Development of a method for the investigation of protein carbamylation by CO2

JANUS, JOANNA,MARY

Development of a method for the investigation of protein carbamylation by CO_2

Joanna Janus

Supervisor: Dr Martin Cann

Master of Research thesis

School of Biological and Biomedical Sciences

Durham University

2016

Development of a method for the investigation of protein carbamylation by CO_2

Joanna Janus

Abstract

Protein carbamylation occurs in biological systems when the nitrogen atom of a neutral amine group makes a nucleophilic attack upon the carbon atom of a CO_2 molecule. This results in the CO_2 becoming bonded to the amine group as a carbamate. Carbamates play important structural and functional roles in several proteins, including haemoglobin, Rubisco and β-lactamases. Despite this and the ubiquity of CO_2, relatively few proteins have been shown to be carbamylated. This is partly due to the labile nature of the carbamate bond, meaning it dissociates readily and can be hard to investigate.

A method was developed to ethylate carbamate modifications on proteins using the reagent triethyloxonium tetrafluoroborate (TEO). The ethylation trapped the carbamate modification so it could be analysed by mass spectrometry. It was first confirmed that TEO could trap a carbamate on an amino acid, then the trapping method was adapted for use on proteins. It was shown that the TEO trapping method was able to stabilise a carbamate on haemoglobin which could be detected using MALDI-TOF and electrospray mass spectrometry. The method was also used to test proteins with unknown carbamate forming abilities. Recombinant proteins comprising the catalytic domains of mammalian transmembrane adenylyl cyclases and the related Rv1625c adenylyl cyclase from *Mycobacterium tuberculosis* were expressed and used in the carbamate trapping experiments. The $Rv1625c_{204-443}$ protein was revealed to potentially possess a carbamylated Lys-296 residue. Due to inconclusive results an alternative method was used to further explore this result. $^{14}CO_2$ was supplied to the wild type $Rv1625c_{204-443}$ recombinant protein and an $Rv1625c_{204-443}$ protein with Lys-296 mutated to Ala, before potential carbamates were trapped with TEO. It was concluded that it was likely that the Lys-296 residue of $Rv1625c_{204-443}$ did not form a carbamate, although further work is needed to confirm this result.

A method has therefore been developed to trap carbamate modifications on proteins and has been used to explore whether adenylyl cyclases undergo carbamylation by CO_2. The method requires further optimisation and some results were inconclusive. However the findings from this study will contribute towards the development of a method to screen many proteins in a cell lysate for carbamate modifications simultaneously, using proteomic techniques. This will be a valuable tool in enabling us to better understand the extent and roles of protein carbamylation in physiology.

Acknowledgements

I would like to thank my supervisors Dr Martin Cann, School of Biological and Biomedical Sciences and Dr David Hodgson, Department of Chemistry for all their support, enthusiasm and patience throughout the project.

Thanks also to Victoria Linthwaite who I worked with closely and was a constant source of support and information throughout, giving up a lot of time to teach me various techniques. Also thanks to Hamish Pegg for guiding me through protein expression and purification, and to Dr Phil Townsend and Chris Dixon for their advice in this area.

I would also like to acknowledge Dr Adrian Brown for all his help and advice with mass spectrometry analysis.

Declaration

I confirm that no part of the material in this thesis has previously been submitted by me for a degree in this or in any other University. Material in section 4.5 was generated through joint work with Victoria Linthwaite; however all experiments in these sections were carried out independently to validate the results. In all other cases material from the work of others has been acknowledged.

Table of Contents

List of Figures and Tables

Table 3.5.1: Summary of approaches taken to optimise the trapping experiment 51

List of Abbreviations

7C1	1st catalytic domain of type 7 adenylyl cyclase
2C2	2nd catalytic domain of type 2 adenylyl cyclase
AC	adenylyl cyclase
Arg	arginine
Asp	aspartate
ATP	adenosine triphosphate
BSA	bovine serum albumin
CA	carbonic anhydrase
cAMP	3',5'-cyclic adenosine monophosphate
CHCA	α-Cyano-4-hydroxycinnamic acid
Ci	inorganic carbon
Cx26	connexin 26
ESI	electrospray ionisation
GC	guanylyl cyclase
Hb	haemoglobin
His	histidine
IPTG	isopropyl-β-D-thiogalactopyranoside
LB	luria broth
LC	liquid chromatography
Lys	lysine
MALDI	matrix assisted laser desorption ionisation
MS	mass spectrometry
mw	molecular weight
Ni-NTA	nickel-nitrilotriacetic acid
PAGE	polyacrylamide gel electrophoresis
pCO_2	partial pressure of carbon dioxide
Rubisco	ribulose-1,5-bisphosphate carboxylase/oxygenase
sAC	soluble adenylyl cyclase
SDS	sodium dodecyl sulphate
Ser	serine
Thr	threonine
TEO	triethyloxonium tetrafluoroborate
TOF	time of flight
Tris	tris(hydroxymethyl)aminomethane
Val	valine

1 Introduction

1.1 Sensing and responding to CO_2

Carbon dioxide (CO_2) is a fundamental component of the both the external and internal environments of living organisms. CO_2 constitutes just 0.04% of the earth's atmospheric gases, and due to its high solubility is found at similar levels in the planets oceans and other large bodies of water (Zeebe, 2012). Plants and other photosynthetic organisms such as cyanobacteria and algae use CO_2 as a substrate for photosynthesis, and so their growth and survival are dependent upon a sufficient amount of CO_2 present in the atmosphere. Many of these photosynthetic organisms also use carbon concentrating mechanisms to increase the local concentration of carbon dioxide around the site of the photosynthetic enzyme Rubisco (Moroney et al., 2013). Aquatic organisms are also sensitive to environmental partial pressures of CO_2, as CO_2 dissolved in bodies of water is capable of forming carbonic acid (Figure 1.1.1). This weak acid can then partly dissociate to form protons and bicarbonate ions, lowering the pH of the water which may impact the physiology of the organisms within it. For example high CO_2 levels have been found to reduce levels of reproductivity and growth in several species of marine and freshwater fish, including juvenile white sturgeon and perch (Portner et al., 2004)

Animals, plants, fungi and bacteria also produce CO_2 as a direct by-product of aerobic respiration or indirectly from the oxidation of lactic acid in anaerobic respiration. This can lead to the creation of localised concentrations of CO_2 within organisms. CO_2 produced

$$CO_2 + H_2O \rightleftharpoons H_2CO_3 \rightleftharpoons HCO_3^- + H^+$$

Figure 1.1.1: Products of CO_2 in solution

CO_2 may combine with water in a reversible reaction to form carbonic acid, which can in turn reversibly dissociate to form bicarbonate and a proton. CO_2 may therefore induce functional effects in organisms via pH, bicarbonate or by molecular CO_2.

through respiration can dissolve in internal fluids within organisms such as blood and cytoplasm, with the resulting acidosis causing potential disruption and damage to cellular systems unless it is regulated and removed. Some organisms are able to make use of the CO_2 emitted by other respiring organisms as an indicator of a potential food source. Parasites including mosquitoes and ticks can use the higher CO_2 concentrations surrounding their respiring hosts as a sensory cue to guide them towards their host (Chaisson and Hallem, 2012), while in other species CO_2 elicits an avoidance response (Bretscher et al., 2011, Jones et al., 2007)

Therefore the ability to detect levels of CO_2 in both their internal and external environments is critical to allow biological organisms to respond accordingly, and there are several mechanisms by which they may do this. Gaseous exchange of CO_2 between an organism and its environment is a direct way of regulating internal CO_2 and oxygen levels. In most animals specialised ventilation mechanisms are required to enable gaseous exchange to take place efficiently, especially if an organism is undergoing higher levels of respiration and so needs to expel greater volumes of CO_2 than at its normal resting rate. A key way by which animals can regulate ventilation and gaseous exchange in response to CO_2 is through the detection of pH levels (Nattie, 1999). As already mentioned dissolved CO_2 will form an acid and produce H^+ (Figure 1.1.1). Therefore CO_2 from respiring tissue will produce protons, which are transported from the cells of the tissue into the blood. In mammals chemoreceptors located in the carotid body respond to the increased concentration of H^+ by stimulating the carotid sinus nerve which relays the impulses to the brainstem, which in turn can increase the amount of ventilation to expel the excess CO_2 and increase the flow of oxygen to the respiring tissues (Nurse and Piskuric, 2013).

While much research has centred on the control of CO_2 responses via pH, there is also evidence of molecular CO_2 being sensed either directly or through the production of bicarbonate ions (Figure 1.1.1). It has been shown that mice can be trained to respond to levels of CO_2 as low as 0.06%, with over 90% accuracy (Hu et al., 2007). This ability to detect CO_2 was shown to occur via activation of a subset of olfactory neurons that express guanylyl cyclase D (GCD) (Hu et al., 2007). The response was shown to be independent of intracellular and extracellular pH but required the presence of carbonic anhydrase II (CAII) activity (Hu et al., 2007). As carbonic anhydrase catalyses the hydration of CO_2 into H+ and HCO_3^- ions (Hewett-Emmett and Tashian, 1996) it was hypothesised that either of these ions could be causing the CO_2 response, although in this study this was not explored further. However as the authors had shown that changing the pH when CO_2 was absent showed no activation of the olfactory neurons, it seems likely that the HCO_3^- ions were causing the response.

The activation of olfactory receptors by CO_2 has also been shown to produce avoidance responses to CO_2 in *Drosophila* and *C. elegans* and to inhibit egg laying in *C. elegans* (Bretscher et al., 2011, Jones et al., 2007, Fenk and de Bono, 2015, Sun et al., 2009). However the molecular mechanisms underpinning these responses are still unclear. Often they appear to be independent of external pH, however it is much more difficult to monitor the internal pH changes within a cell and so it cannot be concluded that pH is not having any effect. Additionally in the cases where bicarbonate and/or molecular CO_2 are found to

be inducing responses it is often hard to distinguish which of these molecules is causing the effect. In the case of the mouse GCD olfactory neurons discussed above it was concluded that bicarbonate ions were likely to induce the response to CO_2 due to a dependence on expression of carbonic anhydrase that hydrates CO_2 into HCO_3^- and H^+ (Hu et al., 2007). However more recently it was shown that the related BAG neurons in *C. elegans,* which also express a membrane bound guanylyl cyclase, were activated by molecular CO_2 and not bicarbonate, although they could also be activated by acid (Smith et al., 2013). Membrane bound guanylyl cyclase 9 but not soluble guanylyl cyclase expression was required for the CO_2 response to occur, and this response was not dependent on the activity of carbonic anhydrase. This suggests that the guanylyl cyclases are acting as a direct receptor for CO_2 rather than bicarbonate ions, which would be more likely to require carbonic anhydrase expression (Smith et al., 2013). In other related studies no attempt is made to distinguish whether bicarbonate or CO_2 is inducing the response, despite papers often implying that bicarbonate is the causative agent, as it is used as the source of CO_2 in many studies (Sun et al., 2009, Hallem et al., 2011)

Therefore it is clear from these examples that if we are to be able to better understand the mechanisms behind cellular responses to CO_2, it is necessary to determine precisely how CO_2 and/or its derivatives are initiating a signalling response. One way in which molecular CO_2 is known to be able to bind to and alter the properties of proteins is via the formation of a carbamate modification. The role of carbamates as a mechanism for CO_2 detection and signalling will be reviewed in the following sections.

1.2 Carbamate bond formation and properties

Carbamate bonds form in biological systems when the nitrogen atom of a neutral amine makes a nucleophilic attack upon the carbon atom of a carbon dioxide molecule (Figure 1.2.1) (Lorimer, 1983), leading to the carbon dioxide molecule becoming attached to the

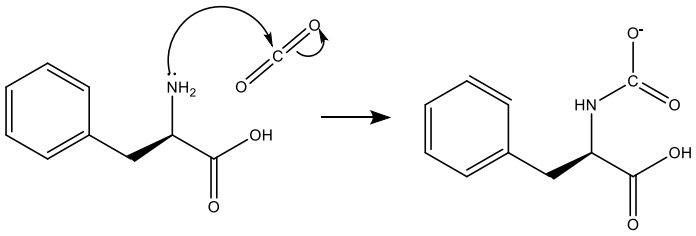


Figure 1.2.1: Mechanism of amine carbamylation by CO_2

A neutral amine on an amino acid such as phenylalanine has a lone pair of electrons, enabling it to make a nucleophilic attack on the less electron dense carbon of the CO_2 molecule. This results in the carbon dioxide becoming bonded to the nitrogen atom of the amine group.

nitrogen. The reaction is readily reversible, meaning that the carbamate is labile and easily dissociates, a property which is necessary to allow regulation of any downstream effects within a system. Whether a carbamate will form depends upon both the pK_a value of the amine group of the amino acid. The pK_a value of an amine group of an amino acid can be defined as the pH at which half of the amine group is ionised. At a physiological pH of 7.4 the majority of amino acids have amine groups with pK_a values of 9.0 or above (Vcelakova et al., 2004), and thus the vast majority of amine groups will be positively charged and unable to form a carbamate. However exceptions can occur due to the fact that the pK_a value of an amino acid can be influenced by its local environment (Kleinboelting et al., 2014). For example the specific environment of an amino acid within the 3D structure of a protein may lower its pK_a value thus allowing carbamate formation to occur. Lysine and arginine amino acids both have amine side chains and so are capable of forming a carbamate. Of these lysine is the most likely candidate due to the lower pK_a value of the amine group in an environment at physiological pH. In addition N-terminal amino acids have free alpha amino groups which tend to have lower pK_a values, meaning they could also be capable of forming a carbamate (Miller et al., 2014, Myers and Nelson, 1990)

In addition to pH, carbamate formation is also influenced by the amount of CO_2 present within a system; with more CO_2 available it may be expected that the amount of

carbamylation increases. However if the concentration of CO_2 is increased a decrease in pH will occur, depleting the supply of uncharged amine groups. This may explain why the photosynthetic enzyme Rubisco, which requires carbamylation in order to become activated, is less carbamylated and therefore less active under high CO_2 concentrations (Roh, 1996). Similarly the availability of molecular CO_2 is also regulated by pH; at the higher pH values favourable for carbamate formation the CO_2 equilibrium as shown in figure 1 will be shifted to the right, so less CO_2 will be in the molecular form and more will be present as HCO_3-. Therefore carbamate formation in a biological system relies upon a balance between CO_2 and pH, though it has been shown that lowering pH by a slight amount decreases carbamate formation more significantly than reducing the concentration of CO_2. (Gros et al., 1976).

The ability of the environment to regulate carbamate formation and the reversibility of the reaction therefore mean the modification is ideally suited to allow regulation of protein function by CO_2 within biological organisms. This has shown to be the case in several proteins, examples of which shall be discussed in the following sections.

1.3 Carbamates with structural and functional roles

Carbamylation is known to be an important protein modification (Meigh, 2015). Discussed here are several examples of proteins for which the formation of a carbamate has proven to be essential to their function. It will be shown that the carbamate modifications are necessary for a variety of different functions, from altering the structure of a protein to allow it to become activated for a particular purpose, to being directly involved as a substrate in reaction mechanisms.

1.3.1 Haemoglobin: formation of an N terminal carbamate for CO2 transportation

Haemoglobin is a globular protein consisting of two alpha and two beta amino acid chains. It has long been known that haemoglobin is able to bind to and transport carbon dioxide as well as oxygen, with each influencing the binding affinity of the other (Bohr, 1904). In well oxygenated environments the structure of haemoglobin allows it to preferentially bind oxygen tightly. However as the concentration of CO_2 increases, for example in respiring muscle tissue, the pH becomes more acidic due to the formation of carbonic acid. This leads to a lysine residue in the haemoglobin alpha subunit becoming protonated and allowing it to interact with the carboxylic acid of histidine 146 in the beta subunit. This in turn can then form a salt bridge with aspartate 94 in the beta subunit. The haemoglobin undergoes a conformational change in structure, meaning that oxygen is less tightly bound and once

released is not able to bind easily again. The result is that oxygen is released into tissue where it is most required (Hsia, 1998).

It has long been established that haemoglobin was also capable of binding and transporting CO_2. Henriques was the first to provide evidence suggesting that carbon dioxide binds to haemoglobin via the formation of a carbamate (Henriques, 1928). This was established simply by measuring the rate at which carbon dioxide was liberated from a mixture of haemoglobin and inorganic carbon (CO_2 and HCO_3^-) during shaking at a physiological pH (Henriques, 1928). Henriques found that carbon dioxide was at first produced rapidly, and then in a much slower second phase. He suggested this pattern was due to a formation of reversible carbamates which rapidly dissociated from haemoglobin upon shaking, followed by slow production of carbon dioxide from the reduction of HCO_3^- to CO_2 and H_2O.

At first Henrique's interpretation of his results was treated with controversy. At the time of his experiment the enzyme carbonic anhydrase had not yet been discovered, and Henriques had unknowingly inhibited it in his experiments, meaning not everyone was able to reproduce his results. Those who did achieve similar results to Henriques discounted the idea of a carbamates and (wrongly) attributed the first rapid production of CO_2 to a 'catalytic action' of haemoglobin itself (Slyke, 1930, Dirken and Mook, 1930). It was not until the discovery of carbonic anhydrase and the subsequent important experiments by Meldrum and Roughton, showing that only when carbonic anhydrase was inhibited could Henriques results be consistently reproduced (Meldrum and Roughton, 1933), that the carbamate hypothesis started to gain more ground.

The first decisive evidence that the CO_2 was forming carbamates on the alpha amino groups of the haemoglobin chains did not appear until the 1960's (Kilmartin and Rossi-Bernardi, 1969). This was achieved by modifying the alpha-amino groups of haemoglobin by reacting them with cyanate, making them unable to form carbamates. Using titration and CO_2 binding studies, they showed that blocking the N-terminals of both the alpha and beta chains of haemoglobin prevented all CO_2 binding. Later experiments on haemoglobin provided further proof of carbamate formation, with X-ray studies confirming the site of the modification to be the N-terminal Valine residue (Arnone, 1974). In addition further studies on the alpha and beta chains found that in deoxyhaemoglobin both have similar binding affinities for CO_2 (Perrella et al., 1975). However there are still no detailed structures showing haemoglobin with a carbamate modification.

Carbamate formation on haemoglobin is important both in terms of CO_2 transport and in altering the confirmation of the molecule. Initial studies used precipitation experiments to estimate the amount of carbamylated haemoglobin formed in mammalian blood (Ferguson,

1936, Ferguson and Roughton, 1934, Stadie, 1935). Though these studies provided further evidence of carbamate formation they were not very accurate in quantifying it, estimating that one third to three quarters of the CO_2 present in reduced blood is transported via carbamate formation. It is probable that an overly high pH led to their large overestimates, as carbamate formation increases greatly with an increase in pH as more of the amine side chains become neutral in charge. We now know that the figure of CO_2 transport by carbamate is closer to between 5-10% and in highly active respiratory tissue will actually drop to lower than this. This is due to the fact that the pH of the blood drops so low that the N terminal valines of the haemoglobin molecule become ionised and are unable to form carbamates. In addition to helping transport CO_2 the carbamate modification is also commonly thought to help stabilise the deoxygenated structure meaning more O_2 is offloaded to respiring tissues (Berg JM, 2002).

The ability of haemoglobin to form a functional carbamate is a clear example of how the environment of an amino acid can change its pK_a value. The pK_a value of the N terminal valine in haemoglobin is unusually low at around 7.0, with the beta chain valine being slightly lower than the alpha chain (Kaplan et al., 1982). Therefore even when more protons are present during acidification of the blood by CO_2, haemoglobin is still able to form a carbamate bond to transport CO_2.

1.3.2 Rubisco: carbamylation to facilitate metal ion recruitment

Rubisco is a crucial enzyme found in aerobic organisms such as plants and cyanobacteria, and is responsible for the conversion of CO_2 into ribulose 1-5 bisphosphate, which can then be used in the synthesis of biomass (Notni et al., 2008). In order for Rubisco to be activated CO_2 must first bind to the protein, facilitating the binding of a magnesium ion to the active site of the protein (Park et al., 2008). It has been shown that the CO_2 binds to Rubisco via the formation of a carbamate. This was shown early on by O'Leary et al who used ^{13}C NMR spectroscopy and $^{13}CO_2$ to identify that a carbamates was formed on an amino group of the enzyme (O'Leary M et al., 1979). However by using this method it was not possible to determine which residue the carbamate formed upon. In addition the alkaline pH used for the study meant that the carbamate could have been formed artificially under non physiological experimental conditions.

A different approach was taken by Lorimer et al, who developed a method to stabilise the specific carbamate onto the Rubisco protein, allowing the carbamylated peptide fragment to be identified using various types of chromatography (Lorimer and Miziorko, 1980). This was performed by supplying radioactively labelled $^{14}CO_2$ to Rubisco and stabilising the

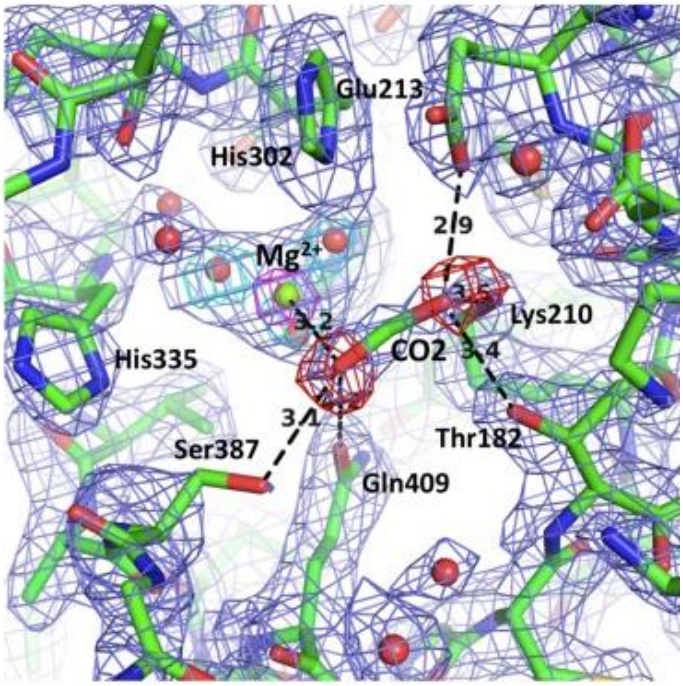


Figure 1.3.2.1: Carbamate in active site of Rubisco

The carbamylated Lys201 residue can be seen here in the pre-activation complex Rubisco from the red algae *Galdieria sulphuraria,* according to a recent crystal structure. The negatively charged carbamate allows recruitment and coordination of the magnesium ion. The interaction with the magnesium ion alongside hydrogen bonding to neighbouring residues also helps to stabilise the carbamate residue. *Image from Stec, 2012*

resulting carbamate using carboxyarabinitol 1, 5-bisphosphate (CABP). Diazomethane was used to methylate the Rubisco-carbamate complex, and the extremely stable methyl bond meant that the carbamate was now trapped and analysis could proceed. In this way they identified the carbamylated residue as lysine 201, found in the catalytic site of Rubisco (Lorimer and Miziorko, 1980).

More recent studies have clearly shown the carbon dioxide molecule in the active site of Rubisco using protein crystallography, as shown in Figure 1.3.2.1 (Stec, 2012, Maeda et al., 2000). These studies have also helped reveal that once formed the carbamate both allows the recruitment of and is stabilised by a Mg^{2+} ion, as well as hydrogen bonding between other key residues in the active site (Maeda et al., 2000). Only once both the carbamate and the magnesium ion are present can Rubisco become an active enzyme. Therefore it can be seen that Rubisco is another protein where carbamylation is crucial to stabilise the active structure of the enzyme in order to allow it to function.

The fact that Rubisco is most active at pH 8.0 or more is a further example of how pH of the environment is critical to carbamate formation. In the presence of light the pH of the stroma becomes more alkaline, and this shows a corresponding increase in Rubisco carbamylation and hence activity of the enzyme (Heldt, 1979). It is important to note that carbonic anhydrases which are required to supply CO_2 to the vicinity of the Rubisco are also most active at more alkaline pH's (Igamberdiev, 2015), so there is more than one contributing factor to the increase in Rubisco activity. However it is still apparent that in the case of Rubisco the stroma pH is regulating the activity of the enzyme through controlling the extent of carbamate formation.

1.3.3 Connexin 26: structural changes independent of metal ion recruitment

The proposed carbamylation of connexin 26 is an example of where a carbamate can interact with neighbouring residues to alter the structure of a protein, independently of interactions with metal ions. Meigh et al have recently shown that under constant pH connexin 26 (Cx26) hemichannels, which help control ventilation rates in mammals, are still able to respond to CO_2 (Meigh et al., 2013). They hypothesised that this was due to the CO_2 directly binding to a lysine residue in one of the Cx26 residues via formation of a carbamate. A 'carbamylation motif' was identified as a sequence of amino acids including a K124 residue that would accommodate a carbamate. This would then subsequently alter the structure of the connexion hemichannel by forming a salt bridge with a neighbouring arginine residue (Figure 1.3.3.1) causing the channel to open (Meigh et al., 2013). Mutational studies where the proposed carbamate forming K124 residue was changed to an alanine prevented the Cx26 channel from opening. In addition when the carbamylation motif was transferred to a non-CO_2 responsive connexin, it gained the ability to respond to CO_2 and opened. Finally these results were shown to occur even when pH remained constant, so was not due to changes in concentrations of H^+.

However this paper offered no structural data to prove that CO_2 was forming a carbamate, though computational modelling suggested that this was likely the case. As no attempt was made to distinguish between the CO_2 and HCO_3^- present in the experiments, it was also possible that HCO_3^- could have been responsible for the changes rather than molecular CO_2, though again this would not have fitted the proposed model. Therefore it has not yet been conclusively established that a carbamate is responsible for the response to CO_2 in the connexin.

