An Umpolung approach for the site-selective and site-specific chemical modification of proteins

Lewis Michael Gooch

Doctor of Philosophy

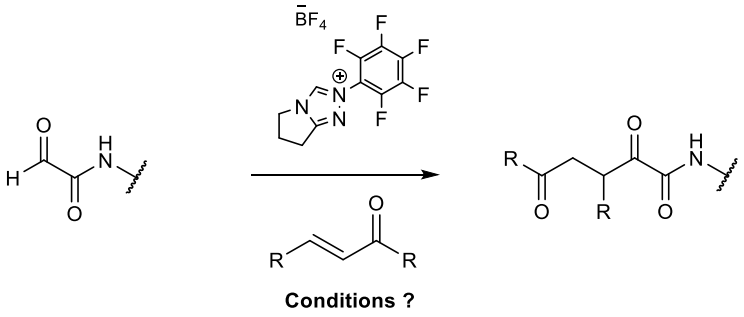
University of York

Chemistry

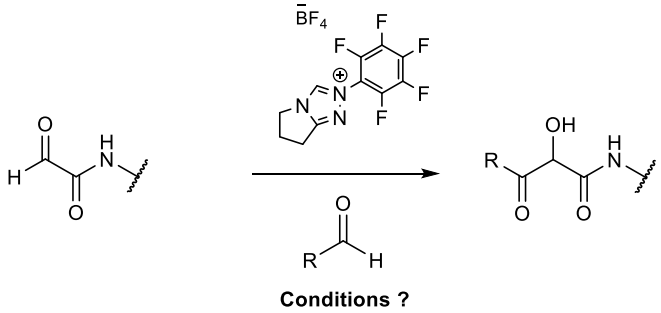
September 2020

Abstract

The rapid expansion of techniques for the chemical modification of proteins has caused a paradigm shift the field of chemical biology owing to their wide-ranging applications in the cell biology and pharmaceutical fields. Many bioconjugation strategies suffer from a lack selectivity, specificity, lack of non-physiological conditions and general vulnerability to hydrolytic cleavage. This thesis describes the design and development of a novel protein bioconjugation strategy, taking advantage of developments from the field of Umpolung chemistry on small molecules. This is achieved utilising proteins bearing previously installed α-oxo-aldehydes, this project developed and optimised the Umpolung modification yielding hydrolytically stable C-C bonds at neutral pH. Chapter 2 describes the investigation of a bio-compatible Stetter reaction, utilising Michael acceptors to generate 1,4 diketone products.



The information from this chapter was used to direct the project going forward, moving towards acyloin-like condensation utilising aliphatic aldehydes to generate 1, 2 diketones, described in Chapters 3 and 4, where several obstacles were encountered and overcome.



This Umpolung strategy was also demonstrated to be able to modify proteins at both *N*-terminal positons and at internal sites. This approach was found to be suited to α-methyl-aldehyde donors which was computationally investigated and discussed in chapter 5. Finally, applications and connectivity comments are addressed in chapter 6.

Contents

List of Tables

List of Figures

Figure 6.7: Schematic showing PES against reaction step for linear aldehyde homo-coupling (red), branched aldehyde homo-coupling (blue), branched aldehyde-dipeptide mimic coupling (green) and linear aldehyde-dipeptide mimic coupling (pink)..... 108

List of Accompanying Material

This thesis, as an electronic copy, forms part of a package which contains the following:

- Raw NMR data for compounds characterised by NMR described in this thesis
- Raw ESI-FTICR-MS data for all protein species described in this thesis
- Raw ESI-LC-MS data for all protein species described in this thesis
- Raw ESI-LC-MS data for trypsin digest experiments described in this thesis

Acknowledgments

I would like to thank Martin Fascione for his encouragement, attention to detail and direction. Without his guidance this research would have simply not been possible. I would also like to thank Professor Peter O'Brien for his academic and personal support throughout this journey. Additionally, I would also like to thank Will Unsworth for being particularly helpful during thesis advisory meetings during the PhD. My writing ability has somewhat improved owing to his detail-oriented feedback.

Next, I would like to thank the past and present members of the MAF group who have had to endure my presence since 2016 and the POB group for their upmost politeness during my group presentations, despite having limited interest. I would also like to thank the few friends I made during my time here for their support and much needed comic relief.

I would be remiss if I did not extend massive thanks to Ed Bergstrom for his mass spectrometry guidance and endless patience. I would also like to thank Karl for running my small molecule mass spectrometry experiments despite their crudeness and Heather Fish, the world's most militant, but proficient NMR technician.

I would also like to thank Dr Rossington and Dr Wilkinson for encouraging me to pursue a PhD. Despite my unlimited respect for you, I am not sure I can forgive you for this just yet.

I thank my friends and family for their support and patience during my PhD. I would also like to thank my wife, caretaker, handler, keeper and general watchperson Mariela for her unwavering support, the likes of which I have never known in my life. You are far too good to marry a ruffian like me.

Finally, I would like to thank my best friend Matt for incredible perspective, strength and resilience, and for showing me every day what a real man looks like. If I can be as half as strong, as loving, as caring, as honest and as selfless as you are daily, I will consider it an honour.

Authors declaration

I, Lewis Michael Gooch, declare that this thesis, to the best of my knowledge, is a presentation of original work and that I am the sole author. I declare that this work has not been previously 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, footnotes in Chapters 5 and 6. The work submitted for this thesis is my own. The contributions made by the other authors towards this work are explicitly stated below. I can confirm that appropriate credit has been given within this thesis where reference has been made to work carried out by others and, above all, I am grateful to those colleagues and friends listed below whose contributions have had a positive impact upon the direction of this thesis.

- Chapter 5: Trypsin digest experiments were performed by A. Dowle; GFP9y39) and GFP(N150) starter cells and ThzK provided by R. Brabham

- Chapter 6: Section 6.2: Ser-Ala dipeptide **84** were synthesised and purified by D. Budhadev, full details can be found in reference 66;

- Chapter 6: Section 6.3 Increased temperature NMR was performed by A. Heyam;

- Chapter 6: Section 6.4 computational PES study carried out by L. Murillo-Herrera

- Chapter 7: All small-molecule HRMS data were recorded by K. Heaton

Chapter 1: Introduction

1.1 Selective Protein Bioconjugation

1.1.1 Cysteine-carbon Formation

Early protein bioconjugation approaches primarily took advantage of simple second-order reactions that selectively target canonical amino acids exposed on a protein surface.[1-3] Early strategies commonly targeted cysteine and lysine residues.

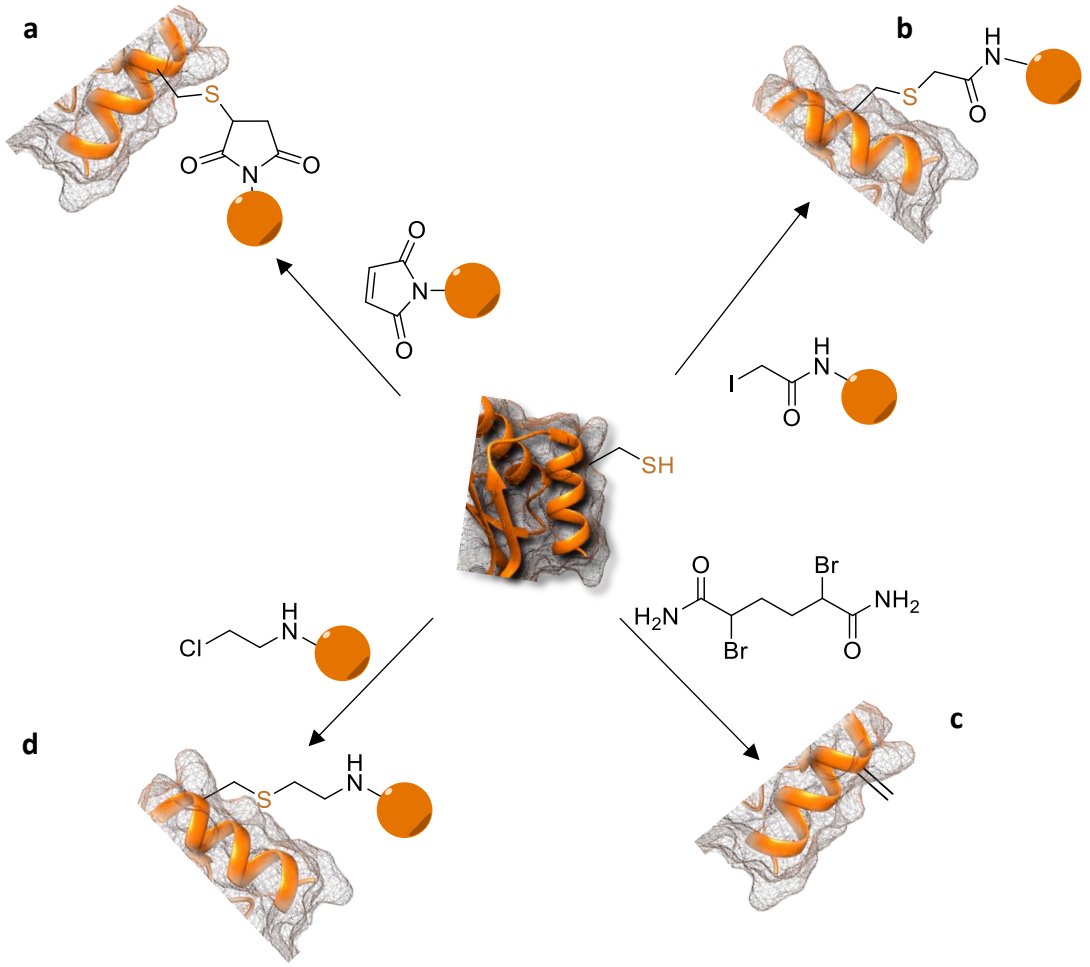
As one of the most nucleophilic (at pH 7) naturally occurring amino acids, the thiol group presented on cysteine was quick to be exploited for protein bioconjugation strategies. Owing to cysteine's naturally low abundance (<2%) within typical protein constructs, cysteine modification can often afford single-site selectivity.[4]

One of the most important methodologies developed is the selective reaction of cysteine with maleimides (Scheme 1.1, a). Widespread adoption of this reaction can be attributed to the ease of synthesis of the maleimide scaffold, and thier wide commercial availability.[5-6] This is highlighted by its use in vaccine[7] and antibody-drug conjugate (ADC) assembly.[8] Use of bromomaleimides by the Caddick group has allowed the *in vivo* study of reversible bioconjugation simultaneously allowing the formation of native disulfide bridges.[5, 9] Cysteine also selectively reacts with other electrophiles including α-halocarbonyls such as iodoacetamides (Scheme 1.1, b).[10]

Shokat and co-workers took inspiration from small molecule aminoethylation chemistry and extended it to histones (Scheme 1.1, c).[11] The thioether product generated termed 'thia-lysine', mimics typical lysine post-translational modifications (PTMs) found on histones. These were then studied to further decipher the role histones play in the packaging and ordering DNA.[11-12]

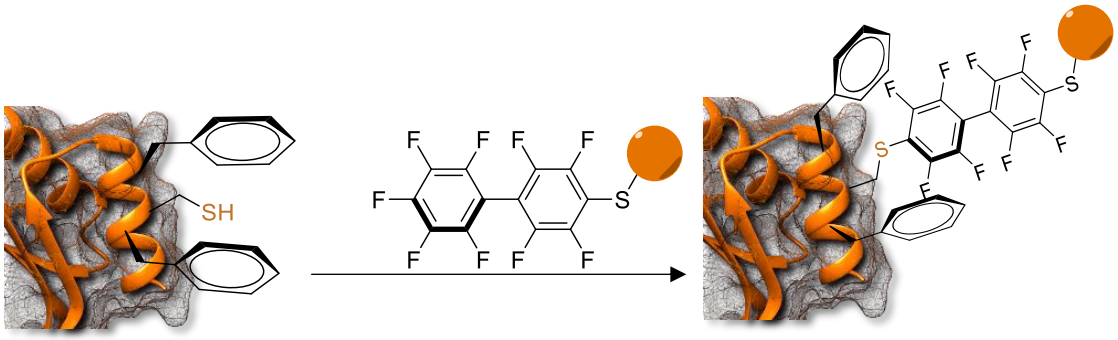
The 21st proteinogenic amino acid selenocysteine can also be employed for selective protein bioconjugation.[13] Selenium has higher nucleophilicity than the thiol group of cysteine. This allows for selective selenocysteine bioconjugation over traditional cysteine residues when reacted with maleimides.[13]

Davis and co-workers developed a two-step method for cysteine modification.[14] Firstly, cysteine is reduced to dehydroalanine (Dha) by treatment with *O*-mesitylenesulfonylhydroxylamine under basic conditions or a bisamide containing 1,4-dibromobutane reagent (Scheme 1.1, c). Subsequent Michael addition with thiol reagents can then yield a thioether bioconjugates (discussed further in chapter 1.1.7).[15]



Scheme 1.1 Outline of Cys-C forming protein bioconjugation strategies: **a**) malemide (**b**) iodoacetamides (**C**) Cys to Dha transformation (**d**) aminoethylation

Recently, Pentalute and co-workers further pushed the boundaries of cysteine modification with the design of a four-amino-acid sequence termed the 'π-clamp'.[16] This ingenuity overcomes selectivity in cysteine targeting, utilising only four-amino-acid sequence (Scheme 1.2) of FCPF serendipitously discovered *via* a library selection approach aimed to identify peptide sequences that promote arylation reactions in water. Based on computationally calculated peptide conformations and energy pathways, the reagents are hypothesized to be recognized by the phenylalanie residues, bringing the perflouroaromatic reagent into the vicinity of the activated cysteine residue.[16]



Scheme 1.2 π-clamp mediated Cys ligation

1.1.2 Cysteine-sulfur Formation

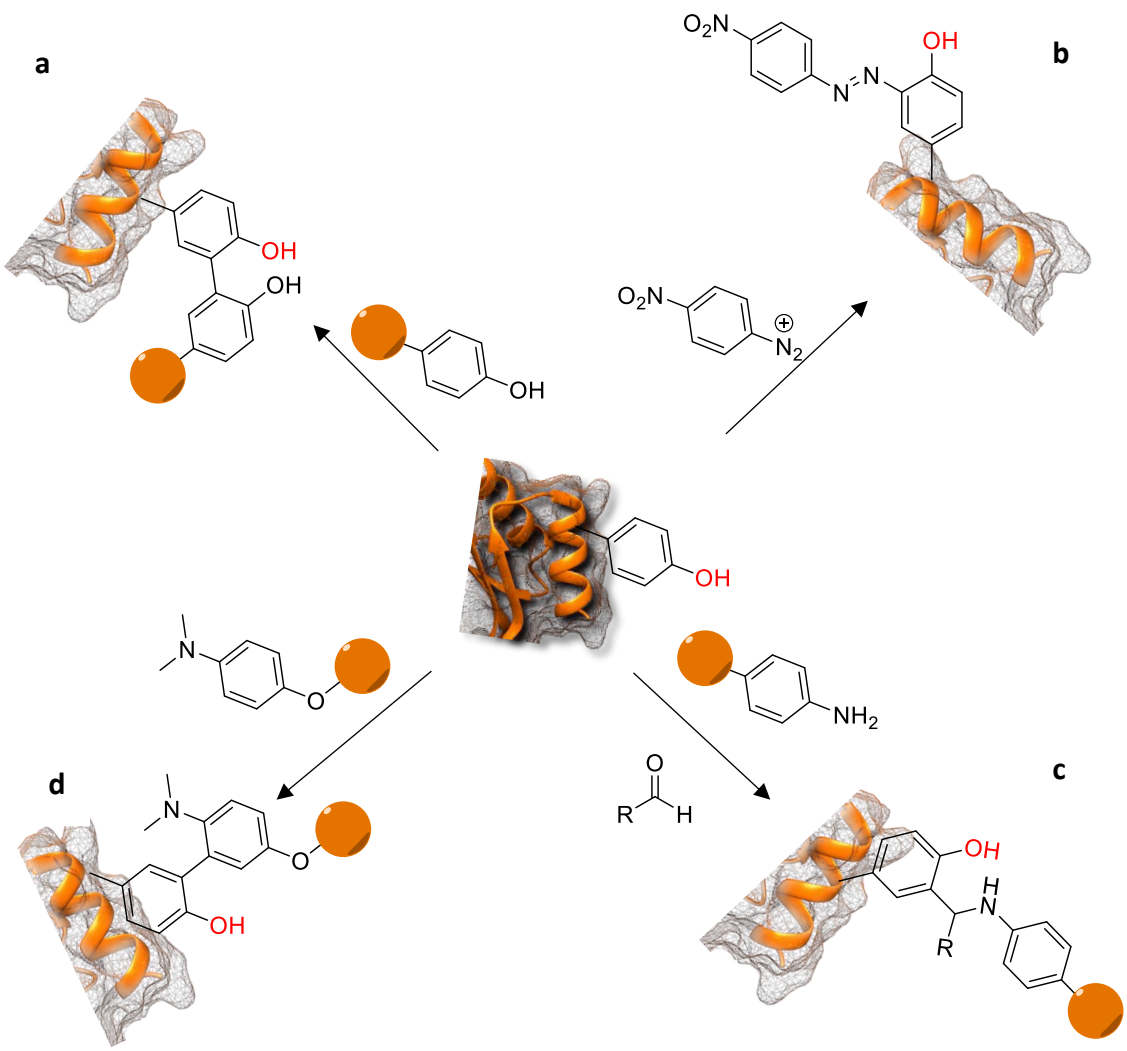
In nature, disulfide bonds are often pivotal in the formation and maintenance of protein tertiary and quaternary superstructure.[17] The propensity of sulfur atoms to change oxidation state is routinely exploited in chemical protein modification strategies, using external thiol bearing moieties.[4] The simplest method to promote disulfide formation between cysteines and thiol containing reporters is air oxidation.[18] Reaction rates for this are slow however[19] and this approach also carries the risk of catalysing protein aggregation.[20] This aggregation can be difficult to control and is dependent upon the accessibility of the desired cysteine residue. To counter this, the development of activated reagents has been key in the furtherment of exploiting disulfide bridges for selective chemical protein modification. The use of Ellman's reagent to quantify free thiols[21] is a classic example of the reliability of activated reagents, and is frequently employed for cysteine activation and subsequent protein modification.[22] This approach has successfully been employed in the development of ADCs,[23-24] enzyme immobilisation [25] and ubiquitination of histones.[26] Additionally, methanethiosulfonate and phenylthiosulfonate reagents are well-established thiol coupling reagents promoting disulfide bond formation. Employing these activated reagents has led to the selective modification of proteins,[27] modified enzymes[28] and glycoconjugates.[29-31]

1.1.3 Tyrosine and tryptophan

Although less common, novel methodologies have been developed for the site-selective modification of tyrosine and tryptophan.[32] These moieties can often be buried within protein 3D structure making surface exposed residues less common.[33]

Early examples of tyrosine modification involved the metal assisted oxidative coupling (Scheme 1.3, a) of two phenol groups. This method was first explored by Bonnafous and co-workers, who employed a nickel(II) catalyst and a magnesium monoperoxyphthalate co-oxidant to cross-link two proteins.[34]

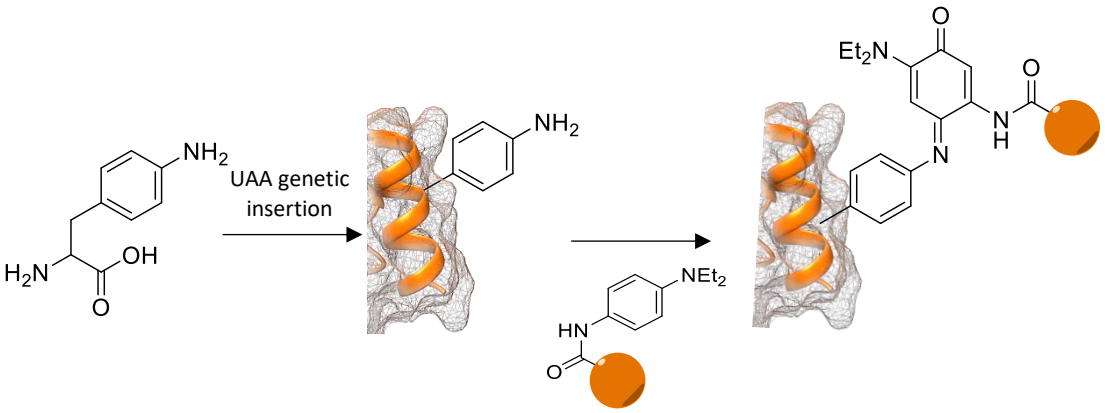
The rarity of tyrosine coupled with its distinctive phenolic hydroxyl group has led the Francis group developing a number of reactions to modify it, typically in the *ortho*-position. Examples include the application of diazonium salts (Scheme 1.3, b) with electron withdrawing *para* substituents to yield diazoarlation,[35-36] a three-component Mannich reaction with aldehydes and anilines (Scheme 1.3, c),[33, 37-39] and the coupling of dialkylanilines (Scheme 1.3, d) *via* cerium(IV)ammonium nitrate (CAN) oxidation.[40] An *ortho*-iodide on tyrosine can also be introduced through employment of Barluenga's reagent, although selectivity issues must be controlled by careful choice of stoichiometry and pH.[41] This, however, has not prevented its subsequent validation as a protein modification candidate. This methodology has been successfully employed in the modification of viral capsids,[41] pharmaceutical and imaging applications.[36]



Scheme 1.3 Outline of Tyr protein bioconjugation strategies: **a**) oxidative coupling (**b**) diazonium reagents (**C**) three way Mannich ligation (**d**) dialkylanilines reagents

Francis also utilised the aniline analogue of tyrosine to develop an oxidative coupling strategy with aryl-diamines (Scheme 1.4). To achieve this, the unnatural amino acid (UAA) *para*-aminophenylalanine was genetically encoded into the protein construct. This mutation generated the previously lacking site selectivity, even in the presence of native tyrosine residues. This methodology was also validated *via* site-selective bioconjugation on the surface of viral capsids.[42-43]

Additionally, the hydroxyl group on tyrosine can also be selectively targeted for modification. The Francis group employed palladium (II) acetate and triphenylphosphine trisulfonate (TPPTS) taking advantage of the formed palladium π-allyl reactivity to selectively alkylate the tyrosine hydroxyl moiety.[44]



Scheme 1.4 UAA insertion of aniline- Tyr analogue and subsequent ligation

In addition, Antos and Francis have developed a bioconjugation reaction selective for tryptophan residues. The authors employed a rhodium carbenoid, generated *in situ* from $[Rh_2(OAc)_4]$ and a diazo compound. The applications of this transformation are limited however owing to the necessity of acidic conditions (pH 2).[45] Improved methodology was developed, requiring less acidic conditions (pH 6) however conversion measured was reduced.[46]

1.1.4 Lysine and N terminal modifications

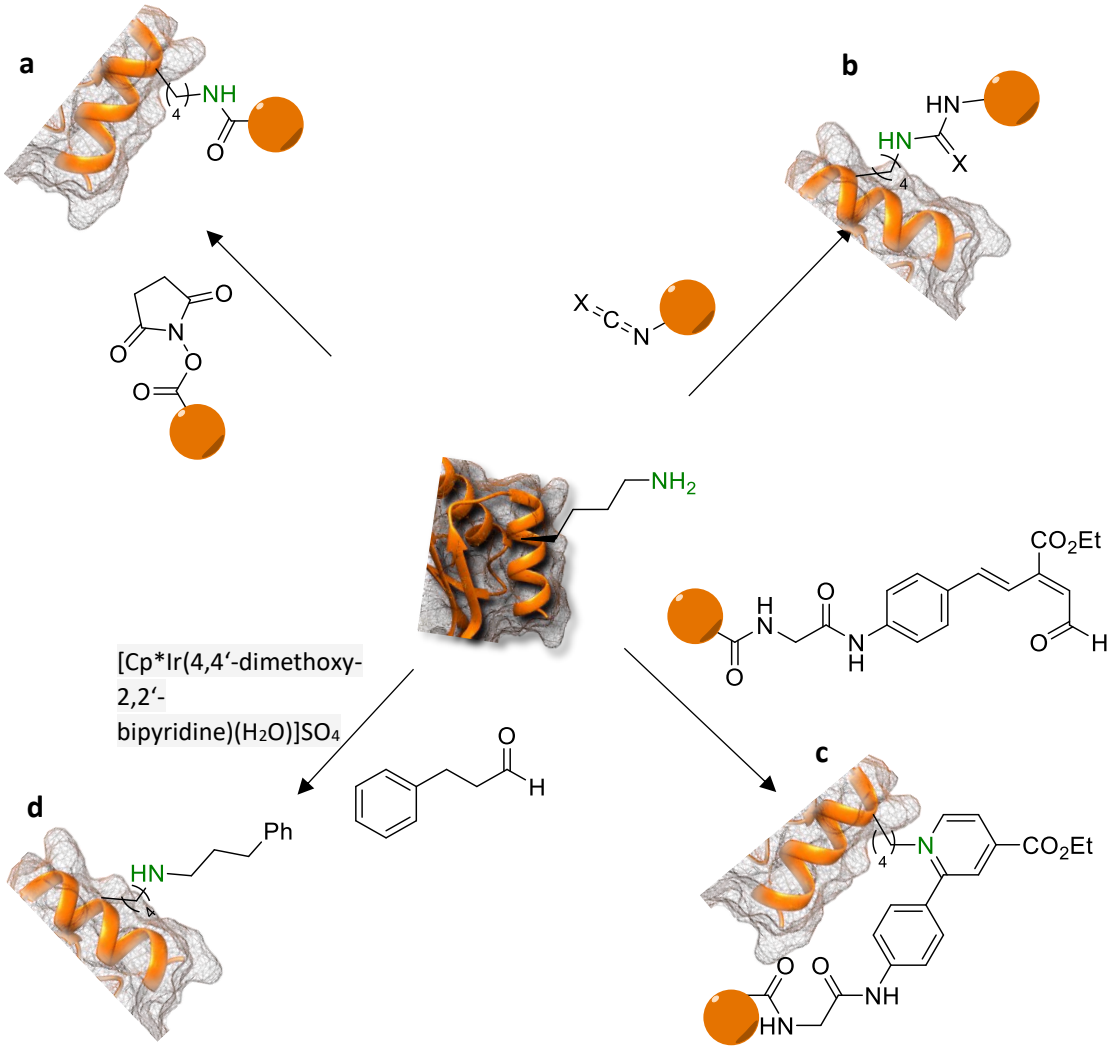
Although far more common than cysteine, lysine residues are also popular targets for bioconjugation, likely a result of the abundance of methods to selectively modify primary amines.[47] Lysine residues are especially favoured for targeted ligation in cases where site specificity is not essential or in cases where several ligations are required. This is indeed true for some vaccines[48] and ADCs, such as Kadcyla.[49]

Selective ligation of the ε-amine on lysine over the thiol group on cysteine can be achieved through the application of strong electrophiles such as sulfonyl chlorides,[6] isothiocyanates or isocyanates[50] and activated esters[51], generating sulfonamides[6], thioureas or ureas[50] and amides[51] respectively (Scheme 1.5, a-c). Notably, the most common approach of amino acid sequencing, Edman degradation, relies on *N*-terminal modification with phenyl isothiocyanate.[52]

Unsaturated aldehyde esters are also employed for ligation *via* lysine targeted protein modification. This is credited to their ability to perform 6π azaelectrocyclization.[53] This has been validated *via* ligation of a positron emitting metal binding ligand and used in conjunction with positron emission tomography (PET) to visualize glycoprotein processes.[54]

McFarland and Francis reported a lysine specific reductive alkylation reaction (Scheme 1.5, d) that proceeds through an iridium-catalyzed transfer hydrogenation.[55] Unlike the classic reaction based on sodium cyanoborohydride,[56] which requires acidic conditions, the iridium-mediated process proceeds in high yield at pH 7.4.[55]

Further methodology for the selective alkylation of lysine have also been developed. Cyclohexene sulfonamide reagents are reported to modify a specific lysine (Lys64) over all other endogenous lysine residues in human serum albumin (HSA).[57] Therefore, the specific domain, HSAdI, has been fused to protein termini as a reactive platform for Lys site-specific modification.[58]



Scheme 1.5 Outline of some chemical approaches to Lys protein bioconjugation strategies: **a**) activated ester (**b**) thio/isothioureas (**C**) 6π azaelectrocyclization (**d**) Ir- catalysed reductive alkylation.

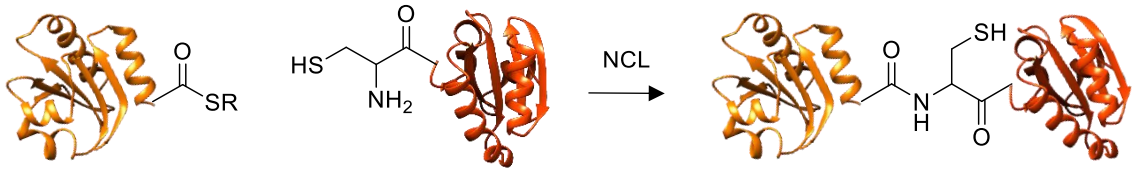
Similarly, the enzyme sortase A has been employed to mediate the formation of an isopeptide bond between a lysine residue in an inserted pilin domain with the threonine carboxyl group from an LPXTG (X = any amino acid) tag-containing substrate (discussed in 1.1.7).[59]

Although the ε-amine on lysine is typically the most nucleophilic amine in proteins, the *N*-terminus can also display unique pH-dependant reactivity.[60] The lower pK_a value of the N-terminus (~ 8) comparative to the ε-amine on lysine (~ 9 – 10) makes selective modification challenging albeit possible. Often this is achieved *via* biomimetic transamination reactions which introduce unique reactive handles. The transamination of proteins can be traced back to 1956, however harsh reaction conditions led to denaturation of proteins.[61] Following this Dixon and co-workers reported a room temperature transamination employing glyoxylate, catalytic base and copper(I).[62] Despite this, transamination of proteins did not become

popular until a biomimetic transamination was introduced by Frances and co-workers. This condensation of *N*-terminal amine employs pyridoxal-5-phosphate under physiological conditions, negating requirements for metal or base additives.[63] This can then be further modified through application of hydrazides,[64] aminooxy[65] reagents or organocatalysis[66] (Discussed further in 1.1.9).

Similarly, to the arene diazonium bioconjugation with Tyr described in the previous section,[36] Carreira and co-workers developed an amine-selective ligation method employing diazonium terephthalates. Product stability is generated through the irreversible formation of a stable triazin-4(3H)-one ring.[67] This methodology was used to nonregioselectively attack lysine residues on myoglobin and offsite attack on the *N*-terminus.

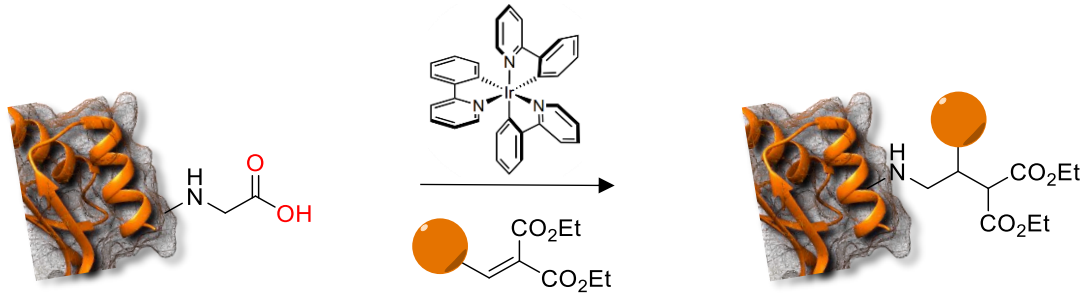
Native chemical ligation (NCL) is the cross coupling of a C terminal thioester with an *N* terminal cysteine.[68] This approach is a key method for the synthesis of polypeptides and proteins, further extending the range of biopolymers and proteins that can be accessed *via* total synthesis. Since its discloser by Kent and co-workers,[69-70] NCL has found numerous applications in protein bioconjugation,[71] polymer synthesis,[72] as well as nanotechnologies.[73-74] The reaction initiates with an intermolecular thiol-thioester exchange generating a transient transthioester. This is the chemo-selective capture step is followed by an intramolecular rearrangement *via* an S,N-acyl shift[75] yielding an amide bond (Scheme 1.6). Furthermore, the successes of NCL have been amplified by both expressed protein ligation (EPL) and protein-*trans*-splicing (PTS). Both EPL and PTS, the biological analogue of NCL, employ self-splicing proteins.[76] Protein self-splicing naturally occurs in an internal protein domain referred to as an intein. The intein is removed in a posttranslational process mechanistically similar to NCL. It is possible to generate recombinant polypeptide C-terminal thioesters. In turn, these can be further modified through employing NCL. PTS takes advantage of the ability to separate inteins into two peptides (Int_N and Int_C) which can produce an intein capable of splicing. PTS exploits these inteins capable of peptide splicing, extending NCL to *in vivo* capabilities and allowing the study of PPIs,[77] *in vivo* synthesis of cyclic peptides[78-79] and *in vivo* semi- synthesis of proteins.[80]



Scheme 1.6 Outline of NCL ligation

1.1.5 C terminal modification

A recent advance in protein modification was reported by MacMillan and co-workers focussing on carboxylic acid moieties at the C-terminus.[81] The authors report a modification approach harnessing visible light mediated single electron transfer (SET) to undergo decarboxylation. This methodology takes advantage of the difference in oxidation potential between C-terminal carboxylic acids and solvent exposed glutamic and aspartic acid residues. This oxidation potential difference allows for selectivity generating the C-terminal carbon centred radical and thus, offsite internal decarboxylation of glutamic and aspartic acid residues is not witnessed. This holds significant advantage over decarboxylative amide coupling and esterification strategies previously reported.[82-84] Following decarboxylation, subsequent treatment with a Michael acceptor yields alkylated product in moderate yields (41-49%) (Scheme 1.7). SET methodology was successfully employed to selectively alkylate human insulin. A significant drawback in this methodology arises however from the requirement of acidic media (pH 3.5), limiting its applications.



Scheme 1.7 Photoredox decarboxylative alkylation of C- terminal proteins

Despite the plethora of chemical approaches to selectively modify proteins, Nature offers several alternatives in the form of enzymes.[84] Typically, enzyme selectivity is achieved *via* recognition of specific amino acid sequences within a target proteins backbone. By employing enzymes, several limitations of the chemical modification of proteins, such as excesses of reagents[85] or harsh conditions[61] are surmounted. Some notable chemoenzymatic approaches to post-translational protein modification are described in this section.

1.1.6 Chemo-enzymatic strategies for protein modification

Ting and co-workers employed biotin ligase (BirA) expressed for from *E. coli*, to selectively biotinylate a lysine residue.[86] This was achieved by inserting the BirA recognition motif into the target protein. The recognition motif is a 15-residue 'acceptor peptide' (AP) sequence engineered into the target protein (Scheme 1.8, a). This allows for *in vivo* application as it will not biotinylate endogenous mammalian proteins.[87] Additionally, Ting and co-workers demonstrated that BirA can also be employed as a 'keto ligase', ligating a ketone containing biotin isotere.[88] This methodology was validated *via* the selective keto-biotin conjugation of epidermal growth factor receptors (EGFRs) expressed on the surface of HeLa cells.[86] This was then subjected to aminooxy or hydrazide reagents containing detectible chemical reporters. In spite of this, BirA has a narrow substrate spectrum limiting its application. This small substrate window however, is not a limitation of ligases from *P. horikoshii*. The wider substrate spectrum of this ligase allowed the Ting group to ligate azido- and alknylbiotin analogues to protein targets.[88] This enabled further modification of the target proteins *via* Staudinger ligation or copper catalysed azide-alkyne 1,3 dipolar cycloaddition (CuAAC) (discussed in 1.1.9).

Following this, Ting and co-workers developed an *E. coli* lipoic acid ligase (LplA) which conjugated lipoic acid analogues to LplA acceptor peptides.[89] The group successfully conjugated an array of alkynyl and azido probes to the AP (Scheme 1.8, b). Once ligated, these probes were further modified with detectible chemical reporters *via* CuAAC and strain promoted azide-alkyne 1,3 dipolar cycloaddition (SPAAC) (discussed in 1.1.9).[90] The Ting group would then develop a mutant LplA which ligated aryl azides into proteins. This was then used for photo cross-linking and subsequent study of protein-protein interactions (PPIs).[91] Inspired by this result, Ting and co-workers then developed the W371 LplA mutant,[92] successfully inserting aryl aldehydes into transmembrane protein neurexin-1β.[93] This aryl aldehyde was then subjected to hyrdazone and aminooxy reagents loaded with detectible chemical reporters. Building upon this methodology, McNaughton and co-workers used the W371 LplA mutant to selectively modify HER2 binding nanobodies.[94]

Ting and co-workers also developed the application of transglutaminases (TGases). These enzymes catalyse amide bond formation between glutamine and lysine residues.[95] Furthermore, the group discovered guinea pig liver transglutaminase (*gp*TGase) possessed wide scope of the amine containing substrate. Coupling this substrate scope with the introduction of a 'Q tag' sequence (PKPQQFM), proteins were successfully bioconjugated by

gpTGase (Scheme 1.8, c).[96] Detectible chemical reporters, including biotin and Alexa Flour568 were successfully ligated to green fluorescent protein (GFP). TGase was also employed to site-selectively PEGylate therapeutic human growth hormone (HGH), improving their half-lives.[97]

Prenylation is a posttranslational protein modification catalysed by a class of enzymes known as prenyltransferases (PFTases). Within this class of enzymes exists farnesyltransferase (FTase).[98] The enzyme identifies C-terminal CaaX motif in target proteins and transfers an isoprenoid group from farnesyl diphosphate (FPP), to the thiol group on cysteine *via* a thioether linkage (Scheme 1.8, d). The four-amino acid sequence is usually denoted as a CaaX, where C is cysteine, 'a' represents aliphatic amino acids (typically leucine, isoleucine or valine), and X representing methionine, serine, alanine or valine.[99] Research into FTase specificity found the enzyme is promiscuous towards both the CaaX recognition motif and the isoprenoid substrate.[100] Further development by Hougland and co-workers revealed that the recognition sequence can be extended to CaaaX, further extending the enzymes applications.[101] Owing to the promiscuity of this enzyme, a wide variety of isoprenoid analogues have been successfully introduced to various protein targets including fluorophores,[102-104] photoaffinity tags,[105-107] biotin,[108] azide,[109-112] alkyne[113-115] and aldehydes.[116-119] Through subsequent downstream modification of these groups, target proteins have been employed for a variety of applications including; *in vivo* protein imaging,[119] proteomic analysis,[108, 115] surface immobilisation[111] and *in vivo* efficacy and pharmacokinetics of protein conjugates for targeted therapy. [116]

Sortase A (SrtA) is a Ca^{2+}-dependant transpeptidase from *Staphylococcus aureus* or *Staphylococcus pyogenes*. These enzymes are used to covalently attach proteins to the bacteria cell wall.[59] SrtA generates an amide bond between the ε-amine on lysine and the carboxyl moiety on threonine located within an LPXTG amino acid sequence. In this sequence, commonly called a 'sortag', 'X' can be any amino acid.[120] SrtA initially recognises the LPXTG sequence and cleaves the threonine-glycine bond generating an acyl-enzyme intermediate. This acyl-enzyme intermediate is subsequently attacked by the free amine of an oligoglycine-terminating peptide generating an amide bond (Scheme 1.8, e). SrtA has been employed extensively for a wide variety site specific posttranslational modification of proteins.[121] Various methodology has been published for the application of SrtA to label the; *N*-terminus,[122] C-terminus[123] as well as internal protein loops[123]. SrtA mutants are common, including a penta-mutant with improved catalytic efficiency,[124] a Ca^{2+}-independent SrtA mutant[125] as well as SrtA homologues from alternative bacterial sources with differing

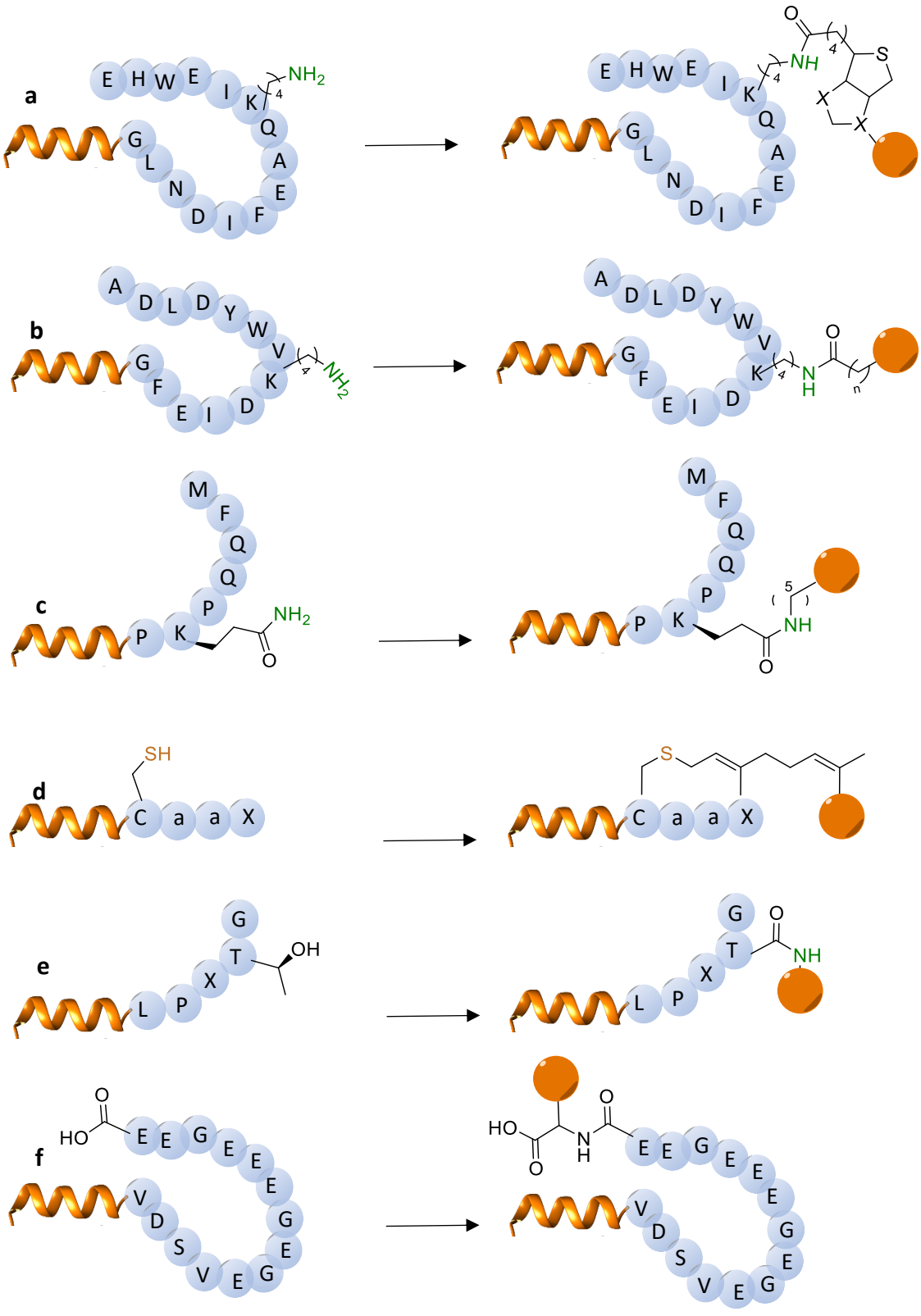
recognition sequences have been reported to successfully modify proteins.[126-129] Initially SrtA was employed for lipid modification,[130] cyclization,[126] and cell surface labelling.[131] More recent applications of SrtA, include the semi-synthesis of proteins with post-translational modifications,[132-133] protein surface immobilization,[134-138] liposome labelling,[139-140] virus-like particle labelling,[141-142] hydrogel formation,[143-144] cell surface labelling and *in vivo* protein labelling.[145-147]

Tubulin tyrosine ligase (TTL) is an ATP dependant enzyme which catalyses the reversible addition of a tyrosine to the C terminus of α -tubulin *via* an amide bond.[148] Once discovered, it was soon found TTL labelling is not limited solely to native tyrosine but can be extended to tyrosine analogues. The promiscuity of TTL ushered in a plethora of various novel biotechnological applications of TTL. The first tyrosine analogue reported to be accepted by TTL was 3-fluorotyrosine.[149] This fluorine ligation enabled the use of 19F NMR to visualise PPIs occurring near the C-terminus of α -tubulin. Acre and co-workers also reported that TTL was able to reversibly ligate 3-nitrotyrosine into tubulin.[150] Inspired by this, Bane and co-workers were able to widen the scope of tyrosine analogues TTL could ligate for *in vivo* application.[151] The Bane group would then go on to demonstrate TTLs ability to ligate 3-azidotyrosine and 3-formyltyrosine (3ForTyr). The azido tyrosine analogue however has limited application, owing to the fact that further downstream functionalisation of azides typically requires use of a copper catalyst. Copper can interfere with proper tubulin assembly, meaning that initial efforts focussed on the 3-formyltyrosine analogue.[151] Hackenberger and co-workers were able to demonstrate that TTL could be sequence specific, targeting a 14 -amino acid, VDSVEGEGEEEGEE recognition sequence, termed 'Tub tag'.[152] This peptide sequence mimics the C-terminal sequence of α -tubulin. Incorporation of Tub tag further extended the enzymes applicability, enabling the targeting of alternative proteins outside of tubulin. The group validated the Tub tag methodology through site specific ligation of azido, nitro, amino and formyl -tyrosine analogues into GFP, ubiquitin and nanobodies (Scheme 1.8, f).[152] Very recently, the Hackenberger group reported a tub tag mediated protein-protein ligation (TuPPL) at the C-terminus of GFP-specific nanobodies.[153] This was achieved through the convergent synthesis of two complimentary click chemistry handles incorporated *via* TTL ligation.

Formylglycine generating enzymes (FGE) are a family of oxidorectases which catalyse the post-translational modification of thiol group on cysteine on sulfatases to formylglycine.[154] The enzyme achieves this *via* oxidation followed by hydrolysis of cysteine located within a 5-amino acid recognition sequence CxPxR (x = any amino acid). Once identified, Bertozzi and

co-workers employed FGE for site specific protein ligation (Scheme 1.8, g).[155] They introduced two types of aldehyde tags site specifically into three different proteins. The first tag was a 13-amino acid sulfatase sequence and the second tag a shorter 6-amino acid motif. They reported that both the shorter and longer peptide tags were effective and conversion occurred both *N*- and C-terminus. Subsequent downstream modification was performed by the employment of aminooxy functionalised reagents including fluorophores, PEG and an affinity tag.[155] Bertozzi and co-workers then explored the protein ligation at internal positions. The CxPxR motif was selected as it minimized protein disruption and generated the highest yields (76%).[156] This was quickly followed by a report of *in vivo* CHO cell conjugation measured using aminooxy-Alexa Fluor 488.[157] Once optimised, *O*-benzylhydroxylamine was employed generating quantitative oxime conjugation.

Despite these advancements, the resulting C=N bond generated through oxime and hydrazine chemistries favours acidic media, and the bond highly susceptible to hydrolysis. This limits the application of this methodology. To circumvent this, Bertozzi and co-workers developed a modified Pictet-Spenglar ligation,[158] followed by a report by the Rabuka group which introduced a hydrazino-iso-Pictet-Spenglar (HIPS) ligation (discussed in 1.1.9).[159] By combining FGE modification and a subsequent HIPS reaction, drug-to-antibody ratios (DAR) was studied *in vivo* revealing efficacy and pharmacokinetics of ADCs.[160] Following this, the Rabuka group reported using FGE recognition tags on antibodies followed by tandem Knoevenagel condensation-Michael addition (discussed in 1.1.7).[161] The group reports a pyrazolone containing maytansine which initially reacts with the FGE generated aldehyde to generate a reactive enone. This Michael acceptor is then subsequently attacked by a second molecule of the starting pyrazolone reagent. As a testament to the power of FGE protein ligation, FGE-mediated protein modification has recently been developed as a commercial platform (SMARTag™) by Redwood Bioscience/Catalent for the synthesis of site-specific ADCs.[162]


a
E H W E I K Q A E F I D N L G
NH2
4
O
NH
S
X
X
b
A D L D Y W V K E I D F G
NH2
O
N
H
n
c
M F Q Q P K P
NH2
O
5
NH
d
SH
C a a X
S
e
G T X L P
OH
NH
f
HO
O
E E G E E E G E G V D S V E G
HO
N
H

Scheme 1.8 Outline of chemoenzymatic bioconjugation strategies: **a**) BirA ligase (**b**)LplA ligase (**c**) gpTgases (**d**) FTase ligation (**e**) SrtA (**f**) TTL ligation (**g**) FGE

1.1.7 Noncanonical amino acid incorporation

The bioconjugation potential of target proteins is typically restricted by the limited functionalities exposed by the 20 canonical amino acid residues. This is further complicated owing to the fact that multiple residues are present in a protein. This can make site- selective and site-specific ligation problematic. Therefore, there has been an increased focus on manipulating the genetic machinery of cells to incorporate noncanonical or unnatural amino acids (UAAs). These methods typically rely on the reassignment of codons to a given UAA. Early methods for UAA incorporation utilize auxotrophic strains for a particular amino acid. Auxotrophic strains are unable biosynthesize the amino acid and instead require uptake from the growth media.[163] tRNA synthetases can then be exploited to insert an UAA instead of the natural amino acid. This has most commonly been employed to introduce azide- and alkyne-containing amino acids into methionine auxotrophic *E. coli* strains.[164] Despite a wide variety of UAAs successfully incorporated *via* this methodology, the reassignment of sense codons results in global incorporation of the UAA, limiting uses *in vivo* and sometimes leading to cellular toxicity at higher concentrations.

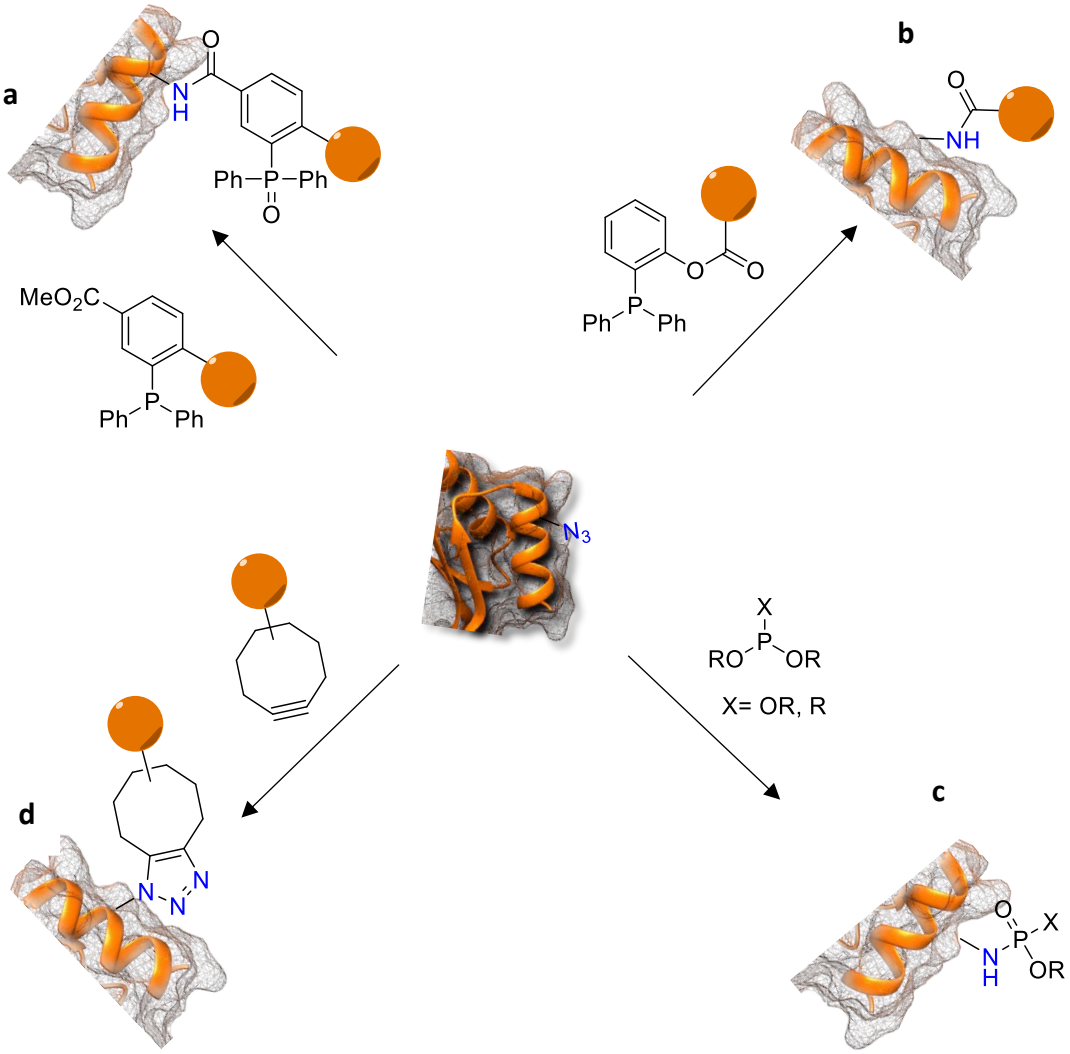
To circumvent the global incorporation that methionine auxotrophic *E. coli* strains generate, an alternative approach stems from reassignment of stop (nonsense) codons, such as the amber codon UAG.[165] The strategy employs orthogonal suppressor tRNA/tRNA synthetase (tRNA/aaRS) pair, which together insert the desired UAA to a tRNA specific for the codon, whilst not impeding the native cellular tRNA machinery. This has been achieved by transferring a tRNA/aaRS from another kingdom into the organism of interest. This has been successfully achieved in both prokaryotic[165] and eukaryotic cells.[166] To achieve this, typically one of two systems is used: tyrosine tRNA/aaRS from the archaebacteria *Methanococcus jannaschii* and the pyrrolysine tRNA/ aaRS of *Methanosarcina barkeri/mazei*.[167] These pairs have been successfully employed to introduce a diverse range of over 150 UAAs into protein constructs. These UAAs possess a variety of structures and key reactive handles necessary for further downstream modification.[168] This technique allows for incredible site- specific and

site- selective incorporation of a UAAs at a single point within a protein expressed *in vivo*. Further advances in the field include amber suppression in living animals[169-170] and increases in suppression efficiency by knockout of the gene for RF1 in *E. coli*.[171] More recently, focus has shifted towards the insertion of photo-uncaged UAAs, negating the need for lengthy, chemical downstream modification.[172-173] Despite this, limitations of this methodology remain. The most important limitation to be overcome in this strategy is the need for more widespread commercial access to the required plasmids, the global applicability towards all protein systems and expanding the range of target codons.[174]

1.1.8 Downstream functionalisation

Often it is not possible to site-selectively modify a protein without activation or pre-treatment *via* insertion of a reactive functional group or chemical handle. A class of bioorthogonal methodology has been developed strictly for this role, promoted by the wide variety of methods to introduce UAAs or unique reactive handles outlined above. The development of novel bioorthogonal reactions as well as UAA/handle insertion is key to surmounting the complimentary limitations of each methodology.[47] Despite requiring two (or more) synthetic steps, downstream functionalisation has played an important role in the development of protein modification as a field. Below is a summary of some key bioorthogonal reactions developed.

The rareness and bioorthogonality of azides in biological systems has led to a significant body of research developed for their use in chemical protein modification. The biologically unique 1, 3 -dipole handle has become routine choice for protein ligation strategies.[175] Early protein ligation utilised azide-functionalized proteins coupled with a classic Staudinger reaction to install amide linked detectible chemical reporters (Scheme 1.9, a).[176] In this approach, the phosphine oxide generated remains incorporated into the protein construct.[177] A significant development to surmount this drawback by Bertozzi and co-workers is the “traceless” Staudinger reaction (Scheme 1.9, b).[178] Other variants developed by Hackenberger and co-workers expanded this type of chemistry to alternative phosphorus functionalities, developing the Staudinger-phosphite[179] and Staudinger-phosphonite[180] ligation protocols generating phosphoramidate or phosphonamidate linkages, respectively (Scheme 1.9, c). These linkages mimic biological phosphoric esters and can ligate one or two chemical reporters into the same protein. This methodology was validated *via* site-selective PEGylation of calcium-binding messenger protein, calmodulin.



Scheme 1.9 Outline of azido bioconjugation strategies: **a**) Staudinger ligation (**b**) Traceless Staudinger ligation (**c**) Phosphite/phosphonate Staudinger ligation (**d**) Strained alkyne cycloaddition

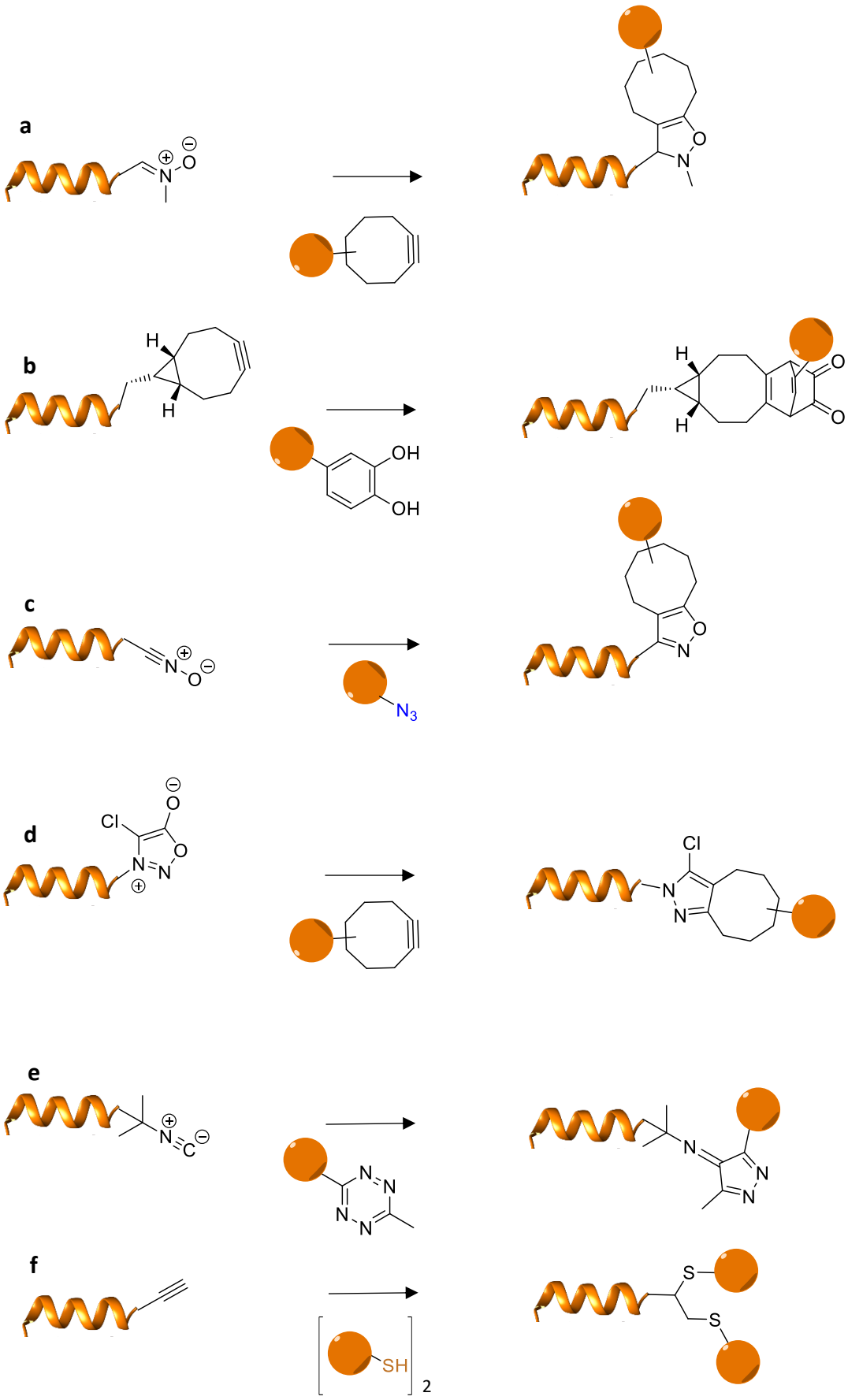
A further extension of azide bioorthogonal ligation is [3 + 2] cycloadditions with alkynes.[181] This is typically referred to as "click chemistry", a term coined by Sharpless and co-workers describing high yielding, highly selective ligations which produce by-products that require no chromatography to remove.[182] Early approaches utilised a modified Huisgen cycloaddition[183] commonly referred to as the CuAAC.[184] Though not a true cycloaddition (proceeds *via*

metallocyclic intermediate)[185] the CuAAC reaction is often regarded as one of the most important developments in the bioconjugation field. Catalytic copper(I) promotes a mild, chemoselective [3 + 2] cycloaddition in aqueous media generating hydrolytically stable 1,4-disubstituted triazole rings. CuAAC methodology has been employed to access and study a wide range of PTMs,[186] as well as site-specific immobilization strategies.[187] Additionally, CuAAC has also been employed to introduce ^{19}F- and ^{18}F- radio-labelled glycosyl reporters on the globular TIM barrel protein S*s*βG (βgalactosidase) and on a bacteriophage Qβ.[188] Despite the success of CuAAC methodology and the continued development of ligands and complexes (e.g., *N*-heterocyclic carbene−Cu catalysts),[189] limitations remain. These include protein damage and associated toxicity rendering *in vivo* application problematic. Toxicity is commonly associated with the reactive oxygen species (ROS) generated, leading to exploration of milder Cu-free protocols.[190]

The Bertozzi group removed copper from the ligation entirely, initially reporting a strain-promoted azide-alkyne cycloaddition (SPAAC) (Scheme 1.9, d).[191] Inspired by 1960s reports by Wittig and Krebs,[192] they reported that highly strained cyclooctynes reacted rapidly at room temperature with azide bearing glycoproteins. In true bioorthogonal fashion, the procedure employs no exogenous ligands or catalysts. This ligation strategy was further validated on the surface of mammalian cells, which also reported no toxicity.[191] To improve the reaction rate of SPAAC, both difluorinated cyclooctynes (DIFO)[193-194] and dibenzocylooctynes[195] were independently reported, allowing the *in vivo* visualisation of dynamic processes. Notably, the Bertozzi group employed DIFOs to visualize the development of membrane-associated glycans in developing Zebrafish embryos.[196] In striving for even faster reaction rates, the Bertozzi group reports utilising biarylazacyclooctynones[197] and cyclopropyl-fused bicyclononynes.[198] Despite removing copper and therefore reactive oxygen species from this approach, limitations remain however, such as the challenging synthesis, handling and storage of unstable compounds. More concerning, a degree of incompatibility towards cysteine has been reported.[199]

Alternatives to azides such as nitrones and nitrile oxides have also become increasingly popular as unique 1, 3-dipole chemical handles for stain-promoted cycloaddition. van Delft, Boons, and co-workers report a strain-promoted alkyne-nitrone cycloaddition (SPANC) in a one-pot, three step ligation strategy (Scheme 1.10, a).[200] In the SPANC approach, a terminal α-oxo-aldehyde handle is inserted *via* serine oxidation with periodate, and subsequent transformation into a nitrone using *N*-methylhydroxylamine. Finally, the nitrone is subjected to a cyclooctyne, completing the reaction. This efficient site-selective ligation is rapid

(typically 10 times faster than SPAAC), and its robustness has been demonstrated on the modification of the chemokine interleukin-8 (IL-8) as well as site-selective conjugation of antibodies to nanoparticles.[201] Upon further exploration on SPAAC- like ligations, the van Delft group reported a strain-promoted oxidation-controlled cyclooctyne-1,2-quinone cycloaddition (Scheme 1.10, b) (SPOCQ).[202] Firstly, 1,2-catechols are oxidized *via* periodate oxidation to 1,2-quinones, these quinones undergo SPOCQ cycloaddition with a protein-BCN construct, inserted *via* amber stop codon expression. This was initially validated on GFP.[202] A further variation of the SPANC ligation employs a protein bearing nitrile oxides as reactive handles to facilitate strain-promoted alkyne−nitrile oxide cycloaddition (SPANOC) (Scheme 1.10, c).[203-204] The SPANOC ligation follows a mechanistically similar pathway to the SPANC ligation (aldehyde insertion- oxime formation- oxidation-cycloaddition). This methodology was successfully employed to site- selectively modify BSA and fetuin. Another novel ligation added to the cycloaddition toolbox is the fast copper-free strain-promoted sydnone−alkyne cycloaddition (SPSAC).[205] This ligation occurs between 4-halosydnones and bicyclo- [6.1.0]-nonyne (BCN). Initially, activated 4-chlorosydnones are inserted into bovine serum albumin (BSA) protein *via* peptide coupling chemistries and purified. Following this, the sydnone containing protein was exposed to BCN- fluorescent tag in aqueous media with 10% DMSO co-solvent (Scheme 1.10, d).[205] Finally, isonitriles can react *via* [4 + 1] cycloaddition with tetrazines in aqueous media (Scheme 1.10, e). This has been developed by Leeper and co-workers who validated this approach by ligating a fluorophore to the membrane trafficking protein, synaptotagmin-I.[206]



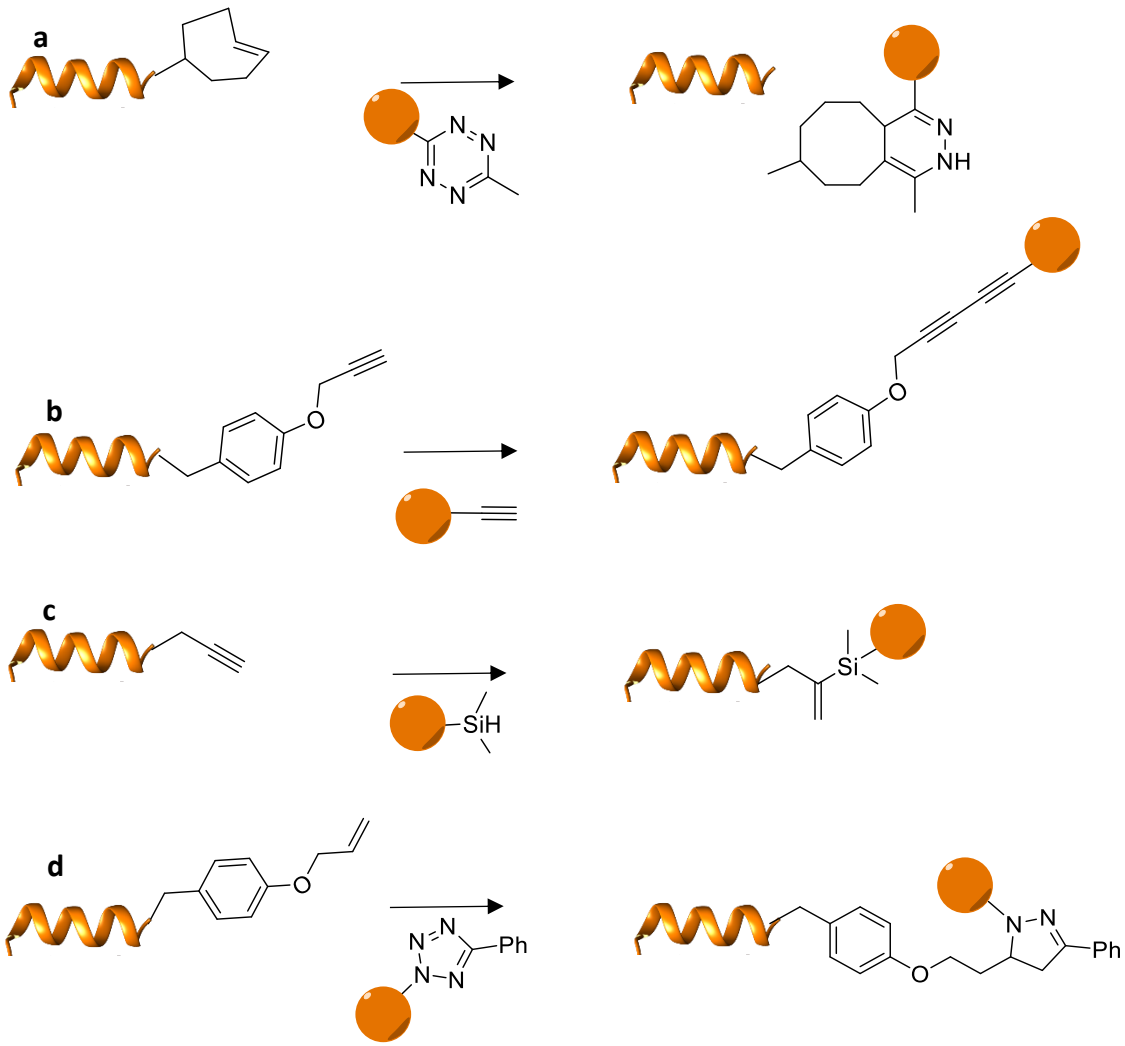
Scheme 1.10 Outline of alternate azide bioconjugation strategies: **a**) Nitrone ligation (**b**) SPOCQ ligation (**c**) Nitrile oxide ligation (**d**) Sydnone addition (**e**) Isonitrile ligations (**f**) Thiol-yne couplings

Alkyne and alkene UAAs can also be selectively incorporated into proteins and exploited in a wide range of bioorthogonal reactions.[207] Relatively simple alkynes such as alkynyl-pyrrolysine analogues can be engineered into protein sequences and transformed without exogenous reagents, using a site-selective thiol-yne coupling (Scheme 1.10, f).[208] Alternatively, cyclooctyne containing UAAS can be introduced and ligated using azides or tetrazine reagents mentioned previously. Katsonis and co-workers report the ligation of alkyne functionalised cowpea chlorotic mottle virus capsids *via* azide-alkyne click chemistry.[209]

Proteins bearing oxanorbornadienes[210] and norbornenes can react with azides and sulfonyl azides respectively, the latter validated on human carbonic anhydrase.[211] A significant advance in alkene and alkyne UAA transformations were reported by the Fox[212] and Hilderbrand[213] groups. They report an inverse-electron demand Diels-Alder (IEDDA) bioconjugation reaction. The dienes *trans*-cyclooctene[212] and norbornene[213] are reported to rapidly conjugate to tetrazine dienophiles (Scheme 1.11, a). This releases nitrogen upon subsequent retro- [4 + 2]-cycloaddition. The IEDDA strategy facilitates protein labelling at rates up to 1,000 times faster than SPAAC (in the case of *trans*-cyclooctene). Further exploration by the van Delft group revealed that employing UAA bearing *trans*-bicyclononene resulted in even faster ligation rates.[198] Despite these tremendous efforts limitation, such as the isomerisation of *trans*-cyclooctenes in the presence of thiols[214] (cf reaction of thiols with cyclooctynes) and the potential instability of tetrazines remain.[215] An inserted alkyne-functionalized UAA can also couple with a second, alkyne-functionalized reagent *via* a Glaser–Hay coupling. Young and co-workers[216] reported a bioorthogonal copper catalysed ligation, validated on GFP, generating a diyne product. Glaser–Hay ligations have been optimized for use in an aqueous solvent, and the ligated product amenable to further modification (Scheme 1.11, b). The Glayser–Hay ligations presents a viable alternative to click chemistry. Using a similar alkyne-functionalized UAA handle, ruthenium-catalysed alkyne hydrosilylation has been reported by Kwan *et al*.[217] This metal mediated ligation generates a C–Si bond and was validated on lysozyme and GFP. The gem-disubstituted vinylsilane product has reactivity to bioorthogonal hydrazine condensation, offering dual ligation possibilities (Scheme 1.11, c).

