Supplementary Data

*Arabidopsis* DXO1 links RNA turnover and chloroplast function independently of its enzymatic activity

Aleksandra Kwasnik, Vivien Ya-Fan Wang, Michal Krzyszton, Agnieszka Gozdek, Monika Zakrzewska-Placzek, Karolina Stepniak, Jaroslaw Poznanski, Liang Tong and Joanna Kufel

**Supplementary Materials and Methods**

| Topic                                                                 | Page |
|----------------------------------------------------------------------|------|
| Sequence analysis                                                    | 2    |
| Yeast strains and growth conditions                                  | 2    |
| Protein mutagenesis, expression and purification                     | 2    |
| Molecular modeling                                                   | 2    |
| Fluorescence polarization                                            | 2    |
| Electrophoretic Mobility Shift Assay (EMSA)                          | 3    |
| Plant material, growth conditions and cordycepin treatment           | 3    |
| Measurement of chlorophyll content                                   | 4    |
| RNA isolation and quantification                                     | 4    |
| Whole transcriptome sequencing                                       | 4    |
| Small RNA sequencing                                                 | 4    |
| Differential expression analysis                                     | 5    |
| Plots and GO enrichment analysis                                     | 5    |
| Capture of NAD-RNA from total RNA                                    | 5    |

**Supplementary Figures**

- Supplementary Figure S1. Sequence and structure of DXO1
- Supplementary Figure S2. Plant-specific features affect DXO1 biochemical activity *in vitro*
- Supplementary Figure S3. Catalytic site and NTE cooperate in RNA binding by DXO1
- Supplementary Figure S4. Chloroplast-associated phenotype of DXO1 deficiency
- Supplementary Figure S5. RNA quality control siRNAs (rqc-siRNAs) accumulate in dxo1 plants
- Supplementary Figure S6. The N-terminal extension is crucial for DXO1 function in chloroplast-related processes

**Supplementary Tables**

- Supplementary Table S1. Summary of crystallographic information
- Supplementary Table S2. Sequences of RNA substrates used in this study
- Supplementary Table S3. List of oligonucleotides used in this study
- Supplementary Table S4. Comparison of dxo1-2 RNA sequencing and microarray analysis of retrograde signaling
- Supplementary Table S5. Summary of high-throughput experiments performed in this study

**Supplementary References**
Supplementary Materials and Methods

**Sequence analysis.** Plant DXO homologs were identified with the blastp algorithm (1). Sequences were aligned using the T-Coffee web server (2) and rendered with Jalview. Disorder probabilities were predicted with GeneSilico MetadisorderMD2 (3).

**Yeast strains and growth conditions.** DXO1 cDNA was expressed under the control of the ADH1 constitutive promoter in a pRS416 vector in wild-type (MATa, leu2, ura3, trp1, his3, abd1::LEU2, p358-ABD1) and rai1Δ (MATa, leu2, ura3, lys2, trp1 his3, abd1::LEU2, RAI1::kanMAX4, p358-ABD1) yeast strains. Yeast were grown at 30°C in YPD medium (1% yeast extract, 2% Bacto-peptone, 2% glucose) or synthetic complete medium (0.67% yeast nitrogen base, 2% glucose or 2% galactose, supplemented with amino acids and nucleotide bases).

**Protein mutagenesis, expression and purification.** Site-specific mutations (E394A/D396A and N298G) were generated using site directed-mutagenesis PCR (see the list of oligonucleotides in Supplementary Table S3) and confirmed by sequencing. Full-length *A. thaliana* AT4G17620, N-terminally truncated (lacking amino acids 1-194) and point-mutated variants were cloned into pET26b with the C-terminal His tag (Novagen). Proteins were overexpressed in *E. coli* BL21 Rosetta (DE3) cells at 20°C for 16 hours after induction with 0.5 mM IPTG. Cells were lysed by sonication in buffer A (20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 10% glycerol and 5 mM β-mercaptoethanol) and recombinant proteins were purified on TALON Metal Affinity Resin (Clontech), eluted in buffer A supplemented with 250 mM imidazol and fractionated by gel filtration chromatography on ENrich TM SEC 70 column (BioRad). Purified proteins were concentrated to 40 µM in a buffer containing 20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 2 mM DTT and 10% glycerol, flash frozen in liquid nitrogen and stored at −80°C. The concentration of the proteins was determined at 280 nm using the molar extinction coefficient calculated on the basis of protein composition by the ProtParam tool (http://web.expasy.org/protparam/).

**Molecular modeling.** Molecular modeling was performed with the aid of Yasara Structure package (ver. 18.04.24) using Yasara2 force-field (4). The structure of DXO1(ΔN194) was modeled in two repetitions by homology, using the template of the crystal structure of mouse DXO in complex with either p-RNA (4J7L, Protein Data Bank) or 3'-NADP (5ULI, Protein Data Bank) in the presence of divalent cation(s). The final model of the NAD⁺-capped polyU substrate (U5) combined the two models above with a fixed chemical bond between 5' phosphorus of p-RNA and 3'O of NAD adenine moiety. The resulting structure of NAD⁺-capped U5 bound to DXO1(ΔN194) was then preoptimized in the neighborhood of the modified phosphate group (5A threshold) and further refined with several cycles of simulated annealing procedure, in high-temperature steps, with the constrained coordinates of all backbone atoms of RNA and protein. Other RNA substrates (p-, ppp- and m⁷Gppp-) positioned in the DXO1 binding site were obtained by local modification of the NAD⁺-capped polyU complex. All complexes were then subjected to 20ns molecular dynamics at 198 K, with the protein in each simulation initially solvated with a water shell of a 4Å thickness. The latter procedure was also repeated for all substrates in complexes with DXO1(ΔN194/N298G), which initial structures were taken from the final snapshots of molecular dynamics (MD) traces recorded for the corresponding complex of the native protein.

