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

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

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

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

242 Overall, RNA editing factors characteristically show less conservation at the 5'-end of their PLS-
243 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

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

Total RNA sequencing and off-target detection

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

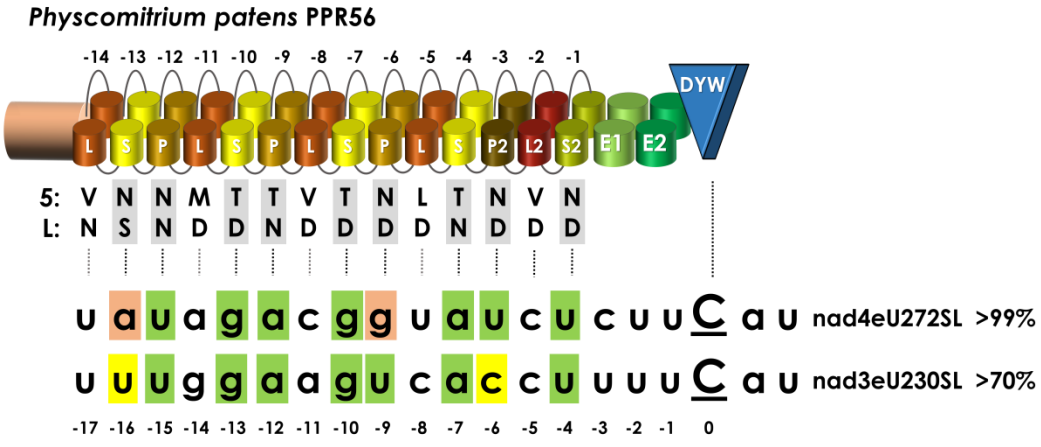
629 editing efficiency was calculated by adding up total RNA reads from all hitting datasets at a site. RNA
630 editing sites were only considered for sites with (i) RNA read coverage of at least 30, (ii) a clear signal
631 for transition in the RNA reads (T+C or G+A > 99%), (iii) a clear DNA reference 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

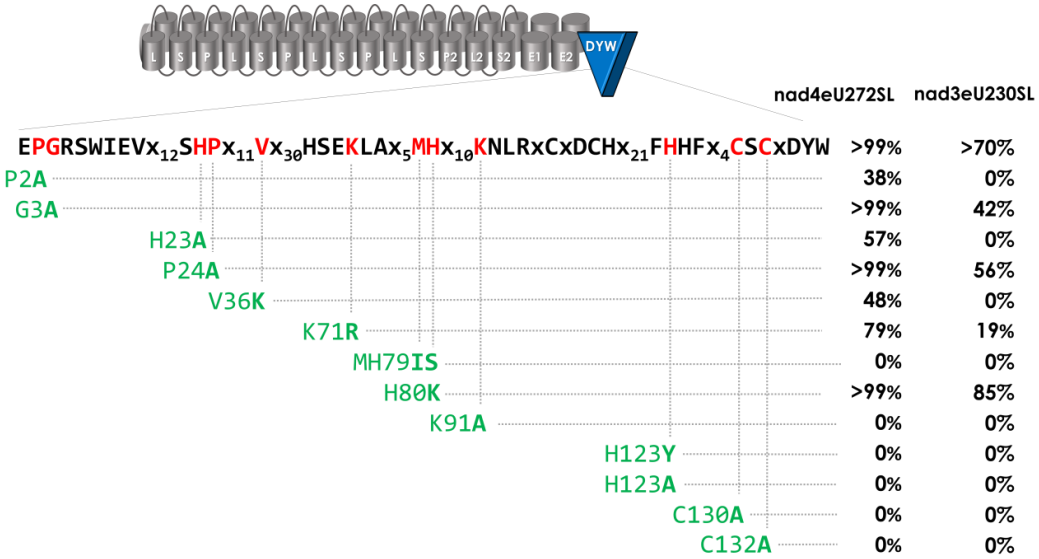
Figure 1. PPR56 and site-directed mutations in its DYW cytidine deaminase domain.

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.



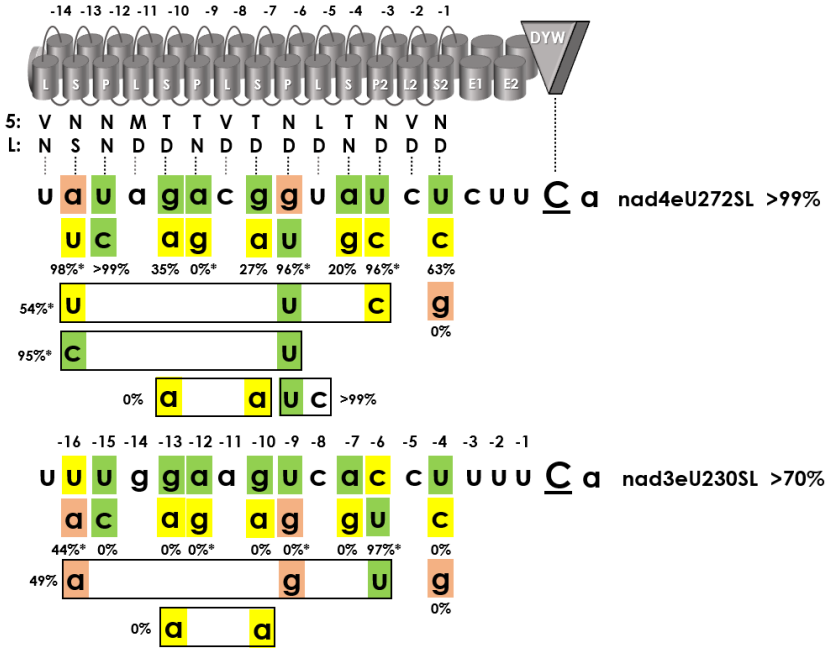
656 ***B.*** *Mutations in the DYW domain of PPR56.*



657

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.

