Plant Cysteine Oxidases are Dioxygenases that Directly Enable Arginyl Transferase-Catalyzed Arginylation of N-End Rule Targets

Mark D. White[1], Maria Klecker[2,3], Richard J. Hopkinson[1], Daan Weits[4], Carolin Mueller[5,6], Christin Naumann[2,3], Rebecca O'Neill[1], James Wickens[1], Tom N. Grossmann[5,6], Nico Dissmeyer[2,3,*], Emily Flashman[1,*]

[1] Chemistry Research Laboratory, University of Oxford, 12 Mansfield Road, Oxford, OX1 3TA, United Kingdom
[2] Independent Junior Research Group on Protein Recognition and Degradation, Leibniz Institute of Plant Biochemistry (IPB), Weinberg 3, D-06120 Halle (Saale), Germany
[3] ScienceCampus Halle – Plant-based Bioeconomy, Betty-Heimann-Str. 3, D-06120 Halle (Saale), Germany
[4] Institute of Biology I, RWTH Aachen University, Worringerweg 1, D-52074 Aachen, Germany
[5] Chemical Genomics Centre of the Max Planck Society, Department of Chemistry and Chemical Biology, Technische Universität Dortmund, Otto-Hahn-Str. 15, D-44227 Dortmund, Germany
[6] VU University Amsterdam, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands

* Corresponding authors: emily.flashman@chem.ox.ac.uk, +44(0)1865 275920
nico.dissmeyer@ipb-halle.de, ++49 (0) 345 5582 1710

Keywords
cysteine dioxygenase; ethylene response factors; plant hypoxia; N-end rule pathway; arginyl transferase; E3 ubiquitin ligase; proteostasis

Abbreviations
PCO, plant cysteine oxidase; ATE, arginyl tRNA transferase; ERF-VII, group VII ETHYLENE RESPONSE FACTOR; 2OG, 2-oxoglutarate; NMR, nuclear magnetic resonance; Met, methionine; NME, N-terminal Met excision; Nt, N-terminal; NO, nitric oxide; HIF, hypoxia-inducible factor; PHD, prolyl hydroxylase; MALDI-MS, matrix-assisted laser desorption/ionization-mass spectrometry; LC-MS, liquid chromatography-mass spectrometry; HRE, HYPOXIA RESPONSIVE ERF; RAP, RELATED TO APETALA2; EBP, ETHYLENE RESPONSE FACTOR 72; CDO, cysteine dioxygenase; MAP, Met-aminopeptidase.

Abstract

Crop yield loss due to flooding is a threat to food security. Submergence-induced hypoxia in plants results in stabilisation of group VII ETHYLENE RESPONSE FACTORS (ERF-VIIs), which aid survival under the adverse conditions. ERF-VII stability is controlled by the N-end rule pathway, which proposes that ERF-VII N-terminal cysteine oxidation in normoxia enables arginylation followed by proteasomal degradation. The PLANT CYSTEINE OXIDASEs (PCOs) have been identified as catalysing this oxidation. ERF-VII stabilisation in hypoxia presumably arises from reduced PCO activity. We directly demonstrate that PCO dioxygenase activity produces Cys-sulfinic acid at the N-terminus of an ERF-VII peptide, which then undergoes efficient arginylation by an arginyl transferase (ATE1). This is the first molecular evidence showing N-terminal cysteine oxidation and arginylation by N-end rule pathway components, and the first ATE1 substrate in plants. The PCOs and ATE1 may be viable intervention targets to stabilise N-end rule substrates, including ERF-VIIs to enhance submergence tolerance in agronomy.

Introduction

All aerobic organisms require homeostatic mechanisms to ensure O_2 supply and demand are balanced. When supply is reduced (hypoxia), a hypoxic response is required to decrease demand and/or improve supply. In animals, this well characterized response is mediated by the Hypoxia-Inducible transcription Factor (HIF), which upregulates genes encoding for vascular endothelial growth factor, erythropoietin and glycolytic enzymes amongst many others.[1-3] Hypoxia in plants is typically a consequence of reduced O_2 diffusion under conditions of waterlogging or submergence, or inside of organs such as seeds, embryos, or floral meristems in buds where the various external cell layers act as diffusion barriers. Although plants can survive temporary periods of hypoxia, flooding negatively impacts on plant growth, and if sustained it can result in plant damage or death[4]. This has a major impact on crop yield, for example, flooding resulted in crop loss costing $3 billion in the U.S. in 2011.[5] As climate change results in increased severe weather events including flooding[4], strategies to address crop survival under hypoxic stress are needed to meet the needs of a growing worldwide population.

The response to hypoxia in rice, Arabidopsis, and barley is known to be mediated by the group VII ETHYLENE RESPONSE FACTORs (ERF-VIIs).[6-11] It has been found that these transcription factors promote the expression of core hypoxia-responsive genes, including those encoding alcohol dehydrogenase and pyruvate decarboxylase that facilitate anaerobic metabolism.[12,13] Crucially, it was shown, initially in Arabidopsis, that the stability of the ERF-VIIs is regulated in an O_2-dependent manner via the Arg/Cys branch of the N-end rule pathway, which directs proteins for proteasomal degradation depending on the identity of their N-terminal amino acid.[14-16] Thus, a connection between O_2 availability and the plant

87 hypoxic response was identified.[11,17,18] The Arabidopsis ERF-VIIs are translated with the
88 conserved N-terminal motif MCGGAI/VSDY/F[4] and co-translational N-terminal methionine
89 excision, catalyzed by Met amino peptidases (MAPs)[19,20], leaves an exposed N-terminal Cys
90 which is susceptible to oxidation.[14-16] N-terminal oxidized Cys residues are then proposed to
91 render the ERF-VII N-termini substrates for arginyl tRNA transferase (ATE)-catalyzed
92 arginylation. The subsequent Nt-Arg-ERF-VIIs are candidates for ubiquitination by the E3
93 ligase PROTEOLYSIS6 (PRT6)[21] which promotes targeted degradation via the 26S
94 proteasome. It has also been shown that degradation of ERF-VIIs by the N-end rule pathway
95 is influenced by NO, and that the ERF-VIIs play a key role in plant NO-mediated stress
96 responses.[22,23] The relationship between O_2- and NO-mediated ERF-VII modification remains
97 to be fully elucidated.
98 The plant hypoxic response mimics the equivalent well-characterized regulatory system in
99 animals, whereby adaptation to hypoxia is mediated by HIF. In normoxic conditions, HIF is
100 hydroxylated at specific prolyl residues targeting it for binding to the von Hippel-Lindau
101 tumour suppressor protein (pVHL), the recognition component of the E3-ubiquitin ligase
102 complex, which results in HIF ubiquitination and proteasomal degradation.[1,3] Thus, while not
103 substrates for the N-end rule pathway of protein degradation, HIF levels are regulated by
104 post-translational modification resulting in ubiquitination, in a manner that is sensitive to
105 hypoxia. HIF prolyl hydroxylation is catalyzed by O_2-dependent enzymes, the HIF prolyl
106 hydroxylases (PHDs 1-3),[2] which are highly sensitive to O_2 availability.[24,] These O_2-sensing
107 enzymes are thus the direct link between O_2 availability and the hypoxic response.[26,27]
108
109 Crucially, a family of five enzymes, the PLANT CYSTEINE OXIDASEs (PCO1-5) were
110 identified in Arabidopsis[28] that were reported to catalyze the O_2-dependent reaction in the
111 plant hypoxic response, specifically the oxidation of the conserved Cys residue at the N-

terminus of the Arabidopsis ERF-VIIs, RAP2.2, RAP2.12, RAP2.3, HRE1 and HRE2. It was found that overexpression of PCO1 and 2 *in planta* specifically led to depleted RAP2.12 protein levels and reduced submergence tolerance, while *pco1 pco2* T-DNA insertion mutants accumulated RAP2.12 protein. Isolated recombinant PCO1 and PCO2 were shown to consume O_2 in the presence of pentameric peptides CGGAI corresponding to the N-termini of various ERF-VIIs **(Supplementary Table 1[28]).** The identification of these enzymes indicates that the hypoxic response in plants is enzymatically regulated,[28] potentially in a similar manner to the regulation of the hypoxic response in animals by the HIF hydroxylases. The PCOs may therefore act as plant O_2 sensors.

