

ARTICLE

DOI: 10.1038/s41467-018-05475-z

OPEN

The identification of carbon dioxide mediated protein post-translational modifications

Victoria L. Linthwaite^{1,2}, Joanna M. Janus^{1,2}, Adrian P. Brown^{1,2}, David Wong-Pascua³, AnnMarie C. O'Donoghue^{2,3,4}, Andrew Porter⁵, Achim Treumann⁵, David R.W. Hodgson^{2,3,4} & Martin J. Cann^{1,2}

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.

¹Department of Biosciences, Durham University, South Road, Durham DH1 3LE, UK. ²Biophysical Sciences Institute, Durham University, South Road, Durham DH1 3LE, UK. ³Department of Chemistry, Durham University, South Road, Durham DH1 3LE, UK. ⁴Centre for Sustainable Chemical Processes, Durham University, South Road, Durham DH1 3LE, UK. ⁵NUPPA, The Protein Facility, Newcastle University, Cookson Building, Newcastle upon Tyne NE2 4HH, UK. Correspondence and requests for materials should be addressed to M.J.C. (email: m.j.cann@durham.ac.uk)

Protein functionalities can be extended and modulated by enzyme-catalysed and spontaneous post-translational modifications (PTMs)¹ such as phosphorylation, nitrosylation, acetylation, methylation, hydroxylation, glycosylation and the attachment of other small proteins. The earliest known PTM, the addition of CO₂ 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 CO₂ while Christiansen and co-workers showed that CO₂ uptake by the blood at constant *p*CO₂ was increased by the presence of O₂^{2,3}. Henriques then used kinetic evidence to postulate the direct combination of CO₂ with the free amino groups on haemoglobin⁴ and Ferguson and Roughton confirmed this through direct chemical analyses⁵. The site of CO₂ binding was demonstrated^{6,7} to occur at the Val-1β site⁸ linked to the O₂ binding state of the β-chain⁹. Physiologically, the reaction between the α-amino group and CO₂ stabilises the deoxygenated form of the protein¹⁰. Reaffirmation of the role of protein carbamylation is observed in Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) which fixes atmospheric CO₂ in plants using Mg²⁺ and inorganic carbon as co-factors¹¹. Experiments with RuBisCO demonstrated that CO₂ is acting not only as a substrate for the carboxylase reaction, but also as a co-factor that binds to an alternative site on the enzyme^{12–14}. Fixation of ¹⁴CO₂ to RuBisCO in complex with Mg²⁺ and a carboxyarabinitol bisphosphate carboxylase reaction intermediate confirmed CO₂-mediated carbamylation of an ε-amino group of lysine within the active site^{14,15}.

The work of Lorimer and co-workers provided a mechanism for carbamate formation whereby nucleophilic attack of a neutral amine on CO₂ converts the amine to an anionic group with the possibility for modulating protein activity¹⁴ (Fig. 1a). Thus the identification of carbamylation as a PTM in haemoglobin and RuBisCO led to the proposal that carbamylation of neutral *N*-terminal α-amino groups and the ε-amino group of lysine side chains could form the basis of a widespread mechanism for biological regulation¹⁶.

The carbamate modification is readily reversible but can be maintained by the protein environment, either through stabilising interactions such as is the case for RuBisCO or through a privileged pH environment such as within haemoglobin. Evidence for the frequency of carbamate formation on protein has been provided by computation that predicts as many as 1.3% of large proteins could bind CO₂ by carbamylation¹⁷. In support of a general biological relevance for such a mechanism of CO₂-binding in protein, several carbamates have been identified on lysine side chains in crystal structures including urease¹⁸, alanine racemase¹⁹, transcarboxylase 5S²⁰, class D β-lactamase²¹, and phosphotriesterase²². A recent carbamate was proposed in the Cx26 connexin hemi-channel, which opens in response to CO₂ in respiratory control and releases ATP²³. Here, carbamate formation on Lys125 is hypothesised to stabilise an electrostatic interaction with Arg104 in the neighbouring subunit of the hexamer thus constraining the hemi-channel in the open state in response to high [CO₂]. The lability of such exchangeable carbamates outside of the cellular environment makes the identification of this PTM a significant analytical challenge. Soft ionisation mass spectrometry techniques have had limited success in demonstrating the presence of carbamate PTMs but were unable to identify locations of modification²⁴.

Here we describe the development and use of a methodology for the trapping of carbamates under physiologically relevant conditions. The development of the carbamate-trapping method was established on amino acids, peptides, and proteins. The results demonstrate that labile carbamates can be selectively derivatised by ethylation thus stabilising them for downstream