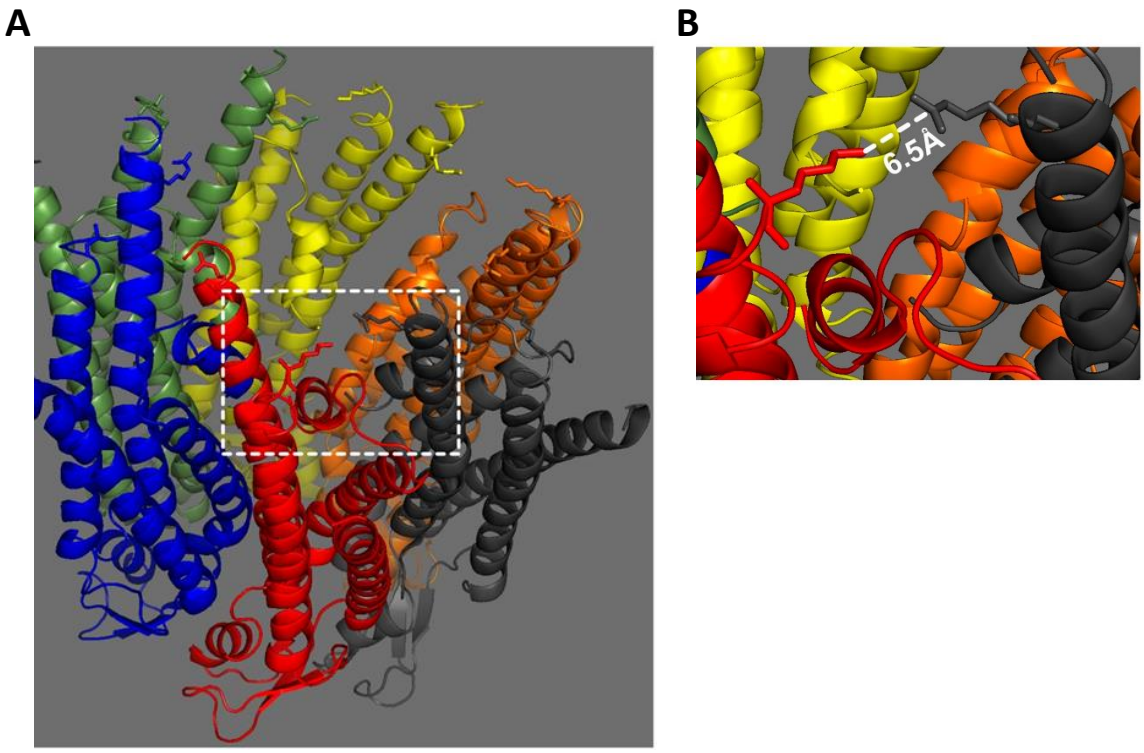


Figure 1.3.3.1: Proposed site of carbamylation in connexin-26

A) Crystal structure of Cx26 showing the site hypothesised to contain a carbamate. B) Magnification of site of carbamylation shown in dashed square. The K125 residue (red) is only a short distance from the R104 residue (grey). If the K125 residue was carbamylated it could potentially form a salt bridge to the R104 residue, activating the hemichannel. *Images from Meigh et al, 2013*

1.3.4 Class D β-Lactamases: examples of mechanistic carbamates

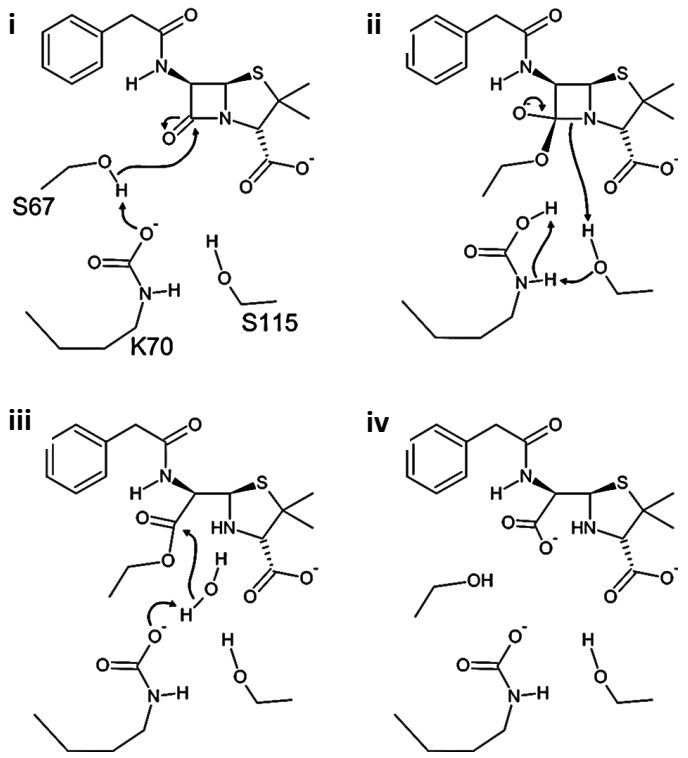
A relatively recent discovery of protein carbamylation is that of the proteins belonging to Class D of the β-Lactamases. β-Lactamases are a type of enzyme found in many bacteria which are capable of hydrolysing the antibiotic β-Lactam, and are grouped into four classes based on molecular structure (Leonard et al., 2013). Classes A, C and D are all serine hydrolases, whilst class B is zinc dependent (Bush et al., 1995, Therrien and Levesque, 2000). β-Lactamases pose a serious challenge to the use of β-Lactam antibiotics to treat diseases. Due to overuse of β-Lactams increasing numbers of bacteria are acquiring β-Lactamases and therefore resistance to the antibiotic (Bush, 2013). The main approach to combat the presence of β-Lactamases has been to inhibit them, though as there are over 1400 different β-Lactamase enzymes this is challenging (Leonard et al., 2013, Toussaint and Gallagher, 2015).

Class D β-Lactamases contain the prefix OXA in their nomenclature, due to their ability to hydrolyse oxacillin, a semi-synthetic penicillin (Leonard et al., 2013). In addition Class D β-Lactamases can also confer resistance against other types of β-Lactam antibiotics that are commonly effective against other classes of β-Lactamases. These include cephalosporins, carbapenems, and combinations of β-Lactam coupled with β-Lactamase inhibitors (Afzal-Shah et al., 2001, Paterson, 2006). Therefore it is particularly important to find new ways of inhibiting this class of enzymes.

It was first discovered that members of Class D β-Lactamases required a carbamate to function from initial crystal structures of the OXA10 enzyme from *Pseudomonas aeruginosa*. (Maveyraud et al., 2000, Golemi et al., 2001). Initially OXA-10 was shown to form a dimer and it was found that the K70 residues of each of the monomers possessed a region of electron density strongly suggestive of a carbamate (Maveyraud et al., 2000). The K70 residues were both found in the active sites of the enzyme, which are located on the same face of the protein dimer. Class D β-Lactamases hydrolyse β-Lactam via a serine acylation mechanism, whereby the substrate undergoes acylation by an active site serine residue in the first step of the catalytic process (Figure 1.3.4.1). It was suggested by Maveyraud et al that the carbamate hydrogen bonded to and was stabilised by the Ser-67 residue which it then activated to allow acylation to proceed (Maveyraud et al., 2000). The carbamate was also proposed to participate in the deacylation of the substrate, by activating the hydrolytic water molecule (Figure 1.3.4.1)

The active site structure showed the carbamate hydrogen bonded to both aTrp-154 nitrogen atom and a water molecule, further stabilising it and allowing it to participate in the hydrolysis mechanism. As further evidence that a carbamate was needed to form an active enzyme the group placed the decarbamylated protein in degassed (CO_2 free) buffer and found that it had no activity. When bicarbonate was added to increase the pH above pH 7.0 then the enzyme became active again (Maveyraud et al., 2000). However this basic experiment, though suggestive of a functional carbamate, did not take into account the possible effect of the change in pH altering the activity.

Experiments by Golemi et al provided additional evidence for the essential function of the carbamate in both the acylation and deacylation of Ser-67, by acting as a general base. It was shown that in addition to the stabilising interactions shown by Maveyraud et al, the amino acids in the immediate vicinity of the carbamate forming K70 caused a hydrophobic environment which would facilitate carbamate formation at a physiological pH (Golemi et al., 2001). This hydrophobicity of the active site region is much greater than in the other serine hydrolase classes A and B of β-Lactamases, which do not form carbamates (Leonard et al., 2013). The enzyme activity experiments were also repeated more thoroughly, where



(Leonard et al., 2013)

Figure 1.3.4.1: Role of the carbamate in the catalytic mechanism of Class D β-lactamases

A proposed role for the carbamate on K70 of Class D β-lactamases is to act as a general base for the H^+ ion belonging to the neighbouring S67 residue, activating the first step of the reaction (i). This causes acylation of the substrate by S67 (ii) resulting in transfer of electrons from the substrate to S115 and onto the carbamate, deprotonating the carboxyl group and activating a water molecule in the active site (iii) The water molecule then hydrolyses the ester linkage of the β-lactam, releasing the inactivated substrate from the enzyme (iv). Mechanism published by Leonard et al., 2013.

increasing activation of OXA-10 was shown with increasing amounts of bicarbonate and correspondingly more carbamate formation, this time controlling for pH effects (Golemi et al., 2001). ^{13}C NMR was also used to prove the location of the carbamate, and mutations of the K70 residue to Ala produced a completely inactive protein. Therefore overall there was a large amount of data showing that a carbamate was forming and influencing the activity of the enzyme. The authors also found that the extent of carbamylation was dependent on pH with full carbamylation at pH 8.0, none at pH 6.0, and intermediate levels at the pH values in between. As we have already seen how pH affects protonation status

and hence carbamylation this is not a surprise; as the pK_a value of K70 was found to be just 5.8 then at physiological pH's it should be fully carbamylated (Golemi et al., 2001).

The combination of the results from these important studies conclusively established for the first time that carbamylation is a critical mechanism for the activation of the OXA-10 β-Lactamase. Since these initial findings several Class D β-Lactamases have been shown to form a carbamate necessary for the acylation of the serine residue to occur, including OXA-1, OXA-24, OXA-46 and OXA-48 (Leonard et al., 2008, Che et al., 2014, Docquier et al., 2010, Docquier et al., 2009). Some of these β-Lactamases are monomeric and some are dimeric. However the hydrophobic active site containing the key residues needed to form a carbamate is conserved among each of the proteins. In contrast to the majority of studies on haemoglobin and Rubisco all of the Class D β-Lactamases were found to bind carbamates from analysing the crystal structures of the proteins. This was likely because having shown this was an effective method for analysing OXA-10 it was then repeated for other members of the class.

Despite all being able to form a carbamate not all the Class D β-Lactamases catalyse the hydrolysis of their substrate by the same mechanisms, highlighting the importance of other components of the active site in influencing the functions of carbamates. OXA-1 and OXA-24 are both monomeric proteins which form carbamates at K70 and K84 respectively, but their ability to do so differs (Che et al., 2014). OXA-10 is carbamylated more easily and decarbamylated with more difficulty than OXA-1 (Che et al., 2014). It has been shown that this is due to a water molecule that is found in the active sites of both OXA-1 and OXA-10, but not OXA-24, which lowers the energy needed for carbamylation by helping to position it correctly for attack by CO_2 also stabilising it (Li et al., 2005, Che et al., 2014). This has implications on the resistance of these β-Lactamases to inhibitors, by influencing their reaction mechanisms when inactivating their substrates. Whilst the penem 1 inhibitor can decarboxylate OXA-1 it is unable to do so to OXA-24 unless conditions where there is a great excess of the antibiotic (Che et al., 2014). Therefore in order to understand how a carbamate influences the structure or reaction mechanism of a protein it is important to take into account how it interacts with the amino acids and molecules surrounding it. In the case of Class D β-Lactamases this is an important aspect of inhibitor design.

Class D β-Lactamases are unique among the other examples of protein carbamylation which have been discussed, in that they form carbamates which participate in a reaction mechanism of an enzyme, rather than just altering the structure of the protein to allow it to function. It would be interesting to see if more protein carbamates with this type of mechanistic function are discovered in future.

1.4 The need to further understand protein carbamylation

There is a clear need for organisms to be able to detect and respond to CO_2, and it is not surprising that carbamate modifications play important roles in the functions of the proteins discussed in the previous section. What is more unexpected is that given the prevalence of CO_2 in most biological systems, relatively few proteins have been shown to be carbamylated. This could be due to a number of reasons. Firstly, research on carbon dioxide detection and signalling has often focused on the response to changes in pH caused by CO_2 rather than any direct interactions with molecular CO_2, as is the case with carbamate formation. There is comparatively less material published where the mechanisms by which carbon dioxide might modify proteins in a biologically relevant manner have been explored. Secondly it is hard to investigate protein carbamate bonds thoroughly due to their labile nature (Lorimer, 1983). Unless it is stabilised by interactions within the protein or with ligands the carbamate bond itself is too unstable to withstand the rigours of analytical techniques such as Mass spectrometry and NMR. This means that it is hard to gain information about the precise positions and interactions of protein carbamate bonds.

Therefore it would be highly advantageous to be able to develop a method for rapid analysis of proteins to determine if they form a carbamate modification, without the risk of the carbamate disappearing during the analysis process. This would allow a greater understanding of the extent of physiological protein carbamylation, and of the role that carbamates play in CO_2 detection and signalling. It would also allow us to better understand the carbamate modification itself, and how likely it is to occur on different amino acids. For example, excluding N terminus carbamates, all the proteins discussed in the previous section were shown to form a functional carbamate do so on a lysine residue, despite arginine also having an amine side chain that could bind a carbamate. This is likely because of the lower pK_a value of lysine making it more likely to form a carbamate at physiological pH values. It would be interesting to see if any functional carbamates do form at arginine residues of proteins, and a rapid and accurate method of investigating carbamylation could ensure this.

An additional point of interest is that in almost all cases of carbamate formation discovered the carbamate has been found in the active site of an enzyme, often coordinating a metal ion (Stec, 2012, Jabri et al., 1995, Benning et al., 1995). In these cases the carbamate tends to play either a structural role in activating the enzyme or in some cases directly participates in the functional mechanism. This is to be expected as the anionic charge that a carbamate creates makes it ideal for interactions with metal ions or hydrogen bonding to other residues.

What is more surprising is that more carbamates have not been discovered bound at the N terminal residues of proteins. Only in haemoglobin has this been shown to occur, where it plays a role in both transport of CO_2 and in stabilising the protein's deoxygenated structure. Given that all proteins contain an N terminal residue with a free amine group it could be expected that a least a few proteins would have an N terminus with a sufficient pK_a value to allow carbamate formation. Perhaps as these hypothetical carbamates may not directly participate in the main known functions of the proteins as occurs in many enzyme carbamates, they have simply not been discovered. Whether they exist and if so have an important functional role or not are both questions that could be answered by further research into carbamate modifications.

1.5 Methods for stable trapping of carbamate modifications

In the past various methods have been used to show that a carbamate is bound to a protein. Most of the techniques have been discussed already in regard to the studies carried out on haemoglobin and Rubisco. Labelling of carbamates using radioactive carbon dioxide and identifying their presence by an NMR shift were originally commonly used methods(O'Leary M et al., 1979), however the conditions used for these experiments are highly artificial in order to stop the carbamate dissociating, and so not representative of a physiological environment. In addition little information can be gathered this way about the precise location of the carbamate on the protein. Methods using crystallography (Stec, 2012) give detailed information about the carbamate's location and its interaction with other proteins, but again conditions used are often artificial and the process can be lengthy. In addition neither of these methods alone are suitable to use for screening to detect carbamate modification, each is used to prove or disprove the concept that a carbamate is already present.

The ideal method for identifying a carbamate would be to first stabilise it on the protein under physiological conditions to prevent artificial results. This would then facilitate analysis using techniques such as Mass Spectrometry and NMR. A method of stabilising a carbamate was first used by Lorimer et al who used diazomethane to successfully methylate a carbamate on Rubisco, allowing it to be rigorously analysed (Lorimer and Miziorko, 1980). A more recent study on trapping of carbamic acid species also showed that diazomethane was capable of methylating carbamates to create highly stable derivatives (Ito and Ushitora, 2006). However there are downsides to the use of diazomethane, chiefly that it is very hazardous to work with due to the risk of explosions

from its extreme reactivity and also because it is very toxic (Moore, 1973). Additionally it can only be used in organic solvents so is not ideal for studying proteins in their native conditions, and could also result in protein denaturation which could cause release of any carbamates before they can be trapped.

An alternative alkylating reagent that has the been shown to be able to ethylate rather than methylate carbamates is Triethyloxonium tetrafluoroborate (TEO), also known as Meerwein's Reagent (Chen and Benoiton, 1977). TEO is less toxic and less reactive than diazomethane under experimental conditions, making it a safer reagent to use to stabilise carbamates. The use of TEO to investigate properties of proteins is not a new one, however the focus of previous studies has been on carboxyl groups rather than carbamates, with TEO being used to ethylate carboxyl groups of insulin, trypsin, and lysozyme (Levy, 1973, Parsons et al., 1969, Ben Avraham and Shalitin, 1985). The related trimethyloxonium tetrafluoroborate has also been used to ethylate carboxyl groups belonging to pepsin (Paterson and Knowles, 1972).

In all these studies the purpose of the ethylation was to subject one or more carboxyl groups of the proteins to a protective modification, in order to prevent any normal functions they may have within the protein. In this way the contribution of the carboxyl groups to the proteins structure and/or functional activity could be assessed. Interestingly despite the additional ethyl groups altering the overall charge status of the protein, all of the proteins were found to retain at least some extent of their functions, although their specific activity was often much reduced. For example trypsin ethylated at its binding site lost 80% of its ability to cleave proteins specifically at the carbonyls of lysine and arginine residues, however it was still able to cleave non-specifically at a reduced rate, so was still at least a partly functional protein (Ben Avraham and Shalitin, 1985). This is important as it means that the ethyl modifications do not necessarily mean that a protein will be completely denatured, which could be a useful feature for assessing the functional properties of a protein with ethylated carbamates. However further investigation would be needed to determine whether the situation would be the same in carbamate ethylation experiments as in carboxyl ethylation experiments. The ability of the protein to retain its function would also be dependent on the extent of the ethylation which would also have to be determines experimentally.

Investigations into the use of TEO as a trapping agent for carbamates have been ongoing at Durham University and have shown that it is capable of stabilising carbamates onto amino acids, which can then be analysed by Mass spectrometry (Vicki Linthwaite, unpublished research). The mechanism of carbamate ethylation by TEO is shown in Figure 1.5.1.

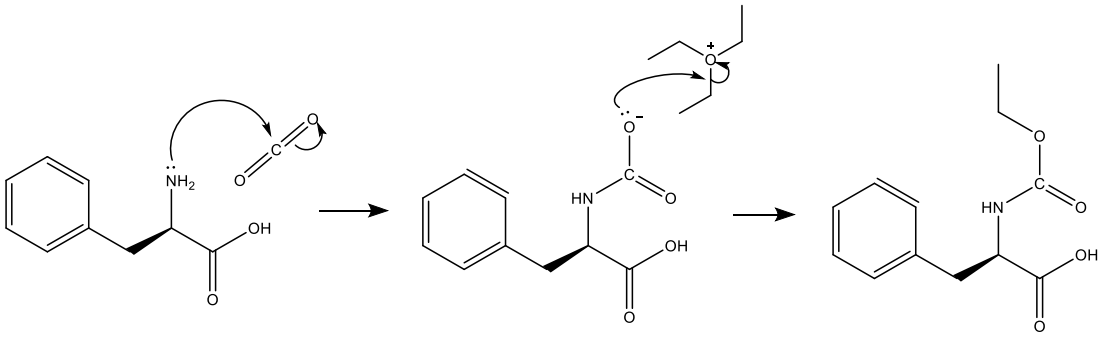


Figure 1.5.1: Trapping of a carbamylated amine group by TEO

The lone pair of electrons of a neutral amine (shown here on phenylalanine) makes a nucleophilic attack upon a CO_2 molecule, forming a negatively charged carbamate. The carbamate group then makes a nucleophilic attack upon a TEO molecule, causing an ethyl group to become stably bound to the carbamate. Therefore the carbamate is now trapped onto the phenylalanine as it is no longer able to dissociate.

The implications of using TEO to trap carbamates are far reaching. The method is rapid and can be used to trap carbamates that have already formed on proteins, without having to alter the pH of the protein or expose it to a non-aqueous potentially denaturing environment. This is not to say that TEO will not denature a protein eventually, the extensive ethylations of not just carbamates but all free carboxylic acid and amine group mean that the protein becomes hydrophobic and forms a non-functional precipitate (unpublished research). However by this point all carbamates should be trapped and so the protein structure and function is no longer needed to be intact. Furthermore, if necessary it may be possible to experimentally determine a concentration of TEO that will trap carbamates without fully denaturing the protein, as discussed earlier (Ben Avraham and Shalitin, 1985).

Stable trapping of carbamates in this way means that instead of having to painstakingly analyse one protein at a time, as has been undertaken for haemoglobin and Rubisco, theoretically many proteins can be exposed to the trapping agent at once, for example in an extract from a cell lysate. Any carbamates will then be trapped and the proteins can be analysed simultaneously using proteomic mass spectrometry techniques, allowing those containing carbamates to be identified. Therefore this would be a much faster approach to identifying proteins whose functioning relies on the formation of a carbamate by screening, opening up a potentially new and large field of carbamate biology.

1.6 Aims of this project

This first main aim of this project was to develop a method to trap carbamate modifications onto proteins using the ethylating chemical reagent Triethyloxonium tetrafluoroborate (TEO). To do this it first had to be confirmed that TEO could trap carbamates on an amino acid. After this was achieved the trapping method could be adapted for use on proteins, using haemoglobin as a control protein for method development as it is known to form a carbamate. At this stage a methodology for analysis of the sample using mass spectrometry was also developed, in addition to a way of analysing the results.

The second main aim was to use the method developed to test proteins unknown to form a carbamate. Adenylyl cyclases were chosen for this, as will be explained later. In addition to mass spectrometry the trapping method was also adapted to use in radioactivity experiments.

The ultimate aim of this project is to contribute towards the development of a method to screen many proteins in a cell lysate for carbamate modifications simultaneously, using mass spectrometry proteomics. This will be a valuable tool in enabling us to better understand the extent and the roles of protein carbamylation in physiology. Although this final aim was beyond the scope of this project, the experiments and results detailed in this thesis will form a key component of its progress.

2 Materials and methods

2.1 Trapping and analysis of carbamate formation

2.1.1 Trapping carbamates on phenylalanine

20 mg of phenylalanine was dissolved in 3 ml deionized H_2O. 20 mM $NaHCO_3$ was added to provide CO_2 at a physiologically relevant concentration. A total of 400 mg of TEO was added in 100 mg quantities. Each 100 mg of solid TEO was dissolved in 250 µl of H_2O and immediately added to the reaction. The final TEO concentration was 0.97 mM. The mixture was incubated at 25^{o}C for one hour, during which pH was maintained as close to 7.4 as possible using a pH stat (section 2.1.2). After one hour the mixture was placed in a round bottomed flask and frozen rapidly using an acetone ice bath. A freeze dryer was used to lyophilise the sample. The powder was then stored at -20^{o}C until analysis by ESI-MS.

2.1.2 Controlling pH using pH stat

A Titralab 854 pH stat was used to maintain the correct pH throughout the trapping experiments. The STAT software was used to set the desired pH and other conditions, with pH 7.4 being used for all proteins except Rubisco. 1 M NaOH was used to adjust the pH. The time constant was set to 5.0 seconds, Gain to 0.1 and maximum addition speed to 5 ml/minute with stirring at 250 rpm. A water bath was used to maintain temperature at 25^{o}C. Both the pH and the volume of NaOH added were recorded throughout the duration of the experiment.

2.1.3 LC/ESI analysis of carbamylated phenylalanine

After the trapping experiment (Section 2.1.1), 0.5 mg of lyophilized phenylalanine was transferred to a glass vial. 1 ml of methanol was used to dissolve the powder to give a concentration of 0.5 mg/ml. The sample was submitted for analysis by Electrospray Ionisation coupled to Liquid Chromatography (LC/ESI). The calculated m/z values of the potential carbamylated and ethylated species produced in the experiment were used to identify whether they were present in the analysed sample.

2.1.4 Trapping carbamates on proteins for Mass Spectrometry analysis

Trapping experiments were performed on haemoglobin, Rubisco, 7C1.2C2 mammalian AC and Rv1625c$_{203\text{-}443}$ proteins. Haemoglobin and Rubisco were purchased as recombinant proteins. The 7C1 and 2C2 mammalian AC proteins and Rv1625c$_{203\text{-}443}$ proteins were recombinantly expressed in *Escherichia coli* (Section 2.2). For all MALDI mass spectrometry experiments 1 mg of each protein was used, except during initial experiments when different quantities of haemoglobin was used for method optimisation. It has been previously shown that mixing the 7C1 and 2C2 subunits of the mammalian tmAC together

prior to an experiment forms a functional catalytic site (Townsend et al., 2009) therefore equal quantities of the two subunits were combined prior to the start of each experiment.

For all experiments 1mg of the relevant protein was used in 3ml phosphate buffer, which was maintained at the physiological pH of the protein's natural environment. This was pH 7.4 for all proteins except Rubisco, which was pH 8.0. 20 mM $NaHCO_3$ was added immediately prior to the start of the experiment to provide CO_2 at a physiologically relevant concentration. Experiments were performed using both ^{12}C and ^{13}C labelled $NaHCO_3$. A total of 280 mg of TEO was added in 70 mg quantities. Each 70 mg of solid TEO was dissolved in 250 µl of H_2O and immediately added to the reaction. The end TEO concentration was 0.68 mM The mixture was incubated at 25^oC for one hour during which pH was maintained as close to correct pH as possible using the pH stat (Section 2.1.2) After one hour the ethylated protein had precipitated out of solution. The whole mixture including the precipitate was dialysed overnight against 4 L of distilled H_2O (1:1000 ratio of sample: dialysis buffer). After dialysis the sample was centrifuged to pellet the precipitated protein. The supernatant was concentrated using spin columns to a volume of 100 µl to concentrate any non-precipitated protein. The pellet of precipitated protein was then resuspended in the 100 µl of concentrated supernatant, and the sample trypsin digested (Section 2.1.5).

2.1.5 Trypsin digestion of proteins

A 100 µl sample containing mostly precipitated protein from the trapping experiment (Section 2.1.4) was used for trypsin digestion. DTT was added to a concentration of 5 mM and the sample incubated at 50^oC for 20 minutes. Ioadacetamide was then added to a concentration of 15 mM and the sample incubated at room temperature, in the dark, for 15 minutes. 10 µl of Mass spectrometry grade Trypsin at a concentration of 0.2 mg/ml was added and the samples incubated shaking, overnight at 37^oC.

2.1.6 Analysis by Matrix Assisted Laser Desorption Ionisation (MALDI) and Tandem Mass Spectrometry (MS/MS)

After trypsin digestion (Section 2.1.5) samples were centrifuged to pellet any remaining precipitated protein. The supernatants were passed through C18 ZipTip pipette tips to both remove unwanted salts and to concentrate the trypsin cleavage fragments. The ZipTip was equilibrated by wetting 2x in 10 µl Acetonitrile followed by 1x 10 µl 0.1% (v/v) TFA. The sample was added to the ZipTip by taking 10 µl of supernatant and pipetting up and down slowly 10 times, then repeating with another 10 µl supernatant. The ZipTip was then washed using 3x10 µl of 0.1% (v/v) TFA. The protein trypsin digest fragments were eluted by pipetting up and down slowly 3x using 5 µl of 30% (v/v) Acetonitrile, 0.1% (v/v) TFA. 1 µl of

each ZipTipped sample was spotted in triplicate onto a MALDI sample plate and mixed immediately with 1 µl of CHCA matrix and allowed to crystallise. The samples were analysed by MALDI/TOF. An Excel database was used to compare expected masses of carbamylated protein fragments to the fragment masses found in the sample using MALDI.

Protein fragments that contained a carbamate as identified by MALDI were then analysed by MS/MS. The selected fragment was further fragmented using collision induced dissociation and the product ions separated by mass in a second mass spectrometer.

2.1.7 $^{14}CO_2$ carbamylation assays

Protein was dissolved or dialysed into 50 mM phosphate buffer, pH 7.4. 5 mg of protein was used in each trapping experiment, with the exception of Rubisco where 2.5 mg was used. 6.8 mg of $NaHCO_3$ was dissolved in 1 ml of the phosphate buffer (final concentration 20 mM) and added immediately prior to the start of the experiment to provide CO_2 at a physiologically relevant concentration. 27 µCi of $NaH^{14}CO_3$ was added to the experiment and a 10 µl sample was immediately taken and added to a scintillation vial for scintillation counting. For scintillation counting 3 ml of Eco-Scint scintillation fluid was mixed with the sample and the ^{14}C radioactivity measured using a scintillation counter. A total of 280 mg of TEO was added to the experiment in 70 mg quantities, with constant stirring. Each 70 mg of solid TEO was dissolved in 250 µl of H_2O and immediately added to the reaction. The end TEO concentration was 0.68 mM. The pH was constantly monitored for 15 minutes from addition of the first amount of TEO using a pH probe and maintained at approximately pH 7.4 by manually adding small volumes of dilute NaOH. For Rubisco pH was maintained at a physiological pH of pH 8.0. After 15 minutes the pH had stabilised and the pH was no longer adjusted. The experiment was left for one hour after which time all TEO was hypothesised to have been hydrolysed. The mixture was centrifuged for 25 minutes at 4000 g to pellet the precipitated protein. The pellet was then washed by resuspending in 1ml of 50 mM phosphate buffer, pH 4.0, and centrifuging again for 2 minutes to pellet the precipitate. The wash step was repeated 2x then the pellet was resuspended in 500 µl of H_2O. The suspension was transferred to a scintillation vial and mixed with 3 ml of scintillation fluid, and then ^{14}C radioactivity measured using a scintillation counter.

Control experiments were performed using no TEO. These experiments were carried out in the same way but 1 ml of water was added to the mixture instead of 4x 250 µl of TEO. After an hour 8 ml of ice cold acetone was added to the protein solution, which was vortexed then left for 20 minutes at -20^{o}C to ensure all protein had precipitated. The pellets were then centrifuged and washed in acidic buffer as described above, and the ^{14}C radioactivity counted.

Additional controls were performed using no radioactivity. These experiments were carried out in an identical way to the main experiments but no $NaH^{14}CO_3$ was added at the start of the experiment.

