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

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

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

[2] Department of Botany Graduate School of Science, Kyoto University, Oiwake-cho, Sakyo-ku, Kyoto, 606-8502, Japan

*Corresponding authors

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

volker.knoop@uni-bonn.de, Phone: +49 228 73-6466

ORCID-IDs

Yingying Yang: 0000-0002-5518-0307

Bastian Oldenkott: 0000-0002-2513-2443

Shyam Ramanathan: 0009-0000-7582-9452

Elena Lesch: 0000-0002-3449-3929

Mizuki Takenaka: 0000-0002-3242-5092

Mareike Schallenberg-Rüdinger: 0000-0002-6874-4722

Volker Knoop: 0000-0002-8485-9423

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

Abstract

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

Introduction

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

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

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

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

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

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

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

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

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

Results

Recombining RNA editing factors PPR56 and PPR65

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

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

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

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

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

Chimeras of PPR56 and PPR45

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

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

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

Chimeras of PPR56 and OTP86

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

Functional chimeras also upon heterologous expression in human cells

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

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

Off-targets for PPR56, PPR65 and a chimeric construct

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

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

Many more off-targets for the PPR56-OTP86 chimera

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

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

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

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

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

Discussion

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

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

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

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

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

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

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

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

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

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

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

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

Materials and Methods

Molecular Cloning

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

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

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 [11]. 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. Expression of chimera protein variants was routinely checked via SDS-PAGE gels. Expression of chimeric constructs in human IMR-90 cell cytosol was executed as described previously [13]. Briefly, MEM (Pan Biotechnologies) media, supplemented with 10% fetal calf serum and 1% Penicillin/Streptomycin was supplemented with 25 µM zinc sulfate prior to PEI MAX (Polyscience) transfection and 20 h incubation. To verify expression, cells were fixed on cover slips, nuclei stained with DAPI (4,6-diamidino-2-phenylindole). The localization of EYFP-tagged PPR proteins was examined on a Zeiss AXIOPHOT microscope using AxioVision software and ImageJ/Fiji version 1.53c for Windows.

Total RNA sequencing and off-target detection

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

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

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

Acknowledgements

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

Funding

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

Author contributions

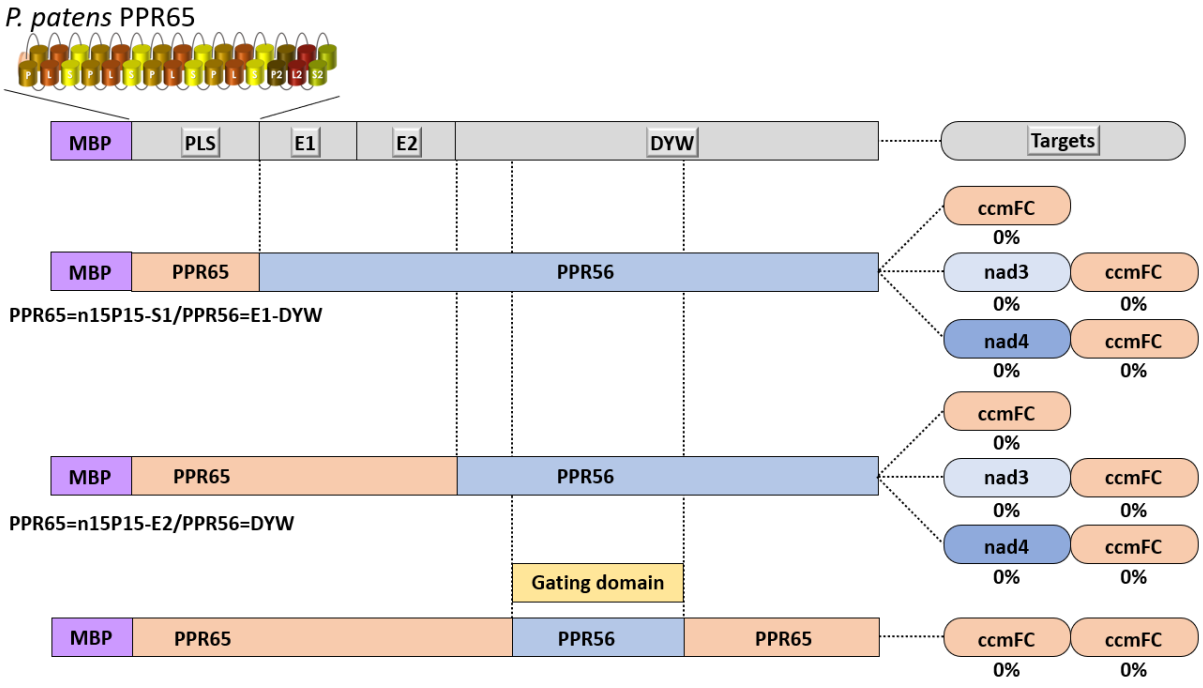
Y.Y. cloned most of the constructs and did most of the RNA editing analyses in the heterologous bacterial system and analyzed RNA-seq data. B.O. and S.R. contributed to construct cloning and RNA editing assays in *E. coli*. E.L. cloned chimeric RNA editing constructs for heterologous expression in human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

