the plant journal

SEB Society for Experimental Biology

The Plant Journal (2016) **85,** 607–621 doi: 10.1111/tpj.13126

SOT1, a pentatricopeptide repeat protein with a small MutS-related domain, is required for correct processing of plastid 23S–4.5S rRNA precursors in *Arabidopsis thaliana*

Wenjuan Wu[1,†], Sheng Liu[2,†], Hannes Ruwe[2,§], Delin Zhang[3], Joanna Melonek[2], Yajuan Zhu[1], Xupeng Hu[1], Sandra Gusewski[2,¶], Ping Yin[3], Ian D. Small[2,4], Katharine A. Howell[2,‡,*] and Jirong Huang[1,‡,*]

[1]*National Key Laboratory of Plant Molecular Genetics, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China,*

[2]*Australian Research Council Centre of Excellence in Plant Energy Biology, The University of Western Australia, Crawley, Western Australia 6009, Australia,*

[3]*National Key Laboratory of Crop Genetic Improvement, National Centre of Plant Gene Research, Huazhong Agricultural University, Wuhan 430070, China, and*

[4]*Centre of Excellence in Computational Systems Biology, The University of Western Australia, Crawley, Western Australia 6009, Australia*

Received 19 December 2015; accepted 12 January 2016; published online 23 January 2016.

*For correspondence (emails kate.howell@uwa.edu.au or huangjr@sibs.ac.cn).

†These authors contributed equally to this work.

‡These authors contributed equally to this work.

§Present address: Molecular Genetics, Institute of Biology, Humboldt University, Berlin D–10115, Germany.

¶Present address: Department of Genetics, Friedrich Schiller University of Jena, Jena D–07743, Germany.

SUMMARY

Ribosomal RNA processing is essential for plastid ribosome biogenesis, but is still poorly understood in higher plants. Here, we show that SUPPRESSOR OF THYLAKOID FORMATION1 (SOT1), a plastid-localized pentatricopeptide repeat (PPR) protein with a small MutS-related domain, is required for maturation of the 23S–4.5S rRNA dicistron. Loss of SOT1 function leads to slower chloroplast development, suppression of leaf variegation, and abnormal 23S and 4.5S processing. Predictions based on the PPR motif sequences identified the 5′ end of the 23S–4.5S rRNA dicistronic precursor as a putative SOT1 binding site. This was confirmed by electrophoretic mobility shift assay, and by loss of the abundant small RNA 'footprint' associated with this site in *sot1* mutants. We found that more than half of the 23S–4.5S rRNA dicistrons in *sot1* mutants contain eroded and/or unprocessed 5′ and 3′ ends, and that the endonucleolytic cleavage product normally released from the 5′ end of the precursor is absent in a *sot1* null mutant. We postulate that SOT1 binding protects the 5′ extremity of the 23S–4.5S rRNA dicistron from exonucleolytic attack, and favours formation of the RNA structure that allows endonucleolytic processing of its 5′ and 3′ ends.

Keywords: pentatricopeptide repeat protein, small MutS-related domain, chloroplast, plastid ribosome, ribosomal RNA, *Arabidopsis thaliana*.

INTRODUCTION

The endosymbiotically derived organelles, the mitochondria and chloroplasts, have retained their own genomes but have been reduced in size through gene transfer and loss over the course of their evolution (Timmis *et al.*, 2004). Retention of these vestigial genomes requires that the genetic machinery necessary for their expression is also present. This genetic machinery in organelles is comparable to that found in eubacteria, but has evolved unique features that distinguish it from its ancestral origins (Barkan, 2011; Hammani and Giege, 2014). For example, involvement of additional nucleus-encoded factors, such as classes of RNA-binding proteins that are not found in bacteria, is one of the defining features of organelle gene expression (Barkan, 2011; Hammani and Giege, 2014). A prominent example of these RNA-binding proteins is the pentatricopeptide repeat (PPR) protein family, which was found to be ubiquitous in eukaryotes (Schmitz-Linneweber and Small, 2008) after its discovery in the model plant *Arabidopsis thaliana* (Aubourg *et al.*, 2000; Small and Peeters, 2000).

PPR proteins (PPRs) are characterized by tandem arrays of degenerate amino acid motifs ranging in length from 31 to 36 amino acids (Lurin *et al.*, 2004). Since their discovery, a growing body of evidence has shown that PPRs are localized almost exclusively to organelles, and that they play a role in a variety of gene expression events including transcription, RNA cleavage, intron splicing, RNA stabilization, RNA editing and translation (Barkan and Small, 2014). Many of these functions may be explained by PPRs simply passively binding to organellar transcripts and either blocking exoribonucleases, and thereby defining RNA precursor ends, or remodelling bound RNA, and exposing or masking ribosome binding sites or *cis*-elements required for other RNA processing events such as cleavage, splicing or editing (Barkan and Small, 2014). Specific examples of these two modes of action have been described (Prikryl *et al.*, 2011; Zhelyazkova *et al.*, 2012). In addition, it has been reported that stable interactions between PPRs and RNA may be tracked as PPR 'footprints' in small RNA sequencing data sets (Ruwe and Schmitz-Linneweber, 2012; Zhelyazkova *et al.*, 2012). Furthermore, structural studies have now confirmed previous models suggesting that PPR motifs form helical sheets that bind RNA in a modular, sequence-specific manner (Yin *et al.*, 2013). In addition, recent breakthroughs in understanding the 'code' that underlies the specificity of how these proteins bind RNA (Barkan *et al.*, 2012; Takenaka *et al.*, 2013; Yagi *et al.*, 2013) have facilitated subsequent studies in which predicted RNA targets have been proven experimentally (Takenaka *et al.*, 2013; Yap *et al.*, 2015) and predictable alteration of sequence recognition has been achieved (Barkan *et al.*, 2012; Kindgren *et al.*, 2015).

In land plants, the PPR protein family has expanded considerably, with 450 members found in Arabidopsis (O'Toole *et al.*, 2008), comprising the P class and PLS class sub-families, defined by the presence of the canonical 35 amino acid motif and variable-length motifs, respectively (Lurin *et al.*, 2004). These sub-families are also separated by function: in most cases, PLS class PPRs are involved in RNA editing, while P class PPRs have more diverse roles in cleavage, splicing, stabilization and/or translation of their target transcripts (Barkan and Small, 2014). In addition, PLS class PPRs are often further distinguished by the presence of additional C–terminal domains: the E, E+ and DYW domains (Lurin *et al.*, 2004). In comparison, most P class PPRs do not contain additional domains. An exception to this is a small sub-group of P class PPRs that contain a C–terminal small MutS-related (SMR) domain (Liu *et al.*, 2013a). The SMR domain was initially described in the cyanobacterium *Synechocystis*, as a C–terminal domain present in the MutS2 protein (Moreira and Philippe, 1999), and was shown to specifically confer DNA nuclease activity (Fukui *et al.*, 2007). Since then, proteins containing SMR domains have been found in both prokaryotic and eukaryotic species (Fukui and Kuramitsu, 2011). However, the coupling of PPR motifs with a C–terminal SMR domain into a single protein is restricted to the Viridiplantae, but as PPR-SMR proteins are found in green alga, they clearly emerged early in the evolution of the PPR protein family and prior to its massive expansion in land plants (Liu *et al.*, 2013a).

In Arabidopsis, just eight of the 450 PPR proteins belong to the PPR-SMR sub-group, and five are localized to the chloroplast (Liu *et al.*, 2013a). Despite its small size, this sub-group of PPRs has been the focus of numerous studies, due to the fact that the enigmatic protein GUN1, a central regulator of chloroplast retrograde signalling (de Dios Barajas-Lopez *et al.*, 2013), was found to encode a PPR-SMR protein (Koussevitzky *et al.*, 2007), and its precise function remains elusive. In addition, estimations of protein abundance using published proteomic datasets indicated that three PPR-SMR proteins dominate the protein mass attributed to all PPRs in plastids of both Arabidopsis and maize (*Zea mays*), specifically the proteins pTAC2/ZmpTAC2, SVR7/ATP4 and AT5G46580/PPR53 (Liu *et al.*, 2013a).

Of these abundant PPR-SMR proteins, PTAC2 and SVR7 in Arabidopsis and ATP4 in maize have been characterized. PTAC2 was identified as a novel component of the plastid transcriptionally active chromosome (pTAC), and *ptac2* mutants are only able to grow with an exogenous sugar supply and show a large reduction in transcripts generated by the plastid-encoded RNA polymerase (Pfalz *et al.*, 2006). The *svr7* and *atp4* mutants have less severe phenotypes but do exhibit plastid gene expression effects. SVR7 is required for plastid rRNA maturation and plastid ATP synthase expression (Liu *et al.*, 2010b; Zoschke *et al.*, 2013a). The SVR7 ortholog in maize, ATP4, is required for *atpB* translation and has additional roles in stabilization of *psaJ* and *rpl16–rpl14* mRNAs (Zoschke *et al.*, 2012, 2013b). The characterization (and naming) of SVR7 came about through a screen for suppressors of *var2*, a mutant lacking the FtsH2 subunit of the FtsH protease, which leads to a variegated leaf phenotype, whereby plants have green- and white-sectored leaves containing normal chloroplasts and non-pigmented plastids, respectively (Chen *et al.*, 2000). Genetic screening for suppressors of *var2* and subsequent characterization of the isolated mutants have shown that mutations suppressing the leaf variegation phenotype directly or indirectly lead to reduced numbers of chloroplast ribosomes (e.g. Yu *et al.*, 2008; Liu *et al.*, 2010a). However, although plastid rRNA biogenesis is mildly disrupted in *svr7* mutants, the direct target of the SVR7 protein has not been determined.

To further characterize the PPR-SMR protein sub-group, we investigated the function of the Arabidopsis protein encoded by the gene AT5G46580, denoted here as SOT1 (SUPPRESSOR OF THYLAKOID FORMATION1). This

protein is orthologous to maize PPR53 (GRMZM2G438524), as described by Zoschke *et al.* (2016). We show that SOT1 binds to the 5′ region of the 23S–4.5S rRNA precursor and blocks exonucleolytic attack, thus facilitating plastid ribosomal RNA maturation and ribosome assembly.

RESULTS

Identification and characterization of the *sot1* mutant

The *sot1* (*suppressor of thf1 1*) mutant was initially identified in a suppressor screen for the leaf variegation phenotype of *thylakoid formation 1* (*thf1*) (Wang *et al.*, 2004), as previously described (Wu *et al.*, 2013; Hu *et al.*, 2015; Ma *et al.*, 2015). The original suppressor line, denoted *40–1*, and the *sot1–1* mutant line, isolated from F_2 progeny of the cross between *40–1* and wild-type (WT), had similar phenotypes: they did not display variegated leaves, but were reduced in size and had paler green cotyledons and leaves compared to WT (Figure 1a and Figure S1). Map-based cloning was performed as previously described (Wu *et al.*, 2013; Hu *et al.*, 2015; Ma *et al.*, 2015), and showed that the suppressor gene was located within a 78 kb interval between the K11I1 and MZA15 markers on the top arm of chromosome 5, which contains 19 annotated genes (Figure 1b). A single nucleotide substitution (C→T) was found to occur 1859 nucleotides downstream of the start codon of AT5G46580, resulting in an amino acid change from serine to leucine (Figure 1b). The map-based cloning result was confirmed, as *40–1* plants expressing *SOT1* under the control of the CaMV 35S promoter displayed a variegated leaf phenotype (Figure 1a). The AT5G46580/SOT1 locus encodes a pentatricopeptide repeat (PPR) protein that contains 11 PPR motifs and a C–terminal SMR domain (Figure 1b). Thus, SOT1 belongs to the small subset of P class PPR proteins known as PPR-SMR proteins.