The Lin group has published a series of papers describing a novel approach to protein modification cycloadditions.[218-220] The group report a light induced tetrazole-alkene 1,3 dipolar cycloaddition, with the tetrazole acting as a latent source of nitrile imines. The generation of nitrile imines requires irradiation with ultraviolet light (termed a ‘photoclick’).

The reaction has been validated on a variety of alkenyl-UAAs, such as homoallylglycine[219] and cyclopropenes.[220] Additionally, genetic incorporation of tetrazoles as a reactive 'photoclick' handle has also been achieved for site selective ligation (Scheme 1.11, d).[218] The ability to spatially and temporally control a bioorthogonal ligation through light makes the 'photoclick' an attractive approach for the site-specific labelling of proteins. Despite this, the true bioorthogonality of this chemistry has recently been disputed in the literature. Yao and co-workers claim the reaction is not bioorthogonal and photoinduced tetrazole possesses reactivity towards biological nucleophiles.[221] The Yao group supports this claim through 'photoclick' labelling studies on human serum albumin protein (HSA), BSA and *in vivo* studies on HepG2 cells. The Yao group claims that despite photoclick ligation not being a bioorthogonal reaction, the chemistry can be repurposed for photoaffinity labelling (PAL), in the form of affinity- based probes (A*f*BPs).[221]



Scheme 1.11 Outline of alternative alkyne/ene bioconjugation strategies: **a**) IEDDA ligation (**b**) Glayser-Hay ligation (**c**) Hydrosilylation ligation(**d**) Photoclick ligation

Bioorthogonal reactions outside of alkyne-, aldehyde-, ketone-, and azides have also received significant development.[222] This is typically a result of the expansion of methodologies developed for site- selective UAA incorporation. A recent novel UAA incorporated is *N*-acryloyl-Lys (AcrK).[223] Alkyl phosphine reagents were used to modify AcrK containing GFP *via* a phospha-Michael addition. This reaction proceeds faster than thiol-based Michael addition. Bertozzi and co-workers recently reported a quadricyclane (QC) ligation strategy.[224] A QC-containing UAA engineered into GFP and BSA. The QC- protein was subjected to nickel bis(dithiolene), leading to a QC [2σ +2σ +2π] ligation. The ligated product can then be cleaved *via* UV irradiation.

In addition to click-type reactions, PTM mimics can be installed *via* downstream functionalisation of Dha residues, formally cysteine. The Dha insertion is a key advance in protein modification as it presents a unique electrophile for downstream functionalisation. Dha can be inserted into protein *via* a number of routes: elimination of active-site serines,[15] the oxidative elimination of unnatural selenocysteine amino acids,[225] or *via* oxidative elimination of cysteine with sulfonylhydroxylamine reagents. These strategies however can suffer from off-site reactions. These unwanted side reactions were surmounted by the Davis group, reporting a bis-alkylation-elimination of cysteine to Dha.[14] The group report initial treatment with a bisamide containing 1,4-dibromobutane core reagent to activate the cysteine residue, which then eliminates, revealing the Dha residue. Dha insertion marks the addition of an electrophilic moiety, a type of functional group not endogenous to proteins, prompting exploration of novel methods to modify Dha residues both at the *N*- terminus and internal positions.

Functionalised thiol containing reagents added to Dha bearing proteins allows rapid and selective ligation. This strategy has been used to install a number of thioether mimics of natural PTMS such as lipidation,[226] glycosylation,[227] phosphorylation[227] and lysine methylation/acetylation[225, 227] as well as installing new reactive handles for subsequent protein modification such as S-allyl cysteine for olefin metathesis.[228] Dha insertion has also allowed for a novel chemical installation of selenocysteine.[229]

Recently, aza-Michael additions have been reported by the Davis and Bernardes groups, employing amine-based nucleophiles generating secondary and tertiary amine products.[230-231] These reactions preclude thiol-based reagents to avoid possible disruption of protein disulfide bonds. These C- N bond forming reactions are stable over a wide pH range however C-C bond formation could produce closer fitting PTM mimics. Two radical-based ligations

were proposed to achieve this. The first approach reported by the Davis group[232] utilises an alkyl-halide (iodide or bromide) for the initiation of radical species in combination with preventative $NaBH_4$ minimalizing unwanted oxidation and disubstitution. This strategy was validated on subtilisin, histones, kinases, annexin V, transmembrane proteins and antibodies. A wide range of side chain scope was also demonstrated, including aliphatic, aryl, nonpolar, polar and PTMs.[232] The second radical- based ligation strategy reported by Park and co-workers[233] used amber stop codon suppression to install O-phosphoserine (Sep) followed by dephosphorylation to obtain Dha. The Dha- protein residue was then subjected to alkyl iodides as a precursor to alkyl radicals. Transmetalation from zinc to copper would then occur generating an organocopper species and produce a radical alkyl reagent. Despite these novel radical ligation strategies, both approaches are limited by their requirement of acidic (pH 4.5) media.

Owing to the development of several site-selective strategies for the insertion of unique carbonyl handles into proteins,[234-235] coupled with the general lack of aldehyde and ketone functionality expressed on protein surfaces, a plethora of bioconjugate methodologies have been developed utilising these bespoke chemical handles.[234]

Early carbonyl ligation strategies used aldehyde handles with hydrazides and aminooxy groups, generating hydrazones and oximes respectively.[236] A related reaction is the 2-amino benzamidoxime (ABAO) ligation, validated on horseradish peroxidase (HRP). This ligation strategy however is limited in use due to the requirement for acidic (pH 4.5) media.[237]

An alternative ligation approach which also takes advantage of iminium formation was a Pictet-Spenglar ligation reported by Bertozzi and co-workers (Scheme 1.12, a).[158] Firstly, an α-oxo-aldehyde handle was inserted to the *N*- terminus of myoglobin, which was then subjected to functionalised tryptamine reagents. This ligation furnishes a C-C bond linked bioconjugation and occurs under mild conditions (pH 6.5). A subsequent variation of the Pictet-Spenglar termed the iso-Pictet-Spenglar ligation was also reported by the Bertozzi group[158] taking advantage of aminooxy containing indoles in favour of tryptamine reagents (Scheme 1.12, b). This ligation strategy successfully incorporated PEGylated, biotinylated, and fluorescent chemical reporters into myoglobin, antibodies and maltose binding protein (MBP).[158] This ligation strategy albeit faster than the Pictet-Spenglar ligation and requiring lower concentration of reagents, is limited by its requirement for acidic media (pH 4.5).

Inspired by this work, the Rabuka group reported a modified iso-Pictet-Spenglar ligation, termed the hydrazino-iso-Pictet-Spenglar (HIPS) ligation (Scheme 1.12, c).[159] In this variation

of the ligation strategy, tryptamines and aminooxy indoles are exchanged in favour of *N*-methyl hydrazine bearing indoles. Owing to the increased reactivity of hydrazines comparative to aminooxy compounds, this subtle change generated improved reaction rates and enabled the ligation to complete in near neutral (pH 6) media. The Rabuka group then employed the HIPS ligation strategy for the site- selective modification of ADCs.[159]

In exploration away from iminium formation, the Rabuka group reported two Knoevenagel like condensations with protein aldehydes. The first example reported utilised a functionalised thiopyrazalone reagent for the site-selective modification of HER2 antibodies.[238] This ligation is often referred to as 'trapped-Knoevenagel' (Scheme 1.12, d) ligation and despite its attractive C-C connectivity, this ligation is limited by the requirement for acid (pH 5.5) media. The second Knoevenagel ligation reported by the Rabuka group was the tandem Knoevenagel-Michael addition.[161] This strategy reports the utilisation of functionalised pyrazolone reporters which react with antibody surface protein aldehydes to generate an enone. This enone is then subjected to a second attack by another molecule of the pyrazolone reagent, generating a bis-pyrazolone conjugate (Scheme 1.12, e). This ligation is an improvement upon the trapped Knoevenagel as it requires milder conditions (pH 7.2) however offers only limited control over ligated products.

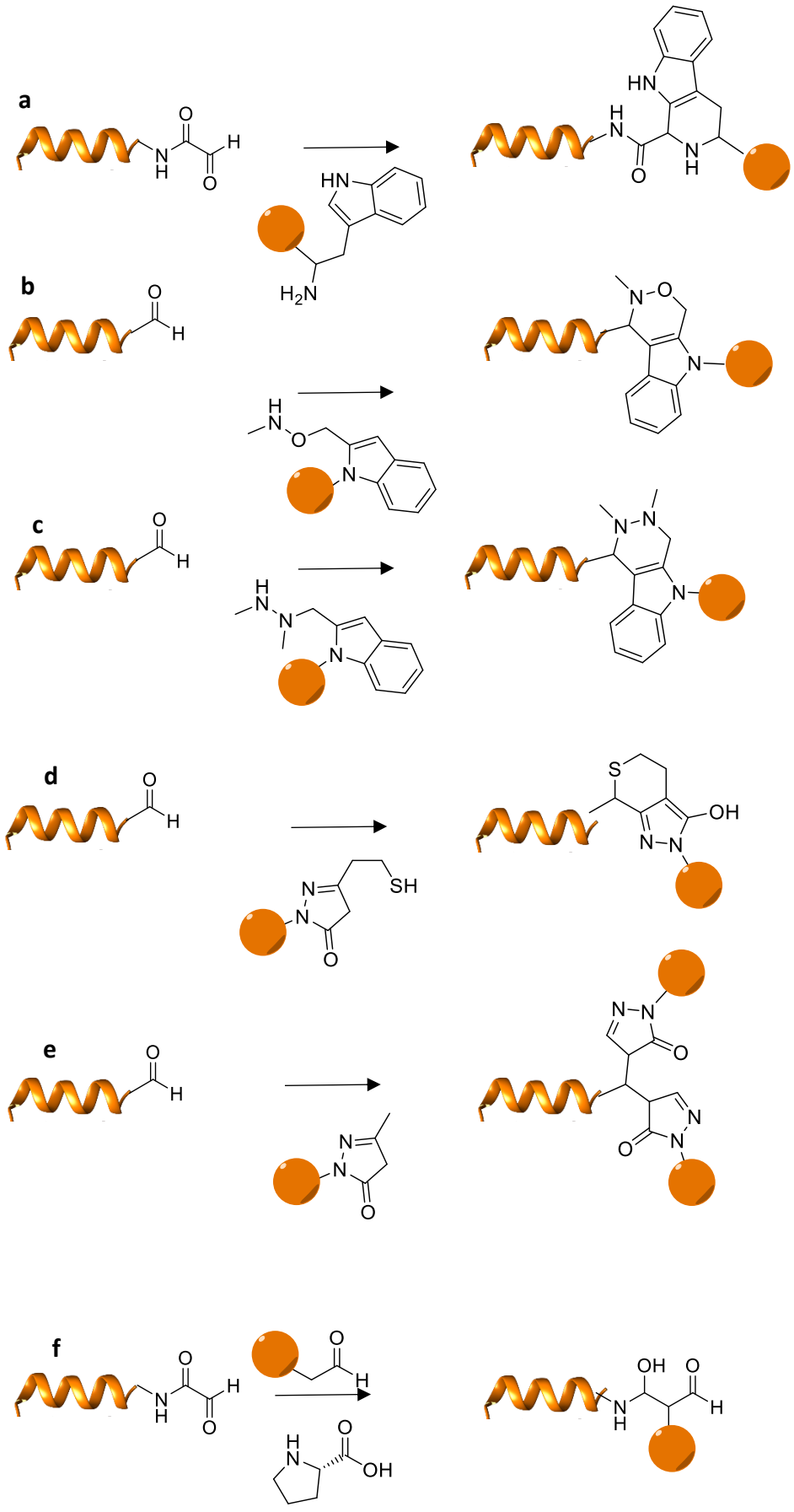
A notable contribution to C–C bond ligation was reported by Davis and co-workers utilising water-soluble and air stable 2-amino-4,6-dihydroxypyrimidine (ADHP) ligands for a Suzuki-Miyaura reaction on cysteine alkylated proteins.[239] This approach was successful in modifying proteins incorporated with a *p*-I-benzyl handles with boronic acid cross-coupling partners to install biaryl and glycosidic moieties.[239]

Another C–C bond ligation utilising carbonyl handles reported is the aldol reaction of 2,4-thiazolidinediones with *N*-terminal aldehydes.[85] The thiazolidinedione ligation has been validated on myoglobin under mild (pH 6.5) conditions. The limitations of this approach however are the requirements for *t*-BuOH co-solvent and substantial reagent excess (1000 equivalents to protein).

Although alternative aldol ligations have been reported[85, 240] a recent notable addition to the aldol ligation toolbox is the L-proline organocatalysed aldol termed 'OPAL' ligation.[66] This extension of protein aldol ligations was reported by Fascione and co-workers, detailing an enamine- type bioconjugation validated on aldehyde containing myoglobin, GFP, thioredoxin and hydrophilic acylated surface proteins (HASPs) from *Leishmania*, ligated to functionalised aldehyde reporters. This strategy successfully incorporated a number of detectible chemical

reporters, including fluorescent probes, folic acid, affinity tags, PEGylation and azide motifs. By employing an aniline catalyst and aminooxy reagents, the furnished β- hydroxy aldehyde OPAL product was susceptible to further downstream modification (Scheme 1.12, g).[66]

Lastly, utilising ketones, the Prescher group reports the ligation between protein bearing cyclopropenones, and bioorthogonal triarylphosphines.[241-242] Cyclopropenone containing lysine was first introduced into GFP *via* amber stop codon suppression. The cyclopropenone containing protein was then subjected to the phosphine reagent at pH 7 and through a ketene-ylid intermediate, generates a carbonyl containing phosphine product. It is possible to subject the phosphine product to a second ligation through azide containing reagents and perform a traceless Staudinger reaction.[242]



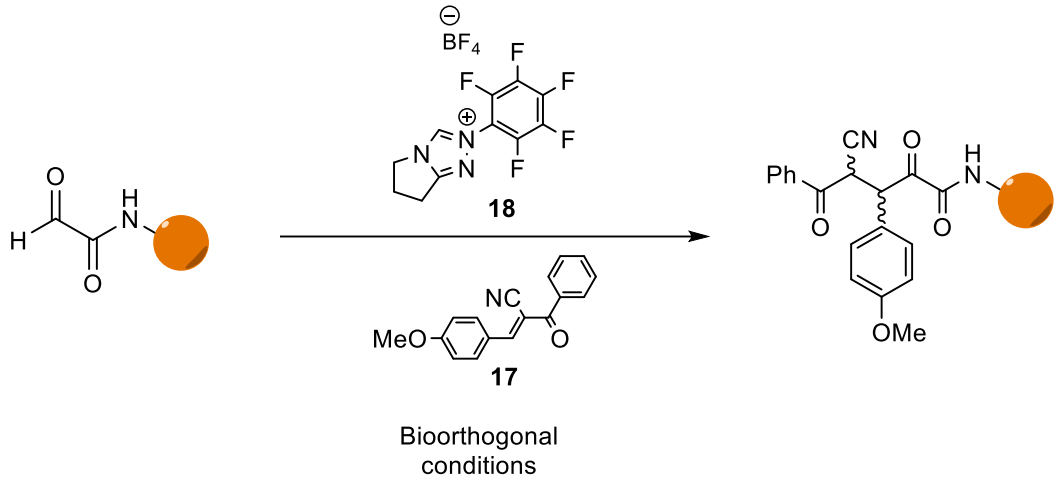
Scheme 1.12 Outline of key aldehyde bioconjugation strategies: **a**) Pictet-Spenglar(**b**) iso-Picter Spenglar (**c**) Hydrazino-Pictet-Spenglar (**d**) Trapped Knoevenagel (**e**) Tandem Knoevenagel (**f**) OPAL

1.1.9 Conclusions

In this chapter key advances and protein ligation strategies are broadly discussed. The introduction and development of amber stop codon suppression as a strategy for selective modification of proteins is bringing a paradigm shift to the field of bioconjugation chemistry. Recently, the field has been revitalised moving away from the 20 endogenous amino acids, instead focussing on introducing unique chemical handles allowing total control over protein ligation chemistries. Despite some tremendous efforts, owing to the recent rise in popularity of this field, reports are increasingly published with little to no regard within the context of bioorthogonal chemistry, straying from applicable and physiological conditions, focussing instead on novelty. Ultimately these reports add little to the current toolbox of techniques available for protein modification.

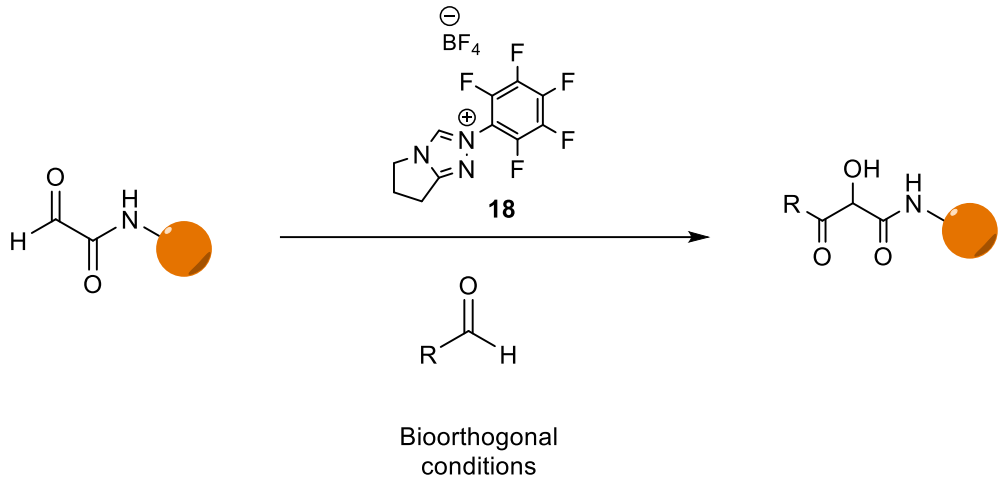
1.2 Project Outline

The primary goal of this project was to develop an Umpolung approach for the site-selective modification of α-oxo protein aldehydes using NHC catalysts under bioorthogonal conditions to generate a stable C-C bond. To achieve this, two approaches were considered and are described in this thesis. The first approach considers the application of Michael acceptors and a pentaflourophenyl substituted triazolium catalyst **18** to perform a bioorthogonal Stetter reactions on α-oxo aldehydes (Scheme 1.13). This constructs a C-C bond to form a 1, 4- dicarbonyl that is hydrolytically stable and is presented in Chapter 2.



Scheme 1.13 Outline of Stetter modification *via* Umpolung reactivity for selective protein modification

The second approach for adapting Umpolung chemistry for the modification of proteins aldehydes is the well documented benzoin condensation. Although well known, it has yet to be reported for the modification of protein aldehydes (Scheme 1.14).



Scheme 1.14 Outline of a benzoin condensation *via* Umpolung reactivity for selective protein modification

Chapters 3-8 describes the progress made towards achieving site-selective modification of protein α-oxo aldehydes *via* NHC catalysed Umpolung chemistry, detailing the scope and current limitations of this methodology.

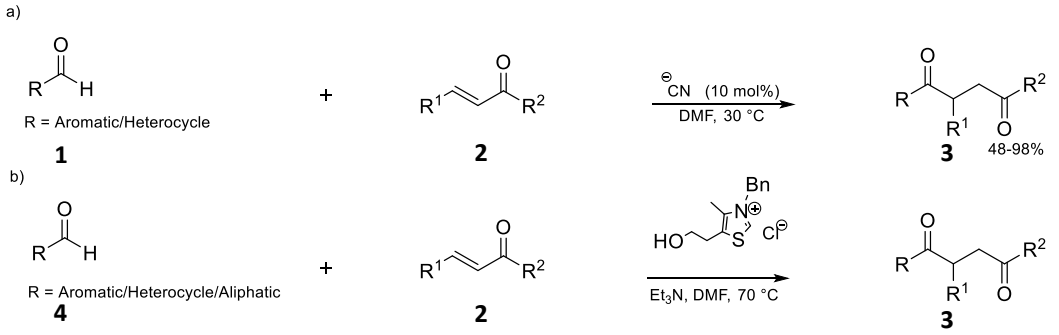
Chapter 2:

Developing a biocompatible Stetter reaction

2.1.1 Stetter Reactions

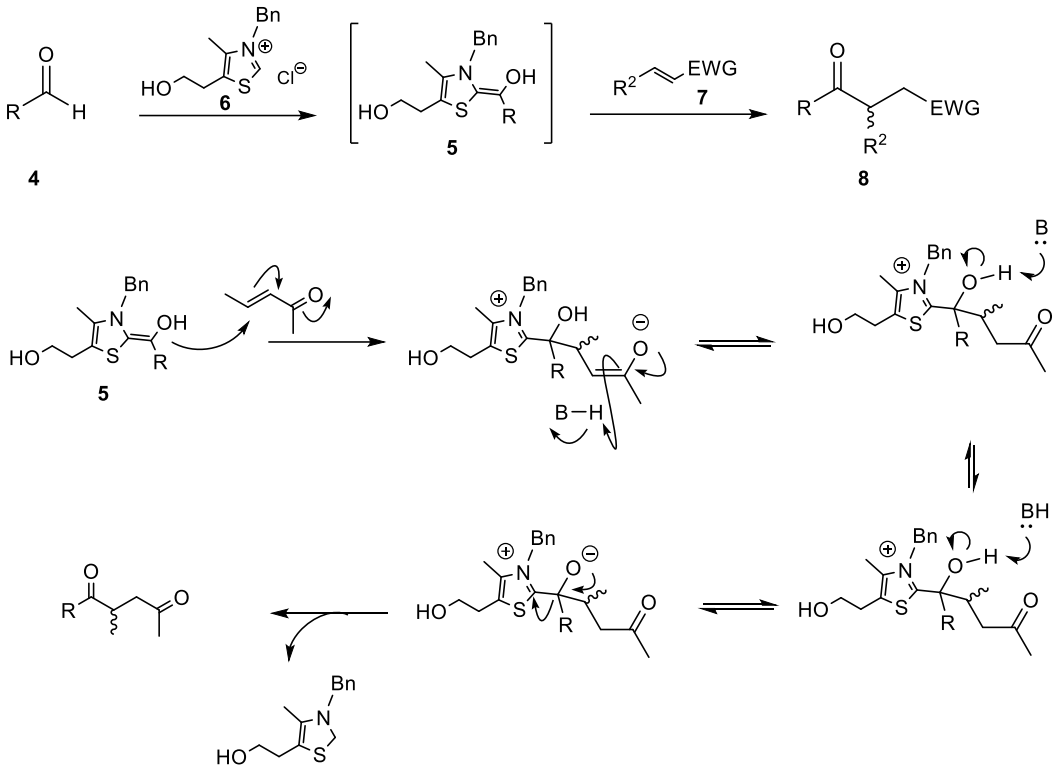
As noted in chapter 1, aldehydes can be inserted into protein motifs and owing to their low abundance can be exploited for site selective and site specific protein modification. The most common and well-developed of these methodologies is the α-oxo aldehyde handle. For this reason, the aldehyde was selected as the handle of choice for the development of an Umpolung bioconjugation investigation.

The term 'Umpolung' refers to the polarity reversal of a functional group.[243] There are several methods of inducing Umpolung reactivity in a molecule including heteroatom exchange which requires additional synthetic steps,[244-245] homologation[246] and aldehyde-aldehyde coupling.[247] This work focussed on modified Stetter reactions[248] which are considered to be the Umpolung coupling of aldehydes, acting as acyl anion synthons, with α,β-unsaturated carbonyl species. The Stetter reaction was first reported in 1973, utilising aldehydes and is well researched on small molecule systems,[249] producing 1,4-dicarbonyl products **3**. Initially this transformation was achieved using Michael acceptor **2** in the presence of a cyanide catalyst exclusively with aromatic or heterocyclic aldehydes **1** (Scheme 2.1 a).[247] However in recent years, in the presence of a triazolium or thiazolium based catalyst, the Stetter reaction has been adapted to utilise a diverse range of aldehyde reaction partners including aliphatic, aromatic and heterocyclic aldehydes **4**. It is also well tolerated with α,β-unsaturated ketones, esters and nitriles as the electrophilic partner (Scheme 2.1 b).



Scheme 2.1: Schematic representation of **a)** Cyanide catalysed Stetter reaction[21] and **b)** thiazolium catalysed

The mechanism of the Stetter reaction is well studied,[250-253] with a key mechanistic intermediate being the nucleophilic aldehyde synthon, commonly referred to as the 'Breslow intermediate' **5**.[246] The Breslow intermediate is formed by the deprotonation of pre-catalyst, generating a carbene **6**, which possesses nucleophilic character**.** The carbene **6** attacks electrophilic aldehydes **4** (Scheme 2.2), to generate the Breslow intermediate **5**. Once formed, the Breslow intermediate **5** then attacks electrophilic alkenes **7** to afford typically irreversible a 1,4-dicarbonyl Stetter product **8**.[247]

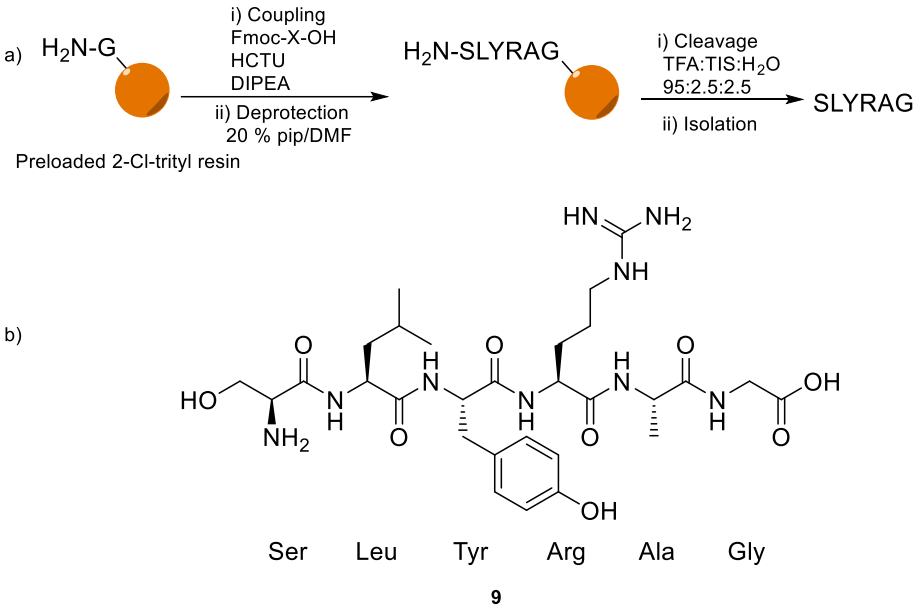


Scheme 2.2: Formation of Breslow Intermediate **5** and subsequent Stetter reaction

2.1.2 Peptide model

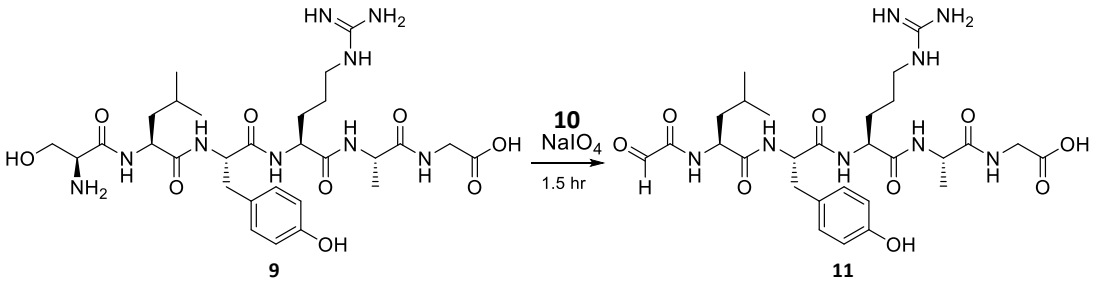
For the development of Umpolung protein bioconjugation chemistries, peptides provide a useful model system. This is as a result of their lower molecular mass compared to proteins making analysis and characterisation simpler, and they are also easier to store and use. Furthermore, in contrast to proteins, peptides can be rapidly synthesised on a large scale *via* solid phase peptide synthesis (SPPS) useful for reaction screening.

For this project the peptide SLYRAG **9** was selected for simplicity and has been well characterised in existing literature.[66] This peptide was synthesised through Fmoc-SPPS methodology (Scheme 2.3) initiating with a pre-loaded Cl-trityl glycine resin.



Scheme 2.3 Representation of **a**) Fmoc SPPS synthesis of SLYRAG **9** and **b**) SLYRAG **9** peptide post resin cleavage

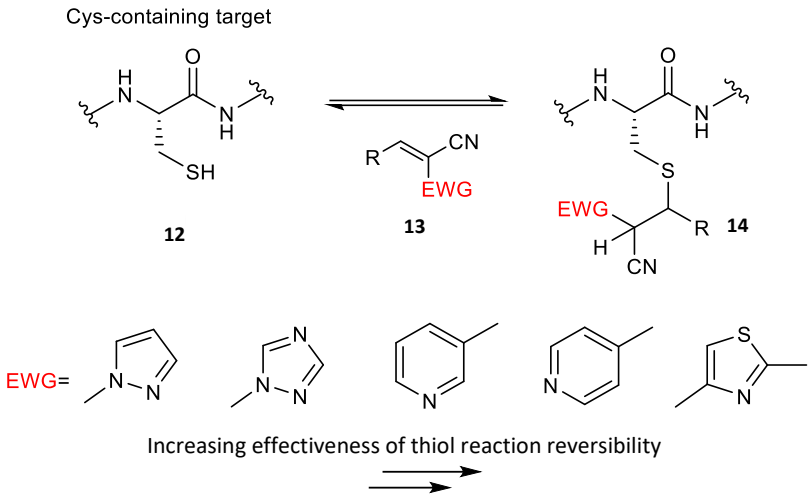
The *N*–terminal serine containing peptide **9** was then oxidised with sodium periodate **10** (Scheme 2.4) to generate the α-oxo aldehyde-LYRAG **11** confirmed *via* LCMS required for methodology screening.



Scheme 2.4: Conditions for oxidation of SLYRAG serine **9** to α-oxo aldehyde, α-oxo aldehyde-LYRAG **11**

2.1.3 Michael Acceptor

Design of reactive Michael acceptors to react with nucleophilic aldehydes, was based around work reported by Taunton and co-workers[254-255] and more recently Houk and co-workers.[256] Both groups probed the observed reversible binding of acrylonitriles containing a carboxylic ester or carboxamide α-substituent with thiols.[256] Taunton demonstrated that the magnitude of thiol **12** reversibility is increased up to 3-fold by incorporating aryl and heteroaryl-activated acrylonitrile Michael acceptors **13** (scheme 2.5).[255] Thiol reversibility in Michael acceptors is preferable as it will ensure no off-site attack at surface exposed cysteine residues occurs.



Scheme 2.5 Taunton's reversibility of thiol-Michael adducts **13** and the effect of EWG substitution[14]

In addition to this, design was also aided by Houk and co-workers' calculations of the free energy profile for the conjugate addition of MeSH into different Michael acceptors. Although heteroatoms were explored, it was found that while an α-EWG makes lowers the energy barrier for thiol addition, it also makes addition less thermodynamically favourable.[256] Upon further analysis it became evident that the substituent on the β-carbon is crucial to the energetics of the reaction (Figure 2.1).

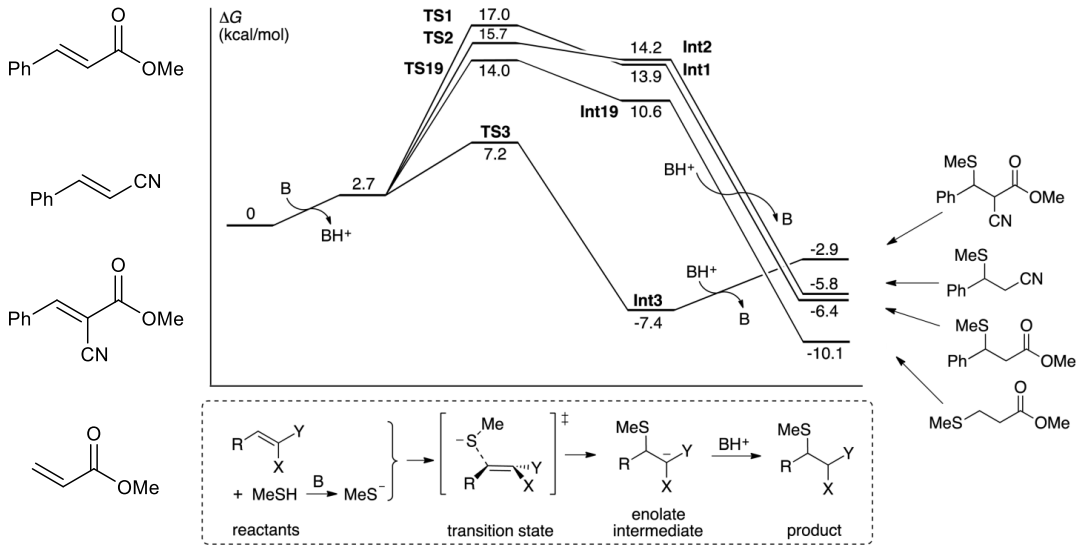
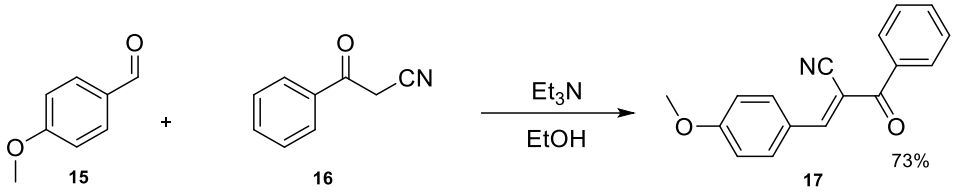


Figure 2.1 Houk's free energy profiles for additions of MeSH to Michael acceptors, adapted from Houk and co-workers [256]

These energy values correspond to a capacity of dual EWGs to firstly lower the energy barrier for thiol addition and allows formation of rapidly reversible adducts. The findings from these

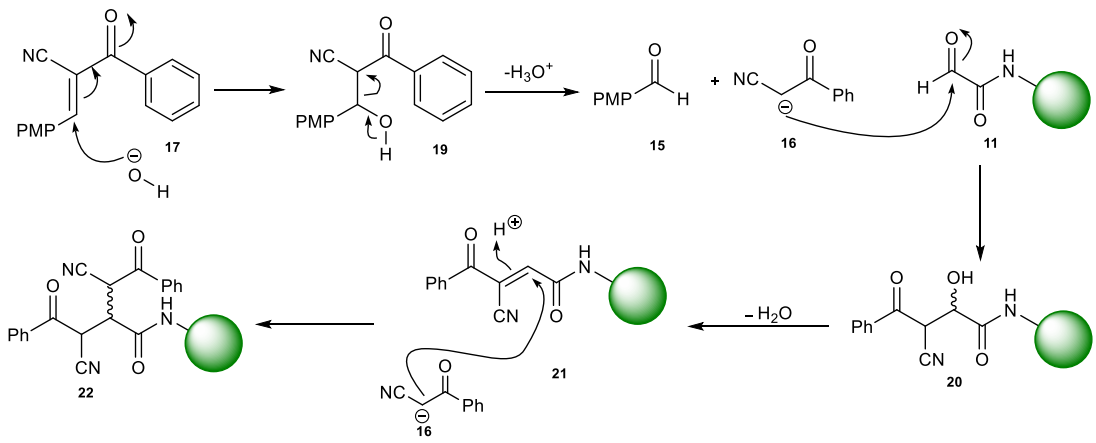
groups regarding the reversible coupling nature of di-substituted Michael acceptors to thiols is vital to the design of Michael acceptors employed in this project. Reversible thiol binding is desirable as it allows for reversible additions to cysteine, thiol containing amino acids present on proteins. Thus, the reversible binding character maintains site selective addition. Taking inspiration from this work, an appropriate Michael acceptor **17** was designed and synthesised from 4-methoxybenzaldehyde **15** and benzoylacetonitrile **16** (Scheme 2.6).



Scheme 2.6 Synthesis of acrylonitrile substituted Michael acceptor **17**

2.1.4 Testing the hypothesis

Initial attempts to perform Stetter reactions employed 20 mM pentaflourophenyl substituted triazolium catalyst **18** and 20 mM Michael acceptor **17** in 25 mM phosphate buffer pH 8 and MeCN, incubated at 37 °C with 1 mM glyoxyl LYRAG **11** over 18 h. No evidence for Stetter product was observed by LC-MS, only a putative tandem Knoevenagel Condensation-Michael addition product (TKM) was afforded, which is thought to be symmetrically di-substituted **22** (Scheme 2.7) at pH 8. The evidence for this was based on mass spectrometry and subsequent tandem mass spectrometry analysis.

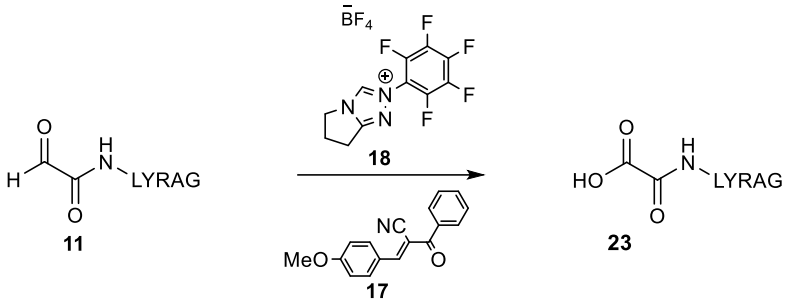


Scheme 2.7 Proposed mechanism for tandem Knoevenagel-Michael addition

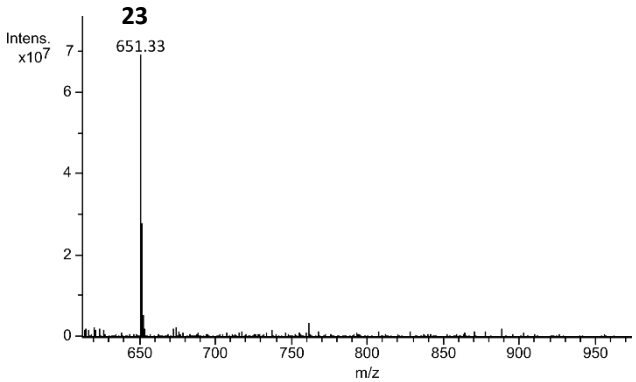
This type of tandem ligation was reported in 2016 by Rabuka's group[161] as a method for protein aldehyde modification.

To avoid this reaction, it was speculated that the TKM reaction could be better controlled under less basic conditions. For this 20 mM pentaflourophenyl substituted triazolium catalyst **18** and 20 mM Michael acceptor **17** in 25 mM phosphate buffer pH 7.4 and MeCN, 1:1 was incubated at 37 °C with 1 mM glyoxyl LYRAG **11** over 18 h. Interestingly, no evidence for Stetter product **20** was observed. Under these conditions the major product generated was that arising from the oxidation of the aldehyde **11** to carboxylic acid **23** as judged by LC-MS (Scheme 2.8).

A)

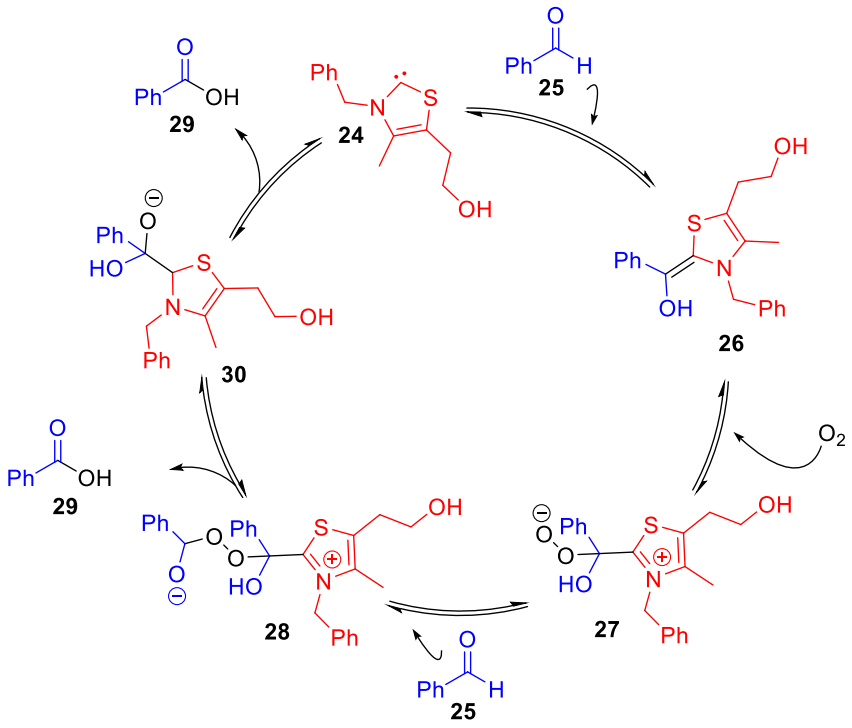


B)



Scheme 2.8: A) Reaction scheme leading to aldehyde oxidation of glyoxyl-LYRAG **11**, B) Mass spectrometry characterisation of aldehyde oxidised-LYRAG **23**

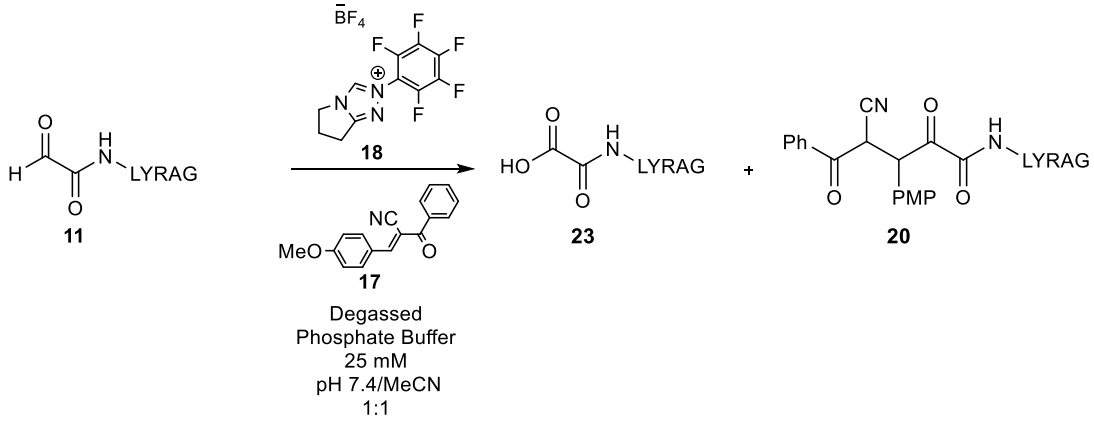
A possible mechanism for this oxidation process (Scheme 2.9) has been noted in the literature based on small molecule studies.[257-260] It is suggested that once the NHC catalyst **24** and aldehyde **25** generate the Breslow intermediate **26**, a single electron transfer event (SET) controls the rapid formation of the peroxidic species **27**. It is then speculated that the formation of intermediate **28** occurs after the peroxidic species **27** attacks a molecule of aldehyde **25** similar to a Baeyer–Villiger oxidation. Following this result, the project sought to negate this oxidation effect. NHC catalyst **24** is then reformed through release of a second carboxylic acid **29**.



Scheme 2.9 Proposed NHC mediated aldehyde oxidation *via* molecular oxygen[260]

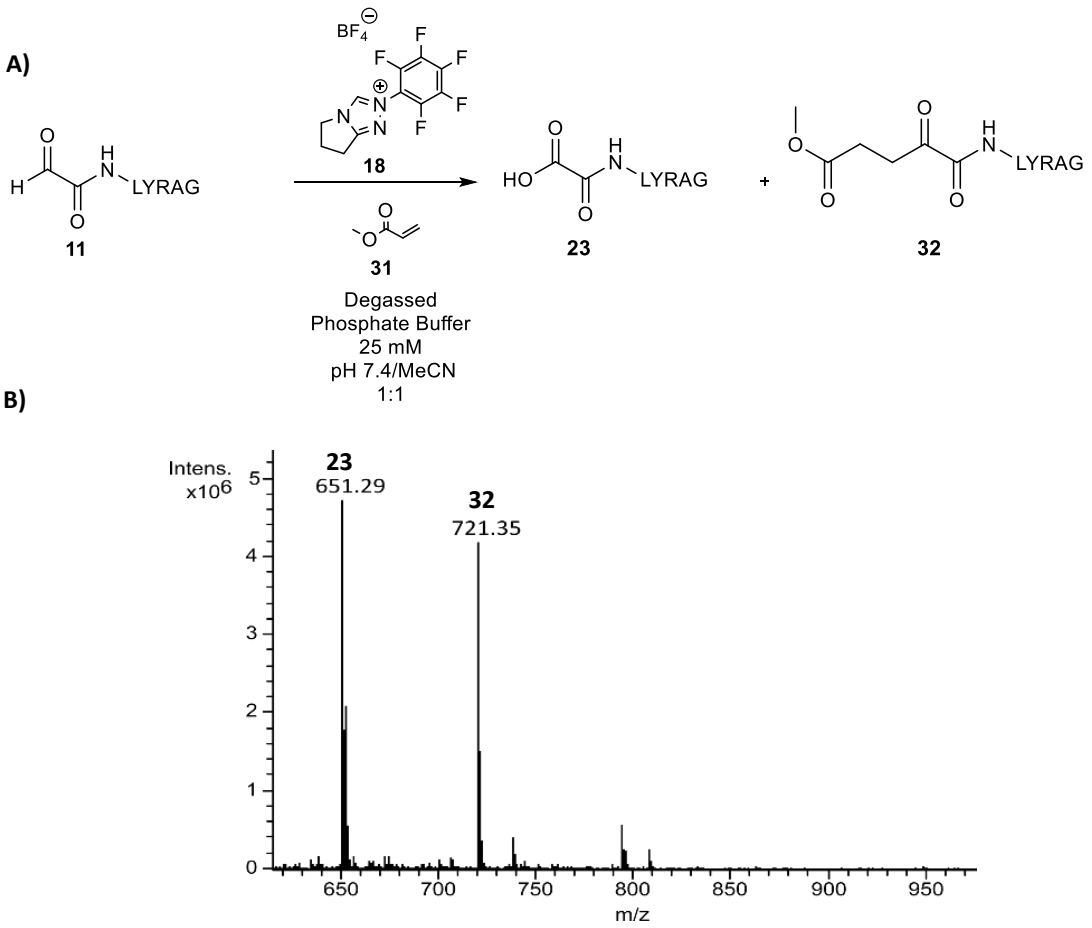
2.1.5 Addressing Aldehyde Oxidation

Attempts to preclude the unwanted aldehyde oxidation therefore focussed initially on preparing degassed buffer at pH 7, in order to reduce the oxygen content in the reaction mixture.[261] For this, 20 mM pentaflourophenyl substituted triazolium catalyst **18** and 20 mM Michael acceptor **17** in 25 mM degassed phosphate buffer pH 7.4 and MeCN, 1:1 was incubated at 37 °C with 1 mM glyoxyl LYRAG **11** over 18 h. Despite minor product generation (11%), degassing did not significantly reduce the level of oxidation observed (89% oxidation) (scheme 2.10).



Scheme 2.10 A) Degassed reaction scheme generating 89% aldehyde oxidation **23**

Degassed buffer did offer low conversion to product so it was proposed that this methodology could be improved, by using a less hindered, more reactive Michael acceptor. To probe steric effects, under inert atmosphere, 20 mM pentaflourophenyl substituted triazolium catalyst **18** and 20 mM methyl acrylate **31** in 25 mM degassed phosphate buffer pH 7.4 and MeCN, 1:1 was incubated at 37 °C with 1 mM glyoxyl LYRAG **11** over 18 h. LC-MS analysis revealed a significant improvement in conversion to Stetter product **32** (45%) (Scheme 2.11).

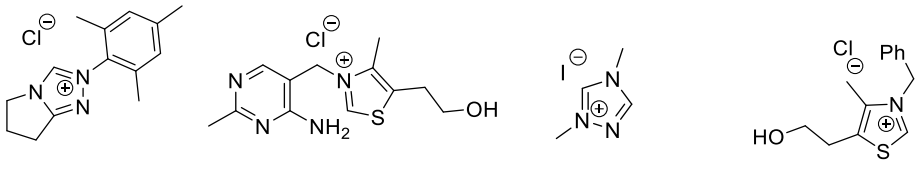


Scheme 2.11 A) Reaction scheme probing steric effects of aldehyde oxidation, B) Mass spectrometry characterisation of Stetter product **32** and aldehyde oxidised-LYRAG **23**

Having improved conversion into the Stetter product **32** significantly, it was next postulated if reduced catalyst load coupled with degassed buffer could fully preclude unwanted

aldehyde oxidation. For this 5 mM pentaflourophenyl substituted triazolium catalyst **18** and 20 mM Michael acceptor **17** in 25 mM degassed phosphate buffer pH 7.4 and MeCN, 1:1 was incubated at 37 °C with 1 mM glyoxyl LYRAG **11** over 18 h. Despite major product generation (79% as judged by LC-MS), unwanted and uncontrollable oxidation of α-oxo aldehyde-LYRAG **11** remained (21%). With this success it was hoped the application of the less hindered, more reactive Michael acceptor methyl acrylate **31** would yield quantitative conversion to the desired Stetter product. For this 5 mM pentaflourophenyl substituted triazolium catalyst **18** and 20 mM methyl acrylate **31** in 25 mM degassed phosphate buffer pH 7.4 and MeCN, 1:1 was incubated at 37 °C with 1 mM glyoxyl LYRAG **11** over 18 h. This generated 78% product **32** as judged by LC-MS however α-oxo aldehyde oxidation remained.

Using this methodology, aldehyde oxidation was seemingly unavoidable with pentaflourophenyl substituted triazolium catalyst **18** so it was next investigated if alternative catalysts could circumvent this issue under the improved reaction conditions described. It was next investigated if employing a different NHC catalyst would prove more effective in generating quantitative yields of the desired product. Showcased by Stetter's 1975 total synthesis of *cis*-jasmone and dihydrojasmone,[262] NHC catalysts emerged as powerful tools for the synthesis of complex molecules. This resulted in the subsequent development of a wide range of various NHC catalysts.[263] Today, a large number of these various NHC catalysts are commercially available. Thiazolium-based carbenes saturate the small molecule market owing to pioneering work by Stetter and co-workers. Stetter reported applying thiazolium catalyst **36** for the generation of 1, 4-diketones with both aliphatic and aromatic aldehydes.[247-248, 262] This novel development would not have been possible without a key publication from Breslow, describing the biochemistry of thiamine **34** (vitamin B1) and its ability to transform aliphatic aldehydes to acyloin products in aqueous conditions.[246] As thiazolium catalyst development continued to grow in popularity, researchers began to probe other areas of chemical space to access more powerful and highly tuned catalysts. A notable example by Olofson was the description of dimethyltriazolium catalyst **35**.[264] Since its initial synthesis, the catalyst has been used for benzoin condensations, Stetter reactions and polymerization.[265-267] Imidizolium based catalysts have also gained significant popularity. Notably, Bode and co-workers described an incredibly powerful *N*-mesityl substituted catalyst **33** in 2005.[268] The hypothesis reported for its high reactivity was the irreversible formation of the Breslow intermediate.[249, 265-266] Based on these successes, and commercial availability, the catalysts selected for this study included mesityl- substituted triazolium **33**, thiamine **34**, dimethyl triazolium **35**, and thiazolium **36** (Figure 2.2).



33 **34** **35** **36**

Figure 2.2 Alternative NHC catalyst employed to preclude unwanted aldehyde oxidation

For this, 1 mM glyoxyl LYRAG **11** was reacted with 20 mM methyl acrylate **31** in 25 mM degassed phosphate buffer pH 7.4 and MeCN, 1:1 and charged with 5 mM of NHC catalyst and left to incubate under inert atmosphere for 18 h then subjected to LC-MS (Figure 2.3). The ratios of oxidation to product (Table 2.1) suggest that regardless which NHC catalyst is employed aldehyde oxidation is unavoidable.

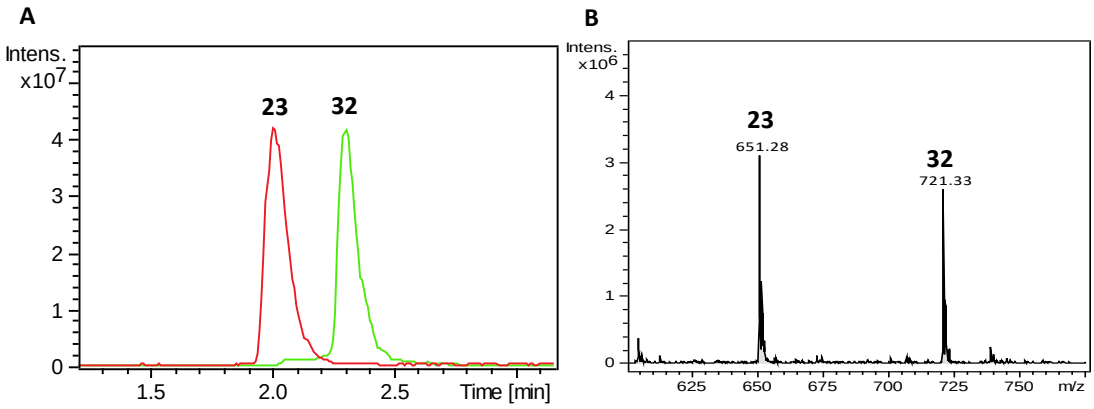


Figure 2.3 Degassed buffer with 5mM loading **33**, a) extracted ion chromatogram oxidised peptide **23** red trace, Stetter product **32**- green trace. b) Mass spectrometry characterisation of **23** and **32**.

Table 2.1: A Summary of the ratio of oxidation to Stetter product utilising a α-oxo aldehyde of SLYRAG in 1:1, pH 7.4 25 mM PB: MeCN

Entry	*Catalyst (5 mM)*	*Ratio of oxidation:Stetter product*
1	***18***	55 : 45
2	***33***	33 : 67
3	***34***	37 : 63
4	***35***	86 : 14
5	***36***	76 : 24

The data revealed that the NHC catalyst capable of generating the highest conversions of desired product was mesityl substituted triazolium **33** (67%) whilst the catalyst which generated the lowest yield of desired product was dimethyl triazolium **35** (14%). Regardless of which NHC catalyst was employed, aldehyde oxidation remained. No starting material remained upon LC-Ms analysis. The Stetter methodology albeit powerful is currently hindered by this unwanted aldehyde oxidation. This could also indicate aryl containing or aryl substituted NHC catalysts are better suited for bioorthogonal Stetter reactions based on the oxidation: product ratios measured.

2.1.6 Conclusions and Future Work

Work detailed in this section showcases the first study of an NHC- catalysed bioconjugation Stetter reaction developed on an α-oxo aldehyde peptide model. This affords a 1,4 dicarbonyl connection. The methodology suffers multiple drawbacks. Firstly, the requirement for bulky nitrile containing Michael acceptors renders reaction times slow. Sluggish reaction times can be avoided if methyl acrylate is employed, but this would undergo non-selective reactions on a protein surface. The methodology also requires 50% organic co-solvent. This requirement for organic solvent limits the application of this methodology in its current state to peptides as typical proteins would likely be denatured under these conditions. Degassing and inert chemistry techniques also limits this methodology as it is not reproducible without Schlenk line experience or a suitable glovebox. Furthermore, this methodology was complicated by a proclivity for irreversible aldehyde oxidation. This can be reduced *via* degassing the phosphate buffer (from 100% to 55% oxidation) but not completely overcome. Coupling degassing methodology with alternative catalysts can further reduce unwanted aldehyde oxidation (up to 33% oxidation) however do not fully out compete this pathway. It is possible that with further study that unwanted

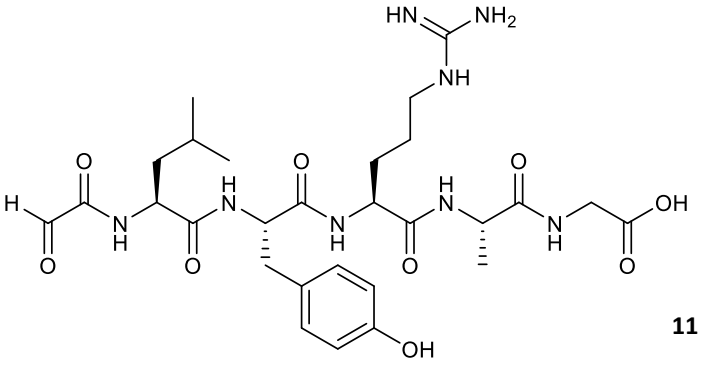
aldehyde oxidation could be surmounted. This could come in the form of employing sacrificial aldehyde, developing new NHC catalysts or the gradual addition of NHC catalyst into the reaction. This is discussed further in Chapter **7**.

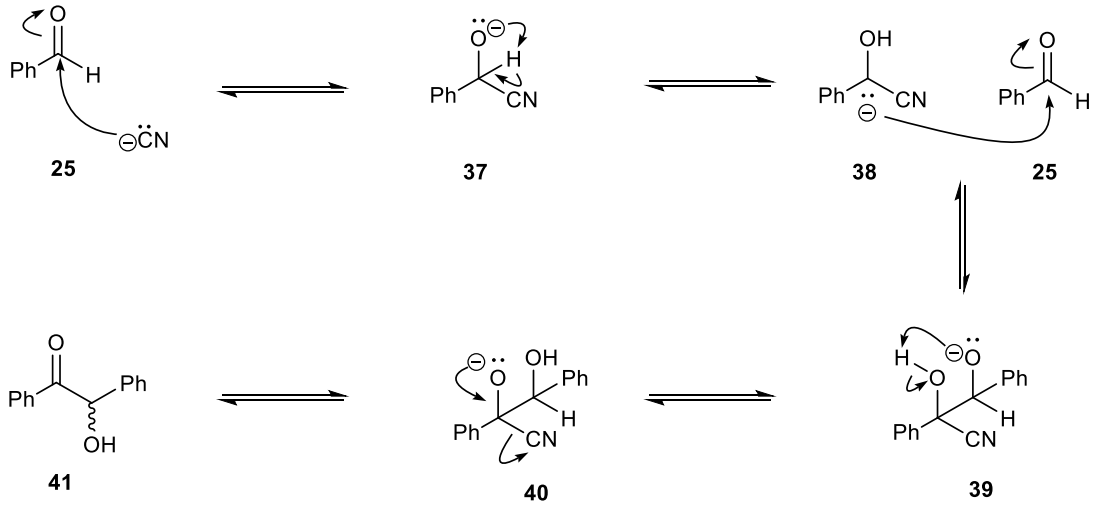
Chapter 3:

Umpolung Benzoin Condensation

3.1.1 The benzoin condensation

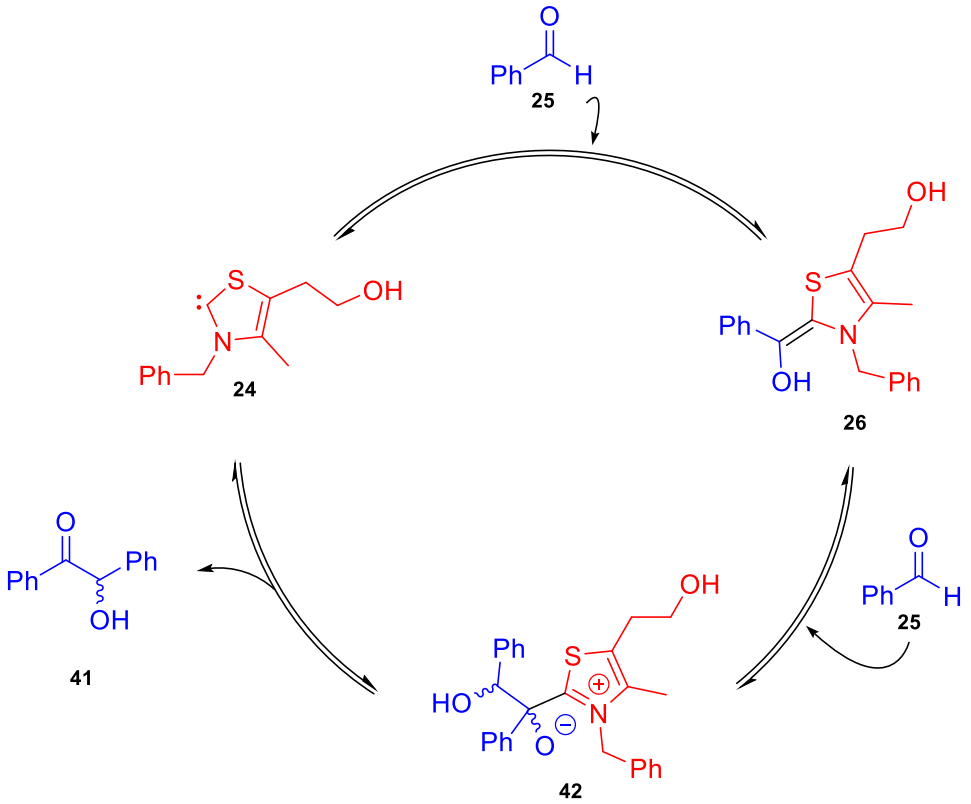
During preliminary Umpolung reaction trials with α-oxo aldehyde-LYRAG **11**, it was hypothesised that a benzoin-like condensation might be accessible under bioorthogonal conditions to selectively modify aldehyde bearing proteins. The benzoin reaction itself was first reported by Wöhler and Liebig in 1832,[269] coupling two molecules of benzaldehyde **25**, using cyanide as a catalyst, and is considered to be the first example of Umpolung chemistry. The mechanism for this was first convincingly described in 1903 by Lapworth[270] (Scheme 3.1) detailing the formation of cyanohydrin **38** and subsequent attack of an electrophilic benzaldehyde **25**.





Scheme 3.1: **A**) α-oxo aldehyde-LYRAG 1**1**; **B**) Proposed mechanism describing cyanide catalysed benzoin condensation[3]

In 1958 Breslow proposed a mechanistic cycle for an NHC mediated variation of this reaction (**Scheme 3.2**),[246] involving an enamine analogue **26** commonly referred to as the 'Breslow Intermediate' akin to the Stetter reaction discussed in chapter 2. This intermediate possesses nucleophilic character at carbon, promoting attack of a second molecule of electrophilic aldehyde. This produces the benzoin product **41** and returns carbene catalyst **24**.

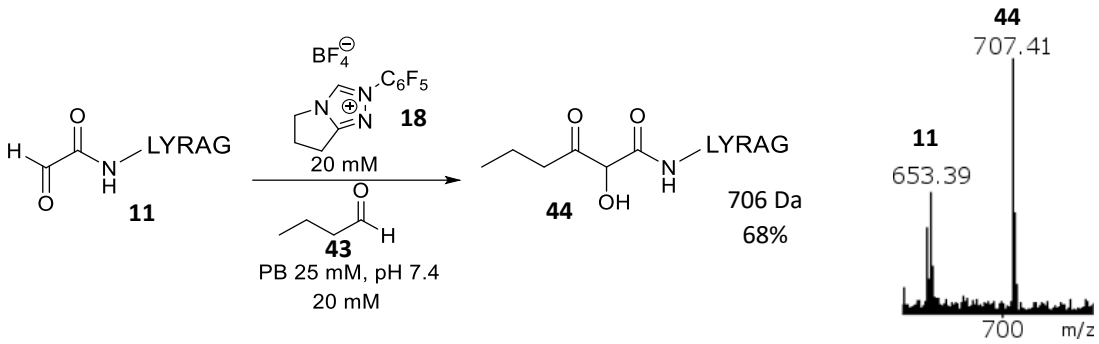


Scheme 3.2: Proposed mechanism of the NHC-catalysed Benzoin reaction by Breslow[3]

Inspired by this work, it was envisioned that a version of a traditional benzoin condensation maybe applicable to α-oxo aldehyde containing proteins, using external small molecule aldehydes as proposed nucleophiles.

3.1.2 Peptide model

To validate the possibility of NHC mediated benzoin- like bioconjugation, an α-oxo aldehyde peptide model was first employed as discussed in chapter 2. For this project, 1 mM α-oxo aldehyde-LYRAG **11** was subjected to butyraldehyde **43** in 25 mM PB at pH 7.4, and allowed to react overnight at 37 °C in the presence of triazolium catalyst **18** (Scheme 3.3).

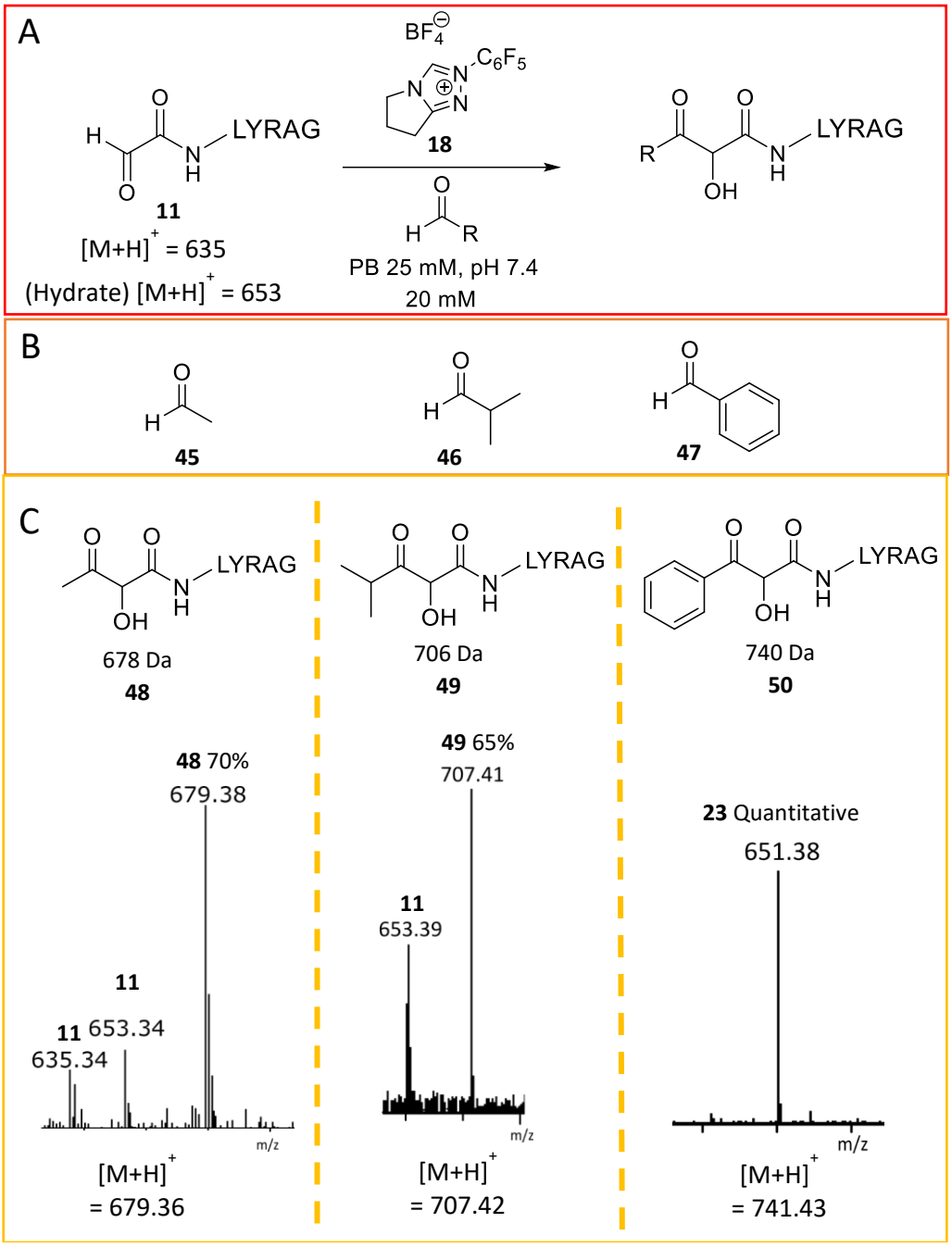


Scheme 3.3: Conditions applied in probing conditions for Umpolung reactivity and ESI LC-MS (+'ve mode)

The reaction mixture was then analysed by LC-MS which revealed *m/z* peaks corresponding to the $[M+H]^+$ species of the expected Umpolung product with conversion of 68%. Furthermore, LC-MS had shown only a single addition had occurred making this reaction suitable for controlled protein modification. Oxidation of α-oxo aldehyde-LYRAG **11** was not witnessed unlike the Stetter experiments described in Chapter 2. It is tentatively presumed that the oxidation is still occurring however on sacrificial excess aldehyde donor as opposed to the α-oxo aldehyde-LYRAG **11**.

3.1.3 Aldehyde donor screen

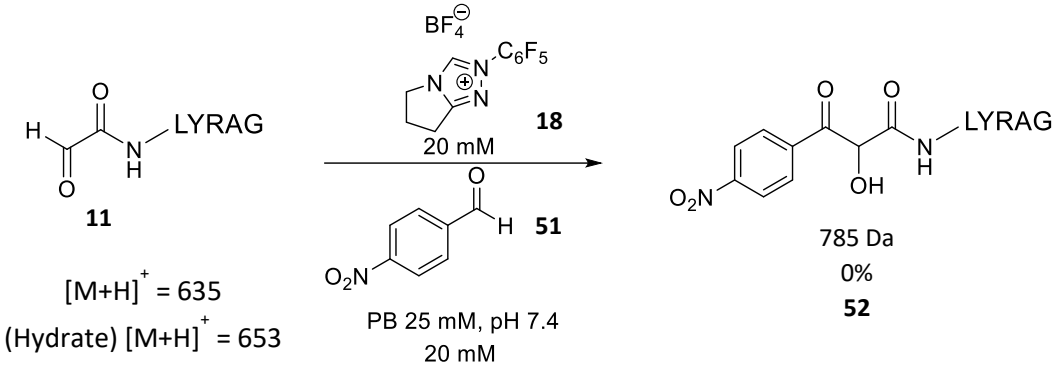
With the success of this initial peptide experiment, α-oxo aldehyde-LYRAG **11** was subjected to an array of various commercially available aldehydes in 25 mM PB at pH 7.4 to establish the scope of the transformation. The mixture was allowed to react over 18 h at 37 °C in the presence of triazolium catalyst **18** (Scheme 3.4). The aldehydes employed were acetaldehyde **45**, isobutyraldehyde **46** and benzaldehyde **47** to accommodate variables such as alkyls, aromatics as well as branched substituents. The LC-MS data generated revealed that when alkyl aldehyde substituents were employed, similar conversions to butyraldehyde **43** were achieved. Acetaldehyde **45** resulted in 70% conversion and isobutyraldehyde **46** resulted in a 65% conversion. When benzaldehyde **47** was employed, LC-MS data gave no evidence of successful ligation to peptide.



