Supporting Information

PPR-SMR protein SOT1 has RNA endonuclease activity

Wen Zhou, Qingtao Lu, Qingwei Li, Lei Wang, Shunhua Ding, Aihong Zhang, Xiaogang Wen, Lixin Zhang and Congming Lu

Materials and Methods

Map-based Cloning and Genetic Complementation of *sot1-3.*

Map-based cloning verified that the mutation was due to a 31-bp sequence deletion in AT5G46580 rather than a T-DNA insertion in the genome. Briefly, *sot1-3* was crossed to wild-type *Arabidopsis thaliana* (ecotype Landsberg) to generate an F1 mapping population. Approximately 200 F2 plants with a high chlorophyll fluorescence phenotype were selected for further analysis. The mutated gene was localized between SSLP (simple sequence length polymorphism) markers K11I1 and MZA15 and subsequently identified by sequencing.

For genetic complementation of *sot1-3*, sequences encoding SOT1 (amino acids 1-710), the SOT1 PPR domain (amino acids 1-613), the SOT1 chloroplast transit peptide (amino acids 1-43)-SMR domain (amino acids 603-710), SOT1 R617A, and SOT1 T655A were subcloned into the pBI121 vector encoding an HA tag at the C-terminus. The resulting constructs were transferred into *Agrobacterium tumefaciens* strain C58, followed by transformation into *sot1-3* plants through the floral dip method (1). The positive transformants were selected on MS medium containing 40 mg L^{-1} kanamycin in MS. The expression of transgenes in the positive transformants was detected by RT-PCR using gene-specific primers (see *SI Appendix*, Table S1) and immunoblots using HA antibody.

Chlorophyll Fluorescence Measurements and P_{700} Absorbance Analysis

Chlorophyll fluorescence images were acquired at room temperature using a modulated imaging fluorometer (FluorCam; Photon System Instruments) (2). The highest and lowest Fv/Fm values were represented by the red and blue extremes of the color scale, respectively.

The middle parts of the first pair of true leaves of 12-d-old wild-type, *sot1-3*, and *sot1-3com* seedlings were used for chlorophyll fluorescence and P_{700} absorbance measurements as previously described (3, 4). The plants were incubated in the dark for 30 min and used to determine the dark-adapted minimum fluorescence (Fo) and maximum fluorescence (Fm) with a PAM-2000 portable chlorophyll fluorometer (Heinz Walz). The

maximal efficiency of PSII photochemistry was calculated as Fv/Fm=(Fm-Fo)/Fm. P_{700} absorbance analysis was performed with a PAM-101 chlorophyll fluorometer equipped with an ED-P700W emitter-detector unit (Heinz Walz).

Transmission Electron Microscopy

Transmission electron microscopy was carried out as described previously (3). The middle parts of the first pair of true leaves of 12-day-old wild-type, *sot1-3*, and *sot1-3com* plants were harvested for chloroplast ultrastructural analysis under a transmission electron microscope (JEM-1230; JEOL).

SDS-PAGE and Immunoblot Analysis

SDS-PAGE and immunoblot analysis were carried out as described (5). To produce the polyclonal antibody for SOT1, the nucleotide sequence (amino acids 401 to 573) of *SOT1* was PCR amplified from wild-type DNA (for primers used, see *SI Appendix*, Table S1). The PCR products were inserted in pET28a encoding a C-terminal His tag, and the resulting construct was transformed into *E. coli* strain BL21 (DE3). The recombinant proteins were purified through a nickel-nitrilotriacetic acid agarose resin matrix (Ni-NTI, Qiagen). The purified proteins were used as an antigen to raise antibodies in rabbit. The SOT1 antibody was diluted 1:2000. The Actin, HA, and MBP antibodies were purchased from MBL Corporation (Japan). Fd and PsaA antibodies were purchased from Agrisera (Vännäs, Sweden). The other antibodies used in this study were produced in our laboratory (5).

***In vivo* Translation Assay of Thylakoid Proteins**

In vivo protein labeling was carried out as previously described (4). In brief, the youngest leaves of 12-d-old wild-type and *sot1-3* seedlings were preincubated in cycloheximide to inhibit nuclear translation, followed by the addition of [^{35}S]Met to the labeling buffer and incubation for 20 min. After labeling, thylakoid membranes were isolated and subjected to SDS-PAGE before autoradiography.

RNA Preparation, RNA Gel Blot Analysis, Circular PCR, and RACE

RNA preparation and RNA gel blot analysis were carried out as described (2). Circular RT-PCR was performed according to Puchta et al. (6). Briefly, 1 µg total RNA was self-ligated using T4 RNA ligase 1 (New England Biolabs M0202) according to the manufacturer's protocol. After ligation, RNA was isolated using standard methods (extraction with phenol: chloroform: isoamyl = 25:24:1 [v/v]). Reverse transcription was performed with M-MLV (Takara) and primers (see *SI Appendix*, Table S1). PCR amplification was performed using primers located in the region of 23S-4.5S rRNA precursor (for primers, see *SI Appendix*, Table S1). The PCR products were then cloned and sequenced. The rapid amplification of cDNA ends (RACE) assay was performed using a SMARTer RACE 5'/3' Kit (Takara 634858). The cDNAs derived from RACE assay were PCR amplified using linker primer and gene-specific primers (see *SI Appendix*, Table S1), followed by cloning and sequencing.

Accession Codes

Sequence data in this article can be found in the GenBank data library under the following accession numbers: *SOT1* (AT5G46580, AED95400.1), *Gm-SOT1* (LOC100780994, XP_003524052*)*, *Arabidopsis thaliana* (AP000423), and *Glycine max* (DQ317523).

Figures

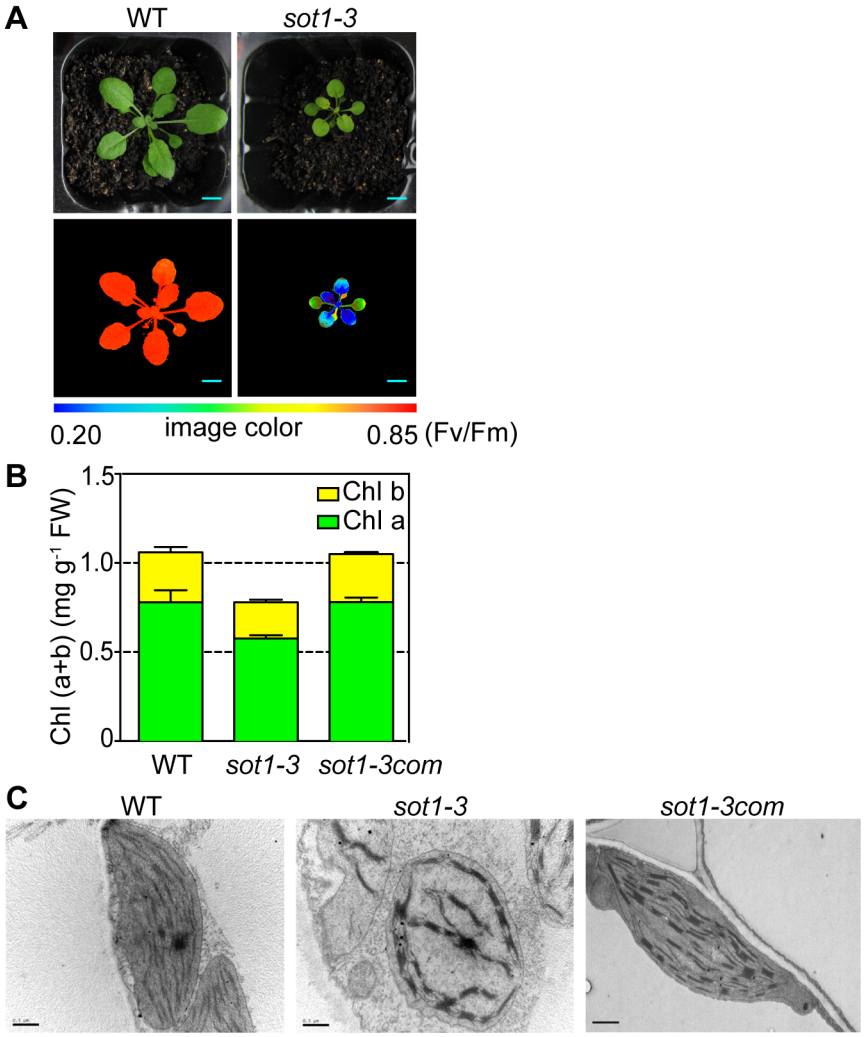

Fig. S1. Phenotypes of wild-type, *sot1-3*, and complemented plants. (*A*) Photographs and chlorophyll fluorescence images of 4-week-old wild type (WT) and *sot1-3*. Chlorophyll fluorescence was shown as pseudo-color according to the index as indicated at the bottom. Scale bar, 0.5 cm. (*B*) Leaf chlorophyll contents and chlorophyll a/b ratio in WT, *sot1-3,* and complemented (*sot1-3com*) plants. Leaf chlorophyll contents and chlorophyll a/b ratio were determined on the leaves of 12-d-old WT, *sot1-3,* and *sot1-3com* grown on MS medium under 100 μmol m^{-2} s^{-1} (14-h-light/10-h-dark cycles) at 23°C, and 50% humidity. Error bars indicate SD (n = 3). (*C*) Chloroplast ultrastructure from the first pair of true leaves in 12-d-old WT, *sot1-3*, and *sot1-3com* seedlings. Scale bar, 0.5 μm.