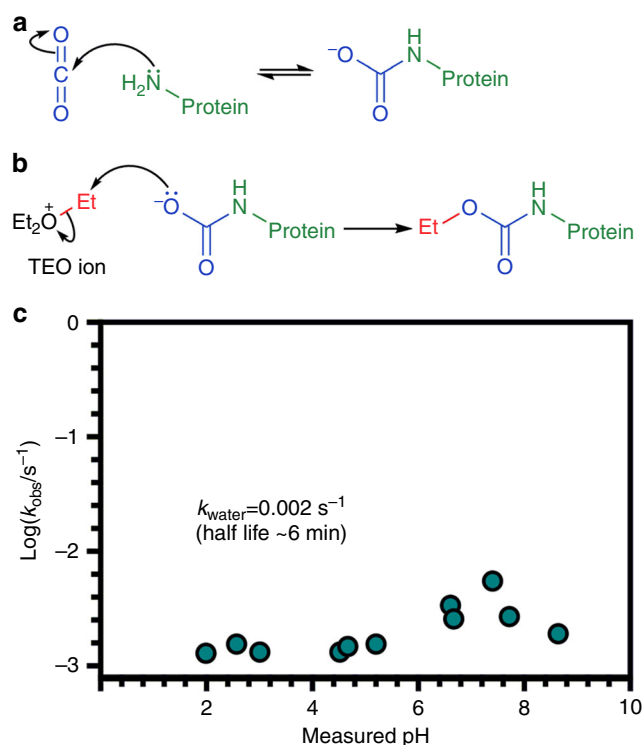


Fig. 1 Protein carbamate formation and trapping with TEO. **a** Carbamates form through the reversible reaction between CO₂ and neutral amine groups. **b** Proposed trapping mechanism of a protein carbamate with TEO. TEO transfers an ethyl group (red) to the anionic carbamate derived from CO₂ (blue) and protein primary amine (green). **c** Observed pseudo first-order rate constants, k_{obs} , for the hydrolysis of TEO plotted as a function of pH (where $k_{\text{obs}} = k_{\text{water}} + [k_{\text{hydroxide}} \cdot K_{\text{W}} / 10^{-\text{pH}}]$)

analysis. We demonstrate that the approach can identify the known carbamate on haemoglobin and go on to apply the method to the proteome of a model organism to identify CO₂-binding sites. We demonstrate the general principle that the activity of an identified protein is altered in response to CO₂ on mutation of the identified CO₂-binding site.

Our results lead us to propose that CO₂-mediated protein binding at exchangeable sites exists beyond haemoglobin. This may have profound consequences for our understanding of cellular responses to CO₂.

Results

Triethyloxonium ion-mediated carbamate trapping on amino acids and peptides.

The diazomethane derivative trimethylsilyldiazomethane (TMS-DAM) has been previously used to alkylate carbamates in CO₂-bubbled solutions containing amines²⁵; however, TMS-DAM is not water soluble so it is unsuitable for work with proteins. We therefore investigated a water-soluble reagent to alkylate and therefore trap labile CO₂ PTMs. Meerwein salts have been used to modify carboxylic groups on proteins through *O*-methylation²⁶. Limited kinetic data on the hydrolysis of the Meerwein reagent triethyloxonium (TEO) suggested a half-life of ~7 min in aqueous solution, and so we explored this system in more detail with a view to developing the reagent for trapping labile carbamates on protein (Fig. 1b). We used an indicating buffer method²⁷ to determine the pH-hydrolysis characteristics of TEO for 2 < pH < 9 and observed $k_{\text{water}} = 2.23 \times 10^{-3} \pm 1.22 \times 10^{-3} \text{ s}^{-1}$ (S.D), corresponding to a $t_{1/2}$ of approximately 6 min at physiological pH (Fig. 1c, Supplementary Figure 1–2, Supplementary Table 1–2, Supplementary Methods). This rate of

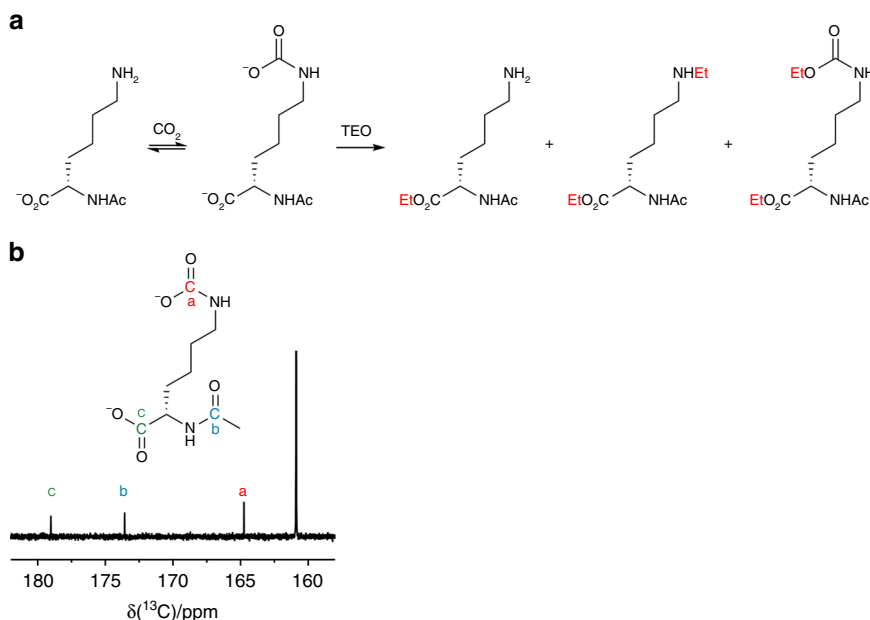


Fig. 2 Carbamate trapping on the ϵ -NH₂ group of α -N-acetyl-lysine. **a** N -acetyl-lysine-carbamate formation and trapping (O -ethylation) by TEO. **b** ^{13}C -NMR spectrum demonstrating the formation of a carbamate on N -acetyl-lysine by the appearance of a peak at 164 ppm

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 $^{13}\text{CO}_2/\text{H}^{13}\text{CO}_3^-$ at pH 8.5 and the formation of a carbamate was confirmed by the presence of a peak at 164 ppm by ^{13}C NMR spectroscopy (Fig. 2b)²⁸. Separately, TEO was added to an α - N -acetyl-lysine/ $\text{CO}_2/\text{HCO}_3^-$ 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 NaHCO_3 solution used to provide CO_2 was replaced with $\text{NaH}^{13}\text{CO}_3$, 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 -ethyloxycarbonyl-lysine (Figure 3ci). The product spectrum shows key signals consistent with N -carboxyethylation at $\delta \sim 4.25$ 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 first on the dipeptide Gly-Phe. A CO_2 -TEO trapping reaction of Gly-Phe at pH 8.5 was investigated by LC-ESI-MS and confirmed successful carbamate trapping on the N -terminus of the Gly-Phe dipeptide (Fig. 4a) alongside C -terminal O -ethylation and N -terminal ethylation side products. Carbamate trapping on tetrapeptide FLKQ was then investigated by LC-ESI-MS (Fig. 4b), yielding a trapped carbamate on either the N -terminus or the lysine side chain. Together, these data demonstrate that it is possible to form carbamates on the α - and ϵ -amino groups of peptides under physiologically relevant aqueous conditions and trap them by O -ethylation with TEO.

