Organocatalyst-mediated cross aldol ligation of proteins

Richard James Spears

Doctor of Philosophy

University of York

Chemistry

June 2018

Abstract

The chemical modification of proteins to produce protein bioconjugates has revolutionised the field of chemical biology, with wide-ranging applications in cell biology and chemical medicine. Of particular note are bioconjugates linked by carbon-carbon (C-C) bonds, which are highly prized due to their hydrolytic stability. Current strategies for their synthesis, however, suffer from a number of practical limitations, such as utilising acidic/basic conditions, showing reduced reactivity in the presence of oxygen, and requiring large concentrations of chemical probe in high molar excess.

This thesis describes the design and development of a novel protein bioconjugation strategy, the Organocatalyst-mediated Protein Aldol Ligation, or "OPAL". During this project, the OPAL was conceived, designed, and rigorously optimised, and subsequently found to site-selectively modify proteins bearing previously installed α-oxo-aldehyde handles *via* hydrolytically stable C-C bonds at neutral pH within an hour using minimal molar equivalents of chemical probe through an organocatalyst-mediated process. The OPAL, along with the palladium mediated decaging of thiazolidine-containing unnatural amino acids, was found to be compatible with site-selective affinity tagging and protein pulldown in cell lysate. Additionally, the OPAL products could be site-selectively functionalised further using alternative protein aldehyde ligation strategies through modification of a newly generated β-hydroxy aldehyde handle that results from the initial OPAL modification. The OPAL strategy was applied to the chemical mimicry of lipidated Leishmania surface proteins previously inaccessible through conventional methods in molecular biology. Finally, within the context of protein bioconjugation, further aldol based C-C bond cleaving and forming protocols were designed and investigated.

Table of Contents

List of Tables

Chapter 7

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

List of Accompanying Materials

This thesis is accompanied by a USB flash drive, which contains the following:

- Raw data files for the mass spectrometry data for all instances described in this thesis. They are organised in folders by Chapter.

- Uncropped pictures of SDS-PAGE and Western Blot analyses described in this thesis

- Raw data files for the NMR data for instances described in this thesis.

- A Microsoft Excel spreadsheet containing the data used to generate the kinetic data described in Chapter 4 and in Chapter 9

Acknowledgements

Firstly, I would like to thank my supervisor Dr Martin Fascione for giving me the opportunity to undertake a PhD, his constant support throughout the project, for sharing with me his never-ending knowledge in both chemistry and biology, and for the insightful discussions we have had on all things science during the PhD. Who would have thought all those years ago (before the Fascione group even existed!) that after our first meeting, followed by a three hour "quick chat" (!) we would have achieved so much in the coming years. I would especially like to express my gratitude towards you for simultaneously allowing me to have creative freedom with the project (particularly when it came to pursuing the identity of the "mystery result" obtained near the beginning of the PhD), whilst also guiding me in the right direction when I had lost my way in the more challenging stages of the project. It has been a pleasure working in your group and I am sure great things will continue to come as a result of your supervision.

I would like to thank both the Fascione lab and Parkin lab (Martin, Alison, Julia, Jing, Harriet, Robin, Emily, Darshita, Tessa, Jenny, Lewis, Natasha, Nick, Mark, along with special mentions to lab alumni Hope, Lindsey, and Gintare) for making the last few years an immensely enjoyable experience, for their enthusiasm for science, and for being friends as well as work colleagues. I would especially like to thank PhD buddy Harriet Chidwick for her support, friendship, and everything in general; I honestly do not think I would have got through the PhD without you. Special mentions also go to Robin Brabham for providing me with much knowledge on phrases and politics I was previously oblivious to, Emily Flack for baking delicious brownies and cake, and for discussions on Yorkshire and the glorious county we hail from, Darshita Budhadev for her kindness, wisdom, sharing her knowledge on synthesis with me, and for her fantastic baking skills, Tessa Keenan sharing her knowledge in molecular biology with me, and sharing her wisdom on writing up the PhD, Nick Yates for being the best "lab son" anyone could ask for and for lively scientific discussions, Alison Parkin for being an inspirational scientist and introducing me to the weird and wonderful world of electrochemistry, and Julia Walton for being an amazing technician, being the "lab mum" to everyone in both groups, and being the glue that holds the lab together. I would also like to thank the

hamsters Rozza, Dozza, Cazza and Mazza for providing lots of cuddles during the course of the PhD.

I would like to extend my sincere gratitude towards Ed Bergström, who provided initial training in use of the mass spectrometry facilities and was always ready to answer any questions I had or repair any problems associated with the instruments; you are a lifesaver and this project would not have been possible without you. I would also like to thank Professor Peter O'Brien and Professor Gideon Grogan for useful discussions during TAP meetings.

I would like to thank my siblings Emily Spears, Sarah Spears, and Paul Spears, and my grandma Nancy Spears, for being the best family I could ask for, and for all their love and support. I would also like to thank my extended family the Forrest family for their love and support as well.

Finally, I would like to thank my mum Helen Spears and my dad Robert Spears for the eternal love, support, and encouragement they have provided throughout my life, for the valuable wisdom they have passed onto me, for all the clothes washing, cooked meals, and financial support they have provided over the years, for the many sacrifices they have made to ensure I have the best possible chance at life, and for making me the man I am today. This thesis is dedicated to you, and I love you both very much.

Authors declaration

I, Richard James Spears, declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References. Alternatively, sources are acknowledged either within the main text, notes in Chapter 8: Experimental, notes in Chapter 9: Appendix, or as references in Chapter 10. The work submitted for this thesis is my own, with the exception of work that has formed part of jointly authored publications and pre-print articles. The contributions made by myself and the other authors towards this work are explicitly stated below. I can confirm that appropriate credit has been given within thesis where reference has been made to work carried out by others:

- Chapter 1 contains content from the publication "Site-selective incorporation and ligation of protein aldehydes" Organic & Biomolecular Chemistry. 14, 7622-7638. [1] The manuscript was composed by R. J. Spears and M. A. Fascione.

- Chapter 2-9 contains content from the pre-print article "Site-Selective C-C Modification of Proteins at Neutral pH Using Organocatalyst-Mediated Cross Aldol Ligations" ChemRxiv, https://doi.org/10.26434/chemrxiv.5872461.v1, [2] which is also currently awaiting publication in Chemical Science. A list of acknowledgements is available with this article, which is repeated here: R. J. Spears. and M. A. Fascione. conceived and designed the work described in this publication; R. J. Spears performed all protein and peptide bioconjugations; D. Budhadev, R. J. Spears, and R. L. Brabham performed chemical synthesis; T. Keenan, R. J. Spears, R L. Brabham, S. McKenna, and J. Walton prepared and characterised proteins, J. A. Brannigan constructed plasmids; S. McKenna and M. J. Plevin performed protein NMR experiments. A. M. Brozozowski, A. J. Wilkinson, M. J. Plevin, and M. A. Fascione supervised the project, and R. J. Spears. and M. A. Fascione composed the manuscript. All authors analysed the data and commented on the paper.

- In Chapter 2, the GFP **33**, sfGFPN150ThzK **39** and GFPY39ThzK **40** were kindly provided by Robin Brabham. The HASPA **35** was kindly provided by Sophie McKenna. The HASPA(G1S) **37** was kindly provided by Tessa Keenan.

- In Chapter 4, dipeptide **81**, NMR/HPLC characterisation of aldol product **83**, and linker **84** were kindly provided by Darshita Budhadev.

- In Chapter 5, indole **104** was kindly provided by Darshita Budhadev.

- In Chapter 6 *N*-myristoylated HASPA **124** was provided by Sophie McKenna, the HASPA NMR experiment and characterisation/processing was performed by Michael Plevin, and the liposome preparation, binding assays, and characterisation were performed by Tessa Keenan. Information on the protein concentration determined by SDS-PAGE was kindly provided by Dr Tessa Keenan.

- In Chapter 8, the plasmid for CTB was provided by Thomas Branson, precursor **135** and **136**, along with HPLC data for peptides was provided by Robin Brabham, and several NMR spectra were recorded and/or processed by Darshita Budhadev. HPLC data for SVARLG **150** was recorded by Amanda Dixon.

In loving memory of my dad Robert Spears, from whom I learned much, and loved a lot.

Chapter 1: Introduction

1.1. Protein bioconjugation

Protein bioconjugation is a term used to collectively describe chemical strategies for linking synthetic groups to proteins *via* covalent bonds.[3] Over the last 25 years (and particularly within the last decade), protein bioconjugation has generated major interest within the field of chemical biology, where the chemical ligation of functional molecules to proteins can vastly enhance their properties and allow for wide-ranging applications (Figure 1.1). For example, conjugation of compounds such as polyethyleneglycol (PEG) to proteins has been well established for improving the half-life of protein probes and therapeutics,[4] with several PEGylated protein therapeutics approved by the U.S Food and Drug Administration (FDA). These drugs include PEGylated interferon alfa-2a (Pegasys) used to treat hepatitis B and C, PEGylated anti-tumour necrosis factor (Certolizumab pegol, or Cimzia) used to treat Crohn's disease and rheumatoid arthritis, and PEGylated asparginase (Oncaspar, or pegaspargase) used to treat leukemia.[5] PEGylation of proteins has also been demonstrated in the preparation of protein-based nanoparticles, which can encapsulate and carry cytotoxic compounds for use in drug delivery. [6] Alternatively, preformed nanoparticles can be functionalised with proteins using bioconjugation strategies for potential use as biomedical diagnostic tools. [7,8] Bioconjugation is also routinely used for covalent immobilisation of proteins to surfaces for applications such as surface plasmon resonance (SPR). [9] Ligation of fluorescent/spectroscopic probes has been utilised for *in vivo* imaging of proteins and other biomolecules, such as tracking glycosylated proteins in developing zebrafish, giving insights into glycan distribution during embryogenesis,[10,11] whereas ligation of affinity tags allows for selective protein pulldown in complex biological mixtures.[12] Additionally, pioneering advances in modern medicine have been achieved through protein bioconjugation, primarily in the construction of antibody-drug conjugates (ADCs) by linking therapeutic antibodies with small molecules drugs *via* chemical ligation strategies.[13] Four ADCs are currently approved by the FDA for therapeutic use; these include trastuzumab emtansine (Kadcyla) used to treat breast cancer and approved for use on the National Health Service (NHS), brentuximab vedotin (Adcetris) used to treat Hodgkin lymphoma, inotuzumab ozogamicin (Besponsa) used to treat acute lymphoblastic leukemia, and gemtuzumab ozogamicin (Mylotarg) used to treat acute myeloid leukemia, with many more in the clinical pipeline. [14] From a cell biology

perspective, the ability to functionalise proteins at specific sites also shows particular promise in the design of new tools for probing post-translational modifications (PTMs), giving access to proteins bearing PTM "mimics".[15] For example, *N*-glycans can be chemically re-installed into non-glycosylated antibodies,[16] whereas histone proteins can be functionalised with a range of different PTM mimics in a process known as "chemical mutagenesis". [17,18]
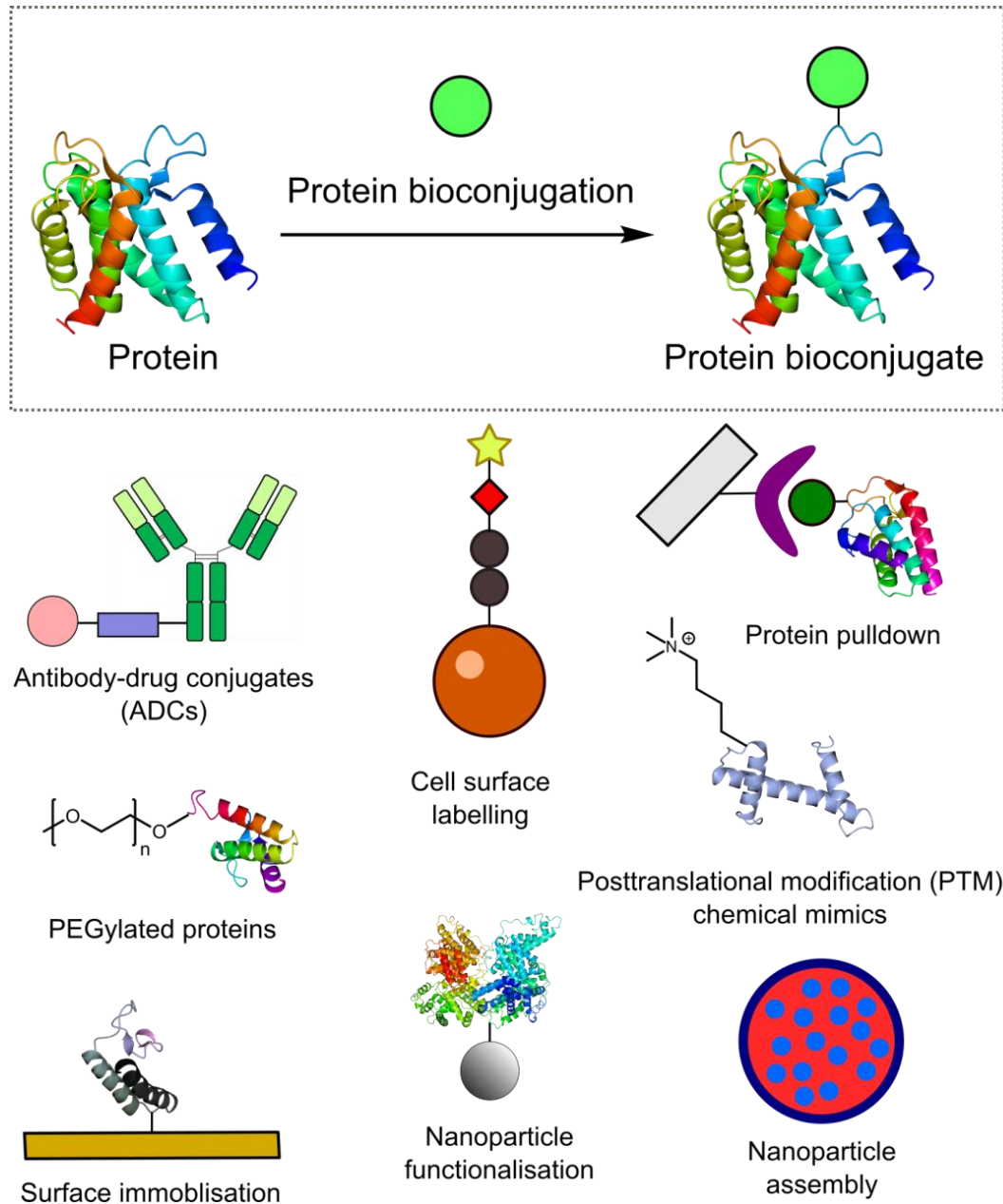
Figure 1.1 Outline of protein bioconjugation and its wide ranging applications in chemistry and biology.

1.1.1 Requirements and challenges of protein bioconjugation

The ultimate goal of protein bioconjugation is akin to that of small molecule organic synthesis; to assemble new compounds by making chemical bonds from one molecule to another. For protein bioconjugation, however, additional considerations must be taken into account in comparison to more traditional approaches to synthesis (Figure 1.2). Proteins are large polypeptides consisting of a sequence of amino acids (primary structure), which adopt local conformations (secondary structure) that fold into a more complex three dimensional structure (tertiary structure). Some proteins also comprise several polypeptide chains that form protein subunits (quaternary structure). Many proteins are capable of carrying out particular functions based on their aforementioned three dimensional structure; enzymes, for example, catalyse reactions and turn over substrates. It is vital, therefore, that any activity of the protein target in question is maintained post modification *i.e.* bioconjugation does not compromise the target proteins original function. Broadly speaking, this requires protein bioconjugation methodology to be compatible with conditions tolerated by proteins, such as having an aqueous solvent system, buffered conditions with a pH that is neutral (or near neutral), and mild temperatures typically ranging between 4 - 37 °C.[19] Additionally, protein bioconjugation strategies that selectively modify a particular site of the target protein in a controlled manner *via* stable linkages are highly prized, as they avoid generating mixtures of bioconjugation products. As proteins comprise amino acids, all protein bioconjugation must also be carried out in an environment containing a vast array of functional groups (a so-called "sea of functionality"),[20] with limited access to protecting group chemistry.
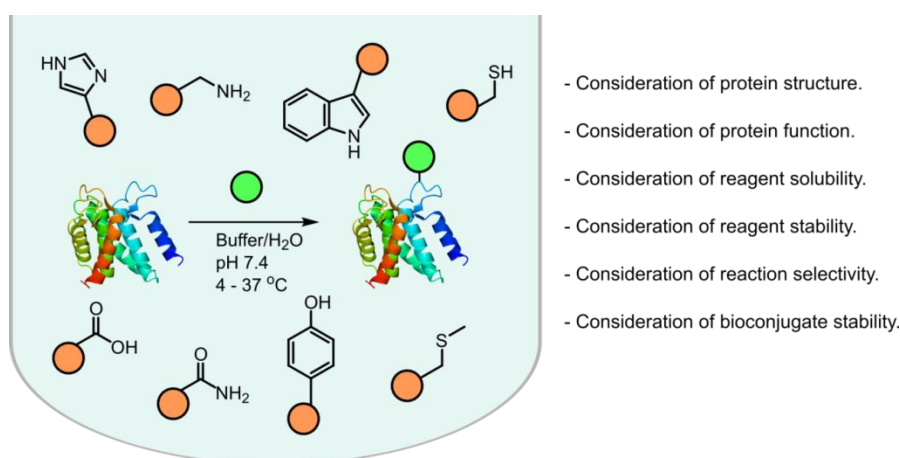
Figure 1.2 Consideration of reaction conditions for the design of successful, synthetically useful protein bioconjugation strategies.

1.1.2. Brief overview of current strategies for protein bioconjugation

From a synthetic perspective, the restrictive nature of the system in which protein bioconjugation must be performed presents significant chemical challenges. Despite these challenges, there are now multiple methods that have been reported in the literature. [19,21,22] Many chemical protein modification strategies involve reaction of a reactive amino acid side chain, the most common of which focus on the ε-amino group of a lysine residue, or the thiol group of a cysteine. Lysine residues, for example, can be covalently modified at the ε-amino group with activated ester compounds to form amide bonds (Figure 1.3A),[23] with isothiocyanates to form ureas (Figure 1.3B),[24] or with aldehyde compounds to form imines that can be subsequently reduced *via* reductive amination (Figure 1.3C). [25] In particular, lysine modification *via* amide bond formation with succinimidyl trans-4-(maleimidylmethyl)cyclohexane-1-carboxylate (SMCC) **1** is instrumental in the two-step preparation of Kadcyla (**2**, Figure 1.3D).[26] Examples of cysteine conjugation include alkylation with iodoacetamides (Figure 1.4A),[27] coupling with maleimides (Figure 1.4B),[28] coupling to allenamides (Figure 1.4C),[29] palladium mediated arylation (Figure 1.4D),[30] and modification with carbonylacrylic reagents (Figure 1.4E) [31] Maleimide coupling, in particular, is used in the synthesis of Adcetris (Figure 1.4F).[26] Cysteine residues can also be converted to dehydroalanine following treatment with α,α′-di-bromo-adipyl(bis)amide,[32] further increasing the range of cysteine-related bioconjugation strategies. The alkene side chain of dehydroalanine is primed to react with thiol based reagents to form thioether bonds,[32] or alkylated to form carbon-carbon bonds *via* addition of alkyl radicals (generated from alkyl halides) in a metal mediated process (Figure 1.5).[17,18] Much of dehydroalanine chemistry has underpinned the synthesis of proteins bearing PTM mimics.[17,18,32] Disulfide bridged cysteines on the other hand can undergo reduction followed by addition[33] (Figure 1.6A) or dibromopyridazinedione addition[34] (Figure 1.6B), or alternatively a one-step reduction-rebridging modification using tris(2-carboxyethyl)phosphine (TCEP)-derived pyridazinediones **5** (Figure 1.6C). [35] Alongside lysine and cysteine modifications, tyrosine side chains can be modified *via* coupling hemin activated luminol reagents (Figure 1.7A), [36] with diazonium chemistry (Figure 1.7B), [37] and with Mannich-type reactions using aniline derived probes and formaldehyde **6** (Figure 1.7C). [38] Additionally, tryptophan side chains act

as reactive groups in rhodium carbenoid mediated reactions (Figure 1.7D), [39] and in organo-radical mediated reactions (Figure 1.7E).[40] Methionine residues can be modified *via* redox–activated chemical tagging using oxaziridine reagents (also known as the "ReACT" strategy (Figure 1.7F).[41] These methods, along with other strategies for the chemical modification of endogenous amino acids, have been extensively reviewed.[21,22]
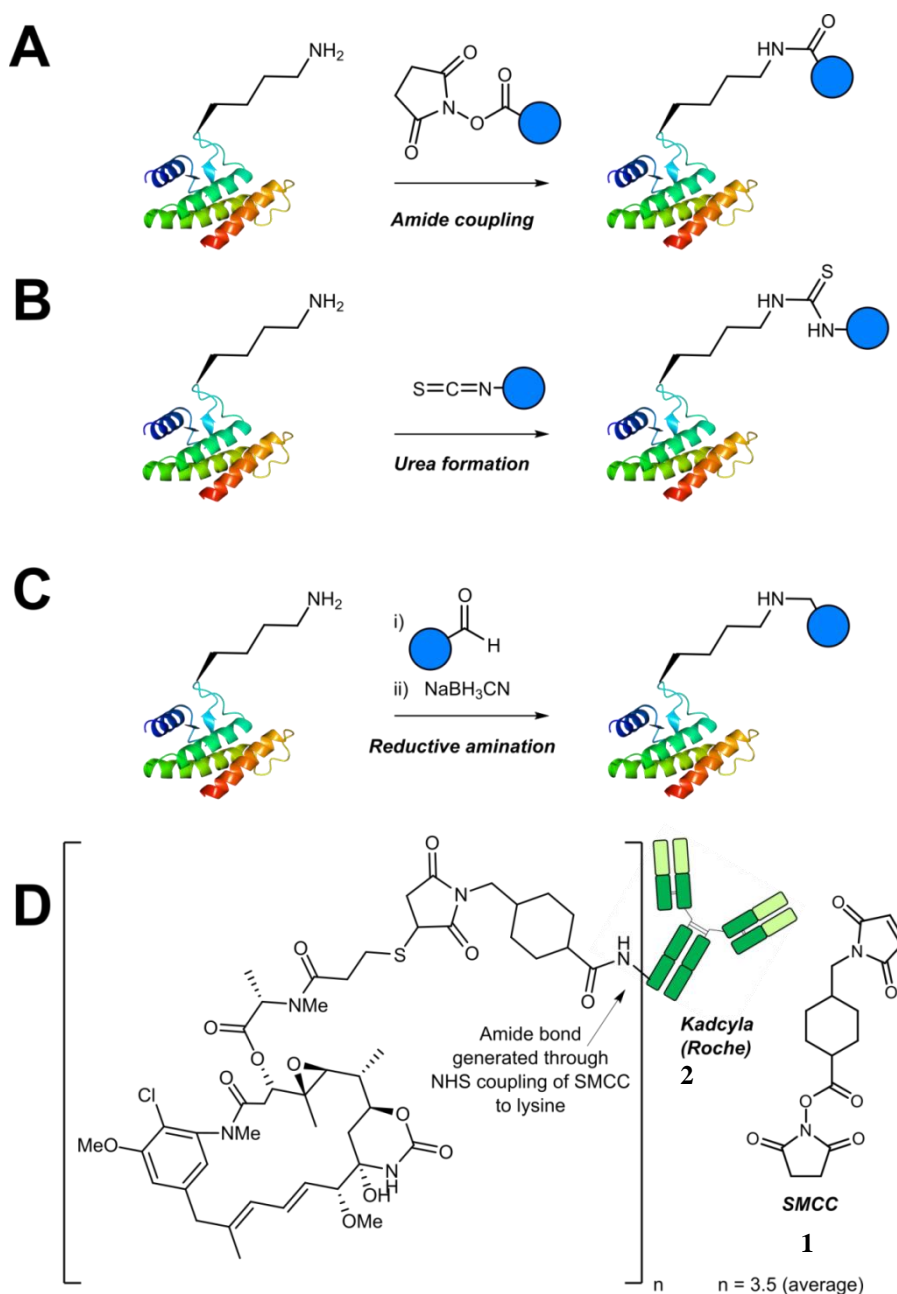
Figure 1.3 Modification of the ε-amino group of lysine. A) Amide bond formation. B) Urea formation. C) Reductive amination. D) Structure of Kadcyla, highlighting the amide bond generated through lysine modification that conjugates the cytotoxic payload to the antibody.
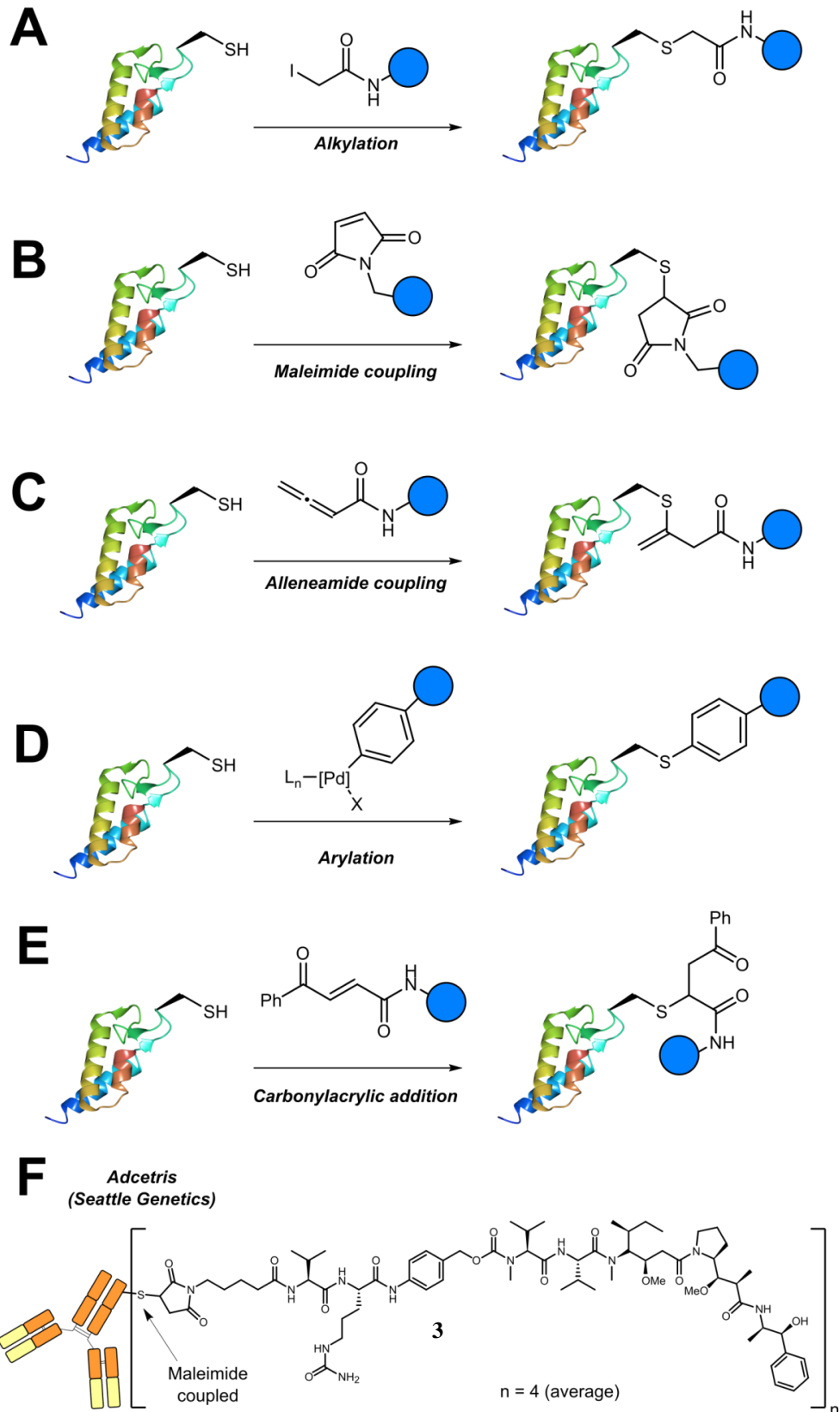
Figure 1.4 Modification of thiol side chain of cysteine. A) Alkylation. B) Maleimide coupling. C) Allenamide coupling. D) Arylation. E) Michael addition of carbonylacrylic reagents. F) Structure of Adcetris, highlighting the thioether bond generated through cysteine modification that conjugates the therapeutic payload to the antibody.
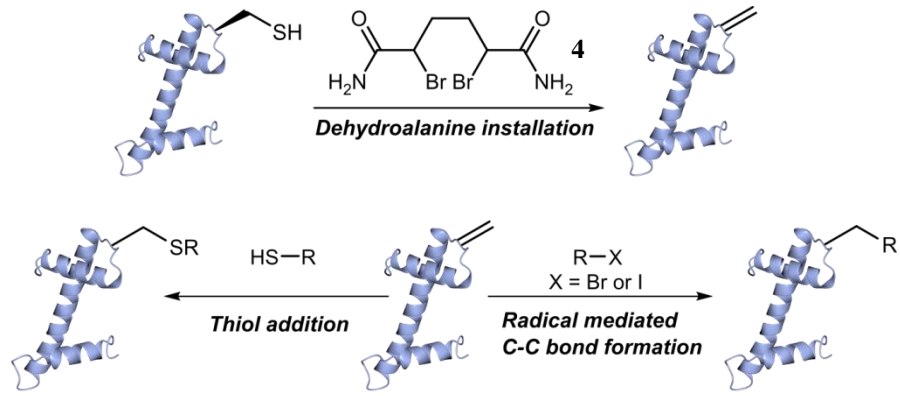
Figure 1.5 Conversion of cysteine to dehydroalanine, and modification via thioether formation or alkylation.
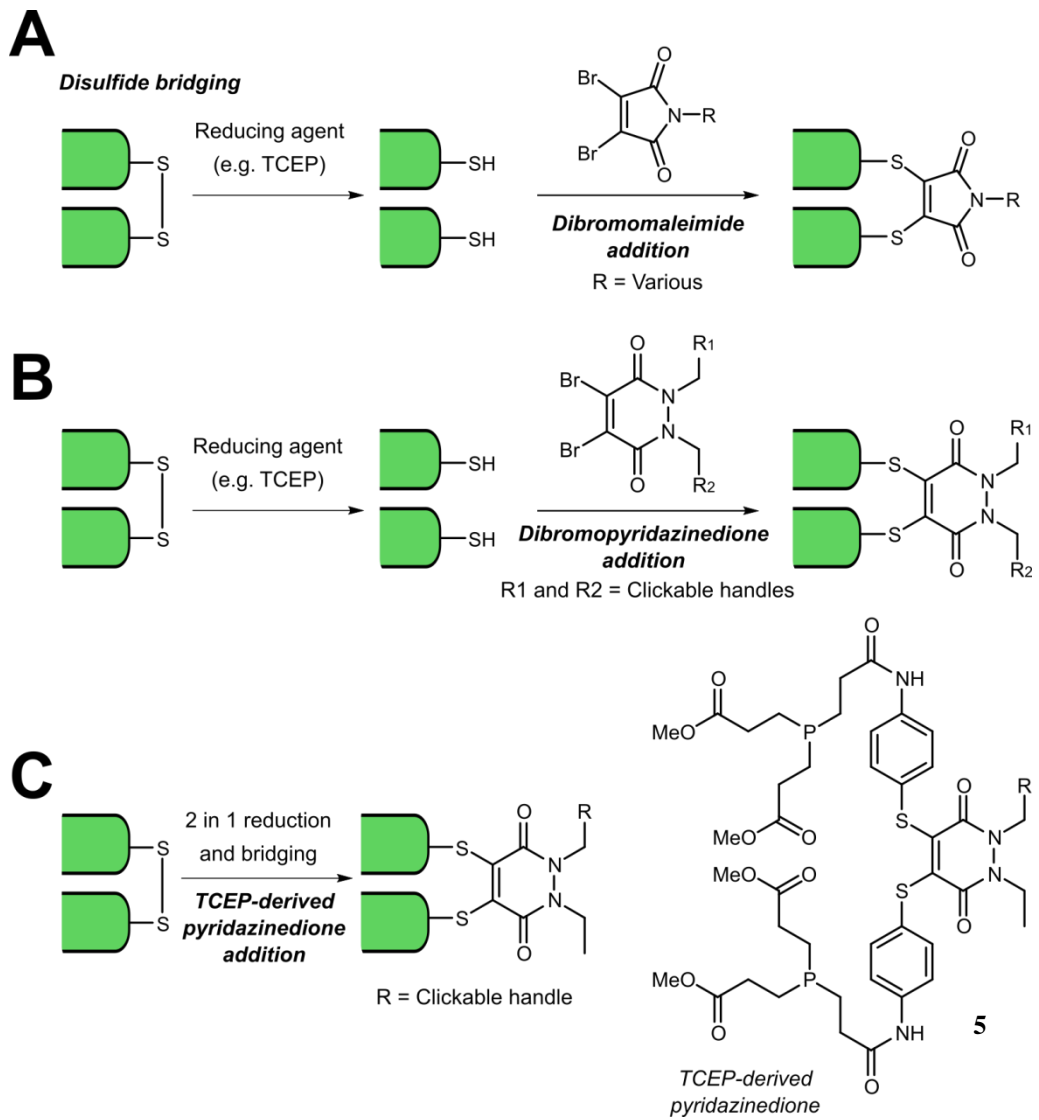
Figure 1.6 Disulfide bridging. A) Reduction, followed by dibromomaleimide addition. B) Reduction, followed by dibromopyridazinedione addition. C) Two in one reduction-bridging using TCEP-derived pyridazinediones.
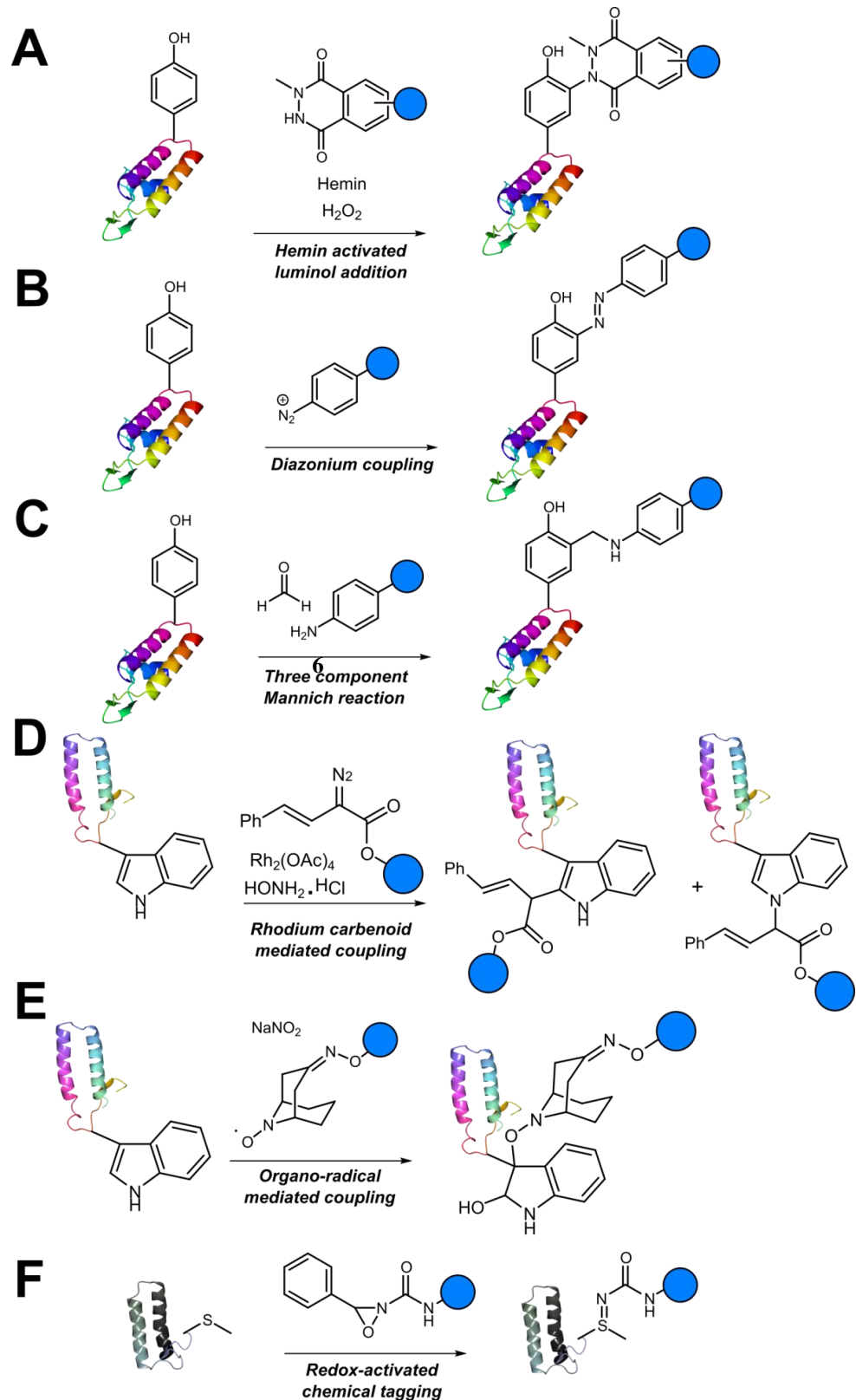
Figure 1.7 Other protein modifications of naturally occurring amino acid side chains . A) Modification of tyrosine with activated luminols B) Modification of tyrosine with diazonium salts. C) Modification of tyrosine via a three-component-Mannich reaction. D) Modification of tryptophan with rhodium carbenoids. E) Modification of tryptophan with organo-radicals. F) Modification of methionine with oxaziridine reagents.

All of these methodologies share one common drawback however, in that site-selectivity cannot be achieved when more than one of these naturally occurring groups is present within a protein. There are exceptions to this; for example, perfluoraromatic reagents can be used to site-selectively modify a single cysteine residue located within a "π-Clamp", a four amino acid sequence comprising of Phe-Cys-Pro-Phe (Figure 1.8 A). [42] Selectivity of the π-clamp strategy is thought to result from energetically favourable interactions between the hydrophobic amino acids surrounding the cysteine residue within the π-clamp, and the hydrophobic perfluroaromatic reagents used in the reaction.[42,43] Equally, site selective lysine modification has been demonstrated through initial modification of a single lysine residue with sulfonyl acrylate **7**, followed by aza-Michael addition (Figure 1.8B); the selectivity of this procedure for a given lysine residue is thought to result from the given residue bearing a lower-than-usual ε-amino group pKa value due to local microenvironment, making that given lysine residue more reactive than other lysines present in the protein. [44] Selective, although incomplete, labelling of lysine can also be achieved through kinetically controlled amide formation with activated esters.[23]
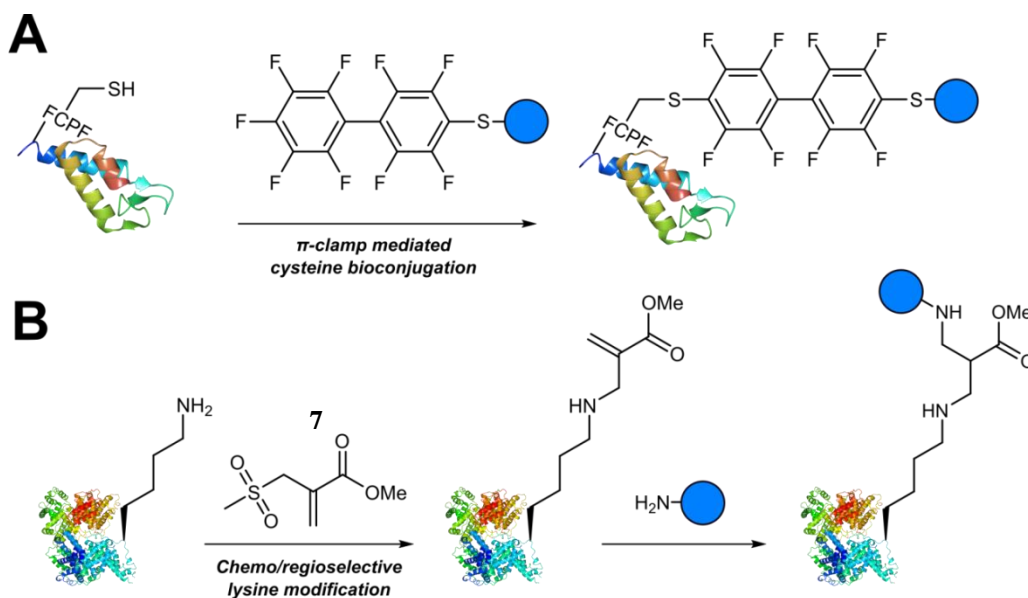
Figure 1.8 Site-selective protein modification of naturally occuring amino acid side chains. A) π-clamp mediated modification of cysteine. B) Modification of lysine using sulfonyl acrylates, followed by aza-Michael addition.

The *N*-terminal residue of proteins often provides a native site suitable for site-selective modification. For single chain proteins, there is only one of these residues,

and the α-amino group displays a different pK$_a$ (~6-8) compared to that of the ε-amino groups of lysine (~10.5), and therefore different reactivity at low-neutral pH.[45] In addition to modification *via* strategies previously described for non-selective lysine modification, the *N*-terminal group of proteins can be site-selectively acylated with ketenes (Figure 1.9A),[46] and undergo site-selective imidazolidinone formation using 2-pyridinecarboxaldehyde (2PCA) reagents (Figure 1.9B) .[47] *N*-terminal cysteines can also be selectively targeted using 2-cyanobenzothiazoles (Figure 1.9C),[48] and are also used in native chemical ligation (Figure 1.9D) ,[49] whilst *N*-terminal proline residues can react with *o*-aminophenols in an oxidative coupling process, using potassium ferricyanide as the oxidant (Figure 1.9E) [50].
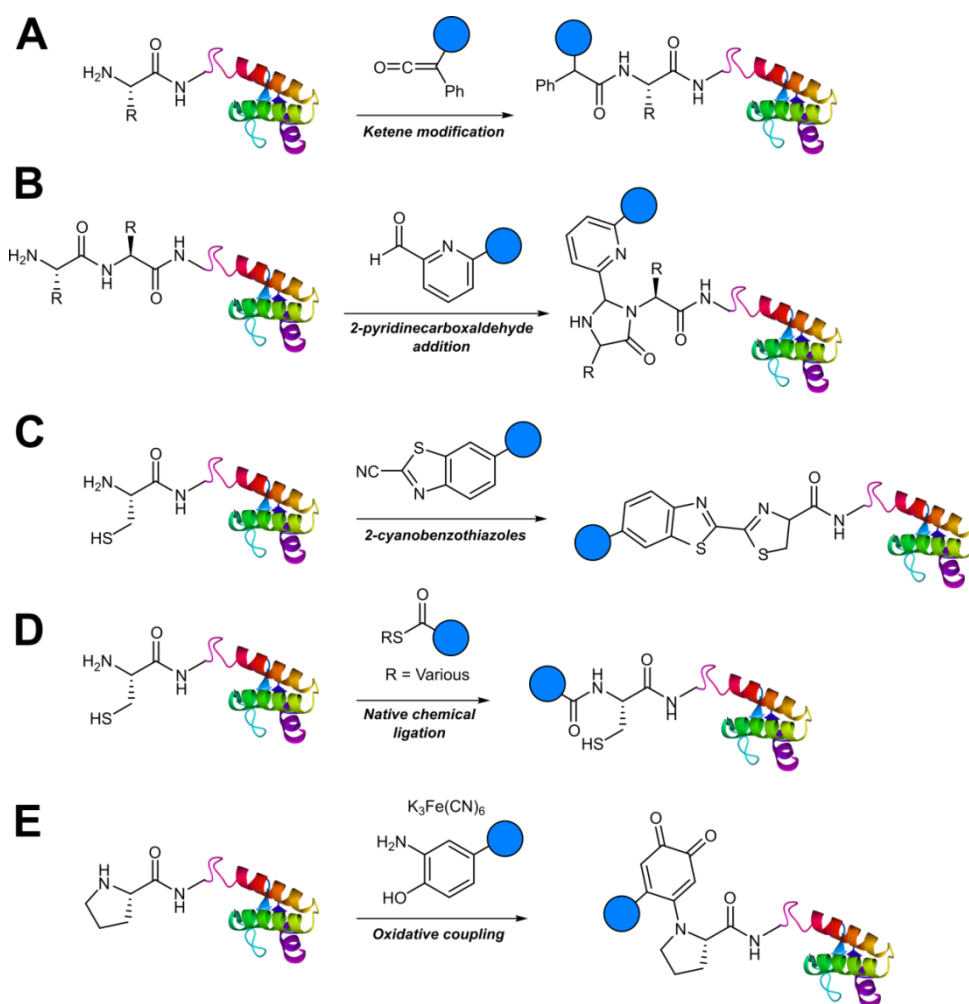
Figure 1.9 Selective *N*-terminal protein modifications A) *N*-terminal modification with ketene. B) *N*-terminal modification with 2PCA reagents. C) *N*-terminal modification with 2-cyanobenzothiazoles. D) Native chemical ligation of *N*-terminal cysteine. E) Oxidation coupling of *o*-aminophenols to *N*-terminal proline.

In a drive to avoid the selectivity issues encountered from targeting naturally occurring amino acid side chains, recent years have seen a growing number of protein labelling techniques emerging that modify non-naturally occurring functionalities in proteins. These strategies primarily focus around incorporating a unique functional 'handle' into a protein that is not naturally present within the proteinogenic amino acid side chains, and then subjecting the tagged protein to a substrate that bears a reactive moiety that specifically reacts with the incorporated handle. Such reactions do not interact with other functional groups present in the protein, therefore generating homogenous bioconjugates. This area of protein bioconjugation has primarily been enabled by advance in methods that allow for the incorporation of unnatural amino acids into proteins, otherwise known as unnatural amino acid mutagenesis, or genetic code expansion. The field of unnatural amino acid mutagenesis has been the subject of several reviews in the literature.[51-53] Arguably the most popular non-naturally occurring handles used in protein bioconjugation are the azide and alkyne functionalities. Azides act as reactive partners with phosphines in the Staudinger ligation (Figure 1.10A) [54] and its traceless variant,[55] and also with alkyne substrates in copper-mediated azide-alkyne cycloadditions (CuAAC, Figure 1.10B).[56] Similarly, azides can also take part in copper-free [3+2] cycloadditions with cyclooctynes, known as strain-promoted azide-alkyne cycloaddition (SPAAC, Figure 1.10C).[57] The latter is more applicable to biological systems, since it does not require the use of copper which exhibits cytotoxicity in cells at high concentrations. Strain-promoted cycloadditions with quinones have also been reported. [58] By extension, alkyne and cyclooctyne handles incorporated into proteins can also participate in CuAAC and SPAAC respectively using azide substrates. Alkynes can also act as handles for metal-mediated cross coupling reactions (Figure 1.10D). [59,60] More recently, terminal alkynes have been demonstrated as handles in copper-mediated Glaser-Hay bioconjugations.[61,62] Rapid protein labelling can also be achieved with strained alkene handles by performing inverse-demand Diels-Alder reactions with tetrazine substrates (Figure 1.10E); similar reactions can also be performed on alkyne handles.[11,63,64] Alongside recombinant protein labelling, these strategies have also been utilised in live cell labelling, and labelling of glycoproteins in complex biological environments, including live Zebrafish embryos,[10,11] without perturbing the native environment. This branch of bioconjugation is known as "bioorthogonal chemistry". These

strategies, alongside other methods for modification of unnatural amino acids, have been extensively reviewed. [15,19]
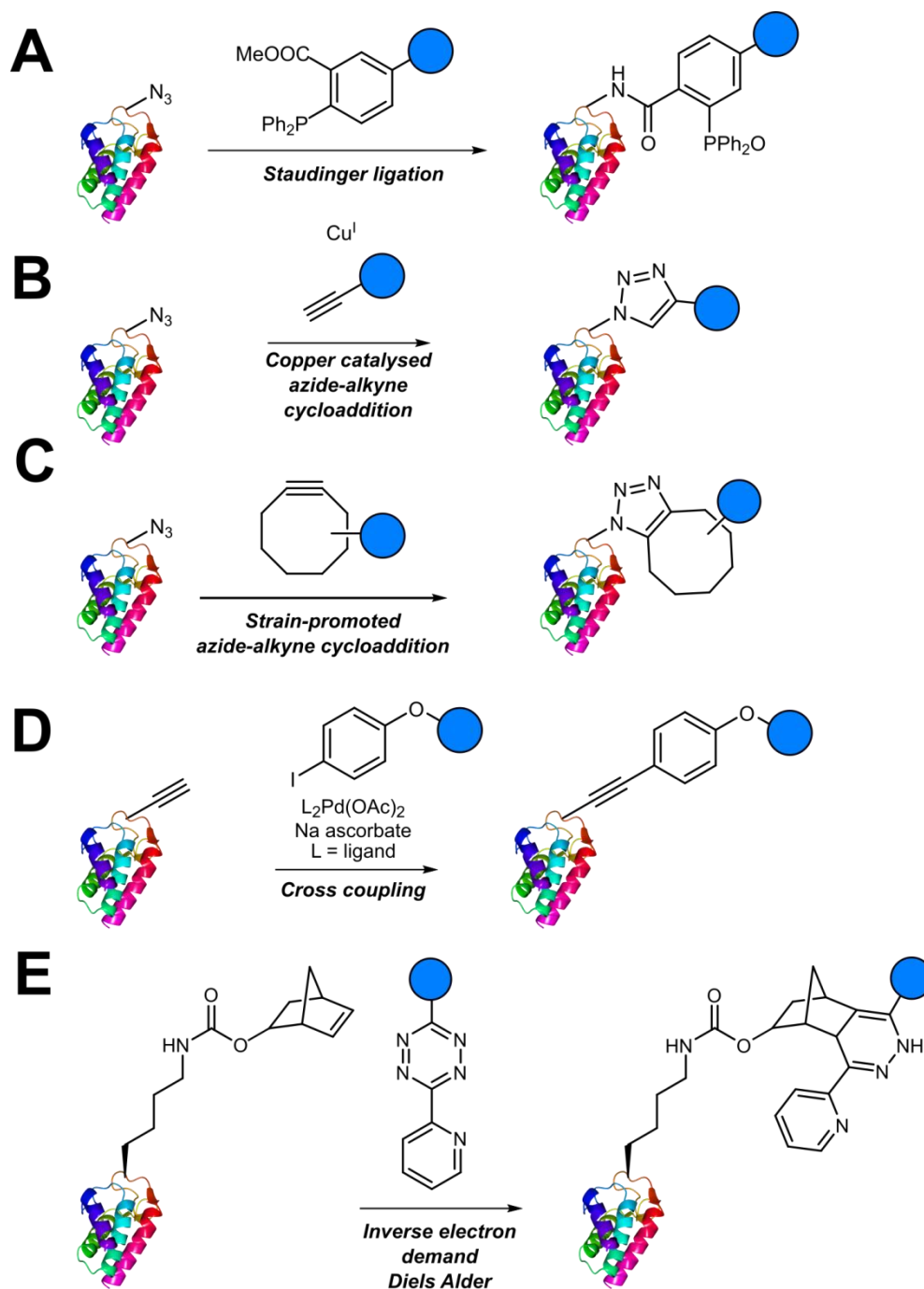
Figure 1.10 Modification of unnatural amino acid containing proteins. A) Staudinger ligation. B) CuAAC. C) SPAAC. D) Cross coupling E) Tetrazine coupling of strained alkene handles.

1.2 Site-selective incorporation of protein aldehydes

One functional group not traditionally found within proteins is the aldehyde functionality. This reactive carbonyl moiety has wide ranging applications in traditional organic synthesis and can participate in a variety of synthetically useful reactions, including synthesis of native peptide bonds.[65-67] Aldehydes also feature notably in protein semisynthesis protocols, including serine/threonine ligations at C-terminal salicylaldehyde esters,[68,69] and a variant of the α-ketoacid-hydroxylamine (KAHA) ligation where incorporated aldehyde side chains can act as purification handles.[70] It is therefore of no surprise that the aldehyde functionality has seen frequent use within the field of site-selective protein modification. Upon installation of an aldehyde handle, it is possible to perform chemistry that reacts with the aldehyde group to site-selectively label a target protein. Many strategies exist for the incorporation an aldehyde handle into a desired protein, along with an increasing selection of ligation protocols for subsequent site-selective labelling. Of particular note is the significant number of chemical probes targeting these handles that are now available commercially. Numerous examples of aldehyde-based ligations have been reported in the literature, including synthesis of antibody-drug conjugates (ADCs), design of proteins bearing PTM 'mimics', and labelling of specific proteins within crude cell lysates (Figure 1.11). This section will provide an overview of the strategies currently reported for incorporating aldehydes into proteins, alongside reported ligation methods exploited in their subsequent chemical modification.
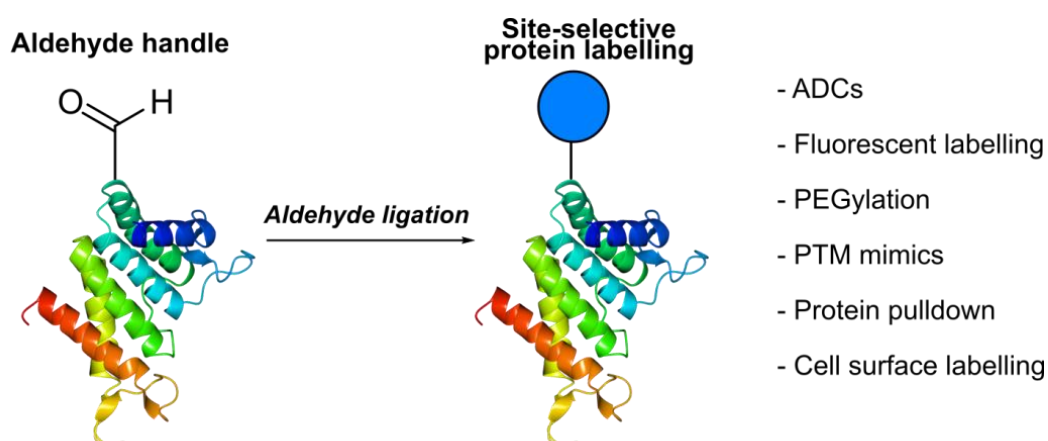
Figure 1.11 Overview of site-selective ligation of an aldehyde incorporated into a protein.

A wide variety of strategies can be employed for the incorporation of aldehyde handles into proteins. It is worth noting at this point that, although all strategies described incorporate an aldehyde functionality, the electronic properties of the aldehyde installed can vastly differ depending on which strategy is used. These range from α,β-unsaturated aldehydes to aryl aldehydes, as well as electron rich to electron deficient aldehydes. The tag can also differ in location and distance from the protein it is attached to. Some of the methods described for installing aldehyde handles into proteins can also be used to install ketone handles into proteins; these will be briefly considered where appropriate.

1.2.1 Oxidation of β amino-alcohols and vicinal diols *via* periodate cleavage

Aldehydes can be installed into proteins via oxidative cleavage of serine and threonine residues located on the *N*-terminus of a protein (Figure 1.12A).[71] This can be achieved through reaction with sodium periodate ($NaIO_4$) and is analogous to vicinal diol cleavage (also known as the Malaprade reaction) [72]. The resulting electron-poor *N*-terminal aldehyde is referred to as an α-oxo aldehyde or glyoxylamide (or glyoxylylamide). An extensive review on glyoxylamides, along with their carboxylic acid counterpart (glyoxylic acids), has recently been published. [73] A particular advantage of generating an aldehyde *via* this method is that the reaction is rapid; oxidation of β amino alcohols is 100-1000 times faster than the oxidation of vicinal diols encountered in carbohydrates, [74] and, from a practical perspective, typically gives complete conversion to the desired aldehyde product within 5-60 minutes. In a similar fashion, aldehydes can also be installed into proteins containing carbohydrate groups (i.e. proteins bearing sialic acid capped *N*-glycans) *via* periodate mediated diol cleavage.
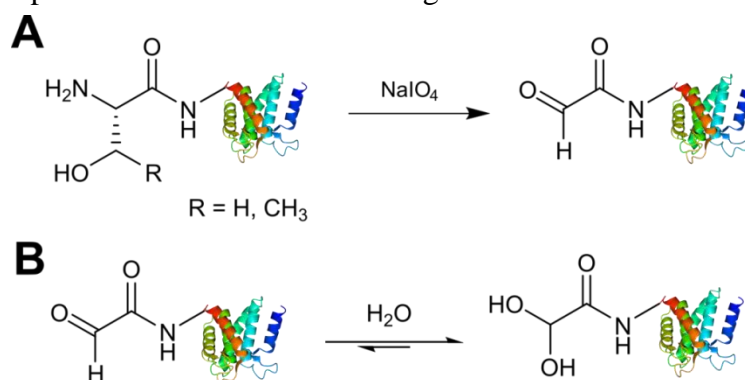
Figure 1.12 (A) Oxidation of serine/threonine N-terminated proteins to generate a glyoxylamide (glyoxyl) species. (B) The aldehyde and hydrate form of the glyoxyl species.

Care must be taken when using this method however as other amino acids can be oxidised under certain conditions, particularly cysteine and methionine residues. [73] For example, periodate oxidation of the peptide SYSMEHFRWG at pH 4.0 results in undesired oxidation of the methionine residue to methionine sulfoxide, whereas oxidation at pH 7.0 gives the desired glyoxyl-YSYMEHFRWG peptide.[75] Therefore, the reaction is usually performed at neutral pH, with another component added, such as excess methionine or ethylene glycol, during or after oxidation to quench unreacted periodate. This is then followed by purification of the oxidised protein,[76,77] which exists in an equilibrium as a mixture of aldehyde and hydrate form (Figure 1.12B).

1.2.2 *N*-terminal transamination strategies

Transamination refers to the transfer of the amino group of an amino acid to a keto acid, and was first established in the 1930s.[78] In nature, this process is catalysed by transaminases, with different transaminases existing for different amino acids.[78] A biomimetic transamination reaction for the installation of both aldehydes and ketones into proteins has also been described in the literature (Figure 1.13A). An initial screen of aldehyde reagents on the peptide antiogensin I identified pyridoxal 5-phosphate (PLP) **8** as a successful reagent for the oxidation of the *N*-terminal aspartate residue (generating a ketone) under mild conditions (pH 6.5, 25 mM phosphate buffer, 18 hours, 37 °C). The *N*-terminal ketone generated from this biomimetic transformation can then be site-selectively modified further. PLP-mediated transamination was successfully demonstrated on a number of proteins bearing different amino acids at their *N*-terminus, including conversion of *N*-terminal glycine to a glyoxylamide in horse heart myoglobin, and oxidation of *N*-terminal lysine to a ketone in RNAase A.[79] The transamination is specific for the amino group of the *N*-terminal amino acid, and does not modify other amine bearing residues such as lysine side chains. Optimisation of PLP-mediated transamination has since undergone extensive investigation utilising solid phase peptide synthesis (SPPS) [80] and combinatorial peptide libraries.[81] PLP **8** mediated transamination of resin bound peptides identified that sequences containing alanine, glycine, aspartate, glutamate, or glutamine residues at the *N*-terminus were transformed into the desired transamination product in the highest yields. Histidine, lysine, tryptophan and proline N-termini containing peptides gave rise to PLP-adduct-type byproducts,

whereas isoleucine and valine *N*-terminated peptides displayed very poor reactivity in PLP-mediated transamination.[80]
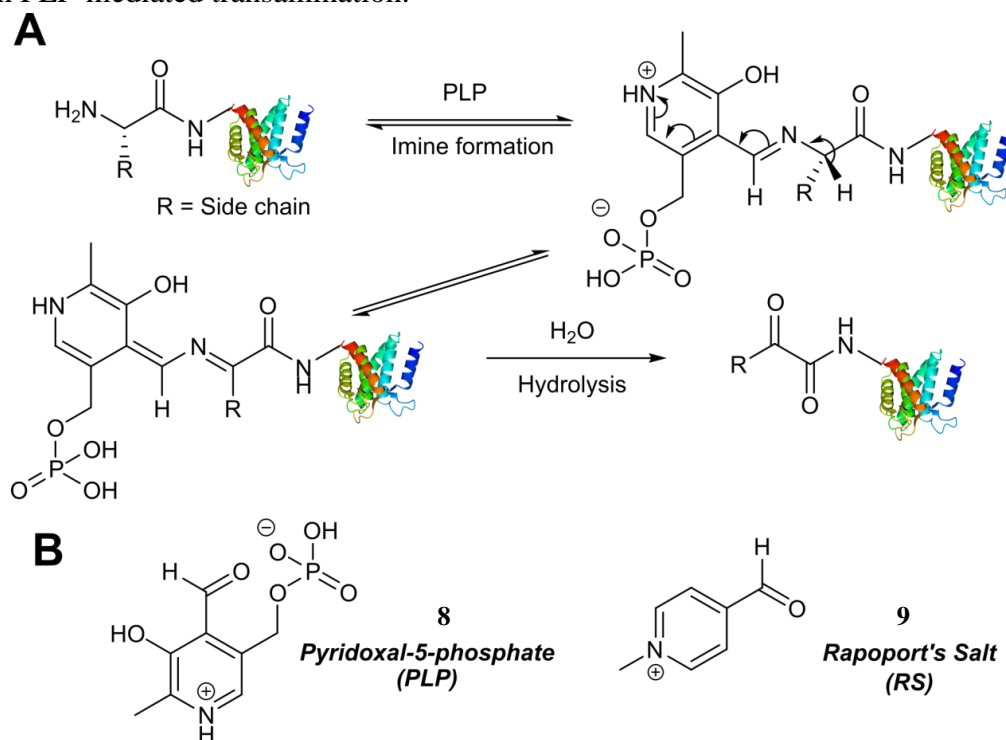
Figure 1.13 (A) Oxidation of the N-terminus using pyridoxal-5-phosphate (PLP) to generate a reactive aldehyde (or ketone). (B) The structure PLP and Rapoport's Salt (RS).

The observation that the valine *N*-termini are unsuitable for transamination is contradictory to the previously reported successful transamination of valine *N*-terminal green fluorescent protein (GFP1V).[79] This was rationalised on the basis that the internal sequence of a peptide/protein may also affect the success of PLP-mediated transamination. Additional screening of resin bound lysine and valine *N*-terminal tetrapeptides demonstrated that the identity of the amino acid at position 2 and 3 can dramatically affect the outcome of the transamination reaction.[80] Combinatorial peptide libraries screens have subsequently revealed that alanine *N*-terminated peptides/proteins, directly followed by a lysine residue, are highly reactive towards PLP-mediated transamination.[81] Transamination of short isoleucine, leucine, and valine *N*-terminal peptides using another pyridoxal derivative, along with the effects of using an imidazole-based buffer system on the transamination reaction, have also been recently reported. [82] An alternative reagent for transamination-mediated reactions on proteins has been reported more recently.[83] Using previously described combinatorial peptide libraries, [81] *N*-methylpyridinium-4-carboxaldehyde **9** (Rapoport's Salt, 'RS', see Figure 1.13B) has since been

identified as an effective reagent for transamination-type modification of proteins; in particular, RS **9** is highly suitable for glutamate *N*-terminated peptides/proteins. [83] Optimisation of the transamination using SPPS has demonstrated that RS is a poor transamination reagent for glycine, serine, threonine, or proline *N*-terminated peptides, but is very suitable for alanine, and glutamate *N*-terminated peptides. Additionally, it is worth noting that for all *N*-terminated amino acids discussed (with the exception of proline), improved yields of the desired transamination product were obtained when using RS under more basic conditions (pH 7.5-8.5 as opposed to pH 6.5). [84]

1.2.3 Genetic incorporation of a formylglycine tag

Formylglycine (fGly) is an electron-rich aldehyde side chain derived from the chemoenzymatic oxidation and subsequent hydrolysis of a cysteine residue that is located within a consensus sequence of CxPxR (where x = any amino acid, Figure 1.14A). Conversion of the thiol to the aldehyde is carried out enzymatically by the Formylglycine Generating Enzyme (FGE, Figure 1.14B). In nature, this transformation is essential for the enzymatic turnover in the active site of sulfatase enzymes; fGly is the active site residue that is responsible for the catalytic hydrolysis of sulfate esters.[85] Use of the fGly residue as a handle for protein bioconjugation was first described in 2007.[86] Short tags containing motifs that are targeted by sulfatases and subjected to fGly conversion in nature were installed at the *N* or *C*-terminus of human growth hormone, mycobacterial sulfotransferase, and maltose-binding protein. Co-expression of the recombinant target proteins in *E. coli* alongside *M. tuberculosis* FGE resulted in >85% conversion of the CxPxR-containing proteins to the desired fGly-tagged proteins (confirmed by MS and subsequent tryptic digest/modification experiments). The fGly tag has since been utilised to install monosaccharides[87] (which can be used as glycosyl acceptors to build *N*-glycans via endoglycosynthase enzymatic derivatisation),[16] biotinylated probes,[88] and therapeutically relevant small molecules.[88-91] The fGly handle has also been used in the specific labelling of tagged proteins in cell extracts.[92] In addition to prokaryotic cell lines, this methodology has also been utilised successfully in mammalian expression systems.[93] Investigation into preactivation of FGE to improve turnover rates using copper has also been reported.[94] For a more detailed discussion on FGEs and the fGly, readers are directed to a recent review in the field.[95]
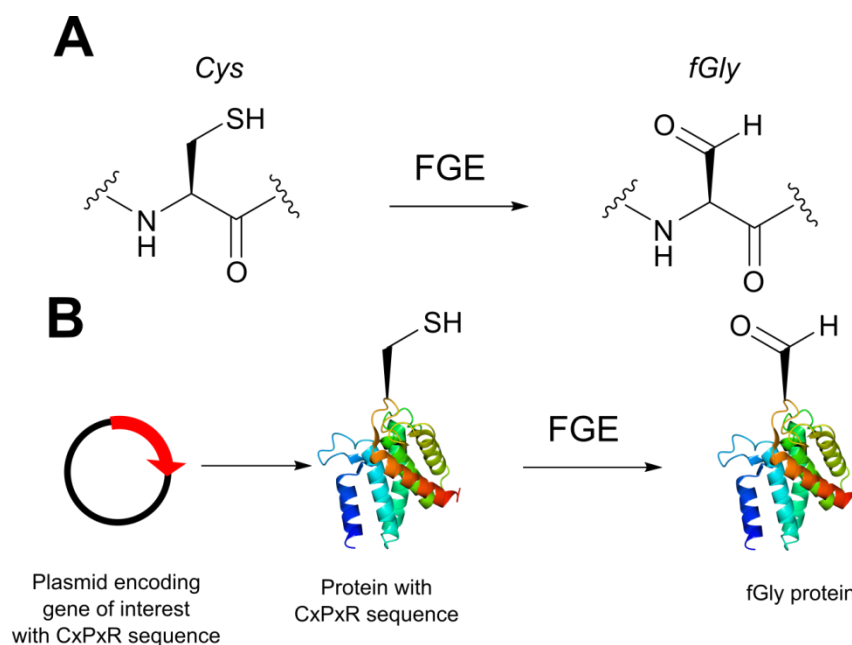
Figure 1.14 (A) Outline of generating the formylglycine (fGly) residue on peptide/proteins bearing a cysteine within a CxPxR sequence. (B) Expression of a plasmid encoding for the desired protein containing the CxPxR sequence, followed by treatment with the formylglycine generating enzyme (FGE) leads to fGly-proteins.

1.2.4 Enzymatic incorporation of farnesyl pyrophosphate analogues

Protein farnesyl-transferase (PFTase) is an enzyme that catalyses the farnesylation of thiol side chains of cysteine residues in proteins. PFTase utilises farnesyl pyrophosphate (FPP) for this transformation and is specific for cysteine residues within a tetrapeptide domain at the C-terminus of a peptide/protein. This tetrapeptide domain is commonly referred to as a CAAX-box, where A = aliphatic amino acid, and X = alanine, serine, methionine, or valine (usually the CAAX-box comprises of the sequence cysteine-valine-isoleucine-alanine).[96,97] Along with its natural substrate FPP, PFTase can also incorporate FPP analogues such as farnesyl aldehyde pyrophosphate (FAPP), allowing for installation of an α,β-unsaturated aldehyde tag into proteins bearing a CAAX-box motif (Figure 1.15A) [98] FAPP installation has been demonstrated on *N*-dansyl peptides and CAAX-box tagged GFP. Following aldehyde installation, tagged proteins can then be specifically isolated *via* modification of the installed aldehyde.[99] PFTase strategies for aldehyde tag incorporation have since been extended to incorporation of an aryl aldehyde analogue (Figure 1.15B),[99] and modification of other proteins such as ciliary neurotrophic factor protein (CNTF).[100] Other aryl aldehyde handles that bear alkyne functionality can also be incorporated using PFTase, allowing for selective, dual

functionalisation of GFP.[100] Mutant variants of PFTase have also been explored in improving the catalytic efficiency of FPP aldehyde analogues; a 43 fold increase, for example, is observed for the incorporation of aryl aldehyde FPP into CAAX-box-bearing CNTF when using a W102A β subunit PFTase variant (as opposed to using wild type PFTase).[101]
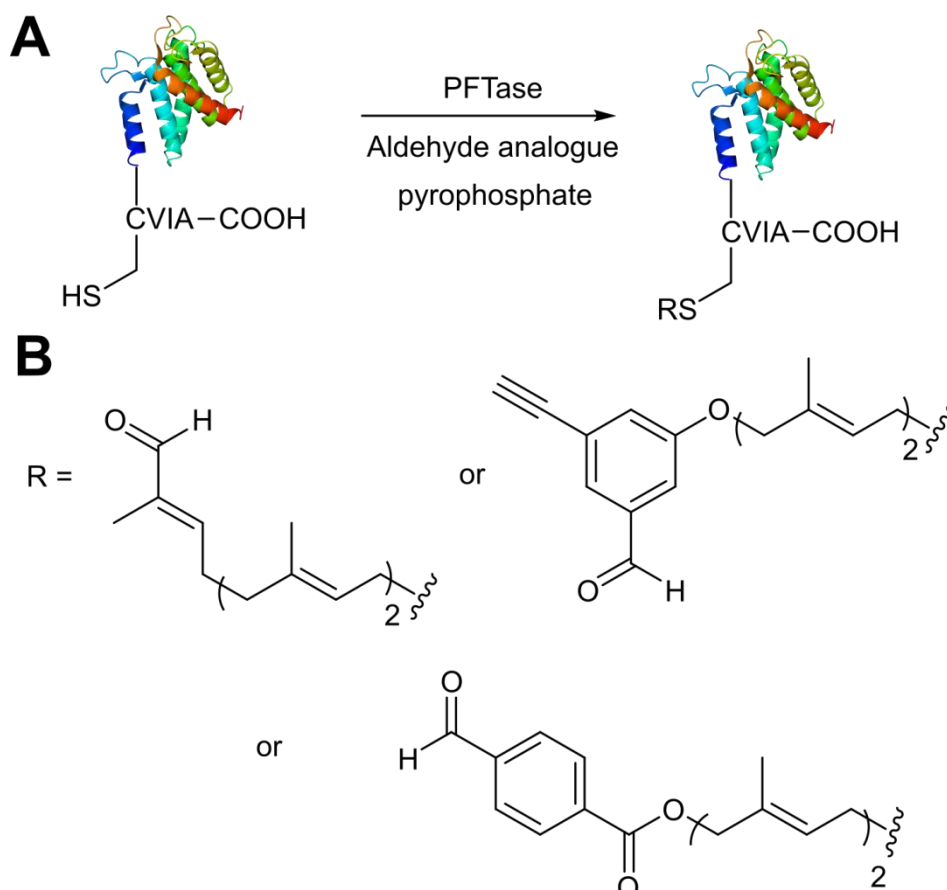
Figure 1.15 (A) Protein farnesyl-transferase (PFTase) mediated installation of farnesyl-derived aldehydes into a protein bearing a C-terminal CAAX box. (B) Aldehyde analogues that can be incorporated.

1.2.5 Enzymatic incorporation using lipoic acid ligase

In *Escherichia coli*, lipoic acid ligase (LplA) is responsible for linking lipoic acid to various proteins involved in oxidative metabolism.[102] LplA specifically targets lysine side chains located within a GFEIDKVWYDLDA sequence known as the LplA acceptor peptide (LAP). The incorporation of aldehyde tags utilising this enzymatic strategy has recently been described (see Figure 1.16A-C). W37I mutants of LplA could be used to incorporate aryl aldehydes into a LAP peptide in a 500:150:1 ratio of lipoic acid-analogue:LAP-peptide:LplA, with a maximum of 62% conversion to

the aldehyde-tagged LAP peptide observed after 5 minutes of incubation. Additionally, proteins bearing LAP tags at the *N*-terminus, including LAP-tagged streptavidin and LAP-tagged alkaline phosphatase, were also successfully modified.[103] More recently, this methodology has been extended to the labelling of LAP-tagged nanobodies. [104]
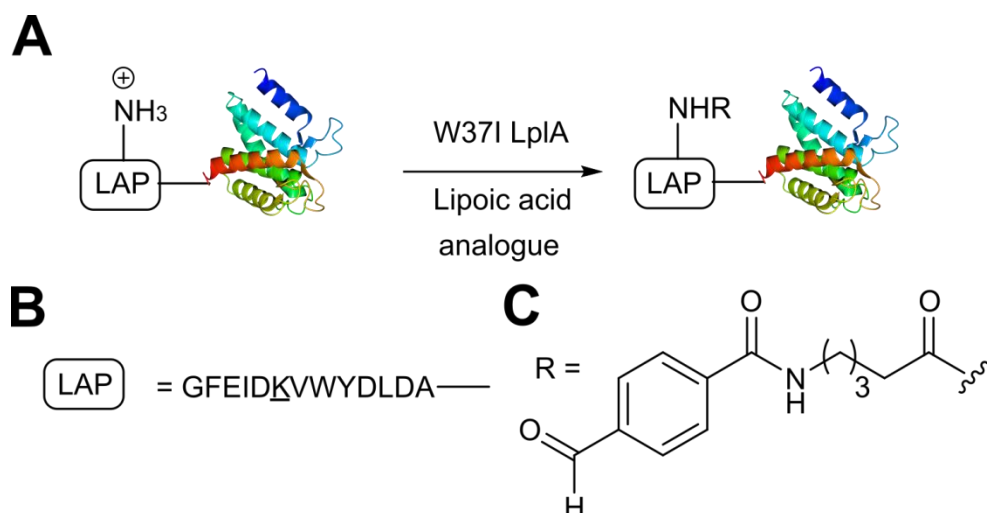
Figure 1.16(A) Outline of using Lipoic acid ligase (LplA) mutants for installing aldehydes into LplA accepter peptide (LAP) bearing proteins. (B) Amino acid sequence of the LAP. (C) Structure of the incorporated aldehyde.

1.2.6 Enzymatic incorporation of 3-formyl-L-tyrosine

A recently described chemoenzymatic method for installation of aldehyde tags into proteins is through enzymatic transformation using tubulin tyrosine ligase (TTL, see Figure 1.17A-C).[105] Although the natural function of TTL is to post-translationally modify the C-terminus of α-tubulin with tyrosine, TTL can also accept tyrosine analogues, such as 3-formyl-L-tyrosine (denoted as '3ForTyr'), in ligation of C-terminal peptides/proteins bearing a VDSVEGEGEEEGEE tag (known as a Tub tag). Mixing of Tub-tagged peptides with TTL in a 200:1 ratio afforded a 63% incorporation of 3ForTyr after 2 h (monitored by HPLC, with >90% incorporation observed after 6 hours). This methodology has also been utilised in the tagging of Tub-tagged GFP, ubiquitin, and GFP-specific nanobodies. Incubation of TTL:Tub-tagged-protein in an optimal ratio of 1:5 for 3 h at 37 °C gave rise to successful labelling with 3ForTyr (confirmed by tryptic digestion experiments). Conversions of 99% were achieved under these conditions (based on SDS-PAGE analysis and observations of other incorporated tyrosine analogues).[105]
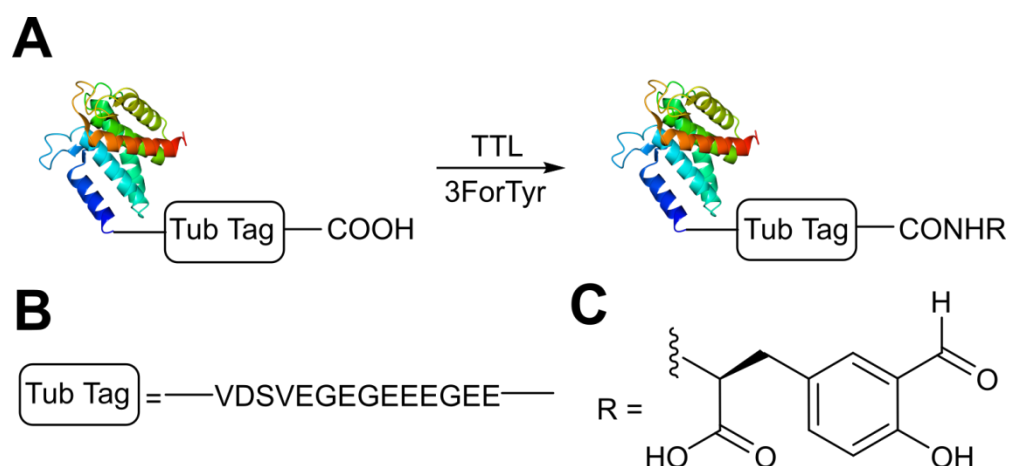
Figure 1. 17 (A) Outline of using Tubulin Tyrosine Ligase (TTL) to for the incorporation of 3-formyl-L-tyrosine (3ForTyr) into proteins bearing a 'Tub Tag'. (B) The amino acid sequence of the Tub Tag. (C) Structure of the incorporated 3ForTyr.

1.2.7 Genetic incorporation of non-native aldehyde containing amino acids

Similarly to installation of other protein handles, another well-publicised strategy[51,52] for installing an aldehyde (or ketone) tag into a protein is to utilise the ever-increasing number of evolved tRNA/tRNA synthetase pairs to perform unnatural amino acid mutagenesis. (Figure 1.18).[106-109] Currently, there a three variants of aldehyde handle that can be incorporated into proteins via unnatural amino acid mutagenesis. The most commonly used unnatural amino acid carbonyl handles are derived from phenylalanine-type side chains, such as phenylalanine derivative **10**. Upon installation, the aldehyde (or ketone) side chain can then be modified with a range of functional handles, including fluorescent probes,[106,109,110] biotinylated probes,[106,107,110] and carbohydrates.[111] ADCs have also recently been developed utilising this methodology.[112,113] More recently described unnatural amino acid – derived aldehyde handles are installed into proteins using precursor unnatural amino acids, followed by deprotection to reveal the desired aldehyde side chain. Installation of azidobenzoxycarbonyl lysine derivative **11** can be deprotected with TCEP to reveal allysine (bearing an alkyl aldehyde),[114] whereas thiazolidine lysine derivative **12** can be deprotected in metal mediated reactions to reveal α-oxo aldehyde side chains.[115,116] Subjecting allysine handles to reductive amination allows access to histone bearing PTM methylated lysine mimics,[114] whereas α-oxo aldehyde side chains can be used for site-selective fluorescent labelling or biotinylation.[116]
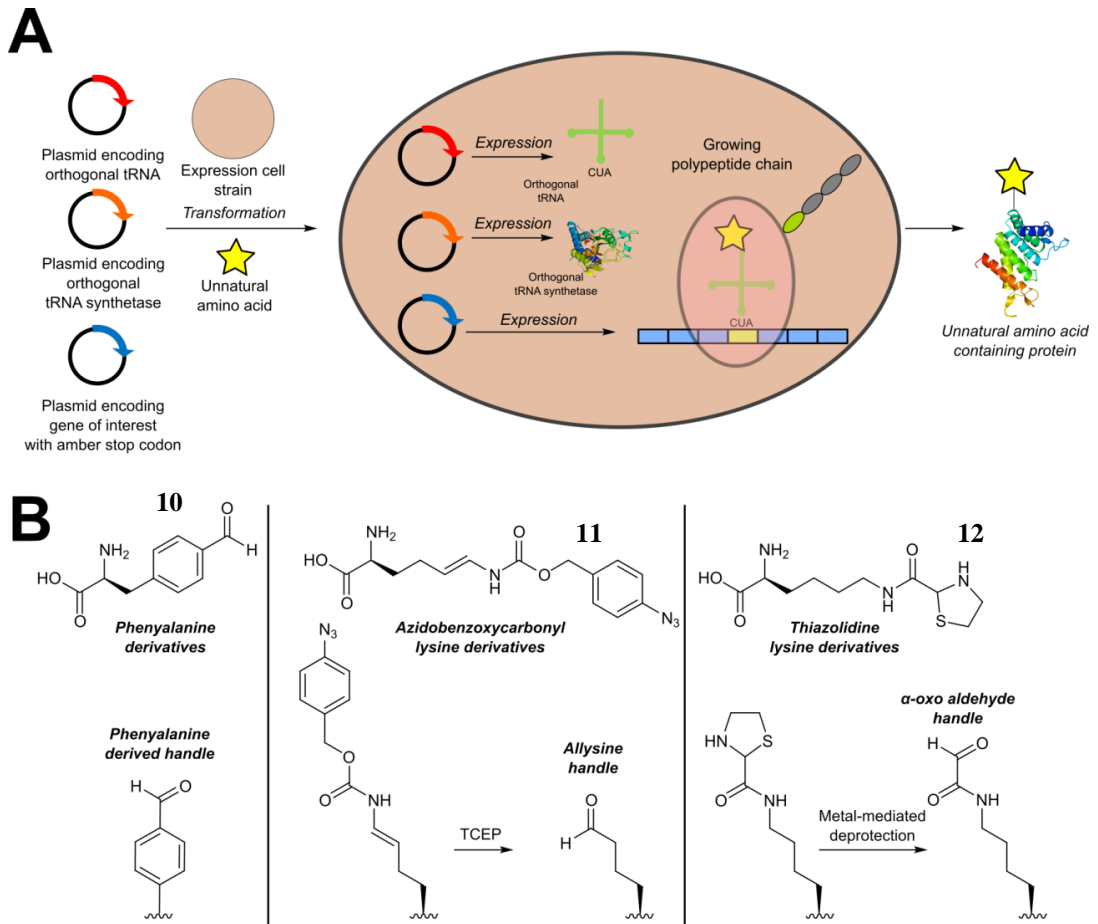
Figure 1.18 (A) Installation of aldehyde into proteins via unnatural amino acid mutagenesis. Expression of a plasmid encoding for the desired protein containing an amber STOP codon in conjunction with an orthogonal tRNA synthetase/tRNA pair and unnatural amino acid allows for incorporation of an aldehyde-containing amino acid at the aforementioned STOP position. (B) Unnatural amino acids and subsequent reactions used to incorporate aldehyde handles into proteins.

1.3 Chemical strategies for the modification of proteins bearing aldehyde handles
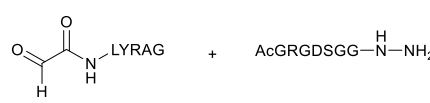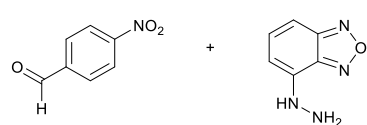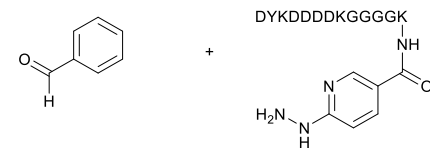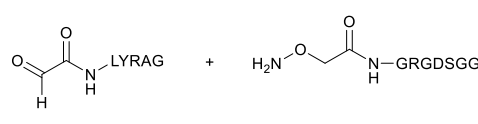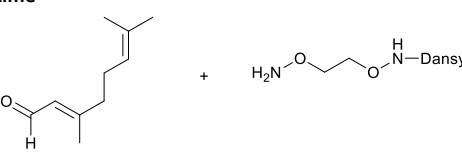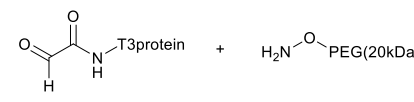
There is now a plethora of strategies for the site-selective modification of proteins following installation of one of the aforementioned aldehyde tags. These range from C=N bond formation to C-C bond formation, and from metal mediated ligations to intramolecular cyclisation ligations. The choice of aldehyde handle, along with the conditions for aldehyde modification (such as pH or use of additional co-solvent) ranges from strategy to strategy. The rate of ligation has also been profiled for the majority of strategies described; often using small molecule aldehydes mimics and, on some occasions, proteins bearing an aldehyde handle. Selected rate constants (and the conditions used) for each ligation strategy discussed herein can be found in Table 1.1.
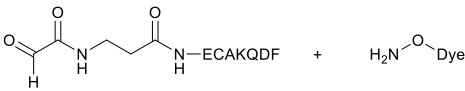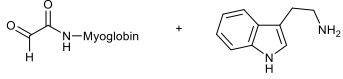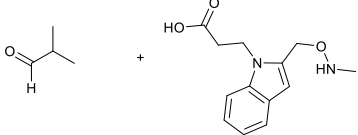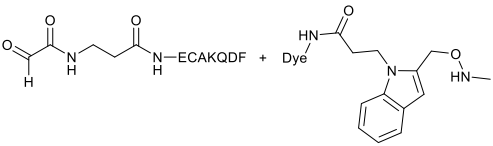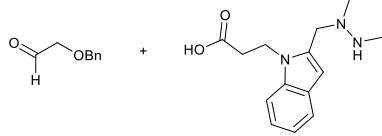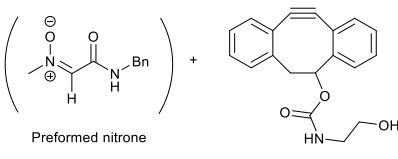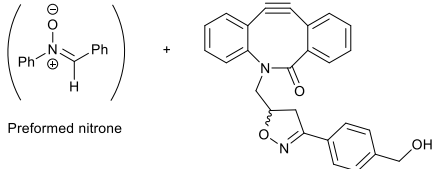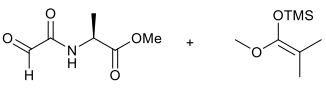
1.3.1 Hydrazone/Oxime ligations
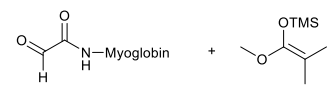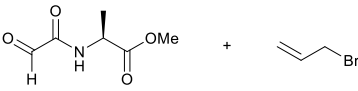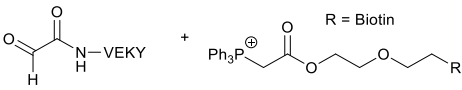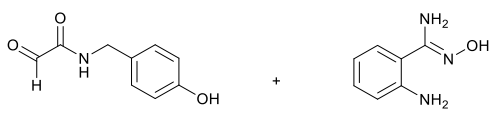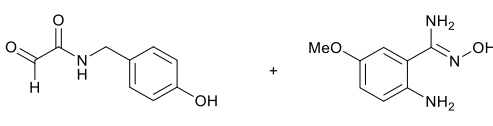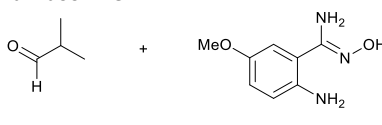
The reaction of hydrazine and hydroxylamine/aminooxy reagents with proteins bearing aldehydes (or ketones) to form hydrazone and oxime bonds, respectively, are among the best established methods for generating homogenous ligation products (Figure 1.19A-B). These ligation strategies proceed via an iminium-type intermediate, and are by far the most popular method of modifying aldehydes installed into proteins. Indeed, a standard method for validating installation of an aldehyde handle into a protein is to perform hydrazone or oxime ligation with small molecule substrates, and then analyse the anticipated ligation product by protein mass spectrometry (protein MS).[71,79,83,86,99,106] Uniquely, to date, they are the only ligation strategies reported in literature that have been demonstrated on all aldehyde handles that can be incorporated into proteins, with the exception of allysine (Figure 1.19C, Table 1.1 entries 1-7). More complex, functionalised hydrazine/hydroxylamine reagents are frequently employed for the purpose of protein bioconjugations. For example, labelling with hydrazine/hydroxylamine functionalised biotin facilitates isolation of the modified protein via binding with streptavidin.[86,110,117] Use of aminooxy functionalised resins can also be applied to isolate aldehyde-tagged proteins via oxime ligation from crude cell lysates.[98,99] Successful incorporation of an aldehyde handle into a protein is also frequently showcased through the reaction with hydrazine or hydroxylamine functionalised
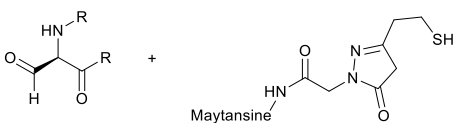
fluorescent probes.[79,83,103,105,109] Analysis of these ligation products via SDS PAGE analysis with fluorescence emission scanning can be used to determine the level of success of the initial aldehyde incorporation method.[16] In a similar manner, hydroxylamine functionalised PEG reagents can also be used to label incorporated

Table 1.1 Comparison of selected rate constants for strategies used in the modification of aldehydes under various conditions. * No rate/yield/conversion given. ** No rate studies performed, reaction time and obtained yield quoted as an alternative. *** No rate studies performed, reaction time and conversion as judged by Liquid Chromatography-Mass Spectrometry (LC-MS) quoted as an alternative.

Entry	Reaction	Conditions	Rate (M⁻¹s⁻¹)	Ref
1	**Hydrazone** (O=CH-C(O)-NH-LYRAG + AcGRGDSGG—N(H)—NH₂)	1 mM aldehyde 1 mM hydrazine pH 5.7, NH₄OAc	0.0031 *(uncatalysed)* 0.21 *(10 mM aniline)*	118
2	**Hydrazone** (4-NO₂-benzaldehyde + benzofurazan-hydrazine)	1 mM aldehyde 18 μM hydrazine pH 7.4, PBS (10% DMF)	0.0013 *(uncatalysed)* 0.11 *(1 mM 5-methoxyanthralic acid)*	119
3	**Hydrazone** (benzaldehyde + DYKDDDDKGGGGK pyridinyl hydrazine)	10 μM aldehyde 10 μM hydrazine pH 7.0, PB	170 *(100 mM aniline)*	120
4	**Oxime** (O=CH-C(O)-NH-LYRAG + H₂N-O-CH₂-C(O)-NH-GRGDSGG)	1 mM aldehyde 1 mM aminooxy pH 4.5, NH₄OAc or pH 7.0, PB	0.057 *(pH 4.5, uncatalysed)* 0.68 *(pH 4.5, 10 mM aniline)* 0.015 *(pH 7.0, uncatalysed)* 0.061 *(pH 7.0, 100 mM p-methoxyaniline)*	121
5	**Oxime** (citronellal-type aldehyde + H₂N-O-CH₂CH₂-O-NH-Dansyl)	50 μM aldehyde 100 μM aminooxy pH 7.0, PB	24 *(25 mM aniline)* 41 *(25 mM m-phenylenediamine)*	122
6	**Oxime** (O=CH-C(O)-NH-T3protein + H₂N-O-PEG(20kDa))	91 μM aldehyde 455 μM aminooxy pH 4.0, Na citrate or pH 7.0, PB	0.22 *(pH 4.0, uncatalysed)* 4.3 *(pH 4.0, 10 mM p-phenylenediamine)* 0.006 *(pH 7.0, uncatalysed)* 0.29 *(pH 7.0, 10 mM p-phenylenediamine)*	123

	Reaction	Conditions	Rate constant	Ref
7	**Oxime**	10 mM aldehyde 10 mM aminooxy pH 7.0, PB:MeOH (1:1)	0.0013	124
8	**Pictet Spengler**	60 µM aldehyde 50 mM indole pH 6.5, PB	N/A*	125
9	*iso*-**Pictet Spengler**	500 µM aldehyde 500 µM indole pH 4.5, NaOAc or pH 7.0, PB	10.6 *(pH 4.5)* 0.26 *(pH 7.0)*	88
10	*iso*-**Pictet Spengler**	2.5 mM aldehyde 2.5 mM aldehyde pH 4, NaOAc:MeOH (1:1)	0.015	124
11	**Hydrazino-*iso*-Pictet Spengler**	50 µM aldehyde 50 µM indole pH 6, PB	4.2	89
12	**Strain-promoted alkyne-nitrone cycloaddition** Preformed nitrone	300 µM nitrone 330 µM alkyne H$_2$O:MeCN (9:1)	12.8	126
13	**Strain-promoted alkyne-nitrone cycloaddition** Preformed nitrone	1:100 molar ratios (nitrone:alkyne) MeCN	58.8	127
14	**Mukaiyama aldol**	200 µM aldehyde 400 µM silyl ketene acetal pH 7.0, PB	24 hours, 55% yield (obtained)**	128

#	Reaction	Conditions	Result	Ref
15	**Mukaiyama aldol**	37.5 µM 7.13 mM silyl ketene acetal H_2O:*t*-BuOH (5:1)	5 hours, 90% (conversion)***	128
16	**Indium mediated allylation**	21 mM aldehyde 150 mg indium 63 mM allyl bromide PB pH 7.0	16 hours, 56% yield (obtained)**	129
17	**Indium mediated allylation**	37.5 µM aldehyde 1 mg indium 836 µM allyl bromide H_2O:*t*-BuOH (6:4)	20 hours, 54% (conversion)***	129
18	**Wittig**	10 mM aldehyde 20 mM ylide H_2O:*t*-BuOH (1:1)	90 min, 67% yield (obtained)**	130
19	**Wittig** R = Biotin	4 mM aldehyde 800 µM ylide 100 mM PBS	0.11 (*pH 6.75*) 0.90 (*pH 7.70*) 5.9 (*pH 8.4*)0	131
20	**2-Amino benzamidooxime**	15 mM aldehyde 3 mM 2-aminobenzamidooxime pH 4.5, NaOAc	0.0086	132
21	**2-Amino benzamidooxime**	2 mM aldehyde 1 mM 2-amino benzamidooxime pH 4.5, NaOAC	1	132
22	**2-Amino benzamidooxime**	2 mM aldehyde 1 mM 2-amino benzamidooxime pH 4.5, NaOAC	40	132
23	**Thiazolidinedione aldol**	10 mM aldehyde 10 mM thiazolidinedione pH 7.0, PB (25% DMSO)	0.0078	133
24	**Thiazolidinedione aldol**	37.5 µM aldehyde 37.5 mM thiazolidinedione pH 7.0, PB	3 hours, 83% (conversion)***	133

25	Trapped Knoevenagel R = Herception	25 µM aldehyde 1 mM thiopyrazolone pH 7.0, Na citrate	0.387	91

aldehyde handles; these ligation products can then be visualised *via* SDS PAGE analysis.[80,84,100,134] As discussed earlier, carbohydrates have also been installed into proteins *via* oxime ligation.[16,87,106] In an alternate approach, glycosylated proteins have also been modified by initial oxidative cleavage of carbohydrate diols with NaIO$_4$, followed by hydrazone ligation of the resultant aldehyde.[135,136] Oxime linked ADCs have also been reported;[104,137] these constructs display both *in vitro*[112] and *in vivo*[134] cytotoxicity. More unconventional uses of oxime ligations strategies involve linking protein species together for the design of new therapeutics.[77,138] A review on oxime chemistry within the field of aldehyde handle labelling has recently been published.[139] Hydrazone and oxime ligations are undoubtedly the most frequently used method for modifying protein aldehyde/ketone handles; indeed, the aforementioned commercially available probes for modifying aldehyde handles are all hydrazine and hydroxylamine-based. However, these strategies are not without their drawbacks. Both ligation strategies result in formation of a C=N bond between protein and substrate, and thus both are susceptible to hydrolysis over time. Of the two, oxime bonds display greater hydrolytic stability.[140] Reductive amination of hydrazone bonds using sodium cyanoborohydride to form a more stable amine bond can be employed; however, this strategy is not commonly used within the context of modifying protein aldehydes (sodium cyanoborohydride is a water sensitive reagent and its use can lead to undesirable reduction of disulfide bonds[22]). Additionally, the optimal pH for hydrazone and oxime ligations is pH 4.5, with poor reactivity observed at neutral pH, typically displaying rate constants of 10^{-3} M^{-1} s^{-1}.[123,141] To address this, several catalysts for performing hydrazone and oxime ligation at neutral pH have been investigated; aniline in particular has received a great deal of attention in the area of hydrazone/oxime ligation catalysis. For hydrazone ligations of peptide substrates in 10 mM anilinium buffer, a 20 fold increase in ligation rate is observed at pH 4.5, and a 70 fold increase in ligation rate is observed at pH 5.7 (Table 1.1, entry 1). Similarly, for oxime ligations of peptide substrates in 100 mM anilinium buffer solutions, a 400 fold increase in ligation rate is observed at pH 4.5, and a 40

fold increase in ligation rate is observed at pH 7.0.[121] Ligations between aryl aldehydes and 6-hydrazinopyridyl reagents are of particular note, displaying rate constants within the range of 10^2-10^3 M^{-1} s^{-1} (Table 1.1, entry 3).[120] The use of aniline in protein bioconjugation at a neutral pH has subsequently been reported.[77,109] Other aniline-based catalysts have also been reported, with *meta*-substituted anilines (Table 1.1, entry 5)[122] and *para*-substituted anilines (Table 1.1, entry 6). displaying improved kinetics to that of aniline for oxime ligations at pH 7.0. 70 In all cases, the catalysis is hypothesised to proceed *via* a 'transimination' type mechanism.[121] In very particular cases, aniline does not appear to improve conjugation
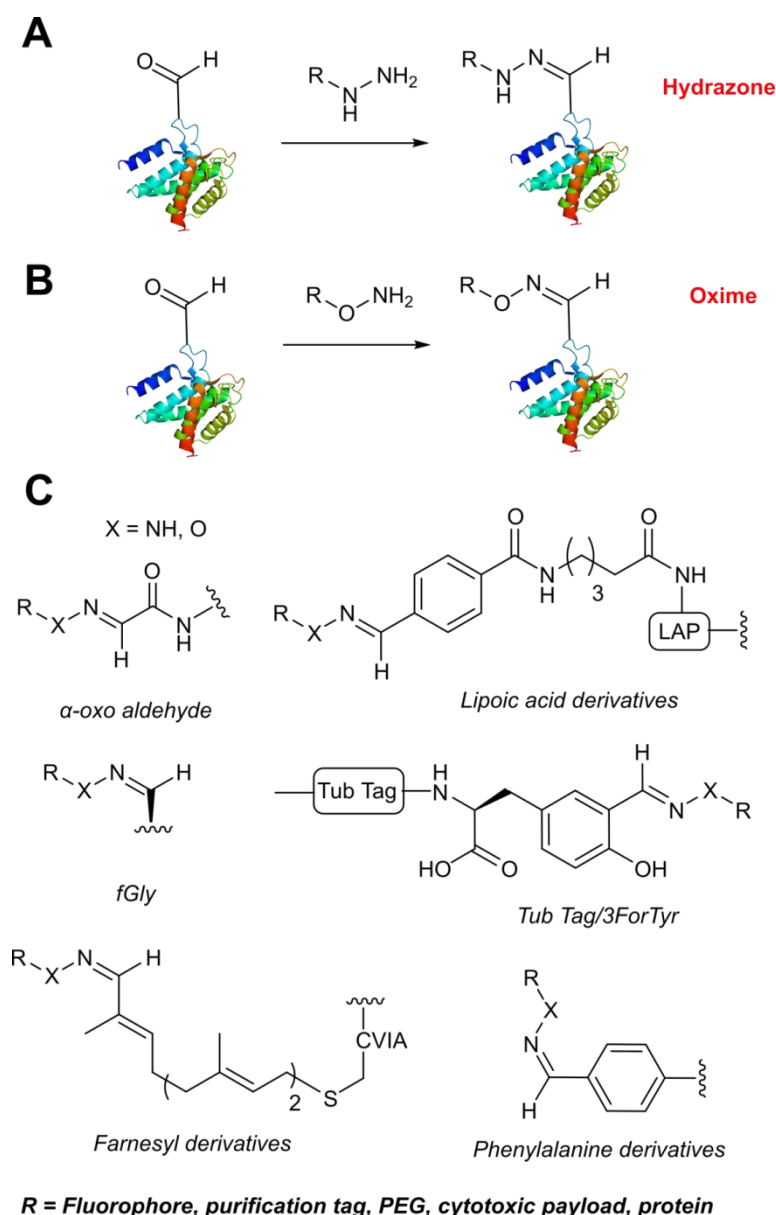
Figure 1.19 (A) Outline of hydrazone ligation on a generic protein bearing an aldehyde handle. (B) Outline of oxime ligation. (C). Structure of hydrazone/oxime bond on protein aldehyde handles.

yields of oxime ligations.[138] Additionally, it has been reported that typical concentrations of aniline used in oxime ligation strategies (i.e. 100 mM aniline)[77,121] exhibit toxicity in live cell labelling experiments.[109] Other catalysts include boronic acids[142] and 2-aminophenols/2-(aminomethyl)benzimidazoles.[143]

1.3.2 Pictet-Spengler, *iso*-Pictet Spengler, and Hydrazino-*iso*-Pictet Spengler

The cyclisation reaction between tryptamines and aldehydes is known as the Pictet-Spengler reaction. Formation of the Pictet-Spengler product occurs through initial iminium formation between the tryptamine amine and aldehyde, followed by intramolecular cyclisation. In the context of protein modification, an *N*-terminal Pictet-Spengler ligation on myoglobin has been reported (Figure 1.20).[125] Following transamination with PLP,[79] the Pictet-Spengler ligation between glyoxylamide-containing myoglobin (glyoxyl-myoglobin) **13**, and tryptamine **14**, tryptophan methyl ester **15**, and tryptamine-functionalised-biotin **16** was successfully carried out at pH 6.5 (confirmed by MS and subsequent proteolytic digest experiments), leading to protein constructs ligated to substrates via a C-C bond.[125]
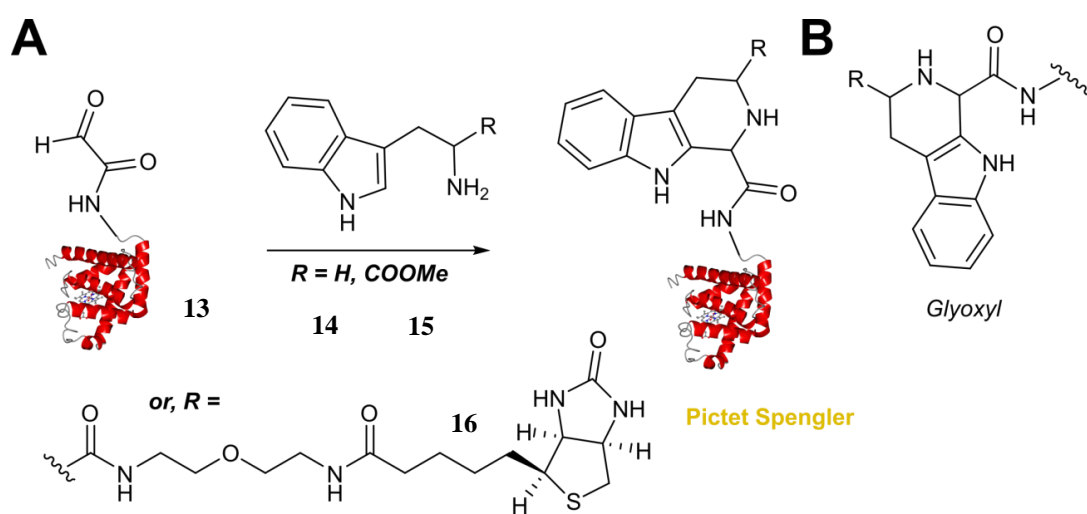
Figure 1.20 (A) The Pictet-Spengler ligation on glyoxyl-myoglobin. (B) The Pictet Spengler ligation at a glyoxyl handle

A variation of the Pictet-Spengler ligation, known as the *iso*-Pictet-Spengler ligation, has recently been reported for chemical modification of aldehyde handles incorporated into proteins (Figure 1.21A).[88] Analysis of the reaction kinetics of the traditional Pictet-Spengler ligation was used as the basis for the design of the *iso*-Pictet-Spengler ligation. Aminooxy-functionalised indoles (as opposed to

tryptamines) were utilised to accelerate the initial iminium formation between the indole and the aldehyde tag. Additionally, the aminooxy side chain was located on the 2-position of the indole ring (as opposed to the 3 position which is the case for tryptamine reagents) to accelerate the final cyclisation step. Based on an isobutyraldehyde system (Table 1.1, entry 9), the optimal pH for the reaction was found to be pH 4.5 (rate constant of 10.6 M^{-1} s^{-1}), although product formation is still observed at pH 6.0-7.0 (observed rate constants of 0.76 M^{-1} s^{-1} at pH
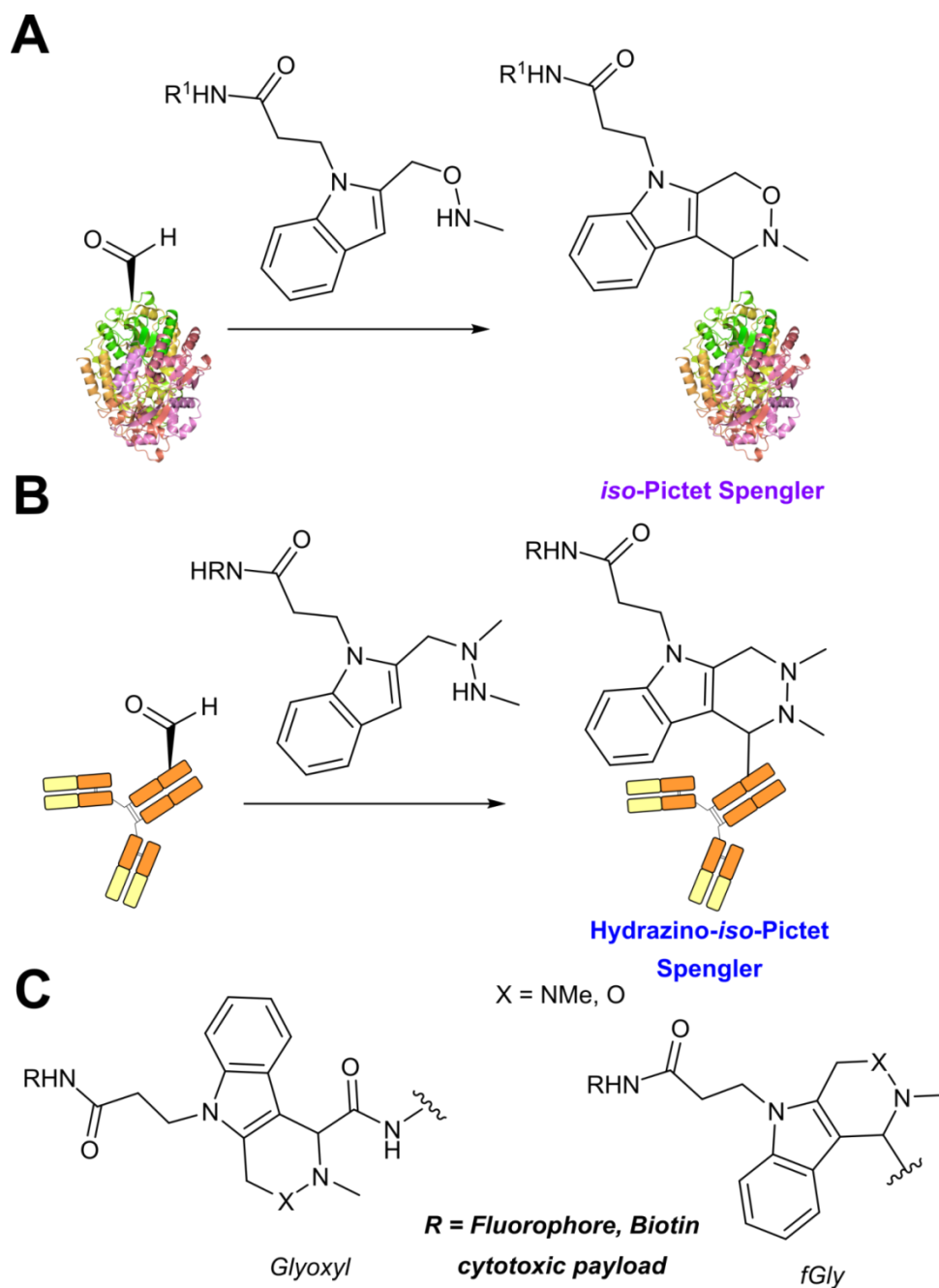
Figure 1.21 (A)The iso-Pictet Spengler ligation on fGly-tagged maltose binding protein. (B). The Hydrainzo-iso-Pictet Spengler ligation on an fGly-tagged antibody. (C). The iso-Pictet/Hydrazino-iso-Pictet Spengler ligation at glyoxyl handle (left) and at an fGly handle (right).

6.0 and 0.26 M^{-1} s^{-1} at pH 7.0). 37 For glyoxyl systems, reduced rate constants for the *iso*-Pictet Spengler ligation have since been reported (Table 1.1, entry 10).[124] The use of an aniline catalyst had no effect on the rate of reaction (and was noted to be inhibitory at pH 5.0-6.5). *Iso*-Pictet-Spengler ligations with PEGylated, biotinylated, and fluorescent probes were successfully demonstrated on glyoxyl-myoglobin, fGly-tagged maltose binding protein (MBP), and fGly-tagged Herceptin antibodies.[88] Compared to the original Pictet-Spengler ligation,[125] both reagent concentrations and reaction time are vastly reduced in the iso-Pictet-Spengler ligation.[88] A minor alteration to the *iso*-Pictet-Spengler ligation, known as the Hydrazino-*iso*-Pictet-Spengler (HIPS) ligation, has also been described in the literature (Figure 1.21B).[89] The concept of the ligation strategy is nearly identical to that of the iso-Pictet-Spengler ligation – except an *N*-methyl hydrazine-functionalised indole is used instead of an aminooxy functionalised indole. This facilitates the ligation at a higher pH of 6.0 (as hydrazines are more reactive than aminooxy compounds at a higher pH).[144] The HIPS ligation shows improved ligation kinetics to that of the iso-Pictet-Spengler ligation at pH 6.0 (observed rate constants of 4.2 M^{-1} s^{-1} using a benzyloxyacetaldehyde system, Table 1.1, entry 11), and has subsequently been utilised in the synthesis of ADCs.[90] The combination of the fGly handle and the Hydrazino-*iso*-Pictet Spengler ligation forms the basis of SMARTag technology by Catalent Pharma Solutions (formerly Redwood Biosciences) for synthesis of next-generation ADCs.[14,90]

1.3.3 Strain-promoted alkyne-nitrone cycloaddition

As discussed previously, alkyne reagents are frequently employed in strategies to selectively modify proteins,[145] and alkyne mediated ligation strategies are commonly utilised in metabolic labelling experiments due to their bioorthogonality.[11] Alkynes can also be used to site-selectively modify glyoxyl-proteins via the Strain-Promoted Alkyne-Nitrone Cycloaddition (SPANC) ligation (Figure 1.22). The SPANC ligation is a one-pot modification strategy that involves formation of a nitrone species between the aldehyde and N-methylhydroxylamine **17**, followed by a cycloaddition reaction between the nitrone and strained cyclooctyne (see Table 1.1, entries 12 and 13 for examples with small molecules).[126] The addition of *p*-anisidine **18** greatly enhances the rate of the SPANC ligation – presumably by a mechanism that is akin to catalysis of hydrazone and oxime ligations. The SPANC ligation is usually

performed at pH 6.8 (NH$_4$OAc buffer) and has been used to site-selectively modify the oxidised chemokine interleukin 8 (glyoxyl-IL-8) with a 2kDa PEG unit **19**. [126] Dual functionalisation of glyoxyl-peptides/proteins incorporating both azide-alkyne click chemistry and the SPANC ligation has also been reported.[146] Additionally, SPANC ligation has been utilised in the design of antibody-nanoparticle constructs to design multivalent nanoparticles (MNPs) using an *N*-terminal serine variant of an anti-HER2 scFv antibody; these MNPs can then be used to selectively, fluorescently label HER2 membrane receptors in breast cancer cells.[8] A variation of the SPANC ligation using preformed nitrones has also been utilised in metabolic labelling of live cells (although these nitrones are susceptible to hydrolysis in aqueous systems).[147] Preformed nitrones have also been the subject of various kinetic studies, identifying biaryl-aza-cyclocyne[126] and acyclic nitrones systems[127] as yielding the highest rate constants for the SPANC (58.8 M^{-1} s^{-1}). A review of the SPANC ligation has recently been published.[148]
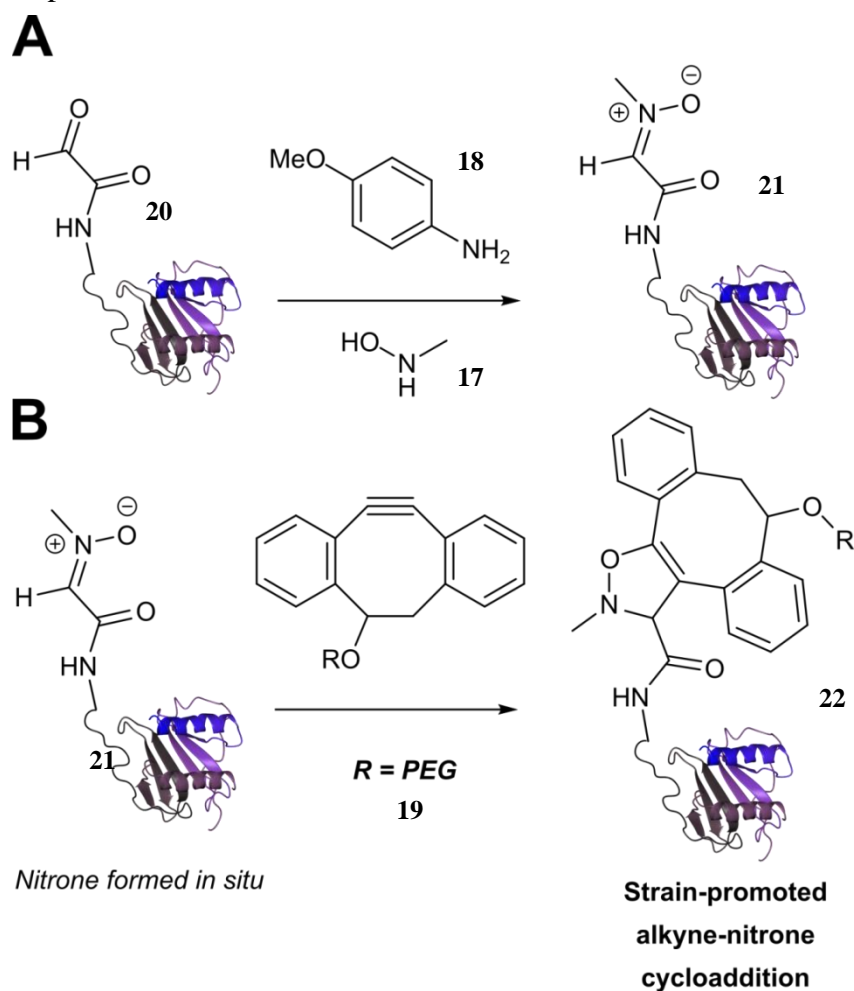
Figure 1.22 (A) Glyoxyl-interleukin 8 **20**, and formation of a nitrone species **21**. (B). Reaction of an alkyne with the pre-formed nitrone via a cycloaddition to generate PEGylated protein **22**.