A T–DNA insertion line GK_840D06, denoted *sot1–2*, was also identified, and plants were confirmed to be homozygous by PCR-based genotyping (Figure 1c). The position of the T–DNA insertion was confirmed by sequencing to be located 5′ of the first PPR motif (Figure 1b). In contrast to *sot1–1*, the *SOT1* transcript was not detected in *sot1–2* (Figure 1d), and the growth phenotype was more severe (Figure 1e). Like *sot1–1*, *sot1–2* was able to suppress *thf1* leaf variegation (Figure 1e), and both *sot1–1* and *sot1–2* were found to also suppress leaf variegation in *var2* (Figure 1f).

The SOT1 protein contains an N–terminal signal peptide that is predicted to target chloroplasts using TargetP 1.1 (Emanuelsson *et al.*, 2000). This is consistent with proteomic datasets indicating that SOT1 is present in the chloroplast (summarized in Liu *et al.*, 2013a). Northern blotting detected *SOT1* transcripts in all green tissues, with high levels in rosette leaves, flowers and siliques, but not in roots (Figure S2a), and GUS reporter assays showed that the *SOT1* promoter was active in various green tissues, such as cotyledons, leaves, flowers and siliques (Figure S2b). These results are in agreement with the chloroplast localization of SOT1 and a putative function in chloroplast development.

Given that SOT1 is one of five chloroplast-targeted PPR-SMR proteins in Arabidopsis (Liu *et al.*, 2013a), which include GUN1, a protein required for chloroplast-to-nucleus retrograde signalling (Koussevitzky *et al.*, 2007), the *sot1* mutant was tested for the *genomes uncoupled* (*gun*) phenotype. Under treatment with norflurazon, *gun* mutants show an inability to repress photosynthesis-associated nuclear genes in response to chloroplast dysfunction. As previously observed for *svr7–3* (Zoschke *et al.*, 2013a), the *sot1–1* mutant was able to repress expression of photosynthesis-associated nuclear genes in the presence of norflurazon (Figure S3), indicating that it is not involved in this retrograde signalling pathway.

Loss of SOT1 affects levels of *ndhA* and chloroplast rRNA species

As SOT1 was found to encode a PPR protein localized to the chloroplast, we deduced that chloroplast gene expression would be affected in the *sot1* mutants and its analysis may reveal putative SOT1 RNA targets. Quantitative RT–PCR was performed to assess transcript levels of all protein-coding genes encoded in the plastid genome (Figure S4). This analysis revealed that transcripts were, in general, mildly up-regulated, with an obvious exception being considerable down-regulation of *ndhA* levels in the *sot1* mutant plants compared to WT. However, as it is known that reduced expression of subunits of the chloroplast NADH dehydrogenase-like complex usually does not result in a clear phenotype when plants are grown under normal conditions (Shikanai, 2007), the reduction in *ndhA* levels cannot explain the *sot1* mutant growth phenotype (Figure 1).

While preparing RNA for analysis, it was noted that samples from the *sot1* mutants displayed differences in chloroplast rRNA accumulation (Figure 2b). Differences in the levels of 23S rRNA species were observed, and these were different to those observed in *svr7–3*, which had been previously reported to show chloroplast rRNA processing defects (Liu *et al.*, 2010b; Zoschke *et al.*, 2013a). To investigate this further, northern blotting was performed using probes corresponding to all rRNA components of the chloroplast ribosome to assess the presence of precursor, intermediate and mature transcripts (Figure 2a,c). This revealed that *sot1–1* mutants had specific differences in the accumulation of 23S and 4.5S rRNAs (generally increased levels of precursors and decreased levels of mature transcripts), with less severe effects on the 16S and 5S rRNAs. Notably, the chloroplast rRNA defects observed were distinct from those seen for *svr7–3*, and in addition to quantitative differences, qualitative differences were also observed. Specifically, the 3.2 knt transcript

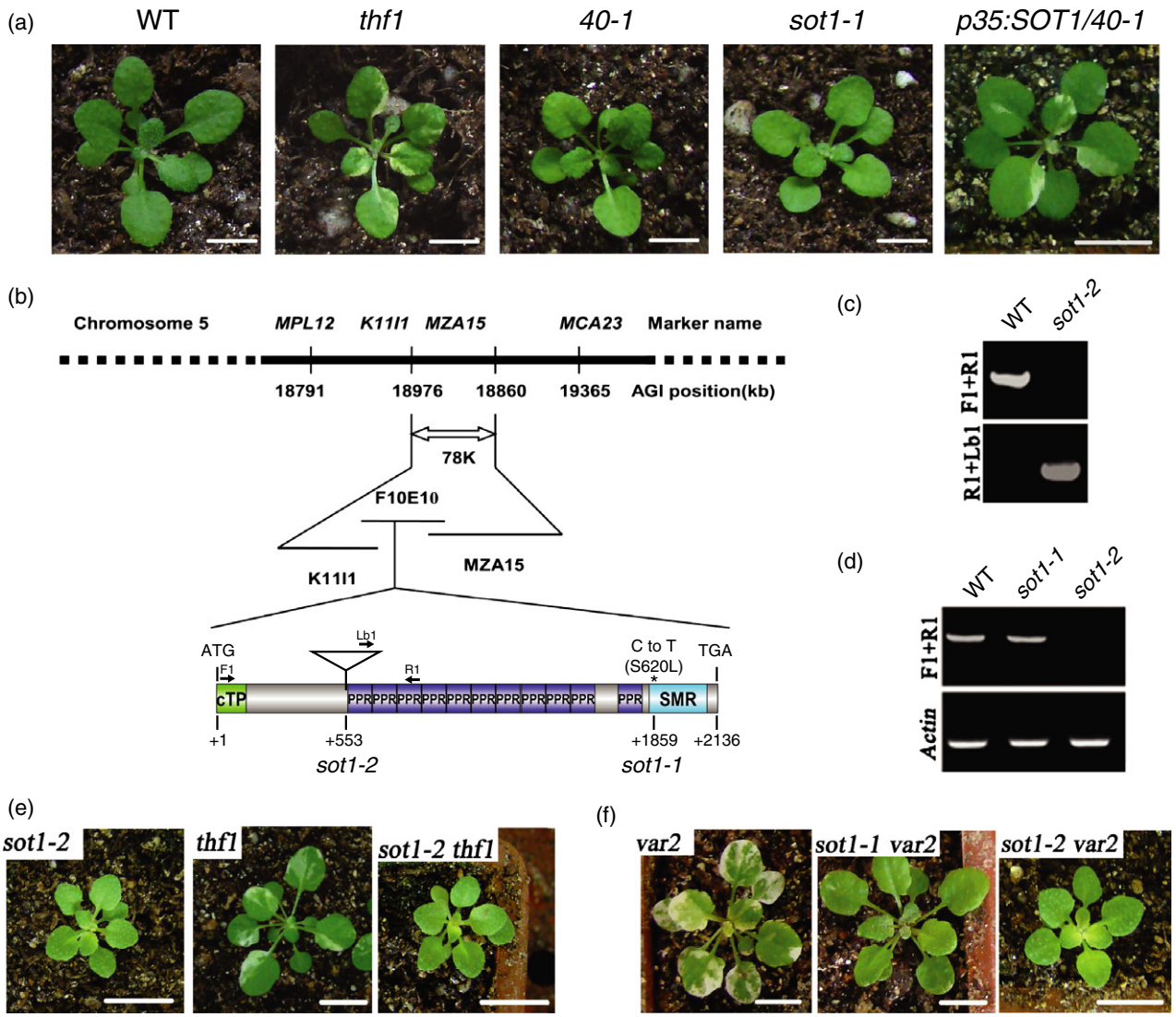

Figure 1. Identification and characterization of the *thf1* suppressor mutant *sot1*.
(a) Phenotypes of 3-week-old plants grown under short-day conditions. Scale bars = 0.5 cm except for *p35:SOT1/40–1*, for which the scale bar = 1 cm.
(b) Map-based cloning of *SOT1*. The *SOT1* locus was mapped between markers K11I1 and MZA15 on chromosome 5. The markers used for fine mapping are shown above the line. The numbers under the line indicate the chromosomal position in kb. The *SOT1* (AT5G46580) gene structure from the initiation (ATG) codon to the termination (TGA) codon within a BAC clone (F10E10) is shown enlarged. A single exon encodes a protein containing a chloroplast transit peptide (cTP), 11 pentatricopeptide repeat (PPR) motifs and a C–terminal small MutS-related (SMR) domain. A point mutation (C→T) was identified in *sot1–1* at position +1859, and results in an amino acid substitution from serine to leucine (S620L) in the SMR domain. Also shown is the position of the T–DNA insertion in *sot1–2* (GK_840D06), which was confirmed via sequencing to be just upstream of the PPR motifs (+553), and of gene-specific primers F1 and R1, used for genotyping and detection of *SOT1* transcripts as shown in (c) and (d). Lb1, T–DNA left border primer.
(c) Confirmation of homozygosity by PCR analysis of the T–DNA insertion in *sot1–2* using the primers shown in (b).
(d) Detection of *SOT1* transcripts in the *sot1–1* and *sot1–2* mutant lines.
(e,f) Forty-day-old plants grown under short-day conditions, showing that *thf1* leaf variegation is also suppressed by *sot1–2*, and *var2* leaf variegation is suppressed by both *sot1–1* and *sot1–2*. Scale bars = 1 cm.

representing the 23S–4.5S dicistronic precursor appeared to be smaller in *sot1–1* (top band in 23S 5′ and 3′ northern panels, Figure 2c) and an additional transcript was detected using the 23S 3′ probe (below the 2.4 kb band, Figure 2c), which we speculate to be a degradation product. The differences in chloroplast rRNAs observed in the *sot1–1* mutant were also seen in *sot1–2* (Figure S5).

In addition to the rRNA component of chloroplast ribosomes, the protein component was also examined by performing western blotting on total leaf protein using antibodies against two nucleus-encoded ribosomal proteins of the small and large subunit respectively, RPS1 and RPL4 (Figure 2d). RPS1 showed a mild reduction in the *sot1* mutants, while RPL4 showed a dramatic reduction.

