PNAS
www.pnas.org

Supplementary Information for

Mining for protein S-sulfenylation in *Arabidopsis* uncovers redox-sensitive sites

Jingjing Huang, Patrick Willems, Bo Wei, Caiping Tian, Renan B. Ferreira, Nandita Bodra, Santiago Agustín Martínez Gache, Khadija Wahni, Keke Liu, Didier Vertommen, Kris Gevaert, Kate S. Carroll, Marc Van Montagu, Jing Yang, Frank Van Breusegem, and Joris Messens

Marc Van Montagu, Jing Yang, Frank Van Breusegem, and Joris Messens
Email: marc.vanmontagu@ugent.be, yangjing54@hotmail.com, frbre@psb.vib-ugent.be, and joris.messens@vub.vib.be

This PDF file includes:

Supplementary text
Figs. S1 to S10
Tables S1 to S3
SI references

Other supplementary materials for this manuscript include the following:

Dataset S1
Dataset S2
Dataset S3

www.pnas.org/cgi/doi/10.1073/pnas.1906768116

MATERIALS AND METHODS

Protein Extraction. After stress treatments and BTD incubation, cells were harvested by filtration and washed 3 times with culture medium. Collected cells were lysed on ice in lysis buffer (25 mM Tris, 15 mM $MgCl_2$, 150 mM NaCl, 0.1% [vol/vol] NP-40, 1 mM NaF, 1 µM *trans*-epoxysuccinyl-L-leucylamido(4-guanidino)butane (E64), 1/50 mL ethylenediaminetetraacetic acid (EDTA)-free Ultra cOmplete tablet [Roche], 5% [vol/vol] ethylene glycol, 0.1 mg/mL 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF), and 1 µg/mL leupeptine, pH 7.6), containing 200 unit/mL catalase and 8 mM dithiothreitol (DTT). The lysate was left on ice for 20 min and incubated at room temperature for 20 min with vortex mixing at 400 rpm. Afterward, the lysate was centrifuged at 20,000*g* for 30 min at 4ºC and the supernatant was collected. Reduced cysteines were alkylated with 32 mM iodoacetamide for 30 min in the dark. Endogenous biotinylated proteins were removed by means of NeutrAvidin agarose beads (Pierce™ NeutrAvidin™ Agarose, Thermo Fisher Scientific) (beads:lysate = 1:6.5 [vol/vol]). Proteins were precipitated with a methanol-chloroform system (aqueous phase/methanol/chloroform, 4:4:1 [vol/vol/vol]) and collected at the aqueous/organic phase interface as a solid disk after centrifugation at 1,700*g* for 20 min at 4°C. Liquid layers were discarded and the proteins were washed twice in methanol/chloroform (1:1, vol/vol), followed by centrifugation at 20,000*g* for 10 min at 4°C to repellet the proteins. The protein pellets were evaporated to dryness and stored at -70°C.

Western blot analysis. Protein pellets were resuspended in PBS buffer containing 0.4% (vol/vol) sodium dodecyl sulfate (SDS). Protein concentrations were determined with the Pierce bicinchoninic acid (BCA) protein kit (Thermo Fisher). Dissolved proteins (100 µg) were reconstituted in a click reaction solution containing 0.1 mM azido-biotin, 0.25 mM $CuSO_4$, 0.5 mM 3-(4-((bis((1-tert-butyl-1H-1,2,3-triazol-4-yl)methyl)amino)methyl)-1H-1,2,3-triazol-1-yl)propan-1-ol (BTTP), and 2.5 mM ascorbate. The mixture was rotated in the dark for 1 h at room temperature and the reaction was stopped by incubation with 1 mM ethylenediaminetetraacetic acid (EDTA) for 5 min at room temperature. The reaction mix was separated by SDS-polyacrylamide gel electrophoresis (PAGE), blotted, and hybridized with a 1:40,000 dilution of horseradish peroxidase (HRP)-conjugated streptavidin (HRP-Strep).

BTD-based Chemoproteomics Approach. *Sample Preparation and Labeling.* Protein pellets were resuspended with 25 mM NH_4HCO_3 buffers, (pH 8.0) containing 0.5 M urea. Protein concentrations were determined with the Pierce bicinchoninic acid (BCA) protein assay kit (Thermo Fisher). Per sample, 4 mg of resuspended protein was digested by mass spectrometry grade trypsin/endoproteinase Lys-C mix (Promega) at a 1:25 (enzyme/substrate) ratio overnight at 37˚C. The tryptic digests were desalted with HLB extraction cartridges (Waters). The desalted samples were vacuum-dried. Desalted tryptic digests were reconstituted in a solution containing 25% (vol/vol) acetonitrile, 1 mM Az-UV–biotin, 10 mM sodium ascorbate, 1 mM Tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine (TBTA), and 10 mM $CuSO_4$, pH 6. Heavy- and light-labeled azido-tagged biotin with a photocleavable linker (Az-UV-biotin) were synthesized as described previously (1, 2). Control and stress-treated samples were labeled with Az-UV-Biotin-heavy and Az-UV-Biotin-light, respectively. The mixture was rotated in the dark for 2 h at room temperature. Excess biotin was removed by strong cation exchange (SCX) chromatography as described (2). In brief, the click mixture was diluted in four volumes of the SCX loading buffer (5 mM KH_2PO_4, 25% [vol/vol]

acetonitrile, pH 3.0) and centrifuged at 16,000*g* for 5 min, after which equal amounts of the supernatant of the heavy- and light-labeled samples were mixed. The mix was loaded in the SCX spin columns (The Nest Group) and washed twice with 150 µL SCX loading buffer by centrifugation at 110*g* for 2 min at room temperature. The peptides were eluted with the SCX elution buffer (5 mM NaH_2PO_4, 0.4 M NaCl, 25% [vol/vol] acetonitrile, pH 3.0). The BTD-labeled peptides were captured by streptavidin as described (2). The streptavidin beads (GE Healthcare) were washed twice with streptavidin-binding buffer (50 mM sodium acetate, pH 4.5), twice with streptavidin-washing buffer (50 mM sodium acetate, 2 M NaCl, pH 4.5), and twice with water. The streptavidin beads were resuspended in 1.2 mL photorelease buffer (25 mM NH_4HCO_3) and the suspensions were transferred to thin-walled glass tubes (VWR International) irradiated at 365 nm UV light (Entela, Upland, CA) for 2 h at room temperature with stirring and vacuum-dried.

LC-MS/MS. A Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific) operated with an Easy-nLC1000 system (Thermo Fisher Scientific) was used for the LC-MS/MS analyses. Samples were reconstituted in 0.1% (vol/vol) formic acid, pressure-loaded onto a 360-µm outer diameter × 75-µm inner diameter microcapillary precolumn packed with Jupiter C18 (5 µm, 300 Å; Phenomenex), and then washed with 0.1% (vol/vol) acetic acid. The precolumn was connected to a 360-µm outer diameter × 50-µm inner diameter microcapillary analytical column packed with ReproSil-Pur C18-AQ (3 µm, 120 Å; Dr Maisch) and equipped with an integrated electrospray emitter tip. The spray voltage was set to 1.5 kV and the heated capillary temperature to 250°C. The LC gradient consisted of 0 min, 7% B; 14 min, 10% B; 51 min, 20% B; 68 min, 30% B; and 69–75 min, 95% B (where A was water with 0.1% formic acid, and B was acetonitrile with 0.1% formic acid) at a flow rate of 600 nL/min. MS1 spectra were measured with a resolution of 120,000, an AGC target of 5E6, a maximum injection time of 100 ms, and a mass range from *m/z* 300 to 1,400. Higher-energy collision dissociation (HCD) MS/MS spectra were acquired with an IonTrap detector and an AGC target of 5E4, a maximum injection time of 35 ms, and collision energy of 32%. Peptide *m/z* that triggered MS/MS scans were dynamically excluded from further MS/MS scans for 18 s.

