each of the two PPR56 mutants with mutations in PPRs P-10TD>TN and S-4TN>TD, respectively (Suppl. Data 2). Intriguingly, the total number of off-targets is more than threefold (449 vs. 133) for the S-4TN>TD mutant (Fig. 7). This variant shows a strong shift in preference from adenosine to guanosine in position -7, exactly as expected from the PPR-RNA code. No further strong shifts of nucleotide preferences are observed for other positions in the conservation profile.

Mysteriously, exactly the opposite is observed for mutation of PPR56|S-10TD>TN where the number of off-targets is now drastically reduced from 133 to only 16. Expectedly, a strong selectivity for adenosine is now seen in position -13 juxtaposed with the mutated PPR as expected (Fig. 7). Further judgements on potential other changes in the conservation profile also at other positions are not evident and should be considered with caution made given the overall small number of only 16 off-targets in this case. It may be noted, however, that adenine or cytidine are prominently present here in position -11, corresponding to the identities in the two native targets opposite of PPR L-8VD, which had turned out to be most sensitive against changes (Fig. 4).

Serial combinations of PPR56 targets

The observation outlined above showing that native target sequences further upstream of the region juxtaposed with the PPR array contributed strongly for higher RNA editing activities (Fig. 6) made us consider the possibility that multiplying targets on a single transcript may affect the respective RNA editing outcomes. The two known targets of PPR56 edited with high (*nad4*) and moderate (*nad3*) efficiencies offered an interesting test case allowing to check upon RNA editing activities at targets of PPR56 in varying combinations (Fig. 8). Cloning the *nad3* target upstream of the *nad4* target led to a further reduction of nad3eU230SL RNA editing activity while leaving editing nad4eU272SL unaffected. A striking result was obtained, however, upon cloning the two targets in the reverse order (Fig. 8). Again, nad4eU272SL editing remained unaffected but editing of nad3eU230SL site now rose to >99% indicating a beneficial effect of the upstream *nad4* target. This surprising enhancing effect of the upstream *nad4* target could even be seen more drastically for the previously tested *nad3* target variant where RNA editing was eradicated with a GC-rich sequence upstream of position -20 (Fig. 6), where RNA editing activity is now boosted to 94% (Fig. 8).

To check whether the enhancing effect of the upstream *nad4* target was dependent on its editability, we converted it into a "pre-edited" state replacing the target cytidine with thymidine (nad4eU272SL|c0u). Notably, the enhancing effect on the downstream *nad3* target remained unaffected, still resulting in >99% conversion at the nad3eU230SL target (Fig. 8). However, introducing mutation nad4eU272SL|a-12g that creates a conceptual mismatch to PPR P-9TN and was found to abolish nad4eU272SL editing (Fig. 2) into either the native or the pre-edited *nad4* target reduced the enhanced editing at the downstream *nad3* target to 93% or 86%, respectively (Fig. 8). We conclude that the "strong" *nad4* editing target can act as an upstream enhancer of downstream editing in its native form but independent of a requirement for the nad4eU272SL editing event.

We wished to check upon a potentially enhancing effect also on two selected off-targets of PPR56 in *E. coli* (Suppl. Data 2). Off- targets yegHeU419SL and folDeU-5 were edited to 38% and 78%, respectively, in the *E. coli* background transcriptome. However, only 38% of editing was observed for

371 folDeU-5 and none at all for yegHeU419SL when cloned individually analogous to the native targets
372 behind the PPR56 coding sequence. RNA editing of >99% or 17%, respectively, was observed when
373 placed in tandem behind the upstream *nad4* target.

374 Finally, we wondered whether such enhancing lateral effect on targets cloned in tandem
375 combinations could also be seen for the moderately efficient edited *nad3* target alone. Indeed, a
376 triplicate arrangement of *nad3* targets resulted in diminished activity at the upstream-most copy, but
377 enhanced RNA editing efficiencies at the middle an 3'-terminal target copy (Fig. 8). Hence, very much
378 like the experimentation with truncation of the upstream extensions of the native targets (Fig. 6)
379 these finding indicate that upstream sequences are necessary to attract an editing factor, which may
380 then scan the transcript in a 5'-3'-direction for the ultimate match of its PPR array to the RNA editing
381 target.

382 ## Placement of targets towards the 5' or 3'-end of a long RNA

383 We wished to test placement of targets in different positions and made use of the newly constructed
384 vector pET41Kmod2 (Suppl. Fig. 2), which allows the alternative cloning of targets also upstream of
385 the editing factor coding sequence into the 5'-UTR. A combination of the *nad4* target in the 5'-UTR
386 with the *nad3* target in the 3'-UTR could not enhance editing of the latter while the former remained
387 unaffected (Fig. 9A). Surprisingly though, cloning in the inverse arrangement led to significant
388 increase in editing at the nad3eU230SL target when cloned into the 5'-UTR (Fig. 9A). This held equally
389 true for tandem cloning of the two targets into the 5'-UTR in either orientation (Fig. 9B). Evidently,
390 providing the "weak" *nad3* target in a 5'- rather than in a 3'-UTR appears to allow for better access
391 and more efficient editing, aside from the enhancing effect of tandem target arrangements.

392 Resulting from the above findings, we tested five additional off-targets identified in *E. coli*
393 (fdhEeU403Q*, paoCeU542TM, rarAeU407TI, arnAeU242SF and cydCeU980PL) that showed variable
394 editing efficiencies at different RNA read coverages and different matches to the PPR array of PPR56
395 (Fig. 10A). Towards that end we tested both for an effect of tandem-cloning with the upstream *nad4*
396 target (Fig. 10B) as a possible enhancer as well as for their placement in the 5'-MCS in wide distance

from the downstream *nad4* target (Fig. 10C). In three cases we found that RNA editing could be strongly enhanced both by placing the respective off-target either in tandem behind the native *nad4* target or alternatively into the 5'-MCS distant from the nad4eU272SL target located in the 3'-MCS: rarAeU407TI from 24% to 66% or 70%, fdheU403Q* from 16% to 75% or 61% and for cydCeU980PL from 50% to over 99% with both placements, respectively. However, a striking reduction was found to only 4% for arnAeU242SF with both cloning strategies and even to the abolishment of editing for paoCeU542TM in the tandem cloning approach (Fig. 10B). Notably, in the latter case RNA editing at the native nad4eU272SL site was concomitantly also reduced to 62% while the usual highly efficient editing was observed in the other nine constructs.

Exploring novel candidate targets

It is important to keep in mind that orthologues of a functionally characterized plant RNA editing factor may have additional or different functions in other species. Intriguingly, the two targets of PPR56 in *Physcomitrium patens* are not conserved in most other available moss mitogenomes (with the exceptions in the Pottiaceae), but rather exist in a pre-edited state with thymidines in the genomic positions. Using the TargetScan option of PREPACT [37] we wished to find alternative targets for PPR56 that may exist in a pre-edited state with thymidine present in the mitogenome of *Physcomitrium*. Indeed we could find cox3eU290SF as such as potential target matching excellently to the RNA binding properties of PPR56 (Fig. 10A). The *E. coli* RNA editing assay setup allows to test such a hypothesis quickly and we accordingly exchanged the T at the potential editing position of the *Physcomitrium* mtDNA sequence into a C. Whereas we could not detect editing of cox3eU290SF when routinely cloned as a single target inserted downstream of the PPR protein coding region, we observed an editing efficiency of 93% when cloned in tandem downstream of nad4eU272SL (Fig. 10B). At present, cox3eU290SF cannot be identified as a candidate editing site in moss mtDNAs but is confirmed as an RNA editing site in the mitochondria of the lycophytes *Isoetes engelmannii* [39] and *Selaginella moellendorffii* [40] and in the fern *Haplopteris ensiformis* [41].

Discussion

Mutating the DYW domain: different effects on two native targets

All of our experimentation showed that the *nad4* target of PPR56 is more resilient towards changes both on the target side and on the protein side than the *nad3* editing target site, which proved to be much more sensitive. Notably, the higher sensitivity of the *nad3* target towards changes also extended to mutations in the DYW domain of PPR56 (Fig. 1B). The carboxy-terminal DYW domain of plant RNA editing factors has long been suspected, and is meanwhile well confirmed, as the catalytic cytidine deaminase domain [23,33,35,42–44]. Many of the highly conserved amino acid residues in the DYW domain are essential for functionality as here again confirmed with a set of mutations in the DYW domain of PPR56. However, while six mutants with single amino acid exchanges in the DYW domain of PPR56 lost RNA editing activity on both targets, seven others affected RNA editing at the nad3eU230SL target more strongly than at the nad4eU272SL site (Fig. 1B). This is all the more striking given that target positions -2 to +2 around the cytidine targeted for C-to-U conversion are identical for the two targets of PPR56. Evidently, the DYW domain is not simply a flexible enzymatic unit that can easily be transplanted but relies on the intricate interactions of the upstream protein regions with different RNA targets. Notably, the *nad4* target of PPR56 not only tolerates exchanges in positions +1 and +2 allowing for the artificial creation of stop codons through C-to-U RNA editing but also for the artificial creation of a start codon after conversion of position -1 to adenosine (Fig. 5).

PPR arrays: The P- and S-type PPRs

It is generally understood that the upstream PPR array of a plant RNA editing factor is responsible for proper target recognition following the established PPR-RNA code rules [28–31,45]. PPR56 is no exception but it should be noted that its P- and S-type PPRs show overall even a slightly better fit to its more weakly edited target nad3eU230SL than to its strongly edited target nad4eU272SL (Fig. 1A). Target selectivity following the PPR code is excellently reflected by the off-target conservation profiles fitting expectations for three P-type and three S-type PPRs of PPR56, including an intended re-targeting after changing key positions in two of these PPRs (Fig. 7). However, exceptions exist as seen for P-type PPR P-6ND which unexpectedly appears to select for guanidines as well as for

449 uridines (Fig. 7), possibly as part of the explanation for efficient editing of nad4eU272SL with a
450 guanidine in the corresponding target position -9. However, RNA editing is lost at the u-9g target
451 mutant of nad3eU230SL (Fig. 2) and this is just one of several examples found in the course of our
452 work showing restricted predictability for RNA editing activities even upon small molecular changes.

453 Another dramatic example is a single u-to-c transition in position -15 of the targets which leaves
454 the high editing efficiency at nad4eU272SL unaffected but abolishes editing completely for the
455 nad3eU230SL target (Fig. 2). This is quite surprising given that the N-terminal PPRs generally
456 contribute more weakly to target selectivity and, fitting this general assumptions, the off-target
457 conservation profiles show no strong preference in these positions (Fig. 7).

458 Similarly, the behavior of PPR56 protein variants is predictable only to a limited degree. For
459 mutations in the crucial positions 5 or L of P- and S-type PPRs of PPR56 we found that ca. 50% of
460 them could be rescued to variable degrees by corresponding mutations on the target side for at least
461 one of the native targets (Fig. 3B). However, this was not the case for the other 50% of mutants
462 tested (Fig. 3A). The PPR protein mutants with successful retargeting included S-10TD>TN and S-
463 4TN>TD that were also tested for off-targets in *E. coli*. Intriguingly, PPR mutant S-4TN>TD not only
464 proved to be more resilient on the *nad4* target and to be rescued by a>g exchanges in native targets
465 (Fig. 3B), but also resulted in a more than threefold amount of 449 off-targets compared to 133 in
466 wild-type PPR56 (Fig. 7). Exactly the opposite is observed for PPR mutant S-10TD>TN having a
467 stronger impact that cannot be rescued on the *nad3* target and resulting in a strictly reduced set of
468 only 16 off-targets (Fig. 7). A similar, although not quite as drastic effect has recently been found for
469 another PPR re-targeting mutant S-7TD>TN in human cells while a huge increase in off-targets was
470 also seen for the S-4 TN>TD mutant [24]. We conclude that the observed effects are very unlikely an
471 effect of the bacterial vs. the eukaryotic expression setups but rather inherent to the PPR array and
472 strongly point to significant impacts on overall protein features even upon changes of single amino
473 acids in a dedicated PPR. Individual PPRs appear to contribute very differently to target recognition
474 or ultimate RNA editing efficiencies and even single amino acid exchanges in position 5 or L of a PPR

may strongly increase or decrease the flexibility of an RNA editing factor for target recognition. In this context it should be remembered that several point mutation alleles also outside of positions 5 or L in PPRs of functionally characterized RNA editing factors strongly affected specific RNA editing functionality in yet unclear ways [e.g. 46]; a G-to-R mutation in the DEK46 protein is a recently reported example along those lines [47].

PPR arrays: The L-type PPRs

The contribution of L-type PPRs for target recognition has been investigated previously, ascribing them a role in RNA editing but not in RNA binding [48]. Notably, the two native targets of PPR56 display different nucleotides opposite of their three central L-type PPRs (Fig. 4). Creating target mutants replacing the nucleotides with the respective other showed clear effects only for PPR L-8VD (Fig. 4). Nevertheless, the cytidine-to-adenosine exchange in the *nad4* target as well as the inverse exchange in the *nad3* target position -11 juxtaposed with PPR L-8VD both reduced RNA editing efficiency (Fig. 4). Remarkably, however, the reduced off-target data set for the PPR mutant S-10TD>TN in particular shows a clear preference for adenosine or cytosine in this position, matching the nucleotide identities in the two native targets (Fig. 7).

RNA editing efficiencies and the wider transcript context

Using target predictions based on the PPR-RNA code generally finds many additional candidate RNA editing sites with equal of even better matches than the documented targets of an RNA editing factor, but these sites remain unedited. To some extent, RNA secondary may play a role to explain this observation. Placing the cytidine to be edited in the context of RNA secondary structures can reduce or even abolish RNA editing altogether (Suppl. Fig. 3). In case of the two closely spaced mitochondrial editing sites ccmFCeU103PS and ccmFCeU122SF in *P. patens*, the upstream located editing site needs to be addressed by PPR65 first, most likely to destabilize a secondary structure to allow PPR71 to bind and edit the downstream site [49]. Such observations can certainly be expected given that binding of a PPR protein to RNA must compete with RNA secondary structure formation. This has been investigated systematically previously, e.g. for the P-type protein PPR10 [50].

Particularly interesting will be the further functional characterization of RNA editing factors like DEK46 acting on edited cytidines naturally embedded in stable secondary structures such as domain V of group II introns [3,22,51]. However, reliable prognoses on a RNA secondary structures are mostly limited to small transcripts while predictions of long-range base-pair formations *in vivo* is questionable.

Maybe more importantly, we here found that several transcript features beyond the region ultimately targeted by the PLS-type PPR array strongly contribute to attract and/or enhance the activity of an editing factor like PPR56. With the benefit of hindsight it has likely been helpful that 5'-extensions beyond the core PPR-targeted region have been included initially in the establishment of the heterologous editing systems [23,24]. We now found that additional native sequences upstream of the RNA sequence ultimately targeted by the PPR array have a significant influence on efficient RNA editing. Progressive 5'-deletions of the native targets and their replacement with foreign sequences results in stark reduction of RNA editing up to complete loss in the case of the "weak" *nad3* target despite retention of native sequence 20 nucleotides upstream of the cytidine to be edited.

Vice versa, we find that within tandem arrangements, an upstream target is able to enhance RNA editing at the downstream targets and this is independent of a cytidine present for conversion to uridine in the upstream "enhancer" target. Notably, it may be interesting to remember that an enhancing effect of multiplied targets had also been observed in early *in vitro* experimentation [52]. Like the series on truncating native sequences further upstream, the new findings may suggest a diffuse enrichment of RNA editing factors like PPR56 near their final destination or possibly even an as yet enigmatic mechanism of 5'-to-3' scanning towards their ultimate target of cytidine deamination. With the enhanced system, we were also able to identify cox3eU290SF as a new additional target in the mitochondrial transcriptome of *P. patens*, which can be recognized by PPR56 and even be edited, when a C is introduced at the editing position.

Designing our setups for heterologous expression, we placed the editing targets into the 3'-UTR behind the editing factor coding sequences, which was intended to test for RNA editing by subsequent cDNA analysis restricted to full length mRNAs. Surprisingly, we now find that not only tandem target arrangements but also their alternative placement of into the 5'-UTR can enhance RNA editing to >99% (Figs. 8-10).

Conclusions and outlook

It is likely unsurprising that heterologous functional expression in prokaryotic and eukaryotic setups and for *in vitro* studies succeeded with evolutionary ancestral RNA editing factors comprising all necessary functionalities in just one polypeptide [23,24,33,35,53,54]. All available data for PPR56 show very similar behavior upon heterologous expression in the bacterial or human cells and even despite differently fused protein tags, indicating its independence from prokaryotic or eukaryotic host factors or from the many other plant organelle RNA maturation factors [55]. Functional heterologous expression will be much more complex for multiprotein editosomes that have to assemble for RNA editing in flowering plants to reconstitute target recognition and a DYW-type cytidine deaminase or to enhance RNA-binding capacities with MORFs/RIPs by protein-protein interactions [56–59].

PPR proteins are frequently investigated by *in vitro* experimentation with REMSAs (RNA electromobility shift assays) using RNA oligonucleotides representing the region bound by the PPR array. Such experimentation has contributed tremendously to understand their mode of binding and may be entirely sufficient for the study of P-type PPR proteins, which largely stabilize transcript ends by tight binding to an RNA, for example. However, scenarios may differ for the PLS-type PPR proteins like RNA editing factors, which are expected to bind only temporarily to allow for cytidine deamination. The *in vivo* experimentation in *E. coli* reported here strongly suggests that the wider transcript environments and the placements of targets matter significantly for the ultimately detected RNA editing frequencies.

We here report that several circumstances affect RNA editing efficiencies even for "simple" single-polypeptide RNA editing factors like PPR56, including (i) the enigmatic L-type PPRs, (ii) the RNA sequences further upstream of the region ultimately bound by the PPR array, (iii) the tandem combination of targets or (iv) their respective placement in long transcripts as here exemplarily shown for the 5'- and 3'-UTRs flanking the PPR56 coding region with our modified vector setup. Whether binding preferences of individual PPRs in plant editing factors can be simply changed via modification of the 5th or last amino acid appears to very much rely on their respective position and/or the overall structure of the PPR array. Hence, any future experimentation with native RNA editing factors or those based on artificial "designer" PPR arrays [53,54,60–67] should take the above into account for testing and conclusions.

Acknowledgements

We gratefully acknowledge the computer resources and support provided by the Paderborn Center for Parallel Computing (PC²). We wish to thank Bastian Oldenkott, Philipp Gerke and Simon Zumkeller in our group for the establishment and help in further development of bioinformatic pipelines and Sarah Brenner for technical assistance. We especially like to thank Elena Lesch for establishing the program to generate construct-specific DNA references. We also like to thank Bastian Oldenkott for designing the initial PPR model for PPR protein figures. We thank Mark Hermann Vegas and Grazia Margherita Willerscheidt for cloning constructs and performing initial *E. coli* experiments as part of their experimental Bachelor theses work.

Author contributions

Y.Y. cloned most of the constructs and did most of the RNA editing analyses in the heterologous system, analyzed RNA-seq data and created figures. J.O. and K.R. significantly contributed to construct cloning and RNA editing assays. M.S-R. and V.K. designed and supervised the study program and contributed to data analyses. V.K. wrote the manuscript, which was read, edited and ultimately approved by all co-authors.

Materials and Methods

Molecular Cloning

Cloning for expression of *Physcomitrium patens* PPR56 variants and targets in *Escherichia coli* was based on vector pET41Kmod as outlined earlier [23]. Protein coding sequences are cloned via gateway cloning downstream of an N-terminal His_6 tag and the maltose-binding protein (MBP) for improved protein solubility [68] behind a T7 promoter controlled by the lac operator. RNA editing target sequences were cloned behind the protein sequence upstream of a T7 terminator. Here, we also created a new vector variant pET41Kmod2 (Suppl. Fig. 2) with further restriction sites allowing for cloning targets also upstream of the respective coding region. To that end, we made use of a former *XbaI* site to create a *NotI-EcoRI-PacI-PstI* multiple cloning site (MCS) upstream of the ribosome binding site (RBS) in pET41Kmod. Target sequences including flanking restriction sites were generated with synthesized oligonucleotides for both DNA strands (Integrated DNA technologies Europe, BVBA, Leuven, Belgium) and ligated into dephosphorylated vectors after hybridization and phosphorylation. All oligonucleotides used in the course of this work are listed in supplementary data 3. To introduce site-directed mutations into PPR56 coding sequence we used an overlap PCR strategy with mutagenizing oligonucleotides. N-terminally truncated PPR56 coding sequences were amplified with classic PCR approach using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific) as described [23] to retain 14 native amino acids upstream of the most N-terminal completely retained PPR (Fig. 3D)

Protein expression and analysis of RNA editing

The setup for the expression of different constructs in the heterologous *E. coli* system and the downstream analysis of RNA editing was done as outlined previously [23]. Briefly, 25 mL of *E. coli* Rosetta 2 (DE3) cultures were pre-grown in 100 mL Erlenmeyer flasks with baffles in LB medium supplemented with 50 µM kanamycin, 17 µM chloramphenicol and 0.4 mM $ZnSO_4$ at 37°C until reaching an OD_{600} of ca. 0.5. The bacterial cultures were then cooled on ice for 5 min. before adding 0.4 mM IPTG for induction of expression and incubation for 20 h at 16°C and 180 rpm. To further explore the expression system, we here also tested elevated incubation temperatures of 24°C

603 instead of the routinely used 16°C for incubation after induction of expression (Suppl. Fig. 4A) and
604 shorter incubation times of only 4 h or 8 h, respectively, instead of the routinely used 20 h incubation
605 time before harvest and analysis of RNA editing (Suppl. Fig. 4B). These experiments suggested to
606 further use a 20 h incubation time at 16°C routinely, although shortened incubation times may be
607 warranted to differentiate between constructs when very high RNA editing activities are observed.
608 PPR56 protein variants were routinely checked for expression on SDS-PAGE gels. Mutant proteins not
609 revealing RNA editing were additionally checked by solubility tests as outlined previously [33] using
610 monoclonal antibodies against His_6 (His.H8, Invitrogen) and secondary antibody Rabbit anti-Mouse
611 IgG (H+L) (Invitrogen).

612 ## Total RNA sequencing and off-target detection

613 To identify off-targets in the *E. coli* transcriptome, total RNA was prepared from individual
614 experiments by using the Nucleo-Spin RNA kit (Macherey-Nagel), followed by DNase I treatment
615 (Thermo Fisher Scientific). Library preparation was done after rRNA depletion (TruSeq Stranded Total
616 RNA with Ribo-Zero), followed by Illumina sequencing (150 bp paired-end with NovaSeq 6000) done
617 by either Novogene or Macrogen. To generate construct-specific DNA reference reads, the simulated
618 reads (by ART MountRainier version 2016-06-05) of pET41Kmod with PPR56 and respective target
619 sequences were merged with genomic DNA reads (WTDNA_SRR941832) of BL21(DE3) cells [69]. The
620 construct-specific reference was made by merging pRARE2 sequence (Rosetta Competent Cells,
621 70953; Millipore, San Diego, CA), pET41Kmod with respective constructs and the *E. coli* BL21 genome
622 (CP010816.1). The datasets obtained are summarized in supplementary data 2. After quantifying the
623 RNA-seq raw data by FastQC (https://www.bioinformatics.babraham.ac.uk/ projects/ fastqc/), the
624 transcriptome reads were aligned with construct-specific DNA reads against the construct-specific
625 reference by GSNAP v2020/04/08 [70] with proposed settings [71]. The SNPs were called by JACUSA
626 v1.3 [72]. The SNPs were further restricted by a custom-made R script (established with kind help
627 provided by S. Zumkeller) restricting to SNPs obtained in at least two datasets from expression of the
628 same protein but not in wild-type or expressing other editing factors like PPR65 [23]. Final RNA

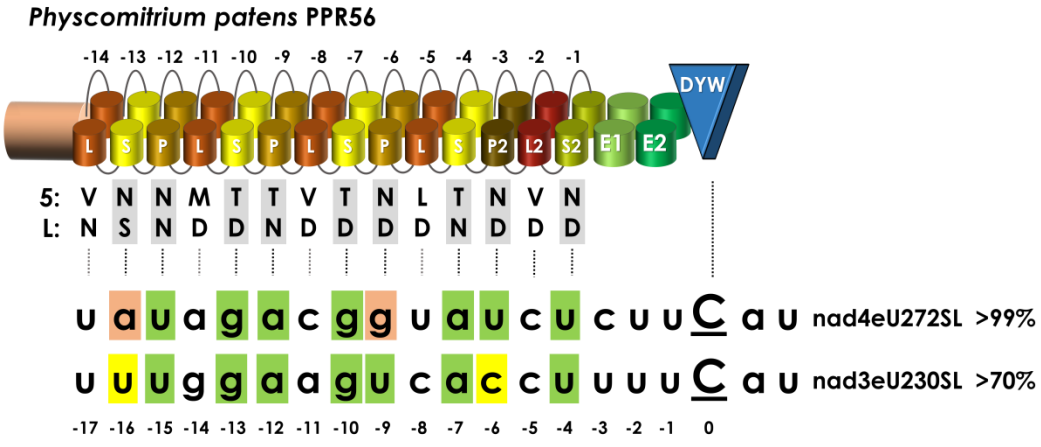
629 editing efficiency was calculated by adding up total RNA reads from all hitting datasets at a site. RNA
630 editing sites were only considered for sites with (i) RNA read coverage of at least 30, (ii) a clear signal
631 for transition in the RNA reads (T+C or G+A > 99%), (iii) a clear DNA reference position (G or C > 98%)
632 and (iv) a C-to-U RNA signal of at least 1%. The original SNP mapping data are given in supplementary
633 data 2.

Figure Legends

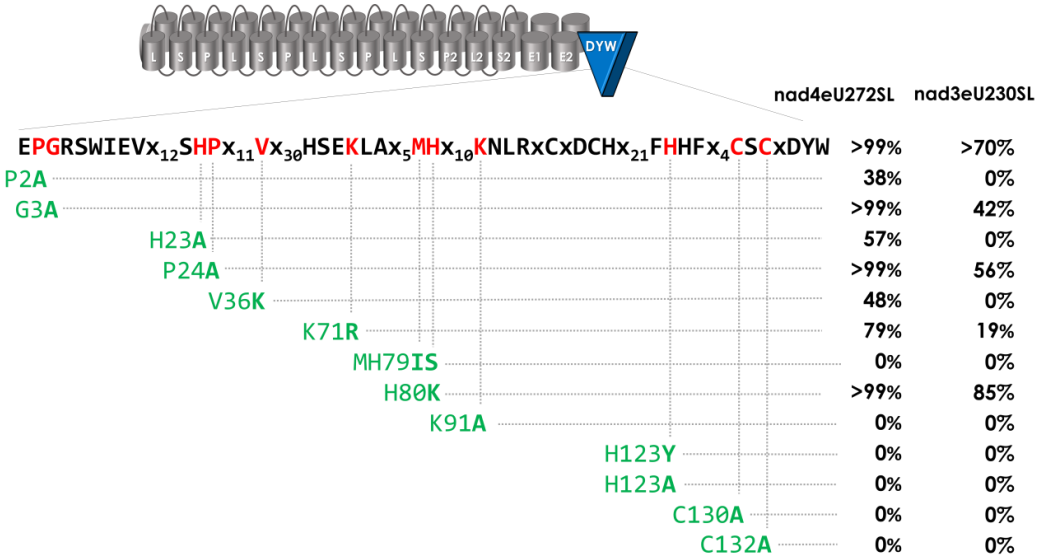
Figure 1. PPR56 and site-directed mutations in its DYW cytidine deaminase domain.

A. PPR56 is a typical plant organelle RNA editing factor featuring a PLS-type PPR array with alternating P-, L- and S-type PPRs followed by extension motifs E1 and E2 and a terminal DYW cytidine deaminase domain. Typically, the most C-terminal PLS triplet of plant editing factors has a deviating consensus and is labeled P2-L2-S2. As suggested previously [73], to account for generally more loosely conserved N-terminal repeats, PPRs are numbered backwards with the terminal PPR S2-1 juxtaposed with position -4 upstream of the editing target cytidine converted into uridine. Shading of matches in green follows the PPR-RNA recognition code based on amino acid identities in positions 5 and L in P- and S-type PPRs: T/S+N:A, T/S+D:G, N+D:U, N+S:C, N+N:Y. PPR56 has two native editing targets in the mitochondria of *Physcomitrium patens*: nad4eU272SL and nad3eU230SL. Near-complete editing (>99%) is generally observed for the nad4eU272SL target, but lower editing (>70%) is variably observed for nad3eU230SL *in planta,* possibly as a result of different strains or cultivation conditions [25,26]. **B**. Thirteen conserved amino acid positions (see Suppl. Fig. 1) in the DYW domain of PPR56 were selected for mutations and tested on both native targets nad4eU272SL and nad3eU230SL in the *E. coli* RNA editing assay system. RNA editing efficiencies are given as the mean of at least three biological replicates (independent primary *E. coli* clones) when RNA editing activity was detected. Initially identified absence of RNA editing for a construct was confirmed with at least one additional independent bacterial clone. All primary data for RNA editing assays are given in supplementary data 1.

A. Makeup of PPR56 and its two native targets.

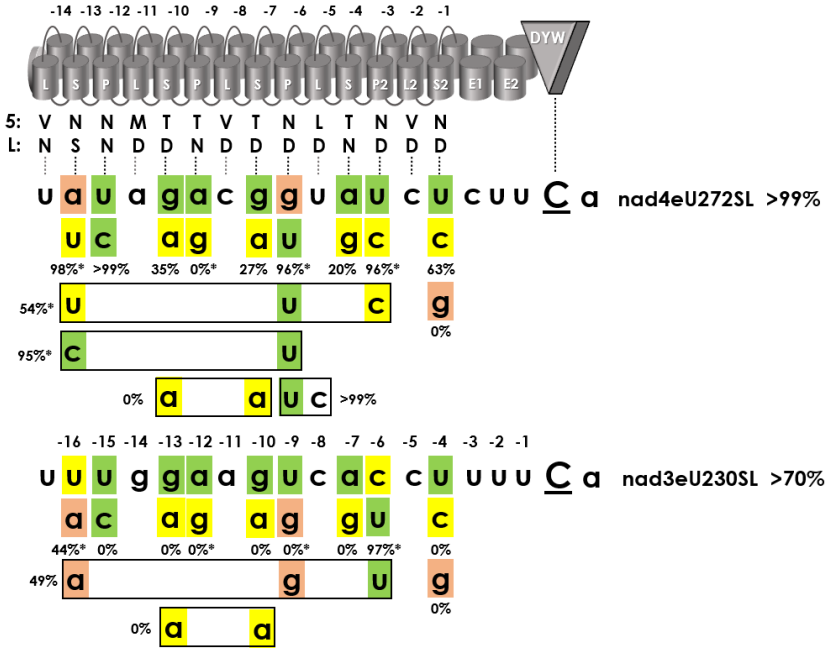


B. Mutations in the DYW domain of PPR56.



658 Figure 2. PPR56 target mutations opposite of P- and S-type PPRs.

659 Mutations have been introduced upstream of the two native PPR56 editing targets nad4eU272SL and
660 nad3eU230SL in positions juxtaposed with P- and S-type PPRs assumed to follow the PPR-RNA code
661 rules for amino acid positions 5 and L. Ten target mutants investigated earlier [23] are indicated with
662 asterisks at the respective percentages (e.g. for nad4eU272SL|a-16u, top left). Designation of PPRs,
663 numbering of positions and shading in target sequences is as in figure 1A. Average RNA editing
664 activities from three replicates are given below individually mutated positions or next to multiple
665 mutations (boxed). Primary data are listed in supplementary data 1.

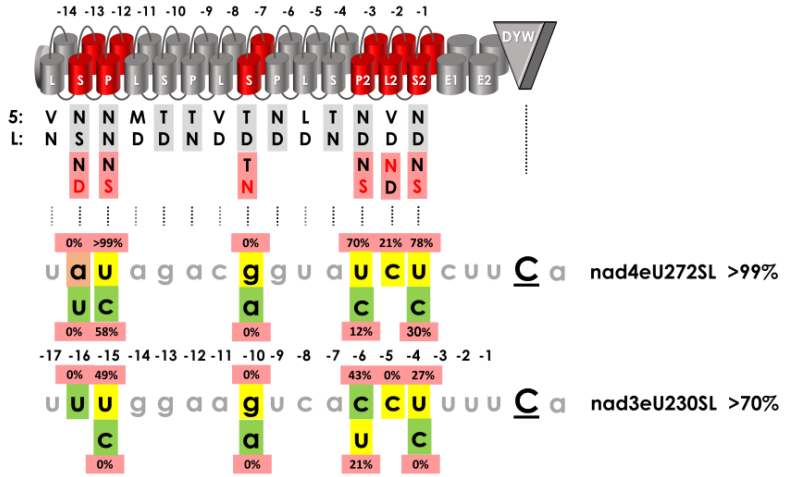


666

667 Figure 3. Compensating and non-compensating PPR and target mutants.

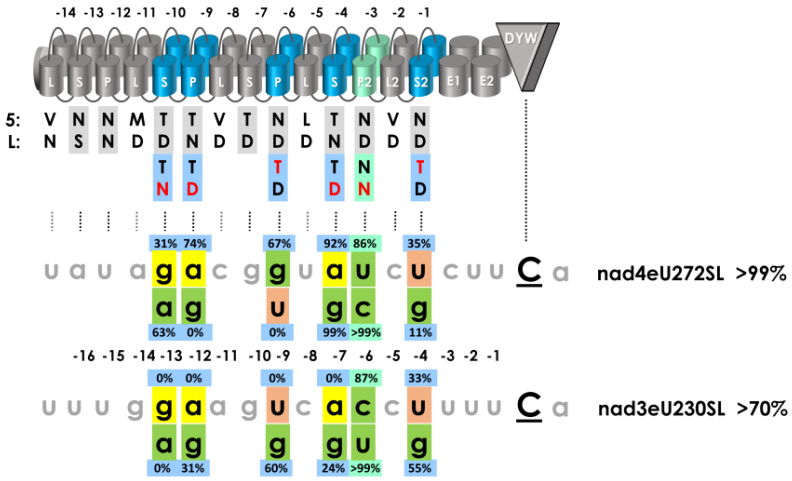
668 Key positions '5' or 'Last' have been altered in individual PPRs of PPR56 (red font) in attempts of re-
669 targeting to modified target sequences with conceptually improved matches in individual positions
670 (green shading) of native targets nad4eU272SL and nad3eU230SL, respectively. RNA editing activities
671 are indicated for the individual PPR mutants next to the respective target position identities. **A**. No
672 re-gain of RNA editing activity is observed for PPR mutations S-13NS>ND, P-12NN>NS, S-7TD>TN,
673 P2-3ND>NS, L2-2VD>ND and S2-1ND>NS (red cylinders) juxtaposed with nucleotide positions -16, -
674 15, -10, -6, -5 and -4 upstream of the edited cytidine in either target. **B**. Moderate re-gains of RNA
675 editing activity are observed for at least one of the two targets for PPR mutations S-10TD>TN, P-
676 9TN>TD, P-6ND>TD, S-4TN>TD and S2-1ND>TD (blue cylinders) opposite of nucleotide positions -13, -
677 12, -9, -7 and -4, respectively. The green cylinder and shading indicates the mutated PPR P2-3ND>NN
678 with a conceptually relaxed selectivity for U over C in position -6. **C**. A double mutant PPR56|S-
679 7TD>TN|S-4TN>TD shows no activity on the native targets but can be rescued to different amounts
680 by the corresponding g-10a|a-7g target double mutants. **D**. Progressive truncation of the two or
681 three terminal PPRs of PPR56 lead to moderate or more drastic reduction of RNA editing efficiencies,
682 respectively.

683 *A. Target mutations not rescued by corresponding PPR mutations.*



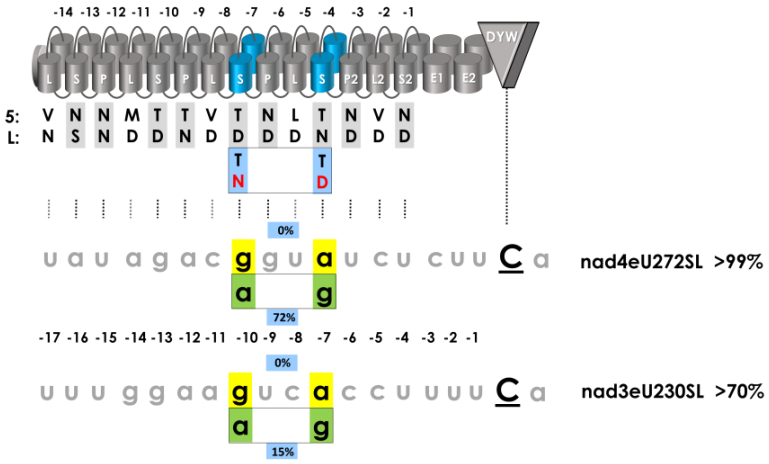
684

685 ***B.*** *Target mutations at least partially rescued by corresponding PPR mutations.*



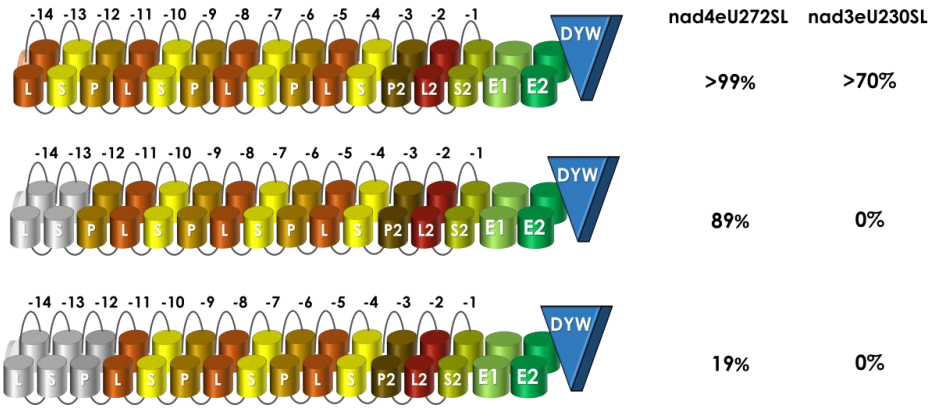
686

687 *C. Double target mutant.*



688

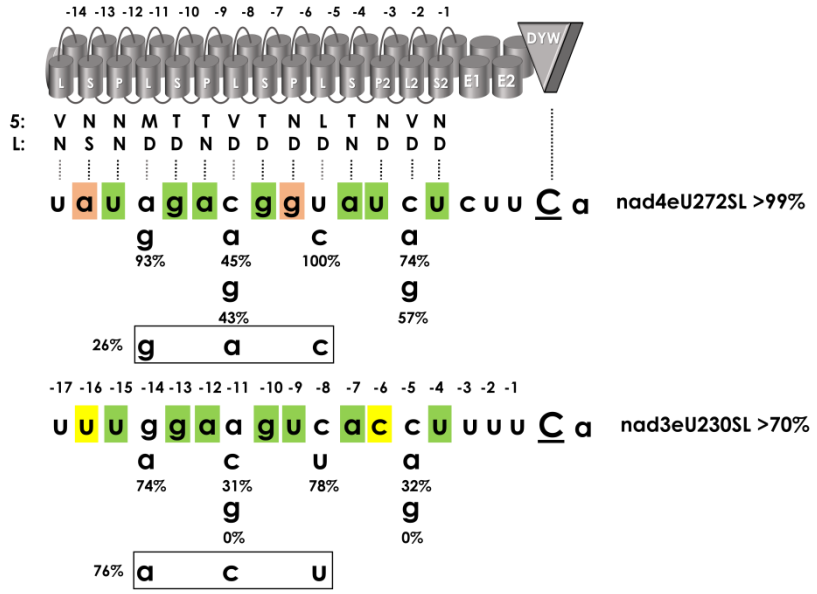
689 ***D.*** *N-terminal PPR truncations of PPR56.*



690

691 Figure 4. Mutation of target positions opposite of L-type PPRs.

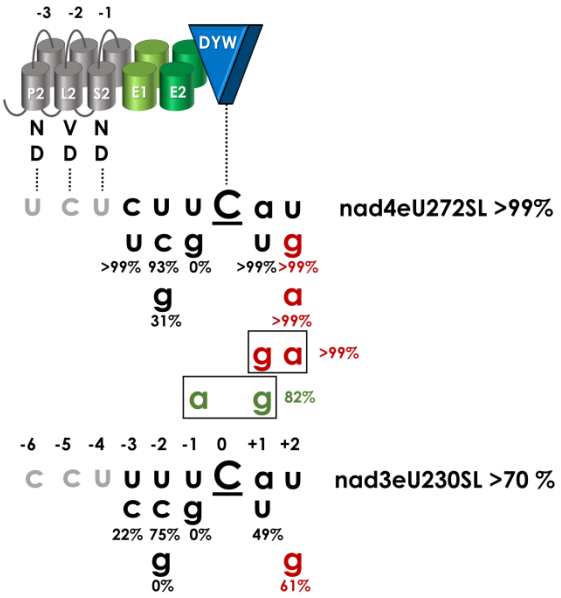
692 Target positions -14, -11 and -8 opposite of L-type PPRs L-11MD, L-8VD and L-5LD have been changed
693 to the nucleotides present in the respective other native target of PPR56. Additional mutations to
694 purines were introduced in positions -11 and -5 opposite of PPRs L-8VD and L2-2VD, which carry the
695 same combination of amino acids in positions 5 and L and are mainly juxtaposed with cytidines in the
696 targets. The strongest effects are seen for nad3eU230SL|a-11g and nad3eU230SL|c-5g abolishing
697 RNA editing completely in the modified *nad3* targets. *Vice versa*, a much stronger effect is seen for
698 the triple mutant nad4eU272SL|a-14g|c-11a|u-8c in the *nad4* target vs. the inverse changes
699 nad3eU230SL|g-14a|a-11c|c-8u in the *nad3* target.



700

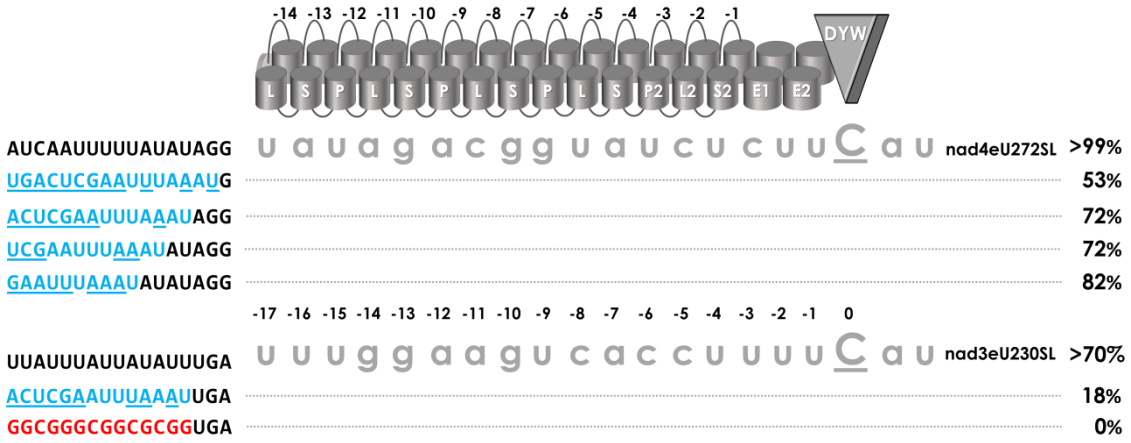
Figure 5. Mutations around the RNA editing sites.

The two native targets of PPR56, nad4eU272SL and nad3eU230SL, feature identical nucleotides in positions -2 to +2 around the edited cytidines (uuCau). With the exception of the exchange u-1g eradicating RNA editing completely at both targets, other exchanges in the upstream region show different outcomes with nad4eU272SL|u-2g (31%) vs. nad3eU230SL|u-2g (0%) or the inverse pyrimidine exchanges in position -3 with no effect for *nad4* editing but reduction to 22% for *nad3*. Changes in positions +1 and +2 do not affect editing of the *nad4* target but reduce editing of *nad3*. The overall tolerance of the *nad4* target region against mutations in positions -1, +1 and +2 allows to engineer all three artificial stop codon identities (red) or an artificial start codon (green) to be created by C-to-U RNA editing.



713 Figure 6. The influence of sequences further upstream of targets.

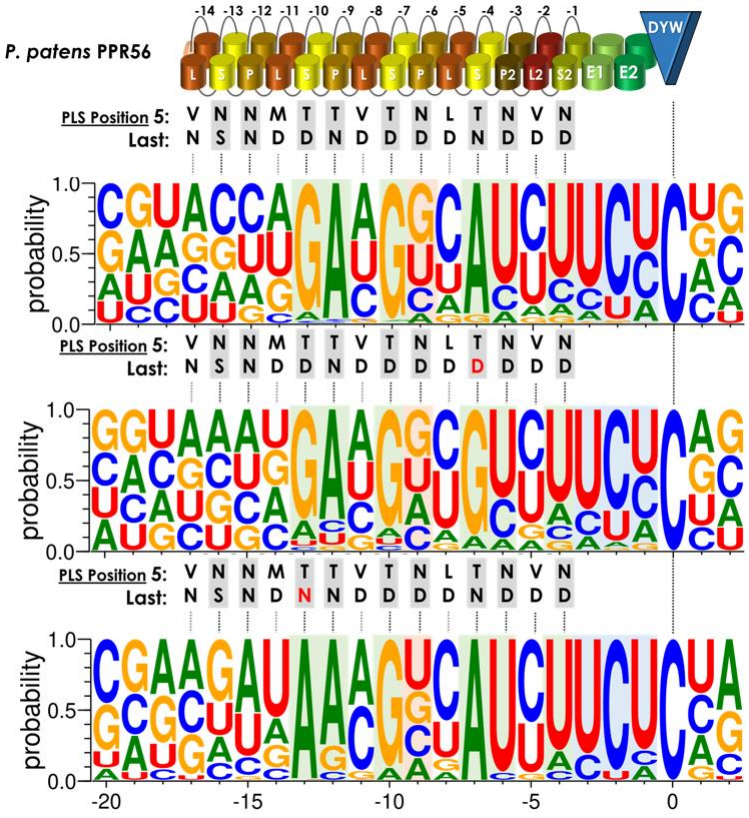
714 PPR56 editing targets were cloned with 17 bp of additional native sequence upstream of the region
715 supposed to be ultimately targeted by the PPR array, with the C-terminal PPR S2-1 juxtaposed with
716 position -4 upstream of the editing site. Progressive 5'-truncations of this upstream sequence to only
717 eight, seven, five or one nucleotide matching the native target behind the *SwaI* cloning site
718 (AUUUAAAU) place them in closer proximity to the upstream vector sequences (blue) with
719 nucleotides not matching the native upstream sequences underlined. The shortening results in
720 serially decreased RNA editing activity to 53% for the *nad4* target. A yet stronger effect is seen for
721 the *nad3* target where a 5'-truncation retaining four native upstream nucleotides reduces editing to
722 18%. Replacing the AU-rich sequence upstream of positions -20 with a GC-rich sequence (red font)
723 abolishes editing at the nad3eU230SL site altogether.



724

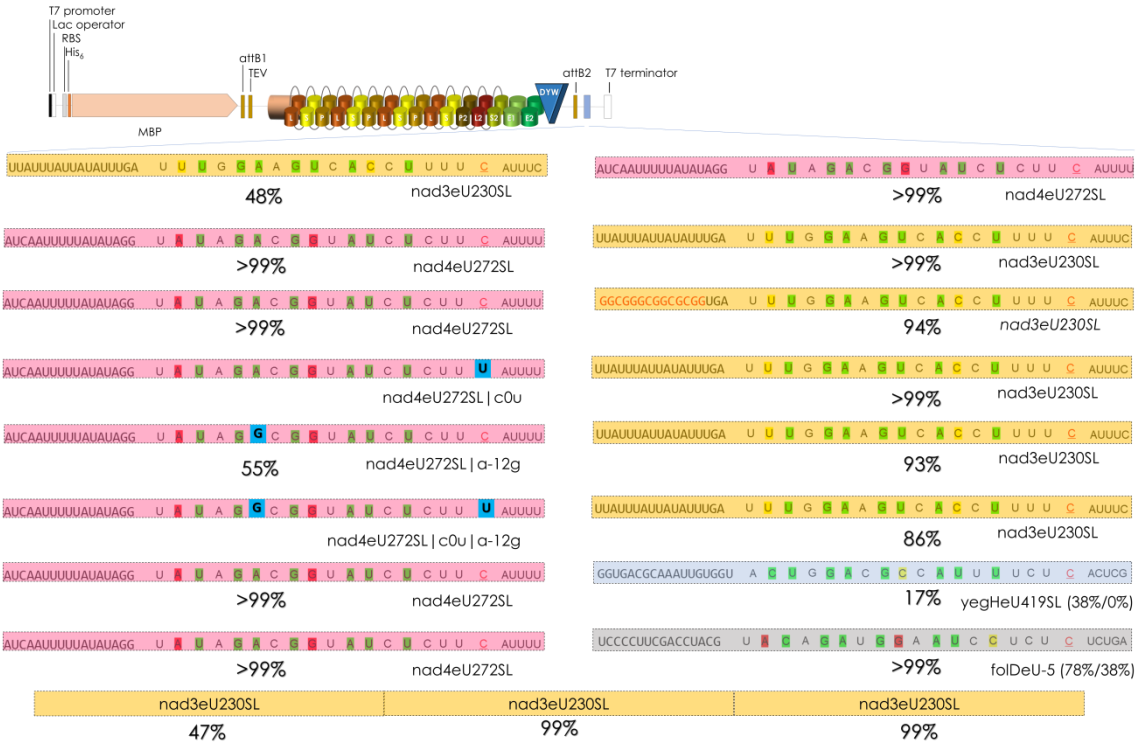
Figure 7. Off-target analyses.

Off-targets of PPR56, PPR56|S4TN>TD and PPR56|S10TD>TN in the *E.coli* transcriptome summarized with Weblogo [74]. Consensus profiles were created from the sequences of 119, 382 and 15 C-to-U RNA editing off-targets, weighted with their respective editing efficiencies. Additional off-targets requiring nucleotide shifts for better binding matches (14, 67 and 1, respectively) were excluded for clarity (Suppl. Data 2). Modified positions in the PPRs are displayed in red. The mutated PPRs have a clear preference to the nucleotides fitting best to the modified binding amino acid pair in positions 5 and L according to the PPR-RNA code. Nucleotide preferences in positions −3, −2 and −1 are highlighted in blue. Nucleotide preferences within the PPR stretch and opposite to P- or S- motifs are highlighted in green.



736 Figure 8. Combining different PPR56 targets.

737 To test for mutual influences of combined targets on the same transcript, a series of tandem
738 constructs and a triplicate arrangement of *nad3* targets was cloned in the multiple cloning site
739 behind the PPR56 coding region. Shading highlights native targets nad3eU230SL (yellow) and
740 nad4eU272SL (pink) and two off-targets identified in the *E. coli* background transcriptome in the
741 transcripts of *yegH* (blue) and *folD* (grey). Numbers in parentheses indicate RNA editing efficiency
742 observed in the off-target analysis and when cloned individually without the upstream *nad4* target,
743 respectively. The series of constructs with the *nad4* upstream of the *nad3* target includes the one
744 with the GC-rich sequence upstream of the latter (red font) that had abolished nad3eU230SL editing
745 altogether.

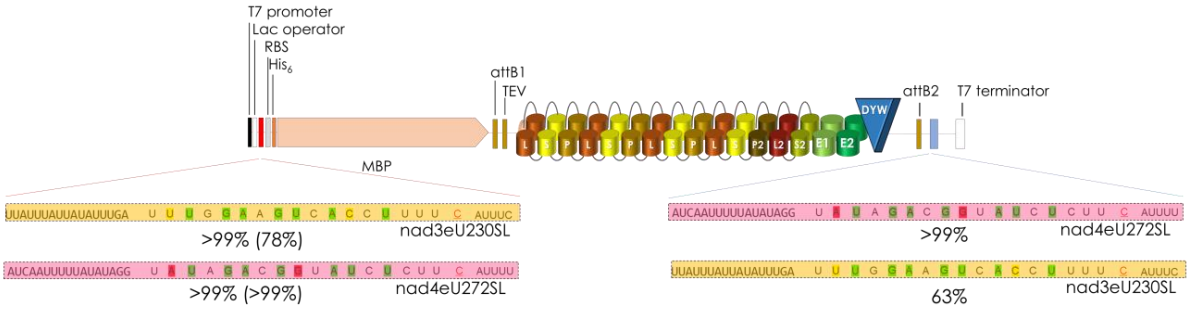


746

747 Figure 9. RNA editing target placement at the 5'- or 3'-end of a long mRNA.

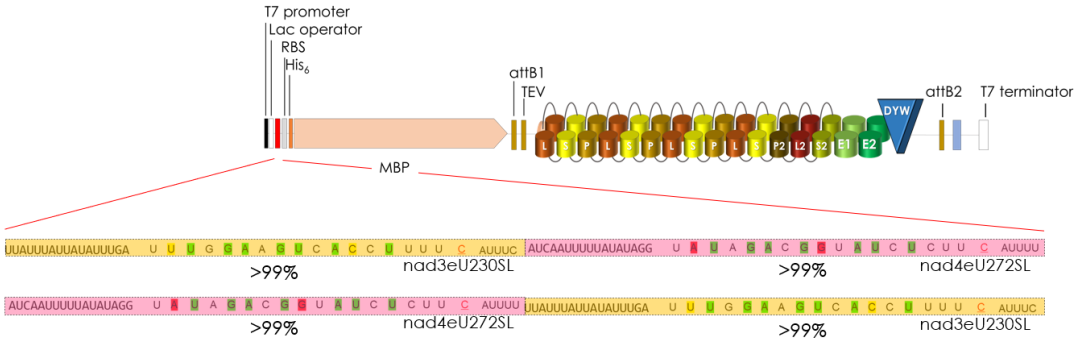
748 **A**. The two native targets of PPR56 were placed separately into the previously used 3'-MCS
749 downstream of the protein coding sequence (blue lines) and into the newly created 5'-MCS (red
750 lines) in pETG41Kmod2 (Suppl. Fig. 2) in both alternative combinations. Cloning is done via *Not*I-*Pac*I
751 in the 5'-MCS and via *Swa*I-*Asc*I in the 3'-MCS. **B**. The tandem combination of the two targets
752 previously tested in the 3'-MCS was now also tested in the 5'-MCS.

753 *A.*



754

755 *B.*



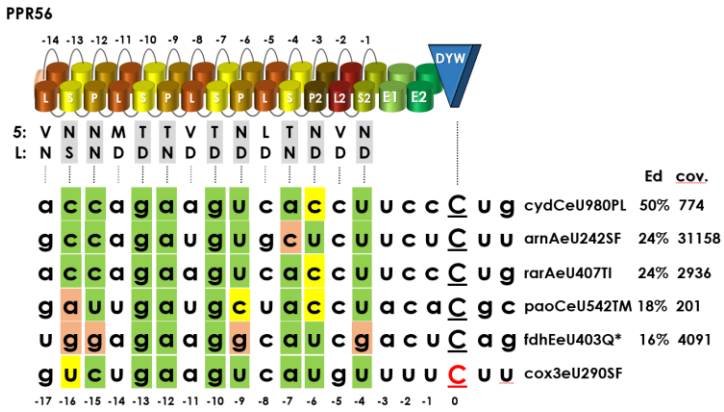
756

757

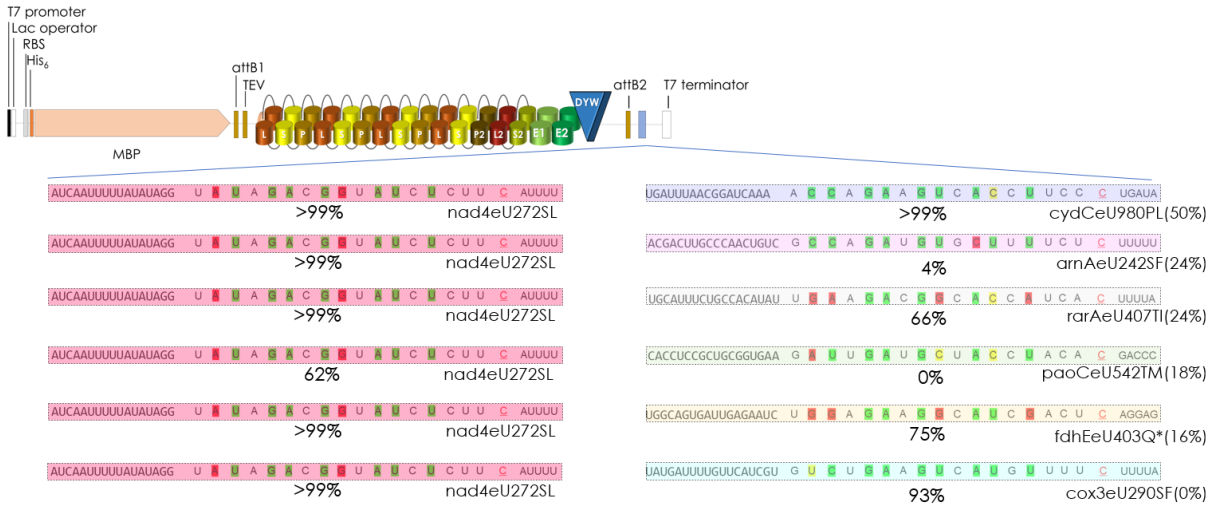
Figure 10. Off-targets in different cloning positions.