Scheme 3.4: A) Outline of conditions employed for aldehyde donor screen using α-oxo aldehyde-LYRAG **11**; **B**) alternative aldehyde donors **45**, **46**, **47** investigated for reactivity; **C**) acquired ESI-MS (+'ve mode) data for each expected Umpolung product (**48**, **49**, and **50**)

LC-MS data from the experiment utilising benzaldehyde **47** implied the overall outcome was quantitative oxidation of the α-oxo aldehyde-LYRAG **11** to a carboxylic acid **23** ($[M+H]^+$ 651). It is likely that the reactivity of benzaldehyde **47** is significantly lower than that of alkyl aldehydes due to the electron donating resonance effect of the aromatic ring, rendering the carbonyl carbon a less reactive electrophile than aliphatic aldehydes. This allows the catalyst ample opportunity to form the Breslow intermediate on the α-oxo aldehyde-LYRAG **11** which, as noted in chapter 2, in aqueous conditions leads to oxidation of the aldehyde to a carboxylic acid.[257-258, 260, 271]

In an attempt to negate the electron withdrawing effect of the aromatic ring on benzaldehyde **47**, α-oxo aldehyde-LYRAG **11** was reacted with *p*-nitrobenzaldehyde **51** in 25 mM phosphate buffer and allowed to react overnight at 37 °C (Scheme 1.5) in the presence of triazolium catalyst **18**. This was then analysed by LC-MS which revealed no evidence to suggest successful ligation to Umpolung product **52**, only starting material **11** was witnessed.

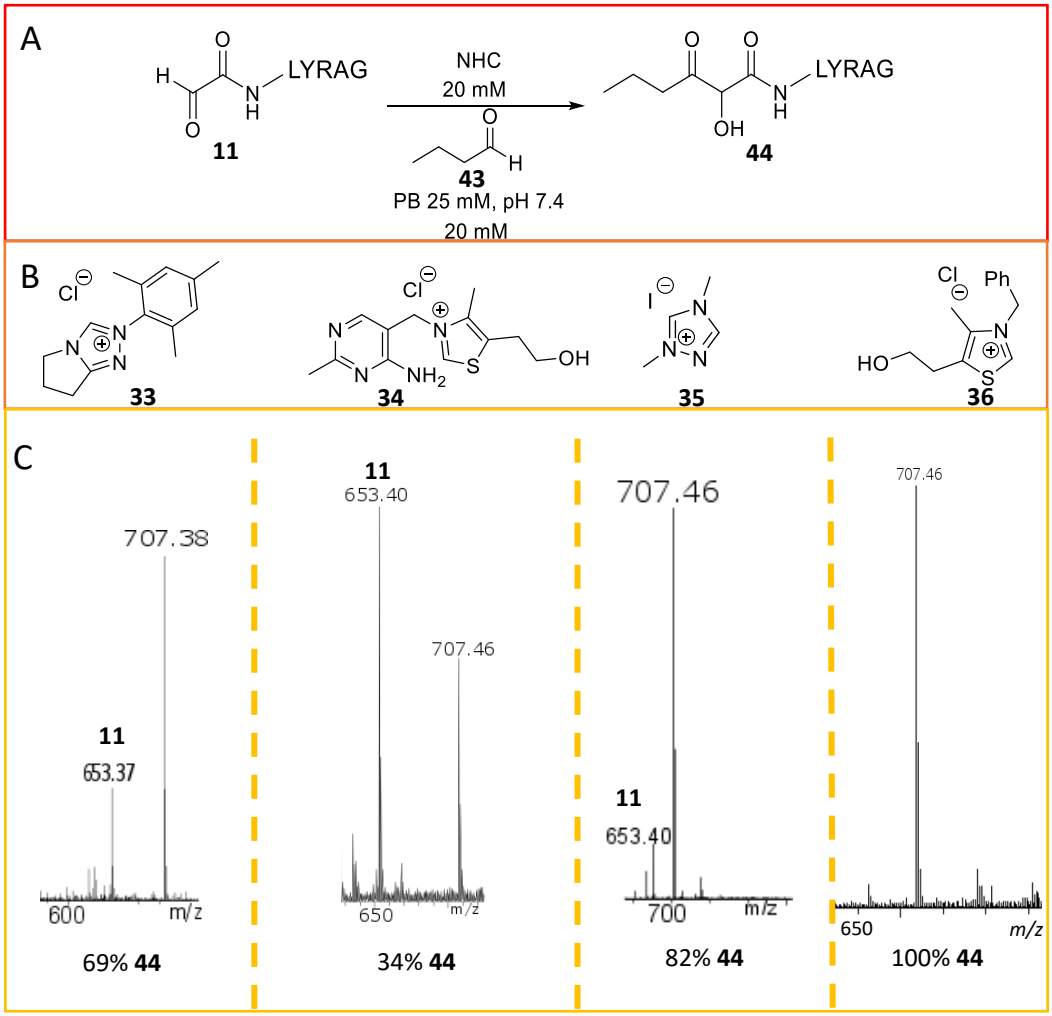


Scheme 3.5 Application of *p*-nitrobenzaldehyde **52** and LC-MS data acquired revealing only starting α-oxo aldehyde-LYRAG **11**

With these data in hand it seemed that using an aliphatic aldehyde, NHC mediated Umpolung catalysis could be employed for the site-selective modification of proteins; however, further optimisation would be required.

3.1.4 Choice of NHC catalyst I

In an approach similar to that outlined in chapter 2, once again alternative catalysts were trialled to view if a more efficient methodology could be found. For this the NHC catalysts mesityl substituted triazolium **33**, thiamine **34**, dimethyl triazolium **35**, and thiazolium **36** were reacted with α-oxo aldehyde-LYRAG **11** in 25 mM PB at pH 7.4 and butyraldehyde **43** and allowed to react overnight at 37 °C.



Scheme 3.6 A) Outline of conditions employed for the NHC catalyst screen on α-oxo aldehyde **11**; **B**) alternative NHC catalysts **33**- **36** investigated for reactivity; **C**) acquired ESI-MS data for each expected Umpolung product

The assessment of the effectiveness of these catalysts was based on their percentage conversion to anticipated Umpolung product. Based on the data gathered, the most effective catalyst was thiazolium **36** generating quantitative conversion of **11** into **44**. The least effective catalyst in this screen was thiamine **34** which promoted 34% conversion judged by

LC-MS. Using these data, it was concluded that aliphatic aldehyde donors employed with thiazolium based catalyst **36** were optimal bioconjugation conditions. The hypothesis for this observation is that the most reactive catalysts quickly generate a homo-coupled aldehyde donor leaving less for donor to couple to the peptide. This is discussed further in chapters 4 and 6.

3.1.5 Conclusions and Future Work

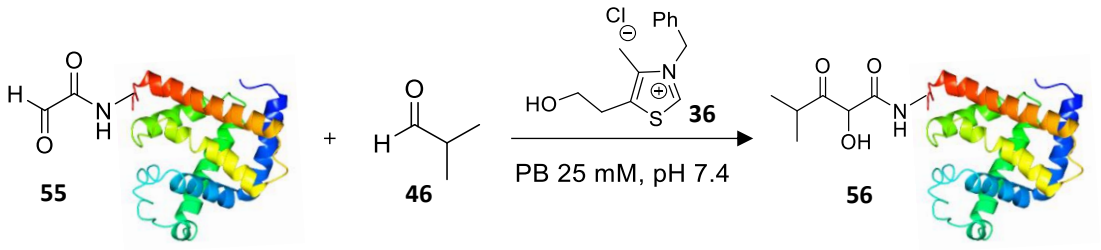
Work detailed in this section describes the first study of an NHC- catalysed acyloin condensation bioconjugation reaction developed on an α-oxo aldehyde peptide model. This affords a 1,3 dicarbonyl connection. The chapter describes the successful ligation between alkyl aldehydes, generating quantitative conversion against α-oxo aldehyde-LYRAG **11** using thiazolium catalyst **36**. The methodology described in this chapter was not able to ligate aromatic aldehydes. Interestingly, it was discovered that under the conditions outlined above, the most powerful catalyst reported in literature (mesityl substituted catalyst **33**) was not the catalyst which produced the highest conversion and that thiamine **36**, which was first studied to give NHC mechanistic insight in buffered aqueous solutions, was also not the catalyst which generated the highest conversion. The methodology holds several advantages over the Stetter methodology developed and described in chapter 2. Firstly, it does not appear to require pre-tuning to limit unwanted side reactions. This methodology simplification removes the potential need for lengthy synthetic steps, prior to bioconjugation applications. The methodology also negates the requirement for 50% organic co-solvent. By removing the need for organic co-solvent, the methodology can be applied to bioconjugation chemistry with reduced risk of denaturing proteins. The methodology described in this chapter also negates the need for the degassing of buffer and Schlenk line techniques, significantly simplifying reaction set up and removing potential need for specialised equipment such as a glove box and, subsequently, the work described in this section does not appear to suffer from unwanted oxidation of the α-oxo aldehyde-LYRAG **11**, as witnessed previously in chapter 2.

Chapter 4:

N-Terminal Protein Modification *via* Umpolung Benzoin Condensation: Discovery of the Importance of α- methyl Substitution

4.1.1 Discerning Optimal Reaction Conditions: Donor Equivalence

Using conditions established in chapter 3 for the site-selective modification of α-oxo aldehyde-LYRAG **11** with aldehyde donors, the methodology was extended to encompass protein modification. To investigate NHC mediated umpolung chemistry on protein constructs, commercially available myoglobin **53** was employed as a model substrate. Following myoglobin **53** *N*-terminal transamination with PLP **54** to generate α-oxo aldehyde myoglobin **55**, 100 µM α-oxo aldehyde myoglobin **55** was treated with 500 µM butyraldehyde **43** and 2.5 mM thiazolium catalyst **36** in 25 mM phosphate buffer pH 7.4 and allowed to react overnight at 37 °C. This experiment returned 41% yield of expected Umpolung product **56** as judged by LC-MS. Despite purification efforts, the raw data for this experiment was unclear, with the charge state ladder generated by LC-MS analysis difficult to distinguish. This was likely a result of multiple non-selective hemi-aminal formation and will be discussed further in this chapter and chapter 5. An identical repeat of this experiment was conducted exchanging butyraldehyde **43** for isobutyraldehyde **46**. Interestingly, this produced a respectable yield (52%) as judged by LC-MS and the raw data generated showed a clear and distinguishable charge state ladder. This caveat will be discussed further in this chapter. As isobutyraldehyde **46** produced the higher yield and offered facile interpretation, it was selected as the reagent of choice in a methodology based screen to deduce optimum conditions for protein modification (Table 4.1) in an attempt to replicate conversion percentages witnessed on peptides modified in chapter 3. For this screen α-oxo aldehyde myoglobin **55** was treated with an increasing equivalence of thiazolium catalyst **36** and isobutyraldehyde **46** (Scheme 4.1).



Scheme 4.1: Outline of conditions employed for methodology equivalence screen

Table 4.1: A Summary of conversion (%) to Umpolung product 56 utilising 100 µM α-oxo aldehyde of myoglobin 55 and isobutyraldehyde 46 in PB at pH 7.0

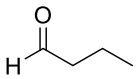
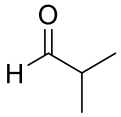
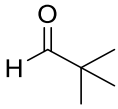
Entry	*Aldehyde loading* ***46***	*Catalyst loading* ***36***	*Results*
1	*2.5 mM*	*2.5 mM*	*Starting material*
2	*2.5 mM*	*12.5 mM*	*Starting material*
3	*2.5 mM*	*25.0 mM*	*Starting material*
4	*12.5 mM*	*2.5 mM*	*Starting material*
5	*12.5 mM*	*12.5 mM*	*62%* ***56***
6	*12.5 mM*	*25.0 mM*	*52%* ***56***
7	*25 mM*	*2.5 mM*	*Starting material*
8	*25 mM*	*12.5 mM*	*64%* ***56***
9	*25 mM*	*25.0 mM*	*75%* ***56***

The results from this methodology screen indicated that a super stoichiometric amount of catalyst and an aldehyde donor at 250 equivalences relative to protein concentration are required to push this reaction towards completion. Having established conditions which resulted in high yields attention shifted to the importance of substitution on the α carbon of aldehyde donors.

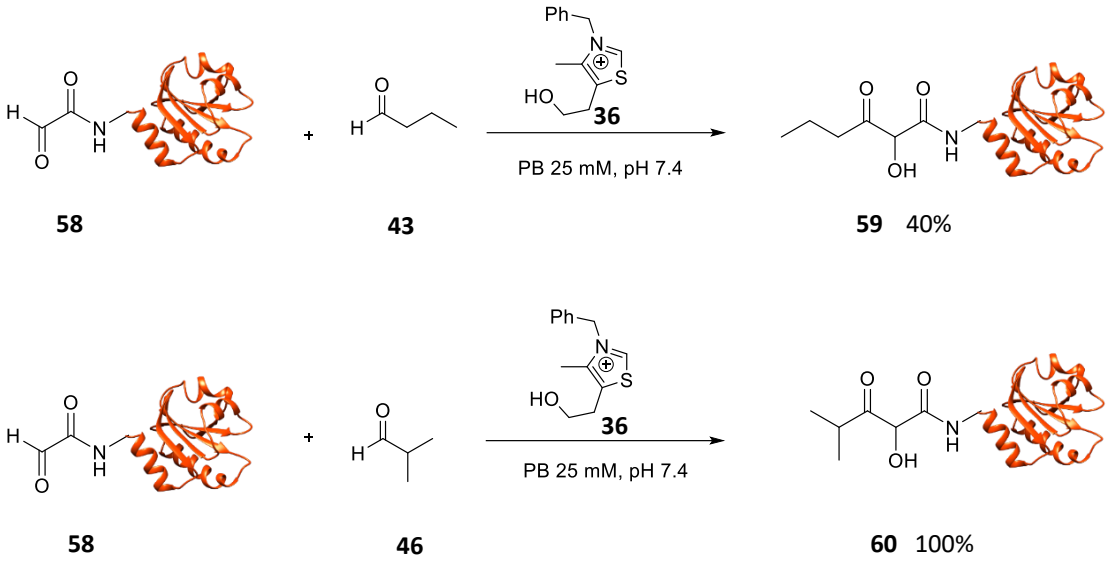
4.1.2 The Importance of α -methyl substitution

As discussed earlier in chapter 4.1, during methodology development for this chemistry, there were apparent differences between the use of butyraldehyde **43** and isobutyraldehyde **46** when modifying a protein that experiments on the peptide system did not highlight. Owing to the improved results utilising isobutyraldehyde **46** it was tentatively proposed that substitution on the α-carbon was important for the stability and reactivity of the Breslow intermediate formed. To further probe the importance of α-substituted aldehydes for this chemistry 25 mM butyraldehyde **43**, isobutyraldehyde **46** and pivaldehyde **57** were reacted with 100 µM α-oxo aldehyde myoglobin **55** in 25 mM phosphate buffer pH 7.4 overnight at 37 °C (Table 4.2).

Table 4.2: A Summary of conversion (%) to umpolung product utilising an α-oxo aldehyde of myoglobin 55 and α-carbon substituted aldehydes

Entry	*Aldehyde*	*Yield*
1	 **43**	*52%*
2	 **46**	*75%*
3	 **57**	*0%*

The results from these series of experiments indicated that substitution at the α-carbon of aldehyde donors was important, generating the higher yield than linear aldehyde donor. Interestingly pivaldehyde **57** gave no evidence of Umpolung product. This is likely due to the thiazolium carbene catalyst **36** being unable to attack the pivaldehyde **57** carbonyl, as it is blocked by the methyl substituents. Such steric hindrance caused by trimethyl substitution of pivaldehyde **57** blocking carbene attack is also in agreement with the literature in a related system.[272] To further validate this mechanistic insight, an experiment utilising a different protein was conducted; 44 µM α-oxo aldehyde thioredoxin **58** was treated with 11 mM butyraldehyde **43** and 11 mM isobutyraldehyde **46** in the presence of 11 mM thiazolium catalyst **36** in 25 mM phosphate buffer pH 7.4 and allowed to react overnight at 37 °C (Scheme 4.2).



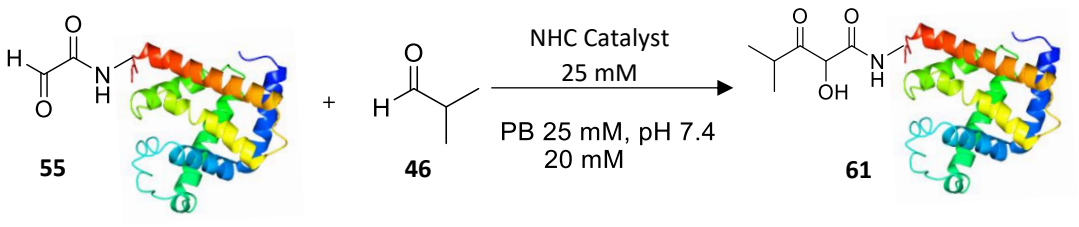
Scheme 4.2: Testing the hypothesis of improved conversions with α-carbon substituted aldehyde donors

These data appear to be in agreement that aldehyde donors bearing α -methyl substitution at the α -carbon are better coupling partners than simple linear alkyl aldehydes alone. Here the use of butyraldehyde **43** led to only 40% conversion, however applying isobutyraldehyde **46** enabled quantitative conversion to desired Umpolung product **60**. This insight however is a contradiction of the methodology data gathered using the peptide model discussed earlier and brought into question the usefulness of these results.

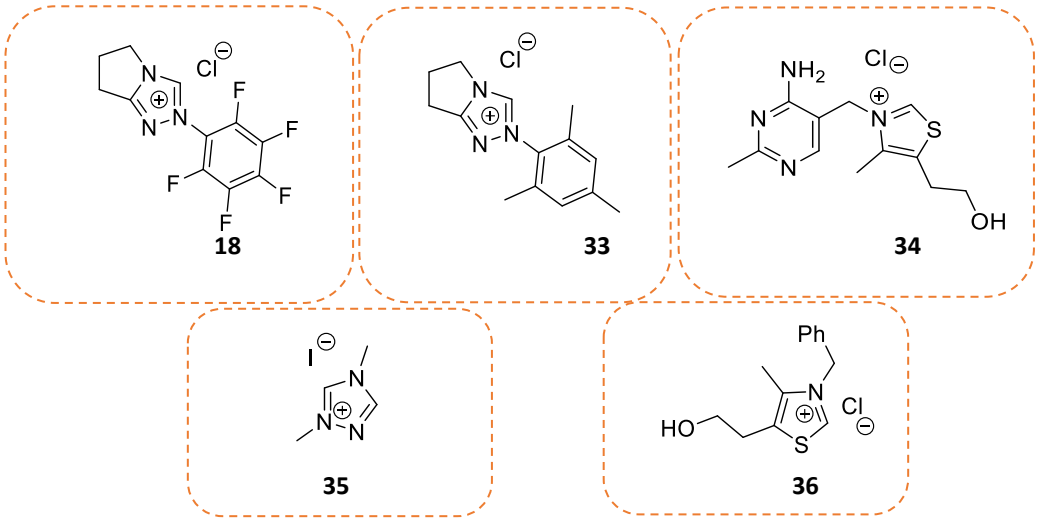
4.1.3 Choice of NHC Catalyst II

Due to the lack of consistency between data gathered on the peptide model and the protein experiments described previously, a repeat of catalyst screening using α-oxo aldehyde myoglobin **55** in favour of α-oxo aldehyde LYRAG **11** was conducted (Scheme 4.3). Here α -oxo aldehyde myoglobin **55** was treated with 25 mM isobutyraldehyde **46** and catalysts **18**, **33**-**36**, in 25 mM phosphate buffer pH 7.4 and left at 37 °C overnight.

A



B



C

Table 4.3: A Summary of conversion (%) to Umpolung product utilising an α-oxo aldehyde of myoglobin and NHC catalysts 18 and 33-36

Entry	*Catalyst*	*Yield of 61*
1	**18**	38%
2	**33**	63%
3	**34**	N/A
4	**35**	100%
5	**36**	75%

Scheme 4.3: Outline of conditions employed for protein catalyst screen; **B**) Catalysts **18, 33**-**36** screened against α -oxo aldehyde myoglobin **55**; **C**) Table of results showing conversion achieved for each catalyst employed

Interestingly the data generated from these experiments, much like the previously discussed aldehyde donors, do not follow the same trends witnessed in the peptide model in chapter 3. For NHC catalysed Umpolung protein ligation strategies, dimethyl triazolium **35** was the most efficient catalyst enabling quantitative conversion to desired product **61** on proteins and not the thiazolium catalyst **36** as previously believed from earlier peptide studies. This is the opposite of what was expected based on the peptide studies discussed in chapter 3. This could be due to the less accessible protein aldehyde handle or change in reaction scale. One study which could provide some insight could be the investigation of the formation and life span of the Breslow intermediate formed by each catalyst as this could significantly affect the reaction. In conclusion these data indicate that a system for site selective modification of aldehyde bearing proteins through Umpolung chemistry should take advantage of a branched alkyl aldehyde donor employed with dimethyl triazolium catalyst **35**.

4.1.4 Assembling the benzoin condensation protein modification protocol

To confirm that the methodology developed in this chapter would indeed be capable of site selective protein modification, isobutyraldehyde **46** was replaced with commercially available 2-methylpentanal **62**. The reasoning for this substitution was to ensure the methodology developed in this chapter could be applied to other α -methyl substituted alkyl aldehydes and not exclusively isobutyraldehyde **46**, and to better replicate the reactive probes intended for protein ligation (Figure 4.1).

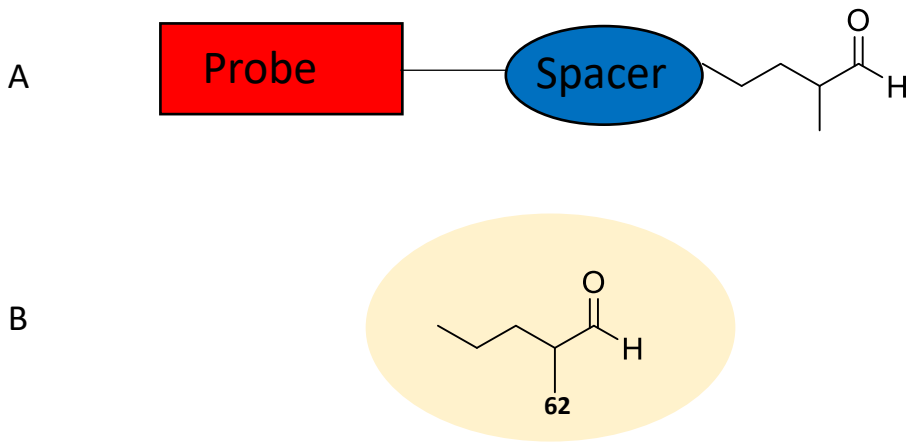
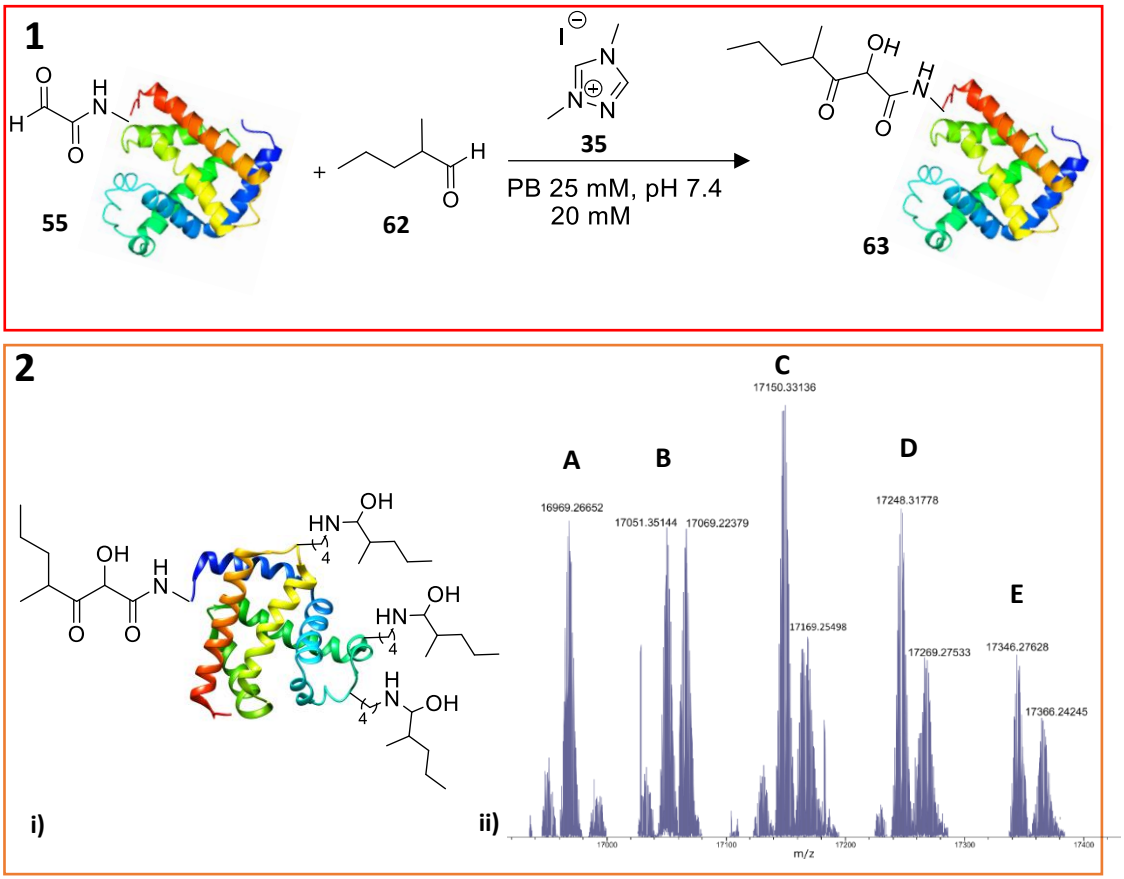


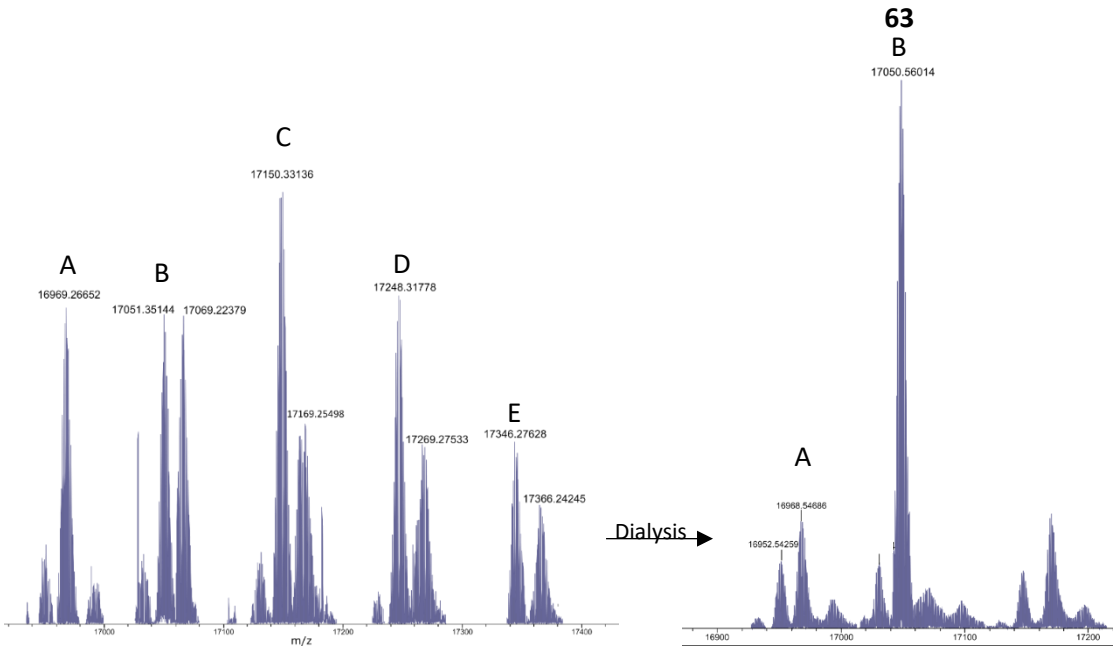
Figure 4.1: Schematic of a probe design based on alkyl α-carbon substituted properties derived from previously discussed myoglobin experiments; B) commercially available 2-methylpentanal **62** to replicate proposed probe.

To confirm the methodology insights discussed in this chapter, 100 µM α-oxo aldehyde myoglobin **55** was subjected to 25 mM 2-methylpentanal **62** and 25 mM dimethyl triazolium catalyst **35** in 25 mM phosphate buffer pH 7.4 and left at 37 °C overnight. Unexpectedly, under these conditions, protein modification was achieved however LC-MS analysis indicated the presence of multiple non-specific ligations (Scheme 4.4).



Scheme 4.4: 1: Outline of conditions for methodology confirmation; **2i**: representation of expected product with non-specific reversible addition, **2ii A**- myoglobin **55** starting material, **B**- desired product **63**, **C**- **63** + 1× hemi-aminal addition, **D**- **63** + 2× hemi-aminal addition, **E**- **63** + 3× hemi-aminal addition

Fortuitously these non-specific ligations are not stable C-C bonds but hydrolytically cleavable hemi-aminals formed from free amines at surface exposed Lys residues on the protein. These exposed amines attack the 2-methylpentanal **62** to form the hemi-aminal bond. To hydrolyse these hemi-aminal bonds, following reaction completion, the protein was dialysed overnight. Following this dialysis period, the protein was subjected ESI-MS revealing quantitative single modification to desired Umpolung product **63** and complete hydrolysis of the hemi-aminal bonds (Scheme 4.5).



Scheme 4.5: Deconvoluted mass spectrum showing **i**) pre-dialysis, **A**- starting material, **B**- desired product **63**, **C**- **63** + 1× hemi-aminal addition, **D**- **63** + 2× hemi-aminal addition, **E**- **63** + 3× hemi-aminal addition; ii) post-dialysis major desired product **63 B**

4.1.5 Probing Specific Ion Effects

Once optimum conditions were established for Umpolung protein modification, salt effects were investigated. In the 1880s research led by Hofmeister probed the ability of different ions to precipitate proteins from blood serum and egg whites.[273] From this, the Hofmeister series was born, which is effectively a ranking system for anions and their effect on proteins in solution (Figure 4.2).

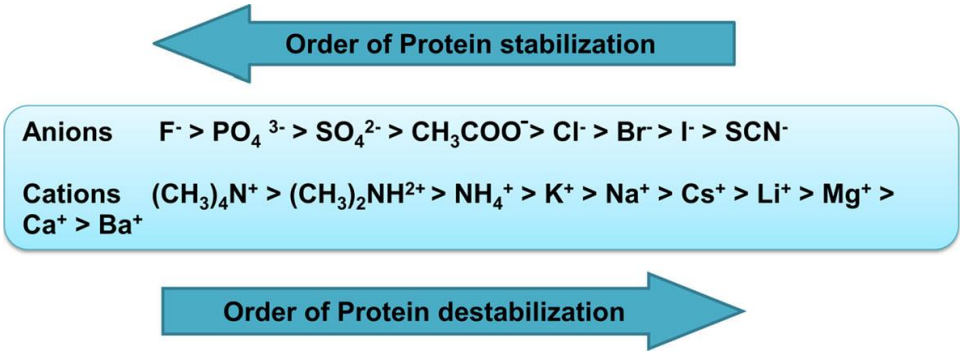


Figure 4.2: Schematic of the Hofmeister series taken from Patel *et al*[274]

Anions on the left side this series salt out (precipitate) solutes and anions sitting toward the right salt in (dissolve or denature) solutes. This is also true for cations, which are also ranked based on the work by Hofmeister. It is largely accepted that this series is a reflection of the specific ion effects on the structure of water. For more detailed explanations on specific ion effects, investigations probed ion behaviour at interfaces. Molecular dynamics simulations assessing anions adsorption at the water-air interface suggest an inverse Hofmeister scale.[275-277] These contradictory data indicate that specific ion effects could reflect differences in ion hydration levels at the air-water interface and biomolecular surfaces compared to a simplistic effect on the overall structure of water. This is supported by Pegrow and co-workers who demonstrated a correlation existing between anion adsorption to the air-water surface and their effect on protein surface hydration.[276] Chen *et al* published similar findings suggesting changes in hydration of ions upon interaction with protein-like surrogates.[275] Exploring specific ion effects is tempting, possibly allowing for the fine tuning of bioconjugation reaction rates whilst adhering to strict biocompatible conditions.

In the 1980s Breslow also reported specific ion effects on the benzoin condensation.[278] Breslow found that in aqueous conditions employing KCl with KCN catalyst, the reaction was 3× faster than without. In addition to this, Breslow also observed the ability of LiCl to increase the reaction whilst lithium perchlorate and lithium iodide effectively shut the reaction down. This is likely due to release of a cation and anion which both sit far right on the Hofmeister

series. Breslow also reported similar trends in salt effects on aqueous Diels-Alder reactions.[279-281]

More recently, Pentelute and co-workers described using specific salt effects to improve their π- clamp bioorthogonal protein modification.[282] This work motivated an investigation into the effect salt addition has on a cross-acyloin bioconjugation described in this chapter (Figure 4.3). To investigate salt effects, 100 µM α-oxo aldehyde myoglobin **55** was subjected to 25 mM isobutyraldehyde **46** and 25 mM dimethyl triazolium catalyst **35** in 25 mM phosphate buffer pH 7.4 with 1M lithium chloride, ammonium sulfate or sodium chloride and left at 37 °C overnight.

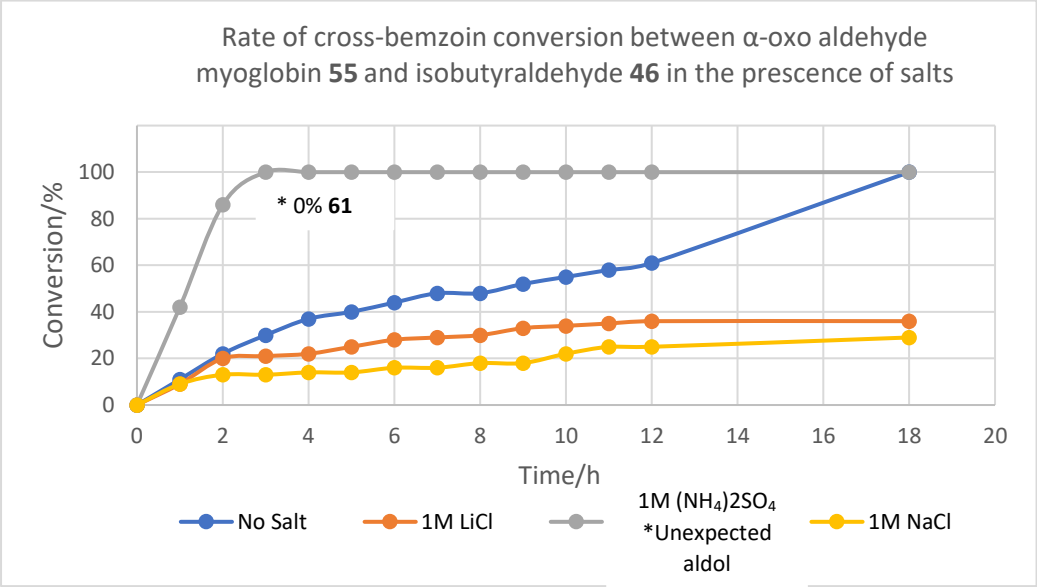
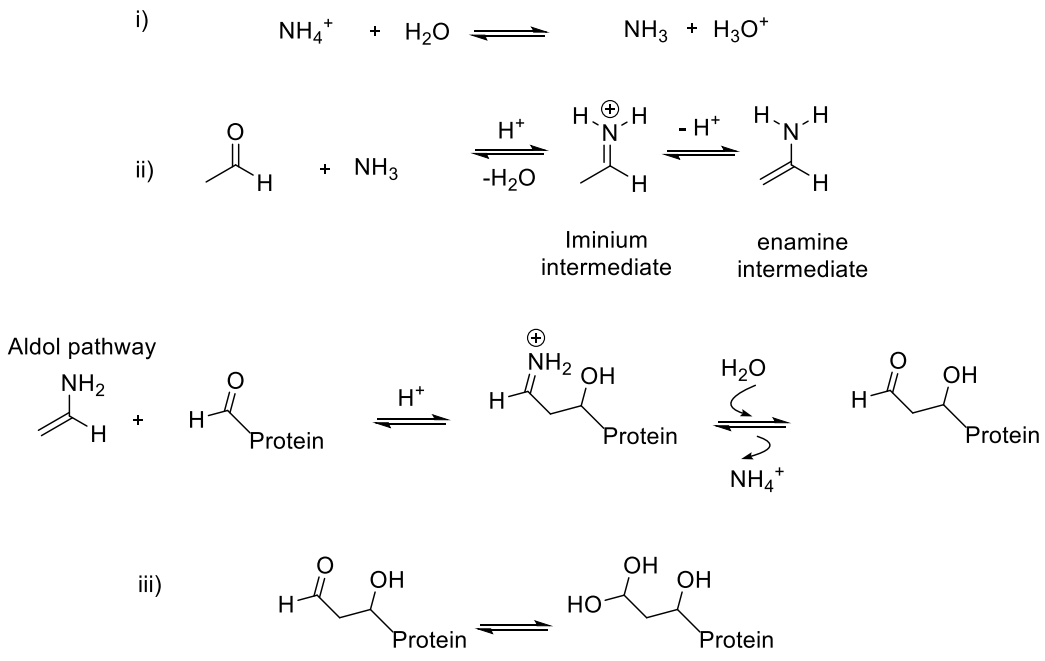


Figure 4.3: Graph summarising various salt effects on the bioorthogonal cross-benzoin over an 18-hour period.

When employing 1 M NaCl (yellow trace) and 1 M LiCl (orange trace) the reactivity is substantially reduced when compared to a control system without salt (blue trace). Interestingly, the application of 2 M LiCl (not shown) completely shuts the reaction down and no product was derived over an 18-hour period. Ammonium sulfate (grey trace), including 2 M (not shown), did not precipitate protein as was expected. This is likely a result of the fine balance that exists between ion-water and ion-protein interactions. Ammonium

sulfate at 2 M, 1 M and 0.5 M produced quantitative product conversion in under 3 hours however, a control experiment precluding triazolium catalyst gave identical results. Analysis of LC-MS data revealed a prominent [M+18] peak. This is associated with the hydration of an aldehyde, as witnessed in previous LC-MS experiments on peptides bearing an α-oxo aldehyde (Chapter 2 + 3) and also on protein aldol reactions previously reported.[234] This is therefore indicative of an aldol product and is attributed to the formation of ammonia and subsequent enamine catalysed aldol reactivity (Scheme 4.6) and is reported in the literature.[283] As it has been reported that the β hydroxy aldehyde generated from the aldol reaction can be further modified, this LCMS insight could be further clarified *via* a reaction with aminooxy detectible chemical reporters.



Scheme 4.6: Probable mechanism from ammonium salt to enamine catalysed aldol reaction.

4.1.6 Conclusions and Future Work

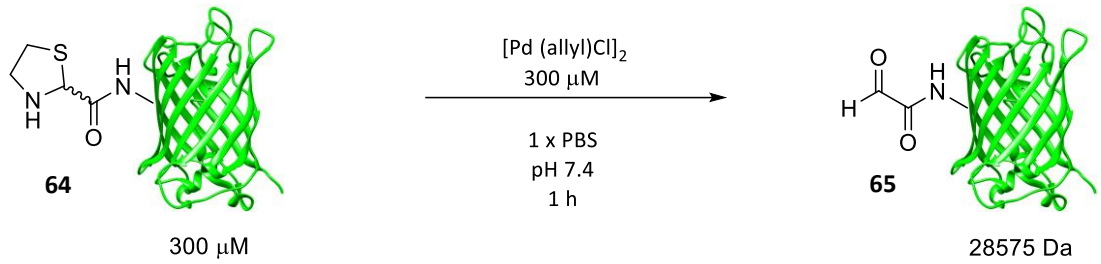
Work detailed in this section describes the successful implementation of an NHC- catalysed benzoin condensation bioconjugation reaction on *N*-terminal α-oxo aldehyde proteins. The chapter describes the discovery and confirmation of the importance of α -methyl substituted alkyl aldehyde donors and that further substation renders the donors unable to participate in the reaction. The requirement for α-methyl substituted alkyl aldehyde donors will be discussed further in chapter 6. This chapter describes the disparity between optimal conditions for Umpolung α-oxo aldehyde peptide modification and α-oxo aldehyde umpolung protein modifications, the latter requiring larger equivalences of catalyst and donor. Unlike chapter 3 describing the requirement of 20 equivalents of aldehyde donor and NHC catalyst to modify peptides, a significant increase to 250 equivalents are needed to promote the protein modification. This is likely a result of proteins being much larger constructs than peptides and therefore slower reaction partners. Additionally, this chapter also describes that the most efficient catalyst for protein ligation is dimethyl triazolium catalyst **35**, and not thiazolium **36**, as the peptide model discussed in chapter 3 suggests. This is tentatively assumed to be a result of the dimethyl triazolium catalyst possessing the least steric hindrance of all the NHC catalysts surveyed. Also described in this chapter is the discovery of multiple, non-specific ligation of excess aldehyde donor to what is suspected to be surface exposed lysine residues. This does not appear to be a limitation of this developed methodology however as these hemi-aminal bonds formed are easily and fully cleaved *via* hydrolysis. Finally, this chapter describes the investigation into specific ion effect on this chemistry in an effort to improve reactivity times. This study revealed the reaction slowed in the presence of 1M LiCl and NaCl. The presence $(NH_4)_2SO_4$ facilitated an unwanted and uncontrollable suspected aldol reaction. This tentative assumption is based on LC-MS analysis and trends seen in similar systems,[66] control experiments without triazolium cataltyst yielding identical results and small molecule reports in literature.[283] Despite this false positive and developing methodology that shuts down the reaction pathway, for completeness it would be intriguing to probe the effect of potassium cations in future studies. Potassium is situated directly below ammonium cations in the Hofmeister series however, would be unable to catalyse unwanted Aldol reactions.

Chapter 5:

Beyond the *N*- terminus: Internal protein modification

5.1.1 Developing methodology for internal bioconjugation

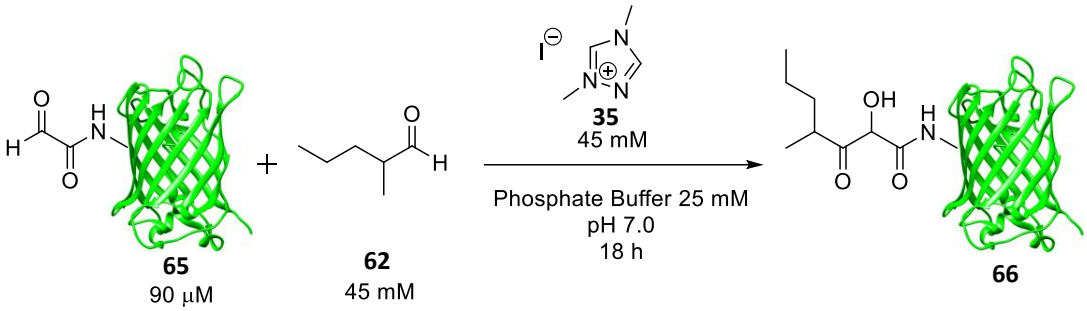
Having confirmed in chapter 4 that NHC mediated Umpolung chemistry could be achieved and employed as an applicable procedure for the selective ligation of proteins bearing *N*-terminal α-oxo aldehydes, its capability to modify aldehydes at internal sites was next assessed. As discussed in chapter 1 UAA mutagenesis is a powerful tool for inserting functionality site-specifically into proteins.[174] First, an internal α-oxo aldehyde was inserted into green fluorescent protein (GFP).[235] This was achieved through transformation of a mutated plasmid carrying a GFPY39 bearing an amber stop codon within its plasmid sequence into *E. coli* Top10s cells. When co-transformed with a plasmid containing a pyrrolysine (Pyl) tRNACUA/pyrrolysyl-tRNA synthetase (RS) pair, which have a high affinity for the amber stop codon, and subsequent addition of unnatural amino acid (UAA) ThzK is inserted into the position of the amber stop codon.[235] Once purified, this UAA containing GFPY39ThzK **64** was treated with palladium allyl chloride dimer to generate the GFPY39 α-oxo K **65** (Scheme 5.1).[1]



Scheme 5.1: Treatment of UAA containing GFP **64** with Pd to unmask GFPY39 α-oxo K **65**

The resulting GFPY39 α-oxo K **65** was reacted with 250 equivalence of dimethyl triazolium catalyst **35** and 2-methylpentanal **62** in 25 mM PB pH 7.4, at 37 °C overnight. Upon LC-MS analysis however it was noted that the same level of conversions previously witnessed on the *N*-terminal trials were not observed, with only 52% conversion into the expected Umpolung product **66** calculated using LC-MS. The experiment however did result in GFP remaining fluorescent green and had not denatured demonstrating the biocompatibility of the reaction. In an effort to improve this yield, the experiment was repeated with double the equivalents of NHC catalyst and aldehyde donor. In this experiment, GFPY39 α-oxo K **65** at 90 μM was reacted with 500 equivalence of dimethyl triazolium catalyst **35** and 2-methylpentanal **62** in 25 mM PB pH 7.4, at 37 °C overnight (Scheme 5.2).

[1] Glycerol starter cells and ThzK provided by Dr Robin Brabham



Scheme 5.2: Conditions employed for the complete modification of non-terminal GFPY39 α-oxo K **65** with 500 eq dimethyl triazolium **35** and 2-methylpentanal **62**

Following an 18-hour period this reaction was dialysed into MiliQ water at approximately 4 °C to cleave the expected non-selective hemi-aminal bonds likely formed and then analysed *via* mass spectrometry. This experiment also resulted in GFP remaining fluorescent green. The LC-MS data indicated however, that one prevailing hemi-aminal had not been hydrolysed (*m/z* 28853) (Figure 5.1). Furthermore, it was concluded that another addition also remained, a hemi-aminal that had likely collapsed to produce an imine (*m/z* 28755) (Figure 5.1). It is suspected that the imine produced is located at the exposed *N*-terminal amine of the GFPY39 α-oxo K **65** and the observed imine in favour of the usually witnessed hemi-aminal is a result of the differences in pKa between the ε-NH_2 group of lysine and the *N*-terminal serine. The exposed amine on Lys has a pKa of approximately 10.54 whereas the *N*-terminal serine in GFP is approximately 9.15.[284] In *N*-terminal α-oxo aldehyde modifications this was not witnessed due to the free amine not being present.

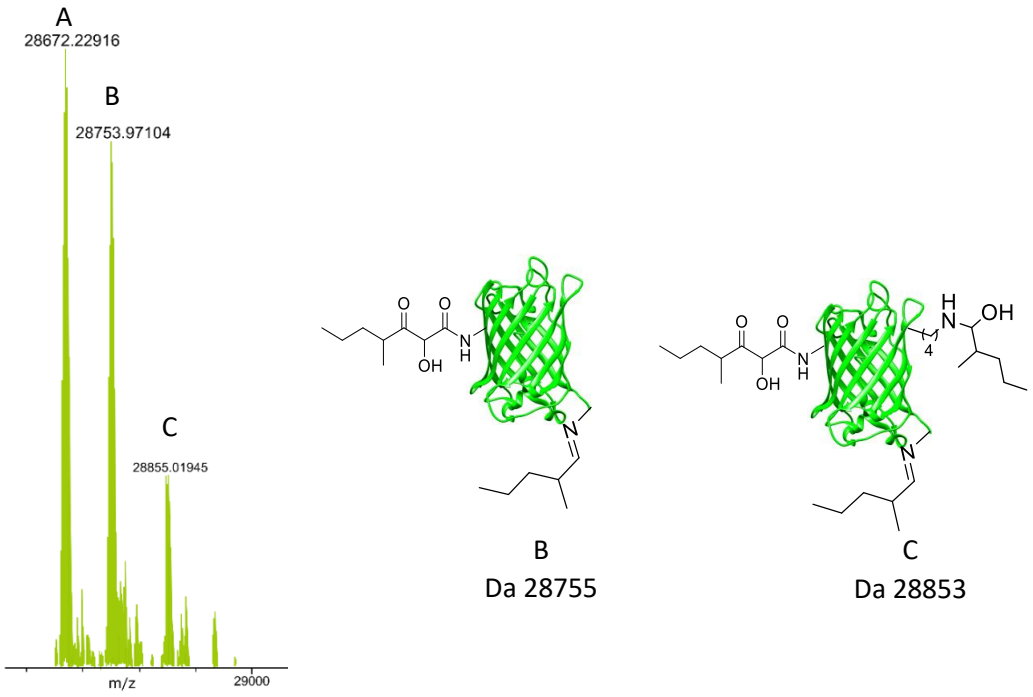


Figure 5.1: Deconvoluted mass spectrum after overnight dialysis showing A- Product **66**, B- **66** + imine addition, C **66** + imine + hemi-aminal addition.

Imine reversibility and equilibriums are well studied within the literature[285-288] and it became apparent that a reagent would need to be added in order to shift this equilibrium and allow hydrolysis of the imine. For this, the experiment was repeated, and glycine was added to the protein during dialysis. Glycine is also less likely to be protonated at pH 7 and therefore a strong nucleophile. It is hypothesised that the glycine traps the aldehyde before it can react with the *N*-terminus and is dialysed out of the system. Following an experimental repeat and dialysis with glycine at 4 °C, complete hydrolysis of all hemi-aminals and *N*-terminal imines (Scheme 5.2) was witnessed.

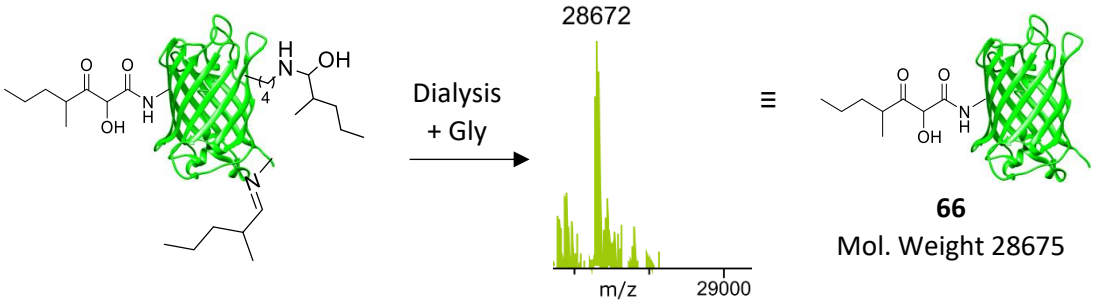
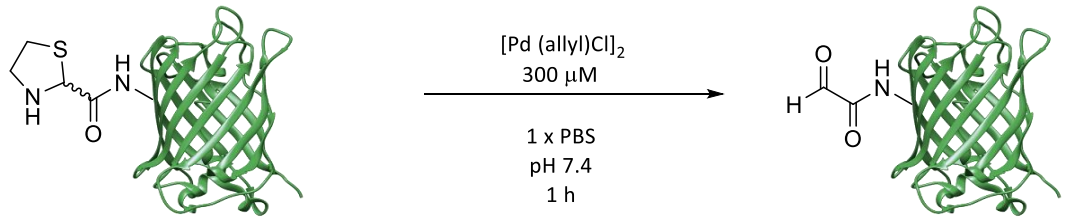


Figure 5.2: Deconvoluted mass spectrum after overnight dialysis with glycine addition yielding Umpolung product **66**

Following this result, an experiment was conducted utilising GFPY39 α-oxo K **65** and 2-methylpentanal **62** with glycine added into the reaction mixture. The motivation for this was to attempt to avoid the necessary dialysis post reaction. This experiment however did not go to completion and non-specific attack remained. Therefore, for this methodology to be successfully employed for selective bioconjugation, glycine must be added after the reaction.

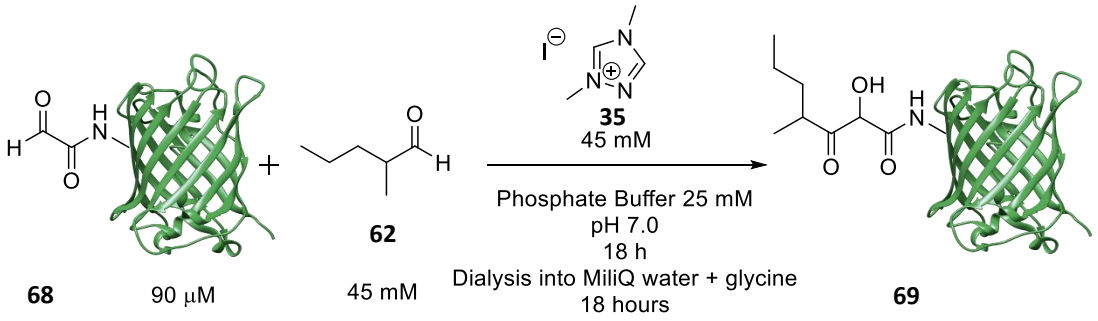
5.1.2 Trypsin Digest Experiments

To confirm that the observed modification was occurring site specifically at the desired position and not a result of non-selective off-site attack, a proteolytic digest using trypsin followed by tandem LC-MS was performed. Firstly, an internal α-oxo aldehyde was inserted into superfolder green fluorescent protein sfGFP. This was achieved through transformation of a mutated sfGFPN150 bearing an amber stop codon within its plasmid sequence into *E. coli* Top10 cells. Again the plasmid bearing the mutated plasmid carrying a sfGFPYN150 was co-transformed with a plasmid containing a pyrrolysine (Pyl) tRNACUA/pyrrolysyl-tRNA synthetase (RS) pair and doped with ThzK. Upon purification, the sfGFPN150Thz **67** was treated with allylpalladium(II) chloride dimer to generate the sfGFPN150 α-oxo K **68** (Scheme 5.3).



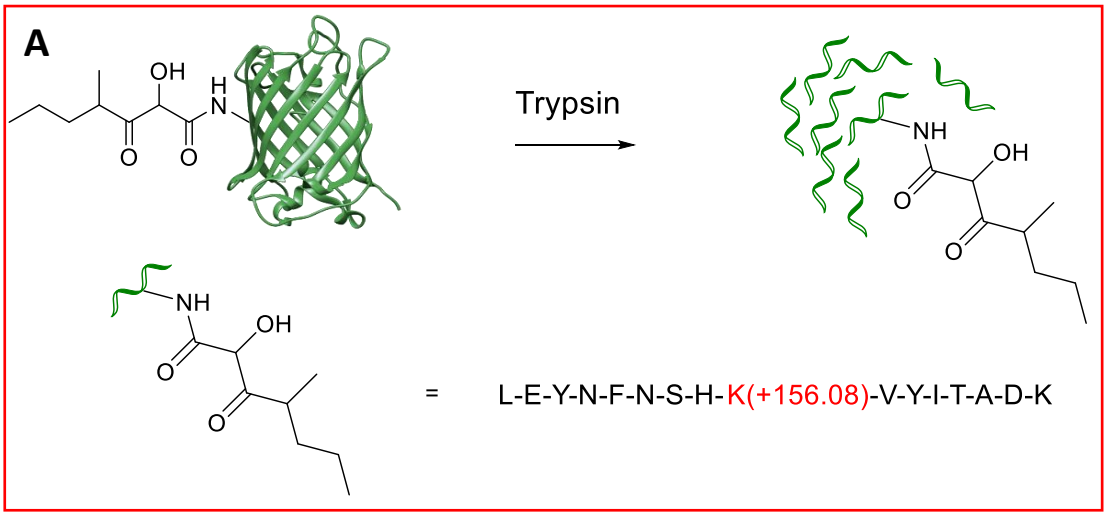
Scheme 5.3: Treatment of UAA containing sfGFPN150 ThzK **67** with Pd to unmask α-oxo aldehyde

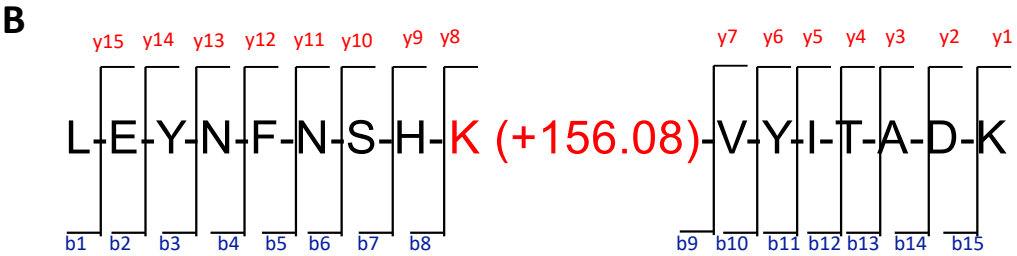
Following the generation of sfGFPN150 α-oxo K **68**, 90 μM was subjected to 500 equivalence of dimethyl triazolium catalyst **35** and 2-methylpentanal **62** in 1 x PBS pH 7.4, at 37 °C overnight. Once incubated, the mixture was dialysed into MiliQ water with glycine at 4 °C for 18 hours to give quantitative conversion to expected Umpolung product **69** (Scheme 5.4).



Scheme 5.4: Conditions employed for the modification of sfGFP(N150) **69** with dimethyl triazolium **35** and 2-methylpentanal **62**

Trypsin is a serine protease that cleaves proteins at the carboxyl end of lysine and arginine residues (unless they are followed by a proline), generating smaller peptide units. With the knowledge of the amino acid sequence of sfGFPN150 α-oxo K **68**, peptide units generated from a trypsin digest of sfGFPN150 α-oxo K **68** can be predicted, based on the position of lysine and arginine residues known within the proteins sequence. One such generated peptide from this cleavage, of the 28 possible, is the peptide LEYNFNSHK(+55.99)VYITADK, where 'K(+55.99)' indicates the position of the α-oxo aldehyde moiety and mass difference (K+ 55.99 Da). Similarly, a trypsin digest of modified sfGFPN150 **69** would give the peptide unit LEYNFNSHK(+156.08)VYITADK, where 'K(+156.08)' indicates the position of the Umpolung modification and its expected mass difference (K+ 156.078644 Da). MS/MS analysis of the tryptic peptides are consistent with modification seen exclusively at position 150 (Figure 5.2). The major fragment observed from the sequence LEYNFNSHK(+156.08)VYITADK was 2080.04 *m/z* corresponding to the mass of the major b fragment. Several fragment masses corresponding to expected breakdown were found within 0.02 Da error. Notably, y fragments 1-7 were recorded in addition to fragments 10 and 11 $[M\text{-}NH_3]^+$. Furthermore, expected b fragments 2-5 and 8 were also recorded. This data not only demonstrates a single site modification but that it was specifically the site of interest that was been modified. All other tryptic peptide fragments generated were also subjected to MS/MS confirming no off-site modification had occurred.





#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	114.09	96.08	97.06	57.55	L					16
2	243.13	225.12	226.12	122.07	E	1984.96	1966.95	1967.94	992.98	15
3	406.20	388.19	389.17	203.60	Y	1855.92	1837.90	1838.90	928.46	14
4	520.24	502.23	503.21	260.62	N	1692.86	1674.84	1675.83	846.93	13
5	667.31	649.30	650.28	334.15	F	1578.82	1560.81	1561.79	789.91	12
6	781.35	763.34	764.33	391.18	N	1431.75	1413.74	1414.74	716.37	11
7	868.38	850.37	851.36	434.69	S	1317.70	1299.69	1300.69	659.35	10
8	1005.44	987.43	988.43	503.22	H	1230.67	1212.66	1213.65	615.84	9
9	1289.62	1271.60	1272.61	645.31	K(+156.08)	1093.61	1075.60	1076.61	547.31	8
10	1388.68	1370.67	1371.67	694.84	V	809.44	791.43	792.41	405.22	7
11	1551.75	1533.74	1534.72	776.37	Y	710.37	692.36	693.34	355.69	6
12	1664.83	1646.82	1647.81	832.92	I	547.31	529.30	530.28	274.15	5
13	1765.88	1747.87	1748.85	883.44	T	434.22	416.21	417.20	217.61	4
14	1836.92	1818.91	1819.89	918.96	A	333.18	315.17	316.15	167.09	3
15	1951.94	1933.93	1934.92	976.47	D	262.14	244.13	245.11	131.57	2
16					K	147.11	129.10	130.09	74.06	1

Figure 5.2 A) Schematic of trypsin digest of modified sfGFPN150 **69** generating peptide fragments, sequence of theoretical fragment generated containing expected Umpolung product; B) Expected peptide fragment breakdown, MS/MS on expected masses found, a list of all possible MS/MS breakdown masses with found masses within 0.02 error bar highlighted (blue representing b fragments, red representing y fragments).

Unexpectedly, unmodified sfGFPN150 α-oxo K **68** was also detected (usually not witnessed) which offered some insight into the conversion levels of this chemistry. The relative intensities of sfGFPN150 α-oxo K **68** (Glyoxyl(ald)) witnessed were compared against modified sfGFPN150 **69** (Dicarbonyl) (Figure 5.3) to estimate a conversion >98%. It is assumed this level of conversion is similar to all protein modification experiments performed employing this chemistry however often not witnessed owing to the disparity between signal ratios.

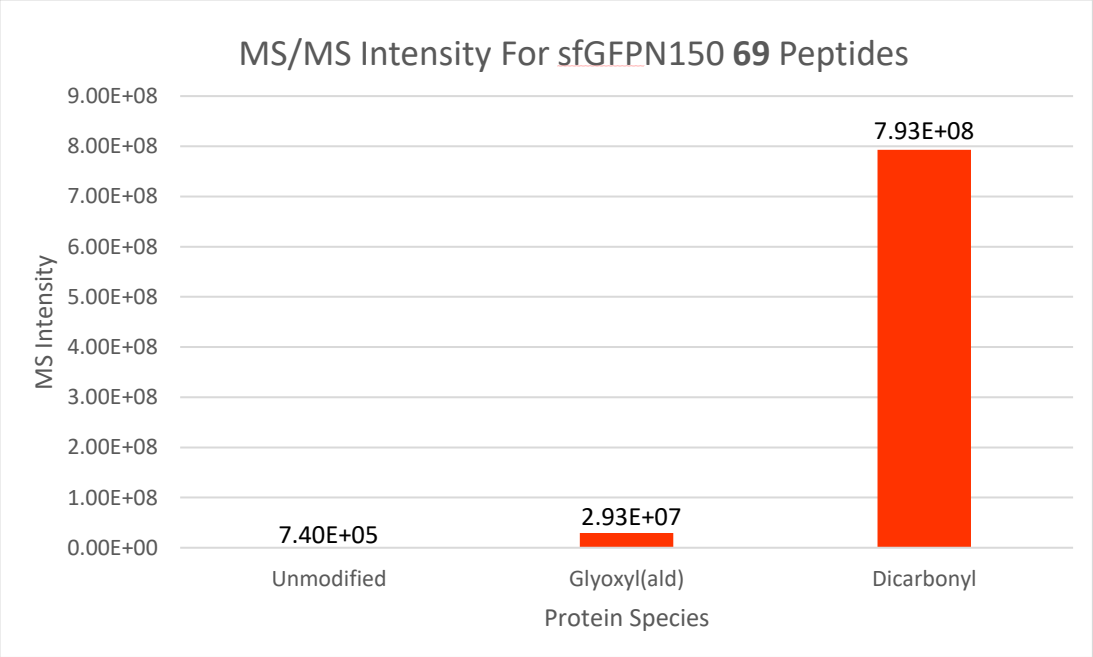
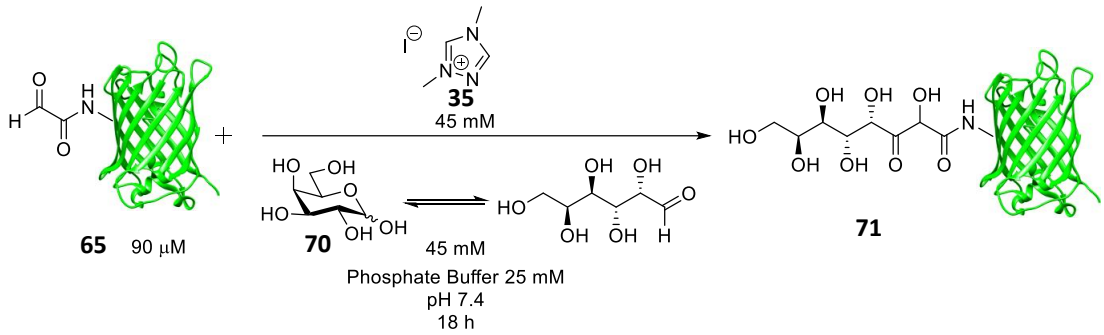


Figure 5.3: A graph comparing the relative intensities of tryptic peptides generated from modified sfGFPN150 **69** containing unmodified sfGFPN150 α-oxo K **68**

5.1.3 Protein Glycosylation

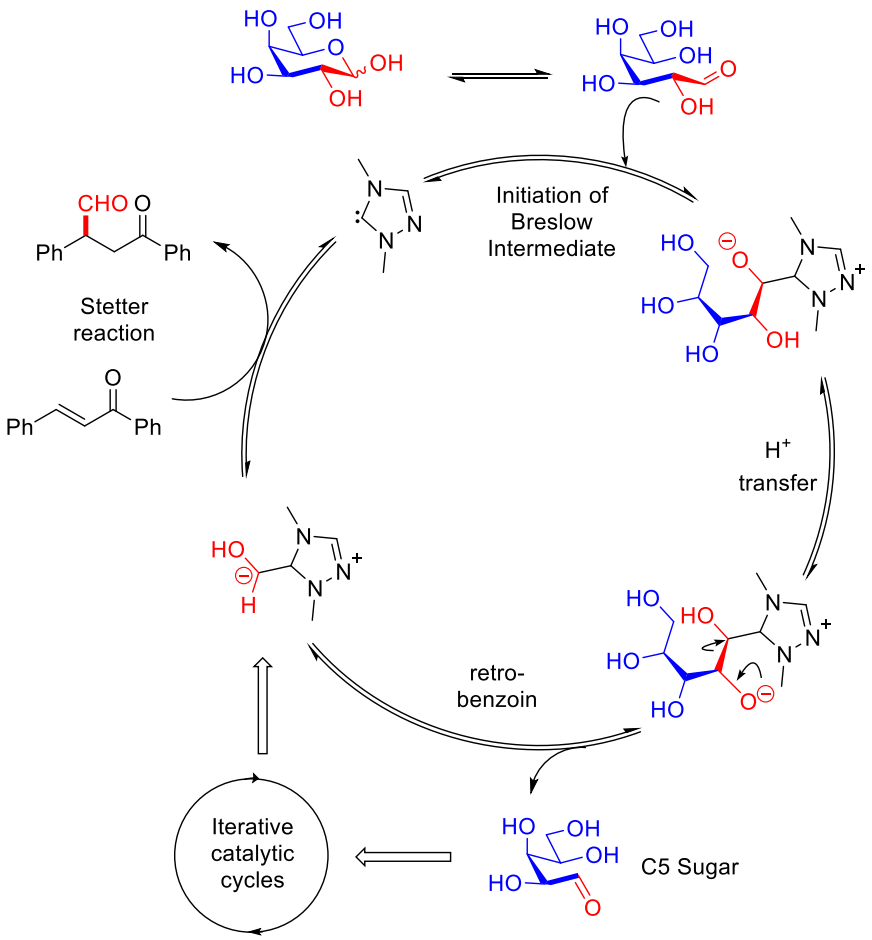
Protein glycosylation is the attachment of sugar moieties to proteins and is one of the most common posttranslational modifications (PTMs) found in all domains of life.[289] This PTM is crucial for a diverse range of processes including: aiding in solubility,[290] cell-to-cell adhesion,[291] cell-ligand interaction,[292] protein folding[293] and many more.[294] Having demonstrated the selectivity and specificity of Umpolung chemistry on proteins, it was hypothesised this chemistry may well be applicable with reducing sugars, such as galactose **70** to modify proteins of interest. To test this, GFPY39 α-oxo K **65** at 90 µM was reacted with 500 equivalents of dimethyl triazolium catalyst **35** and galactose **70** in 25 mM PB pH 7.4, at 37 °C overnight (Scheme 5.6). Following this, the mixture was dialysed into MiliQ water with glycine at 4 °C for 18 hours.



Scheme 5.6: Conditions employed for the modification of α-oxo aldehyde GFPY39 **65** with dimethyl triazolium catalyst **35** and galactose **70**

Initial MS analysis of this reaction revealed minor product formation (~8%) and seemingly multiple non-specific additions. The *m/z* values recorded did not match expected hemi-aminal nor imine products of galactose **70** however, leading to the tentative assumption that competing pathways were present.

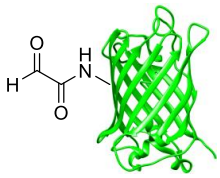
Work by Zhang *et al* offered insight into the result of the reaction when GFPY39 GFPY39 α-oxo K **65** is subjected to dimethyl triazolium catalyst **35** and galactose **70**. In their work, utilising glucose, they report the effective generation of acyl anion Breslow Intermediates through a retro-benzoin pathway (Scheme 5.7), which react with Michael acceptors for subsequent Stetter reactions.[295]



Scheme 5.7: Zhang *et al* demonstrating the pathway of carbohydrate activation to generate formaldehyde equivalents.

Using the data from work by Zhang *et al*[295] and the data generated from LC-MS experiments of α-oxo aldehyde GFPY39 **35**, subjected to dimethyl triazolium catalyst **35** and galactose **70**, it was tentatively proposed that a retro-benzoin-like pathway was occurring generating 6 new reactive aldehyde species *in situ* leading to multiple non-specific ligations of varying carbon length (Figure 5.4). Present in the LC-MS data are *m/z* peaks corresponding to masses for unreacted α-oxo aldehyde GFPY39 **65** 28571 Da and minor expected Umpolung product GFPY39 **71**. Other notable peaks were present however were difficult to clarify what they represented and ultimately could not be fully analysed.