Peptide Identification and Quantification. Raw MS data were searched with the pFind Studio 3.0 software package. RAW files were converted to MGF files by means of the built-in pParse algorithm (3) and spectra were searched against a target-decoy concatenated database of the *Arabidopsis* UniProtKB proteome (UP000006548, 39,389 entries) with the pFind search algorithm (4). A precursor tolerance of 20 ppm and a fragment mass tolerance of 0.02 Da (HCD spectra) were specified. Variable modifications included methionine oxidation, Cys carbamidomethylation, and *S*-sulfenylcysteine-tagged by BTD (+418.1311 Da). No fixed modifications were searched. Search results were filtered at a false discovery threshold of 1% at the spectrum, peptide, and protein levels. For quantification, the BTD-labeled peptides were used with the pQuant algorithm as described (5, 6).

Enrichment Analysis and Subcellular Localization. All 1,394 S-sulfenylated proteins were analyzed for gene set enrichment and overrepresentations were assessed for gene ontology functions, metabolic pathways (7), and custom protein class lists of *Arabidopsis*. Families of kinases (979 proteins) and phosphatases (125 proteins) were extracted from the PlantsP database (8). Transcription factors (2,296 proteins) were derived from the Plant Transcription Factor Database 4.0 (9) and 635 oxidoreductases (enzyme classification 1). In addition, 570 *Arabidopsis* proteases and 1,786 RNA-binding proteins were derived from published curation efforts (10, 11). After comparison of pairwise overlaps between the 1,394 S-sulfenylated proteins, the *P* value of overlap sizes equal to or greater than those observed was calculated with a cumulative hypergeometric test in R, followed by the Bonferroni-Hochberg adjustment for multiple testing correction (see *SI Dataset S2*).

To determine the overrepresentation of Cys-SOHs in protein domains, we applied the same rationale of the S-sulfenylation protein overrepresentation, but on the cysteine level. To this end, the number of cysteines was extracted within the PROSITE profiles (12) of the *Arabidopsis* proteome (UP000006548, 39,389 entries) with BioMart (13). With these cysteine domain sets and the total number of cysteines in the *Arabidopsis* proteome as background, overlaps and statistical overrepresentations of the 1,537 Cys-SOHs identified in the PROSITE profiles was determined by means of the same statistics as described above.

For the subcellular localization of the S-sulfenylated proteins, we used the Multiple Marker Abundance Profiling tool available on the SUBA4 portal (14, 15). We recalculated the subcellular distributions of the assigned high-confidence marker (HCM) proteins, thus leaving the percentage of unassigned proteins out of the distribution. In total, 686 S-sulfenylated proteins were assigned a high-confidence subcellular location.

PDB Structures. For *At*MAPK4 (UniProt Q39024), a structural model was obtained from SWISS-MODEL (16) based on the structure of *At*MAPK6 (PDB identifier 5ci6, 71% sequence identity). For the *Hs*MAPK1/ERK2, *Arabidopsis* ADH1, ADH2/GSNOR, and TENA_E proteins, the PDB structures used were 2y9q, 4rqt, 3uko, and 2f2g, respectively. Structures were visualized and images were exported with Chimera (version 1.12) (17). *At*MAPK4 and *Hs*MAPK1/ERK2 structures were superimposed with the built-in MatchMaker Chimera command.

iceLogo Sequence Overrepresentation Motifs. Of the 1,537 Cys-SOH sites, 1,464 sites (95%) matched a single UniProtKB protein or had six identical amino acids preceding and after the Cys-SOH site in matching UniProtKB proteins. The remaining 5% of the Cys-SOHs matched multiple UniProtKB proteins with differing -6, +6 residue motifs and were excluded for motif analysis. A similar procedure was applied for BTD-modified peptides in human cell lines (18) and 1,273 Cys-SOH sites had unique six-residue--flanking sequences in human UniProtKB proteins (proteome UP000005640, 71,610 entries). The resulting *Arabidopsis* and human Cys-SOH--surrounding sequences were used to determine consensus motifs with iceLogo (version 1.3.8) (19). The respective human and *Arabidopsis* UniProtKB proteomes were used as background, with a regional sampling anchored at the Cys-SOH sites. From the background proteomes, amino acid frequencies can be derived and compared against those from the experimental dataset to assess over- and underrepresentation of amino acids at each position.

Alignment of Human and *Arabidopsis* Cys-SOHs. Human Cys-SOH peptides were parsed from three studies that had used carbon nucleophile probes (6, 18, 20). Peptide sequences were reallocated to the human UniProtKB proteome UP00000654 (93,588 entries, October 2017), keeping track of the Cys-SOH protein positions. The resulting 5,616 Cys-SOHs in 3,584 human proteins were compared with the 1,537 *Arabidopsis* Cys-SOH sites that mapped to the 1,938 identified *Arabidopsis* UniProtKB proteins (*SI Dataset S1*). To this end, human and *Arabidopsis* proteins were considered orthologous when residing in the same eggNOG eukaryotic group (21), where after multiple sequences of homologous proteins were aligned with Clustal Ω (v1.2.2). In the resulting alignments, the relative positions of human and *Arabidopsis* Cys-SOHs were compared and divided over four categories (Fig. 4*B*) by using in-house Perl scripts.

***In vitro* MAPK Expression and Purification.** The *At*MAPK4-coding sequence (AT4G01370), with attL and PreScission (EVLFQ/GP) sites at the N-terminal end, was synthesized by the Gen9 company. The AtMAPK4$_{C181S}$ mutant was generated with the

QuikChange® site-directed mutagenesis kit (Stratagene). The gene fragments were inserted into the pDEST17 vector (Life Technologies). The constructs were transferred into *Escherichia coli* C41(DE3) strain. The protein expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 20°C overnight. Cell pellets were collected by centrifugation and resuspended in 50 mM Tris-HCl, pH 7.8, 200 mM NaCl, 2 mM DTT, 5 mM imidazole, 20 mM $MgCl_2$, 50 µg/mL Dnase I, 1 µg/mL leupeptine, and 0.1 mg/mL AEBSF. Cells were lysed by cell cracker at 20 KPsi. The His-tagged recombinant *At*MAPK4 and *At*MAPK4$_{C181S}$ were purified on Ni^{2+} Sepharose columns (GE Healthcare). The protein concentration was determined at 280 nm.

Dimedone Labeling and Detection by LC-MS/MS. Purified *At*MAPK4 was reduced with 20 mM DTT at room temperature for 1 h. Excess of DTT was removed with Bio-Spin size exclusion columns (Bio-Rad). Reduced *At*MAPK4 (20 µM) was incubated with 2 mM dimedone and 1 mM or 10 mM H_2O_2 at room temperature for 1 h. Excess of H_2O_2 and dimedone were removed by means of a Bio-Spin size exclusion column. The reaction was stopped by incubation with 4 mM N-ethylmaleimide (NEM) in the dark at room temperature for 10 min, blocking the free thiols of *At*MAPK4. To prevent oxidation, the protein samples were prepared in an A35 anaerobic workstation (Don Whitley Scientific). The protein samples were analyzed with sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). S-sulfenylation was visualized on Western blot with an anti-cysteine sulfenic acid antibody (Merck Millipore).

The bands were excised after staining with Coomassie Brilliant Blue (CBB) and digested with in-gel trypsin. The peptides were analyzed by LC-MS/MS as described (22). Peak lists were generated by means of the application spectrum selector in the Proteome Discoverer 2.2 package. From RAW files, MS/MS spectra were exported as individual files in .dta format with the following settings: peptide mass range, 400–3500 Da; minimal total ion intensity, 5000; and minimal number of fragment ions, 12. The resulting peak lists were searched with SequestHT against a custom database containing the *At*MAPK4 sequence. Variable modifications were S-sulfenic dimedone (+138.068 Da), S-sulfinic acid (+31.989 Da) and NEM (+125.047) for cysteine; phosphorylation (+79.966 Da) for Ser, Thr and Tyr; oxidation (+15.994 Da) for Met. Peptide matches were filtered with charge-state against cross-correlation scores (Xcorr) and phosphorylation sites were validated manually. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (23) partner repository with the dataset identifier PXD013588.

