

University College London

**Novel approaches for cysteine
bioconjugation**

By

Ramiz Iqbal Nathani

Submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Declaration

I, Ramiz Iqbal Nathani, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Ramiz Iqbal Nathani

January, 2014.

Abstract

This thesis describes and investigates novel strategies for cysteine modification to achieve protein bioconjugation. Chapter 1 provides an introduction to the research project with an overview of protein modification techniques. Chapter 2 describes the development of a site selective dual labelling strategy based on substrate controlled cysteine modification. The application of this strategy to green fluorescent protein is also detailed. Chapter 3 describes the development and evaluation of thiophosphonium as a platform for protein modification. An in-depth discussion on reaction mechanisms, pathways, stability of thiophosphonium and the utility of this platform is also included. Chapter 4 describes the development of bromomaleimide based reversible cysteine modification, with particular focus on development of improved strategies for conversion of thiomaleimides back to free cysteines. The utility of the approach was demonstrated using a proof of concept experiment.

Contents

Declaration	i
Abstract	ii
Contents	iii
Acknowledgments	iv
Abbreviations	v
Chapter 1 Introduction	1
1.1 Chemical modification of proteins	1
1.2 Protein modification approaches	2
1.2.1 Natural amino acid modification	3
1.2.2 Chemical Modifications of Cysteine	5
1.2.3 Chemical modification of protein termini	12
1.2.4 Insertion of non-natural amino acids	18
1.3 Dual modification of proteins	22
1.3.1 Dual labelling through functionalisation of C- and N-terminus	23
1.3.2 Labelling protein fragments	24
1.3.3 Using non-natural amino acids	26
1.4 Aims	28
Chapter 2 Site-selective homogeneous dual labelling of proteins	30
2.1 Introduction	30
2.2 Aims	33
2.3 Model System	35
2.4 Reaction with <i>N</i> -methylmaleimide	37
2.5 Reactions with 2,5-dibromohexanediamide	39
2.6 Sulfonium stability	44
2.7 Reactivity of dehydroalanine and sulfonium	45
2.8 Dual modification	49
2.8.1 Dual cysteine mutants	49
2.8.2 Reaction of dual mutants with <i>N</i> -methylmaleimide	50
2.8.3 Reaction of Dual Mutants with 2,5-dibromohexanediamide	52
2.8.4 Dual functionalisation of dual modified GFP	55
2.9 Conclusion	60
Chapter 3 Cysteine modification to thiophosphonium	62
3.1 Introduction	62
3.2 Aims	65
3.3 Reaction of disulfides with tris(dimethylamino)phosphine	68
3.3.1 Small molecule model system	68

3.4	Protein model system	74
3.5	Thiophosphonium stability	78
3.6	Mechanistic insights	78
3.6.1	Thiophosphonium reactivity with nucleophiles	83
3.7	Conclusion	84
Chapter 4 Reversible cysteine modification using bromomaleimides		86
4.1	Introduction	86
4.2	Aims	87
4.3	Model system	91
4.3.1	Small molecule model system	91
4.3.2	Protein model system	92
4.4	Reversibility of thiomaleimides	93
4.4.1	Thiomaleimide reversibility with nucleophiles	93
4.4.2	Cycloaddition	98
4.4.3	Tether reagents	100
4.5	Conclusion	106
Conclusions and Further Work		107
Experimental		110
General Experimental		110
Protein Mass Spectroscopy		110
Nuclear Magnetic Resonance (NMR) Spectroscopy		111
Chemicals and Solvents		112
Experimental for Chapter 2		113
Experimental for Chapter 3		154
Experimental for Chapter 4		162
References		171

Acknowledgments

First I would like to thank my supervisor Prof. Steve Caddick for giving me the opportunity to carry out this exciting research, and for his enthusiasm, motivation, guidance and support throughout my PhD. Second I would like to thank Dr. James Baker, Dr. Richard Fitzmaurice and Dr. Mark Smith for their invaluable help, discussions and advice.

I would also like to thank past and present members of the Caddick group for making my time at UCL as fun as it was; in particular Chris, Vijay, Antoine, Rachel, Paul Eifion, Ahmed, Emily, Lourdes and Brian. I would also like to thank my colleagues and other group members for making my stay enjoyable - especially João, Cristina, Elena and Bhavesh. I would also like to thank Dr. Frank King for all the excellent help and advice he gave me.

A big thank you goes to Dr. Abel Aliev for his continued interest and help with all NMR related issues and to Dr. Lisa Harris for her invaluable Mass Spec support.

Finally I would like to thank my family for their support throughout my studies.

Abbreviations

Ala	Alanine
BSA	Bovine Serum Albumin
CHES	<i>N</i> -Cyclohexyl-2-aminoethanesulfonic acid
CI	Chemical Ionisation
CTPR3	<i>Consensus</i> Tetratrico Peptide Repeat
DARPin	Designed Ankyrin Repeat Proteins
DHA	Dehydroalanine
DMF	Dimethylformamide
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDT	1,2-Ethanedithiol
EPL	Expressed Protein Ligation
EI	Electron Ionisation
FRET	Förster Resonance Energy Transfer
GFP	Green Fluorescent Protein
HIV	Human Immunodeficiency Virus
LCMS	Liquid Chromatography - Mass Spectrometry

LRMS	Low Resolution Mass Spectrometry
HRMS	High Resolution Mass Spectrometry
MES	2-(<i>N</i> -Morpholino)ethanesulfonic acid
mRNA	Messenger Ribonucleic Acid
MS	Mass Spectrometry
NBS	<i>N</i> -Bromosuccinimide
NCL	Native Chemical Ligation
NMR	Nuclear Magnetic Resonance
MSH	<i>O</i> -mesitylenesulfonylhydroxylamine
PEG	Polyethylene Glycol
PLP	Pyridoxal-5-phosphate
PSGL-1	P-Selectin-glycoprotein-ligand-1
SPAAC	Strain-promoted azide-alkyne activated click chemistry
SQD	Single Quad Detector
SrtA	Staphylococcus Aureus Sortase
TCEP	Tris(2-carboxyethyl)phosphine
TEV	Tobacco Etch Virus
TFA	Trifluoroacetic Acid
TLC	Thin layer Chromatography
tRNA	Transfer Ribonucleic Acid
UCL	University College London
uPLC	Ultra Performance Liquid Chromatography

UV

Ultraviolet

Chapter 1 Introduction

1.1 Chemical modification of proteins

Decoding the chemistry of life is the ultimate goal for biochemists and chemical biologists. The enormous complexity of chemical and biochemical pathways represents a major obstacle towards achieving this aim. A significant step in this direction was the decoding of the entire human genome sequence, which provided a vast trove of genetic information.^{1,2} Whilst the human genome consists of 25000-30000 genes,¹ the human proteome is much more complex, consisting of more than 1 million proteins.³ It is essential, therefore, to understand the functions of encoded proteins to gain a thorough understanding of the chemistry of life.⁴ While different mRNA transcripts can be generated from a single gene, the true complexity of the human proteome is generated following translation by covalent processing events that change the properties of a protein by proteolytic cleavage or by addition of a modifying group to one or more amino acids.⁵ These chemical changes, referred to as 'Post-Translational Modifications' (PTMs), often fundamentally alter the innate protein function,⁶ e.g. phosphorylation of serine causes conformational switches to (de)activate protein-protein interactions,⁷ ubiquitination generally marks proteins for degradation by 26S proteasome,⁸ and acetylation of lysine regulates DNA-protein interaction for histones (*Figure 1*).⁹

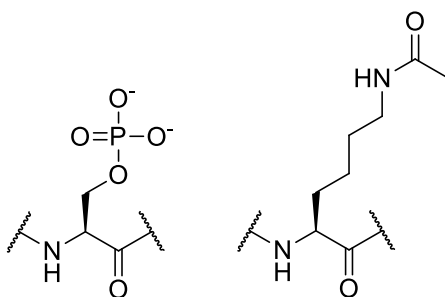


Figure 1 : Phosphorylated serine and N-Acetylated lysine

Consequently, understanding PTMs is hugely significant as it provides invaluable insight into cellular functions and processes. The major challenge in this area is to gain access to post-translationally modified proteins.

Advances in protein modification methods have led to the development of a range of protein bioconjugation technologies and their application in a range of fields such as the imaging of proteins by fluorescent tagging to track proteins *in vitro* and *in vivo*¹⁰ and the development of therapeutic protein conjugates^{11,12} to treat indications such as HIV,¹³ cancer,¹⁴ and malaria.¹⁵ Chemically modified proteins also find applications in diagnostics.¹⁶

Despite these advances, synthetic methodology to access PTMs by developing synthetic mimics is highly desired.^{17,18} The use of synthetic methodology to modify proteins faces some major challenges, the most significant of which is the need for high selectivity and precision, for example, discrimination against the 20 natural amino acids is essential for selective modification.¹⁹ Moreover, these modifications must be carried out in aqueous media, near physiological pH and at ambient temperature to prevent denaturing of proteins. In addition, these reactions should be unaffected by surfactants that are often required for protein stability.²⁰ Despite these challenges, over the years a large number of approaches to protein modification have been developed.^{21,22} Whilst some of these approaches rely on novel reactions designed to work with high specificity, other approaches have relied on expanding the genetic code and introducing unnatural amino acids for site selective modification.²³

This chapter provides an introduction to the various approaches available for protein modification and also outlines the challenges currently being faced. Brief introductory sections are provided at the start of subsequent chapters to place the work described into context.

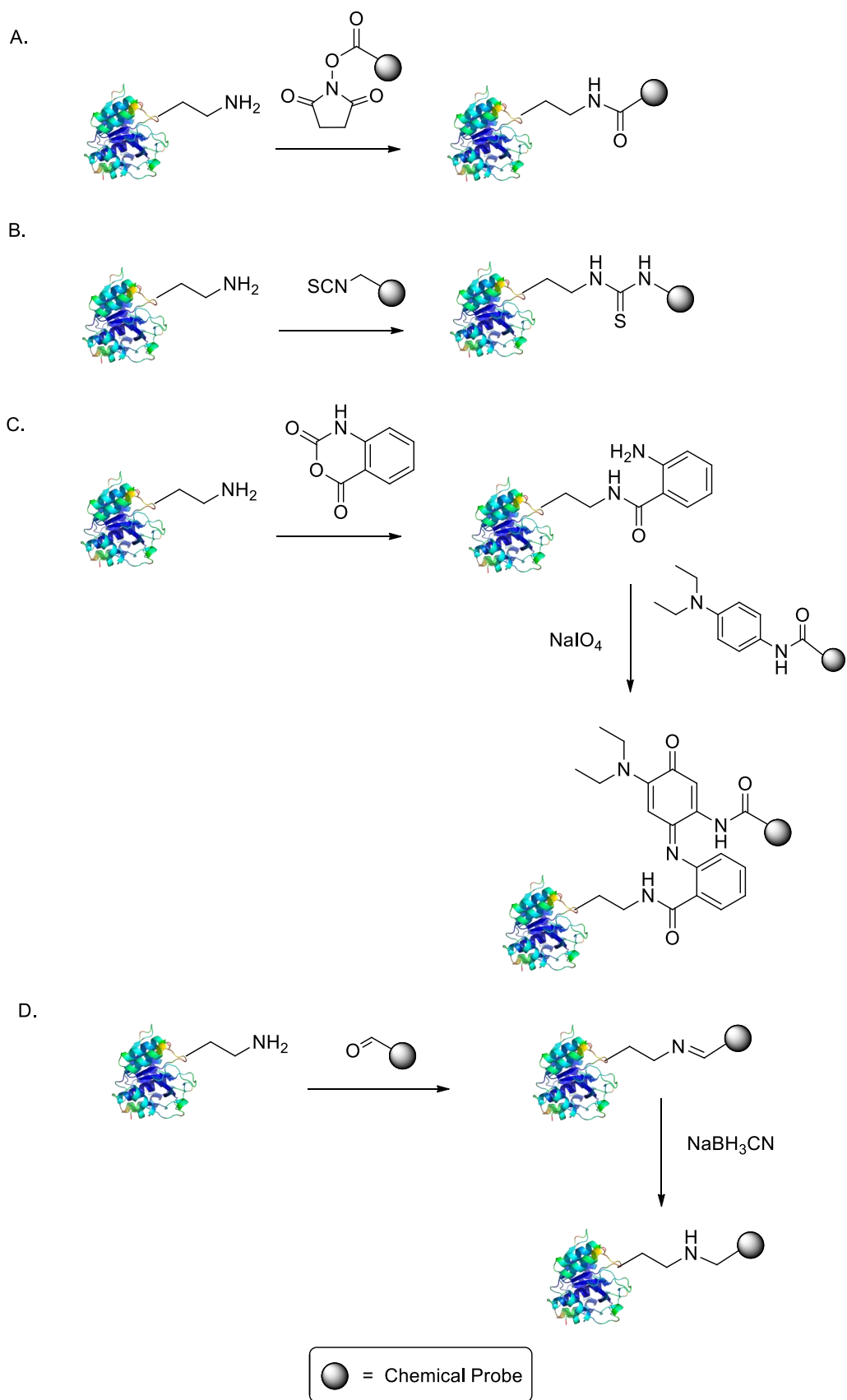
1.2 Protein modification approaches

A plethora of methods for protein modification have been reported over the last decade. These approaches can be divided into two major categories: modification of natural amino acids and modification of non-natural amino acids. Each of these methods have their own unique advantages and disadvantages and are described in further detail below.

1.2.1 Natural amino acid modification

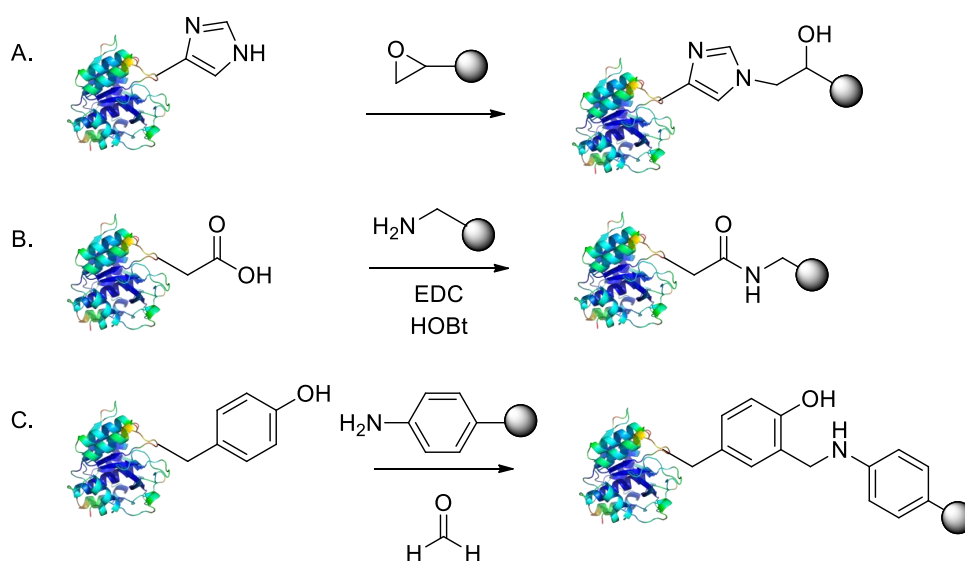
Out of the 20 natural amino acid residues, chemical modification methods have been reported for lysine, cysteine, tyrosine, glutamic acid, aspartic acid, tryptophan, histidine and methionine.²⁴ Lysine and cysteine, the most nucleophilic residues of the naturally occurring amino acids at neutral pH, are the most widely studied and form the major targets for protein modification.

Due to the large natural abundance of lysine, it is often targeted when site selective modification is not essential. Over 30 years ago, Fairbanks *et al.* reported that the reaction of lysine with *N*-hydroxysuccinimide (NHS) esters, at physiological pH and in the absence of any exogeneous reagents, forms irreversible amide linkages (*Scheme 1A*).²⁵ At higher pH (9-9.5), lysine residues react quantitatively with isothiocyanates forming thioureas (*Scheme 1B*).²⁶ Francis *et al.* reported the use of *H*-benzo[d][1,3]oxazine-2,4-dione to functionalize lysines forming orthoaminebenzamide, which can be functionalized further with NaIO₄ and dialkyl acyl phenylenediamine (*Scheme 1C*).²⁷ Glycoconjugates of bovine serum albumin (BSA) have been reported using aldehyde-based reagents to form imines, which can be reduced using sodium borohydride to form an irreversible linkage (*Scheme 1D*).²⁸ Recently, Cal *et al.* developed a reversible lysine modification based on iminoboronates.²⁹ These methods demonstrate the wide range of protein modification methods available for lysine modification.



Scheme 1. Chemical modifications of lysine

Methods to functionalize other amino acid residues have also been developed. Epoxides react with various nucleophiles leading to random surface modification, but have been tuned to react selectively with histidine residues (*Scheme 2A*).³⁰ The carboxylic acid groups present on glutamate and aspartate residues, and the C-terminus can react under peptide coupling conditions using a water soluble carbodiimide and HOBt at neutral pH, to form an amide bond, (*Scheme 2B*).³¹ A highly selective modification of tyrosine has been achieved by a Mannich based multi-component reaction under mild conditions at pH 5.5-6.5 (*Scheme 2C*).³² Phenol addition to diazonium salts, a widely used reaction in the dye industry, has also been used to selectively functionalise the phenol group on tyrosine. Reduction of the azo bond by sodium hydrosulfite to aniline allows further functionalisation with a substituted acrylamide²⁷ and this approach was used to functionalize external surface of tobacco mosaic virus capsid.³³ Preformed imines have also been shown to react efficiently with tyrosine.³⁴



Scheme 2. Chemical modifications of A) Histidine B) Aspartic acid C) Tyrosine

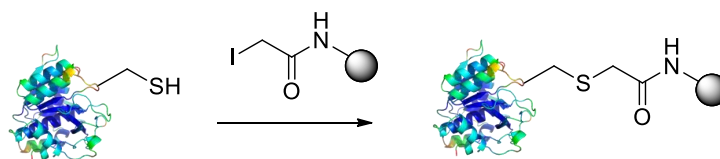
1.2.2 Chemical Modifications of Cysteine

Cysteine is the most nucleophilic residue of the naturally-occurring amino acids and has been widely targeted for protein modification. Owing to the low pK_a of the thiol side chain, cysteines react rapidly with electrophiles, even at low pH. The low natural abundance of free cysteine residues make them very attractive targets for site-selective protein modification.³⁵ By introducing cysteine residues at the position of

interest using site directed mutagenesis,³⁶ protein modification can be carried out with ease using cysteine reactive reagents.

1.2.2.1 α -Halocarbonyls

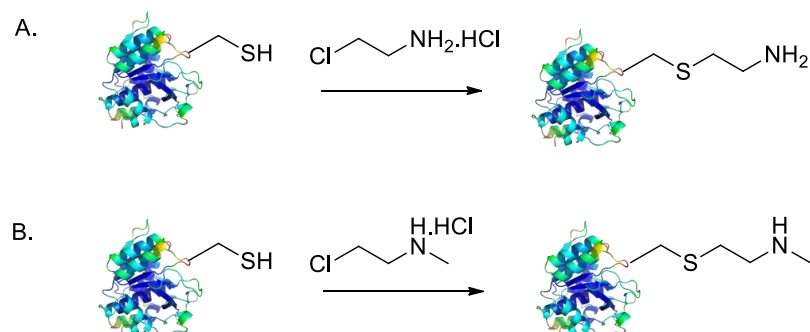
One of the oldest and most commonly used approaches for cysteine modification is through alkylation with α -halocarbonyls (*Scheme 3*), such as iodoacetamides³⁷ and chloroacetamides.³⁸ Although α -halocarbonyls suffer from poor chemoselectivity, *i.e.* they often additionally react with lysine residues, they have been used in a wide range of applications where a high degree of site-selectivity is not required. For example, Brocchini *et al.* used iodoacetamides to covalently attach flavin to the active site of papain, thereby functionally converting a cysteine protease to an oxidoreductase.³⁹ Flitsch *et al.* used carbohydrate tethered iodoacetamides to mimic asparagine-linked glycoproteins.⁴⁰ Homogeneously glycosylated samples of human erythropoietin⁴¹ and dihydrofolate reductase⁴² have been accessed using iodoacetamide-linked carbohydrates.



Scheme 3. Cysteine labelling with iodoacetamide

1.2.2.2 Aminoethylation

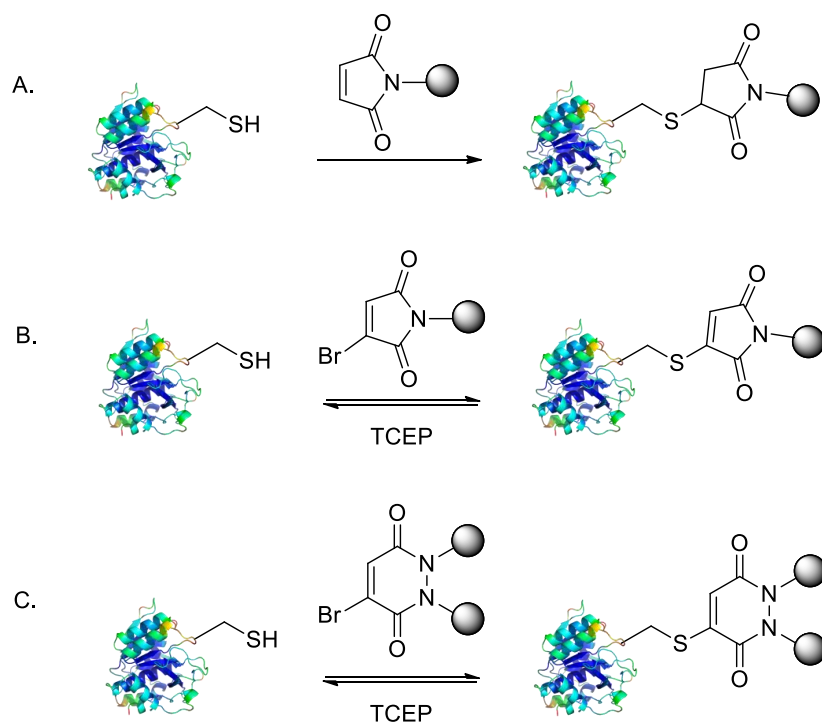
Aminoethylation of cysteine was first described over 50 years ago.⁴³ The resulting moiety acts as a mimic for lysine, with the positive charge being recognised by trypsin as a site for proteolytic action (*Scheme 4*). Methylated lysine analogues afforded by this method have been used to study post-translational modifications of histones, which control the epigenetic status of cells.⁴⁴



*Scheme 4. Cysteine modification by aminoethylation to form A) Lysine mimic
B) Methylated lysine mimic*

1.2.2.3 Michael acceptors

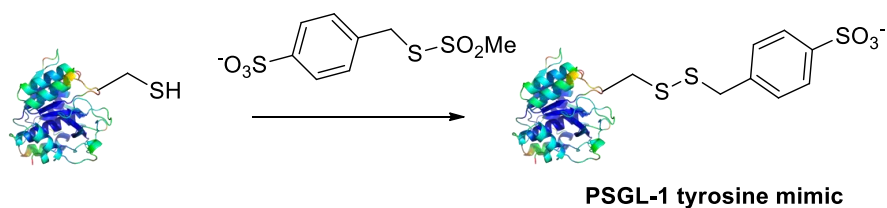
A reliable way to modify cysteine selectively is *via* conjugate addition to Michael acceptors at neutral pH. A number of α,β -unsaturated systems have been investigated and developed as selective cysteine modification reagents. Maleimides⁴⁵ and vinyl sulfones⁴⁶ represent traditional reagents for cysteine modification (*Scheme 5A*). More recently, bromomaleimide and related derivatives have been developed that allow reversible cysteine modification.⁴⁷ Bromomaleimides react as rapidly as conventional maleimides, but revert to free cysteines in the presence of reducing agents such as 2-mercaptoethanol or TCEP (*Scheme 5B*).⁴⁸ Novel pyridazinedione-based reagents have also been developed. These reagents offer the potential for three points of attachment and are completely hydrolytic stable unlike conventional maleimide reagents (*Scheme 5*).⁴⁹ Pyridazinedione-based reagents have also been shown to be highly selective for cysteine over lysine and the resultant bioconjugates are reversible in an excess of reducing reagent.⁵³



*Scheme 5. Cysteine modifications with A) Maleimide B) Bromomaleimide
C) Pyridazinedione*

1.2.2.4. Disulfides

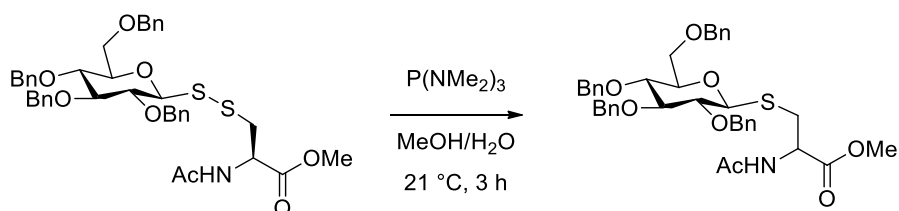
Cysteines react rapidly and selectively with disulfides to undergo disulfide exchange, with the thermodynamic equilibrium governing the final reaction mixture. 5,5'-dithiobis-(2-nitrobenzoic acid), often termed as Ellman's reagent, is widely used to measure protein thiol content.⁵⁰ Thiols react stoichiometrically with Ellman's reagent, resulting in the release of 2-nitro-5-thiobenzoate which exists as a dianion in alkaline pH. The dianion concentration can then be estimated using a spectrophotometer to indirectly measure protein thiol content. More broadly, reagents such as methanethiosulfonates allow efficient modification of cysteines to form desired disulfides.^{51,52} Synthetic modification of a sulfotyrosine-mimic of P-selectin-glycoprotein-ligand-1 (PSGL-1), was carried out using disulfide modification of a cysteine using a methanethiosulfonate reagent (*Scheme 6*). The modified protein displayed strong binding to P-selectin and functioned as an effective mimic.⁵³ Disulfides are versatile covalent linkages, however, they are not stable under reducing conditions and revert back to free cysteine. This property has been used to great effect in the design of a novel drug delivery approach.⁵⁴



Scheme 6. Methanethiosulfonate modification of cysteine to form tyrosine mimic

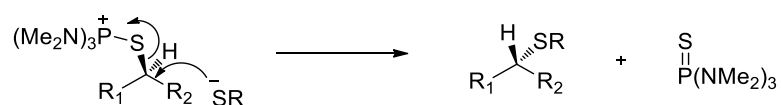
1.2.2.5. Thioethers

Thioether, unlike disulfide, is a robust linkage that is stable under reducing conditions. Davis *et al.* recently reported the use of tris(dimethylamino)phosphine for the conversion of a disulfide linkage to a thioether on a protein.⁵⁵ The conversion of a glycosyl-linked cysteine to a thioether under aqueous conditions was also reported.⁵⁵ The thioether product was obtained as a mixture of stereoisomers with the loss of stereochemistry at the cysteine. This was explained by formation of a dehydroalanine intermediate obtained by the β -elimination of thiophosphonium intermediate, which underwent conjugate addition of thiol to then form thioether (Scheme 7).



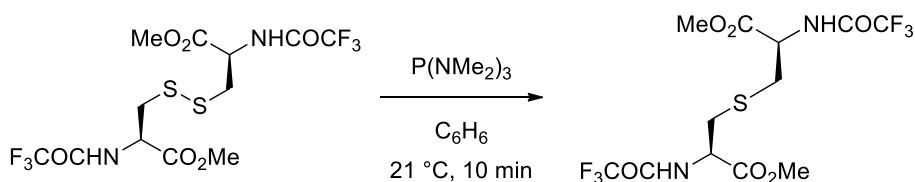
Scheme 7. Disulfide to thioether modification

Under non-aqueous conditions, disulfide substrates without an acidic α -proton have been proposed to proceed *via* the thiophosphonium intermediate.⁵⁶ The reactions proceeded with inversion of configuration at the carbon α to the disulfide linkage indicating the mechanism to be S_N2 (Scheme 8) with the loss of hexamethylthiophosphoramide.



Scheme 8. Proposed mechanism of substitution

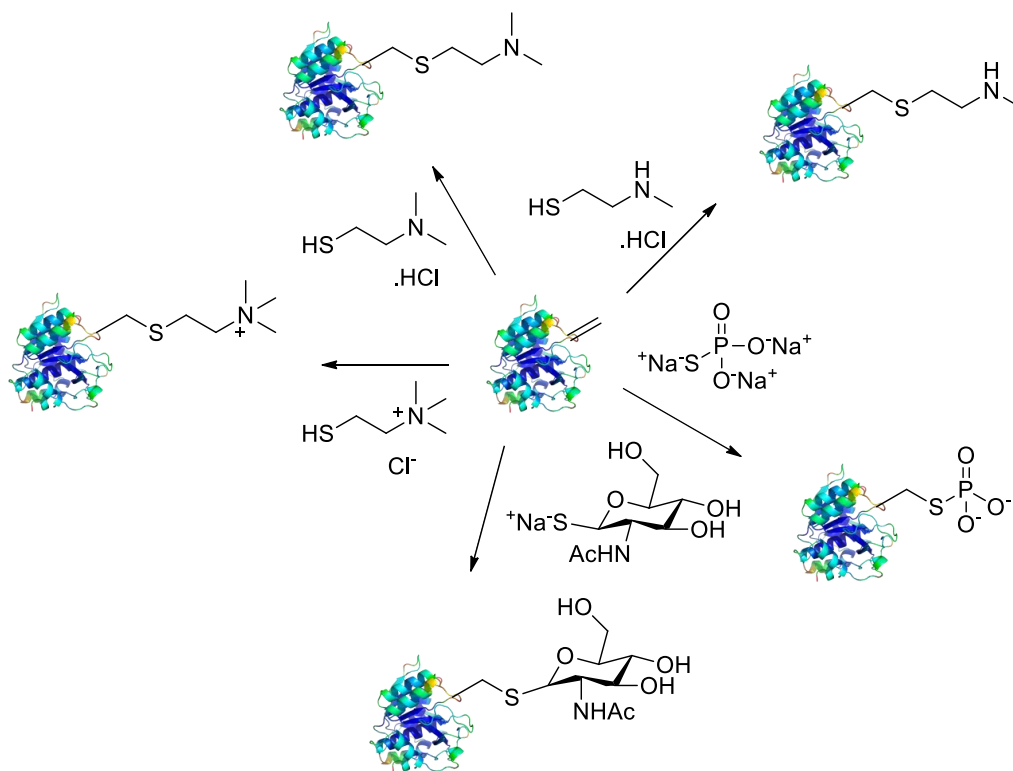
Interestingly, Harpp *et al.* demonstrated conversion of cystine to lanthionine with retention of stereochemistry (*Scheme 9*) in non-aqueous media with a stoichiometric amount of tris(dimethylamino)phosphine, indicating that presence of a base is essential for formation of dehydroalanine, even for substrates with an acidic α -proton.⁵⁶ Desulfurization reactions have also been reported on a number of trisulfide substrates.^{57,58}



*Scheme 9. Cystine derivative to lanthionine derivative using tris(dimethylamino)phosphine*⁵⁶

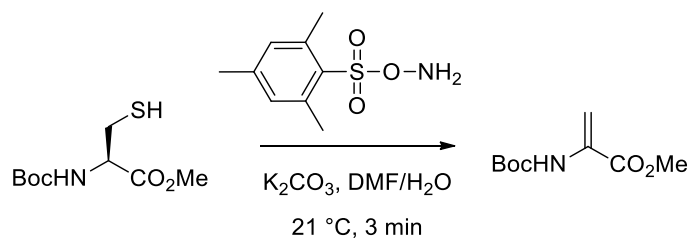
1.2.2.6. Oxidative elimination of cysteine : Dehydroalanine

A distinct approach for cysteine modification is represented by the oxidative elimination of cysteine to yield dehydroalanine (Dha). Dha is seldom observed in nature; rare examples include tyrosine conversion to Dha in thyroglobulin⁵⁹ and the conversion of serine to Dha in lantibiotics, such as nisin.⁶⁰ Chemically, Dha is an excellent electrophile that undergoes efficient conjugate addition with thiols, and thus serves as a versatile handle for protein modification. The reactivity profile of Dha allows access to a wide range of protein modifications that mimic some of the most prevalent PTMs including phosphorylation,⁶¹ glycosylation,^{62,63} methylation⁶⁴ and acetylation.⁶⁵



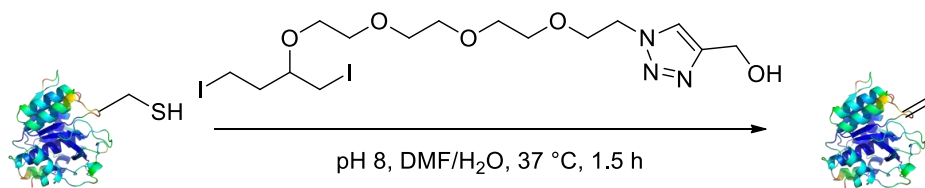
Scheme 10. Modification of dehydroalanine with thiols

Lawton *et al.* reported the first approach to access Dha from cysteine using a double alkylation approach; conjugate addition followed by intramolecular annulation affording a sulfonium intermediate which undergoes β -elimination to yield Dha.⁶⁶ Davis *et al.* has reported a number of approaches for conversion of cysteine to Dha.⁶⁷ A single cysteine mutant of subtilisin *Bacillus Lentus* S156C reacted rapidly with *O*-mesitylenesulfonylhydroxylamine (MSH) to form dehydroalanine under mildly basic conditions (Scheme 11).⁶⁷ While MSH also reacts with methionine, the reaction is reversible in the presence of DTT. Unfortunately, MSH is extremely difficult to handle, with a number of publications reporting an explosive hazard associated with pure and crystalline MSH.⁶⁸



Scheme 11. Conversion of cysteine to dehydroalanine⁷¹

Mild and selective bisalkylation approaches that afford efficient conversion of cysteine residues to Dha have also been reported. Kajihara *et al.* reported that diethyl meso-2,5-dibromoadipate reacts with cysteine on an 11-residue peptide under mildly acidic conditions to give Dha.⁶⁹ A water soluble version of 1,4-diiodobutane (*Scheme 12*) is able to affect the conversion of cysteine to Dha on a camelid single domain antibody scaffold.⁷⁰ This transformation has been observed in murine and human metabolism of anti-cancer drug busulfan and 1,4-dihalobutanes.^{71,72,73}



Scheme 12. Conversion of cysteine to dehydroalanine using diiodobutane derivative.

While Dha modification is a powerful methodology, a significant drawback of this approach is the formation of epimeric products. However, this does not necessarily preclude its applicability. In a study of acetylated and methylated lysine mimics, constructed *via* a Dha intermediate, both epimers were seen to function as enzyme substrates without loss of activity.⁷⁴

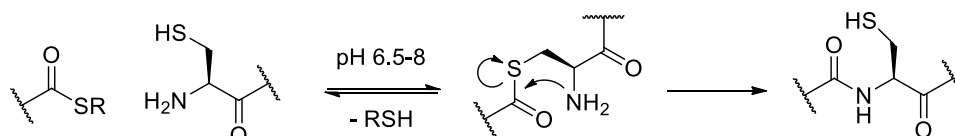
1.2.3 Chemical modification of protein termini

Terminal protein residues afford additional chemical modes of reactivity that have also been exploited to develop site-specific protein modification methods. A number of strategies, both chemical and enzymatically catalysed, revolve around the unique position and reactivity of terminal amino acid residues and will be discussed in detail below.

1.2.3.1 Native Chemical Ligation

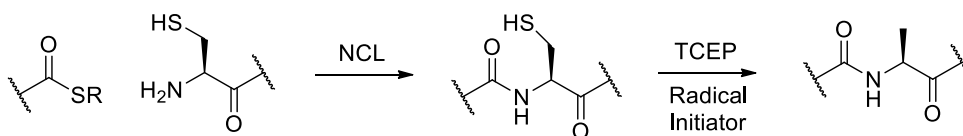
A number of protein ligation strategies have been developed in recent years.⁷⁵ Wieland *et al.* discovered that the reaction of valine-thioester and cysteine led to the formation of valine-cysteine dipeptide.⁷⁶ This reaction laid the foundation for the development of Native Chemical Ligation (NCL).⁷⁷ NCL is a powerful approach that allows for the chemoselective linkage of two peptide or protein fragments. An

N-terminal cysteine and a C-terminal thioester react to form a native peptide linkage under aqueous conditions without the need for any protecting groups (*Scheme 13*). It allows for the installation of protein modifications *via* the ligation of chemically synthesized peptide fragments at the protein termini. Post-translational mimics that have been synthesized by chemical modification methods, although analogous, often differ chemically from natural post-translational modifications. NCL allows site-specific installation of native post-translational modifications with retention of stereochemistry.



Scheme 13. Native Chemical Ligation (NCL)

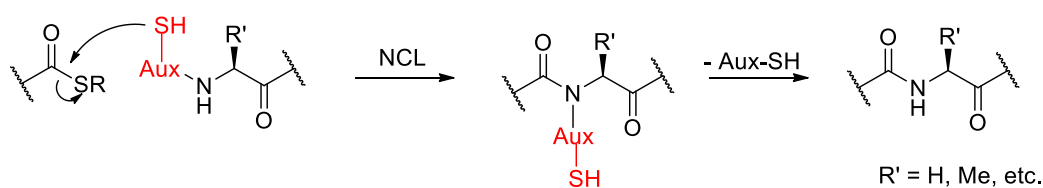
Given the final product contains a cysteine at the ligation position; a major drawback of this approach is the low natural abundance of free cysteine residues in native proteins. Therefore, to broaden the approach, strategies to convert genetically engineered cysteine residue, introduced specifically for NCL, to other naturally occurring amino acids have been developed. Dawson *et al.* reported hydrogenolytic desulfurization of cysteine to afford alanine with complete retention of stereochemistry to synthesize the first cysteine free peptide using an NCL approach.⁷⁸ Danishefsky *et al.* developed a mild, radical based cysteine-selective desulfurization. The reduction is mediated by tris(2-carboxyethyl)phosphine (TCEP).⁷⁹ This approach has been used to convert cysteine to both alanine and valine (*Scheme 14*).^{80,81}



Scheme 14. NCL at alanine using TCEP mediated desulfurization

A major advance was achieved in late 90's with the development of thiol auxiliaries (*Scheme 15*).^{82,83,84,85} These thiol auxiliaries, attached to N-terminus of a synthetic

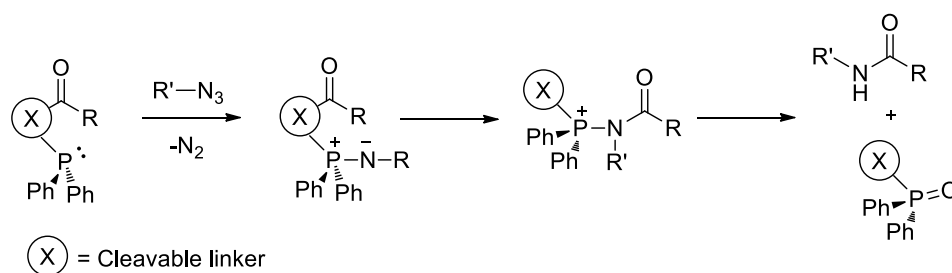
peptide, can facilitate ligation and then be removed by using trifluoroacetic acid (TFA).



Scheme 15. NCL using removable auxiliary

It should be appreciated that the chemical synthesis of peptide thioesters is challenging and represented another major obstacle to the applicability of this approach. To overcome this problem, Muir *et al.* developed ‘Expressed Protein Ligation’ (EPL) as an extension to NCL.^{86,87,88} This technology utilizes inteins to access recombinant polypeptide C-terminal thioesters. Inteins are peptide sequences that mediate self-splicing when inserted within a larger precursor polypeptide. They are eliminated from the precursor sequence as they facilitate conjugation of the N- and C-terminal regions into a new protein. EPL then proceeds through reaction with a chemically synthesized or expressed protein or peptide that contains an N-terminal cysteine residue. The synthesis of proteins which are very large in size that represent a significant challenge with the traditional NCL approach, can thus easily be accessed using EPL, *e.g.* EPL has been used to prepare β' subunit of *Escherichia coli* RNA polymerase containing 1407 amino acids.⁸⁹

Though EPL and NCL are very powerful strategies, they both require cysteine residues or thiol auxiliaries to affect ligation. Strategies that do not proceed via thiol-mediated conjugation have also been developed. One such strategy is Staudinger Ligation,⁹⁰ which is essentially an extension of the Staudinger reaction.⁹¹ Bertozzi *et al.* demonstrated that a phosphine can function as an acyl group donor, forming an amide bond linkage under aqueous conditions (*Scheme 16*).⁹²



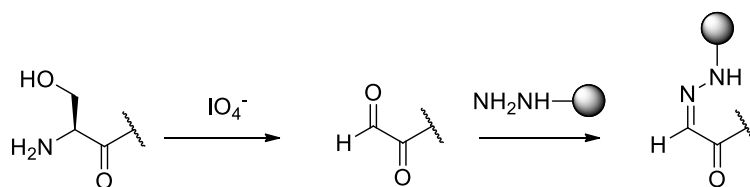
Scheme 16. Staudinger Ligation

Though protein ligation strategies offer a powerful method for the introduction of both natural and non-natural protein modifications, there are issues that make this approach less amenable than chemical strategies. Synthetic and expressed protein and peptide fragments often suffer from poor solubility and large concentrations of denaturants are often required to facilitate efficient ligation.⁹³ The polypeptides formed subsequently require careful re-folding to obtain the desired and correct final protein structures.⁹⁴

1.2.3.2 Chemical modification of protein termini forming non-native linkages

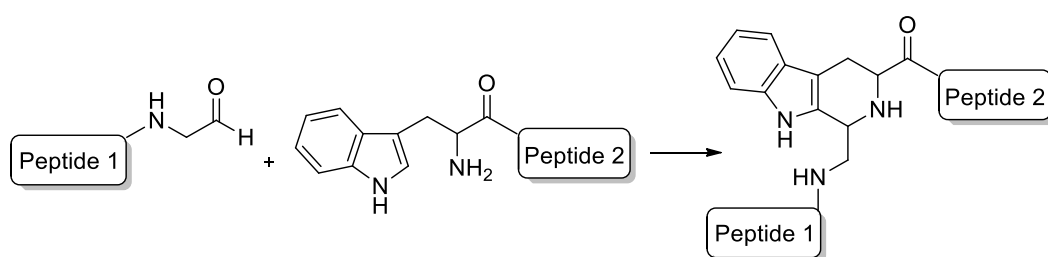
N-Terminal residues often exhibit unique chemical properties. A number of chemical and enzymatic methods have been developed to modify the protein N-terminus with high selectivity.^{95,96} While chemical methods are more general and allow for the introduction of a large number of substrates, enzymatic methods are more selective but restrictive in the size or type of functional groups that can be attached.

The lower pK_a of N-terminal amine residue has been exploited for acetylation by careful control of pH, albeit with limited success.⁹⁵ A number of strategies have been developed that target specific terminal amino acids. Terminal serine and threonine residues react rapidly with periodate under neutral conditions to form an α -N-glyoxyl residue which can be functionalized with aldehyde reactive hydrazide reagents to allow site-selective functionalisation (*Scheme 17*).^{97,98}



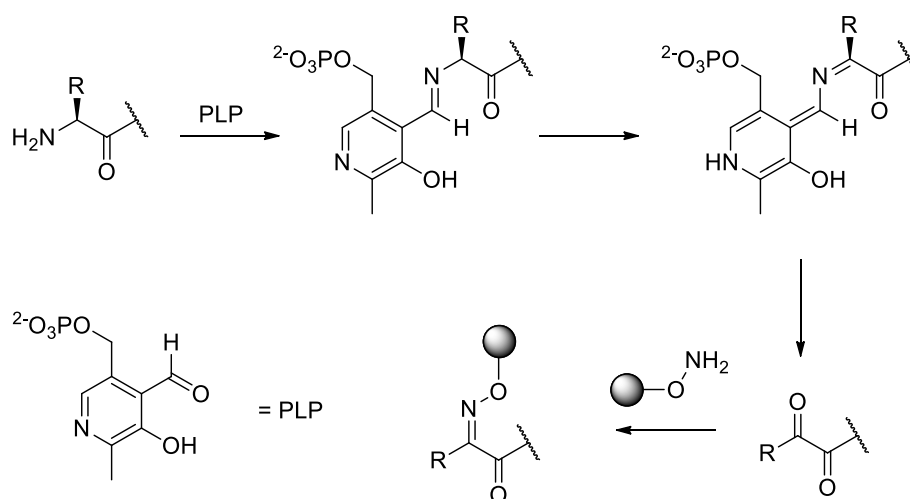
Scheme 17. Periodate modification of N-terminal serine

Also, N-terminal tryptophan residues have been shown to react with C-terminal aldehyde residues in an acid catalyzed condensation reaction, also known as Pictet Spengler reaction (*Scheme 18*).⁹⁹



Scheme 18. Pictet Spengler condensation of tryptophan

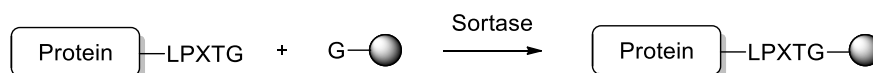
A more general chemoselective strategy that does not depend on the identity of the N-terminal amino acid chain involves oxidation of the N-terminal amino group to an imine followed by hydrolysis to afford a ketone or aldehyde.⁹⁶ However, the initial conditions that were developed for this reaction were harsh, as the reagents, glyoxylic acid, copper(II) salts, pyridine and acetic acid, are unlikely to be tolerated by most proteins. However, Francis *et al.* have reported the same conversion under much milder conditions with high selectivity using pyridoxal-5-phosphate (PLP), a cofactor that affects a variety of metabolic transformations under enzymatic control, including racemization, elimination, decarboxylation, and transamination.¹⁰⁰ Amine reacts with the aldehyde component of PLP to afford an imine; the lower pK_a of N-terminal proton then allows tautomerization to occur uniquely at this site, which, following hydrolysis, yields an aldehyde or a ketone. The resultant aldehyde or ketone can then be modified further with alkoxyamines to install new functionality (*Scheme 19*).¹⁰⁰



Scheme 19. PLP mediated N-terminal protein modification

1.2.3.3. Enzyme-mediated protein termini labelling

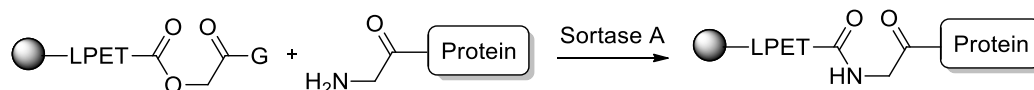
Enzymatic protein labelling methods typically employ an enzyme that selectively attaches a chemical probe to the protein terminus with the aid of a recognition sequence in an unhindered region of the protein. Sortase, a transpeptidase found in the cell envelope of many Gram-positive bacteria that anchors surface proteins to the peptidoglycan cross bridge of the cell wall,¹⁰¹ has been exploited for site-selective protein labelling. Schneewind *et al.* reported *Staphylococcus aureus* sortase (SrtA) mediated catalysis of transpeptidation by cleavage between threonine and glycine at an LPXTG recognition motif and ligation of the carboxyl group of threonine to the amino group of pentaglycine on the cell wall peptidoglycan.¹⁰² Pollock *et al.* developed a SrtA-mediated protein ligation approach allowing the highly site-specific labelling of C-terminal protein residues.¹⁰³ Sortase-mediated reactions are carried out in calcium containing aqueous buffer (pH 7.0-8.5) at nanomolar to micromolar concentration of SrtA. However, this approach requires the engineering of LPXTG into the protein sequence (*Scheme 20*).



Scheme 20. Sortase mediated C-terminal labelling

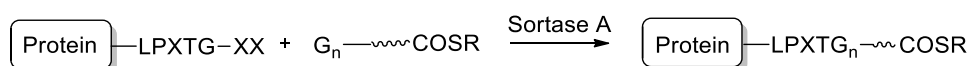
Turnbull *et al.* expanded the scope of sortase-mediated protein labelling by developing an efficient N-terminal labelling approach.¹⁰⁴ This approach uses

depsipeptide substrates for an efficient reaction and proceeds smoothly with equimolar quantities of substrates. As with other enzyme-mediated protein modification approaches, this method is also restricted by the requirement of a four amino acid LPET sequence at the N-terminus of the labelled protein (*Scheme 21*).



Scheme 21. Sortase-mediated N-terminal protein labelling

It should also be noted that Ling *et al.* reported SrtA-mediated synthesis of recombinant protein thioesters which are valuable building blocks for protein semi-synthesis using NCL (*Scheme 22*).¹⁰⁵



Scheme 22. Sortase mediated synthesis of C-terminal thioesters for ligation

1.2.4 Insertion of non-natural amino acids

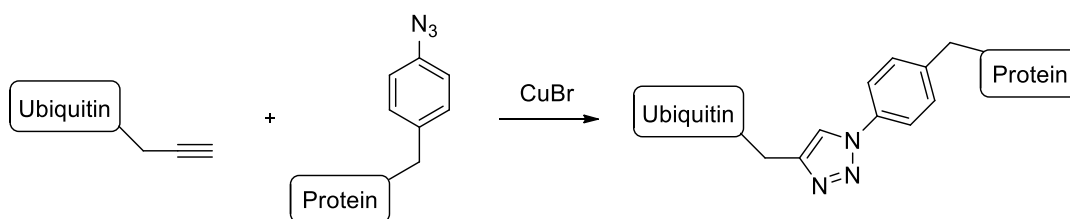
Despite the advances in chemical modification of natural amino acids, multiple copies of canonical amino acid residues in native protein structures can make the site-selective modification of proteins challenging. Chemical modifications of naturally occurring amino acid residues are often based on nucleophilicity, and selectivity delivered *via* careful control of pH and substrate stoichiometry. A fundamentally different approach for protein modification involves the genetically-controlled insertion of unnatural amino acid residues into a protein sequence. These unnatural amino acids are typically designed to present bioorthogonal side-chain functionality, hence allowing the potential for exquisite site-selective and efficient protein modification.

1.2.4.1. Use of auxotrophic strains

The most straightforward method to incorporate non-natural amino acids into recombinant proteins is to use auxotrophic strains of *E. coli*. These strains are unable

to synthesize a specific canonical amino acid and rely on uptake from the growth media. In this growth media, the specific amino acid can be replaced by a structurally similar analogue. The structurally similar analogue is then incorporated by the bacteria. This approach requires careful inducible control, as the cells are first grown in normal media and then transferred to another media containing the non-natural amino acid. Following induction, proteins with this non-natural amino acid incorporated are expressed.

Methionine, one of the rarest naturally occurring amino acid residues, is an excellent target for non-natural amino acid incorporation. Methionine has been successfully replaced with alkyne and azide analogues (*Figure 2*)^{106,107,108} that allow for subsequent protein modification with a copper-catalysed, Huisgen cycloaddition reaction (*Scheme 23*).^{109,110}



Scheme 23. A. Huisgen reaction on azidophenylalanine

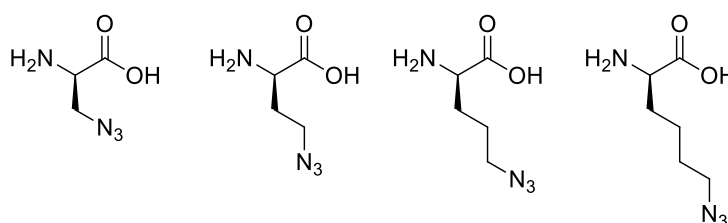


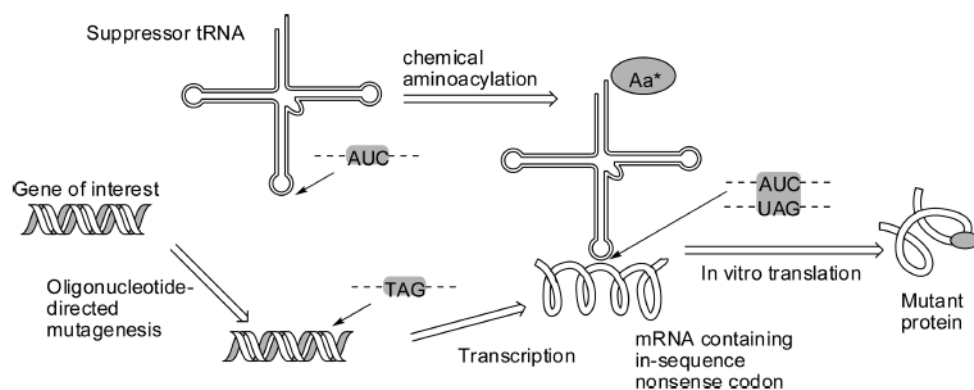
Figure 2. Methionine analogues successfully incorporated in E. coli

This approach has also been extended to label proteins *in vivo*.¹⁰⁸ Apart from methionine, other amino acids such as leucine,¹¹¹ isoleucine,¹¹² phenylalanine,¹¹³ proline,¹¹⁴ and tryptophan¹¹⁵ have been replaced by non-natural amino acids using this strategy. This method allows for the ability to replace all copies of an existing canonical amino acid with a non-natural amino acid analogue. However, it does not allow expansion of the genetic code or incorporation of additional non-natural amino acids into the protein.

1.2.4.2. Stop codon suppressor tRNA

Noren *et al.* developed an even more powerful method to expand the existing set of amino acids by translationally incorporating non-natural amino acids into proteins by expanding the genetic code. This was accomplished by ‘hijacking’ the amber stop codon (UAG) to encode for the non-natural amino acid.¹¹⁶ A stop codon is a nucleotide triplet in mRNA that usually signals the termination of translation.

tRNAs found in yeast efficiently recognise this amber stop codon (UAG), and prevent the release factor from binding to this codon. The ribosome therefore reads through the stop codon and continues protein translation. The other important components in the process of protein elongation, peptidyltransferase and elongation factor, show broad specificity.²³ Incorporation of an additional non-natural amino acid is achieved by aminoacylating the amber codon suppressor tRNA with the non-natural amino of choice (*Figure 3*).



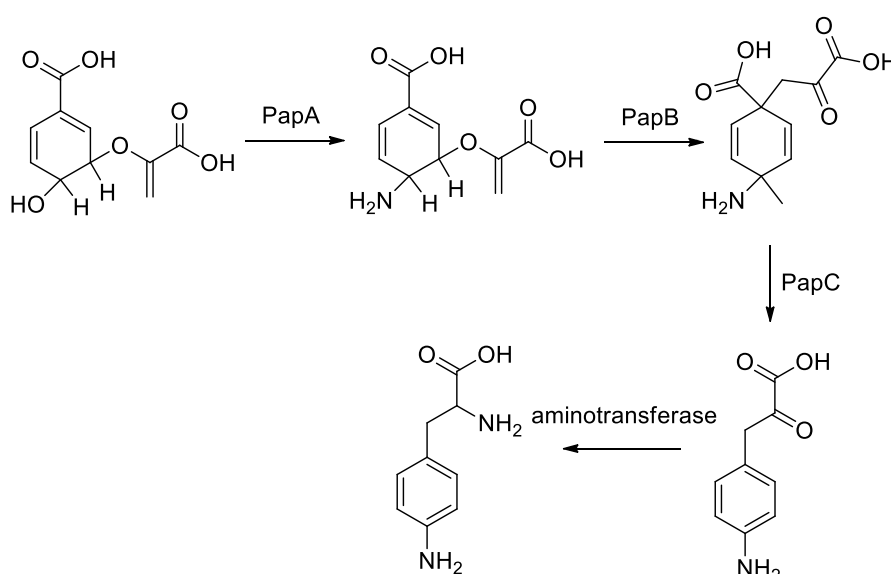
*Figure 3. Amber codon suppressor method for incorporation of non-natural amino acids (figure reproduced from Graff et al.)*²³

One significant drawback of this approach is that the aminoacylated tRNAs are too large to be taken up by the cell. Aminoacylated tRNAs therefore need to be microinjected into the host cell, and large cells such as *Xenopus Oocytes*¹¹⁷ have been used for this purpose. It should be noted that proteins have also been successfully expressed in a cell free system.¹¹⁸ However, while the overall method is very powerful, it presents significant technical challenges, especially for large scale expression, and therefore does not constitute a widespread general protein modification method.

1.2.4.3. Orthogonal aminoacyl-tRNA synthetase

To overcome the problem of microinjection, suppressor tRNA and the corresponding aminoacyl-tRNA synthetase can be produced from the host cell itself by genetic engineering.¹¹⁹ This new pair must be orthogonal to host tRNA/tRNA-synthetase pairs: the introduced tRNA should not be a substrate for any host aminoacyl-tRNA synthetase and must function efficiently, while the introduced aminoacyl-tRNA synthetase must not aminoacylate any host tRNAs and should be selective for the introduced tRNA.¹²⁰ This requirement of orthogonality is met by using tRNA-tRNA synthetase pairs from other organisms, provided the cross species aminoacylation is inefficient.^{121,122}

The non-natural amino acid to be aminoacylated, and subsequently introduced in the protein sequence, must be available in the protein cytoplasm. The non-natural amino acid must therefore be designed to be transported into the cytoplasm, and is added exogenously in the growth media to facilitate this process. Schultz *et al.* reported the biosynthesis of non-natural amino acid *in situ* by modifying the biochemical pathways of *E. coli*. This amino acid was then incorporated by an orthogonal tRNA-aminoacyl tRNA synthetase pair, eliminating the need for non-natural amino acids to be added externally (*Scheme 24*).¹²³



Scheme 24. Biosynthesis of non-canonical aminoacid in E. coli using series of phosphatidate phosphatase enzymes

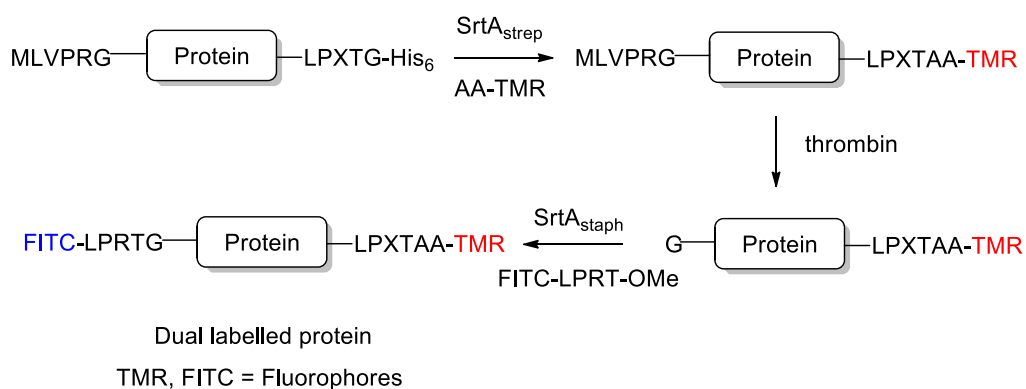
Rapid advances are currently being made in this field, with novel genetic approaches including the development of an orthogonal ribosome-mRNA pair¹²⁴ and the advent of four base codons,¹²⁵ which pave the way for improved efficiency. Despite these advances, however, a number of challenges remain. This approach often suffers from low efficiency and poor yield for two reasons: i) the aminoacylation efficiency of these newly derived tRNA synthetases is often limited and ii) there is a significant competition between the release factor and the suppressor tRNA for the amber codon.¹²⁶ Consequently, incorporation of multiple non-natural amino acids is often difficult as the efficiency decreases multiplicatively with the increase in the number of amber codons in a gene. Another problem associated with this approach is that the fidelity of incorporation of non-natural amino acids varies from system to system, and often with poor reproducibility. Practical challenges also exist, for example, growth media must be supplemented with the non-natural amino acid, often leading to a high cost of expression. Finally, the protein expression system remains technically challenging for most standard biochemistry laboratories and hence the approach remains restricted to specialist laboratories.

1.3 Dual modification of proteins

While a large array of methods are available for site-selective single protein modification, the site-selective dual-labelling of proteins has remained a challenge. Certain applications require that two functional groups be introduced into a protein in site-specific manner. Examples where the site-selective, dual-modification of proteins is desired include Förster resonance energy transfer (FRET)¹²⁷ experiments for folding studies,¹²⁸ targeted theranostic biopharmaceuticals using viral capsids^{129,130} and protein-drug therapeutics that are stabilised with conjugated half-life extension chemistry e.g. poly(ethyleneglycol) (PEG).¹³¹ Cysteine is often the preferred target for the first modification, but a second site for chemoselective functionalization may not always be available. Early methods have tried to exploit differences in cysteine accessibility to achieve site-specific, dual-cysteine labelling using reagent control. However, it is difficult to achieve homogeneous labelling using this approach.^{132,133,134,135}

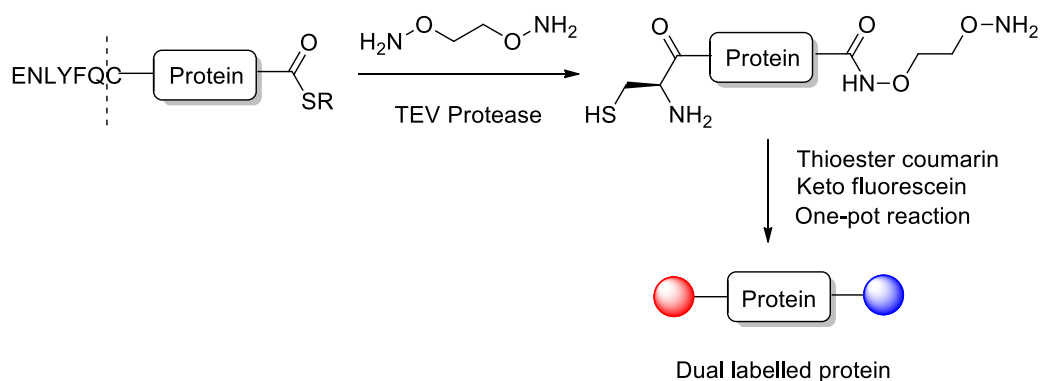
1.3.1 Dual labelling through functionalisation of C- and N-terminus

Site-selective N- and C-terminal labelling methods exploit the unique reactivity of terminal residues to introduce dual functionality. As described earlier, methods that facilitate the sortase-mediated labelling of terminal amino acid residues have been developed. Ploegh *et al.* developed a strategy for introducing different chemical labels at the termini of a protein using different sortase enzymes, SrtA_{strep} and SrtA_{staph}.¹³⁶ A fluorophore-labelled Ala-Ala conjugate served as the nucleophile for the first SrtA_{strep}-mediated protein ligation. Thrombin cleavage was then used to expose a masked glycine residue for a second SrtA_{staph}-mediated ligation to another fluorophore conjugated with the required amino acid recognition sequence (*Scheme 25*).



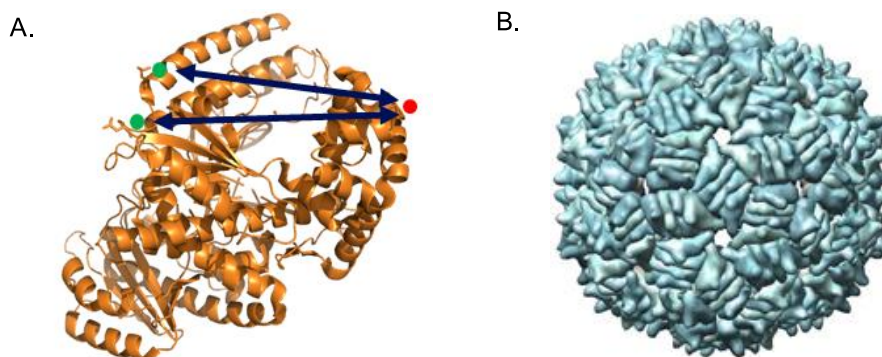
Scheme 25. Dual labelling of proteins

Wu *et al.* developed two chemoselective reactions for a one-pot protein dual labelling of Rab7D3 for protein folding studies.¹³⁷ C-Terminal labelling was accomplished by oxime ligation followed by NCL modification of an N-terminal cysteine. Intein-mediated formation of a C-terminal thioester protein was followed by reaction with a (bis)oxyamine moiety, which was subsequently modified with a keto fluorophore under mild conditions.¹³⁸ An N-terminal cysteine was then exposed by TEV protease cleavage, and chemoselectively-functionalized using NCL to attach another fluorophore for FRET studies.



