The identification of carbon dioxide mediated protein post-translational modifications

Victoria L. Linthwaite1,2, Joanna M. Janus1,2, Adrian P. Brown1,2, David Wong-Pascua3, AnnMarie C. O’Donoghue2,3,4, Andrew Porter5, Achim Treumann5, David R.W. Hodgson2,3,4 & Martin J. Cann1,2

Carbon dioxide is vital to the chemistry of life processes including metabolism, cellular homeostasis, and pathogenesis. CO2 is generally unreactive but can combine with neutral amines to form carbamates on proteins under physiological conditions. The most widely known examples of this are CO2 regulation of ribulose 1,5-bisphosphate carboxylase/oxygenase and haemoglobin. However, the systematic identification of CO2-binding sites on proteins formed through carbamylation has not been possible due to the ready reversibility of carbamate formation. Here we demonstrate a methodology to identify protein carbamates using triethyloxonium tetrafluoroborate to covalently trap CO2, allowing for downstream proteomic analysis. This report describes the systematic identification of carbamates in a physiologically relevant environment. We demonstrate the identification of carbamylated proteins and the general principle that CO2 can impact protein biochemistry through carbamate formation. The ability to identify protein carbamates will significantly advance our understanding of cellular CO2 interactions.
Protein functionalities can be extended and modulated by enzyme-catalysed and spontaneous post-translational modifications (PTMs) such as phosphorylation, nitrosylation, acetylation, methylation, hydroxylation, glycosylation and the attachment of other small proteins. The earliest known PTM, the addition of CO$_2$ to protein amino groups, was uncovered in two classic studies of early physiology. Bohr and co-workers demonstrated that the haemoglobin oxygen saturation curve was responsive to the partial pressure of CO$_2$ while Christiansen and co-workers showed that CO$_2$ uptake by the blood at constant pCO$_2$ was increased by the presence of O$_2$. Henriques then used kinetic evidence to postulate the direct combination of CO$_2$ with the free amino groups on haemoglobin and Ferguson and Roughton confirmed this through direct chemical analyses. The site of CO$_2$ binding was demonstrated by computation that predicts as many as 1.3% of large protein functionalities can be extended and modulated by enzyme-catalysed and spontaneous post-translational modifications (PTMs) such as phosphorylation, nitrosylation, acetylation, methylation, hydroxylation, glycosylation and the attachment of other small proteins. The earliest known PTM, the addition of CO$_2$ to protein amino groups, was uncovered in two classic studies of early physiology. Bohr and co-workers demonstrated that the haemoglobin oxygen saturation curve was responsive to the partial pressure of CO$_2$ while Christiansen and co-workers showed that CO$_2$ uptake by the blood at constant pCO$_2$ was increased by the presence of O$_2$. Henriques then used kinetic evidence to postulate the direct combination of CO$_2$ with the free amino groups on haemoglobin and Ferguson and Roughton confirmed this through direct chemical analyses. The site of CO$_2$ binding was demonstrated by computation that predicts as many as 1.3% of large protein functionalities can be extended and modulated by enzyme-catalysed and spontaneous post-translational modifications (PTMs) such as phosphorylation, nitrosylation, acetylation, methylation, hydroxylation, glycosylation and the attachment of other small proteins. The earliest known PTM, the addition of CO$_2$ to protein amino groups, was uncovered in two classic studies of early physiology. Bohr and co-workers demonstrated that the haemoglobin oxygen saturation curve was responsive to the partial pressure of CO$_2$ while Christiansen and co-workers showed that CO$_2$ uptake by the blood at constant pCO$_2$ was increased by the presence of O$_2$. Henriques then used kinetic evidence to postulate the direct combination of CO$_2$ with the free amino groups on haemoglobin and Ferguson and Roughton confirmed this through direct chemical analyses. The site of CO$_2$ binding was demonstrated by computation that predicts as many as 1.3% of large protein functionalities can be extended and modulated by enzyme-catalysed and spontaneous post-translational modifications (PTMs) such as phosphorylation, nitrosylation, acetylation, methylation, hydroxylation, glycosylation and the attachment of other small proteins. The earliest known PTM, the addition of CO$_2$ to protein amino groups, was uncovered in two classic studies of early physiology. Bohr and co-workers demonstrated that the haemoglobin oxygen saturation curve was responsive to the partial pressure of CO$_2$ while Christiansen and co-workers showed that CO$_2$ uptake by the blood at constant pCO$_2$ was increased by the presence of O$_2$. Henriques then used kinetic evidence to postulate the direct combination of CO$_2$ with the free amino groups on haemoglobin and Ferguson and Roughton confirmed this through direct chemical analyses. The site of CO$_2$ binding was demonstrated by computation that predicts as many as 1.3% of large pro...
hydrolysis allows carbamate trapping and pH control of the experiment to take place on a convenient laboratory timescale.

