Carbon dioxide is essential for life. It is at the beginning of every life process as a substrate of photosynthesis. It is at the end of every life process as the product of post-mortem decay. Therefore, it is not surprising that this gas regulates such diverse processes as cellular chemical reactions, transport, maintenance of the cellular environment, and behaviour. Carbon dioxide is a strategically important research target relevant to crop responses to environmental change, insect vector-borne disease and public health. However, we know little of carbon dioxide’s direct interactions with the cell. The carbamate post-translational modification, mediated by the nucleophilic attack by carbon dioxide on N-terminal α-amino groups or the lysine ε-amino groups, is one mechanism by which carbon dioxide might alter protein function to form part of a sensing and signalling mechanism. We detail known protein carbamates, including the history of their discovery. Further, we describe recent studies on new techniques to isolate this problematic post-translational modification.

**Keywords:** carbon dioxide, post-translational modification, carbamate, haemoglobin, rubisco, connexin, ubiquitin

**INTRODUCTION**

Since its discovery in gas exhaled from the lung in 1757 (West, 2004, 2014) carbon dioxide (CO₂) has been recognised as a critical component of biological processes throughout the biosphere. Its contribution to the essential physiological processes of metabolism, photosynthesis, chemosensing, and cellular homeostasis (Cummins et al., 2020) has meant that organisms across the three domains of life had evolved mechanisms to sense, transport, and respond to CO₂ (Cummins et al., 2014). Although a great deal of knowledge exists about the physiological processes where CO₂ is produced or consumed, less is known about the direct mechanisms of CO₂ interactions with biomolecules.

One way in which CO₂ has been shown to interact with protein directly is through carbamylation of neutral N-terminal α-amino or lysine ε-amino groups (Figure 1). This carbamate post-translational modification (PTM) is critical to regulating oxygen-binding in haemoglobin and the activation of the CO₂-fixing enzyme RuBisCO. It has been suggested that protein carbamylation could form the basis of a widespread mechanism for biological regulation (Morrow et al., 1974; Lorimer, 1983). Computational studies predict that carbamates may be found in more than 1.3% of large proteins (Jimenez-Morales et al., 2014).

In this review, we bring together contextual examples of carbamylation and explore recent computational and experimental approaches with the potential to uncover the distribution of protein carbamylation within proteomes.

**CO₂ and Haemoglobin**

The linked processes of ventilation and metabolism are essential to survival in higher animals. Following ventilation, gas transfer enables oxygen to be provided promptly to the cells and their...
mitochondria, which is used in the final process of aerobic cellular respiration (oxidative phosphorylation). Oxidative phosphorylation coordinates with the tricarboxylic acid (TCA) cycle to form adenosine triphosphate (ATP) and the metabolites required by the organism for survival, producing CO₂ as a waste product (Martínez-Reyes and Chandel, 2020). CO₂ exits the cells to the bloodstream and is transported to the lungs for excretion, thus contributing to pH homeostasis (Cummins et al., 2020).

The link between ventilation and metabolism was first made in 1777 by Antoine Lavoisier with the observation that “Eminently respirable air [oxygen] that enters the lung, leaves it in the form of chalky aeroform acids [carbon dioxide]… in almost equal volume. . .” (West, 2014). At this time, it was presumed that the process of metabolism (“slow combustion”) was performed within the lung (West, 2014). By the mid to late 1800s it was clear that O₂ was transported via the blood to tissues (where metabolism occurred) by the formation of a loose, dissociable interaction with haemoglobin (oxyhaemoglobin) (Barcroft and Hill, 1910; Saha et al., 2014), and CO₂ was returned to the lung by similar means (Severinghaus and Bradley, 1958; Giegé, 2013).

In 1904 Bohr, Hasselbalch and Krogh measured haemoglobin oxygenation in canine blood and described the sigmoidal (rather than hyperbolic) nature of the oxyhaemoglobin dissociation curve (Bohr et al., 1904). This experiment demonstrated that increasing pCO₂ resulted in a lowered affinity of haemoglobin for O₂ (known as the Bohr effect). Conversely, Christensen et al. (1914) described that increasing pO₂ resulted in a decreased affinity of haemoglobin for CO₂ (known as the Haldane effect). The Bohr and Haldane effects were reversible, and observed in various mammalian systems (Christensen et al., 1914; West, 2019). Combined, the Bohr-Haldane effect results in haemoglobin being an efficient O₂ transporter from the lungs to tissues and CO₂ from the tissues to the lungs (Eaton et al., 1999). The nature of the oxyhaemoglobin dissociation curve led to the hypothesis that multiple O₂ binding sites on haemoglobin acted cooperatively (Barcroft and Hill, 1910; Hill, 1913).