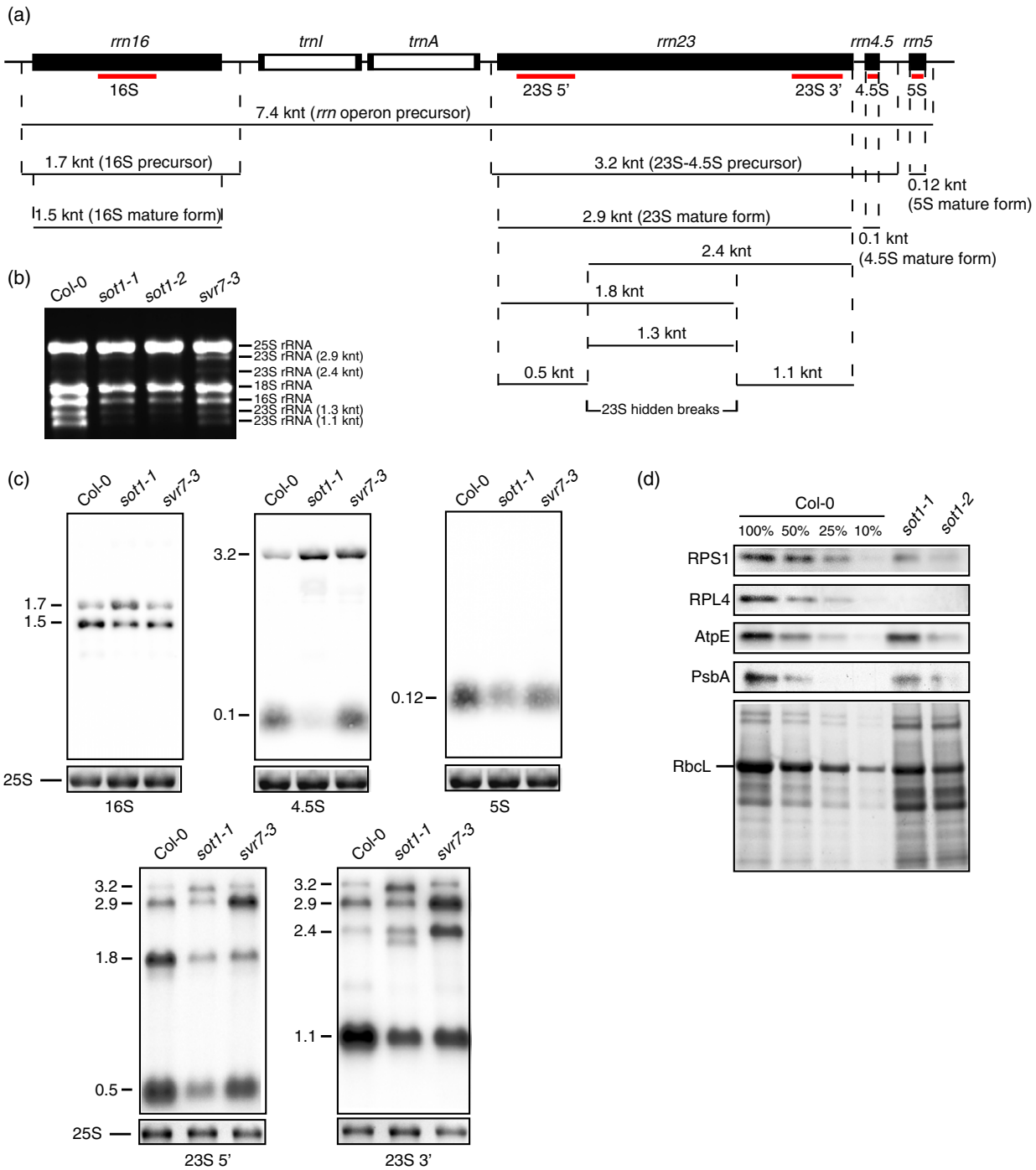

Figure 2. Analysis of plastid ribosome components.

(a) Schematic representation of the Arabidopsis *rrn* operon, which encodes all rRNA components of the plastid ribosome (16S, 23S, 4.5S and 5S rRNA). The mature forms of each rRNA are generated by the action of endo- and exo-ribonucleases on the primary transcript. The sizes of all precursors, intermediates and mature forms are given in kilo-nucleotides (knt).

(b) Differences in plastid rRNA accumulation are seen when total RNA from green tissue is separated on a formaldehyde-containing denaturing agarose gel. The accumulation of various 23S rRNA species is reduced in *sot1* mutant plants, and this is distinct from differences seen in the *svr7–3* mutant.

(c) Northern blot analysis confirms that plastid rRNA processing is affected in the *sot1–1* mutant, and that the differences in 23S rRNA species observed are distinct from those seen in the *svr7–3* mutant. The positions of probes used for northern blotting in (c) are indicated by red lines in (a). Ethidium bromide staining of the 25S rRNA is shown below each blot to indicate differences in gel loading.

(d) Levels of two nucleus-encoded plastid ribosomal proteins, RPS1 and RPL4, and of two plastid-encoded proteins, AtpE and PsbA, were assessed by western blotting of total proteins separated by SDS–PAGE, and were found to be reduced in the *sot1* mutants compared to WT. Visualization of the gel to check loading also shows a reduced amount of RbcL protein in the *sot1* mutant plants.

This indicates that biogenesis of the large chloroplast ribosome subunit is specifically perturbed in the *sot1* mutants, which is consistent with the rRNA deficiencies observed. The reduced amount of RbcL protein observed suggests that plastid translation in general is also perturbed. This is further supported by reduced levels of two plastid-encoded proteins: AtpE and PsbA.

Identifying a putative target site of SOT1

As the size of the 23S–4.5S rRNA precursor appeared smaller in the *sot1* mutants by northern blot analysis, this was investigated further by mapping transcript ends. 5′ RACE was performed using a primer adjacent to the start of the mature transcript (Figure 3a). The resulting PCR products were sequenced to determine the positions of the 5′ ends. This analysis showed that the processed 5′ end of the 23S rRNA precursor, found only in Col–0, maps to a position 73 nt upstream of the mature 23S rRNA end as previously described (Bollenbach *et al.*, 2005; Hotto *et al.*, 2015). This product is absent in the *sot1–2* mutant (Figure 3a), suggesting that SOT1 may be involved in stabilizing this precursor. The larger RACE products, derived from both Col–0 and *sot1–2* RNA, mapped to a region in the *trnA* intron. To our knowledge, there is no previous report of this putative transcript end or of an associated promoter. The effect of tobacco acid pyrophosphatase (TAP) strongly suggests that this band is generated from reverse transcription of a primary transcript, but we have no further evidence that this represents a true transcript end. Moreover, a -10 promoter consensus sequence (TGGTAGAAT) was predicted just upstream of the transcript (de Jong *et al.*, 2012), but additional experimentation is required to prove its activity and is outside the scope of the current study.

We also inspected the 5′ and 3′ ends of the dicistronic 23S–4.5S rRNA using circular RT–PCR (cRT–PCR) combined with DNA sequencing. This analysis showed that unprocessed and/or mis-processed 5′ and 3′ ends were detected in *sot1–2*, but not in WT or *svr7–3* (Figure 3b). Among the 34 23S–4.5S dicistronic molecules analysed for the *sot1–2* sample, only 14 mapped to the mature ends of the 23S–4.5S

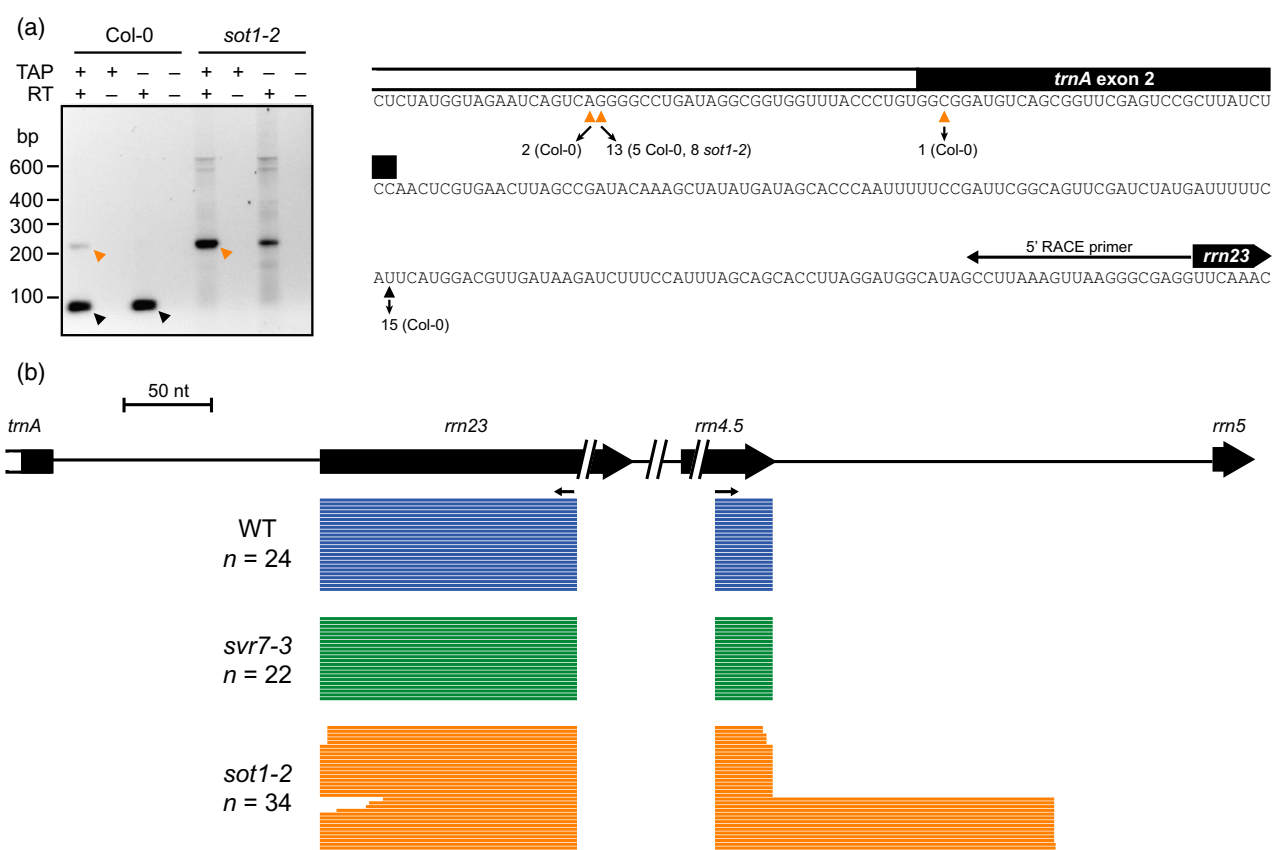

Figure 3. Transcript end mapping of the 23S rRNA precursor.
(a) 5′ RACE was performed for the *rrn23* precursor transcript. WT and *sot1–2* RNA were treated with or without tobacco acid pyrophosphatase (TAP) prior to adapter ligation and reverse transcription (RT). The PCR products generated were separated on an agarose gel, and the bands marked with arrowheads were gel-purified, cloned and sequenced. The positions of the sequenced 5′ ends are indicated by arrowheads, and numbers below the arrowheads indicate the numbers of clones obtained at that position.
(b) cRT–PCR analysis of 23S–4.5S precursors was performed on RNA derived from WT, *svr7–3* and *sot1–2* plants. Small black arrows indicate the position and direction of the cRT–PCR primer pair. Each coloured bar represents a single sequenced clone, and shows the position of the 5′ and 3′ ends relative to the mature 23S–4.5S dicistron, as seen for all clones sequenced for the WT and *svr7–3* samples.

dicistron, whereas all WT/*svr7–3* sequences mapped to the mature ends. Eleven sequenced clones from *sot1–2* had the same 5′ end as WT but a longer 3′ end, while the remaining nine had shorter 5′ and 3′ ends or shorter 5′ ends but longer 3′ ends compared to the mature precursor. These results indicate that loss of SOT1 leads to accumulation of both unprocessed and mis-processed 23S–4.5S dicistrons.