2.2 Protein expression and purification

2.2.1 Preparation of competent cells

E. coli cells were taken from a glycerol stock maintained at -80°C and grown in 5 ml of LB media overnight at 37°C, shaking at 250 rpm. For strains containing antibiotic resistance to ampiciillin and/or kanamycin the LB media contained the antibiotics at a concentration of 100 µg/ml and 50 µg/ml respectively. The overnight culture was diluted 1:100 using fresh LB media to a total volume of 10 ml. The culture was incubated at 37°C with shaking at 250 rpm until it reached an OD_{600} of 0.4. The cells were then harvested by centrifugation at 2700 g for 10 minutes and the supernatant removed. The pellet was resuspended in 5 ml of ice cold 20 mM $CaCl_2$, 80 mM$MgCl_2$ by gentle pipetting. The cells were centrifuged again at 4°C for 10 minutes, 2700 rpm and the supernatant removed. The pellet was resuspended in 5 ml of ice cold 100 mM $CaCl_2$ by gentle pipetting. The cells were centrifuged for a final time at 4°C for 10 minutes, 2700 g and the supernatant removed. The pellet was resuspended in 0.5 ml of ice cold 100 mM $CaCl_2$, containing 15% (v/v) glycerol. The cells were immediately aliquoted into 50 µl aliquots and stored at -80°C for future use.

2.2.2 Isolation of plasmid DNA

Plasmid DNA was isolated using the Qiagen mini prep kit according to the manufacturer's instructions. *E. coli* were grown overnight from glycerol stocks at 30°C with shaking at 200rpm in 5 ml LB containing appropriate antibiotics. The culture was harvested by centrifugation at 5,445 g for 5 minutes at room temperature and the supernatant removed. DNA was then extracted from the bacterial pellet using the ThermoFisher Scientific GeneJet Mini-prep kit. All of the following steps were carried out at room temperature, using a benchtop microcentrifuge at 1200rpm or centrifugations. The bacterial pellet was resuspended in 250 µl of Resuspension Solution containing RNase A, and transferred to a microcentrifuge tube. 250 µl of Lysis Solution was added and the tube inverted four times to mix. 350 µl of Neutralisation Solution was added and the tube inverted four times to mix. The mixture was centrifuged for five minutes to pellet debris. The supernatant was transferred to a GeneJET spin column. The spin column was centrifuged for 1 minute and the flow-through was discarded. 500 µl of wash solution was added to the spin column, the column was centrifuged at for 1 min and the flow-through was discarded. The wash step

was repeated using 500 µl of wash solution and the flow-through discarded. The spin column was centrifuged for an additional 1 minute to remove residual wash solution. The spin column was transferred to a clean 1.5 ml microcentrifuge tube. 50 µL of elution buffer was added to the centre of the resin and left to incubate for 2 minutes. The spin column was then centrifuged for 2 minutes to elute the DNA. The elution step was repeated using 50 µl of deionised H_2O. DNA concentration and purity was measured at 260 nm using a Nanodrop 2000 (Section 2.2.7) Purified plasmid DNA was stored at -20ºC until needed.

2.2.3 Transformation of plasmid DNA into competent cells

1 µl of plasmid DNA corresponding to approximately 20ng of DNA (Section 2.2.2) was added to 50 µl of competent *E. coli* (Section 2.2.1), and the mixture incubated on ice for 20 minutes. The cells were heat shocked for 30 seconds at 42ºC, then placed on ice again for 2 minutes. 450 µl of antibiotic free LB warmed to 30ºC was added to the cells, which were then incubated for 1 hour at 30ºC, shaking at 200rpm. After an hour ~200µl of the *E. coli* were spread on an agar plate containing appropriate antibiotics and incubated overnight at 30ºC to allow colonies to grow. Plates could be stored for up to a week at 4ºC, until a colony was needed.

2.2.4 Expression and Purification of Rv1625c204-443 wild type and mutant proteins

DNA was isolated from *E. coli* BL21 (DE3) containing the pQE30-Rv1625c$_{204-443}$wt plasmid (Guo et al., 2001) and *E. coli* M15 (DE3) containing the Rv1625c$_{204-443}$ mutant plasmids (Section 2.2.2). The DNA was transformed into competent M15 (DE3) cells (Section 2.2.3). 100 µg/mL ampicillin and 50 µg/mL kanamycin were added to the LB agar during selection for antibiotic resistance. One of the colonies produced after transformation was removed from the agar plate and transferred to 150 mL LB containing 100 µg/mL ampicillin and 50 µg/mL kanamycin. The bacteria were incubated overnight at 30 °C, 200 rpm. 10 ml quantities of the overnight culture were added to 12 x 1 L LB, each in 2.5 L baffled flasks containing 100 µg/mL ampicillin and 50 µg/mL kanamycin. The cultures were grown at 30°C, 140 rpm to O.D.$_{600}$ = 0.6. Rv1625c$_{204-443}$ protein production was induced with 45 µM isopropyl-β-D-thiogalactopyranoside (IPTG) and incubated at 30°C for 3 hours. The culture was harvested by centrifugation at 4,000 g for 40 minutes at 4°C using a Beckman JLA 8.1000 rotor. The supernatants were removed and the bacterial pellets washed and resuspended in 20 ml wash buffer per 2 L culture (50 mM Tris-HCl pH 8.5, 1 mM EDTA). The cells were centrifuged at 5,445 g for 30 min at 4 °C using a Beckmann Avanti 30 centrifuge, then supernatants removed. The bacterial pellets were stored overnight at -80°C. Pellets were defrosted rapidly in a 37°C water bath, with continual agitation to allow

them to remain cool, then kept on ice for the rest of the procedure. The pellets were resuspended in 10 mL of cell lysis buffer per original 1 L culture (50 mM Tris-HCl pH 8.5, 300 mM NaCl, 10 mM 1-thioglycerol and 2 mM $MgCl_2$). The cells were lysed by sonication for 2x 90 second cycles at 4°C. Cell debris was removed by centrifugation at 20,000 g for 45 minutes at 4°C, using a Beckmann Avanti 30 centrifuge. 0.5 mL of nickel-nitrilotriacetic acid (Ni-NTA) slurry per original 1 L culture was added to the supernatant and the mixture was incubated for 3 h at 4 °C. The mixture was centrifuged at 400 g for 1 min and the slurry poured into a 12 ml Polyprep column attached to a peristaltic pump. The columns were washed at a flow rate of 1 ml/min with 10x bed volume wash buffer A (50 mM Tris-HCl pH 8.5, 10 mM 1-thioglycerol, 2 mM $MgCl_2$, 400 mM NaCl and 5 mM imidazole), 10x bed volume wash buffer B (wash buffer A + 15 mM imidazole), 10x bed volume wash buffer C (wash buffer B + 10 mM NaCl). The last 1ml of each wash step was saved for analysis by SDS-PAGE. The pump was removed and the protein was eluted in 1ml fractions using elution buffer (wash buffer C + 150 mM imidazole). Elution fractions were collected until the concentration dropped below ~0.2mg/ml, with the nanodrop being used to measure protein concentration (Section 2.2.7). Protein MW and quality was established using SDS-PAGE analysis (Section 2.2.8). 20% (v/v) glycerol was added to all eluates which were frozen at -80ºC for storage.

2.2.5 Expression and Purification of mammalian trans-membrane adenylyl cyclase (tmAC) 7C1a domain

DNA was isolated from *E. coli* Tuner (DE3) containing the pProEx-HAH6-7C1a$_{263-476}$ plasmid (Section 3.2.2). The DNA was transformed into competent Tuner (DE3) cells (Section 3.2.3). 100 µg/mL ampicillin was added to the LB agar during selection for antibiotic resistance. One of the colonies produced after transformation was removed from the agar plate and transferred to 150 mL LB containing 100 µg/mL ampicillin. The bacteria were incubated overnight at 37°C, 200 rpm. 10 ml quantities of the overnight culture were added to 12 x 1 L LB, each in 2.5 L baffled flasks containing 100 µg/mL ampicillin. The cultures were grown at 37°C, 140 rpm to O.D.$_{600}$ = 0.7. tmAC-7C1a protein production was induced with 30 µM isopropyl-β-D-thiogalactopyranoside (IPTG) and the culture incubated at 20°C, 140 rpm overnight for a total time of ~17 hours. The culture was harvested by centrifugation at 4,000 g for 40 minutes at 4 °C using a Beckman JLA 8.1000 rotor. The supernatant was removed and the bacterial pellets washed and resuspended in 20ml wash buffer per 2L culture (50 mM Tris-HCl pH 8.5, 1 mM EDTA). The cells were centrifuged at 5,445 g for 30 minutes at 4°C using a Beckmann Avanti 30 centrifuge, then supernatants removed. The bacterial pellets were stored overnight at -80 °C. Pellets were defrosted rapidly in a 37°C,

with continual agitation to allow them to remain cool, then kept on ice for the rest of the procedure. The pellets were resuspended in 10 mL of cell lysis buffer per original 1 L culture (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM imidazole, 5 mM $MgCl_2$, 2mM DTT and 1 protease inhibitor tablet/20 ml buffer). The cells were lysed by sonication for 2x2 minute cycles at 4°C. Cell debris was removed by centrifugation at 20,000 g for 45 minutes at 4 °C, using a Beckmann Avanti 30 centrifuge. 0.2 mL of nickel-nitrilotriacetic acid (Ni-NTA) slurry per original 1 L culture was added to the supernatant and the mixture was incubated for 3 h at 4 °C. The mixture was centrifuged at 400 g for 1 min and the slurry poured into a 12 ml Polyprep column attached to a peristaltic pump. The columns were washed at a flow rate of 1ml/min with resin wash buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 20 mM imidazole, 2 mM DTT, 5 mM $MgCl_2$), until the protein concentration in the eluted was buffer was less than 0.1 mg/ml. The pump was removed and the protein was eluted in 1 ml fractions using elution buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 20 mM imidazole, 2 mM DTT, 5 mM $MgCl_2$). Elution fractions were collected until the concentration dropped below ~0.1 mg/ml, with the nanodrop being used to measure protein concentration (Section 3.2.7). Protein MW and quality was established using SDS-PAGE analysis (Section 3.2.8). 20% (v/v) glycerol was added to all eluates which were frozen at -80^{o}C for storage.

2.2.6 Expression and Purification of mammalian adenylyl cyclase 2C2a domain

Expression and purification of the tmAC-2C2a domain was carried out in exactly the same way as for the 7C1a domain, with the following exceptions.

- DNA was isolated from E. coli M15 (DE3) containing the pQE60-2C2821-1090 plasmid (Section 2.2.2).
- The DNA was transformed into competent M15 (DE3) cells (Section 2.2.3).
- Antibiotics used in agar and LB media throughout expression were 50 µg/ml Kanamycin and 100 µg/ml Ampicillin.

2.2.7 Quantification of protein concentration using UV absorbance.

Protein concentration was measured using a Nanodrop 2000 spectrophotometer. The Nanodrop was first blanked using 2 µl of the appropriate buffer. 2 µl of protein solution was added to the Nanodrop and absorbance measured at 280 nm. When measuring the concentration of a purified protein the known molecular weight and extinction coefficients were used to calculate the precise concentration.

2.2.8 SDS PAGE analysis

Protein molecular weight and purity was assessed by running the proteins on SDS-PAGE gels. A 15% (w/v) gel resolving gel was used with a 5% (w/v) stacking gel. 4x LDS sample buffer was added in a 1:3 ratio of protein to sample buffer, and the samples incubated at 90^{o}C for 5 minutes to denature the proteins. For analysis of wash and elution fractions from protein purification a 6 µl sample was taken from a 1 ml fraction and mixed with 2 µl of loading buffer. When analysing known amounts of purified protein 5 µg of protein was loaded onto the gel. Proteins were loaded onto the gel along with a molecular weight standard and electrophoresed at 20 V/cm of resolving gel, in Tris-Glycine running buffer (250 mM Tris, 1.92 M Glycine, 1% (w/v) SDS). Protein bands were stained overnight using Coomassie Instant Blue.

2.3 Materials

Specific materials are listed in the tables below. All other chemicals and materials used were obtained from Sigma Aldrich.

2.3.1 Protein production materials

Product	Source
Imidazole	Melford # B4005
Nickel-nitrilotriacetic (NTA) resin	Thermo Fisher # R901-01
IPTG	Melford # MB1008
QIAprep Spin Miniprep kit	Qiagen # 27104
Coomassie Instant Blue	Sigma # ISB1L-1L
LB Miller broth (high salt)	Melford # GL1704
LB Agar	Melford # GL1706
Ampicillin	Melford # A0104
Kanamycin	Melford # K0126

2.3.2 Purchased proteins

Product	Source
Haemoglobin (Human)	Sigma #H7379
Rubisco (Spinach)	Sigma #R8000

2.3.3 Carbamate trapping materials

Product	Source
Triethyloxonium tetrafluoroborate	Sigma #90520
$NaH^{14}CO_3$	American Radiochemicals Inc #0138B

2.3.4 Mass spectrometry materials

Product	Source
CHCA matrix	Sigma #70990
Mass spectrometry grade trypsin	Promega #V5111

3 Development of a method for trapping and analysis of carbamates

3.1 Chapter Introduction

3.1.1 A new approach to investigating protein carbamylation

Carbamates are labile groups which can make their discovery elusive and subsequent investigation difficult (Lorimer, 1983). Previous studies that have found proteins to form a functional carbamate modification have done so in one of two ways. The first of these is based on a hypothesis of the mode of action of the protein, which resulted in a search for a carbamate. Examples of this occurring are the initial discoveries of carbamylated haemoglobin, Rubisco and connexin 26 (Rossi-Bernardi and Roughton, 1967, Lorimer and Miziorko, 1980, Meigh et al., 2013). Alternatively some carbamates have been discovered as a by-product of experiments with different aims. An example of this is the discovery of the carbamylation of OXA 10 of the Class D β-Lactamases, which was found by analysis of the crystal structure of the protein with no prior hypothesis of carbamate formation (Maveyraud et al., 2000). Discoveries such as this one may lead to other proteins of the same family or similar structures also being found to form carbamates through intentional investigations into carbamylation. This is the case for other members of the Class D β-Lactamases and the related BlaR signal transducer protein (Leonard et al., 2013).

There is nothing wrong with the approach taken by any of the above studies; in each the function or structure of a particular protein was being investigated, which either confirmed the hypothesis of a carbamate or revealed the presence of one. However, the discovery of the important role of a carbamate group in the mechanism of Class D β-Lactamases only arose after analysis of the crystal structure of OXA-10. This indicates that there may be many cases of functional carbamate formation that we are yet unaware of. In addition not all carbamates may be discovered as readily as in the case of the Class D β-Lactamases, where the crystallisation conditions were favourable to carbamate formation. As carbamates are unstable it is less likely that they will be discovered by chance, as the experimental conditions would have to be favourable to their formation.

We decided to take an alternative approach to investigating carbamate formation, which was to focus first on identifying whether a protein forms a carbamate and then secondly how this may impact upon its function. It was hypothesised that this approach could be used to screen multiple proteins simultaneously for carbamate formation, potentially leading to the discovery of new cases of functionally important carbamates in a much more efficient manner than previously. To avoid the problem of carbamates dissociating from the protein and impacting the results, it would be necessary to first stabilise or trap the carbamate modifications. In this way proteins could then be analysed without the labile carbamates being lost during the experimental process.

3.1.2 A carbamate trapping method using triethyloxonium tetrafluoroborate

The trapping reagent used in these experiments is the ethylating agent triethyloxonium tetrafluoroborate, (TEO). It has been shown that TEO is capable of reacting with proteins in aqueous solutions despite the fact that it hydrolyses rapidly (King et al., 1986). Previous work carried out by Victoria Linthwaite (University of Durham, unpublished) had found TEO was capable of ethylating carbamylated amine groups, though free carboxyl and neutral amine groups on amino acids were also ethylated. Linthwaite developed a method to carbamylate amino acids by using sodium bicarbonate to provide CO_2 at a physiological concentration. The HCO_3^- then reacts with the N-terminus of the amino acid to form a carbamate. The product was analysed by electrospray mass spectrometry and the presence of ethylated carbamates proved that the TEO reagent was capable of trapping a carbamate as hypothesised.

3.1.3 Method development and optimisation

The initial experiments performed by Linthwaite had been used to develop and validate the trapping method using a simple system of individual amino acids that was easy to analyse. On this basis the first aim in this work was to validate the method on the amino acid phenylalanine and then further develop the method for use on a control protein known to form a functional carbamate. Once this was achieved it would be possible to optimise the method so that it could then be applied to other proteins. For the purposes of method development haemoglobin has been used to trial the trapping method on proteins.

The results obtained from method development and optimisation will be described and discussed. The analysis of the trapped carbamates was performed using mass spectrometry techniques. This was important as mass spectrometry enables many different proteins to be analysed simultaneously through proteomic techniques (Van Riper et al., 2013). This is essential as the ultimate goal of the project is to be able to identify carbamaylated proteins among a large mixture of proteins.

3.2 Establishing a method to control pH during trapping experiments

pH is a critical factor in influencing the outcome of an experiment. TEO hydrolyses rapidly to produce a strong acid, constantly lowering the pH of the experimental conditions in a trapping experiment. If the pH becomes too low during an experiment, for example below pH 6.0, then the amine group on the amino acid or protein will become protonated. Carbamates are unable to form on charged nitrogen atoms so this would prevent

carbamylation of the substrate. Before any trapping experiments could commence, therefore, it was necessary to devise a system to regulate pH during the trapping procedure, To counter the pH-altering effects of hydrolysed TEO a NaOH solution was added to the experiment to raise the pH. However, if the pH rises too high, for example above pH 8.5, there will be very little CO_2 in solution as it will mostly be present in the form of bicarbonate.

Therefore, a Titralab 854 pH Stat was used in order to allow the pH to be closely controlled in a trapping experiment. The pH Stat was set to maintain the pH of the solution at pH 7.4, at 25^{o}C, and programmed to add NaOH solution at a rate which maintained the required pH with the minimum amount of fluctuations possible. Figure 3.2.1A shows an example experiment of how the pH meter maintained the pH over a period of one hour when trapping a carbamate on the α-amino group of phenylalanine, whilst Figure 3.2.1B shows the volume of NaOH added over the period in the same experiment. Though the pH fluctuates at the start of the experiment the fluctuations are kept within a physiological pH range, and the pH becomes more stable over time as the TEO is hydrolysed (Fig 3.2.1A). Over half of the NaOH is added within the first 10 minutes of the experiment and by the end of the experiment no NaOH is being added (Fig 3.2.1B). This demonstrates that half of the TEO is being hydrolysed within the first ten minutes of the experiment and after an hour is completely hydrolysed. These results agree with published data which shows TEO has a half-life of about seven minutes in aqueous solution (King et al., 1986). This experiment was important to determine the length of further experiments as one hour. After this time the TEO would lose all ability to trap a carbamate due to hydrolysis. It is probable that most of the trapping is occurring when the TEO is first added and at its most reactive. It was therefore especially important to control the pH within the first ten minutes.

This ability to precisely control the pH was also essential for experiments using proteins rather than amino acids. Controlling pH is critical to maintain a stable physiological pH for proteins and to permit carbamate formation to be representative of that occurring in vivo.

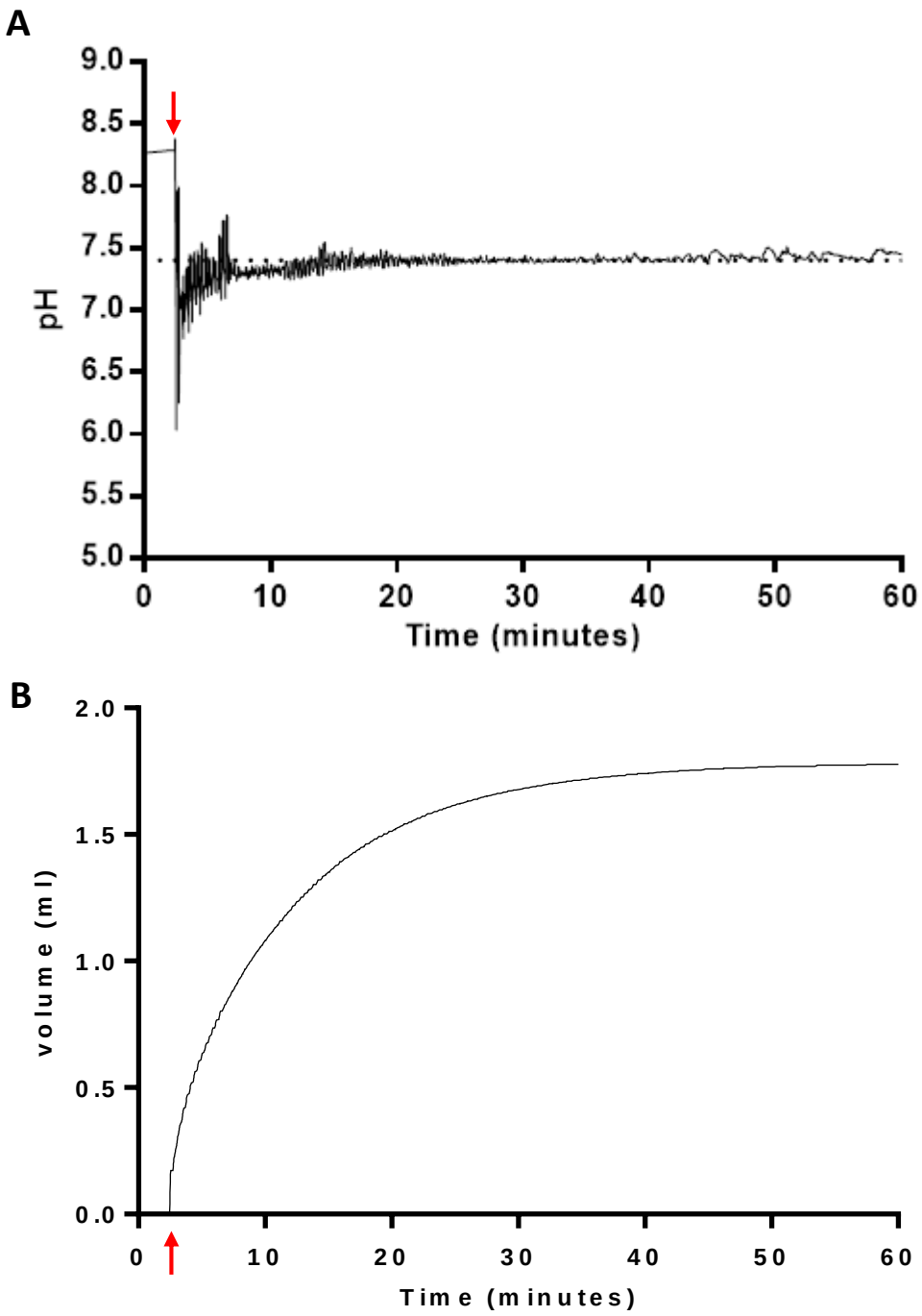


Figure 3.2.1: Regulation of pH during the carbamate trapping experiment

A pH Stat was used to regulate and monitor the pH during TEO trapping experiments. Addition of NaOH was used to neutralise the acidity caused by the hydrolysis of TEO during the experiment. This data presented here is from a trapping experiment on phenylalanine and is representative of all trapping experiments. **A)** The pH values recorded throughout the duration of an hour-long experiment. The arrow indicates the time point at which the TEO was added. The dotted line indicates pH 7.4 **B)** The volume of NaOH added throughout the duration of an hour long experiment. The arrow indicates the time point at which the TEO was added.

3.3 Validating the carbamate trapping method using phenylalanine

Once a method to control pH for trapping experiment had been developed (Section 3.2.1), it was possible to begin an analysis of carbamate trapping. The trapping method was first tested on the amino acid phenylalanine to ensure that the TEO reagent trapped a carbamate as hypothesised. Phenylalanine was chosen as it possesses a single amine group for CO_2 to form a carbamate, making analysis of the product unambiguous. The reaction mechanism of phenylalanine with CO_2 to form a carbamate, followed by the reaction of the carbamate with TEO, is shown in Figure 3.3.1A.

The reaction contained three principle components; phenylalanine, sodium bicarbonate and TEO, as described in the Methods (Section 2.1.1). A sufficient amount of phenylalanine was used in the reaction to ensure that ample product was available for downstream analysis by mass spectrometry. Sodium bicarbonate was added prior to the TEO reagent to provide CO_2 at a physiological concentration. This allowed carbamates to form at the amine group on phenylalanine prior to trapping. A precipitate formed at the end of the reaction. It was hypothesised that the precipitate was due to the ethylated groups causing the amino acid molecules to become hydrophobic and to aggregate. The product was concentrated by lyophilisation. A portion was dissolved in methanol, a suitable solvent for electrospray ionisation mass spectrometry.

The reaction product was analysed using by Liquid Chromatography (LC) coupled to an Electrospray Ionisation (LC-ESI) mass spectrometer. Sample products are first separated using an LC column, then converted into a fine spray of charged droplets, after which the solvent evaporates and the ionised sample is separated according to its m/z values (Ho et al., 2003). As electrospray is a relatively gentle form of ionisation as the sample molecules tend to stay intact, which aids analysis of the components of a sample.

There were several potential products of the experiment that could be hypothesized. Firstly it was possible that phenylalanine with just an ethylated carbamate would be produced, as shown in Figure 3.3.1A. However as TEO is also highly reactive with carboxyl groups, it was also likely that the carbamylated product formed would also contain an ethylated carboxyl group. In addition it was possible that not all of the phenylalanine would be carbamylated. Therefore it was hypothesized that phenylalanine ethylated on either the carboxyl or amine groups or on both would be present in the product. In addition, there was a probability that unmodified phenylalanine would be present in the product, though we hypothesized this was unlikely due to the excess of TEO used. It could be ascertained whether any of these compounds were present in the sample by calculating the expected m/z value of the potential ions of all these species.

Figure 3.3.1B shows a representative analysis from one of three repeats of the trapping experiment with phenylalanine. The same products were found in all three experimental repeats. The figure shows that three products formed from the experiment, represented by three peaks in the order in which the products were eluted from the LC column. The y-axis is not representative of the amount of product formed due to the differing ionisation potential of products. The first two peaks contain no carbamylated phenylalanine; the first peak (i) represents a mass corresponding to phenylalanine ethylated only on the carboxyl group or on the amine group. As TEO has been shown to react more strongly with ionised carboxylate groups than neutral carboxyl groups (Lundblad, 2014) it may also preferentially modify the carboxylate over the neutral amine group, so it is possible that peak (i) consists mainly of phenylalanine ethylated on the carboxyl group. However this hypothesis was not explored further. The second peak (ii) contains phenylalanine ethylated on both the carboxyl and amine groups. The third peak (iii) contains the carbamylated species. As expected the carboxyl group is also ethylated. It was therefore apparent that TEO was ethylating the carbamate as hypothesized, despite also showing reactivity with the carboxyl and non-carbamaylated amine group of the phenylalanine. As no starting product was observed the TEO appeared to be modifying all the phenylalanine present.

Although the sizes of the peaks cannot be used to calculate the amount of the product they represent, they can be used as an estimate. The peak containing the carbamaylated species appeared smaller than the others. This is likely because only a small proportion of the phenylalanine forms a carbamate at pH 7.4, as unlike in proteins the amine group is not in a privileged environment that can promote stabile carbamylation. Therefore the carbamate will be constantly in equilibrium with the charged amine group of phenylalanine, so at any one time only a small amount of carbamate will be trapped.

In summary, TEO was identified as a suitable trapping reagent for carbamates and experiments to adapt the method for work with protein could now proceed.

