Classification: Biological; Plant Biology.

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

Mark D. White[1,2], Laura Dalle Carbonare[3], Mikel Lavilla Puerta[3], Sergio Iacopino[4], Martin Edwards[1], Kate Dunne[1], Elisabete Pires[1], Colin Levy[5], Michael A. McDonough[1]*, Francesco Licausi[3,4]* and Emily Flashman[1]*.

* These authors contributed equally to the work.

Author Affiliations:

[1] Chemistry Research Laboratory, University of Oxford, 12 Mansfield Road, Oxford, OX1 3TA, United Kingdom.
[2] Present address: School of Chemistry, University of Sydney, Sydney NSW 2006 New South Wales, Australia.
[3] Plantlab, Institute of Life Sciences, Scuola Superiore Sant'Anna, Pisa, 56010, Italy.
[4] Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, United Kingdom.
[5] Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom.

ORCIDs:

Mark White: 0000-0002-1431-0339
Laura Dalle Carbonare: 0000-0002-7536-7654.
Mikel Lavilla Puerta: 0000-0001-9920-9020
Sergio Iacopino: 0000-0003-2832-3437
Kate Dunne: 0000-0002-1408-6111
Elisabete Pires: 0000-0001-8646-062X
Colin Levy: 0000-0002-9724-310X
Michael McDonough: 0000-0003-4664-6942
Francesco Licausi: 0000-0003-4769-441X
Emily Flashman: 0000-0002-4169-4278

Corresponding Author:

Emily Flashman; Chemistry Research Laboratory, University of Oxford, 12 Mansfield Road, Oxford, OX1 3TA, United Kingdom; (tel.) +44 1865 275920; emily.flashman@chem.ox.ac.uk.

Keywords

Oxygen-sensing, Plant Cysteine Oxidase, Thiol dioxygenase, Hypoxia, Submergence.

Abstract

In higher plants, molecular responses to exogenous hypoxia are driven by Group VII Ethylene Response Factors (ERF-VIIs). These transcriptional regulators accumulate in the nucleus under hypoxia to activate anaerobic genes but are destabilised in normoxic conditions through the action of oxygen-sensing Plant Cysteine Oxidases (PCOs). The PCOs catalyse the reaction of oxygen with the conserved N-terminal cysteine of ERF-VIIs to form cysteine sulfinic acid, triggering degradation via the Cys/Arg branch of the N-degron pathway. The PCOs are therefore a vital component of the plant oxygen signalling system, connecting environmental stimulus with cellular and physiological response. Rational manipulation of PCO activity could regulate ERF-VII levels and improve flood tolerance, but requires detailed structural information. We report crystal structures of the constitutively expressed PCO4 and PCO5 from *Arabidopsis thaliana* to 1.24 and 1.91 Å resolution, respectively. The structures reveal that the PCOs comprise a cupin-like scaffold, which supports a central metal cofactor coordinated by three histidines. While this overall structure is consistent with other thiol dioxygenases, closer inspection of the active site indicates that other catalytic features are not conserved, suggesting that the PCOs may use divergent mechanisms to oxidise their substrates. Conservative substitution of two active site residues had dramatic effects on PCO4 function both *in vitro* and *in vivo*, through yeast and plant complementation assays. Collectively, our data identify key structural elements that are required for PCO activity and provides a platform for engineering crops with improved hypoxia tolerance.

Significance Statement

The Plant Cysteine Oxidases (PCOs) have been identified as oxygen-sensing enzymes in plants, controlling hypoxia-dependent processes including adaptive responses to flooding. As such, they are potential targets for engineering plants with enhanced flood tolerance. To approach this in a rational manner requires an understanding of how PCO structure relates to their ability to trigger hypoxic adaptation. We report the structures of two PCOs from *Arabidopsis* and show that the effects of mutagenesis of key amino acids at their active site are seen both *in vitro* and, importantly, *in planta*. This work provides a platform for further efforts to manipulate PCO structure and function to improve the ability of crops to withstand future climate extremes.

Introduction

Waterlogged and submerged plants experience reduced oxygen availability (hypoxia), which limits ATP production and threatens survival (1). Flood related crop loss has devastating economic and social impacts, exacerbated by the predicted rise in extreme weather events caused by climate change, which could pose a risk to global food security (2). Accordingly, developing tools and strategies to improve plant survival after flood events is an important goal in addressing this challenge.

Plant physiological responses to hypoxia, including flooding, are driven at the molecular level by Group VII Ethylene Response Factors (ERF-VIIs) (1). These transcription factors upregulate genes that enable adaptation to hypoxic conditions, such as those responsible for anaerobic metabolism (3, 4). Crucially, ERF-VII stability is regulated in an O_2-dependent manner *via* the Cys/Arg branch of the N-degron pathway of protein degradation (5, 6). In normoxia, conserved N-terminal Cys residues are oxidised, rendering the ERF-VIIs targets for N-terminal arginylation by arginyl-tRNA transferases (ATEs). The arginylated N-terminus acts as a signal for ubiquitination, targeting the protein for degradation by the 26S proteasome. In hypoxia, as experienced during flooding, Cys oxidation does not take place and ERF-VIIs are stabilised to initiate an adaptive response.

A direct connection between environmental stimulus and molecular and physiological outcome occurs through the Plant Cysteine Oxidases (PCOs). These enzymes use both atoms of molecular oxygen to catalyse dioxygenation of Nt-Cys to Cys-sulfinic acid (Scheme 1) in ERF-VIIs (as well as other substrates VRN2 and ZPR2 (7, 8)) enabling subsequent N-terminal arginylation (9, 10). PCO activity is sensitive to physiologically-relevant fluctuations in O_2 availability (11), thus these enzymes can act as plant O_2-sensors with a key role in regulating ERF-VII stability. The PCOs share sequence homology and certain characteristics with other thiol dioxygenases (10, 12), but differ (along with cysteamine dioxygenase (13)) in their ability to catalyse oxidation of cysteine residues at the N-termini of polypeptide chains as opposed to the thiol groups of small molecules.

There is evidence that stabilisation of the ERF-VIIs can lead to enhanced plant survival following flooding. Overexpression of the ERF-VII RAP2.12 in Arabidopsis resulted in improved tolerance to submergence and upregulation of genes associated with the hypoxic response (6). When levels of PRT6 (the ubiquitin ligase which targets N-degrons for proteasomal degradation (14, 15)) were decreased in barley using RNA interference, ERF-VII stabilization enabled an upregulation of hypoxic genes which correlated with an improved tolerance to waterlogging (16). Finally, SUB1A-1, an ERF-VII present in certain rare landraces of rice, confers extended submergence tolerance (17) and is reported to avoid degradation by the N-degron pathway *in vitro* (5, 18). While it is important to note that overexpression of stable ERF-VIIs in Arabidopsis under normoxic conditions can be metabolically detrimental to growth and development (6, 19, 20), collectively these observations nevertheless suggest that controlled stabilisation of the ERF-VIIs *via* disruption of the N-degron pathway may provide a suitable mechanism for installing hypoxia-tolerance in plants to facilitate flood survival. The PCOs may be effective targets to manipulate ERF-VII levels in such a controlled fashion, e.g. by modulating their rate of activity, substrate selectivity or their oxygen sensitivity. Structural knowledge of the PCOs is therefore required to enable rational modulation strategies.