Haemoglobins belong to a large family of proteins with members distributed across all three domains of life. The first structures (myoglobin and equine haemoglobin) were determined by X-ray crystallography in the 1950s (Perutz et al., 1960, 1964). Human adult haemoglobin is a tetramer consisting of two α and two β subunits similar in structure and size. The α and β subunits are formed of seven and eight helices, respectively (A–H), joined by non-helical segments. Each subunit binds a heme group consisting of a porphyrin ring that coordinate a Fe²⁺ ion (capable of binding to O₂) by four nitrogen atoms at its centre. The oxygenated and deoxygenated haemoglobin quaternary structures differ. The gap between two polypeptide chains in the haemoglobin molecule narrows when O₂ binds to the Fe²⁺ (Paoli et al., 1996; Park et al., 2006; Perutz et al., 1960, 1964). The binding of the first O₂ to the haemoglobin subunit enhances the ability of subsequent O₂ molecules to bind to the remaining subunits (Adair, 1925; Pauling, 1935). This knowledge was used to develop the Monod Wyman and Changeux “two-state concerted” model for allostery where deoxygenated haemoglobin exists in a tense (T) state (with relatively low O₂ affinity). When Fe²⁺ binds O₂, there is a movement of Fe²⁺ into the heme plane, which triggers a transition to the relaxed (R) state. In this R state, the remaining binding sites are more exposed and have an increased O₂ affinity (Monod and Jacob, 1961; Bringas et al., 2017).

In the 1930s, CO₂ was shown to combine rapidly and reversibly with haemoglobin to form carbaminohaemoglobin (Ferguson and Roughton, 1934; Stadie, 1935; Stadie and O’Brien, 1936). It was suggested that carbaminohaemoglobin formation was possible at multiple amino sites on haemoglobin (Stadie and O’Brien, 1936). Confirmation of carbamate formation at the N-terminal of the valines of each of the four human deoxyhaemoglobin chains was performed using cyanate-based blocking, which inhibited the uptake of CO₂ by haemoglobin (Kilmartin and Rossi-Bernardi, 1969) and confirmed that CO₂-binding was O₂-linked (at constant pCO₂ and pH, deoxyhemoglobin forms more carbamino- CO₂ than haemoglobin-O₂). The CO₂ binding site was demonstrated to occur at the Val-1 site and linked to the O₂-binding state of the β-chain (Matthew et al., 1977). Perrella et al., 1975 found that the adduct formed on the β chain is more prominent than the α chain. Physiologically, the reaction between the a-amino group and CO₂ stabilises the protein’s deoxygenated form. Morrow et al., 1974 suggested that carbamino formation may be a general and functionally important phenomenon throughout biology and not limited to haemoglobin (Morrow et al., 1974).

CO₂ and Ribulose 1,5-Bisphosphate Carboxylase-Oxygenase (RuBisCO)

Thought to have emerged about three billion years ago during the Archaean aeon and before the Great Oxygenation Events (Kacar et al., 2017; Banda et al., 2020), the RuBisCO family of proteins is
one of the most abundant on the planet (Ellis, 1979). Initially discovered in the 1940s by Wildman and Bonner (Wildman, 2002), these enzymes are represented across the three domains of life and grouped as structurally distinct operational forms based on protein sequence and secondary and tertiary structure (Duff et al., 2000; Schneider et al., 1990, 1992). Forms I, II and III catalyse the carboxylation (and oxygenation) of ribulose 1,5-bisphosphate, while form IV contains RuBisCO-like proteins (RLP), which although sequentially and structurally similar, perform distinct biological functions (Watson et al., 1999; Tabita et al., 2008; Kono et al., 2017).

RuBisCO I is found in plants, cyanobacteria, algae, and some proteobacteria and is responsible for the vast majority of atmospheric CO2 fixation through the Calvin-Benson-Bassham (CBB) reductive pentose phosphate pathway (Andersson and Backlund, 2008; the dark reactions of photosynthesis). Most commonly, Form I is a hexadecamer of eight large and eight small subunits (L8S8). There are four Form I subtypes: A and B in cyanobacteria, eukaryotic algae, and higher plants; C and D are found in nongreen algae and phototrophic bacteria (Stec, 2012). It has been demonstrated that although operational RuBisCO forms differ in overall structure, all share a similar catalytic subunit dimer (Duff et al., 2000; Schneider et al., 1990; Tabita et al., 2007, 2008), formed from the interaction of the large subunits (the C-terminus of the first interacting with the N-terminus of the second to form two active sites with residues from both). It is thought that RuBisCO remains in the open state when receiving substrates and eliminating products. Still, when catalytic events occur, it is in a closed state, essentially sequestering the active site from the bulk solvent (Duff et al., 2000).