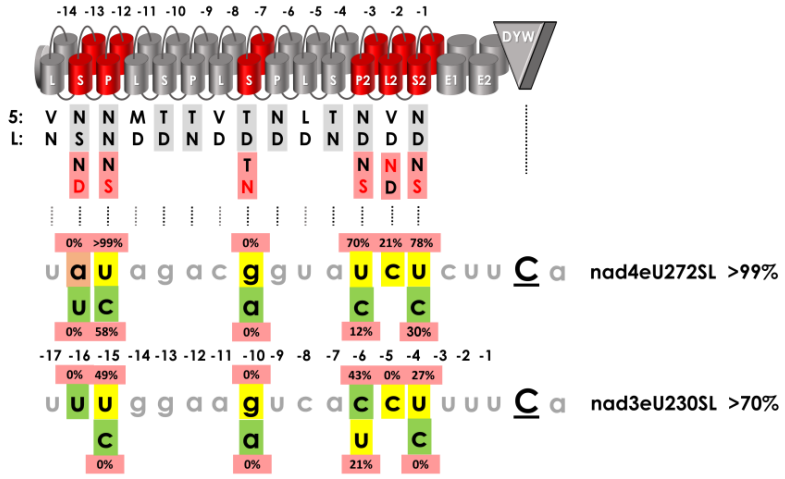


666

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

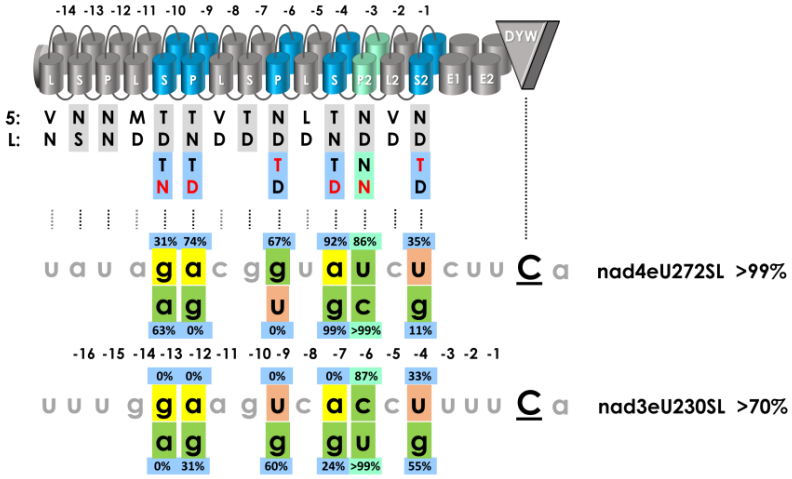
668 Key positions '5' or 'Last' have been altered in individual PPRs of PPR56 (red font) in attempts of re-
669 targeting to modified target sequences with conceptually improved matches in individual positions
670 (green shading) of native targets nad4eU272SL and nad3eU230SL, respectively. RNA editing 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.*