There are 5 PCO proteins in *Arabidopsis thaliana* (AtPCOs1-5), with different expression and catalytic characteristics (11). We report crystal structures of AtPCO4 and AtPCO5, two of the constitutively expressed enzymes, which have the greatest catalytic efficiency *in vitro*. We compared their structures with those of other thiol dioxygenases, focussing on *Rattus norvegicus* Cysteine Dioxygenase, RnCDO (PDB ID: 4IEZ), which catalyses conversion of free L-cysteine to L-cys-sulfinic acid as part of amino acid catabolism (12). While broad similarities exist between AtPCOs 4/5 and other thiol dioxygenases, we found that the AtPCOs have distinct structural features that distinguish them from other members of this family, potentially due to a requirement to bind protein, rather than small molecule, substrates. Mutagenesis of key active site residues of AtPCO4 led to partial or complete inactivation of the enzyme, an effect which we observed both with isolated components, but also in yeast cells and *Arabidopsis* plants. This work demonstrates that structure-guided modifications affecting PCO catalysis can be translated into biological environments opening up the future possibility that selective modifications to substrate specificity or oxygen sensitivity may be a useful method to improve stress tolerance in plants.

Results

Crystallisation and Structure Determination of AtPCO4 and AtPCO5

Full-length and catalytically active AtPCO4 and AtPCO5 were produced to over 95% purity using previously described methods (11) and crystallised under anaerobic conditions. An initial structure of AtPCO5 was determined using the native single-wavelength anomalous diffraction method which was then used as a molecular replacement search model for subsequent data sets. Two different structures of AtPCO4 were determined from independent crystallisation conditions, referred to herein as AtPCO4_1 and AtPCO4_2. Both AtPCO4 and AtPCO5 crystallised as monomers with one molecule per asymmetric unit in each crystal form. AtPCO5, AtPCO4_1 and AtPCO4_2 diffracted to 1.91, 1.82 and 1.24 Å resolution, respectively, having a total of 229 out of 242, 220 out of 241 and 212 out of 241 amino acids of the inherent protein sequence modelled into electron density. The final models are of reliable quality as indicated by the geometry and refinement statistics in Supplementary Table S1.

Overall Architecture of AtPCO4 and 5

The structures of AtPCO4_1 (Figure 1A), AtPCO4_2 (Supplementary Figure S1) and AtPCO5 (Figure 1B) consist of a core double-stranded beta-helix (DSBH, also known as a cupin fold), which supports a facial triad of 3 histidine residues coordinating the active site metal ion involved in catalysis. The DSBH fold is representative of many non-heme iron dependent oxygenases including related thiol dioxygenases (e.g. RnCDO, Figure 1C) (12, 21-25). A comparison of AtPCO4 and AtPCO5 indicates high structural conservation (Supplementary Figure S2) with superimposition based on secondary structure matching resulting in root mean square deviation (RMSD) of 1.02 Å for AtPCO4_1 compared with AtPCO5 (with 213 out of a possible 222 residues aligned). In comparison, AtPCO4_1 structural conservation with RnCDO (PDB ID: 4IEZ) is lower with a RMSD of 2.22 Å (with 145 out of a possible 186 residues aligned), predominantly due to structural deviations in regions flanking the conserved DSBH core.

A significant structural difference between the AtPCOs and RnCDO is a C-terminal loop (residues 220-231), which forms a border along one side of the AtPCO active site opening (Supplementary Figure S3), resulting in a large cavity with the potential to accommodate large polypeptide substrate binding. In contrast, an equivalent loop in RnCDO (residues 174-182) lies across the active site forming a 'lid' that restricts access to the catalytic metal (Supplementary Figure S3). Interestingly, this cavity is occupied by a section of the N-terminal His tag in AtPCO5, demonstrating that peptide binding is possible in this region (Supplementary Figure S4).

Another distinct feature observed in AtPCO4_1 and AtPCO5 is the presence of a hairpin loop (residues 182- 190) projecting from DSBH β-strand VIII at the active site opening opposite to the C-terminal chain 220-231 described above. The loop culminates in a beta turn consisting almost exclusively of charged and polar residues (Tyr_{182}-Ser_{183}-Ser_{184}-Glu_{185}-His_{186}-Asp_{187}-Arg_{188}-His_{189}-Cys_{190} for AtPCO4, Supplementary Figure S5). The vicinity of this hairpin loop close to the active site suggests that it may play a role in substrate binding and recognition. There is sequence divergence in much of this region amongst the AtPCOs (Supplementary Figure S6), which may in turn contribute to their divergent substrate specificities (11). Interestingly, an equivalent hairpin loop is not observed in RnCDO, possibly since its smaller cysteine substrate requires fewer interactions closer to the catalytic centre. In AtPCO4_2 this loop is somewhat disordered, possibly due to the binding of an N-terminal His_6 tag at the active site in this crystal form (Supplementary Figures S1 and S7).

The AtPCO4/5 Active Site

Closer inspection of the AtPCO active sites reveals that the metal cofactor is bound along one face by His98, His100 and His164 in an octahedral arrangement, mirroring the metal coordination of other thiol dioxygenases (Figure 2) (12). Three well-ordered H_2O molecules occupy the remaining ligation sites of AtPCO4_1 and AtPCO5. As described above, electron density corresponding to a 5 amino acid fragment from the N-terminal His_6 tag of an adjacent molecule is observed in the active site of PCO4_2 at approximately 75 % occupancy, with a H_2O molecule satisfying the remaining density (Supplementary Figure S7). Residue His(-)10 (from this His_6 tag) directly interacts with the metal cofactor, displacing the water molecule trans to His164 and ligating the metal through the τ-nitrogen of the imidazole ring (Figure 2B). Attempts to crystallize AtPCO4 and AtPCO5 without the N-terminal His_6 tag were unsuccessful possibly suggesting that the observed intermolecular interactions facilitate crystal formation. AtPCO4_1 and AtPCO4_2 were modelled with catalytically relevant Fe in their active sites, while AtPCO5 was modelled with Ni in the active site, a known contaminant of Ni-affinity chromatography used in AtPCO purification (10, 11).

Comparison of AtPCOs with other classes of thiol dioxygenases

The PCOs are unique among the three types of thiol dioxygenases characterised to date (CDOs, 3MDOs (mercaptopropionate dioxygenases) and MSDOs (mercaptosuccinate dioxygenases)) and represent a new class. In the CDOs, 3MDOs and MSDOs, substrate specificity and reactivity are influenced by the presence and position of an Arg residue in the active site (25, 26). In the CDOs, Arg60 forms a strong salt bridge with the carboxylate group of the active site metal bound L-cysteine, while in other thiol dioxygenases, there is a Gln at position 60 and an Arg residue elsewhere in the active site which contributes to substrate carboxylate binding (e.g. Arg168 in *Pseudomonas aeruginosa* 3MDO)(26). The PCOs do not possess an Arg residue in the active site, which is unsurprising since the cysteinyl substrates of AtPCOs have a peptide bond in place of the charged carboxylate group of free cysteine.