Five off-targets of PPR56 identified in *E. coli* characterized by different RNA coverages and editing efficiencies (A) were selected for cloning in tandem behind the native nad4eU272SL target of PPR56 (B) or separately into the upstream MCS in the 5'-UTR (C). Enhancement of RNA editing was found for three of the off-targets (cydCeU980PL, rarAeU407TI and fdhEeU403Q*) in either cloning arrangement and also for the, hitherto hypothetical, candidate editing cox3eU290SF when cloned downstream of nad4eU272SL (B).

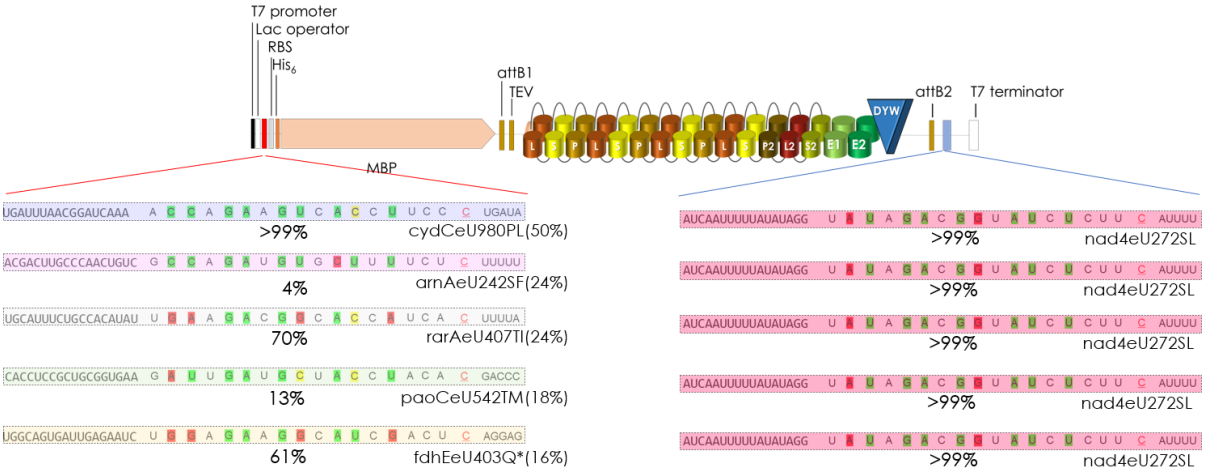
A.



B.



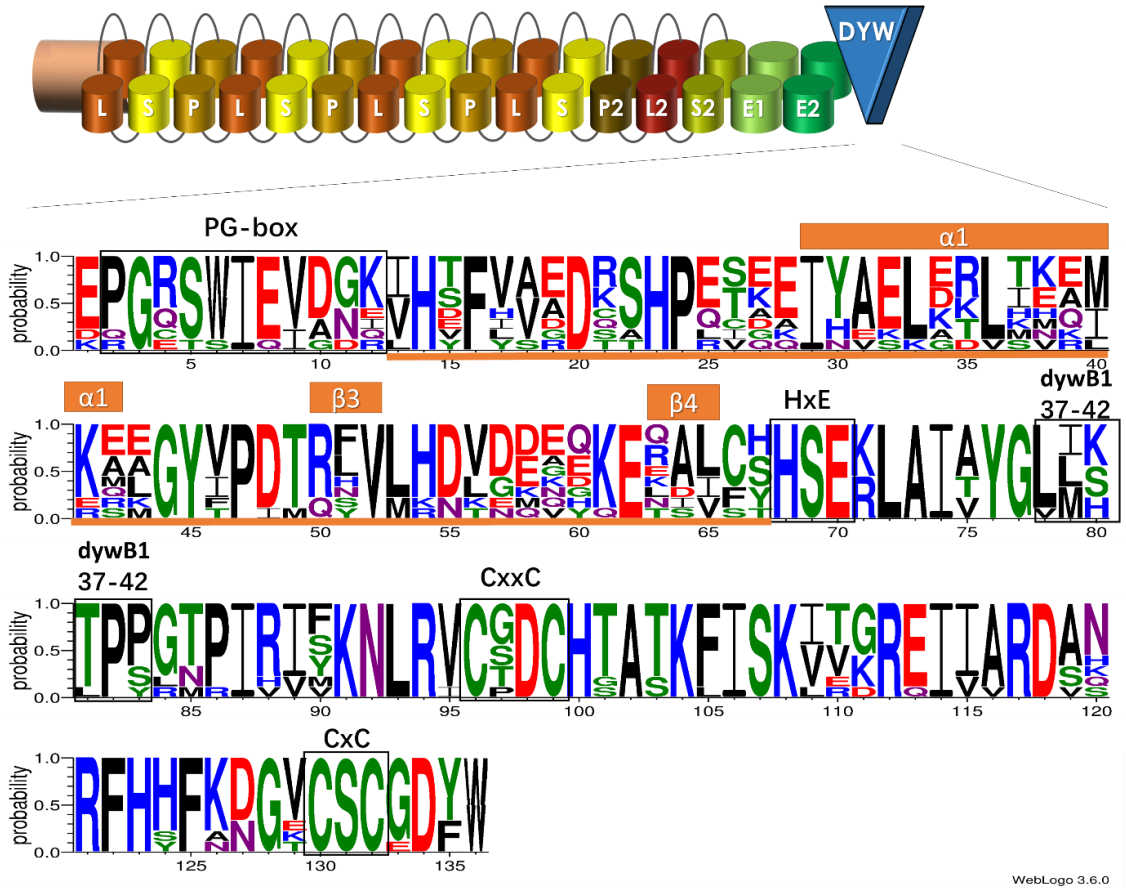
C.



Supplementary information

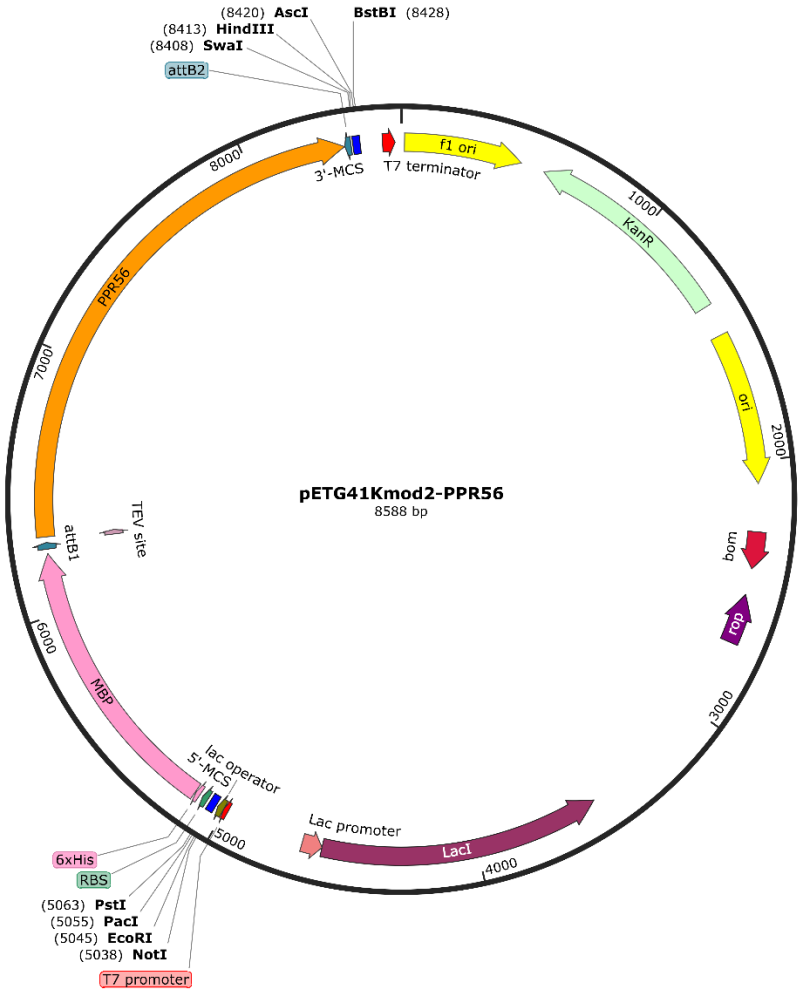
Supplementary Figure 1. WebLogo conservation profile of the DYW domains in nine *Physcomitrium patens* RNA editing factors.

The conservation plot based on the alignment of the DYW domains of nine functionally characterized RNA editing factors of *Physcomitrium patens* has been obtained with WebLogo [74]. Highlighted with frames are the characteristic PG box at the N-terminus of the DYW domain, the signature motifs for coordination of two zinc ions including the catalytic center (HSE) of the cytidine deaminase and the region of amino acids 37-42 discussed as relevant for compatibility for creating protein chimeras [36]. The "gating domain" as recently defined from X-ray structural analysis after crystallization of the OTP86 DYW domain [33] is highlighted in orange. Several residues have been selected for the study of mutants (Fig. 1B).



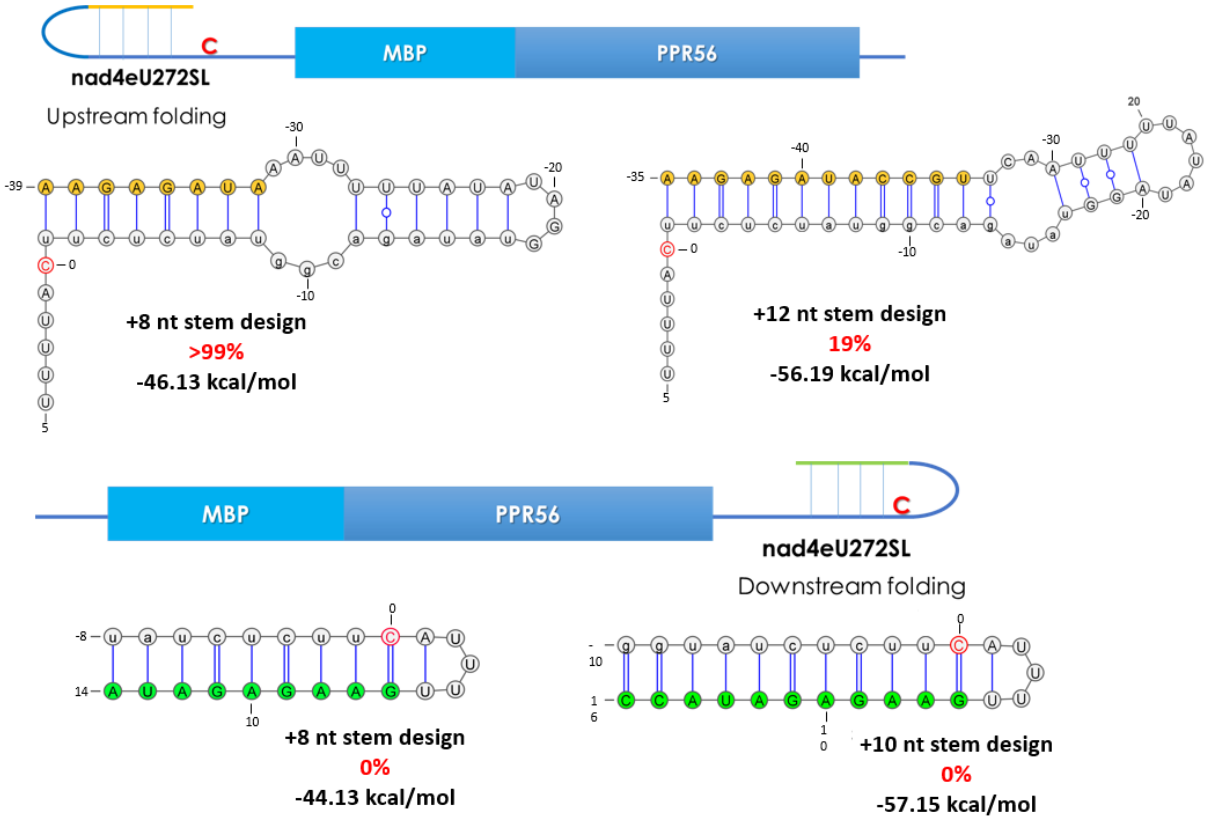
Supplementary Figure 2. Expression vector system pet41Kmod2.

Vector pET41Kmod for expression of RNA editing factors and their targets has been reported previously [23]. Coding sequences of RNA editing factors are inserted by Gateway cloning resulting in flanking attachment attB sequences connecting in-frame via a TEV cleavage site to the upstream maltose binding protein (MBP) and an N-terminal His_6 tag. Transcription is driven from a T7 promoter controlled by a lac operator and translation is initiated by a ribosome binding site (RBS). PPR56 is cloned with an N-terminal extension of 14 native amino acids upstream from its N-terminal PPR L-14. Target sequences were designed with hybridized oligonucleotides inserted by classic cloning into a multiple cloning site (MCS, *Swa*I-*Hin*dIII-*Asc*I-*Bst*BI) in the 3'-UTR between attB2 and a T7 terminator. A new vector variant pET41Kmod2 has been created which also allows for cloning target sequences alternatively upstream into the 5'-UTR in a second MCS (*Not*I-*Eco*RI-*Pac*I-*Pst*I) inserted into a previous *Xba*I site. The vector map was created with SnapGene Viewer 6.2.1 (https://www.snapgene.com).



796 Supplementary Figure 3. The influence of RNA secondary structures embedding the
797 editing site.

798 Artificial sequences have been added upstream (yellow) or downstream (green) to embed the
799 cytidine targeted for RNA editing (red) into secondary structures. The sequence upstream of the
800 cytidine editing target that is supposedly juxtaposed with the PPR array of PPR56 (see Fig. 1A) is
801 shown in small letters. The RNAfold WebServer of the ViennaRNA package [75] was used to predict
802 the secondary structures. RNA structure models were created with VARNAv3-93 (https://varna.lri.fr).

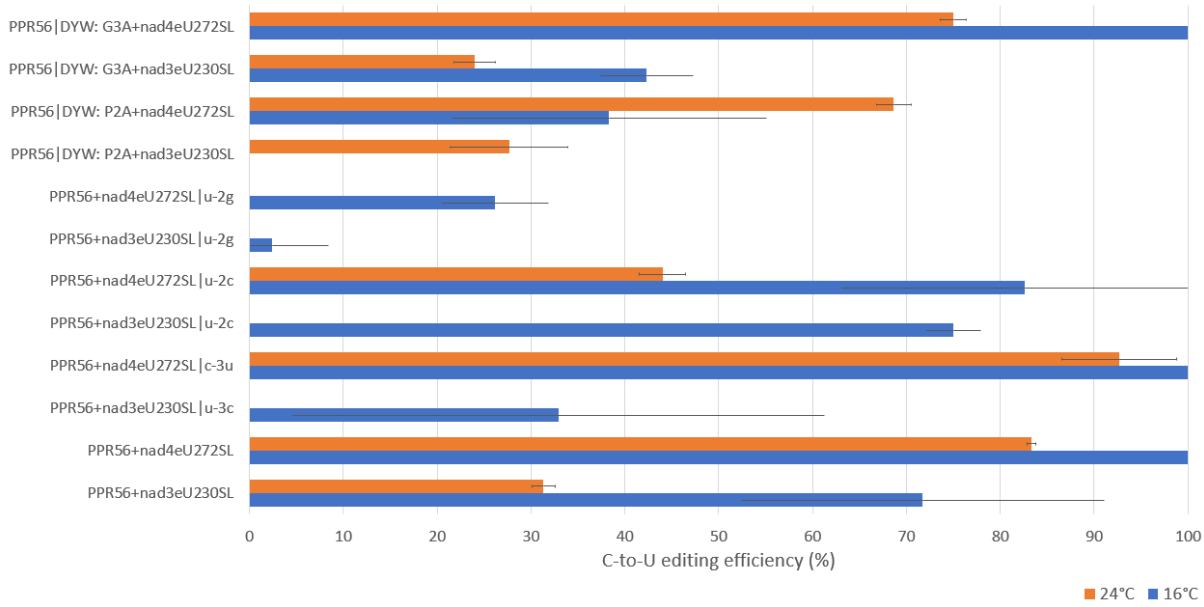


803

804 Supplementary Figure 4. Temperature- and time-dependence of RNA editing.

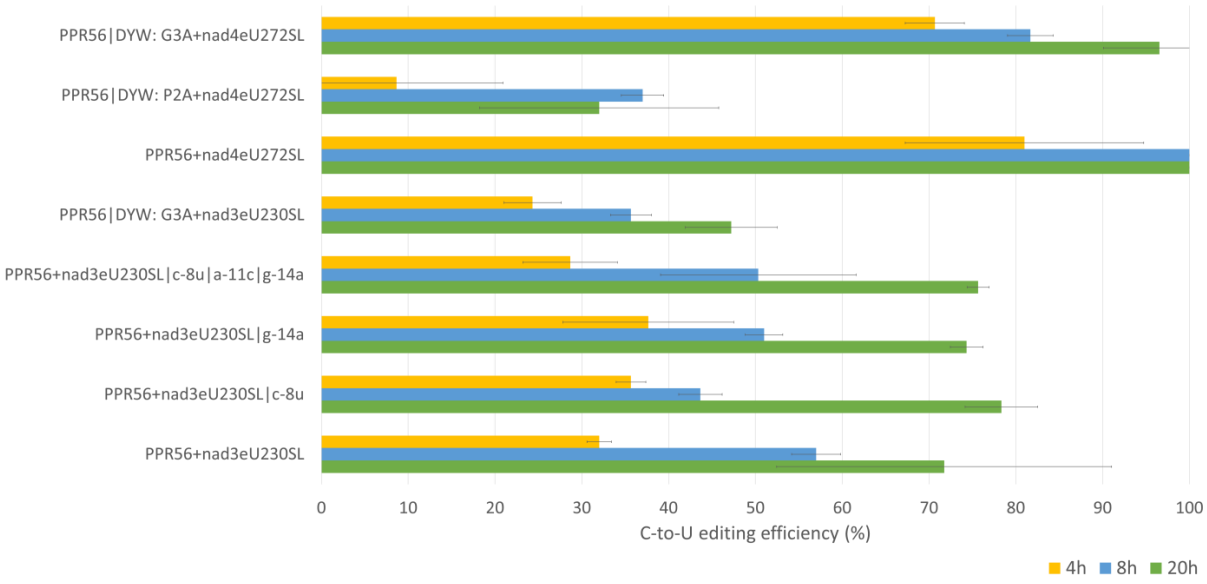
805 **A**. RNA editing was checked at an elevated temperature of 24° (orange bars) instead of the routinely
806 used 16°C (blue bars) for heterologous protein expression in the *E. coli* Rosetta 2 (DE3) arctic express
807 system for a selection of altogether twelve constructs. The elevated temperature of 24°C generally
808 disfavors RNA editing compared to incubation at 16° both on *nad4* and on *nad3* targets with the
809 interesting exception of the PPR56|DYW:P2A mutant. **B**. RNA editing was checked for eight selected
810 constructs also at shorter incubation times of only 4 h or 8 h, respectively, instead of the routinely
811 used 20 h of incubation at 16°C after induction of expression. A reduction of RNA editing is seen in all
812 cases of shorter incubation times except for the efficiently edited nad4 target, which already shows
813 >99% editing after 8 h of incubation.

814 ***A**. Temperature dependence of RNA editing in Escherichia coli.*



815

816 ***B**. Time-dependence of RNA editing in Escherichia coli.*



817

Supplementary Data

Supplementary Data 1. Full set of *E. coli* RNA editing assays.

Full table of results for all individual *E. coli* RNA editing assays including standard deviations. C-to-U RNA editing frequencies are given as 100% when no remaining cytidine signal was detectable upon sequencing of RT-PCR products.

Supplementary Data 2. RNA-seq data sets for analysis of off-targets in *Escherichia coli*.

RNA-seq datasets analyzed for C-to-U RNA editing off-targets. Separate tabs for the summary off-target lists for PPR56, PPR56|S-4TN>TD and PPR56|S-10TD>TN and 13 individual data sets for Jacusa variant calls (*E. coli* wild-type background control for reference, native PPR56 without co-delivered targets (2 replicates), with co-delivered nad3eU230SL target, nad4eU272SL target (2 replicates) and combined *nad4-nad3* target, PPR56|S-10TD>TN without or with co-delivered target nad4eU272SL or nad4eU272SL|-13, and PPR56|S-4TN>TD) analyzed in the course of this study.

Supplementary Data 3. Oligonucleotides.

Oligonucleotides used in this study. All oligonucleotides were synthesized by IDT (Integrated DNA technologies Europe, BVBA, Leuven, Belgium).

References

1. Shikanai T. RNA editing in plant organelles: machinery, physiological function and evolution. Cellular and Molecular Life Sciences. 2006;63: 698–708. doi:10.1007/s00018-005-5449-9

2. Small ID, Schallenberg-Rüdinger M, Takenaka M, Mireau H, Ostersetzer-Biran O. Plant organellar RNA editing: what 30 years of research has revealed. Plant Journal. 2020;101: 1040–1056. doi:10.1111/tpj.14578

3. Knoop V. C-to-U and U-to-C: RNA editing in plant organelles and beyond. Marquardt S, editor. J Exp Bot. 2022 [cited 3 Feb 2023]. doi:10.1093/jxb/erac488

4. Ichinose M, Sugita M. Substitutional RNA Editing in Plant Organelles. Methods Mol Biol. 2021;2181: 1–12. doi:10.1007/978-1-0716-0787-9_1

5. Naramoto S, Hata Y, Fujita T, Kyozuka J. The bryophytes Physcomitrium patens and Marchantia polymorpha as model systems for studying evolutionary cell and developmental biology in plants. Plant Cell. 2022;34: 228–246. doi:10.1093/PLCELL/KOAB218

6. Sun T, Bentolila S, Hanson MR. The unexpected diversity of plant organelle RNA editosomes. Trends Plant Sci. 2016;21: 926–973. doi:10.1016/j.tplants.2016.07.005

7. Takenaka M. How Complex Are the Editosomes in Plant Organelles? Mol Plant. 2014;7: 582–585. doi:10.1093/mp/sst170

8. Sandoval R, Boyd RD, Kiszter AN, Mirzakhanyan Y, Santibańez P, Gershon PD, et al. Stable native RIP9 complexes associate with C-to-U RNA editing activity, PPRs, RIPs, OZ1, ORRM1, and ISE2. The Plant Journal. 2019;99: 1116–1126. doi:10.1111/tpj.14384

9. Sugita M. An Overview of Pentatricopeptide Repeat (PPR) Proteins in the Moss *Physcomitrium patens* and Their Role in Organellar Gene Expression. Plants. 2022;11: 2279. doi:10.3390/plants11172279

10. Schallenberg-Rüdinger M, Knoop V. Coevolution of organelle RNA editing and nuclear specificity factors in early land plants. In: Rensing SA, editor. Genomes and Evolution of Charophytes, Bryophytes and Ferns Advances in Botanical Research, Vol 78. Amsterdam: Elsevier Academic Press; 2016. pp. 37–93. doi:10.1016/bs.abr.2016.01.002

11. Guillaumot D, Lopez-Obando M, Baudry K, Avon A, Rigaill G, Falcon de Longevialle A, et al. Two interacting PPR proteins are major Arabidopsis editing factors in plastid and mitochondria. Proceedings of the National Academy of Sciences of the USA. 2017;114: 8877–8882. doi:10.1073/pnas.1705780114

12. Andrés-Colás N, Zhu Q, Takenaka M, De Rybel B, Weijers D, Van Der Straeten D. Multiple PPR protein interactions are involved in the RNA editing system in Arabidopsis mitochondria and plastids. Proceedings of the National Academy of Sciences of the USA. 2017;114: 8883–8888. doi:10.1073/pnas.1705815114

13. Boussardon C, Salone V, Avon A, Berthome R, Hammani K, Okuda K, et al. Two interacting proteins are necessary for the editing of the *ndhD*-1 site in *Arabidopsis* plastids. Plant Cell. 2012;24: 3684–3694.

14. Boussardon C, Avon A, Kindgren P, Bond CS, Challenor M, Lurin C, et al. The cytidine deaminase signature HxE(x)nCxxC of DYW1 binds zinc and is necessary for RNA editing of ndhD-1. New Phytologist. 2014;203: 1090–1095. doi:10.1111/nph.12928

15. Wang Y, Li H, Huang Z-Q, Ma B, Yang Y-Z, Xiu Z-H, et al. Maize PPR-E proteins mediate RNA C-to-U editing in mitochondria by recruiting the *trans* deaminase PCW1. Plant Cell. 2023;35: 529–551. doi:10.1093/plcell/koac298

16. Malbert B, Burger M, Lopez-Obando M, Baudry K, Launay-Avon A, Härtel B, et al. 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. 2020;9: 280. doi:10.3390/plants9020280

17. Diaz MF, Bentolila S, Hayes ML, Hanson MR, Mulligan RM. A protein with an unusually short PPR domain, MEF8, affects editing at over 60 Arabidopsis mitochondrial C targets of RNA editing. The Plant Journal. 2017;92: 638–649. doi:10.1111/tpj.13709

18. Verbitskiy D, Zehrmann A, Härtel B, Brennicke A, Takenaka M. Two related RNA-editing proteins target the same sites in mitochondria of Arabidopsis thaliana. J Biol Chem. 2012;287: 38064–38072. doi:10.1074/jbc.M112.397992

19. Yang Y-ZZ, Liu X-YY, Tang J-JJ, Wang Y, Xu C, Tan B-CC. GRP23 plays a core role in E-type editosomes via interacting with MORFs and atypical PPR-DYWs in Arabidopsis mitochondria. Proceedings of the National Academy of Sciences of the USA. 2022;119: 2210978119. doi:10.1073/pnas

20. Oldenkott B, Burger M, Hein A-CC, Jörg A, Senkler J, Braun HP, et al. One C-to-U RNA editing site and two independently evolved editing factors: testing reciprocal complementation with DYW-type PPR proteins from the moss Physcomitrium (Physcomitrella) patens and the flowering plants Macadamia integrifolia and Arabidopsis th. Plant Cell. 2020;32: 2997–3018. doi:10.1105/TPC.20.00311

21. Sadamitsu A, Inoue Y, Sakakibara K, Tsubota H, Yamaguchi T, Deguchi H. The complete plastid genome sequence of the enigmatic moss , Takakia lepidozioides (Takakiopsida , Bryophyta): evolutionary perspectives on the largest collection of genes in mosses and the intensive RNA editing. Plant Mol Biol. 2021;107: 431–449. doi:10.1007/s11103-021-01214-z

22. Gerke P, Szövényi P, Neubauer A, Lenz H, Gutmann B, McDowell R, et al. Towards a plant model for enigmatic U-to-C RNA editing: the organelle genomes, transcriptomes, editomes and candidate RNA editing factors in the hornwort *Anthoceros agrestis*. New Phytologist. 2020;225: 1974–1992. doi:10.1111/nph.16297

23. Oldenkott B, Yang Y, Lesch E, Knoop V, Schallenberg-Rüdinger M. Plant-type pentatricopeptide repeat proteins with a DYW domain drive C-to-U RNA editing in *Escherichia coli*. Commun Biol. 2019;2: 85. doi:10.1038/s42003-019-0328-3

24. Lesch E, Schilling MT, Brenner S, Yang Y, Gruss OJ, Knoop V, et al. Plant mitochondrial RNA editing factors can perform targeted C-to-U editing of nuclear transcripts in human cells. Nucleic Acids Res. 2022;50: 9966–9983. doi:10.1093/NAR/GKAC752

25. Ohtani S, Ichinose M, Tasaki E, Aoki Y, Komura Y, Sugita M. Targeted gene disruption identifies three PPR-DYW proteins involved in RNA editing for five editing sites of the moss mitochondrial transcripts. Plant Cell Physiol. 2010;51: 1942–1949. doi:10.1093/pcp/pcq142

26. Rüdinger M, Funk HT, Rensing SA, Maier UG, Knoop V. RNA editing: only eleven sites are present in the *Physcomitrella patens* mitochondrial transcriptome and a universal nomenclature proposal. Molecular Genetics and Genomics. 2009;281: 473–481. doi:10.1007/s00438-009-0424-z

27. Lenz H, Rüdinger M, Volkmar U, Fischer S, Herres S, Grewe F, et al. Introducing the plant RNA editing prediction and analysis computer tool PREPACT and an update on RNA editing site nomenclature. Curr Genet. 2010;56: 189–201. doi:10.1007/s00294-009-0283-5

28. Yan J, Yao Y, Hong S, Yang Y, Shen C, Zhang Q, et al. Delineation of pentatricopeptide repeat codes for target RNA prediction. Nucleic Acids Res. 2019;47: 3728–3738. doi:10.1093/nar/gkz075

29. Barkan A, Rojas M, Fujii S, Yap A, Chong YS, Bond CS, et al. A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. PLoS Genet. 2012;8: e1002910. doi:10.1371/journal.pgen.1002910

30. Yagi Y, Hayashi S, Kobayashi K, Hirayama T, Nakamura T. Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. PLoS One. 2013;8: e57286.

31. Takenaka M, Zehrmann A, Brennicke A, Graichen K. Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. PLoS One. 2013;8: e65343. doi:10.1371/journal.pone.0065343

32. Gutmann B, Royan S, Schallenberg-Rüdinger M, Lenz H, Castleden IR, McDowell R, et al. The Expansion and Diversification of Pentatricopeptide Repeat RNA-Editing Factors in Plants. Mol Plant. 2020;13: 215–230. doi:10.1016/j.molp.2019.11.002

33. Takenaka M, Takenaka S, Barthel T, Frink B, Haag S, Verbitskiy D, et al. DYW domain structures imply an unusual regulation principle in plant organellar RNA editing catalysis. Nat Catal. 2021;4: 510–522. doi:10.1038/s41929-021-00633-x

34. Hayes ML, Giang K, Berhane B, Mulligan RM. Identification of two pentatricopeptide repeat genes required for RNA editing and zinc binding by C-terminal cytidine deaminase-like domains. J Biol Chem. 2013;288: 36519–36529. doi:10.1074/jbc.M113.485755

35. Hayes ML, Santibanez PI. A plant pentatricopeptide repeat protein with a DYW-deaminase domain is sufficient for catalyzing C-to-U RNA editing *in vitro*. J Biol Chem. 2020;295: 3497–3505. doi:10.1074/jbc.ra119.011790

36. Ichinose M, Sugita M. The DYW domains of pentatricopeptide repeat RNA editing factors contribute to discriminate target and non-target editing sites. Plant Cell Physiol. 2018;59: 1652–1659. doi:10.1093/pcp/pcy086

946 37. Lenz H, Hein A, Knoop V. Plant organelle RNA editing and its specificity factors: enhancements
947 of analyses and new database features in PREPACT 3.0. BMC Bioinformatics. 2018;19: 255.
948 doi:10.1186/s12859-018-2244-9

949 38. Ruwe H, Gutmann B, Schmitz-Linneweber C, Small I, Kindgren P, Schmitz-Linneweber C, et al.
950 The E domain of CRR2 participates in sequence-specific recognition of RNA in plastids. New
951 Phytologist. 2019;222: 218–229. doi:10.1111/nph.15578

952 39. Grewe F, Viehöver P, Weisshaar B, Knoop V. A *trans*-splicing group I intron and tRNA-
953 hyperediting in the mitochondrial genome of the lycophyte *Isoetes engelmannii*. Nucleic Acids
954 Res. 2009;37: 5093–5104.

955 40. Hecht J, Grewe F, Knoop V. Extreme RNA editing in coding islands and abundant
956 microsatellites in repeat sequences of *Selaginella moellendorffii* mitochondria: the root of
957 frequent plant mtDNA recombination in early tracheophytes. Genome Biol Evol. 2011;3: 344–
958 358. doi:10.1093/gbe/evr027

959 41. Zumkeller S, Polsakiewicz M, Knoop V. Rickettsial DNA and a trans-splicing rRNA group I intron
960 in the unorthodox mitogenome of the fern Haplopteris ensiformis. Commun Biol. 2023;6: 296.
961 doi:10.1038/s42003-023-04659-8

962 42. Salone V, Rüdinger M, Polsakiewicz M, Hoffmann B, Groth-Malonek M, Szurek B, et al. A
963 hypothesis on the identification of the editing enzyme in plant organelles. FEBS Lett.
964 2007;581: 4132–4138. doi:10.1016/j.febslet.2007.07.075

965 43. Iyer LM, Zhang D, Rogozin IB, Aravind L. Evolution of the deaminase fold and multiple origins
966 of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems.
967 Nucleic Acids Res. 2011;39: 9473–9497. doi:10.1093/nar/gkr691

968 44. Wagoner JA, Sun T, Lin L, Hanson MR. Cytidine deaminase motifs within the DYW domain of
969 two pentatricopeptide repeat-containing proteins are required for site-specific chloroplast
970 RNA editing. J Biol Chem. 2015;290: 2957–2968. doi:10.1074/jbc.M114.622084

971 45. Kobayashi T, Yagi Y, Nakamura T. Comprehensive prediction of target RNA editing sites for
972 PLS-class PPR proteins in *Arabidopsis thaliana*. Plant Cell Physiol. 2019;60: 862–874.
973 doi:10.1093/pcp/pcy251

974 46. Zehrmann A, Verbitskiy D, Härtel B, Brennicke A, Takenaka M. RNA editing competence of
975 trans-factor MEF1 is modulated by ecotype-specific differences but requires the DYW domain.
976 FEBS Lett. 2010;584: 4181–4186. doi:10.1016/j.febslet.2010.08.049

977 47. Ren RC, Lu X, Zhao YJ, Wei YM, Wang LL, Zhang L, et al. Pentatricopeptide repeat protein
978 DEK45 is required for mitochondrial function and kernel development in maize. J Exp Bot.
979 2019;70: 6163–6179. doi:10.1093/jxb/erz391

980 48. Matsuda T, Sugita M, Ichinose M. The L motifs of two moss pentatricopeptide repeat proteins
981 are involved in RNA editing but predominantly not in RNA recognition. PLoS One. 2020;15:
982 e0232366. doi:10.1371/journal.pone.0232366

983 49. Schallenberg-Rüdinger M, Kindgren P, Zehrmann A, Small I, Knoop V. A DYW-protein knockout
984 in *Physcomitrella* affects two closely spaced mitochondrial editing sites and causes a severe
985 developmental phenotype. The Plant Journal. 2013;76: 420–432. doi:10.1111/tpj.12304

986 50. McDermott JJ, Civic B, Barkan A. Effects of RNA structure and salt concentration on the
987 affinity and kinetics of interactions between pentatricopeptide repeat proteins and their RNA
988 ligands. Gerber AP, editor. PLoS One. 2018;13: e0209713. doi:10.1371/journal.pone.0209713

989 51. Xu C, Song S, Yang Y, Lu F, Zhang M, Sun F, et al. DEK46 performs C-to-U editing of a specific
990 site in mitochondrial nad7 introns that is critical for intron splicing and seed development in
991 maize. The Plant Journal. 2020;103: 1767–1782. doi:10.1111/tpj.14862

992 52. Verbitskiy D, van der Merwe JA, Zehrmann A, Brennicke A, Takenaka M. Multiple specificity
993 recognition motifs enhance plant mitochondrial RNA editing in vitro. J Biol Chem. 2008;283:
994 24374–24381. doi:10.1074/jbc.M803292200

995 53. Bernath-Levin K, Schmidberger J, Honkanen S, Gutmann B, Sun YK, Pullakhandam A, et al.
996 Cofactor-independent RNA editing by a synthetic S-type PPR protein. Synth Biol. 2022;7: 1–11.

997 54. Ichinose M, Kawabata M, Akaiwa Y, Shimajiri Y, Nakamura I, Tamai T, et al. U-to-C RNA editing
998 by synthetic PPR-DYW proteins in bacteria and human culture cells. Commun Biol. 2022;5:
999 968. doi:10.1038/s42003-022-03927-3

1000 55. Small I, Melonek J, Bohne A-V, Nickelsen J, Schmitz-Linneweber C. Plant Organellar RNA
1001 Maturation. Barta A, editor. Plant Cell. 2023 [cited 25 Feb 2023]. doi:10.1093/plcell/koad049

1002 56. Yan J, Zhang Q, Guan Z, Wang Q, Li L, Ruan F, et al. MORF9 increases the RNA-binding activity
1003 of PLS-type pentatricopeptide repeat protein in plastid RNA editing. Nat Plants. 2017;3:
1004 17037. doi:10.1038/nplants.2017.37

1005 57. Glass F, Härtel B, Zehrmann A, Verbitskiy D, Takenaka M. MEF13 Requires MORF3 and MORF8
1006 for RNA Editing at Eight Targets in Mitochondrial mRNAs in *Arabidopsis thaliana*. Mol Plant.
1007 2015;8: 1466–77. doi:10.1016/j.molp.2015.05.008

1008 58. Bayer-Császár E, Haag S, Jörg A, Glass F, Härtel B, Obata T, et al. The conserved domain in
1009 MORF proteins has distinct affinities to the PPR and E elements in PPR RNA editing factors.
1010 Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms. 2017;1860: 813–828.
1011 doi:10.1016/j.bbagrm.2017.05.004

1012 59. Zehrmann A, Härtel B, Glass F, Bayer-Császár E, Obata T, Meyer E, et al. Selective homo- and
1013 heteromer interactions between the Multiple Organellar RNA Editing Factor (MORF) proteins
1014 in *Arabidopsis thaliana*. Journal of Biological Chemistry. 2015;290: 6445–6456.
1015 doi:10.1074/jbc.M114.602086

1016 60. Miranda RG, McDermott JJ, Barkan A. RNA-binding specificity landscapes of designer
1017 pentatricopeptide repeat proteins elucidate principles of PPR-RNA interactions. Nucleic Acids
1018 Res. 2018;46: 2613–2623. doi:10.1093/nar/gkx1288

1019 61. Shen C, Zhang D, Guan Z, Liu Y, Yang Z, Yang Y, et al. Structural basis for specific single-
1020 stranded RNA recognition by designer pentatricopeptide repeat proteins. Nat Commun.
1021 2016;7: 11285. doi:10.1038/ncomms11285

1022 62. McDermott JJ, Watkins KP, Williams-Carrier R, Barkan A. Ribonucleoprotein Capture by in Vivo
1023 Expression of a Designer Pentatricopeptide Repeat Protein in Arabidopsis. Plant Cell. 2019;31:
1024 1723–1733. doi:10.1105/tpc.19.00177

1025 63. Filipovska A, Rackham O. Designer RNA-binding proteins: New tools for manipulating the
1026 transcriptome. RNA Biol. 2011;8: 978–983. doi:10.4161/rna.8.6.17907

1027 64. Shen C, Wang X, Liu Y, Li Q, Yang Z, Yan N, et al. Specific RNA Recognition by Designer
1028 Pentatricopeptide Repeat Protein. Mol Plant. 2015;8: 667–670.
1029 doi:10.1016/j.molp.2015.01.001

1030 65. Manavski N, Mathieu S, Rojas M, Méteignier L-V, Brachmann A, Barkan A, et al. *In vivo*
1031 stabilization of endogenous chloroplast RNAs by customized artificial pentatricopeptide
1032 repeat proteins. Nucleic Acids Res. 2021;49: 5985–5997. doi:10.1093/nar/gkab390

1033 66. Royan S, Gutmann B, Colas des Francs-Small C, Honkanen S, Schmidberger J, Soet A, et al. A
1034 synthetic RNA editing factor edits its target site in chloroplasts and bacteria. Commun Biol.
1035 2021;4: 1–10. doi:10.1038/s42003-021-02062-9

1036 67. McDowell R, Small I, Bond CS. Synthetic PPR proteins as tools for sequence-specific targeting
1037 of RNA. Methods. 2022;208: 19–26. doi:10.1016/j.ymeth.2022.10.003

1038 68. Kapust RB, Waugh DS. *Escherichia coli* maltose-binding protein is uncommonly effective at
1039 promoting the solubility of polypeptides to which it is fused. Protein Science. 1999;8: 1668–
1040 1674. doi:10.1110/ps.8.8.1668

1041 69. Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi S-H, et al. Genome Sequences of Escherichia
1042 coli B strains REL606 and BL21(DE3). J Mol Biol. 2009;394: 644–652.
1043 doi:10.1016/j.jmb.2009.09.052

1044 70. Wu TD, Reeder J, Lawrence M, Becker G, Brauer MJ. GMAP and GSNAP for genomic sequence
1045 alignment: enhancements to speed, accuracy, and functionality. In: Mathé E, Davis S, editors.
1046 Statistical Genomics Methods in Molecular Biology. New York, NY: Humana Press; 2016. pp.
1047 283–334. doi:10.1007/978-1-4939-3578-9_15

1048 71. Picardi E, Pesole G. REDItools: high-throughput RNA editing detection made easy.
1049 Bioinformatics. 2013;29: 1813–1814. doi:10.1093/bioinformatics/btt287

1050 72. Piechotta M, Wyler E, Ohler U, Landthaler M, Dieterich C. JACUSA: site-specific identification
1051 of RNA editing events from replicate sequencing data. BMC Bioinformatics. 2017;18: 7.
1052 doi:10.1186/s12859-016-1432-8

1053 73. Hein A, Knoop V. Expected and unexpected evolution of plant RNA editing factors CLB19,
1054 CRR28 and RARE1: retention of CLB19 despite a phylogenetically deep loss of its two known
1055 editing targets in Poaceae. BMC Evol Biol. 2018;18: 85. doi:10.1186/s12862-018-1203-4

1056 74. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome
1057 Res. 2004;14: 1188–1190.

1058 75. Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al.
1059 ViennaRNA Package 2.0. Algorithms for Molecular Biology. 2011;6: 26. doi:10.1186/1748-
1060 7188-6-26

1061

3.3. Yang et al 2023b (submitted)

Yingying Yang, Bastian Oldenkott, Shyam Ramanathan, Elena Lesch, Mizuki Takenaka, Mareike Schallenberg-Rüdinger and Volker Knoop (2023)

Different DYW cytidine deaminase domains strongly expand or restrict the flexibility of chimeric plant C-to-U RNA editing factors to address targets

DYW-type PPR proteins act as editing factors and contain TPR-like E1 and E2 domains between the PLS stretch and DYW domain. Although the TPR-like structure might suggest that they contribute to protein-protein interactions, the exact function of the E domains is unknown. To investigate the function of E domains and compatibility between PPR proteins, PPR56 and PPR65 with a variable chimera point in C-terminal extensions were tested in the *E. coli* system. PPR65 chimera with different versions of the C-terminal extensions of PPR56 could not edit any of the supplied targets, while the opposite chimera, PPR56 with the DYW domain of PPR65, could edit the nad4 target to 78%. Comparing the DYW domain, PPR56 holds a unique sequence compared with other editing factors in *P. patens*, especially in the gating domain suggested by Takenaka et al 2020. Replacing the α1 helix part of PPR56 with PPR65 did not significantly affect editing in the tandem target *nad4* + *nad3*. The chimera of PPR56 with different versions of the DYW domain of PPR45 could edit most of the supplied targets in the *E. coli* system and can edit the *nad4* target in human IMR cells to 54%. Furthermore, the DYW domain of OTP86 from *A. thaliana* was fused with PPR56 and it can edit both the *nad4* and *nad3* targets like native PPR56 in the *E. coli* system. In human cells, the *nad4* target could be edited to 58% with the DYW domain of OTP86. Replacing the gating domain of OTP86 in PPR56 slightly reduced editing of the *nad4* and *nad3* targets

Different chimeras of PPR56 have been sent for RNA-seq for off-target studies. Given that native PPR56 holds 133 off-targets, while native PPR65 holds 6, the PPR56/PPR65 chimera has only 6 obtained. Comparing the nucleotide preferences, it is interesting to note that nucleotides in the editing site positions -1 to -3 show the same preference as PPR56, although these positions are likely to be affected

when changing the C-terminal extensions. These similarities with PPR56 continue until position -10, while position -11 and -14 are dominated by cytidine and uridine as in PPR65 off-targets, respectively. Note that these two positions correspond to L-motifs. Position -13 corresponding to S motif shows adenosine domination (S-10TN in PPR65) instead of TD:g for PPR56. While no off-targets were obtained with PPR56/PPR45 chimera, PPR56/OTP86 chimera holds 472 off-targets. There, the nucleotide selection is much more relaxed in positions -13, -12, -10, and -7. The nucleotides upstream of position -13 do not seem to follow PPR56. Among the 472 off-targets identified, 94 are shared with native PPR56, 3 are shared with PPR56-PPR65 chimera, and another three are shared in all three data sets. In addition, the shared off-targets seem to be edited more efficiently by PPR56/OPT86.

This work is a collaboration with Prof. Mizuki Takenaka's group, and part of the study has been included in the PhD thesis of Bastian Oldenkott (Oldenkott, 2020). I designed and cloned most of the constructs in the *E. coli* system, analyzed the RNA-seq data, and mainly participated in the downstream analysis of the off-target candidate sites.

Different DYW cytidine deaminase domains strongly expand or restrict the flexibility of chimeric plant C-to-U RNA editing factors to address targets

Yingying Yang[1], Bastian Oldenkott[1], Shyam Ramanathan[1], Elena Lesch[1], Mizuki Takenaka[2], Mareike Schallenberg-Rüdinger[1]* and Volker Knoop[1]*

[1] IZMB – Institut für Zelluläre und Molekulare Botanik, Abteilung Molekulare Evolution, Universität Bonn, Kirschallee 1, D-53115 Bonn, Germany
[2] Department of Botany Graduate School of Science, Kyoto University, Oiwake-cho, Sakyo-ku, Kyoto, 606-8502, Japan

*Corresponding authors
mareike.ruedinger@uni-bonn.de, Phone: +49 228 73-6464
volker.knoop@uni-bonn.de, Phone: +49 228 73-6466

ORCID-IDs
Yingying Yang: 0000-0002-5518-0307
Bastian Oldenkott: 0000-0002-2513-2443
Shyam Ramanathan: 0009-0000-7582-9452
Elena Lesch: 0000-0002-3449-3929
Mizuki Takenaka: 0000-0002-3242-5092
Mareike Schallenberg-Rüdinger: 0000-0002-6874-4722
Volker Knoop: 0000-0002-8485-9423

Key words: *Physcomitrium patens,* mitochondrial RNA editing factor, PLS-type pentatricopeptide repeats, DYW-type cytidine deaminase, PPR-RNA recognition code, *Escherichia coli*

Abstract

The protein factors for the specific C-to-U RNA editing events in plant mitochondria and chloroplasts possess unique arrays of RNA-binding pentatricopeptide repeats (PPRs) linked to carboxy-terminal cytidine deaminase DYW domains via the extension motifs E1 and E2. The E1 and E2 motifs have distant similarity to tetratricopeptide repeats (TPRs) known to mediate protein-protein interactions but their precise function is unclear. Here, we investigate the tolerance of PPR56 and PPR65, two functionally characterized RNA editing factors of the moss *Physcomitrium patens,* for creation of chimeras by variably replacing C-terminal protein regions. Making use of a heterologous RNA editing assay system in *Escherichia coli* we find that heterologous DYW domains can strongly restrict or widen the spectrum of off-targets in the bacterial transcriptome for PPR56. Surprisingly, the data suggest that these changes are not only caused by preference of a given DYW domain for the immediate sequence environment of the cytidine to be edited, but also by a long-range impact on the nucleotide selectivity of the upstream PPRs.

Introduction

C-to-U RNA editing in chloroplasts and mitochondria is universally present in all land plants with the unique exception of the marchantiid subclass of complex-thalloid liverworts [1–4]. Despite this wide evolutionary conservation, the molecular machinery for the site-specific deamination of cytidines to create uridines varies in complexity between mosses and seed plants [5–7]. RNA editing factors in the model moss *Physcomitrium patens* are single proteins combining the functions of specific RNA target recognition and a cytidine deaminase function [8,9]. The typical makeup of such proteins includes an N-terminal signal peptide for import into chloroplasts or mitochondria, an array of pentatricopeptide repeats (PPRs) for targeting a specific RNA sequence, the E1 and E2 "extension" motifs and the DYW domain typically exhibiting the eponymous aspartate-tyrosine-tryptophan tripeptide at their carboxy-terminus.

The terminal DYW domain including a highly conserved zinc-coordination site in its catalytic center was suggested early as the cytidine deaminase domain and this function is meanwhile clearly established beyond doubt [10–17]. Similarly, there is a core concept to understand how the arrays of PPRs bind to their RNA target sequences in a one-repeat-per-nucleotide manner following a PPR-RNA binding code [18–21]. However, more understanding is needed here since plant RNA editing factors are characterized by different types of PPRs where L (long) and S (short) variants are present in addition to the canonical P-type PPRs. Yet further PPR variants denoted as "LL" and "SS" have meantime been characterized after extensive analyses of new plant genome data, especially for hornworts, lycophytes and ferns [22,23]. Essentially, in P- and S-type PPRs, amino acid positions 5 and last (L) are crucial to determine ribonucleotide matches (see Suppl. Fig. 1). Threonine or serine (T/S) in position 5 favors purines (A or G) whereas asparagine (N) favors pyrimidines (C or U). In the last (L) position of P- and S-type PPRs, asparagine (N) favors amino-bases (A or C) while aspartate (D) prefers keto-bases (G or U).

Much less understood is the role of the two TPR-like motifs E1 and E2 linking the N-terminal PPR arrays to the carboxyterminal DYW domain. They may, however, play important roles when the ancestral makeup of RNA editing factors comprising the necessary functionalities for RNA target

recognition and cytidine deamination in a single polypeptide get separated during evolution. In contrast to the moss RNA editing factors, only ca. half of the meanwhile characterized RNA editing factors in angiosperms feature the complete suite of protein domains outlined above in one single protein. The first characterized mitochondrial RNA editing factor MEF1 of *Arabidopsis thaliana* is an example for a PLS-type PPR array linked via E1 and E2 motifs to a complete carboxyterminal DYW domain [24]. In contrast, the first characterized chloroplast RNA editing factor in *Arabidopsis* identified earlier is a typical example for truncation behind the E2 motif, relying on a DYW domain to be supplied *in trans* [25,26].

Making things yet more complex in flowering plants are multiple additional editing "helper" factors that are necessary for a complete functional editosome. Featuring prominently among those are the MORFs (Multiple Organelle RNA Factors), alternatively labelled as RIPs (RNA editing Interacting Proteins) – proteins equipped with evolutionary unique MORF domains [27–29]. MORF proteins seem to be multifunctional in aiding both protein-protein interactions [30–32], but also in enhancing the binding of PPR stretches to target RNAs [33,34].

Unsurprisingly, the functional heterologous expression of plant RNA editing factors in a bacterial setup [11] and in cultured human cells [13] has initially succeeded with RNA editing factors from the model moss *Physcomitrium patens*, likely representing an evolutionary ancestral state and not relying on the help of additional proteins. Notably, MORFs/RIPs present in the complex angiosperm editomes [35–41] are not encoded in the *P. patens* genome.

The modular makeup of plant RNA editing factors might suggest that recombination of domains in chimeric proteins could be straightforward. Indeed, the DYW domain of *Arabidopsis* chloroplast RNA editing factor OTP86, for which a crystal structure has recently been obtained, could replace the one of *Physcomitrium* mitochondrial editing factor PPR56 for functional analyses in *E. coli* [12]. However, this functional chimera appeared to be more of an exception than the rule since testing other RNA editing factor chimeras had only moderate success both *in planta* [42] or in the heterologous bacterial system [43].

It remains unclear at present whether some domain combinations of different plant RNA editing factors are incompatible *per se* [42] or whether the specific chimeras simply fail to operate on a given target RNA. To further explore this issue we made use of the *Escherichia coli* RNA editing system and the simplicity of two functional RNA *Physcomitrium* mitochondrial RNA editing factors PPR56 and PPR65 (Suppl. Fig. 1) not relying on additional helper proteins for functionality [11,44,45]. An additional advantage of the heterologous bacterial setup is that it allows for scoring of off-targets in the bacterial transcriptome that may be hit by chimeric RNA editing factors even when no RNA editing could be detected at co-delivered target sequences in the RNA editing assays.

The here presented data for some RNA editing factor chimeras show that not only the observed RNA editing efficiencies are lower but also that the corresponding off-target data sets are much smaller when compared to the native editing factor. We conclude that artificial recombination of PPR-type editing factors connecting PPR arrays with heterologous E1 and E2 motifs and/or a different DYW domain can impair functionality, possibly by a reduced overall protein flexibility necessary for enzymatic C-to-U conversion. However, we find an extraordinary increase to more than 400 off-targets for a chimera of the moss mitochondrial RNA editing factor PPR56 equipped with the DYW domain of angiosperm chloroplast editing factor OTP86. Intriguingly, this increase in off-targets seems to result from relaxed selectivity of the upstream PPR array indicating enhanced flexibility via long-range interaction in the polypeptide. Evidently, the success for creating functional RNA editing factor chimeras is neither dictated by their native organelle environment nor by the phylogenetic distance but rather by yet to be understood intramolecular interactions.

Results

Recombining RNA editing factors PPR56 and PPR65

We started the creation of recombinant chimeras from *Physcomitrium patens* RNA editing factors PPR56 and PPR65 (Suppl. Fig. 1), which had proven to successfully edit their native targets in a heterologous RNA editing assay setup in *Escherichia coli* [11]. A series of targeted point mutations in PPR56 and its two native targets nad3eU230SL and nad4eU272SL has recently revealed a remarkable

sensitivity of nad3eU230SL, but a notable resilience of the nad4eU272SL target against sequence changes [46]. For full comparability, all protein constructs reported here and tested in *E. coli* were cloned in fusion with an N-terminal maltose binding protein (MBP) linked via an attB-TEV sequence as previously described [11,46].

For the designation of protein chimeras we will here use the slash (/) to indicate recombined RNA editing factor components with sequence extensions defined by the well-conserved consensus profiles of the C-terminal domains (Suppl. Fig. 2). Protein sequence continuities will be indicated by hyphens behind the equal (=) symbol or given with the respective amino acid sequence extensions for partial domains behind a colon. Further amino- or carboxy-terminal extensions are given with 'n' or 'c' followed by the number of amino acids or with capital letters N or C when extending to the native N- or C-terminal protein ends, respectively. For example, the designation PPR56=n14L14-E2/PPR65=DYW describes a protein chimera including the CDS of PPR56 starting 14 amino acids upstream of its most N-terminal PPR L-14 and extending to the end of its E2 motif, followed by a downstream fusion to the full DYW domain of PPR65. As a more complex case, PPR65=n15P15-DYW:1-13/PPR56=DYW:14-67/PPR65=DYW:68-C has an internal part of the PPR56 DYW domain representing the gating domain (see suppl. fig. 2 and below) transplanted into PPR65 (Fig. 1A).