1.3.4 Mukaiyama aldol

The Mukaiyama aldol reaction involves the reaction of an aldehyde group and silyl enol ether to give β-hydroxy ketone products. The reaction forms a C-C bond between the aldehyde and silyl enol ether, is catalysed by Lewis acids (such as TiCl$_4$) and can be performed in aqueous solvent systems.[149] A Mukaiyama aldol ligation for peptide and protein modification was first reported in 2010 (Figure 1.23).[128] Glyoxyl-peptides were functionalised with various silyl enol ethers in aqueous conditions to afford the anticipated Mukaiyama aldol ligation products after 24 hours (see Table 1.1, entry 14 for an example). Products bearing alkene functionality could then be modified further by cross metathesis using a Hoveyda-Grubbs second generation catalyst. The Mukaiyama aldol ligation has been utilised to modify glyoxyl-myoglobin **13**, with a maximum of 90% conversion observed as (judged by LC-MS) in a H$_2$O:t-BuOH (5:1) solvent system (Table 1.1, entry 15). Choice of solvent system for the Mukaiyama aldol ligation is of particular importance; 1:1 systems such as H$_2$O:t-BuOH, pH 7 PB:EtOH, and H$_2$O:MeCN all lead to homogenous product formation (76-90% conversion as judged by LC-MS), whereas other solvent systems such as H$_2$O:glycerol and H$_2$O:etheylene glycol lead to no observed conversion to the Mukaiyama aldol ligation product, or solvent systems such as pH 7 PB:EtOH (4:1), pH 7 PB:t-BuOH (1:1), and H$_2$O:t-BuOH (6:1) lead to heterogeneous product mixtures. UV-vis analysis and circular dichroism (CD) analysis of unmodified and modified myoglobin samples confirmed that the tertiary structure of the protein was unaffected by the Mukaiyama aldol ligation.[128]
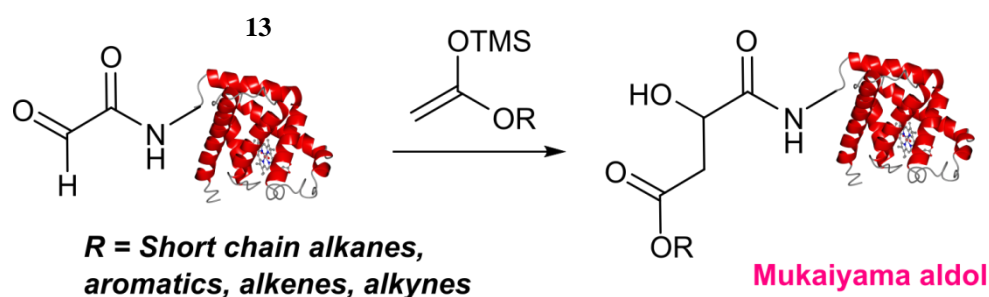
Figure 1.23 Outline of a Mukaiyama ligation on glyoxyl-myoglobin.

1.3.5 Indium mediated allylation

An indium mediated allylation within the context of modifying glyoxyl peptides and proteins with simple allyl bromides has also recently been reported (Figure 1.24).[129] Glyoxyl-peptides were modified by indium mediated allylation with a number of allyl bromides to give racemic mixtures of the desired products after 18 hours (Table 1.1, entry 16). The ligation results in formation of a C-C bond and has also been demonstrated on glyoxyl-myoglobin **13**, affording a maximum of 54% conversion after 20 hours to the desired product (judged by LC-MS analysis, Table 1.1, entry 17). A 6:4 H_2O:t-BuOH solvent system and sonication of the reaction mixture was utilised to achieve conversion to the desired product (these conditions cause partial dissociation of the haem group of myoglobin, although it is possible to reconstitute the protein following modification).[129]
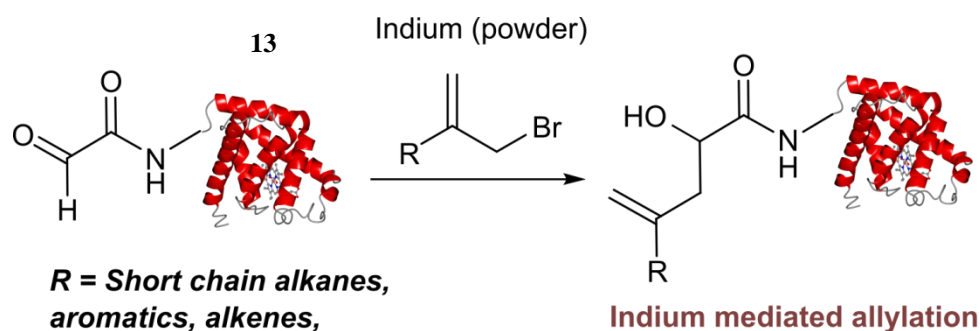
Figure 1.24 Outline an indium mediated allylation of glyoxyl-myoglobin.

1.3.6 Wittig and tandem Wittig/Diel Alder

The Wittig reaction, first reported in 1954, is the reaction between aldehydes and ylide-type reagents to give alkene products.[150] Modification of glyoxyl peptides and proteins via a Wittig ligation have been recently demonstrated (Figure 1.25).[130] Glyoxyl-peptides were successfully modified with a number of small molecule ylides in a 1:1 H_2O:t-BuOH solvent system, generating geometric isomers of the desired Wittig products (see Table 1.1, entry 18 for an example). Glyoxyl-myoglobin **13** and glyoxyl-IL-8 **20** were also modified using this chemistry in a 3:1 H_2O:t-BuOH solvent system. The ligation can also be performed in pH 7-10 buffer:t-BuOH systems, but does not occur in buffer systems below pH 6. The Wittig ligation results in formation of a C-C bond, and gives a 65% conversion complete within 30 minutes

on glyoxyl-myoglobin (judged by LC-MS analysis). UV-vis analysis, CD analysis, and analysis of the proteins ability to store and release oxygen were performed on unmodified and modified myoglobin samples,
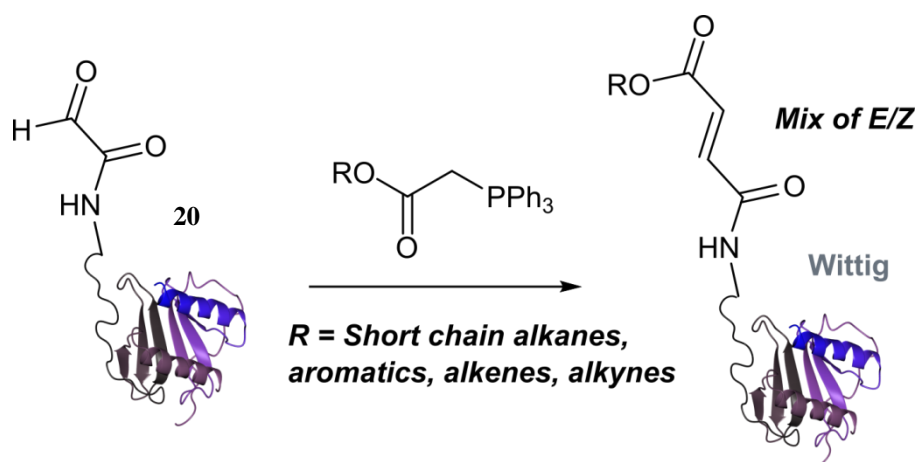
Figure 1.25 Outline of Wittig ligation of glyoxyl-interleukin 8.

confirming that the tertiary structure of the protein was unaffected.[130] More recently a thorough investigation into the kinetics of Wittig olefination within the context of bioconjugation, along with modification of phage libraries has been carried out (Figure 1.26A).[131] In this study, the effects of aldehyde handle choice on both the *E/Z* product ratio (se Figure 1.26B for structures of both isomers) and the rate of Wittig olefination were examined using benzaldehyde, 2-nitrobenzaldehyde, and glyoxyl-VEKY peptide. At neutral pH, lower rate constants and higher selectively for the *E* isomer were observed when using benzaldehyde, whereas higher rate constants and lower selectively for the *E* isomer were observed when the aldehyde handle of choice was either 2-nitrobenzaldehyde or glyoxyl-VEKY peptide These rate constants could be increased further at higher pH values (see Table 1.1, entry 19); however, hydrolysis of the ester backbone of the Wittig products was found to be also accelerated. Total hydrolysis of the *Z*-isomer was observed after 7 h at pH 8, and a half-life of $t_{1/2} = 18$ h was observed for the *E*-isomer, whereas at pH 5, hydrolysis half-life was observed to be $t_{1/2} > 75$ h. Additionally, choice and concentration of buffer salts were also established to affect the rate of Wittig olefination. Subsequently Wittig olefination was then applied towards selective modification of glyoxyl-bearing phage libraries using a ylide ester biotin phosphonium salt precursor (YEB) **23** in MOPS buffer at pH 6.5. These Wittig products could then be site-selectively modified further, either via Michael addition

using functionalised thiols (for example, using a Cys containing FLAG tag), or via a Diels-Alder reaction with cyclopentadiene **24**. The Diels-Alder reaction is selective for the *E*-isomer of the Wittig product (Figure 1.26C), whereas the *Z*-isomer of the Wittig product is unreactive towards Diels-Alder modification.[131]
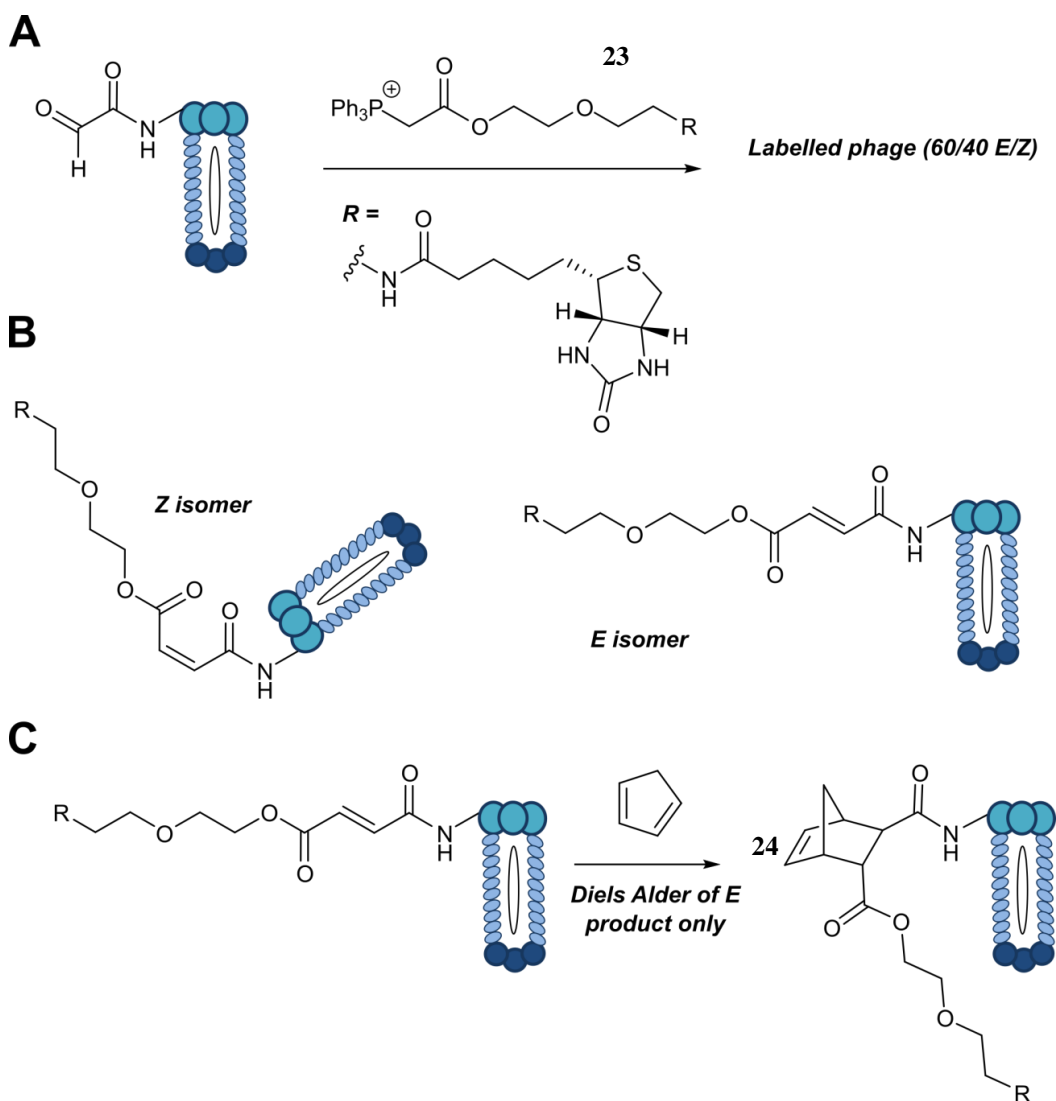
Figure 1.26 (A) Labelling of glyoxyl-phage via a Wittig ligation with ylide ester biotin phosphonium salt precursor. (B) Structure of the Z and E isomers formed during the Wittig ligation. (C) Diels Alder reaction of the Z isomer of the Wittig bioconjugate.

1.3.7 2-Amino benzamidoxime (ABAO)

The use of 2-amino benzamidoxime (ABAO) compounds in the modification of peptides and proteins bearing aldehyde handles has been very recently described (Figure 1.27).[132] Initial screening of a variety of substituted aniline compounds against model glyoxyl substrates identified ABAO compounds as suitable reagents for site-selective modification, giving rise to single, hydrolytically stable dihydroquinazoline products which exhibit fluorescence at 490 nm. The ligation is performed under acidic conditions (pH 4.5), and is thought to proceed through via imine formation between the aldehyde substrate and aniline component.
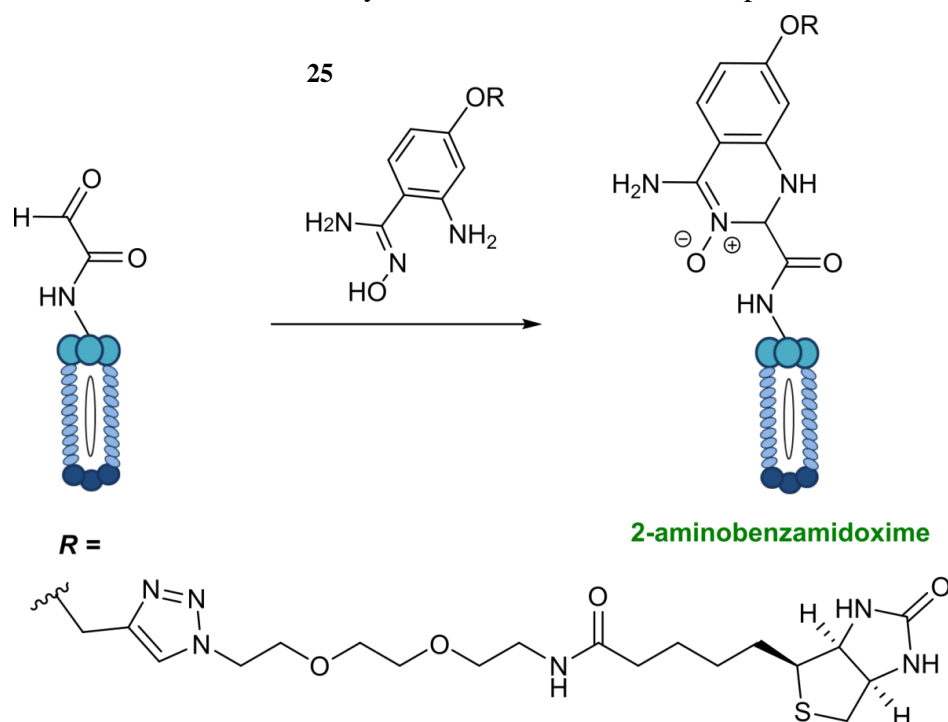
Figure 1.27 Outline of 2-amino benzamidoxime ligation on M13 phage bearing glyoxyl-peptides.

This is followed by intramolecular cyclisation *via* attack of the ABAO ortho amidoxime group at the newly formed imine, completing the ligation. The rate of ABAO ligation can vary depending on the both the choice of aldehyde and ABAO derivative; in particular, higher rate constants are observed in isobutyraldehyde systems compared to glyoxyl systems (40 M-1 s-1 compared to 1 M-1 s-1, Table 1.1, entries 21 and 22), and also in *para*-methoxy ABAO systems (Table 1.1, entry 20). Acetate (from the NaOAc buffer) can also act as a weak catalyst in the ABAO ligation (as observed from altering buffer concentrations). Notably, ABAO ligation is slow when the aldehyde substrate in question is glucose (in a ring opened form),

suggesting that ABAO ligation could be expanded to other selective labelling applications in more complex biological mixtures. The ABAO ligation has been used in the biotinylation of horseradish peroxidase (HRP) bearing aldehydes derived from oxidised glycans, and additionally in the biotinylation of glyoxyl-peptide bearing M13 phage libraries with ABAO-derived biotin **25**.[132]

1.3.8 Thiazolidinedione aldol

Thiazolidinedione compounds have been described as acting as donor compounds in uncatalysed aldol reactions.[151] This methodology has very recently been extended to peptide and protein modification (Figure 1.28).[133] The ligation was demonstrated on C-protected glyoxyl-dipeptides with a number of simple thiazolidinediones bearing various functionality such as alkyne, azide and short chain PEG moieties. Additionally, modification of glyoxyl-myoglobin **13** can be achieved in a 5:1 pH 6.5 PB:t-BuOH solvent system with conversions of 66-83% observed within 3 hours at 37 °C (judged by LC-MS analysis, Table 1.1, entry 24).

R = Small molecules, alkynes, azides

Thiazolidinedione aldol
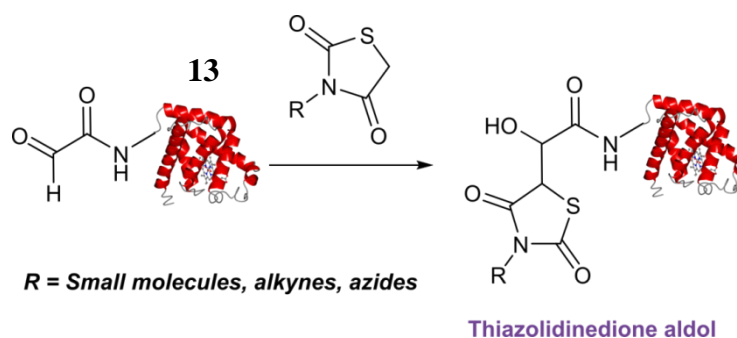
Figure 1.28 Outline of modification of glyoxyl-myoglobin with thiazolidinediones.

In all cases, high concentrations of thiazolidinedione reagent were utilised (1000 molar equivalents with respect to glyoxyl-myoglobin concentration). UV-vis analysis, CD analysis, and analysis of the proteins ability to store and release oxygen were performed on unmodified and modified myoglobin samples, confirming that the tertiary structure of the protein was unaffected.[133] Observed reaction kinetics for the Thiazolidinedione aldol ligation were 0.0078 M-1 s-1 (Table 1.1, entry 23),[133] which are comparable to that of uncatalysed hydrazone and oxime ligations at pH 7.0-7.4.[123,152]

1.3.9 Knoevenagel condensations

A Knoevenagel condensation reaction, dubbed the trapped-Knoevenagel ligation (tK), between fGly peptides/antibodies and compounds containing enolisable protons has recently been described (Figure 1.29). In this ligation, thiopyrazolone reagents are subjected to the fGly species to afford C-C linked tetrahydrothiopyranopyrazolone products.[91] The reaction occurs *via* initial enone formation between the aldehyde and thiopyrazolone, followed by intramolecular cyclisation through Michael-type addition of the thiol side chain. This ligation strategy proceeds at pH 7.0-7.2 and has been validated on benzyloxyacetaldehyde, benzaldehyde and fGly functionalised peptides. The tK ligation has also been explored in the synthesis of ADCs through the tagging of fGly functionalised Herceptin with thiopyrazole-functionalised maytansine; ligation of these species exhibits a rate constant of 0.387 M-1 s-1 (Table 1.1, entry 25), and the resulting constructs display both *in vitro* and *in vivo* cytotoxicity.[91]
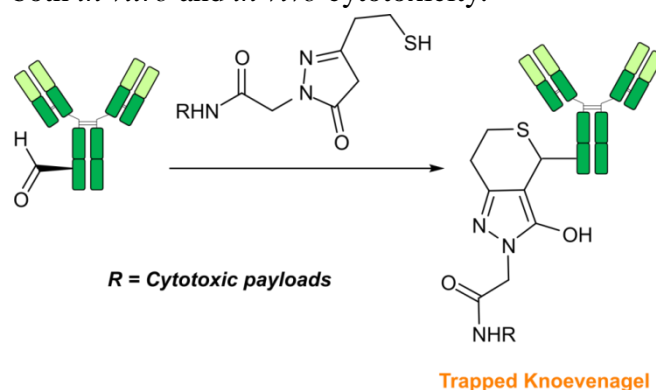
Figure 1.29 Outline of the Trapped-Knoevenagel ligation on fGly-tagged Herceptin.

In addition to the tK ligation, an alternate Knoevenagel condensation, known as the Tandem Knoevenagel Condensation-Michael Addition (TKM) has very recently been reported in the literature (Figure 1.30).[153] The TKM strategy is almost identical to that of the tK, with the key difference lying in the identity of the pyrazolone partner. In constrast to the tK, which uses thiopyrazolone reagents to facilitate both initial Knoevenagel condensation and subsequent Michael-type intramolecular cyclisation, the TKM uses pyrazolone partners that do not contain a thiol side chain, preventing intramolecular cyclisation as seen in the tK. Instead, following initial enone formation as seen in the tK ligation, intermolecular bioconjugation of a second pyrazolone molecule occurs *via* a Michael-type addition, resulting in a second bioconjugation in a one pot process (both steps at pH 7.2). The TKM has been

applied in the synthesis of ADCs bearing drug to antibody ratios (DARs) of four; these constructs have subsequently been used to investigate how the location of conjugated payloads of ADCs can affect ADC half-life, and *in vivo* efficacy.[153]
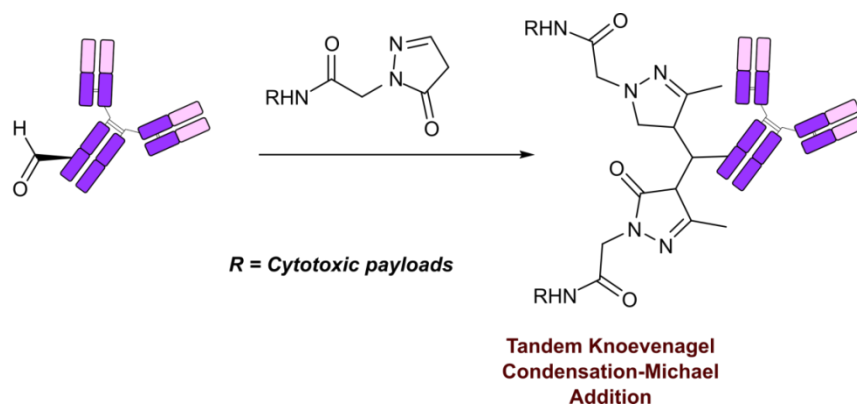
Figure 1.30 Outline of the Tandem Knoevenagel Condensation-Michael addition on fGly-tagged antibodies.

1.4 Developing new methodology for protein bioconjugation

As evident from Chapter 1, there are now numerous elegant methods available for generating protein bioconjugates, revolutionising the way chemists and biologists approach the fields of medicine, chemical biology, and cell biology. However, many of the methods for protein bioconjugation described in the literature suffer from various drawbacks. The majority of conjugation methods using naturally occurring amino acids are non-selective, and lead to a heterogeneous mixture of bioconjugates; this issue can however be avoided by using site-selective strategies. These site-selective strategies must then operate in aqueous conditions, at neutral pH, at mild temperatures, and yield hydrolytically stable bonds, all whilst offering selective modification at a particular site on the protein, and maintaining any integrity and function the unmodified protein may have. [15,19] In particular, strategies that synthesise carbon-carbon (C-C) bonds are highly prized in protein bioconjugation, due to the C-C bond being stable across a range of conditions. Although current methods to construct C-C linked protein bioconjugates have greatly furthered the field of chemical biology, such as in designing new therapeutics or synthesising PTM mimics, [17,18,90,91,153] the majority of C-C bond forming bioconjugation strategies utilise conditions that fall outside of the requirements of a protein bioconjugation strategy, such as conducting ligations outside of neutral pH (usually at acidic pH), displaying poor kinetics of ligation, or using reagents that potentially can denature the target protein. C-C bond forming strategies are also sometimes hampered by the requirement of expensive chemical probes, or by the reduced reactivity displayed in the presence of oxygen. There is therefore a need to develop C-C bioconjugations strategies that are efficient and, most importantly, operate at neutral pH whilst maintain protein integrity. As a bonus, such strategies would also utilise simple chemical probes in small ratios, and be oxygen tolerant.

Site-selective modification of aldehyde handles incorporated into proteins has been at the forefront of the field of chemical biology, with protein aldehyde bioconjugation proving to be a highly versatile method of generating homogenous protein conjugates with wide ranging applications. Particularly notable is that, with the exception of hydrazone/oxime strategies which form hydrolytically unstable C=N bonds, all protein aldehyde bioconjugation strategies lead to C-C bond formation between the target protein and reagent/chemical probe, making the

aldehyde handle an ideal choice for bioconjugation studies regarding C-C linked protein bioconjguates. Each C-C bond forming protein aldehyde bioconjugation strategy has represented a significant landmark in the development of protein bioconjugation strategies for designing hydrolytically stable protein bioconjugates, and their importance in furthering the field of chemical biology cannot be understated. That being said, the protein aldehyde ligation strategies described in the literature suffer from at least one of the aforementioned drawbacks, hampering their overall utility. The *iso*-Pictet Spengler ligations[88,89] and the ABAO ligation[132] show greatly reduced reactivity at neutral pH, requiring acidic conditions, whilst the Pictet Spengler ligation,[125] Mukaiyama aldol ligation,[128] and Thiazolidinedione aldol ligation[133] require very high concentrations of reagent to achieve notable conversions to the desired protein bioconjugate. Indeed, as seen in Table 1.1, the majority of the protein aldehyde ligations strategies display poor kinetics of ligation at neutral pH.[141] The Mukaiyama aldol ligation [128] is heavily dependent on co-solvent to achieve site-selectivity, whilst cyclooctyne reagents, which are used in the SPANC,[126] are known to show off-site reactivity with thiols.[154] Additionally, the TKM[153] leads to two modifications as opposed to one, which may not always be desired depending on the application. Wittig bioconjugates have been reported to show hydrolytic instability (albeit due to the ester backbone from the probe rather than the actual Wittig bond of the bioconjugate itself),[132] whilst the Indium mediated allylation[129] leads to protein denaturation. Undoubtedly, taking all of these factors into account, there is clearly scope for the development of C-C bond forming protein bioconjugation strategies utilising aldehyde handles that operate at neutral pH, use small ratios of chemical probe, show efficient ligation kinetics, and maintain protein integrity. Such developments would represent the next important step in developing useful protein aldehyde bioconjugation chemistry, and contribute towards furthering the fields of protein bioconjugation and chemical biology as a whole.

1.5 Aims of the project

The primary aim of this project was to design new methodology for the site-selective modification of protein aldehydes *via* C-C bonds at neutral pH. Protein scope, diversity of chemical probes, and kinetics of the ligation strategy were also deemed to be of high importance, and were therefore also taken into major consideration during the design process. Additionally, in order to ensure the designed C-C bioconjugation strategy was oxygen tolerant, it was decided that, throughout the course of the project, all experiments were to be performed in the presence of air. As a bonus, the ligation strategy developed would have some downstream application in the more underdeveloped areas within the field of protein bioconjugation, such as in multi-functionalisation of proteins or in the design of useful protein bioconjugates (proteins bearing PTM mimics for example). To achieve these aims, the project was mainly focused on developing an enamine-type aldol reaction that would be suitable for protein bioconjugation. Enamine catalysed aldol reactions are one of the most well documented C-C bond forming reactions in organic synthesis involving aldehydes (and ketones), and have previously been documented for their compatibility with aqueous/buffered conditions without the need for inert conditions (see Section 3.1.7 and references within). The reaction also generates an additional carbonyl group which could be theoretically primed for further reactions, allowing for further protein modification in the context of bioconjugation. Such a strategy has yet to be reported in the literature, and, following its serendipitous discovery during the course of the project (see Section 3.1), an enamine-type aldol reaction for protein bioconjugation was considered to be a worthwhile strategy to investigate in order to achieve the aims outlined above.

Chapter 2: Preparation of α-oxo protein aldehydes

2.1 Obtaining α-oxo aldehyde containing peptides/proteins

As discussed previously in Chapter 1, aldehydes are a functional group not naturally present within the canonical amino acids, making them suitable handles for site-selective protein bioconjugation following their incorporation into a target protein. There are now seven different methods of installing aldehyde handles into proteins reported in the literature, of which there are seven different types of aldehyde handle that all display different electronic/steric properties. Of these handles, the α-oxo aldehyde handle is the most popular choice of aldehyde handle for site-selective protein aldehyde-based bioconjugation, with 11 of the 13 protein aldehyde strategies reported in the literature documented to be compatible with the handle. This aldehyde handle was therefore chosen to be the primary handle for the development of strategies for site-selective protein aldehyde bioconjugation. With this in mind, several protein scaffolds known to be (or postulated to be) compatible with α-oxo aldehyde installation were chosen for further investigation into the development of protein aldehyde bioconjugations.

2.1.1 Glyoxyl-LYRAG

For protein bioconjugation, peptides represent ideal model protein systems, as they are significantly lower in molecular weight (and therefore easier to characterise), and are easy to synthesise and customise *via* SPPS. Therefore, for this project, the six-mer peptide SLYRAG **26** (Figure 2.1A) was synthesised using SPPS. SLYRAG **26** was then oxidised with $NaIO_4$ according to previously reported literature[121] to give glyoxyl-LYRAG **27**, which, akin the glyoxyl-proteins, exists in equilibrium between both an 'aldehyde' and 'hydrate' species (**27** and **27-Hyd**, Figure 2.1B), the latter of which is the dominant form in solution. Purification of this peptide can be easily achieved through reverse phase chromatography, and the purified peptide can be stored long term at -20 $^{\circ}$C as a lyophilised powder. Glyoxyl-LYRAG **27** has previously been described as a test substrate for investigating aniline catalysed oxime ligations, [121] and was therefore considered to be a suitable, model substrate to act as mimic of a protein α-oxo aldehyde species.
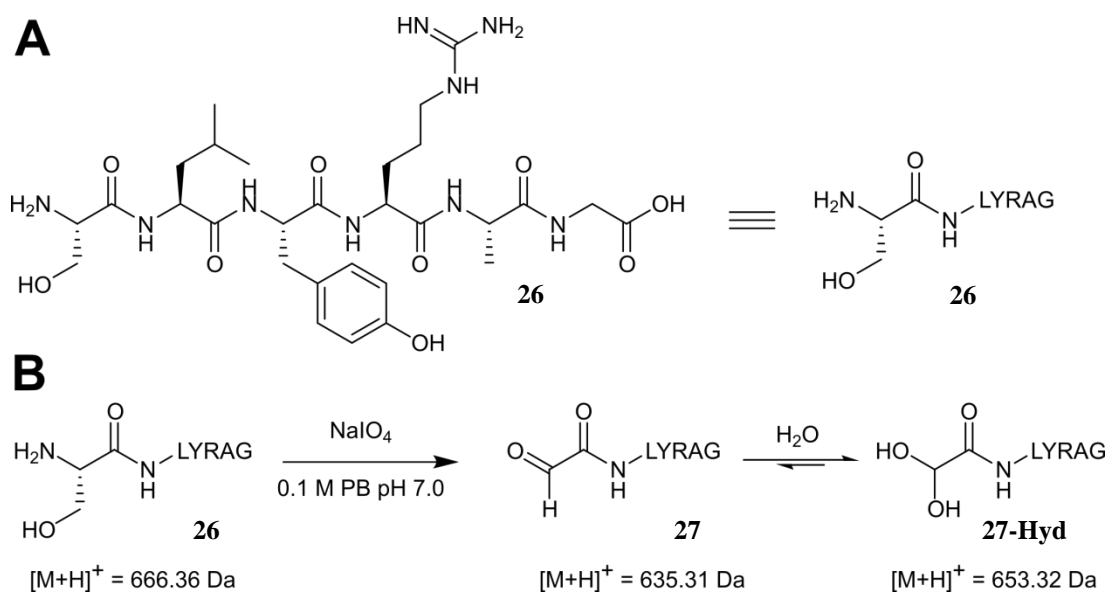
Figure 2.1 (A) Structure of SLYRAG peptide. (B) Oxidation of SLYRAG to Glyoxyl-LYRAG, which is in equilibrium as both the "aldehyde" and "hydrate" form.

2.1.2 Myoglobin

Horse heart myoglobin **28** is a small, 17 kDa protein[155], and is commonly used a model protein substrate in the design of new ligation strategies.[79,88,125,130,133] The protein is commercially available, and glyoxyl functionality can be introduced *via* a transamination reaction with PLP **8** in 25 mM PB pH 6.5 buffer to give glyoxyl-myoglobin **13** (Figure 2.2A) which primarily exists as the hydrate species (**13-Hyd**, Figure 2.2B).[79] Additionally, the protein features a haem group (responsible for carrying oxygen to muscle tissue) which has characteristic absorbance within the region of $\lambda = 410$ nm. This absorbance can be used as an indicator of protein integrity after modifications are made to the protein, as any myoglobin samples that are denatured and lose protein integrity after modification will no longer give an absorption signal within the region of $\lambda = 410$ nm.[79,128,130] There are two methods for installation of the α-oxo aldehyde into myoglobin **28** *via* biomimetic transamination: Gilmore, Francis, and co-workers 2006 method which uses 10 mM loading of PLP **8** and requires 18 h reaction times,[79] and a modified method described in 2013 by Agarwal, Bertozzi, and co-workers that uses a 100 mM loading of PLP **8** and requires 1 h reaction times.[88] Both methods use 50 μM of myoglobin **28** and are performed in 25 mM PB buffer that is adjusted to pH 6.5 using NaOH (pH

adjustment of PLP stock solutions prior to addition to myoglobin is necessary as dissolution of PLP into a 25 mM PB pH 6.5 solution causes a reduction in pH). Initially, the former transamination method was used to obtain glyoxyl-myoglobin in this project; however, as stated in the original 2006 publication and in subsequent publications in the field of biomimetic transamination chemistry, this procedure does not always lead to complete conversion of myoglobin to glyoxyl myoglobin, which affects downstream ligation yields. Through minor optimisation, it was discovered during the project that altering PLP loading from 10 mM to 15 mM resulted in a more reliable conversion of myoglobin to glyoxyl-myoglobin **13** (as judged by LC-MS in subsequent ligation experiment yields). Typical LC-MS data observed for myoglobin **28** and glyoxyl-myoglobin **13** are shown in Figure 2.2C and Figure 2.2D respectively.
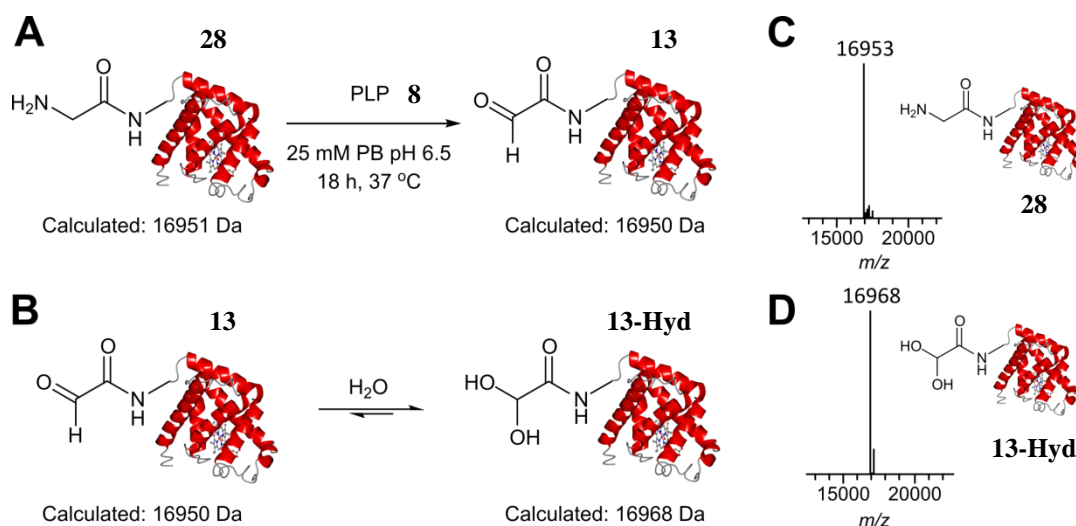
Figure 2.2 (A) Biomimetic transamination of myoglobin **28** to glyoxyl-myoglobin **13** with PLP **8**. (B) Structure of glyoxyl-myoglobin **13**, which is in equilibrium as both the "aldehyde" and "hydrate" form. (C) ESI-MS of myoglobin. (D) ESI-MS of glyoxyl-myoglobin, which is primarily observed as the "hydrate" form **13-Hyd**.

*NB: Myoglobin **28** was obtained from commercial sources as a lyophilised powder with ≥ 90% purity as judged by SDS-PAGE. Subsequent SDS-PAGE analyses of samples of myoglobin **28**, examples of which can be seen in Chapter 5, additionally confirmed the identity and purity of myoglobin **28** used throughout this project.*

2.1.3 Cholera toxin subunit B

In nature, cholera toxin subunit B (CTB) **29** is a pentameric protein that associates with a toxic, monomeric protein cholera toxin subunit A (CTA) to form an AB_5 toxin complex.[156] CTB contains carbohydrate binding sites, which allow the AB_5 complex

to bind to GM1 gangliosides on the surface of epithelial cells.[157] This facilitates endocytosis of the AB$_5$ complex, and eventual release of the toxic CTA, leading to symptoms of the disease cholera.[156] In the context of protein aldehyde chemistry, CTB **29** contains an *N*-terminal threonine residue which, akin to *N*-terminal serine, can be oxidised using NaIO$_4$ to give the corresponding glyoxyl-CTB protein **30** (Figure 2.3A) which initially exists as the hydrated species (Figure 2.3B).[77,158] LC-MS data observed for CTB **29** and glyoxyl-CTB **30** are shown in Figure 2.3C and Figure 2.3D respectively.
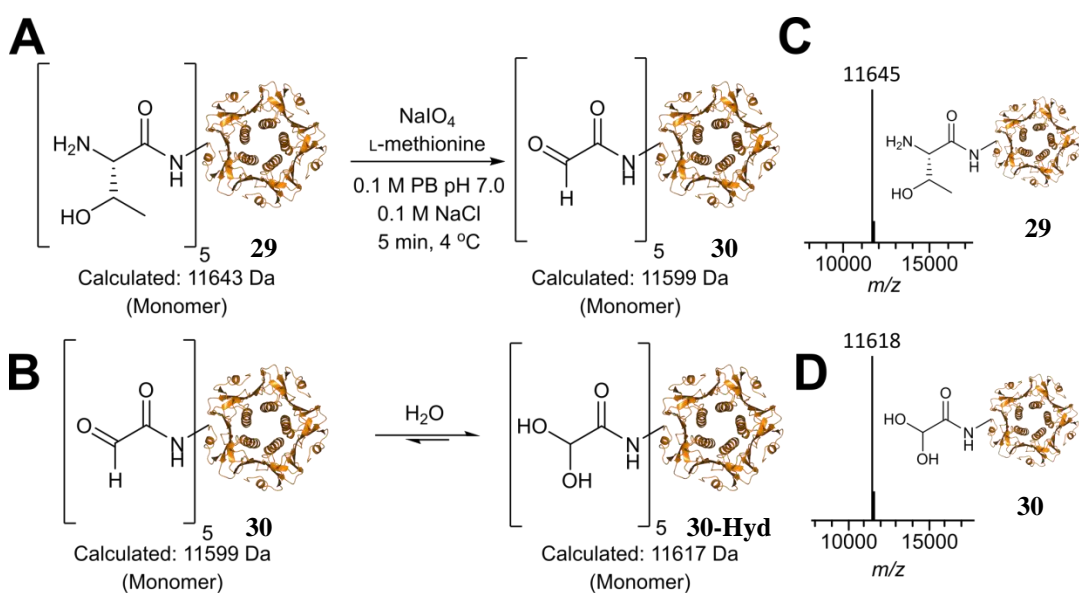
Figure 2.3 (A) Periodate-mediated oxidation of CTB **29** to glyoxyl-CTB **30**. (B) Structure of glyoxyl-CTB **30**, which is in equilibrium as both the "aldehyde" and "hydrate" form. (C) ESI-MS of CTB. (D) ESI-MS of glyoxyl-CTB, which is primarily seen as the "hydrate" form **30-Hyd**.

Conversion of CTB **29** to glyoxyl-CTB **30** is a straightforward procedure; modification of glyoxyl-CTB **30** however is challenging, as the protein contains a proline residue at position 2, which over time can lead to an intramolecular cyclisation reaction to an unreactive hemiaminal as shown in Figure 2.4.[158]
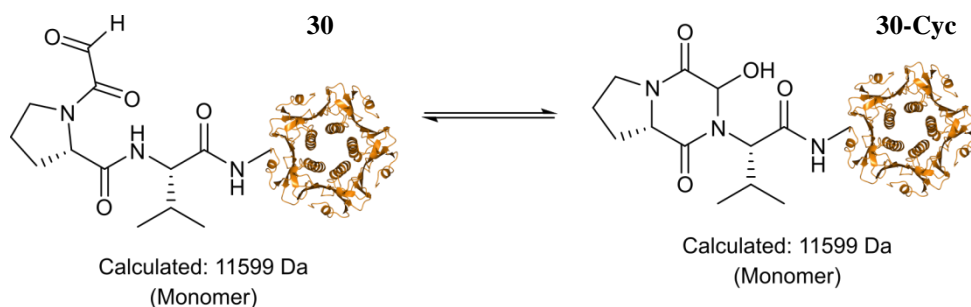
Figure 2.4 Intramolecular cyclisation of glyoxyl-CTB **30**. Nucleophilic attack of the amide bond at the α oxo-aldehyde position occurs over time, resulting in the formation of an unreactive hemiaminal **30-Cyc**.

Experimental details for producing CTB **29**, along with its oxidation, have been previously described; these were used to generate glyoxyl-CTB **30** for use in the project.[77] During this project, CTB **29** was produced under IPTG induction of gene expression, followed by ammonium sulfate precipitation and nickel purification (as outlined in Reference 77, see Experimental for further details). For the production of glyoxyl-CTB **30**, oxidation of CTB **29** was carried out in the presence of L-methionine to prevent the undesired oxidation of thiol side chains in the protein (see Figure 2.3A).

*NB: CTB **29** exists as a stable pentamer (ca. 58 kDa) which breaks down into monomers (ca. 12 kDa) if boiled or subjected to conditions used in the LC MS. SDS-PAGE analysis of unboiled and boiled samples of CTB **29**, examples of which can be seen in the Experimental section, confirmed the identity of CTB **29** and the prescence of the pentameric form following nickel purification.*

2.1.4 Thioredoxin

Thioredoxin is an oxido-reductase protein found in many domains of life, and has various roles including disulfide bond reduction.[159] Thioredoxin **31** from *E. coli* (which is commercially available) contains an *N*-terminal serine and is primed for oxidation by $NaIO_4$ to give the corresponding glyoxyl-thioredoxin **32**. Similarly to CTB, thioredoxin contains cysteine residues, which bear thiol side chains that are prone to undesired oxidation to generate sulfoxide-containing protein by-products. In order to develop a procedure for the selective oxidation of the *N*-terminus of thioredoxin, a screen of conditions with varying temperatures, concentration of $NaIO_4$, and concentration of L-methionine, was carried out as outlined in Table 2.1. Thioredoxin **31** samples for oxidation were kept at final concentration of 80 µM in 102 µL reaction solutions in 0.1 M PB, 0.1 M NaCl pH 7.0 buffer, and allowed to proceed in the dark. All samples were analysed by LC-MS and desalted prior to analysis. Conversion to glyoxyl-thioredoxin **31**, as well as thiol oxidation, could be determined by LC-MS, examples of which are shown in Figure 2.5.

Table 2.1 Conditions screened for the oxidation of thioredoxin to glyoxyl-thioredoxin. Conversions to glyoxyl-thioredoxin and undesired oxidation of thiol residues was assessed by ESI-MS analysis.

Entry	Final NaIO$_4$ conc^n (μM)	Final L-Met conc^n (μM)	Temp. (^{o}C)	Time (min)	Conversion to **31** (%)	Thiol ox (%)
1	88	0	RT	15	>99	50
2	88	0	RT	60	>99	>99
3	400	800	RT	15	>99	50
4	400	800	RT	30	>99	50
5	800	800	RT	15	>99	80
6	200	800	RT	15	>99	25
7	300	600	4	4	>99	0
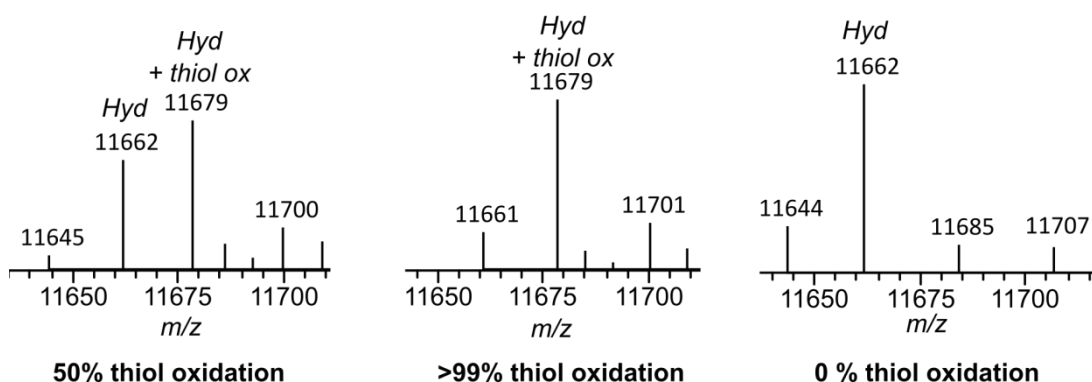
Figure 2.5 ESI-MS data of glyoxyl-thioredoxin **32** bearing different amounts of thiol oxidation following treatment with NaIO$_4$. Calculated mass of **32-Hyd** = 11661 Da, 11677 Da with thiol oxidation.

Successful selective oxidation of the *N*-terminus of thioredoxin **31** was achieved as outlined in Table 2.1, Entry 7. Holding the reaction at 4 ^{o}C, along with keeping the protein, NaIO$_4$ and L-methionine stock solutions on ice prior to performing the oxidation procedure, was found to be critical in avoiding undesired oxidation of thiol residues. This finalised set of conditions for thioredoxin **31** oxidation were taken from Table 2.1, Entry 7, and used to generate glyoxyl-thioredoxin **32** for downstream site-selective modification experiments (Figure 2.6A). Glyoxyl-thioredoxin **32** primarily exists as the hydrate species **32-Hyd** as judged by LC-MS (Figure 2.6B). Examples of LC-MS data observed for both thioredoxin **31** and glyoxyl-thioredoxin **32** are shown in Figure 2.6C and Figure 2.6D respectively.
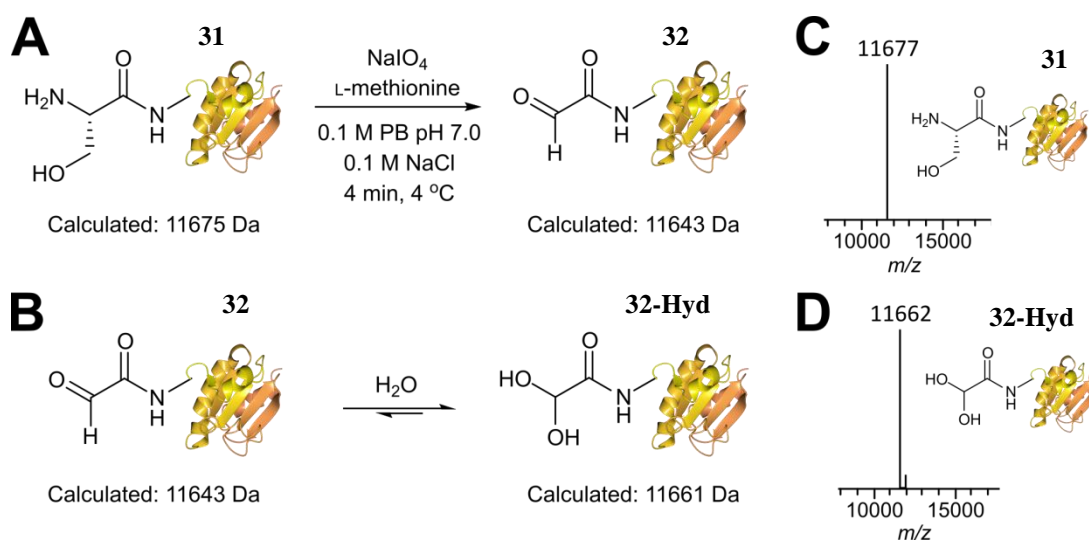
Figure 2.6 (A) Periodate-mediated oxidation of thioredoxin **31** to glyoxyl-thioredoxin **32**. (B) Structure of glyoxyl-thioredoxin **32**, which is in equilibrium as both the "aldehyde" and "hydrate" form. (C) ESI-MS of thioredoxin **31**. (D) ESI-MS of glyoxyl-thioredoxin, which is primarily seen as the "hydrate" form **32-Hyd**.

*NB: Thioredoxin **31** was obtained from commercial sources as a lyophilised powder with ≥ 90% purity as judged by SDS-PAGE. Subsequent SDS-PAGE analyses of samples of thioredoxin **31**, examples of which can be seen in Chapter 5, additionally confirmed the identity and purity of thioredoxin **31** used throughout this project.*

2.1.5 Green fluorescent protein

Green fluorescent protein (GFP) is a 27 kDa protein from the jellyfish *Aequorea victoria* that, upon exposure to blue/uv light, emits green, fluorescent light.[160-162] This fluorescence results from the PTM of amino acids S65, Y66, and G67; following protein folding, these three amino acids undergo cyclisation, dehydration, and oxidation to yield an imidazolidinone chromophore responsible for the proteins characteristic fluorescence.[163] The fluorescent properties of GFP have led to the protein becoming a staple tool in molecular biology,[164] with applications as a marker for gene expression,[161] in visualising subcellular localisation of biomolecules[165,166], and in developing protein bioconjugation strategies.[61] Indeed, the revolutionary discovery and development of GFP by Shimomura, Chalfie, and Tsien, led to them receiving the Nobel Prize in Chemistry in 2008.[167-169] Unsurprisingly, variants of GFP frequently feature as protein scaffolds for the validation and development of methodology within the field of protein aldehyde bioconjugation, including PLP transamination,[79] installation of fGly handles,[93,170] development of PFTase mediated FAPP installation,[98] and Tub tagging.[105] For the development of protein aldehyde bioconjugation strategies during this project, a GFP variant containing an *N*-terminal serine was made available during this project; additionally, this GFP variant also bears an unnatural cyclooctyne-containing amino acid at position 39, meaning that this GFP variant could also be used as a model protein scaffold for testing the orthogonality of any subsequent protein aldehyde ligation with cyclooctyne groups alongside functional groups typically found within the canonical amino acid repertoire. The *N*-terminal serine, cylcooctyne containing GFP (Ser-GFP(Y39CylcooctK), referred to as GFP **33** hereon in)[i] also contains cysteine residues; therefore, a modified procedure for the oxidation of thioredoxin was applied towards the oxidation of GFP **33** to glyoxyl-GFP **34**, and this procedure was used to generate glyoxyl-GFP **34** for use throughout the project (Figure 2.7A, see Experimental for details on preparation). As with other protein α-oxo aldehydes, glyoxyl GFP predominantly exists in the hydrated form **34-Hyd** (Figure 2B). Examples of LC-MS data observed for both GFP **33** and glyoxyl-GFP **34** are shown in Figure 2.7C and Figure 2.7D respectively.

---

[i] GFP **33** was kindly provided by Robin Brabham.
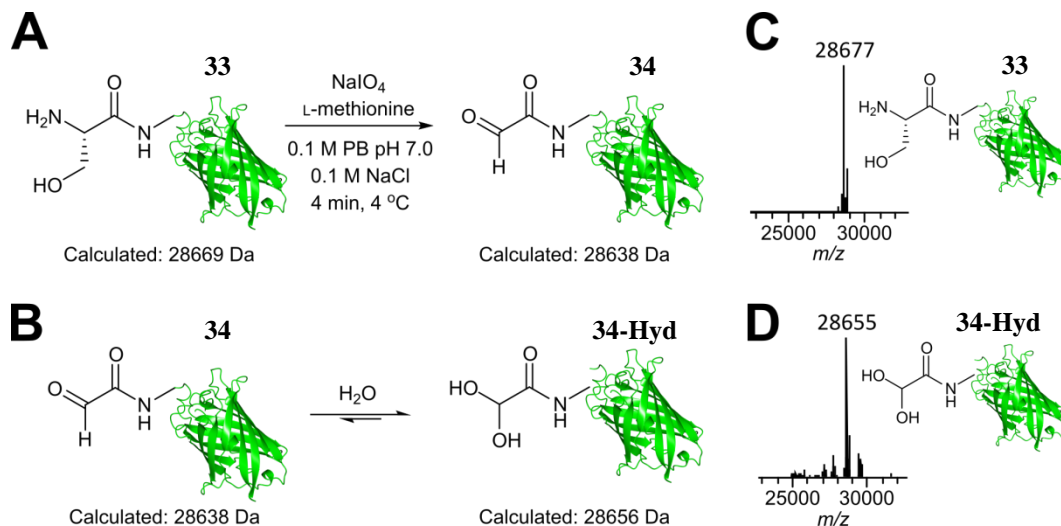
Figure 2.7 (A) Periodate-mediated oxidation of GFP **33** to glyoxyl-GFP **34**. (B) Structure of glyoxyl-GFP **34**, which is in equilibrium as both the "aldehyde" and "hydrate" form. (C) ESI-MS of GFP **33**. (D) ESI-MS of glyoxyl-GFP, which is primarily observed as the "hydrate" form **34-Hyd**.

2.1.6 Hydrophilic acylated surface protein A

Hydrophilic acylated surface proteins (HASPs) are a class of proteins with a currently unknown biological function that are expressed during the infective stages of the parasite *Leishmania donovani*.[171] They are highly immunogenic, and form part of a visceral leishmaniasis vaccine currently undergoing clinical trials in humans.[172] HASPs are enzymatically, post-translationally dually acylated at the *N*-terminal region;[173,174] specifically, HASPs undergo *N*-myristoylation, (covalent attachment of a C14 myristate group) at the amino group an *N*-terminal glycine, and *S*-palmitoylation (covalent attachment of a C16 palmitate group) at the thiol of a cysteine at position 4.[175] *N*-myristoylation is facilitated by the N-myristoyltransferase (NMT) enzyme, whereas *S*-palmitoylation is facilitated by an unknown *S*-palmitoyltransferase. Although not entirely understood, both of these PTMs are thought to facilitate membrane trafficking of the HASPs.[175]

Amongst other criteria, one of the absolute requirements for successful *N*-myristoylation of a target protein by NMT is that the protein in question must contain an exposed *N*-terminal glycine (this is true for HASPs and other proteins that undergo *N*-myristoylation).[176] As a result, HASPs (and other proteins that are targets for *N*-myristoylation) were postulated to be theoretically primed for installation of glyoxyl functionality via *N*-terminal biomimetic transamination using PLP **8**. To test this hypothesis, the procedure for biomimetic transamination of myoglobin was

applied towards installing glyoxyl functionality into HASPA **35**, a 9.5 kDa HASP. [ii]
Successful transamination of HASPA **35** to glyoxyl-HASPA **36** (Figure 2.8A, see
Figure 2.8B for schematic of equilibrium between the "aldehyde" and "hydrate"
form, and Figure 2.8C-D for relevant ESI-MS data) was confirmed as judged by LC-
MS by the appearance of a peak at 9467 Da, corresponding to the hydrated form **36-
Hyd** of glyoxyl-HASPA **36**. An additional peak at 9490 Da was also observed post
biomimetic transamination, which proved to be inert to subsequent site-selective
modification. It was anticipated that this peak was a PLP-protein related by-product,
as PLP related by-products of proteins have been previously reported in the
literature.[80] A peak at 9448 could either correspond to the "aldehyde" form of
glyoxyl-HASPA **36**, or unreacted HASPA **35**.
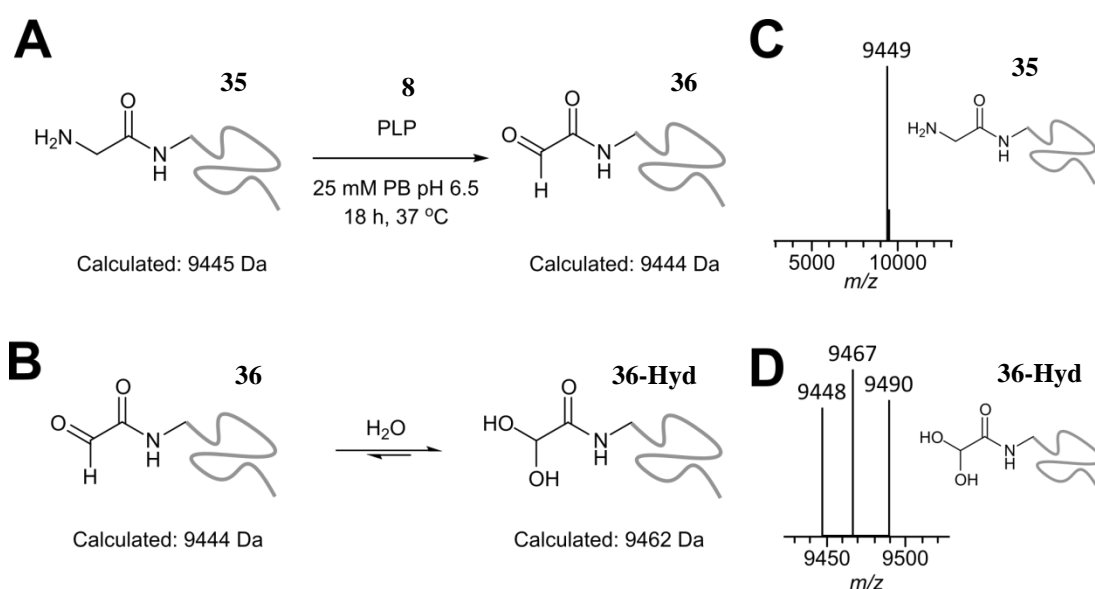
Figure 2.8 (A) Biomimetic transamination of HASPA **35** to glyoxyl-HASPA **36** with PLP **8**. (B) Structure of
glyoxyl-HASPA **36**, which is in equilibrium as both the "aldehyde" and "hydrate" form. (C) ESI-MS of HASPA
35. (D) ESI-MS of glyoxyl-HASPA, which shows the "hydrate" form of the protein **36-Hyd**, with additional
peaks corresponding to unreacted HASPA or "aldehyde" glyoxyl HASPA (9448), and an anticipated HASPA-
PLP by-product (9490).

As seen in Figure 2.8D, a significant portion of HASPA **35** (as judged by LC-MS) is
converted to the aforementioned unknown by-product following biomimetic
transamination. Therefore, in an effort avoid this by-product, and reduce the time
taken for aldehyde installation into HASPA **35**, the *N*-terminal glycine was mutated
into a serine for periodate-based oxidation to give HASAP(G1S) **37**. A modified
procedure for the oxidation of thioredoxin **31** was applied towards the oxidation of

---

[ii] HASPA **35** was kindly provided by Dr Sophie McKenna.

HASPA(G1S) [iii] **37** (Figure 2.9A) to glyoxyl-HASPA **36** (see Figure 2.9B for schematic of equilibrium between the "aldehyde" and "hydrate" form). Both HASPA(G1S) **37** and glyoxyl-HASPA **36** were analysed and confirmed by LC-MS (Figure 2.9C and Figure 2.9D respectively), with the latter compared directly to the ESI-MS data obtained when generated glyoxyl-HASPA **36** *via* biomimetic transamination (Figure 2.9E). As evidenced in Figure 2.9E, generating glyoxyl-HASPA **36** from periodate oxidation of HASPA(G1S) **37**, as opposed to biomimetic transamination of HASPA **35**, avoids formation of PLP-related by-products and gives a greater conversion overall to the desired α-oxo aldehyde.
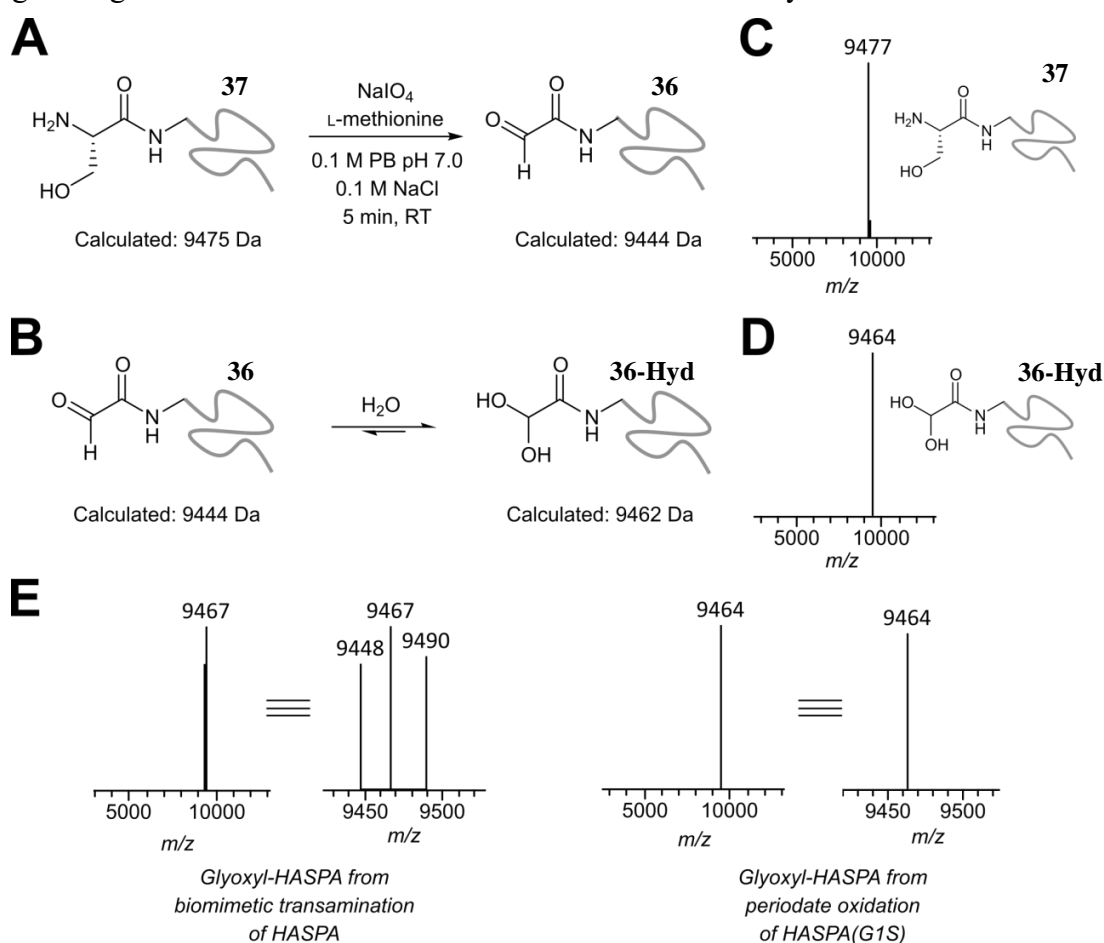
Figure 2 9 (A) Periodate-mediated oxidation of HASPA(G1S) **37** to glyoxyl-HASPA **36**. (B) Structure of glyoxyl-HASPA **36**, which is in equilibrium as both the "aldehyde" and "hydrate" form. (C) ESI-MS of HASPA(G1S) **37**. (D) ESI-MS of glyoxyl-HASPA, which is primarily observed as the "hydrate" form **36-Hyd**. (E) Comparison of ESI-MS data obtained for glyoxyl HASPA **36** generated from either biomimetic transamination of HASPA **35**, or periodate-mediated oxidation of HASPA(G1S) **37**.

*NB: During this project, 15N labelled HASPA **36** and **37** was also used. These species will be referred to as **36-15N** and **37-15N**.*

[iii] HASPA(G1S) **37** was kindly provided by Dr Tessa Keenan. (see Reference 2 preparation of the HASPA protein).

2.1.7 Thiazolidine containing green fluorescent proteins

Until recently, the α-oxo aldehyde handle had been restricted to the *N*-terminus of proteins, limiting the versatility of ligations targeting this handle. It was not until method development for the incorporation of thiazolidine amino acids such as **38** into proteins, and subsequent thiazolidine metal-mediated decaging, that the α-oxo aldehyde handle could be installed at other sites on a target protein. Superfolder GFP (sfGFP) bearing thiazolidine (sfGFPN150ThzK) **39** and GFP bearing thiazolidine (GFPY39ThzK) **40** were made available during this project[iv] (Figure 2.10A), and
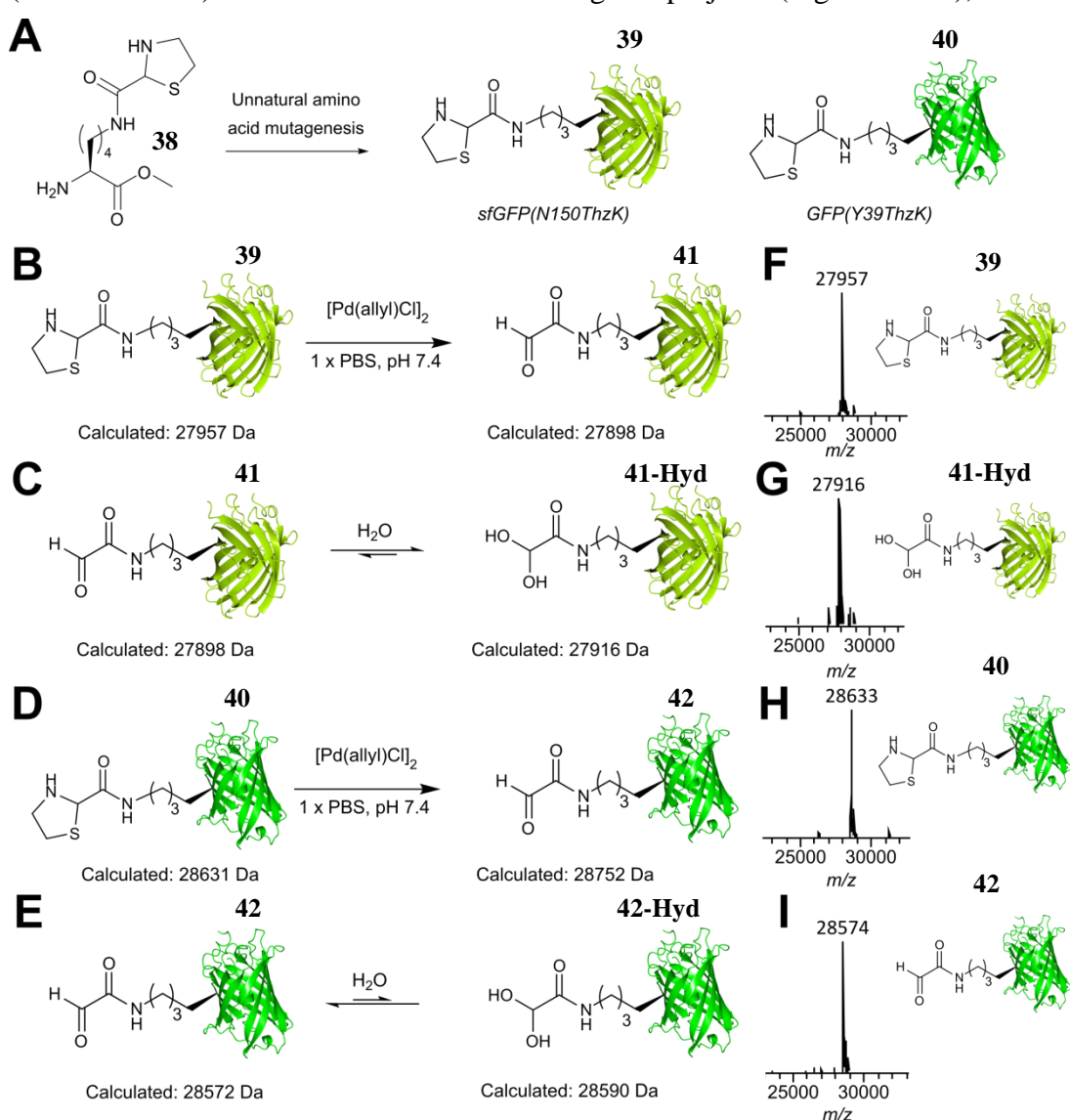
Figure 2.10 (A) Installation of an unnatural thiazolidine side chain using unnatural amino acid mutagenesis to generate sfGFP(N150ThzK) **39** and GFP(Y39ThzK) **40**. (B) Decaging of sfGFP(N150ThzK) **39**. (C) Equilibrium of the 'aldehyde' and 'hydrate' form of sfGFP(N150GlyoxylK) **41**. (D) Decaging of GFP(Y39ThzK) **42**. (E) Equilibrium of the "aldehyde" and "hydrate" form of GFP(Y39GlyoxylK) **42**. (F) ESI-MS of sfGFP(N150ThzK) **39**. (G) ESI-MS of sfGFP(N150GlyoxylK), which is primarily seen as the "hydrate" form **41-Hyd**. (H) ESI-MS of GFP(Y39ThzK) **40**. (I) ESI-MS of GFP(Y39GlyoxylK), which is primarily seen as the "aldehyde" form **42**.

[iv] sfGFPN150ThzK **39** and GFPY39ThzK **40** were kindly provided by Robin Brabham (see Reference 2 for preparation).

could be readily decaged using 1 molar equivalent of palladium (II) allyl chloride dimer ([PdCl(allyl)]$_2$.[116] Decaging of sfGFP(N150ThzK) **39** (Figure 2.10B) yields sfGFP(N150GlyoxylK) **41**, which predominantly exists in the 'hydrate' form **41-Hyd** (Figure 2.10C). Uniquely, decaging of GFP(Y39ThzK) **40** (Figure 2.10D) yields the α-oxo aldehyde-containing protein GFP(Y39GlyoxylK) **42** which, unlike other protein α-oxo aldehydes described in the literature (with the exception of glyoxyl-CTB **30**), was noted to be observed primarily as the 'aldehyde' form **41** (Figure 2.10E). Despite this phenomenon, GFP(Y39GlyoxylK) **41** appears to react readily in subsequent protein ligations,[116] and does not show inactivity in a manner similar to that seen in the 'aldehyde' form of glyoxyl-CTB **30** [158] Examples of ESI-MS data observed for the caged and decaged GFPs are shown in Figure 2.10F-I.

2.2 Conclusions

Multiple proteins bearing α-oxo aldehyde handles at both terminal and internal locations have been prepared throughout this project as, in order to show the versatility of a given protein bioconjugation strategy, it is necessary to demonstrate a given ligation procedure on multiple protein scaffolds and at different sites for conjugation. For scaffolds SLYRAG **26** and CTB **29**, strategies for α-oxo aldehyde incorporation were reported in the literature, and were subsequently used to generate α-oxo aldehyde containing protein scaffolds. For other protein scaffolds, α-oxo aldehyde installation was performed by either altering reported literature procedures (e.g. myoglobin **28**), or by screening conditions for optimal aldehyde installation (e.g. thioredoxin **31**). Altogether, these proteins provide a panel of eight protein scaffolds bearing α-oxo aldehydes which could be used to develop site-selective protein aldehyde bioconjugation strategies.

Chapter 3: Discovery of the Organocatalyst-mediated Protein Aldol Ligation

3.1 The accidental aldol

During experimentation with CTB **29**, an unexpected result was observed upon subjecting glyoxyl-CTB **30**, before purification/desalting, to basic conditions. MS data collected for these experiments suggested a species with a mass identical to that of unoxidised CTB **29** was observed when the protein was incubated at 37 °C with shaking under sufficiently basic conditions. Notably, only one modification was made onto the protein and the result could not be achieved if the protein had, after oxidation, been desalted or purified. Given that desalting/purifying the glyoxyl-CTB **30** following oxidation prevented the return of 'unmodified CTB', it was reasonable to conclude that the oxidation by-products were responsible for the unidentified protein modification. After consideration of the by-products present, a working hypothesis to explain this observation was proposed; theoretically, it was possible that the by-product acetaldehyde **43** reacts with the protein at the newly generated aldehyde position via an aldol-type reaction (Figure 3.1A), giving modified protein **44** with an observed mass (Figure 3.1B) that is almost identical to that of unoxidised CTB **29**. Such a hypothesis was considered entirely plausible, since both a Mukaiyama aldol reaction,[128] and a direct aldol reaction involving thiazolidinediones[133] have previously been described in the literature.

Figure 3.1 (A) Hypothesised reaction of glyoxyl-CTB and acetaldehyde by-product. (B) Associated ESI-MS data obtained when subjected unpurified glyoxyl-CTB to sufficiently basic conditions.
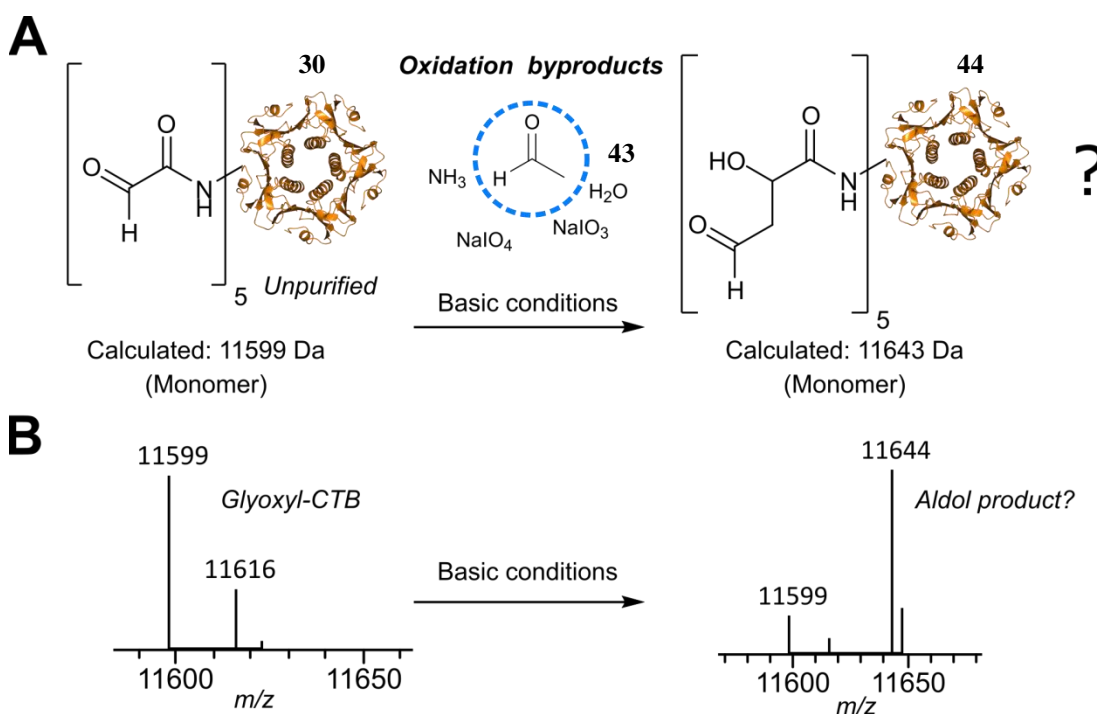
3.1.1 Testing the hypothesis

To confirm whether an aldol-type ligation would be suitable for protein modification, glyoxyl-LYRAG **27** was subjected to excess acetaldehyde **43** in 0.1 M triethanolamine bicarbonate (TEAB) buffer at pH 8.5, and allowed to react overnight at 37 °C (Figure 3.2A). The TEAB buffer system was used as it was anticipated that basic conditions would be required to deprotonate the aldehyde substrates at the α-carbon position to form enolates, and hence facilitate the aldol ligation. The reaction mixture was then analysed by LC-MS. Analysis of the LC-MS data reveal peaks corresponding to the $[M+H]^+$ species of the anticipated aldol product **45**. Gratifyingly, only one modification appeared to have taken place, as no other species that may have corresponded to uncontrolled modification of glyoxyl-LYRAG **27** were present in the LC-MS trace (Figure 3.2B).
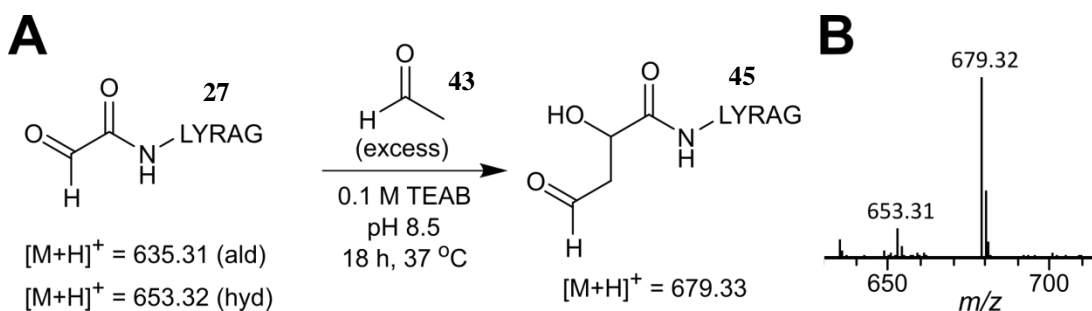
Figure 3.2 (A) Reaction of glyoxyl-LYRAG **27** and acetaldehyde **43** in 0.1 M TEAB pH 8.5 to give anticipated aldol product **45**. (B) Obtained ESI-MS data for (A).

As discussed previously, the anticipated aldol product contains an aldehyde functionality theoretically primed for further site-selective modification. The most common way of testing for aldehydes that have been incorporated into peptides/proteins in the literature is to use hydrazine or hydroxylamine reagents to form hydrazone or oxime bonds (respectively). Therefore, to test for aldehyde functionality, the pH of the crude reaction mixture containing anticipated aldol product **45** was lowered to pH 4 through addition of AcOH, and oxime ligation was attempted using aniline **46** as a catalyst, and aminooxy acetic acid **47** as shown in Figure 3.3A. The reaction was allowed to proceed overnight, and subsequently analysed by LC-MS. Conversion to the anticipated aldol-oxime product **48** was confirmed by the presence of the $[M+H]^+$ peak at 752.34. Remaining glyoxyl-LYRAG **27** also underwent modification to give oxime product **49**, suggesting that the conditions tested were suitable for successful oxime ligation. (see Figure 3.3B for

obtained MS data, and Figure 3.3C for associated structures and [M+H]$^+$ ions of each species).
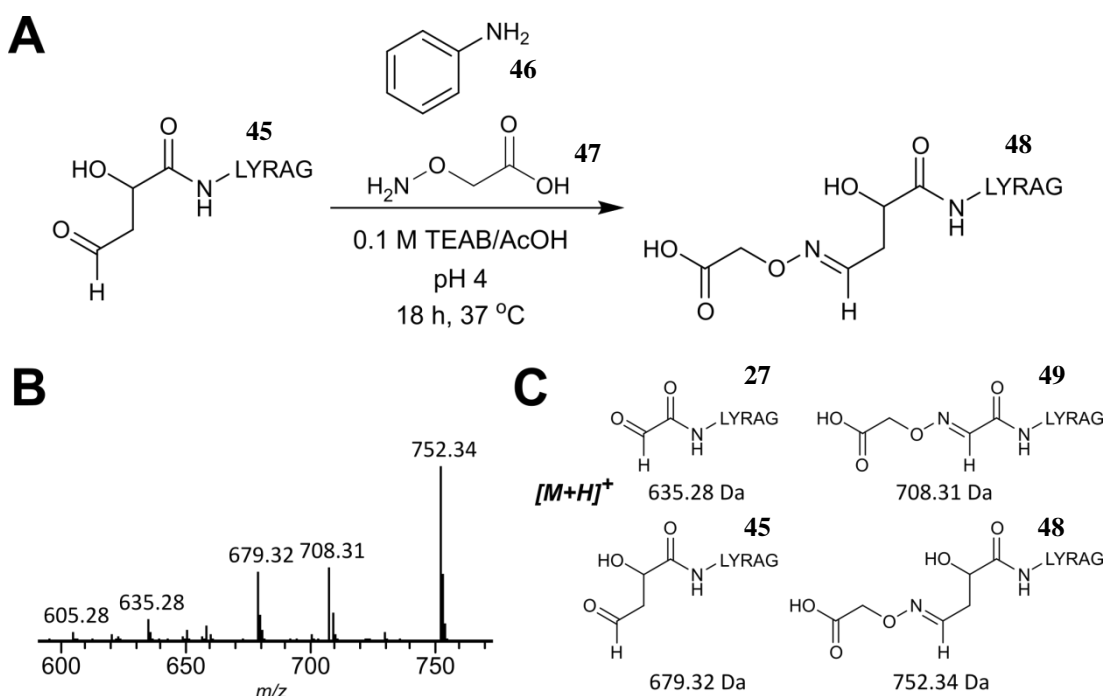
Figure 3.3 (A) Reaction of anticipated aldol product **45** and aminooxy acetic acid **47** using aniline **46** as a catalyst to give anticipated dually modified peptide **48**. (B) Obtained ESI-MS data for (A). (C) Associated structures and [M+H]$^+$ ions for each anticipated species observed in (B).

With these preliminary results in mind, it was concluded that a strategy for site-selective modification of proteins utilising an aldol reaction (of the described manner) was a viable concept, and warranted further investigation.

3.1.2 Other aldehyde donors and reactions at neutral pH

To further establish an aldol-type reaction for site-selective protein modification, aldehydes other than acetaldehyde **43** were investigated for their suitability as donor compounds (Figure 3.4A). Commercially available aldehydes propanal **50**, butanal **51**, isobutyraldehyde **52**, hexanal **53**, and phenylacetaldehyde **54** were chosen as test donor substrates to accommodate a range of donor species bearing different substituents (alkyl vs aromatic, different chain lengths etc.) and differing α proton pKa values (α proton pKa of phenylacetaldehyde = 13, whereas pKa = 17 for the other aldehyde donors). The ligations between 1 mM glyoxyl-LYRAG **27** and the aldehydes **50**-**54** were carried out in TEAB pH 8.5 buffer using 10 molar equivalents (10 equiv.) of aldehyde substrate **50**-**54** (based on glyoxyl-LYRAG **27**
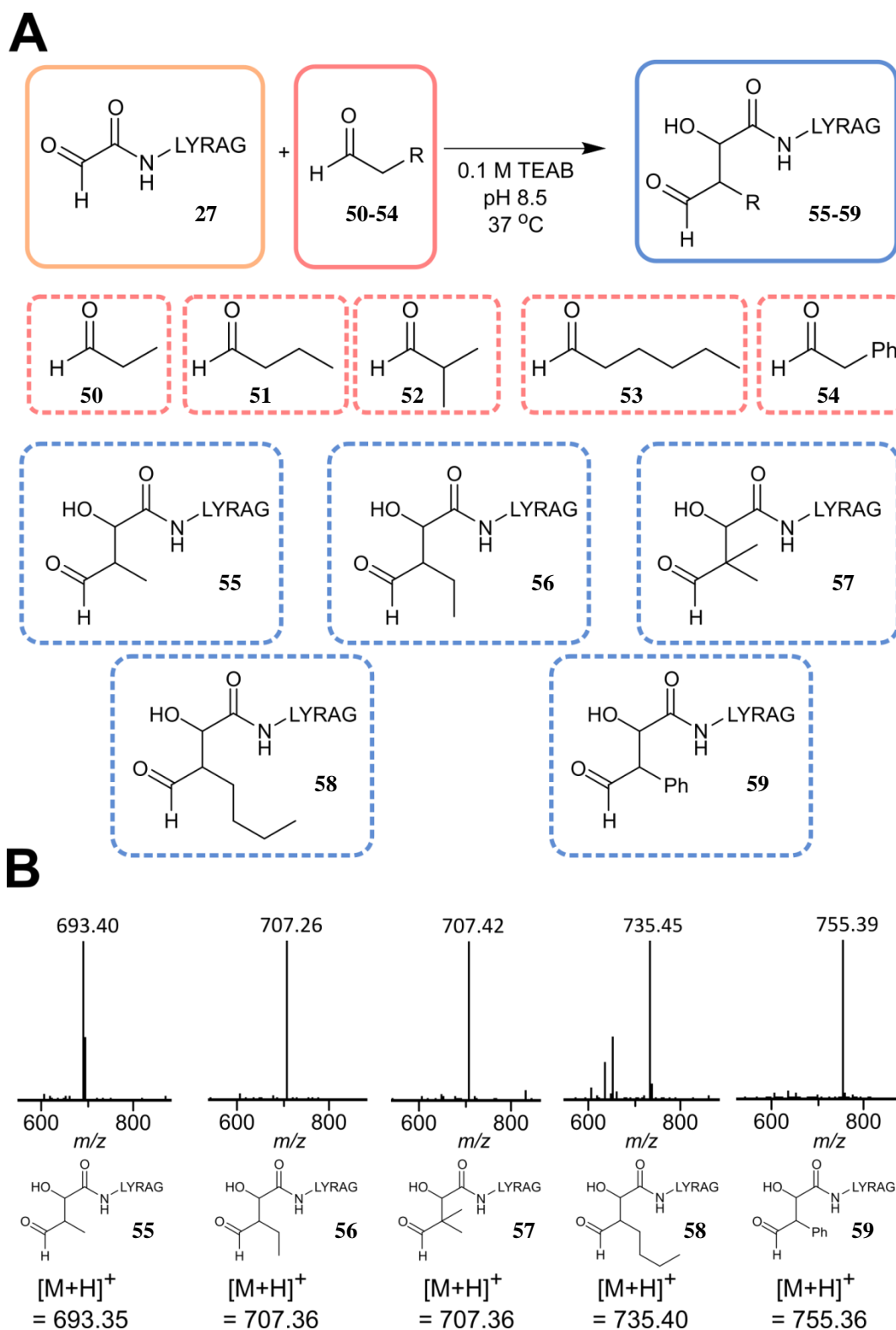
Figure 3.4 (A) Outline of anticipated aldol reaction between glyoxyl-LYRAG **27** and aldehyde donors **50-54**., and structures of associated aldehyde donors **50-54** and anticipated aldol products **55-59**. (B) Obtained ESI-MS data for each anticipated aldol product **55-59**.

concentration), with the exception of phenylacetaldehyde **54** where only one molar equivalent was used (i.e. 1 mM instead of 10 mM). This was due to precipitation

occurring in the reaction mixture when using a 10 mM loading of the donor, which was thought to arise from uncontrolled polymerisation of phenylacetaldehyde **54**. Formation of the anticipated aldol products **55-59** using the commercially available aldehydes was confirmed by LC-MS analysis, with a conversion of >99% for aldehydes **50-52** and **54**. The exception to this was when using hexanal **53** as the donor, which gave a lower conversion of 56%, possibly due to additional steric factors (an aldehyde substrate with a longer, more flexible alkyl chain will be more sterically hindered substrate, and is therefore a more challenging donor to use for modification), or due to decreased water solubility (resulting from the longer, more hydrophobic alkyl chain). Increasing the number of equivalents of hexanal **53** or longer reaction times would likely improve % conversion. Despite using only one molar equivalent of donor, the anticipated aldol ligation between glyoxyl-LYRAG **27** and phenylacetaldehyde **54** proceeded to completion. The same results could not be achieved when attempting an anticipated aldol ligation between glyoxyl-LYRAG **27** and butanal **51** in a 1:1 ratio, with a lower conversion of 61% obtained as judged by LC-MS. A possible explanation for these results is that phenylacetaldehyde **54** contains an aromatic ring substituent, which would stabilise the carbon-carbon double bond formed in the enol tautomer of phenylacetaldehyde **54**, leading a higher 'enol content' of phenylacetaldehyde **54** compared with other aldehydes.[177] Given that enol/enolate formation is required for an aldol-type mechanism to proceed, using a donor that naturally contains a higher 'enol content' would inevitably give higher rates of reaction due to increased enol/enolate availability, likely reducing the concentration of excess donor required to achieve total conversion to the anticipated aldol product.

3.1.3 Anticipated aldol ligations at neutral pH

In order to investigate the suitability of the anticipated aldol ligation in buffer systems closer to neutral pH, as opposed to basic pH, ligation between glyoxyl-LYRAG **27** and aldehydes **51** and **54** were then performed in 25 mM PB at pH 7.0, and at pH 7.5. Conversion to the desired products **56** and **59** was achieved, although the conversion to **56** dropped significantly at a more neutral pH (39% and 42% for PB pH 7.0 and PB pH 7.5 respectively). Total conversion to **59** as judged by LC-MS

was observed despite the change in pH, further highlighting the increased reactivity of phenylacetaldehyde **54** in the anticipated aldol ligation.

3.1.4 Tandem mass spectrometry analysis of anticipated aldol products

Although the obtained LC-MS data contains the $[M+H]^+$ species corresponding to the anticipated aldol product, the data provides no information on the connectivity of the observed reaction products, and therefore no information on the site-selectivity of the ligation. Tandem mass spectrometry (MS/MS) was therefore used to confirm whether any modifications of glyoxyl-LYRAG **27** were occurring site-selectively at the desired position i.e. at the *N*-terminus of the peptide. It is well documented that, for MS/MS, peptides fragment in a predictable manner to give smaller peptide fragments that can be assigned a letter (a, b, c, x, y, or z) and number (dependent on original peptide length) depending on the exact nature of fragmentation. It is then possible to assign the different peptide fragments to different *m/z* values observed in the MS/MS analysis, effectively 'mapping' the structure of the original peptide from which the fragments originated. In order to simplify analysis of modified peptide substrates by MS/MS analysis, the aldol products were treated as 'Modification-LYRAG' peptides (i.e. LYRAG peptides which have been directly modified at the *N*-terminus. Figure 3.5 shows the initial MS/MS analysis of aldol product **56**.
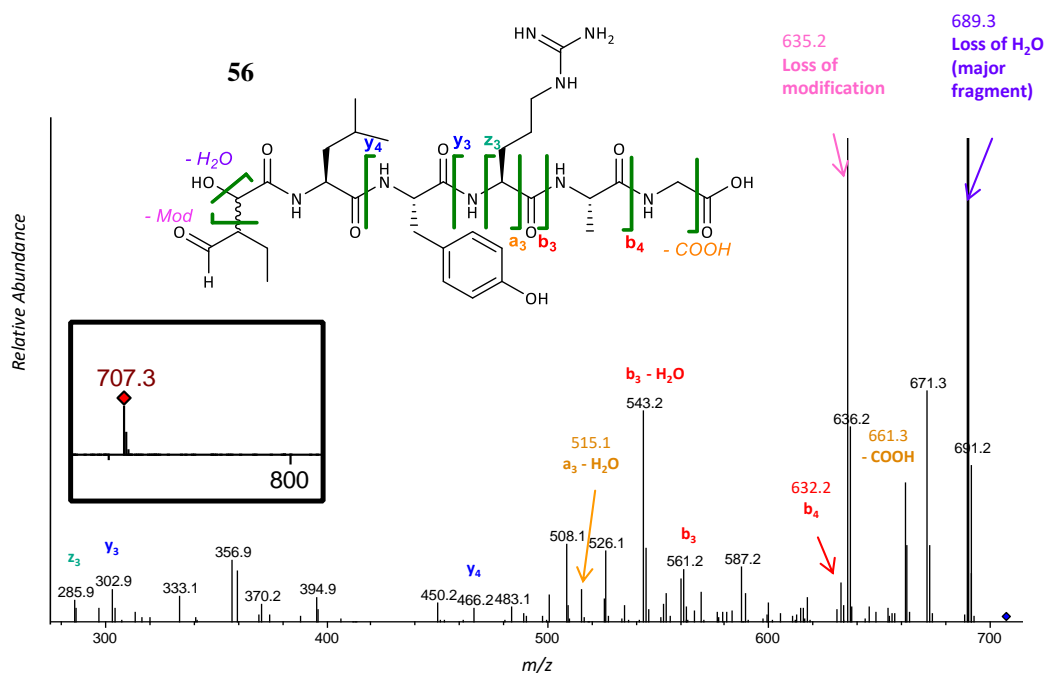
Figure 3.5 MS/MS data of aldol product **56**.

The fragmentation of the 707.3 Da species, corresponding to the anticipated aldol product **56**, was observed to be in good agreement with a site-selective aldol ligation occurring at the *N*-terminus of glyoxyl-LYRAG **27**.

The major fragment peak of **56** is 689.3 Da, corresponding to the loss of H_2O from the aldol product (707.3-18 Da = 689.3 Da). This loss would likely result from fragmentation at the newly generated hydroxyl position, as alcohol groups are known to undergo rearrangement and subsequent fragmentation in MS/MS experiments. Given the position of hydroxyl substituent, it was postulated that subsequent MS/MS analyses of the 689.3 Da major fragment peak would result in fragmentation spectra where the x/y/z fragments remain the same as that of the initial fragmentation spectra of the 707.3 Da species, but the a/b/c fragments will all change by -18 Da. This was indeed found to be the case (Figure 3.6).
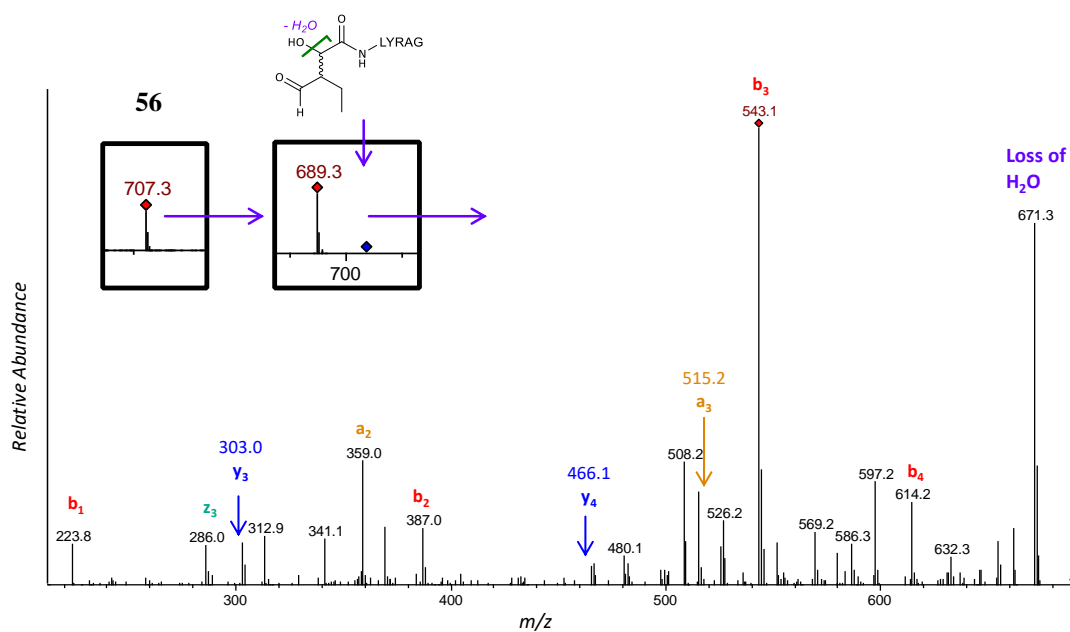
Figure 3.6 MS/MS data of fragmentation of aldol product **56**, followed by fragmentation of the resulting major fragment (689.3 Da).

These observations were considered to be consistent with the successful synthesis of aldol products such as **56**.

3.1.5 Modifying myoglobin

Having established that site-selective modification of glyoxyl-LYRAG **27** using small aldehyde substrates was possible via an aldol-type reaction, the ligation strategy was then extended to modifying glyoxyl-containing proteins. To investigate an aldol ligation on proteins, myoglobin **28** was used a model substrate. As described in Section 2.1.2, the *N*-terminal glycine of myoglobin **28** was oxidised via a transamination reaction with PLP **8** to give glyoxyl-myoglobin **13** in 25 mM PB pH 6.5. Following removal of PLP **8**, glyoxyl-myoglobin **13** was treated with different concentrations of donor **51** at pH 7.0 and pH 7.5, and conversion to the anticipated aldol product **60** was determined as judged by LC-MS (Table 3.1)

Table 3.1 Screening of conditions for aldol reaction between 100 µM glyoxyl-myoglobin **13** and donor **51** with different concentrations of **51** at pH 7.0 and 7.5. Conversion determined as judged by LC-MS.

Entry	**51** conc^n (mM)	pH	Conversion to **60** (%)
1	10	7.0	0
2	25	7.0	22
3	50	7.0	24
4	100	7.0	N/A
5	10	7.5	0
6	25	7.5	24
7	50	7.5	24
8	100	7.5	N/A

Figure 3.7A outlines the reactions described in Table 3.1, and Figure 3.7B shows the LC-MS data obtained for Entry 6 of Table 3.1.
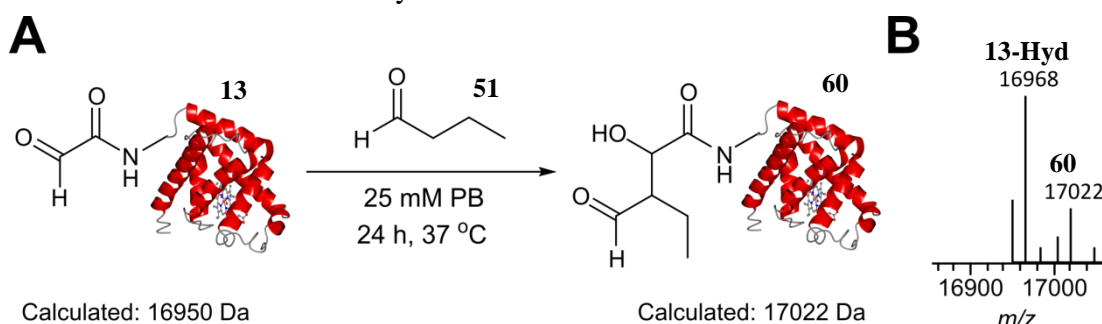
Figure 3.7 (A) Outline of reaction between glyoxyl-myoglobin **13** and donor **51** (B) LC-MS data obtained for reaction 100 µM glyoxyl-myoglobin **13** and 25 mM donor **51** in 25 mM PB pH 7.5 at 37 °C for 24 h. (see Table 3.1, Entry 6.)

Although peaks corresponding to the anticipated aldol product **60** were present in some of the conditions screened described in Table 3.1, extremely low conversions of glyoxyl-myoglobin **13** to the anticipated aldol product **60** were achieved, despite using a large excess of donor **51**. In some cases, such as Table 3.1 Entry 4 and Entry 8, extremely high concentrations of donor **51** resulted in a loss of protein species when analysed by LC-MS. Changing the buffer system from PB pH 7.0 to PB pH 7.5 had negligible effects on the % conversion obtained.

3.1.6 Routes to optimisation

It was evident that, from the results obtained as described earlier, significant optimisation was required in order to make the aldol ligation a viable strategy for site-selective protein bioconjugation. To achieve this, three strategies were considered:

Strategy 1: Raise the pH of the reaction. A higher pH should facilitate deprotonation at the α-carbon position of the donor substrate, assisting in formation of the enolate required for the aldol reaction to proceed, and thus resulting in greater conversion to the desired product. However, this strategy is undesirable as it not only moves the ligation strategy away from the desired conditions of neutral pH, but could also lead to deprotonation of the ε-amino group of lysine residues under sufficiently basic conditions, which may lead to non-selective protein modification through nucleophilic attack of the ε-amino group of lysine at the carbonyl of the aldehyde donor (lysine bioconjugation is frequently performed at basic pH). [21,22]

Strategy 2: Increase the donor concentration. Successful protein bioconjugation usually requires an excess of reagent to drive the reactions to completion. Due to the large excess of donor compound already being used to obtain minimal protein modification, however, this was not considered a viable strategy for optimisation.

Strategy 3: Use a catalyst. Other protein bioconjugation strategies have shown that addition of a catalyst can greatly facilitate the desired ligation, the most notable example arguably seen in aniline catalysed oxime ligations.[121,123] Employing a catalyst in the protein aldol ligation could potentially lead to a similar success. This was considered to be the most viable strategy, and was therefore investigated further.

3.1.7 The Organocatalyst-mediated Protein Aldol Ligation

For small molecules, the amino acid L-proline **61** is well-established as a catalyst for asymmetric aldol reactions. Dating back to the 1970s, the use of L-proline **61** as a catalyst was first reported independently both by the group of Eder, Sauer, and Weichert,[178] and the group of Hajos and Parrish,[179] in what is now known as the Hajos–Parrish–Eder–Sauer–Wiechert reaction. Major developments in proline catalysed asymmetric aldol reactions were then reported in the 2000s. List, Lerner, and Barbas demonstrated direct asymmetric aldol reactions between acetone **62** and aldehyde acceptors under non-inert, room temperature conditions could be performed using catalytic amounts (30 mol%) of L-proline **61** in DMSO. [180] Based on mechanistic observations of aldolase catalytic antibodies,[181] an enamine mechanism for the proline catalysed aldol reaction was proposed, with the noted enantioselectivity of the reaction deduced to arise from a Zimmerman-Traxler type transition state (Figure 3.8A).[180,182] Proline catalysed aldol reactions were further extended to direct, enantioselective cross aldol reactions of non-equivalent aldehydes by Northrup and Macmillan.[183] The proline catalysed cross aldol strategy was also demonstrated in the enantioselective synthesis of protected carbohydrates including glucose, mannose and allose.[184,185] Here, an enantioselective aldol reaction (catalysed by proline) between two α-oxy aldehydes was used to synthesise β-hydroxy aldehydes, which act as reagents in Lewis Acid mediated Mukaiyama Aldol condensations to yield the desired protected sugars. Critically, the β-hydroxy aldehyde intermediates synthesised via this method were found to be inert to further proline catalysed enolisation or enamine aldol additions, preventing uncontrolled polymerisation of α-oxyaldehyde starting materials whilst allowing for further functionalisation in a controlled manner using a different reaction (Figure 3.8B).[185]
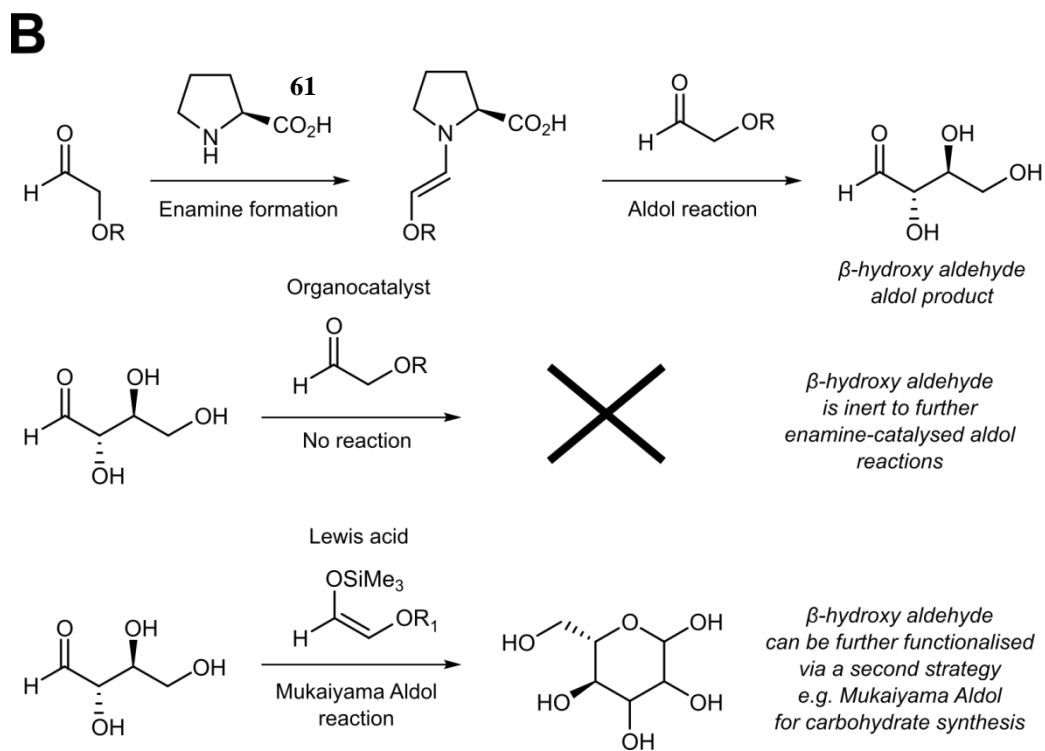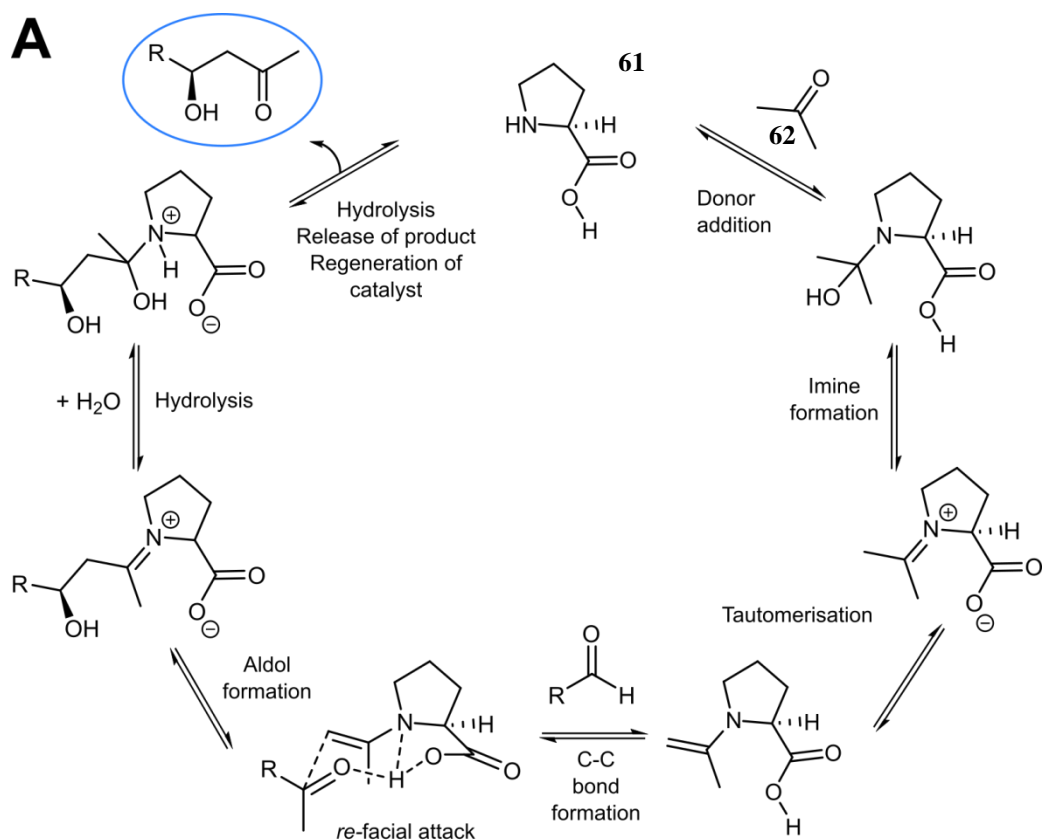
Figure 3.8 (A) A Zimmerman-Traxler type transition state for enantioselective synthesis of β-hydroxy aldehyde products catalysted by L-proline 61. (B) Reactivity of the β-hydroxy aldehyde.

The use of L-proline **61** as catalyst for the OPAL was therefore investigated. In order to test the efficacy of L-proline **61** as a catalyst, an initial screen using glyoxyl-LYRAG **27** and aldehyde **51** in PB pH 7.0, and PB pH 7.5, was performed.

Traditionally, organocatalytic reactions involving catalysts typically employ catalytic amounts (e.g. 20 mol%). However, in the context of peptide/protein bioconjugations, concentration of organocatalyst used is usually much greater than this, as seen in aniline-catalysed hydrazone/oxime ligations.[118,123] Therefore, two reaction protocols were attempted: one employing 20 mol % L-proline **61** (0.2 equiv.), and the other employing 20 equiv. (20 mM) of L-proline **61**. The reaction of 1 mM glyoxyl-LYRAG **27** and 10 mM butanal **51** was carried out using either 0.2 mM or 20 mM L-proline **61** in PB pH 7.0 and PB pH 7.5 for 24 h at 37 °C, and the conversion to aldol product **56** was judged by LC-MS. The results are described in Table 3.2:

Table 3.2 Screening of conditions for aldol reaction between 1 mM glyoxyl-LYRAG **27** and 10 mM donor **51** with different concentrations of **61** at pH 7.0 and 7.5. Conversion determined as judged by LC-MS.

Entry	L-Proline **61** conc^n (mM)	pH	Conversion %
1	0.2	7.0	33
2	20	7.0	>99
3	0.2	7.5	58
4	20	7.5	>99

Figure 3.9A outlines the reactions described in Table 3.2, and Figure 3.9B shows the LC-MS data obtained for Entry 4 of Table 3.2.
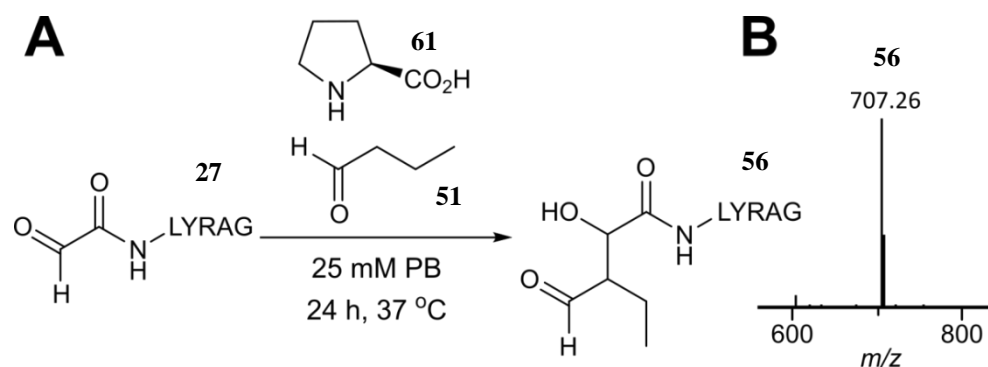
Figure 3.9 (A) Outline of reaction between glyoxyl-LYRAG **27** and donor **51** using L-proline **61** as a catalyst (B) LC-MS data obtained for reaction 1 mM glyoxyl-LYRAG **27** and 10 mM donor **51** using 20 mM L-proline **61** in 25 mM PB pH 7.5 for 24 h at 37 °C. (see Table 3.2, Entry 4).

The addition of 0.2 mM. L-proline **61** to the peptide aldol ligation did not appear to cause any significant improvement to conversion to aldol product **56** after 24 h (see 3.1.3). The addition of 20 mM of L-proline **61**, however, leads to total conversion of glyoxyl-LYRAG **27** to aldol product **56** as judged by LC-MS analysis under identical conditions at both pH 7.0 and pH 7.5, demonstrating the catalytic effects of L-proline **61** in this particular reaction.

The introduction of L-proline **61** within the context of protein modification was next investigated. Aldol reactions between 100 µM glyoxyl-myoglobin **13** and 25 mM butanal **51** were performed with either a 10 mM or 25 mM loading of L-proline **61** at 37 °C. The reactions were monitored after 1 h, 2 h, 4 h, and 6 h of reaction time, and conversion to the anticipated aldol product **60** was confirmed by LC-MS analysis (see Table 3.3).

Table 3.3 Screening of conditions for the reaction between 100 µM glyoxyl-myoglobin **13** and 25 mM donor **51** in PB pH 7.5 with different concentrations of **61** at different time intervals. Conversion determined by LC-MS.