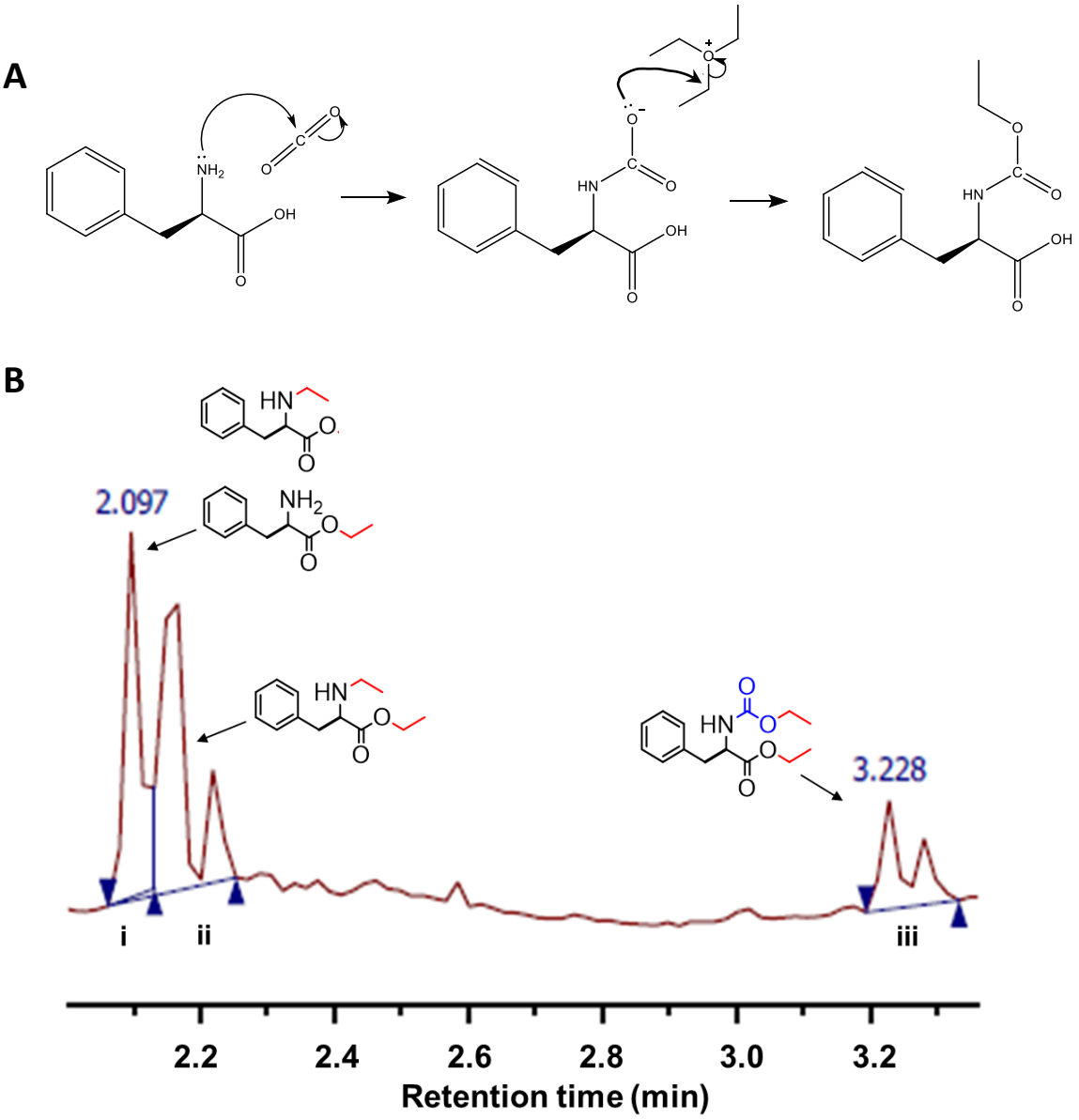


Figure 3.3.1: Carbamylation and ethylation of phenylalanine at pH 7.4

A) The reaction mechanism of CO_2 with the amine group of phenylalanine followed by ethylation of the resulting carbamate by TEO. **B)** A 0.12mM solution of phenylalanine containing 0.02M $NaHCO_3$ and an excess of TEO was maintained at a pH of 7.4 for one hour. The product was analysed by ESI-LC mass spectrometry. Letters and arrows indicate peaks: **i)** m/z =194.110, H^+ adduct, m/z corresponds to either one ethylation on the carboxyl group or one ethylation on the amine group **ii)** m/z=221.142, H^+ adduct, m/z corresponds to ethylations on the carboxyl and amine groups **iii)** m/z = 265.131, Na^+ adduct, m/z corresponding to an ethylated carbamate and an ethylation on the carboxyl group

3.4 Developing the trapping method for use on proteins

3.4.1 Trapping experiment using haemoglobin as the model protein

Once the trapping method had been shown to work with phenylalanine it was adapted for use to trap a carbamate on a protein. Haemoglobin was chosen for this development as it is widely known to trap carbamates on the N-terminal valine residues of both its α and β chains (Kilmartin and Rossi-Bernardi, 1971). It could therefore be used as a positive control to ensure that physiologically-relevant carbamates are being trapped and could be detected in our analysis.

Initial trapping experiments with haemoglobin used the same conditions that were used for the trapping of carbamates on phenylalanine (section 3.3). The addition of 50mM phosphate buffer was used both to maintain physiological pH and stabilise the protein in solution. Additional changes, described below, were also in the processing of the sample after carbamate trapping had been performed and also in the subsequent analysis of carbamate formation.

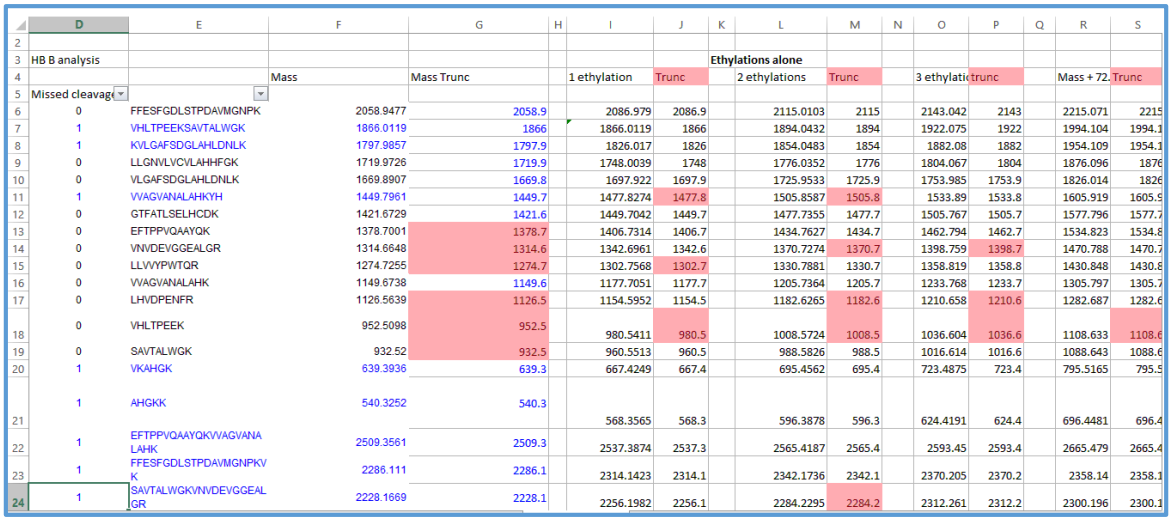
3.4.2 Processing the ethylated protein using proteolytic cleavage

No complex preparation of the phenylalanine sample had been needed prior to mass spectrometry analysis. It was necessary to analyse the trapped haemoglobin using mass spectrometry, but clearly a protein represents a much larger target for mass spectrometry than an amino acid. Further, it was desirable to locate the specific residue on which carbamylation was occurring, in this case the N-terminal valine residues of the α and β chains of haemoglobin. Therefore, after carbamate trapping had occurred, the haemoglobin was proteolytically cleaved into characteristic fragments for analysis by mass spectrometry. Trypsin was used for this purpose as it specifically cleaves protein at arginine and lysine residues (Olsen et al., 2004). Consequently it produces a reliable ‘peptide fingerprint’ which can be used to identify a protein by comparing the masses of the fragments in the experimental sample to a database of masses of known identity. If enough matches occur it can be reasonably deduced that the protein is present in the sample. This method is the basis of protein identification if it is among a large number other proteins in a proteomic experiment (Aebersold and Mann, 2003).

3.4.3 Analysis of the digested protein using MALDI mass spectrometry

Matrix-assisted laser desorption ionisation time of flight (MALDI-TOF) mass spectrometry was used to analyse trapped digested haemoglobin samples. This was because the MALDI technique is both more sensitive and less affected than other mass spectrometry methods by potential sample contaminants including salts and surfactants (Pan et al., 2007). Despite this, the sample was still dialysed at the end of the trapping experiment to remove as many salts as possible. After trypsin digestion the protein fragments were mixed with a matrix and spotted onto a MALDI plate to allow them to crystallise. Similar to electrospray mass spectrometry, MALDI is a gentle ionisation technique that tends to produce ions with a single charge, making it easy to identify the protein fragments by their masses. The ions are separated by Time of Flight (TOF) with ions with smaller m/z values travelling faster to reach the detector sooner. As the charge of all the ions is the same in MALDI the fragments are effectively separated by mass.

The experiment was intended to discover whether the peptide fragment of haemoglobin containing the N-terminal valine residue also contained a trapped carbamate. Any trapped carbamates and also the other ethylations that might occur on the protein would change the masses of the haemoglobin peptide fragments. A database of all the masses of the potential peptide fragments expected was therefore constructed. The database was built using the ExPASy PeptideMass tool to generate the theoretical masses of the peptide fragments of the haemoglobin α and β chains following trypsin digest. An Excel spreadsheet was then designed containing the mass of each of the known peptide fragments plus the mass of an ethyl carbamate (72.029 KDa) together with the masses of up to four additional ethylations (28.0313 KDa) . It was necessary to calculate the masses of all trapping outcomes on a peptide as the potential degree of peptide ethylation was unknown. In addition masses of peptide fragments that were ethylated up to four times but not ethyl carbamylated were also calculated, in order to understand the overall extent of ethylation of the protein. A snapshot of a portion of the Excel spreadsheet is shown in Figure 3.4.3.1. Any masses that were found in the experimental that matched with those calculated to contain one or more ethylations and/or a carbamate were highlighted in pink.



	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
2																
3	HB B analysis							Ethylations alone								
4			Mass	Mass Trunc		1 ethylation	Trunc		2 ethylations	Trunc		3 ethylatic	trunc		Mass + 72.	Trunc
5	Missed cleavage															
6	0	FFESFGDLSTPDAVMGNPK	2058.9477	2058.9		2086.979	2086.9		2115.0103	2115		2143.042	2143		2215.071	2215
7	1	VHLTPEEKSAVTALWGK	1866.0119	1866		1866.0119	1866		1894.0432	1894		1922.075	1922		1994.104	1994.1
8	1	KVLGAFSDGLAHLDNLK	1797.9857	1797.9		1826.017	1826		1854.0483	1854		1882.08	1882		1954.109	1954.1
9	0	LLGNVLVCVLAHHFGK	1719.9726	1719.9		1748.0039	1748		1776.0352	1776		1804.067	1804		1876.096	1876
10	0	VLGAFSDGLAHLDNLK	1669.8907	1669.8		1697.922	1697.9		1725.9533	1725.9		1753.985	1753.9		1826.014	1826
11	1	VVAGVANALAHKYH	1449.7961	1449.7		1477.8274	1477.8		1505.8587	1505.8		1533.89	1533.8		1605.919	1605.9
12	0	GTFATLSELHCDK	1421.6729	1421.6		1449.7042	1449.7		1477.7355	1477.7		1505.767	1505.7		1577.796	1577.7
13	0	EFTPPVQAAYQK	1378.7001	1378.7		1406.7314	1406.7		1434.7627	1434.7		1462.794	1462.7		1534.823	1534.8
14	0	VNVDEVGGEALGR	1314.6648	1314.6		1342.6961	1342.6		1370.7274	1370.7		1398.759	1398.7		1470.788	1470.7
15	0	LLVVYPWTQR	1274.7255	1274.7		1302.7568	1302.7		1330.7881	1330.7		1358.819	1358.8		1430.848	1430.8
16	0	VVAGVANALAHK	1149.6738	1149.6		1177.7051	1177.7		1205.7364	1205.7		1233.768	1233.7		1305.797	1305.7
17	0	LHVDPENFR	1126.5639	1126.5		1154.5952	1154.5		1182.6265	1182.6		1210.658	1210.6		1282.687	1282.6
18	0	VHLTPEEK	952.5098	952.5		980.5411	980.5		1008.5724	1008.5		1036.604	1036.6		1108.633	1108.6
19	0	SAVTALWGK	932.52	932.5		960.5513	960.5		988.5826	988.5		1016.614	1016.6		1088.643	1088.6
20	1	VKAHGK	639.3936	639.3		667.4249	667.4		695.4562	695.4		723.4875	723.4		795.5165	795.5
21	1	AHGKK	540.3252	540.3		568.3565	568.3		596.3878	596.3		624.4191	624.4		696.4481	696.4
22	1	EFTPPVQAAYQKVVAGVANALAHK	2509.3561	2509.3		2537.3874	2537.3		2565.4187	2565.4		2593.45	2593.4		2665.479	2665.4
23	1	FFESFGDLSTPDAVMGNPKVK	2286.111	2286.1		2314.1423	2314.1		2342.1736	2342.1		2370.205	2370.2		2358.14	2358.1
24	1	SAVTALWGKVNVDEVGGEALGR	2228.1669	2228.1		2256.1982	2256.1		2284.2295	2284.2		2312.261	2312.2		2300.196	2300.1

Figure 3.4.3.1: Database of trypsin cleavage fragment masses

After trapping carbamates on haemoglobin using the TEO method, the sample was proteolytically digested using trypsin to generate fragments of characteristic molecular masses, allowing for up to one mis-cleavage in a fragment. The masses of all the different potential fragments were calculated by adding the mass of up to four ethylations to each fragment, as well as the mass of one trapped carbamate in addition to four extra ethylations. The mass of a single ethylation was calculated as an addition of 28.0313, and the mass of a trapped carbamate as 72.029. These masses were then compared to the masses found during mass spectrometry analysis of the haemoglobin, any that matched are highlighted in pink.

3.4.4 Results of haemoglobin trapping experiments

Carbamate trapping experiments performed on haemoglobin revealed that the N-terminal valine residue was carbamylated. Experiments were repeated several times to confirm the findings and optimise the method. For clarity, the results discussed here are those from an experiment performed after all optimisation of the method and analysis had taken place. The main ways by which the method was optimised is summarised in the following section.

The overall mass spectrum of the haemoglobin protein fragments is shown in Figure 3.4.4.1A. The y-axis shows the relative intensity of the fragment ions present in the sample. In general the larger amount of an ion there is present in a sample the more intense the signal will be, however some fragments ionise more easily than others and so intensity is not necessarily an accurate measure of quantity. Clusters of peaks are evident and this effect is caused by differing amounts of ethylations from the TEO reagent on the same peptide fragment. For example one particular peptide fragment may be present with one, two and three ethyl groups attached, creating a cluster of ions that increase in mass in a stepwise fashion by the mass of an ethyl group (28.0313) each time. This pattern is here denoted an “ethylation series”. The ethylation series were useful in determining the reliability of a particular peptide fragment being present in the product sample. If the peptide fragment was observed as an ethylation series, we hypothesized it was more likely to be a true result than if it was just present with one ethyl group. This was because an m/z value corresponding to each additional ethylation on the same fragment effectively represents an additional confirmation that the fragment is present in the sample.

The peptide fragment containing the carbamate on the valine residue of the haemoglobin β chain was hypothesized to have the sequence, VHLTPEEK with a m/z value of 952.5098. The analysis database (Section 3.4.2) was searched for matches with the mass of this peptide fragment plus the mass of an ethylated carbamate group with a potential ethylation series. The data was also examined for the VHLTPEEK peptide fragment in its non carbamylated form. Figure 3.4.4.1B shows a magnified section of the whole spectrum in the appropriate m/z range. All the peaks corresponding to the mass of the VHLTPEEK fragment with its different modifications are highlighted in colour. Peaks shown in red are the ion masses that correspond to a trapped carbamate on the VHLTPEEK fragment with two, three and four additional ethylations. The VHLTPEEK fragment was also observed in a non-carbamylated form with multiple ethylations, which are indicated in blue. The data therefore provided strong evidence that the method had indeed trapped and stabilised the carbamate known to form on the N-terminal valine of haemoglobin. The trapped carbamate was found

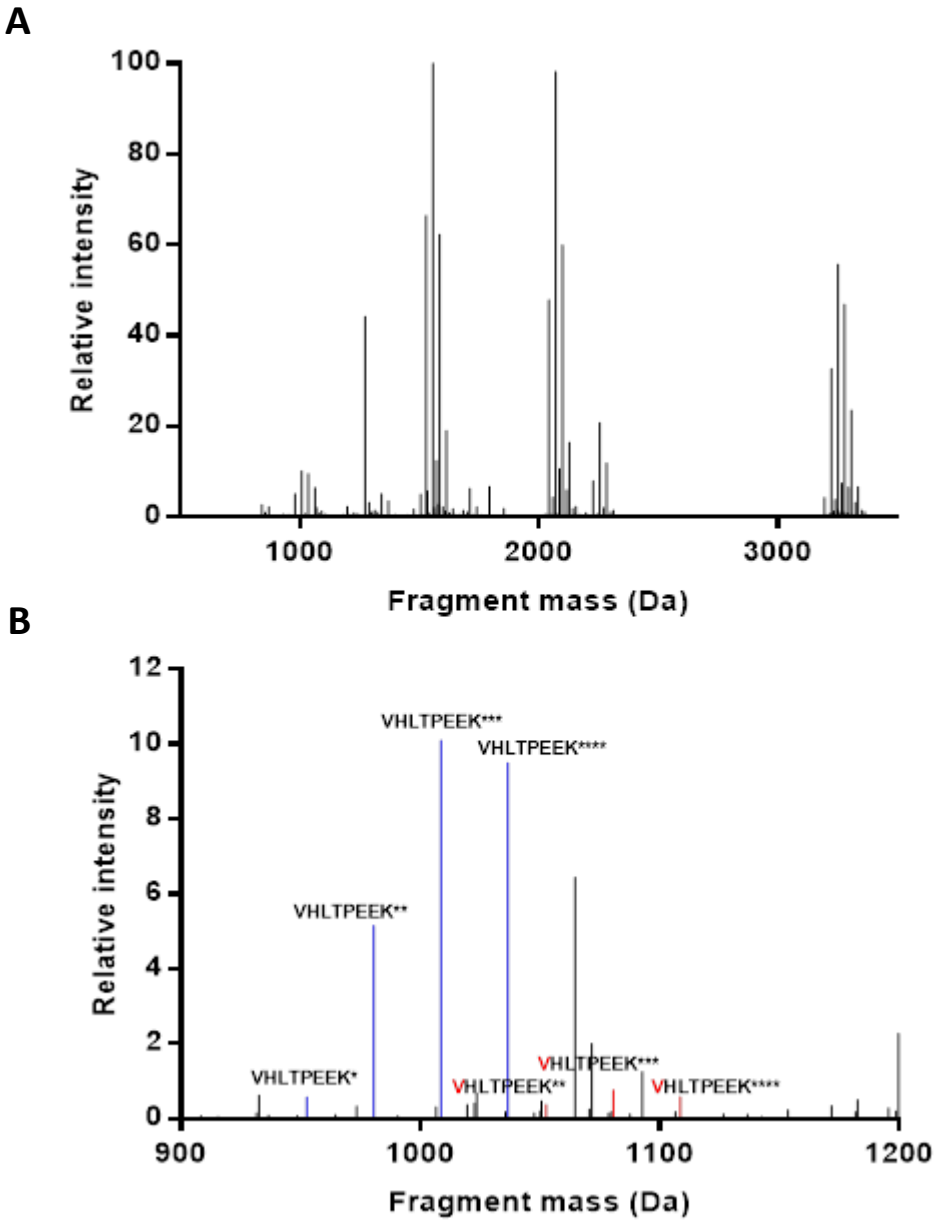


Figure 3.4.4.1: Identity of haemoglobin trypsin digest fragments using MALDI-TOF
The trypsin cleavage fragments of the haemoglobin were mixed with CHCA matrix and then analysed using Matrix assisted laser desorption ionisation time of flight (MALDI-TOF). The resulting fragment ions are displayed showing mass and intensity. **A)** The whole spectrum of haemoglobin fragment ions. **B)** A magnified section of the spectrum shown in A, showing the ion masses corresponding to the fragment of the haemoglobin β chain containing the N-terminal valine residue, VHLTPEEK. Ions with masses corresponding to the non-carbamylated VHLTPEEK fragment are shown in blue, with the number of ethylations indicated by the number of *. Ion masses corresponding to the fragment with a trapped valine carbamate are shown in red, again with additional ethylations indicated by *.

over at least ten different experiments, showing that the observation was reliable and repeatable. The peaks corresponding to the ions of the carbamaylated fragments are low in intensity relative to those of the non-carbamylated fragments. This could suggest that there was a relatively small amount of carbamate formed, however it is equally possible that the carbamylated fragments do not ionise as well, resulting in peaks with a smaller intensity. As the main aim of the experiment was to show that a trapped carbamate was present this was not a concern, although it was seen as a potential issue to address over time.

The data demonstrated that it was highly likely that the identity of the fragment hypothesized to be carbamylated was VHLTPEEK. However, this conclusion was based on the mass of the peptide fragment alone. It was formally possible that the masses observed were a coincidental match and actually belonged to other peptide species entirely. A method was therefore needed to confirm the identity of the peptide fragment beyond reasonable doubt. This was achieved using a technique coupled to the MALDI mass spectrometer called tandem mass spectrometry (MS/MS). This technique is able to split the target peptide fragment ion into smaller constituent ion fragments, called Y and B ions. Similarly to the procedure used to identify the trypsin fragment ions, the Y and B ions formed can be compared to the calculated values of all potential Y and B ions. Figure 3.4.4.2 shows the results when MS/MS was performed on the ion with a mass corresponding to the VHLTPEEK fragment with a trapped carbamate and four additional ethylations. The fragment masses with the top ten highest intensities are displayed. Of these five correspond to fragments of the VHLTPEEK ion with four of these containing the carbamaylated valine. As the presence of just three fragments ions is needed to confirm the identity of a protein fragment, it was concluded that the fragment had been correctly identified.

It can also be seen in Figure 3.4.4.1A that there are three main other collections of peaks corresponding to larger fragment masses. These represent series of ethylations on several other fragments apart from the carbamate forming VHLTPEEK fragment. This is to be expected as there are other carboxylate groups in the protein that will be in a suitable environment to undergo ethylation by TEO. It is likely that there are also many smaller intensity ethylation series which are not as visible, and that the higher intensities of some fragments are due to a combination of a high rate of trypsin cleavage and good fragment ionisation in the mass spectrometer. It has been shown previously through radioactivity studies that carbamates derived from cyanate may also be found at certain lysine residues in haemoglobin, albeit in relatively small quantities. (Lee and Manning, 1973). Cyanate was shown to compete with CO_2 for carbamylation sites, suggesting that CO_2 may also carbamylate several different residues in haemoglobin. However in this investigation no masses corresponding to other potential carbamylated fragments were found.

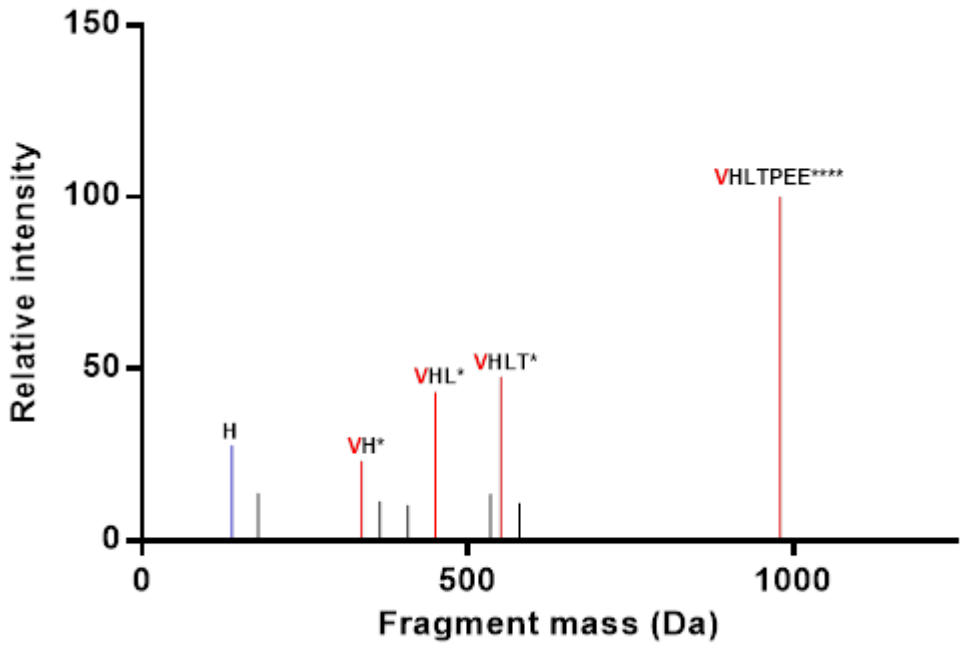


Figure 3.4.4.2: Confirmation of identity of VHLTPEEK fragment using MS/MS

MS/MS was used to confirm the identity of the carbamylated VHLTPEEK fragment, by splitting it into its constituent ions. The ion corresponding to the mass of VHLTPEEK with a trapped carbamate and four additional ethylations, shown in Fig 4.4.4.1B, was chosen for analysis. The resulting fragment ions with the ten highest intensities are shown here. Peaks indicated in colour represent ions belonging to the VHLTPEEK fragment, with the red peaks containing the carbamylated valine residue.

3.5 Optimisation of the carbamate trapping method

The trapping method was optimised to improve the quality of the results throughout the investigation. Although the method was sufficient to identify carbamates on haemoglobin it was necessary to improve the efficiency of the trapping and the sensitivity of the analysis as much as possible, in order to be able to use the method to identify carbamates on much smaller amounts of protein in future.

The different approaches taken to improve the method are summarised in Table 3.4. The main achievements made were to reduce the amount of haemoglobin used from 10 mg to 1 mg. This was important as it would be necessary to use the method on smaller quantities of protein in future. As haemoglobin contains two different chains the actual amount of each chain was 0.5 mg. With further optimisation the analysis should become even more sensitive.

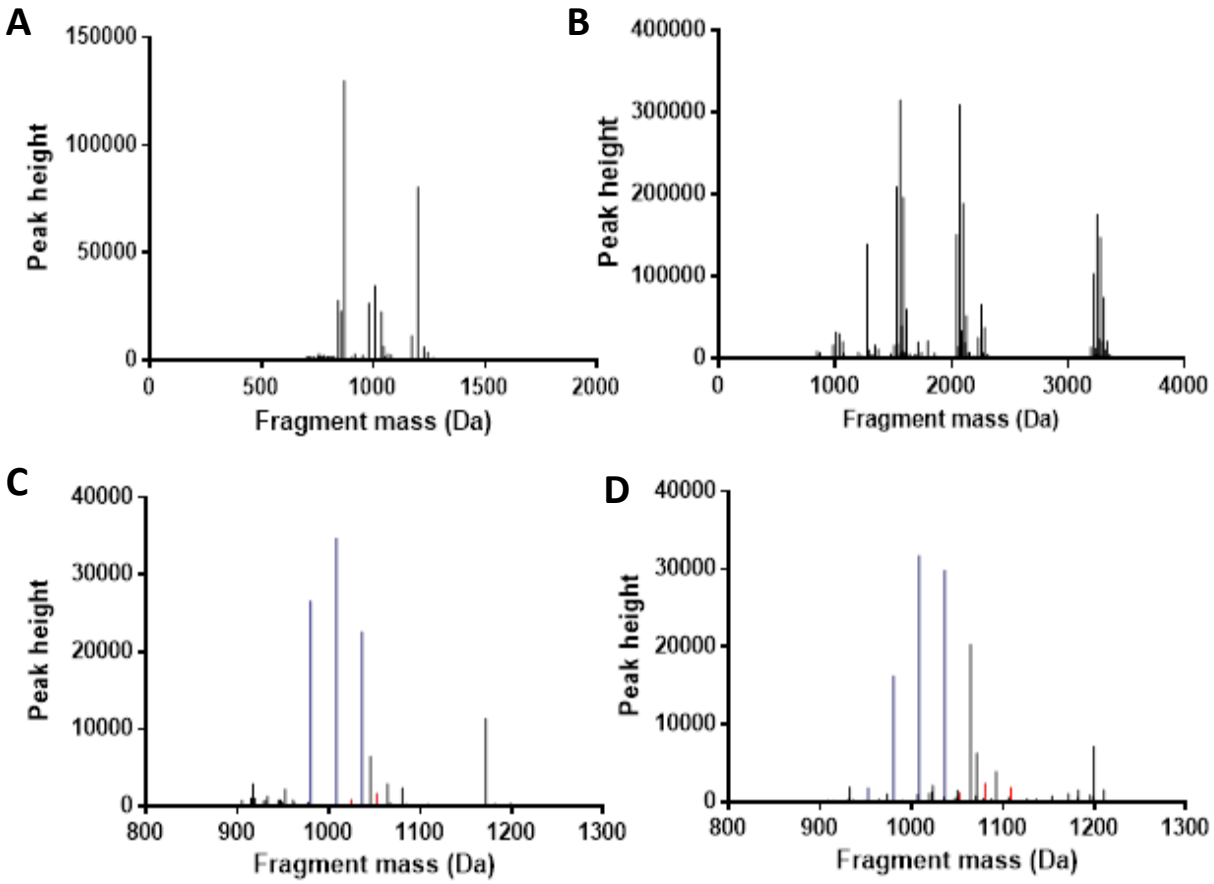
The main drawback of the method, which still remains an obstacle, is the precipitated protein that occurs as a result of the multiple ethylations by the TEO reagent making the protein highly hydrophobic. Though a long trypsin digestion was able to cleave the precipitated protein, cleavage would be much more efficient if the protein was in solution. Several methods were investigated to solubilise the protein including addition of SDS and heat treatment, however these had to be used under conditions compatible with trypsin digestion and mass spectrometry. These methods did not greatly improve solubility of the trapped protein pellet and SDS tended to interfere with mass spectrometry. However simply shaking the sample during trypsin digestion did aid digestion of the precipitate.