Despite PPR56 and PPR65 successfully operating on their native targets in heterologous setups [11,13], a series of constructs reciprocally replacing the DYW domains of PPR56 and PPR65 showed no success in most cases (Fig. 1). The fusions of PPR65 with the DYW domain of PPR56 either including the E1 and E2 motifs of the former or of the latter revealed no RNA editing on the native ccmFCeU103PS target of PPR65 or on the two native targets of PPR56 (Fig. 1A), even when cloned in tandem combinations which were recently found to enhance RNA editing [46]. Likewise, a chimera in which only that part of the PPR56 DYW domain, which is now defined as the gating domain (see suppl. fig. 2) likely having a regulatory role [12], was used to replace the one of PPR65 proved to be non-functional for editing the ccmFCeU103PS target (Fig. 1A).

More success was achieved for reciprocal fusions, replacing C-terminal regions of PPR56 with corresponding sequences of PPR65 (Fig. 1B). While no editing was obtained when the fusion point was located between the PPR array of PPR56 and the E1 motif of PPR65, the chimera with a fusion point between the E2 motif of PPR56 and the DYW domain of PPR65 revealed 78% of RNA editing of the generally more robust nad4eU272SL target of PPR56 [11,44,46]. No RNA editing could be detected for an artificial hybrid target replacing positions -3 to +5 with the corresponding nucleotides of the ccmFCeU103PS editing site to provide a native environment for the DYW domain of PPR65 (Fig. 1B). To further explore the role of the now suggested gating domain (see suppl. fig. 2), we reciprocally replaced the region likely forming its conserved α1 helix in PPR56 with the corresponding one of PPR65. Intriguingly, this chimera revealed strong RNA editing both at the nad3eU230SL target and at the nad4eU272SL target cloned in a tandem arrangement (Fig. 1B).

Chimeras of PPR56 and PPR45

PPR45 is the only RNA editing factor in *Physcomitrium patens* chloroplasts, creating the start codon of *rps14* by RNA editing (cp_rps14eU2TM), also causing a likely collateral minor editing of a cytidine directly upstream of the start, cp_rps14eU-1 [47,48]. The PPR array of PPR45 does not reveal a good match with either target. We selected the DYW domain of PPR45 for testing chimeras because previous experimentation had shown a PPR56/PPR45 fusion to be functional *in planta* [42]. Creation of that chimera was based on a previous, now obsolete, definition of a shorter DYW domain behind previously defined 'E' and 'E+' motifs [49]. For comparability, we replicated the creation of this recombinant protein together with one using the end of the now defined E2 motif as the fusion point (Fig. 2), which had proven to be successful for the PPR56-PPR65 chimera (Fig. 1B). We found low RNA editing activity at the *nad4* target for the "old" construct with the N-terminally truncated DYW domain – with only 9% of C-to-U conversion in *E. coli* much more weakly than previously observed *in planta* [42]. No editing was found for the generally more weakly edited nad3 target of PPR56 for this chimera (Fig. 2). We used the opportunity offered by the *E. coli* assay setup to test an artificial *nad4/rps14* hybrid target replacing positions -3 to +5 relative to the nad4eU272SL editing site with the corresponding positions of the

cp_rps14eU2TM site. Indeed, this construct offering a somewhat more native targeting environment for the DYW domain of PPR45 revealed 28% of RNA editing (Fig. 2).

Testing the new chimeric construct with the fusion point between the E2 motif of PPR56 and the DYW domain of PPR45 resulted in much increased RNA editing efficiencies of 51% at the nad3eU230SL target of PPR56 and more than 99% of C-to-U conversion for the nad4eU272SL and 93% for the hybrid *nad4/rps14* target (Fig. 2).

Chimeras of PPR56 and OTP86

The structure of the DYW cytidine deaminase domain of OTP86, a chloroplast RNA editing factor in *Arabidopsis thaliana* [50], has been obtained by X-ray crystallography [12]. Moreover, a chimera of PPR56 in fusion with the OTP86 DYW domain proved to be functional in the *E. coli* assay system [12,43]. We independently created PPR56/OTP86 chimeras with a fusion point at the end of the E2 motif and, again, replaced only the gating domain of PPR56 with the one of OTP86 in an independent chimera (Fig. 3). Both chimeras proved to be functional for RNA editing of the two native targets of PPR56 with the latter construct transplanting the OTP86 gating domain region alone showing somewhat weaker performance despite the tandem cloning setup recently found to enhance observed RNA editing activities [46].

Functional chimeras also upon heterologous expression in human cells

Given the functionality of both the PPR56/PPR45 (Fig. 2) and the PPR56/OTP86 (Fig. 3) chimeras in *E. coli* we made use of the recently established setup for heterologous expression to check for their functionality also in human cell lines [13]. To that end we cloned the respective protein chimeras in fusion to an upstream EYFP tag behind the CMV promoter as previously reported. EYFP fluorescence was clearly detectable (Suppl. Fig. 3). We observed RNA editing efficiencies of 54% for the PPR56/PPR45 chimera (Fig. 2) and of 58% for the PPR56/OTP86 chimera (Fig. 3), respectively, upon expression in human IMR-90 cells (Suppl. Table 1). Hence, functionality of the two RNA editing factor chimeras is

very likely neither dependent on co-factors in the heterologous prokaryotic or eukaryotic environments nor on the addition of specific protein tags like MBP vs. EYFP.

Off-targets for PPR56, PPR65 and a chimeric construct

A significant benefit of the *E. coli* RNA editing assay setup is the possibility to scan for off-targets in the bacterial background transcriptome. Extending the initially obtained RNA-seq data [11], a set of 133 off-targets is now confirmed for PPR56 upon its expression in *Escherichia coli* [46]. In stark contrast, however, we can verify only six off-targets for PPR65 (Fig. 4). Intriguingly, we have here found PPR56 also to be more flexible against C-terminal sequence replacements than PPR65 (Fig. 1). Profiling of the off-target sequences excellently matches expectations following predictions from the PPR-RNA code (Fig. 4). A lacking selectivity of PPR P-6ND for uridines in PPR56 fits well with guanosine being present in the corresponding position -9 of the efficiently edited native nad4eU272SL target [46] and with the off-target scoring upon expression in human cell lines (Lesch et al. 2022). Similarly, we observe a lacking selectivity for adenosines by PPR P-9TN in PPR65 (Fig. 4).

As for the native PPR65, only six off-targets were identified for the now investigated chimera of PPR56 with the DYW domain of PPR65 (see Fig. 1B), indicating a strong selectivity exerted by the latter, in line with the equally low number of off-targets observed for PPR65 itself (Fig. 4). However, this is in no way reflected by favoring of off-targets featuring GU in positions -3 and -2 fitting the selectivity of native PPR65 (Fig. 4). Instead, the chimera's off-targets completely fit the profile of native PPR56 in the positions upstream of the edited cytidine. Yet more surprisingly, however, the selectivity for guanosine in position -13 opposite of PPR S-10TD is lost and replaced with one for adenine and this shift is accompanied by surprising selectivity for uridine immediately upstream in position -14 opposite of PPR L-11VE, an L-type PPR the roles of which in target selection are presently still unclear (Fig. 4). Evidently, the heterologous DYW domain of PPR65 unexpectedly exerts no selectivity immediately upstream of the editing site but rather affects target selectivity that should be dictated by the upstream PPR stretch of PPR56.

Many more off-targets for the PPR56-OTP86 chimera

Even more drastically than seen for the PPR56/PPR65 chimera discussed above, not a single off-target could be reliably identified in *E. coli* for the PPR56/PPR45 chimera with the truncated DYW domain (PPR56=n14L14-DYW:1-43/PPR45=DYW:44-C) that had shown low RNA editing activity on the tested targets (Fig. 2).

However, totally different picture emerged for the PPR56/OTP86 chimera revealing 472 off-targets, more than threefold the number found for the native PPR56 (Fig. 5). Neither the target conservation profile opposite of the PPR array of PPR56 nor positions immediately upstream of the editing site reveal any change in preferred nucleotide identities. However, the nucleotide selectivity is much relaxed for G in position -13, A in position -12, G in position -10 and A in position -7 opposite of perfectly matching PPRs S-10TD, P-9TN, S-7TD and S-4TN following the PPR-RNA code rules (Fig. 5). Accordingly, a long-range effect seems again to be caused by the terminal DYW domain for interaction of the upstream PPR array with the RNA targets, even when in exactly the opposite way of relaxing rather than restricting target recognition as in the case of the PPR56/PPR65 chimera (Fig. 4).

Shifts in off-target patterns and shifts in RNA editing efficiencies

Of altogether 472 off-targets identified for the PPR56/OTP86 chimera, 94 are shared with the native PPR56, three are shared with the PPR56/PPR65 chimera and another three are shared among all three data sets (Fig. 6A). This leaves 36 and 372 off-targets, respectively, exclusively observed for wild-type PPR56 or the PPR56/OTP86 chimera alone.

Checking upon editing sites shared between the data sets, we find that the PPR56/OTP86 chimera generally results in highest RNA editing at a given off-target (Figure 6B). Hence, the PPR56/OTP86 construct not only results in relaxed identification of off-targets but also in generally higher RNA editing efficiencies at those targets. The opposite is not true, however. Despite the significantly reduced total number of off-targets for the PPR56/PPR65 chimera, it displays higher editing efficiencies at off-targets trpBeU1157TI and recJeU425SF than the native PPR56 protein (Fig. 6B).

Discussion

Plant organelle RNA editing remains a puzzling molecular phenomenon on many levels. Not only is it unclear why it came into existence in the first place, evidently emerging with the earliest land plants. Equally puzzling is how the complex arrangements of RNA editing factors arose that combines their characteristic PLS-type PPR arrays with the E1 and E2 motifs and a carboxy-terminal DYW domain. There is no clear evidence yet that these individual components have been used as combinable building blocks in evolution to initially create the evolutionary ancestral, single-polypeptide RNA editing factors in plants. Rather conversely, the following evolution of the ancestral RNA editing factors gives many examples for their subsequent truncation or disintegration in the course of plant evolution [6]. The sporadic occurrences of plant-type RNA editing factors outside of land plants on the other hand rather points to horizontal gene transfer from plants into protists [51–54].

The meanwhile clearly defined modular makeup of plant RNA editing factors with a PLS-type PPR array terminating in a P2-L2-S2-type PPR triplet, followed by the TPR-like E1 and E2 motifs and ultimately a DYW cytidine deaminase domain [22,23] may *a priori* suggest an interchangeability of their carboxyterminal domains. Indeed, replacing the DYW domain of *Physcomitrium patens* RNA editing factor PPR78 with the one of PPR79 turned out to be functional *in planta* [55] and the DYW domain of PPR56 could even be replaced with the one of flowering plant chloroplast editing factor OTP86 for functional expression in *Escherichia coli* [12]. The maybe most impressive success for a functionally recombined RNA editing factor is the fusion of CRR4 and the “free-standing” DYW domain of DYW1, creating a protein chimera that was able to complement the *Arabidopsis thaliana* crr4/dyw1 double mutant defect for both proteins and correspondingly for RNA editing of the chloroplast editing target ndhDeU2TM [26].

In contrast, however, many other recombinant protein chimeras created similarly could not be shown to be functional [42,43,56,57] suggesting that there is no simple modular concept. The respective DYW domain and/or the respective E1 and E2 motifs may exert further selectivity for target recognition,

especially in the immediate environment of the RNA editing target site and several available data indeed support this conclusion [42,43,56,58].

While ever more plant RNA editing factors have been characterized over the last decades, we have to be careful that many detailed observations may not be generalized. For example, the conserved aspartate (D) - tyrosine (Y) - tryptophan (W) tripeptide at the very terminus of the DYW cytidine domain is quite conserved. Exchanging the central tyrosine with alanine, however, did not abolish editing in DYW1 [14] and the tyrosine side chain was found solvent-exposed in the recently determined crystal structure of DYW1 [59]. In stark contrast, however, exchanging a corresponding phenylalanine in the terminal DFW end of PPR65 to alanine abolished editing completely in *E. coli* whereas its exchange to the more conserved tyrosine had no effect [11]. Evidently, the terminal tripeptides of DYW domains may contribute to functionality in different ways, likely caused by specific interactions with their RNA targets. Similarly, even the exchange of alternatively conserved and chemically similar residues like arginine (R) or lysine (K) in position 71 of the DYW domain (see Suppl. Fig. 2C) results in lowered RNA editing activities for PPR65 and, *vice versa*, for PPR56 [11,46]. Success to create functional chimeras of RNA editing factors will evidently depend on the chosen point for protein fusion within the carboxyterminal P2-L2-S2-E1-E2-DYW arrangement, as is here clearly documented for the PPR56/PPR65 and the PPR56/PPR45 chimeras (Fig. 1B and Fig. 2). Different RNA editing efficiencies were similarly observed when the "stand-alone" DYW1 domain was differently fused to upstream PPR56 sequences as the "carrier protein" [43,59].

We here obtained highly variable outcomes for recombinant chimeric constructs involving components of PPR56 and PPR65, which were both successfully expressed in bacterial and eukaryotic setups using different N-terminal protein tag additions [11,13,46]. Astonishingly, both the upstream region of PPR65 including its PPR array (Fig. 1A) as well as the downstream region including its DYW domain (Fig. 1B) seem to be somewhat recalcitrant against creating functional chimeras or result in strongly restricted off-targeting. This observation is in line with the low number of off-targets upon heterologous expression of PPR65 and may indicate structural rigidity and a lack of flexibility of this

289 RNA editing factor. The observation of strikingly different numbers of off-targets is well corroborated
290 with the recent functional expression of PPR56 and PPR65 also in human cells [13].

291 Exactly the opposite is observed for PPR56 with its high number of off-targets and (variable)
292 flexibility towards replacement of its DYW domain against that of PPR65 as another moss
293 mitochondrial editing factor, against that of PPR45 as a moss chloroplast editing factor and finally
294 against that of OTP86 as an angiosperm chloroplast editing factor. Remarkably, OTP86 is part of a
295 multi-protein editosome complex involving many non-PPR proteins *in planta* [39]. The experimentation
296 with the PPR56/OTP86 chimeras in the heterologous bacterial and eukaryotic setups shows that
297 evidently no (plant-specific) *trans*-acting factors are necessary for the functionality of the OTP86 DYW
298 domain as a cytidine deaminase.

299 The most surprising result of the new data presented here are the highly different numbers of off-
300 targets obtained with the principally functional RNA editing chimeras. Transcript targeting is expected
301 to be mainly determined by the PPR array for the target sequence upstream of position -3 relative to
302 the editing site. Clearly, replacing a DYW domain with another one may cause incompatibilities in the
303 immediate vicinity of an initially investigated target. Nevertheless, on transcriptome level, one could
304 expect just a shift of off-target spectra for the protein chimeras, combining the preferences of the PPR
305 array with the one of the heterologous DYW domain. This *a priori* reasonable hypothesis is, however,
306 contradicted by our data with the striking outcome of much restricted off-targeting upon replacing the
307 DYW domain of PPR56 with those of PPR45 or PPR65 (Fig. 4), but a much extended set of off-targets
308 in stark contrast for the PPR56/OTP86 chimera (Fig. 5). While the small number of only six off-targets
309 identified for the PPR56/PPR65 chimera has to be kept in mind as a cautionary note for conclusions, it
310 is highly surprising that changes in the conservation profile are found for positions juxtaposed with the
311 upstream PPR array of PPR56. Instead, the positions in the immediate environments of the edited off-
312 target cytidines fully match the observations for PPR56 alone without an indication for selectivity
313 exerted by the PPR65 DYW domain (Fig. 4). For the large set of off-targets now identified for the

PPR56/OTP86 chimera in full contrast, we find a relaxation of conservations at four positions conceptually matching the PPR-RNA code excellently (Fig. 5).

Taken together, there may be a combination of factors to explain the new findings: Firstly, different DYW domains may have strongly differing enzymatic efficiencies allowing some to perform cytidine conversion even upon only ephemeral binding to a candidate target. Secondly, there may be yet unclear structural incompatibilities between a DYW domain and the upstream motifs. Finally, there may be long-range impacts of the terminal DYW domain enhancing or reducing flexibility of the upstream protein regions for interactions with appropriate RNA regions.

Care must in any case be taken to use best-comparable setups for further experimentation in heterologous systems. For example, a recent experimental modification of the original heterologous RNA editing assay system in *E. coli* [11] found that the placement of editing targets on separate, constitutively expressed transcripts resulted in overall lower editing efficiencies [43]. In contrast, our very recent study placing "weak" RNA editing targets alternatively into different locations and into variable tandem arrangements with "strong" targets found enhancement of RNA editing for the former [46]. In line with the above ideas, these findings may suggest a scanning mechanism of an RNA editing factor along its target transcript for which a structural flexibility may be highly beneficial indeed.

Materials and Methods

Molecular Cloning

Cloning for expression of *Physcomitrium patens* PPR56 variants and targets in *Escherichia coli* was based on vector pET41Kmod as outlined earlier [11]. Protein coding sequences with an upstream TEV cleavage site are cloned with the gateway system in fusion downstream of an N-terminal His_6 tag and the maltose-binding protein (MBP) for improved protein solubility [60] under control of a T7 promoter controlled by the lac operator. RNA editing target sequences were cloned behind the protein sequence upstream of a T7 terminator. Target sequences including flanking restriction sites were generated with synthesized oligonucleotides for both DNA strands (Integrated DNA technologies Europe, BVBA,

339 Leuven, Belgium) and ligated into dephosphorylated vectors after hybridization and phosphorylation.
340 All oligonucleotides used in the course of this work are listed in supplementary table 3. To create
341 chimeras different overlap extension PCR strategies were used. For expression in the human cell
342 editing essay, constructs were amplified from petG41K based plasmids, using a proof-reading Q5
343 polymerase (New England Biolabs) and cloned into the eukaryotic expression vector pEYFP-C1
344 (Clontech, TaKaRa) to create the final EYFP-tagged fusion protein coding sequences as previously
345 described [13].

346 ## Protein expression and analysis of RNA editing

347 The setup for the expression of different constructs in the heterologous *E. coli* system and the
348 downstream analysis of RNA editing was done as outlined previously [11]. Briefly, 25 mL of *E. coli*
349 Rosetta 2 (DE3) cultures were pre-grown in 100 mL Erlenmeyer flasks with baffles in LB medium
350 supplemented with 50 µM kanamycin, 17 µM chloramphenicol and 0.4 mM $ZnSO_4$ at 37°C until
351 reaching an OD_{600} of ca. 0.5. The bacterial cultures were then cooled on ice for 5 min. before adding
352 0.4 mM IPTG for induction of expression and incubation for 20 h at 16°C and 180 rpm. Expression of
353 chimera protein variants was routinely checked via SDS-PAGE gels. Expression of chimeric constructs
354 in human IMR-90 cell cytosol was executed as described previously [13]. Briefly, MEM (Pan
355 Biotechnologies) media, supplemented with 10% fetal calf serum and 1% Penicillin/Streptomycin was
356 supplemented with 25 µM zinc sulfate prior to PEI MAX (Polyscience) transfection and 20 h incubation.
357 To verify expression, cells were fixed on cover slips, nuclei stained with DAPI (4,6-diamidino-2-
358 phenylindole). The localization of EYFP-tagged PPR proteins was examined on a Zeiss AXIOPHOT
359 microscope using AxioVision software and ImageJ/Fiji version 1.53c for Windows.

360 ## Total RNA sequencing and off-target detection

361 To identify off-targets in the *E. coli* transcriptome, total RNA was prepared from individual replicates
362 by using the NucleoSpin RNA kit (Macherey Nagel), followed by DNase I treatment (Thermo Fisher
363 Scientific). Library preparation and Illumina sequencing (150 bp paired-end with NovaSeq 6000) was
364 done after rRNA depletion (TruSeq Stranded Total RNA with Ribo-Zero) by either Novagene or

Macrogen. To generate construct-specific DNA reference reads, the simulated reads (by ART MountRainier version 2016-06-05) of pET41Kmod with respective constructs and respective target sequences and the pRARE2 tRNA helper plasmid were merged with genomic DNA reads (WTDNA_SRR941832) of BL21(DE3) cells [61]. The datasets are summarized in supplementary table 3.

After quality check of the RNA-seq raw data by FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), the transcriptome reads were aligned with construct-specific DNA reads by GSNAP v2020/04/08 (Wu et al 2010) with standard settings [62]. The SNPs were called by JACUSA v1.3 [63]. RNA editing sites were selected only if clean RNA reads (T + C or G + A > 99%), clean DNA background (G/C > 98%), RNA read coverage of at least 30 and a total RNA editing rate of at least 1% were obtained. The SNPs were further restricted by a custom-made R script (established with kind help provided by S. Zumkeller) to select only those identified in at least two replicates of the respective construct, but exclude false positives called also in WT or in data sets of the respective other editing factors. The final RNA editing efficiency was calculated by adding up total RNA reads at a given site.

Acknowledgements

We gratefully acknowledge the computer resources and support provided by the Paderborn Center for Parallel Computing (PC^2) and the help of Philipp Gerke and Simon Zumkeller in our group for the establishment and help in further development of bioinformatic pipelines. We are grateful to Prof. Oliver Gruß, Dr. Max Schilling and Yannick Riedel at the Institute of Genetics at the University of Bonn for providing resources for experimentation with the human IMR-90 cell line. Finally, we wish to thank Ahmad Noureddine, Kira Ritzenhofen, Jingchan Xie and Sarah Brenner for contributions to molecular cloning and technical assistance, respectively.

Funding

Research was supported by grant SCHA 1952/2-2 from the Deutsche Forschungsgemeinschaft (DFG) to M.S.-R.

Author contributions

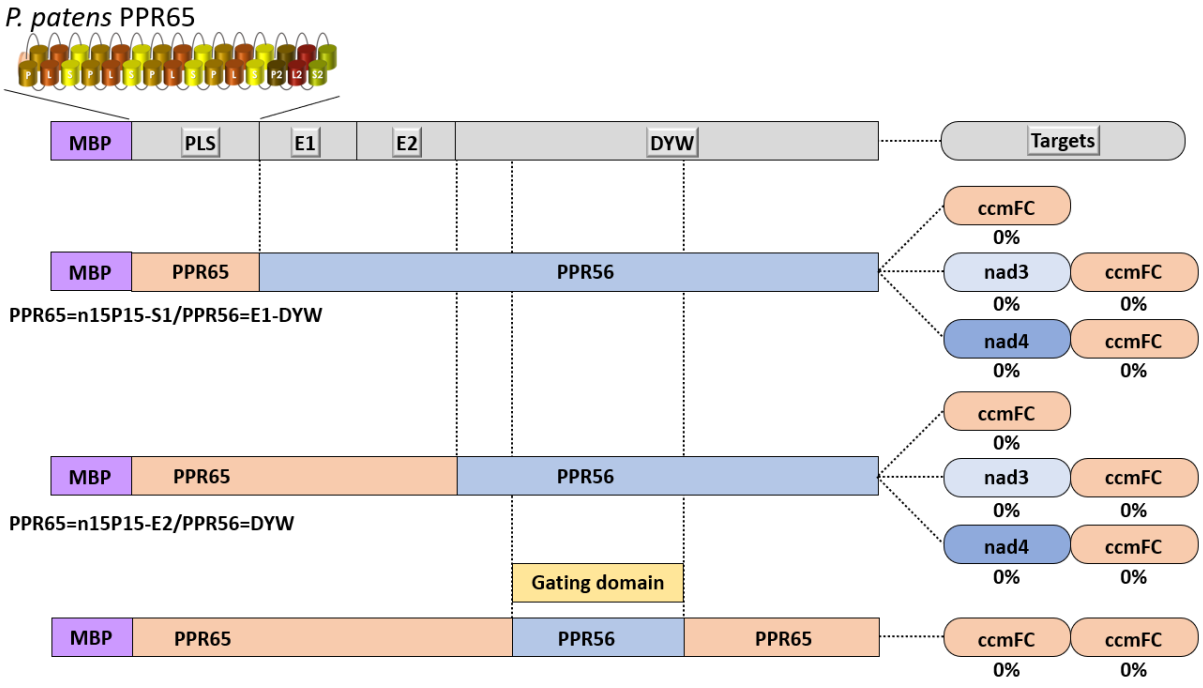
Y.Y. cloned most of the constructs and did most of the RNA editing analyses in the heterologous bacterial system and analyzed RNA-seq data. B.O. and S.R. contributed to construct cloning and RNA editing assays in *E. coli*. E.L. cloned chimeric RNA editing constructs for heterologous expression in human cells, performed the editing essays and helped optimizing the off-target analyses pipeline. M.T. provided the PPR56/OTP86 chimera for off-target analysis. M.S.R. and V.K. designed and supervised the study program and contributed to data analyses. Y.Y. and V.K. created figures. V.K. wrote the manuscript, which was read and approved by all co-authors.

Figure Legends

Figure 1. Reciprocal fusions of PPR arrays and DYW domains of PPR56 and PPR65 across their E1-E2 motifs.

A. PPR65 chimeras with DYW regions of PPR56.

The PLS-type PPR array of PPR65 was fused with downstream regions of PPR56 either between the terminal S2-type PPR and the E1 motif or between the E2 motif and the DYW domain. The N-terminal part of PPR65 including 15 amino acids upstream of its most N-terminal PPR (P-15) was cloned behind an upstream MBP-attB-TEV-sequence in pET41Kmod as previously described [11]. RNA editing activity on the three native targets of the editing factors was analyzed for either construct and no editing activity was likewise observed for a more complex chimera (PPR65=n15P15-DYW:1-13/PPR56=DYW:14-67/PPR65=DYW:68-C) having only the region of the gating domain of PPR65 replaced with the one of PPR56.



B. PPR56 chimeras with DYW regions of PPR65.

The PLS-type PPR array of PPR56 (see suppl. fig. 1) was fused with downstream regions of PPR65 as indicated. The N-terminal part of PPR56 including 14 amino acids upstream of its first PPR (L-14) was cloned behind an upstream MBP-attB-TEV-sequence in pET41Kmod as previously described [11]. RNA editing activity of 78% C-to-U conversion was detected for chimera PPR56=n14-L14-E2/PPR65=DYW on the native nad4eU272SL target of PPR56. Replacing the α1-helix in the gating domain of PPR56 with the corresponding sequence of PPR65 in construct PPR56=n14L14-DYW:1-27/PPR65=DYW:28-43/PPR56=DYW:44-C resulted in RNA editing of 99% and 92%, respectively, for the *nad4* and the *nad3* target of PPR56.

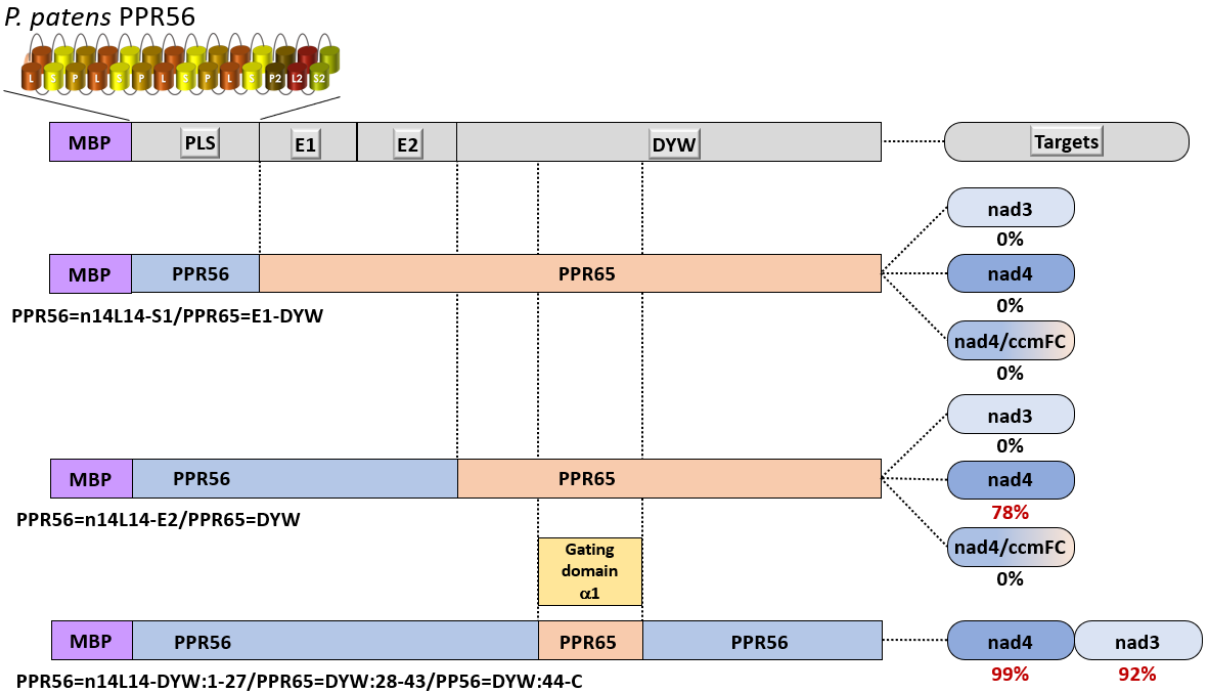


Figure 2. Chimeras of PPR56 and PPR45

Protein chimeras were created to replace C-terminal regions of PPR56 with corresponding sequences of chloroplast RNA editing factor PPR45, natively targeting cp_rps14eU2TM in *Physcomitrium patens*. Fusion points to create chimeras were chosen according to the previous concept of "E" and "E+" domains [42,49] and alternatively following the most recent definition of the TPR-like E1 and E2 motifs [22,23] and tested in *E. coli* (blue bacterial cell icons) on the two native targets (nad3eU230SL and nad4eU272SL) of PPR56 and a hybrid target replacing positions -3 to +5 with the native cp_rps14eU2TM target of PPR45. Rates of editing are indicated below the respective target sequences. The former construct (PPR56=n14L14-DYW:1-43/PPR45=DYW:44-C) was also investigated by RNA-seq analysis in *E. coli* revealing no off-targets. An additional construct with the latter fusion point and replacing the MBP with an EYFP tag (bottom) was tested in parallel in human cells (orange eukaryotic cell icons), revealing 54% of RNA editing at the nad4eU272SL target.

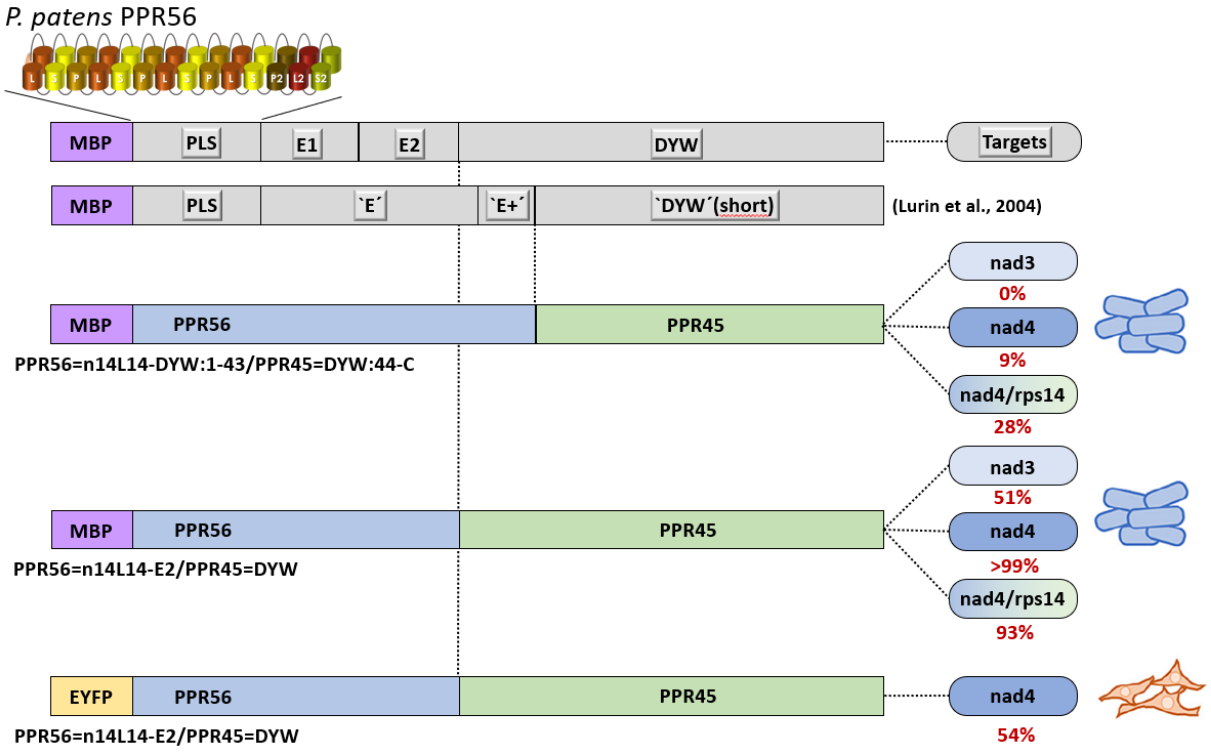


Figure 3. Chimeras of PPR56 and OTP86

Protein chimeras of PPR56 and OTP86 were tested for their editing capacities of the two native targets of PPR56. A chimera with the DYW domain of OTP86 fused behind the E2 motif of PPR56 (PPR56=n14L14-E2/OPT86=DYW) revealed RNA editing to be as efficient as the native PPR56 itself on both targets. An additional construct with the same fusion point and replacing the MBP with an EYFP tag (bottom) was tested in parallel in human cells (orange eukaryotic cell icons), revealing 58% of RNA editing at the nad4eU272SL target. A previously created chimera with the fusion point shifted by one amino acid (PPR56=n14L14-DYW:1/OTP86=DYW:2-C) had previously been tested on the *nad4* target in *E. coli* [12] and was now used for RNA-seq analysis revealing more than 400 off-targets in the bacteria (Fig. 5). Lower RNA editing efficiencies were observed in *E. coli* (blue bacterial cell icons) when only the gating domain of OTP86 was transplanted into the DYW domain of PPR56 (top).

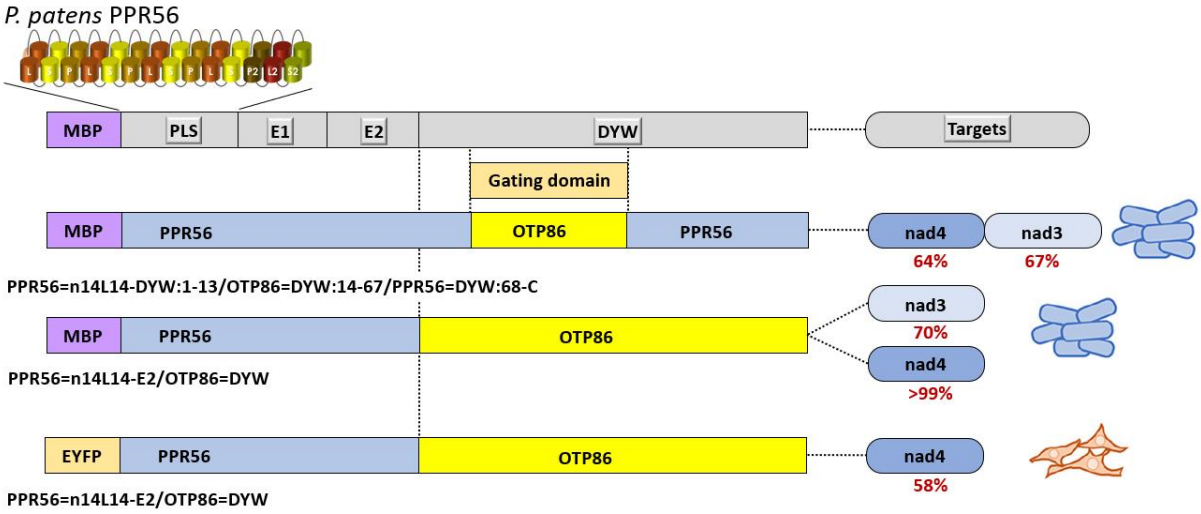


Figure 4. Off-targets for PPR56, PPR65 and a chimera

A total of 119 off-targets were identified for PPR56 (bottom, excluding 14 others requiring shifts for improved match with the PPR array for clarity) with an increased RNA-seq data set in the *Escherichia coli* transcriptome [46], but only six off-targets are confirmed for PPR65 (top). Consensus profiles were created with WebLogo [64] for the off-target sequence environments weighted with their respective editing efficiencies. Off-targets match expectations from the PPR-RNA binding code and fit the native targets of the two RNA editing factors with the exceptions of lacking selectivity by S-10TN for A in PPR65 (top) and by P-6ND for U in PPR56 (bottom), rather favoring G. The now investigated chimera PPR56=n14L14-E2/PPR65=DYW (middle) results in an equally low amount of only 6 off-targets. Yellow background shading highlights positions -3 to -1 where conservation profiles of the chimera match the one of PPR56. The blue arrows point to unexpected matches between the conservation profile of the PPR56/PPR65 chimera and native PPR65. The red arrow points to an unexpected shift of conservation from G to A in position -13 in the PPR56/PPR65 chimera.

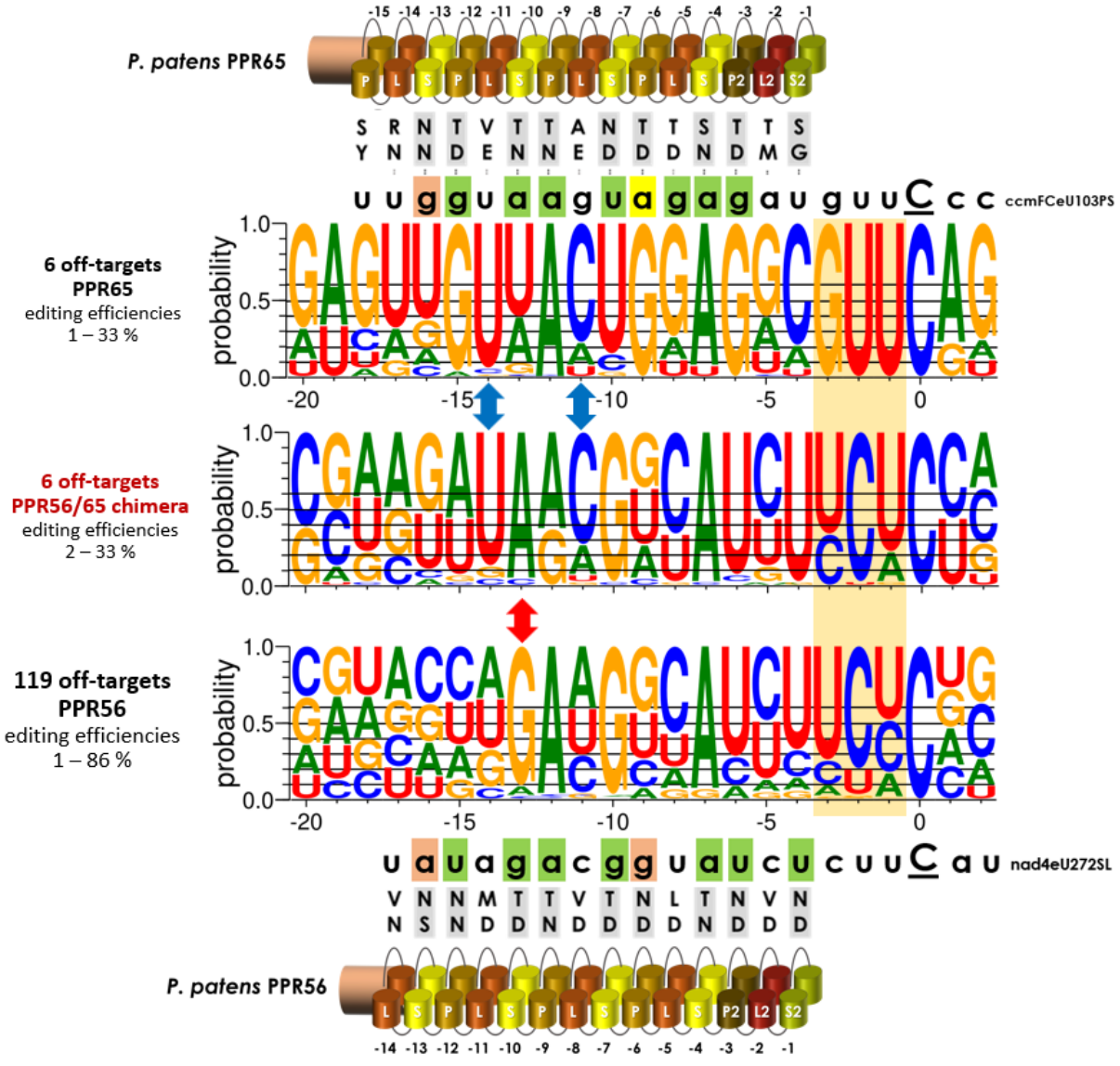


Figure 5. Off-targets of a PPR56-OTP86 chimera

A total of 472 off-targets were identified in RNA-seq data for the *E. coli* transcriptome after expression of the PPR56/OTP86 chimera PPR56=n14L14-DYW:1/OPT86=DYW:2-C. Consensus profiles were created from the sequences of 119 and 358 C-to-U RNA editing off-targets of PPR56 and the PPR56/OTP86 chimera, respectively, weighted with their respective editing efficiencies. Additional off-targets requiring nucleotide shifts for better PPR binding matches (14 and 114, respectively) were excluded for clarity (Suppl. Data 3). A reduced selectivity for nucleotides preferred by wild-type PPR56 is observed for positions -13 (G), -12 (A), -10 (G) and -7(A), all of which match the respective P- or S-type PPRs in PPR56 and this is also seen to a lower extent for positions -16 and -15 matching S-13NS and S-12NN. Slightly enhanced preferences matching the native target cp_rps14eU80SL of OTP86 (bottom) is observed for positions -14 (G), -8 (U), +1 (A) and +2 (A).

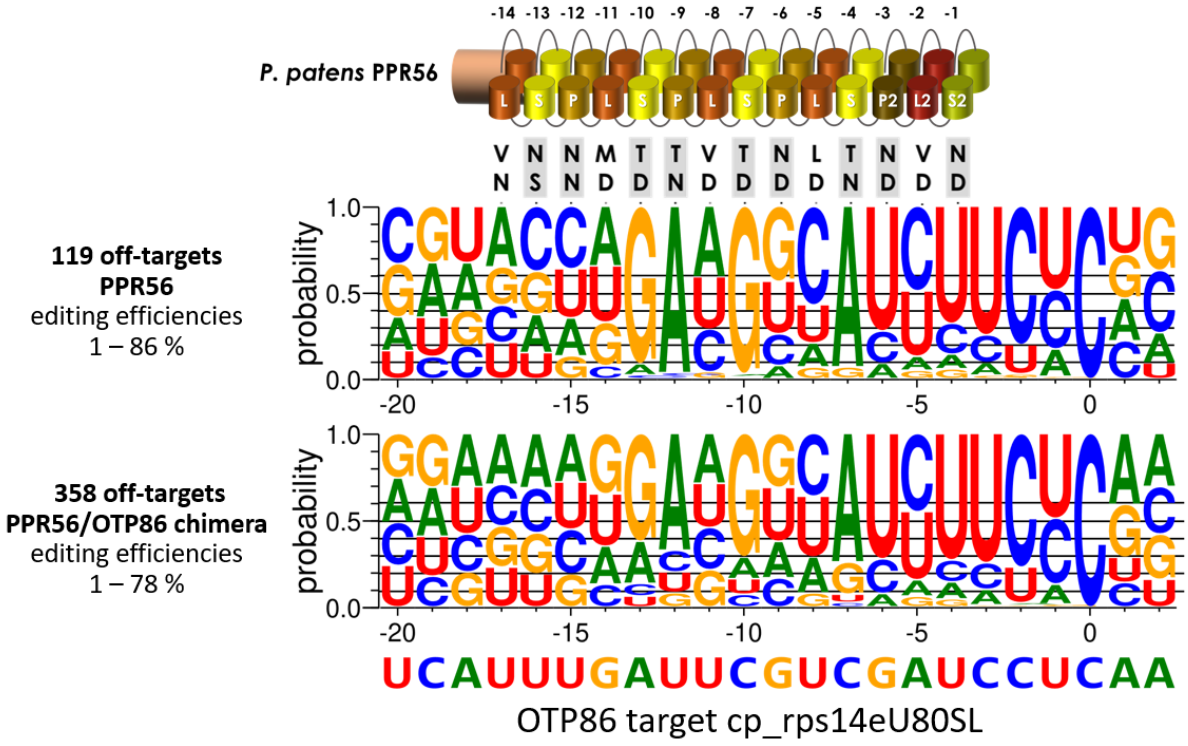
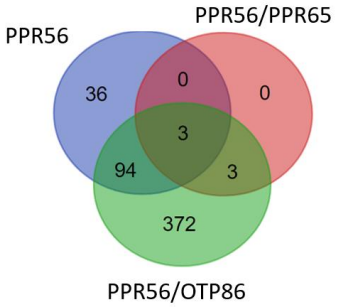


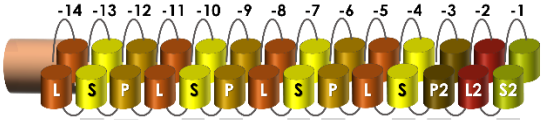
Figure 6. Off-targets of different PPR56 chimeras

A. The sets of off-targets identified in *E. coli* upon expression of native PPR56 (blue) and the chimeras having the DYW domain of PPR56 replaced with the ones of PPR65 (red) or OTP86 (green) are displayed as a Venn diagram (http://bioinformatics.psb.ugent.be/cgi-bin/liste/Venn/calculate_venn.htpl). Only three off-targets are shared between all three data sets whereas 372 off-targets are observed exclusively with the PPR56/OTP86 chimera. **B**. Off-targets shared between the PPR56/PPR65 chimera and the other data sets (top and middle) and the two off-targets with highest and lowest frequencies shared between the PPR56 and PPR56/OTP86 data set alone, respectively (bottom), are listed with the respective RNA editing efficiencies. Bold font highlights the majority of cases in which highest RNA editing is observed with the PPR56/OTP86 chimera. Asterisks indicate cases where RNA editing of 8.8% has recently been observed at the prfBeU-79 off-target for a PPR56|S-10TN mutant and of 2.6% at the rrsEeU-as-2 site for a PPR56|S-4TD mutant [46].

A. Off-targets Venn-Diagram



B. Shared off-targets in detail

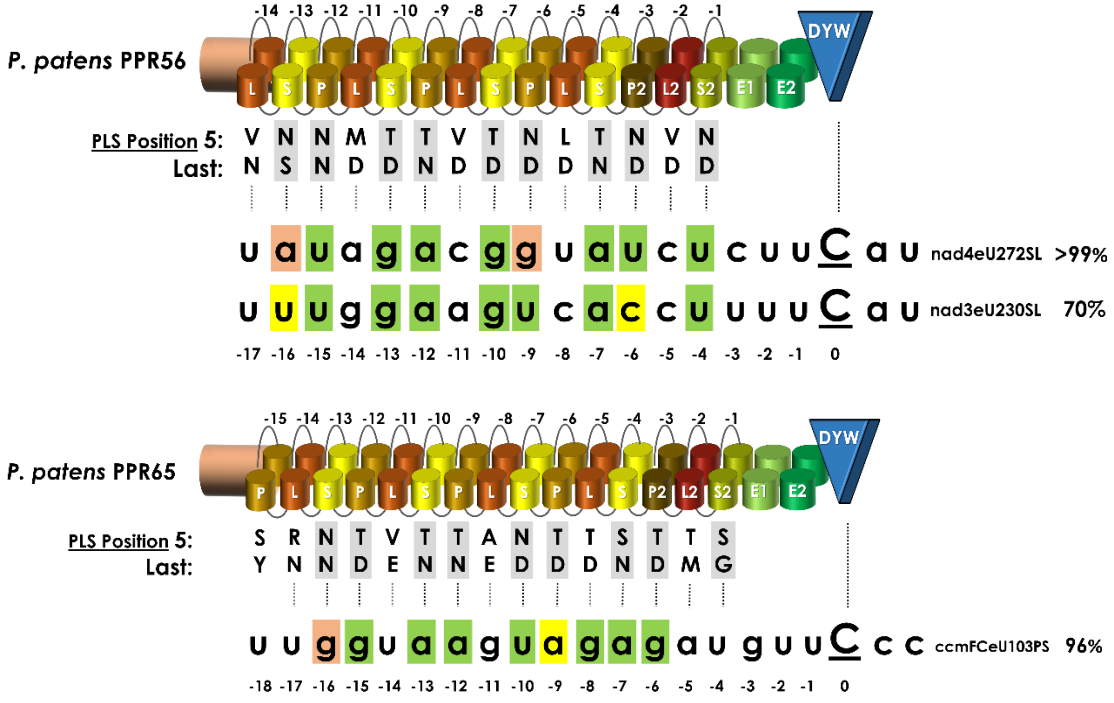


Pos	Name	V N	N S	N N	M D	T D	T N	V D	T D	N D	L D	T N	N D	V D	N D						PPR56	PPR56/ PPR65	PPR56/ OTP86
		-17	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	+1 +2			
989754	aspCeU1046SF	a	c	a	g	a	a	c	g	g	c	a	u	g	u	u	c	u	C	c u	4.3 %	3.5%	**15.0 %**
1302851	trpBeU1157TI	c	g	a	u	a	a	a	g	a	c	a	u	c	u	u	c	a	C	c g	8.7 %	14.3 %	**26.4 %**
2868542	recJeU425SF	g	g	a	u	a	a	c	g	g	u	a	u	u	u	c	c	u	C	c c	11.7 %	19.6 %	**38.5 %**
2867221	recJeU+12	a	u	u	u	a	g	c	g	u	c	a	u	c	u	u	c	u	C	u a	-*	21.4 %	**47.2 %**
3311684	rpsSeU13LF	a	a	g	c	c	a	u	g	c	c	a	c	g	u	u	c	u	C	u c	-	0.8 %	**9.7 %**
4116252	rrsEeU-as-2	c	g	a	a	g	c	a	g	c	a	a	g	c	u	g			C	u u	-*	2.0 %	1.9 %
526833	folDeU-5	a	a	c	a	g	a	u	g	g	a	a	u	c	c	u	c	u	C	u c	85.9 %	-	83.5 %
2188437	ccmFeU118RC	g	c	a	u	g	a	u	g	g	c	g	u	c	u	u	c	c	C	g c	46.9 %	-	**64.6 %**
914422	poxBeU1332LL	u	u	g	a	u	g	g	g	c	g	a	u	u	u	c	c	u	C	u c	1.1 %	-	**4.1 %**
1280500	hnseU215PL	c	g	c	u	g	a	c	g	g	u	a	u	u	g	a	c	c	C	g a	1.4 %	-	**2.7 %**

Supplementary figures

Supplementary Figure 1. Makeup of *Physcomitrium patens* RNA editing factors PPR56 and PPR65

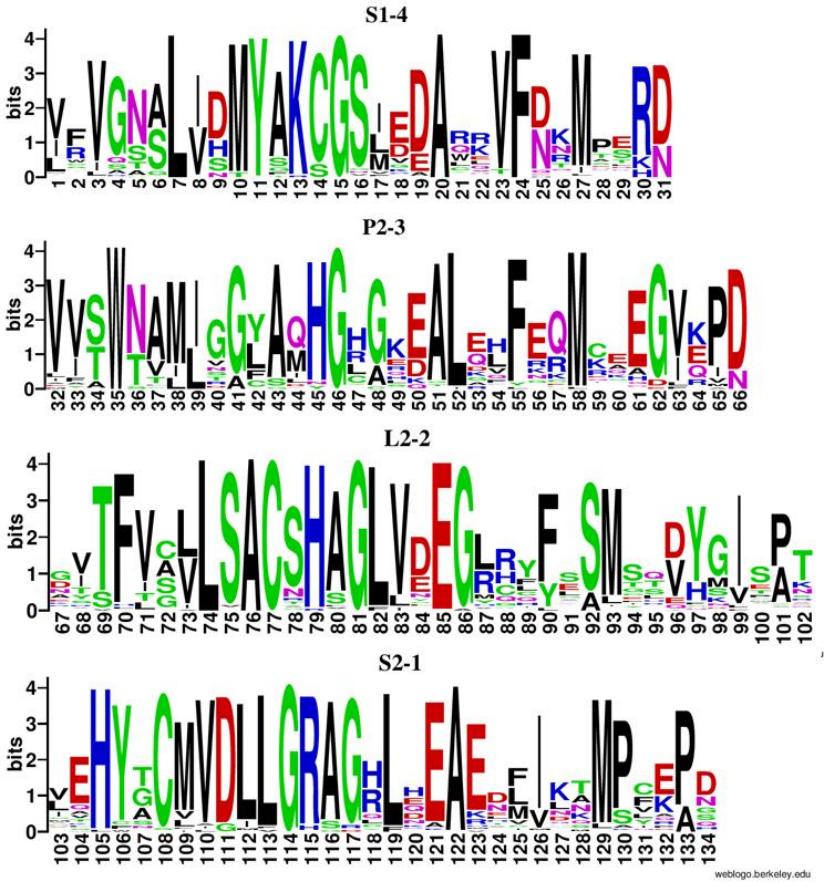
PPR56 and PPR65 are typical RNA editing factors in *Physcomitrium patens* featuring the complete suites of PLS-type PPR arrays linked to a carboxyterminal DYW cytidine deaminase domain via the TPR-like E1 and E2 motifs. Numbering of PPRs is backward as previously suggested [65]. Target recognition follows a PPR-RNA code defined by amino positions 5 and L in P- and S-type PPRS (T/S+N:A, T/S+D:G, N+N:Y, N+D:U). Native targets are indicated below the protein structures with target position -4 juxtaposed with the terminal PPR S2-1. Hybrid target variants nad4eU272SL/ccmFCeU103PS and nad4eU272SL/rps14eU80SL adapting positions -3 to +5 around the cytidine to be edited (C) have been created for testing with the corresponding protein chimeras PPR56/PPR65 (guuCccaca, Fig. 1B) and PPR56/PPR45 (ucaCggcaa, Fig. 2), respectively. Consensus profiles of the four C-terminal PPRs starting with S-4, of the E1 and E2 motifs and of the DYW domain are given in supplementary figure 2.



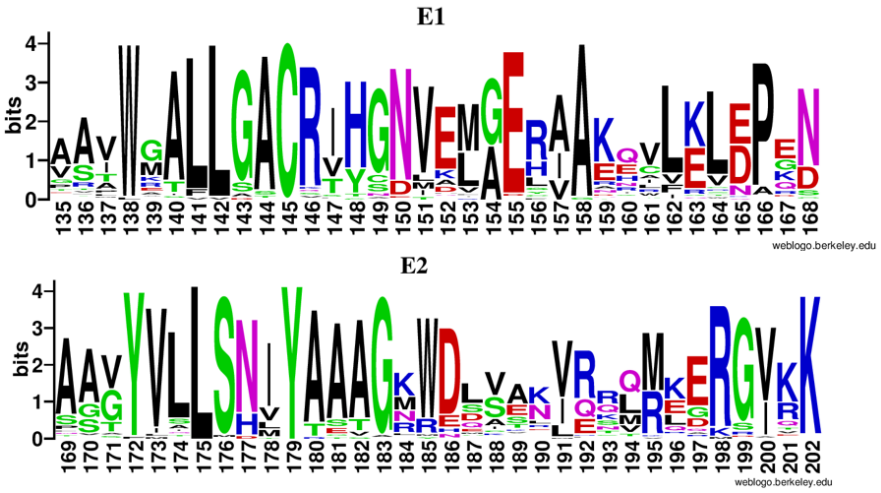
Supplementary Figure 2. WebLogo consensus profiles of C-terminal PPRs S1-4 to S2-1, TPR-like E1 and E2 motifs and DYW domains of moss proteins.