Figure Legends

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

A. PPR65 chimeras with DYW regions of PPR56.

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



B. PPR56 chimeras with DYW regions of PPR65.

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

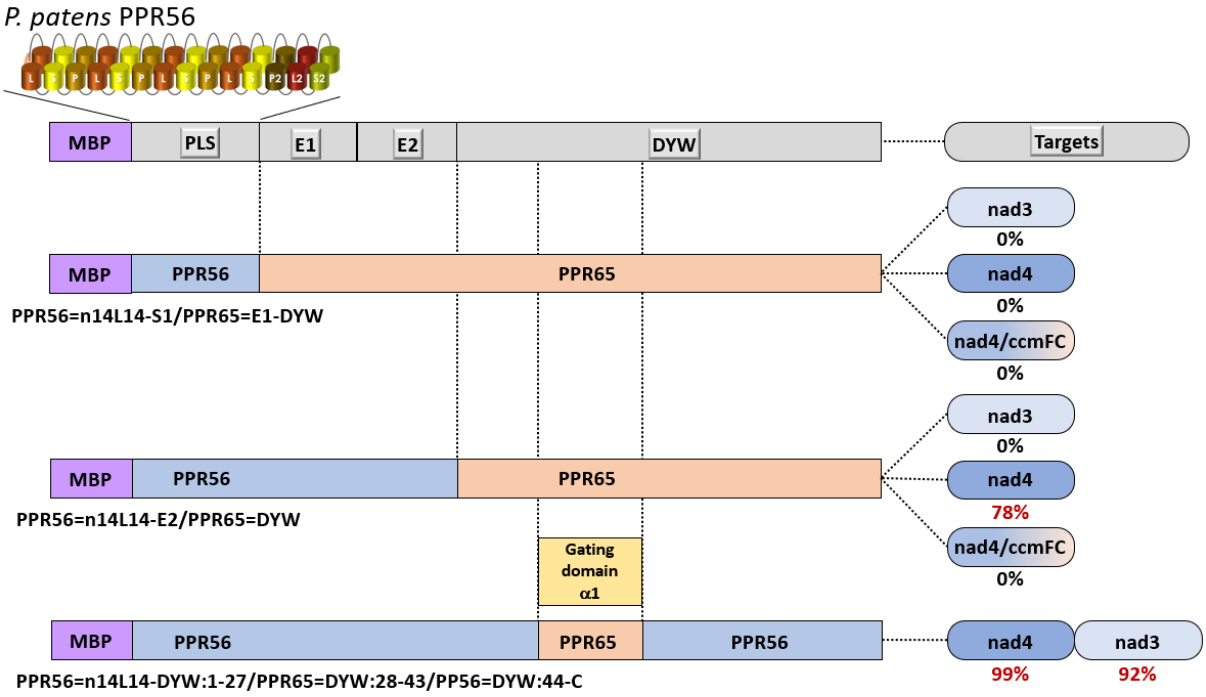


Figure 2. Chimeras of PPR56 and PPR45

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

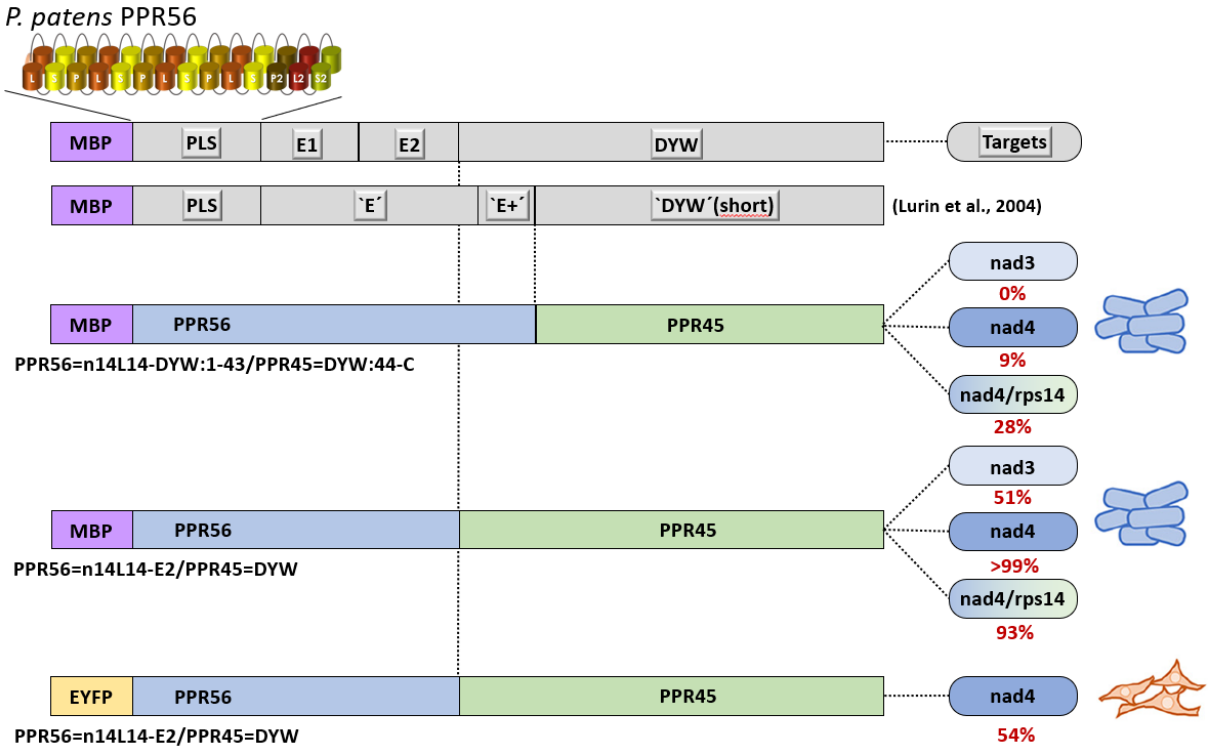


