



# Chemoproteomic identification of CO<sub>2</sub>-dependent lysine carboxylation in proteins

Dustin T. King<sup>1</sup>, Sha Zhu<sup>1</sup>, Darryl B. Hardie<sup>3</sup>, Jesús E. Serrano-Negrón<sup>1</sup>, Zarina Madden<sup>1</sup>, Subramania Kolappan<sup>1</sup> and David J. Vocadlo<sup>1,2</sup>✉

**Carbon dioxide is an omnipresent gas that drives adaptive responses within organisms from all domains of life. The molecular mechanisms by which proteins serve as sensors of CO<sub>2</sub> are, accordingly, of great interest. Because CO<sub>2</sub> is electrophilic, one way it can modulate protein biochemistry is by carboxylation of the amine group of lysine residues. However, the resulting CO<sub>2</sub>-carboxylated lysines spontaneously decompose, giving off CO<sub>2</sub>, which makes studying this modification difficult. Here we describe a method to stably mimic CO<sub>2</sub>-carboxylated lysine residues in proteins. We leverage this method to develop a quantitative approach to identify CO<sub>2</sub>-carboxylated lysines of proteins and explore the lysine 'carboxylome' of the CO<sub>2</sub>-responsive cyanobacterium *Synechocystis* sp. We uncover one CO<sub>2</sub>-carboxylated lysine within the effector binding pocket of the metabolic signaling protein PII. CO<sub>2</sub>-carboxylation of this lysine markedly lowers the affinity of PII for its regulatory effector ligand ATP, illuminating a negative molecular control mechanism mediated by CO<sub>2</sub>.**

Carbon dioxide (CO<sub>2</sub>) has been present in the atmosphere and dissolved in water since life first evolved. Over this time, CO<sub>2</sub> has emerged for some organisms as a vital anabolic metabolite, whereas for others it has become a terminal catabolic waste product. In the CO<sub>2</sub> rich atmosphere of 3 billion years ago, cyanobacteria were likely the first to use CO<sub>2</sub> as their primary carbon source. Since that time, diverse organisms have evolved the ability to sense and adapt to variations in the levels of available CO<sub>2</sub>. Among many known examples<sup>1,2</sup>, cyanobacteria sense low CO<sub>2</sub> levels and respond by activating transcriptional programs that facilitate fixing of carbon<sup>3</sup>. The importance of controlling CO<sub>2</sub> levels within cellular compartments is illustrated by the evolution of elaborate transport systems and specialized organelles, which are used to spatiotemporally regulate subcellular pools of CO<sub>2</sub> (ref. <sup>4</sup>). Despite the important roles played by CO<sub>2</sub> in cellular physiology<sup>1,2</sup>, the fundamental molecular mechanisms underlying how it is sensed remain largely unknown.

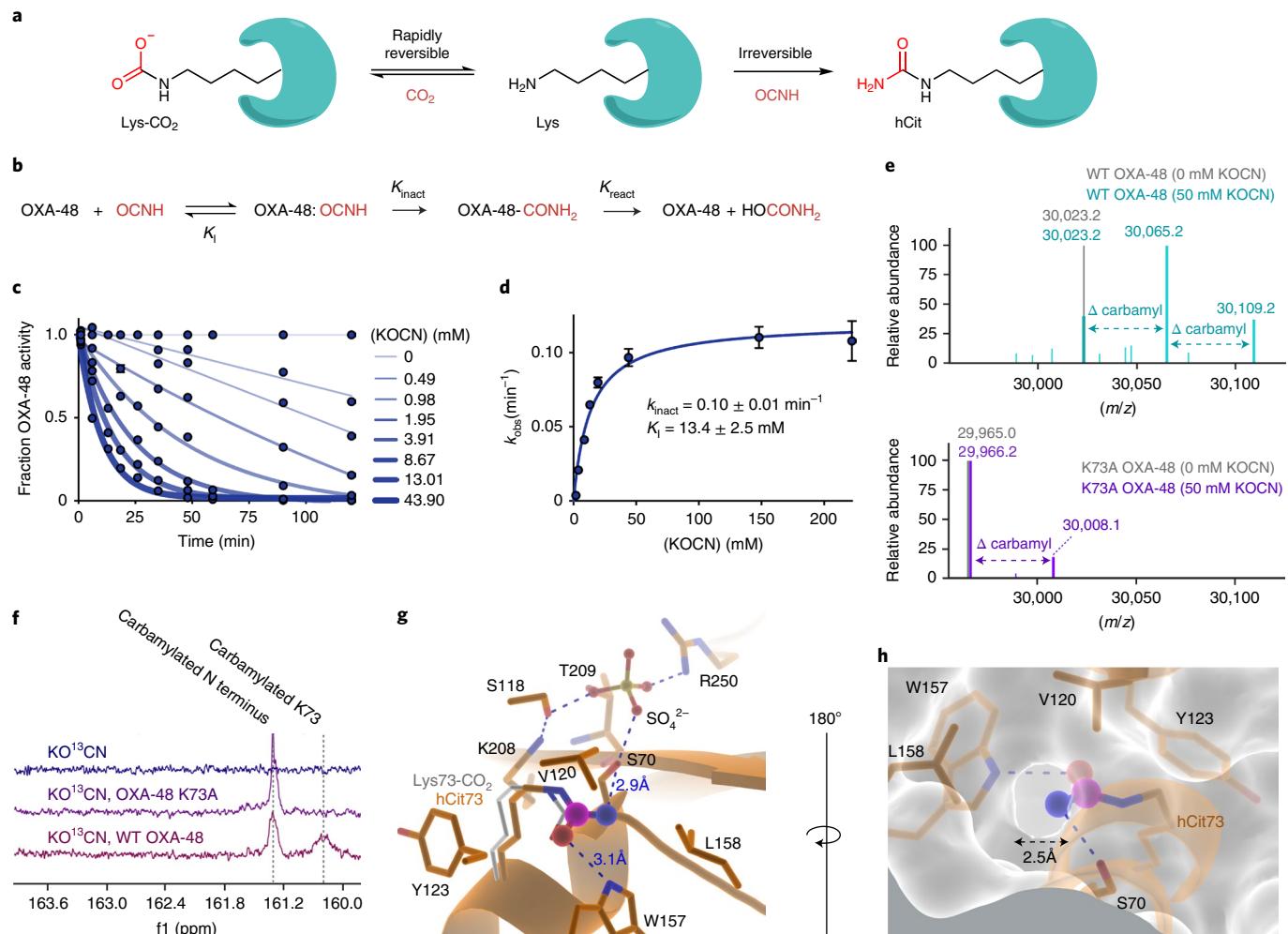
Understanding how organisms sense CO<sub>2</sub> at the biochemical level is complicated by CO<sub>2</sub> gradually equilibrating with water to form bicarbonate (HCO<sub>3</sub><sup>-</sup>). The importance of maintaining this equilibrium within different organelles is manifest by organisms typically expressing multiple organelle-specific carbonic anhydrases that accelerate interconversion of these two inorganic carbon (C<sub>i</sub>) species<sup>5</sup>. Accordingly, interrogation of the biochemical mechanisms by which proteins sense CO<sub>2</sub> is challenging because CO<sub>2</sub> or HCO<sub>3</sub><sup>-</sup> may both be operative molecular species, yet neither forms stable adducts with proteins. The discovery that soluble adenylyl cyclase activity is regulated by HCO<sub>3</sub><sup>-</sup> (ref. <sup>6</sup>) spurred recognition of the role of this enzyme in CO<sub>2</sub> sensing. There have, however, been no compelling reports on the biochemical mechanisms of sensing of HCO<sub>3</sub><sup>-</sup> by proteins. With respect to CO<sub>2</sub>, one often-overlooked mechanism by which CO<sub>2</sub> can directly modulate protein biochemistry is through its ability to covalently carboxylate the ε-NH<sub>2</sub> of lysine residues of proteins to form a carbamic acid (Lys-CO<sub>2</sub>, Fig. 1a), which studies show occurs through CO<sub>2</sub> rather than HCO<sub>3</sub><sup>-</sup> acting as the

electrophile<sup>7</sup>. This posttranslational modification leads to a marked change in molecular structure as well as a complete switch in the physiological charge state of lysine from positive (ε-amino pK<sub>a</sub> roughly 10.5) to negative (carbamic acid, pK<sub>a</sub> roughly 3.9). Notably, Lys-CO<sub>2</sub> on its own decomposes rapidly to give back CO<sub>2</sub> and lysine in a matter of a few seconds (*k*<sub>breakdown</sub> equal to roughly 5 × 10<sup>6</sup> M<sup>-1</sup> s<sup>-1</sup> at pH 7.5)<sup>8</sup>, and Lys-CO<sub>2</sub> sites within proteins are reversible<sup>9–12</sup>. These features make Lys-CO<sub>2</sub> well suited to serve as a molecular sensor that acts by modulating protein function in response to fluctuations in environmental CO<sub>2</sub> levels.

However, Lys-CO<sub>2</sub> has only been well characterized on a handful of proteins<sup>9–13</sup>. Although it is most common as a ligand for coordinating metal ions within proteins<sup>14</sup>, such as in ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO)<sup>15</sup>, Lys-CO<sub>2</sub> residues have also been found in metal-free sites as in class D (OXA) β-lactamases<sup>9</sup>. Although other cases of Lys-CO<sub>2</sub> are known, there are few unambiguous examples where such proteins are shown to contain a Lys-CO<sub>2</sub> site that controls activity in physiological ranges of CO<sub>2</sub>. This limited exploration may stem in part from the prevailing view of lysine residues as being protonated at physiological pH, which would render them unreactive to CO<sub>2</sub>. Global profiling of lysine reactivity has, however, revealed that the human proteome contains several hundred reactive lysine residues<sup>16</sup>. Thus, Lys-CO<sub>2</sub> may play a more prominent role than is currently appreciated in directly sensing CO<sub>2</sub> and reversibly regulating protein biochemistry. Identifying Lys-CO<sub>2</sub> regulated sensors, however, is complicated by the labile nature of this modification, which makes it hard to identify.

Most Lys-CO<sub>2</sub> sites have been identified by serendipitous observation in crystallographic structures<sup>14</sup>. Direct detection of Lys-CO<sub>2</sub> by mass spectrometry (MS) is limited to a few model peptides<sup>17</sup>, probably because Lys-CO<sub>2</sub> rapidly decomposes during analysis<sup>8</sup>. Accordingly, stabilization of Lys-CO<sub>2</sub> by chemical derivatization has been explored. Initial mapping of Lys-CO<sub>2</sub> within RuBisCO was achieved by alkylating the carbamate using diazomethane

<sup>1</sup>Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, Canada. <sup>2</sup>Department of Chemistry, Simon Fraser University, Burnaby, British Columbia, Canada. <sup>3</sup>University of Victoria-Genome BC Proteomics Centre, University of Victoria, Victoria, British Columbia, Canada. ✉e-mail: dvocadlo@sfu.ca



**Fig. 1 | OCNH is a time-dependent inactivator of OXA-48 that selectively forms a stable covalent hCit adduct with the active site Lys73.** **a**, Schematic of CO<sub>2</sub>-dependent lysine carboxylation and OCNH-dependent lysine carbamylation. **b**, Minimal kinetic scheme for inactivation of OXA-48 by OCNH. **c**, Time-dependent inactivation of OXA-48. OXA-48 was preincubated with KOCN before dilution and activity assay using nitrocefin as a substrate. Curves were fit to a standard first order rate equation. **d**, Kinetic analysis of the inactivation of OXA-48 by OCNH. The solid line shows the best fit of the data to a standard saturation kinetic expression. Data in **c,d** are presented as mean values  $\pm$  standard error (s.e.m.) from  $n = 2$  independent experiments. **e**, LC-MS intact mass spectra of WT and K73A OXA-48 in the presence and absence of KOCN. **f**, OXA-48 <sup>13</sup>C NMR in the presence or absence of 50 mM KO<sup>13</sup>CN. **g**, Overlay of the hCit73 OXA-48 structure (PDB ID 7LXG) and the Lys73-CO<sub>2</sub> form (PDB ID 4S2P, ref. 5). The hCit73 OXA-48 backbone atoms are displayed in orange cartoon representation. The hCit73 and Lys73-CO<sub>2</sub> residues are orange and white sticks with heteroatoms colored by type. Hydrogen bonds and electrostatic interactions are illustrated as blue dashes. **h**, Closeup of the OXA-48 solvent accessible surface surrounding hCit73. The surface was calculated with the hCit73 OCNH<sub>2</sub> atoms omitted to enable visualization of the entry channel.

followed by characterization of the stable esterified lysine adduct<sup>10</sup>. This ‘trapping’ strategy was improved on and elegantly applied on a proteome-wide scale<sup>18</sup>. This approach relies on the reactive electrophile triethyloxonium tetrafluoroborate, which alkylates various protein nucleophiles including aspartate and glutamate residues<sup>18</sup>. This broad reactivity complicates MS analyses and could affect protein structure, which may in turn affect the reactivity of lysine residues. Accordingly, new chemical biology tools are needed to help facilitate study and discovery of Lys-CO<sub>2</sub> to illuminate its role in CO<sub>2</sub> sensing in organisms.

Here we describe a quantitative strategy using a simple lysine-selective chemical probe to detect Lys-CO<sub>2</sub> sites through covalent competition with CO<sub>2</sub>. We show that the CO<sub>2</sub> analog isocyanic acid (OCNH) reacts with surprising selectivity, forming an irreversible covalent homocitrulline (hCit) that is structurally analogous to Lys-CO<sub>2</sub> (Fig. 1a). CO<sub>2</sub> competes with OCNH-driven homocitrullination in a concentration dependent manner to allow

detection of known Lys-CO<sub>2</sub> sites within model proteins RuBisCO and OXA-48. We explore Lys-CO<sub>2</sub> within the proteome of the CO<sub>2</sub>-responsive cyanobacterium *Synechocystis* sp<sup>3</sup>. From among the resulting proteomic observations, we validate modification of the central C/N metabolic regulator PII<sup>19</sup> and assess the role of Lys-CO<sub>2</sub> in regulating its biochemical function. We find that Lys-CO<sub>2</sub> formation occurs site specifically within a key ligand binding pocket on PII and that its levels vary over a physiological range of CO<sub>2</sub> concentrations. Furthermore, Lys-CO<sub>2</sub> modification of PII antagonizes its binding to ATP, unveiling a distinct biochemical control mechanism involving negative regulation by CO<sub>2</sub>.

## Results

**OCNH reacts to yield a stable homocitrulline isostere.** The reversible nature of Lys-CO<sub>2</sub> makes it difficult to detect CO<sub>2</sub> reactive lysine residues. Recognizing the complications associated with alkylation-based trapping strategies<sup>10,18</sup>, we decided to pursue

$\text{CO}_2$ -dependent protection of reactive lysines against modification by a lysine-selective chemical probe. We reasoned that small electrophilic analogs of  $\text{CO}_2$ , such as carbon disulfide ( $\text{CS}_2$ ), isothiocyanic acid ( $\text{SCNH}$ ) or isocyanic acid ( $\text{OCNH}$ ), might be useful. Covalent adducts of these molecules with lysine residues should be stable<sup>8</sup> and their similarity to  $\text{CO}_2$  could lend specificity for  $\text{CO}_2$ -reactive lysines. We reasoned that comparing the levels of these stable lysine adducts in the presence and absence of  $\text{CO}_2$  could, accordingly, allow identification of sites of Lys- $\text{CO}_2$ .

As a first step, we examined the effect of these electrophiles on the catalytic activity of the  $\text{CO}_2$ -dependent OXA-48  $\beta$ -lactamase. The activity of this  $\beta$ -lactamase depends on carboxylation of Lys73 (Lys73- $\text{CO}_2$ ) to generate a carbamate that serves as a catalytic general base<sup>9</sup> (Extended Data Fig. 1). Modification of Lys73 by any of these analogs should therefore lead to inhibition of this enzyme. Incubation of OXA-48 with  $\text{CS}_2$ ,  $\text{SCNH}$  and  $\text{OCNH}$  showed us that  $\text{OCNH}$  was the most effective inhibitor (half-maximum inhibitory concentration ( $\text{IC}_{50}$ ) =  $0.82 \pm 0.15$  mM, Extended Data Fig. 2a). This observation is consistent with  $\text{OCNH}$  having a nearly linear structure, thereby making it an excellent mimic of  $\text{CO}_2$  (ref. <sup>20</sup>). Moreover, the nucleophilic primary amine of lysine reacts with  $\text{OCNH}$  to form a stable homocitrulline (hCit) adduct (Fig. 1a and Supplementary Fig. 1), whereas adventitious reactions with other nucleophilic amino acids leads to adducts that rapidly revert<sup>21,22</sup>. The only stable adducts formed by  $\text{OCNH}$  on amino acid and protein sidechains at neutral pH is on lysine<sup>21–23</sup>, consistent with hCit being the only main protein adduct we observed (Supplementary Fig. 2). Notably, hCit is known to be amenable to analysis by MS<sup>24</sup>. Given these observations it was surprising to us that, although  $\text{CS}_2$  and  $\text{OCS}$  have been used to stably modify Lys- $\text{CO}_2$  sites in purified proteins<sup>25,26</sup>, little attention has been given to  $\text{OCNH}$ .