Kinase Assay for *At*MAPK4 and *At*MAPK4$_{C181S}$. *Kinase assay in gel*. Radioactive kinase assays were carried out for the reduced *At*MAPK4 and *At*MAPK4$_{C181S}$ proteins that had been prepared as described above. The kinase assay was adapted from a previous protocol (24). The reaction was performed in 50 mM Tris-HCl, pH 7.5 and with 10 mM $MgCl_2$, 10 µM Myelin Basic Protein (MBP) (Sigma-Aldrich), 50 µM ATP, 1 mM DTT, and 0.1 mCi/mL ^{32}P-ATP (Perkin Elmer). To start the reaction, 2 µM reduced or *At*MAPK4 protein was added to the mixture. The reaction was stopped by adding the loading buffer for the SDS-PAGE analysis after 0, 5, 10, 20, 30, and 45 min. Each sample was analyzed on CBB-stained SDS-PAGE gels. The destained gels were dried on a gel drier and the gels were exposed to the phosphor screen overnight. The phosphorylation signal was detected with a phosphorimager (Bio-Rad).

*Kinase assay in solution. At*MAPK4 (wild-type or mutant) was diluted to a final concentration of 2 µM in kinase buffer solution (50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 200 mM NaCl, 1 mM DTT) with 5 mM ATP, (6000-10000 cpm/nmol [γ-^{32}P] ATP) and

incubated for 1 h at 30 ºC. Reactions were started by adding MBP (200 µM final concentration) and evaluated after 45 min. Reactions were stopped by 1 min boiling in SDS-PAGE loading buffer. Samples were analyzed on SDS-PAGE gel and the bands of *At*MAPK4 (wild type or C181S variant) and MBP were excised and the incorporated ^{32}P radioactivity was determined in a liquid scintillation counter (Perkin Elmer, Tri Carb 2810 TR). When evaluating non-phosphorylated *At*MAPK4 activity, a similar protocol was followed, except for the absence of ATP in the kinase buffer solution during the 1 h incubation step. In this case, ATP was added together with MBP to start of the kinase reaction. Three independent assays were performed to determine *At*MAPK4 activities at fixed time, the results were evaluated with a Student's *t*-test considering a significance level lower than 0.01.

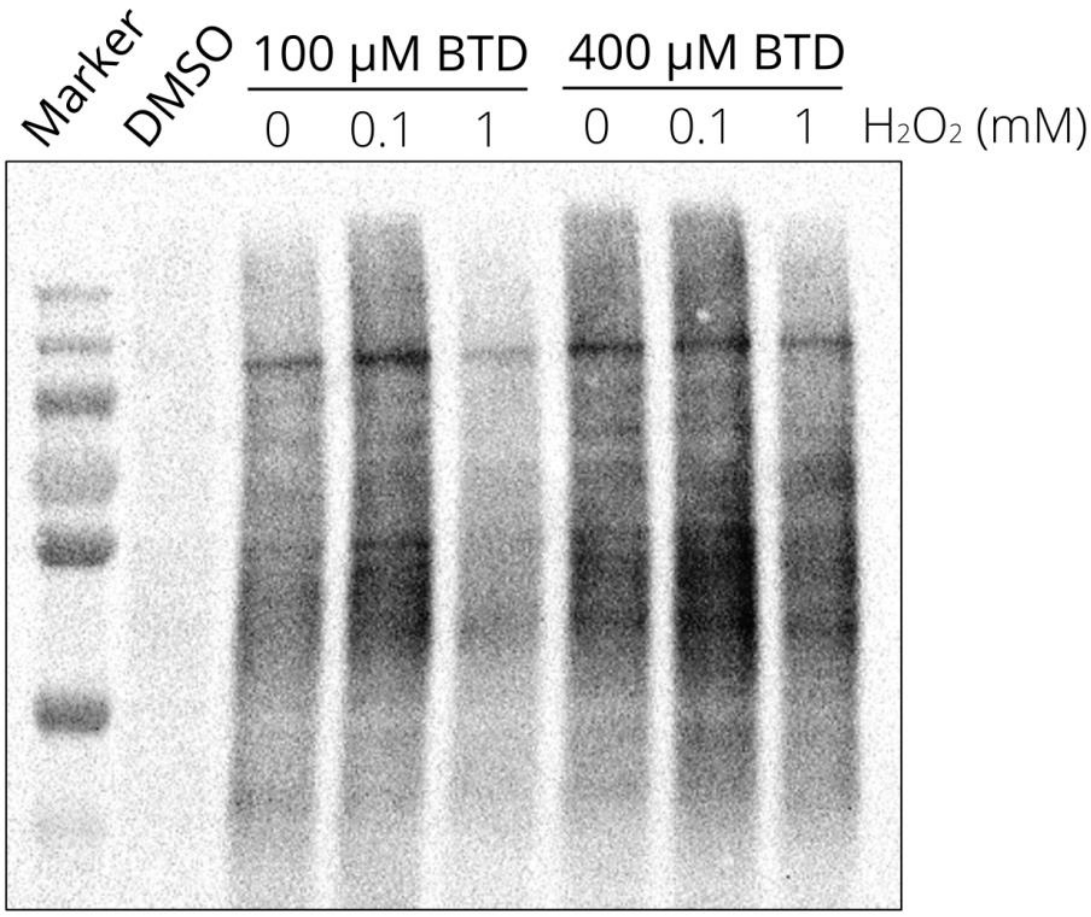

Fig. S1. Western blot detection of S-sulfenylated proteins to optimize the H_2O_2 (0, 0.1, or 1 mM) and BTD probe (100 or 400 µM) concentrations.

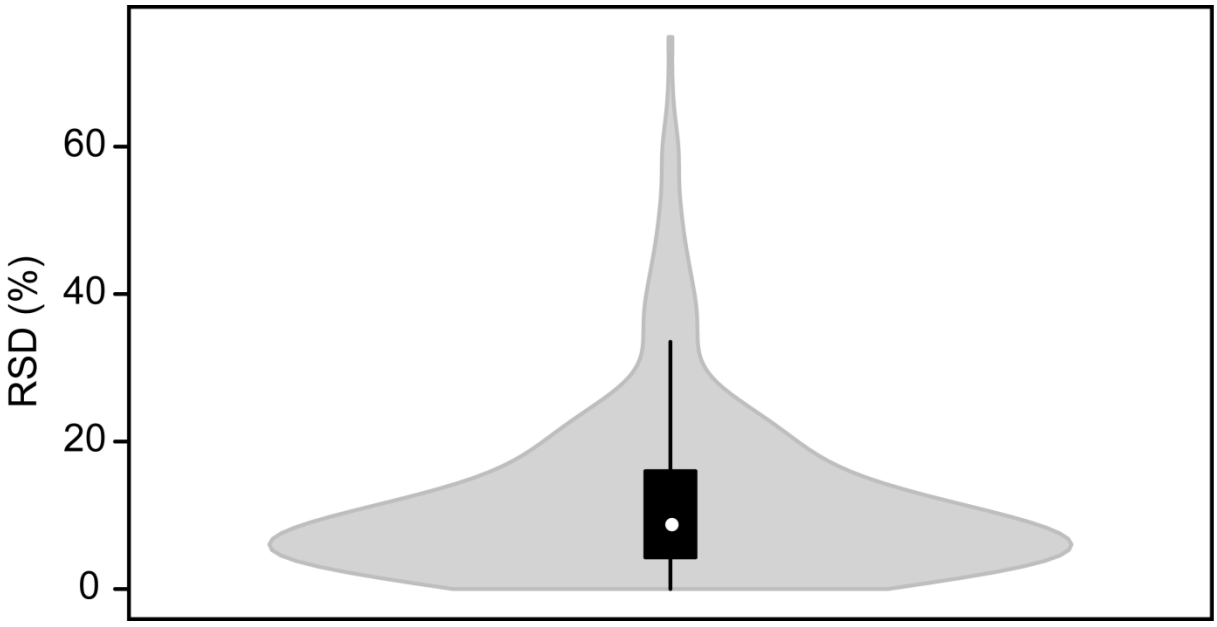

Fig. S2. Violin plot showing the distribution of the relative standard deviation (RSD) of S-sulfenylated cysteines that could be quantified in at least two biological replicates. Median and average RSD were 6.61% and 8.95%, respectively.

