Approaches to rapid protein modification using sortase

Zoe Louise Pike Arnott

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

School of Chemistry

April 2018

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to other workers.

Chapter 2: Use of sortase (7M) alongside depsipeptide substrates includes the use of GGG-CTB which was provided by Darren Machin and His$_6$-TEV, provided by Gemma Wildsmith. The Rhodamine peptide used in FRET studies was synthesised by Kristian Hollingsworth.

Chapter 3: Use of protein-DNA conjugates in structural studies includes the use of an azide-oligonucleotide synthesised by Darren Machin on placement at GlaxoSmithKline. HS-AFM work was done in collaboration with Andrew Lee who prepared and imaged the DNA origami.

Chapter 4: Biological studies of ADC as a potential antibiotic target includes content from the following publication: "The Mechanism of Regulation of Pantothenate Biosynthesis by the PanD–PanZ·AcCoA Complex Reveals an Additional Mode of Action for the Antimetabolite N-Pentyl Pantothenamide (N5-Pan)", Biochemistry, 2017, volume 56, pages 4931-4939, Zoe L. P. Arnott, Shingo Nozaki, Diana C. F. Monteiro, Holly E. Morgan, Arwen R. Pearson, Hironori Niki, and Michael E. Webb. The manuscript was composed by M. E. Webb. The candidate performed ITC binding studies using the PanZ mutants. Produced pentyl pantothenamide for the assay of growth inhibition and ethyldethiacoenzyme A alongside MChem student, Holly Morgan, and carried out purification and subsequent analysis of the interaction with PanZ. The biotin-depsipeptide used for the sortase-mediated labelling of PanD(T57V) was synthesised by Jack Caudwell.

Acknowledgements

Firstly, I would like to thank my supervisor, Mike, for his support over the past three and a half years, he has always been available to answer any questions, no matter how stupid, and provide advice and direction. I would also like to thank my co-supervisor Arwen for putting me up in Hamburg on a number of occasions.

I would like to thank every member of lab 1.49 over the past few years with a few particular mentions: Dan, whose thesis has been a staple on my shelf from day one; Clare, who provided emotional support both in the lab and in the pub assisted by copious amounts of gin; Gemma, who has taught me almost everything I know in the Biolab, for which I am eternally grateful; Darren who dealt with a lot of my stupid questions; Chadamas, whose smiling face could always brighten my day and Holly for being grumpier than me and of course for being hard working throughout her MChem.

I would also like to express my appreciation for Iain Manfield's time and patience when using both the SPR and ITC, particularly at times when I was ready to throw the ITC machine out of the window! And also Martin Huscroft for HPLC assistance and Stuart Warriner for keeping the HRMS alive. Thank you also to the University of Leeds and University of Hamburg for their funding which has allowed me to carry out this work.

I would like to thank all of my friends, particularly Stephen who has had to calm a lot of bad moods as a result of many stressful days in the lab. And finally, thank you to my family, who have always believed in me, especially my mum, who has supported me through far too many years of education, and given me the best advice to enable me to be the person I am today.

Abstract

Sortase A-mediated labelling is a versatile method for protein modification without the need for unnatural amino acid incorporation or complex labelling reagent synthesis. Over the past decade several developments have improved the efficiency of these reactions. These have included the investigation of different types of labelling substrates to reduce the reversible nature of the reactions. In addition to this, the introduction of mutant enzymes has increased the rate of reaction by up to 120-fold, however the labelling is still not quantitative.

In this work one of the faster enzyme mutants, sortase 7M, was used alongside depsipeptide substrates which have been previously reported to take part in irreversible labelling reactions. Optimised conditions have been established to enable proteins to be labelled in under 3 hours using only 2-3 equivalents of depsipeptide substrate. Using this approach, proteins have been labelled with species which have allowed them to be used in a wide variety of biophysical applications including for the characterisation of protein-protein interactions by SPR and FRET, and for the visualisation of protein-protein interactions at a single-molecule level using high-speed AFM.

Abbreviations

AcCoA	Acetyl Coenzyme A
ACP	Acyl carrier proteins
ADP	Adenosine diphosphate
AFM	Atomic force microscopy
AIM	Auto induction media
ATP	Adenosine triphosphate
BCN	Bicylo[6.1.0]non-4-yne
BTTAA	2‑[4‑{(bis[(1-tert-butyl-1H‑1,2,3-triazol-4‑yl)methyl]amino) methyl}-1H-1,2,3-triazol-1-yl] acetic acid
C. elegans	*Caenorhabditis elegans*
C. ternatea	*Clitoria ternatea*
CoaD	Phosphopantetheine adenylyltransferase
CoaE	Dephosphocoenzyme A kinase
cryo-EM	Cryogenic electron microscopy
CTB	Cholera toxin B-subunit
CuAAC	Copper catalysted azide alkyne 1,3 dipolar cycloaddition
CV	Column volume
D. melanogaster	*Drosophila melanogaster*
DIBAC	Aza-dibenzocyclooctyne
DIBO	Dibenzocyclooctynes
DIFO	Difluorocyclooctynes
dNTPs	Deoxyribose nucleoside triphosphates
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EtdtCoA	Ethyldethia Coenzyme A

FITC	Fluorescein isothiocyanate
His$_6$-tag	Hexahistidine affinity tag
HOMO	Highest occupied molecular orbital
HRMS	High resolution mass spectrometry
HS-AFM	High speed atomic force microscopy
IED	Inverse electron-demand
IEDDAC	Inverse electron-demand Diels Alder cycloaddition
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
k_{cat}	Catalytic rate constant
LB	Lysogeny broth
LED	Light emitting diode
LUMO	Lowest unoccupied molecular orbital
MBP	Maltose binding protein
MBP-BCN	Bicylononyne labelled MBP
MBP-DNA	MBP-DNA fusion
MBP-oligo	MBP-oligonucleotide fusion
MIC	Minimal inhibitory concentration
N$_3$-oligo	Azide-functionalised oligonucleotide
N5-Pan	N-pentyl pantothenamide
NCL	Native chemical ligation
NF-depsipeptide	Non-fluorescent depsipeptide
NF-peptide	Non-fluorescent peptide
Ni-NTA	Nickel nitrilotriacetic acid resin
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
SANS	Small angle neutron scattering

SAS	Small angle scattering
SAXS	Small angle X ray scattering
SDM	Site-directed mutagenesis
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
SPAAC	Strain promoted azide alkyne cycloaddition
SPM	Scanning probe microscopy
SPPS	Solid phase peptide synthesis
SPR	Surface plasmon resonance
Srt 5M	Sortase A pentamutant (P94R, D160N, D165A, K190E and K196T)
Srt 7M	Sortase A heptamutant (P94R, D160N, D165A, K190E and K196T, E105K and E108A/Q)
s-TCO	*Trans*-bicyclo[6.1.0]non-4-ene-9-ylmethanol
TBAI	Tetrabutylammonium iodide
TBE	Tris-borate/EDTA acrylamide gel
TCO	T*rans*-cyclooctene
TEV	Tobacco etch virus
TFA	Trifluoro acetic acid
THPTA	Tris[(1-hydroxypropyl-1H-1,2,3-triazol-4-yl)methyl] amine
UV	Ultraviolet light
WT	Wild-type

Contents

1 Background

1.1 Introduction

Site-specific labelling is an important tool for the study of protein structure, function and biophysical characteristics which can assist in the development of biopharmaceutical products. There are many approaches for the modification of proteins including labelling in a residue-specific manner, generally achieved through the chemical modification of cysteine or lysine residues.[1,2] However, specific labelling using this method is limited by the requirement for the protein to only contain one modifiable residue. A better approach is therefore to introduce a new functional group which is able to take part in bio-orthogonal reactions. This term first coined by Bertozzi and co-workers[3] refers to reactions where the reactants and products non-toxic and kinetically, thermodynamically and metabolically stable in living systems. Examples of these reactions include the Staudinger ligation and azide-alkyne cycloadditions (Section 1.2). There are many ways to incorporate the functional groups required for the reactions in a site-specific manner including genetic code expansion and enzymatic methods, both of which will be discussed in more detail in Section 1.3.

1.2 Bio-orthogonal labelling methods

Over the last 20 years, a large number of bio-orthogonal reactions have been developed. The earliest were mainly dependent on azides but now more varied species have been used including; alkenes, alkynes and tetrazines. The development of rapid, quantitative and chemoselective reactions has allowed several different biological systems to be studied over recent years using the chemistry now accessible through these methods.

1.2.1 Staudinger ligation

Azides are a popular functional group for bio-orthogonal labelling as they are not found naturally in any biological systems and are very small and kinetically stable under physiological conditions. In 1919, Staudinger and Meyer[4] discovered that azides and triaryl phosphine species react to form iminophosphoranes (isoelectronic with the aza-ylide form) near quantitatively under mild conditions (Scheme 1.1a). Based upon this, Bertozzi and co-workers developed the Staudinger ligation in 2000.[5] By introducing an ester bond adjacent to the phosphine on an aryl ring, an intramolecular cyclisation reaction takes place between the aza-ylide and the electrophilic ester carbonyl. This intermediate is then hydrolysed to form a stable amide bond (Scheme 1.1b).[4,5] Further to this work, the "traceless Staudinger ligation" was reported which allows the ligation product to be released from the phosphine oxide (Scheme 1.1c).[6] This involves the use of a phosphine reagent containing an acyl phenol group attached to a cleavable linker (**1.05**). The nitrogen (**1.06**) attacks the acyl-carbonyl causing the cleavage of the linker and the phosphonium group, leaving the amide ligation product (**1.07**) and releasing the corresponding phosphine oxide (**1.08**).
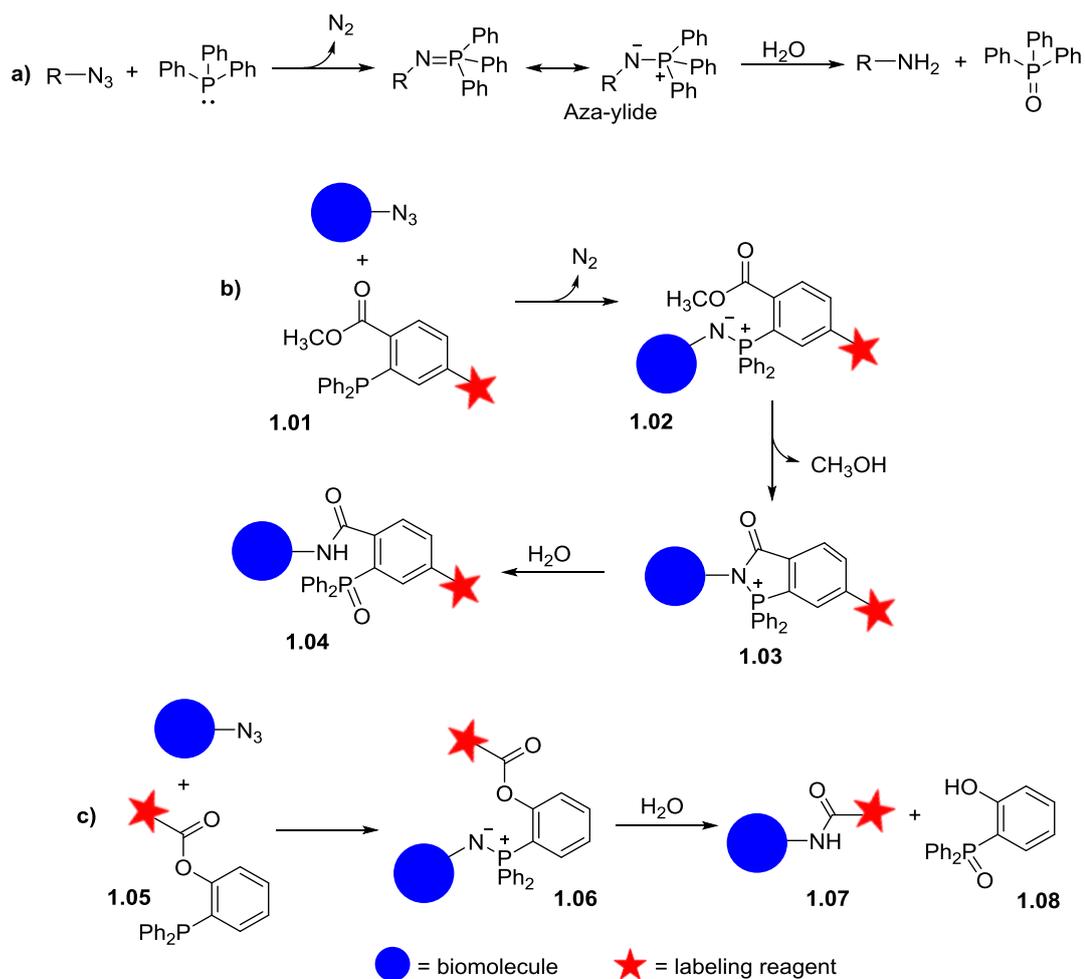
Scheme 1.1: a) Staudinger reduction between an azide and a triphenylphosphine to give a primary amine. b) Staudinger ligation where the product remains attached to the phosphine oxide moiety. c) Traceless Staudinger ligation where the product is able to be released from the phosphine oxide. [5–7]

This approach has been used in a number of protein labelling studies, for example, azide-containing proteins have been labelled with phosphine-fluorophores and phosphine-polyethyleneglycol (PEG).[8] The major limiting factor associated with application of the Staudinger ligation is the slow reaction rate with second-order rate constants $\sim 10^{-3}$ $M^{-1}s^{-1}$.[7] In addition to this, the phosphines are sensitive to oxidative inactivation meaning high concentrations of the phosphine reagents can be required.[9]

1.2.2   Azide-alkyne cycloaddition reactions

1.2.2.1   Copper-catalysed cycloaddition

Azides also react as 1,3 dipolarophiles in [3+2] cycloaddition with terminal alkynes to form stable triazoles using copper (I) catalysts. This copper catalysed azide-alkyne 1,3 dipolar cycloaddition reaction (CuAAC) was initially discovered in 2002 and is a well-known example of "click" chemistry.[10] The reaction proceeds *via* the formation of a copper acetylide which activates the alkyne for reaction with the azide to form a six membered Cu (III) metallacycle intermediate.[11] Recent work in this field has indicated the involvement of two copper (I) ions in the rate determining step with the copper co-ordinating to both the azide and alkyne (Scheme 1.2).[12]
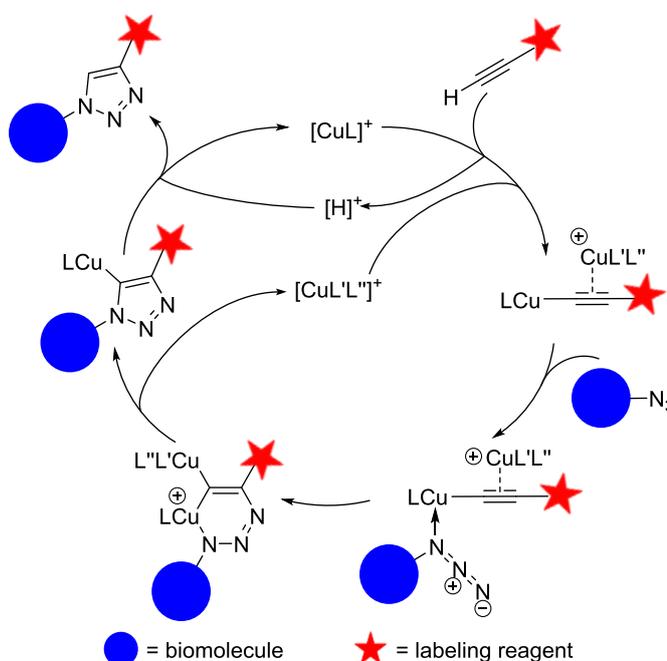
Scheme 1.2: Mechanism for the copper catalysed azide alkyne 1,3 dipolar cycloaddition reaction.[12]

These reactions have been found to proceed up to five orders of magnitude faster than the Staudinger ligation under physiological conditions in cell lysates.[9] Both alkyne and azide containing proteins have been used for biological studies using CuAAC both *in vivo* and *in vitro*.[13–16]

As the copper species required for CuAAC must be maintained as Cu(I), most reactions require the presence of a reducing agent, most commonly sodium ascorbate. This is avoided in most

bioconjugation reactions as copper and sodium ascorbate have been found to be detrimental to biological polymers due to the generation of reactive oxygen species.[13,17,18] In addition to this dehydroascorbate and other ascorbate by-products have been found to potentially cause covalent modifications to arginine, lysine and N-terminal cysteine residues which has the potential to result in protein aggregation.[19,20]

To reduce the concentrations of Cu(I) and sodium ascorbate required for efficient CuAAC reactions, Hong *et al.* introduced water-soluble ligands which coordinate to the Cu(I) to form an activated catalyst for the cycloaddition. In addition to lowering the concentrations of Cu(I) required, they have also been seen to act as a reductant to protect the cells from the reactive oxygen species. One such ligand is BTTAA (**1.09**): the acetate side chain is thought to act as a weak σ-donor, coordinating to the Cu(I) metal centre and increasing the electron density. This ultimately encourages the formation of the strained copper metallacyle intermediate, accelerating the rate of cycloaddition.[21] Another ligand, THPTA (**1.10**), is thought to intercept reactive oxygen species as they are generated in the coordination sphere and act as a sacrificial reductant, this means excess of the ligand is needed. However, this has not been seen to hinder the rate of CuAAC. With regards to the sodium ascorbate, aminoguanidine (**1.11**) has been found to be a useful additive to capture reactive aldehydes which are generated from the hydrolysis of dehydroascorbate.[22]
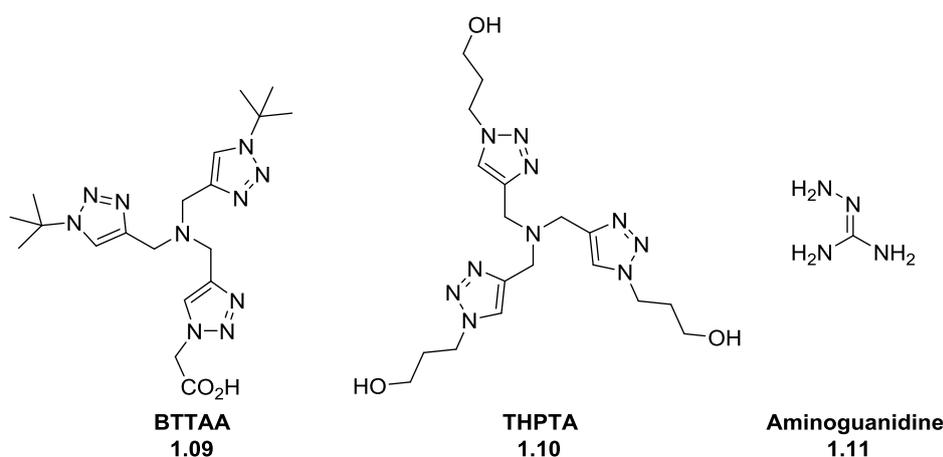
Figure 1.1: Water soluble ligands BTTAA, THPTA and aminoguanidine for use in CuAAC as a method of lowering the Cu(I) concentrations required.[21,22]

Both BTTAA and THPTA (Figure 1.1) have been applied to labelling live mammalian cell surfaces by ligand-assisted CuAAC with Cu(I) concentrations between 100-500 µM which is tolerated for cell surface labelling studies, with overall reaction rate constants being between 10-500 $M^{-1}s^{-1}$. Although these methods greatly improve the potential of CuAAC in biological labelling studies, problems can still be experienced if any groups present bind strongly to Cu, such as hexahistadine-tags (His$_6$-tag), as this binding can lead to protein aggregation. In addition to this, the azide or alkyne incorporated into a biomolecule can be sterically hindered making it inaccessible for the catalyst; however, if the protein is stable enough, increasing reaction temperature can mitigate against this.[22]

1.2.2.2 Strain-promoted azide-alkyne cycloaddition

Codelli *et al*. demonstrated that "click" chemistry could be successfully achieved by reacting azides with ring-strained alkynes.[23] Such strain-promoted azide-alkyne cycloaddition (SPAAC) removes the necessity for toxic catalysts. The ring strain required can be introduced using cycloalkynes, e.g. cyclooctyne, the smallest isolable cyclic alkyne, which has been used to label azides in living systems (Scheme 1.3).[24]
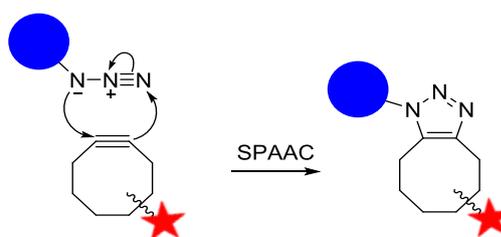
Scheme 1.3: Mechanism for strain-promoted azide-alkyne cycloaddition between an azide and cyclooctyne.[25]

These reactions are slower than CuAAC but more reactive cyclooctynes have been developed to improve the rate of reaction (Figure 1.2). For example, electron-poor fluoride groups were added to yield difluorocyclooctynes (DIFO, **1.12**),[23] benzannulation was introduced in dibenzocyclooctynes (DIBO, **1.13**)[26] and the most reactive of the cyclooctyne structures

developed so far is aza-dibenzocyclooctyne (DIBAC, **1.14**).[27] Bicyclo[6.1.0]non-4-yne (BCN, **1.15**) is a much smaller and stable strained alkyne which is relatively simple to synthesise.[28] However, it is slightly less reactive than the DIBAC compounds. Despite the number of strained alkynes investigated over the years, the maximum rate of these reactions has plateaued at ~ 0.2-0.5 $M^{-1} s^{-1}$.[29]
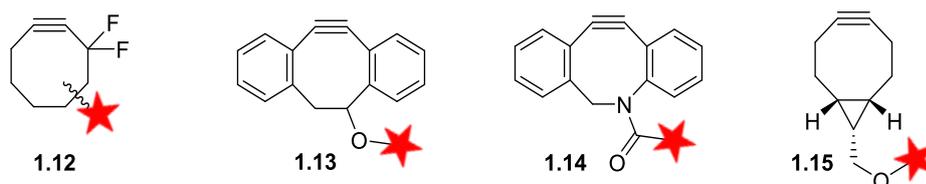
*Figure 1.2: Strained cyclooctynes DIFO (**1.12**), DIBO (**1.13**), DIBAC (**1.14**) and BCN (**1.15**) can be used in SPAAC to improve reaction rates.[24,26–28]*

Dommerholt *et al.* decided to turn their attention to the structure of the azide to further improve the rate of SPAAC reactions. They found that electron-poor aryl azides reacted with BCN 30-fold faster, with a noticeable difference in the rates for the aliphatic cyclooctyne, versus the benzannulated cyclooctynes, with BCN reacting six times faster than DIBAC when reacted with phenylazide. Further investigation showed that the reactions were proceeding *via* an inverse electron-demand (IED) mechanism, the opposite of what is expected for a SPAAC reaction. Through mechanistic studies it was seen that the dominant donor-acceptor orbital interactions involve the π-electrons of BCN. HOMO, and the LUMO of the azide. This is the reason behind the increased reaction rates observed by introducing electron-withdrawing groups to the aromatic azide. Overall, this work opened up new potential for SPAAC reactions with rates much more comparable to copper-catalysed reactions using BCN and aromatic azides which are accessible through synthesis or commercial means.[30]

1.2.3 Inverse electron-demand Diels-Alder cycloaddition

One of the fastest reported bio-orthogonal labelling reactions is that of tetrazines and strained alkenes or alkynes. This reaction is known as inverse electron-demand [4 + 2] Diels-Alder cycloaddition (IEDDAC), in the classical Diels-Alder reaction an electron-rich diene reacts with an electron-poor dienophile. However, in IEDDAC, an electron-poor azadiene (e.g. tetrazine) reacts with an electron-rich dienophile (e.g. strained alkene/alkyne). The rates of these reactions can be finely tuned by manipulating the $LUMO_{diene}$-$HOMO_{dienophile}$ energy gap, for example the LUMO can be lowered by introducing electron withdrawing groups on the tetrazine and the HOMO energy can be increased by introducing strain or adding electron density into the dienophile.[31] The use of tetrazines means that the reaction proceeds initially to form a cycloadduct, which then undergoes a retro Diels-Alder reaction, losing N_2, yielding a 4,5-dihydropyridizine (**1.16**). The product then isomerises to form the 1,4-dihydropyridazine (**1.17**) as the final compound in protic solvents. (Scheme 1.4)
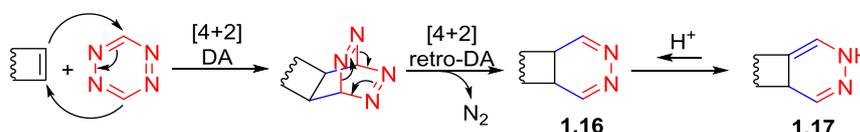
Scheme 1.4: Inverse electron-demand Diels Alder cycloaddition between a tetrazine and an alkene.[31,32]

Tetrazines react with a variety of strained alkenes including cyclopropene[33] and norbornene[34]; however, the cycloadditions with *trans*-cyclooctenes (**1.18**) are the fastest with the alkene double bond twisting the ring, giving it a very high energy HOMO.[32] The ring strain can be increased even further by fusing it to a cyclopropane to give bicyclo[6.1.0]-non-4-ene compounds (**1.19**). This forces the structure into a "half-chair" conformation which has a subsequent effect on the rate of reaction (Figure 1.3). For example, upon reaction with 3,6-diphenyl-*s*-tetrazine (**1.20**), *trans*-bicyclo[6.1.0]non-4-ene-9-ylmethanol (*s*-TCO, **1.21**) reacted 160 times faster than *trans*-cyclooctene (TCO, **1.18**).[35]
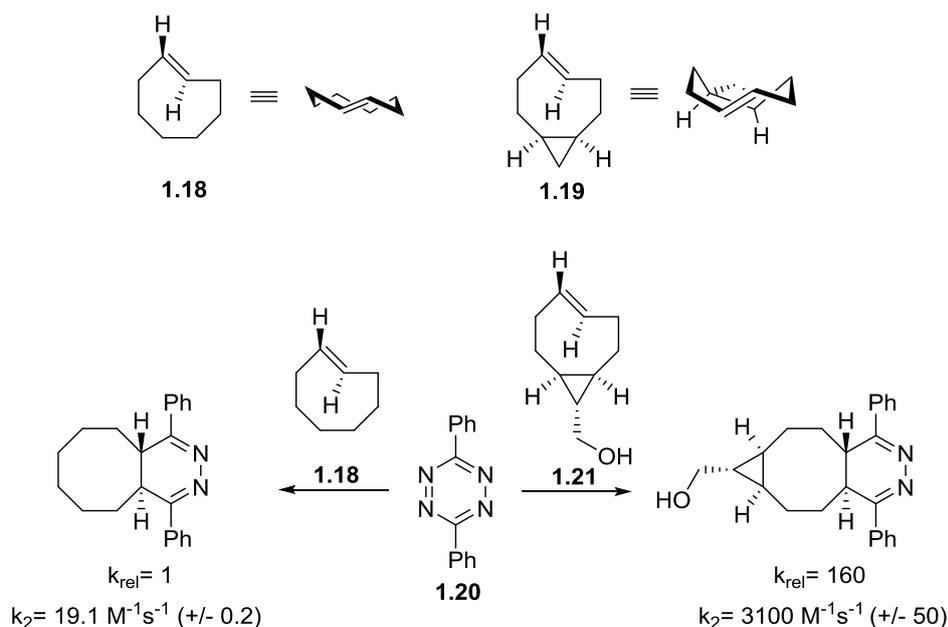
Figure 1.3: The difference in chair conformations of trans-cyclooctene and trans-bicyclo[6.1.0]non-4-ene-9-ylmethanol and the effect the additional ring strain has on the rate of reaction with 3,6-diphenyl-s-tetrazine.[35]

In addition to the strained alkenes, strained alkynes have also been seen to be able to react with tetrazines through IEDDAC as the additional strain and electron density increases the HOMO energy. These reactions also occur quickly with tetrazines but not as fast as *trans*-cyclooctenes. However, in a similar manner, the incorporation of the *cis*-fused cyclopropane improved the reaction rate when comparing cyclooctyne and bicyclononyne structures, reacting with 3,6-diphenyl-*s*-tetrazine 47-fold faster.[36] The benefit of using these alkyne structures with tetrazines is that only one product is formed, unlike the strained alkenes which also form regioisomers. This makes it very useful in situations where the orientation of the probe is important.

Although these are the fastest bio-orthogonal labelling reactions they are often only used for specific *in vivo* applications which involve extremely low concentrations,[37] with SPAAC remaining a popular choice for many copper-free cycloaddition labelling experiments. The synthesis of tetrazines for use in this context is not trivial, generally forming bulky compounds

and requiring the use of anhydrous hydrazine. Additionally, once placed within biological environments, the tetrazines have been found to undergo hydrolysis or decomposition when exposed to some endogenous nucleophiles.[38–40] Ultimately, no matter which method chosen, these unique chemical functionalities all need to be incorporated into the biological system of choice. There are a variety of different approaches that can be used to achieve this, some of which will be discussed in the next part of this chapter.