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 CO_2 binding that are exchangeable with the environment. We hypothesised that selective CO_2 binding to protein through carbamate formation would occur in structurally privileged sites that have evolved to facilitate carbamate formation. For example, CO_2 binding to haemoglobin at the Val-1 β 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 pK_a for 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 CO_2 trapping on the N -terminal valine of the haemoglobin β -chain. Carbamate formation was confirmed using ^{13}C -NMR spectroscopy by the observation of a signal at 164 ppm which matched literature values²⁹ (Fig. 5a).

We trapped CO_2 onto human haemoglobin with TEO and analysed the trypsin-digested products by ESI-MS. A carbamate was identified on the α -amine of the β -chain N -terminal valine (peptide mass 1024.10 Da), consistent with the literature (Fig. 5b)⁸. 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 $[\text{CO}_2]$.

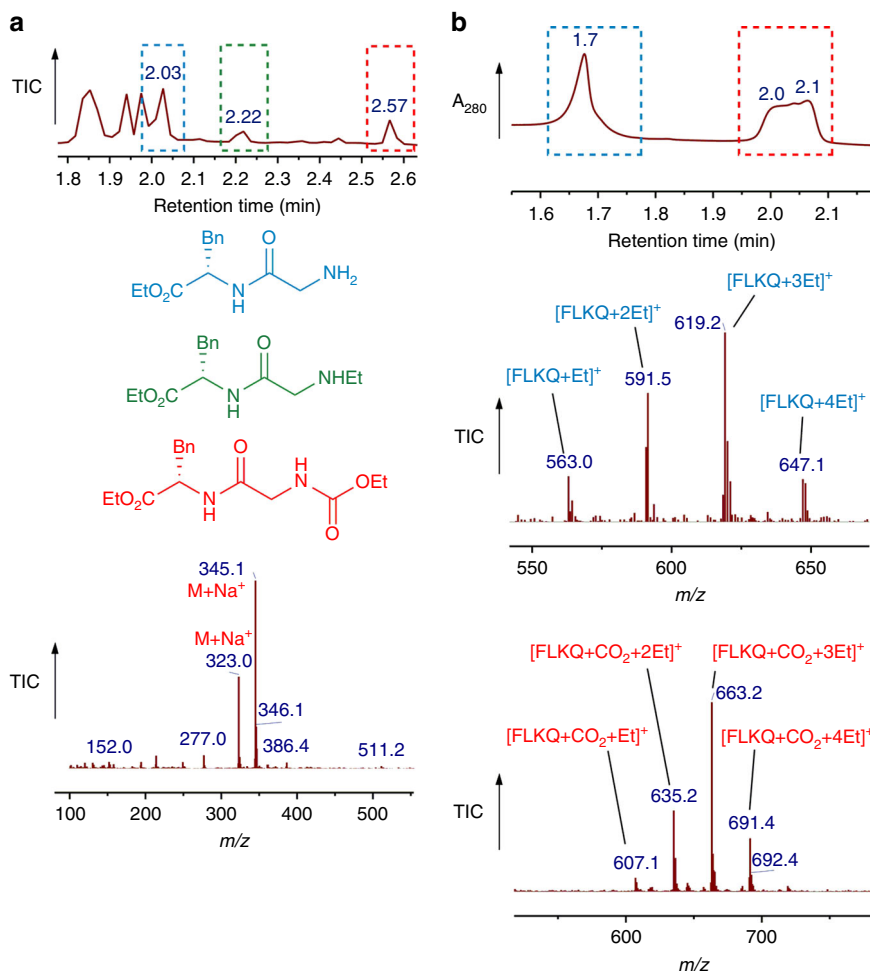


Fig. 4 Carbamate trapping on di- and tetrapeptides. **a** Total ion chromatogram of the carbamate-trapping reaction mixture of Gly-Phe with TEO, and m/z profile for species at retention time 2.57 min. The major products of the trapping reaction are Gly-Phe ethylated on the Phe- α -carboxylate (retention time 2.03 min), Gly-Phe ethylated on the α -carboxylate of Phe and the α -NH₂ group of Gly (retention time 2.22 min) and Phe-Gly ethylated on the carbamate of the α -NH₂-group of Gly and the α -carboxylate group of Phe (retention time 2.57 min). **b** Total ion chromatogram of the carbamate-trapping reaction mixture of FLKQ tetrapeptide with TEO and m/z profiles for species at retention time 1.7 and 2.0–2.1 min. The major products of the trapping reaction are shown as FLKQ with 1–4 ethylation groups (retention time 1.7 min) and FLKQ with 1 trapped carbamate and 1–4 ethylation groups (retention time 2.0–2.1 min)

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). Little ¹⁴CO₂ 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 of ¹⁴C, 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