Small molecule thiol dioxygenases have a Ser-X-His-X-Tyr motif on DSBH β-strand VIII that is proposed to act as a catalytic triad. In mammalian CDOs, a thioether crosslink forms between the Tyr residue in this motif and an adjacent cysteine (Cys93, RnCDO) from DSBH β-strand III following multiple turnovers. This unusual post-translational modification is reported to improve substrate binding and enzyme activity (27-30). Strikingly the distribution of amino acids in the AtPCO4/5 active sites is considerably different (Figure 3): An Asp residue (Asp176 in AtPCO4) replaces His155 of the RnCDO Ser-X-His-X-Tyr motif, while Ser153, Tyr157 and Cys93 are replaced with chemically inert side chains at equivalent positions (Ile174, Leu178 and Val105, respectively, in AtPCO4). These differences are also conserved in AtPCOs1-3 as well as PCOs from other plants (Supplementary Figure S6). This suggests that PCO enzymes cannot form an equivalent thioether crosslink, however residues elsewhere in the AtPCO active sites do have the potential to form a similar post-translational modification. Thioether crosslink formation has been observed in the active site of human cysteamine dioxygenase, HsADO, between Cys220 and Tyr222 (13, 31). HsADO is a thiol dioxygenase which also catalyses oxidation of protein N-terminal cysteine residues and retains sequence homology to the AtPCOs in relevant active site regions (Supplementary Figure S6). We therefore conceived thioether crosslink formation as a possibility for the AtPCOs, specifically via Cys190 (equivalent to Cys220 in HsADO) which is sandwiched between Tyr182 and Tyr192 and borders the hairpin loop proposed to be involved in substrate recognition and binding. However, electron density for a crosslink was not observed in the AtPCO4/5 crystal structures and tandem MS/MS analysis did not show a crosslink involving

Cys190, even following prolonged incubation of AtPCO4 with a Cys-initiating peptide representing its known substrate (equivalent to the conditions under which crosslink formation was observed for CDO) (Supplementary Figure S8). Under the conditions tested therefore the unusual post-translational modification observed to enhance turnover in mammalian CDOs was not identified in the AtPCOs, though we cannot rule out that a cross-link may be observable using different techniques (31).

Modulation of AtPCO4 Activity by Targeted Mutagenesis at the Active Site

To explore the role of AtPCO4/5 active site residues in catalysis we generated variants of AtPCO4 using site directed mutagenesis. We targeted chemically reactive residues which either have the potential to fulfil similar functions to those important in other thiol dioxygenases (Tyr182, Cys190), which project towards the metal centre (Ser107, Asp176) or which contribute to metal binding (His164). The targeted residues are also conserved in all AtPCOs, as well as PCO sequences from other plants (Supplementary Figure S6). Variants were designed conservatively, to investigate the functionality of individual side chains while maintaining structural integrity of the enzymes. AtPCO4 was used as a model enzyme, since its active site is very similar to that of AtPCO5 yet it has a greater specific activity (11) thus the effects of mutagenesis are likely to be more profound.

Variants S107A, H164D, D176N, Y182F and C190A were generated by site-directed mutagenesis then expressed and purified following the same procedure as for wild type AtPCO4, resulting in enzymes that were >90% pure and consistent in secondary structure (Supplementary Figure S9). We initially conducted endpoint assays (30 minutes) to determine the ability of the variants to catalyse oxidation of a 14-mer peptide representing the Cys-initiating N-terminus of the Arabidopsis ERF-VII RAP2.12 (herein referred to as AtRAP2.12$_{2\text{-}15}$). Variants S107A and C190A catalysed AtRAP2.12$_{2\text{-}15}$ oxidation to the same degree as wild type enzyme under these conditions, whereas Y182F activity was reduced to ~60% that of wild type indicating a reduced rate of activity for this variant. D176N and H164D showed minimal and near-zero activity, respectively (Figure 4A).

We next measured initial rates of enzyme activity for the variants (Figure 4B, Supplementary Table S2). C190A and S107A were found to catalyse AtRAP2.12$_{2\text{-}15}$ oxidation at the same rate as wild type enzyme (11.83 (± 1.00), 9.11 (± 0.96) and 11.21 (± 1.48) µmoles/min/mg, respectively) indicating that these amino acid substitutions have limited effect on enzyme function (and providing further evidence that catalytically relevant thioether formation *via* Cys190 does not occur). Y182F demonstrated a significantly lower rate of AtRAP2.12$_{2\text{-}15}$ turnover (1.56 11.21 (± 0.15) µmoles/min/mg) indicating an important role for the hydroxyl group in AtPCO4 activity. Interestingly, the specific activity of D176N was extremely low under equivalent conditions (0.01 (± 0.02) µmoles/min/mg), however a very low level of iron was detected in this variant (Supplementary Figure S10) and when additional iron and ascorbate were included in the assay, the rate of activity increased to 1.58 (± 0.14) µmoles/min/mg. Asp176 may therefore play at least a partial role in maintaining the integrity and/or redox state of the Fe-binding site. H164D specific activity could not be measured as it was completely inactive, even in the presence of exogenous iron and ascorbate (Supplementary Figure S11). While His164 has a key role in iron binding, it was anticipated that its mutation to Asp would retain function given that Asp is a common Fe-ligating residue in iron-dependent dioxygenases (32, 33). Indeed the H164D variant retained some Fe-binding capability (22.4 % occupancy compared to 55.9 % occupancy for wild type, Supplementary Figure S10), suggesting that the mutation negatively impacts substrate binding and/or the ability of the Fe cofactor to activate oxygen for reaction.

We next tested the impact of the two most significant mutations, H164D and D176N, *in vivo*. First, we exploited *Saccharomyces cerevisiae* as an exogenous expression system, by means of a dual luciferase oxygen reporter whose relative luminescence depends on the presence of oxygen and N-terminal cysteine oxidase activity (34). Using this reporter system, AtPCO4/variant activity can be detected through degradation of the luminescent reporter *via* the Arg/Cys N-degron pathway.

As previously observed (34), AtPCO4 expression abated the reporter's output under aerobic conditions but not in hypoxia (1% O_2, 6 h, Figure 4C). Expression of AtPCO4 H164D led to a strong increase of the reporter signal under both normoxia and hypoxia (Figure 4C), consistent with abolition of PCO activity in this variant as seen in the experiments with isolated proteins. Interestingly, for this variant, reporter signal was consistently higher in normoxic than hypoxic conditions. The reason for this is unclear but could be due to non-catalytic interaction of H164D with the reporter protein under these conditions. Expression of AtPCO4 D176N resulted in significant luminescence under aerobic conditions as compared to AtPCO4 wild type, but also considerably less luminescence than was observed with H164D. There was also an increase in luminescence in hypoxia for D176N as compared to normoxia. These results are consistent with the D176N variant retaining a limited amount of cysteine oxidase activity, as observed *in vitro*, while the H164D variant is completely inactive. Unlike the *in vitro* experiments, addition of exogenous iron and ascorbate to aerobic cultures expressing the D176N variant did not enhance the degradation of the oxygen-labile reporter (Supplementary Figure S12), possibly because endogenous levels of (reduced) iron in yeast are sufficient to promote maximal AtPCO4 D176N activity.

Finally, we attempted *in planta* complementation of AtPCO4 function using the H164D and D176N variants. Inactivation of four PCO genes in *Arabidopsis thaliana* by insertional mutagenesis has previously been shown to cause developmental defects, including retarded growth, leaf serration and male and female sterility, along with constitutive expression of low-oxygen response genes (13). We expressed the wild-type AtPCO4 coding sequence, under the control of its own promoter (Supplementary Figure S13), in these *4pco* mutant plants and found that this was sufficient to prevent developmental and molecular defects (Figure 4D,E). In contrast, and consistent with their activity in both isolated protein and yeast assays, expression of both variants AtPCO4 H164D and D176N in the *4pco* mutant plants failed to replicate AtPCO4 wild type function (Figure 4D). Both variants also displayed upregulation of hypoxia-responsive genes under aerobic conditions (Figure 4E). Nonetheless, expression of AtPCO4 D176N suggested marginal improved growth compared to the *4pco* phenotype (Figure 4D) and was able to restore female but not male fertility in *4pco* flowers, as demonstrated by reciprocal crossing with wild-type plants (Supplementary Figure S14). All *PCO4* transgenes exhibited comparable level of expression in two independent lines for each construct, and consistently higher than the wild type gene (Figure 4E), possibly due to the existence of additional transcriptional or post-transcriptional regulation in the introns or 3' untranslated genomic regions of the *PCO4* gene.