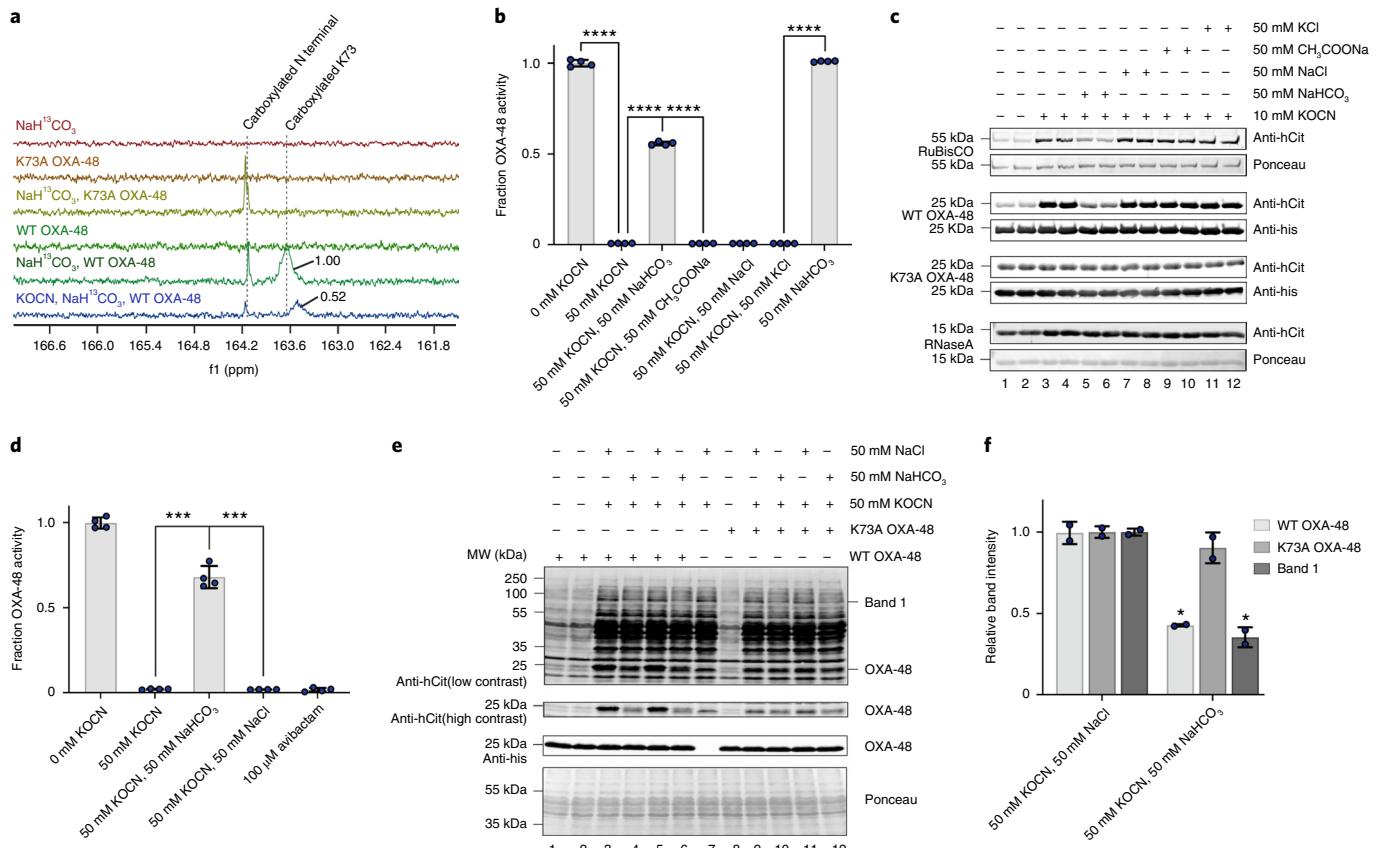
Motivated by the favorable properties of  $\text{OCNH}$ , we next explored its ability to block Lys- $\text{CO}_2$  formation at a known site using as a model OXA-48. We found that  $\text{OCNH}$  rapidly inhibits OXA-48 activity ( $k_{\text{inact}} = 0.10 \pm 0.01 \text{ min}^{-1}$ ,  $K_i = 13.4 \pm 2.5$  mM, Fig. 1c,d). Moreover, although still clearly folded (Extended Data Fig. 2b,c), activity could not be recovered by dialysis, indicating that the adduct is stable over a 48-hour time frame (Extended Data Fig. 2d,e). We next performed LC-MS analysis to determine the number of  $\text{OCNH}$  (+43 Da) molecules incorporated into both the wild-type (WT) OXA-48 and the K73A OXA-48 variant. The WT enzyme (30,023 Da) showed incorporation of two  $\text{OCNH}$  molecules (30,065 ( $\Delta 42$  Da) and 30,109 Da ( $\Delta 86$  Da)) whereas the variant (29,965 Da) incorporated only one (30,008 Da ( $\Delta 43$  Da)), suggesting Lys73 forms a hCit adduct within the WT enzyme (Fig. 1e). We confirmed this using  $^{13}\text{C}$  nuclear magnetic resonance (NMR) experiments to monitor incorporation of  $\text{O}^{13}\text{CNH}$  on OXA-48. We identified two distinct  $^{13}\text{C}$  resonances for  $\text{O}^{13}\text{CNH}$  modification of the WT enzyme (160.5 and 161.3 ppm), whereas only a single resonance (161.3 ppm) was observed for the K73A OXA-48 variant (Fig. 1f). The resonance at 161.3 ppm most likely corresponds to modification of the more nucleophilic N terminus<sup>22</sup>. Modification of the N terminus was confirmed by Edman sequencing, whereby we observed that  $\text{OCNH}$ -treated OXA-48 had a partially blocked N terminus, consistent with carbamylation of its N terminus (Supplementary Fig. 3). As an extra measure, we used a polyclonal pan-specific anti-hCit antibody and found a twofold greater immunoreactivity, as assessed by band intensity, for the WT enzyme reacted with  $\text{OCNH}$  as compared to K73A OXA-48, despite this enzyme having 18 lysine residues (Extended Data Fig. 2f,g). These collective data show  $\text{OCNH}$  stably and selectively modifies the  $\text{CO}_2$ -reactive lysine of OXA-48 to generate hCit. Notably, we found that established lysine-reactive activated esters, and N-alkyl isocyanates, did not compete with  $\text{CO}_2$  for binding to known Lys- $\text{CO}_2$  modified proteins (Supplementary Figs. 4 and 5).

We next set out to confirm formation of hCit73 and examine how closely it resembles the isosteric Lys73- $\text{CO}_2$  that is normally found within the active site of this enzyme. We therefore solved the 2.2 Å resolution crystal structure of OCNH-inactivated OXA-48, which is the first homocitrullinated protein structure in the Protein Data Bank (PDB) (Supplementary Table 1). Most notably, the Fo-Fc omit electron density map confirmed formation of hCit73, with unambiguous density connecting the  $\text{OCNH}_2$  moiety to K73 (Extended Data Fig. 2h). An overlay of our hCit73 OXA-48 coordinates with the known WT OXA-48 structure (PDB ID 4S2P, root mean square deviation (r.m.s.d.) for common alpha-carbon atoms (CA); 0.4 Å) shows hCit73 serves as an almost exact isosteric mimic of Lys73- $\text{CO}_2$  (Fig. 1g). Analysis reveals the  $\text{OCNH}_2$  moiety occupies a confined pocket with the heteroatoms bonding to nearby residues rather than extending into bulk solvent. The hCit73 sidechain O- $\phi$  forms a hydrogen bond with the W157 NH- $\epsilon$  (3.1 Å), and the hCit73 N- $\phi$  hydrogen bonds to the catalytic serine O- $\gamma$  hydroxyl (2.9 Å, Fig. 1g). Notably, hCit73 is separated from bulk solvent by a narrow channel with an aperture of only around 2.5 Å (Fig. 1h). These observations highlight that  $\text{OCNH}$  acts faithfully as a  $\text{CO}_2$  bioisostere that can access constrained protein environments that may be typical of Lys- $\text{CO}_2$  sites.

**$\text{CO}_2$  competes with  $\text{OCNH}$  to block homocitrullination.** We next set out to determine whether  $\text{CO}_2$  could directly compete with irreversible  $\text{OCNH}$  covalent modification using as models both OXA-48 and RuBisCO, both of which have well characterized Lys- $\text{CO}_2$  sites<sup>9,10</sup>. To ensure that  $\text{CO}_2$ -dependent differences in  $\text{OCNH}$  labeling are independent of variation in pH, we performed all  $\text{OCNH}/\text{CO}_2$  competition assays in a buffer that resisted pH shifts on addition of  $\text{HCO}_3^-$  (Supplementary Fig. 6). We first showed using  $^{13}\text{C}$  NMR experiments that incubation of WT OXA-48 with  $\text{H}^{13}\text{CO}_3^-$  resulted in a resonance at 163.6 ppm (Fig. 2a), a chemical shift that is in line with previous  $^{13}\text{C}$  NMR studies of Lys- $\text{CO}_2$  (refs. <sup>14,27</sup>). Notably, this peak was absent in the spectrum obtained for the K73A OXA-48 variant incubated with  $\text{H}^{13}\text{CO}_3^-$ , confirming that the resonance observed at 163.6 ppm for the WT enzyme represents the Lys73- $^{13}\text{CO}_2$  adduct. Preincubation with  $\text{OCNH}$  reduced the relative intensity of this Lys73- $^{13}\text{CO}_2$  peak, supporting site-specific competition between  $\text{OCNH}$  and  $\text{CO}_2$  (Fig. 2a). In addition, incubation with 50 mM  $\text{HCO}_3^-$  substantially protected OXA-48 from inactivation by  $\text{OCNH}$  (Fig. 2b). This  $\text{HCO}_3^-$  concentration was chosen as it yields 2 mM  $\text{CO}_2$  in our assays, a  $\text{CO}_2$  concentration that is at the upper end of the physiological range observed in cyanobacteria (0.1–2.0 mM)<sup>28</sup> and also yields robust protection in our *in vitro* assays using OXA-48 (Extended Data Fig. 3).

Furthermore, immunoblotting using the antihomocitrulline antibody revealed that  $\text{CO}_2$  substantially reduced the band intensity of both  $\text{OCNH}$ -treated RuBisCO and WT OXA-48, yet had no effect on  $\text{OCNH}$ -treated K73A OXA-48 or the negative control protein RNaseA (Fig. 2c). The extent of  $\text{OCNH}$ -dependent hCit formation on OXA-48 varied as a function of  $\text{CO}_2$  concentration, indicating that the method could report on the relative abundance of Lys- $\text{CO}_2$  (Extended Data Fig. 3). These data show that  $\text{OCNH}/\text{CO}_2$  competition assays can be used to identify whether purified proteins are modified *in vitro* by  $\text{CO}_2$  to generate Lys- $\text{CO}_2$ .

We next sought to apply this method to cell lysates from *Escherichia coli*. Before starting experiments, we addressed various technical issues. First, we used a gene knockout strain ( $\Delta \text{CynS}$  *E. coli*) that lacks a detoxifying ‘cyanase’ enzyme that catalyzes breakdown of  $\text{OCNH}$ <sup>29</sup>. Second, we used turbidity assays to show that there was no substantial aggregation of the proteome during the competition assays (Supplementary Fig. 7a). Third, we confirmed a short  $T_{1/2}$  for reaching the bicarbonate equilibrium by showing that the addition of carbonic anhydrase into our buffers had a negligible effect on OXA-48 activity (Supplementary Fig. 7b). With these

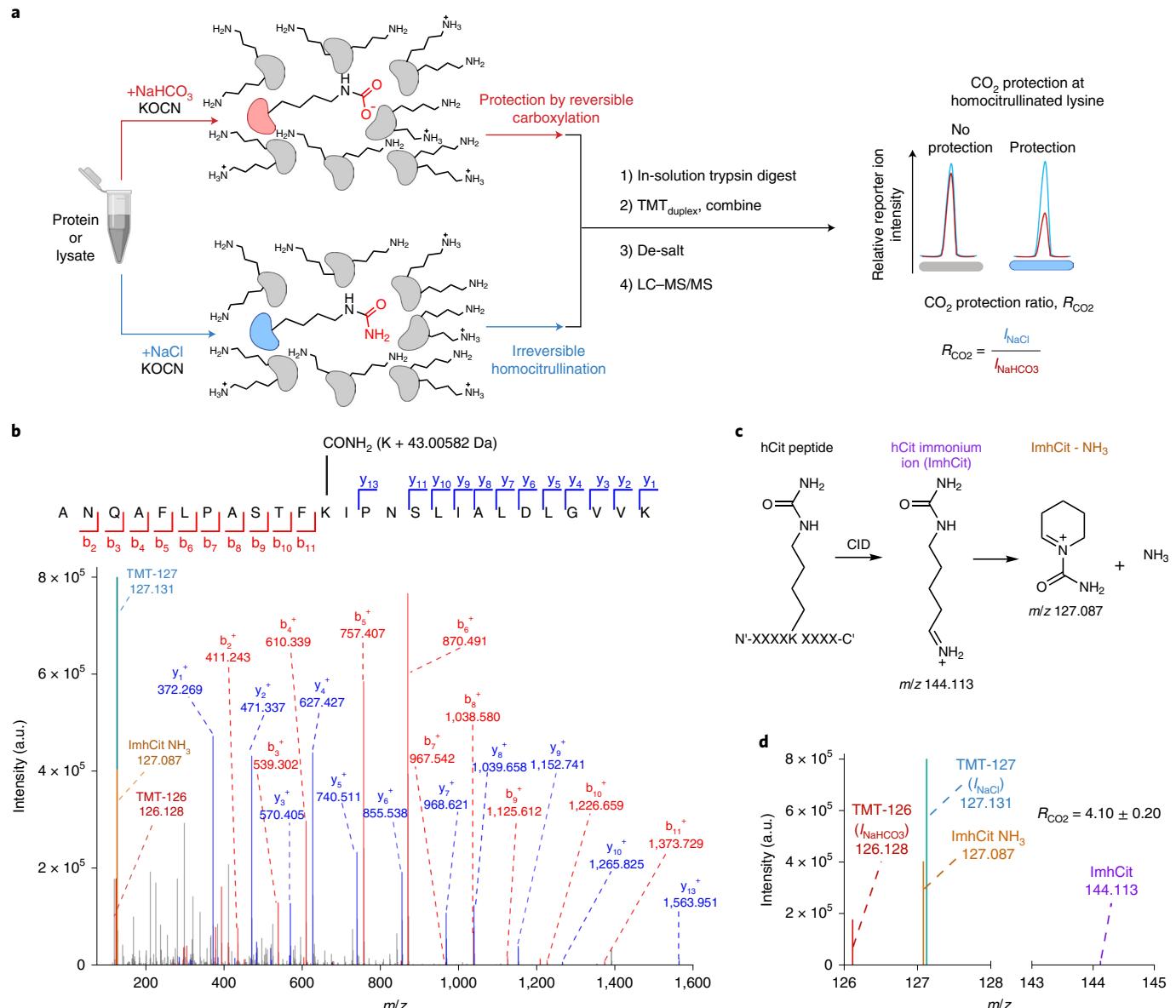


**Fig. 2 | CO<sub>2</sub> competes with OCNH at Lys-CO<sub>2</sub> sites on model proteins in vitro and in cellular lysates.** **a**, OXA-48 <sup>13</sup>C NMR experiments using NaH<sup>13</sup>CO<sub>3</sub>. OXA-48 was preincubated in the presence or absence of 50 mM of the indicated salts before <sup>13</sup>C NMR. Integrated peak intensities for Lys73-<sup>13</sup>CO<sub>2</sub> are shown. **b**, OXA-48 competition activity assays. The indicated salts were added at the same time to OXA-48 and the solution was incubated for 1 h followed by dilution and enzyme kinetic assay. In all bar charts the indicated *P* values correspond to a two tailed Student's *t*-test assuming unequal variance; \**P*<0.05, \*\**P*<0.01, \*\*\**P*<0.001, \*\*\*\**P*<0.0001. The indicated significant *P* values in **b** are as follows: (0 mM KOCN) versus (50 mM KOCN)=1.65×10<sup>-6</sup>, (50 mM KOCN) versus (50 mM KOCN, 50 mM NaHCO<sub>3</sub>)=1.14×10<sup>-6</sup>, (50 mM KOCN, 50 mM NaHCO<sub>3</sub>) versus (50 mM KOCN, 50 mM CH<sub>3</sub>COONa)=1.10×10<sup>-6</sup> and (50 mM KOCN, 50 mM KCl) versus (50 mM NaHCO<sub>3</sub>)=3.70×10<sup>-9</sup>. **c**, Immunoblot competition assays performed on purified proteins with detection using a polyclonal anti-hCit antibody. The image is a representative of three independent experiments. **d**, OXA-48 activity assays following CO<sub>2</sub> competition performed in  $\Delta$ CynS *E. coli* cell lysate. The indicated significant *P* values in **d** are as follows: (50 mM KOCN) versus (50 mM KOCN, 50 mM NaHCO<sub>3</sub>)=2.65×10<sup>-4</sup> and (50 mM KOCN, 50 mM NaHCO<sub>3</sub>) versus (50 mM KOCN, 50 mM NaCl)=2.67×10<sup>-4</sup>. **e**, Representative immunoblot following CO<sub>2</sub> competition assay performed on cell lysates from  $\Delta$ CynS *E. coli*. **f**, Densitometry analysis of select bands from **e**. Immunoblot and analysis in **e,f** are a representative from three independent experiments. The indicated significant *P* values in **f** comparing the NaCl and HCO<sub>3</sub><sup>-</sup> conditions are as follows: WT OXA-48=0.050, Band 1=0.025. Data in bar charts are presented as mean values±s.d. from *n*=4 (**b,d**) or *n*=2 (**f**) independent experiments.

considerations addressed, we validated our OCNH/CO<sub>2</sub> competition assay in cell lysates by spiking in purified WT OXA-48 as a positive control. This enabled us to confirm competition is occurring using either an OXA-48 activity assay or hCit immunoblots. Our kinetic assays using  $\Delta$ CynS *E. coli* lysates revealed OCNH-dependent inhibition of OXA-48 could be overcome by adding CO<sub>2</sub> (Fig. 2d). Next, we found that hCit immunoblot analysis of  $\Delta$ CynS *E. coli* lysates showed a CO<sub>2</sub>-dependent decrease in hCit immunoreactivity for WT OXA-48 and unknown proteins (Fig. 2e,f). Taken together, these results show that OCNH/CO<sub>2</sub> competition assays can be used to study CO<sub>2</sub>-dependent carboxylation of proteins, setting the stage for the development of a proteome-wide approach to map Lys-CO<sub>2</sub>.