We sought to provide molecular evidence that the PCOs catalyze the oxidation step in ERF-VII proteasomal targeting and to determine if this step is required for further molecular priming by arginylation. Using mass spectrometry and NMR techniques we confirm that PCO1 and also PCO4 catalyze dioxygenation of the N-terminal Cys of Arabidopsis ERF-VII peptide sequences to Cys-sulfinic acid ($CysO_2$), rather than the fully oxidized Cys-sulfonic acid ($CysO_3$) (**Supplementary Figure 1**). This oxidation directly incorporates molecular O_2. To our knowledge, these are the first described enzymes that catalyze cysteinyl oxidation, as well as being the first described cysteine dioxygenases in plants. We then verify that the product of the PCO-catalyzed reactions is a substrate for the arginyl tRNA transferase ATE1, demonstrating that PCO activity is relevant and sufficient for the subsequent step of molecular recognition and modification according to the N-end rule pathway. Overall, we thus define the PCOs as plant cysteinyl dioxygenases and ATE1 as an active arginyl transferase, establishing for the first time a direct link between molecular O_2, PCO catalysis and ATE1 recognition and modification of N-end rule substrates.

137 **Results**

138

139 **PCOs catalyze modification of RAP2$_{2-11}$ in an O$_2$-dependent manner**

140 N-terminally hexahistidine-tagged recombinant PCO1 and 4 were purified to ~90% purity, as
141 judged by SDS-PAGE (**Supplementary Figure 2a**). Protein identity was confirmed by
142 comparison of observed and predicted mass by LC-MS (PCO1 predicted mass 36,510 Da,
143 observed mass 36,513 Da; PCO4 predicted mass 30,680 Da, observed mass 30,681 Da,
144 **Supplementary Figure 2b**). The activity of the purified PCO1 and PCO4 was tested towards
145 a synthetic 10-mer peptide corresponding to the methionine excised N-termini of the ERF-
146 VIIs RAP2.2, RAP2.12, and HRE2 (H_2N-CGGAIISDFI-COOH, hereafter termed RAP2$_{2-11}$
147 **Supplementary Table 1**). Assays comprising RAP$_{2-11}$ at 100 μM in the presence or absence
148 of PCO1 or PCO4 at 0.5 μM underwent aerobic or anaerobic coincubation for 30 minutes at
149 30°C prior to analysis of the peptide by matrix-assisted laser desorption/ionization-mass
150 spectrometry (MALDI-MS, **Figure 1a,b**). Only under aerobic conditions and in the presence
151 of PCO1 or PCO4, the spectra revealed the appearance of two species with mass increases of
152 +32 Da and +48 Da, corresponding to two or three added O atoms, highlighting an O_2-
153 dependent reaction for PCOs 1 and 4 (**Figure 1b**), as previously shown for PCOs 1 and 2.[28]
154 These mass shifts were deemed to be consistent with enzymatic formation of Cys-sulfinic
155 ($CysO_2$, +32 Da) and Cys-sulfonic acid ($CysO_3$, +48 Da; **Supplementary Figure 1**), both
156 proposed to be Arg transferase substrates in the Arg/Cys branch of N-end rule mediated
157 protein degradation.[14-16]

158

159 **PCOs are dioxygenases catalyzing the incorporation of both atoms of O$_2$ into RAP2$_{2-11}$**

160 To ascertain whether the PCOs function as dioxygenases, we sought to verify the source of
161 the O atoms in the oxidized RAP2$_{2-11}$ by conducting assays in the presence of $^{18}O_2$ as the

162 cosubstrate or $H_2{}^{18}O$ as the solvent. To probe O_2 as the source of O atoms in the product,
163 anaerobic solutions of $RAP2_{2\text{-}11}$ were prepared in sealed vials before addition of PCO4 using
164 a gas-tight syringe. The vials were then purged with $^{16}O_2$ or $^{18}O_2$ and the reactions were
165 allowed to proceed at 30°C for a subsequent 20 minutes. Upon analysis by MALDI-MS, the
166 mass of the products revealed that molecular O_2 was incorporated into the Cys-sulfinic acid
167 product, but not the Cys-sulfonic acid product (**Figure 2a**). The Cys-sulfinic acid product has
168 a mass of +32 Da in the presence of $^{16}O_2$ and +36 Da in the presence of $^{18}O_2$, demonstrating
169 addition of two ^{18}O atoms and indicating that O_2 is the source of O atoms in this product. The
170 Cys-sulfonic acid product has a mass of +52 Da in the presence of $^{18}O_2$, indicating a third ^{18}O
171 atom has not been incorporated into this product. To probe whether the source of the
172 additional mass in the apparent Cys-sulfonic acid product was an O atom derived from water,
173 an equivalent reaction was carried out under aerobic conditions in the presence of $H_2{}^{18}O$
174 ($H_2{}^{18}O$:H_2O in a 3:1 ratio). No additional mass was observed in the peak corresponding to the
175 Cys-sulfonic acid, raising the possibility that the +48 Da species observed by MALDI-MS is
176 not enzymatically formed. Importantly, no additional mass was observed in the peak
177 corresponding to Cys-sulfinic acid, confirming that this species is a product of a reaction
178 where molecular O_2 is a substrate (**Figure 2b**).
179
180 To further investigate whether the PCO-catalyzed product species observed at +48 Da is
181 enzymatically produced or an artifact of the MALDI-MS analysis method, we turned to liquid
182 chromatography-mass spectrometry (LC-MS) to analyze the products of the PCO-catalyzed
183 reactions. Under these conditions, only peptidic product with a mass increase of +32 Da was
184 observed after incubation with both PCO1 and PCO4, corresponding to the incorporation of
185 two O atoms and the formation of Cys-sulfinic acid (**Figure 2c**), consistent with the products
186 observed using $^{18}O_2$ and $H_2{}^{18}O$ (**Figure 2a,b**). No product was observed with a mass

187 corresponding to Cys-sulfonic acid, which suggested that the +48 Da product detected by
188 MALDI-MS was indeed an artifact. When combined with the observation that significant
189 quantities of Cys-sulfonic acid were not seen in no-enzyme or anaerobic controls (**Figure 1**),
190 it was hypothesized that the Cys-sulfinic acid product of the PCO-catalyzed reaction is non-
191 enzymatically converted to Cys-sulfonic acid during MALDI-MS analysis, potentially as a
192 result of laser exposure. Upon subjecting the products of PCO1 and 4 turnover to MALDI-
193 MS analysis with increasing laser intensity, a direct correlation between laser intensity and
194 the ratio of Cys-sulfonic acid:Cys-sulfinic acid product was observed (**Supplementary**
195 **Figure 3a**). Of note, significant levels of laser induced formation of +32 and +48 Da species
196 upon analysis of unmodified peptide were not observed (**Supplementary Figure 3b**).
197 Together, these results confirm that the +48 Da species observed following incubation of the
198 PCOs with $RAP2_{2\text{-}11}$ are a product of Cys-sulfinic acid exposure to the MALDI-MS laser, and
199 not a product of the PCO-catalyzed reaction. Overall, these data demonstrate that the PCOs
200 are dioxygenase enzymes, similar to the mammalian and bacterial cysteine dioxygenases
201 (CDOs) to which they show sequence homology.[28]
202

203 **PCOs catalyze oxidation of N-terminal Cys of $RAP2_{2\text{-}11}$ to form Cys-sulfinic acid**

204 Recombinant PCO1 and PCO2 were reported to consume O_2 in the presence of pentameric
205 CGGAI peptides corresponding to the methionine-excised N-terminus of the Arabidopsis
206 ERF-VIIs.[28] To definitively verify that the N-terminal cysteinyl residue of $RAP2_{2\text{-}11}$ is indeed
207 the target for the PCO-catalyzed +32 Da modifications, we conducted LC-MS/MS analyses
208 on the reaction products. Fragmentation of $RAP2_{2\text{-}11}$ that had been incubated in the presence
209 and absence of PCO1 and PCO4 revealed b- and y-ion series consistent with oxidation of the
210 N-terminal Cys residue (**Figure 3a**), confirming that the PCOs 1 and 4 act as cysteinyl
211 dioxygenases.

212 As a final confirmation of the nature of the reaction catalyzed by PCO1 and PCO4, their
213 activity was monitored using ^{1}H-NMR. Reactions were initiated by adding 5 μM enzyme to
214 500 μM RAP2$_{2-11}$ (in the presence of 10% D_2O) and products of the reaction were analysed
215 using a 600 MHz NMR spectrometer. In the presence of both PCO1 and PCO4, modification
216 to the cysteinyl residues was observed, as exemplified by the disappearance of the ^{1}H-
217 resonance corresponding to the β-cysteinyl protons (at δ_H 2.88 ppm) and the emergence of a
218 new ^{1}H-resonance at δ_H 2.67 ppm (**Figure 3b**). The chemical shift of the new resonance is
219 similar to that observed for L-Cys conversion to L-Cys-sulfinic acid by mouse CDO,[29] and
220 also to the chemical shift of an L-Cys-sulfinic acid standard measured under equivalent
221 conditions to the PCO assays (**Supplementary Figure 4**). Therefore, the resonance shift
222 observed upon PCO1/4 reaction was assigned to the β-protons of L-Cys-sulfinic acid. Overall
223 these results provide confirmation at the molecular level that Arabidopsis PCOs 1 and 4 act
224 as plant cysteinyl dioxygenases, catalyzing incorporation of O_2 into N-terminal Cys residues
225 on a RAP2 peptide to form Cys-sulfinic acid.