RuBisCO requires the catalytic site to be activated (Tabita et al., 2007, 2008) by sequential carbamylation of a specific lysine side chain followed by Mg2+ binding (Stec, 2012) (Figure 2). Once activated, RuBisCO’s role within the CBB is in catalysing the initial stage of carboxylation of ribulose 1,5-bisphosphate (RuBP) and cleavage to form two molecules of 3-phosphoglycerate (3-PGA) (one of which is used in RuBP regeneration, the other diverted to sustain biosynthesis of sugars and other high energy compounds). However, once activated, RuBisCO also catalyses a competitive oxygenation reaction, leading to the formation of one molecule of 3-PGA and one molecule of 2-phosphoglycerolate (2-PG), which inhibits central carbon metabolism (Tcherkez, 2016). RuBisCO is a relatively inefficient enzyme with a low turnover rate (Bathellier et al., 2018). The turnover rate is further reduced by competition from the oxygenation reaction with the potential to reduce carbon fixation by up to 50% (Andersson and Taylor, 2003; Andersson, 2007). Therefore, optimisation of the carboxylation action of RuBisCO has been explored to improve crop efficiency and climate-resilient photosynthesis (Andersson and Backlund, 2008; Erb and Zarzycki, 2018; Välegård et al., 2018).

Identification of the RuBisCO catalytic site and its activation were, in most part, uncovered through kinetic and physical investigations in the 1960s and 70s. It became apparent that the sequential addition of CO2 to the catalytic site and its activation was, therefore, largely forgotten. However, as observed by Professor Lorimer, there was no available method for trapping these carbamates on protein to enable their identification. The carbamate modification was, therefore, largely forgotten.

FIGURE 2 | The RuBisCO active site showing the K201 carbamate. The figure shows the active site of Arabidopsis thaliana RuBisCO (PDB: 5IU0) with two large subunit chains shown in green and cyan and bound transition-state analogue 2-carboxy-D-arabinitol-1,5-bisphosphate (2-CAB), D203 and E204 coordinate Mg2+ co-factor (grey sphere) in the active site. The carbamate forms on K201 and is stabilised through interactions that include H294 and T173.
DISCOVERY OF OTHER CARBAMATES

Carbamates have been observed on specific lysine side chains in the crystal structures of several proteins, e.g. class D β-lactamase (Golemi et al., 2001), urease (Yamaguchi and Hausinger, 1997), alanine racemase (Morollo et al., 1999), and transcarboxylase 5 S (Hall et al., 2004). Lysine carbamylation can be critical to protein/enzyme activity. Examples include having a direct role in catalytic activity (class D β-lactamase, (Golemi et al., 2001), being a co-catalytic determinant, which bridges metal ions, (e.g. urease, (Yamaguchi and Hausinger, 1997), or in promoting (through hydrogen bonding) active site rearrangements for protein function (e.g. alanine racemase; (Morollo et al., 1999). Figure 3 summarises the carbamate functional role in haemoglobin, RuBisCO and other proteins discussed below.

Class D β-lactamase

β-lactam antibiotics are essential in treating bacterial infection, inhibiting bacterial growth by acylating an active-site serine in essential penicillin-binding proteins and preventing the cross-linking of peptide chains to form peptidoglycan. β-Lactamases enable resistance to β-lactam antibiotics. These enzymes hydrolyse compounds containing a β-lactam ring and are classified based on sequence motifs and mechanism of hydrolysis (Class A–D). Class D OXA enzymes are the most diverse and least well-understood β-lactamases (Maveyraud et al., 2000; Tooke et al., 2019). The class is widely dispersed, found in gram-negative and gram-positive bacterial species and mobilised horizontally predominantly by plasmids and integrons (Bush, 2018; Tooke et al., 2019). In 2000, the first crystal structures of OXA-10 of Pseudomonas aeruginosa were published (Maveyraud et al., 2000; Strynadka et al., 2000). Maveyraud et al., 2000 discovered that OXA-10 is a dimeric beta-lactamase. However, the overall topology of OXA-10 class D β-lactamase is like class A, although amino acid sequence identity is low. Significant structural differences were found in the active site of OXA-10 compared to corresponding regions in class A (and C β-lactamases). In its native state, OXA-10 Lys-70 is carbamylated. It was suggested that this carbamylation offered a possible relationship between enzyme activation by CO₂ and anion inhibition (Maveyraud et al., 2000). Golemi et al., 2001 used activity assays and fluorescence-based approaches to confirm that Lys-70 carbamylation is reversible and that OXA-10 β-lactamase depends on Lys-70 for enzyme acylation and deacylation steps in catalysis. This finding was supported by the inability of Lys-70 mutants to support deacylation. It was suggested that the hydrophobicity of the OXA-10 active site lowered the Lys-70 pKₐ favouring carbamylation (Golemi et al., 2001; Sun, 2003).