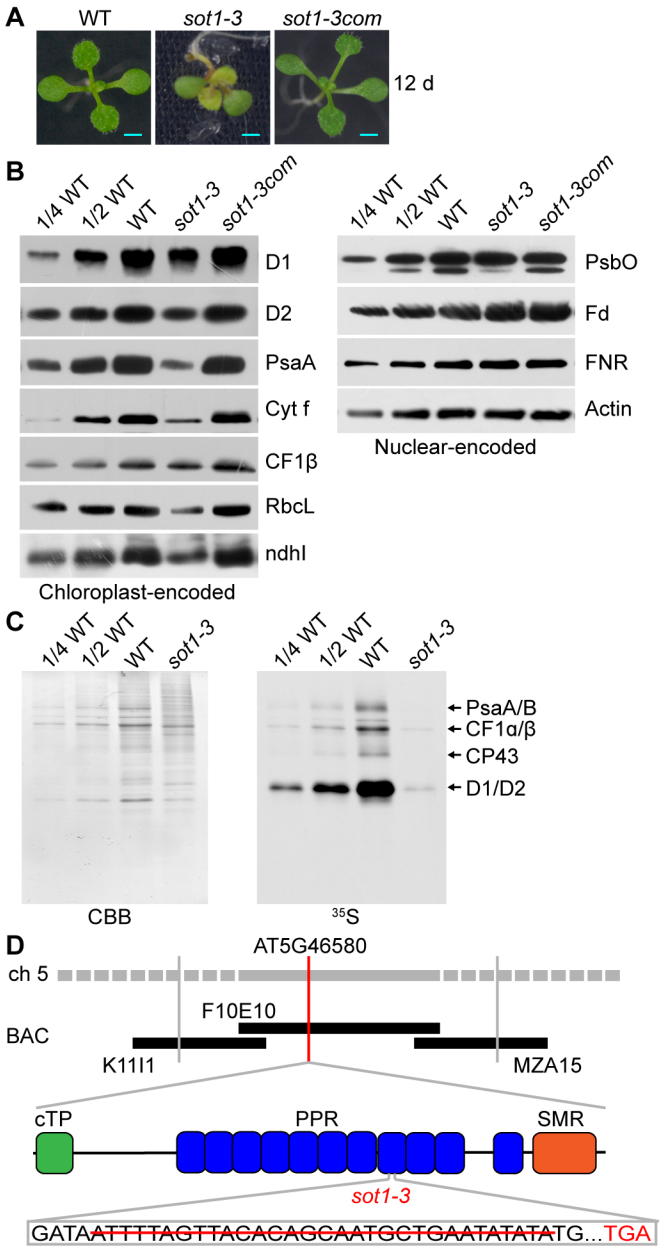

Fig. S2. Identification of the s*ot1-3* mutant and the *SOT1* locus. (*A*) Phenotypes of wild-type (WT), *sot1-3,* and complemented plants (*sot1-3com*). The plants were grown on MS medium for 12 d. Scale bar, 0.2 cm. (*B*) Immunoblot analysis of chloroplast proteins on the basis of equal total leaf proteins. D1, D2, PsaA, Cyt f, CF1β, RbcL, and ndhI are chloroplast-encoded. PsbO, Fd, and FNR are nuclear-encoded and chloroplast-localized. Total leaf proteins were extracted from 12-d-old WT, *sot1-3* and *sot1-3com*

seedlings. β-Actin was used as a loading control. (*C*) *In vivo* pulse labeling analysis of thylakoid membrane proteins. Thylakoid membranes were isolated from 12-d-old WT, *sot1-3* and *sot1-3com* seedlings. After a 20-min pulse of the label, thylakoid membranes were resolved by SDS-PAGE and stained with Coomassie Brilliant Blue (CBB) (left panel) or subjected with autoradiography (right panel). (*D*) Map-based cloning of *SOT1*. The *sot1-3* mutant was mapped to chromosome 5 on BAC clone F10E10. Gray and red vertical lines represent the fine-mapped edges and the *SOT1* locus (AT5G46580), respectively. *SOT1* is represented by a black line. Predicted chloroplast transit peptide (cTP), single PPR motif (PPR), and SMR domain (SMR) are represented by green, blue, and orange boxes, respectively. Red horizontal line indicates the 31-bp deletion (position 1295-1325 with respect to the translation initiation site) in *sot1-3*. TGA, translation stop codon.

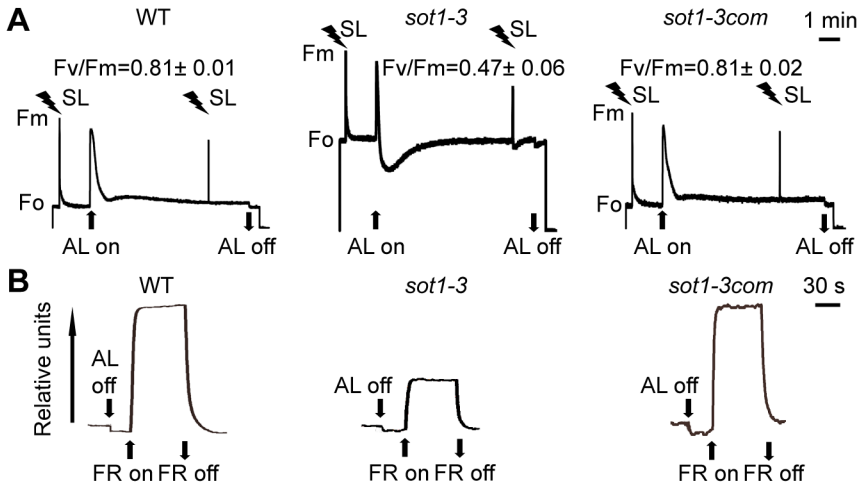

Fig. S3. Functions of PSII and PSI in wild-type, *sot1-3*, and complemented plants. (*A*) Chlorophyll a fluorescence induction analyses. Fm, maximal fluorescence in dark-adapted state; Fo, minimal fluorescence in dark-adapted state; SL, saturating light; AL, actinic light. (*B*) Redox kinetics of P_{700} at 820 nm induced by far-red light (FR; 720 nm). Absorbance of P_{700} at 820 nm is a measurement of the P_{700} redox state. AL, actinic light; FR, far-red light.

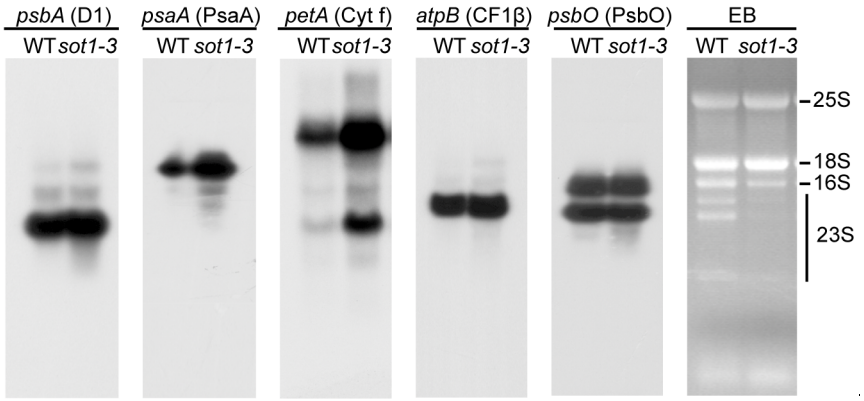

Fig. S4. Accumulation of chloroplast transcripts in WT and *sot1-3*. Total RNAs from 12-d-old WT and *sot1-3* seedlings were subjected to RNA gel blot analysis using the gene-specific probes shown on the top. Ethidium bromide-staining gel was used as a loading control (right panel).

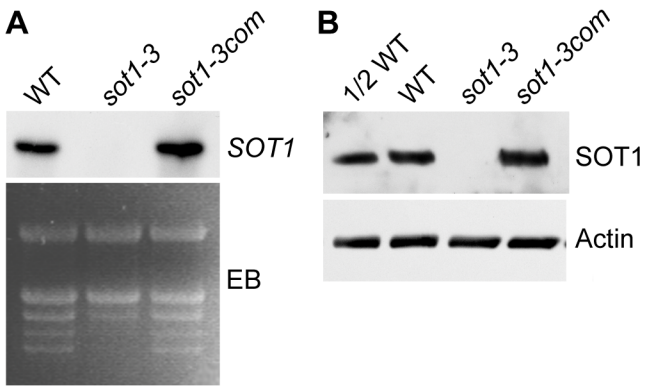

Fig. S5. Expression analyses of *SOT1* in *sot1-3*. (*A*) *SOT1* mRNA expression in WT, *sot1-3*, and *sot1-3com* seedlings. Total RNA isolated from 12-d-old WT, *sot1-3*, and *sot1-3com* seedlings was analyzed by RNA gel blot using gene-specific probes for *SOT1* (top) or by ethidium bromide staining (bottom). EB, ethidium bromide. (*B*) SOT1 protein accumulation in WT, *sot1-3* and *sot1-3com* seedlings. Total proteins extracted from 12-d-old WT, *sot1-3*, and *sot1-3com* seedlings were analyzed by immunoblotting using antibodies against SOT1 (top) and Actin (bottom). Equivalent amounts of total protein (20 μg) were loaded.