Moss protein homologues of *Physcomitrium patens* RNA editing factor PPR56 were identified by BLASTP [66] at https://blast.ncbi.nlm.nih.gov/Blast.cgi (word size = 3, random expectancy cutoff = 1e-70), ultimately retaining 86 proteins after removal of incomplete DYWs and duplicates (as of Feb 22, 2023). WebLogo consensus profiles were created after alignment at https://weblogo.berkeley.edu/logo.cgi for the four carboxyterminal PPRs S-4, P2-3, L2-2 and S2-1 (A), the E1 and E2 motifs (B) and the DYW domains (C). Numbering is continuous from start of PPR S-4 to the end of E2, but starts anew for the DYW domain. The region of the proposed gating domain (pos. 14-67) and the α1 helix (pos. 28-43) in the DYW domain is indicated by underlining in orange and brown, respectively. Critical residues for co-ordination of a zinc ion in the catalytic center are H68, C96, 99 and E70 (via a water molecule) and H100, H123, C130 and C132 for a second zinc ion of structural importance. Residues 5(S/T), I7 and F16 are located in beta sheets β1 and β2 of the PG box and interact with the α2 helix contributing to the catalytic center.

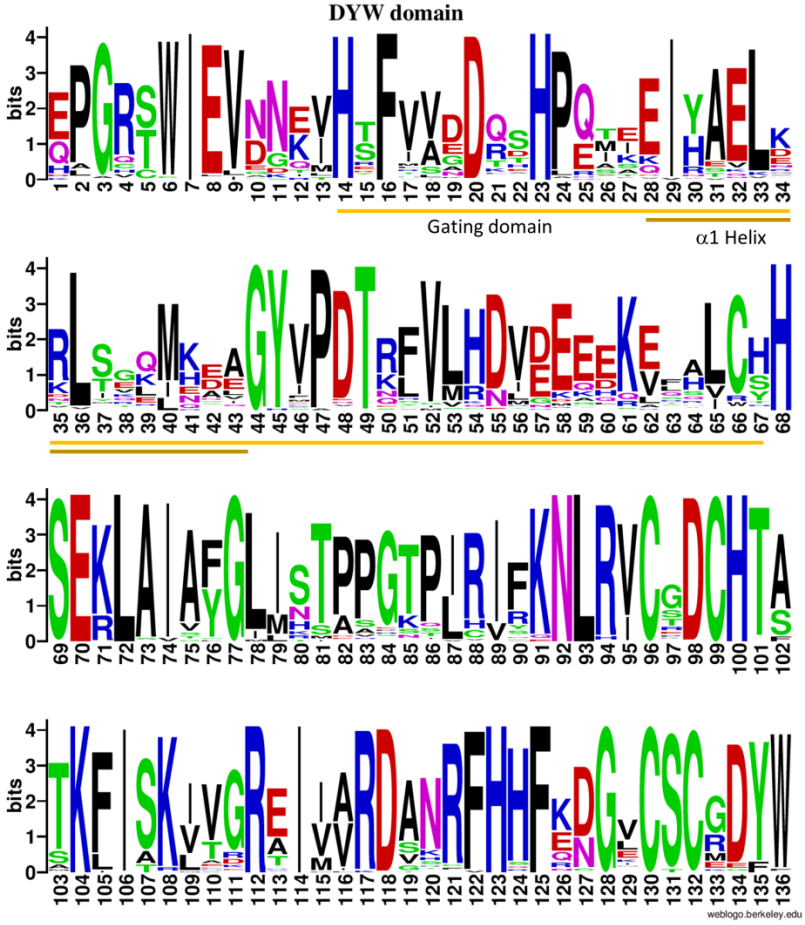
A. Consensus profiles of four terminal PPRs S-4, P2-3, L2-2 and S2-1



B. Consensus profiles of TPR-like motifs E1 and E2

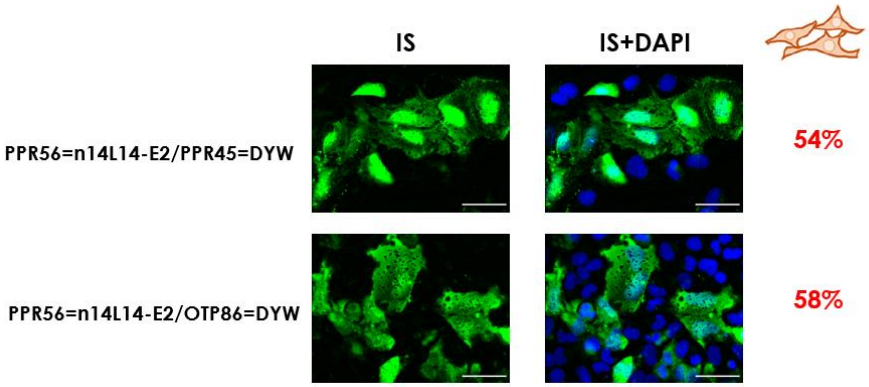


C. Consensus profile of the DYW domain



Supplementary Figure 3. Expression of RNA editing factor chimeras in a human cell line.

Recombinant PPR56/PPR45 and PPR56/OTP86 chimeras tagged with upstream EYFP yielded in fluorescence signals upon expression in human IMR-90 cells and resulted in RNA editing at the co-delivered native target nad4eU272SL of PPR56 as indicated.



Supplementary tables

Supplementary Table 1. Primary results of RNA editing assays in E. coli and human cells.

Results for determination of RNA editing in triplicate assays in Escherichia coli and human IMR cells. The resulting mean and standard deviations are listed. C-to-U RNA editing frequencies are given as 100% when no remaining cytidine signal was detectable upon sequencing of RT-PCR products.

Supplementary Table 2. Summary of RNA-seq data sets for off-target analyses.

RNA-seq datasets analyzed for C-to-U RNA editing off-targets. Separate tabs for the summary off-target lists for 12 individual data sets for Jacusa variant calls (E. coli wild-type background control for reference, native PPR56 without co-delivered targets (2 replicates), with co-delivered nad3eU230SL target, nad4eU272SL target (2 replicates) and combined nad4-nad3 target, PPR56/PPR65 chimera with nad4eU272SL target (2 replicates), and PPR56/OTP86 chimera with nad4eU272SL target (3 replicates)) analyzed in the course of this study.

Supplementary Table 3. Table of identified off-targets.

List of off-targets independently determined for a given protein construct in two independent RNA-seq runs.

Supplementary Table 4. Oligonucleotides.

Oligonucleotides used in this study. All oligonucleotides were synthesized by IDT (Integrated DNA technologies Europe, BVBA, Leuven, Belgium).

References

1. Steinhauser, S., Beckert, S., Capesius, I., Malek, O. & Knoop, V. Plant mitochondrial RNA editing: extreme in hornworts and dividing the liverworts? *J Mol Evol* **48**, 303–312 (1999).

2. Rüdinger, M., Polsakiewicz, M. & Knoop, V. Organellar RNA editing and plant-specific extensions of pentatricopeptide repeat proteins in jungermanniid but not in marchantiid liverworts. *Mol Biol Evol* **25**, 1405–1414 (2008).

3. Freyer, R., Kiefer-Meyer, M.-C. & Kössel, H. Occurrence of plastid RNA editing in all major lineages of land plants. *Proceedings of the National Academy of Sciences of the U.S.A.* **94**, 6285–6290 (1997).

4. Malek, O., Lättig, K., Hiesel, R., Brennicke, A. & Knoop, V. RNA editing in bryophytes and a molecular phylogeny of land plants. *EMBO J* **15**, 1403–1411 (1996).

5. Ichinose, M. & Sugita, M. Substitutional RNA Editing in Plant Organelles. *Methods Mol Biol* **2181**, 1–12 (2021).

6. Knoop, V. C-to-U and U-to-C: RNA editing in plant organelles and beyond. *J Exp Bot* **74**, 2273–2294 (2023).

7. Small, I. D., Schallenberg-Rüdinger, M., Takenaka, M., Mireau, H. & Ostersetzer-Biran, O. Plant organellar RNA editing: what 30 years of research has revealed. *Plant Journal* **101**, 1040–1056 (2020).

8. Ichinose, M., Sugita, C., Yagi, Y., Nakamura, T. & Sugita, M. Two DYW subclass PPR proteins are involved in RNA editing of *ccmFc* and *atp9* transcripts in the moss *Physcomitrella patens*: first complete set of PPR editing factors in plant mitochondria. *Plant Cell Physiol* **54**, 1907–1916 (2013).

9. Schallenberg-Rüdinger, M. & Knoop, V. Coevolution of organelle RNA editing and nuclear specificity factors in early land plants. in *Genomes and Evolution of Charophytes, Bryophytes and Ferns. Advances in Botanical Research, Vol. 78* (ed. Rensing, S. A.) vol. 78 37–93 (Elsevier Academic Press, 2016).

10. Salone, V. *et al.* A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Lett* **581**, 4132–4138 (2007).

11. Oldenkott, B., Yang, Y., Lesch, E., Knoop, V. & Schallenberg-Rüdinger, M. Plant-type pentatricopeptide repeat proteins with a DYW domain drive C-to-U RNA editing in *Escherichia coli*. *Commun Biol* **2**, 85 (2019).

12. Takenaka, M. *et al.* DYW domain structures imply an unusual regulation principle in plant organellar RNA editing catalysis. *Nat Catal* **4**, 510–522 (2021).

13. Lesch, E. *et al.* Plant mitochondrial RNA editing factors can perform targeted C-to-U editing of nuclear transcripts in human cells. *Nucleic Acids Res* **50**, 9966–9983 (2022).

14. Boussardon, C. *et al.* The cytidine deaminase signature HxE(x)nCxxC of DYW1 binds zinc and is necessary for RNA editing of ndhD-1. *New Phytologist* **203**, 1090–1095 (2014).

15. Hayes, M. L. & Santibanez, P. I. A plant pentatricopeptide repeat protein with a DYW-deaminase domain is sufficient for catalyzing C-to-U RNA editing *in vitro*. *J Biol Chem* **295**, 3497–3505 (2020).

16. Iyer, L. M., Zhang, D., Rogozin, I. B. & Aravind, L. Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. *Nucleic Acids Res* **39**, 9473–9497 (2011).

17. Hayes, M. L., Giang, K., Berhane, B. & Mulligan, R. M. Identification of two pentatricopeptide repeat genes required for RNA editing and zinc binding by C-terminal cytidine deaminase-like domains. *J Biol Chem* **288**, 36519–36529 (2013).

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

19. Barkan, A. *et al.* A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet* **8**, e1002910 (2012).

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

21. Yan, J. *et al.* Delineation of pentatricopeptide repeat codes for target RNA prediction. *Nucleic Acids Res* **47**, 3728–3738 (2019).

22. Gutmann, B. *et al.* The Expansion and Diversification of Pentatricopeptide Repeat RNA-Editing Factors in Plants. *Mol Plant* **13**, 215–230 (2020).

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

24. Zehrmann, A., Verbitskiy, D., van der Merwe, J. A., Brennicke, A. & Takenaka, M. A DYW domain-containing pentatricopeptide repeat protein is required for RNA editing at multiple sites in mitochondria of *Arabidopsis thaliana*. *Plant Cell* **21**, 558–567 (2009).

25. Kotera, E., Tasaka, M. & Shikanai, T. A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature* **433**, 326–330 (2005).

26. Boussardon, C. *et al.* Two interacting proteins are necessary for the editing of the *ndhD*-1 site in *Arabidopsis* plastids. *Plant Cell* **24**, 3684–3694 (2012).

27. Takenaka, M. *et al.* Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. *Proceedings of the National Academy of Sciences of the U.S.A.* **109**, 5104–5109 (2012).

28. Bentolila, S., Oh, J., Hanson, M. R. & Bukowski, R. Comprehensive high-resolution analysis of the role of an Arabidopsis gene family in RNA editing. *PLoS Genet* **9**, e1003584 (2013).

29. Haag, S. *et al.* Crystal structures of the *Arabidopsis thaliana* organellar RNA editing factors MORF1 and MORF9. *Nucleic Acids Res* **45**, 4915–4928 (2017).

30. Huang, C. *et al.* MORF2 tightly associates with MORF9 to regulate chloroplast RNA editing in Arabidopsis. *Plant Science* **278**, 64–69 (2019).

31. Zehrmann, A. *et al.* Selective homo- and heteromer interactions between the Multiple Organellar RNA Editing Factor (MORF) proteins in *Arabidopsis thaliana*. *Journal of Biological Chemistry* **290**, 6445–6456 (2015).

32. Bayer-Császár, E. *et al.* The conserved domain in MORF proteins has distinct affinities to the PPR and E elements in PPR RNA editing factors. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1860**, 813–828 (2017).

33. Yan, J. *et al.* MORF9 increases the RNA-binding activity of PLS-type pentatricopeptide repeat protein in plastid RNA editing. *Nat Plants* **3**, 17037 (2017).

34. Royan, S. *et al.* A synthetic RNA editing factor edits its target site in chloroplasts and bacteria. *Commun Biol* **4**, 1–10 (2021).

35. Sun, T., Bentolila, S. & Hanson, M. R. The unexpected diversity of plant organelle RNA editosomes. *Trends Plant Sci* **21**, 926–973 (2016).

36. Takenaka, M. How Complex Are the Editosomes in Plant Organelles? *Mol Plant* **7**, 582–585 (2014).

37. Yang, Y.-Z. Z. *et al.* GRP23 plays a core role in E-type editosomes via interacting with MORFs and atypical PPR-DYWs in Arabidopsis mitochondria. *Proceedings of the National Academy of Sciences of the U.S.A.* **119**, 2210978119 (2022).

38. Gutmann, B., Royan, S. & Small, I. Protein Complexes Implicated in RNA Editing in Plant Organelles. *Mol Plant* **10**, 1255–1257 (2017).

39. Sandoval, R. *et al.* Stable native RIP9 complexes associate with C-to-U RNA editing activity, PPRs, RIPs, OZ1, ORRM1, and ISE2. *The Plant Journal* **99**, 1116–1126 (2019).

40. Guillaumot, D. *et al.* Two interacting PPR proteins are major Arabidopsis editing factors in plastid and mitochondria. *Proceedings of the National Academy of Sciences of the U.S.A.* **114**, 8877–8882 (2017).

41. Andrés-Colás, N. *et al.* Multiple PPR protein interactions are involved in the RNA editing system in Arabidopsis mitochondria and plastids. *Proceedings of the National Academy of Sciences of the U.S.A.* **114**, 8883–8888 (2017).

42. Ichinose, M. & Sugita, M. The DYW domains of pentatricopeptide repeat RNA editing factors contribute to discriminate target and non-target editing sites. *Plant Cell Physiol* **59**, 1652–1659 (2018).

43. Maeda, A. *et al.* DYW deaminase domain has a distinct preference for neighboring nucleotides of the target RNA editing sites. *The Plant Journal* **111**, 756–767 (2022).

44. Ohtani, S. *et al.* Targeted gene disruption identifies three PPR-DYW proteins involved in RNA editing for five editing sites of the moss mitochondrial transcripts. *Plant Cell Physiol* **51**, 1942–1949 (2010).

45. Schallenberg-Rüdinger, M., Kindgren, P., Zehrmann, A., Small, I. & Knoop, V. A DYW-protein knockout in *Physcomitrella* affects two closely spaced mitochondrial editing sites and causes a severe developmental phenotype. *The Plant Journal* **76**, 420–432 (2013).

46. Yang, Y., Ritzenhofen, K., Otrzonsek, J., Schallenberg-Rüdinger, M. & Knoop, V. Beyond a PPR-RNA recognition code: Many aspects matter for the multi-targeting properties of RNA editing factor PPR56. *bioRxiv* 2023.04.05.535663 (2023) doi:10.1101/2023.04.05.535663.

47. Ichinose, M., Uchida, M. & Sugita, M. Identification of a pentatricopeptide repeat RNA editing factor in *Physcomitrella patens* chloroplasts. *FEBS Lett* **588**, 4060–4064 (2014).

48. Miyata, Y. & Sugita, M. Tissue- and stage-specific RNA editing of *rps14* transcripts in moss (*Physcomitrella patens*) chloroplasts. *J Plant Physiol* **161**, 113–115 (2004).

49. Lurin, C. *et al.* Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* **16**, 2089–2103 (2004).

50. Hammani, K. *et al.* A study of new *Arabidopsis* chloroplast RNA editing mutants reveals general features of editing factors and their target sites. *Plant Cell* **21**, 3686–3699 (2009).

51. Rüdinger, M., Fritz-Laylin, L., Polsakiewicz, M. & Knoop, V. Plant-type mitochondrial RNA editing in the protist Naegleria gruberi. *RNA* **17**, (2011).

52. Knoop, V. & Rüdinger, M. DYW-type PPR proteins in a heterolobosean protist: plant RNA editing factors involved in an ancient horizontal gene transfer? *FEBS Lett* **584**, 4287–4291 (2010).

53. Schallenberg-Rüdinger, M., Lenz, H., Polsakiewicz, M., Gott, J. M. & Knoop, V. A survey of PPR proteins identifies DYW domains like those of land plant RNA editing factors in diverse eukaryotes. *RNA Biol* **10**, 1549–1556 (2013).

54. Schaap, P. *et al.* The *Physarum polycephalum* Genome Reveals Extensive Use of Prokaryotic Two-Component and Metazoan-Type Tyrosine Kinase Signaling. *Genome Biol Evol* **8**, 109–125 (2016).

55. Schallenberg-Rüdinger, M. *et al.* A single-target mitochondrial RNA editing factor of *Funaria hygrometrica* can fully reconstitute RNA editing at two sites in *Physcomitrella patens*. *Plant Cell Physiol* **58**, 496–507 (2017).

56. Okuda, K. *et al.* Quantitative analysis of motifs contributing to the interaction between PLS-subfamily members and their target RNA sequences in plastid RNA editing. *The Plant Journal* **80**, 870–882 (2014).

57. Chateigner-Boutin, A.-L. *et al.* The E domains of pentatricopeptide repeat proteins from different organelles are not functionally equivalent for RNA editing. *The Plant Journal* **74**, 935–945 (2013).

58. Ruwe, H. *et al.* The E domain of CRR2 participates in sequence-specific recognition of RNA in plastids. *New Phytologist* **222**, 218–229 (2019).

693 59. Toma-Fukai, S. *et al.* Structural insight into the activation of an Arabidopsis organellar C-to-U
694 RNA editing enzyme by active site complementation. *Plant Cell* doi:10.1093/plcell/koac318
695 (2022) doi:10.1093/plcell/koac318.

696 60. Kapust, R. B. & Waugh, D. S. *Escherichia coli* maltose-binding protein is uncommonly effective
697 at promoting the solubility of polypeptides to which it is fused. *Protein Science* **8**, 1668–1674
698 (1999).

699 61. Jeong, H. *et al.* Genome Sequences of Escherichia coli B strains REL606 and BL21(DE3). *J Mol*
700 *Biol* **394**, 644–652 (2009).

701 62. Picardi, E. & Pesole, G. REDItools: high-throughput RNA editing detection made easy.
702 *Bioinformatics* **29**, 1813–1814 (2013).

703 63. Piechotta, M., Wyler, E., Ohler, U., Landthaler, M. & Dieterich, C. JACUSA: site-specific
704 identification of RNA editing events from replicate sequencing data. *BMC Bioinformatics* **18**, 7
705 (2017).

706 64. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator.
707 *Genome Res* **14**, 1188–1190 (2004).

708 65. Hein, A. & Knoop, V. Expected and unexpected evolution of plant RNA editing factors CLB19,
709 CRR28 and RARE1: retention of CLB19 despite a phylogenetically deep loss of its two known
710 editing targets in Poaceae. *BMC Evol Biol* **18**, 85 (2018).

711 66. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database
712 search programs. *Nucleic Acids Res* **25**, 3389–3402 (1997).

713

3.4. Investigation of PPR proteins as molecular tool

3.4.1. Testing the special feature of *P. patens* PPR65 L-5 motif

The DYW type PPR protein PPR65 from *P. patens* has a “TD” combination in its fifth and last amino acid of L-5 motif, while the ribonucleotide on the *ccmFC* target corresponding to this position stores a guanine. This coincidence fit for the PPR-RNA binding code designed for P- and S- type repeats. Although whether and how the L motif contributing to PPR-RNA recognition is still under debate, the single nucleotide modifications on the target ccmFCeU103PS|g-8a and g-8c abolished RNA editing in *E. coli* system (Oldenkott et al., 2019). Switching TD combination to ND (PPR65|L-5 TD>ND), which would change the ribonucleotide favor from guanine to uridine, together with the target modification (ccmFCeU103PS|g-8u) to complement the protein modification did not restore RNA editing in *E. coli* system (Oldenkott, 2020). This makes it puzzling if the “TD” combination of this L motif pointing to the RNA recognition roll, particularly in this L motif. According to Gutmann et al. (2020), the L1-type PPR repeats in mosses showed higher frequency of “LD” than “TD”, a protein modification on PPR65|L-5 TD>LD was preformed via overlap extension PCR and transferred into the *E. coli* system based on Oldenkott et al. (2019). Both single *ccmFC* target and tandem *ccmFC* targets were tested as described in Yang et al. (2023a), only the upstream *ccmFC* target in tandem targets construct showed 31% editing (Figure 9). This leading to L-5 not linking to target recognition follows the PPR-RNA recognition code, the fitness of the binding code would just be a coincidence.

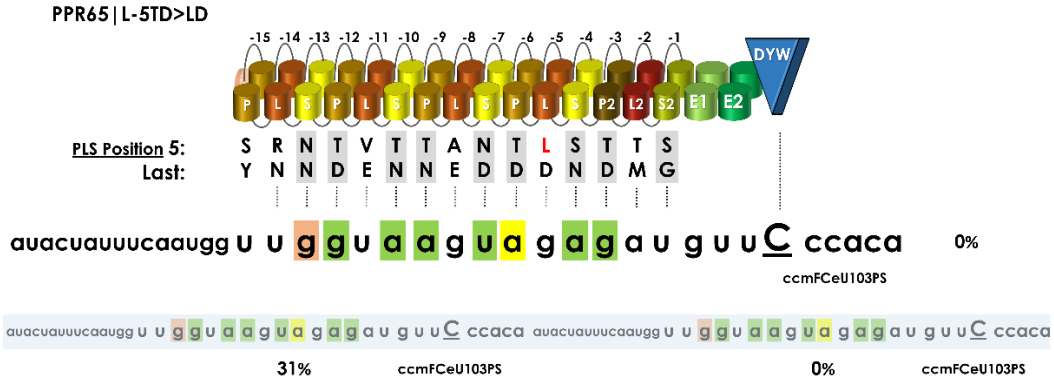


Figure 9. PPR65|L-5TD>LD only mediate upstream site of tandem *ccmFC* targets editing in *E. coli*. The modified amino acid is indicated in red. No editing was observed in the single cloned ccmFCeU103PS, but in tandem *ccmFC* target (shade in blue), 31% editing was observed in the upstream site.

In the next step, complete L-8 AE motif was "copied" in position L-5 (PPR65:n15-P6/L8/S4-DYW), to examine if this L-5 could be replaced by other L motif. Both single and tandem *ccmFC* targets could not be edited by this PPR65 containing artificial repetitive L motif (Figure 10A). Thinking about the PLS motifs showed in triplet in most cases, the P-6, L-5, S-4 was switched with P-9, L-8, S-7 (PPR65:n15-S10/P6-S4/P9-S7/P3-DYW). Accordingly, the ribonucleotide on the corresponding positions were switched (ccmFCeU103PS|a-7u|u-10a) to complement the protein modification. Again, no editing could be gained in both single and tandem *ccmFC* targets (Figure 10B).

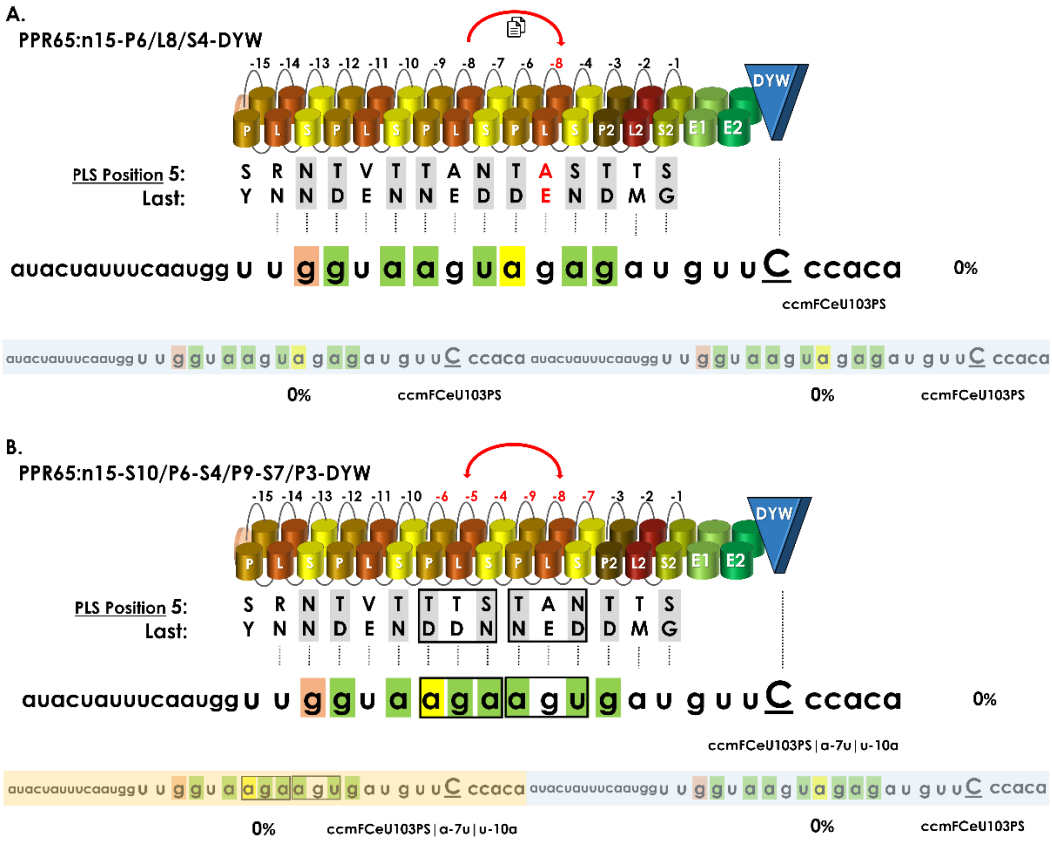


Figure 10. Motif-switch experiment on PPR65 and its target ccmFCeU103PS in *E. coli*. A. The complete L-5 TD motif of PPR65 was replaced by the complete L-8 AD of PPR65 (indicated in red), resulting in two L-8 AD motifs in PPR65. As a result, both single and tandem *ccmFC* targets (shaded in blue) could not be edited by this mutated PPR65. **B.** The complete PLS triple P-6, L-6, and S-4 was replaced with the PLS triple P-9, L-8, and S-7 (indicated in red). As the P- and S- repeats were changed, the corresponding nucleotides were switched as well (indicated in black square). No editing was observed in the mutated single *ccmFC* target. When the mutated ccmFC target (shaded in orange) was placed upstream of the native *ccmFC* target (shaded in blue), no editing was observed in either of them.

3.4.2. Re-positioning PPR motif to examine the possibility of artificial PPR protein

Although synthetic DYW-type PPR proteins with repetitive P motif or S motifs or PLS motifs could edit supplied target successfully (Royan et al., 2021), it is still puzzling why native DYW-type PPR proteins require different PLS motifs combination for target recognition. It is always wondered if the PPR repeats from different PPR

proteins with the same amino acid combination on fifth and last position, although the rest amino acids differs, could be replaced. Among the nine RNA editing factors from *P. patens*, PPR56 and PPR65 stores similar numbers of PLS repeats (PPR56 has 14 and PPR65 has 15). In addition, native PPR56 P-9 motif holds "TN" combination, while native PPR65 P-9 motif, the same position, holds the same. "Coping" the complete P-9 TN motif from PPR56 to replace PPR65 P-9 TN (PPR65:n15-S10/PPR56:P9/PPR65:L8-DYW) leading to 18% of editing on single *ccmFC* target. In the tandem *ccmFC* target construct, 42% editing was obtained in the upstream one, while 24% of cytidine was edited to uridine in the downstream target (Figure 11). Given that PPR65 and its *ccmFC* target is more sensitive to both protein and target mutations, this could be influenced by protein flexibility.

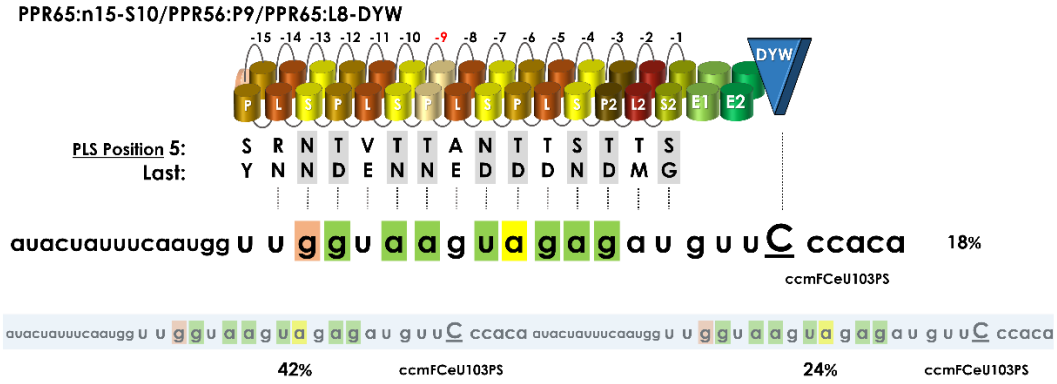


Figure 11. PPR65 with P-9 switched to PPR56 P-9 could edited ccmFCeU103PS in *E. coli*. Since both P-9 of PPR65 and PPR56 has a "TN" combination for target recognition, this motif was switched in PPR65 (labeled in light brown and numbering in red). 18% editing on single cloned ccmFCeU103PS was obtained, while up to 42% editing was seen in tandem *ccmFC* targets.

3.5. Yang et al 2023c (under preparation/revision for submission)

Yingying Yang, Bastian Oldenkott, Elena Lesch, Volker Knoop and Mareike Schallenberg-Rüdinger (2023)

The C-terminal DYW domain of a PPR protein from the protist *Naegleria gruberi* can act as cytidine deaminase in moss mitochondria

Although plant-type RNA editing is mainly restricted to land plants, DYW-type PPR proteins have been found outside of plants as well. *Naegleria gruberi* is one example, which holds 10 DYW-type PPR proteins and 2 editing sites (cox1eU1120HY and cox3eU787RW) in the mitochondrial transcriptome. According to bioinformatic analysis, NgPPR51 and NgPPR45 are the best candidates to edit the two sites based on the PPR-RNA binding code. NgPPR51 has a mitochondrial signal peptide with only one perfect match to each of the editing sites. Ng45423, a pure PLS type PPR protein without C-terminal extensions, also has a mitochondrial signal peptide. The gene Naegr45423 is located only 154 nucleotides away from Naegr45424, and there is also a 324 bp "intron" within Naegr45423. Since Naegr45423 and Naegr45424 show several repetitive regions, PCR was performed on *N. gruberi* DNA to recheck the sequences. However, the "intron" could not be confirmed. The amplicon of Naegr45423 and Naegr45424 is shorter than the annotation. Furthermore, amplification on cDNA further confirmed that Naegr45423 and Naegr45424 are likely one gene that was wrongly annotated before. The rearranged Naegr45423 and Naegr45424 translates into NgPPR45, with a mitochondrial signal peptide, 19 PPR repeats which have 6 perfect matches to the *cox1* site and 5 perfect matches to the *cox3* site, and a complete C-terminal extension.

The analysis of the C-terminal extensions of NgPPR45 and NgPPR51 shows a conserved cytidine deaminase signature with some amino acid differences compared to editing factors in *P. patens*. The most prominent difference is the lack of conservation of the “SW” within the “PG-box”, which has been shown to be important for proper editing function in angiosperms. The terminated DYW tripeptide, which gives the name to the DYW domain, is also modified in both PPR

proteins. While NgPPR51 has “HSE” as all other editing factors do, NgPPR45 has “HAE”, which is usually seen in the reverse editing factors.

Since knockout studies are not yet feasible in *N. gruberi*, different versions of chimeras with the C-terminal extensions of NgPPR45 and NgPPR51 have been fused with PpPPR78 and complemented in the KO PPR78 *P. patens* plant. Only the PpPPR78 fused with the E1E2 and DYW domains of NgPPR45 could edit the cox1eU755SL site (editing site of PPR78) up to 82%. Real-time PCR showed that the editing efficiency of the *cox1* site was related to the level of protein overexpression.

Part of this work has been included in the PhD thesis of Bastian Oldenkott. I mainly designed the PCR for rearrangement of NgPPR45, finished the genotyping and preformed the real-time PCR analysis.

The C-terminal DYW domain of a PPR protein from the protist *Naegleria gruberi* can act as cytidine deaminase in moss mitochondria

Yingying Yang, Bastian Oldenkott, Elena Lesch, Volker Knoop and Mareike Schallenberg-Rüdinger*

IZMB – Institut für Zelluläre und Molekulare Botanik, Abteilung Molekulare Evolution, Universität Bonn, Kirschallee 1, D-53115 Bonn, Germany

*Corresponding author
mareike.ruedinger@uni-bonn.de, Phone: +49 228 73-6464

ORCID-IDs
Yingying Yang: 0000-0002-5518-0307
Bastian Oldenkott: 0000-0002-2513-2443
Elena Lesch: 0000-0002-3449-3929
Volker Knoop: 0000-0002-8485-9423
Mareike Schallenberg-Rüdinger: 0000-0002-6874-4722

Key words*: Naegleria gruberi*, *Physcomitrium patens,* mitochondrial RNA editing factor, PLS-type pentatricopeptide repeat (PPR) protein, DYW-type cytidine deaminase, PPR-RNA recognition code, horizontal gene transfer

Abstract

Organellar cytidine-to-uridine RNA editing driven by RNA-binding DYW-type pentatricopeptide repeat proteins is largely restricted to land plants. One of the rare exceptions is the heterolobosean protist *Naegleria gruberi*, which encodes ten DYW-type PPR proteins in its nuclear genome and features two C-to-U RNA editing sites in its mitochondrial transcriptome. Bioinformatic analyses favored NgPPR45 and NgPPR51 as top candidates for mitochondrial import and targeting those two RNA editing sites. Here, we tested the cytidine deaminase functionality of their C-terminal DYW domains making use of the model moss *Physcomitrium patens*. We designed protein chimeras combining the RNA-binding region of moss editing factor PpPPR78 and different portions of the putative cytidine deaminase domain of the two *N. gruberi* DYW-type PPR proteins, respectively. Chimeras were introduced into a KO plant line of PpPPR78, aiming to restore RNA editing at its assigned sites. Indeed, cox1eU755SL was edited in up to 82 % of transcripts by the chimera of PpPPR78 and the E1-E2-DYW domain of *N. gruberi* NgPPR45. In contrast, neither chimeras with smaller portions of the catalytic C-terminal domain of NgPPR45 nor chimeras using NgPPR51 domains were functional *P. patens*. We conclude that the C-terminal DYW domain of NgPPR45 is a functional cytidine deaminase and that the match of its PPR array to the putative binding regions upstream of the two mitochondrial editing targets of *Naegleria gruberi* make this factor the prime candidate to edit the latter in the protist.

Introduction

Genetic information can be changed on transcript level by RNA editing. Depending on the species, nuclear or organellar transcripts are edited by insertion, deletion or conversion of nucleotides (reviewed in Knoop, 2011; Knoop, 2022). In Metazoa, A-to-I RNA editing of nuclear transcripts is the dominant type of RNA editing (Nishikura, 2010) and only few cases of C-to-U of RNA editing are known (Meier et al., 2005). In plants, mitochondrial and chloroplast transcripts are affected by pyrimidine RNA editing (C-U or U-C) which typically leads to the correction of conserved codon identities or reading frames (reviewed in Knoop, 2022; Small et al., 2020). No single case of C-to-U RNA editing is known in algae by now, leading to the assumption, that the mechanism evolved in the common ancestor of land plants as consequence of colonizing land (Fujii and Small, 2011). In all land plants, except the marchantiid liverworts (Rüdinger et al., 2012), a few sites (moss *Funaria hygrometrica*, with two sites in the chloroplast and eight sites in the mitochondria, (Rüdinger et al., 2011b)) up to several thousands (*Selaginella* with 2700 sites in mitochondria (Hecht et al., 2011), and 3400 sites in the chloroplast (Oldenkott et al., 2014)) need to be edited to ensure translation of the correct protein sequence. In liverworts, mosses and seed plants only cytidines are converted into uridines in organellar transcripts (Dong et al., 2019; Rüdinger et al., 2012). In hornworts, some lycophytes and ferns classic C-to-U RNA editing is also accompanied by the reverse type of editing changing uridines to cytidines (Duff, 2006; Grewe et al., 2011; Gutmann et al., 2020; Knie et al., 2016).

C-to-U RNA editing in land plants is mediated by members of the large family of RNA binding pentatricopeptide repeat (PPR) proteins (Cheng et al., 2016; Gutmann et al., 2020; Ichinose et al., 2022). These alpha solenoid proteins consist of a PPR stretch with two to 30 circa 35 amino acid long PPR repeats (Knoop and Rüdinger, 2010; Lurin et al., 2004). Plant PPR proteins involved in RNA editing belong to the "PLS" subfamily. These proteins were only found to be encoded in species with RNA editing sites identified in their organellar transcriptomes (Gutmann et al., 2020; Rüdinger et al., 2008; Salone et al., 2007). Their

PPR arrays differ from those of “P-type” PPR proteins by exhibiting long (L, 35–36 aa) and short (S, 31–32 aa) PPR variants alternating with the canonical P-type PPRs of 35 amino acids (Cheng et al., 2016; Lurin et al., 2004). P- and S-type PPRs recognize and bind specific nucleotides in an one-repeat-one-nucleotide manner following a PPR-RNA binding code, which was proposed in 2012 and refined in several follow-up publications (Figure 1, Barkan et al., 2012; Takenaka et al., 2013; Yan et al., 2019). The 5th and the last (L) amino acid of each P- and S-type repeat form hydrogen bonds with the assigned nucleotides which places the PPR protein to edit the cytidine four nucleotides downstream of the nucleotide corresponding to the last PPR repeat (Figure 1, Barkan et al., 2012). The PPR array is followed by TPR-like E1 and E2 helix-turn-helix motifs and a DYW domain, named after its conserved C-terminal aspartate, tyrosine and tryptophan tripeptide (Cheng et al., 2016; Lurin et al., 2004).

The DYW domain features the typical $HxE(x)_nCxxC$ deaminase signature (Iyer et al., 2011; Salone et al., 2007) and was recently unequivocally assigned to exert the cytidine deaminase function (Hayes and Santibanez, 2020; Oldenkott et al., 2019; Takenaka et al., 2021). Single moss DYW-type PPR proteins were shown to edit provided targets, when transferred into the bacterium *Escherichia coli* (Oldenkott et al., 2019), *in vitro* (Hayes and Santibanez, 2020) or even in the cytoplasm of human cells (Ichinose et al., 2022; Lesch et al., 2022; Takenaka et al., 2021). Single amino acid changes in the cytidine deaminase signature of the DYW domain of such editing factors led to complete loss of editing, also *in planta* (Boussardon et al., 2014; Ichinose and Sugita, 2018; Wagoner et al., 2015).

Flowering plants encode some complete, but also many C-terminally truncated DYW-type PPR proteins lacking parts of the DYW domain (E+ subgroup), the complete DYW domain (E2 subgroup) or the complete E2 and DYW domain (E1 subgroup; Cheng et al., 2016). They need to recruit a DYW-domain *in trans* to obtain editing functionality (Andrés-Colás et al., 2017; Boussardon et al., 2012; Diaz et al., 2017; Gerke et al., 2020; Guillaumot et al., 2017; Wang et al., 2022). In angiosperms, additional factors like Multiple organellar RNA editing factors (MORFs), also called RNA-editing factor Interacting Protein (RIPs), organelle

RNA recognition motif-containing (ORRM) proteins, organellar zinc finger (OZ) proteins or P-type PPR proteins were also identified to participate in RNA editing of particular editing sites, building a complex editosome (Andrés-Colás et al., 2017; Bentolila et al., 2012; Gipson et al., 2022; Guillaumot et al., 2017; Sun et al., 2016; Sun et al., 2015; Takenaka et al., 2012, reviewed in Small et al., 2023).

The model moss *Physcomitrium patens* encodes only ten complete DYW-type PPR proteins in its nuclear genome (O'Toole et al., 2008), which correlates well with only eleven editing sites identified in its mitochondrial transcriptome and two sites in its plastid transcriptome (Miyata and Sugita, 2004; Rüdinger et al., 2009). Nine of the ten DYW-type PPR proteins were assigned to one up to two editing sites in the chloroplast or mitochondrion, respectively (Ichinose et al., 2014; Schallenberg-Rüdinger et al., 2013a), while the tenth DYW type PPR protein was shown to act in group II intron splicing of *cox1* instead (Ichinose et al., 2012). No RNA editing helper proteins were identified in *P. patens* so far (Uchiyama et al., 2018).

With more genomic and transcriptomic sequence data becoming available, rare cases of DYW domains encoded in genomes of species outside of land plants were discovered (Gutmann et al., 2020; Schallenberg-Rüdinger et al., 2013b). Besides identification of single DYW domain sequences in genomes of fungi (Schallenberg-Rüdinger et al., 2013b), dinoflagellates (Mungpakdee et al., 2014) or algae (Gutmann et al., 2020; Schallenberg-Rüdinger et al., 2013b), DYW type protein families of 10 up to 20 proteins were found to be encoded in the genomes of the slime mold *Physarum* (20, Schaap et al., 2015), the heterolobosean protists *Acrasis* (12, Fu et al., 2014) and *Naegleria gruberi* (10, Knoop and Rüdinger, 2010). In parallel, C-to-U RNA editing sites were found in the mitochondrial transcripts of these species. In *Physarum* four C-to-U RNA editing sites in the mitochondrial transcriptome are accompanied by many other RNA editing events, including insertion of single Cs, Us and dinucleotides as well as deletions (Bundschuh et al., 2011; Gott et al., 2005). In the protists *Acrasis* and *Naegleria* two mitochondrial C-to-U RNA editing sites, cox1eU1120HY and cox3eU787RW in *Naegleria* and atp6eU722SL and cobeU409HY in *Acrasis*, were identified. Editing sites are labeled as proposed in Rüdinger et al. (2009) starting with the

gene name of the affected transcript, eU for editing from C to U, the position of the C to be edited in the coding sequence and the resulting change of the codon identity.

As knockout studies with these heterolobosean protists are not feasible to date (Faktorová et al., 2020), we developed an alternative approach to test the functionality of the deaminase domains of the DYW-type PPR proteins of *Naegleria gruberi*. We used the model moss *Physcomitrium patens* KO line of PpPPR78 (Rüdinger et al., 2011b), which lacks editing of the two mitochondrial editing sites cox1eU755SL and rps14eU137SL. Chimeric protein constructs of PpPPR78 with parts of its C-terminal domain replaced by the counterparts of one of the two selected *N. gruberi* DYW-type PPR proteins NgPPR45 and NgPPR51 were introduced in the KO line to test their ability to restore RNA editing.

Results

Selection of DYW-type PPR proteins from *Naegleria gruberi* for complementation studies

The complete genome sequence of the heterolobosean protist *Naegleria gruberi* was published 2010 (NCBI GenBank entry: NW_003163326.1) and subsequently ten DYW-type PPR proteins and one PLS protein lacking the C-terminal extensions were found to be encoded in the nuclear genome (Fritz-Laylin et al., 2010; Knoop and Rüdinger, 2010).

The predicted encoded proteins vary in the number of PPR repeats and only one DYW-type PPR protein Ng51788 (NgPPR51) and the pure PLS protein (Ng45423) have signal peptides predicted for mitochondrial localization encoded N-terminally of their PPR arrays (Rüdinger et al., 2011a). For nine of the ten DYW-type PPR proteins (except Ng32401), PPR protein models with predicted PPR repeats and respective amino acids five (5) and last (L) for target recognition were already available on the PPR plantenergy webpage ((https://ppr.plantenergy.uwa.edu.au/ppr/). With the tool “Search for PPRs” available on the same webpage PPR repeats can be searched for in input sequences with slightly different parameters. Predicted models for most of the ten DYW type PPR proteins showed differences in comparison to the models available on the webpage, especially in terms of biased PPR (Cheng et al., 2016). To give one example,

PPR repeat P-6 of Ng45424, Ng76708 and Ng69406 showed insertions in the region between the two alpha helixes of the PPR repeat and were only predicted as a PPR repeat using the "Search for PPR" option. We manually inspected the different predictions and decided for the more reliable model with the highest number of PPR repeats predicted but the avoidance of overlapping PPR repeats for each DYW-type PPR protein to be presented in Figure 1 (all predictions are found in Supplementary table 1 and protein sequences are presented in supplementary data 2).

We aligned the PPR repeat region of each DYW-type PPR protein of N. gruberi to the sequence upstream of the editing targets cox1eU1120HY and cox3eU787RW, respectively, to identify the DYW-type PPR proteins, which are most likely to bind to the target RNAs. Interestingly, not the protein NgPPR51, which is predicted to be localized to the mitochondria, but the proteins Ng45424 and Ng32041 were identified to match the respective target sequence of cox1eU1120HY and cox3eU787RW best (Figure 1), respectively. Ng45424 showed six perfectly matching PPR-nucleotide pairs with the cis elements of cox1eU1120HY and five with target cox3eU787RW, respectively. Ng32041 showed seven matches with target cox3eU787RW, but also three mismatches in the C-terminal region of the PPR stretch, which was shown to be more relevant for recognizing the target than the N-terminal region in earlier studies (Oldenkott et al., 2020; Oldenkott et al., 2019; Takenaka et al., 2013). In our predicted gene model, however, the coding sequences of Ng45424 and Ng32041 lack an N-terminal organellar signal peptide.

One single DYW-type PPR protein sequence is buildup of predicted genes Naegr45423 and Naegr45424

To investigate, whether our current gene models are incomplete, we examined the upstream regions of both predicted genes Naegr45424 and Naegr32041 in the available scaffolds of the *N. gruberi* genome.

We indeed found a sequence in the 5' region of the predicted Naegr32041 gene that could be translated into 11 additional PPR repeats, linked to the predicted Ng32041 protein via a 29 aa long linker (Supplementary table 1). However, no clear signal peptide for mitochondrial localization was predicted with the WoLFPSORT tool (Supplementary table 4).

155 The predicted gene Naegr45424 was found to be located only 154 nucleotides downstream of Naegr45423.
156 The latter is predicted to encode for the only pure PLS-type protein and includes an intron of 324 bp. As
157 both predicted genes show numerous repetitive regions (Supplementary figure 1), we re-checked the gene
158 makeup via PCR using primers with unique binding sites outside of the repetitive elements. Sequencing
159 revealed that the coding regions are not separated by an intergenic region and the intron sequence in
160 gene Naegr45423 could not be confirmed (Supplementary data 1, Figure 2). However, the amplicon
161 confirming the connection of Naegr45423 and Naegr45424 was shorter than the calculated size of the
162 combination of the two CDS. The size reduction can be explained by the highly repetitive character of the
163 region between Naegr45423 and Naegr45424, which might have resulted in scaffolding/assembly
164 mistakes in the initial genome annotation.

165 An RT-PCR was performed to prove that Naegr45423 and Naegr45424 are linked by the repetitive region
166 and transcribed jointly (Figure 2). The complete gene will be named NgPPR45 in the following. The N-
167 terminus of NgPPR45 contains a clear signal peptide for mitochondrial localization (predicted with
168 WoLFPSORT and TargetP 2.0, Supplementary table 4). The combined PPR array of NgPPR45 encompasses
169 19 PPR repeats and matches the target sequences of both editing sites in the mitochondrial transcripts of
170 *N. gruberi* equally well as the PPR protein encoded by the original Naegr45424 gene model (Figure 2).

171 ## Differences in the conservation of the C-terminal domains of NgPPR45 and NgPPR51 of *Naegleria gruberi*
172 ## and the nine DYW type PPR editing factors of *Physcomitrium patens*

173 NgPPR45 and NgPPR51 both contain E1, E2 and DYW domains with a conserved cytidine deaminase
174 signature which, however, show particular amino acid differences in comparison to the protein sequences
175 of the PPR editing factors of the moss *Physcomitrium patens* (Figure 3). The most prominent difference is
176 the lack of conservation of the „SW“ within the PgxSWiEv motif in the so-called PG-Box, which was shown
177 to be important for proper RNA editing function in flowering plants (Okuda et al., 2007; Takenaka et al.,

178 2021) and just recently identified to participate in forming the cytidine binding pocket (Toma-Fukai et al.,
179 2022).

180 The C-terminal DYW tripeptide is also modified in both PPR proteins. The change of the Y to an N in
181 NgPPR51 might not have a strong impact, as several DYW proteins, for example PPR65, PPR98 and PPR91
182 of *P. patens* ending with DFW (Figure 3) and OTP86 of *Arabidopsis thaliana* with DSW (Takenaka et al.
183 2021), also show different amino acids in the second position of the tripeptide. The G instead of the D in
184 NgPPR45 might have an influence on editing capacity, as most of the characterized DYW-type PPR proteins
185 have a D at that position and an earlier mutational study with DYW1, a short DYW type PPR protein of
186 *Arabidopsis thaliana* interacting with E+ protein CRR4 on chloroplast editing site ndhDeU2TM (Boussardon
187 et al., 2012), already showed that a change of D into A reduced RNA editing capacity of that editing factor
188 (Boussardon et al., 2014).

189 Other differences from the plant consensus E1E2DYW are either found in the C-terminal domains of
190 NgPPR45 or of NgPPR51, as both proteins only share 46% sequence identity within their E1E2DYW domain.
191 Whereas NgPPR51 has the HSE motif within the zinc binding region HxE(x)$_n$CxxC, which is highly conserved
192 in most of the 400 plant C-U DYW-type PPR editing factors characterized so far (PREPACT Edifacts (Lenz et
193 al., 2018)), NgPPR45 exhibits an HAE motif, which can predominately be found in the only recently
194 identified putative reverse editing factors of hornworts and ferns (Gerke et al., 2020; Gutmann et al., 2020;
195 Ichinose et al., 2022), but also in other bacterial, fungal or mammalian cytidine deaminases for example
196 ADAR 1 and 2 (Iyer et al., 2011; Salone et al., 2007; Takenaka et al., 2021). In four other DYW type PPR
197 proteins of *Naegleria gruberi,* the HAE motif can be identified as well, while five, including NgPPR51, show
198 the dominant triplet HSE. Ng76525 exhibits an HCE instead of HxE, but displays several other mutations
199 including the deletion of the PG box and loss of functionality is likely (Supplementary data 2). NgPPR51 in
200 contrast to NgPPR45 also lacks conservation of the WGAL motif at the start of the E1 domain, also
201 conserved in most, but not all plant type PPR editing factors.

Complementation studies in PpPPR78 KO plant line

The characteristic differences of the DYW domains of the *N. gruberi* DYW type PPR proteins, the lack of information about the functionality of DYW domains outside of land plants and the question of functional compatibility of these domains with plant counterparts separated in evolution by 1.5 billion years brought us to the initial design of our experiments.

We generated different chimeras of PpPPR78 of *P. patens* with C-terminal domains exchanged with those of the *N. gruberi* DYW-type PPR proteins NgPPR45 and NgPPR51. Three different fusion points were chosen based on the assumed functionality of the catalytic C-terminal domain (Figure 4). The first set of chimeras consisted of the PPR stretch of PpPPR78 and the complete C-terminal extensions (E1E2DYW) of NgPPR45 or NgPPR51, respectively. In the second set of chimeras, the PPR region and the E1 and E2 motifs of PpPPR78 were combined with the DYW domain of the two *Naegleria* proteins, respectively. The last set of chimeras was built of the PPR region, E1, E2 and the N-terminal part of the DYW domain of PpPPR78 with the C-terminal part of the DYW domain of the two *Naegleria* proteins, respectively. This C-terminal part of the DYW domain was initially defined as core DYW domain (Cheng et al., 2016; Lurin et al., 2004) and some fusion proteins of different plant editing factors with exchanged „short“ DYW domain (DYW*) were proven to be functional (Ichinose and Sugita, 2018; Schallenberg-Rüdinger et al., 2017).

The different chimeras were transformed into the KO PpPPR78 line using protoplast transformation and inserted into the *P. patens* intergenic (PIG) region via homologous recombination (Okuda et al., 2009; Oldenkott et al., 2020). For each construct a minimum of three stable lines were selected and expression of the transgene, as well as RNA editing ratios, were evaluated (Supplementary table 2).

None of the plant lines expressing chimeras with the N-terminal part of PpPPR78 and the C-terminal part of NgPPR51 showed any editing at the two editing sites of PpPPR78. A different picture emerged for the chimeras with C-terminal domains of NgPPR45. The chimera harboring the complete E1E2DYW domain of NgPPR45 edited cox1eU755SL with 24% - 82 % efficiency (Figure 5), depending on the stable plant line

investigated. The second editing site assigned to PpPPR78, rps14eU137SL, however, was not edited in any of the plant lines. Chimeras of PpPPR78 and shorter parts of the C-terminal domain of NgPPR45 likewise did not show editing in any of the generated stable plant lines investigated.

Equal expression levels of different complementation constructs do not lead to same editing levels

Expression of chimeric proteins was driven by the strong actin promoter (XY). We investigated the expression levels of the different chimeras in stable complementation lines in comparison to the native expression of PpPPR78 in *P. patens* wild type Gransden, which is in general low (Schallenberg-Rüdinger et al., 2017), via quantitative real time PCR. In plants expressing PpPPR78-NgPPR45EDYW, the different editing efficiencies of *cox1* correlated well with the corresponding expression of the fusion protein itself (Figure 5). The plant line with only 2.7fold transgene expression compared to the wild type expression, exhibited 28 % editing of cox1eU755SL. With an increase of expression to up to 1000fold, the editing efficiency increased to 82 %, but did not reach >99 % like in the wild type Gransden.

Other chimeras, which did not show any editing, were expressed to comparable ratios (Figure 5). In case of PpPPR78-NgPPR45DYW even a 1,000-fold higher expression than the PpPPR78 expression in the wild type did not lead to any editing of one of the targets. Thus, different expression levels between lines expressing the same construct do influence editing rates, but differing expression levels are not the driving force for diverging editing capacities of different constructs.

Discussion

With our here presented study, we confirm that the C-terminal domain of DYW-type PPR protein NgPPR45 of the protist *Naegleria gruberi* has cytidine deaminase activity and is able to edit specific cytidines when fused to the upstream PPR array of PpPPR78 for RNA binding *in planta*.

Compatibility of the DYW domain of NgPPR45 of *Naegleria gruberi* with moss editing factor PpPPR78

This is to some extent surprising, as meanwhile studies have shown that the interchangeability of domains between different DYW-type PPR proteins is limited (Ichinose & Sugita, 2018; Maeda et al., 2022; Yang et al., 2023#2). Even the exchange of C-terminal domains of different DYW-type PPR proteins of the same species did not always result in functional editing factors *in planta*. *P. patens'* editing factor PPR56, for example, edits its main target nad4eU272SL with lower efficiencies (35 % instead of 95 %), when its DYW domain is replaced by the DYW domain of the chloroplast editing factor PpPPR45, but not at all when it´s replaced by one of the mosses other eight editing factors (Ichinose and Sugita, 2018, Figure 2).

Ichinose and colleagues identified the conservation of residues 37–42 of the DYW* domain (in Figure 2 positions 166-171) to be important for successful exchangeability, but it remained open, if the motif participates in the recognition of the -3 to -1 region upstream of assigned editing sites (Ichinose and Sugita, 2018; Okuda et al., 2014), or if that region interacts with upstream regions of the PPR protein. Tin contrast, the fusion of the PPR region of PpPPR78 with different portions of the C-terminal domain of editing factor PpPPR79 of *Physcomitrium patens* was shown to result in functional editing factor chimeras (Schallenberg-Rüdinger et al., 2017). This motivated us to use PpPPR78 in this study. In contrast to PpPPR78 and PpPPR79 sharing the same amino acids in position 37-42 of the DYW domain, NgPPR45 and NgPPR51 present another motif than PpPPR78 or any other *P. patens* editing factor (Figure 2). This could possibly be one explanation for the lack of functionality of PpPPR78 being fused to the DYW domain or portions of the DYW domain of NgPPR45 or NgPPR51 only.