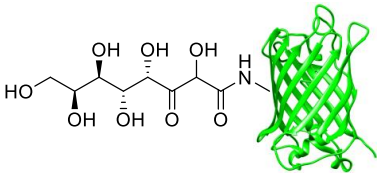
A



GFPY39 α-oxo K **65**

Calc. Mass: 28575

B



Expected Umpolung product **71**

Calc. Mass: 28755

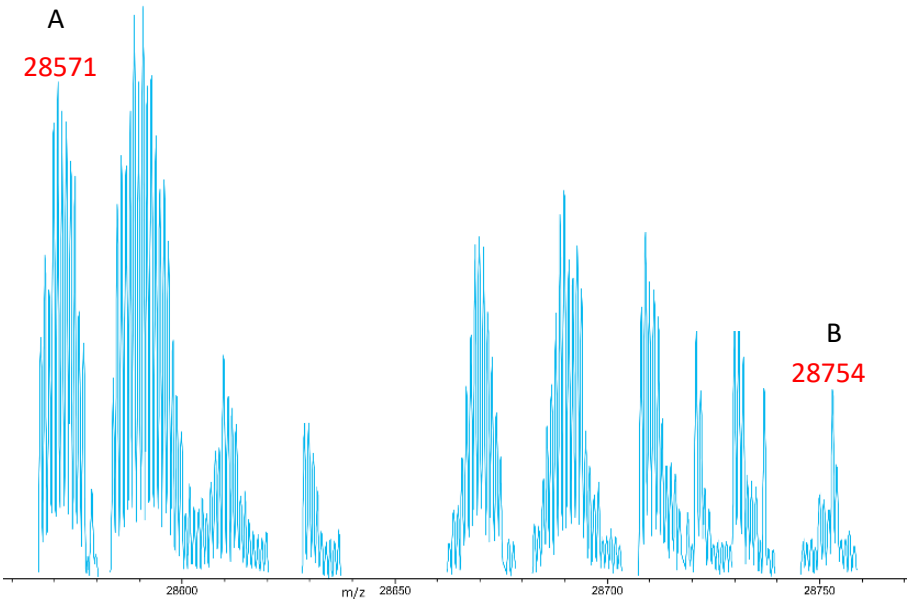
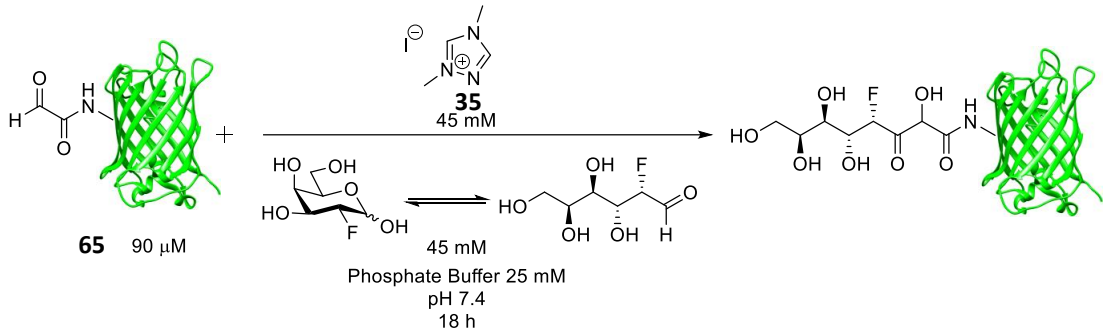


Figure 5.4: Mass spectrum and structures for suspected isolated breakdown products A) α-oxo aldehyde GFPY39 **65**B) Oxidised/sodiated α-oxo aldehyde GFPY39 **72** B) minor expected product **71**

Finally, in an effort to circumvent the suspected retro-benzoin-like pathway, it was theorised that a reducing sugar unable to participate in the intramolecular proton transfer step may block the retro-benzoin pathway, be unable to generate the acyl anion equivalent and thus preclude the competing pathway. Gal-2-F was selected owing to its availability and possessing the properties outlined above, chiefly a fluorine in place of a hydroxy moiety incapable of participating in the intramolecular proton transfer step. To investigate this, GFPY39 α-oxo K **65** at 90 μM was reacted with 500 equivalents of dimethyl triazolium catalyst **35** and gal-2-F in 25 mM PB pH 7.4, at 37 °C overnight (Scheme 5.8). Following this, the mixture was dialysed into MiliQ water with glycine at 4 °C for 18 hours.



Scheme 5.8: Conditions employed to preclude the retro-benzoin-like pathway *via* modification of GFPY39 α-oxo K **65** with dimethyl triazolium catalyst **35** and gal-2-F.

LC-MS data from this experiment revealed no modification had occurred; the data were consistent with GFPY39 α-oxo K **65**, and oxidised/sodiated starting material. It remains unclear if this is a result of the presence of fluorine and its unique reactivity[296] and if glycosylation could remain possible with an alternative heteroatom in the 2 position whilst still blocking the intramolecular proton transfer step, such as a methoxy moiety.

5.2 Conclusions and Future Work

Work detailed in this section describes the successful implementation of the NHC- catalysed acyloin condensation bioconjugation reaction on internal protein sites. The chapter describes the co-transformation of mutated green fluorescent protein containing an amber stop codon with a plasmid containing a pyrrolysine (Pyl) tRNACUA/pyrrolysyl-tRNA synthetase (RS) pair and a UAA. This purified UAA containing GFP was then treated with allylpalladium chloride dimer to generate α-oxo aldehyde GFPY39 **65**. Applying conditions for Umpolung modification developed on *N*-terminal α-oxo aldehydes (250 eq NHC catalyst and 250 eq aldehyde donor) resulted in 52% yield. In an attempt to improve conversion, the equivalences of aldehyde donor and NHC catalyst were doubled to 500. This increase resulted in quantitative conversion to Umpolung product and also the formation of hemi-aminal bonds between aldehyde donor and surface exposed Lys residues. Additionally, these conditions resulted in the formation of an imine moiety, likely to be on the *N*-terminus. These additions were not able to be removed through dialysis alone. Fortuitously, the addition of Gly into the dialysis procedure was able to cleave and remove the unwanted non-specific additions that had occurred.

Also described in this section is the conformation that this strategy introduces a single C-C site specific and site selective modification at the desired residue. This was demonstrated using a different green fluorescent protein, superfolder green fluorescent protein, and was proved using trypsin digest experiments. As before, superfolder green fluorescent protein containing an amber stop codon was co-transformed with a plasmid containing a pyrrolysine (Pyl) tRNACUA/pyrrolysyl-tRNA synthetase (RS) pair and a UAA. Similarly, this was treated with allylpalladium chloride dimer to generate α-oxo aldehyde sfGFPN150 **68**. Using conditions developed on GFPY39 α-oxo K **65**, sfGFPN150 α-oxo K **68** was modified with 2-methylpentanal **62** to quantitative yield, as judged by LC-MS. This modified sfGFPN150 **68** was then subjected to a trypsin digest experiment and subsequent MS/MS analysis of the tryptic peptides produced. The data generated from this experiment agree with the expected results of a single modification at the desired residue, post dialysis and no off-site attack witnessed.

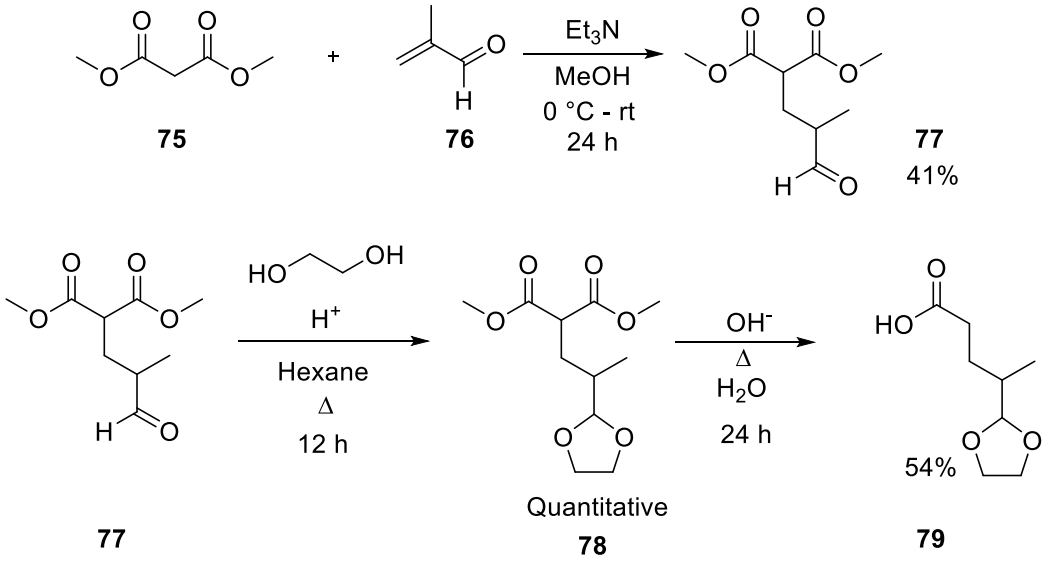
Finally, it was investigated if the methodology described in this chapter could be employed for the glycosylation of proteins. For this investigation, galactose **70** was employed to probe the possibility of glycosylating proteins using the Umpolung methodology described in this

thesis with α-oxo aldehyde GFPY39 **65**. Unfortunately, this resulted in what is tentatively assumed to be the uncontrolled iterative breakdown of galactose **70** and unwanted addition of breakdown products, based on similar results observed in the literature.[295] To block this iterative breakdown pathway, Gal-2-F was employed, however no evidence of successful ligation of Umpolung product was witnessed.

Chapter 6: On Fluorescent probes, connectivity and modelling branched vs linear aldehyde donors

6.1.1 Developing a fluorescent probe

Having established conditions required for the NHC mediated Umpolung ligation chemistry as a method for protein modification, the strategy was next investigated for its ability to modify a target protein with a detectable chemical reporter. For this, it was envisioned that an efficient Fmoc-SPPS approach could be advantageous for facile probe development. This required the introduction of a carboxylic acid moiety into the probe design. To synthesise a viable target molecule, a 3 step synthesis was executed (Scheme 6.1), starting with the triethylamine catalysed Michael addition of methacrolein **76** to dimethyl malonate **75** to generate racemic aldehyde **77**. Whilst asymmetric Michael additions can be performed (such as use of quaternary ammonium salts or organocatalysis) it is of little difference when applied to protein bioconjugation, where stereochemistry is often unimportant and typically does affect protein or reporter functionality. Subsequent protection of the aldehyde group initially proved challenging.



Scheme 6.1: Synthetic route to SPPS compatible probe **79**

Efforts to protect racemic aldehyde **77** first looked to employ methanol and camphorsulfonic acid as catalyst, however the expected product was lost to the polar phase post reaction during separation. A second attempt looking to preclude this by employing ethanol with camphorsulfonic acid proved just as futile. Finally, ethylene glycol was employed with catalytic camphorsulfonic acid to generate the dioxolane protected aldehyde **78**. Once generated, a base promoted ester hydrolysis and subsequent decarboxylation reaction was performed to generate the desired compound **79**.

With a suitable aldehyde handle, possessing the necessary features established in the small molecule ligation development (alkyl chain with methyl substitution at α-carbon), construction of fluorescent probe **80** was attempted incorporating a dansyl moiety[2] as its fluorescent source (Figure 6.1). This was built using Fmoc SPPS methodology, employing H-Gly-2-ClTrt resin as the starting unit. To ensure the dansyl group did not hinder the aldehyde ligation, 2 PEG spacer units were also inserted. This created suitable distance between the probe and the reactive aldehyde handle. This PEG group also aids solubility of the probe ensuring it remains soluble in aqueous solution.

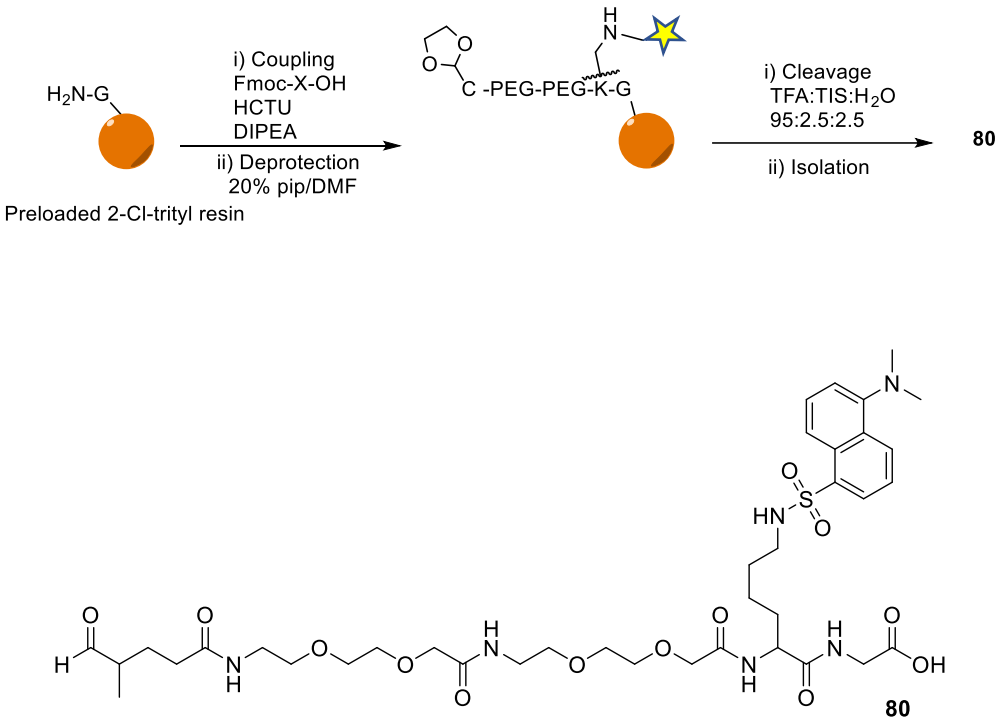


Figure 6.1: Fmoc synthesis of fluorophore probe **80**; Structure of fluorophore probe **80** designed for NHC mediated Umpolung protein ligation

When analysed using LC-MS, the *m/z* value recorded was not the value calculated. It appeared that the terminal aldehyde had cyclised to form **81** (Figure 6.2). This cyclisation to **81** renders the probe unable to partake in ligation as it no longer contains the aldehyde moiety required.

[2] Dansyl Lys provided by Dr Robin Brabham

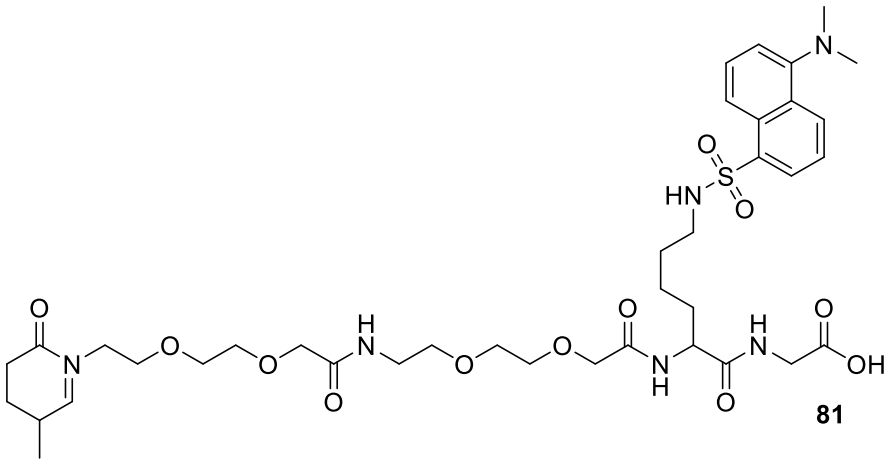


Figure 6.2: Suspected cyclised product **81** from LC-MS analysis

To circumvent this cyclisation, 2 distinct changes were made to the design of the fluorescent probe. Firstly, a secondary amine was introduced in the form of Fmoc-piperidine-4-carboxylic acid (Fmoc INP) before the final coupling of the aldehyde handle, generating fluorescent probe **82**. This aimed to prevent the cyclisation step from occurring. Secondly the cleavage of the peptide from the resin typically requires TFA, ergo unmasking the acid labile protected aldehyde and so was exchanged for a milder method with HFIP which would retain the protected aldehyde (Figure 6.4).[297]

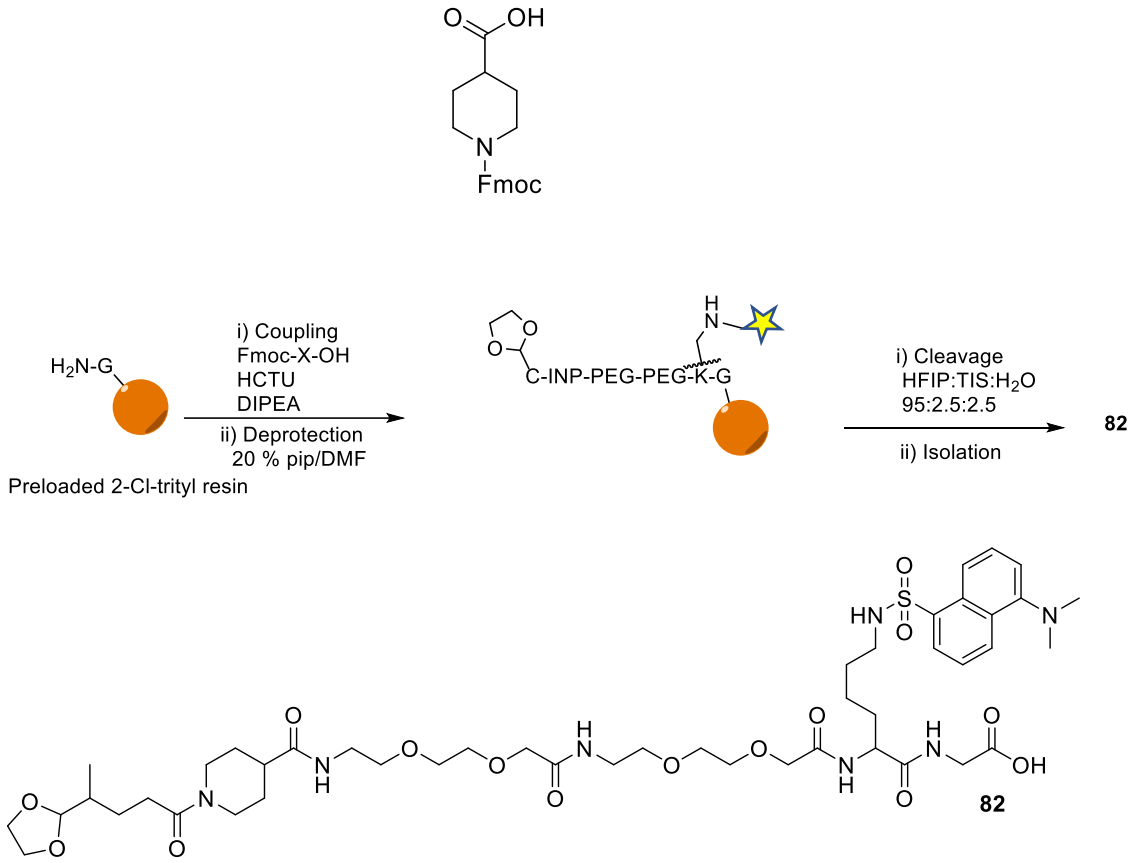
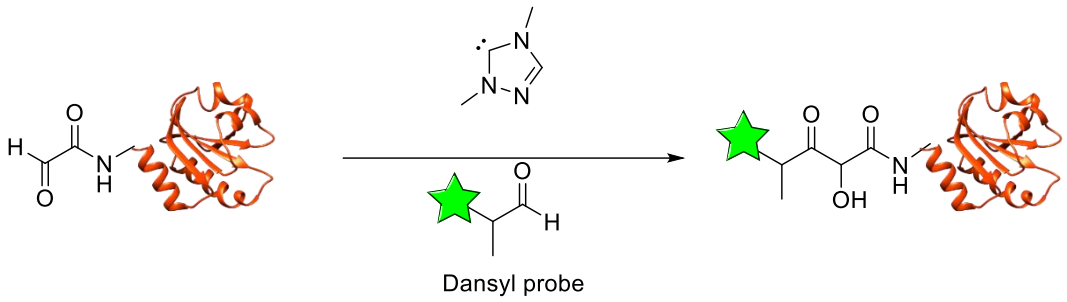


Figure 6.4: Structure of Fmoc-INP; Fmoc SPPS with HFIP resin cleavage: Structure of secondary amine containing fluorophore reporter **82**, retaining the aldehyde protecting group to prevent cyclisation.

To test the applications of the NHC mediated Umpolung ligation strategy, 5 mg fluorescent probe **82** in 140 µL of 25 mM PB pH 7.4 in a 1 mL Eppendorf tube was treated with neat TFA (10 µL) and the dioxolane deprotection monitored *via* LC-MS. Following completion, 0.5 mL 10% AcOH was then added and the solution was lyophilised.

The unmasked aldehyde reporter **82** was then subjected to 44 µM α-oxo aldehyde thioredoxin **58** and 11 mM dimethyl triazolium **35** in 25 mM phosphate buffer at pH 7.4 at 37 °C overnight. A control experiment was also conducted in tandem under identical conditions without the presence of dimethyl triazolium catalyst **35**. Following this reaction, the protein ligation experiment was purified *via* molecular weight cut-off (MWCO) filter followed by 24 h dialysis. The purified protein was then subjected to SDS polyacrylamide gel electrophoresis and visualised using Coomassie Brilliant Blue anionic dye. This was then

exposed to UV light to visualise if any fluorescent probe had ligated to the protein (Figure 6.5).



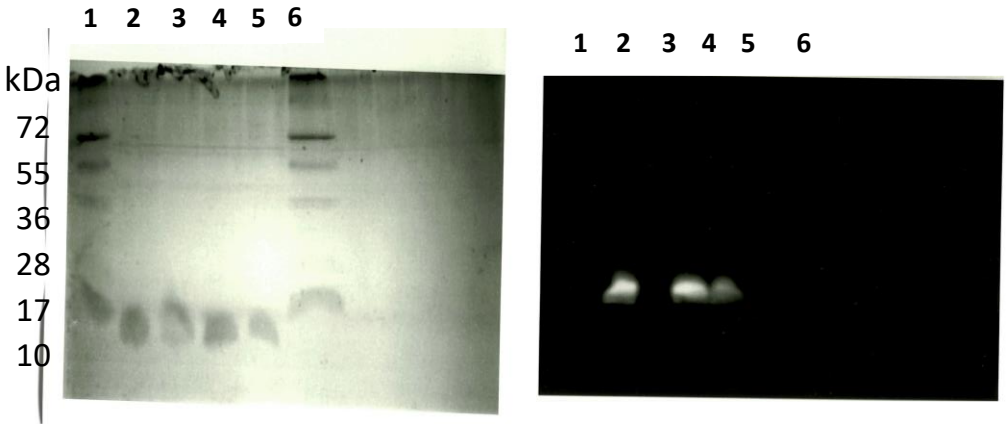
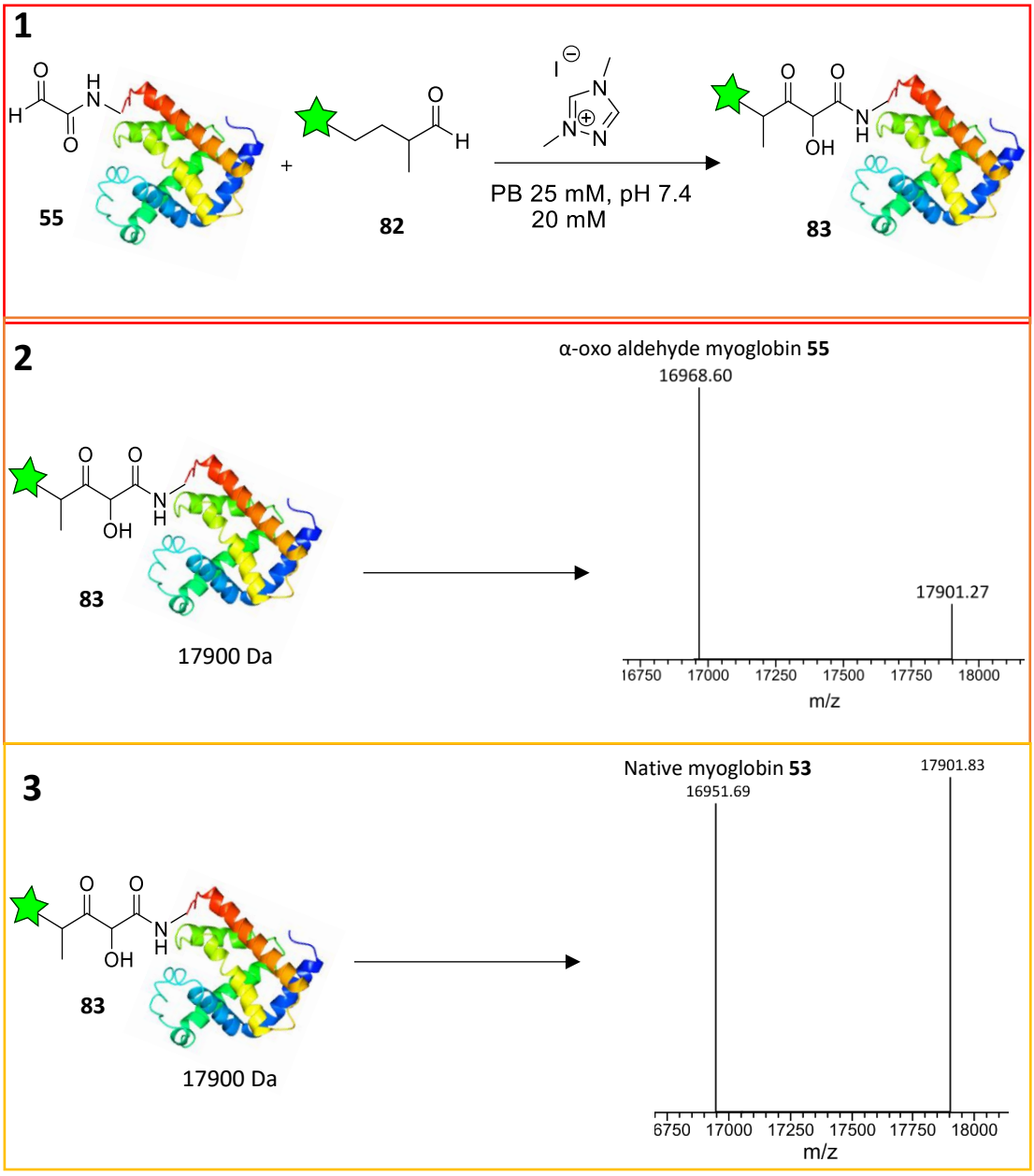


Figure 6.5: Reaction conditions for the fluorescent labelling of α-oxo aldehyde thioredoxin **58**; Polyacrylamide gel; lane 1 = Protein standard ladder, lane 2 = Thioredoxin reaction aliquot, 250 eq, lane 3 = Control experiment, lane 4 = Thioredoxin reaction aliquot, 180 eq, lane 5 = Thioredoxin reaction aliquot, 120 eq protein standard ladder, lane 6 = protein standard ladder.

The results of this experiment suggest successful modification of α-oxo aldehyde thioredoxin **58** through NHC mediated Umpolung catalysis. The negative control without the required dimethyl triazolium catalyst **35** (Figure 6.5, lane **4**) is further validation of the success of this chemistry and confirms the catalyst dependant nature of the reaction. LC-MS analysis of the major product was not possible despite multiple attempts. This difficulty is attributed to closeness of the mass of the protein (~11 KDa) and the parameters of the MWCO purification filters (10 KDa).

For access to LC-MS analysis, a larger protein was selected. For this 100 µM α-oxo aldehyde myoglobin **55** was subjected to 25 mM fluorescent probe **82**, following TFA treatment, and 25 mM dimethyl triazolium catalyst **35** in 25 mM phosphate buffer pH 7.4 and left at 37 °C

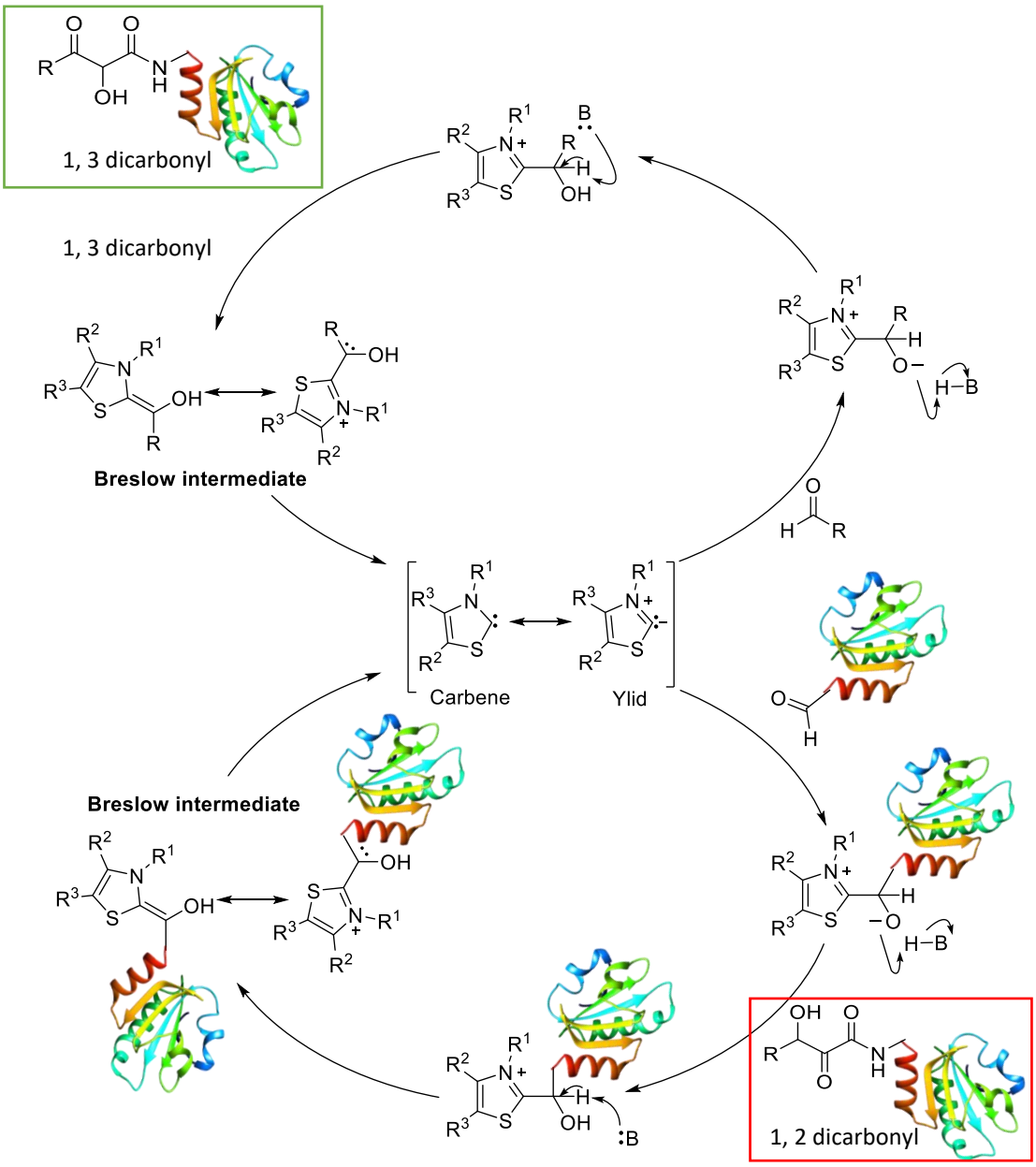
overnight. Following this reaction, the protein ligation experiment was subjected to purification *via* molecular weight cut-off (MWCO) filter followed by 24 h dialysis (Scheme 6.2). This resulted in approximately 17% conversion to expected Umpolung product **83** as judged by LC-MS. This unexpected result prompted an experiment repeat with increased concentrations of catalyst and probe from 250 to 500 equivalence, whereby 100 µM α-oxo aldehyde myoglobin **55** was subjected to 50 mM fluorescent probe **82** and 50 mM dimethyl triazolium catalyst **35** in 25 mM phosphate buffer pH 7.4 and left at 37 °C overnight. Following this reaction, the protein ligation experiment was subjected to purification *via* molecular weight cut-off (MWCO) filter followed by 24 h dialysis. LC-MS analysis appeared to indicate these conditions were successful for quantitative conversion to modified protein.



Scheme 6.2: **1:** Outline of general conditions for Umpolung modification of α-oxo aldehyde myoglobin **55**; **2**: Deconvoluted LC-MS data indicating ~17% conversion when using 250 equivalences of fluorescent probe **82** and NHC catalyst **35**; **3**: Deconvoluted LC-MS data indicating quantitative conversion when using 500 equivalences of fluorescent probe **82** and NHC catalyst **35**

6.2.1 Comments on connectivity

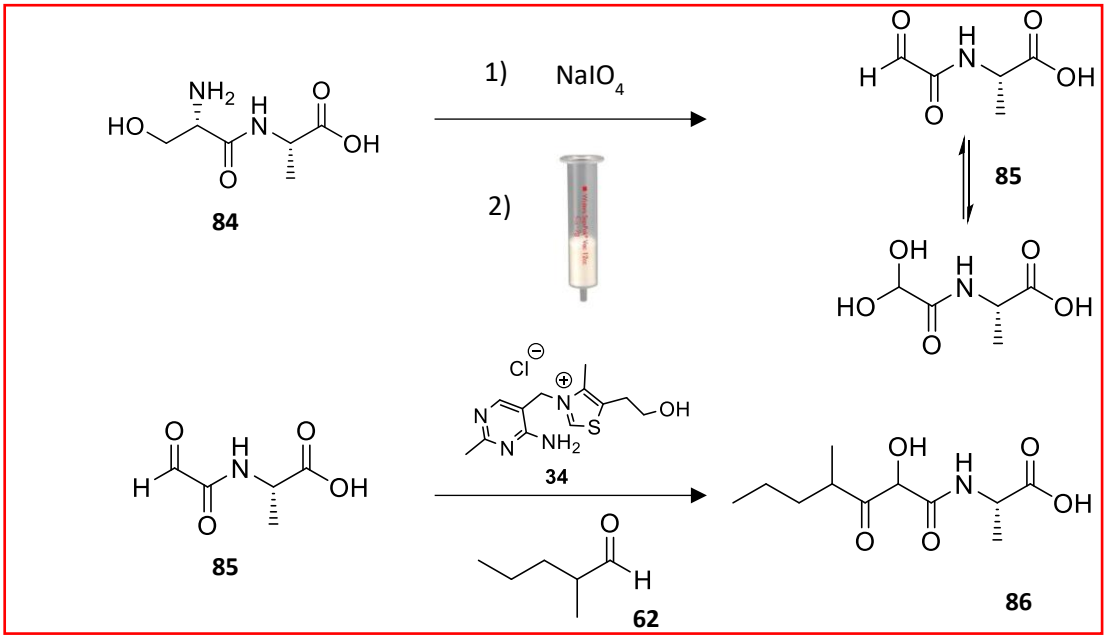
Connectivity in organic chemistry is often critical to methodology development. For this Umpolung bioconjugate chemistry it was decided that the connectivity of protein bioconjugate products was of interest, given that the NHC catalysts employed could, in theory, generate a Breslow intermediate with one or both of the two aldehydes present, the donor aldehyde and the protein α-oxo aldehyde resulting in different products being generated, either a 1, 3 dicarbonyl link or 1, 2 dicarbonyl link respectively (scheme 6.3).



Scheme 6.3: The two possible Breslow Intermediates that could be generated in the Umpolung bioconjugate modification between NHC catalyst with either aldehyde donor (top) to yield a 1,3- dicarbonyl product (green box) or protein α-oxo aldehyde (bottom) to yield a 1,2- dicarbonyl product (red box)

It is tentatively assumed that the Umpolung product generated is a 1,3- dicarbonyl from a Breslow Intermediate generated from the NHC catalyst reaction with aldehyde donor. This assumption is derived from the knowledge gained during the Stetter investigation described in chapter 2. In this instance, only one aldehyde is present for the NHC to couple with. When this process occurs and the Breslow Intermediate is formed with the protein α-oxo aldehyde, uncontrolled total oxidation was witnessed of the α-oxo aldehyde to a carboxylic acid. As this oxidation is not witnessed, it suggests that the Breslow Intermediate is formed on the donor aldehyde. Oxidation of this small molecule aldehyde may also occur but is not easily observed in LC-MS analysis due to the low molecular weight of unwanted oxidised donor aldehyde. This rationale may also explain why high equivalences of aldehyde donor are required, owing to the sacrificial nature of the aldehyde donor. To further investigate the connectivity of the C-C linkage, a Ser-Ala dipeptide **84** was oxidised to α-oxo aldehyde-Ala **85** *via* $NaIO_4$ and purified *via* a reverse phase cartridge. Following this, α-oxo aldehyde-Ala **85** was subjected to 2-methyl pentanal **62** and triazolium catalyst **35** in 25 mM phosphate buffer pH 7.4 and left at 37 °C overnight. The sample was then lyophilized and dissolved in $MeOD_4$ for subsequent NMR analysis. Unfortunately, total NMR characterisation was not possible despite repeated attempts, likely a result of incomplete conversion, tautomerisation and the small quantity of product available (~ 3 mg). Despite this however, it was speculated that a comparison of the most downfield proton on the ^1H NMR (Figure 6.6) between the proposed 1,3- dicarbonyl product **86** and the hydrated α-oxo aldehyde-Ala **85** would aid characterisation. Compounding the complexity of this study further however, it was later realised that the most downfield proton of interest is obscured under the CD_3OD solvent peak. This obscuring from the solvent peak was only discovered in later studies analysing hydrated α-oxo aldehyde-Ala **85** and increasing the temperature in order to observe the most downfield proton. [l] It is tentatively assumed on the lack of split signals upfield or downfield of the solvent peak, which would be expected to be witnessed for the most downfield proton if the configuration was a 1,2-dicarbonyl, and the lack of protein oxidation witnessed *via* LC-MS that the connectivity is likely a 1,3-dicarbonyl.

[l] Ser-Ala dipeptide **84** synthesis and characterisation was performed by Dr Darshita Budhadev (Reference 66) and NMR increased temperature experiments were suggested and performed by Dr Alex Heyam



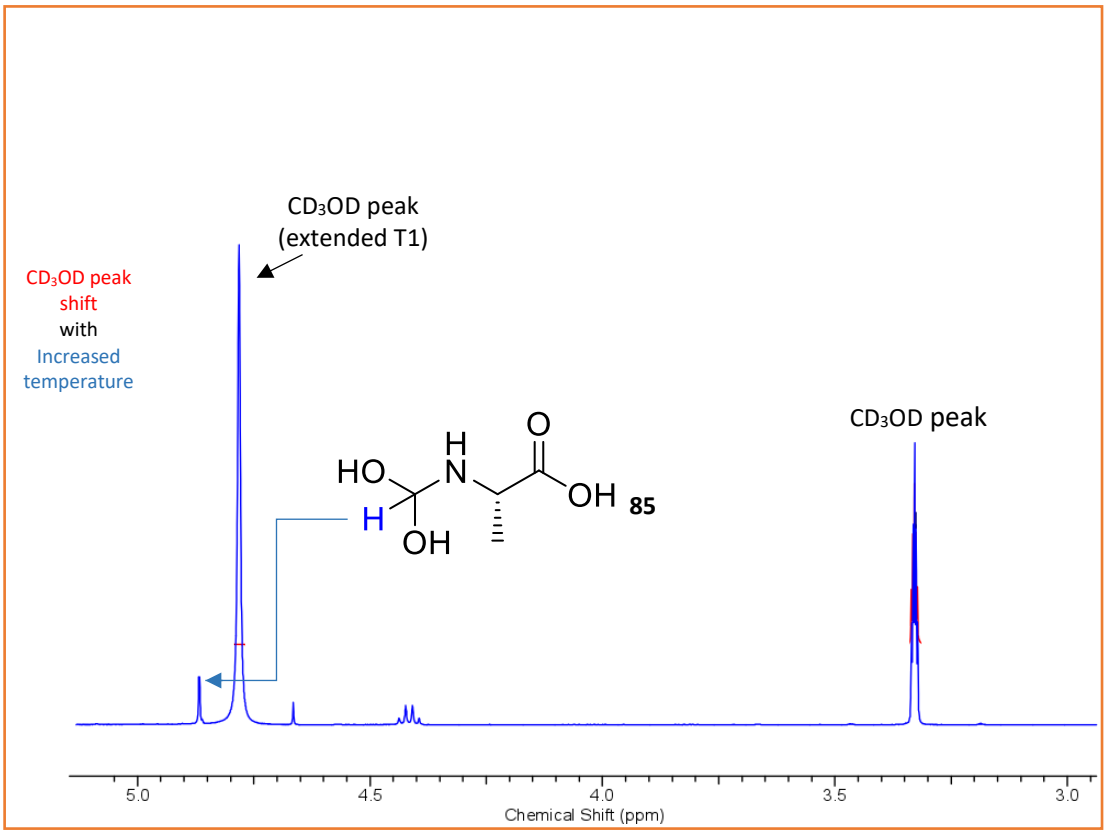
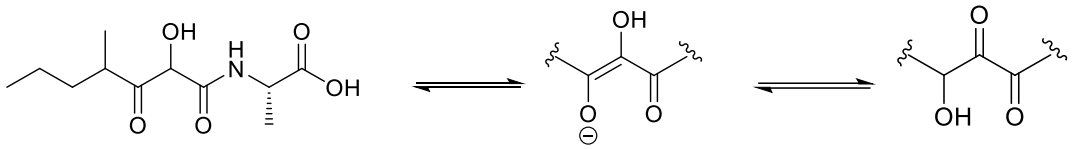


Figure 6.6: Outline of the synthesis of α-oxo aldehyde-Ala **85**; application of 2- methyl pentanal **62** and NHC catalyst for Umpolung coupling; ^{1}H NMR increased relaxation time to observe most downfield proton of the starting α-oxo aldehyde-Ala **85**

Finally, it is noted that in theory, regardless of whether the Breslow Intermediate is formed with the donor aldehyde as is hypothesised, it would still be possible to generate a mixture of 1, 2 and 1,3- dicarbonyl through enol tautomerisation (Scheme 6.4). Although this tautomerisation does not appear to be occurring on the peptide model in the NMR tube, it cannot be ruled out on protein constructs. Despite this, whether 1, 3, connectivity 1, 2 connectivity or both is occurring, what is most important is the generation of a hydrolytically stable C-C bond.



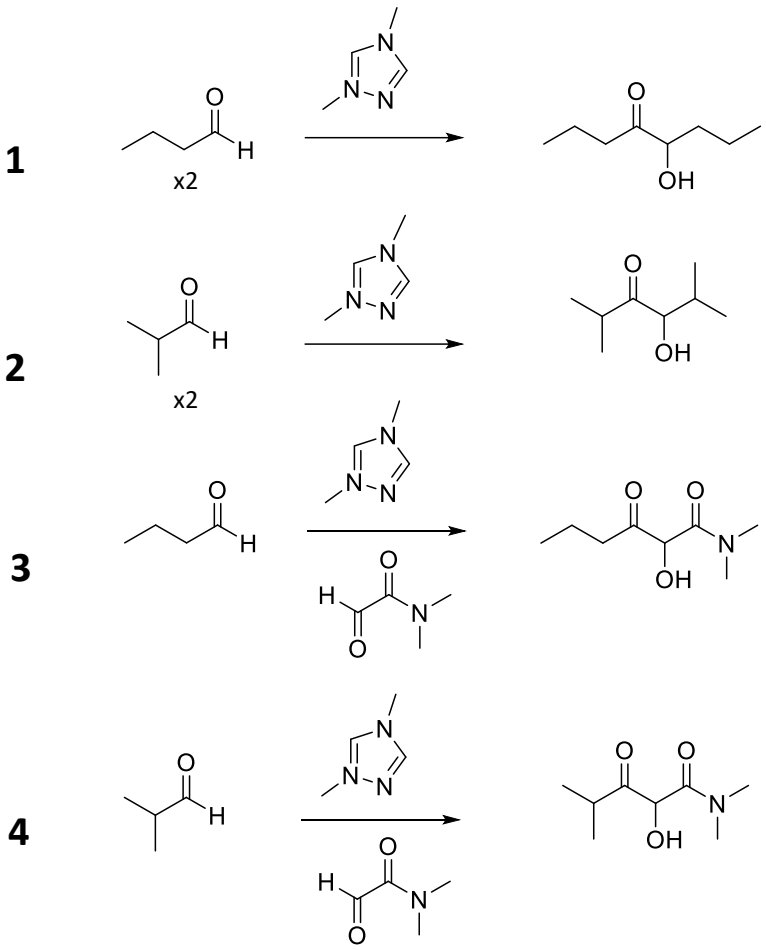
Scheme 6.4: Theoretical reversible enolisation of 1,3- dicarbonyl to generate 1, 2 diketone

6.3.1 Branched vs Linear aldehyde donors: A computational Study

During this project it was discovered that branched aliphatic aldehyde donors proved significantly more valuable in generating quantitative conversion of protein bioconjugates than linear aliphatic aldehydes. It was decided that computational modelling could be employed to explore why aliphatic branched aldehydes undergo Umpolung condensation with proteins more efficiently than aliphatic linear aldehydes. It was tentatively proposed that the increase in apparent reactivity of branched aliphatic aldehydes could be attributed to them being less likely to homo-couple than linear aliphatic aldehydes due to steric hindrance. For this study, the NHC-catalysed acyloin homo-coupling between butyraldehyde **43** and isobutyraldehyde **46** as well as the hetero-coupling of the aldehydes with a dipeptide mimic model (Scheme 6.5) were studied computationally to explore this hypothesis. Calculations were carried out in Turbomole V6.40, with geometry and frequency calculations of (RI)-B3LYP/def2-SVP. Self consistant field (SCF) energies improved at the (RI)-B3LYP-D3-TZVPP level of theory. Implicit solvation in water was considered using COSMO, ε_{water} = 80.1 at 298.15 K.

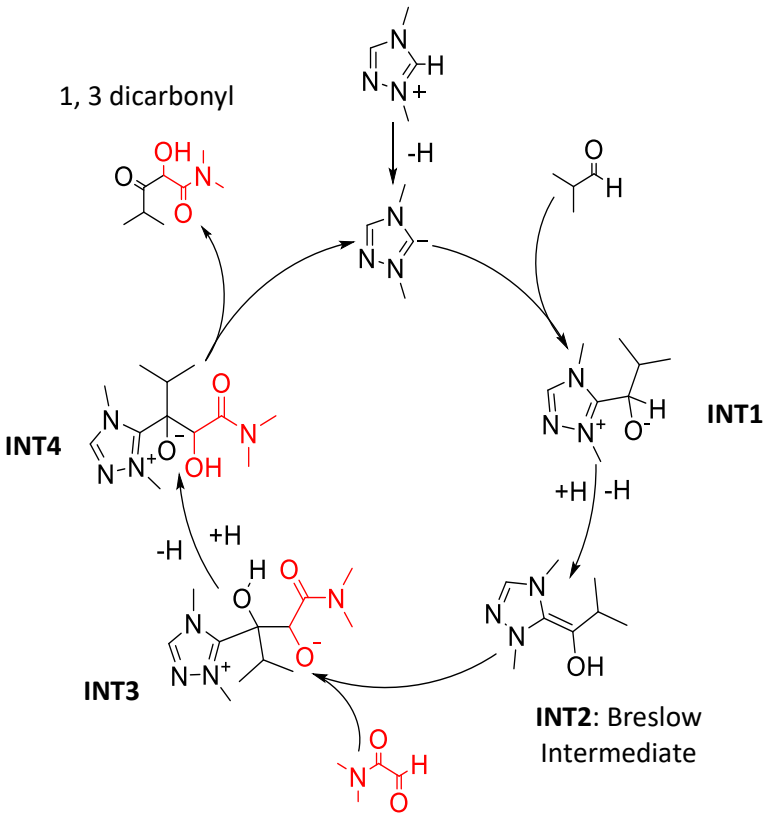
Firstly, the nucleophilic carbene catalyst attacks the carbonyl carbon of the aldehyde through transition state **TS_1**, resulting in a neutral alkoxide adduct, intermediate **INT1**. The relative free Gibbs energy of the activation barrier for this step was calculated to be 59.5 kJ/mol and 58.1 kJ/mol for the butyraldehyde **43** and isobutyraldehyde **46** respectively.

This species then, undergoes a 1,2-proton shift reaction to generate the Breslow enamine intermediate **INT2**.



Scheme 6.5: Schematic showing 1) linear aldehyde homo-coupling, 2) branched aldehyde homo-coupling, 3) linear hetero-coupling with peptide mimic and 4) branched aldehyde hetero coupling with peptide mimic

Following the formation of the Breslow Intermediate **INT2**, coupling with the dipeptide mimic *via* $\mathbf{TS_3}$ occurs to generate **INT3**. The sequence is then completed by cleavage of the catalyst from aldehyde at **INT4** (Scheme 6.6).



Scheme 6.6: Schematic showing proposed catalytic cycle with INT1-INT4 states described in the PES study

The relative free energy of the aldehyde-dipeptide-mimic coupling transition states is significantly lower than the homo-coupling transition states, with values of 55.5 and 65.2 kJ/mol for the linear and branched pathways respectively (Figure 6.7). The four benzoin condensation transition states calculated at **TS_3** provides some evidence as to why the branched aldehydes are favoured in this Umpolung reaction on proteins against linear aldehydes. In the branched reaction pathway, both the *anti* and *syn* homo-coupled transition states (Figure 6.7, blue trace) are significantly higher in energy (+ 55.6 kJ/mol above for *syn* and +55.6 kJ/mol above for *anti*) than the branched aldehyde-dipeptide-mimic coupling transition state (Figure 6.7, green trace). If accurate, this would signify that the desired pathway is energetically more favourable than the homo-coupled self-condensation pathway. Conversely the linear aldehyde *syn* homo-coupling pathway (Figure 6.7, red trace) is only 18 kJ/mol higher than the linear aldehyde-dipeptide-mimic coupling transition state **TS_3** *anti* (Figure 6.7, pink trace).

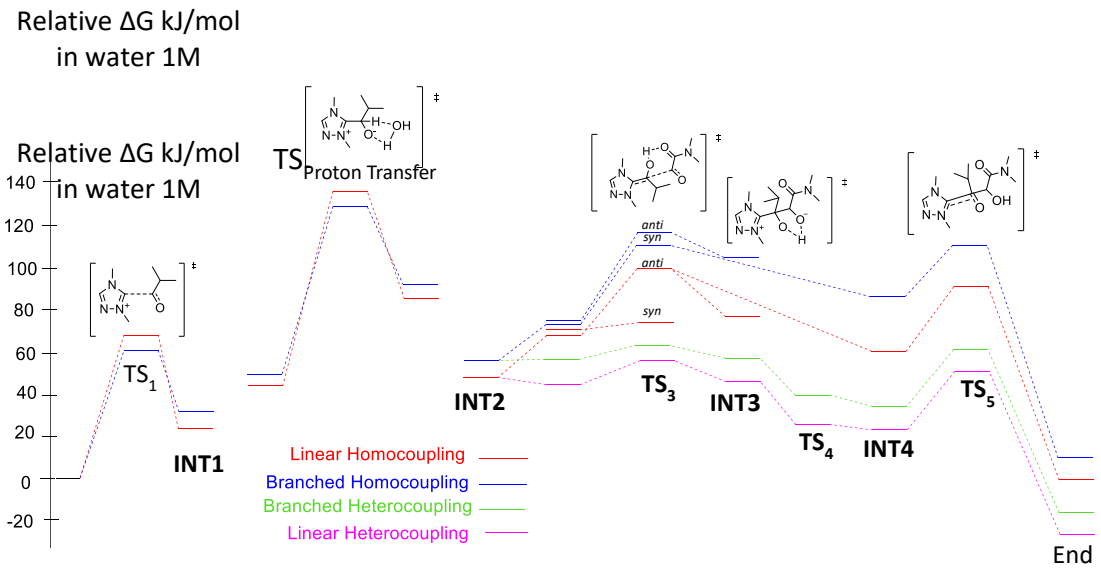


Figure 6.7: Schematic showing PES against reaction step for linear aldehyde homo-coupling (red), branched aldehyde homo-coupling (blue), branched aldehyde-dipeptide mimic coupling (green) and linear aldehyde-dipeptide mimic coupling (pink)

If accurate, this study indicates that the energy barriers for the linear alkyl aldehyde self-coupling pathway to be energetically close enough to each other at a key transition state (**TS_3**) to propose they are in competition with each other. This competition at **TS_3** in the linear aldehyde donor effectively results in less reactive intermediate available to complete the desired reaction owing to the closeness in energy values at **TS_3** of the desired aldehyde-dipeptide-mimic coupling (+55.5 kJ/mol) and the homo-coupling pathway (+74.2 kJ/mol). [4] In the branched system at **TS_3** this competition between homo coupling (+118.1 kJ/mol) and the desired hetero coupling reaction (+62.5 kJ/mol) is not seen.

[4] All computational experiments were designed and executed by Mauricio Murillo-Herrera

6.4.1 Conclusions

The work described in this chapter further confirms the potential usefulness of the Umpolung bioconjugation developed and described in this thesis. This chapter demonstrated the ability to apply this chemistry to protein probing *via* fluorescent tags. The purification and LC-MS analysis of the fluorescent tags used were complicated by the inclusion the PEG spacer units. These units are important for aiding solubility however it is likely a more facile probe could be synthesised and analysed using repeated glycine units in favour of the PEG moieties. In addition, the synthesis is also made sub-optimal owing to the requirement of the secondary amine to prevent cyclisation. Possible solutions for this draw back could be a redesigned linker, possibly a longer adipic α-oxo aldehyde, methyl substituted unit or generate the NHS ester of the final coupling unit to enable facile probe synthesis. Both of these solutions however do further complicate this methodology and risk becoming too complex for the 'plug and play' goal of novel chemical protein modification methodologies. Also described in this chapter are comments on possible connectivity of the synthesised constructs which we unable to definitively prove the true connectivity. It is also noted that the connectivity could be reversible between a 1,3- and a 1,2- dicarbonyl. Finally, the chapter describes the results from a PES study aiming to elucidate why branched alkyl aldehydes are more reactive with the protein aldehyde than linear alkyl aldehydes. This study revealed the energy barriers for the linear alkyl aldehyde self-coupling and desired-coupling pathways to be energetically close enough to each other at a key transition state to propose they are in competition with each other. This closeness of energy barriers is not witnessed in the branched aldehyde self-coupling *versus* desired-coupling pathways.

Chapter 7: Conclusions and Future directions

7.1: Summary

A novel strategy for producing protein bioconjugates is described in this thesis. The strategy takes advantage of small molecule advances in Umpolung chemistry and protein aldehydes. The stable C-C connection generated is a major advantage over hydrazone and oxime ligations which are susceptible to hydrolysis. The requirement for dialysis post reaction to remove unwanted hemi-aminal bonds, leaving on the newly generated C-C bond demonstrates this hydrolytic stability.

Methodology for a Stetter reaction on proteins was not derived owing to the presence of unwanted side reactions which oxidise the α-oxo aldehyde of interest.

The investigation of the acyloin condensation circumvented the unwanted oxidation problems discovered in the Stetter trials.

A range of catalysts were trialled for this reaction and it was found that dimethyl triazolium catalyst **35** was most the effective. A possible explanation for this catalyst being most effective is it was the least sterically hindered of the catalysts investigated and therefore able to access the protein α-oxo aldehyde with least resistance.

This project also revealed that the Umpolung reaction between an aldehyde donor and protein α-oxo aldehyde is more successful when the donor aldehyde is α-methyl substituted. A computational investigation into this caveat is discussed in this thesis which concluded that the substituted donor aldehyde is more amenable to couple with the target protein α-oxo aldehyde than homo-couple with itself. Whereas when a linear, non- α-methyl substituted donor aldehyde is used, the pathways between the desired reaction and homo-coupling are energetically close to each other and therefore in competition with each other. These competing pathways reduce the effectiveness of linear alkyl aldehyde donors and plausibly explain why improved conversions were witnessed when branched aldehydes were employed.

A trypsin digest experiment confirmed that this novel approach generates site-selective and site-specific protein modification.

The work described in this thesis also demonstrates the versatility of this strategy and its ability to not only modify proteins at the *N*-terminus as with myoglobin and thioredoxin but also at internal residues, GFP and sfGFP respectively.

To date, specific ion effects have been shown to only slow the reactivity of this chemistry down, unless employing ammonium sulphate, wherein an undesired aldol reaction is produced.

The strategy developed and discussed in this thesis does suffer from notable drawbacks however, including the requirement for α-methyl substituted aldehyde donors and the aforementioned purification procedure post reaction to hydrolytically cleave the unwanted hemi-aminal bonds formed between donor aldehyde and surface exposed lysine residues. Despite these drawbacks however, the Umpolung approach described in this thesis is a robust method for the chemical modification of proteins and it is hoped that it will not only further expand the chemical protein modification tool box but have real word applications in the fields of chemical biology and medicinal chemistry.

7.2: Future directions

The future of the Umpolung strategy described in this thesis can be split into two distinct paths. The first path is further strategy improvement. As discussed, this approach in its current format, does require dialysis overnight, post reaction. A more detailed study into the aldehyde donor may be able to preclude this requirement. In addition, this thesis was unable to produce a viable bioorthogonal Stetter reaction. It is hoped that further study in this area could resolve the unwanted α-oxo aldehyde oxidation.

The second path is the application of this strategy itself for chemical biology and pharmaceutical studies. This is a broader goal and less specific than the strategy development possibilities but is far more important. It should be the end goal for any effort to expand the toolbox of chemical protein modification that the method is simple, fast and applicable in real world efforts. As noted in chapter 1.1.8, owing to the recent rise in popularity of this field, reports are increasingly published with little to no regard of "bioorthogonal" focussing instead on novelty alone. Novelty aside, these reports add little to the usefulness of the current chemical protein modification toolbox available. It is hoped that the novel strategy described and developed in this thesis does not stray into the territory of novel without purpose. This could be in the form of incorporating a second protein aldehyde as a donor for protein dimerization ad the potential for novel bio-materials.

Experimental

8.1 General Methods

Solvents. All solvents were commercially obtained and dried with molecular sieves (when stated) with the exception of solvents used for flash chromatography purposes, where GPR-grade solvents were used. All commercially-available reagents were used as received.

Characterisation. Electrospray mass spectra for small molecules and peptides were recorded at room temperature on a Bruker Daltonics micrOTOF spectrometer. Protein ESI mass spectra were obtained on a Bruker Solarix XR 9.4 T instrument. ^{1}H and ^{13}C spectra were recorded at 400 MHz and 101 MHz respectively on a JEOL ECS 400 instrument using an internal deuterium lock or 500 MHz and 126 MHz respectively. Chemical shifts are reported in parts per million relatives to $CHCl_3$ (δH 7.27) and CDCl3 (δC 77.0, central line of triplet) The following abbreviations are used in ^{1}H NMR analysis: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, dd = double doublet, dt = double triplet, td = triple doublet, ddd = double double doublet.

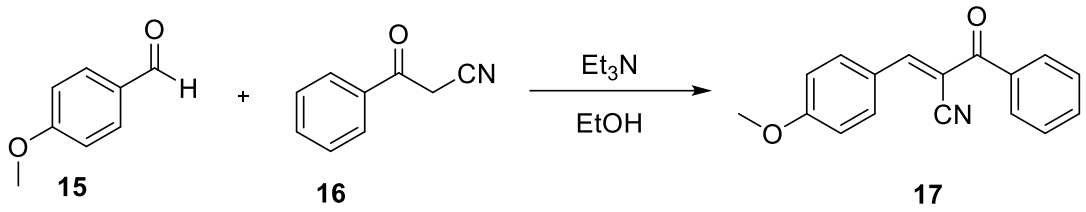
Liquid Chromatography. Liquid chromatography-mass spectrometry (LCMS) was performed on a Dionex UltiMate 3000 Ci Rapid Separation LC system equipped with an UltiMate 3000 photodiode array detector probing at 210-400 nm using a Waters Symmetry C18 3.5 µm column, 4.6 × 75 mm, coupled to a HCT ultra ETD II ion trap spectrometer, in positive ion mode unless stated otherwise.

Chromatography. Flash column chromatography was carried out using silica (Sigma-Aldrich Å 220-440). Thin layer chromatography was carried out using commercially available Merk F_{254} aluminium backed silica plates.

Bioconjugate product conversion. Conversion from starting protein to the anticipated bioconjugates product (conjugation yields, %) was calculated using equation 1, employing starting protein peak intensity and expected product peak intensity.

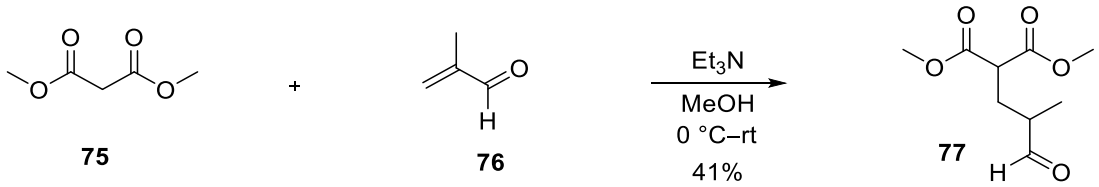
$$\frac{Product\ peak\ intensity}{Starting\ material\ peak\ intensities + Product\ peak\ intensity}\ x\ 100 = \%\ conversion \quad \text{Equ. 1}$$

(E)-2-benzoyl-3-(4-methoxyphenyl)acrylonitrile 17



Commercially available benzoylacetonitrile **16** (200 mg, 1.37 mmol) was added to *p*-methoxy benzaldehyde **15** (300 µL, 1.52 mmol) in absolute ethanol (10 mL) and heated until homogenous. 1 drop of Et_3N was then added and the solution allowed to cool to room temperature. White crystals then precipitated out of solution which were filtered and washed with ice cold ethanol (266 mg, 73%). This compound was used crude without further purification. Spectral data in agreement with literature values.[298]

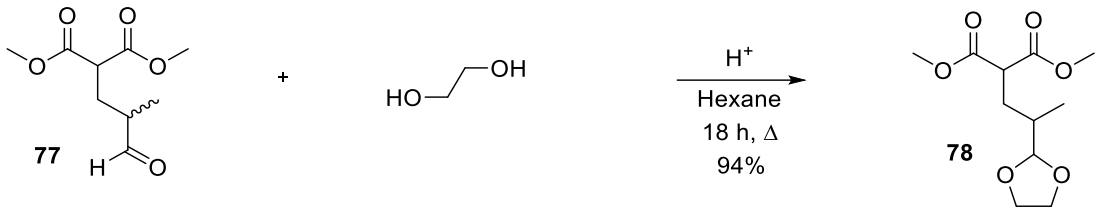
Dimethyl 2-(2-methyl-3-oxopropyl) malonate 77



75 **76** **77**

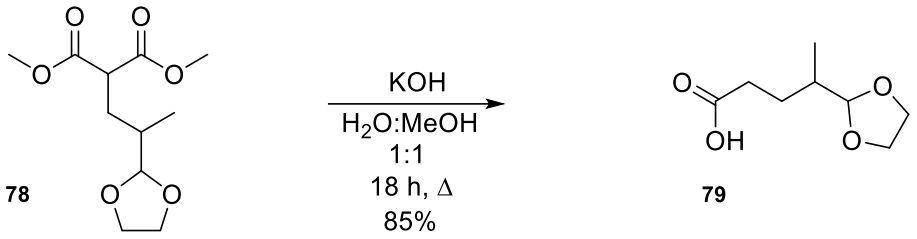
A mixture of methacrolein (7.40 mL, 71.5 mmol), dimethyl malonate (7.05 mL, 60.0 mmol) and Et_3N (8.35 mL, 120 mmol) in MeOH (50 mL) was stirred at 0 °C and allowed to warm to rt over 17 h. EtOAc (30 mL) was added and washed with 3M aqueous HCl (3 x 30 mL), water (3 x 30 mL) and brine (3 x 30 mL). This generated dimethyl 2-(2-methyl-3-oxopropyl) malonate as a pale yellow oil (5.05 g, 41%) used without further purification; **^{1}H NMR** (400 MHz, $CDCl_3$): δ (ppm) 9.61 (d, 1H, J = 1.3, C<u>H</u>O), 3.74 (s, 6H, OC<u>H</u>$_3$), 3.51 (1H, t, J = 7.3, $CH_2C\underline{H}(CO_2Me_2)_2$, 2.47—2.31 (2H, m, <u>H</u>$_2$CH$(CO_2Me_2)_2$, 1.96—1.89 (1H, m, $CH_3C\underline{H}CHO$), 1.14 (3H, d, J = 7.3, C<u>H</u>$_3$CHCHO); **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm), 203.5 (CHO), 169.6 (CO_2Me), 167.3 (CO_2Me), 53.0 (OMe), 49.2 ($CH(CO_2Me)_2$), 44.3 (<u>C</u>HCH$_3$), 29.5 (CH_2), 13.9 (Me); **IR (ATR)**: 2956, 1727, 1435, 1237,1197, 1154, 1059, 969, 735. **ESI-HRMS:** Found $[M+H]^+$ 203.0915, $[M+Na]^+$ 225.0735, $C_9H_{14}O_5$ calculated 203.0914, $C_9H_{14}NaO_5$ calculated 225.0733.

Dimethyl 2-(2-(1,3-dioxolan-2-yl)propyl)malonate 78



To aldehyde **77** (200 mg, 0.990 mmol) stirred in hexane (15 mL) camphor-10-sulfonic acid (2.3 mg) and ethylene glycol (82 µL, 1.49 mmol) were added and stirred at reflux for 18 h. The solution was then washed with water (3 x 30 mL), brine (3 x 30 mL), dried over $MgSO_4$ (~ 3 g) and concentrated under vacuum to yield dimethyl 2-(2-methyl-3-oxopropyl) malonate as a yellow oil (188 mg, 82%) which was used without further purification. **^{1}H NMR** (400 MHz, $CDCl_3$): δ_H (ppm) 4.70 (d, 1H, *J* = 4.7, OC<u>H</u>O), 3.91-3.89 (m, 2H, C<u>H</u>$_2$O), 3.88-3. 79 (m, 2H, C<u>H</u>$_2$O), 3.74 (s, 6H, OC<u>H</u>$_3$), 3.64 (1H, t, *J* = 7.3, $CH_2C\underline{H}(CO_2Me_2)_2$, 2.20-2.10 (2H, m, $C\underline{H}_2CH(CO_2Me_2)_2$, 1.85-1.73 (1H, m, $CH_3C\underline{H}CHO$), 0.97 (3H, d, *J* = 6.8, Me); **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm), 172.6 (CO_2Me), 170.2 (CO_2Me), 107.3 (OCO), 65.2 (CH_2O), 62.4 (CH_2O), 52.8 (2C; OMe), 49.9 ($CH(CO_2Me)_2$), 35.0 (CH_2), 30.8 (CH_2), 14.6 (Me); **IR (ATR):** 2956, 2884, 1730, 1435, 1235, 1153, 944, 832. **ESI-HRMS:** Found $[M+Na]^+$ 269.0995, $C_{11}H_{18}NaO_6$ calculated 269.0996

4-(1,3-Dioxolan-2-yl) pentanoic acid 79



78 79

Dimethyl 2-(2-methyl-3-oxopropyl) malonate (3.00 g, 12.12 mmol) **78** was added to a solution of KOH (3.00 g) dissolved in a 1:1 mixture of H_2O: MeOH (20 mL) and left to stir at reflux for 18 h. The reaction mixture was then washed with DCM (3 x 20 mL) and the aqueous phase acidified to pH 5 with HCl. EtOAc was then added and the organic washed with brine (3 x 20 mL), dried with $MgSO_4$ (~4 g) and concentrated under vacuum to yield 4-(1,3-dioxolan-2-yl) pentanoic acid as a pale yellow oil (1.8 g, 85%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ_H (ppm): 8.5 (1H, bs, OH), 4.70 (d, 1H, *J* = 4.7, OC<u>H</u>O), 3.91-3.89 (m, 2H, C<u>H</u>$_2$O), 3.88-3.78 (m, 2H, C<u>H</u>$_2$O), 3.6 (1H, t, *J* = 7.3, CH_2), 2.20-2.10 (2H, m, C<u>H</u>$_2$CH), 1.85-1.73 (1H, m, C<u>H</u> CH_3), 0.97 (3H, d, *J* = 6.8, Me); **^{13}C NMR** (101 MHz, $CDCl_3$) δ_C (ppm), 174.3 (CO_2OH), 107.4 (OCO), 65.3 (CH_2O), 65.1 (CH_2O), 49.9 (CH), 34.6 (CH_2), 21.2 (CH_2), 14.5 (Me); **IR (ATR)**: 2969, 2884, 1710, 1457, 1379, 1166, 796, 658. **ESI-HRMS:** Found $[M-H]^-$ 173.0274, $C_8H_{14}O_4$ calculated 173.0897.

Protein Purification

The pBAD vector containing ampicillin resistance and either Ser-GFP(Y39TAG) or sfGFP(N150TAG) genes, together with the pEVOL vector containing tRNAPyl, pylRS (*M. mazei*, wild type) and chloramphenicol resistance genes, were co-transformed into electrocompetent *E. coli* Top10 cells and selected on LB agar plates containing ampicillin (100 μg/ml) and chloramphenicol (35 μg/ml). For small-scale expression, 0.5 mL of an overnight culture grown from a single colony was inoculated into 50 mL Terrific Broth Medium containing ampicillin (100 μg/ml) and chloramphenicol (35 μg/ml) in a 250 mL baffled conical flask. At 37 °C with shaking (220 rpm), cells typically grew within 3 h to an OD600 of 0.2-0.3, at which point NCAA (stock solution 80 mM in 0.1 M NaOH (aq.)) was added to a final concentration of 1.5 mM. The cultures were allowed to grow until an OD600 of 0.4-0.6, at which point protein expression was induced by addition of L-arabinose (stock solution 20% (w/w)) at a final concentration of 0.02% (w/w) and left to grow for 16-18 h at 37 °C with shaking (220 rpm). The cultures were harvested by centrifugation (6 000 × g, 4 °C, 20 min). Pellets were re-suspended in 4 × PBS, 10 mM imidazole, pH 8.0 with a Pierce Protease Inhibitor (EDTA-free) tablet and then lysed by sonication on ice for 6 × 30 s with 30 s intervals. The lysate was clarified by centrifugation (20 000 × g, 4 °C, 20 min) and loaded onto a Ni HiTrap Chelating HP column (1 ml, GE Healthcare) pre-equilibrated in 4 × PBS, 10 184 mM imidazole, pH 8.0. The column was washed with 10 column volumes of this buffer and then eluted using a gradient of 0-100% 4 × PBS, 500 mM imidazole, pH 8.0 over 7.5 column volumes, taking 0.5 mL fractions, and the column washed with 7.5 column volumes of 4 × PBS, 500 mM imidazole, pH 8.0, taking 0.5 mL fractions. Fractions containing full-length protein (as determined by SDS-PAGE) were pooled, dialysed into 1 × PBS, pH 7.4 and concentrated (Vivaspin centrifugal concentrator, 10000 MWCO) to a final concentration of 330 μM (as determined by UV-visible spectroscopy, ε280 = 2.0 × 104 dm^3 mol^{-1} cm^{-1}) and stored at -80 °C. For large-scale expressions, the procedure was followed largely as above with 1 L cultures inoculated with 10 mL of an overnight culture, to which NCAA and L-arabinose were added to the same final concentrations. For purification, a larger Ni HiTrap Chelating HP column (5 mL, GE Healthcare) was used, taking 2.5 mL fractions.