Figure 3. Chimeras of PPR56 and OTP86

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

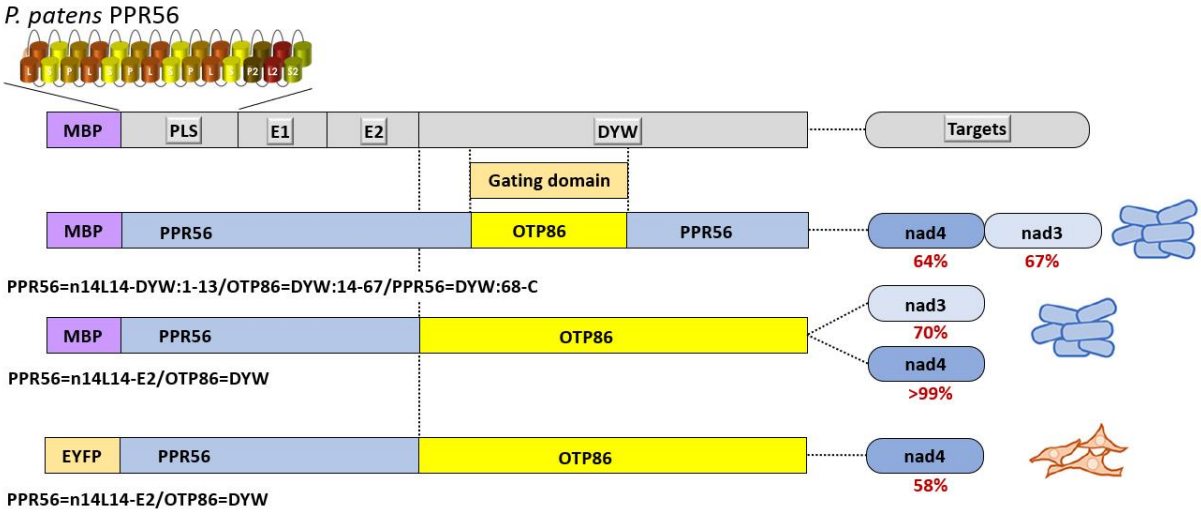


Figure 4. Off-targets for PPR56, PPR65 and a chimera

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