Entry	L-Proline **61** conc^n (mM)	Reaction time (h)	Conversion (%)
1	10	1	30
2	10	2	50
3	10	4	65
4	10	6	75
5	25	1	44
6	25	2	63
7	25	4	82
8	25	6	>99

Figure 3.10A outlines the reactions described in Table 3.3, and Figure 3.10B shows the LC-MS data obtained for Entry 8 of Table 3.3.
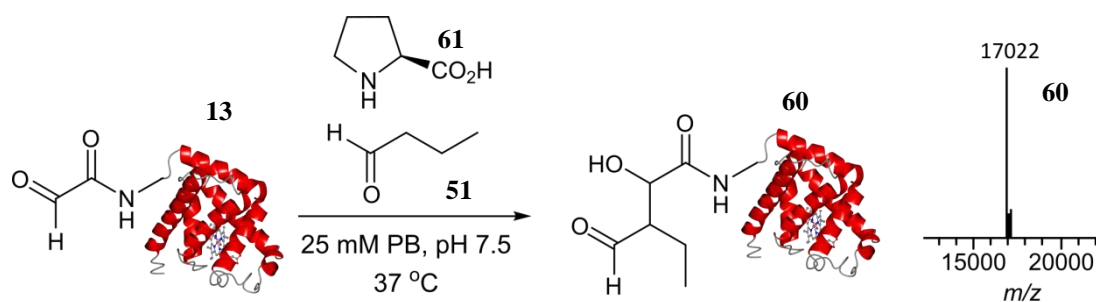
Figure 3.10 (A) Outline of reaction between glyoxyl-myoglobin **13** and donor **51** using L-proline **61** as a catalyst (B) LC-MS data obtained for reaction 100 μM glyoxyl-myoglobin **13** and 25 mM donor **51** using 25 mM L-proline **61** in 25 mM PB pH 7.5 at 37 °C for 6 h. (see Table 3.3, Entry 8.)

As seen in the peptide aldol ligations, the use of L-proline **61** in the anticipated aldol reaction between glyoxyl-myoglobin **13** and butanal **51** was found to vastly improve conversion to the anticipated aldol product **60** as judged by LC-MS analysis, with full conversion observed within 6 h when using a 25 mM loading of L-proline **61**.

3.1.8 Trypsin digest experiments on modified myoglobin

To determine that the observed modification of glyoxyl-myoglobin **13** was occurring at the desired position i.e. the *N*-terminus, a proteolytic digest using trypsin, followed by LC-MS/MS analysis, was performed. Trypsin is a serine protease that cleaves proteins at the carboxyl end of lysine and arginine residues (unless they are followed by a proline),[186] generating smaller peptide units (Figure 3.11A). With the knowledge of the amino acid sequence of myoglobin **28**, peptide units generated from a trypsin digest of myoglobin **28** can be predicted, based on the position of lysine and arginine residues within the proteins sequence. The first occurrence of a lysine or arginine residue in the sequence of myoglobin **28** (reading from the *N*-terminus to the *C*-terminus) is lysine 16 (K16). Cleavage by trypsin at this position would result in the '*N*-terminal' peptide unit GLSDGEWQQVLNVWGK. Similarly, a trypsin digest of glyoxyl myoglobin **13** would give the *N*-terminal peptide unit glyoxyl-LSDGEWQQVLNVWGK. If the anticipated aldol product **60** has been synthesised, then a trypsin digest of the aldol product **60** would give the *N*-terminal peptide unit 'Modification-LSDGEWQQVLNVWGK' with a calculated $[M+H]^+$ of 1887 Da (Figure 3.11B). MS/MS analysis of this peptide unit could then be used to confirm that the modification was occurring at the desired position i.e. at the *N*-terminus of the protein. A trypsin digest was therefore performed on the anticipated aldol product **60** (synthesised using the conditions described in Table 3.3, confirmed

by LC-MS). Following separation and analysis of the tryptic peptides by LC-MS, a species corresponding to 1886.9 Da was observed (Figure 3.11C).
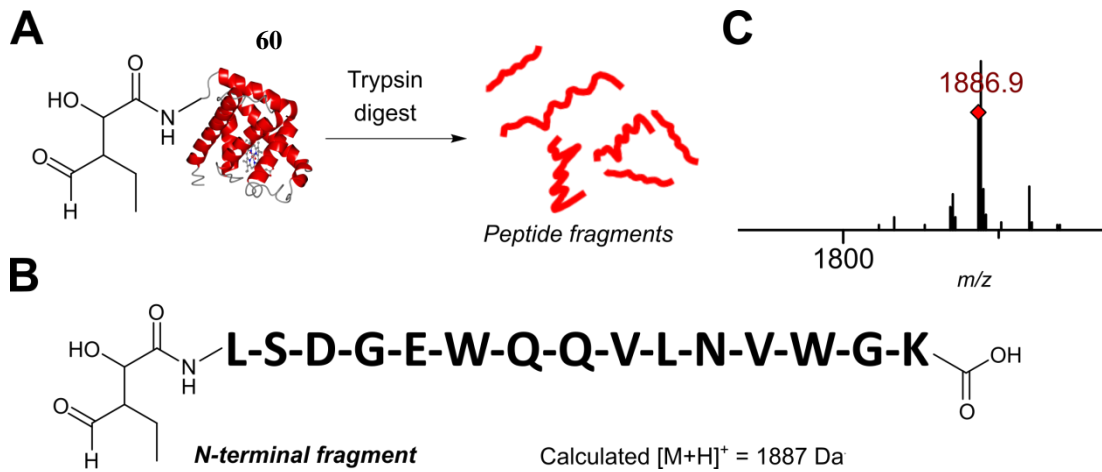
Figure 3.11 (A) Schematic of tryptic digest of aldol product **60** to give peptide fragments. (B) Structure of the *N*-terminal fragment theoretically generated from tryptic digest of anticipated aldol product **60**. (C) Species at 1886.9 Da observed in LC-MS analysis of trypsin digest of anticipated aldol product **60**, which potentially corresponds to the anticipated *N*-terminal fragment described in (B).

This peak was subjected to MS/MS analysis, and the resulting peaks were assigned to potential fragments that would theoretically result from the *N*-terminal fragment of aldol product **60**, as seen in Figure 3.11:
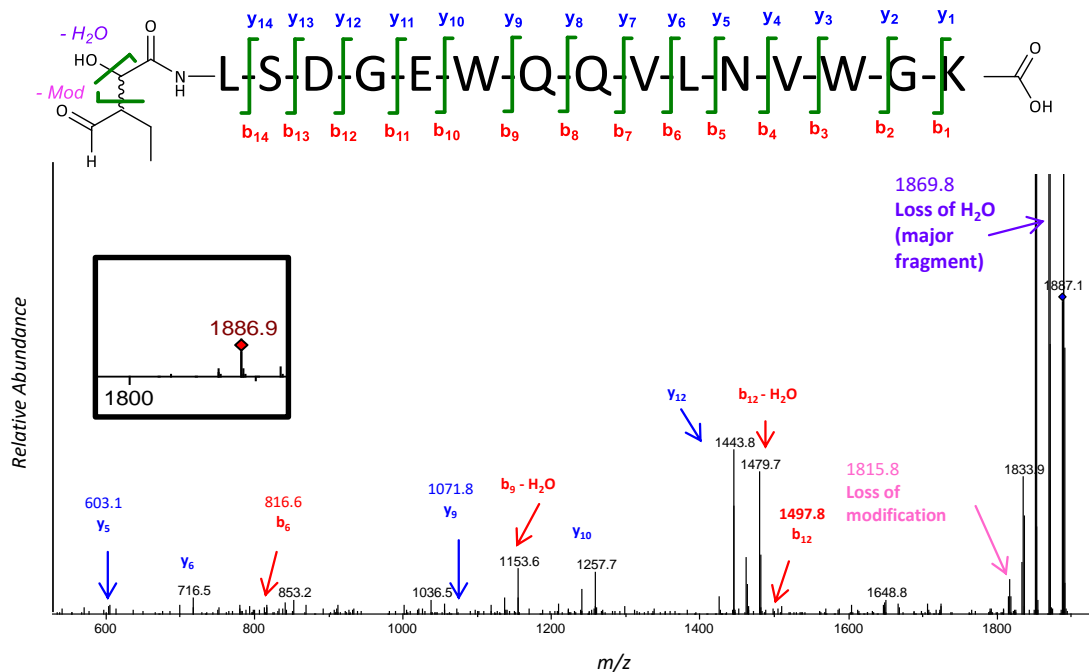
Figure 3.12 MS/MS of anticipated *N*-terminal fragment of aldol product **60** resulting from trypsin digest

The major fragment observed from the initial MS/MS was found to be a peak at 1869.8 Da. Similar to the MS/MS analysis of aldol products, the major fragment peak corresponds to a loss of 18 Da, and presumably results from fragmentation at the newly generated hydroxyl position of the aldol product. As a result, MS/MS analyses of this major fragment peak should result in fragmentation spectra where the predicted x/y/z fragments remain the same compared to predicted fragmentation spectra of the OPAL product **60**, but the predicted a/b/c fragments will all change by -18 Da. As seen in Figure 3.13, this was indeed found to be the case:
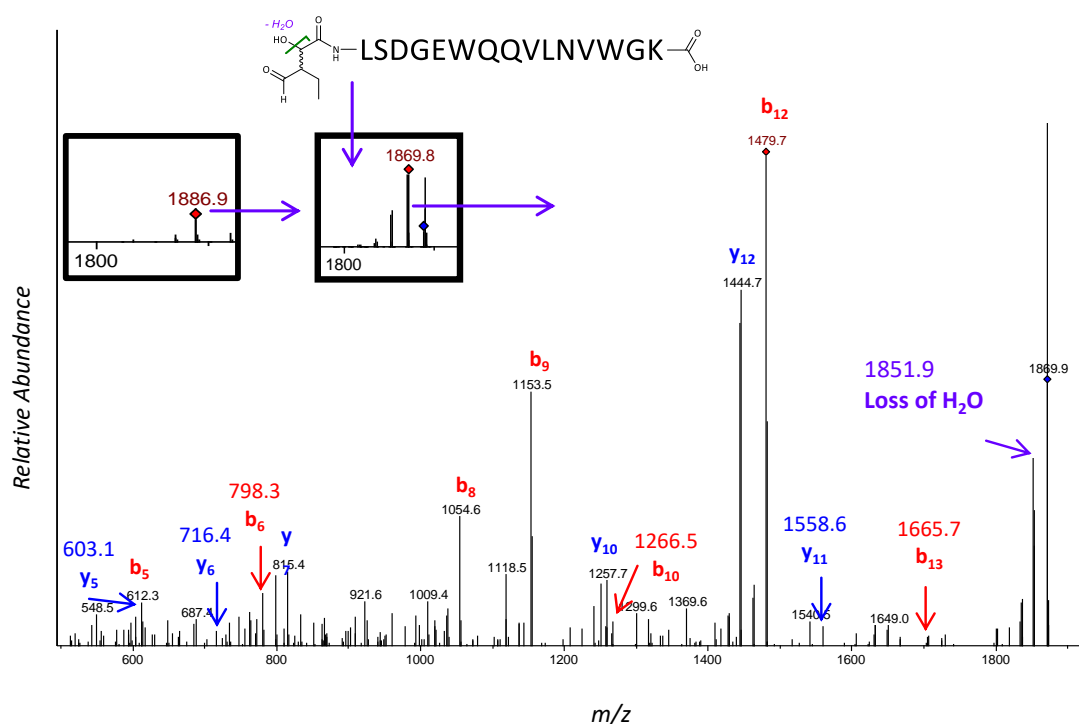
Figure 3.13 MS/MS, followed by MS/MS of the major fragment of the anticipated *N*-terminal fragment of aldol product **60** resulting from trypsin digestion.

Overall, the resulting fragments from the 1869.8 Da fragment confirm both the presence of the β-hydroxy aldehyde group, and that the aldol modification has occurred site-selectively at the glyoxyl position of glyoxyl-myoglobin **13**.

3.1.9 UV/Vis data

As previously discussed, the haem group of myoglobin exhibits a strong absorbance at λ = 410 nm, which can be used as an indicator of protein integrity after modifications are made to the protein.[79,128,130] The UV/vis spectra of unmodified myoglobin **28** and OPAL product **60** were therefore recorded, and compared as shown in Figure 3.14.
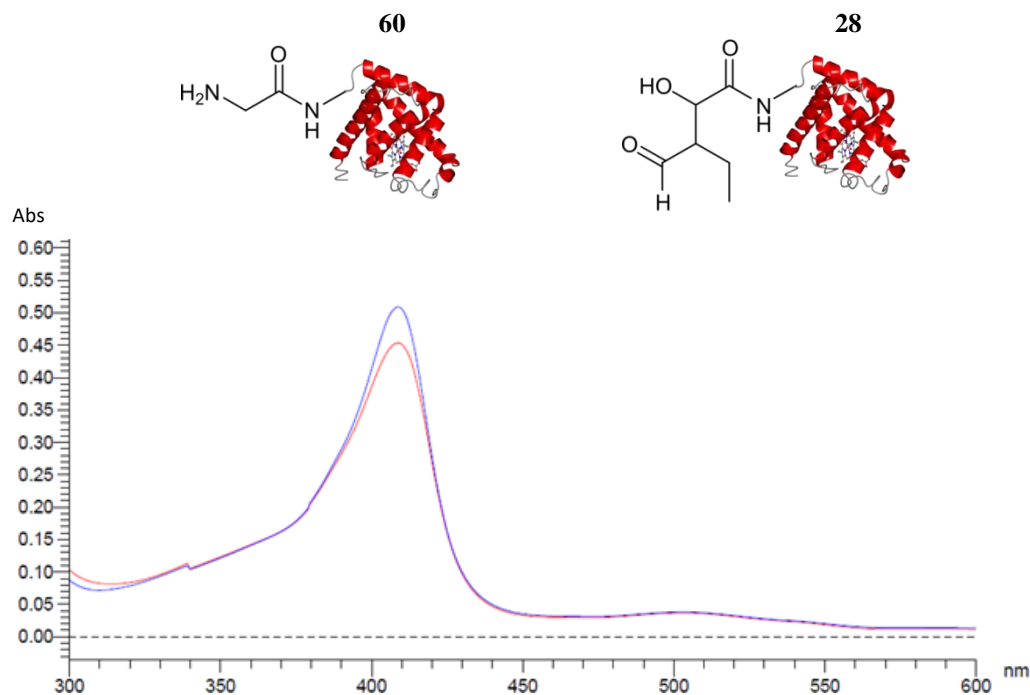
Figure 3.14 UV-Vis measurements of myoglobin **28** (red line) and aldol product **60** (blue line).

The UV/vis spectrum of OPAL product **60** was found to be of an excellent match to that of unmodified myoglobin **28**, suggesting that the protein integrity, and the haem group itself, remained unchanged post modification.

3.2. Further development of the OPAL strategy

From what began as an undesired side reaction, the OPAL has proved to be a suitable procedure for site-selective labelling of α-oxo aldehyde containing proteins. This novel ligation strategy was first validated using the α-oxo aldehyde containing peptide glyoxyl-LYRAG **27** and commercially available small molecule aldehyde, and the site-selectivity of the reaction was confirmed using a combination of LC-MS and MS/MS analysis. Furthermore, the presence of β-hydroxy aldehyde functionality resulting from aldol ligation was confirmed by downstream oxime ligation using the

catalyst aniline **46** and aminooxy acetic acid **47**. By incorporating the catalyst L-proline **61** into the bioconjugation protocol, the aldol ligation (deemed the Organocatalyst mediated Protein Aldol ligation, abbreviated as "OPAL") could be efficiently carried out at neutral pH, allowing for quantitative labelling of peptide and protein α-oxo aldehydes with small molecule aldehydes within 24 h at pH 7.5. The site-selectivity of the OPAL on myoglobin **28** was confirmed by performing trypsin digestion of the aldol product **60**, followed by MS/MS analysis of the resulting peptide fragments. Finally, UV/Vis analysis of the modified myoglobin species **60** confirmed the haem group was intact post modification, suggesting the protein integrity was maintained.

Overall, the OPAL, as described in Chapter 3 offers notable advantages over other protein bioconjugation strategies, particularly in its ability to synthesise hydrolytically stable conjugates *via* C-C bond formation at neutral pH without compromising protein integrity. A significant drawback of the OPAL procedure described in this section, however, are the long reaction times and a high concentration of probe required for efficient labelling. Although many previously reported bioconjugation strategies frequently use long reaction times and large excesses of reagent to force ligations to completion, there has recently been an increasing push towards bioconjugation strategies that use equimolar or near equimolar concentrations of probe to protein. Additionally, from a practical perspective; the longer a protein modification takes the longer the protein spends outside of its optimal storage conditions, and the higher the concentration of label required the more rigorous the purification needed following modification. Consideration of these factors was deemed to be particularly important if the protein in question is a target for further downstream site-selective modifications via the β-hydroxy aldehyde handle, where the target protein will spend even longer outside of its optimal storage conditions.

To further develop the OPAL as an effective bioconjugation strategy, it was clear that optimisation of the reaction conditions would be required. Many protein bioconjugation strategies described in the literature have benefitted from alterations to their reactive partners; for protein aldehyde ligations this has been most evident in the Pictet Spengler series of bioconjugations (higher rate constants in the *iso*-Pictet Spengler ligation, more applicable at pH closer to neutral in the HIPS),[88,89,125] and

has also been reported in the ABAO ligation (higher rate constants).[132] In a similar vein, aniline-catalysted hydrazine/oxime ligations have been the subject of major optimisation through research into alternative catalysts that offer superior ligation kinetics to aniline **46** .[119,139] Therefore, it was evident that the most fruitful areas to research for optimisation of the OPAL would be the identity of both the catalyst and donor used in the ligation protocol.

3.3 Conclusions

To conclude, in the form described within this Chapter, the OPAL proved to be a suitable method for the site-selective modification of peptide and protein α-oxo aldehyde acceptors, using α carbon-containing aldehydes as donors. Protein modification via this method was facilitated using the highly water soluble, commercially available catalyst L-proline **61**, and is thought to proceed by an enamine type mechanism. Finally, the OPAL does not compromise protein integrity post modification, leads to carbon-carbon bond formation between protein and probe, and occurs readily at neutral pH.

Chapter 4: Site-selective protein modification via an optimised OPAL strategy

4.1 Choice of organocatalyst

Since the disclosure of ʟ-proline **61** as an organocatalyst for the aldol reaction, numerous different catalysts have since been reported that offer improvements in catalytic activity, enantioselectivity/diastereoselectivity of product formation, and catalyst solubility. It was therefore envisioned that using a different catalyst to ʟ-proline in the OPAL may lead to a ligation protocol with improved reaction kinetics i.e. higher conversions to the desired OPAL product on a faster timescale. In contrast to other aldol reactions described in the literature, stereochemical control of the OPAL products synthesised by a given catalyst was deemed unimportant; as long as a given compound could theoretically participate in enamine-type catalysis, it could be deemed to be a potential catalyst for the OPAL. This would, however, lead to an enormous amount of compounds to screen in the optimisation of the OPAL, as any primary or secondary amine (which can participate in initial imine formation) would theoretically be a catalyst for the OPAL. Therefore, potential catalysts were also required to have desirable properties, such as having good water solubility, and being commercially available. Catalysts were also chosen on the basis of whether or not they would offer an insight into what properties make a given compound a more (or less) effective organocatalyst compared to ʟ-proline **61** within the context of the OPAL. Based on this, a panel of 10 secondary amines (including ʟ-proline **61**) were chosen to be investigated in their ability to catalyse the OPAL procedure. This panel of catalysts includes pyrrolidine **63**, piperidine **64**, morpholine **65**, ʟ-prolinamide **66**, 2-azetidine carboxylic acid **67**, ᴅ-proline **68**, homomorpholine **69**, (S)-(–)-5-(2-pyrrolidinyl)-1H-tetrazole **70** (referred to as proline tetrazole **70** hereon in), and piperazine **71**. Catalysts **63**-**71**, shown in Figure 4.1 (alongside ʟ-proline **61**), are water soluble, commercially available, and display subtle yet potentially important differences to that of ʟ-proline, such as different side chains, ring size, and amino group pKa.
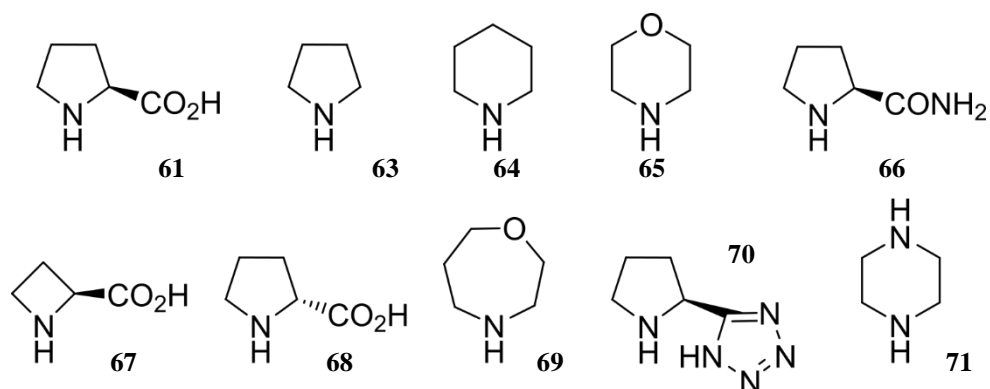
Figure 4.1 Panel of commercially available secondary amines used as potential catalysts for the OPAL.

4.1.1 Practical considerations of catalysts

During acquisition of the kinetic data for the OPAL, it was noted that stock solutions of some of the catalysts prepared deviated significantly from pH 7.5, giving solutions of up to pH 11. For this particular screen, such catalysts included pyrrolidine **63**, piperidine **64**, morpholine **65**, L-prolinamide **66**, and piperazine **71**. Therefore, the pH of stock solutions of all catalysts were confirmed by pH probe or pH test paper, and adjusted to pH 7.5 using concentrated HCl if required.

4.1.2 Initial screening of different organocatalysts

For many protein bioconjugations, performing kinetic studies has proved to be a reliable way of assessing the effectiveness of using alternative catalysts and/or alternative reactive partners for a given ligation strategy. As established in **Chapter 1**, this has typically been done using two small molecules/peptides that contain the reactive groups used in the ligation strategy on proteins, along with any required catalyst for the reaction at a set concentration (catalyst loading). Second order rate constants could then be obtained for the ligation, with the reactions assumed to be bimolecular processes between the two reactive partners. With previously established kinetic studies for protein bioconjugation strategies in mind, the panel of 10 commercially available secondary amines were screened for their ability to catalyse the OPAL between glyoxyl-LYRAG **27** (acceptor mimic) and butanal **51** (donor mimic), and second order rate constants were determined for each catalyst at 1 mM, 10 mM, and 25 mM loadings. Kinetic data was obtained in a manner consistent with previous literature [42,43]; this method obtains kinetic data using an LC-MS-based approach, and was deemed to be ideal for the obtaining kinetic data for

the OPAL which, as described in Chapter 2 and 3, was primarily established using LC-MS analysis.

4.1.3 Kinetic data obtained and comparison of organocatalysts

Figure 4.2 outlines the reaction performed to obtain second order rate constants for the OPAL:
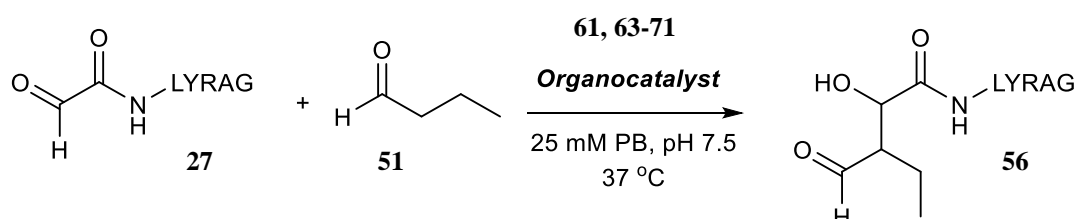
Figure 4.2 Reaction of glyoxyl-LYRAG **27** and butanal **51** using one of the 10 commercially available secondary amines **61** and **63-71** to catalyse the reaction.

Table 4.1 shows the experimentally observed second order rate constants for each catalyst at a set catalyst loading (as outlined in Figure 4.2, see Appendix for details on obtaining the kinetic data).

Table 4.1 Second order rate constants obtained for the reaction of glyoxyl-LYRAG **27** and butanal **51** using different loadings of secondary amines **61** and **63-71**

organocatalyst loading	second order rate constant $(M^{-1}s^{-1})$				
1 mM	0.0009	0.0005	<0.0001	0.0012	0.0112
10 mM	0.0033	0.0037	0.0009	0.0071	0.0670
25 mM	0.0100	0.0058	0.0016	0.0126	0.1101
	61	**63**	**64**	**65**	**66**

organocatalyst loading	second order rate constant $(M^{-1}s^{-1})$				
1 mM	0.0004	0.0007	0.0022	0.0092	0.0022
10 mM	0.0024	0.0034	0.0166	0.0551	0.0057
25 mM	0.0052	0.0067	0.0252	0.0997	0.0157
	67	**68**	**69**	**70**	**71**

From the kinetic data obtained for the 10 catalysts, several observations could be made. In all cases, having a higher catalyst loading was beneficial towards increasing the second order rate constant for the reaction between glyoxyl-LYRAG **27** and

butanal **51**; this was to be expected, as a higher concentration of catalyst will inevitably lead to a higher concentration of active enamine species present at any given time, leading to higher conversions of product on a shorter timescale (compared to using lower concentrations of catalyst). Strikingly, the second order rate constants obtained vary noticeably depending on the catalyst used. Piperidine **64** shows by far the lowest catalytic activity, requiring a 25 mM loading to achieve similar rates to that of L-proline **61** at a 1 mM loading. Indeed, the reaction between glyoxyl-LYRAG **27** and butanal **51** using a 1 mM loading of piperidine **64** led to such low conversions as judged by LC-MS that an accurate rate constant could not be determined on the timescale used for obtaining the kinetic data for the OPAL Pyrrolidine **63**, 2-azetidine carboxylic acid **67**, and D-proline **68** display similar rates to that of L-proline, whilst slightly higher rate constants were observed for morpholine **65**, homomorpholine **69**, and piperazine **71**. The highest rate constants of the 10 catalysts were observed when using L-prolinamide **66**, and proline tetrazole **70**, with a 1 mM loading of **66** or **70** displaying similar rate constants to that of L-proline at a 25 mM loading. Trends relating to specific features of the catalysts investigated, when compared to L-proline **61**, could also be identified. The presence of a carboxylic acid side chain, which is thought to be of major importance in controlling the stereochemistry of aldol products synthesised in small molecule organic synthesis,[180] was found to have little impact on the obtained second order rate constants, as evidenced when comparing the data set for catalysts **63**, **67**, and **68** to catalyst **61**. Likewise, changing the stereochemistry of the carboxylic acid side chain did not lead to any significant changes in obtained second order rate constants, as evidenced when comparing L-proline **61** to D-proline **68**. Changing the ring size to a four membered ring as seen in catalyst **67** had little effect on the second order rate constants obtained. Catalysts **65**, **69**, and **71** yielded higher rate constants compared to L-proline **61**, suggesting that the presence of a heteroatom on the ring assisted in facilitating catalysis. Significantly higher rate constants were obtained when substituting the carboxylic acid group of L-proline **61** for a primary amide or tetrazole group as seen with catalysts **66** and **70** respectively. Despite having twice the number of free amino groups, catalyst **71** did not yield substantially higher second order rate constants to that of L-proline **61** in comparison to catalysts **66** and **70**. Overall, the rate constants obtained for all catalysts in the reaction between glyoxyl-LYRAG **27** and butanal **51** were found to be comparable to rate constants

observed for previously reported uncatalysed/catalysed hydrazine/oxime ligation. In particular, the second order rate constants in Table 4.1 indicate that using L-prolinamide **66** or proline tetrazole **70** over L-proline **61** would lead to an increase in rate of OPAL, and thus a decrease in reaction times required for full conversion to OPAL products.

4.2 Using aryl donors

Although noticeable improvements to the rate of the OPAL were achieved when exchanging L-proline **61** for L-prolinamide **66** or proline tetrazole **70**, the observed second order rate constants were low by literature standards, [1,141,187] ranging within the region of 10^{-2}-10^{-1} M^{-1} s^{-1}. Therefore, to improve the rate of the OPAL further, the identity of the donor used was next investigated. Previous experiments described in Chapter 3 established phenylacetaldehyde **54** as the more reactive of the commercially available donors tested, and was therefore used as a model donor substrate for subsequent kinetic studies. Due to practical considerations (i.e. not having to correct the pH of the stock solution), and due to showing enhanced catalytic activity as seen in Table 4.1, proline tetrazole **70** was used as the catalyst, with an additional screen using L-proline **61** as the catalyst undertaken to provide comparative data on the effects of the rate of the OPAL when changing both the catalyst and donor. Second order rate constants for the reaction between glyoxyl-LYRAG **27** and phenylacetaldehyde **54** were then determined for the two catalysts at 1 mM, 10 mM, and 25 mM in an identical manner as described in section 4.1.

4.2.1 Aryl donors yield higher rate constants

Figure 4.3 outlines the reaction performed to obtain second order rate constants for the OPAL:
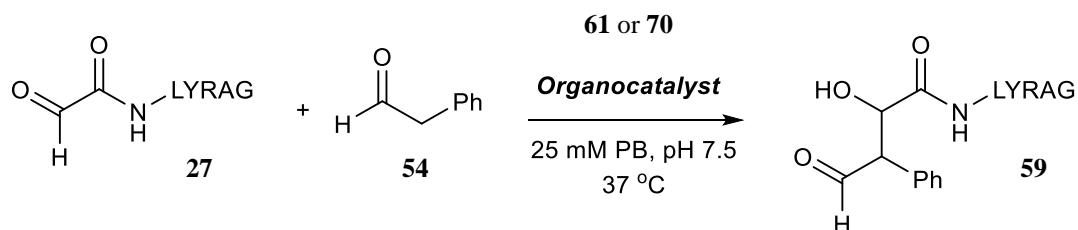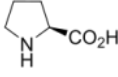
Figure 4.3 Reaction of glyoxyl-LYRAG **27** with phenylacetaldehyde **54** using L-proline **61** or proline tetrazole **70** as an organocatalyst.
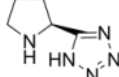
Table 4.2 shows the experimentally observed second order rate constants for the reaction outlined in Figure 4.3, using either L-proline **61** or proline tetrazole **70** as a catalyst. For comparison, the second order rate constants obtained for L-proline **61** and proline tetrazole **70** using butanal **51** (see Table 4.1) are also shown.

Table 4.2 Second order rate constants obtained for the reaction of glyoxyl-LYRAG **27** with either butanal **51** (R = CH_2CH_3) or phenylacetaldehyde **54** (R = Ph) using different loadings of L-proline **61** or proline tetrazole **70**.
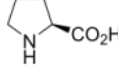
organocatalyst loading	second order rate constant ($M^{-1}s^{-1}$)			
	R = CH_2CH_3 **51**		R = Ph **54**	
1 mM	0.0009	0.0092	1.684	3.792
10 mM	0.0033	0.0551	4.366	11.820
25 mM	0.0100	0.0997	7.899	23.947
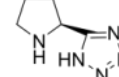
61   **70**   **61**   **70**

In comparison to butanal **51**, significantly higher second order rate constants are observed when using phenylactealdehyde **54** as a donor, with rate constants ranging between 1-24 $M^{-1}s^{-1}$ depending on the catalyst and catalyst loading. Similarly to kinetic studies with butanal **51**, proline tetrazole **70** displays greater catalytic activity compared to that of L-proline **61**. In comparison to previously reported protein aldehyde ligation kinetic studies, the second order rate constants obtained for the OPAL were found to be comparable to specific aniline-catalysed oxime ligations at pH 4.0 [123] and pH 7.0,[122] small molecule *iso*-Pictet Spengler ligations at pH 4.5,[88] small molecule SPANC ligations in H_2O/MeCN,[188] and small molecule ABAO ligations at pH 4.5 [132] (see Section 1, Table 1.1).These rate constants also compare favourably to small molecule oxime ligations performed at neutral pH,[143] and to other protein bioconjugation strategies such as the SPAAC or the Staudinger ligation.[141]

4.3. Putting the two together

Following the kinetic studies outlined in Section 4.1 and 4.2, the use of proline tetrazole **70** compared to L-proline **61** as a catalyst, and phenylacetaldehyde **54** compared to butanal **51** as a donor was investigated within the context of protein modification. For these experiments, glyoxyl-thioredoxin **27** was chosen as a model protein substrate. Glyoxyl-thioredoxin **32** was subsequently treated with four

different combinations of catalyst and donor in 25 mM PB pH 7.5 buffer at 37°C for 1 h as outlined in Figure 4.4A. Conversions to the desired OPAL products were judged by LC-MS analysis (Figure 4.4B).
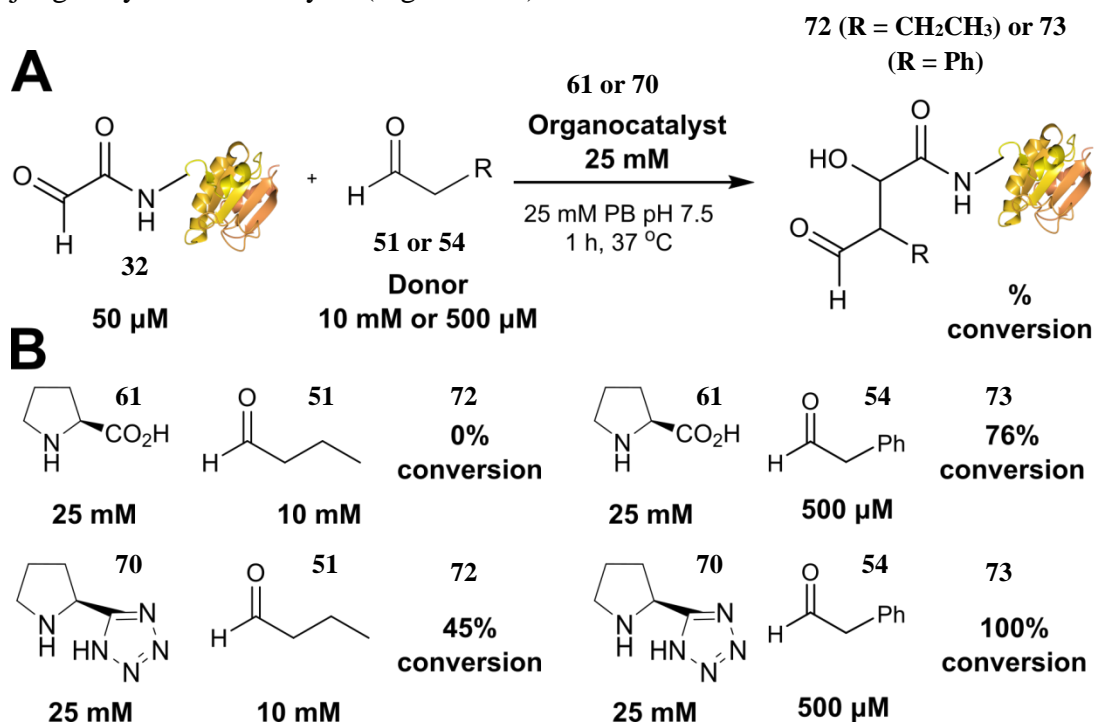
Figure 4.4 (A) Reaction 50 µM glyoxyl-thioredoxin **32** with either 10 mM butanal **51** or 500 µM phenylacetaldehyde **54** using L-proline **61** or proline tetrazole **70** as a catalyst. (B) Observed conversion to the desired OPAL product **72** or **73** for each catalyst/donor combination as judged by LC-MS analysis.

Consistent with the previously acquired kinetic data, proline tetrazole **70** displayed greatly enhanced catalysis compared to L-proline **61** in the OPAL. Similarly, phenylacetaldehyde **54** showed substantially higher reactivity compared to butanal **51**, with significantly higher conversions observed within the 1 h period when using phenylacetaldehyde **54**, despite using 200 molar equivalents less of phenylactealdehyde **54** (500 µM final concentration) compared to butanal **51** (10 mM final concentration). The difference in reactivity of catalyst and donor is especially apparent when comparing the catalyst/donor combination of L-proline **61**/butanal **51** to the proline tetrazole **70**/phenylacetaldehyde **54** combination, with no conversion to the OPAL product observed for the former, and quantitative conversion observed for the latter. Overall, proline tetrazole **70** and phenylacetaldehyde **54** exhibited significantly greater reactivity in the OPAL compared to L-proline **61** and butanal **51** respectively, leading to greater conversions of the desired OPAL product within the 1 h timeframe, along with requiring

significantly lower concentrations of donor to achieve high conversions to the OPAL product.

4.3.1 Rational for enhanced catalyst and donor activity

The enhanced reactivity displayed by proline tetrazole **70** and phenylacetaldehyde **54** in comparison to L-Proline **61** and butanal **51** could be rationalised by previously reported observations made in the small molecule aldol literature. For proline tetrazole **70**, greater compatibility with water and hydrated aldehydes has been previously reported. [189] Focusing on conversion to the aldol product, the intermolecular reaction between chloral **74** and cyclopentanone **75**, using proline tetrazole **70** as a catalyst, (Figure 4.5A) gives an 85% yield of the desired aldol product **76** when performing the reaction in the presence of 100 mol% water. In contrast, the reaction fails to proceed if no water is added, but gives aldol product in an 83% yield if chloral monohydrate is used as the acceptor. Replacing proline tetrazole **70** with L-proline **61** significantly reduced the yield of the desired aldol product **76** (<10%) regardless of whether chloral **74** or chloral monohydrate **77** was used. [189] An additional explanation for the enhanced reactivity of proline tetrazole **70** displayed over L-proline **61** is the tendency for L-proline **61** to form bicyclo-oxazolidinone species **78** with aldehyde substrates e.g. pivaldehyde **79** during a given aldol reaction (Figure 4.5B). These "parasitic" species are thought to sequester active catalyst species, preventing enamine formation and thus reducing catalyst efficiency. Proline tetrazole **70** does not participate in such reactions, and, subsequently, more catalyst is available for enamine formation and downstream aldol product synthesis, leading to a more efficient catalyst. [190]
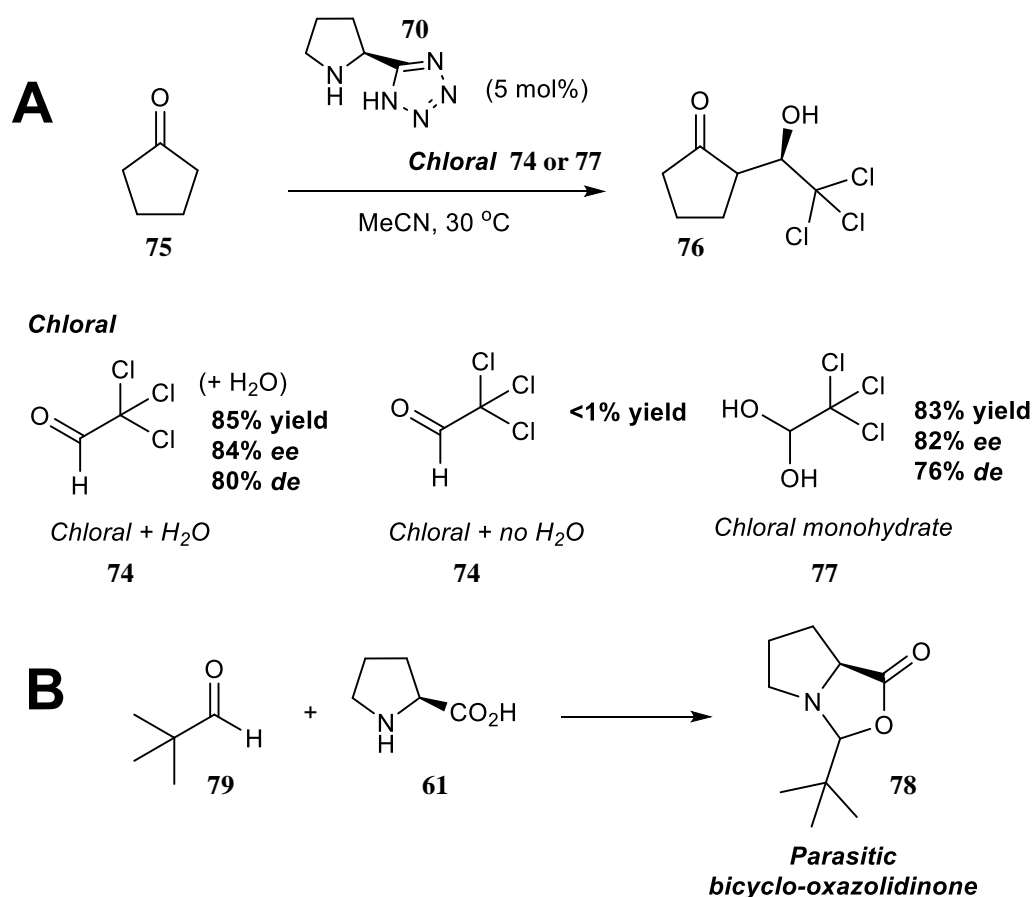
Figure 4.5 Rational for enhanced activity displayed by proline tetrazole **70** over L-proline **61** in the OPAL. (A) Proline tetrazole **70** shows enhanced reactivity in the presence of water/hydrated aldehydes compared to L-proline **61**. (B) L-proline **61** participates in parasitic bicyclo-oxazolidinone formation, reducing catalyst availability and overall reactivity in aldol ligations.

For phenylacetaldehyde **54**, a greater tendency for **54** to form enamines than other aldehyde donors has previously been described in the literature. [191] Equilibrium constants (K_{eq}) for the equilibria between carbonyl compounds and their respective enamines (using O-TBDPS-derived prolinol **80**) were determined by NMR spectroscopy, using proton peak areas of the carbonyl, catalyst and enamine species in a given system. Equilibrium constants for enamine formation between prolinol **80** and phenylacetaldehyde **54** were significantly higher than those observed when using alkyl aldehydes such as isoburtyraldehyde **52** (Figure 4.6), suggesting phenylacetaldehyde **54** shows a much greater tendency to form enamines than alkyl aldehyde donors (likely explained by carbon-carbon double bond stabilising effects provided by the phenylacetaldehyde aromatic ring substituent), and therefore shows significantly enhanced reactivity when used in the synthesis of aldol products. [191]
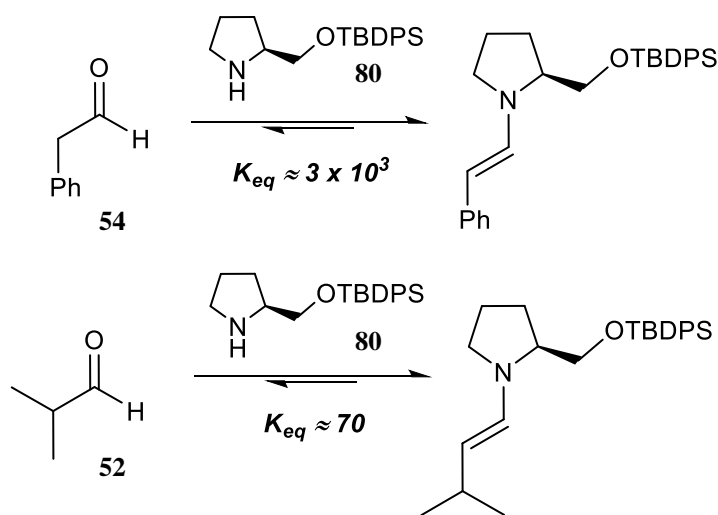
Figure 4.6 Rational for enhanced activity displayed by phenylacetaldehyde **54**. Phenylacetaldehyde **54** shows a greater tendency to form enamines than other aldehyde donor species, increasing the active concentration of reactive enamine species available for aldol ligation, enhancing the overall rate of reaction.

4.3.2 Comments on stereochemistry

The stereochemistry of small molecules is often of immense importance in governing their chemical properties, effects in biological processes, and ultimately their downstream applications. Diastereomers are known to exhibit different chemical properties to each other, such as different reactivity and different retention times in separation techniques. The enantiomers of a given compound can induce profoundly different outcomes when used in a biological context. For example, the drug thalidomide, which interconverts between enantiomers under biological conditions and was used to treat morning sickness in the 1950s, has sedative properties (resulting from the (*R*) enantiomer) but also causes birth defects (resulting from the (*S*) enantiomer).[192] Unsurprisingly, asymmetric reactions that give a high percentage of one diastereomer and/or high levels of enantiomeric excess of a given compound of interest are highly prized in organic synthesis. Organocatalysed aldol reactions in particular have garnered significant interest within the field of asymmetric synthesis and asymmetric catalysis.

Stereochemistry is a factor of great importance when considering synthetic routes towards a desired product in small molecule organic synthesis, and stereochemical control of a given synthetic strategy is often the focal point of many publications within the field of asymmetric synthesis/catalysis. For protein bioconjugation,

however, the stereochemical outcome of the products is rarely a focal point, with only a handful of publications acknowledging the stereochemistry of the products synthesised using a given strategy.[17,128,130,131,133] A possible reason for this is that the stereochemistry of the bond between the non-functional linker segment of a conjugate and between a protein (a large, dynamic biomolecule which can constantly adopt different conformations) is arguably unimportant given that it is unlikely to affect protein/probe functionality, and is therefore not discussed. Alternatively (or additionally), lack of discussion on stereochemistry in protein bioconjugation may be due to the fact that it is difficult to control, or cannot be controlled. Protein bioconjugation must be performed in aqueous media i.e. water, which can greatly interfere with the stereoselectivity certain reactions. The organocatalysed aldol reaction for example has been shown to lead to mixtures of diastereomers when performed in water. [193]

NB: The field of organocatalysed aldol reactions where water is used as an additive/solvent has seen a great deal of discussion. For more information, readers are directed to published material. [194-196]

For the OPAL, it was decided that the stereochemistry of the products was of interest, given that (unlike any previously reported protein bioconjugation strategies) the bioconjugation was being catalysed by enantiomerically pure organocatalysts, which may impact of the stereochemical outcome of the bioconjugates synthesised. Therefore, the stereochemistry of the OPAL was investigated *via* $NaIO_4$ oxidation of Ser-Ala dipeptide **81** to the glyoxyl-alanine **82**, followed by an aldol reaction with phenylactealdehyde **54** using L-proline **61** as a catalyst to synthesise aldol product (2,4,4-trihydroxy- 3-phenylbutanoyl)-L-alanine **83** (Figure 4.7A). Complete conversion to aldol product **83** was confirmed by LC-MS analysis, then characterised through NMR and HPLC studies (Figure 4.7B). The NMR and HPLC data obtained for this small molecule model indicated a mixture of four diastereomers the ratio a:b:c:d = 1: 0.91: 0.89: 0.72 (as obtained from relative area values from HPLC). This lack of stereochemical control was in agreement with other previously described aqueous aldol reactions, and it was therefore anticipated that OPAL modifications of proteins would lead to a mixture of diastereomers.
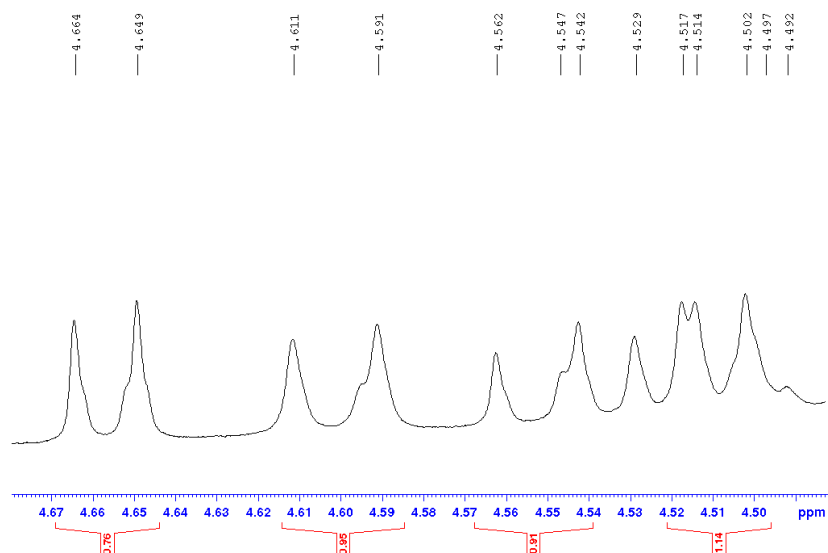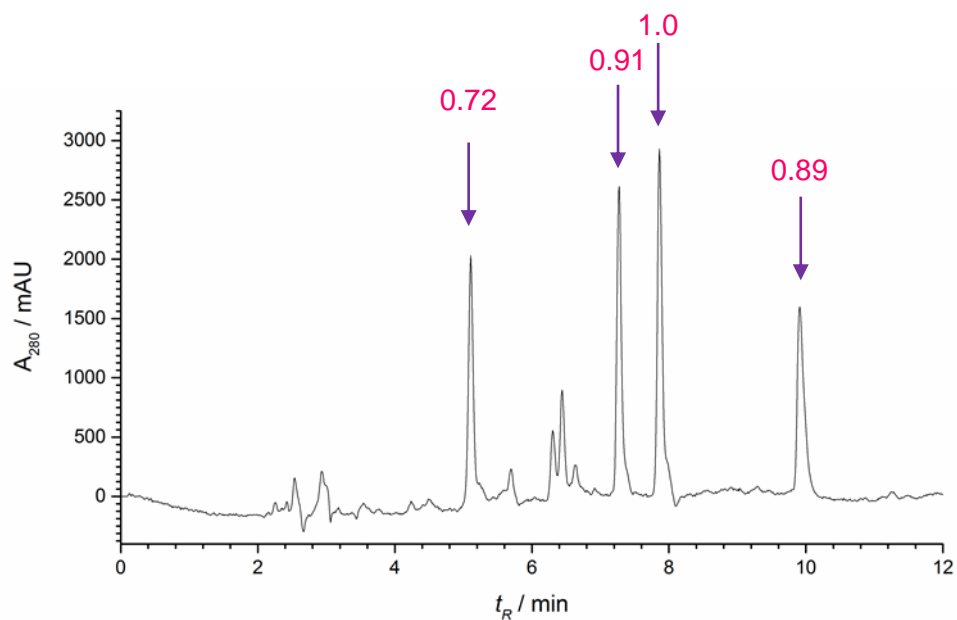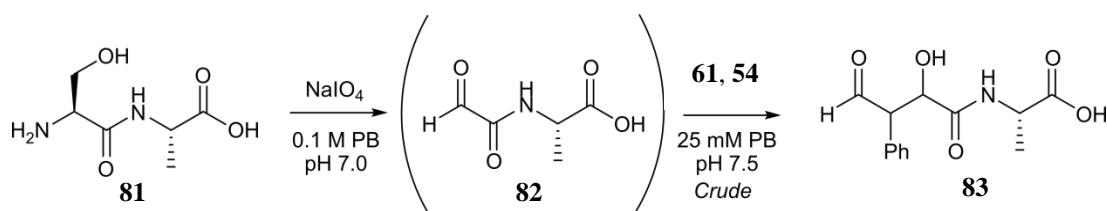
Figure 4.7 (A) Outline of synthesis of aldol product **83**. (B) HPLC data of aldol product **83**, highlighting a mixture of diastereoisomers of **83**. (C) Zoomed in portion of ^{1}H NMR of aldol product **83**, highlighting a mixture of diastereoisomers **83**.

---

v Synthesis of dipeptide **81** and subsequent NMR/HPLC characterisation of aldol product **83** was performed by Dr Darshita Budhadev (see Reference 2 for full characterisation).

4.4 Aryl donor synthesis

Having established an improved OPAL strategy using phenylacetaldehyde **54**, synthesis of probes containing α-aryl aldehyde functionality suitable for the OPAL were next investigated. Probes bearing Boc-protected oxazolidines were constructed using SPPS on resin from readily available building blocks and linker **84** (see Chapter 8, Experimental) for peptide synthesis for synthesis of linker). Figure 4.8 shows the generic structure of the Boc-protected oxazolidine, resin bound probes (where R = functionality of choice e.g. fluorescent group, affinity tag).
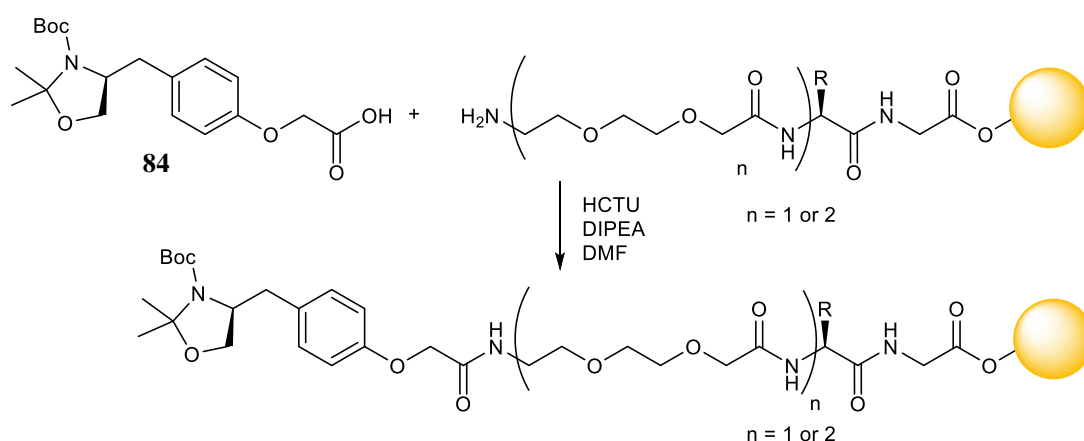
Figure 4.8 Outline of resin bound, Boc-protected oxazolidine probes for the OPAL

Simultaneous resin cleavage and unmasking of the resin bound Boc-protected oxazolidine to give 1,2 amino-alcohol protected probes could then be achieved using TFA (Figure 4.9A). The α-aryl probes constructed contain a polyethylene glycol spacer unit to both increase the water solubility of the probe, and reduce steric bulk at the reactive linker section of the probe. The 1,2 amino alcohol protected probes could then be converted to the desired α-aryl containing probes via rapid periodate oxidation (Figure 4.9B) to yield the α-aryl probes. [vi]

[vi] Linker **84** was kindly provided by Dr Darshita Budhadev (see Reference 2 for details of synthesis).

A

n = 1 or 2

TFA
> Resin cleavage
> Boc-protected
oxazolidine cleavage

Protected linker   Spacer unit   Probe

B

NaIO$_4$
PB, pH 7.0,
2 min, 4 °C
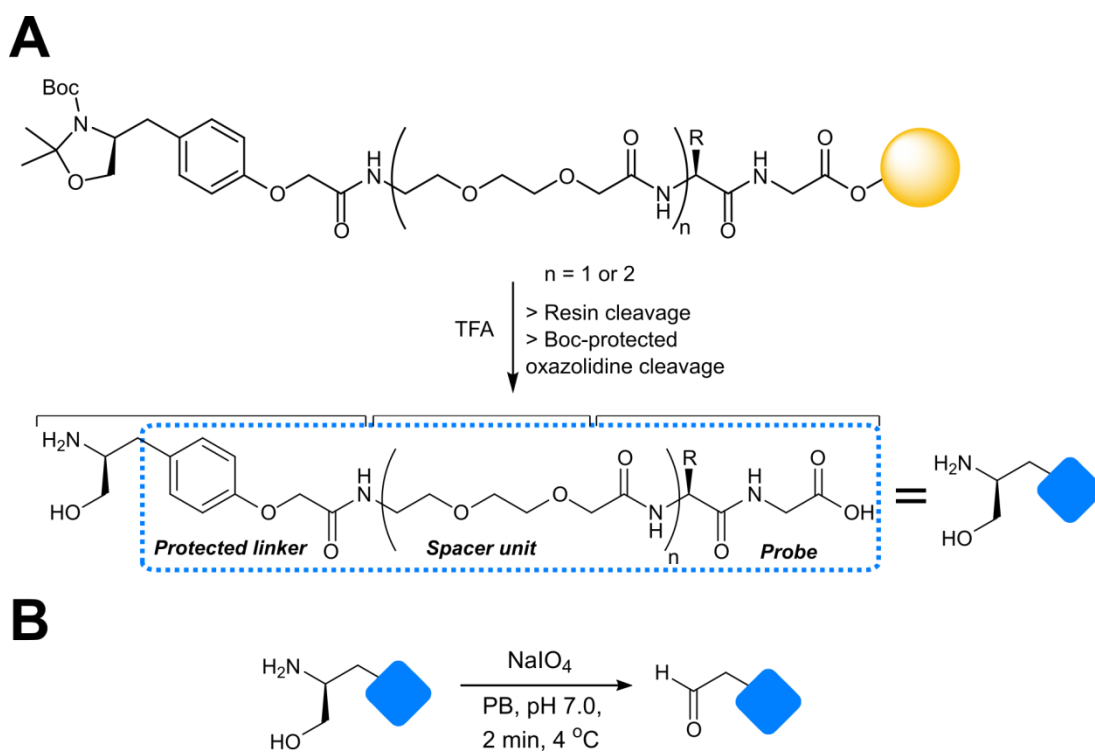
Figure 4.9 (A) Simultaneous resin cleavage and Boc-protected oxazolidine cleavage using TFA to give the amino-alcohol protected OPAL probes (R = functionality of choice). (B) Periodate oxidation of amino-alcohol protected OPAL probes to reveal the α-aryl containing OPAL probes.

Throughout this report, different chemical probes are represented as different symbols depending on the identity of R group shown, as shown in Figure 4.10 (see Chapter 8 for full details on probe preparation).

Fluorescent probe

85

Affinity tag

86

Clickable handle

87

Targeting agent

88
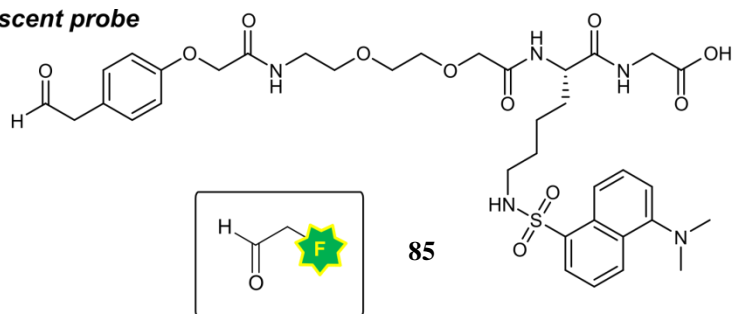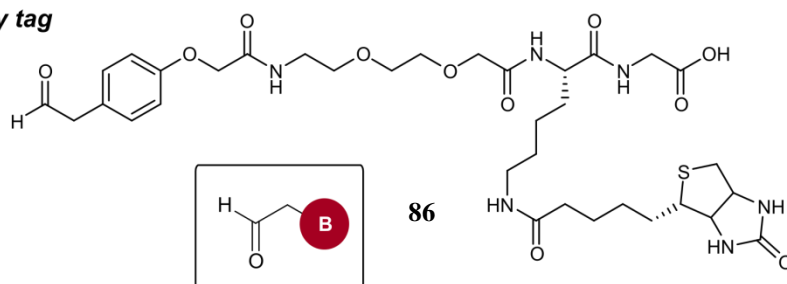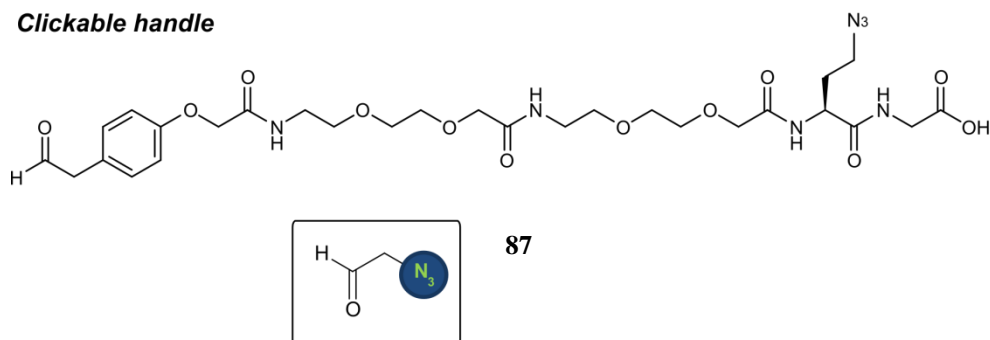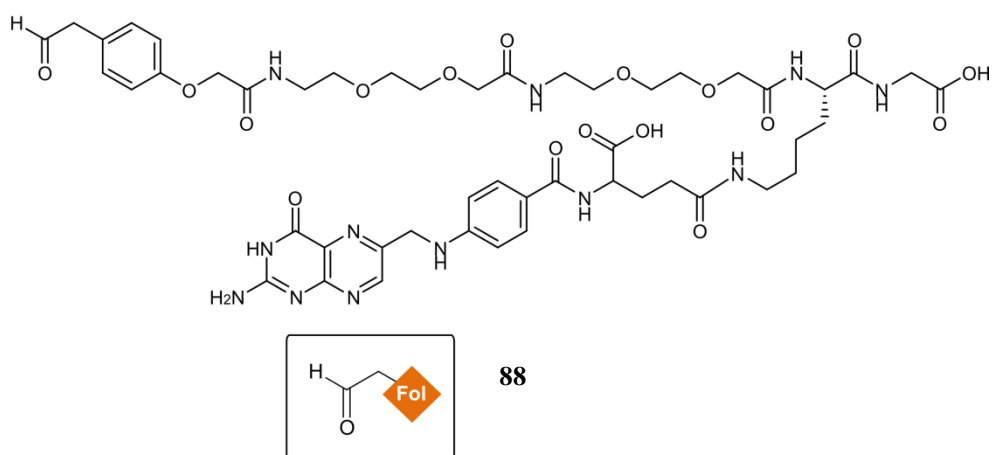
Figure 4.10 Structures of the four α-aryl OPAL probes **85-88** and their representative symbols used throughout the project.

4.5 Protein modification

With α-oxo aldehyde containing proteins, compatible chemical probes and a validated procedure in hand, the optimised OPAL strategy was next investigated within the context of protein modification. It was anticipated that conditions used in the kinetic studies for the OPAL could be used as the basis for site-selective protein modification. For an OPAL with proline tetrazole **70** and aryl probes **85-88**, this would give conditions of 50 μM α-oxo protein, 150 μM probe, and 25 mM proline tetrazole **70** in 25 mM PB pH 7.5 at 37 °C for 30 min. Protein bioconjugations, however, are well known to proceed at a more sluggish rate than their peptide counterparts, requiring higher concentrations of reagent or longer reaction times to achieve similar conjugation results. Therefore, for initial OPAL experiments between glyoxyl-thioredoxin **32** and fluorescent probe **85**, the concentration of probe used was increased from 150 μM to 500 μM, and a 4 µL reaction between 50 µM of glyoxyl-thioredoxin **32** and 500 µM fluorescent probe **85** was performed using 25 mM proline tetrazole **70** in 25 mM PB pH 7.5 for 30 min at 37 °C as outlined in Figure 4.11A. Total conversion to fluorescent thioredoxin **89** was observed as judged by LC-MS analysis, as shown in Figure 4.11B.
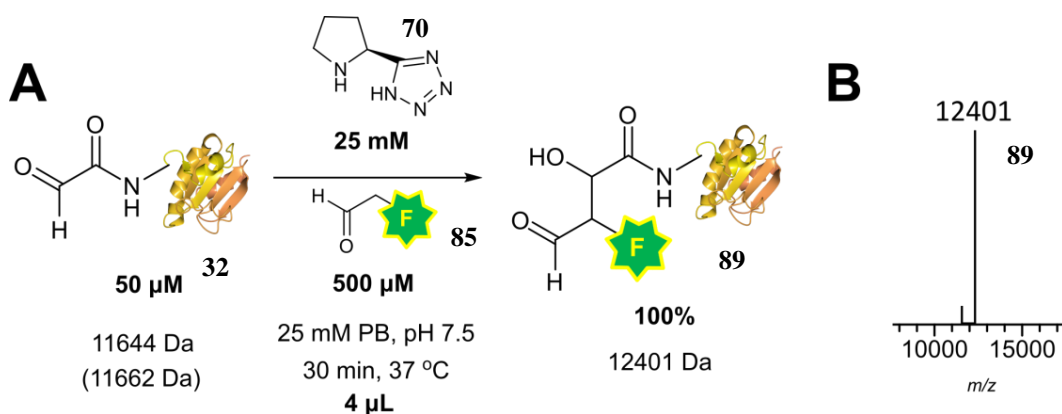
Figure 4.11 (A) Outline of reaction between glyoxyl-thioredoxin **32** and probe **85** using proline tetrazole **70** as a catalyst. Masses quoted are calculated masses, mass in brackets = **32-Hyd**. (B) LC-MS data for the reaction of 50 µM glyoxyl-thioredoxin **32** and 500 µM donor **85** using 25 mM proline tetrazole**70** in 25 mM PB pH 7.5 for 30 min at 37 °C .

Next, the OPAL was performed on glyoxyl-thioredoxin **32** on a 40 µL scale, using 500 µM of fluorescent probe **85** or 500 µM of azide probe **87**, 25 mM proline tetrazole **70** in 25 mM PB pH 7.5 for 30 min at 37 °C. This did not, however, lead to complete conversion to either fluorescent thioredoxin **89** or azide labelled

thioredoxin **90** as judged by LC-MS analysis. Allowing the reactions to proceed for a further 30 min (1 h total) resulted in complete conversion to the desired modified proteins as judged by LC-MS analysis (Figure 4.12).
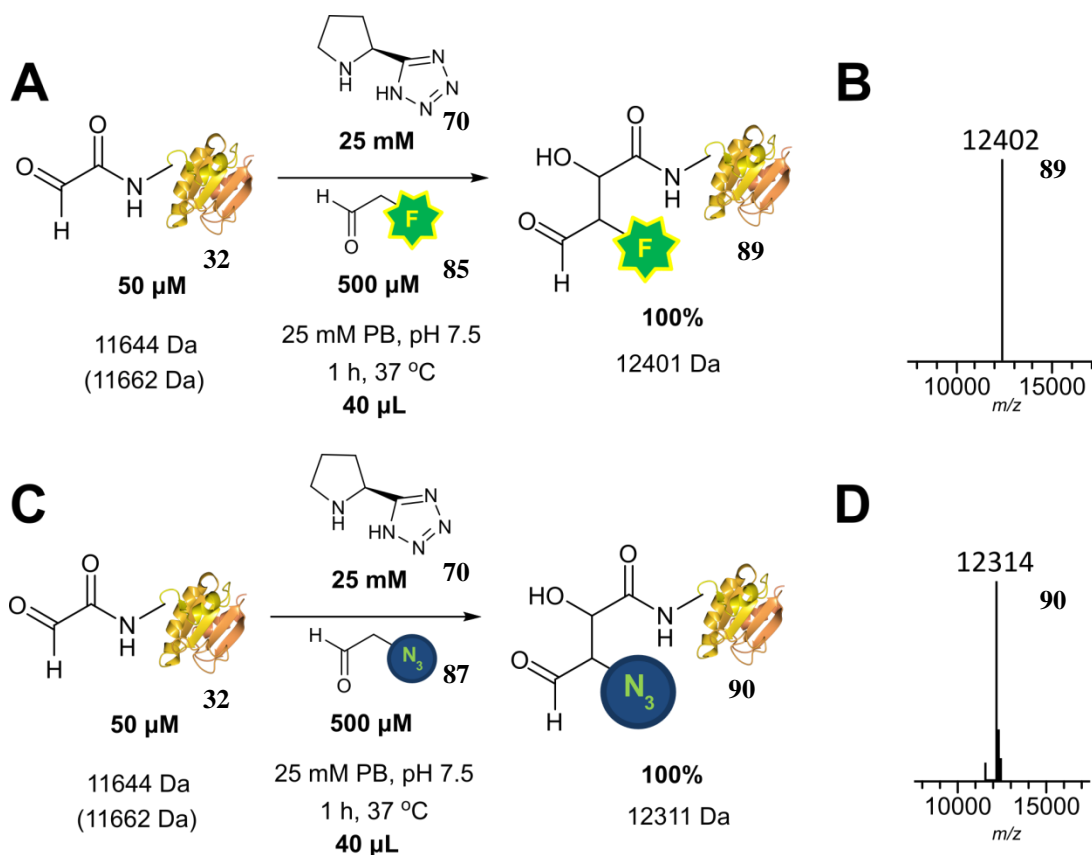
Figure 4.12 (A) Outline of reaction between glyoxyl-thioredoxin **32** and probe **85** using proline tetrazole **70** as a catalyst. Masses quoted are calculated masses, mass in brackets = **32-Hyd** (B) LC-MS data obtained for reaction 50 µM glyoxyl-thioredoxin **32** and 500 µM donor **85** using 25 mM proline tetrazole **70** in 25 mM PB pH 7.5 for 1 h at 37 °C .(C) Outline of reaction between glyoxyl-thioredoxin **32** and probe **87** using proline tetrazole **70** as a catalyst. Masses quoted are calculated masses, mass in brackets = **32-Hyd** (D) LC-MS data obtained for reaction 50 µM glyoxyl-thioredoxin **32** and 500 µM donor **87** using 25 mM proline tetrazole **70** in 25 mM PB pH 7.5 for 1 h at 37 °C .

This procedure was subsequently used for the site-selective fluorescent labelling of glyoxyl-[15N]HASPA **36-15N** and biotinylation of glyoxyl-GFP **34** (Figure 4.13).
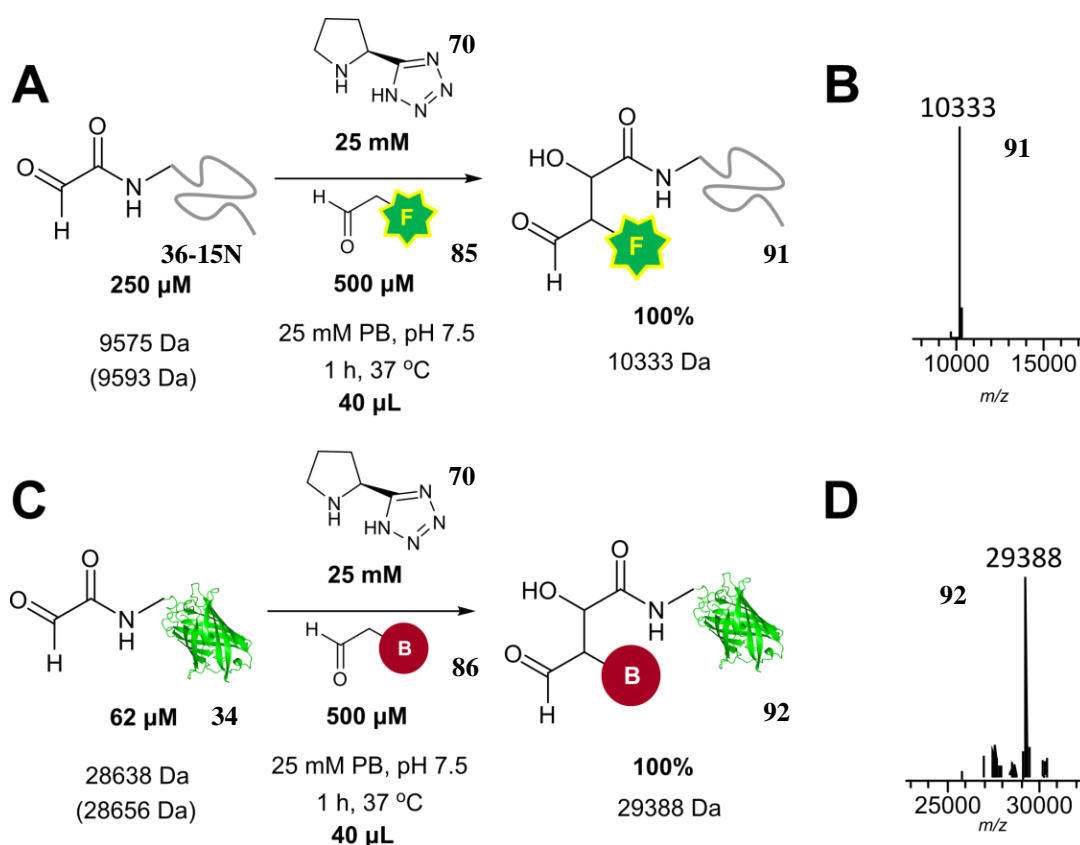
Figure 4.13 (A) Outline of reaction between glyoxyl-[15N]HASPA **36-15N** and probe **85** using proline tetrazole **70** as a catalyst. Masses quoted are calculated masses, mass in brackets = **36-15N-Hyd** (B) LC-MS data obtained for reaction 250 µM glyoxyl-[15N]HASPA **36-15N** and 500 µM donor **85** using 25 mM proline tetrazole **70** in 25 mM PB pH 7.5 for 1 h at 37 °C .(C) Outline of reaction between glyoxyl-GFP **34** and probe **86** using proline tetrazole **70** as a catalyst. Masses quoted are calculated masses, mass in brackets = **34-Hyd** (D) LC-MS data obtained for reaction 62 µM glyoxyl-GFP **34** and 500 µM donor **86** using 25 mM proline tetrazole **70** in 25 mM PB pH 7.5 for 1 h at 37 °C .

In some cases, particularly when using older probes as (see section 4.5.1), or when using sterically bulky folate probe **88**, a higher loading of probe was required to achieve full conversion to the desired OPAL product. With this knowledge in mind, site-selective modification of glyoxyl-thioredoxin **32**, glyoxyl-myoglobin **13**, glyoxyl-GFP **34**, and glyoxyl-HASPA **36** *via* the OPAL was carried out using 1 mM of fluorescent probe **85**, biotin probe **86**, or azide probe **87** (1.5 mM of folate probe **88**), with a 25 mM loading of proline tetrazole **70**, in 25 mM PB pH 7.5 at 37 °C for 1 h, with protein concentrations ranging from 50-250 µM (0.5-2.5 mg mL^{-1}). Successful synthesis of constructs **93-96** was confirmed and analysed by LC-MS (Figure 4.14, see Chapter 8, Experimental for details of modification procedures). These results in particular highlight the versatility of the OPAL, as different protein α-oxo aldehydes and α-aryl donors can be used successfully in the ligation protocol.
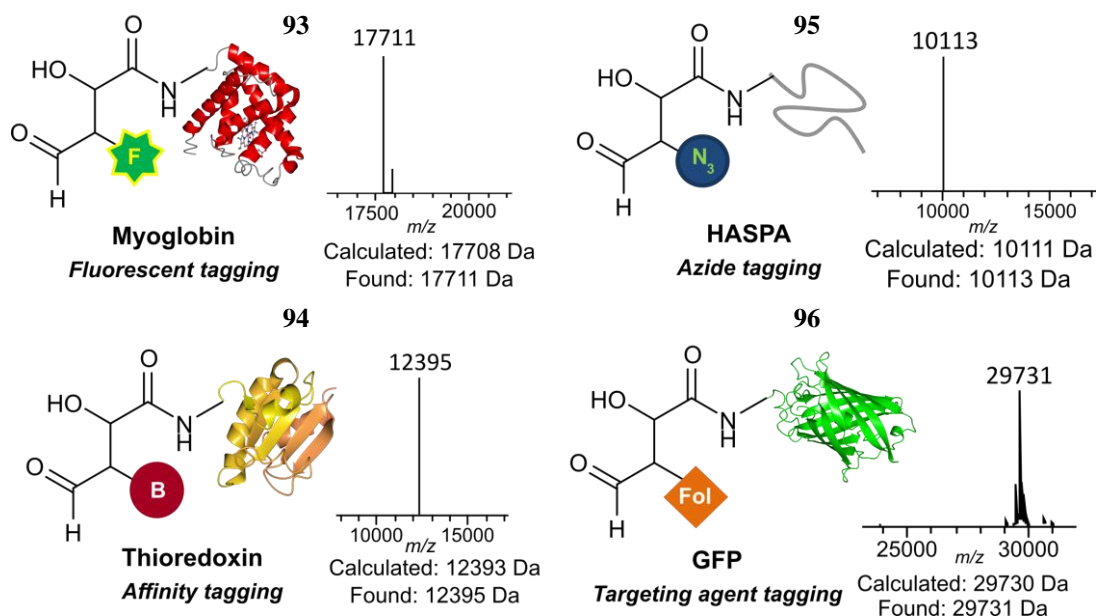
Figure 4.14 Site-selective modification of protein α-oxo aldehydes using different combinations of protein α-oxo aldehydes and α-aryl probes.

In comparison to using L-proline **61** and alkyl aldehyde donor **51**, site-selective protein modification using proline tetrazole **70** and α-aryl donors **85-88** gave faster reaction times (from 6 h to 1 h) and reduced donor concentrations (from 10-25 mM to 0.5 - 1.5 mM) required to achieve full conversion to the desired OPAL product as judged by LC-MS, offering a significantly improved OPAL protocol.

4.5.1. Notes on probe stability and long term storage

During the course of this project, it was found that 1,2 amino alcohol protected probes and α-aryl probes **85-88** could be stored long term as lyophilised powders at -20 °C. The lyophilised chemical probes **85-88** were found to be highly water soluble and could be stored as 50 mM stock solutions in water at -20 °C for at least 3 months (typically as 5 μL aliquots). Stock solutions in this form could then be defrosted and used when required. Throughout the project, defrosted stock solutions of probes **85-88** were kept at 4 °C and were typically used within four days for bioconjugation reactions. It was noted that, after 3 months of storage in solution at -20 °C, a minor decrease in reactivity of probes **85-88** towards protein modification could be observed. Probes **85-88** stored for longer than 6 months in solution at -20 °C could

still be successfully used for site-selective protein modification, but procedures could require higher concentrations of probe to achieve complete conversion to the desired protein bioconjugates after 1 h of incubation at 37 °C. Therefore, if incomplete conversion to the desired modified protein was noted for a given bioconjugation when using old probe, an additional 0.5-1 mM of probe could be added to a given bioconjugation reaction, and allowed to react for a further 30 min at 37 °C to achieve quantitative conversion.

4.5.2 Protein integrity

As discussed in section 3.1.9 the haem group of myoglobin exhibits a strong absorbance at $\lambda = 410$ nm, and could be used as an indicator of protein integrity after modifications are made to the protein. The UV/vis spectra of unmodified myoglobin **28**, glyoxyl-myoglobin **13** and biotinylated myoglobin **97** were therefore recorded, and compared at 410 nM (Figure 4.15).
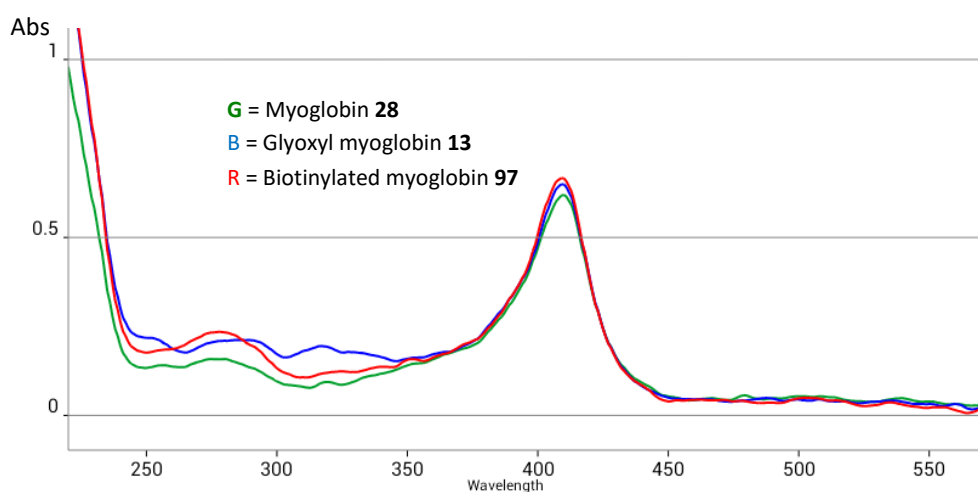
Figure 4.15 UV-Vis measurements of myoglobin **28** (green line), glyoxyl-myoglobin **13** (blue line) and aldol product **97** (red line).

The UV/vis spectrum of OPAL product **97** was found to be of an excellent match to that of unmodified myoglobin **28**, suggesting that the protein integrity, and the haem group itself, remained unchanged post modification.

4.5.3 Modification of glyoxyl-CTB by the OPAL

As described previously in Chapter 2, glyoxyl-CTB **30** is a challenging glyoxyl-protein substrate to modify due to inactivation by intramolecular cyclisation. Multiple attempts to site-selectively modify glyoxyl-CTB **30** via the OPAL were performed throughout the course of this project. Unfortunately, unsuccessful, uninterpretable data, or minimal modification to give the desired OPAL product was observed in all cases, with the most successful OPAL results for glyoxyl-CTB **30** shown in Figure 4.16A and Figure 4.16B using fluorescent probe **85**.
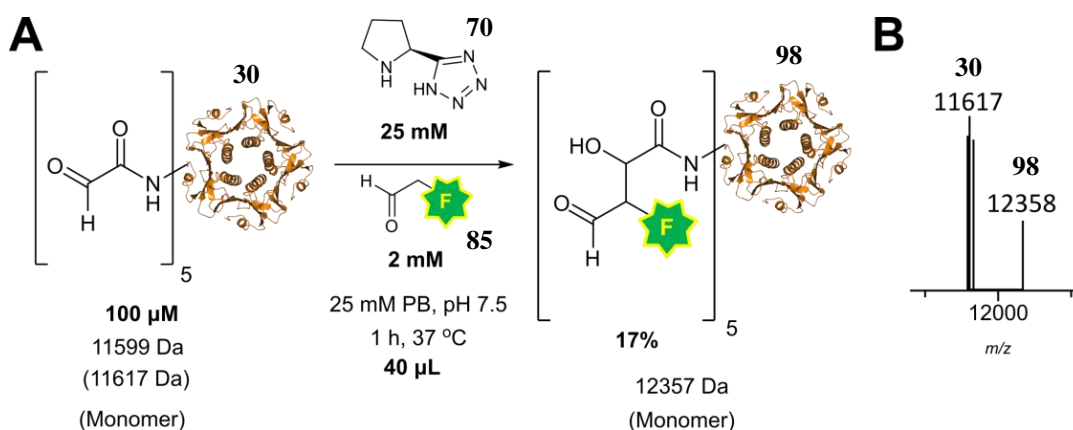
Figure 4.16 (A) Outline of reaction between glyoxyl-CTB **30** and probe **85** using proline tetrazole **70** as a catalyst. Mass quoted = calculated masses, mass in brackets = **30-Hyd** (B) LC-MS data obtained for reaction 100 µM glyoxyl-CTB **30** and 2 mM donor **85** using 25 mM proline tetrazole **70** in 25 mM PB pH 7.5 for 1 h at 37 °C

Allowing the reaction described in Figure 4.15 to proceed longer, or supplementing the reaction with more catalyst, or using high donor concentrations, did not lead to any improvements in conjugation to yield fluorescent CTB **98**. These results further demonstrate the additional challenges associated with modifying protein α-oxo aldehydes containing adjacent proline residues such as glyoxyl-CTB **30**.

4.5.4 Comments on OPAL product stability

Protein bioconjugation strategies should ideally yield conjugates that are hydrolytically stable and do not undergo uncontrolled breakdown. In an effort to assess the stability of the OPAL products, a desalted aliquot of azide labelled thioredoxin **90** was incubated in 25 mM PB pH 7.5 at 37 °C over the course of 72 h with LC-MS data for the sample collected at 24 h intervals (Figure 4.17).
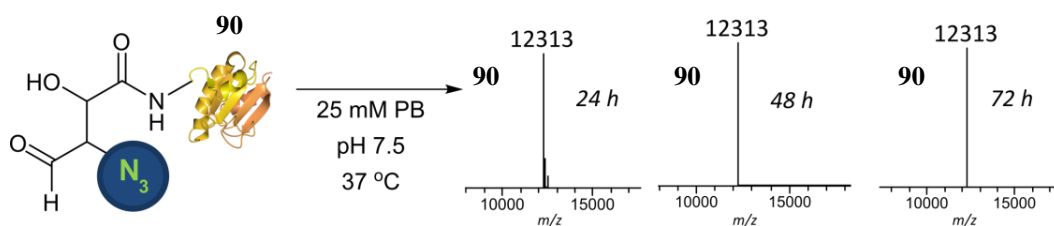
Figure 4.17 Structure of OPAL product azide tagged thioredoxin **90**, and associated MS data after periods of incubation at 37 °C to determine hydrolytic stability.

No hydrolysis of azide labelled thioredoxin **90** to give glyoxyl-thioredoxin **32** was observed as judged by LC-MS. Additionally, OPAL modified protein samples could be stored at 4 °C for several days following purification by desalting, and used at a later date for further site-selective modification experiments. On a related note, purified samples of modified peptides **56** and **59** could be stored as lyophilised powders at -20 °C for at least 1 month, with minimal hydrolysis of the aldol product observed upon solubilisation into buffer.

4.6 Modification of internal protein α-oxo aldehydes

With confirmation of the OPAL as a viable method for site-selective modification of proteins bearing *N*-terminal α-oxo aldehyde, the strategy was next investigated in its compatibility with internal α-oxo aldehyde handles (Figure 4.18A). Following Pd-mediated decaging, sfGFP(N150GlyoxylK) **41** and GFP(Y39GlyoxylK) **42** were subjected to OPAL conditions using 1.5 mM of azide probe **87**, with a 25 mM loading of proline tetrazole **70** in 25 mM PB pH 7.5, at 37 °C 1 h, with sfGFP(N150GlyoxylK) **41** at a final concentration of 100 µM, and GFP(Y39GlyoxylK) **42** at a final concentration of 150 µM, yielding internally azide labelled sfGFP **99** and internally labelled GFP **100** (Figure 4.18B). Additionally, to demonstrate compatibility with other probes, 100 µM sfGFP(N150GlyoxylK) **41** was also site-selectively modified using biotin probe **86** under identical conditions to give internally biotinylated sfGFP **101**.
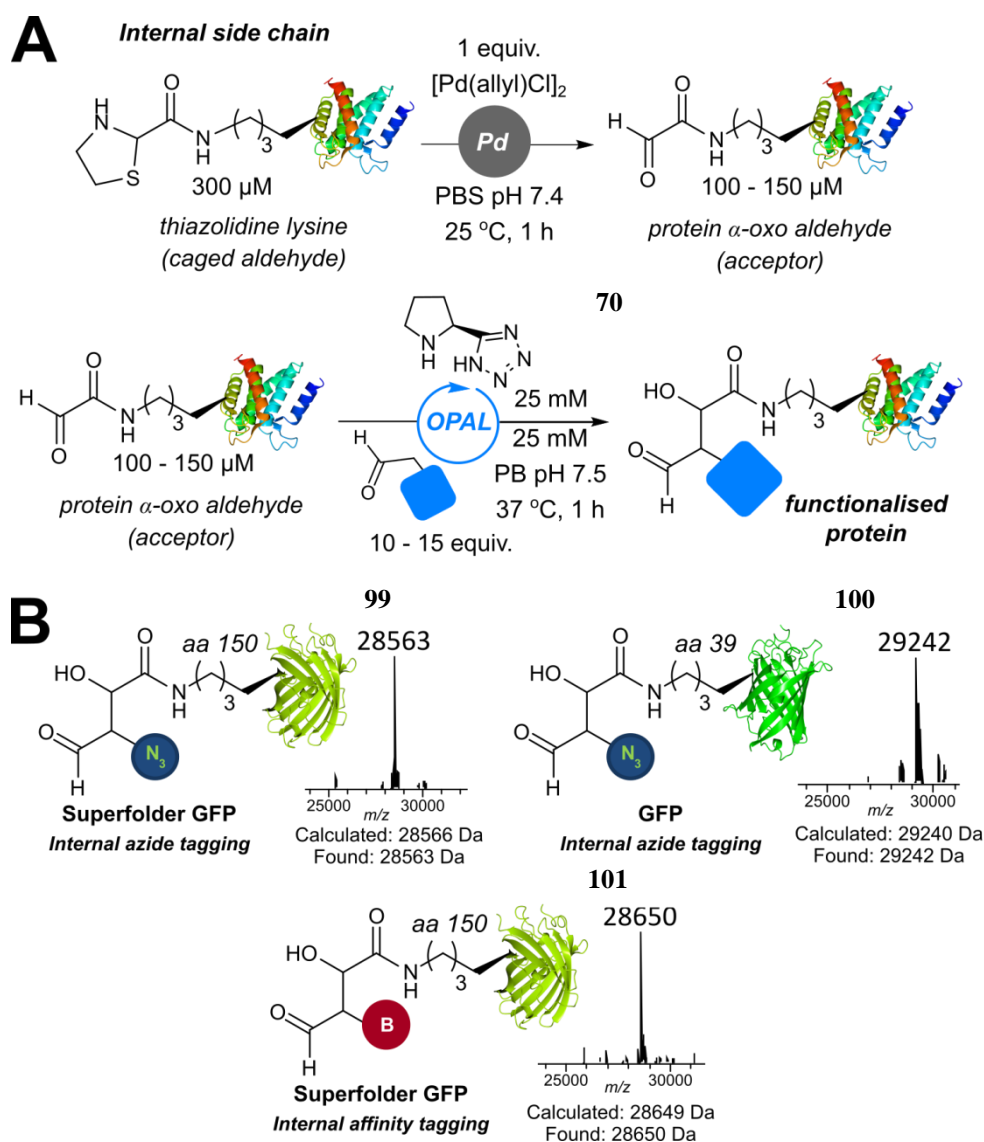
Figure 4.18 (A) Pd-mediated decaging of thiazolidine-containing proteins followed by the OPAL. (B) Site-selective modification of proteins bearing internal glyoxyl handles using azide probe **87** or biotin probe **86**.

4.6.1 Labelling of GFP in cell lysate and selective protein pulldown

Having established the OPAL as a method for the site-selective modification of purified proteins (either from commercial or recombinant sources), the ligation procedure was next investigated in its compatibility to proceed in complex biological mixtures. To demonstrate this, cell lysate containing GFP(Y39ThzK) **42** was first subjected to Pd-mediated decaging, followed by site-selective modification via the OPAL using proline tetrazole **70** and biotin probe **86**. Excess probe was removed using 10kDa MWCOs to give a "post OPAL" cell lysate containing internally biotinylated GFP **102** (Figure 4.19).
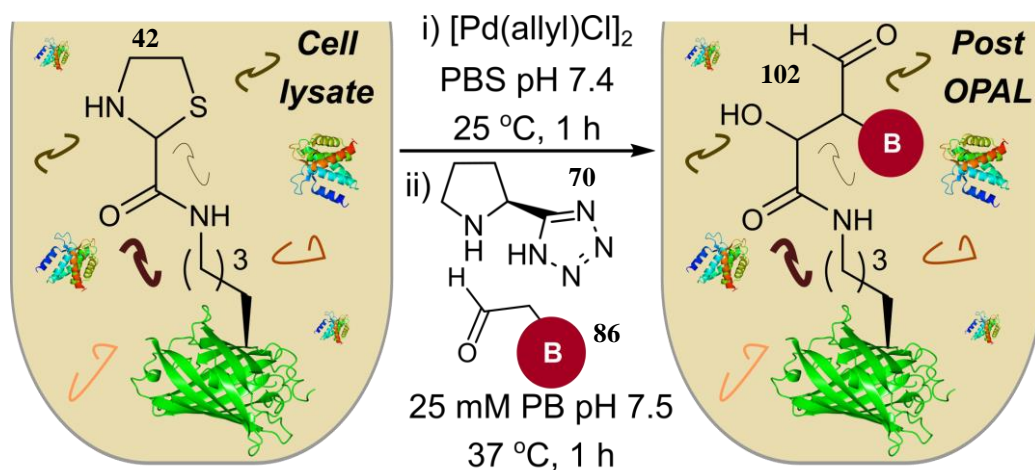
Figure 4.19 Pd-mediated decaging and OPAL biotinylation of GFP(Y39ThzK) **42** in cell lysate using 25 mM proline tetrazole **70** and 1.25 mM biotin probe **86**.

Following removal of excess probe **86**, the lysate was then loaded onto a monomeric avidin column and unbound material was washed off with PBS. Material tagged with biotin i.e. internally biotinylated GFP **102** was then eluted off the monomeric avidin agarose column using an elution buffer comprising of a 2 mM biotin solution in PBS. The monomeric avidin agarose column post lysate loading and post PBS washing, along with the fractions collected during the protein pulldown procedure, were examined for fluorescence (Figure 4.20). Given that the Pd-mediated decaging, OPAL-mediated biotinylation, and elution procedure were performed at pH 7.4-7.5, the integrity and therefore fluorescence of GFP should be retained post modification. In total, one 2 mL fraction of flowthrough was collected (Fraction FT), six fractions of 2 mL washes with 1 x PBS were collected (Fractions 1-6), and 14 fractions of 1 mL washes with 2 mM biotin in 1 x PBS pH 7.4 were collected (Fractions 7-20), as shown in Figure 4.20.
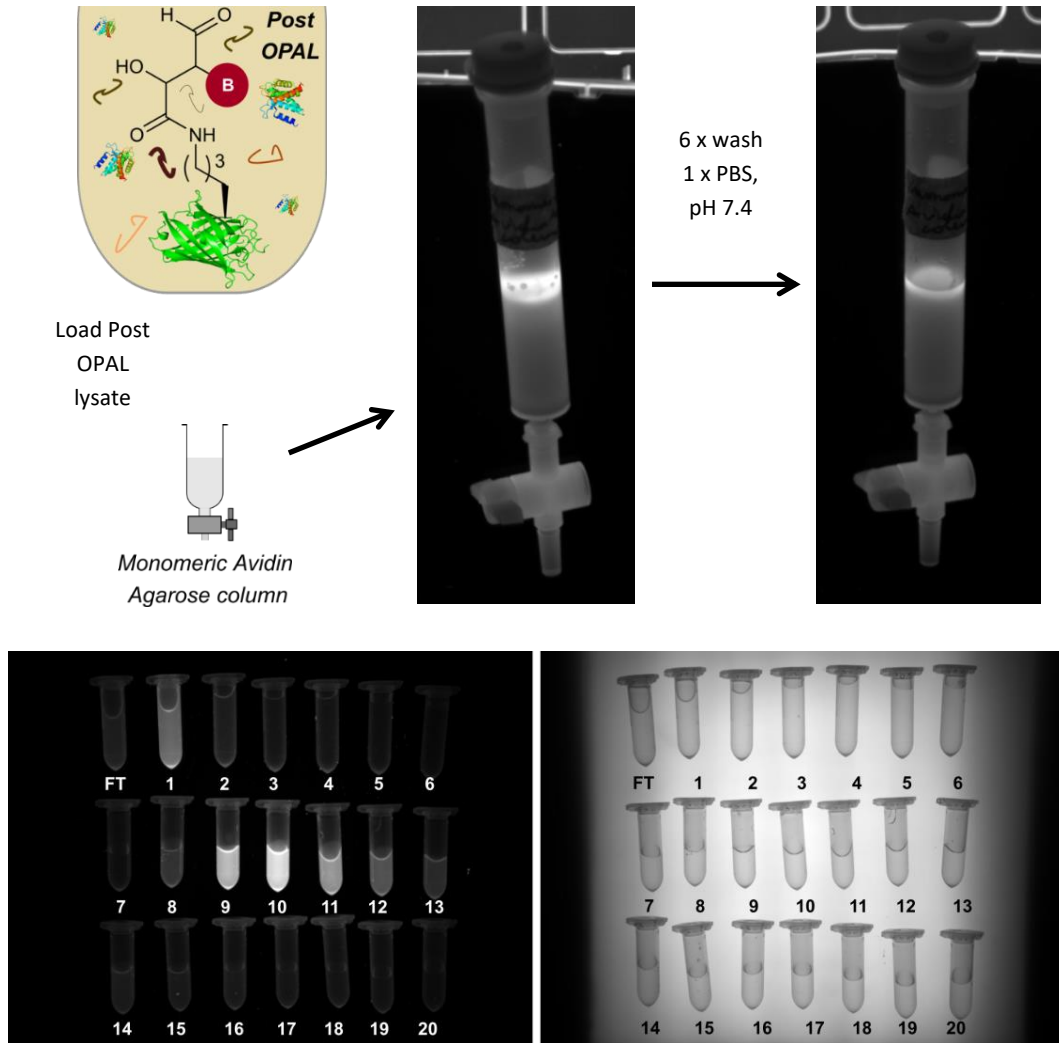
Figure 4.20 Schematic of loading cell lysate containing internally biotinylated GFP **102** onto monomeric avidin agarose column. The column is then washed with PBS. Fluorescent imaging pre and post washing demonstrates retention of GFP fluorescence. Fractions from purification of cell lysate containing internally biotinylated GFP **102** using a monomeric avidin agarose column were then collted. FT = Flowthrough. Fractions 1-6 = Washed with 1 x PBS pH 7.4. Fractions 7-20 = Elution with 2 mM biotin in 1 x PBS pH 7.4. Left: Fluorescent imaging of collected fractions. Right: White light imaging of collected fractions.

As seen in Figure 4.20, fluorescence was retained following loading of the 'post OPAL' cell lysate onto the monomeric avidin agarose column and washing the column with PBS, suggesting successful binding of internally biotinylated GFP **102**. Of the wash fractions 1-6 collected, only fraction 1 displayed fluorescence, suggesting the presence of unbound GFP material. Of the elution fractions 7-20, fractions 8-13 clearly display fluorescence, suggesting the recovery of internally biotinylated GFP. This was accompanied with the expected loss fluorescence displayed by the monomeric avidin agarose column due to elution of bound GFP. To assess the purity of eluted fractions containing internally biotinylated GFP **102** following selective protein pulldown, fractions 7-11 along with 1 and 6 (to assess

purity/GFP presence), were subjected to SDS-PAGE analysis. Samples of the original cell lysate prior to Pd-mediated decaging, and of the 'post-OPAL' lysate, were also subjected to SDS-PAGE analysis for a direct comparison of GFP purity before and after OPAL modification and avidin purification (Figure 4.20)
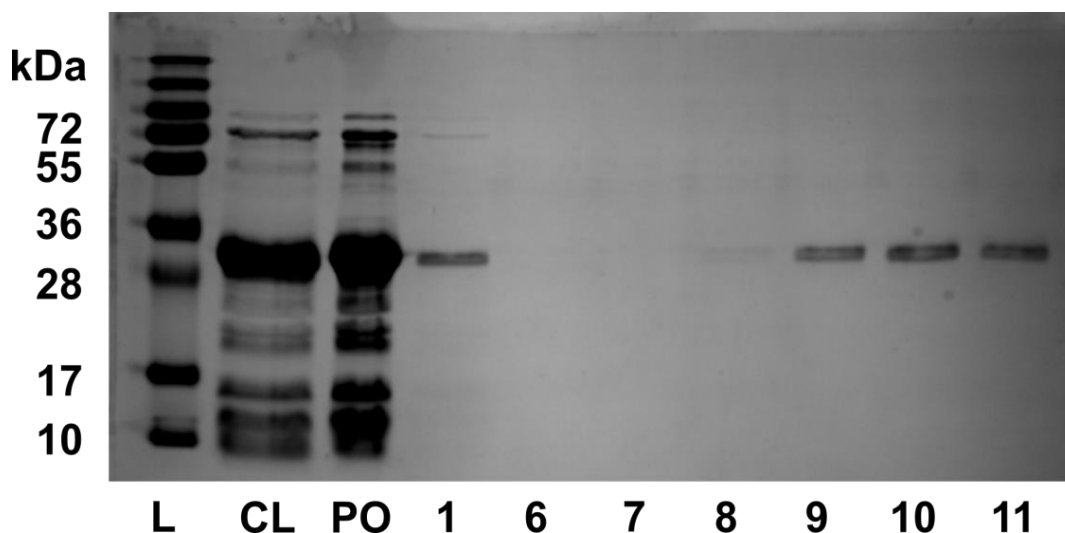
Figure 4.21 SDS-PAGE analysis of cell lysate samples and fractions of interest collected from purification of cell lysate containing internally biotinylated GFP **102** using a monomeric avidin agarose column. L = Ladder. CL = Cell lysate (before addition of allylpalladium(II) chloride dimer). PO = Post OPAL (following removal of excess affinity tag **86** and directly before loading onto the monomeric avidin agarose column). Collected fractions follow same numerical labelling as seen in Figure 4.19.

As shown in Figure 4.21, the cell lysate and 'post-OPAL' lysate contain a complex mixture of biomolecules. Fraction 1 displays GFP with impurities as evidenced by the faint bands above and below the band at ca. 30 kDa (other impurities may have been removed in the fraction FT). No material is seen in fractions 6 and 7. Fractions 8-11 clearly show purified GFP protein, with no other bands corresponding to the impurities seen in the lysate samples or wash samples present. Overall, these results demonstrate that both Pd-decaging of proteins bearing thiazolidine groups, and site-selective modification of proteins via the OPAL strategy, could be successfully carried out in complex biological media without compromising protein integrity, and that selective protein pulldown can be achieved through site-selective biotinylation and subsequent purification using a monomeric avidin agarose column.

4.7 Future work

With regards to optimising the OPAL, the focus was primarily centred on the identity of the catalyst and donor used in the ligation procedure. Through screening a series of commercially available secondary amines and enolisable aldehydes, substantial optimisation of the original OPAL procedure was achieved, with significant reductions in reaction times and donor concentrations required for total conversion to OPAL products from α-oxo aldehyde bearing starting materials. Rigorous kinetic analysis of this panel of catalysts and substrates with a model α-oxo aldehyde acceptor yielded a range of second order rate constants, the highest of which was displayed in the reaction of glyoxyl-LYRAG **27** and phenylacetaldehyde **54** using a 25 mM loading of proline tetrazole **70**. For these conditions, a second order rate constant of ~24 M $^{-1}$s^{-1} was obtained, which compares favourably to protein aldehyde bioconjugation strategies previously reported in the literature.[1] Further improvements to the OPAL procedure could potentially be found through screening of additional catalysts and donors; currently, only 10 different secondary amine catalysts and only two species of aldehyde donor have been investigated. Numerous catalysts have been developed and reported in the literature for use in small molecule aldol reactions, such as peptide-based catalysts, resin bound catalysts, and enzyme-based catalysts. Other amino acids (aside from L-proline **61**) may also show catalytic activity. Aldehyde donors bearing functional groups other than purely alkyl and aryl substituents (e.g. halogens, heteroatoms) may offer superior kinetics when used within the context of the OPAL, and may be worth investigating. There are, however, hundreds of different combinations of catalyst and donor that could be screened, and to investigate every single combination *via* the methods described in this Chapter would be exceptionally time consuming. Therefore, large scale screening of different catalysts and donors would likely require a different set of methodology, such as resin-based strategies for determining optimal sequence for PLP/RS transamination for example. [81,83]

Throughout its development, the OPAL displayed good compatibility with all protein scaffolds regardless of positioning of the α-oxo aldehyde and the donors investigated. The one exception to this was when using glyoxyl-CTB **30** as the α-oxo aldehyde acceptor. For protein scaffolds bearing hindered α-oxo aldehyde handles, such as glyoxyl-CTB **30**, it may be beneficial to investigate different lengths of the

linker component of the donors used for protein bioconjugation. An increased linker length would increase the spacing between the reactive component and bulky, functional moiety of the donor, which may help reduce steric clashes between the protein and donor during a given ligation, thus leading to improved kinetics and conversion yields. Another area that would be worthwhile investigating would be the buffer composition used in the OPAL procedure. Throughout the design and development of the OPAL, the choice of buffer has consistently been kept as a 25 mM PB buffer solution at pH 7.5. However, it has become increasingly apparent within the protein bioconjugation literature that buffer composition and the addition of salts can have significant effects on the rate of a given ligation. This is particularly notable for π-clamp mediated cysteine bioconjugation, where enormous rate enhancement can be achieved by supplementing the reaction with salts such as ammonium sulfate. [43] Therefore, investigating the effect of buffer salt composition within the context of the OPAL would potentially lead to further improvements in the kinetics of ligation.

The selective biotinylation of GFP in cell lysate demonstrated that both Pd-decaging of thiazolidine-bearing amino acid side chains and site-selective modification of proteins *via* the OPAL strategy could be successfully carried out in complex biological media. Importantly, the entire labelling procedure could be performed at neutral pH, maintaining GFP fluorescence (and therefore protein integrity), was site-selective for the protein α-oxo aldehyde over other cell lysate material, and yielded hydrolytically stable, C-C linked bioconjugate. The next stage in this venture would be to apply the OPAL in the labelling of live cells – this would truly assess the bioorthogonality of the OPAL.

To further investigate the stability of OPAL products, it would be worthwhile screening a range of different storage conditions. In particular, observing the stability of OPAL products in human plasma would be of great importance if the OPAL strategy were to be used in the synthesis of ADCs, or other conjugates designed for *in vivo* use.

Finally, site-selective protein modification *via* aldol-type chemistry could be further improved through investigation into metal-based catalysts and more sophisticated donors. Aldol reactions catalysed by metals feature prominently in small molecule

literature, and such procedures could potentially have applications in aldol-based site-selective protein modification. Designing more sophisticated donors could also lead to an improved aldol-type bioconjugation procedure. With the understanding that the OPAL follows an enamine-type mechanism of reaction, it is reasonable to postulate that one of the major mechanistic barriers in synthesis of the aldol product is the required formation of an imine, which subsequently tautomerises to the reactive enamine, between the catalyst and donor. This is because the reaction media used for the OPAL is entirely aqueous, which would result in the rapid hydrolysis of any imine formed between catalyst and donor. Circumventing this issue, either by designing donors that incorporate both catalyst and reactive aldehyde species to allow for intramolecular (as opposed to intermolecular) imine formation, or by designing sufficiently reactive donors containing "preformed enamine" functionality, could lead to an overall improved ligation procedure for site-selective modification of protein aldehydes.

4.8 Conclusion

In conclusion, the OPAL has been validated as a straightforward, organocatalyst mediated bioconjugation strategy that uses commercially available catalysts, and easily accessible probes for the rapid site-selective modification of a range of α-oxo aldehyde containing proteins at neutral pH to give hydrolytically stable bioconjugates. This ligation procedure was designed through kinetic analysis of different organocatalysts and aldehyde donors within the context of an organocatalyst mediated aldol reaction suitable for protein bioconjugation as a framework. Proteins containing *N*-terminal and internal α-oxo aldehydes, and donors bearing different probes, including fluorescent tags, affinity handles, bioorthogonal handles, and targeting agents, were found to be compatible with the OPAL, demonstrating the flexibility of the ligation procedure. The selectivity and effectiveness of the OPAL for protein bioconjugation was demonstrated in the selective labelling, pulldown, and purification of α-oxo aldehyde bearing GFP directly from a cell lysate.

Chapter 5: Site-selective dual modification of proteins

5.1 Beyond single modifications of proteins

As established in Chapter 3, the OPAL yields protein bioconjugates containing an aldehyde handle, which theoretically could participate in further site-selective modification *via* another previously report protein aldehyde ligation strategy.[1] Several strategies for dual/multi-functionalisation of proteins have previously been reported, [197] and are split into two categories: strategies that modify two amino acids in a given protein, and strategies that first modify a given protein with a probe containing an orthogonal handle, followed by modification of that incorporated handle. However, in comparison to strategies for single site-selective modification (which are becoming more prevalent in the literature), strategies for the site-selective dual modification of proteins and beyond are somewhat scarce. Development of strategies for site-selective multi-functionalisation of proteins would be of immense benefit to furthering the field of protein bioconjugation, given that certain applications require dual/multi-functional protein constructs; studying structural and dynamic properties of proteins via Förster Resonance Energy Transfer (FRET) for example requires installation of two fluorescent groups into a target protein.[63] Alternatively, the ability to site-selective modify target protein multiple times would allow for the design of more complex protein bioconjugates, such as those with fluorescent and affinity properties, or enable the conjugation of functionalised proteins to moieties such as surfaces or other proteins. Site-selective dual/multifunctionalisation of proteins would also greatly lend itself to designing next generation ADCs. The drug-antibody ratio (DAR) of ADCs is of critical importance in determining certain parameters of a given ADC. ADCs with sufficiently low DARs may show reduced therapeutic activity, whereas ADCs with sufficiently high DARs potentially suffer detrimental effects to protein structure, stability, and antigen binding.[198] For example, the ADC comprising an anti-CD30 monoclonal antibody caC10 and monomethyl auristatin E with a DAR of 4:1 required lower doses than ADC with a DAR of 2:1, and was found to have comparable *in vivo* activity to that of ADC with DAR 8:1 but twice the maximum tolerated dose in mice and a 3-fold reduction in clearance time.[199] Controlling DAR during current generation ADC synthesis is hampered by the non-specific bioconjugation strategies however, giving a heterogeneous mixture of ADCs with variable DARs. Strategies enabling site-selective dual/multi-functionalisation could

potentially offer a solution to this hurdle in designing next generation therapeutics (see Reference 14 for discussions on current and next generation strategies for preparation of ADCs). Therefore, in an effort to design methodology that could be suitable for such applications, site-selective modification of OPAL products via the β-hydroxy aldehyde was investigated.

5.1.1 The β-hydroxy aldehyde handle

As discussed throughout Chapters 3 and 4, the β-hydroxy aldehyde product of the OPAL displays no reactivity in further aldol reactions. This was in agreement with observations originally made on small molecule aldol products,[185] and was despite the molar excess (compared to the concentration of peptide/protein acceptor) of both catalyst and probe used in the OPAL procedure. However, as established in Section 3.1.1, the β-hydroxy aldehyde can undergo modification via a different protein aldehyde ligation strategy, such as oxime ligation. Therefore, the compatibility of the β-hydroxy aldehyde for site-selective protein aldehyde bioconjugation was investigated using previously reported protein aldehyde ligation strategies. For initial investigation into the suitability of the β-hydroxy aldehyde handle for further site-selective modification, the ligation between aldol-modified peptide **59**, and small molecule reactive partners **103** for oxime ligation, **104** for an *iso*-Pictet Spengler ligation, [88] and **104** [vii] for an ABAO ligation [132] was carried out. These three ligation strategies were chosen for investigation over other potential protein aldehyde ligation strategies as they had previously been established to show good ligation kinetics, and compatibility with a range of aldehyde handles (see Table 1.1 in Chapter 1). As all three ligation strategies have an optimal pH of pH 4.5, all reactions were performed in NaOAc pH 4.5 buffer as outlined in Figure 5.1A. LC-MS data was collected for each ligation to assess conversion to the dually modified peptides **106-108** (Figure 5.1B).

[vii] Indole **104** was kindly provided by Dr Darshita Budhadev (prepared according to Reference 2).
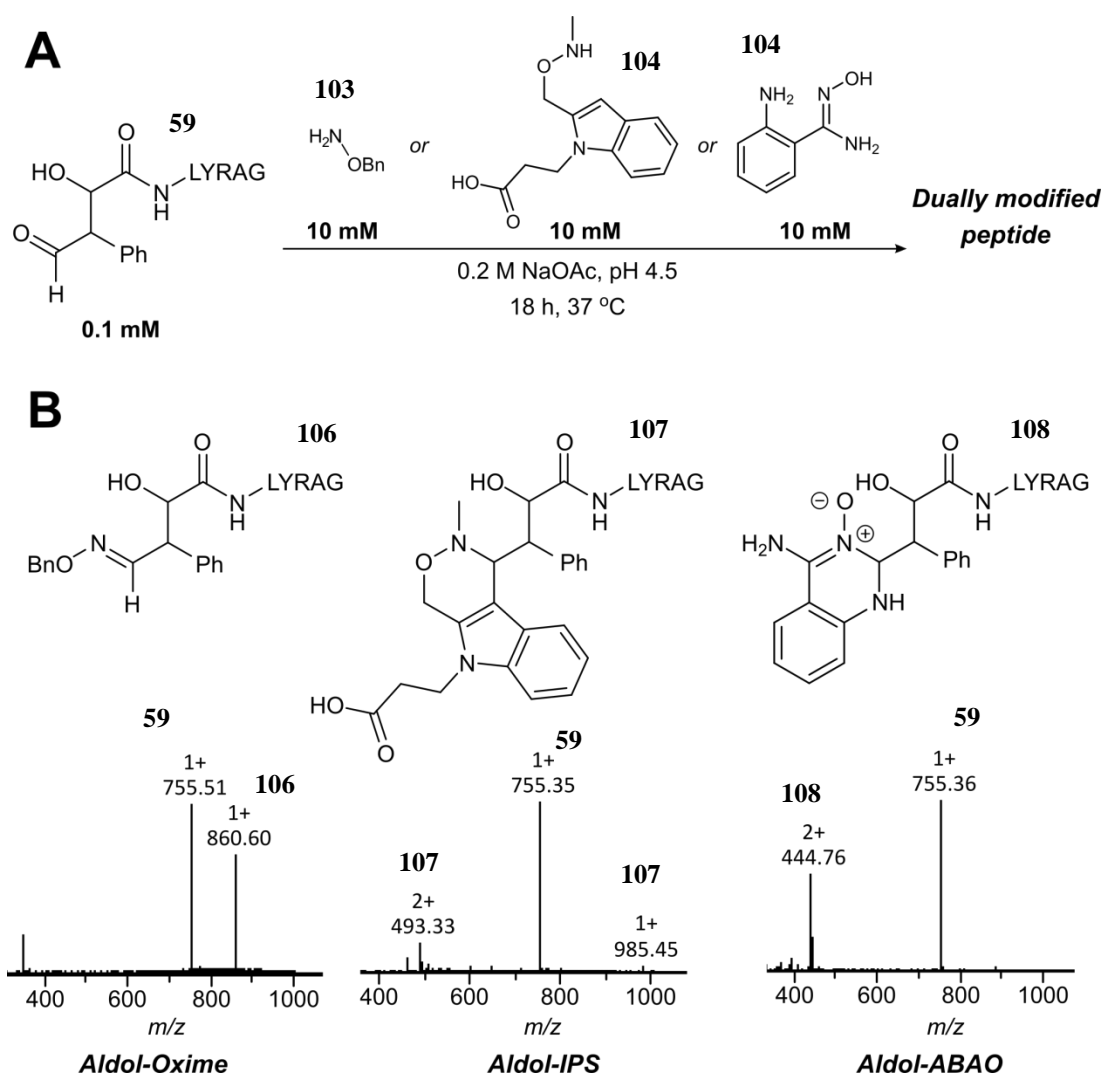
Figure 5.1 (A) Outline of reaction between aldol-modified peptide **59** and small molecule reactive partners **106**, **107**, or **108**. (B) LC-MS data obtained for each reaction of 0.1 mM aldol modified peptide **59** with small molecule **106**, **107**, or **108** in 0.2 M NaOAc, pH 4.5 for 18 h at 37 ºC.