The parameters used to assess experiment improvement were the amount of protein fragments produced, amount of carbamylated fragments identified, and the intensity of the carbamylated fragments. The differences in results before and after the method optimisations shown in Table 3.5 are presented in Figure 3.5.1. It can be seen that although the amount of carbamylated fragments did not increase the intensity did. For analysis of results all intensities of fragments were considered as it was not known how well any particular fragment would ionise, or in what quantity a putative carbamylated fragment would be present. However although reproducibility was considered most important in terms of analysis, optimisation of the method to increase intensity was desirable as it was an indicator of increased efficiency of carbamate ethylation and/or protein digestion, and made it easier to perform MS/MS analysis. In addition, the overall number of protein fragments released by trypsin digestion increased. This increase will be important for future experiments where all the fragments will have to be investigated for carbamates.

Table 3.5.1: Summary of approaches taken to optimise the trapping experiment

Method optimisation	Rational	Result
Dialysis of trapped protein before trypsin digestion	TEO potentially interfering with trypsin digestion or mass spectrometry	MALDI results gave higher fragment intensity,
Different amounts of Hb	10 mg, 1 mg and 0.1 mg of Hb used to investigate minimum amount required for carbamate identification	1 mg was the minimum amount needed to identify a carbamate.
Different amounts of TEO	5, 50, 150, 250, 300, 400 and 500 mg of TEO were tested to see if using a different amount would improve carbamate identification.	250 mg of TEO gave optimal carbamate identification.
Concentration of trapped protein by spin column instead of freeze drying	Freeze drying is a slow process and wastes product. Concentrating is much less wasteful.	No protein lost through concentrating so results more representative. Faster method.
Use of CHCA matrix versus SA matrix	Different matrices crystallise differently and therefore affect fragmentation.	CHCA matrix produced more high quality fragments.
Zip tip sample vs non zip tip preparation for mass spectrometry	Zip tipping can remove unwanted salts and be used to concentrate sample	Zip tipping improved mass spectrometry data with less noise.
Investigation into methods to dissolve pelleted protein	If able to dissolve pellet trypsin digestion would be more efficient, leading to better mass spectrometry results.	Methods investigated not adopted as mass spectrometry results either not improved or made worse.
Length of trypsin digestion	A longer incubation with trypsin could improve sample data	An overnight trypsin digestion was adopted, which increased amount of protein fragments observed.

Shaking during trypsin digestion	Shaking precipitate during trypsin incubation could aid digestion	Shaking increased number of fragments identified.



E

	Before Optimisation	After Optimisation
Number of fragments	114	224
Number of all ethylation series	3	4
Number of fragments in carbamate ethylation series	3	3
Maximum intensity of carbamaylated fragment	1698	2409

Figure 3.5.1: Comparison of trapping method before and after optimisation

Examples of haemoglobin mass spectrum before and after the method optimisation. **A)** Complete mass spectrum of the carbamate trapped haemoglobin analysed by MALDI before any method optimisation had taken place. **B)** Complete mass spectrum of the carbamate trapped haemoglobin analysed by MALDI after all method optimisation had taken place. **C)** Section of the mass spectrum shown in A showing the VHLTPEEK fragment. Blue peaks represent the non-carbamylated fragment and red peaks the trapped carbamylated fragment. **D)** Section of the mass spectrum shown in B showing the VHLTPEEK fragment. Blue peaks represent the non-carbamylated fragment and red peaks the trapped carbamylated fragment. **E)** Table comparing result parameters before and after experiment optimisation.

3.6 Chapter summary

It has been shown here that a general method has been developed to trap a carbamate modification onto haemoglobin by ethylating the carbamate using TEO as an ethylating agent. It is possible to analyse the ethylated protein and identify the trapped carbamate by MALDI and electrospray mass spectrometry. The method has been optimised to identify the optimum conditions for the experiments, although further optimisation may be required in future for different proteins and or different experimental formats.

Therefore the method can now be used to investigate whether other proteins that are regulated by CO_2 also form carbamate modifications, as will be discussed in the next chapter.

4 Investigating carbamate formation on Class III adenylyl cyclases

4.1 Chapter Introduction

The carbamate trapping method described in Section 3.0 had been shown to work on both haemoglobin and Rubisco. Therefore the method could now be used to explore carbamate formation on proteins not previously known to bind CO_2 as a carbamate. Class III adenylyl cyclases (ACs) were investigated as they have previously been shown to be responsive to CO_2 and it was hypothesised that this could be through the formation of a carbamate.

4.1.1 Regulation of Class III adenylyl cyclases by CO2

Adenylyl Cyclases (ACs) are found in both prokaryotes and eukaryotes and catalase the formation of the second messenger cyclic adenosine 3′,5′-monophosphate (cAMP) from ATP. They are therefore an important component of numerous signalling pathways regulating a wide variety of physiological processes (Steegborn, 2014, Sassone-Corsi, 2012).

There are six classes of AC. Class I ACs comprise those found in gram negative bacteria such as *E-coli*, whilst Class 2 ACs are those belonging to pathogens such as Pseudomonas aeruginosa and Bordetella pertussis, and are secreted like toxins into the host organism (McDonough and Rodriguez, 2012, Linder, 2008). All known eukaryotic and a large amount of prokaryotic ACs belong to Class III, which is the largest class. (Danchin, 1993, Steegborn, 2014). The remaining classes IV-VI were more recently discovered and contain as little as one AC per class (Steegborn, 2014).

Mammals possess ten identified class III ACs, nine of which are transmembrane adenylyl cyclases (tmAC) and one which is a soluble adenylyl cyclase (sAC) (Linder, 2006). Following the discovery that the mammalian sAC could be activated by bicarbonate (Chen et al., 2000), the abilities of other class III adenylyl cyclases to respond to inorganic carbon were investigated. ACs from several prokaryotic species were also shown to be regulated by inorganic carbon but initially this was not thought to be the case for mammalian tmACs (Kamenetsky et al., 2006)

Initially there were no attempts to separate the inorganic carbon into CO_2 or HCO_3^- so it was uncertain which species was regulating the ACs, though it was generally assumed to be bicarbonate due to the hypothesis that the ionic molecule was more likely to bind in the active site than CO_2 (Steegborn et al., 2005, Raven, 2006). However a study that did investigate the individual effects of CO_2 and HCO_3^- found that two prokaryotic class III adenylyl cyclases form cyanobacteria, Slr1991 from *Synechocystis* PCC 6803 and CyaB1 from *Anabaena* PCC 7120, responded to CO_2 but not HCO_3^- (Hammer et al., 2006).

On this basis the mammalian tmACs were again investigated to see if they were responsive to CO_2 and/or bicarbonate. It was shown that both a mammalian tmAC and the related Rv1625c AC of *Mycobacterium tuberculosis* were stimulated by CO_2 but not bicarbonate ions, an effect which occurred independently of pH (Townsend et al., 2009). On the other hand sAC was found to respond both to CO_2 and HCO_3. Additionally it was shown that radioactively labelled CO_2 was able to bind to the catalytic region of the Rv1625c AC from *Mycobacterium tuberculosis* independently of any cofactors, suggesting that CO_2 could be activating it by directly binding to the protein (Townsend et al., 2009).

However there is some uncertainty over whether tmACs can be regulated by inorganic carbon alone under physiological conditions in vivo. Experiments investigating the response of both sAC and tmAC in the carotid body found that neither of the adenylyl cyclases was able to functionally respond to inorganic carbon independently of pH (Nunes et al., 2013). As the concentration of bicarbonate/CO_2 used exceeded physiological concentrations it is possible that any response pathway had lost its sensitivity to the inorganic carbon, though this seems unlikely as Townsend et al also used a higher concentration of CO_2 and still saw an increase in cAMP production, though not above that induced by physiological CO_2 concentrations (Townsend et al., 2009) It is possible that the ACs of the carotid body are regulated by pH alone whereas ACs in other locations are additionally regulated by bicarbonate and/or CO_2. Another study investigating the response of mammalian G protein responsive AC isoforms by measuring cAMP output in three different cell types found a reduced response at elevated CO_2 concentrations of 10% compared to 'normal concentrations' of 5% (Cook et al., 2012). This was interesting as previously it was shown that CO_2 stimulation increased cAMP production (Townsend et al., 2009). The reason for this discrepancy is that the CO_2 dependent decrease in cAMP output was reliant on an active inositol 1,4,5-triphosphate (IP_3 receptor), a protein that regulates calcium release from the Endoplasmic reticulum (Cook et al., 2012, Mikoshiba, 2007). This would not have been accounted for in the study by Townsend et al.

Therefore although it is accepted that sAC is regulated by bicarbonate, with the possibility of being regulated in addition by CO_2, the response of the other Class III adenylyl cyclases to CO_2 is a complex one which is still not fully understood. However in two out of three key studies a response of adenylyl cyclase to CO_2 in terms of cAMP output was observed, though it is now known that cAMP amounts are reduced rather than decreased in vivo (Townsend et al., 2009, Cook et al., 2012). In addition the work by Townsend et al showed that CO_2 appeared to be binding to the Rv1625c AC, further supporting the hypothesis that CO_2 may be directly influencing the function of tmACs (Townsend et al., 2009). The final study was in the specific location of the carotid body where it is possible that the effects of

CO_2 are different. In addition the authors were expecting levels of cAMP to increase in response to CO_2 so may have overlooked its potential suppression (Nunes et al., 2013). Overall it appears that CO_2 does have a pH independent role in regulating Class III ACs involving binding of the CO_2 to the protein, but further work is required to enable better understanding of the mechanisms involved.

4.1.2 Investigating the ability of Class III ACs to form carbamate modifications

As shown in section 4.1.1 there is a need for further work into firstly, the precise impact CO_2 has on tmACs, and secondly whether it directly binds to them to influence their functions. To help provide an answer to the second question it was decided to use the carbamate trapping method to explore whether CO_2 was forming a carbamate in the catalytic domains of two ACs; a mammalian tmAC and the related Rv1625c AC from *Mycobacterium tuberculosis*.

The mammalian tmAC consists of two transmembrane domains each containing six membrane spanning segments, and two cytoplasmic domains known as C1 and C2 which form a heterodimer. The C1 and C2 domains each contain a region of highly conserved catalytic residues and are termed C1a and C2a respectively (Whisnant et al., 1996, Tesmer et al., 1997).

Rv1625c is a class III adenylyl cyclase which a useful model of the mammalian tmAC, as it corresponds to exactly one half of the protein, consisting of one transmembrane domain of six repeats and one cytoplasmic catalytic domain (Reddy et al., 2001).

The original aim was to investigate the ability of the mammalian tmAC to form a carbamate, but when results proved inconclusive the related Rv1625c protein was also tested. Both of the proteins have lysine and arginine residues in their catalytic domains with an amine side chain which could potentially bind CO_2 as a carbamate. The residues of K938 and R1079 in the in the C2 domain of tmAC and the substrate specifying K296 in Rv1625c were of particular interest as these are all conserved residues across different classes of AC known to be required for their catalytic activity and function, and as they all have an amine side chain could be capable of binding a carbamate (Guo et al., 2001, Liu et al., 1997).

In order to avoid problems with purifying the membrane domains of the proteins, only the catalytic domains were expressed and used in experiments. These domains have previously been shown to retain the normal activity of the native proteins and were hypothesised to be the most likely to have functional carbamate forming residues, so were suitable for our experiments (Whisnant et al., 1996, Townsend et al., 2009).

4.2 Expression and Purification of Mammalian tmAC C_1 and C_2 domains

Two recombinant proteins comprising the catalytic regions of the mammalian transmembrane AC were expressed and purified separately. The C_1 domain was derived from the human type 7 AC and the C_2 domain from the rat type 2 AC. It has been previously shown that combining these two catalytic domains creates a functional AC (Townsend et al., 2009). Both the $7C_1$ and $2C_2$ recombinant proteins were overexpressed in *E. coli* BL21 (DE3). The hexahistidine tagged protein was then purified using Ni^{2+} affinity chromatography. The protein was eluted from the affinity resin using a buffer containing imidazole which competes with the hexahistidine tag for binding to the affinity resin. Eluted protein fractions were run on a 15 % (w/v) SDS-PAGE gel for analysis of purity, including samples from the wash steps and the original cell lysates before and after incubation with resin.

A representative gel of the purification of the $7C_1$ domain is shown in Figure 4.2.1A. It can be seen that the wash steps were effective in removing the unspecifically bound protein and that the majority of the $7C_1$ protein appeared to be eluted in the first elution fraction. The same results were seen for the $2C_2$ domain. Equal volumes of the purified protein from the first elutions of the tmAC $7C_1$ and tmAC $2C_2$ domains was also loaded onto a separate SDS-PAGE gel. The majority of each protein migrated at a molecular weights of ~25 KDa and ~27 KDa for the C_1 and C_2 domains respectively. This was surprising given that the expected molecular weights of each protein were 42.8 KDa and 31.8 KDa respectively. However when the identity of each protein was analysed by peptide fingerprinting using the MASCOT database it was confirmed that the identities of each of the proteins was correct with over 95% confidence in the resulting scores.

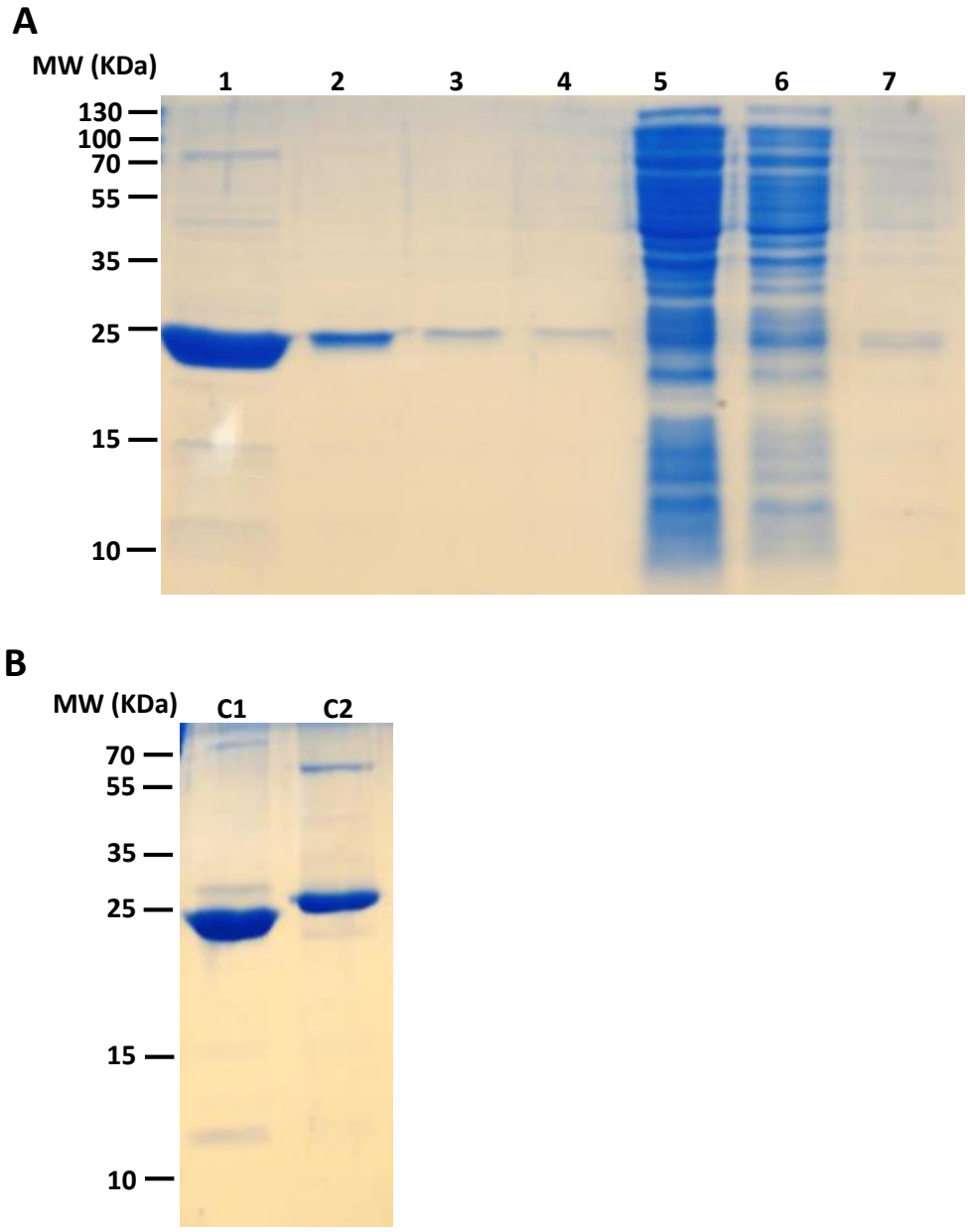


Figure 4.2.1: Purification of tmAC C1 and C2 domains

A) Representative SDS-PAGE gel of different stages of the purification of tmAC C_1 domain from BL21 (DE3) *E. coli* lysate. Purification of the tmAC C_2 domain was carried out in an identical manner. All protein is stained with Coomassie blue and the molecular weight standards are indicated on the left hand side of the gel. Lanes 1-4) Separate fractions after eluting protein with imidazole, with each fraction eluted using 1ml of elution buffer. The molecular weight of tmAC C_1 is 26.2KDa, and the vast majority of protein is seen at ~25KDa. Lane 5) Whole cell lysate before incubation with resin Lane 6) Whole cell lysate after incubation with Nickel resin, pre wash. Lane 7) Nickel resin column flow through after three washes.

B) SDS-PAGE gel of purified tmAC C_1 and C_2 from the first elution steps, with equal volumes of each elution loaded onto the gel. C1= tmAC C_1 domain, C2 = tmAC C_2 domain

4.3 Expression and Purification of Rv1625c$_{204-443}$

The cytoplasmic catalytic domain of Rv1625c was expressed as a recombinant protein, as it has been shown previously that this this region is adequate for normal activity of the protein (Townsend et al., 2009) Rv1625c$_{204-443}$ was overexpressed in *E-coli* M15(DE3). The hexahistidine tagged protein was then purified using Ni^{2+} affinity chromatography. The protein was eluted from the affinity resin using a buffer containing imidazole which competes with the hexahistidine tag for binding to the affinity resin. Eluted protein fractions were run on a 15 % (w/v) SDS-PAGE gel for analysis of purity, including samples from the wash steps and the original cell lysates before and after incubation with resin (Figure 4.3.1).

The majority of the eluted protein migrated at a molecular weight of ~25KDa which has a predicted molecular weight of 26.3KDa. Therefore it was likely the large band at ~25KDa was the expressed Rv1625c$_{204-443}$ protein. Running samples from the wash steps on the gel also showed that the wash methods used during purification were successful in removing the majority of proteins bound unspecifically to the resin (Figure 4.3.1). It was estimated that the protein was at least 90% pure which was deemed acceptable for use in the experiments. The second and sixth elutions shown in Lanes 7 and 11 respectively on figure 4.3.1A were also analysed by western blot, using an anti-His tag antibody (Figure 4.3.2). The tmAC C1 domain protein was also loaded as a positive control for the His tag. A single band was seen at ~26KDa in both samples with the second elution containing a lot more protein than the sixth elution as expected. This confirmed that the His-tagged protein was present in the sample. Finally MASCOT analysis of a MALDI mass spectrometry experiment following trypsin digestion of Rv1625c$_{204-443}$ identified the protein from its peptide fingerprint, with the highest match score of all possible proteins and greater than 95% confidence. This confirmed the protein was the main one present in the sample and was suitable for use in the trapping experiments.

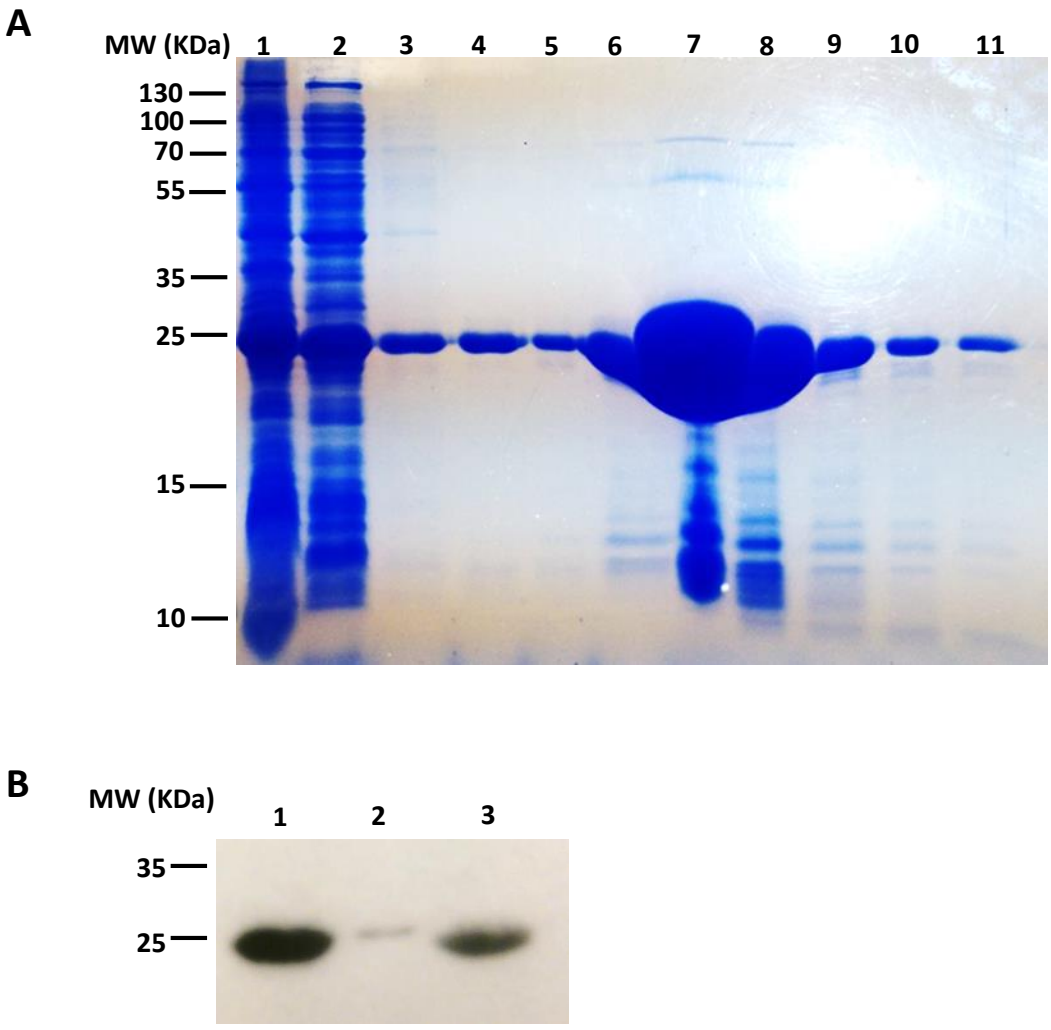


Figure 4.3.1: Purification of Rv1625c$_{204-443}$

A) SDS-PAGE gel of different stages of the purification of RV1625c$_{204-443}$ protein from M15 (DE3) *E. coli* lysate. All protein is stained with Coomassie blue and the molecular weight standards are indicated on the left hand side of the gel. Lane 1) Whole cell lysate before incubation with resin Lane 2) Whole cell lysate after incubation with Nickel resin, pre wash. Lanes 3-5) Nickel resin column flow through after three respective washes. Lanes 6-11) Separate fractions after eluting protein with imidazole, with each fraction eluted using 1ml of elution buffer per fraction. The molecular weight of RV1625c$_{204-443}$ is 26.2KDa, and the vast majority of protein is seen at ~25KDa.

B) Western blot against His tagged protein using HRP conjugated His-probe. Lane 1: Elution 2 of RV1625c$_{204-443}$ protein shown in Figure 4.3.1A, Lane 2: Elution 11 of RV1625c$_{204-443}$ protein shown in Figure 4.3.1B, Lane 3: positive control (tmAC C_1 domain).

4.4 Investigating carbamylation of tmAC catalytic domains using MALDI

The trapping and mass spectrometry methods for analysis for carbamates developed using haemoglobin (Section 3.0) were applied to the tmAC C_1 and C_2 domains to see if either formed any carbamate modifications. Both the C_1 and C_2 domains are required for a functional protein and have previously been shown to spontaneously associate to form an active catalytic region (Townsend et al., 2009). Therefore the domains which were expressed and purified separately as detailed in section 4.2 were mixed together immediately prior to the start of the experiment, to allow for the possibility that the domains may need to be associated for a carbamate modification to form. $NaHCO_3$ was then added to the associated catalytic domains to provide a physiological concentration of CO_2 which would be able to react with any neutral amine groups with the potential to form a carbamate. TEO was added immediately after to allow any carbamates to be ethylated and stabilised. The resulting product was then trypsin digested and the fragments analysed by MALDI mass spectrometry, as described in the methods section (2.1.8).

Unlike with the haemoglobin experiments it was not known upon which residue(s) carbamylation may occur, although it was hypothesised that they could be those previously shown to be important in regulating the activity of the enzyme (Liu et al., 1997, Reddy et al., 2001). Therefore all of the resulting m/z values from the MALDI experiment were compared to all off the theoretical carbamylated and ethylated fragment masses, using a database of the same structure as that used in the haemoglobin experiments. Different masses representing different amounts of ethyl groups on the same fragment could be used as an indicator that that fragment was indeed present and the m/z value not a random occurrence. The experiment was performed using both $NaH^{12}CO_3$ and $NaH^{13}CO_3$. This was to ensure that any fragment masses that appeared to comprise the mass of a carbamate modification with one or more ethylations were true results and not just random masses. If the mass of a potential hit appeared in the $^{12}CO_2$ experiment then if a real result it was expected to also appear in the $^{13}CO_2$ experiment with an additional mass of +1. The fragment could then be analysed by MS/MS to confirm its identity and the location of the carbamate.

Representative images of complete mass spectra from both a $^{12}CO_2$ and a $^{13}CO_2$ experiment are shown in Figure 4.4.1. Both the $^{12}CO_2$ and $^{13}CO_2$ experiment experiments were performed at least two times. In theory the two spectra should be very similar in appearance, however it can be seen that they appear quite different. When analysed more closely it was apparent that many of the same molecular masses were present in the $^{12}CO_2$ and a $^{13}CO_2$ experiments but that the intensities varied greatly. In addition there appeared to be either a

larger number of high molecular weight fragments or of low molecular weight fragments within an experiment, but not an even mixture of both. For example in the spectrum shown in Figure 4.4.1A there are more peaks representing fragments with lower molecular weights and these are also more intense than in the $^{13}CO_2$ spectrum shown in Figure 4.4.1B, where the opposite is true.