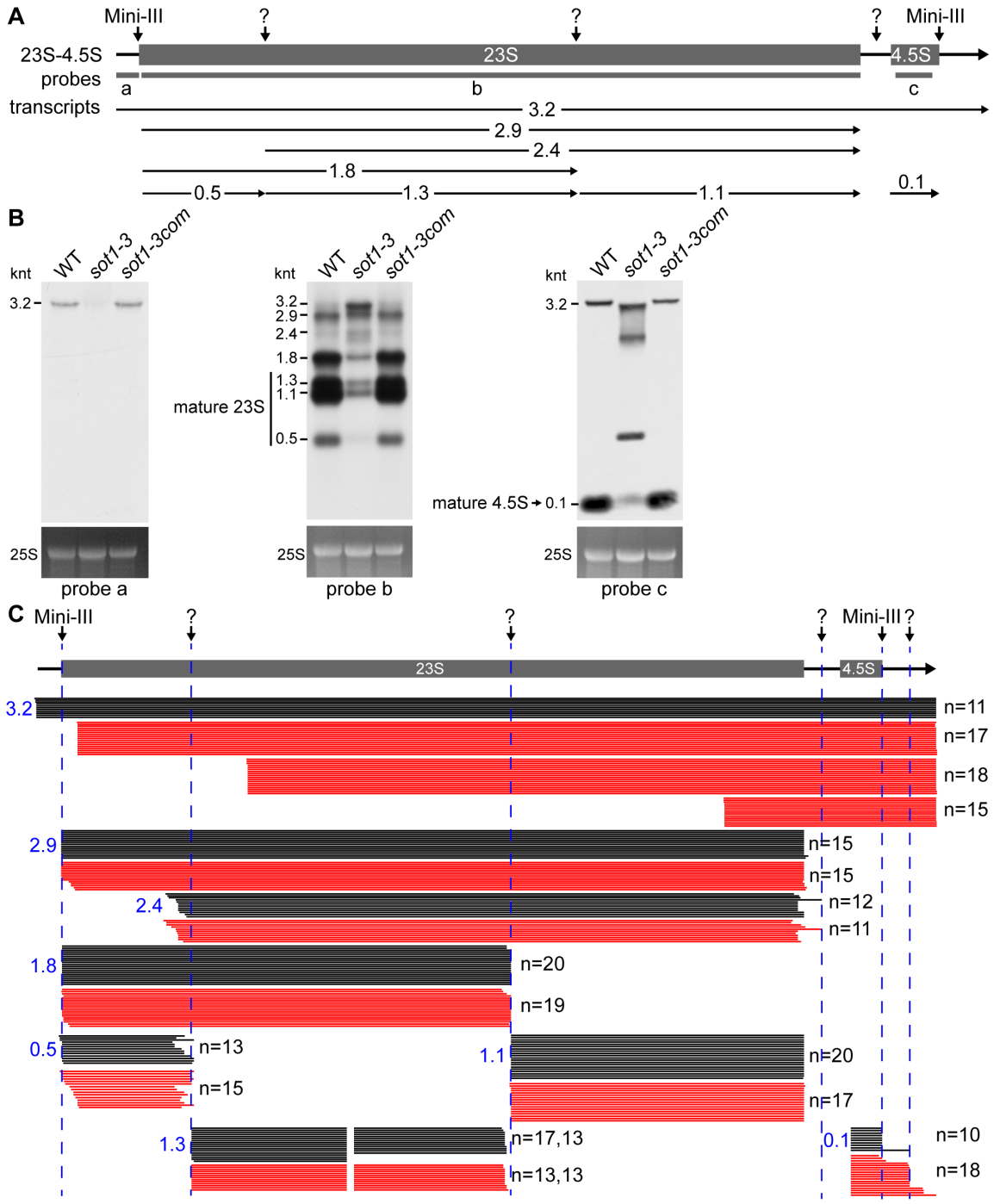

Fig. S6. Analyses of the maturation of 23S and 4.5S rRNA in *sot1-3*. (*A*) Schematic representation of the 23S-4.5S rRNA precursor processing pathway in Arabidopsis chloroplasts. The 23S-4.5S rRNA precursor is shown by a black horizontal arrow (top). Vertical arrows represent the endonucleolytic processing sites in the 23S-4.5S rRNA

precursor (7, 8). Dark boxes represent 23S and 4.5S rRNA. Thin gray boxes under the locus represent the locations of the probes used in (*B*). The locations and sizes (in knt) of accumulated transcripts are marked with horizontal arrows below the probes. (*B*) Loss of SOT1 results in abnormal maturation of 23S and 4.5S rRNA. Mature transcripts and precursors of 23S and 4.5S rRNA were detected by RNA gel blots using probes a to c shown in (*A*). The transcript size is shown in knt to the left of each panel. 25S rRNA (loading control) was stained with ethidium bromide. (*C*) Loss of SOT1 results in staggered 23S 5' ends. The 5' and 3' ends of the transcripts from the 23S-4.5S transcription unit were identified by circular RT-PCR using gene-specific primers (see *SI Appendix,* Table S1), with the exception that the ends of 4.5S and 1.3 knt transcript were identified using RACE. The 5' and 3' ends deduced from circular RT-PCR and RACE clones in WT (black bars) and *sot1-3* (red bars) are shown below the gene model, with each bar indicating a single clone. Mini-III, mini-ribonuclease III.

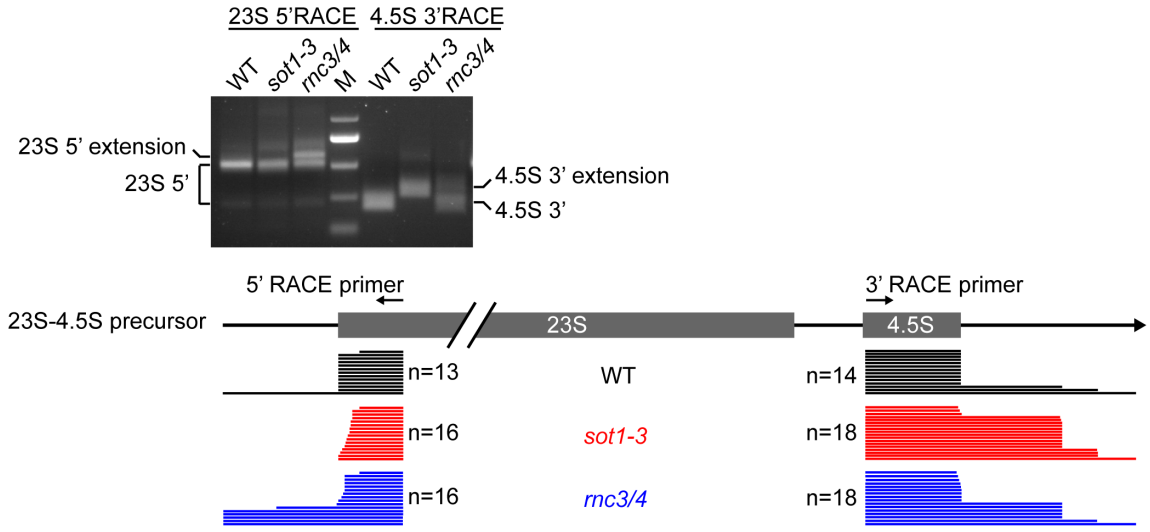

Fig. S7. Staggered 23S rRNA 5' ends and less mature 4.5S rRNA in *sot1-3* and *rnc3/4*. 23S 5' ends and 4.5S 3' ends in WT, *sot1-3* and *rnc3/4* were mapped by RACE using 5' RACE primer and 3' RACE primer, respectively. The PCR products corresponding to 23S 5' and 23S 5' extension in the 5' RACE assay and the PCR products corresponding to 4.5S 3' and 4.5S 3' extension in the 3' RACE assay were separated in agarose gels (top panel), followed by gel purification and sequencing. The 23S 5' ends and 4.5S 3' ends deduced from RACE clones in WT (black bars), *sot1-3* (red bars) and *rnc3/4* (blue bars) are shown below the gene model, with each bar indicating a single clone.