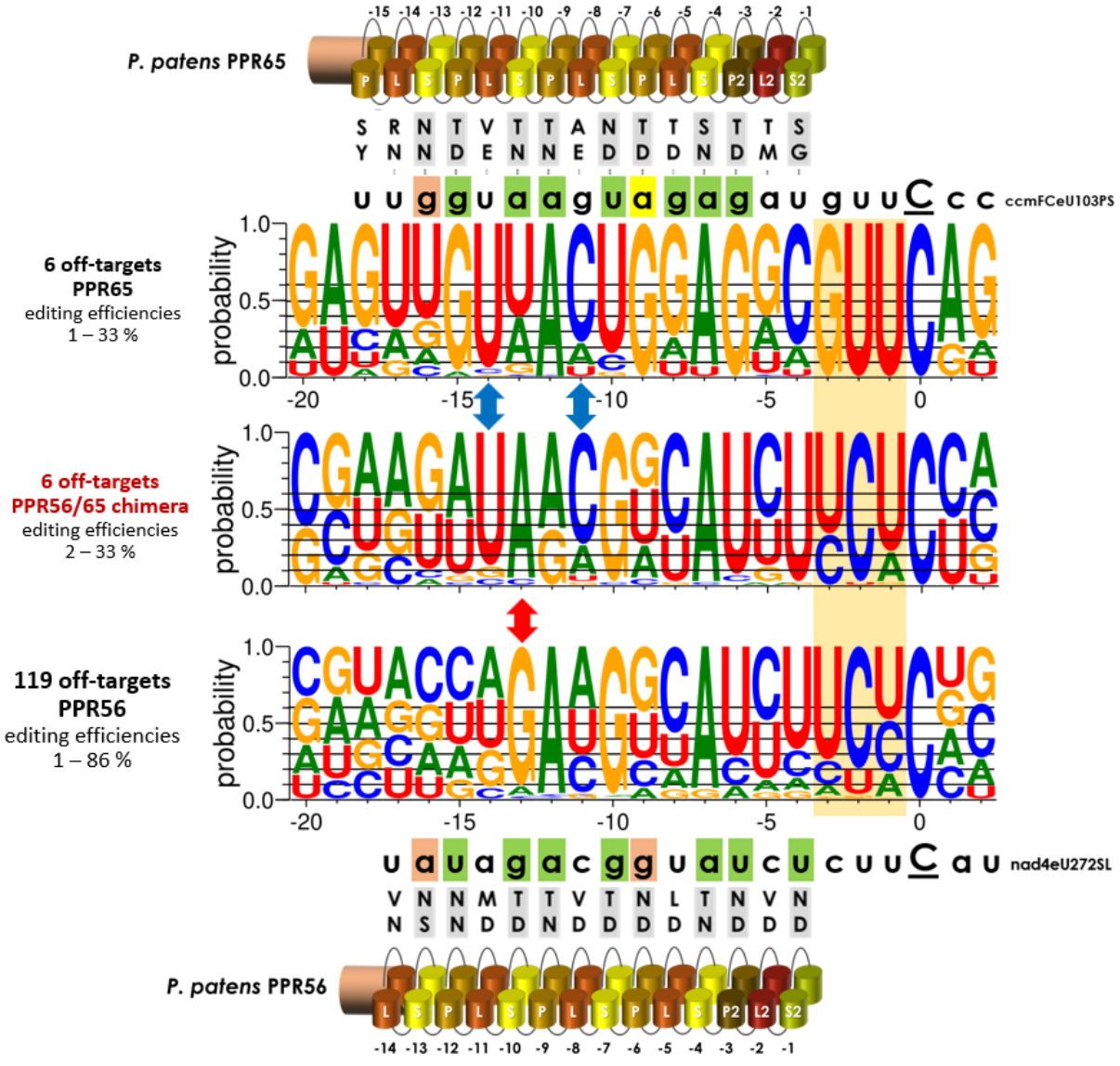


Figure 5. Off-targets of a PPR56-OTP86 chimera

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

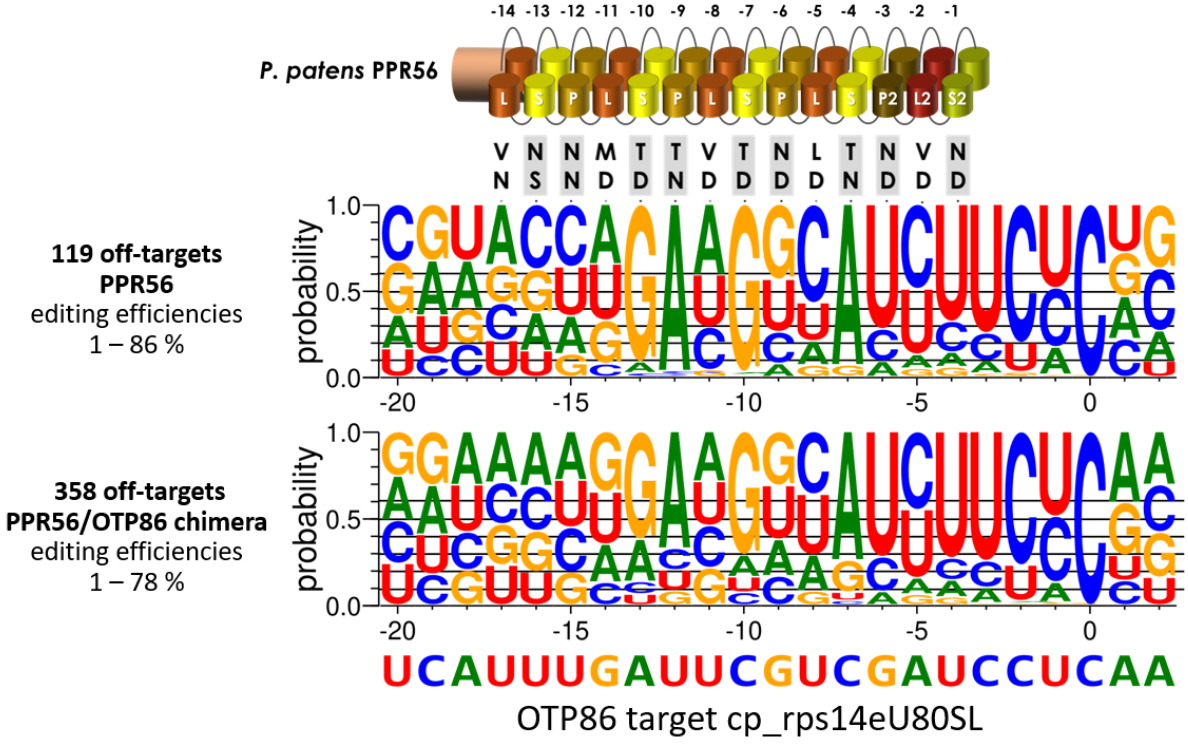
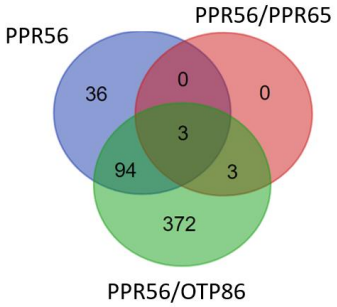