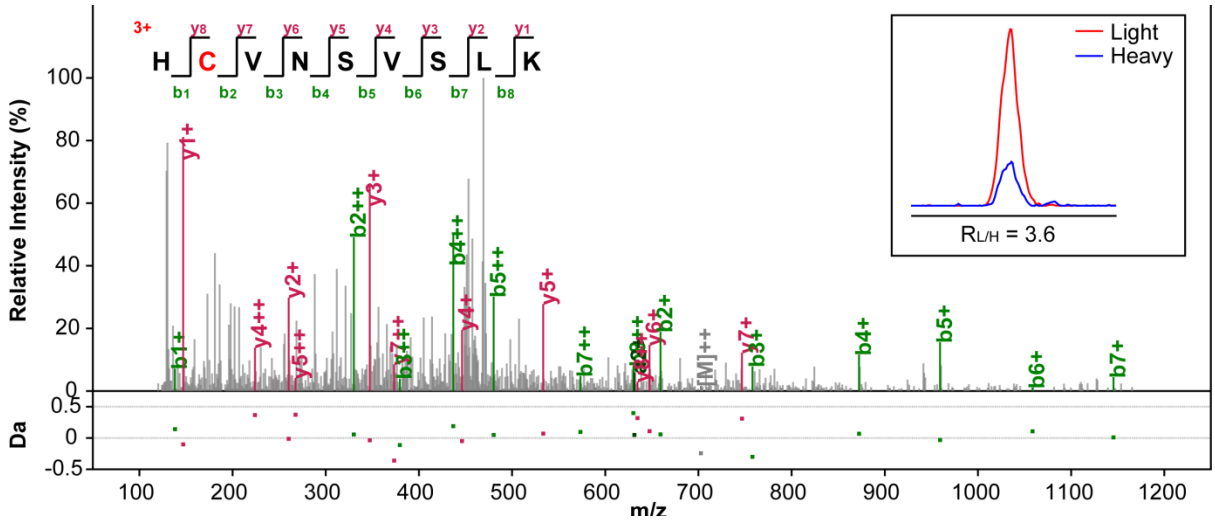

Fig. S3. Annotated spectra of the BTD-modified peptide 'HC$_{BTD}$VNSVSLK' matching the nucleophilic Cys187 of *At*MSRB3/7/8/9 targeting the active sites of METHIONINE SULFOXIDE REDUCTASE B (MsrB) proteins. Extracted-ion chromatograms show the light- (H_2O_2) and heavy-labeled (control) profiles in red and blue, respectively. The BTD-modified peptide. Spectrum and peak annotations were derived from pFind Studio 3.0 result annotation files (25).

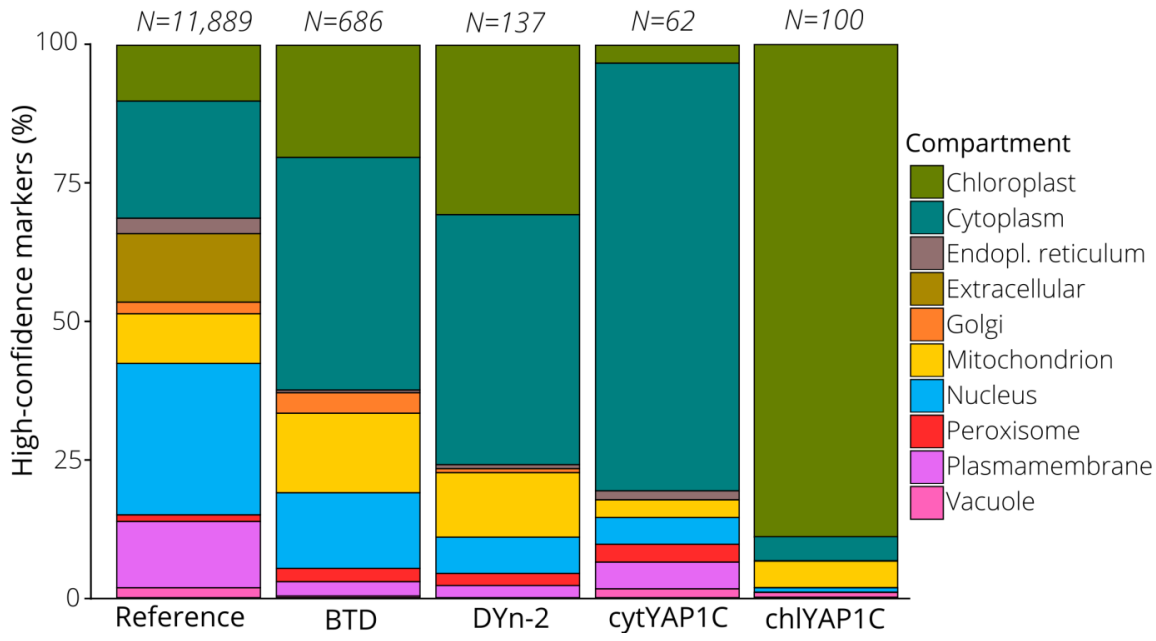

Fig. S4. Subcellular distributions of S-sulfenylated proteins in *Arabidopsis*. Distributions of assigned high-confidence markers (HCMs) were obtained by means of the SUBA4 Multiple Marker Abundance Profiling tool (14). The distribution of the assigned HCMs in the *Arabidopsis* proteome (left, 11,889 protein loci) was compared to the S-sulfenylated proteins reported (686 HCMs; *SI Dataset S1*) or previous studies using the DYn-2 probe (137 HCMs) (26) and YAP1C in the cytosol (cytYAPC1, 62 HCMs) (27), or targeted to the chloroplast (chlYAPC1, 100 HCMs) (28).

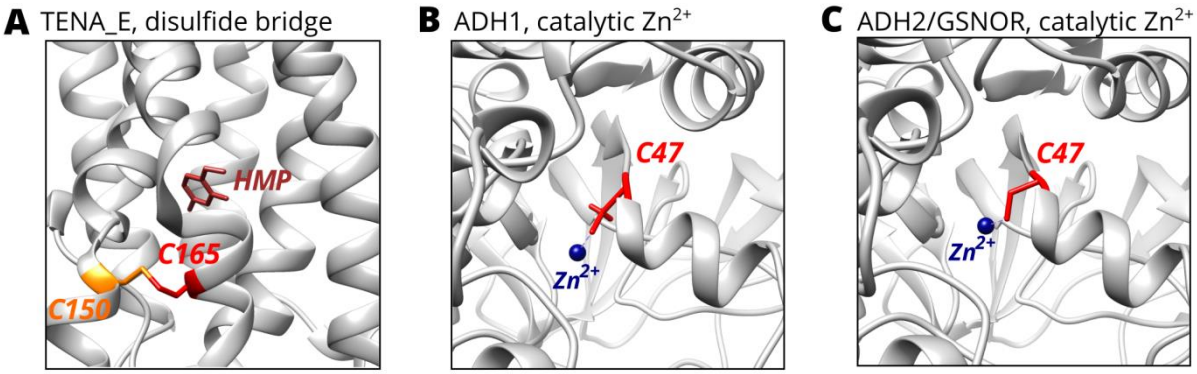

Fig. S5. Cys-SOHs matching annotations derived from PDB structures. (*A*) TENA_E protein structure (PDB 2f2g). The BTD-labeled Cys165 displayed in red forms a disulfide bridge with Cys150 (orange). The substrate 4-amino-5-hydroxymethyl-2-methylpyrimidine (HMP) is in brown. (*B*) ALCOHOL DEHYDROGENASE1 (ADH1) structure (PDB 4rqt). The BTD-labeled Cys47 is in red and the catalytic Zn^{2+} in blue. (*C*) ADH2 structure (PDB 3uko). The BTD-labeled Cys47 is in red and the catalytic Zn^{2+} in blue.