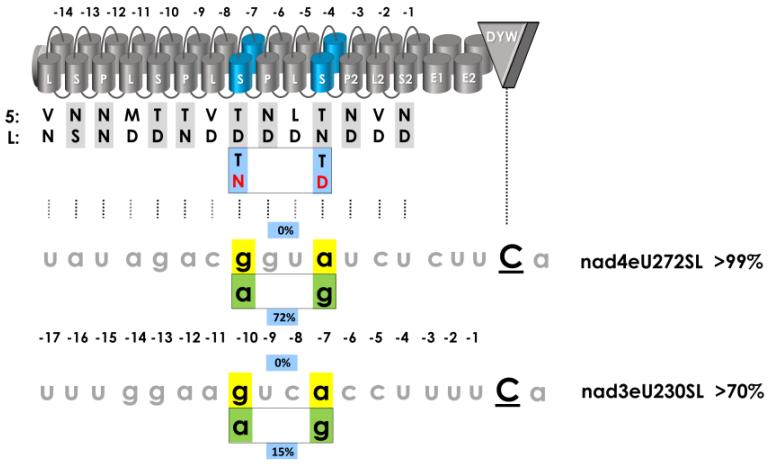
684

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



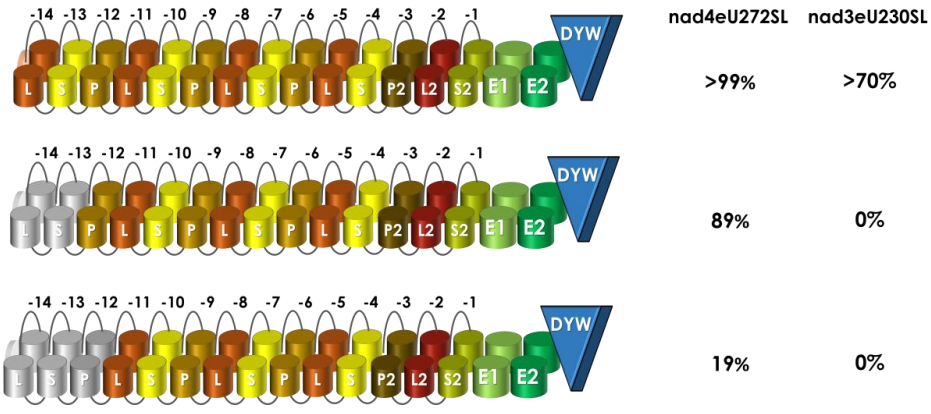
686

687 *C. Double target mutant.*



688

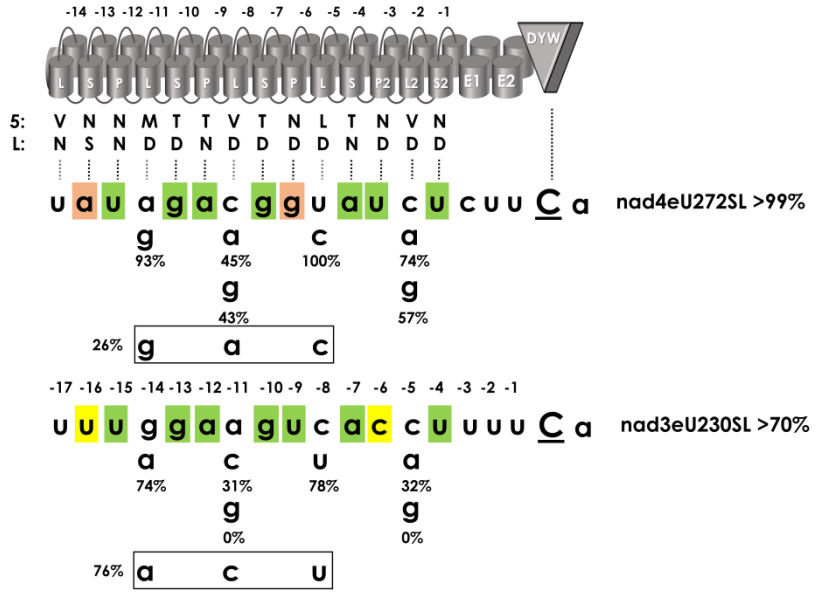
689 ***D.*** *N-terminal PPR truncations of PPR56.*



690

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

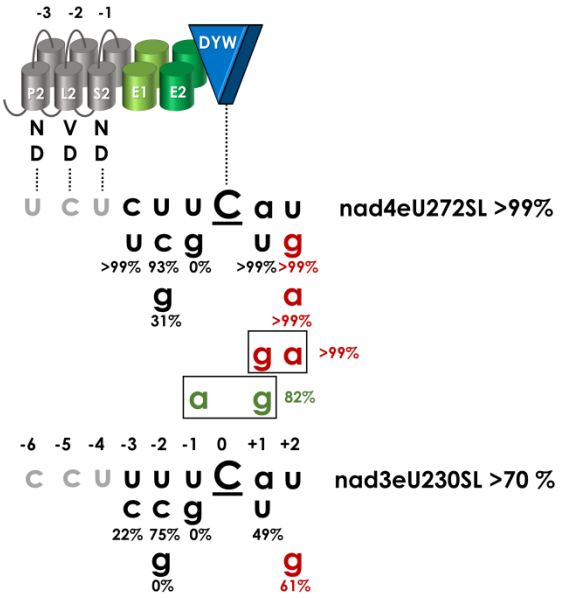
692 Target positions -14, -11 and -8 opposite of L-type PPRs L-11MD, L-8VD and L-5LD have been changed
693 to the nucleotides present in the respective other native target of PPR56. Additional mutations to
694 purines were introduced in positions -11 and -5 opposite of PPRs L-8VD and L2-2VD, which carry 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

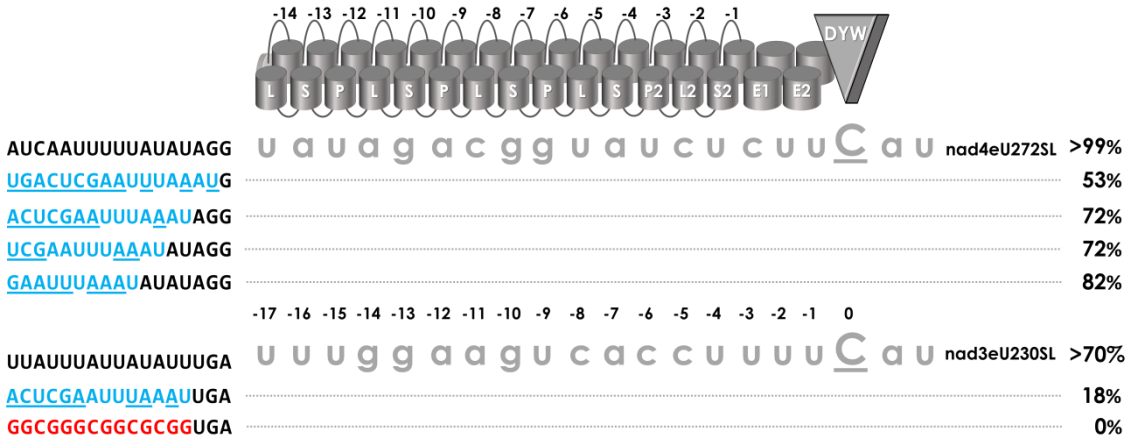
Figure 5. Mutations around the RNA editing sites.

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.