226

227 **ATE1 arginylates acidic N-termini including Cys-sulfinic acid**

228 We next sought to confirm that the PCO-catalyzed Cys-oxidation to Cys-sulfinic acid renders
229 a RAP2 peptide capable of and sufficient for onward modification by ATE1, as proposed by
230 the N-end rule pathway. ATE1 has been suggested to be responsible for transfer of ^{3}H-
231 arginine to bovine α-lactalbumine in highly purified plant extracts *in vitro*[30] and RAP2.12
232 stabilization in *ate1 ate2* double null mutant plant lines implicates ATE1 as an ERF-VII-
233 targeting arginyl transferase *in vivo.*[17,18] To this end, we produced recombinant hexahistidine-
234 tagged Arabidopsis ATE1 (**Supplementary Figure 5**) for use in an arginylation assay which
235 detects incorporation of radiolabeled ^{14}C-Arg into biotinylated peptides. C-terminally
236 biotinylated RAP2$_{2-13}$ peptides (H_2N-XGGAIISDFIPP(PEG)K(biotin)-NH_2) where the N-

237 terminal residue, X, constitutes Gly, Asp, Cys or Cys-sulfonic acid were subjected to the
238 arginylation assay in the presence or absence of PCO1/4 (**Figure 4a**). Peptide with an N-
239 terminal Gly did not accept Arg, while an N-terminal Asp did accept Arg, independent of the
240 presence of PCO1 or 4. A peptide comprising an N-terminal Cys-sulfonic acid was also
241 shown to be a substrate for ATE1, again independent of the presence of PCO1 or 4.

242 Crucially, $RAP2_{2-13}$ with an N-terminal Cys in the absence of PCO1/4 was not an acceptor of
243 arginine transfer by ATE1, yet when either PCO1 or PCO4 was incorporated in the reaction,
244 significant ATE1 transferase activity was observed (**Figure 4a**).

245 To confirm that the increased detection of radiolabel corresponded to arginyl incorporation at
246 the N-termini of the peptides, the experiment was repeated using non-radiolabeled arginine in
247 the presence and absence of PCO4 and ATE1, and peptide products subjected to LC-MS
248 analysis (**Figure 4c**). As with $RAP2_{2-11}$ (**Figure 2c**), the Cys-initiated $RAP2_{2-13}$ peptide
249 displayed a +32 Da increase in mass upon incubation with PCO4 only (**Figure 4c**, red
250 spectrum). Importantly, following incubation of Cys-initiated $RAP2_{2-13}$ with both PCO4 and
251 ATE1, a mass increase equivalent to oxidation coupled to arginylation (+188 Da) was
252 observed (**Figure 4c**, blue spectrum). Subsequent tandem MS analysis of these product ions
253 revealed fragmentation species consistent with the assumption that oxidation and sequential
254 arginylation occur at the N-terminus of PCO4- and ATE1-treated peptides (**Figure 4d**, blue
255 spectrum), strongly suggesting that the PCO-oxidized N-termini of ERF-VIIs are rendered N-
256 degrons via additional arginylation (**Figure 4b**).

257 A +12 Da mass increase was observed in products of control assays lacking PCO4 (**Figure**
258 **4c, d**; purple spectra). This appeared to be related to prolonged incubation in the presence of
259 HEPES and DTT as used in the arginylation assay buffer: The +12 Da modification was not
260 observed if the peptide was dissolved in H_2O (**Figure 4c**, black spectrum) or if incubated in

261 HEPES and DTT for just 1 hour, but was observed when the peptide was incubated in
262 HEPES and DTT overnight (**Supplementary Figure 6**). It is proposed that trace levels of
263 contaminating formaldehyde react with free Nt-Cys residues to form thiazolidine N-termini.[31]

264 These results are in line with proposed arginylation requirements for the Arg/Cys branch of
265 the N-end rule pathway[14-16] including the known Cys-initiated arginylation targets from
266 mammals.[32-34] Importantly, these results demonstrate for the first time Arg transfer mediated
267 by ATE1 dependent on the N-terminal residue of its substrate, and also that both Cys-sulfinic
268 acid (the product of PCO-catalysis) and Cys-sulfonic acid can act as substrates for ATE1.
269 The arginylation observed with PCO-catalyzed Cys-sulfinic acid in particular supports the
270 assumption that N-terminal residues sterically and electrostatically resembling Asp or Glu
271 can serve as Arg acceptors in reactions catalyzed by ATEs,[34] and also confirms the
272 importance of the PCOs as a connection between the stability of their ERF-VII substrates and
273 O_2 availability (**Figure 4b**).

274

275 **Discussion**

276

277 The PCOs were identified in *Arabidopsis thaliana* as a set of five enzymes suggested to
278 catalyze oxidation of N-terminal cysteine residues in ERF-VII transcription factors and
279 oxidation was demonstrated in short peptides corresponding to their N-termini[28]. This
280 oxidation was associated with destabilization of the ERF-VIIs, presumably by rendering them
281 substrates of the Arg/Cys branch of the N-end rule pathway.[14,16] Under conditions of
282 sufficient O_2 availability, ERF-VII protein levels are decreased, while under hypoxic
283 conditions, such as those encountered upon plant submergence or in the context of organ
284 development, ERF-VII levels remain high.[17,18] Importantly, the ERF-VII transcription factors
285 are known to upregulate genes which allow plants to cope with or respond to submergence.[13]

286 The PCOs are proposed to act as potential O_2 sensors involved in regulating the plant hypoxic
287 response.[28]
288
289 We sought to biochemically confirm the role of the PCOs in the plant hypoxic response, and
290 present here mass spectrometry and NMR data that clearly demonstrate that two enzymes
291 from this family, PCOs 1 and 4, are dioxygenases that catalyze direct incorporation of O_2 into
292 $RAP2_{2\text{-}11}$ peptides to form Cys-sulfinic acid. Their direct use of O_2 supports the proposal that
293 these enzymes are plant O_2 sensors.[28] A relationship has been demonstrated between O_2
294 concentration and PCO activity,[28] but it will be of interest to perform detailed kinetic
295 characterization of these enzymes to ascertain their level of sensitivity to O_2 availability, in
296 particular whether their O_2-sensitivity is similar to that of the HIF hydroxylases in
297 animals.[24,25] Although there is functional homology between the PCOs and the HIF
298 hydroxylases, they are apparently mechanistically divergent: The PCOs show sequence
299 homology to the Fe(II)-dependent CDO family of enzymes which do not require an external
300 electron donor for O_2 activation,[28,35] while the HIF hydroxylases are Fe(II)/2OG-dependent
301 oxygenases. Of note, the PCOs are the first identified CDOs in plants. Further, in contrast to
302 the reactions of mammalian and bacterial CDOs which oxidize free L-Cys, the PCOs are also,
303 to our knowledge, the first identified cysteinyl (as opposed to free L-Cys) dioxygenases.
304
305 According to the Arg/Cys branch of the N-end rule pathway, N-terminal Cys oxidation is
306 proposed to enable successive arginylation by ATE1 to render proteins as N-degrons. We
307 therefore sought to demonstrate that PCO-catalyzed ERF-VII N-terminal Cys oxidation to
308 Cys-sulfinic acid promotes arginylation by ATE1. The arginylation assay and mass
309 spectrometry results we present demonstrate that the PCO-catalyzed dioxygenation reaction
310 is sufficient to trigger N-terminal arginylation of RAP2s by ATE1, thus likely rendering

311 ERF-VIIs (at least those comprising the tested N-terminal sequence) as N-degrons, i.e.
312 recognition by PRT6 and other potential E3 ubiquitin ligases, polyubiquitination and possibly
313 transfer to the 26S proteasome for proteolysis.[14-16] Collectively therefore, we present the first
314 molecular evidence confirming the Cys-oxidation and subsequent arginylation steps of the
315 Arg/Cys branch of the N-end rule pathway.[32-34] We also confirm that ATE1 is able to
316 selectively arginylate, as predicted,[34] acidic N-terminal residues of plant substrates, including
317 Cys-sulfonic acid.
318
319 ATE1 is capable of arginylating proteins at both acidic N-termini and midchains, though N-
320 terminal acidic residues are reported to be arginylated via canonical peptide bonds.[36] While
321 reports of midchain arginylation are relatively common,[36,37,38] to date only one physiological
322 and two *in vitro* substrates for the Arg/Cys branch of the N-end rule pathway have been
323 characterized, namely mammalian regulator of G protein signaling (RGS) 4, and RGS5 and
324 10 respectively,[39] where Nt-Cys oxidation was described (to either Cys-sulfinic or Cys-
325 sulfonic acid) as was Nt-Cys arginylation.[32,34] The first three non Cys-branch N-end rule
326 arginylation targets were only very recently identified in mammalian cell culture:[40] the
327 molecular chaperone BiP (GRP78 and HSPA5, heat shock 70 kDa protein 5; Glu-initiated),
328 calreticulin (CRT, Glu-initiated), and the oxidoreductase protein disulphide isomerase (PDI,
329 Asp-initiated). Similarly, data regarding the molecular requirements of plant ATEs is limited.
330 Developmental functions of the single homolog ATE1 in the moss *Physcomitrella patens*
331 were recently described[41] and interaction partners of the enzyme found and arginylated
332 peptides immunologically-detected.[42] Previously, Arg transferase function of Arabidopsis
333 ATE1/2 has been shown using an assay detecting conjugation of ^{3}H-Arg to bovine α-
334 lactalbumin (bearing an N-terminal Glu) in the presence of plant extracts from wild type
335 Arabidopsis, *ate1* and *ate2* single mutants but not from *ate1 ate2* double mutant seedlings.[30]