Plasmid information

The vectors pEVOL, harbouring pylT and pylRS (*M. mazei*, wild type or Y306A Y384F double mutant) genes, and pBAD, harbouring the GFP(Y39TAG) gene and the pBAD vector harbouring the sfGFP(N150TAG) gene was supplied by Dr Robin Brabham.

Sequence of GFP(Y39ThzK)

X represents the position of ThzK.

```
           10          20          30          40          50
60
[M]SYKDDDDKV SKGEELFTGV VPILVELDGD VNGHKFSVSG EGEGDATXGK
LTLKFICTTG
           70          80          90         100         110
120
  KLPVPWPTLV TTLTYGVQCF SRYPDHMKQH DFFKSAMPEG YVQERTIFFK
DDGNYKTRAE
          130         140         150         160         170
180
  VKFEGDTLVN RIELKGIDFK EDGNILGHKL EYNYNSHNVY IMADKQKNGI
KANFKIRHNI
          190         200         210         220         230
240
  EDGSVQLADH YQQNTPIGDG PVLLPDNHYL STQSALSKDP NEKRDHMVLL
EFVTAAGITL
          250
  GMDELYKHHH HHH*
```

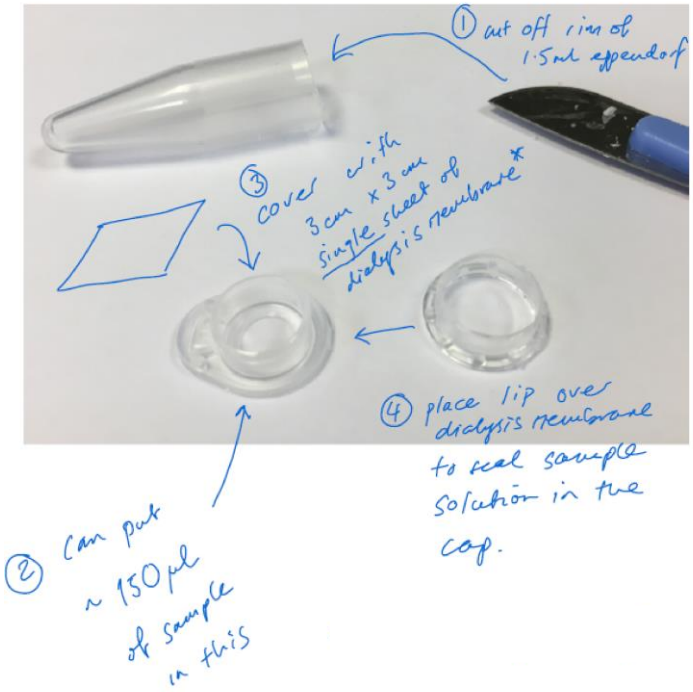
Sequence of sfGFP(N150TAG)-His6

X represents the position of ThzK.

```
          10          20          30          40          50
60
[M]VSKGEELFT GVVPILVELD GDVNGHFSV REGEGDATN GKLTLKFICT
TGKLPVPWPT
          70          80          90         100         110
120
  LVTTLTYGVQ CFSRYPDHMK RHDFFKSAMP EGYVQERRTIS FKDDGTYKTR
AEVKFEGDTL
         130         140         150         160         170
180
 VNRIELKGID FKEDGNILGH KLEYNFNSHX VYITADKQKN GIKANFKIRH
NVEDGSVQLA
         190         200         210         220         230
240
  DHYQQNTPIG DGPVLLPDNH YLSTQSVLSK DPNEKRDHMV LLEFVTAAGI
THGMDELYKG
         250
  SHHH HHH*
```

Protein Modification

Small volume protein dialysis



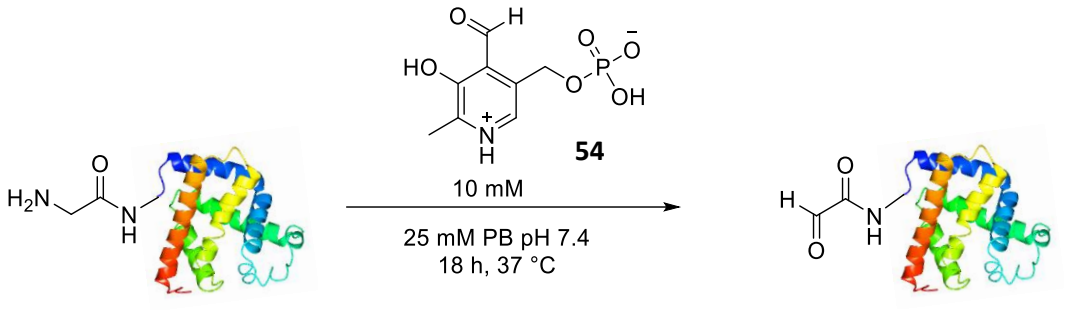
All modified protein dialysis was carried out *via* the following method:

Firstly, lid and lip of a 1.5 mL eppendorf were removed. The lid was inverted and modified protein for dialysis was transferred into the inverted lid.

A section of dialysis tubing was cut to size and left to hydrate in MiliQ water for 5 min. Once hydrated, the dialysis tubing was cut open to give a single membrane and trimmed to size approximately 3 cm x 3 cm.

The dialysis tubing was then carefully placed over the lid containing modified protein. The lip was then attached over the dialysis tubing and lid to form a seal. This was then placed dialysis tubing faced down and left to dialyse overnight.

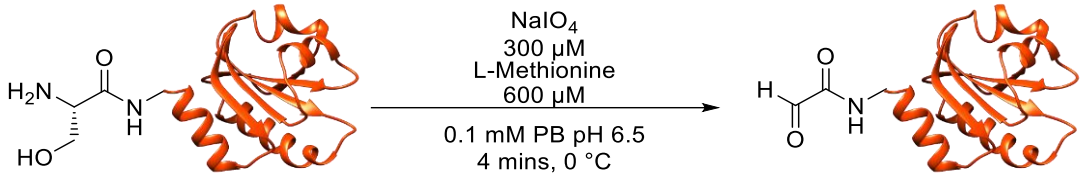
Transamination of myoglobin 54 to α-oxo aldehyde-myoglobin 55



53 **55**

In a 1.5 mL Eppendorf, a 600 µL aliquot of a 250 µM of myoglobin **53** in 25 mM PB pH 6.5 was charged with 600 µL of a 25 mM pyridoxal-5-phosphate **54** solution in 25 mM PB pH 6.5 (pH adjusted as required to pH 6.5 using 4M NaOH). Final pH of solution was confirmed using pH paper. The solution was mixed via repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. The solution was then purified through spin concentration using 10,000 MWCO, and the resulting glyoxyl-myoglobin solution was concentrated to 200 µM, eluting with water. Oxidation to α-oxo aldehyde-myoglobin **55** was confirmed by LC-MS.

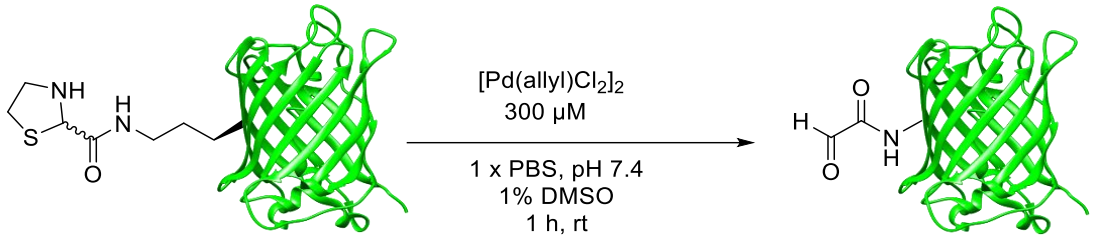
Oxidation of thioredoxin to α-oxo aldehyde-thioredoxin 58



58

In a 0.5 mL eppendorf, a 100 µL aliquot of an 85 µM thioredoxin in 0.1 M 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, followed aby the addition of 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 repeated pipetting, and left on ice in the dark. After 4 mins the reaction was purified using a PD SpinTrap G25 desalting column (GE Healthcare), eluting into 25 mM PB pH 7.5. Oxidation to α-oxo aldehyde thioredoxin **58** was confirmed by LC-MS analysis.

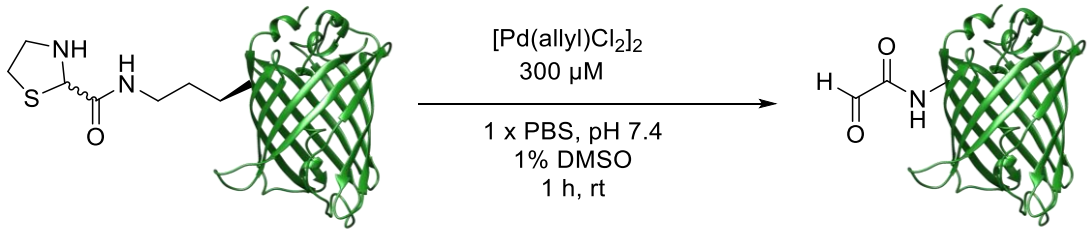
Pd decaging GFPY39ThzK 64 to GFPY39 α-oxo K 65



64 **65**

In a 0.5 mL Eppendorf, a 100 µL aliquot of a 300 µM GFPY39ThzK **64** stock in 1 x PBS, pH 7.4, was charged with 1 µL of a 30 mM allylpalladium(II) chloride dimer solution in DMSO. The solution was mixed immediately through pipette tip swirling, and allowed to sit at rt for 1 h without further agitation. The reaction was quenched by addition of 3-mercaptopropanoic acid, (10 µL, 1% v/v solution, 10 x PBS final concentration = 0.1% v/v) and allowed to sit at 25 °C for 15 min without further agitation. The reaction was diluted up to 500 µL then desalted using a PD MiniTrap G-25 (GE Healthcare), eluting with 25 mM PB pH 7.5. Conversion to GFP α-oxo K **65** was confirmed by ESI-MS analysis.

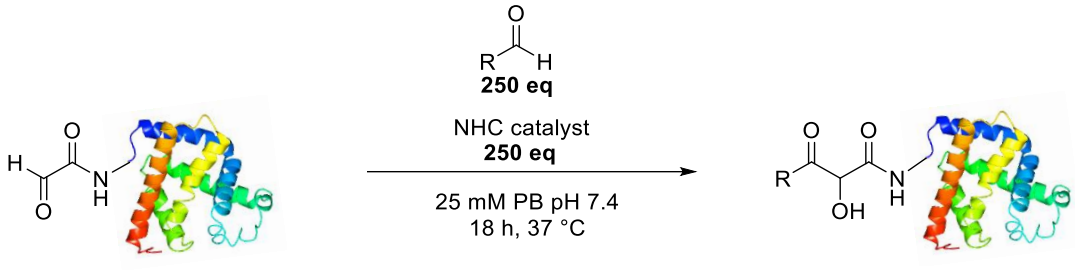
Pd decaging GFPN150ThzK 67 to sfGFPN150 α-oxo K 68



67 **68**

In a 0.5 mL Eppendorf, a 100 µL aliquot of a 300 µM sfGFPN150 **67** stock in 1 x PBS, pH 7.4, was charged with 1 µL of a 30 mM allylpalladium(II) chloride dimer solution in DMSO. The solution was mixed immediately through pipette tip swirling, and allowed to sit at rt for 1 h without further agitation. The reaction was quenched by addition of 3-mercaptopropanoic acid, (10 µL, 1% v/v solution, 10 x PBS final concentration = 0.1% v/v) and allowed to sit at 25 °C for 15 min without further agitation. The reaction was diluted up to 500 µL then desalted using a PD MiniTrap G-25 (GE Healthcare), eluting with 25 mM PB pH 7.5. Conversion to the sfGFP α-oxo K **68** was confirmed by ESI-MS analysis.

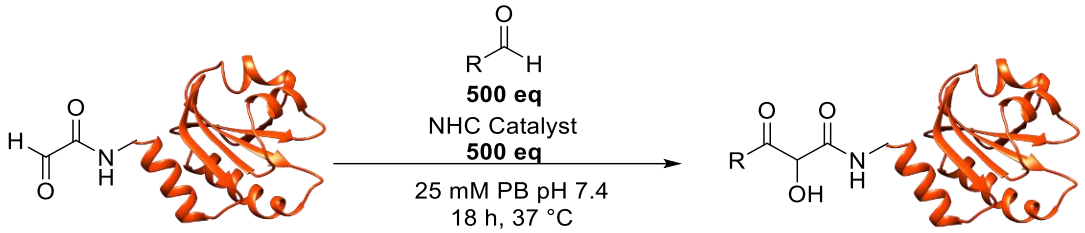
General procedure 1 - Preparation of modified α-oxo aldehyde-myoglobin 55



55

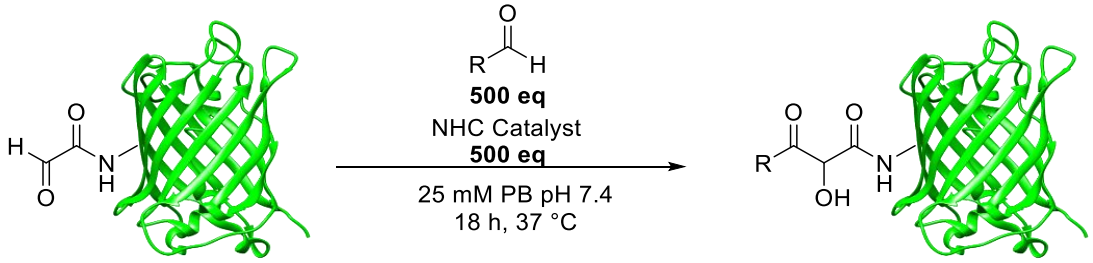
In a 0.5 mL Eppendorf, a 100 µL aliquot of a 200 µM α-oxo aldehyde-myoglobin **55** stock in PB pH 7.4 was charged 25 µL of a 200 mM NHC catalyst stock solution in 25 mM PB pH 7.4. The solution was then charged with 25 µL aliquot of a 200 mM aldehyde donor stock solution in 25 mM PB pH 7.4. Finally, 50 µL of 25 mM PB pH 7.4 was added to the reaction to give a final reaction volume of 200 µL. The solution was mixed via repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Conversion to the expected Umpolung product was judged by LC-MS.

General procedure 2 - Preparation of modified α-oxo aldehyde-Thioredoxin 58



In a 0.5 mL Eppendorf, a 25 µL aliquot of 88 µM α-oxo aldehyde-Thioredoxin **58** stock in PB pH 7.4 was charged 2.9 µL of a 200 mM NHC catalyst stock solution in 25 mM PB pH 7.4. The solution was then charged with 2.9 µL aliquot of a 200 mM aldehyde donor stock solution in 25 mM PB pH 7.4. Finally, 19.2 µL of 25 mM PB pH 7.4 was added to generate a final volume of 50 µL. The solution was mixed via repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Conversion to the expected Umpolung product was judged by LC-MS.

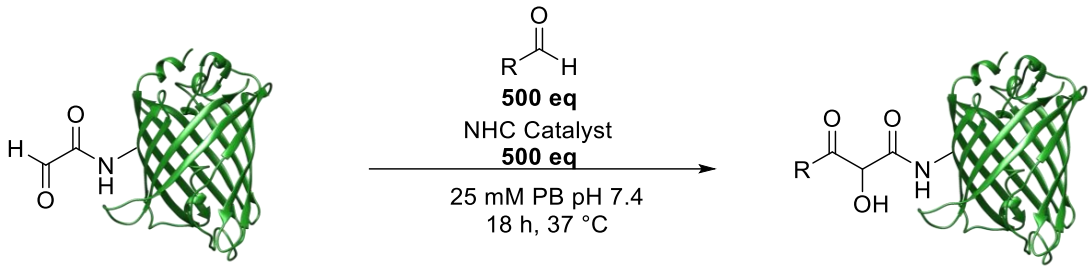
General procedure 3 - Preparation of modified aldehyde-GFPY39 α-oxo K 65



65

In a 0.5 mL Eppendorf, a 25 µL aliquot of a 90 µM GFPY39 α-oxo K **65** stock in PB pH 7.4 was charged 5.6 µL of a 200 mM NHC catalyst stock solution in 25 mM PB pH 7.4. The solution was then charged with 5.6 µL aliquot of a 200 mM aldehyde donor stock solution in 25 mM PB pH 7.4. Finally, 13.8 µL of 25 mM PB pH 7.4 was added to generate a final volume of 50 µL. The solution was mixed via repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. The protein was purified *via* dialysis with the addition of 5.6 µL of glycine in 25 mM PB pH 7.4 and left for 24 h. Conversion to the expected benzoin condensation product was judged by ESI-MS.

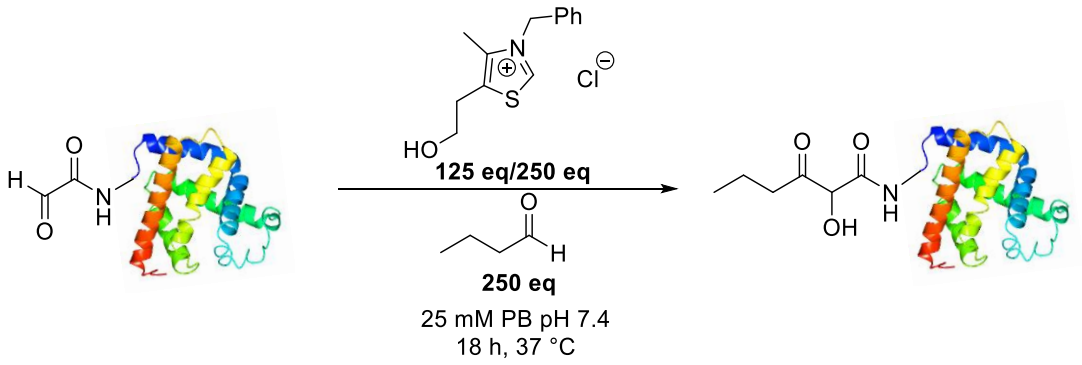
General procedure 3 - Preparation of modified α-oxo aldehyde-sfGFP 68



In a 0.5 mL Eppendorf, a 25 µL aliquot of a 90 µM sfGFPN150 α-oxo K **68** stock in PB pH 7.4 was charged 5.6 µL of a 200 mM NHC catalyst stock solution in 25 mM PB pH 7.4. The solution was then charged with 5.6 µL aliquot of a 200 mM aldehyde donor stock solution in 25 mM PB pH 7.4. Finally, 13.8 µL of 25 mM PB pH 7.4 was added to generate a final volume of 50 µL. The solution was mixed via repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. The protein was purified via dialysis with the addition of 5.6 µL of glycine in 25 mM PB pH 7.4 and left for 24 h. Conversion to the expected Umpolung product was judged by ESI-MS.

Catalyst screen of modified α-oxo aldehyde-myoglobin with butyraldehyde 43

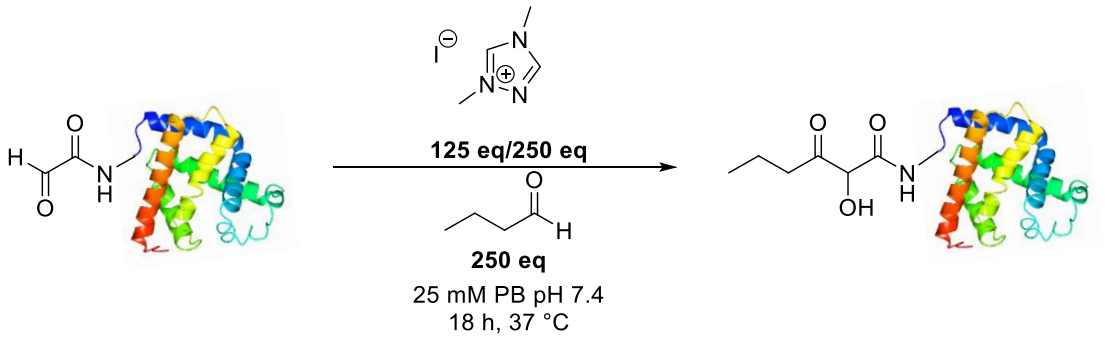
3-Benzyl-5-(2-hydroxyethyl)-4-methylthiazolium chloride 36



For 125 eq catalyst screen, in a 0.5 mL Eppendorf, a 100 µL aliquot of a 200 µM α-oxo aldehyde-myoglobin **55** stock in PB pH 7.4 was charged with 12.5 µL of a 200 mM 3-benzyl-5-(2-hydroxyethyl)-4-methylthiazolium chloride **36** catalyst stock solution in 25 mM PB pH 7.4. The solution was then charged with 25 µL aliquot of a 200 mM butyraldehyde donor stock solution in 25 mM PB pH 7.4. Finally, 62.5 µL of 25 mM PB pH 7.4 was added to the reaction to give a final reaction volume of 200 µL. The solution was mixed *via* repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Conversion to the expected umpolung product was judged by LC-MS.

For 250 eq catalyst screen, in a 0.5 mL Eppendorf, a 100 µL aliquot of a 200 µM α-oxo aldehyde-myoglobin **55** stock in PB pH 7.4 was charged with 25 µL of a 200 mM 3-benzyl-5-(2-hydroxyethyl)-4-methylthiazolium chloride catalyst stock solution in 25 mM PB pH 7.4. The solution was then charged with 25 µL aliquot of a 200 mM butyraldehyde donor stock solution in 25 mM PB pH 7.4. Finally, 50 µL of 25 mM PB pH 7.4 was added to the reaction to give a final reaction volume of 200 µL. The solution was mixed *via* repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Conversion to the expected Umpolung product was judged by LC-MS.

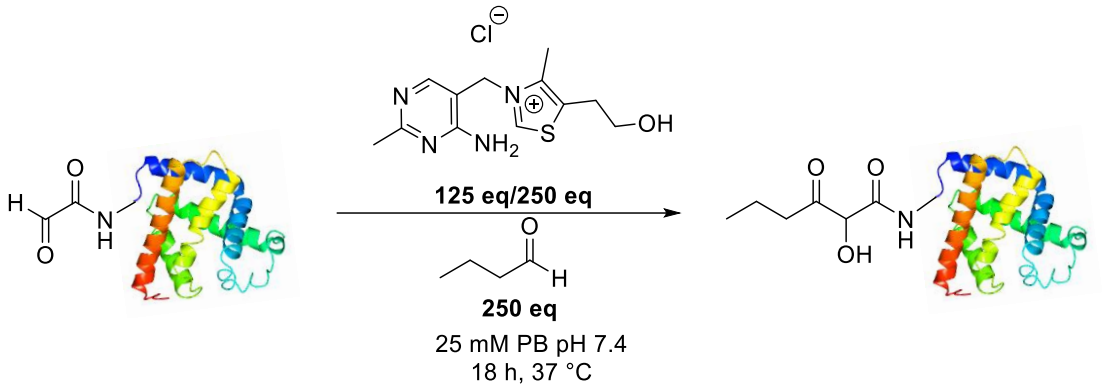
1,4-Dimethyl-4H-1,2,4-triazolium iodide 35



For 125 eq catalyst screen, in a 0.5 mL Eppendorf, a 100 µL aliquot of a 200 µM α-oxo aldehyde-myoglobin **55** stock in PB pH 7.4 was charged with 12.5 µL of a 200 mM 1,4-dimethyl-4H-1,2,4-triazolium iodide **35** catalyst stock solution in 25 mM PB pH 7.4. The solution was then charged with 25 µL aliquot of a 200 mM butyraldehyde donor stock solution in 25 mM PB pH 7.4. Finally, 62.5 µL of 25 mM PB pH 7.4 was added to the reaction to give a final reaction volume of 200 µL. The solution was mixed *via* repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Conversion to the expected Umpolung product was judged by LC-MS.

For 250 eq catalyst screen, in a 0.5 mL Eppendorf, a 100 µL aliquot of a 200 µM α-oxo aldehyde-myoglobin **55** stock in PB pH 7.4 was charged with 25 µL of a 200 mM 1,4-dimethyl-4H-1,2,4-triazolium iodide **35** catalyst stock solution in 25 mM PB pH 7.4. The solution was then charged with 25 µL aliquot of a 200 mM butyraldehyde donor stock solution in 25 mM PB pH 7.4. Finally, 50 µL of 25 mM PB pH 7.4 was added to the reaction to give a final reaction volume of 200 µL. The solution was mixed *via* repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Conversion to the expected Umpolung product was judged by LC-MS.

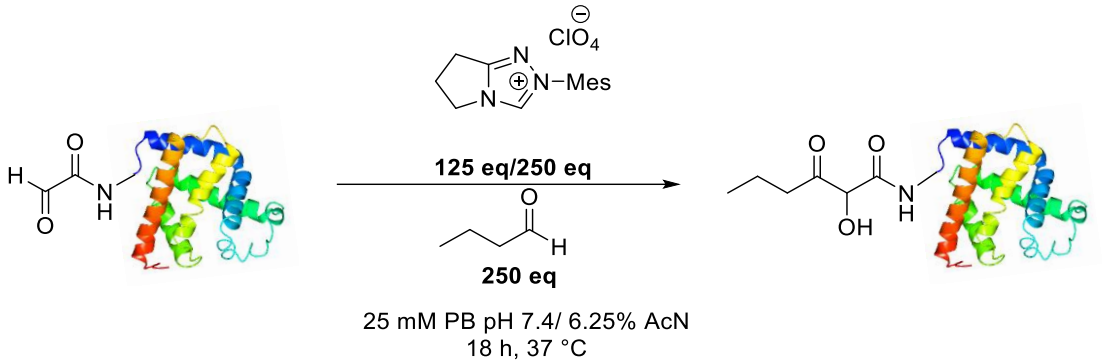
Thiamine hydrochloride 34



For 125 eq catalyst screen, in a 0.5 mL Eppendorf, a 100 µL aliquot of a 200 µM α-oxo aldehyde-myoglobin **55** stock in PB pH 7.4 was charged with 12.5 µL of a 200 mM thiamine hydrochloride catalyst stock solution in 25 mM PB pH 7.4. The solution was then charged with 25 µL aliquot of a 200 mM butyraldehyde donor stock solution in 25 mM PB pH 7.4. Finally, 62.5 µL of 25 mM PB pH 7.4 was added to the reaction to give a final reaction volume of 200 µL. The solution was mixed *via* repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Conversion to the expected Umpolung product was judged by LC-MS.

For 250 eq catalyst screen, in a 0.5 mL Eppendorf, a 100 µL aliquot of a 200 µM α-oxo aldehyde-myoglobin **55** stock in PB pH 7.4 was charged with 25 µL of a 200 mM thiamine hydrochloride in 25 mM PB pH 7.4. The solution was then charged with 25 µL aliquot of a 200 mM butyraldehyde donor stock solution in 25 mM PB pH 7.4. Finally, 50 µL of 25 mM PB pH 7.4 was added to the reaction to give a final reaction volume of 200 µL. The solution was mixed *via* repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Conversion to the expected Umpolung product was judged by LC-MS.

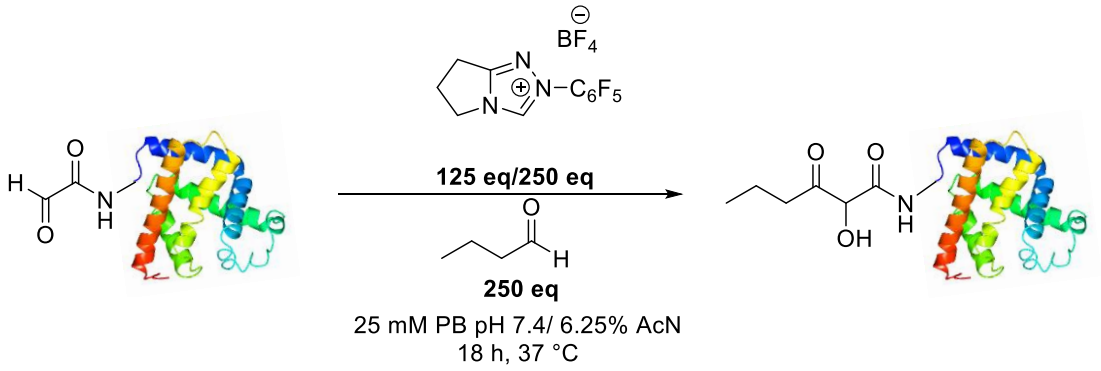
6,7-Dihydro-2-(2,4,6-trimethylphenyl)-5H-pyrrolo[2,1-c]-1,2,4-triazolium Perchlorate 33



For 125 eq catalyst screen, in a 0.5 mL Eppendorf, a 100 µL aliquot of a 200 µM α-oxo aldehyde-myoglobin **55** stock in PB pH 7.4 was charged with 12.5 µL of a 200 mM 6,7-dihydro-2-(2,4,6-trimethylphenyl)-5H-pyrrolo[2,1-c]-1,2,4-triazolium perchlorate **33** catalyst stock solution in acetonitrile. The solution was then charged with 25 µL aliquot of a 200 mM butyraldehyde donor stock solution in 25 mM PB pH 7.4. Finally, 62.5 µL of 25 mM PB pH 7.4 was added to the reaction to give a final reaction volume of 200 µL. The solution was mixed *via* repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Conversion to the expected Umpolung product was judged by LC-MS.

For 250 eq catalyst screen, in a 0.5 mL Eppendorf, a 100 µL aliquot of a 200 µM α-oxo aldehyde-myoglobin **55** stock in PB pH 7.4 was charged with 12.5 µL of a 400 mM 6,7-dihydro-2-(2,4,6-trimethylphenyl)-5H-pyrrolo[2,1-c]-1,2,4-triazolium perchlorate **33** catalyst stock solution in acetonitrile. The solution was then charged with 25 µL aliquot of a 200 mM aldehyde donor stock solution in 25 mM PB pH 7.4. Finally, 62.5 µL of 25 mM PB pH 7.4 was added to the reaction to give a final reaction volume of 200 µL. The solution was mixed *via* repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Conversion to the expected Umpolung product was judged by LC-MS.

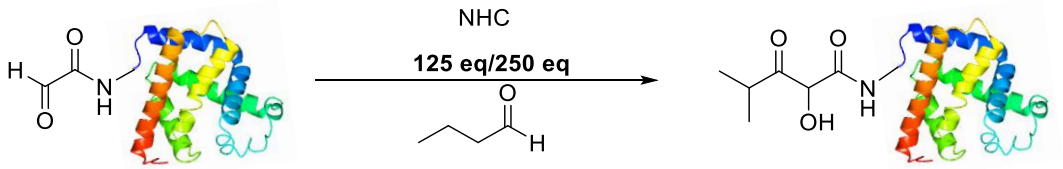
6,7-Dihydro-2-pentafluorophenyl-5H-pyrrolo[2,1-c]-1,2,4-triazolium tetrafluoroborate 18

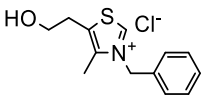
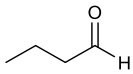
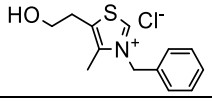
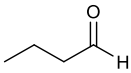
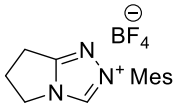
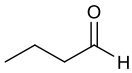
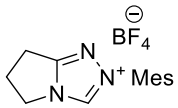
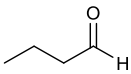
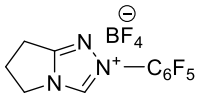
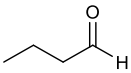
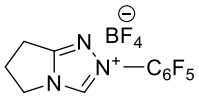
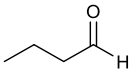
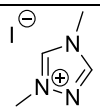
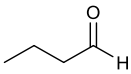
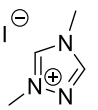
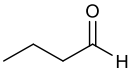
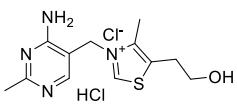
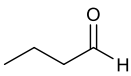
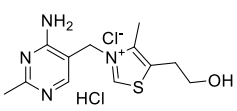
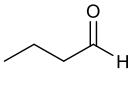


For 125 eq catalyst screen, in a 0.5 mL Eppendorf, a 100 µL aliquot of a 200 µM α-oxo aldehyde-myoglobin **55** stock in PB pH 7.4 was charged with 12.5 µL of a 200 mM 6,7-dihydro-2-pentafluorophenyl-5H-pyrrolo[2,1-c]-1,2,4-triazolium tetrafluoroborate **18** catalyst stock solution in acetonitrile. The solution was then charged with 25 µL aliquot of a 200 mM butyraldehyde donor stock solution in 25 mM PB pH 7.4. Finally, 62.5 µL of 25 mM PB pH 7.4 was added to the reaction to give a final reaction volume of 200 µL. The solution was mixed *via* repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Conversion to the expected Umpolung product was judged by LC-MS.

For 250 eq catalyst screen, in a 0.5 mL Eppendorf, a 100 µL aliquot of a 200 µM α-oxo aldehyde-myoglobin **55** stock in PB pH 7.4 was charged with 12.5 µL of a 400 mM 6,7-dihydro-2-pentafluorophenyl-5H-pyrrolo[2,1-c]-1,2,4-triazolium tetrafluoroborate **18** catalyst stock solution in acetonitrile. The solution was then charged with 25 µL aliquot of a 200 mM aldehyde donor stock solution in 25 mM PB pH 7.4. Finally, 62.5 µL of 25 mM PB pH 7.4 was added to the reaction to give a final reaction volume of 200 µL. The solution was mixed *via* repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Conversion to the expected Umpolung product was judged by LC-MS.

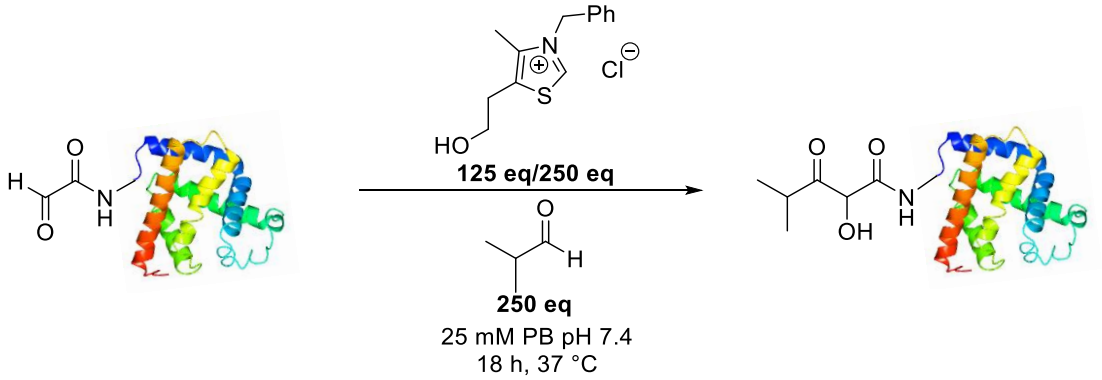
Tabulated results of Myoglobin catalyst screen with butyraldehyde 43



Catalyst	Catalyst eq	Aldehyde eq	Conversion
	**125**	 **250**	**24%**
	**250**	 **250**	**26%**
	**125**	 **250**	**NA%**
	**250**	 **250**	**31%**
	**125**	 **250**	**32%**
	**250**	 **250**	**38%**
	**125**	 **250**	**29%**
	**250**	 **250**	**34%**
	**125**	 **250**	**21%**
	**250**	 **250**	**48%**

Catalyst screen of modified α-oxo aldehyde-myoglobin with isobutyraldehyde 46

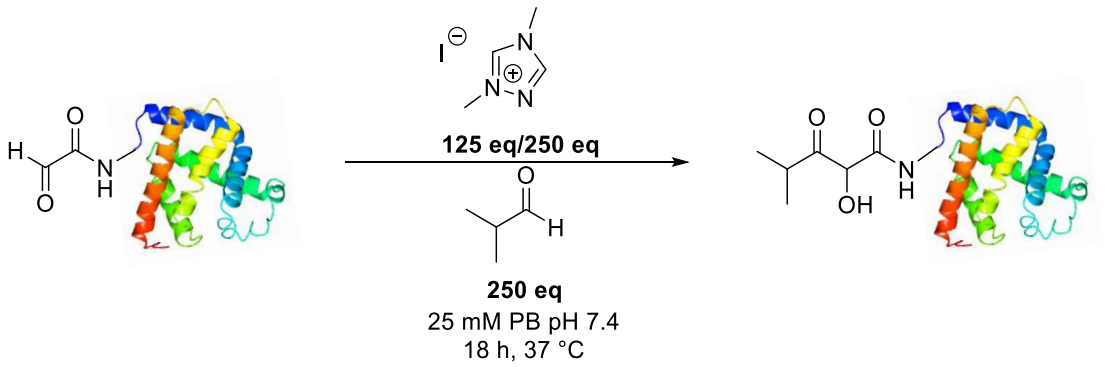
3-Benzyl-5-(2-hydroxyethyl)-4-methylthiazolium chloride 36



For 125 eq catalyst screen, in a 0.5 mL Eppendorf, a 100 µL aliquot of a 200 µM α-oxo aldehyde-myoglobin **55** stock in PB pH 7.4 was charged with 12.5 µL of a 200 mM 3-benzyl-5-(2-hydroxyethyl)-4-methylthiazolium chloride **36** catalyst stock solution in 25 mM PB pH 7.4. The solution was then charged with 25 µL aliquot of a 200 mM isobutyraldehyde donor **46** stock solution in 25 mM PB pH 7.4. Finally, 62.5 µL of 25 mM PB pH 7.4 was added to the reaction to give a final reaction volume of 200 µL. The solution was mixed *via* repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Conversion to the expected Umpolung product was judged by LC-MS.

For 250 eq catalyst screen, in a 0.5 mL Eppendorf, a 100 µL aliquot of a 200 µM α-oxo aldehyde-myoglobin **55** stock in PB pH 7.4 was charged with 25 µL of a 200 mM 3-benzyl-5-(2-hydroxyethyl)-4-methylthiazolium chloride catalyst stock solution in 25 mM PB pH 7.4. The solution was then charged with 25 µL aliquot of a 200 mM isobutyraldehyde donor stock solution in 25 mM PB pH 7.4. Finally, 50 µL of 25 mM PB pH 7.4 was added to the reaction to give a final reaction volume of 200 µL. The solution was mixed *via* repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Conversion to the expected Umpolung product was judged by LC-MS.

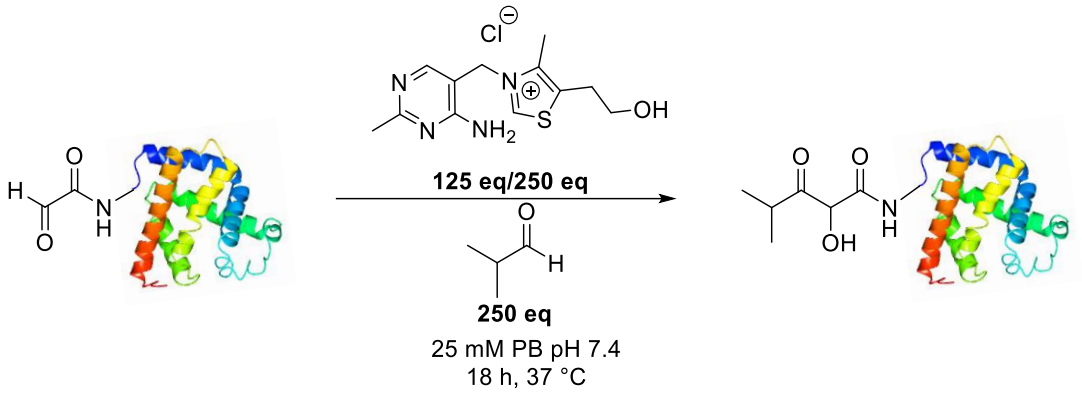
1,4-Dimethyl-4H-1,2,4-triazolium iodide 35



For 125 eq catalyst screen, in a 0.5 mL Eppendorf, a 100 µL aliquot of a 200 µM α-oxo aldehyde-myoglobin **55** stock in PB pH 7.4 was charged with 12.5 µL of a 200 mM 1,4-dimethyl-4H-1,2,4-triazolium iodide **35** catalyst stock solution in 25 mM PB pH 7.4. The solution was then charged with 25 µL aliquot of a 200 mM isobutyraldehyde donor stock solution in 25 mM PB pH 7.4. Finally, 62.5 µL of 25 mM PB pH 7.4 was added to the reaction to give a final reaction volume of 200 µL. The solution was mixed *via* repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Conversion to the expected Umpolung product was judged by LC-MS.

For 250 eq catalyst screen, in a 0.5 mL Eppendorf, a 100 µL aliquot of a 200 µM α-oxo aldehyde-myoglobin **55** stock in PB pH 7.4 was charged with 25 µL of a 200 mM 1,4-dimethyl-4H-1,2,4-triazolium iodide **35** catalyst stock solution in 25 mM PB pH 7.4. The solution was then charged with 25 µL aliquot of a 200 mM isobutyraldehyde donor stock solution in 25 mM PB pH 7.4. Finally, 50 µL of 25 mM PB pH 7.4 was added to the reaction to give a final reaction volume of 200 µL. The solution was mixed *via* repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Conversion to the expected Umpolung product was judged by LC-MS.

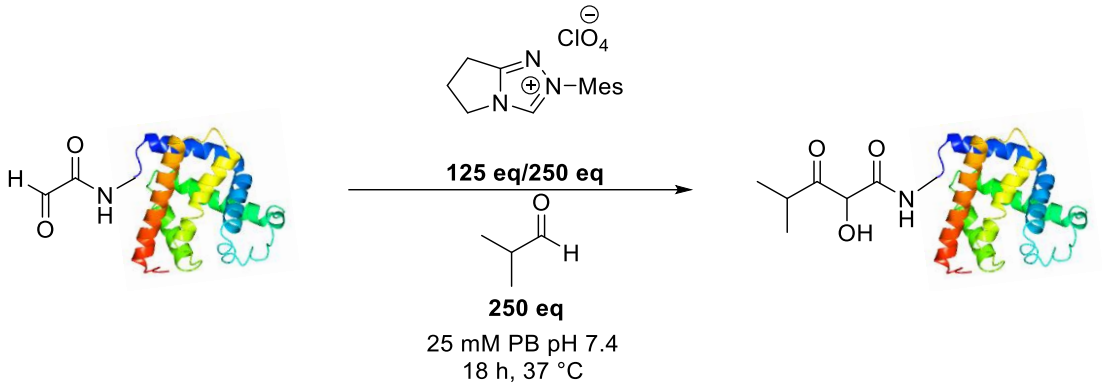
Thiamine hydrochloride 34



For 125 eq catalyst screen, in a 0.5 mL Eppendorf, a 100 µL aliquot of a 200 µM α-oxo aldehyde-myoglobin **55** stock in PB pH 7.4 was charged with 12.5 µL of a 200 mM thiamine hydrochloride catalyst stock solution in 25 mM PB pH 7.4. The solution was then charged with 25 µL aliquot of a 200 mM isobutyraldehyde donor stock solution in 25 mM PB pH 7.4. Finally, 62.5 µL of 25 mM PB pH 7.4 was added to the reaction to give a final reaction volume of 200 µL. The solution was mixed *via* repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Conversion to the expected Umpolung product was judged by LC-MS.

For 250 eq catalyst screen, in a 0.5 mL Eppendorf, a 100 µL aliquot of a 200 µM α-oxo aldehyde-myoglobin **55** stock in PB pH 7.4 was charged with 25 µL of a 200 mM thiamine hydrochloride in 25 mM PB pH 7.4. The solution was then charged with 25 µL aliquot of a 200 mM isobutyraldehyde donor stock solution in 25 mM PB pH 7.4. Finally, 50 µL of 25 mM PB pH 7.4 was added to the reaction to give a final reaction volume of 200 µL. The solution was mixed *via* repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Conversion to the expected Umpolung product was judged by LC-MS.

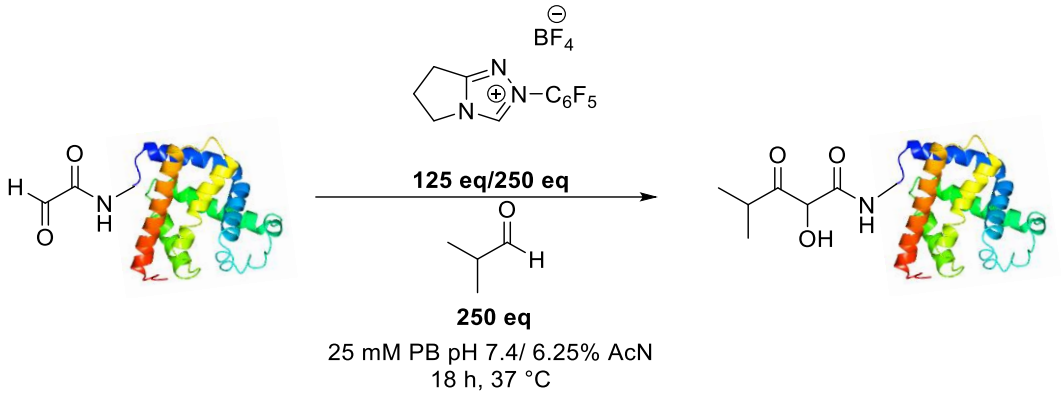
6,7-Dihydro-2-(2,4,6-trimethylphenyl)-5H-pyrrolo[2,1-c]-1,2,4-triazolium perchlorate 33



For 125 eq catalyst screen, in a 0.5 mL Eppendorf, a 100 µL aliquot of a 200 µM α-oxo aldehyde-myoglobin **55** stock in PB pH 7.4 was charged with 12.5 µL of a 200 mM 6,7-dihydro-2-(2,4,6-trimethylphenyl)-5H-pyrrolo[2,1-c]-1,2,4-triazolium perchlorate **33** catalyst stock solution in acetonitrile. The solution was then charged with 25 µL aliquot of a 200 mM butyraldehyde donor stock solution in 25 mM PB pH 7.4. Finally, 62.5 µL of 25 mM PB pH 7.4 was added to the reaction to give a final reaction volume of 200 µL. The solution was mixed *via* repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Conversion to the expected umpolung product was judged by LC-MS.

For 250 eq catalyst screen, in a 0.5 mL Eppendorf, a 100 µL aliquot of a 200 µM α-oxo aldehyde-myoglobin **55** stock in PB pH 7.4 was charged with 12.5 µL of a 400 mM 6,7-dihydro-2-(2,4,6-trimethylphenyl)-5H-pyrrolo[2,1-c]-1,2,4-triazolium perchlorate **33** catalyst stock solution in acetonitrile. The solution was then charged with 25 µL aliquot of a 200 mM aldehyde donor stock solution in 25 mM PB pH 7.4. Finally, 62.5 µL of 25 mM PB pH 7.4 was added to the reaction to give a final reaction volume of 200 µL. The solution was mixed *via* repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Conversion to the expected Umpolung product was judged by LC-MS.

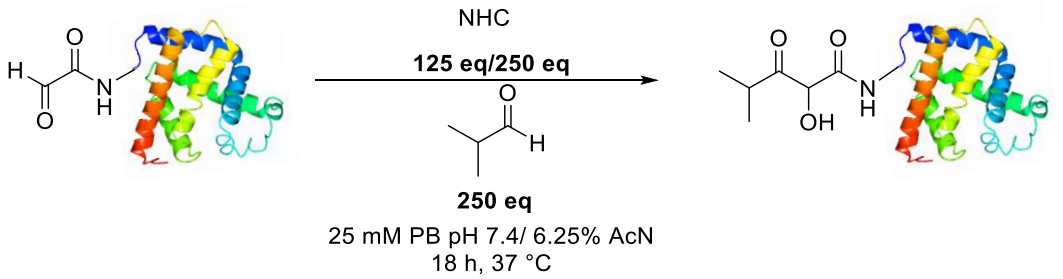
6,7-Dihydro-2-pentafluorophenyl-5H-pyrrolo[2,1-c]-1,2,4-triazolium tetrafluoroborate 18

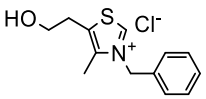
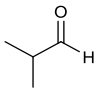
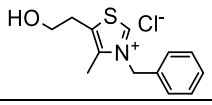
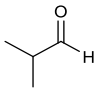
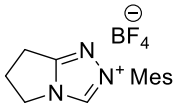
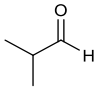
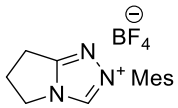
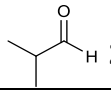
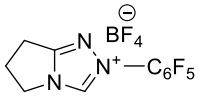
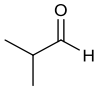
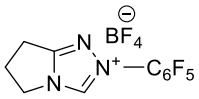
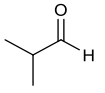
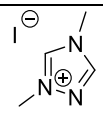
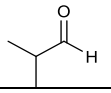
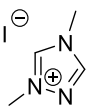
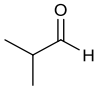
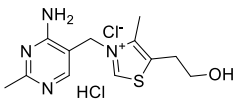
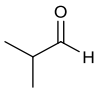
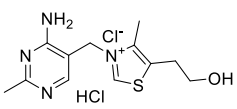
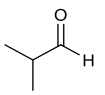


For 125 eq catalyst screen, in a 0.5 mL Eppendorf, a 100 µL aliquot of a 200 µM α-oxo aldehyde-myoglobin **55** stock in PB pH 7.4 was charged with 12.5 µL of a 200 mM 6,7-dihydro-2-pentafluorophenyl-5H-pyrrolo[2,1-c]-1,2,4-triazolium tetrafluoroborate **18** catalyst stock solution in acetonitrile. The solution was then charged with 25 µL aliquot of a 200 mM butyraldehyde donor stock solution in 25 mM PB pH 7.4. Finally, 62.5 µL of 25 mM PB pH 7.4 was added to the reaction to give a final reaction volume of 200 µL. The solution was mixed *via* repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Conversion to the expected umpolung product was judged by LC-MS.

For 250 eq catalyst screen, in a 0.5 mL Eppendorf, a 100 µL aliquot of a 200 µM α-oxo aldehyde-myoglobin **55** stock in PB pH 7.4 was charged with 12.5 µL of a 400 mM 6,7-dihydro-2-pentafluorophenyl-5H-pyrrolo[2,1-c]-1,2,4-triazolium tetrafluoroborate **18** catalyst stock solution in acetonitrile. The solution was then charged with 25 µL aliquot of a 200 mM aldehyde donor stock solution in 25 mM PB pH 7.4. Finally, 62.5 µL of 25 mM PB pH 7.4 was added to the reaction to give a final reaction volume of 200 µL. The solution was mixed *via* repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Conversion to the expected Umpolung product was judged by LC-MS.

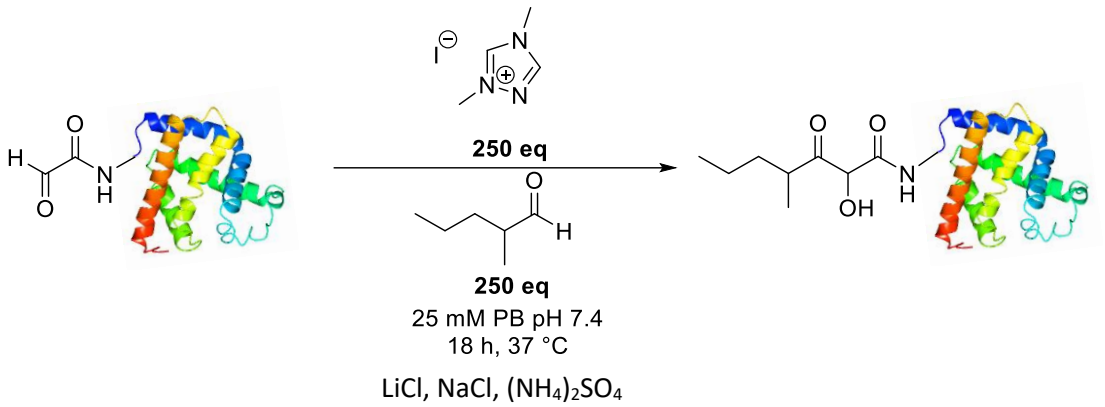
Tabulated results of myoglobin catalyst screen with isobutyraldehyde 46



Catalyst	Catalyst eq	Aldehyde eq	Conversion
	**125**	 **250**	**75%**
	**250**	 **250**	**75%**
	**125**	 **250**	**57%**
	**250**	 **250**	**63%**
	**125**	 **250**	**35%**
	**250**	 **250**	**38%**
	**125**	 **250**	**84%**
	**250**	 **250**	**100%**
	**125**	 **250**	**47%**
	**250**	 **250**	* **100%**

*Caused protein to precipitate

Salt effects on reaction rate of cross-benzoin



In a 0.5 mL Eppendorf, a 100 µL aliquot of a 200 µM α-oxo aldehyde-myoglobin **55** stock in PB pH 7.4 was charged with 25 µL of a 200 mM dimethyl triazolium **35** in 25 mM PB pH 7.4. The solution was then charged with 25 µL aliquot of a 200 mM 2-methylpentanal **62** donor stock solution in 25 mM PB pH 7.4. Finally, 50 µL of buffer was added containing desired concentration of salt to give a final reaction volume of 200 µL. The solution was mixed *via* repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Conversion to the expected Umpolung product was judged by LC-MS at hourly intervals.

NaCl: % conversion over time.

	0.5 M	1 M	2M
0 h	0 %	0 %	0 %
1 h	15 %	9 %	0 %
2 h	15 %	13 %	0 %
3 h	16 %	13 %	0 %
4 h	24 %	14 %	0 %
5 h	24 %	16 %	0 %
6 h	28 %	16 %	0 %
18 h	40 %	29 %	0 %

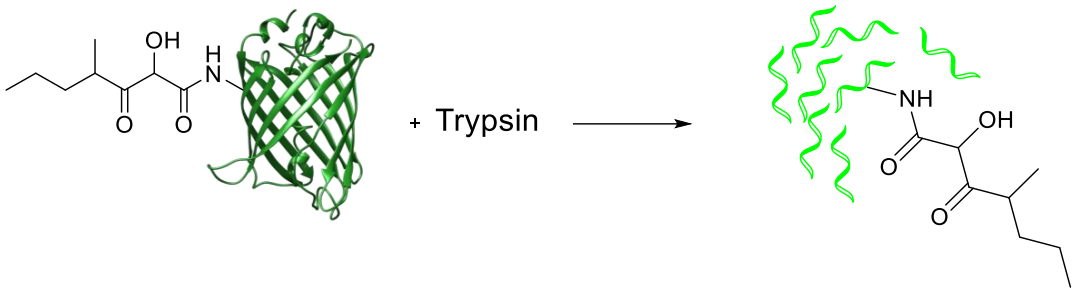
LiCl: % conversion over time.

	0.5 M	1 M	2M
0 h	0 %	0 %	0 %
1 h	13 %	9 %	0 %
2 h	18 %	20 %	0 %
3 h	18 %	21 %	0 %
4 h	18 %	21 %	0 %
5 h	18 %	25 %	0 %
6 h	18 %	28 %	0 %
18 h	21 %	36 %	0 %

$(NH_4)_2SO_4$: % conversion over time

	0.5 M	1 M	2M
0 h	0 %	0 %	0 %
1 h	36 %	51 %	40 %
2 h	54 %	86 %	75 %
3 h	81 %	100 %	93 %
4 h	86 %	100 %	100 %
5 h	100 %	100 %	100 %
6 h	100%	100 %	100 %
18 h	100%	100 %	100 %

Trypsin Digest of alkylated sfGFP α-oxo K



Each protein solution was diluted 1:1 (v:v) with aqueous 100 mM ammonium bicarbonate before reducing with 5 mM DTT for 30 min at 50 °C and alkylating with 15 mM Iodoacetamide for 30 min in the dark at room temperature. Post alkylation, digestion was performed with the addition of 0.2 micrograms of sequencing grade modified trypsin (Promega) and incubation overnight at 37 °C. Protease activity was stopped with aqueous trifluoroacetic acid (1%, v:v) before LC-MS/MS analysis.

The resulting peptides were loaded onto a 50 cm PepMap, 2 µm, 100 Å, C18 EasyNano nanocapillary column (75 um x 150 mm, Thermo). The peptides were eluted onto an Orbitrap Fusion hybrid mass spectrometer (Thermo) over a 35 min gradient of aqueous 3-35% (v:v) acetonitrile. Both MS1 and MS2 spectra were acquired in the Orbitrap mass analyser with Easy-IC internal calibration. Data dependent acquisition was performed in top speed mode using a fixed 1 s cycle, selecting the most intense precursors with charge states 2-5. HCD was then used for peptide fragmentation with 32% activation energy.

The resulting spectral data were searched against the expected protein sequence appended to an in house database using the PEAKS-DB search program. The search criteria specified: Enzyme, trypsin; Peptide tolerance, 3 ppm; MS/MS tolerance, 3 mDa. Carbamidomethylation (C), was set as a fixed modification.

The following modifications were set as variable as fitting sample possibilities:

Lys -> Glyoxyl(ald) +55.989829 Da

Lys -> Dicarbonyl +156.078644 Da

Oxidation of Met was also considered

Solid phase peptide synthesis

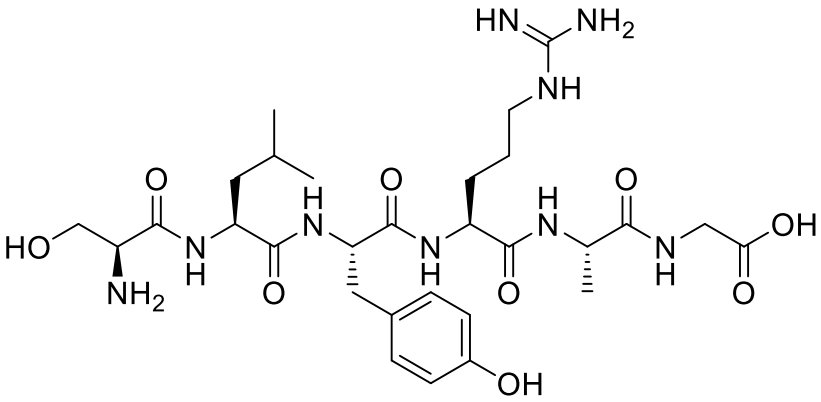
Preloaded. The resin was swollen in DMF for 30 min and then filtered.

Amino acid coupling. DIPEA (11 eq.) was added to a solution of amino acid (5 eq. unless otherwise stated) and HCTU (5 eq. unless otherwise stated) 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 was 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: The resin was 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 was added to the resin and gently agitated by rotation for 60 min. The reaction mixture was drained into ice-cold Et_2O and centrifuged at 4000 rpm at 4 °C until pelleted (ca. 5-10 min). The supernatant was carefully decanted and subsequently re-suspended, centrifuged and supernatant decanted three more times. The precipitated peptide pellet was dissolved in 10% aq. AcOH and lyophilised to obtain a powder.

Cleavage Cocktails: General cleavage cocktail: 95:2.5:2.5 TFA:H_2O:triisopropylsilane. Reductive cleavage cocktail: 88:5:5:2 TFA:H_2O:dithiothreitol:triisopropylsilane. Resin cleavage only: 1:4 1,1,1,3,3,3-HFIP:DCM (all volumes v/v).

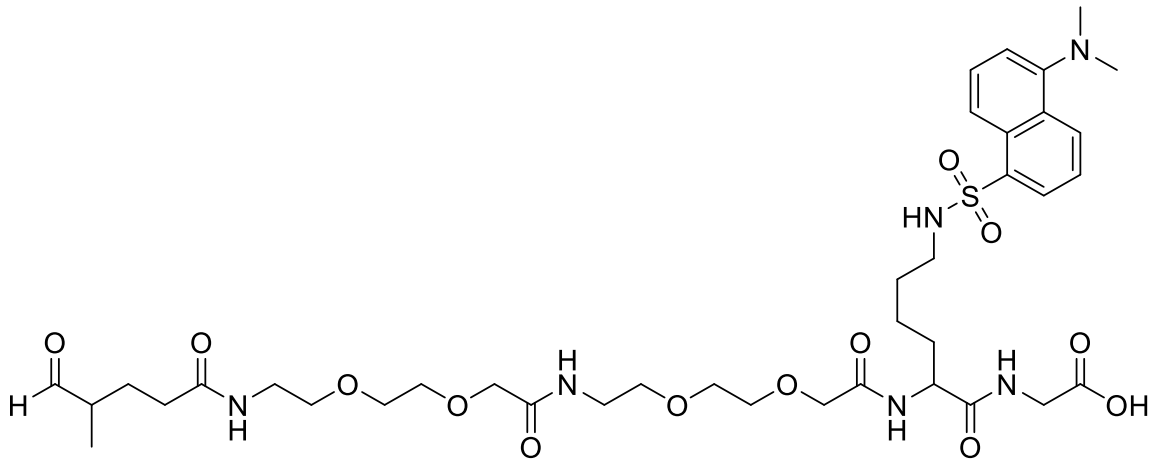


H_2N

Resin:	Preloaded-G	Glycine		Name:	
Loading:	0.53			**FW**	**mg**
mmole:	0.053	**Peptide Mass:**		665.8	35.3
No Residues:	6	**+ N-Mod:**		665.8	35.3

Residue	FW	Eq	mg	CA	Q	DIPEA
Glycine	-		100		mg	µl
Fmoc-Ala-OH	311.17	5	82	HCTU	107	98
Fmoc-Arg (Pbf)-OH	648.80	5	172	HCTU	107	98
Fmoc-Tyr (tBu)-OH	459.27	5	122	HCTU	107	98
Fmoc-Leu-OH	353.25	5	94	HCTU	107	98
Fmoc-Ser (tBu)-OH	383.17	5	102	HCTU	107	98

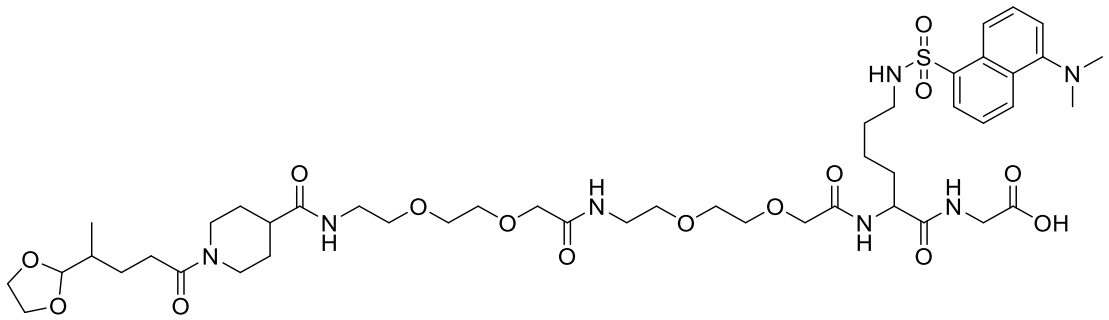
N-Terminus	FW	Eq	mg	CA		Q	DIPEA
none	-	5		HCTU			98
				µL			



Resin:	Preloaded-G	Glycine	**Name:**	
Loading:	0.53		**FW**	**mg**
mmole:	0.053	**Peptide Mass:**	882	35.3
No Residues:	6	**+ N-Mod:**	1012.9	35.3

Residue	**FW**	**Eq**	**mg**	**CA**	**Q**	**DIPEA**
Glycine	-		85		mg	µl
Fmoc-DansLys-OH	311.17	5	132	HCTU	82	90
Fmoc-Peg2-OH	648.80	5	90	HCTU	82	90
Fmoc-Peg2-OH	459.27	5	90	HCTU	82	90
Handle	353.25	5	38	HCTU	82	90

N-Terminus	**FW**	**Eq**	**mg**	**CA**	**Q**	**DIPEA**
none	-	5		HCTU	#REF!	90
				µL		



H_2N-SI

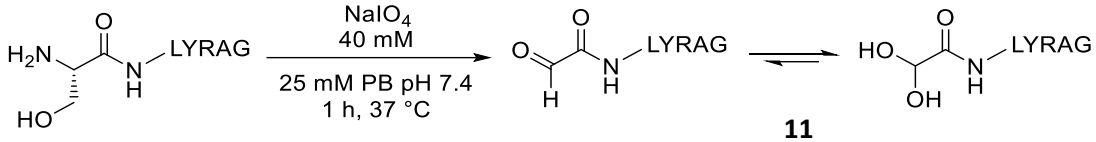
Resin:	Preloaded-G	Glycine	Name:	
Loading:	0.53		**FW**	**mg**
mmole:	0.053	**Peptide Mass:**	665.8	35.3
No Residues:	6	**+ N-Mod:**	665.8	35.3

Residue	FW	Eq	mg	CA	Q	DIPEA
Glycine	-		85		mg	µl
Fmoc-DansLys-OH	311.17	5	132	HCTU	107	90
Fmoc-Peg2-OH	648.80	5	85	HCTU	107	90
Fmoc-Peg2-OH	459.27	5	85	HCTU	107	90
FMOC-INP-OH	353.25	5	90	HCTU	107	90
Handle	383.17	5	38	HCTU	107	90

N-Terminus	FW	Eq	mg	CA	Q	DIPEA
none	-	5		HCTU		98
				µl		

Peptide modifications

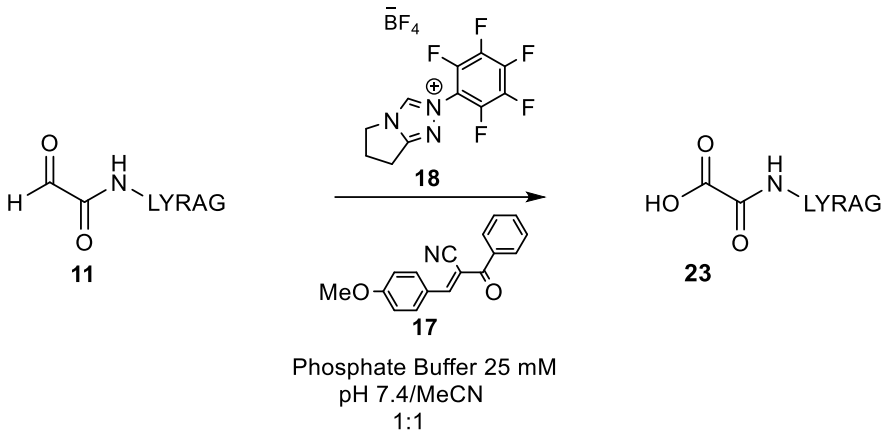
Oxidation of SLYRAG



9 **11**

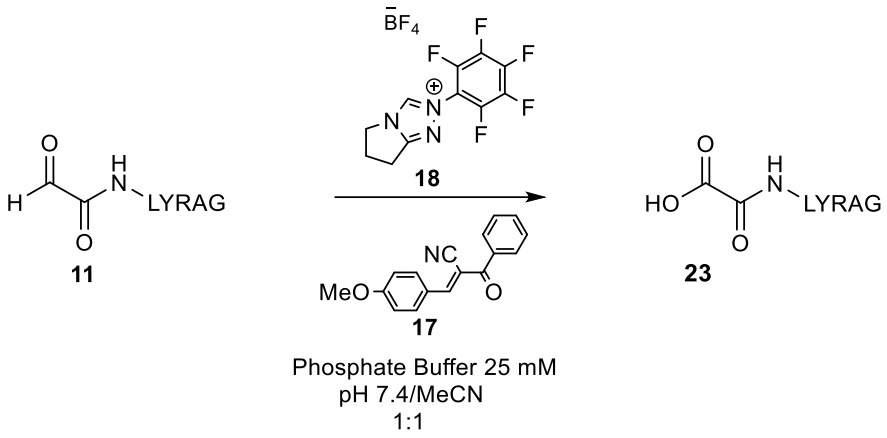
An aliquot of SLYRAG **9** (26.6 mg, 40mM) was charged with sodium periodate (8.5 mg, 40 mM) in 1 mL of 25 mM PB pH 7.4. The solution was vortexed, then allowed to sit at room temperature in the dark for 1 h. The solution was then loaded onto a solid phase extraction C18 cartridge (Supelclean™, 6 mL, LC-18, SUPELCO®) equilibrated with water then acetonitrile. After initial washing with water, the product was eluted over a gradient: H_2O:10% MeCN, H_2O:30% MeCN, H_2O:50% MeCN, H_2O:70% MeCN, and 100% MeCN. The fragments are collected and left to lyophilise overnight and LC-MS used to identify fractions containing α-oxo aldehyde LYRAG **11**. Fractions containing pure α-oxo aldehyde-LYRAG were pooled and subsequently lyophilised to give α-oxo aldehyde-LYRAG **11** as an orange solid, which was stored at -20 °C until required.

Stetter modifications of α-oxo aldehyde-LYRAG 11
20 mM catalyst loading



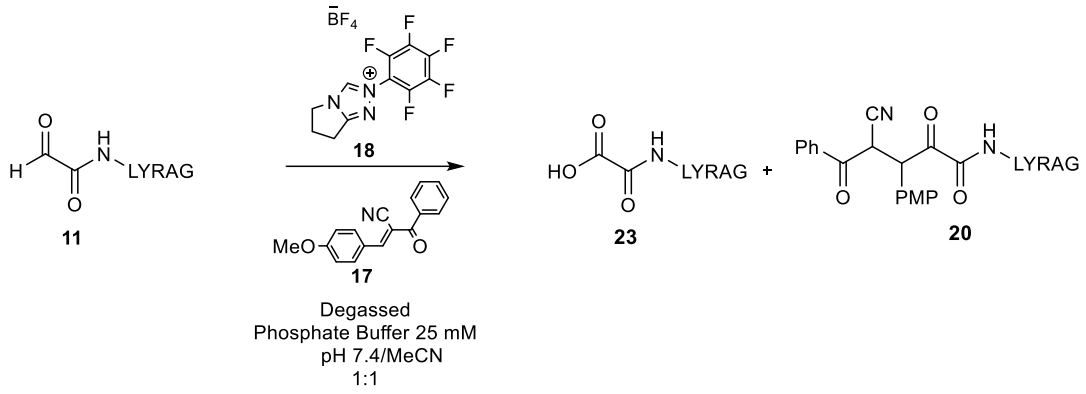
From a 5 mM α-oxo aldehyde-LYRAG **11** stock in 25 mM PB, pH 7.4, a 500 µL aliquot was charged with 500 µL aliquot of a 5 Mm Michael acceptor **17** stock solution in acetonitrile. The reaction was then charged with 6,7-dihydro-2-pentafluorophenyl-5H-pyrrolo[2,1-c]-1,2,4-triazolium tetrafluoroborate **18** (20 mM, 7.25 mg). The solution was briefly vortexed (~10 secs), and allowed to sit at 37 °C overnight without further agitation. After this period the solution was analysed *via* LC-MS without purification.