1.3 Incorporation of bio-orthogonal functionalities into proteins

The incorporation of bio-orthogonal functionalities into proteins can be done globally or site-specifically. Global incorporation is when the modification is introduced residue-specifically, where a natural residue is replaced across the whole protein by an unnatural one. If a protein is labelled site-specifically, then this means the bio-orthogonal species will be incorporated into the protein in a single, specific location. Site-specific incorporation is more challenging than the global approach and will be the main focus of this discussion. In order to achieve this there are a variety of methods which can be employed including genetic code expansion,[14,15,41] native chemical ligation [42] and enzymatic reactions, which will all be discussed in this section.[43]

1.3.1   Amber suppression

Proteins are naturally composed of 20 amino acids, however, unnatural amino acids bearing bio-orthogonal functionalities can be incorporated into the natural proteins (Figure 1.4). This allows the protein to react with probes containing a complementary functional group for bio-orthogonal chemistry to take place which can provide more information about protein structure and function and aid visualisation.[14,44–47]
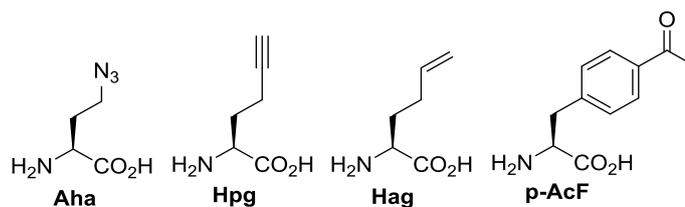
Figure 1.4: Commonly used unnatural amino acids for incorporation into proteins.[48]

Amber suppression allows the site-specific incorporation of unnatural amino acids by hijacking the cell's biosynthetic machinery using the amber stop codon (UAG) which normally directs the termination of protein synthesis. However, for this method, the amber stop codon is instead used to encode an unnatural amino acid loaded onto a complementary tRNA.[49] Within the gene of interest, a codon (XXX) is mutated using site-directed mutagenesis (SDM)[50,51] to the amber codon (TAG). This is translated to the mRNA UAG codon, which is recognised by a tRNA loaded with the unnatural amino acid, incorporating it into the growing peptide and ultimately at a specific location in the protein (Scheme 1.5).
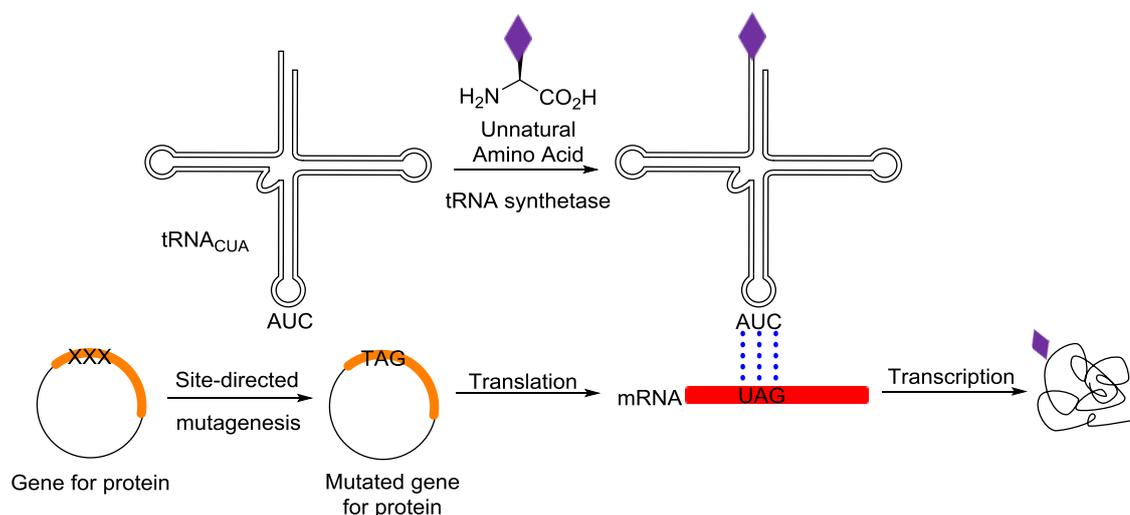
Scheme 1.5: Amber supression for site-specific unnatural amino acid incorporation. The protein is mutated using SDM to contain the amber stop codon (TAG), this is translated to UAG for mRNA which is recognised by the tRNA linked to the unnatural amino acid allowing it to be incorporated into the protein of interest. Adapted from Sletten et al.[52]

For the original *in vitro* approach, the amber stop codon was placed at the desired mutation site and all other amber stop codons removed from the sequence using SDM. A tRNA was synthesised encoding the complementary anticodon with the unnatural amino acid attached at the 3' end which was placed in *E. coli* with the gene encoding the mutated protein, and the endogenous transcription and translation mechanisms of the cell to produce the modified protein.[53,54] However, this approach was low yielding and the synthesis of the aminoacylated tRNA very complex and labour intensive.[55] Wang *et al.* improved this methodology by developing an *in vivo* system which uses an orthogonal aminoacyl-tRNA synthetase-tRNA pair.[56] This consists of a synthetase that is unable to aminoacylate endogenous tRNA, and a tRNA that is not a substrate for endogenous synthetases.[57] The orthogonal tRNA is, however, a substrate for the orthogonal tRNA synthetase, allowing the incorporation of unnatural amino acids which are substrates of the orthogonal aminoacyl-tRNA synthetase in response to the amber stop codon (UAG). To carry out the *in vivo* method, the genes corresponding to the orthogonal tRNA and tRNA synthetase are expressed in a heterologous host, along with the gene encoding the modified protein, with the amber mutation, while the cell culture is supplemented with the unnatural amino acid (Scheme 1.6).[52]
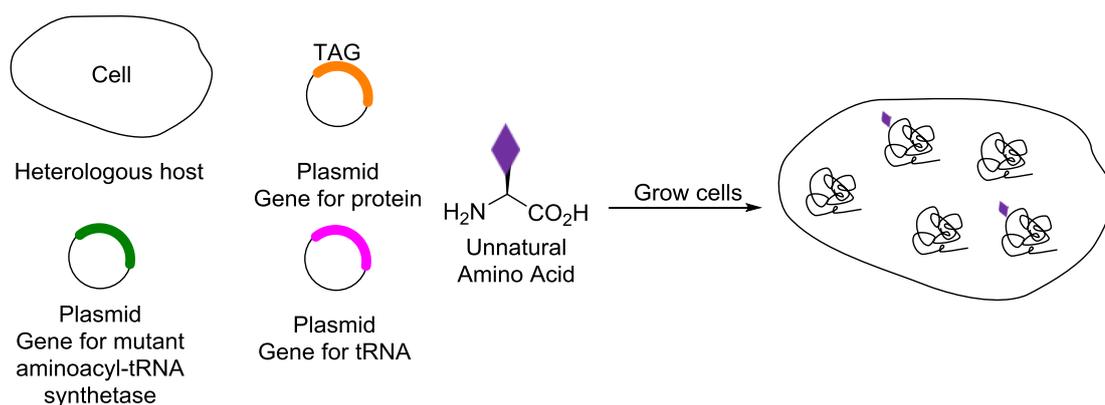
Scheme 1.6: General approach used to introduce an unnatural amino acid site-specifically into a protein using amber suppression. This involves the use of separate genes encoding the mutated protein and orthogonal tRNA and amino-acyl tRNA synthetase within a heterologous host while the culture is supplemented with the unnatural amino acid of choice. Adapted from Sletten et al.[52]

Because orthogonality in this case refers to the interactions between the synthetase tRNA pair and the host translational machinery, a synthetase or tRNA are defined as orthogonal with respect to a specific host. Some examples of orthogonal aminoacyl-tRNA synthetase/tRNA pairs that are now widely used for incorporating unnatural amino acids include:

- Methanococcus jannaschii Tyrosyl-tRNA synthetase (MjTyrRS)/tRNACUA pair which is orthogonal in *E. coli*, but not in eukaryotic cells.[56]

- E. coli Tyrosyl-tRNA synthetase (EcTyrRS)/tRNACUA pair and E. coli Leucyl tRNA synthetase (EcLeuRS)/tRNACUA pair which are both orthogonal in eukaryotic cells, but not *E. coli*.[58,59]

Orthogonal aminoacyl-tRNA synthetases and tRNA pairs have been developed by using directed evolution processes. Incorporation of a given unnatural amino acid requires the substrate specificity of the aminoacyl-tRNA synthetase to be edited. Evolution of the active sites of the synthetase/tRNA pairs can remove the binding sites of the natural amino acids and instead allow activity towards the desired unnatural amino acid.[60] Amber suppression is quite a complex technique, however using orthogonal aminoacyl-tRNA synthetase/tRNA pairs has resulted in the genetic code expansion of yeast,[61] mammalian cells,[62] *C. elegans* [63] and *D. melanogaster* [64] allowing a single site in these systems to be chemoselectively labelled *in vitro* and *in vivo*.

1.3.2 Native chemical ligation

In the 1990s, Kent and co-workers described a novel method for modifying proteins called native chemical ligation (NCL). They found that the chemo-selective reaction of peptides bearing a C-terminal thioester with an N-terminal cysteine substrate resulted in the formation of a thioester intermediate which rapidly rearranges to form the desired amide bond (Scheme 1.7).[65] This ligation reaction has been found to be highly specific, with even non-terminal cysteine residues remaining untouched as they are only able to form transient thioester intermediates due to the lack of the nucleophilic amino group for the final rearrangement.[66]
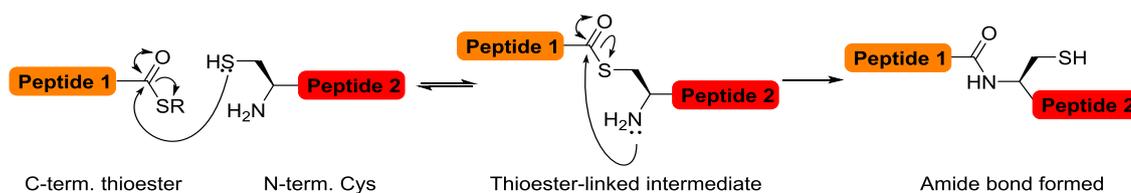
Scheme 1.7: The linkage of two peptide species using native chemical ligation between peptides bearing an N-terminal cysteine residue and C-terminal thioester. [65]

The main limitation of this work is the synthesis of the thioester peptides, originally this was done using Boc solid phase peptide synthesis (SPPS), but now strategies for Fmoc SPPS have been successfully explored using C-terminal N-acylurea functionalities resulting in thioester peptides for use in NCL.[67–69] In addition to this, the reaction conditions required for ligation are highly denaturing and this can lead to problems associated with mis-folding of the proteins. As a result of this, NCL is not generally used as a labelling method due to the requirement to unfold and refold the proteins. However, it has allowed the total synthesis of a number of synthetic proteins; for example, Bang *et al.*[29] were able to carry out a one-pot synthesis of crambin, a small protein consisting of 46 amino acids and Dürek *et al.*[29] synthesised human lysozyme, a 130 amino acid protein, by ligating 4 smaller polypeptide chains together.

1.3.3 Transpeptidases and sortase

Enzymatic approaches to site-specific protein labelling with transpeptidases provide an attractive, alternative approach to incorporating bio-orthogonal functionalities, since their highly selective nature can be readily exploited. Transpeptidase enzymes are generally responsible for linking two peptides together through the formation of a peptide bond. Sortase A has been used widely for protein-modification for a number of years and will be discussed in detail in this section. Additionally, a more recent discovery is that of Butelase, a peptide ligase which has been shown to enable specific and rapid transpeptidation reactions.

1.3.3.1   Sortase A

Sortase A is a transpeptidase enzyme responsible for sorting and reversibly attaching virulence factors to the cell walls of Gram-positive bacteria such as *Staphylococcus aureus*.[70] Sortase recognises an LPXTGX sequence (where X is any amino acid) within the protein of interest and the catalytic cysteine within the active site (Cys184)[71] attacks the amide bond between the threonine and glycine residues to form a thioacyl intermediate. This is subsequently attacked by an N-terminal oligoglycine species covalently attaching the protein to the cell wall (Figure 1.5).[72] The ability for sortase to mediate these ligations has led to it being used as a method for efficient C-terminal, N-terminal and internal modifications of proteins under mild conditions. [73–75]
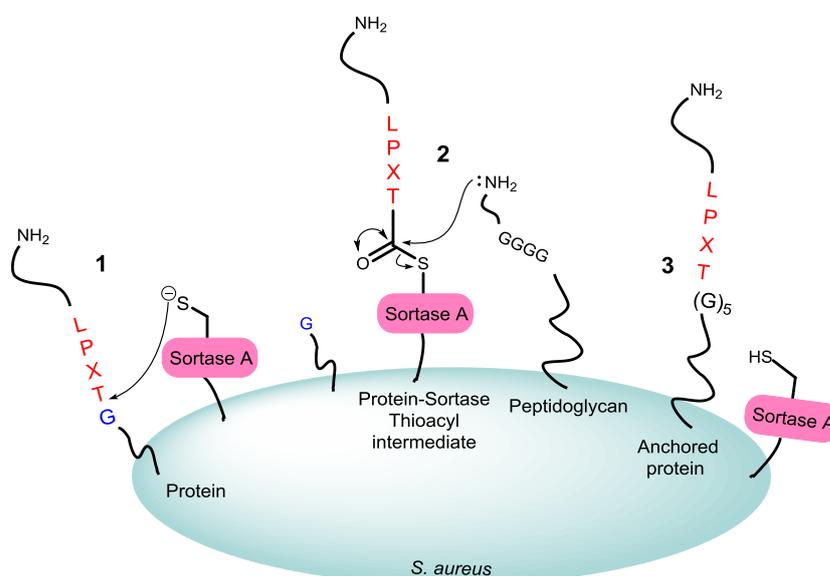
Figure 1.5: Sortase mechanism of reversibly attaching virulence factors to the cell walls of Gram-positive bacteria. Step 1: The catalytic cysteine residue found within the active site of sortase recognises the LPXTGX sequence within the protein of interest and attacks the amide bond between the threonine and glycine residues. Step 2: This forms a thioacyl intermediate which is subsequently attacked by an N-terminal glycine on a peptidoglycan anchored to the cell wall. Step 3: This results in the formation of a covalent bond between the peptidoglycan and the protein of interest, anchoring it to the cell wall. Adapted from Guimares et al.[76]

Initial work using sortase focussed on the use of peptides containing the LPXTGX recognition sequence as the labelling reagents. However, these reactions are reversible due to the formation

of a primary amine as the side product (Scheme 1.8). As a result of this, 20-50 equivalents of the labelling peptide, along with at least 1-2 equivalents of sortase were required in order to drive the equilibrium towards product formation, limiting the applicability of this as a labelling method.[77] A number of methods have been developed in recent years to find ways of making the reaction irreversible or removing the side product from the reaction mixture as soon as it is formed and also trying to improve the efficiency of the enzyme.[78–80]
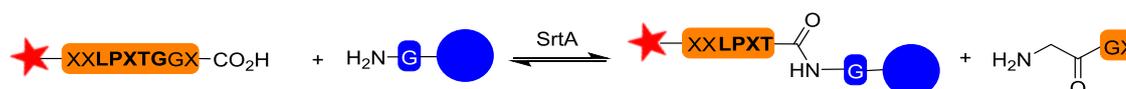
Scheme 1.8: Sortase ligation between a labelling peptide containing the LPXTGX recognition sequence and a protein containing an N-terminal glycine residue.[81]

1.3.3.1.1   The transpeptidation mechanism

Sortase A is thought to work *via* a "reverse protonation" mechanism (Scheme 1.9) with three key active site residues: Cys(184) acting as the nucleophile, His(120) as a general acid and Arg(197) as a transition state stabilising residue. It is understood that less than 0.1 % of the enzyme is in an active form under physiological conditions at any one time. That is for the His(120) and Cys(184) residues to be in a thiolate imidazolium pair as it is believed to have a higher affinity for the substrate than the thiol imidazole form. The deprotonated cysteine is then able to attack the amide bond between the threonine and glycine in the LPXTG motif (Step 1) forming an oxyanion intermediate which is stabilised by hydrogen bonding with Arg(197) (Step 2). Following this, the C-terminal glycine residue is lost through protonation by the histidine imidazolium cation and the thioester intermediate is formed (Step 3). The oligoglycine species then attacks the thioester intermediate leading to the formation of the final transpeptidation product (Steps 4-6).[82,83]
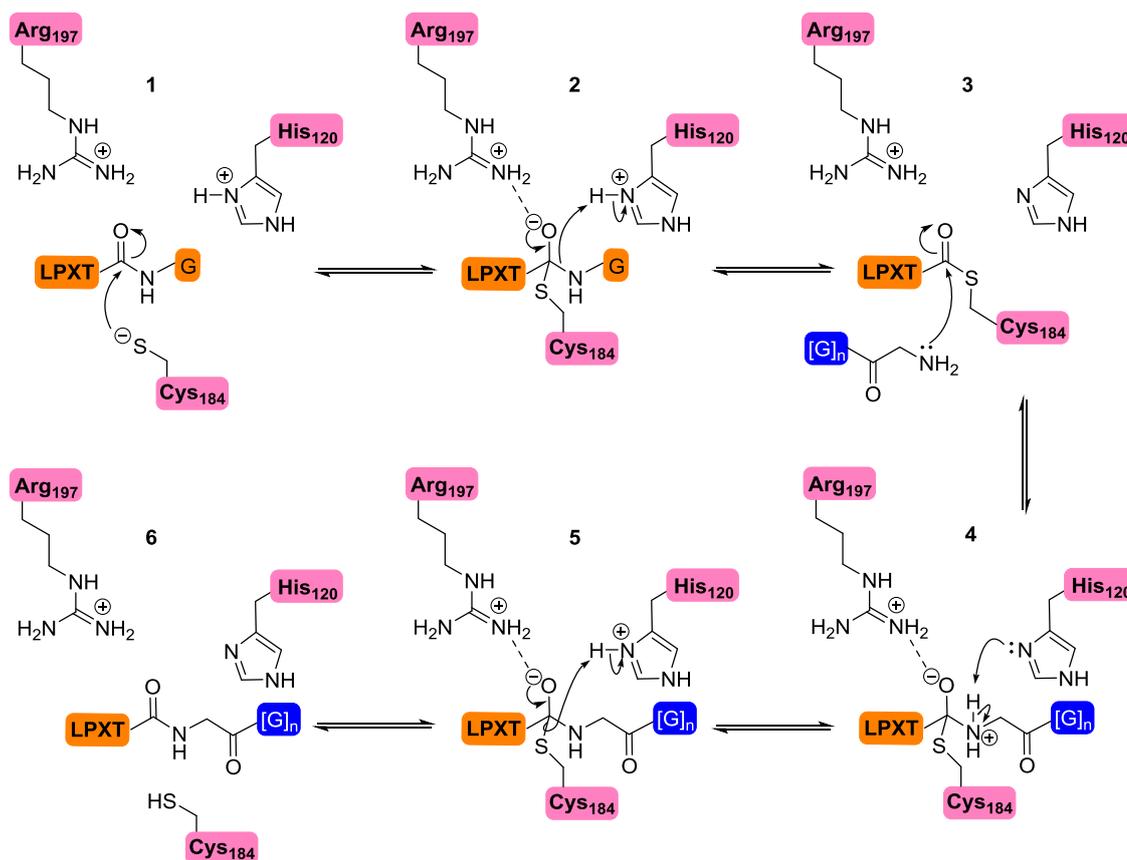
Scheme 1.9: Mechanism for the sortase-mediated transpeptidation reaction.[82,83]

1.3.3.1.2   *Sortase variants*

A lot of research has been done to explore the enzyme itself, altering the catalytic activity through the introduction of selected point mutations. There are two sortase variants in particular which have enhanced activity, a pentamutant (Srt 5M) and a heptamutant (Srt 7M). In 2011, the pentamutant was developed by Chen *et al.*, using directed evolution techniques they found that by introducing five point mutations near the LPXTGX binding site (P94R, D160N, D165A, K190E and K196T) the k_{cat}/K_M increased by ~120-fold. This allowed reactions using a lower quantity of enzyme and lower temperature (≥ 4 °C instead of 37 °C for WT SrtA).[84] Further to this, Hirakawa and co-workers found that by mutating two Glutamic acid residues (E105K and E108A/Q), the enzyme no longer required Ca^{2+} as a co-factor making it compatible with a wider range of buffers including phosphate-buffered saline (PBS).[85,86] As a result of these separate improvements, the Ploegh group combined all seven mutations mentioned above to give Srt 7M,

a rapid and calcium independent form of sortase A.[80] Both the pentamutant and heptamutant enzymes have been used for a variety of labelling applications including on cellular surfaces[87] and intracellular ligations.[88]

In addition to improving the efficiency of the enzyme, work has also been done to investigate changes to the specificity. As mentioned previously the WT sortase A recognises LPXTGX, this stringent substrate specificity limits the use of sortase so any changes made to this could potentially broaden applications of the enzyme. Examples of different substrate recognition sequences are LAXTGX (eSrtA(2A-9)), LPXSGX (eSrtA(4S-9))[89] and FPXTGX[90,91] which were developed using yeast display or phage display screenings. By using sortase variants which recognise different sequences it could be possible to use them for orthogonal labelling reactions on the same system.

1.3.3.1.3   Irreversible sortase labelling

The reversible nature of the sortase reaction can be problematic when using labelling reagents that are difficult to synthesize or expensive as a large excess of these are required. The simplest approach to achieving this is to remove the polyglycine by-product from the reaction as it is formed. This can be achieved by carrying out the reactions in a centrifugal filter or under dialysis conditions and has been seen to enhance the efficiency of the reaction. However, this is only possible when the labelling reagent is large.[78,92,93] A different approach was developed by Policarpo *et al.* for C-terminal labelling whereby sortase 7M was immobilised on Ni-NTA resin to which a protein loading solution of polyglycine and protein-LPXTGG-His$_6$ is applied. The unreacted protein-LPXTGG-His$_6$ binds to the nickel resin, transpeptidation occurs, and the products are released and flowed out of the column. Following this, a peptide "push" solution is applied to the column which contains only the polyglycine peptide, which helps to drive the reaction towards product formation (Figure 1.6a).[79] For N-terminal labelling, Witte and co-workers developed a flow system using sepharose functionalised with sortase 7M packed within a flow column. Here, the column is pre-incubated with the labelling peptide, LPXTGGX to allow the formation of the acyl intermediate on the resin before the $(G)_n$-protein is flowed

through (Figure 1.6b). The benefits of this technique include the selective removal of reaction components and the ability to limit the contact time with the sortase, decreasing the chance of side reactions taking place.[80] In both of these cases the sortase 7M mutant is critical, previous work in the Webb lab found the WT enzyme to be unsuccessful in this type of system as the reaction happens so much faster for Srt 7M. Simply flowing slowly over the sortase 7M-immobilised beads allowed labelling to occur.
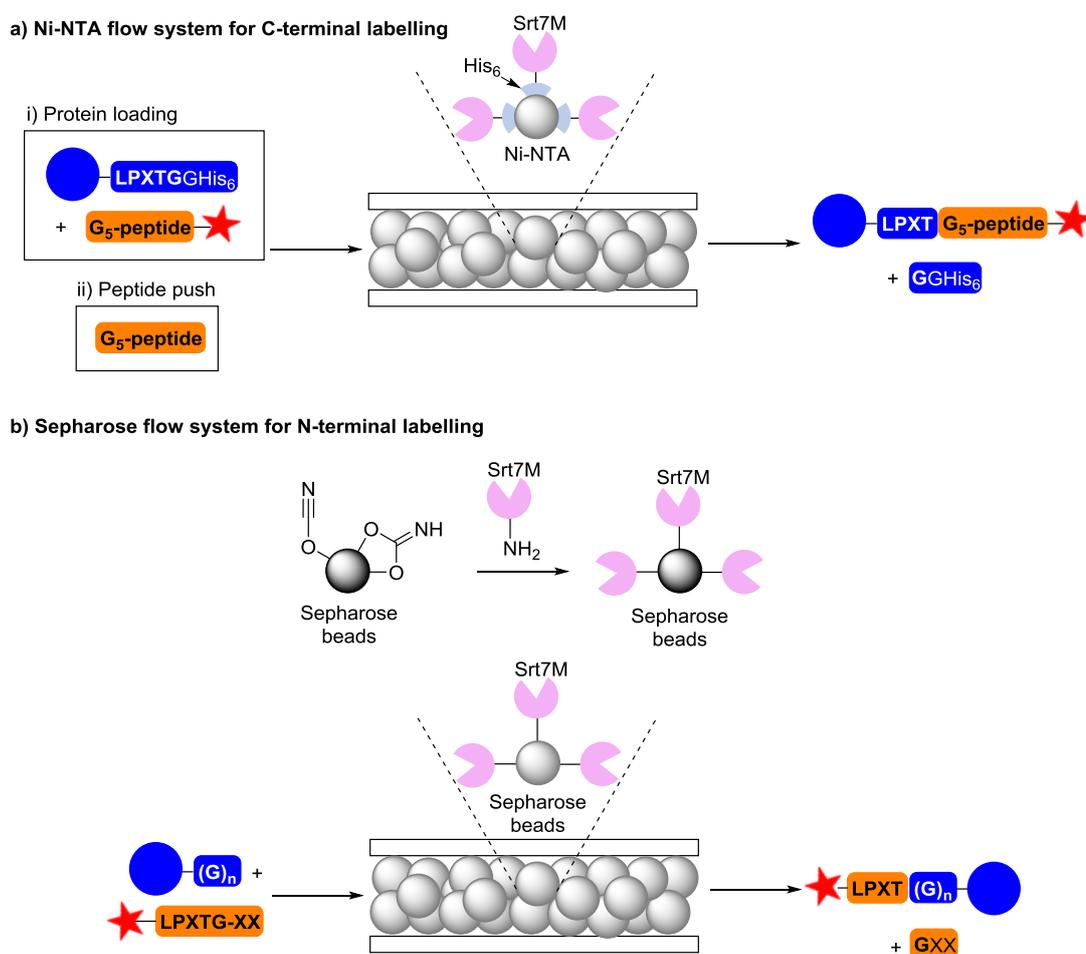
Figure 1.6: By immobilising sortase 7M on to a solid support the ligation reaction is able to take place and the product and by-product flow straight out meaning they are no-longer able to take part in potential reversible reactions. a) Srt 7M on Ni-NTA beads using His$_6$-tagged protein for C-terminal ligation. b) Srt 7M on sepharose beads for N-terminal ligation. Adapted from Policarpo et al.[79] and Witte et al.[80]

Another approach to driving the ligation towards the desired product is to selectively deactivate the ligation product or the by-products to prevent the reverse reaction. The ligation product has been deactivated through the formation of a β-hairpin at the LPXTGX recognition site, making it unreactive (Scheme 1.10a). The secondary structure is formed through the introduction of several additional tryptophan and tyrosine residues on either side of the LPXTGX and it has been shown to inhibit further recognition by sortase, allowing ligation yields to be improved.[94] As for the by-product, one interesting approach has been described by Row and co-workers using nickel chelation. By extending the LPXTGX sequence to LPXTGGH, the by-product would contain the GGH-motif, which is able to bind with high affinity to Ni^{2+} ions. The coordination of the nitrogen lone pair decreases the nucleophilicity of the by-product, minimising the reverse reaction (Scheme 1.10b). These were carried out in the presence of 1-2 equivalents of $NiSO_4$ achieving up to 85% product formation.[95] Since both of these approaches only involve the use of natural amino acids, this can be applied to both C- and N-terminal labelling and have been seen to improve yields for both types of reaction.
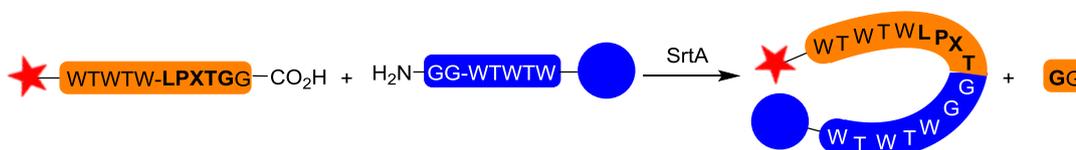
Scheme 1.10: Two methods of product and by-product deactivation to improve sortase labelling efficiency. a) Formation of a hairpin makes the LPXTG site no longer accessible to sortase.[94] b) Nickel chelation decreases the nucleophilicity of the by-product reducing the reverse reaction.[95]

Taking a different approach for N-terminal protein labelling, Williamson *et al.* found that by replacing the amide bond between the threonine and glycine residues within the LPXTGX

sequence with an ester bond, forming a depsipeptide substrate, the reaction could become irreversible. The reason for this is that instead of forming a primary amine as the by-product, a much less reactive hydroxyacetyl was formed instead which was unable to take part in the reverse reaction (Scheme 1.11). These reactions allowed a range of proteins to be labelled with high ligation yields using around 3 equivalents of the depsipeptide with 20 molar% sortase.[81,96]
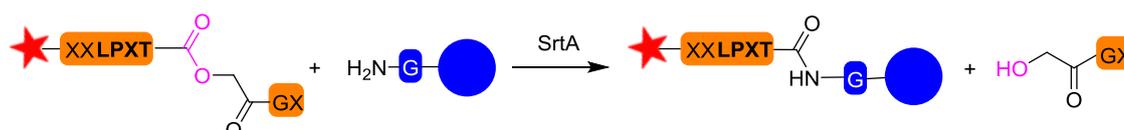
Scheme 1.11: The introduction of an ester bond between the threonine and glycine residues in the LPXTG recognition sequence allows an hydroxyacetyl to be formed as a by-product instead of an amine, rendering the reaction irreversible.[81]

Another approach adopted by Liu *et al.* used the formation of a diketopiperazine as a driving force for irreversible sortase-mediated labelling. They demonstrated the synthesis of two effective substrates for the enzyme, LPETGG-isoacyl-Serine and LPETGG-isoacyl-Homoserine, prepared using standard solid-phase peptide synthesis. Following the successful ligation, the side-product proceeded to undergo an intermolecular rearrangement to form the diketopiperizine preventing it from taking part in the reverse reaction (Scheme 1.12).[97]
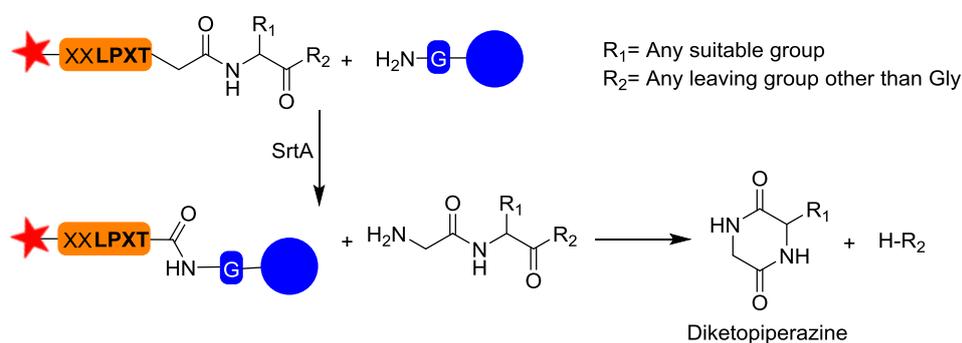
Scheme 1.12: Irreversible sortase-mediated labelling driven by the formation of a diketopiperazine side-product.[97]

1.3.3.2   The kinetic mechanism of sortase A

Sortase A is known to catalyse the transpeptidation reactions through a particularly unusual mechanism, not fully fitting into the initial hypothesis of a ping-pong bi-bi reaction proposed by Huang *et al*.[98] A ping-pong bi-bi mechanism is one where there are two substrates and two products ("bi-bi") and one of the products is released before the second substrate binds ("ping-pong").[99] For sortase this is the case in the manner that the enzyme attacks the amide bond between the threonine and the glycine to form an acyl-enzyme complex (this would be substrate #1). Subsequently the C-terminus of the substrate is released (product #1) before the nucleophilic oligoglycine species (substrate #2) attacks the thioacyl intermediate and the covalent linkage is made to anchor the protein to the cell wall (product #2).[99,100] However, although theoretically the sortase transpeptidation should fit this model, Frankel *et al*. found the parameters to be slightly out, proposing a hydrolytic shunt within the ping-pong mechanism. This hydrolytic shunt means that water is competing with substrate #2 in the second part of the mechanism. Kinetic studies found the catalytic rate constant describing the turnover rate of an enzyme-substrate complex to product and enzyme (k_{cat}) to be much slower for hydrolysis than that of transpeptidation, 0.086 s^{-1} and 0.28 s^{-1} respectively (Figure 1.7). In addition to this it has also been observed that although they both require the formation of the acyl-intermediate, they have different rate-limiting steps. For the transpeptidation this is the initial acylation, whereas for hydrolysis the rate-limiting step is the hydrolysis of the acyl-enzyme intermediate (deacylation), after the first product has been released.[98,101]
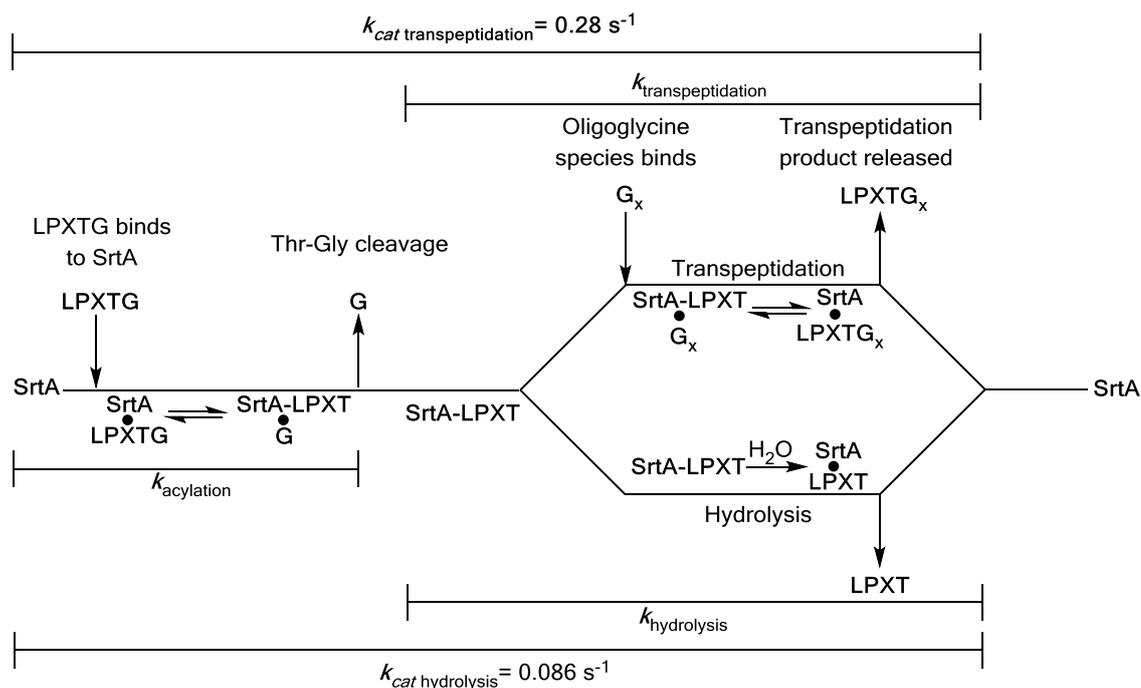
Figure 1.7: Ping-pong bi-bi hydrolytic shunt mechanism for SrtA including the catalytic rate constants calculated by Frankel et al. for the transpeptidation and hydrolysis reactions. SrtA catalyses the cleavage of the T-G bond within the LPXTG recognition sequence to form the acyl intermediate. This can then undergo either transpeptidation or hydrolysis. The acylation step is the rate-limiting step for transpeptidation, whereas the deacylation step is the rate-limiting step for hydrolysis.[101]

Since the hydrolysis step is slower it is likely to take place if the concentration of substrate #2 is low.[99] In addition to this, if the final product (#2) is in excess, the enzyme could attack the LPXTG sequence present in the product leading to the reformation of the acyl-enzyme complex (substrate #1) and causing the product to undergo hydrolysis. Experimentally, it is generally advised to use an excess of the nucleophilic labelling species to prevent the substrate #2 from being in low enough concentrations to allow water to act as an alternative nucleophile. However, all of the sortase-mediated labelling approaches discussed previously lead to the formation of a product containing the LPXTG motif which can be subsequently hydrolysed by the sortase when in excess. The methods which potentially decrease the chance of this from happening are the ones in which sortase is immobilised on resin and used in a flow system developed by Witte *et al.* and Policarpo *et al.*[79,80] By using this approach the length of time the sortase and product are in contact

is reduced as much as possible while also allowing the reaction to take place. Ultimately, by studying the kinetic mechanism for this enzyme, the importance of monitoring the reactions is highlighted. The ability to find a "sweet-spot" where the protein of interest is labelled to a maximum level before the sortase is able to cause the formation of any hydrolysis product would provide a more robust method for protein modification.