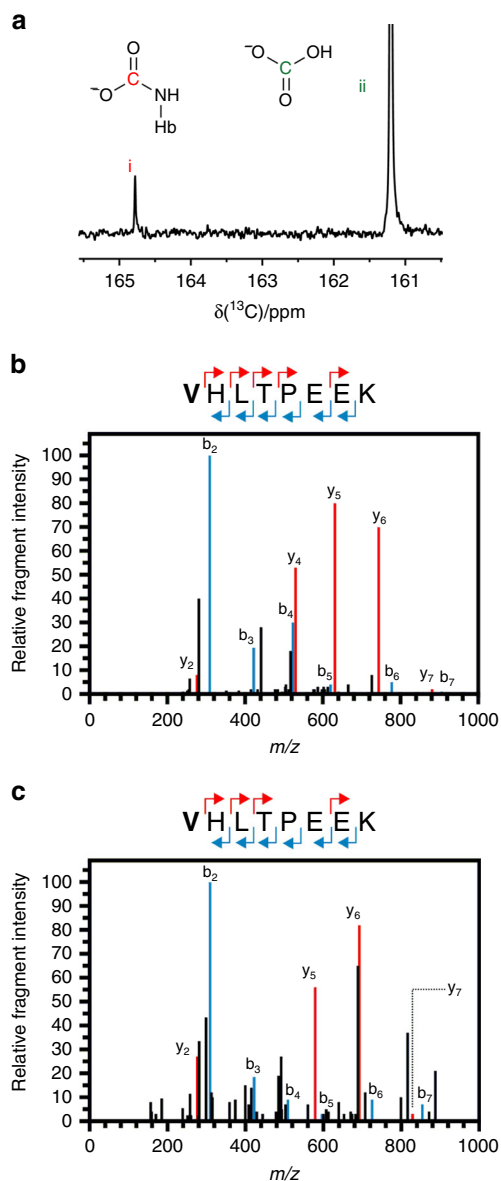


Fig. 5 Identification of the exchangeable CO₂-binding site on haemoglobin. **a** ¹³C-NMR spectrum demonstrating the formation of a carbamate on haemoglobin by the appearance of a peak at 164 ppm (i) together with a peak from H¹³CO₃⁻ in solution (ii). **b** A plot of relative fragment intensity versus 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

before TEO addition and *N*-ethylation of the lysine amino group can only occur on an unmodified lysine so therefore does not compete with carbamate trapping.

Two additional criteria were used to support the identification of carbamylated proteins and eliminate false positives. First, modified residues were discarded unless at least two y-ions and

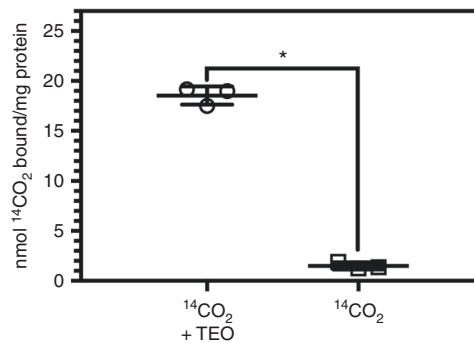


Fig. 6 Identification of exchangeable CO₂-binding site on *Arabidopsis* protein extract. ¹⁴CO₂ trapped onto protein extract of *Arabidopsis thaliana* (**p* < 0.0001, two-tailed *t*-test, *n* = 3 independent replicates, *t* = 29.85, *df* = 4, ±S.E.M.)

Table 1 Carbamylated proteins in *A. thaliana*

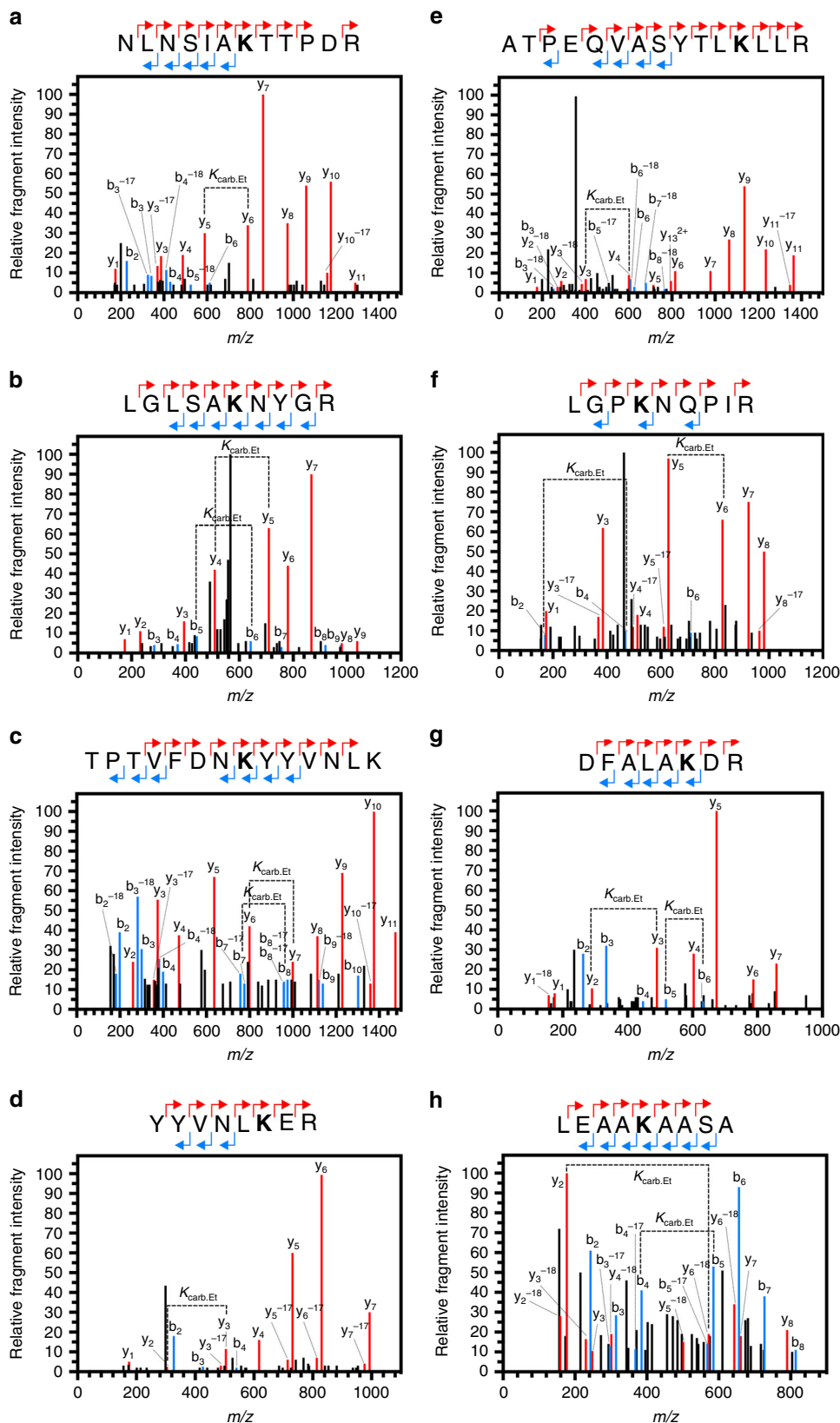
Genome Identification Number	Protein	Residue
At2g38540	Lipid-transfer protein	K65
AtCG00490	Rubisco Large Chain	K183
At3g49120	Peroxidase	K262 K268
At2g21330	FBA1	K293
At3g54400	Eukaryotic aspartyl protease family protein	K251
At4g21280	PSBQA	K109
At4g25100	Fe Superoxide dismutase 1	K208

