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 triethylxonium 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 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 pCO2 was increased by the presence of O2.

Henriques then used kinetic evidence to postulate the direct combination of CO2 with the free amino groups on haemoglobin and Ferguson and Roughton confirmed this through direct chemical analyses. The site of CO2 binding was demonstrated to occur at the Val-1 β site linked to the O2 binding state of the β-chain. Physiologically, the reaction between the α-amino group and CO2 stabilises the deoxygenated form of the protein. 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 cofactors. 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.

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.

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 activity (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 regulation.

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 protein sites linked to the O2 binding state of the protein has been provided by computation that predicts as many as 1.3% of large protein.

Determination of the role of protein carbamylation as a PTM in haemoglobin and glycosylation was uncovered in Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) which fixes atmospheric CO2 in plants using Mg2+ and inorganic carbon as cofactors. 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.

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.

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 activity (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 regulation.

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 carbamylation. In support of a recent carbamate was proposed in the Cx26 channel.

A recent carbamate was proposed in the Cx26 channel, with evidence from direct chemical analyses confirming CO2-mediated carbamylation of an ε-amino group of lysine within the active site.

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 activity (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 regulation.

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 carbamylation. In support of a recent carbamate was proposed in the Cx26 channel, with evidence from direct chemical analyses confirming CO2-mediated carbamylation of an ε-amino group of lysine within the active site.

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 activity (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 regulation.

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 carbamylation. In support of a recent carbamate was proposed in the Cx26 channel, with evidence from direct chemical analyses confirming CO2-mediated carbamylation of an ε-amino group of lysine within the active site.

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 activity (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 regulation.

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 carbamylation. In support of a recent carbamate was proposed in the Cx26 channel, with evidence from direct chemical analyses confirming CO2-mediated carbamylation of an ε-amino group of lysine within the active site.

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 activity (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 regulation.

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 carbamylation. In support of a recent carbamate was proposed in the Cx26 channel, with evidence from direct chemical analyses confirming CO2-mediated carbamylation of an ε-amino group of lysine within the active site.

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 activity (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 regulation.

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 carbamylation. In support of a recent carbamate was proposed in the Cx26 channel, with evidence from direct chemical analyses confirming CO2-mediated carbamylation of an ε-amino group of lysine within the active site.

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 activity (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 regulation.
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 $^{13}\text{CO}_2^{-}/^{13}\text{CO}_3^{-}$ 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/CO$_2$/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 NaH$^{13}$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 $^1$H NMR spectrum (Figure 3cii) was compared to a chemically synthesised standard ε-ethoxy carbonyl-lysine (Figure 3ci). 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 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.

![Fig. 2](image)

**Fig. 2** Carbamate trapping on the ε-NH$_2$ 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.

Triethylxonium 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-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 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)9. 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 [CO$_2$].
Fig. 3 Characterisation of carbamate trapping on the ε-NH₃ 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-acetyllysine ethylated on the α-carboxylate group (retention time ~1.7 min), α-N-acetyl-lysine ethylated on the α-carboxylate and ε-NH₃ 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 CO₂ (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, CO₂ and TEO (key CH₂ 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-amino 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.

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).
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 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.

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 TPTVFDNKKYYVNLK, 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 TPTVFDNKKYYVNLKER, 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 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 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 0.364±0.027, 0.479±0.12 and 0.500±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 observed35. 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 (urease18, alanine racemase19, transcarboxylase SS20, 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 significant14CO2 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, carbamylated 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 mM, 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. Keal.E 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 At4g21000.
**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-c-N-ethyloxycarboxyl-lysine.** α-N-acetyl-lysine (50 mg, 0.25 mmol) and NaHCO3 (50 mg, 0.60 mmol) were dissolved in dH2O (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 H2O (5 mL, pH 2) and the product was extracted into ether (2 × 5 mL). The ether extracts were dried (MgSO4) and the solvent was removed under reduced pressure to afford the c-N-ethyloxycarboxyl-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 α-CNHN), 4.07 (2H, q, J = 7.7 Hz CH2(CH3)2), 3.11 (2H, t, J = 6.6 Hz ε-CH2), 2.03 (3H, s, CH3CO), 1.90–1.68 (2H, m CH2CH2), 1.30 (2H, quintet, J = 6.8 Hz CH2CH2), 1.45–1.32 (2H, m, CH2CH2NH), 1.21 (3H, t, J = 7.1 Hz CH3CH2).