The results described so far suggest that SOT1 is involved in stabilization and/or correct processing of the 23S–4.5S precursor, but is this the direct target of SOT1? Given SOT1 is a PPR protein and prediction of putative binding sites is now possible (Barkan *et al.*, 2012; Takenaka *et al.*, 2013; Yagi *et al.*, 2013), potential targets of SOT1 were determined. Initially, a predicted binding site was generated using the amino acid residues at positions 5 and 35 of the first nine PPR motifs present in the SOT1 protein (Figure 4a). The last two PPR motifs were excluded from the predictions as they are unusual and it is not currently

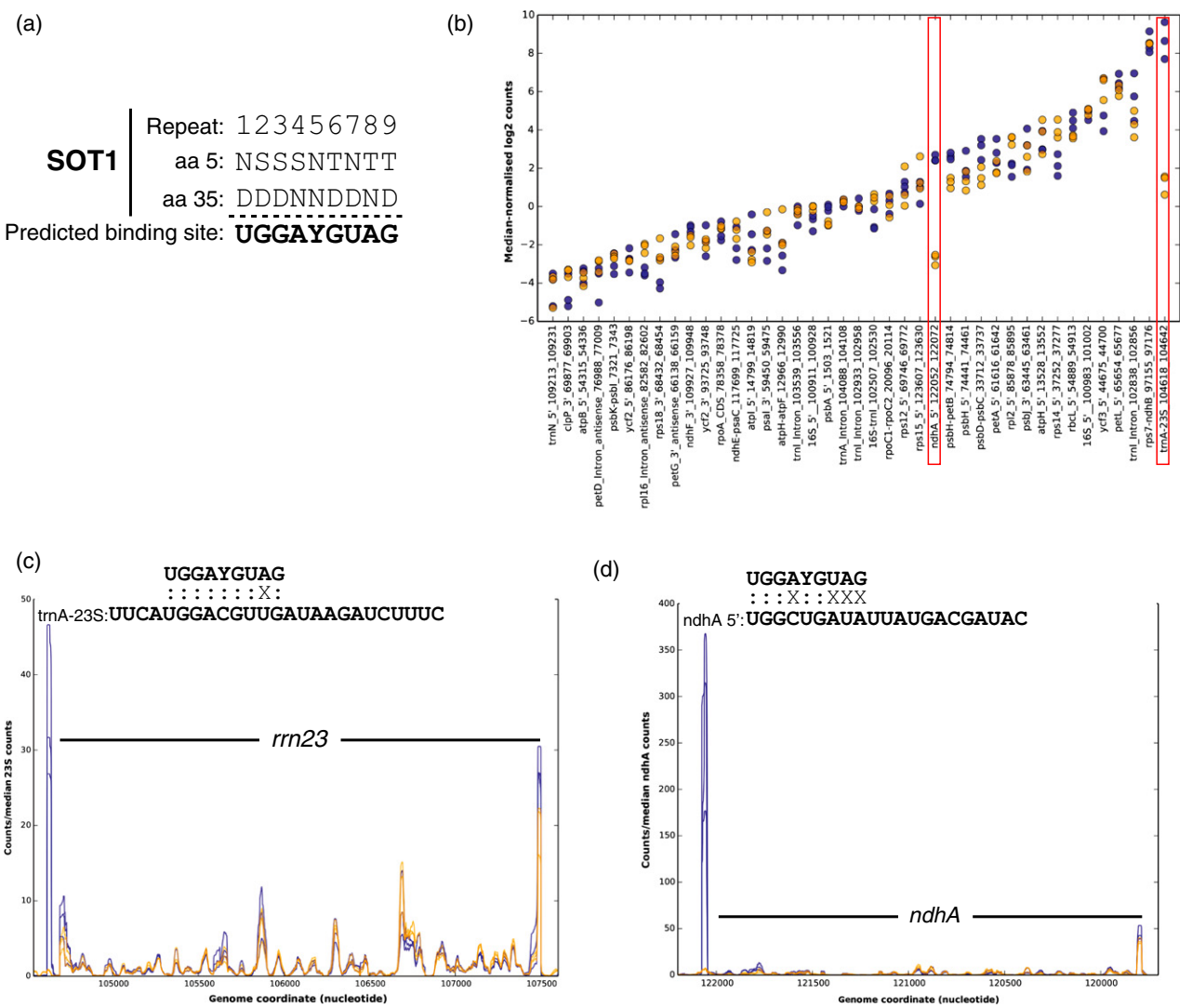

Figure 4. Identification of potential RNA targets of SOT1. (a) Putative binding sites of SOT1 were predicted based on the amino acid residues at positions 5 and 35 of the first nine PPR motifs in the protein. One of the putative binding sites lies within a putative PPR 'footprint' previously described by Ruwe and Schmitz-Linneweber (2012) in the region between *trnA* and *rrn23* (see Table S1).

(b) Putative PPR 'footprints' were analysed by sequencing small RNA libraries derived from WT and *sot1–2* mutant plants. Reads derived from small RNA sequencing (triplicate samples for both WT, shown in blue, and *sot1–2*, shown in orange) were mapped to the chloroplast genome, and putative PPR 'footprints' reported in Ruwe and Schmitz-Linneweber (2012) were analysed. Coverage of these 'footprints' was compared between the genotypes using BEDTools and by plotting $\log_2$(count) minus $\log_2$(median count) for all six samples. The coverage data were then ordered by maximum count for each 'footprint'. Red boxes highlight that the ndhA 5′ and trnA–23S 'footprints' are reduced in the *sot1–2* mutant compared to WT.

(c,d) Read coverage over the *rrn23* and *ndhA* regions for all six samples (blue, WT; orange, *sot1–2*). Coverage across the whole region was normalized using the median coverage across the mature *rrn23* and *ndhA* transcripts, and shows that loss of the ndhA 5′ and trnA–23S 'footprints' in the *sot1–2* mutant is far greater than may be explained by reduction in levels of the corresponding transcript. The sequences of the ndhA 5′ and trnA–23S 'footprints' are shown, and possible SOT1 binding sites are indicated (colons indicate a compatible match, X indicates an incompatible match).

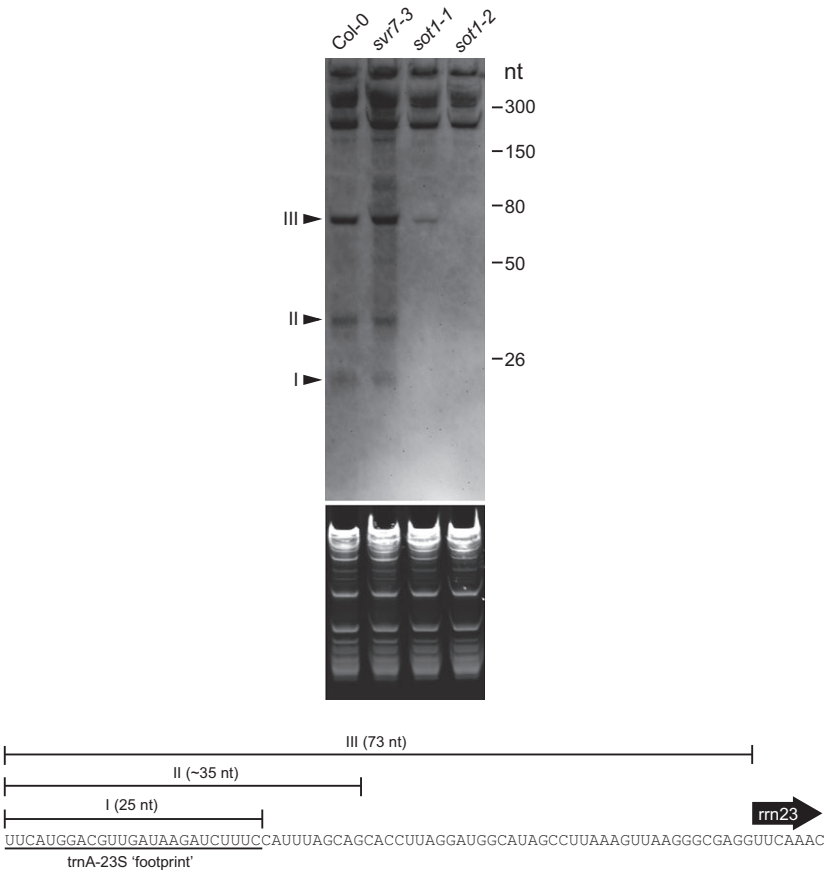

Figure 5. RNA gel blot analysis of the trnA–23S small RNA.
RNA (2.5 μg) from each genotype was separated on denaturing polyacrylamide gels and transferred to nylon membranes. Small RNAs were detected using a digoxigenin-labelled oligonucleotide antisense to the trnA–23S 'footprint'. An ethidium bromide-stained gel is shown to indicate equal loading. The arrowheads indicate three hybridization signals obtained with the *rrn23* 5′ probe. These probably represent: (I) the small RNA corresponding to the trnA–23S 'footprint' (25 nt), (II) a larger isoform of the small RNA with a 3′ extension, which has been found in small RNA sequencing datasets (approximately 35 nt), and (III) the 5′ leader sequence of the 23S ribosomal RNA (73 nt), as indicated in the lower panel.

possible to predict which nucleotides these repeats would recognize. Using this predicted binding site and allowing at most one base mismatch, matches across the chloroplast genome sequence were determined. From these, potential binding sites were identified based on the existing knowledge of PPR binding sites, specifically that they are likely to be found in intergenic regions close to transcript termini or within introns, and on the same strand as the coding sequence (Barkan and Small, 2014). Using these criteria, six putative targets of SOT1 were identified (Table S1), and included a putative binding site (Figure 4a) that not only matched the 5′ region of the 23S–4.5S rRNA but also a putative PPR 'footprint' that has previously been identified (Ruwe and Schmitz-Linneweber, 2012) in a small RNA sequencing dataset (Rajagopalan *et al.*, 2006).

To determine whether the SOT1 protein is linked to accumulation of this specific PPR 'footprint', we analysed PPR 'footprints' by sequencing small RNA libraries prepared from RNA that was extracted from WT and *sot1–2* mutant plants and size-selected. Reads generated from Illumina sequencing of triplicate samples were filtered, trimmed and mapped to the chloroplast genome. BEDTools (Quinlan and Hall, 2010) was used to calculate coverage of the PPR 'footprints' reported by Ruwe and Schmitz-Linneweber (2012) for the various genotypes, and revealed that the trnA–23S and ndhA 5′ 'footprints' were significantly reduced in the *sot1–2* mutant compared to WT (Figure 4b). Local analysis of coverage for the *rrn23* and *ndhA* regions (Figure 4c,d), where coverage across the region was normalized using the median coverage across the mature transcripts, confirmed massive loss of the trnA–23S and ndhA 5′ 'footprints' in the *sot1–2* mutant, and showed that it is far greater than can be explained by reduction in the level of these transcripts.

The reduction in the trnA–23S 'footprint' was also confirmed by RNA gel blot analysis. An oligonucleotide antisense to the trnA–23S 'footprint' was used to detect small RNAs in WT, *svr7–3*, *sot1–1* and *sot1–2* RNA samples (Figure 5). The hybridization signals obtained for WT and *svr7–3* were similar and showed three differently sized bands (I, II and III, Figure 5). These probably represent the small RNA corresponding to the trnA–23S 'footprint' (25 nt, band I), a larger isoform of the small RNA with a 3′ extension that is present in small RNA sequencing datasets (approximately 35 nt, band II) and the 5′ leader sequence of the 23S rRNA (73 nt, band III). All three hybridization signals were lost in *sot1–2*, but low levels of band III were detected in *sot1–1*.

SOT1 binds to the 5′ region of the 23S–4.5S rRNA precursor

As one of the predicted binding sites of SOT1 overlaps with a PPR 'footprint' in the 5′ region of the 23S–4.5S rRNA precusor, and small RNAs corresponding to this 'footprint' are lost in the *sot1* mutant, we set out to test whether SOT1 directly binds this sequence.

The mature form of the SOT1 protein (i.e. without the putative plastid transit peptide) was expressed in *E. coli* as a glutathione-*S*–transferase (GST) fusion protein. The SOT1 protein was purified by fast protein liquid chromatography (FPLC) and detected using SDS–PAGE as a band at approximately 100 kDa (Figure S6). The E10 and E11 fractions were concentrated and then used for RNA electrophoretic mobility shift assays (REMSA) using labelled synthetic RNAs that correspond to various sequences based on the trnA–23S 'footprint' (Figure 6a). The purified recombinant SOT1 was first incubated with a biotin-labelled RNA probe P1 (2 nM) that corresponds to the trnA–23S 'footprint'. The SOT1–RNA complex was detected as a band that migrated more slowly than the free RNA, and the signal strength of the retarded band was in proportion to the level of recombinant SOT1 protein added to the reaction solution (Figure 6b), indicating that SOT1 binds the trnA–23S 'footprint'.

To confirm the sequence specificity of SOT1 binding to P1, unlabelled RNA with the same sequence (C1, 1 μM) was used as an RNA competitor. When the competitor was added to the reaction solution containing P1 and SOT1, the protein–RNA complex was not detectable (Figure 6b).