336 Therefore, the results we present here demonstrate for the first time Arg transferase activity
337 of a plant ATE towards known plant N-end rule substrates.
338
339 Interestingly, as well as O_2, nitric oxide was identified as an RGS oxidizing agent, suggesting
340 a potential role of S-nitrosylation in the Arg/Cys branch of the N-end rule pathway, albeit
341 non-enzymatically controlled.[33] It has also been reported *in planta* that both NO and O_2 are
342 required for ERF-VII degradation, potentially at the Cys oxidation step.[22,23] Although in N-
343 end rule-mediated RGS4/5 degradation it has been proposed that Cys-nitrosylation precedes
344 Cys-oxidation (also currently considered a non-enzymatic process), we find that under the
345 conditions used, the PCO1/4-catalyzed reaction does not require either prior cys-nitrosylation
346 or exogenous NO to proceed efficiently. We cannot rule out that NO plays a role in formation
347 of a Cys-sulfonic acid product, which is also a substrate for ATE1 as shown in our Arg
348 transfer experiments. Alternatively, NO may have a role in ERF-VII degradation *in vivo* via
349 non-enzymatic oxidation or a secondary mechanism. The manner in which NO contributes to
350 Arg/Cys branch of the N-end rule pathway therefore remains to be elucidated.
351
352 ERF-VII stabilization has been shown to result in improved submergence tolerance, elegantly
353 demonstrated in barley by mutation of the candidate E3-ubiquitin ligase *PRT6,*[11] but also in
354 rice containing the *Sub1A* gene; SUB1A is an apparently stable ERF-VII that confers
355 particular flood tolerance in certain rare varieties of rice.[9,17] Overexpression of *Sub1A* in more
356 commonly grown rice varieties has resulted in a 45% increase in yield relative to *sub1a*
357 mutant lines after exposure to flooding.[43] If ERF-VII stabilization is indeed a proficient
358 mechanism for enhancing flood tolerance, then manipulation of PCO or ATE activity may be
359 an efficient and effective point of intervention. This work presents molecular validation of
360 their function, providing the basis for future targeted chemical/genetic inhibition of their

361 activity, but also highlighting genetic strategies for breeding via introgression of variants of
362 N-end rule pathway components or introduction of alleles of enzymatic components of the N-
363 end rule pathway from non-crop species into crops. Any of these strategies has the potential
364 to result in stabilized ERF-VII levels and therefore increase stress resistance and may
365 therefore help to address food security challenges.

366

367

368 **Online Methods**

369

370 **Peptide Synthesis, Purification and Characterization**

371 All reagents used were purchased from Sigma-Aldrich unless otherwise stated. The 10-mer
372 $RAP2_{2\text{-}11}$ peptide (H_2N-CGGAIISDFI-COOH) was purchased from GL Biochem (Shanghai)
373 Ltd, China (**Supplementary Table 1**). The sequence of the 12-mer peptides used in the
374 coupled oxidation-arginylation assay is derived from RAP2.2 and RAP2.12 (H_2N-X-
375 GGAIISDFIPP(PEG)K(biotin)-NH_2) and synthesized by Fmoc-based solid-phase peptide
376 synthesis (SPPS) on NovaSyn®TGR resin (Merck KGaA, **Supplementary Table 2**). Fmoc
377 protected amino acids (Iris Biotech GmbH) were coupled using 4 equivalents of (eq) of the
378 amino acid according relative to initial resin loading. 4 eq amino acid was mixed with 4 eq O-
379 (6-chlorobenzotriazol-1-yl)-N,N,N′,N′-tetramethyluronium hexafluorophosphate (HCTU) and
380 8 eq N,N-diisopropylethylamine (DIPEA; Santa Cruz Biotechnology, sc-293894) and added
381 to the resin for 1 h. In a second coupling, the resin was treated with 4 eq of the Fmoc-
382 protected amino acid mixed with 4 eq benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium
383 hexafluoro-phosphate (PyBOP) and 8 eq 4-methylmorpholine (NMM) for 1 h. After double
384 coupling a capping step to block free amines was performed using acetanhydride and DIPEA
385 in N-methyl-2-pyrrolidinone (NMP) (1:1:10) for 5 min. The *C*-terminal Fmoc-Lys(biotin)-

386 OH, the 8-(9-fluorenylmethyloxycarbonyl-amino)-3.6-dioxaoctanoic acid (PEG) linker and
387 the different Fmoc protected N-terminal amino acids were coupled manually. The remaining
388 peptide sequence was assembled using an automated synthesizer (Syro II, MultiSynTech
389 GmbH). Fmoc deprotection was performed using 20 % piperidine in dimethylformamide
390 (DMF) for 5 min, twice. After each step the resin was washed 5 times with DMF, DCM and
391 DMF, respectively. Final cleavage was performed with 94 % trifluoroacetic acid (TFA),
392 2.5 % 1,2-ethanedithiole (EDT) and 1 % triisopropylsilane (TIPS) in aqueous solution for
393 2 h, twice. The cleavage solutions were combined and peptides were precipitated with diethyl
394 ether (Et_2O) at -20°C for 30 min. Peptides were solved in water/acetonitrile (ACN) 7:3 and
395 purified by reversed-phase high performance liquid chromatography (HPLC; Nucleodur C18
396 culumn; 10×125 mm, 110 Å, 5 µm particle size; Macherey-Nagel) using a flow rate of
397 6 $ml \cdot min^{-1}$ (A: ACN with 1 % TFA, B: water with 1 % TFA). Obtained pure fractions were
398 pooled and lyophilized. Peptide characterization was performed by analytical HPLC (1260
399 Infinity, Agilent Technology; flow rate of 1 $ml \cdot min^{-1}$, A: ACN with 1 % TFA, B: water with
400 1 % TFA) coupled with a mass spectrometer (6120 Quadrupole LC/MS, Agilent Technology)
401 using electro spray ionization (Agilent Eclipse XDB-C18 culumn, 4.6×150 mm, 5 µm
402 particle size). Analytical HPLC chromatograms were recorded at 210 nm (**Supplementary**
403 **Figure 7)**. Quantification was performed by HPLC-based comparison (chromatogram at 210
404 nm) with a reference peptide (**Supplementary Table 2**).

405

406 **Preparation of Recombinant Proteins**

407 Arabidopsis PCO1 and PCO4 sequences in pDEST17 bacterial expression vectors
408 (Invitrogen) were kindly provided by F. Licausi and J. van Dongen.[28] Plasmids were
409 transformed into BL21(DE3) *Escherichia coli* cells, and expression of recombinant protein
410 carrying an N-terminal hexahistidine tag was induced with 0.5 mM IPTG and subsequent

411 growth at 18°C for 18 hours. Harvested cells were lysed by sonication and proteins purified
412 using Ni^{++} affinity chromatography, before buffer exchange into 250 mM NaCl/50 mM Tris-
413 HCl (pH 7.5). Analysis by SDS-PAGE and denaturing liquid-chromatography mass
414 spectrometry (LC-MS) showed proteins with more than 90% purity and predicted molecular
415 weights.