This difference between experiments could occur for several reasons. Firstly the extent of ethylation itself may vary between different experiments as it is a largely variable process producing a differently modified protein each time. Depending on the extent of ethylation the ability of the trypsin to cleave the protein may also be affected differentially between experiments, producing different cleavage fragments. The fact that there are more high molecular weight fragments in the $^{12}CO_2$ spectrum shown in Figure 4.4.1A could have been because this sample was trypsin digested to a lesser extent than the sample which produced the $^{13}CO_2$ spectrum in Figure 4.4.1B. Though the method used was the same for both experiments the fact that the trypsin must digest a precipitated protein will inevitably make the results less reproducible. In addition there are many factors affecting the MALDI mass spectrometry experiment itself that could lead to variable results, for example how well the matrix crystallises or how well and reproducibly each fragment ionises.

Consequently each individual $^{13}CO_2$ and $^{12}CO_2$ was repeated at least twice, resulting in the carbamate trapping procedure being repeated five times overall. It was hypothesised that this should have been adequate to reveal any trapped carbamates. In one $^{12}CO_2$ experiment masses were detected which corresponded to the fragment VGIHSGNVLCGVIGLR from the $tmACC_1$ domain containing a carbamate with up to four additional ethylations (Figure 4.4.2). However the masses did not appear in any of the other $^{12}CO_2$ experiemtns and the corresponding masses did not appear in any of the $^{13}CO_2$ experiments. It was also surprising that the m/z value for the uncarbamaylated fragment did not appear in either its unmodified or ethylated forms. Therefore this m.z value was not analysed further as it was likely an artefact of a single experiment and the result could not be reproduced.

As no other matches were found between the $^{13}CO_2$ and $^{12}CO_2$ expriements for either the tmAC C1 or C_2 domains it was decided to attempt the experiment on the $Rv1625_C$ catalytic domain. This was to attempt to confirm whether carbamate formation was occuring on a structurally similar adenylyl cyclase, and as it has previously been suggested that CO_2 can bind directly to this protein it was hypothesised that this could be occuring via a carbamate modification (Townsend et al., 2009).

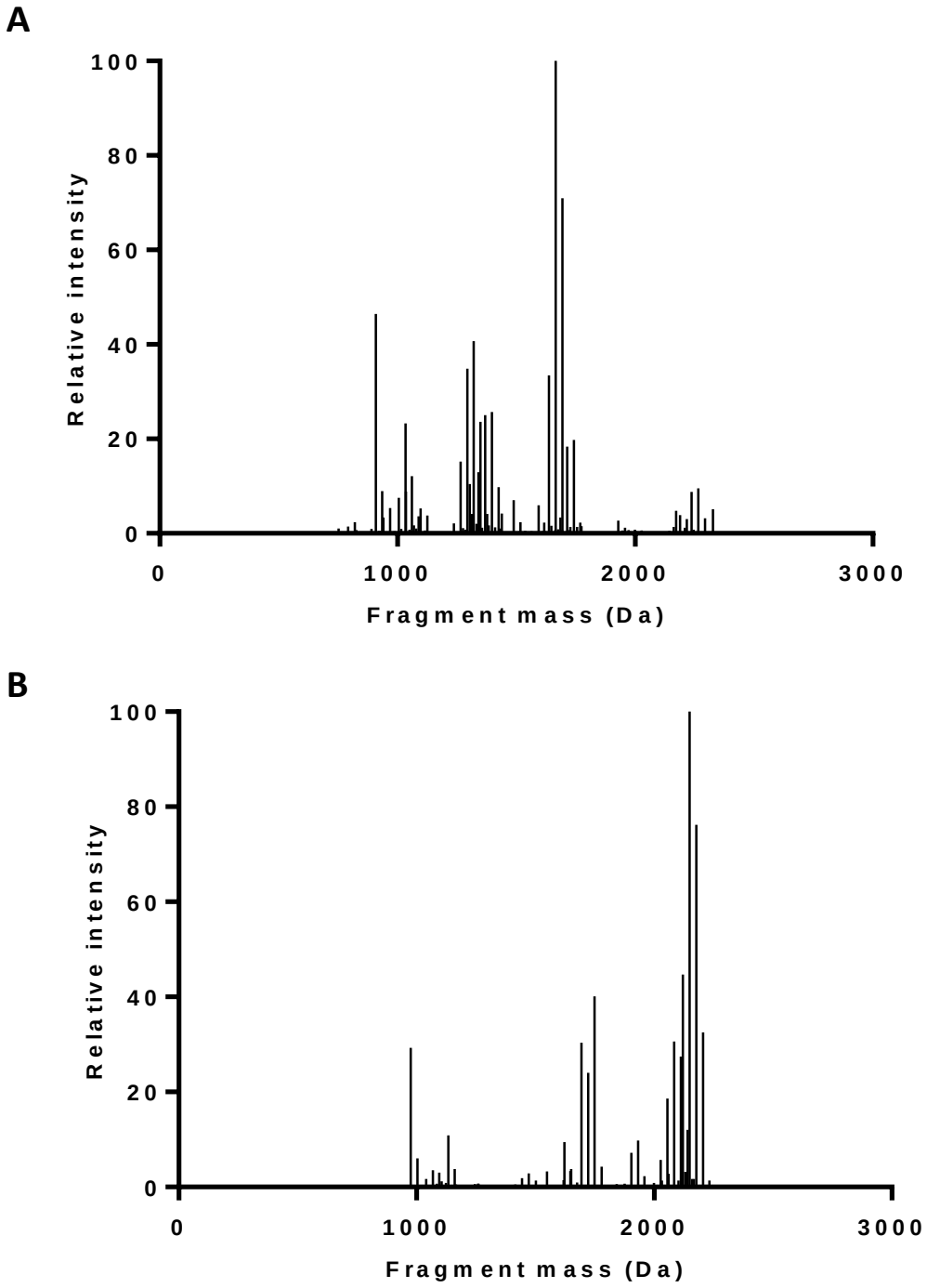


Figure 4.4.1: Representative MALDI mass spectra of tmAC C1 and C2 domains

A) MALDI Mass spectrum of trypsin cleavage fragments of ethylated tmAC C_1 and C_2 domains Equal amounts of each of the domains were combined in a solution containing physiological concentrations of CO_2 and treated with TEO to ethylate and trap any potential carbamates. The products were trypsin digested and analysed by MALDI mass spectrometry.

B) MALDI Mass spectrum of trypsin cleavage fragments of ethylated tmAC C_1 and C_2 domains using same experimental procedure as for (A) but using $^{13}CO_2$ instead of $^{12}CO_2$.

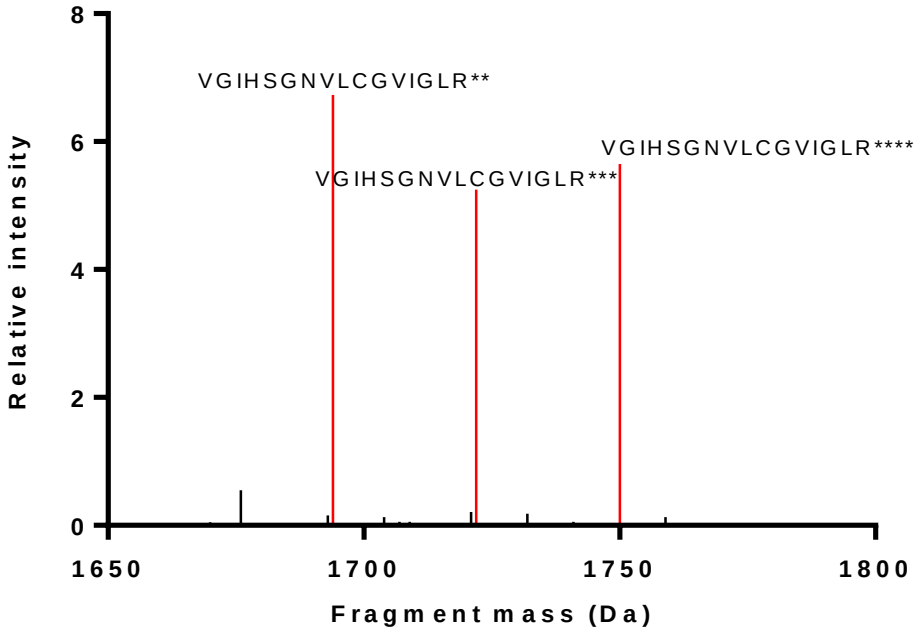


Figure 4.4.2: MALDI mass spectra of tmAC C1 and C2 domains

The MALDI Mass spectrum shown in Figure 5.4.1A was magnified to show m/z values including those corresponding to the mass of the VGIHSGNVLCFVIGLR fragment of the tmAC C_1 domain plus the mass of an ethylated carbamate and up to four additional ethylations. The carbamylated fragments are indicated by the red peaks. The * symbol denotes an ethyl group.

4.5 Investigating carbamylation of Rv1625c catalytic domain using MALDI

It was decided to investigate the carbamate forming capabilities of the Rv1625c adenylyl cyclase as a related alternative adenylyl cyclase to the tmAC protein. Again only the catalytic site was expressed as this was considered the most likely region in which a carbamate would form. Specifically it was hypothesised that the K296 residue was a key candidate for carbamate formation. This was due to the lysine reside being conserved among adenylyl cyclases for playing an important role in ensuring substrate specificity during catalysis (Guo et al., 2001). The advantage of analysing two adenylyl cyclases was that as well as providing additional information about each protein, a carbamate modification that could be harder to detect on one adenylyl cyclase may be easier to detect on another. For example the carbamate forming residues in one protein may be in a less accessible environment for TEO ethylation than in another protein. In particular the K296 residue of the Rv1625c protein corresponds to the K938 residue of the tmAC C2 subunit (Guo et al., 2001), so using Rv1625c was an additional opportunity for investigation of this key lysine residue. As explained earlier only the catalytic domain of Rv1625c was expressed, so the recombinant protein used was the $Rv1625c_{204-443}$ protein.

The $Rv1625c_{204-443}$ protein was expressed as described in section 4.3. The sample preparation and trypsin digestion was performed in the same way as for the tmAC catalytic domains, and the protein analysed by MALDI mass spectrometry. As with the tmAC protein a database was created containing the masses of all the potential cleavage fragments of the $Rv1625c_{204-443}$ protein in addition to the mass of an ethylated carbamate and up to four additional ethylations. At least two experiments were performed where $^{12}CO_2$ was provided as a source for carbamates and at least two experiments where $^{13}CO_2$ was provided.

Representative mass spectra from a $^{12}CO_2$ and a $^{13}CO_2$ experiment are provided in Figures 4.5.1A and 4.5.1B respectively. The spectra should in theory look very similar but as with the tmAC experiments they tended to vary quite widely between different experiments, hence the need for multiple repeats. The reasons for variation are likely the same as those for the variation between the tmAC spectra discussed in section 4.4.

M/Z values corresponding to the mass of the ethyl carbamylated fragment LYSAFDELVDQHGLEKIK with up to three additional ethylations were discovered in one $^{12}CO_2$ experiment (Figure 4.5.2). This at first appeared promising as the last Lysine residue in the sequence was the K296 residue hypothesised to be capable of forming a carbamate. However it was not possible to detect the m/z value of the fragment in any of the other $^{12}CO_2$

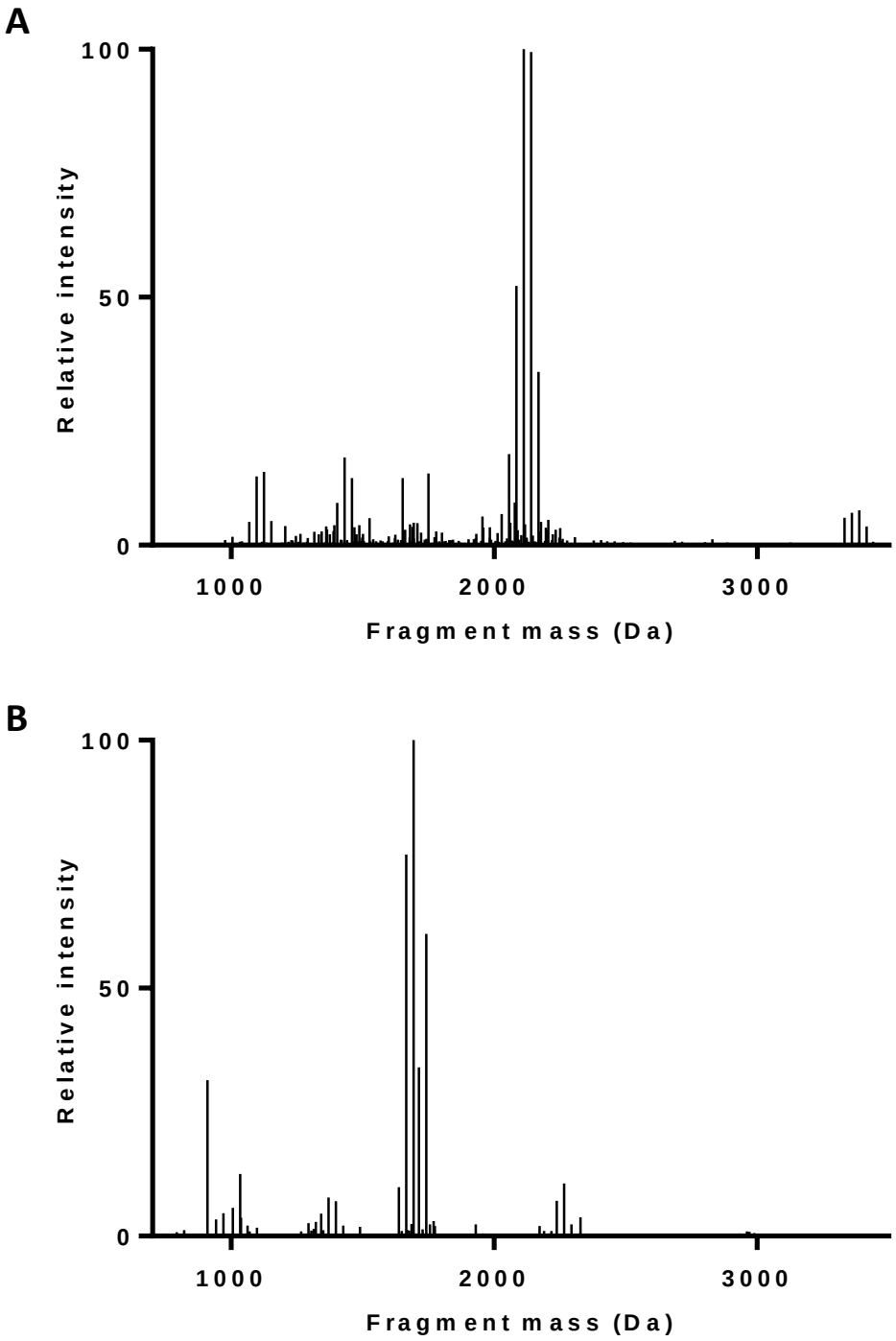


Figure 4.5.1: Representative MALDI mass spectra of Rv1625c$_{204-443}$

A) MALDI Mass spectrum of trypsin cleavage fragments of ethylated Rv1625c$_{204-443}$. The protein was added to a solution containing physiological concentrations of CO_2 and treated with TEO to ethylate and trap any potential carbamates. The products were trypsin digested and analysed by MALDI mass spectrometry.

B) MALDI Mass spectrum of trypsin cleavage fragments of ethylated Rv1625c$_{204-443}$ using same experimental procedure as for (A) but using $^{13}CO_2$ instead of $^{12}CO_2$.

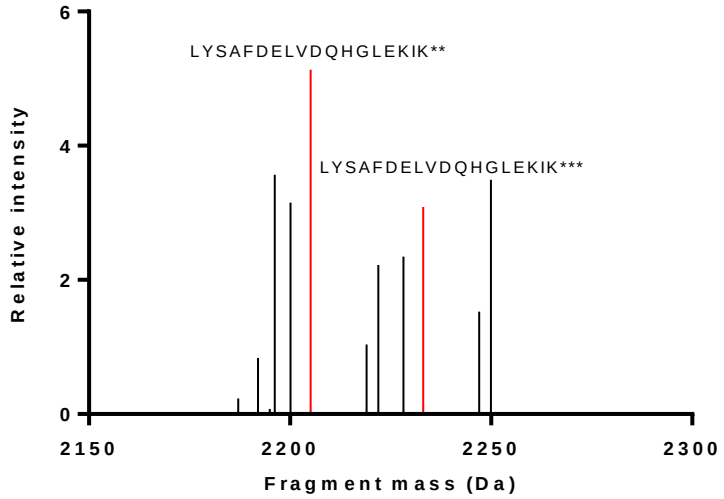


Figure 4.5.2: Magnified MALDI mass spectra of Rv1625c$_{204-443}$

The MALDI Mass spectrum shown in Figure 5.5.1A was magnified to show m/z values including those corresponding to the mass of the LYSAFDELVDQHGLEKIK fragment of the protein plus the mass of an ethylated carbamate and up to three additional ethylations. The carbamylated fragments are indicated by the red peaks. The * symbol denotes an ethyl group. These masses were only found in a single experiment, in subsequent experiments they could not be reproduced.

or $^{13}CO_2$ experiments. It was possible that this was due to the fact that the fragment detected was a trypsin miscleavege fragment, which occurs when trypsin misses a digestion site so what would have been two separate fragments become one. In this case the correctly cleaved fragment would have produced a fragment containing the K296 residue of interest that was just three amino acids long, too small to be detected by the mass spectrometer. Therefore the miscleavege was required so that the K296 residue could be detected, and it is possible that in the other experiments this trypsin miscleavage didn't occur. Therefore for the Rv1625c protein too the mass spectrometry results proved inconclusive as to whether CO_2 formed a carbamate on the protein or not.

4.6 Chapter Summary

Overall the results from the mass spectrometry experiments on both the tmAC and the Rv1625c catalytic domains were inconclusive. As no carbamates were detected it is possible that neither of the proteins forms a carbamate. However for both proteins there were m/z values detected that could correspond to the mass of a carbamylated fragment, but these could not be reproduced across different experiments.

It could be seen that the mass spectra appeared quite different in different experiments when they should have been more similar. Therefore it is possible that a lack of reproducibility in the technique meant that the m/z values of the carbamylated fragments were not present in all experiments. On the other hand the carbamylated fragments appeared consistently in the control haemoglobin experiments discussed in Section 3, therefore the technique was reliable enough to give this results.

Another potential factor that could influence the detection of carbamylated fragments is the position of any potential carbamate modification within the protein itself. It is possible that the haemoglobin carbamate on the N terminus of the protein is very accessible to the TEO reagent and therefore easily ethylated and stabilised. Potentially carbamates in a region of a protein that is less accessible will not be stabilised well and therefore less will be detected in the analysis.

To overcome these problems it was decided to approach the question using a different method not involving mass spectrometry, using $^{14}CO_2$ labelling of the carbamates. This method and the results will be discussed in section 5.0.

5 Investigation of potential carbamylation of Rv1625c using $^{14}CO_2$ labelling

5.1 Chapter Introduction

It has been demonstrated that MALDI mass spectrometry was not successful in showing whether the catalytic domains of either tmAC or Rv1625c are carbamylated. In the case of the Rv1625c protein this was most likely due to the positioning of the trypsin cleavage site (Section 4.5). As $Rv1625c_{204-443}$ protein had previously been shown to bind CO_2 (Townsend et al., 2009) and was also able to be expressed in higher yields than the tmAC domains, it was decided to proceed with this protein. Therefore an alternative method was investigated to demonstrate whether the K296 residue in the catalytic domain of Rv1625c formed a carbamate.

Radioactive labelling using ^{14}C is a technique that has been used in a variety of different situations to investigate whether a material binds or incorporates CO_2 in the form of carbamates (Lorimer and Miziorko, 1980, Golemi et al., 2001, Townsend et al., 2009). It was hypothesized that carbamates formed from $^{14}CO_2$ derived from radioactively labelled $NaH^{14}CO_3$, would be trapped by the TEO reagent, resulting in $^{14}CO_2$ being stably introduced onto the protein. We also hypothesized that radioactive bicarbonate or carbon dioxide not covalently bound to the protein could be washed away using an acidic buffer. This washing should both physically remove the $^{14}CO_2$ as well as causing remaining bicarbonate to convert to gaseous CO_2 because of the low pH, enabling it to diffuse out of the solution. Washing unbound $NaH^{14}CO_3$ from the protein is simplified due to the highly hydrophobic nature of the ethylated protein causing it to precipitate. This means that the precipitated protein can be re-suspended in washing buffer then centrifuged to collect the protein precipitate again. The washed protein precipitate may then be mixed with a suitable scintillation fluid and the amount of ^{14}C radioactivity from trapped carbamates analysed using a scintillation counter.

It was therefore decided to use this approach as an alternative method to investigate whether Rv1625c was able to bind CO_2 as a carbamate. As previously discussed, K296 of Rv1625c is predicted as the most likely residue of the protein to bind a functional carbamate, due to its position in the active site of the protein being conserved with the bicarbonate binding active site in soluble adenylyl cyclase. The same recombinant $RV1625c_{204-443}$ protein (containing the catalytic domain of Rv1625c) that was used in the mass spectrometry experiments of section 4.5 was also used in the $^{14}CO_2$ radioactivity assays to investigate this hypothesis. It was hypothesised that if a carbamate is indeed forming at the K296 residue of $Rv1625c_{204-443}$, mutating this residue to alanine should ablate the ability of $Rv1625c_{204-443}$ to form a carbamate. As an additional control an $Rv1625c_{204-443}$ K397A mutant protein was also used. In this protein the K397 residue is not hypothesized to have

any carbamate forming properties (though the possibility cannot be formally excluded). Therefore when mutated to an alanine the protein should show no differences in carbamate formation to the wild type $Rv1625c_{204-443}$.

In addition to the $Rv1625_{204-443}$ proteins, control proteins were also included in the assay. These were Rubisco, natalase and Rx1. Rubisco was used as a positive control as it is known to form carbamates on Lys201 residues in its large subunits (see section 1.3.2) and has been used in similar radioactivity trapping experiments (Lorimer and Miziorko, 1980, Stec, 2012). Natalase is a synthetic α-amylase enzyme used in detergents (Kasturi et al., 2010) and Rx1 is a plant Nucleotide-Binding Leucine Rich Repeat (NLR) immune receptor (Fenyk et al., 2015). Both natalase and Rx1 were selected as negative controls as they have no identified ability to bind CO_2 (though the possibility of this cannot be excluded), and were readily available for use.

The Rubisco, natalase and Rx1 proteins were all available for purchase or had previously been made. The wild type recombinant protein $Rv1625c_{204-443}$ was expressed and purified as described previously in Section 4.3. The two mutant $Rv1625c^{204-443}$ recombinant proteins that were produced for this experiment were $Rv1625c_{204-443}$ K296A and $Rv1625c^{204-443}$ K397A.

5.2 Expression and Purification of $Rv1625c_{204-443}$ mutants

$Rv1625c_{204-443}$ K296A and $Rv1625c_{204-443}$ K397A proteins were overexpressed in M15 (DE3) E-coli. The hexahistidine tagged Rv1625c proteins were then purified using Ni^{2+} affinity chromatography. The proteins were eluted from the affinity resin using a buffer containing imidazole which competes with the hexahistidine tag for binding to the affinity resin. Eluted protein fractions were run on a 15 % (w/v) SDS-PAGE gel for analysis of purity, including samples from the wash steps and the original cell lysates before and after incubation with resin (Figures 5.2.1 and 5.2.2).

The majority of the eluted protein migrated at a molecular weight of ~25KDa for both the $RV1625c_{204-443}$ K296A and $RV1625c_{204-443}$ K397A proteins, which have predicted molecular weights of 26.3KDa. Therefore it was likely the large band at ~25KDa was the expressed Rv1625c protein. This was confirmed by trypsin digesting the proteins and analysing the fragments using MALDI mass spectrometry. It was confirmed that each of the proteins was derived from the Rv1625c protein from *Mycobacterium tuberculosis.* It was estimated that each of the proteins was at least 90% pure which was deemed acceptable for use in the experiments. Running samples from the wash steps on the gel also showed that the wash

methods used during purification were successful in removing the majority of proteins bound unspecifically to the resin (Figures 5.2.1 and 5.2.2).

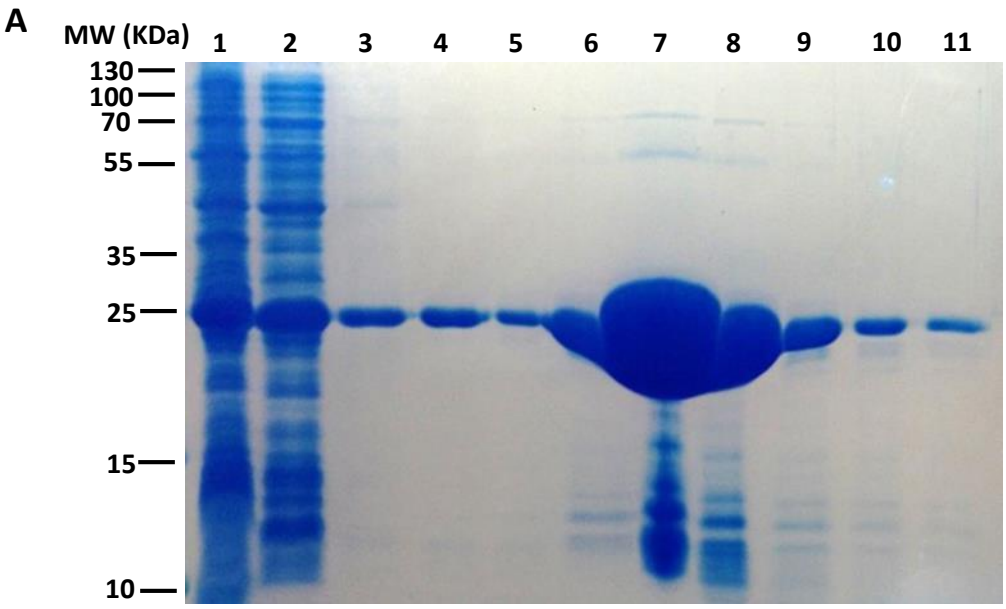


Figure 5.2.1: Purification of Rv1625c$_{204-443}$ K296A mutant

SDS-PAGE gel of different stages of the purification of RV1625c$_{204-443}$ K296A protein from M15 (DE3) *E. coli* lysate. All protein is stained with Coomassie blue and the molecular weight standards are indicated on the left hand side of the gel. Lane 1) Whole cell lysate before incubation with resin Lane 2) Whole cell lysate after incubation with Nickel resin, pre wash. Lanes 3-5) Nickel resin column flow through after three respective washes. Lanes 6-11) Separate fractions after eluting protein with imidazole, with each fraction eluted using 1ml of elution buffer per fraction. The molecular weight of RV1625c$_{204-443}$ K296A is 26.2KDa, and the vast majority of protein is seen at ~25KDa.