Discussion

The structures of AtPCO 4 and 5 presented here provide the first structural characterization of thiol dioxygenases that catalyze dioxygenation of macromolecular substrates, more specifically, cysteine residues positioned at the N-termini of proteins. This work has revealed that although the overall DSBH fold (common to many Fe-dependent dioxygenases) of PCOs is conserved along with the 3 x His Fe-binding motif typical of small molecule thiol dioxygenases , additional features identified as important for substrate binding and catalysis in the small molecule thiol dioxygenases are not conserved in the PCOs. Some of these differences can be rationalized by the requirement for the AtPCOs to bind large polypeptide rather than small molecule substrates. This includes an open active site in the AtPCOs and the loss of active site Arg residues, important for binding substrate carboxylate groups (replaced with peptide bonds in protein substrates) in the small molecule thiol dioxygenases. Other differences include significant divergence amongst the amino acids surrounding the active site: a Ser-His-Tyr catalytic triad and a well-characterized Cys-Tyr thioether cross-link in small molecule thiol dioxygenases are not conserved in AtPCOs. Overall, these differences indicate that the AtPCOs have evolved specifically to enable their function in catalyzing oxidation of Nt-Cys initiating proteins and further work will help to identify amino acids with important roles in substrate binding and oxygen activation in the AtPCO active site.

A thiol dioxygenase in humans (ADO, cysteamine dioxygenase) has recently been reported to also catalyze the oxidation of N-terminal cysteine residue of proteins in an oxygen-sensitive manner (13). While there are no reported structures for ADO to date, sequence alignments with RnCDO, AtPCOs, and PCOs from other plant species (Supplementary Figure S6) suggest that some of the active site features seen in the PCOs are also conserved in HsADO, in contrast to those seen in RnCDO. While larger structural features, such as loop positioning across the active site entrance, are harder to predict, the sequence alignment suggests that the proposed RnCDO Ser-His-Tyr catalytic triad, is not conserved in either PCOs or HsADO. Furthermore, AtPCO Asp176 is conserved both in PCOs from other plants and HsADO, supporting an important role for this residue in catalysis as evidenced by our mutagenesis experiments. Interestingly a Cys-Tyr thioether crosslink, a feature important in RnCDO and other small thiol dioxygenases (24, 25, 35), was detected in HsADO between Cys220 and Tyr222 following incubation with cysteamine (31). These residues are conserved in PCOs (AtPCO4 Cys190 and Tyr192, Supplementary Figure S6), yet we did not observe a crosslink under the conditions used in this study; the thioether crosslink reaction may only occur in mammalian enzymes or upon initiation with small molecule thiols. It is possible that the active site features which appear to be common to the PCOs and HsADO may be related to the polypeptide nature of their substrates in contrast to small molecule thiol dioxygenases.

The structures we report are also the first of oxygen-sensing enzymes from plants. These enzymes were initially reported in relation to their role in catalyzing oxidation of N-terminal cysteine residues of ERF-VII transcription factors in *Arabidopsis*, triggering their degradation under normoxic conditions *via* the Cys/Arg N-degron pathway (9, 10). Since these reports, additional roles have been identified for the PCOs in hypoxic signaling, through targeting the alternative Nt-Cys initiating proteins VRN2 and LITTLE ZIPPER 2 (7, 8). With ~250 Nt-Cys initiating proteins in the Arabidopsis proteome, the list of physiologically relevant PCO substrates is likely to expand. The importance of O_2 as a signaling molecule, and of the PCOs as the transducer of this signaling molecule, is therefore emerging and further structural insight into how these enzymes interact with substrates and oxygen will enable molecular rationalization of these processes.

Particularly attractive is the possibility of manipulating PCO activity to modulate the signaling response to hypoxia. We have demonstrated in this study that it is possible to adopt a structure-guided approach to target functionally important residues (H164 and D176) in the AtPCO4 active site, then to observe consistent effects on the activity of AtPCO4 in both *in vitro* and *in vivo* environments, including in a plant model. While the effect of ablating AtPCO function was not beneficial for the overall plant phenotype (demonstrating the importance of these enzymes to plant physiology), this work lays the foundations for further structure-guided rational manipulation of AtPCO activity to alter specific properties, such as oxygen-sensitivity or substrate selectivity. This is appealing from the perspective of fundamental and applied research, exemplified in this study by the demonstration that selective expression of a partially active PCO variant indicated a requirement for oxygen-sensing in the development of reproductive tissues (Supplementary Fig. 14). Such an approach may be applied with alternative PCO variants to study organ or tissue-specific sensitivity to hypoxia. Moreover, structural characterization of oxygen-sensing enzymes in humans (Hypoxia-Inducible Factor hydroxylases (36, 37)) ultimately enabled development of therapeutic agents for the treatment of hypoxic diseases. Given the role the PCOs play in regulating ERF-VII levels, and the importance of ERF-VII signaling in adapting to submergence, it is possible that manipulation of PCO activity may be an effective strategy to enhance flood tolerance in plants.

Materials and Methods

Preparation of recombinant AtPCO4 and 5

AtPCO4 and AtPCO5 were expressed and purified as previously described (11). Briefly, pET-28a vectors incorporating sequences encoding AtPCO4 and 5 proteins were transformed into BL21 (DE3) *E. coli* and protein expression induced with 0.5 mM IPTG. Following overnight expression at 20 °C, cells were harvested and frozen. To purify the proteins, cells were lysed by sonication and soluble lysate was passed over a Ni-affinity HisTrap column. Elution of His_6-AtPCO4/5 with imidazole was followed by size exclusion chromatography, resulting in protein of >95% purity.

Crystallization and structure determination

All crystals grew within two weeks at room temperature using vapor diffusion methods under a nitrogen atmosphere in an anaerobic glove box (Belle Technology). AtPCO5 crystals were prepared by adding 1 μL protein (25 mg/mL) to 1 μL mother liquor (20% PEG 3350, 0.2 M ammonium nitrate) and equilibrated against a 500 μL precipitant reservoir. AtPCO4_1 crystals were prepared by adding 0.45 μL protein (45 mg/mL, mixed with $RAP2_{2\text{-}15}$ to a final concentration of 8 mM) to 0.45 μL mother liquor (20 % PEG 3350, 0.2 M sodium fluoride) equilibrated against a reservoir containing 30 μL of precipitant and streak seeded using seed stock prepared from AtPCO4_2 crystals. Crystals of the AtPCO4_2 crystal form were prepared by adding 1 μL protein (45 mg/mL) to 1 μL mother liquor (0.2 M ammonium sulfate, 0.1 M Tris (HCl) pH 7.5) and equilibrated against a 500 μL precipitant reservoir using macro-seeding with seed stock obtained from the same conditions.

The AtPCO5 structure was solved using BigEP (38). The anomalous signal was sufficient for Autosol to locate 7 sites (FOM=0.28) followed by solvent flattening and autobuilding in AutoSHARP resulted in 183 residues being built (39, 40). Manual fitting and refinement was carried out using COOT and PHENIX (41, 42). The model of AtPCO5 was then used as a search model for solving AtPCO4 structures by molecular replacement using PHASER (43). Data collection and refinement statistics can be found in Supplementary Table 1.

Preparation of recombinant AtPCO4 variants

Active-site substitutions were introduced by site-directed mutagenesis and the integrity of the mutations was verified by sequencing (Source Bioscience). Variant enzymes were expressed and purified as per the wild type enzyme described above.