The complete C-terminal domain (E1E2DYW domain) of PpPPR78, however, can be replaced by the one of NgPPR45 of *Naegleria gruberi* and led to editing of up to 82% of the cox1eU755SL site in *P. patens*. The E1 and E2 motifs, sharing similarities to TPR repeats which mediate protein-protein interaction (Blatch and Lässle, 1999), might interact with the DYW domain of NgPPR45 and an exchange of the complete C-terminal domain might therefore be successful in contrast to the DYW domain only. Differences of the E1,

E2 and DYW domain of NgPPR45 in comparison to the consensus of the plant E1, E2 and DYW domains might additionally hinder the interaction of the protists E1 and E2 domain with the plant DYW domain or *vice versa*.

Restricted influence of editing factor expression levels on editing rates

Natively DYW-type PPR proteins are only lowly expressed in land plants and the number of individual editing factors in mitochondria was found to be likewise low (Fuchs et al., 2020; Lurin et al., 2004). The more surprising is the high editing efficiency of most editing sites *in planta* (Bentolila et al., 2013; Small et al., 2023). Cox1eU755SL is also one of the sites, which is fully edited *in planta* in different tissues investigated (Rüdinger et al., 2009; Rüdinger et al., 2011b; Uchida et al., 2011), although PpPPR78 is only expressed to low levels under standard growth conditions (Figure 5, Schallenberg-Rüdinger et al., 2017). The second editing site assigned to PpPPR78, rps14eU137SL, is edited in 60-80% of transcripts in *P. patens* (Rüdinger et al., 2009; Rüdinger et al., 2011b). Intriguingly, overexpression of PpPPR78 in the KO PPR78 background of *P. patens* could increase editing of rps14eU137SL to 100%. A chimera of PpPPR78 and PpPPR79 could likewise fully edit the *cox1* site, but rps14eU137SL in only 26 % up to 63 % of transcripts (Schallenberg-Rüdinger et al., 2017). We postulated that the PPR protein RNA interaction is the limiting factor and not the abundance of the protein. The same holds true for our different PpPPR78-NgPPR45 and PpPPR78-NgPPR51 chimeras. Even expressions 1,000-fold higher than PpPPR78 in the wild type did not lead to RNA editing by the chimera with replaced DYW domain only. Highest expression of PpPPR78 with the E1E2DYW domain of PpPPR45 also did not lead to any editing of rps14eU137SL, revealing once more that this site is more difficult to be addressed than its counterpart in the *cox1* transcript.

The expression level of the PpPPR78-NgPPR45EDYW, however, influences the editing efficiency of cox1eU755SL. Five independent stable complementation lines show editing of 24% up to 82% and the editing investigated in each line correlates well with the PPR protein expression levels tested via quantitative RT-PCR (Figure 5).

295 Other studies also confirmed that once a PPR protein or PPR chimera can edit its assigned target, the
296 expression of transcript and subsequently the amount of editing factors in the organelle affects the editing
297 efficiencies at particular sites (Loiacono et al., 2022; Oldenkott et al., 2020).

298 ## The DYW domain of NgPPR45 shares similarity with reverse editing enzymes, but still acts in C-to-U RNA
299 editing

300 DYW-type PPR proteins are mainly restricted to land plants and sporadic appearance of such proteins in
301 species outside of the plant kingdom can most likely be explained by horizontal gene transfer (HGT, (Fu et
302 al., 2014; Knoop and Rüdinger, 2010; Schallenberg-Rüdinger et al., 2013b). Given the divergent sequence
303 conservation of the DYW-type PPR proteins of *Naegleria gruberi* and land plants, an HGT very early in plant
304 evolution some 500 million years ago has most likely seeded the DYW-type PPR protein genes in a protist
305 related to *Naegleria* (Knoop and Rüdinger, 2010). While the cytidine deaminase signature HxE(x)$_n$CxxC is
306 highly conserved in all DYW domains encoded in *N. gruberi*, other motifs which were also identified as key
307 motifs for a fully functional DYW domain were found to be modified in NgPPR45 and NgPPR51 in
308 comparison to the DYW-type PPR proteins of *P. patens* (Figure 2).
309 Whereas the E1E2DYW domain of NgPPR51 indeed did not show cytidine deaminase activity when fused
310 to the PPR array of PpPPR78, the E1E2DYW domain of NgPPR45 did.
311 Interestingly, the C-terminal domain of NgPPR45 shows a relaxed conservation of certain motifs, which
312 are also degenerated in C-terminal domains of the KPAxA PLS-type PPR proteins. This is a subtype of DYW-
313 type PPR proteins exclusively found in hornworts, ferns and lycophytes and suspected to catalyze the
314 reverse RNA editing reaction from U-to-C, which only appears in the organelles of these land plant clades
315 (Gerke et al., 2020; Gutmann et al., 2020).
316 The SW of the PG box is degenerated, as well as the SHP motif, which is completely missing in many of the
317 KPAxA PLS-type PPR proteins (Gerke et al., 2020; Gutmann et al., 2020; Takenaka et al., 2021). The C-
318 terminal DYW triplet is modified into a GYW and the HSExLA motif conserved in all so far characterized C-

to-U editing factors in land plants (citation missing) is modified into a HAExLA, what is also dominantly found in the KPAxA PLS-type PPR proteins.

That this C-terminal domain is now able to edit a cytidine into a uridine *in planta* is surprising, but might also point towards the hypothesis that in early land plant evolution, C-to-U and U-to-C RNA editing was present (Small et al., 2020), with initial RNA editing enzymes that operated in both directions (Knoop, 2022; Small et al., 2023).

Recent phylogenetic concepts assume bryophytes (hornworts, mosses, liverworts) as one monophyletic group (Puttick et al., 2018; Su et al., 2021), what would consequently mean, that U-to-C RNA editing was gained together with C-to-U RNA editing in first land plants, possible linked to the conquest of land (Fujii et al., 2013).

A recent study with synthetic KPAxA PLS-type PPR proteins successfully tested in the two heterologous systems of *Escherichia coli* and humans, also confirmed that these proteins can indeed act in both directions (Ichinose et al., 2022).

Successful transfer of editing factors between different genetic systems

With the proof of functionality of the E1E2DYW domain of NgPPR45 of *Naegleria gruberi*, we have shown that an evolutionary early transfer of a functional editing factor had been successful. This is a further example of the transferability of these editing factors, with huge families of up to several 1000 members in land plants (Banks et al., 2011; Gerke et al., 2020), but with only single exceptional cases in species of other kingdoms accompanied by only low numbers of editing sites identified in their mitochondrial transcriptomes (Bundschuh, 2015; Fu et al., 2014; Rüdinger et al., 2011b).

In recent studies it was shown that DYW-type PPR proteins can be transferred into other genetic systems. Transferred moss editing factors PpPPR56 and PpPPR65 were shown to edit there delivered targets in the bacterium *Escherichia coli*, in human cell cytosol and *in vitro* (Oldenkott et al., 2019; Lesch et al., 2022; Hayes et al., 2020). Another moss editing factor PpPPR79 edits matching editing sites, when transferred

into flowering plant *Arabidopsis thaliana* (Oldenkott et al., 2020). Editing factors of *Arabidopsis thaliana*, however, were not functional in *P. patens* or *E. coli* yet (Maeda et al., 2022; Oldenkott et al., 2020). One reason for this could be the lack of additional helper proteins, which are needed for efficient RNA editing of many sites in flowering plants (Maeda et al., 2022). Indeed, a synthetic editing factor engineered on the basis of DYW-type PPR protein CLB19 of *A. thaliana* was shown to increase its activity in *E. coli* when co-expressed with the seed plant specific RNA editing helper protein MORF9 (Gutmann et al., 2020; Royan et al., 2021). Other PPR protein chimeras with the PPR stretch of PpPPR56 combined with DYW domains of different *Arabidopsis thaliana* editing factors were also tested in bacteria and most of them were shown to be inactive (Maeda et al., 2022; Takenaka et al., 2021).

We also tested PpPPR78 with the E1E2DYW domain of NgPPR45 in the *E. coli* system. Neither the co-transcribed cox1eU755SL nor rps14eU137SL were edited by the chimera in the heterologous system. It remains to be seen, if bacterial factors hinder the editing activity of that protein in bacteria or if other unknown mediators are needed to support the editing in plant organelles.

The reduced editing of cox1eU755SL and the lack of editing of rps14eU137SL by chimera PpPPR78-NgPPR45EDYW could at the end also be a consequence of a reduced compatibility of the C-terminal domain of NgPPR45 with the upstream *cis* element of the *rps14* site (Takenaka et al., 2022).

NgPPR45 might be the best candidate to edit the two endogenous editing sites in *Naegleria gruberi* mitochondria

As NgPPR45 shows a clear target signal for mitochondrial localization and fits best to cox1eU1120HY in the mitochondrial transcriptome of *Naegleria gruberi* in comparison to the other identified DYW-type PPR proteins in *Naegleria,* participation of this editing factor in editing of that site is likely. The PPR stretch of NgPPR45 also moderately fits to the second editing site cox3eU780RW with five matches and only one mismatch in the core region of recognition. Such mismatches were even found to be relevant for proper editing of targets like shown for *P. patens* PPR65. Upon removal of the mismatch by modifying the target

sequence, ccmFCeU103PS is edited less efficiently than the original target in the heterologous *E. coli* system (Oldenkott et al., 2019).

Anyway, the final proof, that this protein edits one or even both editing sites in the mitochondrial transcriptome of Naegleria gruberi is still lacking. The high number of repetitive elements within the PPR stretch (Supplementary figure 1) hindered us to synthesize or amplify the complete protein gene for expression in an heterologous system (Lesch et al., 2022; Oldenkott et al., 2019) or at least to perform electromobility shift assays to test the binding to the appropriate targets (Kindgren et al., 2015; Matsuda et al., 2020; Schallenberg-Rüdinger et al., 2013a).

Unknown functions of the other nine DYW-type PPR proteins in *Naegleria gruberi*

The function of the other DYW-type PPR proteins encoded in the genome of the protist, however, remains puzzling. Protein models for genes Naegr76525, Naegr46207 and Naegr66503 consist of up to six PPR repeats and a degenerated C-terminal domain only. These genes might be pseudogenes or cryptic truncated PPR protein genes as also found sporadically in diverse species outside of land plants such as chlorophyte algae (Gutmann et al., 2020) or in the charophyte algae *Nitella hyalina* (Schallenberg-Rüdinger et al., 2013b). Naegr76708 shows a deletion of the E1 and E2 domain and an incomplete DYW domain and can be excluded as functional RNA editing factor as well. For NgPPR51, a function in mitochondria is likely due to the clear mitochondrial signal peptide of the protein, but no editing target matches the PPR array of that protein. If the protein has another function in RNA processing via interaction with a particular intergenic RNA region like chloroplast DYW-type PPR protein CRR2 in *A. thaliana* (Hashimoto et al., 2003; Ruwe et al., 2018) or participating in splicing like PpPPR43 in *P. patens* (Ichinose et al., 2012) cannot be answered yet.

For the other four DYW-type PPR proteins, a function in RNA editing cannot be ruled out to date. Naegr70351 has 14 PPR repeats which are likely not arranged in the PLS triplet manner, typically found in editing factors. DYW-type PPR proteins Naegr69406, Naegr76708 and Naegr32041, when the N-terminal

391 elongation is included, present long PPR binding regions with 25, 23 and 28 PPR repeats, respectively.
392 These are PPR repeat regions longer than the stretch of most editing factors identified in land plants like
393 *Arabidopsis thaliana* or *Physcomitrium patens* (editing factor overview available via the PREPACT search
394 tool Edifacts (Lenz et al., 2018)). As only NgPPR51 and NgPPR45 possess a clear signal peptide for
395 mitochondrial localization, one may speculate on a function in RNA editing of nuclear-cytosolic transcripts.
396 In nature, no single case of a DYW-type PPR protein acting on a nuclear transcript is known so far, but it
397 was shown recently that moss editing factor PpPPR56 upon expression in human cells not only edits its
398 endogenous delivered target but also numerous off-targets in the cytosolic transcriptome (Lesch et al.,
399 2022). This proves the functionality of DYW-type editing factors in the cytosol in principle.

400 # Materials and Methods

401 ## Amplification and sequence analysis of NgPPR45423 and NgPPR45424

402 Nucleic acids were prepared from *Naegleria gruberi* strain NEG-M as described in Rüdinger et al. (2011a) and
403 kindly provided by Dr. Lillian Fritz-Laylin (Fritz-Laylin et al., 2010). RNA was treated with DNase I (Thermofisher)
404 to remove vestiges of DNA. First strand cDNA was synthesized using the NucleoSpin RNA kit (Macherey-Nagel)
405 and oligodT_{18} primers. Different primers (Supplementary table 3) were used to amplify NgPPR45423 and
406 NgPPR45424 on DNA and cDNA level with Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific).
407 PCR products were gel-purified (BLIRT kit), sequenced (Macrogen Europe) and aligned with the *Naegleria*
408 *gruberi* genome sequence (Fritz-Laylin et al., 2010), 2010, NCBI Genebank entry NW_003163326.1) using
409 MEGA7 (Kumar et al., 2016) and manual adjustment (Supplementary data 1 and 3).

410 ## Plant material and growth conditions

411 *Physcomitrium patens* (Hedw.) Bruch & Schimp., Gransden (Rensing et al., 2020; Rensing et al., 2008) wild
412 type and KO PpPPR78 Gransden (Rüdinger et al., 2011b) were cultivated following Oldenkott et al. (2020).
413 Gametophores were cultivated on modified Knop medium plates (250 mg/L KH_2PO_4, 250 mg/L KCl, 250
414 mg/L $MgSO_4x7H_2O$, 1000 mg/ $Ca(NO_3)_2x4H_2O$, 12.5 mg/L $FeSO_4x7H_2O$, 0.22 mM $CuSO_4$, 0.19 mM $ZnSO_4$,

10 mM H_3BO_3, 0.1 mM Na_2MoO_4, 2 mM $MnCl_2$, 0.23 mM $CoCl_2$, and 0.17 mM KI, pH 5.8, 1% [w/v] agar; (Rüdinger et al., 2011b) at 21°C, with a 16-h-light (photosynthetic photon flux density of 65 mmol/m^2/s, neon tubes, Osram HO 39W/865 Lumilux Cool Daylight)/8-h-dark cycle.

Complementation of *Physcomitrium patens* ppr78 KO line

Physcomitrium patens DNA was prepared using the CTAB preparation method (Doyle and Doyle, 1987). Respective primers were used to amplify parts of PpPPR78 and NgPPR45 or NgPPR51 coding sequences in initial PCRs to be fused subsequently in an overlap extension PCR (Higuchi et al., 1988), primers see supplementary table 2). Fusion PCR products were gel-purified and inserted into plasmid PIG_AN between the rice actin1 promoter and the nos terminator. The constructs are flanked by regions homologous to the *P. patens* intergenic (PIG) region (Okuda et al., 2009) to be inserted into the *P. patens* genome via homologous recombination (Schallenberg-Rüdinger et al., 2017). Constructs were introduced into KO PpPPR78 ecotype Gransden protoplasts using polyethylene glycol–mediated transformation (Hohe et al., 2004) as described in Oldenkott et al. (2020). For selection, mutant lines were cultivated on Knop agar (1% w/v) with 30 µg/ml hygromycin B.

For detection of stable integration of the gene construct, DNA was prepared using the quick extraction method described in Edwards et al. 1991. Genotyping PCRs using TaqNova DNA Polymerase (Blirt) or Q5 High-Fidelity DNA Polymerase (New England Biolabs) were performed following Schallenberg-Rüdinger et al. (2017). Primers Act1Pfor and NosTrev2 were used to confirm the presence of the transgene. The correct orientation of the construct and the insertion into the PIG region was tested with primer combination PpPIG1gen_for and Act1Pfor (for primers see Supplementary table 3). Transgenic lines, which showed expected PCR products with both primer combinations were considered as stable lines.

RNA editing detection and transgene expression

Three to five single gametophores of each stable plant line and control lines (KO PpPPR78, OE PpPPR78 DYW domain truncated, OE PpPPR78, OE PpPPR78 fused with PpPPR79 (E1E2 and) DYW domain, published in Schallenberg-Rüdinger et al. (2017)) were transferred to fresh Knop plates for standardized growth under conditions as described above. After 42 days, equal amounts of plants were harvested for RNA extraction by using the NucleoSpin RNA kit (Macherey-Nagel), followed by DNase I treatment (Thermo Fisher Scientific).

To detect RNA editing, cDNA was synthesized from DNase treated RNA by using random hexamer primer (Roth) and RevertAid reverse transcriptase (Thermo Fisher Scientific). Primer pair 11altfor/16rev and PPrps14for/PPrps14rev was used to amplify the target sequence containing editing positions cox1eU755SL and rps14eU137SL, respectively. PCR assays included cDNA corresponding to 4 ng of RNA, 0.2 µM of each primer, 0.2 mM dNTPs, 1 unit of Taq polymerase with 1x supplied PCR buffer in double-distilled water in total volume of 25 µl. Amplification went through 5 min of initiation followed by 35 cycles, each including 30 sec degeneration at 96 °C, 30 sec annealing at 45 °C, 1 min elongation at 72 °C, and a 5 min finishing step at 72 °C. Purified PCR products were sequenced and RNA editing was detected by using BioEdit 7.0.5.3 (Hall, 1997). The editing efficiency was quantified by the ratio of the thymidine peak height to the sum of thymidine and cytidine peak heights in the chromatogram position, corresponding to the editing site.

Quantitative real time PCR

To quantify the expression of the inserted PPR chimera in the different generated plant lines, 62 ng of DNase treated RNA were used for cDNA synthesis per 20 µl assay with oligodT_{18} primers. The real-time PCR was performed using the SYBR green master mix (Invitrogen) with cDNA corresponding to 3.1 ng initial total RNA per 20 µl assay. cDNA was analyzed on a Bio-Rad CFX96 Real-Time system with the following program: 95°C for 10 min, 50 cycles of 95°C for 15 sec, 58°C for 20 sec and 72°C for 20 sec, finishing with the melt stage: 95°C for 30 sec, 48°C to 95°C with 0.5°C increase each 5 sec. Primer qpcr78-for-all1 and

460 qpcr78-rev-all1 were used to amplify PpPPR78 and PpPPR78 chimeras. Reference gene *Ade PRT*
461 (Phypa_443007) was used for normalization as recommended by Bail et al. (2013). Triplicate
462 measurements were performed for each of three biological replicates. Measurements were analyzed using
463 the delta-delta-Ct method. Melting curves were analyzed to ensure product specificity.

464 ## Acknowledgements

465 We wish to thank Dr. Lilian Fritz-Laylin for providing DNA and RNA of *Naegleria gruberi* and Monika
466 Polsakiewicz and Sarah Brenner for technical assistance. The authors gratefully acknowledge funding by
467 Deutsche Forschungsgemeinschaft (DFG) SCHA 1952/2-2 to M.S.-R.

468 ## Author contributions

469 B.O and M.S-R. cloned the constructs. B.O. and E.L. introduced them into *P. patens* KO PPR78. Y.Y, B.O and
470 E.L. performed the genotyping and RNA editing analysis. Y.Y. did the genome locus analysis and performed
471 the qRT-PCR experiments with initial help of E.L.. V.K and M.S-R designed and supervised the study
472 program and contributed to data analyses. Y.Y. prepared the figures. M.S-R wrote the manuscript, which
473 was read, edited and ultimately approved by all co-authors.

Figures

Figure 1. The ten DYW-type PPR proteins of *Naegleria gruberi*

Coding sequences of the ten DYW-type PPR proteins encoded in the genome of *Naegleria gruberi* were subscribed based on Fritz-Laylin et al. (2010) and motifs and amino acids at binding positions 5th and Last (L) of each PPR repeat were identified using the PPR finder (https://ppr.plantenergy.uwa.edu.au, Cheng et al 2016). N-terminal sequences upstream of the PLS array, which were predicted as signal peptides for mitochondrial localization are colored in orange (Naegr51788), other sequences and spacer sequences within the PPR protein that were not recognized as PPR motifs or C-terminal E1/E2(green)/DYW(blue) domains are displayed in grey (size adjusted by sequence length). The DYW domain of Naegr76525 and Naegr76708 showing amino acid deletions are colored in light blue. 5th and L amino acids of each PLS repeat are shown and P and S motifs, which are important for binding, are shaded in grey. Binding fit of S and P motifs to the corresponding nucleotides upstream of the two mitochondrial editing sites (bold, underlined) of *Naegleria gruberi*, cox1eU1120HY and cox3eU780RW, are highlighted in green for a match, in yellow for the less favored match and in red for a mismatch, based on the PPR-RNA binding code: T/S + N/S: A>G, T/S + D:G>A, N + S: C>U, N + D: U>C, N + N: C/U (Barkan et al., 2012), respectively..

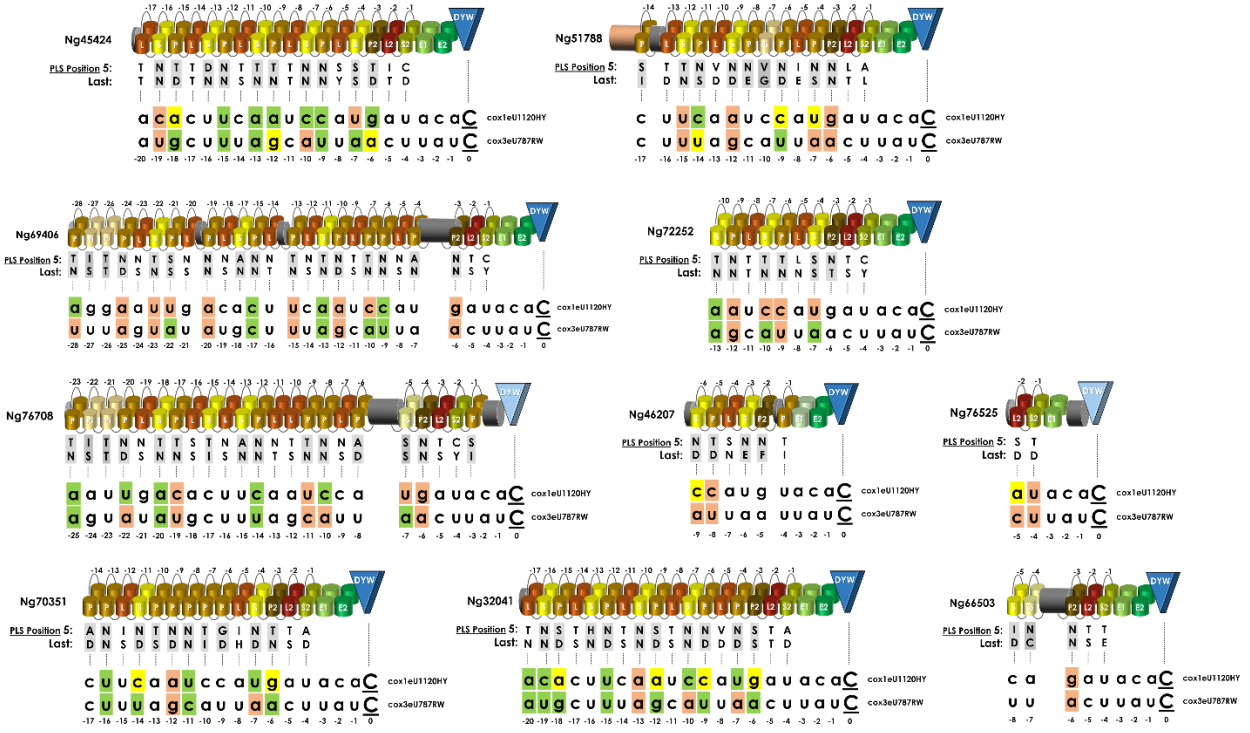
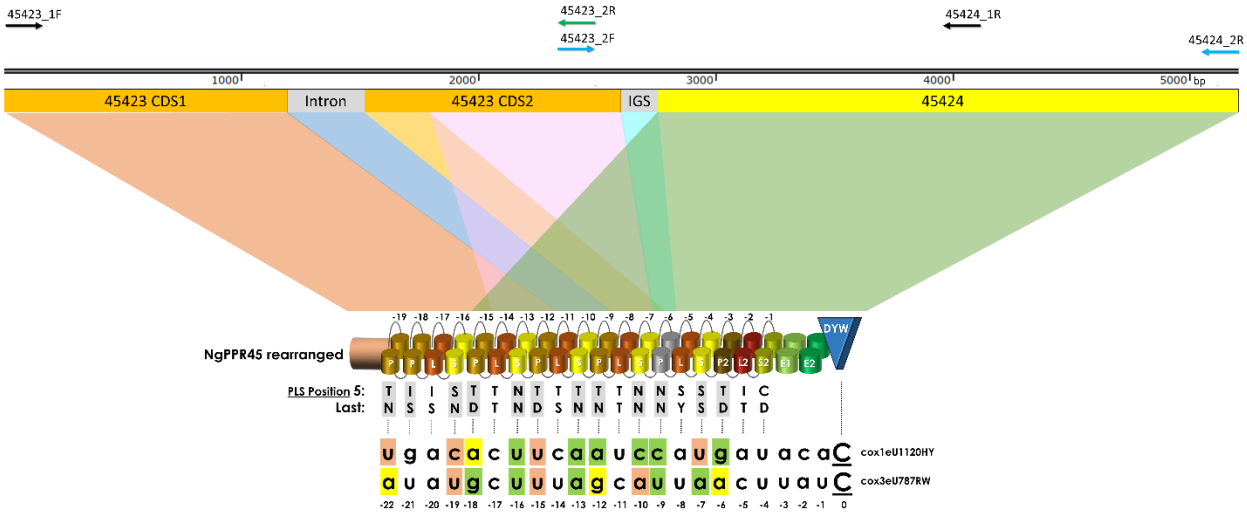


Figure 2. Combination of predicted genes Naegr45423 and Naegr45424 to the gene encoding editing factor NgPPR45.

A. Different primer pairs (arrows) binding to unique parts of predicted genes Nagegr45423 and Naegr45424 were used to amplify the investigated genome region. Sanger sequencing could identify the predicted intron of Naegr45423 and the intergenic region to be part of the coding region. As a result, Naegr45423 and Naegr45424 belong to one reading frame (Supplementary data 1). Sequences within Naegr45423 and Naegr45424 are highly repetitive (shown in different shading), which might have caused difficulties within the initial assembly. The co-transcription of the two predicted genes was finally proven on transcript level. IGS=predicted intergenic region. **B**. Gel electrophoresis of PCR products confirming the proximity of predicted gene regions Naegr45423 and Naegr45424 on genomic DNA level (left) with primer pair 45423_1F and 45423_2R (D1) and 45423_2F and 45424_2R (D2), respectively and on cDNA level with primer pair 45423_1F and 45424_1R (C1).



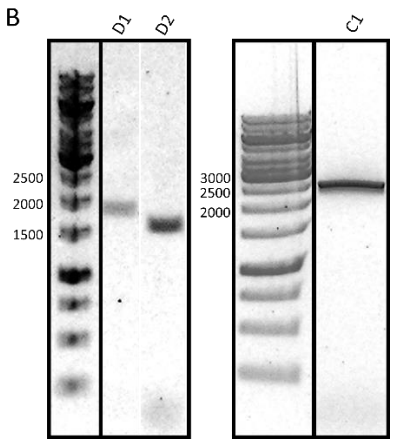


Figure 3. Alignment of C-terminal E1, E2 and DYW domain of NgPPR45 and NgPPR51 of the protist *Naegleria gruberi* and the nine DYW-type PPR editing factors of the moss *Physcomitrium patens*.

Light blue, blue and dark blue indicate amino acid conservation higher than 30%, 50% and 80%, respectively. Conserved motifs associated with the catalytic function of the DYW domain (PG box, zinc binding motifs, DYW motif) or the regulation of activity (gating domain consisting of α1 and ß3 and ß4, (Takenaka et al., 2021) are labeled. Start of the short DYW domain (DYW*) defined by Lurin et al. (2004) is indicated with an arrow, amino acids 37-42 of DYW* suggested to be important for domain compatibility (Ichinose and Sugita, 2018) are highlighted as well. The figure was prepared with Jarview 2.11.2.6.

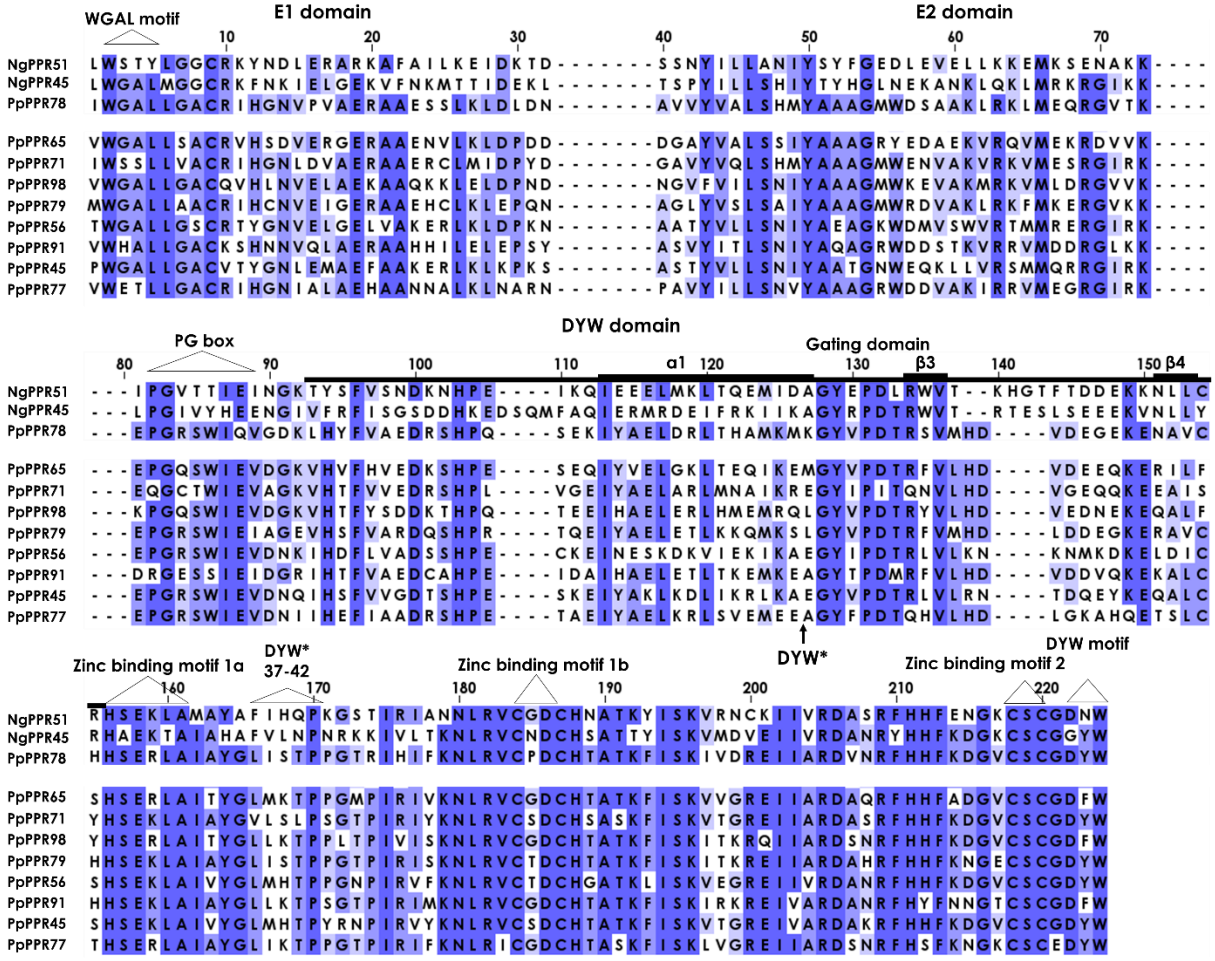


Figure 4. Functional complementation of *P. patens* KO PpPPR78 plants using PpPPR78 chimeras with C-termini of NgPPR45 or NgPPR51 of *Naegleria gruberi*.

Chimeric constructs of PpPPR78 and NgPPR45 with three different fusion points were introduced into the *ppr78* ko line. In wild type *P. patens,* PpPPR78 edits mitochondrial sites cox1eU755SL and rps14eU137SL, with >99% and 60-80% editing efficiency, respectively (shown on top, and >99% presented as 100%, as no C peak is recognizable in the Sanger sequencing chromatogram). Chimeras of PpPPR78 and NgPPR51 did not complement the *ppr78* ko line and editing was not regained in any generated line (see chromatograms of selected lines in the blue box and Supplementary table 2 for the complete data set). (green box) Complementation was successful when the complete C-terminal domain of PpPPR78 including the E1, E2 and DYW domain was replaced by the one of NgPPR45. This resulted in editing of 24% - 82% of cox1eU755SL (shown is the chromatogram of a line with moderate editing of 59%). The rps14eU137SL site was not edited in these complementation lines either.

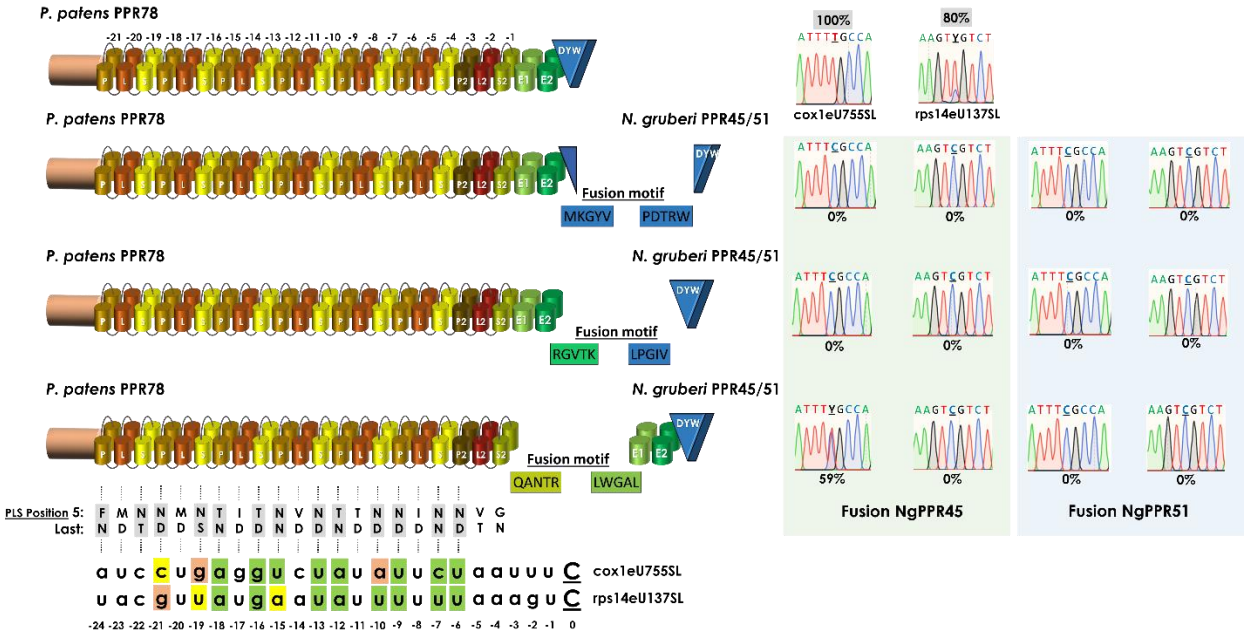
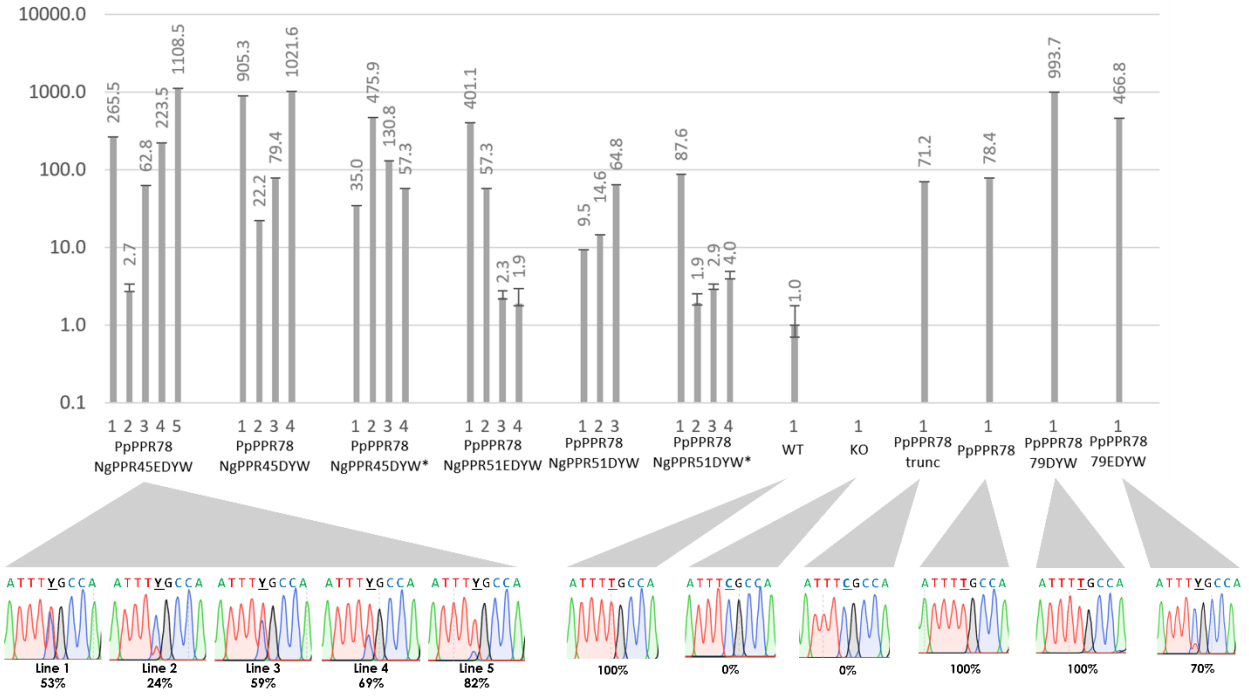


Figure 5. Quantitative real time PCR analysis of complementation lines of *P. patens* KO PpPPR78.

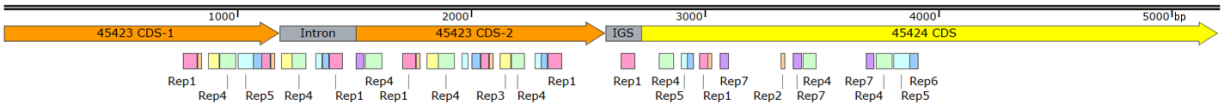
The expression of different PpPPR78 chimeras in each stable overexpression line was analyzed by quantitative RT-PCR in comparison to wildtypic PpPPR78 expression and other complementation lines of an earlier study (right side, (Schallenberg-Rüdinger et al., 2017). KO PpPPR78 was used as the negative control. The values are means of 3 biological replicates (error bars indicate SD). The regain of editing of site cox1eU755SL correlates with the expression levels of introduced PpPPR78-NgPPR45EDYW in the investigated complementation lines.



Supplementary information

Supplementary figure 1. Annotated genome region of Naegr45423 and Naegr45424 contains highly repetitive regions.

Repetitive regions ("Rep") are indicated below in different colored bars. From the end of Naegr45423 CDS-1 until mid of the Naegr45424 CDS, seven different repetitive regions were identified, which are also appearing within the putative intron region of Naegr45423 and the putative intergenic region (IGS). Figure prepared with Snapgene viewer V6.2.1. For details see Supplementary data 1.



Supplementary table 1: Output of PPR Finder for the different PLS-type PPR proteins of *Naegleria gruberi*

PPR models presented on the PPR plantenergy webpage subtool "PPR" (https://ppr.plantenergy.uwa.edu.au/ppr/) and predicted via search tool option "Search for PPR" (https://ppr.plantenergy.uwa.edu.au/fasta/) are displayed for each of the DYW-type PPR proteins from *N. gruberi*, if available. Detected PPR motifs and C-terminal extensions are listed with the region of the motif (start-end), length, score, type, amino acids associated with nucleotide recognition (2nd, 5th and last) and sequence of the motifs detected. End and start of motifs with gaps in between or with an overlap are highlighted in red (column "start-end"). Length number of motifs with unusual length are colored in red (column "length").

Supplementary table 2. *Physcomitrium* KO PPR78 complementation lines

Constructs inserted into KO PPR78 are given with the transgenic line number (Line No) and editing efficiencies (Ed) for both editing sites cox1eU755SL and rps14eU137SL measured for three independent biological replicates (REP) for each line investigated. Average (Ave) and standard deviation (SD) is calculated. Expression levels in relation to WT expression of PPR78 are displayed as well.

Supplementary table 3. Oligonucleotides

Oligonucleotides used in this study. All oligonucleotides were synthesized by IDT (Integrated DNA technologies Europe, BVBA, Leuven, Belgium).

Supplementary table 4. Signal peptide prediction of assembled NgPPR45 and N-terminal extended Ng32041

Results of the prediction of localization of NgPPR45 and N-terminal extended Ng32041 performed with TargetP 2.0 (https://services.healthtech.dtu.dk/service.php?TargetP-2.0) and WoLFPSORT (https://wolfpsort.hgc.jp/).

Supplementary Data

Supplementary data 1. Alignment of sequences to rearrange the annotation of the gene encoding for NgPPR45

The genome region of Naegr45423 and Naegr45424 is aligned with Sanger sequenced PCR products based on DNA and cDNA amplification.

Supplementary data 2. Alignment of the ten DYW-type PPR proteins of *Naegleria gruberi*

Presented are the ten DYW-type PPR proteins of *Naegleria gruberi* aligned with the ten DYW-type PPR proteins of *Physcomitrium patens* for comparison. The rearranged NgPPR45 is presented as well as the N-terminally elongated Naegr32401.

Supplementary data 3. Sanger sequencing file for NgPPR45 assembly and editing analysis of plants

Collection of Sanger sequencing results of PCR products used for NgPPR45 assembly and Sanger sequencing results of PCR products to evaluate the editing of cox1eU755SL and rps14eU137SL in the different transgenic lines generated within this study.

References

Andrés-Colás, N., Q. Zhu, M. Takenaka, B.D. Rybel, D. Weijers, and D.V.D. Straeten. 2017. Multiple PPR protein interactions are involved in the RNA editing system in Arabidopsis mitochondria and plastids. *PNAS*.

Bail, A.L., S. Scholz, and B. Kost. 2013. Evaluation of Reference Genes for RT qPCR Analyses of Structure-Specific and Hormone Regulated Gene Expression in Physcomitrella patens Gametophytes. *PLoS One*. 8.

Banks, J.A., T. Nishiyama, M. Hasebe, J.L. Bowman, M. Gribskov, ..., and I.V. Grigoriev. 2011. The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. *Science*. 332:960-963.

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

Bentolila, S., A.M. Babina, A. Germain, and M.R. Hanson. 2013. Quantitative trait locus mapping identifies REME2, a PPR-DYW protein required for editing of specific C targets in Arabidopsis mitochondria. *RNA Biol*. 10:1520-1525.

Bentolila, S., W.P. Heller, T. Sun, A.M. Babina, G. Friso, K.J. van Wijk, and M.R. Hanson. 2012. RIP1, a member of an Arabidopsis protein family, interacts with the protein RARE1 and broadly affects RNA editing. *Proc Natl Acad Sci U S A*. 109:E1453-1461.

Blatch, G.L., and M. Lässle. 1999. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays*. 21:932-939.

Boussardon, C., A. Avon, P. Kindgren, C.S. Bond, M. Challenor, C. Lurin, and I. Small. 2014. The cytidine deaminase signature HxE(x)n CxxC of DYW1 binds zinc and is necessary for RNA editing of ndhD-1. *New Phytol*. 203:1090-1095.

Boussardon, C., V. Salone, A. Avon, R. Berthome, K. Hammani, K. Okuda, T. Shikanai, I. Small, and C. Lurin. 2012. Two interacting proteins are necessary for the editing of the NdhD-1 site in Arabidopsis plastids. *Plant Cell*. 24:3684-3694.

Bundschuh, R. 2015. The search for the determinants of insertional RNA editing in Physarum polycephalum mitochondrial RNAs. *In* The First International Workshop on Physarum Transport Networks 2015.

Bundschuh, R., J. Altmüller, C. Becker, P. Nürnberg, and J.M. Gott. 2011. Complete characterization of the edited transcriptome of the mitochondrion of Physarum polycephalum using deep sequencing of RNA. *Nucleic Acids Research*. 39:6044-6055.

Cheng, S., B. Gutmann, X. Zhong, Y. Ye, M.F. Fisher, F. Bai, I. Castleden, Y. Song, B. Song, J. Huang, X. Liu, X. Xu, B.L. Lim, C.S. Bond, S.M. Yiu, and I. Small. 2016. Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. *Plant J*. 85:532-547.

Diaz, M.F., S. Bentolila, M.L. Hayes, M.R. Hanson, and R.M. Mulligan. 2017. A protein with an unusually short PPR domain, MEF8, affects editing at over 60 Arabidopsis mitochondrial C targets of RNA editing. *Plant J*. 92:638-649.

Dong, S., C. Zhao, S. Zhang, H. Wu, W. Mu, T. Wei, N. Li, T. Wan, H. Liu, J. Cui, R. Zhu, B. Goffinet, and Y. Liu. 2019. The Amount of RNA Editing Sites in Liverwort Organellar Genes Is Correlated with GC Content and Nuclear PPR Protein Diversity. *Genome Biology and Evolution*. 11:3233-3239.

Doyle, J.J., and J.L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*. 19:11-15.

Duff, R.J. 2006. Divergent RNA editing frequencies in hornwort mitochondrial nad5 sequences. *Gene*. 366:285-291.

Faktorová, D., R.E.R. Nisbet, J.A.F. Robledo, E. Casacuberta, L. Sudek, ..., and J. Lukeš. 2020. Genetic tool development in marine protists: emerging model organisms for experimental cell biology. *Nature Methods*. 17:pages481–494.

Fritz-Laylin, L.K., S.E. Prochnik, M.L. Ginger, J.B. Dacks, M.L. Carpenter, M.C. Field, A. Kuo, A. Paredez, J. Chapman, J. Pham, S. Shu, R. Neupane, M. Cipriano, J. Mancuso, H. Tu, A. Salamov, E. Lindquist, H. Shapiro, S. Lucas, I.V. Grigoriev, W.Z. Cande, C. Fulton, D.S. Rokhsar, and S.C. Dawson. 2010. The genome of Naegleria gruberi illuminates early eukaryotic versatility. *Cell*. 140:631-642.

Fu, C.-J., S. Sheikh, W. Miao, S.G.E. Andersson, and S.L. Baldauf. 2014. Missing Genes, Multiple ORFs, and C-to-U Type RNA Editing in Acrasis kona (Heterolobosea, Excavata) Mitochondrial DNA. *Genome Biol Evol*. 6:2240-2257.

Fuchs, P., N. Rugen, C. Carrie, M. Elsässer, I. Finkemeier, J. Giese, T.M. Hildebrandt, K. Kühn, V.G. Maurino, C. Ruberti, M. Schallenberg-Rüdinger, J. Steinbeck, H.-P. Braun, H. Eubel, E.H. Meyer, S.J. Müller-Schüssele, and M. Schwarzländer. 2020. Single organelle function and organization as estimated from Arabidopsis mitochondrial proteomics. *Plant J*. 101:420-441.

Fujii, S., N. Sato, and T. Shikanai. 2013. Mutagenesis of Individual Pentatricopeptide Repeat Motifs Affects RNA Binding Activity and Reveals Functional Partitioning of Arabidopsis PROTON GRADIENT REGULATION3. *The Plant Cell*. 25.

Fujii, S., and I. Small. 2011. The evolution of RNA editing and pentatricopeptide repeat genes. *New Phytol*. 191:37-47.

Gerke, P., P. Szövényi, A. Neubauer, H. Lenz, B. Gutmann, R. McDowell, I. Small, M. Schallenberg-Rüdinger, and V. Knoop. 2020. Towards a plant model for enigmatic U-to-C RNA editing: the organelle genomes, transcriptomes, editomes and candidate RNA editing factors in the hornwort Anthoceros agrestis. *New Phytol*. 225:1976-1992.

Gipson, A.B., M.R. Hanson, and S. Bentolila. 2022. The RanBP2 zinc finger domains of chloroplast RNA editing factor OZ1 are required for protein-protein interactions and conversion of C to U. *Plant J*. 109:215-226.

Gott, J.M., N. Parimi, and R. Bundschuh. 2005. Discovery of new genes and deletion editing in Physarum mitochondria enabled by a novel algorithm for finding edited mRNAs. *Nucleic Acids Res*. 33:5063-5072.

Grewe, F., S. Herres, P. Viehover, M. Polsakiewicz, B. Weisshaar, and V. Knoop. 2011. A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte Isoetes engelmannii. *Nucleic Acids Res*. 39:2890-2902.

Guillaumot, D., M. Lopez-Obando, K. Baudry, A. Avon, G. Rigaill, A. Falcon de Longevialle, B. Broche, M. Takenaka, R. Berthome, G. De Jaeger, E. Delannoy, and C. Lurin. 2017. Two interacting PPR proteins are major Arabidopsis editing factors in plastid and mitochondria. *Proc Natl Acad Sci U S A*. 114:8877-8882.

Gutmann, B., S. Royan, M. Schallenberg-Rüdinger, H. Lenz, I.R. Castleden, R. McDowell, M.A. Vacher, J. Tonti-Filippini, C.S. Bond, V. Knoop, and I.D. Small. 2020. The Expansion and Diversification of Pentatricopeptide Repeat RNA-Editing Factors in Plants. *Molecular Plant*. 13.

Hall, T. 1997. Biological sequence alignment editor for Win95/98/NT/2K/XP.

Hashimoto, M., T. Endo, G. Peltier, M. Tasaka, and T. Shikanai. 2003. A nucleus-encoded factor, CRR2, is essential for the expression of chloroplast ndhB in Arabidopsis. *Plant J*. 36:541-549.

Hayes, M.L., and P.I. Santibanez. 2020. A plant pentatricopeptide repeat protein with a DYW-deaminase domain is sufficient for catalyzing C-to-U RNA editing in vitro. *J Biol Chem*. 295:3497-3505.

Hecht, J., F. Grewe, and V. Knoop. 2011. Extreme RNA editing in coding islands and abundant microsatellites in repeat sequences of Selaginella moellendorffii mitochondria: the root of frequent plant mtDNA recombination in early tracheophytes. *Genome Biol Evol*. 3:344-358.

Higuchi, R., B. Krummel, and R.K. Saiki. 1988. A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Res*. 16:7351-7367.

Hohe, A., T. Egener, J.M. Lucht, H. Holtorf, C. Reinhard, G. Schween, and R. Reski. 2004. An improved and highly standardised transformation procedure allows efficient production of single and multiple targeted gene-knockouts in a moss, Physcomitrella patens. *Curr Genet*. 44:339-347.

Ichinose, M., M. Kawabata, Y. Akaiwa, Y. Shimajiri, I. Nakamura, T. Tamai, T. Nakamura, Y. Yagi, and B. Gutmann. 2022. U-to-C RNA editing by synthetic PPR-DYW proteins in bacteria and human culture cells. *Communications Biology*. 5.

Ichinose, M., and M. Sugita. 2018. The DYW Domains of Pentatricopeptide Repeat RNA Editing Factors Contribute to Discriminate Target and Non-Target Editing Sites. *Plant Cell Physiol*. 59:1652-1659.

Ichinose, M., E. Tasaki, C. Sugita, and M. Sugita. 2012. A PPR-DYW protein is required for splicing of a group II intron of cox1 pre-mRNA in Physcomitrella patens. *Plant J*. 70.

Ichinose, M., M. Uchida, and M. Sugita. 2014. Identification of a pentatricopeptide repeat RNA editing factor in Physcomitrella patens chloroplasts. *FEBS Lett*. 588:4060-4064.

Iyer, L.M., D. Zhang, I.B. Rogozin, and L. Aravind. 2011. Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. *Nucleic Acids Res*. 39:9473-9497.

Kindgren, P., A. Yap, C.S. Bond, and I. Small. 2015. Predictable alteration of sequence recognition by RNA editing factors from Arabidopsis. *Plant Cell*. 27:403-416.

Knie, N., F. Grewe, S. Fischer, and V. Knoop. 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.

Knoop, V. 2011. When you can't trust the DNA: RNA editing changes transcript sequences. *Cell Mol Life Sci*. 68:567-586.

Knoop, V. 2022. C-to-U and U-to-C: RNA editing in plant organelles and beyond. *Journal of Experimental Botany*. erac488.

Knoop, V., and M. Rüdinger. 2010. DYW-type PPR proteins in a heterolobosean protist: plant RNA editing factors involved in an ancient horizontal gene transfer? *FEBS Lett*. 584:4287-4291.

Kumar, S., G. Stecher, and K. Tamura. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol*. 33:1870-1874.

Lenz, H., A. Hein, and V. Knoop. 2018. Plant organelle RNA editing and its specificity factors: enhancements of analyses and new database features in PREPACT 3.0. *BMC Bioinformatics*. 19:255.

Lesch, E., M.T. Schilling, S. Brenner, Y. Yang, O.J. Gruss, V. Knoop, and M. Schallenberg-Rüdinger. 2022. Plant mitochondrial RNA editing factors can perform targeted C-to-U editing of nuclear transcripts in human cells. *Nucleic Acids Res*. 50:9966-9983.

Loiacono, F.V., D. Walther, S. Seeger, W. Thiele, I. Gerlach, D. Karcher, M.A. Schöttler, R. Zoschke, and R. Bock. 2022. Emergence of Novel RNA-Editing Sites by Changes in the Binding Affinity of a Conserved PPR Protein. *Molecular Biology and Evolution*. 39.

Lurin, C., C. Andres, S. Aubourg, M. Bellaoui, F. Bitton, C. Bruyere, M. Caboche, C. Debast, J. Gualberto, B. Hoffmann, A. Lecharny, M. Le Ret, M.L. Martin-Magniette, H. Mireau, N. Peeters, J.P. Renou, B. Szurek, L. Taconnat, and I. Small. 2004. Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell*. 16:2089-2103.

Maeda, A., S. Takenaka, T. Wang, B. Frink, T. Shikanai, and M. Takenaka. 2022. DYW deaminase domain has a distinct preference for neighboring nucleotides of the target RNA editing sites. *The plant Journal*. 111:756-767.

Matsuda, T., M. Sugita, and M. Ichinose. 2020. The L motifs of two moss pentatricopeptide repeat proteins are involved in RNA editing but predominantly not in RNA recognition. *PLOS ONE*.