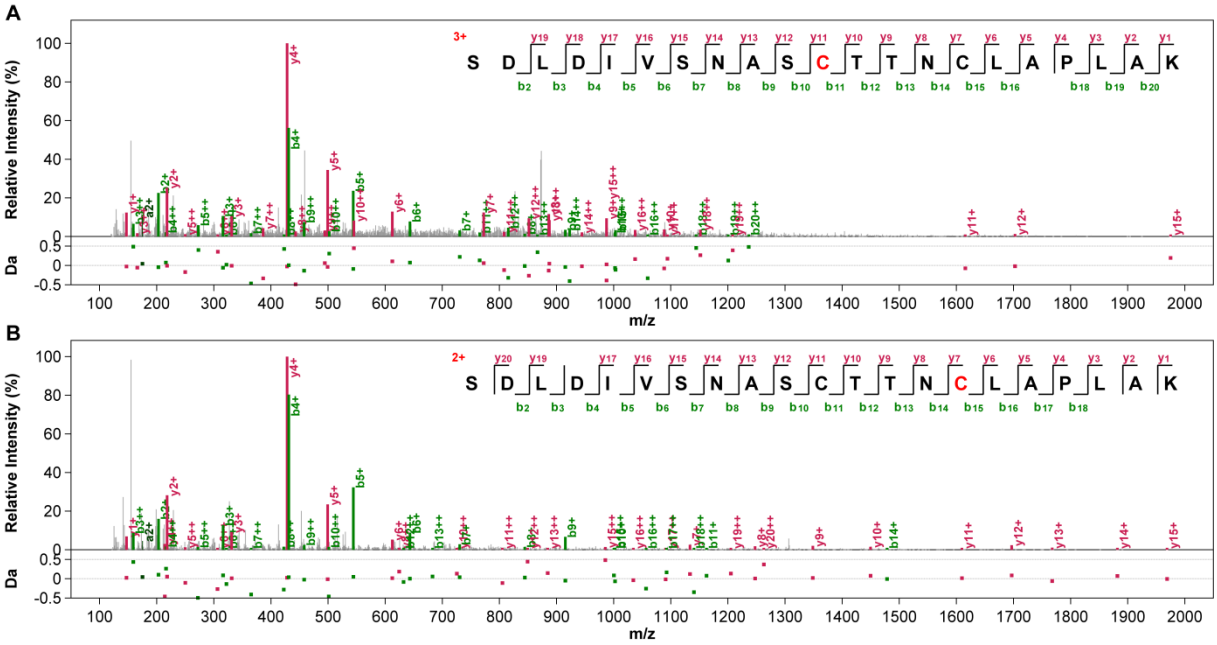

Fig. S6. The GAPDH-modified peptides 'SDLDIVSNASC$_{BTD}$TTNC$_{IAM}$LAPLAK' (*A*) and 'SDLDIVSNASC$_{IAM}$TTNC$_{BTD}$LAPLAK' (*B*), indicating oxidation of Cys156 and Cys160, respectively. Spectrum and peak annotation were derived from pFind Studio 3.0 result annotation files (25).

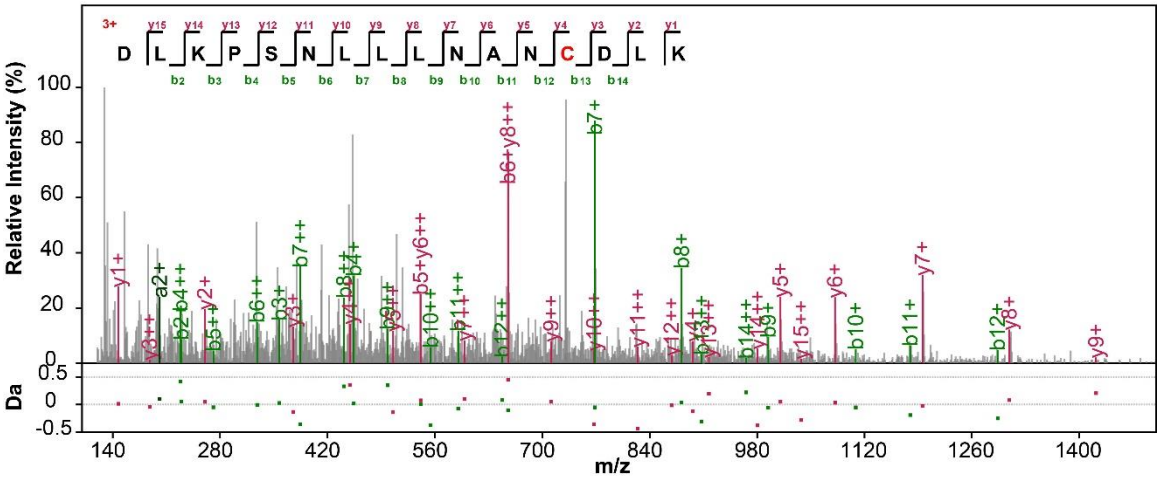

Fig. S7. The BTD-modified peptide 'DLKPSNLLLNANCBTDDLK' matching Cys181 of *At*MAPK4, in addition to matching ambiguously with *At*MAPK6, *At*MAPK7, and *At*MAPK11. Spectrum and peak annotation were derived from pFind Studio 3.0 result annotation files (25).

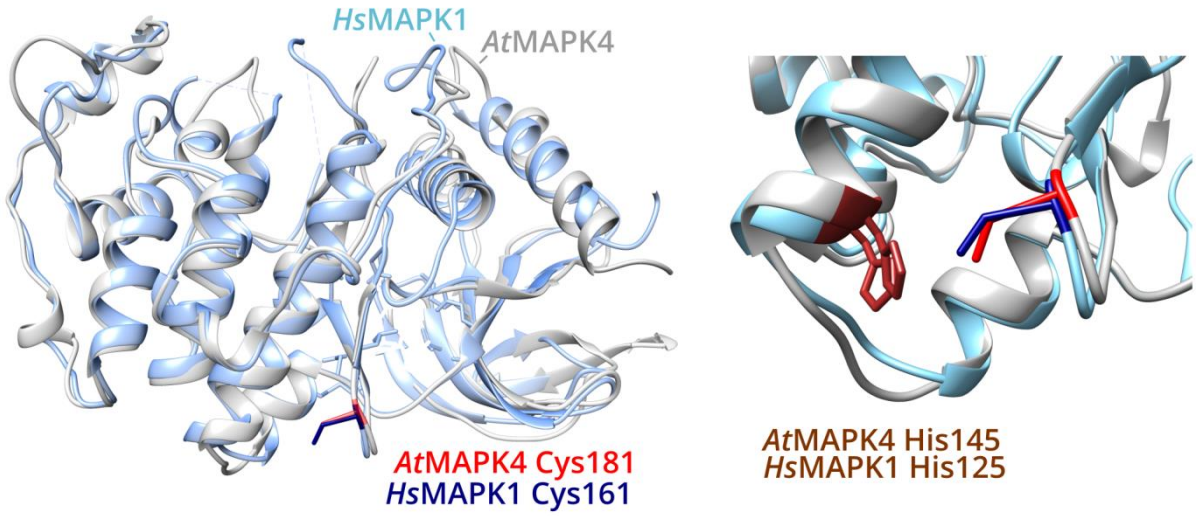

Fig. S8. Overlay of the model of *At*MAPK4 (blue) based on the structure of *At*MAPK6 (PDB identifier 5ci6) with the structure *Hs*MAPK1 (PDB identifier 2y9q) (grey). Structures were superimposed with the default MatchMaker function in Chimera version 1.3 (17). Overall view with the indication of the position of the conserved cysteine and histidine in both structures (*Left*). Detailed view of *At*MAPK4 Cys181 and *Hs*MAPK1 Cys161 with the respective conserved His145 and His125 (*Right*). Cys-His hydrogen bonding lowers the *pK*a of the thiol and makes it more susceptible to oxidation.