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 acetylation³¹. 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 TPTVFDN**K**YYVNLK, proposed carbamylation on K262; MSMS peptide amino acids 263–270 YYVNL**K**ER, proposed carbamylation on K268). We further identified lysine carbamylation on K262 and K268 simultaneously (MSMS peptide amino acids 255–268 TPTVFDN**K**YYVNL**K**ER, 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 plant³². 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*

K262A, and *AtPRX34* K268A proteins by measuring their ability to oxidise 2-methoxyphenol in the presence of H₂O₂ under conditions of atmospheric CO₂ (approximately 12 μM CO₂) and in the absence of CO₂. We compared the ratio of the specific activities for each protein in the presence and absence of CO₂ (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 CO₂ compared to no CO₂. However, both the K262A and K268A mutants demonstrated elevated activity under conditions of atmospheric CO₂ compared to the absence of CO₂. This result suggests a control mechanism similar to that found within haemoglobin. Carbamylation of the two



haemoglobin Val-1 β 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 0.364 \pm 0.027, 0.479 \pm 0.12 and 0.500 \pm 0.099 μ mol 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 observed³³. 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 \pm 0.060, 1.477 \pm 0.027 and 0.933 \pm 0.020 μ mol 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 (urease¹⁸, alanine racemase¹⁹, transcarboxylase 5S²⁰, class D β -lactamase²¹, and phosphotriesterase²²) 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 significant¹⁴ CO₂ binding to protein dependent upon carbamylation 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 enable 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 QqTOF with intermediate resolution and panels **a**, **c**, **d**, **f–h** are CID spectra acquired on a high resolution QqTOF 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. $K_{carb,Et}$ 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

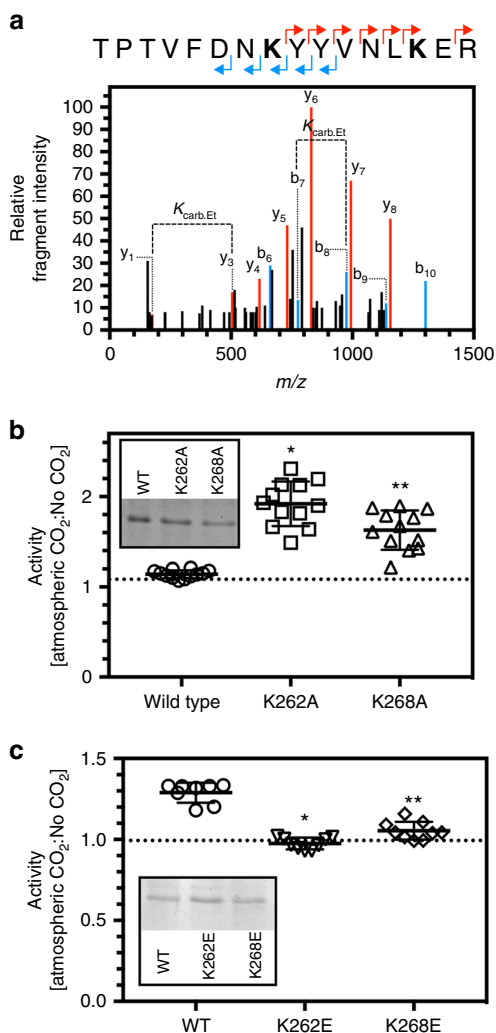


Fig. 8 Biochemical validation of AtPRX34 as a carbamylated protein. **a** MSMS spectra of peptides containing ethyl-trapped carbamates on Lys 262 and Lys 268 of At3g49120. The peptide sequences above the panel indicate the assignment of predominantly singly charged y (red) and b (blue) ions. The modified residue is indicated in bold. $K_{\text{carb.Et}}$ 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 CO_2 or in the absence of CO_2 (* $p < 0.0001$ compared to wild type, one-way ANOVA with Dunn Bonferroni multiple comparison test, $n = 12$ independent replicates, Kruskal-Wallis statistic = 26.4, \pm 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, \pm S.D.). 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 CO_2 or in the absence of CO_2 (* $p < 0.0001$ compared to wild type, one-way ANOVA with Dunn Bonferroni multiple comparison test, $n = 9$ independent replicates, Kruskal-Wallis statistic = 19.53, \pm 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, \pm 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. Triethyloxonium tetrafluoroborate (Et_3OBF_4 ; 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 Et_3OBF_4 addition in order to ensure that all TEO was hydrolysed.