Scheme 26. One pot dual labelling of proteins for single molecule FRET analysis

Single molecule FRET often requires that fluorophores be placed on a specific part of a protein that undergoes dynamic conformational changes. Santoso *et al.* studied the fingers-closing conformational transitions of DNA Polymerase I by placing one fluorophore on a mobile segment of the fingers subdomain (residue 744) and a complementary fluorophore at the base of the thumb subdomain (residue 550), enabling measurement of the inter-fluorophore distance change upon fingers-closing.¹³⁹

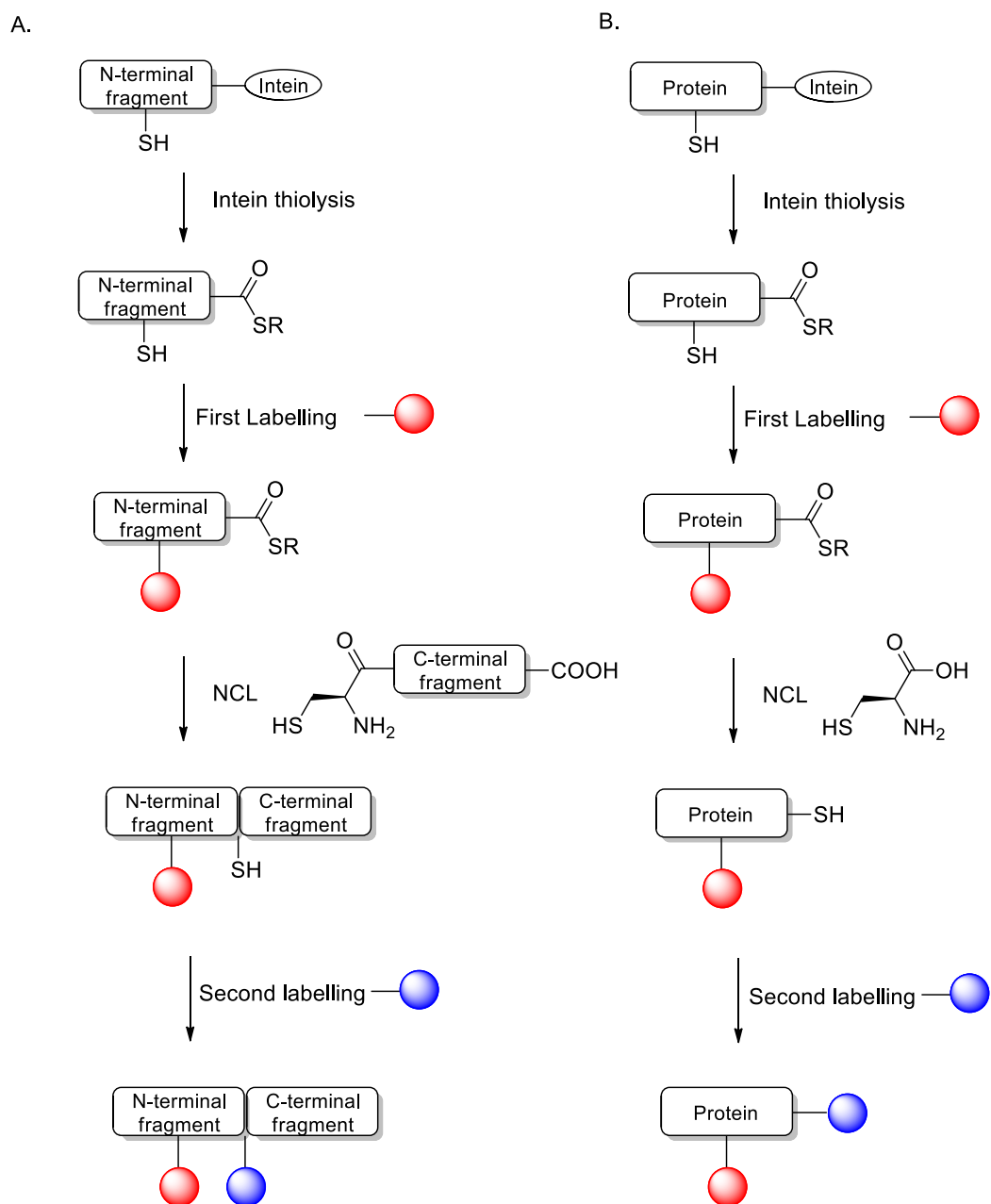


*Figure 4. A. FRET studies of DNA Polymerase – I¹³⁹ B. MS2 virus capsid for drug delivery (figure reproduced from Francis *et. al.*)¹⁴⁰*

1.3.2 Labelling protein fragments

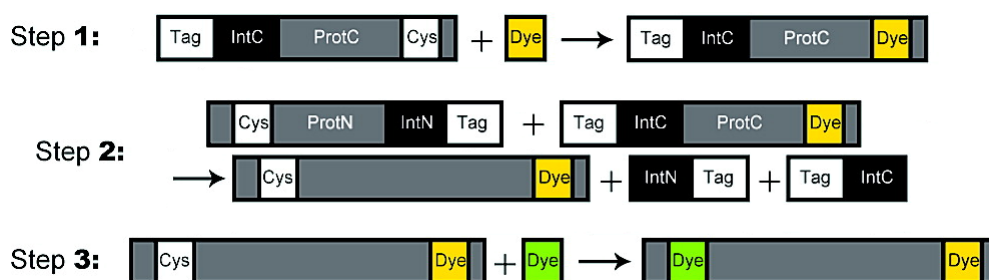
The second approach for site specific dual labelling proteins relies on generating protein fragments, and labelling the fragments individually often *via* cysteine modification, and joining the proteins together through ligation methods.^{141,142} This facilitates non-terminal, dual-modification of proteins.

Rosa *et al.* developed a homogeneous dual-labelling strategy, using an EPL approach, and used it to dual-label CTPR3 (*Consensus Tetratricopeptide Repeat*) with two fluorophores at specific positions in the protein.¹⁴² By using EPL to express either an N-terminal protein fragment or full protein, both as C-terminal thioesters, Rosa was able to differentially position the two fluorophore labels (*Scheme 27*).



*Scheme 27. Dual labelling of proteins using NCL A) Labelling of protein fragments
B) Single labelling of protein followed by Cysteine labelling at C-terminus*

Yang *et al.* reported an alternative approach to cysteine labelling of fragments to afford a dual-labelled diUbiquitin structure, thus facilitating FRET studies.¹⁴¹ The approach relied upon generating protein fragments that contained a single cysteine for fluorophore conjugation, labelling each individual cysteine, and joining the fragments to generate a full-length protein through split intein-mediated protein splicing (*Scheme 28*).¹⁴³



Scheme 28. Dual labelling via split inteins (figure reproduced from Yang et al.)¹⁴¹

Expressing proteins as fragments and labelling the individual fragments to achieve dual-labelling allows homogeneous functionalisation at non-terminal positions. The biggest limitation of this approach is represented by the stability of protein-expressed fragments. These fragments are often insoluble, representing a significant synthetic challenge.⁹³ Finally, this approach requires a large number of steps and often harsh conditions to achieve homogeneous dual labelling, which may not be amenable for proteins which exhibit poor stability and/or sensitive functional groups.⁹⁴

1.3.3 Using non-natural amino acids

A further approach that facilitates homogeneous dual labelling is the introduction of non-natural amino acids, often in conjunction with a traditional site selective method, *e.g.* cysteine modification. The ways of introducing a non-natural amino acid have been discussed in Section 1.2.4.

Pluckthun *et al.* reported a facile dual labelling approach for Designed Ankyrin Repeat Proteins (DARPsins) using a combination of non-natural amino acid and cysteine functionalisation.¹⁴⁴ Azidohomoalanine was introduced, at the N-terminus as an analogue of the initiator methionine using auxotroph techniques. Additionally, a cysteine was introduced at a second position of interest *via* site-directed mutagenesis. Strain-promoted azide-alkyne cycloaddition chemistry was utilized for

orthogonal functionalisation of azidohomoalanine, followed by maleimide modification of the single cysteine residue to obtain homogeneous dual functionalized DARPin (*Figure 5*).¹⁴⁵

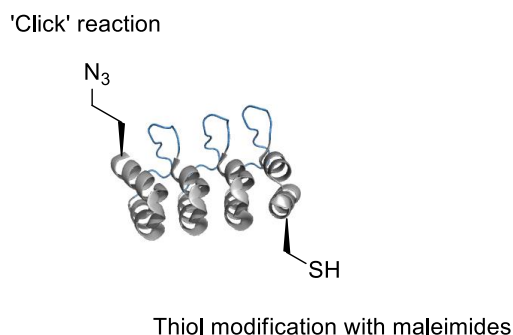


Figure 5. DARPin dual functionalisation using 'click' reaction and thiol modification

Chin *et al.* has reported the incorporation of two non-natural amino acids into a protein by synthetically evolving an orthogonal ribosome that efficiently decodes a series of quadruplet codons and the amber codon, providing several blank codons on an orthogonal mRNA, which it specifically translates. They created mutually orthogonal aminoacyl-tRNA synthetase-tRNA pairs and combined them with ribo-Q1 to affect incorporation of two non-natural amino acids into the protein sequence.¹⁴⁶ At the same time, Liu *et al.* independently developed another method to introduce two non-natural amino acids in the same protein by using two stop codons, amber codon UAG and ochre codon UAA.¹⁴⁷ This method was used to incorporate two fluorophores for FRET analysis of glutamate binding protein (QBP) and superfolder green fluorescent protein (GFP).

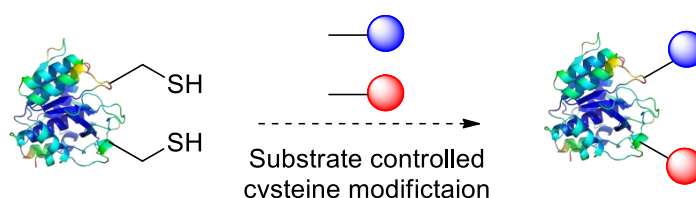
In summary, a range of techniques for protein bioconjugation have been developed over the years for a variety of applications. Each of these techniques has associated advantages and disadvantages. There is undoubtedly a clear need for development of novel bioconjugation methods that facilitate greater homogeneity and site selectivity for single and multiple protein modification with simplicity and efficiency. The development of such methods is certain to allow a better understanding of biological processes and development of improved therapeutics, diagnostics, imaging and related applications.

1.4 Aims

The following thesis aims were identified:

1. Site-selective dual labelling of proteins

Many approaches that facilitate the site-selective, dual-labelling of proteins suffer from a number of drawbacks. Modification of protein termini for dual labelling allows chemical probes to be placed only on N- and C-termini. Fragment based dual labelling methods suffer from poor solubility and stability of protein fragments. Approaches that introduce non-natural amino acids are technically restrictive for most biochemistry and chemical biology laboratories, require access to uncommon bacterial strains, the expression yields are often poor and they are expensive to perform. Thus, there is a serious lack of an accessible and general site-selective, dual modification approach. A simple, robust and effective dual-modification approach was envisaged *via* the different chemical modification of two cysteines selectively in a single protein. The potential for obtaining differential reaction modes on similarly accessible cysteine residues, using protein tertiary structure to control reaction outcome (substrate control), was to be investigated with a view of developing a general site-specific, dual-labelling strategy (*Scheme 29*).

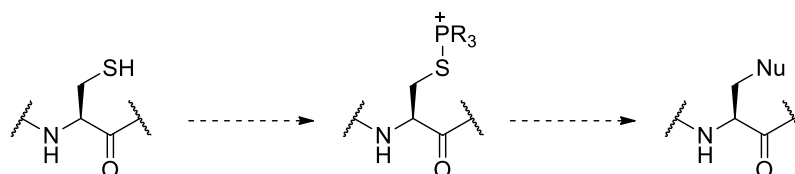


Scheme 29. Proposed site-specific dual labelling of cysteine residue via substrate control

2. Chemical methods for mimicking post-translation modifications (PTMs)

Despite recent progress in the development of techniques for accessing post-translationally modified proteins and their analogues, there remains scope for development of improved chemical modification methods. One powerful method to access PTMs is *via* dehydroalanine.^{6,9} However, products derived from the conjugate addition of a nucleophile onto Dha are likely to be a mixture of

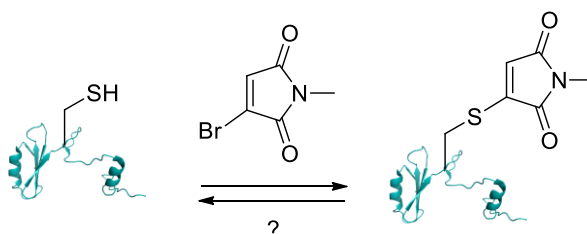
stereoisomers, due to loss of stereochemical information (Section 1.2.2).¹⁵⁵ Cysteine activation *via* thiophosphonium and subsequent nucleophilic displacement of the thiophosphonium adducts⁵⁶ was investigated as a potential method for accessing PTM analogues without loss of native stereochemistry.



Scheme 30. Proposed route for accessing PTM analogues

3. Reversible cysteine modification using bromomaleimides.

Novel bromomaleimide reagents have been developed within the Caddick and Baker groups at UCL, to enable the reversible modification of cysteine. These bromomaleimides react rapidly and selectively with cysteine residues on proteins and peptides to afford thiomaleimides, which can be converted back to free thiols in the presence of excess reducing agents such as tris(2-carboxyethyl)phosphine (TCEP) and 2-mercaptoethanol. This chemistry has considerable potential, particularly in the development of reagents that will facilitate the reversible affinity capture of proteins. The current necessity for a large excess of reducing agent to facilitate cleavage of bromomaleimide-bioconjugates could be detrimental to the structural stability of modified proteins, and therefore poses a challenge to the general applicability of the thiomaleimide reversibility approach. The development of an alternative, efficient and selective cleavage strategy, which would represent a significant advantage over the existing strategy, was explored (*Scheme 31*).



Scheme 31. Development of efficient cleavage strategy for reversible cysteine modification

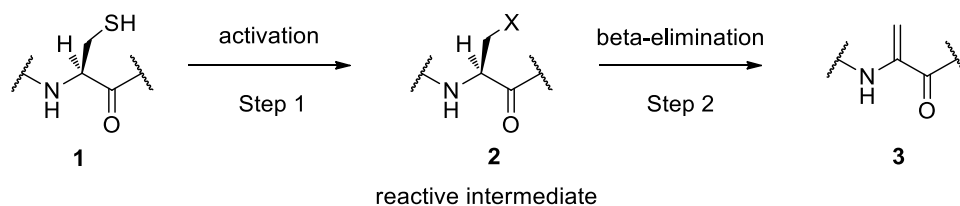
Chapter 2 Site-selective homogeneous dual labelling of proteins

2.1 Introduction

Site-selective protein labelling is widely achieved by reaction of a genetically engineered cysteine residue at a specific position in the protein of interest, with an appropriate label (Section 1.2.2).¹⁴⁸ Site-selective dual labelling of proteins represents a significant challenge due to the lack of reactive, orthogonal, non-abundant natural amino acid residues. Previous methods for site-selective dual modification have relied on introducing the second label by targeting terminal residues, introducing unnatural amino acids or labelling protein fragments (see Section 1.3).

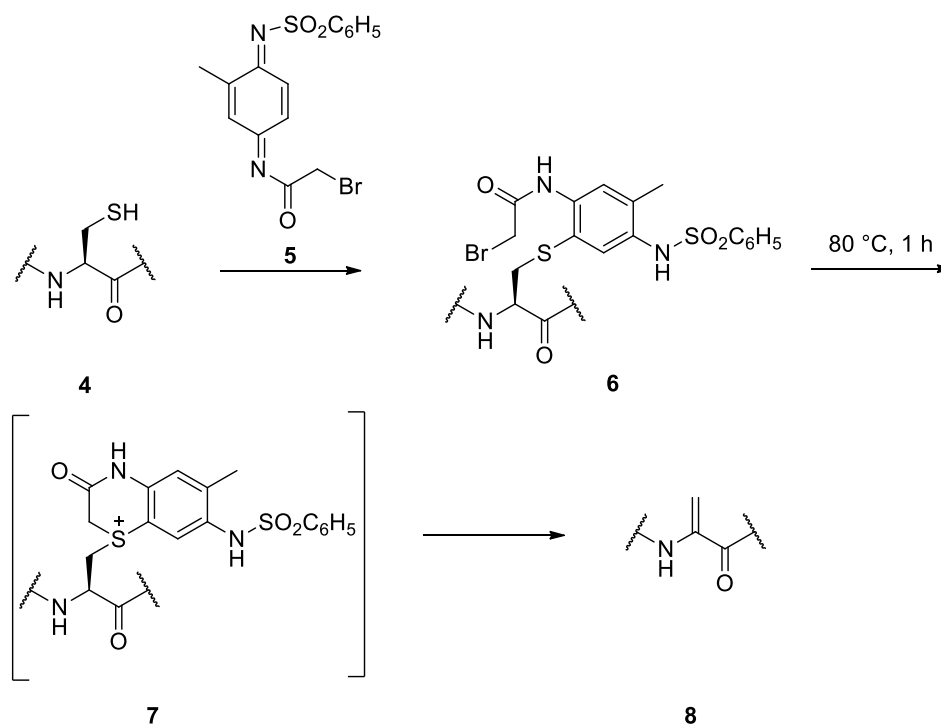
Most protein modification methods designed for natural amino acids (see Section 1.2.1) rely on the nucleophilicity of the target residue. Selectivity is achieved almost exclusively by using protein modification reagents carefully designed to exploit the differences in reactivity of the amino acid side chains under carefully controlled conditions, for example under physiological conditions, *N*-hydroxysuccinimide esters react selectively with lysines²⁵ whilst maleimides react almost exclusively with cysteines.⁴⁵

Cysteine modification is the most widely used approach for site-selective homogenous protein labelling (see Section 1.2.2). One cysteine modification of particular interest proceeds *via* the oxidative elimination of cysteine to dehydroalanine. This method can be broadly described in two steps: Step 1 involves the reaction of the cysteine **1** with an activating reagent to form intermediate **2**; Step 2 involves elimination to form dehydroalanine **3** (*Scheme 32*). Dehydroalanine can then be modified with a variety of thiol nucleophiles to introduce the desired protein modification.



Scheme 32. Conversion of cysteine to dehydroalanine

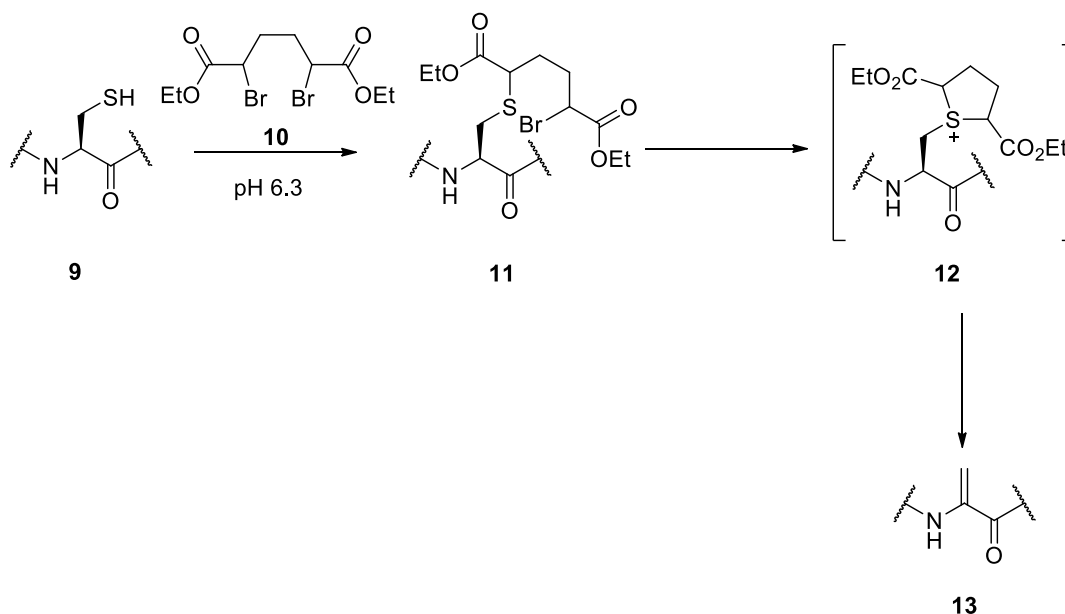
Lawton *et al.* first reported a convenient method for converting cysteine into a suitably activated intermediate *via* the bisalkylation of cysteine.⁶⁶ *N*-Acetylcysteine **4** underwent a conjugate addition reaction with quinonediimide derivative **5** to form **6**. This species then underwent intramolecular annulation to form an activated sulfonium intermediate **7** which eliminated a sulfide to afford dehydroalanine **8** (*Scheme 33*).



*Scheme 33. Conversion of cysteine **4** to dehydroalanine **8** using quinonediimide derivative **5***

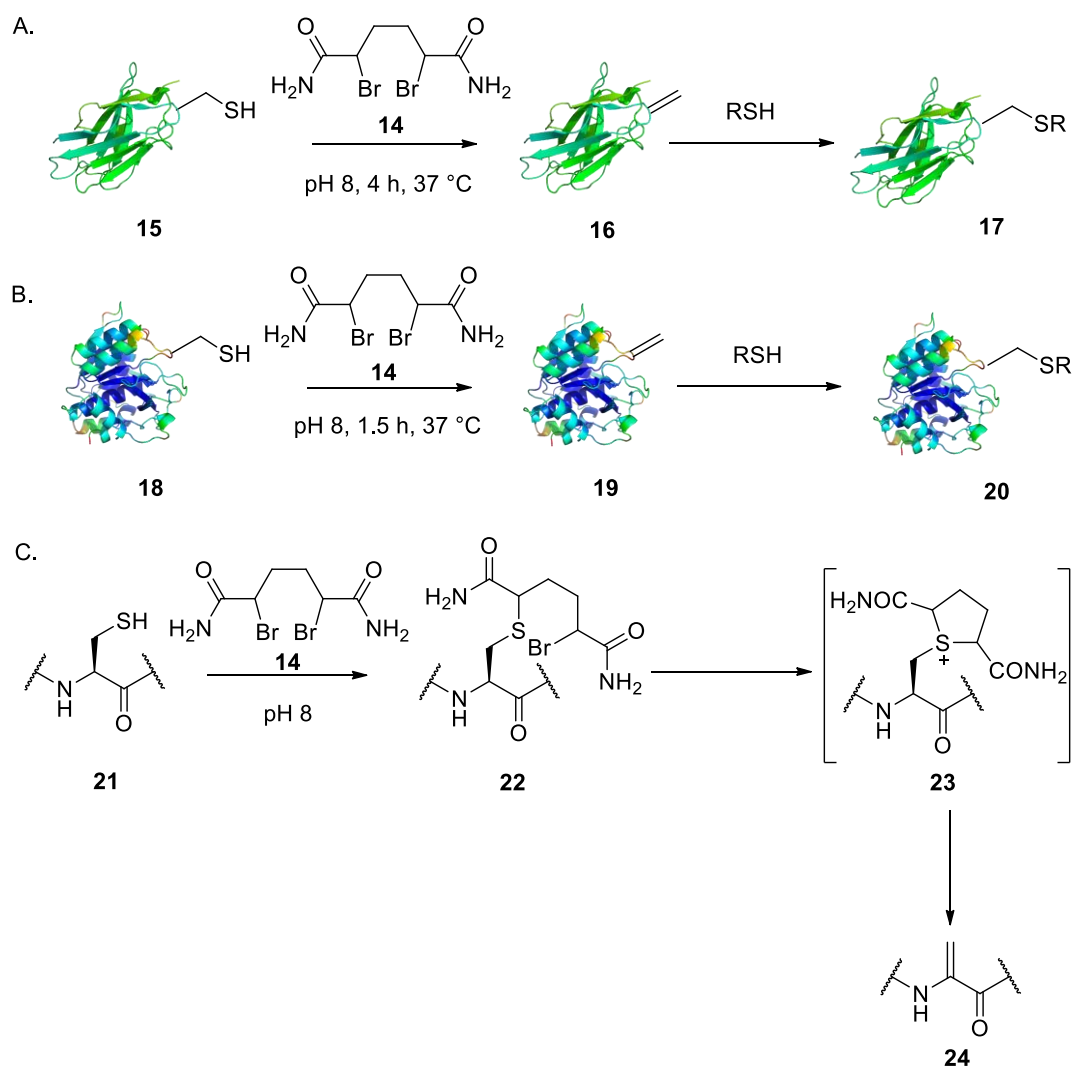
Kajihara *et al.* subsequently reported the reaction of diethyl meso-2,5-dibromoadipate **10** with a cysteine on a 10-residue peptide **9**.⁶⁹ The free thiol of the cysteine residue underwent a nucleophilic displacement reaction with **10** to initially form thioether **11**. This was followed by an intramolecular nucleophilic displacement

reaction to form an activated sulfonium intermediate **12**, which underwent rapid elimination to form dehydroalanine **13** (Scheme 34).



Scheme 34. Conversion of cysteine **9** to dehydroalanine **13** on a peptide using *meso*-2,5-dibromoadipate **10**

Davis *et al.* reported conversion of cysteine to dehydroalanine on proteins using 2,5-dibromohexanediamide **14**.⁷⁰ A cysteine mutant of a single domain camel antibody (A104C) **15** and a cysteine mutant of subtilisin SBL (S156C) **18** underwent reaction with 2,5-dibromohexanediamide **14** to form dehydroalanine derivatives **16** and **19** (pH 8, 37 °C) (Scheme 35A and Scheme 35B). The reaction pathway was observed to be analogous - after the formation of thioether **22**, the reaction is thought to proceed *via* an activated sulfonium intermediate **23** that undergoes rapid elimination to dehydroalanine **24** (Scheme 35C).



*Scheme 35. Reaction of 2,5-dibromohexanediamide **14** with A) Single cysteine mutant of single domain camel antibody **15** B) Single cysteine mutant of subtilisin **18** C) Cysteine derivative **21***

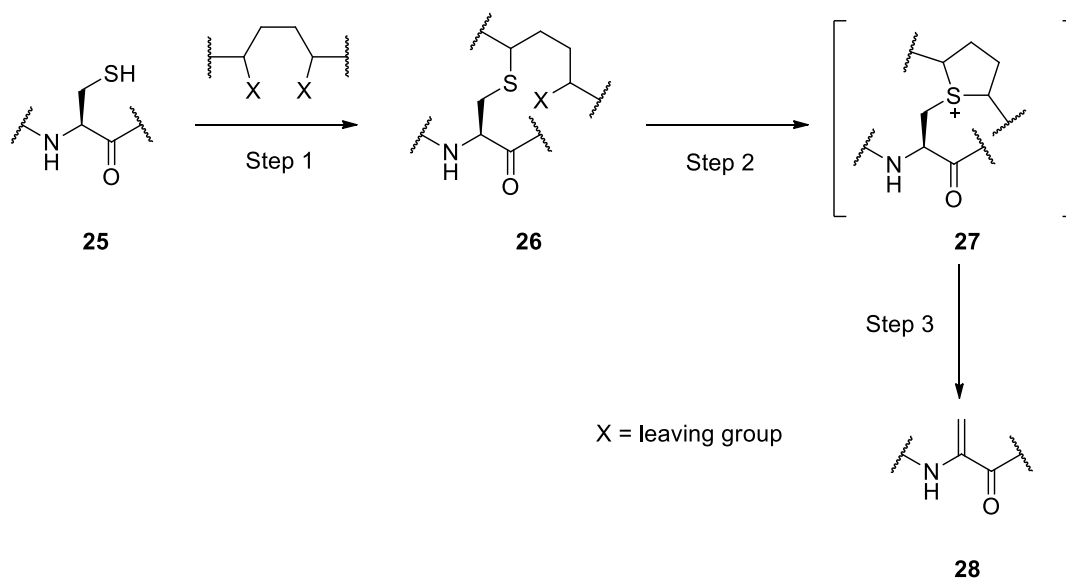
2.2 Aims

Chemical methods for protein modification rely primarily on reagent control for selective protein labelling. However, protein modification reactions *in vivo* are typically substrate-controlled.⁸ The desired modification is usually catalysed by an enzyme that recognises an amino acid sequence or element of tertiary structure on the target protein. It would therefore be of interest to know whether substrate control can be used to effect chemical protein modification and, in particular, whether it can

be used to differentiate the reactivity of two cysteine residues on a protein surface to facilitate the homogeneous, dual modification of a protein.

Oxidative elimination of cysteine to dehydroalanine is one approach that is likely to be amenable to such substrate control. The transformation of cysteine to dehydroalanine *via* bisalkylation can be broadly classified into three steps (*Scheme 36*).

1. Alkylation of cysteine **25** to form thioether **26**;
2. Conversion of thioether **26** to sulfonium **27**;
3. β - Elimination to sulfonium **27** to form dehydroalanine **28**.



Scheme 36. Modification of cysteine 25 to dehydroalanine 28 via bisalkylation

Whilst step 1 of the reaction sequence (*Scheme 36*) is likely to be reagent controlled, steps 2 and 3 may depend on the tertiary structure of the protein, e.g., steric factors may prevent formation of sulfonium at certain positions on the protein (step 2). Similarly, a solvent inaccessible α -proton in a highly structured region of a protein could prevent elimination of the sulfonium intermediate (step 3). It may, therefore, be possible to use ‘substrate-control’ to obtain different reaction outcomes for similarly accessible cysteine residues by controlling the parameters essential for steps 2 and/or 3 of the reaction.

To probe this theory, the following aims were identified:

1. Identify an appropriate protein model system and design cysteine mutants in different regions of protein tertiary structure;
2. Identify an appropriate reagent for bisalkylation of cysteine and investigate the reaction outcomes on the model protein to obtain evidence of substrate-controlled protein modification;
3. Design a strategy to exploit differential cysteine reactivity for the dual-modification of proteins; and
4. Demonstrate the utility of this substrate-controlled homogeneous dual labelling strategy.

2.3 Model System

Superfolder green fluorescent protein (GFP, PDB ID: 2B3P) **29**, was chosen as a model protein to explore bisalkylation of cysteine for substrate controlled chemical modification. Superfolder GFP has a β -barrel structure, displays excellent stability at elevated temperatures and has found wide-ranging applications especially in imaging.¹⁴⁹ It also contains two native cysteine residues, Cys-48 and Cys-70, both of which have been reported to be inaccessible for chemical modifications (*Figure 6*).¹⁴⁹

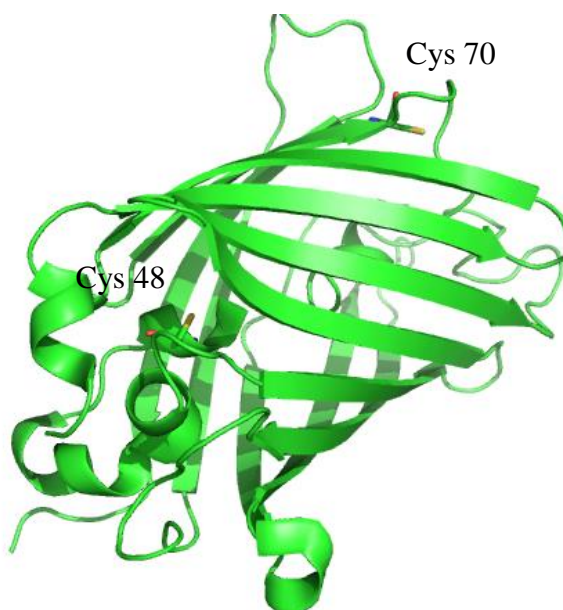


Figure 6. Superfolder Green Fluorescent Protein (Cys48, Cys70)

To explore different reaction outcomes following cysteine bisalkylation, three single cysteine GFP mutants (+CG **30**, T230C **31**, S147C **32**) were designed, cloned and expressed (*Figure 7*) (GFP (T230C) **31** and GFP (S147C) **32** were expressed by Paul Moody, PhD student, Caddick Group). Superfolder GFP **29**, in absence of any reactive cysteine residue was expressed as a control for the reactions. GFP (+CG) **30** was designed with a cysteine residue close to C-terminal end of GFP in a highly flexible region away from the β -barrel. This residue was introduced in a region with no bulky amino acid side chains or tertiary structure likely to inhibit the formation of the sulfonium intermediate (step 2, see *Scheme 36*). Moreover, the α -proton for this residue was likely to be highly solvent accessible, and hence the sulfonium intermediate is likely to undergo rapidly elimination to dehydroalanine (step 3, see *Scheme 36*). This residue was designed to have a similar reaction mode to the examples described in the literature namely SBL (S156C), Histone (K9C),⁷⁴ and the camel antibody (A104C).^{70,74}

GFP (T230C, 233 Δ) **31** was designed with a cysteine residue closer to the β -barrel with a solvent accessible α -proton of residue 230 (solvent accessible surface area = 4.3 \AA^2).¹⁵⁰ Residues around the position 230 could sterically inhibit cyclisation to the sulfonium (step 2, see *Scheme 36*) but the high α -proton accessibility was expected to facilitate elimination to dehydroalanine (step 3, see *Scheme 36*) (*Figure 7*).

GFP (S147C) **32** was designed with a cysteine residue close to the highly structured protein β -barrel to shield the α -proton of S147, thus rendering it inaccessible (solvent accessible surface area = 0.0 \AA^2).¹⁵⁰ This was hypothesized to prevent rapid elimination of sulfonium to dehydroalanine (Step 3, see *Scheme 36*) (*Figure 7*).

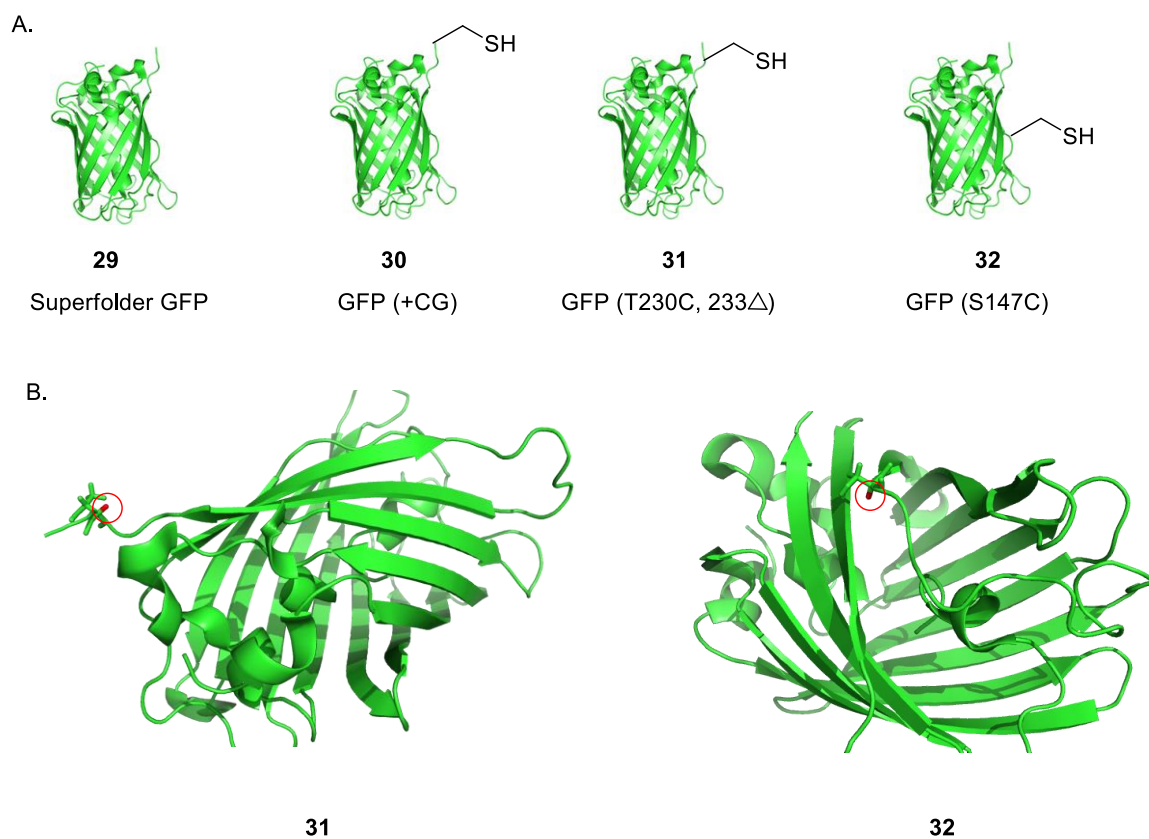


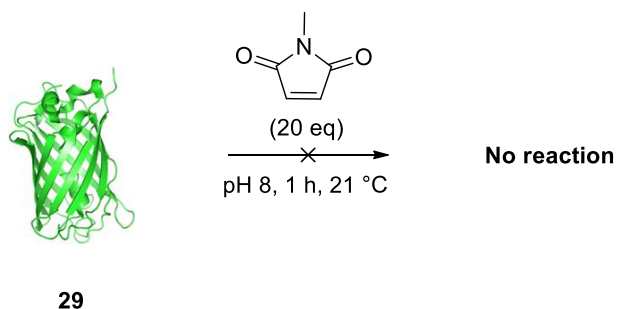
Figure 7. A) GFP mutants with cysteine residues in different environments
 B) α -proton accessibility (α -proton coloured and circled in red)

2.4 Reaction with *N*-methylmaleimide

The reactivity and the accessibility of the cysteine residues of all four GFP variants **29**, **30**, **31** and **32** was determined by treatment with *N*-methylmaleimide as a standard cysteine reactive reagent. As protein structure was unexpected to play any role in the outcome of reaction of GFP single cysteine mutants with *N*-methylmaleimide, the reaction was expected to be ‘reagent-controlled’ and hence identical reaction products were expected.

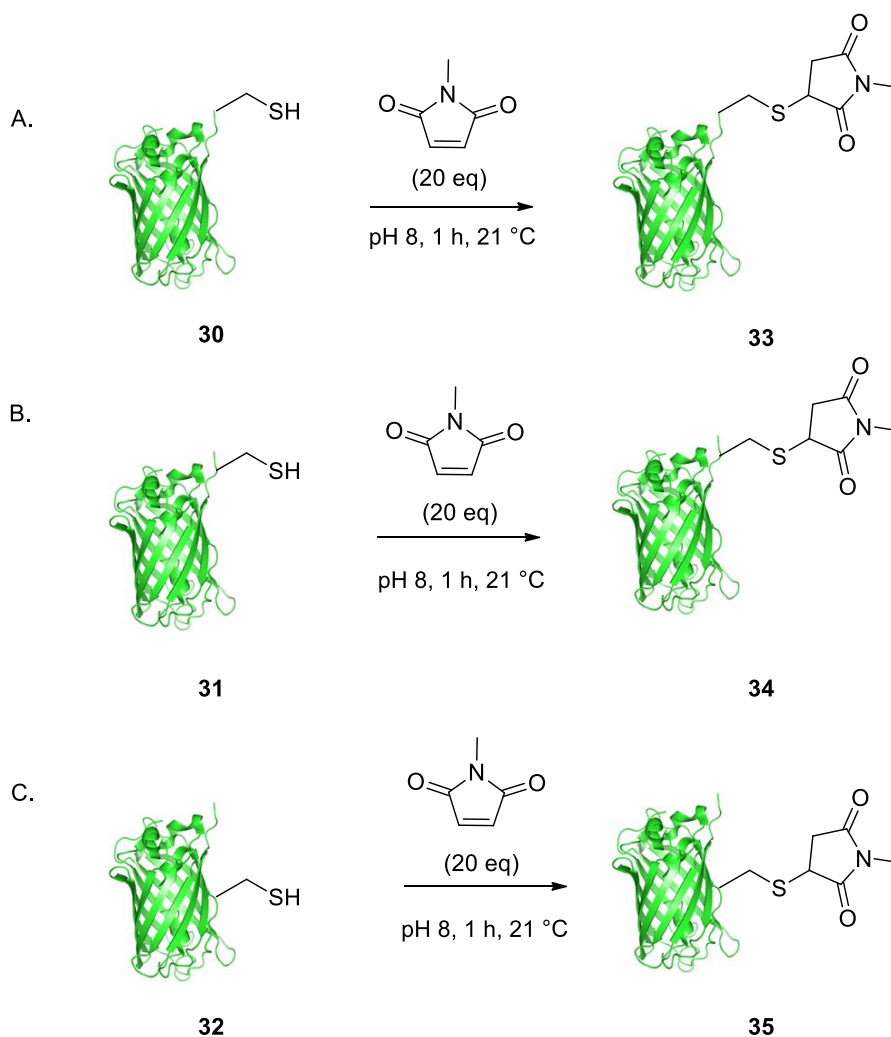
The GFP variants were subjected to a 20-fold excess of *N*-methylmaleimide at room temperature under mildly basic conditions (sodium phosphate buffer, 100 mM, pH 8) and analysed by LCMS after 1 h. GFP (superfolder) **29** served as a control and yielded no reaction with *N*-methylmaleimide, confirming the absence of maleimide

reactive residues at this concentration of maleimide relative to the protein (*Scheme 37*).



Scheme 37. Reaction of GFP 29 with N-methylmaleimide

GFP mutants GFP (+CG) **30**, GFP (T230C, 233Δ) **31** and GFP (S147C) **32** reacted smoothly and rapidly with *N*-methylmaleimide, giving complete reaction with cysteine residues forming the expected reaction products (*Scheme 38*).

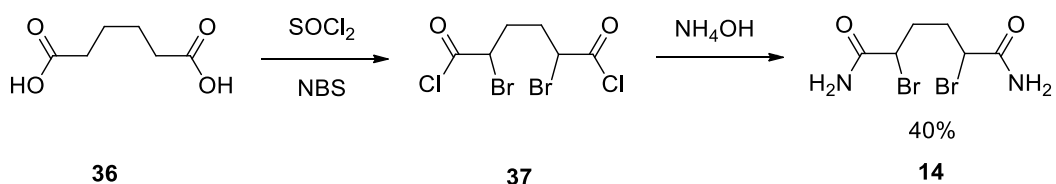


Scheme 38. Reaction of GFP 30, 31 and 32 with N-methylmaleimide

These reactions indicated that the three cysteine mutants behaved in an identical manner when subjected to ‘reagent-controlled’ modification. It can be concluded that reaction outcomes of these single cysteine mutants were likely to be identical with other ‘reagent-controlled’ cysteine modification approaches (see Section 1.2.2).

2.5 Reactions with 2,5-dibromohexanediamide

Reaction of the single cysteine mutants of GFP with a sulfonium-forming reagent was then appraised. The reagent, 2,5-dibromohexanediamide **14**, previously reported for the conversion of cysteine to dehydroalanine on a number of proteins under mild conditions,^{70,74} was chosen as the model reagent for investigating the reaction of cysteine with a sulfonium forming reagent. The reagent was synthesized from adipic acid as described in the literature in a 40% yield (*Scheme 39*).

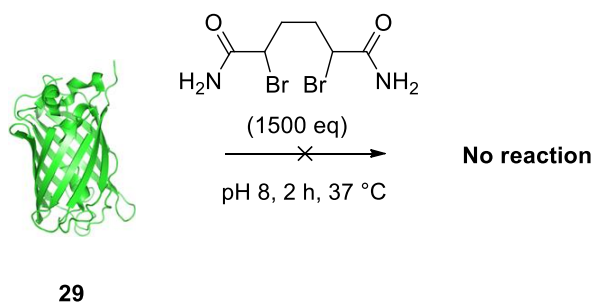


Scheme 39. Synthesis of 2,5-dibromohexanediamide 14

Prior to reacting 2,5-dibromohexanediamide **14** with any of the mutants, the GFP mutants were treated with dithiothreitol (DTT) to reduce any disulfide that may lead to incomplete reactivity. Excess reducing agent was removed by repeated diafiltration into fresh buffer (sodium phosphate, 100 mM, pH 8). Due to the poor solubility of 2,5-dibromohexanediamide **14** in water, the reagent was added as a DMF solution to give a final 10% DMF concentration in the reaction mixture.

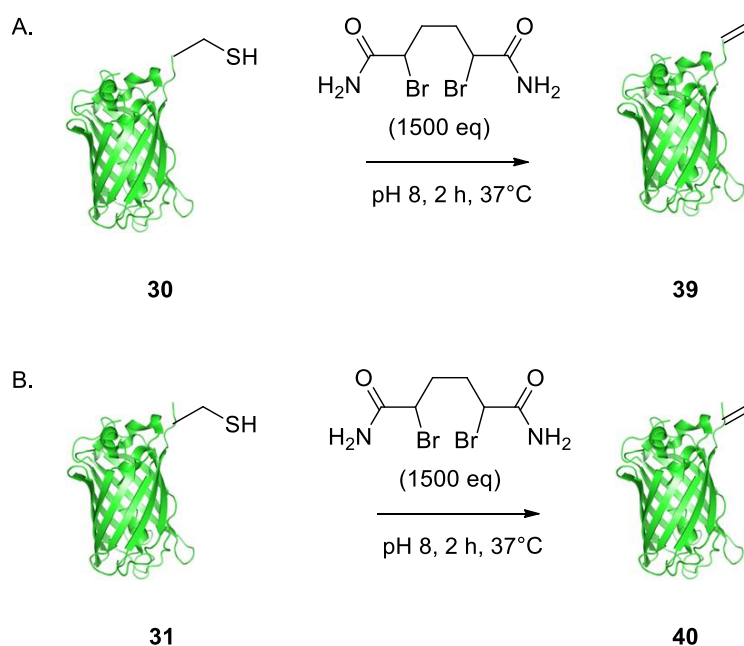
To ensure that DMF did not affect the protein or the reaction, the GFP mutants were incubated in buffer (sodium phosphate, 100 mM, pH 8) containing 10% DMF for 2 h at 37 °C in the absence of any reagent. LCMS analysis indicated that the proteins were stable after 2 h at 37 °C.

Reactions previously reported in the literature on single cysteine protein mutants with 2,5-dibromohexanediamide **14** were carried out under mildly basic conditions (sodium phosphate buffer, 100 mM, pH 8) with a large excess of the reagent (1500 equivalents) at 37 °C. Therefore, similar conditions were employed for reaction of 2,5-dibromohexanediamide **14** with each of the four GFP variants (**29,30,31** and **32**) to ensure the consistency and comparability of the data with that reported in the literature. As expected in the absence of accessible cysteine residues, incubation of superfolder GFP **29** with 2,5-dibromohexanediamide **14** (1500 equivalents, 2 h, 37 °C) showed no reaction by LCMS (*Scheme 40*).



Scheme 40. Reaction of superfolder GFP 29 with 2,5-dibromohexanediamide 14

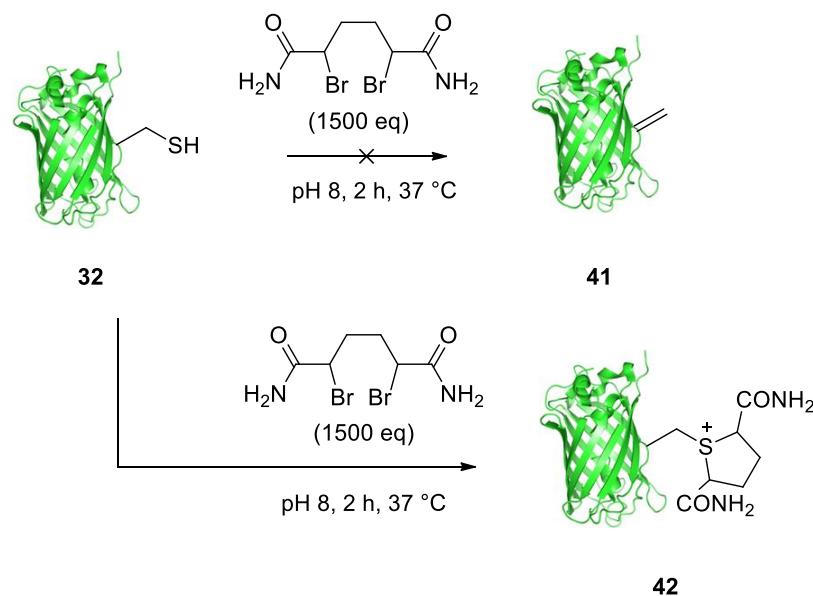
Reaction of single cysteine GFP mutants (**30**, **31**, **32**) with 2,5-dibromohexanediamide **14** provided interesting results. GFP (+CG) **30** reacted smoothly with **14** (1500 equivalents, 2 h, 37 °C) to give clean conversion to GFP (+CG, Dha) **39**. This was the expected outcome as the cysteine residue is in a highly flexible region of the protein with no crystal structure available, hence the α -proton was likely to be accessible (*Scheme 41*). The cysteine residue at position 230 on **31**, closer to β -barrel fold of GFP, also underwent a similar reaction to yield GFP (T230C, 233 Δ , Dha) **40** in complete conversion (*Scheme 41*).



Scheme 41. Reaction of GFP (+CG) 30 and GFP (T230C, 233 Δ) 31 with 2,5-dibromohexanediamide 14

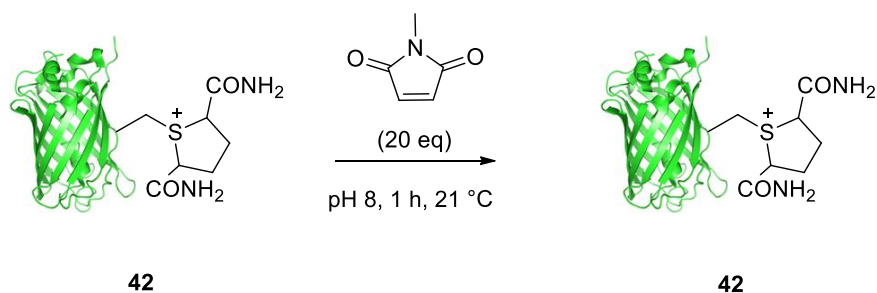
On the other hand, GFP (S147C) **32**, under the same reaction conditions, did not yield GFP (S147C, Dha) **41**. Instead, LCMS indicated quantitative conversion to

GFP (S147C, Sulf) **42** (Scheme 42). The isolation of this sulfonium species as a stable adduct was noteworthy as this intermediate had not previously been isolated. Crucially, the lack of formation of any GFP (S147C, Dha) **41** indicated a level of substrate control in the outcome of the reaction of a cysteine residue with 2,5-dibromohexanediamide **14**.



Scheme 42. GFP(S147C) **32** reaction with 2,5 dibromohexanediamide **14**

In order to gain further evidence for the site of reaction between GFP (S147C) **32** and 2,5-dibromohexanediamide **14**, GFP (S147C, Sulf) **42** was treated with *N*-methylmaleimide (20 eq). No reaction was observed, confirming that cysteine 147 had been modified.



Scheme 43. Reaction of GFP (S147C, Sulf) **42** with *N*-methylmaleimide

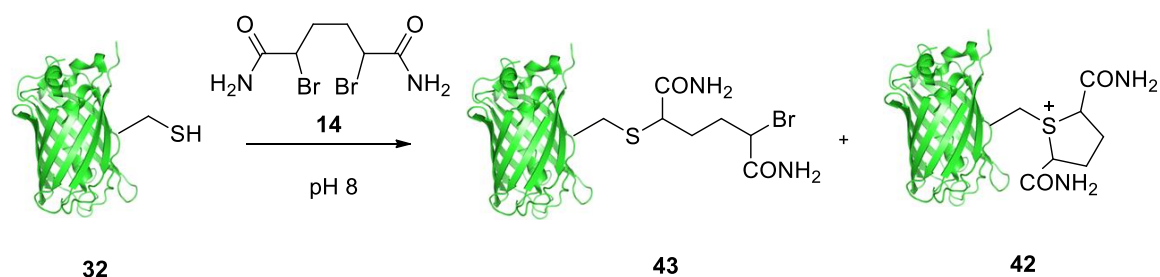
Efforts were made to optimize the reaction conditions for the formation of sulfonium **42** by varying reaction parameters, including number of equivalents of

2,5-dibromohexanediamide **14**, reaction temperature, and reaction time (*Table 1*). The yield of GFP (S147C, Sulf) **42** under different reaction conditions was estimated by ratio of peak heights in the deconvoluted mass spectra.⁷⁴

At 21 °C, no sulfonium **42** was detected after incubation of GFP (S147C) **32** with 10 equivalents of 2,5-dibromohexanediamide **14** for 2 h (*Table 1*, Entry 1). After prolonged incubation for 20 h, sulfonium **42** was formed in 53% yield (*Table 1*, Entry 3). Increasing the equivalents of 2,5-dibromohexanediamide **14** to 25 followed by incubation for 5 h led to the formation of a mixture of sulfonium **42** and thioether **43** with only 50% conversion being observed (*Table 1*, Entry 3). Encouragingly, sulfonium **42** was observed in complete conversion after the incubation of GFP (S147C) **32** with 2,5-dibromohexanediamide **14** for 20 h (*Table 1*, Entry 6). Increasing the equivalents of 2,5-dibromohexanediamide **14** to 50, followed by incubation for 20 h also led to complete conversion to sulfonium **42**, and no other side reactions were detected by LCMS (*Table 1*, Entry 9).

On ice, the reactions proceeded very slowly. After 20 h, no reaction of GFP (S147C) **32** was observed with 10 equivalents of 2,5-dibromohexanediamide **14** (*Table 1*, Entry 10). Increasing the equivalents of 2,5-dibromohexanediamide **14** to 25 led to formation thioether **43** as the sole product in only 20% yield (*Table 1*, Entry 11). Reaction with 50 equivalents of 2,5-dibromohexanediamide **14** after 20 h yielded sulfonium **37** in only 20% yield, with 56% of GFP (S147C) **32** being unreacted (*Table 1*, Entry 12).

At 37 °C, treatment of GFP (S147C) **32** with only 50 equivalents of 2,5-dibromohexanediamide **14** yielded sulfonium **42** after 2 h (*Table 1*, Entry 13).



Entry	14 eq	Temp (°C)	Time (h)	32 (%) ^[a]	43 (%) ^[a]	42 (%) ^[a]
1	10	21	2	>95	0	0
2			5	91	9	0
3			20	47	0	53
4	25	21	2	81	19	0
5			5	50	17	33
6			20	0	0	>95
7	50	21	2	56	23	21
8			5	27	25	48
9			20	0	0	>95
10	10	4	20	>95	0	0
11	25		20	80	20	0
12	50		20	56	24	20
13	50	37	2	0	0	>95

[a] Determined by ratio of peak heights in deconvoluted mass spectrum.

Table 1. Optimization of formation of GFP (S147C, Sulf) 42

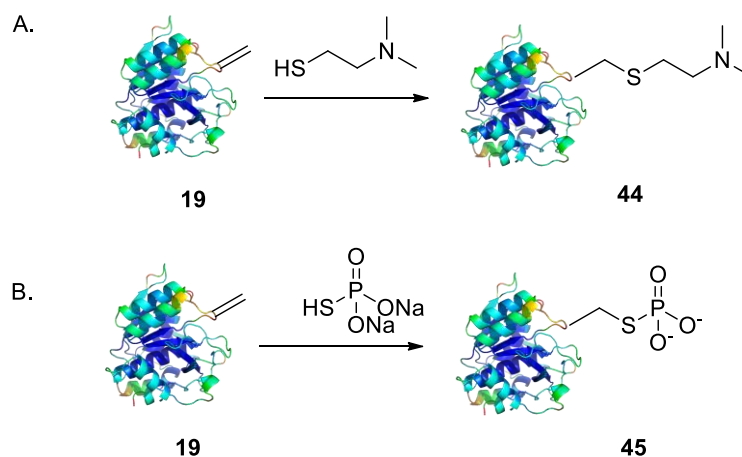
2.6 Sulfonium stability

The stability of the sulfonium species was evaluated by incubation of GFP (S147C, Sulf) **42** at 37 °C in buffer (sodium phosphate, 100 mM, pH 8) for a prolonged period of time. Even after 24 h, no GFP (S147C, Dha) **41** was observed. Instead, 45% of the construct degraded to an unidentifiable product. The mass of this degradation product was lower than GFP (S147C, Sulf) **42**, but higher than GFP (S147C) **28**. No degradation fragments with mass lower than GFP (S147C) **32** were observed.

At lower temperatures, GFP (S147C, Sulf) **42** appeared to be very stable. No degradation was observed after storing the sample at 4 °C in buffer (sodium phosphate, 100 mM, pH 8) for up to 1 month.

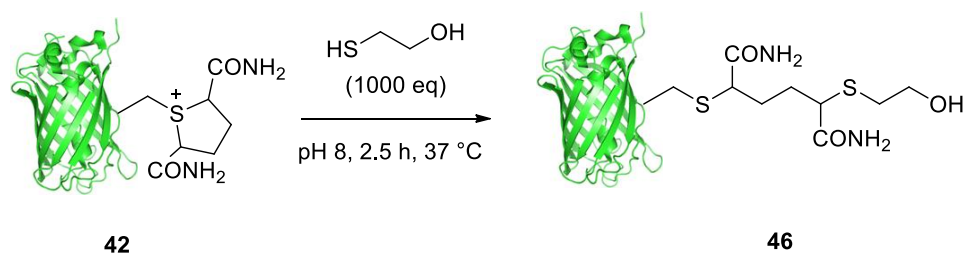
2.7 Reactivity of dehydroalanine and sulfonium

The utility of the dehydroalanine functionality has been amply demonstrated by its ability to react with thiol nucleophiles in a facile manner and has been widely reported in the literature.⁶⁷ The robustness of dehydroalanine mediated functionalisation has been demonstrated using a broad range of thiol nucleophiles including thiophosphate, thioglucose and cysteamine derivatives for synthesis of post-translationally modified protein mimics (*Scheme 44*).



*Scheme 44. Functionalisation of SBL156Dha 19 to methylated lysine mimic 44 and phosphorylated serine mimic 45*⁶⁷

The utility of sulfonium modification is dependent on its ability to serve as a handle for further protein functionalisation. To investigate if sulfonium species could also offer broad thiol reactivity as observed with dehydroalanine, GFP (S147C, Sulf) **42** was evaluated with model thiol nucleophile, 2-mercaptoethanol (1000 eq., 37 °C, 2.5 h). Gratifyingly, a clean conversion to thioether **46**, presumably *via* ring opening of the sulfonium **42** was observed by LCMS (*Scheme 45*).

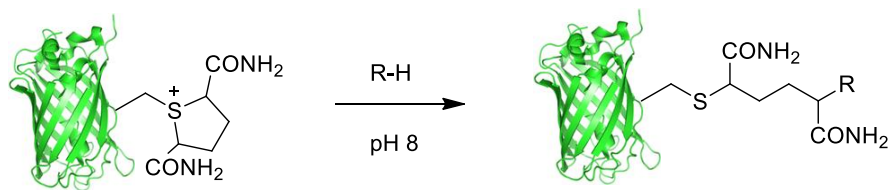


*Scheme 45. Reaction of GFP (S147C, Sulf) **42** with 2-mercaptoethanol*

To evaluate the reactivity of sulfonium moiety further, GFP (S147C, Sulf) **42** was treated with a number of different thiol nucleophiles. The reactions were carried out whilst varying time, temperature and equivalents of nucleophile. At 21 °C, GFP (S147C, Sulf) **37** incubated with a sodium salt of 1-thio- β -D-glucose (100 eq., sodium phosphate buffer, 100mM, pH 8) afforded GFP thioglucose conjugate **47** as the sole product, with complete conversion after 5 h (*Table 2*, Entry 2). Reaction of GFP (S147C, Sulf) **42** with glutathione under similar conditions also successfully led to formation of glutathione conjugate **48** (*Table 2*, Entry 4).

These reactions could also be carried out effectively with fewer equivalents (10 eq.) of 1-thio- β -D-glucose and glutathione at a higher temperature (37 °C), affording the expected products in with complete conversion after 5 h (*Table 2*, Entries 1 and 3). Reaction time could be effectively reduced by increasing the concentration of thiol nucleophile (*Table 2*, Entry 4). Encouragingly, the sulfonium functional group proved to be a versatile and reactive intermediate that tolerated functionalisation with a range of thiol nucleophiles. Thus, sulfonium modification represents a complementary approach for cysteine modification through dehydroalanine for cysteine residues which do not undergo elimination.

The reactivity of the sulfonium was due to the polarization of C-S⁺ bond and the strain of the 5 membered ring system, hence it was expected to be significantly different from dehydroalanine. The reactivity of GFP (S147C, Sulf) **42** towards non sulfur containing nucleophiles was thus investigated.



42

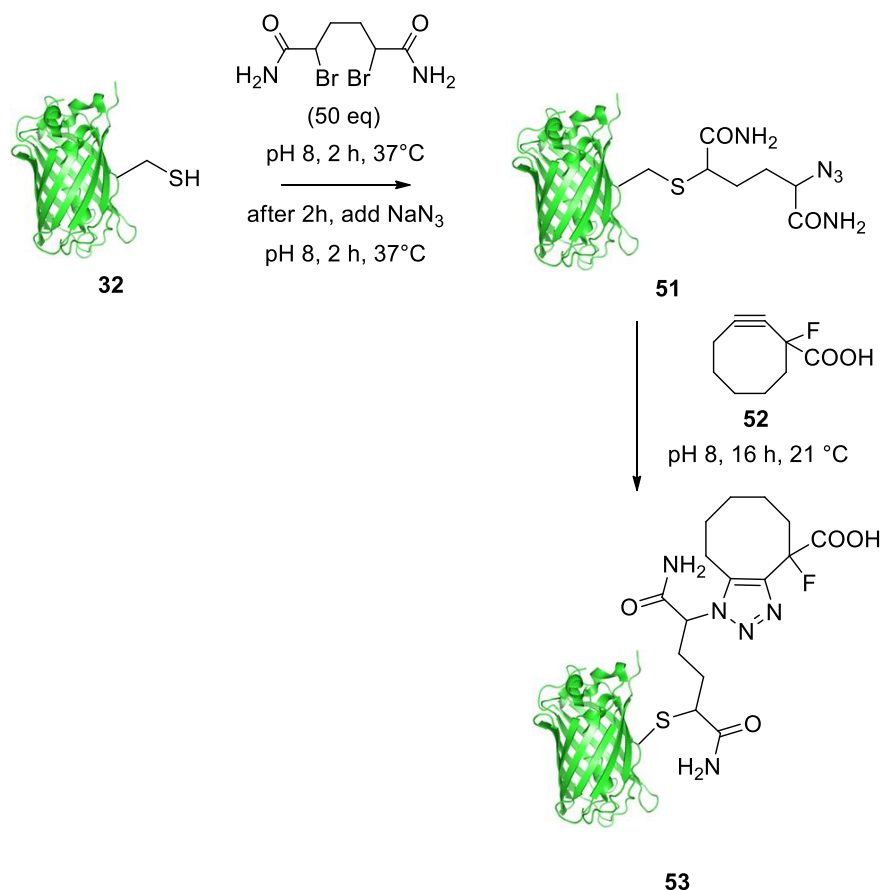
Entry	Nuc	eq.	Temp (°C)	Time (h)	Yield % ^[a]	Prod.
1		10	37	5	>95	47
2		100	21	5	>95	
3		10	37	5	>95	48
4		100	21	5	>95	
5		200	37	2	>95	
6		100	37	5	>95	49
7		1000	37	2.5	40	50
8	NaN ₃	1000	37	2.5	>95	51

[a] Determined by ratio of peak heights in deconvoluted mass spectrum.

Table 2. Conjugation of nucleophiles with GFP (S147C, Sulf) 42

An excess of phenyl selenol (100 eq.), a soft nucleophile similar to thiol, reacted rapidly with GFP (S147C, Sulf) **42** to afford the mixed thio/seleno bisether **49** in excellent conversion (*Table 2*, Entry 6) after incubation for 5 h at 37 °C. Whilst most amines failed to react with sulfonium **42** at pH 8 because of their high pK_a values, phthalimide did undergo reaction, forming **50** in 40 % yield after 2.5 h at 37 °C, while unreacted sulfonium **42** was also observed (*Table 2*, Entry 7). Interestingly, incubation of GFP (S147C, Sulf) **42** with sodium azide (1000 eq.) 2.5 h at 37 °C generated an azide labelled GFP (S147C, Azide) **51** in complete conversion (*Table 2*, Entry 8).

The ability to generate GFP (S147C, Azide) **51** was interesting as the azide group represents a handle for further protein modification using well established ‘click chemistry’.¹⁴⁵



*Scheme 46. One pot conversion of GFP (S147C) **32** to GFP (S147C, Azide) **51**, followed by strain-promoted click reaction*

With this in mind, a one pot method for conversion of the exposed cysteine residue on GFP (S147C) **32** to an azide handle GFP (S147C, Azide) **51** was developed. Incubation of GFP (S147C) **32** with 2,5-dibromohexanediamide **14** (50 eq., 2 h, 37 °C) followed by addition of sodium azide (sat. solution 100 µL/ mg of protein, 2 h, 37 °C) resulted in a clean conversion to GFP (S147C, Azide) **51**. The applicability of the azide reactive handle was analysed using strain-promoted azide-alkyne activated click chemistry (SPAAC).¹⁴⁵ Gratifyingly, treatment of **51** with a model strained octyne **52** (synthesised by Antoine Maruani, PhD Student, Caddick Group) afforded triazole product **53** in complete conversion (*Scheme 46*).

2.8 Dual modification

The ability to control the propensity of sulfonium to form dehydroalanine, by careful positioning of a cysteine residue in the structure of GFP, highlighted the potential of this approach to deliver substrate-controlled chemical protein modification. These initial observations were explored further to ascertain the possibility of using this approach to differentiate between two different cysteine residues on the protein surface, whilst introducing two orthogonal reactive groups for further dual protein modification.

2.8.1 Dual cysteine mutants

Two dual cysteine mutants, GFP (S147C, +CG) **54** and GFP (S147C, T230C, 233Δ) **55** (*Figure 8*), were designed to explore bisalkylation of cysteines using 2,5-dibromohexanediamide **14** to obtain site-selective dual modification.

