INTERFACE
FOCUS

A methodology for carbamate post-translational modification discovery and its application in Escherichia coli

Journal:	*Interface Focus*
Manuscript ID	RSFS-2020-0028.R1
Article Type:	Research
Date Submitted by the Author:	23-Jul-2020

Complete List of Authors:	Linthwaite, Victoria; Durham University, Biosciences Cann, Martin; Durham University, Biosciences

Subject:	Biochemistry < CROSS-DISCIPLINARY SCIENCES, Chemical biology < CROSS-DISCIPLINARY SCIENCES
Keywords:	Carbamylation, carbamate trapping, post-translational modification

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Author-supplied statements

Relevant information will appear here if provided.

Ethics

Does your article include research that required ethical approval or permits?:
This article does not present research with ethical considerations

Statement (if applicable):
CUST_IF_YES_ETHICS :No data available.

Data

It is a condition of publication that data, code and materials supporting your paper are made publicly available. Does your paper present new data?:
Yes

Statement (if applicable):
The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD019606 and 10.6019/PXD019606. The datasets generated during the study are available from the corresponding author on reasonable request.

Conflict of interest

I/We declare we have no competing interests

Statement (if applicable):
CUST_STATE_CONFLICT :No data available.

Authors' contributions

This paper has multiple authors and our individual contributions were as below

Statement (if applicable):
M.J.C conceived the project. V.L.L performed research. Both authors analysed data and wrote the manuscript.

For Review Only

A methodology for carbamate post-translational modification discovery and its application in *Escherichia coli*

Victoria L. Linthwaite[1*] and Martin J. Cann[1]

[1]Department of Biosciences, Durham University, South Road, Durham, DH1 3LE

* Address correspondence to: Victoria L. Linthwaite, Department of Biosciences, Durham University, South Road, Durham, DH1 3LE, United Kingdom. Phone: +44 (191) 3343985. E-mail: v.l.linthwaite@durham.ac.uk

Keywords: Carbamylation, carbamate trapping, post-translational modification

Abstract

Carbon dioxide can influence cell phenotypes through the modulation of signalling pathways. CO_2 regulates cellular processes as diverse as metabolism, cellular homeostasis, chemosensing and pathogenesis. This diversity of regulated processes suggests a broadly conserved mechanism for CO_2 interactions with diverse cellular targets. CO_2 is generally unreactive but can interact with neutral amines on protein under normal intracellular conditions to form a carbamate post-translational modification (PTM). We have previously demonstrated the presence of this PTM in a subset of protein produced by the model plant species, *Arabidopsis thaliana*. Here we describe a detailed methodology for identifying new carbamate PTMs in an extracted soluble proteome under biologically relevant conditions. We apply this methodology to the soluble proteome of the model prokaryote, *Escherichia coli* and identify new carbamate PTMs. The application of this methodology therefore supports the hypothesis that the carbamate PTM is both more widespread in biology than previously suspected and may represent a broadly relevant mechanism for CO_2-protein interactions.

1. Introduction

Carbon dioxide (CO_2) is essential to life[1]. It is at the beginning of every life process as a fundamental substrate of photosynthesis or chemosynthesis and is at the end of every life process as the product of aerobic respiration and post-mortem decay. CO_2 can bind protein via the carbamate post-translational modification (PTM). Carbamylation occurs by the nucleophilic attack of a neutral amine upon CO_2 (Figure 1, a). This reaction has been demonstrated to occur on both protein *N*-terminal α-amino groups (e.g. haemoglobin[2]) and the lysine ε-amino group (e.g. RuBisCO where

For Review Only

CO_2-binding is necessary for enzyme activation[3]). The carbamate PTM has been previously understudied due to the lability and ready reversibility of the modification causing challenges in its analysis. Previous whole-protein mass spectrometry has been carried out to investigate carbamate binding on small proteins, such as insulin, but with limited success[4]. The development of a gentler sampling method allowed the identification of CO_2-binding to small proteins and peptides but without information of the amino acid binding site[4]. PTMs are typically identified and analysed via sequential protease digestion and electrospray ionisation-tandem mass spectrometry (ESI-MSMS). However, these are fundamentally aggressive techniques and the labile carbamate is removed during the process. There has been some experimental success binding CO_2 to amine groups using the compound trimethylsilyl-diazomethane (TMS-DAM) but this work was carried out in organic solvents and under high pH conditions which are not fundamentally applicable to physiological conditions[5].