**Fluorescence polarization.** Fluorescence polarization binding assays were performed to determine the binding affinities of DXO1(wt), DXO1(E394A/D396A), DXO1(ΔN194), DXO1(N298G) and DXO1(ΔN194/N298G) to...
RNA substrates fluorescently 3’-labeled with 6-carboxyfluorescein (FAM) (see sequences of RNA substrates in Supplementary Table S2). The equilibrium dissociation constants values ($K_d$) of RNA substrate with DXO1 variants were determined by monitoring the total fluorescence polarization of mixtures of the fluorescent substrate (10 nM) and proteins at increasing concentrations up to full saturation (16 µM) in a final volume of 100 µl in the binding buffer (30 mM Tris-HCl pH 8.5 mM NH$_4$Cl, 2 mM CaCl$_2$, 0.5 mM DTT, 0.1 mM spermine). Fluorescence polarization at 535 nm (485 nm excitation) was measured using the PARADIGM Detection Platform (Beckman Coulter) in FP fluor 384 black plates (Greiner). The experimentally derived fluorescence data were analyzed according to the two state model for 1:1 RNA binding. The binding affinities and anisotropy levels characteristic for bound and free states, together with the associated standard deviations were estimated using appropriate model implemented in origin 9 (www.originlab.com). For each protein-RNA system the dose-response effect was monitored at increasing protein concentration, while concentration of labeled RNA remained constant. All measurements were at least duplicated. Dissociation constants were estimated for all repetitions globally, whereas the apparent fluorescence polarization of free RNA and that of RNA bound to the protein were fitted independently for each measurement.

**Electrophoretic Mobility Shift Assay (EMSA).** For the detection of protein-RNA complexes gel electrophoretic mobility shift assay (EMSA) was performed with five DXO1 variants (wt, N298G, E394A/D396A, ΔN194 and ΔN194/N298G). Proteins were incubated on ice with radioactively labeled substrates for 60 min in a binding buffer that contained 30 mM Tris-HCl pH 8.0, 50 mM NH$_4$Cl, 2 mM CaCl$_2$ and 0.5 mM DTT. Samples were separated in a native 10% acrylamide gel (37.5:1 acrylamide:bis acrylamide) with 10% glycerol at 10W in 4ºC. Gels were analyzed with PhosphorImager Typhoon FLA 9000 (GE Healthcare).

**Plant material, growth conditions and cordycepin treatment.** Arabidopsis thaliana wild-type ecotype Columbia (Col-0) was used in this study. The following homozygous lines were selected: SALK_103157 (dxo1-1) and SALK_032903 (dxo1-2) with T-DNA insertions in exon II and intron V of DXO1 gene (AT4G17620), respectively. rdr6 plants that contain mutations in the AT3G49500 gene were already characterized (5). The homozygous dxo1-2/rdr6 double mutant was created by crossing and plants were selected by PCR genotyping. Seeds were surface sterilized with 30% bleach 0.02% Triton-X100 solution, stratified for 2 days at 4ºC and grown on soil or on MS medium (6) supplemented with 1% (w/v) sucrose and 0.3% phytagel under long day conditions (16h light/8h dark). DXO1(wt), DXO1(E394A/D396A), DXO1(ΔN194) and DXO1(ΔN194/E394A/D396A) transgenic lines expressing DXO1 variants fused to GFP under the control of the constitutive 35S promoter from the cauliflower mosaic virus (CaMV) were generated by transforming dxo1-2 plants with *Agrobacterium tumefaciens* strain GV3101 carrying pGWB605-DXO1(wt), pGWB605-DXO1(E394A/D396A), pGWB605-DXO1(ΔN194) and pGWB605-DXO1(ΔN194/E394A/D396A) plasmids using the floral-dip method (7). Seeds from *A. tumefaciens*-treated plants were selected based on their resistance to BASTA. Plant binary pGWB605 vectors were obtained by the LR recombination (Invitrogen) with pENTRY1A plasmids containing DXO1 cDNA sequences cloned between Sall and NotI sites (see the list of oligonucleotides in Supplementary Table S3). mRNA half-life measurement experiments were carried out as described (8). Two-week-old seedlings were transferred to flasks containing a buffer (1 mM PIPES, pH 6.25, 1 mM sodium citrate, 1 mM KCl, 15 mM sucrose), and after a 30-min incubation, cordycepin (150 mg/l) was added. Material was collected at 0, 15, 30, 60, 90 and 120 min time points after transcriptional inhibition.
Measurement of chlorophyll content. Chlorophyll (Chl) extraction and quantification were performed using 0.5 g of leaves from tree-week-old plants as previously described (9). Briefly, Chl was extracted with 100% acetone and quantified spectrophotometrically at 662 and 645 nm. Chla and chlb contents were calculated according to the following equations: 
\[ \text{chla} = 12.25 \times A_{662} - 2.79 \times A_{645} \quad (\mu g \text{ per ml solution}) \]
\[ \text{chlb} = 21.50 \times A_{645} - 5.10 \times A_{662} \quad (\mu g \text{ per ml solution}) \]

RNA isolation and quantification. Total yeast RNA was isolated from cultures at OD 0.2 with the hot phenol extraction method (10). Total plant RNA was isolated from two-week-old seedlings or three-week-old plants grown in soil using Trizol reagent (Sigma) according to manufacturer’s instructions. For northern blot analysis, total RNA (15 µg) was resolved in 1% denaturing agarose gel or in 10% urea-polyacrylamide gel and transferred by capillary elution or electrotransfer to a Hybond N+ membrane (Amersham). For siRNA detection, total RNA (15 µg) was resolved in 15% urea-polyacrylamide gel, electrotransferred to a Hybond N membrane (Amersham) and chemically cross-linked with 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) (11).