α -N-acetyl-lysine CO_2 trapping with TEO. α -N-acetyl-lysine (5 mg, 0.03 mmol) was dissolved in phosphate buffer (2 mL, 50 mM, pH 8.5). NaHCO_3 (1.7 mg, 0.02 mmol) was dissolved in phosphate buffer (1 mL, 50 mM, pH 8.5) and added to the α -N-acetyl-lysine solution. The combined mixture was transferred to the Titration Manager, and Et_3OBF_4 (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 Et_3OBF_4 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 α -N-acetyl-lysine product was confirmed. ESI-MS: $[\text{M}+\text{H}^+]$ 289.17.

Synthesis of α -N-acetyl- ϵ -N-ethoxycarbonyl-lysine. α -N-acetyl-lysine (50 mg, 0.25 mmol) and NaHCO_3 (50 mg, 0.60 mmol) were dissolved in dH_2O (1 mL). Ethyl chloroformate (27 mg, 0.25 mmol) in THF (3 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 acidified H_2O (5 mL, pH 2) and the product was extracted into ether (2×5 mL). The ether extracts were dried (MgSO_4) and the solvent was removed under reduced pressure to afford the ϵ -N-ethoxycarbonyl-product (27.7 mg, 40%) as a white solid. ^1H NMR (400 MHz, D_2O) δ /ppm 4.29 (1H, dd, $J = 9.0, 5.0$ Hz α -CHNH), 4.07 (2H, q, $J = 7.2$ Hz CH_2CH_3), 3.11 (2H, t, $J = 6.6$ Hz ϵ - CH_2NH), 2.03 (3H, s CH_3CO), 1.90–1.68 (2H, m CH_2CH), 1.50 (2H, quintet, $J = 6.8$ Hz $\text{CH}_2\text{CH}_2\text{CH}_2$), 1.45–1.32 (2H, m, $\text{CH}_2\text{CH}_2\text{NH}$), 1.21 (3H, t, $J = 7.1$ Hz CH_3CH_2).

Gly-Phe dipeptide trapping. Gly-Phe (8 mg, 0.04 mmol) was dissolved in phosphate buffer (2 mL, 50 mM, pH 8.5). NaHCO_3 (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 Et_3OBF_4 (280 mg, 1.47 mmol) in dH_2O (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 analysed using ESI-MS and the trapped carbamate Gly-Phe product was confirmed. ESI-MS: $[\text{M}+\text{H}^+]$ 323.01.

FLKQ tetrapeptide trapping. FLKQ (5 mg, 0.009 mmol) was dissolved in phosphate buffer (2 mL, 50 mM, pH 8.5). NaHCO_3 (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 Et_3OBF_4 (280 mg, 1.47 mmol) in dH_2O (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 lyophilised and re-dissolved in methanol (1 mg/mL) for MS analyses. The sample was then analysed using ESI-MS and the trapped carbamate FLKQ was confirmed. ESI-MS: $[\text{M}+\text{H}^+]$ 663.16.

Haemoglobin trapping. Human haemoglobin (Hb) (14.5 mg, 0.23 μmol) was dissolved in phosphate buffer (2 mL, 50 mM, pH 7.4). NaHCO_3 (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 Et_3OBF_4 (280 mg, 1.47 mmol) in dH_2O (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 dH_2O (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. NaHCO_3 (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 Et_3OBF_4 (280 mg, 1.47 mmol) in dH_2O (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 dH_2O (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 for one week before planting into jiffy pellet sub 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 N_2 . Pre-chilled extraction

phosphate buffer (4 °C, 100 mM, 15 mL, pH 7.4) was added to the leaves with sand and poly(vinylpyrrolidone) (PVPP) and further grinding was performed. The mixture was passed through Miraclath (Millipore) on ice, and the filtrate was centrifuged at 4500g for 10 min at 4 °C. The supernatant, containing soluble proteins, was used for trapping experiments.

Arabidopsis thaliana leaf lysate 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), added to the protein solution, and the mixture was transferred to the Titration Manager. A freshly made solution of Et₃OBF₄ (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. AtPRX34₃₁₋₃₅₃ wild type, K262A, K268A, K262Q and K268Q encoding open reading frames were cloned into the NdeI and BamHI sites of pET14b by commercial gene synthesis (Genscript). pET14b-AtPRX34₃₁₋₃₅₃ (AtPRX34₃₁₋₃₅₃ wild type, K262A, K268A, K262Q and K268Q mutant proteins) was expressed in *Escherichia coli* BL21 (DE3) pLysS at 20 °C for 16 h with 400 µM isopropyl-β-D-thiogalactoside (IPTG). Pelleted bacteria (20 mL) were suspended in sonication buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 5 mM dithiothreitol, 1% (v/v) Triton-X100), lysed by sonication (180 s) and centrifuged (75,500g, 30 min, 4 °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⁺-nitrilotriacetic 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 column volumes of wash buffer (50 mM Tris pH 7.5, 100 mM NaCl, 10 mM Imidazole). The protein was then eluted with a gradient of imidazole in wash buffer (range 50–250 mM) with detection at 280 nm. The eluted fractions were analysed by SDS-PAGE (Supplementary Figure 3).

AtPRX34 recombinant protein assay. AtPRX34₃₁₋₃₅₃ wild type or mutant protein was added to a mixture of 50 mM Tris HCl pH 7.5, 1.6 mM 2-methoxyphenol 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 protein preparations.

