Supplementary Information for

Structures of the *Arabidopsis thaliana* Oxygen-Sensing Plant Cysteine Oxidases PCO4 and PCO5 enable targeted manipulation of their activity

Mark D. White, Laura Dalle Carbonare, Mikel Lavilla Puerta, Sergio Iacopino, Martin Edwards, Kate Dunne, Elisabete Pires, Colin Levy, Michael A. McDonough, Francesco Licausi and Emily Flashman.

Corresponding Author: emily.flashman@chem.ox.ac.uk, (tel) +44 1865 275920, Chemistry Research Laboratory, 12 Mansfield Road, Oxford, OX1 3TA, United Kingdom.

This PDF file includes:

Supplementary Figures:

Supplementary Tables:

Supplementary References

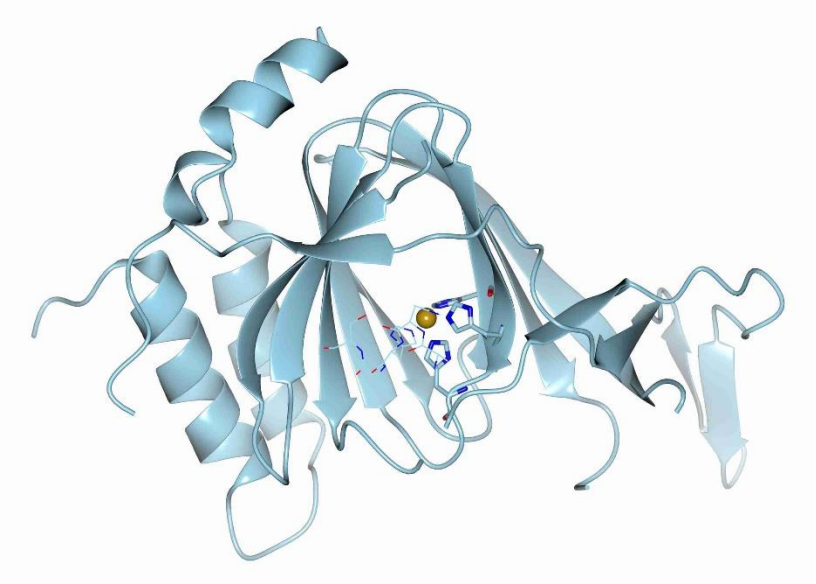


Figure S1. View of the Crystal Structure of AtPCO4_2. The overall structure is very similar to that of AtPCO4_1 but in AtPCO4_2 an N-terminal fragment from an adjacent His_6 tag (lines) occupies the active site (see also Supplementary Figure S5). AtPCO4_2 structure is shown in cartoon form, with the active site metal (gold sphere) coordinated by a facial histidine triad (sticks).

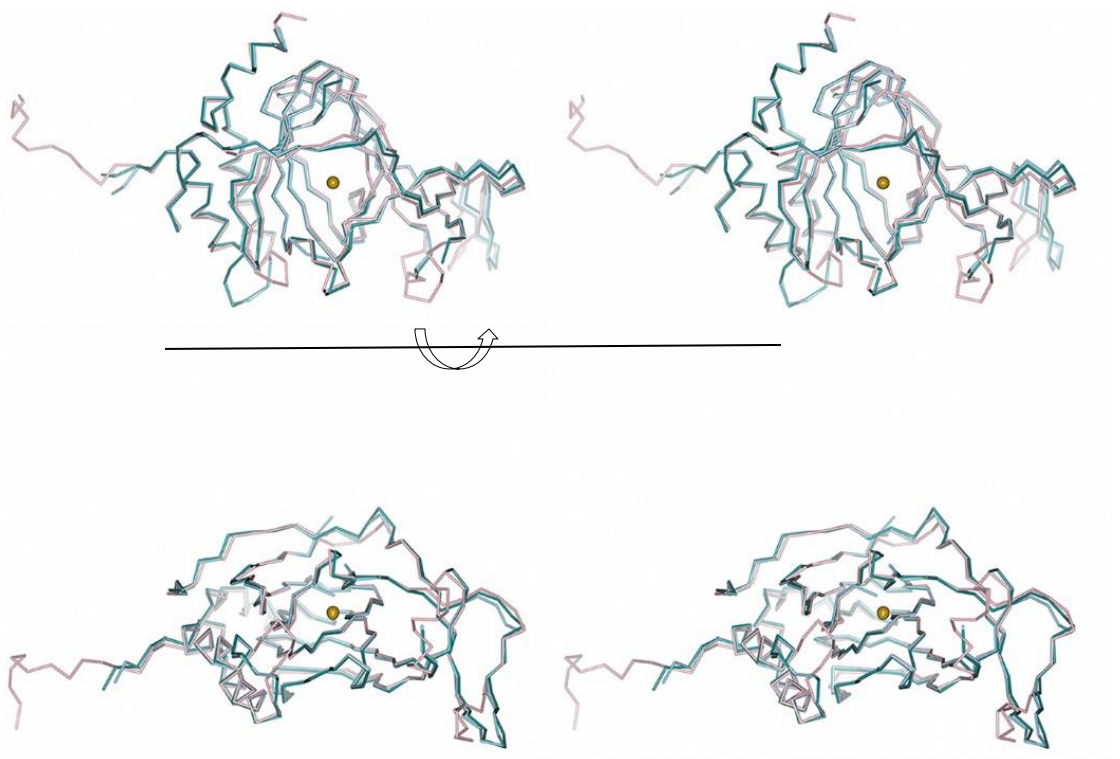


Figure S2. Stereoview of superimposed backbone structures of AtPCO4_1 (teal), AtPCO4_2 (sky blue) and AtPCO5 (light pink). Upper and lower panels are rotated by 90°. Alignment of the assigned structures demonstrates a high degree of conservation of secondary and tertiary structure between the AtPCO homologues.