**Competition-MS enables proteomic mapping of Lys-CO<sub>2</sub>.** We next combined OCNH/CO<sub>2</sub> competition assays with MS analysis to enable quantitative proteome-wide mapping of Lys-CO<sub>2</sub> through a method we term lysine carboxylation competition-MS (LysCarComp-MS, Fig. 3a). First, we developed a MS/MS protocol

using a synthetic hCit-containing tryptic peptide from OXA-48 (ANQAFLPASTFhCit(K<sub>73</sub>+43.01 Da)IPNSLIALDLGVVK, Supplementary Fig. 8). High-energy collisional dissociation (HCD)-based fragmentation enabled mapping of the hCit73 based on the characteristic +43.01 Da increase in the precursor ion mass. The b11 fragment ion lacked this diagnostic mass whereas the b13 ion showed this mass increase, indicating the modification is found between these fragmentation sites (Supplementary Fig. 8c,d). These data enabled high confidence assignment of the hCit modification to amino acid K73 (Xcorr=6.0). Notably, we observed that HCD reproducibly generated two distinctive fragment ions in the low *m/z* range that exactly coincide with the theoretical mass of hCit-specific immonium ions (ImhCit *m/z*=144.113 and ImhCit NH<sub>3</sub> *m/z*=127.087, Fig. 3b-d). We propose these ions should be generally useful as diagnostic markers for mapping hCit residues as seen for similar HCD-induced immonium ions that occur for acetyl-lysine, which are used to identify acetyl-lysine containing peptides in proteomics datasets<sup>30</sup>. Finally, because hCit residues are

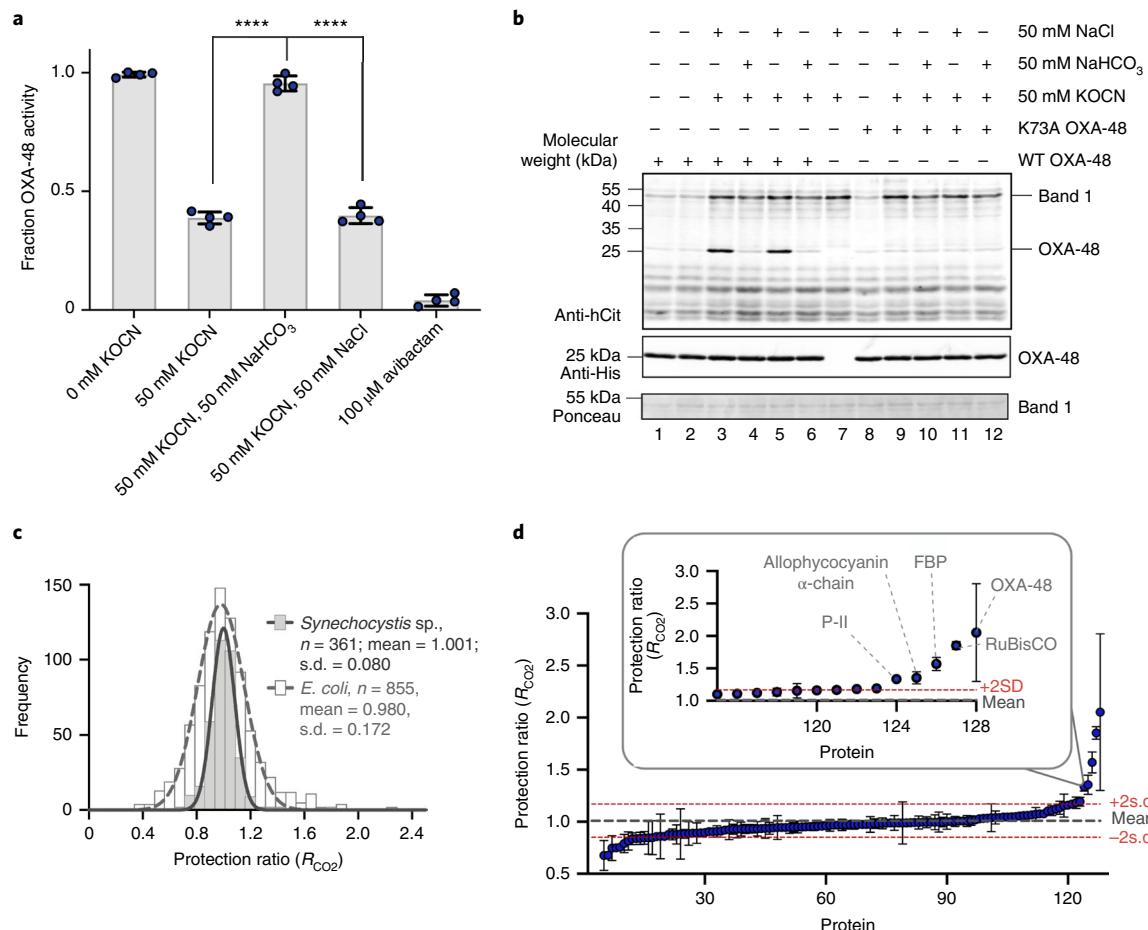


**Fig. 3 | LysCarComp-MS for the quantitative determination of site selective CO<sub>2</sub> protection ratio.** **a**, Schematic diagram outlining the LysCarComp-MS strategy. **b**, MS/MS spectra for TMT<sub>duplex</sub>-labeled hCit-containing tryptic peptide derived from homocitrullinated OXA-48 protein (TMT-ANQAFPLASTFhCit(K<sub>73</sub>+43.01Da)IPNSLIALDLGVVK-TMT, MH+(Da)=3,220.86). **c**, Proposed immonium ion structures derived from hCit. **d**, Reporter ion intensities and corresponding  $R_{CO_2}$  for the hCit73 OXA-48 tryptic peptide. The mean  $R_{CO_2}$  value  $\pm$  s.d. from three independent experimental replicates is shown.

not cleaved by trypsin<sup>31</sup>, the presence of missed tryptic cleavage sites should aid in identifying hCit-containing peptides. These observations indicate that hCit can be readily mapped in tryptic peptides using a series of diagnostic features seen during MS/MS-based peptide sequencing.

We next used OXA-48 as a model protein to assess whether this method can provide relative quantitation of hCit levels in OCNH/CO<sub>2</sub> competition assays. We first incubated WT OXA-48 protein with OCNH in either the presence or absence of physiological concentrations of CO<sub>2</sub> (refs. <sup>28,32</sup>). We predicted that added CO<sub>2</sub> should reduce hCit formation at Lys-CO<sub>2</sub> sites that have physiologically relevant affinities for CO<sub>2</sub>. OXA-48 was then subjected to tryptic digestion followed by isobaric labeling using the tandem mass tag duplex (TMT<sub>duplex</sub>) mass tagging strategy<sup>33</sup>. Labeled peptides were analyzed by LC-MS/MS, and hCit sites were identified using the

strategy outlined above. The hCit73-containing tryptic peptide (ANQAFPLASTFhCit(K<sub>73</sub>+43.01 Da)IPNSLIALDLGVVK) from OXA-48 was confidently identified and mapped, generating a MS/MS spectrum nearly identical to that observed for the synthetic hCit73 peptide (Fig. 3b and Supplementary Fig. 8c). To quantify the extent of protection conferred by CO<sub>2</sub>, we defined a CO<sub>2</sub> protection index ( $R_{CO_2}$ ) that is the ratio of the TMT reporter ion intensities ( $I_{NaCl}/I_{NaHCO_3}$ ), with 1 indicating no protection and >1 indicating CO<sub>2</sub>-dependent protection. Global analysis of all nonhomocitrullinated OXA-48 tryptic peptides revealed an average for  $R_{CO_2}$  of  $1.00 \pm 0.01$ , confirming that overall protein abundance did not vary between conditions. The hCit73-containing peptide, however, displayed the highest  $R_{CO_2}$  of all peptides ( $4.10 \pm 0.20$ , Fig. 3d), with a relative reporter ion intensity (RRI) compared to protein average of  $0.32 \pm 0.09$  (Supplementary Table 2). Two more hCit73-containing



**Fig. 4 | LysCarComp-MS enables proteome-wide identification of lysine carboxylation sites in *Synechocystis* sp.** **a**, OXA-48 activity assays following CO<sub>2</sub> competition performed in *Synechocystis* sp. cell lysates. Chart displays mean values  $\pm$  s.d. from  $n=4$  independent experiments with  $P$  values: two tailed Student's  $t$ -test assuming unequal variance;  $****P < 0.0001$ . The indicated significant  $P$  values in **a** are as follows: (50 mM KOCN) versus (50 mM KOCN, 50 mM NaHCO<sub>3</sub>) =  $2.59 \times 10^{-7}$  and (50 mM KOCN, 50 mM NaHCO<sub>3</sub>) versus (50 mM KOCN, 50 mM NaCl) =  $3.52 \times 10^{-7}$ . **b**, Representative immunoblot following CO<sub>2</sub> competition assay performed on cell lysates from *Synechocystis* sp. The image is a representative of three independent experiments. **c**, Histogram of LysCarComp-MS  $R_{CO_2}$  values for hCit peptides identified in *Synechocystis* sp. and  $\Delta CynS$  *E. coli* lysates. **d**, Distribution of  $R_{CO_2}$  values arranged in increasing order for hCit peptides identified in *Synechocystis* sp. Data are presented as mean values  $\pm$  s.e.m. from  $n=2$  independent experiments. Some data points lack error bars because they were only observed in a single replicate (Supplementary Data 1).

peptides were identified that had two or three missed cleavages ( $RRI = 0.72 \pm 0.1$ ,  $0.10 \pm 0.01$ ), and each showed CO<sub>2</sub>-driven protection against homocitrullination ( $R_{CO_2} = 4.10 \pm 0.18$ ;  $2.80 \pm 0.23$ , Supplementary Table 2). All other identified peptides containing hCit displayed negligible protection ratios with the exception of a tryptic peptide containing K208 ( $R_{CO_2} = 2.3 \pm 0.13$ , Supplementary Table 2), which was observed at 40-fold lower abundance ( $RRI = 0.02 \pm 0.01$ ) as compared to the highest abundance hCit73-containing tryptic peptide ( $RRI = 0.72$ ). Additionally, in our hCit OXA-48 crystal structure, the well-ordered K208 displayed no evidence for homocitrullination (Supplementary Fig. 9). These data show that LysCarComp-MS can identify a known Lys-CO<sub>2</sub> site in a quantitative manner, setting the stage for applying the strategy to complex proteomic mixtures.

We next applied this approach to cell lysates from  $\Delta CynS$  *E. coli* and the well-known CO<sub>2</sub>-responsive cyanobacterium *Synechocystis* sp. 6803 (ref. <sup>3</sup>). We spiked OXA-48 into each sample to serve as an internal positive control. As in  $\Delta CynS$  *E. coli* lysates (Fig. 2d-f), OXA-48 in *Synechocystis* sp. lysates displayed CO<sub>2</sub>-dependent protection in competition assays as determined by assaying both OXA-48 enzyme activity and levels of hCit by immunoblot (Fig. 4a,b).

Furthermore, CO<sub>2</sub>-dependent protection was also observed for a separate protein band (Band 1, Fig. 4b), which could be endogenous RuBisCO (roughly 45 kDa). We subjected these samples to the LysCarComp-MS workflow (Fig. 3a). During MS/MS analyses we considered the diagnostic ImhCit-NH<sub>3</sub> ion ( $m/z$  127.087) to be present if it had a peak intensity value  $>5\%$  of the TMT<sub>127</sub> reporter ion after an initial HCD scan. A subsequent round of HCD involving lower energy deposition was performed on precursor ions meeting this criterion. These combined fragmentation data led to improved identification of modification sites (Supplementary Table 3). We identified 361 and 855 hCit peptides within the respective proteomes of *Synechocystis* sp. and  $\Delta CynS$  *E. coli* (Fig. 4c). For the complete set of hCit sites we observed, as expected, mean  $R_{CO_2}$  values of close to unity ( $\Delta CynS$  *E. coli*,  $0.98 \pm 0.17$ ; *Synechocystis* sp.,  $1.00 \pm 0.08$ ). From among these peptides, we considered a total of eight and 14 candidate Lys-CO<sub>2</sub> sites to be high confidence identifications from *Synechocystis* sp. and  $\Delta CynS$  *E. coli*, respectively, on the basis of five criteria. First, these peptides were observed in at least two experimental replicates. Second, they had a value of Xcorr  $> 2$ . Third, the hCit mapped to a site of missed tryptic cleavage. Fourth, they displayed an  $R_{CO_2}$  value greater than 2 s.d. above the

**Table 1 | Candidate Lys-CO<sub>2</sub> sites from *Synechocystis* sp. proteins identified using LysCarComp-MS**

Sequence	Protein accession	Protein names	<sup>a</sup> R <sub>CO<sub>2</sub></sub> hCit	<sup>c</sup> R <sub>CO<sub>2</sub></sub> non-hCit protein average	<sup>d</sup> RRI	<sup>e</sup> t-test
ANQAFLPASTF( <b>hCit73</b> )IPNSLIALDLGVVKDE HQVFK	Q6XEC0	OXA-48 β-lactamase	2.05 ± 0.75	0.97 ± 0.01	0.02 ± 0.01	0.1
GGLDFT( <b>hCit196</b> )DDENINSQPFMR	P54205	RuBisCO (large chain)	1.85 ± 0.06	1.04 ± 0.02	0.58 ± 0.08	0.0002
LAAPPAA( <b>hCit142</b> )GHVIDIK	P73922	D-fructose 1,6-bisphosphatase class 2	1.57 ± 0.10	1.14 ± 0.04	0.35 ± 0.04	0.009
I( <b>hCit27</b> )AFVTGGAAR	Q01951	Allophycocyanin α-chain	1.35 ± 0.09	1.06 ± 0.04	0.61 ± 0.32	0.008
TGEIGDG( <b>hCit90</b> )IFISPVDSVVR	Q55247	PII	1.34 ± 0.03	0.89 ± 0.03	0.23 ± 0.06	0.05
IASVG( <b>hCit149</b> )TEFVGLMQELYAEPIKPELITK	P73853	Inosine-5'-monophosphate dehydrogenase	1.19 ± 0.04	1.03 ± 0.01	0.05 ± 0.01	0.07
E( <b>hCit177</b> )DGIDFQPITVLLPGEEVPFFFTVK	P10549	Photosystem II manganese-stabilizing polypeptide	1.18 ± 0.07	1.00 ± 0.01	0.03 ± 0.02	0.02
AASVISANAATIV( <b>hCit53</b> )EAVAK	Q01952	Allophycocyanin β-chain	1.17 ± 0.02	1.06 ± 0.04	1.27 ± 0.06	0.004

<sup>a</sup>CO<sub>2</sub> protection ratio ( $R_{CO_2} = I_{NaCl}/I_{NaHCO_3}$ ) for the indicated hCit peptide on the left (in bold, italic), where  $I$  is reporter ion intensity. Note, only hCit peptides with ( $R_{CO_2} > \text{mean global } R_{CO_2} + 2\text{s.d.}$ ) are shown. <sup>b</sup>s.e., standard error (s.d./ $\sqrt{n}$ ), for at least two experimental replicates (Supplementary Data 1). <sup>c</sup>Average  $R_{CO_2}$  for all non-hCit peptides derived from the corresponding protein. <sup>d</sup>RRI is the average  $I_{NaCl}$  for the indicated peptide divided by the global average  $I_{NaCl}$  for all identified peptides. <sup>e</sup>t-test,  $I_{NaHCO_3}$  was normalized to  $I_{NaCl}$ . Subsequently, a one-sample, one-tailed Student's t-test was performed assuming unequal variance. The corresponding P values are shown.

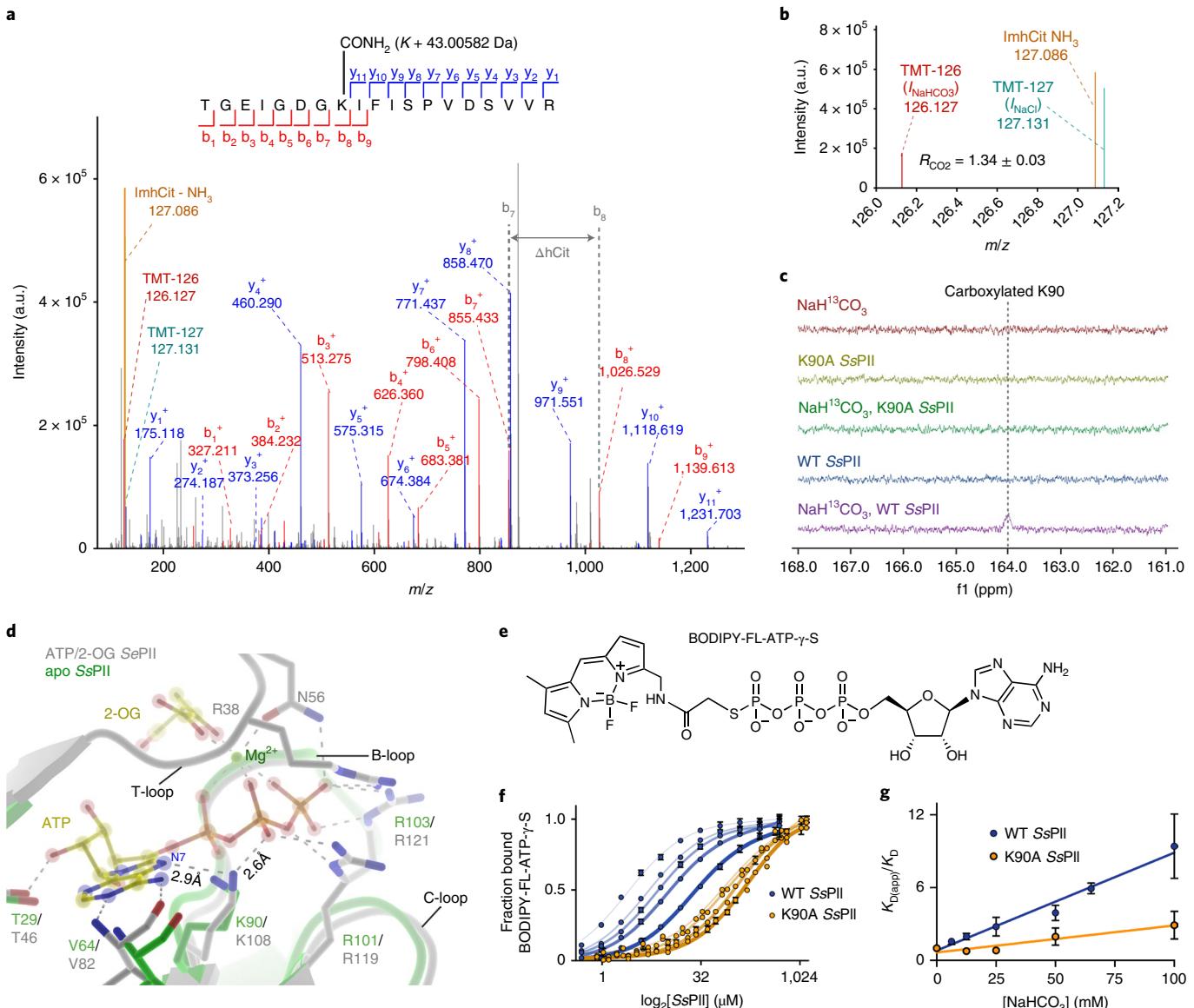
global mean  $R_{CO_2}$  (Fig. 4c,d, Table 1 and Supplementary Table 4). And fifth, they had a statistically significant difference between  $I_{NaCl}$  and  $I_{NaHCO_3}$ . Of note, a hCit73-containing peptide from the added OXA-48 control was observed in the *Synechocystis* sp. dataset and displayed, as expected, a large positive  $R_{CO_2}$  (2.05 ± 0.75). We did not observe the hCit73-containing peptide from the OXA-48 protein control within the  $\Delta CynS$  *E. coli* MS data, but anti-hCit immunoblots confirmed the expected CO<sub>2</sub>-dependent trends in OXA-48 reactivity (Fig. 2e,f). Consistent with our immunoblot experiments (Fig. 4b), hCit was unambiguously mapped to the known Lys-CO<sub>2</sub> site in endogenous RuBisCO from *Synechocystis* sp. (GGLDFThCit( $K_{196} + 43.01$  Da))DDENINSQPFMR, Xcorr = 7.0, Extended Data Fig. 4a). As with OXA-48, a large  $R_{CO_2}$  was observed (1.85 ± 0.06, Table 1 and Extended Data Fig. 4b), consistent with RuBisCO containing a high affinity Lys-CO<sub>2</sub> site that can be observed in crystal structures at high stoichiometry<sup>34,35</sup>. These data show LysCarComp-MS provides precise and accurate detection of sites of Lys-CO<sub>2</sub> within proteomes.

**PII is biochemically regulated by lysine carboxylation.** Given the CO<sub>2</sub>-specific competition for candidate sites identified in our proteomic analyses, we reasoned that CO<sub>2</sub> may modulate the biochemical properties of these modified proteins. From among the 20 total Lys-CO<sub>2</sub> modified proteins identified from both *E. coli* and *Synechocystis* sp. we elected to follow up on one candidate of particularly high interest: the metabolic regulatory protein PII from *Synechocystis* sp. PII proteins comprise an ancient superfamily of signal transduction proteins that are broadly distributed in nature and act to integrate the carbon/nitrogen/energy status of the cell through their ability to form differing protein–protein complexes under different cellular conditions<sup>19</sup>. In cyanobacteria, PII is known to regulate various metabolic pathways including nitrogen uptake systems<sup>36</sup>. Furthermore, PII regulates the global nitrogen transcription factor NtcA by binding its coactivator PipX<sup>37</sup>, arginine biosynthesis by regulating *N*-acetylglutamate kinase<sup>38</sup> and acetyl-CoA levels through regulation of acetyl-CoA carboxylase<sup>39</sup>. Recently, PII has been found to regulate carbon storage metabolism through altering PirC dependent regulation of phosphoglycerate mutase, the enzyme that directs newly fixed CO<sub>2</sub> toward lower glycolysis<sup>40</sup>. The C/N metabolic balance is communicated to PII through its regulation by the effector metabolites ATP, ADP and 2-oxoglutarate

(2-OG), which bind adjacent to a region called the ‘T-loop’ and regulate engagement of PII with binding partners<sup>19</sup>. Therefore, due to the fundamental importance of CO<sub>2</sub> as both an anabolite and regulatory gas in cyanobacteria, coupled with the defined role played by PII as a protein sensor of other regulatory metabolites, we were intrigued by the possibility that it may be biochemically regulated through lysine carboxylation by CO<sub>2</sub>.