We sequentially validated the potential of TEO as a carbamate-trapping agent on amino acid, peptide and protein substrates. Experiments with amino acid and peptide substrates were performed at pH 8.5 to promote carbamate formation in these systems. Initial investigation centred on whether TEO could trap a carbamate on α-N-acetyl-lysine under aqueous conditions. We hypothesised that a carbamate would form on the ε-amino group that could be subsequently trapped with TEO (Fig. 2a). A solution of α-N-acetyl-lysine was incubated with excess 13CO2/H13CO3− at pH 8.5 and the formation of a carbamate was confirmed by the presence of a peak at 164 ppm by 13C NMR spectroscopy (Fig. 2b)28. Separately, TEO was added to an α-N-acetyl-lysine/CO2/HCO3− mixture at constant pH and the reaction products were analysed by LC-ESI-MS. The ethylation mixture was resolved into three major components that demonstrated the trapping of ε-carbamate was successful with side products including C-terminal and N-ethylation (Fig. 3a). To further confirm the formation of the carbamate, the buffered NaHCO3 solution used to provide CO2 was replaced with NaH13CO3, which resulted in the expected 1 Da m/z increase on MS analysis (Fig. 3b). The ethylation product mixture was extracted into ether and its 1H NMR spectrum (Figure 3cii) was compared to a chemically synthesised standard ε-ethoxy carbonyl-lysine (Figure 3c). The product spectrum shows key signals consistent with N-carboxyethylation at δ~4.25 ppm and ~1.2 ppm that corroborate the findings from LC-ESI-MS (Fig. 3a).

After confirming carbamate trapping on α-N-acetyl-lysine we examined peptide systems, focusing on whether TEO could trap a carbamate on α-N-acetyl-lysine-carbamate formation and trapping (O-ethylation) by TEO. b 13C-NMR spectrum demonstrating the formation of a carbamate on α-N-acetyl-lysine by the appearance of a peak at 164 ppm.

Triethyloxonium ion-mediated carbamate trapping on protein. Having demonstrated that TEO is a suitable tool to trap carbamates on amines, we sought to use it for the discovery of protein carbamates that would represent sites for CO2 binding that are exchangeable with the environment. We hypothesised that selective CO2 binding to protein through carbamate formation would occur in structurally privileged sites that have evolved to facilitate carbamate formation. For example, CO2 binding to haemoglobin at the Val-1β site occurs through such a privileged environment. The formation of non-specific carbamates at other sites on protein is proportionately much less likely due to the pKa of the Lys ε-amino group being ~9–10.

The previous experiments with amino acids and peptides had been performed at pH 8.5 to promote carbamate formation by driving the ε-amino group protonation equilibrium towards the uncharged state. However, experiments with protein were performed at pH 7.4 to replicate a cellular environment which does not enhance carbamate formation. Carbamates will therefore only form in privileged environments. We first investigated TEO-mediated CO2 trapping on the N-terminal valine of the haemoglobin β-chain. Carbamate formation was confirmed using 13C-NMR spectroscopy by the observation of a signal at 164 ppm which matched literature values29 (Fig. 5a).

We trapped CO2 onto human haemoglobin with TEO and analysed the trypsin-digested products by ESI-MS. A carbamate was identified on the ε-amine of the β-chain N-terminal valine (peptide mass 1024.10 Da), consistent with the literature (Fig. 5b)30. Peptides carrying a carbamate were identified both with and without ethylation on E7. Ethylation of alternative sites therefore does not influence the ability of the method to trap carbamates. The experiment was repeated after 4% SDS addition and removal to denature haemoglobin and thus destroy the local privileged environment required for carbamate formation. No trapped carbamate was observed in an experiment performed under these conditions. Carbamate formation therefore requires a structure-dependent privileged environment within the protein. Removal of the protein structure by SDS destroyed this privileged environment and thus the carbamate could not form. The trapping methodology is therefore able to identify known functional carbamates on proteins under physiologically relevant conditions of pH and [CO2].
Fig. 3 Characterisation of carbamate trapping on the ε-NH$_2$ group of α-N-acetyl-lysine. 

a. Total ion chromatogram of the carbamate-trapping reaction mixture of α-N-acetyl-lysine with TEO, and m/z profile for species at retention time ~2.6 min. The major products of the trapping reaction are N-acetyl-lysine ethylated on the α-carboxylate group ( retention time ~1.7 min), α-N-acetyl-lysine ethylated on the α-carboxylate and ε-NH$_2$ groups (retention time ~2.1 min) and α-N-acetyl-lysine ethylated on the ε-carbamate and the α-carboxylate groups (retention time ~2.6 min). 

b. MS trace demonstrating the increase of one mass unit from 12C (i) with the use of 13C labelled CO$_2$ (ii). 