416 The coding sequence of Arabidopsis ATE1 was cloned according to gene annotations at
417 TAIR (www.arabidopsis.org) from cDNA. The sequence was flanked by an N-terminal
418 tobacco etch virus (TEV) recognition sequence for facilitated downstream purification ("tev":
419 ENLYFQ-X) using the primers ate1_tev_ss (5'-
420 GCTTAGAGAATCTTTATTTTCAGGGGATGTCTTTGAAAAACGATGCGAGT-3') and
421 ate1_as (5'-
422 GGGGACCACTTTGTACAAGAAAGCTGGGTATCAGTTGATTTCATACACCATTCTC
423 TC-3'). A second PCR using the primers adapter (5'-
424 GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGAGAATCTTTATTTTCAGGGG-3')
425 and ate1_as was performed to amplify the construct to use it in a BP reaction for cloning into
426 pDONR201 (Invitrogen) followed by an LR reaction into the vector pDEST17 (Invitrogen).
427 The N-terminal hexahistidine fusion was expressed in BL21-CodonPlus (DE3)-RIL
428 *Escherichia coli* (*E. coli*) cells. The expression culture was induced with 1 mM IPTG at
429 OD=0.6 and grown for 16 hours at 18°C. After resuspension in LEW buffer (50 mM
430 NaH_2PO_4, pH 8; 300 mM NaCl; 1 mM DTT), the cells were lysed by incubation with 1.2
431 mg/ml lysozyme for 30 min and subsequent sonification in the presence of 1 mM PMSF.
432 Recombinant protein was purified by Ni^{++} affinity chromatography and subjected to Amicon
433 Ultra-15 (30K) (Merck Millipore) filtration for buffer exchange to imidazole-free LEW
434 containing 20% glycerol.

435

436 **PCO Activity Assays and MALDI Analysis**

437 PCO activity assays were conducted under the following conditions unless otherwise stated: 1
438 μM PCO1 or 4 was mixed with 100 or 200 μM $RAP2_{2-11}$ peptide in 250 mM NaCl, 1 mM
439 dithiothreitol (DTT), 50 mM Tris-HCl pH 7.5 and incubated at 30°C for 30-60 minutes.
440 Assays were stopped by quenching 1 μL sample with 1 μL alpha-cyano-4-hydroxycinnamic
441 acid (CHCA) matrix on a MALDI plate prior to product mass analysis using a Sciex 4800
442 TOF/TOF mass spectrometer (Applied Biosystems) operated in negative ion reflectron mode.
443 The instrument parameters and data acquisition were controlled by 4000 Series Explorer
444 software and data processing was completed using Data Explorer (Applied Biosystems).

445

446 To test the activity of PCO4 in the presence of $^{18}O_2$, 100 μL of an anaerobic solution of 100
447 μM $RAP2_{2-11}$ in 250 mM NaCl/50 mM Tris-HCl pH 7.5 was prepared in a septum-sealed
448 glass vial by purging with 100% N_2 for 10 minutes at 100 mL/min using a mass flow
449 controller (Brooks Instruments), as used for previous preparation of anaerobic samples to
450 determine enzyme dependence on O_2.[24] PCO4 was then added using a gas-tight Hamilton
451 syringe, followed by purging with a balloon (approx. 0.7 L) of $^{16}O_2$ or $^{18}O_2$ over the course of
452 10 minutes at room temperature. Reaction vials were then transferred to 30°C for a further 20
453 minutes before products were analysed by MALDI-MS as described above.

454

455 PCO4 activity was additionally tested in the presence of $H_2{}^{18}O$ by conducting an assay in
456 75% $H_2{}^{18}O$, 25% H_2O (with all enzyme/substrate/buffer components comprising a portion of
457 the H_2O fraction). Assays were conducted for 10 minutes at room temperature followed by 20
458 minutes at 30°C for comparison with assays conducted with $^{18}O_2$. Products were analysed by
459 MALDI-MS, as described above.

460

461 **UPLC-MS and MS/MS Analysis of PCO Assay Products**

462 Ultra-high performance chromatography (UPLC) mass spectrometry (MS) measurements
463 were obtained using an Acquity UPLC system coupled to a Xevo G2-S Q-ToF mass
464 spectrometer (Waters) operated in positive electrospray mode. Instrument parameters, data
465 acquisition and data processing were controlled by Masslynx 4.1. Source conditions were
466 adjusted to maximize sensitivity and minimize fragmentation while Lockspray was employed
467 during analysis to maintain mass accuracy. 2 μL of each sample was injected on to a
468 Chromolith Performance RP-18e 100-2 mm column (Merck) heated to 40 ^{o}C and eluted using
469 a gradient of 95 % deionized water supplemented with 0.1 % (v/v) formic acid (analytical
470 grade) to 95 % acetonitrile (HPLC grade) and a flow rate of 0.3 mL/min. Fragmentation
471 spectra of substrate and product peptide ions (MS/MS) were obtained using a targeted
472 approach with a typical collision-induced dissociation (CID) energy ramp of 30 to 40 eV.
473 Analysis was carried out with the same source settings, flow rate and column elution
474 conditions as above.

475

476 **^{1}H NMR Assay**

477 Reaction components (5 μM PCO1 or PCO4 and 500 μM RAP2$_{2-11}$) were prepared to 75 μL
478 in 156 mM NaCl, 31 mM Tris-HCl (pH 7.5) and 10% D_2O (enzyme added last), in a 1.5 mL
479 microcentrifuge tube before being transferred to a 2 mm diameter NMR tube. ^{1}H NMR
480 spectra at 310 K were recorded using a Bruker AVIII 600 (with inverse cryoprobe optimized
481 for ^{1}H observation and running topspin 2 software; Bruker) and reported in p.p.m. relative to
482 D_2O (δ_H 4.72). The deuterium signal was also used as internal lock signal and the solvent
483 signal was suppressed by presaturating its resonance.

484

485

486 **Arginylation Assay**

487 The conditions for arginylation of the 12-mer peptide substrates were modified from [36]. In
488 detail, ATE1 was incubated at 10 µM in the reaction mixture containing 50 mM HEPES, pH
489 7.5; 25 mM KCl; 15 mM $MgCl_2$; 1 mM DTT; 2.5 mM ATP; 0.6 mg/ml *E. coli* tRNA (R1753,
490 Sigma); 0.04 mg/ml *E. coli* aminoacyl-tRNA synthetase (A3646, Sigma); 80 µM (4 nCi/µl)
491 ^{14}C-arginine (MC1243, Hartmann Analytic); 50 µM C-terminally biotinylated 12-mer peptide
492 substrate and, where indicated, 1 µM purified recombinant PCO1 or PCO4 in a total reaction
493 volume of 50 µL. The reaction was conducted at 30°C for 16 to 40 hours. After incubation,
494 each 50 µL of avidin agarose bead slurry (20219, Pierce) equilibrated in PBSN (PBS-
495 Nonidet; 100 mM NaH_2PO_4; 150 mM NaCl; 0.1% Nonidet-P40) were added to the samples
496 and mixed with an additional 350 µL of PBSN. After 2 hours of rotation at room temperature,
497 the beads were washed 4 times in PBSN, resuspended in 4 mL of FilterSafe scintillation
498 solution (Zinsser Analytic) and scintillation counting was performed using a Beckmann
499 Coulter LS 6500 Multi-Purpose scintillation counter.

500

Acknowledgements

Petra Majovsky, Domenika Thieme and Wolfgang Hoehenwarter from the Proteomics Unit of the Leibniz Institute of Plant Biochemistry (IPB), Halle, are acknowledged for mass spectrometry of recombinant ATE1. This work was supported by a Biotechnology and Biological Sciences Research Council (U.K.) New Investigator grant (BB/M024458/1) to E.F., a grant for setting up the junior research group of the *ScienceCampus Halle – Plant-based Bioeconomy* to N.D., by a Ph.D. fellowship of the Landesgraduiertenförderung Sachsen-Anhalt awarded to C.N., a Royal Society Dorothy Hodgkin Fellowship to E.F., a William R. Miller Junior Research Fellowship (St. Edmund Hall, Oxford) to R.H. and grant DI 1794/3-1 by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) to N.D. Financial support came from the Leibniz Association, the state of Saxony Anhalt, the Deutsche Forschungsgemeinschaft (DFG) Graduate Training Center GRK1026 "*Conformational Transitions in Macromolecular Interactions*" at Halle, and the Leibniz Institute of Plant Biochemistry (IPB) at Halle, Germany. We thank Prof. J. van Dongen (RWTH Aachen University, Germany) and Prof. F. Licausi (Scuolo Superiore Sant'Anna, Pisa, Italy) for sharing pDEST-PCO plasmids and helpful discussions.

Author Contributions

M.W. performed the PCO1/4 activity assays and MALDI/LC-/MS/MS analyses. M.K. performed and established arginylation reactions on peptides coupled to biotin pulldown and scintillation measurements and purified ATE1 protein. R.H. performed the NMR assays with E.F. D.W. prepared the pDEST17-PCO1 and 4 plasmids. C.M. synthesized the peptides, T.N.G. supervised and designed the synthesis, C.N. cloned and established purification and activity assays for ATE1. R.O. conducted LC-MS to analyse +12 Da mass shifts. J.W.