The hCit-modified peptide we observed from PII had a high  $R_{CO_2}$  value (1.34 ± 0.03, TGEIGDGhCit( $K_{90} + 43.01$  Da))IFISPVDSVVR, Fig. 5a,b). Examination of our MS/MS data for PII showed the diagnostic ImhCit-NH<sub>3</sub> ion and enabled complete peptide sequencing and accurate localization of hCit to Lys90 (Fig. 5a, Xcorr score = 6.2). To confirm whether Lys90 of PII from *Synechocystis* sp. (SsPII) is carboxylated we performed NaH<sup>13</sup>CO<sub>3</sub> NMR using recombinant WT SsPII and the site-directed variant K90A SsPII. In spectra for WT SsPII we observed a resonance at 164 ppm, which is diagnostic of Lys-<sup>13</sup>CO<sub>2</sub>, that was absent in the spectrum of the K90A variant (Fig. 5c). These data indicate LysCarComp-MS enables accurate identification and precise localization of new Lys-CO<sub>2</sub> sites and maps the site of Lys-CO<sub>2</sub> to the small molecule regulatory pocket of SsPII. Multiple sequence alignment<sup>41</sup> showed that this Lys90 residue is part of an ATP interaction motif that is broadly conserved across PII from various organisms (Extended Data Fig. 5), which suggested to us that it may play an important role in regulating the biochemical functions of PII.

To explore the potential of Lys90-CO<sub>2</sub>-dependent regulation of PII, we performed a structural overlay of apo SsPII (ref. 42) and *Synechocystis elongatus* PII (SePII) bound to ATP and 2-OG (ref. 43) (r.m.s.d. CA; 0.6 Å). This overlay shows that the Lys90 ε-amino group interacts with both the N7 adenosine ring nitrogen and β-phosphate oxygen of ATP (Fig. 5d). Based on these observations, we postulated that formation of Lys90-CO<sub>2</sub> within SsPII may disrupt ATP binding. This would be an important observation since ATP is well known to be a key regulator of various PII–protein complexes<sup>39,40,44</sup>. Accordingly, the ability of CO<sub>2</sub> to antagonize ATP binding would affect the formation of such ATP-regulated complexes. Therefore, to test the effect of Lys90-CO<sub>2</sub> modification on ATP binding, we established a fluorescence polarization (FP) assay for PII using an ATP analog (BODIPY-FL-ATP-γ-S) as a probe (Fig. 5e). We first showed that the K90A variant, in the absence of CO<sub>2</sub>, had a roughly 40-fold lower affinity for this FP probe than WT



**Fig. 5 | The *Synechocystis* sp. metabolic regulatory protein SsPPII is carboxylated at Lys90, inhibiting ATP binding.** **a**, LysCarComp-MS/MS spectra for TMT<sub>duplex</sub>-labeled hCit-containing SsPPII tryptic peptide (TMT-TGEIGDGhCit( $K_{90} + 43.01$ Da))IFISPVDSVVR, MH+(Da) = 2,257.22678. **b**, Closeup of spectra from **a** showing reporter ion intensities and corresponding  $R_{CO_2}$ . The mean  $R_{CO_2} \pm$  s.d. from three experimental replicates is given. **c**, <sup>13</sup>C NMR using purified WT and K90A SsPPII in the presence and absence of 50 mM NaH<sup>13</sup>CO<sub>3</sub>. **d**, Active site overlay of apo SsPPII and ATP/2-OG-bound *Synechococcus elongatus* PII (SePPII) crystal structures (85.7% sequence identity, PDB IDs **1UL3** and **2XUL**). The SsPPII and SePPII protein chains are displayed as green and white cartoons with key residues shown as sticks with atoms colored by type. Bound ATP and 2-OG are yellow sticks, and hydrogen bonds and electrostatic interactions are depicted as gray dashes. **e**, Chemical structure of fluorescent ATP analog used in FP assays. **f**, Influence of NaHCO<sub>3</sub> on FP saturation binding curve of WT and K90A SsPPII. The curves are colored by increasing intensity according to amount of NaHCO<sub>3</sub> including 0, 12.5, 25, 50 and 100 mM. Data are presented as mean values  $\pm$  s.d. from  $n = 4$  independent experiments. **g**, Secondary plots of FP data from **f** showing  $K_{D(\text{app})}/K_D$  for various NaHCO<sub>3</sub> concentrations. Data are presented as mean values with error bars representing s.d. from  $n = 3$  independent experiments.

SsPPII did ( $K_D = 53.6$   $\mu$ M versus  $K_D = 1.3$   $\mu$ M, Extended Data Fig. 6a). In equilibrium competition binding FP experiments performed in the absence of CO<sub>2</sub>, both unlabeled ADP and ATP displayed avid binding to WT SsPPII ( $K_{D(ADP)} = 0.50$   $\mu$ M,  $K_{D(ATP)} = 0.45$   $\mu$ M), whereas our negative control compounds UDP and GTP displayed negligible binding (Extended Data Fig. 6b). Furthermore, the K90A SsPPII variant had impaired binding to all nucleoside di- and triphosphates (Extended Data Fig. 6c). These data validated our FP assay and support its potential for probing the effects of CO<sub>2</sub>-dependent carboxylation of Lys90. Incubation with OCNH revealed time-dependent inhibition of BODIPY-FL-ATP- $\gamma$ -S binding to WT but not to K90A

SsPPII (Extended Data Fig. 6d,e), consistent with homocitrullination of Lys90. Turning to the effect of HCO<sub>3</sub><sup>-</sup>, we observed an increase in the  $K_D$  value for BODIPY-FL-ATP- $\gamma$ -S that depended on the concentration of HCO<sub>3</sub><sup>-</sup> for WT SsPPII, an effect not seen for the K90A variant (Fig. 5f,g). Furthermore, we used biolayer interferometry (BLI) to demonstrate that CO<sub>2</sub> blocks SsPPII binding to an immobilized ATP-biotin probe (Extended Data Fig. 7) to a similar extent as seen by FP, supporting the conclusion that CO<sub>2</sub> impaired binding of ATP to SsPPII. To estimate the apparent affinity of CO<sub>2</sub> for SsPPII we performed equilibrium competition FP experiments and found that CO<sub>2</sub> binds WT SsPPII with a  $K_D$  value of 1.3 mM. In control

experiments, we observed negligible binding of CO<sub>2</sub> to the K90A SsPII variant (Extended Data Fig. 6f). Taken together, these data show that Lys90-CO<sub>2</sub> formation in SsPII is competitive with ATP binding and that its apparent affinity for CO<sub>2</sub> makes it responsive to the physiological concentrations of CO<sub>2</sub> found within cyanobacteria<sup>28,32</sup>. We note that in the absence of CO<sub>2</sub>, when the Lys90 residue is a free amine, the affinity of SsPII for ATP is  $K_D = 2.5 \mu\text{M}$ , whereas when Lys90 exists as Lys90-CO<sub>2</sub> we calculate the affinity for ATP to be  $K_D > 50 \mu\text{M}$  (Fig. 5f,g). We also observed that the equivalent residue within PII from *Arabidopsis thaliana* is carboxylated by CO<sub>2</sub> (Extended Data Fig. 8a–c), and this modification also decreased binding of the BODIPY-FL-ATP-γ-S probe in FP binding assays (Extended Data Fig. 8d). These data support our findings and show that Lys-CO<sub>2</sub> formation provides rheostatic control of ATP binding to SsPII, which would thereby enable CO<sub>2</sub> levels to tune formation of ATP-dependent PII complexes within diverse photosynthetic organisms.

## Discussion

Despite its ubiquitous role as an essential biological gas, little is known regarding the molecular mechanisms by which organisms sense CO<sub>2</sub> (refs. <sup>1,2</sup>). Lys-CO<sub>2</sub> is one appealing mechanism that may dynamically regulate proteins in response to variations in CO<sub>2</sub> (ref. <sup>13</sup>). The rapid and spontaneous decarboxylation of Lys-CO<sub>2</sub> residues, however, complicates study of this modification. Herein, we show that OCNH is selective for native Lys-CO<sub>2</sub> sites, yielding a stable isosteric adduct (hCit) that resembles Lys-CO<sub>2</sub>. Notably, hCit cannot ionize under physiological conditions and should therefore be a useful tool for functional characterization of Lys-CO<sub>2</sub> sites. Leveraging this chemical modification, LysCarComp-MS enables proteome-wide discovery of Lys-CO<sub>2</sub> sites that have CO<sub>2</sub> affinities that fall in a physiologically relevant range. This selective and quantitative method permits ranking of candidate Lys-CO<sub>2</sub> sites and reduces false positives from nonspecific labeling by probes<sup>45</sup>. Nevertheless, downstream validation of candidate sites using orthogonal methods is recommended. In the future, strategies to enrich hCit-modified peptides should increase proteomic coverage and, as such, we anticipate that LysCarComp-MS will enable investigating the ‘lysine carboxylome’, an emerging area that is essential for understanding the molecular basis of CO<sub>2</sub> biology. More generally, we expect this competitive activity-based protein profiling method can serve as a blueprint to monitor other labile posttranslational modifications.

Among the panel of 20 Lys-CO<sub>2</sub> candidate sites we identified using LysCarComp-MS, especially noteworthy from *Synechocystis* sp. are sites within the photosynthetic pigment allophycocyanin, the Calvin cycle enzyme fructose 1,6-bisphosphatase and the metabolic sensor PII. All these proteins play important roles in carbon metabolism<sup>46–48</sup>. We examined the Lys-CO<sub>2</sub> site of SsPII, which regulates SsPII in an unprecedented manner as a negative biochemical effector by antagonizing ATP binding. Given the important role played by ATP in regulating the ability of PII to form protein complexes<sup>39,40,44</sup>, this modification would be expected to enable physiological adaptation through tuning these complexes to meet metabolic demands. Supporting this possibility, we find that levels of Lys-CO<sub>2</sub> within SsPII vary over physiologically relevant ranges of CO<sub>2</sub>. To the best of our knowledge, of the few previously well characterized Lys-CO<sub>2</sub> sites from proteins of any organism<sup>9–13</sup>, few have been shown to have an affinity for CO<sub>2</sub> that would make them responsive to variations in organismal CO<sub>2</sub> levels. Therefore, Lys-CO<sub>2</sub> in SsPII represents a distinct example wherein CO<sub>2</sub> directly regulates protein function over a physiologically relevant range, making it a logical candidate for a physiological sensor of CO<sub>2</sub>.

Given these observations, Lys-CO<sub>2</sub> is likely a largely overlooked molecular mechanism of CO<sub>2</sub> sensing that builds on the prevailing view that sensing of CO<sub>2</sub> occurs through HCO<sub>3</sub><sup>−</sup> (refs. <sup>49,50</sup>). More

generally, we envision that the new chemical modification tools, proteomics datasets and experimental methods described here will help spur rapid advances in uncovering how Lys-CO<sub>2</sub> regulates physiological processes in response to CO<sub>2</sub> in various organisms. Given continuing changes in the levels of atmospheric and oceanic CO<sub>2</sub>, exploiting such methods to advance our understanding of how organisms sense and adapt to variations in CO<sub>2</sub> levels is of vital importance.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41589-022-01043-1>.

Received: 26 April 2021; Accepted: 15 April 2022;

Published online: 16 June 2022

## References

- Cummins, E. P., Strowitzki, M. J. & Taylor, C. T. Mechanisms and consequences of oxygen and carbon dioxide sensing in mammals. *Physiol. Rev.* **100**, 463–488 (2020).
- Cummins, E. P., Selfridge, A. C., Sporn, P. H., Sznajder, J. I. & Taylor, C. T. Carbon dioxide-sensing in organisms and its implications for human disease. *Cell. Mol. Life Sci.* **71**, 831–845 (2014).
- Griffiths, H., Meyer, M. T. & Rickaby, R. E. M. Overcoming adversity through diversity: aquatic carbon concentrating mechanisms. *J. Exp. Bot.* **68**, 3689–3695 (2017).
- Mackinder, L. C. M. et al. A spatial interactome reveals the protein organization of the algal CO<sub>2</sub>-concentrating mechanism. *Cell* **171**, 133–147 (2017).
- Supuran, C. T. Structure and function of carbonic anhydrases. *Biochem. J.* **473**, 2023–2032 (2016).
- Chen, Y. et al. Soluble adenyllyl cyclase as an evolutionarily conserved bicarbonate sensor. *Science* **289**, 625–628 (2000).
- Li, J. et al. Lysine carboxylation in proteins: OXA-10 beta-lactamase. *Proteins* **61**, 246–257 (2005).
- Ewing, S. P., Lockshon, D. & Jencks, W. P. Mechanism of cleavage of carbamate anions. *J. Am. Chem. Soc.* **102**, 3072–3084 (1980).
- 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).
- 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 CO<sub>2</sub> and Mg<sup>2+</sup>. *Biochemistry* **19**, 5321–5328 (1980).
- Jabri, E., Carr, M. B., Hausinger, R. P. & Karplus, P. A. The crystal structure of urease from *Klebsiella aerogenes*. *Science* **268**, 998–1004 (1995).
- Hong, S., Kuo, J. M., Mullins, L. S. & Raushel, F. M. CO<sub>2</sub> is required for the assembly of the binuclear metal center of phosphotriesterase. *J. Am. Chem. Soc.* **117**, 7580–7581 (1995).
- Meigh, L. CO<sub>2</sub> carbamylation of proteins as a mechanism in physiology. *Biochem. Soc. Trans.* **43**, 460–464 (2015).
- Li, Y. et al. Reversible post-translational carboxylation modulates the enzymatic activity of N-acetyl-L-ornithine transcarbamylase. *Biochemistry* **49**, 6887–6895 (2010).
- Stec, B. Structural mechanism of RuBisCO activation by carbamylation of the active site lysine. *Proc. Natl. Acad. Sci. USA* **109**, 18785–18790 (2012).
- Hacker, S. M. et al. Global profiling of lysine reactivity and ligandability in the human proteome. *Nat. Chem.* **9**, 1181–1190 (2017).
- 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).
- Linthwaite, V. L. et al. The identification of carbon dioxide mediated protein post-translational modifications. *Nat. Commun.* **9**, 1–11 (2018).
- Forchhammer, K. & Lüdecke, J. Sensory properties of the PII signalling protein family. *FEBS J.* **283**, 425–437 (2016).
- Evers, J. et al. Molecular structure of isocyanic acid, HNCO, the imide of carbon dioxide. *J. Phys. Chem. A.* **122**, 3287–3292 (2018).
- Lippincott, J. & Apostol, I. Carbamylation of cysteine: a potential artifact in peptide mapping of hemoglobins in the presence of urea. *Anal. Biochem.* **267**, 57–64 (1999).

22. Stark, G. R. Modification of proteins with cyanate. *Enzym. Struct., Part B.* **25**, 590–594 (1967).
23. Stark, G. R., Stein, W. H. & Moore, S. Reactions of the cyanate present in aqueous urea with amino acids and proteins. *J. Biol. Chem.* **235**, 3177–3181 (1960).
24. Steckel, A., Borbely, A., Uray, K. & Schlosser, G. Quantification of the effect of citrulline and homocitrulline residues on the collision-induced fragmentation of peptides. *J. Am. Soc. Mass. Spectrom.* **31**, 1744–1750 (2020).
25. Lorimer, G. H. & Pierce, J. Carbonyl sulfide: an alternate substrate for but not an activator of ribulose-1,5-bisphosphate carboxylase. *J. Biol. Chem.* **264**, 2764–2772 (1989).
26. Yamaguchi, K. & Hausinger, R. P. Substitution of the urease active site carbamate by dithiocarbamate and vanadate. *Biochemistry* **36**, 15118–15122 (1997).
27. Lohans, C. T. et al. (13)C-carbamylation as a mechanistic probe for the inhibition of class D beta-lactamases by avibactam and halide ions. *Org. Biomol. Chem.* **15**, 6024–6032 (2017).
28. Mangan, N. M., Flamholz, A., Hood, R. D., Milo, R. & Savage, D. F. pH determines the energetic efficiency of the cyanobacterial CO<sub>2</sub> concentrating mechanism. *Proc. Natl. Acad. Sci. USA* **113**, E5354–E5362 (2016).
29. Walsh, M. A., Otwinowski, Z., Perrakis, A., Anderson, P. M. & Joachimiak, A. Structure of cyanase reveals that a novel dimeric and decameric arrangement of subunits is required for formation of the enzyme active site. *Structure* **8**, 505–514 (2000).
30. Muroska, J. M., Fu, J. Y., Nguyen, H. H., Ogorzalek Loo, R. R. & Loo, J. A. Leveraging immonium ions for targeting acyl-lysine modifications in proteomic datasets. *Proteomics* **21**, e2000111 (2021).
31. Sanders, J. D., Greer, S. M. & Brodbelt, J. S. Integrating carbamylation and ultraviolet photodissociation mass spectrometry for middle-down proteomics. *Anal. Chem.* **89**, 11772–11778 (2017).
32. Price, G. D., Badger, M. R., Woodger, F. J. & Long, B. M. Advances in understanding the cyanobacterial CO<sub>2</sub>-concentrating-mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants. *J. Exp. Bot.* **59**, 1441–1461 (2008).
33. Thompson, A. et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75**, 1895–1904 (2003).
34. Laing, W. A. & Christeller, J. T. A model for the kinetics of activation and catalysis of ribulose 1,5-bisphosphate carboxylase. *Biochem. J.* **159**, 563–570 (1976).
35. McNevin, D., von Caemmerer, S. & Farquhar, G. Determining RuBisCO activation kinetics and other rate and equilibrium constants by simultaneous multiple non-linear regression of a kinetic model. *J. Exp. Bot.* **57**, 3883–3900 (2006).
36. Watzer, B. et al. The signal transduction Protein PII controls ammonium, nitrate and urea uptake in cyanobacteria. *Front. Microbiol.* **10**, 1428 (2019).
37. Zhao, M. X. et al. Crystal structure of the cyanobacterial signal transduction protein PII in complex with PipX. *J. Mol. Biol.* **402**, 552–559 (2010).
38. Heinrich, A., Maheswaran, M., Ruppert, U. & Forchhammer, K. The *Synechococcus elongatus* PII signal transduction protein controls arginine synthesis by complex formation with *N*-acetyl-L-glutamate kinase. *Mol. Microbiol.* **52**, 1303–1314 (2004).
39. Hauf, W., Schmid, K., Gerhardt, E. C., Huergo, L. F. & Forchhammer, K. Interaction of the nitrogen regulatory protein GlnB (PII) with biotin carboxyl carrier protein (BCCP) controls acetyl-CoA levels in the cyanobacterium *Synechocystis* sp. PCC 6803. *Front. Microbiol.* **7**, 1700 (2016).
40. Orthwein, T. et al. The novel PII-interactor PirC identifies phosphoglycerate mutase as key control point of carbon storage metabolism in cyanobacteria. *Proc. Natl. Acad. Sci. USA* **118**, e2019988118 (2021).
41. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
42. Xu, Y. et al. The structures of the PII proteins from the cyanobacteria *Synechococcus* sp. PCC 7942 and *Synechocystis* sp. PCC 6803. *Acta Crystallogr. D. Biol. Crystallogr.* **D59**, 2183–2190 (2003).
43. Fokina, O., Chellamuthu, V. R., Forchhammer, K. & Zeth, K. Mechanism of 2-oxoglutarate signaling by the *Synechococcus elongatus* PII signal transduction protein. *Proc. Natl. Acad. Sci. USA* **107**, 19760–19765 (2010).
44. Osanai, T., Sato, S., Tabata, S. & Tanaka, K. Identification of PamA as a PII-binding membrane protein important in nitrogen-related and sugar-catabolic gene expression in *Synechocystis* sp. PCC 6803. *J. Biol. Chem.* **280**, 34684–34690 (2005).
45. Wang, S. et al. Advanced activity-based protein profiling application strategies for drug development. *Front. Pharm.* **9**, 353 (2018).
46. Pagels, F., Guedes, A. C., Amaro, H. M., Kijjoa, A. & Vasconcelos, V. Phycobiliproteins from cyanobacteria: chemistry and biotechnological applications. *Biotechnol. Adv.* **37**, 422–443 (2019).
47. De Porcellinis, A. J. et al. Overexpression of bifunctional fructose-1,6-bisphosphatase/sedoheptulose-1,7-bisphosphatase leads to enhanced photosynthesis and global reprogramming of carbon metabolism in *Synechococcus* sp. PCC 7002. *Metab. Eng.* **47**, 170–183 (2018).
48. Selim, K. A., Ermilova, E. & Forchhammer, K. From cyanobacteria to Archaeplastida: new evolutionary insights into PII signalling in the plant kingdom. *N. Phytol.* **227**, 722–731 (2020).
49. Buck, J. & Levin, L. R. Physiological sensing of carbon dioxide/bicarbonate/pH via cyclic nucleotide signaling. *Sensors* **11**, 2112–2128 (2011).
50. Rossetti, T., Jackvony, S., Buck, J. & Levin, L. R. Bicarbonate, carbon dioxide and pH sensing via mammalian bicarbonate-regulated soluble adenylyl cyclase. *Interface Focus* **11**, 20200034 (2021).
51. King, D. T., King, A. M., Lal, S. M., Wright, G. D. & Strynadka, N. C. J. Molecular mechanism of avibactam-mediated beta-lactamase inhibition. *ACS Infect. Dis.* **1**, 175–184 (2015).