1.3.3.3   Butelase

During the course of this project, Nguyen *et al.* discovered butelase, a peptide ligase purified from the pods of a medicinal plant found in Southeast Asia called *Clitoria ternatea*, known locally as "bunga telang", from which the name "butelase" was derived.[102] This peptide ligase is involved in the biosynthesis of cyclotides, which is the largest family of plant cyclic peptides, and has been developed over recent years for use as a method of site-specific protein labelling.[103] Butelase is a very efficient catalyst of these reactions with as little as 0.0001 molar equivalents required for protein labelling, which is an order of magnitude lower than sortase A.[104]

The way in which these ligations proceed is first through the recognition of a tripeptide motif by butelase at the C-terminus, Asx-His-Val, where Asx is either asparagine or aspartic acid. It then mediates cyclisation by cleaving the His-Val and attaching the Asx residue to the N-terminus of the protein where the first residue (X_1) can be any amino acid other than proline, but the second residue (X_2) tends to require a bulky, hydrophobic amino acid, such as isoleucine, leucine, valine or cysteine (Scheme 1.13a).[103] In addition to this intramolecular cyclisation, it has also been used as a very effective method of carrying out intermolecular ligations with similar amino acid requirements. However, this time the Asx-His-Val recognition motif is inserted onto the C-terminus of "peptide 1" and the required X_1X_2 residues are on the N-terminus of "peptide 2", ensuring they are not present in "peptide 1" (Scheme 1.13b).[103]
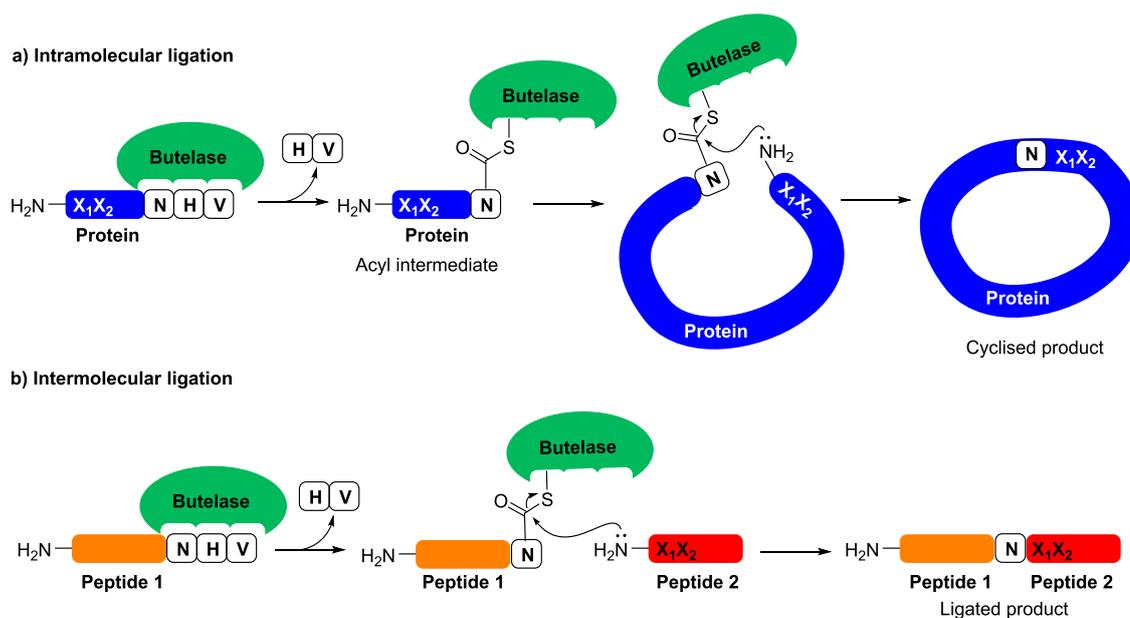
Scheme 1.13: a) The intramolecular ligation mechanism to form a cyclic protein. b) Intermolecular ligation mechanism using butelase. N is either asparagine or aspartic acid, X₁ is any amino acid other than proline and X₂ is a bulky, hydrophobic amino acid. Adapted from Nguyen et al.[103]

Due to the formation of the cleaved peptide, His-Val, the intermolecular reactions are reversible, meaning a 5-fold excess of substrate is required to drive the reaction towards product formation. In an attempt to make the reaction irreversible, Nguyen *et al.* replaced the nitrogen of the amide bond between Asx and His with a sulphur, forming thiodepsipeptide substrates. These species have a thiol as a leaving group and an unacceptable nucleophile which is not recognised by butelase, making the reaction irreversible (Scheme 1.14).[104]
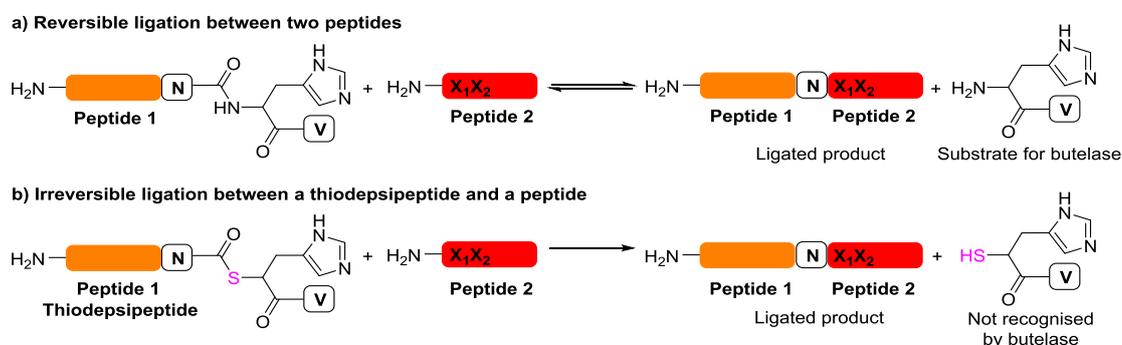
Scheme 1.14: a) The reversible ligation of two peptides using butelase. b) The introduction of a thiodepsipeptide between the asparagine/aspartic acid and histidine allows irreversible ligations to be achieved using butelase.[104]

Although this enzyme is extremely fast and efficient and irreversible methods of labelling have been developed using the thiodepsipeptides, the isolation of butelase itself is not trivial. In order to obtain ~5 mg of enzyme, 1 kg of fresh pods of the *C. ternatea* plant are homogenised before being taken through four chromatographic steps including flash anion exchange, preparative anion exchange HPLC, size-exclusion and analytical anion exchange HPLC.[103] In order to try and make this more simple and applicable to labs across the world, a homologue of butelase was recombinantly expressed in *E. coli* called OaAEP1b by Yang *et al.* However, low expression yields and difficulties associated with finding the enzyme containing fractions after chromatography requiring the use of Western-blotting also makes this quite a laborious method for isolating the enzyme. In addition to this, the enzyme then did not work as expected, with ligations taking place approximately 500 times slower than butelase, requiring up to 22 hours to carry out a ligation which would take butelase 5 minutes.[105]

Butelase is an extremely useful tool for protein labelling. However, it does still require further development to ensure the enzyme is easier to isolate in labs across the world. In addition to this, it seems the enzyme is generally used at 42 °C by Nguyen *et al*., which could potentially limit the number of proteins able to be labelled using this approach.[103]

1.4   Project outline

The aim of this work is to investigate the effect of combining the use of the enhanced sortase variant, sortase 7M, and depsipeptide substrates to determine whether they can be used as a more efficient irreversible labelling method. Initial work was to determine the appropriate conditions for the reaction before applying them to a number of proteins e.g. maltose binding protein (MBP), PanD and Cholera toxin B-subunit (CTB). To gain a greater understanding of these reactions, a fluorescence resonance energy transfer (FRET) monitoring technique was developed to look at how the reactions progressed using sortase 7M. This would also enable the comparison of sortase 7M-mediated labelling using depsipeptides and peptides. Using this as a labelling tool, a number of different substrates were covalently linked to the proteins for use in further biophysical

and structural studies including surface plasmon resonance (SPR), atomic force microscopy (AFM) and small-angle X-ray scattering (SAXS). This involved labelling using a biotin-depsipeptide for immobilisation of PanD on a streptavidin chip. This enabled SPR experiments to be carried out comparing the interactions of PanZ and PanD in the presence of two different coenzyme A species. In addition to this, a bicyclononyne-containing depsipeptide was synthesised to facilitate protein labelling with different azide-functionalised DNA species through strain-promoted azide-alkyne cycloaddition reactions for AFM and SAXS studies.

2 Use of sortase (7M) alongside depsipeptide substrates

This chapter investigates combining the use of depsipeptide substrates and the efficient, calcium independent heptamutant sortase 7M to see if the N-terminal labelling reactions can be improved. A variety of experiments have revealed the optimal concentrations of each component and the temperature at which the reactions prove to be most efficient. A fluorescence resonance energy transfer technique was also developed to monitor the sortase 7M labelling in order to gain a greater understanding of how the reactions progress.

2.1 Reagent preparation

Previous studies have used a combination of SDS-PAGE, HRMS and HPLC to analyse the progression of the WT sortase labelling reactions. For the experiments described in this chapter, SDS-PAGE and HRMS were used, but due to the faster labelling associated with Srt 7M, a more continuous method of monitoring than HPLC was required. This was chosen to be FRET, as the reagents could be mixed and placed in a cuvette for monitoring over a long period of time. This, however, required a fluorescent depsipeptide to be synthesised suitable for use within the FRET system, which will be discussed in more detail in Section 2.3. Fluorescein isothiocyanate (FITC) was chosen as the fluorescent species which would be coupled to the labelling depsipeptide. This would also allow the SDS-PAGE gels to be analysed using UV transillumination in addition to Coomassie blue staining. First, the depsipeptide building block had to be synthesised and incorporated into the longer chain peptide containing the sortase recognition sequence. Then, the FITC could be coupled onto the N-terminus of the peptide.

2.1.1 Synthesis of depsipeptide building block

The first step in this process was to synthesise the depsipeptide building block itself. As the recognition sequence for Srt 7M is LPXTG and the bond attacked by the enzyme is the amide bond between the threonine and glycine residues, the building block required is threonine and glycine connected by an ester bond (ToG). This was done *via* the alkylation of Fmoc-threonine (**2.01**) with benzyl bromide (**2.02**) in the presence of tetrabutylammonium iodide (TBAI) to give benzyl protected ToG (**2.03**) as a colourless solid in good yield (89%).[96] The next step involved the deprotection of the alcohol group using hydrogenation. However, although the first step had proceeded as literature suggested, this second part proved to be problematic. The literature conditions for this were 3:1 THF/H_2O but there seemed to be some polymerised impurity forming in the product instead of the desired deprotected compound. After analysing different batches of palladium catalyst and THF the source of this impurity appeared to be the THF. By changing the solvent system to CH_2Cl_2/MeOH (9:1) the hydrogenation product (**2.04**) was successfully isolated within an hour in high yields of up to 95%.

Scheme 2.1: Synthetic route to the depsipeptide building block required for SPPS.

29

2.1.2 Solid phase peptide synthesis of fluorescein functionalised depsipeptide

The ToG building block was then incorporated into the longer peptide chain using standard solid phase peptide synthesis (SPPS). The overall peptide sequence chosen was H$_2$N-GABA-LPE**ToG**G-CO$_2$H (**2.05**). The resin used for building this peptide was Gly-2-chloro-trityl which is cleaved using TFA to leave an additional glycine residue at the C-terminus of the peptide. Previous work had found the incorporation of the depsipeptide species to occasionally be problematic, with a side product forming without the building block present. To try and avoid this a double coupling of the ToG was carried out, followed by an acetyl anhydride capping step before the next amino acid residue was introduced. This prevents any uncoupled resin from being able to react with any of the other amino acids introduced later in the synthesis. Following use of these approaches only the depsipeptide species could be observed when analysed by mass spectrometry. A short linker group (GABA) was then incorporated before the FITC was coupled to yield the desired peptide without the need for further purification (**2.05**, 53%).

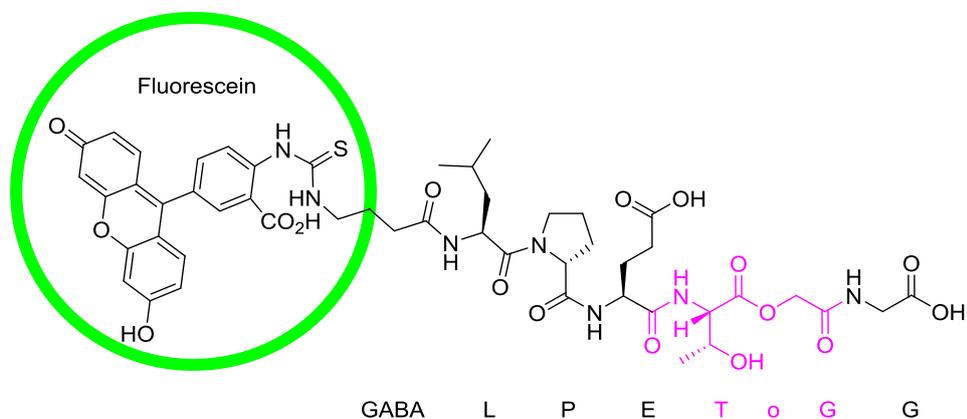*Figure 2.1: FITC-depsipeptide required for visualisation during labelling experiments, **2.05**.*

2.1.3 Preparation of sortase 7M

The sortase 7M plasmid was obtained from Hidde Ploegh and provided by Addgene, it was transformed into commercial *E. coli* BL21 gold (DE3) competent cells and expressed in auto-induction media (AIM) at 30 °C for 5 hours before being left at 25 °C overnight to maximise

cell density while ensuring overexpression of the desired protein. Following cell lysis, the sortase 7M was purified using sequential Ni-NTA affinity chromatography and size-exclusion chromatography using a Superdex® S75 column isocratically eluted with Tris-buffered saline at pH 7.4 (Figure 2.2).
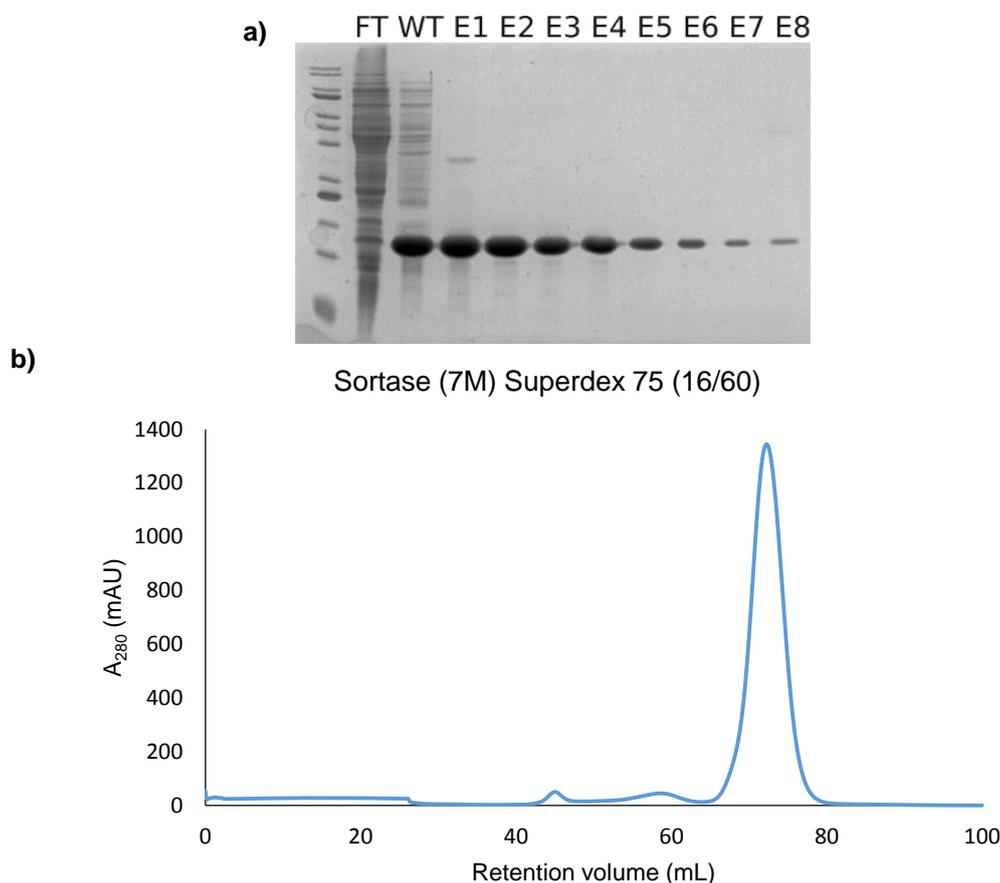
Figure 2.2: Purification of sortase 7M for use in labelling reactions a) SDS-PAGE gel of Ni-NTA affinity chromatography column flow through (FT), wash (WT) and elution fractions (E1-8). b) SEC trace for Srt 7M using Superdex® S75.

2.2 Protein labelling using Srt 7M alongside FITC-depsipeptide

2.2.1 Maltose binding protein

The initial model protein for this work was *E. coli* maltose binding protein (MBP, malE) which is a very stable, globular and monomeric protein often used as a solubility tag for proteins which are difficult to purify and aggregate easily. MBP is approximately 40 kDa. However, for this work

the protein needed to be engineered to contain an N-terminal glycine for the sortase labelling. A short N-terminal linker was introduced, MGVGK, using site-directed mutagenesis (SDM). Following overexpression, methionine aminopeptidase removes the methionine, presenting the glycine residue for labelling. The ability for MBP to bind to amylose resin meant a His$_6$-tag was not required for purification and the Srt 7M could be removed simply using a small amount of Ni-NTA resin. Following the successful mutagenesis, the plasmid was transformed into *E. coli* BL21 Gold cells and expressed in lysogeny broth (LB) at 37 °C before induction of protein expression using IPTG and grown overnight. Once the cells had been lysed, the protein was purified by binding to an amylose affinity column, and eluted using 20% (*w/v*) glycerol in phosphate-buffered saline before being purified using a Superdex® S200 size-exclusion column isocratically eluted with phosphate-buffered saline (pH 7.4) (Figure 2.3).
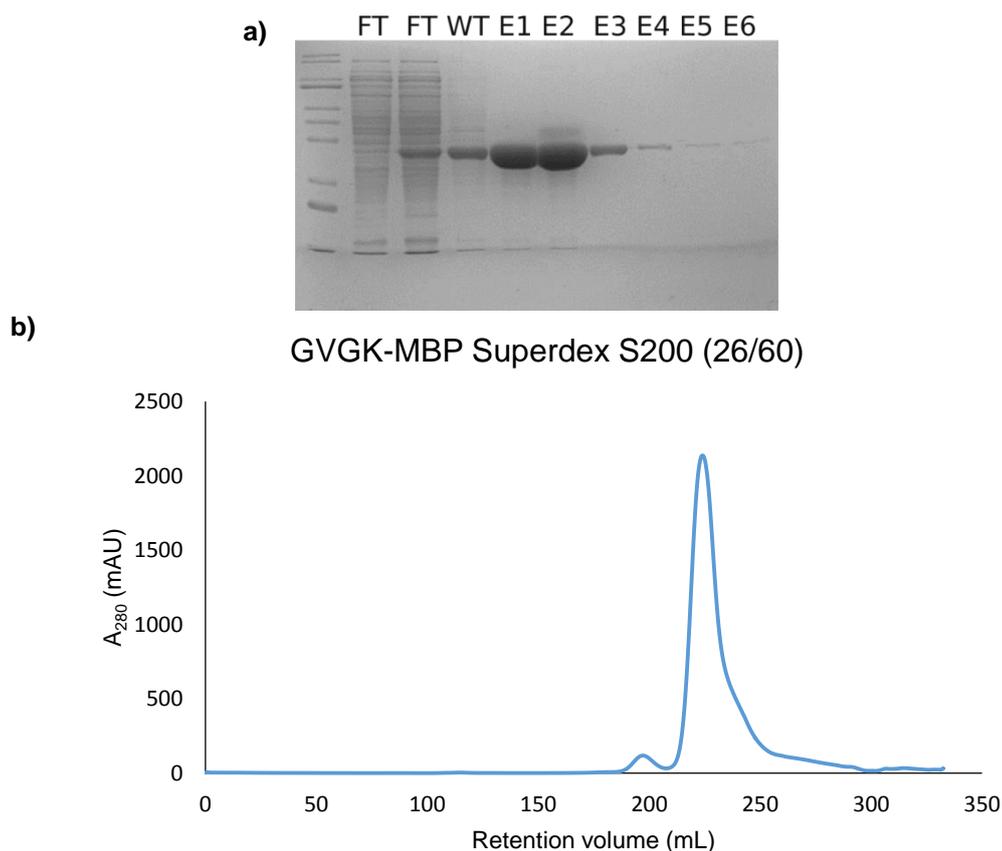

Figure 2.3: Purification of GVGK-MBP for use in labelling reactions a) SDS-PAGE gel of amylose column flow through (FT), wash (WT) and elution fractions (E1-6). b) SEC trace for GVGK-MBP using Superdex® S200.

Since depsipeptides and Srt 7M had not been used together before, the reaction conditions used for depsipeptide labelling using WT SrtA were examined. This was generally 20 molar% enzyme and 3 molar equivalents of the labelling substrate at 37 °C. Since it was known that this enzyme was much more active than the wild-type enzyme, the initial reaction conditions chosen were 5 mol% Srt 7M, 3 eq. depsipeptide on the bench at room temperature. Ensuring the samples were run on an appropriate gel meant that a band shift could be observed from the unlabelled to labelled species when stained using Coomassie Blue. Imaging using UV transillumination also provided confirmation that the protein had been labelled with the fluorescent depsipeptide.

The first reaction, 50 µM MBP, 150 µM FITC-depsipeptide and 2.5 µM sortase 7M, showed a band shift, suggesting that the reaction was complete around 1.5-2 hours using these conditions. Using UV analysis it was possible to confirm that the upper band was fluorescent. However, the intensity of this band seemed to decrease towards the end of the time course. Both methods of analysis implied that the labelled product seemed to form within the first two hours and then the label began to be removed, presumably by hydrolysis, as discussed in Chapter 1, with the lower band reappearing in the 3 and 4 hour time points (Figure 2.4).
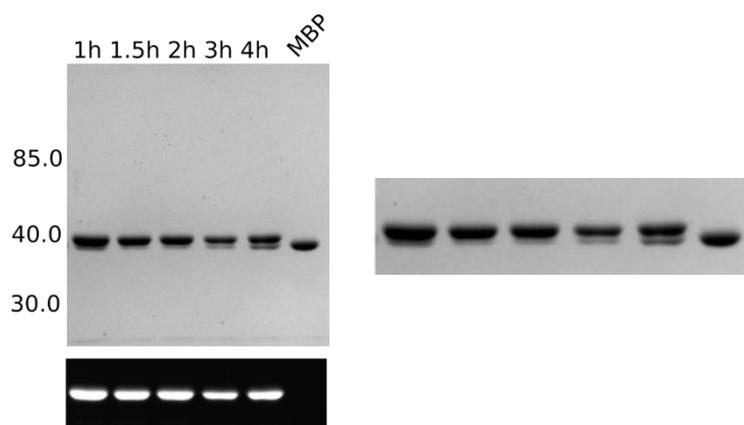
Figure 2.4: MBP labelling over 4 hours using 3 eq. FITC-depsipeptide and 5 mol% Srt 7M with zoomed in image to highlight the lower band reappearing after 3 and 4 hours.

Having observed rapid labelling at room temperature but also hydrolysis, a more strict method of temperature control was employed. This was achieved using the polymerase chain reaction (PCR) thermocycler held at 25 °C and 37 °C while comparing with a sample left on the bench with a room temperature of 21 °C. All three experiments used 50 µM MBP with 150 µM FITC-depsipeptide and 2.5 µM sortase 7M and time points were taken after 1, 2 and 3 hours before analysis by SDS-PAGE. The results showed the sample at room temperature labelled most slowly, with 25 °C achieving higher levels of labelling after approximately 2 hours with only a slight increase in the intensity of the lower band after a further 2 hours. However, the experiment at 37 °C appeared to label to a maximum within the first hour but then the following time points show increasing levels of the lower band presumably due to hydrolysis of the label (Figure 2.5).
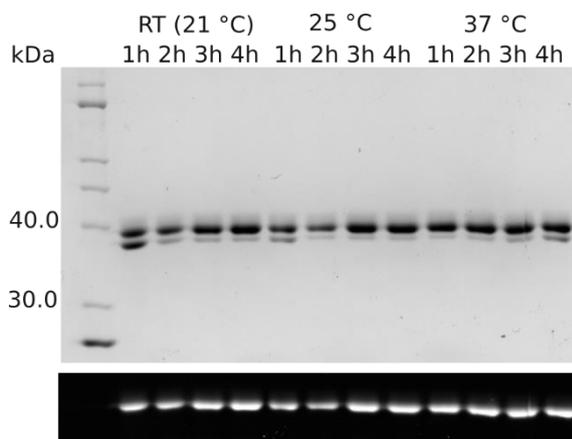
Figure 2.5: Temperature controlled reactions to compare the levels of labelling and hydrolysis products observed when labelling 50 µM MBP with 150 µM FITC-depsipeptide and 2.5 µM sortase 7M at room temperature (21 °C), 25 °C and 37 °C.

These results suggested that 25 °C provides the best balance between the rate of labelling product formation and rate of hydrolysis. To ensure this is the optimal temperature the experiment was repeated using a temperature gradient from 20.7 to 25.1 °C. The reaction samples were placed in wells corresponding to 20.7, 21.8, 22.7, 23.8 and 25.1 °C (Figure 2.6). These results indicated that between 22.7 and 25.1 °C labelling was achieved within 2 hours. However, the samples

maintained at 20.7 °C and 21.8 °C appeared to take ~1 hour longer to achieve similar levels of labelling. This confirmed that 25 °C is the optimal temperature to use for further labelling.
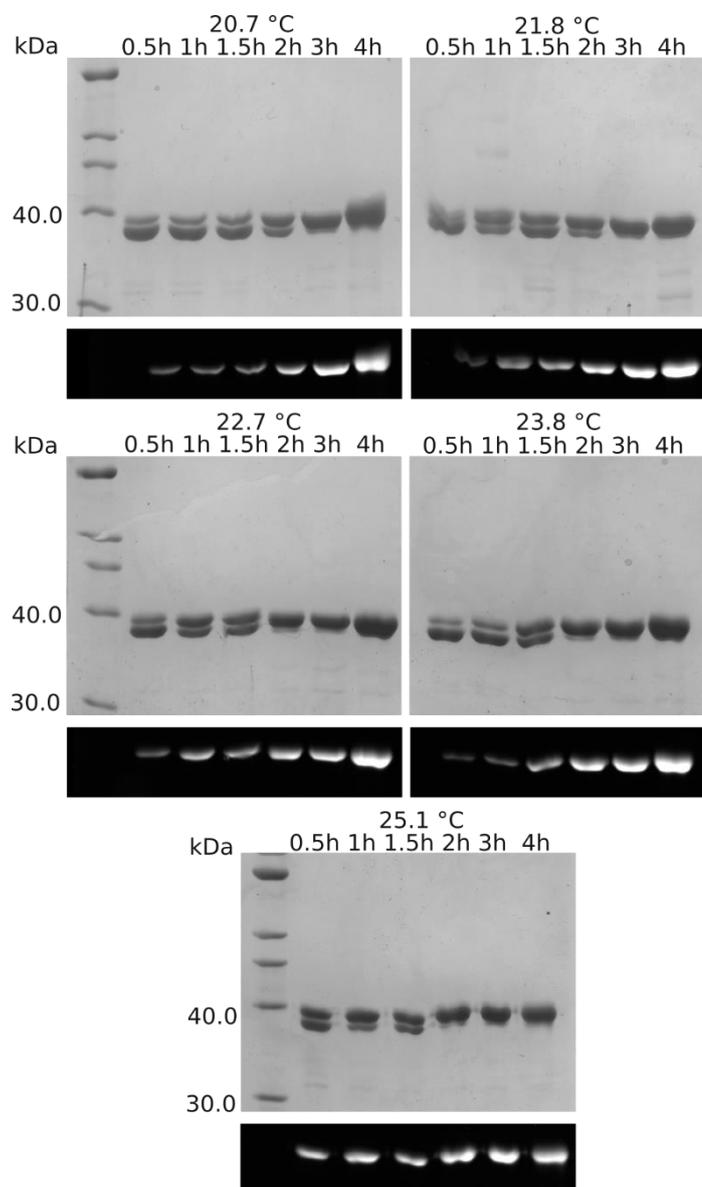
Figure 2.6: Labelling of 50 μM MBP using 150 μM FITC-depsipeptide and 2.5 μM sortase 7M over four hours using a temperature gradient between 20.7 and 25.1 °C.