Recombinant AtPCO4 wild type and variant activity assays

Wild type or variant AtPCO4 activity was measured with synthesized peptide substrate (GL Biochem) representing the Cys-initiating N-terminus of the Arabidopsis ERF-VII RAP2.12 (CGGAIISDFIPPPR), hereafter termed $RAP2_{2\text{-}15}$. For endpoint assays, enzyme (0.4 µM) and $RAP2_{2\text{-}15}$ (200 µM) were incubated in the presence of 1mM TCEP for 30 minutes at 30 °C. For specific activity determination, enzyme (0.1 -1 µM) and $RAP2_{2\text{-}15}$ (200 µM) were incubated at 30°C in the presence of 1 mM TCEP (D176N assays were supplemented with 5µM Fe and 1 mM ascorbate where indicated) and aliquots removed at time points over 0 – 2 minutes. Reactions were quenched with 1% formic acid and $RAP2_{2\text{-}15}$ oxidation was determined by ultrahigh-performance liquid chromatography mass spectrometry, as described in (11).

Measuring AtPCO4 wild type, H164D and D176N activity in *S. cerevisiae* dual luciferase oxygen reporter assay

The *S.cerevisiae* DLOR assay was performed as previously described in (34). Briefly, expression vectors encoding C-DLOR and AtPCO4 wild type/H164D/D176N were transformed into Haploid *S. cerevisiae* MaV2013 strain (Thermo-Fisher Scientific). Successfully transformed colonies were grown at 30°C in liquid yeast synthetic medium (as described in (34)) supplemented with filter-sterilized $FeSO_4$ (15/20 mg/mL) and ascorbic acid (10 mM) for AtPCO4 D176N assays (or double distilled water as mock). Overnight cultures were diluted to OD_{600} 0.1 and grown at 150 rpm for 6 h at 30°C either under aerobic conditions or at 1% O_2 in a gloveless anaerobic chamber (COY). Cells were harvested and luciferase activity was measured using the Dual-Luciferase® Reporter ($DLOR_{TM}$) Assay System (Promega) according to the manufacturer's instructions.

***In planta* complementation assay**

The Columbia-0 (Col-0) ecotype of *Arabidopsis thaliana* was used as the wild-type reference. The genotype *4pco* has been previously described in (13). Seeds were stratified at 4°C in the dark for

48 h and germinated on moist soil at 22°C day/18°C night with a photoperiod of 12 h. For *in vitro* seedlings cultivation, seeds were sterilized and grown on agarized medium composed of half-strength Murashige and Skoog basal salt mixture (Duchefa) and 0.8% plant agar (Duchefa).

Codon substitutions, H164D and D176N, of the PCO4 wild-type sequence were achieved by means of site-directed mutagenesis using the pENTR223:PCO4 vector (ABRC stock code G82286) as template. The 1263-bp-long *PCO4* promoter sequence was amplified from wild-type genomic DNA using the Phusion High-Fidelity DNA Polymerase (New England BioLabs) with forward and reverse primers containing the PmeI and SpeI sites, respectively. These were used to replace the CaMV 35S promoter in the destination vector pH7WG2 to generate pH7WG-promPCO4. A complete list of primers used in this study for cloning of the described fragments is provided in Supplementary Table S3. The wild type and mutant versions of the PCO4 coding sequence were recombined into pH7WG-promPCO4.

Stable transgenic plants were obtained via *Agrobacterium*-mediated transformation using the floral dipping method (44). T_0 seeds were selected on agarized media containing 50 mgl^{-1} hygromicin. The presence of *promPCO4:PCO4wt*, *promPCO4:PCO4(H164D)* or *promPCO4:PCO4(D176N)* T-DNA was tested via PCR using a GoTaq® DNA polymerase (Promega) and the primers listed in Supplementary Table S3.

A T_3 segregating population was used for the phenotypic and molecular characterization of the PCO4-expressing plants, after genotyping for T-DNA insertion in the segregating *PCO5* locus as described in (13). RNA was extracted from fully developed rosette leaves, processed into cDNA as described in (13). Real-time qPCR amplification of housekeeping (UBQ10) and hypoxia-responsive genes was carried out in the ABI Prism 7300 sequence detection system (Applied Biosystems), using the PowerUp™ SYBR® Green Master Mix (Thermo Fisher Scientific) and 10 ng cDNA in each reaction. Relative gene expression values were calculated with the ΔΔCt method.

Wild type and *4pco(D176N)* plants were reciprocally crossed by pollination of emasculated flowers. Photos of siliques bearing seeds with arrested development were taken two weeks after fertilisation. Successfully developing siliques were not dissected and maintained attached to the plant until complete desiccation. F1 plants were grown and tested to validate the success of crossing by identifying the heterozygous state of all T-DNA insertions.

Data Sharing

X-ray diffraction data were deposited in the Research Collaboratory Structural Bioinformatics (RCSB) Protein Data Bank (https://www.rcsb.org) under PDB ID codes 6S7E (AtPCO4_1), 6S0P (AtPCO4_2) and 6SBP (AtPCO5).

Author Contributions

MW – crystallized proteins and solved structures, designed variants, analyzed data, wrote manuscript; LDC – produced and characterized transgenic plants; MLP – carried out experiments with yeast; SI – produced and characterized transgenic plants; ME – generated and characterized AtPCO4 variants; KD – generated AtPCO4 variants; EP – conducted tandem mass spectrometry of cross-linked peptides; CL – supervised anaerobic crystallization and collected data for AtPCO4_1; MAM – supervised structural work; FL – supervised yeast and plant experiments, analyzed plant phenotypes ; EF – conceived and supervised study, analyzed data, wrote manuscript.

Acknowledgments

This work was supported by Biotechnology and Biological Research Council (UK) New Investigator Grant BB/M024458/1, the European Research Council (European Union's Horizon 2020 research and innovation programme grant agreement No. 864888) and Italian Ministry of Education University and Research PRIN grant 20173EWRT9.