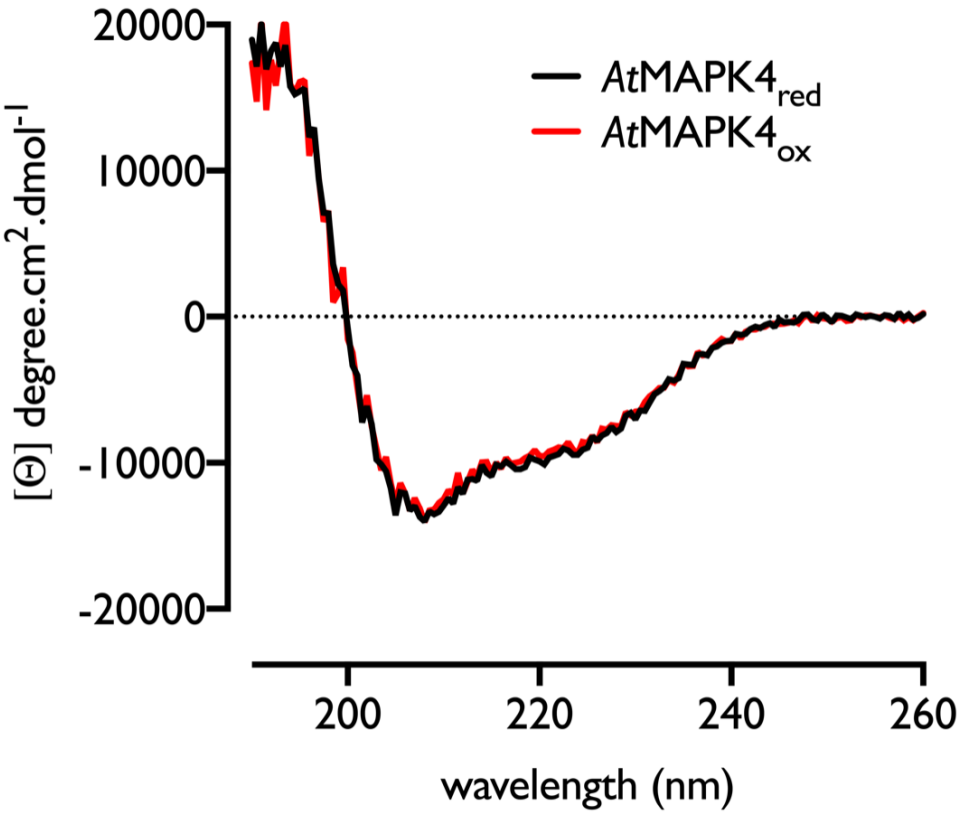

Fig. S9. Secondary structure analysis of reduced (*At*MAPK4$_{red}$) and oxidized *At*MAPK4 (*At*MAPK4$_{ox}$). Circular dichroism spectra are shown for reduced and oxidized *At*MAPK4. *At*MAPK4 recombinant proteins were reduced with 20 mM DTT for 1 h at 4°C (AtMAPK4$_{red}$). AtMAPK4$_{ox}$ was obtained after incubation of 1 mM H_2O_2 at room temperature for 1 h. Excess H_2O_2 and reducing agent were removed by a buffer exchange on a desalting column (Zeba spin column, Thermo Scientific). *At*MAPK4 recombinant proteins were diluted to a final concentration of 2 µM in a 50-mM NaH_2PO4, pH 7.5 at 25ºC, 50 mM NaF, and 1 mM DTT. Spectra were collected at 25ºC in a Jasco J-715 spectropolarimeter.

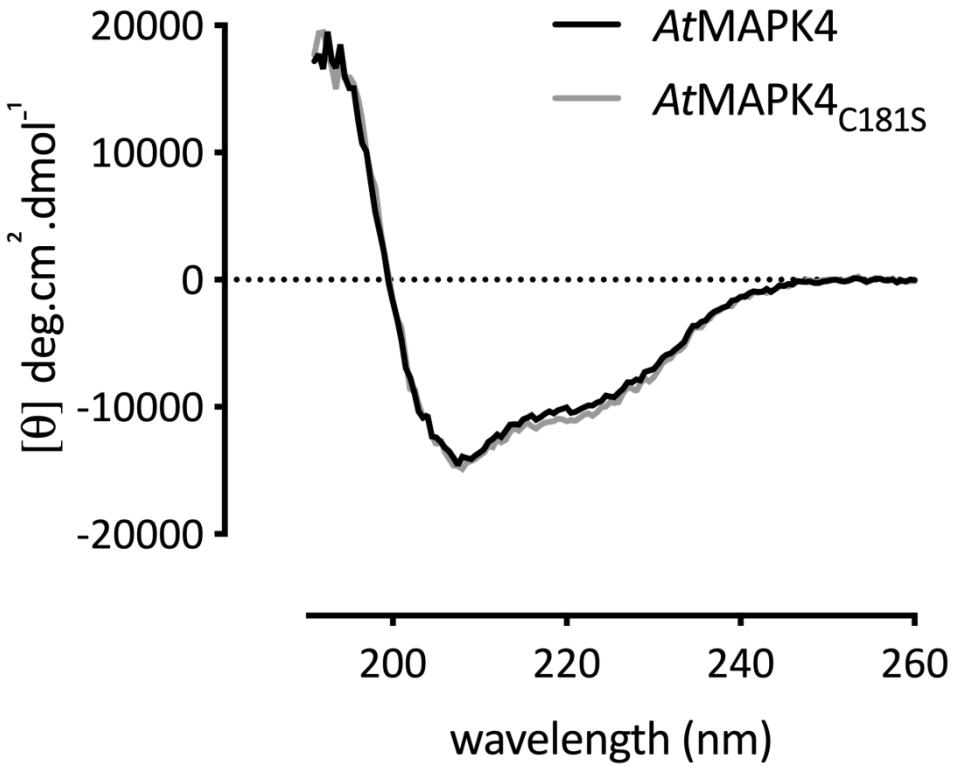

Fig. S10. Secondary structure analysis of *At*MAPK4 and *At*MAPK4$_{C181S}$. Circular dichroism spectra of *At*MAPK4 and *At*MAPK4$_{C181S}$ are shown. *At*MAPK4 recombinant proteins were reduced with 20 mM DTT for 1 h at 4°C. The reducing agent was removed by a buffer exchange on a desalting column (Zeba spin column, Thermo Scientific). AtMAPK4 recombinant proteins were diluted to a final concentration of 2 µM in a 50-mM NaH_2PO4, pH 7.5 at 25ºC, 50 mM NaF, and 1 mM DTT. Spectra were collected at 25ºC in a Jasco J-715 spectropolarimeter.

Table S1. Aminoacyl tRNA-ligases with identified Cys-SOHs. BTD-modified peptides were displayed with '#' indicating the BTD modification and their respective H_2O_2/control ratio. The number of replicates on which the ratio was measured is indicated between parentheses.