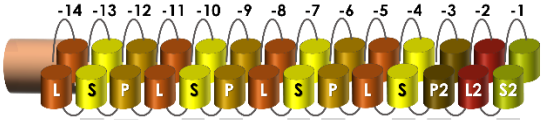
Figure 6. Off-targets of different PPR56 chimeras

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

A. Off-targets Venn-Diagram



B. Shared off-targets in detail

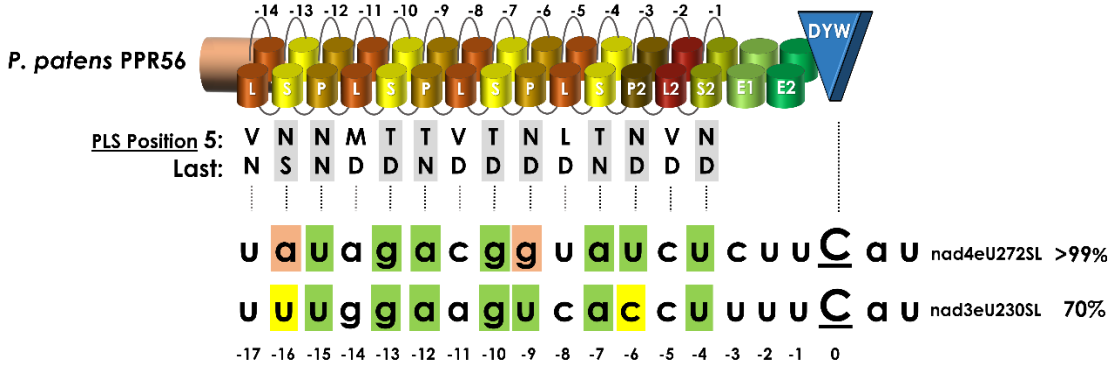


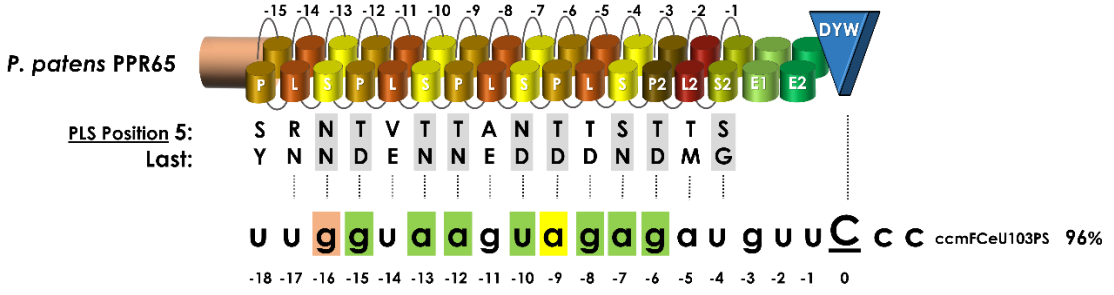
Pos	Name	V N	N S	N N	M D	T D	T N	V D	T D	N D	L D	T N	N D	V D	N D		PPR56	PPR56/ PPR65	PPR56/ OTP86
989754	**aspCeU1046SF**	a	c	a	g	a	a	c	g	g	c	a	u	g	u	u c u C c u	4.3 %	3.5%	**15.0 %**
1302851	**trpBeU1157TI**	c	g	a	u	a	a	a	g	a	c	a	u	c	u	u c a C c g	8.7 %	14.3 %	**26.4 %**
2868542	**recJeU425SF**	g	g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