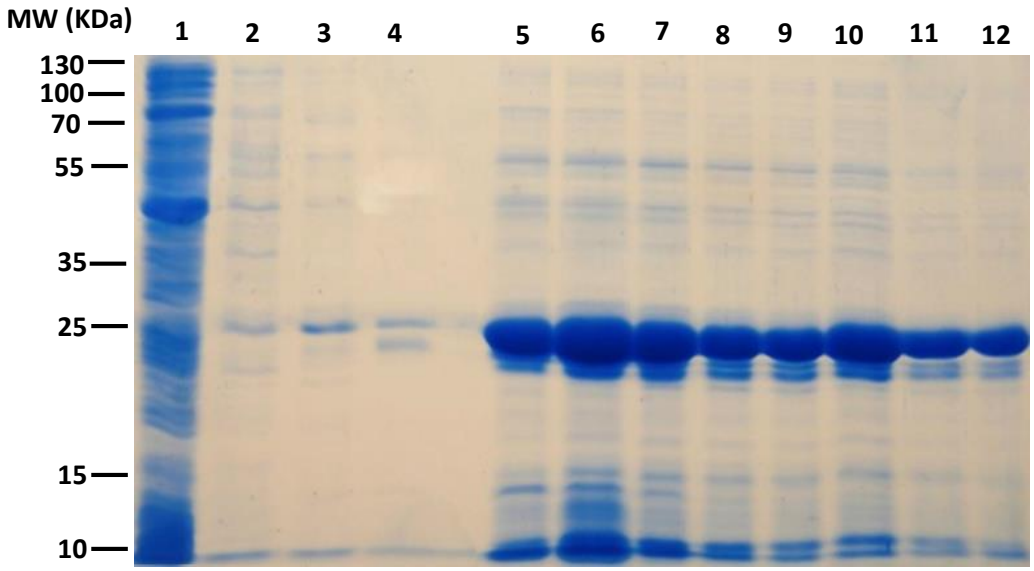


Figure 5.2.2: Purification of Rv1625c$_{204-443}$ K397A mutant

A) SDS-PAGE gel of different stages of the purification of RV1625c$_{204-443}$ K397A protein from M15 (DE3) *E. coli* lysate. All protein is stained with Coomassie blue and the molecular weight standards are indicated on the left hand side of the gel. Lane 1) Whole cell lysate after incubation with Nickel resin, pre wash. Lanes 2-4) Nickel resin column flow through after three respective washes. Lanes 5-12) Separate fractions after eluting protein with imidazole, with each fraction eluted using 1 ml of elution buffer per fraction. The molecular weight of RV1625c$_{204-443}$ K397A is 26.2KDa, and the majority of protein is seen at ~25KDa.

5.3 Trapping experiments on Rv1625c$_{204-443}$ recombinant proteins using $^{14}CO_2$

The abilities of the Rv1625c$_{204-443}$ wild type and mutant proteins to form a carbamate were then explored using the $^{14}CO_2$ binding assay described in the introduction to this chapter (Section 5.1). For each experiment the percentage of $^{14}CO_2$ trapped was calculated by dividing the number of radioactive counts in the washed protein precipitate by the number of radioactive counts at the start of the reaction when the $NaH^{14}CO_3$ was added. The concentration of $NaH^{14}CO_3$ was negligible compared to the 20mM $NaH^{12}CO_3$ in the reaction. Therefore the percentage of $^{14}CO_2$ bound could be used to calculate the total amount in moles of carbamylated protein in a solution containing 20mM $NaH^{12}CO_3$, using the following formula:

Moles of CO_2 bound to protein = % of $^{14}CO_2$ bound x moles of $NaH^{12}CO_3$ used

It was predicted that any $^{14}CO_2$ binding to an amine group as a carbamate would be trapped and stabilised by the TEO reagent, thereby also fixing the radioactivity onto the protein. On this basis Rubisco should form the most carbamate as each of the eight large subunits of a Rubisco molecule can bind a molecule of CO_2. Therefore a fully carbamylated mole of Rubisco can bind eight moles of CO_2 as a carbamate. As Rv1625c$_{204-443}$ is hypothesised to have only one carbamylation site, the number of moles of CO_2 bound to Rubisco was divided by eight, to make the results more comparable to the other proteins used in the experiment.

Background radiation measurements were first carried out by measuring the radioactivity of protein precipitates in experiments where no ^{14}C was added. The average of the raw data values for all proteins (which was 30 Decays per minute) was subtracted from all other experimental results. As a reference the initial ^{14}C radioactivity of a sample was in the region of 20,000-30,000 Disintegrations per Minute (DPM), so the background radiation was about 1000 times less in comparison. It was also necessary to carry out this type of control to ensure that no protein precipitate was creating an abnormally high background reading that would interfere with experimental results. For example, it was found that haemoglobin precipitated by TEO greatly interferes with the readings of the scintillation counter, possibly due to its chromophore, giving high readings of radioactivity counts even when the sample contained no ^{14}C (Figure 5.3.1). Interestingly haemoglobin precipitated with acetone still had higher than average background readings but they were much lower than the TEO precipitate. As the TEO precipitate tended to be much denser than the acetone precipitate, this could explain the results. Therefore although haemoglobin was used as a positive control in the mass spectrometry experiments, it could not be used it in this experiment and so Rubisco was the positive control used.

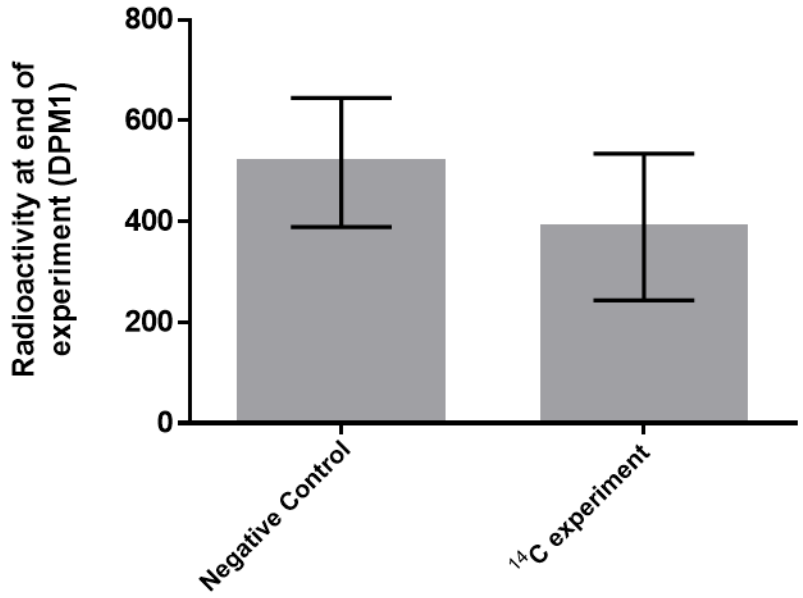


Figure 5.3.1: Haemoglobin precipitated by TEO gives false radioactivity reading

Each protein was tested in the radioactivity experiment for false background reading, as sometimes precipitated protein could artificially increase the number of perceived radioactivity counts by the scintillation counter. Haemoglobin was either mixed with $NaH^{14}CO_3$ and precipitated with TEO before excess radioactivity was washed off , or as a negative control not exposed to any radioactivity and just precipitated with TEO, as described in the methods section. It can be seen that the negative control produced more radioactivity counts than the experimental protein, despite not being exposed to any radioactivity. The experiment was repeated two times for the negative control and four times for the actual trapping experiment. The TEO haemoglobin precipitate was clearly interfering with the readings from the scintillation counter.

Control experiments using no TEO were also repeated three times for each protein, to ensure only $^{14}CO_2$ bound as carbamates was being detected (Figure 5.3.2A and 5.3.2C). The average of the 'no TEO' control values calculated for each proteins were subtracted from the relevant protein's results, to make sure all readings of $^{14}CO_2$ were due to carbamate formation only.

To clarify the terminology used throughout this section with regard to carbamate binding: a 'site specific' or 'specifically bound' carbamate is used to describe a carbamate bound to a residue in a site specifically designed to form a carbamate, and which is therefore likely to have a functional role. The K201 residue in Rubisco is an example of a specific carbamate binding residue, as is the K70 in the Class D B-Lactamase OXA-10 (Li et al., 2005). These sites may be designed to facilitate carbamylation by stabilising interactions with neighbouring residues, or by having properties which favour the neutral state of the lysine residue, for example by creating a hydrophobic environment (Leonard et al., 2008, Leonard et al., 2013). 'Unspecific' or 'background' carbamylation is used as a hypothetical explanation to account for the possibility that carbamates could form at random amine side chains or N-terminal residues. These residues could be capable of binding a carbamate due to having a free amine group, but are likely not specialised to do so. Therefore any carbamate bound will have done so by chance and there will be much lower levels of carbamate formation than would be expected if the residues were in an environment designed to bind functional carbamates.

Results are displayed in units of moles of CO_2 bound per mole of protein and also moles of CO_2 bound per mg of protein. The data from CO_2 bound per mg of protein (Fig 5.3.2D) was used to calculate the molar results (Fig. 5.3.2B). It was found that Rubisco bound on average 0.52 moles of carbamate per mole of protein, after normalising for the eight putative carbamate binding sites as previously described (Fig 5.3.2B). As a protein known to form carbamates and the positive control for the experiment the result for Rubisco was expected, though the amount of carbamate formation was less than the potential maximum of a 1:1 molar ratio of CO_2: protein.

The other control proteins Rx1 and natalase showed binding of 0.76 and 0.14 moles of carbamate per mole of protein, respectively (Fig 5.3.2B). These results are more surprising, as both are proteins with no known ability to bind carbamates. The 0.14 moles of CO_2 binding per mole of natalase showed that the protein was binding a small amount of carbamate, but almost four times less than Rubisco, a difference found to be significant. This result would suggest that the protein has no specific site designed for carbamylation, but that some uncharged amine side chains have bound carbamates in a random manner.

This may have occurred during a temporary pH fluctuation altering the pK_a value of an amine group that wouldn't normally bind carbamates to allow it do so. In this respect natalase is suitable as a negative control displaying some level of 'background' carbamate formation. However Rx1 showed a greater amount of carbamylation than Rubisco, though the difference was not statistically significant. This was unexpected but it is possible that Rx1 does have one or more specific carbamate binding sites that are yet undiscovered. This would be interesting to investigate using mass spectrometry.

It was hypothesised that the wild type $RV1625c_{204-443}$ protein would form a carbamate at its K296 residue, and that the K296A mutant would therefore be unable to form a carbamate. On the other hand the $RV1625c_{204-443}$ K397A mutant was hypothesised to show no difference in carbamate formation to the wild type protein, as there is no evidence that the K397A residue plays a role in carbamate formation. The data demonstrate that all three $Rv1625c_{204-443}$ proteins show a low amount of carbamate formation with values of 0.26, 0.21 and 0.18 moles of CO_2/per mole of protein for the wild type, K296A and K397A proteins respectively (Fig. 5.3.2B). Therefore the wild type was binding slightly more carbamate than the K296A mutant but the K397A mutant was binding less CO_2 than both the wild type and K296A mutant proteins. None of the differences between the three $Rv1625c_{204-443}$ proteins were significant when the p values were set at 0.05 or lower.

The maximum amount of carbamate formation expected for a protein with one specific binding site is 1 mole of carbamate per mole of protein. Therefore the $Rv1625c_{204-443}$ wild type protein was only binding 26% of its proposed maximum carbamate capacity, compared to the 52% exhibited by the positive control Rubisco and the even greater 76% shown by Rx1 (although the number of potential binding sites on Rx1 is unknown). However the wild type $Rv1625c_{204-443}$ protein also almost formed almost double more carbamate than natalase. As neither of the differences in carbamate formation between $RV1625c_{204-443}$ and either Rx1 or Rubisco were significant, it is more difficult to conclude whether the adenylyl cyclase is binding a significant amount of carbamate or not. However the fact that the K296A mutant shows very little decrease in the amount of carbamate formation compared to the wild type protein indicates that the protein is not binding a carbamate at the K296 residue. In addition the K397A mutant showed even less carbamate formation than the K296A mutant, when it was predicted that ablating K397A should have shown no difference from the wild type protein. Together these results suggest that the $Rv1625c_{204-443}$ protein is not forming a carbamate at the hypothesised K296A residue. It is most likely the carbamate binding displayed by the $Rv1625c_{204-443}$ proteins and natalase is from a certain amount of non-specific carbamate formation at free amino groups in the proteins.

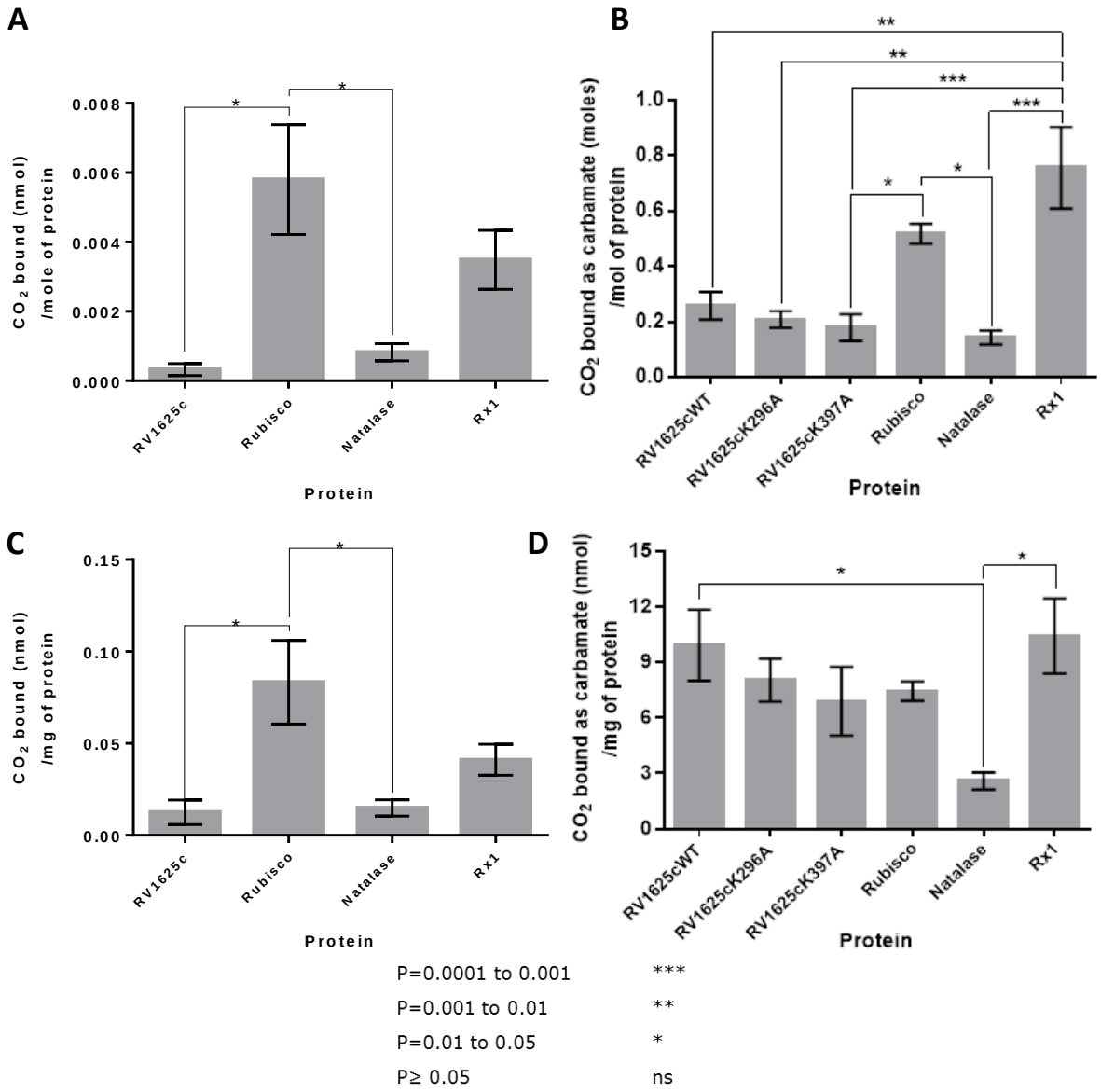


Figure 5.3.2: CO_2 bound to proteins as a carbamate in $^{14}CO_2$ radioactivity assay

Amount of CO_2 bound as a carbamate to RV1625c$_{204-443}$ wild type and mutant proteins along with different control proteins. Each bar represents the mean of three individual experiments, error bars show +/- 1SE of the mean. Control experiments containing no TEO trapping reagent show the amount of remaining CO_2 on each protein after the washing steps in **A)** moles of CO_2 bound per mole of protein and **C)** moles of CO_2 bound per mg of protein. Results from the trapping experiments after subtracting the 'no TEO' control values are displayed as **B)** moles of CO_2 bound per mole of protein and **D)** moles of CO_2 bound per mg of protein. The data displayed in Figures A and B was derived from the data shown in Figures C and D respectively. Significance between different groups was calculated for all the data using Tukey's multiple comparisons test. Only significant differences between proteins have been indicated, all other differences are not significant.

It has to be taken into account that the accuracy of this experimental approach is limited because of the levels of 'unspecific' or 'background' carbamylation that appears to be occurring at unspecific amine sites on all proteins. This could lead to misleading results unless the true extent of unspecific carbamylation vs site specific carbamylation is calculated for the positive control Rubisco. This would be best done by expressing Rubisco in *E. coli* both as a wild type protein and with the carbamate binding K201 residue mutated to a residue unable to form a carbamate, for example alanine. This was unable to be performed in this experiment due to time constraints, but would be essential for anyone using this method in the future. The amount of unspecific carbamylation could then be subtracted from results for other proteins in the assay to represent the amount of carbamylation at specific sites. This result would still be an estimate as it is possible that the amount of unspecific carbamylation may vary depending on the amino acid compositions of each protein. For example proteins containing more lysine residues have the potential to form more carbamates, and the local pK_a values of these residues will also influence this. By using more positive controls proteins to calculate the amount of 'background' carbamate formation a more accurate estimation could be formed to overcome this problem.

Whether the amounts of carbamate forming at random amine groups is an artefact of the experiment or also occurs in vivo would be interesting to investigate. The pH and the amount of CO_2 present were designed to reflect the same conditions as would be found in the proteins physiological environment. However this is very hard to precisely determine and so these factors could have caused unusual carbamates to form at unlikely residues. Similarly the conditions used could affect the amount of carbamate formation at sites specifically designed to bind carbamates. This could explain why not as much carbamate formation was seen for Rubisco as could be expected; it is known that the pH of the plant cytosol increases during the light reaction of photosynthesis, which could directly impact on levels of Rubisco carbamate formation (Yin et al., 1991).

In addition, despite efforts to control the pH, it would occasionally fluctuate to levels as high as pH 8.5 during the experiment, which could have led to 'unspecific' carbamates forming at amine groups that CO_2 wouldn't usually bind to, and being trapped by TEO. Another possibility producing the opposite effect is the competition between CO_2 and the TEO reagent for the neutral amine group, as TEO is able to ethylate an amine with or without a trapped carbamate. Although the carbamates should have formed before the addition of the TEO they are labile. It is possible that a temporary dissociation of a carbamate from an amine group would allow TEO to ethylate the amine, blocking any future carbamylation. However this effect is less likely to occur with carbamates bound in a site specific manner,

as they will be more stable than carbamates bound at non-specific amine sidechains. It is clear that whether that carbamate formation also occurs in vivo at unspecific sites or is an artefact of the experiment, better controls are needed in the experiment to allow all the experimental results to be comparable to each other.

Finally the results discussed have focused on the amount of carbamate formation per mole of each protein (Fig. 5.3.2B) Comparing the amount of carbamate formed using moles as the unit of protein is essential the proteins are binding a carbamate at a specific binding site only. In this case comparing the carbamates bound using the same masses of each protein would be misleading, as those with larger molecular weights would have far fewer binding sites present than those with smaller molecular weights. However as already mentioned at the beginning of this section the amount of carbamate bound per mg of protein (Fig. 5.3.2B), was used to calculate the molar data. This was due to limits on the amount of protein available to use. It has been shown that there is a level of unspecific carbamate formation among all proteins. However it has been impossible to calculate the amount of unspecific carbamate without the use of a Rubisco mutant. Therefore when calculating the amounts of carbamate bound per mole of protein from the mass data the unspecific carbamates have been included in the calculation. By comparing the two sets of data it can be seen that the values of carbamate formation in Figure 5.3.2D are much more similar than those in Figure 5.3.2B. It is possible that when calculating the molar data the results have been skewed in favour of Rx1 and Rubisco which have larger molecular weights forming more carbamate. Alternatively it is possible that the correlation between protein molecular weight and carbamate formation is coincidental. With the use of a mutant of the carbamate binding residue in Rubisco this potential problem could be solved by subtracting 'background' carbamylation from the mass data before it was used to calculate the molar data.

5.4 Chapter Summary

Overall the results from these experiments suggest that the K296 residue in the active site of the Rv1625c protein is not a specific carbamate binding residue. This is based on the evidence showing there was a very small and insignificant difference between the amount of carbamate bound by the $RV1625c_{204-443}$ wild type and K296A mutant proteins (Figure 5.3.2B). If the K296 residue was responsible for carbamate formation a significant drop in carbamate formation would have been expected. In addition the K397A mutant showed even less carbamate formation than the wild type protein, though again this difference was non-significant (Figure 5.3.2B). It is therefore likely that the difference seen between the wild type and K397A proteins is not real. The overall small amount of carbamate formation in all the $Rv1625c_{204-443}$ proteins compared to the positive control Rubisco and also Rx1

further suggests that the Rv1625c protein does not have any specific carbamate binding sites.

The limitations of the experiment in regard to all proteins forming a certain amount of carbamate also have to be taken into account. The obvious way to measure the amount of unspecific carbamylation in a protein is to mutate the site that is carbamylated. Due to the timescale of this project we were unable to create a Rubisco mutant, which would have provided valuable information as to how much of the CO_2 binding is specific. However the putative K296 carbamylation site in $RV1625c_{204-443}$ was mutated, and the differences between the mutant proteins were still not significant. Another possibility is that a potential carbamate in $Rv1625c_{204-443}$ may not be accessible to ethylation. If this is the case it means that using TEO as an ethylating agent will need to be further assessed using more positive control proteins, to establish how accessible carbamates are to ethylation. Therefore the results from this experiment indicate that $Rv1625c_{204-443}$ does not form a carbamate at the K296 residue, however further investigation is required to confirm this result.

These results do not mean that the Rv1625c adenylyl cyclase is not binding to CO_2 at all. There is clear evidence in the literature that both mammalian transmembrane adenylyl cyclase and Rv1625c is activated by CO_2 and that it is probably this occurs by CO_2 binding to the protein (Townsend et al., 2009). It is possible that Rv1625 is binding CO_2 in a form that is not a carbamate, potentially at a different residue to Lys296 of the active site, or in a way that necessitates the interaction of several different residues and not just the single Lys296 residue. In addition, past experiments that have shown an interaction with CO_2 were carried out at a relatively low pH of 6.5 compared to the pH of 7.4 in this experiment (Townsend et al., 2009). The likelihood of a carbamate forming at this pH is slim due to the fact that most lysine residue will be charged due to the increased number of H^+ ions in the environment. Rubisco is known to function best at a pH of 7.8 or higher due to increased carbamylation. Therefore it is not entirely unsurprising that the Rv1625c proteins do not appear to be forming carbamates.

The high amount of carbamate formed on Rx1 is an unexpected and interesting result. Little is known about the effects of CO_2 on the plant innate immune response, however one study found that high atmospheric CO_2 increased pathogen aggressiveness in *Arabidopsis thaliana*, through a proposed pathway involving leaf epidermal changes (Lake and Wade, 2009). Whether CO_2 could also potentially modify immune proteins like Rx1 through carbamylation requires further investigation. This finding highlights the potential of this radioactivity assay using TEO as a trapping agent as a simple method of discovering whether a protein binds carbamate or not, once better negative controls have been

established. However the assay has limitations compared to a mass spectrometry method in that it cannot be used to locate and confirm the identity of a carbamate binding residue. In addition it cannot be used on multiple proteins simultaneously and identify which proteins bind carbamates, which a mass spectrometry based method has the potential for.

An additional benefit of this assay over previous ones is using TEO as a trapping agent that is able to stably bind carbamates whilst allowing the vast majority of other loosely bound radioactivity to be washed off (Figure 5.3.2A and 5.3.2C) This is useful as it results in very low background radioactivity from CO_2 bound in other ways to the protein. This is an advantage over similar radioactivity assays, which often simply separate bound and unbound CO_2 using a sephadex column (Townsend et al., 2009, Golemi et al., 2001). These methods do not guarantee that the radioactivity measured is due to CO_2 bound as carbamates alone. The methylating agent diazomethane has been combined with a radioactivity assay in the past to trap and identify carbamate binding on Rubisco (Lorimer and Miziorko, 1980). However the conditions used in this experiment required organic solvents and low temperatures which were highly artificial, as well as hazardous. Therefore TEO appears is a more desirable trapping agent as it can be used in an aqueous buffer under normal temperatures to represent the physiological environment of the protein, and is also less hazardous.

6 Final discussion and future work

6.1 Final discussion

Carbamate modifications occur when the nitrogen atom of a neutral amine makes a nucleophilic attack upon the carbon atom of a carbon dioxide molecule. This leads to the CO_2 becoming attached to the amine group as a carbamate. It has been shown that carbamates can play important functional and structural roles in proteins. There are also an increasing number of studies showing that inorganic carbon plays an important role in many signalling pathways and processes independently of pH (Jones et al., 2007, Miller et al., 2014, Smith et al., 2013, Hallem et al., 2011, Taylor and Cummins, 2011). Despite this, the mechanisms of inorganic carbon signalling are often not investigated further and in many cases remain a mystery. It is possible that CO_2 could be binding to and influencing the functions of the proteins in these systems via a carbamate modification at lysine residues, the N terminus, or possibly arginine residues (Meigh, 2015).

This first aim of this project was to develop a method to trap carbamate modifications onto proteins by ethylation using the chemical reagent Triethyloxonium tetrafluoroborate (TEO), to allow their analysis by mass spectrometry. It has been shown in Section 3.0 that a TEO trapping method initially developed by a colleague (Victoria Linthwaite) was successfully used to trap an N terminal carbamate on the amino acid phenylalanine. The method was then successfully adapted for use on a protein using haemoglobin as a positive control for carbamate formation. It was shown from these experiments that TEO could effectively trap a carbamates on the N-terminal valine residues of haemoglobin α and β chains, which are known to form the modification(Kaplan et al., 1982) MALDI and Electrospray Mass Spectrometry analysis were used to identify the fragment from its molecular weight, with MS/MS was used to confirm the identity of the fragment from its individual amino acids. Work to refine and optimise the method also succeeded in producing higher quality and more consistent results compared to initial experiments.

The second aim of the project was to use the trapping and mass spectrometry methods developed on a protein shown to respond to CO_2, and with the potential to form a carbamate. Under these criteria two recombinant transmembrane adenylyl cyclases were tested to see if they underwent carbamylation, the C1 and C2 catalytic domains from the mammalian tmAC and the related catalytic domain of the Rv1625c protein from *Mycobacterim tuberculosis*. These proteins were chosen as they have been shown to respond to inorganic carbon and specifically molecular CO_2 (Townsend et al., 2009, Nunes et al., 2013), in addition to possessing conserved lysine and arginine residues in their active sites which could potentially be capable of forming a carbamate (Liu et al., 1997, Linder, 2006). However initial results shown in Section 4.0 did not show any convincing evidence for carbamylation of either of the catalytic domains of the mammalian tmAC. Although one

protein fragment of the C1 domain that appeared in one experiment had a mass that corresponded to the presence of a trapped carbamate, this mass did not appear in any other experiments. Therefore the same experiments were performed on the related $Rv1625c_{204-443}$ protein, as an alternative adenylyl cyclase thought to bind CO_2 (Townsend et al., 2009). Initially results appeared positive as several masses appeared in one experiment corresponding to a carbamylated and ethylated LYSAFDELVDQHGLEKIK fragment, of which the second lysine residue (K296) plays a key role in the active site of the protein (Townsend et al., 2009, Guo et al., 2001). However as with the mammalian tmAC proteins, the results could not be reproduced in subsequent experiments.

Therefore it emerged during the study that an important issue was the reproducibility of the mass spectrometry results, as in all the adenylyl cyclase experiments the results were variable and not very reproducible. Consequently when a fragment mass corresponding to a putative carbamylated residue appeared it was hard to have confidence in its identity. Although the inability to reproduce the results suggests that no carbamate was present, there are many other factors that could have influenced the outcome of the experiment.

Firstly the conditions of the trapping experiment itself could have varied, for example fluctuating pH caused by the hydrolysis of the TEO could have affected the carbamylation status of the protein. Secondly the preparation of the sample for mass spectrometry could have introduced variability, for example during the trypsin digestion of the protein precipitate. Finally the mass spectrometry itself may have affected the results. The MALDI/TOF technique used has many advantages, chiefly that it is rapid, almost exclusively creates intact singly charged ions which aids data analysis, and is not largely affected by small amounts of contaminating salts (Domon and Aebersold, 2006). However it also has drawbacks including background signal from matrix ions, and if the samples to be analysed are not completely uniform in the amounts of salts and other materials they contain this will also influence the results(Pan et al., 2007).