C. ¹H-NMR spectra comparing ethyloxycarbonyl signals between a chemically synthesised α-N-acetyl-ε-N-ethyloxycarbonyl-lysine and b ethylation products formed during the trapping experiments between α-N-acetyl-lysine, CO$_2$ and TEO (key CH$_2$ signals highlighted in red).
We trapped CO₂ onto intact rabbit red blood cells with TEO to confirm that the methodology is able to identify carbamates in the normal cellular environment. Trypsin-digested whole cells were analysed by LC-ESI-MS. The expected carbamate on haemoglobin was again identified on the α-amine of the β-chain N-terminal valine (Fig. 5c). This experiment demonstrates that the developed methodology can also be applied within a cell and that the results obtained are identical to those from isolated protein.

The identification of protein carbamates. We hypothesised that the TEO trapping methodology could be used to isolate previously unidentified CO₂-binding sites on proteins. We therefore performed a small-scale screen of the proteome of a model organism to establish the general principle that CO₂ can form labile interactions with protein through carbamate formation. We selected the CO₂-fixing organism Arabidopsis thaliana for study as we hypothesised it would be most likely to utilise protein carbamylation as a mechanism to couple CO₂ availability to protein function. Extracts of soluble proteins derived from the leaves of A. thaliana were incubated with NaH¹⁴CO₃ and subjected to TEO-trapping (Fig. 6). Little¹⁴CO₂ was incorporated into the protein extracts in the absence of TEO. The inability to identify protein-bound ¹⁴CO₂ in the absence of TEO was due to the ready reversibility of carbamylation that leads to degassing of the sample during preparation for analysis. The trapped proteome contained significant levels of¹⁴C, even when accounting for 50% of the total protein sample being Rubisco. We concluded that Arabidopsis protein extract contains CO₂-interacting proteins carbamylated at labile sites exchangeable with the environment. We therefore proceeded to identify a subset of these carbamylated proteins.

To identify carbamylated proteins within Arabidopsis, soluble leaf protein was equilibrated with CO₂/HCO₃⁻ at pH 7.4 and TEO was added. The trapping reaction mixture was digested with trypsin and samples were analysed by LC-MS-MS. The data were interrogated for variable post-translational modifications on lysine with masses of 72.0211 Da (trapped carbamate) and 28.0313 Da (O-ethylation on glutamate and aspartate side chains). Occasional N-ethylation of the lysine or arginine amino group was also observed (as seen in Fig. 2a, for example). Carbamate formation occurs
two b-ions confirmed the location of the PTM under MS–MS conditions. Second, only peptides that contained an internal lysine residue (missed cleavage) were accepted because carbamylation removes the positive charge on the lysine that is essential for cleavage site recognition by trypsin. This is analogous to the removal of tryptic cleavage sites through lysine acetylation. We identified eight CO2 binding sites in Arabidopsis (Table 1, Fig. 7). Assignment of the MSMS spectra was manually verified and supported by high mass accuracy measurements of the fragment ions in 7 out of the 8 spectra (Supplementary Data 1). Together these data suggest that trapping CO2 with TEO can be used for the discovery of proteins post-translationally modified by CO2.

**Validation of a protein carbamate.** We hypothesised that CO2 would influence the activity of these discovered proteins at the identified site. To demonstrate this, we selected a hit protein for further investigation. The Class III peroxidase PRX34 (AtPRX34; At3g49120) was identified as a hit by MS-MS (Fig. 7c, d). Two proximal lysine carbamylation sites were identified (MSMS peptide amino acids 255–268 TPTVFDNYYVNLK, proposed carbamylation on K262; MSMS peptide amino acids 263–270 YYVNLKER, proposed carbamylation on K268). We further identified lysine carbamylation on K262 and K268 simultaneously (MSMS peptide amino acids 255–268 TPTVFDNYYVNLKER, proposed carbamylation on K262 and K268) indicating that PRX34 carbamylation does not necessarily occur exclusively on K262 or K268 (Fig. 8a). AtPRX34 generates H2O2 in response to microbe-associated molecular patterns suggesting that AtPRX34 has a role in basal defence responses in the plant. We over-expressed the mature coding sequence (amino acids 31–353; without transit peptide sequence) of wild type AtPRX34 and both K262A and K268A single site mutants in E. coli as His-tagged fusion proteins. We assayed the AtPRX34 wild type, AtPRX34

---

**Fig. 5** Identification of the exchangeable CO2-binding site on haemoglobin. a 13C-NMR spectrum demonstrating the formation of a carbamate on haemoglobin by the appearance of a peak at 164 ppm (i) together with a peak from H13CO2~ in solution (ii). b A plot of relative fragment intensity versus mass/charge ratio (m/z) for fragmentation data from MS-MS identifying an ethyl-trapped carbamate on the N-terminal valine of the haemoglobin β-chain. The experiment used purified haemoglobin. The peptide sequence above indicates the identification of predominant +1y (red) +1b (blue) ions by MS-MS shown in the plot. The modified residue is indicated in bold. The experiment also identifies a further ethylation on E7. c A plot of relative fragment intensity versus mass/charge ratio (m/z) for fragmentation data from MS-MS identifying an ethyl-trapped carbamate on the N-terminal valine of the haemoglobin β-chain. The experiment used whole red blood cells. The peptide sequence above indicates the identification of predominant +1y (red) +1b (blue) ions by MS-MS shown in the plot. The modified residue is indicated in bold.