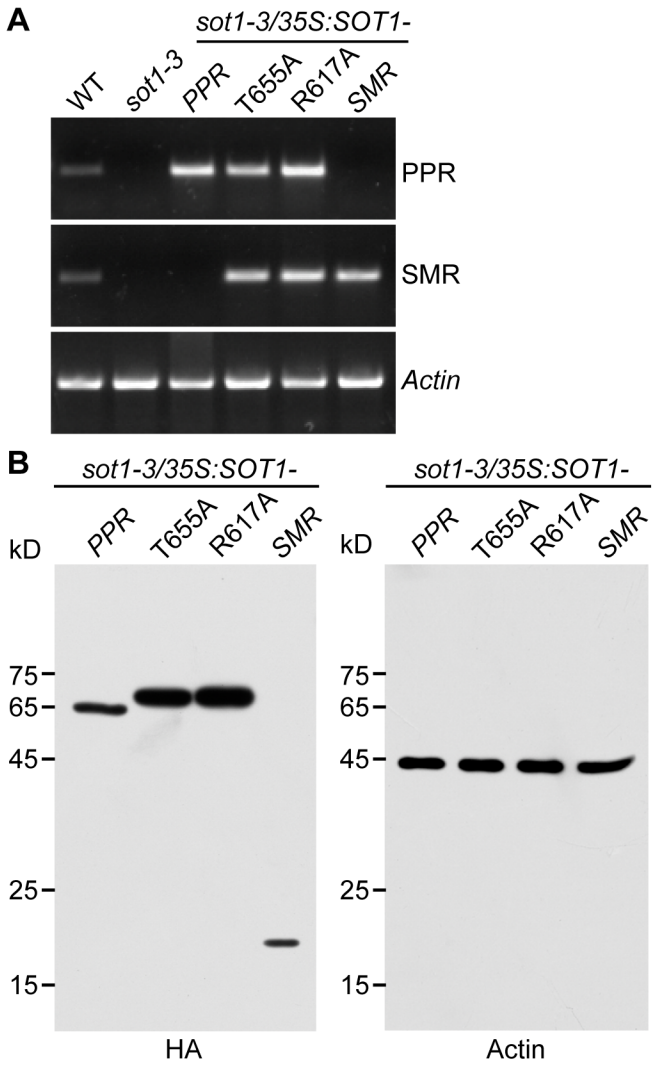

Fig. S8. Expression analyses of the PPR and SMR domain proteins of SOT1 in genetically complemented lines. (*A*) Accumulation of transcripts for the PPR and SMR domains. Total RNAs extracted from *sot1-3*/*35S:SOT1*$_{PPR}$*-HA*, *sot1-3*/*35S:SOT1*$_{\mathrm{T655A}}$*-HA*, *sot1-3*/*35S:SOT1*$_{\mathrm{R617A}}$*-HA* and *sot1-3*/*35S:SOT1*$_{SMR}$*-HA* plants were analyzed by RT-PCR using primers specific to sequences encoding the PPR domain of SOT1 and the SMR domain of SOT1, as well as *Actin2*. (*B*) Accumulation of the PPR and SMR domain proteins. Total proteins extracted from *sot1-3*/*35S:SOT1*$_{PPR}$*-HA*, *sot1-3*/*35S:SOT1*$_{\mathrm{T655A}}$*-HA*, *sot1-3*/*35S:SOT1*$_{\mathrm{R617A}}$*-HA* and *sot1-3*/*35S:SOT1*$_{SMR}$*-HA* plants were analyzed by immunoblotting using antibodies against HA (left panel) and Actin (right panel). Equivalent amounts of total protein (20 μg) were loaded.

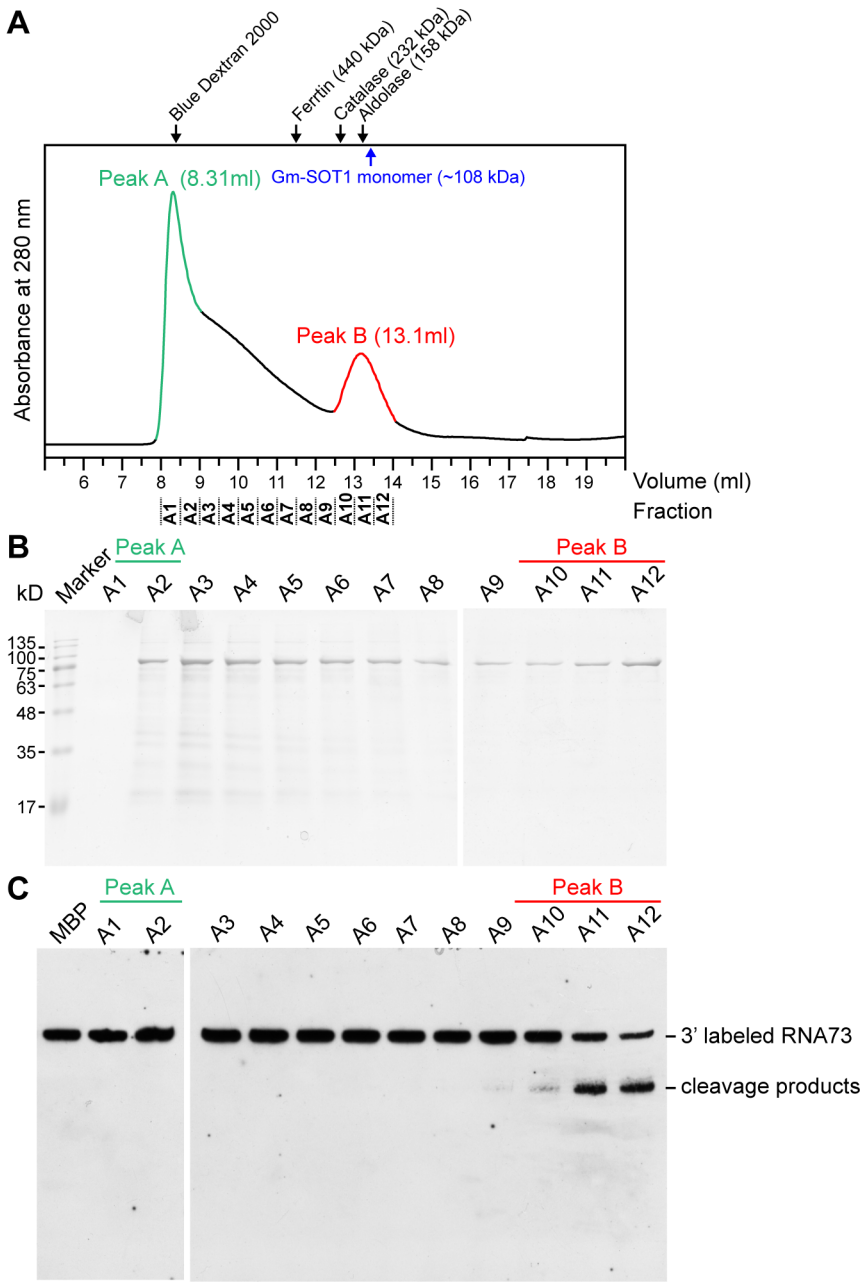

Fig. S9. Preparation of Gm-SOT1 with RNA endonuclease activity. (*A*) Size-exclusion chromatography (SEC) profile of Gm-SOT1. The SEC column was calibrated with standards of blue dextran 2000, ferritin, catalase, and aldolase. Fractions A1 and A2 are labeled as peak A, and fractions A10-A12 are labeled as peak B. (*B*) SDS-PAGE with Coomassie Brilliant Blue staining of fractions A1 to A12. The molecular markers are shown on the left. (*C*) RNA endonuclease activity analysis of fractions A1 to A12. The

RNA73 was used as the substrate for assay of the RNA endonuclease activity. Peak B had the ability to cleave RNA73 efficiently, and the pooled fractions A10-A12 were accordingly used for further analyses.

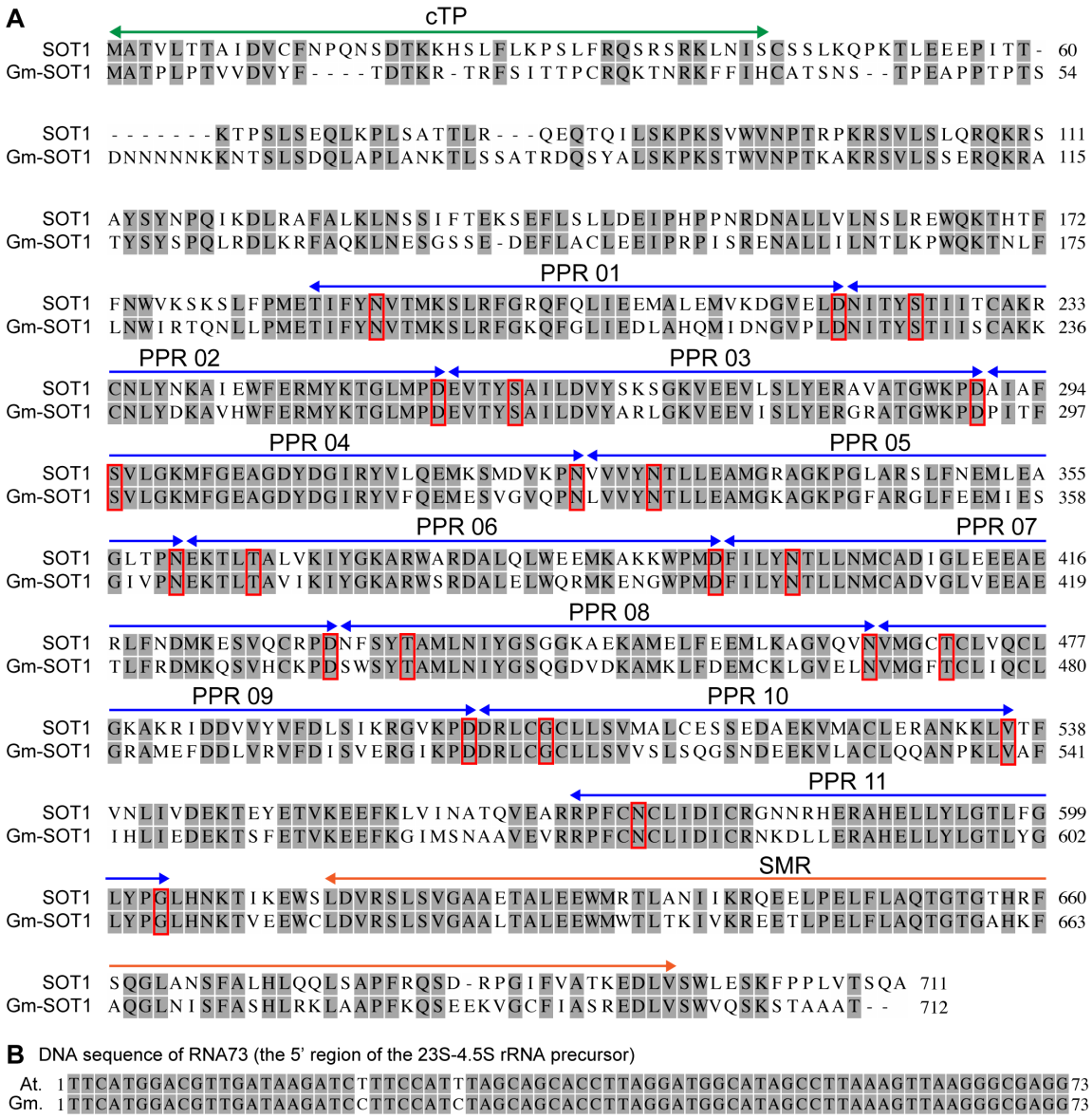

Fig. S10. Sequence alignment of SOT1 and Gm-SOT1, and alignment of RNA73 between *Arabidopsis thaliana* and *Glycine max*. (*A*) Sequence alignment of SOT1 and Gm-SOT1. The amino acids of SOT1 and Gm-SOT1 were aligned using MUSCLE followed by manual adjustments, and visualized by CLC Sequence Viewer soft (www.clcbio.com). Absolutely conserved amino acids are represented in grey boxes. The predicted chloroplast transit peptide (cTP), PPR motifs and SMR domain are indicated. The residues at the 5th and 35th positions of each PPR motif that are predicted to be the molecular determinants for RNA-binding specificity are highlighted in red frames. (*B*)