Random primed probes were amplified using cDNA template with appropriate primers (Supplementary Table S3) and radioactively labeled with DecaPrime kit (Ambion) and \([\alpha-^32P]ATP\) (Hartmann Analytics). Oligoribonucleotide probes (Supplementary Table S3) were radioactively labeled using PNK (Thermo Scientific) and \([\gamma-^32P]ATP\) (Hartmann Analytics). Membranes were hybridized overnight with radioactive probes in PerfectHyb buffer (Sigma), washed, analyzed with PhosphorImager Typhoon FLA 9000 (GE Healthcare) and quantitated with ImageJ software. cDNA for RT-PCR and RT-qPCR was prepared for 1 hour at 55°C with a mix of oligo(dT)20 and 10xRT Random Primers (Applied Biosystems) using a SuperScript® III Reverse Transcriptase Kit (Invitrogen) and 10 µg of total RNA treated with RNase-free TURBO DNase (Ambion) for 60 min at 37°C. qPCR was carried out using LightCycler® 480 SYBR Green I Master (Roche) and primers listed in Supplementary Table S3. UBC9 (AT4G27960) mRNA was used for normalization of expression values. All reactions were performed in three independent biological replicates.

Whole transcriptome sequencing. Libraries were prepared on total RNA isolated from three biological replicates of three-week-old Col-0 and dxa1-2 plants grown at the same time, using Illumina TruSeq Stranded Total RNA with Ribo-Zero Plant rRNA Removal (Plant Leaf) protocol including barcoding and were then paired-end sequenced on HiSeq4000 (DNA Research Centre, Poznan, Poland). Raw sequence data (see Supplementary Table S5 for library read depths) was deposited in the GEO database (GSE95473 and GSE99600). Fastq files were quality checked with fastqc (v0.10.1-http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) with all the replicates showing high quality RNA-seq data. Reads were aligned to TAIR10 A. thaliana genome from ensembl release v29 (12) using HISAT2 v2.0.4 (13) with the following command-line parameters: --fr --rna-strandness RF --known-splicesite-infile. Mapped reads were sorted using samtools sort v1.1 (14).

Small RNA sequencing. Libraries were prepared on total RNA isolated from three biological replicates of three-week-old Col-0 and dxa1-2 plants using NEB Next Small RNA Library Prep Set for Illumina including PAGE selection of small RNA and sequencing using Illumina HiSeq4000 in 50bp single-end mode (BGI, Beijing, China). Obtained fastq files (see Supplementary Table S5 for library read depths) were quality checked using fastqc (v0.10.1 -http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) with all the replicates showing high quality RNA-seq data. Adaptor sequences (Illumina Small RNA Adapter2 TCGTATGCGGTCTTCTGGTGT) were removed using cutadapt (v1.9.dev6 -http://cutadapt.readthedocs.io/en/stable/guide.html) and reads were
quality trimmed with sickle se (v0.940 -https://github.com/najoshi/sickle) with the following command-line parameters: -t illumina -q 20 -l 20. Reads were mapped to TAIR10 A. thaliana genome from ensembl release v29 (12) using bowtie v1.0.0 (15) with the following command-line parameters: -phred33 -n 1 -k 10 -m 10 -l 18 -best -strata. Mapped reads were sorted using samtools sort v1.1 (14).

**Differential expression analysis.** Reads were counted with htseq-count v0.6.0 (16) and the following command-line parameters: -a 0 -s reverse (whole transcriptome sequencing) or -a 0 -s no (small RNA sequencing) using Araport11 gene annotation release 201604 (17). Differential expression (DE) was performed using DESeq2 v1.8.2 (18) R (v3.2.2) package with parameter alpha = 0.05. Genes with FDR < 0.05 were considered significantly changed and those with FDR < 0.05 and absolute log2FC > 1 were considered strongly affected.

To estimate the length of siRNA classes responsible for observed differences, DE analysis was performed on reads grouped according to their length using reformat.sh with parameters minlength and maxlength from bbmap (v35.x -https://sourceforge.net/projects/bbmap/). 20-22-nt and 23-25-nt small RNAs were considered 21-nt and 24-nt length classes, respectively.

**Plots and GO enrichment analysis.** RNA-seq alignments were split to separate read-pairs that originate from transcription on the forward and reverse strands using samtools v1.1 (14). To isolate the forward strand alignments we use the bitflag filters -f 128 -F 16 for R2 reads and -f 64 -F 32 for their R1 pairs, producing two separate alignment files which were then merged. To isolate the reverse strand alignments we use the bitflag filters -f 144 for R2 reads and -f 96 for their R1 pairs, again producing two separate alignment files which were then merged. Alignment coverage graphs were calculated with genomeCoverageBed from bedtools v2.17.0 (19) for all the alignment files with normalization to number of reads and were converted to bigwig format with bedGraphToBigWig (v4) from the UCSC Genome Browser application binaries collection (http://hgdownload.cse.ucsc.edu/admin/exe/linux.x86_64/). Average profiles were created using seqplots (v1.6.0 -http://przemol.github.io/seqplots/) R package. GO enrichment analysis was performed using AmiGO (v1.8 -http://amigo.geneontology.org) with FDR < 0.01.

**Capture of NAD-RNA from total RNA.** NAD capture was performed as previously described (20). Samples of total RNA (100 µg) were incubated at 37°C for 30 min in 100 µl reaction containing 10% 4-pentyn-1-ol, 25 ng/µl ADPRC (Sigma) and buffer (50 mM HEPES pH 7, 5 mM MgCl2). Reaction without ADPRC was used as a negative control. The reaction was stopped by phenol/chloroform extraction and RNA was ethanol precipitated in the presence of glycogen. The subsequent CuAAC reaction was performed in 100 µl of 50 mM HEPES (pH 7) and 5 mM MgCl2 buffer with biotin-PEG3-azide (250 µM) and a freshly prepared mixture of CuSO4 (1 mM), THPTA (0.5 mM) and sodium ascorbate (2 mM) for 30 min at 25°C. Samples were again purified by phenol/chloroform extraction and ethanol precipitated in the presence of glycogen. 20 µl of Dynabeads M-280 Streptavidin (ThermoFisher) for each sample were washed three times with immobilization buffer (10 mM HEPES, 1 M NaCl, 5 mM EDTA, pH 7.2). Beads were blocked with 100 µl of 100 µg/ml BSA and washed three times with immobilization buffer. Each RNA sample, dissolved in 40 µl of immobilization buffer, was incubated with beads at 20°C at 1,000 rpm. for 1h. Beads were washed with streptavidin washing buffer (50 mM Tris-HCl, 8 M urea, pH 7.4) and captured RNA was eluted by phenol/chloroform extraction and ethanol precipitated in the presence of glycogen. cDNA was obtained using SuperScriptIII and random primers as described and then quantified using qPCR.
Supplementary Figure 1