**Fig. 6** Identification of exchangeable CO2-binding site on Arabidopsis protein extract. 14CO2 trapped onto protein extract of Arabidopsis thaliana (p < 0.0001, two-tailed t-test, n = 3 independent replicates, t = 29.85, df = 4, ±S.E.M.)

**Table 1 Carbamylated proteins in A. thaliana**

<table>
<thead>
<tr>
<th>Genome Identification Number</th>
<th>Protein</th>
<th>Residue</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1g238540</td>
<td>Lipid-transfer protein</td>
<td>K65</td>
</tr>
<tr>
<td>AtCG00490</td>
<td>Rubisco Large Chain</td>
<td>K183</td>
</tr>
<tr>
<td>At3g49120</td>
<td>Peroxidase</td>
<td>K262, K268</td>
</tr>
<tr>
<td>At1g21330</td>
<td>FBA1</td>
<td>K293</td>
</tr>
<tr>
<td>At3g54400</td>
<td>Eukaryotic aspartyl protease family protein</td>
<td>K251</td>
</tr>
<tr>
<td>At4g21280</td>
<td>PSBQA</td>
<td>K109</td>
</tr>
<tr>
<td>At4g25100</td>
<td>Fe Superoxide dismutase 1</td>
<td>K208</td>
</tr>
</tbody>
</table>

---
K262A, and AtPRX34 K268A proteins by measuring their ability to oxidise 2-methoxyphenol in the presence of H$_2$O$_2$ under conditions of atmospheric CO$_2$ (approximately 12 μM CO$_2$) and in the absence of CO$_2$. We compared the ratio of the specific activities for each protein in the presence and absence of CO$_2$ (Fig. 8b) in reaction mixtures with measured final pH of 7.4. Wild type protein showed no difference in specific activity under conditions of atmospheric CO$_2$ compared to no CO$_2$. However, both the K262A and K268A mutants demonstrated elevated activity under conditions of atmospheric CO$_2$ compared to the absence of CO$_2$. This result suggests a control mechanism similar to that found within haemoglobin. Carbamylation of the two
haemoglobin Val-1β sites reduced the affinity of haemoglobin for O₂ at a third site. We hypothesise that carbamate formation at K262 and K268 within AtPRX34 provide a control system for peroxidase activity in the presence of CO₂. This reduction in activity is clearly altered when either site is mutated and carbamate formation cannot occur (Fig. 8b). The higher reactivity of the mutants leaves the only explanation to be a role in suppression. We were unable to detect a trapped carbamate at K262 in the K268A mutant protein or a trapped carbamate at K268 in the K262A mutant protein by MSMS. This suggests that the lysine at residue 262 or 268 promote or stabilise carbamate formation at the alternative site as evidenced by the identification of the singly carbamylated sites in the wild type proteins (Fig. 7c–d). Mutation of either lysine to alanine makes it less likely for a carbamate to form at the other site. This is manifested as a loss of sensitivity to CO₂ in the single mutant proteins. The specific activities of the wild type, K262A and K268A proteins at atmospheric CO₂ are 0.36±0.027, 0.479±0.12 and 0.50±0.099 μmol 1.2-benzoquinone mg⁻¹ min⁻¹. These values demonstrate that the altered response to CO₂ in the mutant protein is a true activity change and not due to a change in specific activity caused by the mutation. No other carbamates were identified on AtPRX34 therefore CO₂ is likely able to interact with AtPRX34 at another site by an alternative carbamate-independent mechanism as previously observed. Carbamylation at either K262 or K268 therefore mitigates the effects of CO₂ at a third site which would otherwise activate the enzyme.