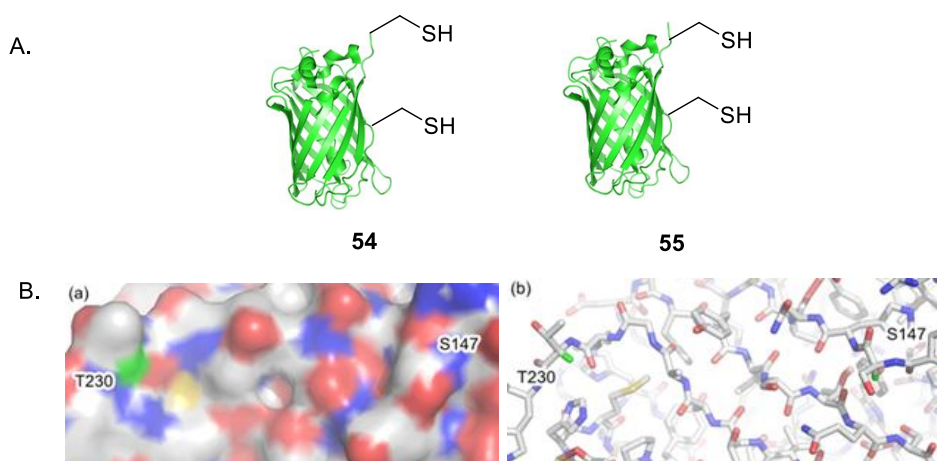


Figure 8. A) Dual Mutants GFP (S147C, +CG) **54** or GFP (S147C, T230C, 233Δ) **55** B) a) Surface and b) Stick representation of GFP (S147C, T230C, 233Δ) **55** (α -proton coloured in green) (PDB: 2B3P)

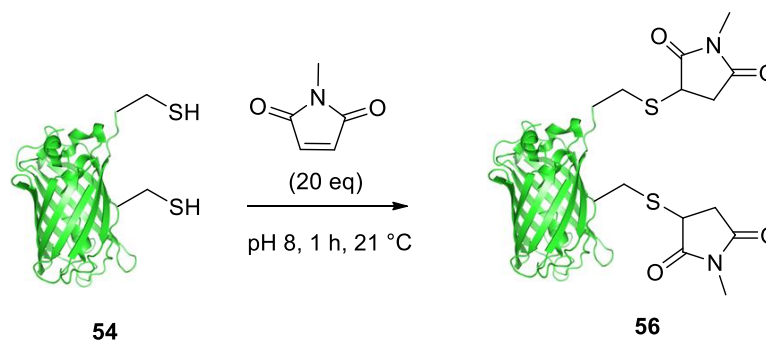
GFP mutants (S147C, +CG) **54** and GFP (S147C, T230C, 233Δ) **55** were designed, cloned and expressed using standard biochemistry techniques (**55** was designed, cloned and expressed by Paul Moody, PhD student, Caddick Group). As discussed in Section 2.3, the α -proton for cysteine at position 147 was expected to be solvent inaccessible (Figure 8B. α -proton coloured in green), whereas the α -protons for cysteine residues at +CG and 230 were solvent accessible (Figure 8B. α -proton coloured in green). Prior to all reactions with 2,5-dibromohexanediamide **14**, the GFP mutants were treated with dithiothreitol (DTT) to reduce any disulfide dimers. Excess reducing agent was removed by repeated diafiltration into buffer (sodium phosphate, 100 mM, pH 8).

2.8.2 Reaction of dual mutants with *N*-methylmaleimide

To probe the relative reactivity and accessibility of the engineered cysteine residues, GFP dual mutants **54** and **55** were incubated with a single equivalent of *N*-methylmaleimide at 21 °C under mildly basic conditions (sodium phosphate buffer, 100 mM, pH 8) and analysed by LCMS after 1 h.

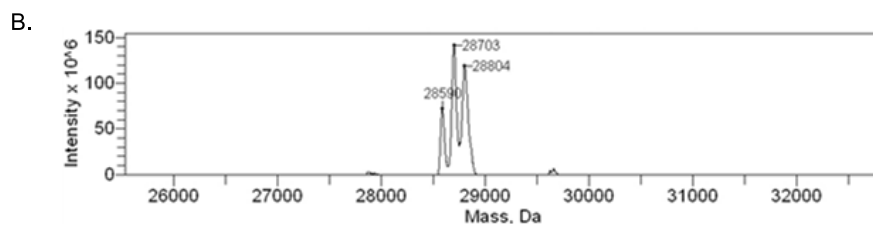
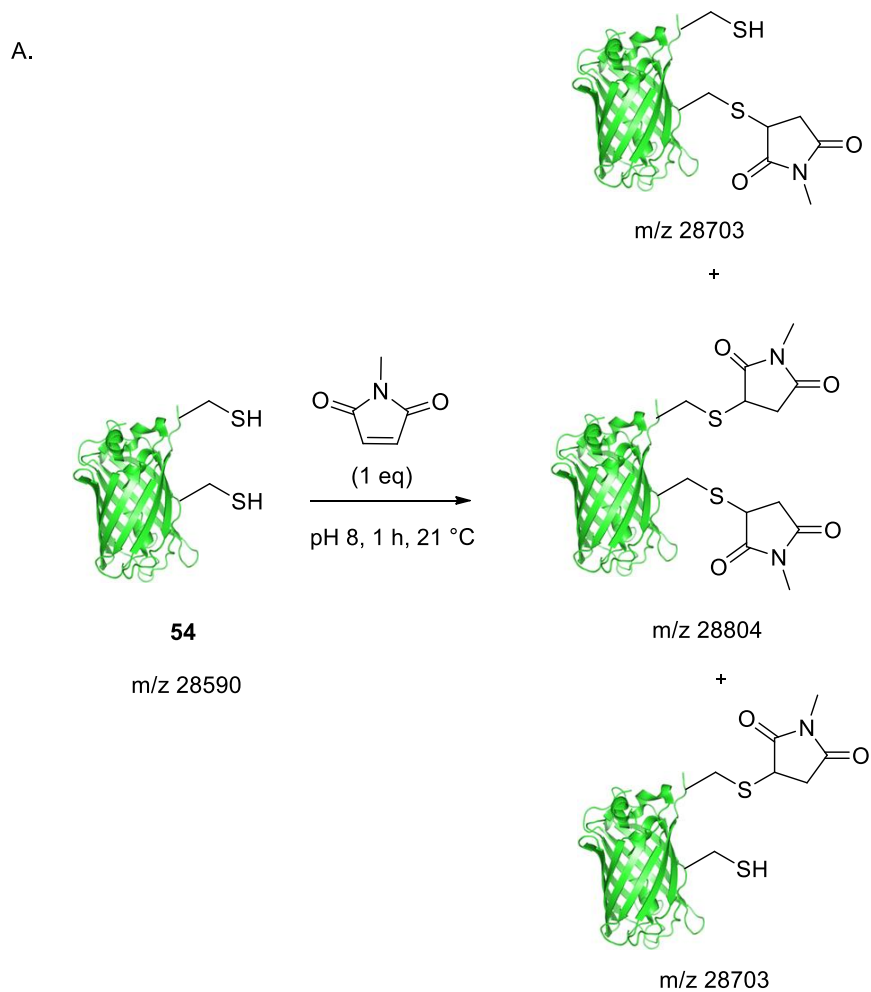
Due to propensity of cysteine residues of GFP (S147C, +CG) **54** to reoxidize rapidly and remain as intramolecular disulfide, reaction with single equivalent of *N*-methylmaleimide did not proceed to completion. However, following reduction with DTT and incubation with excess *N*-methylmaleimide (20 eq.), GFP (S147C, +CG)

54 did react to form dual maleimide labelled GFP (S147C *N*-methylmaleimide, +CG *N*-methylmaleimide) **56** (Scheme 47).



*Scheme 47. Reaction of GFP (S147C, +CG) **54** with *N*-methylmaleimide*

On the other hand, the reaction of GFP (S147C, T230C, 233Δ) **55** with a single equivalent of *N*-methylmaleimide proceeded to completion giving three distinguishable peaks in the mass spectra; i) unreacted GFP starting material, ii) GFP linked to single maleimide; and iii) GFP linked to two maleimides, in a 1:2:1 ratio respectively (Scheme 48). The 1:2:1 ratio demonstrated the essentially equivalent reactivity of the two cysteine residues for GFP (S147C, T230C, 233Δ) **55**.

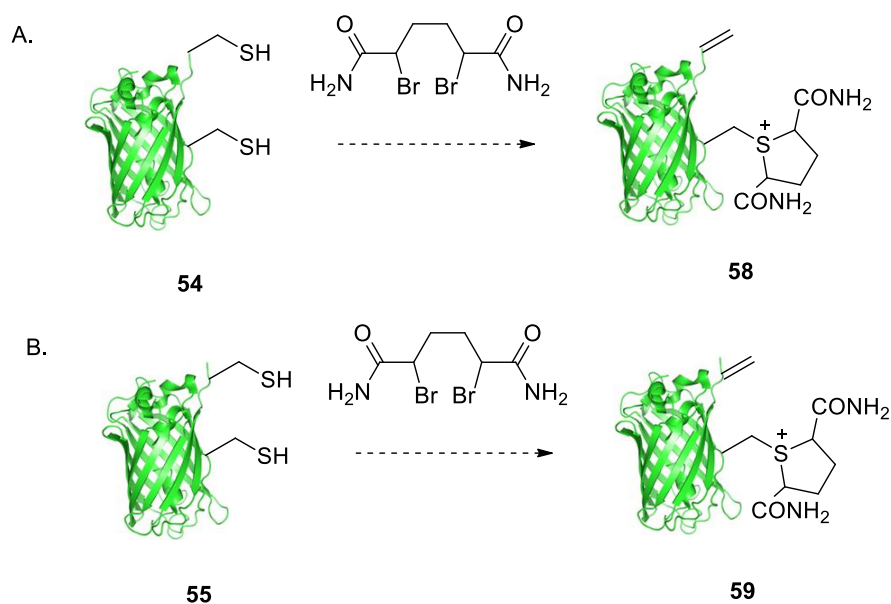


*Scheme 48. Reaction of GFP (S147C, T230C, 233 Δ) **54** with N-methylmaleimide, LCMS spectra and m/z assigned.*

2.8.3 Reaction of Dual Mutants with 2,5-dibromohexanediamide

Extrapolating from the results reported in Section 2.4, it was anticipated that the reaction of cysteine dual mutants GFP (S147C, +CG) **54** and GFP (S147C, T230C,

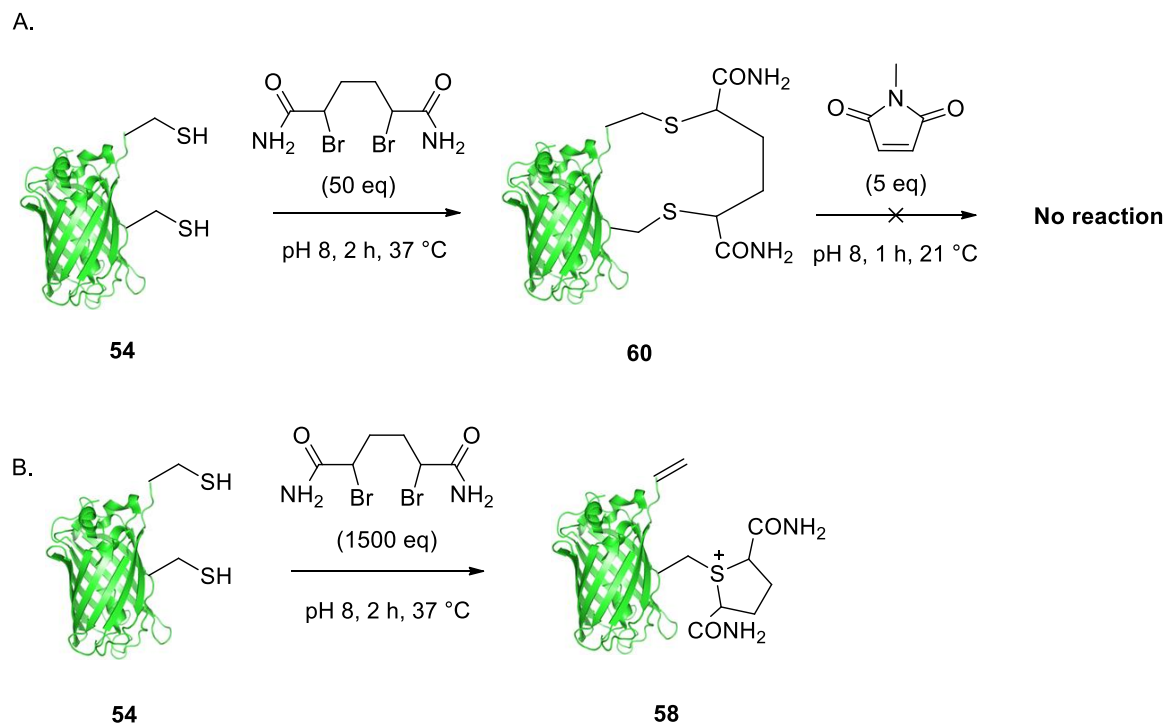
233Δ) **55** with 2,5-dibromohexanediamide **14** would yield sulfonium at position 147 and dehydroalanine at positions +CG and 230 (*Scheme 49*).



*Scheme 49. Expected reaction of mutant GFP (S147C, +CG) **54** and GFP (S147C, T230C, 233A) **55** with 2,5-dibromohexanediamide **14***

GFP (S147C, +CG) **54** was incubated with the 2,5-dibromohexanediamide **14** (50 eq., sodium phosphate buffer, 100 mM, pH 8) for 2 h at 37 °C. Surprisingly, LCMS analysis indicated an unknown product with a mass increase of 142 Da, the expected mass change for conversion of only one cysteine residue to sulfonium. The second cysteine residue appeared to have not reacted at all. This was unexpected as the reaction of GFP (S147C, +CG) **54** with *N*-methylmaleimide indicated the presence of two cysteine residues (*Scheme 47*). To analyse if there was any free cysteine thiol, the unknown product was subjected to reaction with *N*-methylmaleimide. On incubation with 5 equivalents of *N*-methylmaleimide for 1 h at 21 °C (sodium phosphate buffer, 100 mM, pH 8), this unknown product remained completely unchanged. Prior to reduction by DTT, GFP (S147C, +CG) **54** preferred to exist as an intramolecular disulfide not reacting with *N*-methylmaleimide. It was therefore hypothesized that this unknown product could be a bisthioether **60** (*Scheme 50*). This indicated that 2,5-dibromohexanediamide **14** reacted with one of the cysteine thiols first to form a sulfonium or thioether which was then trapped by the remaining free cysteine residue in an intramolecular reaction. To overcome this problem and prevent intraprotein reactivity, the molar equivalents of 2,5-dibromohexanediamide **14** were

increased from 50 to 1500. Gratifyingly, this led to formation of the expected GFP (S147C Sulf, +CG Dha) **58**.

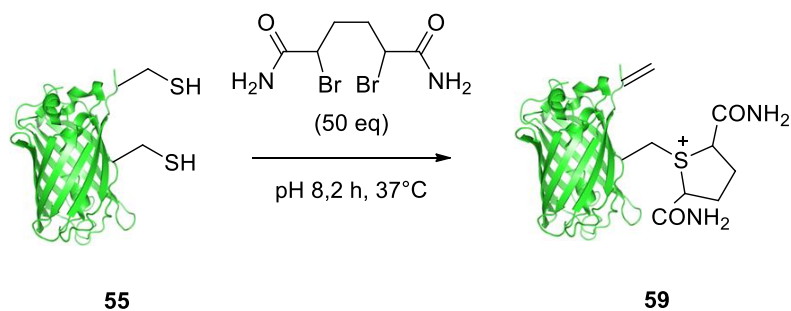


*Scheme 50. Reaction of GFP (S147C, +CG) **54** with 2,5-dibromohexanediamide **14***

A) 50 equivalents B) 1500 equivalents

The second dual mutant, GFP (S147C, T230C, 233Δ) **55** reacted smoothly with just 50 equivalents of 2,5-dibromohexanediamide **14** under similar conditions (sodium phosphate buffer, pH 8, 2 h, 37 °C) to give the expected dual modified product GFP (S147C Sulf, T230C Dha) **59** (*Scheme 51*).

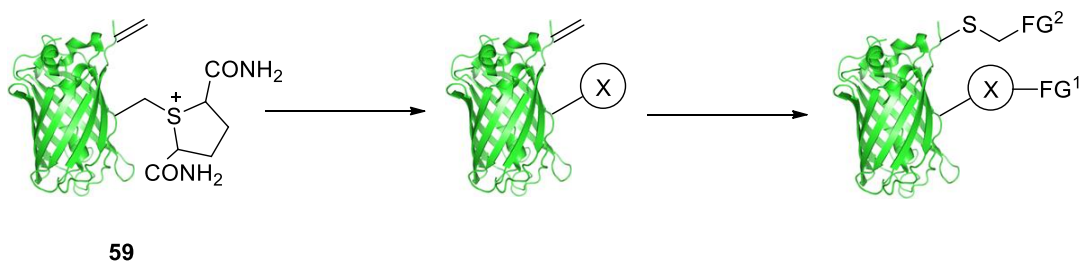
Attempts to identify the position of sulfonium and dehydroalanine on the modified protein were made using tryptic digest followed by mass spectroscopy. However, despite repeated attempts, multiple proteases including trypsin and chymotrypsin were found to be inactive towards GFP.



Scheme 51. Reaction of GFP (S147C, T230C, 233A) 55 with 2,5-dibromohexanediamide 14

2.8.4 Dual functionalisation of dual modified GFP

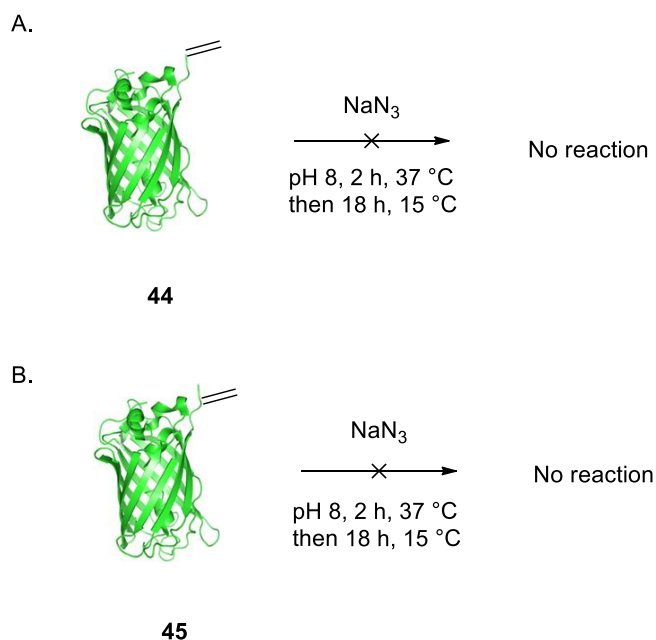
The observed differential reaction of GFP dual cysteine mutants with 2,5-dibromohexanediamide **14** represents the first reported instance of substrate-controlled regioselective, chemical dual-modification of a protein. This differential reactivity could potentially be used for the introduction of two orthogonally reactive functional groups onto a single protein. The exploitation of differential reactivity necessitated a method for orthogonal functionalisation of sulfonium and dehydroalanine.



Scheme 52. Conversion of sulfonium to a reactive intermediate followed by functionalisation

Further modification of sulfonium into a different reactive intermediate that allowed orthogonal protein functionalisation in the presence of dehydroalanine functionality was envisaged (*Scheme 52*). Conversion of sulfonium into a non-thiol reactive functional group would allow facile dual functionalization compatible with thiol modification of dehydroalanine.

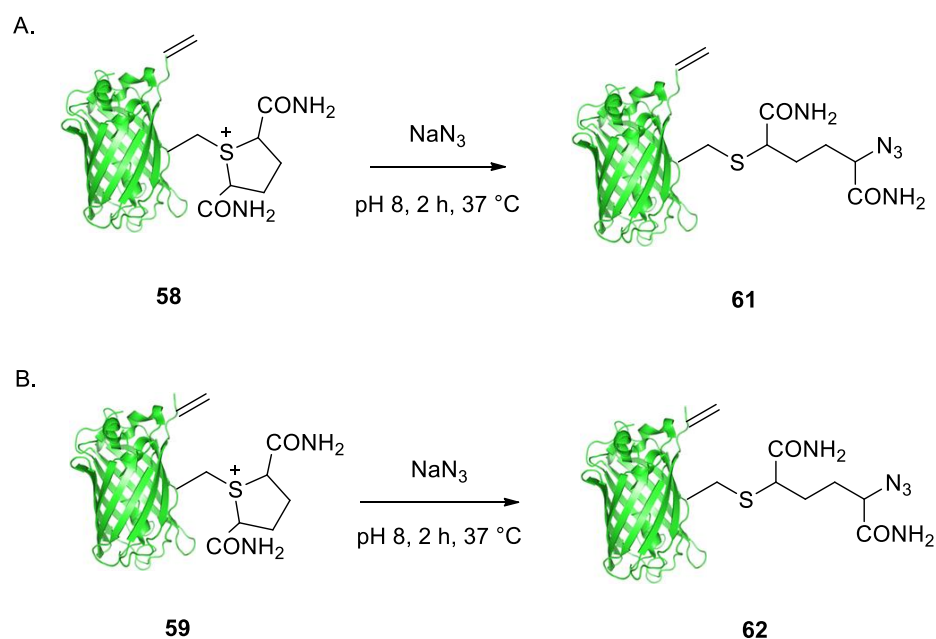
The functional group of immediate interest was azide, as it had already been shown that this functionality could be introduced successfully onto GFP (S147C, Sulf) **42** by ring-opening of sulfonium. Dehydroalanine residues have been previously reported to be functionalized by conjugate addition of thiols,^{67,70} but no previous accounts of reaction with azides exist. The inertness of dehydroalanine to azide, was the key to the success of this approach, and this was evaluated by incubation of GFP (+CG, Dha) **44** and GFP (T230C, Dha) **45** with sodium azide (sodium phosphate buffer, pH 8, 2 h, 37°C). Gratifyingly, no reaction was observed indicating that dehydroalanine was unreactive to azides (*Scheme 53*). No reaction was observed even after a prolonged reaction time period of 18 h at 15 °C.



*Scheme 53. Reaction of GFP (+CG, Dha) **44** and GFP (T230C, Dha) **45** with sodium azide*

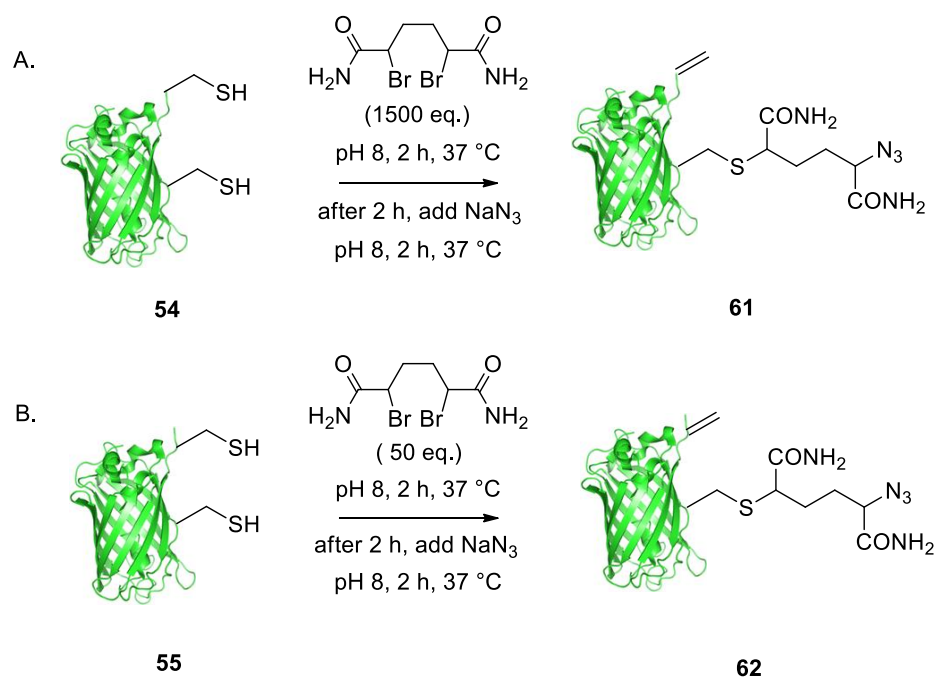
The inertness of dehydroalanine towards sodium azide afforded the opportunity to synthesize site-selectively dual functionalised GFP using dual cysteine mutants GFP (S147C, +CG) **54** and GFP (S147C, T230C, 233Δ) **55**. Incubation of GFP (S147C Sulf, +CG Dha) **58** and GFP (S147C Sulf, T230C Dha) **59** with azide (NaN_3 , >1000 eq, 37 °C, 2 h) resulted in the formation of GFP (S147C Azide, +CG Dha) **61** and GFP (S147C Azide, T230C Dha) **62**, as determined by LCMS (*Scheme 54*). The inertness of dehydroalanine **44** and **45** towards azide (*Scheme 53*) strongly indicated that sodium azide reacts selectively with position 147 sulfonium to generate the

desired dehydroalanine/azide adduct in preference to an azidoalanine/sulfonium adduct.



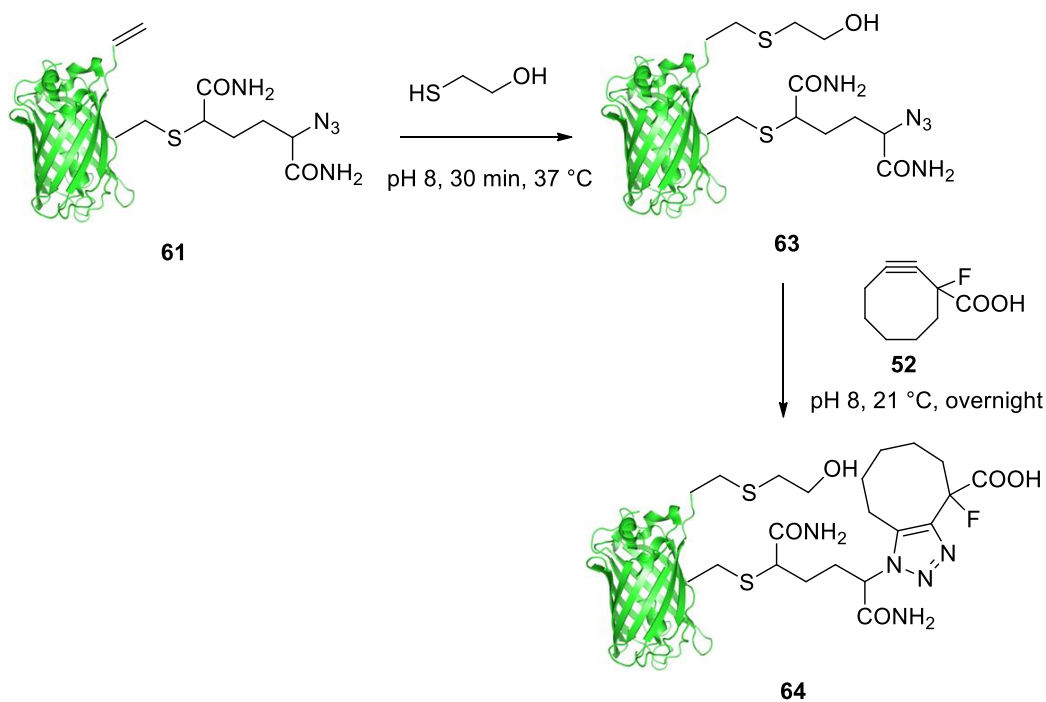
Scheme 54. Reaction of GFP (S147C Sulf, +CG Dha) 58 and GFP (S147C Sulf, T230C Dha) 59 with sodium azide

Azide/Dehydroalanine constructs **61** and **62** represent an elegant platform for orthogonal protein dual functionalisation. The orthogonality and robustness of the reactions to gain access to these constructs allowed these reactions to be carried out in one-pot without need for intermediate purification steps. Incubation of GFP (S147C, +CG) **54** and GFP (S147C, T230C) **55** with 2,5-dibromohexanediamide **14** (50 eq.) for 2 h at 37 °C under basic conditions (sodium phosphate, 100 mM, pH 8.0) followed by treatment with sodium azide (>1000 eq) for 2 h at 37 °C resulted in the GFP (S147C Azide, +CG Dha) **61** and GFP (S147Azide, T230 Dha) **62** in a single pot reaction (*Scheme 55*).



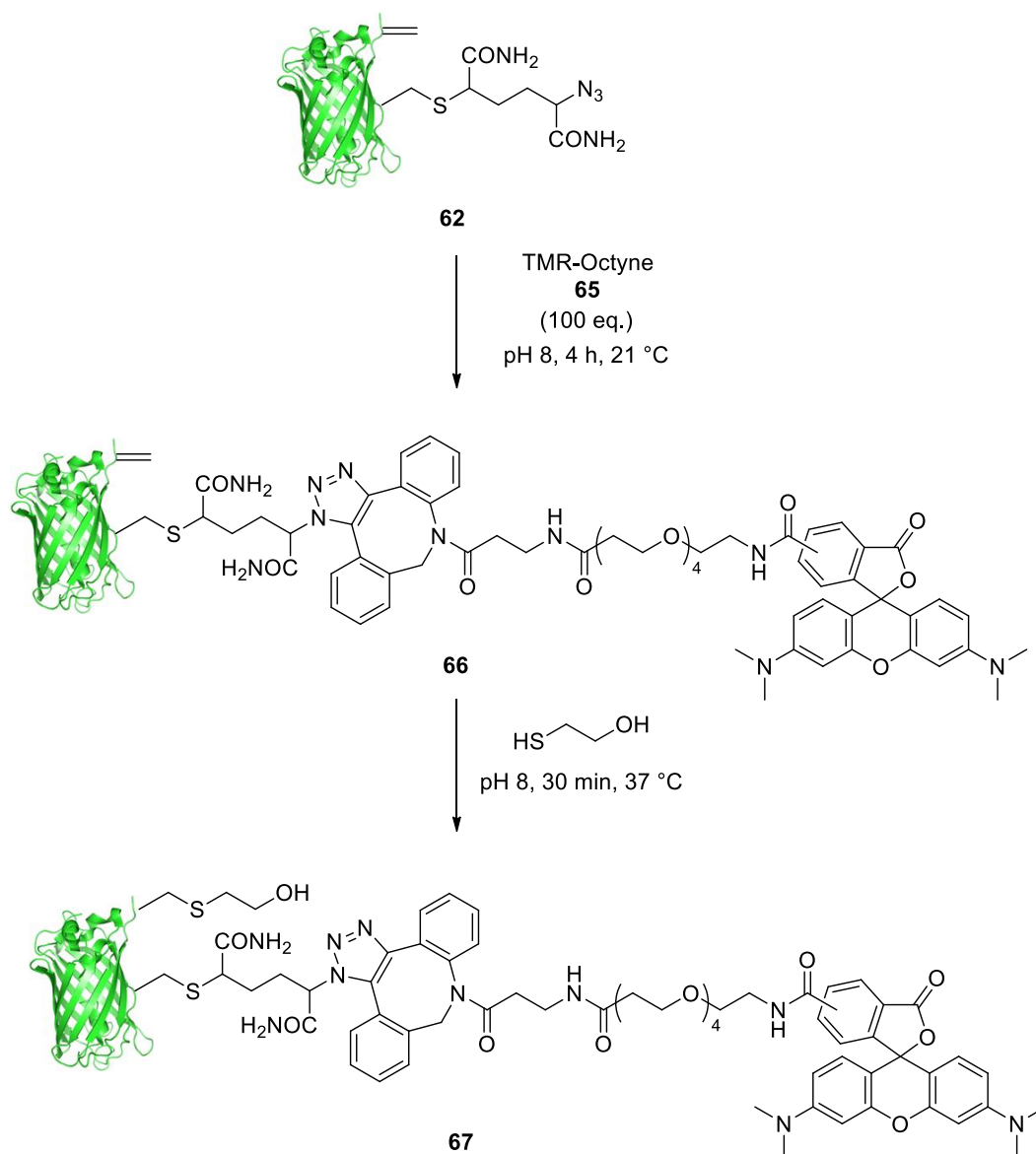
*Scheme 55. One pot conversion of dual mutants GFP (S147C, +CG) **54** and GFP (S147C, T230C, 233Δ) **55** to azide/dehydroalanine constructs GFP (S147C Azide, +CG Dha) **61** and GFP (S147Azide, T230Dha) **62***

The dual functionalisation of azide/dha constructs **61** and **62** was then pursued. Incubation of GFP (S147C Azide, +CG Dha) **61** with excess 2-mercaptoethanol (10 μL/mg protein, sodium phosphate, pH 8, 100 mM) for 30 min at 37 °C resulted in the formation of conjugate **63** (*Scheme 56*). After removal of excess thiol using diafiltration, reaction with octyne **52** (100 eq., overnight, 21 °C) led to the formation of the desired dual functionalized conjugate GFP (S147C Octyne, +CG BME) **64** (*Scheme 56*).



Scheme 56. Dual functionalisation of GFP (S147C Sulf, +CG Dha) 53

To test the robustness of this dual functionalisation approach, GFP (S147C Azide, T230C Dha) **62** was functionalized in the reverse order. The azide functional group was modified using strain-promoted azide-alkyne activated click chemistry (SPAAC)¹⁴⁵ by treatment with commercially available strained alkyne fluorophore, dibenzylcyclooctyne-PEG4-Fluor-545 **65** (100 eq., 4 h, 21 °C). After removal of excess fluorophore using diafiltration, the treatment of the resultant fluorophore-labelled-GFP conjugate **66** with excess 2-mercaptoethanol (10 μL/mg protein, 30 min, 37 °C) resulted in formation of dual functionalised GFP (S147C Azide dibenzylcyclooctyne-PEG4-Fluor-545, T230C Dha 2-mercaptoethanol) conjugate **67** as the sole product (*Scheme 57*). It is important to note that this approach of functionalisation, *i.e.* reaction of azide prior to dehydroalanine generally yielded cleaner conjugates, as evidenced by LCMS.



Scheme 57. Dual functionalisation of GFP (S147C Sulf, T230C Dha) 62

2.9 Conclusion

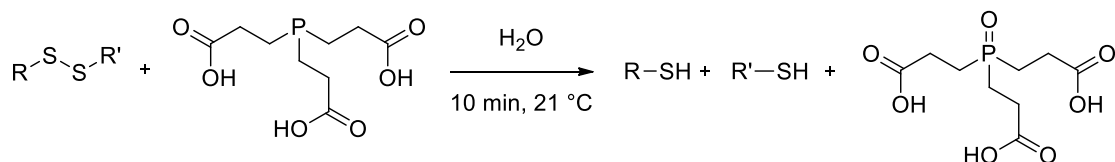
A novel method for the site-selective, dual-labelling of a protein *via* cysteine modification has been developed. The method exploits the reaction of 2,5-dibromohexandiamide with cysteine residues forming either sulfonium or dehydroalanine depending on protein structure and residue microenvironment. It provides the first example of the substrate-controlled, dual-modification of a protein.

A strategy for orthogonal functionalisation of sulfonium and dehydroalanine residues was also developed. The utility of this approach was demonstrated by a model GFP cysteine dual mutant that was site-selectively labelled with commercially available octyne and thiol probes yielding a homogenous product. Overall, the strategy represents a chemoselective dual protein modification method that enables efficient and robust dual functionalisation of cysteine side chains using orthogonal reactivity under mild conditions underpinned by two sequential chemoselective reactions. This approach requires only routine molecular biology techniques and the functionalisation can be affected using a wide range of commercially available reagents, in a simple and robust manner in a single day.

Chapter 3 Cysteine modification to thiophosphonium

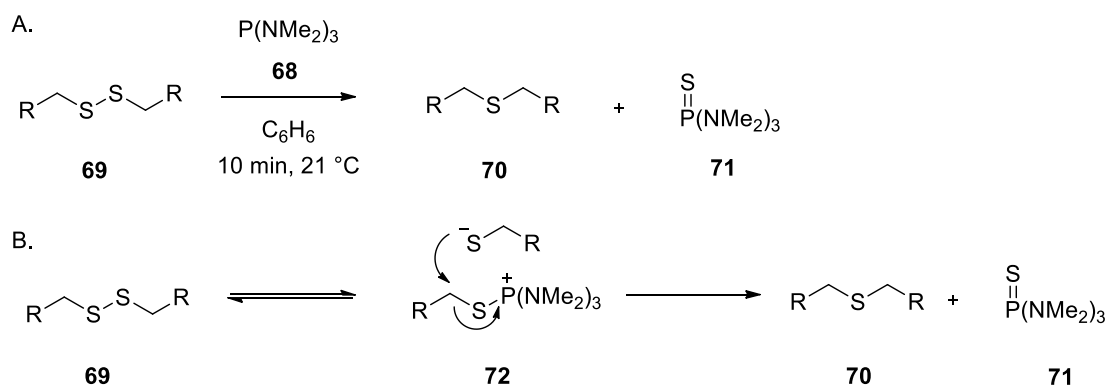
3.1 Introduction

Phosphorus (III)-based reagents exhibit high selectivity towards disulfide bonds and have been widely employed as chemical tools for protein modification. The most common application of such reagents is for selective reduction of disulfide bonds in proteins and peptides, usually achieved using tris(2-carboxyethyl)phosphine (TCEP) (*Scheme 58*).¹⁵¹



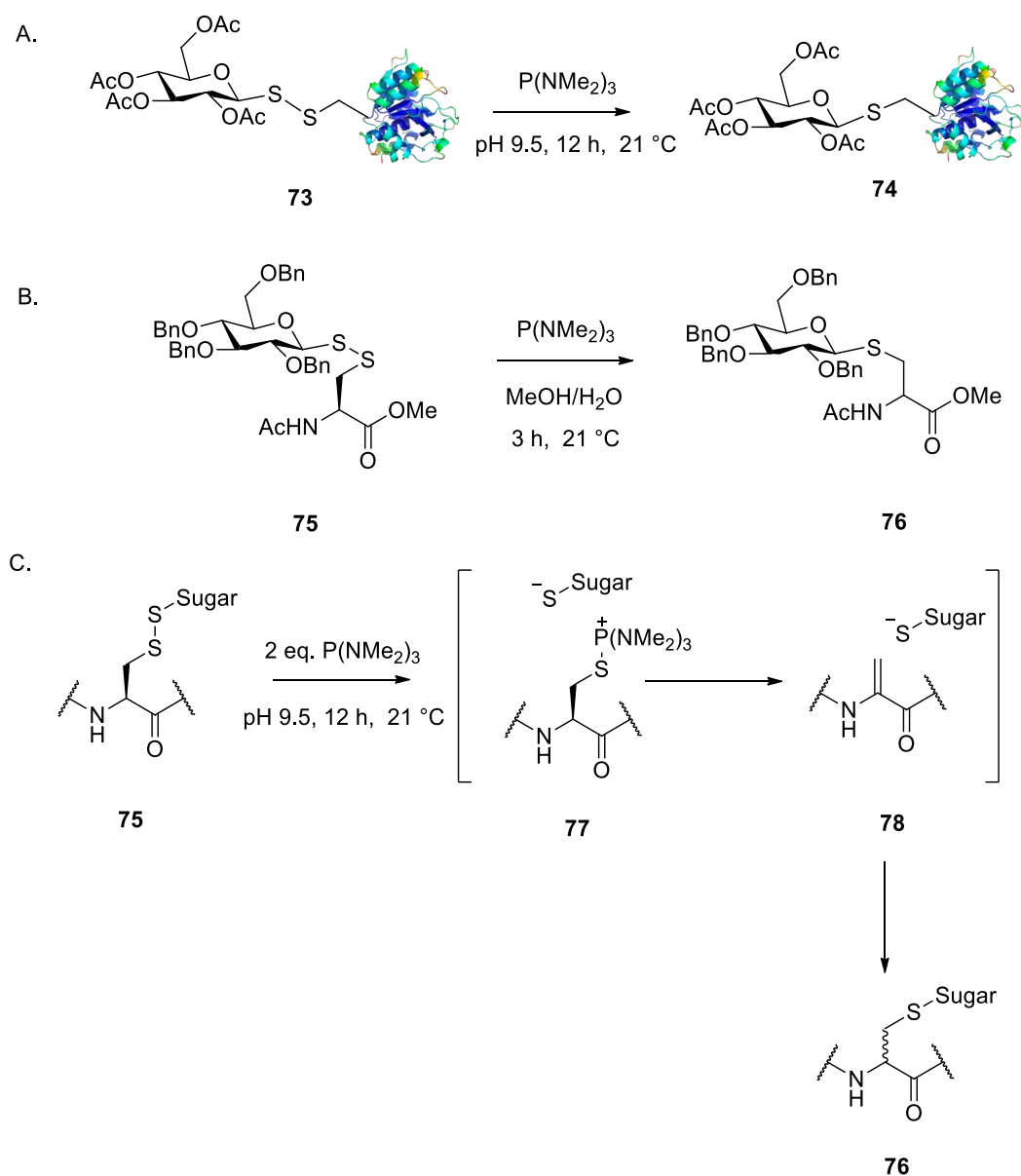
Scheme 58. Tris(2-carboxyethyl)phosphine mediated reduction of disulfides

Another phosphorus(III)-based reagent, tris(dimethylamino)phosphine **68** exhibits similar selectivity for disulfide bonds and has been reported to be an efficient desulfurizing agent for small molecule substrates.^{56,152,153,154} Harpp *et al.* demonstrated rapid and selective desulfurization of disulfides using tris(dimethylamino)phosphine **68** to the corresponding thioethers in near-quantitative yield (*Scheme 59A*).¹⁵⁴ The reaction has been proposed to proceed *via* thiophosphonium intermediate (*Scheme 59B*).



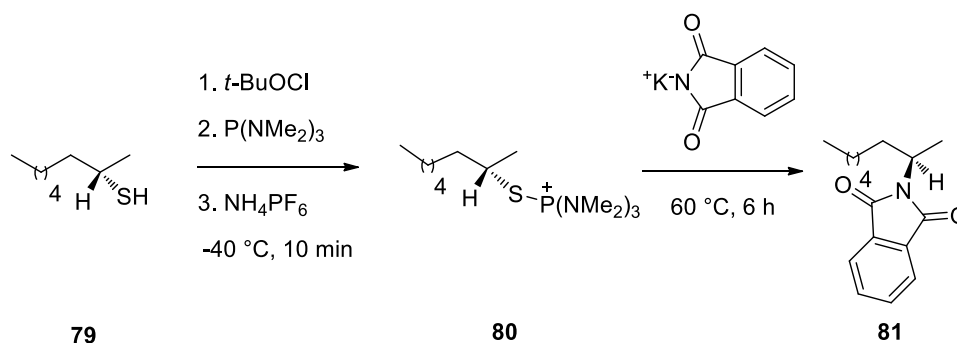
Scheme 59. A) Conversion of disulfides to thioethers B) Proposed mechanism via displacement of thiophosphonium

Davis *et al.* used tris(dimethylamino)phosphine **68** for the conversion of a disulfide linked glycoprotein **73** into thioether linked glycoprotein **74** under basic conditions (pH 9.5) (*Scheme 60A*).¹⁵⁵ Mechanistic investigations of this reaction with a model disulfide linked glycosyl amino acid **75** revealed loss of stereochemical information at the α -carbon, yielding the thioether product **76** as a mixture of diastereoisomers (*Scheme 60B*). The retention of the anomeric configuration suggested that the stereodivergence resulted from the racemization of α -proton of the cysteine residue. Thus, the reaction mechanism was hypothesized to proceed *via* a dehydroalanine intermediate **78** formed by β -elimination of thiophosphonium **77**. The thioether product **76** was obtained presumably *via* conjugate addition of glycosyl thiol onto Dha (*Scheme 60C*).



Scheme 60. Reaction of tris(dimethylamino)phosphine 68 with A) disulfide linked glycoprotein 73 B) disulfide linked glycosyl amino acid 75 C) Mechanism of transformation of disulfide to thioether via Dha intermediate

Thiophosphoniums are activated species that have the potential to provide access to previously unattainable chemical modifications. Krafft *et al.* demonstrated displacement of thiophosphonium **80** from chiral carbon centers by a range of nucleophiles including phthalimide, bromide and azide (*Scheme 61*).¹⁵⁶



*Scheme 61. Displacement of thiophosphonium by phthalimide*¹⁵⁶

The displacement reactions, reported for a variety of secondary and primary thiophosphonium substrates, proceeded with an inversion of configuration at the carbon α to the thiophosphonium linkage. The reaction was proposed to proceed *via* an $\text{S}_{\text{N}}2$ pathway.¹⁵⁶ The displacement reactions were typically carried out under dry conditions in a polar solvent by heating at 60 °C for 6 h, leading to high yields.

3.2 Aims

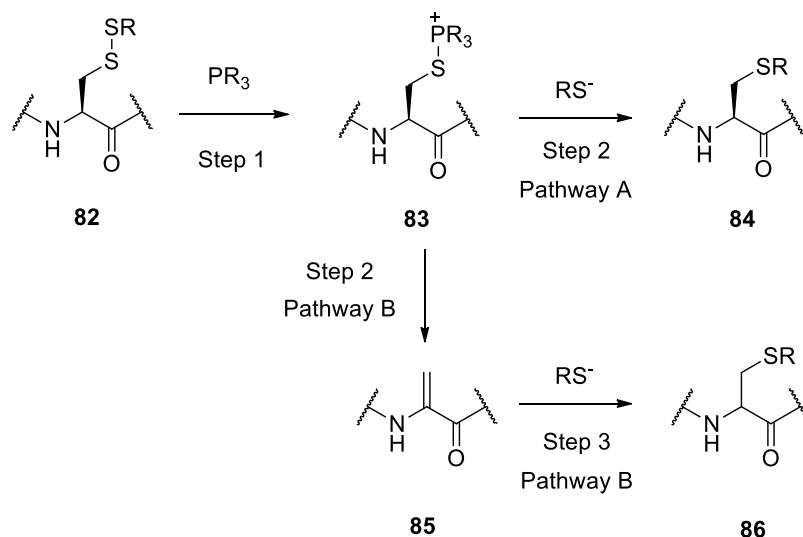
A major aim of this work was to explore the mechanism of the reaction of disulfides with tris(dimethylamino)phosphine **68** and to provide evidence for existence of thiophosphonium species. This work also sought to isolate the thiophosphonium species to evaluate its stability, ease of access and reactivity for protein bioconjugation.

The reaction of cysteine based disulfide substrates with tris(dimethylamino)phosphine **68** to form thioethers can proceed *via* two pathways (*Scheme 62*):

Pathway A: Step 1 involves reaction of phosphine **68** with a disulfide **82**, forming a thiophosphonium **83** intermediate and releasing the corresponding thiol. In step 2, the thiophosphonium **83** then undergoes a $\text{S}_{\text{N}}2$ type substitution reaction forming thioether **84** with retention of stereochemistry (*Scheme 62*); or

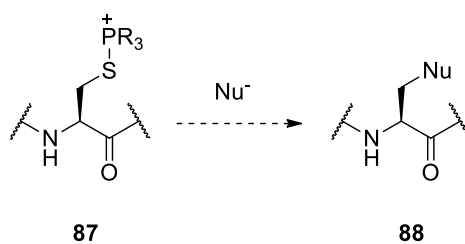
Pathway B: Step 1 again involves reaction of phosphine **68** with a disulfide **82**, forming a thiophosphonium **83** intermediate and releasing the corresponding thiol.

Under basic conditions, however, the thiophosphonium intermediate **83** undergoes β -elimination to form dehydroalanine **85** (Step 2).¹⁵⁵ In step 3, conjugate addition of a thiol to dehydroalanine **85** forms thioether **86** with the loss of stereochemical information (Scheme 62).



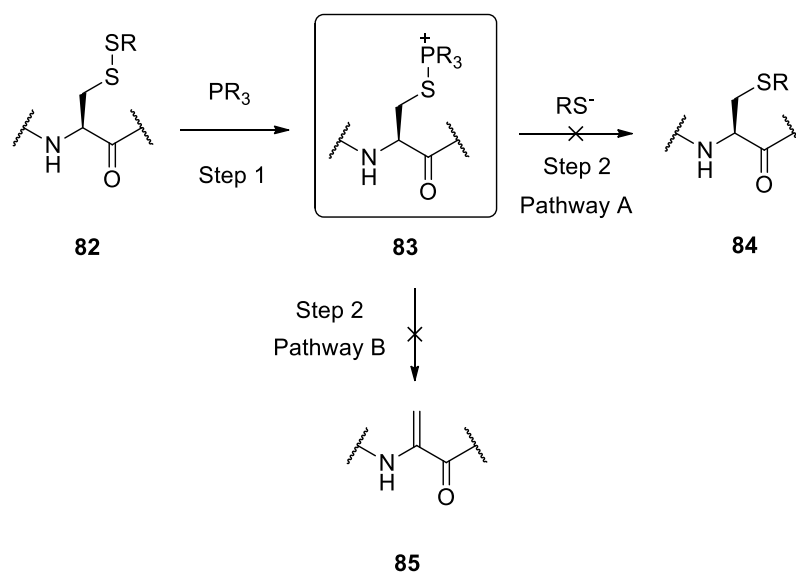
Scheme 62. Pathways for the conversion of disulfide into a thioether

While both reaction pathways have been proposed, no evidence for the existence of the thiophosphonium species, to the best of our knowledge, is available. Further investigations may prove informative and provide access to thiophosphonium as a highly reactive intermediate that could serve as a platform to access to a range of previously unattainable functionalities. This would find plentiful applications in proteomics, protein modification and protein therapeutics. The displacement of thiophosphonium *via* a S_N2 type displacement reaction could also allow access to stereochemically homogenous post-translational mimics (PTMs) that are unattainable by dehydroalanine functionalization.¹⁵⁶



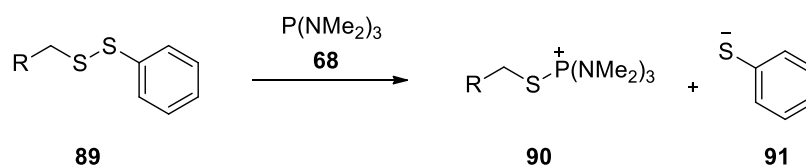
Scheme 63. Access to novel homogenous amino acids or PTMs via thiophosphonium

To access thiophosphonium **83**, it would be essential to prevent step 2 in the reaction sequence for both pathways A and B, as shown below (*Scheme 62*).



Scheme 64. Formation of stable thiophosphonium intermediate

Pathway A could be shut down by employing a suitably biased alkyl-aryl disulfide, such that the released aromatic thiol would be less likely to participate in the displacement reaction owing to its poorer nucleophilicity. This is likely to be a favourable alteration as previous literature reports indicate the preferential attack of phosphine on the sulfur that results in the release of the most stable and therefore the least nucleophilic thiolate (*Scheme 65*).¹⁵⁴



Scheme 65. Reaction of alkyl-aryl disulfide with tris(dimethylamino)phosphine 68

Pathway B, or β -elimination of thiophosphonium **83** to dehydroalanine **85** has been reported to occur under basic conditions (pH 9.5). Hence, we postulate that pH control would allow isolation of a thiophosphonium species. In view of the results obtained in Chapter 2, it should also be noted that substrate control could also prevent the formation of dehydroalanine.

The following aims were identified for this project:

1. To investigate evidence for thiophosphonium intermediate following reaction of a disulfide with tris(dimethylamino)phosphine;
2. To design a suitable alkyl-aryl disulfide system and isolate the subsequent thiophosphonium species;
3. To investigate the stability of thiophosphonium species under aqueous conditions; and
4. To investigate the reactivity of thiophosphonium intermediates towards nucleophiles and thus to determine the suitability of this approach for protein modification.

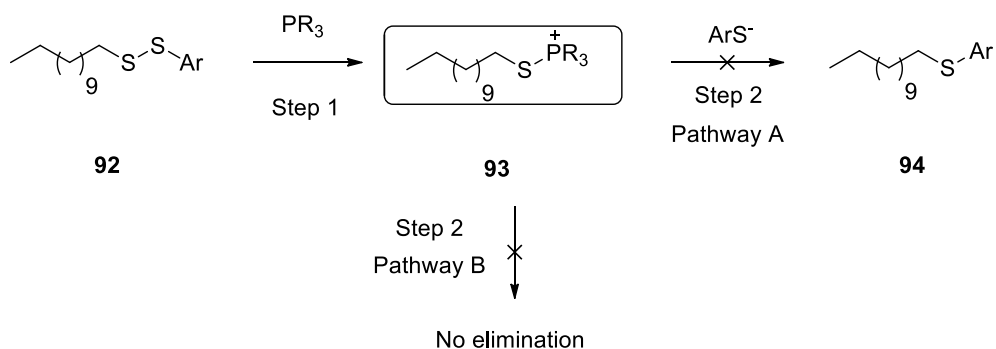
3.3 Reaction of disulfides with tris(dimethylamino)phosphine

3.3.1 Small molecule model system

The reaction of disulfides with tris(dimethylamino)phosphine **68** was investigated initially on small molecule model system as it affords greater mechanistic insight and facile detection of the thiophosphonium species.

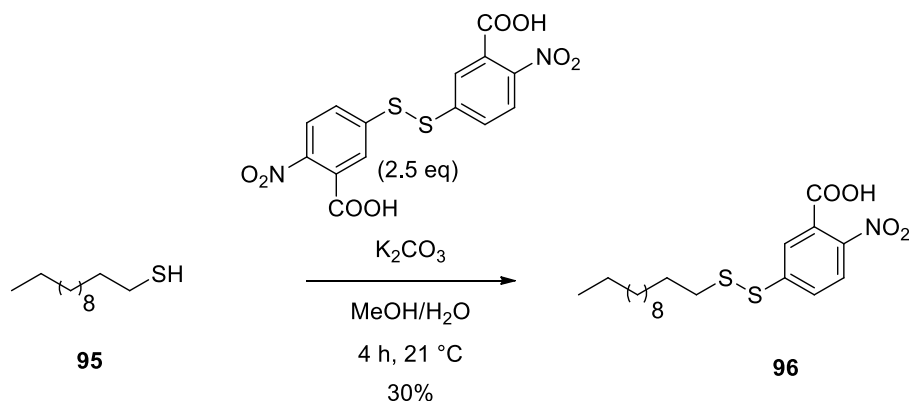
Pathway A

To probe the proposed reaction pathway A (*Scheme 64*), dodecanethiol was chosen as the alkyl component of the model disulfide. The lack of an acidic α -proton blocked reaction pathway B (*Scheme 66*) and therefore allowed exclusive analysis of reaction pathway A. High boiling point of dodecanethiol also simplified isolation of products. To prove that the reaction indeed proceeded through a thiophosphonium intermediate, step 2 was sought to be inhibited by employing a suitable alkyl-aryl disulfide (*Scheme 66*). The aryl thiol component of the disulfide was chosen to be 5-mercapto-2-nitrobenzoic due to its widespread use in biochemistry through Ellman's reagent.



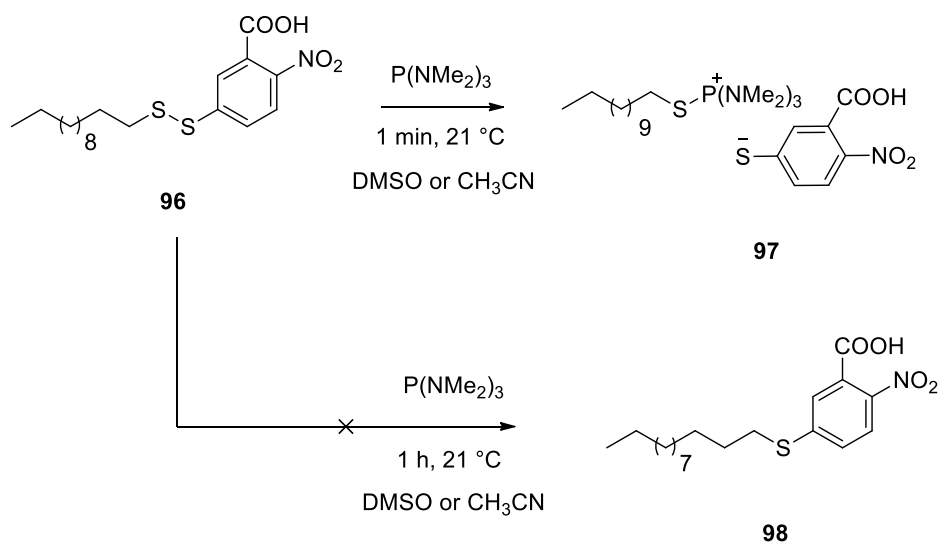
Scheme 66. Formation of stable thiophosphonium intermediate

Disulfide **96** was synthesized in 30% yield by the reaction of dodecanethiol **95** with excess Ellman's reagent in the presence of potassium carbonate as a base (*Scheme 67*).



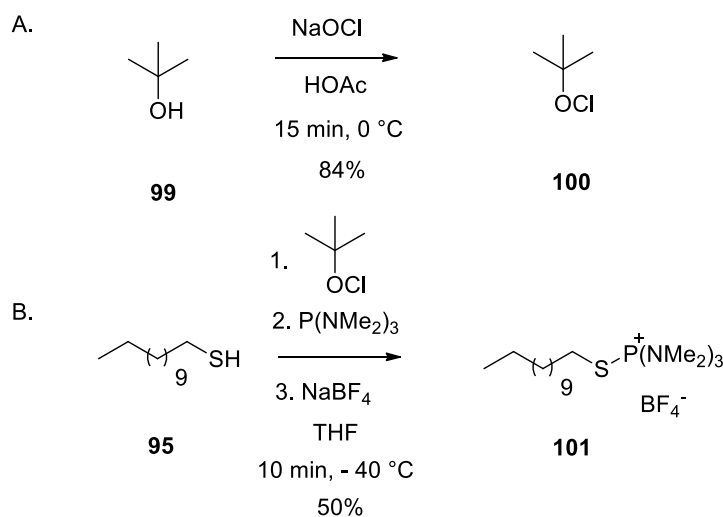
Scheme 67. Synthesis of disulfide 96

The reaction of disulfide **96** with tris(dimethylamino)phosphine **68** at 21 °C (in DMSO or acetonitrile) led to an instantaneous colour change of the reaction mixture from colourless to red, indicating release of 5-mercapto-2-nitrobenzoic acid (*Scheme 68*). Unlike with alkyl-alkyl disulfide substrates,¹⁵⁴ no thioether **98** was isolated or detected in the reaction mixture after 1 h at 21 °C. The reaction indicated that a suitably designed alkyl-aryl disulfide system can inhibit displacement of thiophosphonium species by the released thiol (Pathway A, Step 2, *Scheme 64*).



*Scheme 68. Reaction of disulfide **96** with tris(dimethylamino)phosphine **68** leads to formation of thiophosphonium **97** but no thioether **98***

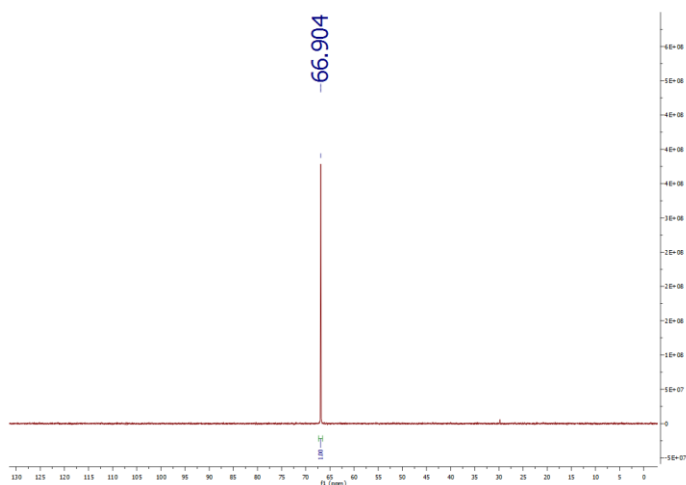
Successful inhibition of reaction pathway A provided an excellent opportunity for potentially detecting thiophosphonium **97** in the reaction mixture. The reaction mixture was probed for evidence of **97** using ^{31}P NMR spectroscopy after 1 min at 21 °C. The ^{31}P NMR spectrum of the reaction mixture was compared against a pure sample of dodecanethiophosphonium tetrafluoroborate **101** that was synthesized via an alternative route (Scheme 69).¹⁵⁷



*Scheme 69. A) Synthesis of t-butyl hypochlorite B) Synthesis of dodecane thiophosphonium tetrafluoroborate salt **101***¹⁵⁷

The reaction mixture positively indicated the presence of a thiophosphonium species **97** (Figure 9) (δ 66.9 ppm). However, the product was not exclusive, with a number of other phosphorus species including tris(dimethylamino)phosphine **68** (δ 123.2 ppm), hexamethylphosphoramide (δ 24.1 ppm) and related oxidation products (δ 13.9 ppm and δ 8.8 ppm) also being observed. An unknown peak (δ 62.5 ppm), was also observed close to the thiophosphonium species which will be discussed later in the chapter.

A.



B.

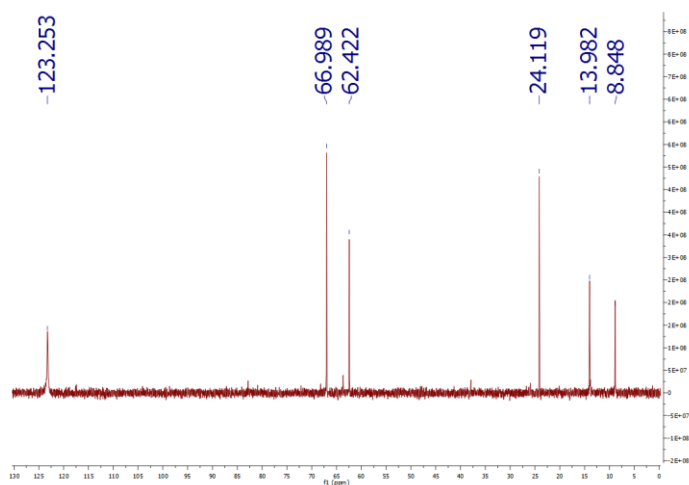
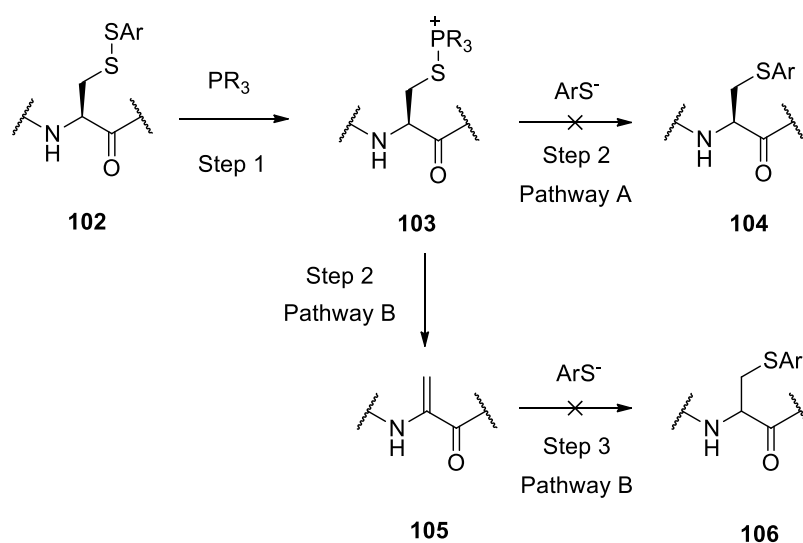


Figure 9. ^{31}P NMR spectra A) Thiophosphonium **97** B) Thiophosphonium **101**

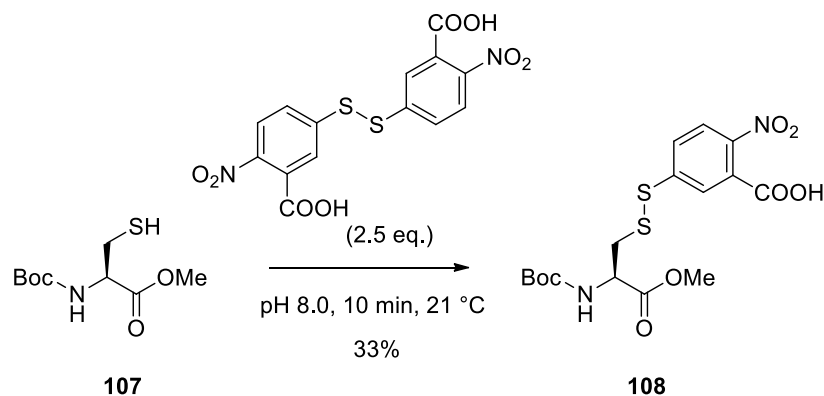
Pathway B

To probe proposed reaction pathway B (*Scheme 64*), protected cysteine was chosen as the alkyl component of the model disulfide. The presence of an acidic α -proton is likely to cause the reaction to proceed *via* pathway B, as previously reported.¹⁵⁵ To prevent the formation of thioether **106** (Step 3, Pathway B), alkyl-aryl disulfide with 5-mercapto-2-nitrobenzoic acid as the aryl thiol component was employed (*Scheme 70*).