Stetter modifications of α-oxo aldehyde-LYRAG 11
5 mM catalyst loading



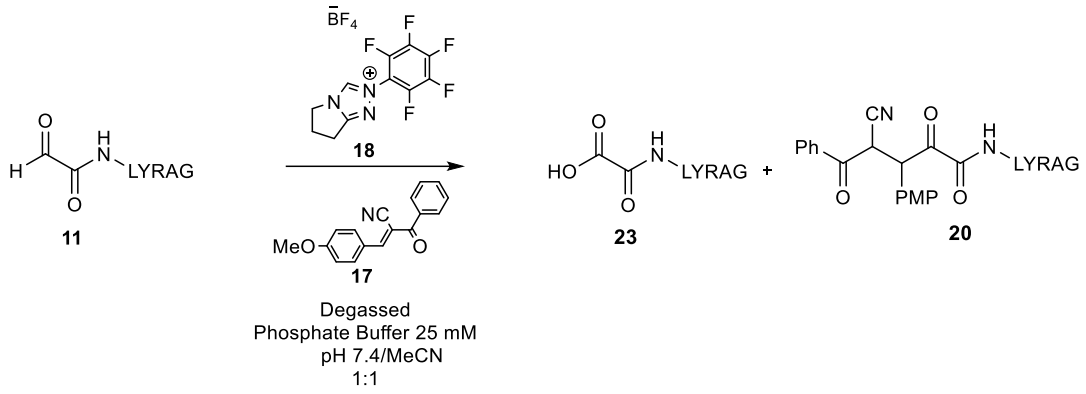
From a 5 mM α-oxo aldehyde-LYRAG **11** stock in 25 mM PB, pH 7.4, a 500 µL aliquot was charged with 500 µL aliquot of a 5 Mm Michael acceptor **17** stock solution in acetonitrile. The reaction was then charged with 6,7-dihydro-2-pentafluorophenyl-5H-pyrrolo[2,1-c]-1,2,4-triazolium tetrafluoroborate **18** (5 mM, 1.8 mg). The solution was briefly vortexed (~10 secs), and allowed to sit at 37 °C overnight without further agitation. After this period conversion to anticipated Umpolung products were assessed by LC-MS without purification.

Stetter modifications of α-oxo aldehyde-LYRAG 11
20 mM catalyst loading-degassed buffer



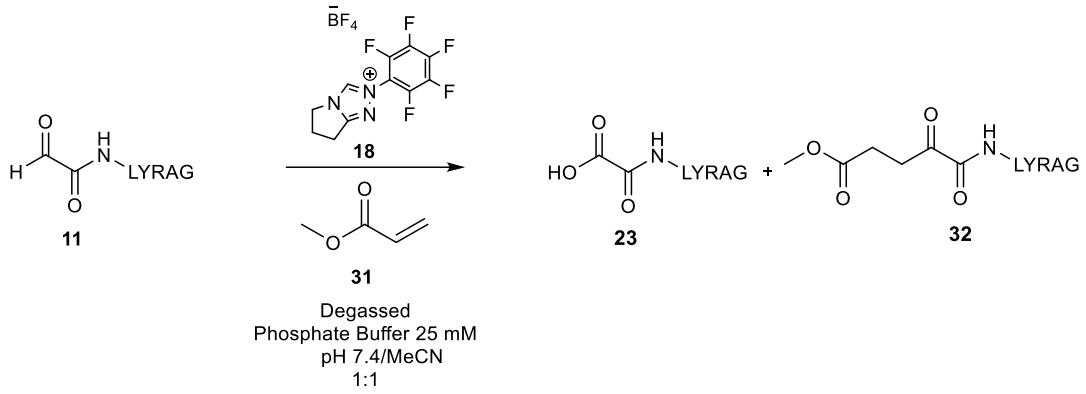
A 5 mM α-oxo aldehyde-LYRAG **11** stock in 25 mM PB, pH 7.4, was degassed under nitrogen for 1 hour. Following this period, a 500 µL aliquot was transferred into a 1.5 mL eppendorf containing 6,7-dihydro-2-pentafluorophenyl-5H-pyrrolo[2,1-c]-1,2,4-triazolium tetrafluoroborate **18** (20 mM, 7.25 mg), under inert atmosphere. This was charged with 500 µL aliquot of a 5 Mm Michael acceptor **17** stock solution in dry acetonitrile. The solution was briefly vortexed (~10 secs), and allowed to sit at 37 °C overnight without further agitation. After this period conversion to anticipated Umpolung products were assessed by LC-MS without purification.

Stetter modifications of α-oxo aldehyde-LYRAG 11
5 mM catalyst loading-degassed buffer



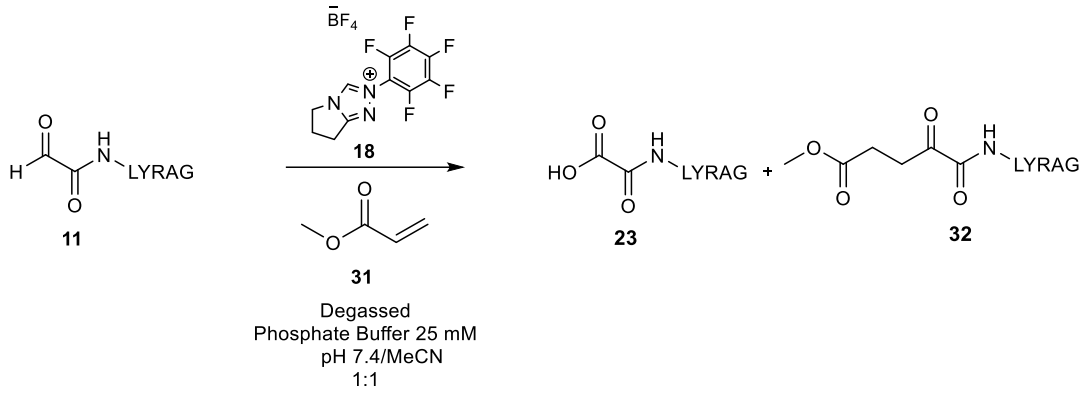
A 5 mM α-oxo aldehyde-LYRAG **11** stock in 25 mM PB, pH 7.4, was degassed under nitrogen for 1 hour. Following this period, a 500 µL aliquot was transferred into a 1.5 mL eppendorf containing 6,7-dihydro-2-pentafluorophenyl-5H-pyrrolo[2,1-c]-1,2,4-triazolium tetrafluoroborate **18** (5 mM, 1.8 mg), under inert atmosphere. This was charged with 500 µL aliquot of a 5 Mm Michael acceptor **17** stock solution in dry acetonitrile. The solution was briefly vortexed (~10 secs), and allowed to sit at 37 °C overnight without further agitation. After this period conversion to anticipated Umpolung products were assessed by LC-MS without purification.

Stetter modifications of α-oxo aldehyde-LYRAG 11
20 mM catalyst loading-degassed buffer



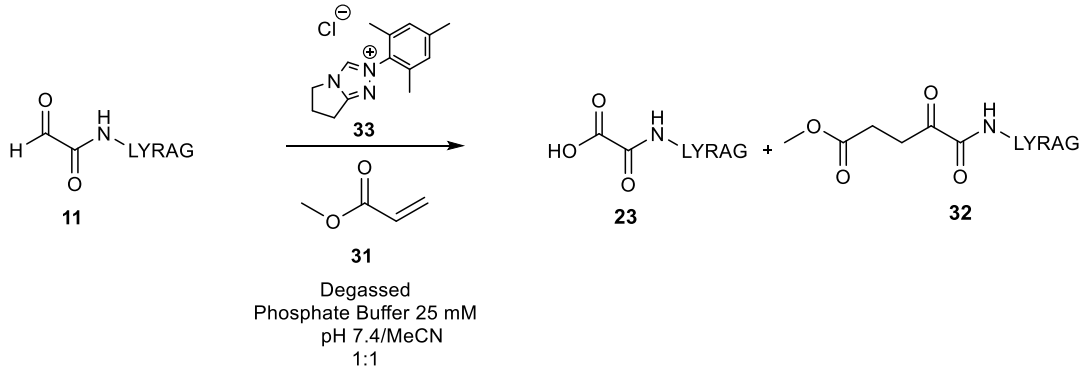
A 5 mM α-oxo aldehyde-LYRAG **11** stock in 25 mM PB, pH 7.4, was degassed under nitrogen for 1 hour. Following this period, a 500 µL aliquot was transferred into a 1.5 mL eppendorf under inert atmosphere containing 6,7-dihydro-2-pentafluorophenyl-5H-pyrrolo[2,1-c]-1,2,4-triazolium tetrafluoroborate **18** (20 mM, 7.25 mg). This was charged with a 500 µL aliquot of a 40 mM methyl acrylate **31** stock solution in dry acetonitrile. The solution was briefly vortexed (~10 secs), and allowed to sit at 37 °C overnight without further agitation. After this period conversion to anticipated Umpolung products were assessed by LC-MS without purification.

Stetter modifications of α-oxo aldehyde-LYRAG 11
5 mM catalyst loading-degassed buffer



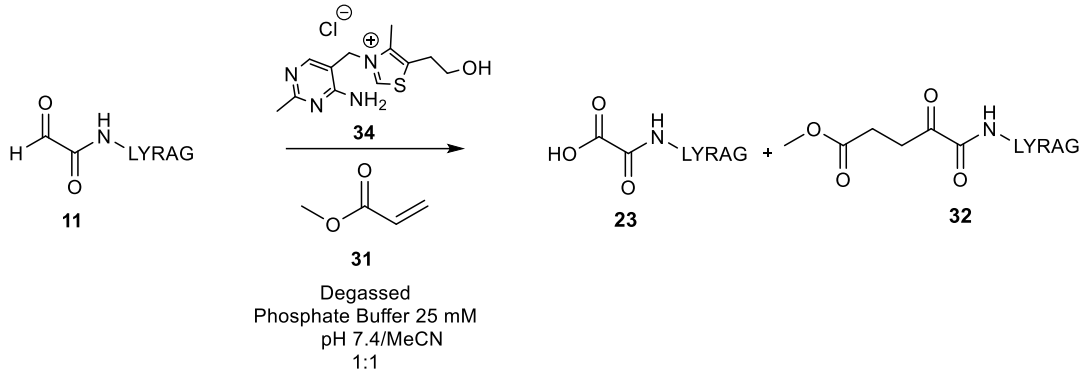
A 5 mM α-oxo aldehyde-LYRAG **11** stock in 25 mM PB, pH 7.4, was degassed under nitrogen for 1 hour. Following this period, a 500 µL aliquot was transferred into a 1.5 mL eppendorf under inert atmosphere containing 6,7-dihydro-2-pentafluorophenyl-5H-pyrrolo[2,1-c]-1,2,4-triazolium tetrafluoroborate **18** (5 mM, 1.8 mg). This was charged with a 250 µL aliquot of an 80 mM methyl acrylate **31** stock solution in dry acetonitrile. The solution was briefly vortexed (~10 secs), and allowed to sit at 37 °C overnight without further agitation. After this period conversion to anticipated Umpolung products were assessed by LC-MS without purification.

Stetter modifications of α-oxo aldehyde-LYRAG 11
5 mM catalyst loading-degassed buffer



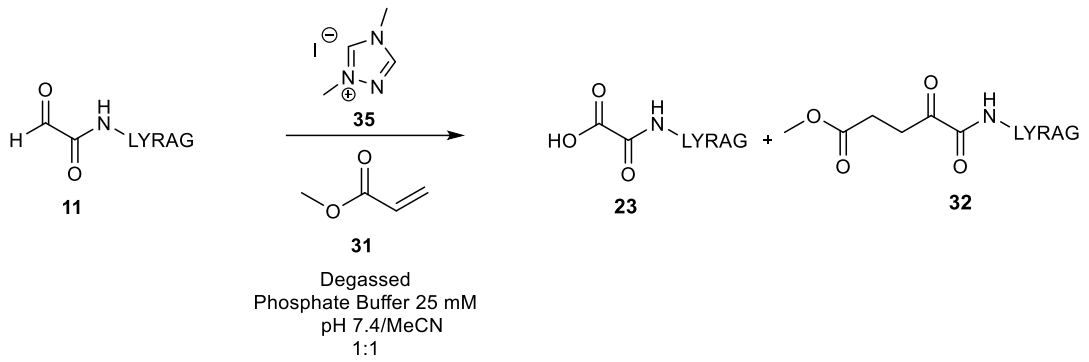
A 5 mM α-oxo aldehyde-LYRAG **11** stock in 25 mM PB, pH 7.4, was degassed under nitrogen for 1 hour. Following this period, a 500 µL aliquot was transferred into a 1.5 mL eppendorf under inert atmosphere. To this, a 125 µL aliquot from a 40 mM stock solution containing mesityl substituted triazolium **33** in dry acetonitrile was then transferred into the reaction. This was charged with a 250 µL aliquot of an 80 mM methyl acrylate **31** stock solution in dry acetonitrile. Finally, 125 µL of dry acetonitrile was added to generate a final reaction volume of 1 mL. The solution was briefly vortexed (~10 secs), and allowed to sit at 37 °C overnight without further agitation. After this period conversion to anticipated Umpolung products were assessed by LC-MS without purification.

Stetter modifications of α-oxo aldehyde-LYRAG 11
5 mM catalyst loading-degassed buffer



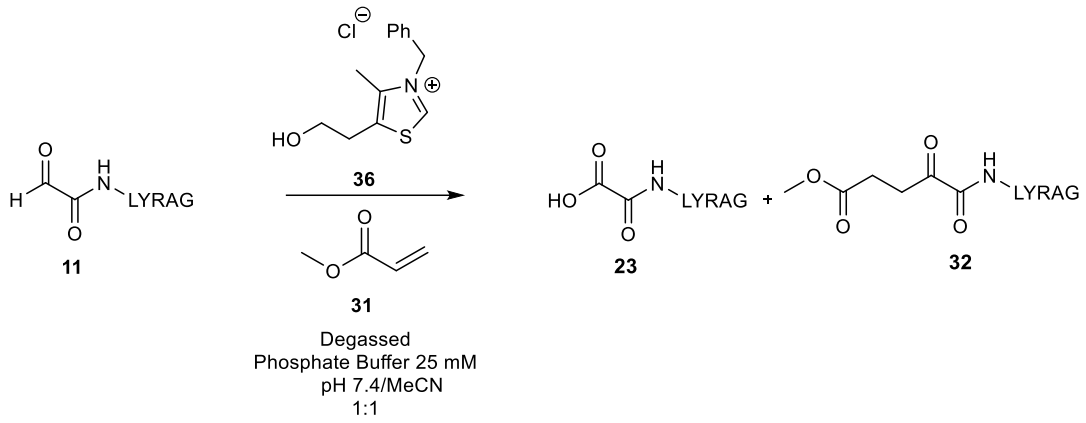
A 5 mM α-oxo aldehyde-LYRAG **11** stock in 25 mM PB, pH 7.4, was degassed under nitrogen for 1 hour. Following this period, a 500 µL aliquot was transferred into a 1.5 mL eppendorf under inert atmosphere. To this, a 125 µL aliquot from a 40 mM stock solution containing mesityl substituted thiamine **34** in dry acetonitrile was then transferred into the reaction. This was charged with a 250 µL aliquot of an 80 mM methyl acrylate **31** stock solution in dry acetonitrile. Finally, 125 µL of dry acetonitrile was added to generate a final reaction volume of 1 mL. The solution was briefly vortexed (~10 secs), and allowed to sit at 37 °C overnight without further agitation. After this period conversion to anticipated Umpolung products were assessed by LC-MS without purification.

Stetter modifications of α-oxo aldehyde-LYRAG 11
5 mM catalyst loading-degassed buffer



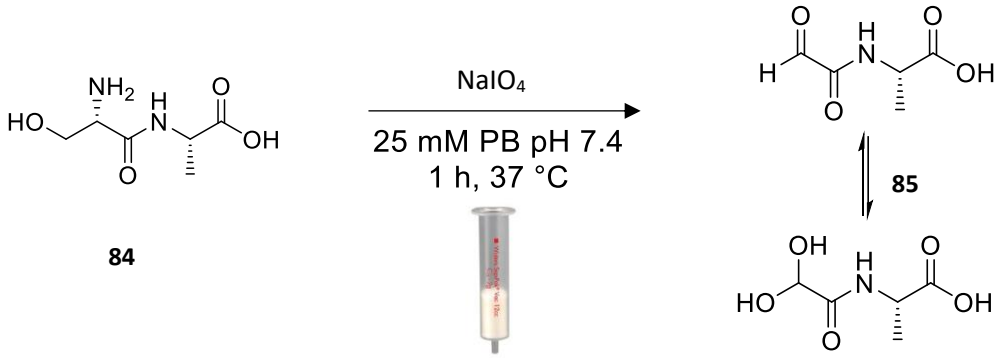
A 5 mM α-oxo aldehyde-LYRAG **11** stock solution in 25 mM PB, pH 7.4, was degassed under nitrogen for 1 hour. Following this period, a 500 µL aliquot was transferred into a 1.5 mL eppendorf under inert atmosphere. To this, a 125 µL aliquot from a 40 mM stock solution containing mesityl substituted dimethyl triazolium **35** in dry acetonitrile was then transferred into the reaction. This was charged with a 250 µL aliquot of an 80 mM methyl acrylate **31** stock solution in dry acetonitrile. Finally, 125 µL of dry acetonitrile was added to generate a final reaction volume of 1 mL. The solution was briefly vortexed (~10 secs), and allowed to sit at 37 °C overnight without further agitation. After this period conversion to anticipated Umpolung products were assessed by LC-MS without purification.

Stetter modifications of α-oxo aldehyde-LYRAG 11
5 mM catalyst loading-degassed buffer



A 5 mM α-oxo aldehyde-LYRAG **11** stock solution in 25 mM PB, pH 7.4, was degassed under nitrogen for 1 hour. Following this period, a 500 µL aliquot was transferred into a 1.5 mL eppendorf under inert atmosphere. To this, a 125 µL aliquot from a 40 mM stock solution containing mesityl substituted thiazolium catalyst **36** in dry acetonitrile was then transferred into the reaction. This was charged with a 250 µL aliquot of an 80 mM methyl acrylate **31** stock solution in dry acetonitrile. Finally, 125 µL of dry acetonitrile was added to generate a final reaction volume of 1 mL. The solution was briefly vortexed (~10 secs), and allowed to sit at 37 °C overnight without further agitation. After this period conversion to anticipated Umpolung products were assessed by LC-MS without purification.

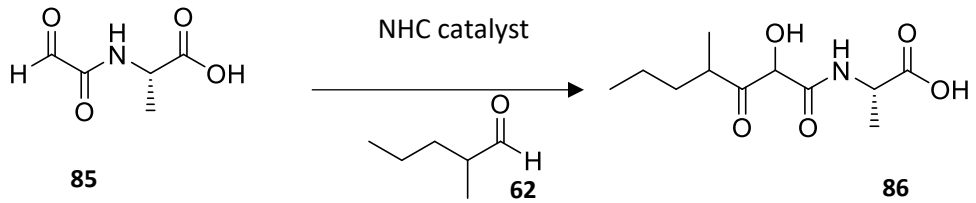
Oxidation of Ser-Ala Dipeptide



84

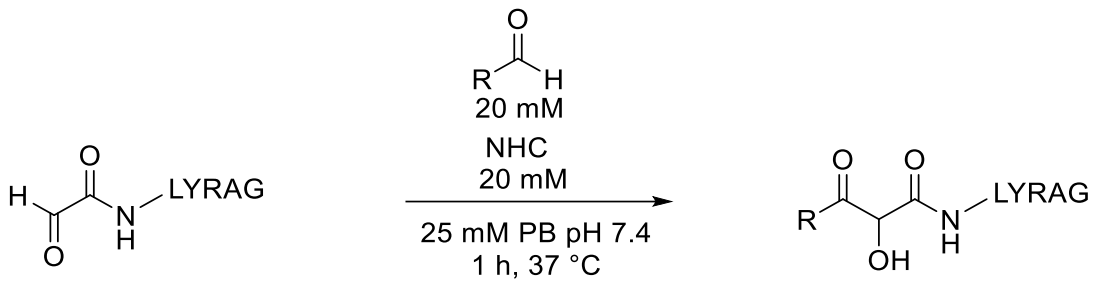
To a solution of compound **84** (0.05 g, 0.27 mmol) in 0.1 M PB pH 7.0 (700 µL) was added $NaIO_4$ (0.06 g, 0.28 mmol). The reaction was mixed until complete dissolution was achieved, and then allowed to sit at rt for 1 h. Complete oxidation to **85** was observed by LC-MS analysis. Upon completion, he solution was then loaded onto a solid phase extraction C18 cartridge (Supelclean™, 6 mL, LC-18, SUPELCO®) equilibrated with water then acetonitrile. After initial washing with water, the product was eluted over a gradient: H_2O:10% MeCN, H_2O:30% MeCN, H_2O:50% MeCN, H_2O:70% MeCN, and 100% MeCN. The fragments were collected and left to lyophilise overnight and LC-MS used to identify fractions containing desired product **85**.

Umpolung of α-oxo aldehyde -Ala Dipeptide



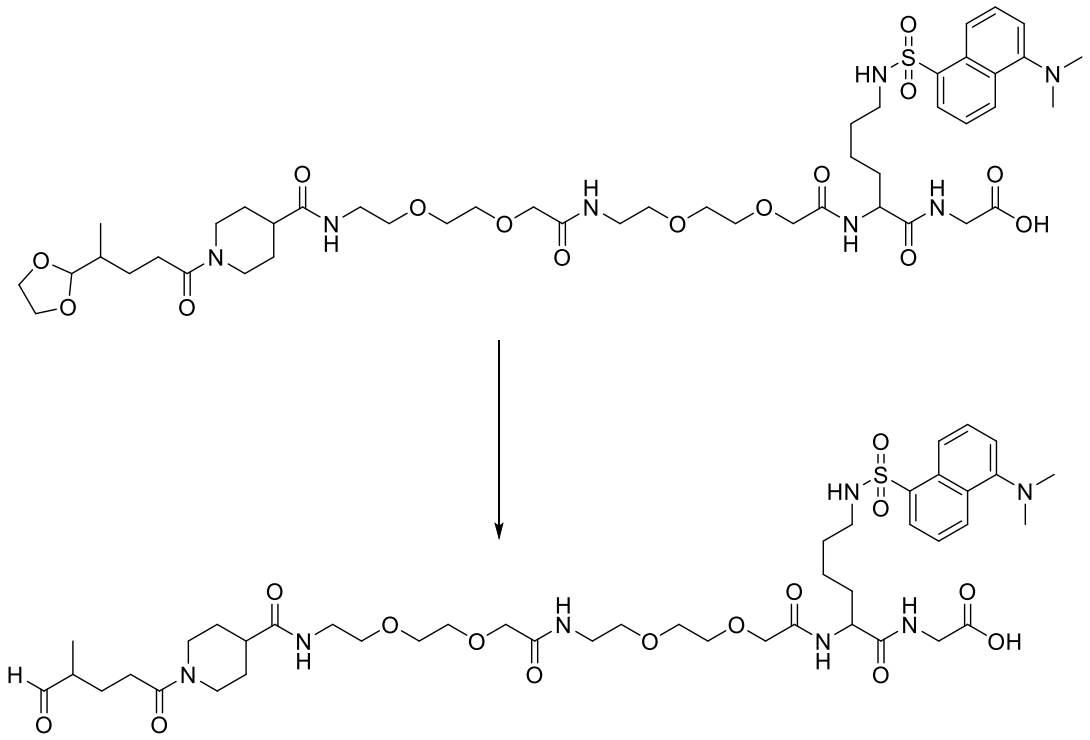
To a solution of compound **85** (25 mg, 0.04 mmol) in 3 mL 25 Mm PB pH 7.4 was added thiamine **34** (40 mg, 0.15 mmol) and 2-methylpentanal **62** (15 mg, 0.15 mmol). The solution was briefly vortexed (~10 secs), and allowed to sit at 37 °C. without further agitation. Conversion to **86** was monitored by LC-MS analysis. Upon completion, he solution was then loaded onto a solid phase extraction C18 cartridge (Supelclean™, 6 mL, LC-18, SUPELCO®) equilibrated with water then acetonitrile. After initial washing with water, the product was eluted over a gradient: H_2O:10% MeCN, H_2O:30% MeCN, H_2O:50% MeCN, H_2O:70% MeCN, and 100% MeCN. The fragments were collected and left to lyophilise overnight and LC-MS used to identify fractions containing desired product **86**.

General procedure 4 – Umpolung reaction of α-oxo aldehyde-LYRAG 11 with small molecule aldehyde donors



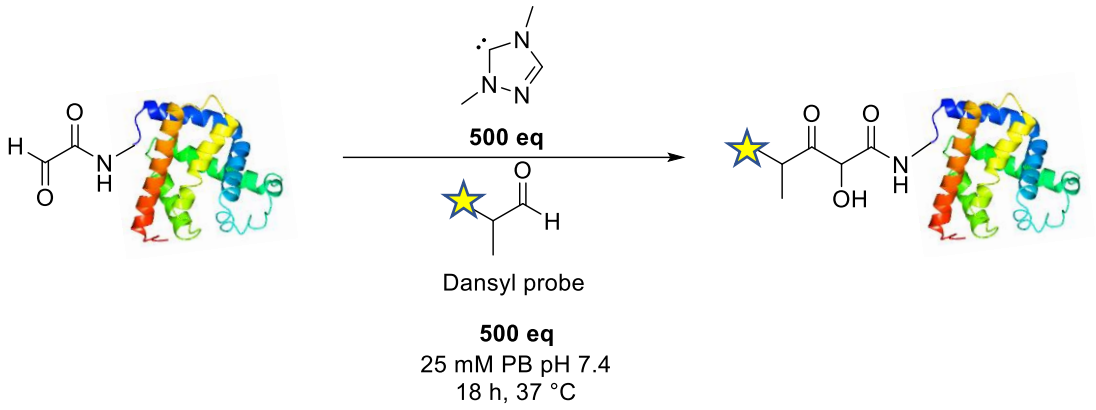
11

From a 5 mM α-oxo aldehyde-LYRAG **11** stock in 25 mM PB, pH 7.4, a 200 µL aliquot was charged with 40 µL of a 0.5 M aldehyde donor stock solution in 25 mM PB pH 7.4. The reaction was then charged with 100 µL of a 200 mM NHC catalyst in 25 mM PB pH 7.4. Finally, 690 µL 25 mM PB pH 7.4 was added to create a final volume of 1 mL. The solution was briefly vortexed (~10 secs), and allowed to sit at 37 °C overnight without further agitation. After this period Conversion to anticipated Umpolung products were confirmed by LC-MS without purification.



1,3 dioxolane protected dansyl probe **83** (5 mg), was charged with 140 µL of 25 mM PB pH 7.4 in a 1 mL Eppendorf tube and then neat TFA (10 µL) was added to give a final volume of 150 µL. The solution was vortexed, then allowed to sit at room temperature for 1 h. The generation of dansyl probe was monitored using LC-MS analysis.

Following completion, 0.5 mL 10% AcOH was added and the solution was lyophilised.



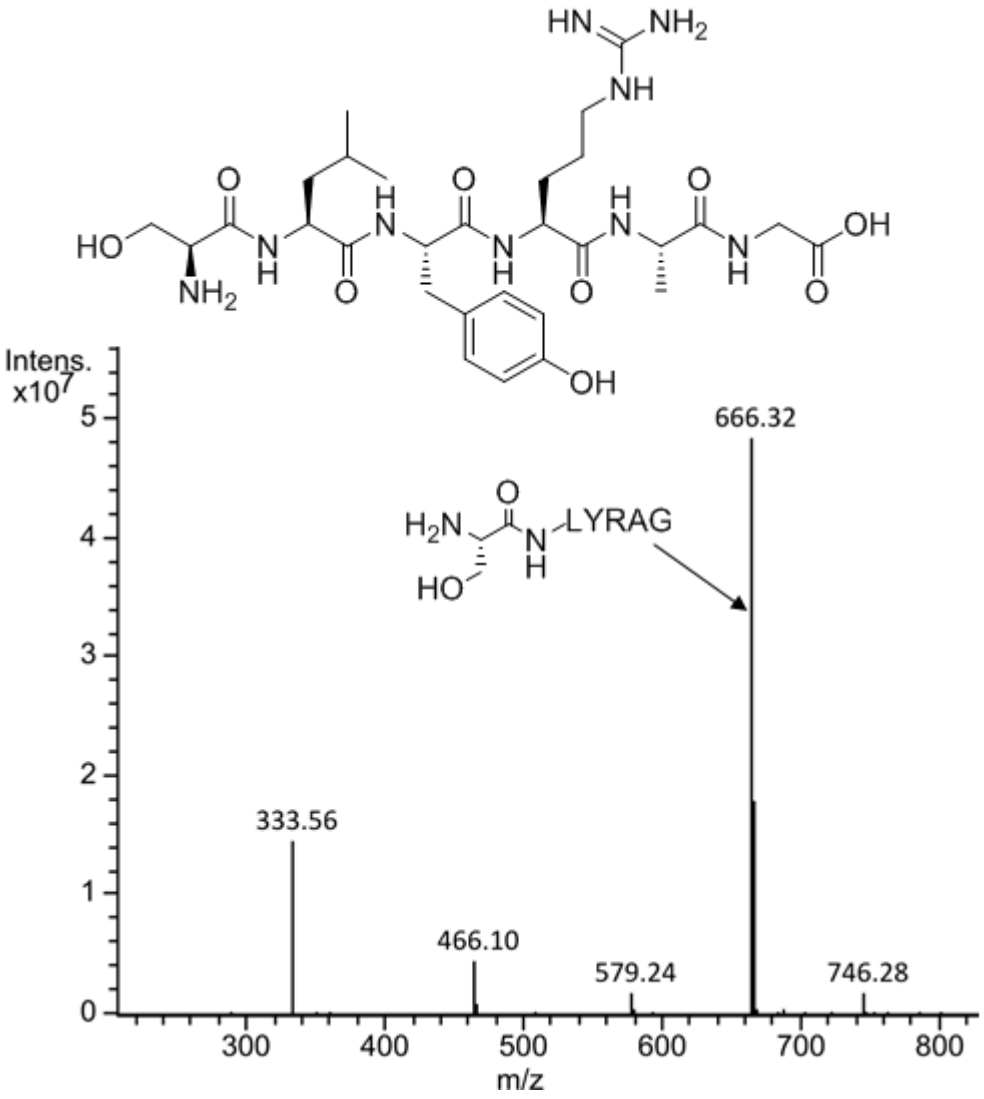
In a 0.5 mL Eppendorf, a 53 µL aliquot of a 200 µM α-oxo aldehyde-myoglobin **55** stock in PB pH 7.4 was charged with 26.6 µL of a 200 mM triazolium catalyst in 25 mM PB pH 7.4. To this, 5 mg dansyl probe was dissolved in 50 µL 25 mM PB pH 7.4 and added to the solution. Finally, 295.4 µL of 25 mM PB pH 7.4 was added to the reaction to give a final reaction volume of 425 µL. The solution was mixed *via* repeatedly pipetting up and down, then incubated at 37 °C without further agitation for 18 h. Following this the solution was dialysed at 4 °C overnight and concentrated in an 10 KDa MWCO. Conversion to the expected Umpolung product was judged by LC-MS.

Mass spectrometry data of modified peptides

SLYRAG **9**

Calculated $[M+H]^+$ 666.36

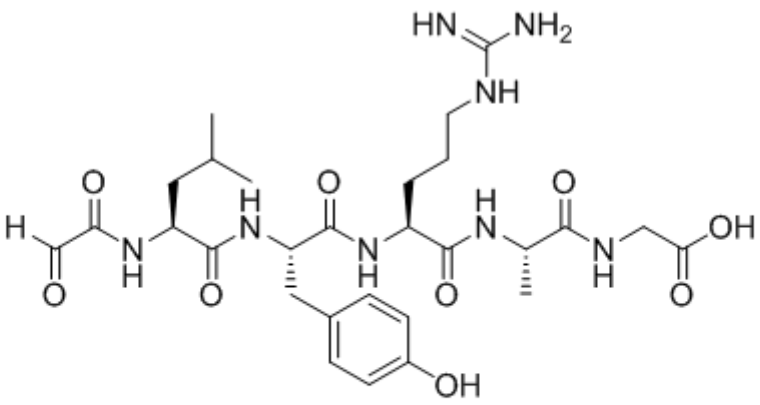
Found $[M+H]^+$ 666.32

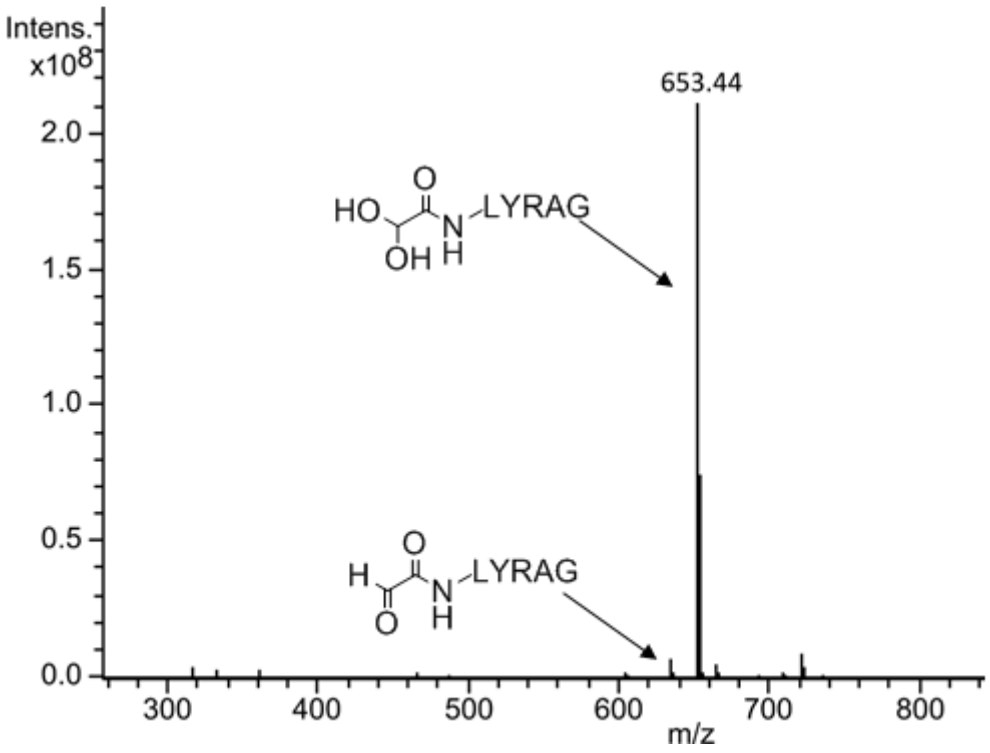


α-Oxo-LYRAG **11**

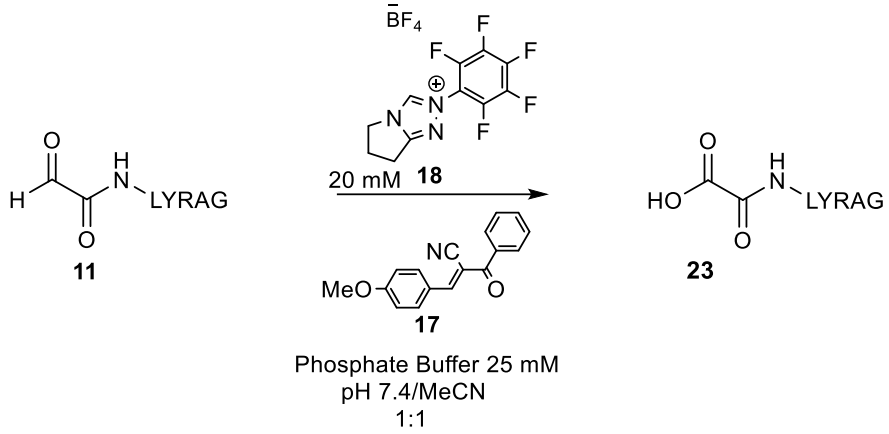
Calculated $[M+H]^+$ 635.31, 653.32 (hydrate)

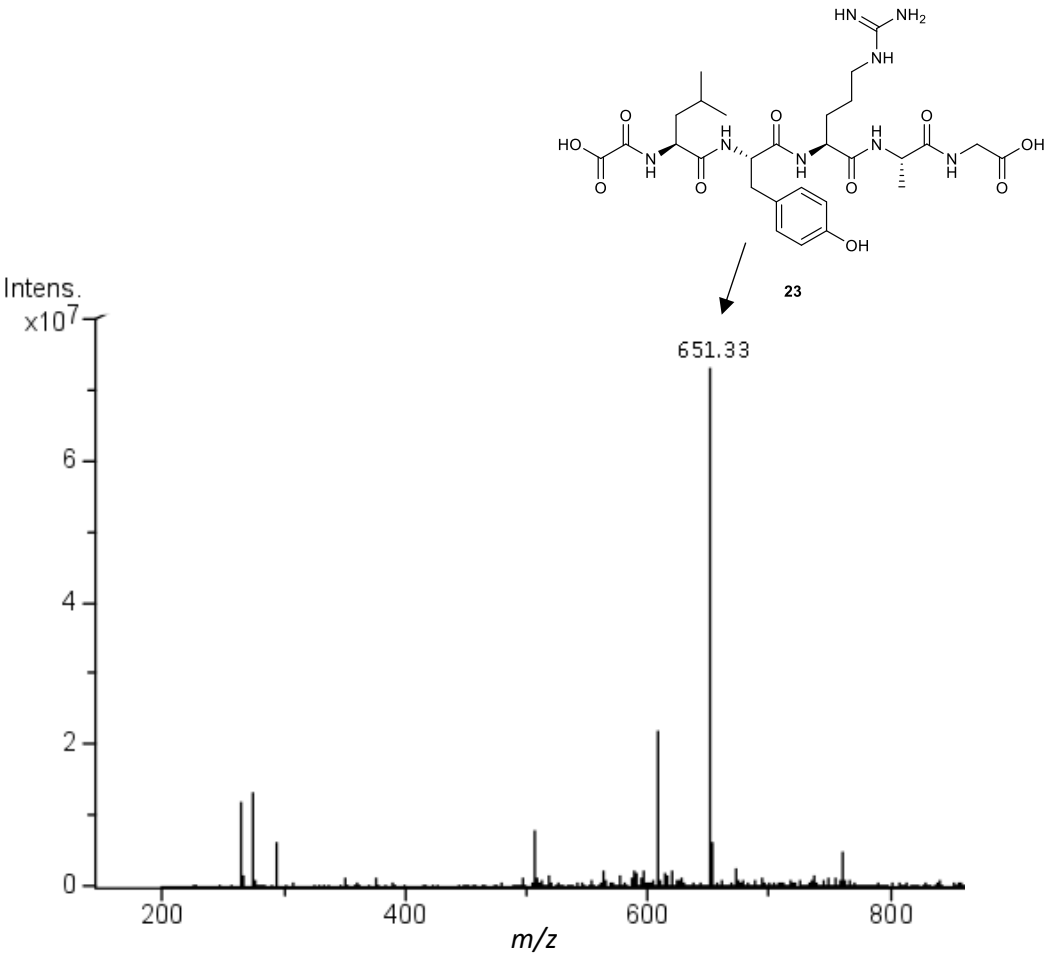
Found $[M+H]^+$ 635.42, 653.44 (hydrate)

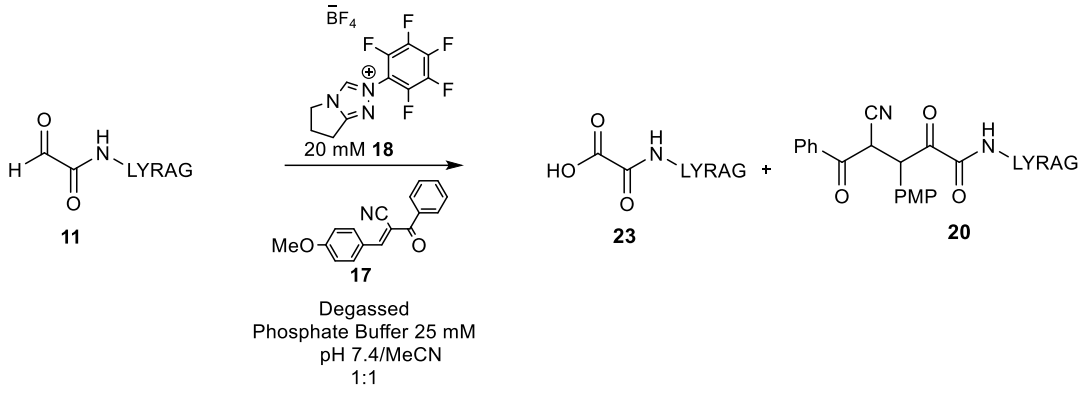


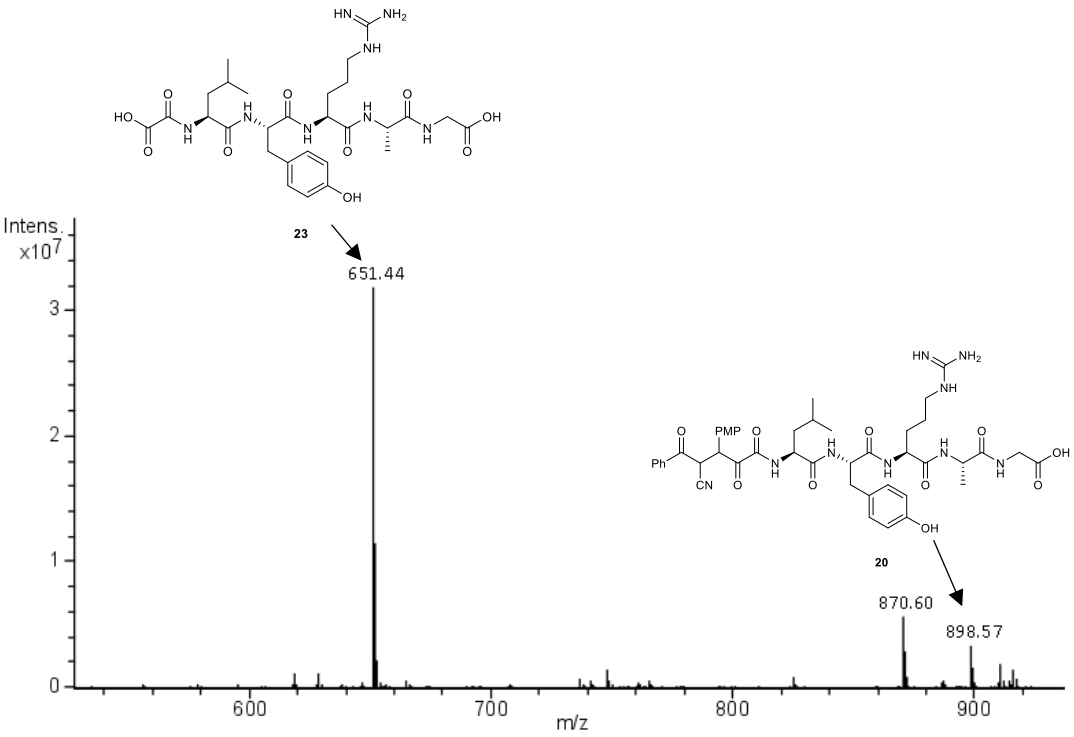


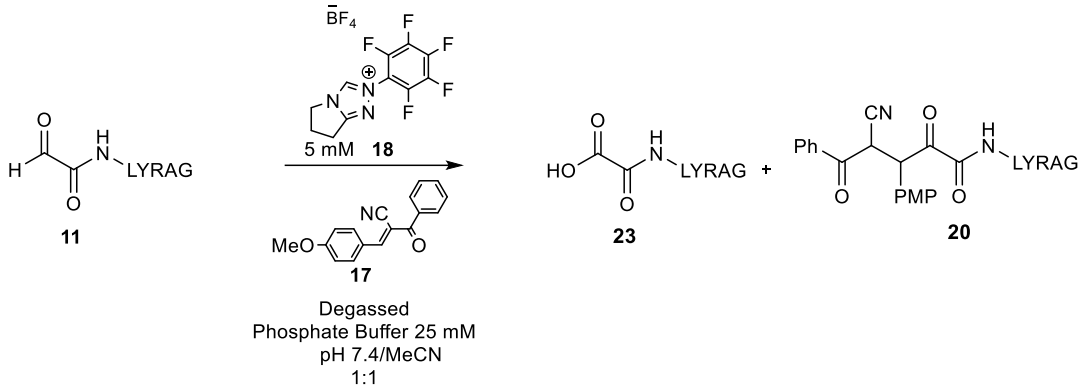
Stetter modifications of α-oxo aldehyde-LYRAG 11

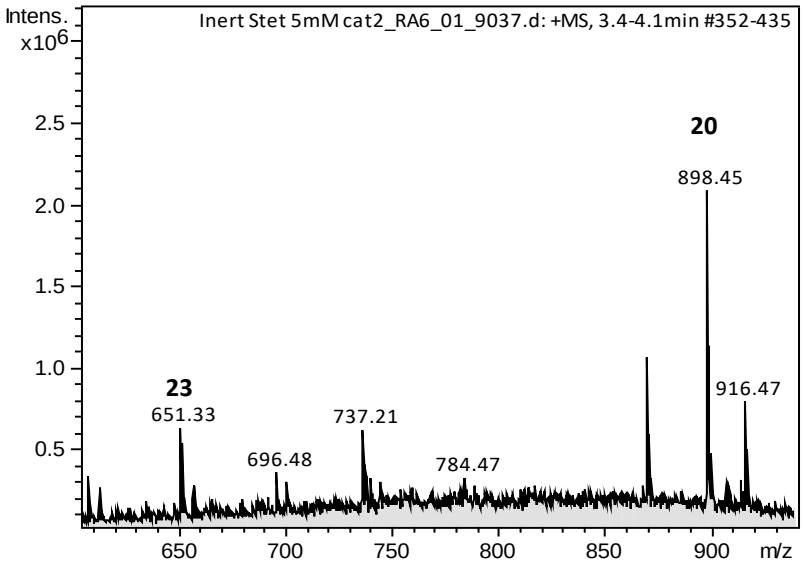


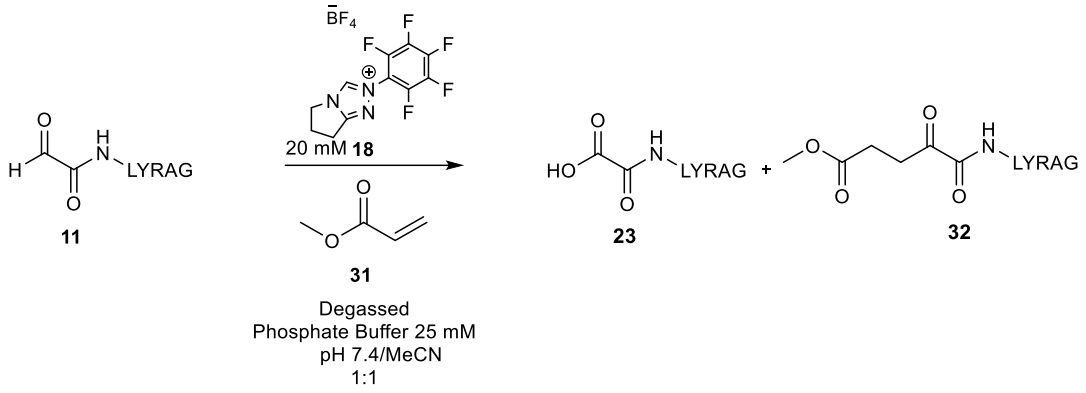


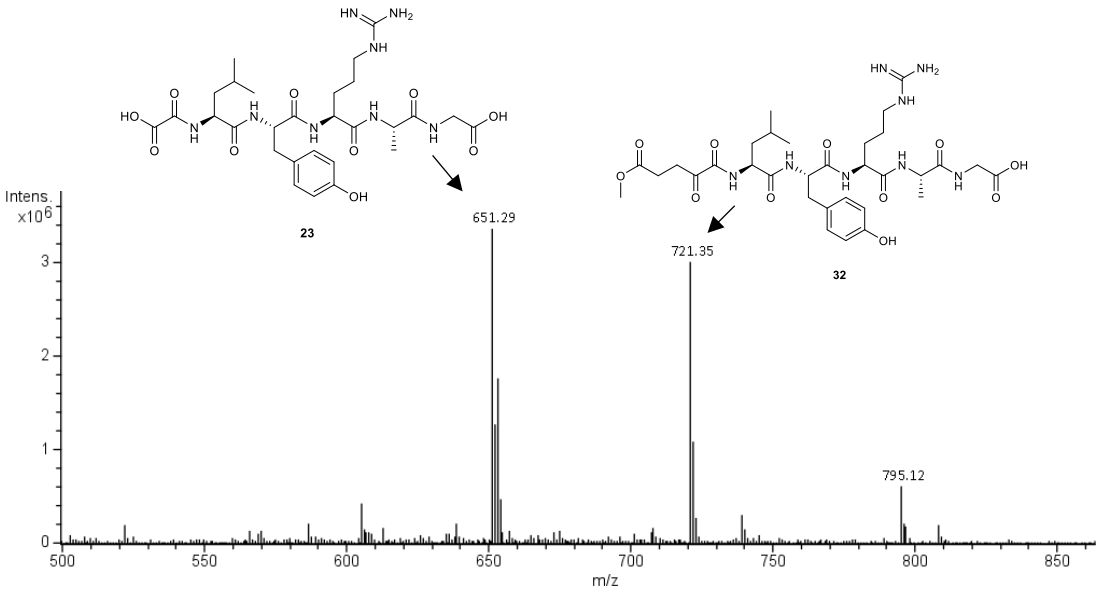

11
20 mM 18
17
Degassed
Phosphate Buffer 25 mM
pH 7.4/MeCN
1:1
23
20

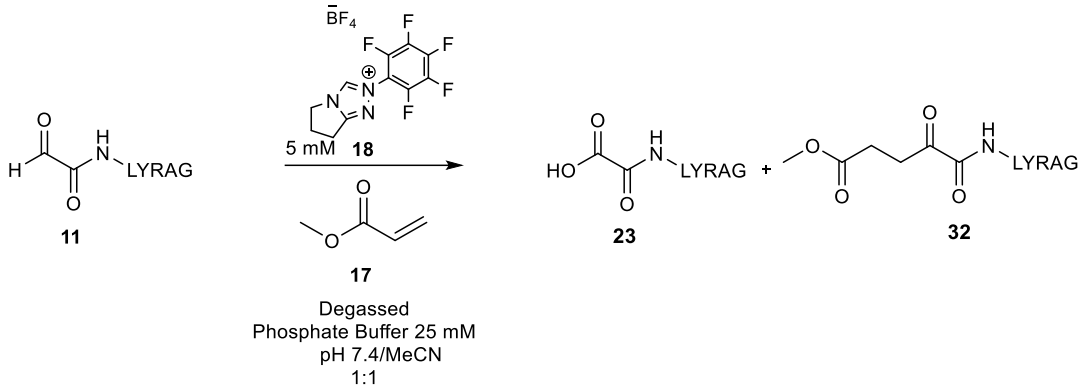

23
651.44
20
870.60
898.57
Intens.
x10^7
m/z

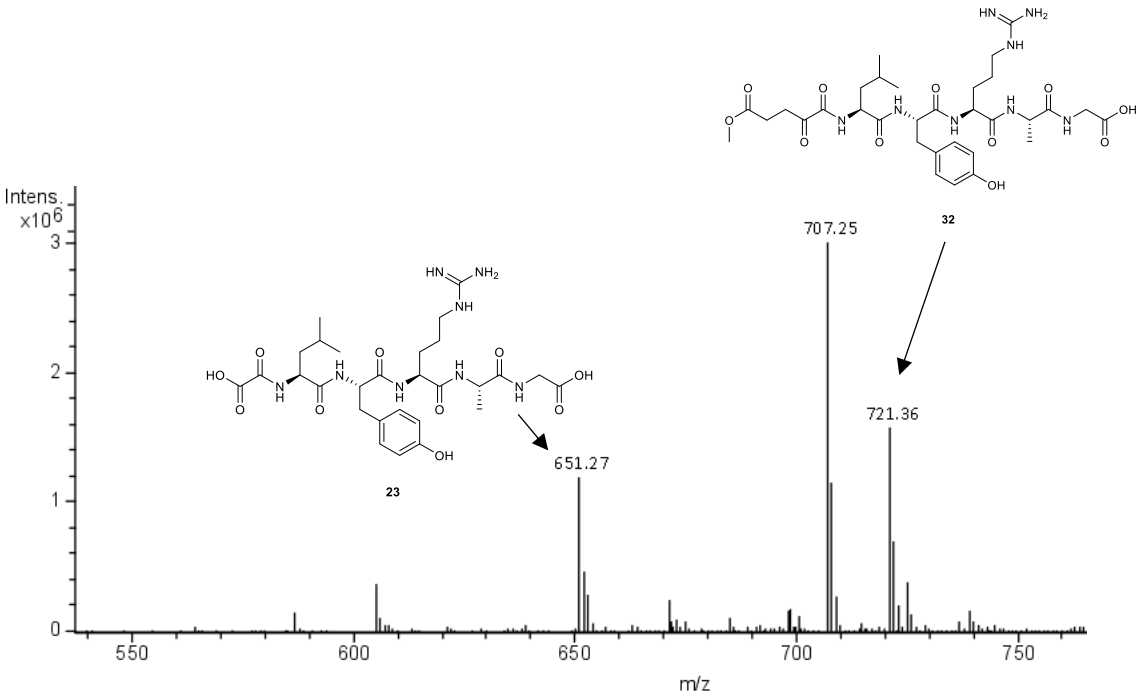

11
5 mM 18
17
Degassed
Phosphate Buffer 25 mM
pH 7.4/MeCN
1:1
23
20


Intens.
x10^6
Inert Stet 5mM cat2_RA6_01_9037.d: +MS, 3.4-4.1min #352-435
20
898.45
23
651.33
737.21
696.48
784.47
916.47
m/z


BF4
F
F
F
F
F
N
N
N
20 mM 18
H
O
N
LYRAG
O
11
O
O
31
Degassed
Phosphate Buffer 25 mM
pH 7.4/MeCN
1:1
HO
O
N
LYRAG
23
+
O
O
O
N
LYRAG
32


HN
NH2
NH
HO
OH
23
651.29
721.35
795.12
32
Intens.
x10^6
3
2
1
0
500
550
600
650
700
750
800
850
m/z

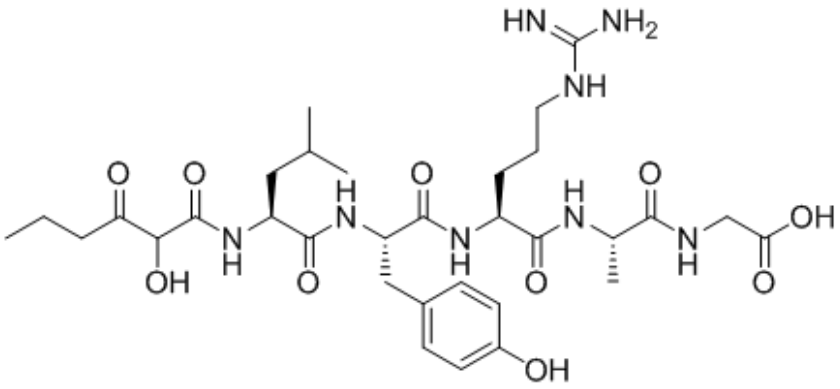

BF4
5 mM 18
LYRAG
11
17
23
32
Degassed
Phosphate Buffer 25 mM
pH 7.4/MeCN
1:1

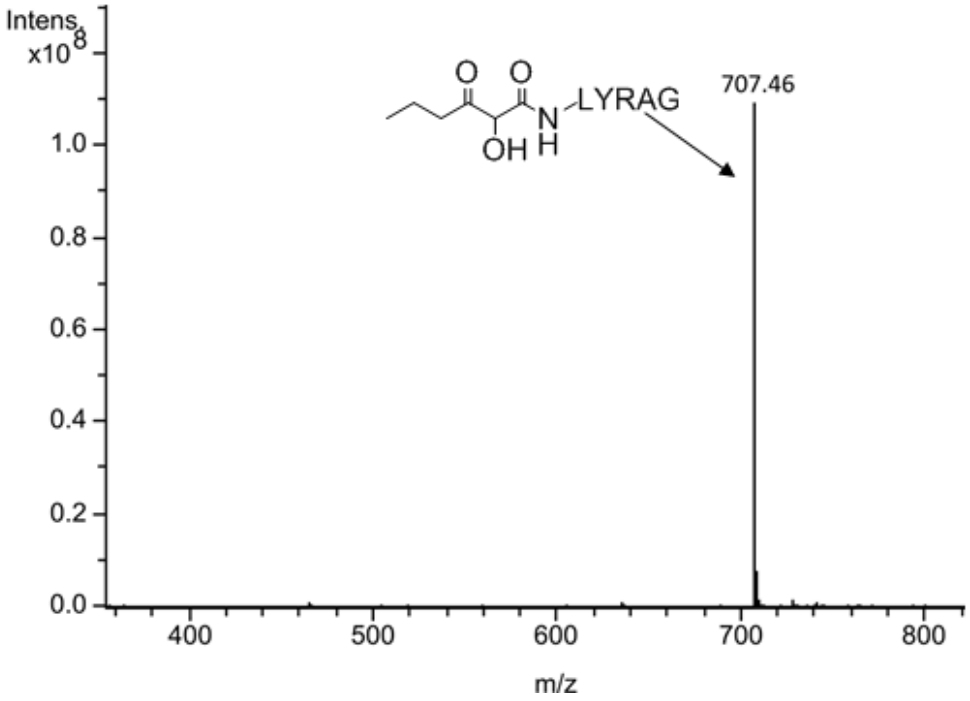

Intens.
x10^6
707.25
721.36
651.27
23
32
m/z

α-Oxo-LYRAG **44**

Calculated $[M+H]^+$ 707.36

Found $[M+H]^+$ 707.46

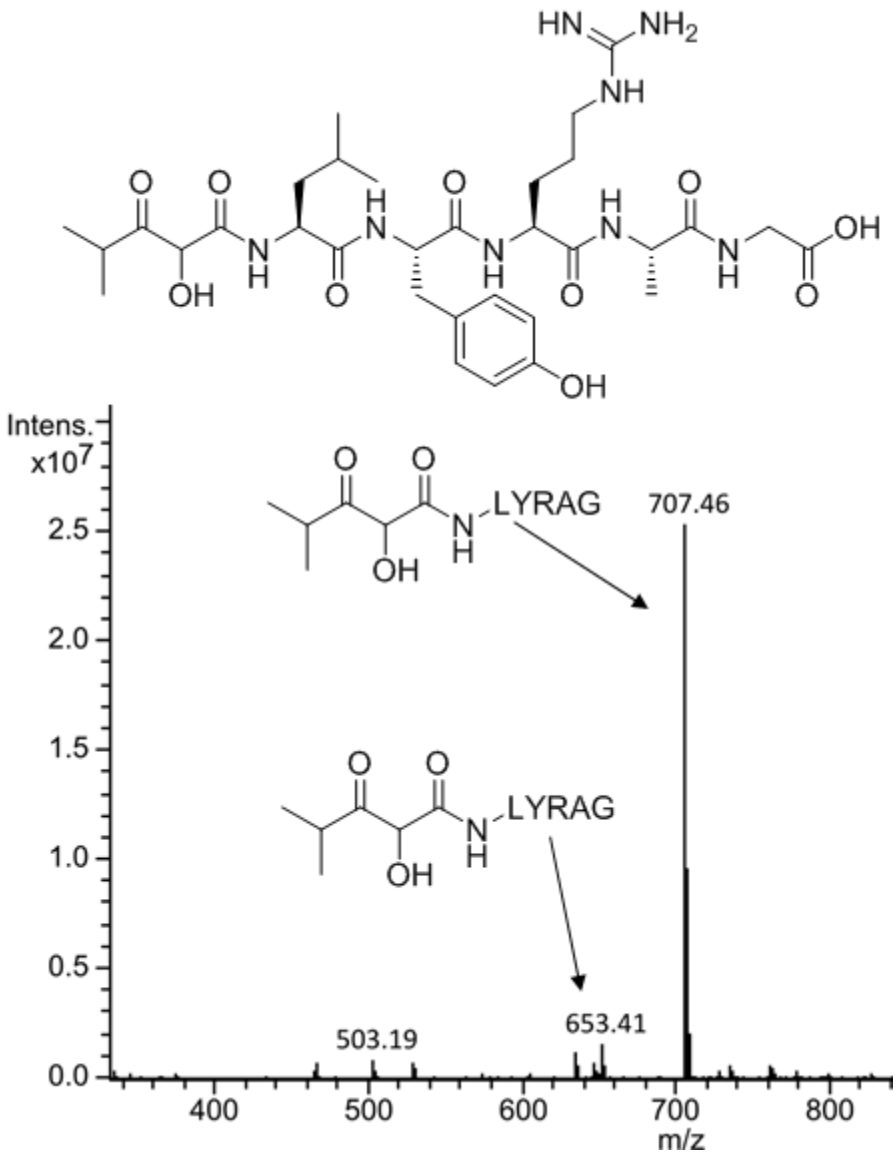




α-Oxo-LYRAG **49**

Calculated $[M+H]^+$ 707.36

Found $[M+H]^+$ 707.46

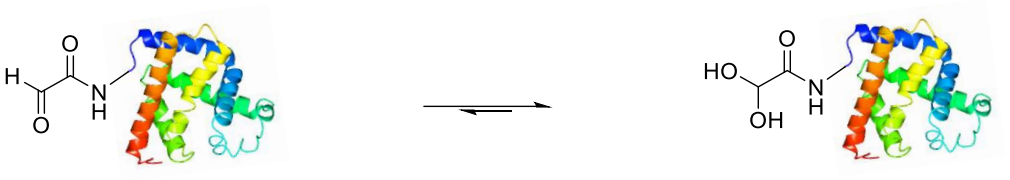


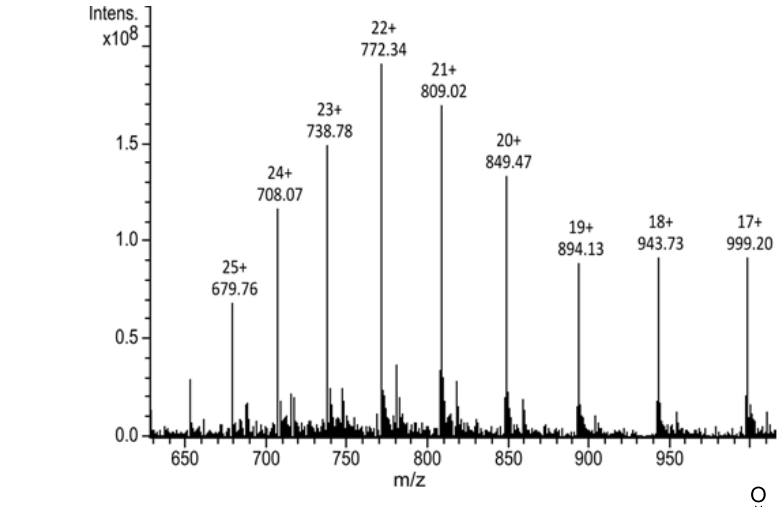
α-oxo-myoglobin **55**

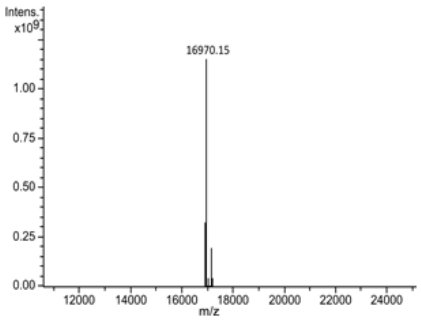
Calculated $[M+H]^+$ 16951, 16969 (hydrate)

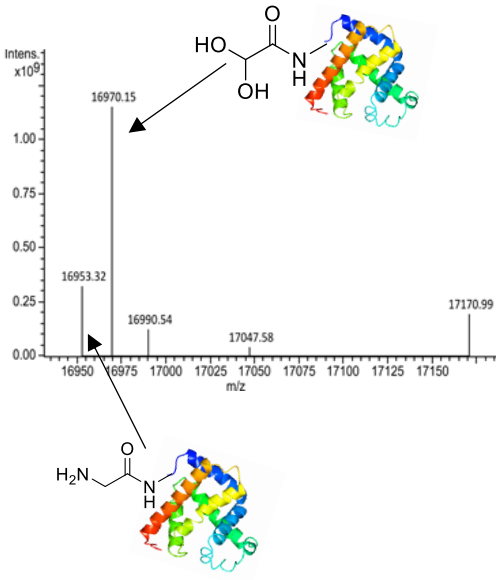
Found $[M+H]^+$ 16970

Peaks 1690, 17047 and 17170 correspond to unknown PLP oxidation by products.





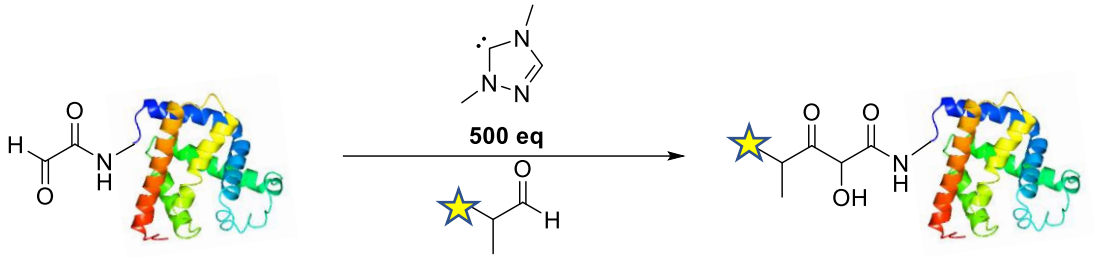




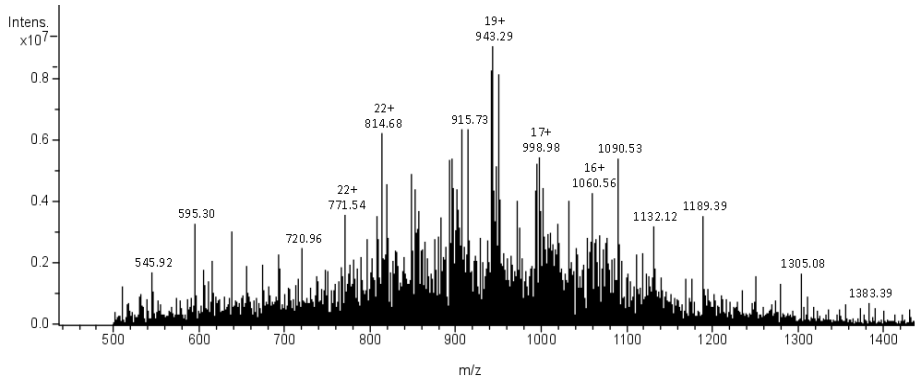
Probe -myoglobin **83**

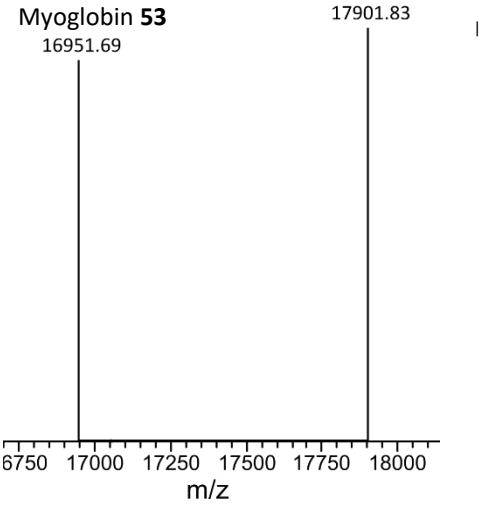
Calculated $[M+H]^+$ 17901

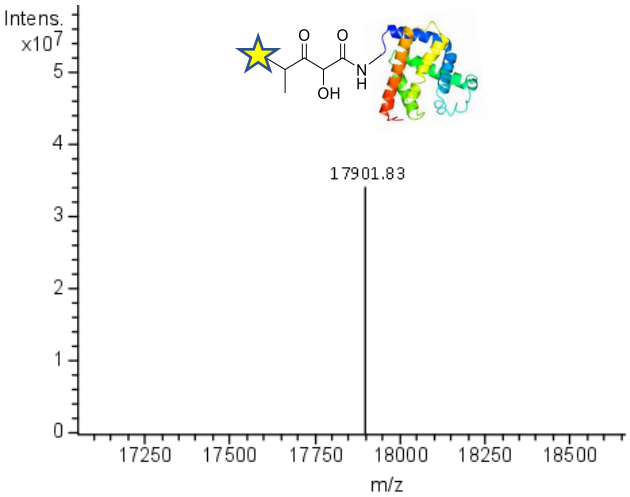
Found $[M+H]^+$ 17901



500 eq
25 mM PB pH 7.4
18 h, 37 °C



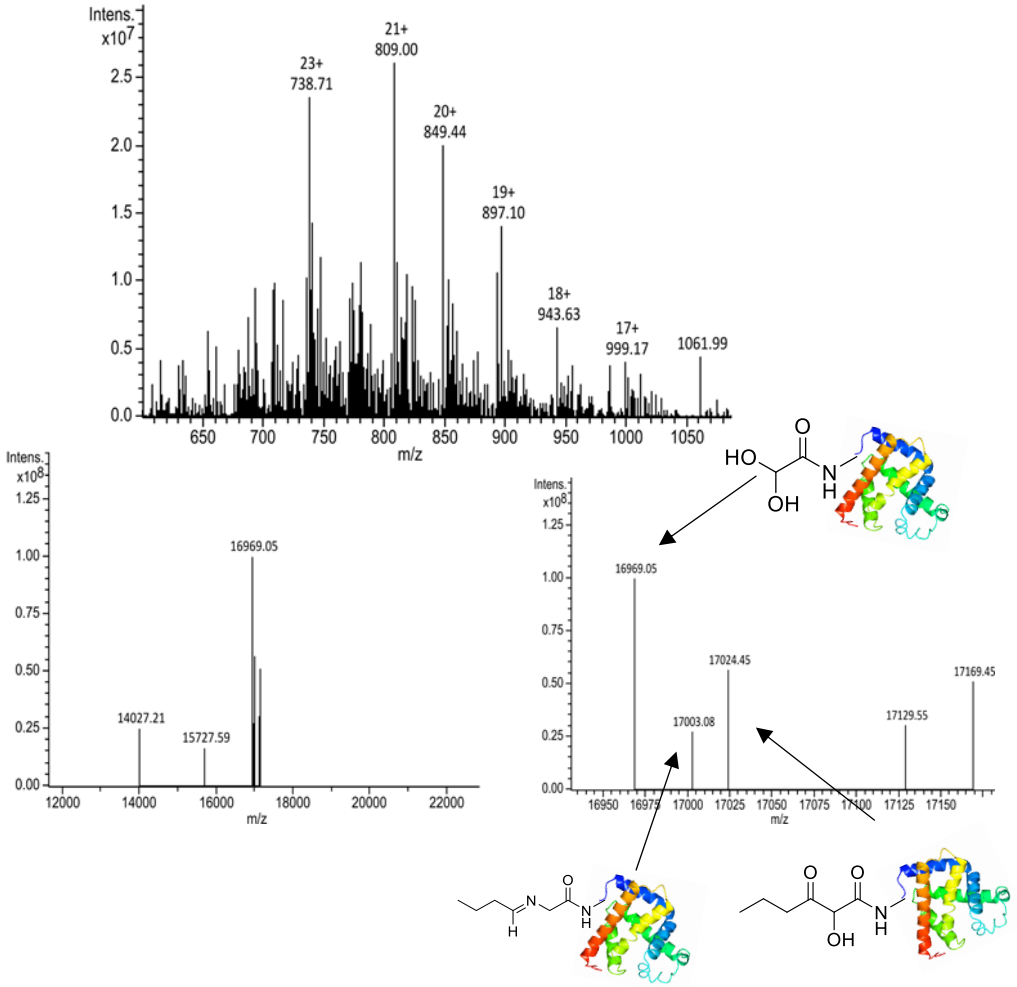




Butyr-myoglobin **57**

Calculated $[M+H]^+$ 17023

Found $[M+H]^+$ 17024



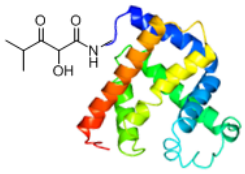
Wild type myoglobin, *n*-terminal Gly forming Schiff base with donor aldehyde.