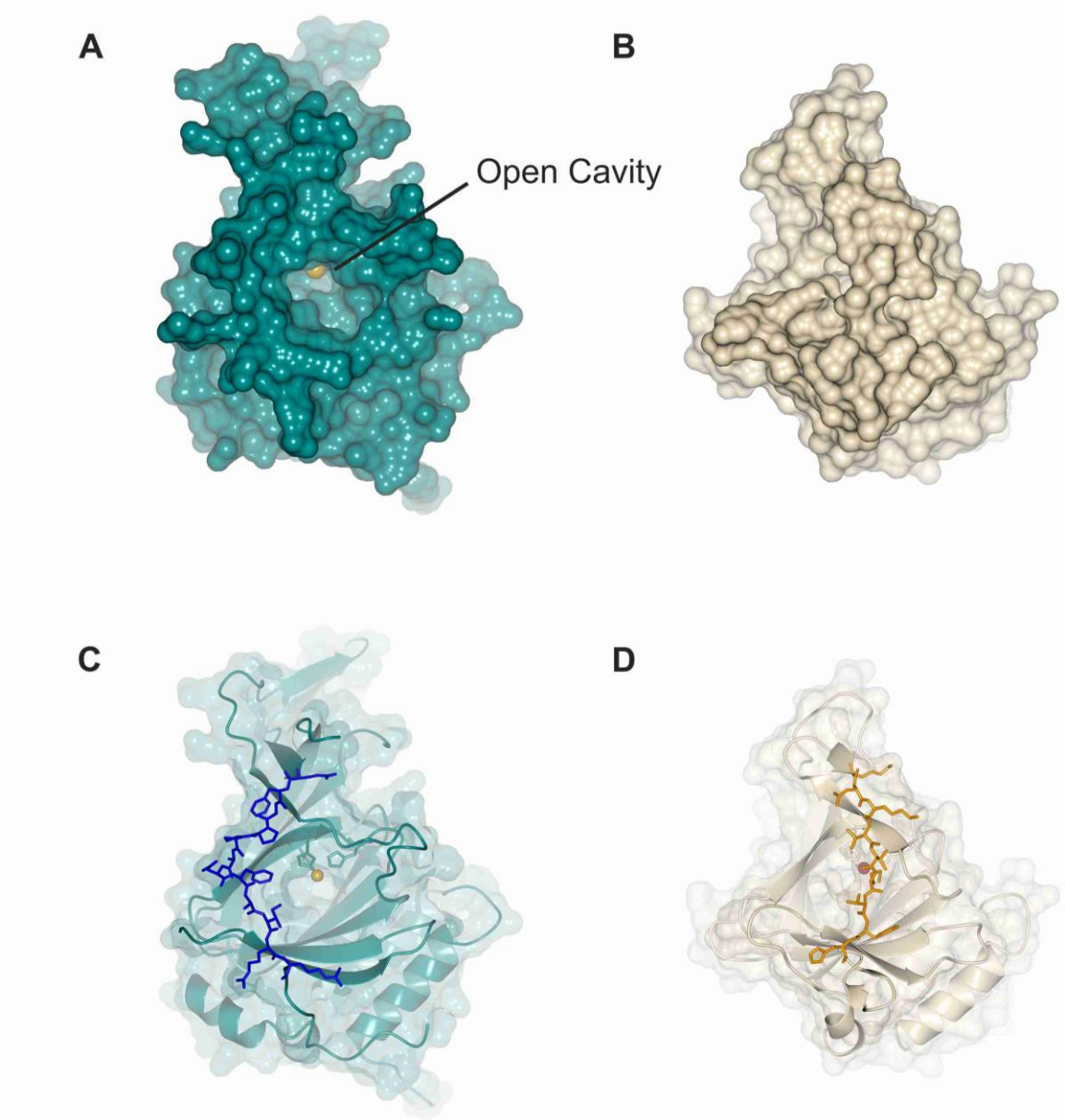


Figure S3. Comparison of active site entrance of AtPCO4_1 (teal) and RnCDO (wheat, PDB:4IEZ). (**A**) and (**B**) show surface representations of the two enzymes, revealing that AtPCO4_1 active site consists of a relatively large cavity. (**C**) and (**D**) show that this cavity exists due to a differing arrangement of equivalent chains (bold sticks) at the top of the double stranded beta helix structure in the two enzymes. This open structure in AtPCO4_1 (similar in AtPCO4_2 and AtPCO5, see Supplementary Figure S2) may facilitate the binding of larger, polypeptide based substrates to the AtPCOs. Views were aligned using Cα backbone structural alignments.

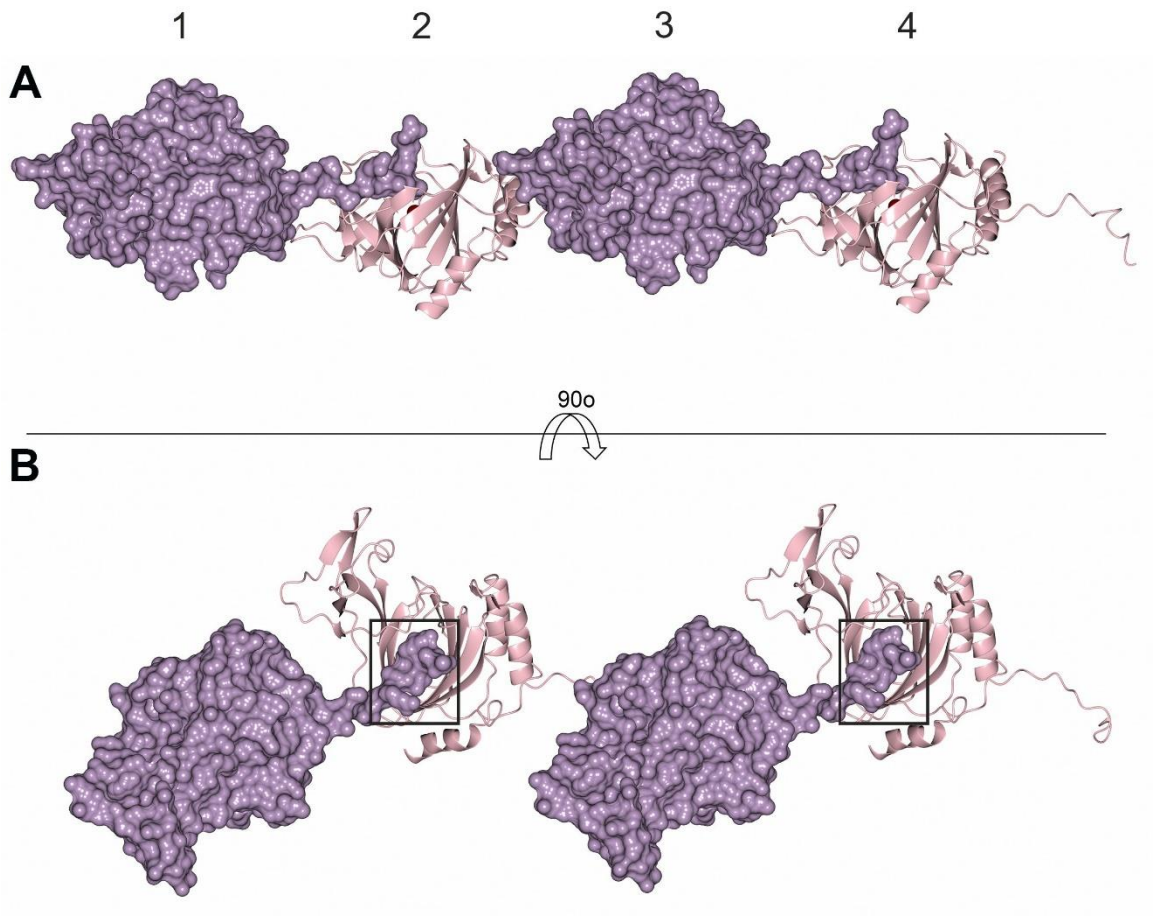


Figure S4. AtPCO5 crystal packing is facilitated by an interaction between the N-terminal His_6 tag and the cavity at the active site entrance of adjacent molecules. A side on (**A**) and aerial (**B**) view of four adjacent, symmetry related AtPCO5 molecules (1-4) packed in the crystal lattice, represented as alternating space filling (purple, 1 and 3) and ribbon (pink, 2 and 4) diagrams. The uncleaved tag of one unit overlaps the active site entrance, occupying the cavity (that is, the top of the central beta barrel motif) of an adjacent unit. The uncleaved N-terminal tag of molecules 1 and 3 is highlighted in panel B with a black box.

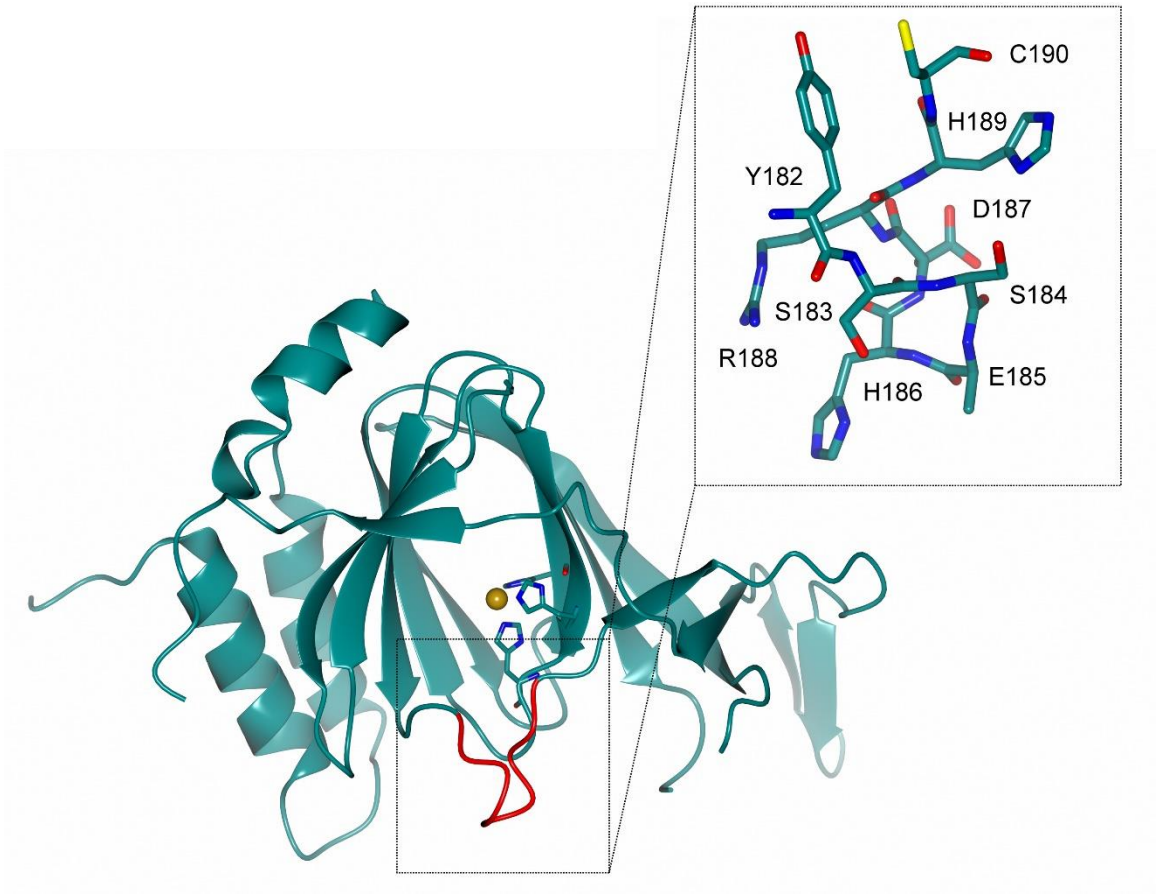


Figure S5. View of AtPCO4_1 crystal structure highlighting a loop at the entrance to the active site. The presence of charged and polar residues in this loop suggests a role in substrate binding and/or recognition.