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2022

## Methods

**Plasmid construction, protein expression and purification.** The codon optimized *Klebsiella pneumoniae* OXA-48 gene (National Center for Biotechnology Information (NCBI) accession API82700, see Supplementary Table 5 for codon optimized sequence) corresponding to amino acids K22-P265 with the N-terminal signal peptide removed was synthesized and cloned into the NcoI/XbaI restriction sites in the pET28a(+) expression vector by GenScript. Similarly, the full length codon optimized gene sequence for SsPII and AtPII (NCBI accessions CAA66127.1, OAP00825.1; Supplementary Table 5) was cloned into the NcoI/XbaI restriction sites of the pET28a(+) vector with a C-terminal 6X-His tag by GenScript. The K73A OXA-48 and PII carboxylation deficient mutants (SsPII; K90A, AtPII; K164A) were generated by site-directed mutagenesis by GenScript. Purified *Spinacia oleracea* RuBisCO and human carbonic anhydrase I were purchased from MyBioSource and Sigma, dialyzed into storage buffer (50 mM potassium phosphate pH 7.2), flash frozen and stored at -80 °C at 3.5 mg ml<sup>-1</sup>.

Vectors containing the above OXA-48 and PII variants were transformed into chemically competent BL21(DE3) *E. coli* host cells (Invitrogen, C6000-03). Transformed cells were grown at 37 °C with 200 r.p.m. of shaking until an optical density ( $OD_{600}$ ) of 0.7 was reached. Samples were cooled to 25 °C, expression was induced by addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside and the cultures were incubated with 200 r.p.m. shaking at 25 °C for 16 h. Cell pellets from 61 of culture were resuspended in 100 ml lysis buffer (OXA-48; 50 mM HEPES pH 7.5, 350 mM NaCl, 20 mM imidazole, 1 mg ml<sup>-1</sup> lysozyme, 2 EDTA free protease inhibitor tablets (PIT) from Roche and PII; 50 mM HEPES pH 7.5, 300 mM NaCl, 2 PIT, 1 mg ml<sup>-1</sup> lysozyme) and lysed by sonication. Cell lysates were centrifuged at 14,600g for 20 min to remove unbroken cells and inclusion bodies. Supernatants were then centrifuged at 45,000g for 45 min to remove membranes. Proteins were purified from the supernatant using nickel chelation chromatography (5 ml HisTrap FF, GE Healthcare). The column was equilibrated with 10 column volumes of equilibration buffer (OXA-48; 50 mM HEPES pH 7.5, 350 mM NaCl, 20 mM imidazole, and PII; 50 mM HEPES pH 7.5, 300 mM NaCl, 10 mM imidazole) and eluted using a linear gradient of imidazole from 20–500 mM over 50 min. Fractions containing purified OXA-48 and PII variants were exchanged by overnight dialysis (10kDa cutoff membrane) into assay buffer (OXA-48; 100 mM potassium phosphate pH 7.2, 150 mM NaCl or SsPII; 100 mM potassium phosphate pH 7.2) or crystallization buffer (50 mM HEPES pH 7.5, 150 mM NaCl). Purified proteins were flash frozen and stored at -80 °C until use.

**OXA-48 kinetic assays.** Nitrocefin was purchased from Calbiochem. Kinetic parameters for WT and K73A OXA-48 catalyzed nitrocefin hydrolysis were determined at 30 °C in assay buffer in a final volume of 100 µl unless otherwise stated. This assay buffer resists changes in pH (>0.1) induced by addition of NaHCO<sub>3</sub> over the concentration range used (Supplementary Fig. 6). Unless otherwise stated, all buffer components were thoroughly degassed under vacuum before kinetic or competition assays to minimize residual dissolved CO<sub>2</sub>. Rates of hydrolysis were monitored continuously at 486 nm in 384-well clear microplates (NUNC, 262160) using a Synergy Neo2 hybrid multi-mode plate reader operated using the Gen5 (v.3.0) software package. Unless indicated otherwise, individual assays were set up such that OXA-48 and NaHCO<sub>3</sub> were preincubated in a sealed mixing plate for 15 min before addition of nitrocefin to enable equilibration of Lys-CO<sub>2</sub> and prevent CO<sub>2</sub> escape before the start of the reaction. Unless otherwise stated, plates were sealed using Corning universal optical tape before kinetics experiments to prevent gas leakage. Reaction velocities were assessed by linear regression of the progress curves over a 10-min period.

For nitrocefin and NaHCO<sub>3</sub> concentration response experiments, a final enzyme concentration of 100 pM was used. Inhibition (IC<sub>50</sub>) experiments involved preincubation of OXA-48 with inhibitor for 1 h in a 90-µl volume followed by addition of 10 µl of (2 mM nitrocefin, 500 mM NaHCO<sub>3</sub>) to start the reaction. IC<sub>50</sub> is defined as the concentration of inhibitor required to reduce substrate hydrolysis by 50% and was determined by fitting the data to a sigmoidal four-parameter dose-response curve using GraphPad Prism (v.9). In jump-dilution time course inhibition experiments, 1 µM of OXA-48 was preincubated for various timepoints with the indicated concentrations of various salts (KOCN, NaHCO<sub>3</sub>, NaCl) in a 100 µl reaction volume. Following preincubation, the protein was jump diluted 10,000-fold into read buffer (150 µM nitrocefin, 50 mM NaHCO<sub>3</sub>, 100 mM potassium phosphate pH 7.2) and activity was monitored as described above. Reaction rates for the various KOCN treatments were normalized to the minus KOCN control at each timepoint to derive fractional activity, and the resulting data were fit to a one phase exponential decay to determine k<sub>obs</sub> at each KOCN concentration. Nonlinear least squared regression was used to fit these data to the Michaelis-Menten equation using GraphPad Prism (v.9) to determine k<sub>inact</sub> and K<sub>i</sub>. For decarbamylation experiments, either 50 mM KOCN or 50 mM NaCl was incubated with 10 µM OXA-48 for 2 h to facilitate full enzyme inactivation. The inhibited enzyme was then desaltsed using Zebra spin columns according to the manufacturer's instructions and jump diluted 1,000-fold into assay buffer and incubated at 25 °C. Following dilution, the return of activity was monitored by periodically adding 10 µl aliquots of the protein to 90 µl of read buffer at various timepoints over a 3-d period. Reaction rates were normalized to the 50 mM NaCl control at each timepoint and data were plotted in GraphPad Prism (v.9).

Competition assays were performed by mixing nonprotein reaction components in a 10 µl volume in competition buffer (200 µM potassium phosphate pH 7.2). The competition reaction was started by addition of an equal volume containing either 10 µM purified OXA-48/RuBisCO or 400 µg of cell lysate containing 10 µM OXA-48. Competition assays were incubated for 1 h before 10,000-fold dilution in read buffer and monitoring of progress curves as described above. During incubation and kinetic monitoring, the reactions were covered using gas impermeable sealing tape to prevent CO<sub>2</sub> escape. Reaction rates were determined as above and fractional activities were determined by normalization to the zero-competitor condition.

**Circular dichroism spectroscopy.** OXA-48 protein (5 mg ml<sup>-1</sup>) was incubated with various salts (50 mM KOCN, 50 mM NaHCO<sub>3</sub>) for 2 h and then desaltsed by two passes through a Zebra spin column. Following desalting, the protein concentration was determined by nanodrop and then diluted to 0.1 mg ml<sup>-1</sup> in circular dichroism buffer (10 mM potassium phosphate pH 7.2). Far-ultraviolet circular dichroism experiments were performed using a Chirascan Plus circular dichroism spectrometer (Applied Photophysics) operated using the Chirascan Pro-Data SX software (v.2.5.0), using quartz cuvettes with a 2-mm path length. Circular dichroism spectral scanning data were collected at 25 °C, with a wavelength interval of 1.0 nm. Wavelength scans were taken from 195 to 300 nm for 500 µl OXA-48 samples prepared as described above. The millidegree output units were converted to mean residue ellipticity (εMR) and plotted as a function of wavelength using GraphPad Prism (v.9).

**Intact protein MS.** A total of 100 µl of 5 mg ml<sup>-1</sup> WT or K73A OXA-48 were incubated in the presence or absence of 50 mM KOCN for 2 h. Following incubation, samples were desaltsed using Zebra desalting columns according to the manufacturer's instructions and dialyzed for 4 h into 10 mM Tris pH 7. Following dialysis, samples were diluted to 0.2 mg ml<sup>-1</sup> in 0.1% formic acid. The Waters Acquity chromatography system was coupled online with an LTQ Velos Pro Orbitrap Mass spectrometer (Thermo Fisher). A 1-µl aliquot of each protein solution was separated by reversed-phase liquid chromatography using an Acquity UPLC BEH C8 Column (130 Å, 1.7 µm, 2.1 × 50 mm) heated to 35 °C. Solvents were: A, 0.1% formic acid and B, 100% acetonitrile, 0.1% formic acid. Samples were separated using an 18-minute gradient (0–2 min 2% B; 15 min 80% B, hold 2 min 80% B) at 200 µl min<sup>-1</sup>.

The Orbitrap instrument parameters were as follows: nano-electrospray ion source with spray voltage 3.6 kV, capillary temperature 325 °C, S-lens radio frequency level 68, heater temperature 350 °C and sheath gas and auxiliary gas flowrates 25 and 10, respectively. The MS scan mass range was 300–2,000 m/z, 18 min acquisition, resolution was 60,000 full-width at half-maximum at 200 m/z, automatic gain control (AGC) target was  $1 \times 10^5$  and maximum fill time was 500 ms and 1 µs scan time. The lock mass for siloxane (391.28428 Da) was used for internal calibration. The Thermo protein deconvolution software package (v.3) was used to analyze the intact mass data. The following auto Xtract isotopically resolved settings were used: resolution at 400 m/z, 60,000; signal to noise threshold, 3; m/z range, 600–2,000 and charge carrier, H<sup>+</sup>; and fit factor, 80%; remainder threshold, 25%; charge range, 5–50; minimum intensity, 1 and expected intensity error, 3. Data were exported to Excel and depicted graphically using GraphPad Prism (v.9).

**X-ray crystallography.** Before crystallization, 20 mg ml<sup>-1</sup> OXA-48 was incubated in the presence of 50 mM KOCN for 1 h. Subsequently, nitrocefin kinetic experiments confirmed OXA-48 was fully inactivated. Homocitrullinated OXA-48 crystals were then grown using the sitting drop vapor diffusion method at 25 °C. Drops contained 1 µl of the homocitrullinated OXA-48 protein combined with an equal volume of precipitant (0.2 M lithium sulfate, 0.1 M Bis-Tris pH 5.5, 25% w/v PEG 3350). Drops were streak seeded with finely crushed apo OXA-48 crystals pregrown in the same condition using a cat whisker. Crystals were soaked for 30 s in cryoprotectant solution (mother liquor, 30% glycerol, 50 mM KOCN), and flash vitrified in liquid nitrogen. The hCitr OXA-48 crystals diffracted to 2.2 Å resolution at beamline CMCF-08ID-1 of the Canadian Light Source in Saskatoon Saskatchewan using the MxDc data collection software package (v.2020.8.21). All crystallographic data in this study were collected from a single crystal at a temperature of 100 K and wavelength of 1.0 Å.

Data were processed using IMOSFLM (v.7.4.0)<sup>52</sup> and CCP4 (v.7.1.018)<sup>53</sup>. During refinement, a total of 5% of reflections was set aside for cross validation. The structure was solved by molecular replacement using the program Phaser-MR (v.1.20.1)<sup>54</sup>, with chain A of the native crystal structure (PDB ID 4S2P)<sup>55</sup> with lysine in place of Lys-CO<sub>2</sub> at amino acid 73 as a starting model. Several iterations of manual rebuilding in coot (v.0.9.3)<sup>56</sup>, followed by refinement using Phenix (v.1.18.2)<sup>57</sup> were carried out. The structure was refined with isotropic B factors. Water, ions, sulfate, chloride and hCitr were added manually by examination of the Fo-Fc and 2Fo-Fc electron density maps, and all hCitr atoms were refined at full occupancy. The final refined model has excellent Ramachandran statistics (favored 96.84%, allowed 2.95%, outliers 0.21%). Of note, there are two Ramachandran outliers corresponding to residue S155 from each protein chain. We note that the S155 electron density is well defined and was also observed as an outlier in earlier published structures and, thus, the observed phi/psi angles are likely facilitated

by the local hydrogen bonding network. The hCit OXA-48 crystal structure has been deposited in the PDB with accession code 7LXG. PyMol (v.2.3.5) was used to construct the following structural figures; Figs. 1g–h and 5d, Extended Data Figs. 2h and 8a and Supplementary Fig. 9.

**13C NMR spectroscopy.** <sup>13</sup>C NMR spectra were acquired using an AVANCE II 600 spectrometer equipped with a 5-mm QNP cryoprobe. <sup>13</sup>C spectra consisted of 18,432 scans, using a 2.0-s relaxation delay. Protein samples consisted of 1 mM OXA-48 (WT/K73A), S<sub>n</sub>PII (WT/K90A) or A<sub>t</sub>PII (WT/K164A) protein variants prepared in NMR buffer (100 mM potassium phosphate pH 7.2, 10% v/v D<sub>2</sub>O) in the presence or absence of 50 mM NaH<sup>13</sup>CO<sub>3</sub>. For NMR competition assays, both 50 mM KOCN and 50 mM NaH<sup>13</sup>CO<sub>3</sub> were added at the same time and incubated for 2 h at 25 °C before NMR. In KO<sup>13</sup>CN labeling experiments, 2 mM WT or K73A OXA-48 was preincubated for 1 h in NMR buffer in the presence or absence of 50 mM KO<sup>13</sup>CN before <sup>13</sup>C NMR as described above. All experiments were measured in a 300-μl Shigemi tube at 298 K. Data were analyzed in topspin (v.3.6) and MestReNova (v.14.1).

**N-terminal sequencing.** Purified recombinant OXA-48 was prepared as described above. A total of 100 μg OXA-48 was then incubated in the presence or absence of 50 mM KOCN for 1 h. Following incubation, the protein samples were desalted using Zebra spin columns according to the manufacturer's instructions and dialyzed overnight into PBS. These protein samples (2.5 mg ml<sup>-1</sup>) were then mixed with an equal amount of *Gallus gallus* lysozyme internal control protein (Bioshop, product number LYS702.25) and sent to Creative Proteomics for N-terminal sequencing analysis. A total of 9 μg of these protein samples were blotted onto polyvinyl difluoride membranes and separately analyzed by Edman degradation sequencing on a PPSQ-33A (Shimadzu) sequencer using the standard protocol<sup>57</sup>.

**Turbidity assays.** *E. coli* lysates (20 mg ml<sup>-1</sup>) were incubated in the presence or absence of 50 mM KOCN, 50 mM NaHCO<sub>3</sub> or 50 mM NaCl in a 100-μl reaction volume in buffer (50 mM sodium phosphate pH 7.2). Samples were immediately added to NUNC 384-well flat bottom clear assay plates and sealed using Corning universal optical sealing tape. Absorbance was monitored continuously at 405 nm at a rate of one read per minute using a Synergy Neo2 hybrid multi-mode plate reader operated using the Gen5 (v.3.0) software package. These assays were repeated at two different temperatures (25 and 40 °C). Absorbances were plotted as a function of time in GraphPad Prism (v.9).

**Cell culture.** Cyanase single gene knockout *E. coli* cells ( $\Delta$ CynS, clone catalog number OEC4987-213605890) were attained from the Keio collection (Dharmacon). *E. coli* cells were grown in a 1 l culture of Luria Bertani medium with 200 r.p.m. shaking at 37 °C overnight and then collected. *Synechocystis* sp. 6803 (ATCC 27184) cells were attained from CedarLane. *Synechocystis* sp. cells were cultured using 4-l Erlenmeyer flasks in 1 l of BG-11 media (Thermo Fisher, A1379901) retrofitted with a fish tank pump in line with a high-efficiency particulate absorbing filter to facilitate clean air exchange. *Synechocystis* sp. cells were grown under 4–5 klux illumination using a 150-W LED growth light and were collected at late log phase after 18 d of continuous growth. Collected *E. coli* or *Synechocystis* sp. cells were resuspended in lysis buffer (50 ml PBS pH 7.2, 2 PIT, 1 mg ml<sup>-1</sup> lysozyme) and lysed by sonication followed by 20 passes through a Dounce homogenizer. Lysed cells were centrifuged at 45,000 g for 1 h and the supernatant was collected, proteins were quantified and samples were flash frozen using liquid nitrogen.