Secondly although the experiment produced consistent results for haemoglobin this was after method optimisation, and it is possible that different conditions are needed for different proteins, depending on the positions of any potential carbamates. However it is hard to optimise an experiment for a protein unknown to form a carbamate. It is also possible that the trapping method will not work on some proteins, for example if the carbamate is inaccessible to ethylation by TEO. Therefore in future the experiment should be tested using more positive controls for carbamylation such as Rubisco and β-lactamases to see how reliable it is in identifying trapped carbamates.

However the additional work needed to improve the trapping method and the mass spectrometry analysis was beyond the time constraints of this project. It was decided instead to further investigate the $Rv1625c_{204-443}$ protein that was shown to potentially be carbamylated, using a different technique to mass spectrometry. The trapping experiments described in Section 5.0 were performed on $Rv1625c_{204-443}$ proteins using radioactively labelled $NaH^{14}CO_3$ as the source of CO_2. Mutating the K397 residue of $Rv1625c_{204-443}$ to alanine showed no difference in the amount of carbamate bound, as would be expected if the K296 residue was responsible for carbamate formation as hypothesised. However mutating the K296 residue to alanine also showed no decrease in carbamate formation. This suggests that no carbamate was forming at this residue. Therefore it seems likely that the Rv1625c protein doesn't form a carbamate at the substrate defining K296A residue, although the possibility of TEO not being able to modify this amino acid cannot be ruled out.

It is also possible that the RV1625c protein could be forming a carbamate at another position. To establish this a robust negative control is needed for the $^{14}CO_2$ trapping experiment, such as a mutation of the carbamate forming K201 residue of spinach Rubisco. Although other proteins not hypothesised to form a carbamate were used in the experiment as negative controls, the possibility that they did actually form one or more carbamates could not be ruled out. This was especially in this experiment the case as the results showed that all proteins did appear to form carbamates to some extent compared to controls where no TEO was present. This suggests that some level of 'unspecific' carbamylation may occur on all proteins, perhaps at very low levels. If this is the case it also requires further investigation, especially as no carbamates were detected by the MALDI experiments in Section 4.0, contradicting this theory.

The high level of trapped carbamate on the Rx1 nucleotide binding protein was also an interesting find, as it is has not been previously found to have any association with CO_2 signalling or carbamate formation, and was originally intended as a negative control. This needs further investigation in future.

Overall it has been shown that a carbamate trapping method has been developed for use on proteins and has been shown to work reproducibly on haemoglobin, though experiments on adenylyl cyclases were inconclusive. The method has several advantages over previous carbamate detection strategies; it removes the complications caused by trying to analyse unstable carbamate modifications, and being able to trap carbamates in the native conditions of the protein means results are more physiologically relevant and less prone to experimental artefacts.

6.2 Future work

Future work still needs to be done to determine how to make results more reproducible. This should begin with the testing of more proteins known to form carbamates as positive controls in the mass spectrometry experiments, to enable the reproducibility and accuracy of the method to be determined. At least one negative control protein where the specific carbamate binding amino acid is mutated should also be used in future $^{14}CO_2$ experiments, to establish whether proteins still show 'background' protein carbamylation. If so it should be determined whether this is a true result or if it is an artefact of the assay. The potential of the Rx1 protein to form a carbamate should also be investigated further, as it is possible that a new role for CO_2 interaction with this protein could be discovered.

The ultimate goal which the results from these experiments contribute towards is to be able to trap potential carbamates on multiple proteins at a time under physiological conditions, for example in the lysate of a cell. The trapped proteins could then be screened simultaneously using mass spectrometry proteomic techniques, and those found to form carbamate modifications identified and subjected to further analysis to elucidate the functions of the identified carbamates. Use of this screening technique will enable rapid discovery of proteins which form carbamate modifications, facilitating our understanding of the extent of the modification in physiology and the ways in which it functions.

7 References

References

AEBERSOLD, R. & MANN, M. 2003. Mass spectrometry-based proteomics. *Nature,* 422**,** 198-207.

AFZAL-SHAH, M., WOODFORD, N. & LIVERMORE, D. M. 2001. Characterization of OXA-25, OXA-26, and OXA-27, molecular class D beta-lactamases associated with carbapenem resistance in clinical isolates of Acinetobacter baumannii. *Antimicrob Agents Chemother,* 45**,** 583-8.

ARNONE, A. 1974. X-ray studies of the interaction of CO2 with human deoxyhaemoglobin. *Nature,* 247**,** 143-5.

BEN AVRAHAM, R. & SHALITIN, Y. 1985. The specificity and kinetic properties of trypsin ethylated at the binding site. *FEBS Lett,* 180**,** 239-42.

BENNING, M. M., KUO, J. M., RAUSHEL, F. M. & HOLDEN, H. M. 1995. Three-dimensional structure of the binuclear metal center of phosphotriesterase. *Biochemistry,* 34**,** 7973-8.

BERG JM, T. J., STRYER L. 2002. *Biochemistry. 5th edition, Section 10.2* [Online]. New York: W H Freeman. Available: https://www.ncbi.nlm.nih.gov/books/NBK22596/ [Accessed].

BOHR, C., HASSELBALCH,K., AND KROGH,A 1904. Über einen in biologischer Beziehung wichtigen Einfluss, den die Kohlensäurespannung des Blutes auf dessen Sauerstoffbindung übt *Skand. Arch. Physiol,* 16**,** 401-412.

BRETSCHER, A. J., KODAMA-NAMBA, E., BUSCH, K. E., MURPHY, R. J., SOLTESZ, Z., LAURENT, P. & DE BONO, M. 2011. Temperature, oxygen, and salt-sensing neurons in C. elegans are carbon dioxide sensors that control avoidance behavior. *Neuron,* 69**,** 1099-113.

BUSH, K. 2013. Proliferation and significance of clinically relevant beta-lactamases. *Ann N Y Acad Sci,* 1277**,** 84-90.

BUSH, K., JACOBY, G. A. & MEDEIROS, A. A. 1995. A functional classification scheme for beta-lactamases and its correlation with molecular structure. *Antimicrob Agents Chemother,* 39**,** 1211-33.

CHAISSON, K. E. & HALLEM, E. A. 2012. Chemosensory behaviors of parasites. *Trends Parasitol,* 28**,** 427-36.

CHE, T., BETHEL, C. R., PUSZTAI-CAREY, M., BONOMO, R. A. & CAREY, P. R. 2014. The different inhibition mechanisms of OXA-1 and OXA-24 beta-lactamases are determined by the stability of active site carboxylated lysine. *J Biol Chem,* 289**,** 6152-64.

CHEN, F. M. F. & BENOITON, N. L. 1977. N-Ethylamino Acid Synthesis and N-Acylamino Acid Cleavage Using Meerweins Reagent. *Canadian Journal of Chemistry-Revue Canadienne De Chimie,* 55**,** 1433-1435.

CHEN, Y., CANN, M. J., LITVIN, T. N., IOURGENKO, V., SINCLAIR, M. L., LEVIN, L. R. & BUCK, J. 2000. Soluble adenylyl cyclase as an evolutionarily conserved bicarbonate sensor. *Science,* 289**,** 625-8.

COOK, Z. C., GRAY, M. A. & CANN, M. J. 2012. Elevated carbon dioxide blunts mammalian cAMP signaling dependent on inositol 1,4,5-triphosphate receptor-mediated Ca2+ release. *J Biol Chem,* 287**,** 26291-301.

DANCHIN, A. 1993. Phylogeny of adenylyl cyclases. *Adv Second Messenger Phosphoprotein Res,* 27**,** 109-62.

DIRKEN, M. N. & MOOK, H. W. 1930. The carriage of carbon dioxide by blood. *J Physiol,* 70**,** 373-84.

DOCQUIER, J. D., BENVENUTI, M., CALDERONE, V., GIULIANI, F., KAPETIS, D., DE LUCA, F., ROSSOLINI, G. M. & MANGANI, S. 2010. Crystal structure of the narrow-spectrum OXA-46 class D beta-lactamase: relationship between active-site lysine carbamylation and inhibition by polycarboxylates. *Antimicrob Agents Chemother,* 54**,** 2167-74.

DOCQUIER, J. D., CALDERONE, V., DE LUCA, F., BENVENUTI, M., GIULIANI, F., BELLUCCI, L., TAFI, A., NORDMANN, P., BOTTA, M., ROSSOLINI, G. M. & MANGANI, S. 2009. Crystal structure of the OXA-48 beta-lactamase reveals mechanistic diversity among class D carbapenemases. *Chem Biol,* 16**,** 540-7.

DOMON, B. & AEBERSOLD, R. 2006. Mass spectrometry and protein analysis. *Science,* 312**,** 212-7.

FENK, L. A. & DE BONO, M. 2015. Environmental CO2 inhibits Caenorhabditis elegans egg-laying by modulating olfactory neurons and evokes widespread changes in neural activity. *Proc Natl Acad Sci U S A,* 112**,** E3525-34.

FENYK, S., TOWNSEND, P. D., DIXON, C. H., SPIES, G. B., DE SAN EUSTAQUIO CAMPILLO, A., SLOOTWEG, E. J., WESTERHOF, L. B., GAWEHNS, F. K., KNIGHT, M. R., SHARPLES, G. J., GOVERSE, A., PALSSON, L. O., TAKKEN, F. L. & CANN, M. J. 2015. The Potato Nucleotide-Binding Leucine-Rich Repeat (NLR) Immune Receptor Rx1 is a Pathogen Dependent DNA-Deforming Protein. *J Biol Chem*.

FERGUSON, J. K. 1936. Carbamino compounds of CO(2) with human haemoglobin and their role in the transport of CO(2). *J Physiol,* 88**,** 40-55.

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

GOLEMI, D., MAVEYRAUD, L., VAKULENKO, S., SAMAMA, J. P. & MOBASHERY, S. 2001. Critical involvement of a carbamylated lysine in catalytic function of class D beta-lactamases. *Proceedings of the National Academy of Sciences of the United States of America,* 98**,** 14280-14285.

GROS, G., FORSTER, R. E. & LIN, L. 1976. The carbamate reaction of glycylglycine, plasma, and tissue extracts evaluated by a pH stopped flow apparatus. *J Biol Chem,* 251**,** 4398-407.

GUO, Y. L., SEEBACHER, T., KURZ, U., LINDER, J. U. & SCHULTZ, J. E. 2001. Adenylyl cyclase Rv1625c of Mycobacterium tuberculosis: a progenitor of mammalian adenylyl cyclases. *EMBO J,* 20**,** 3667-75.

HALLEM, E. A., SPENCER, W. C., MCWHIRTER, R. D., ZELLER, G., HENZ, S. R., RATSCH, G., MILLER, D. M., HORVITZ, H. R., STERNBERG, P. W. & RINGSTAD, N. 2011. Receptor-type guanylate cyclase is required for carbon dioxide sensation by Caenorhabditis elegans. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 254-259.

HAMMER, A., HODGSON, D. R. & CANN, M. J. 2006. Regulation of prokaryotic adenylyl cyclases by CO2. *Biochem J,* 396**,** 215-8.

HELDT, H. W. 1979. Light-Dependent Changes of Stromal H+ and Mg2+ Concentrations Controlling CO2 Fixation. *Photosynthesis II.* Berlin-Heidelberg: Springer-Verlag.

HENRIQUES, O. M. 1928. Die Bindungsweise des Kohlendioxyds im Blute. *Biochem.Z,* 200**,** 1-24.

HEWETT-EMMETT, D. & TASHIAN, R. E. 1996. Functional diversity, conservation, and convergence in the evolution of the alpha-, beta-, and gamma-carbonic anhydrase gene families. *Mol Phylogenet Evol,* 5**,** 50-77.

HO, C. S., LAM, C. W., CHAN, M. H., CHEUNG, R. C., LAW, L. K., LIT, L. C., NG, K. F., SUEN, M. W. & TAI, H. L. 2003. Electrospray ionisation mass spectrometry: principles and clinical applications. *Clin Biochem Rev,* 24**,** 3-12.

HSIA, C. C. 1998. Respiratory function of hemoglobin. *N Engl J Med,* 338**,** 239-47.

HU, J., ZHONG, C., DING, C., CHI, Q., WALZ, A., MOMBAERTS, P., MATSUNAMI, H. & LUO, M. 2007. Detection of near-atmospheric concentrations of CO2 by an olfactory subsystem in the mouse. *Science,* 317**,** 953-7.

IGAMBERDIEV, A. U. 2015. Control of Rubisco function via homeostatic equilibration of CO2 supply. *Front Plant Sci,* 6**,** 106.

ITO, Y. & USHITORA, H. 2006. Trapping of carbamic acid species with (trimethylsilyl)diazomethane. *Tetrahedron,* 62**,** 226-235.

JABRI, E., CARR, M. B., HAUSINGER, R. P. & KARPLUS, P. A. 1995. The Crystal-Structure of Urease from Klebsiella-Aerogenes. *Science,* 268**,** 998-1004.

JONES, W. D., CAYIRLIOGLU, P., KADOW, I. G. & VOSSHALL, L. B. 2007. Two chemosensory receptors together mediate carbon dioxide detection in Drosophila. *Nature,* 445**,** 86-90.
KAMENETSKY, M., MIDDELHAUFE, S., BANK, E. M., LEVIN, L. R., BUCK, J. & STEEGBORN, C. 2006. Molecular details of cAMP generation in mammalian cells: a tale of two systems. *J Mol Biol,* 362**,** 623-39.
KAPLAN, H., HAMEL, P. A., CHAN, A. M. & ODA, G. 1982. Chemical properties of the N-termini of human haemoglobin. *Biochem J,* 203**,** 435-43.
KASTURI, C., WANDSTRAT, M. E. & SONG, B. X. 2010. Liquid detergent composition exhibiting enhanced alpha-amylase enzyme stability. Google Patents.
KILMARTIN, J. V. & ROSSI-BERNARDI, L. 1969. Inhibition of CO2 combination and reduction of the Bohr effect in haemoglobin chemically modified at its alpha-amino groups. *Nature,* 222**,** 1243-6.
KILMARTIN, J. V. & ROSSI-BERNARDI, L. 1971. The binding of carbon dioxide by horse haemoglobin. *Biochem J,* 124**,** 31-45.
KING, G. J., GAZZOLA, C., BLAKELEY, R. L. & ZERNER, B. 1986. Triethyloxonium Tetrafluoroborate as an Ethylating Agent in Aqueous-Solution. *Inorganic Chemistry,* 25**,** 1078-1078.
KLEINBOELTING, S., DIAZ, A., MONIOT, S., VAN DEN HEUVEL, J., WEYAND, M., LEVIN, L. R., BUCK, J. & STEEGBORN, C. 2014. Crystal structures of human soluble adenylyl cyclase reveal mechanisms of catalysis and of its activation through bicarbonate. *Proc Natl Acad Sci U S A,* 111**,** 3727-32.
LAKE, J. A. & WADE, R. N. 2009. Plant-pathogen interactions and elevated CO2: morphological changes in favour of pathogens. *J Exp Bot,* 60**,** 3123-31.
LEE, C. K. & MANNING, J. M. 1973. Kinetics of the carbamylation of the amino groups of sickle cell hemoglobin by cyanate. *J Biol Chem,* 248**,** 5861-5.
LEONARD, D. A., BONOMO, R. A. & POWERS, R. A. 2013. Class D beta-lactamases: a reappraisal after five decades. *Acc Chem Res,* 46**,** 2407-15.
LEONARD, D. A., HUJER, A. M., SMITH, B. A., SCHNEIDER, K. D., BETHEL, C. R., HUJER, K. M. & BONOMO, R. A. 2008. The role of OXA-1 beta-lactamase Asp(66) in the stabilization of the active-site carbamate group and in substrate turnover. *Biochem J,* 410**,** 455-62.
LEVY, D. 1973. The esterification of insulin with triethyloxonium tetrafluoroborate. *Biochim Biophys Acta,* 310**,** 406-15.
LI, J., CROSS, J. B., VREVEN, T., MEROUEH, S. O., MOBASHERY, S. & SCHLEGEL, H. B. 2005. Lysine carboxylation in proteins: OXA-10 beta-lactamase. *Proteins-Structure Function and Bioinformatics,* 61**,** 246-257.
LINDER, J. U. 2006. Class III adenylyl cyclases: molecular mechanisms of catalysis and regulation. *Cell Mol Life Sci,* 63**,** 1736-51.
LINDER, J. U. 2008. Structure-function relationships in Escherichia coli adenylate cyclase. *Biochem J,* 415**,** 449-54.
LIU, Y., RUOHO, A. E., RAO, V. D. & HURLEY, J. H. 1997. Catalytic mechanism of the adenylyl and guanylyl cyclases: modeling and mutational analysis. *Proc Natl Acad Sci U S A,* 94**,** 13414-9.
LORIMER, G. H. 1983. Carbon-Dioxide and Carbamate Formation - the Makings of a Biochemical Control-System. *Trends in Biochemical Sciences,* 8**,** 65-68.
LORIMER, G. H. & MIZIORKO, H. M. 1980. Carbamate formation on the epsilon-amino group of a lysyl residue as the basis for the activation of ribulosebisphosphate carboxylase by CO2 and Mg2+. *Biochemistry,* 19**,** 5321-8.
LUNDBLAD, R. L. 2014. *Chemical Reagents for Protein Modication Fourth Edition,* Florida, CRC Press.
MAEDA, S., PRICE, G. D., BADGER, M. R., ENOMOTO, C. & OMATA, T. 2000. Bicarbonate binding activity of the CmpA protein of the cyanobacterium Synechococcus sp. strain PCC 7942 involved in active transport of bicarbonate. *J Biol Chem,* 275**,** 20551-5.

MAVEYRAUD, L., GOLEMI, D., KOTRA, L. P., TRANIER, S., VAKULENKO, S., MOBASHERY, S. & SAMAMA, J. P. 2000. Insights into class D beta-lactamases are revealed by the crystal structure of the OXA10 enzyme from Pseudomonas aeruginosa. *Structure,* 8**,** 1289-98.
MCDONOUGH, K. A. & RODRIGUEZ, A. 2012. The myriad roles of cyclic AMP in microbial pathogens: from signal to sword. *Nat Rev Microbiol,* 10**,** 27-38.
MEIGH, L. 2015. CO2 carbamylation of proteins as a mechanism in physiology. *Biochem Soc Trans,* 43**,** 460-4.
MEIGH, L., GREENHALGH, S. A., RODGERS, T. L., CANN, M. J., ROPER, D. I. & DALE, N. 2013. CO2 directly modulates connexin 26 by formation of carbamate bridges between subunits. *Elife,* 2.
MELDRUM, N. U. & ROUGHTON, F. J. 1933. Carbonic anhydrase. Its preparation and properties. *J Physiol,* 80**,** 113-42.
MIKOSHIBA, K. 2007. The IP3 receptor/Ca2+ channel and its cellular function. *Biochem Soc Symp***,** 9-22.
MILLER, S., POLLACK, J., BRADSHAW, J., KUMAI, Y. & PERRY, S. F. 2014. Cardiac responses to hypercapnia in larval zebrafish (Danio rerio): the links between CO2 chemoreception, catecholamines and carbonic anhydrase. *J Exp Biol,* 217**,** 3569-78.
MOORE, J. A., AND DONALD E. REED 1973. "Diazomethane." *Organic Syntheses* 16.
MORONEY, J. V., JUNGNICK, N., DIMARIO, R. J. & LONGSTRETH, D. J. 2013. Photorespiration and carbon concentrating mechanisms: two adaptations to high O2, low CO2 conditions. *Photosynth Res,* 117**,** 121-31.
MYERS, T. G. & NELSON, S. D. 1990. Neuroactive carbamate adducts of beta-N-methylamino-L-alanine and ethylenediamine. Detection and quantitation under physiological conditions by 13C NMR. *J Biol Chem,* 265**,** 10193-5.
NATTIE, E. 1999. CO2, brainstem chemoreceptors and breathing. *Prog Neurobiol,* 59**,** 299-331.
NOTNI, J., SCHENK, S., GORLS, H., BREITZKE, H. & ANDERS, E. 2008. Formation of a unique zinc carbamate by CO2 fixation: implications for the reactivity of tetra-azamacrocycle ligated Zn(II) complexes. *Inorg Chem,* 47**,** 1382-90.
NUNES, A. R., HOLMES, A. P., SAMPLE, V., KUMAR, P., CANN, M. J., MONTEIRO, E. C., ZHANG, J. & GAUDA, E. B. 2013. Bicarbonate-sensitive soluble and transmembrane adenylyl cyclases in peripheral chemoreceptors. *Respir Physiol Neurobiol,* 188**,** 83-93.
NURSE, C. A. & PISKURIC, N. A. 2013. Signal processing at mammalian carotid body chemoreceptors. *Seminars in Cell & Developmental Biology,* 24**,** 22-30.
O'LEARY M, H., JAWORSKI, R. J. & HARTMAN, F. C. 1979. C nuclear magnetic resonance study of the CO(2) activation of ribulosebisphosphate carboxylase from Rhodospirillum rubrum. *Proc Natl Acad Sci U S A,* 76**,** 673-5.
OLSEN, J. V., ONG, S. E. & MANN, M. 2004. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics,* 3**,** 608-14.
PAN, C., XU, S., ZHOU, H., FU, Y., YE, M. & ZOU, H. 2007. Recent developments in methods and technology for analysis of biological samples by MALDI-TOF-MS. *Anal Bioanal Chem,* 387**,** 193-204.
PARK, H., JUNG, Y. M., YOU, J. K., HONG, W. H. & KIM, J. N. 2008. Analysis of the CO2 and NH3 reaction in an aqueous solution by 2D IR COS: formation of bicarbonate and carbamate. *J Phys Chem A,* 112**,** 6558-62.
PARSONS, S. M., JAO, L., DAHLQUIST, F. W., BORDERS, C. L., JR., RACS, J., GROFF, T. & RAFTERY, M. A. 1969. The nature of amino acid side chains which are critical for the activity of lysozyme. *Biochemistry,* 8**,** 700-12.
PATERSON, A. K. & KNOWLES, J. R. 1972. The number of catalytically essential carboxyl groups in pepsin. Modification of the enzyme by trimethyloxonium fluoroborate. *Eur J Biochem,* 31**,** 510-7.

PATERSON, D. L. 2006. Resistance in gram-negative bacteria: Enterobacteriaceae. *Am J Infect Control,* 34**,** S20-8; discussion S64-73.
PERRELLA, M., KILMARTIN, J. V., FOGG, J. & ROSSI-BERNARDI, L. 1975. Identification of the high and low affinity CO2-binding sites of human haemoglobin. *Nature,* 256**,** 759-61.
PORTNER, H. O., LANGENBUCH, M. & REIPSCHLAGER, A. 2004. Biological impact of elevated ocean CO2 concentrations: Lessons from animal physiology and earth history. *Journal of Oceanography,* 60**,** 705-718.
RAVEN, J. A. 2006. Sensing inorganic carbon: CO2 and HCO3. *Biochem J,* 396**,** e5-7.
REDDY, S. K., KAMIREDDI, M., DHANIREDDY, K., YOUNG, L., DAVIS, A. & REDDY, P. T. 2001. Eukaryotic-like adenylyl cyclases in Mycobacterium tuberculosis H37Rv: cloning and characterization. *J Biol Chem,* 276**,** 35141-9.
ROH, K. S., J.K. KIM, S.D. SONG, H.S. CHUNG AND J.S. SONG 1996. Decrease of the activation and carbamylation of rubisco by high CO2 in kidney bean. *Korean J. Biotechnol. Bioeng,* 11**,** 295-302.
ROSSI-BERNARDI, L. & ROUGHTON, F. J. 1967. The specific influence of carbon dioxide and carbamate compounds on the buffer power and Bohr effects in human haemoglobin solutions. *J Physiol,* 189**,** 1-29.
SASSONE-CORSI, P. 2012. The cyclic AMP pathway. *Cold Spring Harb Perspect Biol,* 4.
SLYKE, V. D., D. AND HAWKINS, J,A. 1930. STUDIES OF GAS AND ELECTROLYTE EQUILIBRIA IN BLOOD XVI. THE EVOLUTION OF CARBON DIOXIDE FROM BLOOD AND BUFFER SOLUTIONS. *Journal of Biological Chemistry,* 87**,** 265-279.
SMITH, E. S., MARTINEZ-VELAZQUEZ, L. & RINGSTAD, N. 2013. A Chemoreceptor That Detects Molecular Carbon Dioxide. *Journal of Biological Chemistry,* 288**,** 37071-37081.
STADIE, W. C. 1935. The Role of the Carbamino Compounds in the Transport of Co2 by the Blood. *Science,* 81**,** 207-8.
STEC, B. 2012. Structural mechanism of RuBisCO activation by carbamylation of the active site lysine. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** 18785-18790.
STEEGBORN, C. 2014. Structure, mechanism, and regulation of soluble adenylyl cyclases - similarities and differences to transmembrane adenylyl cyclases. *Biochim Biophys Acta,* 1842**,** 2535-47.
STEEGBORN, C., LITVIN, T. N., LEVIN, L. R., BUCK, J. & WU, H. 2005. Bicarbonate activation of adenylyl cyclase via promotion of catalytic active site closure and metal recruitment. *Nat Struct Mol Biol,* 12**,** 32-7.
SUN, L., WANG, H., HU, J., HAN, J., MATSUNAMI, H. & LUO, M. 2009. Guanylyl cyclase-D in the olfactory CO2 neurons is activated by bicarbonate. *Proc Natl Acad Sci U S A,* 106**,** 2041-6.
TAYLOR, C. T. & CUMMINS, E. P. 2011. Regulation of gene expression by carbon dioxide. *Journal of Physiology-London,* 589**,** 797-803.
TESMER, J. J., SUNAHARA, R. K., GILMAN, A. G. & SPRANG, S. R. 1997. Crystal structure of the catalytic domains of adenylyl cyclase in a complex with Gsalpha.GTPgammaS. *Science,* 278**,** 1907-16.
THERRIEN, C. & LEVESQUE, R. C. 2000. Molecular basis of antibiotic resistance and beta-lactamase inhibition by mechanism-based inactivators: perspectives and future directions. *FEMS Microbiol Rev,* 24**,** 251-62.
TOUSSAINT, K. A. & GALLAGHER, J. C. 2015. beta-lactam/beta-lactamase inhibitor combinations: from then to now. *Ann Pharmacother,* 49**,** 86-98.
TOWNSEND, P. D., HOLLIDAY, P. M., FENYK, S., HESS, K. C., GRAY, M. A., HODGSON, D. R. & CANN, M. J. 2009. Stimulation of mammalian G-protein-responsive adenylyl cyclases by carbon dioxide. *J Biol Chem,* 284**,** 784-91.

VAN RIPER, S. K., DE JONG, E. P., CARLIS, J. V. & GRIFFIN, T. J. 2013. Mass spectrometry-based proteomics: basic principles and emerging technologies and directions. *Adv Exp Med Biol,* 990**,** 1-35.

VCELAKOVA, K., ZUSKOVA, I., KENNDLER, E. & GAS, B. 2004. Determination of cationic mobilities and pKa values of 22 amino acids by capillary zone electrophoresis. *Electrophoresis,* 25**,** 309-17.

WHISNANT, R. E., GILMAN, A. G. & DESSAUER, C. W. 1996. Interaction of the two cytosolic domains of mammalian adenylyl cyclase. *Proc Natl Acad Sci U S A,* 93**,** 6621-5.

YIN, Z. H., SIEBKE, K. & HEBER, U. 1991. Light-dependent pH changes in leaves of C3 plants : IV. Action spectra indicate indirect energization of proton transport into mesophyll vacuoles by cyclic photophosphorylation. *Planta,* 184**,** 30-4.

ZEEBE, R. E. 2012. History of Seawater Carbonate Chemistry, Atmospheric CO2, and Ocean Acidification. *Annu. Rev. Earth Planet. Sci,* 40**,** 141-65.