Mitochondrion 19 (2014) 191–197

Contents lists available at ScienceDirect

Mitochondrion

journal homepage: www.elsevier.com/locate/mito

RNA editing in plant mitochondria —Connecting RNA target sequences and acting proteins

Mizuki Takenaka *, Daniil Verbitskiy, Anja Zehrmann, Barbara Härtel, Eszter Bayer-Császár, Franziska Glass, Axel Brennicke

Molekulare Botanik, Universität Ulm, 89069 Ulm, Germany

ARTICLE INFO

Available online 13 April 2014

Keywords:
Plant mitochondria
RNA editing
PPR proteins
Protein-RNA code
MORF proteins

ABSTRACT

RNA editing changes several hundred cytidines to uridines in the mRNAs of mitochondria in flowering plants. The target cytidines are identified by a subtype of PPR proteins characterized by tandem modules which each binds with a specific upstream nucleotide. Recent progress in correlating repeat structures with nucleotide identities allows to predict and identify target sites in mitochondrial RNAs. Additional proteins have been found to play a role in RNA editing; their precise function still needs to be elucidated. The enzymatic activity performing the C to U reaction may reside in the C-terminal DYW extensions of the PPR proteins; however, this still needs to be proven. Here we update recent progress in understanding RNA editing in flowering plant mitochondria.

1. Introduction: RNA editing – the process

A comparison of genomic and the thereof transcribed RNA sequences in the two plant extranuclear organelles with resident genomes, chloroplasts and mitochondria, reveals numerous sequence discrepancies in almost all land plants. These differences are caused by RNA editing, which changes specific nucleotide identities (Bock and Knoop, 2012; Chateigner-Boutin and Small, 2010; Finster et al., 2012; Shikanai, 2012). This type of RNA editing, which converts cytidines to uridines, evolved in land plants (Castandet and Araya, 2012; Gray, 2012; Knoop and Rüdinger, 2010) and was most likely subsequently lost in some marchantiid liverworts (Groth-Malonek et al., 2007; Knoop, 2013). The numbers of nucleotides altered vary between the plant lineages; only eight events occur in mitochondria of the moss *Funaria hygrometrica* and eleven in *Physcomitrella patens* (Rüdinger et al., 2009, 2011; Sugita et al., 2013; Ichinose et al., 2013). At the other end, the lycophytes *Isoetes engelmanii* and *Selaginella moellendorfii* modify more than 1.700 and 2.100 nucleotides, respectively (Grewe et al., 2011; Hecht et al., 2011). In these lycophytes as well as in ferns and hornworts reverse reactions converting U to C are observed in addition to the C to U alterations. In flowering plants, several hundred exclusively C to U alterations are seen in the two organelles, the majority occurring in mitochondria. Precise numbers are difficult to determine (Giegé and Brennicke, 1999; Schuster and Brennicke, 1991), if not impossible or rather inappropriate, since all degrees of percentages of altered nucleotides are found as elegantly demonstrated in a recent analysis of deep cDNA sequencing in the model plant *Arabidopsis thaliana* (Bentolila et al., 2013). In most recent mitochondrial genome analyses, RNA editing sites have been estimated by prediction programs such as PREP-mt (Mower, 2009) or the improved PREPACT (Lenz and Knoop, 2013), but have been probed only sporadically by cDNA analysis for selected genes. Nevertheless, the numbers of mitochondrial editing sites given for *Cycas taitungensis* with over 1000 (Chaw et al., 2008; Salmans et al., 2010), *Phoenix dactylifera* with almost 600 (Fang et al., 2012), *Spirodela polyrhiza* with 540 (Wang et al., 2012), and *Butomus umbellatus* with about 560 (Cuenca et al., 2013) reflect the relative frequencies of editing. For the recently analysed mitochondrial and plastid genomes in *Boea hygrometrica* and the very large mitochondrial genome in *Picea abies*, the number of edits was unfortunately not given (Nystedt et al., 2013; Zhang et al., 2012). Interestingly, extensive editing has been observed in a non-eudicot or -monocot angiosperm species, *Liriodendron tulipifera*, where more than 700 editing sites have been documented by cDNA analysis in protein coding sequences (Richardson et al., 2013). For *Amborella trichopoda*, the sister to all other angiosperms, experimental cDNA analyses identified many more editing sites in bona fide *Amborella* genes than in any other eudicot or monocot species, but a total number would be difficult to assign in this plant due to the extensive integration of sequences from other plant species which are also partly transcribed and edited (Rice et al., 2013). Unfortunately, the predictions as well as cDNA probings are often limited to protein coding