Urease

Urease plays an essential role in nitrogen metabolism in archaea, bacteria, fungi, plants and invertebrates (Dixon et al., 1975; Krajewska, 2009; Rutherford, 2014). It catalyses urea hydrolysis to form carbonic acid (H₂CO₃) and ammonia (NH₃) and is important in human disease, plant metabolism, and agricultural ammonia emissions (Krajewska, 2009; Svane et al., 2020). In bacteria, ureases’ active site and primary structure are well studied in organisms such as Proteus mirabilis, Helicobacter pylori, and Klebsiella aerogenes (Jabri et al., 1995; Jabri and Karplus, 1996). The mechanism of urease activation is CO₂ dependent in vitro (Pearson et al., 1998). This dependence has been explained by the active site containing two Ni²⁺ ions, which are bridged (stabilised) by the carboxyl group of a carbamylated lysine residue (Lys-217) that is essential for urease activation (Jabri and Karplus, 1996; I.-S. Park and Hausinger, 1995).

COMPUTATIONAL APPROACHES FOR THE IDENTIFICATION OF CARBAMATE SITES

To enhance protein carbamate discovery, Jimenez-Morales et al. (2014) developed a model to predict the uncarboxylated and carboxylated status of lysine residues in proteins. They initially used a training set of 251 proteins (identified by X-ray crystallography) which contained at least one protein subunit with a carbamylated lysine residue (KCX) to investigate the characteristics of the carbamate microenvironment when compared to uncarboxylated (LYS) sites. They observed that the crucial feature of the KCX microenvironment “was the large numbers of packed atoms, water molecules and ions found in proximity to the KCX site chain.” Additionally, all KCX sites were buried (inaccessible from the surface), with residues converging structurally at the KCX site dispersed along the protein’s primary sequence (so no sequence motif was associated with lysine carbamylation). The KCX residue was often in contact with positively charged ions with mono-nuclear or bi-nuclear interactions with divalent ions (Zn²⁺, Mg²⁺, Ca²⁺, Fe²⁺, Ni²⁺, Mn²⁺). His and Asp residues were present in all analyses of KCX containing metal-binding sites. No more than one KCX site was observed on any single protein chain. The
RuBisCO carbamylated lysine (used in the training set) is typical of a structure where the carbamate binds Mg<sup>2+</sup> coordinated by Asp and Glue residues (Figure 2).

This information was used to develop a naïve Bayesian model (to predict potential KCX and LYS sites (Predictor of Lysine Carbonylation: PreLysCarb)). The 251-protein data set was used for training/testing (along with sub-sets with redundancy reduction implemented at 40 and 90% sequence identity). They carried out “leave-one-out cross-validation tests” on the three data sets. At 90% sequence identity PreLysCar correctly classified 54/62 KCX sites (87% sensitivity) and 4255/4259 LYS sites (99.7% specificity). Investigating false-positive rates in high-resolution protein structures, the model indicated that 11 to 19/575 proteins were incorrectly predicted to have a KCX residue (false positive rate of between 1.9 and 3.3%). When PreLysCar was applied to a subset of solved protein structures from the PDB (structures greater than 200 residues, solved by X-ray crystallography, containing 14,261 protein chains after 90% redundancy removal), it predicted that at least 1.3% of proteins with more than 200 amino acids in the PDB could potentially be subject to spontaneous lysine carboxylation. As the model has been trained using previously identified stable carbamates, which are predominantly buried, it may only represent a subset of possible KCX sites (Linthwaite et al., 2020).

What is clear from this analysis is that no consensus sequence enables easy CO2-binding site prediction as has proven so successful for phosphorylation sites, for example. This observation is borne out by analysing the growing numbers of experimentally observed carbamates (see following sections) with no clear primary consensus sequence. Future machine learning approaches as experimental data sets increase in size will, therefore, be essential as predictive tools.

**MASS SPECTROMETRY APPROACHES FOR THE IDENTIFICATION OF CARBAMATE SITES**

Mass spectrometry-based proteomics is a powerful tool for discovering and exploring PTMs. Although carbamylation of neutral N-terminal α-amino or lysine ε-amino groups may be a “general and functionally important phenomenon throughout biology” (Morrow et al., 1974) and “form the basis for a widespread mechanism of biological regulation” (Lorimer, 1983), the transient and readily reversible nature of this PTM all but renders impossible the use of mass spectrometry-based proteomics workflows (even when soft ionisation techniques are utilised; Terrier and Douglas, 2010). This ready reversibility limits our discovery of the distribution and contribution of carbamylamion to biological functions.

Linthwaite et al., 2018 developed a chemical mechanism of covalently “trapping” carbamates under physiologically relevant conditions to overcome this limitation. In this technique, the triethylxonium ion (TEO; a crystalline salt soluble under aqueous conditions) stabilises pre-formed carbamate modifications (formed through incubation of proteins or cellular lysates with CO2) by transferring an ethyl group from the oxonium ion to the negatively charged carbamate, forming a covalent bond which can withstand downstream protease digestion and enables CO2-binding protein identification by HPLC-ESI-MS/MS analysis. This approach has been validated for non-buried exchangeable carbamate binding sites (where presumably the carbamate is labile and the bound CO2 exchangeable into the bulk solvent) using individual amino acids and single proteins (previously known to be carbamylated; Linthwaite et al., 2018). This approach has recently contributed to studies of CO2 interaction with connexin 26 and ubiquitin (Sections 6.1 and 6.2) and shows potential in carbamate discovery through screening cell lysates (Section 6.3).