It is clear that a methodology is required to identify CO_2-binding carbamate sites compatible with cellular environments. Such a method would allow the investigation of the many pathways known to respond to CO_2 and identify their component ~~CO_2-binding~~ proteins where carbamylation might be the underlying mechanism of regulation. In order to investigate the extent of the carbamate PTM within a proteome we require a new methodology capable of improving the stability of this modification under biologically relevant conditions to generate a robust group for downstream analysis.

We have developed a method capable of trapping the carbamate modification using the reagent triethyloxonium tetrafluoroborate (TEO) (Figure 1, b). TEO is a crystalline salt which is soluble under aqueous conditions and therefore can be used in biologically relevant buffers for trapping experiments. We are therefore able to identify sites of ~~CO_2-binding~~carbamylation under conditions more representative of a cellular environment. The method involves the transfer of an ethyl group from the TEO reagent to a pre-formed carbamate and converts it into a group robust enough for tryptic digest and ESI-MSMS (Figure 1, c). The method was validated by identifying the carbamate known to form on the human haemoglobin beta chain *N*-terminus and then used to discover previously unknown ~~CO_2-binding~~carbamylated proteins in the soluble proteome of the model plant *Arabidopsis thaliana*[6].

Here we describe in detail how to use this methodology to identify new ~~CO_2-binding~~carbamylated proteins and provide evidence for the presence of exchangeable CO_2-binding carbamate sites in the soluble proteome of *Escherichia coli*. This supports the hypothesis that carbamylation is a broadly biologically relevant CO_2-mediated PTM.

2. Methods

2.1 Protein sample preparation

Wild type *E.coli* BL21 (DE3) overnight culture (5 mL in LB broth, 37 °C) was centrifuged at 13000 g for 5 minutes. The supernatant was discarded, and the cell pellet resuspended in phosphate buffer (3 mL, 50 mM sodium phosphate, 150 mM NaCl, pH

For Review Only

7.4) containing protease inhibitors (SigmaFAST protease inhibitor cocktail). The sample was sonicated for 5 s twice at 70 watts (Cole-Parmer, Ultrasonic processor) and centrifuged at 13000 g for 15 minutes to remove insoluble cellular material. The extracted soluble protein sample (500 µg) was pre-incubated with 20 mM sodium bicarbonate (pH 7.4) for 10 minutes at room temperature before carbamate trapping, to allow carbamates to form.

2.2 Carbamate trapping reaction

Extracted *E.coli* protein (500 µg, 3 mL 50 mM phosphate buffer) was added to a pH stat 5 mL cell (Titrando 902, Metrohm) with a pH probe and burette. TEO (280 mg, Sigma UK) was added in three step-wise increments in phosphate buffer (1 mL, 50 mM). The reaction pH was maintained at 7.4 (physiological for *E.coli* cells[7]) with the addition of 1 M NaOH (freshly prepared from pellets stored under N_2) throughout the addition of the trapping reaction. The reaction was monitored, and pH stabilised for 60 minutes until completion. The trapped protein sample was dialysed (SnakeSkin™ dialysis tubing, 3.5K MWCO, 22mm) overnight (1 L, dH_2O, 4 °C).

2.3 Tryptic digestion

Post-dialysis, the trapped reaction sample supernatant was removed using vacuum centrifugation for 1 hr at 30 °C. After drying, the protein sample was resuspended in 8 M urea (500 µL) and reduced with addition of dithiothreitol (DTT, 25 mM final concentration) at 37 °C for 1 h. The sample was then alkylated with iodoacetamide (40 mM final concentration) in the dark for 1 h at room temperature. This sample was centrifuged at 1000 g for 5 min at room temperature and the soluble supernatant removed. The sample was diluted to 1 M urea and digested with Trypsin gold (mass spectrometry grade, Promega) in a 1:25 (w/w) ratio overnight at 37 °C. The digested solution was desalted on a C18 column (solid phase extraction cartridge, Agilent) prior to injection and analysis~~ed~~ by ESI-MSMS on an LTQ Orbitrap XL mass spectrometer (Thermo) coupled to an Ultimate 3000 nano-HPLC instrument. Peptides eluted from the LC gradient were injected online to the mass spectrometer (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 capillary electrophoresis rejection of singly charged ions).