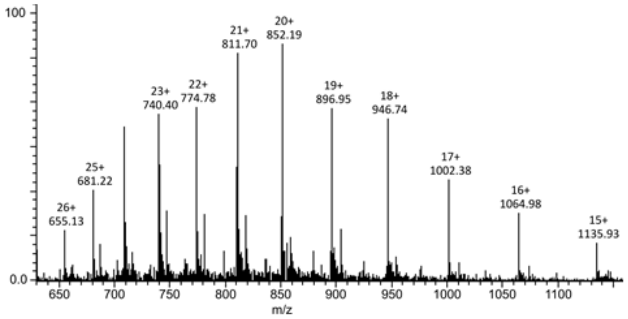
Wild type myoglobin, *n*-terminal Gly forming Schiff base with donor aldehyde.

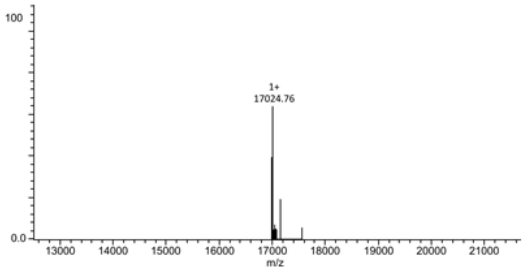
Isobutyr-myoglobin **56**

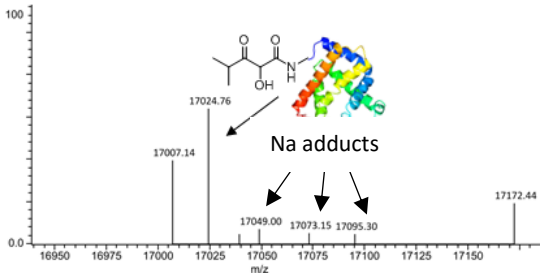
Calculated $[M+H]^+$ 17023

Found $[M+H]^+$ 17024





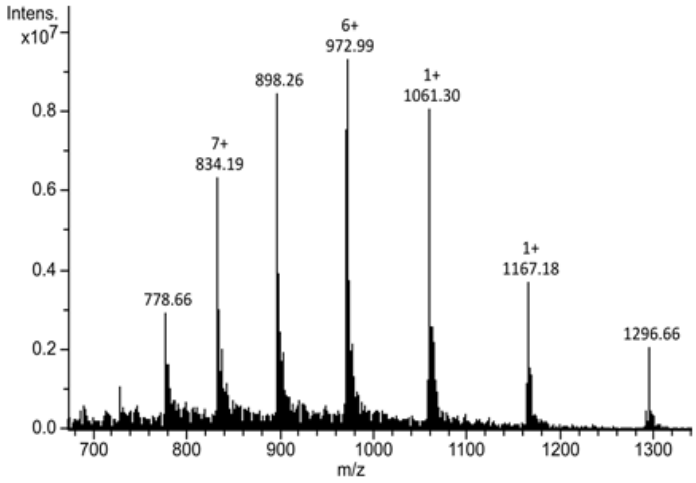


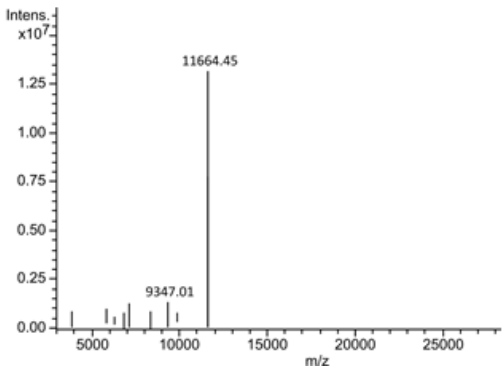


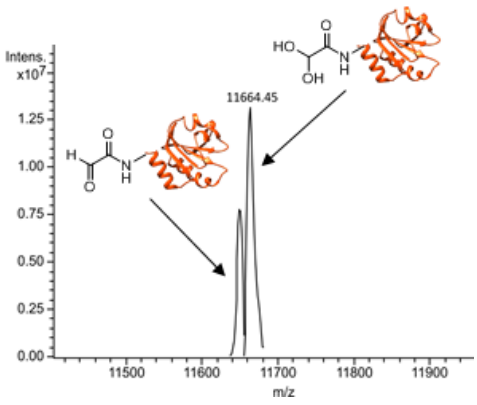
Oxo- Thioredoxin **58**

Calculated $[M+H]^+$ 11643, 11661 (hydrate)

Found $[M+H]^+$ 11661 (hydrate)



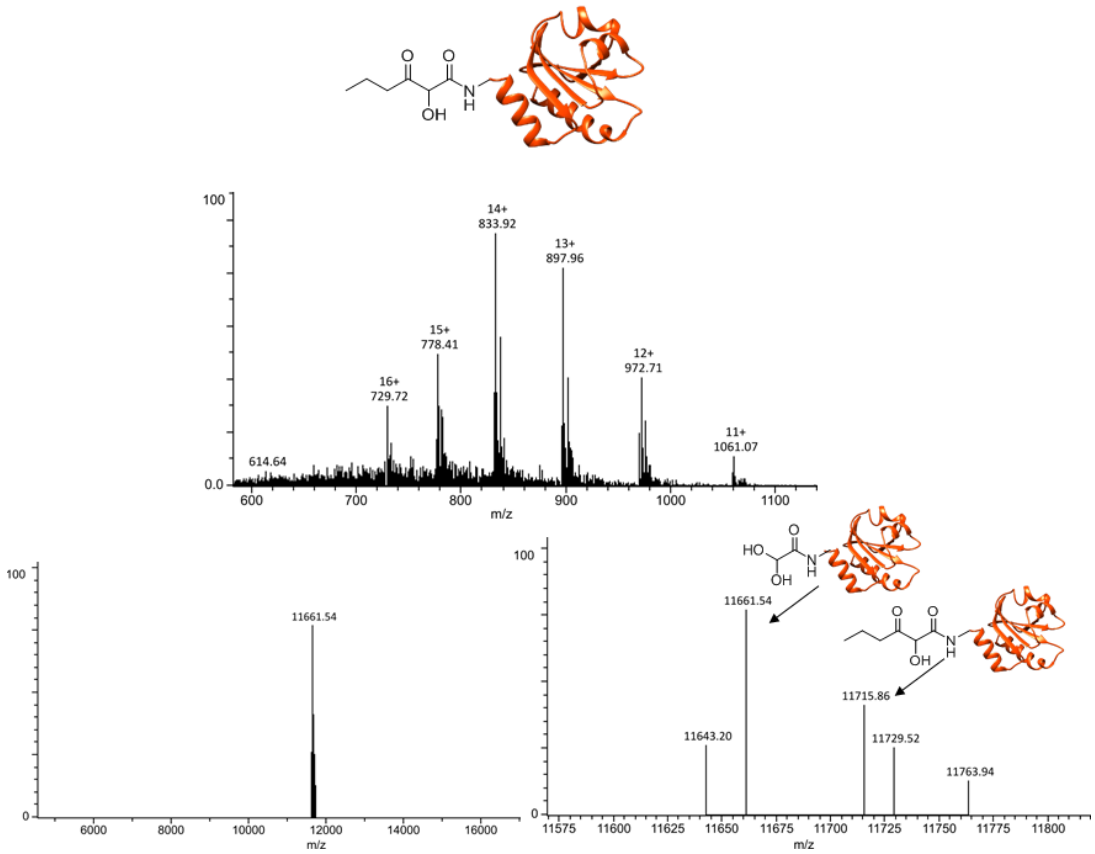




Oxo- butyr Thioredoxin **59**

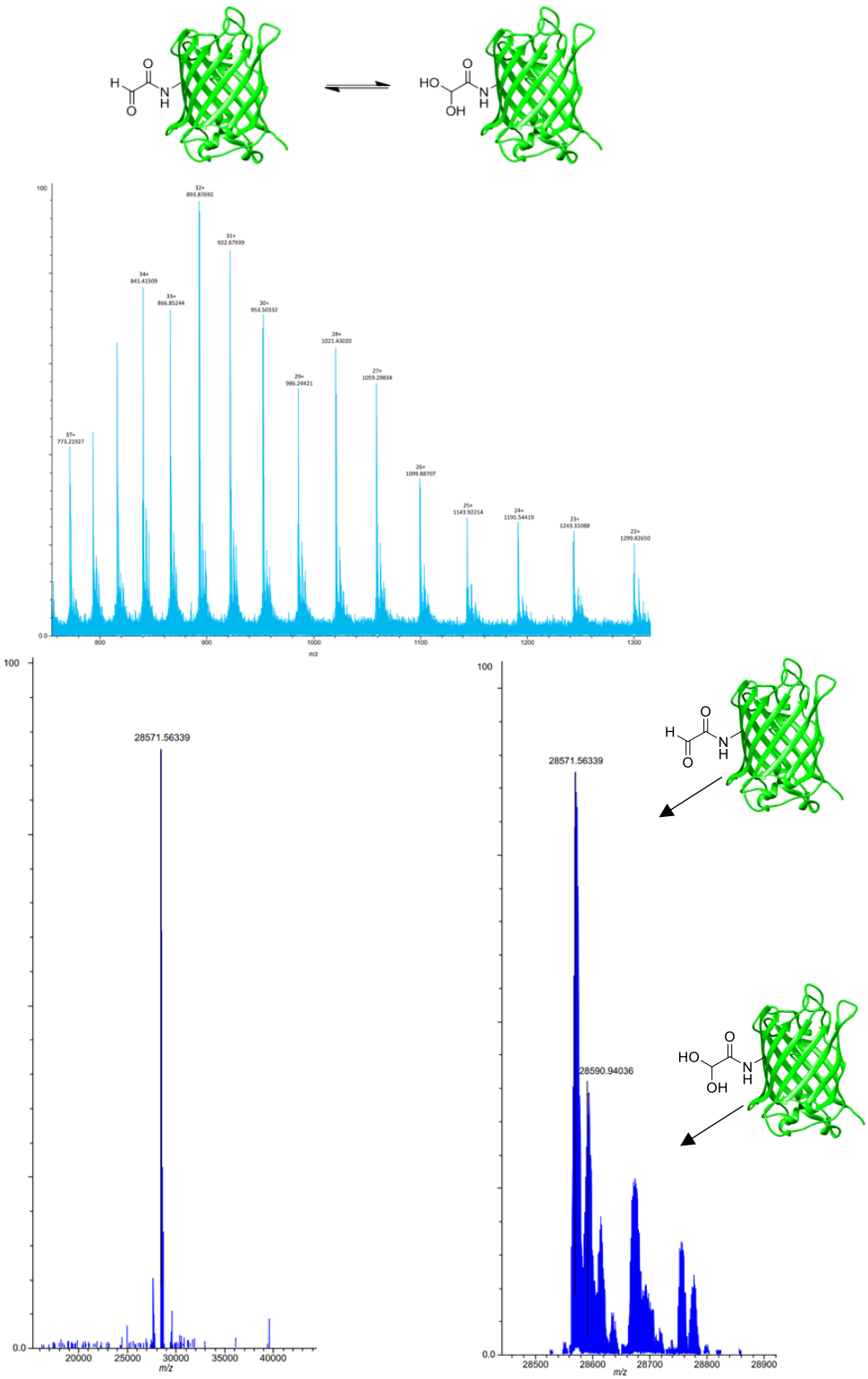
Calculated $[M+H]^+$ 11715

Found $[M+H]^+$ 11715

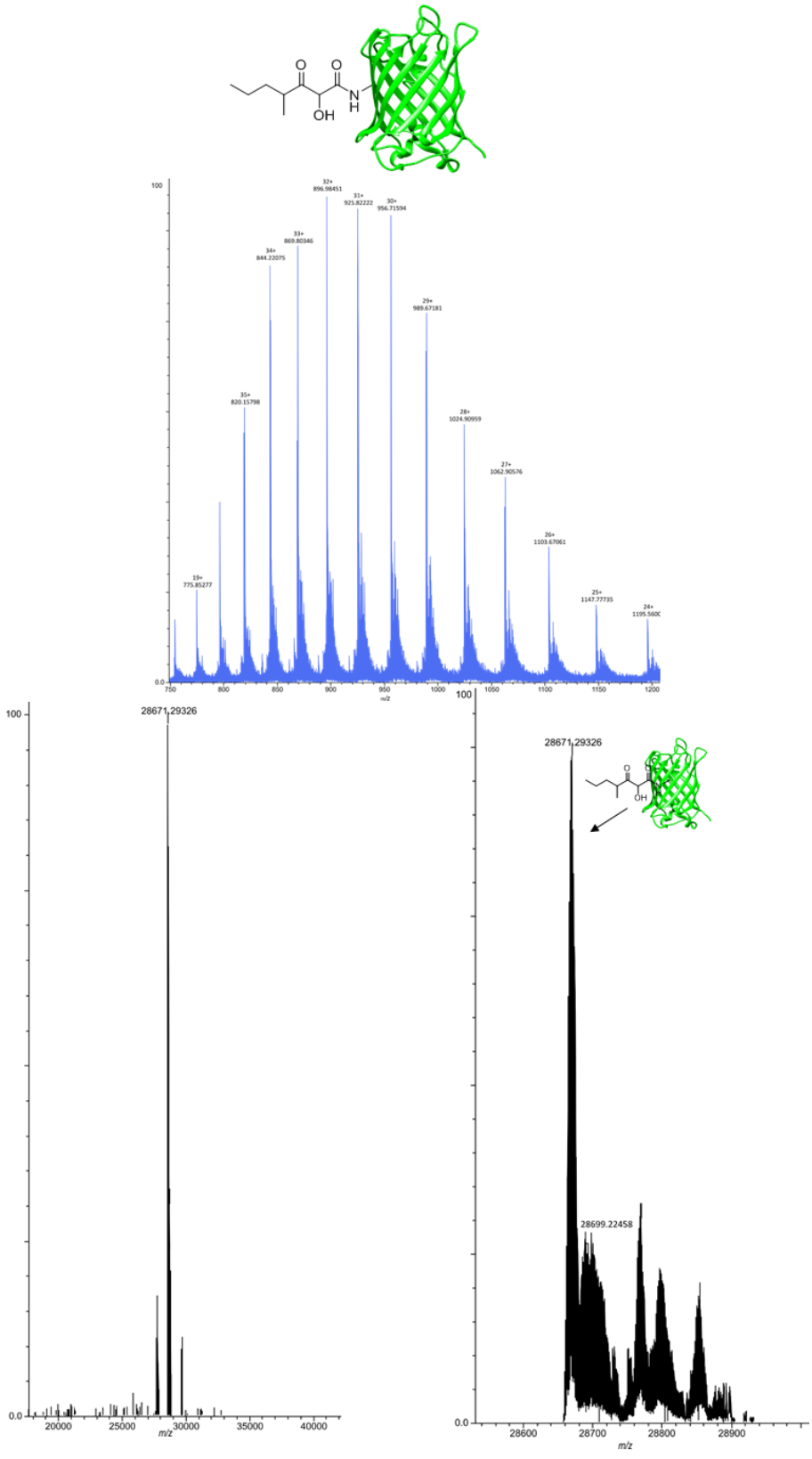


GFPY39 α-oxo K **65** Calculated $[M+H]^+$ 28575, 28593 (hydrate)

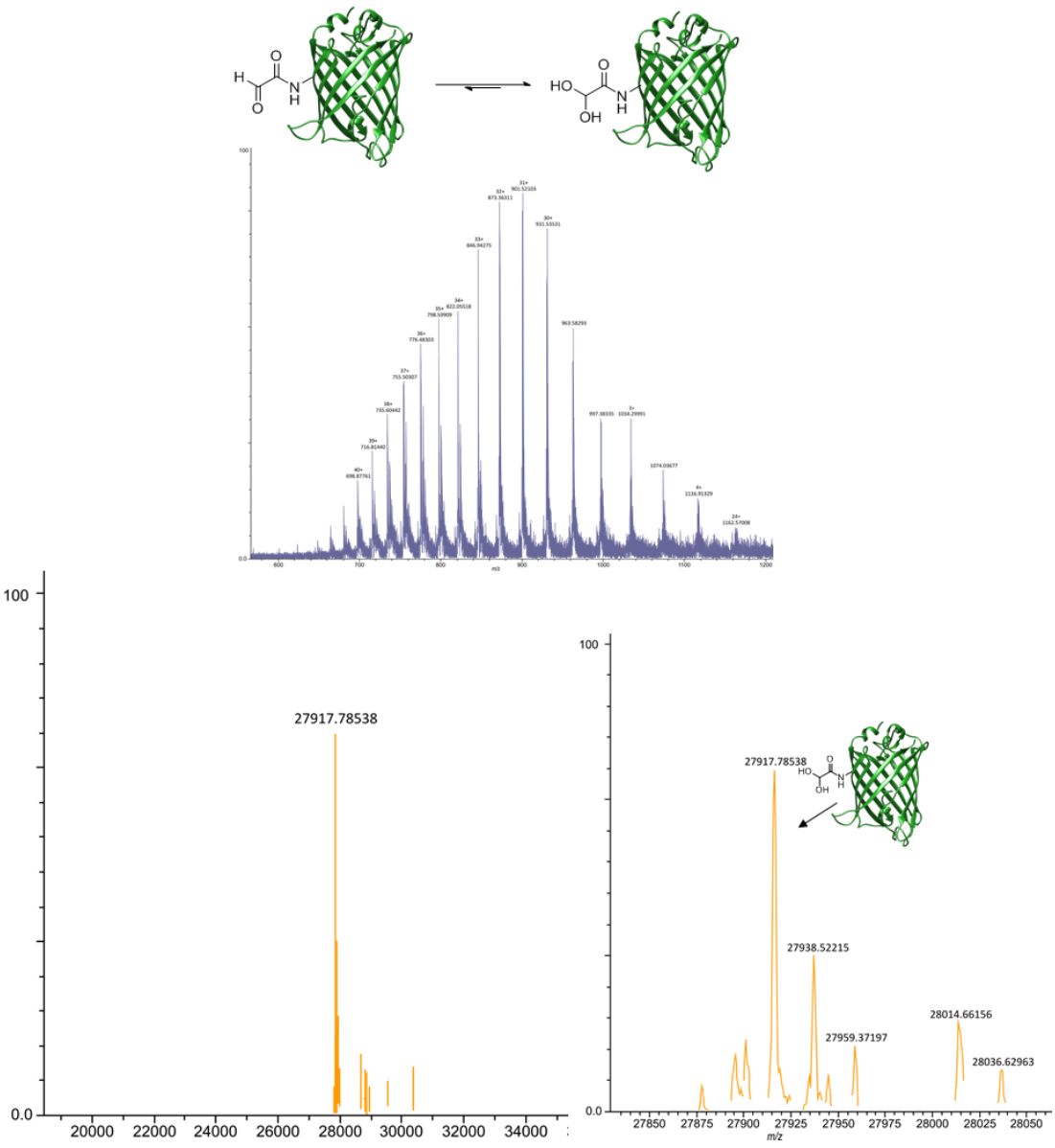
Found $[M+H]^+$ 28571, 28590 (hydrate)



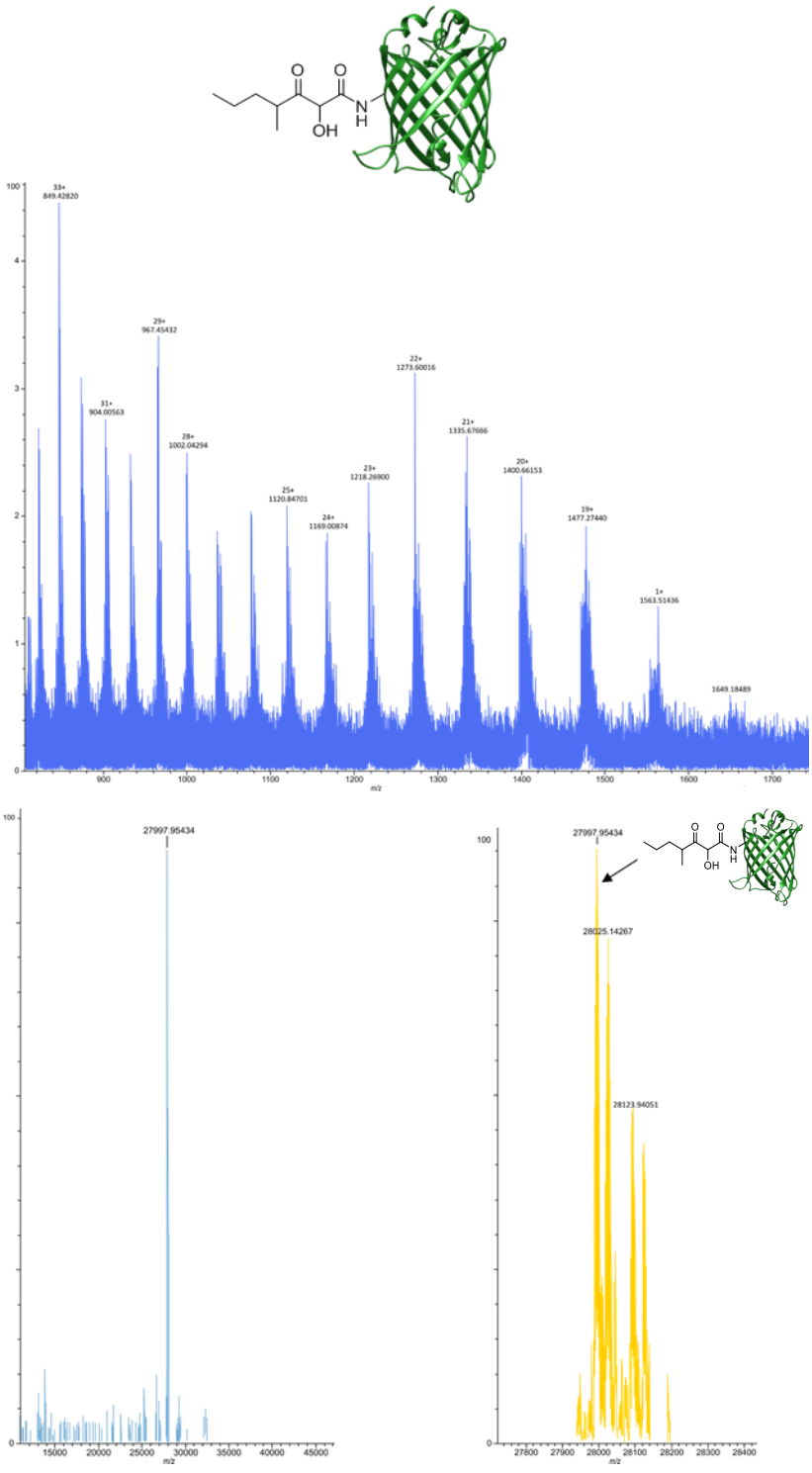
alkyl-GFPY39 **66** Calculated $[M+H]^+$ 28671, Found $[M+H]^+$ 28671

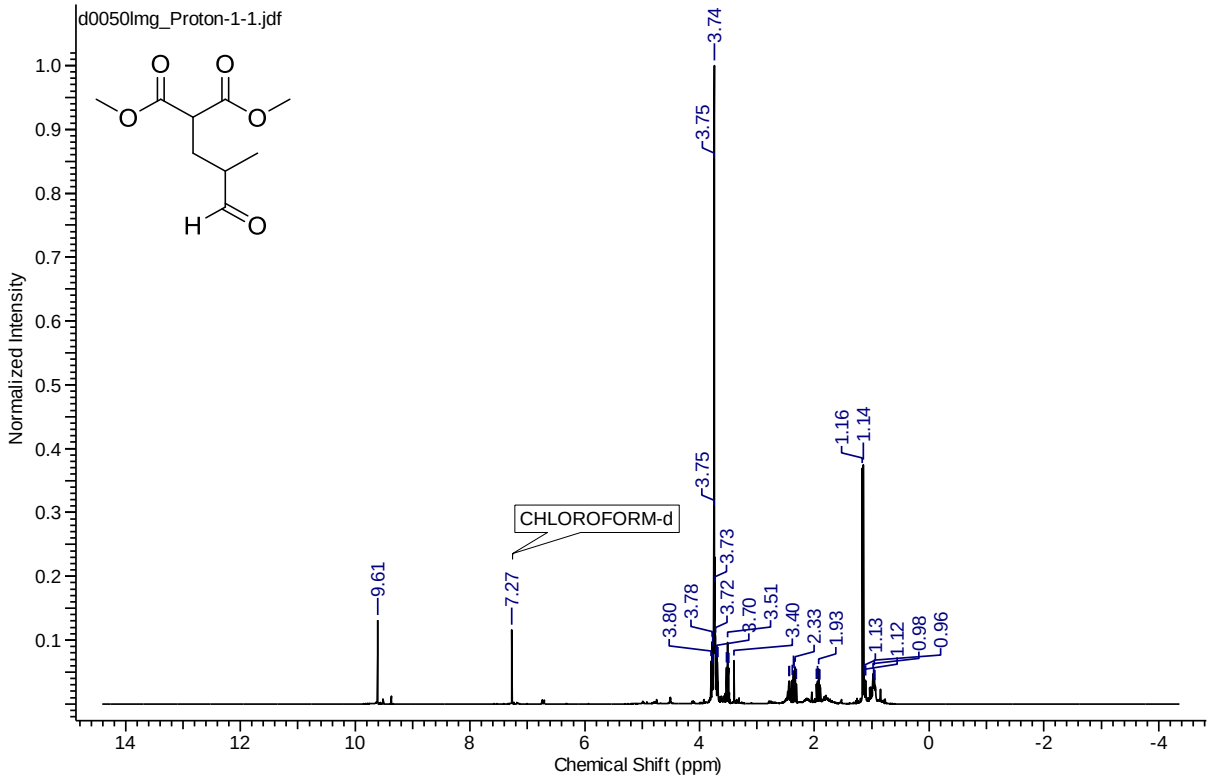


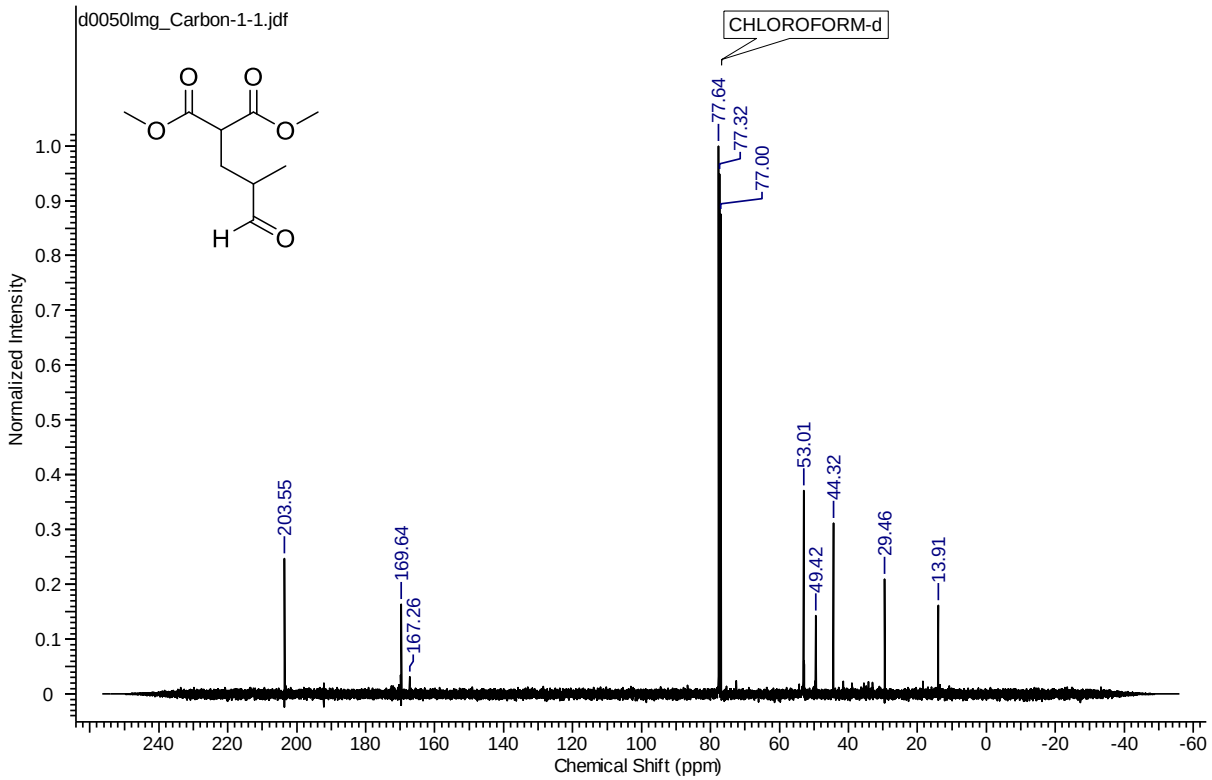
sfGFPN150 **67** Calculated $[M+H]^+$ 27900, 27918 (hydrate) Found $[M+H]^+$ 27917

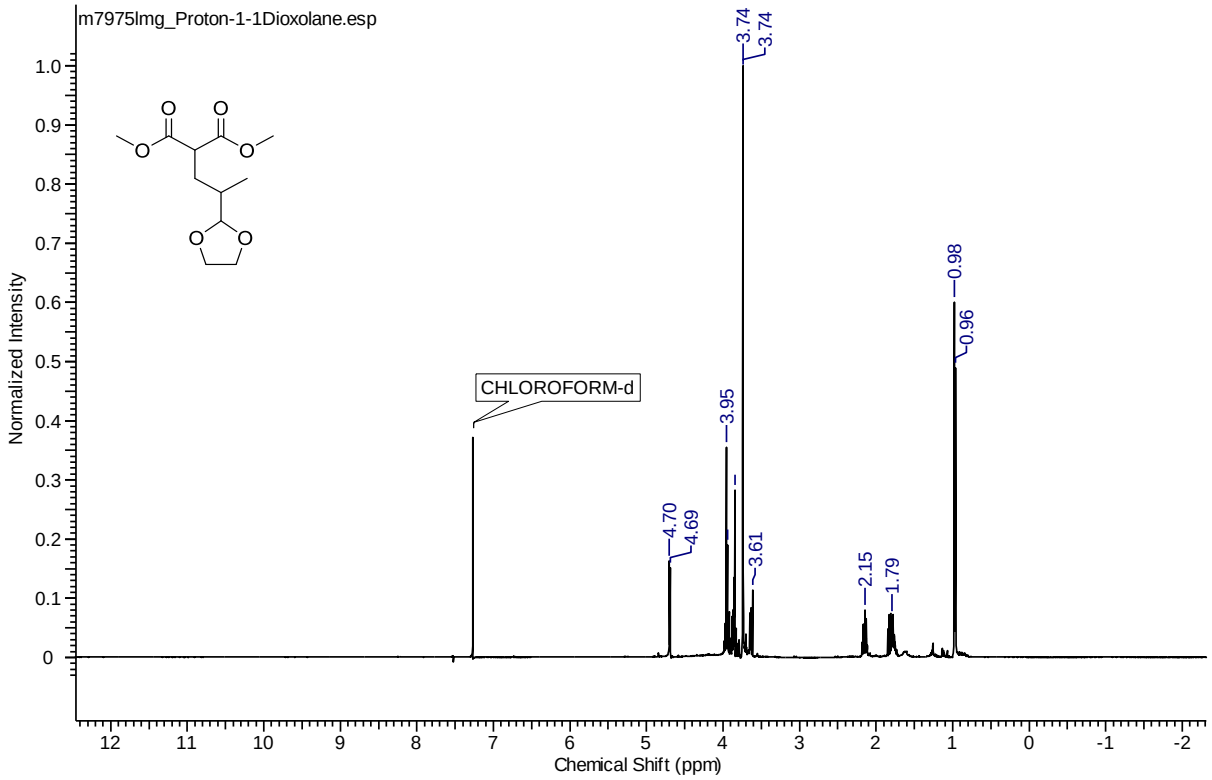


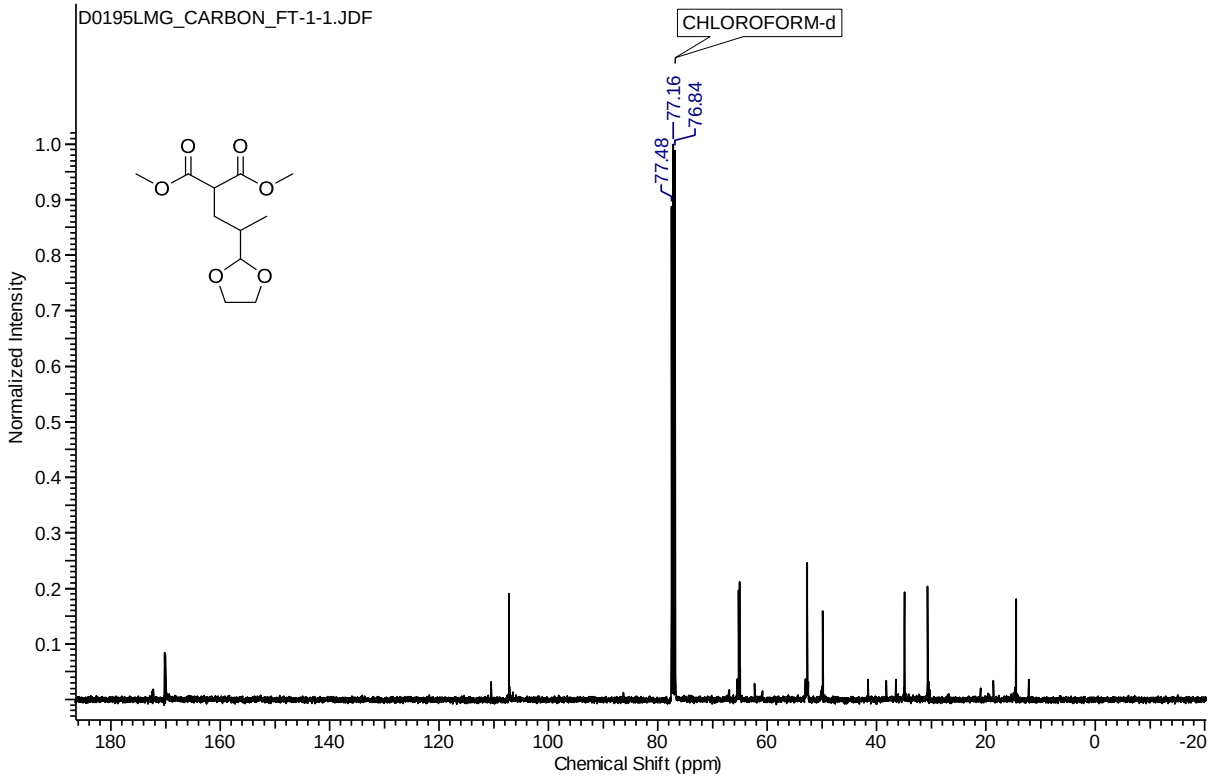
Oxo-sfGFPN150 **69** Calculated $[M+H]^+$ 28000 Found $[M+H]^+$ 27997

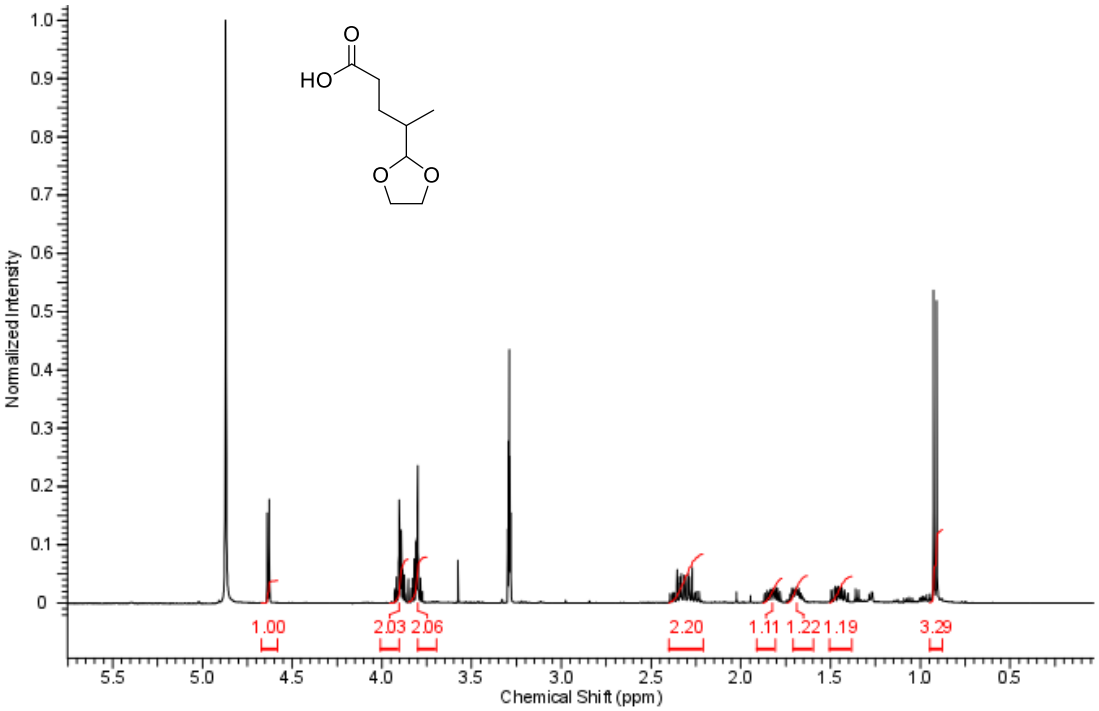


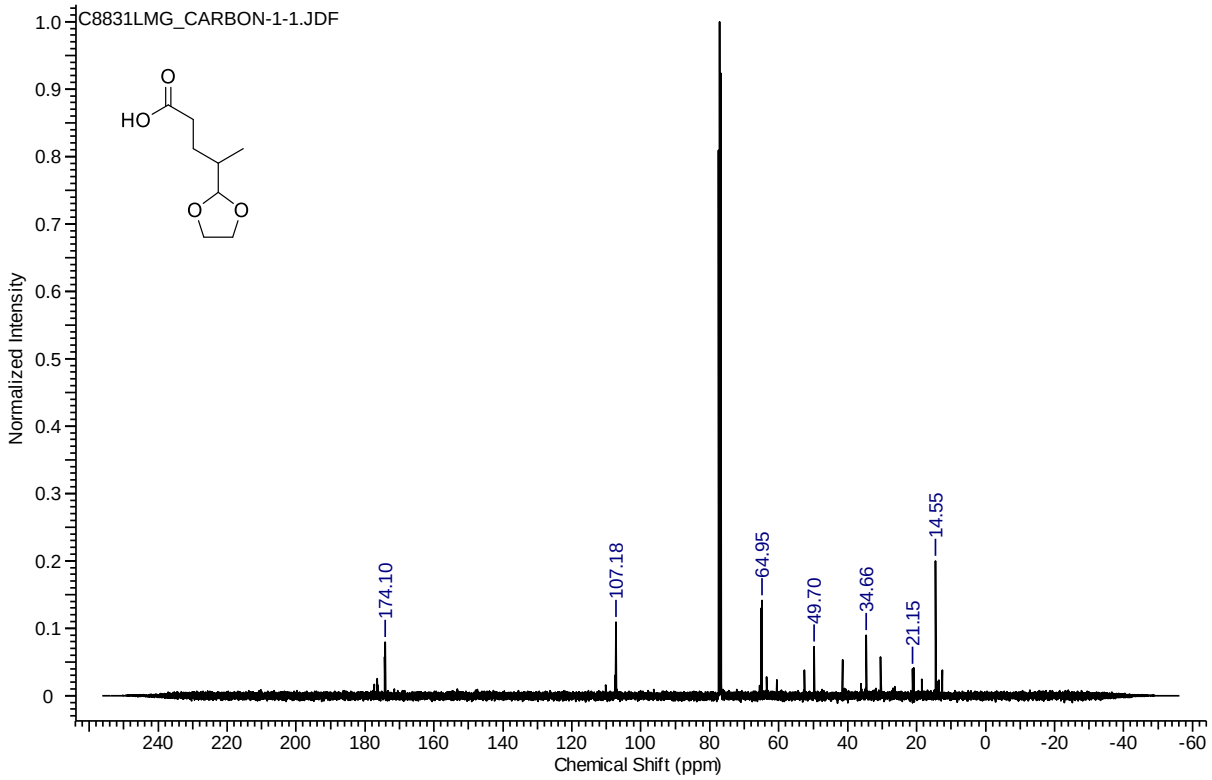

d0050lmg_Proton-1-1.jdf
Normalized Intensity
Chemical Shift (ppm)
CHLOROFORM-d
9.61
7.27
3.74
3.75
3.75
3.73
3.72
3.80
3.78
3.70
3.51
3.40
2.33
1.93
1.16
1.14
1.13
1.12
0.98
0.96

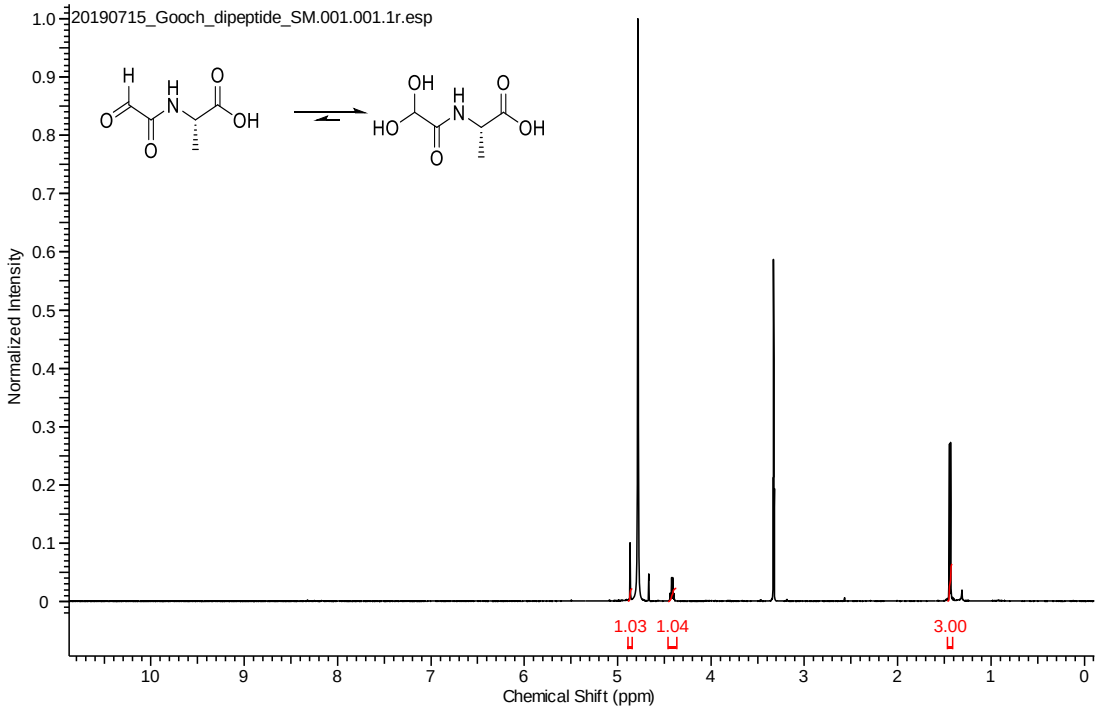

d0050lmg_Carbon-1-1.jdf
Normalized Intensity
Chemical Shift (ppm)
CHLOROFORM-d
77.64
77.32
77.00
203.55
169.64
167.26
53.01
49.42
44.32
29.46
13.91

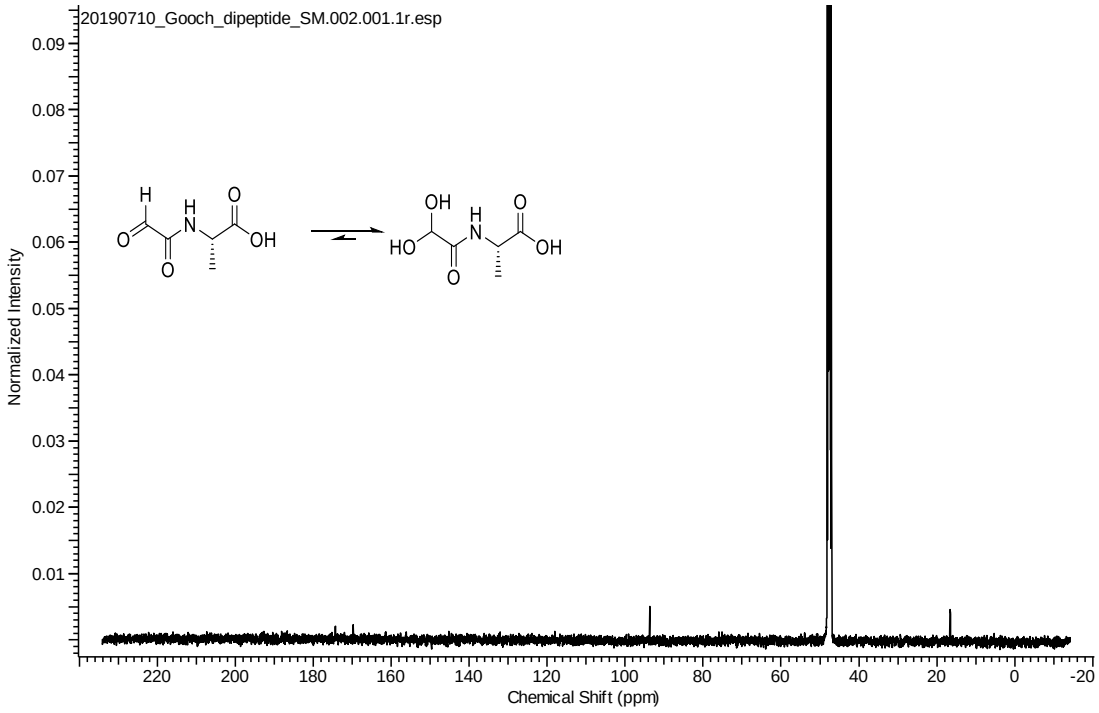

m7975lmg_Proton-1-1Dioxolane.esp
Normalized Intensity
Chemical Shift (ppm)
CHLOROFORM-d
3.74
3.74
0.98
0.96
3.95
4.70
4.69
3.61
2.15
1.79

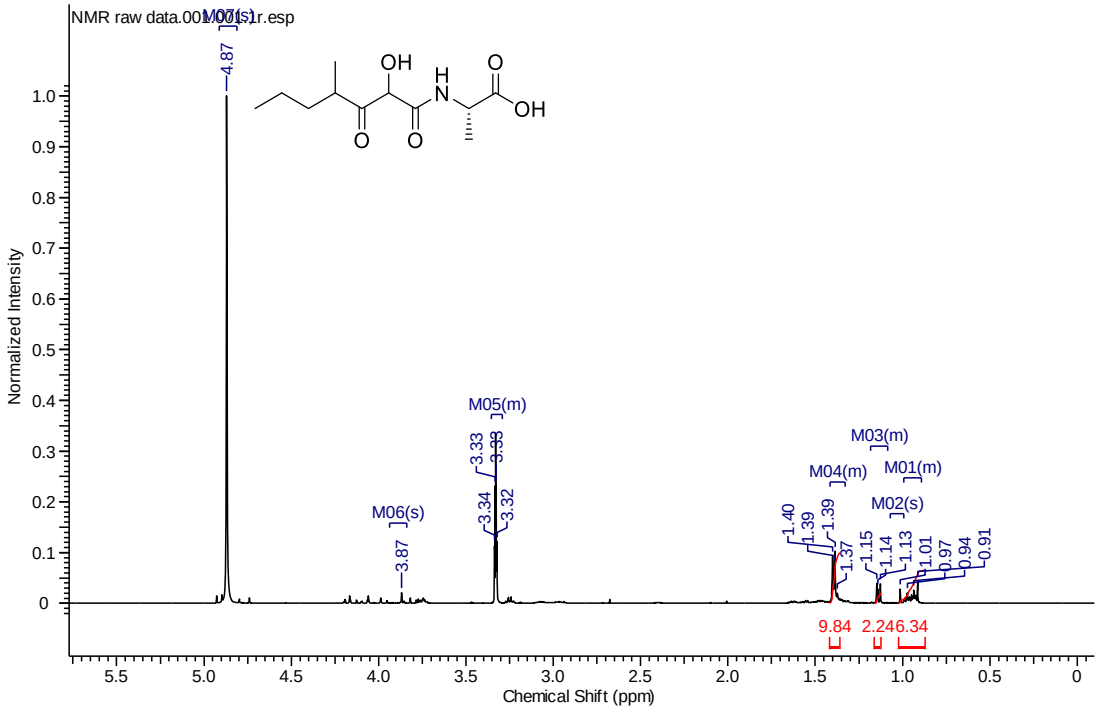

D0195LMG_CARBON_FT-1-1.JDF
Normalized Intensity
Chemical Shift (ppm)
CHLOROFORM-d
77.16
76.84
77.48

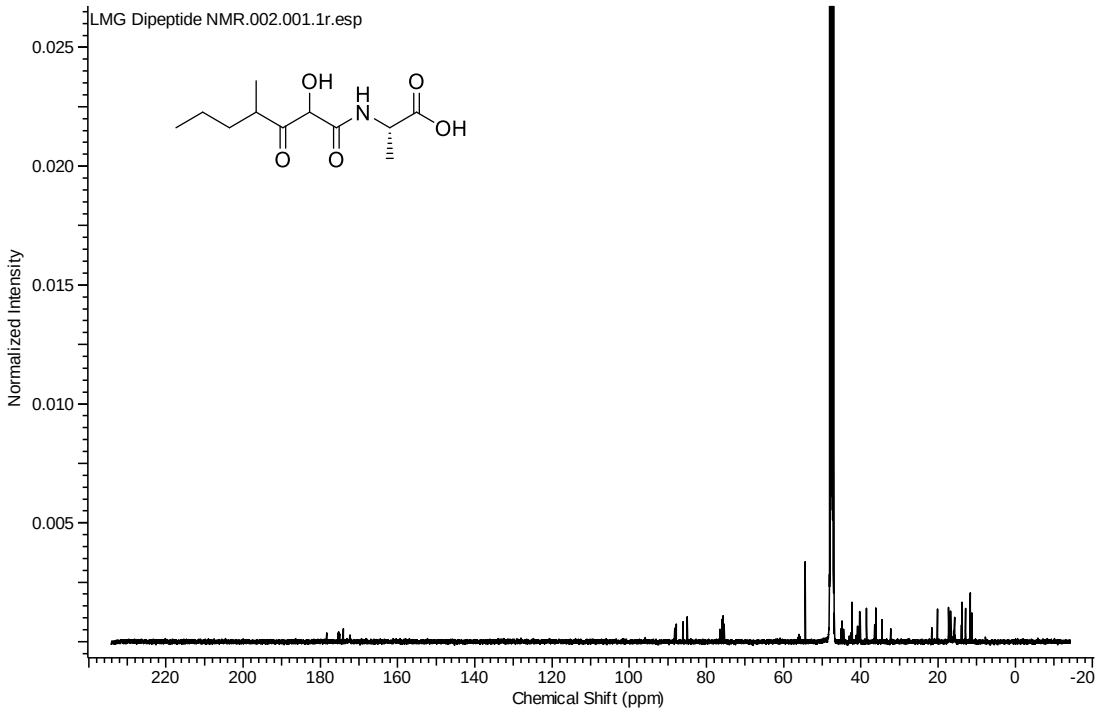

HO
O
O
O
Normalized Intensity
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1.00
2.03
2.06
2.20
1.11
1.22
1.19
3.29
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
Chemical Shift (ppm)

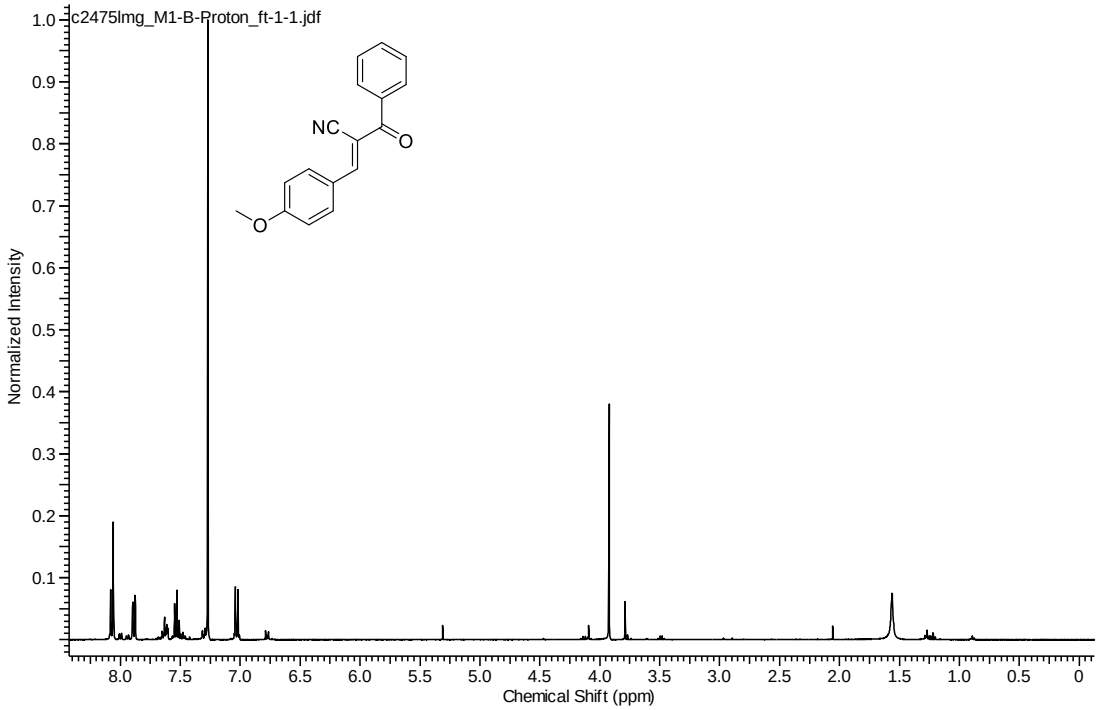

C8831LMG_CARBON-1-1.JDF
HO
O
O
O
Normalized Intensity
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
174.10
107.18
64.95
49.70
34.66
21.15
14.55
240
220
200
180
160
140
120
100
80
60
40
20
0
-20
-40
-60
Chemical Shift (ppm)

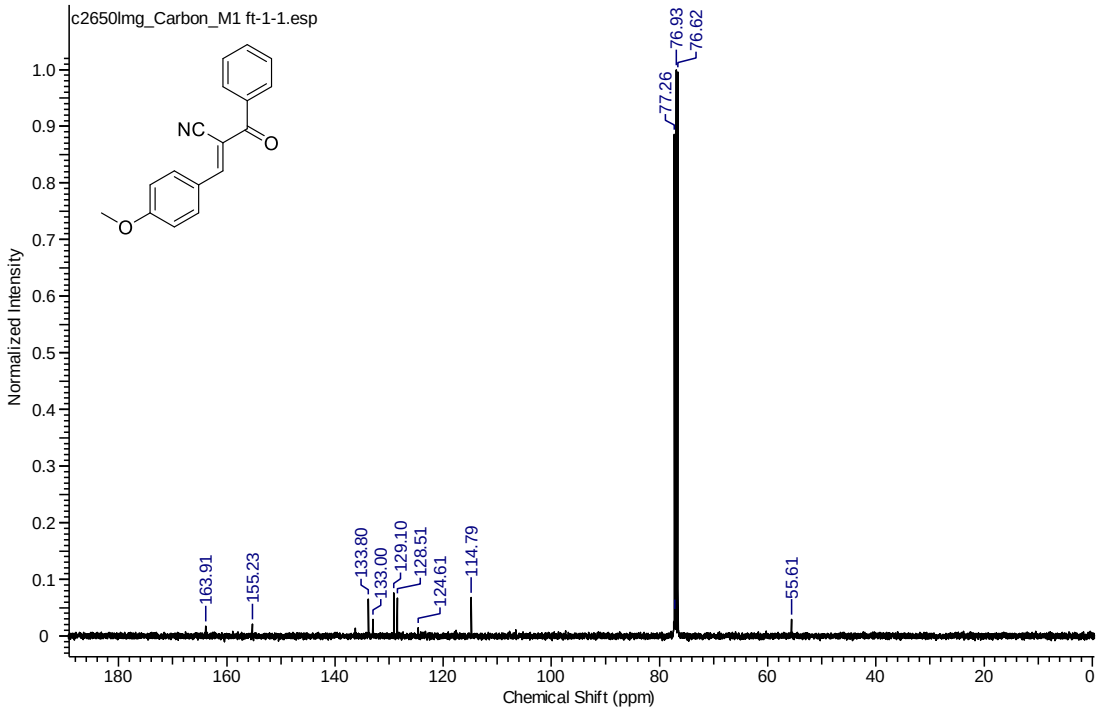

20190715_Gooch_dipeptide_SM.001.001.1r.esp
Normalized Intensity
Chemical Shift (ppm)
1.03
1.04
3.00


20190710_Gooch_dipeptide_SM.002.001.1r.esp
Normalized Intensity
Chemical Shift (ppm)


NMR raw data.001.001.1r.esp
M07(s)
4.87
M06(s)
3.87
M05(m)
3.33
3.32
3.34
M04(m)
M03(m)
M02(s)
M01(m)
1.40
1.39
1.37
1.15
1.14
1.13
1.01
0.97
0.94
0.91
9.84
2.24
6.34
Normalized Intensity
Chemical Shift (ppm)


LMG Dipeptide NMR.002.001.1r.esp
Normalized Intensity
Chemical Shift (ppm)


c2475lmg_M1-B-Proton_ft-1-1.jdf
NC
O
O
Normalized Intensity
Chemical Shift (ppm)


c2650lmg_Carbon_M1 ft-1-1.esp
NC
O
O
76.93
76.62
77.26
163.91
155.23
133.80
133.00
129.10
128.51
124.61
114.79
55.61
Normalized Intensity
Chemical Shift (ppm)

List of Abbreviations

ABAO 2-Amino benzamidoxime

AcrK *N*-Acryloyl-lysine

ADC Antibody Drug Conjugate

A*f*BPs Affinity- based probes

BCN Bicyclo- [6.1.0]-nonyne

BSA Bovine serum albumin

CAN Ceric ammonium (IV) nitrate

C-C Carbon-Carbon bond

CD Circular dichroism

CuAAC Copper-mediated azide-alkyne cycloaddition

DAR Drug-antibody ratio

DCM Dichloromethane

Dha Dehydroalanine

DIFO Difluorinated cyclooctynes

DIPEA N, N-Diisopropylethylamine

DMF Dimethylforamide

DMSO Dimethylsulfoxide

E. coli Escherichia Coli

EGFRs Epidermal growth factor receptors

EPL Expressed protein ligation

ESI Electrospray ionisation

EtOAc Ethyl acetate

EWG Electron withdrawing group

FGE Formylglycine Generating Enzyme

fGly Formylglycine

FAPP Farnseyl aldehyde pyrophosphate

FPP Farnseyl pyrophosphate

GFP Green fluorescent protein

gpTGase Guinea pig liver transglutaminase

h Hour(s)

HCTU 2-(6-Chloro-1-H-benzotriazole-1-yl)-1,1,3,3- tetramethylaminium hexafluorophosphate

HER2 Human epidermal growth factor receptor 2

HFIP Hexafluoro-2-propanol

HGH Human growth hormone

HIPS Hydrazino-iso-Pictet Spengler

HPLC High performance liquid chromatography

HRMS High resolution mass spectrometry

HRP Horseradish peroxidase

HSA Human serum albumin

Hz Hertz

IEDDA Inverse electron demands Diels–Alder

IL-8 Interleukin-8

KOH Potassium hydroxide

LAP Lipoic acid ligase acceptor peptide

LB Lysogeny broth

LC-MS Liquid chromatography-mass spectrometry

LiCl Lithium chloride

LplA Lipoic acid ligase

LRMS Low resolution mass spectrometry

MBP Maltose binding protein

MeCN Acetonitrile

MeOH Methanol

MeSH Methanethiol

MNP Multivalent nanoparticle

MS Mass spectrometry

MS/MS Tandem mass spectrometry

MWCO Molecular weight cut off

m/z Mass to charge ratio

NaCl Sodium chloride

NCL Native chemical ligation

NHC N-heterocyclic carbene

NHS ester N-hydroxysuccinimide ester

NMR Nuclear magnetic resonance

OPAL Organocatalyst-mediated protein aldol ligation

PAL Photoaffinity labelling

PB Phosphate buffer

PBS Phosphate buffer saline

PEG Polyethylene glycol

PES Potential energy surface

PET Positron emission tomography

PFTase Protein farnesyl-transferase

PLP Pyridoxal-5-phosphate

PMP *para*-methoxyphenol

PTM Post-translational modification

PTS Protein-*trans*-splicing

QC Quadricyclane

scFv Single chain Fragment variable

SET Single electron transfer

sfGFP Superfolder green fluorescent protein S

SPAAC Strain-promoted azide-alkyne cycloaddition

SPANC Strain-promoted alkyne nitrone cycloaddition

SPANOC strain-promoted alkyne–nitrile oxide cycloaddition

SPOCQ Strain-promoted oxidation-controlled cyclooctyne-1,2-quinone cycloaddition

SPPS Solid phase peptide synthesis

SPSAC Copper-free strain-promoted sydnone–alkyne cycloaddition

SrtA Sortase A

TCEP Tris(2-carboxyethyl)phosphine

TEAB Triethanolamine bicarbonate T

TFA Trifluoroacetic acid

ThzK Thiazolidine lysine

TIS Triisopropylsilane

tK Trapped Knoevenagel

TKM Tandem Knoevenagel Condensation Michael Addition

TPPTS Triphenylphosphine trisulfonate

tRNA Transfer ribonucleic acid

TTL Tubulin tyrosine ligase

UAA Unnatural amino acids

UV-Vis Ultraviolet-visible

YEB Ylide ester biotin phosphonium salt precursor 2PCA 2-pyridinecarboxaldehyde

3ForTyr 3-Formyl-L-tyrosine

References

1. Glazer, A. N., Specific Chemical Modification of Proteins. *Annu. Rev. Biochem.* **1970**, *39* (1), 101-130.
2. Francis, M. B., Chemical Biology. Wiley: **2007**; p 593.
3. Hermanson, G. T., Immunotoxin Conjugation Techniques. In *Bioconj. Tech.*, Elsevier: **1996**; pp 494-527.
4. Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G., Chemical Modification of Proteins at Cysteine: Opportunities in Chemistry and Biology. *Chem. Asian. J.* **2009**, *4* (5), 630-640.
5. Smith, M. E. B.; Schumacher, F. F.; Ryan, C. P.; Tedaldi, L. M.; Papaioannou, D.; Waksman, G.; Caddick, S.; Baker, J. R., Protein Modification, Bioconjugation, and Disulfide Bridging Using Bromomaleimides. *J. Am. Chem. Soc.* **2010**, *132* (6), 1960-1965.
6. Hermanson, G. T., Bioconj. Tech. Elsevier: **2008**.
7. Betting, D. J.; Kafi, K.; Abdollahi-Fard, A.; Hurvitz, S. A.; Timmerman, J. M., Sulfhydryl-Based Tumor Antigen-Carrier Protein Conjugates Stimulate Superior Antitumor Immunity against B Cell Lymphomas. *J. Immunol.* **2008**, *181* (6), 4131-4140.
8. Sussman, D.; Westendorf, L.; Meyer, D. W.; Leiske, C. I.; Anderson, M.; Okeley, N. M.; Alley, S. C.; Lyon, R.; Sanderson, R. J.; Carter, P. J.; Benjamin, D. R., Engineered cysteine antibodies: an improved antibody-drug conjugate platform with a novel mechanism of drug-linker stability. *Protein Eng. Des. Sel.* **2018**, *31* (2), 47-54.
9. Moody, P.; Smith, M. E. B.; Ryan, C. P.; Chudasama, V.; Baker, J. R.; Molloy, J.; Caddick, S., Bromomaleimide-Linked Bioconjugates Are Cleavable in Mammalian Cells. *ChemBioChem* **2012**, *13* (1), 39-41.
10. Lundell, N.; Schreitmüller, T., Sample Preparation for Peptide Mapping— A Pharmaceutical Quality-Control Perspective. *Anal. Biochem.* **1999**, *266* (1), 31-47.
11. Simon, M. D.; Chu, F.; Racki, L. R.; de la Cruz, C. C.; Burlingame, A. L.; Panning, B.; Narlikar, G. J.; Shokat, K. M., The Site-Specific Installation of Methyl-Lysine Analogs into Recombinant Histones. *Cell* **2007**, *128* (5), 1003-1012.
12. Chatterjee, C.; Muir, T. W., Chemical Approaches for Studying Histone Modifications. *J. Biol. Chem.* **2010**, *285* (15), 11045-11050.
13. Hofer, T.; Thomas, J. D.; Burke, T. R.; Rader, C., An engineered selenocysteine defines a unique class of antibody derivatives. *PNAS.* **2008**, *105* (34), 12451-12456.
14. Bernardes, G. J. L.; Chalker, J. M.; Errey, J. C.; Davis, B. G., Facile Conversion of Cysteine and Alkyl Cysteines to Dehydroalanine on Protein Surfaces: Versatile and Switchable Access to Functionalized Proteins. *J. Am. Chem. Soc.* **2008**, *130* (15), 5052-5053.
15. Wang, J.; Schiller, S. M.; Schultz, P. G., A Biosynthetic Route to Dehydroalanine-Containing Proteins. *Angew. Chem. Int. Ed.* **2007**, *46* (36), 6849-6851.
16. Zhang, C.; Welborn, M.; Zhu, T.; Yang, N. J.; Santos, M. S.; Van Voorhis, T.; Pentelute, B. L., π-Clamp-mediated cysteine conjugation. *Nat. Chem.* **2015**, *8*, 120.
17. Spicer, C. D.; Davis, B. G., Selective chemical protein modification. *Nat. Comm.* **2014**, *5* (1), 4740.
18. van Kasteren, S. I.; Kramer, H. B.; Jensen, H. H.; Campbell, S. J.; Kirkpatrick, J.; Oldham, N. J.; Anthony, D. C.; Davis, B. G., Expanding the diversity of chemical protein modification allows post-translational mimicry. *Nature* **2007**, *446* (7139), 1105-1109.
19. Bulaj, G., Formation of disulfide bonds in proteins and peptides. *Biotechnol. Adv.* **2005**, *23* (1), 87-92.
20. Gunnoo, S. B.; Madder, A., Chemical Protein Modification through Cysteine. *ChemBioChem* **2016**, *17* (7), 529-553.
21. Habeeb, A. F. S. A., [37] Reaction of protein sulfhydryl groups with Ellman's reagent. In *Meth. Enzymol.*, Academic Press: 1972; Vol. 25, pp 457-464.

22. Trivedi, M. V.; Laurence, J. S.; Siahaan, T. J., The role of thiols and disulfides on protein stability. *Curr. protein peptide sci.* **2009**, *10* (6), 614-625.
23. Bernardes, G. J. L.; Casi, G.; Trüssel, S.; Hartmann, I.; Schwager, K.; Scheuermann, J.; Neri, D., A Traceless Vascular-Targeting Antibody–Drug Conjugate for Cancer Therapy. *Angew. Chem. Int. Ed* **2012**, *124* (4), 965-968.
24. List, T.; Casi, G.; Neri, D., A Chemically Defined Trifunctional Antibody–Cytokine–Drug Conjugate with Potent Antitumor Activity. *Mol. Cancer Ther.* **2014**, *13* (11), 2641-2652.
25. López-Gallego, F.; Abian, O.; Guisán, J. M., Altering the Interfacial Activation Mechanism of a Lipase by Solid-Phase Selective Chemical Modification. *Biochem.* **2012**, *51* (35), 7028-7036.
26. Chatterjee, C.; McGinty, R. K.; Fierz, B.; Muir, T. W., Disulfide-directed histone ubiquitylation reveals plasticity in hDot1L activation. *Nat. Chem. Biol.* **2010**, *6* (4), 267-269.
27. Kenyon, G. L.; Bruice, T. W., [40] Novel sulfhydryl reagents. In *Meth. Enzymol.*, Academic Press: 1977; Vol. 47, pp 407-430.
28. Berglund, P.; DeSantis, G.; Stabile, M. R.; Shang, X.; Gold, M.; Bott, R. R.; Graycar, T. P.; Lau, T. H.; Mitchinson, C.; Jones, J. B., Chemical Modification of Cysteine Mutants of Subtilisin Bacillus lentus Can Create Better Catalysts Than the Wild-Type Enzyme. *J. Am. Chem. Soc.* **1997**, *119* (22), 5265-5266.
29. Davis, B. G.; Maughan, M. A. T.; Green, M. P.; Ullman, A.; Jones, J. B., Glycomethanethiosulfonates: powerful reagents for protein glycosylation. *Tet. Asymm.* **2000**, *11* (1), 245-262.
30. Gamblin, D. P.; Garnier, P.; Ward, S. J.; Oldham, N. J.; Fairbanks, A. J.; Davis, B. G., Glycosyl phenylthiosulfonates (Glyco-PTS): novel reagents for glycoprotein synthesis. *Org. Biomol. Chem.* **2003**, *1* (21), 3642-3644.
31. van Kasteren, S. I.; Kramer, H. B.; Gamblin, D. P.; Davis, B. G., Site-selective glycosylation of proteins: creating synthetic glycoproteins. *Nat. Prot.* **2007**, *2* (12), 3185-3194.
32. Antos, J. M.; Francis, M. B., Transition metal catalyzed methods for site-selective protein modification. *Curr. Opin. Chem. Bio.* **2006**, *10* (3), 253-262.
33. Joshi, N. S.; Whitaker, L. R.; Francis, M. B., A Three-Component Mannich-Type Reaction for Selective Tyrosine Bioconjugation. *J. Am. Chem. Soc.* **2004**, *126* (49), 15942-15943.
34. Kodadek, T.; Duroux-Richard, I.; Bonnafous, J.-C., Techniques: Oxidative cross-linking as an emergent tool for the analysis of receptor-mediated signalling events. *Tren. Pharmacol. Sci.* **2005**, *26* (4), 210-217.
35. Schlick, T. L.; Ding, Z.; Kovacs, E. W.; Francis, M. B., Dual-Surface Modification of the Tobacco Mosaic Virus. *J. Am. Chem. Soc.* **2005**, *127* (11), 3718-3723.
36. Hooker, J. M.; Kovacs, E. W.; Francis, M. B., Interior Surface Modification of Bacteriophage MS2. *J. Am. Chem. Soc.* **2004**, *126* (12), 3718-3719.
37. Le Breton, N.; Martinho, M.; Mileo, E.; Etienne, E.; Gerbaud, G.; Guigliarelli, B.; Belle, V., Exploring intrinsically disordered proteins using site-directed spin labeling electron paramagnetic resonance spectroscopy. *Front. mol. biosci.* **2015**, *2*, 21-21.
38. McFarland, J. M.; Joshi, N. S.; Francis, M. B., Characterization of a Three-Component Coupling Reaction on Proteins by Isotopic Labeling and Nuclear Magnetic Resonance Spectroscopy. *J. Am. Chem. Soc.* **2008**, *130* (24), 7639-7644.
39. Romanini, D. W.; Francis, M. B., Attachment of Peptide Building Blocks to Proteins Through Tyrosine Bioconjugation. *Bioconj. Chem.* **2008**, *19* (1), 153-157.
40. Seim, K. L.; Obermeyer, A. C.; Francis, M. B., Oxidative Modification of Native Protein Residues Using Cerium(IV) Ammonium Nitrate. *J. Am. Chem. Soc.* **2011**, *133* (42), 16970-16976.