References

1. Bailey-Serres J*, et al.* (2012) Making sense of low oxygen sensing. *Trends Plant Sci* 17(3):129-138.
2. Bailey-Serres J, Parker JE, Ainsworth EA, Oldroyd GED, & Schroeder JI (2019) Genetic strategies for improving crop yields. *Nature* 575(7781):109-118.
3. Mustroph A*, et al.* (2010) Cross-kingdom comparison of transcriptomic adjustments to low-oxygen stress highlights conserved and plant-specific responses. *Plant Physiol* 152(3):1484-1500.
4. Giuntoli B & Perata P (2018) Group VII Ethylene Response Factors in Arabidopsis: Regulation and Physiological Roles. *Plant Physiol* 176(2):1143-1155.
5. Gibbs DJ*, et al.* (2011) Homeostatic response to hypoxia is regulated by the N-end rule pathway in plants. *Nature* 479(7373):415-418.
6. Licausi F*, et al.* (2011) Oxygen sensing in plants is mediated by an N-end rule pathway for protein destabilization. *Nature* 479(7373):419-422.
7. Gibbs DJ*, et al.* (2018) Oxygen-dependent proteolysis regulates the stability of angiosperm polycomb repressive complex 2 subunit VERNALIZATION 2. *Nature Commun* 9(1):5438.
8. Weits DA*, et al.* (2019) An apical hypoxic niche sets the pace of shoot meristem activity. *Nature* 569(7758):714-717.
9. Weits DA*, et al.* (2014) Plant cysteine oxidases control the oxygen-dependent branch of the N-end-rule pathway. *Nature Commun* 5:3425.
10. White MD*, et al.* (2017) Plant cysteine oxidases are dioxygenases that directly enable arginyl transferase-catalysed arginylation of N-end rule targets. *Nature Commun* 8:14690.
11. White MD, Kamps JJAG, East S, Taylor Kearney LJ, & Flashman E (2018) The Plant Cysteine Oxidases from Arabidopsis thaliana are kinetically tailored to act as oxygen sensors. *J Biol Chem*.
12. Stipanuk MH, Simmons CR, Karplus PA, & Dominy JE, Jr. (2011) Thiol dioxygenases: unique families of cupin proteins. *Amino acids* 41(1):91-102.
13. Masson N*, et al.* (2019) Conserved N-terminal cysteine dioxygenases transduce responses to hypoxia in animals and plants. *Science* 365(6448):65-69.
14. Garzon M*, et al.* (2007) PRT6/At5g02310 encodes an Arabidopsis ubiquitin ligase of the N-end rule pathway with arginine specificity and is not the CER3 locus. *FEBS Lett* 581(17):3189-3196.
15. Graciet E*, et al.* (2009) The N-end rule pathway controls multiple functions during Arabidopsis shoot and leaf development. *Proc Natl Acad Sci USA* 106(32):13618-13623.
16. Mendiondo GM*, et al.* (2016) Enhanced waterlogging tolerance in barley by manipulation of expression of the N-end rule pathway E3 ligase PROTEOLYSIS6. *Plant Biotech J* 14(1):40-50.
17. Xu K*, et al.* (2006) Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* 442(7103):705-708.
18. Lin C-C*, et al.* (2019) Regulatory cascade involving transcriptional and N-end rule pathways in rice under submergence. *Proc Natl Acad Sci USA* 116(8):3300-3309.
19. Paul MV*, et al.* (2016) Oxygen Sensing via the Ethylene Response Transcription Factor RAP2.12 Affects Plant Metabolism and Performance under Both Normoxia and Hypoxia. *Plant Physiol* 172(1):141-153.
20. Giuntoli B*, et al.* (2017) Age-dependent regulation of ERF-VII transcription factor activity in Arabidopsis thaliana. *Plant, Cell & Environ* 40(10):2333-2346.
21. Driggers CM*, et al.* (2013) Cysteine dioxygenase structures from pH4 to 9: consistent cys-persulfenate formation at intermediate pH and a Cys-bound enzyme at higher pH. *J Mol Biol* 425(17):3121-3136.

22. Driggers CM, Hartman SJ, & Karplus PA (2015) Structures of Arg- and Gln-type bacterial cysteine dioxygenase homologs. *Protein Sci* 24(1):154-161.
23. Joseph CA & Maroney MJ (2007) Cysteine dioxygenase: structure and mechanism. *Chem Commun (Camb)* (32):3338-3349.
24. McCoy JG*, et al.* (2006) Structure and mechanism of mouse cysteine dioxygenase. *Proc Natl Acad Sci USA* 103(9):3084-3089.
25. Simmons CR*, et al.* (2006) Crystal structure of mammalian cysteine dioxygenase. A novel mononuclear iron center for cysteine thiol oxidation. *J Biol Chem* 281(27):18723-18733.
26. Aloi S, Davies CG, Karplus PA, Wilbanks SM, & Jameson GNL (2019) Substrate Specificity in Thiol Dioxygenases. *Biochemistry* 58(19):2398-2407.
27. Davies CG, Fellner M, Tchesnokov EP, Wilbanks SM, & Jameson GN (2014) The Cys-Tyr cross-link of cysteine dioxygenase changes the optimal pH of the reaction without a structural change. *Biochemistry* 53(50):7961-7968.
28. Driggers CM*, et al.* (2016) Structure-Based Insights into the Role of the Cys-Tyr Crosslink and Inhibitor Recognition by Mammalian Cysteine Dioxygenase. *J Mol Biol* 428(20):3999-4012.
29. Li W, Blaesi EJ, Pecore MD, Crowell JK, & Pierce BS (2013) Second-sphere interactions between the C93-Y157 cross-link and the substrate-bound Fe site influence the O(2) coupling efficiency in mouse cysteine dioxygenase. *Biochemistry* 52(51):9104-9119.
30. Li W & Pierce BS (2015) Steady-state substrate specificity and O2-coupling efficiency of mouse cysteine dioxygenase. *Arch Biochem Biophys* 565:49-56.
31. Wang Y*, et al.* (2018) Cofactor Biogenesis in Cysteamine Dioxygenase: C-F Bond Cleavage with Genetically Incorporated Unnatural Tyrosine. *Angew Chem Int Ed Engl* 57(27):8149-8153.
32. de Visser SP & Straganz GD (2009) Why do cysteine dioxygenase enzymes contain a 3-His ligand motif rather than a 2His/1Asp motif like most nonheme dioxygenases? *J Phys Chem A* 113(9):1835-1846.
33. Leitgeb S & Nidetzky B (2008) Structural and functional comparison of 2-His-1-carboxylate and 3-His metallocentres in non-haem iron(II)-dependent enzymes. *Biochem Soc Transact* 36(Pt 6):1180-1186.
34. 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.
35. 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.
36. Chowdhury R*, et al.* (2009) Structural basis for binding of hypoxia-inducible factor to the oxygen-sensing prolyl hydroxylases. *Structure* 17(7):981-989.
37. McDonough MA*, et al.* (2006) Cellular oxygen sensing: Crystal structure of hypoxia-inducible factor prolyl hydroxylase (PHD2). *Proc Natl Acad Sci USA* 103(26):9814-9819.
38. Sikharulidze I, Winter G, & Hall DR (2016) Big EP: Automated structure solution pipeline deployment at Diamond Light Source. *Acta Crystallogr A* 72:S193-S193.
39. Terwilliger TC*, et al.* (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr D* 65(6):582-601.
40. Vonrhein C, Blanc E, Roversi P, & Bricogne G (2007) Automated Structure Solution With autoSHARP. *Macromolecular Crystallography Protocols: Volume 2: Structure Determination*, ed Doublié S (Humana Press, Totowa, NJ), pp 215-230.
41. Adams PD*, et al.* (2011) The Phenix software for automated determination of macromolecular structures. *Methods* 55(1):94-106.
42. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D* 60(12 Part 1):2126-2132.
43. McCoy AJ*, et al.* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4):658-674.
44. Clough SJ & Bent AF (1998) Floral dip: a simplified method for Agrobacterium -mediated transformation of Arabidopsis thaliana. *Plant J* 16(6):735-743.

Figures Legends

Scheme 1. Plant Cysteine Oxidases catalyse dioxygenation of conserved cysteine residues at the N-terminus of Group VII Ethylene Response Factors. Both atoms of molecular O_2 (red) are incorporated into the product to generate Cys-sulfinic acid.

Figure 1. Views of crystal structures of AtPCO4, AtPCO5 and RnCDO. (**A**) AtPCO4 (teal), (**B**) AtPCO5 (light pink) and (**C**) RnCDO (wheat, PDB ID: 4IEZ (21)) proteins each consist of a conserved double stranded beta helix (DSBH, or cupin) core which binds an active site metal (gold and tan sphere for AtPCO and RnCDO structures, respectively) using a facial histidine triad (sticks).

Figure 2. Views of the active sites of AtPCO4_1, AtPCO4_2 and AtPCO5. Representative electron density is displayed around the active sites of (**A**) AtPCO4_1 (blue-green sticks), (**B**) AtPCO4_2 (cyan sticks) and (**C**) AtPCO5 (salmon sticks). A facial triad consisting of protein derived residues H98, H100 and H164 and 3 water molecules (red spheres, W1-3) coordinate to a metal ion (orange sphere). Note that His-10 (thin sticks) from the His_6-tag of a symmetry related molecule displaces W3 trans to H164 in AtPCO4_2. 2mFo-Fc contoured to 1.2σ for the protein (grey mesh) and mFo-Fc OMIT electron density contoured to 5σ for the metal (green mesh).