**CO2 and Connexin 26**

Connexins (Cx) have an essential role in intercellular and extracellular communication and, therefore, homeostasis in multicellular organisms. These transmembrane proteins form hexameric connexons or hemichannels (HCs) in the plasma membrane, allowing low molecular weight molecules to transfer across the plasma membrane (Meigh et al., 2014). In closely apposed membranes, two HCs can dock together to form homomeric or heteromeric gap junctions (GJs), which mediate intercellular communication through the transfer of low molecular weight molecules (<1.5 kD), e.g. ions, metabolites and second messengers, between cells (Giepmans, 2004; Meigh et al., 2014; Willebrords et al., 2016). GJ intercellular communication can be regulated by pH (closed by acidification), transmembrane voltage (opened by voltage potential greater than −20 mV) and calcium concentration (opened by removal of extracellular Ca<sup>2+</sup>). PTMs such as S-nitrosylation, sumoylation and phosphorylation can directly regulate GJ opening (Willebrords et al., 2016).

Connexin 26 (Cx26) HCs and GJs are found in cells throughout the body, including in the cochlear epithelial network of the ear, keratinocytes of the skin, alveolar epithelium of the lungs, epithelial cells of the GI tract, and chemosensory areas of the brain (Cohen-Salmon et al., 2002; Huckstepp et al., 2010a; Huckstepp et al., 2010b; Meigh et al., 2014; Willebrords et al., 2016; Srinivas et al., 2018; van de Wiel et al., 2020). Mutations in the gene encoding Cx26 (GJB2) are relatively common (frequency of carriers ~2–4% of the human population) and are linked to nine non-syndromic and syndromic deafness disorders (SDD; Cook et al., 2019; Srinivas et al., 2018). SDDs are associated with visual impairment and dermatological abnormalities and, in some instances, Keratitis-Icthyosis-Deafness syndrome is underpinned by Cx26-A88V and Cx26-G45E missense mutations (Meigh et al., 2014).

Huckstepp et al., 2010b and Wenker et al., 2012 suggested that Cx26 has a role in mediating the central CO2-dependent drive to breathe with HCs enabling ATP release from the medulla oblongata in the absence of extracellular acidification. Exploring the potential molecular mechanism, Huckstepp et al., 2010a found that increasing pCO2 at fixed pH opens Cx26 HCs (and HCs of two related beta-connexins, Cx30 and Cx32). Meigh et al., 2013 compared the amino acid sequences of Cx26, Cx30 and Cx32 with Cx31 (a connexin that has no
sensitivity to pCO$_2$ and hypothesised a carbamylation motif present in Cx26, 30, 32 that was absent from Cx31. Using the existing Cx26 crystal structure (Maeda et al., 2009), the authors noted that the carbamylation “motif” contained K125 at the end of a subunit’s alpha helix where K125 is oriented towards R104 on a neighbouring subunit of the hexamer. Therefore, they hypothesised that if K125 was carbamylated, it could feasibly form a salt bridge (“carbamate bridge”) with R104, linking the subunits, and preventing hemichannel closure. By inserting the identified carbamylation “motif” into Cx31, they demonstrated that it was sufficient to form a CO$_2$-sensitive hemichannel. If the K125 residue was substituted for an amino acid that could not be carbamylated, CO$_2$ sensitivity was lost. Using Cx26, they confirmed that K125 and R104 were essential for forming the carbamate bridge. Meigh et al., 2015 used a similar mutation-based approach to explore the potential for alternative mechanisms of bridge formation between residues 125 and 104 in adjacent hexamer subunits. They found they could convert the CO$_2$-sensitive hemichannels to a NO/NO$^+$-sensitive hemichannel using Cx26-K125C with Cx26-R104 or a redox-sensitive hemichannel using the combination of Cx26-K125C and Cx26-R104C, thus suggesting that distinct mechanisms of bridging involving residues 125 and 104 on adjacent hexamer subunits of Cx26 was possible (Meigh et al., 2015).

Using a combination of carbamate trapping, high-resolution cryo-EM and classification of particles, Brotherton et al., 2020 proposed that under physiologically relevant high pCO$_2$ conditions (90 mmHg), a carbamate was formed on Lys-125, and additionally at two other positions Lys-108 and Lys-122, but not under physiologically relevant low pCO$_2$ conditions (20 mmHg). They suggested that the shared environment within the cytoplasmic TM2 and TM3 regions of the Cx26 mobile loop favoured CO$_2$ modification. Classification of particles indicated that the positions of TM2 and TM3 could influence the conformation of the N-terminal helix and found that under high pCO$_2$ conditions, the N-terminus was more defined than under low pCO$_2$ conditions. The authors hypothesised that gating is mediated by the movement of the N-terminal helix and its ability to plug the channel under physiologically relevant conditions (Brotherton et al., 2020).