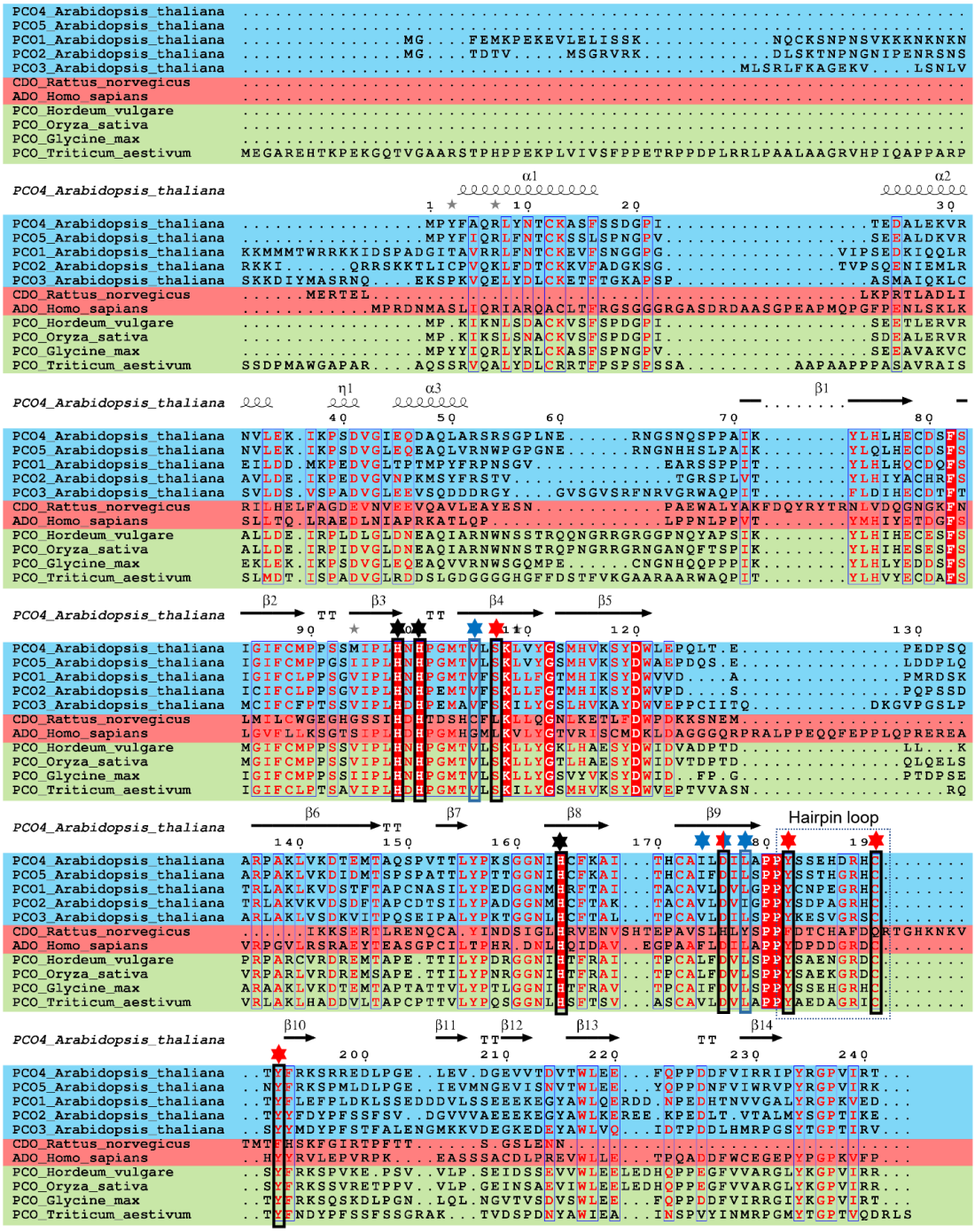


Figure S6. Sequence alignment comparing the amino acid composition of AtPCOs, RnCDO, HsADO and PCO sequences from barley, rice, soybean and wheat. The amino acid compositions of AtPCOs 1-5 (blue background), RnCDO and HsADO (both red background), and representative PCO sequences from various crop species (green background) were aligned using Clustal Omega (1). PCO sequences from *Hordeum vulgare, Oryza sativa, Glycine max and Triticum aestivum* were identified using a BlastP search with AtPCO4 as the input sequence. Secondary structure annotation was completed using the coordinates of AtPCO4_1 (PDB ID 6S7E)

with ESPript (2). Black boxes identify the position of conserved PCO residues discussed in the text, which are divided into metal binding or other active site residues by black and red asterisks, respectively. Blue boxes and asterisks identify the position of CDO residues contributing to the cross link and catalytic triad, respectively.

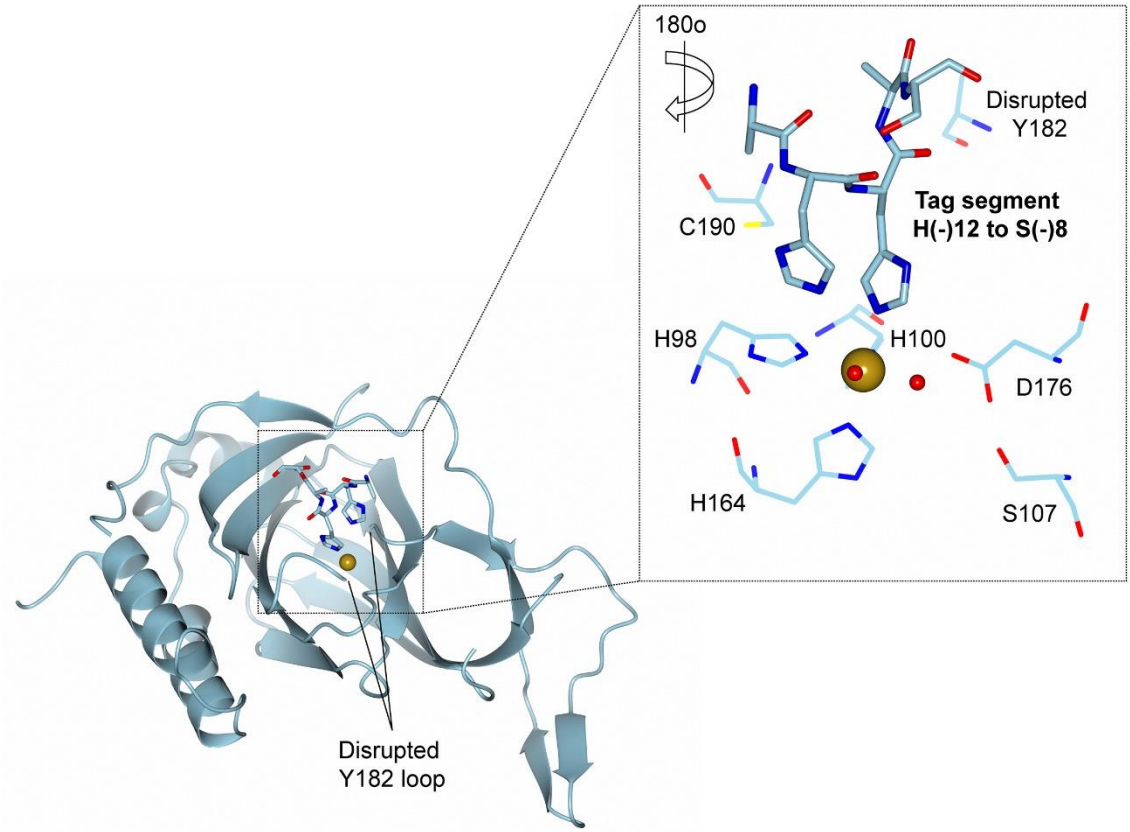
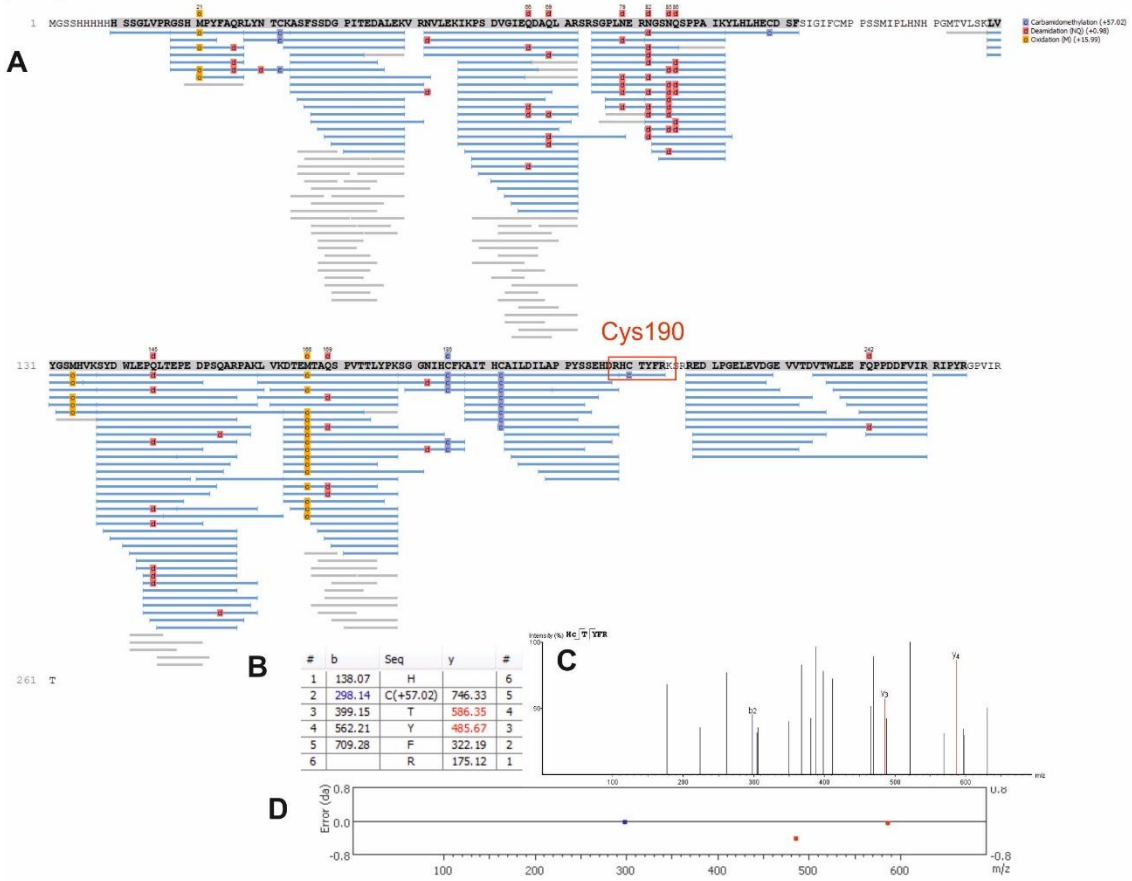