DNA sequence alignment of RNA73 between *Arabidopsis thaliana* and *Glycine max*. The sequence alignment was performed using ClustalW, followed by manual adjustments, and visualized using CLC Sequence Viewer. The grey boxes indicate absolutely conserved nucleic acids.

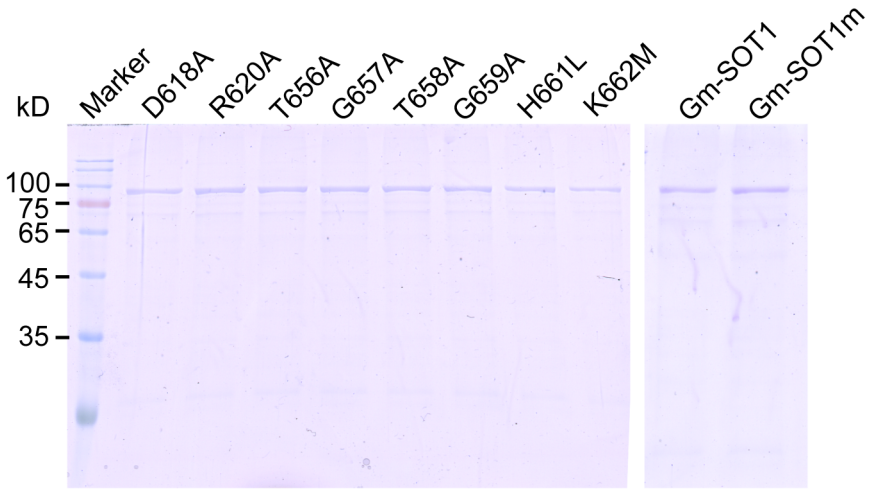

Fig. S11. Site-directed mutagenesis of Gm-SOT1. To investigate the critical amino acids of the SMR domain of SOT1 for RNA endonuclease activity, eight amino acids (D618, R620, T656, G657, T658, G659, H661, and K662) were mutated. In addition, to investigate programmable RNA recognition and cleavage by Gm-SOT1, mutagenesis at the 5th amino acids in the PPR motifs 1 to 4 of Gm-SOT1 was performed and the mutated Gm-SOT1 was termed as Gm-SOT1m. The mutated proteins were resolved by SDS-PAGE and stained with CBB.

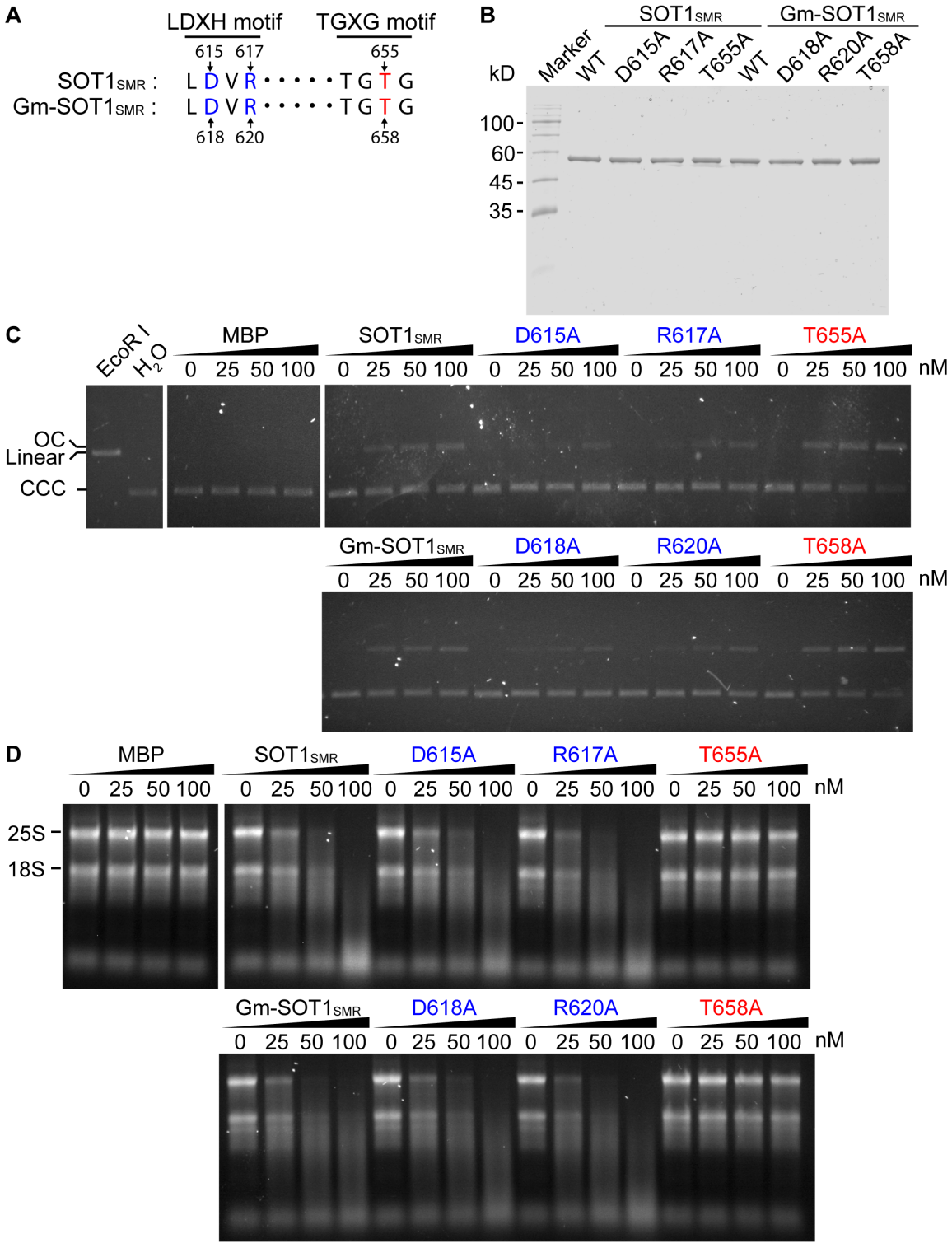

Fig. S12. DNA nicking and RNA nuclease activities of the SMR domain of SOT1 and Gm-SOT1 depend on distinct amino acids. (*A*) The amino acids used for site-

directed mutagenesis in LDXH (indicated in blue) and TGXG (indicated in red) motifs of $SOT1_{SMR}$ and Gm-$SOT1_{SMR}$. The number indicates the amino acid position with respect to the translation initiation site. (*B*) SDS-PAGE with Coomassie Brilliant Blue staining of the variants of $SOT1_{SMR}$ and Gm-$SOT1_{SMR}$. The variants were obtained from the D615A, R617A, and T655A substitutions for $SOT1_{SMR}$ and the D618A, R620A, and T658A substitutions for Gm-$SOT1_{SMR}$. The molecular markers are shown on the left. (*C*) DNA nicking activities of the variants of $SOT1_{SMR}$ and Gm-$SOT1_{SMR}$. MBP, $SOT1_{SMR}$, Gm-$SOT1_{SMR}$, and the variants with different concentrations were incubated with 5 ng pUC19 plasmid DNA at 25°C for 60 min in the presence of 3 mM $MgCl_2$. The reactions were stopped by loading buffer and were electrophoresed on 1.2% agarose gels. Parallel experiments were carried out with EcoR I to linearize pUC19 or with H_2O as a control. OC, Linear, and CCC indicate open circular, linear, and covalently closed-circular forms of pUC19, respectively. (*D*) RNA nuclease activities of the variants of $SOT1_{SMR}$ and Gm-$SOT1_{SMR}$. MBP, $SOT1_{SMR}$, Gm-$SOT1_{SMR}$, and the variants with different concentrations were incubated with total wild-type Arabidopsis RNA at 25°C for 30 min. The reaction products were separated in agarose/formaldehyde gels and observed by ethidum bromide-staining.