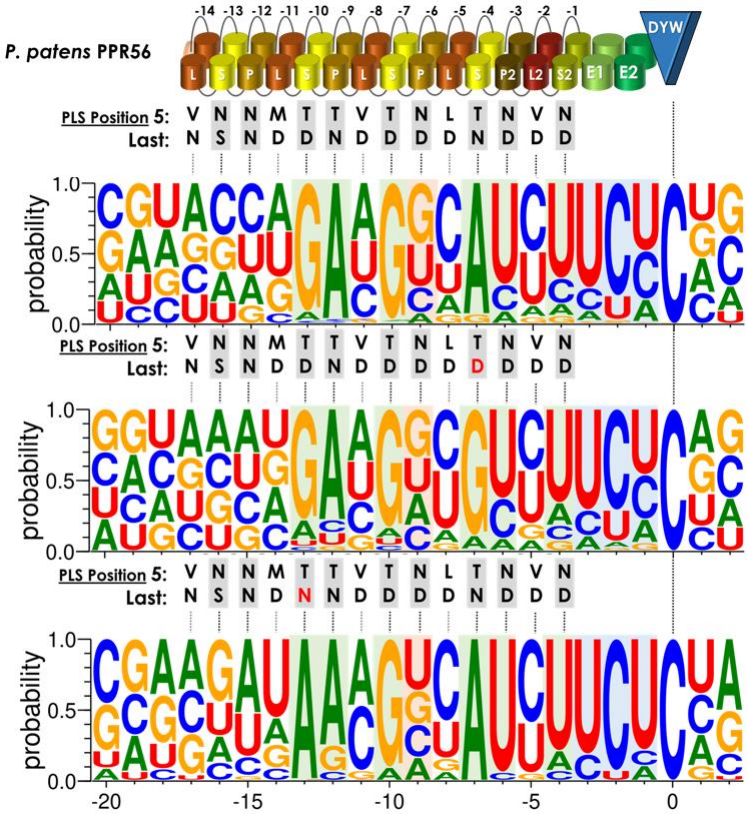
714 PPR56 editing targets were cloned with 17 bp of additional native sequence upstream of the region
715 supposed to be ultimately targeted by the PPR array, with the C-terminal PPR S2-1 juxtaposed with
716 position -4 upstream of the editing site. Progressive 5'-truncations of this upstream sequence 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

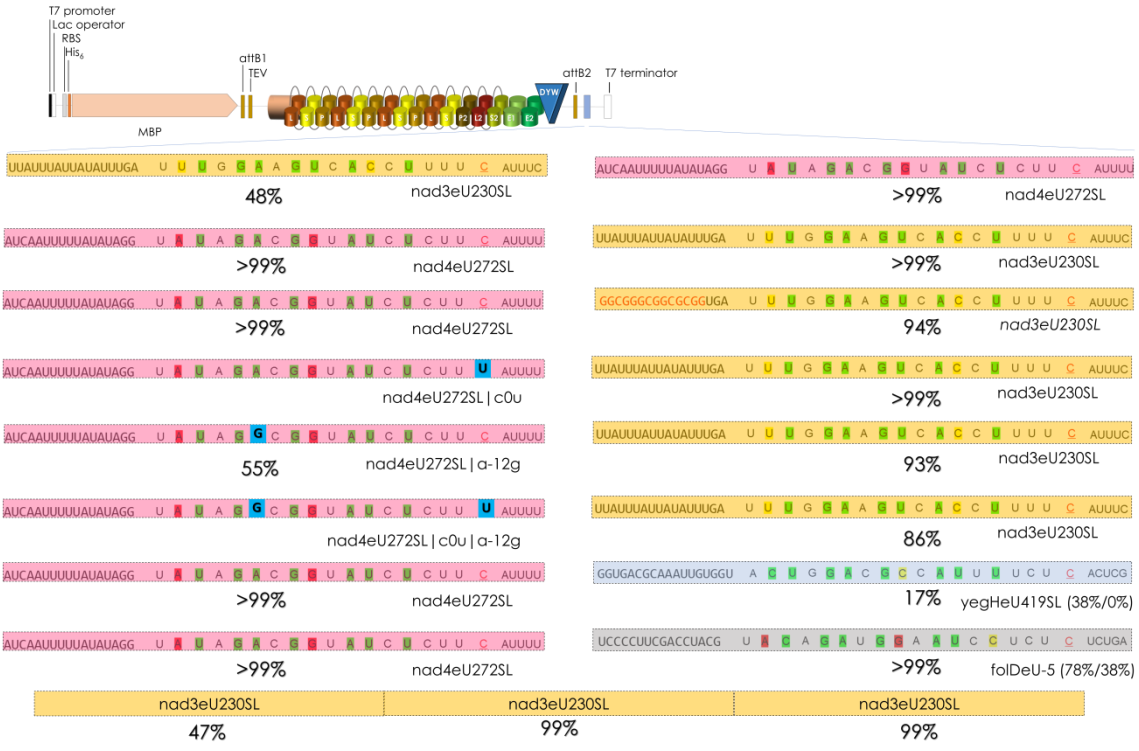
Figure 7. Off-target analyses.

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.