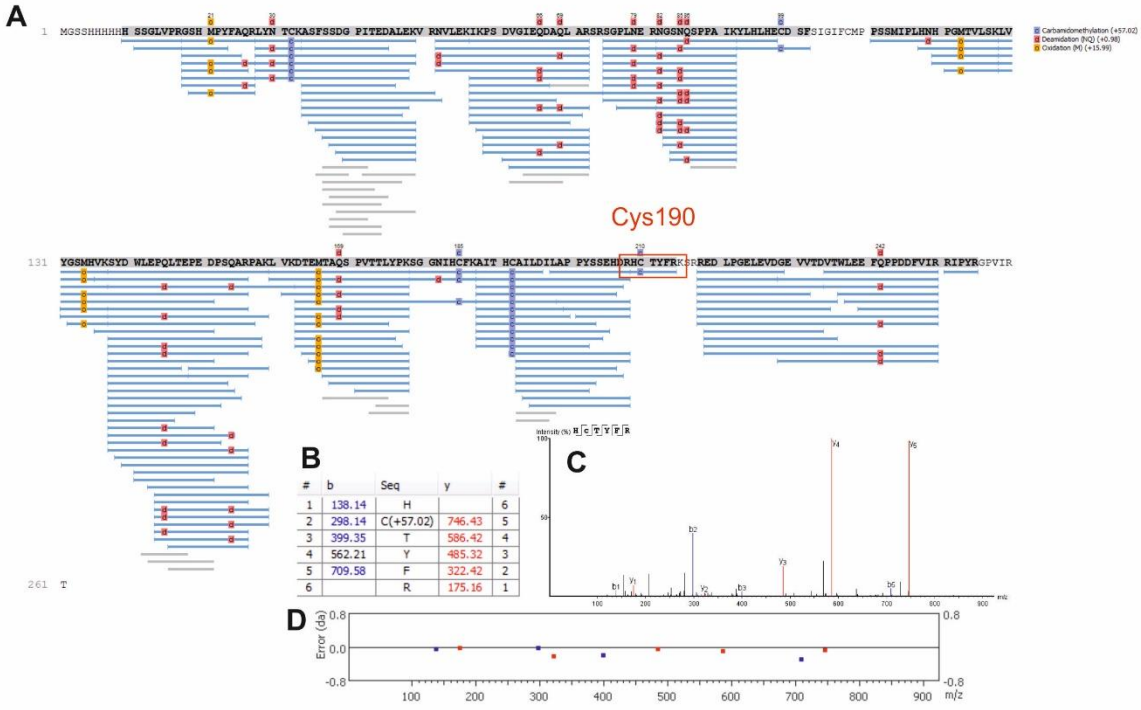
Figure S7. A His_6-tag fragment occupies the active site of AtPCO4_2 and disrupts loop Y182-C190. A His_6-tag fragment consisting of residues His(-)12 to Ser(-)8 (sticks and inset) occupies the active site of AtPCO4_2 with the τ-nitrogen of His(-)10 coordinating the iron cofactor and displacing the water molecule trans to His164. This binding appears to disrupt the Tyr182-Cys190 loop which has no clear electron density relative to AtPCO4_1 and AtPCO5 (Supplementary Figure S5).

His_6-AtPCO4



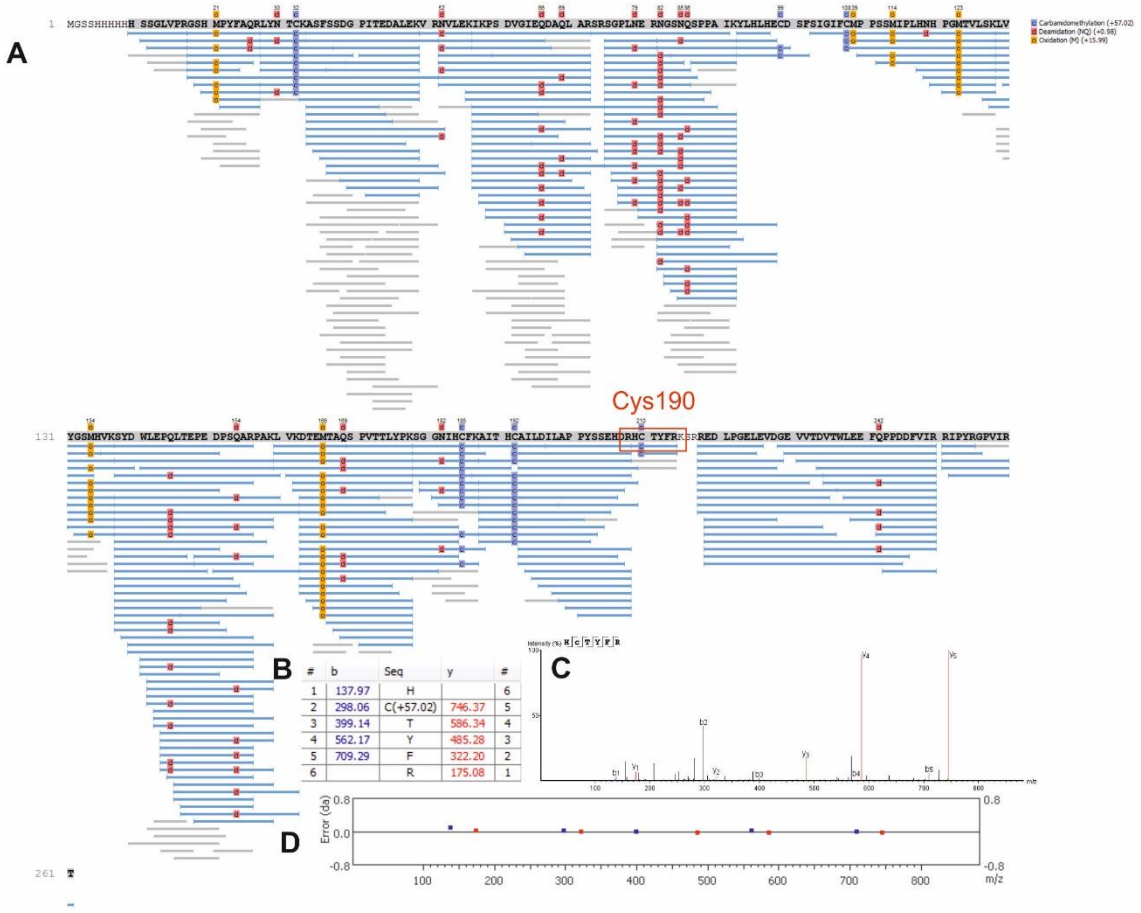
#	b	Seq	y	#
1	138.07	H		6
2	298.14	C(+57.02)	746.33	5
3	399.15	T	586.35	4
4	562.21	Y	485.67	3
5	709.28	F	322.19	2
6		R	175.12	1