A

ADXO1 residues modified in this study (Asn268, Gln394, Asp396)

- Boarder of DXO1 N-terminal extension deleted in this study (ΔN194)

- Six conserved motifs that form DXO/Rat1 active site

- DXO1 phosphorylation site (Ser119)

- Residues that contact mG cap in MmDXO

- Residues that contact p-RNA in MmDXO

- Residues responsible for SpRat1 binding to Rat1

B

C

WT

N298G
dN194
E344A/D396A

Residues

0.0
0.5
1.0

DXY disorder probability (dp)

0
250
500

Residues

0
250
500

Ser119

GOXO (II)

EIO (III)

KIQ (IV)

R (I)
Supplementary Figure S1. Sequence and structure of DXO1. (A) T-Coffee sequence alignment of *Arabidopsis thaliana* DXO1, GmDXO1 (*Glycine max*), VvDXO1 (*Vitis vinifera*), OsDXO1 (*Oryza sativa*), MmDXO1 (*Mus musculus*), KIDxo1 (*Kluyveromyces lactis*) and SpRai1 (*Schizosaccharomyces pombe*). Six motifs that form DXO active site are indicated (21). Sequence identities are given at the end of alignment. Bold highlight, conserved residues (threshold 0.7); red asterisks, residues modified in this study; green arrow, estimated border of plant-specific N-terminal extension; legend below the alignment explains marking used for residues that contact m7G cap and 5’ monophosphate RNA in MmDXO (22), responsible for the interaction of SpRai1 with Rat1 (23) and phosphorylation site in DXO1 (24). (B) Analysis of *S. cerevisiae rai1Δ* phenotype complementation by DXO1. Northern blot of 5.8S and 25S rRNA maturation. Schematic representation of polycistronic 35S pre-rRNA with regions corresponding to mature 18S, 5.8S and 25S rRNAs (gray), internally transcribed sequences (white) and position of oligonucleotide probes. (C) Disorder probabilities predicted with GeneSilico MetadisorderMD2 method (dp > 0.5 disordered regions, dp < 0.5 ordered regions) (3). Six motifs that form DXO1 active site are indicated.
Supplementary Figure S2. Plant-specific features affect DXO1 biochemical activity in vitro. (A-B) DXO1 activity toward in vitro transcribed uniformly $^{32}$P-labeled 5' triphosphate RNA substrates with 33% (A) and 3% (B) guanosine content. Reaction products resolved by 8% urea-PAGE. (C) DXO1 activity toward in vitro transcribed uniformly $^{32}$P-labeled 5' monophosphate RNA substrate with 3% guanosine content. Reaction products resolved by 8% urea-PAGE. (D) DXO1 activity toward uniformly $^{32}$P-labeled NAD$^+$-capped RNA substrate resolved by 8% urea-PAGE.
Supplementary Figure S3. Catalytic site and NTE cooperate in RNA binding by DXO1. (A-C) Structural modeling of DXO1(ΔN194) and DXO1(ΔN194/N298G) with U5 substrates that contain m7G-cap (A), 5’ triphosphate in the orientation that favors triphosphohydrolase activity (TPH) (B) or 5’ triphosphate in the orientation that favors pyrophosphohydrolase activity (PPH) (C). RNA oligonucleotide and functional residues are labeled and shown as stick models. Hydrogen bonds are represented as pink arrows. Position of the magnesium ions in DXO1(ΔN194) is marked with green circles in both structures. Note that, the change in the position of magnesium ions in DXO1(ΔN194/N298G) occurs for the activities that are stimulated by this modification in vitro, namely m7G-cap hydrolysis (A) and triphosphonucleotide hydrolysis (B). Structural rearrangements are accompanied with the significant reduction in the number of hydrogen bonds, from six to four for the m7G-cap (A) and from eight to four for 5’-ppp when the complex has the TPH orientation (B). Positioning of magnesium ions and number of hydrogen bonds (three) are essentially unaffected in DXO1(ΔN194/N298G) if the complex with 5’ triphosphate RNA that has the PPH orientation (C). In accordance, PPH activity is not enhanced upon the N298G modification (Figure 2B). (D) Electrophoretic mobility shift assay (EMSA) of DXO1 variants and radioactively labeled m7Gppp-, Gppp-, ppp- and NppA-RNA substrates. Incubation periods, variants and concentrations of recombinant proteins and types of RNA substrates are indicated at each panel. All assays were repeated at least three times.
Supplementary Figure S4. Chloroplast-associated phenotype of DXO1 deficiency. (A) 28-day-old Col-0, dxo1-1 and dxo1-2 plants. (B-C) Comparison of levels of selected mRNAs (B) and mRNA level downstream of the T-DNA insertion in AT4G17620 (C) in dxo1-1 and dxo1-2 measured by RT-qPCR and quantified relative to Col-0 (normalized to 1). Gene identifiers are indicated below the graph. Results are the mean of three independent biological replicates with error bars representing standard deviation; *P ≤ 0.05; **P ≤ 0.01; ***P ≤ 0.001 (t-test); FC, fold change. (D) 9,725 Araport11 annotated genes show differential expression in the dxo1-2 mutant. MA-plot of sequencing results (|log₂FC| > 0 as a function of mean expression level on a logarithmic scale); statistically significant hits (FDR < 0.05) are shown in blue; FC, fold change.
Supplementary Figure 5