Recent work by Nijjar et al., 2020 has shown in contrast to the opening of HCs at high physiological CO$_2$ that intact Cx26 GJs connecting HeLa cells are closed when pCO$_2$ is increased from 35 to 55 mmHg. This closing effect is dependent on the same residues (K125 and R104) involved in the CO$_2$-dependant opening of Cx26 hemichannels. Further, the action is also directly attributed to a change in pCO$_2$ rather than to changes in pH. They explained the contrasting action based on the free energy difference between CO$_2$-bound and unbound states for HC and GJ (Nijjar et al., 2020). Specifically, it was energetically more favourable for HCs to bind CO$_2$ in the open state and less energetically unfavourable to close. In contrast, it was energetically more favourable for the gap junction to bind CO$_2$ in the closed state and then energetically unfavourable to open. The docking of two connexons as an HC provided a close interaction that constrained the conformation of hemichannels docked in a gap junction. Nijjar et al., 2020 proposed the need for more information about the structures of Cx26 as free HCs and as components of GJs to investigate conformational differences and provide a greater understanding of the differential modulation of hemichannels and gap junctions by CO$_2$. What is clear from these studies is that CO$_2$ can have a signalling role via carbamate formation.

**CO$_2$ and Ubiquitin**

Ubiquitin (Ub) is a highly conserved protein found in all eukaryotic cells. It plays a crucial role in regulating protein activity and degradation through Ub protein covalent conjugation to a lysine side chain on target proteins (Komander, 2009). Ubiquitylation conjugation occurs through the sequential activity of Ub-activating enzymes (E1), Ub-conjugating enzymes (E2), and Ub ligases (E3) (Yau and R ape, 2016; Heap et al., 2017) and can be reversed by deubiquitylating enzymes (DUBs). Ub can be conjugated into poly-Ub chains through the N-terminus (M1) or at any of the seven conserved lysine residues (K6, K11, K27, K29, K33, K48, and K63) forming single, mixed or branched Ub-chains, resulting in a diverse range of Ub-chain structures (Swatek and Komander, 2016; Yau and R ape, 2016; Heap et al., 2017). Distinct Ub-chains are involved in different biological functions (Heap et al., 2017). The most abundant and well-studied of these are K48 single linked Ub-chains which target proteins for proteasomal degradation (Heap et al., 2017), and K63 single linked Ub-chains involved in cell signalling, trafficking and lysosomal degradation (Ohtake et al., 2015; Swatek and Komander, 2016; Michel et al., 2017). Roles of other less abundant single and mixed linked Ub-chains include regulation of enzymatic activity (K33), cell cycle regulation (K11), inflammation and immune response (M1, K63/M1, K48) and post-replication repair (K6 and K33) (Komander, 2009; Michel et al., 2017).

Ub activity can be further regulated by additional PTMs, e.g. acetylation, phosphorylation and SUMOylation (Koyano et al., 2014; Morimoto and Shirakawa, 2016; Lamoliatte et al., 2017; Lacoursiere et al., 2020). For example, all Ub lysine residues (except K29) are acetylated under differing cellular conditions (Choudhary et al., 2009; Elia et al., 2015; Lacoursiere et al., 2020). Ohtake et al., 2015 demonstrated that when endogenous Ub is acetylated at K6 and K48, this does not affect the ability of Ub to conjugate with substrate protein but inhibits the elongation of K11, K48 and K63-linked Ub-chains by several E2 enzymes (by neutralising the lysine residue positive charge involved in non-covalent interaction of Ub with specific E2s). This acetylation results in the accumulation of monoubiquitylated substrates in the cell. Using a SILAC based approach, Ohtake et al., 2015 confirmed that acetylation of K6 and K48 was linked to the enrichment of chromosome or chromatin related factors, including histone H2B, and that monoubiquitylated H2B was stabilised by the expression of acetylated Ub (Ohtake et al., 2015; Morimoto and Shirakawa, 2016; Michel et al., 2017). It has recently been suggested that Ub PTMs add an extra layer of complexity to the ‘ubiquitin code’ and that this extends beyond currently identified PTMs (Ohtake and Tsuchiya, 2016).
Linthwaite et al., 2021 recently explored the potential modification of human Ub by CO2. They demonstrated that Ub K33 and K48 could be carbamylated under physiological conditions by CO2 (25 mM CO2/HCO3 at pH 7.4), using the “trapping” method (Figure 4) outlined in Linthwaite et al., 2018. Ub carbamylation was confirmed independently by 13C-NMR, identifying K6 and K63 as additional carbamylation sites. Using in vitro conjugation assays, Linthwaite et al. (2021) found that di-Ub formation at K48 was significantly decreased when the CO2 concentration was increased from the normal physiological reference range (1.8–2.3 mM dissolved CO2) to the hypercapnic range (>2.3 mM CO2).