We hypothesised that mutation of K262 or K268 to glutamate would represent the local charge state of a carbamate at 100% occupancy. We therefore compared the ratio of the specific activities for PRX34 K262E or K268E to the wild type protein in the presence and absence of CO₂ (Fig. 8c). As before, wild type protein showed only negligible difference in specific activity under conditions of atmospheric CO₂ compared to no CO₂. In addition to this both the K262E and K268E mutants demonstrated no change in activity. However, these values were significantly different from the wild type which highlights the possible variability in occupancy of the carbamate in the wild type protein. This variability is not present in the fully occupied glutamate mutants. The specific activities of the wild type, K262E and K268E proteins at atmospheric CO₂ are 1.653±0.060, 1.477±0.027 and 0.933±0.020 μmol 1.2-benzoquinone mg⁻¹ min⁻¹. Specific activities are different between independent preparations of refolded proteins. The ratio of specific activities (Atmospheric CO₂:No CO₂) is independent of absolute specific activity and is comparable across preparations and repeatable across experiments.

The data are consistent with a model in which the two carbamate sites present are cooperative in maintaining protein activity levels. If the ability to carbamylate at one site is removed (K262A, K268A) then the difference of activity with changes in CO₂ levels significantly increases. In mutants that mimic carbamates at 100% occupancy (K262E, K268E) any change to activity due to changes in CO₂ level is removed.

AtPRX34 is closely related to four additional Class III peroxidases encoded in the *Arabidopsis* genome (AtPRX32, AtPRX33, AtPRX37 and AtPRX38). The carbamylated lysines are conserved in all five peroxidases. A future task, therefore, will be to elucidate the physiological function of the individual peroxidases in *Arabidopsis*, the role of CO₂ in these physiological processes and the impact of the individually carbamylated residues. We therefore demonstrate that under physiologically relevant conditions protein carbamates can be identified in which CO₂-binding site influences protein biochemistry in vitro.

**Discussion**

It is remarkable that so little is known about how CO₂ influences the function of the proteome, despite its fundamental importance within the cellular environment. Here we describe the general principle that CO₂ can reversibly bind protein through carbamate formation. The carbamates identified to date by design (haemoglobin, RubisCO) or fortuitously as stable modifications in crystal structures (urease, alamine racemase, transcarboxylase SS, class D β-lactamase, and phospho-riesterase) have clear functional roles. This confirms that carbamate formation is a candidate mechanism for protein activity to be directly responsive to environmental CO₂. However, the majority of these carbamates were discovered incidentally due to the lack of a tool for their direct investigation. We have presented a route to identify such CO₂-binding sites and provide evidence that such a site can influence protein biochemistry in a CO₂-dependent manner.

Our method operates under physiologically relevant conditions and successfully identified the known site of carbamate formation in haemoglobin dependent upon its local privileged environment. Analyses of the proteome of *A. thaliana* demonstrated significant CO₂ binding to protein dependent upon carbamylisation as evidenced by the requirement for TEO to trap CO₂ on protein. A small-scale proteomics screen identified eight carbamylation sites from 3614 proteins. Several other potential sites were ruled out by the stringent conditions used to eliminate potential false positives. Further developments in chromatography should enable us to increase the coverage of the proteome in such CO₂-trapped samples.

Our trapping method provides the capability for identifying proteins targeted by CO₂ in any system, which should in turn allow the construction of models for how cellular functions detect and therefore respond to CO₂. Protein carbamylation is likely to be more widespread than previously suspected and can represent a mechanism by which CO₂ availability is coupled to protein function. The challenge for the future is to identify protein targets for CO₂ and the functional roles of the resulting carbamate. It is highly likely that many, if not all, carbamylation sites will be functionally relevant.

In conclusion, we present a method for the identification of carbamylated proteins at the proteome level and show that this PTM is likely to be of biological significance. This method will allow a significant expansion of our current understanding of protein regulation by CO₂ and provide information concerning the extent to which CO₂ interacts with the proteome.

**Methods**

**CO₂ trapping.** All CO₂ trapping experiments were carried out in phosphate buffer (4 mL, 50 mM, pH 7.4). This solution was transferred to a TIM856 Titration

---

**Fig. 7** The identification of CO₂-binding proteins. MSMS spectra of peptides that were identified with ethyl-trapped carbamates on Lys residues. Panel b is a CID spectrum acquired on an LTQ Orbitrap XL mass spectrometer (low resolution), panel e was acquired on a QStar Pulsar mass spectrometer QToF with intermediate resolution and panels a, c, d, f–h are CID spectra acquired on a high resolution QToF mass spectrometer (Sciex TT6600). The peptide sequences above each panel indicate the assignment of predominantly singly charged y (red) and b (blue) ions. The modified residue is indicated in bold. K<sub>ε-carbE</sub> indicates the molecular weight difference between ions diagnostic of the modified lysine. a Lysine 65 of At2g38540. b Lysine 183 of AtCG00490. c Lysine 262 of At3g49120. d Lysine 268 of At3g49120. e Lysine 293 of At2g21330. f Lysine 251 of At3g54400. g Lysine 109 of At4g21280. h Lysine 208 of At4g25100.
**N-acetyl-lysine CO2 trapping with TEO.** α-N-acetyl-lysine (5 mg, 0.03 mmol) was dissolved in phosphate buffer (2 mL, 50 mM, pH 8.5). NaHCO3 (1.7 mg, 0.02 mmol) was dissolved in phosphate buffer (1 mL, 50 mM, pH 8.5) and added to the N-acetyl-lysine solution. The combined mixture was transferred to the Titration Manager, and Et3OBF4 (100 mg, 0.53 mmol) was added in three equal portions while the pH of the solution was maintained via the automated addition of NaOH solution (1 M). The mixture was stirred for 1 h after the final Et3OBF4 addition, then lyophilised and re-dissolved in methanol (1 mg/mL) for MS analysis. The sample was analysed using ESI-MS and the trapped carbamylated N-acetyl-lysine product was confirmed. ESI-MS: [M+H+] 289.17.