A

B

C

D

E

F

Stabilized and nqc-siRNA-accumulating mRNAs

Rqc-siRNA-accumulating mRNAs
Supplementary Figure S5. RNA quality control siRNAs (rqc-siRNAs) accumulate in dxo1 plants. (A) Half-lives of selected mRNAs by northern blot after transcriptional inhibition by cordycepin. T1/2, mean values of transcript levels relative to untreated plants; T0, relative level of transcript in dxo1-2 versus Col-0 at zero time point; based on three biological replicates. 18S rRNA; loading control. (B) Venn diagrams depicting the overlap between genes with up-regulated or down-regulated siRNA levels and siRNAs in 21-nt and 24-nt length categories (FDR < 0.05; |log2FC| > 1). DE analysis was performed separately for two length categories of filtered small RNA-seq data. Numbers represent genes identified in each experiment (outside the diagrams) and overlapping genes (inside the diagrams). (C) Overlap in siRNA accumulation between dxo1-2, dcp2-1 and vcs-6 depicted in Venn diagram that includes siRNAs significantly up-regulated in dxo1-2 (FDR < 0.05; log2FC > 1) and genes published in (25). Numbers represent siRNA-accumulating genes identified in each mutant (outside the diagrams) and overlapping genes (inside the diagrams). (D-E) Stabilized (D) and rqc-siRNA-accumulating (E) mRNAs do not show defects in deNADding and capping. Total RNA treated with indicated combination of DXO1, Dcp2 and Terminator was analysed by RT-qPCR. mRNA levels were normalized to Pol III-transcribed capless pre-tRNA-snoRNA (tsnoRNA) negative control and calculated relative to untreated Col-0 and dxo1-2. Gene identifiers and levels of siRNA accumulation (log2FC) are indicated below the graph. Results are the mean of three biological replicates, error bars represent standard deviation; *P ≤ 0.05; **P ≤ 0.01; ***P ≤ 0.001 (t-test). (F) Levels of mRNA enriched by NAD capture in Col-0 and dxo1-2 plants. Gene identifiers and levels of siRNA accumulation (log2FC) are indicated below and above the graph, respectively. (D-F) Results are the mean of three biological replicates, error bars represent standard deviation; *P ≤ 0.05; **P ≤ 0.01; ***P ≤ 0.001 (t-test).
Supplementary Figure S6. The N-terminal extension is crucial for DXO1 function in chloroplast-related processes. (A) mRNA levels of DXO1 in dxo1-2 and dxo1-2 expressing DXO1 variants. Results are the mean of three independent biological replicates with error bars representing standard deviation. (B) Comparison of dxo1-2 RNA-seq data with microarray analysis of genome-wide changes upon light acclimation in wild-type plants and stn7/psad1-1 plants with altered redox-dependent retrograde signaling (26). Low light, 8 μmol quanta m⁻² s⁻¹; normal light, 80 μmol quanta m⁻² s⁻¹; high light, 800 μmol quanta m⁻² s⁻¹. See Supplementary Table S4 for the full comparison of the datasets.
Supplementary Table S1

| Structure                        | DXO1(ΔN194) |
|---------------------------------|-------------|
| Data collection                 |             |
| Space group                     | P2 1 2 1    |
| Cell dimensions                 |             |
| \( a, b, c (\text{Å}) \)       | 69.1, 80.9, 125.9 |
| \( \alpha, \beta, \gamma (\text{°}) \) | 90, 90, 90 |
| Resolution (Å) \(^1\)          | 50-1.8 (1.86-1.8) |
| \( R_{\text{merge}} \) (%)     | 5.8 (40.0)  |
| \( I/\sigma I \)               | 20.7 (3.6)  |
| Completeness (%)                | 98.9 (98.3) |
| Redundancy                      | 4.5 (4.3)   |
| Refinement                      |             |
| Resolution (Å)                  | 50-1.8      |
| No. reflections                 | 61,925      |
| \( R_{\text{work}} / R_{\text{free}} \) (%) | 20.1 / 24.2 |
| No. atoms                       |             |
| Protein                         | 5,476       |
| Ligand/Ion                      | –           |
| Water                           | 467         |
| B-factors                       |             |
| Protein                         | 29.4        |
| Ligand/Ion                      | –           |
| Water                           | 35.9        |
| R.m.s. deviations               |             |
| Bond lengths (Å)                | 0.011       |
| Bond angles (°)                 | 1.4         |

Supplementary Table S1: Summary of crystallographic information. The numbers in parentheses are for the highest resolution shell. One crystal was used for each data collection.

Supplementary Table S2

| RNA substrate                                      | Sequence                                                                 |
|----------------------------------------------------|--------------------------------------------------------------------------|
| FAM-labeled                                        | ACUCACUCACUCACCCCAAAAAAA[\text{FAM}]                                   |
| Guanosine-rich (from p-GEM template)               | GGGCGGAAUUCGAGCUCGGUACCCGGGAUCCUCUAGAGUCGACC UGCAGGGCAUGCAAGCUUGAGAGUAUUCU \text{AUGUGUACCUAA} |
| Guanosine-deficient (from yeast oriIV-1 DNA template) | GCACCCACACACCAACACACACCUACCCCUAACACAUCCCCUACCUAAC ACUGAUUUCUACCUGUCUCCAAACCUACCCUCACAUUACCUACCUCCCUACCU CCCACUCGUUACCC |
| Adenosine-deficient                                | ACUCUCUCUCUCUCUCUCUCUCUCUUCUCUCUCUUGCUUCU CCUCUCUCUCUCUUCCCGGCUUCU |