The effect of elevated CO2 on Ub-dependant processes was explored and focused on the regulation of nuclear factor kβ (NF-kB) (Iwai, 2014). NF-kB is a transcription factor central to inflammation, innate and acquired immune responses, nervous system function, and cell survival (Iwai, 2014; Albensi, 2019; Song and Li, 2021). Under elevated pCO2, NF-kB mediated transcription is suppressed (Cummins et al., 2010). NF-kB is inactive in resting cells, bound in the cytoplasm to the inhibitor of NF-KBs (IκB) proteins, blocking its transport to the nucleus. There are two known pathways of NF-kB activation, the canonical and non-canonical pathways. In the canonical pathway, the IκB kinase complex (IKK1, IKK2 and NF-kB essential modulator (NEM)) is activated upon interaction with, e.g., inflammatory cytokines or Toll-like receptor (TLR) ligands. This activation results in specific phosphorylation of IκB Ser residues within IκBα. Phosphorylated IκBα is recognised by the Ub ligase complex SCFSR in order to generate K48-linked poly-Ub chains and are subsequently degraded by the proteasome. NF-kB is released from IκBα and translocates to the nucleus, where it binds to the DNA consensus sequence of a target gene. TNF-receptor family proteins activate the non-canonical pathway. This activation results in the stabilisation of NIK and its phosphorylation of IKKα. This phosphorylated IKKα induces further phosphorylation of NF-kB2/p100, which forms a complex with RelB. Once phosphorylated NF-kB/p100 is ubiquitinated by SCFSR E3s generating K48-linked poly-Ub chains. This ubiquitinylation is followed by the partial degradation to p52 by the proteasome. The resulting RelB/p52 heterodimer is translocated to the nucleus (Iwai, 2014).

When cells carrying an NF-kB dependent GFP reporter were exposed to increasing TNFα concentrations under normal (5% (v/v) CO2 in air) and elevated CO2 conditions (10% (v/v) CO2 in air), NF-kB dependent GFP reporter activity was significantly decreased under elevated CO2 (Linthwaite et al., 2021). Transfection of the same cells with plasmids encoding wild-type Ub, mutant K48R Ub, mutant K63R Ub, or an empty vector was used to address the hypothesis that overexpression of K48R Ub would alter the relative response of the NF-kB pathway to elevated CO2. Cells transfected with an empty vector, Wt Ub or K63R Ub, showed an unaltered CO2 response. In contrast, when cells were transfected with K48R Ub, the effect of increased CO2 on the inhibition of the NF-kB reporter was ablated, suggesting that Ub K48 may be the target for CO2 in the NF-kB-dependent transcriptional response to hypercapnia in human cells.

**CONCLUSION**

We can draw some clear conclusions from the studies presented. First, carbamylation is a non-enzymatic PTM that can also be
readily reversible. Second, the biological consequences of the carbamate PTM can vary discretely with changing CO₂. Third, the carbamate PTM is more widespread among proteomes than suspected. Several future challenges arise from these conclusions. First, the extent of carbamylation in a proteome is unknown. How do we identify those that might have a functional role in CO₂ sensing and signalling instead of forming through standard physicochemical mechanisms and being functionally neutral? Third, is carbamylation the sole mechanism for CO₂ detection? If not, what other means exist, and how do we identify them? It is clear; there is plenty more to this review.

**REFERENCES**

Adair, G. (1925). A Critical Study of the Direct Method of Measuring the Osmotic Pressure of Hemoglobin. *Proc. R. Soc. Lond. Ser. B* 98 (692), 523. doi:10.1098/rspb.1925.0054

Akoyunoglou, G., Argiroudi-Akoyunoglou, J.-H., and Methenitou, H. (1967). Studies on the Active Site of the Enzyme Ribulose-5-phosphate Carboxylase. *Biochim. Biophys. Acta (Bba) - Enzymol.* 132 (2), 481–491. doi:10.1016/0005-2744(67)90166-0

Albenis, B. C. (2019). What Is Nuclear Factor Kappa B (NF-Kb) Doing in and to the Mitochondrion? *Front. Cell Dev. Biol.* 7, 154. doi:10.3389/fcell.2019.00154

Andersson, I., and Taylor, T. C. (2003). Structural Framework for Catalysis and Regulation in Ribulose-1,5-Bisphosphate Carboxylase/oxygenase. *Arch. Biochem. Biophys.* 414 (2), 130–140. doi:10.1016/S0003-9861(03)00164-4

Banda, D. M., Pereira, J. H., Liu, A. K., Orr, D. J., Hammel, M., He, C., et al. (2020). Bringas, M., Petruk, A. A., Estrin, D. A., Capece, L., and Martí, M. A. (2017). Tertiary and Quaternary Structural Basis of Oxygen Affinity in Human Hemoglobin as Revealed by Multiscale Simulations. *BioRxiv* doi:10.1101/847845

Blake and Cann Carbamate Post-Translational Modification

**AUTHOR CONTRIBUTIONS**

LB and MC conceived, designed and wrote the review manuscript.