His_6-AtPCO4 + $AtRAP2_{2-15}$



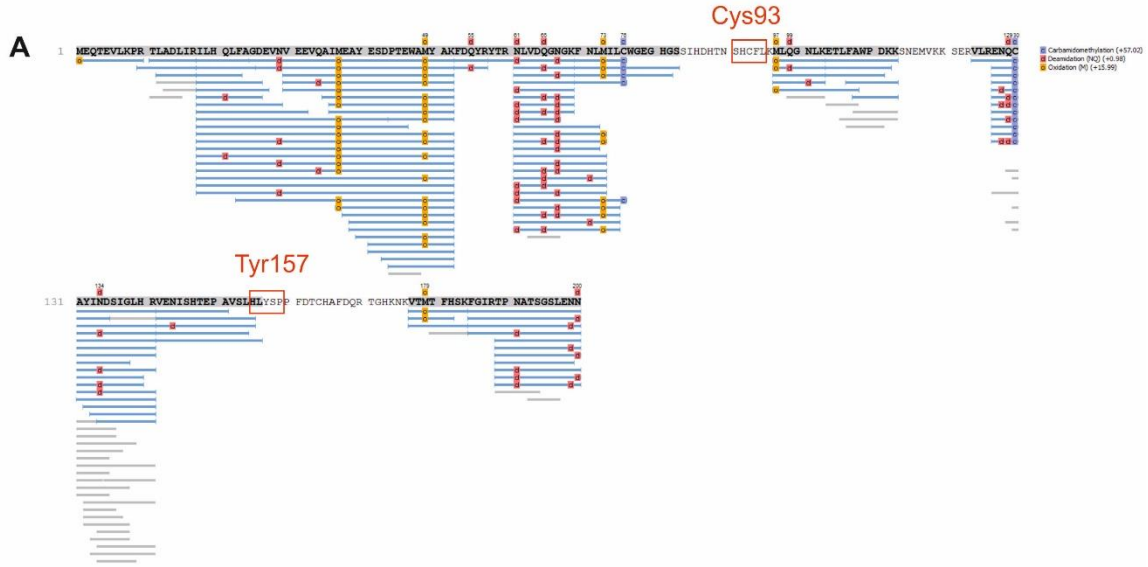
#	b	Seq	y	#
1	138.14	H		6
2	298.14	C(+57.02)	746.43	5
3	399.35	T	586.42	4
4	562.21	Y	485.32	3
5	709.58	F	322.42	2
6		R	175.16	1

His_6-AtPCO4 + Cysteine

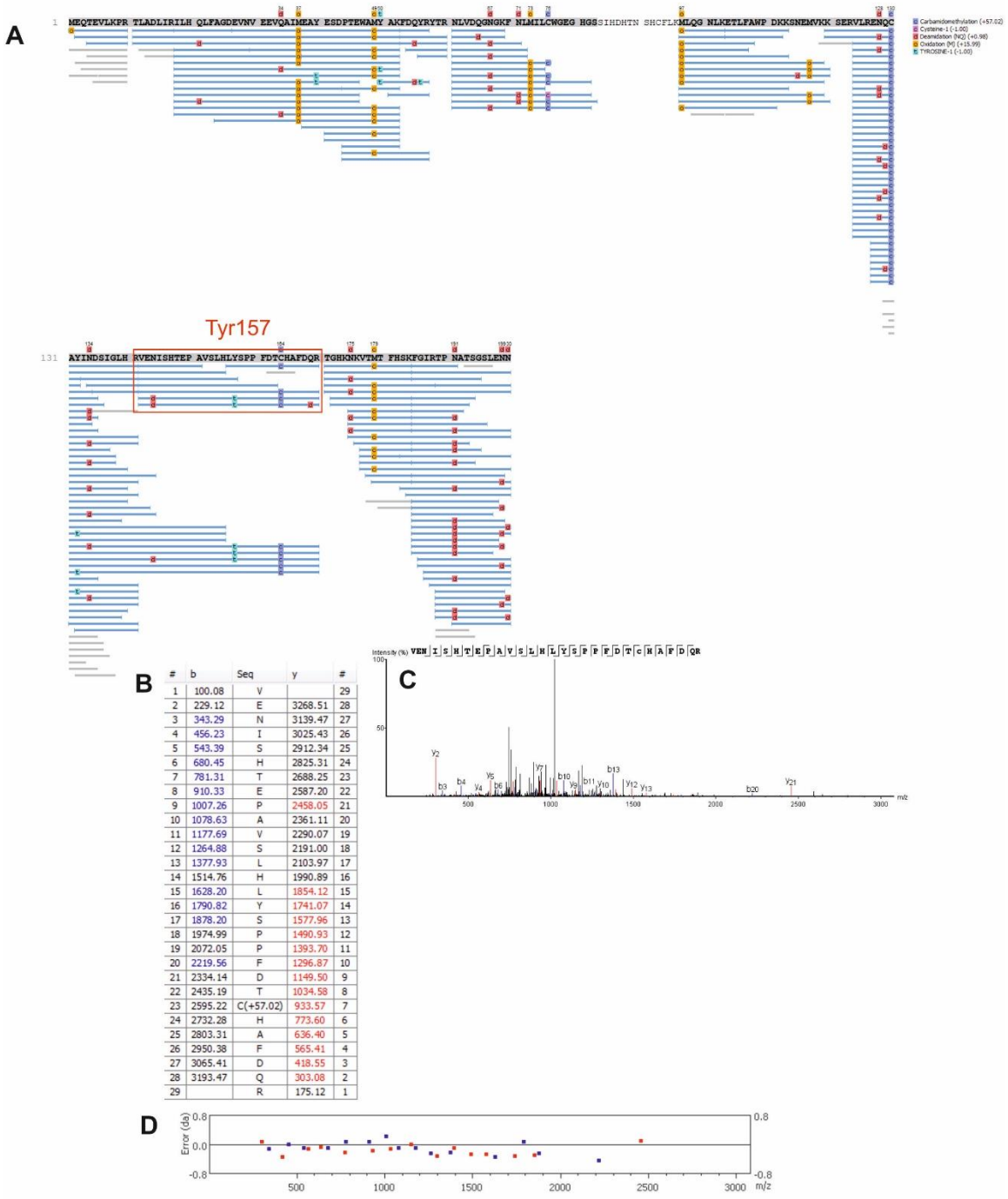


#	b	Seq	y	#
1	137.97	H		6
2	298.06	C(+57.02)	746.37	5
3	399.14	T	586.34	4
4	562.17	Y	485.28	3
5	709.29	F	322.20	2
6		R	175.08	1

HsCDO + Cysteine



HsCDO



#	b	Seq	y	#
1	100.08	V		29
2	229.12	E	3268.51	28
3	343.29	N	3139.47	27
4	456.23	I	3025.43	26
5	543.39	S	2912.34	25
6	680.45	H	2825.31	24
7	781.31	T	2688.25	23
8	910.33	E	2587.20	22
9	1007.26	P	2458.05	21
10	1078.63	A	2361.11	20
11	1177.69	V	2290.07	19
12	1264.88	S	2191.00	18
13	1377.93	L	2103.97	17
14	1514.76	H	1990.89	16
15	1628.20	L	1854.12	15
16	1790.82	Y	1741.07	14
17	1878.20	S	1577.96	13
18	1974.99	P	1490.93	12
19	2072.05	P	1393.70	11
20	2219.56	F	1296.87	10
21	2334.14	D	1149.50	9
22	2435.19	T	1034.58	8
23	2595.22	C(+57.02)	933.57	7
24	2732.28	H	773.60	6
25	2803.31	A	636.40	5
26	2950.38	F	565.41	4
27	3065.41	D	418.55	3
28	3193.47	Q	303.08	2
29		R	175.12	1

Figure S8. LC-MS/MS analyses of trypsin-digested AtPCO4 and HsCDO, with and without substrate, indicate Cys-Tyr cross-link formation does not occur in AtPCO4.

Results for each protein incubation are presented as: (**A**) Peptides observed following tryptic digest;peptide(s) containing amino acids of interest highlighted with a red box. (**B**) Table of theoretical masses of expected ions for fragmentation of peptide with observed *b*- and *y*-series ions

highlighted in blue and red, respectively. (**C**) LC-MS/MS spectrum of peptide of interest showing identified *b*- and *y*-series ions. (**D**) The "Error Map" shows the mass errors of matched fragment ions. The m/z ratio is displayed on the x-axis and the error is listed on the y-axis in Daltons. Each matched fragment ion is represented by a dot.

Cys190 is the only cysteine residue in the active site of AtPCO4/5 therefore searches for Cys-Tyr cross-link formation focused on identifying this residue in peptides derived from tryptic digest of the protein. Cross-link formation would prevent observation of the predicted linear peptide fragment. When AtPCO4 was incubated in the presence or absence of either free L-cysteine or its known peptidic substrate AtRAP2$_{2-15}$, peptides containing Cys190 were consistently observed, and indeed Cys190 was consistently carbamidomethylated (during the tryptic digest procedure) which would be limited in the event of Cys-Tyr cross-link formation. Cys-Tyr cross-link formation was therefore not detected in AtPCO4 under the conditions used.