527 performed LC-MS analysis. E.F. performed the PCO1 and PCO4 protein purification and
528 selected activity assays. E.F., M.W., M.K. and N.D. designed the study, E.F. and N.D. wrote
529 the manuscript. M.W., M.K., N.D. and E.F. designed the figures. All authors read and
530 approved the final version of this manuscript.
531

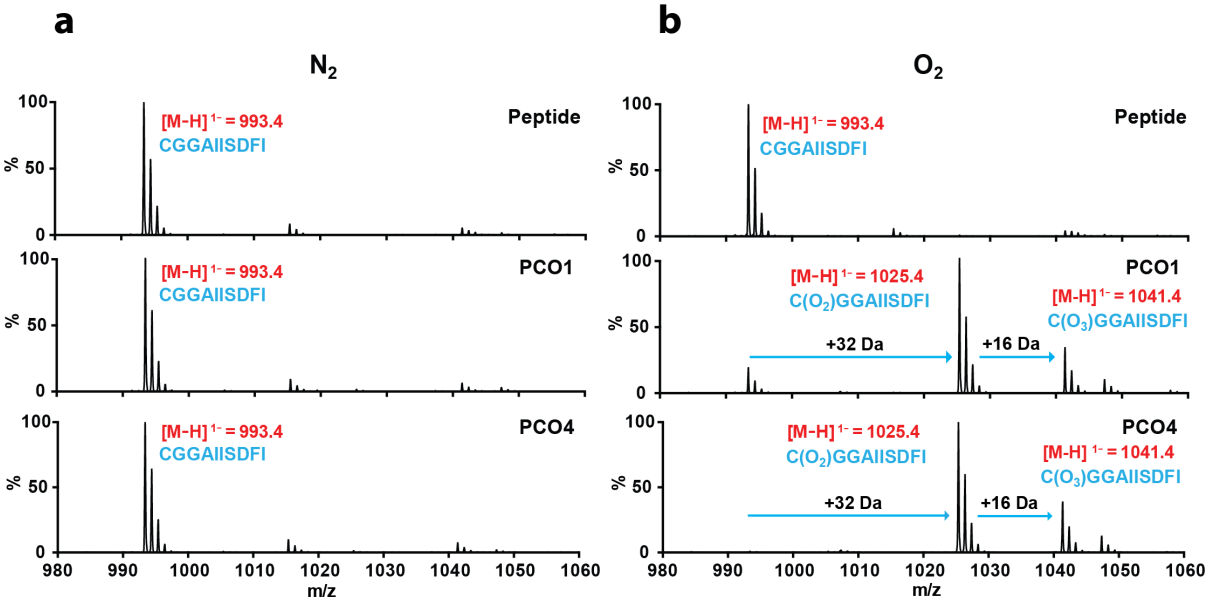
532 **References**

533 1. Kaelin, W.G., Jr. & Ratcliffe, P.J. Oxygen sensing by metazoans: the central role of the HIF
534 hydroxylase pathway. *Mol Cell* **30**, 393-402 (2008).
535 2. Myllyharju, J. Prolyl 4-hydroxylases, master regulators of the hypoxia response. *Acta Physiol*
536 *(Oxf)* **208**, 148-65 (2013).
537 3. Semenza, G.L. Oxygen sensing, homeostasis, and disease. *N Engl J Med* **365**, 537-47 (2011).
538 4. Bailey-Serres, J. et al. Making sense of low oxygen sensing. *Trends Plant Sci* **17**, 129-38
539 (2012).
540 5. Bailey-Serres, J., Lee, S.C. & Brinton, E. Waterproofing crops: effective flooding survival
541 strategies. *Plant Physiol* **160**, 1698-709 (2012).
542 6. Hattori, Y. et al. The ethylene response factors SNORKEL1 and SNORKEL2 allow rice to adapt
543 to deep water. *Nature* **460**, 1026-U116 (2009).
544 7. Hinz, M. et al. Arabidopsis RAP2.2: an ethylene response transcription factor that is
545 important for hypoxia survival. *Plant Physiol* **153**, 757-72 (2010).
546 8. Licausi, F. et al. HRE1 and HRE2, two hypoxia-inducible ethylene response factors, affect
547 anaerobic responses in Arabidopsis thaliana. *Plant J* **62**, 302-15 (2010).
548 9. Xu, K. et al. Sub1A is an ethylene-response-factor-like gene that confers submergence
549 tolerance to rice. *Nature* **442**, 705-8 (2006).
550 10. Papdi, C. et al. Functional identification of Arabidopsis stress regulatory genes using the
551 controlled cDNA overexpression system. *Plant Physiol* **147**, 528-42 (2008).
552 11. Mendiondo, G.M. et al. Enhanced waterlogging tolerance in barley by manipulation of
553 expression of the N-end rule pathway E3 ligase PROTEOLYSIS6. *Plant Biotechnol J* **14**, 40-50
554 (2016).
555 12. Lee, S.C. et al. Molecular characterization of the submergence response of the Arabidopsis
556 thaliana ecotype Columbia. *New Phytol* **190**, 457-71 (2011).
557 13. Mustroph, A. et al. Profiling translatomes of discrete cell populations resolves altered
558 cellular priorities during hypoxia in Arabidopsis. *Proc Natl Acad Sci U S A* **106**, 18843-8
559 (2009).
560 14. Varshavsky, A. The N-end rule pathway and regulation by proteolysis. *Protein Sci* (2011).
561 15. Tasaki, T., Sriram, S.M., Park, K.S. & Kwon, Y.T. The N-end rule pathway. *Annu Rev Biochem*
562 **81**, 261-89 (2012).
563 16. Gibbs, D.J., Bacardit, J., Bachmair, A. & Holdsworth, M.J. The eukaryotic N-end rule pathway:
564 conserved mechanisms and diverse functions. *Trends Cell Biol* (2014).
565 17. Gibbs, D.J. et al. Homeostatic response to hypoxia is regulated by the N-end rule pathway in
566 plants. *Nature* **479**, 415-8 (2011).
567 18. Licausi, F. et al. Oxygen sensing in plants is mediated by an N-end rule pathway for protein
568 destabilization. *Nature* **479**, 419-22 (2011).
569 19. Ross, S., Giglione, C., Pierre, M., Espagne, C. & Meinnel, T. Functional and developmental
570 impact of cytosolic protein N-terminal methionine excision in Arabidopsis. *Plant Physiol* **137**,
571 623-37 (2005).
572 20. Giglione, C., Boularot, A. & Meinnel, T. Protein N-terminal methionine excision. *Cell Mol Life*
573 *Sci* **61**, 1455-74 (2004).
574 21. Garzon, M. et al. PRT6/At5g02310 encodes an Arabidopsis ubiquitin ligase of the N-end rule
575 pathway with arginine specificity and is not the CER3 locus. *FEBS Lett* **581**, 3189-96 (2007).
576 22. Gibbs, D.J. et al. Nitric oxide sensing in plants is mediated by proteolytic control of group VII
577 ERF transcription factors. *Mol Cell* **53**, 369-79 (2014).
578 23. Gibbs, D.J. et al. Group VII Ethylene Response Factors Coordinate Oxygen and Nitric Oxide
579 Signal Transduction and Stress Responses in Plants. *Plant Physiol* **169**, 23-31 (2015).
580 24. Tarhonskaya, H. et al. Investigating the contribution of the active site environment to the
581 slow reaction of hypoxia-inducible factor prolyl hydroxylase domain 2 with oxygen. *Biochem*
582 *J* (2014).