**Synthesis of α-N-acetyl-ε-N-ethylxycarbonyl-lysine.** α-N-acetyl-lysine (50 mg, 0.25 mmol) and NaHCO3 (50 mg, 0.60 mmol) were dissolved in D2O (1 mL). Ethyl chloroformate (27 mg, 0.25 mmol) in THF (5 mL) was added with stirring. The mixture was stirred overnight at room temperature, then the solvents were removed under reduced pressure. The precipitate was dissolved in acidic H2O (5 mL, pH 2) and the product was extracted into ether (2 x 5 mL). The ether extracts were dried (MgSO4) and the solvent was removed under reduced pressure to afford the e-N-ethylxycarbonyl-product (27.7 mg, 40%) as a white solid. 1H NMR (400 MHz, D2O) δ ppm: 4.29 (1H, dd, J = 9.0, 5.0 Hz a-CHNH), 4.07 (2H, q, J = 7.2 Hz CH2CH3), 3.11 (2H, t, J = 6.6 Hz ε-CH2NH), 2.03 (3H, s CH3CO), 1.90-1.68 (2H, m CH2CH3), 1.50 (2H, quintet, J = 6.8 Hz CH2CH3), 1.45-1.32 (2H, m, CH2CH2NH), 1.21 (3H, t, J = 7.1 Hz CH2CH3).

**Gly-Phe dipeptide trapping.** Gly-Phe (8 μg, 0.04 mmol) was dissolved in phosphate buffer (2 mL, 50 mM, pH 8.5). NaHCO3 (1.7 mg, 0.02 mmol) was dissolved in phosphate buffer (1 mL, 50 mM, pH 8.5), added to the dipeptide, and the mixture was transferred to the Titration Manager. A freshly made solution of Et3OBF4 (280 μg, 1.47 mmol) in D2O (1 mL) was added to the mixture in three portions while the pH was maintained by the automated addition of NaOH solution (1 M). The reaction mixture was stirred for 1 h, lyophilised and re-dissolved in methanol (1 mg/mL) for MS analysis. The sample was then analysed using ESI-MS and the trapped carbamylated Gly-Phe product was confirmed. ESI-MS: [M+H+] 323.01.

**FLKQ tetrapeptide trapping.** FLKQ (5 mg, 0.09 mmol) was dissolved in phosphate buffer (2 mL, 50 mM, pH 8.5). NaHCO3 (1.7 mg, 0.02 mmol) was dissolved in phosphate buffer (1 mL, 50 mM, pH 8.5), added to the tetrapeptide solution, and the mixture was transferred to the Titration Manager. A freshly made solution of Et3OBF4 (280 μg, 1.47 mmol) in D2O (1 mL) was added to the mixture in three portions while the pH was maintained by the automated addition of NaOH solution (1 M). The reaction mixture was stirred for 1 h then dialysed against D2O (1 L) overnight. The sample was then centrifuged, an aliquot (100 μL) was taken from the supernatant and digested using trypsin. ESI-MS data confirmed a trapped carbamate on the N-terminal peptide of the Hb β-chain.

**Haemoglobin trapping.** Human haemoglobin (Hb) (14.5 mg, 0.23 μmol) was dissolved in phosphate buffer (2 mL, 50 mM, pH 7.4). NaHCO3 (1.7 mg, 0.02 mmol) was dissolved in phosphate buffer (1 mL, 50 mM, pH 7.4), added to the protein solution, and the mixture was transferred to the Titration Manager. A freshly made solution of Et3OBF4 (280 μg, 1.47 mmol) in D2O (1 mL) was added to the mixture in three portions while the pH was maintained by the automated addition of NaOH solution (1 M). The reaction mixture was stirred for 1 h then dialysed against D2O (1 L) overnight. The sample was then centrifuged, an aliquot (100 μL) was taken from the supernatant and digested using trypsin. ESI-MS data confirmed a trapped carbamate on the N-terminal peptide of the Hb β-chain.