**Anti-hCit Immunoblots.** Competition assays were performed exactly as described in the above kinetics section, with the exception that reactions were stopped by addition of an equal volume of 2× Laemmli buffer followed by boiling. A total of 1 μg of purified protein or 20 μg of cellular lysate per well was loaded onto a 10% SDS-PAGE unless otherwise stated. Gels were electrophoresed at 70 V for 1.5 h and protein bands were transferred to 0.45-μm nitrocellulose membranes using the TransBlot Turbo (Bio-Rad) semidry transfer system according to manufacturer's instructions. Blots were incubated in blocking buffer (4% bovine serum albumin, 1× PBS pH 7.4, 0.1% tween 20) on a rocker at 4 °C for 2 h before addition of a 1:2,000 dilution of 1 mg ml<sup>-1</sup> primary antibody (goat polyclonal pan-specific antihomocitrulline antibody (MyBioSource, MBS390103) and mouse monoclonal anti-His antibody (Invitrogen, MA1-21315)). The blots were incubated with primary antibody overnight and imaged using a 1:20,000 dilution of the appropriate LI-COR fluorescent secondary antibodies (IRDye 800CW donkey anti-goat IgG (LI-COR, 926-32214) and IRDye 680LT goat anti-mouse (LI-COR, 926-68020)), and a gel imager system using the LI-COR Image Studio software (v.3.0) according to the manufacturer's instructions. Densitometry analysis was carried out using ImageJ (v.1.51), and charts were made in GraphPad Prism (v.9).

**In-solution tryptic digests and Isobaric labeling.** Competition assays were performed as described above in 50-μl reaction volumes. Following competition, samples were immediately desalted using Zebra spin desalting columns according to the manufacturer's instructions (Thermo Scientific, 89882). Samples were

reduced (5 mM dithiothreitol, 50 °C, 45 min) and alkylated (10 mM iodoacetamide, 25 °C, 30 min, dark), followed by acetone precipitation. Precipitated pellets were resuspended in digestion solution (50 mM HEPES pH 8, 4 μg of trypsin) and incubated for 18 h at 37 °C. Following incubation, peptides were dialyzed for 2 h into 0.1% trifluoroacetic acid using 1 kDa cutoff Pur-A-Lyzer Midi dialysis kits (Sigma, PURD10005). Peptides were then dried down by SpeedVac. Peptides were then resuspended in 100 μl of 50 mM HEPES pH 8. A total of 160 μg of the appropriate TMT<sub>duplex</sub> isobaric tags were then added to the 100 μl of tryptic digest (50 μg) from the above competition samples. Reactions were incubated at 25 °C for 1 h with shaking at 600 r.p.m. Subsequently, 8 μl of 5% hydroxylamine was added and the samples were incubated for 15 min at 25 °C to quench the reaction. Quenched NaCl and NaHCO<sub>3</sub> reactions were then mixed together and desalted by overnight dialysis in 0.1% trifluoroacetic acid using the above-mentioned 1 kDa cutoff dialysis units. Peptides were then quantified by nanodrop and dried down.

**Liquid chromatography–MS/MS (LC-MS/MS).** MS experiments were performed at UVic-Genome BC Proteomics Centre, Victoria, Canada. Each dried peptide sample was rehydrated to 1 μg μl<sup>-1</sup> with (2% acetonitrile, 0.1% formic acid). A 1-μl aliquot of the peptide solution was separated by online reversed-phase liquid chromatography using a Thermo Scientific EASY-nanoLC II system with an in-house prepared Microsorb 300-3 C8 reversed-phase precolumn (100 μm inner diameter, 1.5 cm long, 3 μm, 300 Å), and a reversed-phase nano-analytical column Microsorb 300-3 C8 column (75 μm inner diameter, 20 cm long, 3 μm, 300 Å), at a flow rate of 300 nl min<sup>-1</sup>. The chromatography system was coupled online with an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific) equipped with a Nanospray Flex NG source (Thermo Fisher Scientific). Solvents were: A, 0.1% formic acid and B, 90% acetonitrile and 0.1% formic acid. After a 248 bar (roughly 4 μl) precolumn equilibration and 248 bar (roughly 3 μl) nanocolumn equilibration, samples were separated using a 140-min gradient (0 min 5% B; 5 min 8% B; 120 min 55% B; 5 min 100% B; hold 10 min 100% B).

The Orbitrap Fusion instrument parameters (Fusion Tune v.3.3 software) for the Orbitrap-Orbitrap HCD induced fragmentation scan (MS/MS) for peptide identification and generation of TMT reporter ions were as follows: nano-electrospray ion source with a spray voltage of 2.55 kV and capillary temperature of 275 °C. The survey MS1 scan *m/z* range was set to 350–200 in profile mode, resolution 120,000 full-width at half-maximum at 200 *m/z*, AGC target standard, one microscan with auto maximum inject time for maximum sensitivity and scan-rate. The lock mass for siloxane (391.28428 Da) was used for internal calibration. Data-dependent acquisition Orbitrap survey spectra were scheduled at least every 3 s, with the software determining 'maximum' number of MS/MS acquisitions during this period. The most intense ions with a 2–7 charge state exceeding 50,000 counts were selected for HCD MS/MS fragmentation with detection in centroid mode. Dynamic exclusion settings are, repeat count of 2; if occurring in 5 s and exclusion duration 5 s with a 10 ppm mass window. The ddMS/MS OT HCD scan used a quadrupole isolation window of 1.6 *m/z*; Orbitrap detector resolution 50,000, auto *m/z* scan range mode, first *m/z* 100, centroid detection, 1 microscan, AGC target value  $1 \times 10^5$ , 105 ms maximum injection time and fixed HCD collision energy 38%. An additional lower HCD collision energy (32%) mass triggered Orbitrap HCD MS/MS scan occurred if the carbamyl-lysine [Immonium-NH<sub>3</sub><sup>+</sup>] ion at 127.0863 *m/z* (+10 ppm) was observed at 5% relative intensity to the 127.131079 TMT reporter ion to give further confirmation for the desired homocitrulline modification.

**LysCarComp–MS peptide and protein identification.** The *Synechocystis* sp. and  $\Delta$ CynS *E. coli* proteome datasets were processed with raw files being created by XCalibur v.3.0.63 software and analyzed with the Proteome Discoverer software (v.1.4.0.228, Thermo Scientific). Data were searched against the corresponding Uniprot reference proteomes (*E. coli* K12: UP00000625, *Synechocystis* sp. 6803: UP000001425), with *K. pneumoniae* OXA-48 sequence added (UniProt accession Q6XEC0) in FASTA format. Default search settings were used unless stated otherwise. Search parameters used a MS/MS tolerance of 15 mmu with fixed modifications: methylthio (C), TMT<sub>duplex</sub> (N terminus) and variable modifications: TMT<sub>duplex</sub> (K), carbamylation (K + 43.00582 Da). Of note, we do not include N-terminal carbamylation events in our proteomic analysis. The processing step workflow was set up with spectrum files being run through a reporter ion quantifier (integration tolerance 20 ppm, most confident centroid mode) and a separate spectrum selector followed by two branching scan event filters with corresponding SEQUEST-HT database searches. The scan event filters had the following common parameters: Fourier transform MS mass analyzer, MS/MS order and HCD activation. However, the event filters differed with respect to collision energy range (30–34% or 36–39%). The parallel SEQUEST-HT searches have the following parameters in common, maximum missed tryptic sites: 3, peptide length range: 6–150 amino acids, precursor ion mass tolerance 10 ppm and fragment mass tolerance 0.05 Da. However, the two searches had different fragment mass tolerances of 0.05 or 0.3 Da. Peptide spectrum match (PSM) identified proteins were filtered using a target-decoy approach at a false discovery rate of 1%. After search, SEQUEST-HT cross correlation scores (Xcorr) were calculated and PSMs filtered using a fixed value PSM validator with default settings. Consensus workflow for reporter ion quantification was performed using default settings.

Peptides corresponding to contaminant proteins were manually removed from the dataset. Results were then exported to Excel files for further processing.

**LysCarComp–MS data analysis.** Each reporter ion intensity was normalized to the global nonhomocitrullinated average for each experimental condition separately. Peptides bearing a homocitrulline residue located at a missed tryptic site with Xcorr scores >2 and having detectable reporter ion intensities in at least two different experimental replicates were taken forward for further analysis. Reporter ion intensities were then used to calculate a lysine CO<sub>2</sub> protection ratio ( $R_{\text{CO}_2}$ ) according to the following equation:

$$R_{\text{CO}_2} = \frac{I_{\text{NaCl}}}{I_{\text{NaHCO}_3}} \quad (1)$$

where  $I_{\text{NaHCO}_3}$  and  $I_{\text{NaCl}}$  represent the normalized reporter ion intensities for the NaHCO<sub>3</sub> and NaCl treated samples for any given peptide. Homocitrullinated peptides with an  $R_{\text{CO}_2}$  greater than 2 s.d. above the global mean  $R_{\text{CO}_2}$  for all peptides were considered candidate Lys-CO<sub>2</sub> sites. For each peptide meeting the above criteria, when possible, the combined average  $R_{\text{CO}_2}$  from all nonhomocitrullinated peptides from within the same protein was calculated. This enabled us to detect any overall differences in protein abundance between treatments. Furthermore, we performed a statistical comparison of  $I_{\text{NaCl}}$  and  $I_{\text{NaHCO}_3}$ . For this analysis, the  $I_{\text{NaHCO}_3}$  reporter ion intensity was normalized to  $I_{\text{NaCl}}$  within each TMT<sub>duplex</sub> replicate for each identified peptide so as to account for variation between runs. Subsequently, we performed a one-sample, one-tailed Student's *t*-test assuming unequal variance to test whether there is a significant difference ( $P \leq 0.1$ )<sup>38</sup> between  $I_{\text{NaCl}}$  and  $I_{\text{NaHCO}_3}$  for each identified peptide. All processed MS/MS data corresponding to peptides observed in at least two experimental replicates including the above-mentioned parameters are supplied in a comprehensive supplementary Excel spreadsheet (Supplementary Data 1) and displayed for top Lys-CO<sub>2</sub> candidates in Table 1 and Supplementary Table 4.

Additionally, OCNH reacted OXA-48 LC–MS/MS data were analyzed for the presence of carbamylated adducts with the various nucleophilic amino acids. For this analysis, the OXA-48 LC–MS/MS spectra for each carbamylated (+43.00582 Da) peptide was assessed for the presence of OCNH-containing immunion ions at the following characteristic *m/z* values (Lys 127.1, Glu 145.1, Asp 131.1, Arg 172.2, His 152.2, Cys 119.2, Met 149.2, Ser 103.1, Thr 117.1 and Tyr 178.2). All ions having a relative intensity greater than 5% of the  $I_{\text{NaCl}}$  reporter ion were counted. These results are displayed as a bar graph in Supplementary Fig. 2.

**FP.** BODIPY-FL-ATP- $\gamma$ -S was purchased from Thermo Fisher. Saturation binding FP experiments were performed in a 100- $\mu$ l volume in assay buffer (100 mM HEPES pH 7.0) at a constant probe concentration of 20 nM, and SsPII and AtPII protein concentrations ranged from 0 to 1.3 mM. Assays were performed at 30 °C in Nunc 384-well low volume black plates using a Synergy Neo2 HTS multi-mode microplate reader (BioTek) using the Gen5 (v.3.0) software package. The fluorescein filter set was used (excitation wavelength 485 nm, emission wavelength 528 nm), with a 20-nm bandwidth and 50% gain setting. All reported data were corrected for background of free probe in the absence of protein. To determine probe  $K_D$  values, the data were fit to a single site-specific binding model in GraphPad Prism (v.9).

FP equilibrium competition binding experiments were performed in reaction buffer using 20 nM probe and concentrations of WT SsPII, K90A SsPII, WT AtPII and K164A AtPII equal to their respective  $K_D$  (1.3, 53.6, 0.6 and 1.5  $\mu$ M) in the presence of varying amounts of competitors (ATP, ADP, GTP, UDP, 2-OG and NaHCO<sub>3</sub>). For all FP assays, the fully assembled experimental samples containing all assay components were incubated at room temperature for 15 min in sealed plates to enable time for equilibration before taking reads. Furthermore, we ensured that relative FP was stable over time. Equilibrium competition data were fit to a four-parameter sigmoidal function in GraphPad Prism to determine IC<sub>50</sub>s and values were input into a standard FP equilibrium binding equation to determine  $K_D$  for the unlabeled competitor molecule as previously described<sup>39</sup>. Here and throughout, wherever applicable, the [CO<sub>2</sub>] was determined from total C<sub>i</sub> using well-established standard equations for the carbonate equilibria<sup>40</sup>.

**BLI.** BLI was performed using the Octet RED96e eight-channel system (Satorius) with high-precision streptavidin biosensors (catalog no. 18-5019). The  $\gamma$ -PEG8-ATP-biotin probe was purchased from Jena Bioscience (catalog no. NU-926-BIO). BLI was performed at 30 °C in a 96-well plate (Greiner Bio-One, catalog no. 655209) in a 200- $\mu$ l well volume. At the start of the experiments, the sensors were preequilibrated in read buffer (200 mM HEPES pH 7.2, 0.02% Tween 20, 0.1% BSA, 5 mM MgCl<sub>2</sub>) for 30 min. Next, the  $\gamma$ -PEG8-ATP-biotin probe was loaded onto the sensors for 5 min at 500 nM in the presence of 2  $\mu$ M biotin. Following loading, the sensors were rinsed in read buffer and a baseline was collected for 60 s. Subsequently, the loaded sensors were dipped into read buffer plus various concentrations of WT- or K90A SsPII (10 nM–16  $\mu$ M), either in the presence or absence of 50 mM NaHCO<sub>3</sub>, and association was monitored for 240 s. Subsequently, the SsPII bound sensors were dipped into fresh buffer either in the presence or absence of 50 mM NaHCO<sub>3</sub>, and dissociation was monitored for 300 s in read buffer. A no SsPII control was included in each experiment, and this baseline

was used for normalization of all other reads at each timepoint. As an additional control, a well lacking the  $\gamma$ -PEG8-ATP-biotin probe in the presence of 16  $\mu$ M SsPII was included to demonstrate minimal nonspecific binding to the sensor. For each experimental sample, the response in equilibrium ( $R_{\text{eq}}$ ) was calculated using the data analysis software of the Octet system (v.12). Data were fit to either a one-analyte or two-analyte binding model. The association/dissociation curves and the corresponding relative wavelength shifts were plotted using the Octet Satorius software package as displayed in Extended Data Fig. 7.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

All raw proteomics data were deposited to the ProteomeXchange Consortium via the PRIDE partner repository, with the dataset identifier (PXD031976). The hCit OXA-48 crystal structure has been deposited in the PDB with accession code 7LXG. The PDB codes for previously determined X-ray crystallographic structures used in the preparation of this manuscript are as follows: OXA-48 (4S2P), apo *Synechocystis* sp. PII (1UL3), ATP/2-OG bound *S. elongatus* PII (2XUL) and *A. thaliana* PII (2O66). The NCBI accession codes used in generating recombinant proteins for this study are as follows: *K. pneumoniae* OXA-48 (AP182700), *Synechocystis* sp. PII (CAA6127.1) and *A. thaliana* PII (OAP00825.1). UniProt reference proteomes used in LysCarComp–MS analysis are as follows: *E. coli* K12 (UP000000625), *Synechocystis* sp. 6803 (UP000001425) and *K. pneumoniae* OXA-48 (Q6XEC0). All other data needed to evaluate the conclusions in the manuscript are available within the main text or supplementary materials. Source data are provided with this paper.

## Code availability

No custom code was generated.

## References

52. Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 271–281 (2011).
53. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 235–242 (2011).
54. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
55. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).
56. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D. Biol. Crystallogr.* **68**, 352–367 (2012).
57. Speicher, K. D., Gorman, N. & Speicher, D. W. N-terminal sequence analysis of proteins and peptides. *Curr. Protoc. Protein Sci.* Chapter: Unit-11.10 (NIH, 2001).
58. Neilson, K. A. et al. The influence of signals from chilled roots on the proteome of shoot tissues in rice seedlings. *Proteomics* **13**, 1922–1933 (2013).
59. Rossi, A. M. & Taylor, C. W. Analysis of protein-ligand interactions by fluorescence polarization. *Nat. Protoc.* **6**, 365–387 (2011).
60. Butler, J. N. *Carbon Dioxide Equilibria and their Applications* (Lewis Publishers Inc., 1991).

## Acknowledgements

We are grateful to the Natural Sciences and Engineering Research Council of Canada (NSERC) (grant no. RGPIN298406) and the Human Frontiers Science Program (grant no. RGP0058/2020) for supporting this research. D.J.V. thanks the Canada Research Chairs program for support as a Tier I Canada Research Chair in Chemical Biology. D.T.K. is supported by a postdoctoral fellowship from the Canadian Institutes of Health Research (CIHR) and a Trainee Award from the Michael Smith Foundation for Health Research and Pacific Alzheimer Research Foundation. We acknowledge the UVic-Genome BC Proteomics Centre, Victoria, Canada for performing all MS experiments. X-ray crystallography was performed at the Canadian Light Source synchrotron facility, a national research facility of the University of Saskatchewan, which is supported by the Canada Foundation for Innovation (CFI), NSERC, the National Research Council (NRC), CIHR, the Government of Saskatchewan and the University of Saskatchewan. We are grateful to L. Craig, S. Cecioni, Y. Zhu and M. Alteeen for their expert input. Parts of the schematics were generated using BioRender.com.

## Author contributions

D.J.V. and D.T.K. conceived and designed experiments. D.T.K., Z.M. and S.K. performed plasmid construction and protein purification. D.T.K. performed biochemical experiments, MS preparations and data analysis. D.T.K. and J.E.S.N. performed X-ray crystallography and data analysis. D.B.H. performed MS, and D.T.K. and D.B.H. analyzed the MS data. S.Z. performed the NMR experiments and analyzed the data. D.J.V. and D.T.K. analyzed other data and wrote the manuscript with input from all.

**Competing interests**

The authors declare no competing interests.

**Additional information**

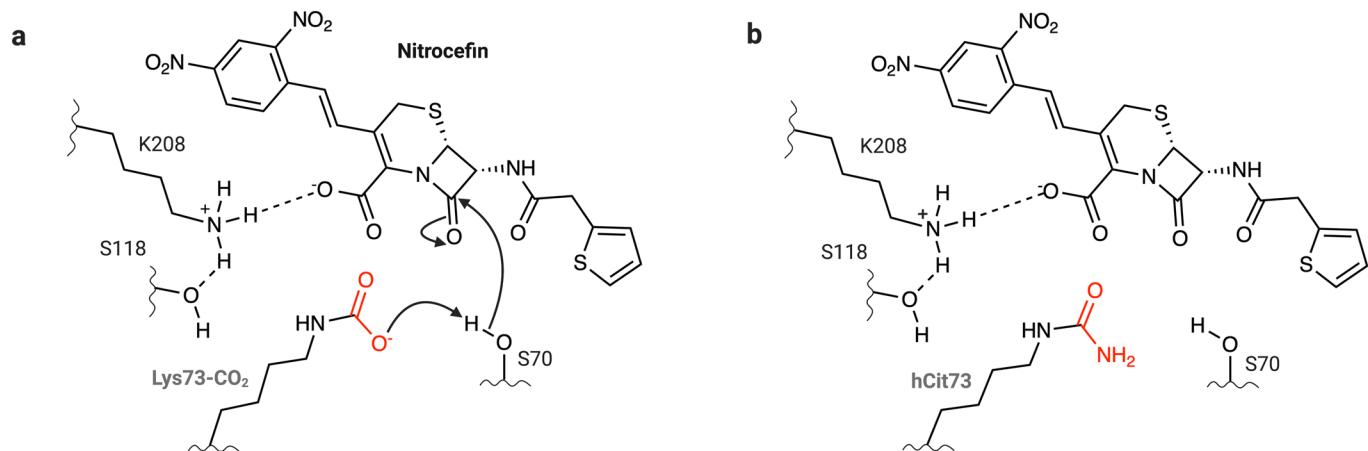
Extended data are available for this paper at <https://doi.org/10.1038/s41589-022-01043-1>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41589-022-01043-1>.