718 Meier, J.C., C. Henneberger, I. Melnick, C. Racca, R.J. Harvey, U. Heinemann, V. Schmieden, and R. Grantyn.
719 2005. RNA editing produces glycine receptor α3P185L, resulting in high agonist potency. *Nature*
720 *Neuroscience*. 8:736-744.
721 Miyata, Y., and M. Sugita. 2004. Tissue- and stage-specific RNA editing of rps 14 transcripts in moss
722 (Physcomitrella patens) chloroplasts. *J Plant Physiol*. 161:113-115.
723 Mungpakdee, S., C. Shinzato, T. Takeuchi, T. Kawashima, R. Koyanagi, K. Hisata, M. Tanaka, H. Goto, M.
724 Fujie, S. Lin, N. Satoh, and E. Shoguchi. 2014. Massive gene transfer and extensive RNA editing of
725 a symbiotic dinoflagellate plastid genome. *Genome Biol Evol*. 6:1408-1422.
726 Nishikura, K. 2010. Functions and regulation of RNA editing by ADAR deaminases. *Annu Rev Biochem*.
727 79:321-349.
728 O'Toole, N., M. Hattori, C. Andres, K. Iida, C. Lurin, C. Schmitz-Linneweber, M. Sugita, and I. Small. 2008.
729 On the expansion of the pentatricopeptide repeat gene family in plants. *Mol bio Evol.* 25:1120-
730 1128.
731 Okuda, K., A.L. Chateigner-Boutin, T. Nakamura, E. Delannoy, M. Sugita, F. Myouga, R. Motohashi, K.
732 Shinozaki, I. Small, and T. Shikanai. 2009. Pentatricopeptide repeat proteins with the DYW motif
733 have distinct molecular functions in RNA editing and RNA cleavage in Arabidopsis chloroplasts.
734 *Plant Cell*. 21:146-156.
735 Okuda, K., F. Myouga, R. Motohashi, K. Shinozaki, and T. Shikanai. 2007. Conserved domain structure of
736 pentatricopeptide repeat proteins involved in chloroplast RNA editing. *Proc Natl Acad Sci U S A*.
737 104:8178-8183.
738 Okuda, K., H. Shoki, M. Arai, T. Shikanai, I. Small, and T. Nakamura. 2014. Quantitative analysis of motifs
739 contributing to the interaction between PLS-subfamily members and their target RNA sequences
740 in plastid RNA editing. *Plant J*. 80:870-882.
741 Oldenkott, B., M. Burger, A.-C. Hein, A. Jörg, J. Senkler, H.-P. Braun, V. Knoop, M. Takenaka, and M.
742 Schallenberg-Rüdinger. 2020. One C-to-U RNA Editing Site and Two Independently Evolved Editing
743 Factors: Testing Reciprocal Complementation with DYW-Type PPR Proteins from the Moss
744 Physcomitrium (Physcomitrella) patens and the Flowering Plants Macadamia integrifolia and
745 Arabidopsis. *Plant Cell*. 32:2997-3018.
746 Oldenkott, B., K. Yamaguchi, S. Tsuji-Tsukinoki, N. Knie, and V. Knoop. 2014. Chloroplast RNA editing going
747 extreme: more than 3400 events of C-to-U editing in the chloroplast transcriptome of the
748 lycophyte Selaginella uncinata. *RNA*. 20:1499-1506.
749 Oldenkott, B., Y. Yang, E. Lesch, V. Knoop, and M. Schallenberg-Rüdinger. 2019. Plant-type
750 pentatricopeptide repeat proteins with a DYW domain drive C-to-U RNA editing in Escherichia coli.
751 *Communications Biology*. 2.
752 Puttick, M.N., J.L. Morris, T.A. Williams, C.J. Cox, D. Edwards, P. Kenrick, S. Pressel, C.H. Wellman, H.
753 Schneider, D. Pisani, and P.C.J. Donoghue. 2018. The Interrelationships of Land Plants and the
754 Nature of the Ancestral Embryophyte. *Curr Biol*. 28:733-745.
755 Rensing, S.A., B. Goffinet, R. Meyberg, S.-Z. Wu, and M. Bezanilla. 2020. The Moss Physcomitrium
756 (Physcomitrella) patens: A Model Organism for Non-Seed Plants. *Plant Cell*. 32:1361-1376.
757 Rensing, S.A., D. Lang, A.D. Zimmer, A. Terry, A. Salamov, H. Shapiro, T. Nishiyama, P.F. Perroud, E.A.
758 Lindquist, Y. Kamisugi, T. Tanahashi, K. Sakakibara, T. Fujita, K. Oishi, I.T. Shin, Y. Kuroki, A. Toyoda,
759 Y. Suzuki, S. Hashimoto, K. Yamaguchi, S. Sugano, Y. Kohara, A. Fujiyama, A. Anterola, S. Aoki, N.
760 Ashton, W.B. Barbazuk, E. Barker, J.L. Bennetzen, R. Blankenship, S.H. Cho, S.K. Dutcher, M. Estelle,
761 J.A. Fawcett, H. Gundlach, K. Hanada, A. Heyl, K.A. Hicks, J. Hughes, M. Lohr, K. Mayer, A.
762 Melkozernov, T. Murata, D.R. Nelson, B. Pils, M. Prigge, B. Reiss, T. Renner, S. Rombauts, P.J.
763 Rushton, A. Sanderfoot, G. Schween, S.H. Shiu, K. Stueber, F.L. Theodoulou, H. Tu, Y. Van de Peer,
764 P.J. Verrier, E. Waters, A. Wood, L. Yang, D. Cove, A.C. Cuming, M. Hasebe, S. Lucas, B.D. Mishler,
765 R. Reski, I.V. Grigoriev, R.S. Quatrano, and J.L. Boore. 2008. The Physcomitrella genome reveals
766 evolutionary insights into the conquest of land by plants. *Science*. 319:64-69.

Royan, S., B. Gutmann, C.C.d. Francs-Small, S. Honkanen, J. Schmidberger, A. Soet, Y.K. Sun, L.V.P. Sanglard, C.S. Bond, and I. Small. 2021. A synthetic RNA editing factor edits its target site in chloroplasts and bacteria. *Communications Biology*. 4.

Rüdinger, M., L. Fritz-Laylin, M. Polsakiewicz, and V. Knoop. 2011a. Plant-type mitochondrial RNA editing in the protist Naegleria gruberi. *RNA*. 17:2058-2062.

Rüdinger, M., H.T. Funk, S.A. Rensing, U.G. Maier, and V. Knoop. 2009. RNA editing: only eleven sites are present in the Physcomitrella patens mitochondrial transcriptome and a universal nomenclature proposal. *Mol Genet Genomics*. 281:473-481.

Rüdinger, M., M. Polsakiewicz, and V. Knoop. 2008. Organellar RNA editing and plant-specific extensions of pentatricopeptide repeat proteins in jungermanniid but not in marchantiid liverworts. *Mol Biol Evol*. 25:1405-1414.

Rüdinger, M., P. Szovenyi, S.A. Rensing, and V. Knoop. 2011b. Assigning DYW-type PPR proteins to RNA editing sites in the funariid mosses Physcomitrella patens and Funaria hygrometrica. *Plant J*. 67:370-380.

Rüdinger, M., U. Volkmar, H. Lenz, M. Groth-Malonek, and V. Knoop. 2012. Nuclear DYW-type PPR gene families diversify with increasing RNA editing frequencies in liverwort and moss mitochondria. *J Mol Evol*. 74:37-51.

Ruwe, H., B. Gutmann, C. Schmitz-Linneweber, I. Small, and P. Kindgren. 2018. The E-domain of CRR2 participates in sequence-specific recognition of RNA in plastids. *New Phytol*.

Salone, V., M. Rüdinger, M. Polsakiewicz, B. Hoffmann, M. Groth-Malonek, B. Szurek, I. Small, V. Knoop, and C. Lurin. 2007. A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Lett*. 581:4132-4138.

Schaap, P., I. Barrantes, P. Minx, N. Sasaki, R.W. Anderson, M. Bénard, K.K. Biggar, N.E. Buchler, R. Bundschuh, X. Chen, C. Fronick, L. Fulton, G. Golderer, N. Jahn, V. Knoop, L.F. Landweber, C. Maric, D. Miller, A.A. Noegel, R. Peace, G. Pierron, T. Sasaki, M. Schallenberg-Rüdinger, M. Schleicher, R. Singh, T. Spaller, K.B. Storey, T. Suzuki, C. Tomlinson, J.J. Tyson, W.C. Warren, E.R. Werner, G. Werner-Felmayer, R.K. Wilson, T. Winckler, J.M. Gott, G. Glöckner, and W. Marwan. 2015. The Physarum polycephalum Genome Reveals Extensive Use of Prokaryotic Two-Component and Metazoan-Type Tyrosine Kinase Signaling. *Genome Biol Evol*. 8:109-125.

Schallenberg-Rüdinger, M., P. Kindgren, A. Zehrmann, I. Small, and V. Knoop. 2013a. A DYW-protein knockout in Physcomitrella affects two closely spaced mitochondrial editing sites and causes a severe developmental phenotype. *Plant J*. 76:420-432.

Schallenberg-Rüdinger, M., H. Lenz, M. Polsakiewicz, J.M. Gott, and V. Knoop. 2013b. A survey of PPR proteins identifies DYW domains like those of land plant RNA editing factors in diverse eukaryotes. *RNA Biol*. 10:1549-1556.

Schallenberg-Rüdinger, M., B. Oldenkott, M. Hiss, P.L. Trinh, V. Knoop, and S.A. Rensing. 2017. A Single-Target Mitochondrial RNA Editing Factor of Funaria hygrometrica Can Fully Reconstitute RNA Editing at Two Sites in Physcomitrella patens. *Plant Cell Physiol*. 58:496-507.

Small, I., J. Melonek, A.-V. Bohne, J. Nickelsen, and C. Schmitz-Linneweber. 2023. Plant Organellar RNA Maturation. *The Plant Cell*.

Small, I.D., M. Schallenberg-Rüdinger, M. Takenaka, H. Mireau, and O. Ostersetzer-Biran. 2020. Plant organellar RNA editing: what 30 years of research has revealed. *The Plant journal: for cell and moelecular biology*. 101:1040-1056.

Su, D., L. Yang, X. Shi, X. Ma, X. Zhou, S.B. Hedges, and B. Zhong. 2021. Large-Scale Phylogenomic Analyses Reveal the Monophyly of Bryophytes and Neoproterozoic Origin of Land Plants. *Molecular Biology and Evolution*. 38:3332–3344.

Sun, T., S. Bentolila, and M.R. Hanson. 2016. The Unexpected Diversity of Plant Organelle RNA Editosomes. *Trends Plant Sci*. 21:962-973.

Sun, T., X. Shi, G. Friso, K. Van Wijk, S. Bentolila, and M.R. Hanson. 2015. A zinc finger motif-containing protein is essential for chloroplast RNA editing. *PLoS Genet*. 11:e1005028.

Takenaka, M., S. Takenaka, T. Barthel, B. Frink, S. Haag, D. Verbitskiy, B. Oldenkott, M. Schallenberg-Rüdinger, C.G. Feiler, M.S. Weiss, G.J. Palm, and G. Weber. 2021. DYW domain structures imply an unusual regulation principle in plant organellar RNA editing catalysis. *Nature Catalysis*. 4:510-522.

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

Takenaka, M., A. Zehrmann, D. Verbitskiy, M. Kugelmann, B. Hartel, and A. Brennicke. 2012. Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. *Proc Natl Acad Sci U S A*. 109:5104-5109.

Toma-Fukai, S., Y. Sawada, A. Maeda, H. Shimizu, T. Shikanai, M. Takenaka, and T. Shimizu. 2022. Structural insight into the activation of an Arabidopsis organellar C-to-U RNA editing enzyme by active site complementation. *The Plant Cell*.

Uchida, M., S. Ohtani, M. Ichinose, C. Sugita, and M. Sugita. 2011. The PPR-DYW proteins are required for RNA editing of rps14, cox1 and nad5 transcripts in Physcomitrella patens mitochondria. *FEBS Lett*. 585:2367-2371.

Uchiyama, H., M. Ichinose, and M. Sugita. 2018. Chloroplast ribonucleoprotein-like proteins of the moss Physcomitrella patens are not involved in RNA stability and RNA editing. *Photosynthetica*. 56:62-66.

Wagoner, J.A., T. Sun, L. Lin, and M.R. Hanson. 2015. Cytidine deaminase motifs within the DYW domain of two pentatricopeptide repeat-containing proteins are required for site-specific chloroplast RNA editing. *J. Biol. Chem.* 290:2957-2968.

Wang, Y., H. Li, Z.-Q. Huang, B. Ma, Y.-Z. Yang, Z.-H. Xiu, L. Wang, and B.-C. Tan. 2022. Maize PPR-E proteins mediate RNA C-to-U editing in mitochondria by recruiting the trans deaminase PCW1. *The Plant Cell*. 35:529-551.

Yan, J., Y. Yao, S. Hong, Y. Yang, C. Shen, Q. Zhang, D. Zhang, T. Zou, and P. Yin. 2019. Delineation of pentatricopeptide repeat codes for target RNA prediction. *Nucleic Acids Res*. 47:3728-3738.

4. Discussion

4.1. The *E. coli* system to study plant-type RNA editing

4.1.1. Advantages of the *E. coli* system in comparison to plant-based systems

Physcomitrium patens, the model moss, has a relatively simple RNA editing mechanism involving only 13 RNA editing sites fully assigned to 9 DYW-type PPR proteins (Figure 6)(Schallenberg-Rüdinger and Knoop, 2016). Although PEG-mediated protoplast transformation and particle bombardment are efficient methods widely used in moss research, it is mainly applied to the nuclear genome (Kamisugi et al., 2005; Sugita, 2022). Following transformation, the plant must be regenerated for several months before genotyping can occur (Rensing et al., 2020). Additionally, since plant-type RNA editing takes place in organelle transcripts, organelle transformation methods would be helpful for modifying targets. The plastid transformation method, which uses particle bombardment, has been available since 1999 (Cho et al., 1999) and was recently improved with using PEG-mediated transformation (Sugita, 2021). However, the mitochondrial genome is not yet stably transformable in mosses. For three other mosses, in addition to *P. patens*, transformation methods have been established. Gene targeting efficiency with protoplast transformation in *Ceratodon purpureus*, another model moss, is comparable to that in *P. patens* (Thümmler et al., 1992; Trouiller et al., 2007). The well-known CRISPR/Cas9 system has been established in the non-model moss *Scopelophila cataractae* (Nomura et al., 2016), while PEG-mediated protoplast transformation was recently applied to Arctic *Bryum* sp. KMR5045 (Byun et al., 2021). In contrast to mosses, in late-evolved angiosperms like *A. thaliana*, nuclear genome transformation has been well-developed since 1984 (Block et al., 1984; Horsch et al., 1984; Paszkowski et al., 1984) whereas organelle genome editing has only become available more recently, with the first reported methods in 2007 (Verma and Daniell, 2007; Wang et al., 2009).

In a broad view of plant-type RNA editing, different kinds of PPR proteins, helper proteins, and editing mechanisms vary in different clades (see section 1.2.1). Target modifications are not possible or difficult to preform when studying organellar RNA editing mechanisms, and it takes much more time for plant growth and regeneration. The study of plant-type RNA editing is not only limited by plant transformation methods, but a strong phenotype of the plant can lead researchers to struggle due to a lack of plant materials. Among the 9 editing factors in P. patens, 2 shows lethal phenotype (PPR45 and PPR98), 4 shows strong phenotype (PPR65, PPR71, PPR77 and PPR91) (see section 1.4.1). This makes the follow-up complementation studies difficult to continue.

The heterologous *E. coli* system provides a well-developed and easy transformation method with short regeneration periods (see section 1.4.2). In 2013, the P-type PPR protein from maize, PPR10, was successfully expressed in *E. coli* BL21(DE3) and purified. The crystallization structure of PPR10 reveals on how PPR proteins interact with its target (Yin et al., 2013). However, expression of other PPR proteins in *E. coli* often results in high numbers of inclusion bodies, making the work and purification of PPR proteins difficult (Coquille et al., 2014; Okuda and Shikanai, 2012; Yin et al., 2013). To overcome this problem, the N-terminal of the PPR proteins was fused with different tags (e.g. MBP, thioredoxin) to increase protein solubility (Kindgren et al., 2015; Okuda et al., 2014; Schallenberg-Rüdinger et al., 2013a). Later on, capped with Met-Gly-Asn-Ser (MGNS), making the synthetic PPR protein expression in *E. coli* easier (Coquille et al., 2014; Royan et al., 2021). In 2019, the maltose binding protein (MBP)-fused PPR proteins PPR65 and PPR56 were successfully expressed in the *E. coli* system (Figure 7) and edited their supplied targets at a level comparable to the *in planta* situation (Oldenkott et al., 2019). The protein solubility was largely ensured by the maltose binding protein tag (Figure 8). Most notably, the *E. coli* system allows modifications in both, the PPR protein and organellar target transcripts (Oldenkott et al., 2019).

4.1.2. Native RNA editing in *E. coli*

Although it has been proven with the *E. coli* system that a DYW-type PPR protein

alone is sufficient for plant-type RNA editing (Oldenkott et al., 2019), it is important to note that RNA editing events also occur in *E. coli*. In 1996, A-to-I RNA editing of tRNA-Arg at position A_{34} was discovered in *E. coli* extract (Auxilien et al., 1996). The corresponding editing factor, tRNA adenosine deaminase A (tadA), was identified for this site via a *tadA* mutant strain (Wolf et al., 2002). TadA shares a conserved sequence with the human ADAR1, ADAT, and APOBEC1 deaminases, which are responsible for A-to-I editing in human cells (Bar-Yaacov et al., 2018; Gerber and Keller, 2001; Wolf et al., 2002). In deep sequencing of *E. coli*, additional 15 A-to-I editing events on mRNA were reported, with 12 of them occurring in coding sequences and resulting in a change from a tyrosine (TAC) to a cysteine (TGC) codon (Bar-Yaacov et al., 2017). TadA was shown to be responsible for editing these sites as well.

In addition to A-to-I editing, *E. coli* is also capable of C-to-U editing on one site, which occurs at a low frequency of 1-10% and results in a silent change (GGC>GGT, glycine) on *tatA* (Bar-Yaacov et al., 2017; Oldenkott, 2020). However, the *tadA* mutant strain has shown that tadA is not responsible for this site (Bar-Yaacov et al., 2017). While no PPR proteins are encoded in the *E. coli* genome, other cytidine deaminase proteins have been found, which are involved in nucleotide synthesis rather than RNA transcriptional processes (Betts et al., 1994; Danielsen et al., 1992; O'Donovan et al., 1971; Yang et al., 1992). To determine if the C-to-U conversion is due to these deamination factors from *E. coli*, mutant cytidine deaminase and/or *tadA* strains could be used. However, since only one single lowly edited site has been found in the nuclear transcriptome of *E. coli*, it is likely that the influence of the introduced editing factors is minor. Furthermore, no interaction between the native editing factors of *E. coli* and the introduced editing factors can be expected.

4.1.3. Three of nine DYW-type PPR proteins of *P. patens* function in *E. coli*: Possible reasons and improvements

Apart from PPR65 and PPR56, which show comparable editing efficiency on their targets in *E. coli*, PPR78 from *P. patens* is the third PPR protein successfully transferred into the *E. coli* system (Lesch, 2020). However, PPR78 shows different

editing efficiency than in its native situation. In plants, PPR78 edits its targets cox1eU755SL and rps14eU137SL up to >99% and 80%, respectively (Rüdinger et al., 2011a; Schallenberg-Rüdinger et al., 2017; Uchida et al., 2011). In the *E. coli* system, the *rps14* site was edited in 99% of the transcripts, while only 69% was edited in the *cox1* site (Lesch, 2020). It is important to note that the *cox1* site is only edited when the target is elongated to 200 bp; the standard 46 bp target shown in Oldenkott et al. (2019) did not gain editing (Lesch, 2020; Yang, 2019). The length of the target supplied in the *E. coli* system would influence the RNA secondary structure, which could make PPR proteins less accessible to bind RNA molecules properly and/or to the editing site (Lesch, 2020; Yang, 2019; Yang et al., 2023a). Introducing a longer sequence from the native target would largely keep the RNA structure similar to that in the native situation.

It could be argued that PPR56 and PPR65, which have the shortest length and least number of PPR repeats among the 9 DYW-type PPR editing factor proteins (Figure 6), are easier to transfer and function more efficiently in other systems. In contrast, all the other editing factors and their targets from *P. patens* has been tried with the same strategy as PPR56 and PPR65, but none of the others worked (Yang, 2019). However, the fact that native PPR78, which has a comparable length of N-terminal sequence before the first PLS repeats to PPR56 (183 aa vs. 200 aa for PPR78), edits the *cox1* site more efficiently with an elongated N-terminal (up to 89 aa vs. 16 aa for PPR56 and 15 aa for PPR65), suggests that PPR proteins with longer PPR repeats may require a longer N-terminal sequence for proper folding (Lesch, 2020). To address this, the extension of the N-terminal sequence was applied to PPR79, which has the standard 46 bp nad5eU598RC target, but no editing was obtained (Oldenkott, 2020). It may be that the extension of the N-terminal sequence needs to cooperate with an elongated target, as PPR78 did, to achieve proper folding and efficient editing. However, the elongation of the *nad5* target to even 300 bp did not result in any editing by PPR79 with the additional N-terminal sequence (Ramanathan, 2021). Similarly, PPR71 was unable to edit the 200 bp long ccmFCeU122SF target despite having an extended N-terminal sequence (Ramanathan, 2021), indicating that the N-terminal sequence alone may not be

sufficient to address all of the folding issues for the six other PPRs that did not function in the *E. coli* system (Oldenkott, 2020; Yang, 2019). The reason why only PPR56, PPR65, and PPR78 are able to perform editing in *E. coli* remains unclear. However, it is worth noting that expressing PPR proteins in a heterologous system is not an easy task, as they tend to form high inclusion bodies (Coquille et al., 2014; Okuda and Shikanai, 2012; Yan et al., 2017b; Yin et al., 2013).

On the other hand, in the *E. coli* system, additional editing targets were identified for *P. patens* editing factors. One such target is cox3eU290SF, which was pre-edited not only in *P. patens* but also in all available mosses' mitochondrial genomes (Liu et al., 2019; Ritzenhofen, 2021). It has been identified as a true editing site in lycophytes *Isoetes engelmannii* (Grewe et al., 2009), *Selaginella moellendorffii* (Hecht et al., 2011), and the fern *Haplopteris ensiformis* (Zumkeller and Knoop, 2023). When the *nad4* target is cloned upstream of the *cox3* site, PPR56 is able to edit the site up to 97% in the *E. coli* system (Yang et al., 2023a). Intriguingly, PPR56 orthologs could not be found in the organisms mentioned above (Ritzenhofen, 2021). Another target, cox3eU355RW, was edited by the ortholog of PPR65 in *Dicranum scoparium* and the site is also shown as an editing site in other mosses, while remaining pre-edited in *P. patens* (Liu et al., 2019; Ritzenhofen, 2021). After switching the T back to C, PPR65 of *P. patens* is able to edit up to 44% of the targets with the ccmFCeU103PS cloned downstream of it (Ritzenhofen, 2021). The editing site ccmFNeU1465RC predicted by PREPACT was found pre-edited in *P. patens* and is conserved in some of the mosses (Lesch, 2020; Liu et al., 2019). Intriguingly, PPR78 was able to fully edit this site in *E. coli*, even more efficiently than its native sites cox1eU755SL and rps14eU137SL (Lesch, 2020).

In addition to the maltose binding protein, other tags have been shown to improve protein solubility in *E. coli* (Kimple et al., 2013). The MGNS cap was added to synthetic PPR proteins, and the proteins were successfully expressed in *E. coli* BL21(DE3) (Coquille et al., 2014; Royan et al., 2021). In human cells, fusion of the EYFP and HA tags with *P. patens* PPR56 and PPR65 resulted in different editing efficiencies (Lesch et al., 2022). Similarly, the EYFP, HA, and MGNS tags were fused

with both PPR56 and PPR65 in *E. coli*. Interestingly, PPR56 and PPR65 showed a preference for different tags. PPR56 with the MGNS cap showed full editing of the supplied *nad4* target, while PPR65 favored the HA tag and achieved 67% editing on its *ccmFC* target (Willerscheidt, 2022). It is also important to note that PPR56 and PPR65 were able to edit their targets even without any tag, with 92% editing of nad4eU272SL and 18% editing of ccmFCeU103PS, respectively (Willerscheidt, 2022).

In addition to allowing modification of PPR proteins and their targets, the *E. coli* system also enables co-expression of PPR proteins. In a yeast-2-hybrid experiment, PPR65 was found to have the tendency to form weak homodimers and strong heterodimers with PPR71 and PPR79 (Schallenberg-Rüdinger et al., 2013a). Since the E domains in PPR proteins are TPR-like and serve for protein-protein interaction (Cheng et al., 2016; Schallenberg-Rüdinger and Knoop, 2016), angiosperm PPR proteins with different E domains alter specificity of interaction with MORF proteins (helper proteins, see section 1.3.3) (Bayer-Csaszar et al., 2017), it would be interesting to investigate whether DYW-type PPR proteins from the basal land plant *P. patens* can interact with each other. For example, could the un-functional DYW domain of PPR56 be complemented by PPR65's DYW domain? Would such complementation transiently influence the editing of PPR65 itself? Previous experiments showed that cloning both PPR proteins containing MBP tags on one vector was difficult (unpublished data), but this could be further investigated using different tags fused in the future. In addition, editing factors and co-factors cloned on different vector function nicely in *E. coli* system already (Maeda et al., 2022; Royan et al., 2021).

4.1.4. The *E. coli* system to investigate editing factors of other plant species

Besides the editing factors from *P. patens*, attempts were made to transfer PPR proteins from *A. thaliana* into the *E. coli* system. The knockout study assigned the maize DYW-type PPR protein PPR2263 and its *A. thaliana* ortholog mitochondrial editing factor 29 (MEF29) to two mitochondrial sites, nad5eU1550TI and cobeU908PS, with >99% and 80% editing, respectively (Sosso et al., 2012). MEF29

was chosen as the first candidate PPR protein from *A. thaliana* to test in the *E. coli* system for several reasons. It has a comparable number of PPR repeats (17) to PPR56 (14) and PPR65 (15), a comparable number of matches and mismatches to PPR56 and PPR65, serves a comparable number of editing sites (2), and has a complete, conserved DYW domain ending with "DYW" (Figure 12A). However, no editing could be obtained in both the supplied standard 46 bp *nad5* and *cob* targets (Yang, 2019).

Another *A. thaliana* PPR protein, cell wall maintainer 1 (CWM1), also known as MEF41, has been shown in a knockout study to be responsible for editing four mitochondrial sites, including nad5eU598RC, which is the only target assigned to PPR79 in *P. patens* (Figure 12B) (Hu et al., 2016). Phylogenetic analysis has revealed that AtCWM1 is the functional analog of PpPPR79, and *Macadamia integrifolia* (Mi)CWM1, another functional analog, is a DYW-type PPR protein with a complete DYW domain (Oldenkott et al., 2020). When the N-terminal mitochondrial signal peptide was replaced with that of AtCWM1, MiCWM1 was found to be functionally expressed in both *A. thaliana* and *P. patens*, indicating that it does not require additional helper proteins for editing (Oldenkott et al., 2020). However, the same chimera was unable to edit the supplied 46 bp and 200 bp *M. intergrifolia nad5* target in *E. coli* system (Oldenkott, 2020).

The quintuple editing 1 (QED1) factor, also known as Organelle Transcript Processing 81 (OTP81), was found to be involved in editing five chloroplast sites in *A. thaliana* with varying efficiencies (Figure 12C)(Wagoner et al., 2015). QED1 was transferred to the *E. coli* system along with its two highest edited targets, ndhBeU872SL and rpoBeU2432SL, using both standard 46 bp and 200 bp targets. However, none of the constructs were able to perform editing successfully (Noureddine, 2022). Similarly, the plastid editing factor OTP86 was found to be responsible for editing at least rps14eU805SL (Figure 12D)(Hammani et al., 2009). Transferring the complete OTP86 did not result in editing of the supplied *rps14* target (Mizuki Takenaka pers. comm.). In an approach with an chimeric protein composed of PpPPR56 PLS motifs with E domains fused with AtOTP86, both *nad3*

and *nad4* targets were edited to the same level as by the complete PpPPR56 (Noureddine, 2022; Takenaka et al., 2021). Given that in *A. thaliana*, it was shown that mutations or knockouts of helper proteins, for example *rip1*, *rip2*, and *rip9* can affect the editing of *ndhB* and *ropB* sites to varying degrees (Bentolila et al., 2013), it would be interesting to investigate whether co-expressing these helper proteins with the editing factor in *E. coli* could help to maintain editing on the supplied sites.

In addition to native PPR proteins, synthetic PPR proteins have also been tested in the *E. coli* system. One such protein is the DYW-type PPR protein dsn3PLS-DYW, which was designed based on consensus sequences from 38 plant species. This protein was optimized for binding and editing one of the native AtCLB19 target sites, rpoAeU200SF (Royan et al., 2021). In the *E. coli* system, around 12% of transcripts with the *rpoA* site were edited by dsn3PLS-DYW. Interestingly, dsn3PLS-DYW was able to restore 40% editing in the *A. thaliana clb19* knockout plant, while the second editing site of CLB19, clpPeU559HY, remained unedited in both *A. thaliana clb19* KO and *E. coli* (Royan et al., 2021). When co-expressed with the helper proteins MORF2 and MORF9 placed on a separate vector, the editing efficiency of the *rpoA* site was increased to 33% and 37%, respectively (Royan et al., 2021). Another synthetic PPR protein, TRX-9S-DYW, was recently expressed in the *E. coli* system (Bernath-Levin et al., 2022). This S-type PPR protein was designed based on consensus sequences from 37 plant species, plus the P2L2S2E1E2DYW domain from dsn3PLS-DYW, and again targeted the *rpoA* site of CLB19. Around 50% editing could be obtained with TRX-9S-DYW. However, co-expression of MORF2 did not significantly affect the editing efficiency (Bernath-Levin et al., 2022) since MORF proteins did not show interaction with S motifs (figure 5)(Yan et al., 2017b).

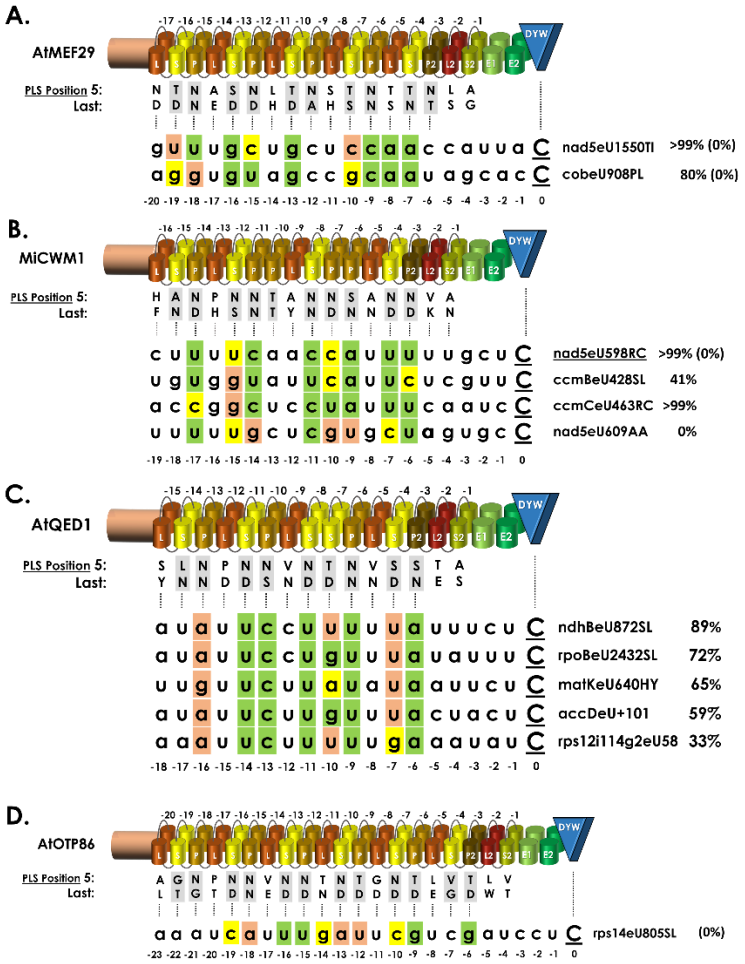


Figure 12. Attempts of editing factors from angiosperms *A. thaliana* and *M. integrifolia* investigated in the *E. coli* system and their targets edited *in planta*. Editing efficiency in bracket shows the editing in *E. coli* system. **A.** The DYW-type PPR protein MEF29 from *A. thaliana* was found to edit nad5eU1550TI and cobeU908PL sites with over 99% and 80% efficiency, respectively (Sosso et al., 2012). **B.** The *M. intergrifolia* DYW-type PPR protein CWM1 is the functional analogo of PPR79 from *P. patens* and can edit the nad5eU598RC site. The editing site shared between MiCWM1 and PpPPR79 nad5eU598RC is underlined. Oldenkott et al. (2020) reported no editing on the nad5eU609AA site in *Macadamia*. However, when using a chimera of AtCWM1 signal peptide fused with MiCWM1, the silent nad5eU609AA site could be edited along with other sites in *A. thaliana*. **C.** *A. thaliana* plastid editing factor QED1 is assigned to five targets, including the rps12i114g2eU58 editing site in the group II intron of the *rps12* gene (Longevialle et al., 2010), nomenclature follows Dombrovska and Qiu (2004). **D.** Plastid editing factor OTP86 from *A. thaliana* is known to be responsible for rps14eU805SL. Based on a phylogenetic study of OTP86 and the *rps14* target, it is believed that OTP86 may edit additional sites (Noureddine, 2022).

4.1.5. Evolution aspect of C-to-U editing and U-to-C editing

While C-to-U RNA editing is widely distributed in plants, the reverse U-to-C editing event has only been observed in hornworts, some lycophytes, and ferns (Figure 2, see section 1.2.1). In 2016, the genome of the hornwort *Anthoceros agrestis* was published (Szövényi et al., 2015), allowing for the identification of potential editing factors that convert U-to-C (Gerke et al., 2020). Through this analysis, a group of 734 DYW-type PPR proteins with conserved sequences in the N-terminal of the DYW domain was identified and named as KAPAxA group (Figure 13A). Two subclades were specified, one with the last three amino acids as "GRP" and the other as "DRH", differing from the "DYW" subclade associated with forward editing factors (Gerke et al., 2020). Notably, the key motifs associated with zinc ion (HxE, CxDCH, and CSC) are conserved in the KAPAxA group as well (Figure 13A). Furthermore, KAPAxA group DYW domains were found in all other clades with U-to-C editing events, but absent in those without reverse editing (Gerke et al., 2020; Gutmann et al., 2020). This strongly suggests that KAPAxA group DYW-type PPR proteins are the editing factors responsible for reverse U-to-C editing. However, none of the plant species with reverse editing has been well established as a model plant, and the amino donor for uridine amination remains puzzling. Therefore, studying reverse editing in plants is challenging.

Despite the challenges in studying reverse U-to-C RNA editing in plants, efforts have been made to establish this process in the *E. coli* system. Using the bioinformatic tool PREPACT, reverse editing sites in *A. agrestis* were assigned to corresponding PPR proteins based on the PPR-RNA binding code. However, despite testing several putative reverse editing factors in the *E. coli* system, none of them were able to perform editing (Bruns, 2022; Oldenkott, 2020; Rüth, 2019). Chimeras of the reverse editing factors with the PLS stretch from forward editing factors *P. patens* PpPPR56 and PpPPR78 were also tested, but failed to achieve editing as well (Bruns, 2022; Oldenkott, 2020) Interestingly, forward editing factor PpPPR56 with included "PAKA" modification, was still able to convert C to U in the nad4 target with 99% efficiency, but could not convert U to C in the nad4eU272SL|c0u target, even with

additional possible amino donors in the medium (Figure 13B)(Bruns, 2022) This finding suggests that the key motif responsible for uridine amination may reside within the DYW domain.

In addition to native PPR proteins, researchers have also investigated reverse U-to-C RNA editing using synthetic PPRs in the *E. coli* system (Ichinose et al., 2022). Based on dsn3PLS-DYW (see section 4.1.4), PLS motifs were modified to be fused to different KAPAxA DYW domains and adapted them to bind to the modified editing site *rpoA* for *A. thaliana* CLB19. In the *E. coli* system, they were able to achieve approximately 50% reverse editing on the modified *rpoA* site. Interestingly, one of the KAPAxA constructs was able to perform both forward and reverse editing in human cells without any additional amino donor added to the medium (Ichinose et al., 2022). The occurrence of both C-to-U and U-to-C editing by a single construct suggests that the difference between the DYW domain of forward editing factors and reverse editing factors could be easily adapted possibly with an extra amino donor in the catalytic pocket created by the gating domain, thus allowing for RNA editing in both directions (Ichinose et al., 2022; Takenaka et al., 2021).

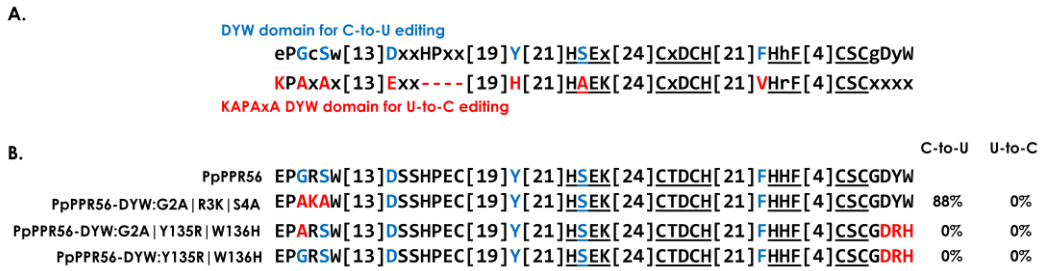


Figure 13. Conserved key motifs found in the forward C-to-U editing DYW domain and the reverse U-to-C editing KAPAxA DYW domain. A. Features of forward editing (blue) is compared with the reverse editing (red). Most conserved (uppercase letter) and conserved (lowercase letter) is labeled. Key catalytic motifs and related amino acids are underlined. **B.** N-terminal and C-terminal amino acids in forward editing *P. patens* factor PPR56 is mutated as KPAxA DYW domain and tested for both directions of editing on the *nad4* target (Bruns, 2022). Figure modified from Knoop (2022).

When we consider the evolution of forward and reverse editing, it is not surprising that C-to-U editing is more widely distributed than U-to-C editing, since the water molecules are generally present in all living environments and act as a co-factor for

zinc ions to catalyze the cytidine deaminase process (Knoop, 2022; Takenaka et al., 2021). Although U-to-C editing would have been advantageous for plant UV defense during the transition from water to land, there is currently no evidence to support the idea that reverse editing was gained prior to forward editing.

4.1.6. Other heterologous systems

Plant knockout studies often result in a strong phenotype, making downstream analysis difficult (see section 4.1.1). The use of a heterologous system can gives the possibility to work with this proteins and targets and assign them. In addition, using a system outside the native environment, which may have the "side-effect" of rescuing knocked out proteins with unknown interactions, provides a transparent picture of the RNA editing mechanism. Although redundant editing factors with the same targets are rare, other proteins like MORFs supporting the editing may make the clear analysis of modified PPR protein difficult. For example, nad5eU609VV still has up to 27% editing in the *cwm1* knockout plant (Oldenkott et al., 2020). The established heterologous *E. coli* system has demonstrated that a single DYW-type PPR protein is sufficient for efficient C-to-U editing and that the DYW domain acts as a cytidine deaminase on RNA transcripts (Oldenkott et al., 2019). Successful expression of PPR65 and observed editing *in vitro* further support this statement (Hayes and Santibanez, 2020).

However, *E. coli*, being a prokaryotic organism, has its obvious drawbacks when expressing eukaryotic PPR proteins. Although plant-type RNA editing was found only in the two prokaryote-like organelles, expressing PPR proteins from a variety of plant backgrounds is not easily achieved (see section 4.1.4). Recently, plant-type RNA editing was also established in human cells (Lesch et al., 2022). Both PPR56 and PPR65 from *P. patens* were found to perform editing on the supplied targets (56% on *nad4* target for PPR56, 52% on *ccmFC* target for PPR65) in a human cell cytosol. That might be surprising, as to date, no single case of C-to-U RNA editing has been reported on plant nuclear transcripts in the cytosol. However, there is one case of a PPR protein that serves its targets in both mitochondria and chloroplasts (Hein et al., 2020; Yap et al., 2015). Since the editing factors in plants are always

organellar-located, it is possible that the protein needs to have a correct folding after cleavage the signal peptide only after passing through the organellar membrane. The achievement of plant-type RNA editing in human cells clearly against this hypothesis (Lesch et al., 2022). In addition, same mutation results in different editing efficiencies, as PPR56 S7 mutant did not gain editing in E. coli system with both native and complemented *nad4* target (Yang et al., 2023a), but could edit around 30% of complemented *nad4* target in human cells (Lesch et al., 2022). Moreover, recent experiment shows that PPR56 and PPR65 could edit delivered targeted in plant cytosol (Per. Comm. Mirjam Thielen).

Another successful heterologous system established for plant-type RNA editing is the unicellular eukaryote, *Saccharomyces cerevisiae* (Hanraths, 2022). As *S. cerevisiae* has a smaller genome and is already established in our lab, the yeast system provides a good alternative as an easy and widely used model organism (Hoffman et al., 2015). PPR56 and PPR78 of *P. patens* have been successfully expressed in *S. cerevisiae* and efficiently edited their targets (PPR56 >99% on *nad4* target (Hanraths, 2022), PPR78 44% on *rps14* target (Per. Comm. Shyam Ramanathan)). Although PPR65 did not work in the yeast system yet, optimization of the just recently established system and testing other factors might bring an improvement in the future. As an advantage of this heterologous system, the yeast nuclear genome can be easily manipulated, providing the opportunity to redirect editing factor PPR proteins to yeast targets. Given the successful establishment of the CRISPR-Cas gene modification tool for yeast mitochondrial targeting (Yoo et al., 2020), natively organellar-located PPR proteins could be even investigated for yeast mitochondrial transcripts in the future.

4.2. Redirecting PPR proteins to novel targets

It has been proposed that the PPR-RNA binding code applies mainly to the P- and S- motifs (Barkan et al., 2012; Cheng et al., 2016), but recent computational studies suggest that the amino acid combination of L-motifs have a slightly preference on RNA recognition as well (Gutmann et al., 2020; Yan et al., 2019). In addition to the fifth and last amino acid, the second amino acid has been proposed to be important

for RNA recognition as well (Kobayashi et al., 2012; Kobayashi et al., 2019; Yagi et al., 2013). This could be a hint for further investigations on the PPR-RNA interaction for a 3-letter code (2nd, 5th and last). However, modifying the L motif to match the suggested combination did not confirm the code significantly (Figure 9), leading to the hypothesis that the positions of PLS repeats may contribute differently to RNA recognition and binding (Yang et al., 2023a).

4.2.1. L motifs might act as spacer and help P and S motif for proper binding to RNA

One of the first questions to address is whether the L motifs in PPR proteins act as RNA recognition elements or spacers. Although most of the target modifications shows that most of the L motifs do not alter the RNA recognition function (Yang et al., 2023a), however, on the protein side, one of the L motifs, L-2, a L2-type repeat in the terminal triplet of PLS repeat, was modified by changing the "VD" combination to "ND", which follows the PPR-RNA binding code for P and S motifs (Yang et al., 2023a). This led to a reduction of editing, although an additional match to the target was generated. This could also be reasoned by the position of the L-motif, as L2 and even S2 motifs usually do not conform to the binding code. In contrast, modifying the L-2 motif from "TM" to "TN" in PPR65, which corresponds to the adenine on *ccmFC* target, did not show a significant difference (69% vs. 70%) (Oldenkott, 2020). However, changing the N-terminal L motif L-14 from "RN" to "NN" to match the uridine on the corresponding position in *ccmFC* target decreased editing efficiency to 36%. Similarly, destroying the "VE" preference for uridine in L-11 motif of PPR65 by changing it to "VA" also reduced editing to 37% (Oldenkott, 2020). One exception is the L-5 TD motif in PPR65, which has a strong influence on RNA editing efficiency (Oldenkott et al., 2019).

On the other hand, mutations in the nucleotides of the *nad4* and *nad3* targets corresponding to the L motif did not completely eliminate RNA editing, except for mutations in nucleotides opposite to L-8 and L-2 from adenine to guanine. Interestingly, the triple mutation on targets corresponding to L-5, L-8, and L-11 on the *nad3* target even restored editing to 76% (Yang et al., 2023a). Influences of

these mutations in the L motif's amino acid combination and on the corresponding nucleotides suggest that the L motifs act more as spacers, maintaining the protein structure for proper interaction of the P and S motifs.

REMSA experiments using mutations on all L-motifs in PPR56 that fit the PPR-RNA binding code for at least one target did not result in significant binding to the *nad3* target and even less binding to the *nad4* target. Furthermore, this mutated PPR56 failed to restore RNA editing in the knockout *ppr56* moss (Matsuda et al., 2020). Similarly, an REMSA experiment on *A. thaliana* CLB19 showed that the L motif did not participate in RNA target interaction (Kindgren et al., 2015). Yan and colleagues generated a synthetic PLS-type PPR protein based on duplicated the consensus sequences of *A. thaliana* PLS-type PPR proteins and co-expressed it with MORF9. The crystallization structure showed that MORF9 interacts with the PLS-type PPR protein by binding to the L motifs, resulting in a significant conformational change in the L motif that might make the PPR protein more accessible for RNA interaction (Yan et al., 2017b).

Since PLS-type PPR proteins often need to serve more than one target, a "kiss and run" mechanism has been hypothesized for RNA editing factors (Knoop, 2020). This mechanism suggests that PPR proteins would need to interact with RNA molecules less tightly than P-type PPR proteins. In addition, L-motifs are most often opposite to the third codon position, which are not as conserved as other positions. The L motif would provide a more flexible structure, with most L motifs involved in structure maintenance to interact with different targets and quickly switch from one transcript to another.

4.2.2. PPR repeats could be distinct in different positions

Among the mutations of the PPR56 PPR repeats on the P and S motifs in the *E. coli* system, most strong editing reductions on the N-terminal and C-terminal repeats could not be restored by complementing the corresponding target. In contrast, most of the motifs in the center of PPR repeats with one amino acid modification could be redirected to a new target (Yang et al., 2023a). Studies on the P-type PPR protein PPR10 from maize pointed out that PPR repeats in the center of the PPR

stretch have decreased nucleotide selectivity (Barkan et al., 2012; Miranda et al., 2017). However, since P-type PPR proteins are not related to RNA editing function, instead, PPR10 is mainly for RNA stabilization, and differences between P-type and PLS-type PPR proteins exist in sequence, structure and RNA binding activities (Barkan and Small, 2014), it is not surprising to see that a PLS-type PPR protein shows more selectivity on its central stretch. A study on the PLS-type PPR protein CLB19 from *A. thaliana* confirmed this statement. With REMSA experiments, it was shown that the P and S motifs corresponding to target positions -10, -9, -7 and -6 have specific interactions with their RNA targets (Kindgren et al., 2015). N-terminal truncation experiment on PPR56 shows that even without the first three L-, S-, P-motif, still 19% of *nad4* target could be edited in *E. coli* system (Yang et al., 2023a). Furthermore, PpPPR79 could edit the targets of CWM1 sites in the complementation study, although the N-terminal of the PPR repeats did not match to the additional targets (Oldenkott et al., 2020).

Within the central stretch of PPR motifs, there are also differences between them. A motif switch experiment on PPR65 shows that duplicating L-8 to replace L-5 could not edit the *ccmFC* target (Figure 10A). Switching complete P-6, L-5 and S-4 with P-9, L-8 and S-7 also abolished editing (Figure 10B). Thinking about the protein's 3D structure, PPR repeats that are distinct in different positions would mainly be because they need to form a proper structure for catching RNA molecules. The central part of the PPR stretch, which is important for RNA recognition, would influence editing more than the N- and C-terminal repeats. This is especially true when C-terminal motifs are linked with TPR-like E domains likely for protein-protein interaction and DYW domain for cytidine deaminase function. Modifications on the P2L2S2 motifs are limited under a certain degree, and often do not resulting in full editing of the targets (Yang et al., 2023a).

Intriguingly, in *P. patens* PPR56, L-8 and S-7 store seven repetitive regions with L-5 and S-4 (Figure 14). This "duplicated" PLS triplet is not found in other editing factors of *P. patens* which makes it the best candidate for testing motif switch. It is important to note that the 5th and last amino acid in these repetitive PLS triples

are different (Figure 4). Mutations in the last amino acid of S-7 and S-4, which switched the amino acid combination, could not restore editing in the S-7 motif but could do so in the S-4 motif in the *E. coli* system. However, double mutations on both S-4 and S-7 motifs switching their amino acid combination could redirect to a new target (Yang et al., 2023a). Another example is replacing the P-9 motif of PPR65 with that of PPR56; in that case up to 42% of *ccmFC* transcript was edited (Figure 11). Given that PPR65 is more strict for target selection and more sensitive to mutations (Oldenkott et al., 2019; Yang, 2019), this switch motif construct surprisingly succeeded. These experiments highlighted that even though complete PPR repeats are conserved with only minor differences, they could be not switched in their positions. Others like P-9 motif from the same position could be switched, which leads to the hypothesis that this might depends on the protein and the position of the motif.

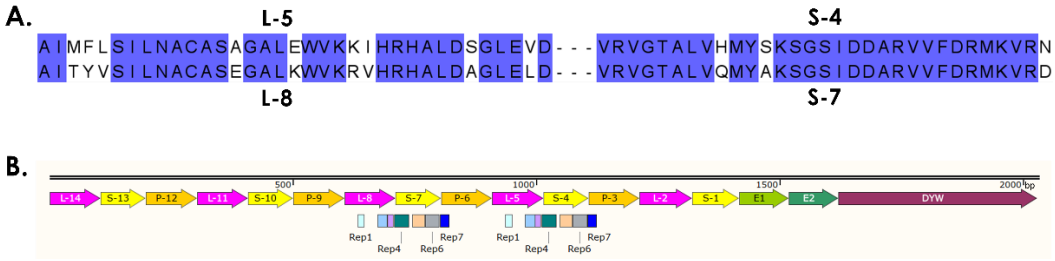


Figure 14. *P. patens* PPR56 stores "duplicated" PPR repeats. A. Alignment of L-5 and S-4 with L-8 and S-7 motif of PPR56. Conserved amino acids are shaded in blue. Figure prepared by Jarview. **B.** PPR56 L-8, S-7 and P-6 motifs are likely duplicated as L-5, S-4 and P-3, as 7 repetitive regions found. Figure prepared by Snapgene (https://www.snapgene.com/).

4.2.3. A wilder target context is influenced by PPR proteins

Aside from the nucleotides opposite to the PLS motif for PPR-RNA recognition, it seems like the sequence further upstream in the RNA transcript also contributes to efficient RNA editing. When up to 16 native nucleotides upstream of the PPR binding region were replaced with a foreign sequence, editing on both targets of PPR56 and PPR65 was reduced in *E. coli* system (Yang, 2019; Yang et al., 2023a). Elongating the sequence from the native target also helps to improve editing (see section 4.1.3). Although one argument could be that the upstream sequence

influences RNA secondary structure, there is no secondary structure prediction tool developed yet that represents biologically meaningful structures for complete transcripts with editing sites. At least general predictions show that there are no obvious strong secondary structures formed in case of upstream binding regions leading to the reduced or even impaired editing (Yang, 2019), and strong secondary structures indeed have influence on the RNA editing (Yang et al., 2023a). This brought about the hypothesis that PPR proteins might need some extra sequences further upstream of the PPR binding site for landing before they bind to the proper site and are ready for editing.

The influence of the upstream sequence could go further. With another editing site cloned next to the target, the editing efficiency could be enhanced (section 3.4) (Ritzenhofen, 2021; Yang et al., 2023a). Notably, the additional target does not even need to have a cytidine on the editing site, although unperfect binding of the additional target does not enhance as much as the perfect binding one (Yang et al., 2023a), suggesting that the bound PPR has an easier accessibility to the next target. An early *in vitro* experiment showed that a tandem atp4eU248PL site from cauliflower with 20-23 nt upstream and 3-5 nt downstream of the editing site could enhance editing efficiency from 2-3% to 50-80%. In addition, increasing the distance between targets would decrease the enhancement of editing (Verbitskiy et al., 2008). With the crystallization structure suggesting active and inactive forms of gating domains in the DYW domain (Takenaka et al., 2021), this could be explained by activated PPR proteins checking surrounding transcripts for other editing sites.

Aside from the sequence upstream of the PPR binding part, the nucleotides between the editing site and the binding part (positions -3, -2 and -1) also contribute to editing efficiency (Oldenkott et al., 2019; Yang et al., 2023a). In general, guanosine is avoided in position -1, as this has been widely observed across different editing sites (Lenz et al., 2018). Switching the DYW domain of PPR56 with another DYW domain in *E. coli* can change the nucleotide preference at the -1 and +1 positions of the editing site (Maeda et al., 2022). The DYW domain most likely

interacts with these positions (Okuda et al., 2014; Takenaka et al., 2021). Although the RNA footprint of CRR2 from *A. thaliana*, a DYW-type PPR protein that does not participate in RNA editing, points to a nucleotide preference at positions -3 and -2 (Ruwe et al., 2018), switching the position -3, -2 and -1 of the editing site to as native target increase the chimera PPR56 with C-terminals of PPR45 editing on nad4 target in *E. coli* system (Yang et al., 2023b). The crystallization structure of the DYW domain suggests that it is rather that the nucleotide might influence conformational changes in the gating domain (Takenaka et al., 2021). This would make it easier for the editing site to be caught by the catalytic center.

4.3. Off-targets in the *E. coli* transcriptome

4.3.1. Dramatic difference in number of off-targets obtained

One advantage of using the *E. coli* system is that its background transcriptome provides an opportunity to study off-target effects. While the native dual-targeted PPR56 obtained 133 off-targets in the *E. coli* transcriptome (Yang et al., 2023a), single-targeted PPR65 only edited 6 off-targets (Oldenkott et al., 2019). This difference in off-target effects may be explained by the flexibility of single and multi-targeted PPR proteins. However, the other dual-targeted PPR78 had only 4 off-targets in the *E. coli* system (Lesch et al unpublished). In the synthetic DYW-type PPR protein tested in *E. coli* system, no off-target could be obtained (Royan et al., 2021). Given that PPR56 also edits 906 off-targets in human cells (Lesch et al., 2022), it makes PPR56 a fascinating candidate for off-target studies.

Despite the large number of off-targets obtained by PPR56, recent studies have shown that the sequences it recognizes still follow a certain pattern. The key motifs for target selection include S-10TD:g, P-9TN:a, S-7TD:g, S-4TN:a, P2-3ND:u, and S2-1ND>u (Yang et al., 2023a). However, the P-6ND motif does not show a significant nucleotide preference, and the native target nad4eU272SL contains a guanine at position -9, which does not follow the PPR-RNA binding code (Figure 4). Although the L motif does not contribute to RNA recognition (as discussed in section 4.2.1), L-11MD and L-8VD are opposite to cytidine and guanine, respectively, while L-5LD

and L2-VD show a preference for pyrimidines. Additionally, a "ucu" preference has been observed at positions -3, -2, and -1. When PPR56 is mutated, the number and identity of off-targets change dramatically. For example, the PPR56|S4TN>TD mutant has about four times more off-targets (449) than the native version (133), while the PPR56|S10TD>TN mutant has only 16 off-targets (Yang et al., 2023a). These changes have also been observed in human cells (Lesch et al., 2022). The S4TN>TD mutation leads to 2265 off-targets, while the S7TD>TN mutation leads to only 323 off-targets. Amino acid changes in the P and S repeats lead to significant changes in nucleotide preference, with S4TN>TD favoring a over g and S10TD>TN favoring g over a (Yang et al., 2023a). These nucleotide preferences have also been confirmed in human cells (Lesch et al., 2022). This change could be due to the structural difference of the PLS stretch which leads to lose or tight binding of RNA molecules.

In the analysis of native PPR56 off-targets, five biological replicates were used with different targets supplied to PPR56, including no target, *nad4*, *nad3*, and *nad4* + *nad3*. Interestingly, out of the 133 off-targets obtained, 117 were found in PPR56 with no target supplied, while 104 were found in PPR56 with *nad4* target, 83 for PPR56 with *nad3* target, and only 52 for PPR56 with tandem *nad4* + *nad3* target. Since the targets were cloned on the same transcript of PPR proteins and overexpressed, it is hypothesized that the correlation between off-targets and supplied target suggests that PPR56 without a high amount of target scans deeper in the *E. coli* transcriptome to perform its job. In the *E. coli* system, the ratio between the protein and the supplied target is 1:1. However, in plants, PPR proteins are typically lowly expressed and multiple targets need to be edited. The high number of off-targets obtained in the *E. coli* system suggests that overexpression of the protein may cause the protein to scan for more targets than intended. This may explain why PPR proteins are expressed at low levels in plants (Fuchs et al., 2020; Lurin et al., 2004).