In all cases, site-selective modification of the β-hydroxy alkyl aldehyde to the desired dually modified peptide **106**, **107**, or **108** was observed. Conversion yields were low, however, despite using long reaction times and large concentrations of reaction partner. Attention was therefore turned to improving the conjugation yield of oxime ligation by addition of aniline which, as described in Chapter 1.3.1, can act as a nucleophilic catalyst for oxime ligation *via* an imine-based mechanism. The reaction between aldol modified peptide **59**, and aminooxy **103** was therefore performed in the presence of 100 mM aniline in NaOAc at pH 4.5. A notable increase in conjugation yields was observed, resulting in a 61% conversion to **106.**

5.1.2 Reversal of pH dependence

The optimal pH of non-catalysed and catalysed oxime ligations has previously been reported as pH 4.5.[1,139] Acidic pH, however, is incompatible with certain biological applications and bioconjugation certain protein scaffolds; GFP, for example, denatures and loses its characteristic fluorescence under acidic conditions.[200] Kinetics of non-catalysed oxime ligation at neutral pH are generally sluggish, with rate constants typically in the region of 10^{-3} M^{-1} s^{-1}, but can be improved through addition of aniline, making the ligation suitable for bioconjugation at neutral pH. Therefore, an additional screen of aniline–catalysed oxime ligation of aldol-modified peptides **56** and **59** and aminooxy **103** was performed in 0.2 M PB at pH 7.5 using aniline **46** as a catalyst as outlined in Figure 5.2.
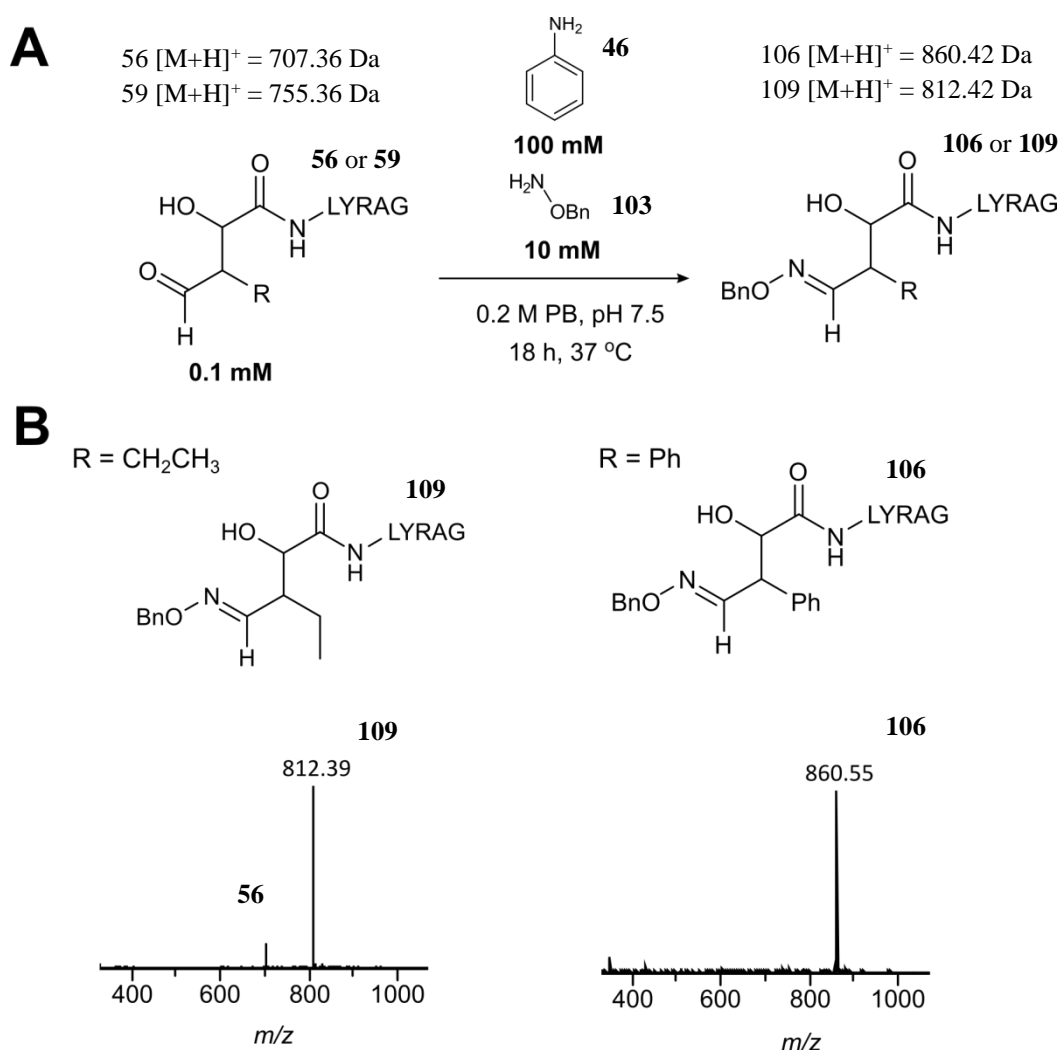
Figure 5.2 (A) Outline of reaction between aldol modified peptides **56** or **59** with aminooxy **103** using aniline **46** as a catalyst. (B) LC-MS data obtained for each reaction of 0.1 mM of aldol modified peptides **56** or **59** with 10 mM aminooxy **103** using 100 mM aniline **46** in 0.2 M PB pH 7.5 for 18 h at 37 °C.

In both cases, the conjugation yields obtained at pH 7.5 were higher than those obtained at pH 4.5, suggesting that pH 7.5 was the more optimum pH for the aniline-catalysed oxime ligation, demonstrating an unexpected reversal in pH dependence in comparison to other aniline-catalysed oxime ligations reported in the literature. It was decided this phenomenon warranted further investigation. To achieve this, previously reported aniline-based catalyst **18**, along with **110** and **111**, [119,152] were investigated for activity in the aniline-catalysed oxime ligation of aldol-modified peptides **56** and **59** and aminooxy **103** both at pH 4.5 and pH 7.5 (Figure 5.3).

pH	Conjugation yield (%)				
	No cat	46	18	110	111
4.5	29	46	56	42	Precipitates
7.0	18	88	96	95	99

109
R = CH$_2$CH$_3$

pH	Conjugation yield (%)				
	No cat	46	18	110	111
4.5	42	61	81	34	Precipitates
7.0	67	95	92	97	97

106
R = Ph
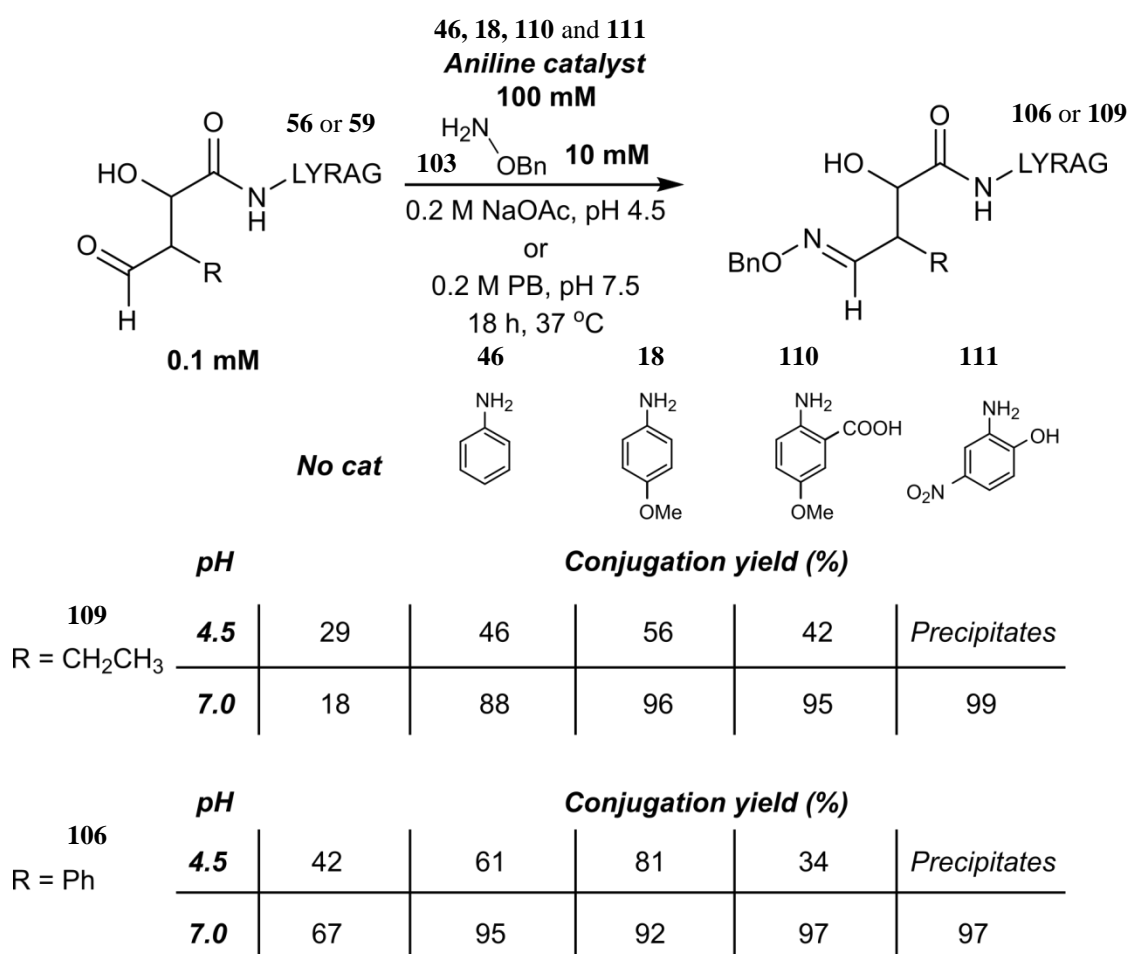
Figure 5.3 Outline of investigation into pH effect on the outcome of oxime ligation on aldol-modified peptides **56** or **59**. Conjugation yields to give dually modified peptide **106** or **109** were obtained as judged by LC-MS.

*N.B: LC-MS data was not collected for experiments using catalyst **111** at pH 4.5, as the catalyst precipitates under acidic conditions. In some experiments, small peak at 740 Da is observed in all cases – see Chapter 7 for an explanation as to the identity of this peak and why it was present.*

The trend of improved conjugation yields at pH 7.5 as opposed to pH 4.5 was again evident during the screening of the alternative aniline catalysts, with higher conversions to the desired dually modified peptides obtained at pH 7.5. To confirm this phenomenon on proteins, an aniline-catalysed oxime ligation was also performed between OPAL-modified myoglobin **60** and aminooxy **103** at pH 4.5 and pH 7.5 using aniline **46** as a catalyst (Figure 5.4).
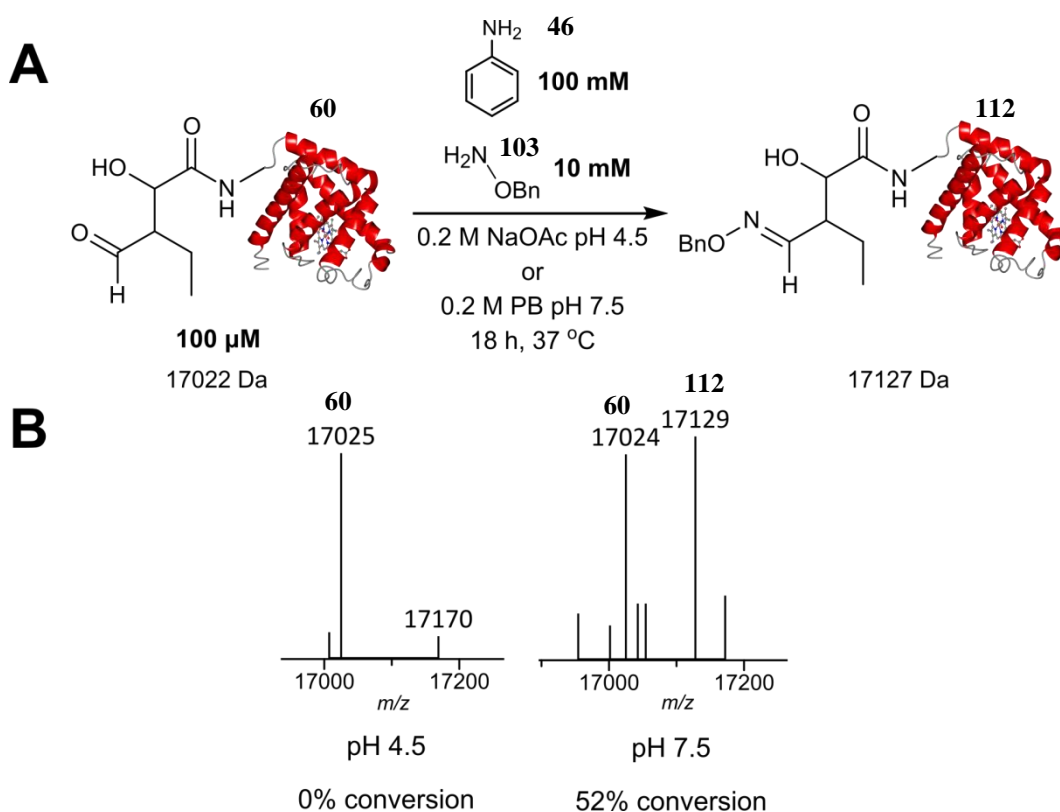
Figure 5.4 (A) Outline of reaction between OPAL-modified myoglobin **60** and aminooxy **103** using aniline **46** as a catalyst at pH 4.5 or pH 7.5. Masses quoted = calculated masses. (B) LC-MS data obtained for reaction at pH 4.5 (left) or pH 7.5 (right) for 18 h at 37 °C between 100 µM aldol myoglobin **60** and 10 mM aminooxy **103** using 100 mM aniline **46**.

For the ligation of OPAL-modified myoglobin **60**, no dually modified protein product **112** was observed at pH 4.5, whereas a 52% conversion to the dually modified protein product **112** was observed at pH 7.5, further confirming the reversal in pH dependence of aniline-catalysed oxime ligation in this particular system.

5.1.3 Hypothesis of pH dependence

The pH dependence of aniline-catalysed oxime ligation for the β-hydroxy aldehyde containing peptides/proteins, where neutral pH was seemingly favoured over acidic pH, differs significantly from the established pH dependence of aniline-catalysed

oxime ligation typically observed for other aldehyde handles. In previously reported literature, aniline-catalysed and uncatalysed oxime ligations have always yielded superior kinetics at acidic pH (pH 4.5), regardless of the efficiency a given aniline catalyst displays in oxime ligation at neutral pH (pH 7.0).[121,123] As shown in Figure 5.3 and 5.4 however, the reverse was true when using the β-hydroxy aldehyde handle. The current working hypothesis for this reversal of pH dependence relates to anticipated intramolecular H-bonding contributions from the hydroxyl functional group of the β-hydroxy aldehyde handle, and takes inspiration from previously reported aniline catalyst oxime ligations, and imine formation of *o*-hydroxyaryl aldehyde. Firstly, it has increasingly been shown that rate enhancements in aniline-catalysed oxime ligations at neutral pH can be achieved if the aniline catalyst used contains an *ortho* proton donor functional group (e.g. a carboxylic acid as seen in catalyst **113**).[152] Catalysts bearing an *ortho* proton donor are thought to alter the pKa of the aniline amino group, and facilitate the rate determining step at neutral pH (dehydration of the intermediate hemiaminal)[201] through a cyclic transition state (Figure 5.5A).[143] It is thought the most efficient catalysts for catalysed oxime ligations have a pKa value close to that of the solution pH [143]; indeed, protonated aniline **46** and aldehyde-aniline imines have pKa values of 4.6 and 2.6 respectively, which are similar values to the optimal pH of 4.5 for aniline catalysed oxime ligations.[121,202] Therefore, at higher pH values (i.e. pH 7.0) catalysts with *ortho* proton donors are theorised to be more active than catalysts without *ortho* proton donors due to the higher aniline amino pKa values. Secondly, it has previously been suggested that imine formation of glycine **114** with *o*-hydroxyaryl aldehydes,[203] such as salicylaldehyde **115** and PLP **8**, is stabilised in aqueous solution by intramolecular H-bonding between the iminium nitrogen and *ortho* phenoxide group (Figure 5.5B). Additionally, the presence of the *ortho* phenoxide groups causes a significant increase in the determined pKa of the aforementioned iminium ions. [203] Based on these observations, it is possible that the hydroxyl group of the β-hydroxy aldehyde handle acts as a pseudo *ortho* proton donor during imine formation, as outlined in Figure 5.5C, resulting in a higher-than-usual pKa value of the aldehyde-aniline iminium ion . This in turn would result in aniline-catalysed oxime ligations of the β-hydroxy aldehyde displaying higher reactivity at higher pH values than expected. Alternatively (or additionally), the presence of a pseudo *ortho* proton donor may also facilitate the rate determining dehydration step at neutral pH. Therefore, improved

conjugation yields at the β-hydroxy aldehyde would occur at neutral pH compared to acidic pH due to these factors.

Catalysts with
***ortho* proton donor**

113
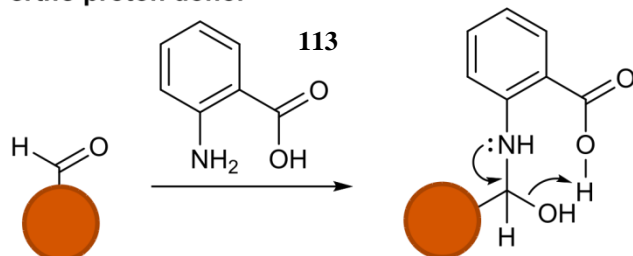
Facilitation of
dehydration
step
Reference 190

Aldehyde with
***ortho* proton donor**

114

115
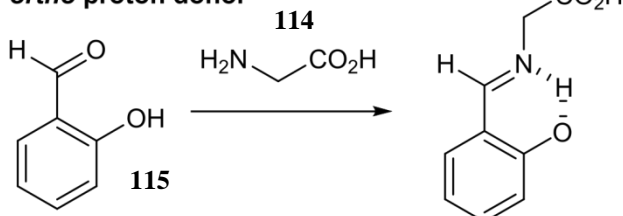
pKa elevation
Reference 192

β-hydroxy aldehyde
pseudo *ortho* proton donor

46
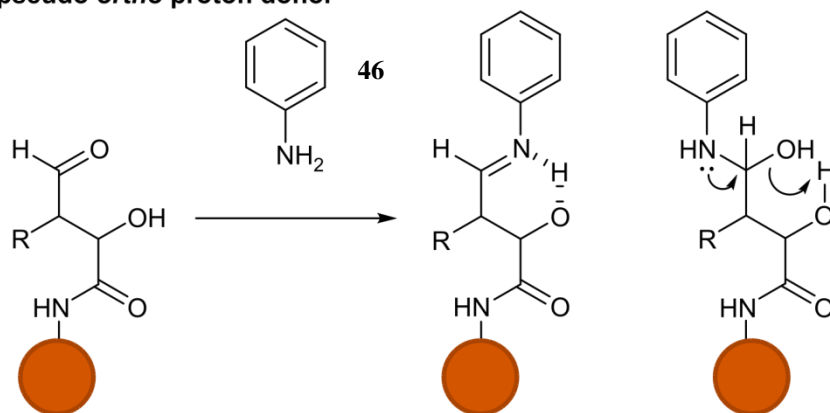
Facilitation of
dehydration
pKa elevation

Figure 5.5 A) Hypothesised effects of anilines bearing ortho proton donors on dehydration step of imine formation between an aldehyde and aniline catalyst. (B) Hypothesised pKa elevation of imine formed between an aromatic aldehyde bearing ortho proton donors and an aniline catalyst. (C) Hypothesised effects of contributions from the *ortho* hydroxyl of the β-hydroxy aldehyde on aniline catalysis.

5.1.4 Tandem mass spectrometry of dually modified peptide

To ensure site-selective modification was occurring at the β-hydroxy aldehyde position, MS/MS of the anticipated dually modified species was performed. Figure 5.6 shows the initial MS/MS analysis of dually modified **109**.
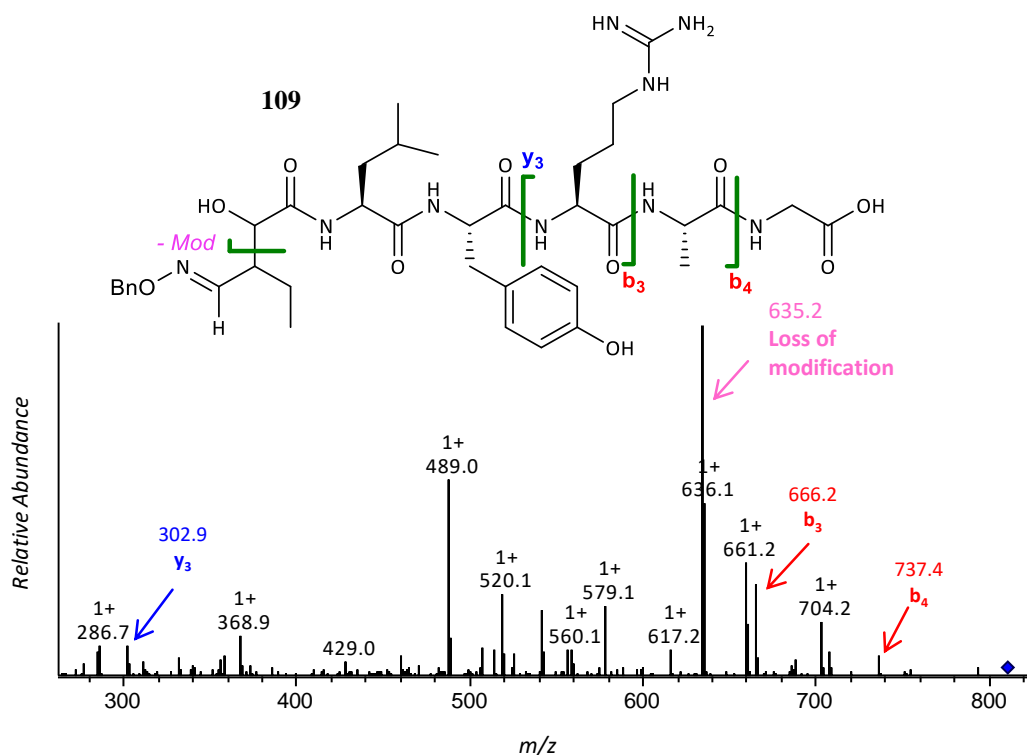
Figure 5.6 (A) MS/MS data of dually modified peptide **109.**

Although several peptide fragments corresponding to the anticipated dually modified **109** were observed, the major peak observed from the initial MS/MS was found to be a peak at 635.1 Da. For dually modified **109**, this fragment would likely occur from fragmentation of the C-C bond between the aldol-oxime segment and β-hydroxy segment of the peptide. As a result, further MS/MS analysis of the peak at 635.1 Da should give a fragmentation pattern that can be predicted based on this hypothesis. Figure 5.7 shows the MS/MS analysis of the peak at 651.3 Da.
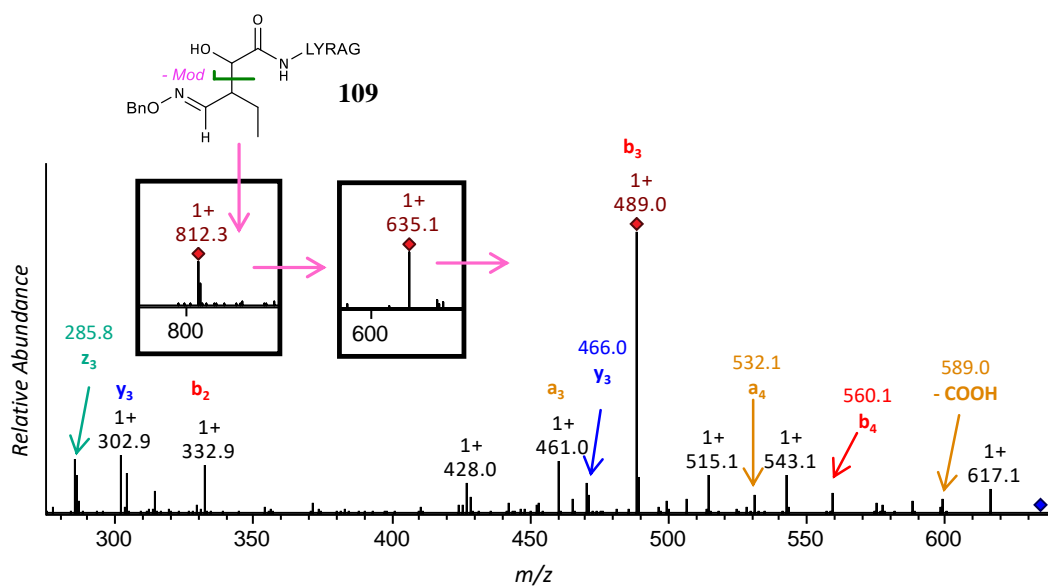
Figure 5.7 MS/MS data of fragmentation of dually modified peptide **109**, followed by fragmentation of major fragment (635.1 Da).

The fragmentation pattern obtained for the fragment corresponding to 635.1 Da were considered to be consistent with the successful synthesis of the aldol-oxime dually modified peptide **109** in a site-selective manner.

5.1.5 Hydrolytic stability of dually modified peptides

To investigate the hydrolytic stability of the OPAL-oxime products, dually modified β-hydroxy oxime peptide **109** was synthesised, purified, lyophilised, and incubated in either 0.2 M NaOAc pH 4.5 buffer or in 0.2 M PB pH 7.5 buffer at 37 °C. The stability of both the aldol and oxime bond of peptide **109** was monitored by LC-MS. Figure 5.8 shows the results of incubating peptide **109** after 1 day, 2 days, 4 days, and 30 days at either pH 4.5 or pH 7.5.
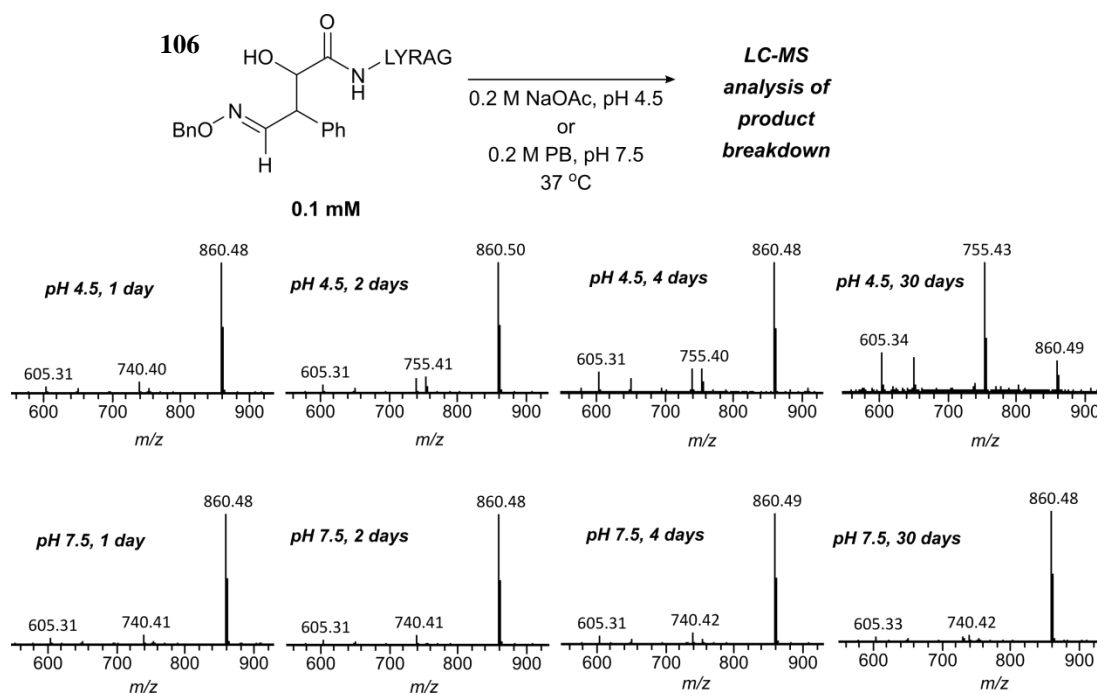
Figure 5.8 Stability testing of dually modified peptide **106**, and associated LC-MS data for periods of incubation at pH 4.5 or pH 7.5. Calculated [M+H]+ of **106** = 860.42 Da, calculated [M+H]+ of 59 = 755.36 Da.

NB: A small peak at 740 Da is observed in all cases – see Chapter 7 for an explanation as to the identity of this peak and why it was present.

At both pH 4.5 and pH 7.5, no hydrolysis of the aldol bond of the OPAL-oxime product to glyoxyl-LYRAG **27** was observed. Additionally, minimal hydrolysis of the dually modified peptide **109** to aldol-modified peptide **59** was observed at pH 7.5. However, notable hydrolysis of the dually modified peptide **109** to aldol-modified peptide **59** was observed at pH 4.5, with 3%, 10%, 15% and 80% conversion to aldol modified peptide **59** after 1 day, 2 days, 4 days, and 30 days of incubation at 37 °C. These results suggest that dually modified OPAL-oxime products show good hydrolytic stability in environments at neutral pH, but would undergo slow hydrolysis back to the OPAL product in acidic environments, whilst OPAL products *i.e.* **59** show excellent hydrolytic stability at both pH 4.5 and pH 7.5.

5.1.6 Site-selective dual protein modification *via* an OPAL-oxime strategy

Having established a pH dependence for the aniline-catalysed oxime ligation of β-hydroxy aldehyde scaffolds, the synthesis of dually, site-selective modified protein bioconjugates via a tandem OPAL-oxime procedure was investigated. Using azide

labelled myoglobin (stored in 5 mM PB pH 7.5), conditions for aniline-catalysed oxime ligation between 100 µM azide labelled myoglobin **116** and commercially available aminooxy biotin **117** in 50 mM PB, pH 7.5, were screened, with varying concentrations of aniline catalysts, varying concentrations of aminooxy biotin, and varying reaction times. The conditions, along with the observed conversions (as judged by LC-MS) to the dually modified protein species **118** for each set of conditions, are shown in Table 5.1

Table 5.1 Screening of conditions for aniline catalysed oxime ligation of azide labelled myoglobin **116** using aminooxy biotin **117**.

Entry	Catalyst	Catalyst loading (mM)	Biotin **117** (mM)	Reaction time (h)	Conversion to **118** (%)
1	-	-	15	18	0
2	**110**	10	15	18	0
3	**110**	100	15	18	*Precipitates*
4	**111**	10	15	18	0
5	**111**	100	15	18	*Precipitates*
6	**110**	10	50	18	16
7	**111**	10	50	18	0
8	**18**	10	15	18	47
9	**18**	10	15	42	65
10	**18**	10	30	18	38
11	**18**	20	15	18	50
12	**18**	20	15	42	68
13	**18**	20	30	18	38

Figure 5.9A outlines the reaction described in Table 5.1 Entry 9, and Figure 5.9B shows the LC-MS data obtained for Entry 9 of Table 5.1.
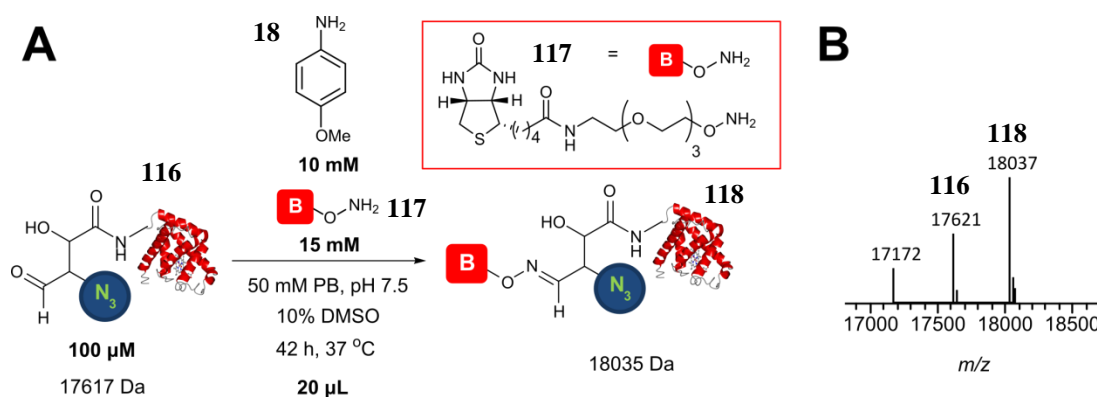
Figure 5.9 (A) Outline of reaction between azide labelled myoglobin **116** and aminooxy biotin **117** using *p*-anisidine **18** as a catalyst. Masses quoted = calculated masses. (B) LC-MS data obtained for the reaction of 100 µM azide labelled myoglobin **116** and 15 mM aminooxy biotin **117** using 10 mM p-anisidine in 50 mM PB pH 7.5, 10 % DMSO, for 42 h at 37 ºC.

NB: An unidentified peak at 17172 Da has been theorized to correspond to a protein-PLP by-product, as it has frequently been noted in samples of glyoxyl-myoglobin 13 post transamination. This peak was not taken into consideration when calculating conversions to the dually modified protein species 118.

Despite their compatibility with previously reported small molecule oxime ligations, along with oxime ligations of aldol-modified peptides **56** and **59**, catalysts **110** and **111** showed poor activity in the ligation between azide labelled myoglobin **116** and aminooxy biotin **117**. Greater catalytic activity was observed when using *p*-anisidine **18**, with a maximum of 65% and 68% conversion to the desired dually modified protein obtained when using a 10 mM and 20 mM catalyst loading respectively. Doubling the loading of catalyst **18** had minimal effect on the obtained conjugation yields whereas, surprisingly, conjugation yields dropped when increasing the concentration of aminooxy biotin **117**. A current working hypothesis for this observation is that the increased concentration of aminooxy biotin influences the pH of the reaction medium (as stock solutions of aminooxy biotin **117** are initially acidic, and must be neutralised with NaOH), resulting in a more acidic solution, and thus leading to a slower rate of oxime ligation.

The conditions in Table 5.1, Entry 12 were then used for the site-selective modification of fluorescently labelled thioredoxin **89** with aminooxy biotin **117**, using *p*-anisidine **18** as a catalyst (Figure 5.10A). A 75% conversion to fluorescent, biotinylated thioredoxin was observed after 42 h of incubation in 50 mM PB, pH 7.5, at 37 ºC as judged by LC-MS analysis (Figure 5.10B).
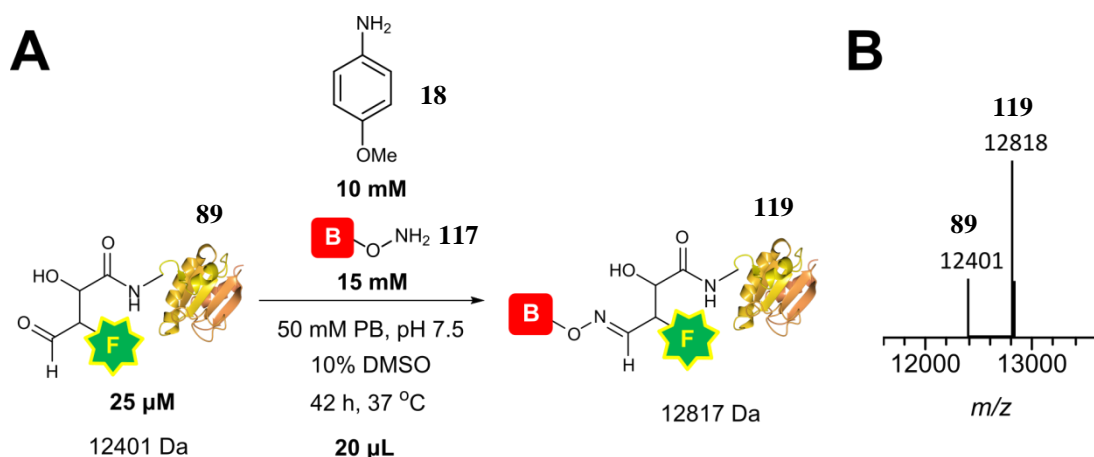
Figure 5.10 (A) Outline of reaction between fluorescent thioredoxin **89** and aminooxy biotin **117** using *p*-anisidine **18** as a catalyst. Mass quoted = calculated masses. (B) LC-MS data obtained for the reaction of 100 µM fluorescent thioredoxin **89** and 15 mM aminooxy biotin **117** using 10 mM *p*-anisidine **18** in 50 mM PB pH 7.5, 10 % DMSO, for 42 h at 37 °C.

5.2 SDS-PAGE and Western Blot analysis of dually modified protein constructs

Alongside mass spectrometry, protein bioconjugates can be further characterised using other analytical techniques depending on the groups attached during the ligation procedure as discussed in Section 1. Samples of azide labelled, biotinylated thioredoxin **120** were therefore subjected to SDS-PAGE analysis and Western Blot analysis (Antibiotin-alkaline phosphatase antibody detection) to detect for successful biotinylation, along with thioredoxin **31**, glyoxyl-thioredoxin **32**, and azide labelled thioredoxin **90** as controls, as shown in Figure 5.11
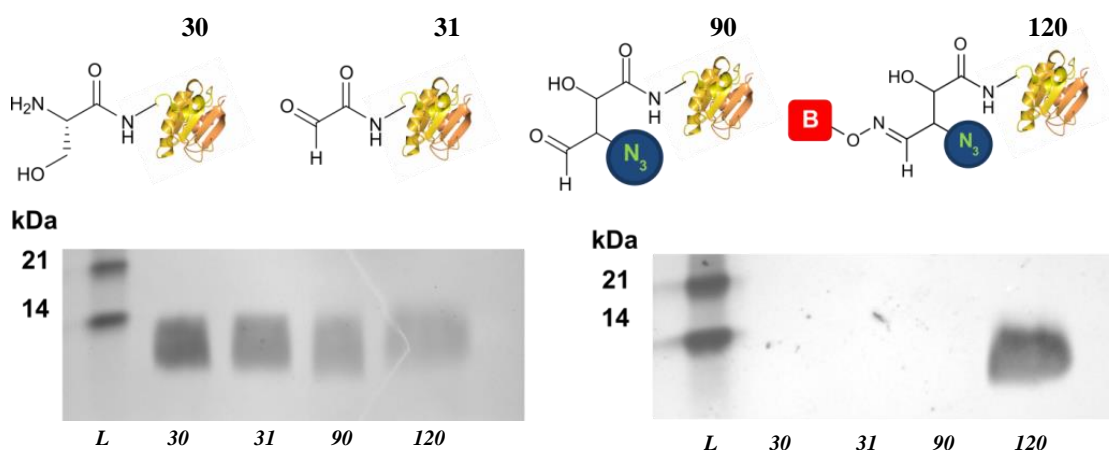
Figure 5.11 SDS-PAGE analysis (left) and Western Blot analysis (right, antibiotin alkaline phosphatase antibody detection) of thioredoxin **30**, glyoxyl-thioredoxin **31**, azide labelled thioredoxin **90**, and azide labelled, biotinylated thioredoxin **120**. L = Ladder (molecular weight marker).

Coomassie staining of thioredoxin **30**, glyoxyl-thioredoxin **31**, azide labelled thioredoxin **90**, and azide labelled, biotinylated thioredoxin **120** confirm the presence

of each protein respectively. As anticipated, only biotinylated protein **120** was detected in the Western Blot experiment

To further demonstrate the ability to dually, site-selectively modify proteins via an OPAL-oxime procedure, fluorescent, biotinylated thioredoxin **119** was also subjected to SDS-PAGE analysis and Western Blot analysis, along with fluorescent visualisation. Thioredoxin **30**, glyoxyl-thioredoxin **31**, and fluorescently labelled thioredoxin **89** were also used and analysed as control samples. (Figure 5.12).
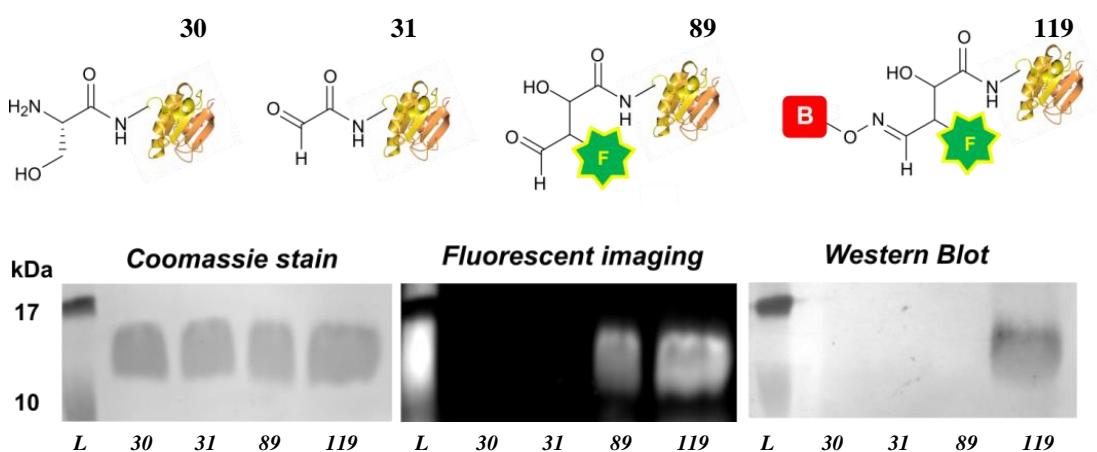
Figure 5.12 SDS-PAGE analysis (left) and Western Blot analysis (right, antibiotin alkaline phosphatase antibody detection) of thioredoxin **30**, glyoxyl-thioredoxin **31**, fluorescent thioredoxin **89**, and fluorescent, biotinylated thioredoxin **119**. L = Ladder (molecular weight marker).

Coomassie staining of thioredoxin **30**, glyoxyl-thioredoxin **31**, fluorescently labelled thioredoxin **89**, and fluorescently labelled, biotinylated thioredoxin **119** confirm the presence of each protein respectively. As anticipated, only fluorescent proteins **89** and **119** were detected in the fluorescent imaging experiment, and only biotinylated protein **119** was detected in the Western Blot experiment (detecting for biotin).

To further demonstrate the versatility of dual, site-selective protein bioconjugation via an OPAL-oxime strategy, aminooxy PEG 2kDa **121**, was synthesised as previously described [204] and used to site-selectively modify fluorescently labelled myoglobin **93**. The dually labelled protein **122** was subsequently analysed using SDS-PAGE analysis and fluorescent visualisation, along with myoglobin **28**, glyoxyl-myoglobin **13**, fluorescently labelled myoglobin **93**, and mono-PEGylated myoglobin **123** as controls. (Figure 5.13).
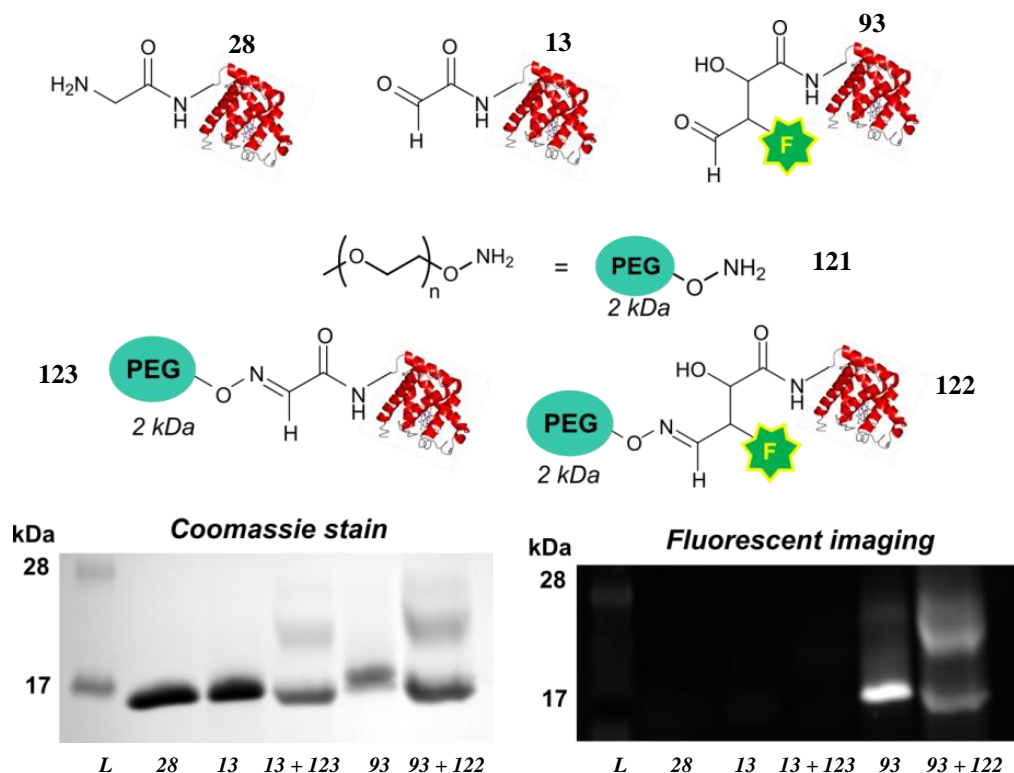
Figure 5.13 SDS-PAGE analysis (left) and fluorescent imaging (right) of myoglobin **28**, glyoxyl-myoglobin **13**, PEGylated myoglobin **123** (upper band in Lane 4, lower band = **13**) fluorescent myoglobin **93**, and fluorescent, PEGylated myoglobin **122** (upper band, lower band = **93**). L = Ladder (molecular weight marker).

Coomassie staining of myoglobin **28**, glyoxyl-myoglobin **13**, mono-PEGylated myoglobin **123** (which also contains unreacted glyoxyl-myoglobin **13**), fluorescently labelled myoglobin **93**, and fluorescently labelled, PEGylated myoglobin **122** (which also contains unreacted fluorescently labelled **93**) confirm the presence of each protein respectively. As anticipated, only samples containing fluorescently labelled proteins **93** and **122** were detected in the fluorescent imaging experiment. For protein samples treated with aminooxy PEG 2k **121** two protein bands were observed, with the upper band corresponding to a single addition of the polymer unit to the protein. The results obtained for this experiment were consistent with samples containing both unmodified protein (**13** or **93**) and PEGylated protein (**123** or **122** respectively).

It was notable, in this set of experiments, that the lower of the two protein bands in samples treated with aminooxy PEG 2k **121** runs slightly lower than expected compared to samples containing unmodified protein that have not been treated with aminooxy PEG **121**. To investigate this phenomenon, three samples of myoglobin **28**, glyoxyl myoglobin **13**, and fluorescently labelled myoglobin **93** were treated

with aminooxy PEG 2k **121** , allowed to sit at 37 °C for 3 h, and then analysed by SDS PAGE analysis as shown in Figure 5.14. These samples were also observed at a slightly lower molecular weight than expected, suggesting this phenomenon was an artefact of SDS-PAGE analysis when analysing samples containing both protein and aminooxy PEG 2kDa **121** (Figure 5.14).
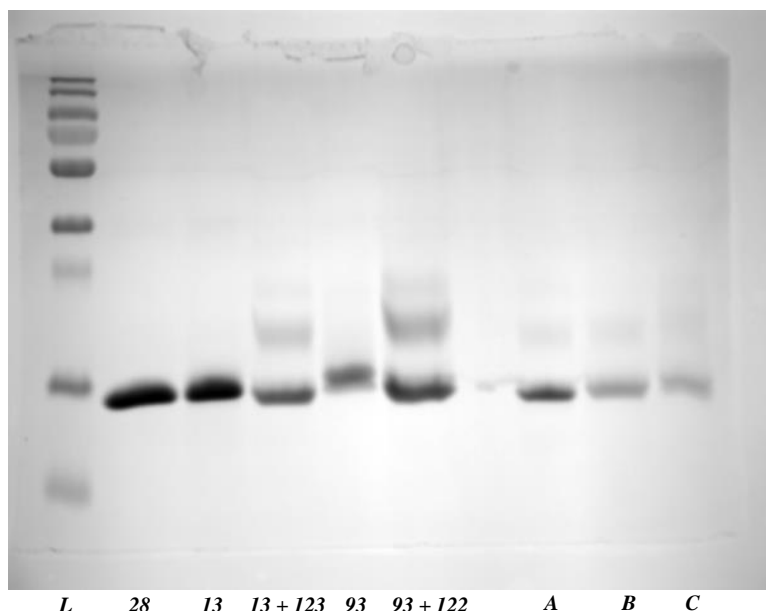
Figure 5.14 Coomassie stained SDS-PAGE analysis of myoglobin constructs, and associated myoglobin-PEG related controls. A = **28** treated with **121** for 3h. B = **13** treated with **121** for 3h. C = **93** treated with **121** for 3h.

5.3 Future work

Although notable site-selective modification of OPAL products can be achieved through aniline-catalysed oxime ligation at neutral pH, the reaction does not go to completion, requires long reaction times, and requires a large excess of aminooxy reagent. This is likely because of three reasons: 1) Oxime ligations are sluggish by nature. 2) the β-hydroxy aldehyde handle suffers notable steric hindrance from the functionality attached from previous the OPAL procedure, and 3) the β-hydroxy aldehyde is less reactive in comparison previously described aldehyde handles (such as α-oxo aldehydes). For a more efficient dual, site-selective protein bioconjugation strategy, it may be beneficial to investigate a different reaction for the second step. A good candidate for the second step of the tandem bioconjugation strategy would be the SPANC; this ligation procedure mechanistically follows the same initial pathway as aniline-catalysed oxime ligation, but offers vastly superior ligation kinetics, and is regularly performed at neutral pH. Alternatively, designing a ligation procedure

which displays significant selectively/reactivity with β-hydroxy aldehyde handles over other aldehyde handles could potentially lead to a highly efficient tandem bioconjugation strategy. Further exploration of aniline catalysts within the context of oxime ligation of β-hydroxy aldehydes may prove fruitful in optimisation of an OPAL-oxime strategy for dual, site-selective biconjugation. Finally, the observed stability of the oxime bond of the β-hydroxy oxime products warrants further investigation. This stability at pH 7.5 and slow hydrolysis at pH 4.5 of the oxime bond may be of use in applications where conjugates remain stable at neutral pH but release bound cargo at acidic pH, such as ADC design.

5.4 Conclusion

In conclusion, it has been demonstrated that dually, site-selective modified protein bioconjugates can be synthesised at neutral pH *via* a tandem OPAL-oxime procedure. This was achieved by first using the OPAL to site-selectively modify α-oxo aldehyde containing proteins, and then site-selectively modifying the resulting β-hydroxy aldehyde handle *via* aniline-catalysed oxime ligation. Remarkably, aniline-catalysed oxime ligation of β-hydroxy-aldehyde handles was found to proceed more readily at neutral pH in comparison to acidic pH, in a direct reversal of pH dependence previously reported for aniline-catalysed oxime ligation of aldehyde handles. Aminooxy reagents bearing functional moieties including therapeutically relevant PEG, and affinity handles, were found to be compatible with site-selective modification of a number of OPAL bioconjugate scaffolds. Bi-functionalisation of protein bioconjugates was confirmed by LC-MS analysis, along with characterisation by SDS-PAGE/fluorescent visualisation/Western Blot analysis where appropriate.

Chapter 6: Post translational modification mimics of myristoylated and palmitoylated proteins

6.1 Myristoylation and palmitoylation of HASPs

Following the sequencing of the human genome,[205,206] it was discovered that an estimated 20,000-25,000 protein-encoding genes are present in the human genome. However, it is thought the number of proteins present in human proteome is significantly higher. A major contributing factor to this increase in complexity from genome to proteome is thought to be the PTM of proteins, which (depending on the PTM) are responsible for roles such as protein localisation, biomolecule interactions, and activity regulation. One of the greatest challenges faced in studying the role of specific PTMs is obtaining protein samples bearing the desired PTMs in sufficient purity and in large quantities. For conventional molecular biology, the recombinant production of proteins bearing specific PTMs, such as complex glycans, typically requires eukaryotic cell strains such as yeast or mammalian (prokaryotic cell strains lack the organelles to perform these kinds of PTMs). Eukaryotic cells are much more challenging to culture than prokaryotic cells, and (depending on the PTM) can lead to heterogeneous mixtures of proteins bearing different PTMs at multiple sites, making the study of one particular PTM at a designated site on a given protein immensely challenging.

Due to significant advances in synthetic biology and protein bioconjugation over the last two decades, it is now possible to site-selectively install chemical functionalities at a particular protein site, generating homogenously modified protein constructs.[15,19] Alongside chemical probes such as drugs, fluorescent moieties and affinity tags, protein bioconjugation can also be used to design chemical 'mimics' of proteins bearing specific PTMs. This is typically achieved through recombinant expression of the desired protein in a prokaryotic cell strain system, followed by site-selective modification to install the chemical PTM mimic of interest at the site where the natural PTM would occur. This methodology offers significant advantages in the production of proteins bearing PTMs over conventional molecular biology methodology; recombinant expression is usually less time consuming and higher yielding when using prokaryotic cell strains as opposed to eukaryotic cell stains, and the heterogeneity associated with type of PTM/PTM location can be avoided through the use of site-selective bioconjugation chemistry. Chemical installation of PTM mimics through protein bioconjugation has been primarily investigated within the context of sulfur chemistry.[30] In particular, dehydroalanine has been utilised in the

synthesis of a vast array of PTM protein mimics, including lysine/arginine methylation, glycosylation, and phosphorylation.[17,18,32] Aldehyde handles have also been exploited in the synthesis of chemical mimics of *N*-glycans on antibodies, [16] and methylated lysine mimics on histones. [114]

As discussed in Chapter 5, proteins can exhibit multiple PTMs at once, and would require protein bioconjugation strategies that allow for dual/multifunctionalisation to enable their chemical synthesis. It was therefore envisioned that an OPAL-oxime strategy could be used for designing proteins bearing two PTMs. However, given the close proximity of the OPAL and oxime modification, the OPAL-oxime strategy would likely only be suitable for designing chemical PTM mimics where the two PTMs were naturally close in space. Bearing this in mind, attention was turned towards applying the OPAL-oxime strategy for mimicking dual acylation of hydrophilic acylated surface proteins (HASPs). As discussed in Section 2.1.6, HASPs such as HASPA **35** are highly immunogenic proteins present in all human infective Leishmania parasites that are expressed during the infective stages of the parasite. They are enzymatically dually lipidated at the *N*-terminal region, which is thought to facilitate their binding to plasma membranes.[175] The *N*-myristoylation of HASPA can be achieved *in vitro* to give *N*-myristoylated HASPA **124**, as the NMT required is known, allowing for generation of mono-acylated HASPA species.[207] However, the *S*-palmitoyltransferase required for the second acylation of HASPs, generating dually acylated HASPA **125**, is currently unknown, meaning that *in vitro* dual acylation of HASPA, and obtaining dually acylated HASPA **125** species for further study is not possible. Given the close proximity of these two modifications, it was hypothesised that an OPAL-oxime strategy using a suitable C14 probe **126** to generate a chemically "myristoylated" species **127**, and a suitable C16 probe **128** to generate a chemically, dually lipidated HASPA mimic **129** could be achieved, and therefore help facilitate the study of post-translationally modified HASPA (Figure 6.1).
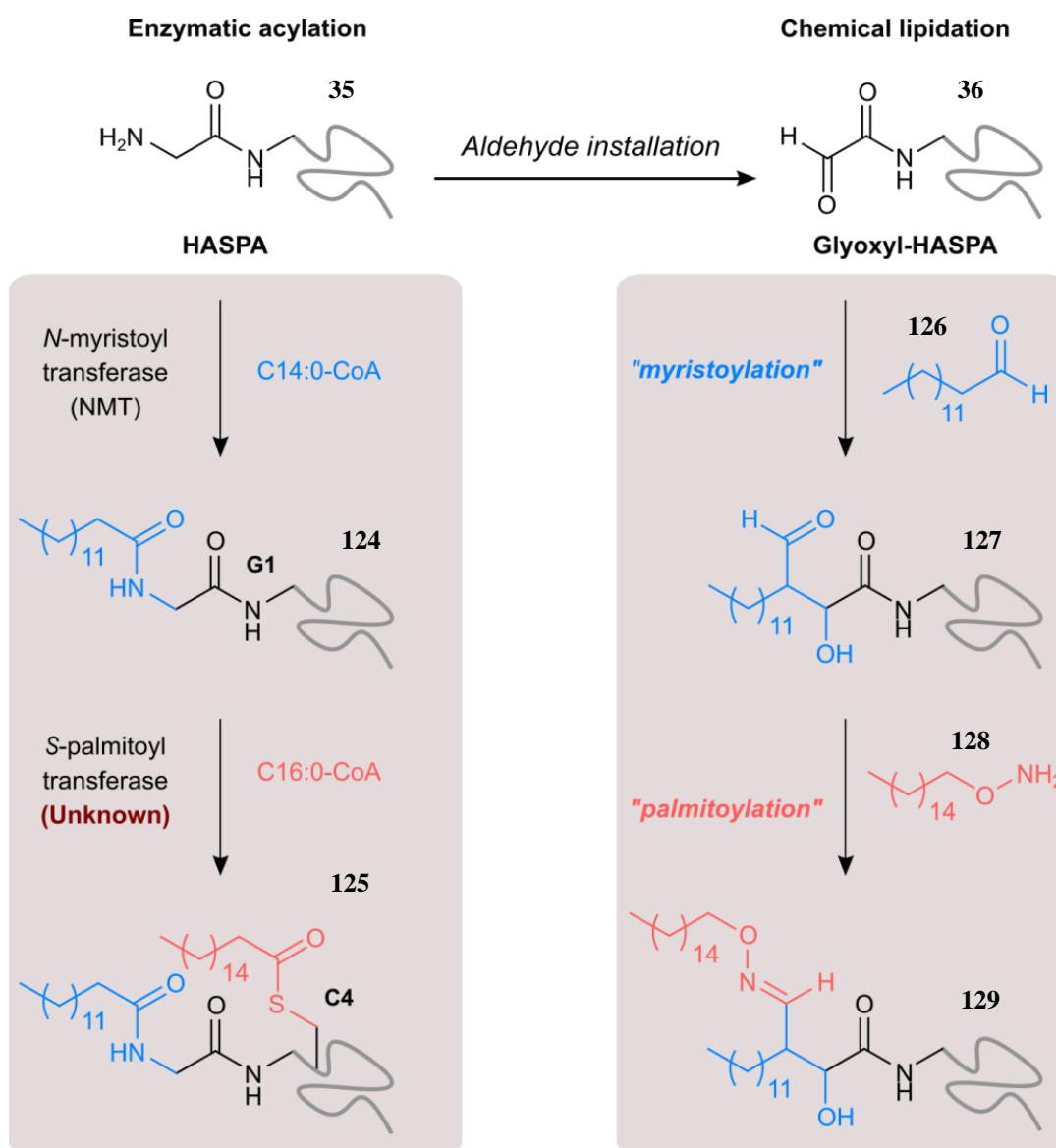
Figure 6.1 Enymatic and chemical post-translational modification of HASPA **35**. The left hand side shows the enzymatic route for the *N*-myristoylation of HASPA **35** (catalysed by a known NMT), followed by *S*-palmitoylation of *N*-myristoylated HASPA **124** (catalysed by a currently unknown *S*-palmitoyltransferase). The right hand side shows the proposed chemical route constructing lipidated mimics of HASPA **35**, starting with aldehyde installation to generate glyoxyl-HAPA **36**, followed by OPAL with donor **126** to generate the chemically myristoylated HASPA **127** mimic, followed by oxime ligation with aminooxy **128** to generate dually lipidated HASPA **129** mimic.

6.1.1 Validation of chemical myristoylation

To begin with, the OPAL modification with the C14 aldehyde myristyaldehyde **126** to generate mono lipidated HASPA mimics was investigated. Unlike previously described aldehyde donors, myristaldehyde **126** was found to very insoluble in water or buffered solutions; this was undoubtedly due to the hydrophobic nature of the C14 chain. Therefore, the solubility of myristaldehyde **126** in MeCN, t-BuOH, and DMSO, was investigated; these three solvents have previously been successfully employed as co-solvents for other protein aldehyde bioconjugation strategies reported in the literature.[128,130,133] At a final concentration of 25 mM, myristaldehyde proved to be insoluble in both a 1:1 H_2O:MeCN solvent system, and a 1:1 H_2O:t-BuOH solvent system, but was found to be soluble in a 1:1 H_2O:DMSO. DMSO was therefore used as the co-solvent of choice of subsequent chemical myristoylation experiments. To confirm whether a chemical myristoylation strategy via the OPAL could be achieved, the aldol reaction between glyoxyl-LYRAG **27** and myristaldehyde **126**, using L-proline **61** as a catalyst, performed in 25 mM PB, pH 7.5, for 24h at 37 °C, was investigated. DMSO was used both to prepare myristaldehyde **126** stock solutions, and as a co-solvent for the reaction. Three chemical myristoylation reactions of glyoxyl-LYRAG **27** to give chemically myristoylated peptide **130** were performed, with differing amounts of DMSO content used in each experiment (Figure 6.2)
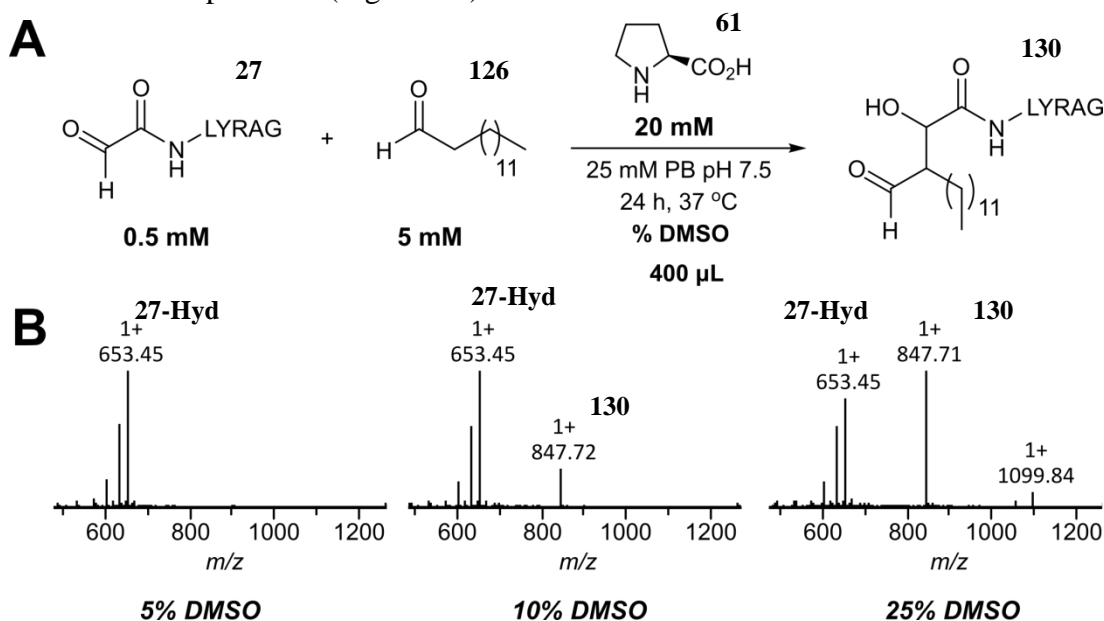
Figure 6.2 (A) Chemical myristoylation of glyoxyl-LYRAG **27** using myristaldehyde **126** and L-proline **61** as a catalyst in solvent systems containing different amounts of DMSO. (B) Results for the chemical myristoylation of glyoxyl-LYRAG **27** in solvent systems with different DMSO content. Calculated [M+H]$^+$ of **130** = 847.52 Da

Successful chemical myristoylation of glyoxyl-LYRAG **27** was confirmed by LC-MS and MS/MS analysis with a 14% conversion to **130** observed when performing the reaction in 10% DMSO, and a 42% conversion to **130** observed when performing the reaction in 25% DMSO. For the reaction performed in 25% DMSO, a very small amount of doubly lipidated peptide was observed at 1059.96 Da was observed (<2%), and an unknown peak at 1099.84 Da was also observed. No conversion to the desired chemical myristoylated peptide **130** was observed when performing the reaction in 5% DMSO. The results from the chemical myristoylation of glyoxyl-LYRAG **27** suggested that the DMSO content of the reaction was a limiting factor in the rate of the ligation. This will undoubtedly be related to the insolubility of the myristaldehyde **126** reagent; the higher the DMSO content of the reaction, the higher concentration of myristaldehyde **126** reagent in solution available for reaction with the glyoxyl-LYRAG peptide **27**.

6.1.2 Chemical myristoylation of HASPA

The chemical myristoylation procedure was next extended towards the modification of HASPA. Conditions for chemical myristoylation of HASPA were first screened using glyoxyl-HASPA generated from HASPA bearing an *N*-terminal glycine as outlined in Table 6.1.

*NB: In all cases of chemical lipidation of HASPA, peaks at M -18 were noted in the LC-MS experiments, which likely corresponds to the aldol condensation product with a loss of H_2O. This was only found for chemical lipidation via the OPAL and was not noted in any other OPAL modifications with probes **85**-**88**. It was hypothesised that this dehydration was related to the hydrophobicity of the attached lipid chain, which is in close proximity to the β-hydroxy group.*

Table 6.1 Outline of conditions screened for the chemical myristoylation of glyoxyl-HASPA **36** (generated from PLP **8** transamination of HASPA **35**). * Conversion calculation based on **35**, **36-Hyd**, anticipated PLP by product, and **127** peak intensity. ** Incomprehensible data obtained.

Entry	Protein **36** (µM)	Myristaldehyde **126** (mM)	% DMSO	Conversion to **127** (%)
1	120	25	50	15*
2	120	5	65	50*
3	120	15	65	63*
4	120	25	65	63*
5	120	15	90	N/A**
6	120	25	90	N/A**
7	30	15	70	N/A**
8	60	15	70	N/A**
9	240	15	70	53%*
10	480	15	70	71%*

Figure 6.3A outlines the reaction described in Table 6.1, Entry 10, and Figure 6.3B shows the LC-MS data obtained for Entry 10 of Table 6.1.
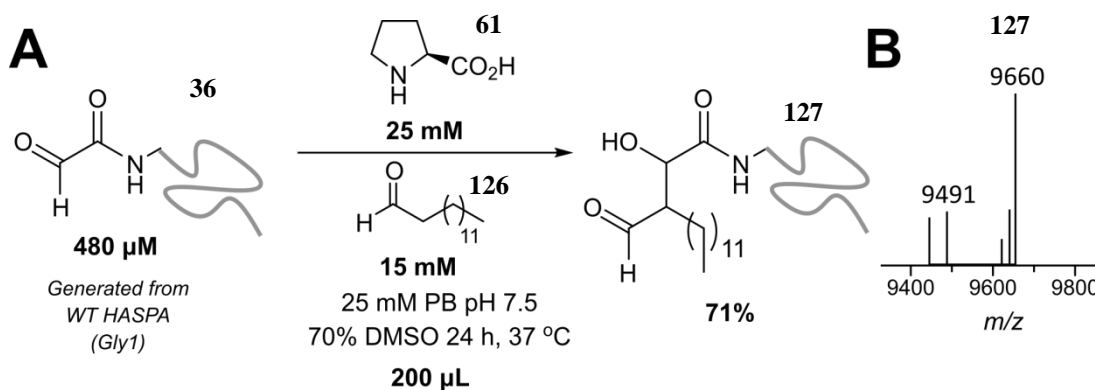
Figure 6.3 (A) Chemical myristoylation of glyoxyl-HASPA **36** (generated from PLP **8** transamination of HASPA **35**). (B) LC-MS data obtained for Table 6.1, Entry 10. Peak at 9491 Da was anticipated to be a PLP-related by-product. Calculated mass of **127** = 9657 Da

Conditions for successful chemical myristoylation as outlined in Entry 10 of Table 6.1 were then applied towards chemical myristoylation of glyoxyl-HASPA **61** generated from HASPA(G1S) **37** bearing an *N*-terminal serine. This was also applied to 15N labelled glyoxyl-HASPA **36-15N** (Figure 6.4). Full conversion to the desired

chemically myristoylated HASPA **127** and **127-15N** was achieved as judged by LC-MS.
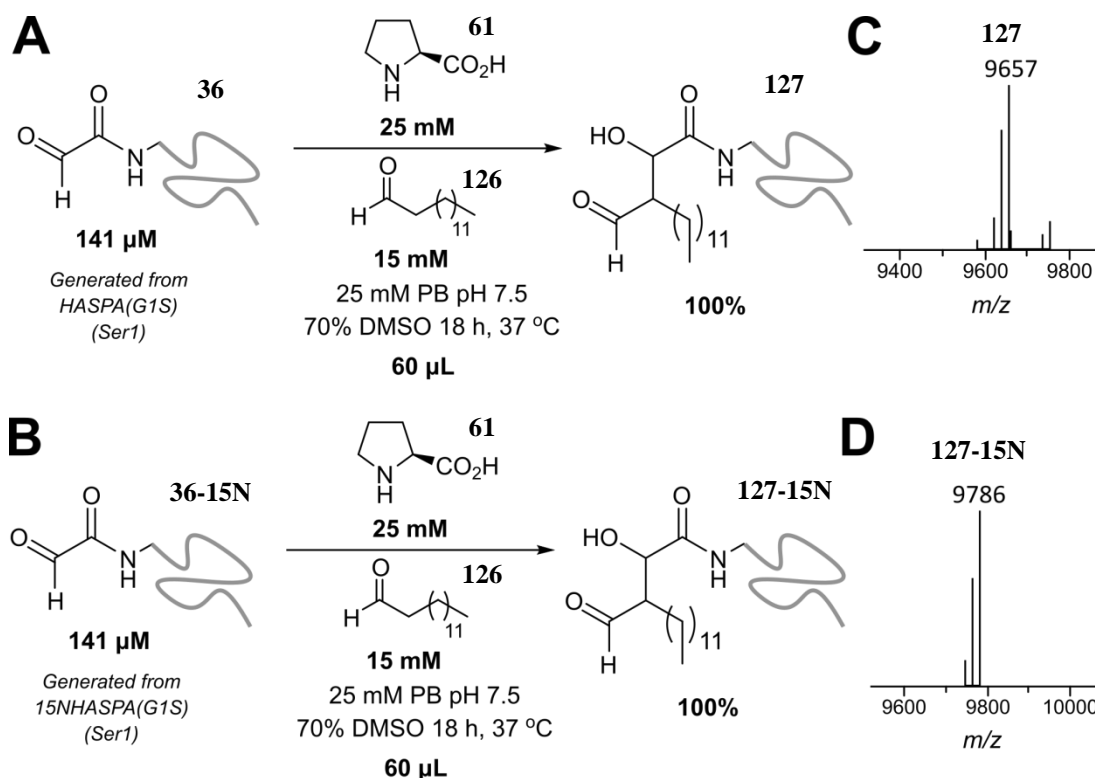
Figure 6.4 (A) Chemical myristoylation of glyoxyl-HASPA **36** (generated from NaIO$_4$ oxidation of **37**) with myristaldehyde **126** using L-proline **61** as an organocatalyst. (B) Chemical myristoylation of glyoxyl-HASPA **36-15N** (generated from NaIO$_4$ oxidation of **37-15N**) with myristaldehyde **126** using L-proline **61** as an organocatalyst. (C) LC-MS data obtained for (A). Calculated mass of **127** = 9657 Da. (D) LC-MS data obtained for (B). Calculated mass of **127-15N** = 9787 Da.

6.1.3 Isolation of chemically myristoylated protein

In order to carry out downstream applications of the chemically myristoylated HASPA (e.g. modification of the β-hydroxy aldehyde, protein NMR), removal of the excess myristaldehyde **126** and high DMSO content post reaction was first required. To achieve this, multiple 60 µL chemical myristoylation reactions were pooled to give an estimated protein content of 640 µg (based on initial concentration of the unmodified HASPA(G1S) **37** stock). The pooled reactions were then diluted with water to give solutions <20% DMSO. This solution was then desalted using multiple PD Minitrap G-25 columns (GE Healthcare), and the desalted protein was lyophilised to give the chemically myristoylated HASPA **127** as a white powder. The protein could then be stored at -80 °C until required for downstream use. Typically, a

25% protein recovery (determined by SDS-PAGE analysis,[viii] based on initial concentration of the unmodified HASPA(G1S) **37** stock) was obtained when following this procedure.

6.1.4 NMR characterisation of chemically myristoylated HASPA

NMR spectroscopy was used to characterise the chemically myristoylated HASPA **125**, and to assess the similarity of its structural and solution based properties to that of enzymatically *N*-myristoylated HASPA **122** (the characterisation of which has been previously reported).[207] Following NMR characterisation of chemically myristoylated HASPA **125**, heteronuclear single quantum coherence (HSQC) spectra revealed the chemical mimic **125** to be highly similar to that of the enzymatically *N*-myristoylated HASPA **122**. These data demonstrated that chemical myristoylation of HASPA effectively mimics that of enzymatically *N*-myristoylated HASPA.[ix]

NB: For a detailed discussion on protein NMR of HASPs, readers are directed to reference 207.
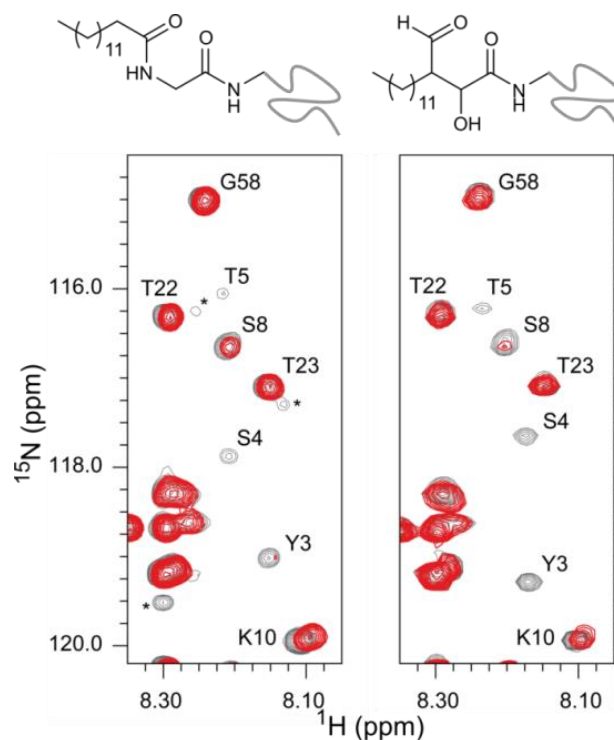
Figure 6.5 Regions of 2D (1H, 15N) HSQC NMR spectra of [15N]labelled HASPA. Comparison of unmodified **35** (grey) and enzymatically myristoylated (red) HASPA **124** (left). Comparison of unmodified **37** (grey) and chemically myristoylated (red) HASPA **127** (right). (1H,15N) Resonance assignments are indicated. Unassigned peaks are denoted by asterisks.

6.2 Dual chemical lipidation of HASPA

With a procedure in place for the chemical myristoylation of HASPA, further modification of the chemically myristoylated HASPA **127** with a second alkyl chain via the β-hydroxy aldehyde handle was investigated. First, a suitable C16 aminooxy reagent **128** was synthesised for use in the oxime ligation (see **Experimental**). Palmitoyl aminooxy **128** was then prepared as a 25 mM stock solution in DMSO for use in the modification of chemically myristoylated HASPA **127-15N** samples. Reactions of the isolated samples were performed in 100 mM PB, pH 6.5, 70% DMSO at 37 °C, using 100 mM aniline **46** as a catalyst (Figure 6.6A). Conversions after 48 h and 96 h of reaction time were judged by LC-MS analysis (Figure 6.6B).

Figure 6.6 (A) Oxime ligation of chemically myristoylated HASPA **127-15N** with palmitoyl aminooxy **128** using aniline **46** as a catalyst in a PB/DMSO solvent system. (B) LC-MS data obtained for (A) after 48 h (left) and 96 h (right). Calculated mass of **129** = 9787 Da. Calculated mass of **129-15N** = 10026 Da.

Despite using long reaction times and high concentrations of palmitoyl aminooxy **128**, poor conversions to the desired dually modified HASPA were obtained, with a maximum conversion of 30% obtained after 4 days. During the reaction, notable precipitation had occurred, which was hypothesised to be the reason for the low conversion observed. This precipitation was established to be due to the palmitoyl aminooxy **128** reagent; control experiments of incubating **128** in 100 mM PB, pH

6.5, 70% DMSO at 37 °C with 100 mM aniline **46** led to precipitation after 18 h. Therefore, to avoid precipitation issues and low conversions to the dually modified protein, EtOH was next investigated as a co-solvent for the second alkylation step of HASPA. EtOH is miscible with water, and has previously been used as a solvent for reactions involving palmitoyl aminooxy **128**. [208] Initially, reactions of chemically myristoylated HASPA with palmitoyl aminooxy **128** were set up in 100 mM PB pH 6.5, 70% EtOH at 37 °C with 100 mM of aniline;333 these conditions however led to immediate precipitation of phosphate salts. Reducing the concentration of PB from 100 mM to 25 mM circumvented this issue, and reactions of chemically myristoylated HASPA with palmitoyl aminooxy **128** were performed up in 25 mM PB pH 6.5, 70% EtOH at 37 °C using 100 mM aniline as a catalyst (Figure 6.7A). Conversions after 48 h and 96 h of reaction time under these conditions were judged by LC-MS analysis (Figure 6.7B).
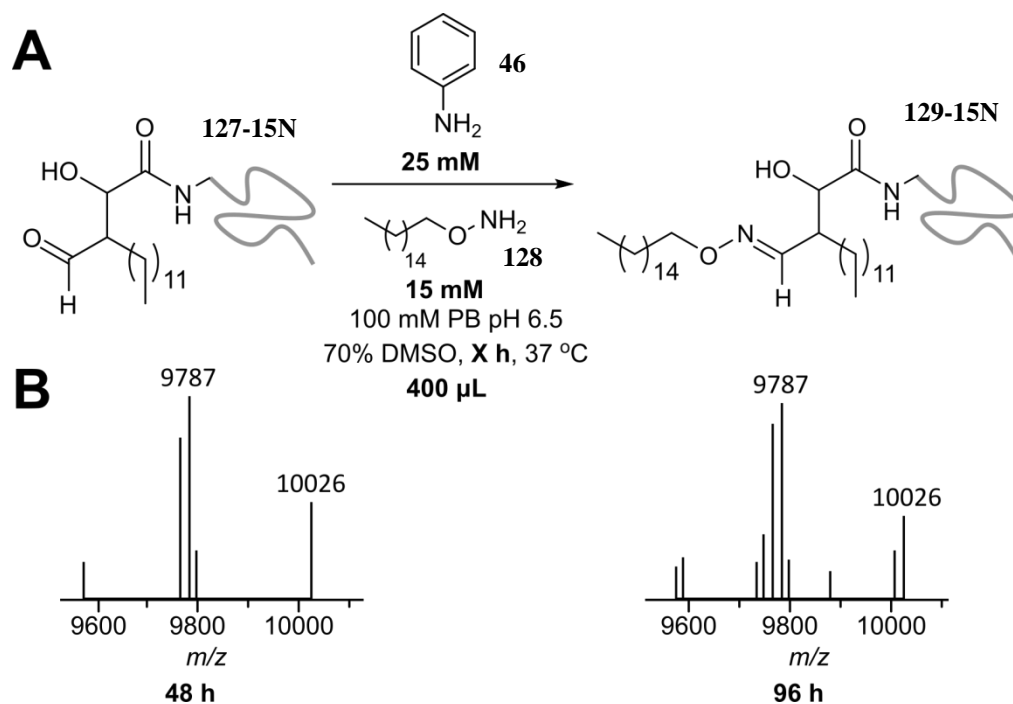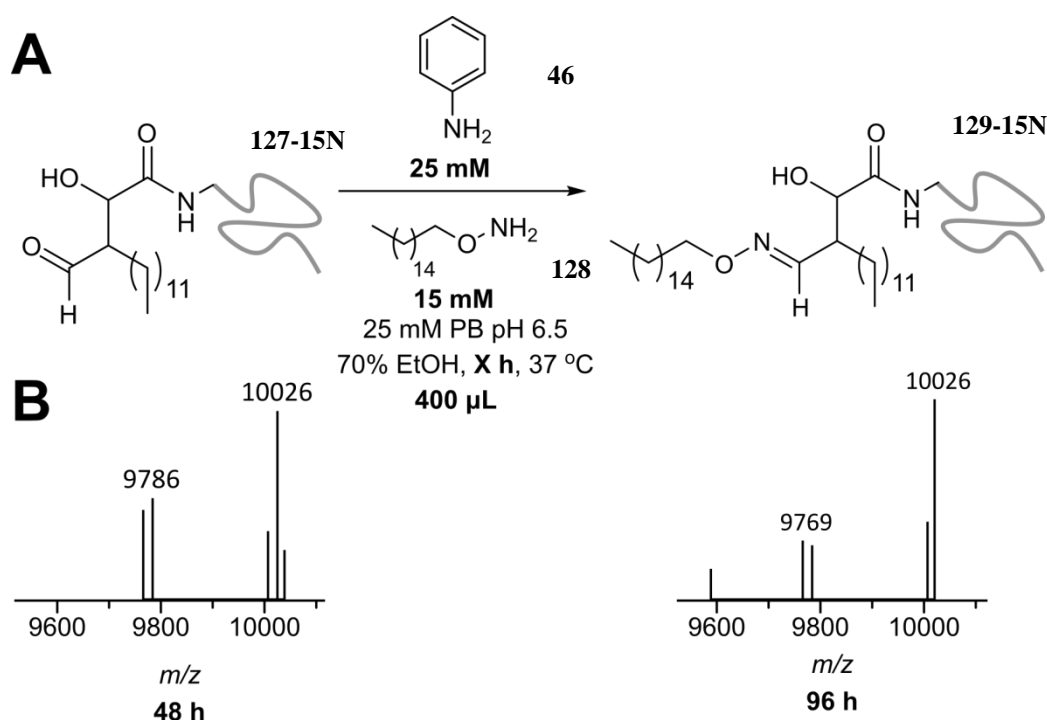

Figure 6.7 (A) Oxime ligation of chemically myristoylated HASPA **127-15N** with palmitoyl aminooxy **128** using aniline **46** as a catalyst in a PB/EtOH solvent system. (B) LC-MS data obtained for (A) after 48 h (left) and 96 h (right). Calculated mass of **129** = 9787 Da. Calculated mass of **129-15N** = 10026 Da.

As shown in Figure 6.7, significantly higher conversions to dually modified HASPA were obtained when using EtOH as a co-solvent compared to using DMSO, with a maximum of 75% conversion obtained after 4 days of incubation.

6.2.1 Liposome binding of chemically lipidated HASPA

As discussed earlier, the HASP lipidation is postulated to facilitate binding to the plasma membrane. To investigate this, liposome sedimentation assays of chemical myristoylated HASPA **127** and dually lipidated HASPA **129** (along with unmodified HASPA(G1S) **37** (as a control) were performed to assess the proteins ability to bind to membranes. [x] SDS PAGE analysis (Figure 6.8) revealed no liposome binding by HASPA(G1S) **37** throughout a 30 h time period, whereas liposome binding was observed with lipidated HASPA **127** and **129** (see lanes P in Figure 6.8). The results of this *in vitro* system suggest that the addition of a lipid chains to HASPs facilitate binding of the protein to membranes, supporting the hypothesis that these PTMS assist trafficking of HASPSs to the plasma membrane *in vivo*.[175]

Figure 6.8 SDS-PAGE analysis of liposome sedimentation assay of unmodified HASPA **37**, chemically myristoylated HASPA **127** and dually lipidated HASPA **129**. The supernatant (S) and resuspended liposome pellet (P) were analysed. (M = Molecular weight marker, ladder).

[x] The liposome preparation, binding assays, and characterisation were performed by Dr Tessa Keenan (see Reference 2 for details).

6.3 Future work

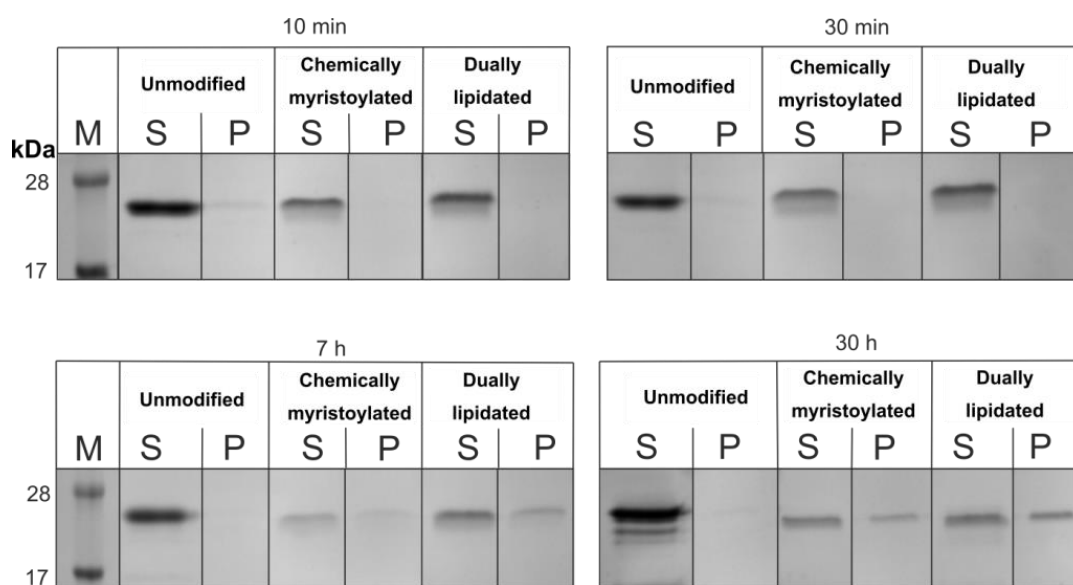Through aldehyde installation and subsequent OPAL, chemical mimics of *N*-myristoylated HASPA have been described. This "chemical myristoylation" procedure offers certain benefits over that of the natural, enzymatic procedure of *N*-myristoylation. First, chemical myristoylation does not require NMT, avoiding both the requirement of expressing/purifying recombinant NMT and the necessity of removing the enzyme post *N*-myristoylation of HASPA. Secondly, chemical myristoylation utilises myristaldehyde **126** as the source of the myristyl group. At the time of writing, myristaldehyde **126** is significantly more commercially available compared to that of myristoyl coA used in the enzymatic *N*-myristoylation procedure, making chemical myristoylation more attractive of the two procedures from a scaling and costing perspective. Finally, chemical myristoylation lends itself to lipidation of other protein targets. Enzymatic *N*-myristoylation is limited to certain proteins due to substrate specificity of NMT. Alongside the absolute requirement of an *N*-terminal glycine, as many as 17 *N*-terminal residues can be involved in determining whether or not a protein (or peptide) will undergo *N*-myristoylation.[176] This selectivity is understandably important in nature to avoid offsite *N*-myristoylation of other intracellular proteins; from a bioconjugation perspective however, it limits NMTs applicability towards myristoylation of other protein targets. The only requirement for chemical myristoylation is a glyoxyl handle, which can be installed both at the *N*-terminus and at internal positions of protein, suggesting this procedure could be applied to protein targets which do not/cannot undergo conventional myristoylation.

The chemical myristoylation procedure described in this chapter for the chemical myristoylation of HASPA does, however, have notable drawbacks. The myristaldehyde **126** substrate is an alkyl donor, requiring high concentrations of donor similar to that of butanal **51**. As a result, an exceptionally large percentage of DMSO co-solvent was required to solubilise donor **126**, which in turn was also required to obtain full conversion to the chemically myristoylated HASPA **127**. Employing α-aryl aldehyde based myristyl donors would likely reduce the required donor concentration. Alternatively, different organocatalysts, such as L-proline **61**

derivatives bearing long alkyl chains, may prove more compatible with the chemical myristoylation procedure. Different buffer composition may also lead to an improved chemical myristoylation procedure.

With respect to isolation of the chemically myristoylated HASPA, although the procedure described in this chapter does allow for purification/isolation of chemically myristoylated HASPA, the process of diluting samples to >20% DMSO, purification by multiple PD MiniTrap G-25 columns (GE Healthcare), and lyophilisation to obtain the protein for downstream applications is time consuming costly in terms of resources, and leads to poor protein recovery. The isolation step of the chemical myristoylation process in particular would benefit from significant optimisation (although part of this would be reliant on optimisation of the reaction conditions). Since the HASPA bears a histidine tag, nickel affinity chromatography would likely be a good candidate for purifying the chemically myristoylated proteins species. Alternatively, a more standard method of dialysis may also be applicable towards isolation of chemically myristoylated HASPA **127**.

The β-hydroxy aldehyde handle of the chemically myristoylated HASPA **127** proved to be compatible with aniline-catalysed oxime ligation for the addition of a second long alkyl chain, using 100 mM aniline **46** as the catalyst and palmitoyl aminooxy **128** as the aminooxy reagent. Similarly to other dual bioconjugations described in Chapter 5, aniline-catalysed oxime ligation of chemically myristoylated HASPA suffers from slow ligation kinetics, and requires high concentrations to obtain >70% conversions to the dually alkylated HASPA. The palmitoyl aminooxy reagent **128** also proved to be a challenging substrate to work with requiring a high percentage of EtOH co-solvent to solubilse **128**. Different buffer compositions, or using the SPANC ligation (which has previously been utilised for labelling glyoxyl-peptides with C16 chains) [146] may potentially lead to a more efficient procedure for conjugation of the second alkyl chain to chemically myristoylated HASPA **126**. Alternatively, dually modified HASPA species could be generated by first installing the ThzK unnatural amino acid at position 4 of HASPA(G1S) **37** (the site that undergoes S-palmitoylation in nature), subjecting the protein to Pd-mediated decaging and subsequent periodate oxidation to reveal α-oxo aldehyde handles at both position 1 and 4 of HASPA, followed by OPAL using L-proline **61** and myristaldehyde **126** to simultaneously install two long alkyl chains into HASPA.

This procedure is arguably more attractive than the current procedure of installing the second chain at the β-hydroxy aldehyde position, as it would be a one pot procedure, and would avoid the necessity of using both the less reactive β-hydroxy aldehyde group and the highly insoluble palmitoyl aminooxy reagent **128**.

6.4 Conclusions

To conclude, a strategy for the chemical myristoylation and dual lipidation of HASPA has been described. Following installation of glyoxyl functionality, HASPA could be "chemically myristoylated" *in vitro* via the OPAL using L-proline **61** as the organocatalyst and myristaldehyde **126** as the donor to give the *N*-myristoylated HASPA mimic **127**. This chemically myristoylated mimic **127** displayed near identical structural and solution properties to that of enzymatically myristoylated HASPA **124** as determined by NMR spectroscopy. Chemically myristoylated HASPA **127** could then be further modified via the β-hydroxy aldehyde handle using palmitoyl aminooxy **128** to synthesise a HASPA bearing two long chain alkyl groups. Finally, both mono and dually chemically lipidated HASPA were found to show improved binding to liposomes in comparison to unmodified HASPA **37**, reaffirming the hypothesis of HASP lipidation facilitating membrane binding.

Chapter 7: Beyond the OPAL and closing remarks

7.1 Discovery of a retro-aldol mediated breakdown of OPAL products

The OPAL procedure synthesises a hydrolytically stable C-C bond between protein and conjugate, which is a major advantage over other protein bioconjugation strategies, such as hydrazone/oxime ligation for example which are prone to hydrolysis.[140] For certain applications, however, controlled hydrolysis/release of the covalently attached moiety is desirable; this is highly apparent in ADCs, which are required to release their payload upon arrival at the targeted site. Traditional methods of bioconjugate cleavage include the use of hydrazones (hydrolysed over time under acidic conditions)[209] and disulfides (cleaved by high intracellular levels of thiols such as glutathione).[210] However, these methods require the use of labile linkers, which can lead to uncontrolled/off site release of conjugate *in vivo* (as seen in the ADC Mylotarg, which uses a hydrazone linker).[211] Alternative strategies focus on incorporating a peptide linker into the chemical probe which can undergo *in vivo* enzymatic cleavage as seen in cathepsin cleavable linkers, providing a more stable alternative linker. [212]

In the context of the aldol reaction, the C-C bond cleavage of the β-hydroxy carbonyl product to generate two carbonyl species (i.e. the reverse of the aldol reaction) is known as the retro-aldol reaction. The primary focus of retro-aldol research has been in the study of aldolase enzymes. In nature, aldolases can catalyse both the forward and retro-aldol processes, feature in several biologically important pathways such as sugar metabolism, [213] and are classified into two different categories based on their mechanism of action; Class I aldolases utilise an enamine based mechanism which is thought to involve a catalytically active lysine residue, whereas Class II aldolases operate *via* a metal-mediated process utilising a transition metal such as Zn^{2+}. In organic synthesis, natural and engineered aldolases have been used as catalysts to promote both C-C bond forming aldol reactions and C-C bond breaking retro-aldol reactions.[214,215] In particular, the retro-aldol C-C bond cleavage of methodol has seen much interest within the context of computationally designing and engineering protein biocatalysts. [215-217] Small molecule induced retro-aldol reactions for metal-mediated C-C activation has also been described in the literature.[218,219]

As discussed earlier, the protein bioconjugates synthesised *via* the OPAL are hydrolytically stable and do not breakdown. It was envisaged, however, that it might

be possible to induce retro-aldol breakdown of the OPAL products by some other means, such as treating the bioconjugate with a small molecule catalyst. Bearing this in mind, attention was drawn towards the aniline-catalysed oxime ligations performed on aldol-modified peptides (Figure 7.1A). As discussed in Chapter 5, in some cases a small peak at 740 Da, which was hypothesised to correspond to a singly modified oxime-LYRAG species **137** was identified in these samples. This was initially thought to occur from using stock samples of aldol-modified peptide that contained small amounts of unreacted glyoxyl-LYRAG **27**. However, the peak intensity of oxime peptide **137** was noted to be different in each sample depending on the catalyst and pH used; in particular, a significant peak corresponding to oxime peptide **137** was noted in the LC-MS analysis of the aniline-catalysed oxime ligation of aldol modified peptide **59** at pH 7.5 using catalyst **18**, whereas no such species was present when using catalyst **110** (Figure 7.1B). This suggests that, during the dual modification, breakdown of the aldol-modified peptides back to glyoxyl-LYRAG **27**, followed by oxime ligation, may have occurred in order to generate oxime peptide **137** (Figure 7.1C).
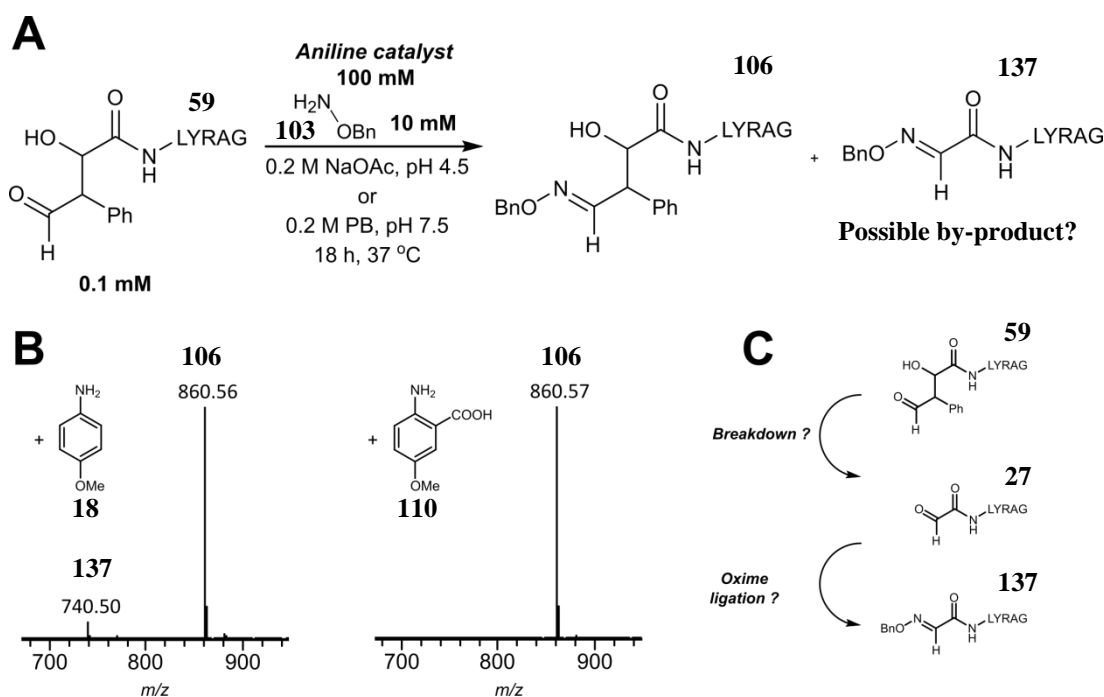

Figure 7.1 (A) Outline of aniline catalysed oxime ligation of aldol product **59** to generate dually modified peptide **106** and byproduct oxime product **137**. (B) LC-MS data obtained when using catalyst **18** or catalyst **110** at pH 7.5, showing the presence of the anticipated byproduct **137** present when using catalyst **18** but not when using catalyst **110**. (C) Proposed pathway for generation of oxime product **137**.

It was therefore hypothesised that incubating OPAL products with an aniline catalyst without the addition of an aminooxy nucleophile may lead to breakdown of the OPAL products to the α-oxo aldehyde species, thus inducing cleavage of the otherwise stable C-C bond and regenerating the original α-oxo aldehyde species *via* a retro-aldol mediated process. To investigate this hypothesis, aldol product **59** was incubated with each of the aniline catalysts **18**, **46**, **110**, and **111** (or with no catalyst as a control) in either pH 4.5 or pH 7.5 buffer for 18 h at 37 °C, with the outcome analysed by LC-MS (Figure 7.2). Notably different results were obtained depending on the pH used and choice of catalyst. For example, incubation of aldol product **59** with aniline **46** at pH 4.5 lead to no breakdown of aldol product **59**. However, analysis by LC-MS of the incubation of aldol product **59** with aniline **46** at pH 7.5 revealed the appearance of a peak at 653 Da, which would correspond to the hydrated glyoxyl-LYRAG **27-Hyd**. Based on peak intensity, a maximum of 52% conversion to glyoxyl-LYRAG **27** was observed when using catalyst **18** at pH 7.5. Similar to oxime ligation of the β-hydroxy aldehyde of OPAL products, this breakdown appears to be facilitated at pH 7.5 in comparison to pH 4.5 (Figure 7.2).
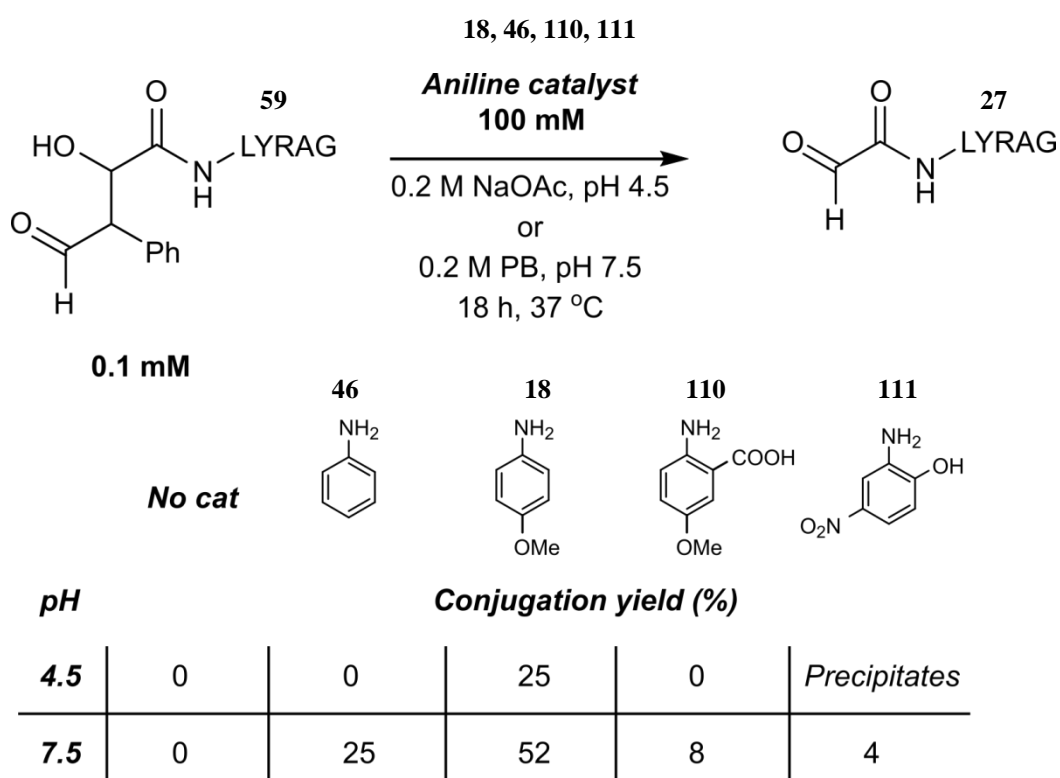

pH	No cat	46	18	110	111
4.5	0	0	25	0	Precipitates
7.5	0	25	52	8	4

Figure 7.2 Outline of investigation into pH effect on the outcome of aniline-catalysed breakdown of aldol product **59**. Conversions glyoxyl-LYRAG **27** were obtained as judged by LC-MS.

NB: An additional aniline catalyst, p-phenylenediamine (previously described for aniline catalysed oxime ligation), [123] was also screened for retro-aldol activity, along with catalysis in oxime ligation of aldol product 59. However, data obtained when using this catalyst yielded multiple unknown species when judged by LC-MS, and was subsequently not investigated further.

The results shown in Figure 7.2 suggest that significant, controlled breakdown of OPAL products could be achieved through incubation with catalyst **18** (and, to a lesser extent, with catalyst **46**), which likely occurs through a retro-aldol mediated process.

7.1.1 Mechanism for the retro-aldol mediated breakdown of OPAL products

There are isolated examples of small molecule-mediated retro-aldol reactions,[218,219] with the primary focus of research into retro-aldol chemistry focused within the area of engineering Class I aldolases to catalyse the breakdown of small molecule β-hydroxy carbonyls.[220] The retro-aldol mediated cleavage of C-C bonds through these retro-aldolases is hypothesised to occur through initial imine formation between the ε-amino group of a catalytically active lysine residue in the aldolase active site and the β-hydroxy carbonyl of the aldol molecule.[215] Previously reported studies of aldolases have determined that this catalytic lysine residue resides in a hydrophobic pocket of the enzyme, resulting in a "highly perturbed" ε-amino pKa value of ~6 (in contrast to the typical pKa of ~9-10 seen in other lysine residues).[221] The similarity in this pKa value compared to the pKa of aniline (pKa = 4.6) and *p*-anisidine (pKa = 5.3) and the propensity for anilines to form imines may explain why these molecules are effective at promoting a retro-aldol mediated C-C cleavage of aldol modified peptides. Based on these previous observations, a potential mechanism for the aniline-catalysed retro-aldol mediated breakdown of OPAL products was proposed, and is shown in Figure 7.3.
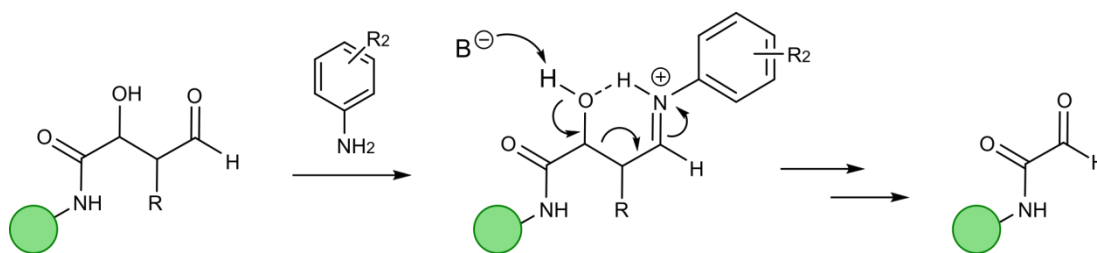
Figure 7.3 Proposed mechanism for the aniline-catalysed retro-aldol mediated breakdown of OPAL products. Imine formation between the β-hydroxy aldehyde, followed by decomposition and hydrolysis, leads to breakdown of the OPAL product to regenerate an α-oxo aldehyde species. (B = unspecified base).

7.1.2 Controlling the pathway to oxime ligation or retro-aldol

As shown in Chapter 5, the addition of an aniline catalyst was found to be essential for facilitating the oxime ligation of OPAL products. In particular, catalyst **18** was especially effective at catalysing this process at pH 7.5. During these experiments no observed breakdown of OPAL products back to the protein α-oxo aldehyde (or to singly oxime modified product) was observed. However, as shown in Figure 7.2, catalyst **18**, along with catalyst **46** (which was used to generate dually lipidated HASPA **129**), were found to be effective at promoting breakdown of aldol-modified products **59**. Taking both the results described in this Chapter and in Chapters 5 and 6 into account, it was hypothesised that an additional factor in addition to catalyst loading, catalyst choice, and pH choice, must influence whether or not retro-aldol mediated breakdown of OPAL products occurred. Notably, it has previously been shown for the π-clamp mediated site-selective bioconjugation[42] of cysteines that alteration of the buffer salt composition can dramatically impact on the outcome of bioconjugation.[43] Throughout this thesis, the aniline-catalysed oxime ligation of peptide aldol products was consistently performed in 200 mM buffer, whereas aniline-catalysed oxime ligation of protein aldol products was consistently performed in 50 mM buffer (the exception being for synthesis of dually lipidated HASPA **129** which was performed in 25 mM buffer). To check the potential impact of buffer salt concentration on the outcome of retro-aldol mediated breakdown of OPAL products, aldol product **59** was incubated with catalyst **18** at pH 7.5 with different concentrations of PB salts, and conversion to glyoxyl-LYRAG **27** was judged by LC-MS (Figure 7.4).
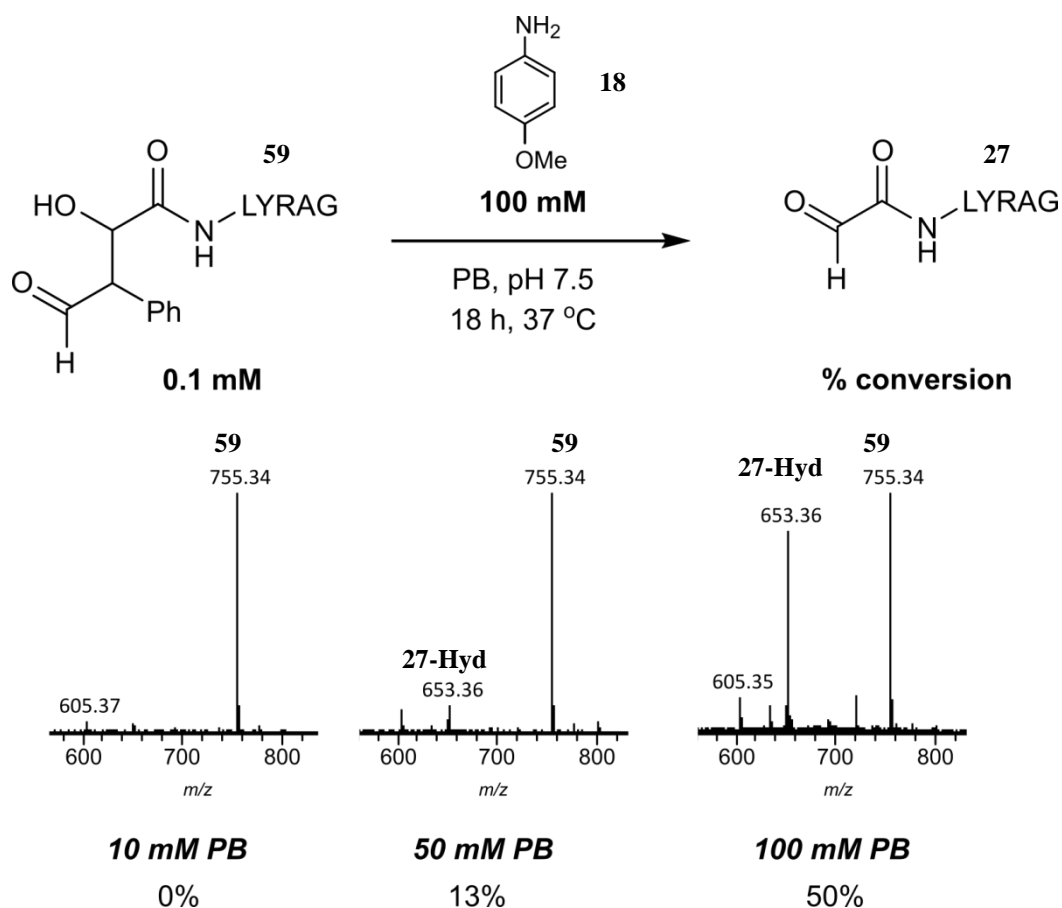
Figure 7.4 Outline of investigation into the effects of concentration of buffer salts used in the retro-aldol mediated breakdown of **59** using catalyst **18**. Conversion to glyoxyl-LYRAG **27** was assessed as judged by LC-MS.

NB: Addition of an aniline catalyst to the reaction medium can sometimes alter pH of a given reaction (particularly in weakly buffering solutions). The actual pH for these reactions was therefore assessed by pH probe, and found to be pH 7.4-7.6 in all cases.

As shown in Figure 7.4, significantly higher conversion to glyoxyl-LYRAG **27** was observed when using higher concentrations of PB salts, suggesting that the aniline catalysed retro-aldol mediated breakdown of OPAL products was facilitated in systems with higher concentrations of PB salts. Comparisons of Figure 7.2 and Figure 7.4 suggest that this was limited up to 100 mM of PB, as moving from 100 mM PB salts to 200 mM PB salts had a negligible effect on conversion to glyoxyl-LYRAG **27** These results would explain why retro-aldol mediated breakdown was observed during aniline-catalysed oxime ligation experiments of aldol product **59** (performed with 200 mM PB) but not in aniline-catalysed oxime ligation

experiments of OPAL described in Chapter 5 (performed with 50 mM PB) or in the dual lipidation of HASPA described in Chapter 6 (performed with 25 mM PB).

7.1.3 C-C bond cleavage of protein bioconjugates

Finally, to begin to investigate retro-aldol mediated cleavage of functionalised OPAL products, fluorescently labelled thioredoxin **89** (25 µM) was incubated with catalysts **18**, **46**, **110**, and **111**, in 0.2 M PB pH 7.5 at 37 ºC (0.2 M, as opposed to 0.1 M, PB was chosen in an attempt to maximise conversion). The anticipated breakdown to glyoxyl-thioredoxin **32** was judged by LC-MS after 18 h or 42 h intervals, as outlined in Table 7.1.

Table 7.1 Outline of screening catalysts in the retro-aldol mediated breakdown of fluorescently labelled thioredoxin **89**.

Entry	Catalyst	Time (h)	Conversion to glyoxyl-thioredoxin **32** (%)
1	**46**	18	24*
2	**18**	18	45*
3	**110**	18	18*
4	**111**	18	*Precipitate****
5	-	18	0*
6	**18**	42	57*

* A small, additional peak at 11676 Da was also present in all samples treated for retro-aldol, which currently has not been identified. ** In this particular set of experiments, a small amount of precipitate was observed, and for this reason the data was not analysed.

Figure 7.5 shows the results from Table 7.1, Entry 6:
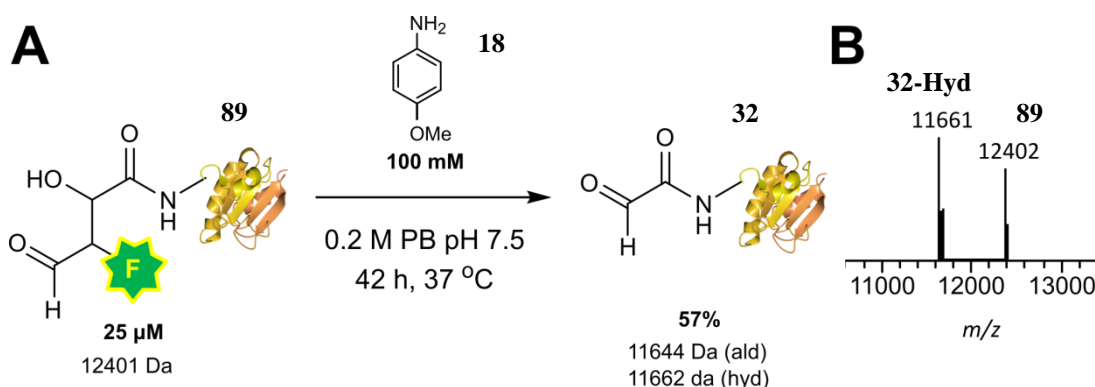
Figure 7.5 (A) Outline of the retro-aldol mediated breakdown of OPAL product **89** using catalyst **18**. (B) LC-MS data obtained for Table 7.1, Entry 6.

Similar to retro-aldol mediated reactions of aldol product **59**, catalyst **18** was found to be most effective at promoting retro-aldol mediated breakdown of fluorescently labelled thioredoxin **89** to glyoxyl-thioredoxin **32**, with a 57% conversion to glyoxyl-thioredoxin **32** observed after 42 h of incubation. These results demonstrate that cleavage of an otherwise stable C-C linked OPAL bioconjugate can be achieved through aniline catalysis, and that the original protein α-oxo aldehyde can be regenerated in the process.