Scheme 70. Reaction pathway for reaction of disulfide 102 with tris(dimethylamino)phosphine 68

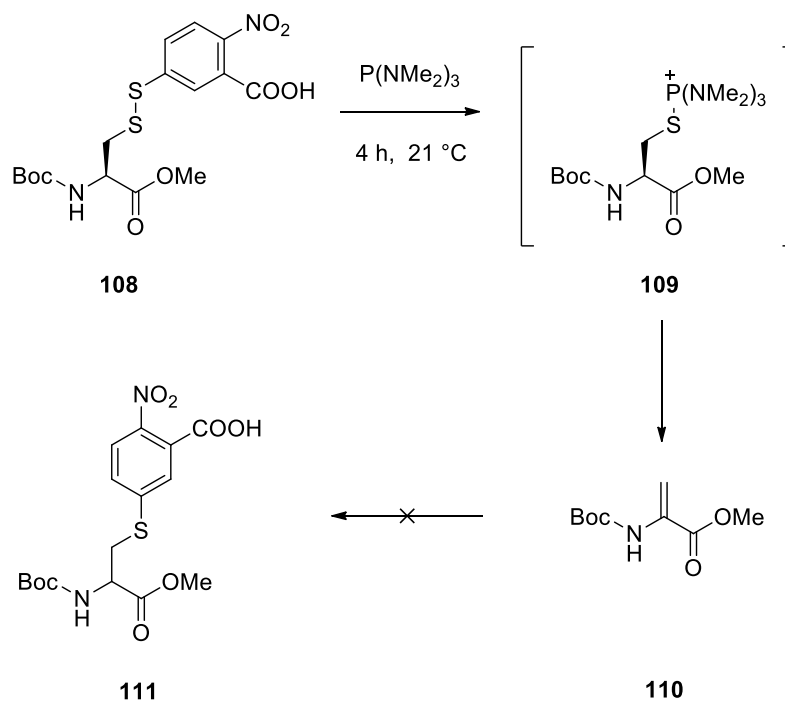
Disulfide **108** was synthesized by reaction of cysteine **107** with excess Ellman's reagent in buffer (phosphate buffer, 100 mM, pH 8.0) in a modest yield of 33% (*Scheme 71*). Excess Ellman's reagent was required to minimize the reaction of unreacted cysteine **107** with product disulfide **108**.



Scheme 71. Synthesis of model disulfide 108

The treatment of disulfide **108** with tris(dimethylamino)phosphine **68** under basic conditions (CHES/MES buffer, 70 mM, pH 9.5) yielded dehydroalanine **110** in a respectable 45% yield (*Scheme 72*), indicating that the reaction proceeds *via* pathway B as expected. The poor nucleophilicity of the released aryl thiolate obviates step 3, and thus no thioether **111** was detected in the reaction mixture (*Scheme 72*). ^{31}P NMR indicated no evidence of thiophosphonium species.

To isolate thiophosphonium **109**, Step 2 (Pathway B, *Scheme 70*) was sought to be blocked by control of pH to prevent elimination of thiophosphonium **109** to dehydroalanine **110** (*Scheme 72*). The reaction of disulfide **108** with tris(dimethylamino)phosphine **68** was repeated at lower pH (phosphate buffer, 100 mM, pH 8.0). Unfortunately, however, no trace of thiophosphonium **109** could be observed or isolated, and dehydroalanine **110** was isolated yet again in 50% yield (*Scheme 72*). Various attempts to prevent elimination of thiophosphonium **109** to dehydroalanine **110** under different pHs (5.0, 6.0, 7.0) failed. Surprisingly, pH measurement of the reaction mixture subsequent to addition of tris(dimethylamino)phosphine **68** indicated that the reactions were under highly basic conditions (pH 14). At the concentration required for small molecule reactions, buffer strength was found to be insufficient for these experiments, hence no evidence of thiophosphonium **109** could be obtained.



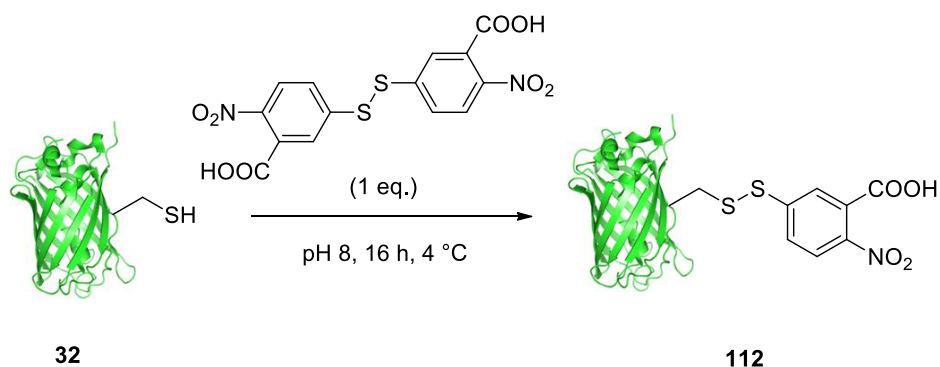
Scheme 72. Conversion of disulfide 108 to dehydroalanine 110

3.4 Protein model system

Unlike small molecule reactions, protein bioconjugation reactions are carried out at much lower (micromolar) concentrations and therefore afford greater control over pH variations. The formation of thiophosphonium was hence analysed on a protein model system. Single cysteine mutant GFP (S147C) **32** (Section 2.3) was used as a model system to investigate the reactivity of a disulfide with tris(dimethylamino)phosphine **68** and explore thiophosphonium formation on a protein. The reaction was unlikely to proceed *via* Pathway B (*Scheme 64*), i.e., elimination of thiophosphonium to dehydroalanine, due to the presence of a solvent inaccessible α -proton as observed in Chapter 2 in case of the strong resistance of activated sulfonium at position 147 to undergo β -elimination. As with the small molecule system, 5-mercapto-2-nitrobenzoic acid was chosen as the aryl component of the the disulfide.

GFP (S147C, 5-mercapto-2-nitrobenzoic acid) disulfide **112** was synthesized by treatment of GFP (S147C) **32** with a single equivalent of Ellman's reagent under basic conditions (phosphate buffer, 100 mM, pH 8.0) for 16 h at 4 °C. LCMS

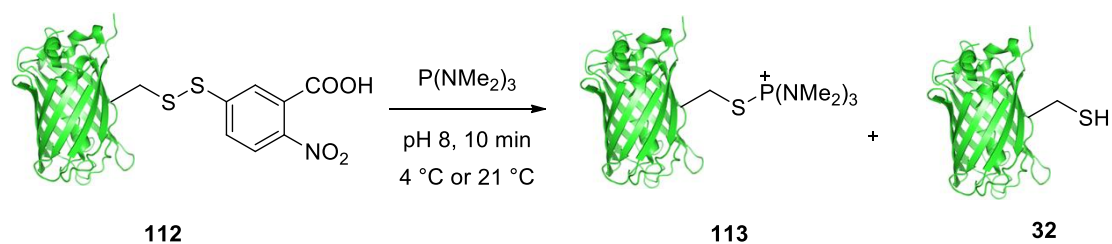
analysis indicated complete formation of GFP (S147C, 5-mercapto-2-nitrobenzoic acid) disulfide **112** (Scheme 73).



Scheme 73. Reaction of GFP (S147C) **32** with Ellman's reagent

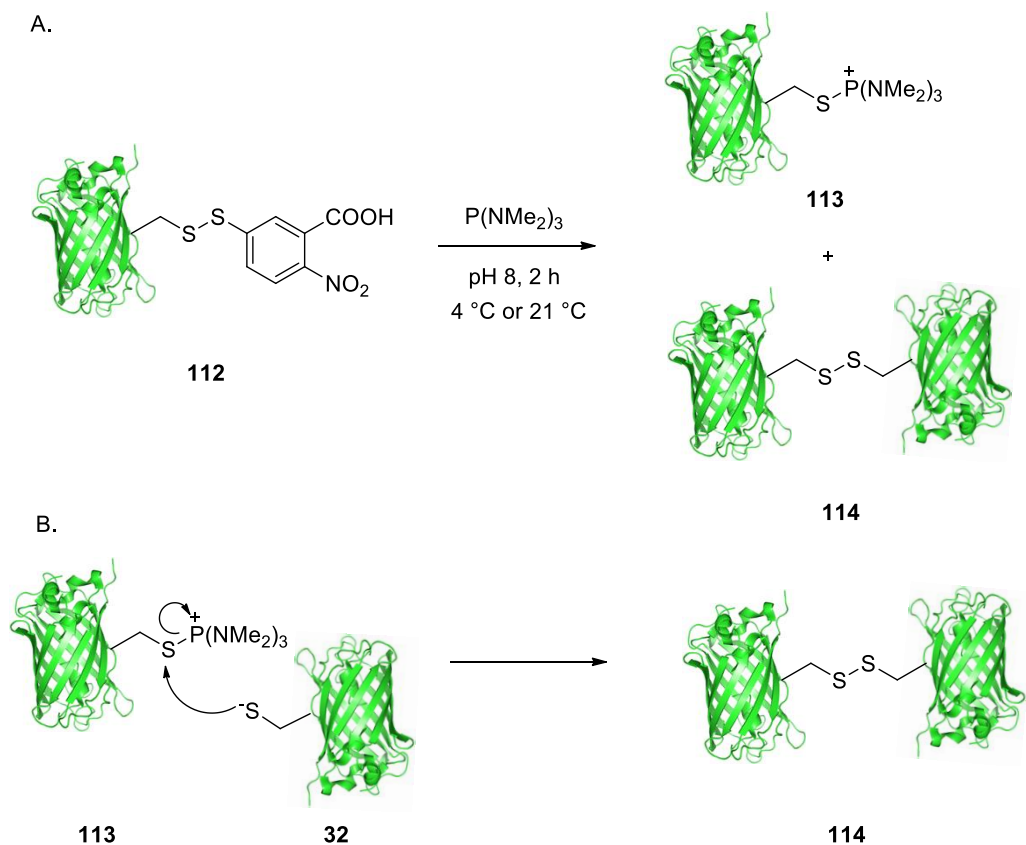
GFP (S147C, 5-mercapto-2-nitrobenzoic acid) disulfide **112** (1 mg/ml, phosphate buffer, 100 mM, pH 8.0) was then treated with varying equivalents of tris(dimethylamino)phosphine **68** (1, 10, 25 or 50 eq.), but LCMS indicated no reaction after incubation at 21 °C for 16 h. Increasing the number of molar equivalents of **68** (100 and 125 eq.) provided some evidence of GFP (S147C, thiophosphonium) **113**; however, simultaneous and rapid degradation of the protein sample was also observed, evidenced by fragmentation and poor signal in LCMS. Increasing the molar equivalents further (200 eq.) led to almost complete degradation of the protein and no detection of any protein sample was observed LCMS.

Interestingly, increasing the concentration of the protein GFP (S147C, 5-mercapto-2-nitrobenzoic acid) disulfide **112** (5 mg/ml, phosphate buffer, 100 mM, pH 8.0) led the reaction to proceed successfully, indicating presence of GFP (S147C, thiophosphonium) **113** (65% yield by LCMS) after incubation with **68** (40, 60 and 80 equivalents) at either 4 °C or 21 °C for 10 min (Scheme 74). In addition to thiophosphonium **113**, GFP (S147C) **32** was also detected. No significant difference in yield was observed when 40, 60 or 80 equivalents of tris(dimethylamino)phosphine **68** were used.



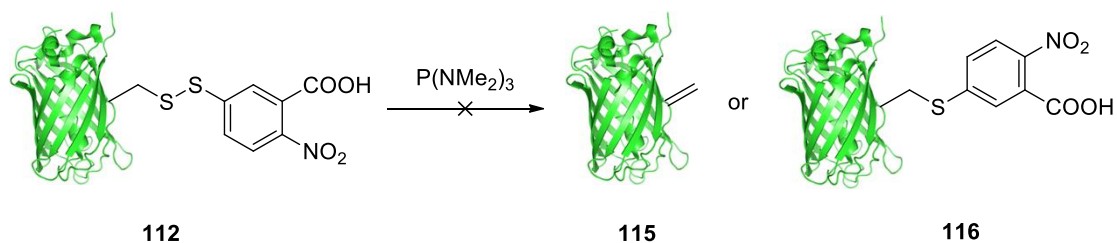
*Scheme 74. Reaction of GFP (S147C, 5-mercapto-2-nitrobenzoic acid) disulfide **112** with tris(dimethylamino)phosphine **68***

Furthermore, the LCMS analysis of the reaction mixture after a 2 h period indicated complete consumption of GFP (S147C) **32** and the formation of GFP (S147C) disulfide dimer **114** (*Scheme 75A*). No formation of dimer **114** was detected by incubating GFP (S147C) **32** in the absence of **68** under the reaction conditions (phosphate buffer, 100 mM, pH 8.0), indicating that the formation of GFP (S147C) disulfide dimer **114** proceeded *via* the reaction of GFP (S147C) **32** with GFP (S147C, thiophosphonium) **113** (*Scheme 75B*).



Scheme 75. A) Reaction of GFP (S147C, 5-mercapto-2-nitrobenzoic acid) disulfide 112 with tris(dimethylamino)phosphine 68 after 2 h B) Mechanism of formation of GFP (S147C) disulfide dimer 114

Significantly, no GFP (S147C, Dha) 115 or GFP (S147, 5-mercapto-2-nitrobenzoic acid) thioether 116 was observed (*Scheme 76*) indicating that step 2 of both reaction pathways A and B were inhibited (*Scheme 64*).



Scheme 76. Reaction of GFP (S147C, 5-mercapto-2-nitrobenzoic acid) disulfide 112 with tris(dimethylamino)phosphine 68

3.5 Thiophosphonium stability

The stability of thiophosphonium species under aqueous conditions was critical for its applicability as a reactive intermediate for protein bioconjugation. Tris(2-carboxyethyl)phosphine relies on rapid hydrolysis of thiophosphonium formed on reaction with a disulfide for efficient reduction to free cysteine.

Interestingly, ^{31}P NMR spectroscopy in D_2O revealed that thiophosphonium species **97** and dodecanethiophosphonium tetrafluoroborate **101** were strongly resistant to hydrolysis with no change being observed after 24 h at 21 °C, perhaps indicating that the thiophosphonium is stabilized by the three amino group ligands on phosphorus. GFP (S147C, thiophosphonium) **113** exhibited similar stability, and could be detected after 24 h following storage at 4 °C. Under similar conditions, however, tris(dimethylamino)phosphine **68** degraded rapidly, indicating that the phosphine species was very short lived in an aqueous environment, thus explaining the need for a large excess of tris(dimethylamino)phosphine **68** in the reaction with model disulfides **96**, **108** and **112**.

3.6 Mechanistic insights

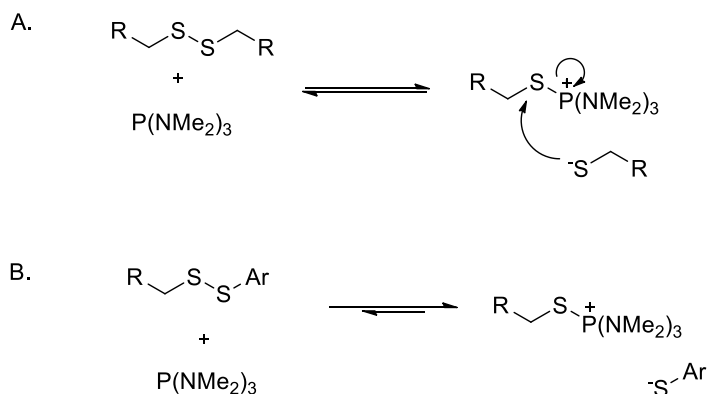
The aqueous stability of the thiophosphonium species indicated its suitability for protein functionalisation. The poor aqueous stability of tris(dimethylamino)phosphine **68**, however, made it difficult to access the thiophosphonium species. Thus, aqueous conditions often necessitated the use of excess tris(dimethylamino)phosphine **68**, leading to a rapid increase in pH of the reaction mixture and therefore making it unsuitable for general protein modification.

Furthermore, thiophosphonium **97** and GFP (S147C, thiophosphonium) **113** were often accompanied by a number of side products including GFP (S147C) **32** and GFP (S147C) disulfide dimer **114** respectively.

In addition to *Scheme 62*, the following reaction mechanism parameters can be used to account for the observed reaction outcomes:

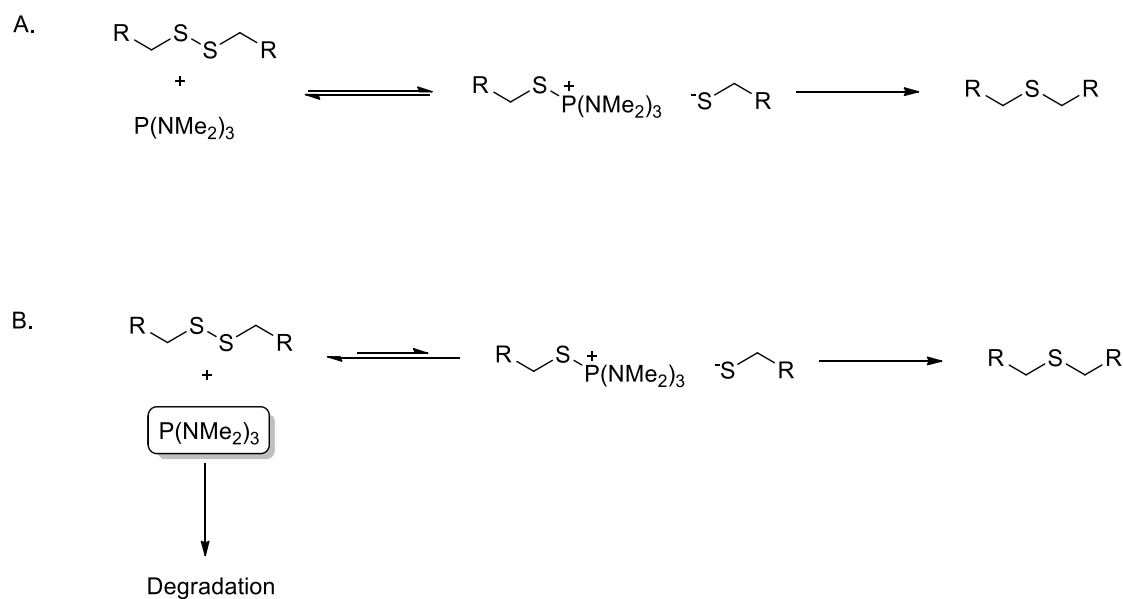
1. Reversibility: The reaction of GFP (S147C) **32** with GFP (S147C, thiophosphonium) **113** to form GFP (S147C) disulfide dimer **114** (*Scheme*

75B) indicated that step 1, i.e. reaction of tris(dimethylamino)phosphine **68** with disulfides, involves reversible formation thiophosphonium intermediate and a free thiol (*Scheme 62*). To achieve a greater concentration of thiophosphonium, it was essential that the thiol released from step 1 is a poor nucleophile (aromatic thiol) (*Scheme 77*).



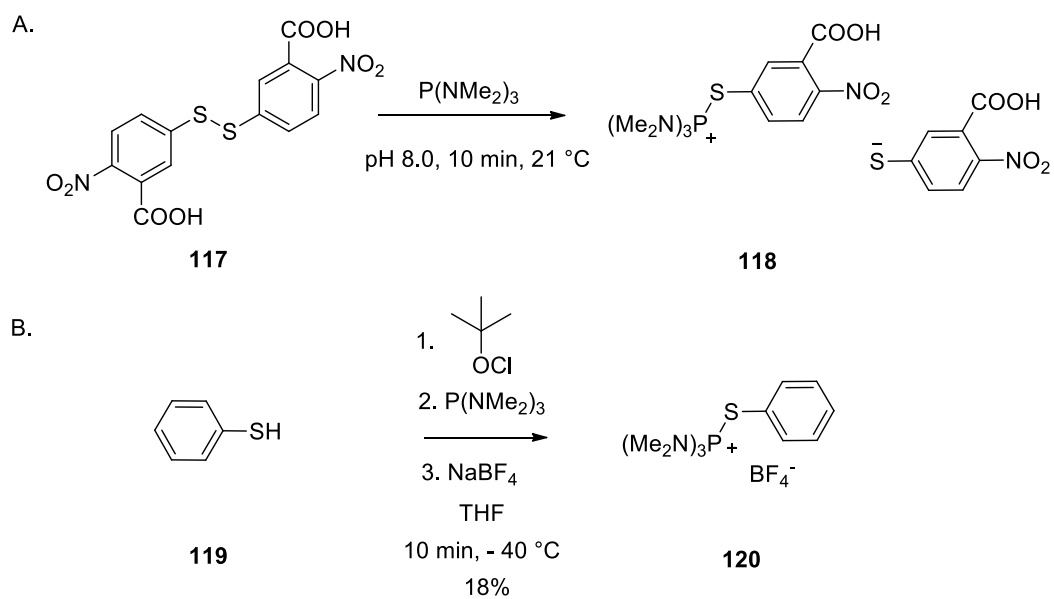
Scheme 77. Reaction of tris(dimethylamino)phosphine with disulfide

2. Stability: While thiophosphonium species exhibit excellent stability, tris(dimethylamino)phosphine **68** degrades rapidly under aqueous conditions. For alkyl-alkyl disulfides, under non-aqueous conditions, step 1 of the reaction with **68** is under equilibrium until thiophosphonium gets trapped as the irreversible thioether product. Under aqueous conditions however, the rapid degradation of **68** shifts the equilibrium to the left, leading to a lower concentration of thiophosphonium available for displacement. For alkyl aryl disulfides, this has limited effect, provided that no free alkyl thiols are present in the system.



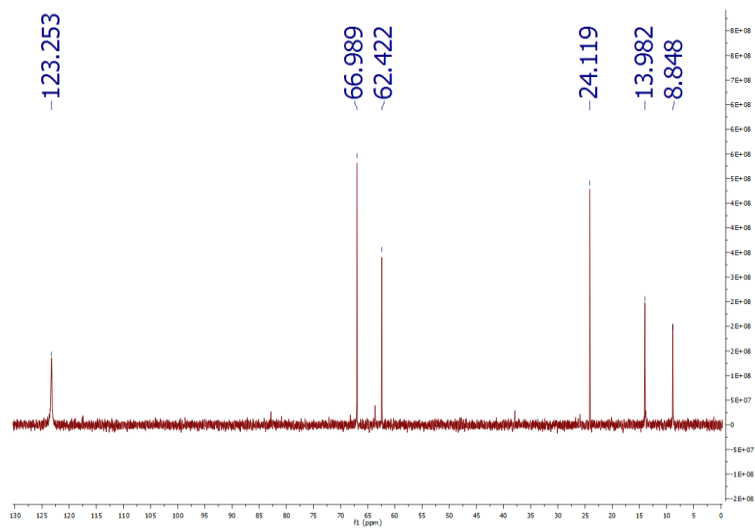
Scheme 78. Reaction pathway of tris(dimethylamino)phosphine 68 with disulfide A) Non-aqueous environment B) Aqueous environment

- Position of attack: The formation of GFP (S147C) **32** in the reaction of GFP (S147C, 5-mercapto-2-nitrobenzoic acid) disulfide **112** with tris(dimethylamino)phosphine **68** indicates that the attack of phosphine does not occur exclusively on the alkyl sulfur. In retrospect, the unidentified species (δ 62.4 ppm) observed in the reaction between disulfide **96** and **68** (*Scheme 67B*) was thought to be arylthiophosphonium **121** (*Scheme 80B*). To confirm that this unidentified species was indeed arylthiophosphonium **121**, Ellman's disulfide was treated with tris(dimethylamino)phosphine **68** and incubated for 10 min at 21 °C (*Scheme 79A*). ^{31}P NMR of the reaction mixture indicated the formation of the major peak very close to the unidentified species at δ 61.6 ppm. In absence of any ^{31}P NMR data for arylthiophosphonium species in the literature, further evidence was obtained about the unidentified ^{31}P NMR signal by synthesizing benzenethiophosphonium salt **120** of the commercially available thiophenol *via* alternative pathway¹⁵⁷ which gave ^{31}P NMR signal in a similar region (δ 61.9 ppm) (*Scheme 79B*).

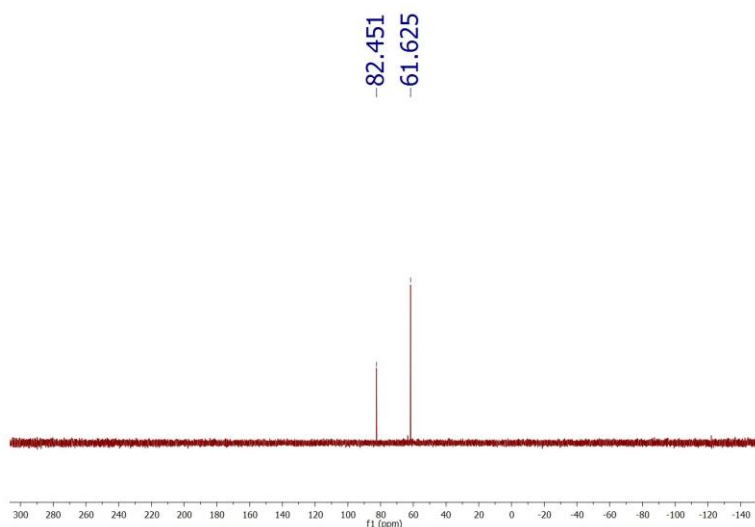


Scheme 79. A) Reaction of Ellman's reagent with tris(dimethylamino)phosphine **68**
 B) Synthesis of benzenethiophosphonium tetrafluoroborate **120**

A.



B.



C.

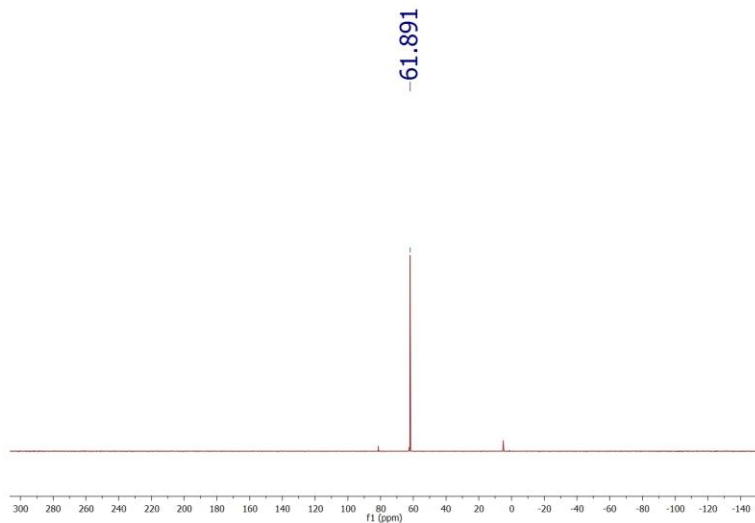
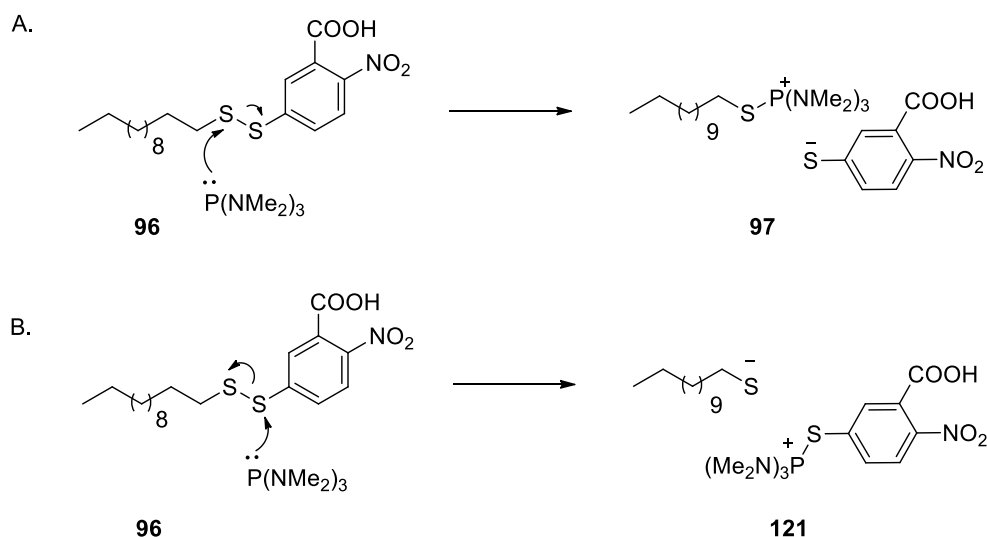


Figure 10. ^{31}P NMR spectra for A) Thiophosphonium **97** B) Reaction of Ellman's disulfide with **68** C) Benzenethiophosphonium tetrafluoroborate **120**

Whilst previous literature reports indicate the preference of the phosphine attack on the sulfur that results in the release of the most stable thiolate, and not on the most electropositive sulfur atom, the detection of two different thiophosphonium species by ^{31}P NMR indicated that the reaction proceeds *via* attack at both sulfur atoms (*Scheme 80*). In fact, a subsequently released publication indicated a similar mode of reactivity.⁷⁰

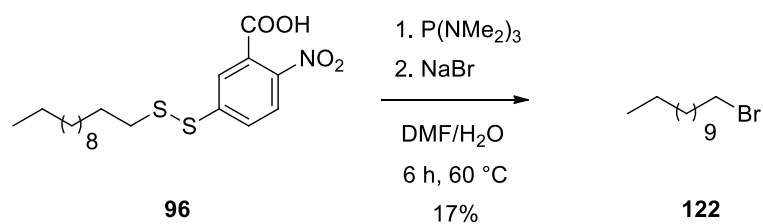


Scheme 80. Possible reaction pathways – attack of tris(dimethylamino)phosphine 68 on disulfide 96

This leads to the release of free alkyl thiol in the reaction mixture, and the amount of available thiophosphonium is reduced by the stability effect discussed in the previous section.

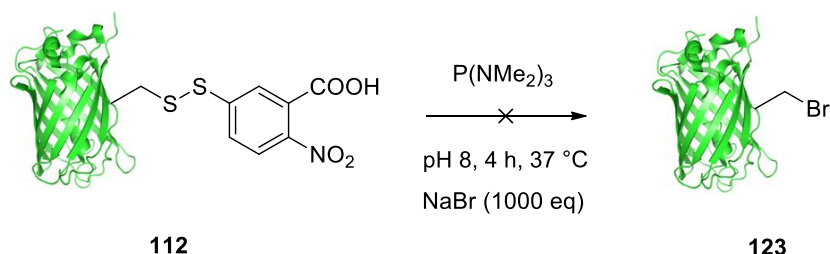
3.6.1 Thiophosphonium reactivity with nucleophiles

Due to the inability to isolate any thiophosphonium species cleanly, displacement reaction with nucleophiles (*Scheme 63*) was likely to proceed sub-optimally. Despite this, displacement of thiophosphonium **97** with bromide as nucleophile was attempted as a proof of principle reaction. Disulfide **96** was treated with tris(dimethylamino)phosphine **68**, followed by excess sodium bromide under aqueous conditions in a one-pot reaction. Following incubation of the reaction at $60\text{ }^\circ\text{C}$ for 6 h dodecanebromide **122** was successfully isolated in 17% yield (*Scheme 81*).



*Scheme 81. One pot conversion of disulfide **96** to dodecanebromide **122***

This reaction demonstrated the proof of concept of converting a disulfide into a stable thiophosphonium species which could undergo a displacement reaction in aqueous conditions, albeit in poor yield. The displacement of thiophosphonium with bromide was then attempted on the protein system. GFP (S147C, 5-mercapto-2-nitrobenzoic acid) disulfide **112** (1 mg/ml, phosphate buffer, 100 mM, pH 8.0) was incubated with tris(dimethylamino)phosphine **68** (60 equivalents, in 5 μ L DMF) for 2 hours at 21 $^\circ$ C, followed by incubation with sodium bromide (1000 equivalents, 5 μ L in water) for 4 h at 37 $^\circ$ C. Unfortunately, no GFP (S147C, bromoalanine) **123** could be detected, and a number of unidentifiable products were detected by LCMS (*Scheme 82*).



*Scheme 82. Reaction of GFP (S147C, 5-mercapto-2-nitrobenzoic acid) disulfide **112** with tris(dimethylamino)phosphine **68** and sodium bromide*

3.7 Conclusion

A novel method for protein functionalisation *via* a thiophosphonium species was evaluated. The reaction of disulfides with tris(dimethylamino)phosphine forming thioethers was investigated as a platform to provide access to thiophosphonium species.

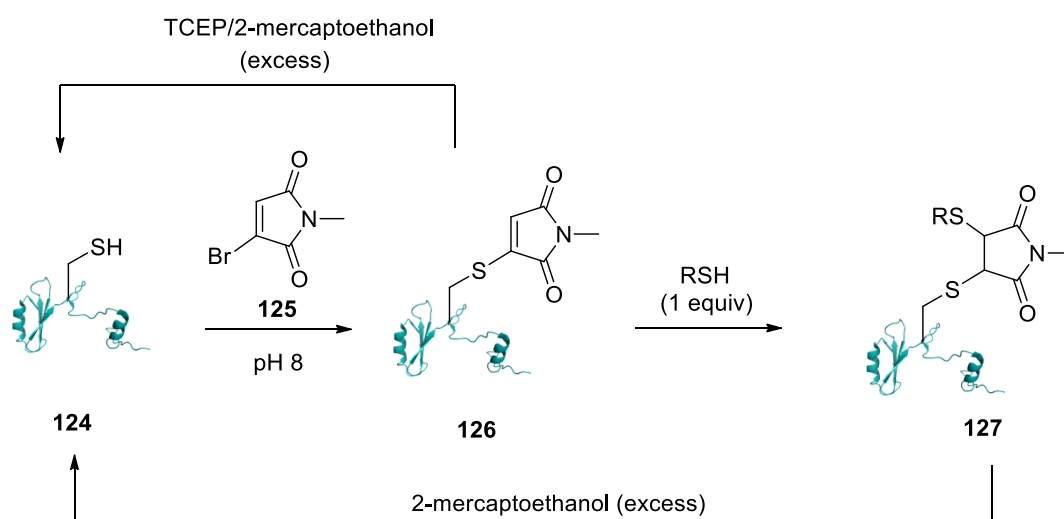
Novel strategies to isolate thiophosphonium species were designed and evaluated on model small molecule and protein systems. Alkyl-aryl disulfides were found to provide facile access to thiophosphonium species as a reactive intermediate, albeit in a non-selective manner. The reaction of tris(dimethylamino)phosphine with unsymmetrical alkyl-aryl disulfides was observed to be preferential, but not completely selective for attack on the least electrophilic sulfur atom.

The stability and reactivity of the thiophosphonium intermediate was studied on small molecule and protein model systems. Thiophosphonium species was found to be stable for at least 24 h under slightly basic, aqueous conditions. The viability of thiophosphonium-derived functionalisation, including the various challenges involved in synthesising and functionalising thiophosphonium intermediates was also investigated. Whilst the thiophosphonium species has potential as a platform for protein functionalisation, the difficulty in accessing the thiophosphonium species in a selective and high yielding manner presents a major challenge.

Chapter 4 Reversible cysteine modification using bromomaleimides

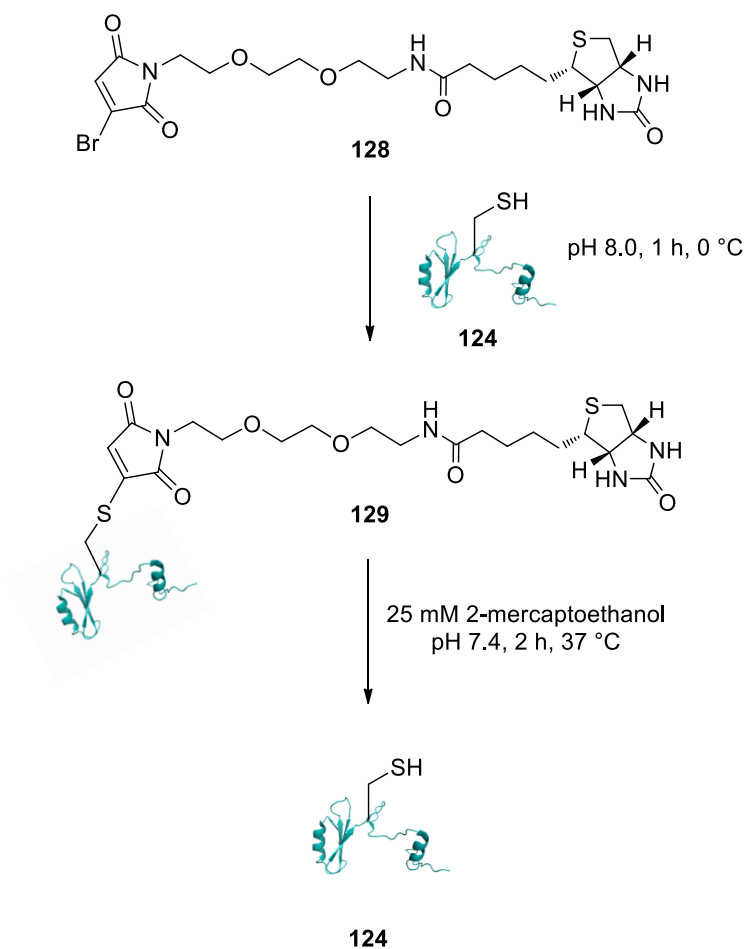
4.1 Introduction

Bromomaleimides represent a novel class of reagents that react rapidly and selectively with cysteine residues on proteins and peptides to afford thiomaleimides (*Scheme 83*). In a reducing environment, such as an excess of tris(2-carboxyethyl)phosphine (TCEP) or 2-mercaptoethanol, thiomaleimides can be converted back to free thiols (*Scheme 83*).⁴⁷



Scheme 83. Reversibility of thiomaleimides 126 and dithiosuccinimides 127

This quantitative, reversible protein modification is of general interest for applications in chemical biology and proteomics. Previous work in this laboratory has demonstrated the utility of the thiomaleimide reversibility by introducing a bromomaleimide-based biotin affinity tag **128** onto a single cysteine mutant Grb2-SH2 domain (L111C) **124** and validating the concept of pull down experiments using neutravidin beads (*Scheme 84*).¹⁵⁸



Scheme 84. Cleavable biotinylation of Grb2-SH2 domain (L111C) **124** using thiomaleimides¹⁵⁸

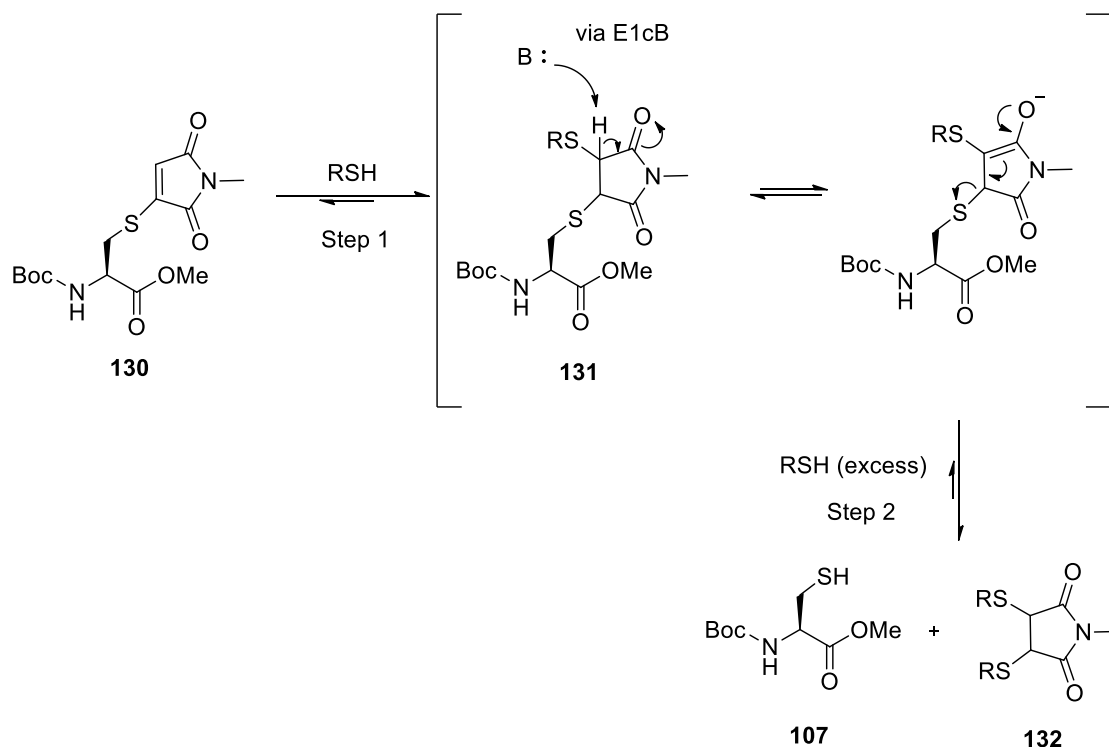
4.2 Aims

Despite the utility of reversible protein modification of cysteine *via* thiomaleimide, their application is restricted by the requirement for a large excess of reducing agent to facilitate reversibility. These conditions are likely to be incompatible with disulfide bonds which play a significant role in maintaining the structural integrity of proteins.¹⁵⁹ Employment of reducing conditions to affect reversibility could also lead to unfavourable protein folding, aggregation or disulfide scrambling.¹⁶⁰ The use of such harsh conditions for cleaving thiomaleimide to free cysteine poses a challenge to the general applicability of the thiomaleimide reversibility approach. Thus, the development of efficient, selective and milder reversibility conditions would represent a significant improvement over the existing methodology.

Mechanistically, the reversibility of thiomaleimides with excess thiols can be explained by the proposed mechanism outlined below (*Scheme 85*):

Step 1: Reaction of thiomaleimide **130** with a single equivalent of thiol proceeds rapidly to form the dithiosuccinimide **131**.⁴⁸ In an environment of excess thiol, however, dithiosuccinimides have been reported to be unstable, indicating that the process is under equilibrium.⁴⁸ It is proposed that the dithiosuccinimides are unlikely to be stable because of the low pK_a of the succinimide proton. The reverse reaction can proceed *via* the elimination of the acidic succinimide proton under mildly basic conditions.

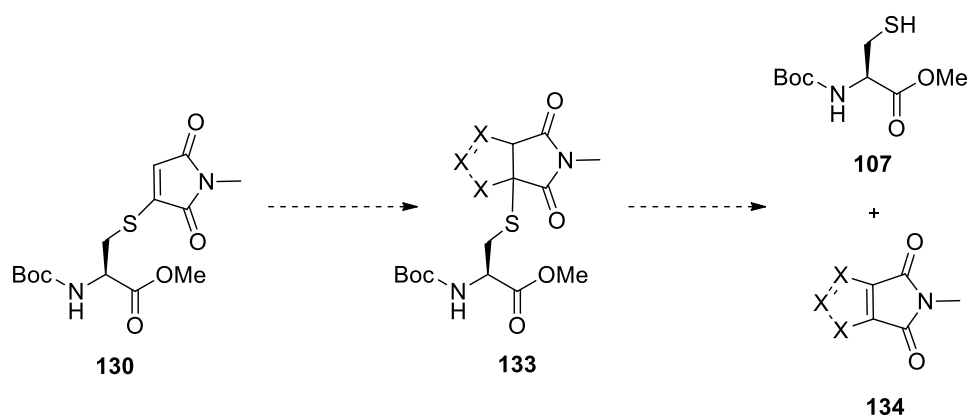
Step 2: In the presence of an excess thiol, it is proposed that unstable dithiosuccinimide **131** undergoes thiol exchange. The reaction is proposed to proceed *via* elimination of acidic succinimide proton releasing the free cysteine **107** and formation of the dithiosuccinimide **132**. An excess of thiol thus pushes the equilibrium to the right (*Scheme 79*).



Scheme 85. Proposed mechanism for reversibility of thiomaleimide 130

Based on this understanding of the reaction mechanism, three distinct approaches were designed for the development of an improved thiomaleimide reversibility approach:

1. Nucleophile screen: A significant drawback of the current methodology is use of disulfide reactive reagents such as 2-mercaptoethanol or TCEP. Reagents which do not react with disulfide bonds, but can convert thiomaleimide to free cysteine would represent a significant improvement over the existing methodology. A number of nucleophiles, including non-thiol nucleophiles such as carbon nucleophiles, amines and azides were to be evaluated as potential reagents for conversion of thiomaleimide to free cysteine. It is proposed that these nucleophiles may offer differential reactivity patterns towards thiomaleimides and disulfides which could be exploited for the development of an improved reversibility protocol.
2. Cycloaddition reactions: The true potential of thiomaleimide based cysteine modification can be realised with the development of a reversibility protocol that is orthogonal to disulfide reduction. As maleimides are excellent cycloaddition partners that undergo cycloaddition reactions with 1,3-dipoles including azides¹⁶¹ and nitrile oxides,¹⁶² this affords an opportunity for the development of a thiomaleimide reversibility protocol. It is proposed that cycloaddition reaction of 1,3-dipoles with thiomaleimide might lead to the formation of an intermediate cycloadduct **133** that could undergo rapid aromatisation releasing free cysteine **107** (*Scheme 86*). This approach was expected to be orthogonal to disulfide reduction and likely to react exclusively with thiomaleimide.



Scheme 86. Proposed route for conversion of thiomaleimides to cysteine using [3+2] cycloaddition reaction

3. Tether reagents: The optimisation of the existing protocol for conversion of thiomaleimide to free cysteine by decreasing the amount of reducing agent would also represent a significant improvement and lead to enhanced applicability of this approach, especially for protein systems with sensitive disulfide bonds. The proposed mechanistic considerations indicate that the pK_a of the dithiosuccinimide protons play a significant role in the reversibility of thiomaleimides to free thiols (*Scheme 85*). The control of the equilibrium in step 2 of the proposed mechanism is vital to developing an improved cleavage approach (see step 2 - *Scheme 85*). It is therefore proposed that thiosuccinimide conjugates lacking acidic proton would be stable to any further transformation by thiols. Thus, a tether approach is envisaged using bidentate thiol nucleophiles which could form spiro-cyclic adducts **136** (*Figure 11*) by trapping the reactive dithiosuccinimides into a stable, unreactive adduct.

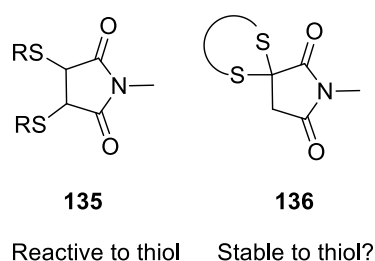


Figure 11. Reactivity of thiosuccinimides

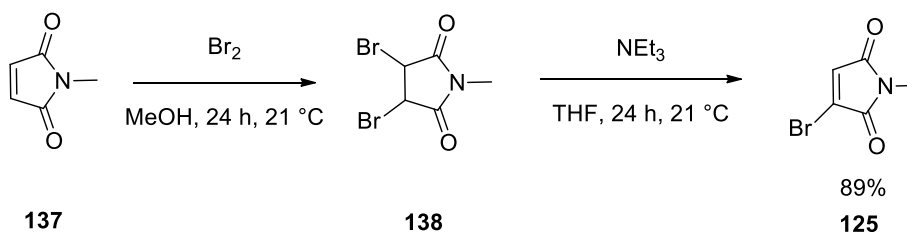
The following aims were identified:

1. To design and develop a small molecule and a protein model system to investigate the conversion of thiomaleimide to free thiol;
2. To screen non-thiol nucleophiles for reactivity with thiomaleimide to afford free thiol;
3. To develop a cycloaddition strategy for an orthogonal thiomaleimide reversibility protocol.
4. To develop a tether-type strategy to improve the existing thiomaleimide reversibility protocol;

4.3 Model system

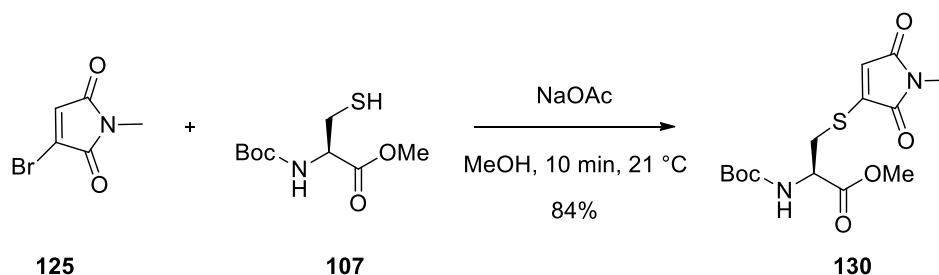
4.3.1 Small molecule model system

Protected cysteine derivative **107** was chosen as a small molecule model system (*Scheme 88*). Its solubility in organic and aqueous solvents allowed simple isolation and purification by standard chromatographic techniques. Bromo-*N*-methylmaleimide **125** was chosen as a model bromomaleimide reagent for evaluation. It was synthesized by bromination of *N*-methylmaleimide followed by elimination of HBr in 89% yield (*Scheme 87*).⁴⁷



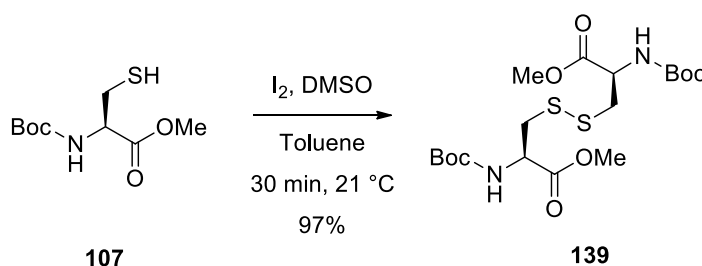
Scheme 87. Synthesis of bromo-N-methylmaleimide 125

Small molecule model thiomaleimide **130** was synthesized by the reaction of cysteine derivative **107** with bromo-*N*-methylmaleimide **125** in methanol with sodium acetate as base. The reaction was complete in 10 min affording thiomaleimide **130** in 84% yield (*Scheme 88*).



*Scheme 88. Reaction of bromo-*N*-methylmaleimide **125** with cysteine **107***

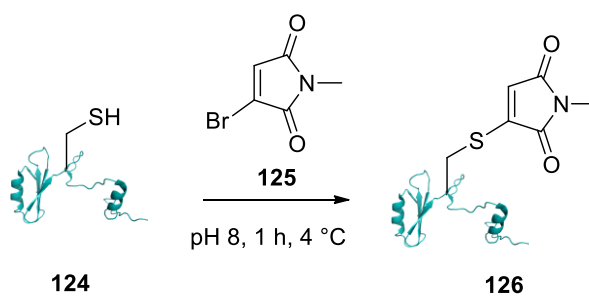
A model disulfide, cystine **139**, was synthesized from the corresponding cysteine **107** in 97% yield using iodine as oxidant (*Scheme 89*).¹⁶³



*Scheme 89. Cysteine **107** oxidation to cystine **139***

4.3.2 Protein model system

Previous investigations of bromomaleimide and thiomaleimide reactivity have been carried out using single cysteine mutant Grb2-SH2 Domain (L111C) **124**.⁴⁸ For consistency and ease of direct comparison of results, all investigations were carried out using the same protein. Thiomaleimide **126** was synthesized by treatment of Grb2-SH2 domain (L111C) **124** with a single equivalent of bromo-*N*-methylmaleimide **125** for 1 hour at 4 °C (*Scheme 90*).

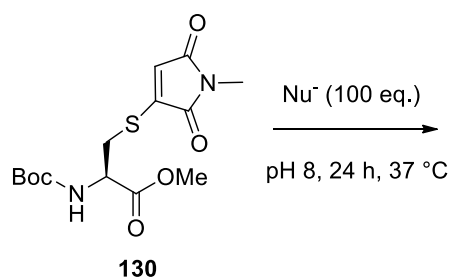


*Scheme 90. Reaction of Grb2-SH2 domain (L111C) **124** with bromo-N-methylmaleimide **125***

4.4 Reversibility of thiomaleimides

4.4.1 Thiomaleimide reversibility with nucleophiles

The reaction of a wide range of non-thiol based nucleophiles with thiomaleimides was investigated. Model thiomaleimide **130** was incubated with excess nucleophile (100 eq.) for 24 h at 37 °C under mildly basic conditions (phosphate buffer, 100 mM, pH 8.0) (*Table 3*) and the reaction was analysed by TLC for release of free cysteine **107**. DMF was used for solubilizing organic molecules into the buffer solution.



Entry	Nucleophile	Reaction outcome / Product
1	H ₂ N-OH 140	No reaction
2	 141	No reaction
3	 142	No reaction
4	 143	No reaction
5	 144	 145 64%
6	 146	No reaction

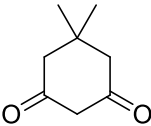
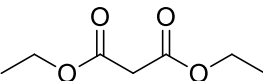
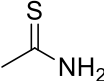
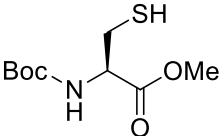
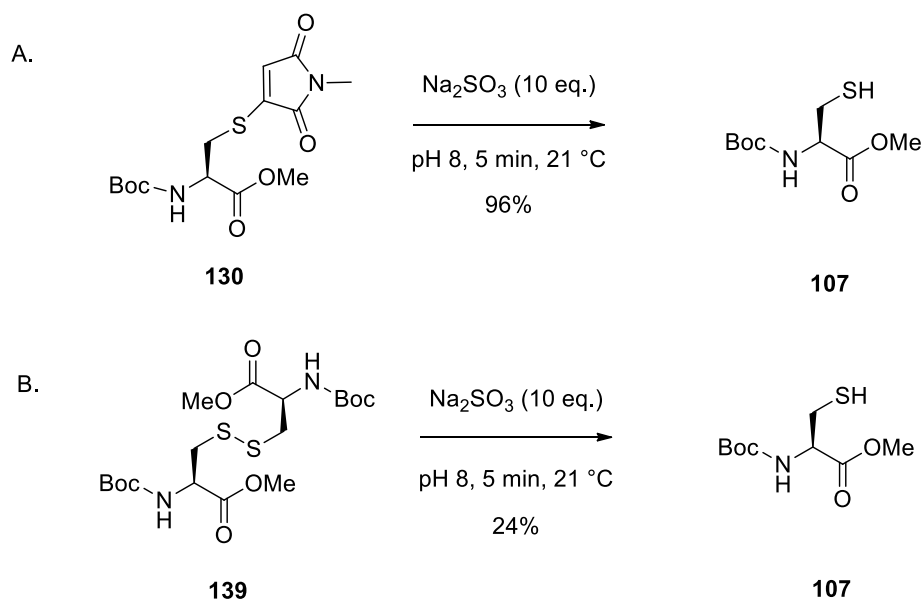
7		No reaction
	147	
8		No reaction
	148	
9		No reaction
	149	
10	NaN ₃	No reaction
11	Na ₂ SO ₃ ^a	
		107
		96%

Table 3. Reaction of thiomaleimide 130 with nucleophiles. ^a10 eq, 21 °C, 2h ^bCrude

Amines represented a class of reagents that are excellent nucleophiles and have not previously been reported to react with disulfides (*Table 3* – Entry 1-3). Unfortunately, however, no reaction was detected following incubation with thiomaleimide **130** with amines **140**, **141** and **142** after 24 hours at 37 °C.

Carbon-based nucleophiles such as nitroethane **143**, 2-nitropropane **144** and diethylmalonate **148** have been reported to undergo conjugate addition reactions owing to their low pKa.¹⁶⁴ Meldrum's acid **146** and Dimedone **147** are also known to have an unusually low pKa of 5.¹⁶⁵ While no reaction was observed on reaction of thiomaleimide **130** with most carbon nucleophiles, conjugate addition product of 2-nitropropane **145** was identified (*Table 3* – Entry 4-8) after incubation for 24 hours at 37 °C. Despite repeated attempts, this product could not be purified. Nonetheless, as no release of cysteine **107** was detected, the reaction of 2-nitropropane **144** with thiomaleimide **130** was not attempted further.

No reaction of thiomaleimide **130** with thioacetamide **149** or sodium azide was observed (*Table 3* – Entry 9-10). Interestingly, reaction of thiomaleimide **130** with sodium sulfite (10 equivalents) led to release of cysteine **107** in excellent yield (96%) after 2 h at 21 °C (*Table 3* – Entry 11). Reaction with sodium sulfite was very rapid, with starting material thiomaleimide **130** completely consumed after a matter of minutes. Sodium sulfite, however, has been reported to react with disulfide bonds.¹⁶⁶ The reaction of sodium sulfite was evaluated for any preferential reactivity towards thiomaleimide **130** over disulfide **139**.



Scheme 91. Reaction of sodium sulphite with A) thiomaleimide 130 B) disulfide 139

Comparative reaction of sodium sulfite with both substrates (*Scheme 91*) indicated some preferential reactivity for thiomaleimide **130** over disulfide **139** (*Table 4*). In the case of thiomaleimide **130**, the starting material was completely consumed with 96% of cysteine **107** being liberated within 5 min. The reaction was relatively slower with the disulfide **139**, with only 40% conversion of disulfide **139** after the same time period (*Table 4* - Entry 2). Although some preferential reactivity was clearly observed, any selectivity was not expected due to rapid rate of reaction in both cases.

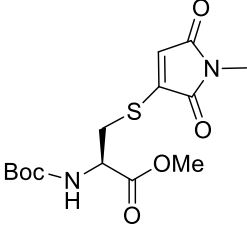
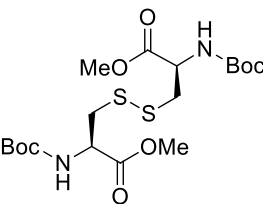
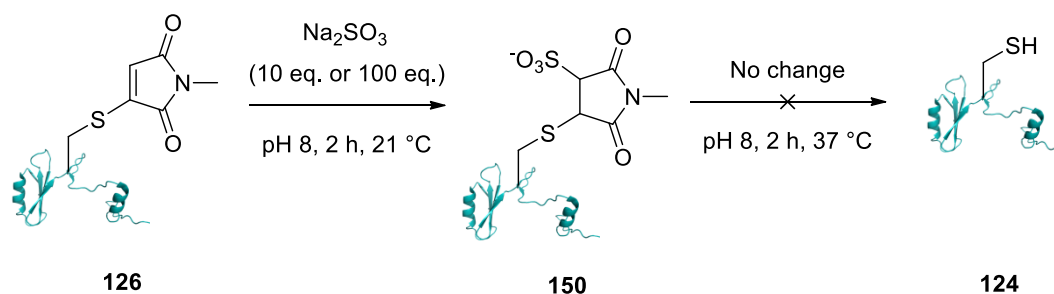
Reactant	Equivalents of Na ₂ SO ₃	Time (min)	Conversion	Yield
 130	10	5	100%	96%
 139	10	5	40%	24%

Table 4. Reaction of sodium sulfite with thiomaleimide **130** and disulfide **139**

The reaction of sodium sulfite with a Grb2-SH2 domain (L111C) derived thiomaleimide **126** was next explored. Thiomaleimide **126** was incubated with sodium sulfite under similarly basic conditions (phosphate buffer, 100 mM, pH 8.0). The only the product observed by LCMS was that resulting from the addition of sulfite to the thiomaleimide **150**. No cleavage of the thiomaleimide **126** to yield free thiol was observed, even after heating the conjugate for 2 h at 37 °C (Scheme 92).

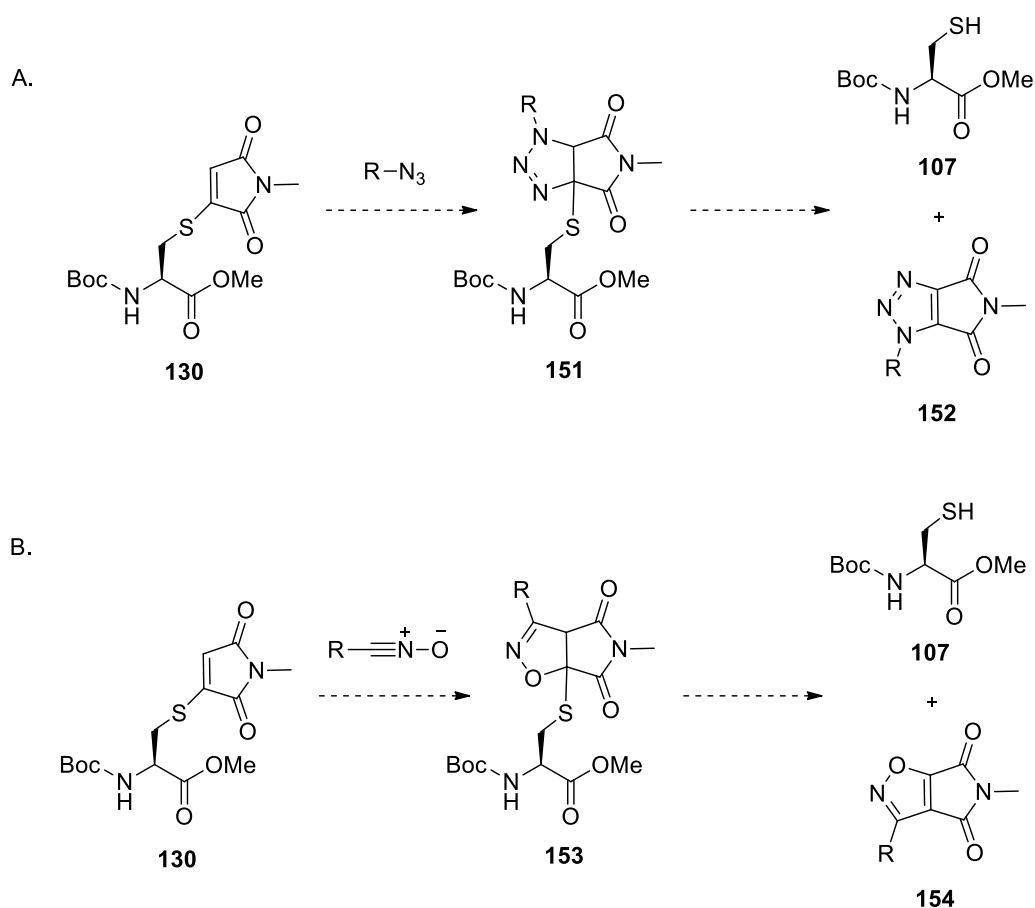


Scheme 92. Reaction of thiomaleimide **126** with sodium sulfite

The reaction of Grb2-SH2 domain (L111C) derived thiomaleimide **126** with 2-nitropropane **144** was also attempted. However, no reaction was observed even after incubation for 24 h at 37 °C. Thus, this reaction was not studied further.

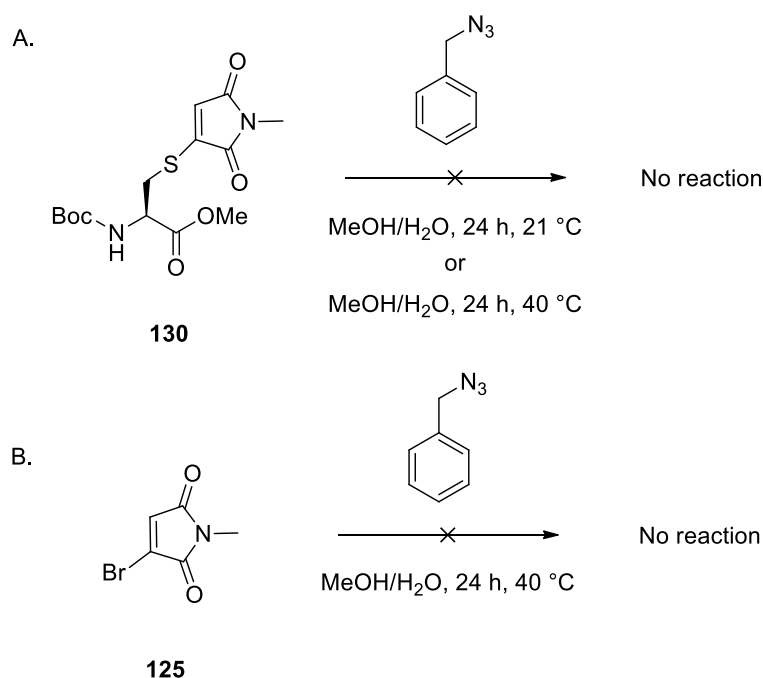
4.4.2 Cycloaddition

Maleimides are electron poor dienophiles/dipolarophiles and therefore excellent partners for cycloaddition reactions. Propargyl-allenyl dipoles like azides or nitrile oxides react with maleimide in aqueous media forming non-aromatic cyclic adducts.¹⁶⁷ It was thus envisaged that thiomaleimides might undergo similar reaction with 1,3-dipoles. It was proposed that the reaction of thiomaleimides with nitrile oxide and azide may form non-aromatic cycloadducts **151** or **153**, which could undergo aromatisation *via* the loss of the free cysteine **107** (Scheme 93).



Scheme 93. Proposed cycloaddition reaction outcome of thiomaleimide **130** with A) azide B) nitrile oxide

The reactivity of thiomaleimide **130** with benzyl azide was evaluated under a range of conditions (*Scheme 94*).

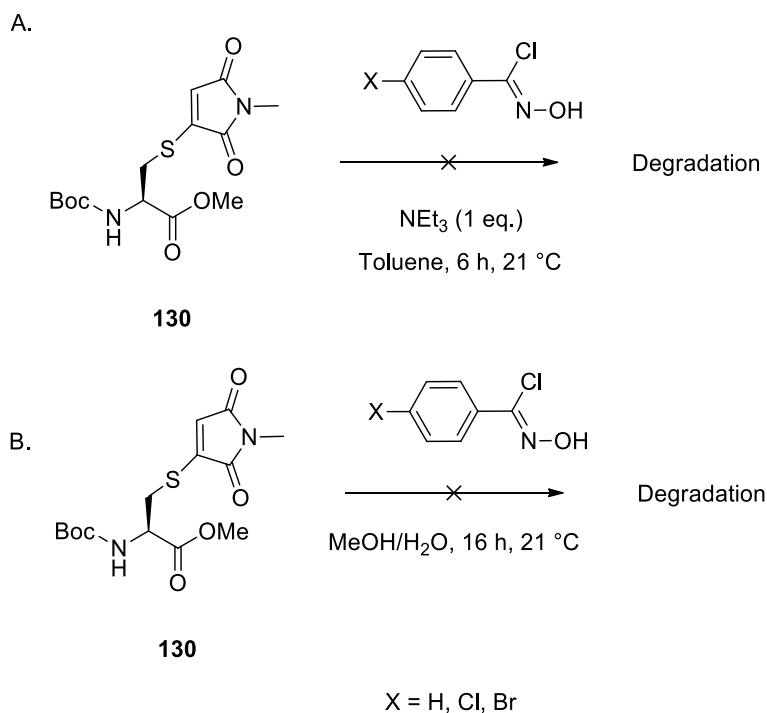


Scheme 94. Reaction of thiomaleimide 130 and bromo-N-methylmaleimide 125 with benzylazide

The reaction of thiomaleimide **130** with the benzyl azide in equimolar amounts at room temperature in mixture of methanol and water indicated no reaction following incubation for 24 h at 21 °C (*Scheme 94A*). Increasing the number of equivalents of benzylazide to 10 also led to no reaction, even after increasing the temperature to 40 °C (*Scheme 94A*). Similar reaction conditions were applied to the reaction of bromo-*N*-methylbromomaleimide **125** with benzyl azide to assess if it was the thioether functionality that precluded the reaction (*Scheme 94B*). Disappointingly, no conversion was observed in this case too.