**Red blood cell trapping.** Red blood cells were separated from a rabbit blood sample by centrifugation. The red blood cells were dialysed into phosphate buffer (100 mM, pH 7.4) overnight. NaHCO3 (6.8 mg) was dissolved in phosphate buffer (1 mL, 50 mM, pH 7.4), added to the red blood cell solution (representing 3.88 mg total red blood cell protein), and the mixture was transferred to the Titration Manager. A freshly made solution of Et3OBF4 (280 μg, 1.47 mmol) in D2O (1 mL) was added to the mixture in three portions while the pH was maintained by the automated addition of NaOH solution (1 M). The reaction mixture was stirred for 1 h then dialysed against D2O (1 L) overnight. The sample was then centrifuged, an aliquot (100 μL) was taken from the supernatant and digested using trypsin. ESI-MS data confirmed a trapped carbamate on the N-terminal peptide of the Hb β-chain.

**Arabidopsis thaliana plant growth.** Arabidopsis seeds were plated onto 0.8% (w/v) plant agar containing 4.4 g/L Murashige and Skoog salt mixture and incubated at 4°C for 48 h in the dark. The seeds were then incubated at 22°C with 12 h of light per day for 5 weeks before planting into jiffy pellet soil plugs (LBS Horticulture) and grown at 22°C with 12 h of daylight for 5 weeks.

**Arabidopsis protein extraction.** Arabidopsis leaves (5 g dry weight) were ground in a pestle and mortar in the presence of liquid N2. Pre-chilled extraction

---

**Fig. 8** Biochemical validation of ATPRX34 as a carbamylated protein.

a) MS/MS spectra of peptides containing ethyl-trapped carbamates on Lys 262 and Lys 268 of At3g49120. The peptide sequences above the panel indicate the assignment of predominantly singly charged y (red) and z (blue) ions. The modified residue is indicated in bold. K_{carb.} indicates the molecular weight difference between ions diagnostic of the modified Lys. b) The ratios of the specific activities of wild type, K262A or K268A ATPRX34 protein at atmospheric CO2 or in the absence of CO2 (* p<0.0001 compared to wild type, one-way ANOVA with Dunn Bonferroni multiple comparison test, n=12 independent replicates, Kruskal-Wallis statistic = 26.4, ±S.D.; ** p < 0.0028 compared to wild type, one-way ANOVA with Dunn Bonferroni multiple comparison test, n=12 independent replicates, Kruskal-Wallis statistic = 26.4, ±S.D.). c) Inset—purification of recombinant AtPRX34 proteins (1.0 μg; SDS/PAGE analysis and Coomassie Blue staining). c) The ratios of the specific activities of wild type, K262E or K268E ATPRX34 protein at atmospheric CO2 or in the absence of CO2 (* p<0.0001 compared to wild type, one-way ANOVA with Dunn Bonferroni multiple comparison test, n=9 independent replicates, Kruskal-Wallis statistic= 19.53, ±S.D.; ** p<0.0194 compared to wild type, one-way ANOVA with Dunn Bonferroni multiple comparison test, n=9 independent replicates, Kruskal-Wallis statistic= 19.53, ±S.D.). Inset- Purification of recombinant PRX34 proteins (1.0 μg; SDS/PAGE analysis and Coomassie Blue staining)

Manager (Radiometer Analytical) and incubated at 25°C with stirring. Triethylxylon tetrafluoroborate (Et3OBF4, various amounts; details below) was added stepwise with a constant pH being maintained (pH 7.4) through the slow addition of 1 M NaOH solution via the automatic burette. The reaction mixture was stirred, and the pH maintained, for 1 h after the final Et3OBF4 addition in order to ensure that all TEO was hydrolysed.
phosphate buffer (4 °C, 100 mM, 15 mL, pH 7.4) was added to the leaves with sand and poly(vinylpoly)pyrrolidone (PVPP) and further grinding was performed. The mixture was passed through Miracloth (Millipore) on ice, and the filtrate was centrifuged at 4500 g for 10 min at 4 °C. The supernatant, containing soluble proteins, was used for trapping experiments.

**Arabidopsis thaliana leaf lyase trapping.** Extracted protein solution (3 mg, Bradford Assay) was dissolved in phosphate buffer (2 mL, 50 mM, pH 7.4). NaHCO3 (1.7 mg, 0.02 mmol) was dissolved in phosphate buffer (1 mL, 50 mM, pH 7.4). The protein solution, and the mixture was transferred to the Titration Manager. A freshly made solution of EtOBEA (280 mg, 1.47 mmol) in dH2O (1 mL) was added to the mixture in three portions while the pH was maintained by the automated addition of NaOH solution (1 M). The reaction mixture was stirred for 1 h then dialysed against dH2O (1 L) overnight. The sample was then centrifuged, an aliquot (100 µL) was taken from the supernatant. This was diluted to 1 µg mL⁻¹ and taken forward for trypsin digestion.