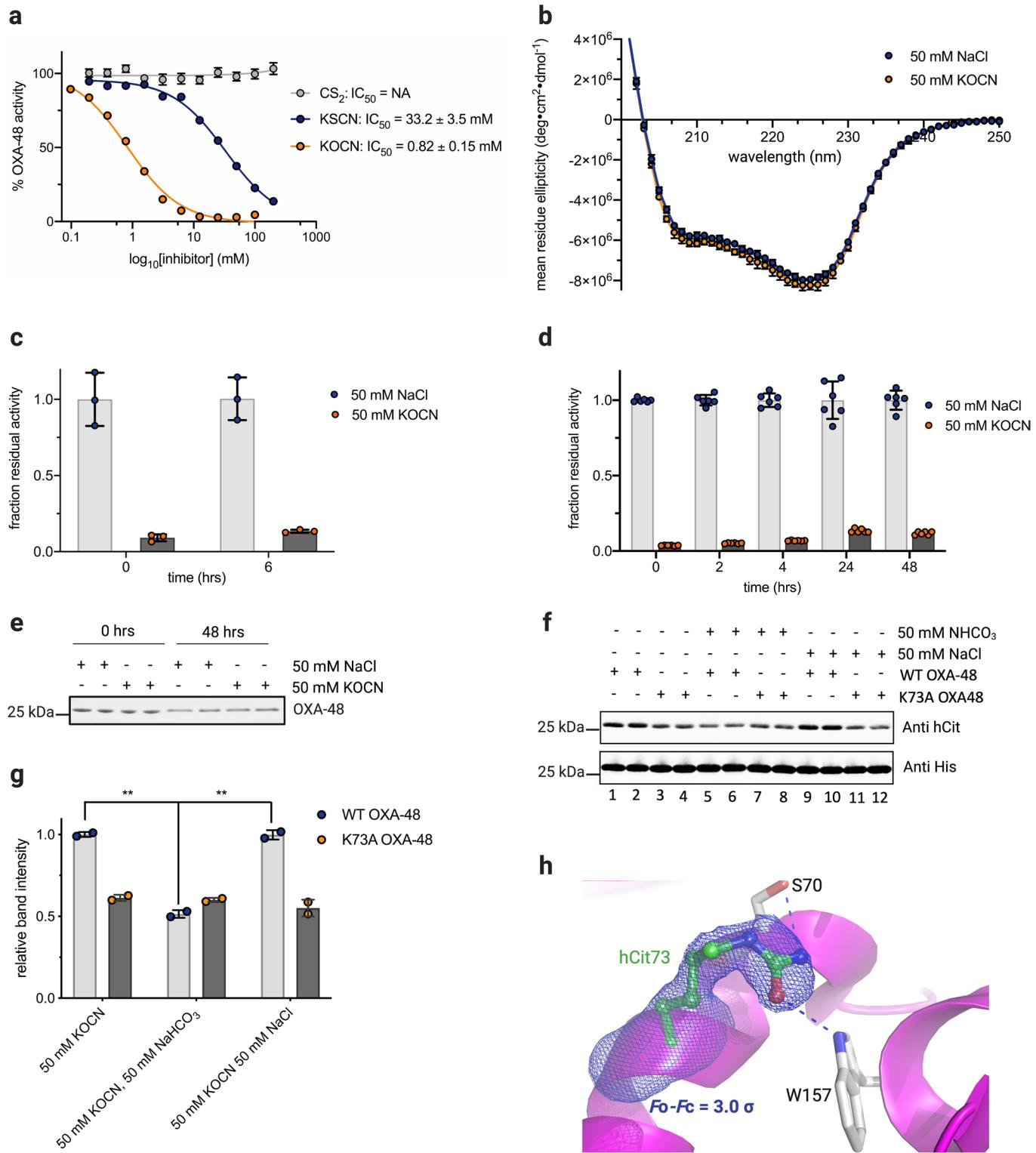
**Correspondence and requests for materials** should be addressed to David J. Voccadillo.

**Peer review information** *Nature Chemical Biology* thanks Cong-Zhao Zhou and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

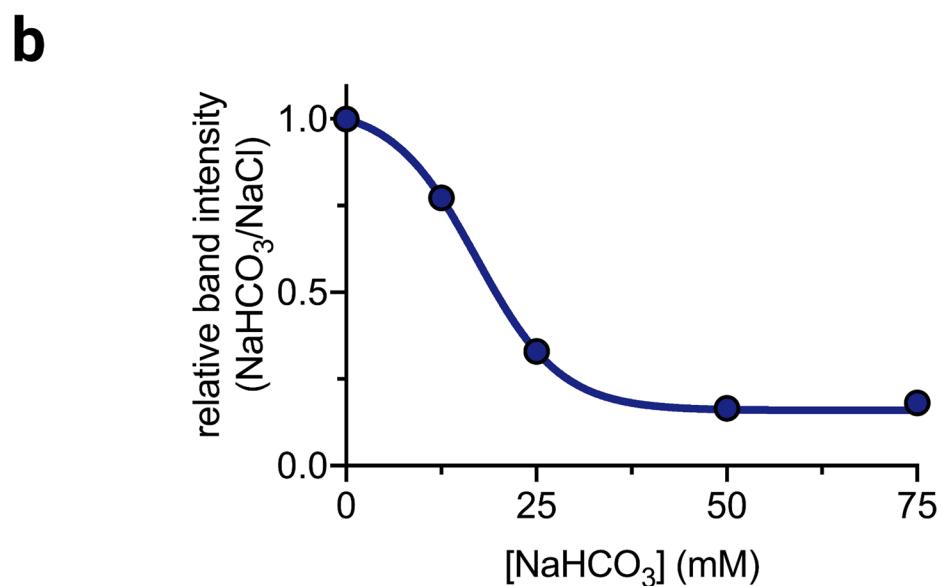
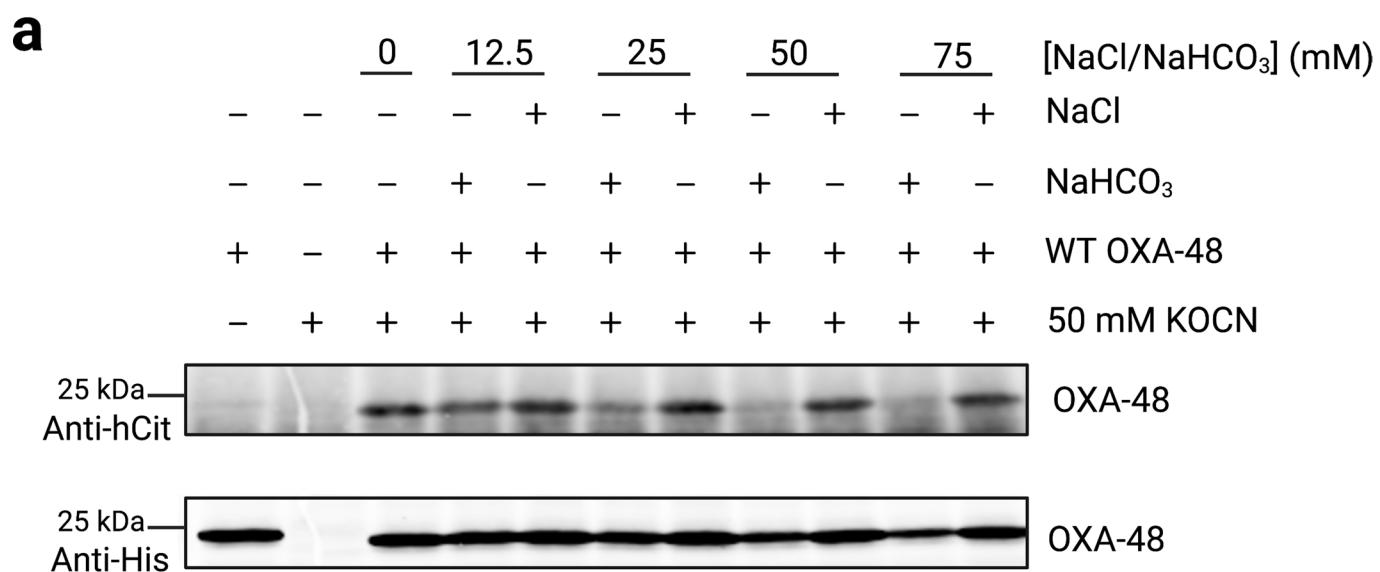


**Extended Data Fig. 1 | Homocitrullination of OXA-48 Lys73 blocks catalytic activity.** **a-b**, Mechanistic diagram of WT (Lys73-CO<sub>2</sub>) and homocitrullinated (hCit73) OXA-48 showing how hCit73 is unable to act as a general base and activate the nucleophilic serine.

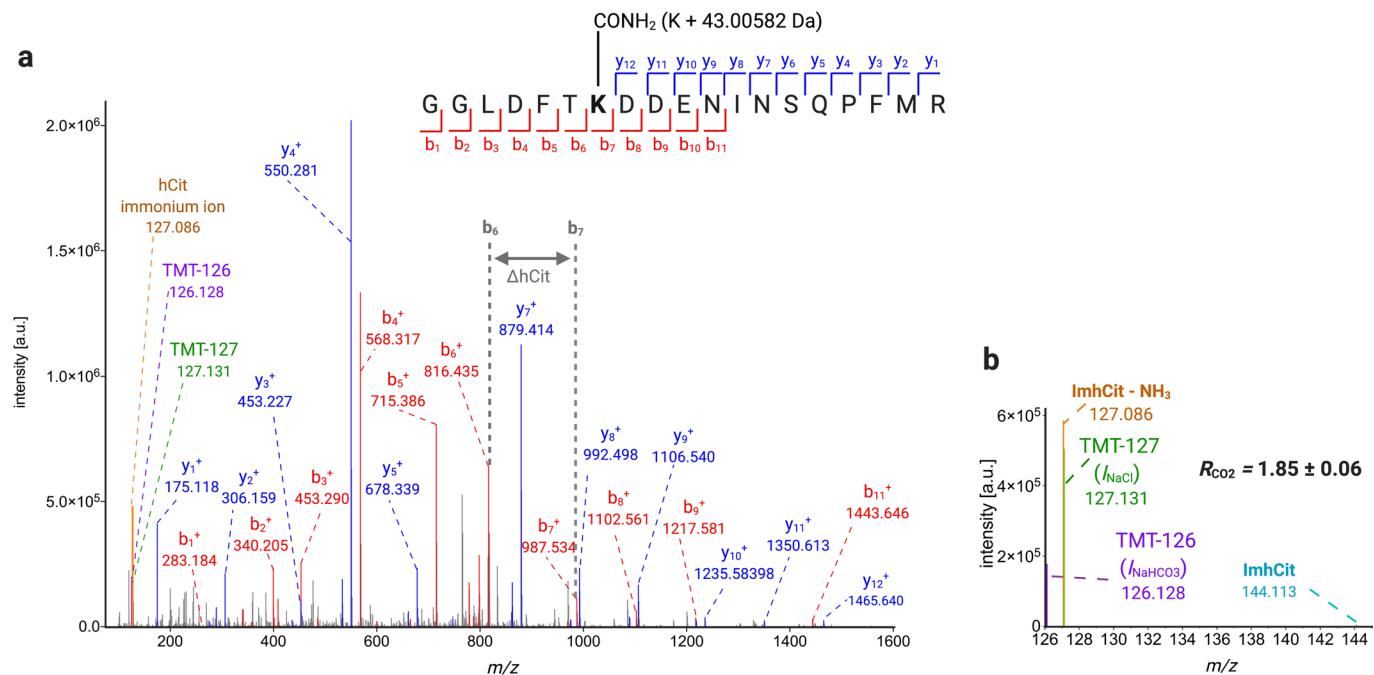


Extended Data Fig. 2 | See next page for caption.

**Extended Data Fig. 2 | In vitro OCNH-mediated homocitrullination of OXA-48, related to Fig. 1.** **a**, IC<sub>50</sub> experiments comparing OXA-48 inhibition by CO<sub>2</sub> analogues. Data represent mean values  $\pm$  SD from n=4 independent experiments. **b**, Far-UV CD spectra of OXA-48 pre-incubated in the presence of 50 mM NaCl or 50 mM KOCN. Data represent mean values  $\pm$  SD from n=4 independent experiments. **c**, Activity assays on samples from (b) taken immediately before (0 hrs) and after CD (6 hrs). OXA-48 remains inactivated throughout the CD experiment. Data represent means  $\pm$  SD from n=3 independent experiments. **d**, Jump-dilution kinetics experiments to detect time-dependant decarbamylation of OXA-48. Data represent mean values  $\pm$  SD from n=6 independent experiments. **e**, Anti-His immunoblot for OXA-48 samples from (d) incubated in the presence of either 50 mM NaCl or 50 mM KOCN at 0- and 48-hour timepoints. Samples were centrifuged to remove insoluble aggregates prior to immunoblotting. This immunoblot was performed once. **f**, Immunoblot competition assays for WT and K73A OXA-48. The image shown in f is a representative image from three independent experiments. **g**, Densitometry analysis of select bands from (f). The chart displays mean  $\pm$  SD from two experimental replicates with P values: two tailed student's t-test assuming unequal variance; \*\*P < 0.01. The indicated significant P values in (g) are as follows: (50 mM KOCN) vs. (50 mM KOCN, 50 mM NaHCO<sub>3</sub>) = 0.00346, (50 mM KOCN, 50 mM NaHCO<sub>3</sub>) vs. (50 mM KOCN, 50 mM NaCl) = 0.00337. **h**, Crystallographic evidence for hCit73 OXA-48. The hCit73 omit Fo-Fc electron density map contoured at 3.5  $\sigma$  is shown overlaid on the final refined coordinates. The protein backbone is displayed in magenta cartoon with key active site residues displayed as white sticks with heteroatoms coloured by type. The hCit73 residue is displayed as green sticks. Select hydrogen bonds are shown as blue dashes.



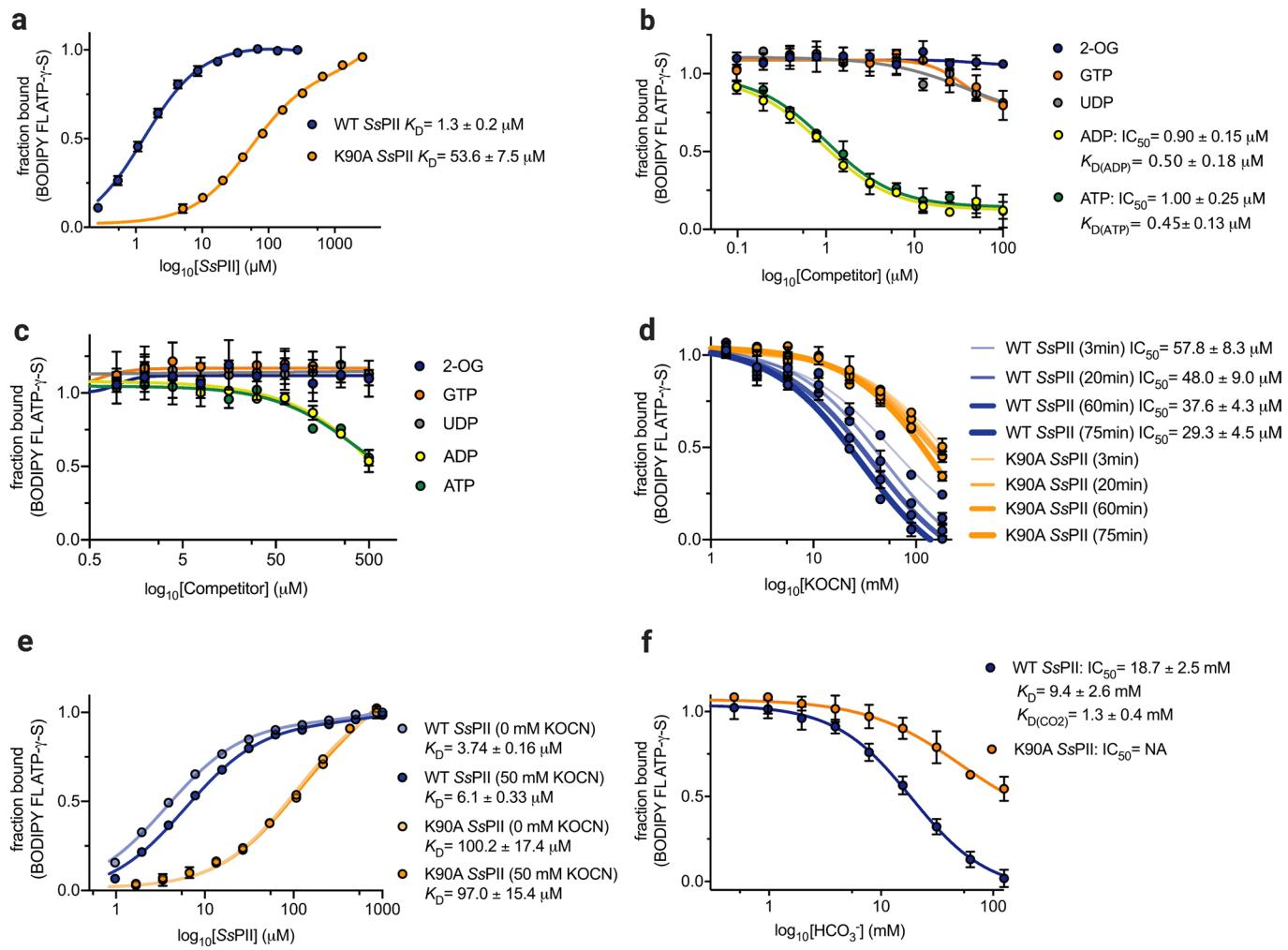
**Extended Data Fig. 3 | CO<sub>2</sub> competes with OCNH on OXA-48 in a concentration dependant manner, related to Fig. 2.** **a**, Immunoblot competition assay performed on purified OXA-48 with detection using a polyclonal anti-hCit antibody. **b**, Densitometry analysis of relative OXA-48 band intensities in a. Data points are normalized to the Anti-His control band and then again to the zero NaHCO<sub>3</sub> control. Data are fit to a three-parameter sigmoidal dose-response curve in GraphPad Prism.