Figure 3. Comparison of the active sites of AtPCO4 and RnCDO. The number and distribution of chemically reactive amino acids is considerably different between (**A**) AtPCO4 and (**B**) RnCDO, with the catalytic triad of RnCDO (Ser153-His155-Tyr157) replaced in AtPCO4 with an overall less reactive triad (Ile174-Asp176-Leu178). AtPCO4 Val105 also replaces RnCDO Cys93. Other amino acids in the AtPCO4 active site which have a potential role in catalysis and are of the subject of further investigation in this study are highlighted. The inset in panel (**B**) shows a clearer view of the RnCDO Cys93-Tyr157 thioether crosslink.

Figure 4. Effect of active site mutations on AtPCO4 function *in vitro* and *in vivo*. A, Endpoint assays comparing the ability of recombinant AtPCO4 and variants thereof (0.4 µM) to catalyze oxidation of the N-terminal cysteine of a $RAP2_{2\text{-}15}$ peptide substrate (200 µM) over 30 minutes. Letters indicate statistically significant differences (1-way ANOVA followed by Tukey post-hoc test, n = 3). **B,** Time course assays comparing initial rates of recombinant AtPCO4 WT and variant activity (0.1 – 1 µM) towards the $RAP2_{2\text{-}15}$ peptide substrate (200 µM); data shown as mean ± SD (n=3). **C,** Relative luciferase activity (Fluc/Rluc) in *Saccharomyces cerevisiae* cells expressing C-DLOR reporter under aerobic and hypoxic conditions in the presence of PCO4 WT, PCO4 H164D and PCO4 D176N. Letters indicate statistically significant differences (1-way ANOVA followed by Holm-Sidak post-hoc test, n = 5). **D**, Expression of AtPCO4 (wt) restores the wild-type phenotype in *4pco A. thaliana* mutants; in contrast, expression of AtPCO4 (H164D) or AtPCO4 (D176N) fails to reconstitute the wild-type phenotype in *4pco* (scale bar = 1 cm). **E,** Relative mRNA level of hypoxia-inducible genes in wild-type and 4*pco* mutant plants expressing PCO4(wt), PCO4(H164D) or PCO4(D176N). Expression of *ADH*, *PDC1*, *LBD41*, *SAD6* and *Hb1* was significantly increased in *4pco* mutant plants and those expressing the inactivating PCO4 variants, compared to wild type and *4pco* plants expressing the wild type PCO4 enzyme. Red and blue dots represent data from independent transgenic lines. Letters distinguish statistically different genotypes (1-way ANOVA followed by Holm-Sidak post-hoc test for multiple comparisons for *ADH*, *PDC1*, *LBD41*, *SAD6* and *Hb1,* ANOVA on ranks and Tukey post-hoc test for transgenic *PCO4*).

Figures

Scheme 1.

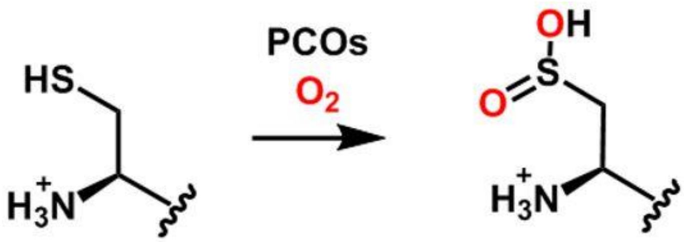


Figure 1.