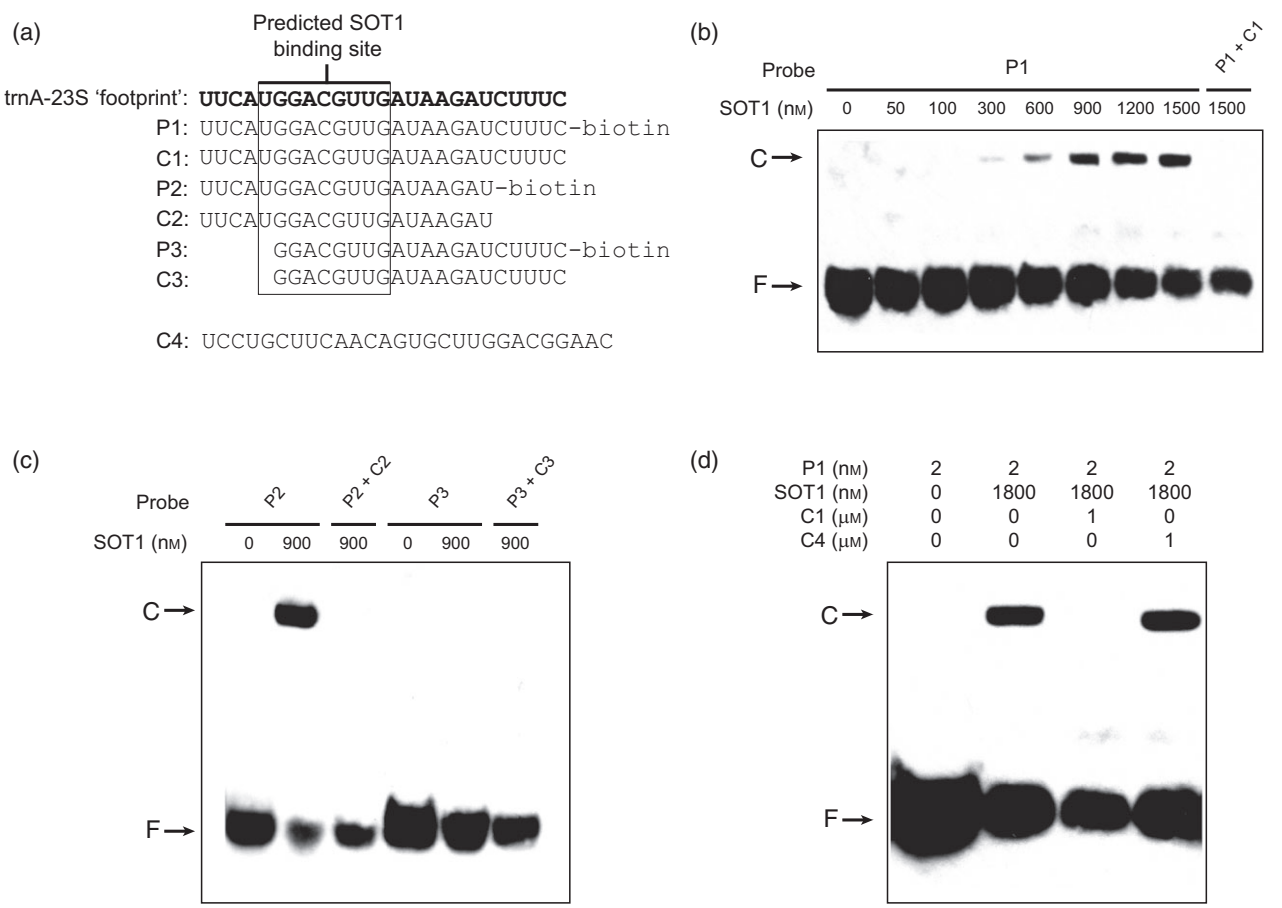

Figure 6. SOT1 binds to the 5′ region of the 23S–4.5S precursor.
(a) Schematic representation of RNA sequences used for *in vitro* binding assays with respect to the trnA–23S 'footprint' and the predicted SOT1 binding site.
(b) REMSA of SOT1 binding to 2 nM of biotin-labeled P1. The concentration of the recombinant SOT1 protein is indicated above each lane. The positions of protein–RNA complexes and free RNA are indicated by C and F, respectively. In the last lane, unlabelled RNA C1 (1 μM) was added to compete for binding between SOT1 and P1.
(c) A 5 nt deletion at the 5′ end of the trnA–23S 'footprint' affects SOT1 binding to RNA. P2 has a 5 nt deletion at the 3′ end and P3 has a 5 nt deletion at the 5′ end. In the competitive assay, the concentration of the labelled probe (P2 and P3) is 2 nM and the concentration of unlabelled probe is 1 μM. The positions of protein–RNA complexes and free RNA are indicated by C and F, respectively.
(d) Analysis of the binding specificity of SOT1. Unlabelled RNA corresponding to the trnA–23S 'footprint' (C1) competes with P1 for binding to SOT1, but a competitor with a sequence unrelated to the 'footprint' sequence (C4) does not.

These results indicate that SOT1 specifically binds to its target. We also tested SOT1–RNA binding affinity using RNAs with a 5 nt deletion at either the 3′ end (P2) or the 5′ end (P3) of the 'footprint'. Our results show that the interaction between SOT1 and the 'footprint' was still detected if the 5 nt at the 3′ end were deleted, but was completely abolished if the 5 nt at the 5′ end were removed, indicating that the predicted core sequence is critical for the interaction (Figure 6c). Furthermore, we showed that, when a competitor with a sequence unrelated to the 'footprint' sequence (C4) was added, the protein–RNA complex was unaffected (Figure 6d), providing additional support for sequence specificity of SOT1 binding to its target. Taken together, these results suggest that SOT1 binds to the 5′ region of the 23S–4.5S precursor in a sequence-specific manner.

DISCUSSION

SOT1 is involved in chloroplast ribosome biogenesis

The chloroplast ribosome resembles prokaryotic 70S ribosomes with regard to function, composition and structure, and comprises a small 30S subunit and a large 50S subunit (Tiller and Bock, 2014). The biogenesis of ribosomes is a complicated and finely regulated process, which begins with transcription of the ribosomal gene operon, followed by the coordinated action of many proteins involved in rRNA processing, and assembly of the rRNA with ribosomal proteins (Kaczanowska and Ryden-Aulin, 2007). In bacteria, the three rRNAs are transcribed as a single large precursor, which undergoes co-transcriptional folding, processing and modification to form the mature rRNAs (Kaczanowska and Ryden-Aulin, 2007; Gutgsell and Jain, 2010). Although the chloroplast 50S subunit contains three rRNAs (23S, 4.5S and 5S) compared to two rRNAs (23S and 5S) in bacteria, the plastid 23S and 4.5S rRNAs are functionally equivalent to the 5′- and 3′-terminal regions of the bacterial 23S rRNA, respectively (Edwards and Kossel, 1981; Machatt *et al.*, 1981).

In higher plants, RNA-binding proteins that are essential for maturation of plastid rRNAs have been identified and probably play an important role in defining the cleavage site for RNases and protecting rRNA from exoribonucleolytic digestion, as RNases usually lack sequence specificity. In the model plant Arabidopsis, the chloroplast stem-loop binding proteins (CSP41a/b) have been suggested to be involved in the 5′ end maturation of 23S rRNA, and a DEAD box protein, RH39, has been suggested to be involved in introduction of the hidden break in the 23S rRNA (Beligni and Mayfield, 2008; Nishimura *et al.*, 2010). RHON1, which interacts with RNase E, was reported to bind the inter-cistronic region of the 23S–4.5S dicistronic rRNA precursor via its C–terminal Rho–N domain, and is presumed to confer sequence specificity to RNase E (Stoppel *et al.*, 2012). Recently, an octotricopeptide repeat protein, RAP, was found to be required for chloroplast 16S rRNA maturation, most likely by defining the trimming site of RNase J (Kleinknecht *et al.*, 2014). However, until now, there have been no reports of PPR proteins being directly involved in plastid rRNA maturation, despite their clear role in defining mRNA termini (Prikryl *et al.*, 2011).

The 23S–4.5S dicistronic rRNA precursor is released by cleavage of the primary polycistronic transcript of the plastidic *rrn* operon by an uncharacterized endoribonuclease, and its 5′ and 3′ ends are then processed to generate the mature 23S–4.5S rRNA (Edwards and Kossel, 1981; Walter *et al.*, 2002; Bollenbach *et al.*, 2005). This mature dicistron is further endonucleotytically cleaved into the 23S and 4.5S precursors (Bollenbach *et al.*, 2005). Recent studies reported that chloroplast mini-ribonuclease III participates in rRNA maturation, including that of the 23S–4.5S precursor, and intron recycling (Hotto *et al.*, 2015), and that loss of the endoribonuclease YbeY also affects plastid rRNA maturation (Liu *et al.*, 2015), but the precise steps involved in maturation of the unique plastid 23S–4.5S dicistronic precursor remain unclear.

Chloroplast RNA processing has been reported to be perturbed in many mutants that are defective in ribonucleases and RNA helicases (*dal*, *dcl*, *rhon1*, *rnr1*, *rh3* and *rh22*) (Bisanz *et al.*, 2003; Bellaoui and Gruissem, 2004; Bollenbach *et al.*, 2005; Asakura *et al.*, 2012; Chi *et al.*, 2012; Stoppel *et al.*, 2012; Liu *et al.*, 2015), ribosomal proteins (Liu *et al.*, 2013b; Wu *et al.*, 2013; Ma *et al.*, 2015), and factors involved in rRNA maturation and modification (Yu *et al.*, 2008; Liu *et al.*, 2010b, 2013b; Wu *et al.*, 2013; Kleinknecht *et al.*, 2014; Han *et al.*, 2015). However, in many cases, these irregularities in rRNA processing are probably due to secondary effects of the mutation. This study provides evidence for a direct role of the PPR-SMR protein, SOT1, in binding the 5′ region of the 23S–4.5S rRNA intermediate and protecting it from exonucleolytic attack. We have shown that rRNA processing is defective in the *sot1* mutant lines, and that the predicted SOT1 binding site corresponds to the 5′ end mapped by 5′ RACE and cRT–PCR as well as a PPR 'footprint' which is missing in the mutants. These results, combined with the observation that purified SOT1 binds to the 'footprint' *in vitro*, strongly support a direct role for SOT1 in chloroplast rRNA maturation and thus in chloroplast ribosome biogenesis. Our approach also highlights the power of combining binding site predictions, using the PPR 'code' (Barkan *et al.*, 2012), with a high-throughput method of screening PPR 'footprints' using small RNA analysis (Ruwe and Schmitz-Linneweber, 2012; Zhelyazkova *et al.*, 2012; Loizeau *et al.*, 2014), to discover novel PPR protein–RNA interactions.