41. Espuña, G.; Andreu, D.; Barluenga, J.; Pérez, X.; Planas, A.; Arsequell, G.; Valencia, G., Iodination of Proteins by IPy2BF4, a New Tool in Protein Chemistry. *Biochem.* **2006**, *45* (19), 5957-5963.
42. Carrico, Z. M.; Romanini, D. W.; Mehl, R. A.; Francis, M. B., Oxidative coupling of peptides to a virus capsid containing unnatural amino acids. *Chem. Comm.* **2008**, (10), 1205-1207.
43. Tong, G. J.; Hsiao, S. C.; Carrico, Z. M.; Francis, M. B., Viral Capsid DNA Aptamer Conjugates as Multivalent Cell-Targeting Vehicles. *J. Am. Chem. Soc.* **2009**, *131* (31), 11174-11178.
44. Tilley, S. D.; Francis, M. B., Tyrosine-Selective Protein Alkylation Using π-Allylpalladium Complexes. *J. Am. Chem. Soc.* **2006**, *128* (4), 1080-1081.
45. Antos, J. M.; Francis, M. B., Selective Tryptophan Modification with Rhodium Carbenoids in Aqueous Solution. *J. Am. Chem. Soc.* **2004**, *126* (33), 10256-10257.
46. Antos, J. M.; McFarland, J. M.; Iavarone, A. T.; Francis, M. B., Chemoselective Tryptophan Labeling with Rhodium Carbenoids at Mild pH. *J. Am. Chem. Soc.* **2009**, *131* (17), 6301-6308.
47. Sletten, E. M.; Bertozzi, C. R., Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angew. Chem. Int. Ed.* **2009**, *48* (38), 6974-6998.
48. Smith, M. L.; Lindbo, J. A.; Dillard-Telm, S.; Brosio, P. M.; Lasnik, A. B.; McCormick, A. A.; Nguyen, L. V.; Palmer, K. E., Modified Tobacco mosaic virus particles as scaffolds for display of protein antigens for vaccine applications. *Virology* **2006**, *348* (2), 475-488.
49. Chudasama, V.; Maruani, A.; Caddick, S., Recent advances in the construction of antibody-drug conjugates. *Nat Chem* **2016**, *8* (2), 114-119.
50. Nakamura, T.; Kawai, Y.; Kitamoto, N.; Osawa, T.; Kato, Y., Covalent Modification of Lysine Residues by Allyl Isothiocyanate in Physiological Conditions: Plausible Transformation of Isothiocyanate from Thiol to Amine. *Chem. Res. Toxicol.* **2009**, *22* (3), 536-542.
51. Kalkhof, S., Chances and Pitfalls of Chemical Cross-Linking with Amine-Reactive N-Hydroxy Succinimide Esters. *Anal. Bioanal. Chem.* **2008**, *392*, 305-12.
52. Edman, P., Method for Determination of the Amino Acid Sequence in Peptides. *Acta chemica Scan.* **1950**.
53. Tanaka, K.; Katsumura, S.; Fukase, K., Discovery and Application of 6π-Azaelectrocyclization to Natural Product Synthesis and Synthetic Biology. *Sci. China Chem.* **2012**, *55*.
54. Tanaka, K.; Masuyama, T.; Hasegawa, K.; Tahara, T.; Mizuma, H.; Wada, Y.; Watanabe, Y.; Fukase, K., A Submicrogram-Scale Protocol for Biomolecule-Based PET Imaging by Rapid 6π-Azaelectrocyclization: Visualization of Sialic Acid Dependent Circulatory Residence of Glycoproteins. *Angew. Chem. Int. Ed.* **2008**, *47* (1), 102-105.
55. McFarland, J. M.; Francis, M. B., Reductive Alkylation of Proteins Using Iridium Catalyzed Transfer Hydrogenation. *J. Am. Chem. Soc.* **2005**, *127* (39), 13490-13491.
56. Jentoft, N.; Dearborn, D. G., Labeling of proteins by reductive methylation using sodium cyanoborohydride. *J. Biol. Chem.* **1979**, *254* (11), 4359-65.
57. Asano, S.; Patterson, J. T.; Gaj, T.; Barbas III, C. F., Site-Selective Labeling of a Lysine Residue in Human Serum Albumin. *Angew. Chem. Int. Ed.* **2014**, *53* (44), 11783-11786.
58. Patterson, J. T.; Wilson, H. D.; Asano, S.; Nilchan, N.; Fuller, R. P.; Roush, W. R.; Rader, C.; Barbas, C. F., Human Serum Albumin Domain I Fusion Protein for Antibody Conjugation. *Bioconj. Chem.* **2016**, *27* (10), 2271-2275.
59. Mazmanian, S. K.; Liu, G.; Ton-That, H.; Schneewind, O., Staphylococcus aureus Sortase, an Enzyme that Anchors Surface Proteins to the Cell Wall. *Science* **1999,** *285* (5428), 760-763.
60. Dixon, H. B. F., N-terminal modification of proteins—a review. *J. Protein Chem.* **1984**, *3* (1), 99-108.

61. Cennamo, C.; Carafoli, B.; Bonetti, E. P., Non-enzymatic Transamination between Peptides and Pyridoxal. Isolation of the 2,4-Dinitrophenylhydrazones of Some Ketopeptides. *J. Am. Chem. Soc.* **1956**, *78* (14), 3523-3527.
62. Dixon, H., Transamination of peptides. *Biochem. J.* **1964**, *92* (3), 661-666.
63. 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. *J. Am. Chem. Soc.* **2006**, *118* (32), 5433-5437.
64. O'Shannessy, D. J.; Dobersen, M. J.; Quarles, R. H., A novel procedure for labeling immunoglobulins by conjugation to oligosaccharide moieties. *Immunol. Lett.* **1984**, *8* (5), 273-277.
65. 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. *Bioconj. Chem.* **1992**, *3* (2), 138-146.
66. Spears, R. J.; Brabham, R. L.; Budhadev, D.; Keenan, T.; McKenna, S.; Walton, J.; Brannigan, J. A.; Brzozowski, A. M.; Wilkinson, A. J.; Plevin, M.; Fascione, M. A., Site-selective C–C modification of proteins at neutral pH using organocatalyst-mediated cross aldol ligations. *Chem. Sci.* **2018**, *9* (25), 5585-5593.
67. Diethelm, S.; Schafroth, M. A.; Carreira, E. M., Amine-Selective Bioconjugation Using Arene Diazonium Salts. *Org. Lett.* **2014**, *16* (15), 3908-3911.
68. Conibear, A. C.; Watson, E. E.; Payne, R. J.; Becker, C. F. W., Native chemical ligation in protein synthesis and semi-synthesis. *Chem. Soc. Rev.* **2018**, *47* (24), 9046-9068.
69. Schnolzer, M.; Kent, S., Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protease. *Science* **1992**, *256* (5054), 221-225.
70. Dawson, P.; Muir, T.; Clark-Lewis, I.; Kent, S., Synthesis of proteins by native chemical ligation. *Science* **1994**, *266* (5186), 776-779.
71. Spokoyny, A. M.; Zou, Y.; Ling, J. J.; Yu, H.; Lin, Y.-S.; Pentelute, B. L., A Perfluoroaryl-Cysteine SNAr Chemistry Approach to Unprotected Peptide Stapling. *J. Am. Chem. Soc.* **2013**, *135* (16), 5946-5949.
72. Paramonov, S. E.; Gauba, V.; Hartgerink, J. D., Synthesis of Collagen-like Peptide Polymers by Native Chemical Ligation. *Macromol.* **2005**, *38* (18), 7555-7561.
73. Coffinier, Y.; Olivier, C.; Perzyna, A.; Grandidier, B.; Wallart, X.; Durand, J. O.; Melnyk, O.; Stiévenard, D., Semicarbazide-Functionalized Si(111) Surfaces for the Site-Specific Immobilization of Peptides. *Langmuir* **2005**, *21* (4), 1489-1496.
74. Dendane, N.; Melnyk, O.; Xu, T.; Grandidier, B.; Boukherroub, R.; Stiévenard, D.; Coffinier, Y., Direct Characterization of Native Chemical Ligation of Peptides on Silicon Nanowires. *Langmuir* **2012**, *28* (37), 13336-13344.
75. Wieland, T.; Bokelmann, E.; Bauer, L.; Lang, H. U.; Lau, H., Über Peptidsynthesen. 8. Mitteilung Bildung von S-haltigen Peptiden durch intramolekulare Wanderung von Aminoacylresten. *Justus Liebigs Annalen der Chemie* **1953**, *583* (1), 129-149.
76. Muralidharan, V.; Muir, T. W., Protein ligation: an enabling technology for the biophysical analysis of proteins. *Nat. Meth.* **2006**, *3* (6), 429-438.
77. Ozawa, T.; Nogami, S.; Sato, M.; Ohya, Y.; Umezawa, Y., A Fluorescent Indicator for Detecting Protein−Protein Interactions in Vivo Based on Protein Splicing. *Anal. Chem.* **2000**, *72* (21), 5151-5157.
78. Scott, C. P.; Abel-Santos, E.; Wall, M.; Wahnon, D. C.; Benkovic, S. J., Production of cyclic peptides and proteins in vivo. *PNAS* **1999**, *96* (24), 13638-13643.
79. Evans, T. C.; Martin, D.; Kolly, R.; Panne, D.; Sun, L.; Ghosh, I.; Chen, L.; Benner, J.; Liu, X.-Q.; Xu, M.-Q., Protein trans-Splicing and Cyclization by a Naturally Split Intein from the dnaE Gene ofSynechocystis Species PCC6803. *J. Biol. Chem.* **2000**, *275* (13), 9091-9094.
80. Giriat, I.; Muir, T. W., Protein Semi-Synthesis in Living Cells. *J. Am. Chem. Soc.* **2003**, *125* (24), 7180-7181.

81. Bloom, S.; Liu, C.; Kölmel, D. K.; Qiao, J. X.; Zhang, Y.; Poss, M. A.; Ewing, W. R.; MacMillan, D. W. C., Decarboxylative alkylation for site-selective bioconjugation of native proteins via oxidation potentials. *Nat. Chem.* **2017**, *10*, 205.
82. Zuo, Z.; MacMillan, D. W. C., Decarboxylative Arylation of α-Amino Acids via Photoredox Catalysis: A One-Step Conversion of Biomass to Drug Pharmacophore. *J. Am. Chem. Soc.* **2014**, *136* (14), 5257-5260.
83. McGrath, N. A.; Andersen, K. A.; Davis, A. K. F.; Lomax, J. E.; Raines, R. T., Diazo compounds for the bioreversible esterification of proteins. *Chem. Sci.* **2015**, *6* (1), 752-755.
84. Totaro, K. A.; Liao, X.; Bhattacharya, K.; Finneman, J. I.; Sperry, J. B.; Massa, M. A.; Thorn, J.; Ho, S. V.; Pentelute, B. L., Systematic Investigation of EDC/sNHS-Mediated Bioconjugation Reactions for Carboxylated Peptide Substrates. *Bioconj. Chem.* **2016**, *27* (4), 994-1004.
85. Wang, P.; Zhang, S.; Meng, Q.; Liu, Y.; Shang, L.; Yin, Z., Site-Specific Chemical Modification of Peptide and Protein by Thiazolidinediones. *Org. Lett.* **2015**, *17* (6), 1361-1364.
86. Chen, I.; Howarth, M.; Lin, W.; Ting, A. Y., Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat. Meth.* **2005**, *2* (2), 99-104.
87. Beckett, D., Specificity and selectivity in post-translational biotin addition. *Biochem. Soc. Trans.* **2018**, *46* (6), 1577-1591.
88. Slavoff, S. A.; Chen, I.; Choi, Y.-A.; Ting, A. Y., Expanding the Substrate Tolerance of Biotin Ligase through Exploration of Enzymes from Diverse Species. *J. Am. Chem. Soc.* **2008**, *130* (4), 1160-1162.
89. Fernández-Suárez, M.; Baruah, H.; Martínez-Hernández, L.; Xie, K. T.; Baskin, J. M.; Bertozzi, C. R.; Ting, A. Y., Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nat. Biotechnol.* **2007**, *25* (12), 1483-1487.
90. Yao, J. Z.; Uttamapinant, C.; Poloukhtine, A.; Baskin, J. M.; Codelli, J. A.; Sletten, E. M.; Bertozzi, C. R.; Popik, V. V.; Ting, A. Y., Fluorophore Targeting to Cellular Proteins via Enzyme-Mediated Azide Ligation and Strain-Promoted Cycloaddition. *J. Am. Chem. Soc.* **2012**, *134* (8), 3720-3728.
91. Baruah, H.; Puthenveetil, S.; Choi, Y.-A.; Shah, S.; Ting, A. Y., An Engineered Aryl Azide Ligase for Site-Specific Mapping of Protein–Protein Interactions through Photo-Cross-Linking. *Angew. Chem. Int. Ed.* **2008**, *47* (37), 7018-7021.
92. Uttamapinant, C.; White, K. A.; Baruah, H.; Thompson, S.; Fernández-Suárez, M.; Puthenveetil, S.; Ting, A. Y., A fluorophore ligase for site-specific protein labeling inside living cells. *PNAS* **2010**, *107* (24), 10914-10919.
93. Cohen, J. D.; Zou, P.; Ting, A. Y., Site-Specific Protein Modification Using Lipoic Acid Ligase and Bis-Aryl Hydrazone Formation. *ChemBioChem* **2012**, *13* (6), 888-894.
94. 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* **2016,** *17* (2), 155-158.
95. Zhang, Y.; Park, K.-Y.; Suazo, K. F.; Distefano, M. D., Recent progress in enzymatic protein labelling techniques and their applications. *Chem. Soc. Rev.* **2018**, *47* (24), 9106-9136.
96. Lin, C.-W.; Ting, A. Y., Transglutaminase-Catalyzed Site-Specific Conjugation of Small-Molecule Probes to Proteins in Vitro and on the Surface of Living Cells. *J. Am. Chem. Soc.* **2006**, *128* (14), 4542-4543.
97. Fontana, A.; Spolaore, B.; Mero, A.; Veronese, F. M., Site-specific modification and PEGylation of pharmaceutical proteins mediated by transglutaminase. *Adv. Drug Del. Rev.* **2008**, *60* (1), 13-28.
98. Jiang, H.; Zhang, X.; Chen, X.; Aramsangtienchai, P.; Tong, Z.; Lin, H., Protein Lipidation: Occurrence, Mechanisms, Biological Functions, and Enabling Technologies. *Chem. Rev.* **2018**, *118* (3), 919-988.

99. Gauchet, C.; Labadie, G. R.; Poulter, C. D., Regio- and Chemoselective Covalent Immobilization of Proteins through Unnatural Amino Acids. *J. Am. Chem. Soc.* **2006**, *128* (29), 9274-9275.
100. Micali, E.; Chehade, K. A. H.; Isaacs, R. J.; Andres, D. A.; Spielmann, H. P., Protein Farnesyltransferase Isoprenoid Substrate Discrimination Is Dependent on Isoprene Double Bonds and Branched Methyl Groups. *Biochem.* **2001**, *40* (41), 12254-12265.
101. Blanden, M. J.; Suazo, K. F.; Hildebrandt, E. R.; Hardgrove, D. S.; Patel, M.; Saunders, W. P.; Distefano, M. D.; Schmidt, W. K.; Hougland, J. L., Efficient farnesylation of an extended C-terminal C(x)3X sequence motif expands the scope of the prenylated proteome. *J. Biol. Chem.* **2018**, *293* (8), 2770-2785.
102. Owen, D. J.; Alexandrov, K.; Rostkova, E.; Scheidig, A. J.; Goody, R. S.; Waldmann, H., Chemo-Enzymatic Synthesis of Fluorescent Rab 7 Proteins: Tools to Study Vesicular Trafficking in Cells. *Angew. Chem. Int. Ed.* **1999**, *38* (4), 509-512.
103. Chen, A. P. C.; Chen, Y.-H.; Liu, H.-P.; Li, Y.-C.; Chen, C.-T.; Liang, P.-H., Synthesis and Application of A Fluorescent Substrate Analogue to Study Ligand Interactions for Undecaprenyl Pyrophosphate Synthase. *J. Am. Chem. Soc.* **2002**, *124* (51), 15217-15224.
104. Dursina, B.; Reents, R.; Delon, C.; Wu, Y.; Kulharia, M.; Thutewohl, M.; Veligodsky, A.; Kalinin, A.; Evstifeev, V.; Ciobanu, D.; Szedlacsek, S. E.; Waldmann, H.; Goody, R. S.; Alexandrov, K., Identification and Specificity Profiling of Protein Prenyltransferase Inhibitors Using New Fluorescent Phosphoisoprenoids. *J. Am. Chem. Soc.* **2006**, *128* (9), 2822-2835.
105. Chehade, K. A. H.; Kiegiel, K.; Isaacs, R. J.; Pickett, J. S.; Bowers, K. E.; Fierke, C. A.; Andres, D. A.; Spielmann, H. P., Photoaffinity Analogues of Farnesyl Pyrophosphate Transferable by Protein Farnesyl Transferase. *J. Am. Chem. Soc.* **2002**, *124* (28), 8206-8219.
106. Turek, T. C.; Gaon, I.; Distefano, M. D.; Strickland, C. L., Synthesis of Farnesyl Diphosphate Analogues Containing Ether-Linked Photoactive Benzophenones and Their Application in Studies of Protein Prenyltransferases. *J. Org. Chem.* **2001**, *66* (10), 3253-3264.
107. Turek-Etienne, T. C.; Strickland, C. L.; Distefano, M. D., Biochemical and Structural Studies with Prenyl Diphosphate Analogues Provide Insights into Isoprenoid Recognition by Protein Farnesyl Transferase. *Biochem.* **2003**, *42* (13), 3716-3724.
108. Nguyen, U. T. T.; Guo, Z.; Delon, C.; Wu, Y.; Deraeve, C.; Fränzel, B.; Bon, R. S.; Blankenfeldt, W.; Goody, R. S.; Waldmann, H.; Wolters, D.; Alexandrov, K., Analysis of the eukaryotic prenylome by isoprenoid affinity tagging. *Nat. Chem. Bio.* **2009**, *5* (4), 227-235.
109. Kho, Y.; Kim, S. C.; Jiang, C.; Barma, D.; Kwon, S. W.; Cheng, J.; Jaunbergs, J.; Weinbaum, C.; Tamanoi, F.; Falck, J.; Zhao, Y., A tagging-via-substrate technology for detection and proteomics of farnesylated proteins. *PNAS* **2004**, *101* (34), 12479-12484.
110. Rose, M. W.; Rose, N. D.; Boggs, J.; Lenevich, S.; Xu, J.; Barany, G.; Distefano, M. D., Evaluation of geranylazide and farnesylazide diphosphate for incorporation of prenylazides into a CAAX box-containing peptide using protein farnesyltransferase*. *J. Peptide Res.* **2005**, *65* (6), 529-537.
111. Duckworth, B. P.; Xu, J.; Taton, T. A.; Guo, A.; Distefano, M. D., Site-Specific, Covalent Attachment of Proteins to a Solid Surface. *Bioconj. Chem.* **2006**, *17* (4), 967-974.
112. Nguyen, U. T. T.; Cramer, J.; Gomis, J.; Reents, R.; Gutierrez-Rodriguez, M.; Goody, R. S.; Alexandrov, K.; Waldmann, H., Exploiting the Substrate Tolerance of Farnesyltransferase for Site-Selective Protein Derivatization. *ChemBioChem* **2007**, *8* (4), 408-423.
113. Duckworth, B. P.; Chen, Y.; Wollack, J. W.; Sham, Y.; Mueller, J. D.; Taton, T. A.; Distefano, M. D., A Universal Method for the Preparation of Covalent Protein–DNA Conjugates for Use in Creating Protein Nanostructures. *Angew. Chem. Int. Ed.* **2007**, *46* (46), 8819-8822.
114. Das, D.; Tnimov, Z.; Nguyen, U. T. T.; Thimmaiah, G.; Lo, H.; Abankwa, D.; Wu, Y.; Goody, R. S.; Waldmann, H.; Alexandrov, K., Flexible and General Synthesis of Functionalized

Phosphoisoprenoids for the Study of Prenylation in vivo and in vitro. *ChemBioChem* **2012**, *13* (5), 674-683.
115. DeGraw, A. J.; Palsuledesai, C.; Ochocki, J. D.; Dozier, J. K.; Lenevich, S.; Rashidian, M.; Distefano, M. D., Evaluation of Alkyne-Modified Isoprenoids as Chemical Reporters of Protein Prenylation. *Chem. Biol. Drug Des.* **2010**, *76* (6), 460-471.
116. Rashidian, M.; Dozier, J. K.; Lenevich, S.; Distefano, M. D., Selective labeling of polypeptides using protein farnesyltransferase via rapid oxime ligation. *Chem. Comm.* **2010**, *46* (47), 8998-9000.
117. 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.* **2012**, *134* (20), 8455-8467.
118. Zhang, Y.; Blanden, M. J.; Sudheer, C.; Gangopadhyay, S. A.; Rashidian, M.; Hougland, J. L.; Distefano, M. D., Simultaneous Site-Specific Dual Protein Labeling Using Protein Prenyltransferases. *Bioconj. Chem.* **2015**, *26* (12), 2542-2553.
119. Dozier, J. K.; Khatwani, S. L.; Wollack, J. W.; Wang, Y.-C.; Schmidt-Dannert, C.; Distefano, M. D., Engineering Protein Farnesyltransferase for Enzymatic Protein Labeling Applications. *Bioconj. Chem.* **2014**, *25* (7), 1203-1212.
120. Pishesha, N.; Ingram, J. R.; Ploegh, H. L., Sortase A: A Model for Transpeptidation and Its Biological Applications. *Ann. Rev. Cell Dev. Biol.* **2018**, *34* (1), 163-188.
121. Schmohl, L.; Schwarzer, D., Sortase-mediated ligations for the site-specific modification of proteins. *Curr. Op. Chem. Biol.***2014**, *22*, 122-128.
122. Theile, C. S.; Witte, M. D.; Blom, A. E. M.; Kundrat, L.; Ploegh, H. L.; Guimaraes, C. P., Site-specific N-terminal labeling of proteins using sortase-mediated reactions. *Nat. Prot.* **2013**, *8*, 1800.
123. Guimaraes, C. P.; Witte, M. D.; Theile, C. S.; Bozkurt, G.; Kundrat, L.; Blom, A. E. M.; Ploegh, H. L., Site-specific C-terminal and internal loop labeling of proteins using sortase-mediated reactions. *Nature Protocols* **2013**, *8*, 1787.
124. Beerli, R. R.; Hell, T.; Merkel, A. S.; Grawunder, U., Sortase Enzyme-Mediated Generation of Site-Specifically Conjugated Antibody Drug Conjugates with High In Vitro and In Vivo Potency. *PLOS ONE* **2015,** *10* (7), e0131177.
125. Hirakawa, H.; Ishikawa, S.; Nagamune, T., Design of Ca2+-independent Staphylococcus aureus sortase A mutants. *Biotechnol. Bioeng.* **2012**, *109* (12), 2955-2961.
126. Popp, M. W.; Dougan, S. K.; Chuang, T.-Y.; Spooner, E.; Ploegh, H. L., Sortase-catalyzed transformations that improve the properties of cytokines. *PNAS* **2011**, *108* (8), 3169-3174.
127. Matsumoto, T.; Takase, R.; Tanaka, T.; Fukuda, H.; Kondo, A., Site-specific protein labeling with amine-containing molecules using Lactobacillus plantarum sortase. *Biotechnol. J.* **2012**, *7* (5), 642-648.
128. Das, S.; Pawale, V. S.; Dadireddy, V.; Singh, A. K.; Ramakumar, S.; Roy, R. P., Structure and specificity of a new class of Ca2+-independent housekeeping sortase from Streptomyces avermitilis provide insights into its non-canonical substrate preference. *J. Biol. Chem.* **2017**, *292* (17), 7244-7257.
129. Nikghalb, K. D.; Horvath, N. M.; Prelesnik, J. L.; Banks, O. G. B.; Filipov, P. A.; Row, R. D.; Roark, T. J.; Antos, J. M., Expanding the Scope of Sortase-Mediated Ligations by Using Sortase Homologues. *ChemBioChem* **2018**, *19* (2), 185-195.
130. Antos, J. M.; Miller, G. M.; Grotenbreg, G. M.; Ploegh, H. L., Lipid Modification of Proteins through Sortase-Catalyzed Transpeptidation. *J. Am. Chem. Soc.* **2008**, *130* (48), 16338-16343.
131. Popp, M. W.-L.; Antos, J. M.; Ploegh, H. L., Site-Specific Protein Labeling via Sortase-Mediated Transpeptidation. *Curr. Prot. Protein Sci.* **2009**, *56* (1), 15.3.1-15.3.9.

132. Staus, D. P.; Wingler, L. M.; Choi, M.; Pani, B.; Manglik, A.; Kruse, A. C.; Lefkowitz, R. J., Sortase ligation enables homogeneous GPCR phosphorylation to reveal diversity in β-arrestin coupling. *PNAS* **2018**, *115* (15), 3834-3839.
133. Tan, X.-L.; Pan, M.; Zheng, Y.; Gao, S.; Liang, L.-J.; Li, Y.-M., Sortase-mediated chemical protein synthesis reveals the bidentate binding of bisphosphorylated p62 with K63 diubiquitin. *Chem. Sci.* **2017**, *8* (10), 6881-6887.
134. Uth, C.; Zielonka, S.; Hörner, S.; Rasche, N.; Plog, A.; Orelma, H.; Avrutina, O.; Zhang, K.; Kolmar, H., A chemoenzymatic approach to protein immobilization onto crystalline cellulose nanoscaffolds. *Angew. Chem. Int. Ed.* **2014**, *53* (46), 12618-12623.
135. Le, R. K.; Raeeszadeh-Sarmazdeh, M.; Boder, E. T.; Frymier, P. D., Sortase-Mediated Ligation of PsaE-Modified Photosystem I from Synechocystis sp. PCC 6803 to a Conductive Surface for Enhanced Photocurrent Production on a Gold Electrode. *Langmuir* **2015**, *31* (3), 1180-1188.
136. Qafari, S. M.; Ahmadian, G.; Mohammadi, M., One-step purification and oriented attachment of protein A on silica and graphene oxide nanoparticles using sortase-mediated immobilization. *RSC Adv.* **2017**, *7* (88), 56006-56015.
137. Garg, S.; Singaraju, G. S.; Yengkhom, S.; Rakshit, S., Tailored Polyproteins Using Sequential Staple and Cut. *Bioconj. Chem.* **2018**, *29* (5), 1714-1719.
138. Ham, H. O.; Qu, Z.; Haller, C. A.; Dorr, B. M.; Dai, E.; Kim, W.; Liu, D. R.; Chaikof, E. L., In situ regeneration of bioactive coatings enabled by an evolved Staphylococcus aureus sortase A. *Nat. Comm.* **2016**, *7* (1), 11140.
139. Tabata, A.; Ohkubo, Y.; Anyoji, N.; Hojo, K.; Tomoyasu, T.; Tatematsu, Y.; Ohkura, K.; Nagamune, H., Development of a Sortase A-mediated Peptide-labeled Liposome Applicable to Drug-delivery Systems. *Anticancer Res.* **2015**, *35* (8), 4411-4417.
140. Silvius, J. R.; Leventis, R., A Novel "Prebinding" Strategy Dramatically Enhances Sortase-Mediated Coupling of Proteins to Liposomes. *Bioconj. Chem.* **2017**, *28* (4), 1271-1282.
141. Schoonen, L.; Pille, J.; Borrmann, A.; Nolte, R. J. M.; van Hest, J. C. M., Sortase A-Mediated N-Terminal Modification of Cowpea Chlorotic Mottle Virus for Highly Efficient Cargo Loading. *Bioconj. Chem.* **2015**, *26* (12), 2429-2434.
142. Patterson, D.; Schwarz, B.; Avera, J.; Western, B.; Hicks, M.; Krugler, P.; Terra, M.; Uchida, M.; McCoy, K.; Douglas, T., Sortase-Mediated Ligation as a Modular Approach for the Covalent Attachment of Proteins to the Exterior of the Bacteriophage P22 Virus-like Particle. *Bioconj. Chem.* **2017**, *28* (8), 2114-2124.
143. Cambria, E.; Renggli, K.; Ahrens, C. C.; Cook, C. D.; Kroll, C.; Krueger, A. T.; Imperiali, B.; Griffith, L. G., Covalent Modification of Synthetic Hydrogels with Bioactive Proteins via Sortase-Mediated Ligation. *Biomacromol.* **2015**, *16* (8), 2316-2326.
144. Gau, E.; Mate, D. M.; Zou, Z.; Oppermann, A.; Töpel, A.; Jakob, F.; Wöll, D.; Schwaneberg, U.; Pich, A., Sortase-Mediated Surface Functionalization of Stimuli-Responsive Microgels. *Biomacromol.* **2017**, *18* (9), 2789-2798.
145. Pasqual, G.; Chudnovskiy, A.; Tas, J. M. J.; Agudelo, M.; Schweitzer, L. D.; Cui, A.; Hacohen, N.; Victora, G. D., Monitoring T cell–dendritic cell interactions in vivo by intercellular enzymatic labelling. *Nature* **2018**, *553*, 496.
146. Glasgow, J. E.; Salit, M. L.; Cochran, J. R., In Vivo Site-Specific Protein Tagging with Diverse Amines Using an Engineered Sortase Variant. *J. Am. Chem. Soc.* **2016**, *138* (24), 7496-7499.
147. Wu, Q.; Ploegh, H. L.; Truttmann, M. C., Hepta-Mutant Staphylococcus aureus Sortase A (SrtA7m) as a Tool for in Vivo Protein Labeling in Caenorhabditis elegans. *ACS Chem. Biol.* **2017**, *12* (3), 664-673.
148. Ersfeld, K.; Wehland, J.; Plessmann, U.; Dodemont, H.; Gerke, V.; Weber, K., Characterization of the tubulin-tyrosine ligase. *J. Cell Biol.* **1993**, *120* (3), 725-732.

149. Monasterio, O.; Nova, E.; López-Brauet, A.; Lagos, R., Tubulin-tyrosine ligase catalyzes covalent binding of 3-fluoro-tyrosine to tubulin: kinetic and [19F]NMR studies. *FEBS lett.* **1995**, *374* (2), 165-168.
150. Bisig, C. G.; Purro, S. A.; Contín, M. A.; Barra, H. S.; Arce, C. A., Incorporation of 3-nitrotyrosine into the C-terminus of α-tubulin is reversible and not detrimental to dividing cells. *Euro. J. Biochem.* **2002**, *269* (20), 5037-5045.
151. Banerjee, A.; Panosian, T. D.; Mukherjee, K.; Ravindra, R.; Gal, S.; Sackett, D. L.; Bane, S., Site-Specific Orthogonal Labeling of the Carboxy Terminus of α-Tubulin. *ACS Chem. Biol.* **2010**, *5* (8), 777-785.
152. Schumacher, D.; Helma, J.; Mann, F. A.; Pichler, G.; Natale, F.; Krause, E.; Cardoso, M. C.; Hackenberger, C. P. R.; Leonhardt, H., Versatile and Efficient Site-Specific Protein Functionalization by Tubulin Tyrosine Ligase. *Angew. Chem. Int. Ed.* **2015**, *54* (46), 13787-13791.
153. Stengl, A.; Gerlach, M.; Kasper, M.-A.; Hackenberger, C. P. R.; Leonhardt, H.; Schumacher, D.; Helma, J., TuPPL: Tub-tag mediated C-terminal protein–protein-ligation using complementary click-chemistry handles. *Org. Biomol. Chem.* **2019**, *17* (20), 4964-4969.
154. Dierks, T.; Schmidt, B.; Borissenko, L. V.; Peng, J.; Preusser, A.; Mariappan, M.; von Figura, K., Multiple Sulfatase Deficiency Is Caused by Mutations in the Gene Encoding the Human C_α-Formylglycine Generating Enzyme. *Cell* **2003**, *113* (4), 435-444.
155. Carrico, I. S.; Carlson, B. L.; Bertozzi, C. R., Introducing genetically encoded aldehydes into proteins. *Nat. Chem. Biol.* **2007**, *3* (6), 321-322.
156. Rush, J. S.; Bertozzi, C. R., New Aldehyde Tag Sequences Identified by Screening Formylglycine Generating Enzymes in Vitro and in Vivo. *J. Am. Chem. Soc.* **2008**, *130* (37), 12240-12241.
157. Wu, P.; Shui, W.; Carlson, B. L.; Hu, N.; Rabuka, D.; Lee, J.; Bertozzi, C. R., Site-specific chemical modification of recombinant proteins produced in mammalian cells by using the genetically encoded aldehyde tag. *PNAS* **2009**, *106* (9), 3000-3005.
158. Agarwal, P.; van der Weijden, J.; Sletten, E. M.; Rabuka, D.; Bertozzi, C. R., A Pictet-Spengler ligation for protein chemical modification. *PNAS* **2013**, *110* (1), 46-51.
159. Agarwal, P.; Kudirka, R.; Albers, A. E.; Barfield, R. M.; de Hart, G. W.; Drake, P. M.; Jones, L. C.; Rabuka, D., Hydrazino-Pictet-Spengler Ligation as a Biocompatible Method for the Generation of Stable Protein Conjugates. *Bioconj. Chem.* **2013**, *24* (6), 846-851.
160. Drake, P. M.; Albers, A. E.; Baker, J.; Banas, S.; Barfield, R. M.; Bhat, A. S.; de Hart, G. W.; Garofalo, A. W.; Holder, P.; Jones, L. C.; Kudirka, R.; McFarland, J.; Zmolek, W.; Rabuka, D., 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. *Bioconj. Chem.* **2014**, *25* (7), 1331-1341.
161. Kudirka, R. A.; Barfield, R. M.; McFarland, J. M.; Drake, P. M.; Carlson, A.; Bañas, S.; Zmolek, W.; Garofalo, A. W.; Rabuka, D., Site-Specific Tandem Knoevenagel Condensation–Michael Addition To Generate Antibody–Drug Conjugates. *ACS Med. Chem. Lett.* **2016**, *7* (11), 994-998.
162. Hallam, T. J.; Wold, E.; Wahl, A.; Smider, V. V., Antibody Conjugates with Unnatural Amino Acids. *Molecul. Pharma.* **2015**, *12* (6), 1848-1862.
163. de Graaf, A. J.; Kooijman, M.; Hennink, W. E.; Mastrobattista, E., Nonnatural Amino Acids for Site-Specific Protein Conjugation. *Bioconj. Chem.* **2009**, *20* (7), 1281-1295.
164. Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R., Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *PNAS* **2002**, *99* (1), 19-24.
165. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G., Expanding the Genetic Code of *Escherichia coli*. *Science* **2001**, *292* (5516), 498-500.

166. Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G., An Expanded Eukaryotic Genetic Code. *Science* **2003**, *301* (5635), 964-967.
167. Liu, C. C.; Schultz, P. G., Adding New Chemistries to the Genetic Code. *Ann. Rev. Biochem.* **2010**, *79* (1), 413-444.
168. Ovaa, H.; wals, k., Unnatural amino acid incorporation in E. coli: current and future applications in the design of therapeutic proteins. *Front. Chem.* **2014**, *2* (15).
169. Greiss, S.; Chin, J. W., Expanding the Genetic Code of an Animal. *J. Am. Chem. Soc.* **2011**, *133* (36), 14196-14199.
170. Parrish, A. R.; She, X.; Xiang, Z.; Coin, I.; Shen, Z.; Briggs, S. P.; Dillin, A.; Wang, L., Expanding the Genetic Code of Caenorhabditis elegans Using Bacterial Aminoacyl-tRNA Synthetase/tRNA Pairs. *ACS Chem. Biol.* **2012**, *7* (7), 1292-1302.
171. Johnson, D. B. F.; Wang, C.; Xu, J.; Schultz, M. D.; Schmitz, R. J.; Ecker, J. R.; Wang, L., Release Factor One Is Nonessential in Escherichia coli. *ACS Chem. Biol.* **2012**, *7* (8), 1337-1344.
172. Yamaguchi, A.; Matsuda, T.; Ohtake, K.; Yanagisawa, T.; Yokoyama, S.; Fujiwara, Y.; Watanabe, T.; Hohsaka, T.; Sakamoto, K., Incorporation of a Doubly Functionalized Synthetic Amino Acid into Proteins for Creating Chemical and Light-Induced Conjugates. *Bioconj. Chem.* **2016**, *27* (1), 198-206.
173. Tian, Y.; Jacinto, M. P.; Zeng, Y.; Yu, Z.; Qu, J.; Liu, W. R.; Lin, Q., Genetically Encoded 2-Aryl-5-carboxytetrazoles for Site-Selective Protein Photo-Cross-Linking. *J. Am. Chem. Soc.* **2017**, *139* (17), 6078-6081.
174. Kato, Y., Translational Control using an Expanded Genetic Code. *Int. J. Mol. Sci.***2019**, *20* (4), 887.
175. Schilling, C. I.; Jung, N.; Biskup, M.; Schepers, U.; Bräse, S., Bioconjugation viaazide–Staudinger ligation: an overview. *Chem. Soc. Rev.* **2011**, *40* (9), 4840-4871.
176. King, M.; Wagner, A., Developments in the Field of Bioorthogonal Bond Forming Reactions—Past and Present Trends. *Bioconj. Chem.* **2014**, *25* (5), 825-839.
177. Saxon, E.; Bertozzi, C. R., Cell Surface Engineering by a Modified Staudinger Reaction. *Science* **2000**, *287* (5460), 2007-2010.
178. Saxon, E.; Armstrong, J. I.; Bertozzi, C. R., A "Traceless" Staudinger Ligation for the Chemoselective Synthesis of Amide Bonds. *Org. Lett.* **2000**, *2* (14), 2141-2143.
179. Serwa, R.; Majkut, P.; Horstmann, B.; Swiecicki, J.-M.; Gerrits, M.; Krause, E.; Hackenberger, C. P. R., Site-specific PEGylation of proteins by a Staudinger-phosphite reaction. *Chem. Sci.* **2010**, *1* (5), 596-602.
180. Vallée, M. R. J.; Majkut, P.; Wilkening, I.; Weise, C.; Müller, G.; Hackenberger, C. P. R., Staudinger-Phosphonite Reactions for the Chemoselective Transformation of Azido-Containing Peptides and Proteins. *Org. Lett.* **2011**, *13* (20), 5440-5443.
181. Debets, M. F.; van Berkel, S. S.; Dommerholt, J.; Dirks, A. J.; Rutjes, F. P. J. T.; van Delft, F. L., Bioconjugation with Strained Alkenes and Alkynes. *Acc. Chem. Res.* **2011**, *44* (9), 805-815.
182. Kolb, H. C.; Finn, M. G.; Sharpless, K. B., Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem. Int. Ed.* **2001**, *40* (11), 2004-2021.
183. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B., A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes. *Angew. Chem. Int. Ed.* **2002**, *41* (14), 2596-2599.
184. Wang, Q.; Chan, T. R.; Hilgraf, R.; Fokin, V. V.; Sharpless, K. B.; Finn, M. G., Bioconjugation by Copper(I)-Catalyzed Azide-Alkyne [3 + 2] Cycloaddition. *J. Am. Chem. Soc.* **2003**, *125* (11), 3192-3193.
185. Himo, F.; Lovell, T.; Hilgraf, R.; Rostovtsev, V. V.; Noodleman, L.; Sharpless, K. B.; Fokin, V. V., Copper(I)-Catalyzed Synthesis of Azoles. DFT Study Predicts Unprecedented Reactivity and Intermediates. *J. Am. Chem. Soc.* **2005**, *127* (1), 210-216.

186. Spicer, C. D.; Davis, B. G., Selective chemical protein modification. *Nat. Comm.* **2014**, *5*, 4740.
187. Raliski, B. K.; Howard, C. A.; Young, D. D., Site-Specific Protein Immobilization Using Unnatural Amino Acids. *Bioconj. Chem.* **2014**, *25* (11), 1916-1920.
188. Boutureira, O.; D'Hooge, F.; Fernández-González, M.; Bernardes, G. J. L.; Sánchez-Navarro, M.; Koeppe, J. R.; Davis, B. G., Fluoroglycoproteins: ready chemical site-selective incorporation of fluorosugars into proteins. *Chem. Comm.* **2010**, *46* (43), 8142-8144.
189. Díaz Velázquez, H.; Ruiz García, Y.; Vandichel, M.; Madder, A.; Verpoort, F., Water-soluble NHC-Cu catalysts: applications in click chemistry, bioconjugation and mechanistic analysis. *Org. Biomol. Chem.***2014**, *12* (46), 9350-9356.
190. Li, S.; Cai, H.; He, J.; Chen, H.; Lam, S.; Cai, T.; Zhu, Z.; Bark, S. J.; Cai, C., Extent of the Oxidative Side Reactions to Peptides and Proteins During the CuAAC Reaction. *Bioconj. Chem.* **2016**, *27* (10), 2315-2322.
191. 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.* **2004**, *126* (46), 15046-15047.
192. Wittig, G.; Krebs, A., Zur Existenz niedergliedriger Cycloalkine, I. *Chemische Berichte* **1961**, *94* (12), 3260-3275.
193. Baskin, J. M.; Prescher, J. A.; Laughlin, S. T.; Agard, N. J.; Chang, P. V.; Miller, I. A.; Lo, A.; Codelli, J. A.; Bertozzi, C. R., Copper-free click chemistry for dynamic *in vivo* imaging. *PNAS* **2007**, *104* (43), 16793-16797.
194. Kim, E. J.; Kang, D. W.; Leucke, H. F.; Bond, M. R.; Ghosh, S.; Love, D. C.; Ahn, J.-S.; Kang, D.-O.; Hanover, J. A., Optimizing the selectivity of DIFO-based reagents for intracellular bioorthogonal applications. *Carb. Res.* **2013**, *377*, 18-27.
195. Ning, X.; Guo, J.; Wolfert, M. A.; Boons, G.-J., Visualizing Metabolically Labeled Glycoconjugates of Living Cells by Copper-Free and Fast Huisgen Cycloadditions. *Angew. Chem. Int. Ed.* **2008**, *47* (12), 2253-2255.
196. Laughlin, S. T.; Baskin, J. M.; Amacher, S. L.; Bertozzi, C. R., In Vivo Imaging of Membrane-Associated Glycans in Developing Zebrafish. *Science* **2008**, *320* (5876), 664-667.
197. Jewett, J. C.; Sletten, E. M.; Bertozzi, C. R., Rapid Cu-Free Click Chemistry with Readily Synthesized Biarylazacyclooctynones. *J. Am. Chem. Soc.* **2010**, *132* (11), 3688-3690.
198. Dommerholt, J.; Schmidt, S.; Temming, R.; Hendriks, L. J. A.; Rutjes, F. P. J. T.; van Hest, J. C. M.; Lefeber, D. J.; Friedl, P.; van Delft, F. L., Readily Accessible Bicyclononynes for Bioorthogonal Labeling and Three-Dimensional Imaging of Living Cells. *Angew. Chem. Int. Ed.* **2010**, *49* (49), 9422-9425.
199. Conte, M. L.; Staderini, S.; Marra, A.; Sanchez-Navarro, M.; Davis, B. G.; Dondoni, A., Multi-molecule reaction of serum albumin can occur through thiol-yne coupling. *Chem. Comm.* **2011**, *47* (39), 11086-11088.
200. Ning, X.; Temming, R. P.; Dommerholt, J.; Guo, J.; Ania, D. B.; Debets, M. F.; Wolfert, M. A.; Boons, G.-J.; van Delft, F. L., Protein Modification by Strain-Promoted Alkyne–Nitrone Cycloaddition. *Angew. Chem. Int. Ed.* **2010**, *49* (17), 3065-3068.
201. Colombo, M.; Sommaruga, S.; Mazzucchelli, S.; Polito, L.; Verderio, P.; Galeffi, P.; Corsi, F.; Tortora, P.; Prosperi, D., Site-Specific Conjugation of ScFvs Antibodies to Nanoparticles by Bioorthogonal Strain-Promoted Alkyne–Nitrone Cycloaddition. *Angew. Chem. Int. Ed.* **2012**, *51* (2), 496-499.
202. Borrmann, A.; Fatunsin, O.; Dommerholt, J.; Jonker, A. M.; Löwik, D. W. P. M.; van Hest, J. C. M.; van Delft, F. L., Strain-Promoted Oxidation-Controlled Cyclooctyne–1,2-Quinone Cycloaddition (SPOCQ) for Fast and Activatable Protein Conjugation. *Bioconj. Chem.* **2015**, *26* (2), 257-261.
203. Sanders, B. C.; Friscourt, F.; Ledin, P. A.; Mbua, N. E.; Arumugam, S.; Guo, J.; Boltje, T. J.; Popik, V. V.; Boons, G.-J., Metal-Free Sequential [3 + 2]-Dipolar Cycloadditions using

Cyclooctynes and 1,3-Dipoles of Different Reactivity. *J. Am. Chem. Soc.* **2011**, *133* (4), 949-957.
204. Wallace, S.; Chin, J. W., Strain-promoted sydnone bicyclo-[6.1.0]-nonyne cycloaddition. *Chemical Science* **2014**, *5* (5), 1742-1744.
205. Plougastel, L.; Koniev, O.; Specklin, S.; Decuypere, E.; Créminon, C.; Buisson, D.-A.; Wagner, A.; Kolodych, S.; Taran, F., 4-Halogeno-sydnones for fast strain promoted cycloaddition with bicyclo-[6.1.0]-nonyne. *Chem. Comm.* **2014**, *50* (66), 9376-9378.
206. Stöckmann, H.; Neves, A. A.; Stairs, S.; Brindle, K. M.; Leeper, F. J., Exploring isonitrile-based click chemistry for ligation with biomolecules. *Org. Biomol. Chem.* **2011**, *9* (21), 7303-7305.
207. Hoyt, E. A.; Cal, P. M. S. D.; Oliveira, B. L.; Bernardes, G. J. L., Contemporary approaches to site-selective protein modification. *Nat. Rev. Chem.* **2019**, *3* (3), 147-171.
208. Li, Y.; Pan, M.; Li, Y.; Huang, Y.; Guo, Q., Thiol–yne radical reaction mediated site-specific protein labeling via genetic incorporation of an alkynyl-l-lysine analogue. *Org. Biomol. Chem.* **2013**, *11* (16), 2624-2629.
209. Hommersom, C. A.; Matt, B.; van der Ham, A.; Cornelissen, J. J. L. M.; Katsonis, N., Versatile post-functionalization of the external shell of cowpea chlorotic mottle virus by using click chemistry. *Org. Biomol. Chem.* **2014**, *12* (24), 4065-4069.
210. van Berkel, S. S.; Dirks, A. J.; Debets, M. F.; van Delft, F. L.; Cornelissen, J. J. L. M.; Nolte, R. J. M.; Rutjes, F. P. J. T., Metal-Free Triazole Formation as a Tool for Bioconjugation. *ChemBioChem* **2007**, *8* (13), 1504-1508.
211. Gattner, M. J.; Ehrlich, M.; Vrabel, M., Sulfonyl azide-mediated norbornene aziridination for orthogonal peptide and protein labeling. *Chem. Comm.* **2014**, *50* (83), 12568-12571.
212. Blackman, M. L.; Royzen, M.; Fox, J. M., Tetrazine Ligation: Fast Bioconjugation Based on Inverse-Electron-Demand Diels−Alder Reactivity. *J. Am. Chem. Soc.* **2008**, *130* (41), 13518-13519.
213. Devaraj, N. K.; Weissleder, R.; Hilderbrand, S. A., Tetrazine-Based Cycloadditions: Application to Pretargeted Live Cell Imaging. *Bioconj. Chem.* **2008**, *19* (12), 2297-2299.
214. Lang, K.; Davis, L.; Wallace, S.; Mahesh, M.; Cox, D. J.; Blackman, M. L.; Fox, J. M.; Chin, J. W., Genetic Encoding of Bicyclononynes and trans-Cyclooctenes for Site-Specific Protein Labeling in Vitro and in Live Mammalian Cells via Rapid Fluorogenic Diels–Alder Reactions. *J. Am. Chem. Soc.* **2012**, *134* (25), 10317-10320.
215. Karver, M. R.; Weissleder, R.; Hilderbrand, S. A., Bioorthogonal Reaction Pairs Enable Simultaneous, Selective, Multi-Target Imaging. *Angew. Chem. Int. Ed.* **2012**, *51* (4), 920-922.
216. Lampkowski, J. S.; Villa, J. K.; Young, T. S.; Young, D. D., Development and Optimization of Glaser–Hay Bioconjugations. *Angew. Chem. Int. Ed.* **2015**, *54* (32), 9343-9346.
217. Kwan, T. T. L.; Boutureira, O.; Frye, E. C.; Walsh, S. J.; Gupta, M. K.; Wallace, S.; Wu, Y.; Zhang, F.; Sore, H. F.; Galloway, W. R. J. D.; Chin, J. W.; Welch, M.; Bernardes, G. J. L.; Spring, D. R., Protein modification via alkyne hydrosilylation using a substoichiometric amount of ruthenium(ii) catalyst. *Chem. Sci.* **2017**, *8* (5), 3871-3878.
218. Song, W.; Wang, Y.; Qu, J.; Madden, M. M.; Lin, Q., A Photoinducible 1,3-Dipolar Cycloaddition Reaction for Rapid, Selective Modification of Tetrazole-Containing Proteins. *Angew. Chem. Int. Ed.* **2008**, *47* (15), 2832-2835.
219. Song, W.; Wang, Y.; Qu, J.; Lin, Q., Selective Functionalization of a Genetically Encoded Alkene-Containing Protein via "Photoclick Chemistry" in Bacterial Cells. *J. Am. Chem. Soc.* **2008**, *130* (30), 9654-9655.
220. Yu, Z.; Pan, Y.; Wang, Z.; Wang, J.; Lin, Q., Genetically Encoded Cyclopropene Directs Rapid, Photoclick-Chemistry-Mediated Protein Labeling in Mammalian Cells. *Angew. Chem. Int. Ed.* **2012**, *51* (42), 10600-10604.

221. Li, Z.; Qian, L.; Li, L.; Bernhammer, J. C.; Huynh, H. V.; Lee, J.-S.; Yao, S. Q., Tetrazole Photoclick Chemistry: Reinvestigating Its Suitability as a Bioorthogonal Reaction and Potential Applications. *Angew. Chem. Int. Ed.* **2016**, *55* (6), 2002-2006.
222. Chen, X.; Wu, Y.-W., Selective chemical labeling of proteins. *Org. Biomol. Chem.* **2016**, *14* (24), 5417-5439.
223. Lee, Y.-J.; Kurra, Y.; Liu, W. R., Phospha-Michael Addition as a New Click Reaction for Protein Functionalization. *ChemBioChem* **2016**, *17* (6), 456-461.
224. Tomlin, F. M.; Gordon, C. G.; Han, Y.; Wu, T. S.; Sletten, E. M.; Bertozzi, C. R., Site-specific incorporation of quadricyclane into a protein and photocleavage of the quadricyclane ligation adduct. *Bioorg. Med. Chem.* **2018**, *26* (19), 5280-5290.
225. Guo, J.; Wang, J.; Lee, J. S.; Schultz, P. G., Site-Specific Incorporation of Methyl- and Acetyl-Lysine Analogues into Recombinant Proteins. *Angew. Chem. Int. Ed.* **2008**, *47* (34), 6399-6401.
226. Chalker, J. M.; Gunnoo, S. B.; Boutureira, O.; Gerstberger, S. C.; Fernández-González, M.; Bernardes, G. J. L.; Griffin, L.; Hailu, H.; Schofield, C. J.; Davis, B. G., Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chem. Sci.* **2011**, *2* (9), 1666-1676.
227. 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.* **2012**, *51* (8), 1835-1839.
228. Lin, Y. A.; Chalker, J. M.; Floyd, N.; Bernardes, G. J. L.; Davis, B. G., Allyl Sulfides Are Privileged Substrates in Aqueous Cross-Metathesis: Application to Site-Selective Protein Modification. *J. Am. Chem. Soc.* **2008**, *130* (30), 9642-9643.
229. Lin, Y. A.; Boutureira, O.; Lercher, L.; Bhushan, B.; Paton, R. S.; Davis, B. G., Rapid Cross-Metathesis for Reversible Protein Modifications via Chemical Access to Se-Allyl-selenocysteine in Proteins. *J. Am. Chem. Soc.* **2013**, *135* (33), 12156-12159.
230. Freedy, A. M.; Matos, M. J.; Boutureira, O.; Corzana, F.; Guerreiro, A.; Akkapeddi, P.; Somovilla, V. J.; Rodrigues, T.; Nicholls, K.; Xie, B.; Jiménez-Osés, G.; Brindle, K. M.; Neves, A. A.; Bernardes, G. J. L., Chemoselective Installation of Amine Bonds on Proteins through Aza-Michael Ligation. *J. Am. Chem. Soc.* **2017**, *139* (50), 18365-18375.
231. Dadová, J.; Wu, K.-J.; Isenegger, P. G.; Errey, J. C.; Bernardes, G. J. L.; Chalker, J. M.; Raich, L.; Rovira, C.; Davis, B. G., Precise Probing of Residue Roles by Post-Translational β,γ-C,N Aza-Michael Mutagenesis in Enzyme Active Sites. *ACS Cent. Sci.* **2017**, *3* (11), 1168-1173.
232. Wright, T. H.; Bower, B. J.; Chalker, J. M.; Bernardes, G. J. L.; Wiewiora, R.; Ng, W.-L.; Raj, R.; Faulkner, S.; Vallée, M. R. J.; Phanumartwiwath, A.; Coleman, O. D.; Thézénas, M.-L.; Khan, M.; Galan, S. R. G.; Lercher, L.; Schombs, M. W.; Gerstberger, S.; Palm-Espling, M. E.; Baldwin, A. J.; Kessler, B. M.; Claridge, T. D. W.; Mohammed, S.; Davis, B. G., Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science* **2016**, *354* (6312),
233. Yang, A.; Ha, S.; Ahn, J.; Kim, R.; Kim, S.; Lee, Y.; Kim, J.; Söll, D.; Lee, H.-Y.; Park, H.-S., A chemical biology route to site-specific authentic protein modifications. *Science* **2016**, *354* (6312), 623-626.
234. Spears, R. J.; Fascione, M. A., Site-selective incorporation and ligation of protein aldehydes. *Org. Biomol. Chem.* **2016**, *14* (32), 7622-38.
235. Brabham, R. L.; Spears, R. J.; Walton, J.; Tyagi, S.; Lemke, E. A.; Fascione, M. A., Palladium-unleashed proteins: gentle aldehyde decaging for site-selective protein modification. *Chem. Comm.* **2018**, *54* (12), 1501-1504.
236. Kölmel, D. K.; Kool, E. T., Oximes and Hydrazones in Bioconjugation: Mechanism and Catalysis. *Chem. Rev.* **2017**, *117* (15), 10358-10376.

237. 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.* **2014**, *136* (23), 8149-8152.
238. Kudirka, R.; Barfield, Robyn M.; McFarland, J.; Albers, Aaron E.; de Hart, Gregory W.; Drake, Penelope M.; Holder, Patrick G.; Banas, S.; Jones, Lesley C.; Garofalo, Albert W.; Rabuka, D., Generating Site-Specifically Modified Proteins via a Versatile and Stable Nucleophilic Carbon Ligation. *Chem. Biol.* **2015**, *22* (2), 293-298.
239. Chalker, J. M.; Wood, C. S. C.; Davis, B. G., A Convenient Catalyst for Aqueous and Protein Suzuki−Miyaura Cross-Coupling. *J. Am. Chem. Soc.* **2009**, *131* (45), 16346-16347.
240. Alam, J.; Keller, T. H.; Loh, T.-P., Functionalization of Peptides and Proteins by Mukaiyama Aldol Reaction. *J. Am. Chem. Soc.* **2010**, *132* (28), 9546-9548.
241. Shih, H.-W.; Prescher, J. A., A Bioorthogonal Ligation of Cyclopropenones Mediated by Triarylphosphines. *J. Am. Chem. Soc.* **2015**, *137* (32), 10036-10039.
242. 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.* **2017**, *139* (21), 7370-7375.
243. Seebach, D., Methods of Reactivity Umpolung. *Angew. Chem. Int. Ed.* **1979**, *18* (4), 239-258.
244. Griera, R.; Rigat, L.; Alvarez, M.; Joule, J. A., Reactions of 1-methyl-4-quinolone with 2-lithio-1,3-dithianes. *J. Am. Chem. Soc. Perkin Transactions 1* **1992**, (10), 1223-1227.
245. Wadi, A.; López-Calahorra, F.; Alvarez, M.; Joule, J. A., An improved annelation method with methyl2-(1,3-dithian-2-yl)benzoate as a bidentate synthon. *Tet. Lett.* **1992**, *33* (25), 3679-3682.
246. Breslow, R., On the Mechanism of Thiamine Action. IV.1 Evidence from Studies on Model Systems. *J. Am. Chem. Soc.* **1958**, *80* (14), 3719-3726.
247. Stetter, H., Catalyzed Addition of Aldehydes to Activated Double Bonds—A New Synthetic Approach. *Angew. Chem. Int. Ed.* **1976**, *15* (11), 639-647.
248. Stetter, H.; Schreckenberg, M., A New Method for Addition of Aldehydes to Activated Double Bonds. *Angew. Chem. Int. Ed.* **1973**, *12* (1), 81-81.
249. Bugaut, X.; Glorius, F., Organocatalytic umpolung: N-heterocyclic carbenes and beyond. *Chem. Soc. Rev.* **2012**, *41* (9), 3511-22.
250. Bharadwaj, A. R.; Scheidt, K. A., Catalytic Multicomponent Synthesis of Highly Substituted Pyrroles Utilizing a One-Pot Sila-Stetter/Paal-Knorr Strategy. *Org. Lett.* **2004,** *6*, 2465-2468.
251. Cardinal-David, B.; Scheidt, K. A., Carbene catalysis. Beyond the benzoin and Stetter reactions. *Top. Organomet. Chem.* **2013**, *44* (, 233-260.
252. Mattson, A. E.; Bharadwaj, A. R.; Scheidt, K. A., The Thiazolium-Catalyzed Sila-Stetter Reaction: Conjugate Addition of Acylsilanes to Unsaturated Esters and Ketones. *J. Am. Chem. Soc.* **2004**, *126,* 2314-2315.
253. Mattson, A. E.; Bharadwaj, A. R.; Zuhl, A. M.; Scheidt, K. A., Thiazolium-Catalyzed Additions of Acylsilanes: A General Strategy for Acyl Anion Addition Reactions. *J. Org. Chem.* **2006**, *71*, 5715-5724.
254. Serafimova, I. M.; Pufall, M. A.; Krishnan, S.; Duda, K.; Cohen, M. S.; Maglathlin, R. L.; McFarland, J. M.; Miller, R. M.; Frödin, M.; Taunton, J., Reversible targeting of noncatalytic cysteines with chemically tuned electrophiles. *Nat. Chem. Biol.* **2012**, *8* (5), 471-476.
255. Krishnan, S.; Miller, R. M.; Tian, B.; Mullins, R. D.; Jacobson, M. P.; Taunton, J., Design of Reversible, Cysteine-Targeted Michael Acceptors Guided by Kinetic and Computational Analysis. *J. Am. Chem. Soc.* **2014**, *136* (36), 12624-12630.
256. Krenske, E. H.; Petter, R. C.; Houk, K. N., Kinetics and Thermodynamics of Reversible Thiol Additions to Mono- and Diactivated Michael Acceptors: Implications for the Design of Drugs That Bind Covalently to Cysteines. *J. Org. Chem.* **2016**, *81* (23), 11726-11733.

257. Chiang, P.-C.; Bode, J. W., On the Role of CO2 in NHC-Catalyzed Oxidation of Aldehydes. *Org. Lett.* **2011**, *13* (9), 2422-2425.
258. Delany, E. G.; Fagan, C.-L.; Gundala, S.; Mari, A.; Broja, T.; Zeitler, K.; Connon, S. J., NHC-catalysed aerobic aldehyde-esterifications with alcohols: no additives or cocatalysts required. *Chem. Comm.* **2013**, *49* (58), 6510-6512.
259. De Sarkar, S.; Biswas, A.; Samanta, R. C.; Studer, A., Catalysis with N-Heterocyclic Carbenes under Oxidative Conditions. *Chem. Eur. J.* **2013**, *19* (15), 4664-4678.
260. Bortolini, O.; Chiappe, C.; Fogagnolo, M.; Massi, A.; Pomelli, C. S., Formation, Oxidation, and Fate of the Breslow Intermediate in the N-Heterocyclic Carbene-Catalyzed Aerobic Oxidation of Aldehydes. *J. Org. Chem.* **2017**, *82* (1), 302-312.
261. Butler, I. B.; Schoonen, M. A. A.; Rickard, D. T., Removal of dissolved oxygen from water: A comparison of four common techniques. *Talanta* **1994**, *41* (2), 211-215.
262. Stetter, H.; Kuhlmann, H., Addition von Aldehyden an aktivierte Doppelbindungen; VII1. Eine neue einfache Synthese von cis-Jasmon und Dihydrojasmon. *Synthesis* **1975**, *1975* (06), 379-380.
263. Flanigan, D. M.; Romanov-Michailidis, F.; White, N. A.; Rovis, T., Organocatalytic Reactions Enabled by N-Heterocyclic Carbenes. *Chem. Rev.* **2015**, *115* (17), 9307-87.
264. Kohn, H.; Olofson, R. A., 1,4-Dimethyl-1,4-dihydro-1,2,4,5-tetrazine and its N-alkyl salt. Synthesis, structure, and chemistry. *J. Org. Chem.* **1972**, *37* (22), 3504-3509.
265. Marion, N.; Diez-Gonzalez, S.; Nolan, S. P., N-heterocyclic carbenes as organocatalysts. *Angew. Chem. Int. Ed.* **2007**, *46* (17), 2988-3000.
266. Nair, V.; Vellalath, S.; Babu, B. P., Recent advances in carbon-carbon bond-forming reactions involving homoenolates generated by NHC catalysis. *Chemical Society Reviews* **2008**, *37* (12), 2691-2698.
267. Bugaut, X.; Glorius, F., Organocatalytic umpolung: N-heterocyclic carbenes and beyond. *Chem. Soc. Rev.* **2012**, *41* (9), 3511-3522.
268. Sohn, S. S.; Bode, J. W., Catalytic Generation of Activated Carboxylates from Enals: A Product-Determining Role for the Base. *Org. Lett.* **2005**, *7* (18), 3873-3876.
269. Wöhler; Liebig, Untersuchungen über das Radikal der Benzoesäure. *Annalen der Pharmacie* **1832**, *3* (3), 249-282.
270. Lapworth, A., XCVI.-Reactions involving the addition of hydrogen cyanide to carbon compounds. *J. Am. Chem. Soc. Trans.* **1903**, *83* (0), 995-1005.
271. Delany, E. G.; Fagan, C.-L.; Gundala, S.; Zeitler, K.; Connon, S. J., Aerobic oxidation of NHC-catalysed aldehyde esterifications with alcohols: benzoin, not the Breslow intermediate, undergoes oxidation. *Chem. Comm.* **2013**, *49* (58), 6513-6515.
272. O'Toole, S. E.; Rose, C. A.; Gundala, S.; Zeitler, K.; Connon, S. J., Highly Chemoselective Direct Crossed Aliphatic−Aromatic Acyloin Condensations with Triazolium-Derived Carbene Catalysts. *J. Org. Chem.* **2011**, *76* (2), 347-357.
273. Hofmeister, F., Zur Lehre von der Wirkung der Salze. *Archiv für experimentelle Pathologie und Pharmakologie* **1888**, *24* (4), 247-260.
274. Patel, R.; Kumari, M.; Khan, A. B., Recent Advances in the Applications of Ionic Liquids in Protein Stability and Activity: A Review. *App. Biochem. Biotechnol.* **2014**, *172* (8), 3701-3720.
275. Chen, X.; Yang, T.; Kataoka, S.; Cremer, P. S., Specific Ion Effects on Interfacial Water Structure near Macromolecules. *J. Am. Chem. Soc.* **2007**, *129* (40), 12272-12279.
276. Pegram, L. M.; Record, M. T., Hofmeister Salt Effects on Surface Tension Arise from Partitioning of Anions and Cations between Bulk Water and the Air−Water Interface. *J. Phys. Chem. B* **2007**, *111* (19), 5411-5417.
277. Tobias, D. J.; Hemminger, J. C., Getting Specific About Specific Ion Effects. *Science* **2008**, *319* (5867), 1197-1198.

278. Kool, E. T.; Breslow, R., Dichotomous salt effects in the hydrophobic acceleration of the benzoin condensation. *J. Am. Chem. Soc.* **1988**, *110* (5), 1596-1597.
279. Breslow, R.; Maitra, U.; Rideout, D., Selective diels-alder reactions in aqueous solutions and suspensions. *Tet. Lett.* **1983**, *24* (18), 1901-1904.
280. Rideout, D. C.; Breslow, R., Hydrophobic acceleration of Diels-Alder reactions. *J. Am. Chem. Soc.* **1980**, *102* (26), 7816-7817.
281. Breslow, R.; Maitra, U., On the origin of product selectivity in aqueous diels-alder reactions. *Tet. Lett.* **1984**, *25* (12), 1239-1240.
282. Dai, P.; Zhang, C.; Welborn, M.; Shepherd, J. J.; Zhu, T.; Van Voorhis, T.; Pentelute, B. L., Salt Effect Accelerates Site-Selective Cysteine Bioconjugation. *ACS Cent. Sci.* **2016**, *2* (9), 637-646.
283. Dziedzic, P.; Bartoszewicz, A.; Córdova, A., Inorganic ammonium salts as catalysts for direct aldol reactions in the presence of water. *Tet. Lett.* **2009**, *50* (52), 7242-7245.
284. Oregioni, A.; Stieglitz, B.; Kelly, G.; Rittinger, K.; Frenkiel, T., Determination of the pKa of the N-terminal amino group of ubiquitin by NMR. *Scientific Reports* **2017,** *7* (1), 43748.
285. Dirksen, A.; Hackeng, T. M.; Dawson, P. E., Nucleophilic catalysis of oxime ligation. *Angew. Chem. Int. Ed.* **2006**, *45* (45), 7581-4.
286. Layer, R. W., The Chemistry of Imines. *Chemical Reviews* **1963,** *63* (5), 489-510.
287. Cordes, E. H.; Jencks, W. P., Nucleophilic Catalysis of Semicarbazone Formation by Anilines. *J. Am. Chem. Soc.* **1962**, *84* (5), 826-831.
288. Sayer, J. M.; Jencks, W. P., Imine-forming elimination reactions. 2. Imbalance of charge distribution in the transition state for carbinolamine dehydration. *J. Am. Chem. Soc.* **1977**, *99* (2), 464-474.
289. Varki, A., Biological roles of glycans. *Glycobiology* **2017,** *27* (1), 3-49.
290. Tams, J. W.; Vind, J.; Welinder, K. G., Adapting protein solubility by glycosylation.: N-Glycosylation mutants of Coprinus cinereus peroxidase in salt and organic solutions. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1999**, *1432* (2), 214-221.
291. Gu, J.; Isaji, T.; Xu, Q.; Kariya, Y.; Gu, W.; Fukuda, T.; Du, Y., Potential roles of N-glycosylation in cell adhesion. *Glycoconj. J.* **2012**, *29* (8), 599-607.
292. Rudd, P. M.; Merry, A. H.; Dwek, R. A. In *Roles for Glycosylation in Receptor-Ligand Interactions in the Immune System*, Animal Cell Technology Meets Genomics, Dordrecht, 2005//; Gòdia, F.; Fussenegger, M., Eds. Springer Netherlands: Dordrecht, 2005; pp 31-42.
293. Xu, C.; Ng, D. T. W., Glycosylation-directed quality control of protein folding. *Nature Reviews Mol. Cell Biol.* **2015**, *16* (12), 742-752.
294. Wang, Y.-C.; Peterson, S. E.; Loring, J. F., Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Res.* **2014**, *24* (2), 143-160.
295. Zhang, J.; Xing, C.; Tiwari, B.; Chi, Y. R., Catalytic Activation of Carbohydrates as Formaldehyde Equivalents for Stetter Reaction with Enones. *J. Am. Chem. Soc.* **2013**, *135* (22), 8113-8116.
296. Ni, C.; Hu, J., The unique fluorine effects in organic reactions: recent facts and insights into fluoroalkylations. *Chem. Soc. Rev.* **2016**, *45* (20), 5441-5454.
297. Bollhagen, R.; Schmiedberger, M.; Barlos, K.; Grell, E., A new reagent for the cleavage of fully protected peptides synthesised on 2-chlorotrityl chloride resin. *J. Chem. Soc. Chem. Comm.* **1994**, (22), 2559-2560.
298. Bogdanowicz-Szwed, K.; Pałasz, A., Hetero-Diels-Alder reaction of 3-aryl-2-benzoyl-2-propenenitriles with enol ethers. Synthesis of 2-alkoxy-3,4-dihydro-2H-pyran-5-carbonitriles. *Monatshefte für Chemie / Chemical Monthly* **1997**, *128* (11), 1157-1172.