7.2 Variants on aldol bioconjugation

Throughout Chapters 3-6 it has been demonstrated that protein α-oxo aldehydes can be site-selectively dually modified first through an organocatalyst-mediated process that uses a proline-derived catalyst, followed by another organocatalyst-mediated process that uses an aniline-derived catalyst, in a two-step tandem process. Intriguingly, in small molecule literature, the use of both a proline-derived organocatalyst and an aniline-derived organocatalyst in the context of aldol reactions leads to different process known as the Mannich reaction.[222] As discussed in Chapter 1, a Mannich ligation strategy for modifying proteins has been previously reported; this ligation strategy is selective for surface exposed tyrosine residues through a three component reaction involving both aniline-type and aldehyde-type reagents, resulting in formation of stable C-C bond between the substrates and the ortho position of the tyrosine phenol side chain.[223] Similar bioconjugations involving preformed imines and phenols/tyrosine containing peptides and proteins has also been reported.[224] Whilst the reported Mannich ligation is selective for surface

exposed tyrosine residues it does not distinguish between multiple surface exposed tyrosine residues, leading to heterogeneous protein constructs.[223]. Both *ortho* positions (with respect to the hydroxyl group) of the tyrosine phenol side chain can be modified, particularly for small peptides, further leading to heterogeneous product formation.[224]

7.2.1 A 'modified Mannich' bioconjugation

It is clear from the published literature that it is possible to take phenol (i.e. tyrosine) containing peptides/proteins, and chemically modify them selectively with aldehyde and aniline reagents. Additionally, since the publication of the three component Mannich ligation (nearly 15 years ago), significant advancements have been made within the field of protein bioconjugation; in particular, incorporation of handles such as aldehydes into proteins has advanced significantly.[1] Therefore, it should theoretically be possible to take aldehyde containing peptides/proteins, and chemically modify them selectively with phenol and aniline reagents via an alternative 'modified-Mannich' ligation strategy. By installing one aldehyde into the protein and having the 'aldehyde' component of the reaction as the limiting agent, product heterogeneity should be avoided. Additionally, by using functionalised phenol and aniline reagents in this reaction, it should be possible to achieve dual functional modification of the protein in question *via* a one-step process (Figure 7.6).
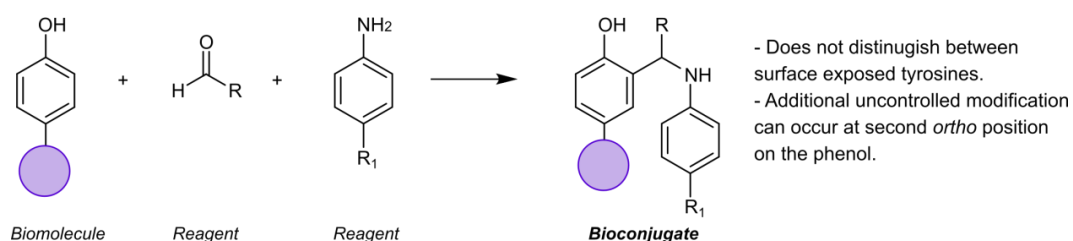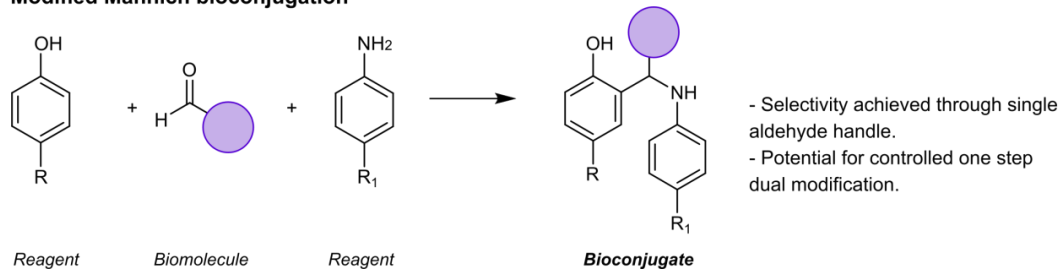
Figure 7.6 Comparison of the three component Mannich bioconjugation strategy to the hypothesised modified Mannich bioconjugation strategy.

To investigate this hypothesis, two modified-Mannich ligations between the model peptide glyoxyl-VARLG **132** (chosen because it does not contain any tyrosine residues which may interfere with the initial validation stages), commercially available aniline **133** and substituted phenol reagents **134** or **135** were performed in 0.1 M PB pH 6.5 for 24 h at 37 °C. Conversion to the anticipated Mannich products **136** and **137** respectively was assessed as judged by LC-MS analysis (Figure 7.7).
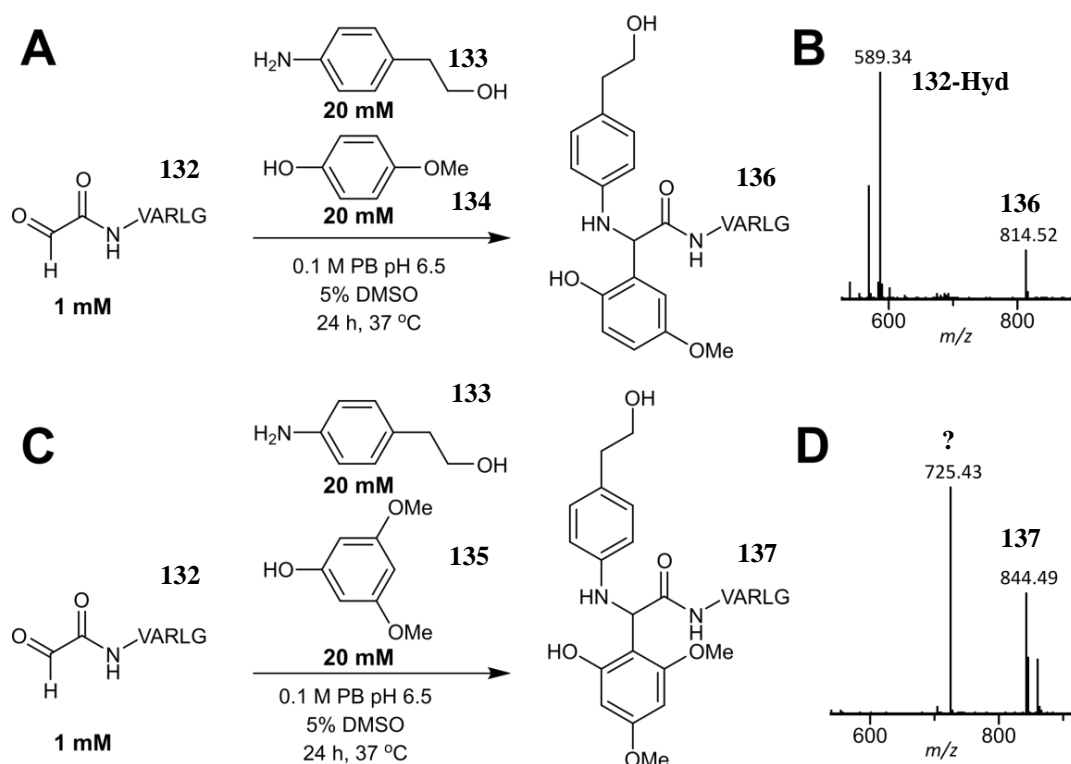
Figure 7.7 (A) Outline of reaction of glyoxyl-VARLG **132** with aniline **133** and phenol **134**. Calculated [M+H]$^+$ of product **136** = 814.44 Da. (B) LC-MS data obtained for (A). (C) Outline of reactyion of glyoxyl-VARLG **132** with aniline **133** and phenol **135**. Calculated [M+H]$^+$ of product **137** = 844.45 Da. (D) LC-MS data obtained for (C).

As shown in Figure 7.7, conversion to the anticipated Mannich products was observed in both cases (these species were then identified as the desired Mannich products through MS/MS analysis, see Figure 7.8-7.11). However, only a 12% conversion to Mannich product **136**, or a significant amount of unanticipated by-product produced alongside Mannich product **137**, was observed in these experiments. These results suggest that significant optimisation would be required in order to make the modified Mannich ligation a viable strategy for bioconjugation.
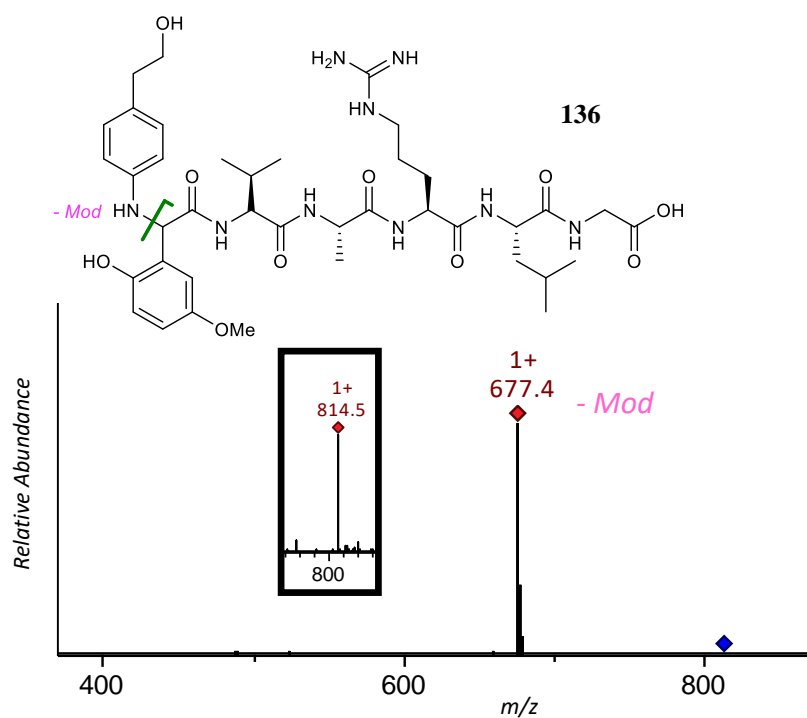
Figure 7.8 MS/MS data of Mannich product **136**.
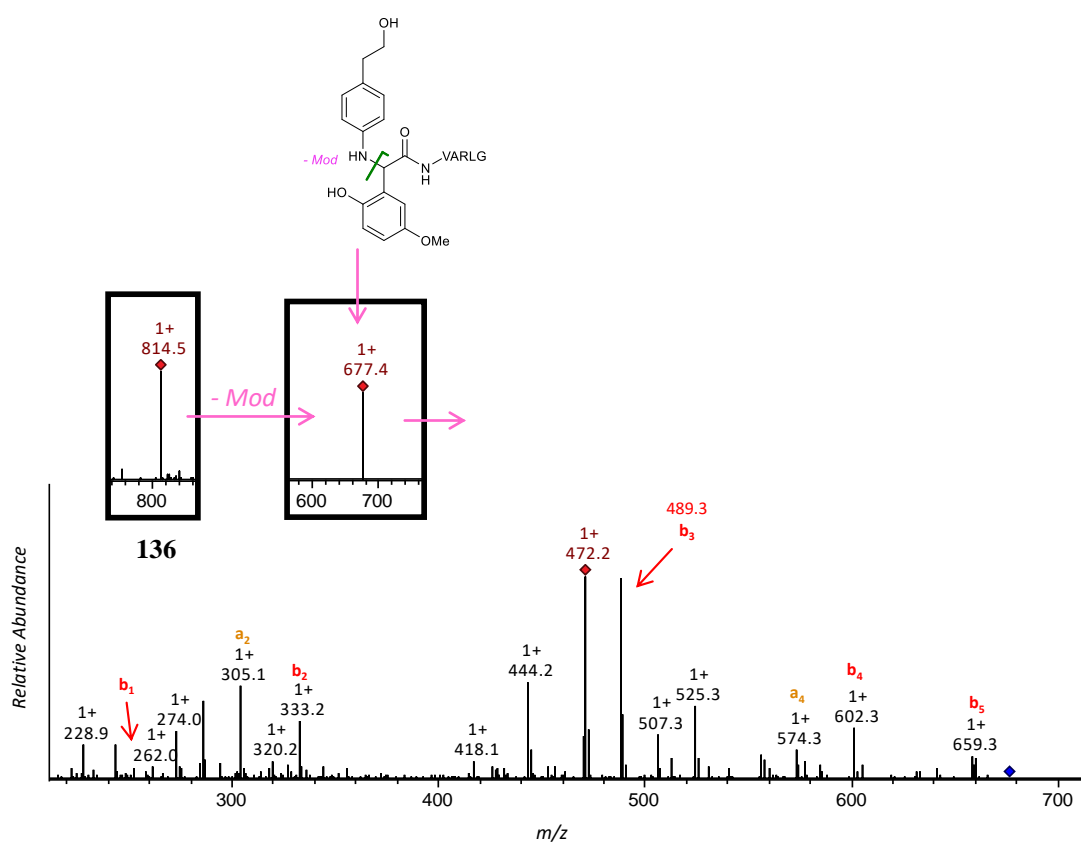
Figure 7.9 MS/MS data of fragmentation of Mannich product **136**, followed by fragmentation of the resulting major fragment (677.4 Da).
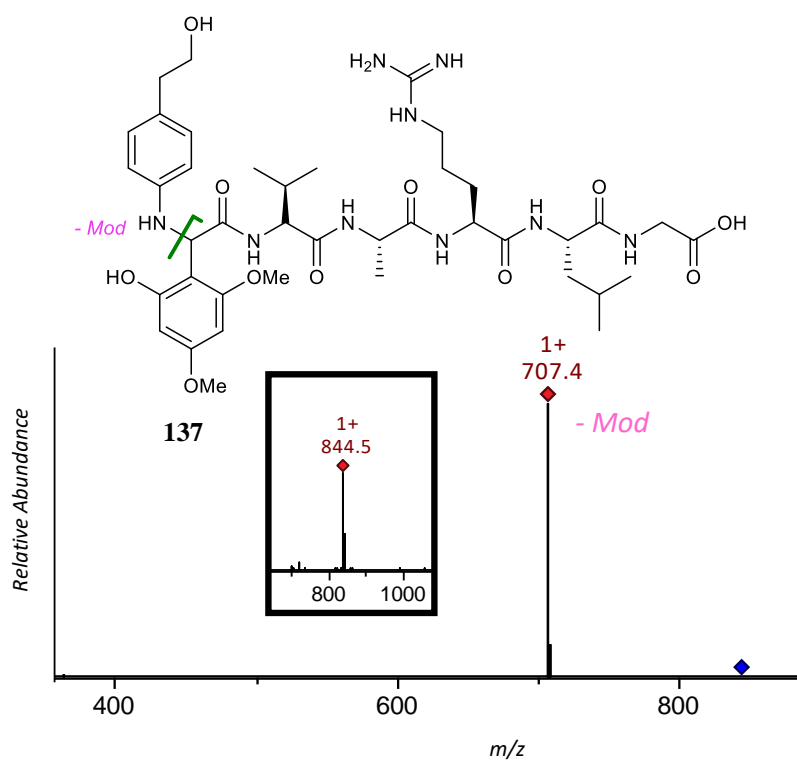
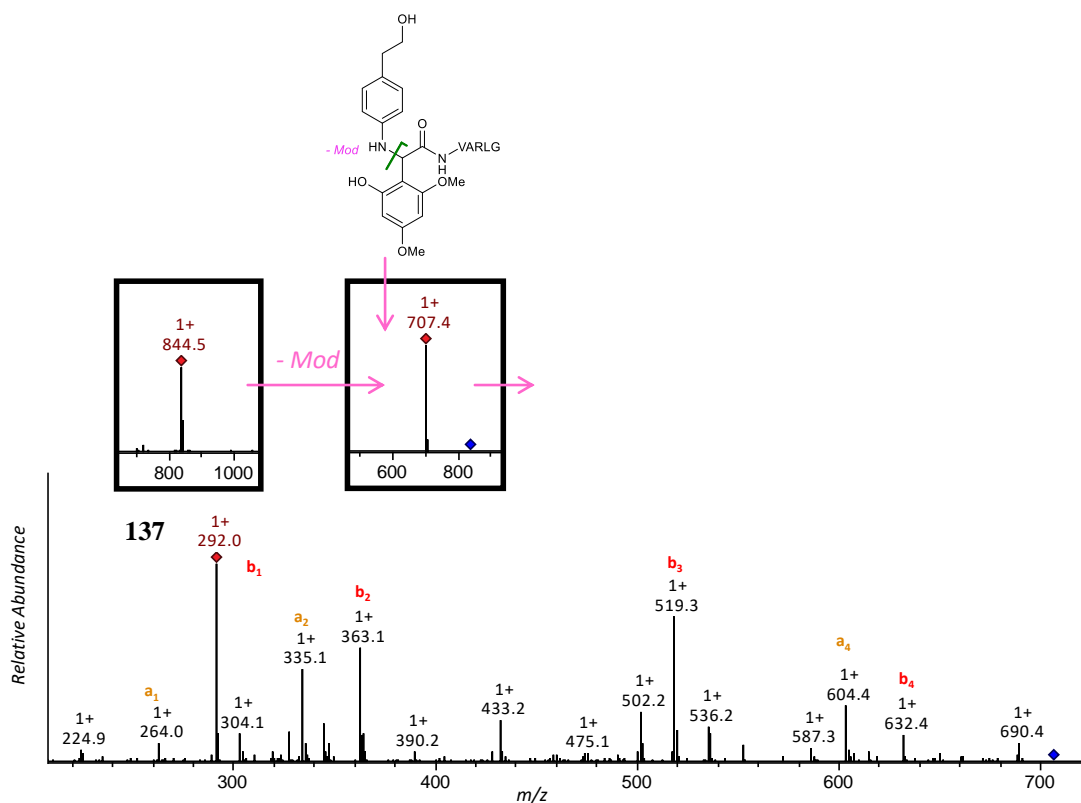**Figure 7.10** MS/MS data of Mannich product **137**.

Figure 7.11 MS/MS data of fragmentation of Mannich product **136**, followed by fragmentation of the resulting major fragment (707.4 Da).

7.2.1 A catalyst free aldol ligation

The results described in Figure 7.7 both show low conversions to the desired Mannich product, but for different reasons; for the reaction with phenol **134**, this was due to significant quantities of starting material remaining, but for the reaction with phenol **135**, this was due to the presence of significant quantities of an unanticipated species at 725.43 Da. Incidentally, the *m/z* value of this species corresponds to the mass of glyoxyl-VARLG **132** + the mass of phenol **135** (+1 Da for [M+H]$^+$), and MS/MS analysis of this species yields a fragmentation pattern similar that described for aldol-modified peptides (modification identified at the *N*-terminus, highly intense peak of species -18 Da, which upon fragmentation yields x/y/z fragments corresponding to the "VARLG" peptide in a similar manner to Mannich product **137**, see Figure 7.12 and Figure 7.13). Modifications of aldehydes in aqueous solvents with phenol reagents that act as enolate equivalents have previously been reported,[225] and it was therefore hypothesised that phenol **135** may be participating as a preformed enolate equivalent in an aldol-type direct conjugation with glyoxyl-VARLG **132**. Such a ligation would be of significant interest within the field of bioconjugation, as the procedure would lead to C-C bond formation between phenol and biomolecule. To validate this hypothesis, glyoxyl-VARLG **132** was incubated with different concentrations of phenol **135** in 0.1 M PB pH 6.5 for 24 h at 37 ºC, and conversion to the anticipated phenol-peptide product **138** was assessed by LC-MS analysis (Figure 7.14).
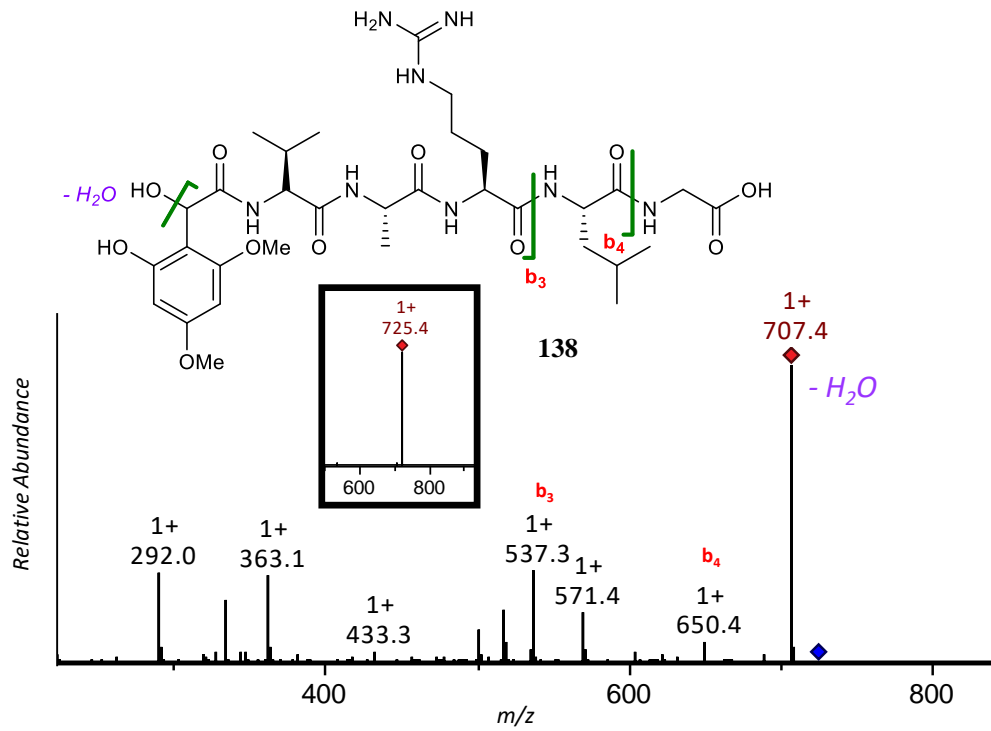
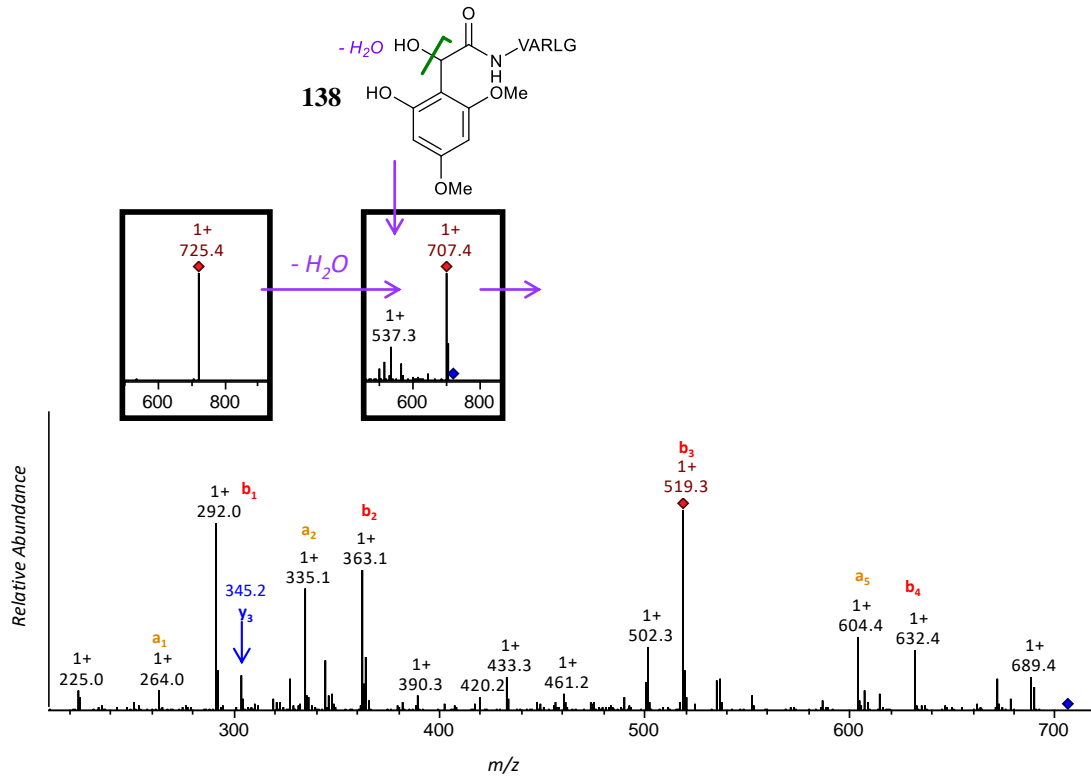**Figure 7.12** MS/MS data of hypothesised phenol-peptide product **138**.

Figure 7.13 MS/MS data of fragmentation of phenol-peptide product **138**, followed by fragmentation of the resulting major fragment (707.4 Da).
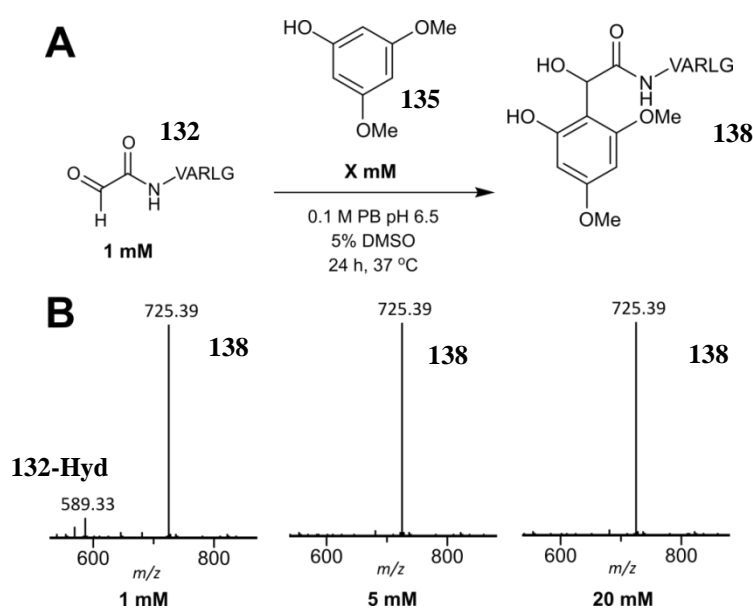
Figure 7.14 (A) Outline of reaction of glyoxyl-VARLG **132** with different concentrations of phenol **135**. Calculated [M+H]⁺ of anticipated phenol-peptide product = 725.38 Da. (B) LC-MS data obtained for each loading of phenol **138** after 24 h of reaction.

As shown in Figure 7.14, almost complete conversion (95%) to the hypothesised phenol-peptide product **138** can be achieved using only 1 molar equiv. of phenol **135**, with total conversion achieved when using 5 molar equiv. of phenol **135**.

7.2.2 Controlling reaction pathways through pH

The effect of pH of both the modified Mannich bioconjugation and the unanticipated phenol-mediated ligation was next investigated. The three-component Mannich bioconjugation has been reported to have an optimal reactivity observed at pH 5.5-6.5 and poor reactivity observed above pH 8.[223] It has been hypothesised that at pH 6.5 (the optimal pH) a cyclic transition state between the phenol and aldehyde-aniline imine can be formed, which simultaneously activates both the aldehyde-aniline imine and the phenol aromatic towards product formation. [223] On peptides however Mannich bioconjugations have also been reported to proceed over a much wider pH range, with reactions proceeding at a pH as low as 2 and as high as 10. [224] It was therefore hypothesised that both the modified Mannich ligation procedure and the unanticipated phenol-mediated ligation procedure could proceed at different pH values i.e. neutral pH. Therefore, the modified Mannich ligation between glyoxyl-VARLG **132**, aniline **133**, and phenol **135**, along with the phenol-mediated ligation between glyoxyl-VARLG **132** and phenol **135**, was performed in 0.1 M PB pH 7.5

for 24 h at 37 °C to test this hypothesis. Similar reactions were also performed in 0.1 M NaOAc pH 4.5 to provide a comparison of performing both reactions at neutral and acidic pH. Both reactions were analysed by LC-MS, as shown in Figure 7.15
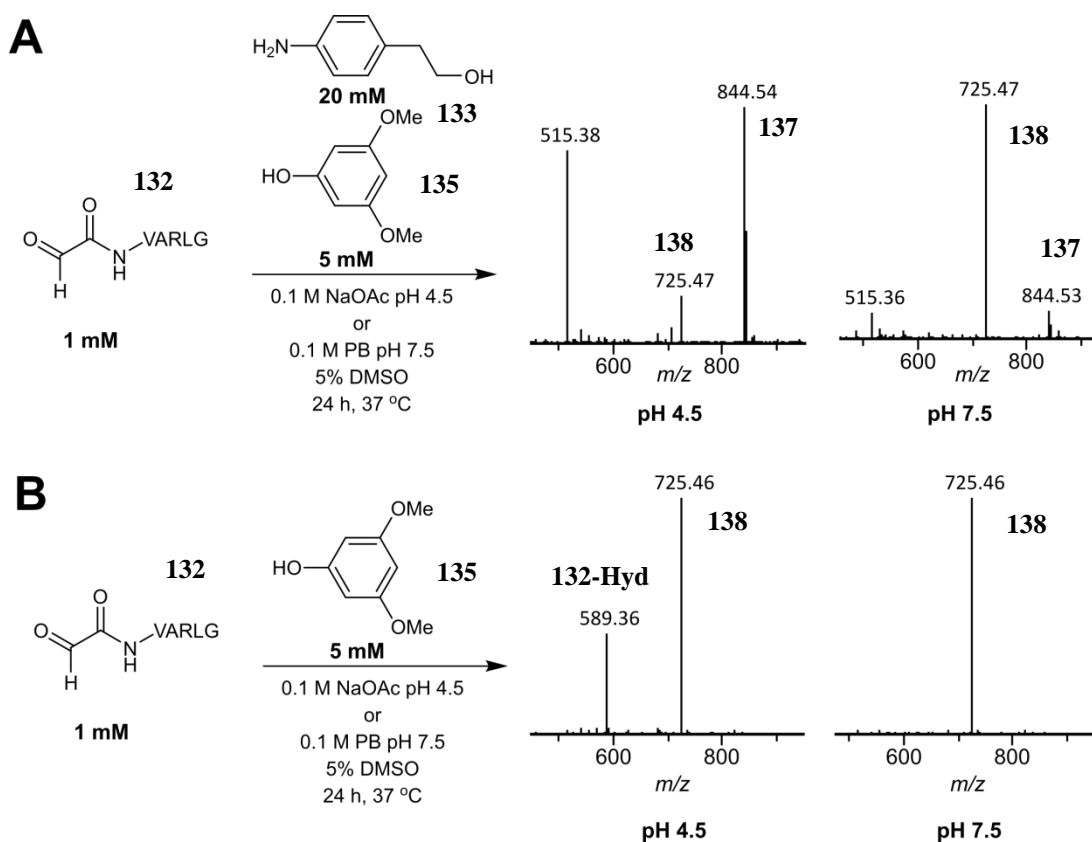

Figure 7.15 (A) Outline of reaction of glyoxyl-VARLG **132** with aniline **133** and phenol **135** at pH 4.5 or pH 7.5, with each outcome analysed by LC-MS. (B) Outline of reaction of glyoxyl-VARLG **132** with phenol **135** at pH 4.5 or pH 7.5, with each outcome analysed by LC-MS.

*NB: For the reaction of glyoxyl-VARLG **132** with aniline **133** and phenol **135** at pH 4.5, an intense peak at 515.38 Da was observed. It was unclear as to where this peak originated from – it was hypothesised that it could possibly be the peptide VARLG which may have originated from decomposition during the reaction or from fragmentation of product **137** during ESI-MS analysis.*

The results of performing the modified Mannich and phenol-mediated bioconjugation protocols varied significantly depending on if they were performed at pH 4.5 or pH 7.5. For the modified mannich bioconjugation protocol, a greater proportion of Mannich-peptide product **137** compared to phenol-peptide product **138** was observed at pH 4.5, with phenol-peptide product **138** dominating at pH 7.5. For the phenol-mediated bioconjugation protocol, a 60% conversion to phenol-peptide product **138** was observed at pH 4.5, whereas full conversion to phenol-peptide

product **138** was noted at pH 7.5. The results shown in Figure 7.15A suggest that for the modified mannich bioconjugation protocol, it would be possible to facilitate mannich bioconjugation (and avoid phenol bioconjugation by-products in this case) by performing the reaction at pH 4.5 as opposed to pH 6.5 or pH 7.5. Additionally, the results shown in Figure 7.15B suggest that the phenol-mediated bioconjugation strategy proceeds efficiently at pH 7.5

7.2.3 Dissecting the phenol-based ligation

In the interest of designing ligation strategies that operate at neutral pH, the phenol-mediated bioconjugation protocol was subjected to further investigation. As seen in Figure 7.8, phenol **134** and phenol **135** show significantly different reactivity towards ligation with glyoxyl-VARLG **132**. These results suggest that the choice of phenol reagent and substituent position appears to greatly impact whether or not modification of glyoxyl-VALRG **132** occurs. Therefore, to dissect the intricacies of the ligation, a series of small molecule aromatics were screened for their activity in the site-selective modification of glyoxyl-VARLG **132** (Figure 7.16). These compounds were chosen primarily to investigate the importance of the phenol OH group and the effects of substitution on the aromatic ring on the outcome of the phenol-based ligation.
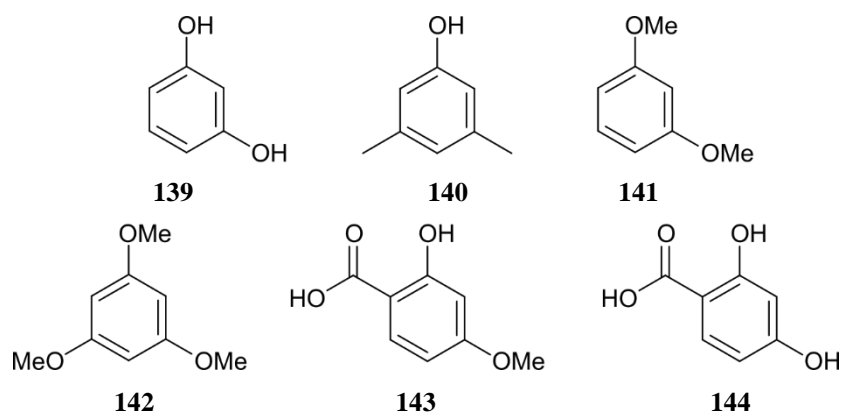
Figure 7.16 Aromatics screened for reactivity with glyoxyl-VARLG **132**.

Figure 7.17A outlines the reactions of 1 mM of glyoxyl-VARLG **132** in 0.1 M PB pH 7.5, 5% DMSO, at 37 °C for 24 h with 5 mM of aromatics **139-144**. Conversion to modified peptide species was then assessed by LC-MS analysis, with different aromatics displaying different reactivity towards glyoxyl-VARLG **132** (Figure

7.17B, green tick indicates reaction of aromatic with glyoxyl-VARLG **132**, whereas red cross indicates no reaction with glyoxyl-VARLG **132** was observed). Structures of the anticipated aromatic-peptide products **145-147** that were observed in the screen of aromatic compounds **139-144**, along with relevant LC-MS data, are shown in Figure 7.17C.
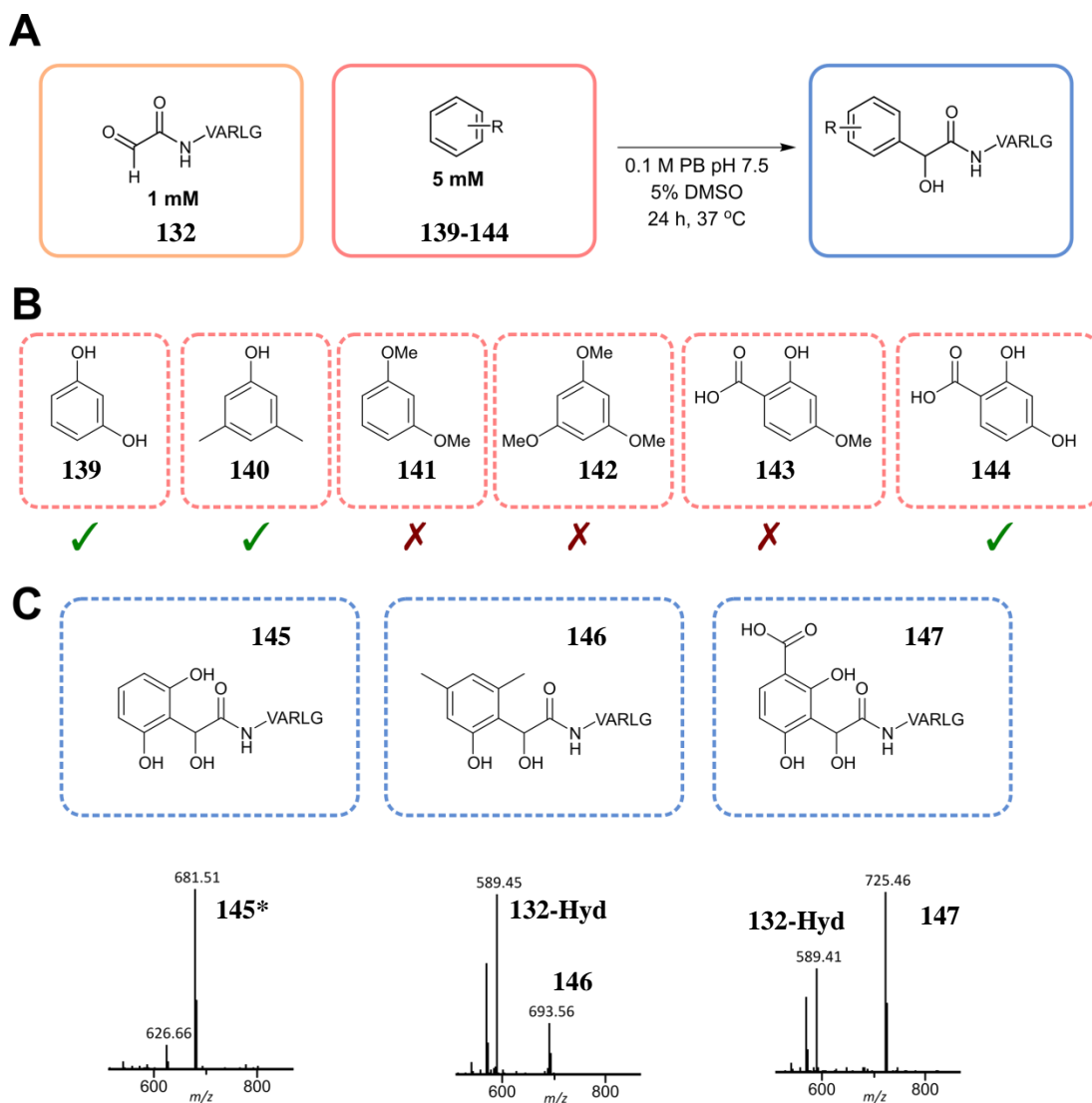
Figure 7.17 (A) Outline of reaction of glyoxyl-VARLG **132** with aromatics **139-144** to give anticipated modified peptide products. (B) Structure of aromatics **139-144** screened in (A). A green tick below the structure indicates that some degree of modification with glyoxyl-VARLG **132** occurred, whereas a red cross indicates that no reaction with glyoxyl-VARLG **132** was observed whatsoever. (C) Structures of anticipated aromatic peptide products **145**, **146**, and **147** observed when using aromatic **139**, **140** and **144** respectively. LC-MS data for these experiments is also shown. Calculated [M+H]$^+$ of **145** = 681.35 Da, calculated [M+H]$^+$ of **146** = 693.39 Da, calculated [M+H]$^+$ of **147** = 725.34 Da

*NB: For the reaction with aromatic **139** and aromatic **144**, it was hypothesised that multiple phenol-peptide products would be possible, as aromatic **139** and **144** have additional possible ortho positions (with respect to phenol OH) available for*

ligation. However, all the products generated would show up as the same mass when analysed by ESI-MS. It would therefore require an additional analytical technique (i.e. NMR) to determine the actual connectivity and product distribution seen for the peak at 681.51 Da in Figure 7.17C. Therefore, although the peaks at 681.51 Da and 725.46 Da have been labelled as product 145 and 147 respectively, in this instance, these peak likely comprises of a mixture of products.

As seen in Figure 7.17, successful modification of glyoxyl-VARLG **132** with each of aromatics **139-144** varied greatly depending on which compound was used. Total conversion was achieved when using resorcinol **139** (NB: a small unidentified peak at 626.66 Da was also observed), with a lower conversion of 49% observed when using phenol **144**, and a significantly lower conversion of 14% observed when using phenol **140**. No modification occurred when using aromatics **141-143**, with only unreacted glyoxyl-VARLG **132** observed in each case. The results of the screen of aromatic compounds **139-144** described in Figure 7.9 provides significant insights into the ligation strategy described thus far. Firstly, a phenol group was found to be absolutely vital for successful modification, as using aromatics **141** and **142**, which do not contain phenol groups, lead to no modification of glyoxyl-VARLG **132**. Substitution of the aromatic ring was also found to be of major importance in dictating the outcome of the ligation; phenols bearing strongly electron donating groups (such as phenol **135**, phenol **139**, and phenol **139**) were found to give high conversions to the desired phenol-peptide product, whereas phenols bearing weakly electron donating groups (such as phenol **140**) or those containing electron withdrawing groups (such as phenol **143** and phenol **144**) were found to give low conversions to the desired product. Finally, the presence of a carboxylic acid group *ortho* to the hydroxyl group of the phenol was found to have significant effects on the outcome of the ligation, with only a 49% conversion to the desired product observed when using phenol **144**, and no conversion observed when using phenol **143**. These results were deemed particularly interesting, as both compounds have the required phenol group and both contain a strongly electron donating group in the *meta* position, but only conversion with phenol **144** was achieved. Phenols that contain *ortho* carboxylic acid groups are known to participate in intramolecular hydrogen bonding. Given that the phenol group is vital for successful ligation, it was hypothesised that the presence of an *ortho* hydrogen bonding group to the phenol

group may lead to "consumption" of the phenol moiety, preventing it from participating as an enol equivalent in aldol-type ligations. This would explain why phenol **143** (which has only one phenol group and can theoretically participate in intramolecular hydrogen bonding to the adjacent carboxylic acid group) was unreactive towards glyoxyl-VARLG **132** but why phenol **144** (which has two phenol groups present, one of which will not participate in intramolecular hydrogen bonding) was reactive towards glyoxyl-VARLG **132**.

7.2.4 Hypothesised mechanism for the catalyst-free aldol bioconjugation

The use of phenols as enol/enolate equivalent has been previously documented in the literature. For example, one of the most well-known examples of phenols acting as enolate equivalents is in the Kolbe-Schmitt carboxylation, which can be utilised to synthesise salicylic acids.[226,227] Phenols as enolate equivalents have also been reported in the modification of aldehydes.[225] It was therefore hypothesised that the phenol reagents described thus far would also operate *via* a similar mechanism in the modification of α-oxo aldehydes as shown in Figure 7.18. This mechanism would explain the absolute requirement of a phenol for successful conjugation, and also explain improved conjugation yields obtained at pH 7.5 compared to pH 4.5 (higher pH = higher proportion of deprotonated phenol, thereby increasing the active content of phenol).
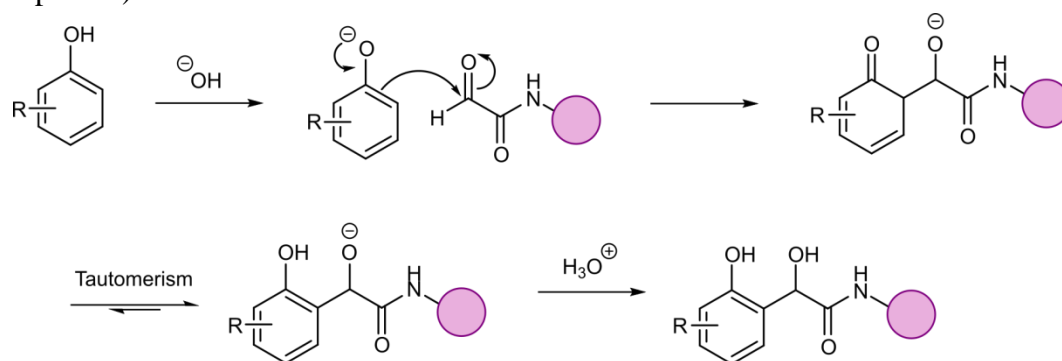
Figure 7.18 Proposed mechanism for the phenol-mediated ligation of α-oxo aldehydes. Following deprotonation, phenol acts as an enolate equivalent and participates in nucleophilic attack of the α-oxo aldehyde. Tautomerisation followed by protonation generates the final phenol product.

7.2.5 Towards catalyst free protein bioconjugation

Finally, to begin to demonstrate the aforementioned phenol ligation was compatible with site-selective protein bioconjugation, the reaction between 75 μM of glyoxyl-myoglobin **13** and 1 mM of phenol **135** in 0.1 M PB pH 7.5, 10% DMSO, at 37 °C

for 1 h was performed as outlined in Figure 7.19A. Conversion to the anticipated modified myoglobin **148** was assessed as judged by LC-MS (Figure 7.19B).
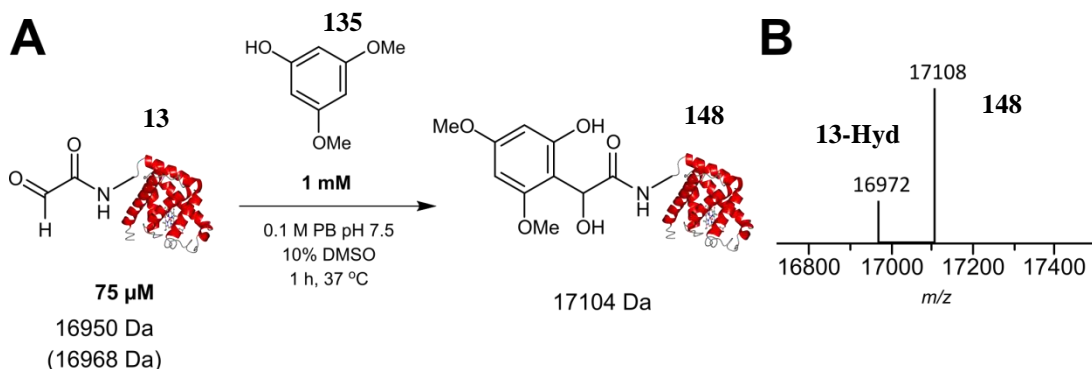
Figure 7.19 (A) Outline of reaction between glyoxyl-myoglobin **13** and phenol **135** (B) LC-MS data obtained for reaction of 75 µM **13** and 1 mM phenol **135** in 0.1M PB pH 7.5 (10% DMSO) at 37 °C for 1 h.

As shown in Figure 7.19, a 79% conversion to the anticipated phenol-myoglobin product **148** was observed. This result suggests that suitable phenol reagents such as **135** can be used to in protein bioconjugation. Further studies, such as UV/Vis analysis and trypsin digestion experiments, would be required to fully characterise this product, and determine the suitability of using phenol reagents in this manner for site-selective protein bioconjugation. Additionally, one could anticipate that by using a suitably functionalised phenol compound, one could enable highly efficient, catalyst free, site-selective protein bioconjugation of protein aldehydes in a similar manner to the OPAL. This phenol-induced ligation of protein aldehydes (or 'PHILIPA' for short) could provide an efficient, straightforward way to site-selectively modify protein aldehydes *via* a hydrolytically stable C-C bond at neutral pH without the need for large concentrations of small molecule organocatalyst (Figure 7.20).
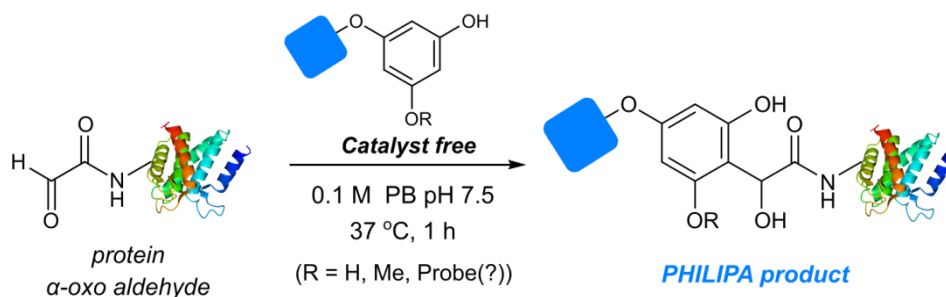
Figure 7.20 Proposal of a novel protein bioconjugation strategy for the site-selective modification of protein α-oxo aldehydes using phenols at neutral pH to generate C-C linked bioconjugates.

7.3 Concluding statement

In conclusion, the work described in this thesis has led to the development of a novel protein bioconjugation strategy, the Organocatalyst-mediated Protein Aldol Ligation (abbreviated to "OPAL"). Following its serendipitous discovery, the OPAL was validated on model peptide and protein systems, and optimised through rigorous kinetic analysis of different small molecule organocatalysts and aldehyde donors. Throughout this project, the OPAL has proved to be a robust, versatile, and powerful C-C bond forming bioconjugation strategy, which offers site-selective modification of a range of protein aldehydes at both internal and *N*-terminal positions. Alongside palladium-mediated decaging of thiazolidine-bearing proteins, compatibility and selectivity of the OPAL was demonstrated through selective labelling and protein pulldown of GFP within cell lysates, which maintains its fluorescence as a result of modification and purification at neutral pH. In addition, the newly generated β-hydroxy aldehyde handle that results from OPAL modification was found to be amenable to site-selective modification via a different protein aldehyde bioconjugation strategy. Screening of aniline-catalysed oxime ligations on aldol-modified peptides revealed an unexpected switch in pH dependence, in which neutral pH was favoured over acidic pH. This information was then used in the synthesis of dually modified protein bioconjugates *via* an OPAL-oxime at neutral pH. The OPAL was then applied in the chemical mimicry of *N*-terminal myristoylation of the Leishmania protein HASPA, and was demonstrated to mimc structural and solution based properties of its enzymatically *N*-myristoylated counterpart *via* characterisation by protein NMR. The tandem OPAL-oxime strategy was also utilised to generate a dually lipidated HASPA species previously inaccessible through conventional enzymatic methods, which was further used to explore the effects of HASPA lipidation and membrane binding. Finally, an aniline-catalysed retro-aldol mediated process for the controlled breakdown of OPAL products, along with a 'catalyst free' aldol-type bioconjugation, was discussed. It is hoped the work described in this thesis will further contribute towards the bioconjugation toolbox in chemical biology, assist in the study of lipidated proteins such as HASPs, and facilitate research towards the discovery of new organocatalyst mediated and catalyst free bioconjugation strategies for wide ranging applications in chemical biology.

Chapter 8: Experimental

General procedures and materials

Unless otherwise specified, all chemical reagents were obtained from commercial sources and used without further purification. Myoglobin from equine heart (M1882) and Thioredoxin from *Escherichia coli* (T0910) were purchased from Sigma and used without further purification. Sequencing Grade Modified Trypsin V5111 was purchased from Promega. Pierce™ Monomeric Avidin Agarose was purchased from ThermoFisher Scientific.

Small molecule NMR spectroscopy

^{1}H and ^{13}C NMR spectra were measured using either Jeol 400-MR or Bruker 500-MR spectrometer at The University York Centre for Magnetic Resonance at 400/500 MHz for ^{1}H and 100/125 MHz for ^{13}C with Me$_4$Si as the internal standard. Multiplicities are given as singlet (s), broad singlet (br s), doublet (d), doublet of doublets (dd), triplet (t), quartet (q), pentet (p) or multiplet (m). Resonances were assigned using HH-COSY and CH-HSQC. All NMR chemical shifts (δ) were recorded in ppm and coupling constants (J) are reported in Hz. ACD-NMR Processor Academia Edition and MestReNova were used for processing spectral data.

UV/Vis analysis

UV/Vis analysis was performed using either a U1900 spectrometer (HITACHI) in line with UV solutions 2.2 software (HITACHI), or using a DS-11 FX+ spectrophotometer/fluoremter (DeNovix) in line with DS-11/DS-11 FX Software (Version 3.15).

High performance liquid chromatography instrumentation

Analytical HPLC was performed using either an Accucore C18 coulmn 2.6 μm column, 2.1 × 150 mm, or a Phenomenex Kinetex 5μ phenyl-hexyl 100 A column of dimension 250 ^{x} 4.6 mm. Shimadzu Prominence LC-20AD pump and SDP-M20A Diode Array detector were used during the analysis.

Liquid Chromatography Mass Spectrometry instrumentation

High Performance Liquid Chromatography-Electrospray Ionisation Mass Spectrometry (LC-MS) was accomplished using a Dionex UltiMate® 3000 LC

system (ThermoScientific) equipped with an UltiMate® 3000 Diode Array Detector (probing 250-400 nm) in line with a Bruker HCTultra ETD II system (Bruker Daltonics), using Chromeleon® 6.80 SR12 software (ThermoScientific), Compass 1.3 for esquire HCT Build 581.3, esquireControl version 6.2, Build 62.24 software (Bruker Daltonics), and Bruker compass HyStar 3.2-SR2, HyStar version 3.2, Build 44 software (Bruker Daltonics) at The University York Centre of Excellence in Mass Spectrometry (CoEMS). All mass spectrometry was conducted in positive ion mode unless stated otherwise. Data analysis was performed using ESI Compass 1.3 DataAnalysis, Version 4.1 software (Bruker Daltonics).

LC-MS analysis of peptide and protein ligations

Prior to analysis by LC-MS, peptide or protein ligation mixture was diluted 1:3 in water and then further diluted 1:1 in acetonitrile with 1 % (v/v) formic acid. Peptide samples were analysed using an Accucore™ C18 2.6 μm column (50 x 2.1 mm) (ThermoScientific). Water with 0.1 % (v/v) formic acid (solvent A) and acetonitrile with 0.1 % (v/v) formic acid (solvent B) were used as the mobile phase at a flow rate of 0.3 mL/min at room temperature (RT). A multi-step gradient of 6.5 min was programmed as follows: 90% A for 0.5 min, followed by a linear gradient to 95% B over 3.5 min, followed by 95% B for an additional 0.5 min. A linear gradient to 95% A was used to re-equilibrate the column Under these conditions all peptides typically eluted between 2-5 min**.** Protein samples were analysed without the use of a column at RT. Water with 0.1 % (v/v) formic acid (solvent A) and acetonitrile with 0.1 % (v/v) formic acid (solvent B) were used as the mobile phase at a 1:1 ratio over the course of 3 min as follows: 0.05 mL/min to 0.25 mL/min for 1 min, 0.025 mL/min for 1 min, followed by 1.0 mL/min for 1 min. Under these conditions, all proteins typically eluted between 0.1-1.5 min.

Green fluorescent protein and Superfolder green fluorescent protein Mass Spectrometry instrumentation

Electrospray ionisation mass spectrometry (ESI-MS) of samples relating to green fluorescent protein (GFP) or superfolder green fluorescent protein (sfGFP) were obtained using a solariX XR FTMS 9.4T (Bruker) using ftms Control, ftmsControl 2.1.0 Build: 98 software (Bruker Daltonics) at The University York Centre of Excellence in Mass Spectrometry (CoEMS). All mass spectrometry was conducted

in positive ion mode unless stated otherwise. Data analysis was performed using ESI Compass 1.3 DataAnalysis Version 4.1 software.

Green fluorescent protein and Superfolder green fluorescent protein Mass Speectrometry analysis

Prior to analysis, samples containing GFP or sfGFP were desalted using either a PD SpinTrap G25 column (GE Healthcare Life Sciences), or using PD MiniTrap G25 column (GE Healthcare Life Sciences), eluting with water. The desalted protein sample (50 µL) was then diluted by addition of 50 µL of a 1:1 solution of water:acetonitrile 1% (v/v) formic acid for analysis.

Analysis of trypsin digest samples

Tryptic digestion samples were analysed using a Symmetry® C18 5µm 3.0 x 150 mm reverse-phase column (Waters). Water with 0.1% (v/v) formic acid (solvent A), and acetonitrile with 0.1% (v/v) formic acid (solvent B) were used as the mobile phase at a flow rate of 0.08 ml/min at 40 $^{\circ}$ C. A multi-step gradient of 45 min was programmed as follows: 95% A for 0.1 min, followed by a linear gradient to 80% B over 40 min, followed by a linear gradient to 95% B for 1 min, followed by a linear gradient to 95% A for 4 min.

Analysis of conjugation yields v1

Conversion from the designated starting material to the desired material (conjugation yields, %) was calculated by analysing the peak intensities of starting material and product species, and using Equation 1:

$$\frac{Product\ peak\ intensity}{Starting\ material\ peak\ intensities + Product\ peak\ intensity}\ x\ 100 = \%\ conversion \quad \text{Equ. 1}$$

Analysis of conjugation yields v2

Conversion from the designated starting material to the desired material (conjugation yields, %) was calculated by analysing the peak area of the starting material and product species, and using Equation 2:

$$\frac{Peak\ area\ of\ product}{Peak\ area\ of\ starting\ material + Peak\ area\ of\ product}\ x\ 100 = \%\ conversion \qquad \text{Equ. 2}$$

Kinetic studies

Kinetic data was obtained from reactions performed on a 20 µL scale using an adapted LC-MS method.[42] Reactions were quenched by addition of 80 µL 1:1 H_2O:MeCN (1% formic acid) at time points 2 min, 5 min, 10 min, 20 min, and 30 min, and then analysed by LC-MS. Reaction yields at each time point were calculated and the second-order rate constants (k_2) were determined by fitting the data to the following equation 3:

$$y = \frac{\left(ln\frac{[A]_0[D]_t}{[A]_t[D]_0}\right)}{([D_0]-[A]_0)} = k_2t \qquad \text{Equ. 3}$$

where $[A]_0$ and $[D]_0$ are the initial concentrations of the acceptor (peptide aldehyde) and donor (small molecule aldehyde) respectively, and $[A]_t$ and $[D]_t$ are the concentrations of the acceptor and donor at time t. Full details on the conditions used for each experiment can be found in the **Kinetic Data** section

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis

For expression, purification, site-selective modification experiments, and liposome assay experiments relating to hydrophilic surface acylated protein A (HASPA), green fluorescent protein (GFP) and superfolder green fluorescent protein (sfGFP),

all SDS-PAGE analysis was performed using 12% or 4-20% gradient polyacrylamide gels[228]. For experiments relating to dually modified proteins, all SDS-PAGE analysis was performed using 15% acrylamide gels. Samples were reduced by boiling for 5-10 min (2% SDS, 2mM 2-mercaptoethanol, 4% glycerol, 40mM Tris-HCl pH 6.8, 0.01% bromophenol blue). Molecular weight markers used were either PageRuler Plus Prestained Protein Ladder (ThermoScientific) or SDS-PAGE Molecular Weight Standards, Low Range (Bio-Rad). Each gel was run at 200 volts for 45-80 min.

Coomassie stain

For Coomassie stain experiments, the gel was washed with fixing solution (40% MeOH, 10% AcOH), stained with 0.1% Coomassie Brilliant Blue R-250 (50% MeOH, 10% AcOH), and finally destained with solution (50% MeOH, 10% AcOH). Images of the resulting gels were captured and analysed using a Syngene G:BOX Chemi XRQ equipped with a Synoptics 4.0 MP camera, with GeneSys software (Version 1.5.7.0).

Fluorescent imaging

For fluorescent imaging of fluorescently modified proteins, the SDS PAGE gel was washed with fixing solution (40% MeOH, 10% AcOH). Visualisation of protein fluorescence, and images of the resulting gels, were captured using a Syngene G:BOX Chemi XRQ equipped with a Synoptics 4.0 MP camera in line with GeneSys software (Version 1.5.7.0).

Western Blot analysis

For western blot analysis of biotinylated protein samples (12 µg) were run on 15% SDS–PAGE and transferred onto a nitrocellulose membrane filter (0.45 µm, Amersham Protran Sandwich, GE Healthcare) using an electroblot apparatus (Bio-Rad, Hercules, CA) at 100V, 350 mA for 1 h in cooled transfer buffer (25 mM Tris–

HCl pH 8.3, 192 mM glycine, 0.1% SDS, 20% (v/v) methanol). The membrane was incubated in blocking solution (Phosphate buffered saline (PBS) tablets, Sigma)) containing 5% non-fat dry milk powder for 16 hours at 4 °C. The membrane was processed through sequential incubations with primary antibody, alkaline phosphatase anti-biotin (goat, Vector Labs, CA) 1:1000 dilution in PBS for 1 hour at room temperature, followed by washing in PBS, 0.01% Tween-20, and then incubation with visualising substrate BCIP/NBT Alkaline Phosphatase Substrate Kit (Vector Labs, CA) until immunoreactive proteins on the membrane were visible. The reaction was stopped by washing the membrane in distilled water. The membranes were imaged using a Syngene G:BOX Chemi XRQ equipped with a Synoptics 4.0 MP camera, with GeneSys software (Version 1.5.7.0).

Procedure for trypsin digestion

A 100 µL solution of OPAL product **60** with a total protein content of 1 mg was prepared. The product was analysed by MS before being subjected to the trypsin digest. The 100 µL solution of OPAL product **60** was dialysed into 50 mM Tris-HCl, pH 8.0. 36 mg of solid urea was then added to the solution, giving a final concentration of 6M Urea. To this solution was added DTT (5 µL of a 200 mM solution in 50 mM Tris-HCl, pH 8.0). The mixture was allowed to stand at room temperature for 1 h. The solution was then charged with iodoacetamide (20 µL of a 200 mM solution in 50 mM Tris-HCl, pH 8.0), gently vortexted, and allowed to stand at room temperature in the dark for 1 h. DTT (20 µL of a 200 mM solution in 50 mM Tris-HCl, pH 8.0) was then added to consume any unreacted iodoacetamide, and the solution was allowed to stand at room temperature in the dark for 1 h. 775 µL of a 50 mM Tris-HCl, 1 mM $CaCl_2$ (pH 7.6) was then added to reduce the urea concentration to >0.6 M. Trypsin solution (0.2 µg µL^{-1}, 100 µL in resuspension buffer, 50 mM acetic acid) was then added. The reaction mixture was gently vortexted and incubated for 16 h at 37 °C. *(NB: As much of this procedure as possible was performed in a laminar flowhood).* To stop the trypsin digest procedure, 1 µL of formic acid was added to bring the pH of the solution to pH 3-4 (checked by pH paper). A 50 µL aliquot of this solution was analysed directly by LC-MS. The remaining solution was stored in a -20 °C freezer and defrosted if more samples were required.

Sequence of proteins used

*NB: The initiatator methionine for all proteins has not been included in the protein sequences, as this is not present in the final full length protein. For proteins bearing an unnatural amino acid, the unnatural amino acid has been assigned as **X**.*

Horse heart myoglobin 28

GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKT
EAEMKASEDLKKHGTVVLTALGGILKKKGHHEAELKPLAQSHATKHKIPIK
YLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG

CTB 30

TPQNITDLCAEYHNTQIHTLNDKIFSYTESLAGKREMAIITFKNGATFQVEVP
GSQHIDSQKKAIERMKDTLRIAYLTEAKVEKLCVWNNKTPHAIAAISMAN

Thioredoxin 31

SDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGK
LTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLD
ANLA

GFP (Y39CycloOctK) 33

SYKDDDDKVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDAT**X**GKLT
LKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQ
ERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYN
SHNVYIMADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPD
NHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKHHHHHH

HASPA 35

GAYSTKDSAKEPQKRADNIDTTTRSDEKDGIHVQESAGPVQENFGDAQEKN
EDGHNVGDGANGNEDGNDDQPKEHAAGNLEHHHHHH

HASPA(G1S) 37

SAYSTKDSAKEPQKRADNIDTTTRSDEKDGIHVQESAGPVQENFGDAQEKN
EDGHNVGDGANGNEDGNDDQPKEHAAGNLEHHHHHH

GFP(Y39ThzK) 40

SYKDDDDKVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDAT**X**GKLT
LKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQ
ERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYN
SHNVYIMADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPD
NHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKHHHHHH

sfGFP(N150K) 39

VSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGK
LPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDD
GTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSH**X**VYITAD
KQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSV
LSKDPNEKRDHMVLLEFVTAAGITHGMDELYKGSHHHHHH

CTB 30 preparation

The protocol for obtaining this protein is adapted from a published protocol. [77] A stab from a stock of *Escherichia coli* BL21 DE23 containing a pSAB2.2 plasmid (kindly provided by Dr Thomas Branson). encoding CTB **29** was used to inoculate LB growth media (5 mL) containing ampicillin (100 μg mL-1) and this starter culture was incubated at 37 °C for 20 h. Starter culture (1 mL) was added to LB growth media (4 x 1 L) which was then incubated at 37 °C and monitored until the OD600 reached ca. 0.5 before IPTG (1 M final concentration) was added to each flask in induce protein over-expression. The incubation was continued for 24 h at 30 °C before isolating the cells by centrifugation at 10000 x g for 10 min. The bacterial cell pellet was discarded and the supernatant was retained. Solid ammonium sulfate was added to the supernatant (final concentration 60% w/v) and stirred for 1 h. The solution was then centrifuged at 17000 x g for 25 min and the supernatant discarded. The protein pellet was resuspended in 0.1 M phosphate buffer and centrifuged at 15000 x g for 8 mins to remove any insoluble material. The supernatant was filtered through a 0.45 μm filter (Sartorius Minisart) and loaded onto 2 x 5 mL HisTrap FF

Ni columns. Protein purification was then performed by first washing the Ni column with 0.1 M phosphate buffer containing 20 mM imidazole (pH 7, dubbed Buffer A) and then eluting bound protein with 0.1 M phosphate buffer containing 500 mM imidazole (pH 7, dubbed Buffer B) over a gradient of 0-100% Buffer B over 30 min using an AKTA Start (GE Healthcare Life Technologies). Protein presence in collected fractions was monitored by UV spectroscopy at 280 nm. Fractions thought to contain the desired protein were first subjected to SDS-PAGE analysis.
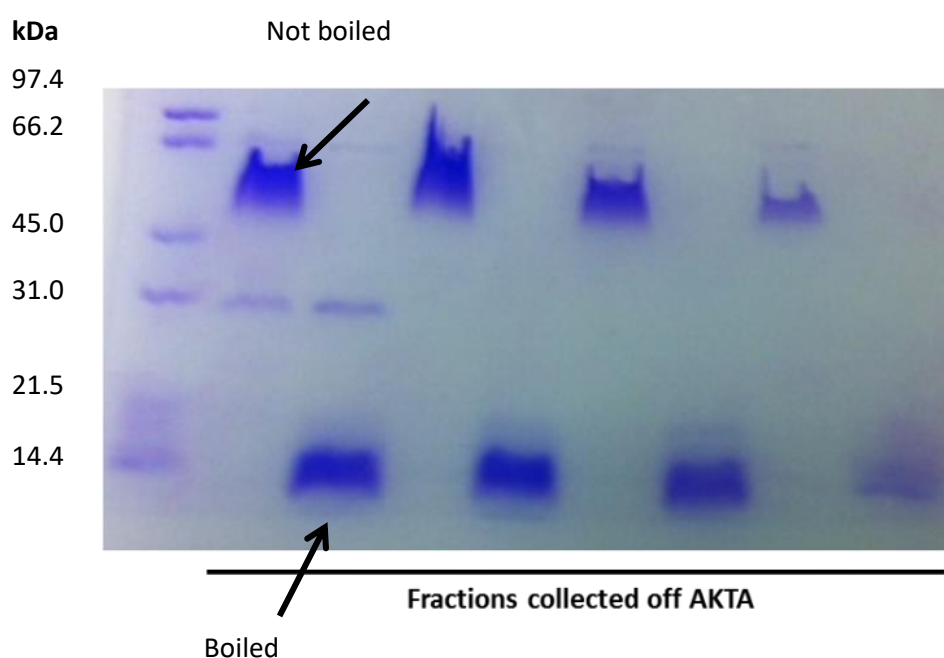
NB: CTB exists as a stable pentamer (ca. 58 kDa) which breaks down into monomers (ca. 12 kDa) if boiled or subjected to conditions used in the LC MS

Fractions containing the desired, sufficiently pure protein were then pooled and buffer exchanged into 0.1 M phosphate buffer via dialysis (using SnakeSkin Dialysis Tubing, Life Technologies). Concentration of CTB **29** was measured by UV spectroscopy at 280 nm ($\varepsilon = 11585$ mol^{-1} dm^3 cm^{-1}, taken from reference 229).

Synthesis of small molecules

Benzamidoxime 105:
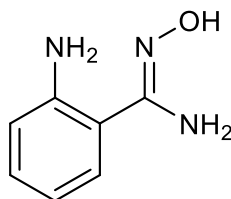
Synthesis of benzamidoxime **105** was performed as previously reported[132]. To the ethanolic solution (40 mL) of 2-amino benzonitrile (2 g, 17 mmol) and hydroxylamine hydrochloride (1.3 g, 18.7 mmol), aqueous $NaHCO_3$ (1.71 g, 20.4 mmol) solution (12 mL) was added. The mixture was refluxed overnight, allowed to cool to room temperature, and diluted with 40 mL ethanol. The solid was filtered off and washed with cold ethanol (2 x 10 mL). All ethanol fractions were pooled and concentrated *in vacuo*. The crude solution was then purified by flash chromatography (DCM:MeOH 95:5) to give the pure benzamidoxime **105** as a light orange, flaky solid (1.7 g, 65%). **¹H NMR** (400 MHz, DMSO-d_6): δ 9.57 (s, 1H), 7.37-7.34 (d, J = 7.79, 1H), 7.04-6.99 (t, J = 9.62 1H), 6.67-6.4 (d, J = 8.24 1H), 6.54-6.50 (t, J = 6.87 1H), 6.21 (br, 2H), 5.72 (br, 2H). **¹³C NMR** (100 MHz, DMSO-d_6): 152.88, 146.79, 129.00, 127.29, 115.46, 114.85, 114.19. **ESI-HRMS**: Found [M+H]⁺ 152.0817, $C_7H_{10}N_3O$, requires 152.0818.

Palmitoyl phthalimide 149
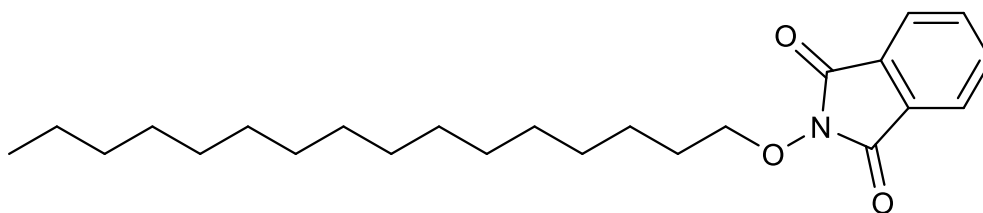
Synthesis of this compound was adapted from a protocol previously described for the synthesis of alkyl phthalimides[230]. To a stirred solution of cetyl alcohol (1.19 g, 4.90 mmol), N-hydroxyphthalimide (0.96 g, 5.88 mmol), and PPh_3 (1.70 g, 6.47 mmol) in THF (17 mL) was added DIAD (1.27 mL, 5.88 mmol). The solution was stirred under a nitrogen atmosphere at room temperature overnight. The solvent was removed in vacuo and the resultant white powder was dissolved in hexane and filtered. The solvent was removed in vacuo to give compound **131** as a white powder (0.61 g, 32%). **^{1}H NMR** (400 MHz, $CDCl_3$): δ 7.83 , 7.76 (2d, 4H, J = 3.1 Hz, ArH), 4.20 (t, 2H, J = 6.8 Hz, OCH_2), 1.79 (p, 2H, J = 6.9 Hz, CH_2), 1.48 (p, 2H, J = 6.9 Hz, CH_2), 1.26 (bs, 24H, CH_2), 0.88 (t, 1H, J = 6.9 Hz, CH_3). **^{13}C NMR** (125 MHz, $CDCl_3$): δ 163.67 (2) (C=O), 134.39 (2), 128.95 (2), 123.44 (2) (ArC), 78.63 (OCH_2), 31.90, 29.65, 29.63, 29.61, 29.54, 29.47, 29.34, 29.30, 28.12, 25.50, 22.67 (all CH_2), 14.1 (CH_3). **ESI-HRMS:** Found $[M+Na]^+$ 410.2662, $C_{24}H_{37}NNaO_3$, requires 410.2666.

Palmitoyl aminooxy 128
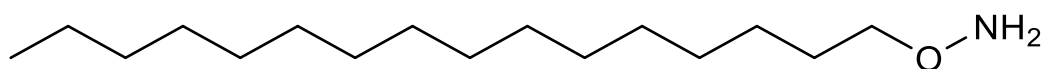
To a solution of **149** (0.039 g, 0.151 mmol) in DCM (1 mL) was added hydrazine monohydrate (76 μl, 2.47 mmol). The solution was stirred vigorously for 45 minutes, during which time a white solid appeared. The solution was filtered through cotton wool, and the filtrate was collected. The resulting filtrate then concentrated under a stream of nitrogen to give **128** as a white solid in quantitative yield that was used without further purification. **^{1}H NMR** (500 MHz, $CDCl_3$): δ 3.64 (t, 2H, J = 6.7 Hz, OCH_2), 1.56 (p, 2H, J = 6.8 Hz, CH_2), 1.25 (bs, 26H, CH_2), 0.87 (t, 1H, J = 6.9 Hz, CH_3). **^{13}C NMR** (125 MHz, $CDCl_3$): δ 76.39 (OCH_2), 32.07, 29.83 (2), 29.80 (2), 29.73 (2), 29.66, 28.55, 26.15, 22.83 (all CH_2), 14.25 (CH_3). **ESI-HRMS:** Found $[M+H]+$ 258.2793, $C_{16}H_{36}NO$, requires 258.2791

PEG2K phthalimide 150

(2kDa)
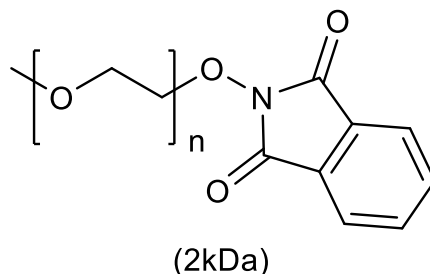
Synthesis of this compound was performed as previously reported[204]. Under an atmosphere of nitrogen, a solution of poly(ethylene glycol) monomethyl ether, average molecular weight 2000 g mol^{-1} (2.00 g, 0.994 mmol), *N*-hydroxyphthlimide (194 mg, 1.19 mmol), and PPh$_3$ (312 mg, 1.19 mmol) in DCM (10 mL) was charged with diisopropyl azodicarboxylate (212 µL, 1.09 mmol) via dropwise addition. The reaction mixture was then allowed to stir under nitrogen for 18 h at room temperature. The solution was then directly added to 400 mL of diethyl ether, and the suspension was stirred vigorously for 20 min. The suspension was filtered, and the resulting solid was washed with diethyl ether (3 x 70 mL), and residual solvent was removed *in vacuo*. The dry solid was then subjected to the same procedure a second time to give the product as a white powder that was used without further purification (1.5 g, 75%). **^{1}H NMR** (400 MHz, CDCl$_3$): δ 7.82, 7.75 (2d, 4H, *J* = 3.1 Hz, ArH), 3.70-3.50 (poly(ethylene glycol) signals), 3.37 (s, 3H, OCH$_3$). **^{13}C NMR** (125 MHz, CDCl$_3$): δ 163.4 (2) (C=O), 134.4 (2), 128.9 (2), 123.4 (2) (ArC), 71.8, 70.5 (OCH$_2$), 58.9 (OCH$_3$).

PEG2K aminooxy 121
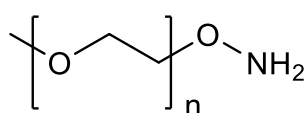
(2kDa)

Synthesis of this compound was performed as previously reported[204]. To a solution of **132** (1.00 g, 0.494 mmol) in DCM (10 ml) was added hydrazine hydrate (76 µL, 2.47 mmol). The solution was stirred vigorously for 30 min, during which time a white solid appeared. The solution was filtered through cotton wool, and the filtrate was collected. The resulting filtrate was then concentrated under a stream of nitrogen to give **121** as a white solid in quantitative yield that was used without further purification. **^{1}H NMR** (500 MHz, CDCl$_3$): δ 3.44 (poly(ethylene glycol) signals), 3.16 (s, 3H, OCH$_3$). **^{13}C NMR** (125 MHz, CDCl$_3$): δ 70.08 (OCH$_2$, poly(ethylene glycol) signals), 58.51 (OCH$_3$).

Solid Phase Peptide Synthesis (SPPS) and donor synthesis

Peptides were synthesised via manual solid phase peptide synthesis (SPPS) using an *in situ* neutralisation/HCTU activation procedure for Fmoc chemistry on an H-Gly-2-ClTrt resin (Sigma) using Fmoc protected amino acids as described below:

Preloaded resin preparation. The preloaded 2-chlorotrityl resin was weighed out into a 2 mL SPPS cartridge fitted with a PTFE stopcock, swollen in DMF for 30 min and then filtered.

Amino acid coupling. DIPEA (11.0 eq.) was added to a solution of amino acid (5.0 eq.) and HCTU (5.0 eq.) dissolved in the minimum volume of DMF and the solution added to the resin. The reaction mixture was gently agitated by rotation for 1 h, and the resin filtered off and washed with DMF (3×2 min with rotation).

Fmoc deprotection. A solution of 20% piperidine in DMF was added to the resin and gently agitated by rotation for 2 minutes. The resin was filtered off and repeated four more times, followed by washes with DMF (5×2 min with rotation).

Cleavage and Isolation. Resins containing full synthesised peptides were washed with DCM (3×2 min with rotation) and MeOH (3×2 min with rotation). The resin was dried on a vacuum manifold and further dried on a high vacuum line overnight. A solution of cleavage cocktail 95:2.5:2.5 (v/v) TFA:H_2O:triisopropylsilane was then added to the resin, and the resulting mixture was gently agitated by rotation for 60 min. The reaction mixture was drained into ice-cold Et_2O and centrifuged at 6000 rpm at 4 °C until pelleted (*ca*. 5-10 min). The supernatant was carefully decanted and subsequently resuspended, centrifuged and supernatant decanted three more times. The precipitated peptide pellet was then either dissolved 10% MeCN or in 10% aq. AcOH and lyophilised. Lyophilised peptides were then stored at -20 °C until required.

Notes on folate containing peptides

For designing peptides containing lysine modified at the N_ε position with folic acid, Fmoc-Lys (Dde)-OH was incorporated into the peptide chain as described above. Upon synthesising the desired peptide chain, and prior to **Cleavage and Isolation**, the resin bound peptide was treated with $NH_2NH_2.H_2O$ (2% in DMF) and gently

agitated by rotation for 5 min. This process was repeated, and the resin bound peptide was washed with DMF (3 × 2 min with rotation). A solution of folic acid (2.5 eq), HCTU (2.5 eq.), and DIPEA (5.0 eq.) in 1:1 DMSO:DMF was then added to the resin, and the resulting mixture was gently agitated by rotation for 8 h. The resin was then filtered off, washed with DMF (9 x 2 min with rotation), and the desired peptide was then obtained following the **Cleavage and Isolation** step mentioned prior. The desired peptide was then further purified via size-exclusion chromatography (Sephadex LH-20 in water), and fractions containing pure, desired peptide were lyophilised and stored at -20 °C until required.

NB: Precursor **135** and **136**, along with HPLC data for the peptides, was kindly provided by Robin Brabham. HPLC data for SVARLG **150** was kindly provided by Amanda Dixon.

Synthesis of SLYRAG 26

Synthesised using 400 mg resin (1.1 mmol g^{-1} loading)

Yield = 280 mg (95 %)
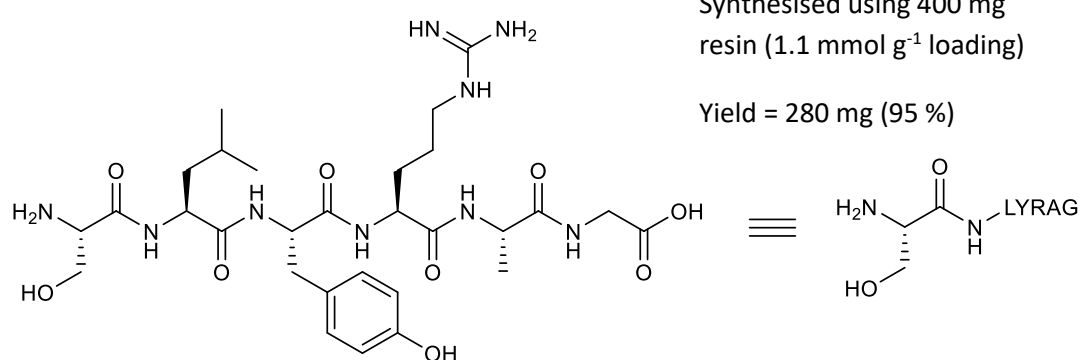
SLYRAG **26** was synthesised as previously described[121].

Synthesis of SVARLG 150
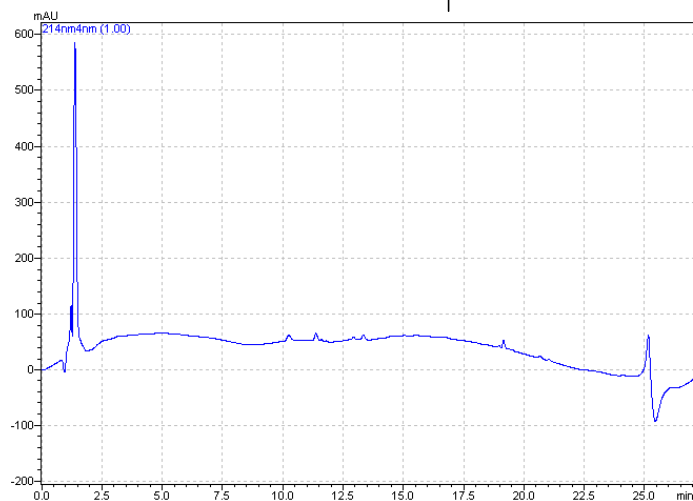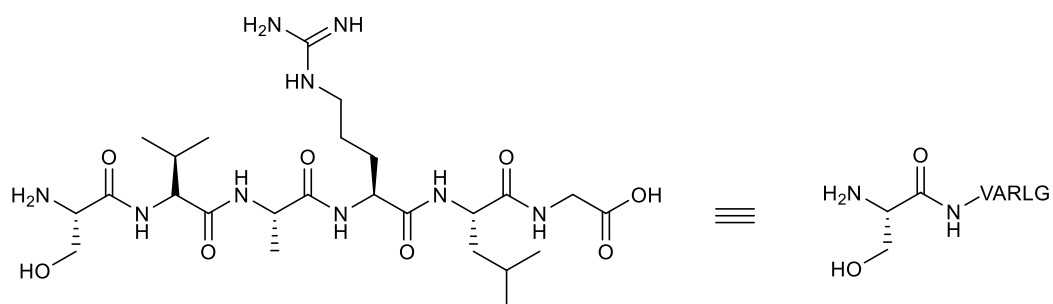
Synthesised using 400 mg resin (0.63 mmol g^{-1} loading)

Yield = 130 mg (86 %)

Synthesis of fluorescent label precursor 151
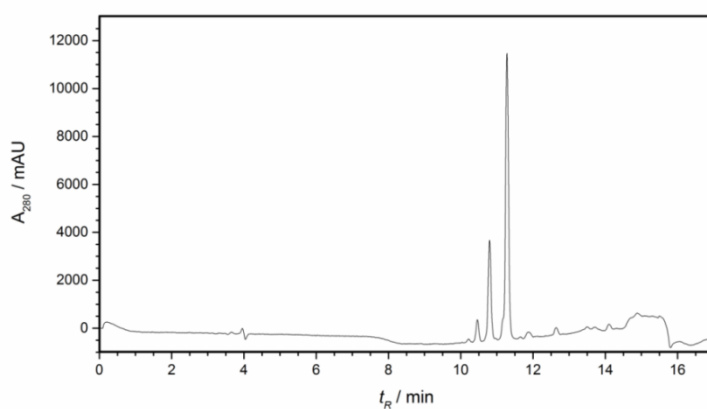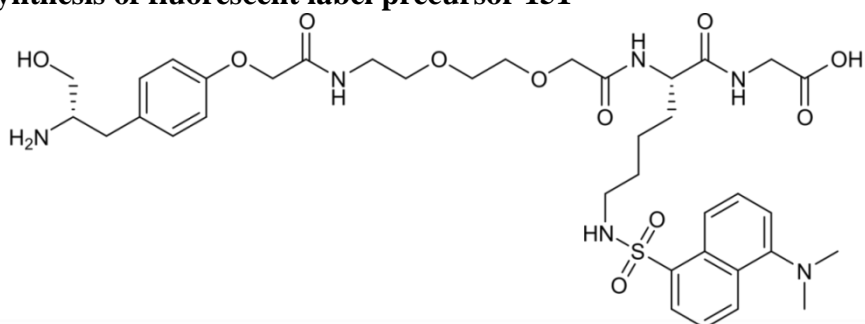
Synthesised using 85 mg resin (0.54 mmol g^{-1} loading)

Yield = 33 mg (92 %)

HRMS: Found [M+H]$^+$ 789.3511, C$_{37}$H$_{53}$N$_6$O$_{11}$S, requires 789.3488. **HPLC**:t_R 11.28 min.

Synthesis of fluorescent probe 85
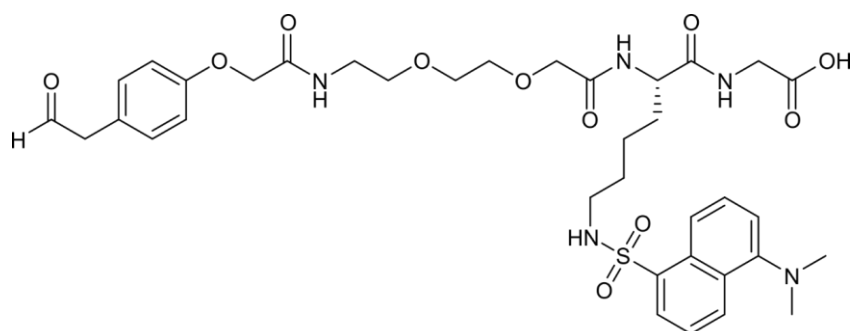
To a solution of **133** (10 mg in 500 µL, 10 mM, 0.1 M PB, 0.1 M NaCl pH 7.0) was added methionine (250 µL, 200 mM, 0.1 M PB, 0.1M NaCl pH 7.0) and $NaIO_4$ (210 µL, 112 mM, 0.1 M PB, 0.1 M NaCl pH 7.0). The reaction was mixed thoroughly, and allowed to sit for 2 min on ice in the dark. The solution was then loaded onto a solid phase extraction cartridge (Grace Davison Extract Clean, 8 ml reservoir, Fisher Scientific) equilibrated with water/acetonitrile. After initial washing with water, the product was eluted over a gradient of acetonitrile. The product was then diluted with water, and subsequently lyophilised to give **11** as a pale yellow, fluffy powder (4 mg, 40%). **LRMS**: Found $[M+H]^+$ 758.34, $C_{36}H_{48}N_5O_{11}S$, requires 758.34.

Synthesis biotin affinity tag precursor 152
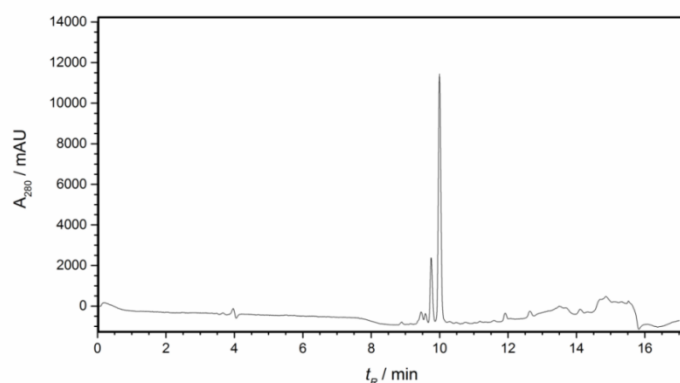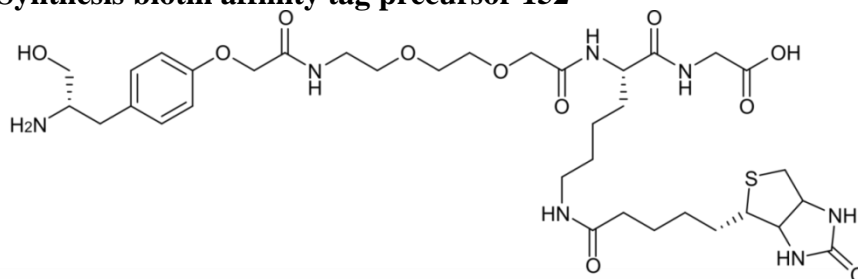
Synthesised using 100 mg resin (0.54 mmol g^{-1} loading)

Yield = 41 mg (98 %)

HRMS: Found $[M+H]^+$ 782.3781, $C_{35}H_{56}N_7O_{11}S$, requires 782.3753. **HPLC**:t_R 9.99 min.

Synthesis of biotin affinity tag 86
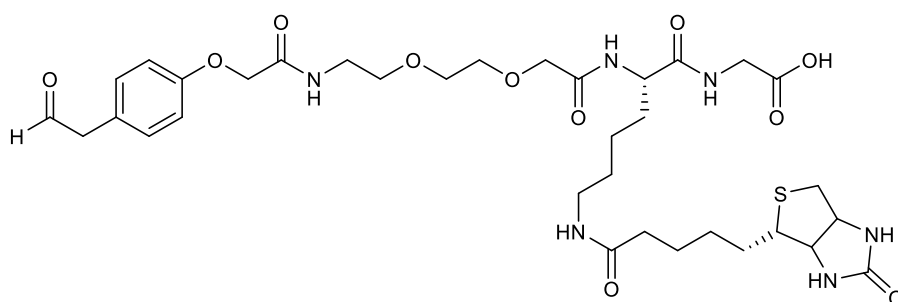
To a solution of **134** (10 mg in 500 µL, 10 mM, 0.1 M PB, 0.1 M NaCl pH 7.0) was added methionine (250 µL, 200 mM, 0.1 M PB, 0.1M NaCl pH 7.0) and $NaIO_4$ (210 µL, 112 mM, 0.1 M PB, 0.1 M NaCl pH 7.0). The reaction was mixed thoroughly, and allowed to sit for 2 min on ice in the dark. The solution was then The solution was then loaded onto a solid phase extraction cartridge (Grace Davison Extract Clean, 8 ml reservoir, Fisher Scientific) equilibrated with water/acetonitrile. After initial washing with water, the product was eluted over a gradient of acetonitrile. The product was then diluted with water, and subsequently lyophilised to give **86** as a white, fluffy powder (9 mg, 84%). **LRMS**: Found $[M+H]^+$ 751.39, $C_{34}H_{51}N_6O_{11}S$, requires 751.40.

Synthesis of bioorthogonal azide handle precursor 153

Synthesised using 100 mg resin (0.54 mmol g^{-1} loading)

Yield = 32 mg (84 %)
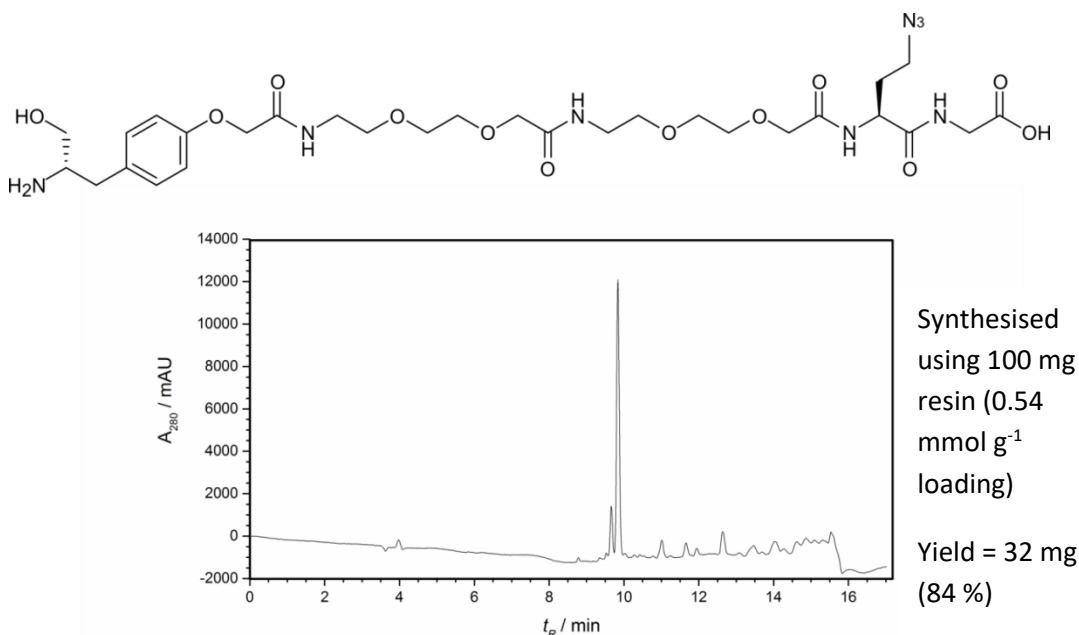
HRMS: Found $[M+H]^+$ 699.3322, $C_{29}H_{47}N_8O_{12}$, requires 699.3308. **HPLC**:t_R = 9.83 min.

Synthesis of bioorthogonal azide handle 87
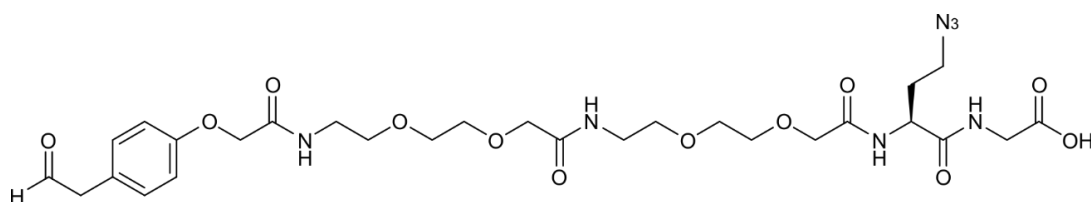
To a solution of **135** (10 mg in 500 µL, 10 mM, 0.1 M PB, 0.1 M NaCl pH 7.0) was added methionine (250 µL, 200 mM, 0.1 M PB, 0.1M NaCl pH 7.0) and NaIO$_4$ (210 µL, 112 mM, 0.1 M PB, 0.1 M NaCl pH 7.0). The reaction was mixed thoroughly, and allowed to sit for 2 min on ice in the dark. The solution was then loaded onto a solid phase extraction cartridge (Grace Davison Extract Clean, 8 ml reservoir, Fisher Scientific) equilibrated with water/acetonitrile. After initial washing with water, the product was eluted over a gradient of acetonitrile. The product was then diluted with water, and subsequently lyophilised to give **14** as a pale green, fluffy powder (8 mg, 80%). **LRMS**: Found [M+Na]$^+$ 690.29, C$_{28}$H$_{41}$N$_7$NaO$_{12}$, requires 690.27.

Synthesis of folate targeting moiety precursor 154
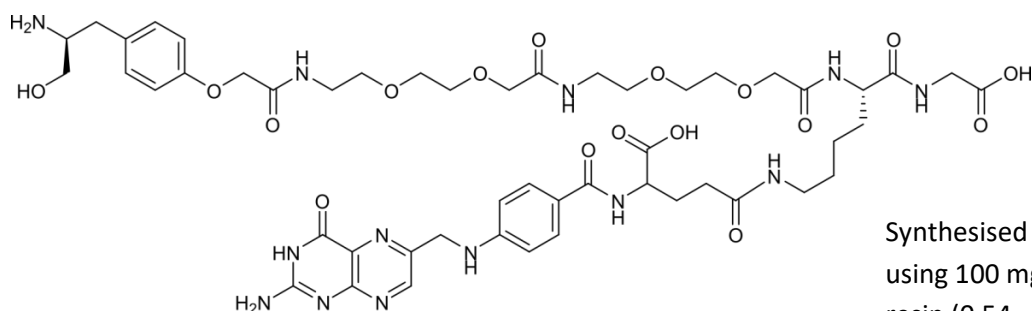
Synthesised using 100 mg resin (0.54 mmol g^{-1} loading)
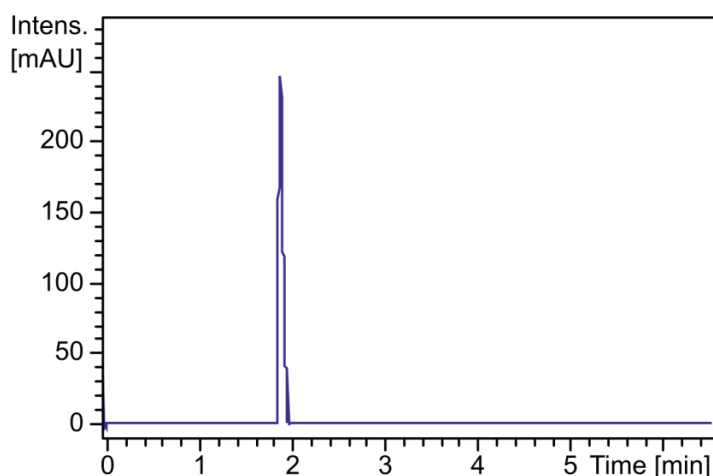
*Note: HPLC analysis of **136** was instead performed using the '**LC-MS analysis of peptide and protein ligations**' method as described previously for peptide analysis.*

HRMS: Found [M+H]$^+$ 1124.5037, C$_{50}$H$_{70}$N$_{13}$O$_{17}$, requires 1124.5007. **HPLC**:t_R 1.9 min.

Synthesis of folate targeting moiety 88
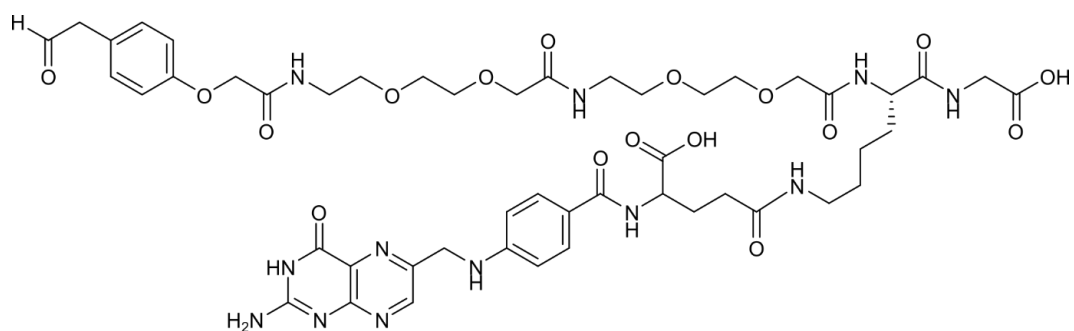
To a solution of **135** (10 mg in 500 µL, 9 mM, 0.1 M PB, 0.1 M NaCl pH 7.0) was added methionine (250 µL, 200 mM, 0.1 M PB, 0.1M NaCl pH 7.0) and $NaIO_4$ (210 µL, 112 mM, 0.1 M PB, 0.1 M NaCl pH 7.0). The reaction was mixed thoroughly, and allowed to sit for 2 min on ice in the dark. The solution was then loaded onto a solid phase extraction cartridge (Grace Davison Extract Clean, 8 ml reservoir, Fisher Scientific) equilibrated with water/acetonitrile. After initial washing with water, the product was eluted over a gradient of acetonitrile. The product was then diluted with water, and subsequently lyophilised to give **13** as a yellow, fluffy powder (3 mg, 31%). **LRMS**: Found $[M+2H]^{2+}$ 547.33, $C_{49}H_{66}N_{12}O_{17}$, requires 547.72.

Aldol product 83
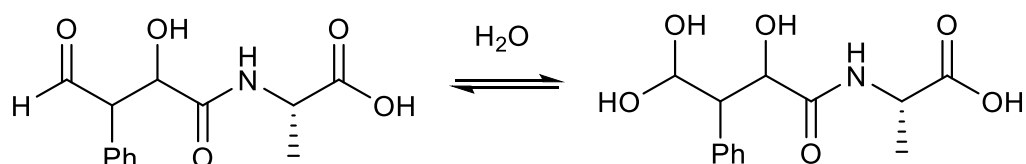
To a solution of compound **81** (0.05 g, 0.27 mmol) in 0.1 M PB pH 7.0 (700 µL) was added $NaIO_4$ (0.06 g, 0.28 mmol). The reaction was mixed until complete dissolution was achieved, and then allowed to sit at rt in the dark for 45 min**.** Complete oxidation of **81** to **82** was observed by LC-MS analysis**.** To this solution, L-proline **61** (0.006 g, 0.05 mmol) and phenylacetaldehyde **54** (0.031 mL, 0.032 g, 0.027 mmol) were added. The reaction mixture was mixed thoroughly, and then allowed to sit at 37 °C for 1 hour. Conversion to the desired aldol product **83** was monitored by LC-MS. The solvent was removed *in vacuo*, and the residue resuspended in ethyl acetate, resulting in the precipitation of L-proline **61** which was

filtered off. The filtrate was evaporated *in vacuo* and the crude reaction mixture was subjected to HPLC, ESI-MS and NMR analysis. ESI-MS confirmed full conversion of the starting material to the aldol product (see below for ESI-MS trace).
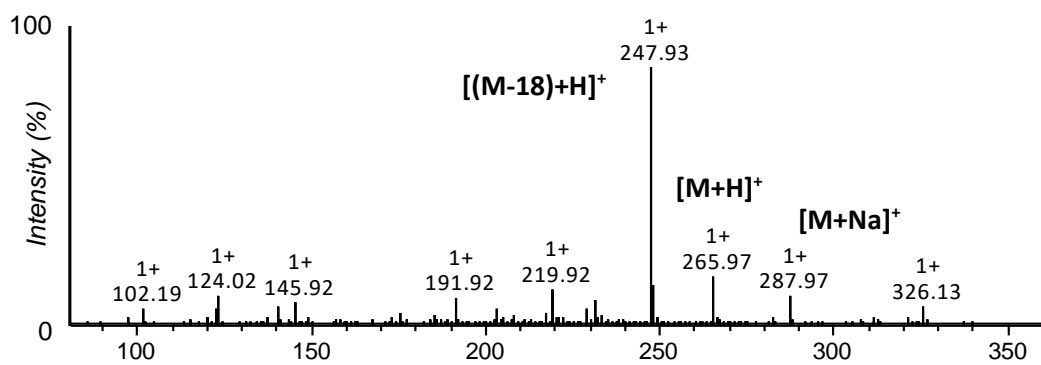
Peptide and protein chemical modifications

Oxidation of SLYRAG 26 to glyoxyl-LYRAG 27
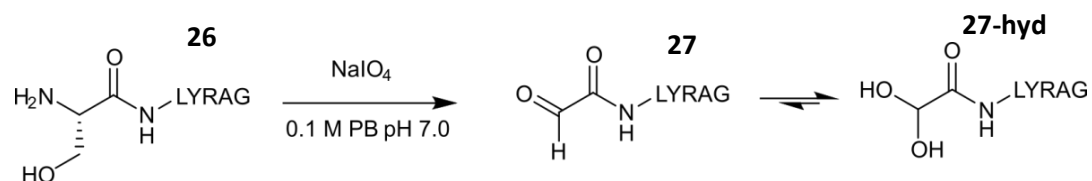
Oxidation of SLYRAG **26** to glyoxyl-LYRAG **27** was carried out by dissolving a desired amount of peptide in 1 mL of 25 mM phosphate buffer (PB) pH 7.0, followed by addition of 2 equivs. of $NaIO_4$. The solution was vortexed, then allowed to sit at room temperature in the dark for 1 h. The solution was then loaded onto a solid phase extraction cartridge (Grace Davison Extract Clean, 8 mL reservoir, Fisher Scientific) equilibrated with water/acetonitrile. After initial washing with water, the product was eluted over a gradient of acetonitrile. Fractions containing pure, oxidised peptide (as judged by LC-MS analysis) were pooled and subsequently lyophilised to give glyoxyl-LYRAG **27** as an orange solid, which was stored at -20 ºC until required.

Oxidation of SVARLG 155 to glyoxyl-VARLG 132
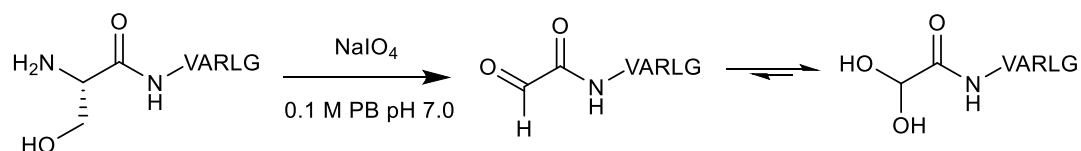
Oxidation of SVARLG **155** to give glyoxyl-VARLG **132** as a white powder was performed in the same manner as described for the preparation of glyoxyl-LYRAG **27**.

Transamination of horse heart myoglobin 28 to glyoxyl myoglobin 13
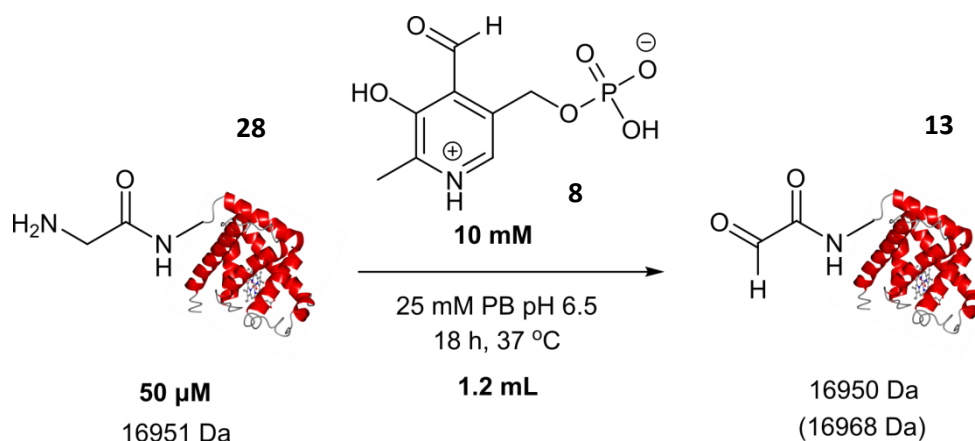
A 480 µL aliquot of a 250 µM of myoglobin **S1** stock solution in 25 mM PB pH 6.5 was charged with 1.2 mL of a 25 mM pyridoxal-5-phosphate **8** solution in 25 mM PB pH 6.5 (pH adjusted to pH 6.5 using 2M NaOH), and then charged with 720 µL of 25 mM PB pH 6.5. Final pH of solution was checked either by pH probe or pH paper. The mixture was briefly agitated, and incubated at 37 °C without further agitation for 24 h. The solution was then purified via spin concentration using 10,000 MWCO, and the resulting glyoxyl-myoglobin solution was concentrated to 200 µM, eluting with water. Oxidation to glyoxyl-myoglobin **13** was confirmed by LC-MS.

Oxidation of CTB 29 to glyoxyl-CTB 30
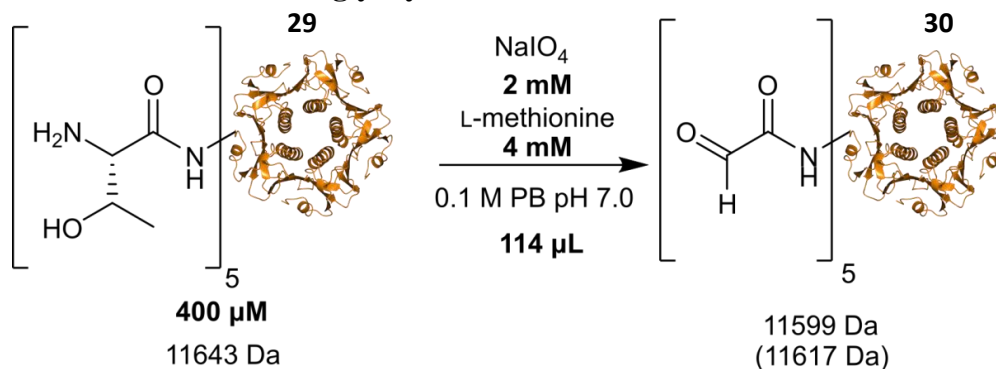
A 100 µL aliquot of an 453 µM CTB **29** stock in 0.1 M PB, 0.1 M NaCl, pH 7.0 was charged with 7 µL of a 66 mM L-methionine stock solution in 0.1 M PB, 0.1 M NaCl, pH 7.0, and 7 µL of a 33 mM NaIO₄ stock solution in 0.1 M PB, 0.1 M NaCl, pH 7.0. The solution was mixed by gentle pipetting, and allowed to sit on ice in the dark for 4 min. The reaction was immediately purified using a PD SpinTrap G25 desalting column (GE Healthcare Life Sciences), eluting into 25 mM PB pH 7.5. Quantitative oxidation to glyoxyl-CTB **30** was confirmed by LC-MS analysis.

Oxidation of thioredoxin 31 to glyoxyl-thioredoxin 32

31

NaIO$_4$
300 µM
L-methionine
600 µM

0.1 M PB pH 7.0

102 µL

32

H$_2$N

O

N
H

HO

80 µM
11675 Da

O

O

N
H

H

11644 Da
(11662 Da)
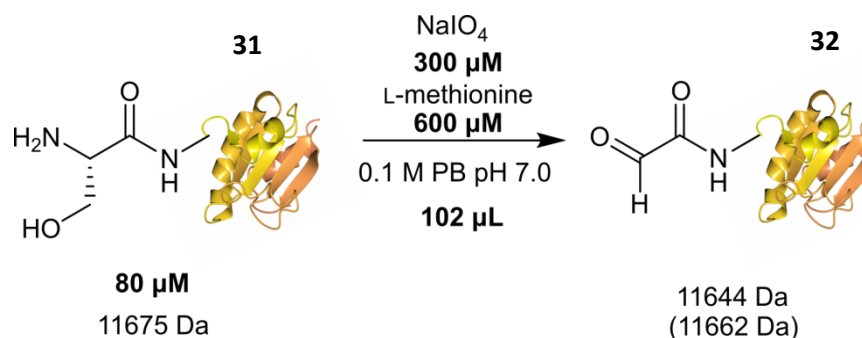
A 100 µL aliquot of an 85 µM thioredoxin **31** stock in 25 mM PB pH 7.5 was charged with 1 µL of a 66 mM L-methionine stock solution in 0.1 M PB, 0.1 M NaCl, pH 7.0, and 1 µL of a 33 mM NaIO$_4$ stock solution in 0.1 M PB, 0.1 M NaCl, pH 7.0. The solution was mixed by gentle pipetting, and allowed to sit on ice in the dark for 4 min. The reaction was immediately purified using a PD SpinTrap G25 desalting column (GE Healthcare Life Sciences), eluting into 25 mM PB pH 7.5. Quantitative oxidation to glyoxyl-thioredoxin **32** was confirmed by LC-MS analysis.

Oxidation of GFP (Y39CycloOctK) 33 to glyoxyl-GFP 34

33

NaIO$_4$
1.2 mM
L-methionine
2.4 mM

0.1 M PB pH 7.0

120 µL

34

H$_2$N

O

N
H

HO

100 µM
28669 Da

O

O

N
H

H

28638 Da
(28656 Da)
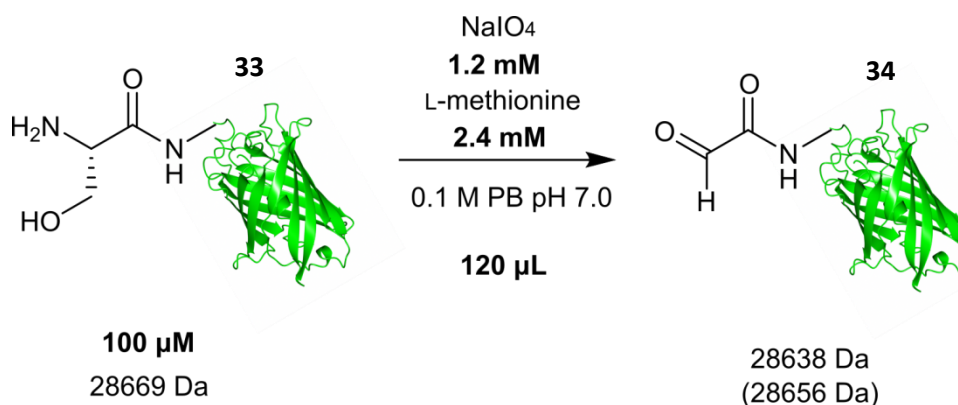
A 100 µL aliquot of a 100 µM GFP **33** (Y39CycloOctK) stock in 1 x PBS, pH 7.4, was charged with 3 µL of a 66 mM L-methionine stock solution in 0.1 M PB, 0.1 M NaCl, pH 7.0, and 2 µL of a 33 mM NaIO$_4$ stock solution in 0.1 M PB, 0.1 M NaCl, pH 7.0. The solution was mixed by gentle pipetting, and allowed to sit on ice in the dark for 4 min. The reaction was immediately purified using a PD SpinTrap G25 desalting column (GE Healthcare Life Sciences), eluting into 25 mM PB pH 7.5. Quantitative oxidation to glyoxyl-GFP (Y39CycloOctK) **34** was confirmed by LC-MS analysis.

Transamination of HASPA 35 to glyoxyl-HASPA 36
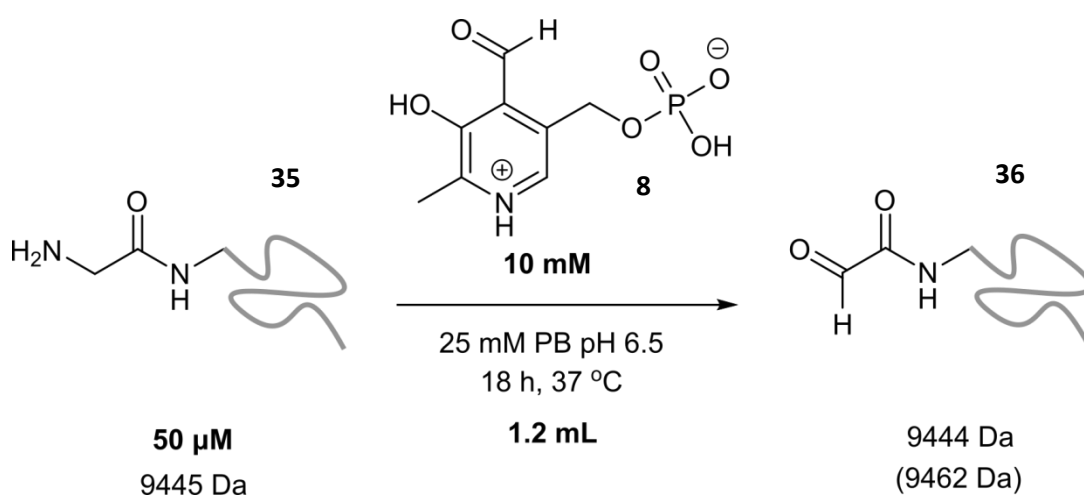
Transamination of HASPA **35** was performed in an identical fashion to that described for **Transamination of horse heart myoglobin 28 to glyoxyl myoglobin 13**.