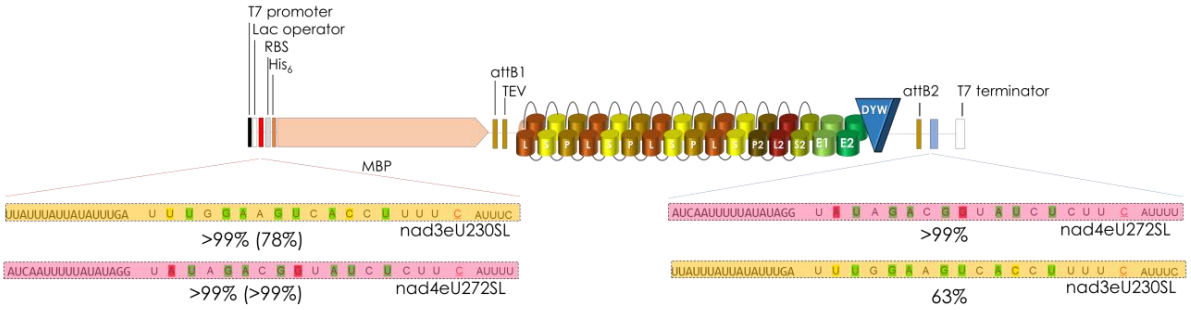


746

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

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

A.



B.

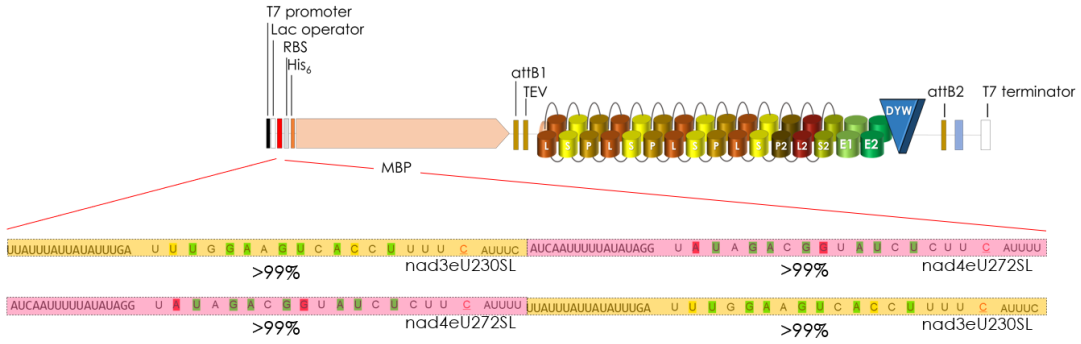
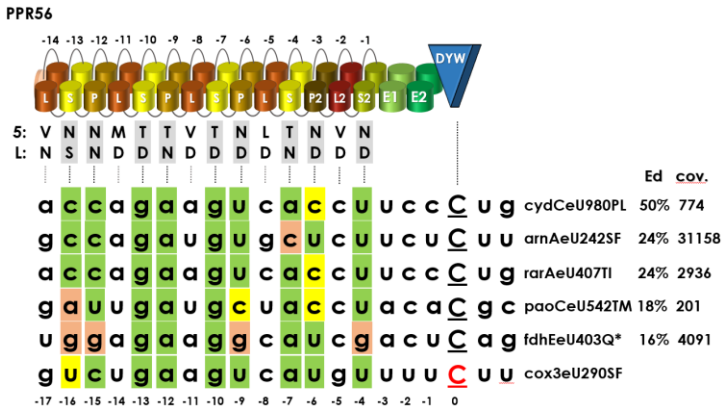


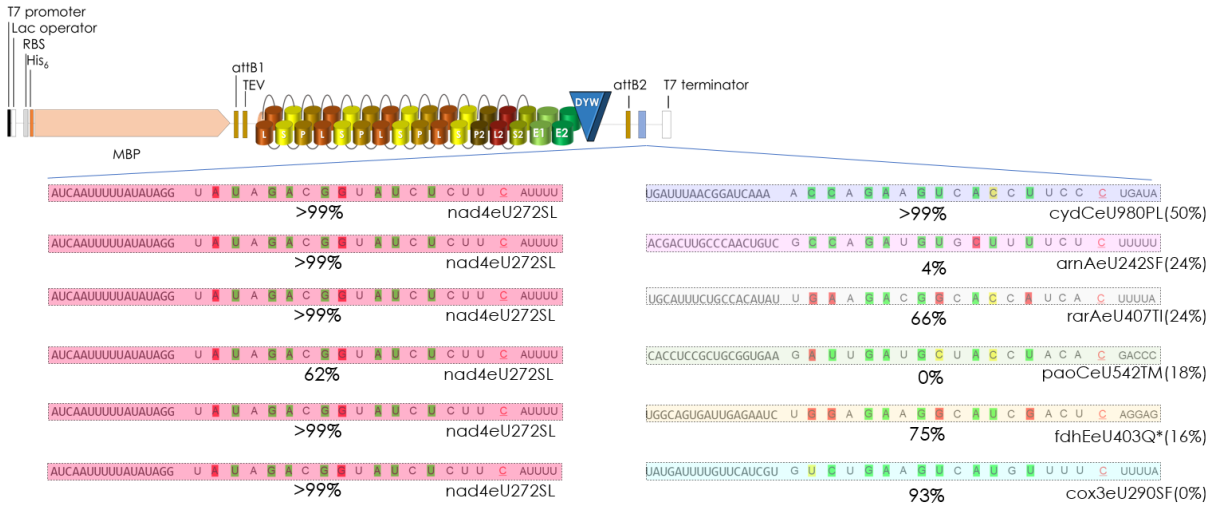
Figure 10. Off-targets in different cloning positions.

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).