Nitrile oxides are more reactive dipoles than azides and are known to react with maleimides at room temperatures in aqueous media.¹⁶² Nitrile oxides can be generated *in situ* from α -chloroaloximes either by deprotonation in water or by use of a base in an organic solvent. Thus the reaction of these 1,3-dipoles with thiomaleimides was evaluated. *N*-hydroxybenzimidoyl chloride, 4-chloro-*N*-hydroxybenzimidoyl chloride and 4-bromo-*N*-hydroxybenzimidoyl chloride (*Scheme*

94) were used as a source of nitrile oxides (synthesised by Dr C.C.Lee, Caddick Group).



Scheme 95. Reactions of thiomaleimide 130 with α -chloroaldoximes

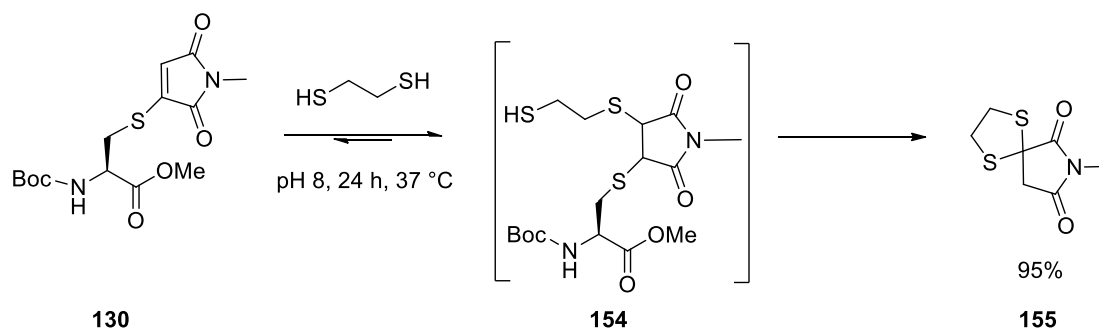
Unfortunately, all attempts of nitrile oxide cycloaddition with thiomaleimide **130** proved to be unsuccessful with a number of unidentifiable products being obtained. No cysteine **107** was detected and the approach was not explored further.

4.4.3 Tether reagents

Optimisation of the protocol for the conversion of thiomaleimide to free cysteine would represent a significant enhancement for the general applicability of thiomaleimide reversibility approach. The drawback of the current protocol rests with the use of excess reducing agents to affect reversibility. A novel bis-thiol based tether strategy was designed to trap the dithiosuccinimide intermediates into spiro-cyclic adducts **136** (*Figure 11*). As spiro-cyclic adducts lack an acidic proton, use of bis-thiols could eliminate the need for employing excess reducing conditions.

The utility of bis-thiols as reagents for converting thiomaleimides to cysteine was evaluated, initially on a small molecule model thiomaleimide. 1,2-Ethanedithiol (2 equivalents) was incubated with thiomaleimide **130** for 24 h at 37 °C under basic

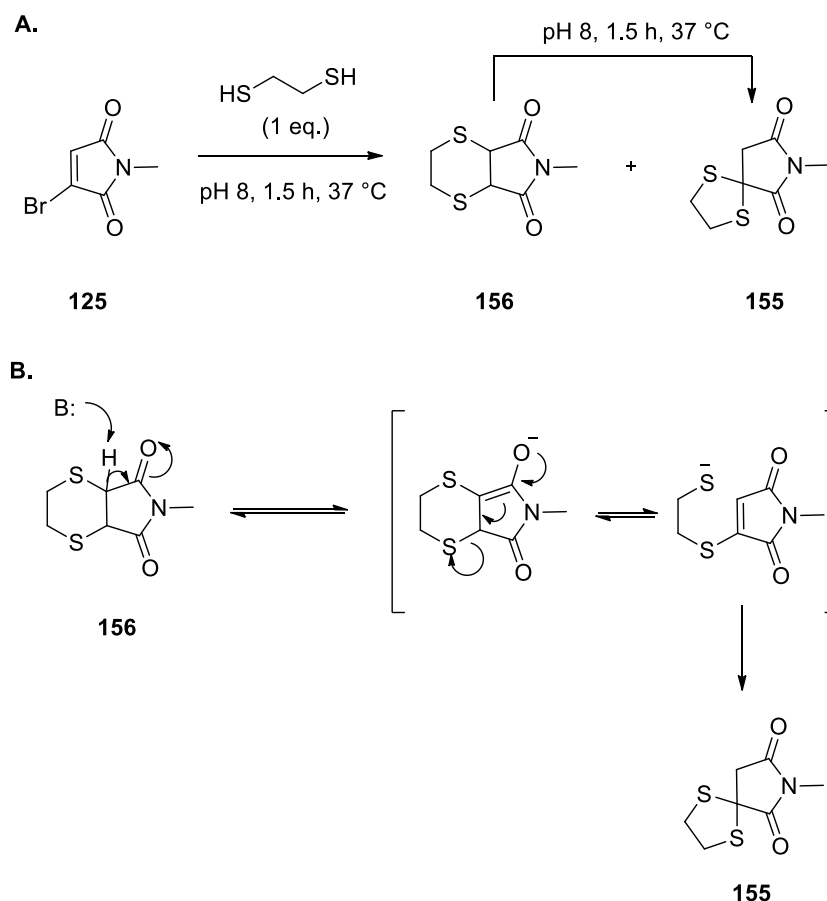
conditions (phosphate buffer, 100 mM, pH 8.0). DMF was used for solubilizing organic molecules into the buffer solution (*Scheme 96*).



Scheme 96. Reaction of thiomaleimide 130 with 1,2-ethanedithiol

Whilst free cysteine **107** could not be isolated perhaps due to re-oxidation and/or reaction with 1,2-ethanedithiol, stable spirocyclic dithiolane **155** was isolated in excellent yields. This transformation demonstrated the proof of concept of using bis-thiol reagents to trap thiomaleimides into stable dithiolanes, despite having to employ a modest excess of the dithiol reagents to compensate for the tendency of dithiols to oxidize rapidly to disulfides.

Further insight into the mechanism was gained by reaction of 1,2-ethanedithiol with bromo-*N*-methylmaleimide **125**, avoiding the interference of released cysteine **107** from thiomaleimide **130**. Interestingly, incubation of bromo-*N*-methylmaleimide **125** with 1,2-ethanedithiol yielded a mixture of spirocyclic dithiolane **155** and bicyclic bithioether **156** in a 65:35 ratio after 90 min at 37 °C (*Scheme 97*). Isolation of bithioether **156** followed by heating under the same reaction conditions, however, led to complete conversion to the thermodynamic product, dithiolane adduct **155** in quantitative yield.

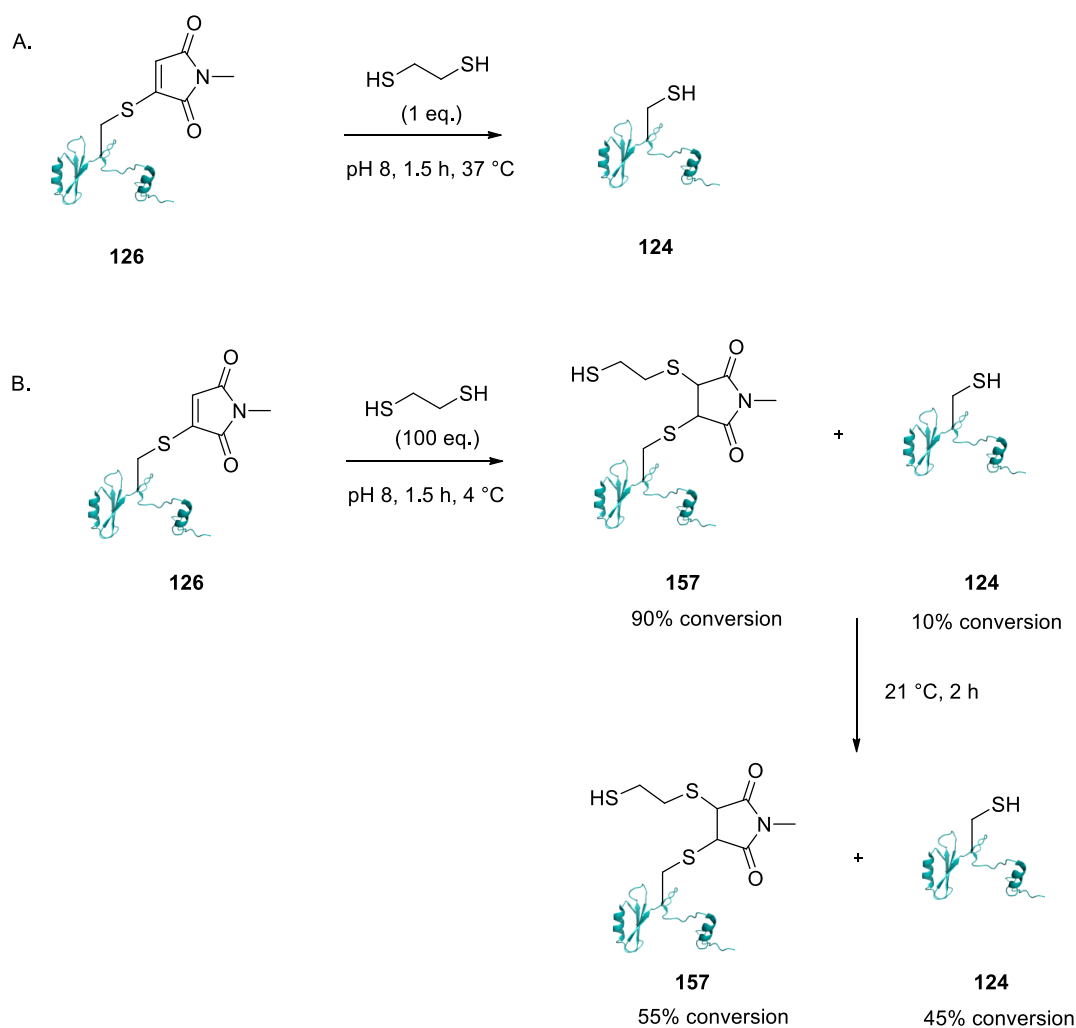


*Scheme 97. Reaction of A. Bromo-N-methylmaleimide **125** reaction with 1,2-ethanedithiol B. Conversion of bithioether **156** to dithiolane **155***

The conversion of bithioether **156** to dithiolane **155** is proposed to proceed through elimination-addition (*Scheme 97B*).

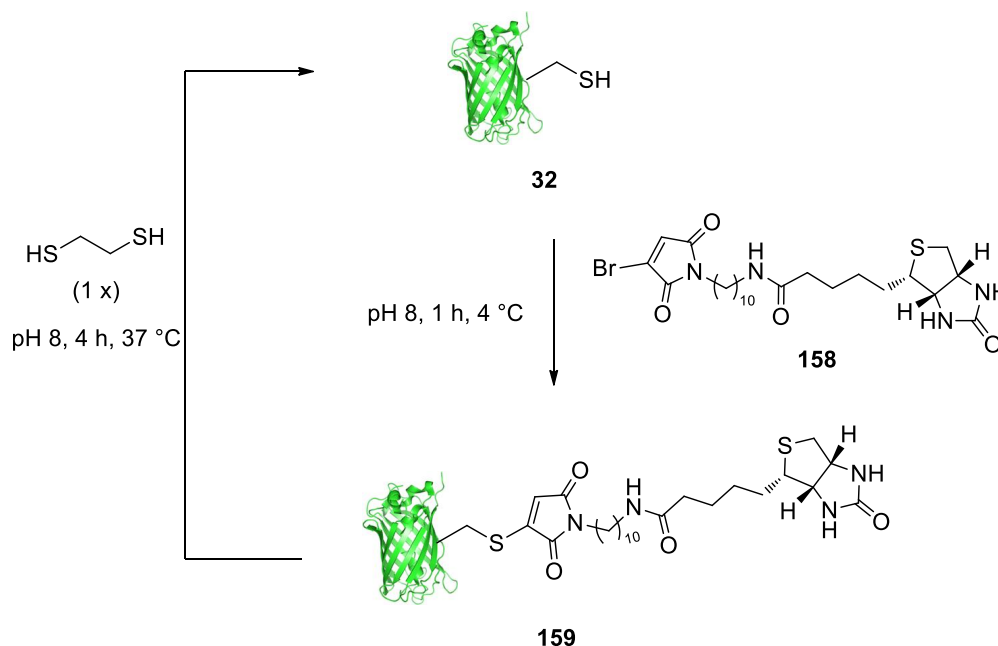
The reaction of 1,2-ethanedithiol with thiomaleimides was then investigated on a protein model system. Gratifyingly, incubation of thiomaleimide **126** with a single equivalent of 1,2-ethanedithiol for 4 h at 37 °C resulted in complete conversion to free Grb2-SH2 domain (L111C) **124** (*Scheme 98A*).

Incubation of Grb2-SH2 domain (L111C) derived thiomaleimide **126** with one equivalent of 1,2-ethanedithiol at 21 °C or 4 °C indicated some liberation of Grb2-SH2 domain (L111C) **124**, however the reaction was far slower. Intermediate dithiosuccinimide **157** was more stable at 4 °C and at 21 °C consistent with previous literature reports (*Scheme 98*).⁴⁸



Scheme 98. Reversibility of thiomaleimide 126 using 1,2-ethanedithiol at A) 37 °C B) 4 °C to 21 °C

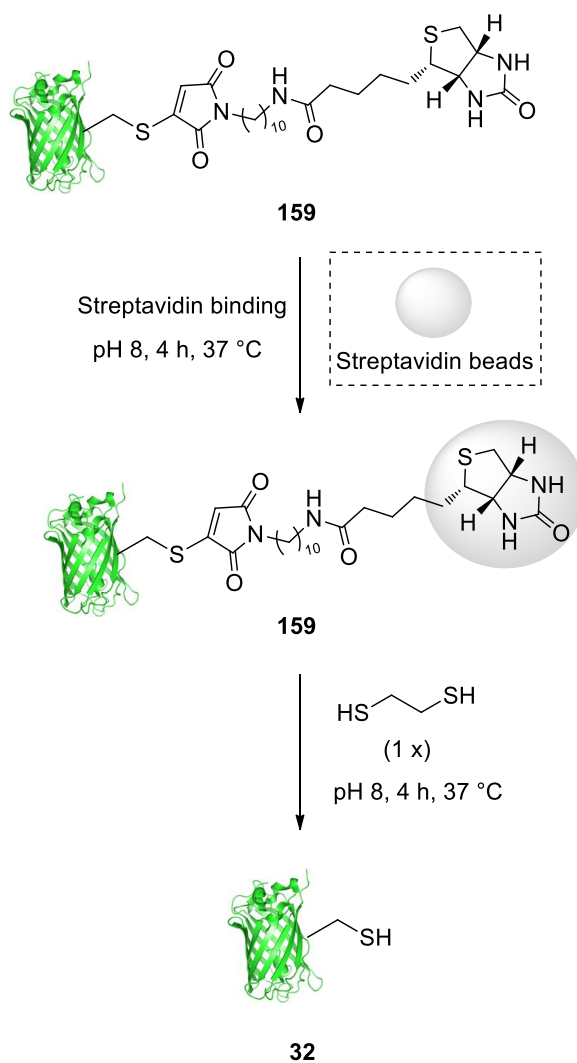
The utility of bis-thiol based protein modification approach was subsequently demonstrated with another model protein, single cysteine mutant GFP (S147C) **32** (cloned and expressed by Paul Moody, PhD student, Caddick group). Incubation of GFP (S147C) **32** with bromomaleimide-dec-biotin **158** (synthesised by Chris Ryan, Post-Doctoral Research Scientist, Caddick Group) for 1 h at 4 °C led to the formation of GFP (S147C, biotin-dec-maleimide) **159** in complete conversion. Thiomaleimide GFP (S147C, biotin-dec-maleimide) **159** was then treated with stoichiometric amount of 1,2-ethanedithiol to facilitate release of free protein. Gratifyingly, incubation for 4 h at 37 °C led to a complete release for GFP (S147C) **32** as observed by LCMS (*Scheme 99*).



*Scheme 99. Reversible modification of GFP (S147C) **32** with biotin-dec-maleimide **158***

The validity of bisthiol approach was then appraised with a proof of concept pull down experiment to establish that 1,2-ethanedithiol was well tolerated for protein purification using streptavidin beads. The pull down experiments were carried out in collaboration with Paul Moody, PhD student, Caddick group. Fluorescence of GFP provided a convenient parameter for the rapid analysis, hence was used as the read-out for released protein.

GFP (S147C, biotin-dec-maleimide) **159** was incubated directly with pre-washed streptavidin beads at 37 °C for 4 h (*Scheme 100*). After the incubation period, the unbound GFP (S147C) **32** was removed by thoroughly washing the beads with a large volume of PBS buffer solution (3×10 -fold dilution by diafiltration). Incubation of the washed GFP (S147C) loaded streptavidin beads with EDT at 37 °C for 4 h resulted in cleavage of GFP (S147C) from the beads. The released GFP (S147C) **32** was collected by washing with PBS buffer solution (*Scheme 100*). Fluorescence measurement of the supernatant revealed release of GFP (S147C) **32** protein from conjugate GFP (S147C, biotin-dec-maleimide) **159**, thereby indicating the suitability of use of 1,2-ethanedithiol for pull down experiments with the bromomaleimide reversibility approach (*Figure 12*).



Scheme 100. Proof of concept pull down experiment to release GFP (S147C) 32 from GFP (S147C, biotin-dec-maleimide) 159 with stoichiometric amount of 1,2-ethanedithiol

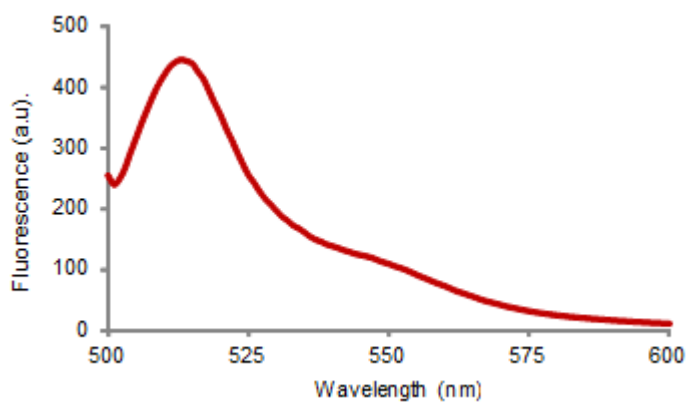


Figure 12. Fluorescence spectra of GFP (S147C) 32 following pull down

4.5 Conclusion

The utility of bromomaleimide based protein modification approach for reversible cysteine modification is compromised due to the necessity of using excess reducing conditions. Novel strategies for converting thiomaleimide to free cysteine were designed and evaluated on model small molecule and protein systems to develop improved reversibility protocols.

A nucleophile screen indicated rapid conversion of thiomaleimide to free cysteine on treatment with sodium sulfite. Whilst sodium sulfite was shown to preferentially react with thiomaleimide over disulfide, the reaction was found to be non-selective.

A cycloaddition based approach using 1,3-dipoles was also evaluated. The presence of bromo and thio functionalities on maleimide led to complete inhibition of cycloaddition reaction under conditions suited for protein modification.

A bis-thiol based tether strategy was designed based on proposed mechanism for conversion of thiomaleimide to cysteine, and validated on a small molecule model system and a model protein system. Bisthiol reagent 1,2-ethanedithiol was found to be highly efficient for conversion of thiomaleimides to free cysteines leading to formation of spirocyclic dithiolane. An insight into the reaction mechanism was obtained by analysis on small molecule model system. A proof of concept pull down experiment was carried out to demonstrate the utility of 1,2-ethanedithiol mediated conversion of thiomaleimide to free cysteine. This approach represented a significant improvement on the existing methodology and requires only a single equivalent of 1,2-ethanedithiol.

Conclusions and Further Work

The thesis has described and explored three distinct approaches for cysteine modification, each of which has been elaborated in separate chapters.

1. Site-selective dual labelling of proteins

The first approach for protein modification described in the thesis involves a novel method for the site-selective, dual-labelling of green fluorescent protein based on 'substrate controlled' cysteine modification. Two equally reactive cysteine residues are modified selectively in a single protein which is then dual labelled with orthogonal functional groups. The method exploits the protein micro-environment to distinguish reaction outcome of 2,5-dibromohexandiamide with cysteine residues forming either sulfonium or dehydroalanine. Further functionalisation of sulfonium moiety was carried out selectively *via* reaction with sodium azide. This provides a platform for dual functionalisation. The protein was then labelled using commercially available octyne and thiol probes to yield a homogeneously dual labelled protein. The reactions were observed to be highly selective and proceed to completion. The evidence for the reactions is obtained throughout using LCMS. Overall, the strategy represents an efficient, accessible and robust site-selective, dual modification approach.

Future work on this project should include exploring green fluorescent protein single mutants at various positions with differing solvent accessibility values across structured and non-structure protein regions with a view to establish a clear correlation between solvent accessibility and elimination of sulfonium to dehydroalanine. This applicability of this approach for other therapeutic proteins, including anticalins and DARPins is likely to be of interest.

2. Cysteine modification to thiophosphonium

The second approach for protein modification described in the thesis involves evaluation of thiophosphonium as a platform to access homogeneously modified proteins. The reaction of alkyl-aryl disulfides with tris(dimethylamino)phosphine was investigated as a simple way to access thiophosphonium species, on small molecule and protein systems. Additionally, the evidence for the formation of stable alkyl thiophosphonium in reaction of alkyl-aryl disulfides with tris(dimethylamino)phosphine was obtained using ^{31}P NMR spectroscopy. The formation of alkyl thiophosphonium was the predominant pathway; however, it was not found to be exclusive as aryl thiophosphonium was also observed. The suitability of thiophosphonium as an effective scaffold for protein modification was highlighted with stability studies. The viability of thiophosphonium-derived functionalisation, including the various challenges involved in synthesising and functionalising thiophosphonium intermediates was also investigated. Significant difficulty remains in accessing the thiophosphonium species in a selective and high yielding manner.

It is worthy to note that Davis *et al.* reported a similar method to access thiophosphonium functionality in a recent publication, whilst the work was undergoing.⁷⁰ Future work on this project should include exploring the use of alternative phosphines including tris(diethylamino)phosphine for more selective formation of the thiophosphonium species. In addition, functionalisation of thiophosphonium by reaction with various nucleophiles including thiol and azide is likely to be of interest.

3. Reversible cysteine modification using bromomaleimides.

The third approach for cysteine modification described in the thesis involves development of bromomaleimide based reversible cysteine modification. Bromomaleimide reagents, developed within the Caddick and Baker groups at UCL, allow rapid and selective functionalisation with cysteine residues to afford thiomaleimides, that can be converted back to free thiols in the presence of excess reducing agents such as tris(2-carboxyethyl)phosphine (TCEP) and 2-mercaptoethanol. The majority of this work focussed on the development of improved strategies for conversion of thiomaleimides back to free cysteines.

Multiple strategies were designed and evaluated on small molecule and protein model system. Of particular note is the discovery that the reaction of stoichiometric amount of 1,2-ethanedithiol with thiomaleimide proceeds to form stable spirocyclic dithiolanes in near quantitative yield, releasing free thiol. The utility of using bistiols to affect thiomaleimide reversibility was demonstrated in a proof of concept pull-down experiment using streptavidin beads. In addition to bistiols, sodium sulfite was shown to preferentially react with thiomaleimide over disulfide but the reaction was found to be non-selective.

Future work on this project should include utility of bithiol reagents for conversion of thiomaleimides to free thiols for applications in chemical biology and proteomics.

Experimental

General Experimental

Organic Chemistry

All reactions were carried out at ambient temperature and under a nitrogen atmosphere and at atmospheric pressure unless otherwise stated. Air- and moisture-sensitive liquids and solutions were transferred *via* syringe into the reaction vessels through rubber septa. All reaction mixtures were stirred magnetically. Reaction progress was monitored by analytical TLC using Merck aluminium coated plates covered with a 0.2 mm layer of silica gel 60 F254. Product spots were visualised by UV irradiation at 254 nm and subsequent staining with potassium permanganate solution, followed by heating. Concentration *in vacuo* refers to distillation on a Buchi rotary evaporator, and where appropriate under high vacuum. Infrared spectra were obtained on a Perkin Elmer Spectrum 100 FTIR Spectrometer operating in ATR mode with frequencies given in reciprocal centimetres (cm^{-1}). Melting points were measured with a Gallenkamp apparatus and are uncorrected. Optical rotations were measured using a Perkin Elmer 343 polarimeter. Fluorescence was measured using Starna Scientific 26.100-F quartz fluorescence cuvette with a 10mm path length. The fluorescence spectra were obtained at room temperature using a Cary Eclipse Fluorescence Spectrophotometer. The sample was excited at 494nm, and the emission intensity was scanned at 120nm/min with an averaging time of 0.5s and a data interval of 1nm. The excitation and emission slit was set to 5nm.

Protein Mass Spectroscopy

LCMS was performed on protein samples using one of the following:

1. Thermo Scientific uPLC connected to MSQ Plus Single Quad Detector (SQD). Column: Hypersil Gold C4 1.9 μm 2.1 x 50 mm. Wavelength: 254 nm. Mobile Phase: 99:1 Water (0.1% formic acid): MeCN (0.1% formic acid)

to 1:9 Water (0.1% formic acid): MeCN (0.1% formic acid) gradient over 4 min. Flow Rate: 0.3 mL/min. MS Mode: ES+. Scan Range: $m/z = 500-2000$. Scan time: 1.5 s. Data obtained in continuum mode. The electrospray source of the MS was operated with a capillary voltage of 3.5 kV and a cone voltage of 50 V. Nitrogen was used as the nebulizer and desolvation gas at a total flow of 600 L/h. Total mass spectra for protein samples were reconstructed from the ion series using the pre-installed ProMass software using default settings for large proteins in m/z range 500-1500.

2. Waters Acquity uPLC connected to Waters Acquity Single Quad Detector (SQD). Column: Acquity uPLC BEH C18 1.7 μm 2.1 x 50 mm. Wavelength: 254 nm. Mobile Phase: 95:5 Water (0.1% Formic Acid): MeCN (0.1% Formic Acid) Gradient over 4 min to 5:95 Water (0.1% Formic Acid): MeCN (0.1% Formic Acid). Flow Rate: 0.6 mL/min. MS Mode: ES+. Scan Range: $m/z = 85-2000$. Scan time: 0.25 sec. Data obtained in continuum mode. The electrospray source of the MS was operated with a capillary voltage of 3.5 kV and a cone voltage of 50 V. Nitrogen was used as the nebulizer and desolvation gas at a total flow of 600 L/h. Total mass spectra for protein samples were reconstructed from the ion series using the MaxEnt 1 algorithm pre-installed on MassLynx software

Quoted yields determined by ratio of peak heights in deconvoluted mass spectrum unless otherwise stated.

Nuclear Magnetic Resonance (NMR) Spectroscopy

^1H NMR spectra were recorded at 300 MHz, 400 MHz, 500 MHz and 600 MHz and ^{13}C NMR at 75 MHz, 100 MHz, 125 MHz and 150 MHz on Bruker AMX300, AMX400, AMX500 and AMX600 at 25 °C in a deuterated solvent as described below. The chemical shifts (δ) for ^1H and ^{13}C are quoted relative to residual signals of the solvent on the parts per million (ppm) scale. In the case of multiple amide rotamers, only the major rotamer has been assigned. Coupling constants (J values) are reported in Hertz (Hz) and are reported as J (H-H) couplings unless otherwise stated. Units for optical rotation are $10^{-1} \text{ deg cm}^2 \text{ g}^{-1}$. Infrared spectra were obtained on a Perkin Elmer Spectrum 100 FTIR Spectrometer operating in ATR mode. Mass

spectra were obtained at UCL on either a VG70-SE (FAB), Thermo Finnigan MAT900Xp (EI and CI) or Waters LCT Premier XE (ES) mass spectrometer.

Chemicals and Solvents

Solvents were used as received unless otherwise stated. Petrol refers to petroleum ether (b.p. 40-60 °C). All reagents were purchased from Sigma-Aldrich or Fisher Scientific and were used without further purification. Dibenzylcyclooctyne-PEG₄-Fluor-545 **65** was purchased from Jena Bioscience GmbH. Octyne **52** was synthesised by Antoine Marauani (PhD student, Caddick Group). Bromomaleimide-dec-biotin **158** was synthesized by Dr. Chris Ryan (Post-Doctoral Research Scientist, Caddick Group).

Cloning and expression of proteins

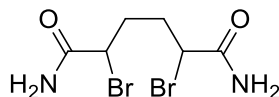
Superfolder GFP **29**, GFP (T230C, 233Δ) **30**, GFP (S147C) **32** and GFP (S147C, T230C,233Δ) **55** were cloned and expressed by Paul Moody (PhD Student, Caddick Group). Grb2-SH2 (L111C) **124** was obtained from Waksman group.

The gene for GFP (+CG) **31** and GFP (S147C, +CG) **54** in the vector pNIC28-Bsa4 were generated as described previously.¹⁴⁹ All proteins were expressed with polyhistidine-tag at the N-terminus. Protein concentration was determined using UV-absorbance at 280 nm with estimated extinction coefficient generated using ExPASy ProtParam tool.

The vector was transformed into BL21(DE3)plysS cells, and proteins were expressed at 15 °C. Proteins were purified by nickel chromatography, and the buffer was exchanged by repeated diafiltration with sodium phosphate (100 mM, pH 8.0) at 4 °C. Prior to analysis by LCMS, dithiothreitol (10 μL, 340 mM as a solution in water, 1000 equivalents per 100 μL of protein (1 mg/mL solution)) was added, the mixture incubated at 21 °C for 1 h and excess reducing agent was removed by repeated diafiltration with sodium phosphate (100 mM, pH 8.0).

Experimental for Chapter 2

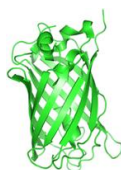
2,5-Dibromohexanediamide **14**⁷⁰



To a stirring solution of SOCl_2 (15.0 mL, 206 mmol) was added adipic acid (5.00 g, 34.2 mmol) at 21 °C. The reaction mixture was refluxed for 90 min. CHCl_3 (20 mL) was added to the reaction followed by NBS (14.6 g, 82.2 mmol). The reaction was stirred vigorously and 4 drops of HBr (48% aq.) were added. The reaction was then heated to reflux for 2 h, cooled to 4 °C and stirred for 30 min. Precipitate from the reaction mixture was filtered and washed with Et_2O (20 mL). The filtrate was concentrated *in vacuo* to obtain the crude acid chloride. In a separate reaction flask, 40 mL of NH_4OH (25% aq.) was cooled to 4 °C and crude acid chloride was added dropwise over 20 min with rapid stirring. The reaction mixture was further stirred for another 1 h. A white solid precipitated was obtained that was filtered and dried *in vacuo*. The crude product was resuspended in H_2O (20 mL) and MeOH (20 mL), heated to 60 °C and stirred for 30 min. The suspension was cooled, filtered and washed with cold MeOH (20 mL) to afford **14** as a white solid (4.13 g, 13.7 mmol, 40%).

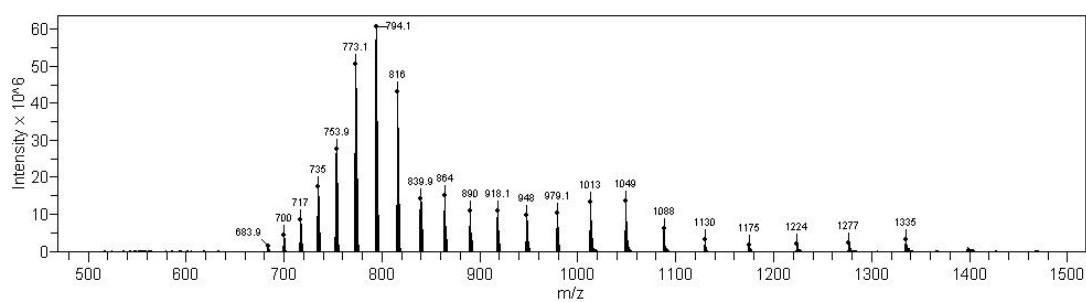
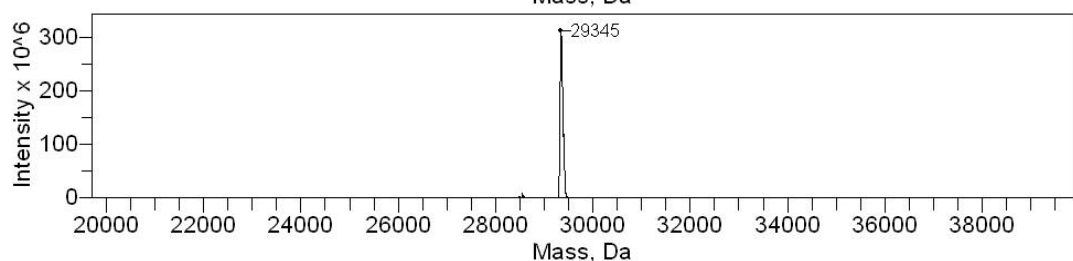
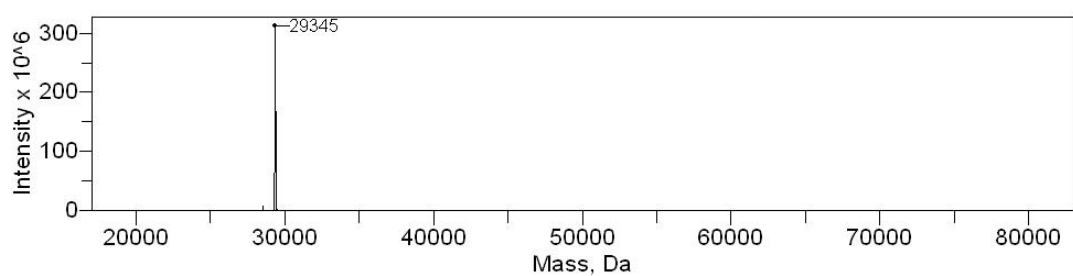
^1H NMR (600MHz, DMSO) δ 7.70 (s, 2H), 7.33 (s, 2H), 4.35 – 4.29 (m, 2H), 2.06-1.81 (m, 4H); ^{13}C NMR (150MHz, DMSO) (Both diastereomers 1:1) δ 169.9 (C), 169.8 (C), 48.5 (CH), 48.2 (CH), 32.6 (CH_2), 32.5 (CH_2); IR (solid) ν_{max} 3332, 3176, 1657, 1443, 1411, 765, 693 cm^{-1} ; LRMS (m/z) (CI) 305 ($^{81,81}\text{M}^+$, 29), 303 ($^{81,79}\text{M}^+$, 54), 301 ($^{79,79}\text{M}^+$, 26), 286 (100); HRMS $\text{C}_6\text{H}_{10}^{79}\text{Br}_2\text{N}_2\text{O}_2$ calcd. 300.9187 found 300.9191 m.p. 192-194 °C.

Superfolder GFP 29¹⁴⁹

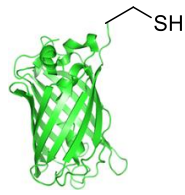


Sequence

MHHHHHSSGVDLGTENLYFQSMRKGEELFTGVVPILEVELDGDVNGHKFSVRGEGEG
DATNGKLTLLKFICTTGKLPVPWPTLVTTLTYGVCFARYPDHMKQHDFFKSAMPEGY
VQERTISFKDDGTYKTRAEVKFECDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHN
VYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSV
LSKDPNEKRDHMLLEFVTAAGITHGMDELYK (Expected: 29341, Observed
29345)

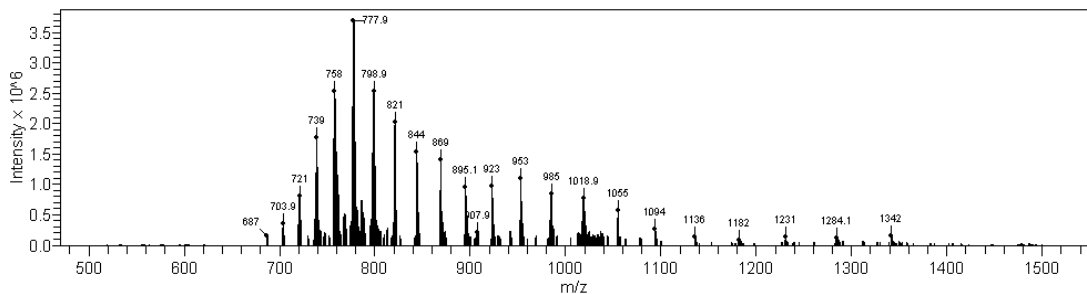
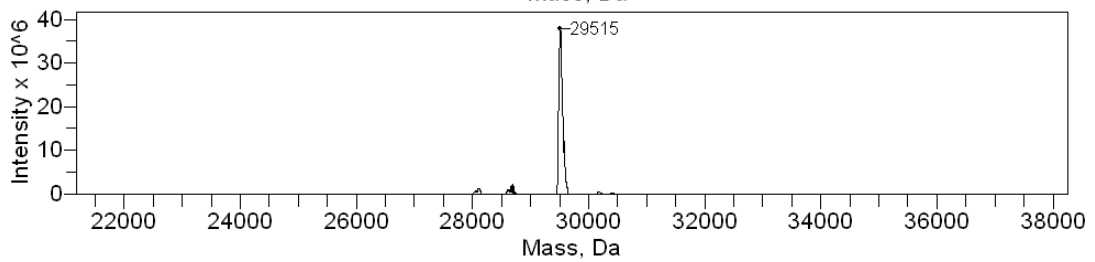
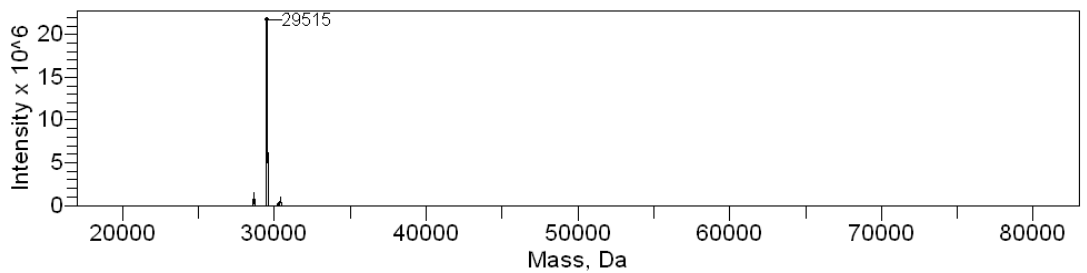


GFP (+CG) 30

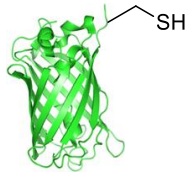


Sequence

MHHHHHSSGVDLGTENLYFQSMRKGEELFTGVVPILEVELDGDVNGHKFSVRGEGEG
DATNGKLTLLKFICTTGKLPVPWPTLVTTLTYGVCFAFYDPDHMKQHDFFKSAMPEGY
VQERTISFKDDGTYKTRAEVKFEGLTLVNRIELKGI DFKEDGNILGHKLEYNFNSHN
VYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSV
LSKDPNEKRDHMLLEFVTAAGITHGMDELYKCG (Expected 29519, Observed
29515)

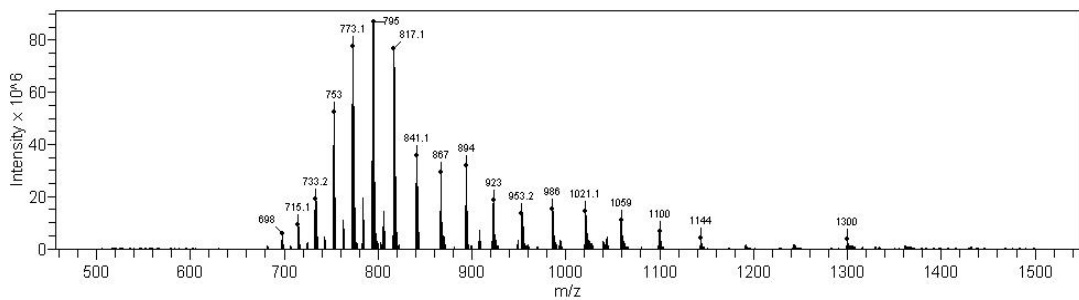
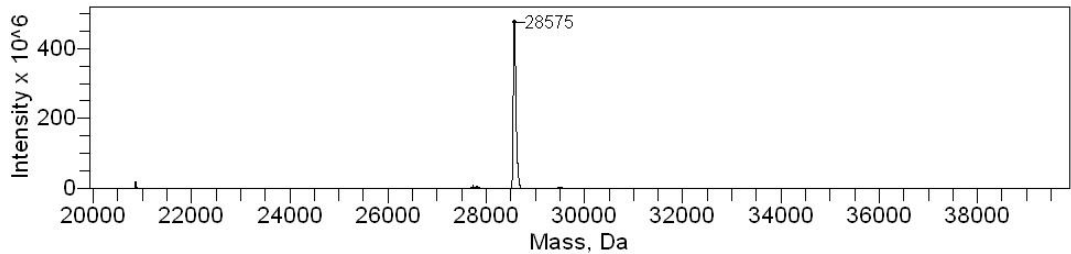
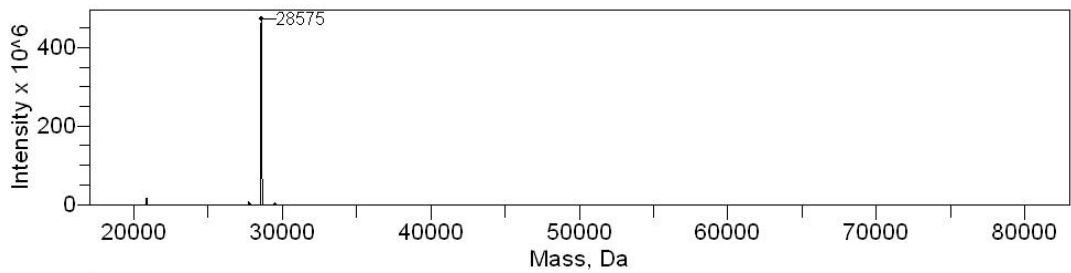


GFP (T230C, 233Δ) 31

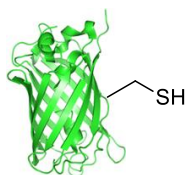


Sequence

MHHHHHSSGVDLGTENLYFQSMRKGEELFTGVVPILEVELDGDVNGHKFSVRGEGEG
DATNGKLTTLKFICTTGKLPVPWPTLVTTLTYGVCFARYPDHMKQHDFFKSAMPEGY
VQERTISFKDDGTYKTRAEVKFEGLTLVNRIELKGIDFKEDGNILGHKLEYNFNFSHN
VYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSV
LSKDPNEKRDHMLLEFVTAAGICHG (Expectd 28581, Observed 28575)

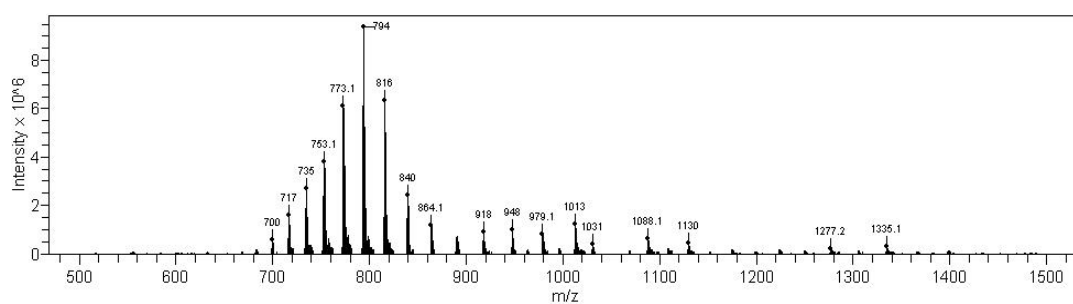
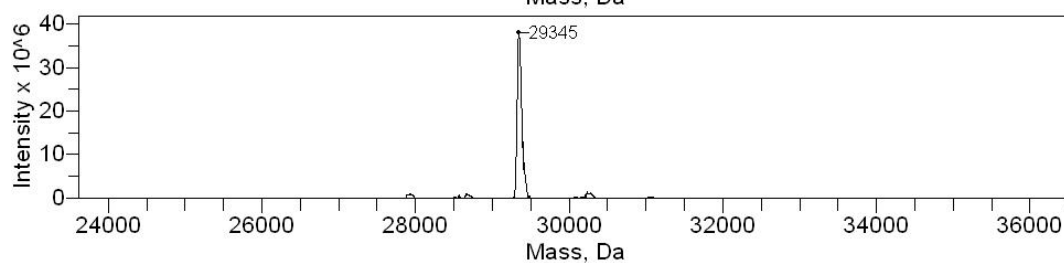
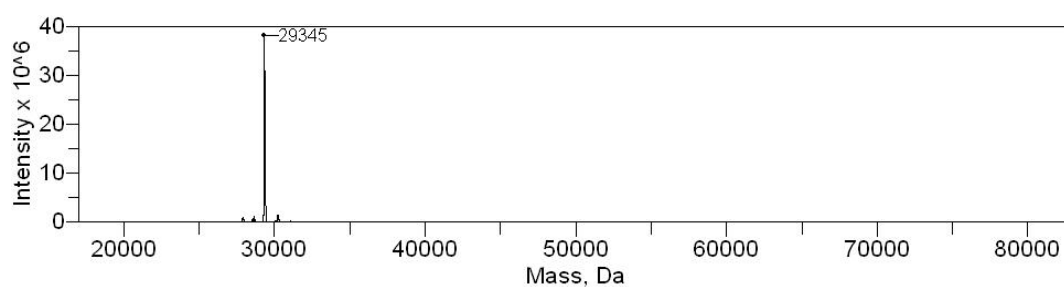


GFP (S147C) 32¹⁴⁹

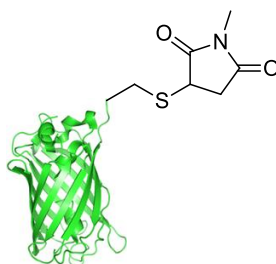


Sequence

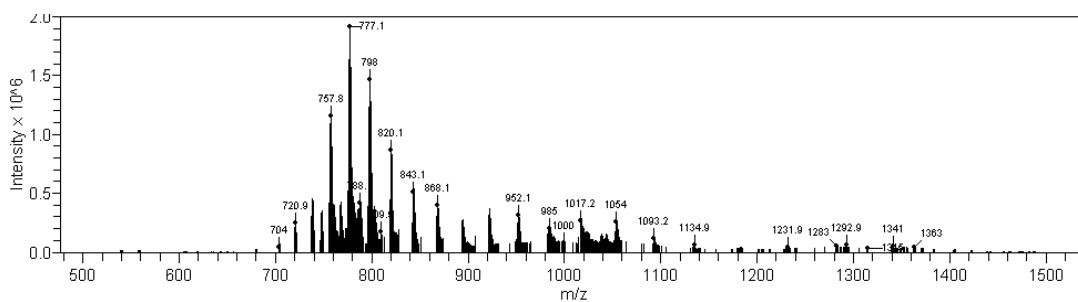
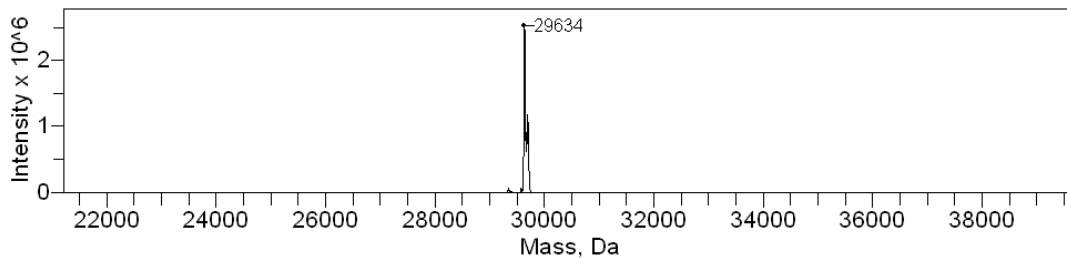
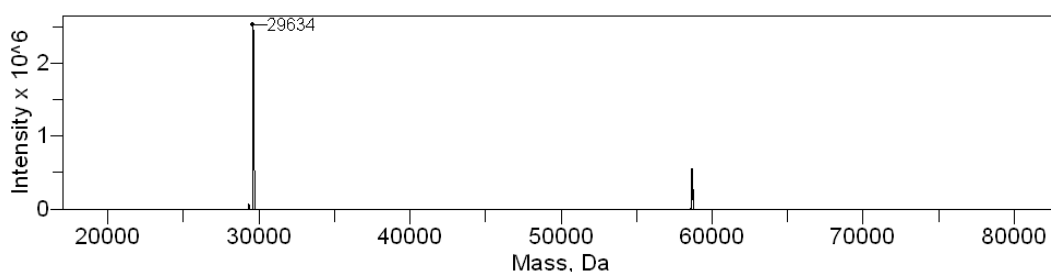
MHHHHHSSGVDLGTDNLYFQSMRKGEELFTGVVPI LVELDGDVNGHKFSVRGEGEG
DATNGKLT LKFICTTGKLPVPWPTLVTTLT YGVQCFARYPDHMKQHDFFKSAMPEGY
VQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGI DFKEDGNILGHKLEYNFNCHN
VYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSV
LSKDPNEKRDHMLLEFVTAAGITHGMDELYK (Expected 29343, Observed 29345)



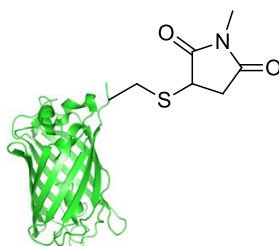
GFP (+CG, *N*-methylmaleimide) **33**



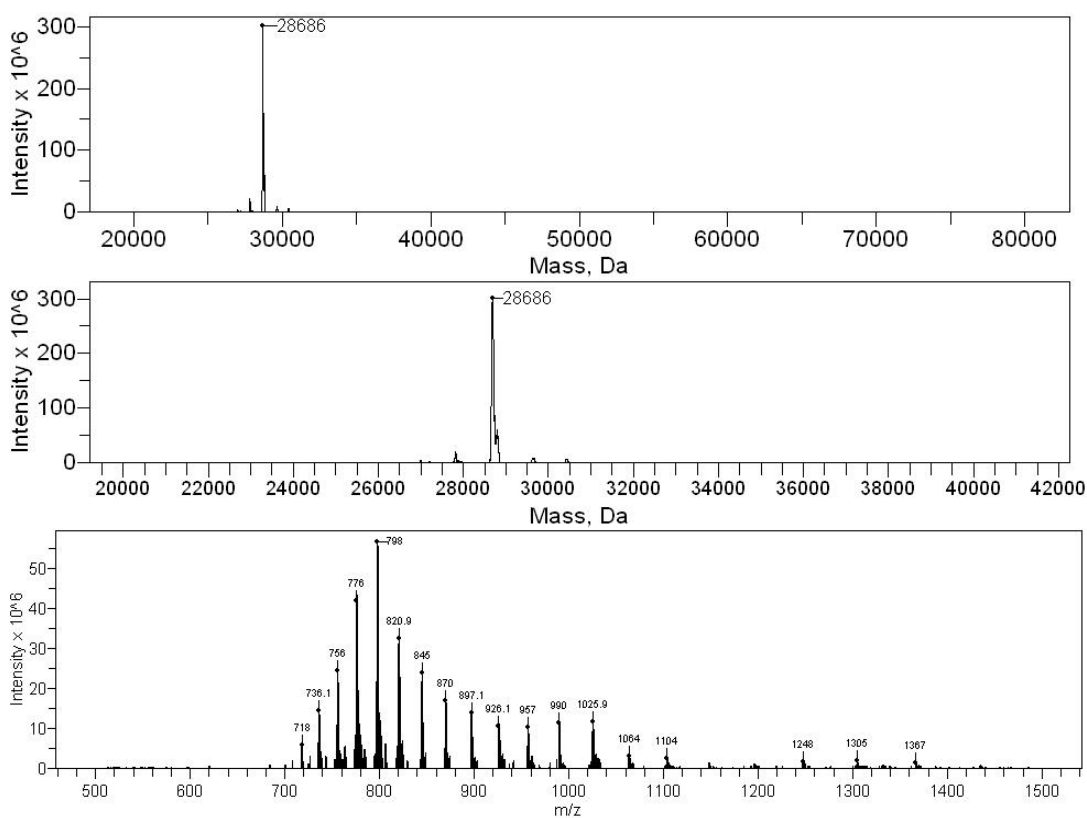
N-methylmaleimide (5 μ L, 138 mM solution in DMF, 20 equivalents) was added to a solution of GFP (+CG) **30** (100 μ L, 1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 21 $^{\circ}$ C. The mixture was vortexed for 1 s, maintained at 21 $^{\circ}$ C for 1 h and analysed by LCMS (Expected 29624, Observed 29634)



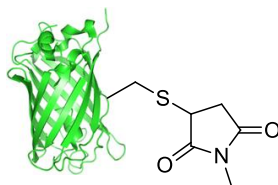
GFP (T230C, *N*-methylmaleimide) **34**



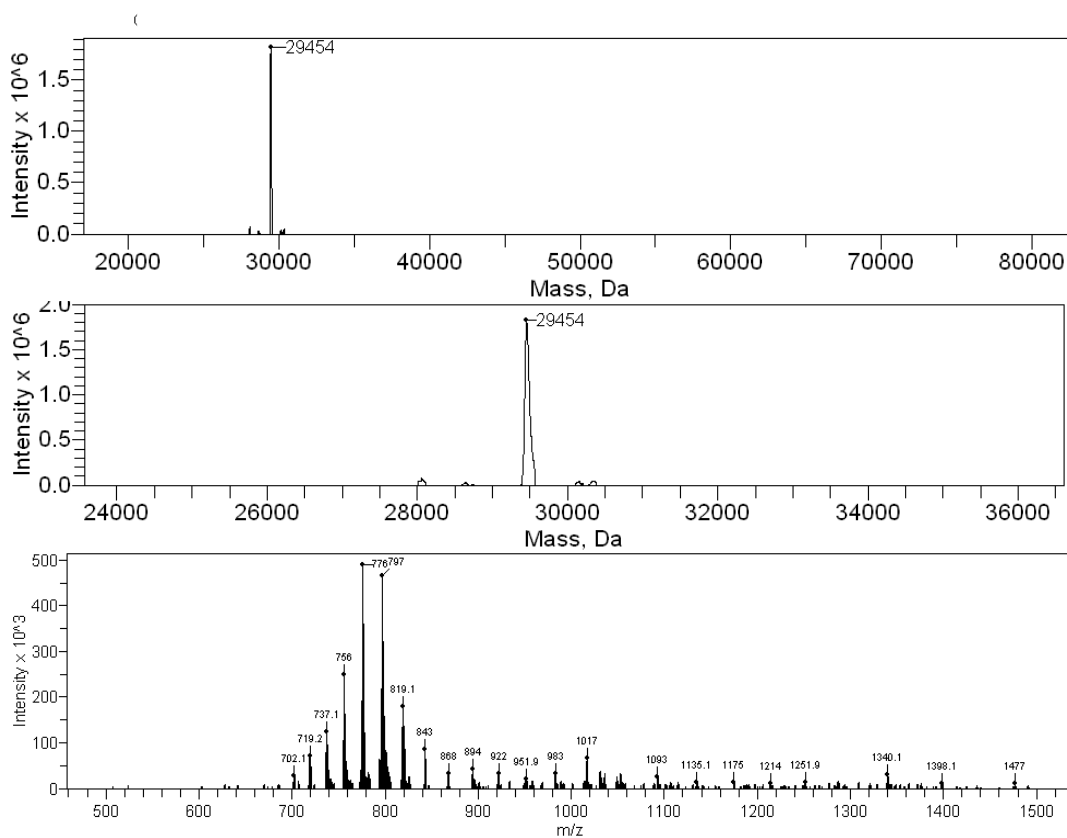
N-methylmaleimide (5 μ L, 138 mM solution in DMF, 20 equivalents) was added to a solution of GFP (T230C, 233 Δ) **31** (100 μ L, 1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 21 $^{\circ}$ C. The mixture was vortexed for 1 s, maintained at 21 $^{\circ}$ C for 1 h and analysed by LCMS (Expected 28686, Observed 28686).



GFP (S147C, *N*-methylmaleimide) **35**



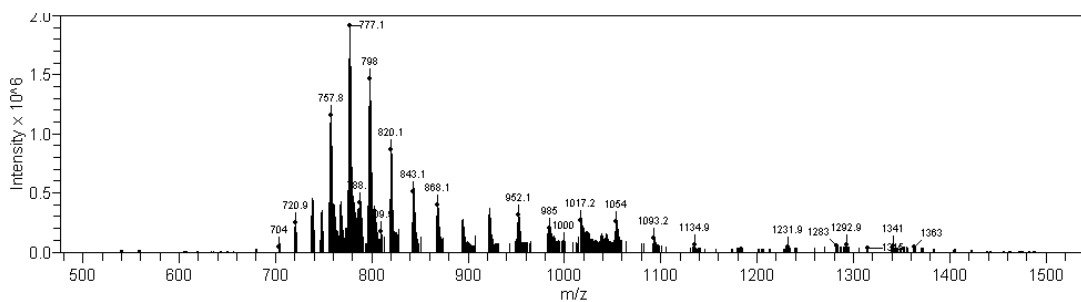
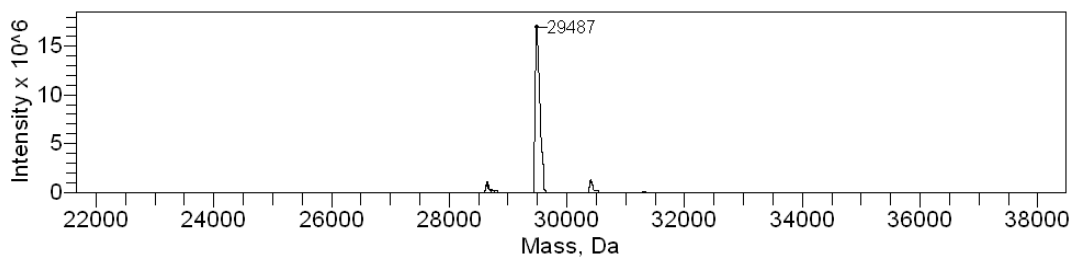
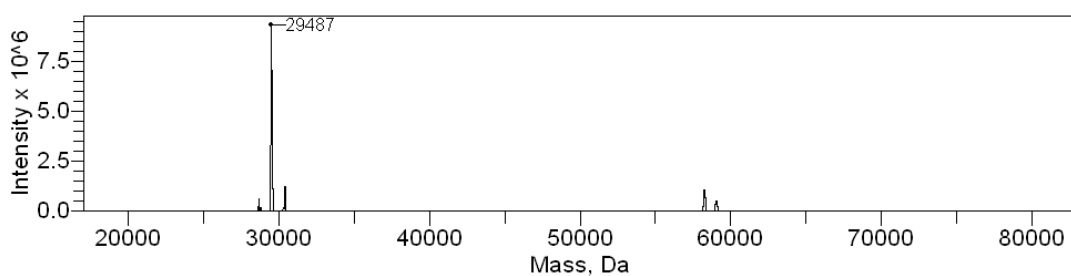
N-methylmaleimide (5 μ L, 138 mM solution in DMF, 20 equivalents) was added to a solution of GFP (S147C) **32** (100 μ L, 1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 21 $^{\circ}$ C. The mixture was vortexed for 1 s, maintained at 21 $^{\circ}$ C for 1 h and analysed by LCMS (Expected 29456, Observed 29454).



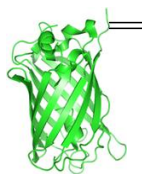
GFP (+CG, Dha) **39**



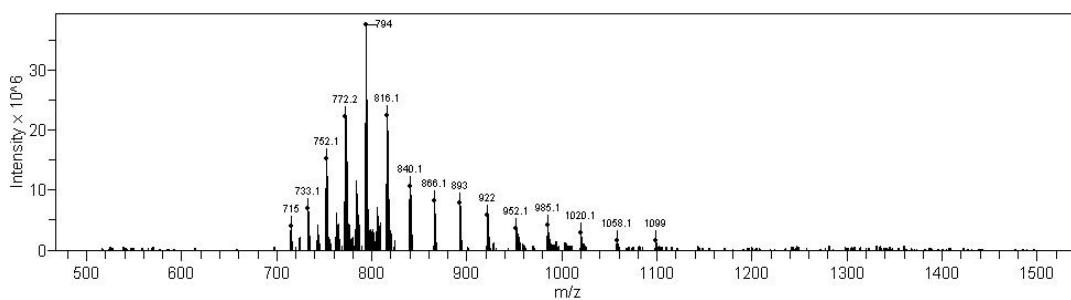
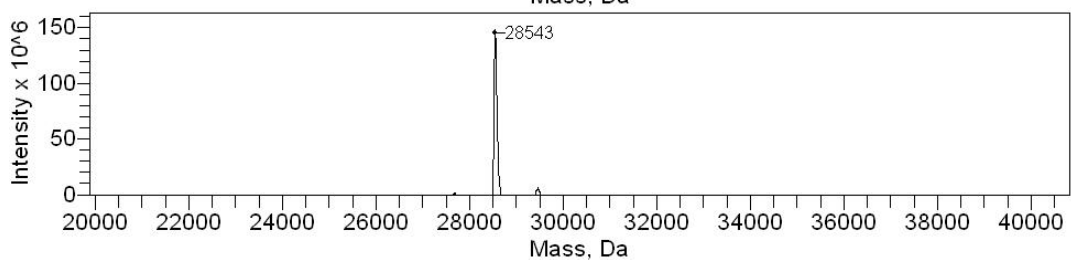
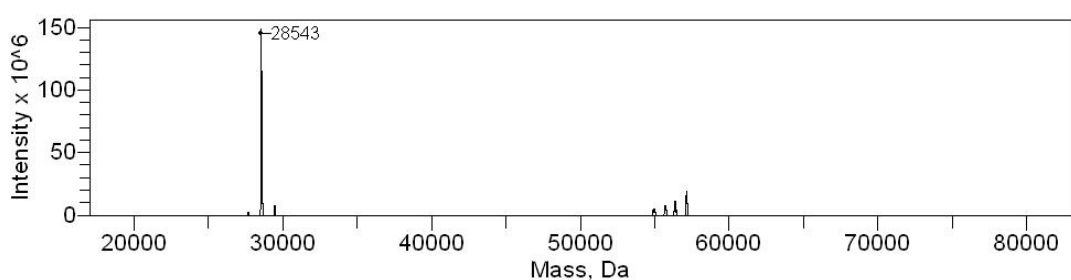
2,5-Dibromohexanediamide (**14**, 10 μ L, 17 mM solution in DMF, 50 equivalents) was added to a solution of GFP (+CG) **30** (1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 21 °C. The mixture was vortexed for 1 s, maintained at 37 °C for 2 h and analysed by LCMS (Expected 29481, Observed 29487).