UniProt	Cys	Symbol	BTD-modified peptide	$R_{H_2O_2/Ctrl}$ (reps)
F4HWL4	435	TyrRS2	KPAPIC#YDGFEPSGR	1.35 (1)
F4I116	838	LeuRS	LSC#GTGGMHHDLLLK	1.01 (2)
F4I116	449	LeuRS	VC#LDLK	0.85 (1)
F4I4Z2	947	ALATS	AVVC#AGVPEK	1.155 (2)
F4IYF8	135	HisRS	SCASLIGIC#SIIDHK	1.64 (1)
F4JLM5	88	IleRS	FGWDC#HGLPVENEIDR	1.56 (1)
F4JLM5	621	IleRS	NLIC#NGLVLAEDGKK	1.12 (2)
F4JLM5	830	IleRS	GSEESVHYC#SIPPR	1.09 (2)
O04630	255	ThrRS	CGPLVDLC#R	H_2O_2
O04630	618	ThrRS	QAIVC#PISEK	1.03 (2)
O04630	461	ThrRS	FQQDDAHIFC#TTEQVK	-
O23247	620	EMB1027/ArgRS	LLLC#EATAIVMR	1.14 (1)
O23247	302	EMB1027/ArgRS	IC#DISR	0.78 (1)
O48593	215	SYNO/AsnRS	IIVVGEC#DSSYPIQK	H_2O_2
O82462	696	GluRS	GYFRC#DVPFVK	1 (2)
O82462	520	GluRS	IIDPVC#PR	0.74 (1)
P93736	433	TWN2	ITPAHDPNDC#EVGK	2.13 (1)
P93736	496	TWN2	LGLC#SR	-
Q39230	405	SerRS	TIC#CILENYQR	1.1 (2)
Q39230	267	SerRS	YAGYSSC#FR	0.99 (2)
Q8H104	117	AspRS	ESGFTVQC#VVEETR	1.53 (3)
Q8H104	170	AspRS	MYC#LSR	1.19 (1)
Q8RWT8	473	OVA7/SerRS	FVHTLNATAC#AVPR	1.23 (1)
Q8S9J2	101	TyrRS1	MTSAGC#R	0.98 (2)
Q8S9J2	53	TyrRS1	SIGEEC#IQEEELK	-
Q8W4F3	51	OVA9/GlnRS	VTANLTAVIHEAAVTDGC#DR	1.79 (1)
Q94K73	160	PheRSchl,mit	C#HTSAHQAELLR	1 (3)
Q9C713	553	ArgRS	FYSNC#QVNGSAEETSR	H_2O_2
Q9C713	631	ArgRS	LLLC#EATAIVMR	1.14 (1)
Q9C713	250	ArgRS	IC#EISR	1.06 (1)
Q9FFC7	758	EMB86/AlaRS	VVEVPGVSMELC#GGTHVGNTAEIR	H_2O_2
Q9M084	140	IBI1/AspRS2	ESGSTVQC#VVSQSEK	H_2O_2
Q9M084	302	IBI1/AspRS2	GQPAC#LAQSPQLHK	0.89 (3)
Q9M1R2	276	ProRS	GVQGATSHC#LGQNFAK	0.93 (3)
Q9M2T9	137	MetRSchl,mit	NPPEHC#DLISQSYR	1.13 (2)
Q9M2T9	437	MetRSchl,mit	KNC#ESTLVVDSTVAAEGVPLK	0.93 (1)
Q9SGE9	152	PheRScyt	LHQNIC#R	1.14 (1)
Q9SSK1	101	SYNC3/AsnRS3	LVATGTC#VTVDGVLK	1.09 (1)
Q9SSK1	538	SYNC3/AsnRS3	HC#GFGLGFER	0.93 (1)
Q9SVN5	497	MetRScyt	EDKPLC#AVVIR	1.52 (3)
Q9ZPI1	530	LysRS	HELC#NAYTELNDPVVQR	1.75 (3)
Q9ZPI1	468	LysRS	FDVKC#PPPQTTAR	0.93 (2)
Q9ZPI1	461	LysRS	YLIDAC#AR	0.88 (1)
F4HWL4	435	TyrRS2	KPAPIC#YDGFEPSGR	1.35 (1)

Table S2. Cys-SOHs with matching positional UniProtKB protein annotation. The UniProtKB (41) protein identifier is given alongside the Araport11 description. The table is subdivided in three evidence types: (i) cysteine activity studied via mutagenesis studies, (ii) cysteine interactions derived from protein structure, and (iii) cysteines electronically annotated via sequence analysis or similarity in UniProtKB, often via PROSITE-ProRule (12), High-quality Automated and Manual Annotation of Proteins (HAMAP) signatures (42), and the Protein Information Resource SuperFamily (PIRSF) signatures (43).

Protein	Symbol	Cys	Site Description	Reference/Source
Mutagenesis studies				
O64517	MC4	139	C→A: Loss of protease activity.	29
P16127	CHLI1	193	C→S: Reduces ATPase activity 2-fold	30
		354	C→S: Reduces ATPase activity 5-fold	
P25858, Q9FX54	GAPC1, GAPC2	155	C→ S: Loss of activity and *S*-GSH.	31
P34791	CYP20-3	131	C→S: Reduced PPIase activity, lower sensitivity to redox regulation	32
P46011	NIT4	197	C→A: Loss of nitrilase activity	33
Q8GY23	UPL1	3648	C→S or A: Loss of conjugation ability	34
Q8H7F6	GRXS16	123	C→S: Increased glutaredoxin activity.	35
Q9FEB5	SEX4	198	C→S: Loss of glucan phosphatase	36
Q9SRK5	LSF2	193	C→S: Abolishes phosphatase activity	37
Structural studies, PDB protein structure				
P06525	ADH1	47	Catalytic Zinc	38; PDB 4rqt
Q96533	ADH2	47	Catalytic Zinc	PDB 3uko
Q9ASY9	TENA_E	165	Disulfide	39; PDB 2f2g
Electronic sequence annotation				
Q9C5C8	MsrB2	187	Nucleophilic Cys	PRU01126, ProRule
Q9M0Z6,Q8VY86, O49707,Q84JT6	MsrB3, MsrB7, MsrB8,MsrB9	161, 129, 128, 129		
Q94A65	AtSTR14	166	Cysteine persulfide intermediate	PRU00173, ProRule
Q8S904	MFDX2	165	2Fe-2S ferredoxin-type domain	PRU00465, ProRule
O03042	RBCL	247	Interchain disulfide	MF_01338, HAMAP rule
Q9FFC7	EMB86/AlaRS	758	Zinc-binding	MF_03134, HAMAP rule
Q9SJT1, F4IHI1	SAE2	156	Zinc-binding	PIRSF039133, PIR superfamily
Q8LGG0	FKBP12	26	*Vicia faba* mutagenesis study, intramolecular disulphide required for complex formation	40
Q9CAI3	CAD1	101	Structural Zinc, similarity to CAD5	PDB: 2cf
Q9FGI6	EMB1467	117	Iron-sulfur 1 (2Fe-2S)	By similarity
Q9FJ95	SDH	54	Catalytic Zinc	By similarity
P47924	RIBA1	384	Catalytic Zinc	By similarity
Q94AR8	IIL1	376	Iron-sulfur (4Fe-4S)	By similarity
Q9SV64	FATA2	300	Active site	By similarity

Table S3. Identified phosphorylation sites and modified cysteine sites in recombinant *At*MAPK4 treated with 0, 1, or 10 mM H_2O_2 (see *SI Material and Methods*). For the assigned cysteine modification and phosphorylation sites, the number of peptide-to-spectrum matches (PSMs) of trypsin and chymotrypsin-digested samples were summed per H_2O_2 treatment. PSMs larger than 2, 5, and 10 are highlighted in grey, brown, and orange, respectively. For a full overview, see *SI Dataset S3*.

***At*MAPK4 Site**	**Modification**	**Peptide-to-spectrum matches**		
		0 mM H_2O_2	**1 mM H_2O_2**	**10 mM H_2O_2**
Cysteine modifications				
Cys6	Cys-SH	14	5	-
	Cys-SOH	2	-	-
	SO_2H	-	1	1
Cys58	Cys-SH	43	40	9
Cys146	Cys-SH	75	54	34
Cys181	Cys-SH	60	36	5
	Cys-SOH	2	3	2
	SO_2H	2	-	1
Cys218	Cys-SH	3	4	-
Cys232	Cys-SH	7	6	3
	Cys-SOH	-	4	-
	SO_2H	-	-	2
Cys341	Cys-SH	84	70	8
	Cys-SOH	-	4	3
	SO_2H	-	-	2
Phosphorylation sites				
Ser2		3	2	-
Thr21		1	-	-
Thr61		-	-	2
Thr65		-	-	2
Tyr159		8	-	2
Ser173		1	1	2
Thr193		1	-	1
Ser195		1	-	-
Thr197		1	-	-
Tyr203		33	10	6
Thr222		-	1	-
Ser229		1	2	-
Thr237		1	-	1
Thr239		1	-	2
Ser267		1	1	-
Thr355		-	1	-
Ser375		1	1	-

Dataset S1
Detailed overview of 1,537 identified Cys-SOHs in this study with matching meta-data.