Supplementary figures

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

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

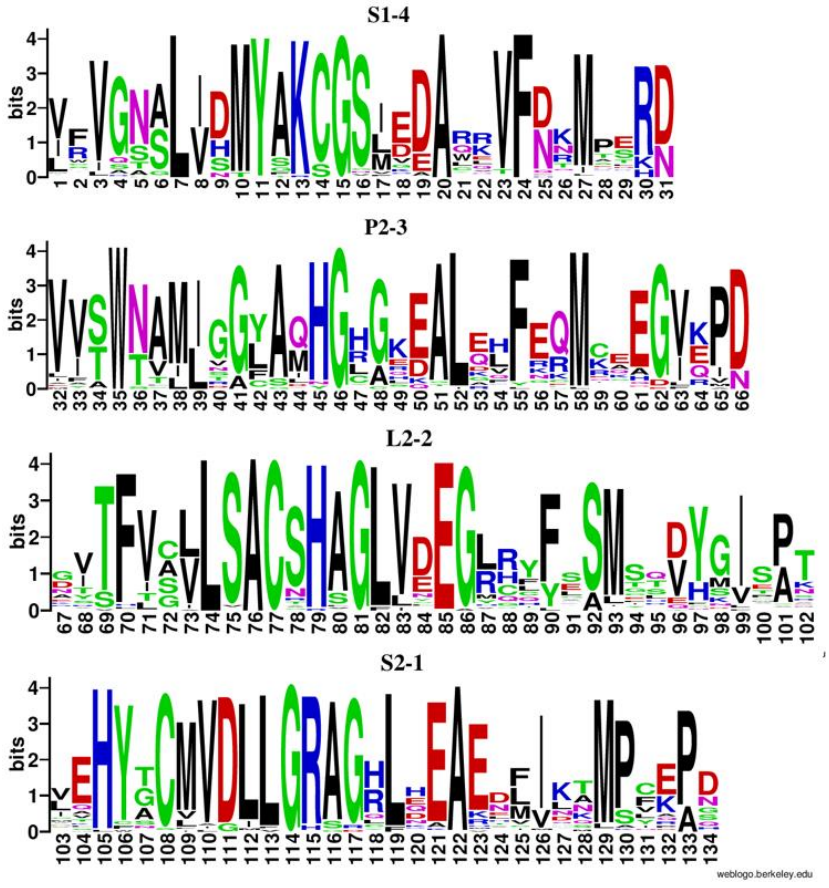




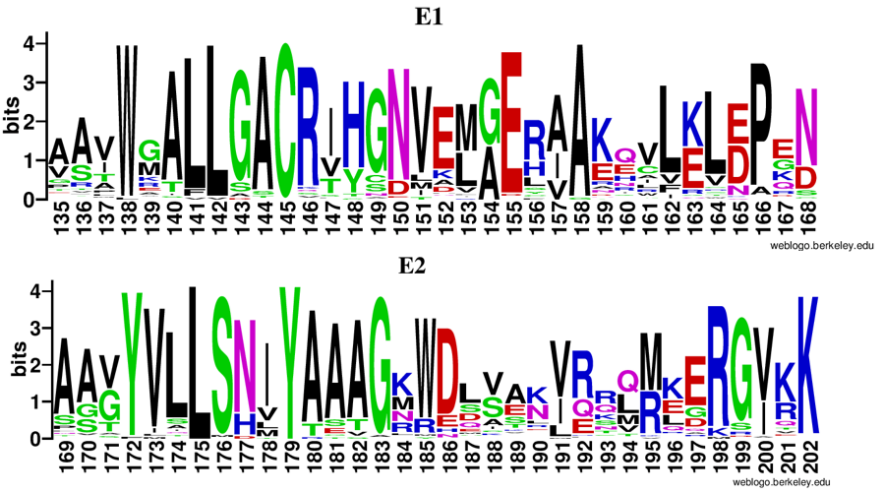
Supplementary Figure 2. WebLogo consensus profiles of C-terminal PPRs S1-4 to S2-1, TPR-like E1 and E2 motifs and DYW domains of moss proteins.