Based on the knowledge gained from this study, as well as the recent report on the role of mini-ribonuclease III in rRNA maturation (Hotto *et al.*, 2015), a model for SOT1

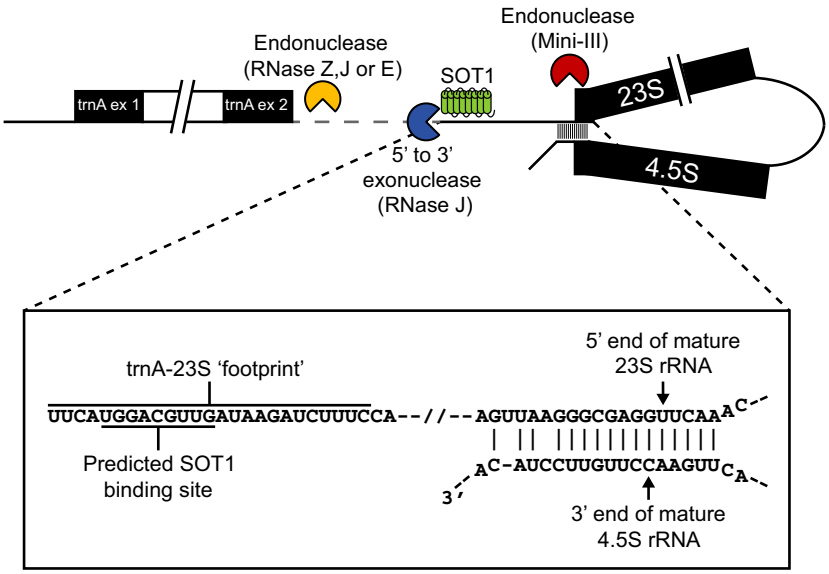

Figure 7. Model of SOT1 action on the chloroplast 23S–4.5S rRNA precursor.
The endonuclease activity of RNase Z, J or E creates entrance sites for RNase J 5′→3′ exonucleolytic activity upstream of the SOT1 binding site. When SOT1 is present, RNase J is blocked 73 nt upstream of the 5′ end of the mature 23S rRNA, allowing proper 5′ maturation of 23S and 3′ maturation of 4.5S by mini-ribonuclease III. If SOT1 is missing, RNase J progresses into the 23S rRNA, causing rRNA processing defects as observed in this study.

action on the chloroplast 23S–4.5S rRNA precursor is proposed (Figure 7). By binding to the 23S–4.5S rRNA precursor, SOT1 protects the RNA from exonucleolytic processing for a sufficient amount of time for the RNA to fold back on itself to form the double-stranded structure required for processing by mini-ribonuclease III. The disruption to mini-ribonuclease III processing is evident from our results. The cRT–PCR results (Figure 3b) demonstrate that both 5′ and 3′ processing of the 23S–4.5S dicistron are disrupted in the *sot1* mutant (but not in the similar mutant *svr7*), indicating disruption of simultaneous processing of both ends, catalysed by mini-ribonuclease III. Further direct evidence is shown in Figure 5, in which the approximately 75 nt RNA fragment that hybridizes to the trnA–23S 'footprint' probe presumably represents the 23S 5′ extension that is endonucleolytically removed by mini-ribonuclease III, and, whilst relatively abundant in wild-type and *svr7* RNA, is much reduced in *sot1–1* and absent in *sot1–2*. Thus, we propose that SOT1 not only protects the 23S–4.5S precursor, but also facilitates endonucleolytic processing of both its 5′ and 3′ ends.

SOT1 may have additional roles in plastid gene expression

The *sot1* mutants show additional perturbations to chloroplast RNA processing (Figure 4 and Figure S4). Here we have described the role of SOT1 in rRNA processing, but cannot rule out the possibility that it has additional functions. In particular, our systematic analyses of plastid transcripts and small RNA 'footprints' consistently demonstrated an effect on *ndhA* transcripts, that we have not fully evaluated as the sequence of the small RNA lacks similarity to the predicted binding site of SOT1 (Figure 4d) and loss of *ndhA* expression does not explain the growth and molecular phenotype of the *sot1* mutant (Figures 1 and 2). Moreover, Zoschke *et al.* (2016) have shown that the SOT1 ortholog in maize, PPR53, does not bind synthetic RNA corresponding to the ndhA 5′ putative PPR 'footprint', and propose that the effect on *ndhA* expression must involve an additional protein. Therefore, whilst acknowledging that SOT1 may have additional roles in plastids, we have focused on its action in processing 23S–4.5S precursor RNAs, as we believe this is its major and direct function. This is supported by the fact that SOT1 is one of most abundant PPRs in plastids, as judged by spectral counts from mass spectrometry (Liu *et al.*, 2013a), and the trnA–23S 'footprint' that it binds is also one of the most abundant small RNAs in the organelle, present at levels that are approximately 250-fold higher than the level of most proposed PPR 'footprints' (Figure 4b).

Does SOT1 act as an endoribonuclease?

Although it seems most plausible that the endonucleolytic processing of the 23S–4.5S dicistron is catalysed by the mini-ribonuclease III, it is worth considering whether SOT1 may play a direct role in catalysing this event. SOT1 contains an SMR domain (Liu *et al.*, 2013a), which has been shown to exhibit nicking endonuclease activity in other proteins (Fukui and Kuramitsu, 2011). The mRNA cycling sequence binding protein from *Leishmania donovani* is one of the few SMR-containing RNA-binding proteins to have been studied. Its SMR domain was reported to possess endoribonuclease activity (Bhandari *et al.*, 2011). However, to date, there are no reports of plant SMR domain proteins with endoribonuclease activity. The mutation in *sot1–1* is within the SMR domain, but we have not been able to ascertain whether the effect of the mutation is limited to the action of the SMR domain or whether it affects the activity or stability of the whole protein. In the

future, it will be important to investigate whether the plant SMR domain has any endonuclease activity, or whether it acts as a protein–protein or protein–nucleic acid interaction domain.

Parallels with other plastid PPR-SMR proteins

SOT1 is the fourth of the Arabidopsis chloroplast-localized PPR-SMR proteins to be studied in detail. Despite their similar structure and clear evolutionary relatedness (Figure S7) (Liu *et al.*, 2013a), these four proteins appear to play very different roles in plastid gene expression. PTAC2 is required for plastid RNA polymerase function (Pfalz *et al.*, 2006), GUN1 is involved in retrograde signalling and suppression of nuclear gene expression in response to plastid dysfunction (Koussevitzky *et al.*, 2007), and SVR7 is required for correct expression of the plastid ATP synthase (Zoschke *et al.*, 2013a), as is its maize orthologue, ATP4 (Zoschke *et al.*, 2012). These apparent differences hide some intriguing parallels, particularly between SOT1 and SVR7. Both were initially found as *suppressor of variegation* mutants (Liu *et al.*, 2010b; this study), both are apparently highly abundant (relative to other PPR proteins) (Liu *et al.*, 2013a), both affect rRNA processing (Liu *et al.*, 2010b; this study), and both are predicted to bind very similar, even overlapping, sequences: the predicted binding site of SOT1 is UGGAYGUAG, that of SVR7 is UGGAYRURG.

We previously reported that *thf1* variegation may be attributed to a reduced level of FtsH protease activity (Zhang *et al.*, 2009). In this study, we provide genetic evidence suggesting that *var2* and *thf1* share the same mechanism controlling leaf variegation. The *sot1* mutant is able to suppress *var2* leaf variegation (Figure 1f), and the previously identified *var2* suppressor *svr7* (Liu *et al.*, 2010b) also suppresses *thf1* leaf variegation (Figure S8). To date, a number of suppressors have been identified from genetic screens for the *thf1* and *var2* leaf variegation phenotype (Park and Rodermel, 2004; Miura *et al.*, 2007; Yu *et al.*, 2008, 2011; Liu *et al.*, 2010a,b, 2013b; Wu *et al.*, 2013; Hu *et al.*, 2015; Ma *et al.*, 2015). Interestingly, many of these variegation suppressors have defects in chloroplast rRNA processing, suggesting a link between plastid ribosome biogenesis and leaf variegation. On this basis, an undiscovered direct role for SVR7 in some aspect of ribosome biogenesis appears possible.

Stretching the parallels further, leaf variegation must involve some aspect of plastid retrograde signalling, with nuclear expression of plastid genes involved in photosynthesis being shut down in the white sectors (Kato *et al.*, 2007). Viewed in this way, the action of the *sot1* and *svr7* mutations to suppress leaf variegation in response to certain forms of chloroplast dysfunction resembles the action of the *gun1* mutation in suppressing the nuclear gene expression response to other forms of plastid dysfunction. Although neither *svr7* nor *sot1* show a *gun* phenotype under the classical conditions used to discover *gun1* (Figure S3), it is tempting to speculate that further research will discover some unifying mode of action that links these PPR-SMR proteins.

EXPERIMENTAL PROCEDURES

Plant materials and growth conditions

Arabidopsis thaliana Col–0 plants were used as the wild-type control for this work, except for map-based cloning, for which Landsberg *erecta* (L*er*) was used. Arabidopsis T–DNA insertion lines were obtained from the Nottingham Arabidopsis Stock Centre (http://arabidopsis.info). *sot1–2* corresponds to the GABI-Kat T–DNA insertion line GK_840D06 (Kleinboelting *et al.*, 2012; stock ID N480586) and *svr7–3* corresponds to the SAIL line SAIL_423_G09 (Sessions *et al.*, 2002; stock ID N819547) that has been described previously (Zoschke *et al.*, 2013a). Homozygous plants were identified by PCR-based genotyping using the primers listed in Table S2, and locations of the insertion were confirmed by sequencing. The *sot1–1* mutant was isolated from *thf1* suppressors that had been generated by EMS mutagenesis and screening of 8000 independent M_2 progeny (see below). Double mutants were generated by standard crossing procedures. Seeds generated from crossings were germinated and the plants were allowed to self-fertilize. Double homozygotes were selected from the resulting progeny.

Seeds were surface-sterilized and stratified at 4°C for 2–3 days, and then germinated on either half-strength Murashige and Skoog (MS) medium containing 1% sucrose, or directly on soil, and then grown under long-day conditions (16 h light/8 h dark) at 22°C with a light intensity of 100 μmol photons m^{-2} sec^{-1}, unless otherwise specified.

Map-based cloning, plasmid constructs and generation of transgenic plants

F_2 populations from the cross between *40–1* and L*er* were used for mapping the *sot1–1* locus. F_2 seeds were germinated on MS plates, and seedlings without leaf variegation were transferred to soil for further growth. Genomic DNA was isolated from *40–1* lines in which the *thf1* background was identified by PCR. Linkage analysis was performed using SSLP markers (http://www.arabidopsis.org), and fine mapping was undertaken using the primers listed in Table S2. Plasmid construction and plant transformation for complementation testing were performed as previously described (Zhou *et al.*, 2008).

Northern blotting

Northern blotting was performed as previously described (Chateigner-Boutin *et al.*, 2011). The primers used to amplify the 16S, 25S 5′, 25S 3′, 4.5S and 5S rRNA sequences for cloning and subsequent probe generation are given in Table S2.

Protein analyses

Total protein was extracted from the aerial tissue of 2-week-old seedlings, separated by SDS–PAGE using a 4-15% TGX Stain-Free gel (Bio–Rad, http://www.bio-rad.com/) and transferred onto a poly(vinylidene difluoride) membrane (GE Healthcare Life Sciences, http://www.gelifesciences.com/). Protein loading was checked by visualizing the gels on a Gel Doc™ EZ imager (Bio–Rad). Western blotting was performed using anti-RPS1 (1:1500; Uniplastomic, http://www.uniplastomic.com/), anti-RPL4 (1:1500; Uniplastomic), anti-AtpE (1:5000; provided by Alice Barkan,

Department of Biology, University of Oregon) and anti-PsbA antibodies (1:50 000, Agrisera, http://www.agrisera.com/). Anti-rabbit antibodies conjugated to horseradish peroxidase (Sigma-Aldrich, http://www.sigmaaldrich.com/) was used as a secondary antibody, and a chemiluminescent signal was generated using ECL reagents (Bio–Rad).