2.4 ESI-MSMS Analysis

The post-ESI-MSMS raw data files were converted into .mgf files using the freeware MSConvert provided by Proteowizard[8]. Once converted to .mgf files these files were analysed using PEAKS Studio 10.5 software[9] including the variable modifications ethylation (28.03 at D or E), carboxyethylation (72.02 at K or protein *N*-terminal groups), oxidation (M), acetylation (N-terminal) and the fixed modification

1
2
3 carbimidomethyl (C). This data was then refined using a false discovery rate (FDR) of
4 1 %, 2 unique peptides per protein and a PTM AScore of 50.

3. Application of the Method

Carbamates are covalently trapped onto protein by their modification with the reagent TEO. This reagent modifies the carbamate to produce a group sufficiently robust for downstream analysis. There are several variables to be considered for application with the trapping methodology. The most important of these is to maintain a biologically relevant pH throughout the trapping reaction to approximate the cellular environment. In order to facilitate this, the trapping reaction is performed within a pH stat (Titrando 902, Metrohm). The other important factor to consider is that the CO_2 concentration used in the trapping reaction is of biological relevance to the cell type being investigated. In this investigation we were probing *E.coli* cells using 20 mM HCO_3^-/CO_2. Cellular PCO_2 in *E.coli* in its native environment is relatively understudied. However, the PCO_2 in the natural *E.coli* growth environment (luminal walls of the colon) is estimated to be up to 40 kPa[10]. As arterial PCO_2 is approximately 5.0 kPa we are almost certainly working at a PCO_2 to which *E.coli* is physiologically exposed.

Due to its solubility, TEO reacts with hydroxide ions within the solvent (H_2O), leading to the production of H^+ during the reaction and a reduction in reaction pH. To prevent this, the pH stat is capable of counteracting acid group formation with the slow addition of 1 M NaOH throughout the trapping reaction. The work discussed here was carried out at pH 7.4. Due to the reactivity of the TEO reagent, it is also important to use a buffer which does not carry free amine groups or detergents, both of which could be a target for and deplete the TEO reagent. All trapping reactions are therefore performed in phosphate buffers to avoid this.

The TEO reagent has a half-life under aqueous conditions of 6 min, therefore the reaction time required for complete hydrolysis of the reagent, and thus safe handling of the reaction is 60 min[6]. After the experiment has reached completion the sample is dialysed into 1 L dH_2O. This dialysis is to remove reaction side products produced by the hydrolysis of the TEO reagent (ethanol and dimethyl ether) and all buffer salts which would otherwise interfere with protease digestion and ESI-MSMS analysis. Proteins may precipitate out of solution by the removal of the buffer salts, but this is unimportant as the reaction is complete and the protein groups for analysis already modified.

After dialysis, the sample is digested with the protease trypsin, analysed by ESI-MSMS and the data probed using the mass spectrometry software. The data presented here was analysed using the software PEAKS Studio 10.5. There are also a number of other MS software that can be used for data analysis depending on the search criteria of the user. For analysis of carbamate trapping data, it is important to prioritise software capable of handling a large number of variable modifications, a freeware alternative to PEAKS Studio is MaxQuant[11].

In the first instance carbamates are searched for via the variable addition of a 72.02 mass on any possible lysine side chain or protein *N*-terminus within the target

For Review Only

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

proteome. Carbamates are only formed at these specific sites due to their lower p*K*a values allowing for dissociation to a neutral amine and therefore CO_2 binding[12]. A software-based confidence score is used to remove false positives identified in the search. These carbamate ID's are then examined to remove the remaining false positives. The first step in this manual curation arises as a function of the choice of trypsin protease used in the digest. Trypsin cleaves proteins after a lysine or arginine side chain due to their positive charge. Removal of this charge and increased size of the side chain by the presence of a trapped carbamate thereby prevents trypsin cleavage. This method of searching for lysine PTMs by virtue of a missed cleavage site has been employed previously to search for lysine acetylation[13]. This method acts as an internal control for the analysis of possible carbamates identified by the mass spectrometry software. Therefore, if a carbamate is located at the *C*-terminus of a peptide, this carbamate is almost certainly a false positive (Figure 2, a). There can also be peptides that have several matching ions but do not have any fragment ions surrounding the carbamate modification itself. If the majority of the matching ions within a peptide spectrum are all located *N*- or *C*-terminal to the proposed carbamate, this is unlikely to be a true carbamate. Figure 2 b shows a spectrum with many matching y ions (red) however these ions are all matching masses *C*-terminal the site of the possible carbamate. Therefore, there is low confidence in the sequence on the peptide containing the possible carbamate.