The next variable to be investigated was the concentration of enzyme, using 50 µM MBP, 150 µM FITC-depsipeptide and 5 µM sortase 7M held at 25 °C with time points taken over four hours. These results are similar to those obtained using 2.5 µM Srt 7M at 37 °C, the maximum level of labelling was achieved within the first hour followed by the gradual appearance of the hydrolysis

product over the course of the reaction (Figure 2.7a). Following this result, overnight reactions (~ 16 hours) were compared for 5 and 10 mol% sortase 7M in the fridge at 4 °C. Very similar results were obtained for both experiments, with potentially slightly more labelled product obtained for the higher sortase concentration. Although the labelling achieved at this much colder temperature is not quantitative, it demonstrates that this could be a useful way of attempting to label unstable and temperature-sensitive proteins (Figure 2.7b).
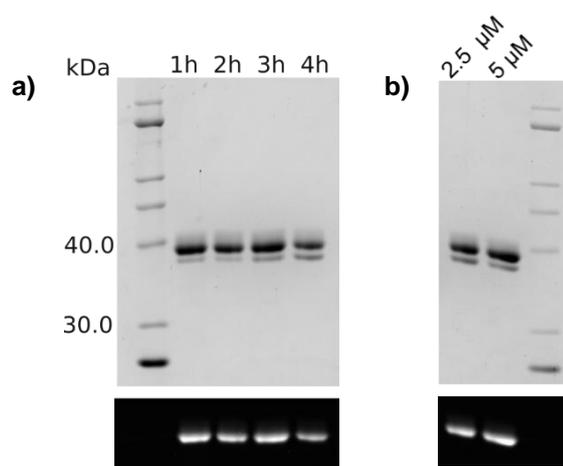

Figure 2.7: a) 50 μM MBP using 150 μM FITC-depsipeptide and 5 μM sortase 7M at 25 °C.
b) Experiments at 4 °C overnight for labelling 50 μM MBP using 150 μM FITC-depsipeptide and either 2.5 or 5 μM sortase 7M.

Labelling experiments using sortase have previously been shown to be strongly dependent on concentration. The lowest concentration of protein labelled using WT sortase is 50 μM which was the reason for using this concentration in the previous experiments.[96] However, the ability to label very low concentrations of protein could be a useful tool if samples are precious, so to check the labelling was still possible at decreased concentrations 10 and 25 μM MBP were labelled at a fixed ratio of MBP:Srt 7M to see if this influenced the relative intensities of the labelled and unlabelled bands on the SDS-PAGE gels. The results illustrated the successful labelling of both concentrations of protein with very similar ratios of labelled and unlabelled bands (Figure 2.8). This means Srt 7M successfully labelled much lower concentrations, using much less enzyme and at lower temperatures than the WT enzyme using depsipeptides.
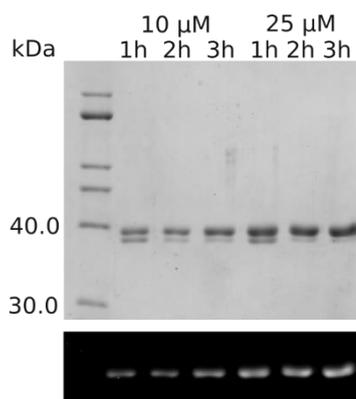
Figure 2.8: Varied concentrations of protein labelling reactions at 25°C. 10 µM MBP, 30 µM FITC-depsipeptide with 0.5 µM sortase 7M and 25 µM MBP, 75 µM FITC-depsipeptide with 1.25 µM sortase 7M.

In protein labelling experiments, the labelling substrate is often the most precious part of the reaction mixture having either been chemically synthesised or purchased. Experiments in this chapter so far have used 3 equivalents of the depsipeptide substrate (150 µM when labelling 50 µM protein). This was the general concentration used for labelling with WT SrtA, so to investigate whether this could be reduced when using the more active enzyme 50 µM and 100 µM depsipeptide (1 and 2 eq.) were tested for the labelling of 50 µM protein. The results showed that using 50 µM substrate failed to achieve high levels of labelling. However, the 100 µM depsipeptide example appeared to label to near-quantitative levels (Figure 2.9). If the label is particularly precious it may be beneficial to use this lower concentration in labelling experiments.
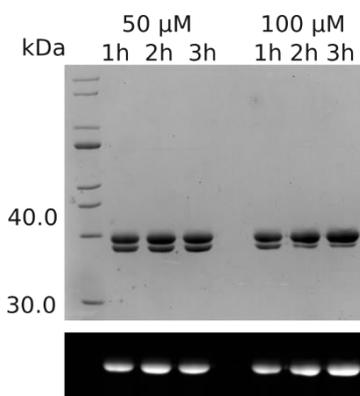
Figure 2.9: Comparison of using a reduced amount of depsipeptide substrate. 50 µM MBP with 50 or 100 µM FITC-depsipeptide or peptide using 2.5 µM Srt 7M at 25 °C.

Two final control experiments were performed to confirm that labelling is dependent on both the depsipeptide and the sortase 7M. To analyse the difference between the way Srt 7M and WT SrtA work when used with depsipeptides an experiment was carried out using 5 μM WT SrtA (20 mol%) to label 25 μM MBP (Figure 2.10). The results showed that the depsipeptides do work well with the WT for the labelling of MBP. However, after 4 hours little more than 50 % labelling had occurred, demonstrating that the Srt 7M is superior to the WT Srt.
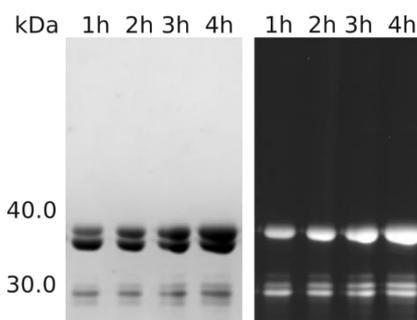
Figure 2.10: Labelling of 25 μM MBP with 75 μM FITC-depsipeptide using 5 μM WT sortase at 37 °C. The labelling appears less efficient than Srt 7M MBP labelling and the UV image highlights the increased levels of non-specific labelling taking place on sortase fragments near the bottom of the gel.

Finally, to confirm that the depsipeptide is required for faster labelling with Srt 7M and it was not due to the enzyme alone, the rate of labelling using the analogous peptide was investigated. The two reactions were performed in parallel using 25 μM MBP, 5 mol% Srt (1.25 μM) and 3 eq. of either the peptide or depsipeptide (75 μM) at 18 °C (room temperature in the mass spectrometry facility). These reactions were monitored by HRMS in addition to the standard SDS-PAGE analysis (Figure 2.11). The results show a significant difference between the two labelling substrates; the peptide fails to achieve >50 % labelling while the depsipeptide is very close to completion by the end of the experiment.
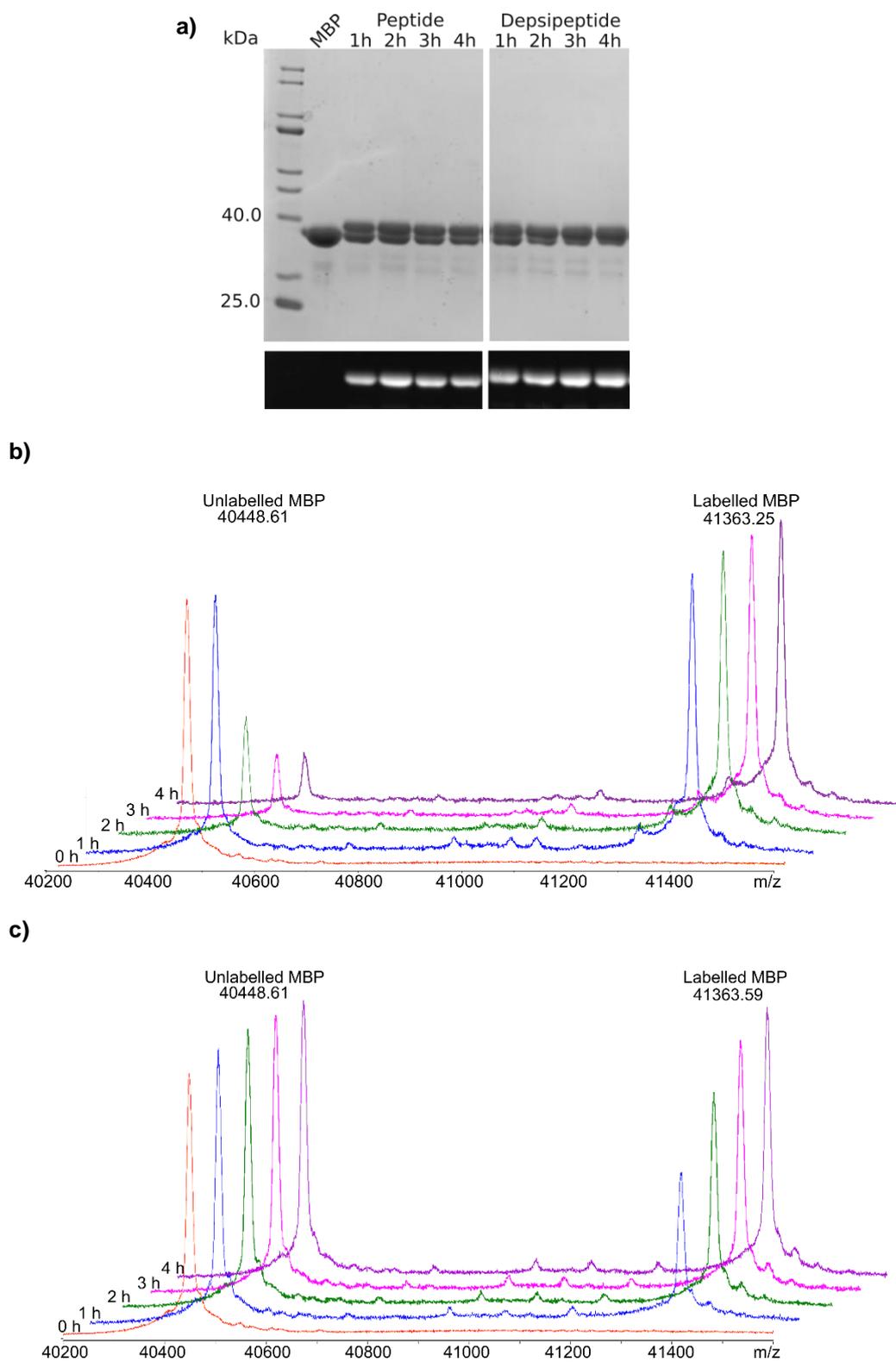
Figure 2.11: a) 25 µM MBP with 75 µM FITC-depsipeptide or peptide using 1.25 µM Srt 7M at 18 °C. b) Corresponding time points were analysed using HRMS for both the depsipeptide and c) peptide.

Although the synthesis of depsipeptides is slightly more complicated due to the need to synthesise the ToG building block, these results support the suggestion that labelling efficiency is greatly improved using them. In addition, Srt 7M is a clearly more efficient labelling reagent than the WT SrtA- less enzyme is required, fewer non-specific reactions take place and the target protein (MBP) can be labelled at temperatures as low as 4 ℃.

2.2.2   Cholera toxin B-subunit

Following confirmation that a monomeric protein could be labelled, more complex multimeric proteins were investigated. The first such system was Cholera toxin B-subunit (CTB) (Figure 2.12a). Labelling of this protein has previously been studied using WT SrtA alongside a FITC-depsipeptide and was unable to be quantitatively labelled using 120 µM CTB, 20 mol% enzyme and 3 eq. of FITC-depsipeptide at 37 °C for 3 hours.[106] For this experiment, CTB (with an N-terminal triglycine motif) was provided as purified protein by Darren Machin. The labelling reaction was then carried out using 50 µM CTB (monomer concentration) with 150 µM FITC-depsipeptide and 2.5 µM Srt 7M over 3 hours at room temperature. When analysed by both SDS-PAGE and HRMS, CTB was found to be fully labelled within 3 hours (Figure 2.12b&c).
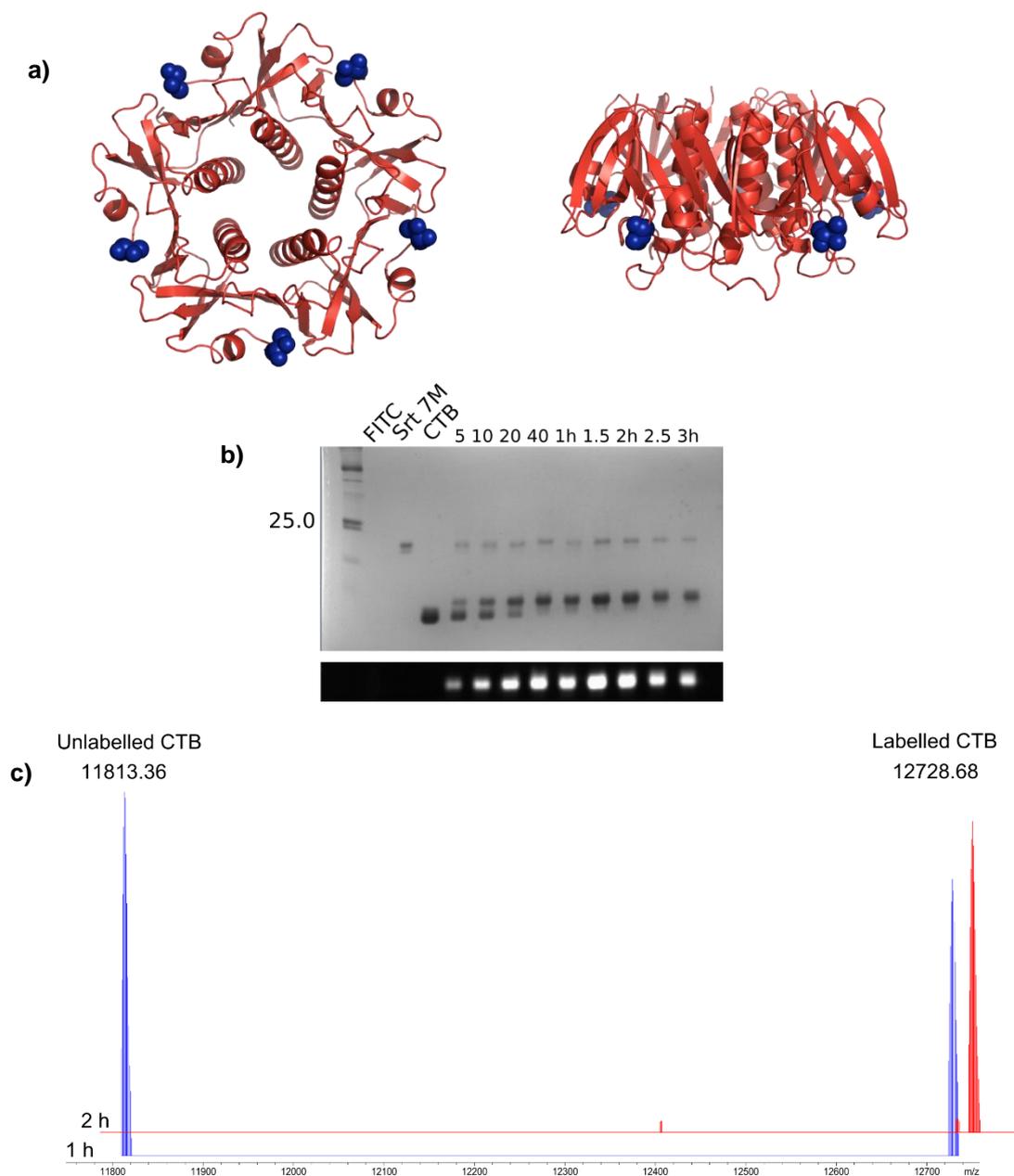
Figure 2.12: a) PYMOL image of pentameric CTB structure with N-termini pointing out of the sides of the complex. PDB ID: 3CHB[107] b) SDS-PAGE gel of time points taken during sortase-mediated CTB labelling reaction with FITC-depsipeptide c) HRMS of 1 hour and 2 hour time points leading to complete labelling of CTB.

2.2.3   PanD

Another multimeric protein frequently used within the group is PanD, a tetrameric protein used alongside PanZ as part of studying the biosynthetic pathway towards the production of

pantothenate which will be discussed in greater detail later in Chapter 4. PanD is purified using a His$_6$-tag through Ni-NTA affinity chromatography, meaning a TEV cleavage site had to be introduced onto the N-terminus along with a short GGSSS tail which would become available once the His$_6$-tag had been removed. This short tail was introduced because, unlike CTB, the N-termini are clustered within the central pore of the tetrameric structure (Figure 2.14a). Following the successful mutagenesis, the GGSSS-PanD(T57V) was transformed into *E. coli* BL21 ΔpanD ΔpanZ (DE3) cells before growth in lysogeny broth and induction of protein overexpression using IPTG. Having been grown overnight at 37 °C the cells were lysed and purified using Ni-NTA affinity chromatography and size-exclusion chromatography (Superdex® S75 in Tris-buffered saline with 0.1 mM DTT at pH 7.4) (Figure 2.13).
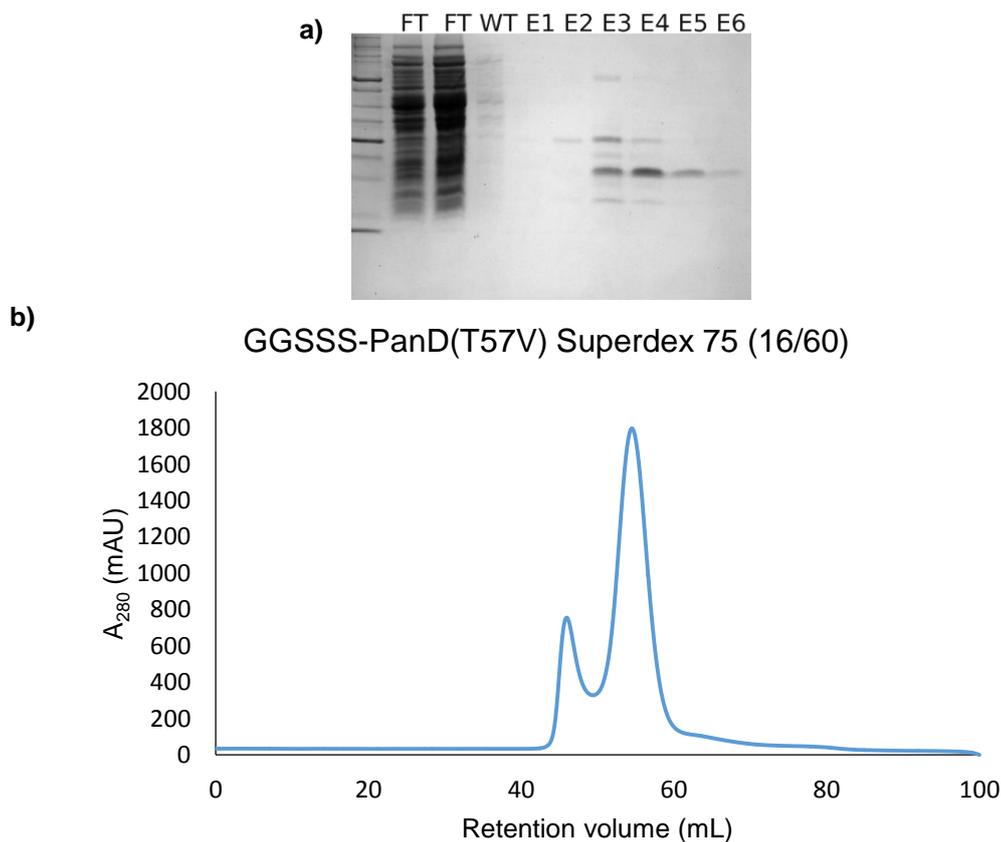

Figure 2.13: Purification of GGSSS-PanD(T57V) for use in labelling reactions a) SDS-PAGE gel of Ni-NTA affinity chromatography column flow through (FT), wash (WT) and elution fractions (E1-6). b) SEC trace for GGSSS-PanD(T57V) using Superdex® S75.

The labelling reaction was analysed over four hours at 25 °C using 50 µM PanD, 150 µM FITC-depsipeptide and 2.5 µM sortase 7M (Figure 2.14b). The progression of this time course appears to be much slower than that for either MBP or CTB, likely due to the close proximity of the N-termini. Although the short GGSSS tail had been introduced in an attempt to improve the accessibility, upon labelling their close proximity may be causing the remaining glycine residues to become sterically hindered causing the labelling to take place at a slower rate. However, the majority of the protein was labelled demonstrating that multimeric proteins can be labelled using Srt 7M.
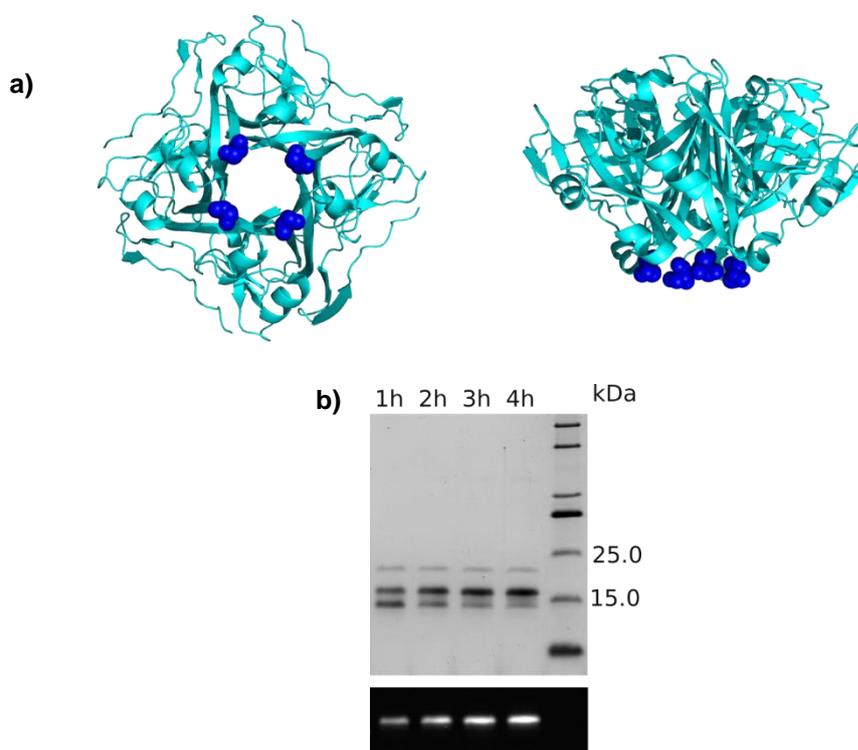
Figure 2.14: a) PYMOL image of tetrameric PanD structure with clustered N-termini within the central pore. PDB ID: 4CRZ[108] b) Time points of the labelling of 50 µM PanD, 150 µM FITC-depsipeptide and 2.5 µM sortase 7M at 25 °C.

This work has illustrated the capability of sortase 7M to label proteins with optimised conditions at 25 °C requiring 3 equivalents of depsipeptide substrate and only 5 molar percent of enzyme. By exploring these conditions further, it has been found that by either increasing the enzyme concentration or temperature can allow proteins to be labelled within 1 hour. However, it is

important to note that with increased enzyme activity, comes increased levels of hydrolysis product which has been faintly observed in both cases. It could also be said that there is evidence of this hydrolysis taking place for the lower sortase concentrations at 25 °C after 4 hours. This highlights the importance of monitoring the reaction progression either using HRMS or SDS-PAGE to ensure the maximum levels of labelling are achieved, which is also reflected in the slower rate of labelling observed for the more complex structure, PanD. Another method for monitoring the reaction progression was desired, one which could be continuous and potentially allow a greater understanding of the balance between the labelling and hydrolysis reactions. For this, fluorescence resonance energy transfer (FRET) was chosen, using the FITC species already synthesised as a donor with an appropriate acceptor.

2.3 Using FRET as a method of monitoring the sortase 7M labelling reaction

In chemical biology FRET is commonly used to monitor interactions between fluorescent molecules. For this application, the labelling was already using a fluorescent substrate in the form of the FITC-depsipeptide. However, this needed to be used to label another fluorescent species to act as a FRET pair. This chapter includes a brief explanation of FRET, followed by two different approaches employed as a method for monitoring the Srt 7M labelling system.

2.3.1   A brief introduction to fluorescence resonance energy transfer

Fluorescence resonance energy transfer (FRET) is a common tool used in chemical and biological studies to gain a greater understanding about where two fluorescent molecules are in relation to each other. This non-invasive technique can provide information about structural dynamics and protein-protein interactions which influence the fluorescence observed as a function of time.

This phenomenon involves the radiationless transfer of energy between an excited donor fluorophore and an acceptor molecule. This radiationless characteristic means the energy is not transferred *via* photons, instead the energy is transferred through dipole-dipole coupling through space. The donor absorbs energy from an external source, causing it to become excited and subsequently transfer this excitation energy on to the near-by acceptor species which does release a photon in the form of fluorescence.[109] This can be observed as a decrease in the donor fluorescence and an increase in the acceptor fluorescence. The transfer can be described using a Jablonski diagram, illustrating the transitions involved between the donor and acceptor species resulting in the acceptor fluorescence (Figure 2.15a). In order for the energy to be transferred between the two species, they have to be considered a FRET pair. This means that the emission spectrum of the donor species (e.g. FITC) overlaps with the absorbance spectrum of the acceptor (e.g. mCherry), with the degree of overlap being referred to as the spectral overlap integral (J) (Figure 2.15b).
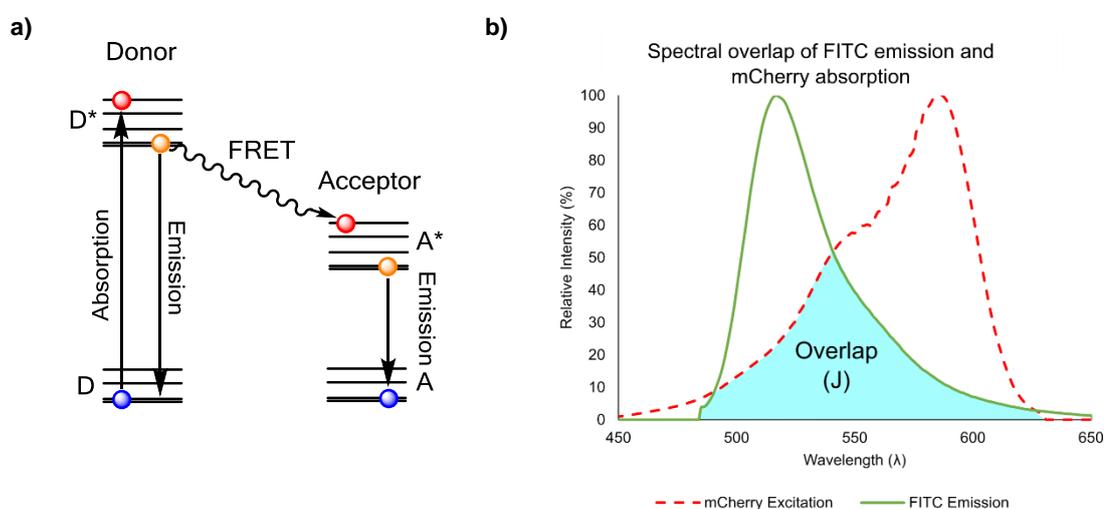
Figure 2.15: a) Jablonski diagram used to describe the energy transitions involved in FRET where an excited donor molecule is able to transfer energy to a near-by acceptor molecule, exciting it and leading to fluorescence to be observed. b) Spectral overlap of FITC emission and mCherry absorbance to illustrate the spectral overlap integral (J).

In addition to this requirement, the energy transfer is only able to take place when the donor and acceptor are in close proximity to one another, typically ranging from 1-10 nm. In 1948, Förster

showed that the efficiency of the energy transfer is inversely proportional to the distance between the donor and acceptor to the sixth power (Equation 2.1).[110] In the equation, R_0 is known as the Förster radius which is where the efficiency of the energy transfer taking place is 50%. This distance is typically between 20-60 Å. However, it is specific to the FRET pair being investigated and information on this can be found within literature.[111]

$$E = \frac{1}{1 + (\frac{r}{R_0})^6}$$

*Equation 2.1: FRET efficiency (E) allows the distance between the donor and acceptor (r) to be determined using the above equation due to the dipole-dipole coupling mechanism where R_0 is the Förster radius of the donor/acceptor pair (this is the distance where energy transfer efficiency = 50%).*[110]

The rate at which FRET can occur depends on several variables including the extent of the spectral overlap (J), the quantum yield of the donor, the orientation of the transition dipole moments and the distance separating the donor and acceptor. This allows any changes in distance between the donor and acceptor to be measured through the effect it has on the measured FRET. Because of this, FRET has been used extensively as a tool in molecular biology for studying the structure, dynamics and molecular interactions of proteins and nucleic acids both *in vitro* and *in vivo*.[112–115]

2.3.2 Applying FRET to the sortase 7M labelling system

In this work, the FITC-functionalised species was used as the donor, excited at ~494 nm and the hypothesis was that as the labelling reaction occurred, the acceptor would come into close enough proximity to allow the energy to be transferred, subsequently causing a decrease in FITC fluorescence and increase in acceptor fluorescence when monitored using a fluorimeter. In this chapter two approaches were taken; one was to label a fluorescent protein, mCherry (Scheme 2.2), and the other was the labelling of a Rhodamine-functionalised peptide. By monitoring the fluorescence of the FITC and mCherry/Rhodamine, any changes as a result of the reaction

progression would be observed and hopefully provide information about the way in which these reactions progress.
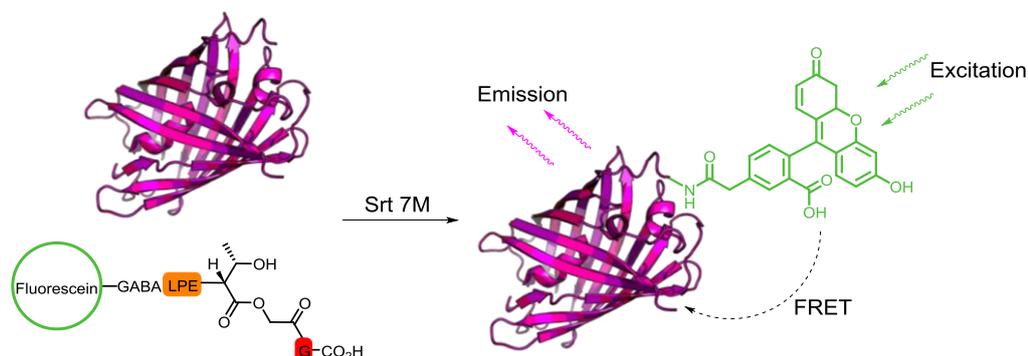
Scheme 2.2: Schematic of proposed labelling reaction to monitor any FRET observed by exciting FITC and analysing changes in mCherry and FITC emission.