Supplementary Table S2: Sequences of RNA substrates used in this study.
| Application | Primer sequence |
|-------------|-----------------|
| **dxo1-1 genotyping**  
(T-DNA insert detection) | F: ATTTTGCCGATTTCGGACC  
R: AGGTCTGCTTATACCCATC |
| **dxo1-1 genotyping**  
(wild-type allele detection) | F: CTGGCTGCTTATACCCATC  
R: ATTTTGCCGATTTCGGACC |
| **dxo1-2 genotyping**  
(T-DNA insert detection) | F: CTGGCTGCTTATACCCATC  
R: ATTTTGCCGATTTCGGACC |
| **dxo1-2 genotyping**  
(wild-type allele detection) | F: CTGGCTGCTTATACCCATC  
R: ATTTTGCCGATTTCGGACC |
| **rdr6 genotyping**  
(T-DNA insert detection) | F: ATTTTGCCGATTTCGGACC  
R: CTAGTAGAATCACTCCAATCCG |
| **DXO1 E394A/D396A mutagenesis**  
(T-DNA insert detection) | F: ATTTTGCCGATTTCGGACC  
R: CTAGTAGAATCACTCCAATCCG |
| **DXO1 N298G mutagenesis**  
(T-DNA insert detection) | F: ATTTTGCCGATTTCGGACC  
R: CTAGTAGAATCACTCCAATCCG |
| **EcoRI/XhoI DXO1(wt) cloning for yeast transformation** | F: ACTGGAATTCATGGATTCACCGCCGAAGAAAAAT  
R: ACTGCTCGAGTTACTGAGGTGGTGGTGGTG |
| **EcoRI/XhoI DXO1(ΔN194) cloning for yeast transformation** | F: ACTGGAATTCATGGATTCACCGCCGAAGAAAAAT  
R: ACTGCTCGAGTTACTGAGGTGGTGGTGGTG |
| **NdeI/XhoI DXO1(wt) cloning for expression in E. coli** | F: ACTGGAATTCATGGATTCACCGCCGAAGAAAAAT  
R: ACTGCTCGAGTTACTGAGGTGGTGGTGGTG |
| **NdeI/XhoI DXO1(ΔN194) cloning for expression in E. coli** | F: ACTGGAATTCATGGATTCACCGCCGAAGAAAAAT  
R: ACTGCTCGAGTTACTGAGGTGGTGGTGGTG |
| **SalI/NotI DXO1(wt) cloning for Arabidopsis transformation** | F: ATCGGTCGACATGGATTCACCGCCGAAGAAAAAT  
R: ATCGCGGCGCCGCGCGGCTAGGTTGTGTGTGTGT |
| **SalI/NotI DXO1(ΔN194) cloning for Arabidopsis transformation** | F: ATCGGTCGACATGGATTCACCGCCGAAGAAAAAT  
R: ATCGCGGCGCCGCGCGGCTAGGTTGTGTGTGTGT |
| **Guanosine-rich template for transcription in vitro** | F: TGTAATACGACTCACTATAGGG  
R: TTAGTGACACTATAGGAGG |
| **Guanosine-deficient template for transcription in vitro** | F: TGTAATACGACTCACTATAGGG  
R: TTAGTGACACTATAGGAGG |
| **Adenosine-deficient with A as the initiating nucleotide** | F: TGTAATACGACTCACTATAGGG  
R: TTAGTGACACTATAGGAGG |
| **AT1G06880 RT-qPCR** | F: GTTCGATTCCCGGCTGGTG  
R: GAGCTCAGAGTAGGCGAAATC |
| **AT1G13440 RT-qPCR** | F: TCTCGATCTCAATTTCGCAAAA  
R: CGAAACCGTTGATTCCGATTC |
| **AT1G35580 RT-qPCR** | F: CAGAGTGTTATAATGGAAAGGTGTG  
R: CTTCCTTTCGATCTTCAATCTTGG |
| **AT1G37130 RT-qPCR** | F: ACGATTTCTTATGGAAAGGTGTG  
R: CTTCCTTTCGATCTTCAATCTTGG |
| **AT1G68520 RT-qPCR** | F: ATTTTGAGAGAGAACCCATAGTT  
R: ATCGCAGGCGGCAAGAAAAAT |
| **AT1G75660 RT-qPCR** | F: CCCGGATTCAACGTTCTTGAC  
R: GAGGATAGGGAGGATAGGGGC |
| **AT1G78080 RT-qPCR** | F: GTCCGATTCAACGTTCTTGAC  
R: GAGGATAGGGAGGATAGGGGC |
| **AT2G32150 RT-qPCR** | F: CTGGCCGCAACAGTATTTTGTG  
R: TCTTCCTTTCGATCTTCAATCTTGG |
| **AT2G40000 RT-qPCR** | F: TTGATAGAGAGAACCCATAGT  
R: ATCGCAGGCGGCAAGAAAAAT |
| **AT2G41430 RT-qPCR** | F: TGAGAATGGCGATGGTATCAGG  
R: AGGTGACAGATGGGTATCAGG |
| **AT2G41430 RT-qPCR** | F: TGAGAATGGCGATGGTATCAGG  
R: AGGTGACAGATGGGTATCAGG |
| Gene ID       | Technique               | Forward Primer                         | Reverse Primer                          |
|--------------|-------------------------|----------------------------------------|-----------------------------------------|
| AT2G46820    | RT-qPCR                 | F: TGTTTCTGTGAGATTCTTCCGAGGC          | R: AGAAGAGGCAGGAGCCGGAGG               |
| AT3G15210    | RT-qPCR                 | F: GCTGAGTGACCAGAAAGGTTGC            | R: GTCCGAAGGAAAGACACAGTC              |
| AT3G18780    | RT-qPCR                 | F: GGAACATTTGTGCTCAGTG              | R: CTGAGCCTTGGAGATCCACA               |
| AT3G23820    | RT-qPCR                 | F: AGCGTCTCTCAGTCTAGTC             | R: AGACTCCATCGGCACAGTTTT            |