Dataset S2
Gene set enrichment analysis.

Dataset S3
Peptide identification results of reduced and oxidized recombinant *At*MAPK4.

SI References

1. Lin D, Saleh S, Liebler DC. (2008) Reversibility of covalent electrophile-protein adducts and chemical toxicity. *Chem Res Toxicol* 21:2361-2369.
2. Yang J, Tallman KA, Porter NA, Liebler DC. (2015) Quantitative chemoproteomics for site-specific analysis of protein alkylation by 4-hydroxy-2-nonenal in cells. *Anal Chem* 87:2535-2541.
3. Yuan Z-F, et al. (2012) pParse: a method for accurate determination of monoisotopic peaks in high-resolution mass spectra. *Proteomics* 12:226-235.
4. Wang L-h, et al. (2007) pFind 2.0: a software package for peptide and protein identification via tandem mass spectrometry. *Rapid Commun Mass Spectrom* 21:2985-2991.
5. Liu C, et al. (2014) pQuant improves quantitation by keeping out interfering signals and evaluating the accuracy of calculated ratios. *Anal Chem* 86:5286-5294.
6. Akter S, et al. (2018) Chemical proteomics reveals new targets of cysteine sulfinic acid reductase. *Nat Chem Biol* 14:995-1004.
7. Kanehisa M, Goto S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28:27-30.
8. Gribskov M, et al. (2001) PlantsP: a functional genomics database for plant phosphorylation. *Nucleic Acids Res* 29:111-113.
9. Jin J, et al. (2017) PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res* 45:D1040-D1045.
10. Lallemand J, et al. (2015) Extracellular peptidase hunting for improvement of protein production in plant cells and roots. *Front Plant Sci* 6:37.
11. Van Ruyskensvelde V, Van Breusegem F, Van Der Kelen K. (2018) Post-transcriptional regulation of the oxidative stress response in plants. *Free Radic Biol Med* 122:181-192.
12. Sigrist CJA, et al. (2005) ProRule: a new database containing functional and structural information on PROSITE profiles. *Bioinformatics* 21:4060-4066.
13. Smedley D, et al. (2015) The BioMart community portal: an innovative alternative to large, centralized data repositories.*Nucleic Acids Res* 43:W589-W598.
14. Hooper CM, et al. (2017) Multiple marker abundance profiling: combining selected reaction monitoring and data-dependent acquisition for rapid estimation of organelle abundance in subcellular samples. *Plant J* 92:1202-1217.
15. Hooper CM, Castleden IR, Tanz SK, Aryamanesh N, Millar AH. (2017) SUBA4: the interactive data analysis centre for Arabidopsis subcellular protein locations. *Nucleic Acids Res* 45:D1064-D1074.
16. Schwede T, Kopp J, Guex N, Peitsch MC. (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res* 31:3381-3385.
17. Pettersen EF, et al. (2004) UCSF Chimera--A visualization system for exploratory research and analysis. *J Comput Chem* 25:1605-1612.
18. Gupta V, Yang J, Liebler DC, Carroll KS (2017) Diverse redoxome reactivity profiles of carbon nucleophiles. *J Am Chem Soc* 139:5588-5595.
19. Colaert N, Helsens K, Martens L, Vandekerckhove J, Gevaert K (2009) Improved visualization of protein consensus sequences by iceLogo. *Nat Methods* 6:786-787.
20. Yang J, Gupta V, Carroll KS, Liebler DC (2014) Site-specific mapping and quantification of protein *S*-sulphenylation in cells. *Nat Commun* 5: 4776.
21. Huerta-Cepas J, et al. (2016) eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* 44:D286-D293.

22. Pyr Dit Ruys S, et al. (2012) Identification of autophosphorylation sites in eukaryotic elongation factor-2 kinase. *Biochem J* 442:681-692.
23. Perez-Riverol Y, et al. (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* 47:D442-D450.
24. Leissing F, et al. (2016) Substrate thiophosphorylation by Arabidopsis mitogen-activated protein kinases. *BMC Plant Biol* 16:48.
25. Chi H, et al. (2018) Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. *Nat Biotechnol* 36:1059-1061.
26. Akter S, et al. (2015) DYn-2 based identification of *Arabidopsis* sulfenomes. *Mol Cell Proteomics* 14:1183-1200.
27. Waszczak C, et al. (2014) Sulfenome mining in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 111:11545-11550.
28. De Smet B, et al. (2019) *In vivo* detection of protein cysteine sulfenylation in plastids. *Plant J* 97:765-778.
29. Watanabe N, Lam E. (2011) Calcium-dependent activation and autolysis of *Arabidopsis* metacaspase 2d. *J Biol Chem* 286:10027-10040.
30. Ikegami A, et al. (2007) The CHLI1 subunit of *Arabidopsis thaliana* magnesium chelatase is a target protein of the chloroplast thioredoxin. *J Biol Chem* 282:19282-19291.
31. Bedhomme M, et al. (2012) Glutathionylation of cytosolic glyceraldehyde-3-phosphate dehydrogenase from the model plant *Arabidopsis thaliana* is reversed by both glutaredoxins and thioredoxins *in vitro*. *Biochem J* 445:337-347.
32. Motohashi K, Koyama F, Nakanishi Y, Ueoka-Nakanishi H, Hisabori T. (2003) Chloroplast cyclophilin is a target protein of thioredoxin. Thiol modulation of the peptidyl-prolyl *cis-trans* isomerase activity. *J Biol Chem* 278:31848-31852.
33. Piotrowski M, Schönfelder S, Weiler EW. (2001) The *Arabidopsis thaliana* isogene *NIT4* and its orthologs in tobacco encode *β*-cyano-L-alanine hydratase/nitrilase. *J Biol Chem* 276:2616-2621.
34. Bates PW, Vierstra RD. (1999) UPL1 and 2, two 405 kDa ubiquitin--protein ligases from *Arabidopsis thaliana* related to the HECT-domain protein family. *Plant J* 20:183-195.
35. Liu X, et al. (2013) Structural insights into the N-terminal GIY-YIG endonuclease activity of *Arabidopsis* glutaredoxin AtGRXS16 in chloroplasts. *Proc Natl Acad Sci USA* 110:9565-9570.
36. Sokolov LN, Dominguez-Solis JR, Allary A-L, Buchanan BB, Luan S. (2006) A redox-regulated chloroplast protein phosphatase binds to starch diurnally and functions in its accumulation. *Proc Natl Acad Sci USA* 103:9732-9737.
37. Meekins DA, et al. (2015) Mechanistic insights into glucan phosphatase activity against polyglucan substrates. *J Biol Chem* 290:23361-23370.
38. Chen F, Wang P, An Y, Huang J, Xu Y. (2015) Structural insight into the conformational change of alcohol dehydrogenase from *Arabidopsis thaliana* L. during coenzyme binding. *Biochimie* 108:33-39.
39. Blommel PG, et al. (2004) Crystal structure of gene locus At3g16990 from *Arabidopsis thaliana*. *Proteins* 57:221-222.
40. Xu Q, Liang S, Kudla J, Luan S (1998) Molecular characterization of a plant FKBP12 that does not mediate action of FK506 and rapamycin. *Plant J* 15:511-519.
41. UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 47: D506-D515.
42. Pedruzzi I, et al. (2015) HAMAP in 2015: updates to the protein family classification and annotation system. *Nucleic Acids Res* 43:D1064-1070.

43. Wu CH, et al. (2004) PIRSF: family classification system at the Protein Information Resource. *Nucleic Acids Res*. 32:D112-D114.