Control untrapped (same experimental conditions but without the addition of TEO trapping reagent) experiments were performed on *E.coli* proteome samples and the MS-MS data searched for carbamates without an alkylation but none were found. This is unsurprising as the mass spectrometry conditions are known to be harsh enough to remove the untrapped CO_2 bound modification. Carbamate discovery is therefore dependent upon TEO-trapping.

To date, carbamate PTMs that are freely exchangeable with the environment and thus potential control sites to mediate CO_2 responses have been identified in haemoglobin and *Arabidopsis*[6,12]. In addition, a carbamate PTM on the connexin 26 gap junction is hypothesised to mediate mammalian respiratory responses to CO_2 although the PTM awaits direct analysis[14]. We hypothesise that the carbamate PTM is widespread in biology. We therefore deployed the TEO-trapping methodology to investigate whether exchangeable CO_2-binding carbamate sites might exist in *E. coli.* We screened the soluble proteome of *E.coli* (DE3) for new carbamate PTMs using the methodology and analysis criteria described. 0.5 mg of soluble protein from an overnight *E.coli* culture was trapped with TEO reagent, digested with trypsin and analysed by ESI-MSMS. Over five sample injections 294 proteins were identified with a 1 % FDR. We identified six CO_2-binding carbamate sites on five proteins in this small-scale screen of the *E. coli* proteome (Table 1, Figure 3). Carbamate PTMs that are exchangeable with the environment are therefore evident in prokaryotes.

4. Discussion and conclusion

Previous bacterial carbamates have been discovered on several proteins; these can be divided into either exchangeable or non-exchangeable carbamates. An

For Review Only

exchangeable carbamate binding site exists in a labile state and carbamate occupancy is presumed dependent on environmental PCO_2. A non-exchangeable binding site is metal ion coordinated which greatly increases its stabilisation and therefore is also not amenable to alkylation with the TEO reagent.

Several previously discovered carbamates are non-exchangeable metal ion-coordinated sites. Examples include allantoinase[15] which upon crystallisation was discovered to contain two metal ions (Fe bridged by a carbamate within the protein active site and MurD[16] which was demonstrated to contain a carbamylated lysine which helps stabilise interaction with Mg^{2+}. Other known examples include urease[17] and phosphotriesterase[18]. Any disruption of these stabilised carbamates through mutation of the lysine residue resulted in a loss of protein activity[15,16], thus demonstrating the importance of carbamate formation for protein function.

Other bacterial proteins which are known to form carbamates include alanine racemase[19] and beta-lactamases[20] where neither of the carbamates are stabilised by metal ion coordination. The carbamates are proposed to be stabilised by hydrogen bonding side chains (arginine for alanine racemase and tryptophan for beta-lactamase). Study of OXA-1 beta-lactamase demonstrated that increases in bicarbonate increased enzyme activity. Despite being exchangeable sites, these carbamates were not identified in this mass spectrometry screen and further work on increasing the proteome coverage is ongoing.

All of these previously discovered carbamate binding sites are located within protein active sites where their presence has been found to have a functional role. This emphasises the importance of the investigation of CO_2-binding carbamate sites as regulators of protein activity within a cellular environment.

Here we provide a detailed explanation of a recent methodology to trap CO_2 bound to protein under biologically relevant conditions. This work is an extension of our previous studies on *Arabidopsis* and demonstrates reversible ~~CO_2~~-carbamate binding within another organism.

Within *E.coli* both CO_2 and bicarbonate ions are essential for metabolic cellular processes[21]. We have identified carbamate PTMs on five *E.coli* proteins. These proteins represent a range of roles within the bacterial physiological functions, some of them exist in locations already known to interact with a changing CO_2 environment.