2.3.2.1   Investigating the FRET observed between mCherry and FITC

By viewing the fluorescence and emission spectra of FITC and mCherry using the online tool Fluorescence SpectraViewer provided by ThermoFisher Scientific, it appeared that they should be able to behave as a FRET pair for these labelling experiments as the spectral profiles overlap sufficiently for the FITC emission and mCherry absorbance (Figure 2.16).[116]
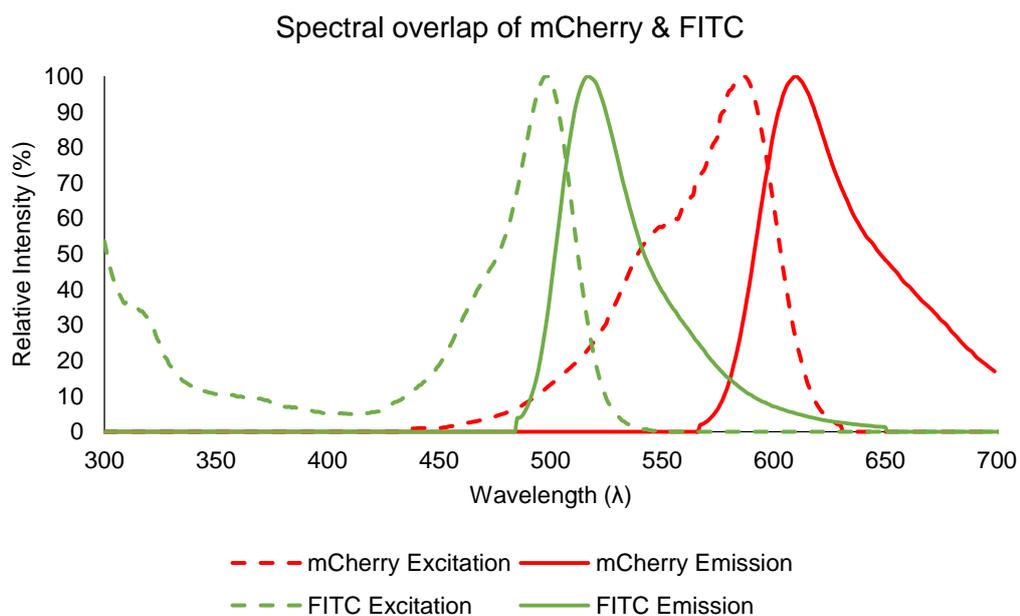
Figure 2.16: Overlapping fluorescence and absorbance spectra of FITC and mCherry.[116]

To allow labelling to be achieved, two amino acids were added to the N-terminus, GV, by Matthew Balmforth using site-directed mutagenesis. The mCherry was then expressed in auto-induction media at 37 °C for 24 hours before being lysed and purified using Ni-NTA affinity chromatography and size-exclusion chromatography (Tris-buffered saline (pH 7.4)). In an initial labelling reaction monitored by HRMS purified protein (50 µM) was almost completely labelled using 2.5 µM Srt 7M and 100 µM depsipeptide at room temperature after 1 hour (Figure 2.17).
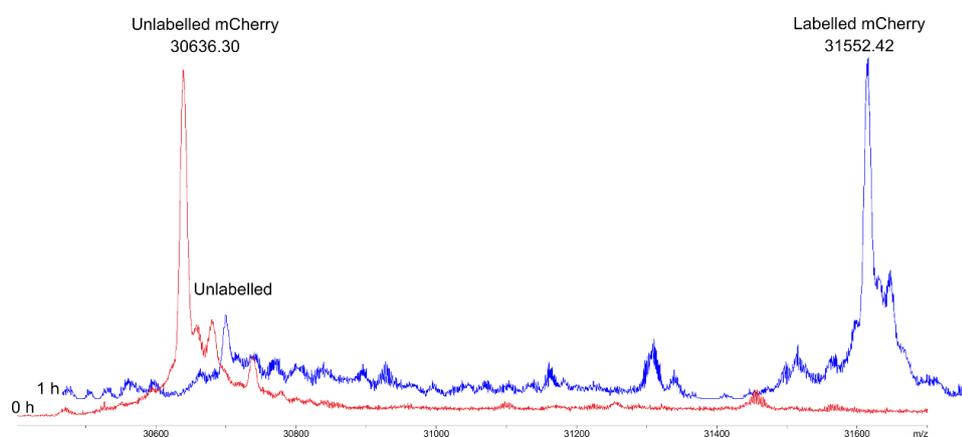
Figure 2.17: HRMS trace of before and 1 hour into mCherry labelling using 2.5 µM Srt 7M and 100 µM FITC-depsipeptide.

One important point to mention is that in early experimental investigations using the FITC as the donor, it was found to be behaving in an unusual manner at high concentrations. The higher the equivalents of FITC-depsipeptide used in the experiments, the lower the fluorescence seemed to be. To investigate this further, the FITC-depsipeptide was measured alone as a dilution series to see how the concentration affected the emission at 518 nm when excited at 470 nm. In Figure 2.18 the levels of fluorescence observed behave in a normal manner until the concentration of FITC exceeds 12.5 µM. At this point, the fluorescence begins to decrease with increasing concentration. This is likely to be due to the phenomenon of self-quenching, often observed when molecules have an overlap in their own absorbance and emission spectra and are at a high concentration and therefore close to one another in the solution.[117]
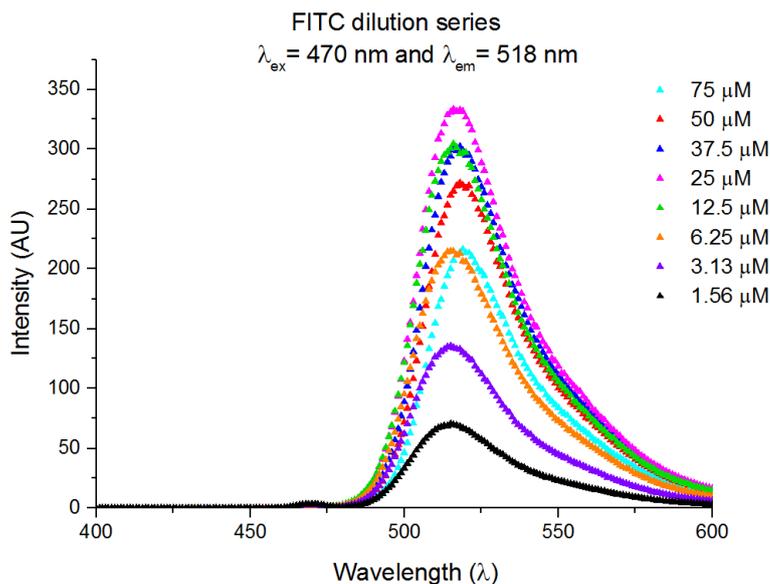
Figure 2.18: FITC dilution series using the FITC-depsipeptide from 1.56-75 µM to observe the effect of self-quenching at the higher concentrations resulting in the decision to keep the concentration ≤10 µM during the FRET experiments.

By referring to Figure 2.16 it is possible to see that this could be having an effect when using the FITC-depsipeptide. To avoid this problem, the concentration of the FITC species was maintained below 10 µM. To achieve this another non-fluorescent analogue was synthesised using SPPS, which will be referred to as the "NF-depsipeptide" (Figure 2.19a). The NF-depsipeptide was used as the main substrate in the reactions and the mixture was effectively "spiked" with the FITC species. There is a risk that the non-fluorescent peptide would behave differently within the system, however a comparison of the data from before and after spiking showed similar behaviour with regards to the shapes of the curves (Figure 2.19b&c)
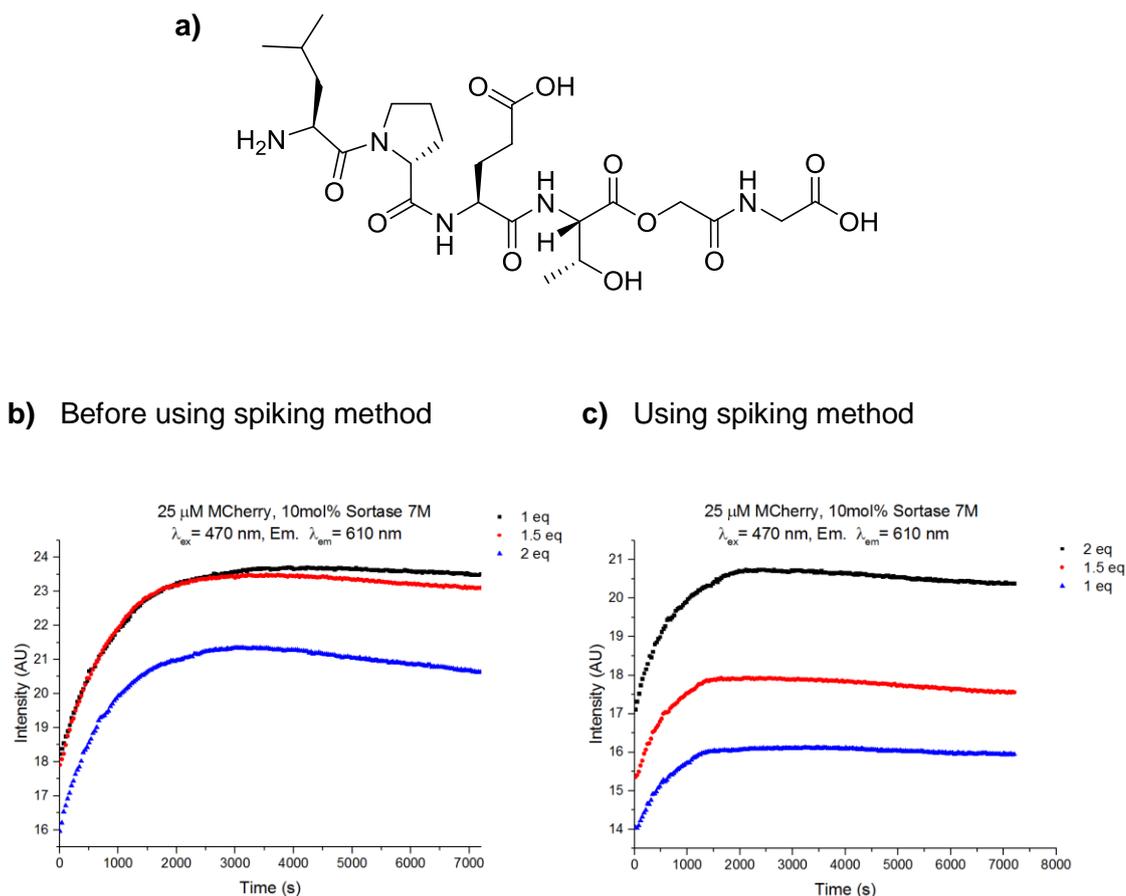
Figure 2.19: a) NF-depsipeptide used to spike the FRET experiments to avoid FITC self-quenching b) The higher concentrations of FITC-depsipeptide were appearing as lower intensities thought to be due to self-quenching c) By introducing the non-fluorescent depsipeptide the expected pattern of intensities was observed.

The mCherry (10 µM) was mixed with the non-fluorescent and FITC depsipeptides in a ratio of 4:1 at varied concentrations (10-20 µM) before the Srt 7M (1-5 molar%) was added and quickly transferred into the cuvette ensuring no bubbles were present. The first 10 minutes of each reaction was monitored, looking at the changes in emission intensities at 518 nm (FITC emission) and 610 nm (mCherry emission) when the FITC was excited at 487 nm. As expected, the percentage of sortase added to the reaction affected the rate of intensity change seen using the fluorimeter. In Figure 2.20 each graph depicts a different concentration of sortase added to the experiment with the gradient of the downward slope for FITC emission becoming greater for the higher concentration of sortase. Graphs depicting the mCherry emission appear to show an

increase in the mCherry emission as the labelling reaction progresses. These experiments show that Srt 7M is successfully labelling the low concentration of mCherry and FRET is taking place. The depsipeptide substrate concentration appears to have only a small effect on the slope gradient. However, the change in sortase concentration increases this, implying an increased rate in labelling as would be expected.
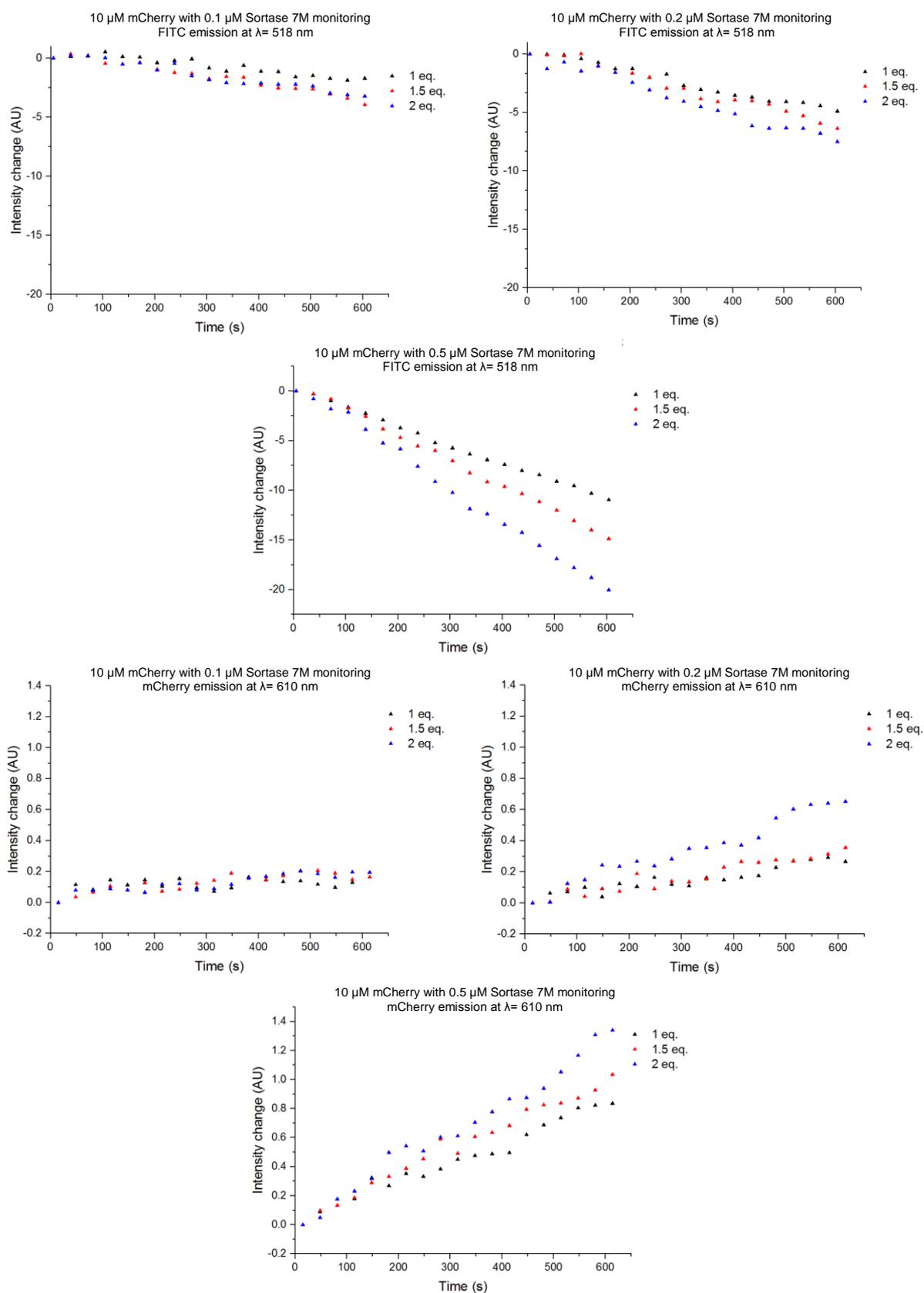
Figure 2.20: Graphs showing the decrease in FITC emission and increase in mCherry emission observed over the first 10 minutes of a labelling experiment between 10 µM mCherry and varied quantities of FITC-depsipeptide while using 1-5 mol% Srt7M.

Further experiments were carried out using a higher concentration of mCherry (25 µM) and Srt 7M (2.5 µM). In these experiments the NF-depsipeptide was used alongside the FITC-depsipeptide in a ratio of 9:1 due to the higher overall reaction concentration. Monitoring the FITC emission, the reactions appeared to all start in a similar manner with a steep initial gradient for the first 15 minutes before beginning to plateau (Figure 2.21). This point of plateau differed depending on the concentration of depsipeptide with the higher the depsipeptide concentration leading to a greater intensity change implying more of the mCherry had been labelled. When analysing the equivalent data for the emission in mCherry, additional information seems to be present. The curve is not as smooth, but the most noticeable characteristic is the way in which the curves reach a maximum at the ~15 minute mark, plateau for a few minutes, and then begin to come back down again. The fact that the FITC curves don't reflect this raised questions about whether this could be due to the dark-state phenomenon and irreversible photobleaching of mCherry.[118,119] This is a characteristic of fluorescent proteins that can limit the brightness of the fluorescence and can be useful in super-resolution microscopy, however for this application it can limit the reliability of the results obtained. As a more robust method, a Rhodamine-peptide containing an N-terminal glycine residue was chosen as the acceptor molecule in experiments discussed in the next section.
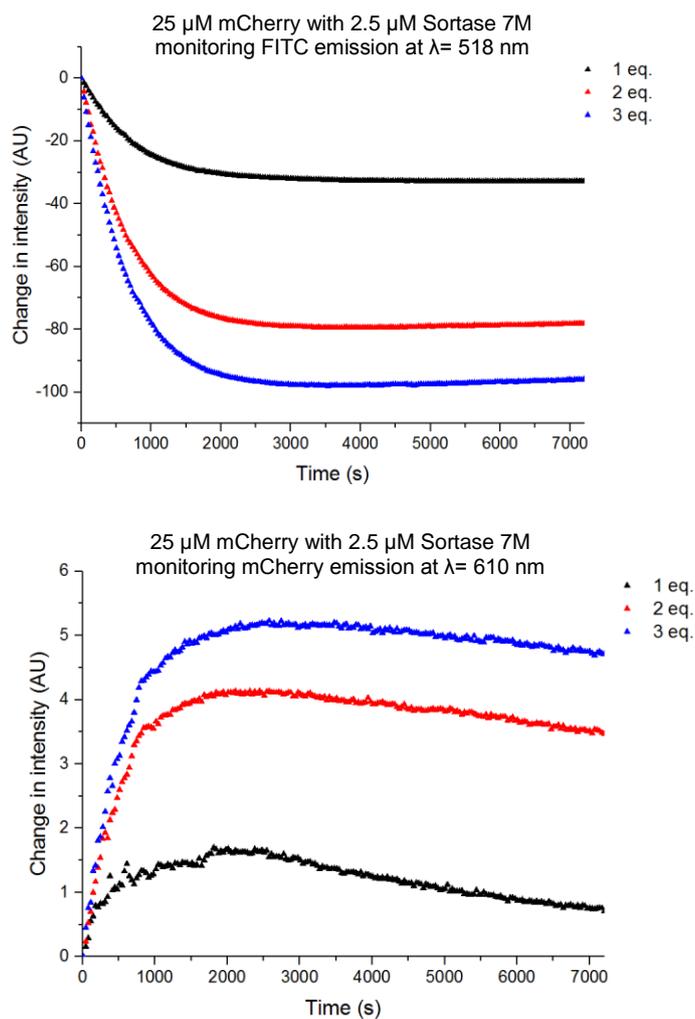
Figure 2.21: FRET experiments monitoring the change in fluorescence intensity observed over 2 hours of sortase-mediated labelling reactions between 25 µM mCherry and FITC using 10 mol% Srt 7M and varied concentrations of FITC-depsipeptide. A decrease in fluorescence was observed through excitation of FITC at 470 nm while monitoring fluorescence of FITC at 518 nm with an increase observed in the emission from mCherry at 610 nm.

2.3.2.2 Comparing the labelling of a Rhodamine-peptide using FITC-depsipeptide and FITC-peptide

The spectral data for FITC and Rhodamine overlap in a manner which should allow FRET to be observed when exciting the FITC (Figure 2.22).

Spectral overlap of Rhodamine & FITC
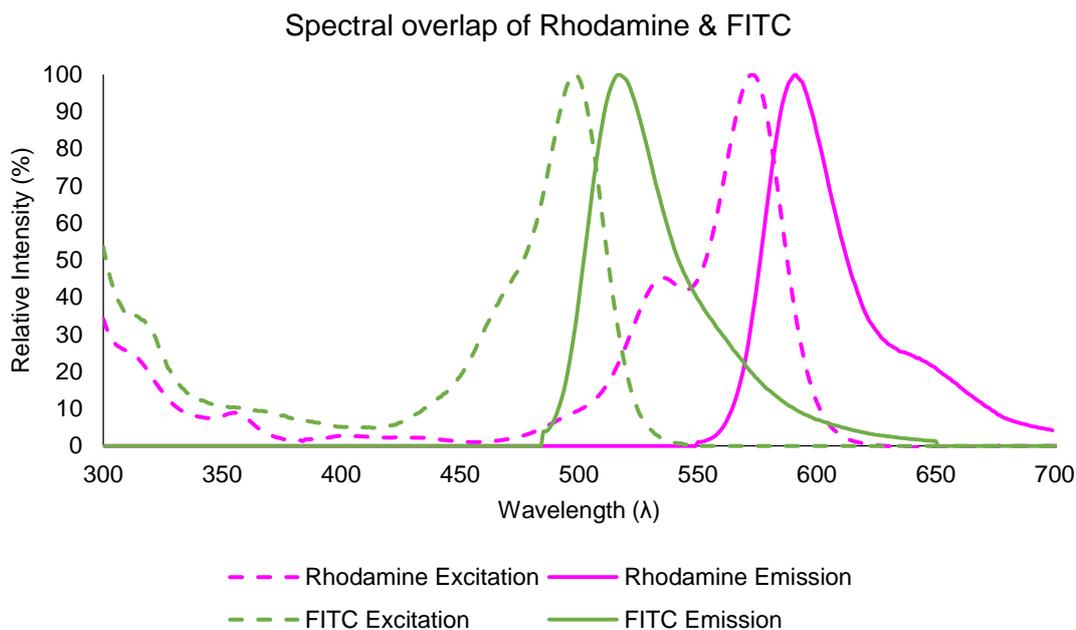
Figure 2.22: Graph of spectral overlap between Rhodamine and FITC absorbance and emission spectra.[116]

As mentioned in the previous discussion, a non-fluorescent version of the donor species was used alongside the FITC-species to prevent the possibility of self-quenching. Due to this being a comparative study between the peptide and depsipeptide labelling reactions, both the FITC and non-fluorescent peptide (NF-peptide) analogues were synthesised using SPPS. The Rhodamine peptide used in these experiments was synthesised by Kristian Hollingsworth with a sequence of H_2N-GVSK-Rhod-OH. The Rhodamine species was used at 25 µM with 3 molar equivalents of the depsipeptide or peptide species (in a 9:1 NF/FITC ratio) and 10 mol% of Srt 7M. The sample was excited at 468 nm and data was collected for Rhodamine emission at 580 nm. By comparing the data obtained for the different labelling substrates, it was clear to see that by using the depsipeptide species increased the overall amount of FRET observed than when the peptide was used (Figure 2.23). This provides additional evidence, alongside the HRMS study in the previous section, for the depsipeptide being a more efficient labelling substrate when using the Srt 7M. This data also showed signs of Rhodamine signal decreasing over the longer periods of time which could be signs of the threonine-glycine bond hydrolysing. If so, this FRET method could

potentially be used as a tool for analysing ways of reducing the levels of hydrolysis. Methodically

changing conditions and reagents could be find an even more efficient way of using Srt 7M and

depsipeptides for protein labelling.
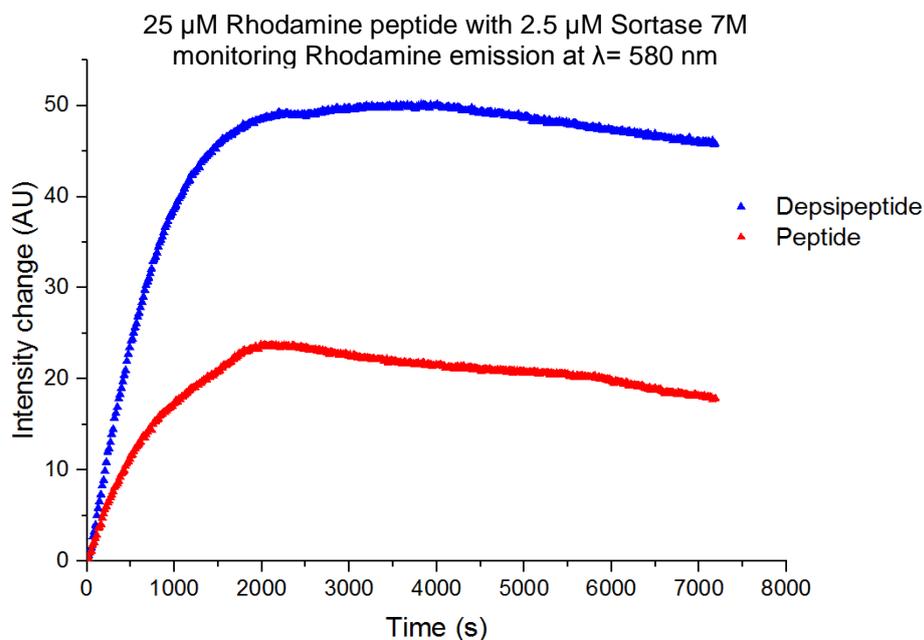
Figure 2.23: Average plot of three repeats comparing the FRET observed when labelling 25 μM Rhodamine peptide using 3 molar equivalents of either the FITC-depsipeptide of FITC-peptide species alongside 10 mol% Srt7M.

2.4   Conclusions and future work

The results discussed in this chapter have investigated the combination of two methods for

improving sortase-mediated labelling as a tool for chemical biology. By bringing together the

highly active sortase mutant, Srt 7M, and the irreversible nature of experiments using

depsipeptide substrates, it has been possible to achieve quantitative labelling of a monomeric

species, MBP and pentameric species, CTB. In addition to this, both of these reactions took place

using significantly reduced enzyme concentrations in comparison to the wild-type sortase, with

the best results observed at 25 °C. Work using PanD also illustrated the effect steric hindrance of

the N-termini can have on the rate of the enzymatic reaction, highlighting the importance of monitoring the progress of labelling for systems of differing complexity.

In addition to this, at high concentrations of sortase 7M, high temperatures or longer time scales, signs of formation of a hydrolysis product were visible. However, by balancing these three variables appropriately depending on reagent and time availability, it can be possible to choose conditions where this unwanted product is minimised. By using FRET, both the labelling and hydrolysis processes were able to be monitored for the labelling of mCherry and a Rhodamine peptide in a continuous manner. The reactions appeared to have reached a maximum level of labelling after ~30 minutes before a decrease in Rhodamine emission was measured, potentially caused by the hydrolytic effect. FRET proved itself to be a very useful technique for monitoring the reaction progression and can be used to ensure maximum levels of labelling are achieved for the system of choice.

Experiments monitored by SDS-PAGE, HRMS and FRET all illustrated the improvement in labelling achieved using the sortase 7M alongside depsipeptide substrates instead of using the WT sortase or peptide substrates. Overall, by pulling together these two improvements in the area of sortase-mediated protein labelling an efficient and useful labelling tool has been described using 3 equivalents of depsipeptide, 5 molar% sortase 7M with incubation at 25 °C proving to be the most effective labelling strategy.

3 Use of protein-DNA conjugates in structural studies

The work described in this chapter discusses the generation of protein-DNA conjugates using sortase 7M-mediated protein labelling with applications to high speed atomic force microscopy (HS-AFM) and small angle X-ray scattering (SAXS).