A Cys-Tyr cross-link is known to occur between Cys-93 and Tyr-157 of human CDO (3) following multiple turnovers. Peptidic fragments containing HsCDO Cys93 and Tyr157 were not observed following incubation of the enzyme with L-cysteine, consistent with formation of a thio-ether cross-link. In contrast, when HsCDO was incubated in the absence of L-cysteine, a peptidic fragment containing Tyr157 was observable by LC-MS/MS, indicating that the thio-ether cross-link had not formed. This is consistent with observations reported in the literature (3-5) and supports formation of a Cys-Tyr cross-link in the active site of HsCDO under the conditions used.

Method: Proteins (AtPCO4 and HsCDO) were incubated +/- L-cysteine or +/- AtRAP2$_{2-15}$ peptide for 16 hours at 25 °C (50 mM Tris, 400 mM NaCl pH7.5, 10 mM Tris(2-carboxyethyl)phosphine hydrochloride). Proteins were visualized by SDS-PAGE and bands corresponding to HsCDO/AtPCO4 were excised and subjected to in-gel trypsin digest. Resulting tryptic peptides were analyzed on a NanoAcquity-UPLC system (Waters) connected to an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific) possessing an EASY-Spray nano-electrospray ion source (Thermo Fisher Scientific). The peptides were trapped on an in-house packed guard column (75 µm i.d. x 20 mm, Acclaim PepMap C18, 3µm, 100 Å) using solvent A (0.1% Formic Acid in water) at a pressure of 140 bar. The peptides were separated on an EASY-spray Acclaim PepMap® analytical column (75 µm i.d. × 50 mm, RSLC C18, 3 µm, 100 Å) using a linear gradient (length: 100 minutes, 3 % to 60 % solvent B (0.1% formic acid in acetonitrile), flow rate: 300 nL/min). The separated peptides were electrospray directly into the mass spectrometer operating in a data-dependent mode using a CID based method. Full scan MS spectra (scan range 350-1500 m/z, resolution 120000, AGC target 1e6, maximum injection time 250 ms) and subsequent CID MS/MS spectra (AGC target 5e4, maximum injection time 100 ms) of the 10 most intense peaks were acquired in the Ion Trap. CID fragmentation was performed at 35 % of normalized collision energy and the signal intensity threshold was kept at 500 counts. The CID method used performs beam-type CID fragmentation. The analysis was performed with Peaks 8.5. The raw MS file was searched against the respective protein sequence. Trypsin with a maximum number of 3 missed cleavages and one unspecific end was selected as the protease. Carbamidomethylation (cysteine), oxidation (methionine) and deamination (asparagine, glutamine) were set as variable modifications. Precursor mass tolerance was set as 15 ppm. Fragment mass tolerances for CID were set to 0.8 Da respectively. All spectra were manually validated. All peptides present a -10lgP >20, spectra were manually checked and validated or disqualified.

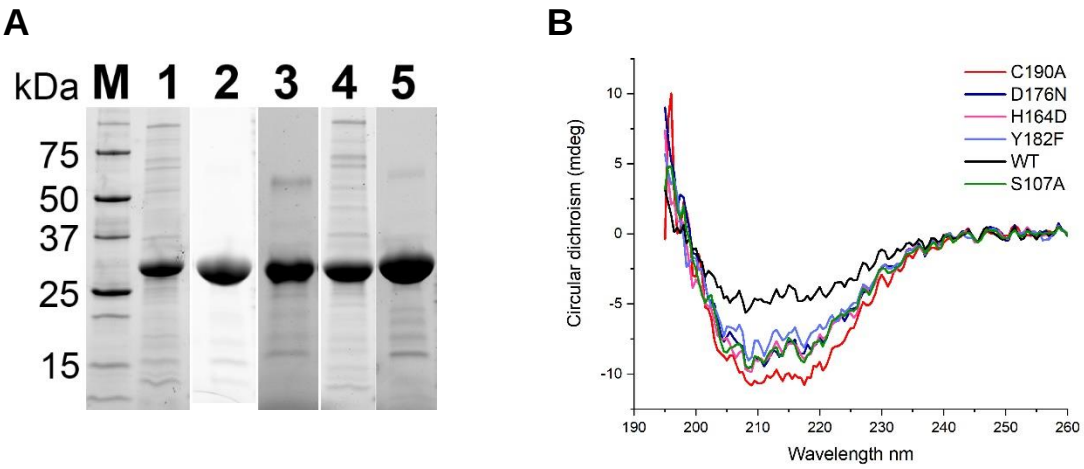
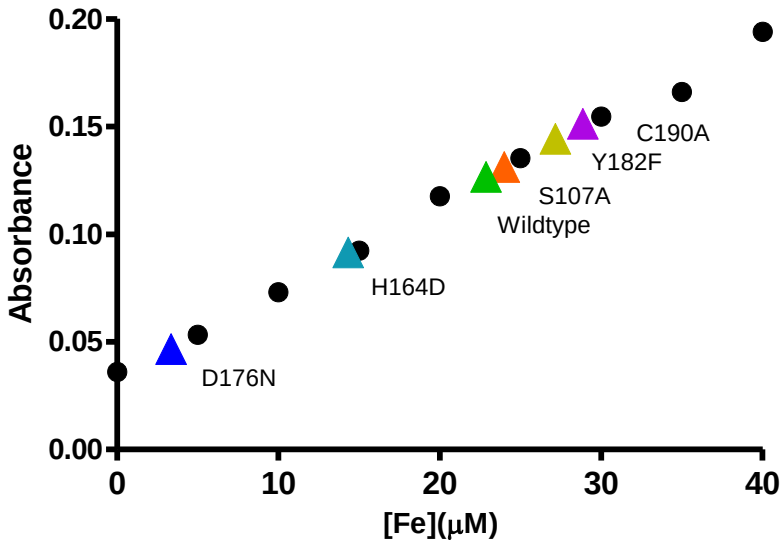


Figure S9. Purification of AtPCO4 variants. (A) SDS PAGE gel demonstrating purity of recombinant AtPCO4 variants. M, molecular weight marker (masses indicated); 1, S107A; 2, H164D; 3, D176N; 4, Y182F; 5, C190A. **(B)** Circular dichroism spectra for AtPCO4 WT and variants demonstrating overall retention of secondary structure in variant proteins; proteins at 0.2 mg/mL in phosphate buffer were scanned at 30 °C in an Applied Photophysics Chirascan™ at wavelengths from 260-195nm in 0.5 nm increments. Data shown represent the average of 3 scans, smoothed (Savitsky-Golay, 5-point window, polynomial order 2). Differences in amplitude are indicative of variability in protein concentration measurement.



AtPCO4 variant	[Fe] (μM)	% occupancy
Wild type	22.85	55.87
S107A	23.98	51.73
H164D	15.53	22.42
D176N	3.33	10.16
Y182F	27.15	57.60
C190A	28.86	47.57

Figure S10. Quantification of Fe in AtPCO4 variants. Fe was extracted from known concentrations of AtPCO4 wild type and variants by denaturing the protein (HCl and heat), then Fe was quenched with bathophenanthroline. Absorbance of the bathopenanthroline:Fe complex was measured at 535nm and compared to a standard curve generated using the same method and known Fe concentrations (black circles). This allowed determination of Fe concentration and, by comparison to protein concentration, calculation of % occupancy.

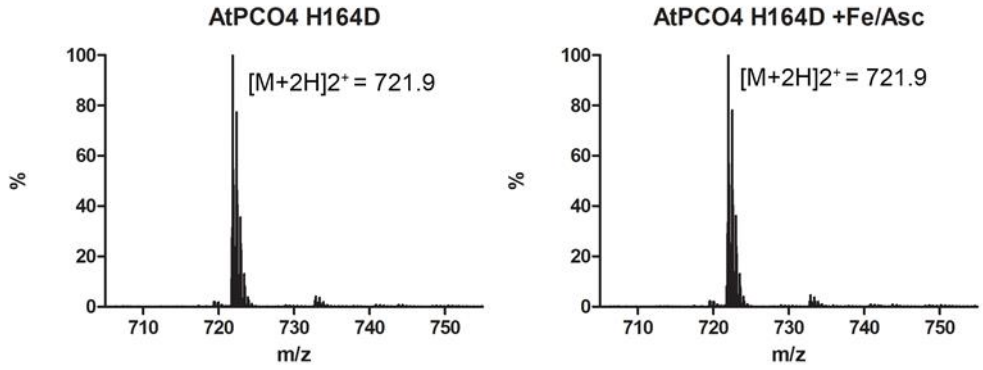


Figure S11. AtPCO4 H164D is inactive including in the presence of exogenous iron and ascorbate. AtPCO4 H164D (0.4 µM) was incubated for 30 minutes with AtRAP2$_{2-15}$ (200 µM) at 30 °C in the presence of 1 mM tris (2-carboxyethyl)phosphine. This took place either in the absence (left panel) or presence (right panel) of 1 mM sodium ascorbate and 40 µM ammonium iron (II) sulphate. Quenched reaction mixtures were analysed using a Xevo G2-S Q-ToF mass spectrometer (Waters) operated in positive electrospray mode. No species were observed which corresponded to the expected m/z for dioxygenated AtRAP2$_{2-15}$ ([M+2H]$^{2+}$ = 737.9), indicating inactive enzyme.