Additionally, it was found that fusion with different DYW domains surprisingly changes the target selection of PPR56. The chimeric protein PPR56 fused with

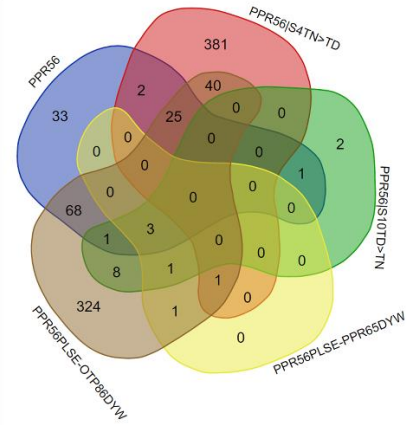
PPR65 DYW domain only had 6 off-targets, while no off-targets could be obtained in the one with PPR45 DYW domain. In contrast, PPR56 fused with *A. thaliana* OTP86 DYW domain showed 472 off-targets in the *E. coli* transcriptome (Yang et al., 2023b). Examining the nucleotide selection, the OTP86 DYW domain did not significantly change the target preference, while the PPR65 DYW domain fished out completely different targets. While target positions from -1 to -10 remained similar to native PPR56, positions -11 to -16 switched to PPR65 preference, particularly in positions -15 and -16, where PPR56 had no clear preference on P-12NN and S-13NS, but PPR65 showed P-12TD:g and S-13NN:u favor. Additionally, the position corresponding to the L motif in positions -11 and -14 switched to PPR65 preference as well. Although only the cytidine deaminase function domain was changed, and no mutations were made on the PLS motifs, which serve for target selection and binding, the results suggest that the DYW domain might influence target selection, particularly on the faraway N-terminal side, and this might due to the structural change of the complete PPR protein. Nevertheless, it should be kept in mind that statistically, only low numbers of off-targets were obtained in both native PPR65 and PPR56-PPR65 chimera proteins.

4.3.2. Off-targets shared between PPR56 native and mutants

It is interesting to observe that some of the off-targets obtained by the native PPR56 and mutants in the *E. coli* system are shared between them (Figure 15). Only 25% (33) of the off-targets edited by the native PPR56 could not be edited by other mutants, while 85% (381) of the off-targets of the more flexible S4 mutant are unique (Figure 15A). Interestingly, 73% of the targets from the native PPR56 are shared with the OTP86 chimera. When comparing with the editing factors from *P. patens*, it is observed that the DYW domain of PPR56 and OTP86 has less conservation to the consensus DYW domain of the *P. patens* editing factors (see section 4.5.1, Figure 18). This indicates that the DYW of the mitochondrial factor PPR56 from early evolved moss is highly compatible with the later evolved angiosperm DYW domain of plastid factor OTP86. As for the PPR56 chimera with PPR65, although a nucleotide preference shift is observed in the N-terminal side,

none of these six targets can be edited by the native PPR65 (Oldenkott et al., 2019).

A.



B.

Name	POS	Ed_S4	Ed_native	Change
truBeU199RW	3181078	44.40%	3.32%	41.08%
sufABeU902PL	1707900	27.56%	3.30%	24.26%
plaPeU965PL	1982342	41.48%	20.00%	21.48%
yqjFeU287PL	3117147	28.31%	7.34%	20.97%
kbleU554SL	3661199	28.46%	11.73%	16.74%
purEeU10RC+2	522178	42.02%	28.43%	13.59%
rarAeU407TI	943491	24.15%	11.70%	12.44%
gyrBeU1393RW	3767523	15.86%	4.31%	11.55%
ccmFeU118Q*	2188437	58.11%	46.88%	11.23%
ftsleU668SF	94885	31.52%	20.28%	11.23%
cydAeU979LL	731097	8.87%	1.63%	7.23%
arnAeU242PL	2258454	23.93%	17.95%	5.98%
ilvGeU545PL	3841344	26.20%	20.53%	5.67%
yqXeU940Q++2	1371425	6.16%	1.16%	4.99%
fabBeU1037AV	2327822	17.30%	13.80%	3.51%
ddlAeU311PL	365238	23.02%	20.93%	2.09%
speEeU470SF	128799	33.03%	31.36%	1.67%
argCeU686TI	4063987	20.63%	20.54%	0.08%
entBeU844RC	589024	21.85%	23.14%	-1.29%
aspAeU1373SF	4274041	17.18%	18.62%	-1.45%
yrbGeU170PL	3208944	51.68%	53.15%	-1.47%
nupCeU1031PL	2391958	13.02%	14.59%	-1.56%
fdhEeU403PS	3987413	16.30%	18.11%	-1.81%
gntKeU126HH	3440616	4.66%	7.92%	-3.25%
folDeU-5	526833	78.18%	85.92%	-7.74%
yegHeU419PL	2037327	38.12%	47.85%	-9.73%
cydCeU980PL	933272	50.00%	62.70%	-12.70%

C.

Name	POS	Ed_S10	Ed_native	Change
argTeU-14	2868542	27.25%	11.67%	15.58%
trpBeU1157TI	1302851	17.12%	8.70%	8.42%
glpXeU842SF	4022660	13.65%	11.58%	2.07%
aspCeU1046SF	989754	5.90%	4.33%	1.57%
recJeU425SF	2315058	1.67%	1.45%	0.23%

Figure 15. Off-targets could be shared between native PPR56 and mutants in *E. coli* system. A. Venn graph of 5 sets of off-targets obtained in *E. coli* transcripts. Figure generated by http://bioinformatics.psb.ugent.be/cgi-bin/liste/Venn/calculate_venn.htpl. **B.** Off-targets shared between native PPR56 and S4 TN>TD mutant. Amino acid mutation indicated in red and bold. Off-targets listed with name, location in BL21(DE3) genome, and editing in S4 mutant (Ed_S4) and in native PPR56 (Ed_native). Off-targets with a guanine in position 7 shade in dark green under "Ed_S4". Editing increased shade in green under "Change". **C.** Off-targets shared between native PPR56 and S10 TD>TN mutant.

Off-targets shared between different proteins are edited with varying efficiencies. Among the 27 off-targets shared between native PPR56 and the S4 mutant, 18 of them show increased editing in the S4 mutant. Interestingly, only half of these targets have a guanine in position -7 that fits the TN>TD switch in the binding code (Figure 15B). However, this is not observed in the S10 mutant (Figure 15C). All 5 off-targets shared between native PPR56 and the S10 mutant show increased editing with the S10 mutant, and all of them have an adenine in position -13 that fits the TD>TN switch. It is important to note that the change in editing efficiency is not solely due to the binding code switch, but also due to the switch in DYW domain (Yang et al., 2023b). A similar editing efficiency shift could be obtained in the off-targets in PPR56 native and the S4, S7 mutants human cells as well (Lesch et al., 2022).

4.3.3. Editing site positions can be shifted in off-targets

Generally, PPR proteins bind to the 4th nucleotide upstream of the cytidine to be edited. However, there are cases where the editing site can be shifted. PPR77 from *P. patens*, for example, most likely binds to both cox2eU370RW and cox3eU733RW sites from the 5th nucleotide upstream of the editing site (Schallenberg-Rüdinger et al., 2013a). This phenomenon can also be observed in some of the off-targets obtained in PPR56 and its mutants in the *E. coli* transcriptome (Figure 16A) and in human cell off-targets (Lesch et al., 2022). These editing shifts occur within the range of +2 and -2 of the usual editing position (see Appendix Supplementary Table 3).

In addition to the shift of the editing site, some cytidines located adjacent to the target cytidine are co-edited as well. However, the editing efficiency of the co-edited targets is usually lower than that of the main candidate sites. This phenomenon is especially observed in off-targets obtained with native PPR56, in the S4 mutant, and for the OTP86 chimera (Figure 16A). Although the obtained off-target set could be influenced statistically by the number of off-targets, it can be seen that the OTP86 chimera holds more co-edited sites than the one with a native PPR56 DYW domain. Interestingly, there is one example in the OTP86 chimera

where two editing sites are located 9 nucleotides away, which is outside the usual range of ±2 nucleotides for an editing shift (Figure 16B). The downstream target binding sequence poorly matches with the PPR56 PLS stretch. This downstream site could be a result of PPR56 scanning the transcript for alternative editing candidates.

A.

Construct	Off-targets	Editing shift	Shift %	Co-edit
PPR56	133	14	11%	1
PPR56|S4TN>TD	449	67	15%	6
PPR56|S10TD>TN	16	1	6%	-
PPR56PLSE-PPR65DYW	6	1	17%	-
PPR56PLSE-OTP86DYW	472	114	24%	15

B.

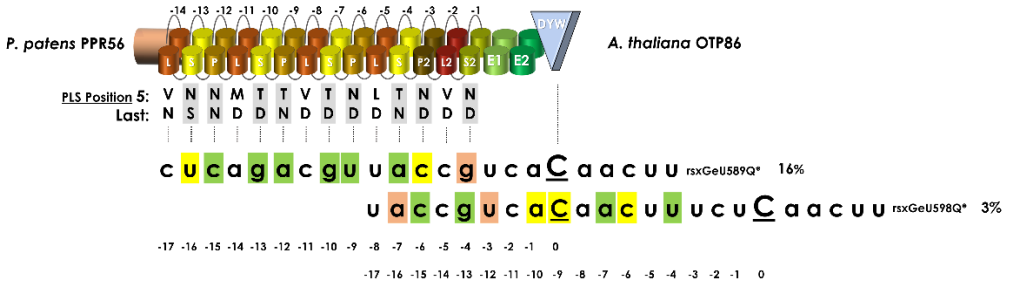


Figure 16. Editing sites in off-targets could be shifted or co-edited by native PPR56 and its mutants in *E. coli* transcriptome. A. Number of editing sites shifted in the range of ± 2 and co-edited was counted in different off-target sites of native PPR56 and its mutants. **B.** An extreme example of two targets located 9 nucleotides away in PPR56-OTP86 chimera. Nucleotides shaded based on binding code in Figure 3.

4.4. Plant-type RNA editing outside of land plants

4.4.1. Pyrimidine editing sites are found outside of land plants

In addition to C-to-U changes in the organellar transcripts of land plants, nucleotide substitutions at the RNA level have rarely been observed in mitochondrial transcripts of non-embryophyte organisms. Although several DYW-type PPR proteins have been discovered in these organisms (see section 1.3.4), editing sites could not be identified due to the lack of organellar transcriptome data (Gutmann et al., 2020; Schallenberg-Rüdinger et al., 2013b). However, editing sites have been detected in protists such as *Acrasis kona* and *Naegleria gruber* (see section 1.3.4). *Physarum polycephalum* is an acellular slime mold that belongs to the Amoebozoa

protist group (Oettmeier et al., 2020). Through deep RNA sequencing, four C-to-U conversions in *cox1* transcripts were identified, along with one U-to-G and one C-to-G substitution in its mitochondrial transcriptome (Bundschuh et al., 2011). These, along with the DYW domain PPR proteins found in *P. polycephalum*, may represent plant-type RNA editing events outside of land plants (Schallenberg-Rüdinger et al., 2013b). Using the powerful bioinformatic tool PREPACT, seven weak C-to-U editing candidates were identified in *Acanthamoeba castellanii* mitochondrial DNA, while 17 strong candidates were found in the *Malawimonas jakobiformis* mitogenome to restore conserved codons (Schallenberg-Rüdinger et al., 2013b). Since DYW-type proteins are found in the genomes of *A. castellanii* and *M. jakobiformis*, plant-type RNA editing may occur in these protists as well.

4.4.2. Candidate RNA editing factors in *N. gruberi*

In the heterolobosean protist *Naegleria gruberi*, a total of 10 DYW-type PPR proteins have been identified in its nuclear genome (Knoop and Rüdinger, 2010). NgPPR45 and NgPPR51, show a clear mitochondrial signal peptide and are therefore the best candidates for serving two mitochondrial editing sites, cox1eU1120HY and cox3eU780RW (Yang et al., 2023c). The functions of other DYW-type PPR proteins, which have either a short PLS stretch or some degenerated motifs within the binding domain, are still unknown. Unlike angiosperms, which have numerous helper proteins to aid in editing function (Sun et al., 2016), protists may not have homologs of co-factors. P-type PPRs, which are involved in RNA stability, RNA splicing, and other post-transcriptional processes, could be an alternative function for these DYW-type PPR proteins, as the example of PpPPR43 for intron splicing. On the other hand, these proteins may be by-products when transferring the true editing factors from plants and remain unspecified and undergo degeneration in protists nuclear genome, similar to other horizontal gene transfer events in mitochondrial genome, as seen in the extreme example of *Amborella* (Rice et al., 2013).

Upon closer examination of the DYW domains of NgPPR45 and NgPPR51, it was found that the central cytidine catalytic domain HSE was altered to HAE, which is

commonly seen in reverse U-to-C editing factors (see section 4.1.5 and Figure 13). Furthermore, NgPPR45 features a "GYW" motif at the end of its DYW domain, which is a modification towards a reverse editing factor. Conversely, NgPPR51 has a well-conserved HSE and DNW, making it a potential functional forward editing factor. However, the WGAL motif at the beginning of its E1 domain is not conserved. When considering the binding code, NgPPR45 appears more promising than NgPPR51, as it has 7 fitting PPR nucleotide pairs in cox1eU1120HY and 5 for cox3eU787RW (Yang et al., 2023c). Nevertheless, additional research is required to confirm this finding.

4.4.3. Transferring plant-type RNA editing factor of a protist back to plant

Although transient transformation protocols, such as electroporation, have been applied to some protists (Faktorová et al., 2020), stable transformation for knockout or knockdown study has not been easily achieved to date. Genome assembly of the protist *N. gruberi* presents another challenge, as the highly repetitive sequence may have led to the incorrect assembly of the DYW-type PPR protein NgPPR45 (Yang et al., 2023c). To overcome these challenges, a heterologous *E. coli* system was used to express PPR proteins from *N. gruberi*. However, the highly repetitive sequences of NgPPR45 makes the molecular cloning difficult to perform. The chimera of C-terminal domains of NgPPR45 with PpPPR78 was able to perform editing on the supplied targets (Ramanathan, 2021). Further investigation is needed to understand the reasons for the unsuccessful editing and to develop new strategies for stable transformation in protists.

Nonetheless, obtaining DYW-type PPR proteins in *N. gruberi* is likely a result of horizontal gene transfer from an early plant ancestor (Knoop and Rüdinger, 2010; Rüdinger et al., 2011b), and transferring them back to plants could be an alternative for studying their functionality. Since neither cox1eU1120HY nor cox3eU787RW from *N. gruberi* are present in *P. patens*, chimeric proteins with the PLS stretch of *P. patens* editing factors are used for target binding. A chimera of PpPPR78 with complete C-terminal extensions (E domains plus DYW domain) of NgPPR45 was able to edit the cox1eU755SL site up to 82% in a *ppr78* knockout plant, providing the first evidence that the C-terminal extensions carried by a protist can perform

cytidine deamination function (Yang et al., 2023c). It is also noteworthy that PPR proteins from moss are compatible with the C-terminal domains of a protist PPR protein, as changing the DYW domain with a PPR protein from the same original plant often results in no or low editing restoration in *P. patens* (Ichinose and Sugita, 2018). This chimera, with complete C-terminal domains, is the only one restoring editing in a *P. patens* knockout plant, leading to the hypothesis that the TPR-like E domains may need to be compatible with the DYW domain for a functional catalytic reaction.

Real-time quantitative PCR analysis revealed that the *cox1* target editing efficiency of PpPPR78 with NgPPR45 C-terminal extensions was correlated with the expression level of the chimera protein in all five transformed lines (Yang et al., 2023c). Other studies on PpPPR78-PpPPR79 chimera, PpPPR79 and QED1 with have likewise shown that there is a correlation of editing efficiencies and expression of the editing factor (Loiacono et al., 2022; Oldenkott et al., 2020; Schallenberg-Rüdinger et al., 2017).

4.5. Using PPR protein as a molecular tool

4.5.1. Expressing chimera PPR proteins in the *E. coli* system

Expressing the complete native PPR protein is not easy due to the difficulties in proper expression and solubility. Creating chimeras of different motifs or building blocks of PPR proteins could provide an alternative way to understand the function of motifs and the combination and compatibility of different domains. This is especially easy to achieve with a heterologous system.

Switching motifs in the PLS stretch

The P-9 motif of PPR65 was successfully switched, resulting in an 18% increase in editing on a single *ccmFC* target. When the tandem *ccmFC* target was used, up to 42% of transcripts were converted from cytidine to uridine (see section 3.4.2, Figure 11). Although gaining editing with mutated PPR65 is not easy due to the strict target selection for only one target, it appears that the P-9 motif from PPR65

and PPR56 does not have a significant influence on the overall RNA mechanism. Among the 9 DYW-type PPR proteins that act as RNA editing factors in *P. patens*, 9 P-type PPR repeats with a "TN" combination were found and aligned (Figure 17). It can be observed that the main conserved amino acids are found in the start and end parts of the repeat. The second position in the P-9 motif of PPR56 contains an isoleucine, while most others have a valine. Since the second amino acid is also proposed to be part of the 3-letter PPR-RNA binding code (Kobayashi et al., 2012; Kobayashi et al., 2019; Yagi et al., 2013), switching a “VTN” motif such as PPR91 P-14 with the P-9 could be further investigated in the future.

Figure 17. Alignment of P motifs with “TN” combination from editing factors of *P. patens*. Conserved amino acids are shaded in blue. Bottom line shows the conserved amino acid from all P motifs as in Cheng et al. (2016). Capital letters indicate most conserved amino acid among P motifs. Alpha helix regions are shaded. Amino acid with same conservation in P “TN” motif of *P. patens* colored in red. The figure was prepared with Jarview 2.11.2.6.

However, switching the motifs of native PPR proteins is not always successful. Technically, the usual method for linking multiple pieces together is through overlap extension PCR (Ho et al., 1989). When the sequences are highly similar, as in the case of PPR56 (Figure 14), achieving this can be difficult. The modern technique of golden gate cloning allows for efficient DNA assembly using type IIs restriction enzymes, which cut outside of the recognized site (Engler et al., 2009). With only a four-nucleotide sticky end in each fragment, this could be an alternative way to construct chimeras in PPR stretches.

Switching PPR motifs between orthologs with the same fifth and last amino acid like QED1 from *A. thaliana* and cacao was tested in an *A. thaliana qed1* knockout plant. Different chimeras with variable combination of PPR stretch was tested with QED1 from the two plants. However, these combinations could only partially restore editing in some QED1 targets (Loiacono et al., 2022). Moreover, work with synthetic PPR repeats derived from consensus sequences of native PPR repeats also seem to face this problem. For instance, a synthetic PLS-type PPR protein targeting the *rpoA* site of *A. thaliana* CLB19 was only able to restore 40% of the editing in a *clb19* knockout plant, compared to around 80% editing in a wildtype plant. However, when co-expressed with MORF2 in *E. coli*, full editing was achieved at the *rpoA* site (Royan et al., 2021). This suggests that PPR repeats may not always be compatible with each other for building a fully functional editing factor.

Chimeras PPR proteins to investigate C-terminal extensions

The C-terminal extensions include the E domains and the DYW domains, which play a crucial role in the cytidine deamination catalytic function. Through improved understanding of different motifs and the use of powerful bioinformatic tools, the definition of C-terminal extensions has been revised (Figure 18A) (Cheng et al., 2016; Lurin et al., 2004). Interestingly, based on the crystal structure (Takenaka et al., 2021), the previous definition of the DYW domain (DYW*) starts in the middle of the gating domain, which requires a conformational change before the catalytic center can bind to the editing site.

PPR56 and PPR65 have been reconstructed on the C-terminal extensions based on two different definitions. However, PPR65 PLS motifs fused with different versions of PPR56 C-terminal extensions could not achieve any editing in the *E. coli* system (Yang et al., 2023b). Nonetheless, PPR56 with PPR65 DYW domain was able to edit 78% of the *nad4* target. Interestingly, replacing the α1 motif (Figure 18A) of PPR56 with PPR65 could edit tandem targets of *nad4* and *nad3* up to 53%. When fused with the plastid DYW* domain of PPR45, PPR56 edited only 9% of the *nad4* target, but 28% when the target position was changed from -3 to +5, as in the *rps14* target,

the native target assigned to PPR45. Replacing the DYW domain with the *A. thaliana* plastid editing factor OTP86 did not affect the editing of PPR56 on both of its targets (Noureddine, 2022; Takenaka et al., 2021; Yang et al., 2023b). Furthermore, replacing the gating domain of PPR56 with OTP86 edited both *nad4* and *nad3* to approximately 65% in tandem enhancement (Yang et al., 2023b). Note that OTP86 stores an extra glutamate within the α1 motif (Figure 18B). When fused with DYW domains of other DYW-type PPR proteins from *A. thaliana*, PPR56 was able to edit only the *nad4* target in rare cases (Maeda et al., 2022). Co-factors that are recruited to complete DYW domains for editing events in *A. thaliana* were tested with PPR56 chimera as well. Only DYW1, which serves specifically for CRR4 PPR protein, could perform editing up to 90%, when fused from the beginning of the α1 motif of the gating domain. Interestingly, breaking the α1 motif as DYW* reduced editing to only 20% (Maeda et al., 2022). The importance of the complete α1 motif in forming a functional DYW domain in chimera proteins is highlighted. In general, it appears that PPR56 is more tolerant of different types of DYW domains, not only from the same origin *P. patens*, but also from angiosperm, where the editing mechanism is more complex.

In contrast, attempts to engineer chimeric proteins by fusing the PLS stretch of PPR78 with C-terminal extensions of PPR79, with or without E domains, did not result in successful editing of the *rps14* and *cox1* targets in *E. coli* system. When the DYW* domain of PPR79 was fused with PPR78, only the *rps14* target could be edited to a limited extent of approximately 12% (Ramanathan, 2021). However, when PPR78 was fused with the E1E2DYW domain of PPR79, full restoration of editing was observed on the cox1eU755SL site, with partial restoration on the rps14eU137SL site in a *ppr78* knockout *P. patens* plant. Notably, fusion of PPR78 with the DYW* domain of PPR79 fully restored editing on both the *rps14* and *cox1* sites (Schallenberg-Rüdinger et al., 2017). Most of the chimeras seems to work in plant system better than in the heterologous system.

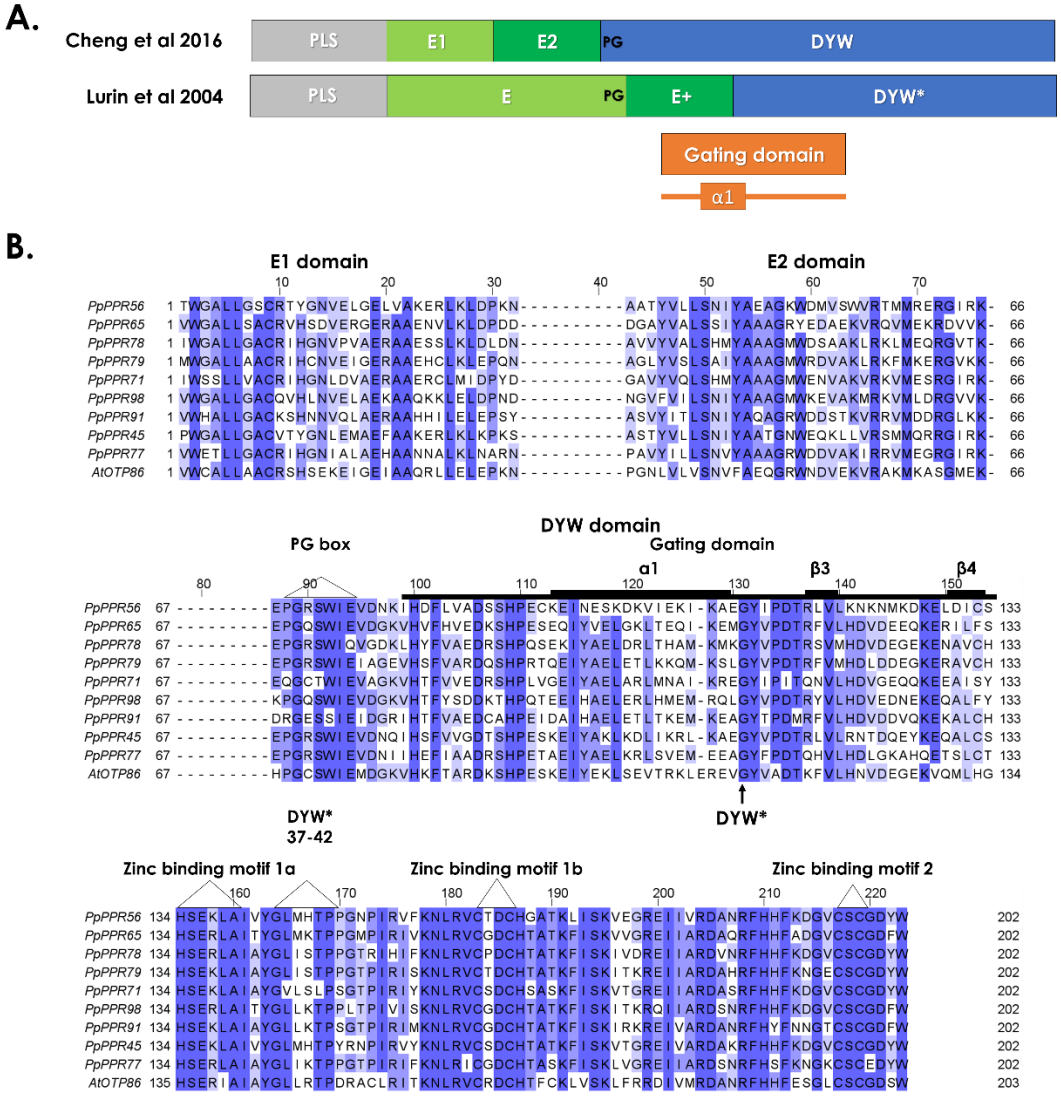


Figure 18. The C-terminal extensions of DYW-type PPR proteins. A. Different definition of C-terminal extensions. C-terminal extensions were differently defined in (Lurin et al., 2004) and in (Cheng et al., 2016). The orange block shows the position of the gating domain and the α1 motif. **B.** Alignment of C-terminal domains from *P. patens* editing factors and *A. thaliana* OTP86. C-terminal domains are labeled as in Cheng et al. (2016). Conserved amino acids are highlighted as in Figure 17. Key motifs are labeled as in Figure 3. Gating domain contains one α helix (α1) and two β sheets (β3 and β4).

Upon examination of the alignment in Figure 18B, it becomes apparent that the α1 motif in PpPPR56 is less conserved compared to other motifs, and it seems like it is the PpPPR56 has the unique sequence from others (Gutmann et al., 2020).

Surprisingly, experimental results have demonstrated that this unique sequence can be substituted with others, as seen in the cases of PpPPR65 and AtOTP86 (Yang et al., 2023b). In 2018, Ichinose and his colleagues identified a crucial region within the DYW* domain, spanning amino acids 37 to 42, that plays a vital role in C-terminal compatibility (Ichinose and Sugita, 2018). However, this finding fails to explain the contrasting results obtained with the chimera of PPR56 and PPR65, as they share similar amino acid sequences with only one difference in position 167 (Figure 18B). Notably, the substitution of this different histidine with lysine (PPR56|DYW:H80K) did not significantly affect the editing of both *nad4* and *nad3* targets (Yang et al., 2023a).

4.5.2. PPR56 as an outstanding editing factor

Among the three complete editing factors successfully transferred into the *E. coli* system (see section 4.1.3), PPR56 stands out due to its high flexibility in mutations, off-targets and chimera experiments. As one of the two PPR proteins that show comparable editing efficiencies in both plants and *E. coli*, PPR56 has two native targets with different editing efficiencies. The *nad4* target is almost fully edited and shows high tolerance to different point mutations, whereas the moderately edited *nad3* target is more sensitive to changes. However, these two targets are not conserved in most of the available moss mitogenomes (except Pottiaceae) and already exist as pre-edited sites with thymidine on the DNA level (Ritzenhofen, 2021). In most moss species, with available genome or transcriptome data, PPR56 ortholog could not be identified, what fits to the lack of the two assigned sites, but the PLS stretch with a partial DYW domain could be amplified from *Pottia truncata* total DNA. PPR56 exhibits a similar behavior in both heterologous systems, *E. coli* and human cells, tolerating most of the mutations, flexibly fusing with different tags, and actively editing its *nad4* target with different C-terminal chimeras from at least some proteins. It also shows a high amount of off-targets in the *E. coli* transcriptome (133) (Yang et al., 2023a) and in human cells (more than 900)(Lesch et al., 2022). Compared to the off-targets of single-targeted PPR65 (6)(Oldenkott et al., 2019) and another dual-targeted PPR78 (4) (Lesch et al. unpublished) in the *E. coli*

transcriptome, PPR56 displays extremely high flexibility in target selection. In addition to its flexibility, mutations on *nad4* target position -3 to -1 region could be edited by PPR56, which results in the option to create start or stop codons via editing (Yang et al., 2023a). PPR56 has been shown to be able to efficiently edit point mutations that create a start codon (AUG) or stop codons (UAA, UAG, UGA) on its targets in the *E. coli* system. This suggests that PPR56 could potentially be used as a molecular tool to switch on and off a transcript (Yang et al., 2023a).

As PPR56 is found in a sister clade of PPR45 (Ritzenhofen, 2021; Rüdinger et al., 2011a), a plastid editing factor, and has duplicated PLS motifs (Figure 14), it is possible that PPR56 was duplicated from PPR45, specifically to serve the editing of nad4eU272SL and nad3eU230SL in *Physcomitrium patens* and Pottiaceae. Note that the nad4eU272SL and nad3eU230SL sites stores a cytidine on DNA level only in 6 of 9 species investigated within the *Physcomitrium-Physcomitrella* species complex (Beike et al., 2014). Furthermore, since the potential target cox3eU290SF from *P. patens* obtained for PPR56 in the *E. coli* system (see section 4.1.3) has been shown to be a true editing site in lycophytes *Isoetes engelmannii* (Grewe et al., 2009), *Selaginella moellendorffii* (Hecht et al., 2011), and the fern *Haplopteris ensiformis* (Zumkeller and Knoop, 2023), it is possible that functional analogs of PPR56 could be found in those species.

4.5.3. PPR protein as a promising molecular tool

Genetic modification has rapidly advanced in recent years with the development of DNA-targeting systems such as Zinc-finger nucleases (ZFN), transcription activator-like effector nucleases (TALEN), and CRISPR-Cas endonucleases, enabling the manipulation of genomes and control of gene expression (Miller et al., 2010; Porteus and Carroll, 2005; Wang et al., 2016). Programmable DNA construction has become particularly popular in the biomedical field, especially among living cells (Pei and Lu, 2019). However, these molecular tools often mediate DNA repair mechanisms, either homology-directed recombination (HDR) or non-homologous end joining (NHEJ), which can increase the risk of inducing unexpected mutations. Additionally, manipulating the DNA level, which stores heritable information,

increases the possibility of transferring incorrect information to the next generation.

RNA manipulations, especially during post-transcriptional processes, have the advantage of not affecting heritable information, while providing more possibilities by keeping the original information intact. There are successful approaches with PUF proteins as RNA manipulation tools. The PUF protein is derived from the truncated RNA-binding domain of human Pumilio homology 1 (PUM1) (Cheong and Hall, 2006; Wang et al., 2002). After engineering, the PUF protein can recognize RNA targets based on a recognition code and target cytosine (Zhao et al., 2018). By fusing with proteins such as endonucleases, splicing factors, and splicing enhancers, the PUF protein can achieve various modifications (Pei and Lu, 2019). Up to date, most of these modifications are based on biomedical studies conducted in human cells.

In addition to nuclear genome manipulation, editing in organelles can help to address organellar control events. Two new methods for plant organelle genome editing have recently been developed: the DddA-derived cytosine base editor (DdCBE) (Kang et al., 2021) and the mitochondria-targeting transcription activator-like effector cytidine deaminase (mitoTALECD) (Nakazato et al., 2022). The mitoTALECD method has been tested on two sites assigned to the editing factor OTP87, specifically atp1eU1178SL and nad7eU24LL. Using mitoTALECD, otp87 mutant plants could be rescued from a strong phenotype, with full editing of atp1eU1178SL achieved (Nakazato et al., 2022). However, these TALEN-based tools are targeting organellar DNA rather than transcripts.

PPR proteins, especially PLS-type PPR proteins, are among the best molecular tools for targeting the organellar transcriptome. They have a target-selecting PLS stretch that efficiently recognizes and binds to sequences upstream of the editing site in a one motif-one nucleotide manner, and a DYW domain that serves as a cytidine deaminase (see section 1.3). Compared to the recently established RNA-based CRISPR-Cas system (Huang et al., 2020; Komor et al., 2016; Nishida et al., 2016), only one protein is required to complete the catalytic event, making it easier to transport the protein through organellar membranes. Synthetic PLS-type PPR

protein dsn3PLS-DYW was able to successfully target only the rpoAeU200SF editing site assigned to *A. thaliana* CLB19 and not the other site clpPeU559HY in the *clb19* knockout plant (Royan et al., 2021). Notably, no off-targets were detected by this PLS-type PPR protein in the *E. coli* system, and 10 off-targets obtained in *A. thaliana* , with one being edited in 44% of transcripts, while the rest of the nine off-targets were edited below 1.5%.

Although PPR proteins were originally thought to exclusively target organelles, recent study has shown that they are also capable of editing supplied targets in the cytosol of human cells (Lesch et al., 2022) and even in *P. patens* cytosol (Per. Comm. Mirjam Thielen). This expands the potential applications of PPR proteins beyond organellar RNA editing, to also include nucleo-cytosolic editing. The numerous off-targets observed with PPR56 could be greatly reduced by a single amino acid change (Yang et al., 2023a), which suggests the possibility of targeting the transcript of interest more specifically. The promising potential of PPR proteins as molecular tools has been under development for commercial application with the company "Editforce" since 2015 (https://www.editforce.co.jp/en/).

5. Summary

Plant-type C-to-U RNA editing is a post-transcriptional process that converts cytidines to uridines in mitochondrial and chloroplast transcripts. This process is mediated by pentatricopeptide repeat (PPR) proteins, which recognize specific targets via their PLS repeats and perform cytidine deamination via their C-terminal extension, the DYW domain. In the model moss *Physcomitrium patens*, all 13 editing sites are fully assigned to 9 DYW-type PPR proteins, making it an ideal model for studying RNA editing mechanisms. Additionally, plant-type RNA editing has been successfully established in other heterologous systems, such as *Escherichia coli*. The faithful editing efficiency in comparison to plants makes it a powerful system for further investigating RNA editing mechanisms.

The mitochondrial editing factor PPR56 from *P. patens* efficiently edits its two targets, nad3eU230SL and nad4eU272SL, in both plant and *E. coli* systems. However, when compared to other PPR proteins such as PPR65 and PPR78, PPR56 exhibits over 100 off-targets in the *E. coli* transcriptome. Detailed mechanistic studies in the bacterial system have shown that: (i) single amino acid modifications on the PLS motif can redirect PPR proteins to new targets, (ii) the RNA editing activity of PPR proteins is influenced by the target context, and (iii) the editing efficiency is enhanced by tandem targets. A new candidate site, cox3eU290SF, was found to be influenced by PPR56, which suggests a scanning mechanism of PPR proteins along transcripts.

Although C-to-U RNA editing is mainly restricted to land plants, one of the exceptions is the heterolobosean protist *Naegleria gruberi*, which stores 10 DYW-type PPR proteins in its nuclear genome and has two editing sites in its mitochondrial transcriptome. In this study, NgPPR45 was rearranged, resulting in a DYW-type PPR protein with a mitochondrial signal peptide that nicely fits the two editing sites according to the PPR-RNA binding code, making it an ideal candidate. Since knockout studies could not be performed in *N. gruberi*, different chimeras of

PpPPR78 with NgPPR45 were tested in the *ppr78* knockout *P. patens*. The chimera with the PLS stretch of PpPPR78 and the complete C-terminal extensions of NgPPR45 could edit one of the targets for PpPPR78, cox1eU755SL, up to 82%. This proves that the C-terminal extensions of NgPPR45 have a functional cytidine deaminase and can act in fusion with the independently evolved moss PPR stretch. The failure of other chimeras points out that complete C-terminal extensions might be important for fusion proteins from different systems.

PPR proteins, with their target-specific cytidine deamination process, make them an ideal molecular tool for transcriptomic engineering. However, chimeras of PPR proteins do not always work. The PLS stretch of PPR65 is not competitive with the C-terminal extensions of PPR56, while the opposite is true. Changing the DYW domains of PPR56 from different heterologous sources results in restricted or widened off-target numbers in the *E. coli* transcriptome. The experiments show that the preference of the DYW domains does not influence the immediate sequence of the editing site environment, but rather has a long-range impact on the upstream PPR stretch.

To further investigate the mechanism of RNA editing, PPR proteins can be overexpressed in the cytosol of plants along with a supplied target in knockout plants. Additionally, off-targets in the nuclear transcriptome can be examined to provide a wider range of options for transcriptomic engineering. Due to its high activity and flexibility, PPR56 is an excellent editing factor and could serve as a solid foundation for a molecular tool.

6. Acknowledgment

It would not have been possible for me to complete this work without the support of my colleagues, family, and friends, especially during the stressful times of the pandemic. After more than ten years of studying abroad, I am finally able to return to my beloved home country and continue the rest of my life. It is a great honor to have completed my studies from the master's to doctoral level in the university of Bonn, and I am deeply grateful to my examination committee members. Their initial interest in biology and evolution has led to a renowned career, and I am privileged to have been a part of it.

I would like to thank my two supervisors, Prof. Dr. Volker Knoop and Dr. Mareike Schallenberg-Rüdinger, for giving me the chance to learn, understand, and investigate in the field of plant-type RNA editing. Thank you for your knowledge and support over the last few years. Being alone abroad and far away from my family, I feel grateful to have had you guide me as my academic parents.

Thanks to Prof. Peter Dormann for graciously agreeing to co-review my thesis on a tight timeline before my visa expires. I thoroughly enjoyed the lab courses and lectures in my master's program, which piqued my interest in the molecular plant field.

I would like to thank all the members of the AG Molekulare Evolution for creating a great working atmosphere, especially Dr. Bastian Oldenkott, Elena Lesch, Dr. Simon Zumkeller, Kira Ritzenhofen, and Sarah Brenne for their help and support in various ways.

I would like to thank my friend Jie Zhang and my boyfriend Hongcan Wu for patiently helping and encouraging me to finish this thesis. Thanks to my family for their loving support and for backing me throughout the doctoral program. It's time for me to go back to them and enjoy the coming days!

7. Appendix

All appendix data is stored on a compact disc enclosed to this thesis.

Supplementary tables

Supplementary Table 1: List of oligonucleotides

Supplementary Table 2: Overview of *E. coli* experiments

Supplementary Table 3: Off-targets obtained from native PPR56 and mutants.

Supplementary datasets

Supplementary Dataset 1: PDF version and Supplementary Information: Yang et al. 2023a

Supplementary Dataset 2: PDF version and Supplementary Information: Yang et al. 2023b

Supplementary Dataset 3: PDF version and Supplementary Information: Yang et al. 2023c

8. Abbreviation

%	percent
°C	degree Celsius
m	micro
A	adenosine / alanine
Aa	amino acid
Amp	ampicillin
BLAST	Basic Local Alignment Search Tool
bp	base pair
C	cytidine / cysteine
cDNA	complementary DNA
D	aspartate
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
E	glutamic acid
F	phenylalanine
G	guanine / glycine
h	hour
H	histidine
I	isoleucine
IPTG	isopropyl b-D-1-thiogalactopyranoside
IUPAC	International Union of Pure and Applied Chemistry
K	lysine
kb	kilobase pairs
L	leucine
LB	Luria-Bertani
m	milli
M	molarity / methionine
mRNA	messenger RNA
n	nano
N	normality / asparagine
NCBI	National Center for Biotechnology Information
no.	number
nt	nucleotide
P	proline
PCR	polymerase chain reaction
pos.	position

PPR	pentatricopeptide repeat
Q	glutamine
R	arginine
RNA	ribonucleic acid
rRNA	ribosomal RNA
S	serine
SDS	sodium dodecyl sulfate
T	thymidine / threonine
tRNA	transfer RNA
U	unit / uridine
UV	ultraviolet
W	tryptophan
Y	tyrosine

9. Bibliography

Main work

Yang, Y., K. Ritzenhofen, J. Otrzonsek, M. Schallenberg-Rüdinger, and V. Knoop. 2023a. Beyond a PPR-RNA recognition code: Many aspects matter for the multi-targeting properties of RNA editing factor PPR56. *PLOS Genetics* https://doi.org/10.1101/2023.04.05.535663

Yang, Y., B. Oldenkott, K. Ritzenhofen, S. Ramanathan, M. Schallenberg-Rüdinger, and V. Knoop. 2023b. Highly different behaviors of plant C-to-U RNA editing factors upon creation of chimeras and strikingly different off-target spectra. *Communications Biology*

Yang, Y., B. Oldenkott, E. Lesch, S. Ramanathan, V. Knoop, and M. Schallenberg-Rüdinger. 2023c. Functionality of the cytidine deaminase domain of DYW-type PPR protein NgPPR45 of the protist Naegleria gruberi in moss mitochondria.

Reference

Andres-Colas, N., Q. Zhu, M. Takenaka, B. De Rybel, D. Weijers, and D. Van Der Straeten. 2017. Multiple PPR protein interactions are involved in the RNA editing system in Arabidopsis mitochondria and plastids. *Proc Natl Acad Sci U S A*. 114:8883-8888 https://doi.org/10.1073/pnas.1705815114

Auxilien, S., P.F. Crain, R.W. Trewyn, and H. Grosjean. 1996. Mechanism, specificity and general properties of the yeast enzyme catalysing the formation of inosine 34 in the anticodon of transfer RNA. *J Mol Biol*. 262:437-458 https://doi.org/10.1006/jmbi.1996.0527.

Bar-Yaacov, D., Y. Pilpel, and O. Dahan. 2018. RNA editing in bacteria: occurrence, regulation and significance. *RNA Biol*. 15:863-867 https://doi.org/10.1080/15476286.2018.1481698

Bar-Yaacov, D., E. Mordret, R. Towers, T. Biniashvili, C. Soyris, S. Schwartz, O. Dahan, and Y. Pilpel. 2017. RNA editing in bacteria recodes multiple proteins and regulates an evolutionarily conserved toxin-antitoxin system. *Genome Res*. 27:1696-1703 https://doi.org/10.1101/gr.222760.117

Barkan, A., and I. Small. 2014. Pentatricopeptide repeat proteins in plants. *Annu Rev Plant Biol*. 65:415-442 https://doi.org/10.1146/annurev-arplant-050213-040159

Barkan, A., M. Rojas, S. Fujii, A. Yap, Y.S. Chong, C.S. Bond, and I. Small. 2012. A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet*. 8:e1002910 https://doi.org/10.1371/journal.pgen.1002910

Bayer-Csaszar, E., S. Haag, A. Jorg, F. Glass, B. Hartel, T. Obata, E.H. Meyer, A. Brennicke, and M. Takenaka. 2017. The conserved domain in MORF proteins has distinct affinities to the PPR and E elements in PPR RNA editing factors. *Biochim Biophys Acta Gene Regul Mech*. 1860:813-828 https://doi.org/10.1016/j.bbagrm.2017.05.004

Becker, J., and C. Wittmann. 2016. Systems metabolic engineering of Escherichia coli for the

heterologous production of high value molecules-a veteran at new shores. *Curr Opin Biotechnol*. 42:178-188 https://doi.org/10.1016/j.copbio.2016.05.004

Begu, D., B. Castandet, and A. Araya. 2011. RNA editing restores critical domains of a group I intron in fern mitochondria. *Curr Genet*. 57:317-325 https://doi.org/10.1007/s00294-011-0349-z

Beike, A.K., D. Lang, A.D. Zimmer, F. Wüst, D. Trautmann, G. Wiedemann, P. Beyer, E.L. Decker, and R. Reski. 2015. Insights from the cold transcriptome of Physcomitrella patens: global specialization pattern of conserved transcriptional regulators and identification of orphan genes involved in cold acclimation. *New Phytol*. 205:869-881 https://doi.org/10.1111/nph.13004

Beike, A.K., M.v. Stackelberg, M. Schallenberg-Rüdinger, S.T. Hanke, M. Follo, D. Quandt, S.F. McDaniel, R. Reski, B.C. Tan, and S.A. Rensing. 2014. Molecular evidence for convergent evolution and allopolyploid speciation within the Physcomitrium-Physcomitrellaspecies complex. *BMC Evolutionary Biology*. 14 https://doi.org/10.1186/1471-2148-14-158

Benne, R. 1994. RNA-editing in trypanosome mitochondria. *European Journal of Biochemistry*. 221:9-23 https://doi.org/10.1111/j.1432-1033.1994.tb18710.x.

Benne, R., J.V.D. Burg, J.P.J. Brakenhoff, P. Sloof, J.H.V. Boom, and M.C. Tromp. 1986. Major transcript of the frameshifted coxll gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell*. 46:819-826 https://doi.org/10.1016/0092-8674(86)90063-2

Bentolila, S., A.M. Babina, A. Germain, and M.R. Hanson. 2013. Quantitative trait locus mapping identifies REME2, a PPR-DYW protein required for editing of specific C targets in Arabidopsis mitochondria. *RNA Biol*. 10:1520-1525 https://doi.org/10.4161/rna.25297

Bentolila, S., W.P. Heller, T. Sun, A.M. Babina, G. Friso, K.J. van Wijk, and M.R. Hanson. 2012. RIP1, a member of an Arabidopsis protein family, interacts with the protein RARE1 and broadly affects RNA editing. *Proc Natl Acad Sci U S A*. 109:E1453-1461 https://doi.org/10.1073/pnas.1121465109

Bernath-Levin, K., J. Schmidberger, S. Honkanen, B. Gutmann, Y.K. Sun, A. Pullakhandam, C.C.d. Francs-Small, C.S. Bond, and I. Small. 2022. Cofactor-independent RNA editing by a synthetic S-type PPR protein. *Synthetic Biology*. 7:1-11 https://doi.org/10.1093/synbio/ysab034

Betts, L., S. Xiang, S.A. Short, R. Wolfenden, and C.W.C. Jr. 1994. Cytidine deaminase. The 2.3 A crystal structure of an enzyme: transition-state analog complex. *J Mol Biol*. 235:635-656 https://doi.org/10.1006/jmbi.1994.1018.

Bezanilla, M., A. Pan, and R.S. Quatrano. 2003. RNA interference in the moss Physcomitrella patens. *Plant Physiol*. 133:470-474 https://doi.org/10.1104/pp.103.024901

Binder, S., A. Marchfelder, and A. Brennicke. 1994. RNA editing of tRNA(Phe) and tRNA(Cys) in mitochondria of Oenothera berteriana is initiated in precursor molecules. *Mol Gen Genet*. 244:67-74

Binder, S., A. Marchfelder, A. Brennicke, and B. Wissinger. 1992. RNA editing in trans-splicing intron sequences of nad2 mRNAs in Oenothera mitochondria. *J Biol Chem*. 267:7615-7623

Blanc, V., S. Litvak, and A. Araya. 1995. RNA editing in wheat mitochondria proceeds by a deamination mechanism. *FEBS Lett*. 373:56-60

Block, M.D., L. Herrera-Estrella, M.V. Montagu, J. Schell, and P. Zambryski. 1984. Expression of foreign genes in regenerated plants and in their progeny. *EMBO J.* 3:1681-1689 https://doi.org/10.1002/j.1460-2075.1984.tb02032.x

Bobik, K., T.N. McCray, B. Ernest, J.C. Fernandez, K.A. Howell, T. Lane, M. Staton, and T.M. Burch-Smith. 2017. The chloroplast RNA helicase ISE2 is required for multiple chloroplast RNA processing steps in Arabidopsis thaliana. *Plant J.* 91:114-131 https://doi.org/10.1111/tpj.13550

Borner, G.V., M. Morl, B. Wissinger, A. Brennicke, and C. Schmelzer. 1995. RNA editing of a group II intron in Oenothera as a prerequisite for splicing. *Mol Gen Genet*. 246:739-744

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

Boussardon, C., V. Salone, A. Avon, R. Berthome, K. Hammani, K. Okuda, T. Shikanai, I. Small, and C. Lurin. 2012. Two interacting proteins are necessary for the editing of the NdhD-1 site in Arabidopsis plastids. *Plant Cell*. 24:3684-3694 https://doi.org/10.1105/tpc.112.099507

Bruns, N. 2022. Pentatricopeptide repeat (PPR) Proteine der KPAxA-Klade in Hornmoosen. *In* Mathematisch-Naturwissenschaftliche Fakultät. Vol. Bachelor of Science. Rheinische Friedrich-Wilhelms-Universität Bonn.

Bundschuh, R., J. Altmüller, C. Becker, P. Nürnberg, and J.M. Gott. 2011. Complete characterization of the edited transcriptome of the mitochondrion of Physarum polycephalum using deep sequencing of RNA. *Nucleic Acids Research*. 39:6044-6055 https://doi.org/10.1093/nar/gkr180

Byun, M.Y., S. Seo, J. Lee, Y.-H. Yoo, and H. Lee. 2021. Transfection of Arctic Bryum sp. KMR5045 as a Model for Genetic Engineering of Cold-Tolerant Mosses. *Front. Plant Sci.* 11 https://doi.org/10.3389/fpls.2020.609847

Cai, M., Y. Huang, R. Yang, R. Craigie, and G.M. Clore. 2016. A simple and robust protocol for high-yield expression of perdeuterated proteins in Escherichia coli grown in shaker flasks. *J Biomol NMR*. 66:85-91 https://doi.org/10.1007/s10858-016-0052-y

Carpenter, E.J., N. Matasci, S. Ayyampalayam, S. Wu, J. Sun, J. Yu, F.R.J. Vieira, C. Bowler, R.G. Dorrell, M.A. Gitzendanner, ..., and G.K. Wong. 2019. Access to RNA-sequencing data from 1173 plant species: The 1000 plant transcriptomes initiative (1KP). *GigaScience*. 8 https://doi.org/10.1093/gigascience/giz126

Cazalet, C., L. Gomez-Valero, C. Rusniok, M. Lomma, D. Dervins-Ravault, H.J. Newton, F.M. Sansom, S. Jarraud, N. Zidane, L. Ma, C. Bouchier, J. Etienne, E.L. Hartland, and C. Buchrieser. 2010. Analysis of the Legionella longbeachae Genome and Transcriptome Uncovers Unique Strategies to Cause Legionnaires' Disease. *PLoS Genet*. 6 https://doi.org/10.1371/journal.pgen.1000851

Chateigner-Boutin, A.L., and I. Small. 2010. Plant RNA editing. *RNA Biol*. 7:213-219

Chateigner-Boutin, A.L., C. Colas des Francs-Small, S. Fujii, K. Okuda, S.K. Tanz, and I. Small. 2013. The E domains of pentatricopeptide repeat proteins from different organelles are not functionally equivalent for RNA editing. *Plant J.* 74:935-945 https://doi.org/10.1111/tpj.12180

Chen, X., L. Zhou, K. Tian, A. Kumar, S. Singh, B.A. Prior, and Z. Wang. 2013. Metabolic engineering of Escherichia coli: a sustainable industrial platform for bio-based chemical production. *Biotechnol Adv*. 31:1200-1223 https://doi.org/10.1016/j.biotechadv.2013.02.009

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

Cheong, C.-G., and T.M.T. Hall. 2006. Engineering RNA sequence specificity of Pumilio repeats. *Proc Natl Acad Sci U S A*. 103:13635-13639 https://doi.org/10.1073/pnas.0606294103.

Cho, S.H., Y.S. Chung, S.K. Cho, Y.W. Rim, and J.S. Shin. 1999. Particle bombardment mediated transformation and GFP expression in the moss Physcomitrella patens. *Mol Cells.* 9:14-19

Choudhary, M., X. Zanhua, Y.X. Fu, and S. Kaplan. 2007. Genome Analyses of Three Strains of Rhodobacter sphaeroides: Evidence of Rapid Evolution of Chromosome II. *Journal of Bacteriology*. 189 https://doi.org/10.1128/JB.01498-06

Collonnier, C., A. Epert, K. Mara, F. Maclot, A. Guyon-Debast, F. Charlot, C. White, D.G. Schaefer, and F. Nogué. 2017. CRISPR-Cas9-mediated efficient directed mutagenesis and RAD51-dependent and RAD51-independent gene targeting in the moss Physcomitrella patens. *Plant Biotechnol J*. 15:122-131 https://doi.org/10.1111/pbi.12596

Coquille, S., A. Filipovska, T. Chia, L. Rajappa, J.P. Lingford, M.F. Razif, S. Thore, and O. Rackham. 2014. An artificial PPR scaffold for programmable RNA recognition. *Nat Commun*. 5:5729 https://doi.org/10.1038/ncomms6729

Cove, D.J., P.-F. Perroud, A.J. Charron, S.F. McDaniel, A. Khandelwal, and R.S. Quatrano. 2009. Transformation of the moss Physcomitrella patens using T-DNA mutagenesis. *Cold Spring Harb Protoc*. 2 https://doi.org/10.1101/pdb.prot5144

Covello, P.S., and M.W. Gray. 1989. RNA editing in plant mitochondria. *Nature*. 341:662-666 https://doi.org/10.1038/341662a0

D'Andrea, L., and L. Regan. 2003. TPR proteins: the versatile helix. *Trends Biochem Sci.* 28 https://doi.org/10.1016/j.tibs.2003.10.007

Danielsen, S., M. Kilstrup, K. Barilla, B. Jochimsen, and J. Neuhard. 1992. Characterization of the Escherichia coli codBA operon encoding cytosine permease and cytosine deaminase. *Mol Microbiol*. 6:1335-1344 https://doi.org/10.1111/j.1365-2958.1992.tb00854.x.

Das, A., P. Cohen, and D. Barford. 1998. The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *EMBO J.* 17:1192-1199 https://doi.org/10.1093/emboj/17.5.1192

Delannoy, E., M. Le Ret, E. Faivre-Nitschke, G.M. Estavillo, M. Bergdoll, N.L. Taylor, B.J. Pogson, I. Small, P. Imbault, and J.M. Gualberto. 2009. Arabidopsis tRNA adenosine deaminase arginine edits the wobble nucleotide of chloroplast tRNAArg(ACG) and is essential for efficient chloroplast translation. *Plant Cell*. 21:2058-2071 https://doi.org/10.1105/tpc.109.066654

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

Dombrovska, O., and Y.-L. Qiu. 2004. Distribution of introns in the mitochondrial gene nad1 in land plants: phylogenetic and molecular evolutionary implications. *Mol Phylogenet Evol*. 32:246-263 https://doi.org/10.1016/j.ympev.2003.12.013.

Edera, A.A., C.L. Gandini, and M.V. Sanchez-Puerta. 2018. Towards a comprehensive picture of C-to-U RNA editing sites in angiosperm mitochondria. *Plant Mol Biol* https://doi.org/10.1007/s11103-018-0734-9

Engler, C., R. Gruetzner, R. Kandzia, and S. Marillonnet. 2009. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLoS ONE*. 4 https://doi.org/10.1371/journal.pone.0005553

Faktorová, D., R.E.R. Nisbet, J.A.F. Robledo, E. Casacuberta, L. Sudek, ..., and J. Lukeš. 2020. Genetic tool development in marine protists: emerging model organisms for experimental cell biology. *Nature Methods*. 17:pages481–494 https://doi.org/10.1038/s41592-020-0796-x

Freyer, R., Kiefer-Meyer, M.-C., Kössel H. 1997. Occurrence of plastid RNA editing in all major lineages of land plants. *Proc Natl Acad Sci U S A.*

Fu, C.-J., S. Sheikh, W. Miao, S.G.E. Andersson, and S.L. Baldauf. 2014. Missing Genes, Multiple ORFs, and C-to-U Type RNA Editing in Acrasis kona (Heterolobosea, Excavata) Mitochondrial DNA. *Genome Biol Evol*. 6:2240-2257 https://doi.org/10.1093/gbe/evu180

Fuchs, P., N. Rugen, C. Carrie, M. Elsässer, I. Finkemeier, J. Giese, T.M. Hildebrandt, K. Kühn, V.G. Maurino, C. Ruberti, M. Schallenberg-Rüdinger, J. Steinbeck, H.-P. Braun, H. Eubel, E.H. Meyer, S.J. Müller-Schüssele, and M. Schwarzländer. 2020. Single organelle function and organization as estimated from Arabidopsis mitochondrial proteomics. *Plant J*. 101:420-441 https://doi.org/10.1111/tpj.14534

Fujii, S., C.S. Bond, and I.D. Small. 2011. Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. *Proc Natl Acad Sci U S A*. 108:1723-1728 https://doi.org/10.1073/pnas.1007667108

Fujii, S., N. Sato, and T. Shikanai. 2013. Mutagenesis of Individual Pentatricopeptide Repeat Motifs Affects RNA Binding Activity and Reveals Functional Partitioning of Arabidopsis PROTON GRADIENT REGULATION3. *The Plant Cell*. 25 https://doi.org/10.1105/tpc.113.112193

Gerber, A.P., and W. Keller. 2001. RNA editing by base deamination: more enzymes, more targets, new mysteries. *Trends Biochem Sci*. 26:376-384 https://doi.org/10.1016/s0968-0004(01)01827-8.