The 60 kDa chaperonin protein assists in the refolding of stress-denatured proteins[22]; one condition that causes stress is increased levels of CO_2 altering cellular pH levels[23]. This provides a potential link between stress responses and a CO_2-sensing mechanism. The DNA-binding protein HU-alpha is a histone-like DNA-binding protein which introduces negative supercoiling to DNA to prevent its denaturation under extreme conditions[24]. This HU regulon also regulates acid-stress genes[24].

Other proteins such as tryptophanase which is the protein responsible for synthesising indole and pyruvate from L-tryptophan[25], the glutamine-binding periplasmic protein, which is involved in transporting glutamine across the periplasmic space[26] and the ribose important binding protein involved in the ABC transporter complex made of three subunits. RbsB delivers ribose to the inner membrane complex

For Review Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

of RbsAC[27] are not involved in any cellular processes specifically identified as responsive to CO_2. This intriguing possibility awaits future investigation.

Computational modelling has been used to suggest that as many as 1.3% of proteins could bind CO_2 by carbamylation[28]. The discovered carbamates arose from a model trained using previously identified stable carbamates and are almost all entirely buried within their respective protein structures. Our discovered carbamates are, by definition, not buried as they must be in contact with bulk solvent for alkylation. The two methods are therefore likely to identify different subsets of carbamylation sites. This modelling, along with our screens of *Arabidopsis* and *E.coli* demonstrate that carbamylation is likely to be broadly relevant mechanism for protein-CO_2 interactions within both prokaryotes and eukaryotes.

For Review Only

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[29] partner repository with the data set identifier PXD019606 and 10.6019/PXD019606. The datasets generated during the study are available from the corresponding author on reasonable request.

Authors' contributions

M.J.C conceived the project. V.L.L performed research. Both authors analysed data and wrote the manuscript.

Competing interests

We declare no competing interests.

Funding

This work was supported by Leverhulme Trust grant RPG-2016-017 and Biotechnology and Biological Sciences Research Council grant BB/S015132/1.

Acknowledgements

We thank Andrew Porter for assistance with mass spectrometry.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

1. Sharabi, K. *et al.* Sensing, physiological effects and molecular response to elevated CO2 levels in eukaryotes. *J Cell Mol Med* **13**, 4304–4318 (2009).

2. Ferguson, J. K. & Roughton, F. J. The direct chemical estimation of carbamino compounds of CO(2) with haemoglobin. *J Physiol* **83**, 68–86 (1934).

3. Lorimer, G. H. & Miziorko, H. M. Carbamate formation on the .epsilon.-amino group of a lysyl residue as the basis for the activation of ribulosebisphosphate carboxylase by carbon dioxide and magnesium(2+). *Biochemistry* **19**, 5321–5328 (1980).

4. Terrier, P. & Douglas, D. J. Carbamino group formation with peptides and proteins studied by mass spectrometry. *Journal of The American Society for Mass Spectrometry* **21**, 1500–1505 (2010).

5. Ito, Y. & Ushitora, H. Trapping of carbamic acid species with (trimethylsilyl)diazomethane. *Tetrahedron* **62**, 226–235 (2006).

6. Linthwaite, V. L. *et al.* The identification of carbon dioxide mediated protein post-translational modifications. *Nat Commun* **9**, 3092–3092 (2018).

7. Zilberstein, D., Agmon, V., Schuldiner, S. & Padan, E. Escherichia coli intracellular pH, membrane potential, and cell growth. *J Bacteriol* **158**, 246–252 (1984).

8. Chambers, M. C. *et al.* A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* **30**, 918–920 (2012).

9. Ma, B. *et al.* PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Communications in Mass Spectrometry* **17**, 2337–2342 (2003).

10. Kolkman, J. J., Otte, J. A. & Groeneveld, A. J. Gastrointestinal luminal PCO2 tonometry: an update on physiology, methodology and clinical applications. *British journal of anaesthesia* **84**, 74–86 (2000).

11. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* **26**, 1367–1372 (2008).

12. Lorimer, G. H. Carbon dioxide and carbamate formation: the makings of a biochemical control system. *Trends in Biochemical Sciences* **8**, 65–68 (1983).