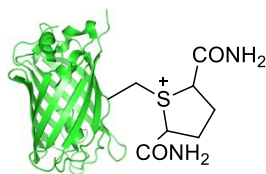
GFP (T230C, Dha) **40**



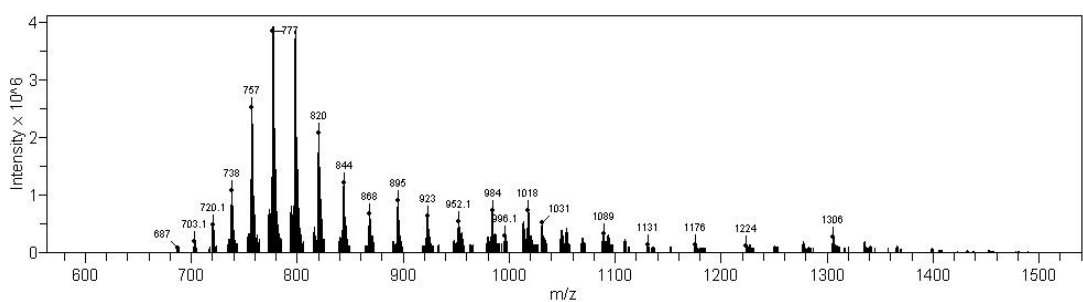
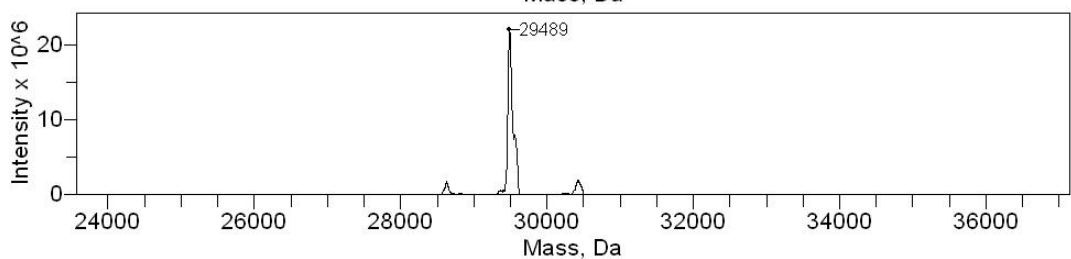
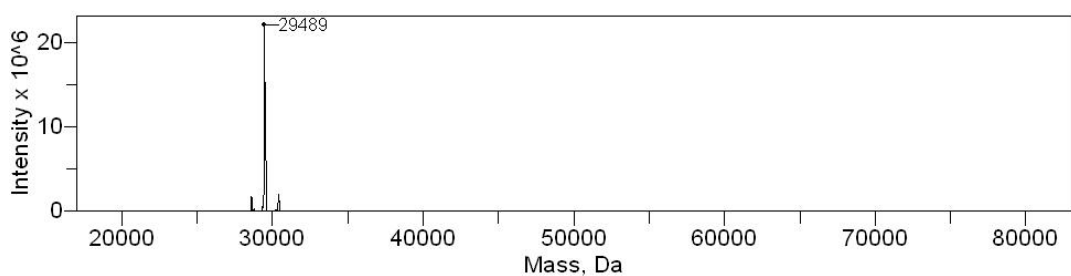
2,5-Dibromohexanediamide (**14**, 10 μ L, 17 mM solution in DMF, 50 equivalents) was added to a solution of GFP (T230C, 233 Δ) **31** (1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 21 $^{\circ}$ C. The mixture was vortexed for 1 s, maintained at 37 $^{\circ}$ C for 2 h and analysed by LCMS (Expected 28541, Observed 28543).



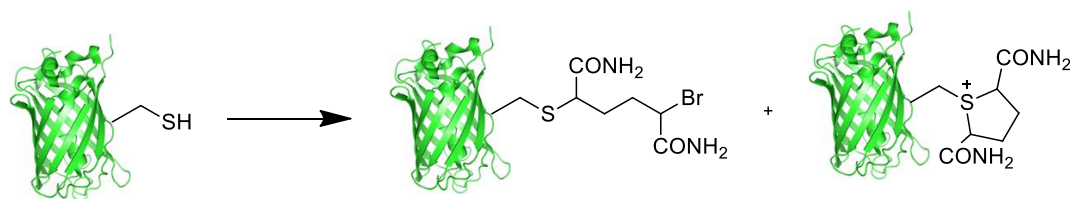
GFP (S147C, Sulf) **42**



2,5-Dibromohexanediamide (**14**, 10 μ L, 17 mM solution in DMF, 50 equivalents) was added to a solution of GFP (S147C) **32** (1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 21 °C. The mixture was vortexed for 1 s, maintained at 37 °C for 2 h and analysed by LCMS (Expected 29486, Observed 29489).



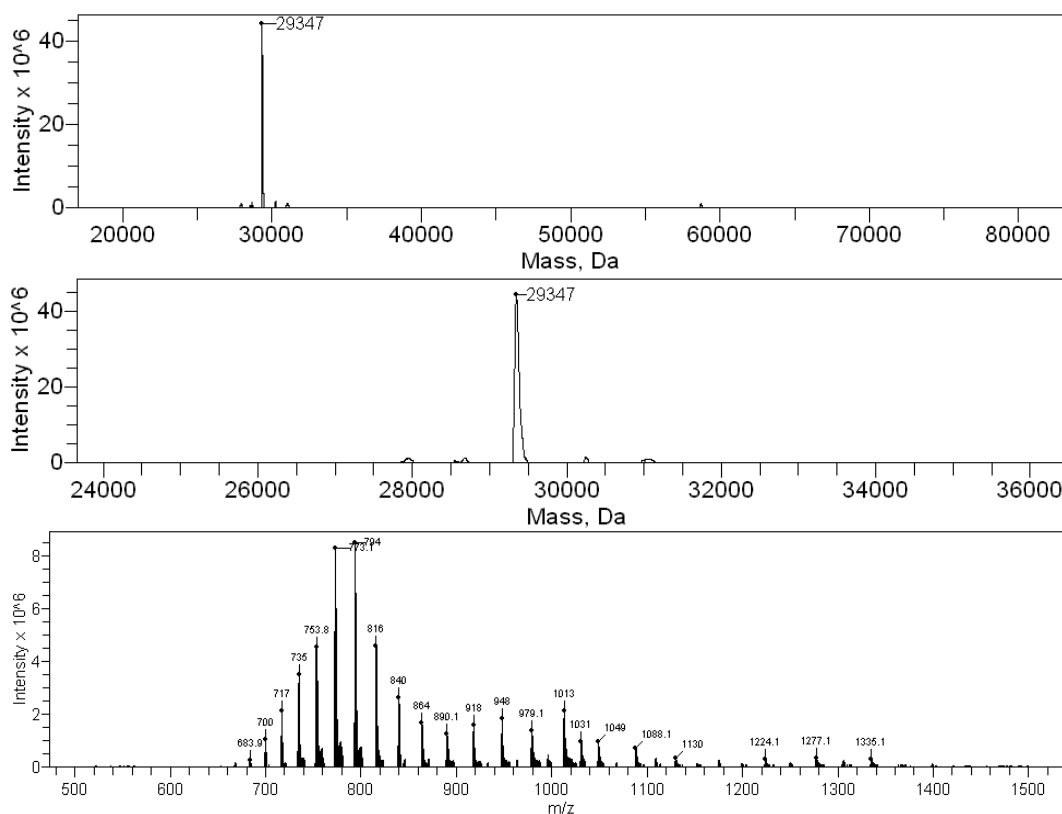
Optimization of GFP (S147C, Sulf) **42** formation



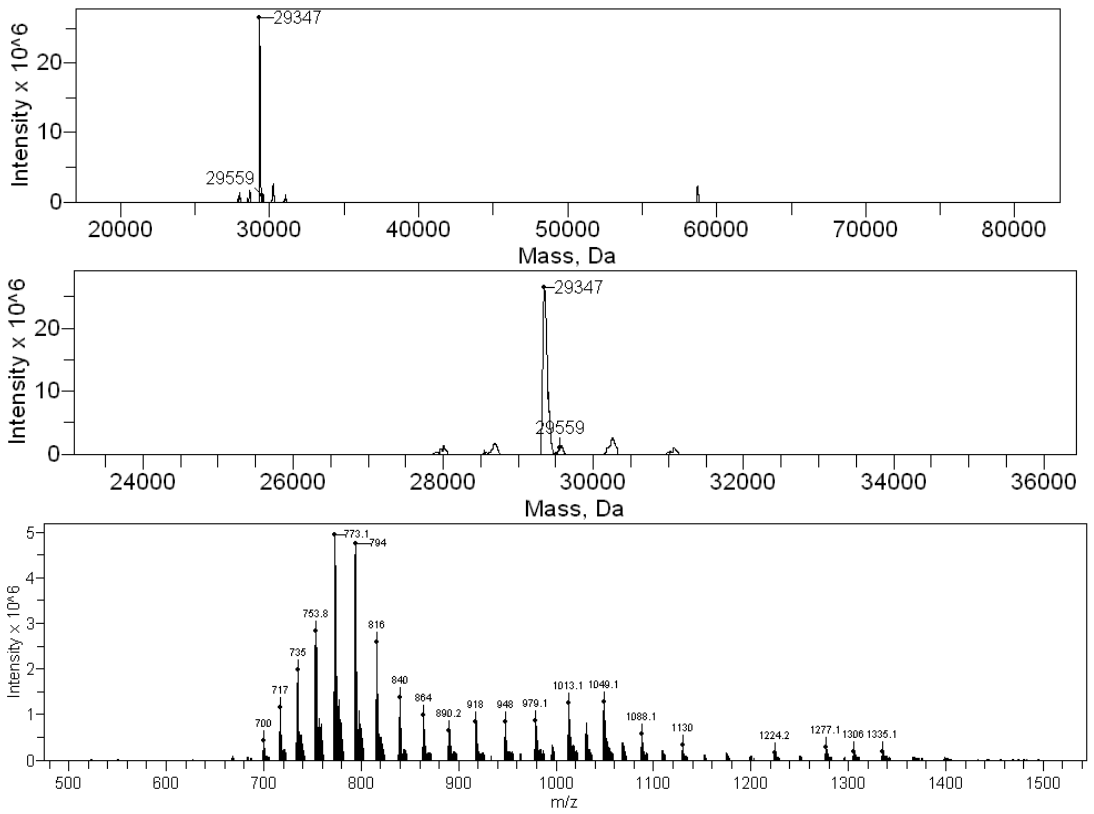
2,5-Dibromohexanediamide **14** (10, 25 or 50 equivalents), as a solution in DMF, was added to a solution of GFP (S147C) (100 μ L, 1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 21 $^{\circ}$ C. The mixture was vortexed for 1 s, maintained at the required temperature (4 $^{\circ}$ C, 21 $^{\circ}$ C or 37 $^{\circ}$ C) for the prescribed time (2 h, 5 h or 20 h) and analysed by LCMS (Expected Mass: 29566 and 29486).

10 equivalents, 21 $^{\circ}$ C

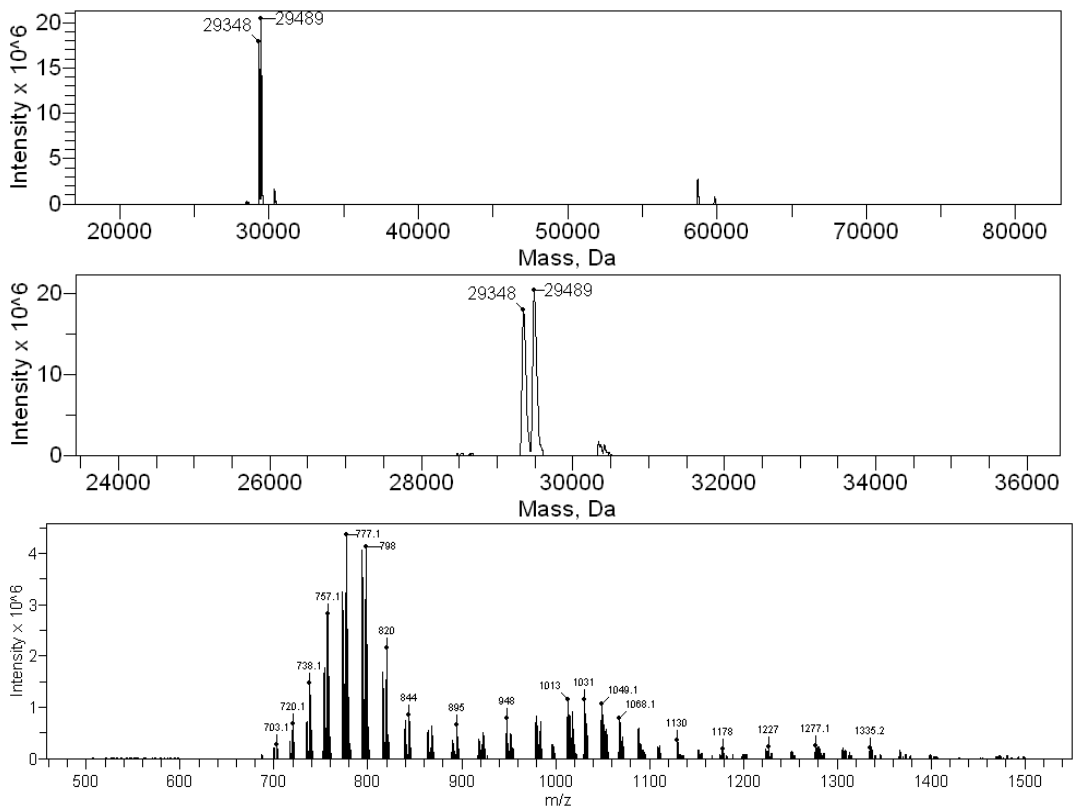
2 h



5 h

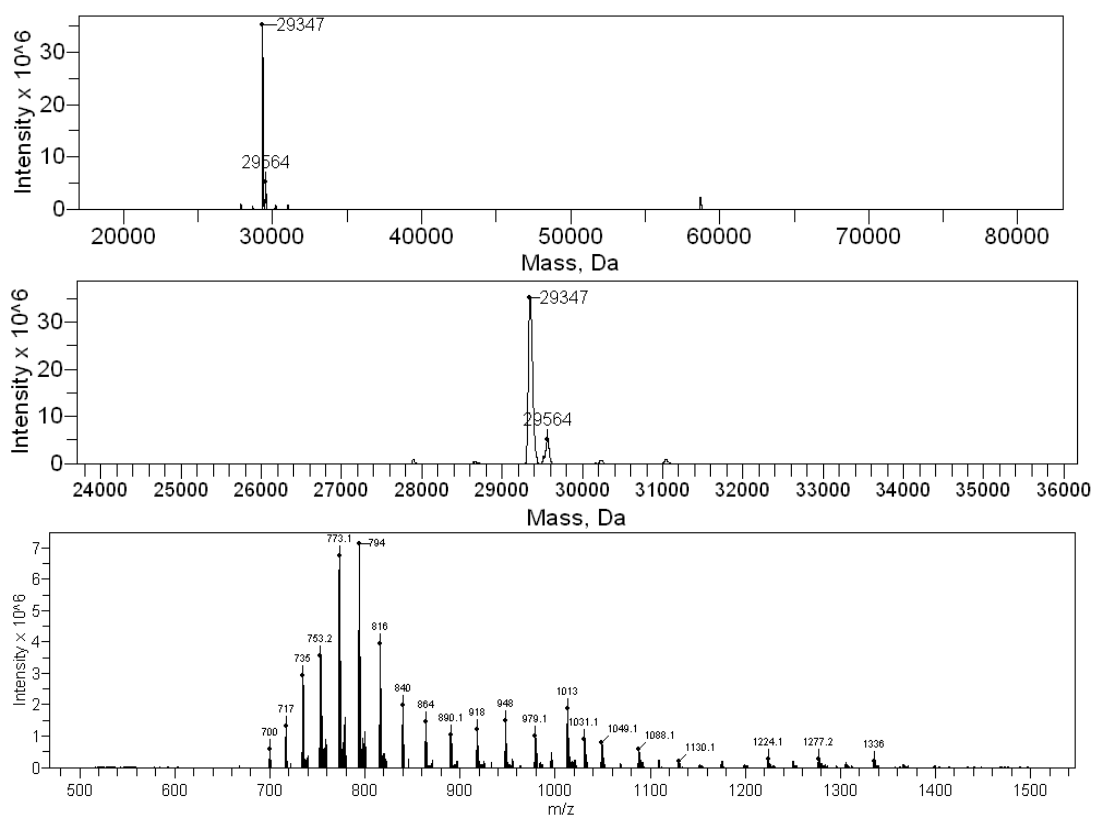


20 h

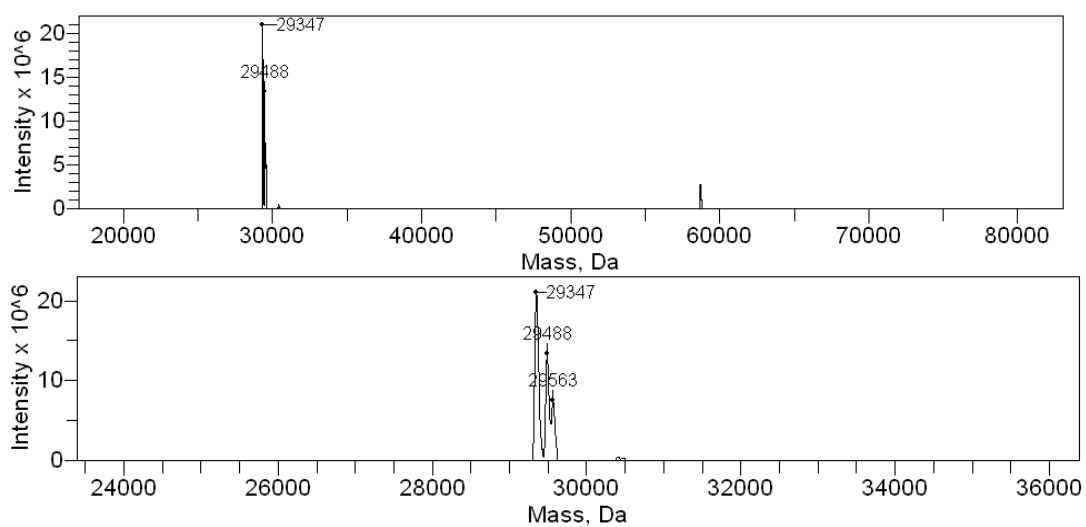


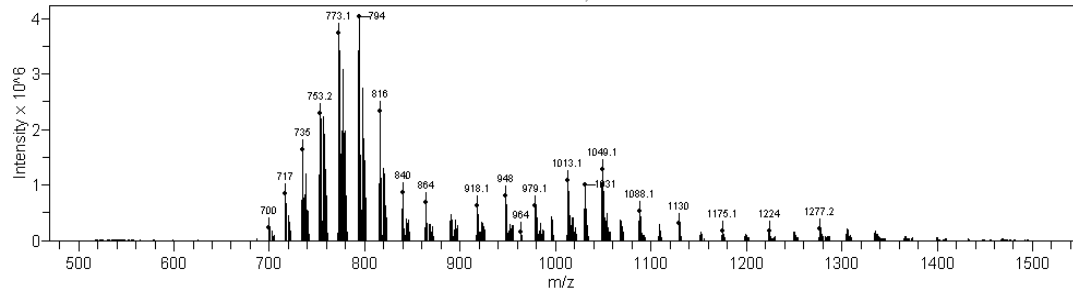
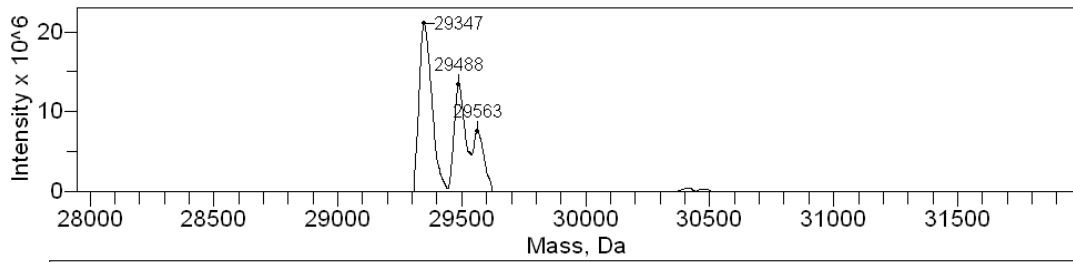
25 equivalents, 21 °C

2 h

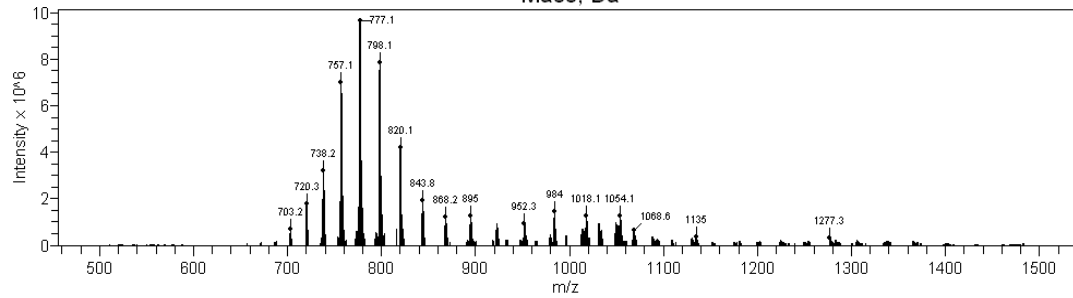
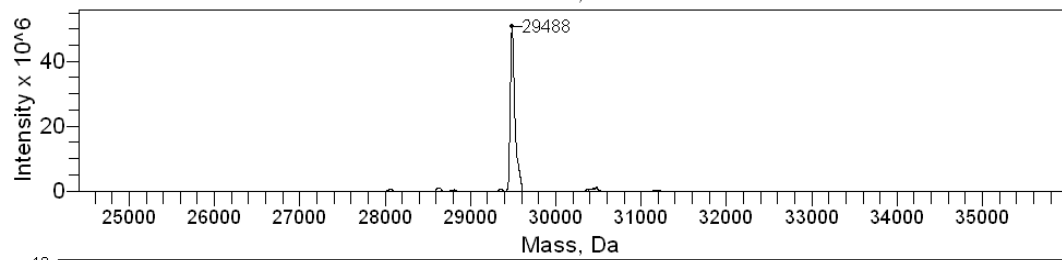
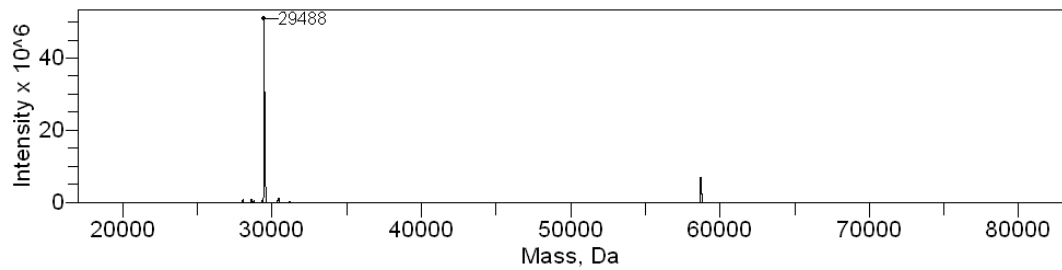


5 h



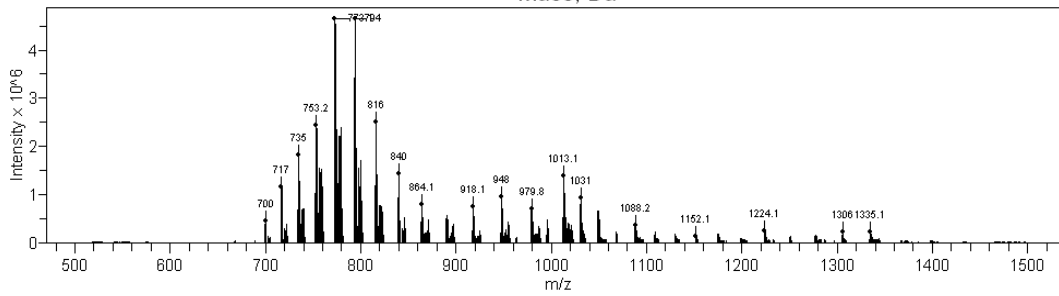
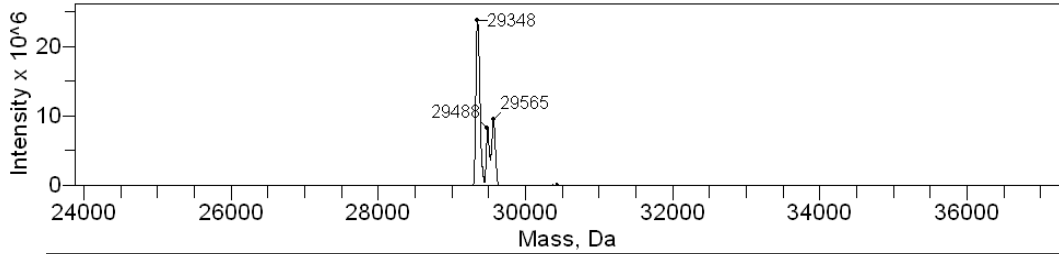
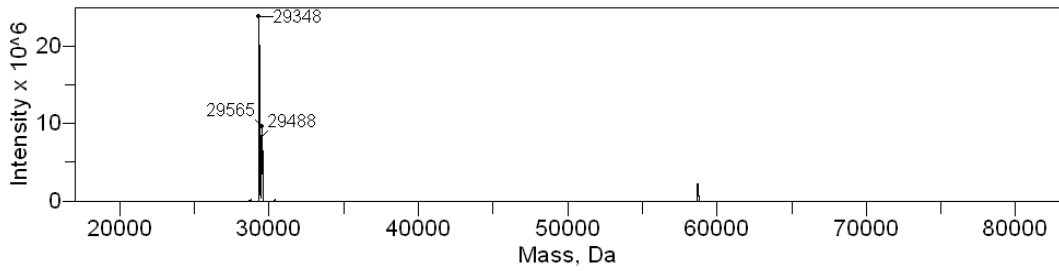


20 h

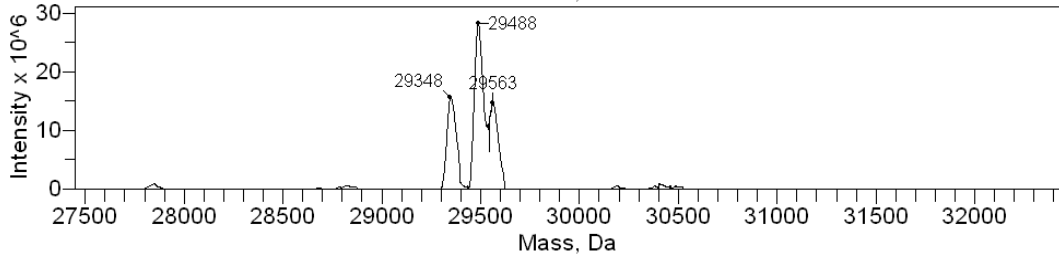
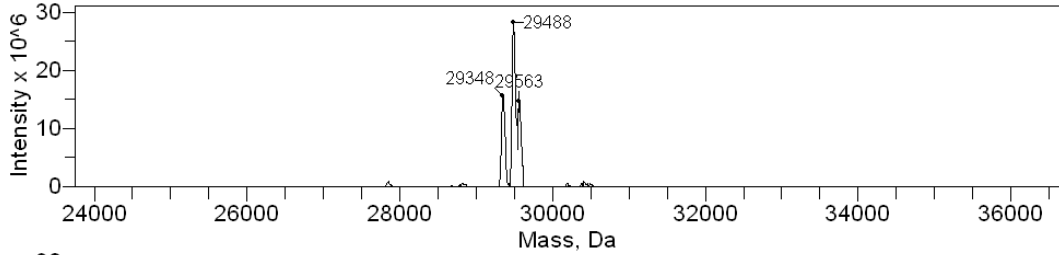
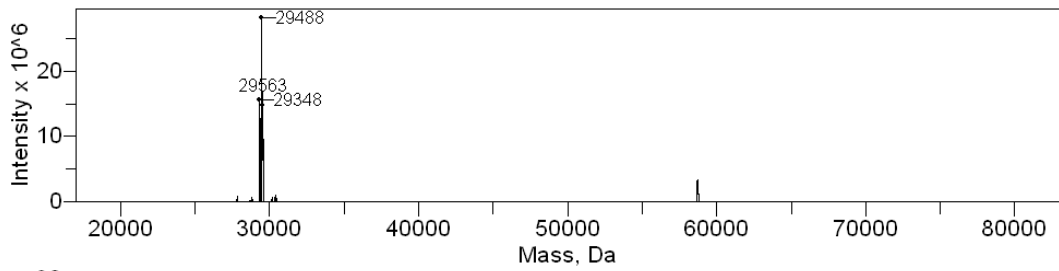


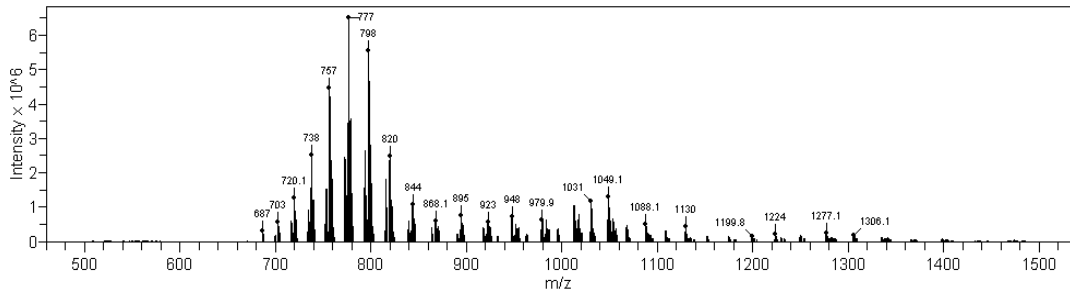
50 equivalents, 21 °C

2 h

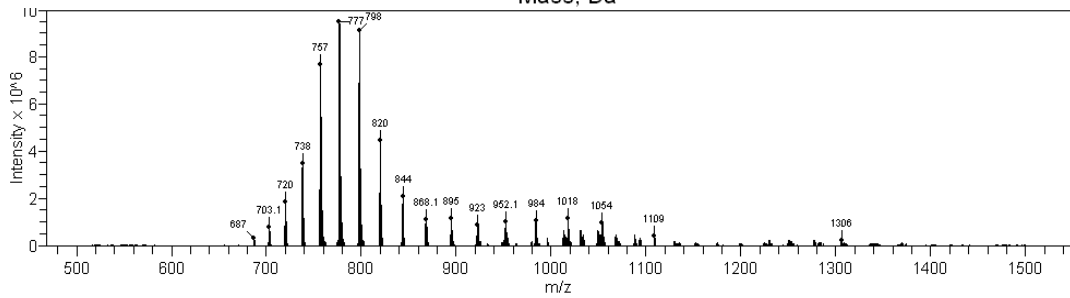
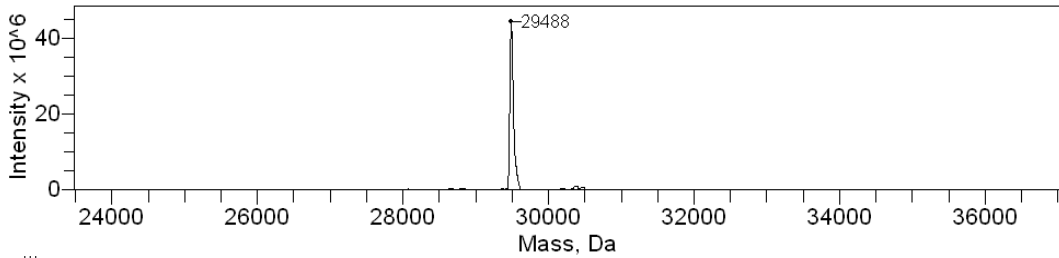
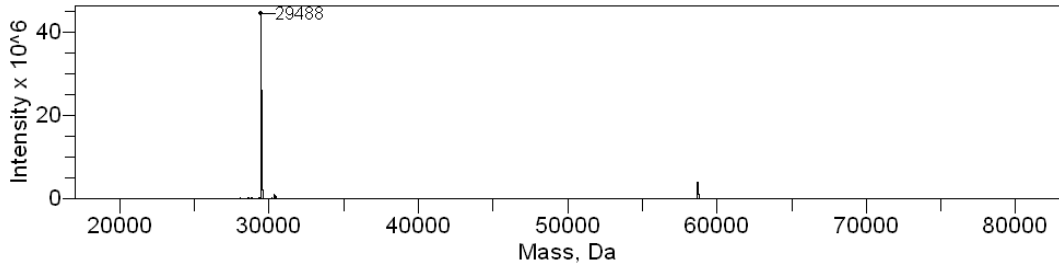


5 h

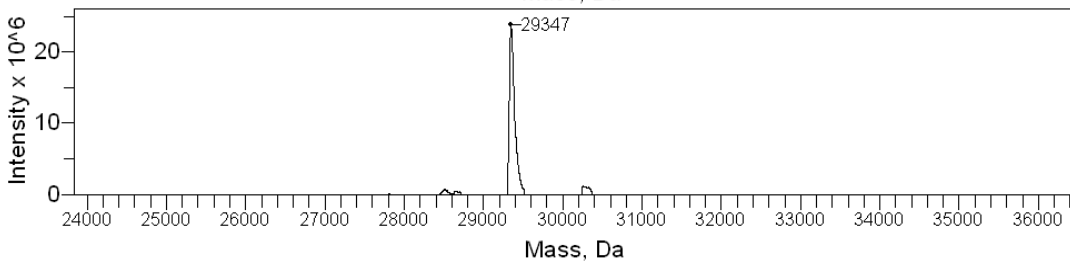
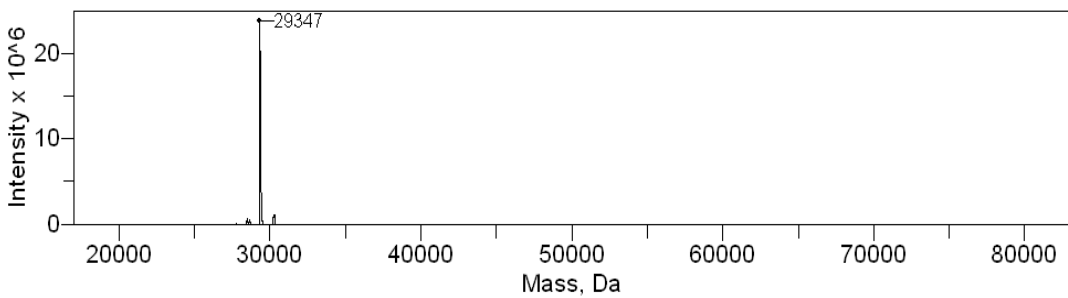


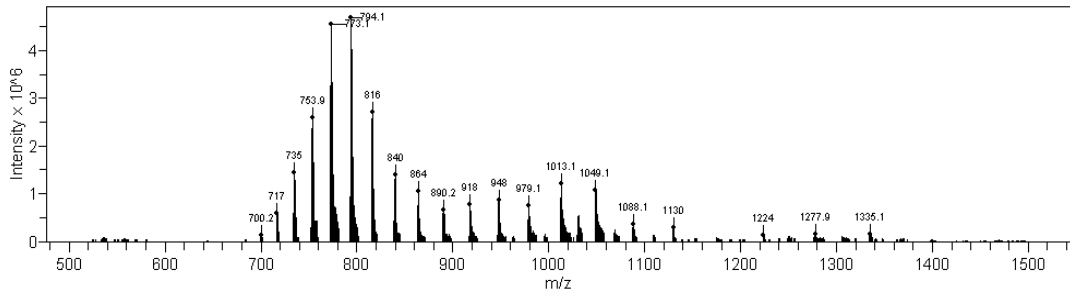


20 h

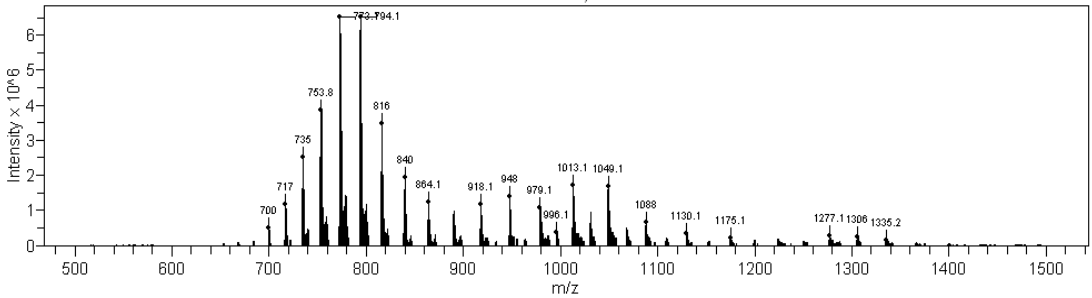
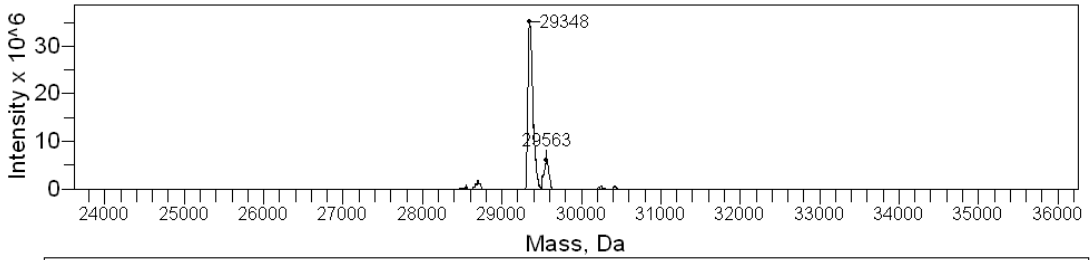
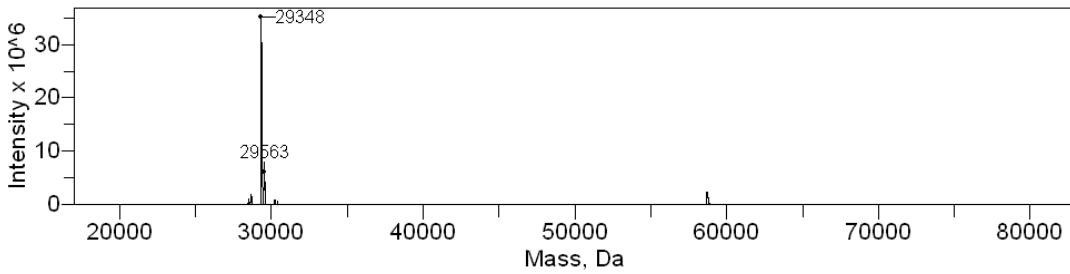


10 equivalents, 4 °C, 20 h

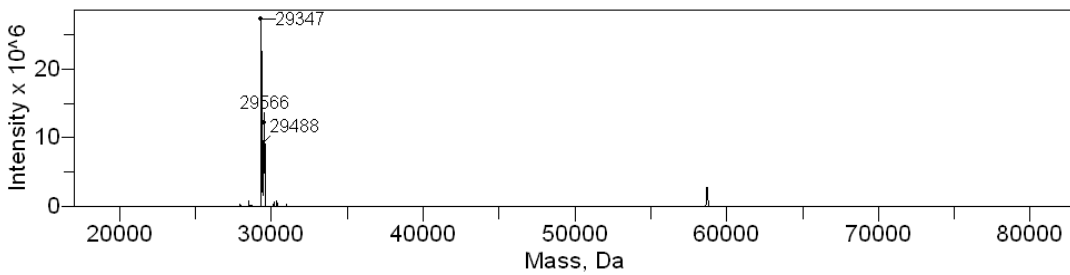


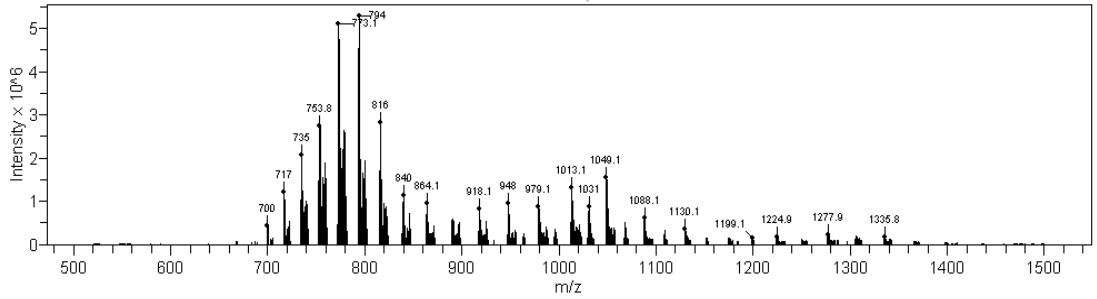
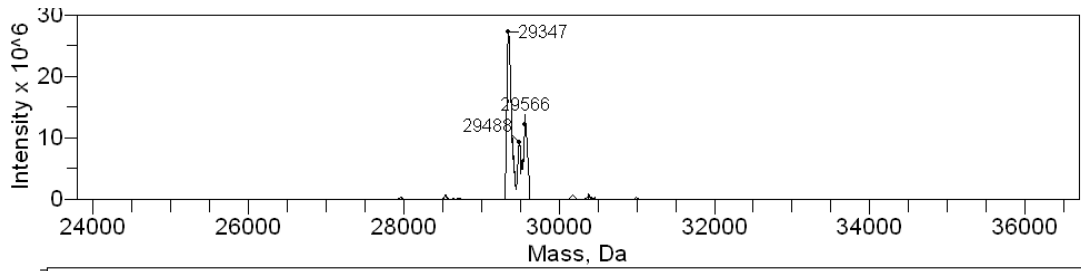


25 equivalents, 4 °C, 20 h

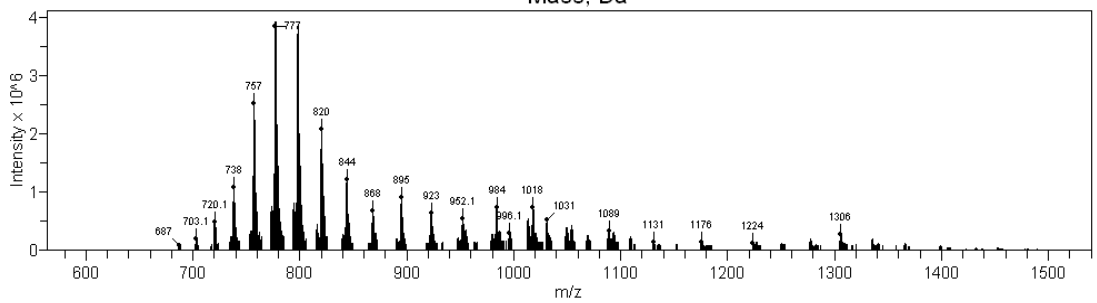
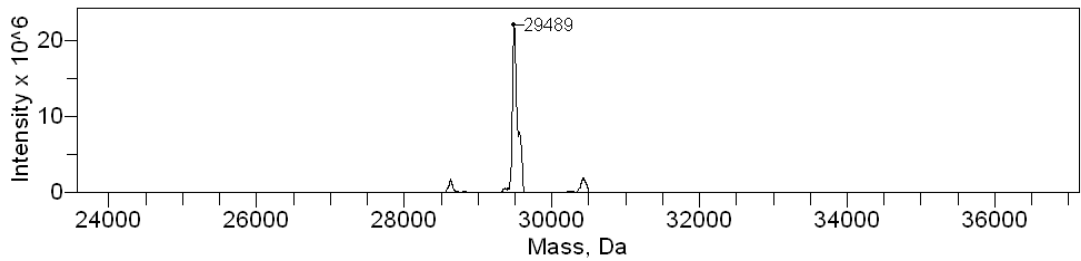
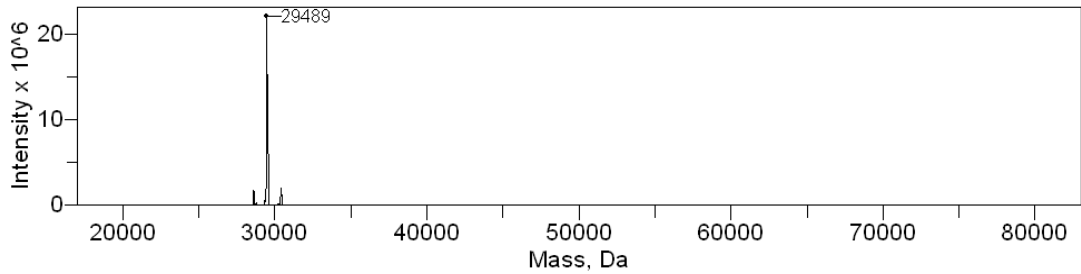


50 equivalents, 4 °C, 20 h

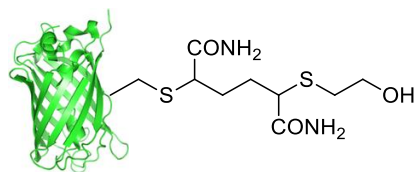




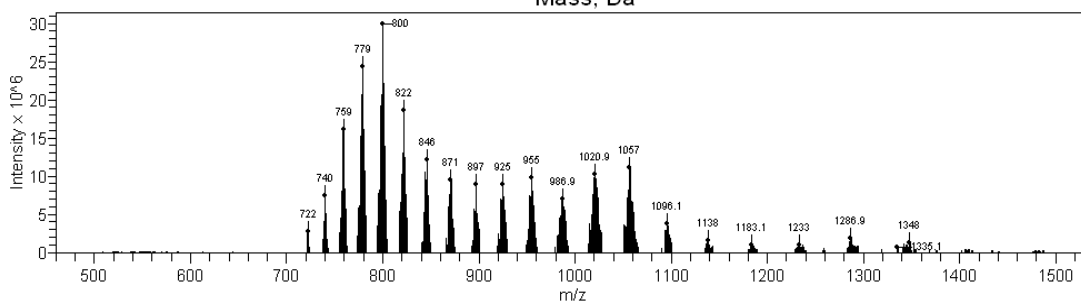
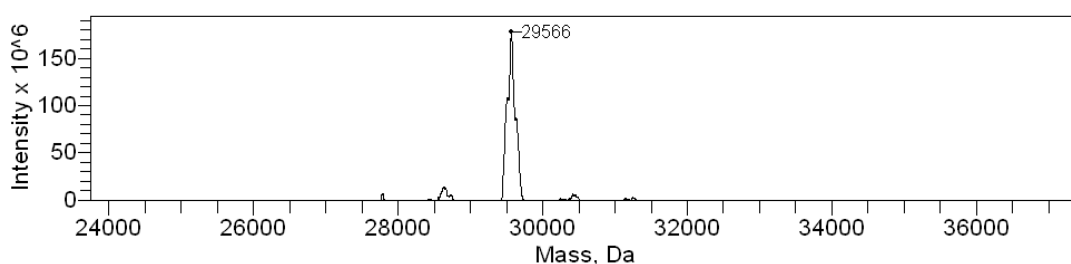
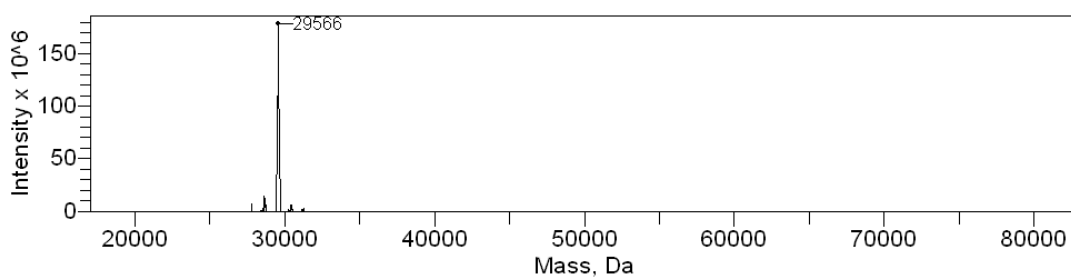
50 equivalents, 37 °C, 2 h



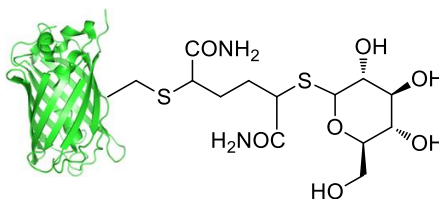
GFP (S147C, Sulf) 2-mercaptoethanol conjugate **46**



2-Mercaptoethanol (5 μ L, 680 mM solution in H₂O, 1000 equivalents) was added to a solution of GFP (S147C, Sulf) **42** (1.0 mg/mL). The mixture was vortexed for 1 s and maintained at 37 °C for 2.5 h and analysed by LCMS (Expected 29563, Observed 29566).

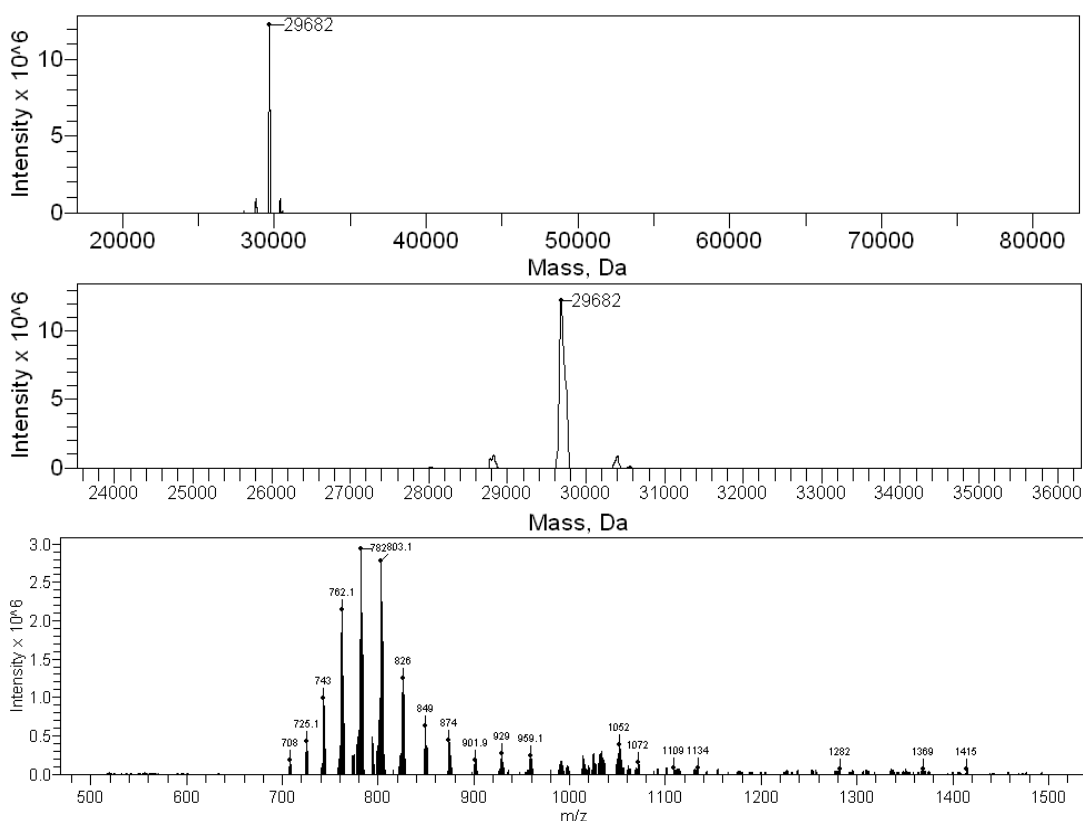


GFP (S147C, Sulf) thioglucose conjugate **47**

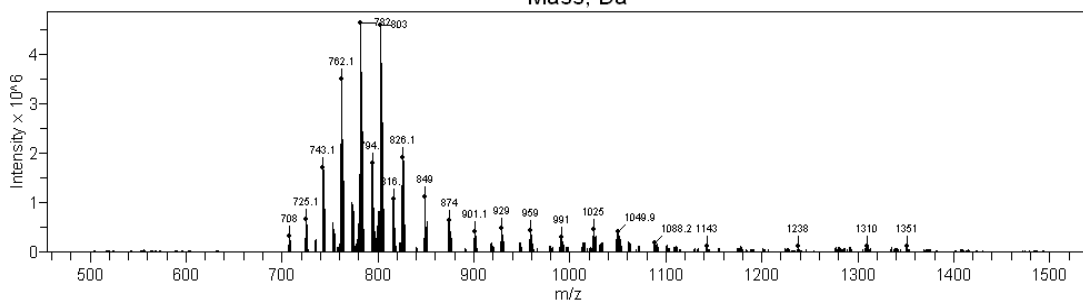
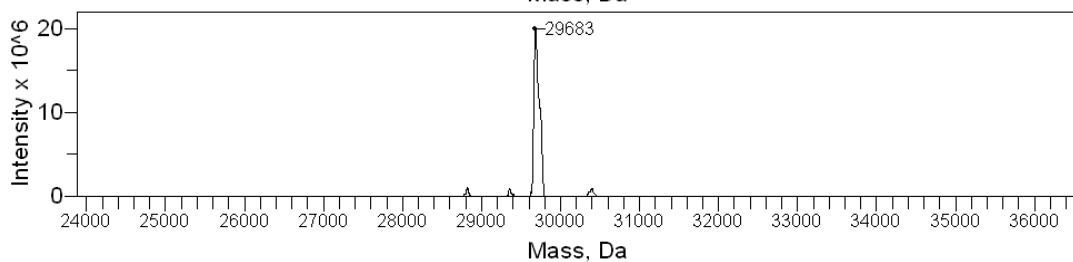
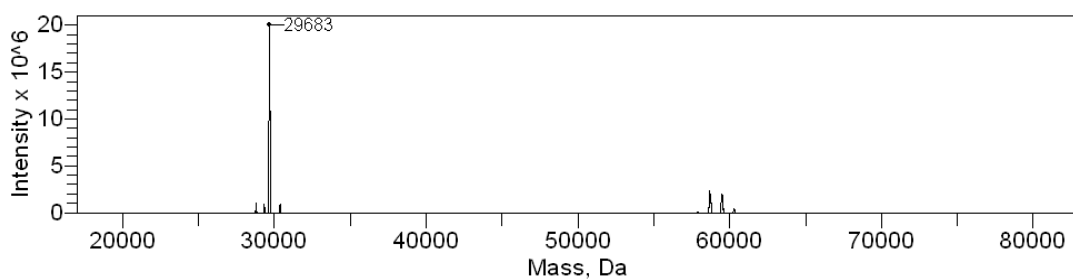


Thioglucose (10 or 100 equivalents), as a solution in water, was added to a solution of GFP (S147C, Sulf) **42** (100 μ L, 1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 21 $^{\circ}$ C. The mixture was vortexed for 1s, maintained at the required temperature (21 or 37 $^{\circ}$ C) for 5 h and analysed by LCMS (Expected 29681, Observed 29682).

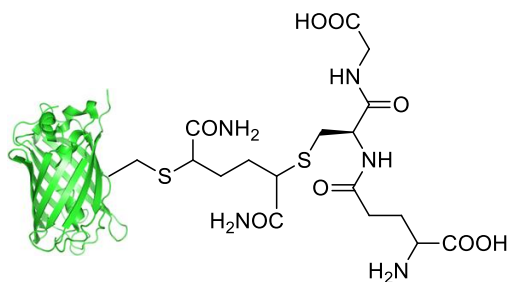
10 equivalents, 37 $^{\circ}$ C, 5 h



100 equivalents, 21 °C, 5 h

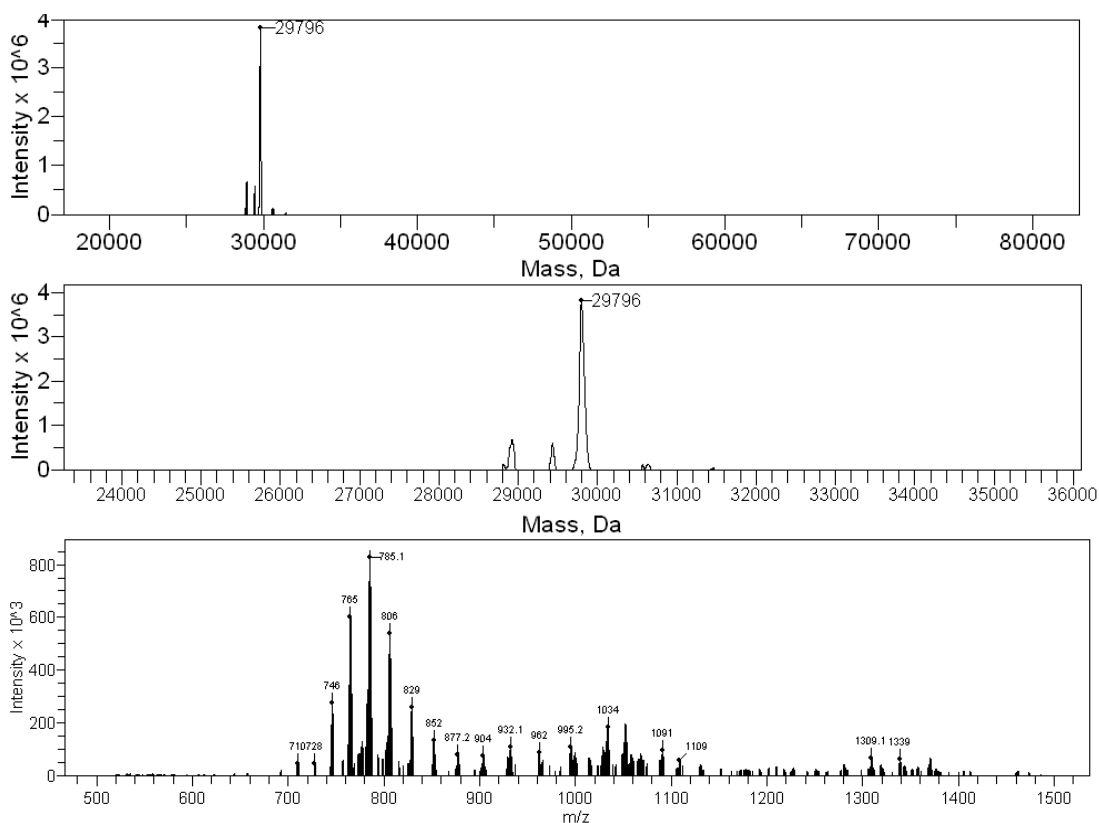


GFP (S147C, Sulf) Glutathione conjugate **48**

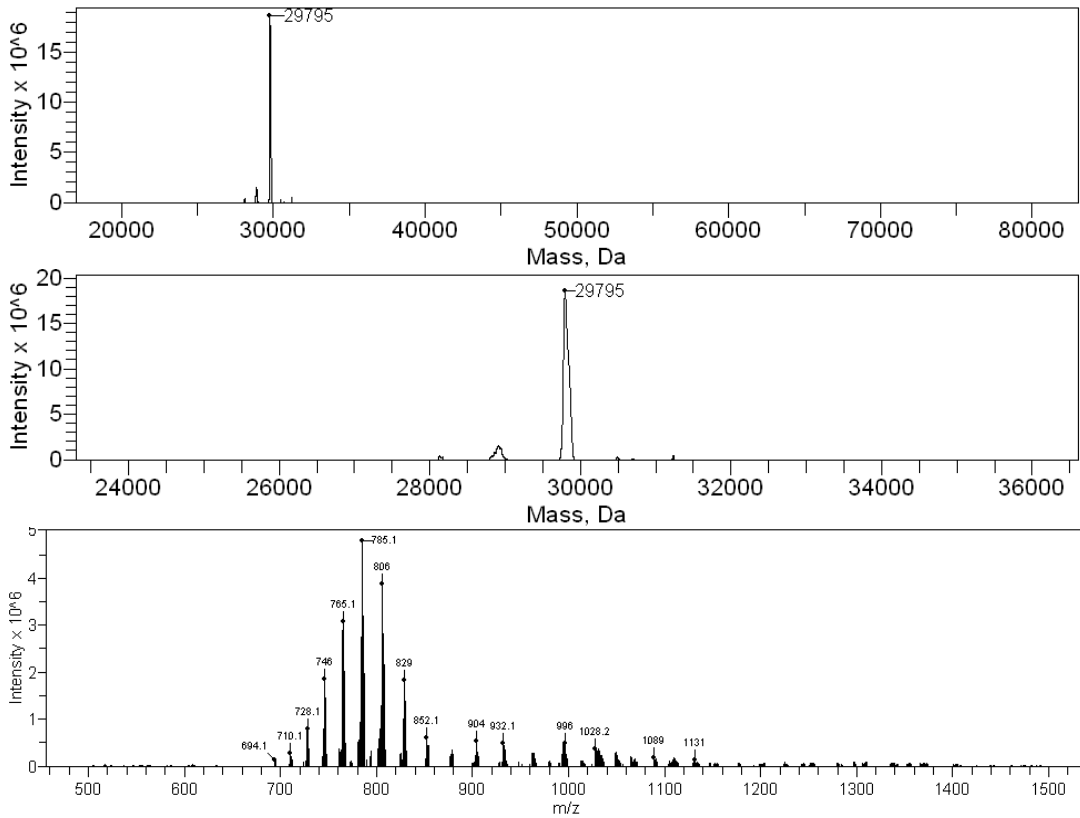


Glutathione (10, 100 or 200 equivalents), as a solution in water, was added to a solution of GFP (S147C, Sulf) **42** (100 μ L, 1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 21 $^{\circ}$ C. The mixture was vortexed for 1s, maintained at the required temperature (21 or 37 $^{\circ}$ C) for the prescribed time (2 h or 5 h) and analysed by LCMS (Expected 29792, Observed 29795/29796).

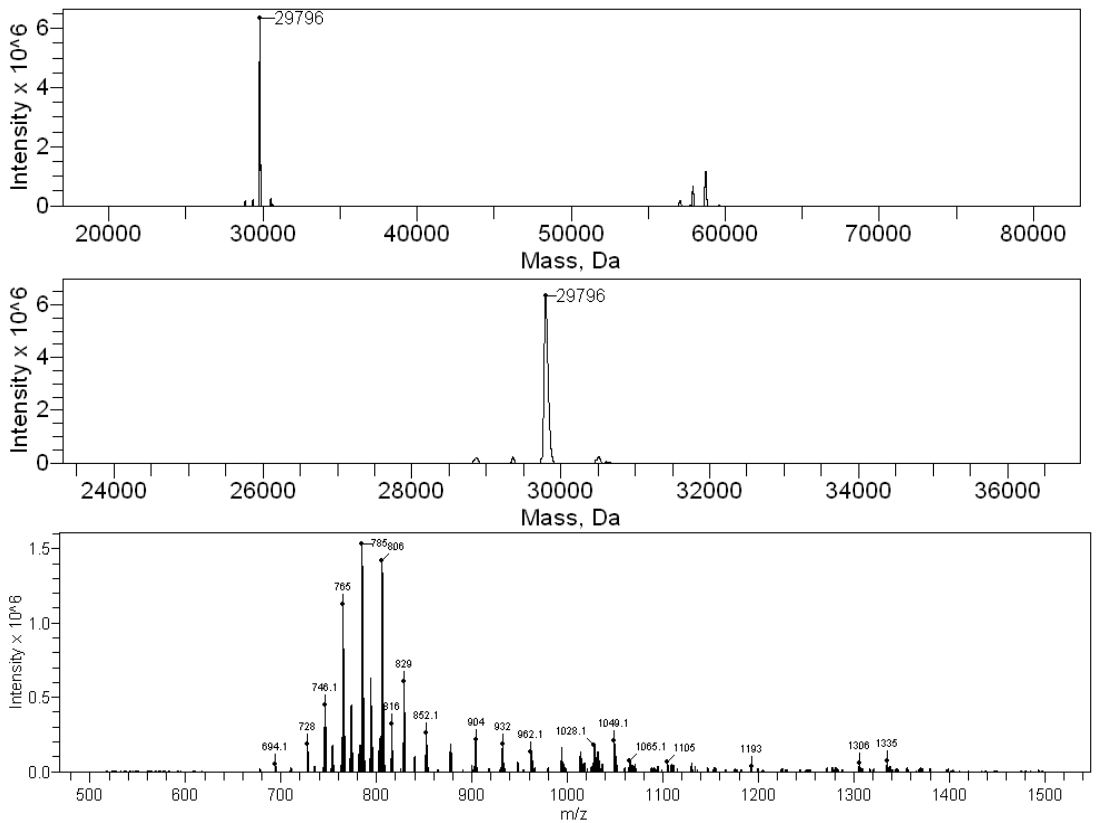
10 equivalents, 37 $^{\circ}$ C, 5 h



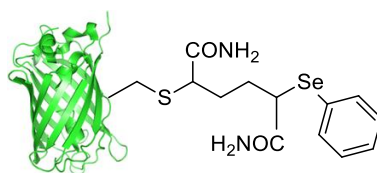
100 equivalents, 21 °C, 5 h



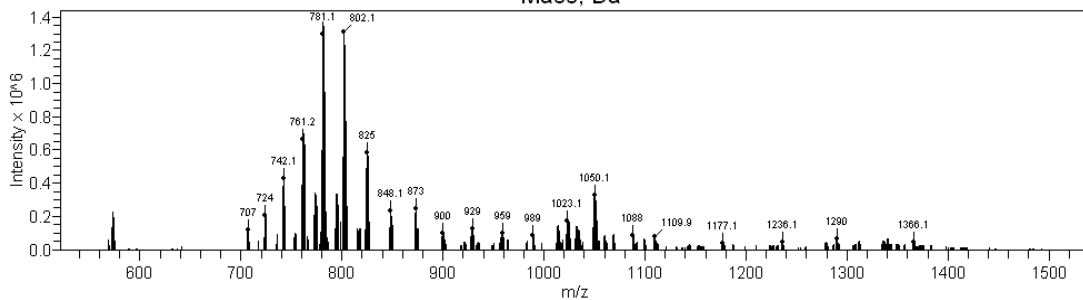
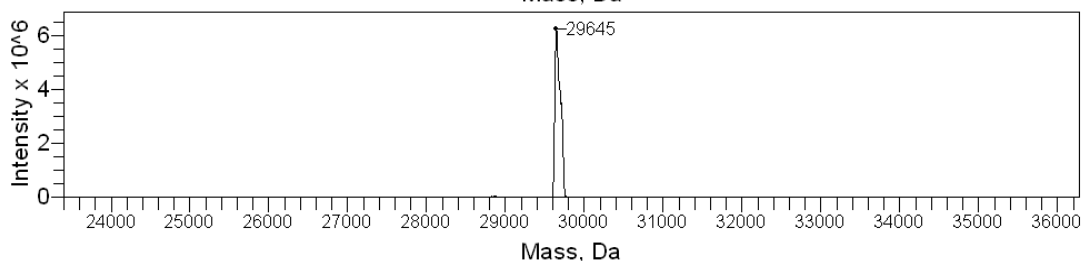
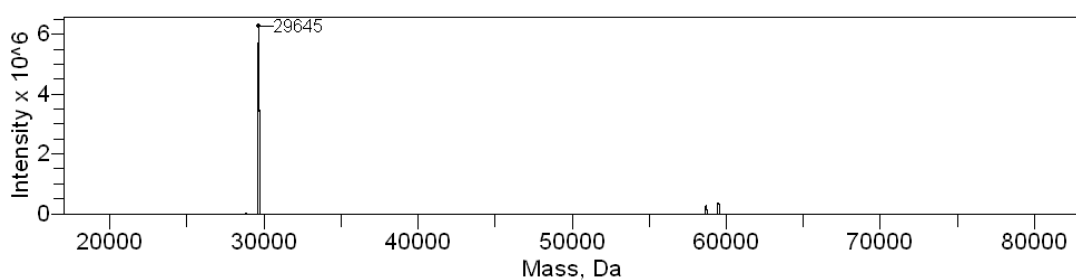
200 equivalents, 37 °C, 2 h



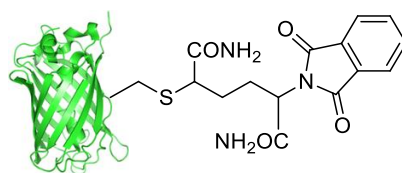
GFP (S147C, Sulf) benzeneselenol conjugate **49**



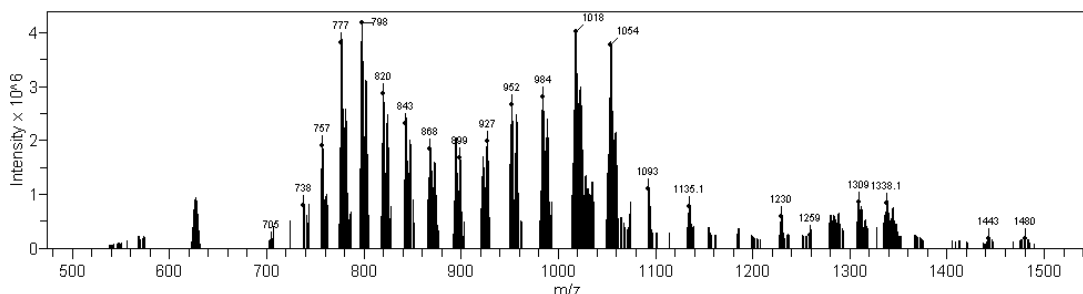
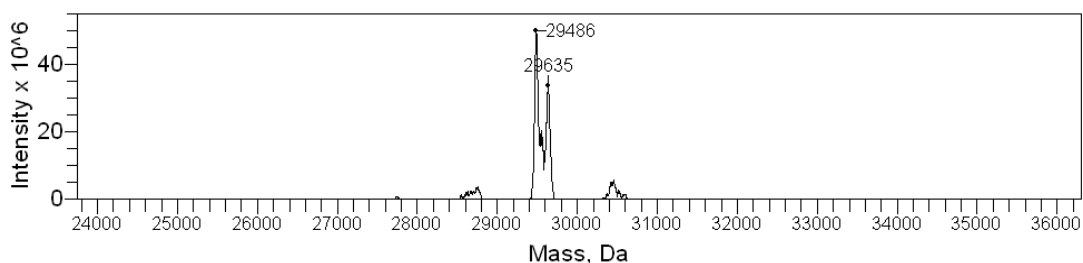
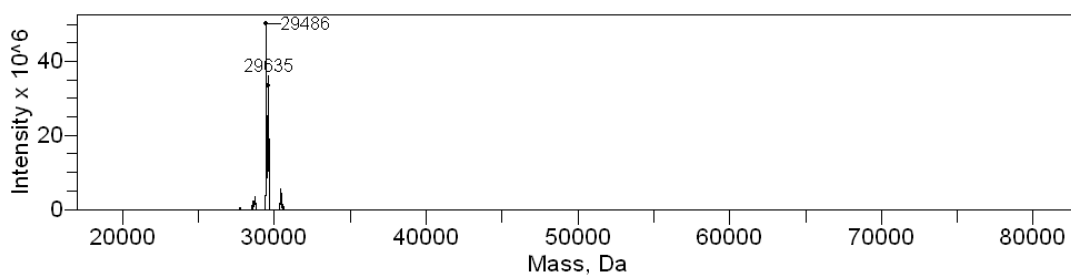
Benzeneselenol (5 μ L, 68 mM solution in DMF, 100 equivalents) was added to a solution of GFP (S147C, Sulf) **42** (100 μ L, 1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 21 $^{\circ}$ C. The mixture was vortexed for 1s, maintained at the 37 $^{\circ}$ C for 5 h and analysed by LCMS (Expected 29642, Observed 29645).