Oxidation of HASPA(G1S) 37 to glyoxyl-HASPA 36
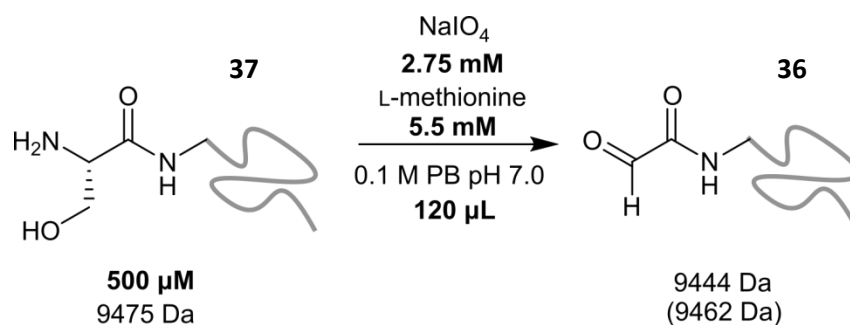
A 100 μL aliquot of an 600 μM HASPA(G1S) **37** stock in 0.1 M PB, 0.1 M NaCl, pH 7.0 was charged with 10 μL of a 66 mM L-methionine stock solution in 0.1 M PB, 0.1 M NaCl, pH 7.0, and 10 μL of a 33 mM NaIO$_4$ stock solution in 0.1 M PB, 0.1 M NaCl, pH 7.0. The solution was mixed by gentle pipetting, and allowed to sit on ice in the dark for 4 min. The reaction was immediately purified using a PD SpinTrap G25 desalting column (GE Healthcare Life Sciences), eluting into 25 mM PB pH 7.5. Quantitative oxidation to glyoxyl-HASPA **36** was confirmed by LC-MS analysis.

Oxidation of [15N]HASPA(G1S) 37-15N to glyoxyl-[15N]HASPA 36-15N
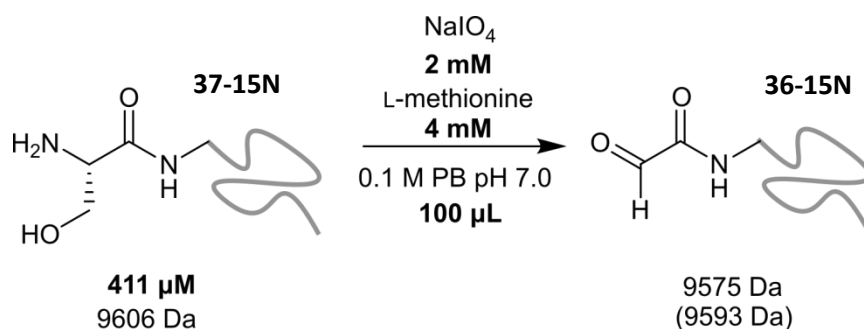
A 50 µL aliquot of an 822 µM [15N]HASPA(G1S) **37-15N** stock in 0.1 M PB, 0.1 M NaCl, pH 7.0 was charged with 36 µL of 0.1 M PB, 0.1 M NaCl, pH 7.0 buffer, 7 µL of a 66 mM L-methionine stock solution in 0.1 M PB, 0.1 M NaCl, pH 7.0, and 7 µL of a 33 mM NaIO₄ stock solution in 0.1 M PB, 0.1 M NaCl, pH 7.0. The solution was mixed by gentle pipetting, and allowed to sit on ice in the dark for 4 min. The reaction was immediately purified using a PD SpinTrap G25 desalting column (GE Healthcare Life Sciences), eluting into 25 mM PB pH 7.5. Quantitative oxidation to glyoxyl-HASPA **36-15N** was confirmed by LC-MS analysis.

Palladium decaging of sfGFP(N150ThzK) 39[116]
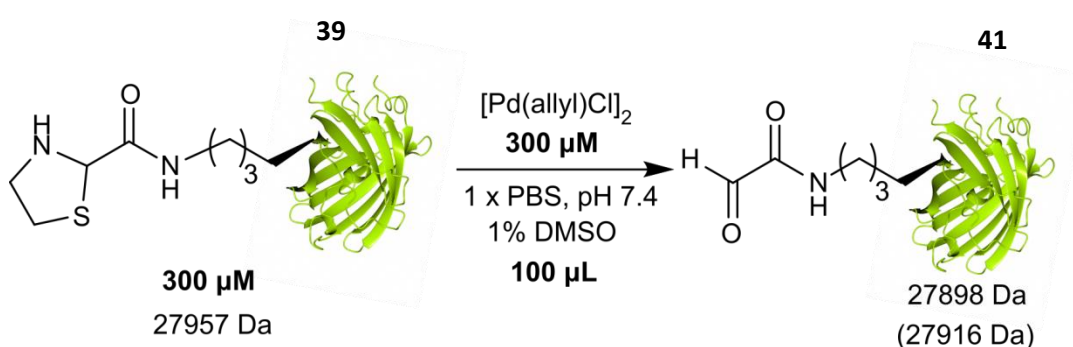
A 99 µl aliquot of a 300 µM sfGFP(N150ThzK) **39** stock in 1 x PBS, pH 7.4, was charged with 1 µl of a 30 mM allylpalladium(II) chloride dimer. The solution was mixed by gentle pipetting, and allowed to sit at room temperature for 60 min without further agitation. The reaction was then quenched by addition of 10 µL of a 3-mercaptopropanoic acid solution, 1% v/v solution, 10 x PBS (final conc^n = 0.1% v/v) to each aliquot, and allowed to sit at 25 °C for 15 min without further agitation. The reaction was then desalted using a PD MiniTrap G-25 (GE Healthcare Life

Sciences), eluting with 25 mM PB pH 7.5. Conversion to the decaged protein aldehyde **41** was confirmed by ESI-MS analysis.

Palladium decaging of GFP(Y39ThzK) 40 [116]
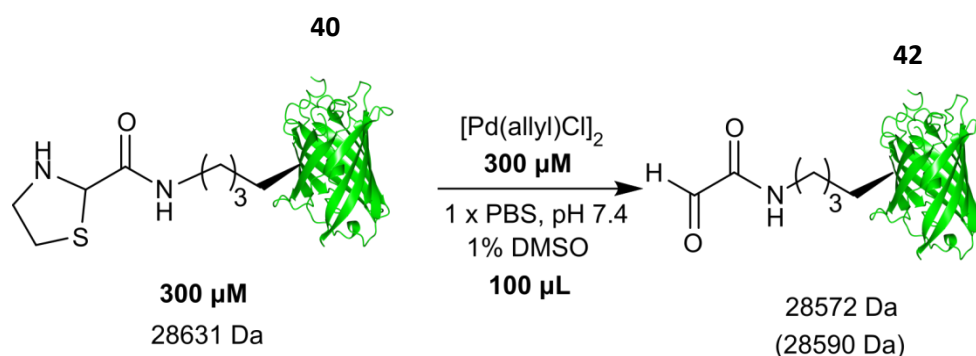
A 99 µl aliquot of a 300 µM GFP(Y39ThzK) **40** stock in 1 x PBS, pH 7.4, was charged with 1 µl of a 30 mM allylpalladium(II) chloride dimer stock solution in DMSO. The solution was mixed by gentle pipetting, and allowed to sit at room temperature for 60 min without further agitation. The reaction was then quenched by addition of 10 µL of a 3-mercaptopropanoic acid solution, 1% v/v solution, 10 x PBS (final conc^n = 0.1% v/v) to each aliquot, and allowed to sit at 25 °C for 15 min without further agitation. The reaction was then desalted using a PD MiniTrap G-25 (GE Healthcare Life Sciences), eluting with 25 mM PB pH 7.5. Conversion to the decaged protein aldehyde **42** was confirmed by ESI-MS analysis.

Modification of glyoxyl-LYRAG 27 with acetaldehyde 43
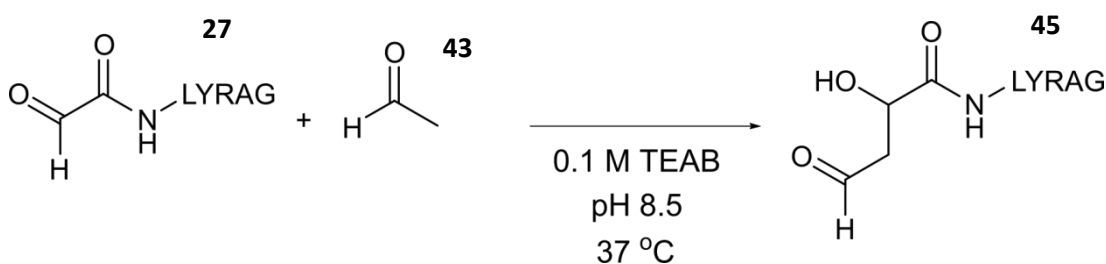
A 1 mL aliquot of a 1 mM glyoxyl-LYRAG **27** stock in 0.1 M TEAB pH 8.5 was charged with a drop of acetaldehyde **43**. The reaction was vortexed, and allowed to sit at 37 °C overnight without further agitation. The anticipated aldol product **45** was confirmed by LC-MS.

Synthesis of anticipated dually modified peptide 48
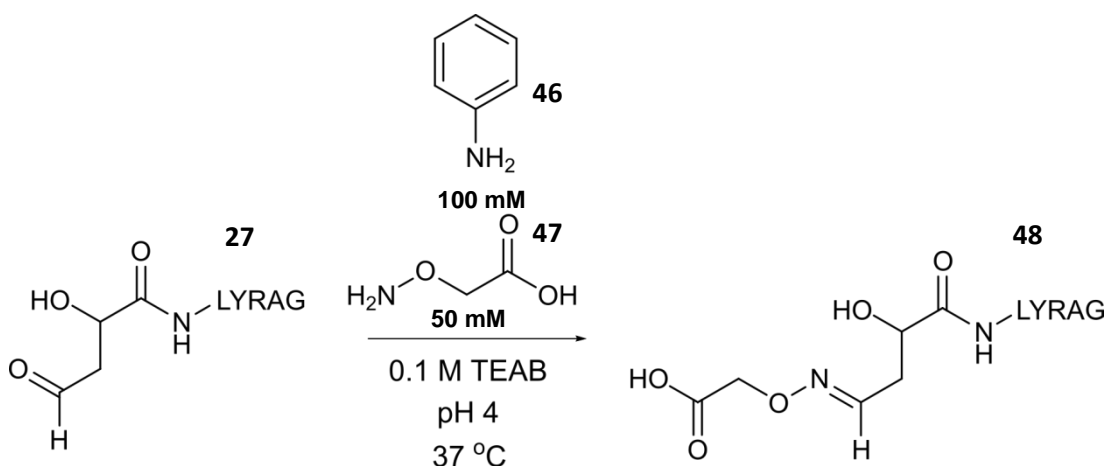
A 200 µL aliquot of aldol-modified-LYRAG **45** (prepared as described earlier) in 0.1 M TEAB, pH 8.5, was acidified to pH 4 with glacial acetic acid (checked by pH paper). The solution was then charged with 799 µL of Milli-Q (MQ) H_2O, and 1 µL of aniline **46** A 1 mg sample of aminooxy acetic acid was then added to the solution, vortexted, and allowed to sit at 37 ℃ overnight without further agitation. The anticipated dually-modified product **48** was confirmed by LC-MS.

Synthesis of anticipated aldol products 55-59
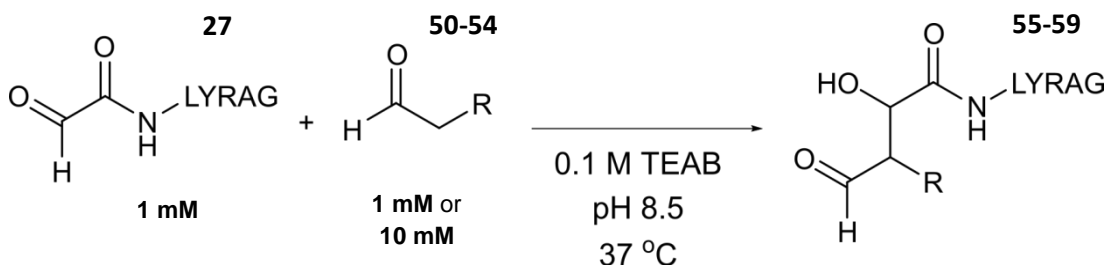
A 200 µL aliquot of a 5 mM glyoxyl-LYRAG **27** stock in MQ H_2O was charged with 10 µL of a 1M donor **50**, **51**, **52** or **53** stock solution in MQ H_2O (or 100 µL of a 10 mM donor **54** stock solution in MQ H_2O). The reaction was then charged with 690 µL (or 600 µL) of MQ H_2O, followed by addition of 100 µL of 1M TEAB pH 8.5. The solution was vortexed, and allowed to sit at 37 ℃ overnight without further agitation. The anticipated aldol products **55-59** were confirmed by LC-MS.

Synthesis of anticipated aldol products 56 and 59 at neutral pH
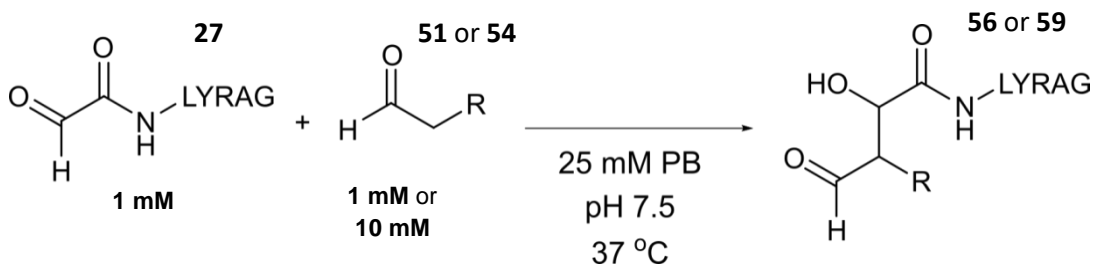
A 200 µL aliquot of a 5 mM glyoxyl-LYRAG **27** stock in 25 mM PB, pH 7.5, was charged with 10 µL of a 1M donor **50**, **51**, **52** or **53** stock solution in 25 mM PB pH 7.5 (or 100 µL of a 10 mM donor **54** stock solution in 25 mM PB pH 7.5). The reaction was then charged with 790 µL (or 700 µL) of 25 mM PB pH 7.5. The solution was vortexed, and allowed to sit at 37 °C overnight without further agitation. Conversion to the anticipated aldol products **56** or **59** were confirmed by LC-MS.

Screening of conditions for the uncatalysed synthesis of aldol-product 60
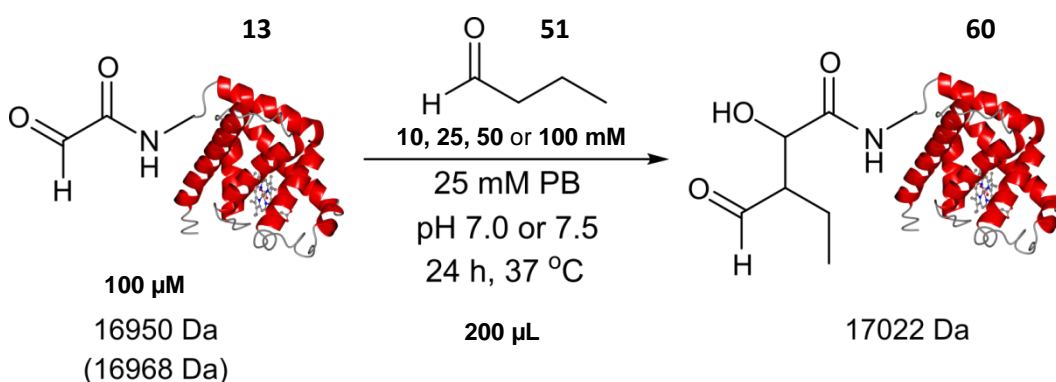
A 100 µL aliquot of a 200 µM glyoxyl myoglobin **13** stock (prepared as described earlier) in MQ H$_2$O was charged with 75 µL of 50 mM PB, and then charged with 25 µL of a 80 mM, 200 mM, 400 mM, or 800 mM donor **51** stock solution in 50 mM PB. Following mixing by pipetting, the reaction was allowed to sit at 37 °C for 24 h without further agitation. Conversion to the anticipated aldol product **60** was judged by LC-MS.

Screening of L-proline 61 concentration in the synthesis of 56
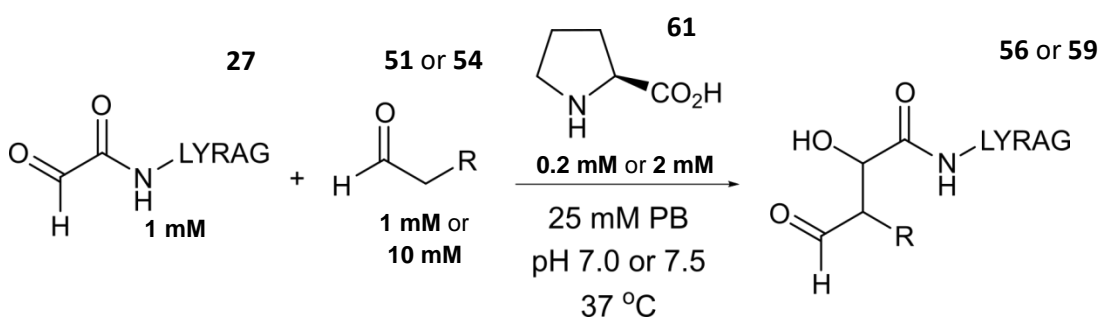
A 200 µL aliquot of a 5 mM glyoxyl-LYRAG **27** stock in 25 mM PB, pH 7.5, was charged with 10 µL of a 1M donor **50**, **51**, **52** or **53** stock solution in 25 mM PB pH 7.5 (or 1.12 µL of **54**). The reaction was then charged with 690 µL (or 698.88 µL) of 25 mM PB pH 7.5, followed by addition of 100 µL of a 20 mM or 200 mM L-proline **61** stock solution. The solution was vortexed, and allowed to sit at 37 ℃ overnight without further agitation. Conversion to the anticipated aldol products **56** or **59** were confirmed by LC-MS.

Screening of L-proline 61 concentration in the synthesis of 60
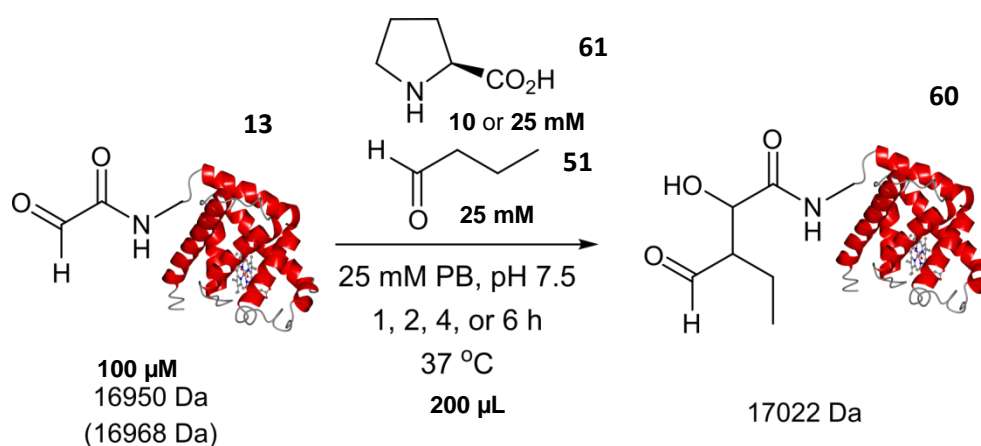
A 100 µL aliquot of a 200 µM glyoxyl myoglobin **13** stock (prepared as described earlier) in MQ H₂O was charged with 65 µL or 50 µL 50 mM PB pH 7.5. The solution was then charged with 10 µL or 25 µL of a 200 mM L-proline **61** stock solution in 50 mM PB pH 7.5. The solution was then charged with 25 µL of a 200 mM donor **51** stock solution in 50 mM PB pH 7.5 Following mixing by pipetting, the reaction was allowed to sit at 37 ℃ for 1 h, 2h, 4h, or 6h, without further agitation. Conversion to the anticipated aldol product **60** was judged by LC-MS.

Screening catalysts and donors for the synthesis of aldol product 72 and 73
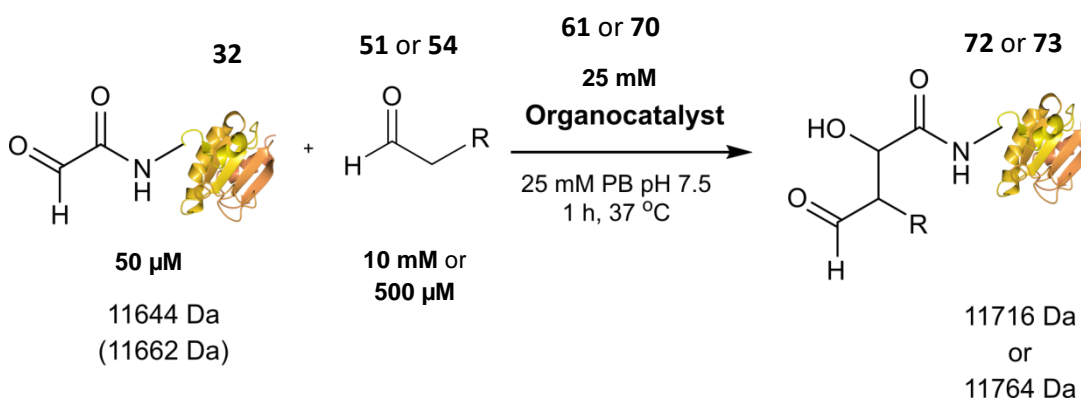
A 12.5 µL aliquot of a 80 µM glyoxyl-thioredoxin **32** stock (prepared as described earlier) in 25 mM PB pH 7.5 was charged with 2.5 µL of a 200 mM L-proline **61** or proline tetrazole **70** stock solution in 25 mM PB pH 7.5. The solution was then charged with 1 µL of a 200 mM donor **51** or 10 mM donor **54** stock solution in 25 mM PB pH 7.5, followed by addition of 5 µL of 25 mM PB pH 7.5. Following mixing by pipetting, the reaction was allowed to sit at 37 ℃ for 1 h without further agitation. Conversion to the anticipated aldol product **72** or **73** was judged by LC-MS.

## Synthesis of fluorescently labelled thioredoxin 89
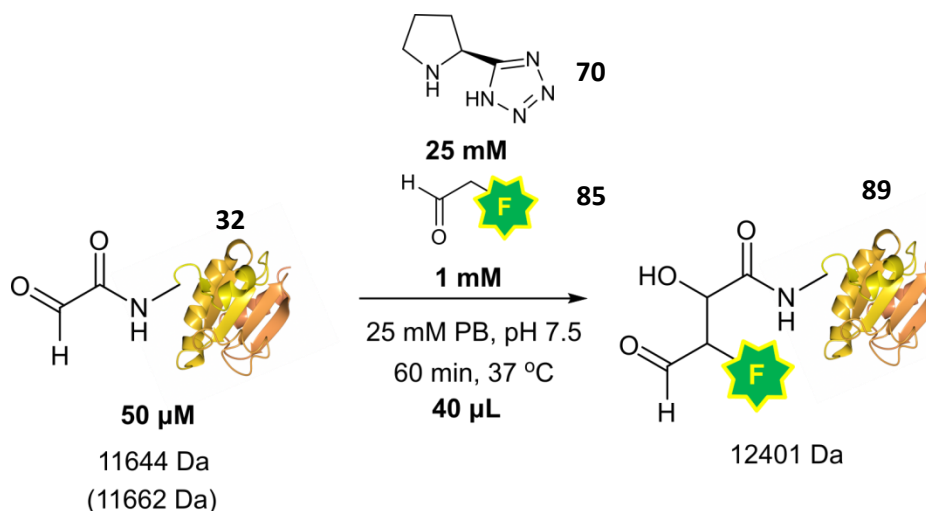
A 25 µL aliquot of 80 µM glyoxyl-thioredoxin **32** stock prepared as described earlier in 25 mM PB pH 7.5 was charged with 5 µL of a 200 mM proline tetrazole **70** stock solution in 25 mM PB pH 7.5. The solution was then charged with 10 µL of a 4 mM fluorescent label **85** stock solution in 25 mM PB pH 7.5. Following mixing by

pipetting, the reaction was allowed to sit at 37 °C for 60 min without further agitation. Quantitative labelling to fluorescently labelled thioredoxin **89** was confirmed by LC-MS analysis.

Synthesis of biotinylated thioredoxin 94
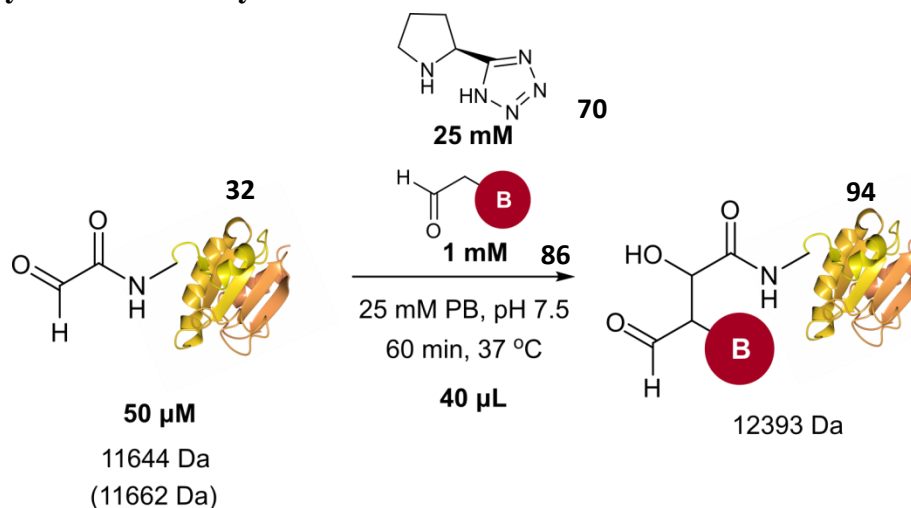
A 25 µL aliquot of 80 µM glyoxyl-thioredoxin **32** stock (prepared as described earlier) in 25 mM PB pH 7.5 was charged with 5 µL of a 200 mM proline tetrazole **70** stock solution in 25 mM PB pH 7.5. The solution was then charged with 10 µL of a 4 mM biotin affinity tag **86** stock solution in 25 mM PB pH 7.5. Following mixing by pipetting, the reaction was allowed to sit at 37 °C for 60 min without further agitation. Quantitative labelling to biotinylated thioredoxin **94** was confirmed by LC-MS analysis.
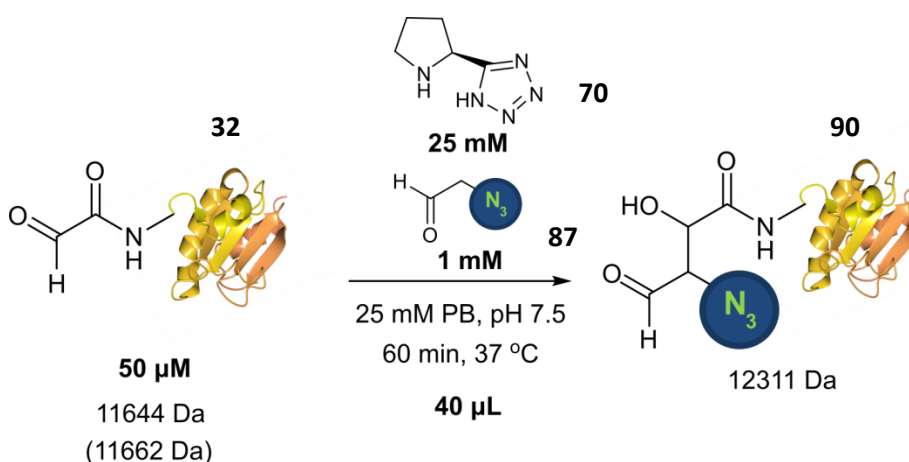
Synthesis of azide labelled thioredoxin 90

A 25 µL aliquot of 80 µM glyoxyl-thioredoxin **32** stock (prepared as described earlier) in 25 mM PB pH 7.5 was charged with 5 µL of a 200 mM proline tetrazole

70 stock solution in 25 mM PB pH 7.5. The solution was then charged with 10 µL of a 4 mM bioorthogonal azide handle **87** stock solution in 25 mM PB pH 7.5. Following mixing by pipetting, the reaction was allowed to sit at 37 °C for 60 min without further agitation. Quantitative labelling to azide labelled thioredoxin **90** was confirmed by LC-MS analysis.
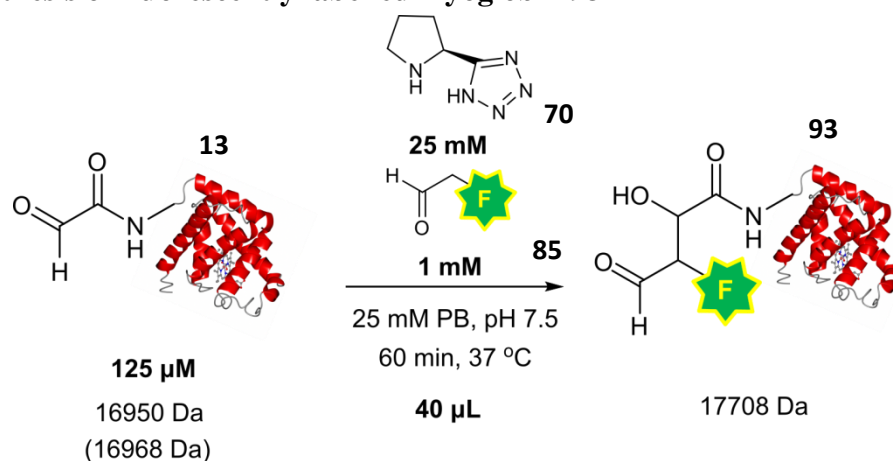
Synthesis of fluorescently labelled myoglobin 93

A 25 µL aliquot of 200 µM glyoxyl-myoglobin **13** stock (prepared as described earlier) in 25 mM PB pH 7.5 was charged with 5 µL of a 200 mM proline tetrazole **70** stock solution in 25 mM PB pH 7.5. The solution was then charged with 10 µL of a 4 mM fluorescent label **85** stock solution in 25 mM PB pH 7.5. Following mixing by pipetting, the reaction was allowed to sit at 37 °C for 60 min without further agitation. Quantitative labelling to fluorescently labelled myoglobin **93** was confirmed by LC-MS analysis.

Synthesis of folate labelled GFP (Y39CycloOctK) 96
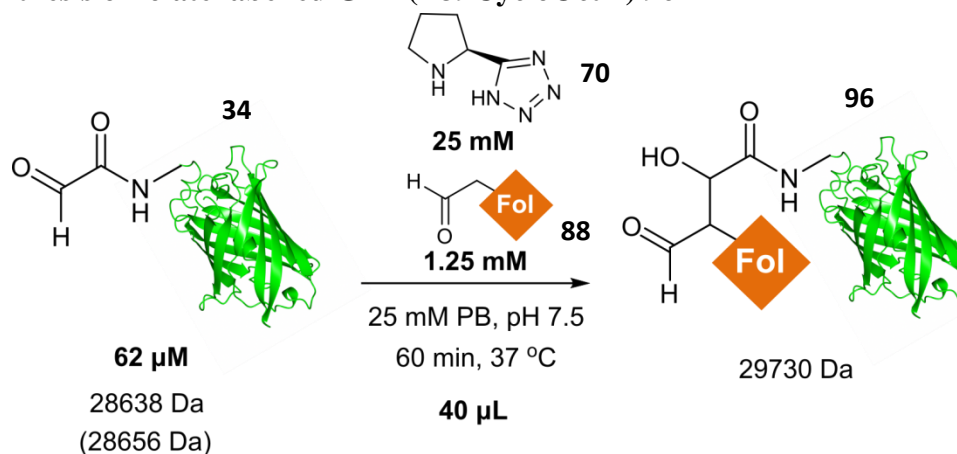
A 25 µL aliquot of 100 µM glyoxyl-GFP **34** stock (prepared as described earlier) in 25 mM PB pH 7.5 was charged with 5 µL of a 200 mM proline tetrazole **70** stock solution in 25 mM PB pH 7.5. The solution was then charged with 10 µL of a 5 mM

folate targeting moiety **88** stock solution in 25 mM PB pH 7.5. Following mixing by pipetting, the reaction was allowed to sit at 37 °C for 60 min without further agitation. Quantitative labelling to folate labelled GFP **96** confirmed by ESI-MS analysis.

Synthesis of biotinylated GFP 92
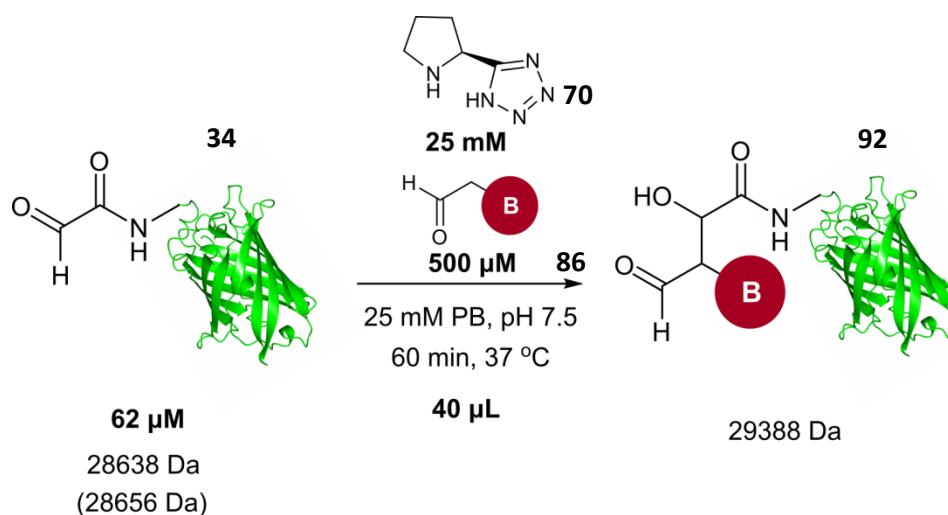
A 25 µL aliquot of 100 µM glyoxyl-GFP (Y39CycloOctK) **34** stock (prepared as described earlier) in 25 mM PB pH 7.5 was charged with 5 µL of a 200 mM proline tetrazole **70** stock solution in 25 mM PB pH 7.5. The solution was then charged with 10 µL of a 2 mM biotin affinity tag **86** stock solution in 25 mM PB pH 7.5. Following mixing by pipetting, the reaction was allowed to sit at 37 °C for 60 min without further agitation. Quantitative labelling to biotinylated GFP **92** was confirmed by ESI-MS analysis.

Synthesis of internally azide labelled sfGFP 99
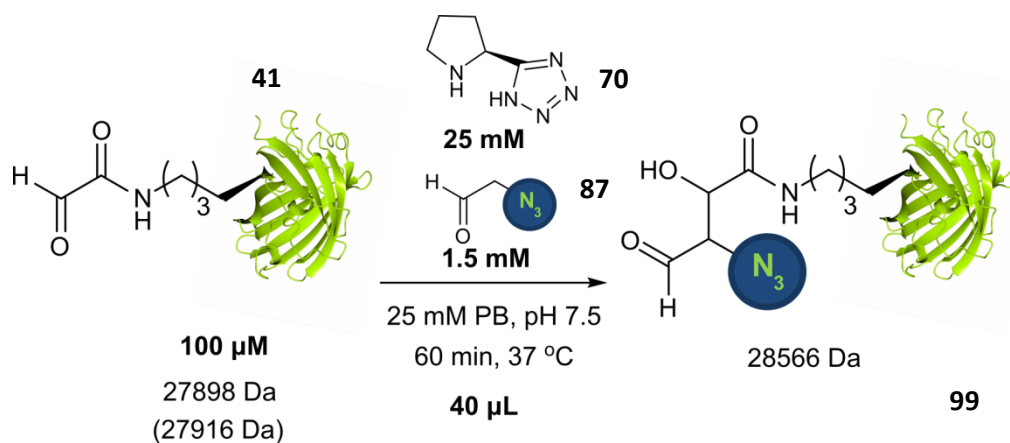
A 25 µL aliquot of 160 µM sfGFP(ThzK150Oxo) **41** (prepared as described earlier) in 25 mM PB pH 7.5 was charged with 5 µL of a 200 mM proline tetrazole **70** stock solution in 25 mM PB pH 7.5. The solution was then charged with 10 µL of a 5 mM bioorthogonal azide handle **87** stock solution in 25 mM PB pH 7.5. Following mixing by pipetting, the reaction was allowed to sit at 37 °C for 60 min without further agitation. Quantitative labelling to internally azide labelled sfGFP **99** was confirmed by ESI-MS analysis

Synthesis of internally azide labelled GFP 100
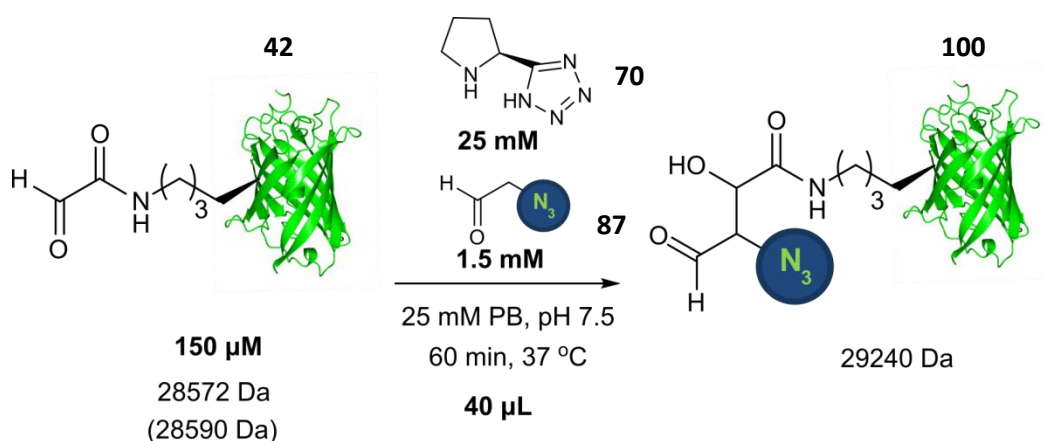
A 25 µL aliquot of 240 µM GFP(ThzK39Oxo) **42** (prepared as described earlier) in 25 mM PB pH 7.5 was charged with 5 µL of a 200 mM proline tetrazole **70** stock solution in 25 mM PB pH 7.5. The solution was then charged with 10 µL of a 5 mM bioorthogonal azide handle **87** stock solution in 25 mM PB pH 7.5. Following mixing by pipetting, the reaction was allowed to sit at 37 °C for 60 min without further agitation. Quantitative labelling to internally azide labelled GFP **100** was confirmed by ESI-MS analysis

Synthesis of azide labelled HASPA 95
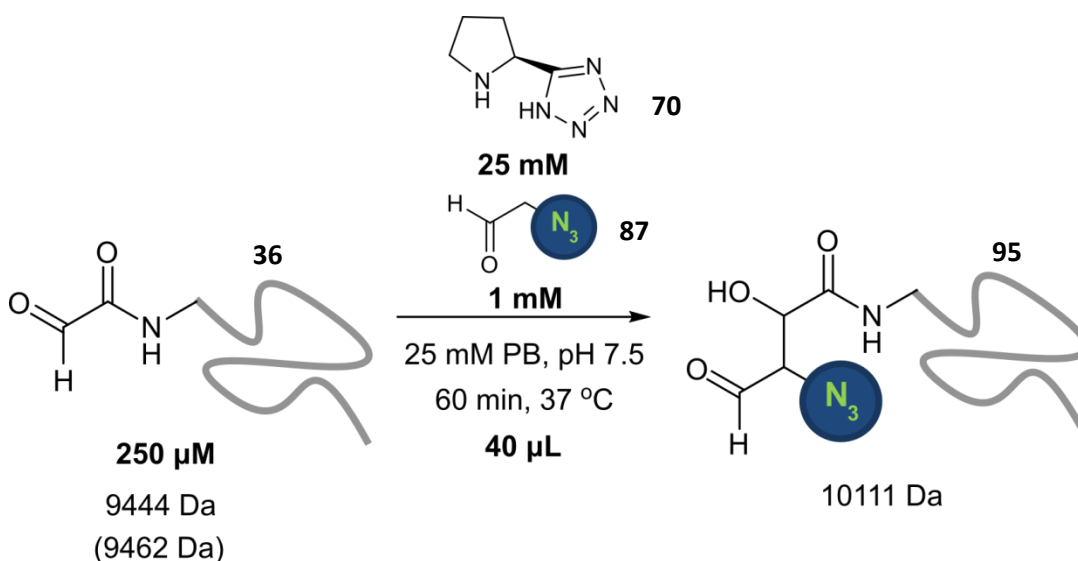
A 25 µL aliquot of 400 µM glyoxyl-HASPAG1S **36** (prepared as described earlier) in 25 mM PB pH 7.5 was charged with 5 µL of a 200 mM proline tetrazole **70** stock solution in 25 mM PB pH 7.5. The solution was then charged with 10 µL of a 5 mM bioorthogonal azide handle **87** stock solution in 25 mM PB pH 7.5. Following mixing by pipetting, the reaction was allowed to sit at 37 °C for 60 min without further agitation. Quantitative labelling to azide labelled HASPA **S57** was confirmed by LC-MS analysis.

Fluorescent labelling of glyoxyl-CTB 30
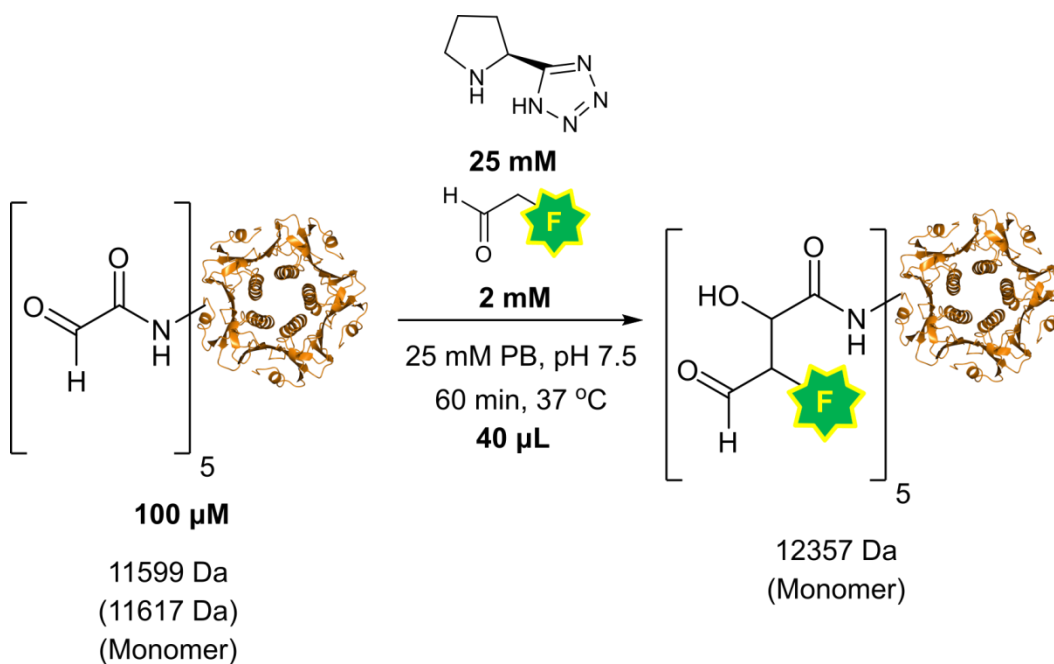
A 25 µL aliquot of 400 µM [15N]HASPA(G1S) **36-15N** (prepared as described earlier) in 25 mM PB pH 7.5 was charged with 5 µL of a 200 mM proline tetrazole **70** stock solution in 25 mM PB pH 7.5. The solution was then charged with 10 µL of a 2 mM fluorescent label **85** stock solution in 25 mM PB pH 7.5. Following mixing by pipetting, the reaction was allowed to sit at 37 ºC for 60 min without further agitation. Quantitative labelling to fluorescently labelled [15N]HASPA **91-15N** was confirmed by LC-MS analysis.

Synthesis of biotinylated myoglobin 97
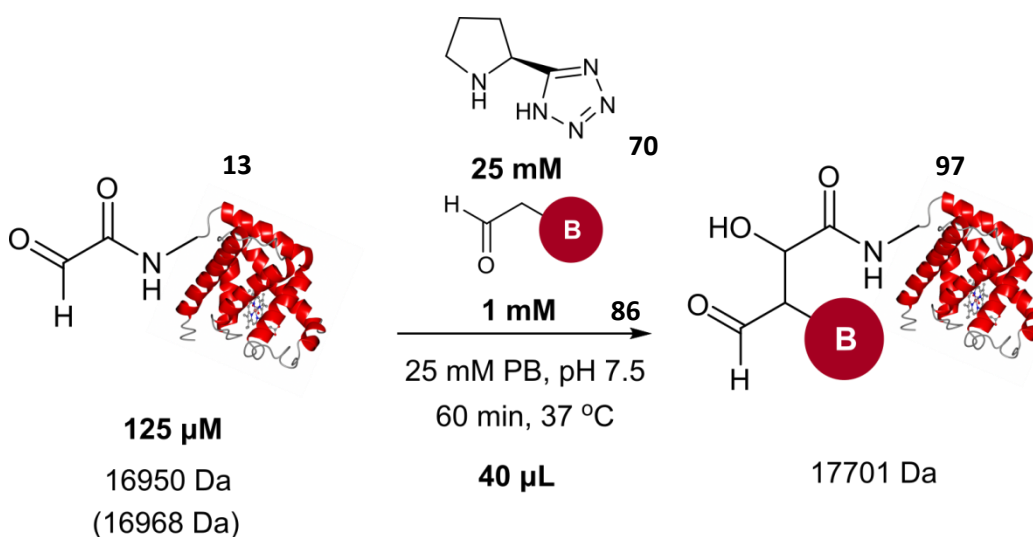
A 25 µL aliquot of 200 µM glyoxyl-myoglobin **5** stock (prepared as described earlier) in 25 mM PB pH 7.5 was charged with 5 µL of a 200 mM proline tetrazole **9** stock solution in 25 mM PB pH 7.5. The solution was then charged with 10 µL of a 4 mM biotin affinity tag **12** stock solution in 25 mM PB pH 7.5. Following mixing by pipetting, the reaction was allowed to sit at 37 ºC for 60 min without further agitation. Quantitative labelling to biotinylated myoglobin **S58** was confirmed by LC-MS analysis. Structural integrity of the myoglobin protein was determined by UV/Vis analysis.

Synthesis of azide labelled myoglobin 116
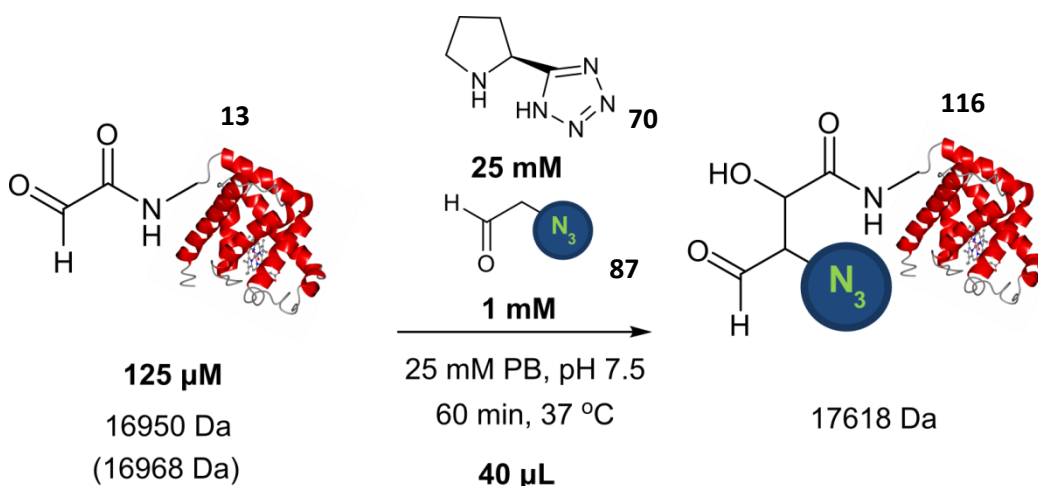
A 25 µL aliquot of 200 µM glyoxyl-myoglobin **13** stock (prepared as described earlier) in 25 mM PB pH 7.5 was charged with 5 µL of a 200 mM proline tetrazole **70** stock solution in 25 mM PB pH 7.5. The solution was then charged with 10 µL of a 4 mM bioorthogonal azide handle **87** stock solution in 25 mM PB pH 7.5. Following mixing by pipetting, the reaction was allowed to sit at 37 °C for 60 min without further agitation. Quantitative labelling to azide labelled myoglobin **116** was confirmed by LC-MS analysis.
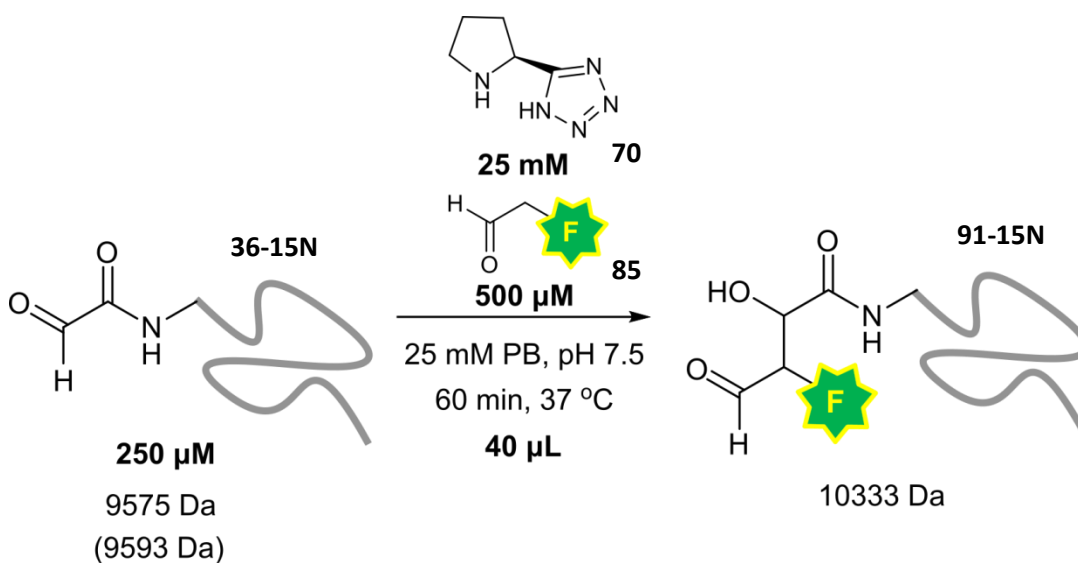
Synthesis of fluorescently labelled [15N] HASPA 91

A 25 µL aliquot of 400 µM [15N]HASPA(G1S) **36-15N** (prepared as described earlier) in 25 mM PB pH 7.5 was charged with 5 µL of a 200 mM proline tetrazole **70** stock solution in 25 mM PB pH 7.5. The solution was then charged with 10 µL of a 2 mM fluorescent label **85** stock solution in 25 mM PB pH 7.5. Following mixing by pipetting, the reaction was allowed to sit at 37 ℃ for 60 min without further agitation. Quantitative labelling to fluorescently labelled [15N]HASPA **91-15N** was confirmed by LC-MS analysis.

Synthesis of internally biotinylated sfGFP 101

A 25 µL aliquot of 160 µM sfGFP(150GlyoxylK) **41** (prepared as described earlier) in 25 mM PB pH 7.5 was charged with 5 µL of a 200 mM proline tetrazole **70** stock solution in 25 mM PB pH 7.5. The solution was then charged with 10 µL of a 5 mM biotin affinity tag **86** stock solution in 25 mM PB pH 7.5. Following mixing by pipetting, the reaction was allowed to sit at 37 ℃ for 60 min without further agitation. Quantitative labelling to internally biotinylated sfGFP **101** was confirmed by ESI-MS analysis

Site-selective biotinylation of GFP in cell lysate and subsequent protein pulldown

A 10 mL culture of cells expressing GFP(Y39ThzK) **40** (prepared as described previously [116]) was harvested by centrifugation. The resulting cell pellets were resuspended in 1.25 mL of 4 x PBS and lysed by sonication on ice for 9 x 30s with 30 s intervals. The cell lysate was clarified by centrifugation (17000 x g, 4 ℃, 15

min), and the pelleted, insoluble matter was discarded. The supernatant was retained, and both the concentration and content of GFP **40** was determined by UV/Vis absorbance measured at 488 nm assuming a molar extinction coefficient of ε = 55,000 M^{-1} cm^{-1} for GFP (GFP $conc^n$ = 0.643 mg mL^{-1}, GFP content = 803 µg). A 1 mL sample of the supernatant was then carried forward for palladium-mediated decaging (GFP $conc^n$ = 0.643 mg mL^{-1}, GFP content = 0.643 µg). Ten 100 µL aliquots of cell lysate were charged with 1 µL of a 30 mM allylpalladium(II) chloride dimer stock solution in DMSO (final $conc^n$ = 300 µM). Following mixing by pipetting, the reactions were allowed to sit at 25 $^\circ$C for 1 h without further agitation. The reactions were then quenched by addition of 10 µL of a 3-mercaptopropanoic acid solution, 1% v/v solution, 10 x PBS (final $conc^n$ = 0.1% v/v) to each aliquot, and allowed to sit at 25 $^\circ$C for 15 min without further agitation. The reactions were pooled, desalted using PD MiniTrap G-25 columns (GE Healthcare Life Sciences), eluting with 25 mM PB pH 7.5, and concentrated to 180 µL using 10,000 MWCOs (Amicon Ultra-0.5 mL Centrifugal Filters) to give the 'post-decaged' lysate containing GFP(ThzK39Oxo) **S14** (GFP $conc^n$ = 2.5 mg mL^{-1}, GFP content = 451 µg, protein recovery from initial supernatant sample used= 70%). A 125 µL aliquot of 'post-decaged' lysate (GFP $conc^n$ = 2.5 mg mL^{-1}, GFP content = 313 µg) was then carried forward for site-selective biotinylation. Five 25 µL aliquots of 'post-decaged' lysate in 25 mM PB pH 7.5 were charged with 5 µL of a 200 mM proline tetrazole **9** stock solution in 25 mM PB pH 7.5. The five solutions were then charged with 10 µL of a 5 mM biotin affinity tag **12** stock solution in 25 mM PB pH 7.5. Following mixing by pipetting, the reactions were allowed to sit at 37 $^\circ$C for 60 min without further agitation. Excess affinity tag **12** was removed via spin concentration using 10,000 MWCOs (Amicon Ultra-0.5 mL Centrifugal Filters) to give 100 µL of the 'post-OPAL' lysate containing internally biotinylated GFP **S16** (GFP $conc^n$ =1.73 mL^{-1}, GFP content = 173 µg, protein recovery from 'post-decaged' lysate sample used = 55%). A 'post-OPAL' lysate sample containing a 5 µg GFP content was retained for SDS-PAGE analysis. The remaining 'post-OPAL' lysate (GFP content = 168 µg) was loaded onto a 2 mL monomeric avidin agarose column (prepared in house using Pierce™ Monomeric Avidin Agarose according to the user guide provided, ThermoFisher Scientific), washed with 1 x PBS pH 7.4, and eluted using 2 mM biotin in 1 x PBS pH 7.4, collecting 1 mL fractions, according to the user guide provided. In total, one 2 mL fraction of flowthrough was collected

(Fraction FT), six fractions of 2 mL washes with 1 x PBS were collected (Fractions 1-6), and 14 fractions of 1 mL washes with 2 mM biotin in 1 x PBS pH 7.4 were collected (Fractions 7-20). Fractions were first visualised for protein fluorescence using a Syngene G:BOX Chemi XRQ equipped with a Synoptics 4.0 MP camera in line with GeneSys software (Version 1.5.7.0), and fractions of interest were subsequently analysed via SDS-PAGE. Fractions 1, 7, 8, 9, 10, 11, 12, 13, and 14, were also analysed by UV/Vis absorbance measured at 488 nm for GFP content. Fraction 1 was determined to contain 40 µg total GFP content, whereas fractions 7, 8, 9, 10, 11, 12, 13, and 14, were determined to contain 123 µg total GFP content, leaving 5 µg of 168 µg 'post-OPAL' GFP material unaccounted for. Overall, the pooled fractions 7-14 resulted in a 73% recovery of internally biotinylated GFP **102** that was originally loaded onto the monomeric avidin agarose column.

Synthesis of aldol-oxime-LYRAG 106
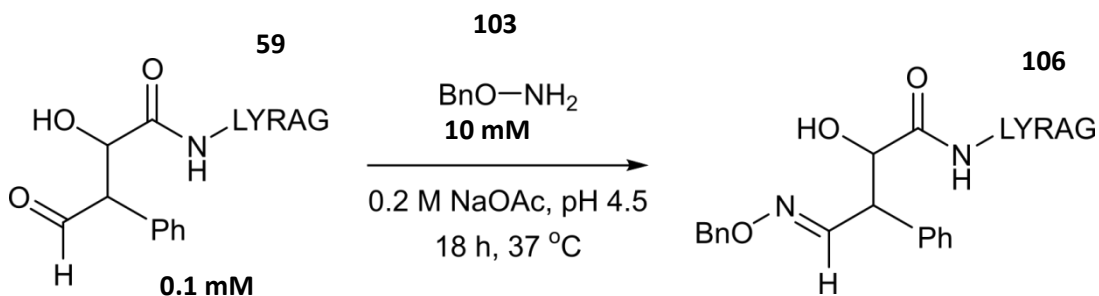
A 20 µL aliquot of a 5 mM aldol-modified-LYRAG **59** stock in MQ H_2O was charged with 879 µL of 0.1 M NaOAc, pH 4.5. The solution was then charged with 1 µL of *O*-benzylhydroxylamine **103**. The reaction was vortexed, and incubated 37 °C overnight without further agitation. Successful conversion to dually modified peptide **106** was confirmed by LC-MS analysis.

Synthesis of aldol-*iso*-Pictet-Spengler-LYRAG 107
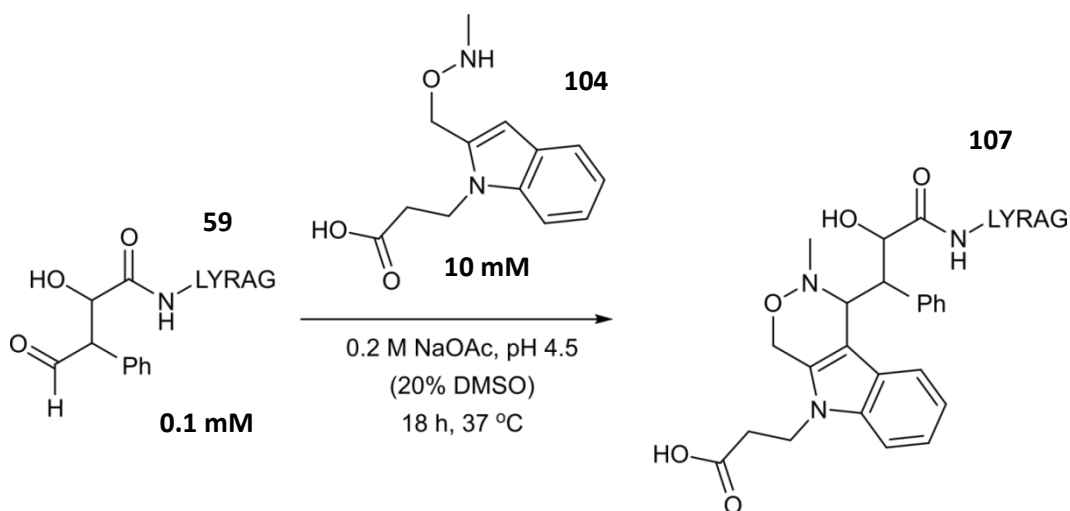
A 20 µL aliquot of an aldol-modified-LYRAG **59** stock in MQ H_2O was charged with 780 µL of 0.1 M NaOAc. The solution was then charged with 200 µL of a 50 mM indole **104** stock solution in DMSO. The reaction was vortexed, and incubated

37 °C overnight without further agitation. Successful conversion to dually modified peptide **107** was confirmed by LC-MS analysis.

Synthesis of aldol-ABAO-LYRAG 108
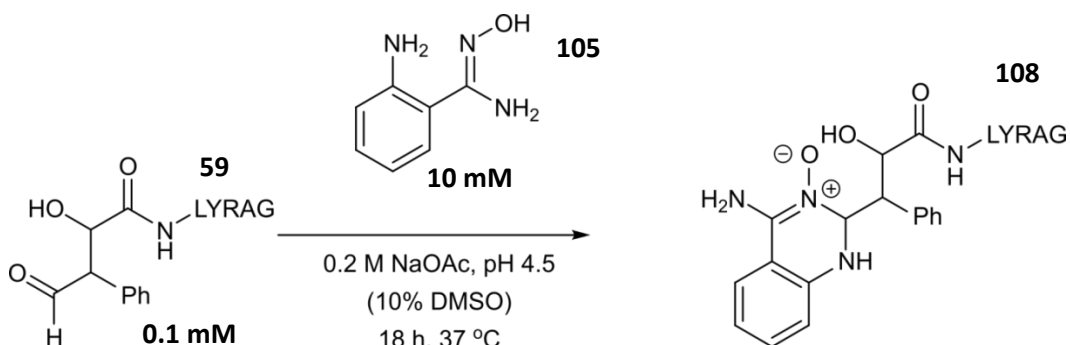
A 20 µL aliquot of a 5 mM aldol-modified-LYRAG **59** stock in MQ H_2O was charged with 780 µL of 0.1 M NaOAc, pH 4.5. The solution was then charged with 100 µL of ABAO **105** stock solution in DMSO. The reaction was vortexed, and incubated 37 °C overnight without further agitation. Successful conversion to dually modified peptide **108** was confirmed by LC-MS analysis.

Screening of aniline catalysts for oxime ligation
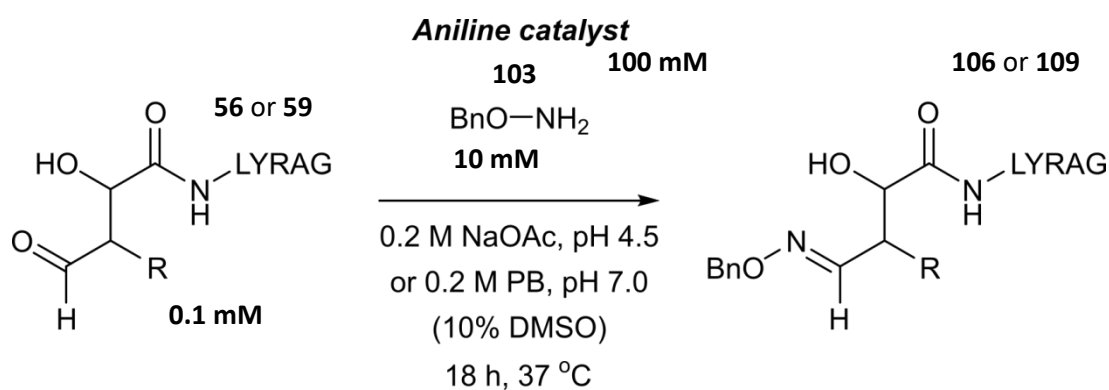
A 10 µL aliquot of a 5 mM **56** or **59** stock in MQ H_2O was charged with 879 µL of 0.2 M NaOAc, pH 4.5, or with 879 µL of 0.2 M PB pH 7.5. The solution was then charged with 1 µL of *O*-benzylhydroxylamine **103**, and then charged with 100 µL of

1M aniline catalyst in DMSO. The reaction was vortexed, and incubated 37 °C overnight without further agitation. Successful conversion to dually modified peptide **106** or **109** was confirmed by LC-MS analysis.

Screening effects of pH on aniline catalysed oxime ligation of 60
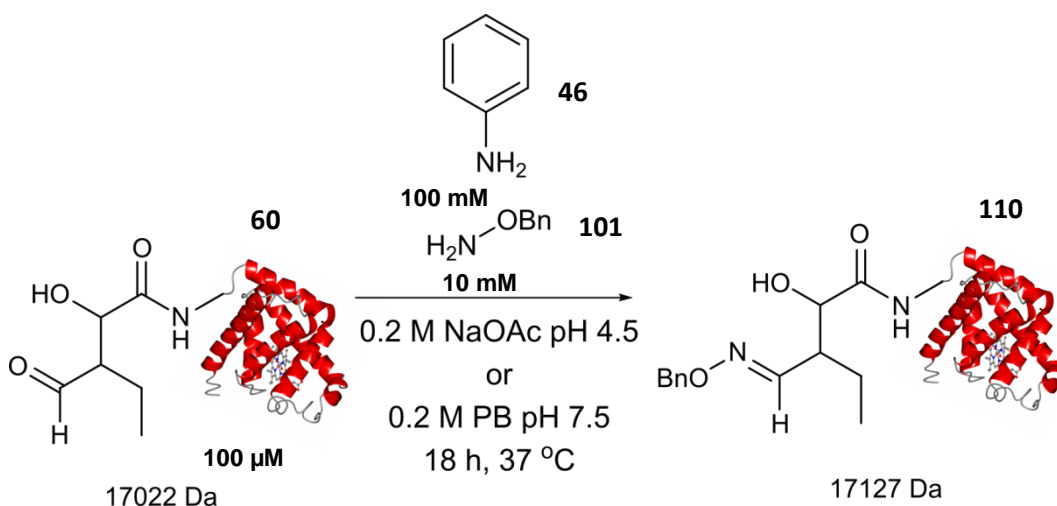
A 100 µL aliquot of aldol-modified myoglobin **60** (prepared as described earlier) in 0.2 NaOAc pH 4.5, or in 0.2 M PB pH 7.5, containing 100 mM aniline **46**, was charged with 100 µL a 20 mM aminooxy **101** stock solution in 0.2 M NaOAc pH 4.5, or in 0.2 M PB pH 7.5. Following mixing by pipetting, the reaction was allowed to sit at 37 °C for 18 h without further agitation. Conversion to the dually modified product **101** was assessed by LC-MS analysis.

Synthesis of fluorescently labelled, biotinylated thioredoxin 119
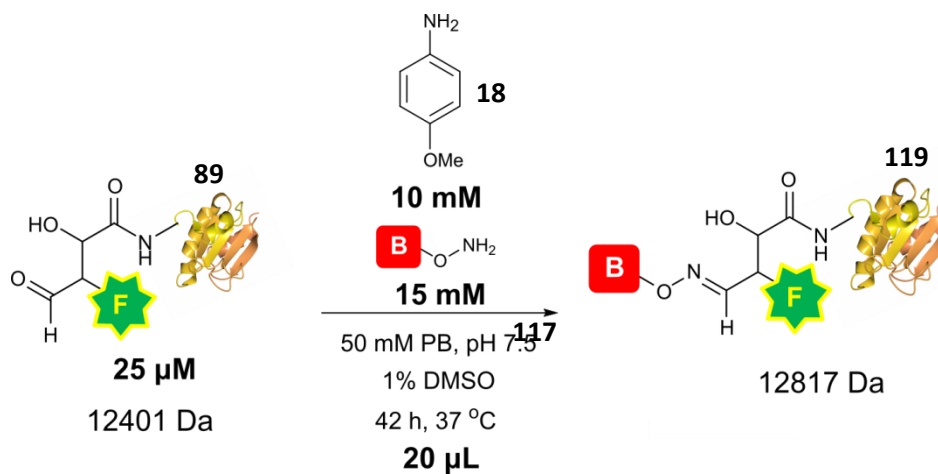
A 120 µL aliquot of fluorescently labelled thioredoxin **89** (prepared as described previously) was desalted using a PD SpinTrap G-25 column (GE Healthcare Life Sciences), eluting with 5 mM PB pH 7.5. A 10 µL aliquot of desalted protein was then charged with 3.8 µL of 0.2 M PB pH 7.5, and 4.8 µL of MQ H_2O. The solution was then charged with 1.2 µL of a 250 mM aminooxy biotin **117** stock solution in 50 mM PB pH 7.5 (pH adjusted to pH 7.5 using 2M NaOH), and then charged with 0.2 µL of a 1M *p*-anisidine **18** stock solution in DMSO. Following mixing by pipetting, the reaction was allowed to sit at 37 °C for 42 h without further agitation. Successful conversion to dually modified protein **119** was confirmed by LC-MS.

Synthesis of azide labelled, biotinylated thioredoxin 120
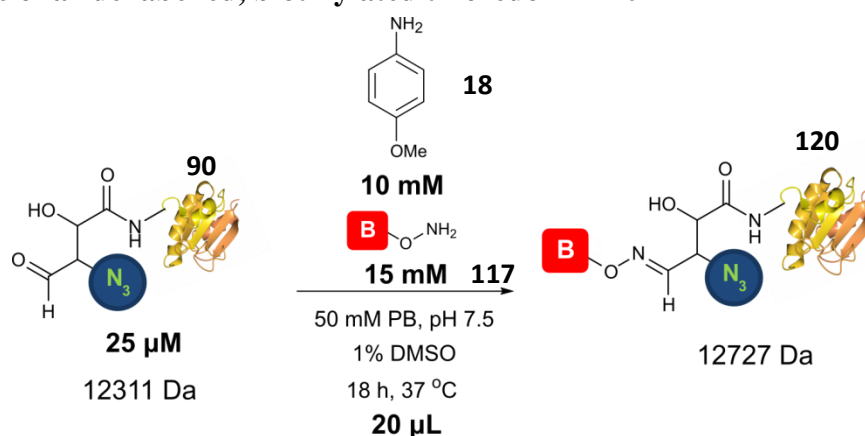
A 120 µL aliquot of 50 µM azide labelled thioredoxin **90** (prepared as described previously) was desalted using a PD SpinTrap G-25 column (GE Healthcare Life Sciences), eluting with 5 mM PB pH 7.5. A 10 µL aliquot of desalted protein was then charged with 3.8 µL of 0.2 M PB pH 7.5, and 4.8 µL of MQ H_2O. The solution was then charged with 1.2 µL of a 250 mM aminooxy biotin stock **117** solution in 50 mM PB pH 7.5 (pH adjusted to pH 7.5 using 2M NaOH), and then charged with 0.2 µL of a 1M *p*-anisidine **18** stock solution in DMSO. Following mixing by pipetting, the reaction was allowed to sit at 37 °C for 18 h without further agitation. Successful conversion to dually modified protein **120** was confirmed by Western Blot.

Synthesis of fluorescently labelled, PEGylated myoglobin 26
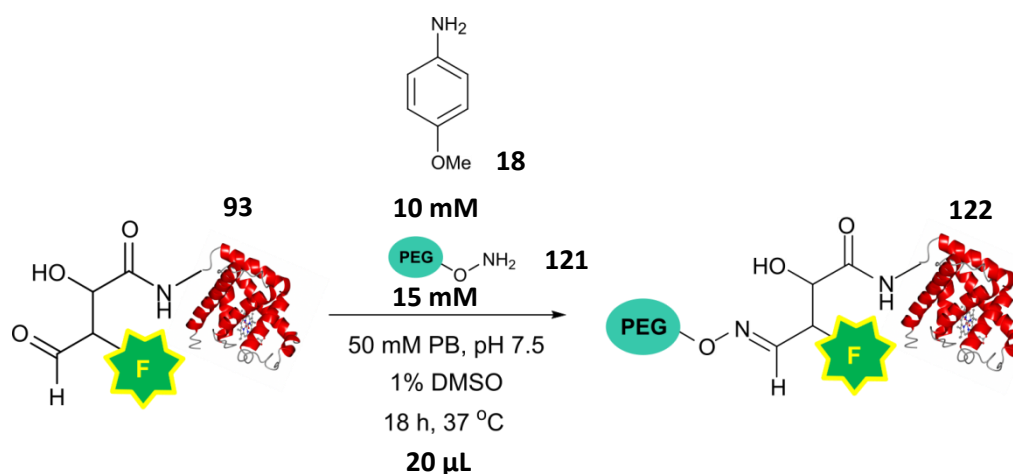
A 120 µL aliquot of 100 µM fluorescently labelled myoglobin **93** (prepared as described previously) was desalted using a PD SpinTrap G-25 column (GE Healthcare Life Sciences), eluting with 5 mM PB pH 7.5. A 10 µL aliquot of desalted protein was then charged with 3.8 µL of 0.2 M PB pH 7.5, and 4.8 µL of MQ H$_2$O. The solution was then charged with 1.2 µL of a 250 mM aminooxy PEG 2K **121** stock solution in 50 mM PB pH 7.5 (pH adjusted to pH 7.5 using 2M HCl), and then charged with 0.2 µL of a 1 M *p*-anisidine **18** stock solution in DMSO. Following mixing by pipetting, the reaction was allowed to sit at 37 °C for 18 h without further agitation. Successful labelling to give dually modified protein **122** was confirmed SDS PAGE analysis.

Synthesis of azide labelled, biotinylated myoglobin 118
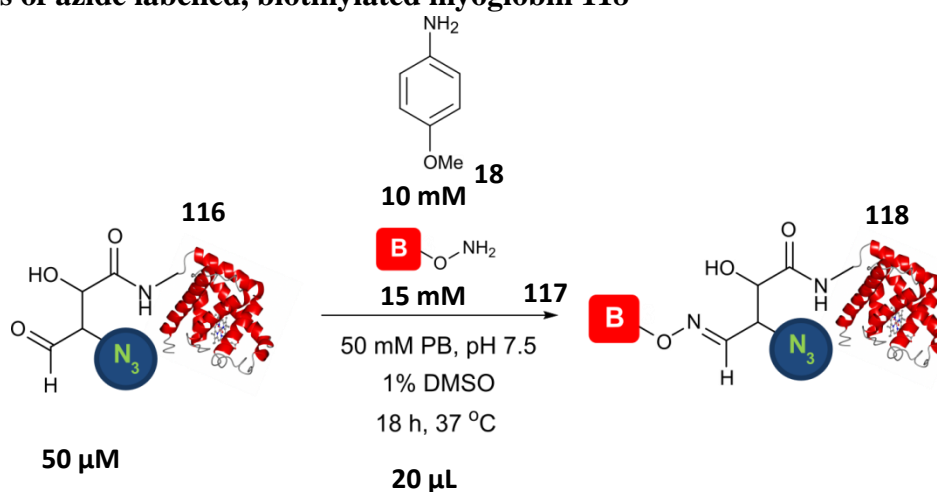
A 120 µL aliquot of 100 µM azide labelled myoglobin **116** (prepared as described previously) was desalted using a PD SpinTrap G-25 column (GE Healthcare Life Sciences), eluting with 5 mM PB pH 7.5. A 10 µL aliquot of desalted protein was then charged with 3.8 µL of 0.2 M PB pH 7.5, and 4.8 µL of MQ H_2O. The solution was then charged with 1.2 µL of a 250 mM aminooxy biotin stock **117** solution in 50 mM PB pH 7.5 (pH adjusted to pH 7.5 using 2M HCl), and then charged with 0.2 µL of a 1 M *p*-anisidine **21** stock solution in DMSO. Following mixing by pipetting, the reaction was allowed to sit at 37 °C for 18 h without further agitation. Successful coversion to give dually modified protein **118** was confirmed by LC-MS analysis.

Chemical myristoylation of HASPA
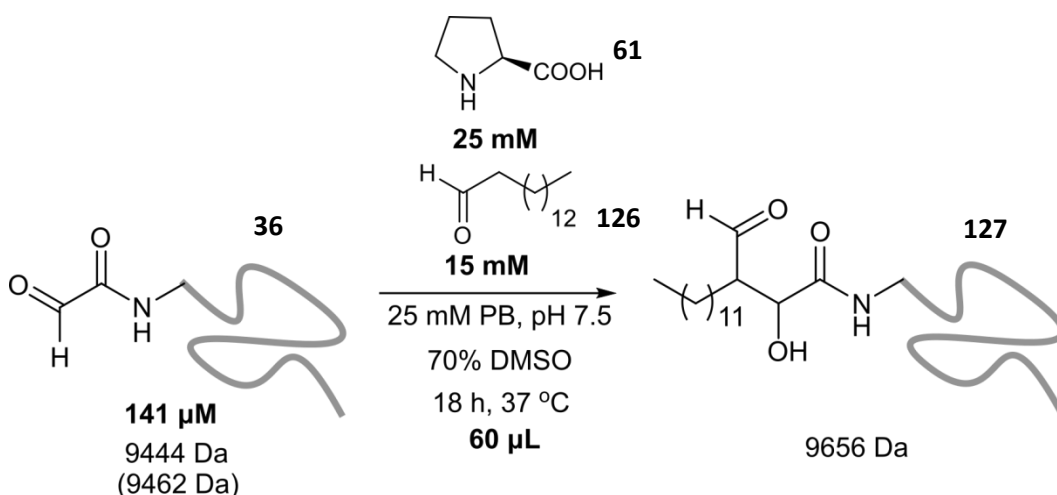
A 17 µL aliquot of 500 µM glyoxyl-HASPA **36** stock (prepared as described earlier) in 25 mM PB pH 7.5 was charged with 1 µL of a 1.5 M L-proline **61** stock solution in 25 mM PB pH 7.5. The solution was then charged with 6 µL of DMSO, and then charged with 36 µL of a 25 mM tetradecanal **126** stock solution in DMSO. Following mixing by pipetting, the reaction was incubated at 37 °C overnight without further agitation. Quantitative labelling to chemically myristoylated HASPA **127** was confirmed by LC-MS analysis (note elimination of β-hydroxyl to afford enone of **127** is also observed). Samples were then diluted to >20% DMSO content,

purified via PD MiniTrap G-25 columns (GE Healthcare Life Sciences), eluting MQ H$_2$O, and subsequently lyophilised to give a white powder (stored at -80 $^{\circ}$C).

Chemical myristoylation of 15N labelled HASPA

Chemical myristoylation of glyoxyl-[15N]HASPA(G1S) **36-15N** was identical to that of chemical myristoylation of unlabelled glyoxyl-HASPA **36**.

Synthesis of dually acylated HASPA 33
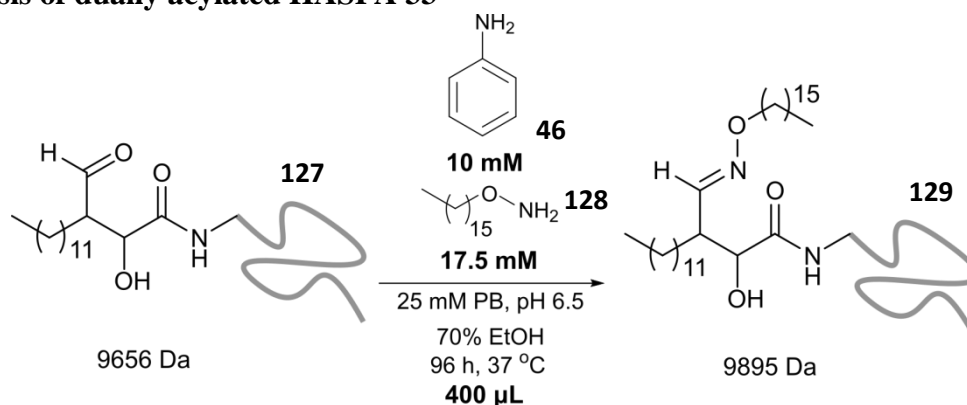
Prior to dual modification, samples of chemically myristoylated HASPA **127** (prepared as described earlier) were pooled to give an estimated maximum protein content of 160 µg (based on initial HASPA protein concentration). The pooled samples were diluted to >20% DMSO content, purified via PD MiniTrap G-25 columns (eluting into water), and subsequently lyophilised to give a white powder that was stored at -80 $^{\circ}$C until required. For the dual acylation of HASPA, the lyophilised aliquot of chemically myristoylated HASPA **127** was resuspended in 1 x PBS buffer (120 µL, pH 7.4), and then buffered exchanged using a PD SpinTrap G-25 column (GE Healthcare Life Sciences, eluting into 25 mM PB pH 6.5). The solution was then charged with 280 µL of 25 mM palmitoyl aminooxy **128** in EtOH, and then charged with 3.6 µL of aniline **46**. The solution was briefly vortexed, and the reaction was allowed to sit at 37 $^{\circ}$C for 96 h without further agitation. After 96h successful conversion to **129** was achieved as judged by LC-MS.

Screening of aniline catalysts for retro-aldol activity on peptides
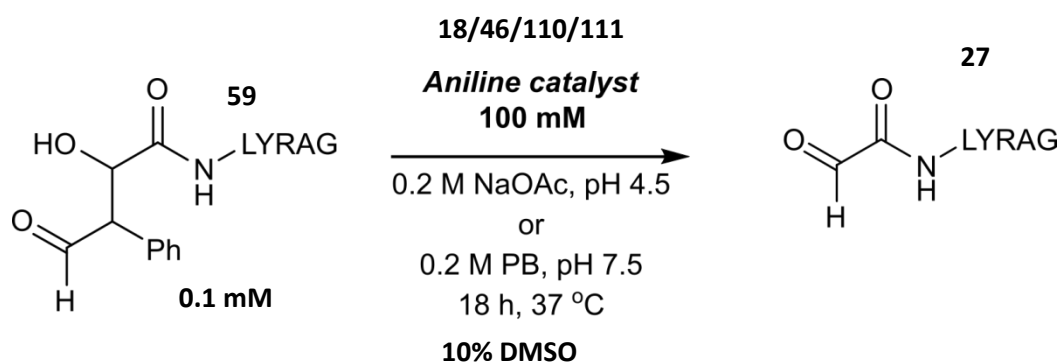
A 10 µL aliquot of a 5 mM stock of aldol-modified-LYRAG **59** in 0.2 M NaOAc pH 4.5 or in 0.2 M PB pH 7.5 was charged with 440 µL of 0.2 M NaOAc pH 4.5 or 440 µL of 0.2 M PB pH 7.5. The solution was then charged with 50 µL of a 1M stock solution of aniline catalyst **18**, **46**, **110**, or **111** in DMSO. The reaction was vortexed, and incubated at 37 °C overnight without further agitation. Successful conversion to glyoxyl-LYRAG **27** was judged by LC-MS analysis.

Screening of buffer concentration and the retro aldol activity on peptides
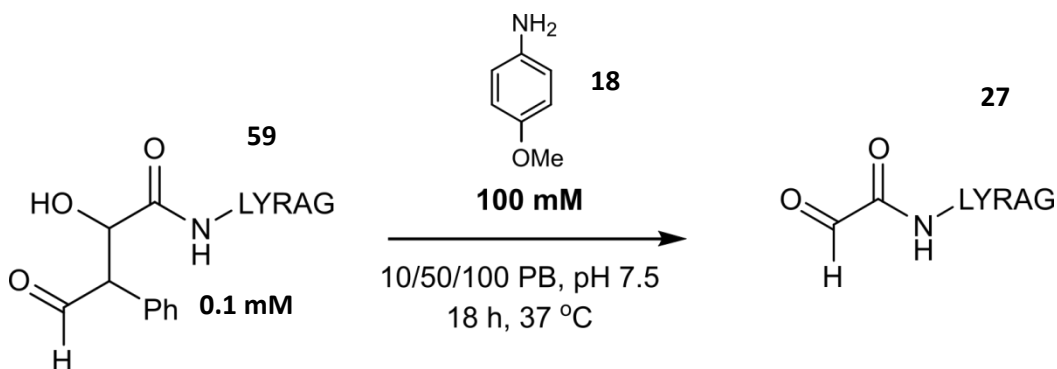
A 1 µL aliquot of a 5 mM stock of aldol-modified-LYRAG **59** in H_2O was charged with a volume 0.2 M PB pH 7.5 and a volume of H_2O, followed by addition of 5 µL of a 1 M stock solution of *p*-anisidine **18**, in such a manner that the final concentration of PB was 10 mM, 50 mM, or 100 mM. The reactions were vortexed, and incubated at 37 °C overnight without further agitation. Successful conversion to glyoxyl-LYRAG **27** was judged by LC-MS analysis.

Screening of aniline catalyst for retro-aldol activity on proteins
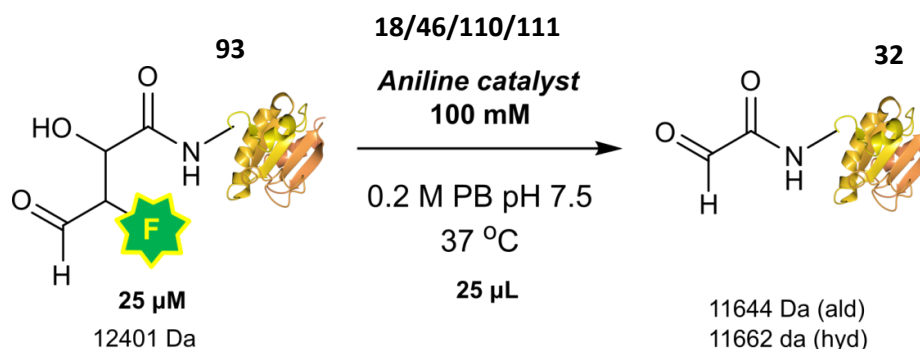
A 12.5 µL aliquot of fluorescently labelled thioredoxin **93** in 0.2 M PB pH 7.5 was charged with 10 µL 0.2 M PB pH 7.5, followed by addition of 5 µL of a 1M stock solution of aniline catalyst **18**, **46**, **110**, or **111** in DMSO. Following mixing by pipetting, the reaction was incubated at 37 ºC for 18 h (or 42 h) without further agitation. Successful conversion to glyoxyl-thioredoxin **32** was judged by LC-MS analysis.

Synthesis of modified mannich product 136
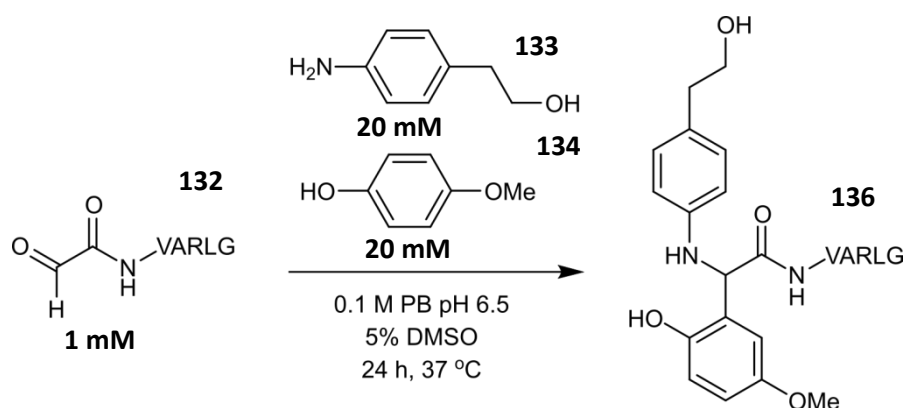
A 200 µL aliquot of a 5 mM stock of glyoxyl-VARLG **132** was charged with 500 µL of 0.2 M PB pH 6.5, followed by addition of 250 µL of H$_2$O. The solution was then charged with 10 µL of DMSO, followed by addition of 20 µL of a 1 M solution of aniline **133** in DMSO and 20 µL of a 1 M solution of phenol **134** in DMSO. The reaction was vortexed, and incubated at 37 ºC for 24 h without further agitation. Modification of glyoxyl-VARLG **132** was assessed by LC-MS analysis.

Synthesis of modified mannich product 137

A 200 µL aliquot of a 5 mM stock of glyoxyl-VARLG **132** in MQ H$_2$O was charged with 500 µL of 0.2 M PB pH 6.5, followed by addition of 250 µL of H$_2$O. The solution was then charged with 10 µL of DMSO, followed by addition of 20 µL of a 1 M solution of aniline **133** in DMSO and 20 µL of a 1 M solution of phenol **135** in DMSO. The reaction was vortexed, and incubated at 37 ℃ for 24 h without further agitation. Modification of glyoxyl-VARLG **132** was assessed by LC-MS analysis

Initial experiments towards a catalyst free aldol ligation
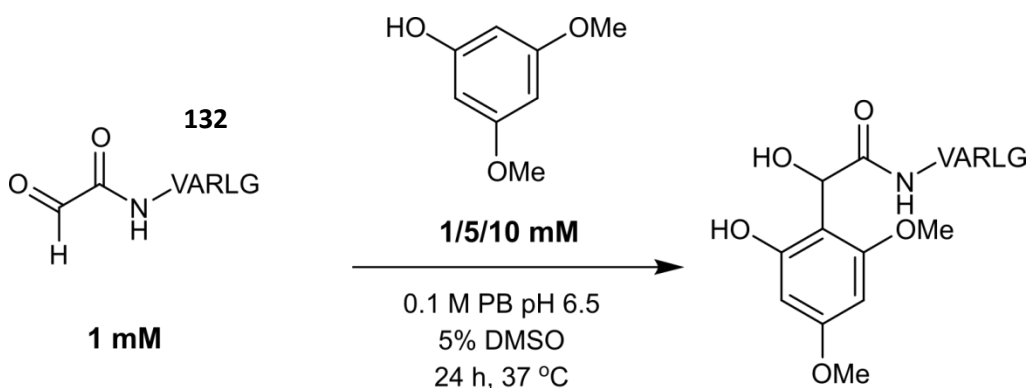
A 200 µL aliquot of a 5 mM stock of glyoxyl-VARLG **132** in MQ H$_2$Owas charged with 500 µL of 0.2 M PB pH 6.5, followed by addition of 250 µL of H$_2$O. The solution was then charged with a volume of DMSO followed by addition of a volume of of a 1 M solution of phenol **135** in DMSO, in such a manner that the final concentration of **135** was 1 mM, 5 mM, or 10 mM. The reaction was vortexed, and incubated at 37 ℃ for 24 h without further agitation. Modification of glyoxyl-VARLG **132** was assessed by LC-MS analysis.

Screening of pH in the modified Mannich bioconjugation protocol
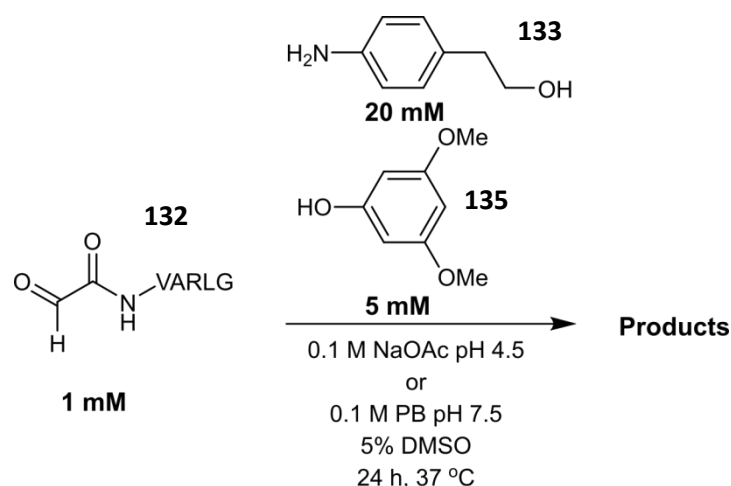
A 200 µL aliquot of a 5 mM stock of glyoxyl-VARLG **132** in MQ H_2O was charged with 500 µL of 0. 2 M NaOAc pH 4.5, or with 250 µL of 0.2M PB pH 7.5, followed by addition of 250 µL of H_2O. The solution was then charged with 25 µL of DMSO, followed by addition of 20 µL of a 1 M solution of aniline **133** in DMSO and 5 µL of a 1 M solution of phenol **135** in DMSO. The reaction was vortexed, and incubated at 37 °C for 24 h without further agitation. Modification of glyoxyl-VARLG **132** was assessed by LC-MS analysis.

Screening of pH in the phenol-mediated ligation
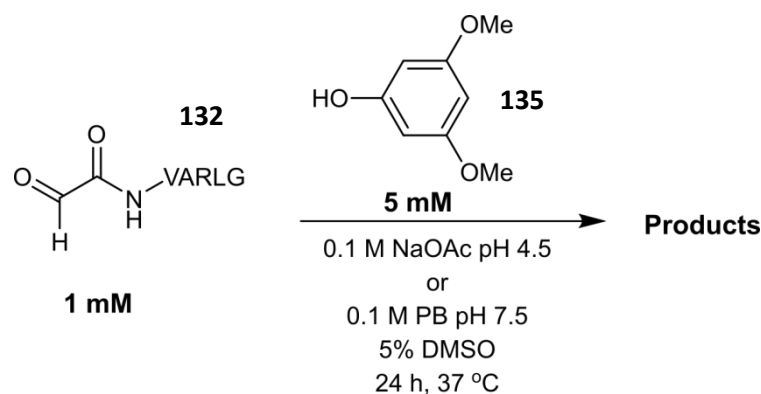
A 200 µL aliquot of a 5 mM stock of glyoxyl-VARLG **132** in MQ H_2O was charged with 500 µL of 0.2 M NaOAc pH 4.5, or with 500 µL of 0.2M PB pH 7.5, followed by addition of 250 µL of H_2O. The solution was then charged with 45 µL of DMSO, followed by addition of 5 µL of a 1 M solution of phenol **135** in DMSO. The reaction was vortexed, and incubated at 37 °C for 24 h without further agitation. Modification of glyoxyl-VARLG **132** was assessed by LC-MS analysis

Screening aromatic compounds for modification of α-oxo aldehydes
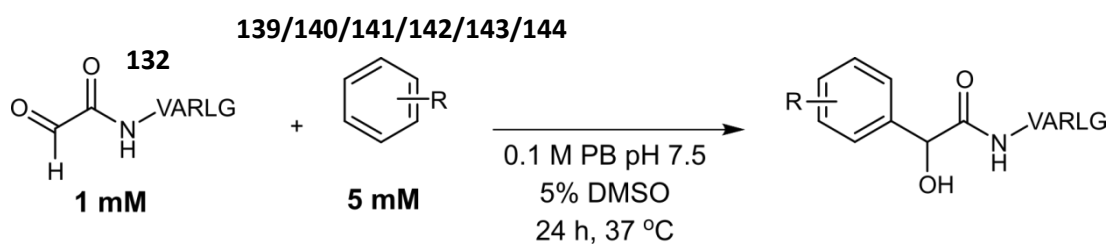
A 200 µL aliquot of a 5 mM stock of glyoxyl-VARLG **132** in MQ H$_2$O was charged with 500 µL of 0.2 M NaOAc pH 4.5, or with 500 µL of 0.2M PB pH 7.5, followed by addition of 250 µL of H$_2$O. The solution was then charged with 45 µL of DMSO, followed by addition of 5 µL of a 1 M solution of phenol in DMSO. The reaction was vortexed, and incubated at 37 °C for 24 h without further agitation. Modification of glyoxyl-VARLG **132** was assessed by LC-MS analysis.

Synthesis of anticipated PHILIPA product 148
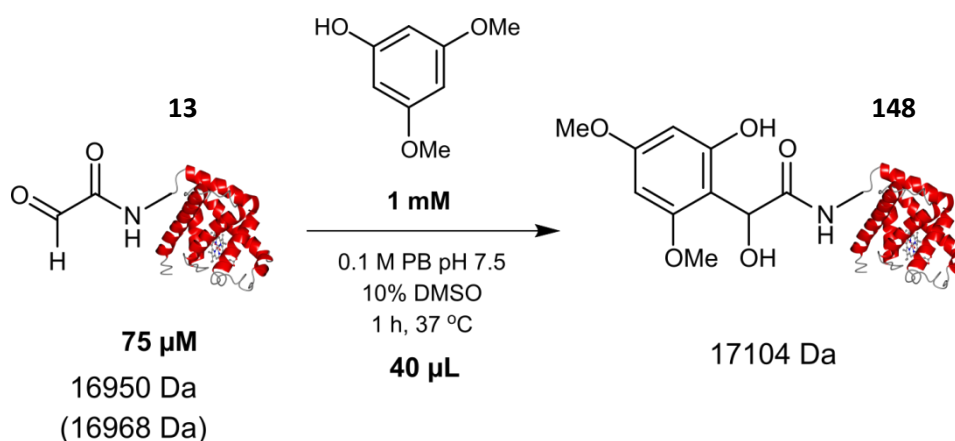
A 5 µL aliquot of 600 µM glyoxyl-myoglobin **13** stock (prepared as described earlier) in H$_2$O was charged with 20 µL of 0.2 M PB pH 7.5. The solution was then charged with 3.8 µL of DMSO, followed by addition of 200 mM phenol **135** stock solution in 25 DMSO. Following mixing by pipetting, the reaction was allowed to sit at 37 °C for 60 min without further agitation. Conversion to the hypothesised phenol product **148** was assessed by LC-MS analysis.

Mass spectrometry data of modified peptides

SLYRAG *26*
Calculated [M+H]+ = 666.36
Found [M+ H]+ = 666.32
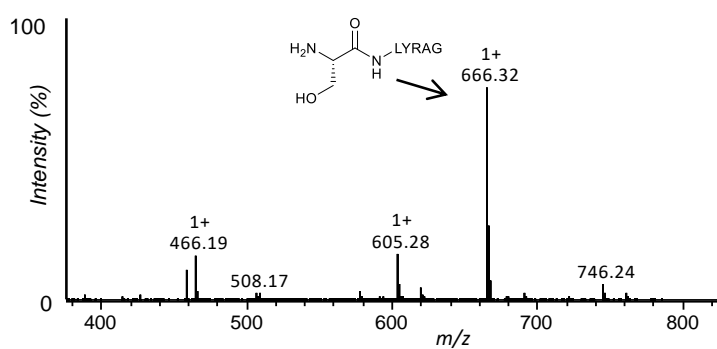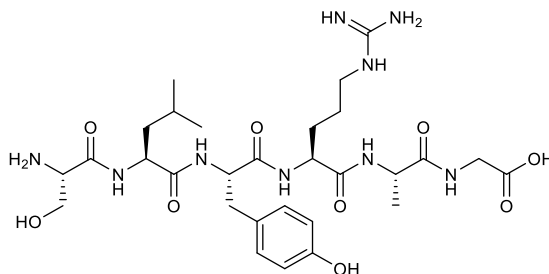
Glyoxyl-LYRAG *27*
Calculated [M+H]+ = 635.31 (ald), 653.32 (hyd)
Found [M+ H]+ = 635.41 (ald), 653.44 (hyd)
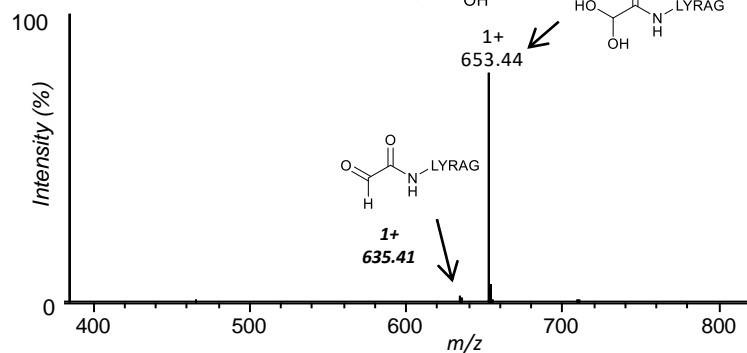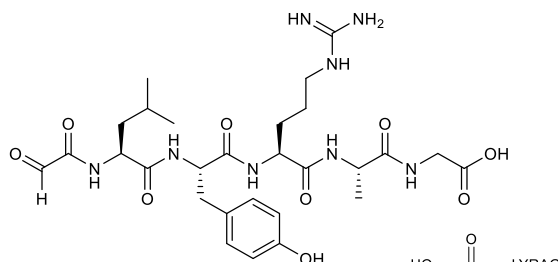
*Aldol-LYRAG **45***
Calculated [M+H]$^+$ = 679.33
Found [M+ H]$^+$ = 679.32
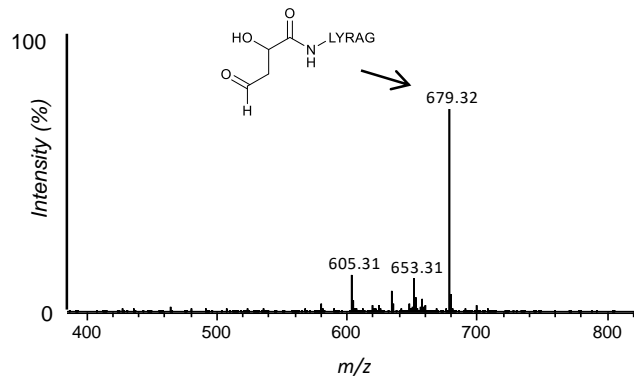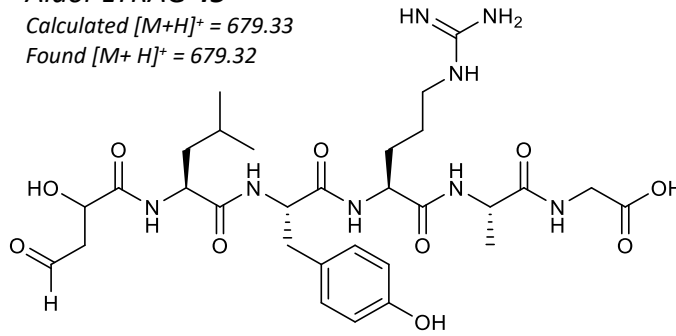
*Aldol-Oxime-LYRAG **48***
Calculated [M+H]$^+$ = 752.36
Found [M+ H]$^+$ = 752.34
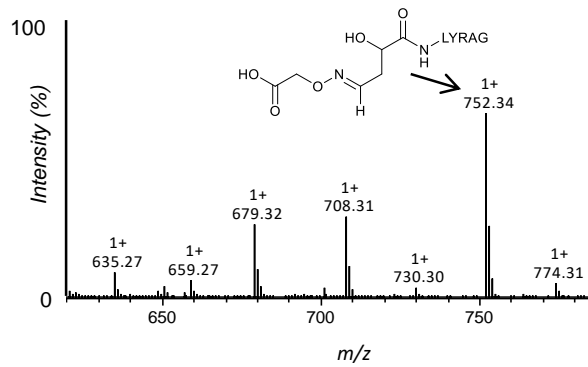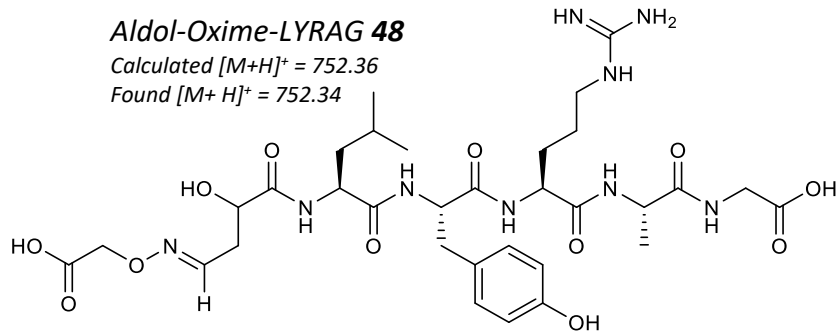
*Aldol-LYRAG **55***
Calculated [M+H]+ = 693.35
Found [M+ H]+ = 693.38
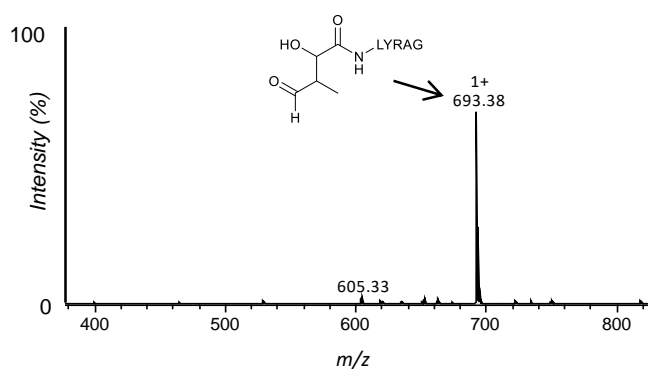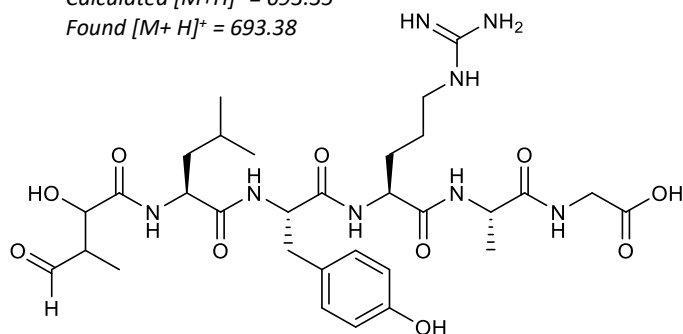
*Aldol-LYRAG **56***
Calculated [M+H]+ = 707.36
Found [M+ H]+ = 707.26
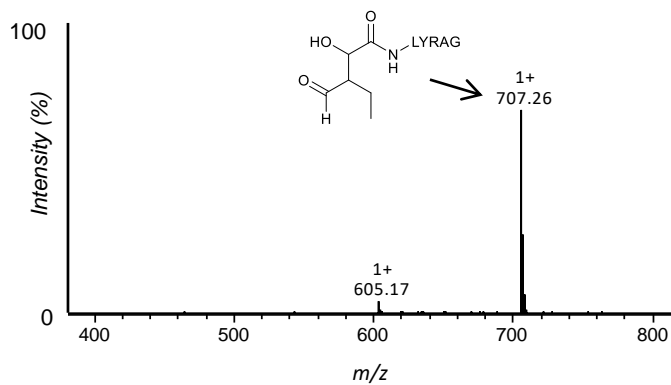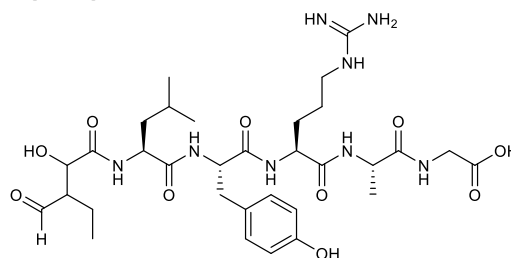
*Aldol-LYRAG **57***

Calculated [M+H]⁺ = 707.36
Found [M+ H]⁺ = 707.39
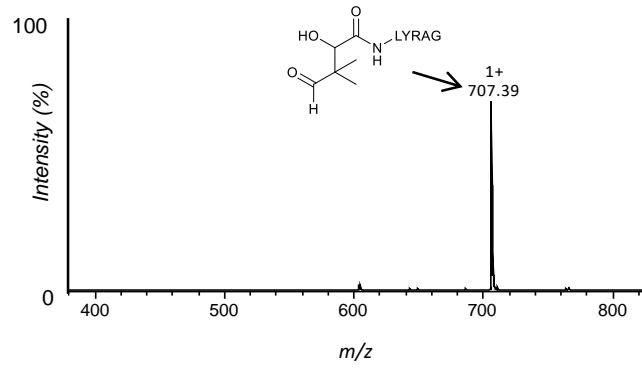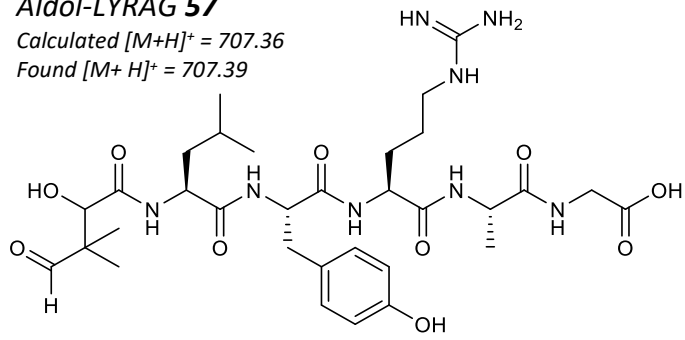
*Aldol-LYRAG **58***

Calculated [M+H]⁺ = 735.340
Found [M+ H]⁺ = 735.45
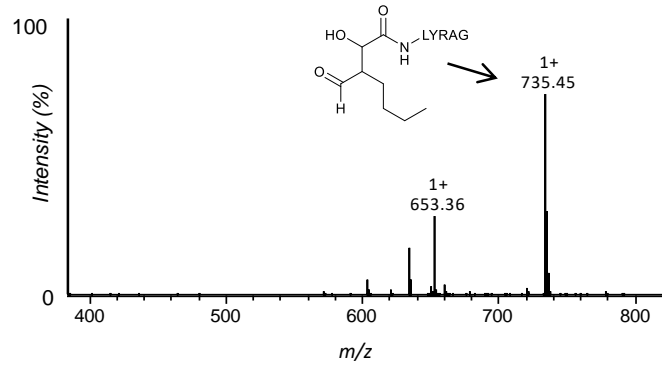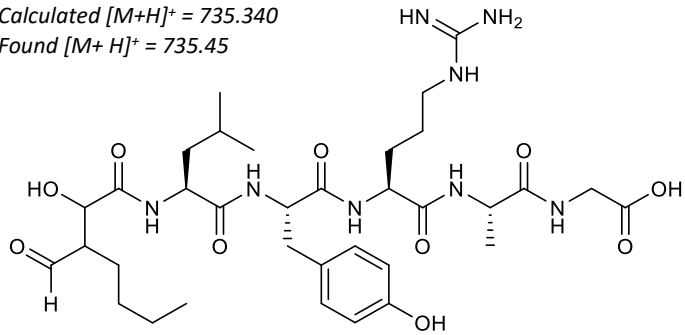
*Aldol-LYRAG **59***
Calculated [M+H]+ = 755.36
Found [M+ H]+ = 755.39
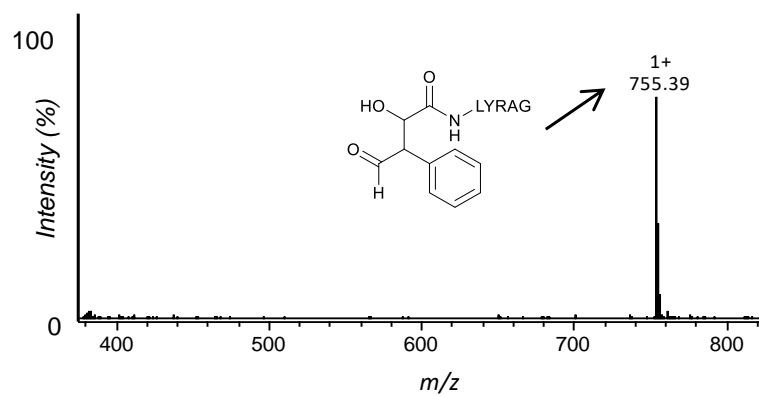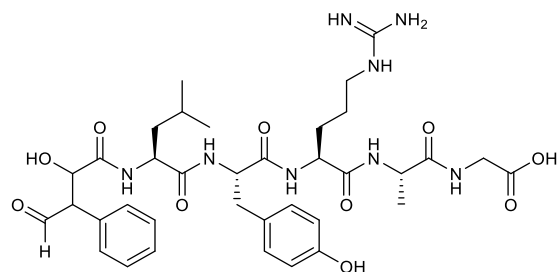
*Aldol-Oxime-LYRAG **106***
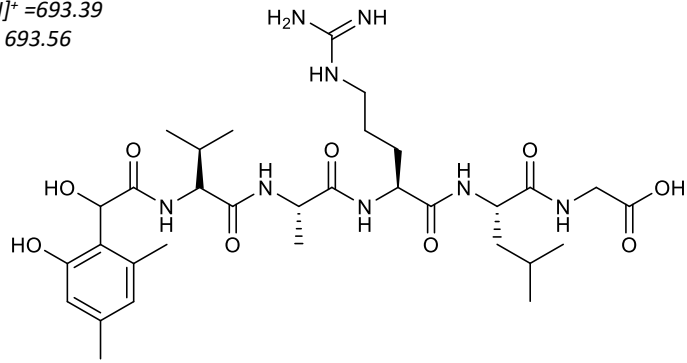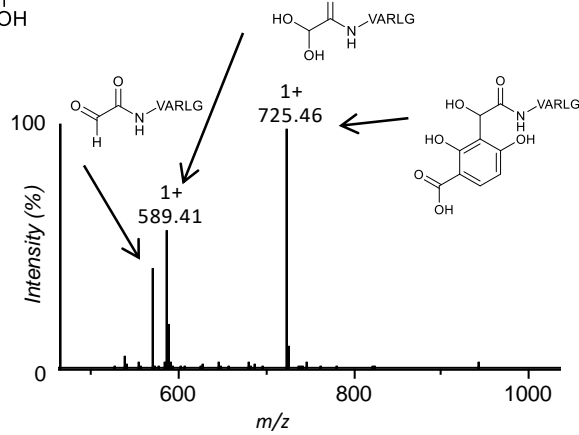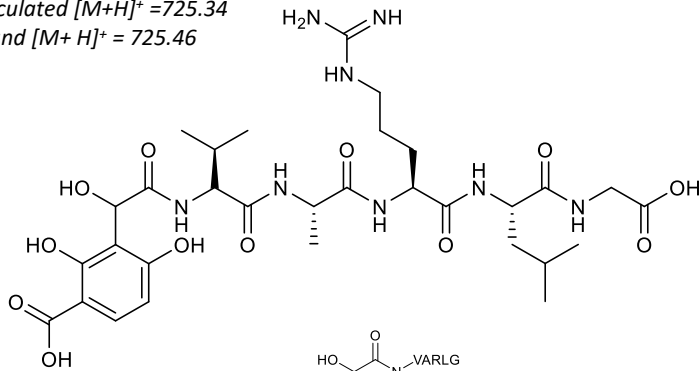*Calculated [M+H]+ = 860.42*
Found [M+ H]+ = 860.57
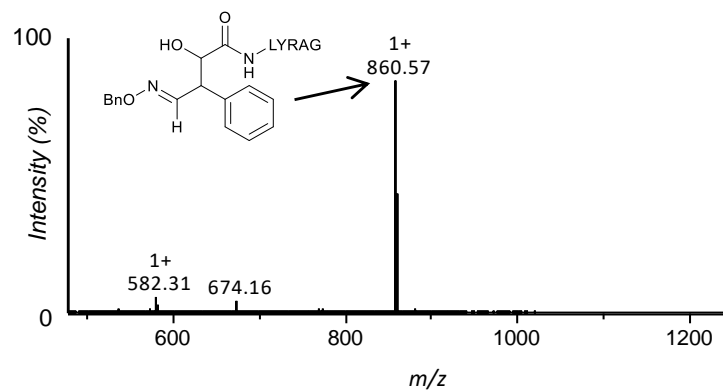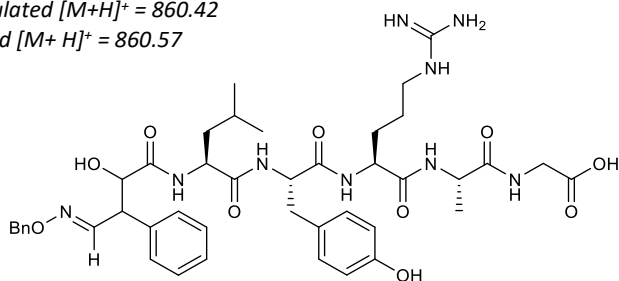
*Aldol-IPS **107***