Mass spectrometry and data handling. Following the trapping reaction proteins were either digested using the filter aided sample preparation method (FASP) or using gel-aided sample preparation (GASP) as described without modifications^{34,35}. 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 either on a Qstar Pulsar i QqTOF mass spectrometer (Sciex) coupled to an Ultimate 3000 nano-HPLC instrument, an LTQ Orbitrap XL mass spectrometer (Thermo) also coupled to an Ultimate 3000 nano-HPLC instrument or a TripleTOF 6600 QqTOF mass spectrometer (Sciex) coupled to an Eksigent ekspert nanoLC 425 instrument. Peptides were concentrated on a Pepmap C18 trap column (300 µm ID × 5 mm) and separated on a Pepmap C18 reversed phase analytical column (Dionex, UK) (3 µm particles, 75 µm ID × 250 mm). Separations were carried out at 300 nL min⁻¹ on a linear gradient from 96% A (0.05% formic acid), 4% B (0.05% formic acid, 80% acetonitrile) to 35% B over 70 min followed by a second linear gradient from 35% B to 65% B over 7 min. Peptides eluted from the LC gradient were injected online to the mass spectrometers using the following parameters: QStar (IDA mode, mass range 300–1600 Da, MS accumulation time 1 s, ion source voltage 2300 V, 3 MSMS spectra per cycle, MSMS mass range 100–1600 Da, MSMS accumulation time 3 s), Orbitrap LTQ XL (lock mass enabled, mass range 400–1800 Da, resolution 60,000 at 400 Da, 10 MSMS spectra per cycle, collision induced dissociation (CID) at 35% normalised CE, rejection of singly charged ions), TripleTOF 6600 (IDA mode, mass range 400–1250 Da, MS accumulation time 250 ms, ion source voltage 2400 V, 30 MSMS spectra per cycle, MSMS mass range 100–1500 Da, MSMS accumulation time 50 ms).

Raw data (.raw and .wiff) were converted to *.mgf peak lists using MSconvert³⁷. Peak lists were then searched using X!Tandem³⁰ against a concatenated database consisting of the *A. thaliana* proteome (Uniprot entry version 139 updated 07/07/

2017), cRAP (version 01/01/2012)³⁸ and the reverse fasta files for both³⁰. The following search parameters were used: Parent mass error, 20 ppm. Fragment mass error, 30 ppm (0.4 Da for QStar and Orbitrap XL data). Missed cleavage sites allowed, (1) Fixed modification of + 57.0215@C. Oxidation of methionine was allowed for as a variable modification. Three rounds of refinement were carried out to detect further modified peptides permitting additional variable modifications in round 1 (0.98401@N, 79.9663@S, 79.9663@T, 79.9663@Y, 15.999@M, 15.999@W, 0.98401@Q, 14.016@C, 14.016@D, 14.016@E, 14.016@H, 28.0313@K, 28.0313@E, 28.0313@D), in round 2 (14.016@D, 31.9898@M, 31.9898@W, -18.011@S, -18.011@T, -57.021@C, 14.016@Q, 57.021@H, 57.021@D, 57.021@E, 72.0211@K) and in round 3 (56.0626@K). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository³⁹ 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 Kruskal–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.proteomexchange.org>) via the PRIDE partner repository³⁸ 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

- Prabakaran, S., Lippens, G., Steen, H. & Gunawardena, J. Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdiscip. Rev. Syst. Biol. Med* **4**, 565–583 (2012).
- Bohr, C., Hasselbalch, K. & Krogh, A. About a new biological relation of high importance that the blood carbonic acid tension exercises on its oxygen binding. *Skand. Arch. Physiol.* **16**, 402–412 (1904).
- Christiansen, J., Douglas, C. G. & Haldane, J. S. The absorption and discussion of carbon dioxide by human blood. *J. Physiol.* **48**, 244–271 (1914).
- Henriques, O. M. The bonding method of carbon-dioxide in the blood. V. Preliminary announcement: some physiological observations on the carbohaemoglobin problem. *Biochem. Z.* **200**, 22–24 (1928).
- Ferguson, J. K. W. & Roughton, F. J. W. The direct chemical estimation of carbamino compounds of CO₂ with haemoglobin. *J. Physiol.* **83**, 68–86 (1935).
- Kilmartin, J. V. & Rossi-Bernardi, L. Inhibition of CO₂ combination and reduction of the Bohr effect in haemoglobin chemically modified at its alpha-amino groups. *Nature* **222**, 1243–1246 (1969).
- Kilmartin, J. V. & Rossiber, L. Binding of carbon dioxide by horse haemoglobin. *Biochem. J.* **124**, 31–3 (1971).
- Matthew, J. B., Morrow, J. S., Wittebort, R. J. & Gurd, F. R. Quantitative determination of carbamino adducts of alpha and beta chains in human adult hemoglobin in presence and absence of carbon monoxide and 2,3-diphosphoglycerate. *J. Biol. Chem.* **252**, 2234–2244 (1977).
- Morrow, J. S., Matthew, J. B., Wittebort, R. J. & Gurd, F. R. Carbon 13 resonances of ¹³CO₂ carbamino adducts of alpha and beta chains in human adult hemoglobin. *J. Biol. Chem.* **251**, 477–484 (1976).
- Louie, G., Tran, T., Englander, J. J. & Englander, S. W. Allosteric energy at the hemoglobin beta chain C terminus studied by hydrogen exchange. *J. Mol. Biol.* **201**, 755–764 (1988).
- Andrews, T. J., Badger, M. R. & Lorimer, G. H. Factors affecting interconversion between kinetic forms of ribulose diphosphate carboxylase-oxygenase from spinach. *Arch. Biochem. Biophys.* **171**, 93–103 (1975).
- Cleland, W. W., Andrews, T. J., Gutteridge, S., Hartman, F. C. & Lorimer, G. H. Mechanism of Rubisco: the carbamate as general base. *Chem. Rev.* **98**, 549–562 (1998).
- Lorimer, G. H., Badger, M. R. & Andrews, T. J. The activation of ribulose-1,5-bisphosphate carboxylase by carbon dioxide and magnesium ions. Equilibria, kinetics, a suggested mechanism, and physiological implications. *Biochemistry* **15**, 529–536 (1976).
- Lorimer, G. H. & Mizioroko, H. M. Carbamate formation on the epsilon-amino group of a lysyl residue as the basis for the activation of ribulosebisphosphate carboxylase by CO₂ and Mg²⁺. *Biochemistry* **19**, 5321–5328 (1980).