Gerke, P., P. Szövényi, A. Neubauer, H. Lenz, B. Gutmann, R. McDowell, I. Small, M. Schallenberg-Rüdinger, and V. Knoop. 2020. Towards a plant model for enigmatic U-to-C RNA editing: the organelle genomes, transcriptomes, editomes and candidate RNA editing factors in the hornwort Anthoceros agrestis. *New Phytol*. 225:1976-1992 https://doi.org/10.1111/nph.16297

Glass, F., B. Härtel, A. Zehrmann, D. Verbitskiy, and M. Takenaka. 2015. MEF13 Requires MORF3 and MORF8 for RNA Editing at Eight Targets in Mitochondrial mRNAs in Arabidopsis thaliana. *Mol Plant*. 8 https://doi.org/10.1016/j.molp.2015.05.008.

Grewe, F., P. Viehoever, B. Weisshaar, and V. Knoop. 2009. A trans-splicing group I intron and tRNA-hyperediting in the mitochondrial genome of the lycophyte Isoetes engelmannii. *Nucleic Acids Res*. 37:5093-5104 https://doi.org/10.1093/nar/gkp532

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

Groth-Malonek, M., U. Wahrmund, M. Polsakiewicz, and V. Knoop. 2007. Evolution of a Pseudogene: Exclusive Survival of a Functional Mitochondrial nad7 Gene Supports Haplomitrium as the Earliest Liverwort Lineage and Proposes a Secondary Loss of RNA Editing in Marchantiidae. *Molecular Biology and Evolution*. 24:1068–1074

Gualberto, J.M., L. Lamattina, G. Bonnard, J.H. Weil, and J.M. Grienenberger. 1989. RNA editing in wheat mitochondria results in the conservation of protein sequences. *Nature*. 341:660-662 https://doi.org/10.1038/341660a0

Guillaumot, D., M. Lopez-Obando, K. Baudry, A. Avon, G. Rigaill, A. Falcon de Longevialle, B. Broche, M. Takenaka, R. Berthome, G. De Jaeger, E. Delannoy, and C. Lurin. 2017. Two interacting PPR proteins are major Arabidopsis editing factors in plastid and mitochondria. *Proc Natl Acad Sci U S A*. 114:8877-8882 https://doi.org/10.1073/pnas.1705780114

Gutmann, B., S. Royan, M. Schallenberg-Rüdinger, H. Lenz, I.R. Castleden, R. McDowell, M.A. Vacher, J. Tonti-Filippini, C.S. Bond, V. Knoop, and I.D. Small. 2020. The Expansion and Diversification of Pentatricopeptide Repeat RNA-Editing Factors in Plants. *Molecular Plant*. 13 https://doi.org/10.1016/j.molp.2019.11.002.

Hammani, K., K. Okuda, S.K. Tanz, A.-L. Chateigner-Boutin, T. Shikanai, and I. Small. 2009. A Study of New Arabidopsis Chloroplast RNA Editing Mutants Reveals General Features of Editing Factors and Their Target Sites. *The Plant Cell*. 21:3686–3699 https://doi.org/10.1105/tpc.109.071472

Hanraths, A.E. 2022. Investigating Plant-type C-U RNA Editing Mediated by PPR56 Variants in the Moss Physcomitrium patens and in the Heterologous System Saccharomyces cerevisiae. *In* Faculty of Mathematics and Natural Sciences. Vol. Bachelor of Science. Rheinische Friedrich-Wilhelms-Universität Bonn.

Härtel, B., A. Zehrmann, D. Verbitskiy, J.A.v.d. Merwe, A. Brennicke, and M. Takenaka. 2013. MEF10 is required for RNA editing at nad2-842 in mitochondria of Arabidopsis thaliana and interacts with MORF8. *Plant Mol Biol*. 81:337-346 https://doi.org/10.1007/s11103-012-0003-2

Hashimoto, M., T. Endo, G. Peltier, M. Tasaka, and T. Shikanai. 2003. A nucleus-encoded factor, CRR2, is essential for the expression of chloroplast ndhB in Arabidopsis. *Plant J*. 36:541-549 https://doi.org/10.1046/j.1365-313x.2003.01900.x.

Hayes, M.L., and P.I. Santibanez. 2020. A plant pentatricopeptide repeat protein with a DYW-deaminase domain is sufficient for catalyzing C-to-U RNA editing in vitro. *J Biol Chem*. 295:3497-3505 https://doi.org/10.1074/jbc.RA119.011790

Hayes, M.L., K. Giang, B. Berhane, and R.M. Mulligan. 2013. Identification of two pentatricopeptide repeat genes required for RNA editing and zinc binding by C-terminal cytidine deaminase-like domains. *J Biol Chem*. 288:36519-36529 https://doi.org/10.1074/jbc.M113.485755

Hecht, J., F. Grewe, and V. Knoop. 2011. Extreme RNA editing in coding islands and abundant microsatellites in repeat sequences of Selaginella moellendorffii mitochondria: the root of frequent plant mtDNA recombination in early tracheophytes. *Genome Biol Evol*. 3:344-358

https://doi.org/10.1093/gbe/evr027

Hein, A., and V. Knoop. 2018. Expected and unexpected evolution of plant RNA editing factors CLB19, CRR28 and RARE1: retention of CLB19 despite a phylogenetically deep loss of its two known editing targets in Poaceae. *BMC Evolutionary Biology*. 18 https://doi.org/10.1186/s12862-018-1203-4

Hein, A., M. Polsakiewicz, and V. Knoop. 2016. Frequent chloroplast RNA editing in early-branching flowering plants: pilot studies on angiosperm-wide coexistence of editing sites and their nuclear specificity factors. *BMC Evolutionary Biology*. 16 https://doi.org/10.1186/s12862-016-0589-0

Hein, A., S. Brenner, and V. Knoop. 2019. Multifarious Evolutionary Pathways of a Nuclear RNA Editing Factor: Disjunctions in Coevolution of DOT4 and Its Chloroplast Target rpoC1eU488SL. *Genome Biology and Evolution*. 11:798-813 https://doi.org/10.1093/gbe/evz032

Hein, A., S. Brenner, M. Polsakiewicz, and V. Knoop. 2020. The dual-targeted RNA editing factor AEF1 is universally conserved among angiosperms and reveals only minor adaptations upon loss of its chloroplast or its mitochondrial target. *Plant Mol Biol*. 102:185-198 https://doi.org/10.1007/s11103-019-00940-9.

Hiesel, R., B. Wissinger, W. Schuster, and A. Brennicke. 1989. RNA editing in plant mitochondria. *Science*. 246:1632-1634

Hirose, T., and M. Sugiura. 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 J.* 20:1144-1152 https://doi.org/10.1093/emboj/20.5.1144

Ho, S.N., H.D. Hunt, R.M. Horton, J.K. Pullen, and L.R. Pease. 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*. 77:51-59

Hoch, B., R.M. Maier, K. Appel, G.L. Igloi, and H. Kossel. 1991. Editing of a chloroplast mRNA by creation of an initiation codon. *Nature*. 353:178-180 https://doi.org/10.1038/353178a0

Hoffman, C.S., V. Wood, and P.A. Fantes. 2015. An Ancient Yeast for Young Geneticists: A Primer on the Schizosaccharomyces pombe Model System. *Genetics*. 201:403-423 https://doi.org/10.1534/genetics.115.181503

Hollósy, F. 2002. Effects of ultraviolet radiation on plant cells. *Micron*. 33:179-197 https://doi.org/10.1016/s0968-4328(01)00011-7

Horsch, R.B., R.T. Fraley, S.G. Rogers, P.R. Sanders, A. Lloyd, and N. Hoffmann. 1984. Inheritance of functional foreign genes in plants. *Science*. 223:496-498 https://doi.org/10.1126/science.223.4635.496.

Hu, Z., R. Vanderhaeghen, T. Cools, Y. Wang, I. De Clercq, O. Leroux, L. Nguyen, K. Belt, A.H. Millar, D. Audenaert, P. Hilson, I. Small, G. Mouille, S. Vernhettes, F. Van Breusegem, J. Whelan, H. Hofte, and L. De Veylder. 2016. Mitochondrial Defects Confer Tolerance against Cellulose Deficiency. *Plant Cell*. 28:2276-2290 https://doi.org/10.1105/tpc.16.00540

Huang, C., Z.-R. Li, Q.-B. Yu, L.-S. Ye, Y.-L. Cui, D.P. Molloy, and Z.-N. Yang. 2019. MORF2 tightly associates with MORF9 to regulate chloroplast RNA editing in Arabidopsis. *Plant Sci* https://doi.org/10.1016/j.plantsci.2018.10.020

Huang, X., J. Lv, Y. Li, S. Mao, Z. Li, Z. Jing, Y. Sun, X. Zhang, S. Shen, X. Wang, M. Di, J. Ge, X. Huang, E. Zuo, and T. Chi. 2020. Programmable C-to-U RNA editing using the human APOBEC3A deaminase. *EMBO J*. 39:e104741 https://doi.org/10.15252/embj.2020104741

Ichinose, M., and M. Sugita. 2018. The DYW Domains of Pentatricopeptide Repeat RNA Editing Factors Contribute to Discriminate Target and Non-Target Editing Sites. *Plant Cell Physiol*. 59:1652-1659 https://doi.org/10.1093/pcp/pcy086

Ichinose, M., M. Uchida, and M. Sugita. 2014. Identification of a pentatricopeptide repeat RNA editing factor in Physcomitrella patens chloroplasts. *FEBS Lett*. 588:4060-4064 https://doi.org/10.1016/j.febslet.2014.09.031

Ichinose, M., E. Tasaki, C. Sugita, and M. Sugita. 2012. A PPR-DYW protein is required for splicing of a group II intron of cox1 pre-mRNA in Physcomitrella patens. *Plant J*. 70 https://doi.org/10.1111/j.1365-313X.2011.04869.x.

Ichinose, M., C. Sugita, Y. Yagi, T. Nakamura, and M. Sugita. 2013. Two DYW subclass PPR proteins are involved in RNA editing of ccmFc and atp9 transcripts in the moss Physcomitrella patens: first complete set of PPR editing factors in plant mitochondria. *Plant Cell Physiol*. 54:1907-1916 https://doi.org/10.1093/pcp/pct132

Ichinose, M., A. Ishimaru, C. Sugita, K. Nakajima, Y. Kawaguchi, and M. Sugita. 2020. Two Novel PLS-Class Pentatricopeptide Repeat Proteins Are Involved in the Group II Intron Splicing of Mitochondrial Transcripts in the Moss Physcomitrella patens. *Plant Cell Physiol*. 61:1687-1698 https://doi.org/10.1093/pcp/pcaa070

Ichinose, M., M. Kawabata, Y. Akaiwa, Y. Shimajiri, I. Nakamura, T. Tamai, T. Nakamura, Y. Yagi, and B. Gutmann. 2022. U-to-C RNA editing by synthetic PPR-DYW proteins in bacteria and human culture cells. *Communications Biology*. 5 https://doi.org/10.1038/s42003-022-03927-3

Kamisugi, Y., A.C. Cuming, and D.J. Cove. 2005. Parameters determining the efficiency of gene targeting in the moss Physcomitrella patens. *Nucleic Acids Res*. 33 https://doi.org/10.1093/nar/gni172

Kang, B.-C., S.-J. Bae, S. Lee, J.S. Lee, A. Kim, H. Lee, G. Baek, H. Seo, J. Kim, and J.-S. Kim. 2021. Chloroplast and mitochondrial DNA editing in plants. *Nature Plants*. 7:899-905 https://doi.org/10.1038/s41477-021-00943-9

Karcher, D., and R. Bock. 2002. Temperature sensitivity of RNA editing and intron splicing reactions in the plastid ndhB transcript. *Curr Genet*. 41:48-52 https://doi.org/10.1007/s00294-002-0278-y

Karcher, D., and R. Bock. 2009. Identification of the chloroplast adenosine-to-inosine tRNA editing enzyme. *RNA*. 15:1251-1257 https://doi.org/10.1261/rna.1600609

Karpenahalli, M.R., A.N. Lupas, and J. Söding. 2007. TPRpred: a tool for prediction of TPR-, PPR- and SEL1-like repeats from protein sequences. *BMC Bioinformatics*. 8 https://doi.org/10.1186/1471-2105-8-2

Kimple, M.E., A.L. Brill, and R.L. Pasker. 2013. Overview of affinity tags for protein purification. *Curr Protoc Protein Sci*. 73:9.9.1-9.9.23 https://doi.org/10.1002/0471140864.ps0909s73

Kindgren, P., A. Yap, C.S. Bond, and I. Small. 2015. Predictable alteration of sequence recognition by RNA editing factors from Arabidopsis. *Plant Cell*. 27:403-416 https://doi.org/10.1105/tpc.114.134189

Knie, N., F. Grewe, S. Fischer, and V. Knoop. 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 https://doi.org/10.1186/s12862-016-0707-z

Knoop, V. 2011. When you can't trust the DNA: RNA editing changes transcript sequences. *Cell Mol Life Sci*. 68:567-586 https://doi.org/10.1007/s00018-010-0538-9

Knoop, V. 2020. Plant PPRs Come in Multiple Flavors—But Why? *Plant and Cell Physiology*. 61:1685-1686 https://doi.org/10.1093/pcp/pcaa096

Knoop, V. 2022. C-to-U and U-to-C: RNA editing in plant organelles and beyond. *Journal of Experimental Botany*. erac488 https://doi.org/10.1093/jxb/erac488

Knoop, V., and M. Rüdinger. 2010. DYW-type PPR proteins in a heterolobosean protist: plant RNA editing factors involved in an ancient horizontal gene transfer? *FEBS Lett*. 584:4287-4291 https://doi.org/10.1016/j.febslet.2010.09.041

Kobayashi, K., M. Kawabata, K. Hisano, T. Kazama, K. Matsuoka, M. Sugita, and T. Nakamura. 2012. Identification and characterization of the RNA binding surface of the pentatricopeptide repeat protein. *Nucleic Acids Res*. 40:2712-2723 https://doi.org/10.1093/nar/gkr1084

Kobayashi, T., Y. Yagi, and T. Nakamura. 2019. Comprehensive Prediction of Target RNA Editing Sites for PLS-Class PPR Proteins in Arabidopsis thaliana. *Plant Cell Physiol*. 60:862-874 https://doi.org/10.1093/pcp/pcy251.

Komor, A.C., Y.B. Kim, M.S. Packer, J.A. Zuris, and D.R. Liu. 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*. 533:420-424 https://doi.org/10.1038/nature17946

Kotera, E., M. Tasaka, and T. Shikanai. 2005. A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature*. 433:326-330 https://doi.org/10.1038/nature03229

Kunzmann, A., A. Brennicke, and A. Marchfelder. 1998. 5' end maturation and RNA editing have to precede tRNA 3' processing in plant mitochondria. *Proc Natl Acad Sci U S A*. 95:108-113

Lang, D., K.K. Ullrich, F. Murat, ..., and S.A. Rensing. 2018. The Physcomitrella patens chromosome-scale assembly reveals moss genome structure and evolution. *The plant Journal*. 93:515-533 https://doi.org/10.1111/tpj.13801

Lee, J.H., S.-C. Jung, L.M. Bui, K.H. Kang, J.-J. Song, and S.C. Kim. 2013. Improved Production of l-Threonine in Escherichia coli by Use of a DNA Scaffold System. *Appl Environ Microbiol*. 79:774-782 https://doi.org/10.1128/AEM.02578-12

Leebens-Mack, J.H., M.S. Barker, E.J. Carpenter, M.K. Deyholos, M.A. Gitzendanner, S.W. Graham, I. Grosse, Z. Li, M. Melkonian, S. Mirarab, ..., and G.K. Wong. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*. 574:679-685 https://doi.org/10.1038/s41586-019-1693-2

Lenz, H., A. Hein, and V. Knoop. 2018. Plant organelle RNA editing and its specificity factors: enhancements of analyses and new database features in PREPACT 3.0. *BMC Bioinformatics*. 19:255 https://doi.org/10.1186/s12859-018-2244-9

Lesch, E., M.T. Schilling, S. Brenner, Y. Yang, O.J. Gruss, V. Knoop, and M. Schallenberg-Rüdinger. 2022. Plant mitochondrial RNA editing factors can perform targeted C-to-U editing of nuclear

transcripts in human cells. *Nucleic Acids Res*. 50:9966-9983 https://doi.org/10.1093/nar/gkac752

Lesch, E.I.P. 2020. Evolution of moss RNA editing factors and their functions tested in a variety of model systems. *In* Faculty of Mathematics and Natural Sciences. Vol. Master of Science. Rheinische Friedrich-Wilhelms-Universität Bonn.

Li, X., M. Sun, S. Liu, Q. Teng, S. Li, and Y. Jiang. 2021. Functions of PPR Proteins in Plant Growth and Development. *Int. J. Mol. Sci.* 22 https://doi.org/10.3390/ijms222011274

Liu, Y., M.G. Johnson, C.J. Cox, R. Medina, N. Devos, A. Vanderpoorten, L. Hedenäs, N.E. Bell, J.R. Shevock, B. Aguero, D. Quandt, N.J. Wickett, A.J. Shaw, and B. Goffinet. 2019. Resolution of the ordinal phylogeny of mosses using targeted exons from organellar and nuclear genomes. *Nature Communications*. 10 https://doi.org/10.1038/s41467-019-09454-w

Loiacono, F.V., D. Walther, S. Seeger, W. Thiele, I. Gerlach, D. Karcher, M.A. Schöttler, R. Zoschke, and R. Bock. 2022. Emergence of Novel RNA-Editing Sites by Changes in the Binding Affinity of a Conserved PPR Protein. *Molecular Biology and Evolution*. 39 https://doi.org/10.1093/molbev/msac222

Longevialle, A.F.d., I.D. Small, and C. Lurin. 2010. Nuclearly encoded splicing factors implicated in RNA splicing in higher plant organelles. *Mol Plant*. 3:691-705 https://doi.org/10.1093/mp/ssq025

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

Lutzoni, F., M.D. Nowak, M.E. Alfaro, V. Reeb, J. Miadlikowska, M. Krug, A.E. Arnold, L.A. Lewis, D.L. Swofford, D. Hibbett, K. Hilu, T.Y. James, D. Quandt, and S. Magallón. 2018. Contemporaneous radiations of fungi and plants linked to symbiosis. *Nature Communications*. 9 https://doi.org/10.1038/s41467-018-07849-9

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

Marchfelder, A., A. Brennicke, and S. Binder. 1996. RNA editing is required for efficient excision of tRNA(Phe) from precursors in plant mitochondria. *J Biol Chem*. 271:1898-1903

Marechal-Drouard, L., D. Ramamonjisoa, A. Cosset, J.H. Weil, and A. Dietrich. 1993. Editing corrects mispairing in the acceptor stem of bean and potato mitochondrial phenylalanine transfer RNAs. *Nucleic Acids Res*. 21:4909-4914

Matsuda, T., M. Sugita, and M. Ichinose. 2020. The L motifs of two moss pentatricopeptide repeat proteins are involved in RNA editing but predominantly not in RNA recognition. *PLOS ONE* https://doi.org/10.1371/journal.pone.0232366

McDaniel, S.F., M.v. Stackelberg, S. Richardt, R.S. Quatrano, R. Reski, and S.A. Rensing. 2010. The speciation history of the Physcomitrium--Physcomitrella species complex. *Evolution*. 64:217-231 https://doi.org/10.1111/j.1558-5646.2009.00797.x

Medina, R., M.G. Johnson, Y. Liu, N.J. Wickett, A.J. Shaw, and B. Goffinet. 2019. Phylogenomic

delineation of Physcomitrium (Bryophyta: Funariaceae) based on targeted sequencing of nuclear exons and their flanking regions rejects the retention of Physcomitrella, Physcomitridium and Aphanorrhegma. *Journal of Systematics and Evolution*. 57:404-417 https://doi.org/10.1111/jse.12516

Mehta, A., and D.M. Driscoll. 2002. Identification of domains in apobec-1 complementation factor required for RNA binding and apolipoprotein-B mRNA editing. *RNA*. 8:69-82

Miller, J.C., S. Tan, G. Qiao, K.A. Barlow, J. Wang, D.F. Xia, X. Meng, D.E. Paschon, E. Leung, S.J. Hinkley, G.P. Dulay, K.L. Hua, I. Ankoudinova, G.J. Cost, F.D. Urnov, H.S. Zhang, M.C. Holmes, L. Zhang, P.D. Gregory, and E.J. Rebar. 2010. A TALE nuclease architecture for efficient genome editing. *Nature Biotechnology*. 29:143–148 https://doi.org/10.1038/nbt.1755

Miranda, R.G., M. Rojas, M.P. Montgomery, K.P. Gribbin, and A. Barkan. 2017. RNA-binding specificity landscape of the pentatricopeptide repeat protein PPR10. *RNA*. 23:586-599 https://doi.org/10.1261/rna.059568.116

Miyata, Y., and M. Sugita. 2004. Tissue- and stage-specific RNA editing of rps 14 transcripts in moss (Physcomitrella patens) chloroplasts. *J Plant Physiol*. 161:113-115 https://doi.org/10.1078/0176-1617-01220

Miyata, Y., C. Sugiura, Y. Kobayashi, M. Hagiwara, and M. Sugita. 2002. Chloroplast ribosomal S14 protein transcript is edited to create a translation initiation codon in the moss Physcomitrella patens. *Biochim Biophys Acta*. 1576:346-349

Moreira, S., M. Valach, M. Aoulad-Aissa, C. Otto, and G. Burger. 2016. Novel modes of RNA editing in mitochondria. *Nucleic Acids Res*. 44:4907-49019 https://doi.org/10.1093/nar/gkw188

Nakazato, I., M. Okuno, C. Zhou, T. Itoh, N. Tsutsumi, M. Takenaka, and S.-i. Arimura. 2022. Targeted base editing in the mitochondrial genome of Arabidopsis thaliana. *PNAS*. 119 https://doi.org/10.1073/pnas.2121177119.

Nishida, K., T. Arazoe, N. Yachie, S. Banno, M. Kakimoto, M. Tabata, M. Mochizuki, A. Miyabe, M. Araki, K.Y. Hara, Z. Shimatani, and A. Kondo. 2016. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science*. 353 https://doi.org/10.1126/science.aaf8729.

Nomura, T., T. Sakurai, Y. Osakabe, K. Osakabe, and H. Sakakibara. 2016. Efficient and Heritable Targeted Mutagenesis in Mosses Using the CRISPR/Cas9 System. *Plant Cell Physiol*. 57:2600-2610 https://doi.org/10.1093/pcp/pcw173

Noureddine, A. 2022. When the Moss World Meets the Angiosperm World: Testing the editing efficiency of the angiosperm PPR protein QED1 and of other moss-angiosperm chimeras with PPR56 and PPR78 in the heterologous E. coli system. *In* Faculty of Mathematics and Natural Sciences. Vol. Master of Science. Rheinische Friedrich-Wilhelms-Universität Bonn.

O'Donovan, G.A., G. Edlin, J.A. Fuchs, J. Neuhard, and E. Thomassen. 1971. Deoxycytidine triphosphate deaminase: characterization of an Escherichia coli mutant deficient in the enzyme. *J Bacteriol*. 105:666-672 https://doi.org/10.1128/jb.105.2.666-672.1971.

O'Toole, N., M. Hattori, C. Andres, K. Iida, C. Lurin, C. Schmitz-Linneweber, M. Sugita, and I. Small. 2008. On the expansion of the pentatricopeptide repeat gene family in plants. *Mol bio Evol.* 25:1120-1128 https://doi.org/10.1093/molbev/msn057.

O Malek, K.L., R Hiesel, A Brennicke, V Knoop. 1996. RNA editing in bryophytes and a molecular

phylogeny of land plants. *EMBO J.* 1996 Mar 15; 15(6): 1403–1411.

Oettmeier, C., T. Nakagaki, and H.-G. Döbereiner. 2020. Slime mold on the rise: the physics of Physarum polycephalum. *J. Phys. D: Appl. Phys*. 53 https://doi.org/10.1088/1361-6463/ab866c

Ohtani, S., M. Ichinose, E. Tasaki, Y. Aoki, Y. Komura, and M. Sugita. 2010. Targeted gene disruption identifies three PPR-DYW proteins involved in RNA editing for five editing sites of the moss mitochondrial transcripts. *Plant Cell Physiol*. 51:1942-1949 https://doi.org/10.1093/pcp/pcq142

Okuda, K., and T. Shikanai. 2012. A pentatricopeptide repeat protein acts as a site-specificity factor at multiple RNA editing sites with unrelated cis -acting elements in plastids. *Nucleic Acids Res*. 40:5052–5064 https://doi.org/10.1093/nar/gks164

Okuda, K., F. Myouga, R. Motohashi, K. Shinozaki, and T. Shikanai. 2007. Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing. *Proc Natl Acad Sci U S A*. 104:8178-8183 https://doi.org/10.1073/pnas.0700865104

Okuda, K., H. Shoki, M. Arai, T. Shikanai, I. Small, and T. Nakamura. 2014. Quantitative analysis of motifs contributing to the interaction between PLS-subfamily members and their target RNA sequences in plastid RNA editing. *Plant J*. 80:870-882 https://doi.org/10.1111/tpj.12687

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

Okuda, K., K. Hammani, S.K. Tanz, L. Peng, Y. Fukao, F. Myouga, R. Motohashi, K. Shinozaki, I. Small, and T. Shikanai. 2010. The pentatricopeptide repeat protein OTP82 is required for RNA editing of plastid ndhB and ndhG transcripts. *Plant J*. 61:339-349 https://doi.org/10.1111/j.1365-313X.2009.04059.x

Oldenkott, B. 2020. Plant-type RNA editing in Physcomitrium patens: An ancient RNA editing machinery can operate in other genetic systems. *In* Mathematisch-Naturwissenschaftlichen Fakultät. Vol. Erlangung des Doktorgrades (Dr. rer. nat.). Rheinischen Friedrich-Wilhelms-Universität Bonn.

Oldenkott, B., K. Yamaguchi, S. Tsuji-Tsukinoki, N. Knie, and V. Knoop. 2014. Chloroplast RNA editing going extreme: more than 3400 events of C-to-U editing in the chloroplast transcriptome of the lycophyte Selaginella uncinata. *RNA*. 20:1499-1506 https://doi.org/10.1261/rna.045575.114

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

Oldenkott, B., M. Burger, A.-C. Hein, A. Jörg, J. Senkler, H.-P. Braun, V. Knoop, M. Takenaka, and M. Schallenberg-Rüdinger. 2020. One C-to-U RNA Editing Site and Two Independently Evolved Editing Factors: Testing Reciprocal Complementation with DYW-Type PPR Proteins from the Moss Physcomitrium (Physcomitrella) patens and the Flowering Plants Macadamia integrifolia and Arabidopsis. *Plant Cell*. 32:2997-3018 https://doi.org/10.1105/tpc.20.00311

Otrzonsek, J. 2020. Untersuchung des Einflusses modifizierter Erkennungssequenzen und Mutation der konservierten PG-Box auf die C-zu-U-RNA-Editing-Effizienz des Pentatricopeptide Repeat Protein 56 aus Physcomytrium patens im heterologen Escherichia coli System. *In* Institut für Zelluläre und Molekulare Botanik Bonn. Vol. Bachelorarbeit. Rheinische Friedrich-Wilhelms-Universität Bonn.

Paszkowski, J., R.D. Shillito, M. Saul, V. Mandák, T. Hohn, B. Hohn, and I. Potrykus. 1984. Direct gene transfer to plants. *EMBO J.* 3:2717-2722 https://doi.org/10.1002/j.1460-2075.1984.tb02201.x

Pei, Y., and M. Lu. 2019. Programmable RNA manipulation in living cells. *Cellular and Molecular Life Sciences*. 76:4861–4867 https://doi.org/10.1007/s00018-019-03252-9

Porteus, M.H., and D. Carroll. 2005. Gene targeting using zinc finger nucleases. *Nat Biotechnol*. 23:967-973 https://doi.org/10.1038/nbt1125.

Puttick, M.N., J.L. Morris, T.A. Williams, C.J. Cox, D. Edwards, P. Kenrick, S. Pressel, C.H. Wellman, H. Schneider, D. Pisani, and P.C.J. Donoghue. 2018. The Interrelationships of Land Plants and the Nature of the Ancestral Embryophyte. *Curr Biol*. 28:733-745 https://doi.org/10.1016/j.cub.2018.01.063

Qiu, Y.L., L. Li, B. Wang, Z. Chen, V. Knoop, M. Groth-Malonek, O. Dombrovska, J. Lee, L. Kent, J. Rest, G.F. Estabrook, T.A. Hendry, D.W. Taylor, C.M. Testa, M. Ambros, B. Crandall-Stotler, R.J. Duff, M. Stech, W. Frey, D. Quandt, and C.C. Davis. 2006. The deepest divergences in land plants inferred from phylogenomic evidence. *Proc Natl Acad Sci U S A*. 103:15511-15516 https://doi.org/10.1073/pnas.0603335103

Rajasekhar, V.K., and R.M. Mulligan. 1993. RNA Editing in Plant Mitochondria: [alpha]-Phosphate Is Retained during C-to-U Conversion in mRNAs. *Plant Cell*. 5:1843-1852 https://doi.org/10.1105/tpc.5.12.1843

Ramanathan, S. 2021. Evaluating the C-to-U RNA editing efficiencies of PPR43 and other moss DYW-type PPR protein chimeras in the heterologous Escherichia coli system. *In* Faculty of Mathematics and Natural Sciences. Vol. Master of Science. Rheinische Friedrich-Wilhelms-Universität Bonn.

Rensing, S.A., B. Goffinet, R. Meyberg, S.-Z. Wu, and M. Bezanilla. 2020. The Moss Physcomitrium (Physcomitrella) patens: A Model Organism for Non-Seed Plants. *Plant Cell*. 32:1361-1376 https://doi.org/10.1105/tpc.19.00828

Rensing, S.A., D. Lang, A.D. Zimmer, A. Terry, A. Salamov, H. Shapiro, T. Nishiyama, P.F. Perroud, E.A. Lindquist, Y. Kamisugi, T. Tanahashi, K. Sakakibara, T. Fujita, K. Oishi, I.T. Shin, Y. Kuroki, A. Toyoda, Y. Suzuki, S. Hashimoto, K. Yamaguchi, S. Sugano, Y. Kohara, A. Fujiyama, A. Anterola, S. Aoki, N. Ashton, W.B. Barbazuk, E. Barker, J.L. Bennetzen, R. Blankenship, S.H. Cho, S.K. Dutcher, M. Estelle, J.A. Fawcett, H. Gundlach, K. Hanada, A. Heyl, K.A. Hicks, J. Hughes, M. Lohr, K. Mayer, A. Melkozernov, T. Murata, D.R. Nelson, B. Pils, M. Prigge, B. Reiss, T. Renner, S. Rombauts, P.J. Rushton, A. Sanderfoot, G. Schween, S.H. Shiu, K. Stueber, F.L. Theodoulou, H. Tu, Y. Van de Peer, P.J. Verrier, E. Waters, A. Wood, L. Yang, D. Cove, A.C. Cuming, M. Hasebe, S. Lucas, B.D. Mishler, R. Reski, I.V. Grigoriev, R.S. Quatrano, and J.L. Boore. 2008. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science*. 319:64-69 https://doi.org/10.1126/science.1150646

Rice, D.W., A.J. Alverson, A.O. Richardson, G.J. Young, M.V. Sanchez-Puerta, J. Munzinger, K. Barry,

J.L. Boore, Y. Zhang, C.W. dePamphilis, E.B. Knox, and J.D. Palmer. 2013. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm Amborella. *Science*. 342:1468-1473 https://doi.org/10.1126/science.1246275.

Ritzenhofen, K. 2021. Physcomitrium patens C-to-U RNA editing factors PPR56 and PPR65 in planta and in the Escherichia coli system: Investigation of their target specificity and their modifiability towards recognizing new targets. *In* Faculty of Mathematics and Natural Sciences. Vol. Bachelor of Science. Rheinische Friedrich-Wilhelms-Universität Bonn.

Rodriguez, A., J.A. Martnez, N. Flores, A. Escalante, G. Gosset, and F. Bolivar. 2014. Engineering Escherichia coli to overproduce aromatic amino acids and derived compounds. *Microbial Cell Factories*. 13 https://doi.org/10.1186/s12934-014-0126-z

Royan, S., B. Gutmann, C.C.d. Francs-Small, S. Honkanen, J. Schmidberger, A. Soet, Y.K. Sun, L.V.P. Sanglard, C.S. Bond, and I. Small. 2021. A synthetic RNA editing factor edits its target site in chloroplasts and bacteria. *Communications Biology*. 4 https://doi.org/10.1038/s42003-021-02062-9

Rüdinger, M., M. Polsakiewicz, and V. Knoop. 2008. Organellar RNA editing and plant-specific extensions of pentatricopeptide repeat proteins in jungermanniid but not in marchantiid liverworts. *Mol Biol Evol*. 25:1405-1414 https://doi.org/10.1093/molbev/msn084

Rüdinger, M., P. Szovenyi, S.A. Rensing, and V. Knoop. 2011a. Assigning DYW-type PPR proteins to RNA editing sites in the funariid mosses Physcomitrella patens and Funaria hygrometrica. *Plant J*. 67:370-380 https://doi.org/10.1111/j.1365-313X.2011.04600.x

Rüdinger, M., L. Fritz-Laylin, M. Polsakiewicz, and V. Knoop. 2011b. Plant-type mitochondrial RNA editing in the protist Naegleria gruberi. *RNA*. 17:2058-2062 https://doi.org/10.1261/rna.02962911

Rüdinger, M., H.T. Funk, S.A. Rensing, U.G. Maier, and V. Knoop. 2009. RNA editing: only eleven sites are present in the Physcomitrella patens mitochondrial transcriptome and a universal nomenclature proposal. *Mol Genet Genomics*. 281:473-481 https://doi.org/10.1007/s00438-009-0424-z

Rüdinger, M., U. Volkmar, H. Lenz, M. Groth-Malonek, and V. Knoop. 2012. Nuclear DYW-type PPR gene families diversify with increasing RNA editing frequencies in liverwort and moss mitochondria. *J Mol Evol*. 74:37-51 https://doi.org/10.1007/s00239-012-9486-3

Rugen, N., H. Straube, L.E. Franken, H.P. Braun, and H. Eubel. 2019. Complexome profiling reveals association of PPR proteins with ribosomes in the mitochondria of plants. *Mol. Cell. Proteomics*. 18:1345-1362

Rüth, P. 2019. U-zu-C RNA Editing im Hornmoos Anthoceros agrestis: Analyse putativer reverser Editing-Faktoren aus der Pentatricopeptide repeat Protein Familie im heterologen Escherichia coli System. *In* Mathematisch-Naturwissenschaftliche Fakultät. Rheinischen Friedrich-Wilhelms-Universität Bonn, Bachelor of Science.

Ruwe, H., B. Castandet, C. Schmitz-Linneweber, and D.B. Stern. 2013. Arabidopsis chloroplast quantitative editotype. *FEBS Lett*. 587:1429-1433 https://doi.org/10.1016/j.febslet.2013.03.022

Ruwe, H., B. Gutmann, C. Schmitz-Linneweber, I. Small, and P. Kindgren. 2018. The E-domain of CRR2 participates in sequence-specific recognition of RNA in plastids. *New Phytol*

https://doi.org/10.1111/nph.15578

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

Sandoval, R., R.D. Boyd, A.N. Kiszter, Y. Mirzakhanyan, P. Santibańez, P.D. Gershon, and M.L. Hayes. 2019. Stable native RIP9 complexes associate with C-to-U RNA editing activity, PPRs, RIPs, OZ1, ORRM1 and ISE2. *Plant J*. 99:1116-1126 https://doi.org/10.1111/tpj.14384

Sauer, M., D. Porro, D. Mattanovich, and P. Branduardi. 2008. Microbial production of organic acids: expanding the markets. *Trends in Biotechnology*. 26:100-108 https://doi.org/10.1016/j.tibtech.2007.11.006

Schaap, P., I. Barrantes, P. Minx, N. Sasaki, R.W. Anderson, M. Bénard, K.K. Biggar, N.E. Buchler, R. Bundschuh, X. Chen, C. Fronick, L. Fulton, G. Golderer, N. Jahn, V. Knoop, L.F. Landweber, C. Maric, D. Miller, A.A. Noegel, R. Peace, G. Pierron, T. Sasaki, M. Schallenberg-Rüdinger, M. Schleicher, R. Singh, T. Spaller, K.B. Storey, T. Suzuki, C. Tomlinson, J.J. Tyson, W.C. Warren, E.R. Werner, G. Werner-Felmayer, R.K. Wilson, T. Winckler, J.M. Gott, G. Glöckner, and W. Marwan. 2015. The Physarum polycephalum Genome Reveals Extensive Use of Prokaryotic Two-Component and Metazoan-Type Tyrosine Kinase Signaling. *Genome Biol Evol*. 8:109-125 https://doi.org/10.1093/gbe/evv237

Schaefer, D.G., and J.P. Zrÿd. 1997. Efficient gene targeting in the moss Physcomitrella patens. *The plant Journal*. 11:1195-1206 https://doi.org/10.1046/j.1365-313x.1997

Schallenberg-Rüdinger, M., and V. Knoop. 2016. Coevolution of Organelle RNA Editing and Nuclear Specificity Factors in Early Land Plants. *In* Genomes and Evolution of Charophytes, Bryophytes, Lycophytes and Ferns. Vol. 78. S.A. Rensing, editor. Elsevier Ltd. 37–93.9780128011027

Schallenberg-Rüdinger, M., P. Kindgren, A. Zehrmann, I. Small, and V. Knoop. 2013a. A DYW-protein knockout in Physcomitrella affects two closely spaced mitochondrial editing sites and causes a severe developmental phenotype. *Plant J*. 76:420-432 https://doi.org/10.1111/tpj.12304

Schallenberg-Rüdinger, M., H. Lenz, M. Polsakiewicz, J.M. Gott, and V. Knoop. 2013b. A survey of PPR proteins identifies DYW domains like those of land plant RNA editing factors in diverse eukaryotes. *RNA Biol*. 10:1549-1556 https://doi.org/10.4161/rna.25755

Schallenberg-Rüdinger, M., B. Oldenkott, M. Hiss, P.L. Trinh, V. Knoop, and S.A. Rensing. 2017. A Single-Target Mitochondrial RNA Editing Factor of Funaria hygrometrica Can Fully Reconstitute RNA Editing at Two Sites in Physcomitrella patens. *Plant Cell Physiol*. 58:496-507 https://doi.org/10.1093/pcp/pcw229

Shen, C., D. Zhang, Z. Guan, Y. Liu, Z. Yang, Y. Yang, X. Wang, Q. Wang, Q. Zhang, S. Fan, T. Zou, and P. Yin. 2016. Structural basis for specific single-stranded RNA recognition by designer pentatricopeptide repeat proteins. *Nat Commun*. 7:11285 https://doi.org/10.1038/ncomms11285

Shi, X., S. Bentolila, and M.R. Hanson. 2016. Organelle RNA recognition motif-containing (ORRM) proteins are plastid and mitochondrial editing factors in Arabidopsis. *Plant Signal Behav*. 11:e1167299 https://doi.org/10.1080/15592324.2016.1167299

Simpson, M.G. 2010. Plant Systematics. Academic Press.978-0-12-374380-0

Small, I., J. Melonek, A.-V. Bohne, J. Nickelsen, and C. Schmitz-Linneweber. 2023. Plant Organellar RNA Maturation. *The Plant Cell* https://doi.org/10.1093/plcell/koad049

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

Small, I.D., M. Schallenberg-Rüdinger, M. Takenaka, H. Mireau, and O. Ostersetzer-Biran. 2020. Plant organellar RNA editing: what 30 years of research has revealed. *The Plant journal: for cell and moelecular biology*. 101:1040-1056 https://doi.org/10.1111/tpj.14578

Smith, H.C., J.M. Gott, and M.R. Hanson. 1997. A guide to RNA editing. *RNA*. 3:1105-1123

Sosso, D., S. Mbelo, V. Vernoud, G. Gendrot, A. Dedieu, P. Chambrier, M. Dauzat, L. Heurtevin, V. Guyon, M. Takenaka, and P.M. Rogowsky. 2012. PPR2263, a DYW-Subgroup Pentatricopeptide repeat protein, is required for mitochondrial nad5 and cob transcript editing, mitochondrion biogenesis, and maize growth. *Plant Cell*. 24:676-691 https://doi.org/10.1105/tpc.111.091074

Sper-Whitis, G.L., J. L.Moody, and J. C.Vaughn. 1996. Universality of mitochondrial RNA editing in cytochrome-c oxidase subunit I (coxI) among the land plants. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*. 1307:301-308

Steinhauser, S., S. Beckert, I. Capesius, O. Malek, and V. Knoop. 1999. Plant mitochondrial RNA editing. *J Mol Evol*. 48:303-312

Su, A., and L. Randau. 2011. A-to-I and C-to-U Editing within Transfer RNAs. *Biochemistry (Moscow)*. Vol. 76, No. 8:932-937

Su, D., L. Yang, X. Shi, X. Ma, X. Zhou, S.B. Hedges, and B. Zhong. 2021. Large-Scale Phylogenomic Analyses Reveal the Monophyly of Bryophytes and Neoproterozoic Origin of Land Plants. *Molecular Biology and Evolution*. 38:3332–3344 https://doi.org/10.1093/molbev/msab106

Sugita, M. 2021. Plastid Transformation in Physcomitrium (Physcomitrella) patens: An Update. *Methods Mol Biol*. 2317:321-331 https://doi.org/10.1007/978-1-0716-1472-3_19

Sugita, M. 2022. An Overview of Pentatricopeptide Repeat (PPR) Proteins in the Moss Physcomitrium patens and Their Role in Organellar Gene Expression. *Plants*. 11 https://doi.org/10.3390/plants11172279

Sugiura, C., and M. Sugita. 2004. Plastid transformation reveals that moss tRNA(Arg)-CCG is not essential for plastid function. *Plant J*. 40:314-321 https://doi.org/10.1111/j.1365-313X.2004.02202.x

Sugiura, C., Y. Kobayashi, S. Aoki, C. Sugita, and M. Sugita. 2003. Complete chloroplast DNA sequence of the moss Physcomitrella patens: evidence for the loss and relocation of rpoA from the chloroplast to the nucleus. *Nucleic Acids Res*. 31:5324-5331

Sun, T., S. Bentolila, and M.R. Hanson. 2016. The Unexpected Diversity of Plant Organelle RNA Editosomes. *Trends Plant Sci*. 21:962-973 https://doi.org/10.1016/j.tplants.2016.07.005

Sun, T., X. Shi, G. Friso, K. Van Wijk, S. Bentolila, and M.R. Hanson. 2015. A zinc finger motif-containing protein is essential for chloroplast RNA editing. *PLoS Genet*. 11:e1005028 https://doi.org/10.1371/journal.pgen.1005028

Szövényi, P., E. Frangedakis, M. Ricca, D. Quandt, S. Wicke, and J.A. Langdale. 2015. Establishment of Anthoceros agrestis as a model species for studying the biology of hornworts. *BMC Plant Biology*. 15 https://doi.org/10.1186/s12870-015-0481-x

Takahashi, A., and T. Ohnishi. 2004. The Significance of the Study about the Biological Effects of Solar Ultraviolet Radiation using the Exposed Facility on the International Space Station. *Biological Sciences in Space*. 18:255-260 https://doi.org/10.2187/bss.18.255

Takenaka, M., A. Zehrmann, A. Brennicke, and K. Graichen. 2013. Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. *PLoS One*. 8:e65343 https://doi.org/10.1371/journal.pone.0065343

Takenaka, M., A. Jörg, M. Burger, and S. Haag. 2019. RNA editing mutants as surrogates for mitochondrial SNP mutants. *Plant Physiol Biochem*. 135:310-321 https://doi.org/10.1016/j.plaphy.2018.12.014.

Takenaka, M., A. Zehrmann, D. Verbitskiy, M. Kugelmann, B. Hartel, and A. Brennicke. 2012. Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. *Proc Natl Acad Sci U S A*. 109:5104-5109 https://doi.org/10.1073/pnas.1202452109

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

Tasaki, E., and M. Sugita. 2010. The moss Physcomitrella patens, a model plant for the study of RNA editing in plant organelles. *Plant Signal Behav*. 5:727-729 https://doi.org/10.4161/psb.5.6.11664

Tasaki, E., M. Hattori, and M. Sugita. 2010. The moss pentatricopeptide repeat protein with a DYW domain is responsible for RNA editing of mitochondrial ccmFc transcript. *Plant J*. 62:560-570 https://doi.org/10.1111/j.1365-313X.2010.04175.x

Teichert, I. 2018. Adenosine to inosine mRNA editing in fungi and how it may relate to fungal pathogenesis. *PLoS Pathog*. 14 https://doi.org/10.1371/journal.ppat.1007231

Terasawa, K., M. Odahara, Y. Kabeya, T. Kikugawa, Y. Sekine, M. Fujiwara, and N. Sato. 2007. The mitochondrial genome of the moss Physcomitrella patens sheds new light on mitochondrial evolution in land plants. *Mol Biol Evol*. 24:699-709 https://doi.org/10.1093/molbev/msl198

Thümmler, F., H. Schuster, and J. Bonenberger. 1992. Expression of oat phyA cDNA in the moss Ceratodon purpureus. *Photochem Photobiol*. 56:771-776 https://doi.org/10.1111/j.1751-1097.1992.tb02233.x.

Tillich, M., S.L. Hardel, C. Kupsch, U. Armbruster, E. Delannoy, J.M. Gualberto, P. Lehwark, D. Leister, I.D. Small, and C. Schmitz-Linneweber. 2009. Chloroplast ribonucleoprotein CP31A is required for editing and stability of specific chloroplast mRNAs. *Proc Natl Acad Sci U S A*. 106:6002-6007 https://doi.org/10.1073/pnas.0808529106

Trouiller, B., F. Charlot, S. Choinard, D.G. Schaefer, and F. Nogué. 2007. Comparison of gene targeting efficiencies in two mosses suggests that it is a conserved feature of Bryophyte transformation. *Biotechnol Lett*. 29:1591-1598 https://doi.org/10.1007/s10529-007-9423-

5

Uchida, M., S. Ohtani, M. Ichinose, C. Sugita, and M. Sugita. 2011. The PPR-DYW proteins are required for RNA editing of rps14, cox1 and nad5 transcripts in Physcomitrella patens mitochondria. *FEBS Lett*. 585:2367-2371 https://doi.org/10.1016/j.febslet.2011.06.009

Verbitskiy, D., A. Zehrmann, A. Brennicke, and M. Takenaka. 2010. A truncated MEF11 protein shows site-specific effects on mitochondrial RNA editing. *Plant Signal Behav*. 5:558-560 https://doi.org/10.4161/psb.11095.

Verbitskiy, D., J.A. van der Merwe, A. Zehrmann, A. Brennicke, and M. Takenaka. 2008. Multiple specificity recognition motifs enhance plant mitochondrial RNA editing in vitro. *J Biol Chem*. 283:24374-24381 https://doi.org/10.1074/jbc.M803292200

Verma, D., and H. Daniell. 2007. Chloroplast vector systems for biotechnology applications. *Plant Physiol*. 145:1129-1143 https://doi.org/10.1104/pp.107.106690.

Wagner, R.W., J.E. Smith, B.S. Cooperman, and K. Nishikura. 1989. A double-stranded RNA unwinding activity introduces structural alterations by means of adenosine to inosine conversions in mammalian cells and Xenopus eggs. *Proc Natl Acad Sci U S A*. 86:2647-2651

Wagoner, J.A., T. Sun, L. Lin, and M.R. Hanson. 2015. Cytidine deaminase motifs within the DYW domain of two pentatricopeptide repeat-containing proteins are required for site-specific chloroplast RNA editing. *J. Biol. Chem.* 290:2957-2968

Waltz, F., T.-T. Nguyen, M. Arrivé, A. Bochler, J. Chicher, P. Hammann, L. Kuhn, M. Quadrado, H. Mireau, Y. Hashem, and P. Giegé. 2019. Small is big in Arabidopsis mitochondrial ribosome. *Nature Plants*. 5:106-117 https://doi.org/10.1038/s41477-018-0339-y

Wang, H.-H., W.-B. Yin, and Z.-M. Hu. 2009. Advances in chloroplast engineering. *J Genet Genomics*. 36:387-398 https://doi.org/10.1016/S1673-8527(08)60128-9.

Wang, H., M.L. Russa, and L.S. Qi. 2016. CRISPR/Cas9 in Genome Editing and Beyond. *Annu Rev Biochem*. 85:227-264 https://doi.org/10.1146/annurev-biochem-060815-014607

Wang, X., J. McLachlan, P.D. Zamore, and T.M.T. Hall. 2002. Modular recognition of RNA by a human pumilio-homology domain. *Cell*. 110:501-512 https://doi.org/10.1016/s0092-8674(02)00873-5.

Wang, Y., H. Li, Z.-Q. Huang, B. Ma, Y.-Z. Yang, Z.-H. Xiu, L. Wang, and B.-C. Tan. 2022. Maize PPR-E proteins mediate RNA C-to-U editing in mitochondria by recruiting the trans deaminase PCW1. *The Plant Cell*. 35:529-551 https://doi.org/10.1093/plcell/koac298

Wickett, N.J., S. Mirarab, N. Nguyen, ..., and J. Leebens-Mack. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc Natl Acad Sci U S A*. 111:4859-4868 https://doi.org/10.1073/pnas.1323926111

Willerscheidt, G.M. 2022. Plant-type C-U RNA editing in Escherichia coli: Investigating the RNA editing functionality and target specificity of modified DYW-type PPR proteins of Physcomitrium patens. *In* Faculty of Mathematics and Natural Sciences. Vol. Bachelor of Science. Rheinische Friedrich-Wilhelms-Universität Bonn.

Wolf, J., A.P. Gerber, and W. Keller. 2002. tadA, an essential tRNA-specific adenosine deaminase from Escherichia coli. *EMBO J*. 21:3841-3851 https://doi.org/10.1093/emboj/cdf362

Yagi, Y., S. Hayashi, K. Kobayashi, T. Hirayama, and T. Nakamura. 2013. Elucidation of the RNA

recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS One*. 8:e57286 https://doi.org/10.1371/journal.pone.0057286

Yan, J., Q. Zhang, and P. Yin. 2017a. RNA editing machinery in plant organelles. *Science China Life Sciences*. 61:162-169 https://doi.org/10.1007/s11427-017-9170-3

Yan, J., Q. Zhang, Z. Guan, Q. Wang, L. Li, F. Ruan, R. Lin, T. Zou, and P. Yin. 2017b. MORF9 increases the RNA-binding activity of PLS-type pentatricopeptide repeat protein in plastid RNA editing. *Nat Plants*. 3:17037 https://doi.org/10.1038/nplants.2017.37

Yan, J., Y. Yao, S. Hong, Y. Yang, C. Shen, Q. Zhang, D. Zhang, T. Zou, and P. Yin. 2019. Delineation of pentatricopeptide repeat codes for target RNA prediction. *Nucleic Acids Res*. 47:3728-3738 https://doi.org/10.1093/nar/gkz075

Yang, C., D. Carlow, R. Wolfenden, and S.A. Short. 1992. Cloning and nucleotide sequence of the Escherichia coli cytidine deaminase (ccd) gene. *Biochemistry*. 31:4168-4174 https://doi.org/10.1021/bi00132a003,

Yang, Y. 2019. Investigating plant-type C-to-U RNA editing in Escherichia coli. *In* Faculty of Mathematics and Natural Sciences. Vol. Master of Science. Rheinische Friedrich-Wilhelms-Universität Bonn.

Yap, A., P. Kindgren, C.C.d. Francs-Small, T. Kazama, S.K. Tanz, K. Toriyama, and I. Small. 2015. AEF1/MPR25 is implicated in RNA editing of plastid atpF and mitochondrial nad5, and also promotes atpF splicing in Arabidopsis and rice. *Plant J*. 81:661-669 https://doi.org/10.1111/tpj.12756.

Yin, P., Q. Li, C. Yan, Y. Liu, J. Liu, F. Yu, Z. Wang, J. Long, J. He, H.-W. Wang, J. Wang, J.-K. Zhu, Y. Shi, and N. Yan. 2013. Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature*. 504:168-171 https://doi.org/10.1038/nature12651

Yoo, B.-C., N.S. Yadav, E.M.O. Jr, and H. Sakai. 2020. Cas9/gRNA-mediated genome editing of yeast mitochondria and Chlamydomonas chloroplasts. *PeerJ* https://doi.org/10.7717/peerj.8362

Yu, W., and W. Schuster. 1995. Evidence for a site-specific cytidine deamination reaction involved in C to U RNA editing of plant mitochondria. *J Biol Chem*. 270:18227-18233

Zehrmann, A., D. Verbitskiy, J.A. van der Merwe, A. Brennicke, and M. Takenaka. 2009. A DYW domain-containing pentatricopeptide repeat protein is required for RNA editing at multiple sites in mitochondria of Arabidopsis thaliana. *Plant Cell*. 21:558-567 https://doi.org/10.1105/tpc.108.064535

Zehrmann, A., D. Verbitskiy, B. Hartel, A. Brennicke, and M. Takenaka. 2011. PPR proteins network as site-specific RNA editing factors in plant organelles. *RNA Biol*. 8:67-70

Zehrmann, A., B. Härtel, F. Glass, E. Bayer-Császár, T. Obata, E. Meyer, A. Brennicke, and M. Takenaka. 2015. Selective homo- and heteromer interactions between the multiple organellar RNA editing factor (MORF) proteins in Arabidopsis thaliana. *J Biol Chem*. 290:6445-6456 https://doi.org/10.1074/jbc.M114.602086

Zhang, H.-D., Y.-L. Cui, C. Huang, Q.-Q. Yin, X.-M. Qin, T. Xu, X.-F. He, Y. Zhang, Z.-R. Li, and Z.-N. Yang. 2015. PPR protein PDM1/SEL1 is involved in RNA editing and splicing of plastid genes in Arabidopsis thaliana. *Photosynth Res*. 126:311-321 https://doi.org/10.1007/s11120-015-0171-4

Zhang, Y., U. Werling, and W. Edelmann. 2014. Seamless Ligation Cloning Extract (SLiCE) cloning method. *Methods Mol Biol*. 1116:235-244 https://doi.org/10.1007/978-1-62703-764-8_16

Zhao, D., S. Yuan, B. Xiong, H. Sun, L. Ye, J. Li, X. Zhang, and C. Bi. 2016. Development of a fast and easy method for Escherichia coli genome editing with CRISPR/Cas9. *Microbial Cell Factories*. 15 https://doi.org/10.1186/s12934-016-0605-5

Zhao, Y.-Y., M.-W. Mao, W.-J. Zhang, J. Wang, H.-T. Li, Y. Yang, Z. Wang, and J.-W. Wu. 2018. Expanding RNA binding specificity and affinity of engineered PUF domains. *Nucleic Acids Res*. 46:4771-4782 https://doi.org/10.1093/nar/gky134.

Zumkeller, S., and V. Knoop. 2023. Categorizing 161 plant (streptophyte) mitochondrial group II introns into 29 families of related paralogues finds only limited links between intron mobility and intron-borne maturases. *BMC Ecol Evol*. 23:5 https://doi.org/10.1186/s12862-023-02108-y