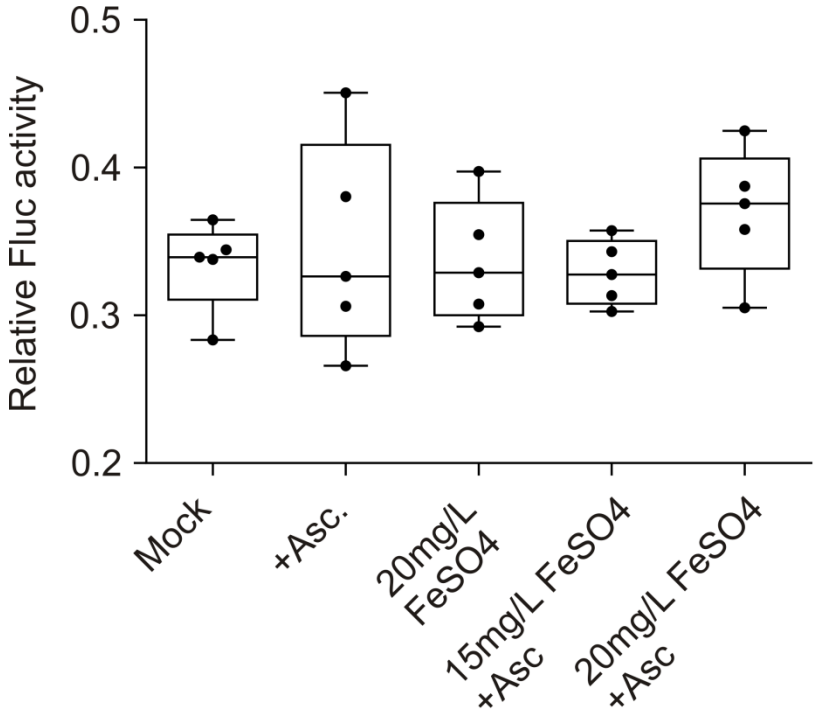


Figure S12. Exogenous Fe (II) and ascorbate does not promote AtPCO4 D176N activity in a *S.cerevisiae* luminescence reporter assay. A dual luciferase oxygen reporter assay (6) was used to determine whether exogenous ascorbate (10 mM) and/or iron could enhance the activity of AtPCO4 D176N. No significant difference was observed in Fluc activity compared to the mock experiment (no supplementation), indicating maximal AtPCO4 D176N under endogenous conditions.

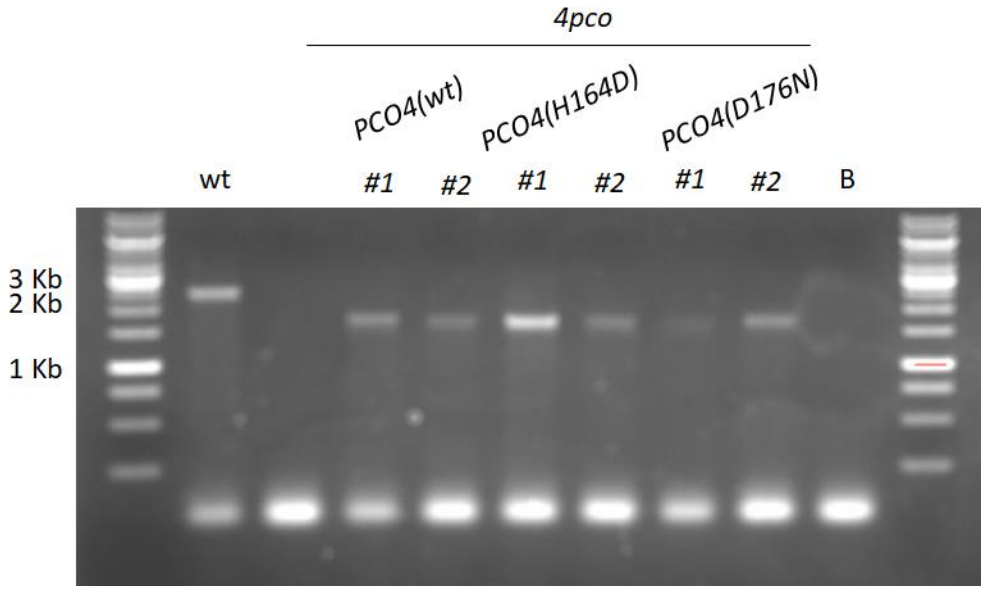


Figure S13. Expression of AtPCO4 transgenes in *4pco* Arabidopsis plants. PCR-amplification of wild type and mutant AtPCO4 sequences in *4pco Arabidopsis thaliana* plants to confirm presence of the transgenes. Lane wt indicates amplification of endogenous *AtPCO4* from the wild type (Col-0) plant. Lane B indicates blank control lane. Other bands represent amplification of transgenes encoding AtPCO4 wild type, H164D and D176N (in two independent transgenic lines, see main text Figure 4D and E). Wild type and transgenic amplicons differ in size due to the absence of introns in the latter.

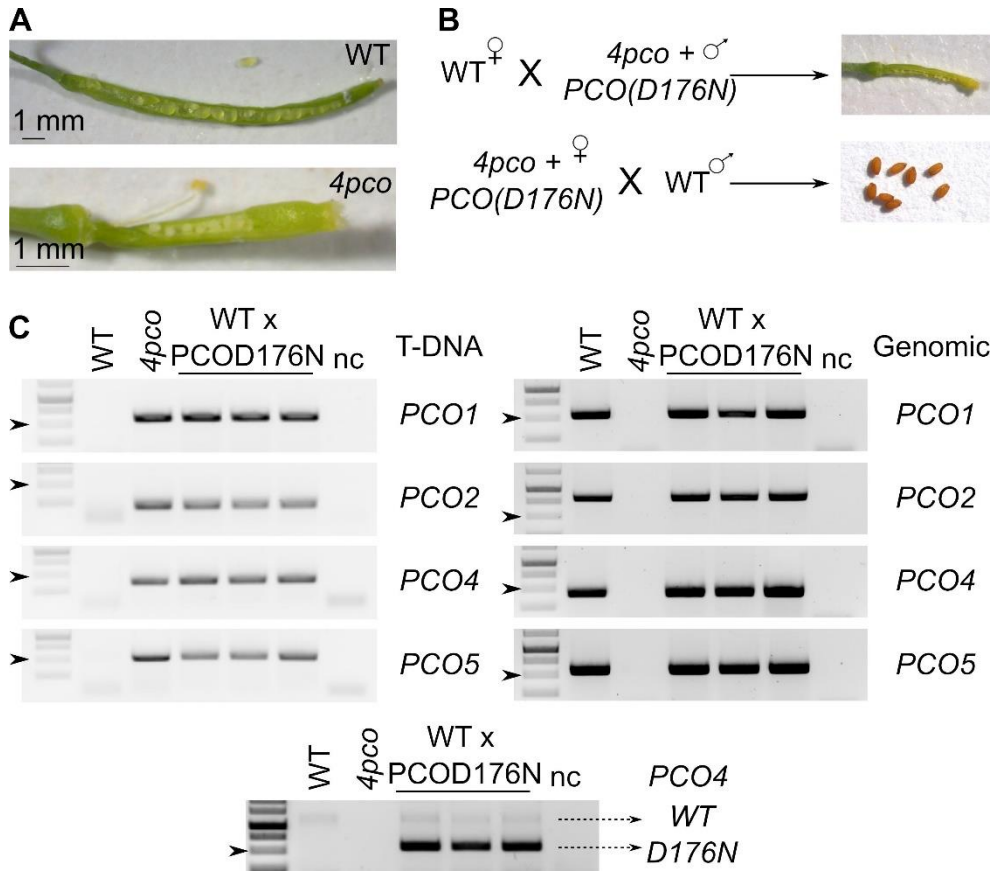


Figure S14. Restoration of *pco4* female fertility by *PCO4(D176N)* expression. (**A**) Comparison of siliques from wild-type (Col-0) and *4pco* plants. (**B**) Scheme depicting the results of reciprocal pollination of *4pco+PCO4(D176N)* and wild type plants. Only wild type pollen could fertilise *4pco+PCO4(D176N)* egg cells. All other crosses or selfing did not lead to seed production. (**C**) PCR-amplification of AtPCO1, 2, 4 and 5 T_DNA insertions (left) and genomic sequences (right) in F1 plants derived from *4pco+PCO4(D176N) x Wt* crossing. The bottom panel shows amplification of full genomic wild type and transgenic versions of PCO4. Arrowheads indicate the 500 bp of the molecular weight marker.