Rapid amplification of cDNA ends (RACE)

Total RNA was isolated using a Qiagen miRNeasy kit (https://www.qiagen.com/). Duplicate samples were treated with and without tobacco acid pyrophosphatase (Epicentre, http://www.epibio.com/) to allow ligation of the adapter RNA by converting 5′ triphosphorylated primary transcript ends into monophosphate ends. After phenol/chloroform extraction and ethanol precipitation, RNA was ligated with T4 RNA ligase 1 (NEB, https://www.neb.com/) to the 5′ SRA adapter (NEB). cDNA was synthesized using random primers and SuperScript III reverse transcriptase (Thermo Fisher Scientific, http://www.thermofisher.com/). 5′ RACE was performed using the primers listed in Table S2. PCR products were gel-eluted, cloned into pGEM®-T Easy (Promega, http://www.promega.com/) and sequenced.

cRT–PCR

Total RNA was isolated using the RNAgents system (Promega), and 5 μg of total RNA were circularized using T4 RNA ligase (Thermo Fisher Scientific) as described by Perrin *et al.* (2004). After phenol/chloroform extraction and ethanol precipitation, cDNA was synthesized using a reverse transcription system (Promega) and specific primers (Table S2). The region containing the junction of the 5′ and 3′ ends was then amplified, cloned into PCR4 TOPO vectors (Thermo Fisher Scientific), and sequenced.

Prediction of RNA binding sites

Possible RNA binding sequences of SOT1 were generated according to the 'code' governing PPR–nucleotide interactions described by Barkan *et al.* (2012). The nucleotide patterns representing the putative target sequences were searched against the Arabidopsis chloroplast genome (NC_000932, http://www.ncbi.nlm.nih.gov/nuccore/7525012) using FUZZNUC from the European Molecular Biology Open Software Suite (EMBOSS) (Rice *et al.*, 2000).

Small RNA sequencing and gel blot analysis

To retain small RNAs, RNA was extracted from frozen, ground seedling tissue using a Qiagen miRNeasy kit including the optional homogenization step using QIAshdredder spin columns (Qiagen). RNA was denatured prior to being separated by size on a 12% urea polyacrylamide gel, before either being transferred to a Hybond N+ membrane (GE Healthcare Life Sciences) or being used to obtain small RNAs for library preparation (see below). After UV cross-linking of the RNA to the membrane (240 mJ/cm^2), the membrane was hybridized at 60°C to a digoxigenin-labelled LNA oligonucleotide probe antisense to the trnA–23S 'footprint' (2.5 nM in PerfectHyb™ Plus hybridization buffer, Sigma-Aldrich) for 48 h. The membrane was washed and the digoxigenin-labelled probe detected as previously described (Kim *et al.*, 2010). LNA oligonucleotides were purchased from Exiqon (http://www.exiqon.com/), and labelled using a DIG oligonucleotide tailing kit (Roche, http://www.roche.com/index.htm).

To prepare small RNA libraries, the size-selected RNA (15-50 nt) was extracted from gel pieces by passive elution, ethano-precipitated overnight and then used in the NEBNext® Multiplex Small RNA Library Prep Set for Illumina (NEB) according to the manufacturer's instructions. Libraries were pooled, quantified and then sequenced (1 x 61 bp) on a HiSeq 1500 system (Illumina, https://www.illumina.com/) using the rapid-run output mode with on-board cluster generation. Sequencing reads were filtered (> 15 bp) and trimmed of any adapter sequenced using cutadapt (Martin, 2011). Reads were mapped to the Col–0 plastid genome (with one copy of the inverted repeat removed) using Bowtie 2 version 2.2.5 (Langmead and Salzberg, 2012) with parameter -L 30 to ensure stringency. Strand-specific read coverage of the chosen segments of the genome was calculated using BEDTools version 2.25.0 (Quinlan and Hall, 2010).

Protein expression, purification and RNA electrophoretic mobility shift assay (REMSA)

The sequence encoding mature SOT1 was cloned into the expression vector PDEST15 (Thermo Fisher Scientific), and expression of the protein was induced using 0.5 mM isopropyl-β–D–thiogalactopyranoside at 28°C for 6 h in *E. coli* strain Rosetta. The protein was purified as described by Cai *et al.* (2011). REMSA experiments were performed using a LightShift™ chemiluminescent RNA EMSA kit (Thermo Fisher Scientific) according to the manufacturer's instructions. All the biotin-labelled probes (P1, P2 and P3) were used at 2 nM, and the unlabelled competitors and control RNA (C1, C2, C3 and C4) were used at 1 μM. SOT1 protein and RNAs were incubated at room temperature for 30 min in the reaction solution (10 mM HEPES pH 7.3, 20 mM KCl, 1 mM $MgCl_2$, 1 mM dithiothreitol, 5% glycerol and 2 μg tRNA). The reactants were analysed using a native polyacrylamide gel, and signals were detected according to the manufacturer's instructions.

Chlorophyll extraction, generation of transgenic plants for GFP localization and GUS expression, gun phenotype test, and quantitative RT–PCR

Chlorophyll extraction, generation of transgenic plants for GFP localization and GUS expression, gun phenotype test, and quantitative RT–PCR were performed as described in Appendix S1.

ACKNOWLEDGEMENTS

We thank Steven Rodermel (Department of Botany, Iowa State University) for providing the *var2* seeds, Alice Barkan (Department of Biology, University of Oregon) for providing the AtpE antibody, and Reimo Zoschke (Department of Organelle Biology, Biotechnology and Molecular Ecophysiology, Max Planck Institute of Molecular Plant Physiology) and Alice Barkan for helpful discussions and sharing data. This work was supported by funds from the National Science Foundation of China (31570231), the 973 Program (2015CB150104) and the National Special Grant for Transgenic Crops (2015ZX08009-003-005) to J.H., an Australian Research Council Discovery Early Career Researcher Award (DE120101117) to K.A.H. and an Australian Research Council Centre of Excellence award (CE140100008) to I.D.S. S.L. was the recipient of a joint PhD scholarship from the China Scholarship Council and the University of Western Australia. H.R. was supported by a PhD stipend from the Deutscher Akademischer Austauschdienst.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Mutation in *SOT1* rescues the *thf1* phenotype.

Figure S2. Expression patterns of *SOT1*.

Figure S3. Quantitative PCR analysis of photosynthesis-associated nuclear genes to determine whether the *sot1* mutant has a *gen-*

omes uncoupled (*gun*) phenotype after norflurazon treatment.
Figure S4. Quantitative RT–PCR was performed to assess transcript levels of all protein-coding genes encoded in the plastid genome in the *sot1–1* and *sot1–2* mutant background.
Figure S5. Changes to chloroplast rRNA transcripts observed in *sot1–1* are also seen in *sot1–2*.
Figure S6. Purification of the recombinant SOT1 protein used for RNA binding assays.
Figure S7. Alignment of PPR-SMR proteins in Arabidopsis.
Figure S8. *svr7* rescues *thf1* leaf variegation.
Table S1. Predicted binding sites of SOT1.
Table S2. Primers used in this study.
Appendix S1. Supplementary methods.

REFERENCES

Asakura, Y., Galarneau, E., Watkins, K.P., Barkan, A. and van Wijk, K.J. (2012) Chloroplast RH3 DEAD box RNA helicases in maize and Arabidopsis function in splicing of specific group II introns and affect chloroplast ribosome biogenesis. *Plant Physiol.* **159**, 961–974.

Aubourg, S., Boudet, N., Kreis, M. and Lecharny, A. (2000) In *Arabidopsis thaliana*, 1% of the genome codes for a novel protein family unique to plants. *Plant Mol. Biol.* **42**, 603–613.

Barkan, A. (2011) Expression of plastid genes: organelle-specific elaborations on a prokaryotic scaffold. *Plant Physiol.* **155**, 1520–1532.

Barkan, A. and Small, I. (2014) Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.* **65**, 415–442.

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

Beligni, M.V. and Mayfield, S.P. (2008) *Arabidopsis thaliana* mutants reveal a role for CSP41a and CSP41b, two ribosome-associated endonucleases, in chloroplast ribosomal RNA metabolism. *Plant Mol. Biol.* **67**, 389–401.

Bellaoui, M. and Gruissem, W. (2004) Altered expression of the Arabidopsis ortholog of DCL affects normal plant development. *Planta*, **219**, 819–826.

Bhandari, D., Guha, K., Bhaduri, N. and Saha, P. (2011) Ubiquitination of mRNA cycling sequence binding protein from *Leishmania donovani* (LdCSBP) modulates the RNA endonuclease activity of its Smr domain. *FEBS Lett.* **585**, 809–813.

Bisanz, C., Begot, L., Carol, P., Perez, P., Bligny, M., Pesey, H., Gallois, J.L., Lerbs-Mache, S. and Mache, R. (2003) The Arabidopsis nuclear DAL gene encodes a chloroplast protein which is required for the maturation of the plastid ribosomal RNAs and is essential for chloroplast differentiation. *Plant Mol. Biol.* **51**, 651–663.

Bollenbach, T.J., Lange, H., Gutierrez, R., Erhardt, M., Stern, D.B. and Gagliardi, D. (2005) RNR1, a 3′→5′ exoribonuclease belonging to the RNR superfamily, catalyzes 3′ maturation of chloroplast ribosomal RNAs in *Arabidopsis thaliana*. *Nucleic Acids Res.* **33**, 2751–2763.

Cai, W., Okuda, K., Peng, L. and Shikanai, T. (2011) PROTON GRADIENT REGULATION 3 recognizes multiple targets with limited similarity and mediates translation and RNA stabilization in plastids. *Plant J.* **67**, 318–327.

Chateigner-Boutin, A.L., des Francs-Small, C.C., Delannoy, E., Kahlau, S., Tanz, S.K., de Longevialle, A.F., Fujii, S. and Small, I. (2011) OTP70 is a pentatricopeptide repeat protein of the E subgroup involved in splicing of the plastid transcript *rpoC1*. *Plant J.* **65**, 532–542.

Chen, M., Choi, Y., Voytas, D.F. and Rodermel, S. (2000) Mutations in the Arabidopsis VAR2 locus cause leaf variegation due to the loss of a chloroplast FtsH protease. *Plant J.* **22**, 303–313.

Chi, W., He, B., Mao, J., Li, Q., Ma, J., Ji, D., Zou, M. and Zhang, L. (2012) The function of RH22, a DEAD RNA helicase, in the biogenesis of the 50S ribosomal subunits of Arabidopsis chloroplasts. *Plant Physiol.* **158**, 693–707.

de Dios Barajas-Lopez, J., Blanco, N.E. and Strand, A. (2013) Plastid-to-nucleus communication, signals controlling the running of the plant cell. *Biochim. Biophys. Acta*, **1833**, 425–437.

Edwards, K. and Kossel, H. (1981) The rRNA operon from *Zea mays* chloroplasts: nucleotide sequence of 23S rDNA and its homology with *E. coli* 23S rDNA. *Nucleic Acids Res.* **9**, 2853–2869.

Emanuelsson, O., Nielsen, H., Brunak, S. and von Heijne, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005–1016.

Fukui, K. and Kuramitsu, S. (2011) Structure and function of the small MutS-related domain. *Mol Biol Int* **2011**, 691735.

Fukui, K., Kosaka, H., Kuramitsu, S. and Masui, R. (2007) Nuclease activity of the MutS homologue MutS2 from *Thermus thermophilus* is confined to the Smr domain. *Nucleic Acids Res.* **35**, 850–860.

Gutgsell, N.S. and Jain, C. (2010) Coordinated regulation of 23S rRNA maturation in *Escherichia coli*. *J. Bacteriol.* **192**, 1405–1409.

Hammani, K. and Giege, P. (2014) RNA metabolism in plant mitochondria. *Trends Plant Sci.* **19**, 380–389.

Han, J.H., Lee, K., Lee, K.H., Jung, S., Jeon, Y., Pai, H.S. and Kang, H. (2015) A nuclear-encoded chloroplast-targeted S1 RNA-binding domain protein affects chloroplast rRNA processing and is crucial for the normal growth of *Arabidopsis thaliana*. *Plant J.* **83**, 277–289.