13. Zee, B. M. & Garcia, B. A. Discovery of lysine post-translational modifications through mass spectrometric detection. *Essays Biochem* **52**, 147–163 (2012).

14. Meigh, L. *et al.* CO_2directly modulates connexin 26 by formation of carbamate bridges between subunits. *Elife* **2**, e01213–e01213 (2013).

For Review Only

15. Kim, K., Kim, M.-I., Chung, J., Ahn, J.-H. & Rhee, S. Crystal structure of metal-dependent allantoinase from Escherichia coli. *J Mol Biol* **387**, 1067–1074 (2009).

16. Šink, R. *et al.* MurD enzymes: some recent developments. *Biomolecular Concepts* **4**, 539–556 (2013).

17. Lippard, S. J. At last--the crystal structure of urease. *Science* **268**, 996 (1995).

18. Buchbinder, J. L. *et al.* Biochemical Characterization and Crystallographic Structure of an Escherichia coli Protein from the Phosphotriesterase Gene Family. *Biochemistry* **37**, 5096–5106 (1998).

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. Leonard, D. A. *et al.* The role of OXA-1 β-lactamase Asp66 in the stabilization of the active-site carbamate group and in substrate turnover. *Biochemical Journal* **410**, 455–462 (2008).

21. Merlin, C., Masters, M., McAteer, S. & Coulson, A. Why is carbonic anhydrase essential to Escherichia coli? *J Bacteriol* **185**, 6415–6424 (2003).

22. Kerner, M. J. *et al.* Proteome-wide Analysis of Chaperonin-Dependent Protein Folding in Escherichia coli. *Cell* **122**, 209–220 (2005).

23. Baez, A., Flores, N., Bolívar, F. & Ramírez, O. T. Metabolic and transcriptional response of recombinant Escherichia coli to elevated dissolved carbon dioxide concentrations. *Biotechnology and Bioengineering* **104**, 102–110 (2009).

24. Oberto, J., Nabti, S., Jooste, V., Mignot, H. & Rouviere-Yaniv, J. The HU Regulon Is Composed of Genes Responding to Anaerobiosis, Acid Stress, High Osmolarity and SOS Induction. *PLOS ONE* **4**, e4367 (2009).

25. Vederas, J. C., Schleicher, E., Tsai, M. D. & Floss, H. G. Stereochemistry and mechanism of reactions catalyzed by tryptophanase Escherichia coli. *Journal of Biological Chemistry* **253**, 5350–5354 (1978).

26. Hsiao, C.-D., Sun, Y.-J., Rose, J. & Wang, B.-C. The Crystal Structure of Glutamine-binding Protein fromEscherichia coli. *Journal of Molecular Biology* **262**, 225–242 (1996).

27. Clifton, M. C. *et al.* In vitro reassembly of the ribose ATP-binding cassette transporter reveals a distinct set of transport complexes. *J Biol Chem* **290**, 5555–5565 (2015).

28. Jimenez-Morales, D., Adamian, L., Shi, D. & Liang, J. Lysine carboxylation: unveiling a spontaneous post-translational modification. *Acta Crystallographica Section D* **70**, 48–57 (2014).

For Review Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

29. Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **47**, D442–D450 (2019).

For Review Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figures

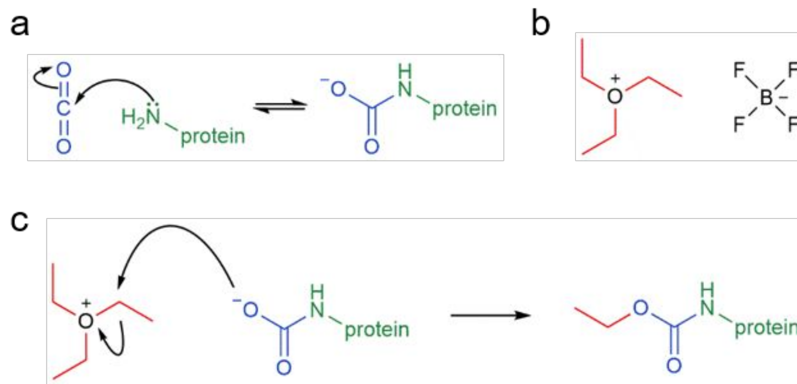


Figure 1. Formation of carbamate on a neutral amine group and the subsequent trapping with TEO. a Reversible reaction of CO_2 (blue) binding to a neutral amine on a protein (green). b TEO, ethyl groups highlighted in red. c Ethylation of carbamate by oxonium ion from TEO. An ethyl group is transferred from the oxonium ion to the negatively charged carbamate.