Table S1. Crystallographic data and refinement statistics.

Sample	AtPCO4_1	AtPCO4_2	AtPCO5
PDB ID	6S7E	6S0P	6SBP
Data collection			
Wavelength (Å)	0.9795	0.9762	0.9762
Resolution range (Å)	26.46 - 1.82 (1.89 - 1.82)	37.27 - 1.24 (1.29 - 1.24)	48.73 - 1.91 (1.98 - 1.91)
Space group	P 21 21 21	P 21 21 21	P 41 21 2
Cell dimensions:			
***a*, *b*, *c* (Å)**	46.41, 71.71, 73.22	43.71, 68.99, 71.31	68.91, 68.91, 123.93
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Total reflections	138834 (13767)	741972 (69345)	306302 (31359)
Unique reflections	22311 (2146)	61758 (6069)	23879 (2320)
Multiplicity	6.2 (6.4)	12.0 (11.4)	12.8 (13.5)
Completeness (%)	98.91 (97.68)	98.15 (95.76)	99.49 (95.86)
$I/\sigma I$	12.16 (1.27)	12.68 (0.94)	21.07 (2.01)
Wilson B-factor	32.91	15.03	40.75
R_{merge}	0.076 (1.602)	0.062 (1.471)	0.062 (1.394)
R_{meas}	0.083 (1.743)	0.064 (1.542)	0.065 (1.449)
R_{pim}	0.033 (0.679)	0.018 (0.455)	0.018 (0.391)
$CC_{1/2}$	0.998 (0.627)	0.999 (0.842)	0.999 (0.797)
CC*	0.999 (0.878)	1.000 (0.956)	1.000 (0.942)
Refinement			
Resolution (Å)	26.46 - 1.82 (1.89 - 1.82)	37.27 - 1.24 (1.29 - 1.24)	48.73 - 1.91 (1.98 - 1.91)
No. reflections	22311 (2146)	61758 (5844)	23879 (2224)
No. reflections (R_{free})	1256 (130)	3115 (288)	1164 (115)
R_{work}	0.188 (0.310)	0.179 (0.369)	0.205 (0.608)
R_{free}	0.237 (0.344)	0.209 (0.376)	0.219 (0.560)
CC_{work}	0.959 (0.830)	0.962 (0.892)	0.961 (0.253)
CC_{free}	0.857 (0.756)	0.963 (0.885)	0.949 (0.265)
No. atoms:			
Total	1927	1954	1963
Macromolecules	1778	1765	1855
Ligands	1	1	7
Solvent	148	188	101
Protein residues	222	219	241
R.M.S deviations:			
Bond length (Å)	0.012	0.008	0.004
Bond angles (°)	0.92	1.18	1.22
Ramachandran favoured (%)	97.69	98.56	95.78
Ramachandran allowed (%)	2.31	1.44	3.8
Ramachandran outliers (%)	0	0	0.42
Rotamer outliers (%)	1.56	0.51	3.5
Clash score	1.43	3.99	2.19
B-factors:			
Average	44.09	26.73	62.83
Macromolecules	43.86	25.85	63.24
Ligands	29.9	14.33	73.14
Solvent	46.95	35.04	54.53
Number of TLS groups	10		5

Statistics for the highest-resolution shell are shown in parentheses.

Table S2. Specific activity of AtPCO4 wild type and variants. Initial rates of AtPCO4 wild type and variant activity were measured at 30 °C with 200 µM $RAP2_{2-15}$, 1 mM TCEP and variable (0.1-1 µM) concentrations of enzyme (Figure 5B). Where indicated, FeSO4 (5 µM) and sodium ascorbate (1 mM) were added to D176N assays. Specific activities were calculated as mean ± SD (n=3) based on gradients of linear data fits from 0-1 minute.

Enzyme	Specific activity (µmoles/min/mg)
AtPCO4 WT	11.21 ± 1.48
AtPCO4 S107A	9.11 ± 0.96
AtPCO4 H164D	N.D.
AtPCO4 D176N	0.01 ± 0.02
AtPCO4 D176N + Fe/Asc	1.58 ± 0.14
AtPCO4 Y182F	1.56 ± 0.15
AtPCO4 C190A	11.83 ± 1.00

Table S3. Primers used in this study

Primer name	Forward sequence	Reverse sequence
Variant C190A	TTCAGAGCATGATCGGCATGC CACTTACTTCCGGAAATCC	GGATTTCCGGAAGTAAGTGGCA TGCCGATCATGCTCTGAAG
Variant Y182F	ACATCTTAGCTCCACCTTTCTC TTCAGAGCATGATC	GATCATGCTCTGAAGAGAAAGG TGGAGCTAAGATGT
Variant D176N	CACCCATTGTGCTATTCTTAAC ATCTTAGCTCCACC	GGTGGAGCTAAGATGTTAAGAA TAGCACAATGGGTG
Variant H164D	CGAAAAGTGGTGGCAACATTG ACTGTTTCAAAGCC	GGCTTTGAAACAGTCAATGTTGC CACCACTTTTCG
Variant S107A	CGGGCATGACCGTGCTAGCCA AGCTCGTTTATGGTT	AACCATAAACGAGCTTGGCTAG CACGGTCATGCCCG
PromPCO4	GGGGTTTAAACGTTCTGCAAC CGCCTTTTGAC	GGGACTAGTGATCTGAATCCAAT TCAAAAG
PCO4-H164D_F	AGTGGTGGCAACATTGACTGT TTCAAAGCCATC	GATGGCTTTGAAACAGTCAATGT TGCCACCACT
PCO4-D176N_F	CATTGTGCTATTCTTAACATCT TAGCTCCACCT	AGGTGGAGCTAAGATGTTAAGA ATAGCACAATG
Test of PCO4 insertion on genomic DNA	GGGGAGCTCGTTCTGCAACCG CCTTTTGAC	GCCGTCATCTCAGTATCCTTCA
Test of PCO4 presence on cDNA	caccATGCCTTACTTTGCTCAGAGG	CTGGTAAGTCTTCTCTTCTGGA
PCO4 qPCR	CATGAGTGTGACAGTTTCTCTA TAG	TGGTTCGGTCAGTTGAGGCTCT A
ADH qPCR	TATTCGATGCAAAGCTGCTGTG	CGAACTTCGTGTTTCTGCGGT
HB1 qPCR	TTTGAGGTGGCCAAGTATGCA	TGATCATAAGCCTGACCCCAA
PDC1 qPCR	CACAGAATCTTCAATGTTCTTA CC	CCATGATAAAGCGTACATGGAA
LBD41 qPCR	TGAAGCGCAAGCTAACGCA	ATCCCAGGACGAAGGTGATTG
SAD6 qPCR	TTGGCAACCCGCTTCTTTCTTA CC	TTTCCCTCAGCTCACGAACCTG
UBQ10 qPCR	GGCCTTGTATAATCCCTGATGAATAAG	AAAGAGATAACAGGAACGGAAA CATAGT
genomic_PCO1	AATGGTGGTCCTGGTGTTATTC	GCAAGGTAACAACGACAAACAA
genomic_PCO2	TGTTCTTTTGCCCTCTTCTCTC	TCCGGGTGATGTACAAATACAA
genomic_PCO4	CATGAGCCTGAAGTCTGCAAAA	GCCGTCATCTCAGTATCCTTCA
genomic_PCO5	GCCCATTTAGGTAGCTGCAGTG	AGCTTCCTGTTCGAGACCAA
LBb1	GCGTGGACCGCTTGCTGCAACT	
LGBK	ATATTGACCATCATACTCATTGC	
dspm5	CGGGATCCGACACTCTTTAATTAAC TGACACTC	

Supplementary References

1. Madeira F*, et al.* (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* 47(W1):W636-W641.
2. Robert X & Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42(W1):W320-W324.
3. Ye S*, et al.* (2007) An insight into the mechanism of human cysteine dioxygenase. Key roles of the thioether-bonded tyrosine-cysteine cofactor. *J Biol Chem* 282(5):3391-3402.
4. Dominy JE, Jr.*, et al.* (2008) Synthesis of amino acid cofactor in cysteine dioxygenase is regulated by substrate and represents a novel post-translational regulation of activity. *J Biol Chem* 283(18):12188-12201.
5. Stipanuk MH, Simmons CR, Karplus PA, & Dominy JE, Jr. (2011) Thiol dioxygenases: unique families of cupin proteins. *Amino acids* 41(1):91-102.
6. Puerta ML*, et al.* (2019) A Ratiometric Sensor Based on Plant N-Terminal Degrons Able to Report Oxygen Dynamics in Saccharomyces cerevisiae. *J Mol Biol* 431(15):2810-2820.