Hotto, A.M., Castandet, B., Gilet, L., Higdon, A., Condon, C. and Stern, D.B. (2015) Arabidopsis chloroplast mini-ribonuclease III participates in rRNA maturation and intron recycling. *Plant Cell*, **27**, 724–740.

Hu, F., Zhu, Y., Wu, W., Xie, Y. and Huang, J. (2015) Leaf variegation of *Thylakoid formation1* is suppressed by mutations of specific σ–factors in Arabidopsis. *Plant Physiol.* **168**, 1066–1075.

de Jong, A., Pietersma, H., Cordes, M., Kuipers, O.P. and Kok, J. (2012) PePPER: a webserver for prediction of prokaryote promoter elements and regulons. *BMC Genom.* **13**, 299.

Kaczanowska, M. and Ryden-Aulin, M. (2007) Ribosome biogenesis and the translation process in *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* **71**, 477–494.

Kato, Y., Miura, E., Matsushima, R. and Sakamoto, W. (2007) White leaf sectors in *yellow variegated2* are formed by viable cells with undifferentiated plastids. *Plant Physiol.* **144**, 952–960.

Kim, S.W., Li, Z., Moore, P.S., Monaghan, A.P., Chang, Y., Nichols, M. and John, B. (2010) A sensitive non-radioactive Northern blot method to detect small RNAs. *Nucleic Acids Res.* **38**, e98.

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

Kleinboelting, N., Huep, G., Kloetgen, A., Viehoever, P. and Weisshaar, B. (2012) GABI-Kat SimpleSearch: new features of the Arabidopsis thaliana T-DNA mutant database. *Nucleic Acids Res.* **40**, D1211–D1215.

Kleinknecht, L., Wang, F., Stube, R., Philippar, K., Nickelsen, J. and Bohne, A.V. (2014) RAP, the sole octotricopeptide repeat protein in Arabidopsis, is required for chloroplast 16S rRNA maturation. *Plant Cell*, **26**, 777–787.

Koussevitzky, S., Nott, A., Mockler, T.C., Hong, F., Sachetto-Martins, G., Surpin, M., Lim, J., Mittler, R. and Chory, J. (2007) Signals from chloroplasts converge to regulate nuclear gene expression. *Science*, **316**, 715–719.

Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.

Liu, X., Rodermel, S.R. and Yu, F. (2010a) A *var2* leaf variegation suppressor locus, *SUPPRESSOR OF VARIEGATION3*, encodes a putative chloroplast translation elongation factor that is important for chloroplast development in the cold. *BMC Plant Biol.* **10**, 287.

Liu, X., Yu, F. and Rodermel, S. (2010b) An Arabidopsis pentatricopeptide repeat protein, SUPPRESSOR OF VARIEGATION7, is required for FtsH-mediated chloroplast biogenesis. *Plant Physiol.* **154**, 1588–1601.

Liu, S., Melonek, J., Boykin, L.M., Small, I. and Howell, K.A. (2013a) PPR-SMRs: ancient proteins with enigmatic functions. *RNA Biol.* **10**, 1501–1510.

Liu, X., Zheng, M., Wang, R., Wang, R., An, L., Rodermel, S.R. and Yu, F. (2013b) Genetic interactions reveal that specific defects of chloroplast translation are associated with the suppression of *var2*-mediated leaf variegation. *J. Integr. Plant Biol.* **55**, 979–993.

Liu, J., Zhou, W., Liu, G. *et al.* (2015) The conserved endoribonuclease YbeY is required for chloroplast ribosomal RNA processing in Arabidopsis. *Plant Physiol.* **168**, 205–221.

Loizeau, K., Qu, Y., Depp, S., Fiechter, V., Ruwe, H., Lefebvre-Legendre, L., Schmitz-Linneweber, C. and Goldschmidt-Clermont, M. (2014) Small RNAs reveal two target sites of the RNA-maturation factor Mbb1 in the chloroplast of *Chlamydomonas*. *Nucleic Acids Res.* **42**, 3286–3297.

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

Ma, Z., Wu, W., Huang, W. and Huang, J. (2015) Down-regulation of specific plastid ribosomal proteins suppresses *thf1* leaf variegation, implying a role of THF1 in plastid gene expression. *Photosynth. Res.* **126**, 301–310.

Machatt, M.A., Ebel, J.P. and Branlant, C. (1981) The 3′-terminal region of bacterial 23S ribosomal RNA: structure and homology with the 3′-terminal region of eukaryotic 28S rRNA and with chloroplast 4.5s rRNA. *Nucleic Acids Res.* **9**, 1533–1549.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMB net.Journal*, **17**, 10–12.

Miura, E., Kato, Y., Matsushima, R., Albrecht, V., Laalami, S. and Sakamoto, W. (2007) The balance between protein synthesis and degradation in chloroplasts determines leaf variegation in *Arabidopsis yellow variegated* mutants. *Plant Cell*, **19**, 1313–1328.

Moreira, D. and Philippe, H. (1999) Smr: a bacterial and eukaryotic homologue of the C–terminal region of the MutS2 family. *Trends Biochem. Sci.* **24**, 298–300.

Nishimura, K., Ashida, H., Ogawa, T. and Yokota, A. (2010) A DEAD box protein is required for formation of a hidden break in Arabidopsis chloroplast 23S rRNA. *Plant J.* **63**, 766–777.

O'Toole, N., Hattori, M., Andres, C., Iida, K., Lurin, C., Schmitz-Linneweber, C., Sugita, M. and Small, I. (2008) On the expansion of the pentatricopeptide repeat gene family in plants. *Mol. Biol. Evol.* **25**, 1120–1128.

Park, S. and Rodermel, S.R. (2004) Mutations in ClpC2/Hsp100 suppress the requirement for FtsH in thylakoid membrane biogenesis. *Proc. Natl Acad. Sci. USA*, **101**, 12765–12770.

Perrin, R., Lange, H., Grienenberger, J.M. and Gagliardi, D. (2004) AtmtPNPase is required for multiple aspects of the 18S rRNA metabolism in *Arabidopsis thaliana* mitochondria. *Nucleic Acids Res.* **32**, 5174–5182.

Pfalz, J., Liere, K., Kandlbinder, A., Dietz, K.J. and Oelmuller, R. (2006) pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. *Plant Cell*, **18**, 176–197.

Prikryl, J., Rojas, M., Schuster, G. and Barkan, A. (2011) Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proc. Natl Acad. Sci. USA*, **108**, 415–420.

Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

Rajagopalan, R., Vaucheret, H., Trejo, J. and Bartel, D.P. (2006) A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes Dev.* **20**, 3407–3425.

Rice, P., Longden, I. and Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277.

Ruwe, H. and Schmitz-Linneweber, C. (2012) Short non-coding RNA fragments accumulating in chloroplasts: footprints of RNA binding proteins? *Nucleic Acids Res.* **40**, 3106–3116.

Schmitz-Linneweber, C. and Small, I. (2008) Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci.* **13**, 663–670.

Sessions, A., Burke, E., Presting, G. *et al.* (2002) A high-throughput Arabidopsis reverse genetics system. *Plant Cell*, **14**, 2985–2994.

Shikanai, T. (2007) Cyclic electron transport around photosystem I: genetic approaches. *Annu. Rev. Plant Biol.* **58**, 199–217.

Small, I.D. and Peeters, N. (2000) The PPR motif – a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci.* **25**, 46–47.

Stoppel, R., Manavski, N., Schein, A., Schuster, G., Teubner, M., Schmitz-Linneweber, C. and Meurer, J. (2012) RHON1 is a novel ribonucleic acid-binding protein that supports RNase E function in the Arabidopsis chloroplast. *Nucleic Acids Res.* **40**, 8593–8606.

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

Tiller, N. and Bock, R. (2014) The translational apparatus of plastids and its role in plant development. *Mol Plant*, **7**, 1105–1120.

Timmis, J.N., Ayliffe, M.A., Huang, C.Y. and Martin, W. (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **5**, 123–135.

Walter, M., Kilian, J. and Kudla, J. (2002) PNPase activity determines the efficiency of mRNA 3′-end processing, the degradation of tRNA and the extent of polyadenylation in chloroplasts. *EMBO J.* **21**, 6905–6914.

Wang, Q., Sullivan, R.W., Kight, A., Henry, R.L., Huang, J., Jones, A.M. and Korth, K.L. (2004) Deletion of the chloroplast-localized *Thylakoid formation1* gene product in Arabidopsis leads to deficient thylakoid formation and variegated leaves. *Plant Physiol.* **136**, 3594–3604.

Wu, W., Zhu, Y., Ma, Z. *et al.* (2013) Proteomic evidence for genetic epistasis: ClpR4 mutations switch leaf variegation to virescence in Arabidopsis. *Plant J.* **76**, 943–956.

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

Yap, A., Kindgren, P., Colas des Francs-Small, C., Kazama, T., Tanz, S.K., Toriyama, K. and Small, I. (2015) AEF1/MPR25 is implicated in RNA editing of plastid *atpF* and mitochondrial *nad5*, and also promotes *atpF* splicing in Arabidopsis and rice. *Plant J.* **81**, 661–669.

Yin, P., Li, Q., Yan, C. *et al.* (2013) Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature*, **504**, 168–171.

Yu, F., Liu, X., Alsheikh, M., Park, S. and Rodermel, S. (2008) Mutations in *SUPPRESSOR OF VARIEGATION1*, a factor required for normal chloroplast translation, suppress *var2*-mediated leaf variegation in Arabidopsis. *Plant Cell*, **20**, 1786–1804.

Yu, F., Park, S.S., Liu, X. *et al.* (2011) *SUPPRESSOR OF VARIEGATION4*, a new *var2* suppressor locus, encodes a pioneer protein that is required for chloroplast biogenesis. *Mol Plant*, **4**, 229–240.

Zhang, L., Wei, Q., Wu, W., Cheng, Y., Hu, G., Hu, F., Sun, Y., Zhu, Y., Sakamoto, W. and Huang, J. (2009) Activation of the heterotrimeric G protein α–subunit GPA1 suppresses the ftsh-mediated inhibition of chloroplast development in Arabidopsis. *Plant J.* **58**, 1041–1053.

Zhelyazkova, P., Hammani, K., Rojas, M., Voelker, R., Vargas-Suarez, M., Borner, T. and Barkan, A. (2012) Protein-mediated protection as the predominant mechanism for defining processed mRNA termini in land plant chloroplasts. *Nucleic Acids Res.* **40**, 3092–3105.

Zhou, W., Cheng, Y., Yap, A., Chateigner-Boutin, A.L., Delannoy, E., Hammani, K., Small, I. and Huang, J. (2008) The Arabidopsis gene *YS1* encoding a DYW protein is required for editing of *rpoB* transcripts and the rapid development of chloroplasts during early growth. *Plant J.* **58**, 82–96.

Zoschke, R., Kroeger, T., Belcher, S., Schottler, M.A., Barkan, A. and Schmitz-Linneweber, C. (2012) The pentatricopeptide repeat-SMR protein ATP4 promotes translation of the chloroplast *atpB/E* mRNA. *Plant J.* **72**, 547–558.

Zoschke, R., Qu, Y., Zubo, Y.O., Borner, T. and Schmitz-Linneweber, C. (2013a) Mutation of the pentatricopeptide repeat-SMR protein SVR7 impairs accumulation and translation of chloroplast ATP synthase subunits in *Arabidopsis thaliana*. *J. Plant. Res.* **126**, 403–414.

Zoschke, R., Watkins, K.P. and Barkan, A. (2013b) A rapid ribosome profiling method elucidates chloroplast ribosome behavior *in vivo*. *Plant Cell*, **25**, 2265–2275.

Zoschke, R., Watkins, K.P., Miranda, R.G. and Barkan, A. (2016) The PPR-SMR protein PPR53 enhances the stability and translation of specific chloroplast RNAs in maize. *Plant J.* **85**, 593–605.