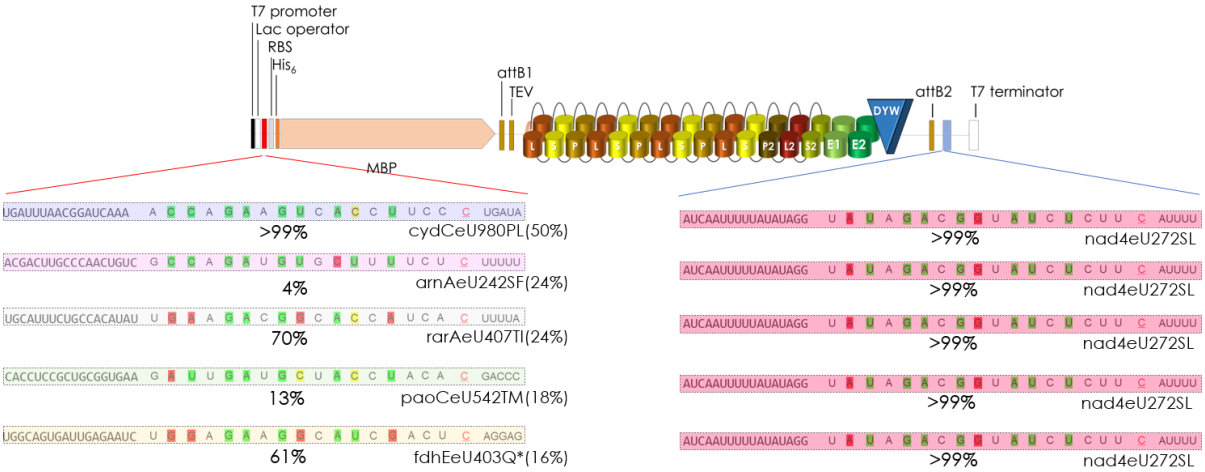
A.



B.