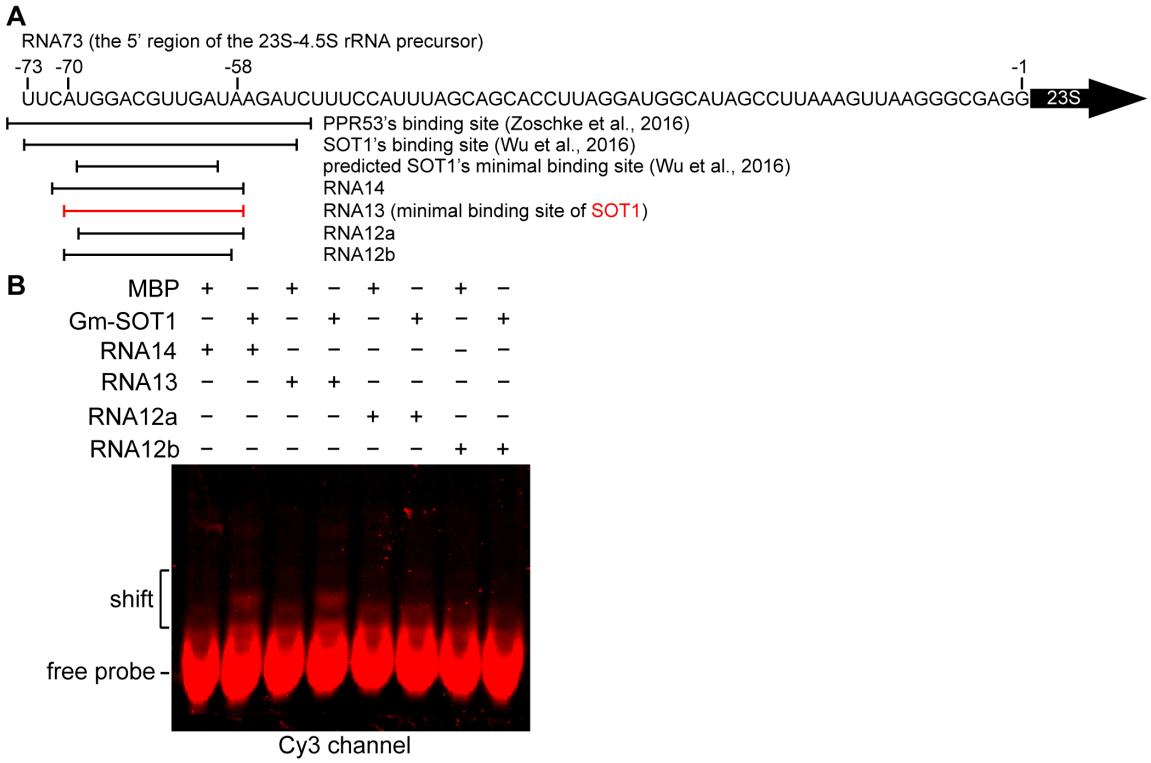

Fig. S13. Fine-mapping of the minimal binding sequence of SOT1. (*A*) Schematic illustration of the potential binding sequence of SOT1. Experimentally identified binding sequence (20-nt/24-nt segments by SOT1/PPR53) and the predicted minimal binding sequence of SOT1 are marked below the sequence of RNA73 (9, 10). The positions of RNA probes (RNA14, RNA13, RNA12a and RNA12b) used to identify the minimal binding sequence in (*B*) are also presented below the sequence of RNA73. (*B*) Fine-mapping the minimal binding sequence of SOT1 by electrophoretic mobility shift assays (EMSA). MBP and Gm-SOT1 were incubated with RNA14, RNA13, RNA12a, and RNA12b, respectively. The affinity between Gm-SOT1 and RNA13 indicates that RNA13 is the minimal binding sequence of SOT1.

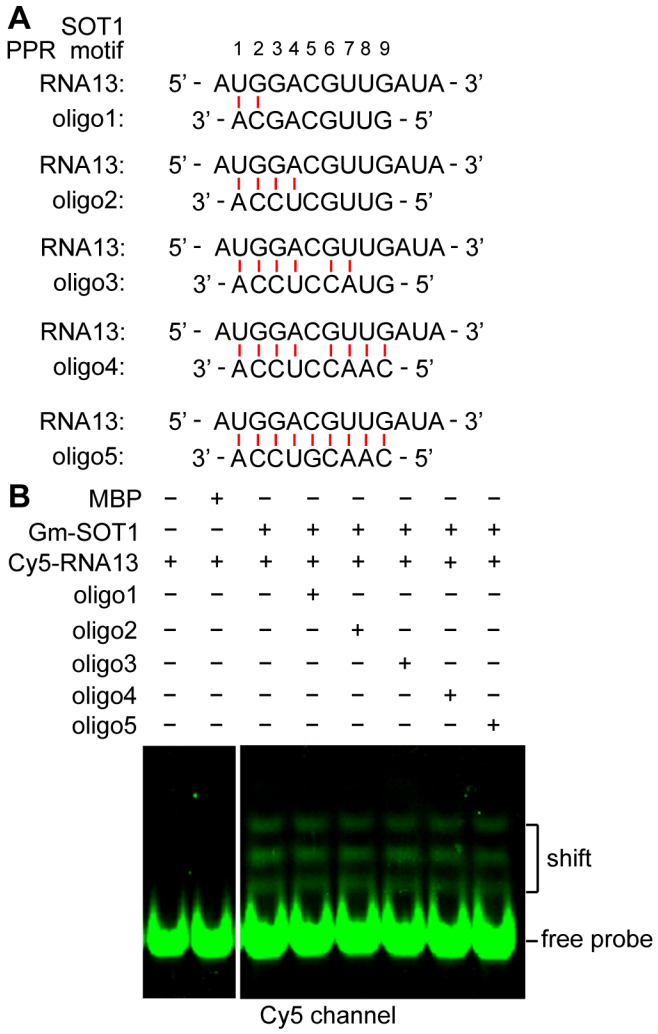

Fig. S14. Effects of varying structural content of RNA13 on RNA binding by Gm-SOT1. (*A*) Schematic illustration of short, unlabeled oligos 1-5 that progressively base pair with RNA13. According to the PPR code (11-13), the "UGGACGUUG" sequence in RNA13 is the binding site corresponding to typical PPR motifs 1 to 9 of SOT1, respectively. Oligos 1-5 represent the oligos that base pair with 2, 4, 6, 8, and 9 nucleotides in RNA13, respectively. (*B*) Analyses of RNA binding of SOT1 by adding short, unlabeled oligos. Cy5-labeled RNA13 (40 nM) was pre-incubated with oligos 1-5 (40 nM), respectively, before the addition of Gm-SOT1 (400 nM).

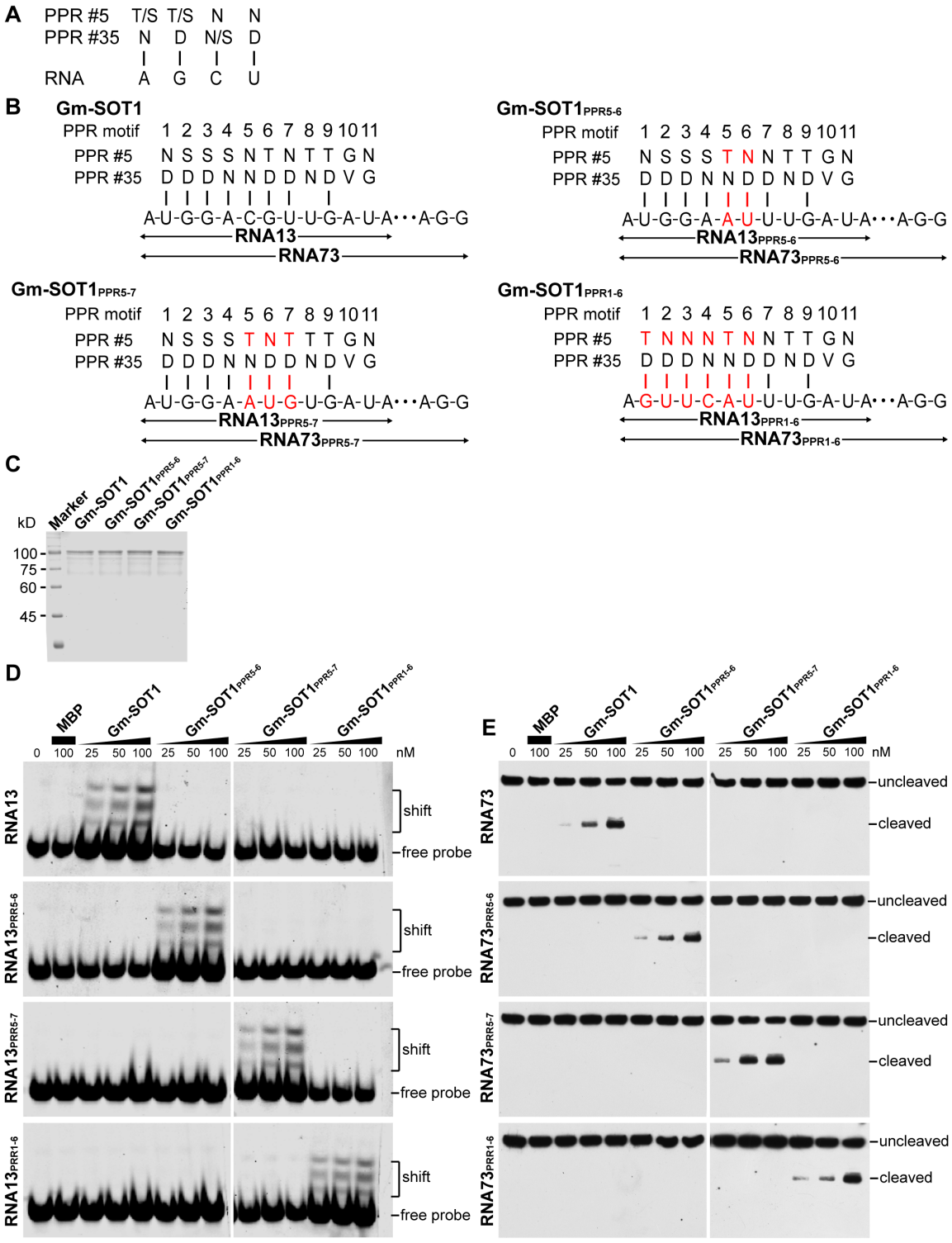

Fig. S15. Programmable RNA recognition and cleavage by SOT1. (*A*) The PPR code (11-13). (*B*) Alignment between Gm-SOT1, Gm-SOT1$_{PPR5-6}$, Gm-SOT1$_{PPR5-7}$, Gm-