For Review Only

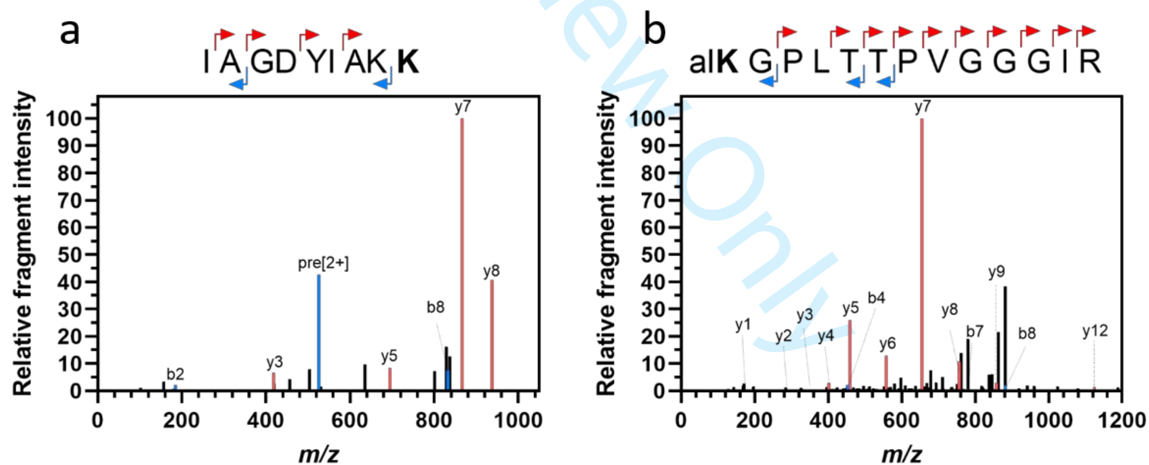
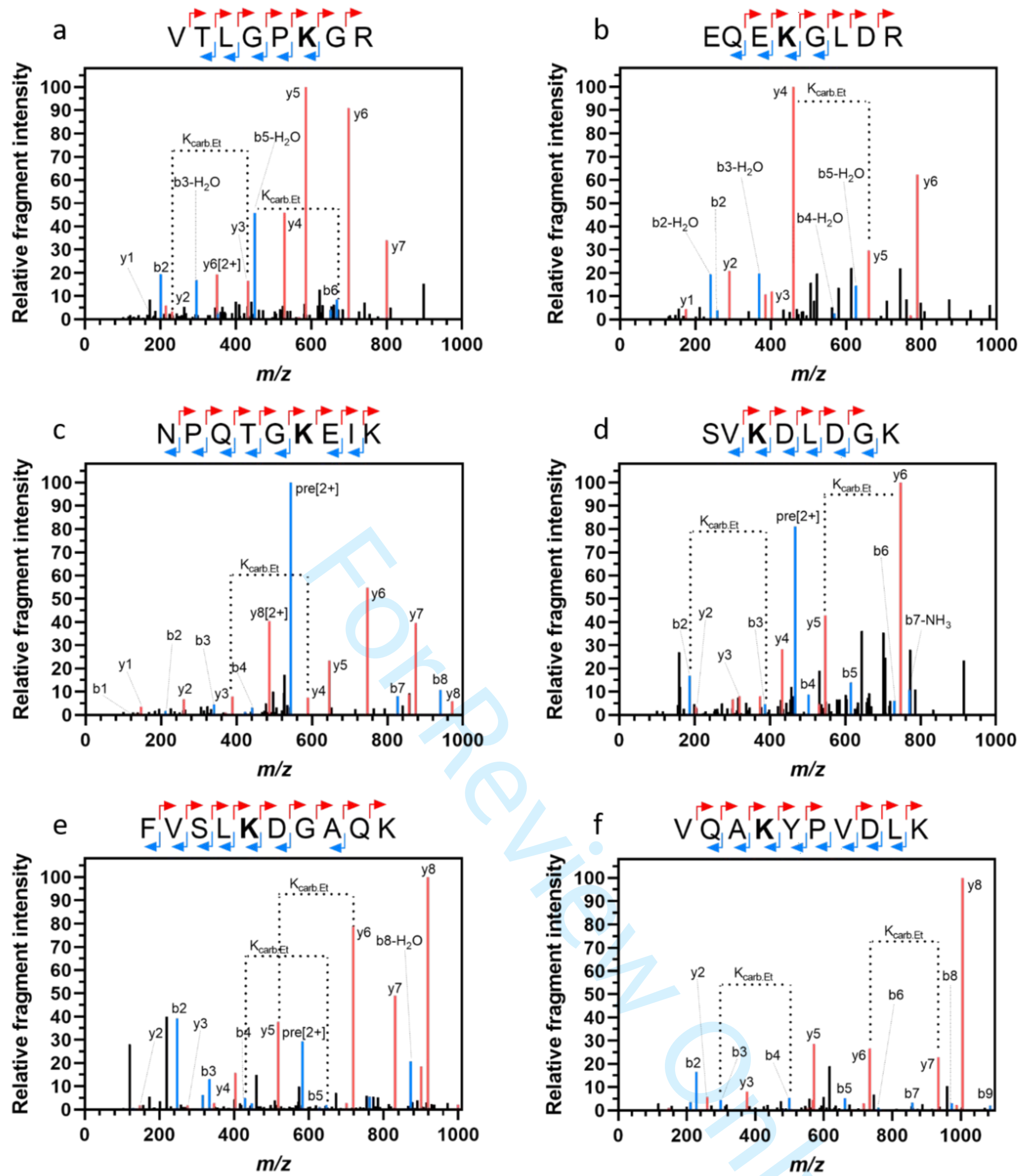