* Corresponding author. Tel.: +49 731 5022610; fax: +49 731 5022610.
E-mail address: mizuki.takenaka@uni-ulm.de (M. Takenaka).

http://dx.doi.org/10.1016/j.mito.2014.04.005
1567-7249/

sequences and the extent of editing in rRNA and tRNA molecules as well as in introns remains open in most plant species. Deep transcriptome sequencing as was first done for *Vitis vinifera* (Picardi et al., 2010) detected 44 editing events in structural RNA sequences of tRNAs and introns. The recent deep sequence analysis of *Nicotiana tabacum* mitochondrial transcripts identified five editing events in tRNAs and 73 in non-coding regions in addition to 557 editing events in open reading frames (Grimes et al., 2014).

The biochemical results of RNA editing in plants are a C to U deamination and in lycophytes also the U to C amination, but the underlying reactions and their catalysts are as yet unclear. Recent reviews summarize general and specific aspects of RNA editing and the reader is referred to these for general overviews (Several chapters in: Bock and Knoop, 2012; Fujii and Small, 2011; Hammani and Giegé, 2014; Knoop, 2004, 2011, 2013; Shikanai, 2006; Takenaka et al., 2013b; Takenaka, in press). We here focus on advances in understanding the proteins involved in editing and in deciphering the contacts between these proteins and the RNA. Major progress has been made in understanding how the C nucleotides to be edited to U are identified in the RNA population of the organelles in angiosperms. The proteins recognizing and binding to a specific RNA sequence are a subgroup of the PPR proteins (pentatricopeptide repeat proteins), which are involved in targeting the RNA editing reactions to specific nucleotides as detailed below (Barkan and Small, 2014; Lurin et al., 2004; O'Toole et al., 2008; Schmitz-Linneweber and Small, 2008; Small and Peeters, 2000; Takenaka et al., 2008). Additional proteins, unrelated to PPRs, are also involved in organellar RNA editing, suggesting that the process is mediated by complex editosomes consisting of several different proteins in chloroplasts as well as in mitochondria. A number of excellent reviews have been written on the functions, actions, structures and specializations of PPR proteins (Barkan and Small, 2014; Schmitz-Linneweber and Small, 2008; Yagi et al., 2013b). We therefore focus here on selected aspects to summarize recent developments in our understanding of the RNA editing process and point out several present lines of investigations.

2. Specific nucleotides are edited—the RNA-protein code

In the organellar RNAs, nucleotides to be edited must be distinguished from those that have to be left unaltered. In flowering plants, no editing appears to alter the rRNAs (Schuster et al., 1991), but some tRNAs are edited (Binder et al., 1994; Grewe et al., 2009), suggesting that in these instances secondary structure folding of the RNA can be opened for editing and/or tRNA precursors are not as tightly folded (Grimes et al., 2014; Marchfelder et al., 1996). *In vitro* and *in organello* assays had shown that the sequence pattern 5–20 nucleotides upstream of the edited nucleotide contain the determining address (Bock et al., 1996; Farré et al., 2001; Hegeman et al., 2005; Neuwirt et al., 2005; Takenaka et al., 2004; van der Merwe et al., 2006). These unique RNA sequences are recognized and bound by the complementary unique arrangement of variant tandem elements of specific PPR proteins (Barkan and Small, 2014).