Calculated [M+H]⁺ = 985.47
Found [M+ H]⁺ = 985.45

493.33 corresponds to [M+2H]²⁺

985.45 corresponds to [M+H]⁺
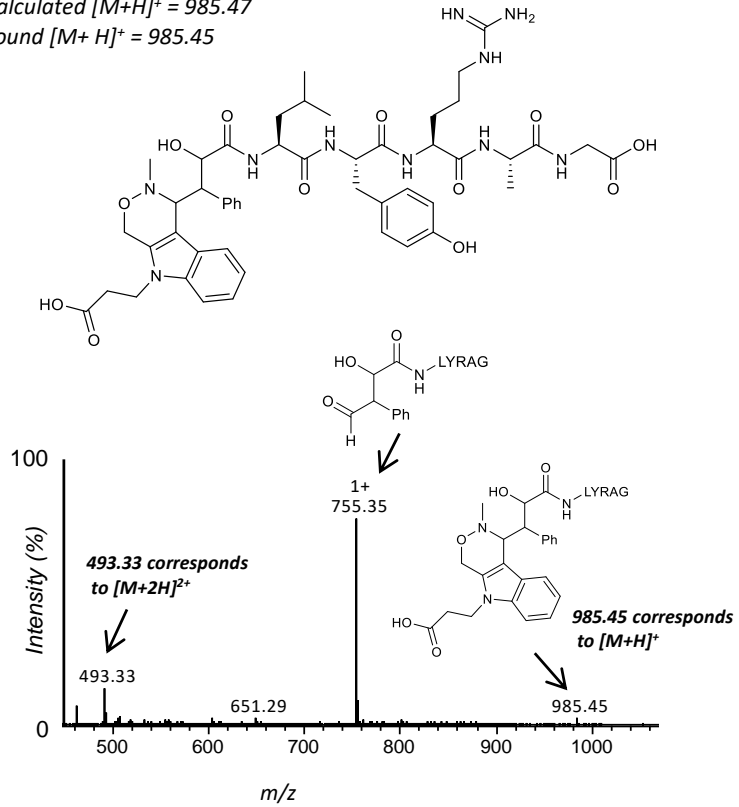
*Aldol-ABAO **108***

Calculated [M+H]⁺ = 888.43
Found [M+ H]⁺ = 888.44

444.76 corresponds to [M+2H]²⁺

888.44 corresponds to [M+H]⁺
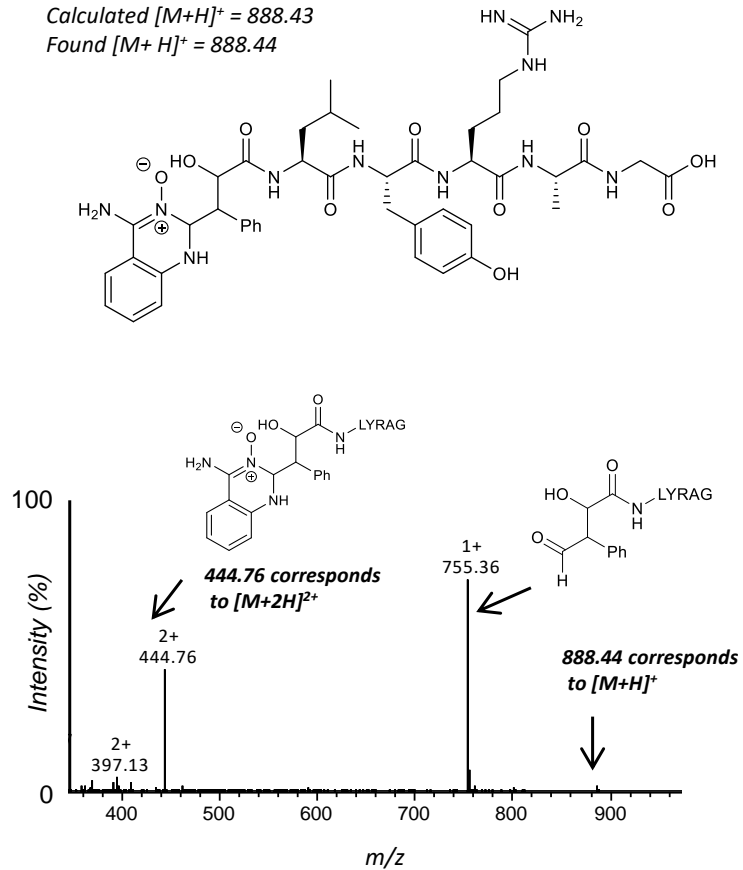
*Aldol-Oxime-LYRAG **109***
Calculated [M+H]$^+$ = 812.42
Found [M+ H]$^+$ = 812.40
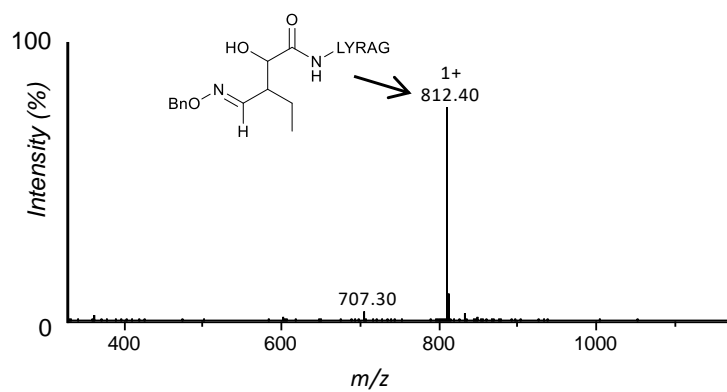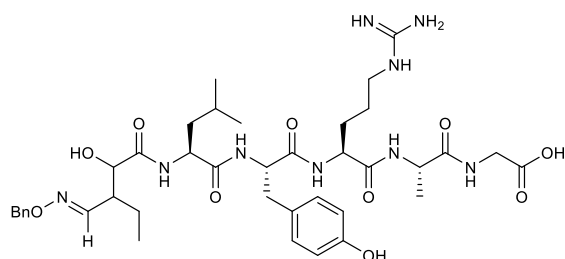
*Myristoylated-LYRAG **130***
Calculated [M+H]$^+$ = 847.52
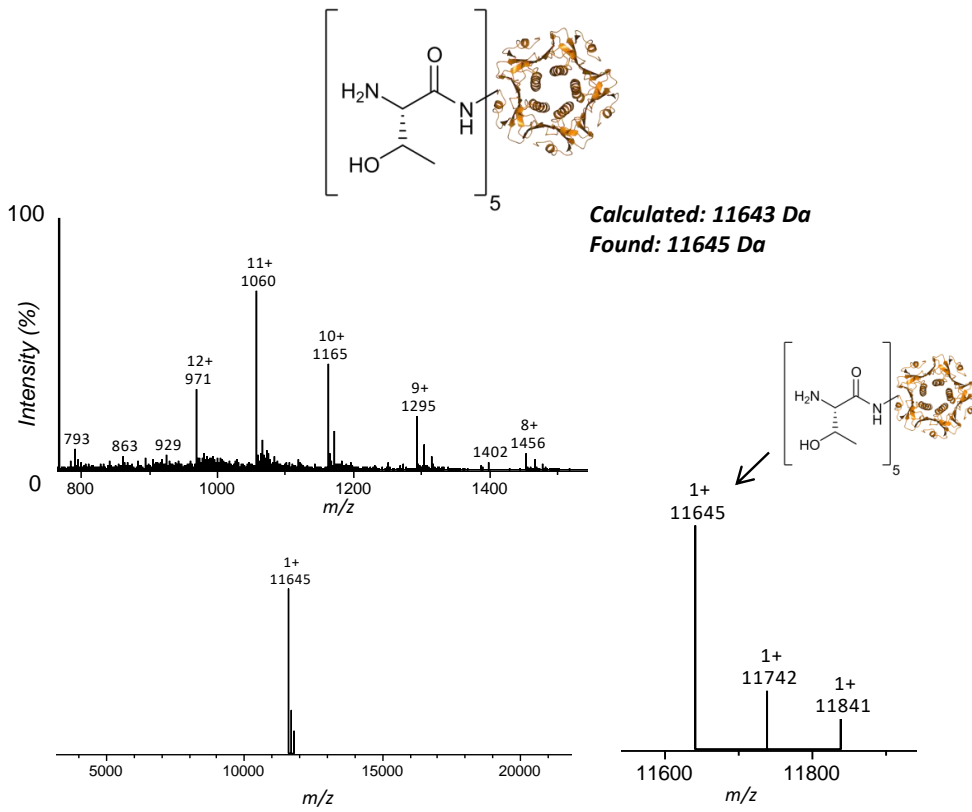*Found [M+ H]$^+$ = 847.72*
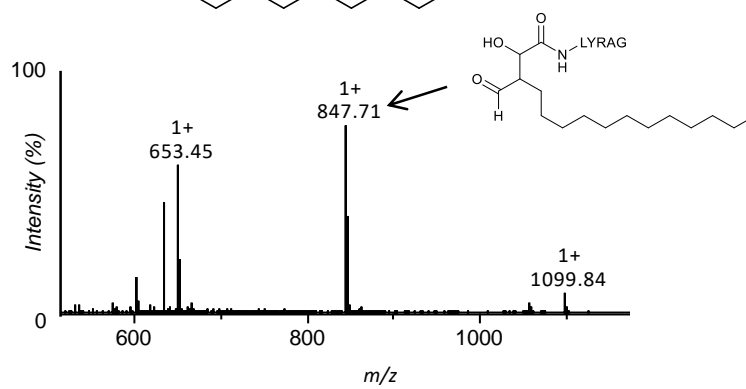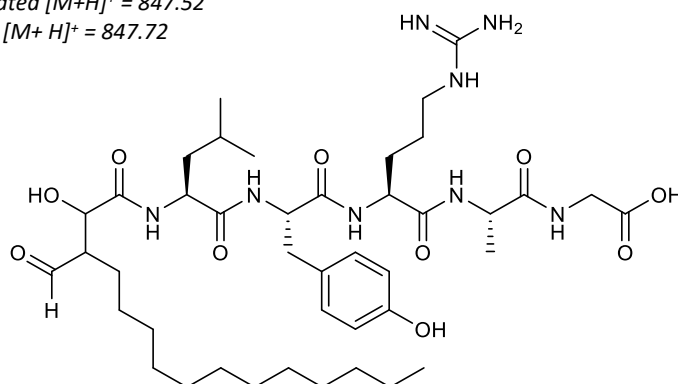
*SVARLG **151***
Calculated [M+H]$^+$ = 602.35
Found [M+ H]$^+$ = 602.26
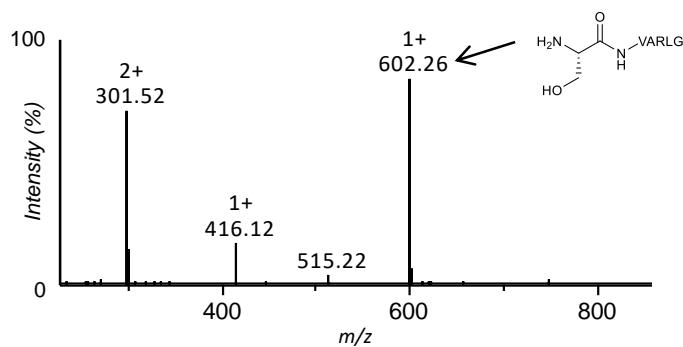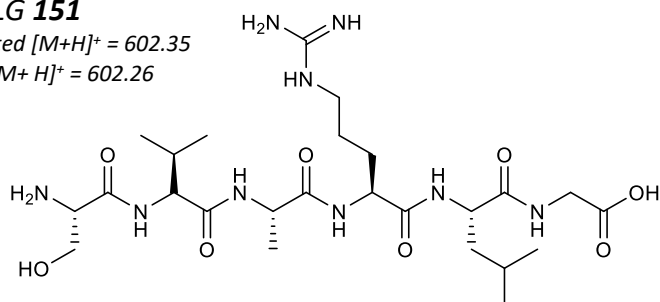
*Glyoxyl-VARLG **132***
Calculated [M+H]$^+$ = 571.31 (ald), 589.32 (hyd)
Found [M+ H]$^+$ = 571.32 (ald), 589.33 (hyd)
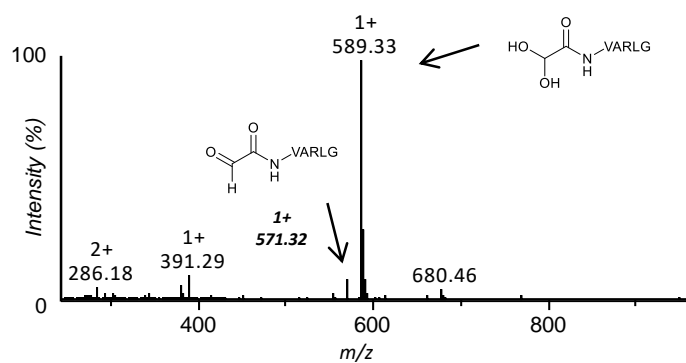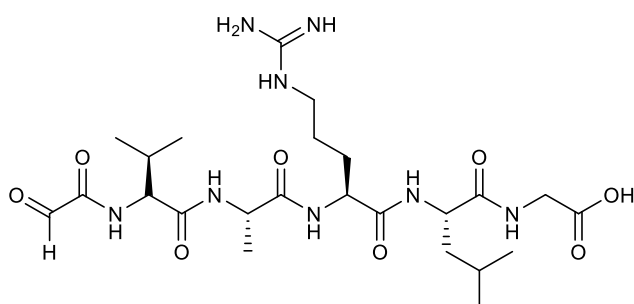
Mannich product **136**
Calculated [M+H]⁺ = 814.44
Found [M+ H]⁺ = 814.52
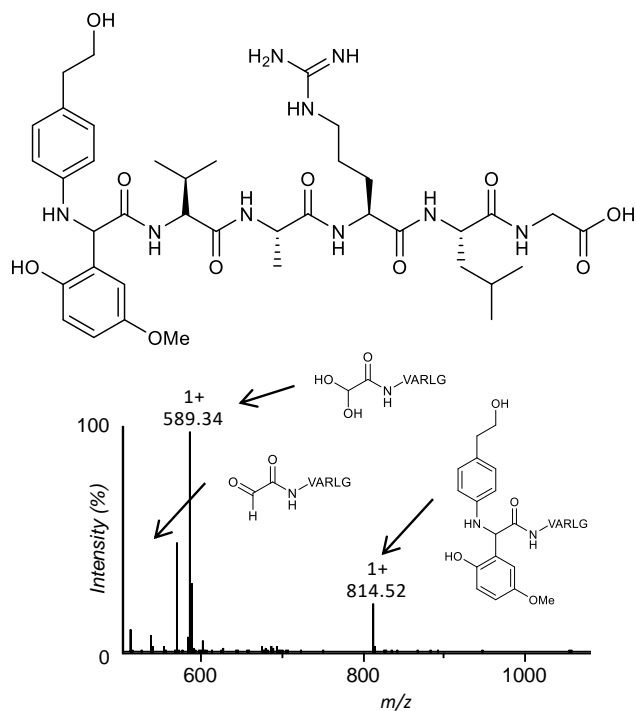
Mannich product **137**
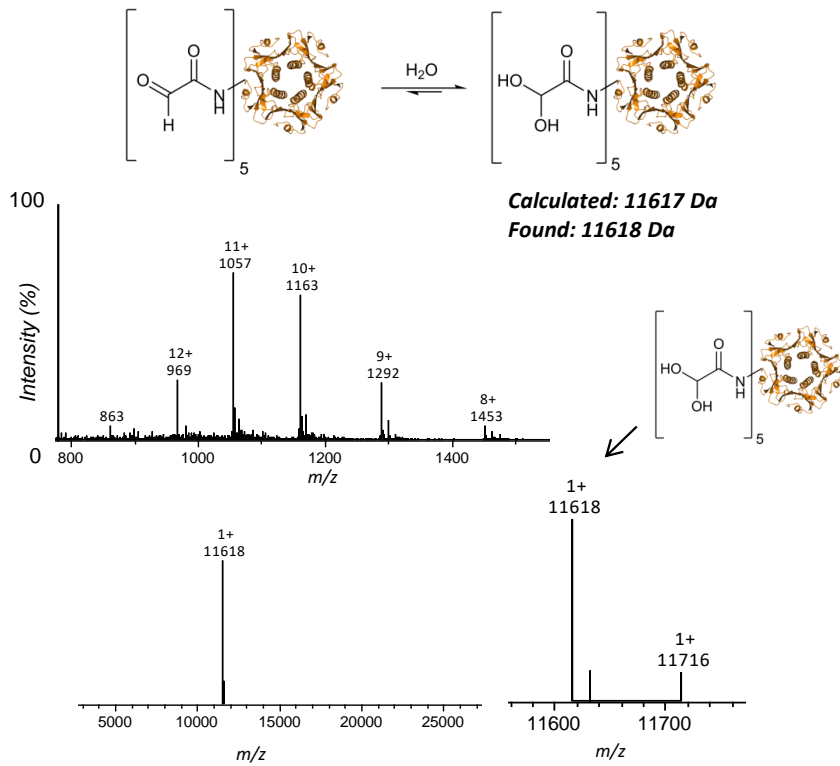*Calculated [M+H]⁺ = 844.45*
Found [M+ H]⁺ = 844.49
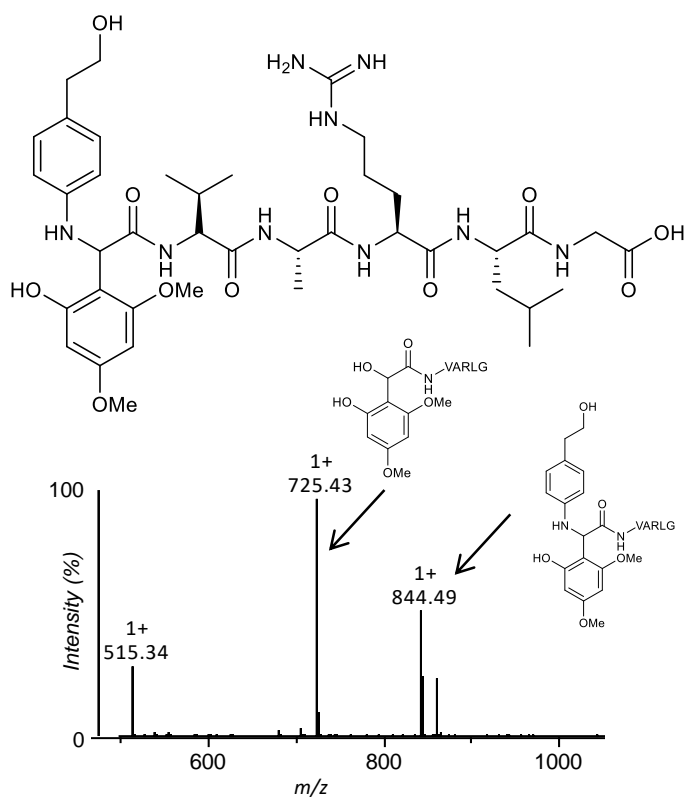
Phenol product 138
Calculated [M+H]⁺ =725.38
Found [M+ H]⁺ = 725.49
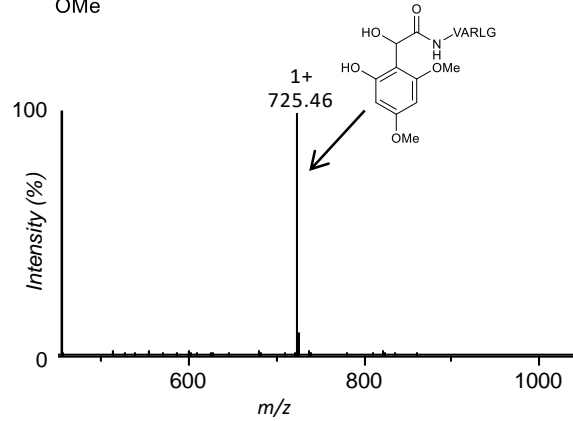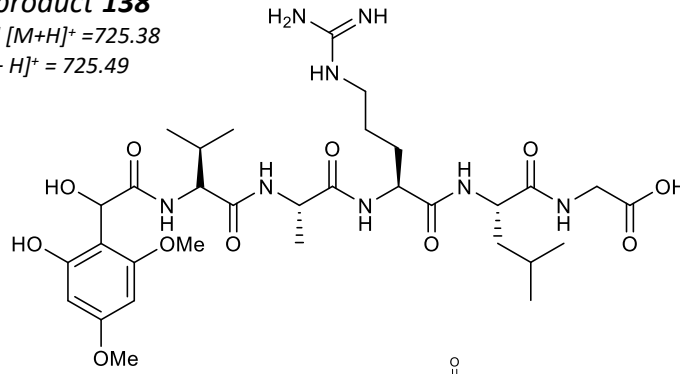
Phenol product 145
Calculated [M+H]⁺ =681.35
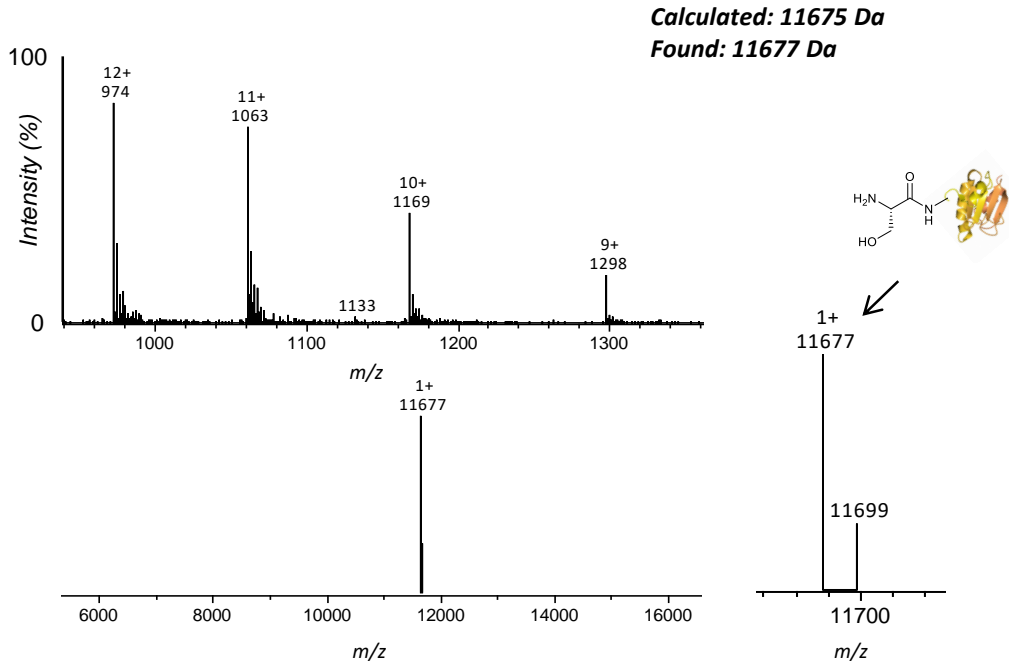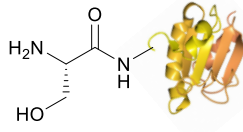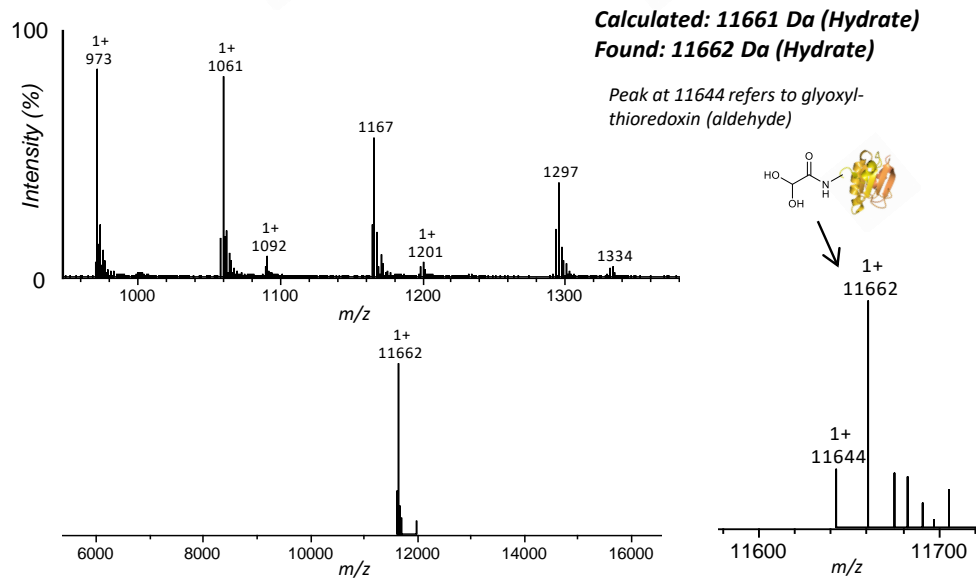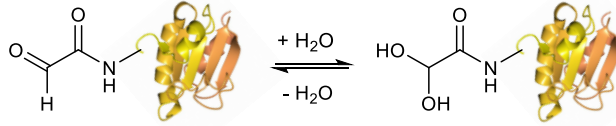*Found [M+ H]⁺ = 681.52*
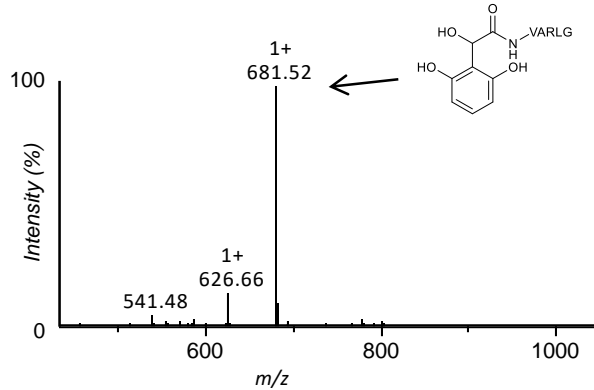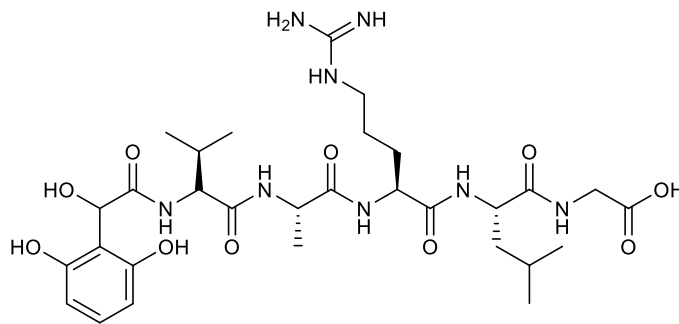
Possible products
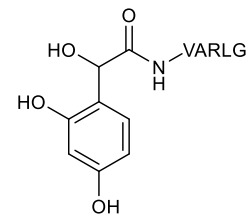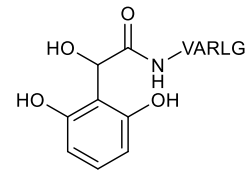
*Phenol product **146***

Calculated [M+H]⁺ =693.39
Found [M+ H]⁺ = 693.56
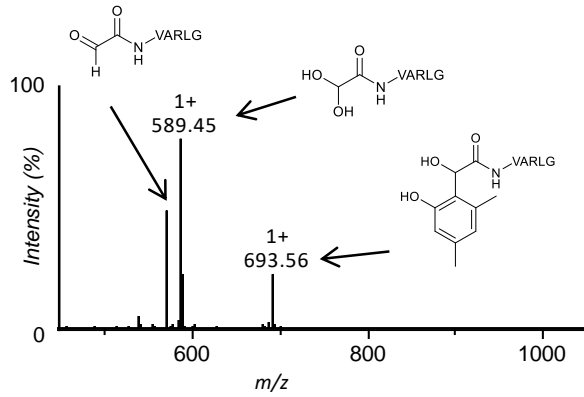
*Phenol product **147***

Calculated [M+H]⁺ =725.34
Found [M+ H]⁺ = 725.46

Possible products

Mass spectrometry data of proteins and modified proteins

Horse Heart Myoglobin **28**

Calculated: 16951 Da
Found: 16953 Da

Glyoxyl-myoglobin **13**

Calculated: 16968 Da (Hydrate)
Found: 16968 Da (Hydrate)

Peak at 16951 Da can refer to either myoglobin or glyoxyl-myoglobin (aldehyde) Peak at 17170 Da is an unknown, PLP related byproduct

CTB 29

Calculated: 11643 Da
Found: 11645 Da

Glyoxyl-CTB 30

Calculated: 11617 Da
Found: 11618 Da

Thioredoxin **31**

Calculated: 11675 Da
Found: 11677 Da

Glyoxyl-thioredoxin **32**

Calculated: 11661 Da (Hydrate)
Found: 11662 Da (Hydrate)

Peak at 11644 refers to glyoxyl-thioredoxin (aldehyde)

GFP 33

Calculated: 28669 Da
Found: 28677 Da

Glyoxyl-GFP 34

Calculated: 28656 Da (Hydrate)
Found: 28655 Da (Hydrate)

Hydrophilic acylated surface protein A (HASPA) 35

Calculated: 9445 Da
Found: 9449 Da

HASPA(G1S) 37

Calculated: 9475 Da
Found: 9477 Da

*Glyoxyl-HASPA **36** (generated from NaIO₄ oxidation of **37**)*

Calculated: 9462 Da (Hydrate)
Found: 9464 Da (Hydrate)

*[15N]HASPA(G1S) **37-15N***

Calculated: 9606 Da
Found: 9607 Da

*Glyoxyl-[15N]HASPA **S36-15N** (generated from NaIO₄ oxidation of **37-15N**)*
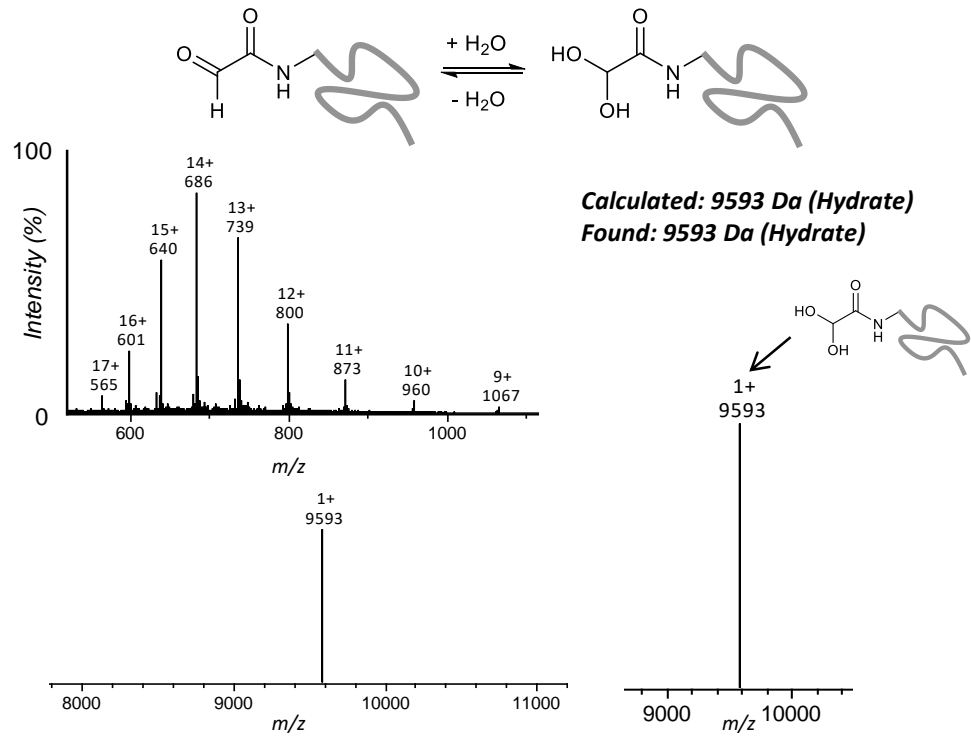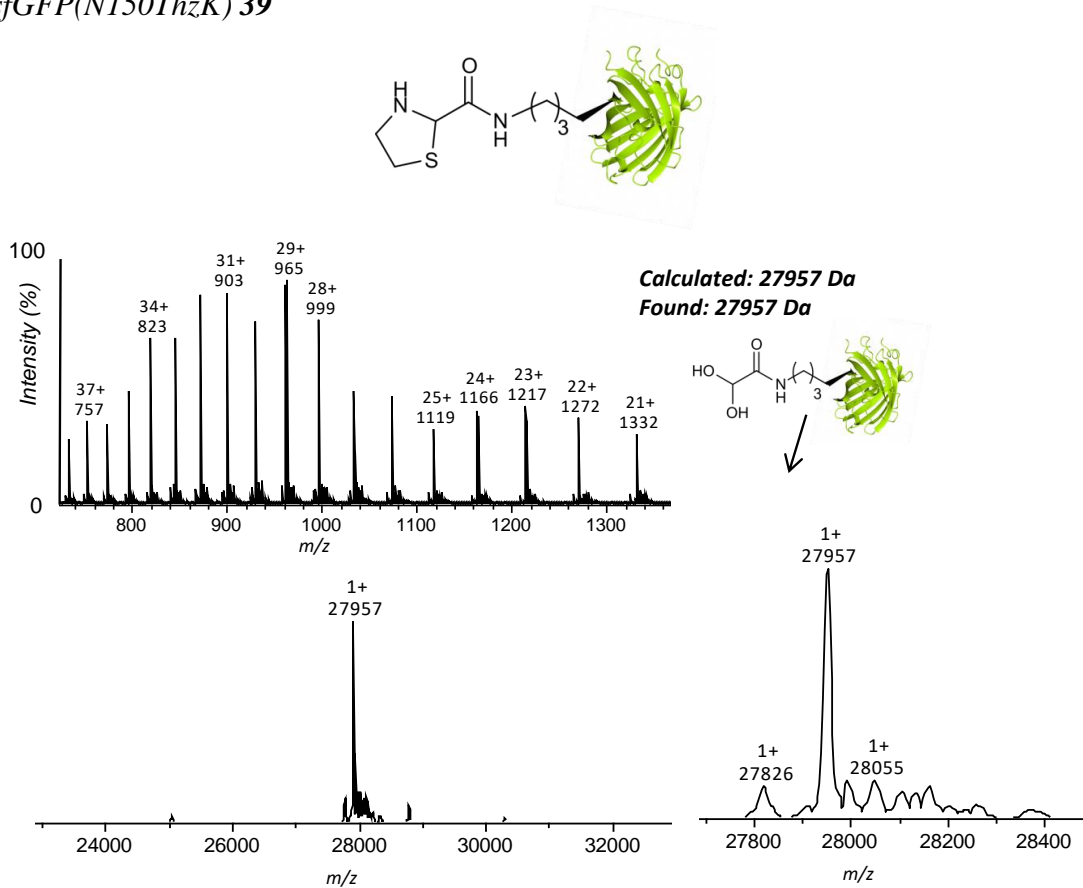
Calculated: 9593 Da (Hydrate)
Found: 9593 Da (Hydrate)

*sfGFP(N150ThzK) **39***

Calculated: 27957 Da
Found: 27957 Da

GFP(Y39ThzK) **40**

Calculated: 28631 Da
Found: 28633 Da

sfGFP(N150GlyoxylK) **41**

Calculated: 27916 Da (Hydrate)
Found: 27916 Da (Hydrate)

GFP(Y39GlyoxylK) **42**

Calculated: 28572 Da (Ald)
Found: 28574 Da (Ald)

Aldol-myolglobin **60**

Calculated: 17022 Da
Found: 17023 Da

Peak at 17170 Da is an
unknown, PLP related byproduct
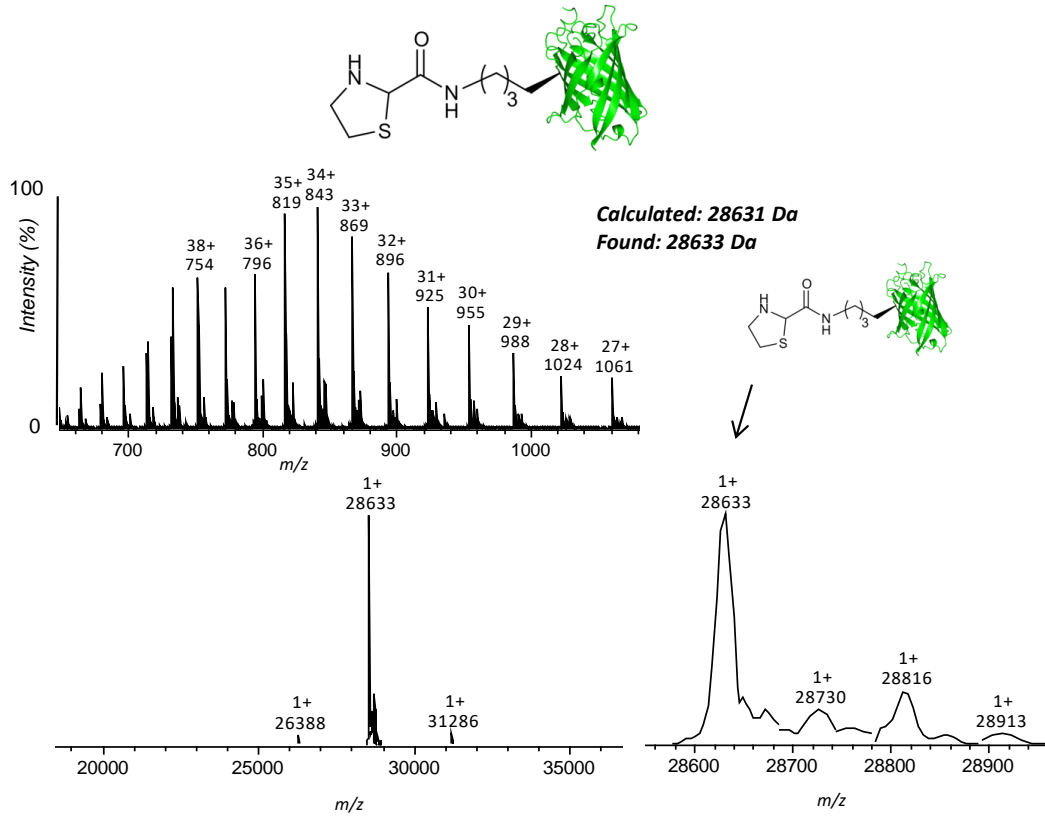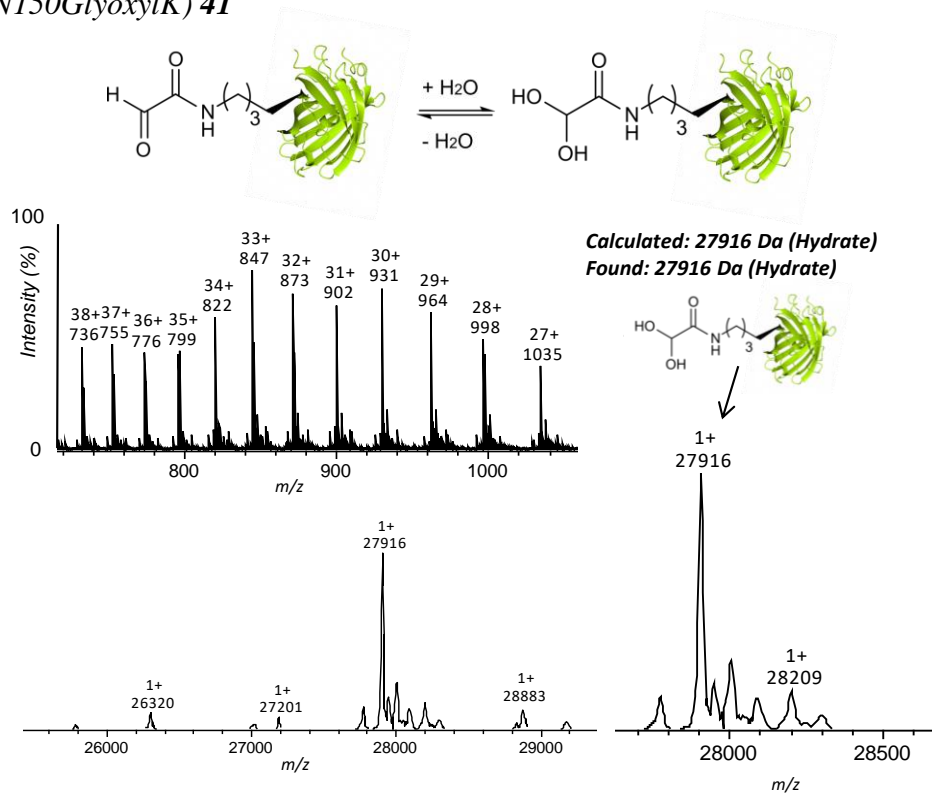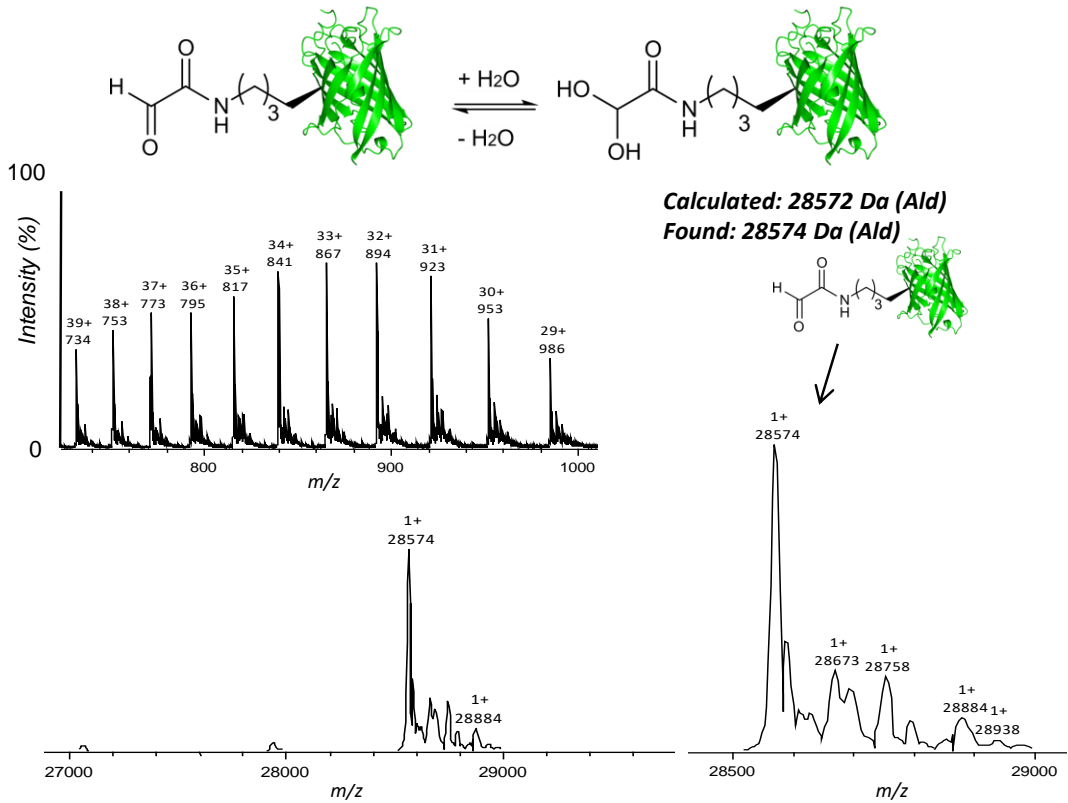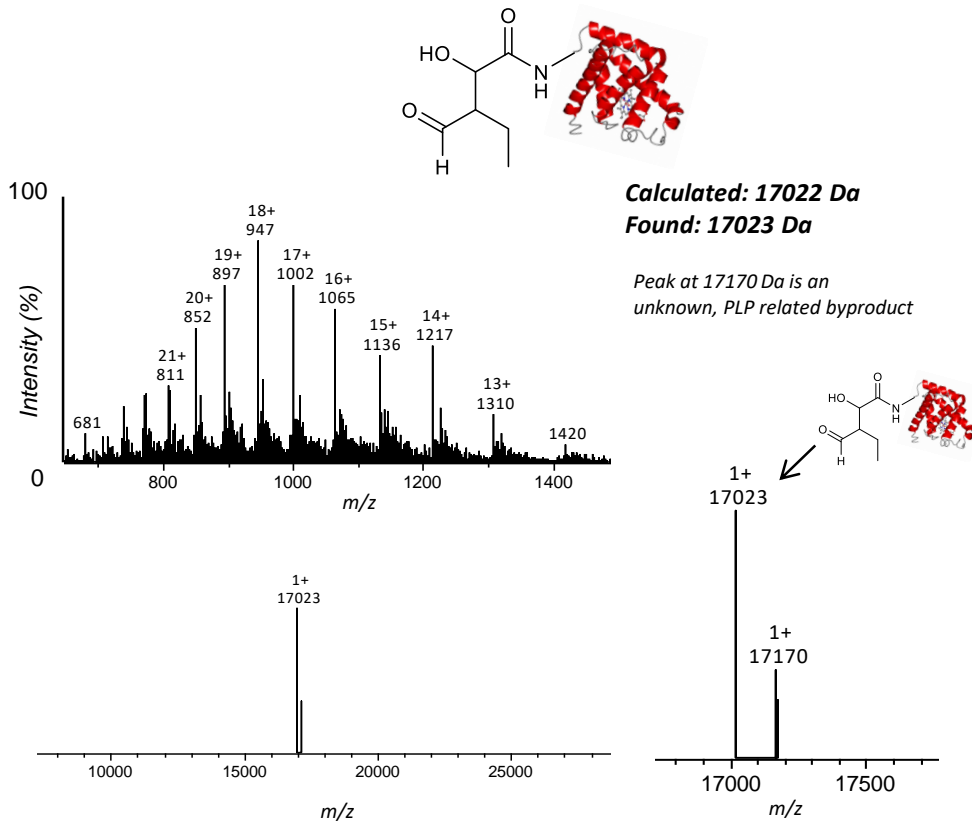
Fluroescently tagged thioredoxin **89**

Calculated: 12401 Da
Found: 12401 Da
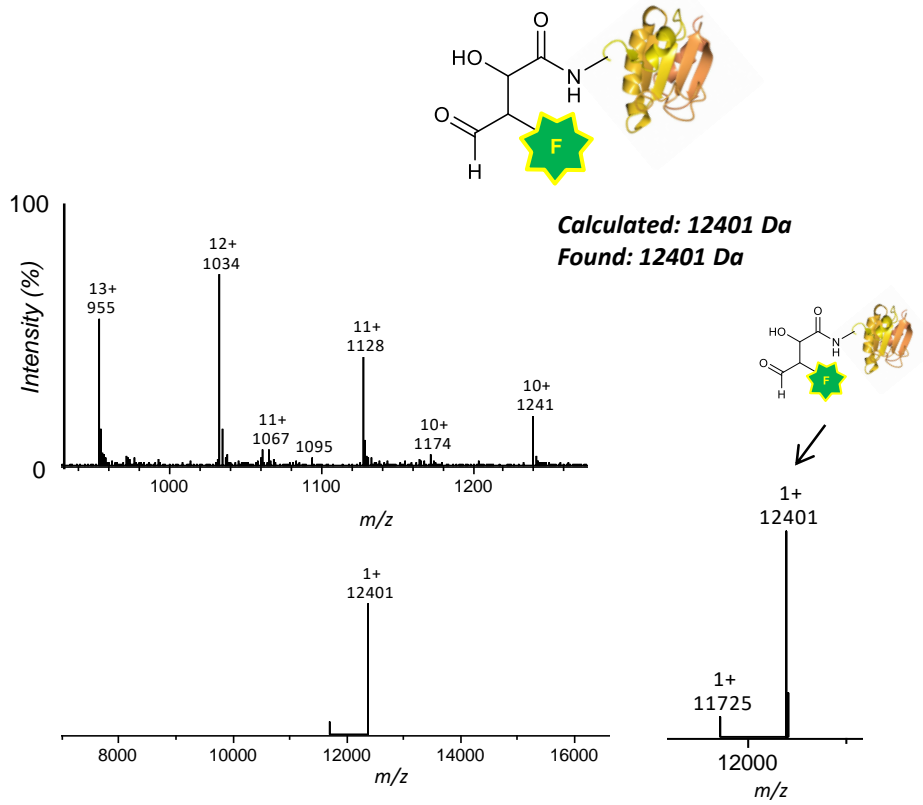
Biotin tagged thioredoxin **94**

Calculated: 12393Da
Found: 12395 Da

Azide tagged thioredoxin **90**

Calculated: 12311 Da
Found: 12314 Da
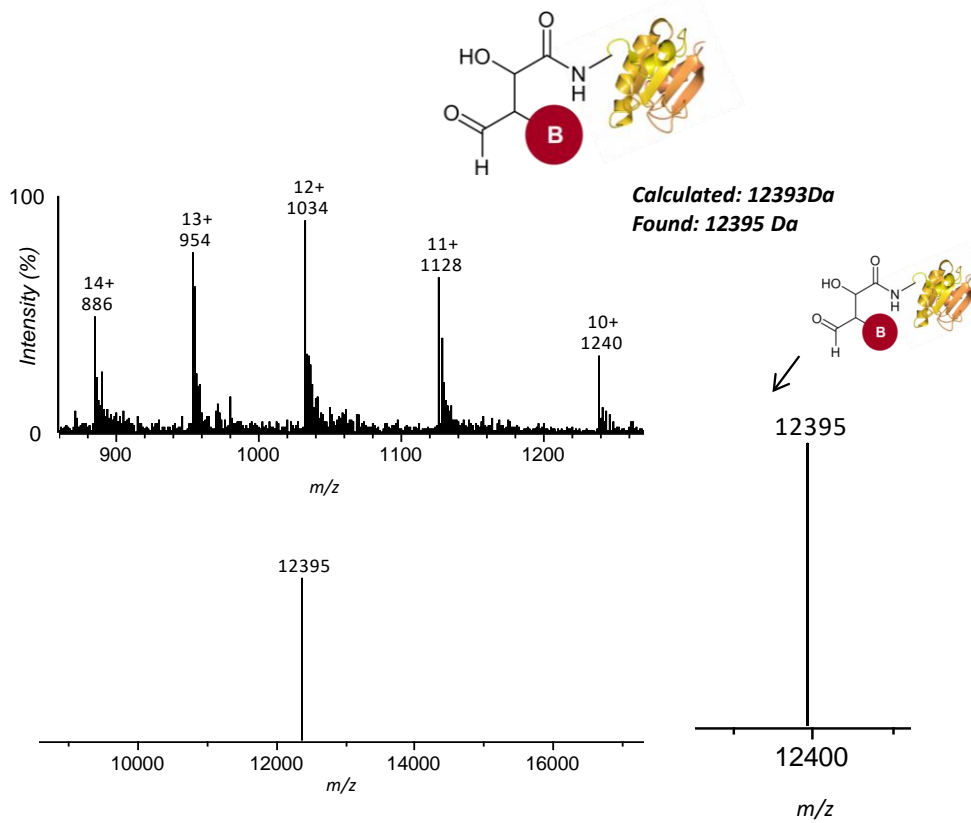
Fluorescently labelled myoglobin **93**
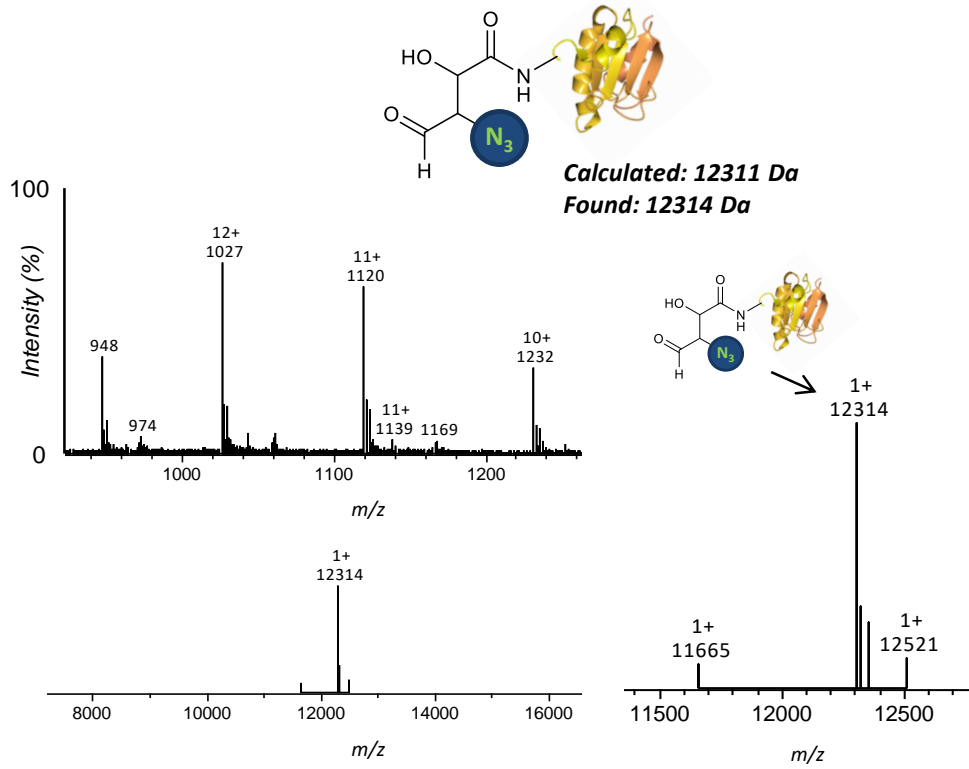
Calculated: 17708 Da
Found: 17711 Da
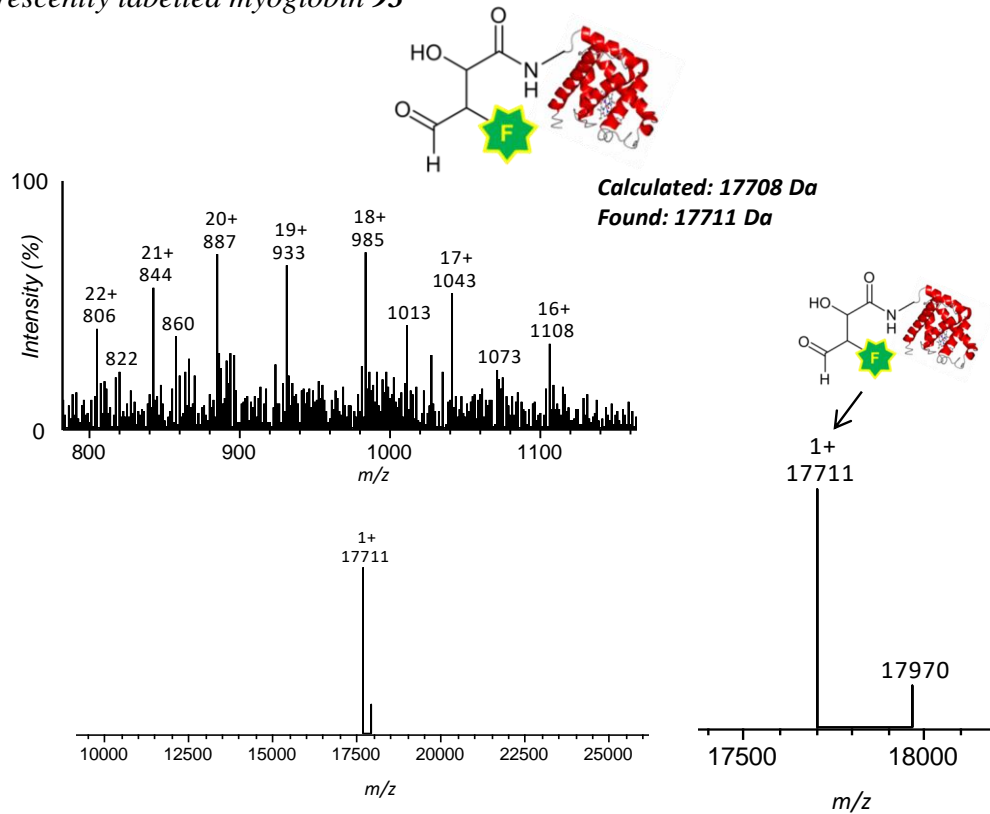
Azide tagged myoglobin **116**
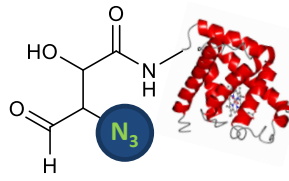
Calculated: 17617 Da
Found: 17617 Da

Peak at 17169 Da is an unknown, PLP related byproduct
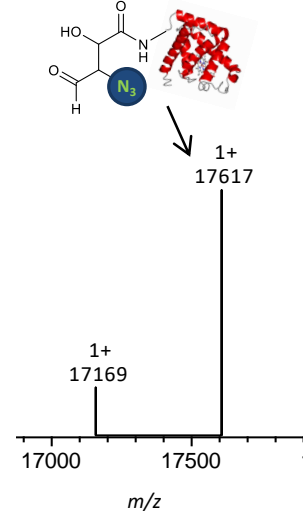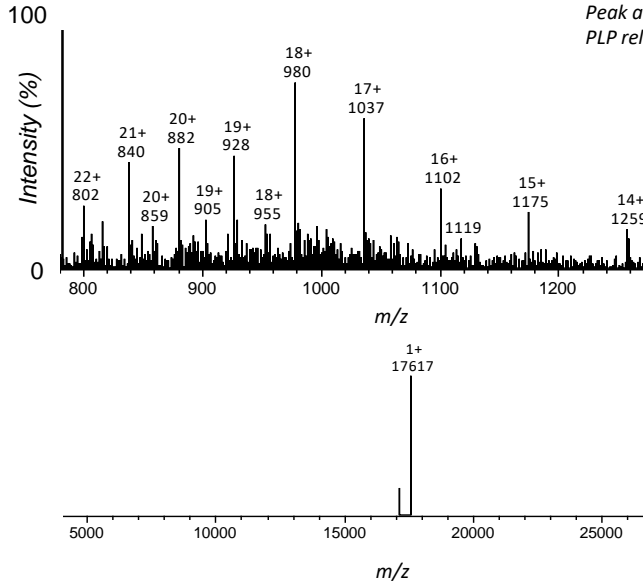
Fluorescently labelled CTB **98**
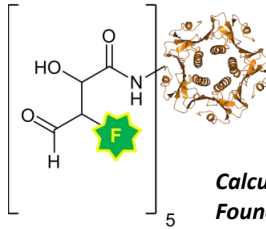
Calculated: 12357 Da
Found: 12358 Da
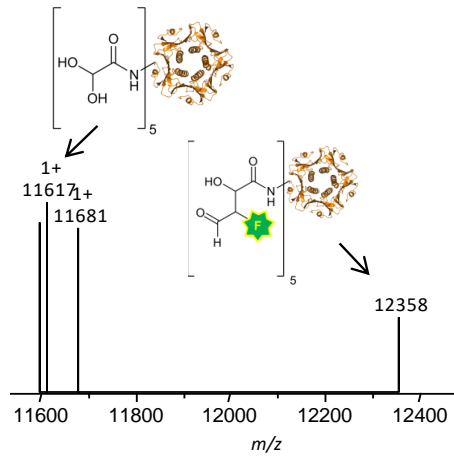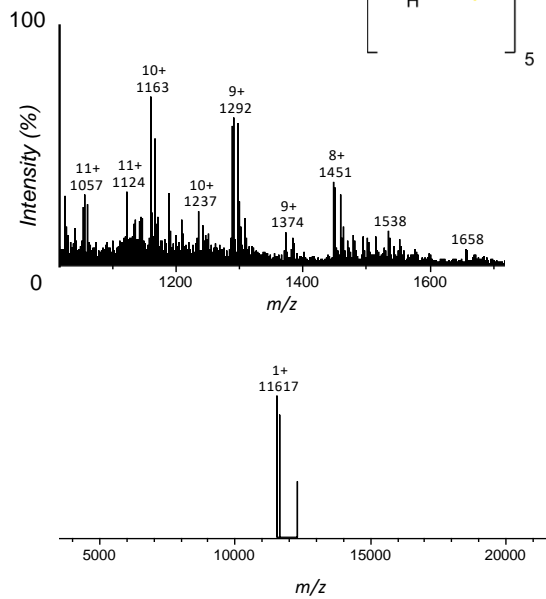
Biotin tagged GFP **92**
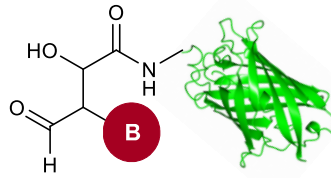
Calculated: 29388 Da
Found: 29388 Da
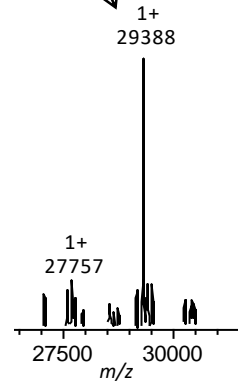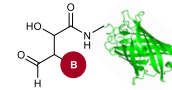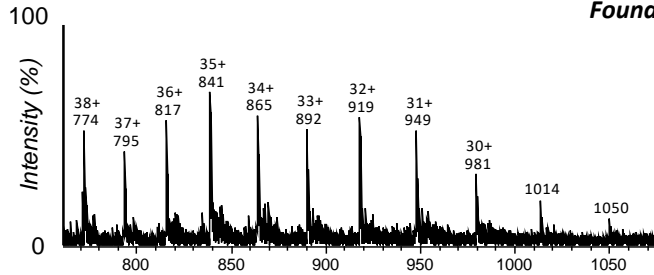
Folate tagged GFP (Y39CycloOctK) **96**
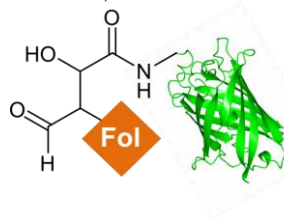
Calculated: 29730 Da
Found: 29731 Da

Peak at 1093 Da corresponds to unreacted probe that is present after SpinTrap purification. Consider further methods of purification if purer protein samples are required for downstream use.
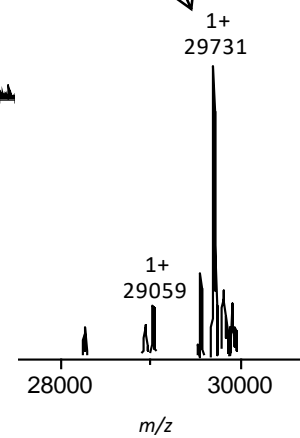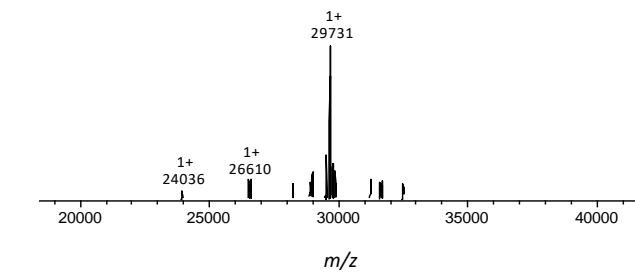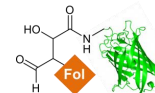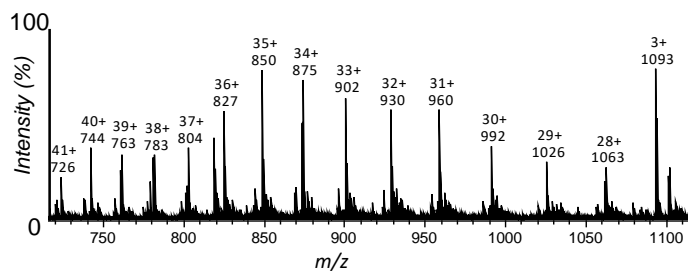
*Internal (position 150) azide labelled sfGFP **99***

Calculated: 28566 Da
Found: 28563 Da
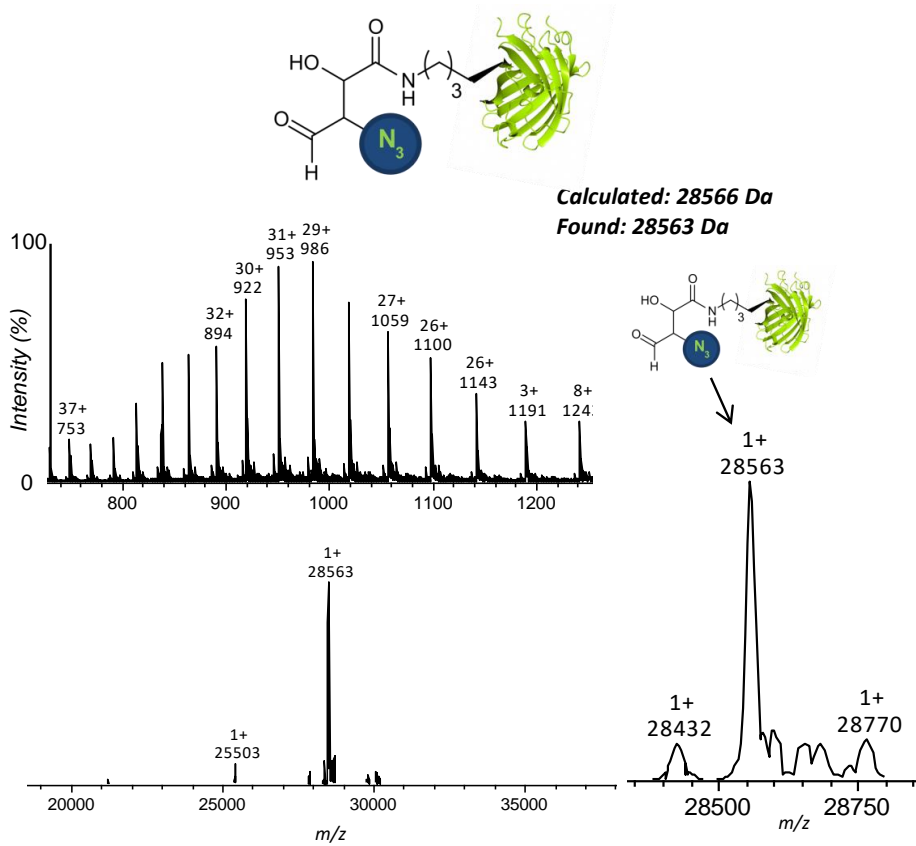
*Internal (position 39) azide labelled GFP **100***

Calculated: 29240 Da
Found: 29242 Da
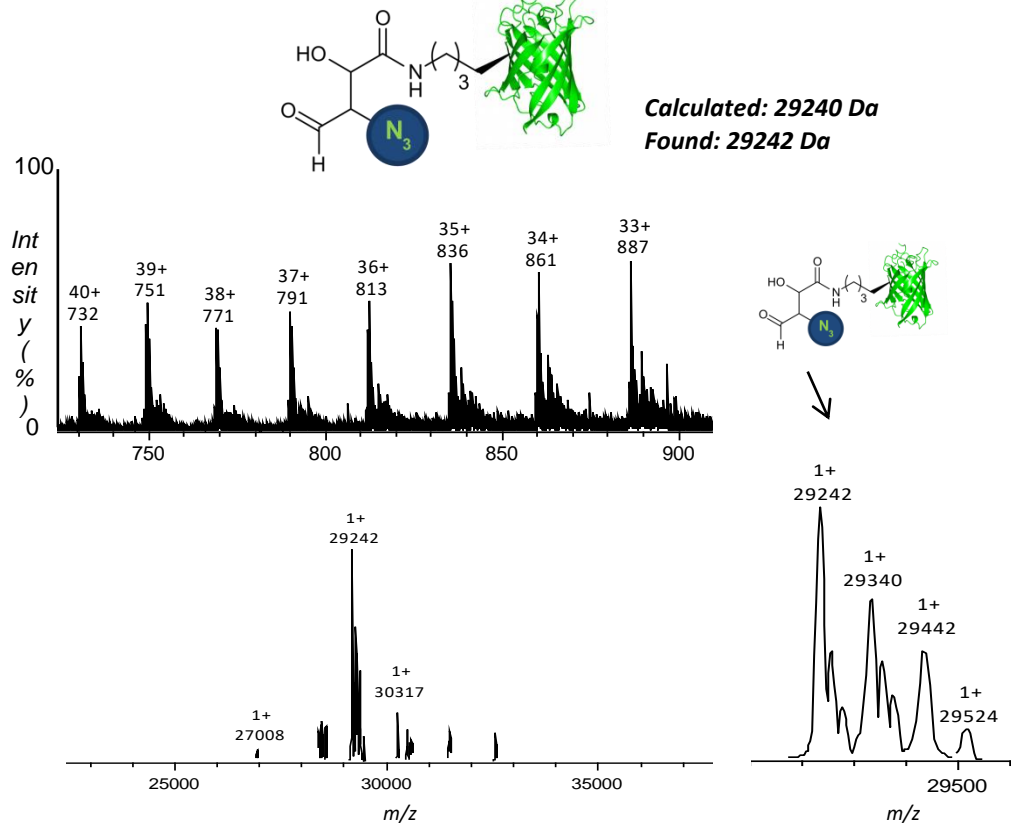
*Chemically myristoylated HASPA(G1S) **127***
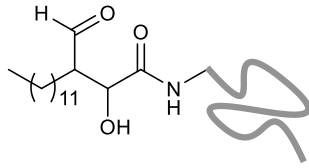
Calculated: 9656 Da
Found: 9657 Da
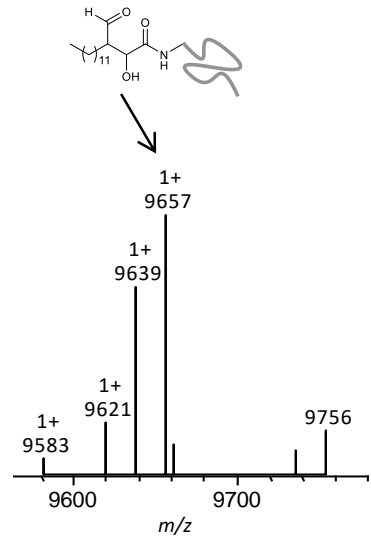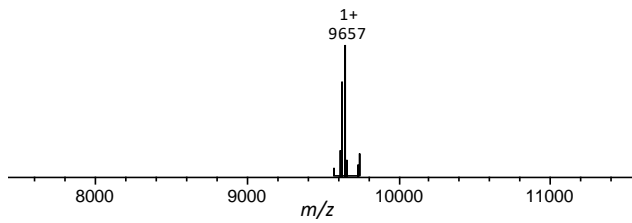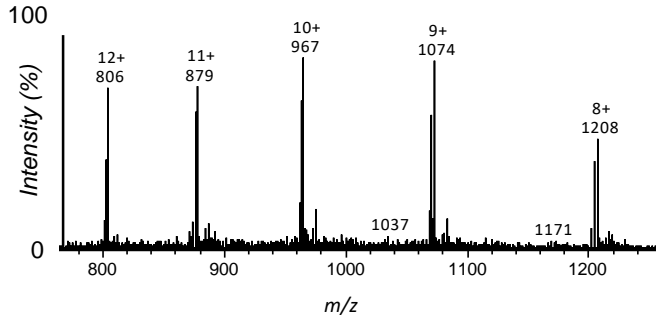
*Chemically myristoylated [15N]HASPA(G1S) **127-15N***
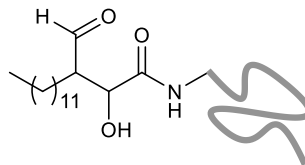
Calculated: 9787 Da
Found: 9786 Da
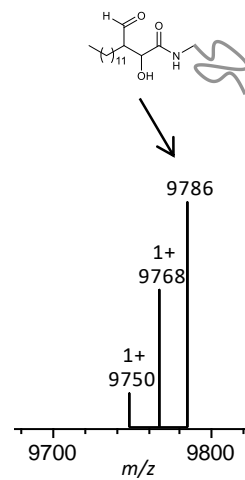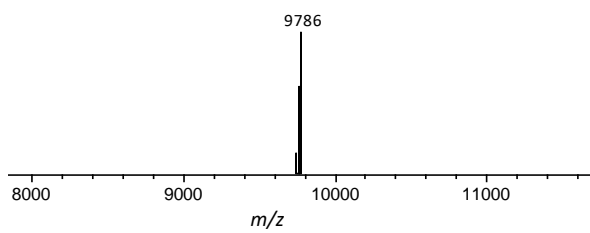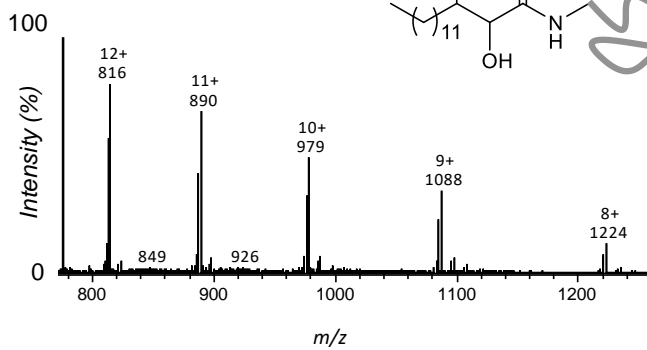
Azide tagged HASPA(G1S) **95**
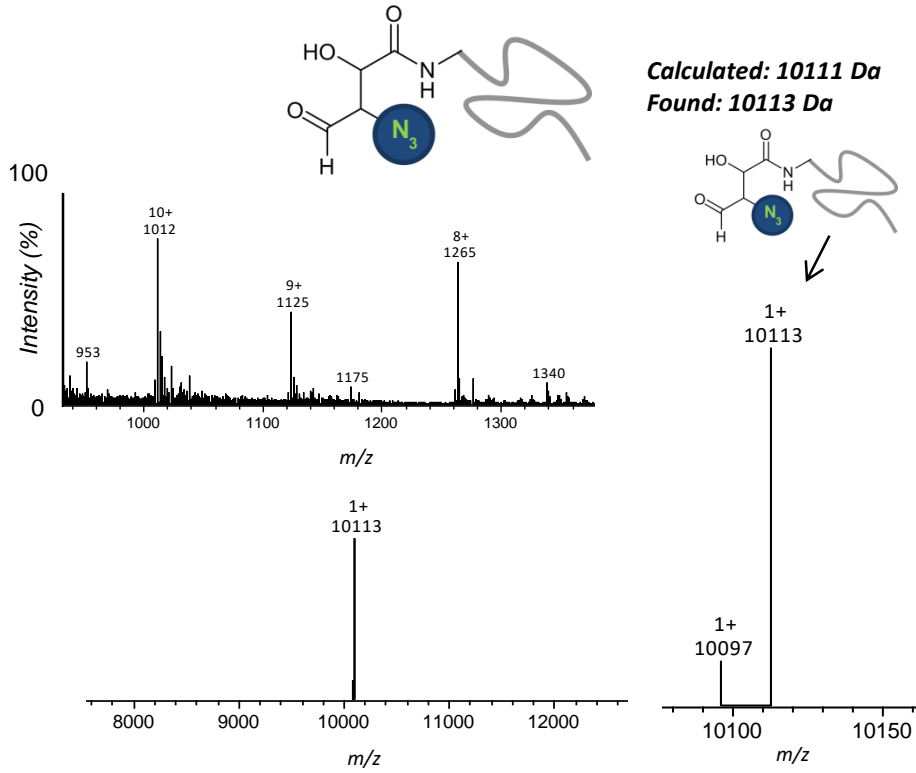
Fluorescently tagged [15N]HASPA(G1S) **91-15N**
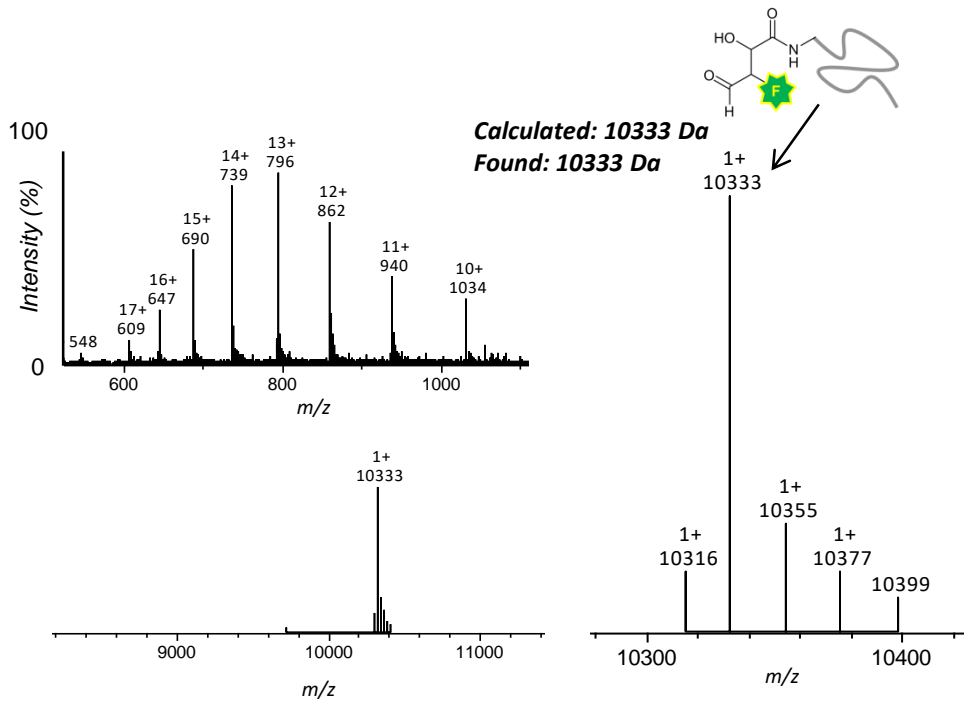
*Internal (position 150) biotinylated sfGFP **102***
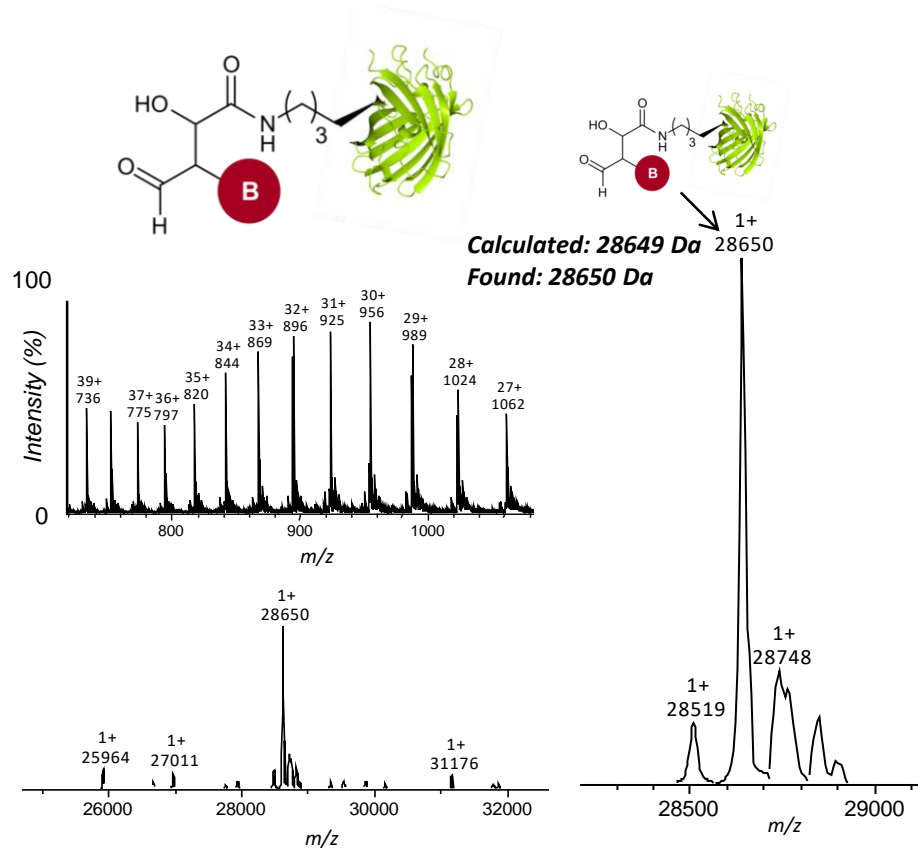
Calculated: 28649 Da
Found: 28650 Da

*Fluorescently labelled, biotinylated thioredoxin **119***

Calculated: 12817 Da
Found: 12818 Da

Azide labelled, biotinylated myoglobin **118**

Calculated: 18035 Da
Found: 18038 Da

Peak at 17174 Da is an unknown, PLP related byproduct

Dually acylated HASPA **129**

Calculated: 9895 Da
Found: 9898 Da
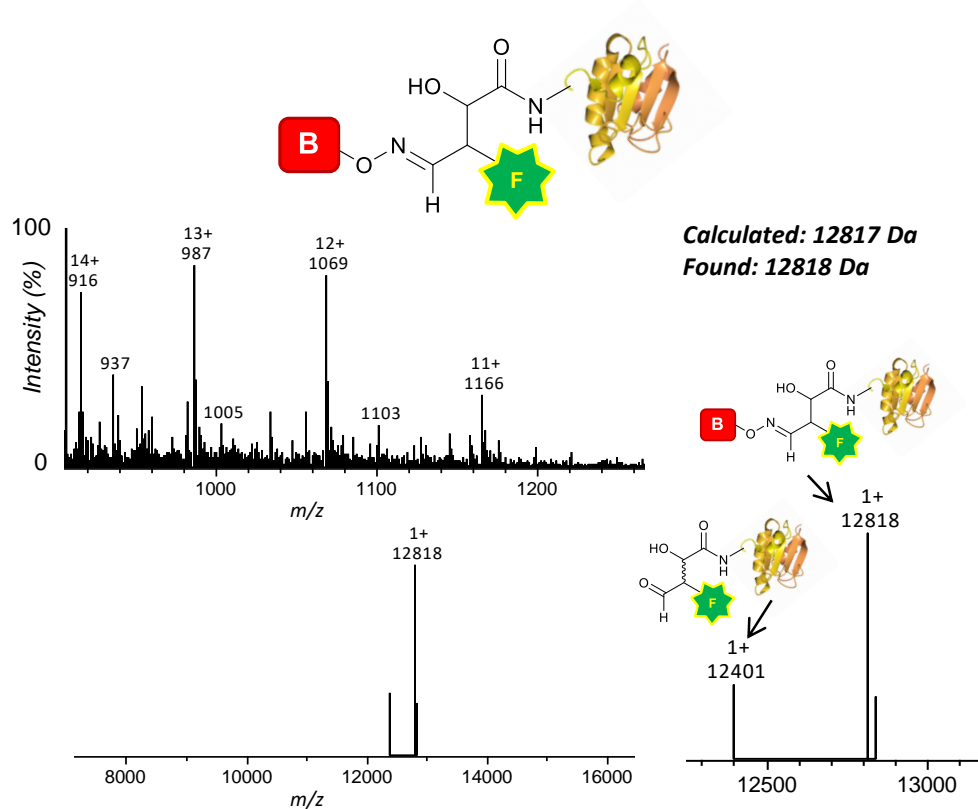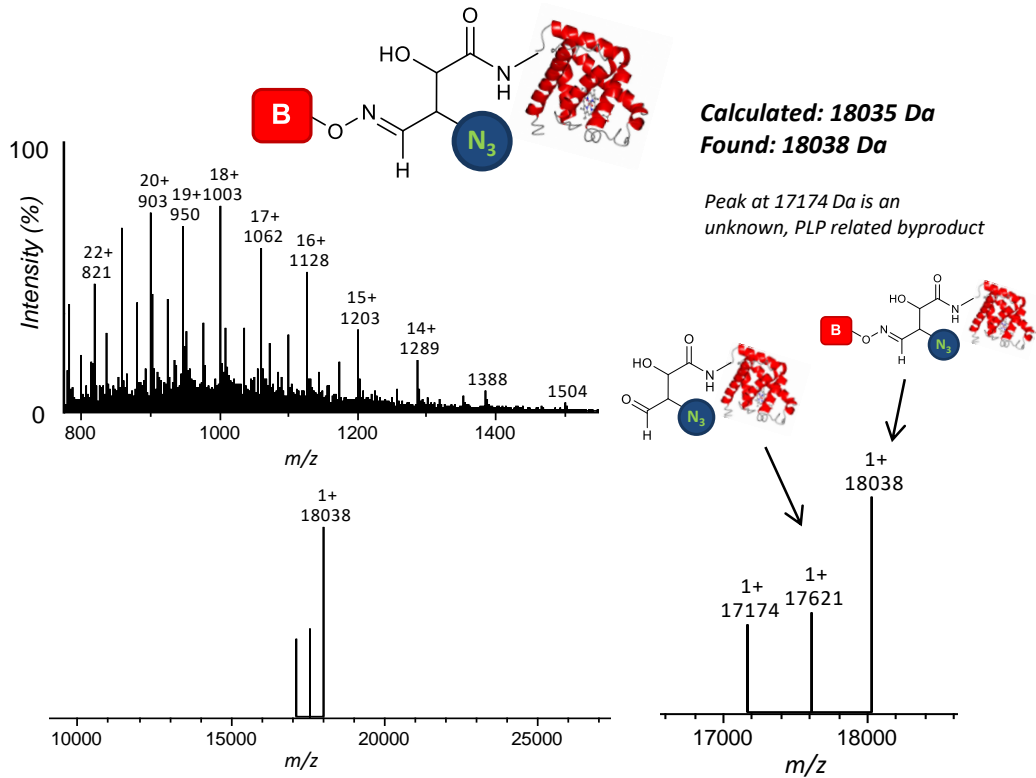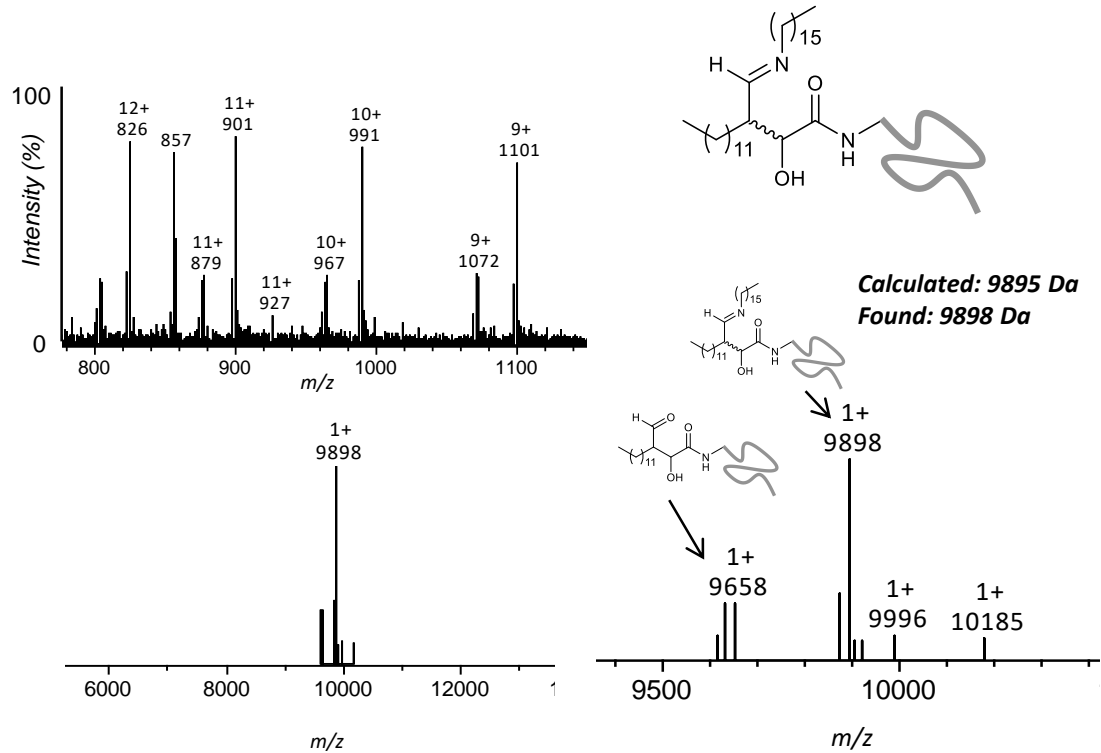
*PHILIPA myoglobin product **148***

Calculated: 17014 Da
Found: 17018 Da
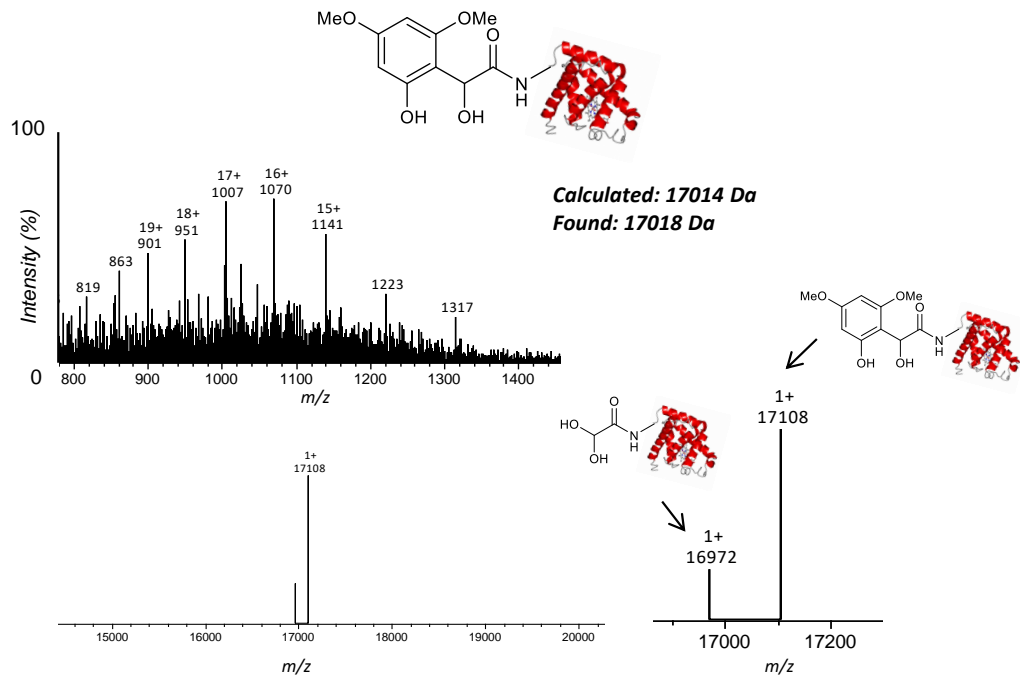
Kinetic data for OPAL

Donor	Organo catalyst	Organocatalyst loading (mM)	Rate Constant ($M^{-1} s^{-1}$)	Error of Rate Constant ($M^{-1}s^{-1}$)
51	61	1	0.0009	0.00006
		10	0.0033	0.0001
		25	0.0100	0.0006
	63	1	0.0005	0.00003
		10	0.0037	0.0002
		25	0.0058	0.0005
	64	1	>0.00001	-
		10	0.0009	0.00006
		25	0.0016	0.00016
	65	1	0.0010	0.00008
		10	0.0062	0.0006
		25	0.0108	0.00115
	66	1	0.0097	0.000788
		10	0.0581	0.00512
		25	0.0942	0.0108
	67	1	0.0004	0.00003
		10	0.0024	0.00008
		25	0.0052	0.00008
	68	1	0.0006	0.000063
		10	0.0030	0.0003
		25	0.0057	0.0006
	69	1	0.0022	0.00015
		10	0.0166	0.0012
		25	0.0252	0.0035
	70	1	0.0092	0.001
		10	0.0551	0.004
		25	0.0977	0.009
	71	1	0.0009	0.0002
		10	0.0049	0.00066
		25	0.0103	0.00121
54	61	1	1.6840	0.106
		10	4.6620	0.219
		25	7.8990	0.579
	70	1	3.7920	0.294
		10	11.8200	1.18
		25	23.9470	1.98

Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (100 mM) using 1 mM of catalyst **61**.
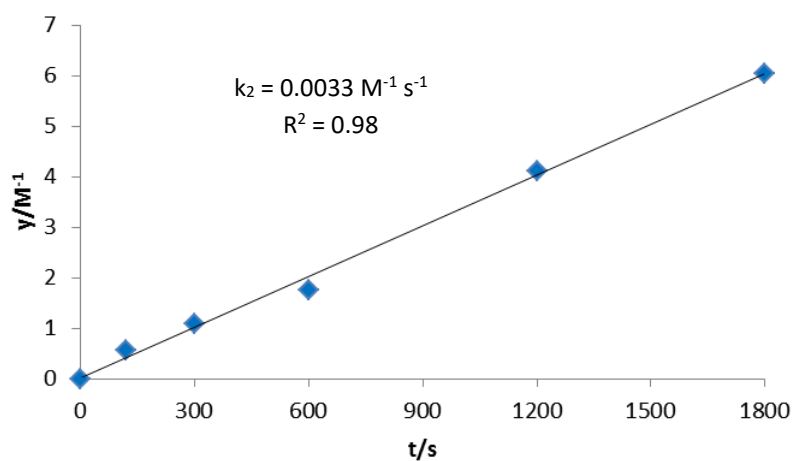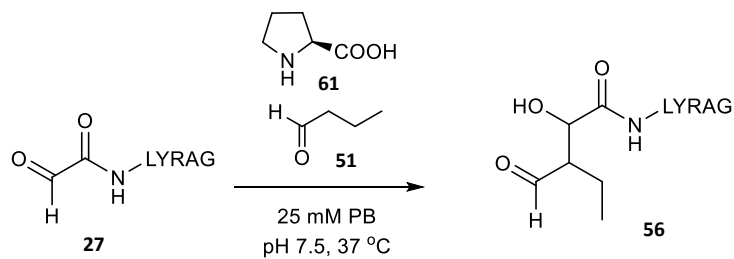
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (100 mM) using 10 mM of catalyst **61**.
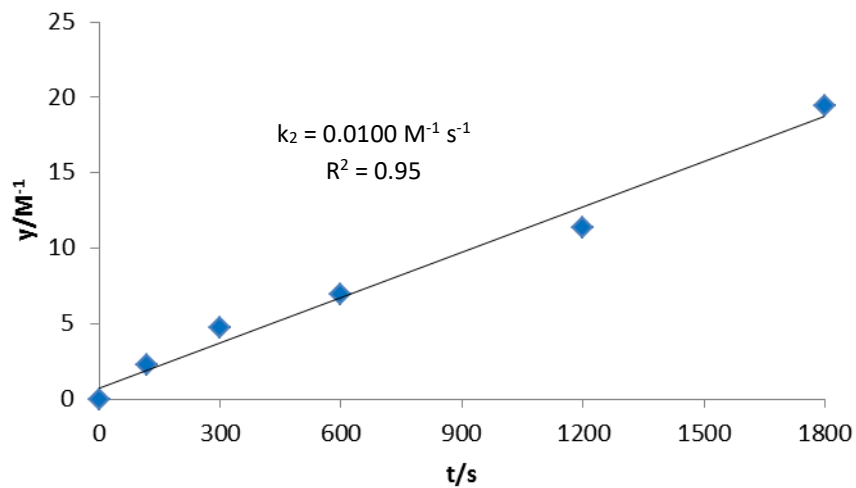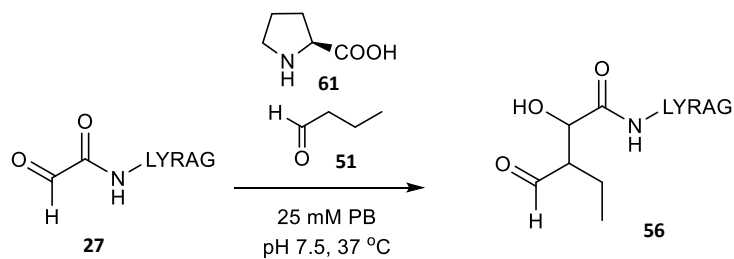
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (100 mM) using 25 mM of catalyst **61**.
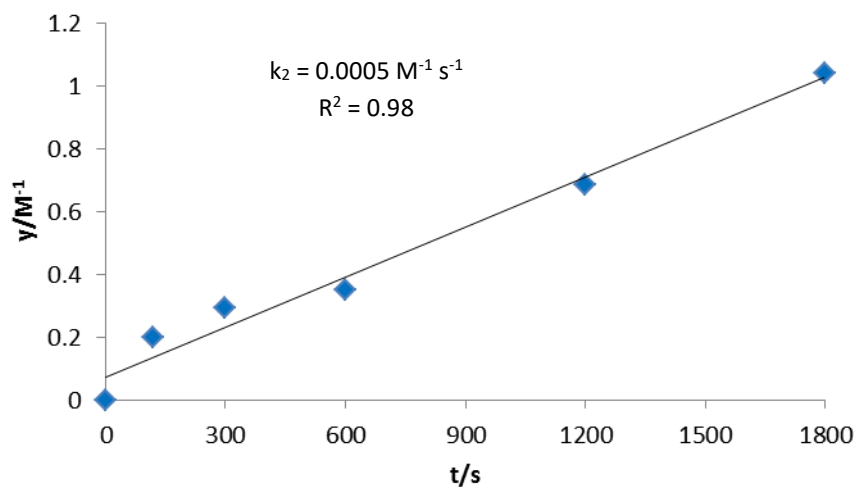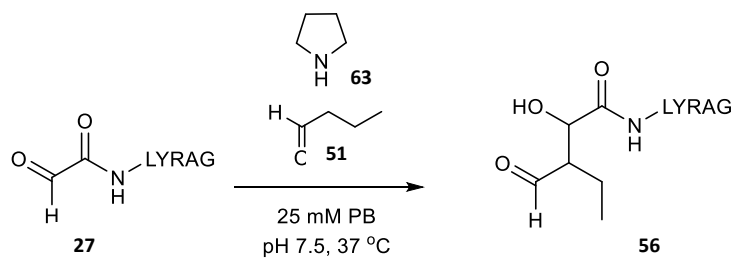
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (100 mM) using 1 mM of catalyst **63**.
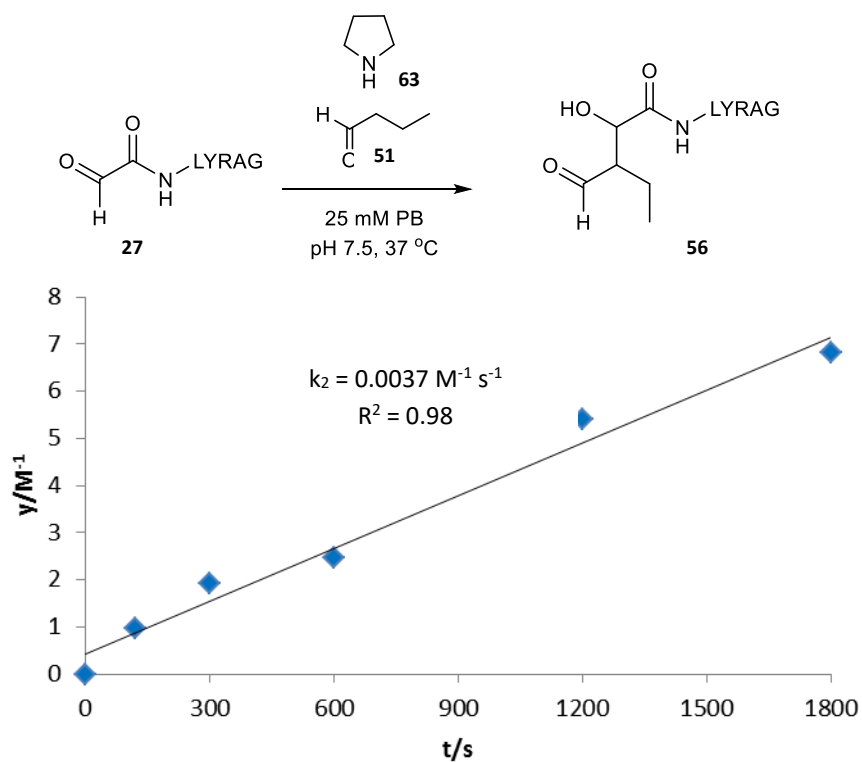
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (100 mM) using 10 mM of catalyst **63**.
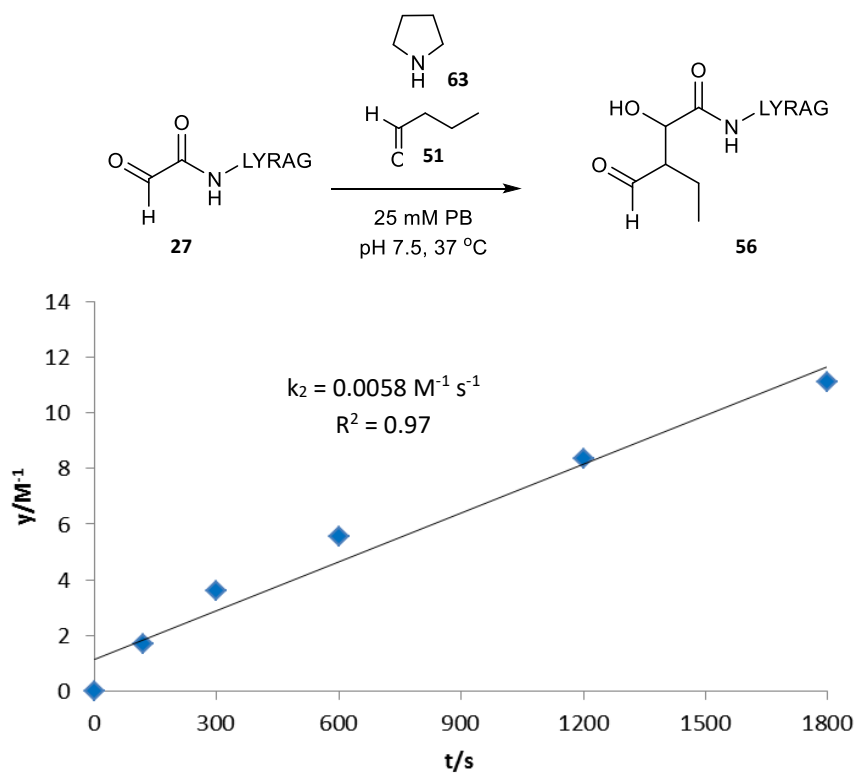
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (100 mM) using 25 mM of catalyst **63**.