| AT3G44260    | RT-qPCR                 | F: GCCTGATTGATCAGGTGGC            | R: CTGCTCGAGACTCAACAGC               |
| AT4G02440    | RT-qPCR                 | F: GGTGACTCGAGAAGTACTCC            | R: GGAATACGAGGAGTAGACACTG            |
| AT4G08950    | RT-qPCR                 | F: CGCTGAGATTTGCTGTTG              | R: AAACCTCCTGGATTTGCGA               |
| AT4G11280    | RT-qPCR                 | F: CGGCCGAAATTCTCTTTATTTG           | R: ACGCATCAATACCTCAAAAGAC           |
| AT4G17490    | RT-qPCR                 | F: ATCGCACCAGGAAACCAGA            | R: ATTTCGCCACAGGCTCATC              |
| AT4G17620    | RT-qPCR                 | F: GCCCATCTTAAACTCGGCTAGG         | R: GCCACCATCAGAAGGCTCTG             |
| AT4G25100    | RT-qPCR                 | F: TCGATCCTTGGCTTCTCAAC           | R: AAGCAAGAAGCTCTCTGATGAG          |
| AT4G27960    | RT-qPCR                 | F: TCCATGTAGCGAGGACCCGTG        | R: ACTTCTCCAGAATAGGGGCTATCCG       |
| AT4G29905    | RT-qPCR                 | F: AACAGACCAGAGGATGTG            | R: ACAGAAACACACTTCGCTCTC           |
| AT4G36040    | RT-qPCR                 | F: GTACTGCCATCAGACGTTG           | R: CGTACAGTAAACGCGGATGGA          |
| AT5G12140    | RT-qPCR                 | F: AAATGGCGGATCAACAAGG          | R: TTGTATGCTGACAGAGCG              |
| AT5G20160    | RT-qPCR                 | F: ACCCTATCTCTCTGAAAACCC         | R: CGATACAGGATGACAGGCAAGCG         |
| AT5G47640    | RT-qPCR                 | F: AGCCGCACACTTTAACAATTACACC    | R: TGAATGGAAAGCTGAAGCGAAGCG       |
| AT5G49100    | RT-qPCR                 | F: ACACCAGTGAGAGGAAATGTGCG        | R: ACAGCTCTCTTCCACACCACACAGC       |
| AT5G51174    | RT-qPCR                 | F: ATGATGATGGGAGATTTTGGAAG       | R: TCAGTAATGTGGCTGATTTTGAAG        |
| AT5G53902    | RT-qPCR                 | F: CTGGATAGCTCAGCTCAGCT          | R: TAACAGATATCTACAGACACTG          |
| AT5G67300    | RT-qPCR                 | F: CGATGTACAGGACACTCTCGT        | R: CTGTCGCTAGCGTTTGAGCCTG         |
| ATCC80540    | RT-qPCR                 | F: GCGACTTGCGGTATTGTGT          | R: GTCGGACCTCATAATCCACGGG         |
| AT1G37130    | random priming probe    | F: CATACACTAGGCTCCTCTCC       | R: GATGAAATGGCTGACGCGCTCTC         |
| AT2G32150    | random priming probe    | F: TTCTGAGAAGATGATTGTGTGAG       | R: GATGAATATTGTGCTACACCACACTA       |
| AT2G40000    | random priming probe    | F: CATCAACACCAGTGAAATGGAAC       | R: CGGTTGAGTCTCACTGCAACCA          |
| AT3G15210    | random priming probe    | F: GCCGAAGATGGGCTTTGAACC         | R: CAGGCTCTTCCGAGTGGAGG           |
| AT3G23820    | random priming probe    | F: TTGGGCTCTCTAGTCGCTG          | R: TCCCTCCAAAGTGAGGACCA           |
| AT3G44260    | random priming probe    | F: CGCTGACTTCCAGCTGCGGCTG       | R: CCAGCATGATTTCTCCGGCCCATC        |
| AT4G02440 random priming probe for northern blot | F: GCGTACCGGACCCATAAC  | R: CACCTCGGAATCTACTCGC |
| AT4G08950 random priming probe for northern blot | F: CAAACCCTAGCTCGAAAGG  | R: CAAAAACTTCCTGCCGTTCG |
| AT4G29905 random priming probe for northern blot | F: CGTTTCTTTCCTCCACACTTC  | R: CATTGCACAAAACCTCAGCGAGG |
| AT4G36040 random priming probe for northern blot | F: GCTTTCTCTCATCTCCAACTTC  | R: GATGAATATTGTCCACCACCCTA |
| AT5G67300 random priming probe for northern blot | F: GAGATCGGTGAGTGCGGATCTCC  | R: CCTCGACGAAATCGCCGCCAATT |
| 18SRRNA oligonucleotide probe for northern blot | GATCCTTTCCGAGGTCACCTACCG |
| RDN18-1 oligonucleotide probe for northern blot | CATGGCTTAATCTTGGAGAC |
| RDN25-1 oligonucleotide probe for northern blot | CTCCGCTTATGTGATGTC |
| RDN37-1 oligonucleotide probe for northern blot | CGCCTAGACGCTCTCTTTTA |
| RDN58-1 oligonucleotide probe for northern blot | CGATGCGAGAACCAGAGATCCG |