PPR proteins are found in all eukaryotes, but their numbers have greatly increased in land plants in comparison to fungi and animals and also to alga. A recent survey of sequenced genomes identified 8–17 PPR proteins in Rhodophyta and 14–25 in species of the Chlorophyta (Tourasse et al., 2013). In flowering plant genomes, several hundred PPR proteins are encoded in the nucleus, *e.g.* 450 in *A. thaliana*, 477 in *Oryza sativa*, 365 in *Medicago trunculata*, 629 in *Glycine max*, more than 1.000 in *S. moellendorfii* and 106 in *P. patens*, but only 10 in *Schizosaccharomyces pombe* and 15 in *Saccharomyces cerevisiae* (Fujii and Small, 2011; Lurin et al., 2004). Their PPR name-giving substructures are the variant elements of about 35 amino acids arranged in tandem in the protein. Elements in PPR proteins involved in processes other than editing are usually 35 amino acids long (P-type) and PPR proteins for RNA editing contain elements of variable length (PLS-type). Each of these modules can bind to one nucleotide and the number of elements defines the maximal length of RNA sequence that can be recognized. The number of elements in plant PPR proteins is usually larger than the minimal RNA sequence of about 6–8 nucleotides required in the RNA population to define a unique site. However, some PPR proteins have fewer such elements and would thus not be able to target a specific site without further additional guiding. The largest PPR protein in *A. thaliana* in terms of number of PPR motifs is MEF12 which contains 25 such elements (Härtel et al., 2013), much larger than the average number of 13–14 such PPR motifs in *A. thaliana* (13.7 average) and in *Oryza sativa* (13.1 average) (Fig. 1). In the moss *P. patens*, average numbers of PPR elements are higher than in flowering plants with about 20 such repeats (Fig. 1). The reason for these differences is not clear; the mitochondrial and plastid genomes in the moss are not more complex than in the flowering plants and thus do not require an extended specificity through a larger recognition motif.

The coding system in the PPR elements basically relies on the amino acid identities at two specific positions in the PPR elements. These were first identified by *in silico* analyses comparing target RNA patterns and coinciding amino acid identities in non-editing P-type and editing PLS-type PPR proteins (Barkan et al., 2012; Takenaka et al., 2013a; Yagi et al., 2013a). These coincidences have been confirmed by crystal structural analyses of the PPR repeats with and without the target RNA sequence (Ban et al., 2013; Ke et al., 2013; Yin et al., 2013). Although these were obtained from a strongly, and possibly irreversibly, RNA binding P-type PPR protein (PPR10), the similarity between the code parameters suggests that each element in the RNA editing PPR proteins will analogously fold into two helical repeats and expose the same amino acid positions to the respective ribonucleotide. Still, the different lengths of the variant PLS elements in the RNA editing PPR proteins between 32 and 38 amino acids may exert an influence in the molecular interactions and contribute to the difference in affinity distinct from the group of tightly binding PPR proteins such as PPR10. PPR10 and other such P-type PPR proteins involved in stopping exonucleolytic DNases must attach strongly if not irreversibly to their target to resist dissociation by the progressive DNA degrading enzyme and to resolve secondary structures (reviewed in Barkan and Small, 2014). While the target sequence bound by *e.g.* PPR10 is detected as a small RNA molecule, none of the editing site motifs have been seen in such assays or by primer extension probings (Ruwe and Schmitz-Linneweber, 2012). It will be important in the near future to identify the parameters allowing an RNA editing PPR protein to attach to specific target sites and yet to be able to dissociate again rapidly from the RNA so as not to inhibit translation or further editing by other factors.

The protein–protein connection in the PPR10 dimers suggests that the repeats not only bind to RNA but can also present a protein binding surface and undergo connections to other proteins in the editosome. The surprising observation of the dimer formation of PPR10 in the crystal structure analysis raises the question about its relevance *in vivo*. Does such a homodimer form *in vivo* and does it bind to two RNA molecules? How should we imagine this close vicinity of two mRNA molecules? Alternatively, is the dimer only formed *in vitro* in the absence of other proteins which would *in vivo* dissociate the PPR protein homodimers? Is the PPR homodimer then the non-physiological product of sticky protein surfaces of the PPR elements? In any case, while one face of the PPR elements clearly attaches to the RNA, the other is accessible to protein–protein interactions which may be involved in the assembly of the editosome *in vivo* (Fig. 2).

3. Additional domains in the RNA editing PPR proteins—signatures of a deaminase

In addition to the variable length of the PPR elements, the RNA editing PPR proteins are characterized by C-terminal extensions. Adjacent to the PPR elements is the so-called E-domain which is present in all of the RNA editing PPR proteins. Structural and some sequence similarities suggest that this region evolved from two PPR repeat modules.