3.1 Preparation of bio-orthogonal labels

Previous work in the group has used WT sortase to partially label proteins with bicyclononyne labels followed by SPAAC with azido-functionalised RNA oligonucleotides to generate protein-fusions. While effective, we initially wanted to use a more rapid labelling method and therefore investigated the use of the more reactive *trans*-bicyclononene. It was attempted to be generated following the photochemical isomerisation protocol reported by Fox and co-workers to form *trans*-cyclooctene.[120] This approach involves the UV-irradiation of a sample of *cis*-cyclooctene within a flow system driven by the complexation of the *trans*-isomer to silver nitrate (Figure 3.1a). This has allowed them to generate a number of *trans*-cyclooctene species to be formed including *trans*-bicyclononene, while avoiding the complex chemical synthesis route otherwise required (Figure 3.1b).[120,121]

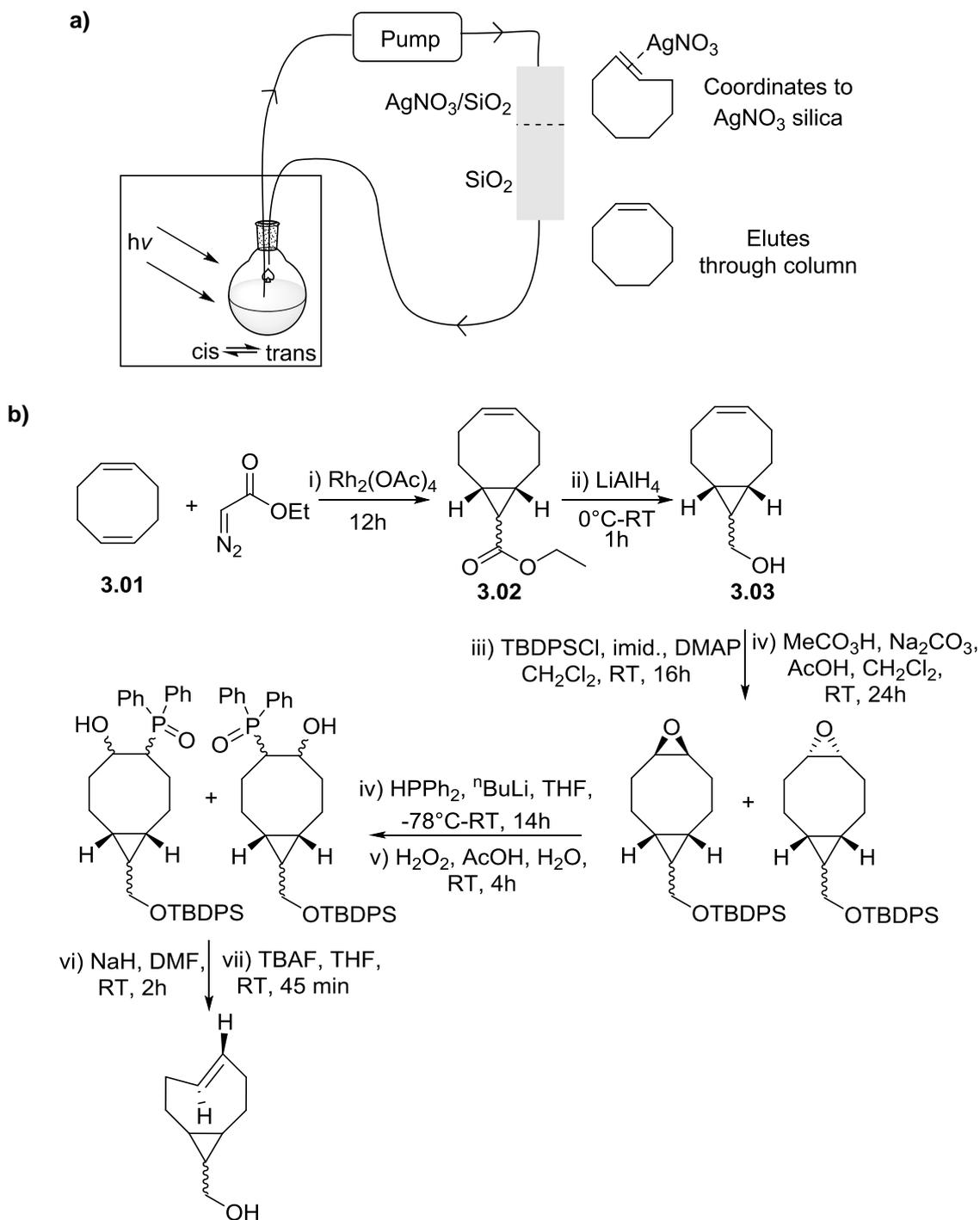*Figure 3.1: a) Photochemical isomerisation set up developed by Royzen et al. for the formation of trans-cyclooctene from cis-cyclooctene through UV-irradiation. b) Synthetic route of trans-bicylo[6.1.0]non-4-ene-9-ylmethanol from cyclooctadiene.*[120,121]

The initial aim was to recreate this photochemical system using an HPLC pump and mercury lamp with UPVC tubing coiled around a quartz immersion well to produce

trans-bicylo[6.1.0]non-4-ene-9-ylmethanol. Silver nitrate impregnated silica was made by mixing the silica thoroughly with a silver nitrate solution in the dark before being dried and packed into a column which was incorporated into the system. This was subsequently flushed with hexane to ensure the silica column was packed fully and could tolerate the 10 mL min^{-1} flow rate of the HPLC pump. Test reactions used *cis*-cyclooctene with an excess of methyl benzoate (a singlet sensitiser required to allow the isomerisation to take place), within a 1% ether in hexane solvent system. The reaction mixture was stirred to maintain a homogenous solution while being pumped through the tubing around the mercury lamp (λ= 254 nm) and onto the AgNO$_3$-silica. Following 12 hours of continuous flow, the silica was removed, dried and stirred with a mixture of ammonium hydroxide and dicholoromethane before being filtered. The NH$_4$OH layer was separated and washed with CH$_2$Cl$_2$ before combining the CH$_2$Cl$_2$ layers and evaporating under reduced pressure. However, when analysed by ^{1}H NMR it was apparent that only starting material was present. Following several unsuccessful attempts, comparison to the work by Royzen *et al.* indicated that the problem could be associated with the flow rates being used. The HPLC pump available in the laboratory had a maximum flow rate of 10 mL min^{-1}, however in the literature they used flow rates of 100 mL min^{-1} and a light box containing eight mercury lamps. This is a substantially higher intensity of illumination than can be achieved using the available equipment. Due to the small conversion of the *cis*-isomer to *trans*-, it seemed the equipment they used was vital for the conversion levels they observed. As a result of this, the decision was made to return to the use of the bicyclononyne functionality for which the synthetic approach was already well established within the laboratory.

### 3.1.1 Synthesis of the bicyclo[6.1.0]nonyne functionalised depsipeptide

The bicyclononyne was synthesised *via* several steps described by Dommerholt *et al.* and illustrated in Scheme 3.1.[28] This resulted in a carbonate form which could be easily coupled to the peptide by stirring in DMF at room temperature in the presence of NEt$_3$ for 3 hours.
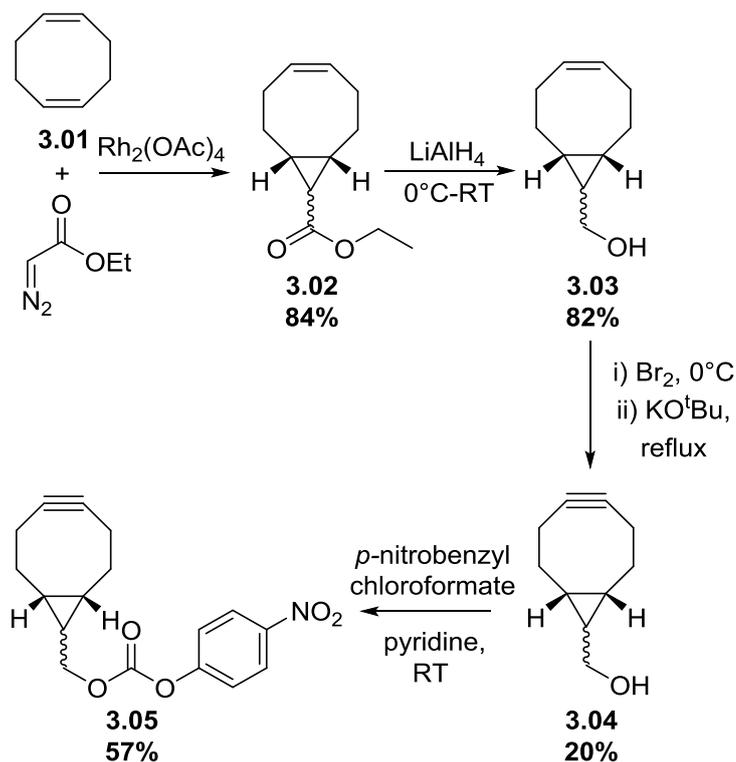
Scheme 3.1: Synthetic route for bicyclo[6.1.0]nonyne carbonate for coupling to depsipeptide species. As described by Dommerholt et al.[28]

The synthesis of the BCN is not trivial; although it was successful and allowed the isolation of a carbonate product, over the four steps the yield was extremely poor (<10%) due to the lengthy and complex route. In the first step, ethyl diazoacetate is added to cyclooctadiene (**3.01**) over 12 hours using a syringe pump, in the presence of a Rhodium (II) catalyst, this affords the formation of the bicyclononene species (**3.02**) as a mixture of *syn-* and *anti*-isomers. They were separated by flash column chromatography and the following steps were carried out on the isomers separately. The bicyclononene was reduced using LiAlH₄, before forming the alkyne bond through dibromination and double *syn* elimination. The formation of the alkyne bond using this method failed to achieve yields >20%. This is generally believed to be due to the formation of side products (Scheme 3.2). One potential side reaction could be the formation of a dimeric species (**3.07**); under basic conditions, the hydroxyl group on the bicyclononane dibromide species would be deprotonated, in this form it could displace one of the bromine atoms on another

dibromide forming a dimer. In addition to this, another potential side product could be the result of the elimination reaction taking place to form the diene (**3.09**).
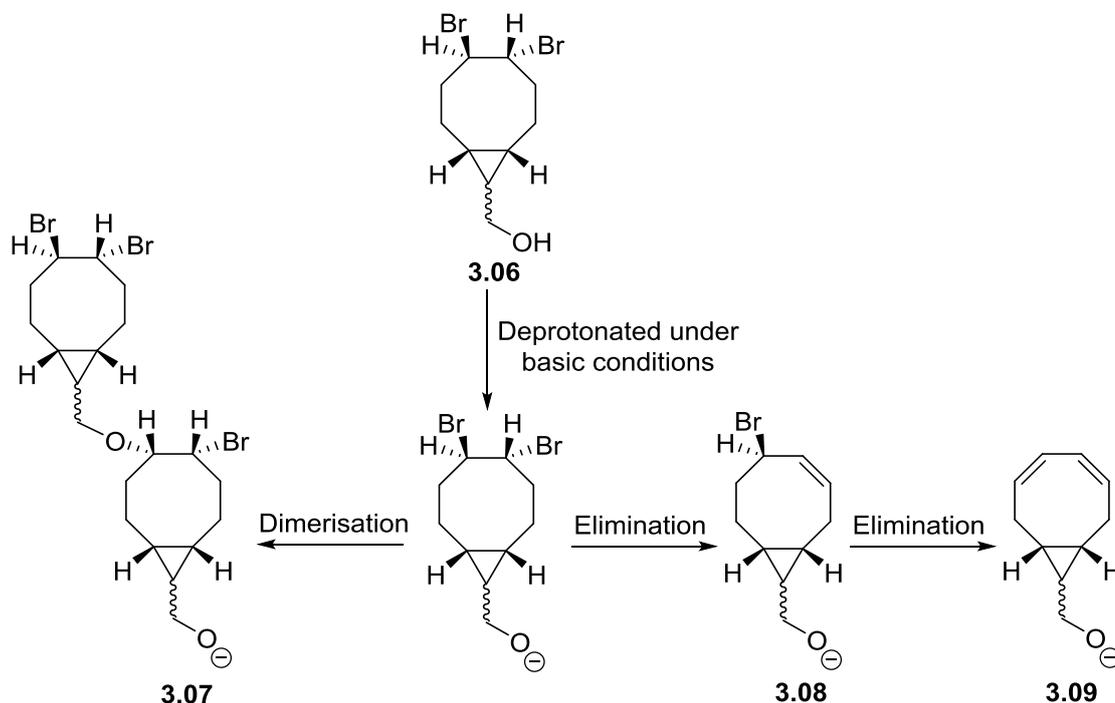
Scheme 3.2: Proposed side products as a result of the debromination and double elimination required to form the triple bond in bicyclononyne.

Once the bicyclononyne has been isolated, a room temperature stir in the presence of pyridine and *para*-nitrobenzyl chloroformate yields the final product (**3.05**) for coupling to the depsipeptide. To achieve this the depsipeptide and BCN carbonate were stirred overnight at room temperature in solution (DMF) in the presence of trimethylamine (Scheme 3.3). This simple coupling step was successful, achieving high yields (80%) and allowing the BCN-depsipeptide (**3.12**) to be isolated for future sortase-mediated labelling reactions. The synthetic route to this BCN-depsipeptide is lengthy and low yielding. Although the reasons for this have been discussed above, the overall process is time consuming and expensive with the Rhodium (II) acetate catalyst costing ~£800/g.[122] While carrying out this piece of work, a commercially available mixture of the BCN isomers became commercially available, they were provided in the alcohol form which

was turned into the carbonate species by Katherine Horner before being coupled to the peptide in the same way. The experiments in this section used both the commercial and synthesised BCN.
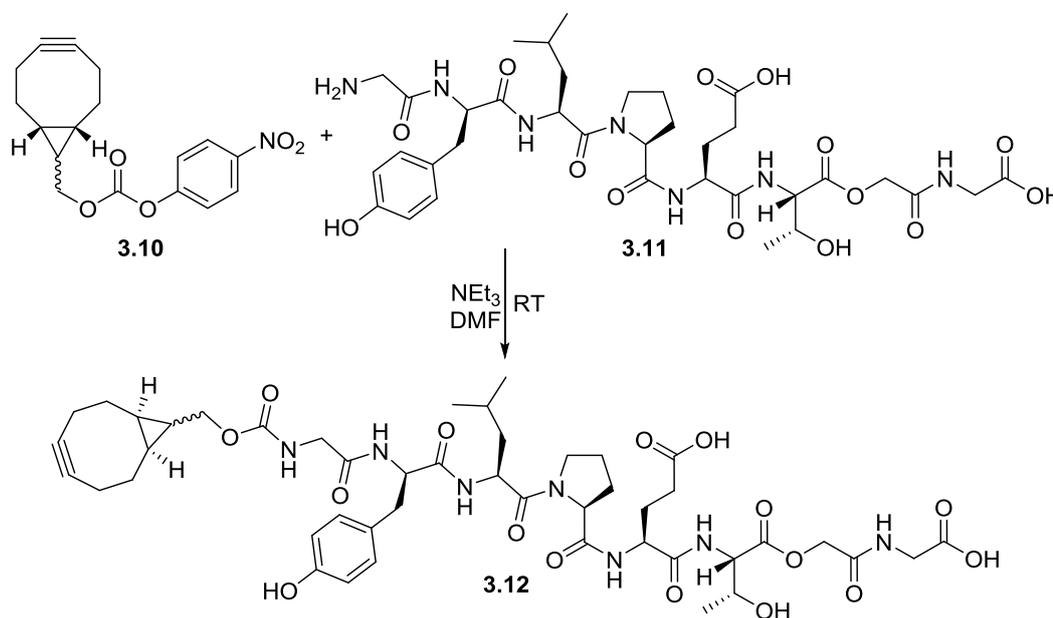
Scheme 3.3: Simple coupling step to attach the bicyclononyne carbonate to the depsipeptide species synthesised using solid phase peptide synthesis.

3.2 Labelling with a short oligonucleotide for use in HS-AFM

The initial application for the BCN-functionalised protein was to provide a method for attaching proteins to DNA origami frames for use in high-speed atomic force microscopy (HS-AFM) to monitor enzymatic reactions. This work was done in collaboration with Andrew Lee and the aim was to create a DNA origami frame with an internal cavity containing an anchor point for the protein to attach to. It was hoped that by using this approach it would be possible to watch the Srt-mediated interchange of proteins in real-time.

3.2.1 Introduction to atomic force microscopy

Atomic force microscopy (AFM) is a high-resolution type of scanning probe microscopy (SPM), first used experimentally in 1986, which has become a commonly used technique across

chemistry, physics and biology.[123] The AFM set-up consists of a cantilever with a small tip on the end, called the probe; a detector for monitoring the motion of the cantilever and a form of motion control to allow the cantilever or sample to move in three dimensions (*x, y, z*). The cantilever scans over the surface of interest allowing the tip-surface interaction to be mapped using the atomic forces experienced by the probe. The AFM measures the vertical and lateral deflections of the probe by using an optical lever, this is done through the reflection of a laser beam off the cantilever. The reflected laser beam hits a photodetector which picks up changes in the position of the cantilever and allows the deflections to be recorded (Figure 3.2). Depending on the sample being studied, the probe can interact with the surface in different ways, for example, when studying biological specimens, the probe is oscillated at a constant amplitude which can then be used to provide topographical information about the surface.[124]
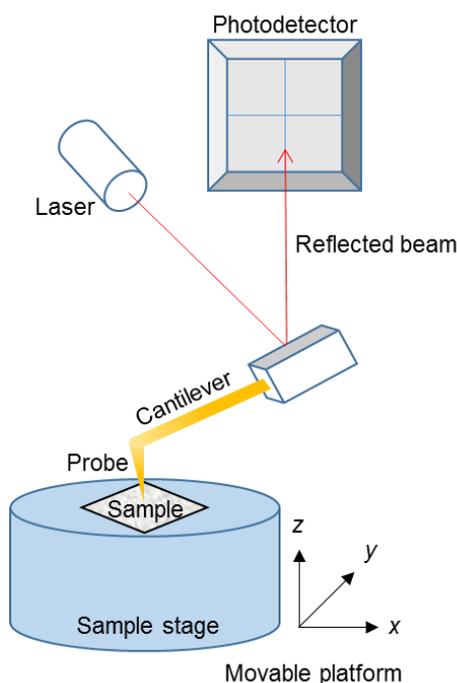
Figure 3.2: Simple AFM set up with a movable sample stage and the probe controlled by the movement of the cantilever controlled by a laser. The deflection of the probe causes the reflected laser to move which is detected by a photodetector.

3.2.2   How high-speed AFM works

High-speed AFM was developed as a method of analysing a surface with sub-100 ms time resolution, which has proved itself to be a valuable technique for monitoring dynamic protein environments, capturing molecular videos of proteins and providing an insight into how they function. Other single-molecule techniques used for studying protein dynamics like single-molecule fluorescence microscopy have been used extensively, however it actually monitors the behaviour of an optical marker, and the protein itself is invisible.[125] By using HS-AFM it is possible to directly observe the structure and dynamics of proteins undergoing conformational changes alongside function without disturbing the sample. First developed in 1990s,[126,127] this technique has advanced over the last few decades and now has the ability to image >10 frames per second.[128,129] The HS-AFM instrumentation set-up involves the cantilever oscillating in the z-direction with optimised parameters to ensure a weak tip-sample interaction is maintained to reduce the chance of sample damage.[130,131]

HS-AFM has been used to visualise a multitude of dynamic protein events from conformational changes, DNA-protein interactions and motor actions, with a particularly ground-breaking piece of work on myosin V. Kodera *et al*. provided a molecular movie of the myosin V walking along the actin filaments, confirming the mode of motion (Figure 3.3).[132] The group responsible for this piece of work also demonstrated the conformational changes Bacteriorhodopsin exhibits in response to light[133] and the rotary catalysis of the rotorless F_1-ATPase[134] using HS-AFM. The unique feature HS-AFM provided to assist these pieces of research is the ability to create a movie of the multiple dynamic motions taking place within the biological sample.
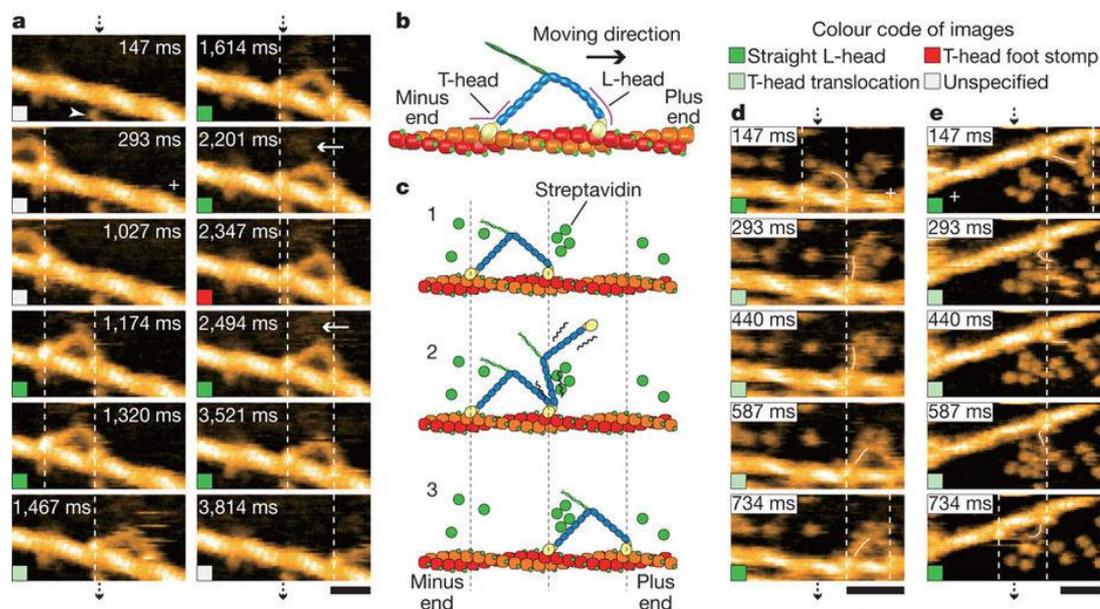
Figure 3.3: HS-AFM images of Myosin walking along a filament. a) Successive AFM images showing movement of the myosin in 1 µM ATP. b) Schematic of when the two heads are bound. c) Schematic explaining e and d, d & e) Successive AFM images showing a hand-over-hand movement in 1 µM ATP (d) and 2 µM ATP (e) with the swinging lever highlighted with a thin white line. All scale bars are 30 nm apart from d (50 nm) and images were taken at 146.7 ms per frame. This work by Kodera et al. was ground breaking in using HS-AFM to monitor the motion of myosin in such detail and confirm hypothesis of how this protein moves. [132]

3.2.3 DNA origami

DNA origami was first developed by Rothemund in 2006 and was based on the desire to find a method to create highly complex nanostructures using a "one-pot" approach. This was achieved using numerous short strands of DNA to act as "staples", allowing the folding of a long single strand of DNA in the form of a large single stranded circular viral genome called M13mp18. This allowed the formation of complex nanoscale shapes including a star and smiley face which were ~100 nm across and could form very intricate structures (Figure 3.4). Further to this, Rothemund described a three-dimensional approach, whereby DNA loops were incorporated in a manner which meant they were able to extend out of the surface of a DNA "tile" allowing writing and images to be built on top of a flat DNA origami surface (Figure 3.4c).[135]
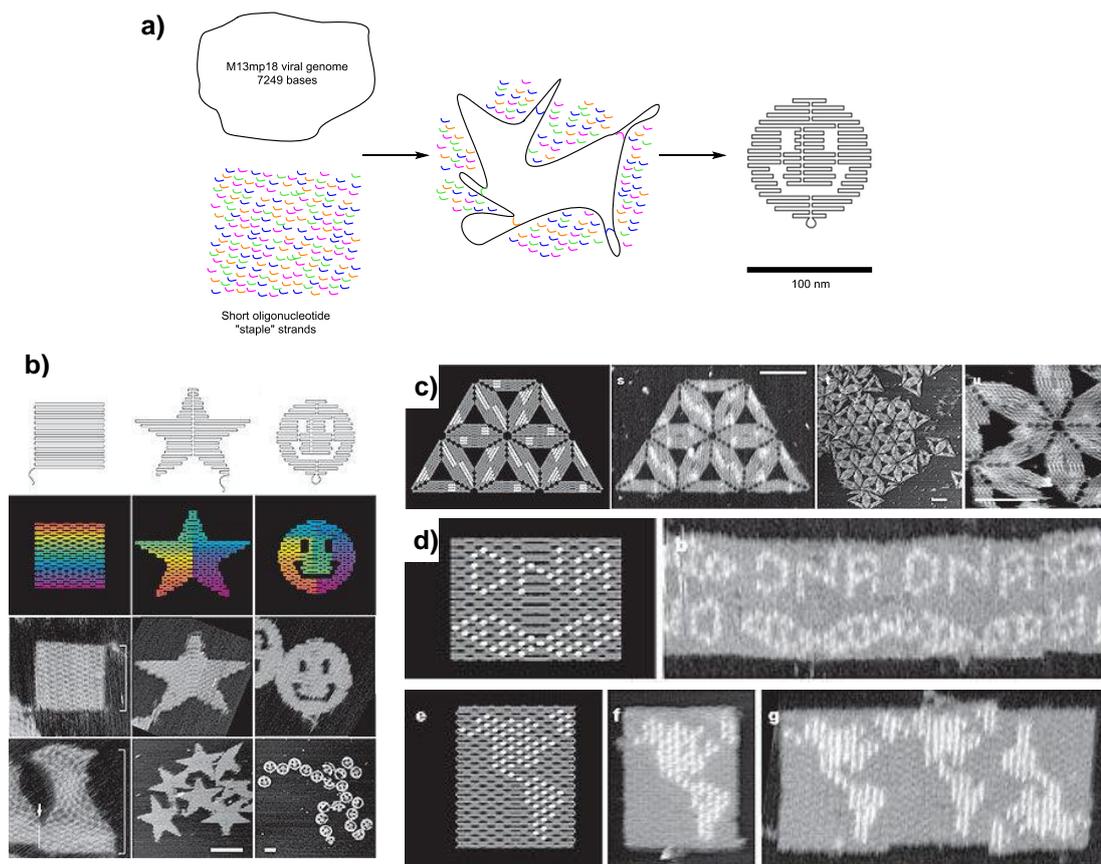
Figure 3.4: a) Schematic depicting the way in which the DNA is folded starting from a large viral genome and hundreds of DNA "stapling" strands. They are mixed and slowly cooled from 95 °C to 20 °C which allows the staples to fold the DNA into the desired pattern. b) Some examples of the intricate shapes made. c) The pin-board approach used by Rothemund whereby DNA loops protrude out of the surface of a flat DNA "tile". d) Another pattern formed from folded DNA displaying the level of intricacy which can be achieved using this approach. Adapted from Rothemund.[135]

3.2.4 The DNA origami frame design for the incorporation of MBP

To incorporate the MBP into the origami frame, a short azide-functionalised oligonucleotide (N_3-oligo) was provided by Darren Machin following large scale synthesis while on placement at GlaxoSmithKline. First the MBP was labelled with the BCN-depsipeptide and then this was reacted with the azide-oligonucleotide through SPAAC to create an MBP-oligonucleotide fusion (MBP-oligo, Figure 3.5). The oligo is complementary to the "sticky end" anchor points on the DNA frame, which should allow them to anneal to each other when mixed together. Firstly, it was necessary to see if the MBP-oligo stuck to the anchor DNA within the frame, placing the MBP within the window. Then the effect of adding more sortase to the sample needed to be

analysed to see if it lead to the recognition of the sequence present between the oligo and MBP and cause the MBP to be removed from the window, all while monitoring the dynamics using HS-AFM. Finally, it was hoped that the addition of both sortase and an excess of a second, larger protein, can lead to an exchange of proteins.

The origami frame required was designed and created by Andrew Lee (see Section 6.3.9), the structure consists of 32 parallel DNA helices of 288 bp in length, with an internal square cavity of 128 bp, which is approximately 43 nm across. The external dimensions of the frame are ~100 nm × 80 nm with a small notch taken out of one corner to act as a polarity marker. It is important to note that small variations in the dimensions can be observed due to the structure expanding or contracting depending on salt concentrations which can affect the extent of repulsion exhibited between the DNA molecules.

Two anchor points for attaching the protein were designed to symmetrically stick out of the main frame into the cavity, consisting of a single strand of DNA complementary to the N_3-oligo used for the protein labelling. This would hopefully result in the formation of a 36 bp double stranded piece of DNA pointing into the cavity with the protein of interest attached to the end. One important requirement was for there to be no excess N_3-oligo left within the product, this was to avoid the possibility of it attaching to the frames without any protein, leaving an empty window (Figure 3.5).
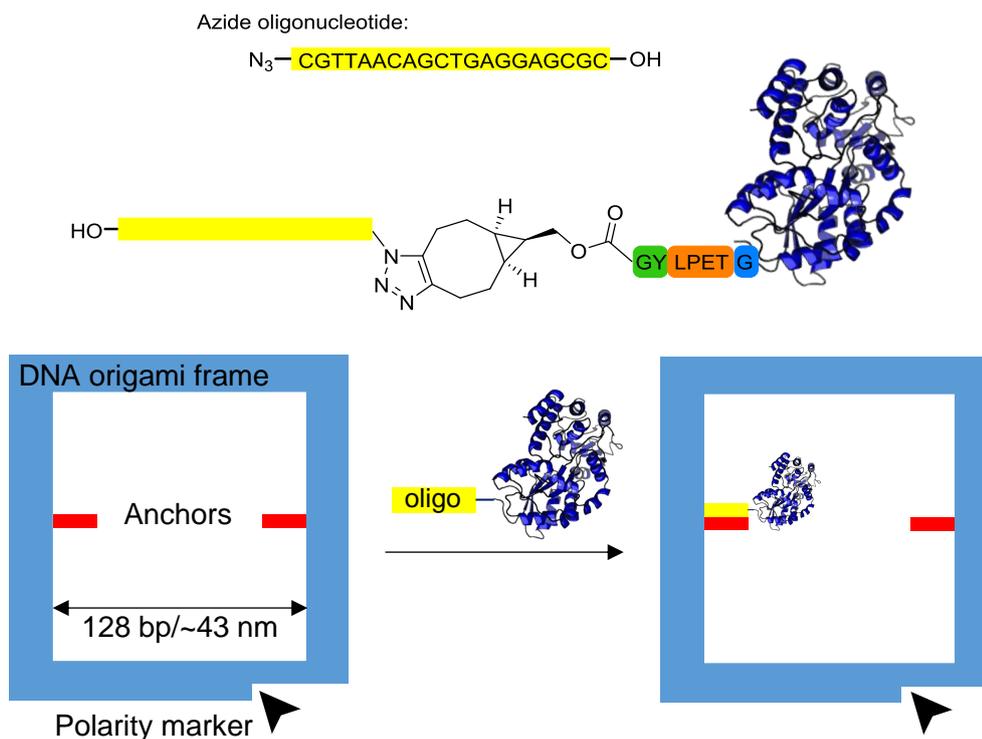
Figure 3.5: The BCN-depsipeptide incorporated into the MBP was reacted with an azide-terminated oligonucleotide to create a MBP-oligo fusion. This was then added to a specially designed DNA origami frame containing specific anchor points to allow the protein to become attached to the frame and become placed within the window.

When sortase is used in this labelling, the peptide recognition sequence, LPETG, is left behind between the protein and oligonucleotide. Because of this, by adding more enzyme to the protein-decorated frames, it would be possible to use the HS-AFM technique to observe whether the sortase recognises the sequence and interacts with it in some way.

3.2.5 N-terminal labelling of proteins using BCN-depsipeptide alongside sortase 7M

Maltose binding protein was selected as the model protein for the biophysical applications, as discussed in Chapter 2 it was labelled well using the FITC-depsipeptide. For the purpose of this work, the MBP was labelled with the BCN-depsipeptide using 50 µM MBP, 3 eq. depsipeptide and 5 mol% Srt 7M. First, the reaction was performed at room temperature for 2 hours and followed by HRMS, as can be seen in the MS images the MBP appears to label well within the

first hour, and by the second hour the amount of unlabelled protein appears to rise again which is

likely to be a sign of the labelled MBP undergoing hydrolysis (Figure 3.6).
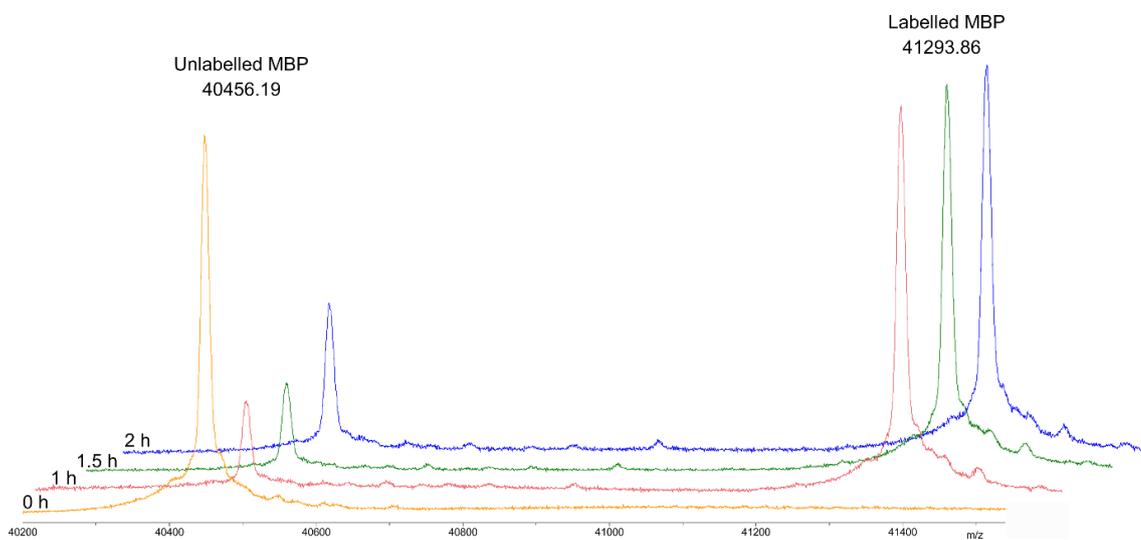

Figure 3.6: Sortase 7M mediated labelling of MBP with the BCN-depsipeptide followed by HRMS.