25. Hirsila, M., Koivunen, P., Gunzler, V., Kivirikko, K.I. & Myllyharju, J. Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor. *J Biol Chem* **278**, 30772-80 (2003).
26. Bruick, R.K. Oxygen sensing in the hypoxic response pathway: regulation of the hypoxia-inducible transcription factor. *Genes Dev* **17**, 2614-23 (2003).
27. Epstein, A.C. et al. C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell* **107**, 43-54 (2001).
28. Weits, D.A. et al. Plant cysteine oxidases control the oxygen-dependent branch of the N-end-rule pathway. *Nat Commun* **5**, 3425 (2014).
29. Li, W. & Pierce, B.S. Steady-state substrate specificity and O2-coupling efficiency of mouse cysteine dioxygenase. *Arch Biochem Biophys* **565**, 49-56 (2015).
30. Graciet, E. et al. The N-end rule pathway controls multiple functions during Arabidopsis shoot and leaf development. *Proc Natl Acad Sci U S A* **106**, 13618-23 (2009).
31. Kallen, R.G. The mechanism of reactions involving Schiff base intermediates. Thiazolidine formation from L-cysteine and formaldehyde. *J Am Chem Soc* **93**, 6236-48 (1971).
32. Davydov, I.V. & Varshavsky, A. RGS4 is arginylated and degraded by the N-end rule pathway in vitro. *J Biol Chem* **275**, 22931-41 (2000).
33. Hu, R.G. et al. The N-end rule pathway as a nitric oxide sensor controlling the levels of multiple regulators. *Nature* **437**, 981-6 (2005).
34. Kwon, Y.T. et al. An essential role of N-terminal arginylation in cardiovascular development. *Science* **297**, 96-9 (2002).
35. Joseph, C.A. & Maroney, M.J. Cysteine dioxygenase: structure and mechanism. *Chem Commun (Camb)*, 3338-49 (2007).
36. Wang, J. et al. Arginyltransferase ATE1 catalyzes midchain arginylation of proteins at side chain carboxylates in vivo. *Chem Biol* **21**, 331-7 (2014).
37. Eriste, E. et al. A novel form of neurotensin post-translationally modified by arginylation. *J Biol Chem* **280**, 35089-97 (2005).
38. Wong, C.C. et al. Global analysis of posttranslational protein arginylation. *PLoS Biol* **5**, e258 (2007).
39. Lee, M.J. et al. RGS4 and RGS5 are in vivo substrates of the N-end rule pathway. *Proc Natl Acad Sci U S A* **102**, 15030-5 (2005).
40. Cha-Molstad, H. et al. Amino-terminal arginylation targets endoplasmic reticulum chaperone BiP for autophagy through p62 binding. *Nat Cell Biol* **17**, 917-29 (2015).
41. Schuessele, C. et al. Spatio-temporal patterning of arginyl-tRNA protein transferase (ATE) contributes to gametophytic development in a moss. *New Phytol* **209**, 1014-27 (2016).
42. Hoernstein, S.N. et al. Identification of Targets and Interaction Partners of Arginyl-tRNA Protein Transferase in the Moss Physcomitrella patens. *Mol Cell Proteomics* **15**, 1808-22 (2016).
43. Dar, M.H., de Janvry, A., Emerick, K., Raitzer, D. & Sadoulet, E. Flood-tolerant rice reduces yield variability and raises expected yield, differentially benefitting socially disadvantaged groups. *Sci Rep* **3**, 3315 (2013).

627 **Figures**

628



629

630

631 **Figure 1. PCO1, PCO4 and O_2-dependent modification of a RAP2$_{2\text{-}11}$ peptide substrate,**

632 **consistent with Cys-oxidation.** MALDI-MS spectra showing products following PCO1 and

633 PCO4 incubation with RAP2$_{2\text{-}11}$ under anaerobic (**a**) or aerobic (**b**) conditions. Products with

634 mass increases of +32 Da and +48 Da were only observed in the presence of PCO1 or PCO4

635 and O_2.

636

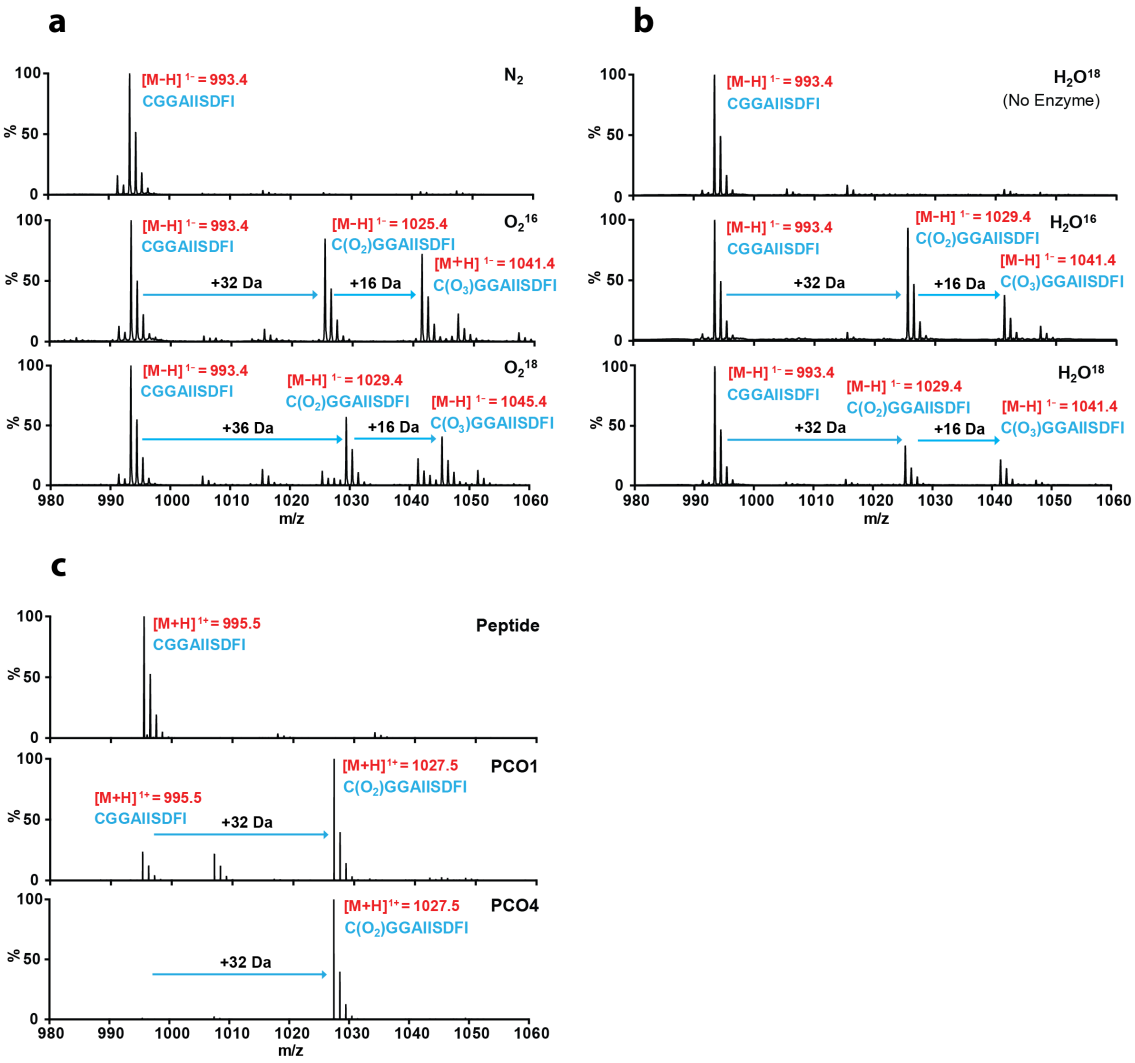


Figure 2. The PCOs are dioxygenases which catalyze incorporation of molecular O_2 into $RAP2_{2-11}$. (a) MALDI-MS spectra showing that PCO4-catalyzed reactions carried out in the presence of $^{18}O_2$ result in a +4 Da increase in the mass of the putative Cys-sulfinic acid product, however a +6 Da increase in the size of the putative Cys-sulfonic acid product is not observed; **(b)** MALDI-MS spectra showing that PCO4-catalyzed reactions carried out in the presence of $H_2{}^{18}O$ show no additional incorporation of mass compared with products of reactions in the presence of $H_2{}^{16}O$; **(c)** LC-MS spectra confirm that the +48 Da reaction product is an artifact of MALDI-MS analysis (**Supplementary Figure 3**) and incubation of PCO1 and PCO4 with $RAP2_{2-11}$ results in formation of a single product with a mass increase of +32 Da, consistent with Cys-sulfinic acid formation.