**Gly-Phe dipeptide trapping.** Gly-Phe (8 mg, 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 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, 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, 0.009 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 7.4), added to the tetrapeptide, 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 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 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 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 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 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 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
phosphate buffer (4°C, 100 mM, 1.5 mL, pH 7.4) was added to the leaves with sand and poly(vinylpolypyrrolidone) (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.

**Arobidopsis 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) and poly(vinylpolypyrrolidone) (PVPP) and further grinding was performed. The phosphate buffer (4 °C, 100 mM, 15 mL, pH 7.4) was added to the leaves with sand and the mixture was transferred to the Tritation Manager. A freshly made solution of EtOBF3 (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.

**AfpXR34 recombinant protein expression.** AfpXR34L1−355, 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-thigalactoside (IPTG). Pelleted bacteria (20 mL) were suspended in sonic 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 (5 min, 75% power) and poly(vinylpolypyrrolidone) (PVPP) and further grinding was performed. The phosphate buffer (4 °C, 100 mM, 15 mL, pH 7.4) was added to the leaves with sand and the mixture was transferred to the Tritation Manager. A freshly made solution of EtOBF3 (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.

**AfpXR34 recombinant protein assay.** AfpXR34L1−355, 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-thigalactoside (IPTG). Pelleted bacteria (20 mL) were suspended in sonic 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 (5 min, 75% power) and poly(vinylpolypyrrolidone) (PVPP) and further grinding was performed. The phosphate buffer (4 °C, 100 mM, 15 mL, pH 7.4) was added to the leaves with sand and the mixture was transferred to the Tritation Manager. A freshly made solution of EtOBF3 (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.

**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 modifications35,36. The resulting peptide solution was desalted with home packed C18 stage tips36. The resulting peptide mixture was dried down and dissolved in 10 mM oxidised glutathione, 0.21 mM reduced glutathione, 10 µM hemin) at 4 °C for 30 min. The C18 stage tips36 were washed with three column volumes of 250 mM diphosphoglycerate. The column was washed with three column volumes of 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.

**References**

1. 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).
2. Bohr, C., Hasselbach, 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).
3. Christiansen, J., Douglas, C. G. & Haldane, J. S. The absorption and discussion of carbon dioxide by human blood. J. Physiol. 48, 244–271 (1914).
4. Henriques, O. M. The bonding method of carbon-dioxide in the blood. V. Preliminary announcement: some physiological observations on the carbon haemoglobin problem. Biochem. Z. 200, 22–24 (1928).
5. Ferguson, J. K. W. & Roughton, F. J. W. The direct chemical estimation of carboximino compounds of CO2 with haemoglobin. J. Physiol. 83, 68–86 (1935).
6. 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).
7. Kilmartin, J. V. & Rossiber, L. Binding of carbon dioxide by horse haemoglobin. Biochem. J. 124, 31–3 (1971).
8. Matthew, J. B., Morrow, J. S., Wittebort, R. J. & Gurd, F. R. Quantitative determination of carboximino 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).
9. Morrow, J. S., Matthew, J. B., Wittebort, R. J. & Gurd, F. R. Carbon 13 resonances of 13CO2 carboximino adducts of alpha and beta chains in human adult hemoglobin. J. Biol. Chem. 251, 477–484 (1976).
10. Moe, G. H., Tran, T., Englander, J. & Englander, S. W. Allosteric energy at the transferrin receptor 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
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-A 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. 65, 548–555 (2009).

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 carbamate site of alanine racemase. Biochemistry 38, 3293–3301 (1999).

20. Hall, P. R. et al. Transcarboxylase 55 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 carboxylated 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., Raushel, 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 Spectrom. 21, 1500–1505 (2010).

25. Ito, Y. & Ushitora, H. Trapping of carbamic acid species with (trimethylsilyl) phosphodichloridate and thiophosphodichloridate ions. J. Org. Chem. 59, 6144–6146 (1994).

26. Craig, R. & Beavis, R. C. TANDEM: matching proteins with tandem mass spectrometry. Trends Biochem. Sci. 36, 181–184 (2011).

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. Employing proteowizard to convert published maps and institutional affiliations.

32. Donoghue, A. C. & Hodgson, D. R. Hydrolysis studies of phosphodichloridate and thiophosphodichloridate ions. J. Org. Chem. 77, 5829–5831 (2012).

33. Rappsilber, J., Ryder, U., Lamond, A. I. & Mann, M. Large-scale proteomic analysis of the human spliceosome. Genome Res. 12, 1231–1245 (2002).

34. Vizcaíno, 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 support 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.

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