Figure 2. False positive spectra for CO_2-binding carbamate sites identified by mass spectrometry analysis software. The charts are plots of relative fragment intensity versus mass/charge ratio (m/z) for fragmentation data from ESI-MSMS identifying ethyl-trapped carbamates in the presence of $^{12}CO_2$. Peptide sequences indicate the identification of predominant +1y (red) +1b (blue) ions by ESI-MSMS shown in the plot. The potential modified residue is indicated in bold. **A.** Identification of a potential carbamate at a tryptic digest site. **B.** Identification of a carbamate without any surrounding +1y or +1b ions and the suggestion of an unlikely ethyl modification on the peptide N-terminal alanine .



For Review Only

Figure 3. The identification of CO_2-binding proteins~~.~~ The charts are plots of relative fragment intensity versus mass/charge ratio (m/z) for fragmentation data from ESI-MSMS identifying ethyl-trapped carbamates in the presence of $^{12}CO_2$. ~~Plots of relative fragment intensity versus mass/charge ratio (m/z) for fragmentation data from ESI-MSMS identifying ethyl-trapped carbamate on Ub K33 (**B**, **D**) and K48 (**C**, **E**) in the presence of $^{12}CO_2$ (**B**, **C**) or $^{13}CO_2$ (**D**, **E**).~~ Peptide sequences indicate the identification of predominant +1y (red) +1b (blue) ions by ESI-MSMS shown in the plot. The modified residue is indicated in bold. $K_{carb.Et}$ indicates the molecular weight difference between ions diagnostic of the modified Lys. **A.** Lysine 34 of *groL*. **B.** Lysine 121 of *tnaA*. **C.** Lysine 67 of *hupA*. **D.** Lysine 127 of *glnH*. **E.** Lysine 45 of *rbsB*. **F.** Lysine 285 of *rbsB*.

For Review Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1

Gene name	**Protein**	**Residue**
groL	60 kDa chaperonin	K34
tnaA	Tryptophanase	K121
hupA	DNA-binding protein HU-alpha	K67
glnH	Glutamine-binding periplasmic protein	K127
rbsB	Ribose import binding protein RbsB	K45 K285

Table 1. Summary of *E.coli* proteins carrying CO_2-binding carbamate sites and the lysine where this binding occurs.

For Review Only