769 *C.*

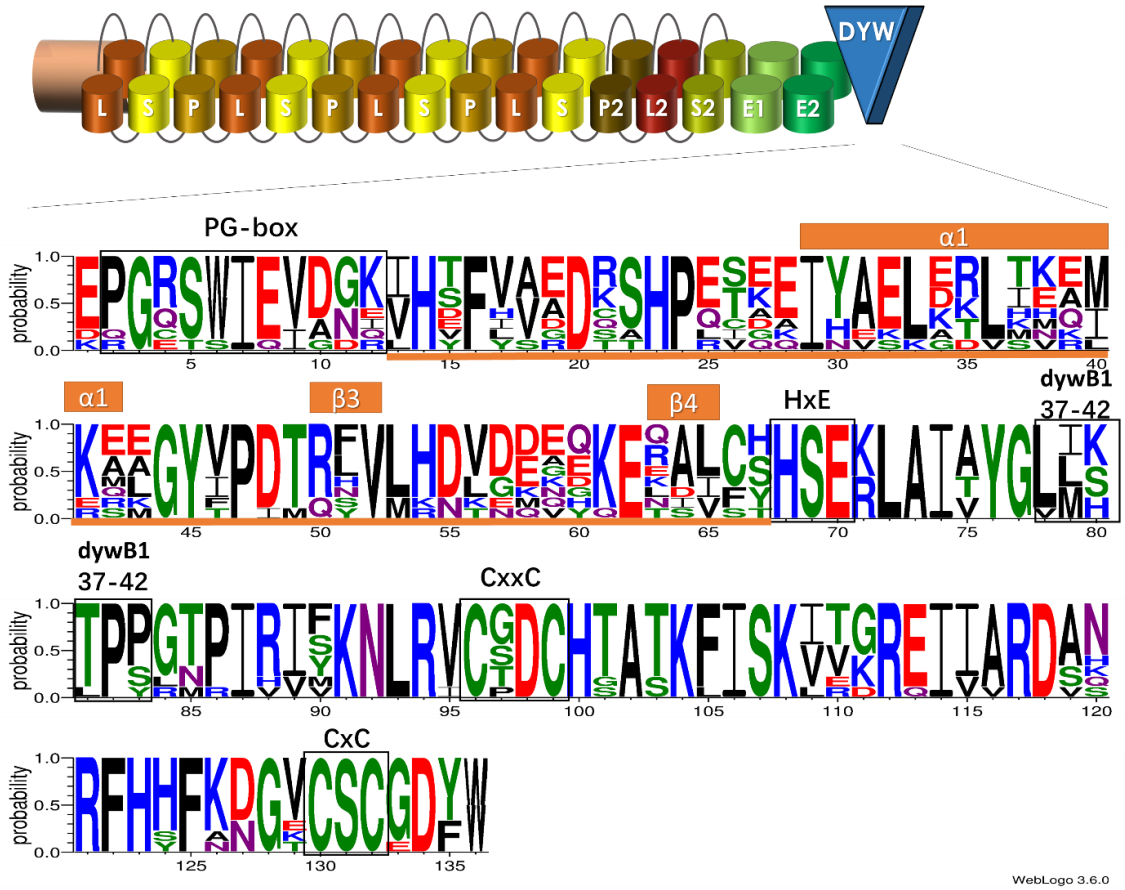


770

Supplementary information

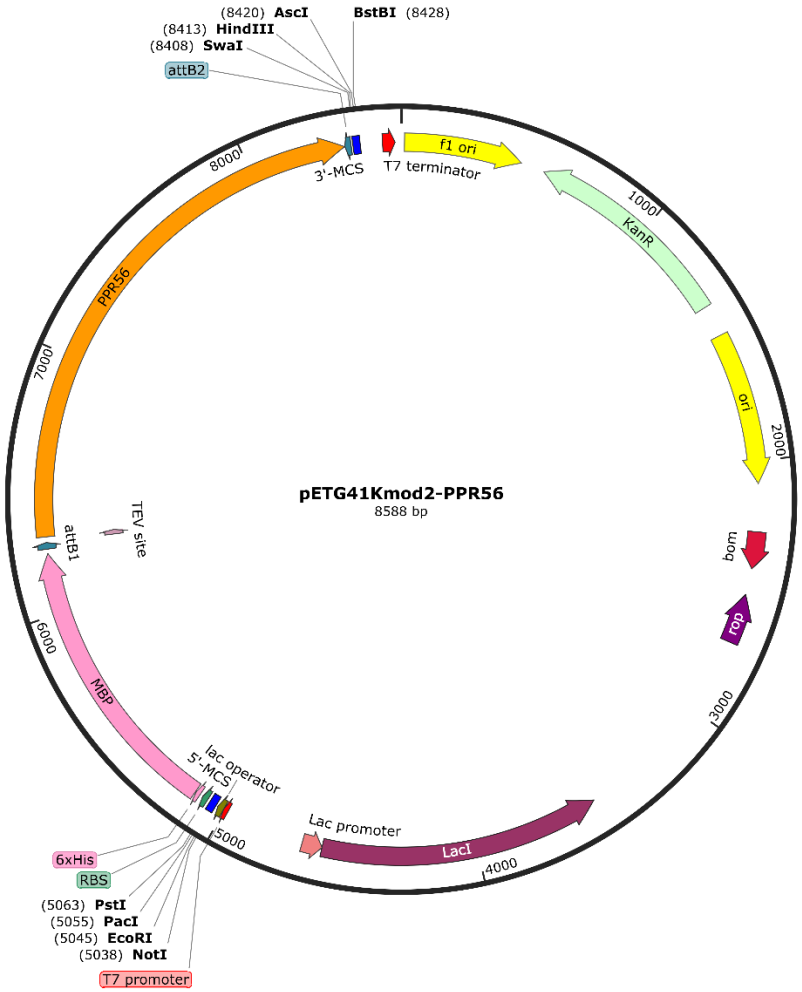
Supplementary Figure 1. WebLogo conservation profile of the DYW domains in nine *Physcomitrium patens* RNA editing factors.

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).



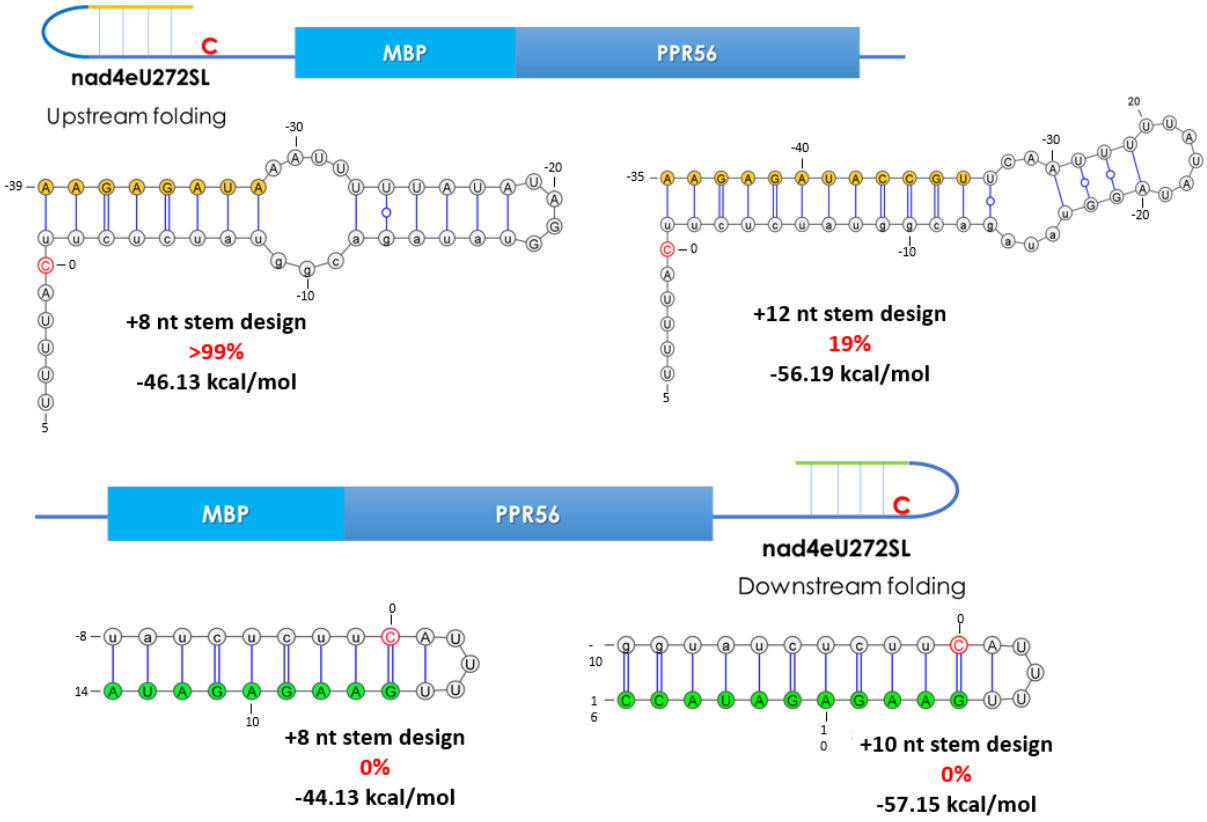
Supplementary Figure 2. Expression vector system pet41Kmod2.