15. Lundqvist, T. & Schneider, G. Crystal structure of the ternary complex of ribulose-1,5-bisphosphate carboxylase, Mg(II), and activator CO₂ at 2.3-Å resolution. *Biochemistry* **30**, 904–908 (1991).
16. Lorimer, G. H. Carbon dioxide and carbamate formation—the makings of a biochemical control-system. *Trends Biochem. Sci.* **8**, 65–68 (1983).
17. Jimenez-Morales, D., Adamian, L., Shi, D. & Liang, J. Lysine carboxylation: unveiling a spontaneous post-translational modification. *Acta Crystallogr. D. Biol. Crystallogr.* **70**, 48–57 (2014).
18. Yamaguchi, K. & Hausinger, R. P. Substitution of the urease active site carbamate by dithiocarbamate and vanadate. *Biochemistry* **36**, 15118–15122 (1997).
19. Morollo, A. A., Petsko, G. A. & Ringe, D. Structure of a Michaelis complex analogue: propionate binds in the substrate carboxylate site of alanine racemase. *Biochemistry* **38**, 3293–3301 (1999).
20. Hall, P. R. et al. Transcarboxylase 5S structures: assembly and catalytic mechanism of a multienzyme complex subunit. *EMBO J.* **23**, 3621–3631 (2004).
21. Golemi, D., Maveyraud, L., Vakulenko, S., Samama, J. P. & Mobashery, S. Critical involvement of a carbamylated lysine in catalytic function of class D beta-lactamases. *Proc. Natl Acad. Sci. USA* **98**, 14280–14285 (2001).
22. Benning, M. M., Kuo, J. M., Rauschel, F. M. & Holden, H. M. Three-dimensional structure of the binuclear metal center of phosphotriesterase. *Biochemistry* **34**, 7973–7978 (1995).
23. Meigh, L. et al. CO₂ directly modulates connexin 26 by formation of carbamate bridges between subunits. *eLife* **2**, e01213 (2013).
24. Terrier, P. & Douglas, D. J. Carbamino group formation with peptides and proteins studied by mass spectrometry. *J. Am. Soc. Mass Spec.* **21**, 1500–1505 (2010).
25. Ito, Y. & Ushitora, H. Trapping of carbamic acid species with (trimethylsilyl) diazomethane. *Tetrahedron* **62**, 226–235 (2006).
26. Paterson, A. K. & Knowles, J. R. The number of catalytically essential carboxyl groups in pepsin. Modification of the enzyme by trimethyloxonium fluoroborate. *Eur. J. Biochem.* **31**, 510–517 (1972).
27. Delley, R. J., O'Donoghue, A. C. & Hodgson, D. R. Hydrolysis studies of phosphodichloridate and thiophosphodichloridate ions. *J. Org. Chem.* **77**, 5829–5831 (2012).
28. O'Leary M., H., Jaworski, R. J. & Hartman, F. C. C nuclear magnetic resonance study of the CO₂ activation of ribulosebisphosphate carboxylase from *Rhodospirillum rubrum*. *Proc. Natl Acad. Sci. USA* **76**, 673–675 (1979).
29. Morrow, J. S., Keim, P. & Gurd, F. R. CO₂ adducts of certain amino acids, peptides, and sperm whale myoglobin studied by carbon 13 and proton nuclear magnetic resonance. *J. Biol. Chem.* **249**, 7484–7494 (1974).
30. Craig, R. & Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20**, 1466–1467 (2004).
31. Zee, B. M. & Garcia, B. A. Discovery of lysine post-translational modifications through mass spectrometric detection. *Essays Biochem.* **52**, 147–163 (2012).
32. Daudi, A. et al. The apoplastic oxidative burst peroxidase in *Arabidopsis* is a major component of pattern-triggered immunity. *Plant Cell* **24**, 275–287 (2012).
33. Cundari, T. R. et al. CO₂-formatics: how do proteins bind carbon dioxide? *J. Chem. Inf. Model.* **49**, 2111–2115 (2009).
34. Wisniewski, J. R., Zougman, A. & Mann, M. Combination of FASP and stagetip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J. Proteome Res.* **8**, 5674–5678 (2009).
35. Fischer, R. & Kessler, B. M. Gel-aided sample preparation (GASP)—a simplified method for gel-assisted proteomic sample generation from protein extracts and intact cells. *Proteomics* **15**, 1224–1229 (2015).
36. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906 (2007).
37. Holman, J. D., Tabb, D. L. & Mallick, P. Employing proteowizard to convert raw mass spectrometry data. *Curr. Protoc. Bioinforma.* **46**, 13.24.11–13.24.19 (2014).
38. Rappsilber, J., Ryder, U., Lamond, A. I. & Mann, M. Large-scale proteomic analysis of the human spliceosome. *Genome Res.* **12**, 1231–1245 (2002).
39. Vizcaino, J. A. et al. The Proteomics Identifications database: 2010 update. *Nucleic Acids Res.* **38**, D736–D742 (2010).

Acknowledgements

The authors thank George Lorimer and Nigel Robinson for helpful discussion. The authors thank Rebecca Eno for experimental assistance. This work was supported by EPSRC grant EP/K504336/1 and Leverhulme Trust grant RPG-2016-017.

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., J.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/>.

© The Author(s) 2018