Experimentally determined k_2 is extremely low (>0.0001 M^{-1} s^{-1})

Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (100 mM) using 1 mM of catalyst **64**

$k_2 = 0.0009$ M^{-1} s^{-1}

$R^2 = 0.98$
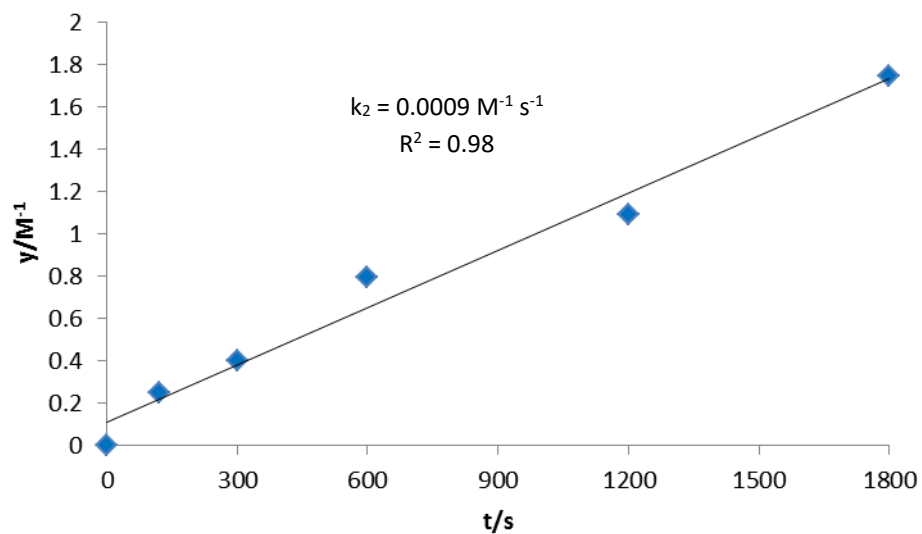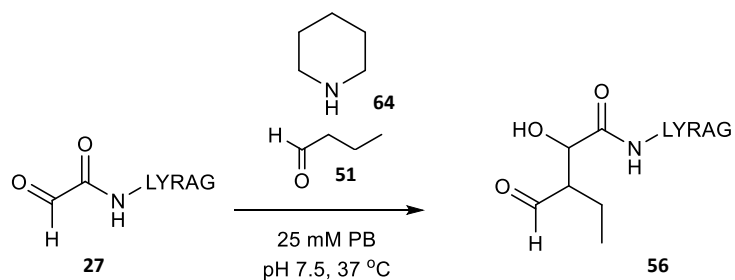
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (100 mM) using 10 mM of catalyst **64**.
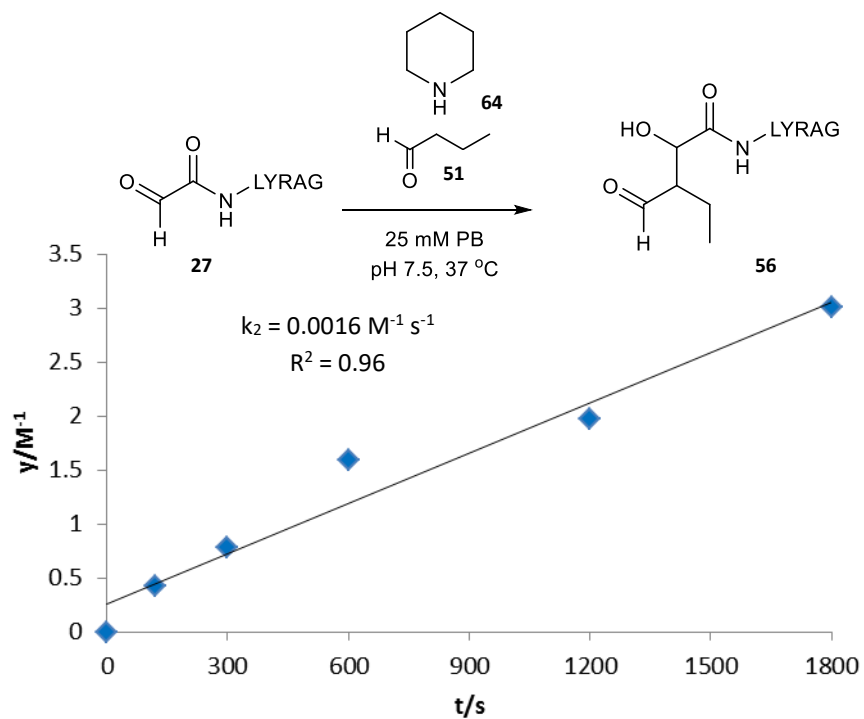
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (100 mM) using 25 mM of catalyst **64**.
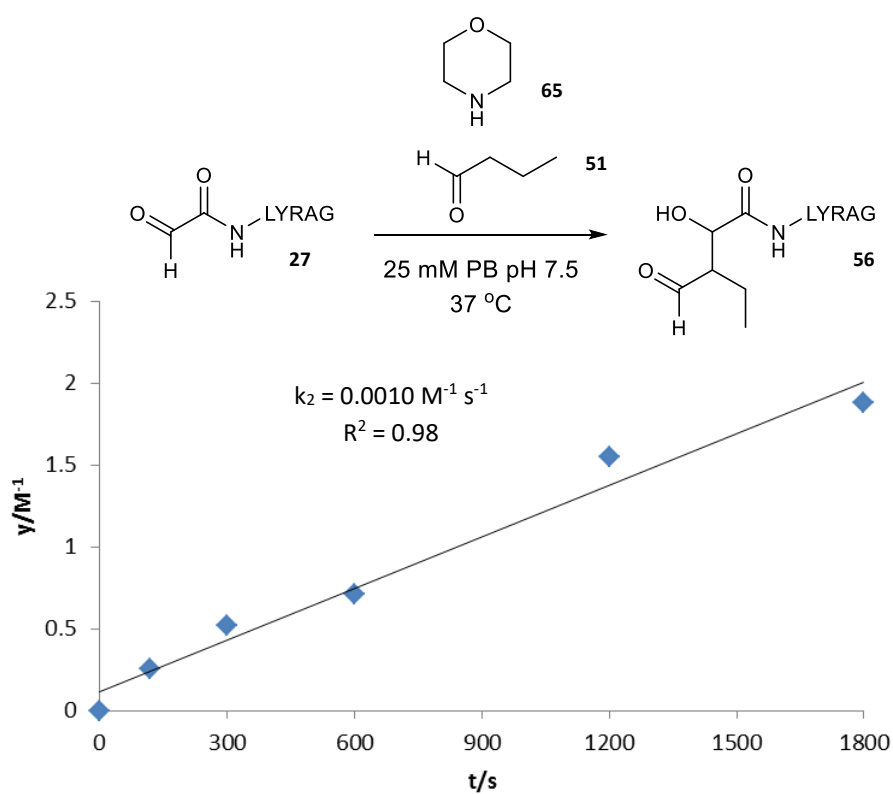
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (50 mM) using 1 mM of catalyst **65**.
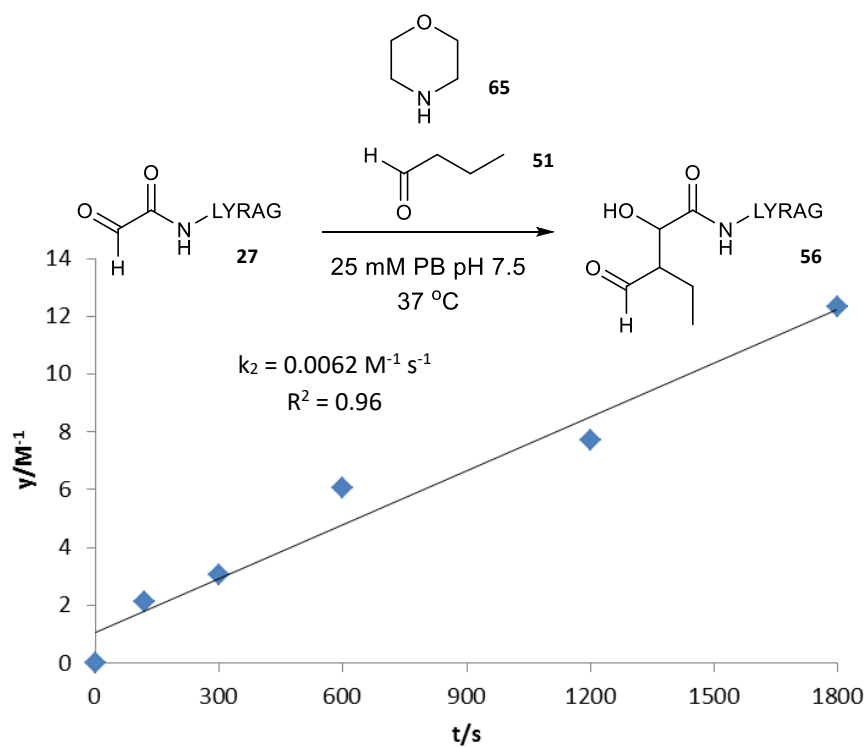
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (50 mM) using 10 mM of catalyst **65**.
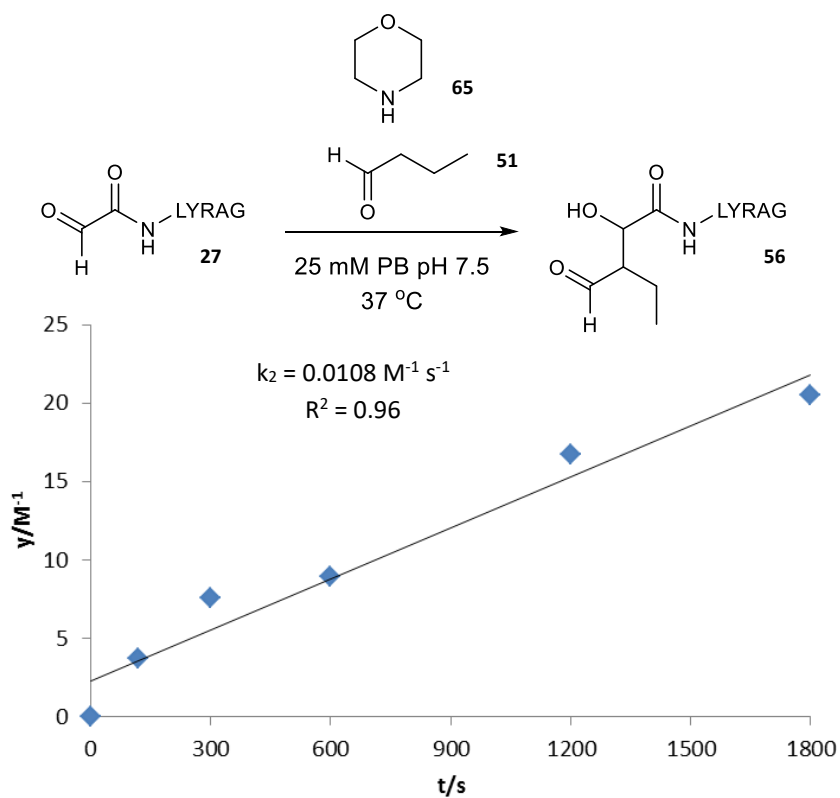
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (50 mM) using 25 mM of catalyst **65**.
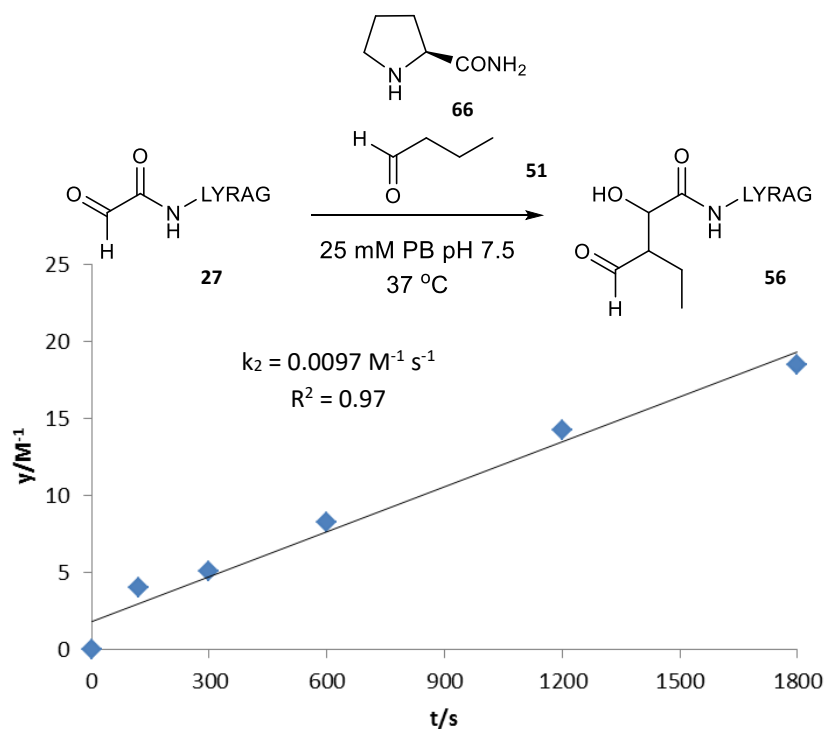
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (10 mM) using 1 mM of catalyst **66**.
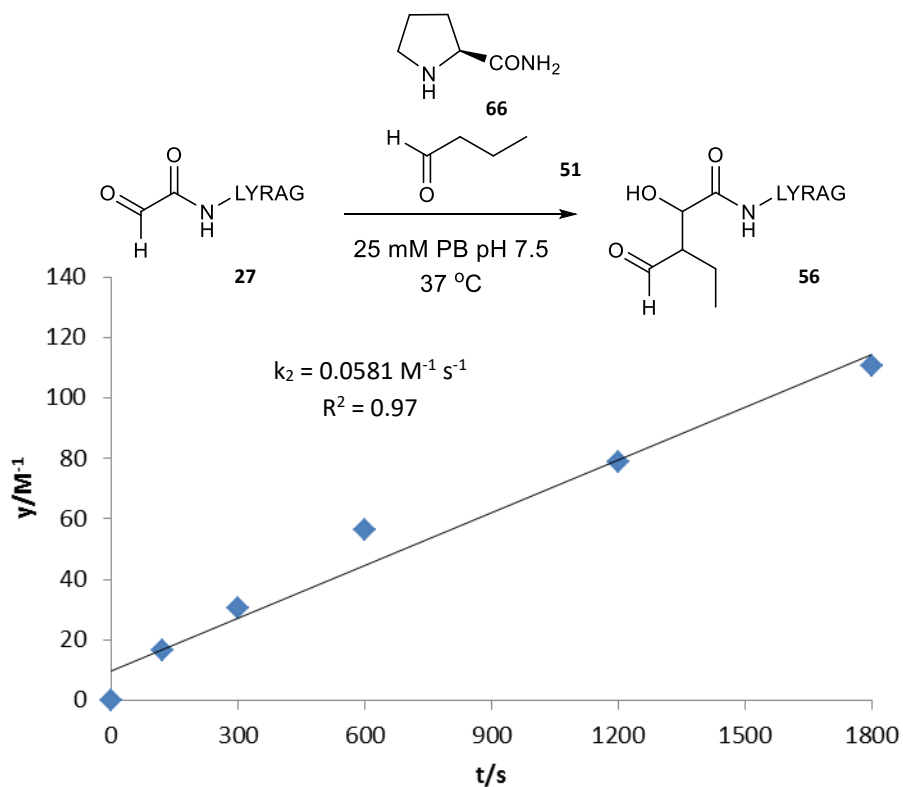
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (10 mM) using 10 mM of catalyst **66**.
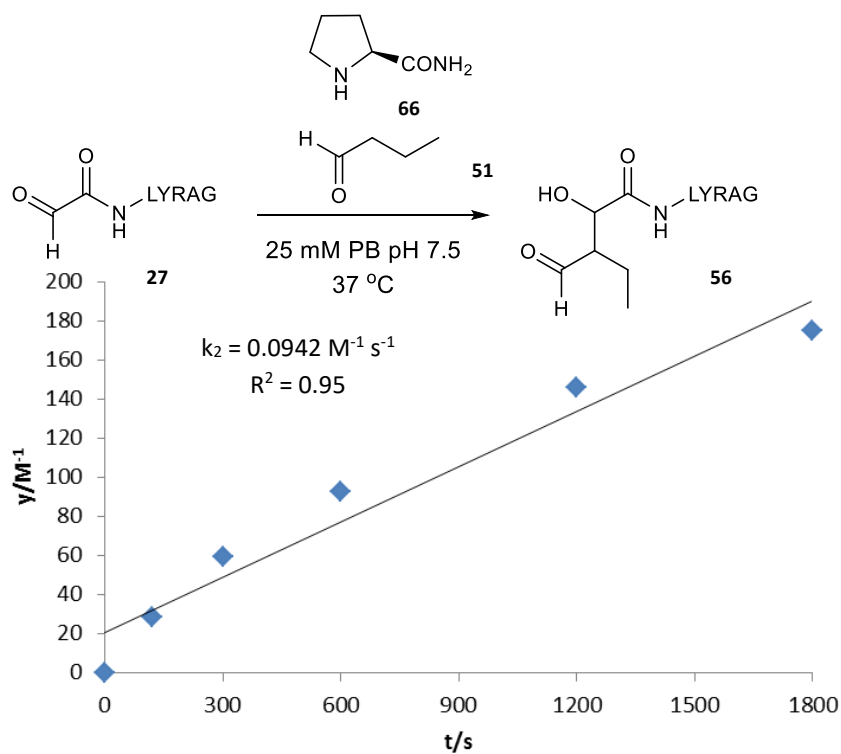
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (10 mM) using 25 mM of catalyst **66**.
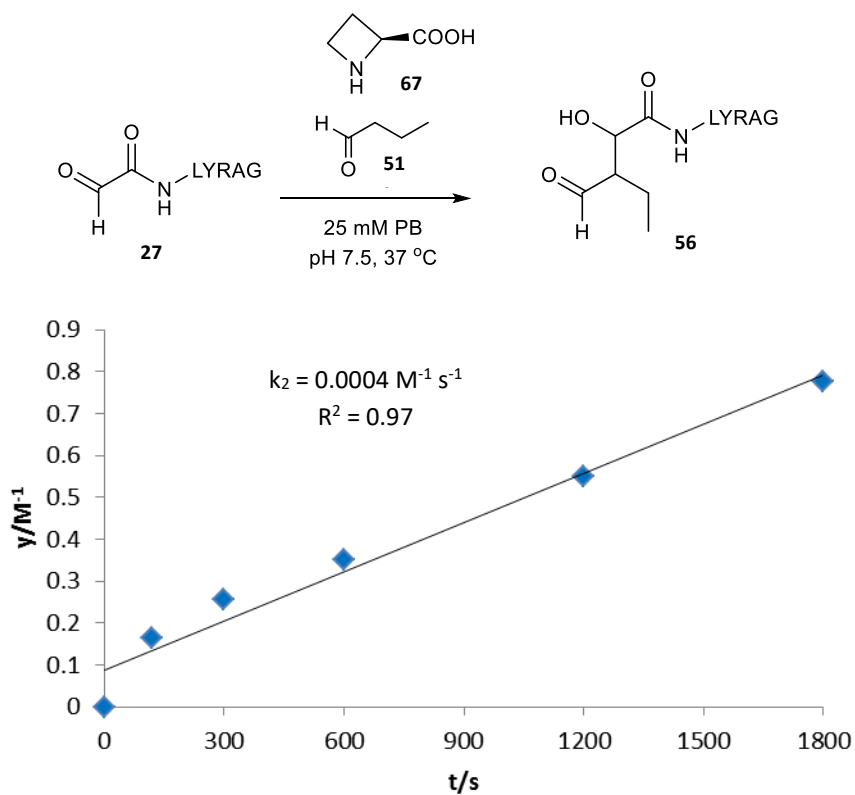
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (100 mM) using 1 mM of catalyst **67**.
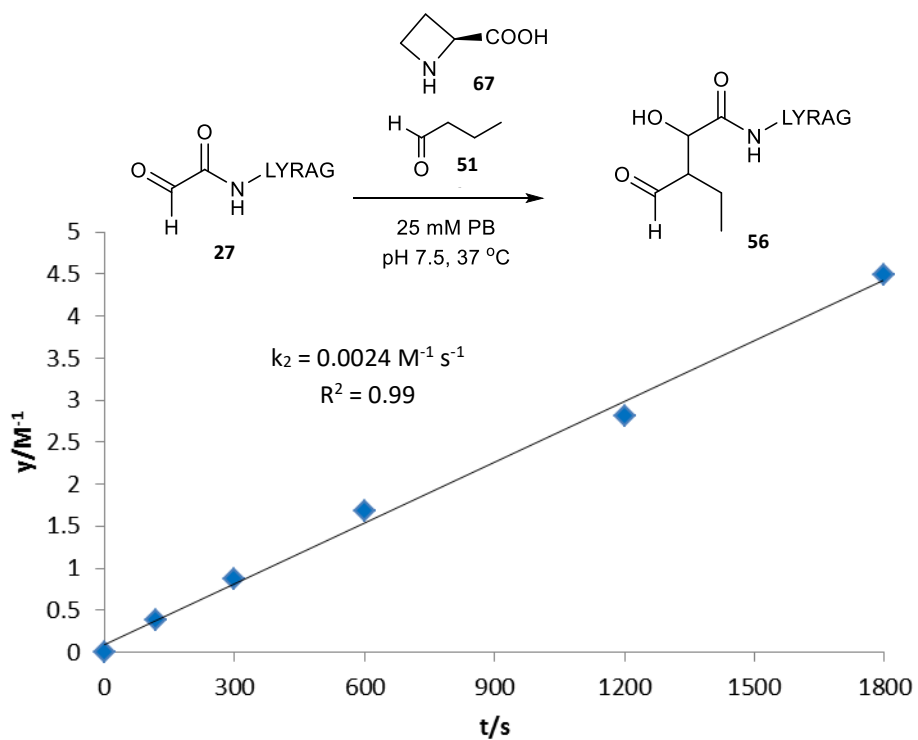
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (100 mM) using 10 mM of catalyst **67**.
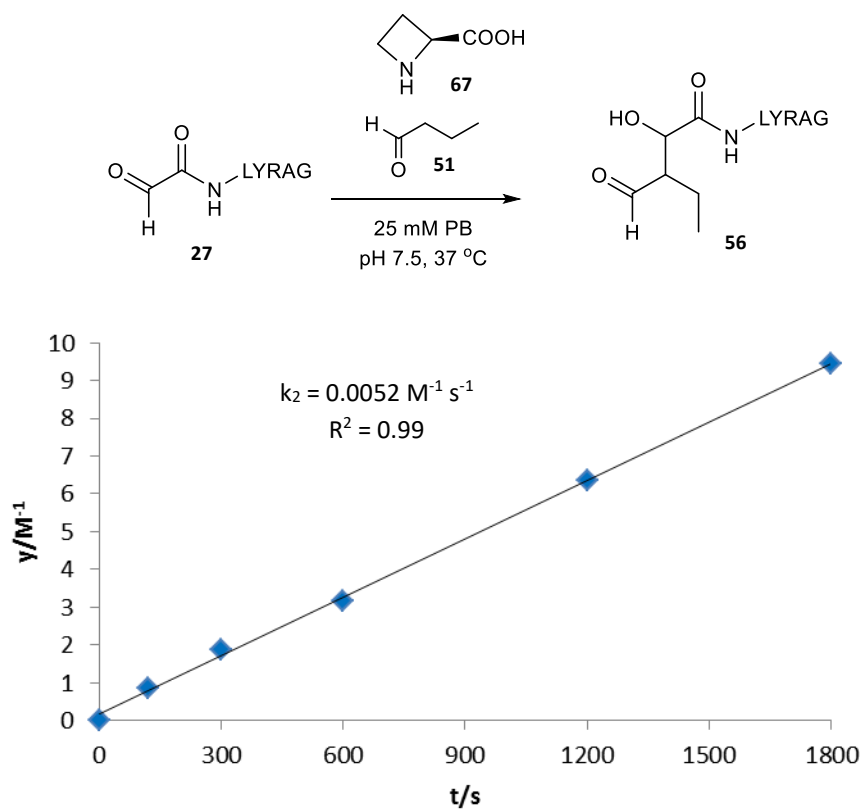
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (100 mM) using 25 mM of catalyst **67**.
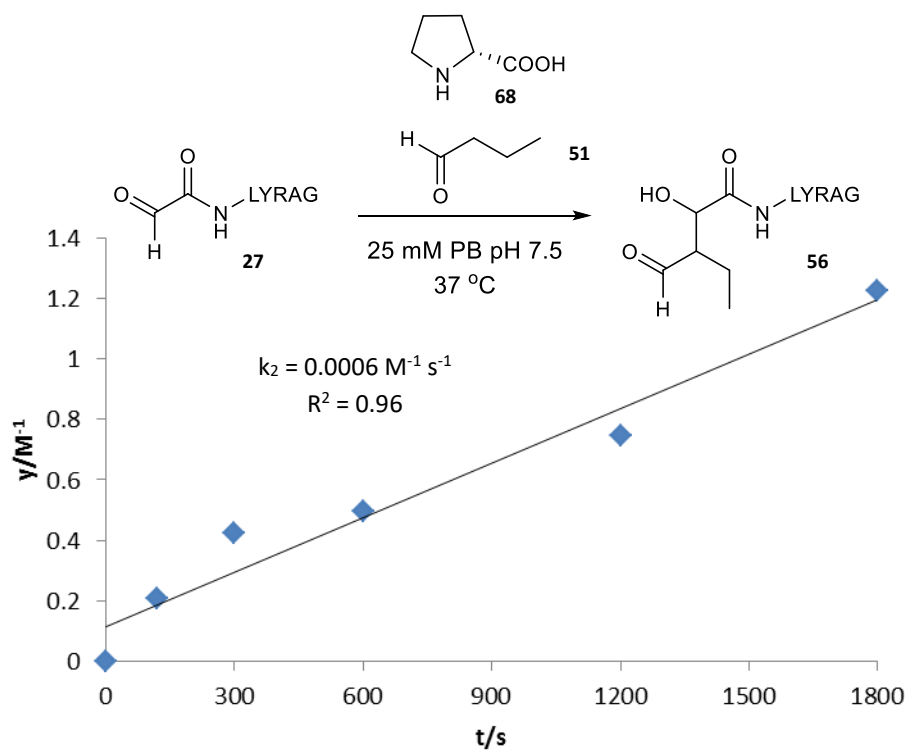
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (100 mM) using 1 mM of catalyst **68**.
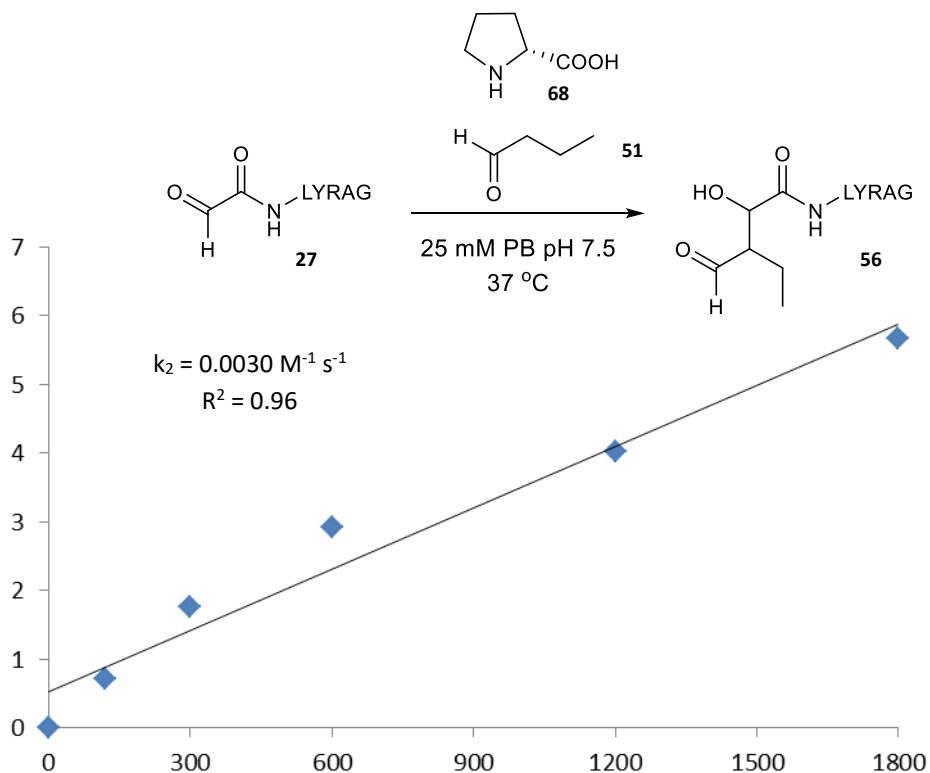
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (100 mM) using 10 mM of catalyst **68**.
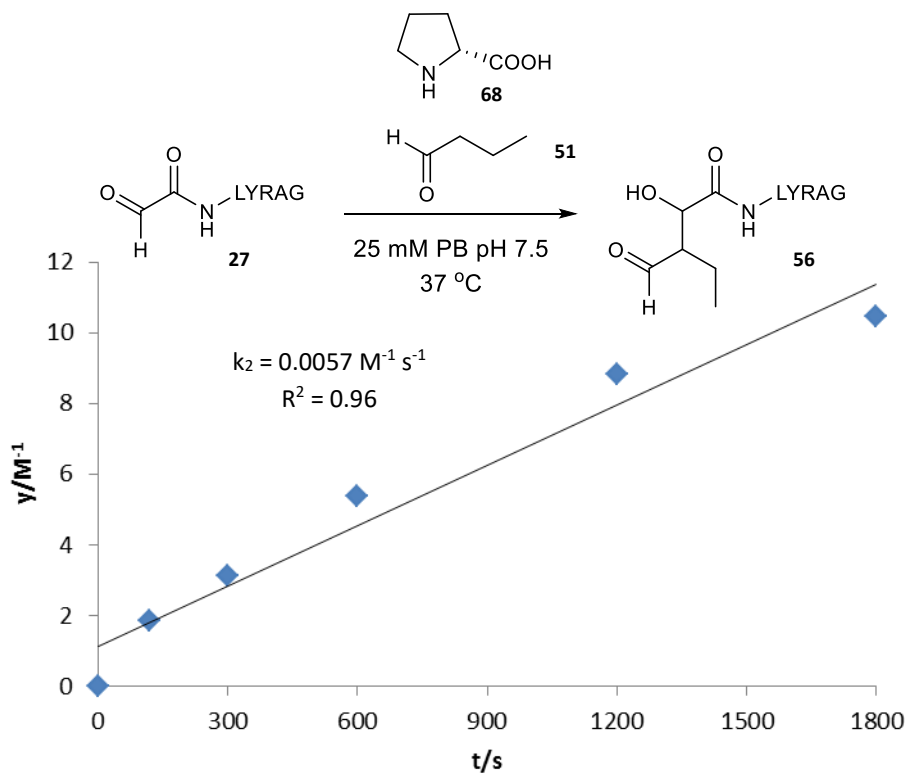
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (100 mM) using 25 mM of catalyst **68**.
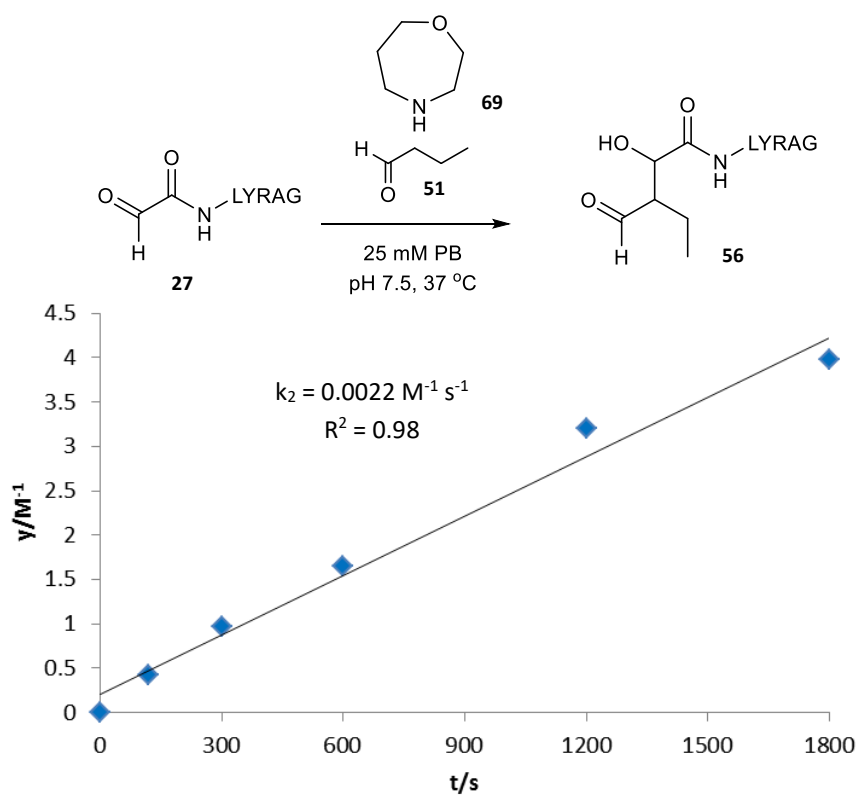
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (100 mM) using 1 mM of catalyst **69**.
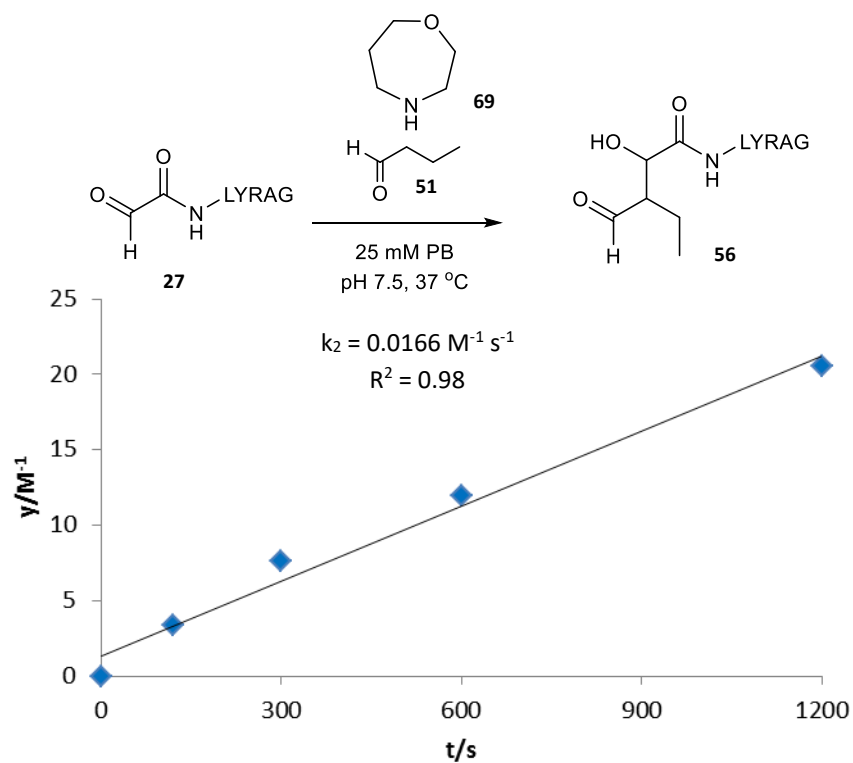
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (100 mM) using 10 mM of catalyst **69**.
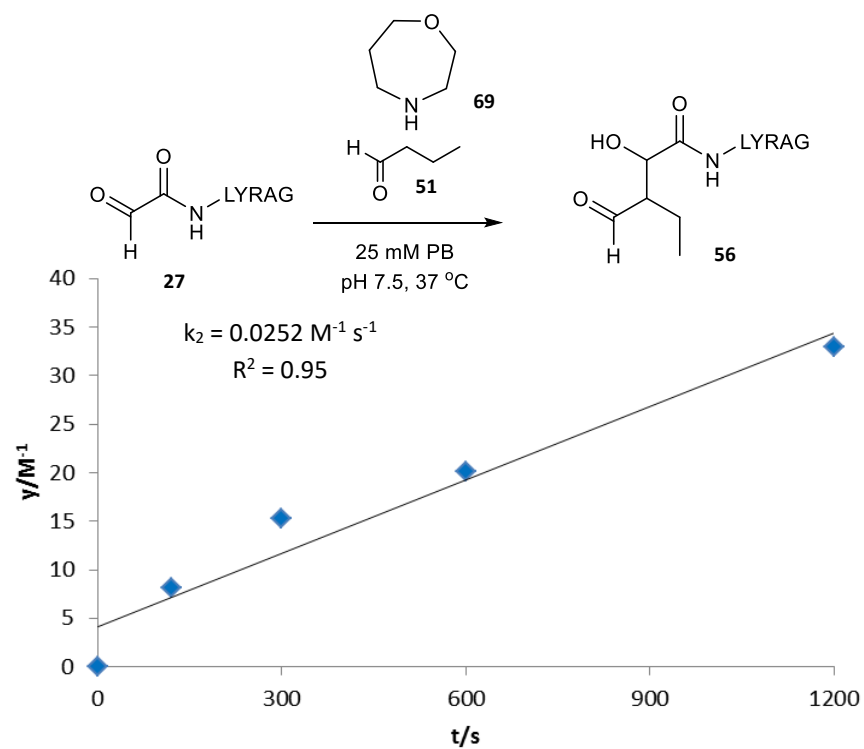
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (100 mM) using 25 mM of catalyst **69**.
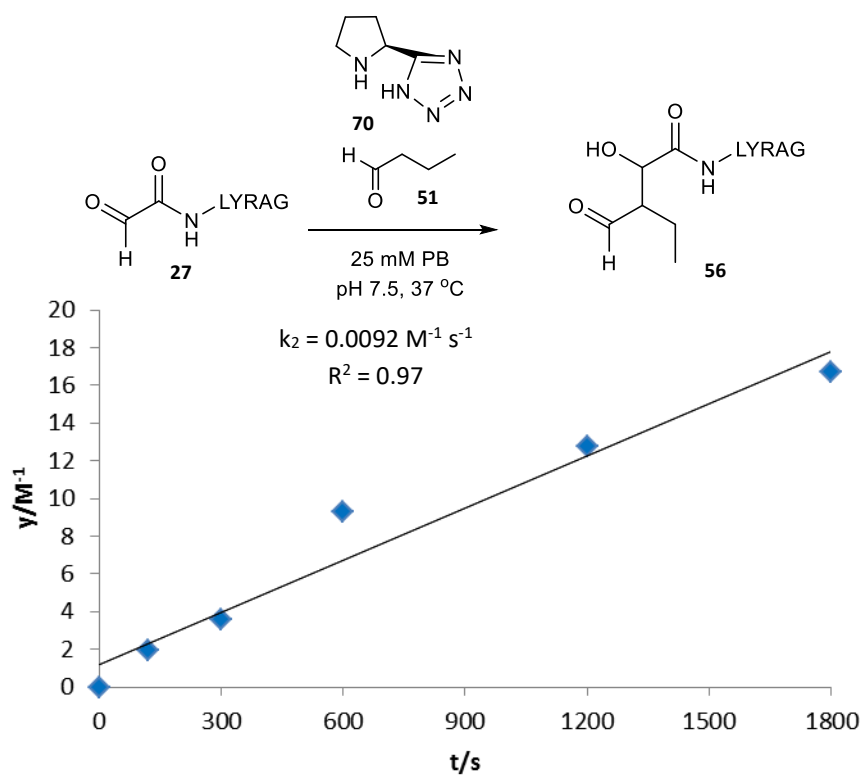
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (10 mM) using 1 mM of catalyst **70**.
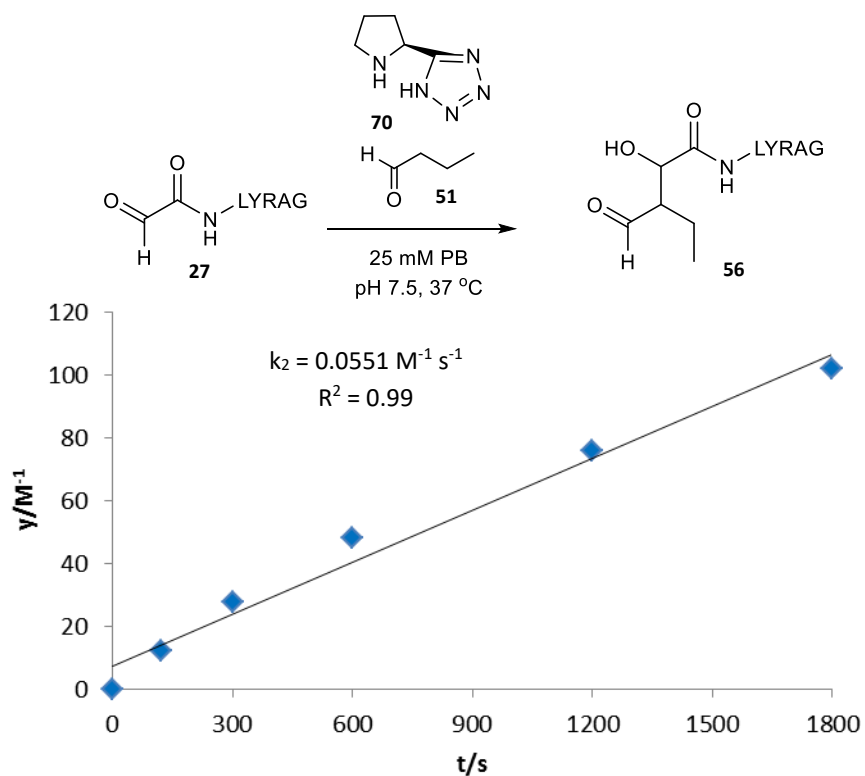
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (10 mM) using 10 mM of catalyst **70**.
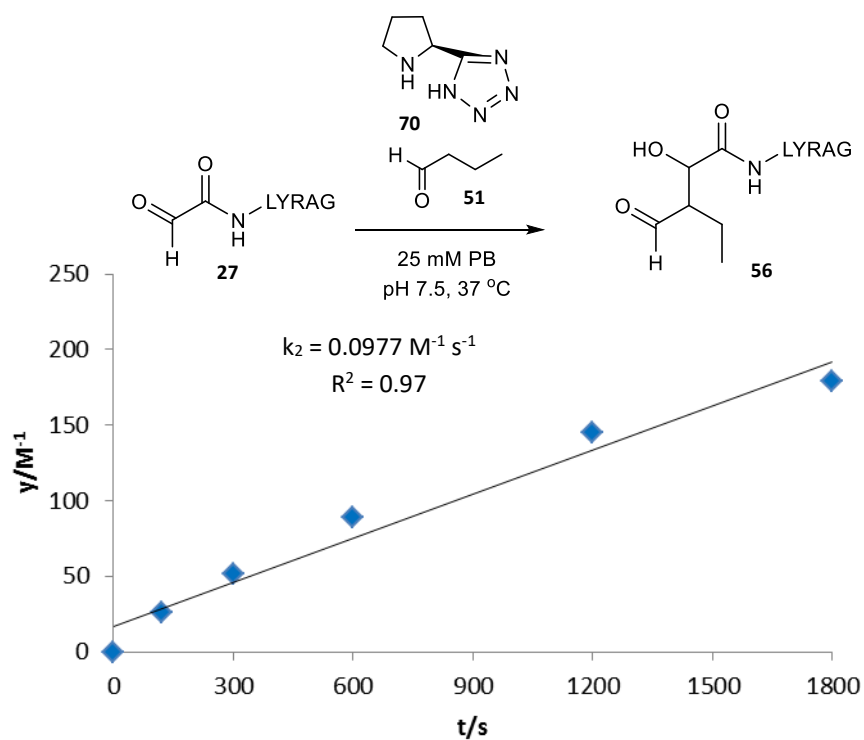
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (10 mM) using 25 mM of catalyst **70**.
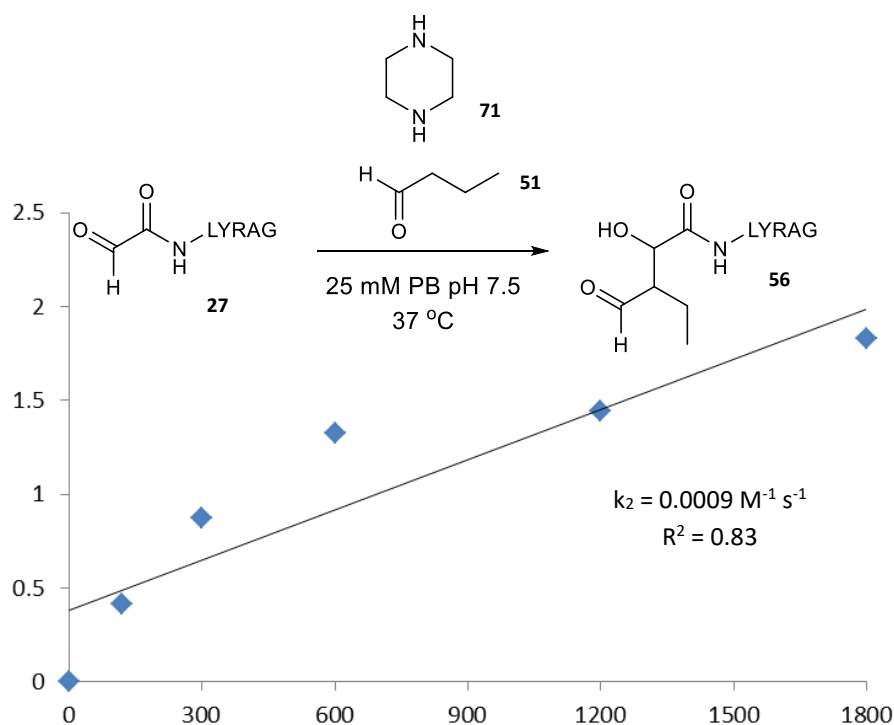
Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (100 mM) using 1 mM of catalyst **71**.

Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (100 mM) using 10 mM of catalyst **71**.

Kinetic data for OPAL of glyoxyl-LYRAG **27** (0.5 mM) with donor **51** (100 mM) using 25 mM of catalyst **71**.

Kinetic data for OPAL of glyoxyl-LYRAG **27** (50 μM) with donor **54** (150 μM) using 1 mM of catalyst **61**.

Kinetic data for OPAL of glyoxyl-LYRAG **27** (50 μM) with donor **54** (150 μM) using 10 mM of catalyst **61**.
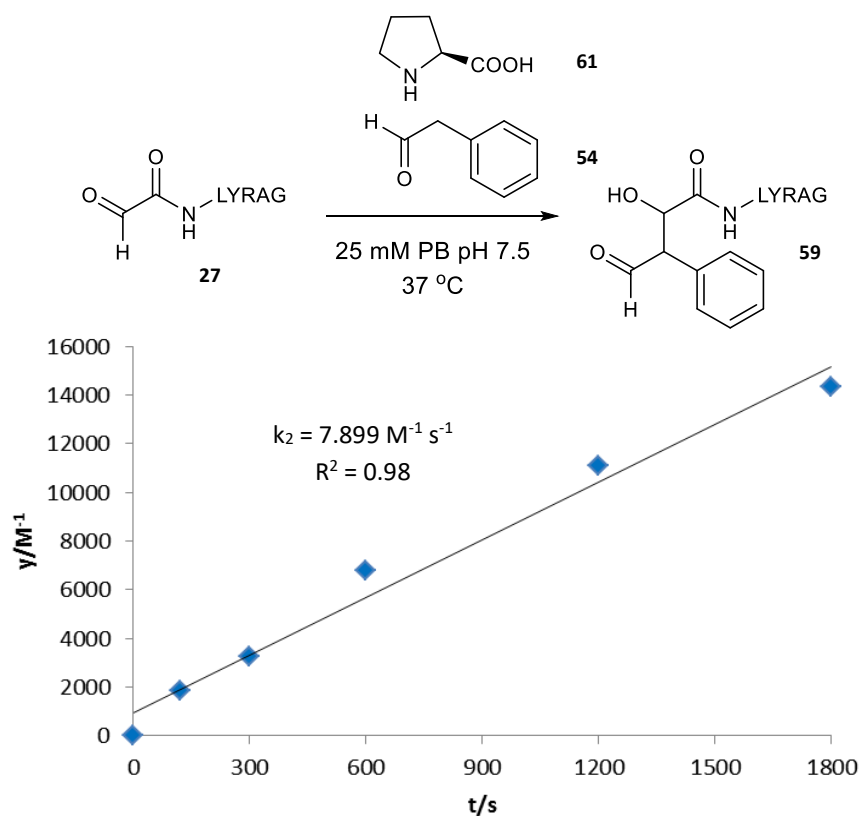
Kinetic data for OPAL of glyoxyl-LYRAG **27** (50 µM) with donor **54** (150 µM) using 25 mM of catalyst **61**.
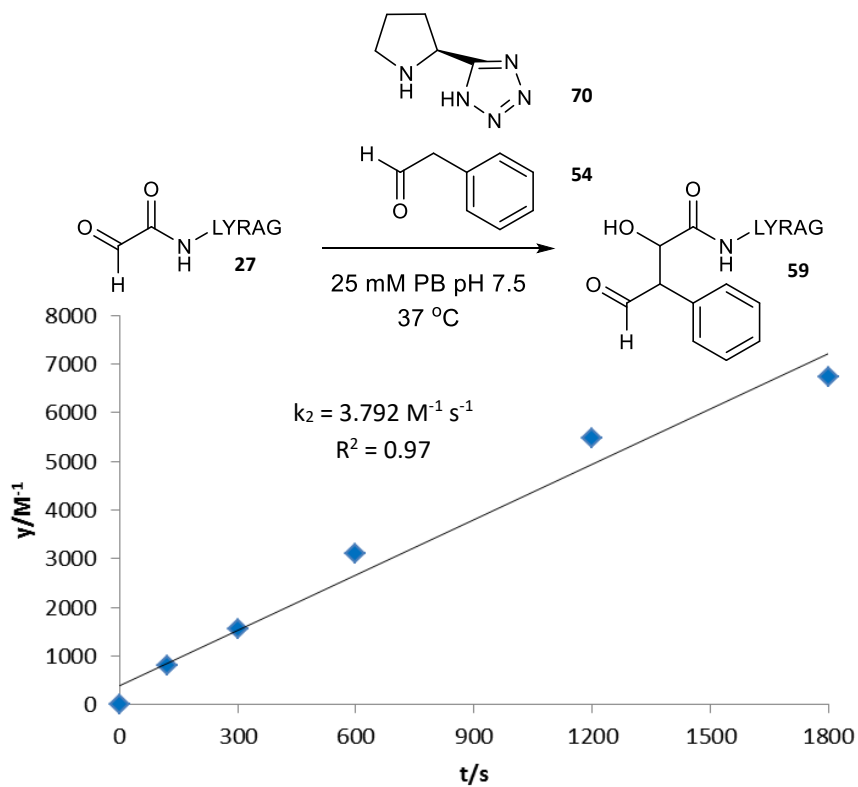
Kinetic data for OPAL of glyoxyl-LYRAG **27** (50 µM) with donor **54** (150 µM) using 1 mM of catalyst **70**.

Kinetic data for OPAL of glyoxyl-LYRAG **27** (50 μM) with donor **54** (150 μM) using 10 mM of catalyst **70**.

Kinetic data for OPAL of glyoxyl-LYRAG **27** (50 μM) with donor **54** (150 μM) using 25 mM of catalyst **70**.

Chapter 9: Appendix

NMR Data
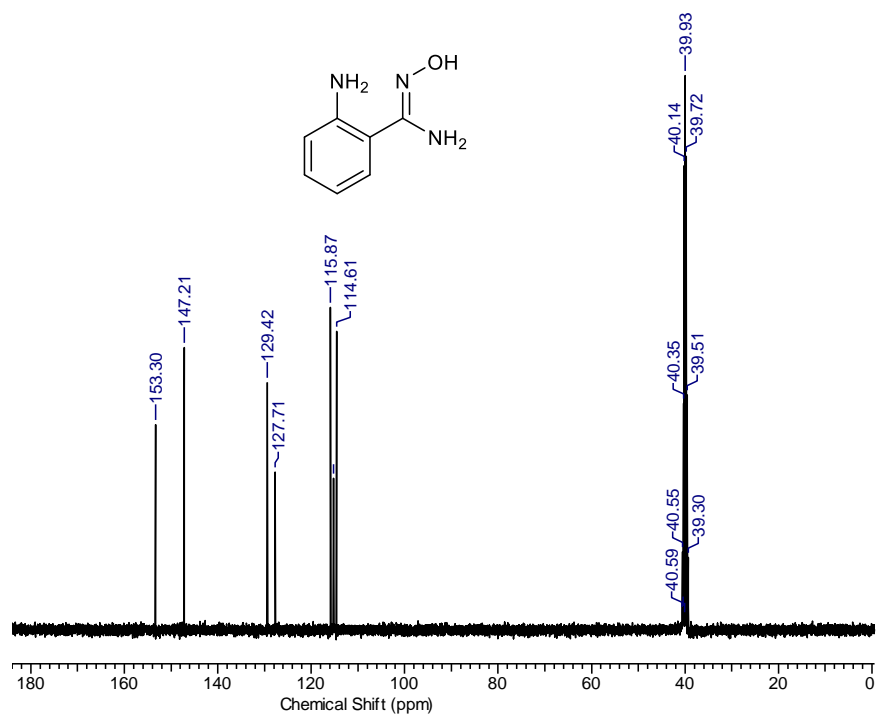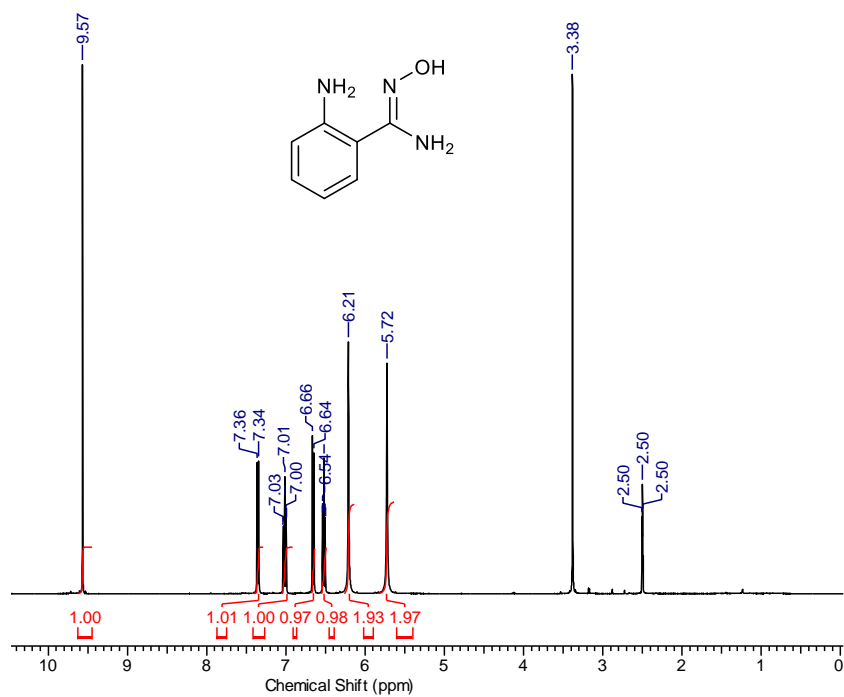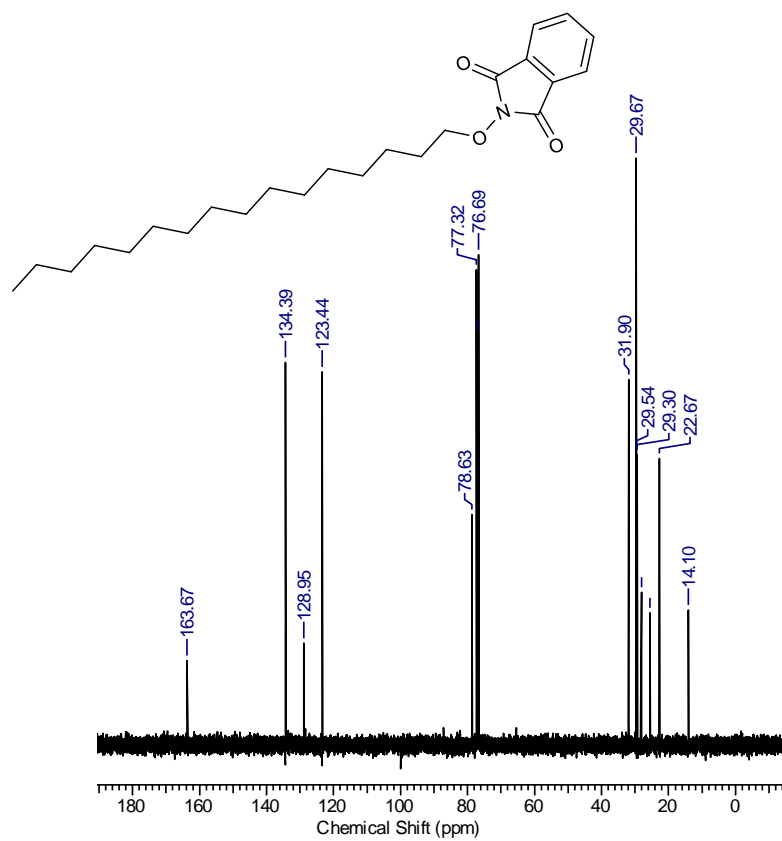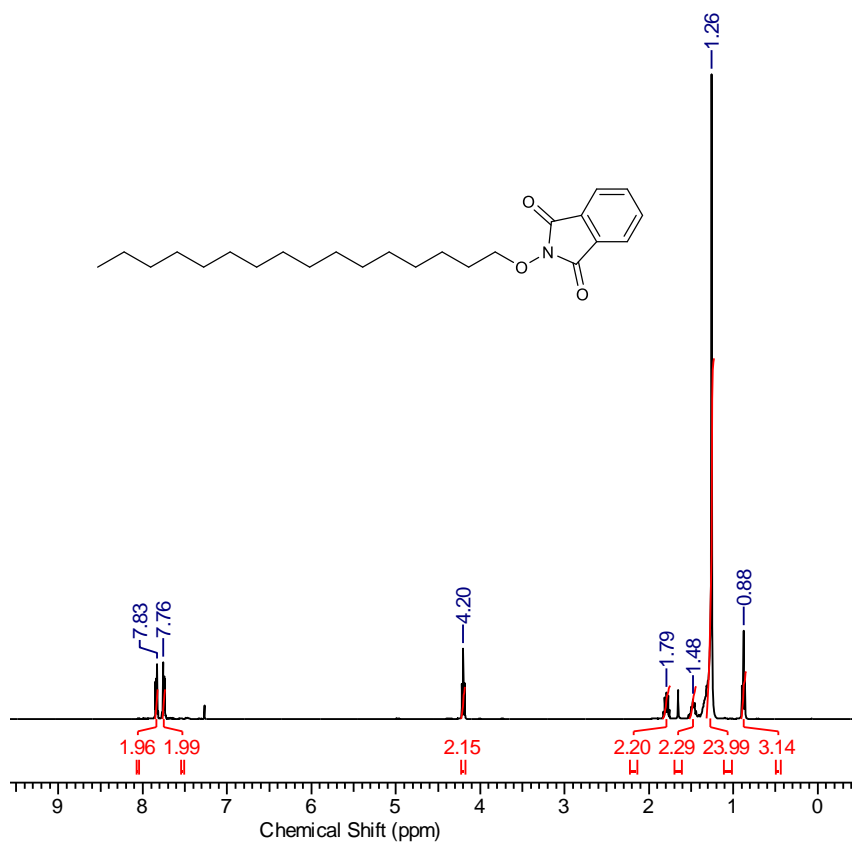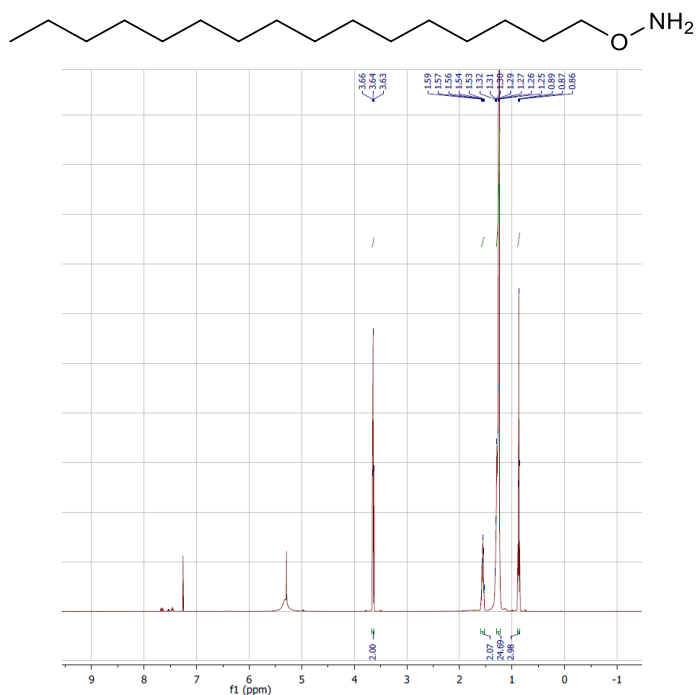
NB: This spectra was kindly recorded by Darshita Budhadev
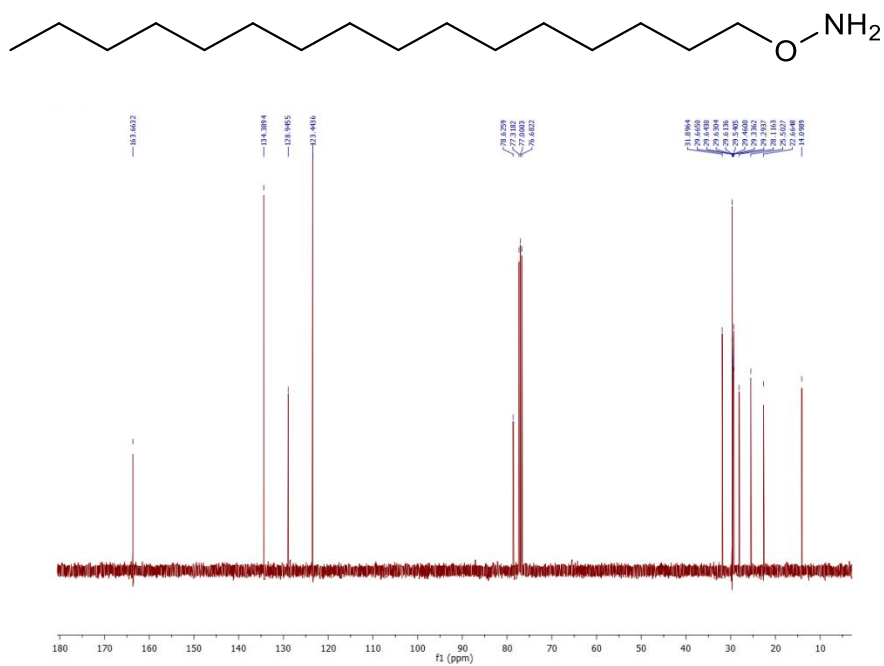
NB: This spectra was kindly recorded and processed by Darshita Budhadev.

(2kDa)

Chemical Shift (ppm)

7.82
7.75
2.01 2.00

3.63
3.37
3.08

163.3752
134.3920
128.8865
123.4265
77.3181
77.0001
76.6816
71.8390
70.4677
58.9619

f1 (ppm)

NB: This spectra was kindly recorded and processed by Darshita Budhadev.
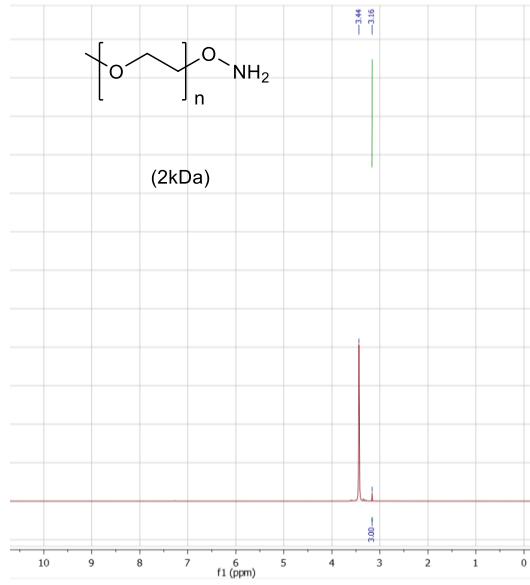
NB: This spectra was kindly recorded by Darshita Budhadev
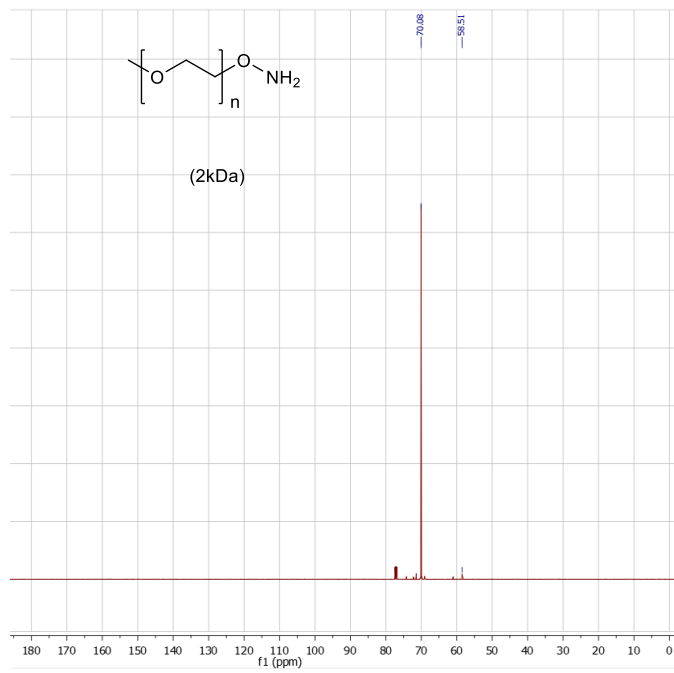
NB: This spectra was kindly recorded by Darshita Budhadev

Chapter 10: Abbreviations

ABAO	2-amino benzamidoxime
ADC	Antibody Drug Conjugate
C-C	Carbon-Carbon
CD	Circular dichroism
CNTF	Ciliary neurotrophic factor protein
CTA	Cholera toxin subunit A
CTB	Cholera toxin subunit B
CuAAC	Copper-mediated azide-alkyne cycloaddition
DAR	Drug-antibody ratio
DMF	Dimethylforamide
DMSO	Dimethylsulfoxide
E. Coli	*Escherichia Coli*
FDA	Food and Drug Administration
FGE	Formylglycine Generating Enzyme
fGly	Formylglycine
FAPP	Farnseyl aldehyde pyrophosphate
FPP	Farnseyl pyrophosphate
GFP	Green fluorescent protein
HASP	Hydrophilic acylated surface protein
HIPS	Hydrazino-*iso*-Pictet Spengler
HCTU	2-(6-Chloro-1-H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate
HPLC	High performance liquid chromatography

HRMS	High resolution mass spectrometry
HRP	Horseradish peroxidase
IL-8	Interleukin-8
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KAHA	α-keto acid hydroxylamine
LAP	Lipoic acid ligase acceptor peptide
LB	Lysogeny broth
LC-MS	Liquid chromatography-mass spectrometry
LplA	Lipoic acid ligase
LRMS	Low resolution mass spectrometry
MBP	Maltose binding protein
MNP	Multivalent nanoparticle
M. tuberculosis	*Mycobacterium tuberculosis*
MS/MS	Tandem mass spectrometry
MWCO	Molecular weight cutoff
$NaIO_4$	Sodium periodate
NaOAc	Sodium acetate
NHS	National Health Service
NMR	Nuclear magnetic resonance
NMT	*N*-myristoyltransferase
OPAL	Organocatalyst-mediated protein aldol ligation
PB	Phosphate buffer
PBS	Phosphate buffer saline

PEG	Polyethylene glycol
PFTase	Protein farnesyl-transferase
PLP	Pyridoxal-5-phosphate
PTM	Post-translational modification
ReACT	Redox-activated chemical tagging
RS	Rapoport's Salt
scFv	Single chain Fragment variable
sfGFP	Superfolder green fluorescent protein
SMCC carboxylate	Succinimidyl trans-4-(maleimidylmethyl) cyclohexane-1-
SPAAC	Strain-promoted azide-alkyne cycloaddition
SPANC	Strain-promoted alkyne nitrone cycloaddition
SPPS	Solid phase peptide synthesis
TCEP	Tris(2-carboxyethyl)phosphine
TEAB	Triethanolamine bicarbonate
TFA	Trifluoroacetic acid
tK	Trapped Knoevenagel
TKM	Tandem Knoevenagel Condensation Michael Addition
TTL	Tubulin tyrosine ligase
UV-Vis	Ultraviolet-visible
YEB	Ylide ester biotin phosphonium salt precursor
2PCA	2-pyridinecarboxaldehyde
3ForTyr	3-formyl-L-tyrosine

Chapter 11: References

1   Spears, R. J. & Fascione, M. A. Site-selective incorporation and ligation of protein aldehydes. *Org. Biomol. Chem.* **14**, 7622-7638, doi:10.1039/C6OB00778C (2016).

2   Spears, R. J. *et al. Site-Selective C-C Modification of Proteins at Neutral pH Using Organocatalyst-Mediated Cross Aldol Ligations*. (2018).

3   Stephanopoulos, N. & Francis, M. B. Choosing an effective protein bioconjugation strategy. *Nat. Chem. Biol.* **7**, 876-884 (2011).

4   Fee, C. J. Size comparison between proteins PEGylated with branched and linear poly(ethylene glycol) molecules. *Biotechnol. Bioeng.* **98**, 725-731, doi:10.1002/bit.21482 (2007).

5   Dozier, J. & Distefano, M. Site-Specific PEGylation of Therapeutic Proteins. *Int. J. Mol. Sci.* **16**, 25831 (2015).

6   Fach, M., Radi, L. & Wich, P. R. Nanoparticle Assembly of Surface-Modified Proteins. *J. Am. Chem. Soc.* **138**, 14820-14823, doi:10.1021/jacs.6b06243 (2016).

7   Polito, L. *et al.* One-step bioengineering of magnetic nanoparticles via a surface diazo transfer/azide-alkyne click reaction sequence. *Chem. Commun.*, 621-623, doi:10.1039/B716113A (2008).

8   Colombo, M. *et al.* Site-Specific Conjugation of ScFvs Antibodies to Nanoparticles by Bioorthogonal Strain-Promoted Alkyne–Nitrone Cycloaddition. *Angew. Chem. Int. Ed.* **51**, 496-499, doi:10.1002/anie.201106775 (2012).

9   Wammes, A. E. M. *et al.* Site-specific peptide and protein immobilization on surface plasmon resonance chips via strain-promoted cycloaddition. *Lab Chip* **13**, 1863-1867, doi:10.1039/C3LC41338A (2013).

10   Laughlin, S. T., Baskin, J. M., Amacher, S. L. & Bertozzi, C. R. In Vivo Imaging of Membrane-Associated Glycans in Developing Zebrafish. *Science* **320**, 664-667, doi:10.1126/science.1155106 (2008).

11   Agarwal, P., Beahm, B. J., Shieh, P. & Bertozzi, C. R. Systemic Fluorescence Imaging of Zebrafish Glycans with Bioorthogonal Chemistry. *Angew. Chem. Int. Ed.* **54**, 11504-11510, doi:10.1002/anie.201504249 (2015).

12   Row, R. D., Shih, H.-W., Alexander, A. T., Mehl, R. A. & Prescher, J. A. Cyclopropenones for Metabolic Targeting and Sequential Bioorthogonal

Labeling. *J. Am. Chem. Soc.* **139**, 7370-7375, doi:10.1021/jacs.7b03010 (2017).

13   Agarwal, P. & Bertozzi, C. R. Site-Specific Antibody–Drug Conjugates: The Nexus of Bioorthogonal Chemistry, Protein Engineering, and Drug Development. *Bioconjugate Chem.* **26**, 176-192, doi:10.1021/bc5004982 (2015).

14   Beck, A., Goetsch, L., Dumontet, C. & Corvaia, N. Strategies and challenges for the next generation of antibody-drug conjugates. *Nat. Rev. Drug Discov.* **16**, 315-337, doi:10.1038/nrd.2016.268 (2017).

15   Krall, N., da Cruz, F. P., Boutureira, O. & Bernardes, G. J. L. Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem.* **8**, 103-113, doi:10.1038/nchem.2393 (2016).

16   Smith, E. L. *et al.* Chemoenzymatic Fc Glycosylation via Engineered Aldehyde Tags. *Bioconjugate Chem.* **25**, 788-795, doi:10.1021/bc500061s (2014).

17   Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science* **354**, doi:10.1126/science.aag1465 (2016).

18   Yang, A. *et al.* A chemical biology route to site-specific authentic protein modifications. *Science* **354**, 623-626, doi:10.1126/science.aah4428 (2016).

19   Boutureira, O. & Bernardes, G. J. L. Advances in Chemical Protein Modification. *Chem. Rev.* **115**, 2174-2195, doi:10.1021/cr500399p (2015).

20   Sletten, E. M. & Bertozzi, C. R. Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angew. Chem. Int. Ed.* **48**, 6974-6998, doi:10.1002/anie.200900942 (2009).

21   Baslé, E., Joubert, N. & Pucheault, M. Protein Chemical Modification on Endogenous Amino Acids. *Chem. Biol.* **17**, 213-227, doi:http://dx.doi.org/10.1016/j.chembiol.2010.02.008 (2010).

22   Koniev, O. & Wagner, A. Developments and recent advancements in the field of endogenous amino acid selective bond forming reactions for bioconjugation. *Chem. Soc. Rev.* **44**, 5495-5551, doi:10.1039/C5CS00048C (2015).

23   Chen, X., Muthoosamy, K., Pfisterer, A., Neumann, B. & Weil, T. Site-Selective Lysine Modification of Native Proteins and Peptides via Kinetically

Controlled Labeling. *Bioconjugate Chem.* **23**, 500-508, doi:10.1021/bc200556n (2012).

24   Tuls, J., Geren, L. & Millett, F. Fluorescein isothiocyanate specifically modifies lysine 338 of cytochrome P-450scc and inhibits adrenodoxin binding. *J. Biol. Chem.* **264**, 16421-16425 (1989).

25   Moosmann, A., Blath, J., Lindner, R., Müller, E. & Böttinger, H. Aldehyde PEGylation Kinetics: A Standard Protein versus a Pharmaceutically Relevant Single Chain Variable Fragment. *Bioconjugate Chem.* **22**, 1545-1558, doi:10.1021/bc200090x (2011).

26   Jackson, D. Y. Processes for Constructing Homogeneous Antibody Drug Conjugates. *Org. Process Res. Dev.* **20**, 852-866, doi:10.1021/acs.oprd.6b00067 (2016).

27   Sechi, S. & Chait, B. T. Modification of Cysteine Residues by Alkylation. A Tool in Peptide Mapping and Protein Identification. *Anal. Chem.* **70**, 5150-5158, doi:10.1021/ac9806005 (1998).

28   Kim, Y. *et al.* Efficient Site-Specific Labeling of Proteins via Cysteines. *Bioconjugate Chem.* **19**, 786-791, doi:10.1021/bc7002499 (2008).

29   Abbas, A., Xing, B. & Loh, T.-P. Allenamides as Orthogonal Handles for Selective Modification of Cysteine in Peptides and Proteins. *Angew. Chem. Int. Ed.* **53**, 7491-7494, doi:10.1002/anie.201403121 (2014).

30   Chalker, J. M., Bernardes, G. J. L. & Davis, B. G. A "Tag-and-Modify" Approach to Site-Selective Protein Modification. *Acc. Chem. Res.* **44**, 730-741, doi:10.1021/ar200056q (2011).

31   Bernardim, B. *et al.* Stoichiometric and irreversible cysteine-selective protein modification using carbonylacrylic reagents. *Nat. Commun.* **7**, 13128, doi:10.1038/ncomms13128

https://www.nature.com/articles/ncomms13128#supplementary-information (2016).

32   Chalker, J. M., Lercher, L., Rose, N. R., Schofield, C. J. & Davis, B. G. Conversion of Cysteine into Dehydroalanine Enables Access to Synthetic Histones Bearing Diverse Post-Translational Modifications. *Angew. Chem. Int. Ed.* **51**, 1835-1839, doi:10.1002/anie.201106432 (2012).

33   Smith, M. E. B. *et al.* Protein Modification, Bioconjugation, and Disulfide Bridging Using Bromomaleimides. *J. Am. Chem. Soc.* **132**, 1960-1965, doi:10.1021/ja908610s (2010).

34   Maruani, A. *et al.* A plug-and-play approach to antibody-based therapeutics via a chemoselective dual click strategy. *Nat. Commun.* **6**, doi:10.1038/ncomms7645 (2015).

35   Lee, M. T. W., Maruani, A., Baker, J. R., Caddick, S. & Chudasama, V. Next-generation disulfide stapling: reduction and functional re-bridging all in one. *Chem. Sci.* **7**, 799-802, doi:10.1039/C5SC02666K (2016).

36   Sato, S., Nakamura, K. & Nakamura, H. Tyrosine-Specific Chemical Modification with in Situ Hemin-Activated Luminol Derivatives. *ACS Chem. Biol.* **10**, 2633-2640, doi:10.1021/acschembio.5b00440 (2015).

37   Addy, P. S., Erickson, S. B., Italia, J. S. & Chatterjee, A. A Chemoselective Rapid Azo-Coupling Reaction (CRACR) for Unclickable Bioconjugation. *J. Am. Chem. Soc.* **139**, 11670-11673, doi:10.1021/jacs.7b05125 (2017).

38   Omura, Y. *et al.* Regioselective Mannich reaction of phenolic compounds and its application to the synthesis of new chitosan derivatives. *Tetrahedron Lett.* **42**, 7273-7275, doi:http://dx.doi.org/10.1016/S0040-4039(01)01491-5 (2001).

39   Antos, J. M. & Francis, M. B. Selective Tryptophan Modification with Rhodium Carbenoids in Aqueous Solution. *J. Am. Chem. Soc.* **126**, 10256-10257, doi:10.1021/ja047272c (2004).

40   Seki, Y. *et al.* Transition Metal-Free Tryptophan-Selective Bioconjugation of Proteins. *J. Am. Chem. Soc.* **138**, 10798-10801, doi:10.1021/jacs.6b06692 (2016).

41   Lin, S. *et al.* Redox-based reagents for chemoselective methionine bioconjugation. *Science* **355**, 597-602, doi:10.1126/science.aal3316 (2017).

42   Zhang, C. *et al.* π-Clamp Mediated Cysteine Conjugation. *Nature Chem.* **8**, 120-128, doi:10.1038/nchem.2413 (2016).

43   Dai, P. *et al.* Salt Effect Accelerates Site-Selective Cysteine Bioconjugation. *ACS Central Science* **2**, 637-646, doi:10.1021/acscentsci.6b00180 (2016).

44   Matos, M. J. *et al.* Chemo- and Regioselective Lysine Modification on Native Proteins. *J. Am. Chem. Soc.* **140**, 4004-4017, doi:10.1021/jacs.7b12874 (2018).

45   Rosen, C. B. & Francis, M. B. Targeting the N terminus for site-selective protein modification. *Nat. Chem. Biol.* **13**, 697, doi:10.1038/nchembio.2416 (2017).

46   Chan, A. O.-Y. *et al.* Modification of N-Terminal α-Amino Groups of Peptides and Proteins Using Ketenes. *J. Am. Chem. Soc.* **134**, 2589-2598, doi:10.1021/ja208009r (2012).

47   MacDonald, J. I., Munch, H. K., Moore, T. & Francis, M. B. One-step site-specific modification of native proteins with 2-pyridinecarboxyaldehydes. *Nat. Chem. Biol.* **11**, 326-331, doi:10.1038/nchembio.1792 (2015).

48   Jeon, J. *et al.* Efficient Method for Site-Specific 18F-Labeling of Biomolecules Using the Rapid Condensation Reaction between 2-Cyanobenzothiazole and Cysteine. *Bioconjugate Chem.* **23**, 1902-1908, doi:10.1021/bc300273m (2012).

49   Dawson, P. E. & Kent, S. B. H. SYNTHESIS OF NATIVE PROTEINS BY CHEMICAL LIGATION 1. *Annu. Rev. Biochem.* **69**, 923-960, doi:doi:10.1146/annurev.biochem.69.1.923 (2000).

50   Seim, K. L., Obermeyer, A. C. & Francis, M. B. Oxidative Modification of Native Protein Residues Using Cerium(IV) Ammonium Nitrate. *J. Am. Chem. Soc.* **133**, 16970-16976, doi:10.1021/ja206324q (2011).

51   Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the Genetic Code of Escherichia coli. *Science* **292**, 498-500, doi:10.1126/science.1060077 (2001).

52   Wang, L. & Schultz, P. G. Expanding the Genetic Code. *Angew. Chem. Int. Ed.* **44**, 34-66, doi:10.1002/anie.200460627 (2005).

53   Brabham, R. & Fascione, M. A. Pyrrolysine Amber Stop-Codon Suppression: Development and Applications. *ChemBioChem* **18**, 1973-1983, doi:10.1002/cbic.201700148 (2017).

54   Saxon, E. *et al.* Investigating Cellular Metabolism of Synthetic Azidosugars with the Staudinger Ligation. *J. Am. Chem. Soc.* **124**, 14893-14902, doi:10.1021/ja027748x (2002).

55   Saxon, E., Armstrong, J. I. & Bertozzi, C. R. A "Traceless" Staudinger Ligation for the Chemoselective Synthesis of Amide Bonds. *Org. Lett.* **2**, 2141-2143, doi:10.1021/ol006054v (2000).

56   Speers, A. E. & Cravatt, B. F. Profiling Enzyme Activities In Vivo Using Click Chemistry Methods. *Chem. Biol.* **11**, 535-546, doi:https://doi.org/10.1016/j.chembiol.2004.03.012 (2004).

57   Agard, N. J., Prescher, J. A. & Bertozzi, C. R. A Strain-Promoted [3 + 2] Azide−Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems. *J. Am. Chem. Soc.* **126**, 15046-15047, doi:10.1021/ja044996f (2004).

58   Borrmann, A. *et al.* Strain-Promoted Oxidation-Controlled Cyclooctyne–1,2-Quinone Cycloaddition (SPOCQ) for Fast and Activatable Protein Conjugation. *Bioconjugate Chem.* **26**, 257-261, doi:10.1021/bc500534d (2015).

59   Li, N., Lim, R. K. V., Edwardraja, S. & Lin, Q. Copper-Free Sonogashira Cross-Coupling for Functionalization of Alkyne-Encoded Proteins in Aqueous Medium and in Bacterial Cells. *J. Am. Chem. Soc.* **133**, 15316-15319, doi:10.1021/ja2066913 (2011).

60   Hauke, S. *et al.* Two-Step Protein Labeling Utilizing Lipoic Acid Ligase and Sonogashira Cross-Coupling. *Bioconjugate Chem.* **25**, 1632-1637, doi:10.1021/bc500349h (2014).

61   Lampkowski, J. S., Villa, J. K., Young, T. S. & Young, D. D. Development and Optimization of Glaser–Hay Bioconjugations. *Angew. Chem. Int. Ed.* **54**, 9343-9346, doi:10.1002/anie.201502676 (2015).

62   Maza, J. C., McKenna, J. R., Raliski, B. K., Freedman, M. T. & Young, D. D. Synthesis and Incorporation of Unnatural Amino Acids To Probe and Optimize Protein Bioconjugations. *Bioconjugate Chem.* **26**, 1884-1889, doi:10.1021/acs.bioconjchem.5b00424 (2015).

63   Wang, K. *et al.* Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. *Nature Chem.* **6**, 393, doi:10.1038/nchem.1919

https://www.nature.com/articles/nchem.1919#supplementary-information (2014).

64   Machida, T., Lang, K., Xue, L., Chin, J. W. & Winssinger, N. Site-Specific Glycoconjugation of Protein via Bioorthogonal Tetrazine Cycloaddition with a Genetically Encoded trans-Cyclooctene or Bicyclononyne. *Bioconjugate Chem.* **26**, 802-806, doi:10.1021/acs.bioconjchem.5b00101 (2015).

65   Groth, T. & Meldal, M. N-Terminal Peptide Aldehydes as Electrophiles in Combinatorial Solid Phase Synthesis of Novel Peptide Isosteres. *J. Comb. Chem.* **3**, 45-63, doi:10.1021/cc000058+ (2001).

66   Pattabiraman, V. R. & Bode, J. W. Rethinking amide bond synthesis. *Nature* **480**, 471-479 (2011).

67   Raj, M., Wu, H., Blosser, S. L., Vittoria, M. A. & Arora, P. S. Aldehyde Capture Ligation for Synthesis of Native Peptide Bonds. *J. Am. Chem. Soc.* **137**, 6932-6940, doi:10.1021/jacs.5b03538 (2015).

68   Levine, P. M., Craven, T. W., Bonneau, R. & Kirshenbaum, K. Semisynthesis of Peptoid–Protein Hybrids by Chemical Ligation at Serine. *Org. Lett.* **16**, 512-515, doi:10.1021/ol4033978 (2014).

69   Lee, C. L. & Li, X. Serine/threonine ligation for the chemical synthesis of proteins. *Curr. Opin. Chem. Biol.* **22**, 108-114, doi:http://dx.doi.org/10.1016/j.cbpa.2014.09.023 (2014).

70   Murar, C. E., Thuaud, F. & Bode, J. W. KAHA Ligations That Form Aspartyl Aldehyde Residues as Synthetic Handles for Protein Modification and Purification. *J. Am. Chem. Soc.* **136**, 18140-18148, doi:10.1021/ja511231f (2014).

71   Geoghegan, K. F. & Stroh, J. G. Site-directed conjugation of nonpeptide groups to peptides and proteins via periodate oxidation of a 2-amino alcohol. Application to modification at N-terminal serine. *Bioconjugate Chem.* **3**, 138-146, doi:10.1021/bc00014a008 (1992).

72   Sklarz, B. Organic chemistry of periodates. *Q. Rev., Chem. Soc.* **21**, 3-28, doi:10.1039/QR9672100003 (1967).

73   El-Mahdi, O. & Melnyk, O. α-Oxo Aldehyde or Glyoxylyl Group Chemistry in Peptide Bioconjugation. *Bioconjugate Chem.* **24**, 735-765, doi:10.1021/bc300516f (2013).

74   Barlow, C. B., Guthrie, R. D. & Prior, A. M. Periodate oxidation of amino-sugars. *Chem. Commun. (London)*, 268-269 (1966).

75   Gaertner, H. F. *et al.* Construction of protein analogs by site-specific condensation of unprotected fragments. *Bioconjugate Chem.* **3**, 262-268, doi:10.1021/bc00015a010 (1992).

76   Chen, J., Zeng, W., Offord, R. & Rose, K. A Novel Method for the Rational Construction of Well-Defined Immunogens: The Use of Oximation To

Conjugate Cholera Toxin B Subunit to a Peptide−Polyoxime Complex. *Bioconjugate Chem.* **14**, 614-618, doi:10.1021/bc025651u (2003).

77   Branson, T. R. *et al.* A Protein-Based Pentavalent Inhibitor of the Cholera Toxin B-Subunit. *Angew. Chem. Int. Ed.* **53**, 8323-8327, doi:10.1002/anie.201404397 (2014).

78   Torchinsky, Y. M. Transamination: its discovery, biological and chemical aspects (1937–1987). *Trends Biochem. Sci.* **12**, 115-117, doi:http://dx.doi.org/10.1016/0968-0004(87)90052-1 (1987).

79   Gilmore, J. M., Scheck, R. A., Esser-Kahn, A. P., Joshi, N. S. & Francis, M. B. N-Terminal Protein Modification through a Biomimetic Transamination Reaction. *Angew. Chem. Int. Ed.* **45**, 5307-5311, doi:10.1002/anie.200600368 (2006).

80   Scheck, R. A., Dedeo, M. T., Iavarone, A. T. & Francis, M. B. Optimization of a Biomimetic Transamination Reaction. *J. Am. Chem. Soc.* **130**, 11762-11770, doi:10.1021/ja802495w (2008).

81   Witus, L. S. *et al.* Identification of Highly Reactive Sequences For PLP-Mediated Bioconjugation Using a Combinatorial Peptide Library. *J. Am. Chem. Soc.* **132**, 16812-16817, doi:10.1021/ja105429n (2010).

82   Zhang, M., Zhang, X., Li, J., Guo, Q. & Xiao, Q. A New Pyridoxal Derivative for Transamination of N-Terminus of Proteins. *Chin. J. Chem.* **29**, 1715-1720, doi:10.1002/cjoc.201180306 (2011).

83   Witus, L. S. *et al.* Site-Specific Protein Transamination Using N-Methylpyridinium-4-carboxaldehyde. *J. Am. Chem. Soc.* **135**, 17223-17229, doi:10.1021/ja408868a (2013).

84   Palla, K. S., Witus, L. S., Mackenzie, K. J., Netirojjanakul, C. & Francis, M. B. Optimization and Expansion of a Site-Selective N-Methylpyridinium-4-carboxaldehyde-Mediated Transamination for Bacterially Expressed Proteins. *J. Am. Chem. Soc.* **137**, 1123-1129, doi:10.1021/ja509955n (2015).

85   Lukatela, G. *et al.* Crystal Structure of Human Arylsulfatase A: The Aldehyde Function and the Metal Ion at the Active Site Suggest a Novel Mechanism for Sulfate Ester Hydrolysis. *Biochemistry* **37**, 3654-3664, doi:10.1021/bi9714924 (1998).

86   Carrico, I. S., Carlson, B. L. & Bertozzi, C. R. Introducing genetically encoded aldehydes into proteins. *Nat. Chem. Biol.* **3**, 321-322,

doi:http://www.nature.com/nchembio/journal/v3/n6/suppinfo/nchembio878_S1.html (2007).

87   Hudak, J. E., Yu, H. H. & Bertozzi, C. R. Protein Glycoengineering Enabled by the Versatile Synthesis of Aminooxy Glycans and the Genetically Encoded Aldehyde Tag. *J. Am. Chem. Soc.* **133**, 16127-16135, doi:10.1021/ja206023e (2011).

88   Agarwal, P., van der Weijden, J., Sletten, E. M., Rabuka, D. & Bertozzi, C. R. A Pictet-Spengler ligation for protein chemical modification. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 46-51, doi:10.1073/pnas.1213186110 (2013).

89   Agarwal, P. *et al.* Hydrazino-Pictet-Spengler Ligation as a Biocompatible Method for the Generation of Stable Protein Conjugates. *Bioconjugate Chem.* **24**, 846-851, doi:10.1021/bc400042a (2013).

90   Drake, P. M. *et al.* Aldehyde Tag Coupled with HIPS Chemistry Enables the Production of ADCs Conjugated Site-Specifically to Different Antibody Regions with Distinct in Vivo Efficacy and PK Outcomes. *Bioconjugate Chem.* **25**, 1331-1341, doi:10.1021/bc500189z (2014).

91   Kudirka, R. *et al.* Generating Site-Specifically Modified Proteins via a Versatile and Stable Nucleophilic Carbon Ligation. *Chem. Biol.* **22**, 293-298, doi:http://dx.doi.org/10.1016/j.chembiol.2014.11.019 (2015).

92   Shi, X. *et al.* Quantitative fluorescence labeling of aldehyde-tagged proteins for single-molecule imaging. *Nat. Meth.* **9**, 499-503, doi:http://www.nature.com/nmeth/journal/v9/n5/abs/nmeth.1954.html#supplementary-information (2012).

93   Wu, P. *et al.* Site-specific chemical modification of recombinant proteins produced in mammalian cells by using the genetically encoded aldehyde tag. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 3000-3005, doi:10.1073/pnas.0807820106 (2009).

94   Holder, P. G. *et al.* Reconstitution of Formylglycine-generating Enzyme with Copper(II) for Aldehyde Tag Conversion. *J. Biol. Chem.* **290**, 15730-15745, doi:10.1074/jbc.M115.652669 (2015).

95   Appel, M. J. & Bertozzi, C. R. Formylglycine, a Post-Translationally Generated Residue with Unique Catalytic Capabilities and Biotechnology Applications. *ACS Chem. Biol.* **10**, 72-84, doi:10.1021/cb500897w (2015).

96   Gauchet, C., Labadie, G. R. & Poulter, C. D. Regio- and Chemoselective Covalent Immobilization of Proteins through Unnatural Amino Acids. *J. Am. Chem. Soc.* **128**, 9274-9275, doi:10.1021/ja061131o (2006).

97   Nguyen, U. T. T. *et al.* Exploiting the Substrate Tolerance of Farnesyltransferase for Site-Selective Protein Derivatization. *ChemBioChem* **8**, 408-423, doi:10.1002/cbic.200600440 (2007).

98   Rashidian, M., Dozier, J. K., Lenevich, S. & Distefano, M. D. Selective labeling of polypeptides using protein farnesyltransferase via rapid oxime ligation. *Chem. Commun.* **46**, 8998-9000, doi:10.1039/C0CC03305G (2010).

99   Rashidian, M., Song, J. M., Pricer, R. E. & Distefano, M. D. Chemoenzymatic Reversible Immobilization and Labeling of Proteins without Prior Purification. *J. Am. Chem. Soc.* **134**, 8455-8467, doi:10.1021/ja211308s (2012).

100   Rashidian, M. *et al.* Simultaneous Dual Protein Labeling Using a Triorthogonal Reagent. *J. Am. Chem. Soc.* **135**, 16388-16396, doi:10.1021/ja403813b (2013).

101   Dozier, J. K. *et al.* Engineering Protein Farnesyltransferase for Enzymatic Protein Labeling Applications. *Bioconjugate Chem.* **25**, 1203-1212, doi:10.1021/bc500240p (2014).

102   Puthenveetil, S., Liu, D. S., White, K. A., Thompson, S. & Ting, A. Y. Yeast Display Evolution of a Kinetically Efficient 13-Amino Acid Substrate for Lipoic Acid Ligase. *J. Am. Chem. Soc.* **131**, 16430-16438, doi:10.1021/ja904596f (2009).

103   Cohen, J. D., Zou, P. & Ting, A. Y. Site-Specific Protein Modification Using Lipoic Acid Ligase and Bis-Aryl Hydrazone Formation. *ChemBioChem* **13**, 888-894, doi:10.1002/cbic.201100764 (2012).

104   Gray, M. A., Tao, R. N., DePorter, S. M., Spiegel, D. A. & McNaughton, B. R. A Nanobody Activation Immunotherapeutic that Selectively Destroys HER2-Positive Breast Cancer Cells. *ChemBioChem* **17**, 155-158, doi:10.1002/cbic.201500591 (2016).

105   Schumacher, D. *et al.* Versatile and Efficient Site-Specific Protein Functionalization by Tubulin Tyrosine Ligase. *Angew. Chem. Int. Ed.* **54**, 13787-13791, doi:10.1002/anie.201505456 (2015).

106   Wang, L., Zhang, Z., Brock, A. & Schultz, P. G. Addition of the keto functional group to the genetic code of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 56-61, doi:10.1073/pnas.0234824100 (2003).

107   Huang, Y. *et al.* Genetic incorporation of an aliphatic keto-containing amino acid into proteins for their site-specific modifications. *Bioorg. Med. Chem. Lett.* **20**, 878-880, doi:http://dx.doi.org/10.1016/j.bmcl.2009.12.077 (2010).

108   Tuley, A. *et al.* The genetic incorporation of thirteen novel non-canonical amino acids. *Chem. Commun.* **50**, 2673-2675, doi:10.1039/C3CC49068H (2014).

109   Tuley, A., Lee, Y.-J., Wu, B., Wang, Z. U. & Liu, W. R. A genetically encoded aldehyde for rapid protein labelling. *Chem. Commun.* **50**, 7424-7426, doi:10.1039/C4CC02000F (2014).

110   Zeng, H., Xie, J. & Schultz, P. G. Genetic introduction of a diketone-containing amino acid into proteins. *Bioorg. Med. Chem. Lett.* **16**, 5356-5359, doi:http://dx.doi.org/10.1016/j.bmcl.2006.07.094 (2006).

111   Liu, H., Wang, L., Brock, A., Wong, C.-H. & Schultz, P. G. A Method for the Generation of Glycoprotein Mimetics. *J. Am. Chem. Soc.* **125**, 1702-1703, doi:10.1021/ja029433n (2003).

112   Axup, J. Y. *et al.* Synthesis of site-specific antibody-drug conjugates using unnatural amino acids. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 16101-16106, doi:10.1073/pnas.1211023109 (2012).

113   Tian, F. *et al.* A general approach to site-specific antibody drug conjugates. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 1766-1771, doi:10.1073/pnas.1321237111 (2014).

114   Wang, Z. A. *et al.* A Genetically Encoded Allysine for the Synthesis of Proteins with Site-Specific Lysine Dimethylation. *Angew. Chem. Int. Ed.* **56**, 212-216, doi:10.1002/anie.201609452 (2017).

115   Bi, X., Pasunooti, K. K., Lescar, J. & Liu, C.-F. Thiazolidine-Masked α-Oxo Aldehyde Functionality for Peptide and Protein Modification. *Bioconjugate Chem.* **28**, 325-329, doi:10.1021/acs.bioconjchem.6b00667 (2017).

116   Brabham, R. L. *et al.* Palladium-unleashed proteins: gentle aldehyde decaging for site-selective protein modification. *Chem. Commun.* **54**, 1501-1504, doi:10.1039/C7CC07740H (2018).

117   Chelius, D. & Shaler, T. A. Capture of Peptides with N-Terminal Serine and Threonine: A Sequence-Specific Chemical Method for Peptide Mixture Simplification. *Bioconjugate Chem.* **14**, 205-211, doi:10.1021/bc025605u (2003).

118   Dirksen, A., Dirksen, S., Hackeng, T. M. & Dawson, P. E. Nucleophilic Catalysis of Hydrazone Formation and Transimination: Implications for Dynamic Covalent Chemistry. *J. Am. Chem. Soc.* **128**, 15602-15603, doi:10.1021/ja067189k (2006).

119   Crisalli, P. & Kool, E. T. Water-Soluble Organocatalysts for Hydrazone and Oxime Formation. *J. Org. Chem* **78**, 1184-1189, doi:10.1021/jo302746p (2013).

120   Dirksen, A. & Dawson, P. E. Rapid Oxime and Hydrazone Ligations with Aromatic Aldehydes for Biomolecular Labeling. *Bioconjugate Chem.* **19**, 2543-2548, doi:10.1021/bc800310p (2008).

121   Dirksen, A., Hackeng, T. M. & Dawson, P. E. Nucleophilic Catalysis of Oxime Ligation. *Angew. Chem. Int. Ed.* **45**, 7581-7584, doi:10.1002/anie.200602877 (2006).

122   Rashidian, M. *et al.* A Highly Efficient Catalyst for Oxime Ligation and Hydrazone–Oxime Exchange Suitable for Bioconjugation. *Bioconjugate Chem.* **24**, 333-342, doi:10.1021/bc3004167 (2013).

123   Wendeler, M., Grinberg, L., Wang, X., Dawson, P. E. & Baca, M. Enhanced Catalysis of Oxime-Based Bioconjugations by Substituted Anilines. *Bioconjugate Chem.* **25**, 93-101, doi:10.1021/bc400380f (2014).

124   Saito, F., Noda, H. & Bode, J. W. Critical Evaluation and Rate Constants of Chemoselective Ligation Reactions for Stoichiometric Conjugations in Water. *ACS Chem. Biol.* **10**, 1026-1033, doi:10.1021/cb5006728 (2015).

125   Sasaki, T., Kodama, K., Suzuki, H., Fukuzawa, S. & Tachibana, K. N-terminal labeling of proteins by the Pictet–Spengler reaction. *Bioorg. Med. Chem. Lett.* **18**, 4550-4553, doi:http://dx.doi.org/10.1016/j.bmcl.2008.07.033 (2008).

126   Ning, X. *et al.* Protein Modification by Strain-Promoted Alkyne–Nitrone Cycloaddition. *Angew. Chem. Int. Ed.* **49**, 3065-3068, doi:10.1002/anie.201000408 (2010).

127   McKay, C. S., Chigrinova, M., Blake, J. A. & Pezacki, J. P. Kinetics studies of rapid strain-promoted [3 + 2]-cycloadditions of nitrones with biaryl-aza-cyclooctynone. *Org. Biomol. Chem.* **10**, 3066-3070, doi:10.1039/C2OB07165G (2012).

128   Alam, J., Keller, T. H. & Loh, T.-P. Functionalization of Peptides and Proteins by Mukaiyama Aldol Reaction. *J. Am. Chem. Soc.* **132**, 9546-9548, doi:10.1021/ja102733a (2010).

129   Alam, J., Keller, T. H. & Loh, T.-P. Indium mediated allylation in peptide and protein functionalization. *Chem. Commun.* **47**, 9066-9068, doi:10.1039/C1CC12926K (2011).

130   Han, M.-J., Xiong, D.-C. & Ye, X.-S. Enabling Wittig reaction on site-specific protein modification. *Chem. Commun.* **48**, 11079-11081, doi:10.1039/C2CC35738K (2012).

131   Triana, V. & Derda, R. Tandem Wittig/Diels-Alder diversification of genetically encoded peptide libraries. *Org. Biomol. Chem.* **15**, 7869-7877, doi:10.1039/C7OB01635B (2017).

132   Kitov, P. I., Vinals, D. F., Ng, S., Tjhung, K. F. & Derda, R. Rapid, Hydrolytically Stable Modification of Aldehyde-Terminated Proteins and Phage Libraries. *J. Am. Chem. Soc.* **136**, 8149-8152, doi:10.1021/ja5023909 (2014).

133   Wang, P. *et al.* Site-Specific Chemical Modification of Peptide and Protein by Thiazolidinediones. *Org. Lett.* **17**, 1361-1364, doi:10.1021/acs.orglett.5b00005 (2015).

134   Kularatne, S. A. *et al.* A CXCR4-Targeted Site-Specific Antibody–Drug Conjugate. *Angew. Chem. Int. Ed.* **53**, 11863-11867, doi:10.1002/anie.201408103 (2014).

135   O'Shannessy, D. J., Dobersen, M. J. & Quarles, R. H. A novel procedure for labeling immunoglobulins by conjugation to oligosaccharide moieties. *Immunol. Lett.* **8**, 273-277, doi:http://dx.doi.org/10.1016/0165-2478(84)90008-7 (1984).

136   Wolfe, C. A. C. & Hage, D. S. Studies on the Rate and Control of Antibody Oxidation by Periodate. *Anal. Biochem.* **231**, 123-130, doi:http://dx.doi.org/10.1006/abio.1995.1511 (1995).

137   Thompson, P. *et al.* Hydrolytically Stable Site-Specific Conjugation at the N-Terminus of an Engineered Antibody. *Bioconjugate Chem.* **26**, 2085-2096, doi:10.1021/acs.bioconjchem.5b00355 (2015).

138   Hudak, J. E. *et al.* Synthesis of Heterobifunctional Protein Fusions Using Copper-Free Click Chemistry and the Aldehyde Tag. *Angew. Chem. Int. Ed.* **51**, 4161-4165, doi:10.1002/anie.201108130 (2012).

139   Agten, S. M., Dawson, P. E. & Hackeng, T. M. Oxime conjugation in protein chemistry: from carbonyl incorporation to nucleophilic catalysis. *J. Pept. Sci.* **22**, 271-279, doi:10.1002/psc.2874 (2016).

140   Kalia, J. & Raines, R. T. Hydrolytic Stability of Hydrazones and Oximes. *Angew. Chem. Int. Ed.* **47**, 7523-7526, doi:10.1002/anie.200802651 (2008).

141   Lang, K. & Chin, J. W. Bioorthogonal Reactions for Labeling Proteins. *ACS Chem. Biol.* **9**, 16-20, doi:10.1021/cb4009292 (2014).

142   Schmidt, P., Stress, C. & Gillingham, D. Boronic acids facilitate rapid oxime condensations at neutral pH. *Chem. Sci.* **6**, 3329-3333, doi:10.1039/C5SC00921A (2015).

143   Larsen, D., Pittelkow, M., Karmakar, S. & Kool, E. T. New Organocatalyst Scaffolds with High Activity in Promoting Hydrazone and Oxime Formation at Neutral pH. *Org. Lett.* **17**, 274-277, doi:10.1021/ol503372j (2015).

144   Nigst, T. A., Antipova, A. & Mayr, H. Nucleophilic Reactivities of Hydrazines and Amines: The Futile Search for the α-Effect in Hydrazine Reactivities. *J. Org. Chem* **77**, 8142-8155, doi:10.1021/jo301497g (2012).

145   Best, M. D. Click Chemistry and Bioorthogonal Reactions: Unprecedented Selectivity in the Labeling of Biological Molecules. *Biochemistry* **48**, 6571-6584, doi:10.1021/bi9007726 (2009).

146   Temming, R. P., Eggermont, L., van Eldijk, M. B., van Hest, J. C. M. & van Delft, F. L. N-terminal dual protein functionalization by strain-promoted alkyne-nitrone cycloaddition. *Org. Biomol. Chem.* **11**, 2772-2779, doi:10.1039/C3OB00043E (2013).

147   McKay, C. S., Blake, J. A., Cheng, J., Danielson, D. C. & Pezacki, J. P. Strain-promoted cycloadditions of cyclic nitrones with cyclooctynes for labeling human cancer cells. *Chem. Commun.* **47**, 10040-10042, doi:10.1039/C1CC13808A (2011).

148   MacKenzie, D. A., Sherratt, A. R., Chigrinova, M., Cheung, L. L. W. & Pezacki, J. P. Strain-promoted cycloadditions involving nitrones and alkynes — rapid tunable reactions for bioorthogonal labeling. *Curr. Opin. Chem. Biol.* **21**, 81-88, doi:http://dx.doi.org/10.1016/j.cbpa.2014.05.023 (2014).

149   Kitanosono, T. & Kobayashi, S. Mukaiyama Aldol Reactions in Aqueous Media. *Adv. Synth. Catal.* **355**, 3095-3118, doi:10.1002/adsc.201300798 (2013).

150   Wittig, G. & Schöllkopf, U. Über Triphenyl-phosphin-methylene als olefinbildende Reagenzien (I. Mitteil. *Chem. Ber.* **87**, 1318-1330, doi:10.1002/cber.19540870919 (1954).

151   Paladhi, S., Chauhan, A., Dhara, K., Tiwari, A. K. & Dash, J. An uncatalyzed aldol reaction of thiazolidinediones. *Green Chem.* **14**, 2990-2995, doi:10.1039/C2GC35819K (2012).

152   Crisalli, P. & Kool, E. T. Importance of ortho Proton Donors in Catalysis of Hydrazone Formation. *Org. Lett.* **15**, 1646-1649, doi:10.1021/ol400427x (2013).

153   Kudirka, R. A. *et al.* Site-Specific Tandem Knoevenagel Condensation–Michael Addition To Generate Antibody–Drug Conjugates. *ACS Medicinal Chemistry Letters* **7**, 994-998, doi:10.1021/acsmedchemlett.6b00253 (2016).

154   van Geel, R., Pruijn, G. J. M., van Delft, F. L. & Boelens, W. C. Preventing Thiol-Yne Addition Improves the Specificity of Strain-Promoted Azide–Alkyne Cycloaddition. *Bioconjugate Chem.* **23**, 392-398, doi:10.1021/bc200365k (2012).

155   Zaia, J., Annan, R. S. & Biemann, K. The correct molecular weight of myoglobin, a common calibrant for mass spectrometry. *Rapid Commun. Mass Spectrom.* **6**, 32-36, doi:10.1002/rcm.1290060108 (1992).

156   Merritt, E. A. & Hol, W. G. J. AB5 toxins. *Curr. Opin. Struct. Biol.* **5**, 165-171, doi:https://doi.org/10.1016/0959-440X(95)80071-9 (1995).

157   Zhang, R.-G. *et al.* The Three-dimensional Crystal Structure of Cholera Toxin. *J. Mol. Biol.* **251**, 563-573, doi:http://dx.doi.org/10.1006/jmbi.1995.0456 (1995).

158   Rose, K. *et al.* New Cyclization Reaction at the Amino Terminus of Peptides and Proteins. *Bioconjugate Chem.* **10**, 1038-1043, doi:10.1021/bc9900587 (1999).

159   Holmgren, A. Thioredoxin structure and mechanism: conformational changes on oxidation of the active-site sulfhydryls to a disulfide. *Structure* **3**, 239-243, doi:https://doi.org/10.1016/S0969-2126(01)00153-8 (1995).

160   Shimomura, O., Johnson, F. H. & Saiga, Y. Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea. *J. Cell. Comp. Physiol.* **59**, 223-239, doi:10.1002/jcp.1030590302 (1962).

161   Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. & Prasher, D. Green fluorescent protein as a marker for gene expression. *Science* **263**, 802-805, doi:10.1126/science.8303295 (1994).

162   Tsien, R. Y. The Green Fluorescent Protein. *Annu. Rev. Biochem.* **67**, 509-544, doi:10.1146/annurev.biochem.67.1.509 (1998).

163   Barondeau, D. P., Kassmann, C. J., Tainer, J. A. & Getzoff, E. D. Understanding GFP Chromophore Biosynthesis: Controlling Backbone Cyclization and Modifying Post-translational Chemistry. *Biochemistry* **44**, 1960-1970, doi:10.1021/bi0479205 (2005).

164   Gerdes, H.-H. & Kaether, C. Green fluorescent protein: applications in cell biology. *FEBS Lett.* **389**, 44-47, doi:https://doi.org/10.1016/0014-5793(96)00586-8 (1996).

165   Wang, S. & Hazelrigg, T. Implications for bcd mRNA localization from spatial distribution of exu protein in Drosophila oogenesis. *Nature* **369**, 400, doi:10.1038/369400a0 (1994).

166   Vartiainen, M. K., Guettler, S., Larijani, B. & Treisman, R. Nuclear Actin Regulates Dynamic Subcellular Localization and Activity of the SRF Cofactor MAL. *Science* **316**, 1749-1752, doi:10.1126/science.1141084 (2007).

167   Osamu, S. Discovery of Green Fluorescent Protein (GFP) (Nobel Lecture). *Angew. Chem. Int. Ed.* **48**, 5590-5602, doi:doi:10.1002/anie.200902240 (2009).

168   Martin, C. GFP: Lighting Up Life (Nobel Lecture). *Angew. Chem. Int. Ed.* **48**, 5603-5611, doi:doi:10.1002/anie.200902040 (2009).

169   Y., T. R. Constructing and Exploiting the Fluorescent Protein Paintbox (Nobel Lecture). *Angew. Chem. Int. Ed.* **48**, 5612-5626, doi:doi:10.1002/anie.200901916 (2009).

170   Knop, M., Lemnaru, R. & Seebeck, F. P. Mutation of Conserved Residues Increases in Vitro Activity of the Formylglycine-Generating Enzyme. *ChemBioChem* **18**, 1755-1761, doi:10.1002/cbic.201700174 (2017).

171   McKean, P. G., Denny, P. W., Knuepfer, E., Keen, J. K. & Smith, D. F. Phenotypic changes associated with deletion and overexpression of a stage-regulated gene family in Leishmania. *Cell. Microbiol.* **3**, 511-523, doi:doi:10.1046/j.1462-5822.2001.00135.x (2001).

172   Osman, M. *et al.* A third generation vaccine for human visceral leishmaniasis and post kala azar dermal leishmaniasis: First-in-human trial of ChAd63-KH. *PLoS Negl. Trop. Dis.* **11**, e0005527, doi:10.1371/journal.pntd.0005527 (2017).

173   MacLean, L. M. *et al.* Trafficking and release of Leishmania metacyclic HASPB on macrophage invasion. *Cell. Microbiol.* **14**, 740-761, doi:10.1111/j.1462-5822.2012.01756.x (2012).

174   Kumar, M. *et al.* Hydrophilic Acylated Surface Protein A (HASPA) of Leishmania donovani: Expression, Purification and Biophysico-Chemical Characterization. *Protein J.* **36**, 343-351, doi:10.1007/s10930-017-9726-x (2017).

175   Denny, P. W., Gokool, S., Russell, D. G., Field, M. C. & Smith, D. F. Acylation-dependent Protein Export inLeishmania. *J. Biol. Chem.* **275**, 11017-11025, doi:10.1074/jbc.275.15.11017 (2000).

176   Wright, M. H., Heal, W. P., Mann, D. J. & Tate, E. W. Protein myristoylation in health and disease. *J. Chem. Biol.* **3**, 19-35, doi:10.1007/s12154-009-0032-8 (2010).

177   Chiang, Y., Kresge, A. J., Walsh, P. A. & Yin, Y. Phenylacetaldehyde and its cis- and trans-enols and enolate ions. Determination of the cis : trans ratio under equilibrium and kinetic control. *J. Chem. Soc., Chem. Commun.*, 869-871, doi:10.1039/C39890000869 (1989).

178   Eder, U., Sauer, G. & Wiechert, R. New Type of Asymmetric Cyclization to Optically Active Steroid CD Partial Structures. *Angewandte Chemie International Edition in English* **10**, 496-497, doi:10.1002/anie.197104961 (1971).

179   Hajos, Z. G. & Parrish, D. R. Asymmetric synthesis of bicyclic intermediates of natural product chemistry. *J. Org. Chem* **39**, 1615-1621, doi:10.1021/jo00925a003 (1974).

180   List, B., Lerner, R. A. & Barbas, C. F. Proline-Catalyzed Direct Asymmetric Aldol Reactions. *J. Am. Chem. Soc.* **122**, 2395-2396, doi:10.1021/ja994280y (2000).

181   Wagner, J., Lerner, R. A. & Barbas, C. F. Efficient Aldolase Catalytic Antibodies That Use the Enamine Mechanism of Natural Enzymes. *Science* **270**, 1797-1800, doi:10.1126/science.270.5243.1797 (1995).

182   Sakthivel, K., Notz, W., Bui, T. & Barbas, C. F. Amino Acid Catalyzed Direct Asymmetric Aldol Reactions: A Bioorganic Approach to Catalytic Asymmetric Carbon−Carbon Bond-Forming Reactions. *J. Am. Chem. Soc.* **123**, 5260-5267, doi:10.1021/ja010037z (2001).

183   Northrup, A. B. & MacMillan, D. W. C. The First Direct and Enantioselective Cross-Aldol Reaction of Aldehydes. *J. Am. Chem. Soc.* **124**, 6798-6799, doi:10.1021/ja0262378 (2002).

184   Northrup, A. B., Mangion, I. K., Hettche, F. & MacMillan, D. W. C. Enantioselective Organocatalytic Direct Aldol Reactions of α-Oxyaldehydes: Step One in a Two-Step Synthesis of Carbohydrates. *Angew. Chem. Int. Ed.* **43**, 2152-2154, doi:10.1002/anie.200453716 (2004).

185   Northrup, A. B. & MacMillan, D. W. C. Two-Step Synthesis of Carbohydrates by Selective Aldol Reactions. *Science*, doi:10.1126/science.1101710 (2004).

186   Olsen, J. V., Ong, S.-E. & Mann, M. Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues. *Mol. Cell. Proteomics* **3**, 608-614, doi:10.1074/mcp.T400003-MCP200 (2004).

187   Chen, X. & Wu, Y.-W. Selective chemical labeling of proteins. *Org. Biomol. Chem.*, doi:10.1039/C6OB00126B (2016).

188   Dommerholt, J. *et al.* Readily Accessible Bicyclononynes for Bioorthogonal Labeling and Three-Dimensional Imaging of Living Cells. *Angew. Chem. Int. Ed.* **49**, 9422-9425, doi:10.1002/anie.201003761 (2010).

189   Torii, H., Nakadai, M., Ishihara, K., Saito, S. & Yamamoto, H. Asymmetric Direct Aldol Reaction Assisted by Water and a Proline-Derived Tetrazole

Catalyst. *Angew. Chem. Int. Ed.* **43**, 1983-1986, doi:10.1002/anie.200352724 (2004).

190   Hartikka, A. & Arvidsson, P. I. 5-(Pyrrolidine-2-yl)tetrazole: Rationale for the Increased Reactivity of the Tetrazole Analogue of Proline in Organocatalyzed Aldol Reactions. *Eur. J. Org. Chem.* **2005**, 4287-4295, doi:10.1002/ejoc.200500470 (2005).

191   Sánchez, D. *et al.* Relative Tendency of Carbonyl Compounds To Form Enamines. *Org. Lett.* **14**, 536-539, doi:10.1021/ol203157s (2012).

192   Thalidomide - American Chemical Society. *https://www.acs.org/content/acs/en/molecule-of-the-week/archive/t/thalidomide.html* (2014).

193   Dickerson, T. J. & Janda, K. D. Aqueous Aldol Catalysis by a Nicotine Metabolite. *J. Am. Chem. Soc.* **124**, 3220-3221, doi:10.1021/ja017774f (2002).

194   Hayashi, Y. In Water or in the Presence of Water? *Angew. Chem. Int. Ed.* **45**, 8103-8104, doi:10.1002/anie.200603378 (2006).

195   Brogan, A. P., Dickerson, T. J. & Janda, K. D. Enamine-Based Aldol Organocatalysis in Water: Are They Really "All Wet"? *Angew. Chem. Int. Ed.* **45**, 8100-8102, doi:10.1002/anie.200601392 (2006).

196   Jimeno, C. Water in asymmetric organocatalytic systems: a global perspective. *Org. Biomol. Chem.* **14**, 6147-6164, doi:10.1039/C6OB00783J (2016).

197   Maruani, A., Richards, D. A. & Chudasama, V. Dual modification of biomolecules. *Org. Biomol. Chem.* **14**, 6165-6178, doi:10.1039/C6OB01010E (2016).

198   Tang, Y. *et al.* Real-Time Analysis on Drug-Antibody Ratio of Antibody-Drug Conjugates for Synthesis, Process Optimization, and Quality Control. *Scientific Reports* **7**, 7763, doi:10.1038/s41598-017-08151-2 (2017).

199   Hamblett, K. J. *et al.* Effects of Drug Loading on the Antitumor Activity of a Monoclonal Antibody Drug Conjugate. *Clin. Cancer Res.* **10**, 7063-7070, doi:10.1158/1078-0432.ccr-04-0789 (2004).

200   Patterson, G. H., Knobel, S. M., Sharif, W. D., Kain, S. R. & Piston, D. W. Use of the green fluorescent protein and its mutants in quantitative

fluorescence microscopy. *Biophys. J.* **73**, 2782-2790, doi:https://doi.org/10.1016/S0006-3495(97)78307-3 (1997).

201   Jencks, W. P. Studies on the Mechanism of Oxime and Semicarbazone Formation1. *J. Am. Chem. Soc.* **81**, 475-481, doi:10.1021/ja01511a053 (1959).

202   Cordes, E. H. & Jencks, W. P. Nucleophilic Catalysis of Semicarbazone Formation by Anilines. *J. Am. Chem. Soc.* **84**, 826-831, doi:10.1021/ja00864a030 (1962).

203   Crugeiras, J., Rios, A., Riveiros, E. & Richard, J. P. Substituent Effects on the Thermodynamic Stability of Imines Formed from Glycine and Aromatic Aldehydes: Implications for the Catalytic Activity of Pyridoxal-5′-phosphate. *J. Am. Chem. Soc.* **131**, 15815-15824, doi:10.1021/ja906230n (2009).

204   Schlick, T. L., Ding, Z., Kovacs, E. W. & Francis, M. B. Dual-Surface Modification of the Tobacco Mosaic Virus. *J. Am. Chem. Soc.* **127**, 3718-3723, doi:10.1021/ja046239n (2005).

205   International Human Genome Sequencing, C. Initial sequencing and analysis of the human genome. *Nature* **409**, 860, doi:10.1038/35057062

https://www.nature.com/articles/35057062#supplementary-information (2001).

206   International Human Genome Sequencing, C. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931, doi:10.1038/nature03001

https://www.nature.com/articles/nature03001#supplementary-information (2004).

207   McKenna, S. Functional and structural studies of the hydrophilic acylated surface proteins from Leishmania donovani. (2015).

208   Huang, J.-X. *et al.* Synthesis and Fungicidal Activity of Macrolactams and Macrolactones with an Oxime Ether Side Chain. *J. Agric. Food Chem.* **55**, 10857-10863, doi:10.1021/jf072733+ (2007).

209   Lee, M. H. *et al.* Acid-triggered release of doxorubicin from a hydrazone-linked Gd3+-texaphyrin conjugate. *Chem. Commun.* **52**, 10551-10554, doi:10.1039/C6CC05673C (2016).

210   Zhao, R. Y. *et al.* Synthesis and Evaluation of Hydrophilic Linkers for Antibody–Maytansinoid Conjugates. *J. Med. Chem.* **54**, 3606-3623, doi:10.1021/jm2002958 (2011).

211   Webb, S. Pharma interest surges in antibody drug conjugates. *Nat. Biotechnol.* **29**, 297, doi:10.1038/nbt0411-297 (2011).

212   Dubowchik, G. M. *et al.* Cathepsin B-Labile Dipeptide Linkers for Lysosomal Release of Doxorubicin from Internalizing Immunoconjugates: Model Studies of Enzymatic Drug Release and Antigen-Specific In Vitro Anticancer Activity. *Bioconjugate Chem.* **13**, 855-869, doi:10.1021/bc025536j (2002).

213   Puckett, S. *et al.* Inactivation of Fructose-1,6-Bisphosphate Aldolase Prevents Optimal Co-catabolism of Glycolytic and Gluconeogenic Carbon Substrates in Mycobacterium tuberculosis. *PLoS Pathog.* **10**, e1004144, doi:10.1371/journal.ppat.1004144 (2014).

214   M., D. S., A., G. W. & Chi-Huey, W. Recent Advances in Aldolase-Catalyzed Asymmetric Synthesis. *Adv. Synth. Catal.* **349**, 1308-1320, doi:doi:10.1002/adsc.200700115 (2007).

215   Giger, L. *et al.* Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat. Chem. Biol.* **9**, 494, doi:10.1038/nchembio.1276

https://www.nature.com/articles/nchembio.1276#supplementary-information (2013).

216   Jiang, L. *et al.* De Novo Computational Design of Retro-Aldol Enzymes. *Science* **319**, 1387-1391, doi:10.1126/science.1152692 (2008).

217   Zeymer, C., Zschoche, R. & Hilvert, D. Optimization of Enzyme Mechanism along the Evolutionary Trajectory of a Computationally Designed (Retro-)Aldolase. *J. Am. Chem. Soc.* **139**, 12541-12549, doi:10.1021/jacs.7b05796 (2017).

218   Zhang, S.-L. & Yu, Z.-L. C–C Activation by Retro-Aldol Reaction of Two β-Hydroxy Carbonyl Compounds: Synergy with Pd-Catalyzed Cross-Coupling To Access Mono-α-arylated Ketones and Esters. *J. Org. Chem* **81**, 57-65, doi:10.1021/acs.joc.5b02098 (2016).

219   Zhang, S.-L. & Deng, Z.-Q. Copper-catalyzed retro-aldol reaction of [small beta]-hydroxy ketones or nitriles with aldehydes: chemo- and stereoselective access to (E)-enones and (E)-acrylonitriles. *Org. Biomol. Chem.* **14**, 7282-7294, doi:10.1039/C6OB01198E (2016).

220   Windle, C. L., Müller, M., Nelson, A. & Berry, A. Engineering aldolases as biocatalysts. *Curr. Opin. Chem. Biol.* **19**, 25-33, doi:https://doi.org/10.1016/j.cbpa.2013.12.010 (2014).

221   Karlstrom, A. *et al.* Using antibody catalysis to study the outcome of multiple evolutionary trials of a chemical task. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 3878-3883, doi:10.1073/pnas.97.8.3878 (2000).

222   Notz, W., Sakthivel, K., Bui, T., Zhong, G. & Barbas, C. F. Amine-catalyzed direct asymmetric Mannich-type reactions. *Tetrahedron Lett.* **42**, 199-201, doi:https://doi.org/10.1016/S0040-4039(00)01908-0 (2001).

223   Joshi, N. S., Whitaker, L. R. & Francis, M. B. A Three-Component Mannich-Type Reaction for Selective Tyrosine Bioconjugation. *J. Am. Chem. Soc.* **126**, 15942-15943, doi:10.1021/ja0439017 (2004).

224   Minakawa, M., Guo, H.-M. & Tanaka, F. Imines that React with Phenols in Water over a Wide pH Range. *J. Org. Chem* **73**, 8669-8672, doi:10.1021/jo8017389 (2008).

225   Saimoto, H. *et al.* Effect of Calcium Reagents on Aldol Reactions of Phenolic Enolates with Aldehydes in Alcohol. *J. Org. Chem* **61**, 6768-6769, doi:10.1021/jo961352k (1996).

226   H., K. Ueber Synthese der Salicylsäure. *Justus Liebigs Ann. Chem.* **113**, 125-127, doi:doi:10.1002/jlac.18601130120 (1860).

227   R., S. Beitrag zur Kenntniss der Kolbe'schen Salicylsäure Synthese. *Journal für Praktische Chemie* **31**, 397-411, doi:doi:10.1002/prac.18850310130 (1885).

228   Laemmli, U. K. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227**, 680-685 (1970).

229   Branson, T. R. The self-assembly of nanoarchitectures via protein-ligand interactions. (2013).

230   Qi, X. *et al.* A solid-phase approach to DDB derivatives. *Eur. J. Med. Chem.* **40**, 805-810, doi:http://dx.doi.org/10.1016/j.ejmech.2005.03.024 (2005).