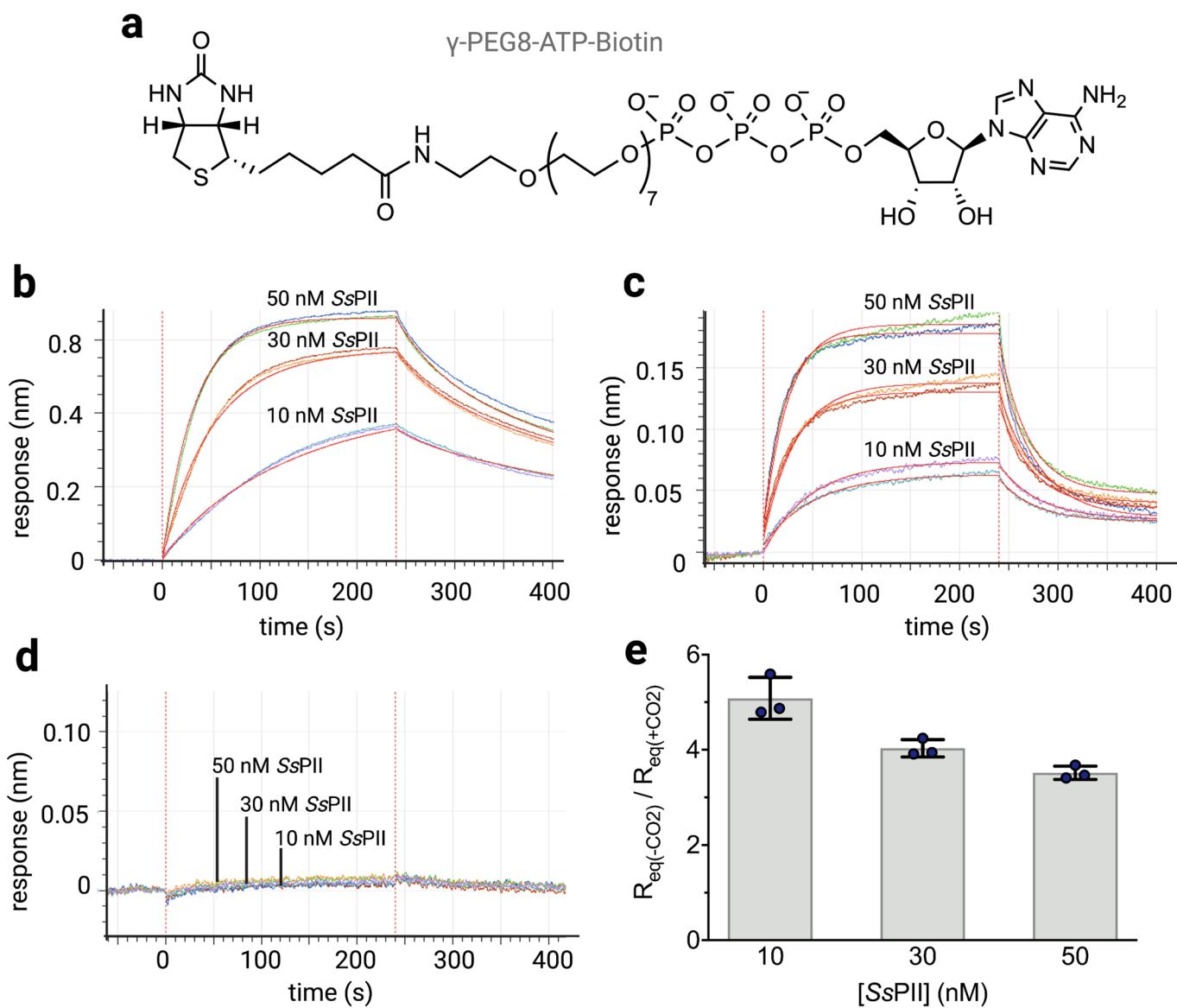
**Extended Data Fig. 4 | Lys-CO<sub>2</sub> modification of RuBisCO large subunit identified by proteome wide LysCarComp-MS performed on *Synechocystis* sp., related to Fig. 4.** **a**, LC-MS/MS spectra showing unambiguous evidence for hCit196 residue within TMT<sub>duplex</sub>-labelled RuBisCO large subunit peptide [TMT-GGLDFThCit(K<sub>196</sub> + 43.01 Da)DDENINSQPFMR, MH + [Da] = 2452.16]. **b**, Close-up of TMT<sub>duplex</sub> reporter ions and immonium ions derived from (a). The mean  $R_{\text{CO}_2}$  with SD from three independent experimental replicates is given.

		B-loop	K90	R101	C-loop
	<b><i>Synechocystis</i> sp.</b>	<b>mvvdklvsaartgeigdgkifispvdsvvrirtgekdteai</b>			<b>112</b>
	<i>Anabaena</i> sp.	mvvdkiiaaartgeigdg <b>kifispveqvirirtgeknteav</b>			112
	<i>Prochlorococcus marinus</i>	svidaiaeaktgeigdg <b>kifispidsvvrrirtgtdkeal</b>			112
cyanobacteria	<i>Synechococcus elongatus</i>	tvidkivaaartgeigdg <b>kifvspvdqtirirtgeknadai</b>			112
	<i>Trichodesmium erythraeum</i>	tvidkiinasrtgeigdg <b>kifvspvdkivrrirtgedndai</b>			112
	<i>Crocospaera watsonii</i>	mvidklrgaartgeigdg <b>kifvspvdqvirirtgesdleai</b>			112
	<i>Synechococcus</i> sp.	tvidkivaaarpgeigdg <b>kifvspvdqtirirtgeknadai</b>			112
microalgae	<i>Chlamydomonas reinhardtii</i>	avvrlvaasaytgeigdg <b>kifvhpvaevvrritaetgleae</b>			191
	<b><i>Arabidopsis thaliana</i></b>	<b>svintiiegartgeigdgkifvlpvsdvirrvrtgergekae</b>			<b>186</b>
	<i>Macleaya cordata</i>	evidklleeartgeigdg <b>kiflvpvsdiirrvrtgergekae</b>			97
plant	<i>Bacillus subtilis</i>	qvtetakrvlktgspgdg <b>kifvyeisntinirtdeegpeal</b>			116
	<b><i>Escherichia coli</i></b>	<b>tcvdttiirtaqtgkigdgkifvfdvarvirirtgeeddaai</b>			<b>112</b>
firmicute	<i>Marinomonas posidonica</i>	qvveaisgaantgkigdg <b>kifivnleqavrirtgetveav</b>			112
$\gamma$ -proteobac.		*	*****:	:	: : **

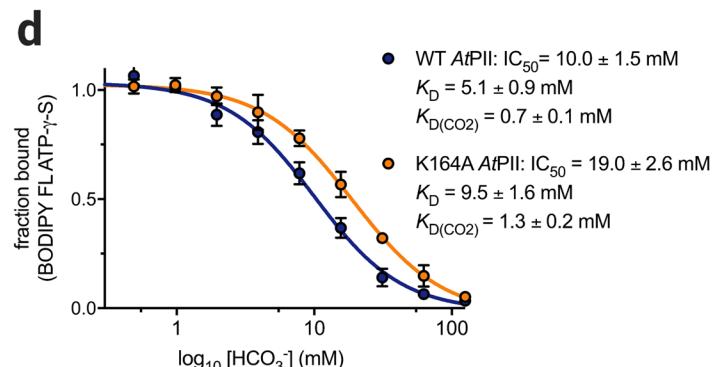
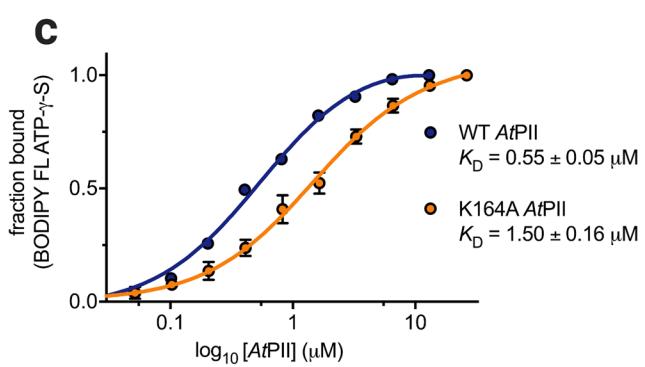
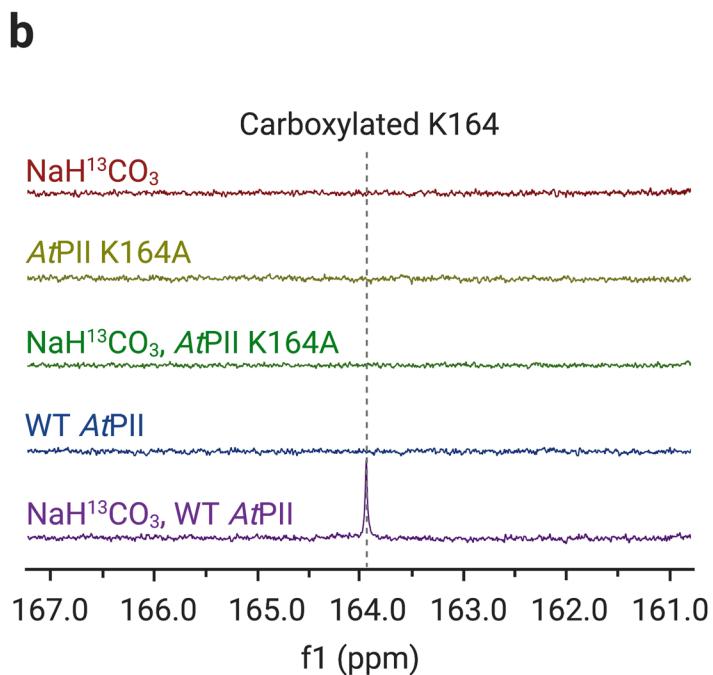
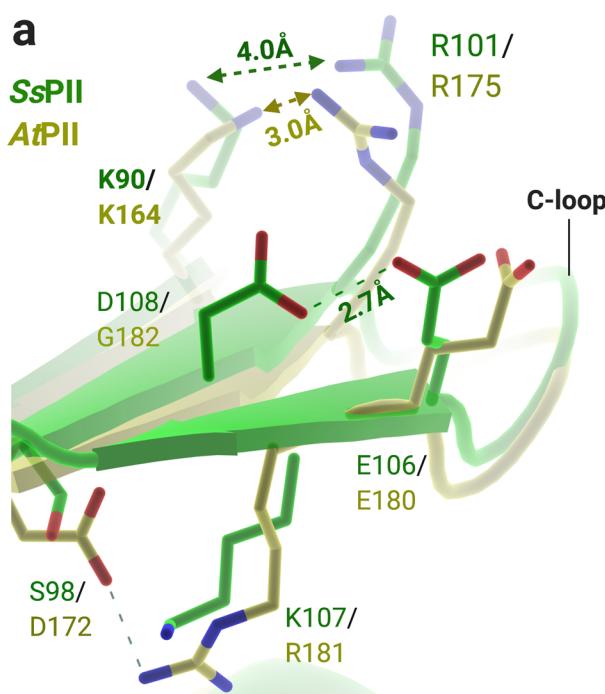
**Extended Data Fig. 5 |** Partial multiple sequence alignment for PII proteins from select species constructed using Clustal Omega<sup>41</sup>, related to Fig. 5.



**Extended Data Fig. 6 | CO<sub>2</sub> blocks BODIPY FL ATP- $\gamma$ -S binding to PII from *Synechocystis* sp., related to Fig. 5.** **a**, FP equilibrium dissociation binding of BODIPY FL ATP- $\gamma$ -S to WT and K90A SsPPII. **b–c**, FP equilibrium competition binding experiments for WT (b) and K90A SsPPII (c) in the presence of varying concentrations of small molecules (2-OG, GTP, UDP, ADP, ATP). **d**, Time-dependant effect of KOCN on FP equilibrium probe binding to WT and K90A SsPPII. **e**, FP equilibrium binding of BODIPY FL ATP- $\gamma$ -S following pre-incubation of SsPPII in the presence and absence of KOCN followed by jump dilution and desalting to remove residual KOCN. In (a–e), all data is presented as mean values  $\pm$  SD from  $n=2$  independent experiments. **f**, FP equilibrium competition binding experiments for WT and K90A SsPPII in the presence of varying concentrations of HCO<sub>3</sub><sup>−</sup>. Data in (f) are presented as mean values  $\pm$  SD from  $n=4$  independent experiments. In (b) and (f),  $K_D$ 's for competitor compounds were calculated from IC<sub>50</sub> values using a standard FP equilibrium binding formula as previously described<sup>59</sup>.  $K_{D(\text{CO}_2)}$  was determined from  $K_D$  using standard equations for carbonate equilibria as previously described<sup>60</sup>.



**Extended Data Fig. 7 |** Biolayer interferometry confirms that  $CO_2$  blocks ATP binding to SsPPII, related to Fig. 5. **a**, Chemical structure of the biotin-conjugated ATP analogue used in BLI. **b-c**, Representative independent duplicate BLI association/dissociation curves for WT SsPPII run in the absence (**b**) and presence (**c**) of 50 mM  $HCO_3^-$ . **d**, Representative BLI association/dissociation curves for K90A SsPPII run in the absence of  $HCO_3^-$ . SsPPII protein concentrations and curves are exactly as shown in **b-c**. **e**, Equilibrium response ratio for SsPPII binding to the ATP-biotin probe in the presence and absence of 50 mM  $HCO_3^-$ .  $R_{eq}$  values correspond to the response value at the 240 s timepoint in **b** and **c**. Data are presented as mean values  $\pm$  SD from three independent experiments.



**Extended Data Fig. 8 | Lys164-CO<sub>2</sub> modification blocks ATP binding to PII from *Arabidopsis thaliana*.** **a**, Active site overlay of apo SsPPII and apo *Arabidopsis thaliana* PII (AtPPI, 53.6% sequence identity, PDB IDs: 1UL3 and 2O66). The SsPPII and AtPPI protein chains are displayed as green and yellow cartoons with select residues shown as sticks with atoms coloured by type. **b**, <sup>13</sup>C NMR using purified WT and K164A AtPPI in the presence and absence of 50 mM NaH<sup>13</sup>CO<sub>3</sub>. **c**, FP equilibrium dissociation binding of BODIPY FL ATP-γ-S to WT and K164A AtPPI. **d**, FP equilibrium competition binding experiments for WT and K164A AtPPI in the presence of varying concentrations of HCO<sub>3</sub><sup>-</sup>. In c-d, data are presented as mean values  $\pm$  SD from n=2 independent experiments. In d, K<sub>D</sub>'s for competitor compounds were calculated from IC<sub>50</sub> values using a standard FP equilibrium binding formula as previously described<sup>59</sup>. K<sub>D(CO2)</sub> was determined from K<sub>D</sub> using standard equations for carbonate equilibria as previously described<sup>60</sup>.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give P values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Gen5 (version 3.0) software for enzyme kinetics and fluorescence polarization. Fusion Tune (version 3.3) software for mass spectrometry. Li-COR Image Studio (version 3.0) for immunoblotting. MxDc data collection software (version 2020.8.21) for X-ray crystallography data collection. Chirascan Pro-Data SX software (version 2.5.0) for circular dichroism data collection. Satorius Octet systems data analysis software (version 12) for bilayer interferometry.

Data analysis

Proteome Discoverer (version 1.4.0.228) and XCalibur software (3.0.63) for processing of raw proteomics data. Prism 9.0 (GraphPad) for all curve fitting and statistical tests. ImageJ (version 1.51) for densitometry analysis of immunoblots. Coot (version 0.9.3), Phenix (version 1.18.2), Phaser (version 1.20.1), IMOSFLM (version 7.4.0), CCP4 (version 7.1.018), and PyMOL (version 2.3.5) for processing and analysis of X-ray crystallography data. Thermo protein deconvolution software package (version 3.0) for analysis of intact mass spectrometry data. Topspin (version 3.6) and MestReNova (version 14.1) for NMR data analysis. Excel (version 16.54) for analysis of enzyme kinetics data, mass spectrometry, fluorescence polarization, and bilayer interferometry.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All raw proteomics data were deposited to the ProteomeXchange Consortium via the PRIDE partner repository, with the dataset identifier (PXD031976). The hCitr-OXA-48 crystal structure has been deposited in the PDB with accession code 7LXG. The PDB codes for previously determined X-ray crystallographic structures used in the preparation of this manuscript are as follows: OXA-48 (4S2P), apo Synechocystis sp. PII (1UL3), ATP/2-OG bound Synechococcus elongatus PII (2XUL), and Arabidopsis thaliana PII (2O66). The ncbi accession codes used in generating recombinant proteins for this study are as follows: Klebsiella pneumoniae OXA-48 (AP182700), Synechocystis sp. PII (CAA66127.1), Arabidopsis thaliana PII (QAP00825.1). UniProt reference proteomes used in LysCarComp-MS analysis are as follows: E. coli K12 (UP000000625), Synechocystis sp. 6803 (UP000001425), Klebsiella pneumoniae OXA-48 (Q6XEC0). All other data needed to evaluate the conclusions in the manuscript are available within the main text or supplementary materials.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculations were performed. Sample sizes were selected according to standards in the field (for enzyme kinetics, all biochemical assays, and mass spectrometry).
Data exclusions	No data were excluded from analysis.
Replication	All attempts at replication were successful. All experiments were replicated at least twice with the exception of large scale LysCarComp-MS proteomics experiments which were performed only once for practical reasons but were conducted using at least three independent experimental replicates.
Randomization	Not applicable
Blinding	Not applicable

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

### Antibodies used

Goat pan-specific anti-homocitrulline, MyBioSource, polyclonal, Cat # MBS390103, dilution = 1:2,000  
 Mouse monoclonal anti-6X-His tag, Invitrogen, clone: HIS.H8, Cat # MA1-21315, dilution = 1:2,000  
 IRDye 800CW donkey anti-goat IgG secondary antibody, polyclonal, LI-COR, Cat # 926-32214, dilution = 1:20,000  
 IRDye 680LT goat anti-mouse IgG secondary antibody, polyclonal, LI-COR, Cat # 926-68020, dilution = 1:20,000

## Validation

Goat pan-specific anti-homocitrulline: In this study, we show that this antibody detects homocitrullinated proteins and displays comparatively low or negligible binding to their non-homocitrullinated controls in immunoblots. Wherever appropriate, when immunoblotting for homocitrullinated proteins, we include the non-homocitrullinated controls for direct comparison on the same blot. Furthermore, this antibody has been shown to detect various homocitrullinated proteins and not their unmodified or citrullinated counterparts [Autoimmunity Reviews 17 (2018) 136-141].

Mouse monoclonal anti-6X-His tag: In this study, we verify specific detection of 6X-His tagged OXA-48 in immunoblots by including controls comprised of the identical lysates lacking the histidine tagged protein. Furthermore, the manufacturer validated this antibody by conducting relative expression tests by immunoblot analysis using poly-histidine tagged proteins as well as a panel of appropriate negative control proteins. These tests clearly show that this antibody binds specifically to the antigen stated (6X-His). Please see the suppliers advanced verification data: [https://www.thermofisher.com/order/genome-database/dataSheetPdf?producttype=antibody&productsubtype=antibody\\_primary&productId=MA1-21315&version=137](https://www.thermofisher.com/order/genome-database/dataSheetPdf?producttype=antibody&productsubtype=antibody_primary&productId=MA1-21315&version=137)

LI-COR secondary antibodies were used according to the manufacturers well-established protocols.