**AtPRX34 recombinant protein expression.** AtPRX341-353 wild type, K263A, K267Q and K268Q encoding open reading frames were cloned into the Ndel and BamHI sites of pET14b by commercial gene synthesis (Genscript). pET14b–AtPRX341-353 (AtPRX341-353 wild type, K263A, K267Q, K268Q and mutant proteins) was expressed in Escherichia coli BL21(DE3) cells at 20 °C for 16 h with 400 µM isopropyl-ß-D-thiogalactoside (IPTG). Pelleted bacteria (20 mL) were suspended in sonication buffer (137 mM NaCl, 2.7 mM KCl, 10 mM NaH2PO4, 1.8 mM KH2PO4, 5 mM dithiothreitol, 1% (v/v) Triton-X100), lysed by sonication (1.2 kW, 30 min, 3°C). The pellet was suspended in sonication buffer and sonication and pellet wash repeated three times. The pellet was subsequently suspended in denaturation buffer (6 M guanidine hydrochloride, 100 mM Tris HCl pH 8.5, 10 mM EDTA), further sonicated (180 s), and incubated at 4°C with gentle rocking. Protein was dialysed against refolding buffer (3.25 M urea, 50 mM Tris HCl pH 9.5, 5% (v/v) glycerol, 40 mM CaCl2, 0.7 mM oxidised glutathione, 0.21 mM reduced glutathione, 10 µM hemin) at 4°C for 16 h. Dialysed protein was incubated with Ni2⁺-nitritotriacetic acid resin (Qiagen) for 1 h. The resin was transferred to a column and allowed to settle while the flow-through was collected. The column was washed with three volume columns of wash buffer (50 mM Tris HCl pH 7.5, 100 mM NaCl, 10 mM Imidazole). The protein was then eluted with a gradient of imidazole in wash buffer (range 25–250 mM) with detection at 280 nm. The eluted fractions were analysed by SDS-PAGE (Supplementary Figure 3).

**AtPRX34 recombinant protein assay.** AtPRX341-353 wild type or mutant protein was added to a mixture of 50 mM Tris HCl pH 7.5, 1.6 mM 2-methoxophenol and 1.2 mM hydrogen peroxide. The mixture was incubated at 25°C for 30 min and activity monitored through the production of 1,2-benzoquinone by absorbance readings taken at 470 nm. CO2 minus experiments were performed in a CO2-scrubbed atmospheric chamber with reagents pre-equilibrated to remove CO2. Biologically independent experiments were performed on independently made preparations.

**Mass spectrometry and data handling.** Following the trapping reaction proteins were either digested using the filter aid sample preparation method (FASP) or using gel-aided sample preparation (GASP) as described without modifications34,35. The resulting peptide solution was desalted with home packed C18 stage tips36. The resulting peptide mixture was dried down and dissolved in 15 µL 4% acetonitrile (MeCN), 0.05% trifluoroacetic acid (TFA). 20% of this was analysed by LCMSMS carried out on a Qstar Pulsar J QTOF mass spectrometer (Sciex) coupled to an Ultimate 3000 nano-ESI source (Thermo) also coupled to an Orbitrap XL mass spectrometer (Thermo) that was allowed for as a variable modification in the mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomeexchange.org) via the PRIDE partner repository37 with the data set identifier PXD007753.

**Statistical analysis.** Data was analysed by two-tailed t-test or one-way ANOVA as indicated in the text. Normal distribution of the data for ANOVA of Fig. 8 was confirmed by the Shapiro–Wilk test. Homogeneity of variances was rejected for all factors examined using the Levene Test. Data were therefore examined using a non-parametric Kruskall–Wallis test with Dunn Bonferroni post-hoc test for pairwise comparisons with unequal variances.

**Data availability.** The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomeexchange.org) via the PRIDE partner repository37 with the data set identifier PXD007753. The datasets generated during the study are available from the corresponding author on reasonable request.

Received: 31 January 2018 Accepted: 3 July 2018
Published online: 06 August 2018

**References**


34. Wisniewski, J. R., Zougman, A. & Mann, M. Combination of FASP and stage-tip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. J. Proteome Res. 8, 5674–5678 (2009).


Acknowledgements

The authors thank George Lorimer and Nigel Robinson for helpful discussion. The authors thank Rebecca Eno for experimental assistance.

Author contributions

M.J.C. conceived the project. V.L.L., D.R.W.H., A.M.O.D., A.T., and M.J.C. designed the research. V.L.L., L.M.J., A.P.B., A.P., A.T., and D.W. performed the experiments. V.L.L., D.R.W.H., and M.J.C. analysed the data. M.J.C. wrote the manuscript with input from all authors.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-05475-z.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.