**FUNDING**

This review was supported by the Biotechnology and Biological Sciences Research Council grant no. BB/S015132/1.

**ACKNOWLEDGMENTS**

We thank Tim Blower for assistance with analysis of the RuBisCO structure.

Cohen-Salmon, M., Ott, T., Michel, V., Hardelin, J.-P., Perfettini, I., Eybalin, M., et al. (2002). Targeted Ablation of Connexin26 in the Inner Ear Epithelial Gap Junction Network Causes Hearing Impairment and Cell Death. *Carr. Biol.* 12 (13), 1106–1111. doi:10.1016/S0960-9822(02)00904-1

Cook, J., de Wolf, E., and Dale, N. (2019). Cx26 Keratitis Ichthyosis Deafness Syndrome Mutations Trigger Alternative Splicing of Cx26 to Prevent Expression and Cause Toxicity In Vitro. *R. Soc. Open Sci.* 6, 191128(8). doi:10.1098/rsos.191128

Cummins, E. P., Oliver, K. M., Lenihan, C. R., Fitzpatrick, S. F., Bruning, U., Scholz, C. C., et al. (2010). NF-κB Links CO2 Sensing to Innate Immunity and Inflammation in Mammalian Cells. *J. Immunol.* 185, 4439–4445. doi:10.4049/jimmunol.1001701

Cummins, E. P., Selfridge, A. C., Sporn, P. H., Snajder, J. I., and Taylor, C. T. (2014). Carbon Dioxide-Sensing in Organisms and its Implications for Human Disease. *Cell. Mol. Life Sci.* 71 (5), 831–845. doi:10.1007/s00018-013-1470-6

Cummins, E. P., Strowitzki, M. J., and Taylor, C. T. (2020). Mechanisms and Consequences of Oxygen and Carbon Dioxide Sensing in Mammals. *Physiol. Rev.* 100 (1), 463–488. doi:10.1152/physrev.00003.2019

Dixon, N. E., Gazzola, C., Blakeley, R. L., and Zernner, B. (1975). Jack Bean Urease (EC 3.5.1.5). Metalloenzyme. Simple Biological Role for Nickel. *J. Am. Chem. Soc.* 97 (14), 4131–4133. doi:10.1021/ja00747a045

Duff, A. P., Andrews, T. J., and Curmi, P. M. (2000). The Transition between the Open and Closed States of Rubisco Is Triggered by the Inter-phosphate Distance of the Bound Bisphosphatate. *J. Mol. Biol.* 298, 903–916. doi:10.1006/jmbi.2000.3724

Eaton, W. A., Henry, E. R., Eric, J., and Mozzarelli, A. (1999). Is Cooperative Oxygen Binding by Hemoglobin Really Understood? *Nat. Struct. Biol.* 6 (4), 351–358. doi:10.1038/7586

Elia, A. E. H., Boardman, A. P., Wang, D. C., Huttlin, E. L., Everley, R. A., Dephoure, N., et al. (2015). Quantitative Proteomic Atlas of Ubiquitination and Acetylation in the DNA Damage Response. *Mol. Cell 59* (5), 867–881. doi:10.1016/j.molcel.2015.05.006

Ellis, R. J. (1979). The Most Abundant Protein in the World. *Trends Biochem. Sci.* 4 (11), 241–244. doi:10.1016/0968-0004(79)90212-3

Erb, T. J., and Zarrzycki, J. (2018). A Short History of RubisCO: the Rise and Fall (?) of Nature’s Predominant CO₂ Fixing Enzyme. *Curr. Opin. Biotechnol.* 49, 100–107. doi:10.1016/j.cpbio.2017.07.017

Ferguson, J. K. W., and Roughton, F. J. W. (1934). The Direct Chemical Estimation of Carbamino Compounds of CO2 with Haemoglobin. *J. Physiol.* 83 (1), 68–86. doi:10.1113/jphysiol.1934.sp033212

Giegié, R. (2013). A Historical Perspective on Protein Crystallization from 1840 to the Present Day. *Fems J.* 280 (24), 6456–6497. doi:10.1111/fibs.12580

Giemps, B. (2004). Gap junctions and Connexin-Interacting Proteins. *Cardiovasc. Res.* 62 (2), 233–245. doi:10.1016/j.cardiores.2003.12.009

Golemi, D., Maveysra, L., Vakulenko, S., Samama, J.-P., and Mohashery, S. (2001). Critical Involvement of a Carbamylated Lysine in Catalytic Function of
Yau, R., and Rape, M. (2016). The Increasing Complexity of the Ubiquitin Code. Nat. Cel Biol 18, 579. doi:10.1038/ncb3358
Yuan, Y., Tam, M. F., Simplaceanu, V., and Ho, C. (2015). New Look at Hemoglobin Allostery. Chem. Rev. 115 (4), 1702–1724. doi:10.1021/cr500495x

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Blake and Cann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.