SOT1$_{PPR1-6}$ and their RNA targets. The amino acids at position 5 and 35 of each PPR motif are shown. Mutated amino acids and the cognate RNA targets are indicated in red. Gm-SOT1$_{PPR5-6}$, Gm-SOT1$_{PPR5-7}$ and Gm-SOT1$_{PPR1-6}$ are the Gm-SOT1 variants in which the 5th amino acids were mutated in the PPR motifs either 5 to 6, or 5 to 7, or 1 to 6. The cognate RNA target for each Gm-SOT1 variant was predicted according to the "PPR code". (*C*) Coomassie Brilliant Blue staining for purified Gm-SOT1, Gm-SOT1$_{PPR5-6}$, Gm-SOT1$_{PPR5-7}$, and Gm-SOT1$_{PPR1-6}$. The marker sizes are shown to the left. (*D*) Programmable RNA recognition by Gm-SOT1. EMSA shows the binding of Gm-SOT1 and the variants (Gm-SOT1$_{PPR5-6}$, Gm-SOT1$_{PPR5-7}$, and Gm-SOT1$_{PPR1-6}$) to biotin-labeled RNA13 and mutated-RNA13 (RNA13$_{PPR5-6}$, RNA13$_{PPR5-7}$, and RNA13$_{PPR1-6}$). Increasing concentrations of protein (25, 50, and 100 nM) were incubated with 10 nM probes. (*E*) Programmable RNA cleavage by Gm-SOT1. RNA cleavage analysis shows the cleavage of biotin-labeled RNA73 and mutated-RNA73 (RNA73$_{PPR5-6}$, RNA73$_{PPR5-7}$, and RNA73$_{PPR1-6}$) by Gm-SOT1 and the variants. Increasing concentrations of protein (25, 50, and 100 nM) were incubated with 10 nM probes.

Tables

Table S1. A list of primers used in this study

Purpose	Primer name	Primer sequence (5'-3')
Complementation vectors construction	*SOT1*(1-710)-forward	GGATCCATGGCGACTGTTCTTACCACA
	SOT1(1-710)-reverse	CTCGAGCTGCTTGAGAAGTAACTAAAGG
	$SOT1_{PPR}$(1-613)-forward	CTAGTCTAGAATGGCGACTGTTCTTACCAC
	$SOT1_{PPR}$(1-613)-reverse	CGCGGATCCGGCTCCATTCTTTTATGGTC
	$SOT1_{\text{signal peptide}}$(1-43)-forward	CTAGTCTAGAATGGCGACTGTTCTTACCAC
	$SOT1_{\text{signal peptide}}$(1-43)-reverse	CCGGAATTCAGAAATGTTAAGCTTTCTGG
	$SOT1_{SMR}$(603-710)-forward	CCGGAATTCGGACTACACAACAAGACCAT
	$SOT1_{SMR}$(603-710)-reverse	CGCGGATCCATGCTTGAGAAGTAACTAAAGG
	$SOT1_{\text{R617A}}$(1-710)-forward	TGGAGCCTAGATGTGGCGTCATTATCAGTT
	$SOT1_{\text{R617A}}$(1-710)-reverse	GCCACATCTAGGCTCCATTCTTTTATG
	$SOT1_{\text{T655A}}$(1-710)-forward	TTGGCGCAAACTGGGGCAGGAACTCAC
	$SOT1_{\text{T655A}}$(1-710)-reverse	CCCCAGTTTGCGCCAAAAACAGCTC
Protein expression vectors construction	SOT1 antibody-forward	GGATCCACGTTGTTGAATATGTGTGCT
	SOT1 antibody-reverse	CGGTCGACTCTGCATATATCTATCAAACA
	MBP-forward	CCGGAATTCGATGAAAATCGAAGAAGGT
	MBP-reverse	CGGGGTACCGAAATCCTTCCCTCGATC
	$SOT1_{SMR}$(603-710)-forward	CCGGAATTCGGGACTACACAACAAGAC
	$SOT1_{SMR}$(603-710)-reverse	ACGCGTCGACGCTTGAGAAGTAACTAAAG
	$SOT1_{SMR\text{-D615A}}$(603-710)-forward	AAAGAATGGAGCCTAGCTGTGAGGTCA
	$SOT1_{SMR\text{-D615A}}$(603-710)-reverse	GCTAGGCTCCATTCTTTTATGGTCTTG
	$SOT1_{SMR\text{-R617A}}$(603-710)-forward	TGGAGCCTAGATGTGGCGTCATTATCAGTT
	$SOT1_{SMR\text{-R617A}}$(603-710)-reverse	GCCACATCTAGGCTCCATTCTTTTATG
	$SOT1_{SMR\text{-T655A}}$(603-710)-forward	TTGGCGCAAACTGGGGCAGGAACTCAC
	$SOT1_{SMR\text{-T655A}}$(603-710)-reverse	CCCCAGTTTGCGCCAAAAACAGCTC
	Gm-SOT1$_{SMR}$(606-711)-forward	CCGGAATTCAGGTTTGCACAACAAAACTGTC
	Gm-SOT1$_{SMR}$(606-711)-reverse	ACGCGTCGACGTAGCAGCTGCAGTAGACTT
	Gm-SOT1(182-706)-forward	CGCGGATCCCCAGAACTTGCTTCCCAT

	Gm-SOT1(182-706)-reverse	CATGCCATGGTCTTTGATTGCACCCAAG
	Gm-SOT1$_{D618A}$(182-706)-forward/*Gm-SOT1*$_{SMR-D618A}$(606-711)-forward	GAAGAATGGTGTTTAGCCGTTCGGTCA
	Gm-SOT1$_{D618A}$(182-706)-reverse/*Gm-SOT1*$_{SMR-D618A}$(606-711)- reverse	GCTAAACACCATTCTTCGACAGTTTTG
	Gm-SOT1$_{R620A}$(182-706)-forward/*Gm-SOT1*$_{SMR-R620A}$(606-711)-forward	TGGTGTTTAGACGTTGCGTCATTGTCG
	Gm-SOT1$_{R620A}$(182-706)-reverse/*Gm-SOT1*$_{SMR-R620A}$(606-711)- reverse	GCAACGTCTAAACACCATTCTTCGAC
	Gm-SOT1$_{T656A}$(182-706)-forward	TTATTCCTAGCCCAAGCTGGTACTGG
	Gm-SOT1$_{T656A}$(182-706)-reverse	CTTGGGCTAGGAATAACTCAGGCAGT
	Gm-SOT1$_{G657A}$(182-706)-forward	TCCTAGCCCAAACTGCTACTGGTGCTC
	Gm-SOT1$_{G657A}$(182-706)-reverse	GCAGTTTGGGCTAGGAATAACTCAGG
	Gm-SOT1$_{T658A}$(182-706)-forward/*Gm-SOT1*$_{SMR-T658A}$(606-711)-forward	CTAGCCCAAACTGGTGCTGGTGCTCAT
	Gm-SOT1$_{T658A}$(182-706)-reverse/*Gm-SOT1*$_{SMR-T658A}$(606-711)- reverse	CACCAGTTTGGGCTAGGAATAACTCAG
	Gm-SOT1$_{G659A}$(182-706)-forward	GCCCAAACTGGTACTGCTGCTCATAAG
	Gm-SOT1$_{G659A}$(182-706)-reverse	GCAGTACCAGTTTGGGCTAGGAATAAC
	Gm-SOT1$_{H661L}$(182-706)-forward	CTGGTACTGGTGCTCTTAAGTTTGCC
	Gm-SOT1$_{H661L}$(182-706)-reverse	AGAGCACCAGTACCAGTTTGGGCTAG
	Gm-SOT1$_{K662M}$(182-706)-forward	GTACTGGTGCTCATATGTTTGCCCAAG
	Gm-SOT1$_{K662M}$(182-706)-reverse	ATATGAGCACCAGTACCAGTTTGGGC
	Gm-SOT1$_{PPR1\text{-}PPR4\ variant}$ (182-706)-forward1	TTCTACACTGTCACAATGAAGTCTCTGAGGT
	Gm-SOT1$_{PPR1\text{-}PPR4\ variant}$ (182-706)-reverse4	CTTCATTGTGACAGTGTAGAATATGGTTTCC
	Gm-SOT1$_{PPR1\text{-}PPR4\ variant}$ (182-706)-forward2	ACTTACAACACCATCATTTCCTGTGCCAAGA
	Gm-SOT1$_{PPR1\text{-}PPR4\ variant}$ (182-706)-reverse1	GGAAATGATGGTGTTGTAAGTGATGTTGTCC
	Gm-SOT1$_{PPR1\text{-}PPR4\ variant}$ (182-706)-forward3	GATGAGGTAACTTACAATGCTATTCTAGATG
	Gm-SOT1$_{PPR1\text{-}PPR4\ variant}$ (182-706)-reverse2	AGCATTGTAAGTTACCTCATCAGGCATCAAA
	Gm-SOT1$_{PPR1\text{-}PPR4\ variant}$ (182-	ACTTTCAATGTGTTGGGGAAGATGTTT