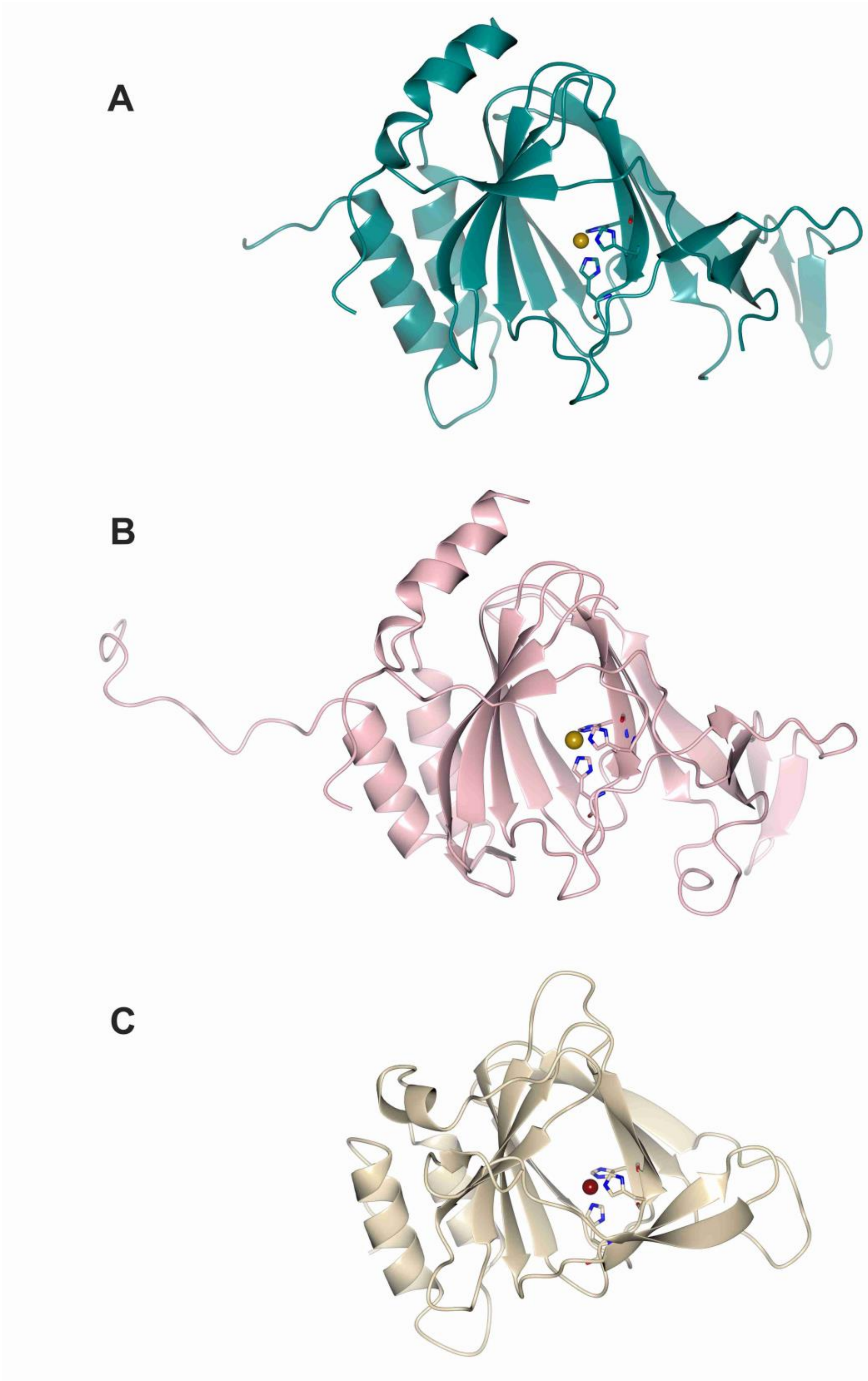


Figure 2

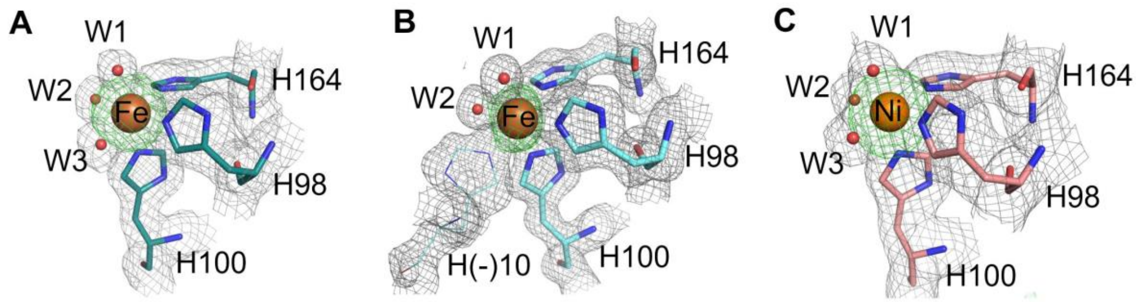


Figure 3

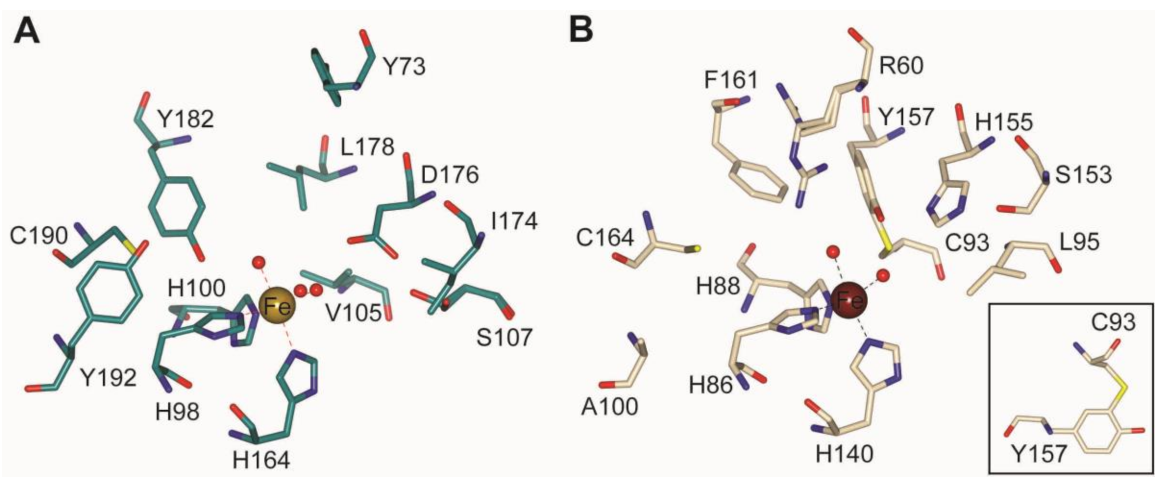
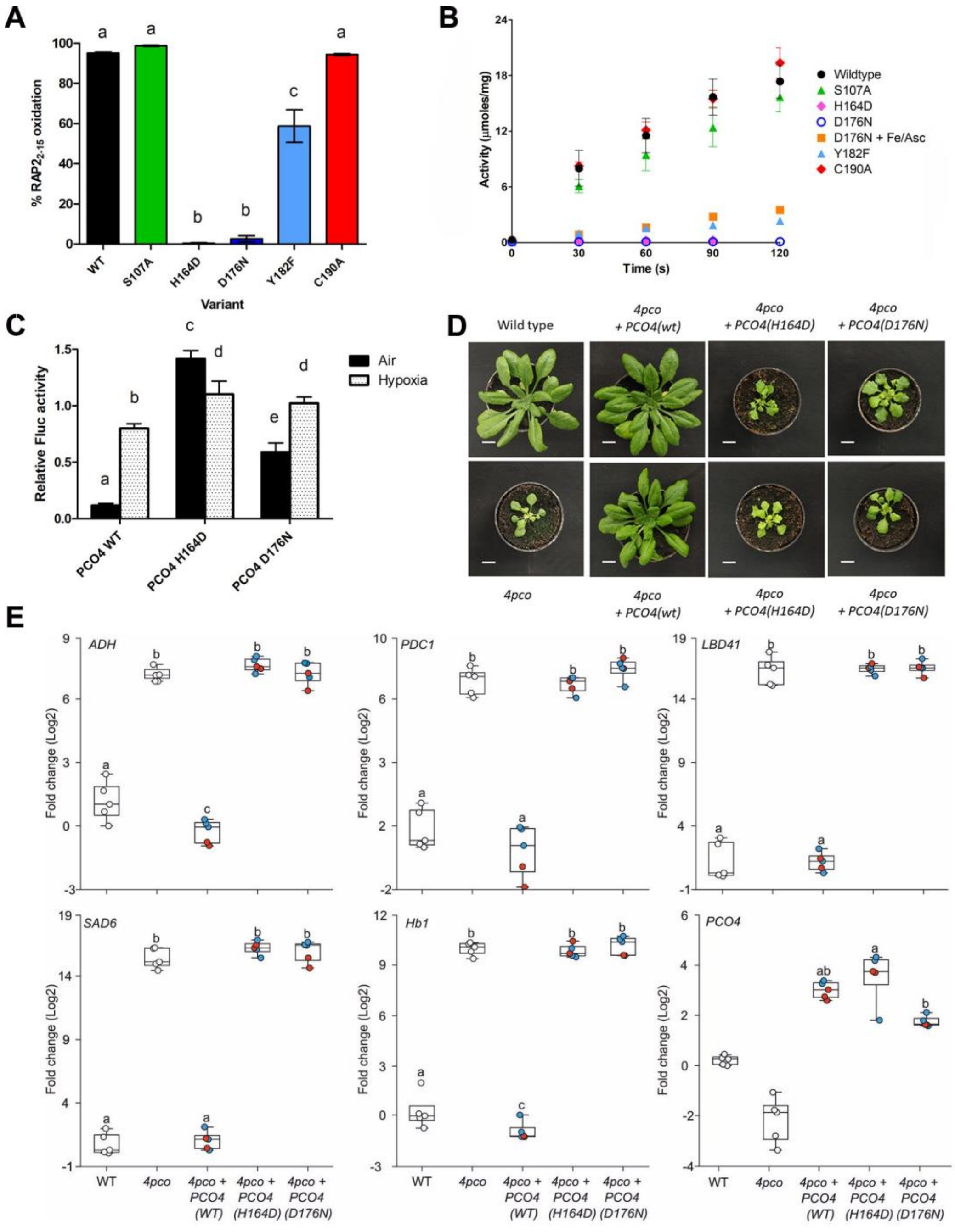


Figure 4


A
% RAP2₂₋₁₅ oxidation
Variant
WT
S107A
H164D
D176N
Y182F
C190A
B
Activity (µmoles/mg)
Time (s)
Wildtype
S107A
H164D
D176N
D176N + Fe/Asc
Y182F
C190A
C
Relative Fluc activity
Air
Hypoxia
PCO4 WT
PCO4 H164D
PCO4 D176N
D
Wild type
4pco + PCO4(wt)
4pco + PCO4(H164D)
4pco + PCO4(D176N)
4pco
E
ADH
PDC1
LBD41
SAD6
Hb1
PCO4
Fold change (Log2)
WT
4pco
4pco + PCO4 (WT)
4pco + PCO4 (H164D)
4pco + PCO4 (D176N)