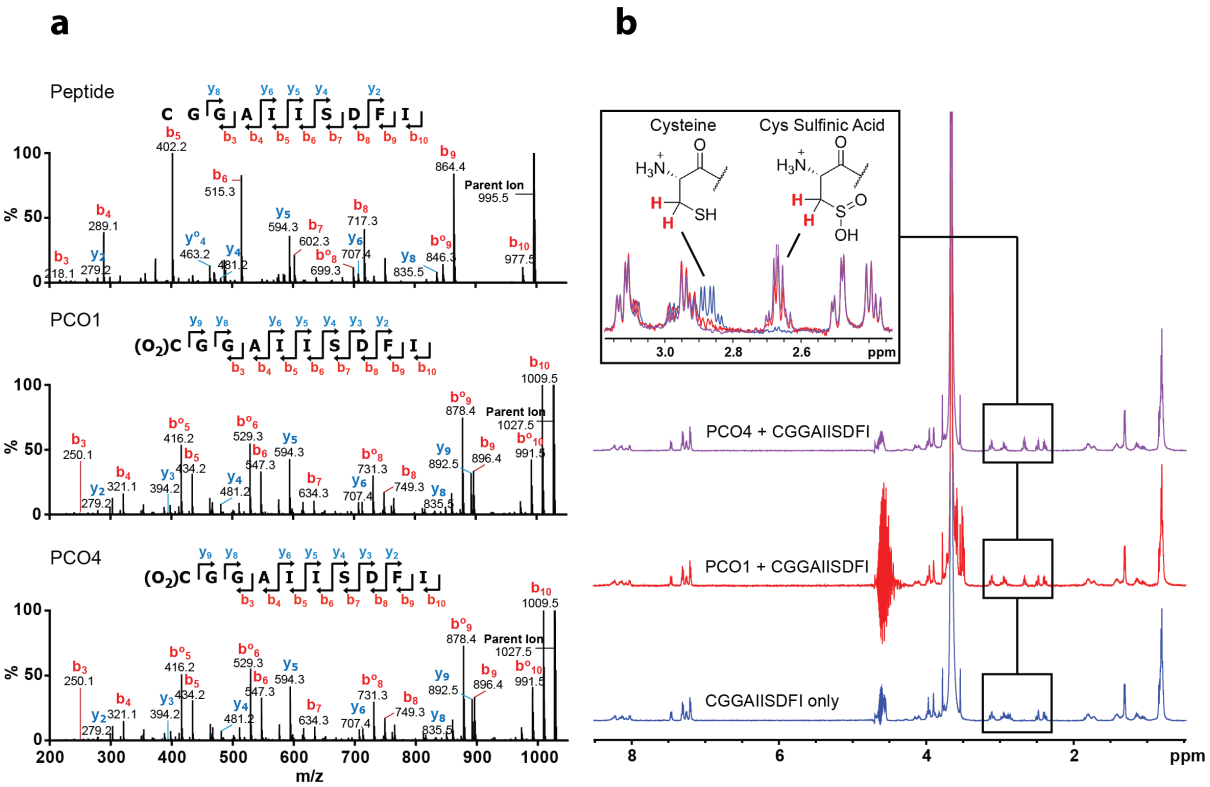


Figure 3. PCO1 and PCO4 oxidize the N-terminal Cys of RAP2$_{2\text{-}11}$ to Cys-sulfinic acid as confirmed by (a) LC-MS/MS and (b) ^{1}H-NMR. (**a**) Peptidic products of PCO-catalyzed reactions were subjected to LC-MS/MS analysis. In the presence of enzyme, fragment assignment was consistent with expected b- and y-series ion masses for RAP2$_{2\text{-}11}$ with N-terminal Cys-sulfinic acid. (**b**) ^{1}H-NMR was used to monitor changes to RAP2$_{2\text{-}11}$ (500 μM) upon incubation with enzyme (5 μM). In the presence of PCO1 (red) and PCO4 (purple), the ^{1}H-resonance at δ_H 2.88 ppm (assigned to the β-cysteinyl protons of RAP2$_{2\text{-}11}$, blue) was observed to decrease in intensity, with concomitant emergence of a resonance at δ_H 2.67 ppm. This new resonance was assigned to the β-protons of Cys-sulfinic based on chemical shift analysis (see **Supplementary Figure 4**).

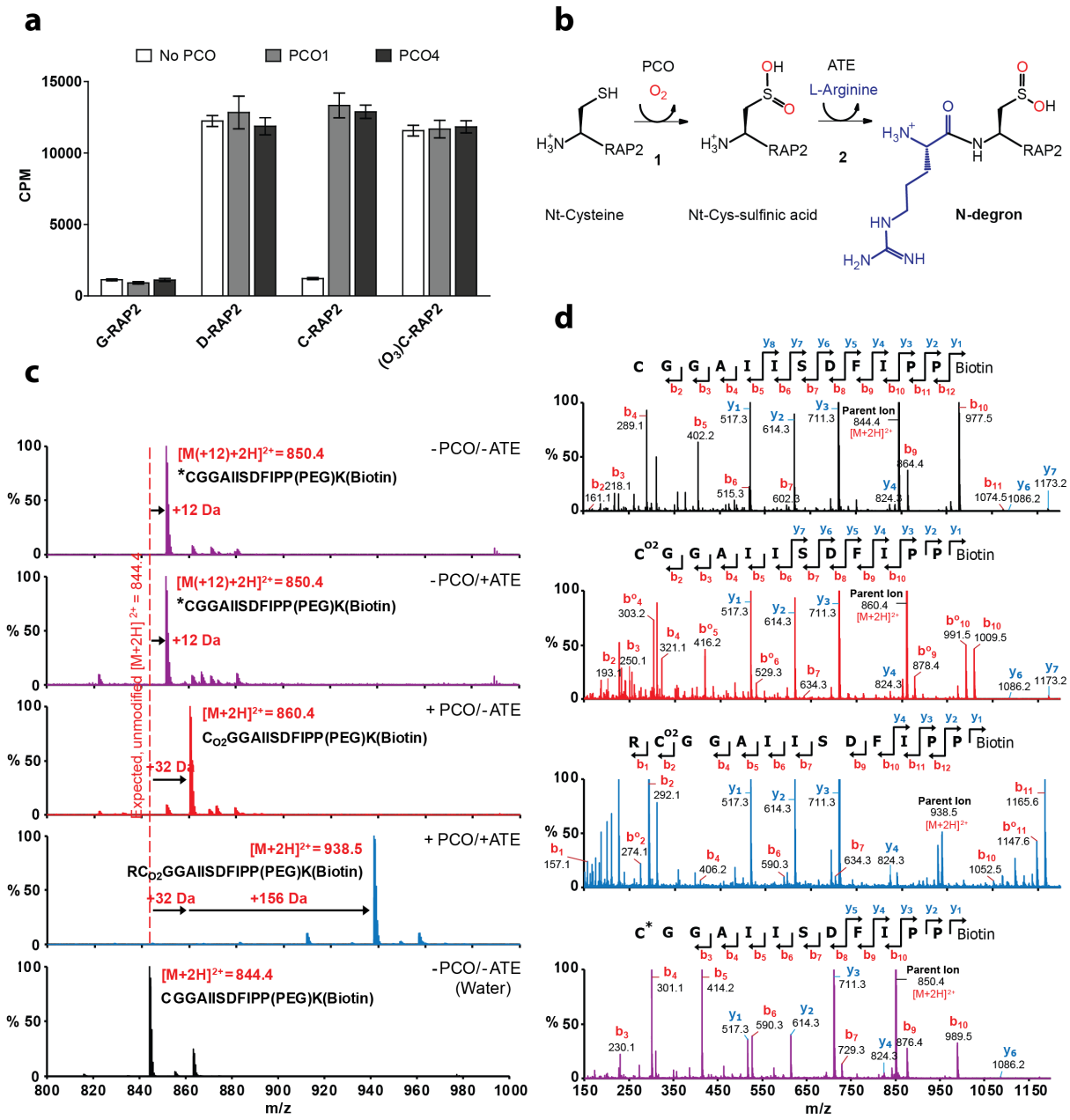


Figure 4. PCO-Catalyzed Cys-sulfinic acid formation renders RAP2$_{2\text{-}13}$ a substrate for ATE1 catalyzed arginylation. (a) ^{14}C-Arg incorporation by ATE1 into the 12-mer N-terminal RAP2 peptide (H_2N-XGGAIISDFIPP(PEG)K(biotin)-NH_2, X = Gly, Asp, Cys or Cys-sulfonic acid (C(O_3))), was assayed by liquid scintillation counting of immobilized biotinylated peptides after the arginylation reaction and removal of unreacted ^{14}C-Arg (n=3). In the case of the Cys-starting peptide (RAP2$_{2\text{-}13}$), ATE1 activity was strongly dependent on the presence of PCO1 or PCO4. **(b)** Scheme showing PCO- and ATE1-catalyzed reactions on

671 Nt-Cys RAP2.2, as validated in this study. **(c)** LC-MS spectra of products of equivalent
672 assays with Cys-initiated $RAP2_{2-13}$ using non-radiolabelled Arg, revealing a sequential mass
673 increase of +32 (corresponding to oxidation) and +156 Da (corresponding to arginylation)
674 only in the presence of PCO and ATE1 (blue spectrum). The red spectrum shows a +32 Da
675 mass increase for Cys-$RAP2_{2-13}$ incubated +PCO/-ATE, demonstrating Cys-sulfonic acid
676 formation as expected. Purple spectra show +12 Da products formed upon incubation of Cys-
677 $RAP2_{2-13}$ in the absence of PCO +/-ATE (for explanation of this mass increase see text and
678 **Supplementary Figure 6**); the black spectrum shows Cys-$RAP2_{2-13}$ dissolved in H_2O. **(d)**
679 b- and y-ion series spectra generated by MS/MS analysis of Cys-$RAP2_{2-13}$ only (no
680 incubation; black), Cys-$RAP2_{2-13}$ incubated +PCO/-ATE (red), Cys-$RAP2_{2-13}$ incubated with
681 PCO and ATE1 (blue) and Cys-$RAP2_{2-13}$ incubated without PCO or ATE1 (purple),
682 confirming arginylation only at the N-terminus of PCO-modified $RAP2_{2-13}$.