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).

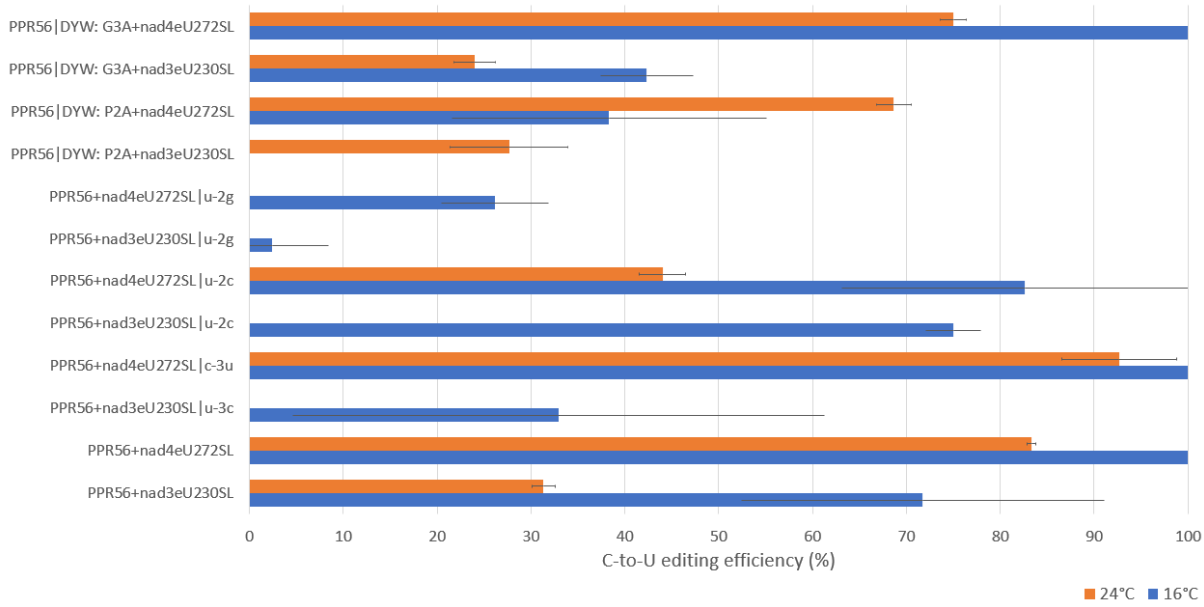


803

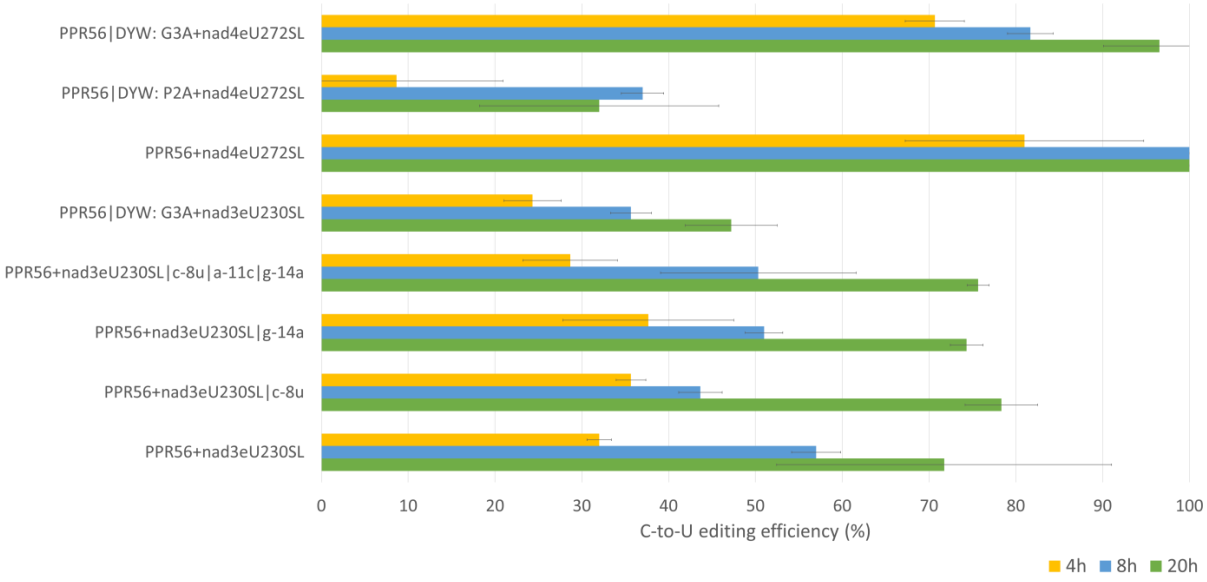
Supplementary Figure 4. Temperature- and time-dependence of RNA editing.

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

A. Temperature dependence of RNA editing in Escherichia coli.



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

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

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

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

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

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

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

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

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

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

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

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

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

73. Hein A, Knoop V. 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 Evol Biol. 2018;18: 85. doi:10.1186/s12862-018-1203-4

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

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