	706)-forward4	GGGG
	Gm-SOT1$_{\text{PPR1-PPR4 variant}}$ (182-706)-reverse3	CTTCCCCAACACATTGAAAGTAATAGGGTCA
	Gm-SOT1$_{\text{PPR5 variant}}$ (182-706)-forward	ACTACCTTGTTGGAGGCAATGGGAAAGGC
	Gm-SOT1$_{\text{PPR5 variant}}$ (182-706)-reverse	TGCCTCCAACAAGGTAGTGTACACAACC
	Gm-SOT1$_{\text{PPR6 variant}}$ (182-706)-forward	AACGCGGTTATTAAGATATACGGCAAGGCC
	Gm-SOT1$_{\text{PPR6 variant}}$ (182-706)-reverse	TATCTTAATAACCGCGTTCAGAGTCTTCTC
	Gm-SOT1$_{\text{PPR7 variant}}$ (182-706)-forward	ACTACTTTGTTGAATATGTGTGCGGATG
	Gm-SOT1$_{\text{PPR7 variant}}$ (182-706)-reverse	CATATTCAACAAAGTAGTATACAAGATA
RNA gel blot	*psbA*-forward	TATGGGTCGTGAGTGGGA
	psbA-reverse	GCCGCTAAGAAGAAATGTAA
	psaA-forward	CCCATTTCGGCCAACTCTCT
	psaA-reverse	AAGGGAGTTGCTCCTTCAGC
	petA-forward	GGCGACTGGGCGTATTGTAT
	petA-reverse	ATCTGACCCCTTCCCCTGTT
	atpB-forward	CTGGGACGTATCGCCCAAAT
	atpB-reverse	TTCCTTCACGAGTCCGTTCG
	psbO-forward	GACCCTTCGAGGTTGCTTCA
	psbO-reverse	GACCATACCACACCCCTTGG
	RNA73-forward (for probe a)	TTCATGGACGTTGATAAGATC
	RNA73-reverse (for probe a)	CCTCGCCCTTAACTTTAAGGC
	23S-forward (for probe b)	TTCAAACGAGGAAAGGCTTACGG
	23S-reverse (for probe b)	AGGAGAGCACTCATCTTGGGG
	4.5S-forward (for probe c)	GAAGGTCACGGCGAGACGAG
	4.5S-reverse (for probe c)	GTTCAAGTCTACCGGTCTGTTAG
	SOT1-forward	CTTAGTTTGTACGAAAGAGCG
	SOT1-reverse	CTTCTTCCTCCAATCCAATA
	Actin-forward	GGTAACATTGTGCTCAGTGGTGG
	Actin-reverse	CTCGGCCTTGGAGATCCACATC
Circular RT-PCR	3.2RT	CCATCCTTTGGATAGAAGGGC
	3.2F	GTTCAACTCTTTGACAACACG
	3.2R1	CAGTTCGCCAGGTTGTCT
	3.2R2	GCGTTACTCAAGCCGACATTCT
	3.2R3	GTTATCCGCTCCGCACTT
	2.9RT	GGTGGGCTTACTACTTAGATGC
	2.9F	GGGCTGTAGTATGTTCCAAG
	2.9R	CAGTTCGCCAGGTTGTCT
	2.4RT	GGTGGGCTTACTACTTAGATGC
	2.4F	GGGCTGTAGTATGTTCCAAG
	2.4R	GCGTTACTCAAGCCGACATTCT
	1.8RT	GGTACTCGAACATTGGCTC

	1.8F	GATGGTTATCGGTTCAAGG
	1.8R	CAGTTCGCCAGGTTGTCT
	0.5RT	TGGTCCTTGCTGATTCAC
	0.5F	ATGGGAGCAGCCTAAACC
	0.5R	CAGTTCGCCAGGTTGTCT
	1.1RT	GGTGGGCTTACTACTTAGATGC
	1.1F	GGGCTGTAGTATGTTCCAAG
	1.1R	CGGCTCCCCTGTCATCAG
RACE	Linker primer 1	CTAATACGACTCACTATAGGGCAAGCA GTGGTATCAACGCAGAGT
	Linker primer 2	CTAATACGACTCACTATAGGGC
	1.3 5’	GGTACTCGAACATTGGCTC
	1.3 3’	GATGGTTATCGGTTCAAGG
	23S 5’	TTACGCCAAGCTTTGGTCCTTGCTGATT CACACGGGATTC
	4.5S 3’	GATTACGCCAAGCTTGAAGGTCACGGC GAGACGAGCC
	RNA73 cleavage site	GATTACGCCAAGCTTCCTCGCCCTTAA CTTTAAGGCTATGC
RNA endonuclease activity	RNA73	Bio/FAM-UUCAUGGACGUUGAUAAGAUCUUUCC AUUUAGCAGCACCUUAGGAUGGCAUA GCCUUAAAGUUAAGGGCGAGG/-Bio
	RNA73m	Cy3-UUCAGUUCCGUUGAUAAGAUCUUUCC AUUUAGCAGCACCUUAGGAUGGCAUA GCCUUAAAGUUAAGGGCGAGG
	$RNA73_{PPR5-6}$	UUCAUGGAAUUUGAUAAGAUCUUUCC AUUUAGCAGCACCUUAGGAUGGCAUA GCCUUAAAGUUAAGGGCGAGG-Bio
	$RNA73_{PPR5-7}$	UUCAUGGAAUGUGAUAAGAUCUUUCC AUUUAGCAGCACCUUAGGAUGGCAUA GCCUUAAAGUUAAGGGCGAGG-Bio
	$RNA73_{PPR1-6}$	UUCAGUUCAUUUGAUAAGAUCUUUCC AUUUAGCAGCACCUUAGGAUGGCAUA GCCUUAAAGUUAAGGGCGAGG-Bio
EMSA	RNA14	Cy3-CAUGGACGUUGAUA
	RNA13	Cy3-AUGGACGUUGAUA
	RNA12a	Cy3-UGGACGUUGAUA
	RNA12b	Cy3-AUGGACGUUGAU
	RNA13	Cy5/Bio-AUGGACGUUGAUA
	RNA13m	Cy3-AGUUCCGUUGAUA
	oligo1	GUUGCAGCA
	oligo2	GUUGCUCCA
	oligo3	GUACCUCCA
	oligo4	CAACCUCCA
	oligo5	CAACGUCCA
	$RNA13_{PPR5-6}$	Bio-AUGGAAUUUGAUA
	$RNA13_{PPR5-7}$	Bio-AUGGAAUGUGAUA
	$RNA13_{PPR1-6}$	Bio-AGUUCAUUUGAUA

References

1. Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16(6):735-743.
2. Chi W, et al. (2008) The pentratricopeptide repeat protein DELAYED GREENING1 is involved in the regulation of early chloroplast development and chloroplast gene expression in Arabidopsis. *Plant Physiol* 147(2):573-584.
3. Peng L, et al. (2006) LOW PSII ACCUMULATION1 is involved in efficient assembly of photosystem II in *Arabidopsis thaliana*. *Plant Cell* 18(4):955-969.
4. Liu J, et al. (2012) PsbP-domain protein1, a nuclear-encoded thylakoid lumenal protein, is essential for photosystem I assembly in *Arabidopsis*. *Plant Cell* 24(12):4992-5006.
5. Zhong L, et al. (2013) Chloroplast small heat shock protein HSP21 interacts with plastid nucleoid protein pTAC5 and is essential for chloroplast development in *Arabidopsis* under heat stress. *Plant Cell* 25(8):2925-2943.
6. Puchta O, et al. (2010) DMR1 (CCM1/YGR150C) of *Saccharomyces cerevisiae* encodes an RNA-binding protein from the pentatricopeptide repeat family required for the maintenance of the mitochondrial 15S ribosomal RNA. *Genetics* 184(4):959-973.
7. Hotto AM, et al. (2015) Arabidopsis chloroplast mini-ribonuclease III participates in rRNA maturation and intron recycling. *Plant Cell* 27(3):724-740.
8. Stoppel R, Meurer J (2012) The cutting crew - ribonucleases are key players in the control of plastid gene expression. *J Exp Bot* 63(4):1663-1673.
9. Zoschke R, Watkins KP, Miranda RG, Barkan A (2016) The PPR-SMR protein PPR53 enhances the stability and translation of specific chloroplast RNAs in maize. *Plant J* 85(5):594-606.
10. Wu W, et al. (2016) 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. Plant J* 85(5):607-621.
11. Barkan A, et al. (2012). A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet* 8(8):e1002910.
12. Takenaka M, Zehrmann A, Brennicke A, Graichen K (2013) Improved computational

target site prediction for pentatricopeptide repeat RNA editing factors. *PLoS ONE* 8(6):e65343.

13. Yagi Y, Hayashi S, Kobayashi K, Hirayama T, Nakamura T (2013) Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS ONE* 8(3):e57286.