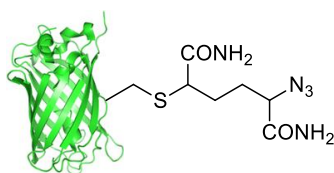
GFP (S147C, Sulf) phthalimide conjugate **50**



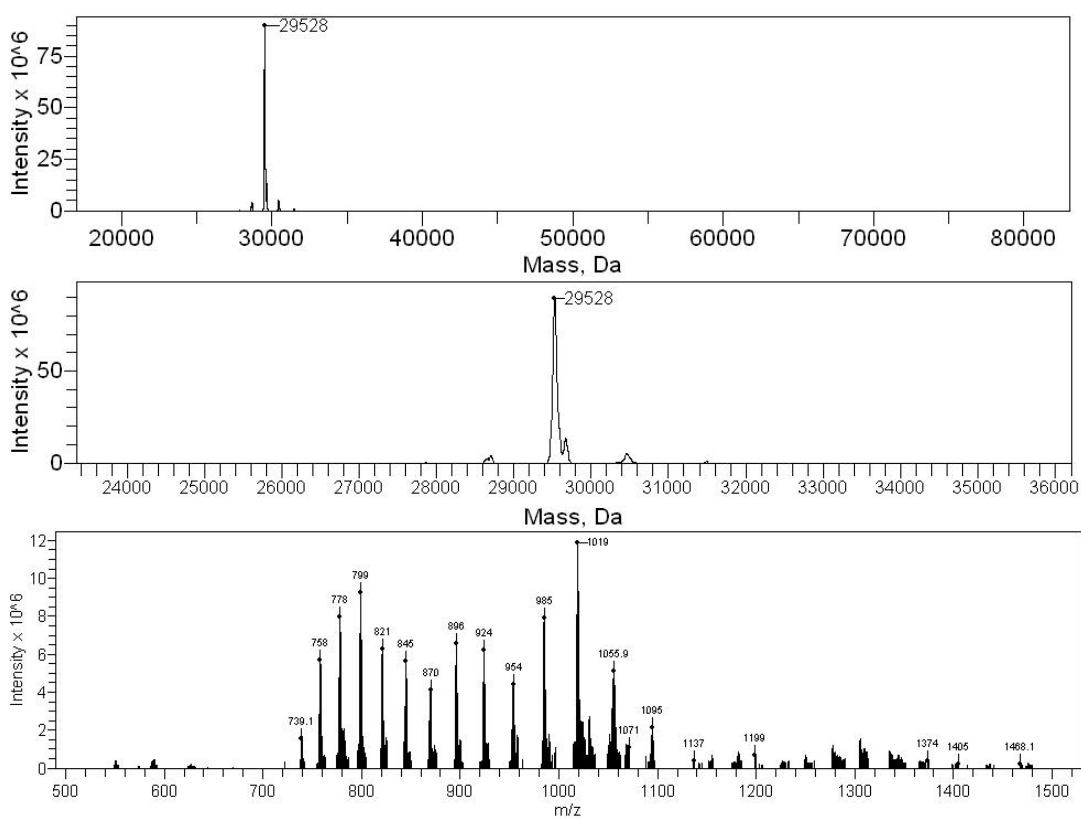
Potassium phthalimide (10 μ L, 340 mM solution in DMF, 1000 equivalents) was added to a solution of GFP (S147C, Sulf) **42** (100 μ L, 1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 21 $^{\circ}$ C. The mixture was vortexed for 1s and maintained at 37 $^{\circ}$ C for 2.5 h. Excess reagents were removed by repeated diafiltration into fresh buffer (100mM sodium phosphate, pH 8.0) using VivaSpin sample concentrators (GE Healthcare, 10,000 MWCO) and the mixture analysed by LCMS (Expected 29632, Observed 29635).



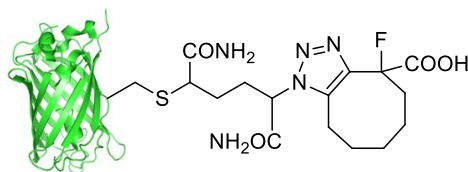
GFP (S147C, Azide) **51**



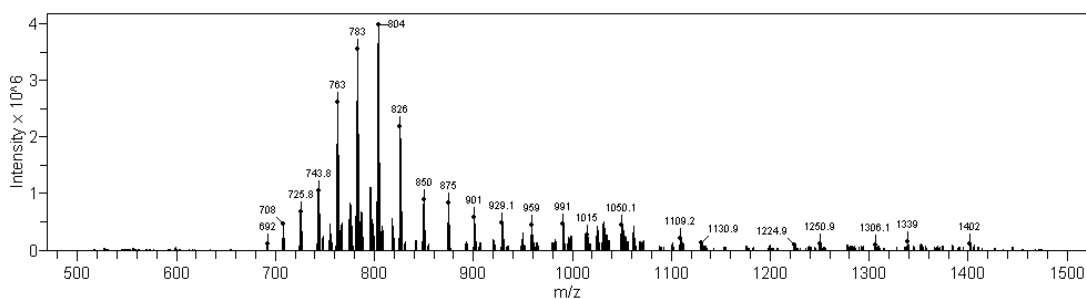
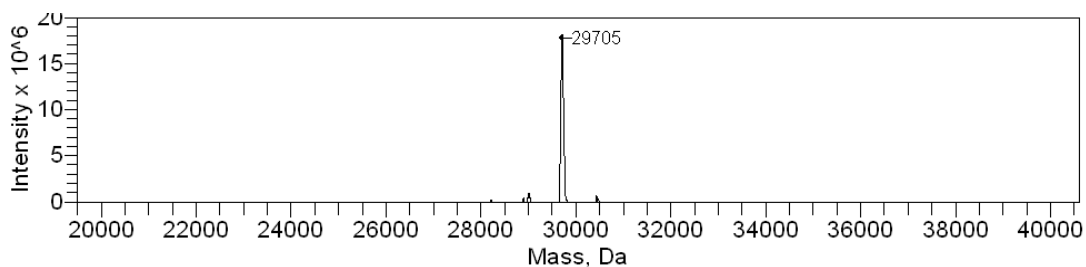
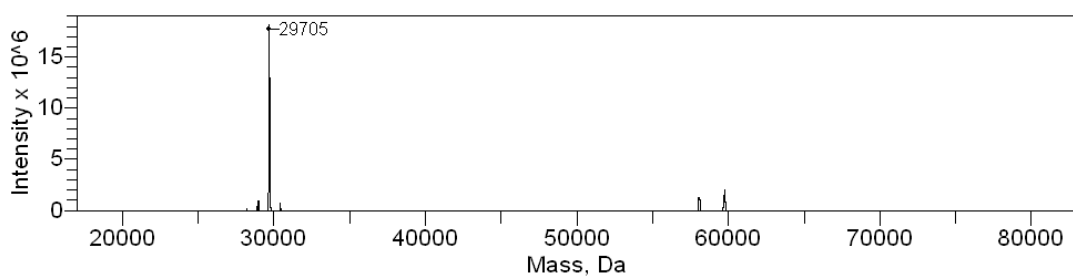
Sodium azide (10 μ L, saturated solution in water) was added to a solution of GFP (S147C, Sulf) **42** (100 μ L, 1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 21 $^{\circ}$ C. The mixture was vortexed for 1s, maintained at 37 $^{\circ}$ C for 2.5 h and analysed by LCMS (Expected 29528, Observed 29528).



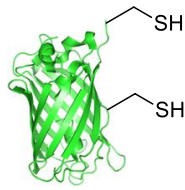
GFP (S147C, azide) octyne conjugate **53**



Octyne (**52**, 10 μ L, 34 mM solution in DMF, 100 equivalents) was added to a solution of GFP (S147C, azide) **42** (100 μ L, 1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 21 $^{\circ}$ C. The mixture was vortexed for 1 s and then maintained at 21 $^{\circ}$ C for 16 h. Excess reagents were removed by repeated diafiltration into fresh buffer (100mM sodium phosphate, pH 8.0) using VivaSpin sample concentrators (GE Healthcare, 10,000 MWCO) and the mixture analysed by LCMS (Expected 29698, Observed 29705).

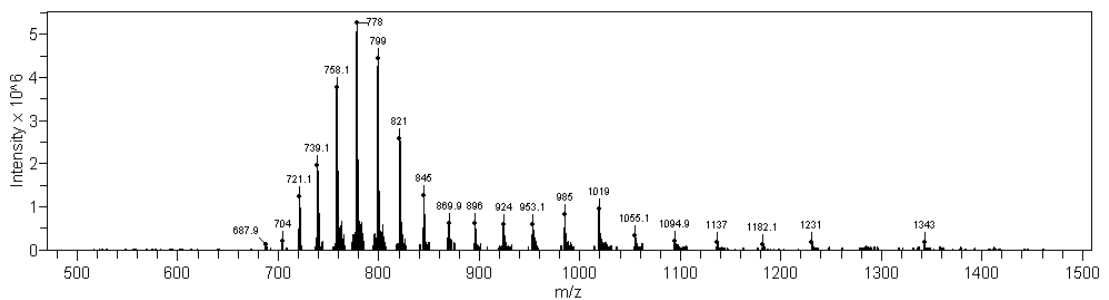
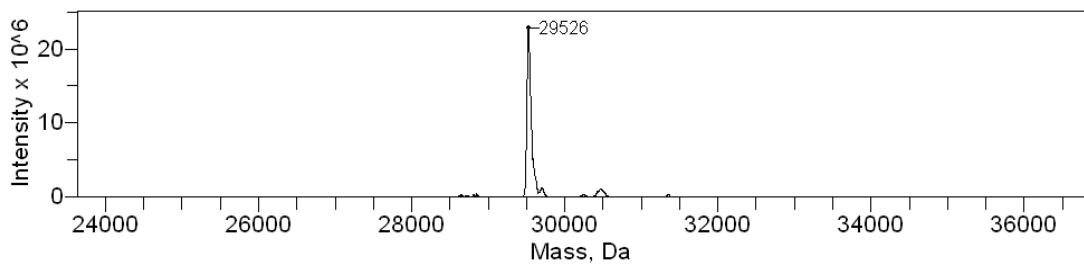
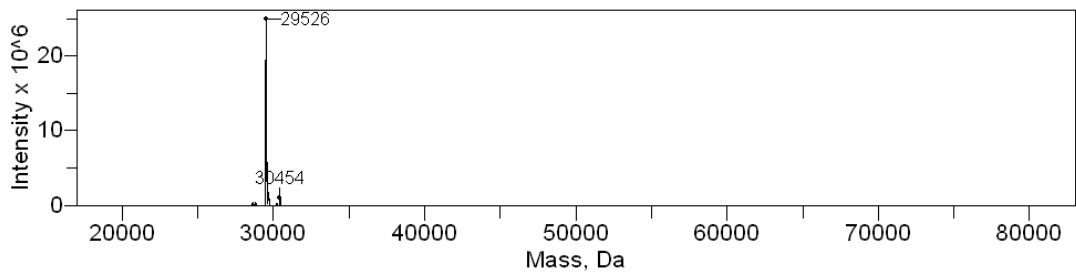


GFP (S147C, +CG) 54

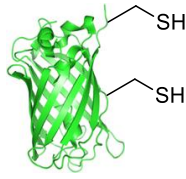


Sequence

MHHHHHSSGVDLG TENLYFQSMRKGEELFTGVVPILEVELDGDVNGHKFSVRGEGEG
DATNGKLT LKFICTTGKLPVPWPTLVTTLT YGVQCFARYPDHMKQHDFFKSAMPEGY
VQERTISFKDDGTYKTRAEVKFEGDTLVNRIELK GIDFKEDGNILGHKLEYNFNCHN
VYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSV
LSKDPNEKRDMVLLLEFVTAAGITHGMDELYKCG (Expected 29535, Observed
29526)

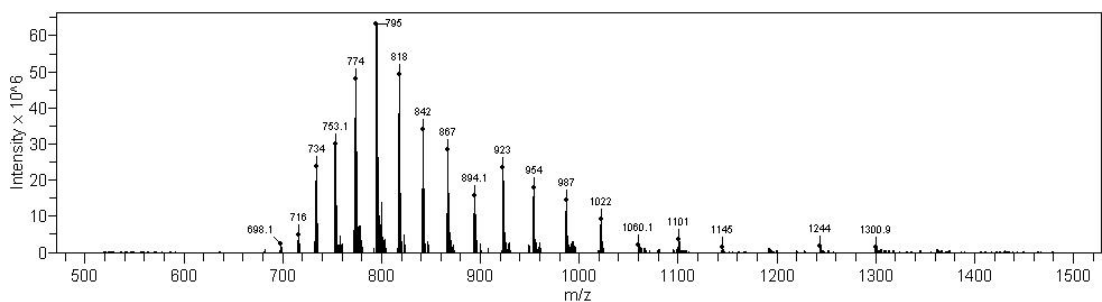
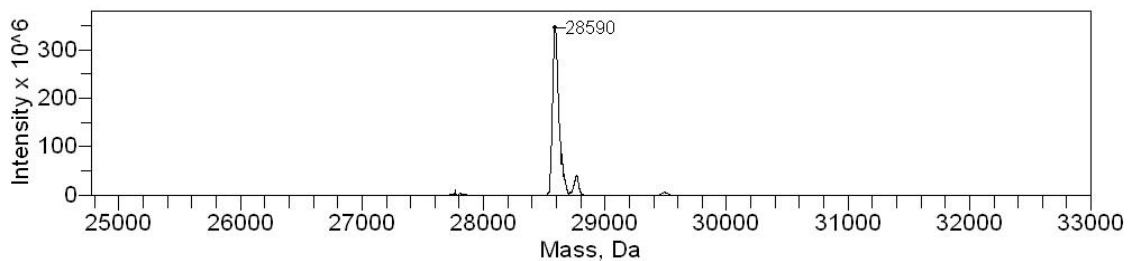
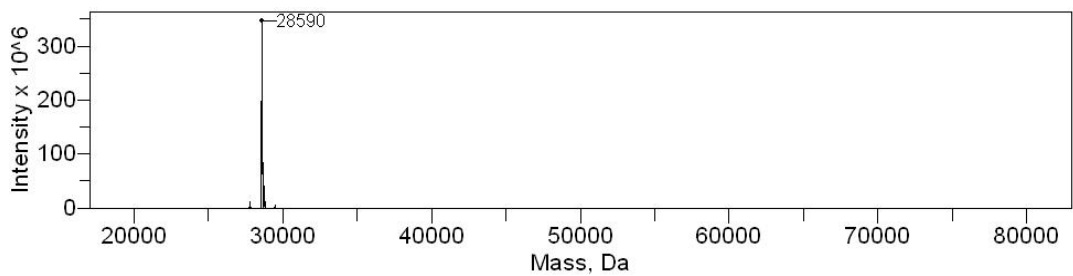


GFP (S147C, T230C, 233Δ) 55

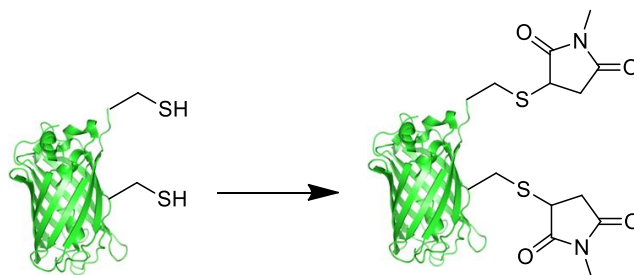


Sequence

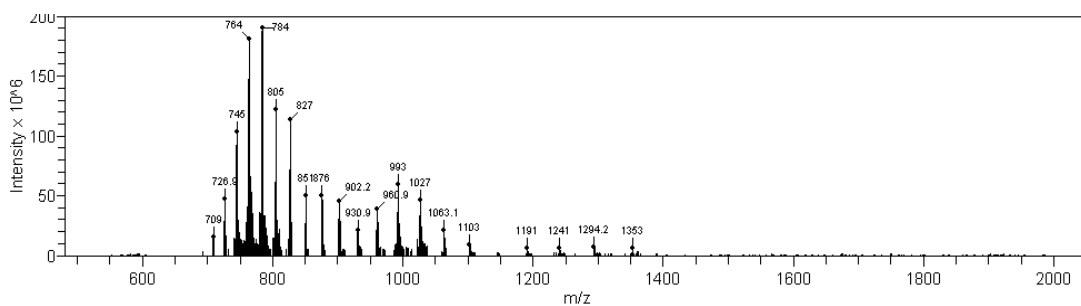
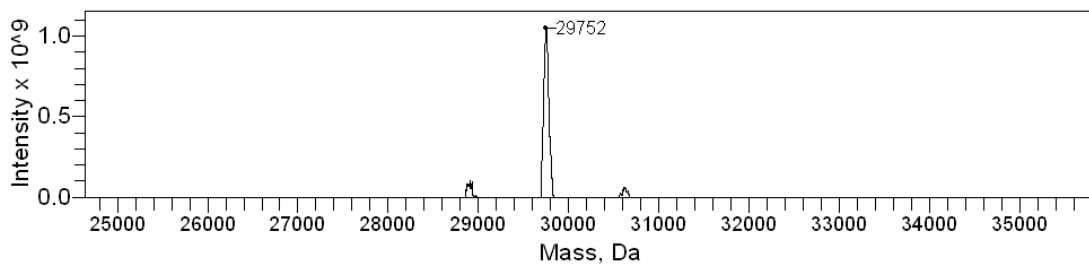
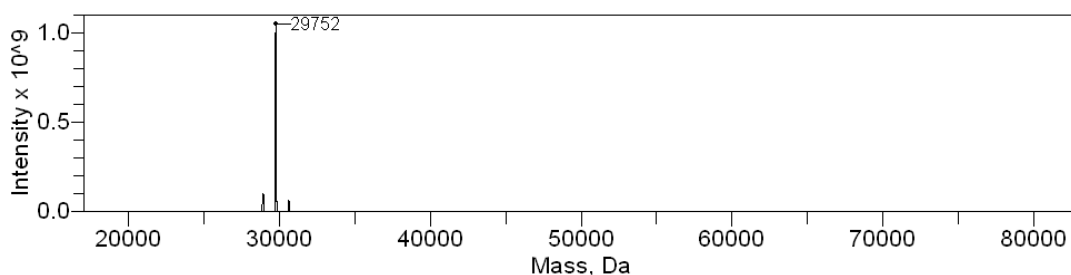
MHHHHHSSGVDLGTENLYFQSMRKGEELFTGVVPI LVELDGDVNGHKFSVRGEGEG
DATNGKLTLLKFICTTGKLPVPWPPTLVTTLTYGVQCFARYPDHMKQHDFFKSAMPEGY
VQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNCHN
VYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSV
LSKDPNEKRDHMLLEFVTAAGICHG (Expected 28597, Observed 28590)



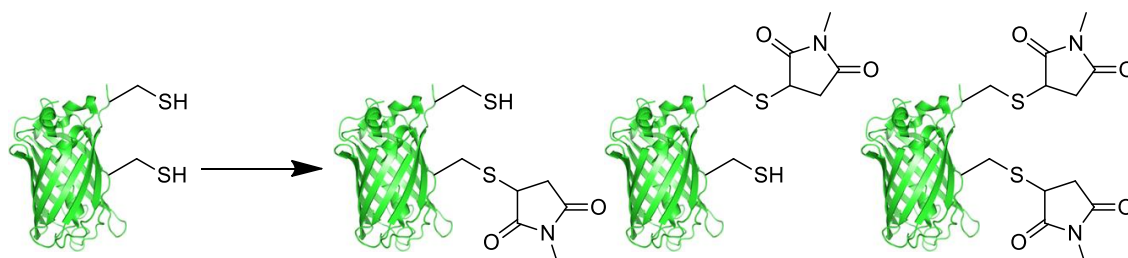
GFP (S147C *N*-methylmaleimide, +CG *N*-methylmaleimide) **56**



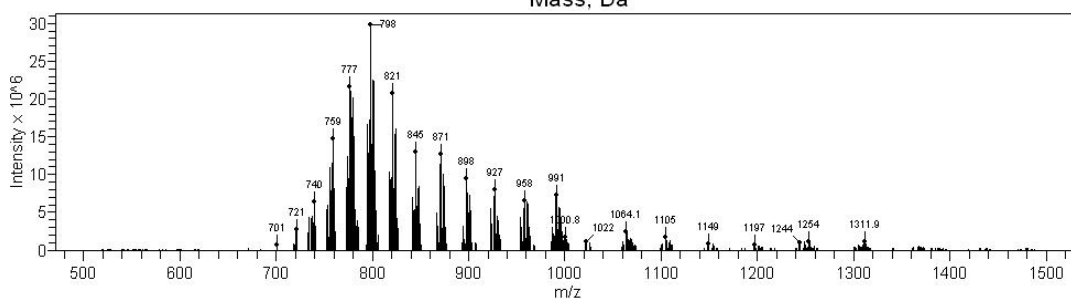
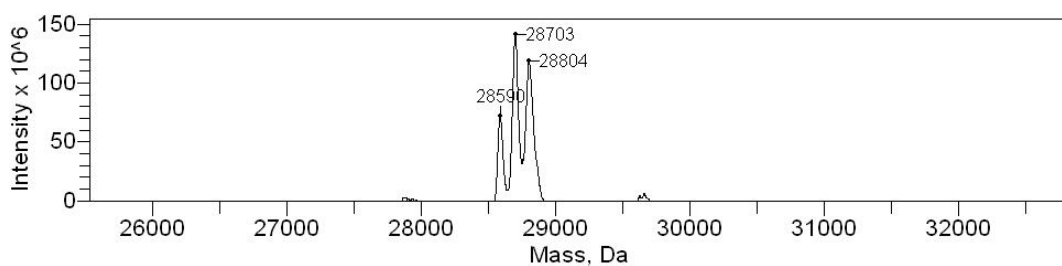
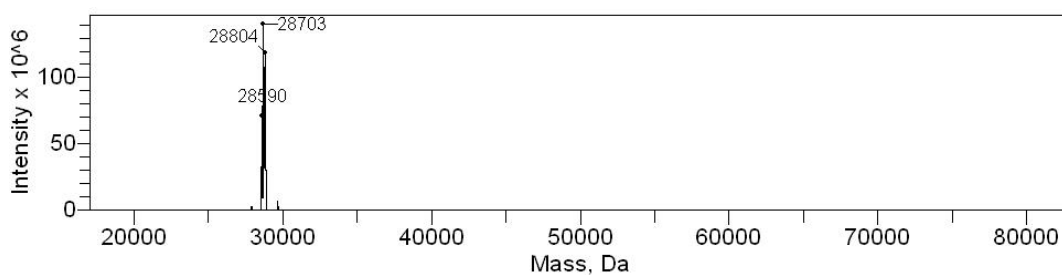
N-methylmaleimide (5 μ L, 6.8 mM solution in DMF, 1 equivalent) was added to a solution of GFP (S147C, +CG) **54** (100 μ L, 1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 21 $^{\circ}$ C. The mixture was vortexed for 1 s, maintained at 21 $^{\circ}$ C for 1 h and analysed by LCMS (Expected 29744, Observed 29752).



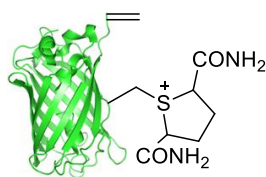
Reaction of GFP (S147C, T230C, 233Δ) **55** with *N*-methylmaleimide



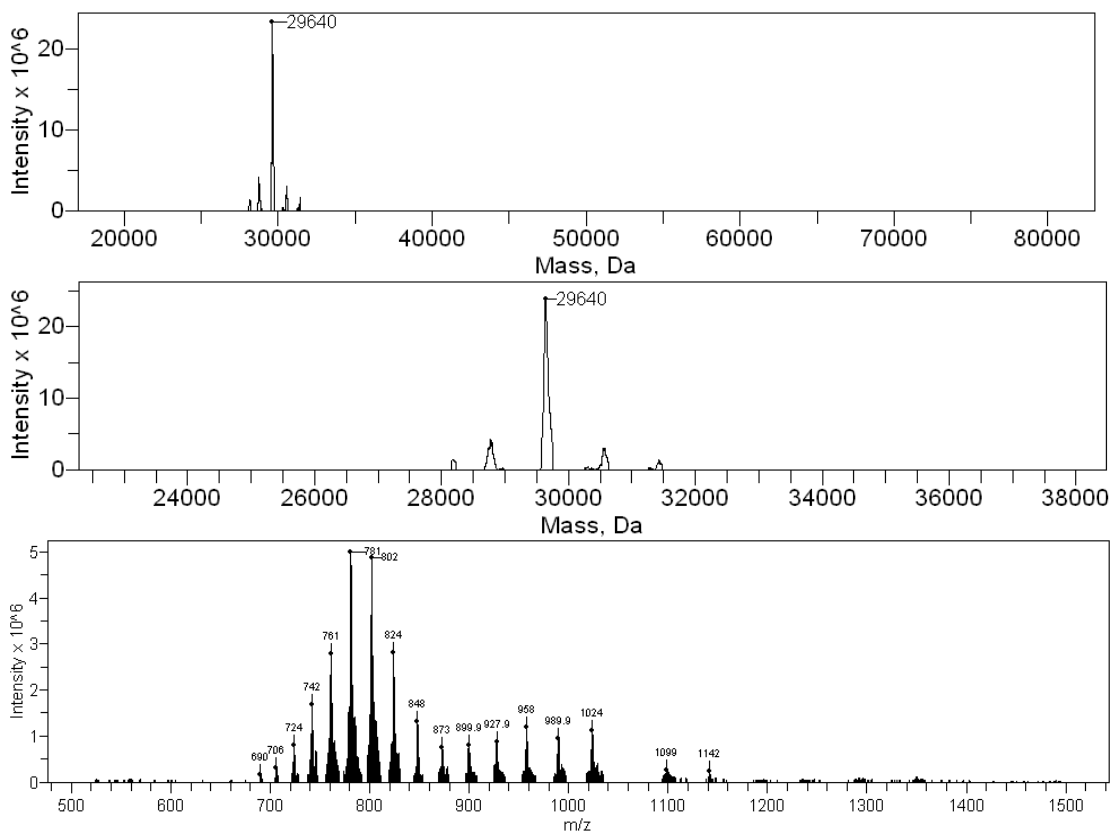
N-methylmaleimide (5 μ L, 6.8 mM solution in DMF, 1 equivalent) was added to a solution of GFP (S147C, T230C, 233 Δ) **55** (100 μ L, 1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 21 $^{\circ}$ C. The mixture was vortexed for 1 s, maintained at 21 $^{\circ}$ C for 1 h and analysed by LCMS (Expected 28701 and 28812, Observed 28703 and 28804).



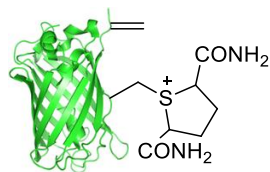
GFP (S147C Sulf, +CG Dha) **58**



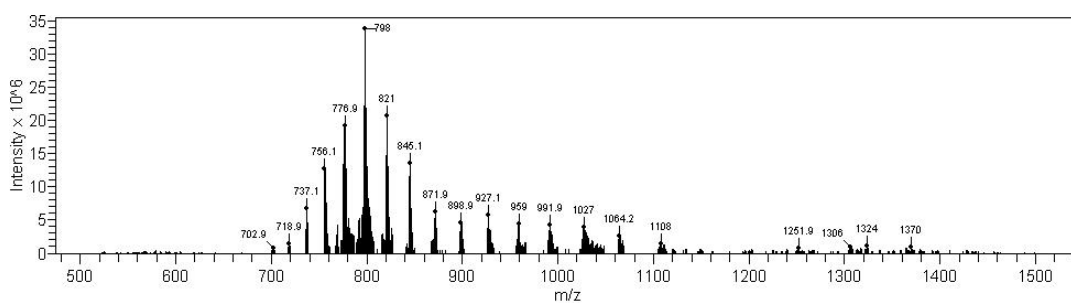
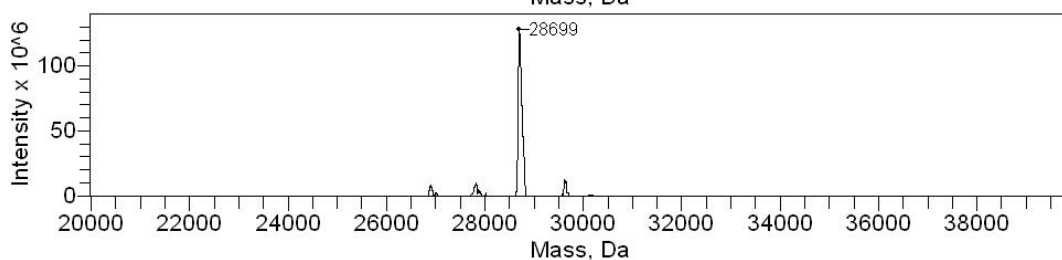
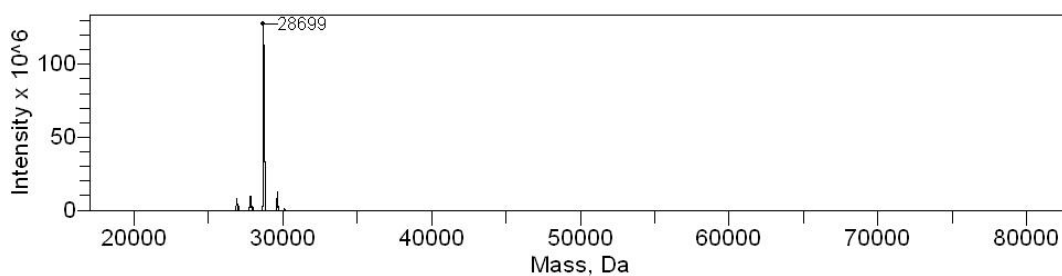
2,5-Dibromohexanediamide (**14**, 30 μL , 510 mM solution in DMF, 1500 equivalents) was added to a solution of GFP (S147C, +CG) **54** (300 μL , 1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 21 $^{\circ}\text{C}$. The mixture was vortexed for 1 s, maintained at 37 $^{\circ}\text{C}$ for 2 h and analysed by LCMS (Expected 29633, Observed 29640).



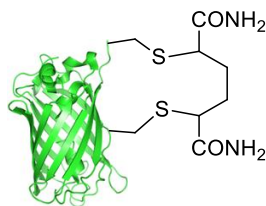
GFP (S147C Sulf, T230C Dha) **59**



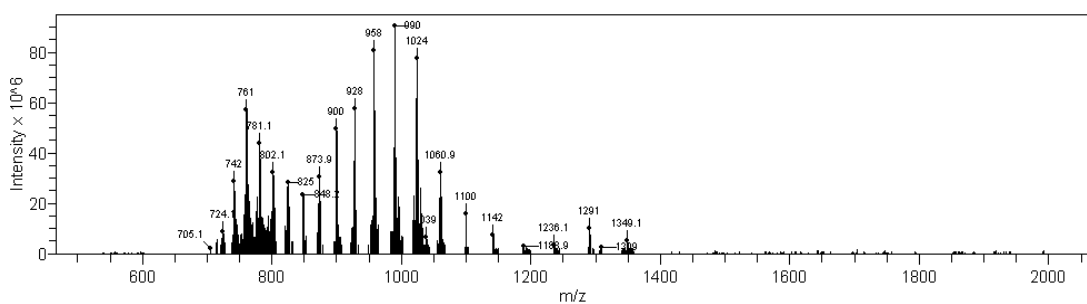
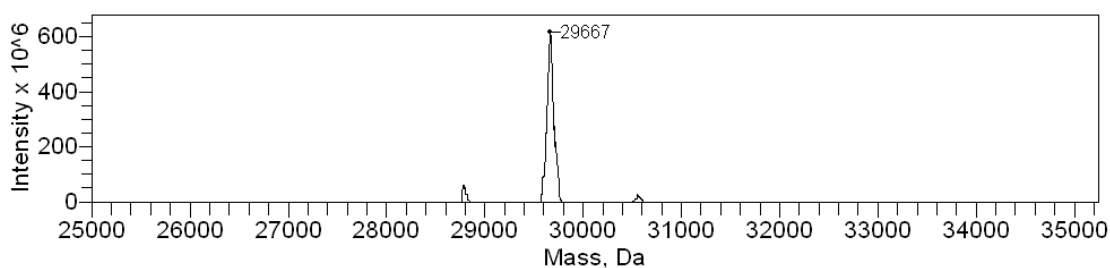
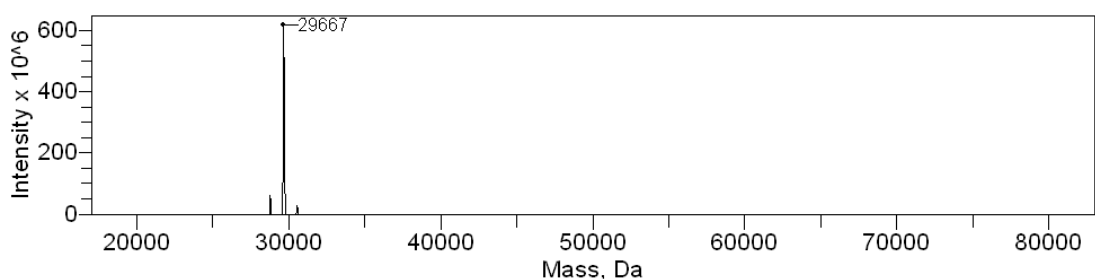
2,5-Dibromohexanediamide **14** (30 μ L, 17 mM solution in DMF, 50 equivalents) was added to a solution of GFP (S147C, T230C, 233 Δ) **55** (300 μ L, 1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 21 $^{\circ}$ C. The mixture was vortexed for 1 s, maintained at 37 $^{\circ}$ C for 2 h and analysed by LCMS (Expected 28697, Observed 28699).



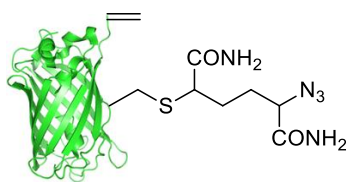
GFP (S147C, +CG) hexanediamide bisthioether **60**



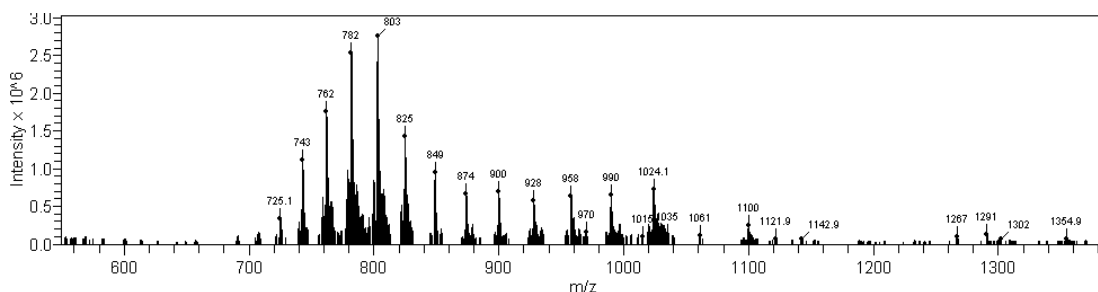
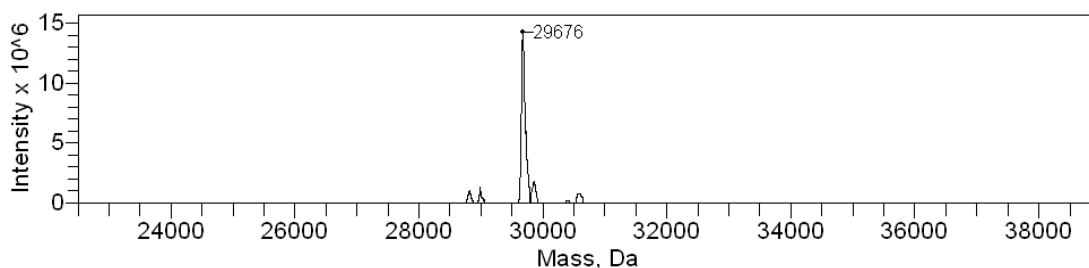
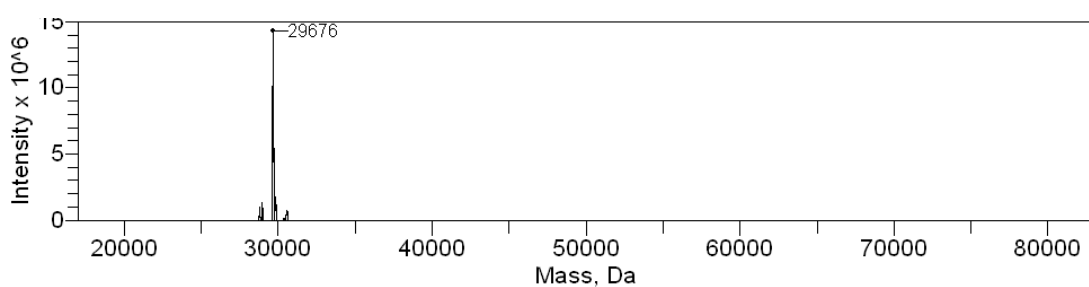
2,5-Dibromohexanediamide (**14**, 30 μ L, 17 mM solution in DMF, 50 equivalents) was added to a solution of GFP (S147C, +CG) **54** (300 μ L, 1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 21 $^{\circ}$ C. The mixture was vortexed for 1 s, maintained at 37 $^{\circ}$ C for 2 h and analysed by LCMS (Expected 29666, Observed 29667).



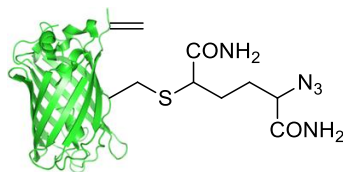
GFP (S147C Azide, +CG Dha) **61**



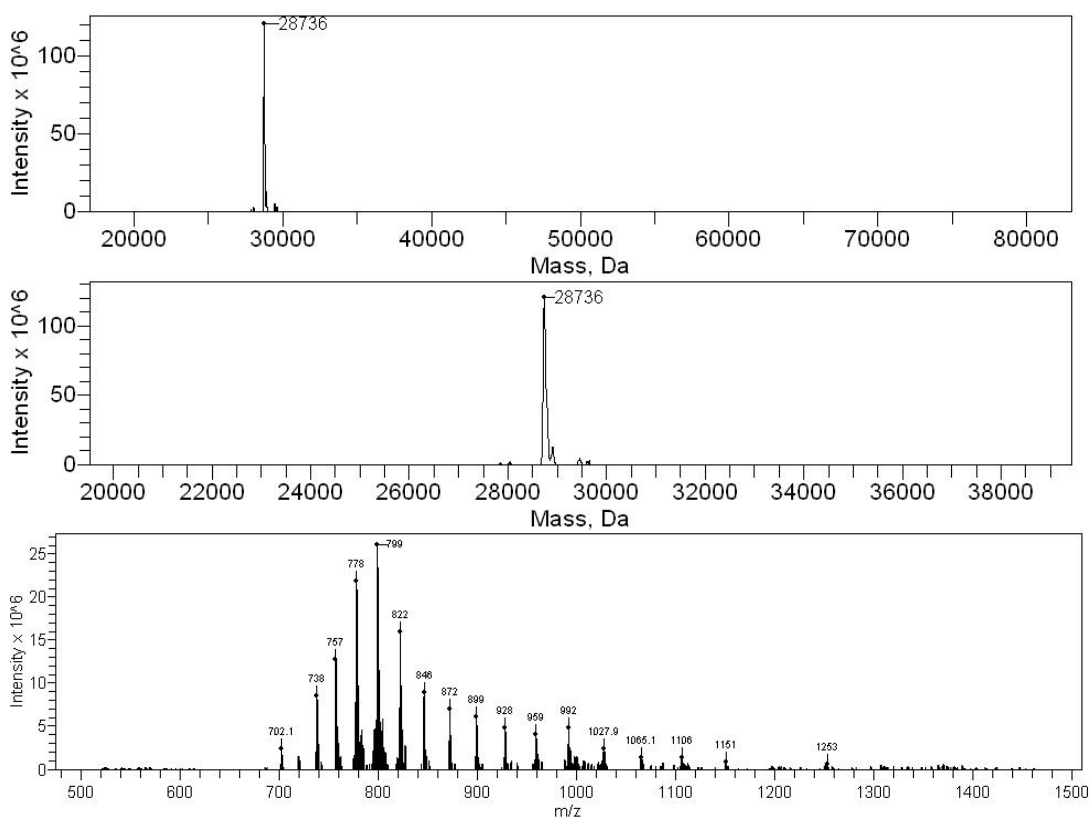
Sodium azide (30 μ L, saturated solution in water) was added to a solution of GFP (S147C Sulf, +CG Dha) **58** (300 μ L, 1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 21 $^{\circ}$ C. The mixture was vortexed for 1 s and maintained at 37 $^{\circ}$ C for 2 h. Excess reagents were removed by repeated diafiltration into fresh buffer (100mM sodium phosphate, pH 8.0) using VivaSpin sample concentrators (GE Healthcare, 10,000 MWCO) and the mixture analysed by LCMS (Expected 29675, Observed 29676).



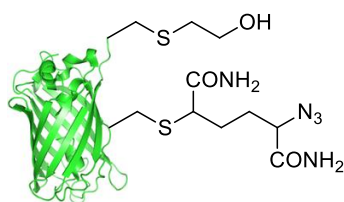
GFP (S147C Azide, T230 Dha) **62**



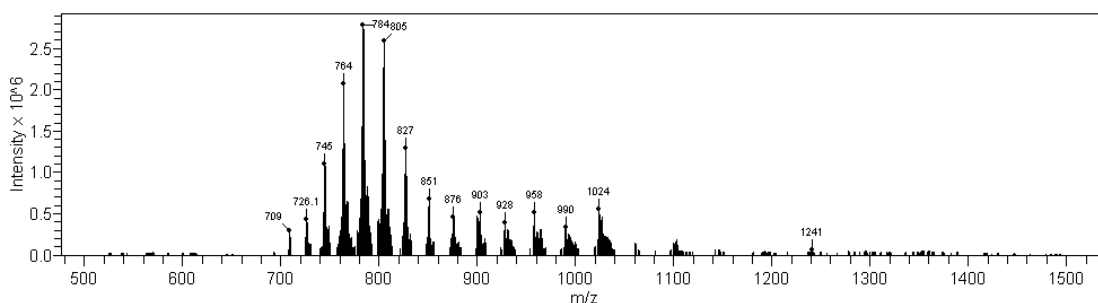
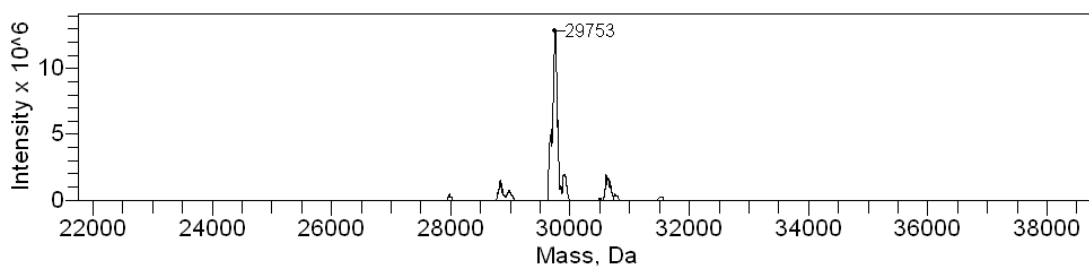
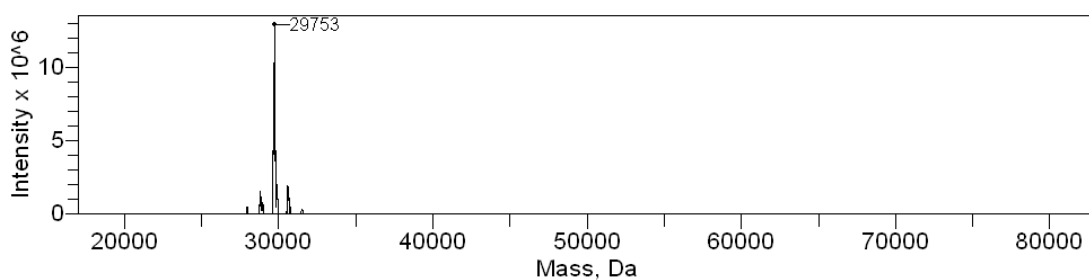
Sodium azide (30 μ L, saturated solution in water) was added to a solution of GFP (S147C Sulf, T230C Dha) **59** (300 μ L, 1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 21 $^{\circ}$ C. The mixture was vortexed for 1 s and maintained at 37 $^{\circ}$ C for 2 h. Excess reagents were removed by repeated diafiltration into fresh buffer (100mM sodium phosphate, pH 8.0) using VivaSpin sample concentrators (GE Healthcare, 10,000 MWCO) and the mixture analysed by LCMS (Expected 28739, Observed 28736).



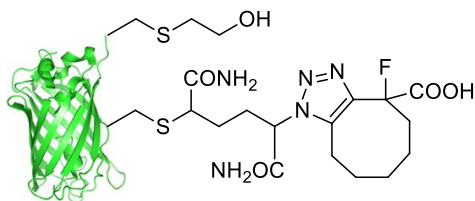
GFP (S147C Azide, +CG Dha 2-mercaptoethanol) conjugate **63**



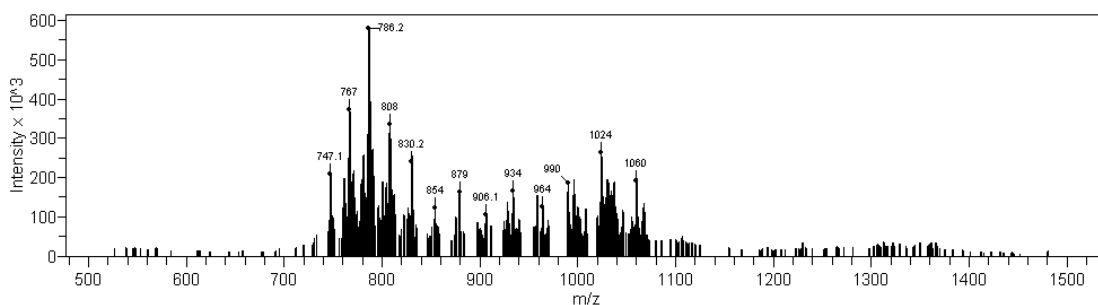
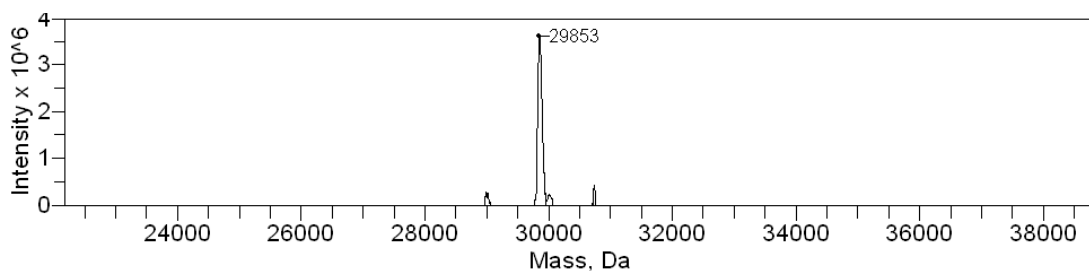
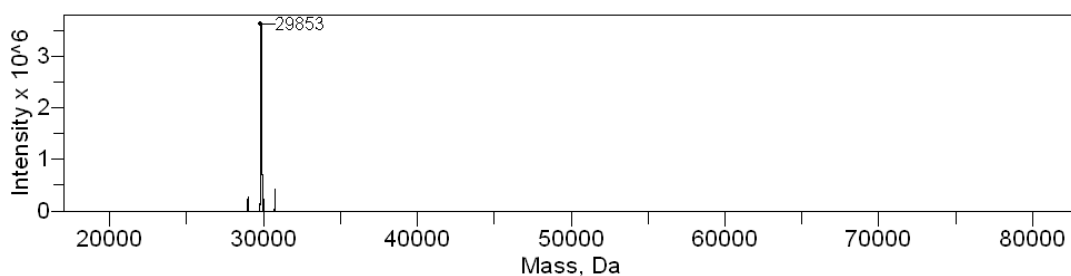
2-mercaptoethanol (2 μ L) was added to a solution of GFP (S147C Azide, +CG Dha) **61** (200 μ L, 1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 37 $^{\circ}$ C. The mixture was vortexed for 1 s and maintained at 37 $^{\circ}$ C for 30 min. Excess reagents were removed by repeated diafiltration into fresh buffer (100mM sodium phosphate, pH 8.0) using VivaSpin sample concentrators (GE Healthcare, 10,000 MWCO) and the mixture analysed by LCMS (Expected 29754, Observed 29753).



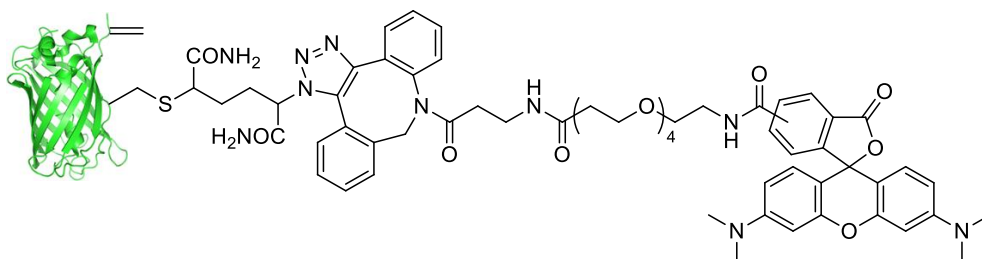
GFP (S147C Azide Octyne, +CG Dha 2-mercaptoethanol) conjugate **64**



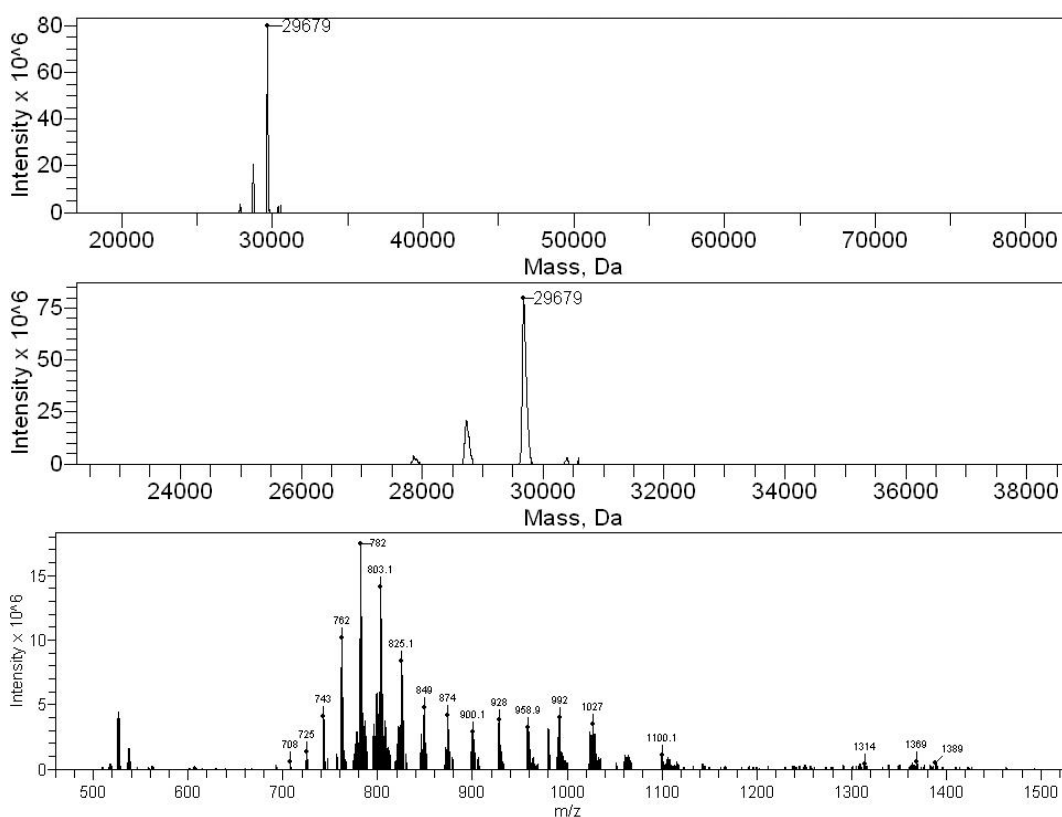
Octyne **52** (10 μ L, 34 mM solution in DMF, 100 equivalents) was added to a solution GFP (S147C Azide, +CG Dha 2-mercaptoethanol) conjugate **63** (100 μ L, 1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 21 $^{\circ}$ C. The mixture was vortexed for 1 s and maintained at 21 $^{\circ}$ C for 16 h. Excess reagents were removed by repeated diafiltration into fresh buffer (100mM sodium phosphate, pH 8.0) using VivaSpin sample concentrators (GE Healthcare, 10,000 MWCO) and the mixture analysed by LCMS (Expected 29845, Observed 29853).



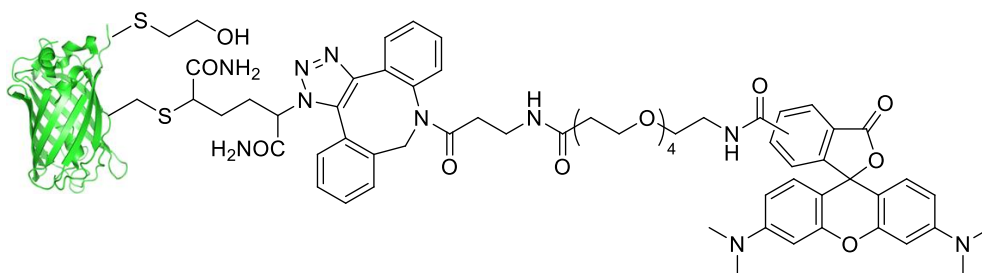
GFP (S147C Azide dibenzylcyclooctyne-PEG₄-Fluor-545,
T230C Dha) conjugate **66**



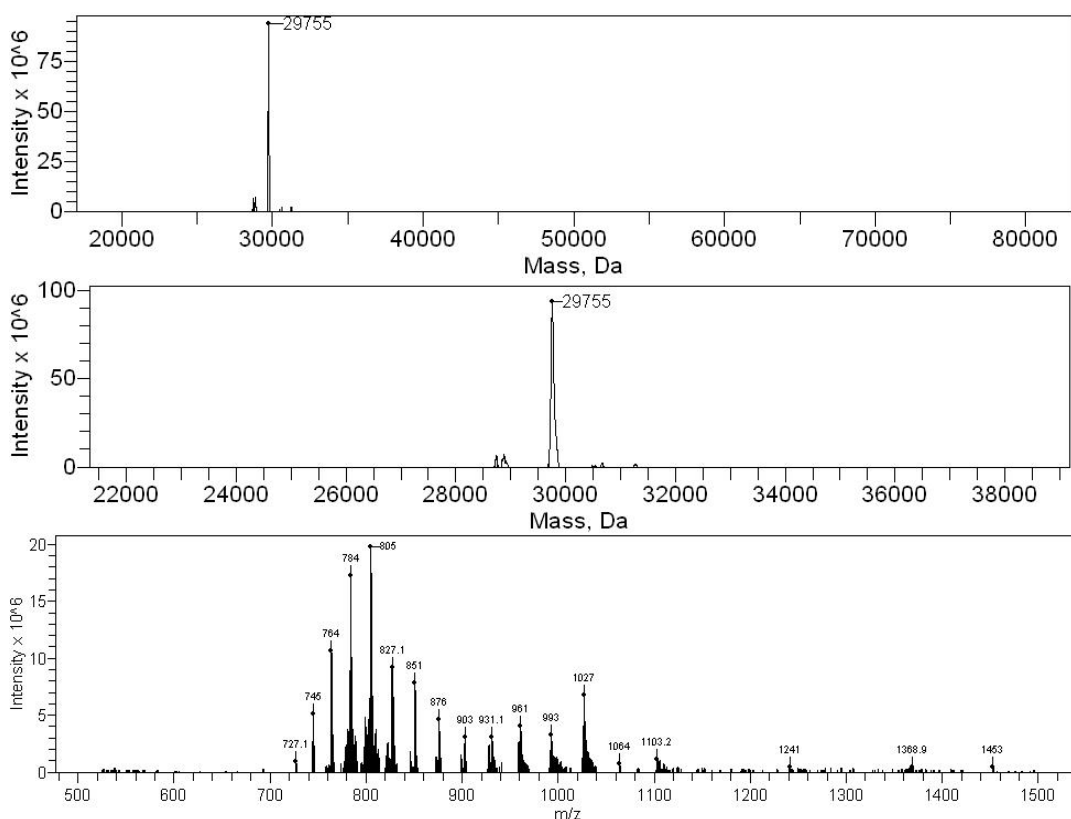
Dibenzylcyclooctyne-PEG₄-Fluor-545 **65** (10 μ L, 34 mM solution in DMF, 100 equivalents) was added to a solution of GFP (S147C Azide, T230C Dha) **62** (100 μ L, 1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 21 $^{\circ}$ C. The mixture was vortexed for 1 s, maintained at 21 $^{\circ}$ C for 4 h. Excess reagents were removed by repeated diafiltration into fresh buffer (100mM sodium phosphate, pH 8.0) using VivaSpin sample concentrators (GE Healthcare, 10,000 MWCO) and the mixture analysed by LCMS (Expected 29675, Observed 29679).



GFP (S147C Azide dibenzylcyclooctyne-PEG4-Fluor-545, T230C Dha 2-mercaptoethanol) conjugate **67**

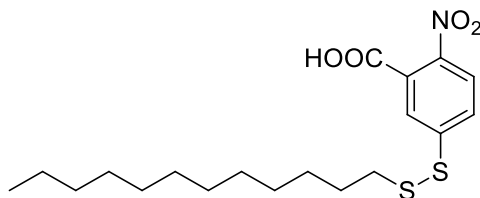


2-Mercaptoethanol (2 μL) was added to a solution of GFP (S147C Azide dibenzylcyclooctyne-PEG₄-Fluor-545, T230C Dha) conjugate **66** (100 μL , 1.0 mg/mL) in sodium phosphate (100 mM, pH 8.0) at 21 °C. The mixture was vortexed for 1 s, maintained at 37 °C for 30 min. Excess reagent was removed by repeated diafiltration into fresh buffer (100mM sodium phosphate, pH 8.0) using VivaSpin sample concentrators (GE Healthcare, 10,000 MWCO) and the mixture analysed by LCMS (Expected 29753, Observed 29755).



Experimental for Chapter 3

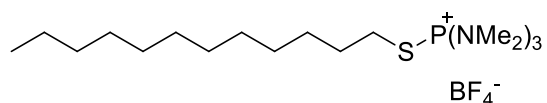
5-(Dodecylsulfanyl)-2-nitrobenzoic acid **96**



To a stirring solution of 5,5'-dithiobis-(2-nitrobenzoic acid) (2.00 g, 4.94 mmol) in MeOH (30 mL) at 21 °C was added dodecanethiol (591 μ L, 2.47 mmol) premixed with K_2CO_3 (340 mg, 2.47 mmol) in a mixture of water (5mL) and MeOH (15 mL). The addition of dodecanethiol was carried out dropwise over 20 min followed by stirring at 21 °C for 4 h. After this, the reaction mixture was concentrated *in vacuo* and diluted with water (25 mL), followed by extraction with EtOAc (2 x 25 mL). The organic layers were combined, dried ($MgSO_4$) and concentrated *in vacuo*. The crude product was purified by flash column chromatography on silica gel (35% EtOAc/petrol) to afford **96** as a pale yellow solid (296 mg, 0.740 mmol, 30%).

1H NMR (600MHz, MeOD): δ 7.94 (d, 1H, $J = 8.4$), 7.70 (d, 1H, $J = 1.8$), 7.63 (dd, 1H, $J = 8.4$ and 1.8), 2.80 (t, 2H, $J = 7.2$), 1.68 (quintet, 2H, $J = 7.8$), 1.40-1.38 (m, 2H), 1.31-1.24 (m, 16H), 0.90 (t, 3H, $J = 7.2$); ^{13}C NMR (125MHz, MeOD) δ 173.0 (C), 147.4 (ArC), 146.7 (ArC), 130.8 (ArC), 129.1 (ArC), 127.1 (ArC), 125.8 (ArC), 40.0 (CH_2), 33.1 (CH_2), 30.8 (CH_2), 30.7 (CH_2), 30.6 (CH_2), 30.5 (CH_2), 30.3 (CH_2), 29.9 (CH_2), 29.4 (CH_2), 28.9 (CH_2), 23.8 (CH_2), 14.5 (CH_3); IR (solid) ν_{max} 2900, 2851, 1464 cm^{-1} ; LRMS (m/z) (ES+) 422 ($[M+Na]^+$, 100); HRMS $C_{19}H_{29}O_4NS_2Na$ calcd. 422.1430 found 422.1635; m.p. 42-44 °C.