Subsequent labelling reactions were therefore based on these findings, and MBP-BCN was successfully produced by leaving the reaction for 1.5 hours at room temperature (Figure 3.7a). Most of the MBP became labelled with a small amount of MBP-BCN dimer forming at ~80 kDa (Figure 3.7b).

70

a)

b)
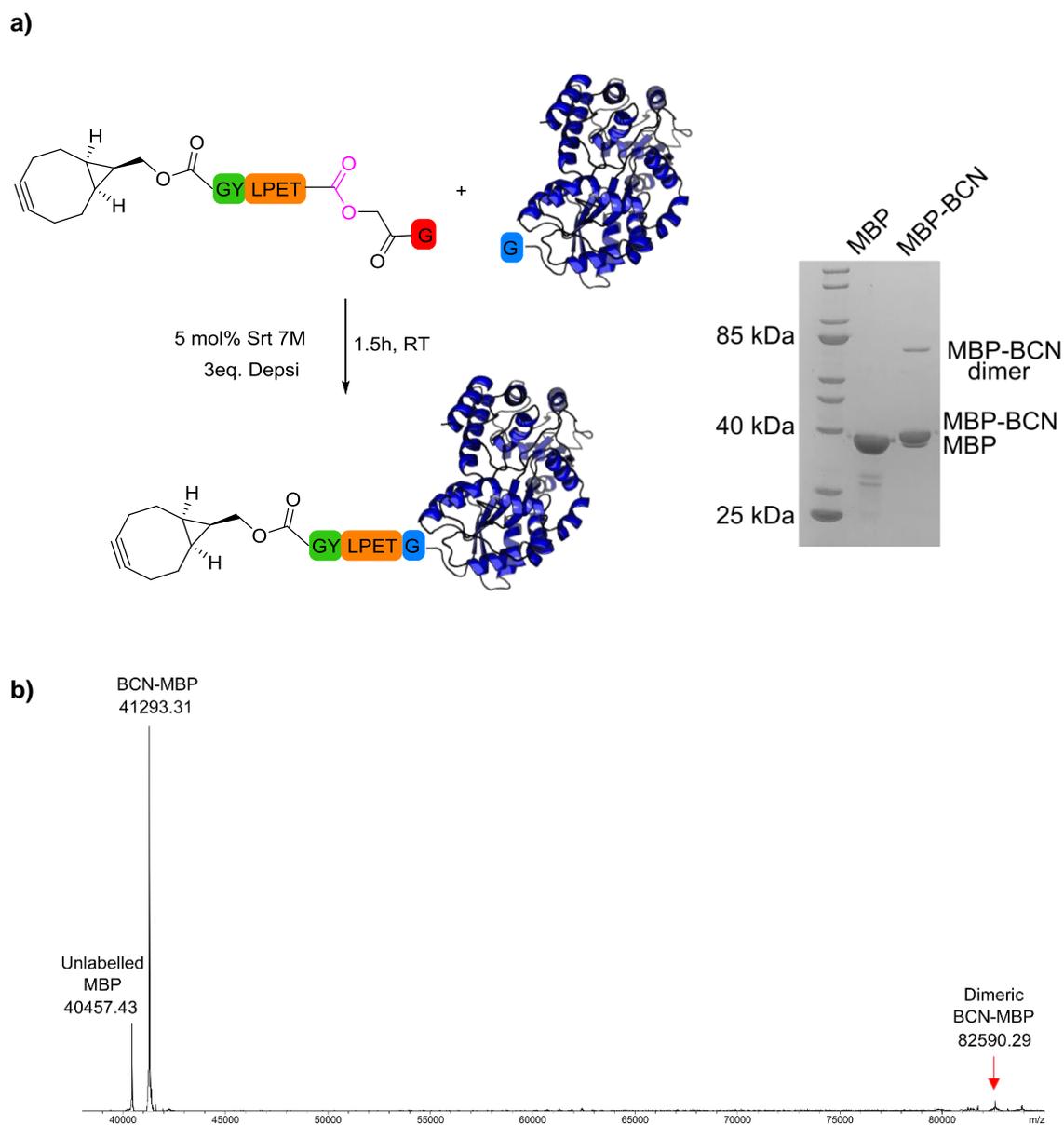
Figure 3.7: a) Schematic depicting the sortase 7M labelling of MBP with the BCN-depsipeptide substrate and an image of an SDS-PAGE gel showing a band shift along with the appearance of a dimeric species at ~80 kDa. This suggests the successful labelling of the protein using 3 equivalents of depsipeptide and 5 mol% Srt 7M within 1.5 hours at room temperature. b) HRMS trace of dimeric species found within the reaction mixtures after 1.5 hours.

3.2.6   Labelling with the oligonucleotide and product purification

The next step was to carry out the SPAAC reaction using the N_3-oligo, generally these reactions are incubated at room temperature for ~24 hours using a 1.2 equivalent excess of the N_3-oligo

concentration. The SDS-PAGE displayed in Figure 3.8 shows the successful "click" reaction, although the conversion is not quantitative. Experiments were incubated for a longer time period (up to 48 hours) and also at 37 °C to see if any improvements could be made, however the results appeared to remain the same.
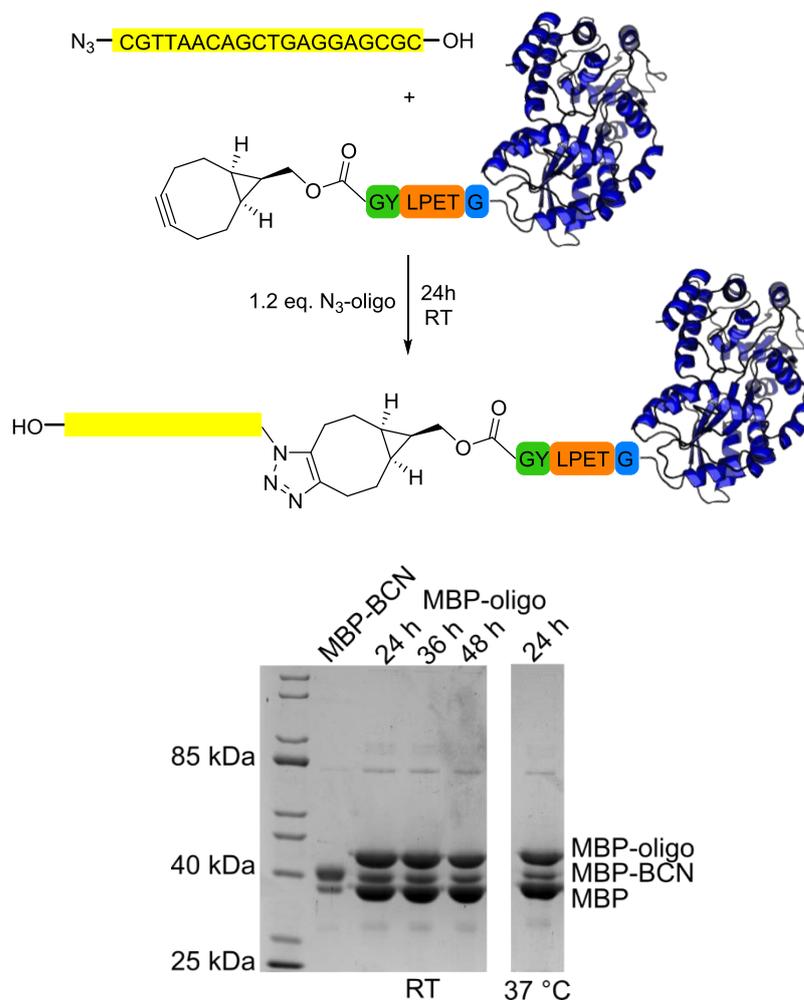
Figure 3.8: Schematic of MBP-BCN labelling with 1.2 equivalents of N₃-oligo, the SDS-PAGE gel shows the optimal results were found when the sample was incubated at room temperature for 24 hours.

Excess N₃-oligo in the product used for HS-AFM was removed using amylose affinity chromatography. Due to MBP's inherent affinity for amylose resin, the reaction mixture could be applied to the column, washed with buffer to remove the excess oligonucleotide before elution of the desired MBP-oligo and MBP-BCN mixture using 20% (*w/v*) glucose in phosphate-buffered

saline. To confirm that this had removed the N_3-oligo, the samples from the amylose column were mixed with an oligonucleotide complementary to the N_3-oligo to facilitate visualisation using ethidium bromide. The samples were analysed using a Tris-borate-EDTA (TBE) gel stained with EtBr. The flow through samples illustrate that the excess N_3-oligo did not stick to the resin, however the MBP-oligo is only present when the glucose buffer had been applied to the column. The absence of a band further down the gel for the elution sample confirmed the successful removal of the excess oligonucleotide (Figure 3.9).
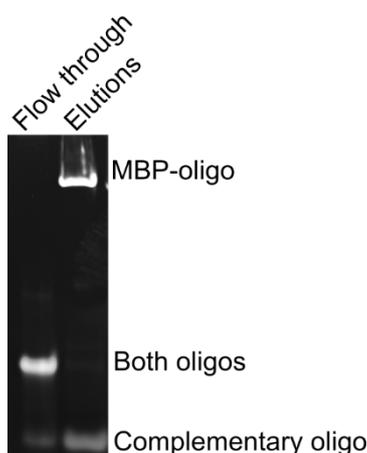
Figure 3.9: TBE gel showing the concentrated flow through and elution samples from an amylose column which have been mixed with another oligonucleotide, complementary to the N_3-oligo used to label the MBP to ensure visibility when staining using EtBr and imaged using UV.

Size-exclusion chromatography using a small Superdex® S75 increase column allowed a reasonable separation at a flow rate of 0.2 mL min^{-1} (Figure 3.10a). Analysis by SDS-PAGE suggested that MBP-oligo species were present in fractions B-E, with fraction F containing the mixture of MBP, MBP-BCN and a small amount of MBP-oligo (Figure 3.10b). The two peaks within the trace could be a result of the formation of an MBP-oligo dimer. This could be forming through the oligonucleotides annealing to each other due to the presence of two palindromic sequences; GTTAAC and CAGCTG, both of which have low T_m values of 16 and 20 °C respectively. To double check there was no excess oligo in the product, the samples were mixed with the complementary oligo and analysed using a Tris-borate-EDTA (TBE) gel. Subsequent

staining with EtBr confirmed the absence of any excess oligonucleotide present in the product (Figure 3.10c).

a)

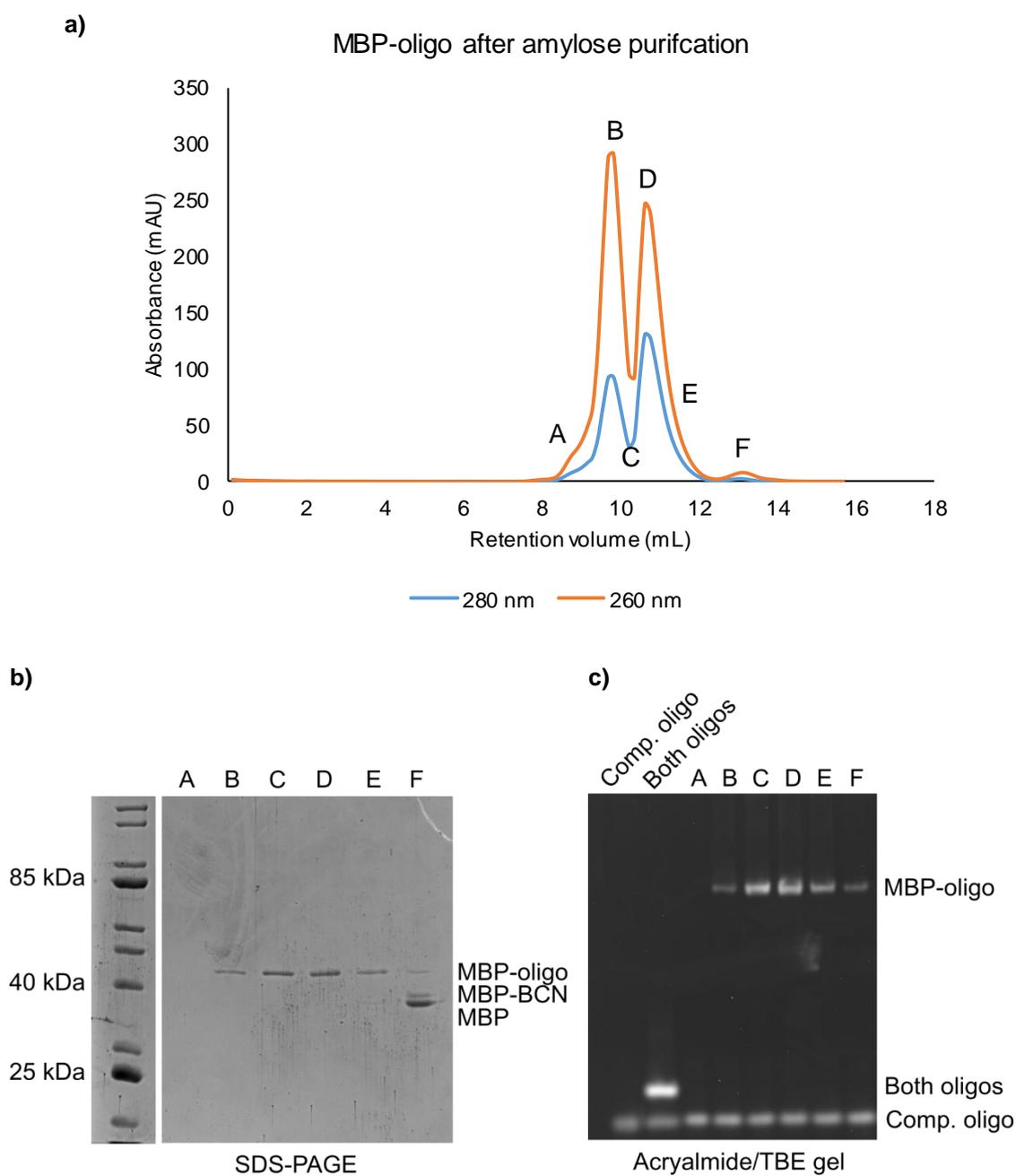**b)**   **c)**

Figure 3.10: a) Size-exclusion trace for MBP-oligo using a Superdex® S75 increase b) SDS-PAGE analysis of the labelled fractions A-F c) TBE gel stained with EtBr of SEC fractions mixed with the complementary oligonucleotide to ensure there isn't any N_3-oligo left in the samples.

3.2.7 HS-AFM experiments and results

Once the MBP-oligo had been purified, it was passed on to Andrew Lee in Bioelectronics. The following data detailed in this chapter was obtained by him. The DNA origami frame was formed using the M13mp18 single stranded scaffold, and the "stapling" oligonucleotides. They were mixed in a buffer containing Tris/magnesium acetate and EDTA before undergoing the folding process, consisting of a cooling gradient from 95 °C to 15 °C at -1 °C min^{-1}. Following this, the DNA scaffolds were purified from the excess oligonucleotides using a Sephacryl® S-400 spin column, before the MBP-oligo was introduced. This was done by mixing an excess of MBP-oligo with the DNA scaffold, and subjecting it to another cooling gradient of -0.5 °C min^{-1} from 30 °C to 4 °C and was seen to achieve >90% incorporation when assessed by AFM (Figure 3.11).
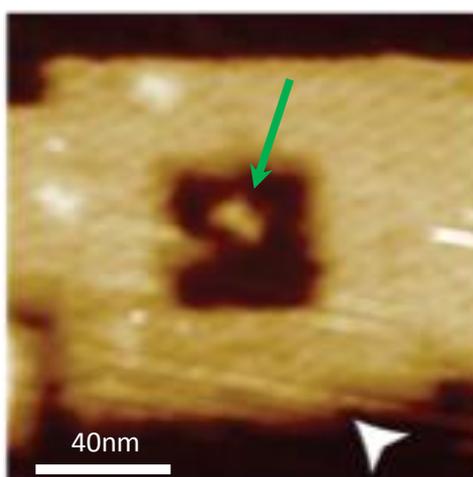
Figure 3.11: AFM image of the DNA origami frame with the MBP-oligo incorporated, placing MBP (green arrow) within the internal cavity of the frame.

The next stage in this process was to deposit the frames onto a muscovite mica surface: this consists of layers of an aluminium phyllosilicate lattice which are ionically bonded through potassium ions. The regular arrangement of the atoms and ions within lattice structures provides planes of weakness which allow the material to split along structural planes; in mineralogy this separation of the planes is called cleavage. When a mica surface is cleaved, the K$^+$ ions between the planes become highly mobile and can be readily exchanged with divalent cation species such

as Ni^{2+}. This results in a positive overcharge of the mica surface, enabling molecules holding a net negative charge, like DNA, to become deposited on to the surface.[136] This is generally achieved by freshly cleaving the mica and incubating it in a solution of 10 mM $NiCl_2$. The DNA is then deposited on the surface by incubating a solution containing the frames on the mica for 15 minutes, before being washed off with water and immersed in a Tris/magnesium acetate buffer.[137]

Once the DNA had been anchored to the surface, it was placed on the AFM platform and an elongated bubble with buffer was formed between the surface and the cantilever of ~200 µL in volume (Figure 3.12a). The sortase 7M was then injected into this bubble in a ratio of 95:5 of DNA to sortase, however this ratio is not absolute. In practise, the relative concentration of sortase will be much higher, and the reason for this is that although the approximate number of frames in the solution added to the mica surface is known, it is not possible to know exactly how many stick to the surface. This means the 95:5 ratio is based on the number of frames added to the surface and not the number which adhere to the surface, ultimately increasing the effective concentration of sortase to MBP. This is the possible reason for what appears to be extremely rapid cleavage of the MBP from the frame seen in Figure 3.12b. The MBP (shown by a green arrow) is joined by sortase 7M (red arrow) in the frame at 6 seconds, they are present together for between 19-21 seconds, before both leaving the frame completely. This illustrated that the sortase recognised the LPETG and cleaved the T-G bond holding the MBP to the oligo. As there is no other substrate within the system, this cleavage is the hydrolysis reaction discussed in Section 1.3.3.2. Once cleaved, the MBP and sortase become free from the surface which is why they are no longer visible by AFM, however, they are floating freely within the 200 µL buffer bubble. In order to remove them from the experimental platform completely the buffer would need to be constantly exchanging; for example, a flow system would be required.
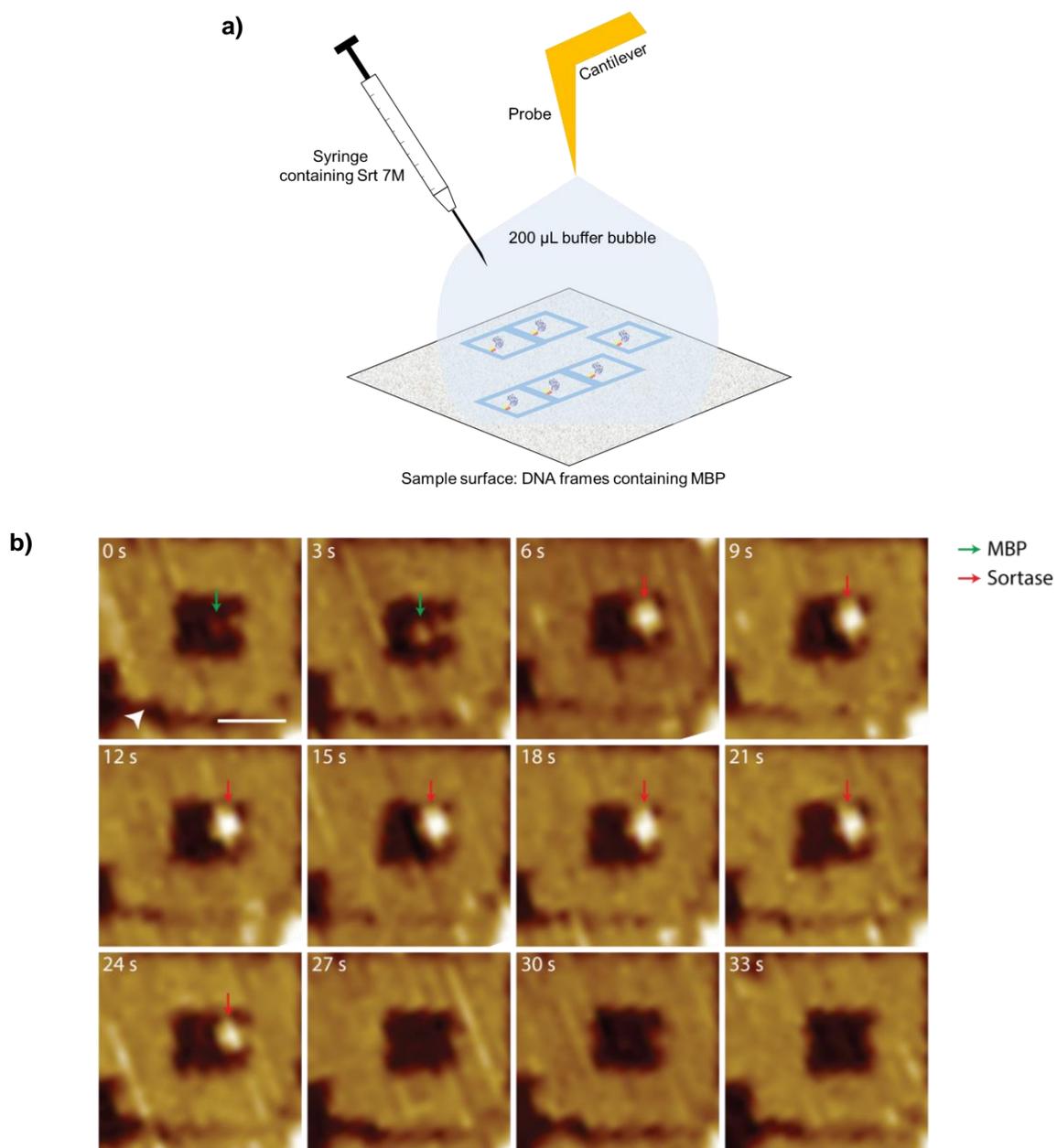
Figure 3.12: a) Schematic of practical approach to applying the sortase 7M to the DNA origami frames containing MBP by injecting into an elongated bubble between the tip of the probe and the surface. b) HS-AFM images of sortase 7M being injected into the sample and causing the apparent cleavage of the MBP from the frame.

3.2.8   Conclusions and future work

This work provided proof that a protein chemically liked to a short single stranded piece of DNA could be incorporated into a larger DNA scaffold using built in complementary anchor points. In addition to this, by using sortase to incorporate the SPAAC handle into the protein, the recognition sequence remained in the MBP-oligo fusion for further reaction while being monitored. The results of the MBP cleavage from the frame through the addition of sortase are very promising, showing that this could be a useful method with many potential applications.

Firstly, a flow system is currently being developed to allow the rapid addition of both the sortase and the second substrate to potentially allow the observation of sortase exchanging one protein for another. In order for the protein exchange to be visible, the MBP-fusion would be anchored to the frame and a much larger protein, for example MupB, containing an N-terminal glycine residue would be added as the second substrate. MupB is a protein involved in the resistance mechanism associated with the antibiotic Mupirocin, and is approximately ~120 kDa in size.[138] The *E.coli* strain required for the expression of this large and monomeric protein was provided by Jennifer Tomlinson (O'Neill group), from this the protein was expressed in lysogeny broth in the presence of trace elements and induced using IPTG. Following cell lysis, the protein was purified using nickel affinity chromatography and subsequent size-exclusion chromatography on a Superdex® S200 column with 10 mM Tris, 100 mM NaCl, 10 mM $MgCl_2$ and 1 mM DTT at pH 7.4. The His_6-tag was removed using TEV to leave the N-terminal glycine residue which underwent successful labelling using sortase 7M and the fluorescein depsipeptide (Figure 3.13). The efficiency of this labelling is hard to gauge as the large size causes problems with visualising band shifts on SDS-PAGE gels and is also difficult to analyse by mass spectrometry. However, it is very useful for this application as it is 3 times larger than MBP, which means any exchange in the proteins attached to the frame will be observed by HS-AFM.
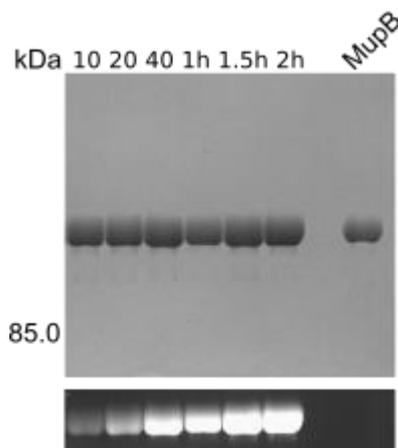
Figure 3.13: Labelling of 30 µM MupB using 60 µM FITC-depsipeptide and 1.5 µM sortase 7M at room temperature.

The proof of concept described in this chapter provides a method for attaching different proteins to the DNA origami frames and allowing the sortase reaction to be monitored. There are many different sortase species available which recognise different sequences e.g. eSrtA(2A-9) and eSrtA(4S-9) recognise LAXTG and LPXSG respectively.[89] In the future it could even be possible to decorate the frame with different proteins and observe protein cleavage or replacement using the different enzymes side-by-side on one DNA frame.

3.3   Labelling with a long double stranded piece of DNA for SAS studies

Small-angle scattering (SAS) techniques are often used in biochemistry to explore biomolecular structures and their dynamics within solution, providing a method for studying large biomolecules which can be difficult to crystallise. SAS involves a sample within solution being put in an X-ray (SAXS) or neutron beam (SANS) and allows them to be explored using buffers which provide more physiological environments, and unlike crystallography and cryo-EM, the molecules are free to undergo conformational changes which can be linked to protein function. This ability to move and rotate randomly within the solution means when the sample is hit by the X-ray or neutron beam it is impossible to know what orientation the biomolecule is in when the data is

acquired. The problem with this is that it means calculations based on the data obtained from SAS experiments must be averaged over all possible orientations the biomolecule could have been in which can provide information about the general shape, but detail obtained is limited with maximum resolution rarely exceeding 10 Å.[139]

To improve the quality of data obtained through SAS techniques, the aim of this piece of work was to reduce the number of degrees of freedom that the biomolecule has within the solution. This alignment of biomolecules was hoped to be achieved by N-terminally labelling a protein with an aligning tail group and subjecting the sample to shear forces in a flow system. By doing this, the goal was to improve the accuracy of the data obtained and provide a more detailed understanding about the structure and dynamics of proteins studied in solution by inducing an anisotropic effect (Figure 3.14).
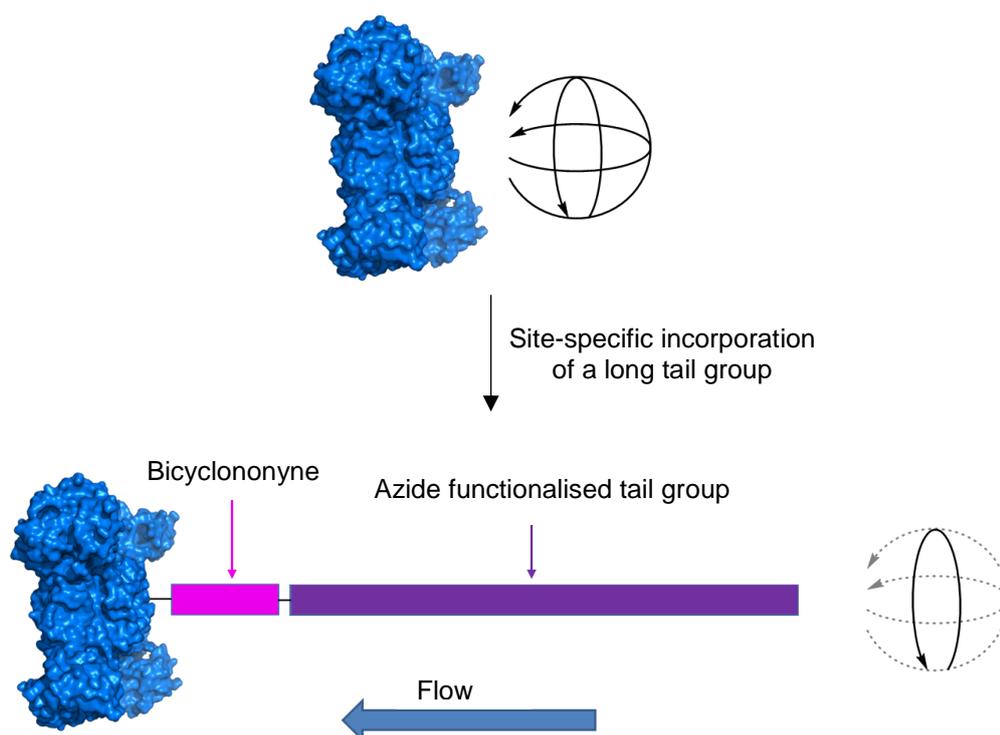
Figure 3.14: Using N-terminal labelling to introduce a BCN to the protein of interest, a long azide-terminated tail group will be incorporated using SPAAC. The aim of this it to reduce the number of degrees of freedom the protein has within solution and induce an anisotropic effect under flow conditions to improve the quality of data obtained from SAS experiments.