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

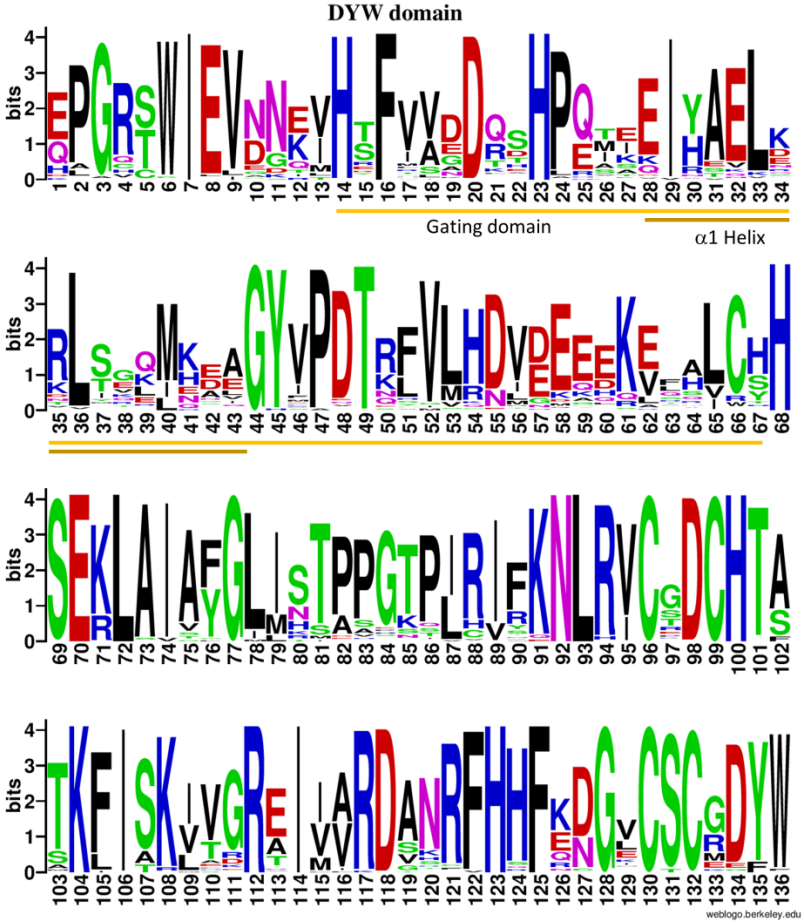
A. Consensus profiles of four terminal PPRs S-4, P2-3, L2-2 and S2-1



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

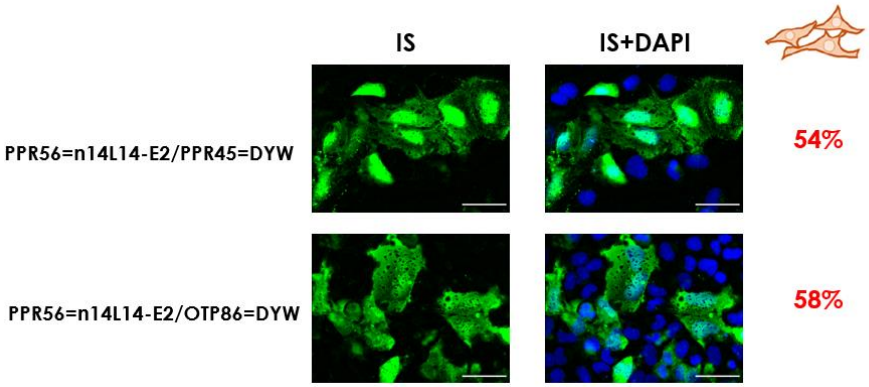


C. Consensus profile of the DYW domain



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

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



Supplementary tables

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

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

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

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

Supplementary Table 3. Table of identified off-targets.

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

Supplementary Table 4. Oligonucleotides.

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

713