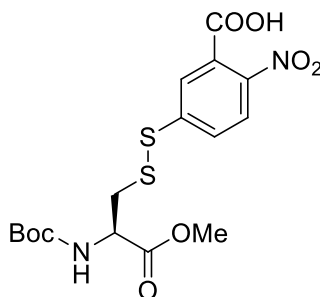
Tris(dimethylamino) dodecanethiophosphonium
tetrafluoroborate **101**



To a stirring solution of dodecanethiol (2.50 g, 12.3 mmol) in THF (35 mL) cooled to $-40\text{ }^\circ\text{C}$ was added *t*-butylhypochlorite (1.51 mL, 12.9 mmol). The reaction was stirred for 10 min followed by addition of tris(dimethylaminophosphine) (2.24 mL, 12.3 mmol). After stirring for 5 min, a white precipitate was observed. The reaction was diluted with water (25 mL) and washed with Et_2O (2 x 25 mL). The aqueous layer was treated with NaBF_4 (2.00 g, 18.2 mmol) in water (25 mL). The solution became clear and the aqueous layer was extracted with DCM (3 x 25 mL), dried (MgSO_4) and concentrated *in vacuo*. The crude product was purified by recrystallization from DCM/ Et_2O to afford **101** as a white solid (2.80 g, 6.20 mmol, 50%).

^1H NMR (600MHz, CDCl_3) δ 2.94-2.90 (m, 2H), 2.82 (d, 18 H, $J_{\text{H-P}} = 11.2$), 1.68 (quintet, 2H, $J = 7.8$), 1.35 (quintet, 2H, $J = 7.2$), 1.26-1.21 (m, 16H), 0.83 (t, 3H, $J = 7.1$); ^{13}C NMR (150MHz, CDCl_3) δ 37.5 (d, CH_3 , $J_{\text{C-P}} = 3.8$), 32.0 (CH_2), 31.1 (d, CH_2 , $J_{\text{C-P}} = 4.2$), 30.3 (CH_2), 30.2 (CH_2), 29.7 (d, CH_2 , $J_{\text{C-P}} = 1.8$), 29.6 (CH_2), 29.5 (CH_2), 29.4 (CH_2), 29.1 (CH_2), 28.5 (CH_2), 22.8 (CH_2), 14.2 (CH_3); ^{31}P NMR (121MHz, CDCl_3) δ 66.9; IR ν_{max} (solid) 2923, 2854, 1468, 1300, 1297, 1171, 1049, 1036 cm^{-1} ; LRMS (m/z) (ES+) 364 (M^+ , 100); HRMS $\text{C}_{18}\text{H}_{43}\text{N}_3\text{PS}$ calcd. 364.2915, found 364.2908; m.p. 126-128 $^\circ\text{C}$.

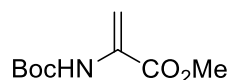
(*R*)-5-((2-((tert-butoxycarbonyl)amino)-3-methoxy-3-oxopropyl)disulfanyl)-2-nitrobenzoic acid **108**



To a stirring solution of 5,5'-dithiobis-(2-nitrobenzoic acid) (1.26 g, 3.19 mmol) in DMF (10 mL) and buffer (100 mM sodium phosphate, pH 8.0) (10 mL) at 21 °C was added a solution of *N*-Boc-Cys-OMe **107** (250 mg, 1.06 mmol) in DMF (2 mL) dropwise over 30 min. The reaction was stirred at 21 °C for 10 min. The pH was adjusted to 5 with HCl (1M), followed by extraction with EtOAc (2 x 25 mL). The organic layer was washed with buffer (100 mM sodium phosphate, pH 8.0) (2 x 25 mL). The aqueous layers were combined and the pH was adjusted to pH 4 by the addition of HCl (1M) followed by extraction with EtOAc (2 x 25 mL). The organic layers were combined and washed with saturated aqueous LiCl (2 x 25 mL), dried (MgSO₄) and concentrated *in vacuo* to afford **108** as a yellow oil (150 mg, 0.350 mmol, 33%).

¹H NMR (600MHz, MeOD) δ 7.92 (d, 1H), 7.89-7.88 (m, 1H), 7.85-7.83 (m, 1H), 4.43 (dd, 1H, *J* = 9.0 and 4.2), 3.70 (s, 3H), 3.26 (dd, 1H, *J* = 13.8 and 4.5), 3.07 (dd, 1H, *J* = 14.2 and 8.9), 1.42 (s, 9H); ¹³C NMR (150MHz, MeOD) δ 173.1 (C), 168.0 (C), 157.7(C), 147.7 (ArC), 145.6 (ArC), 130.7 (ArC), 129.5 (ArC), 127.6 (ArC), 125.9 (ArC), 81.0 (C), 54.2 (CH), 53.0 (CH₃), 41.2 (CH₂), 28.7 (CH₃); IR ν_{\max} (oil) 2983, 1709, 1601, 1521, 1360, 1159 cm⁻¹; LRMS (ES+) (m/z) 455([M+Na]⁺, 30), 311 (100); HRMS C₁₆H₂₀N₂O₈NaS₂ calcd. 455.0559, found 455.0532; [α]_D²⁰ +56.6 (c 1.0, MeOH); m.p. 89-91 °C.

Methyl 2-((tert-butoxycarbonyl)amino)acrylate **110**

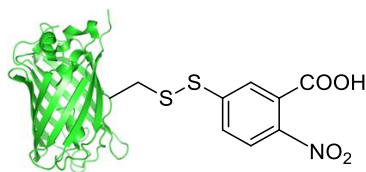


Method 1: To a stirring solution of disulfide **108** (30.0 mg, 0.069 mmol) in DMF (6 mL) and buffer (100 mM sodium phosphate, pH 8.0) (20 mL) at 21 °C was added tris(dimethylamino)phosphine (25 μ L, 0.14 mmol). The reaction was stirred at 21 °C for 4 h. The reaction mixture was extracted with EtOAc (2 x 25 mL). The organic layers were combined and washed with saturated aqueous LiCl solution (3 x 25 mL), saturated aqueous NaCl (25 mL), dried (MgSO₄) and concentrated *in vacuo*. The crude residue was purified by flash column chromatography on a silica gel (5% EtOAc/Petrol) to afford **110** as a colourless oil (7.00 mg, 0.034 mmol, 50%).

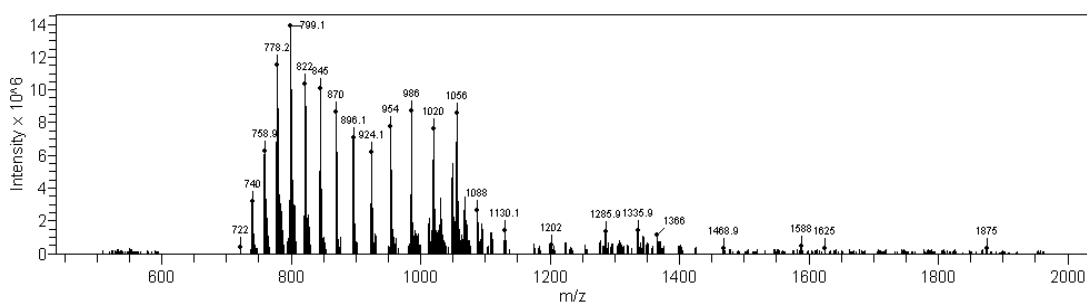
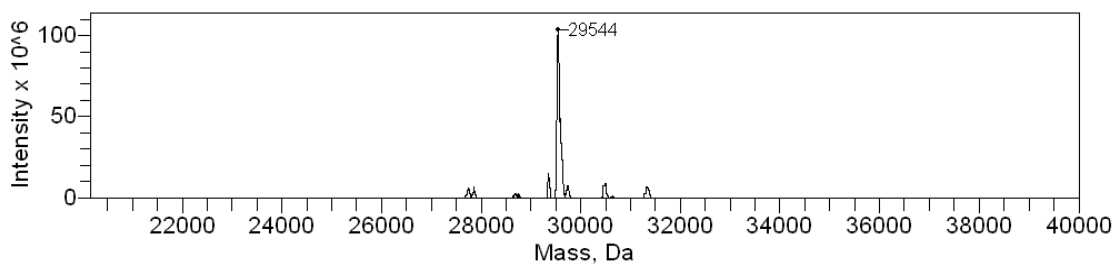
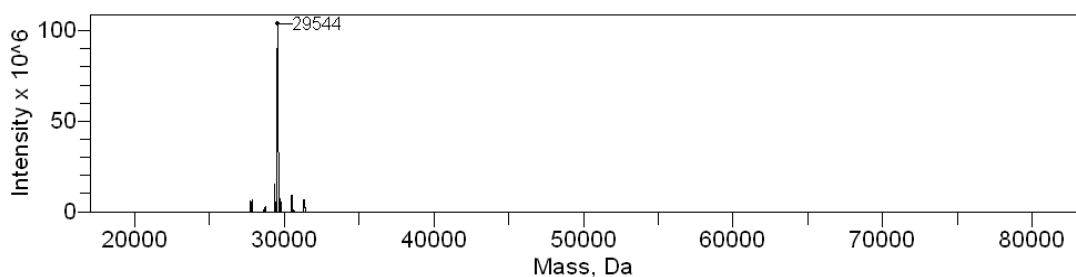
Method 2: To a stirring solution of disulfide **108** (30.0 mg, 0.069 mmol) in DMF (6 mL) and buffer (70 mM CHES/MES, pH 9.5) (20 mL) at 21 °C was added tris(dimethylamino)phosphine (25.0 μ L, 0.140 mmol). The reaction was stirred at 21 °C for 4 h. The reaction mixture was extracted with EtOAc (2 x 25 mL). The organic layers were washed with saturated aqueous LiCl solution (3 x 25 mL), saturated aqueous NaCl (25 mL), dried (MgSO₄) and concentrated *in vacuo*. The crude residue was purified by flash column chromatography on a silica gel (5% EtOAc/Petrol) to afford **110** as a colourless oil (6.30 mg, 0.031 mmol, 45%).

¹H NMR (500MHz, CDCl₃) δ 7.00 (s, 1H), 6.15 (s, 1H), 5.72 (s, 1H), 3.82 (s, 3H), 1.48 (s, 9H); ¹³C NMR (125MHz, CDCl₃) δ 164.6 (C), 152.6 (C), 131.4 (CH), 105.3 (CH), 80.8 (CH₃), 52.9 (CH₃), 28.3 (CH₃); IR ν_{\max} (oil) 3423, 2979, 1717, 1630 cm⁻¹; LRMS (m/z) (EI) 201 (M⁺, 23), 174 (40), 145 (100); HRMS C₉H₁₅O₄N calcd. 201.0996, found 201.0991.

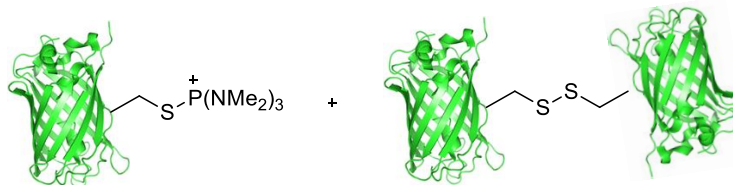
GFP (S147C, 5-mercapto-2-nitrobenzoic acid)disulfide **112**



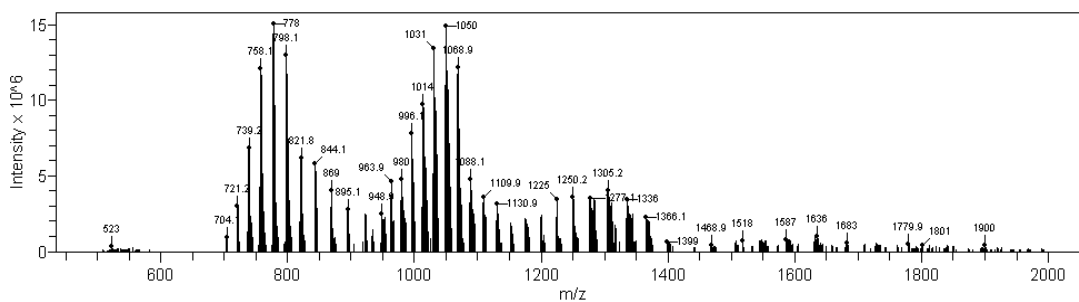
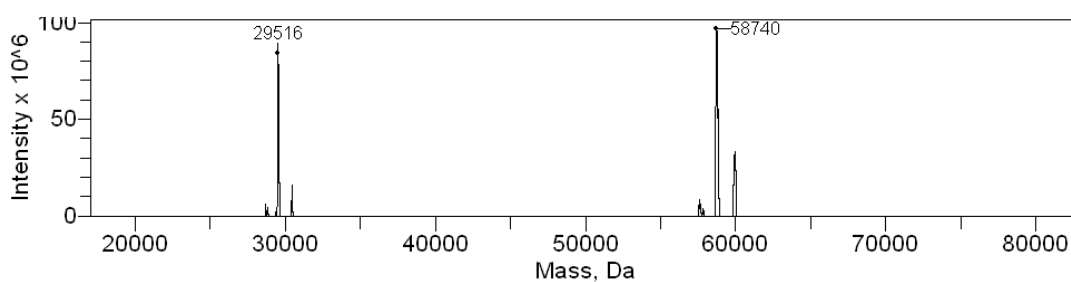
To a solution of GFP (S147C) **32** (100 μ L, 1 mg/ml, 100 mM sodium phosphate, pH 8.0) at 4 $^{\circ}$ C was added 5,5'-dithiobis-(2-nitrobenzoic acid) (5 μ L, 0.67 mM solution in DMF, 1 equivalent). The mixture was vortexed for 1 s, maintained at 4 $^{\circ}$ C for 16 h and analysed by LCMS (Expected 29543, Observed 29544).



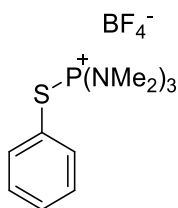
GFP (S147C, thiophosphonium) **113** and GFP(S147C) Dimer **114**



To a solution of GFP (S147C, 5-mercapto-2-nitrobenzoic acid) disulfide **112** (100 μL , 5 mg/ml, 100 mM sodium phosphate, pH 8.0) was added tris(dimethylamino)phosphine (5 μL , 205 mM solution in DMF, 60 equivalents) at 4 $^\circ\text{C}$. The reaction was maintained at 21 $^\circ\text{C}$ for 2 h and analysed by LCMS (Expected 29508 and 58690, Observed 29516 and 58740).



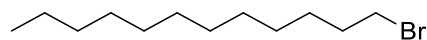
Tris(dimethylamino) benzenethiophosphonium
tetrafluoroborate **120**



To a stirring solution thiophenol (1.00 g, 9.09 mmol) in THF (35 mL) cooled to $-40\text{ }^\circ\text{C}$ was added *t*-butylhypochlorite (1.14 mL, 9.54 mmol). The reaction was stirred for 10 minutes followed by addition of tris(dimethylaminophosphine) (1.65 mL, 9.09 mmol). After stirring for 5 minutes, a white precipitate was observed. The reaction was diluted with water (25 mL) and washed with Et_2O (2 x 25 mL). The aqueous layer was treated with NaBF_4 (2.00 g, 1.82 mmol) in water (25 mL). The solution became clear and the aqueous layer was extracted with DCM (3 x 25 mL), dried (MgSO_4) and concentrated *in vacuo*. The crude product was purified by recrystallization from DCM/ Et_2O to afford **120** as a white solid (600 mg, 1.67 mmol, 18%).

^1H NMR (600MHz, CDCl_3) δ 7.59-7.57 (m, 2H), 7.51-7.45 (m, 3H), 2.73 (d, 18 H, $J = 10.8$); ^{13}C NMR (150MHz, CDCl_3) δ 136.9 (d, ArC, $J_{\text{C-P}} = 4.6$), 131.7 (d, ArC, $J_{\text{C-P}} = 3.6$), 130.8 (d, ArC, $J_{\text{C-P}} = 3$), 121.4 (d, ArC, $J_{\text{C-P}} = 6.8$), 37.8 (d, CH_3 , $J_{\text{C-P}} = 3.2$); ^{31}P NMR (121MHz, CDCl_3) δ 61.9; IR ν_{max} (solid) 2923, 1473, 1297, 1171, 1046, 985 cm^{-1} ; LRMS (m/z) (ES+) 272 (M^+ , 100); HRMS $\text{C}_{12}\text{H}_{23}\text{N}_3\text{PS}$ calcd. 272.1350, found 272.1346; m.p. 98-100 $^\circ\text{C}$.

1-Bromododecane **122**

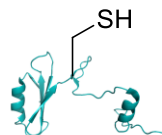


To a stirring solution of 5-(dodecyldisulfanyl)-2-nitrobenzoic acid **96** (50.0 mg, 0.125 mmol) in DMF (1.50 mL) at 21 °C was added tris(dimethylamino)phosphine **68** (50.0 μ L, 0.275 mmol). The solution turned red within 1 min and NaBr (64.0 mg, 0.625 mmol) in water (0.5 mL) was subsequently added. The reaction was heated to 60 °C for 6 h. The aqueous reaction mixture was extracted with EtOAc (2 x 10 mL). The organic layer was then washed with aqueous LiCl (3 x 10 mL), aqueous NaCl (15 mL), dried (MgSO₄) and concentrated *in vacuo*. The residue was purified by flash column chromatography on a silica gel (1% to 5% EtOAc/Petrol) to afford **122** as a colourless oil (5.00 mg, 0.021 mmol, 17%).

¹H NMR (600MHz, CDCl₃) δ 3.40 (t, 2H, *J* = 6.9), 1.85 (quintet, 2H, *J* = 6.9), 1.44-1.39 (m, 2H), 1.29-1.26 (m, 16H), 0.88 (t, 3H, *J* = 6.8); ¹³C NMR (150MHz, CDCl₃) δ 34.2 (CH₂), 33.0 (CH₂), 32.0 (CH₂), 29.8 (CH₂), 29.7 (CH₂), 29.7 (CH₂), 29.6 (CH₂), 29.5 (CH₂), 28.9 (CH₂), 28.3 (CH₂), 22.8 (CH₂), 14.3 (CH₃); IR ν_{\max} (oil) 2923, 2853, 1465, 906 cm⁻¹; LRMS (m/z) (EI) 250 (⁸¹M⁺, 4), 248 (⁷⁹M⁺, 4), 137 (95), 135 (100); HRMS C₁₂H₂₅⁷⁹Br calcd. 248.1134, found 248.1136.

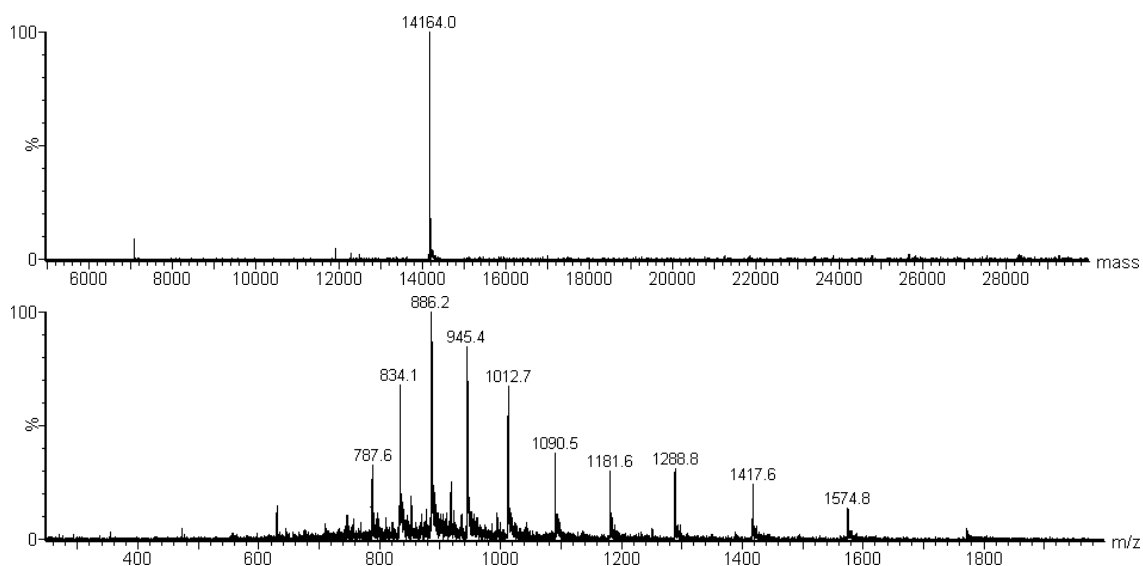
Experimental for Chapter 4

Grb2-SH2 Domain (L111C) 124

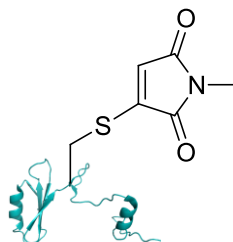


Sequence

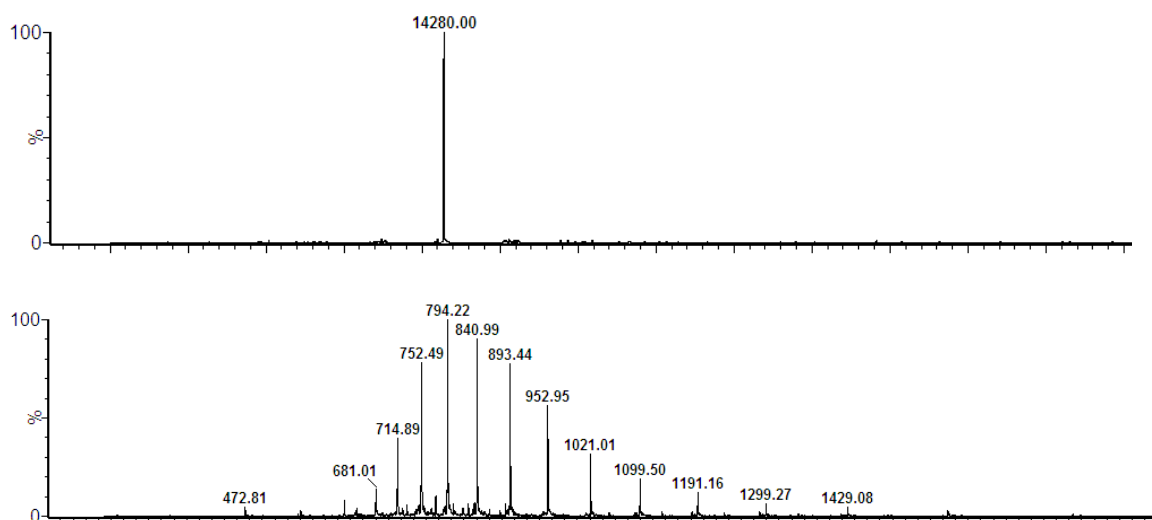
MGIEMKPHPWFFGKIPRAKAEEMLSKQRHDGAFLIRESESAPGDFSLSVKFG
NDVQHFVCRDGAGKYFLWVVKFNSLNELVDYHRSTS VSRNQQIFLRDIEQV
PQQPTYVQAGSRSHHHHHH (Expected 14170 Observed 14164).



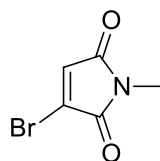
Grb2-SH2 (L111C, *N*-methylmaleimide) **126**



To a solution of Grb2-SH2 (L111C) **124** (100 μ L, 1 mg/ml, 100 mM sodium phosphate, pH 8.0) at 4 $^{\circ}$ C was added bromo-*N*-methylmaleimide **125** (5 μ L, 2.82 mM solution in DMF, 1 equivalent). The mixture was vortexed for 1 s, maintained at 4 $^{\circ}$ C for 1 h then analysed by LCMS (Expected 14275, Observed 14280).



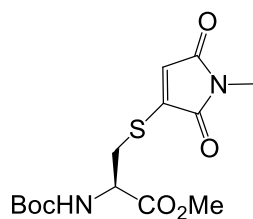
3-Bromo-1-methyl-1*H*-pyrrole-2,5-dione **125** ⁴⁷



To a stirring solution of *N*-methylmaleimide **137** (500 mg, 4.50 mmol) in MeOH (22.5 mL) at 21 °C was added bromine (520 μ L, 10.0 mmol) dropwise. The reaction mixture was stirred at 21 °C for 24 h. The reaction mixture was concentrated *in vacuo* and the crude product was dissolved in THF (20 mL). To this was added NEt₃ (0.800 mL, 5.85 mmol) and the reaction was stirred for 24 h at 21 °C. The reaction mixture was concentrated *in vacuo* and the crude residue was purified by flash column chromatography on silica gel (30% EtOAc/Petrol) to afford **125** as a pale white powder (0.760 g, 4.00 mmol, 89%).

¹H NMR (500MHz, CDCl₃) δ 6.90 (s, 1H), 3.09 (s, 3H); ¹³C NMR (125MHz, CDCl₃) δ 168.6 (C), 165.4 (C), 131.9 (CH), 131.4 (C), 24.7 (CH₃); IR ν_{\max} (solid) 3106, 1777, 1708, 1590, 1440, 1391, 1231, 970 cm⁻¹; LRMS (m/z) (EI) 192 (⁸¹M⁺, 99), 190 (⁷⁹M⁺, 100); HRMS C₅H₅O₂N⁷⁹Br calcd. 189.9504, found 189.9505; m.p. 77-79 °C.

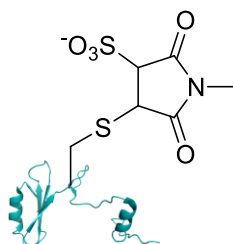
(*R*)-Methyl 2-((tert-butoxycarbonyl)amino)-3-((1-methyl-2,5-dioxo-2,5-dihydro-1H-pyrrol-3-yl)thio)propanoate **130**⁴⁷



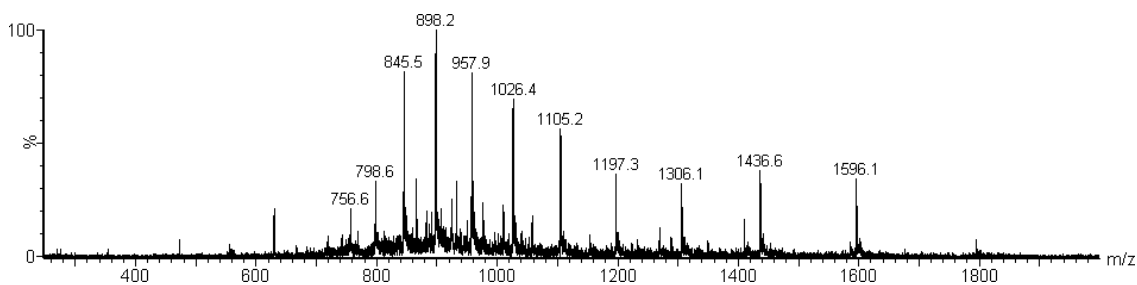
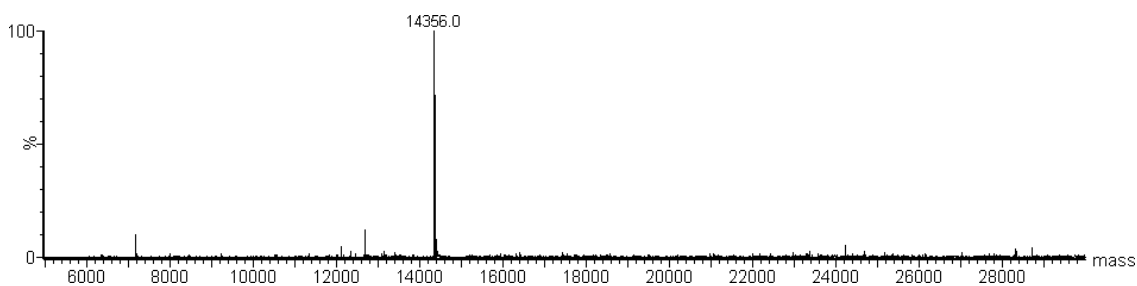
To a stirring solution of *N*-Boc-Cys-OMe **107** (320 mg, 1.36 mmol) and NaOAc (820 mg, 4.08 mmol) in MeOH (10 mL) at 21 °C was added a solution of **125** (258 mg, 1.36 mmol) in MeOH (10 mL) over a period of 10 min. The reaction mixture was concentrated *in vacuo* and the crude residue was purified by column chromatography (10% to 30% EtOAc/Petrol) to afford **130** as a pale white powder (393 mg, 1.14 mmol, 84%).

¹H NMR (600 MHz, CDCl₃) δ 6.29 (s, 1H), 5.41 (d, 1H, *J* = 6.5), 4.69-4.67 (m, 1H), 3.81 (s, 3H), 3.49 (dd, 1H, *J* = 13.5 and 5.0), 3.37 (dd, 1H, *J* = 13.5 and 5.0), 3.02 (s, 3H), 1.46 (s, 9H); ¹³C NMR (125 MHz, CDCl₃) δ 170.5 (C), 169.8 (C), 168.0 (C), 155.3 (C), 150.3 (C), 118.7 (CH), 81.2 (C), 53.5 (CH₃), 53.0 (CH), 34.1 (CH₂), 28.6 (CH₃), 24.4 (CH₃); IR ν_{\max} (solid) 3368, 2977, 1746, 1695, 1162 cm⁻¹; LRMS (m/z) (ES+) 367 ([M+Na]⁺, 46), 344 (M⁺, 100); HRMS C₁₄H₂₀N₂O₆NaS calcd. 367.0940, found 367.0931; [α]_D²⁰ -18.6 (c 1.0, MeOH); m.p. 101-103 °C.

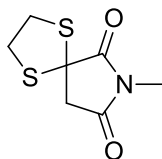
Grb2-SH2 (L111C, *N*-methylmaleimide, Sulfite) **155**



To a solution of Grb2-SH2 (L111C, *N*-methylmaleimide) **126** (100 μ L, 1 mg/ml, 100 mM sodium phosphate, pH 8.0) at 4 $^{\circ}$ C was sodium sulfite (5 μ L, 28.2 mM solution in water, 10 equivalents). The mixture was vortexed for 1 s, maintained at 21 $^{\circ}$ C for 2 h and analysed by LCMS (Expected 14355, Observed 14356).



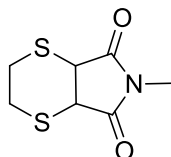
7-methyl-1,4-dithia-7-azaspiro[4.4]nonane-6,8-dione **155**



To a stirring solution of thiomaleimide **130** (40.0 mg, 0.211 mmol) in DMF (1.5 mL) and buffer (100 mM sodium phosphate, pH 8.0) (13 mL) at 21 °C was added 1,2-ethanedithiol (52.0 mg, 0.620 mmol) in DMF (1.5 mL). The reaction was heated at 37 °C for 24 h. The reaction mixture was extracted with EtOAc (2 x 25 mL). The organic layers were combined and washed with saturated aqueous LiCl (3 x 25 mL), saturated aqueous NaCl (25 mL), dried (MgSO₄) and concentrated *in vacuo*. The residue was purified by flash column chromatography on silica gel (10% to 30% EtOAc/Petrol) to afford a white solid (40.0 mg, 0.200 mmol, 95%).

¹H NMR (600 MHz, CDCl₃) δ 3.77-3.71 (m, 2H), 3.59-3.52 (m, 2H), 3.26 (s, 2H), 3.02 (s, 3H); ¹³C NMR (150 MHz, CDCl₃) δ 178.1 (C), 173.1 (C), 60.0 (C), 42.2 (CH₂), 41.1 (CH₂), 25.7 (CH₃); IR (solid) ν_{max} 2935, 1736, 1434, 1376, 1285, 1242 cm⁻¹; LRMS (m/z) (ES+) 203 (M⁺, 54), 144 (100); HRMS C₇H₉NO₂S₂ calcd. 203.0069, found 203.0067; m.p. 71-73 °C.

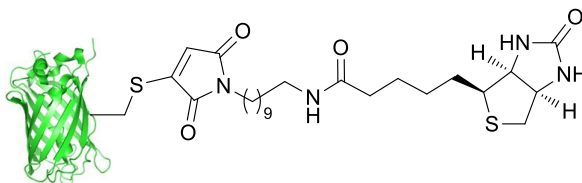
6-Methyldihydro-4aH-[1,4]dithiino[2,3-c]pyrrole-5,7(6H,7aH)-
dione **156**



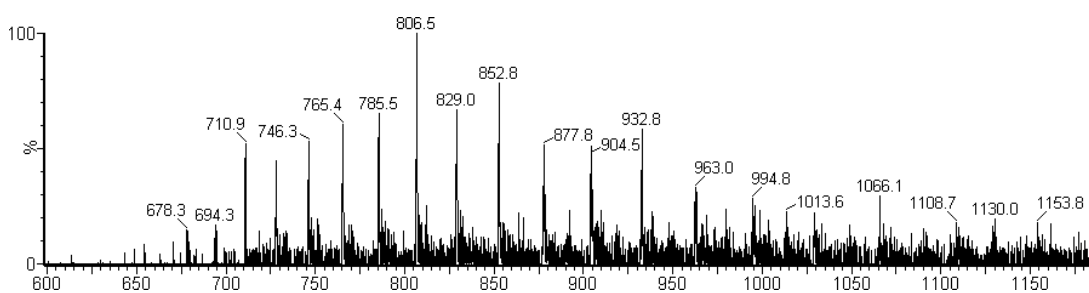
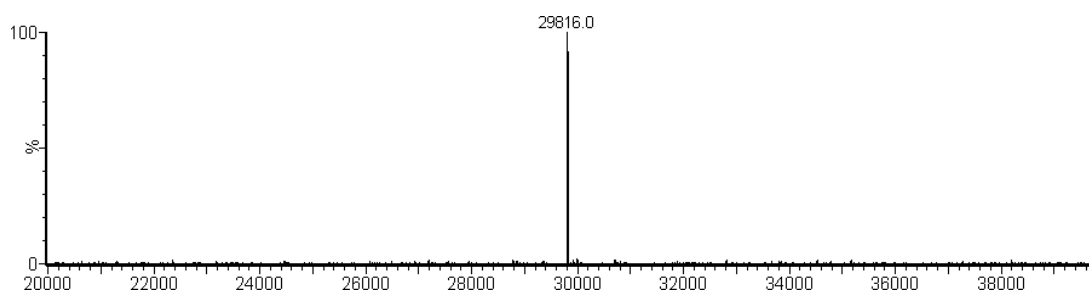
To a stirring solution of bromo-*N*-methylmaleimide **125** (40.0 mg, 0.210 mmol) in a mixture of DMF (1.50 mL) and buffer (100 mM sodium phosphate, pH 8.0) (13 mL) at 21 °C was added 1,2-ethanedithiol (20.0 mg, 0.210 mmol) in DMF (1.50 mL). The reaction mixture was heated at 37 °C for 90 min and then extracted with EtOAc (2 x 25 mL). The organic layers were combined and washed with saturated aqueous LiCl (3 x 25 mL), saturated aqueous NaCl (25 mL), dried (MgSO₄), concentrated *in vacuo* and the crude residue purified by flash column chromatography on silica gel (10% to 30% EtOAc/petrol) to afford **156** as a white solid (15.0 mg, 0.070 mmol, 35%).

¹H NMR (CDCl₃, 600 MHz) δ 4.05 (s, 2H), 3.10 (s, 3H), 2.88-2.79 (m, 4H); ¹³C NMR (CDCl₃, 150 MHz) δ 174.3 (C), 40.8 (CH₃), 25.8 (CH), 24.8 (CH₂); IR ν_{max} (solid) 2985, 1737, 1447, 1373 cm⁻¹; LRMS (m/z) (ES+) 203 (M⁺, 100); HRMS C₇H₉NO₂S₂ calcd. 203.0069, found 203.0070; m.p. 69-71 °C.

GFP (S147C, biotin-dec-maleimide) **159**



To a solution of GFP (S147C) (100 μ L, 1 mg/ml, 100 mM sodium phosphate, pH 8.0) at 4 $^{\circ}$ C was added bromomaleimide-dec-biotin **158** (5 μ L, 2.82 mM solution in DMF, 1 equivalent). The mixture was vortexed for 1 s, maintained at 4 $^{\circ}$ C for 1 h then analysed by LCMS (Expected 29823, Observed 29816).



References

-
- ¹ Lander, E.S.; Linton, L.M.; Birren, B.; Nusbaum, C.; Zody, M.C. *et al. Nature*, **2001**, *409*, 860-921.
- ² Venter, J.; Adams, M.; Myers, E.; Li, P.; Mural, R. *et al. Science*, **2001**, *291*, 1304–1351.
- ³ Jensen, O. *Curr. Opin. Chem. Biol.*, **2004**, *8*, 33-41.
- ⁴ Kasteren, S.I; Kramer, H.B.; Jensen, H.H.; Campbell, S.J.; Kirkpatrick, J.; Oldham, N.J.; Antony, D.C.; Davis, B.G. *Nature*, **2007**, *446*, 1105-1109.
- ⁵ Mann, M.; Jensen, O. *Nat. Biotech.*, **2003**, *21*, 255 – 261.
- ⁶ Davis, B.G. *Science*, **2004**, *303*, 480-482.
- ⁷ Nowak, S., Corces, C. *Trends in Genetics*, **2004**, *20*, 4-7.
- ⁸ Basle, E.; Joubert, N.; Pucheault, M. *Cell*, **2010**, *17*, 213-227.
- ⁹ Chalker, J.; Lercher, L.; Rose, N.; Schofield, C.; Davis, B. *Angew. Chem. Int. Ed.*, **2012**, *51*, 1835-1839.
- ¹⁰ Girouard, S.; Houle, M.H.; Grandbois, A.; Keillor, J. W.; Michnick, S.W. *J. Am. Chem. Soc.* **2005**, *127*, 559-561.
- ¹¹ Dube D.; Bertozzi, C.R. *Nat. Rev. Drug Discov.*, **2005**, *4*, 477-488.
- ¹² Doore, K.; Gamblin, D.P.; Davis, B.G. *Chem. Euro. J.*, **2006**, *12*, 656-665.
- ¹³ Astronomo, R.D.; Lee, H.K.; Scanlan, C.N.; Pantophlet, R.; Huang, C.Y.; Wilson I.A.; Blixt, O.; Dwek, R.A.; Wong, C.H.; Burton, D.R. *J. Virol.*, **2008**, *82*, 6359-6368.
- ¹⁴ Krug, L.M.; Ragupathi, G.; Ng, K.K.; Hood, C.; Jennings, H. J.; Guo, Z.; Kris, M.G.; Miller, V.; Pizzo, B.; Tyson, L.; Baez, V.; Livingston, P.O. *Clin. Cancer Res.*, **2004**, *10*, 916-923.
- ¹⁵ Seeberger, P.H.; Werz, D.B. *Nature*, **2007**, *446*, 1046-1051.
- ¹⁶ Heller, A.; Feldman, B. *Chem. Rev.*, **2008**, *108*, 2482-2505
- ¹⁷ Hackenberger, C. P. R.; Schwarzer, D. *Angew. Chem. Int. Ed.*, **2008**, *47*, 10030–10074.
- ¹⁸ Lundblad, R.L., Chemical reagents for Protein Modification, 3rd ed., CRC Press, Boca Reton, RL, 2005.
- ¹⁹ Carrico, I. S. *Chem. Soc. Rev.*, **2008**, *37*, 1423-1431.
- ²⁰ Chi, E.Y.; Krishnan, S.; Randolph, T.W.; Carpenter, J.F. *Pharm. Res.*, **2003**, *20*, 1325-1336.
- ²¹ Hackenberger, C. P. R.; Schwarzer, D. *Angew. Chem. Int. Ed.*, **2008**, *47*, 10030–10074.

-
- ²² Lundblad, R.L., Chemical reagents for Protein Modification, 3rd ed., CRC Press, Boca Reton, RL, 2005.
- ²³ de Graaf, A.; Kooijman, M; Hennink, W; Mastrobattista, E. *Bioconjugate Chem.*, **2009**, *20*, 1281-1295.
- ²⁴ Kalia, J.; Raines, R. *Curr. Org. Chem.*, **2010**, *14*, 138-147
- ²⁵ Lomant, A.; Fairbanks, G. *J. Mol. Biol.*, **1976**, *104*, 243-247.
- ²⁶ Tuls, J.; Geren, L.; Millett, F. *J. Biol. Chem.*, **1989**, *264*, 16421-16425.
- ²⁷ Hooker, J.M.; Esser- Kahn, A.P.; Francis, M.B. *J. Am. Chem. Soc.*, **2006**, *128*, 15558-15559.
- ²⁸ Gildersleeve, J.C.; Oyelaran, O.; Simpson, J.T.; Allred, B. *Bioconjugate Chem.*, **2008**, *19*, 1485-1490.
- ²⁹ Cal, P.M.; Vicente, J.B.; Pires, E.; Coelho, A.V.; Veiros, L.F.; Cordeiro, C.; Gois, P.M. *J. Am. Chem. Soc.*, **2012**, *134*, 10299–10305.
- ³⁰ Chen, G.; Heim, A.; Riether, D.; Yee, D.; Milgrom, Y.; Gawinowicz, M.A.; Sames, D. *J. Am. Chem. Soc.*, **2008**, *125*, 8130–8133.
- ³¹ Schlick, T.L.; Ding, Z.; Kovacs, E.W.; Francis, M.B. *J. Am. Chem. Soc.* **2005**, *127*, 3718–3723.
- ³² Joshi, N.S.; Whitaker, L.R.; Francis, M.B.; *J. Am. Chem. Soc.* **2004**, *126*, 15942–15943.
- ³³ Hooker, J.M.; Kovacs, E.W.; Francis, M.B. *J. Am. Chem. Soc.* **2004**, *126*, 3718–3719.
- ³⁴ Guo, H.-M.; Minakawa, M.; Ueno, L.; Tanaka, F. *Bioorg. Med. Chem. Lett.*, **2009**, *19*, 1210–1213.
- ³⁵ Griep, M. A.; Mesman, T. N. *Bioconjugate Chem.* **1995**, *6*, 673-678.
- ³⁶ Wong, L. S.; Khan, F.; Micklefield, J. *Chem.Rev.* **2009**, *109*, 4025-4032.
- ³⁷ Goddard, D.R; Michaelis, L. *J. Biol. Chem.* **1953**, *112*, 361-371.
- ³⁸ Nielsen, M. L.; Vermeulen, M.; Bonaldi, T.; Cox, J.; Moroder, L.; Mann, M. *Nat. Methods.* **2008**, *5*, 459.
- ³⁹ Shaunak, S.; Godwin, A.; Choi, J.W.; Balan, S.; Pedone, E.; Vijayrangam, D.; Heidelberger, I.; Teo, I.; Zloh, M.; Brocchini, S.; *Nat. Chem. Biol.*, **2006**, *2*, 312 – 313.
- ⁴⁰ Davis, N.; Flitsch S.; *Tetrahedron Lett.*, **1991**, *32*, 6793-6796.
- ⁴¹ Macmillan, D.; Bill, R.M.; Sage, K. A.; Fern, D.; Flitsch, S. L. *Chem. Biol.*, **2001**, *8*, 133 – 145.
- ⁴² Swanwick, R.; Daines, A.M.; Tey L.; Flitsch, S.L.; Allemann, R.L. *ChemBioChem*, **2005**, *6*, 1338 –1340.

-
- ⁴³ Lindley, H. *Nature*, **1956**, *178*, 647 –648.
- ⁴⁴ Martin, C.; Zhang, Y. *Nat. Rev. Mol. Cell Biol.*, **2005**, *6*, 838 –849.
- ⁴⁵ G. T. Hermanson, *Bioconjugation Techniques*, 2nd edition, Academic Press, Inc. , 2008.
- ⁴⁶ Weerapana, E.; Simon, G. M. ; Cravatt, B. F. *Nat. Chem. Biol.*, **2008**, *4*, 405 –407.
- ⁴⁷ Tedaldi, L. M.; Smith, M. E. B.; Nathani, R. I.; Baker, J. R. *Chem. Commun.* **2009**, *43*, 6583-6585.
- ⁴⁸ Smith, M.; Caddick, S.; Baker, J. R.; Waksman, G.; Schumacher, F.; Papaioannou, D.; Tedaldi, L.; Ryan, C. *J. Am. Chem. Soc.*, **2010**, *132*, 1960-1965.
- ⁴⁹ Chudasama, V.; Smith, M. E.B.; Schumacher, F.; Papaioannou, D.; Waksman, G.; Baker, J. R.; Caddick, S. *Chem. Commun.*, **2011**, *47*, 8781-8783.
- ⁵⁰ Ellman, G.L. *Arch. Biochem. Biophys.*, **1959**, *82*, 70– 77.
- ⁵¹ Kenyon, G.; Bruice, T. W. *Methods Enzymol.*, **1977**, *47*, 407 – 430.
- ⁵² Davis, B.G.; Maughan, M. A. T.; Green, M. P. ; Ullman, A.; Jones, J. B.; *Tetrahedron: Asymmetry*, **2000**, *11*, 245 –262.
- ⁵³ van Kasteren, S.I.; Kramer, H. B. ; Jensen, H. H.; Campbell, S. J. ; Kirkpatrick, J.; Oldham, N. J.; Anthony, D. C.; Davis, B. G. *Nature*, **2007**, *446*, 1105 –1109.
- ⁵⁴ Wender, P. A.; Goun, E. A. ; Jones, L. R.; Pillow, T. H.; Rothbard, J. B.; Shinde, R.; Contag, C. H. *Proc. Natl. Acad. Sci. USA*, **2007**, *104*, 10340 – 10345.
- ⁵⁵ Davis, B.; Claridge, T.; Farid, O.; Errey, J.; Chalker, J.; Thompson, S.; Grayson, E.; Bernardes, G. *Angew. Chem. Int. Ed.*, **2008**, *47*, 2244 –2247.
- ⁵⁶ Harpp, D.; Gleason, J. *J. Org. Chem.*, **1971**, *36*, 73.
- ⁵⁷ Jacquier, R.; Daunis, J.; Cavelier-Frontin, F. *Tetrahedron: Asymmetry*, **1992**, *3*, 85-94.
- ⁵⁸ Menez, A.; Dugave, C. *Tetrahedron: Asymmetry*, **1997**, *9*, 1453.
- ⁵⁹ Gavaret, M.; Nunez, J.; Cahnmann, H. *J. Biol. Chem.*, **1980**, *255*, 5281–5285.
- ⁶⁰ Palmer, D.E.; Pattaroni, C; Nunami, K.; Chadha, R.K.; Goodman, M.; Wakamiya, T.; Fukase, K. *J. Am. Chem. Soc.*, **1992**, *114*, 5634-5642.
- ⁶¹ Cohen, P. *Nat. Cell Biol.*, **2002**, *4*, E127–E130.
- ⁶² Dwek, R.A. *Chem. Rev.*, **1996**, *96*, 683–720.
- ⁶³ Ohtsubo, K.; Marth, J. D. *Cell*, **2006**, *126*, 855–867.
- ⁶⁴ Martin, C.; Zhang, Y. *Nat. Rev. Mol. Cell Biol.*, **2005**, *6*, 838–849.

-
- ⁶⁵ Kim, S.C.; Sprung, R.; Chen, Y.; Xu, Y.; Ball, H.; Pei, J.; Cheng, T.; Kho, Y.; Xiao, H.; Xiao, L.; Grishin, N.V.; White, M.; Yang, X.; Zhao, Y.; *Mol. Cell*, **2006**, *23*, 607–618.
- ⁶⁶ Holmes, T.J.; Lawton, R.G. *J. Am. Chem. Soc.* **1977**, *99*, 1984–1986.
- ⁶⁷ Bernardes, G.J.; Chalker, J.M.; Errey, J.C.; Davis, B.G. *J. Am. Chem. Soc.*, **2008**, *130*, 5052–5053.
- ⁶⁸ Mendiola, J.; Rincón, J.A.; Mateos, C.; Soriano, J.F.; Frutos, O.; Niemeier, J.K.; Davis, E.D. *Org. Process Res. Dev.*, **2009**, *13*, 263–267.
- ⁶⁹ Okamoto, R.; Souma, S.; Kajihara, Y. *J. Org. Chem.*, **2009**, *74*, 2494–2501
- ⁷⁰ Chalker, J.M.; Gunnoo, S.; Boutureira, O.; Gerstberger, S.C.; Gonzalez, M.F.; Bernardes, G.L.; Griffin, L.; Hailu, H.; Schofield, C.; Davis, B.G.; *Chem. Sci.*, **2011**, *2*, 1666–1676.
- ⁷¹ Roberts, J.; Warwick, G.P.; *Biochem. Pharmacol.*, **1961**, *6*, 205–216.
- ⁷² Onkenhout, W.; van Loon, W.M.; Buijs, W.; van der Gen, A.; Vermeulen, N.P. *Drug Metab. Dispos.*, **1986**, *14*, 608–612.
- ⁷³ Marchand, D.H.; Rimmel, R. P.; Abdel-Monem, M. M. *Drug Metab. Dispos.*, **1988**, *16*, 85–92.
- ⁷⁴ Chalker, J.; Lercher, L.; Rose, N.; Scholfield, C.; Davis, B. *Angew. Chem. Int. Ed.*, **2012**, *51*, 1835–1839.
- ⁷⁵ Macmillan, D. *Angew. Chem. Int. Ed.*, **2006**, *45*, 7668 – 7672.
- ⁷⁶ Wieland, V.T.; Bokelmann, E.; Bauer, L.; Lang, H.U.; Lau, H. *Liebigs Ann. Chem.*, **1953**, *583*, 129 – 149.
- ⁷⁷ Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. *Science*, **1994**, *266*, 776 – 779.
- ⁷⁸ Yan, L.; Dawson, P.E. *J. Am. Chem. Soc.*, **2001**, *123*, 526 – 533.
- ⁷⁹ Wan, Q.; Danishefsky, S.J. *Angew. Chem. Int. Ed.*, **2007**, *46*, 9248 – 9252.
- ⁸⁰ Haase, C.; Rohde, H.; Seitz, O. *Angew. Chem. Int. Ed.* **2008**, *47*, 6807 – 6810.
- ⁸¹ Chen, J.; Wan, Q.; Yuan, Y.; Zhu, J.; Danishefsky, S.J. *Angew. Chem. Int. Ed.*, **2008**, *47*, 8521 – 8524.
- ⁸² Botti, P.; Carrasco, M.R.; Kent, S.B.H. *Tetrahedron Lett.*, **2001**, *42*, 1831–1833.
- ⁸³ Offer, J.; Boddy, C.; Dawson, P.E. *J. Am. Chem. Soc.*, **2002**, *124*, 4642–4646.
- ⁸⁴ Kemp, D.; Galakatos, N.G.; Bowen, B.; Tan, K. *J. Org. Chem.* **1986**, *51*, 1821–1829.
- ⁸⁵ Brik, A.; Yang, Y.; Ficht, S.; Wong, C.H. *J. Am. Chem. Soc.*, **2006**, *128*, 5626–5627.
- ⁸⁶ Muir, T.W.; Sondhi, D.; Cole, P.A. *Proc. Natl. Acad. Sci. USA*, **1998**, *95*, 6705–6710.
- ⁸⁷ Severinov, K.; Muir, T.W. *J. Biol. Chem.*, **1998**, *273*, 16205–16209.

-
- ⁸⁸ Hofmann, R.M.; Muir, T.W. *Curr. Opin. Biotechnol.*, **2002**, *13*, 297–303.
- ⁸⁹ Muir, T.W. *Annu. Rev. Biochem.*, **2003**, *72*, 249–289.
- ⁹⁰ Saxon, E.; Armstrong, J.; Bertozzi, C.R. *Org. Lett.*, **2000**, *2*, 2141–2143.
- ⁹¹ Staudinger, H.; Meyer, J. *Helv. Chim. Acta*, **1919**, *2*, 635–646.
- ⁹² Saxon, E.; Bertozzi, C. R. *Science*, **2000**, *287*, 2007–2010.
- ⁹³ Richardson, J.; Chan, C.; Blanc, J.; Saadi, M.; Macmillan, D. *Org. Biomol. Chem.* **2010**, *8*, 1351–1360.
- ⁹⁴ Richardson, J.; Macmillan, D. *Org. Biomol. Chem.* **2008**, *6*, 3977–3982.
- ⁹⁵ Baker, D.P.; Lin, E.Y.; Lin, K.; Pellegrini, M.; Petter, R. C.; Chen, L.; Arduini, R.M.; Brickelmaier, M.; Wen, D.; Hess, D.; Chen, L.; Grant, D.; Whitty, A.; Gill, A.; Lindner, A.R.; Pepinsky, B. *Bioconjugate Chem.*, **2006**, *17*(1), 179 – 188.
- ⁹⁶ Jencks, W.P. *J. Am. Chem. Soc.*, **1959**, *81*, 475 – 481.
- ⁹⁷ Geoghegan, K.; Emery, M.J.; Martin, W.H.; McColl, A. S.; Daumy, G. *Bioconjugate Chem.* **1993**, *4*, 537–544.
- ⁹⁸ Geoghegan, K.F.; Stroh, G.J. *Bioconjugate Chem.* **1992**, *3*, 138–146.
- ⁹⁹ Li, X.; Zhang, L.; Hall, S. E.; Tam, J.P. *Tetrahedron Lett.*, **2000**, *41*, 4069–4073.
- ¹⁰⁰ Gilmore, J.M.; Scheck, R.A.; Esser-Kahn, A.P.; Joshi, N.S.; Francis, M.B. *Angew. Chem. Int. Ed.*, **2006**, *45*, 5307 –5311.
- ¹⁰¹ Mazmanian, S. K.; Liu, G.; Ton-That, H.; Schneewind, O. *Science*, **1999**, *285*, 760–763.
- ¹⁰² Perry, A. M.; Ton-That, H.; Mazmanian, S. K.; Schneewind, O. *J. Biol. Chem.*, **2002**, *277*, 16241–16248.
- ¹⁰³ Mao, H.; Hart, S.A; Schink, A.; Pollok, B.A.; *J. Am. Chem. Soc.*, **2004**, *126*, 2670–2671.
- ¹⁰⁴ Williamson, D.J.; Fascione, M.A.; Webb, M.E.; Turnbull, W. *Angew. Chem. Int. Ed.*, **2012**, *51*, 9377 –9380.
- ¹⁰⁵ Ling, J.L.; Policarpo, R.L; Rabideau, A.E.; Liao, X.; Pentelute, B.L. *J. Am. Chem. Soc.*, **2012**, *134*, 10749–10752.
- ¹⁰⁶ Kiick, K. L.; Weberskirch, R.; Tirrell, D. A. *FEBS Lett.*, **2001**, *502*, 25–30.
- ¹⁰⁷ Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. *Proc.Natl. Acad. Sci U.S.A.*, **2002**, *99*, 19–24.
- ¹⁰⁸ Link, A. J.; Vink, M. K.; Tirrell, D. A. *J. Am. Chem. Soc.*, **2004**, *126*, 10598–10602.

-
- ¹⁰⁹ Rostovtsev, V.V; Green, L.G.; Fokin, V.V.; Sharpless, K.B. *Angew. Chem. Int. Ed.*, **2002**, *41*, 2596-2599.
- ¹¹⁰ Sommer, S.; Weikart, N.; Brockmeyer, A.; Janning, P.; Mootz H.D. *Angew. Chem. Int. Ed.*, **2011**, *42*, 9888-9892.
- ¹¹¹ Tang, Y.; Tirrell, D. A. *J. Am. Chem. Soc.*, **2001**, *123*, 11089–11090.
- ¹¹² Wang, Q.; Chan, T. R.; Hilgraf, R.; Fokin, V. V.; Sharpless, K. B.; Finn, M. G. *J. Am. Chem. Soc.*, **2003**, *125*, 3192–3193.
- ¹¹³ Yoshikawa, E.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules*, **1994**, *27*, 5471–5475.
- ¹¹⁴ Renner, C.; Alefelder, S.; Bae, J. H.; Budisa, N.; Huber, R.; Moroder, L. *Angew. Chem. Int. Ed.*, **2001**, *40*, 923–925.
- ¹¹⁵ Minks, C.; Huber, R.; Moroder, L.; Budisa, N. *Biochemistry*, **1999**, *38*, 10649 - 10659.
- ¹¹⁶ Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. *Science*, **1989**, *244*, 182–188.
- ¹¹⁷ Nowak, M. W.; Kearney, P. C.; Sampson, J. R.; Saks, M. E.; Labarca, C. G.; Silverman, S. K.; Zhong, W.; Thorson, J. *Science*, **1995**, *268*, 439–442.
- ¹¹⁸ Kiga, D.; Sakamoto, K.; Kodama, K.; Kigawa, T.; Matsuda, T.; Yabuki, T.; Shirouzu, M.; Harada, Y.; Nakayama, H.; Takio, K.; Hasegawa, Y.; Endo, Y.; Hirao, I.; Yokoyama, S. *Proc. Natl. Acad. Sci. U.S.A.*, **2002**, *99*, 9715–9720.
- ¹¹⁹ Xie, J.; Schultz, P. G. *Curr. Opin. Chem. Biol.*, **2005**, *9*, 548–554.
- ¹²⁰ Hendrickson, T. L.; Crecy-Lagard, V.; Schimmel, P. *Annu. Rev. Biochem.*, **2004**, *73*, 147–176.
- ¹²¹ Liu, D.R.; Schultz, P.G. *Proc. Natl. Acad. Sci. U.S.A.*, **1999**, *96*, 4780-4785.
- ¹²² Sakamoto, K.; Hayashi, A.; Sakamoto, A.; Kiga, D.; Nakayama, H.; Soma, A.; Kobayashi, T.; Kitabatake, M.; Takio, K.; Saito, K.; Shirouzu, M.; Hirao, I.; Yokoyama, S. *Nucleic Acids Res.*, **2002**, *30*, 4692–4699.
- ¹²³ Ryan, A.; Mehl, J.; Anderson, C.; Santoro, S.W.; Wang, L.; Martin, A.; King, D.; Horn, D.M.; Schultz, P.G. *J. Am. Chem. Soc.*, **2003**, *125*, 935-939.
- ¹²⁴ Rackham, O.; Chin, J. W. *Nat. Chem. Biol.*, **2005**, *1*, 159–166.
- ¹²⁵ Hohsaka, T.; Ashizuka, Y.; Taira, H.; Murakami, H.; Sisido, M. *Biochemistry* *40*, 11060–11064.
- ¹²⁶ Wang, K.; Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. *Nat. Biotechnol.*, **2007**, *25*, 770–777.

-
- ¹²⁷ Joo, C.; Balci, H.; Ishitsuka, Y.; Buranachai, C.; Ha, T. *Annu. Rev. Biochem.*, **2008**, *77*, 51–76.
- ¹²⁸ Cecconi, C.; Shank, E.A.; Bustamante, C.; Marqusee, S. *Science*, **2005**, *309*, 2057–2060.
- ¹²⁹ Stephanopoulos, N.; Tong, G.J.; Hsiao, S.C.; Francis, M.B. *ACS Nano*, **2010**, *4*, 6014–6020.
- ¹³⁰ Banerjee, D.; Liu, A.P.; Voss, N.R.; Schmidt, S.L.; Finn, M.G. *ChemBioChem*, **2010**, *11*, 1273–1279.
- ¹³¹ Oh, I.K.; Mok, H.; Park, T.G. *Bioconjug. Chem.*, **2006**, *17*, 721–727.
- ¹³² Ratner, V.; Kahana, E.; Eichler, M.; Haas, E. *Bioconjugate Chem.*, **2002**, *13*, 1163–1170.
- ¹³³ Kao, M.W.; Yang, L.L.; Lin, J.C.; Lim, T.S.; Fann, W.; Chen, R.P. *Bioconjugate Chem.*, **2008**, *19*, 1124–1126.
- ¹³⁴ Smith, J.J.; Conrad, D.W.; Cuneo, M.J.; Hellinga, H.W. *Protein Sci.*, **2005**, *14*, 64–73.
- ¹³⁵ Jäger, M.; Michalet, X.; Weiss, S. *Protein Sci.*, **2005**, *14*, 2059–2068.
- ¹³⁶ Antos, J. M.; Chew, G.-L.; Guimaraes, C. P.; Yoder, N. C.; Grotenbreg, G. M.; Popp, M. W.L.; Ploegh, H. L. *J. Am. Chem. Soc.* **2009**, *131*, 10800–10801.
- ¹³⁷ Yi, L.; Sun, H.; Itzen, A.; Triola, G.; Waldmann, H.; Goody, R. S.; Wu, Y. W., *Angew. Chem. Int. Ed.*, **2011**, *50*, 8287–8290.
- ¹³⁸ Dawson, P.E.; Muir, T.W.; Clark-Lewis, I.; Kent, S.B. *Science*, **1994**, *266*, 776 – 779.
- ¹³⁹ Santoso, Y.; Joyce, C.; Potapova, O.; Reste, L.; Hohlbein, J.; Torella, J.; Grindley, N.; Kapanidis, A. *Proc. Natl. Acad. Sci. USA*, **2010**, *107*, 715–720.
- ¹⁴⁰ Kovacs, E.; Hooker, J.; Romanini, W.; Holder, P.; Berry, K.; Francis, M. *Bioconjugate Chem.*, **2007**, *18*, 1140–1147.
- ¹⁴¹ Yang, J.Y.; Yang, W. Y., *J. Am. Chem. Soc.* **2009**, *131*, 11644–11645.
- ¹⁴² De Rosa, L.; Cortajaren A.; Romanelli, A.; Regan, L.; D'Andrea, L. *Org. Biomol. Chem.*, **2012**, *10*, 273–280.
- ¹⁴³ Muralidharan, V.; Muir, T. W. *Nat. Methods*, **2006**, *3*, 429–438.
- ¹⁴⁴ Simon, M.; Zangemeister-Wittke, U.; Pluckthun, A. *Bioconjugate Chem.* **2012**, *23*, 279–286.
- ¹⁴⁵ Sletten, E. M.; Bertozzi, C. R. *Org. Lett.*, **2008**, *10*, 3097–3099.
- ¹⁴⁶ Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J.W. *Nature*, **2010**, *464*, 441–444.
- ¹⁴⁷ Wan, W.; Huang, Y.; Wang, Z.; Russell, W. K.; Pai, P. J.; Russell, D.H.; Liu, W. R. *Angew. Chem. Int. Ed.*, **2010**, *49*, 3211– 3214.

-
- ¹⁴⁸ Chalker, J.; Bernardes, G.; Lin, Y.; Davis, B. *Chem. Asian. J.*, **2009**, *4*, 630-640.
- ¹⁴⁹ Moody, P.; Smith, M. E. B.; Ryan, C. P.; Chudasama, V.; Baker, J. R.; Molloy, J.; Caddick, S. *ChemBioChem*, **2012**, 38-41.
- ¹⁵⁰ S. Hubbard, J. Thornton, *Naccess* **1993**, University of Manchester, Manchester, U.K based on B. Lee, F. M. Richards, *J. Mol. Biol.* **1971**, *55*, 379-400 by Paul Moody Department of Chemistry, University College London
- ¹⁵¹ Whitesides, G.; Moran, J.; Butler, J.; Burns, J. *J. Org. Chem.*, **1991**, *56*, 2648-2650.
- ¹⁵² Harpp, D.; Gleason, J. *J. Am. Chem. Soc.*, **1971**, *19*, 2437.
- ¹⁵³ Harpp, D.; Ash, D.; Smith, R. *J. Org. Chem.*, **1980**, *45*, 5155.
- ¹⁵⁴ Harpp, D.; Gleason, J. *J. Org. Chem.*, **1971**, *36*, 73.
- ¹⁵⁵ Davis, B.; Claridge, T.; Farid, O.; Errey, J.; Chalker, J.; Thompson, S.; Grayson, E.; Bernardes, G. *Angew. Chem. Int. Ed.* **2008**, *47*, 2244–2247.
- ¹⁵⁶ Krafft, G.; Siddall, T. *Tetrahedron Lett.*, **1985**, *26*, 4867-4870.
- ¹⁵⁷ Mintz, M.; Walling, C. *Org. Syn.*, **1969**, *49*, 184.
- ¹⁵⁸ Ryan, C.; Smith, M.E.B.; Schumacher, F.; Grohmann, D.; Papaioannou, D.; Waksman, G.; Werner, F.; Baker, J.; Caddick, S. *Chem. Commun.*, **2011**, *47*, 5452-5455.
- ¹⁵⁹ Volkin, D.; Mach, H.; Middaugh, C. *Mol. Biotechnol.*, **1997**, *8*, 105.
- ¹⁶⁰ Rudinger, J.; Ruegg, U. *Methods in Enzymology*, **1977**, *47*, 111-116.
- ¹⁶¹ Washburne, S.; Peterson, W.; Berman, D. *J. Org. Chem.*, **1972**, *37*, 1738.
- ¹⁶² Engbets, J.; Rispens, T. *J. Phys. Org. Chem.* **2005**, *18*, 908–917.
- ¹⁶³ Labidalle, S.; Souchard, J.; Hoffman, P.; Gennison, V.; Petit, C. *Chem. Pharm. Bull.*, **2000**, *48*, 1634-1638.
- ¹⁶⁴ Valverde, S.; Bao, R.; Fenude, E.; Herradon, B. *J. Org. Chem.*, **1996**, *61*, 1143-1147.
- ¹⁶⁵ Nakamura, S.; Hirao, H.; Ohwada, T. *J. Org. Chem.*, **2004**, *69*, 4309-4316.
- ¹⁶⁶ Raftery, M. J. *Rapid Commun. Mass Spectrom.*, **2005**, *19*, 674–682.
- ¹⁶⁷ Washburne, S.; Peterson, W.; Berman, D. *J. Org. Chem.*, **1972**, *37*, 1738.