Supplementary Table S3. List of oligonucleotides used in this study.
### Supplementary Table S4

| Experiment/repetition | Libraries Col-0 (M reads) | Libraries dxo1-2 (M reads) |
|-----------------------|--------------------------|---------------------------|
| RNA-seq/1             | 42.6                     | 44.3                      |
| RNA-seq/2             | 45.3                     | 36.7                      |
| RNA-seq/3             | 38.9                     | 39.7                      |
| small RNA-seq/1       | 9.5                      | 11.5                      |
| small RNA-seq/2       | 9.9                      | 10.0                      |
| small RNA-seq/3       | 9.8                      | 10.0                      |

**Supplementary Table S4.** Comparison of *dxo1-2* RNA sequencing and microarray analysis of retrograde signaling (|log_{2} FC| > 0) (26). P-values of overlaps were calculated using GeneOverlap R package. Low light, 8 µmol quanta m^{-2} s^{-1}; normal light, 80 µmol quanta m^{-2} s^{-1}; high light 800 µmol quanta m^{-2} s^{-1}.

### Supplementary Table S5

| Experiment/repetition | Libraries Col-0 (M reads) | Libraries dxo1-2 (M reads) |
|-----------------------|--------------------------|---------------------------|
| low → high light vs low light | 1736 | 1419 |
| dxo1-2 downregulated (dxo_down) | 2208 | 2403 |
| dxo1-2 upregulated (dxo_up) | 1692 | 476 |
| low light vs normal light | 2643 | 2222 |
| normals (M reads) | 1431 | 220 |
| overlap dxo_down vs experiment down (exp_down) | 1333 | 470 |
| overlap dxo_up vs experiment up (exp_up) | 3.2E-73 | 744 |
| normal light vs low light | 2329 | 2218 |
| normal light vs low light | 1591 | 1536 |
| dxo1-2 upregulated (dxo_up) | 997 | 578 |
| overlap dxo_up vs experiment up (exp_up) | 1063 | 744 |
| overlap dxo_down vs experiment down (exp_down) | 3.5E-73 | 744 |
| overlap dxo_down vs experiment down (exp_down) | 0.518 | 3561 |
| normal light vs low light | 2329 | 2218 |
| double psad stn7 mutant vs Col-0 | 2811 | 2679 |
| overlap dxo_up vs experiment up (exp_up) | 931 | 415 |
| overlap dxo_up vs experiment up (exp_up) | 4.19E-106 | 999 |
| overlap dxo_down vs experiment down (exp_down) | 0.999 | 9997 |
| overlap dxo_down vs experiment down (exp_down) | 9.07E-135 | 9997 |
| overlap dxo_down vs experiment down (exp_down) | 0.002 | 3590 |
| overlap dxo_down vs experiment down (exp_down) | 73 | 3590 |

**Supplementary Table S5.** Summary of high-throughput experiments performed in this study.
Supplementary References

1. Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 1997; 25:3389–3402.

2. Tommaso P., Moretti S., Xenarios I., Orobilg M., Montanyola A., Chang J.-M., Taly J.-F., Notredame C. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* 2011; 39:W13–W17.

3. Kozlowski L.P., Bujnicki J.M. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics.* 2012; 13:111.

4. Krieger E., Vriend G. New ways to boost molecular dynamics simulations. *J. Comput. Chem.* 2015; 36: 996–1007.

5. Mourrain P., Béclin C., Elmany T., Feuerbach F., Godon C., Morel J.B., Jouette D., Lacombe A.M., Nikic S., Picault N., et al. Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell.* 2000; 26:533–42.

6. Murashige T., Skoog F. A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant.* 1962; 15:473–497.

7. Clough S.J., Bent A.F. Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J.* 1998; 16:735–743.

8. Souret F.F., Kastenmayer J.P., Green P.J. AtXRN4 degrades mRNA in Arabidopsis and its substrates include selected miRNA targets. *Mol. Cell.* 2004; 15:173–183.

9. Lichtenthaler H.K. Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes. *Methods Enzymol.* 1987; 148:350–382.

10. Schmitt M.E., Brown T.A., Trumpower B.L. A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae. *Nucleic Acids Res.* 1990; 118:3091–3092.

11. Pall G.S., Hamilton A.J. Improved northern blot method for enhanced detection of small RNA. *Nat. Protoc.* 2008; 3:1077–84.

12. Kersey P.J., Allen J.E., Armean I., Boddu S., Bolt B.J., Carvalho-Silva D., Christensen M., Davis P., Falin L.J., Grabmueller C., et al. Ensembl Genomes 2016: More genomes, more complexity. *Nucleic Acids Res.* 2016; 44:D574–D580.

13. Kim D., Langmead B., Salzberg S.L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods.* 2015; 12:357–60.

14. Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009; 25:2078–2079.

15. Langmead B., Trapnell C., Pop M., Salzberg S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009; 10:R25.

16. Anders S., Pyl P.T., Huber W., Camp J.G., Vernet B., Köhler K., Voigt B., Okita K., Mantic T., He Z., et al. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015; 31:166–169.

17. Cheng C.Y., Krishnakumar V., Chan A.P., Thibaud-Nissen F., Schobel S., Town C.D. Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant J.* 2017; 89:789–804.
18. Love M.I., Huber W., Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014; 15:550.

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

20. Winz M.L., Cahová H., Nübel G., Frindert J., Höfer K., Jäschke A. Capture and sequencing of NAD-capped RNA sequences with NAD captureSeq. *Nat Protoc.* 2017; 12:122–149.

21. Wang V.Y.F., Jiao X., Kiledjian M., Tong L. Structural and biochemical studies of the distinct activity profiles of Rai1 enzymes. *Nucleic Acids Res.* 2015; 43:6596–6606.

22. Jiao X., Chang J.H., Kilic T., Tong L., Kiledjian M. A mammalian pre-mRNA 5’ end capping quality control mechanism and an unexpected link of capping to pre-mRNA processing. *Mol. Cell.* 2013; 50:104–115.

23. Xiang S., Cooper-Morgan A., Jiao X., Kiledjian M., Manley J.L., Tong L. Structure and function of the 5’--->3’ exoribonuclease Rat1 and its activating partner Rai1. *Nature.* 2009; 458:784–8.

24. Reiland S., Messerli G., Baerenfaller K., Gerrits B., Endler A., Grossmann J., Gruissem W., Baginsky S. Large-scale Arabidopsis phosphoproteome profiling reveals novel chloroplast kinase substrates and phosphorylation networks. *Plant Physiol.* 2009; 150:889–903.

25. De Alba A.E.M., Moreno A.B., Gabriel M., Mallory A.C., Christ A., Bounon R., Balzerge S., Aubourg S., Gautheret D., Crespi M.D., et al. In plants, decapping prevents RDR6-dependent production of small interfering RNAs from endogenous mRNAs. *Nucleic Acids Res.* 2015; 43:2902–2913.

26. Gläßer C., Haberer G., Finkemeier I., Pfannschmidt T., Kleine T., Leister D., Dietz K.J., Häusler R.E., Grimm B., Mayer K.F.X. Meta-analysis of retrograde signaling in Arabidopsis thaliana reveals a core module of genes embedded in complex cellular signaling networks. *Mol. Plant.* 2014; 7:1167–90.