3.3.1 How small-angle scattering works

SAXS and SANS provide different information about the sample being studied. SAXS involves the elastic scattering of X-rays when they interact with the electrons within the sample. The detector picks up the change in direction of the X-ray which is defined as the scattering vector, s, and results in a scattering profile for the sample (Figure 3.15a). In addition to the sample scatter profile, the buffer alone must also be analysed, this then allows the buffer measurements to be subtracted from the sample measurements to effectively leave the scattering profile for the biomolecule of interest (Figure 3.15b). Structural information can then be extracted using a number of different computational tools which go through multiple algorithms to provide a "SAXS envelope" containing information about the molecular mass and shape of the molecule being studied.[140–146]
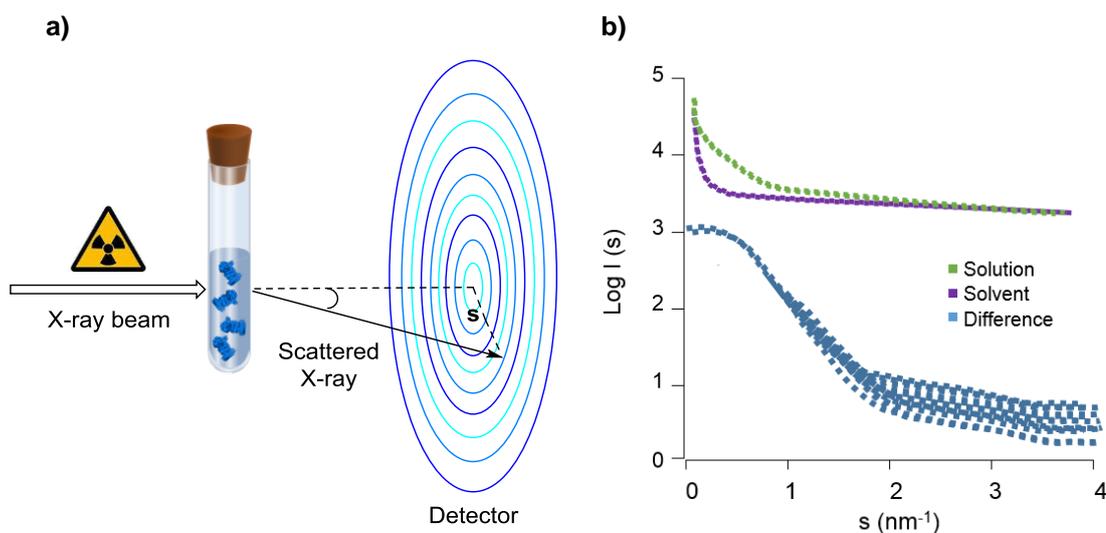
Figure 3.15: a) Basic set up for a SAXS experiment, the sample is placed between the X-ray source and a detector, when the X-ray beam interacts with the electrons within the sample. The detector picks up the scattered X-ray and the scattering vector, s, is obtained. b) The intensity of the scattered beam, I(s), is plotted against the scattering vector for the sample and the buffer alone. The buffer scatter is subtracted from that of the sample in buffer to leave the scattering profile for the protein alone.[147]

The method of processing SAXS data is the same for SANS, however the physical mechanism behind obtaining the scattering data differs. Instead of interacting with the electrons within the

sample, the neutrons interact with the atomic nuclei. This makes it sensitive to isotopic content, for example, hydrogen and deuterium have very different scattering lengths.[148,149] This characteristic is extremely useful as it allows contrast variation experiments whereby changing the H_2O and D_2O content within the surrounding buffer of the sample allows parts of the biomolecular system being studied to be "invisible" to the detector. This is because in biological systems, hydrogen and deuterium are readily interchangeable, having little to no effect on the structure, but dramatically changing the scattering profile in SANS. At certain ratios of H_2O and D_2O the scatter of the molecule will be equal to that of the surrounding buffer, this means that when the solvent pattern is subtracted from that of the system, the biomolecule will be invisible (Figure 3.16). The points where this happens are where the red line for water crosses the other lines on the graph and they are called match points (indicted by an arrow). For example, the match point for protein is ~40% D_2O in H_2O and the match point for DNA is ~70% D_2O in H_2O.[150]
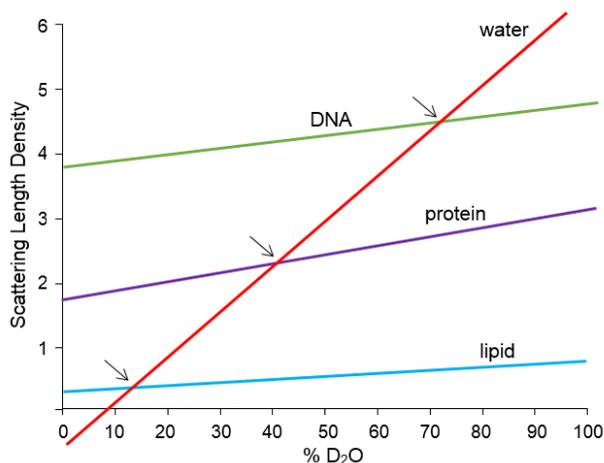
Figure 3.16: Graph to show the scattering lengths of several biomolecules as a function of D_2O concentration, arrows point out the match points. Adapted from Jacrot.[150]

SAXS and SANS have been used side-by-side as complementary techniques to allow specific parts of more complex biological samples to be studied. This has been seen to be particularly useful for membrane proteins where they are studied in deuterated nanodiscs to mimic the membrane and see them in a more native-like environment. Using the contrast variation

experiments it is possible to make the nanodiscs "invisible" in SANS, so only the protein scattering profile is visible. SAXS experiments are used in conjunction to analyse the overall structure of the nanodisc-protein complex as everything is visible at the same time and it allows any changes in the structure to be seen.[151]

3.3.2   The tail group

Since the group attached to the protein needed to be able to align the proteins under flow conditions when being subjected to shear forces, the tail needed to be very long in comparison to the dimensions of the protein being investigated. In addition to this, the tail needed to be able to remain strong and rigid in a rod-like manner when in the physiological buffer environment. Initially polymers were considered for this purpose, however it is challenging to control the polydispersity to ensure all tails are the same length. It was also estimated that the shear forces required to ensure they were not in a partial globular state would be too great for the protein to remain intact. Instead, it was decided that double-stranded DNA would be the best approach. DNA is an inherently strong structure held together with many hydrogen bonding interactions and is less likely to lose a rod-like shape when introduced to physiological buffers. By using DNA, it could also be used in SANS experiments and made "invisible" using the contrast variation experiments (Figure 3.17) allowing both SAS techniques to be used side-by-side to gain even more insight into the structure and function of the protein of interest. Like the HS-AFM experiments discussed earlier in this chapter, the azide-oligonucleotide was instrumental for the way in which the DNA tail attaches to the protein through SPAAC.
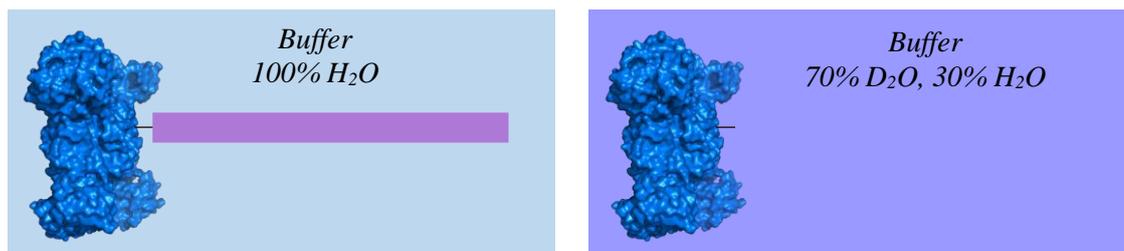
Figure 3.17: Using the match point information found within the graph in the previous figure, it is possible to adjust the ratio of D_2O and H_2O in the surrounding buffer of a system to allow parts of it to become effectively invisible. In this case, this is because the scatter from the buffer alone will be equal to the scatter for the DNA but not the protein, so when the sample-buffer subtraction is carried out, only the scatter of the protein will be obtained.

Since this approach has not been attempted before, the ideal length of the piece of DNA was unknown, however the ratio of protein:DNA size needed to be significantly weighted towards the tail group. Due to this, a 1 kb fragment with an approximate length of 320 nm was initially selected. MBP is approximately 5 nm across, therefore this would give a ratio of ~ 1:60 nm.

A major requirement of this piece of work was to synthesise very large quantities of DNA for attaching to the protein itself. Two strategies were explored, firstly, an enzymatic approach using a combination of restriction enzymes and nicking endonucleases to create a tail group with an overhang complementary to the N_3-oligo. The initial plan was to link the oligonucleotide to the BCN-functionalised protein and subsequently mix it with the long strand of DNA, allowing the oligonucleotide to stick to the complementary "sticky" end. Generation of the fragment required mutagenesis of a plasmid to contain specific restriction sites to allow multiple enzymes to cut the DNA in the desired locations in an efficient manner to allow high quantities of the overhanging DNA to be produced (Figure 3.18a). The second approach was to use the N_3-oligo as a primer in a polymerase chain reaction (PCR) leading to amplification of a desired segment of DNA giving an azide-terminated DNA product (Figure 3.18b).

Azide oligonucleotide:

N$_3$— CGTTAACAGCTGAGGAGCGC —OH
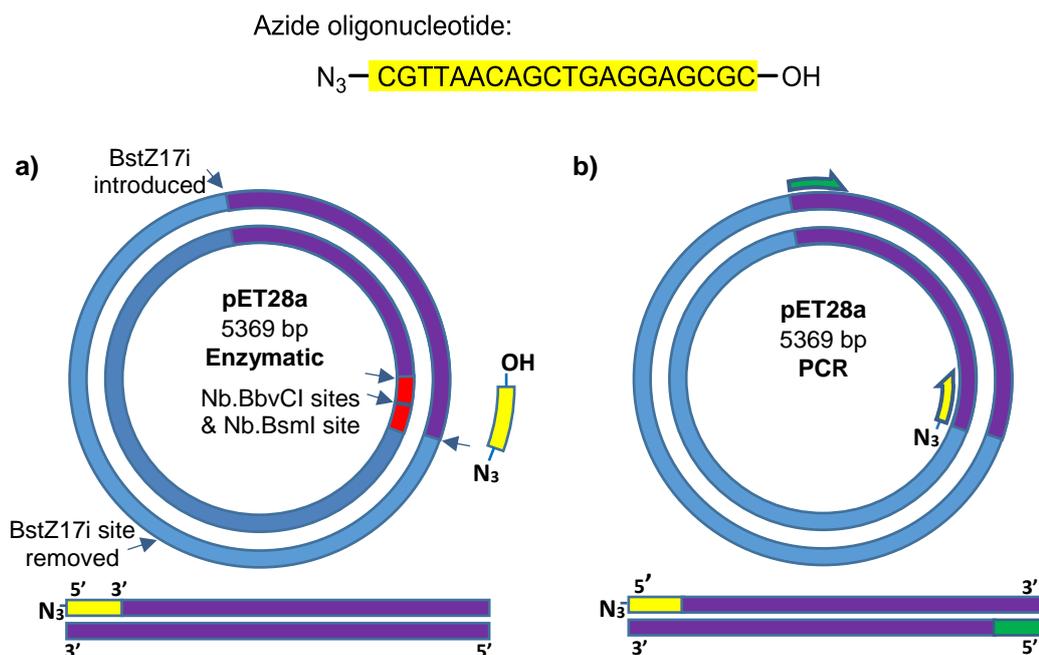
Figure 3.18: The two separate routes to be investigated for making 1 kb of azide-functionalised DNA. a) The enzymatic route required the use of nicking endonucleases and restriction enzymes to cut a whole plasmid into the 1 kb strand with a 20 bp overhang for the small N$_3$-oligo to anneal to. b) The PCR route will use the N$_3$-oligo in the reaction as the reverse primer, allowing it to be exponentially incorporated into the 1 kb piece of DNA automatically.

3.3.2.1 Enzymatic approach

For the plasmid to be cut in the specific locations using the restriction enzymes and nicking endonucleases, the plasmid chosen (pET28a) was mutated to contain the specific recognition sequences for the enzymes. The restriction enzyme BstZ17i was chosen for cutting the blunt end of the tail piece, as this prevented the need to use the more expensive PsiI site by just changing two residues, however another BstZ17i site also had to be removed to prevent cutting at another site of the plasmid. The nicking endonucleases chosen were Nb.BbvCI and Nb.BsmI for creating the overhanging "sticky" end. Two Nb.BbvCI sites were introduced in order to create a small fragment which would allow the overhanging section to pull apart more easily. Again, the specific sequences required for these enzymes were inserted approximately 1 kb downstream of the

BstZ17i site. All of these required recognition sequences were introduced using sequential Quikchange site-directed mutagenesis.[152]

Plasmid DNA was then produced on large scale before identification of the optimal conditions for restriction digestion to form the desired strand. Each enzyme was used at a concentration dependent on the number of sites to be cut within the plasmid and their reported activity. There was just one site for BstZ17i, two for Nb.BbvCI and three for Nb.BsmI (there is one in the area of the plasmid of interest and two more further around), meaning the enzymes were always used in a ratio of 1:2:3 units. Following digestion, the DNA appeared to be staying in the well of the agarose gel. This was likely to be the enzyme remaining stuck to the DNA after the site has been cut, so to ensure this was no longer a problem, a 10 minute deactivation step was included at the end at 80 °C. This was carried out for the nicking endonucleases separately to ensure the enzymes were both individually cutting the plasmid. The addition of the final 80 °C step proved to be successful and the DNA moved out of the well and began running as it should, showing three bands at 1 kb, 4.5 kb and 5.5 kb corresponding to the expected sizes for the digested DNA and also some uncut plasmid (Figure 3.19).
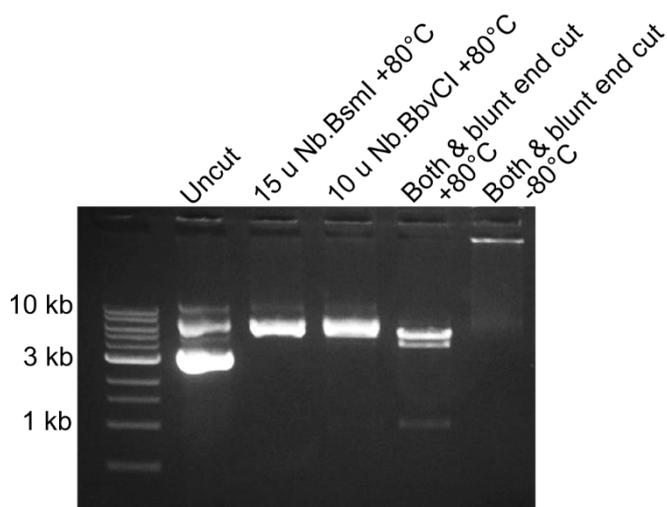
Figure 3.19: Agarose gel of digested DNA reactions using 400 ng plasmid DNA in a total volume of 20 μL, heated at 65 °C for 1h followed by 37 °C for 1h. Nicking enzymes were used in the following concentrations; 15 units Nb.BsmI 10 units Nb.BbvCI. A deactivation step of 80 °C for 10 minutes was used for all digested samples apart from the final well, which remained at the top of the gel, indicating the enzymes were still attached to the digested plasmid.

Following this, the concentration of the nicking endonucleases were doubled which caused the uncut plasmid band to become much fainter (Figure 3.20a). The temperature of the final step was increased to 95 °C to see if the uncut plasmid band could in fact be cut plasmid which is failing to fall apart. This successfully led to the appearance of only two bands, the larger 5.5 kb piece was no longer visible, leaving only the 4.5 kb and very faint 1 kb bands (Figure 3.20b). To check if this was the reason for three bands with half the amounts of enzyme the reactions were tested again however the larger band reappeared if less than 15 units Nb.BsmI, 10 units Nb.BbvCI and 2.5 units of BstZ17i were used per 200 ng DNA to be digested. These high quantities of enzyme required to fully digest the plasmid into the desired 1 kb overhanging strand made this approach unfavourable for continued exploration. The enzymes were much less efficient than expected and a large amount of plasmid DNA and enzymes would be needed for the scale required in SAS experiments therefore work turned towards the development of a PCR method whereby the 1 kb strand would be amplified exponentially.
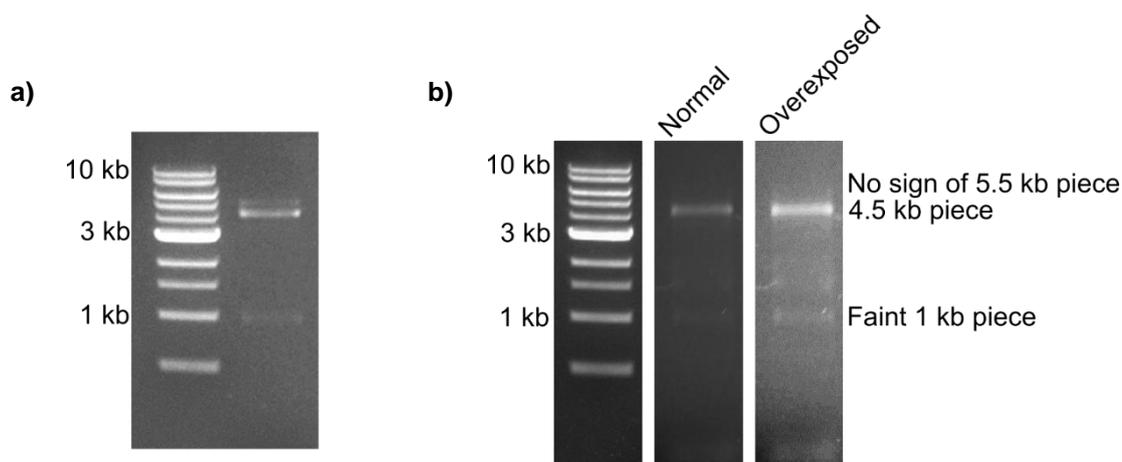
*Figure 3.20: Digestion of 200 ng DNA using 2.5 units BstZ17i, 15 units Nb.BsmI and 10 units of Nb.BbvCI, in a total volume of 30 μL heated at, 65 °C for 1h and 37 °C for 1h. a) Still signs of uncut plasmid present in the sample after a final 80 °C heating step for 25 minutes. b) A final heating step of 95 °C for 2 minutes showed the plasmid to be fully digested forming the desired 1 kb DNA alongside a 4.5 kb piece.*

3.3.2.2   Polymerase chain reaction approach

As an alternative to the enzymatic method, PCR was investigated for the large-scale synthesis of the azide-functionalised piece of DNA. PCR involves the cycling of temperatures to firstly allow the double stranded DNA to pull apart into single strands (denaturing step, 95 °C). The temperature is then changed to allow specially designed primers to bind to the DNA at their specific, complementary region (annealing step, temperature dependent on the primers being used). Then at 72 °C, the nucleotide building blocks in the form of deoxyribose nucleoside triphosphates (dNTPs) along with the polymerase enzyme build the complementary strand along from the primer location from 5' to 3' (extension step). These steps are repeated for 30-35 cycles, exponentially increasing the amount of the desired piece of DNA, until the primers and dNTPs have been consumed.

In this work, the N_3-oligo was used as the reverse primer in the PCR mixture, this had a calculated T_m of 62.5 °C. A forward primer was designed to bind approximately 1 kb upstream of the N_3-oligo binding site with a similar T_m value. Using this method every strand of DNA amplified should be functionalised with the azide group and be equal in length.

The polymerase chosen at first to carry out these reactions was Pfu as there was a large in-house stock of the enzyme. The reactions were based on the guidelines from Roche for the concentrations of the reaction components and the thermocycling program.[153] Initial optimisation experiments were done using a reverse primer identical to the N_3-oligo but not functionalised. Using this standard primer, the maximum number of cycles was found to be 30, along with the ideal Mg^{2+} concentration within the enzyme buffer, which appeared to be 15 mM (Figure 3.21).
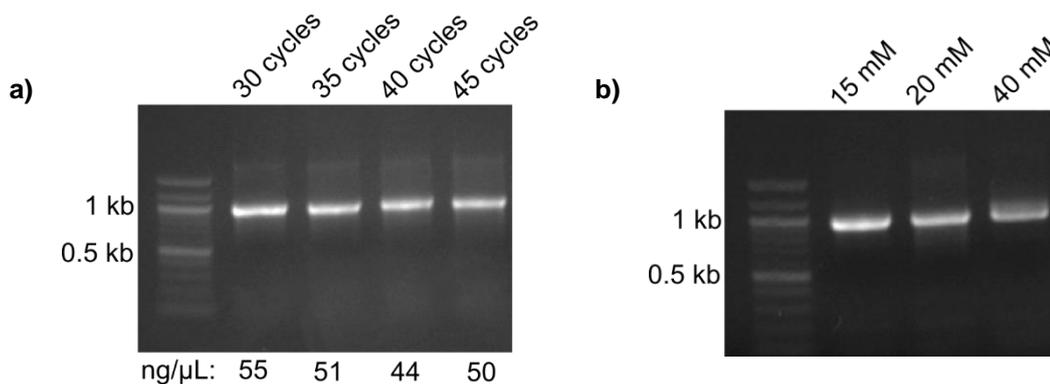
Figure 3.21: Analysis of the optimal number of cycles and concentration of Mg^{2+} in the PCR reaction buffer using 50 ng template DNA and 50 µL total volume. a) These results showed that following PCR clean-ups, the reaction using 30 cycles yielded the most DNA. b) The band intensities on the different concentrations of $MgSO_4$ indicate 15 mM to yield the most 1 kb DNA.

After this, the annealing temperature was investigated as the only information on this so far was the calculated T_m, however in practise there could be a more effective temperature for the reaction. With the prediction being 62.5 °C, a gradient of annealing temperatures were investigated from 59-64 °C. This showed that the original temperature being used (62 °C) was yielding much less DNA than if the annealing temperature was set to 60 °C. To double check this, the 62 °C and 60 °C samples were taken through a PCR clean-up step with the 60 °C sample yielding more than double the amount of DNA (Figure 3.22a). Now the PCR had been optimised, the N_3-oligo was then tested as the reverse primer and successfully amplified the desired DNA, visible as one clean band on the agarose gel (Figure 3.22b). The scale of the individual reactions was investigated comparing 100 µL and 200 µL total volumes. Interestingly following a PCR clean-up, the 100 µL reaction resulted in 4.7 µg DNA whereas the 200 µL reaction yield was only 1.8 µg DNA. This reduced efficiency of the PCR reaction is likely to be due to the upper part of the tube not being in direct contact with the heating block and in fact being closer to the heated lid which remains at a constant 95 °C throughout the thermocycling program. For the 100 µL reaction, all of the liquid remains within the heating block, therefore this was used as the maximum volume for all subsequent reactions.
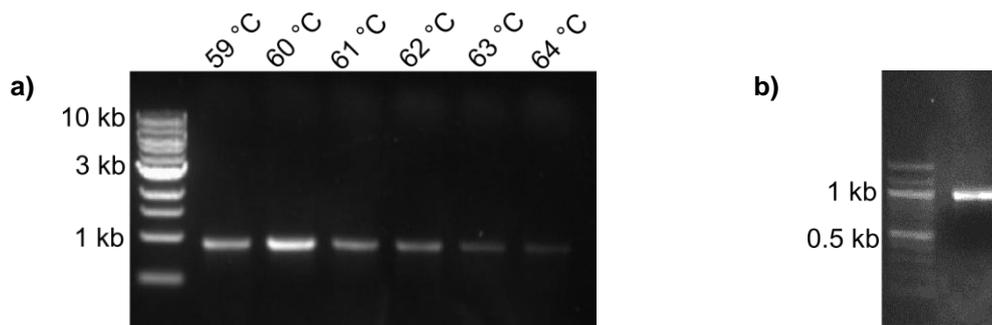
Figure 3.22: a) Temperature gradient carried out on the reactions (6 × 50 ng template, 50 µL reactions) showed 60 °C to yield the most DNA, this was also analysed by carrying out PCR clean-ups with 60 °C allowing 88 ng/µL to be isolated while 62 °C allowed only 36 ng/µL to be purified. b) Successful PCR using the N_3-oligo as the reverse primer using 50 ng template DNA and 50 µL total volume.

As the amount of DNA required from this for labelling the protein is very high, KOD Hot Start polymerase was chosen for further experiments as it is a much faster and more efficient polymerase and the Hot Start capabilities mean non-specific amplification can be avoided as the polymerase is inactive at lower temperatures.[154] Using KOD also meant that the amount of primer and dNTPs required in each reaction was significantly lower, and in addition to this, the length of time required at each temperature was reduced within the cycling part of the program, especially for the extension section which goes from 120 s using Pfu to 15 s using KOD. When compared directly to Pfu it was visible that more DNA could be obtained from the same amount of template, less enzyme and in a much shorter amount of time (Figure 3.23).

a)

	KOD	Pfu
Template DNA	50 ng	50 ng
Fwd primer	0.3 µM	0.6 µM
Rev primer	0.3 µM	0.6 µM
dNTPs	0.2 mM ea.	0.2 mM ea.
Enzyme	1 unit	2.5 units
Buffers	8 µL	5 µL
H_2O	Up to 50 µL	Up to 50 µL

b)

	KOD	Pfu
Initial denaturation	120s	120s
Denature	20 s	15 s
Anneal	15 s	30 s
Extension	15 s	60 s
Final Extension	300 s	300 s
Total time (30 cycles)	32 min	60 min

Figure 3.23: a) Table depicting the standard protocols used for KOD and Pfu polymerases in PCR reactions to highlight the difference in concentrations of primers, dNTPs and enzyme required for 50 ng of template DNA. b) Time scales for each part of the PCR reaction required by both KOD and Pfu polymerases including an estimated total time required to carry out 30 cycles.[153,155]

The only optimisation experiment with KOD was to see if it would be possible to use less of the enzyme but carry out more cycles. This was tested by directly comparing a reaction mixture with the recommended amount of KOD for 30 cycles against half of the recommended amount of KOD for 35 cycles in 100 µL reactions. By looking at the gel it is possible to see the band is brighter for the half KOD, 35 cycles, but in addition to this, a PCR clean-up was performed on the reaction mixture and compared the yields (Figure 3.24). This confirmed that half the amount of recommended KOD could be used for each reaction by using 35 cycles as this gave 5.3 µg of DNA *versus* 5.0 µg using the recommended amount of KOD and 30 cycles. Once a number of reactions had been successfully scaled up carrying out 48×100 µL reactions at a time, a purification method was next to be investigated.
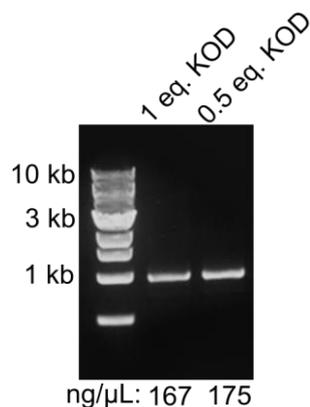
Figure 3.24: Comparison of using 1 equivalent of KOD for 30 cycles and 0.5 equivalents for 35 cycles. The band intensities appear very similar and the yield was confirmed to be almost identical following purification by PCR clean-up.

3.3.2.3 Purification of N_3-DNA

The 1 kb DNA needed to be purified from the reaction mixture containing the excess dNTPs, primers and enzyme. For this, size-exclusion chromatography was investigated as a potential method of purifying the large piece of DNA from these much smaller contaminants. To test this out, initially the DNA was applied to a Superose® 6 10/300 column which had been thoroughly washed with 0.1 M NaOH, and autoclaved water to ensure no DNase from previous protein samples was present. The sample was initially eluted in autoclaved water as this would allow concentration of the DNA after freeze-drying. The column allowed some separation of DNA species, however when the fractions were analysed by agarose gel there appeared to be some ~500 bp pieces of DNA. This was likely to be due to the use of water to run the column, the lack of salts or buffer system could have been the cause of the DNA falling apart in this manner (Figure 3.25).
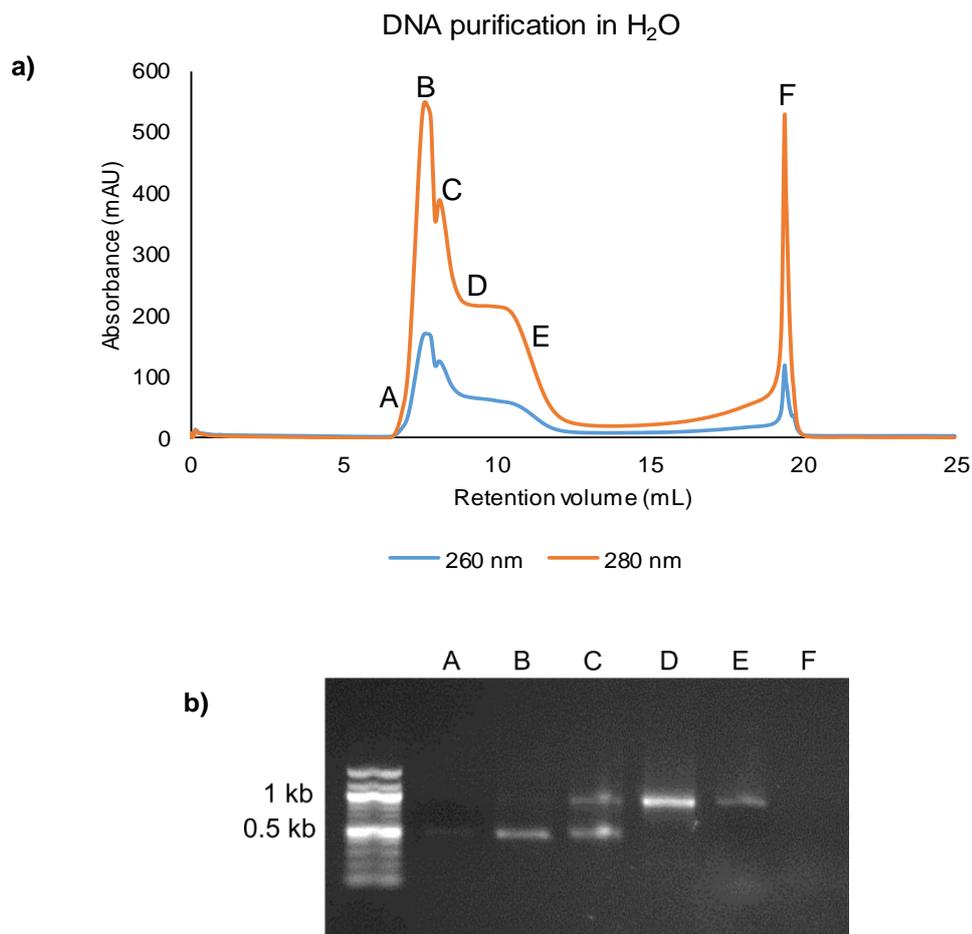
Figure 3.25: a) Size-exclusion trace for the PCR reaction mixture shows a level of resolution between the DNA species using a Superose® 6 10/300 column. b) Agarose gel showing the different DNA sizes present in the fractions, however there should only be 1 kb pieces in there. The appearance of the 0.5 kb piece of DNA is thought to be due to the lack of buffer system used causing the DNA to fall apart.

The subsequent study used 10 mM Tris HCl (pH 7.4) which is often used as a DNA storage buffer so should be a suitable buffer system for the DNA during chromatography. As can be seen in the SEC trace, the separation was much better and no single stranded DNA was present in the fractions when analysed by agarose gel (Figure 3.26). However, there were some large peaks in the trace which didn't show up as DNA species on the agarose gel electrophoresis. To investigate this further, 20% acrylamide-TBE gel analysis allowed very small fragments of DNA to be visualised. Using this approach it was possible to see that these peaks were likely to be the primers, confirming that they were able to be separated from the desired DNA using this method (Figure 3.26c).
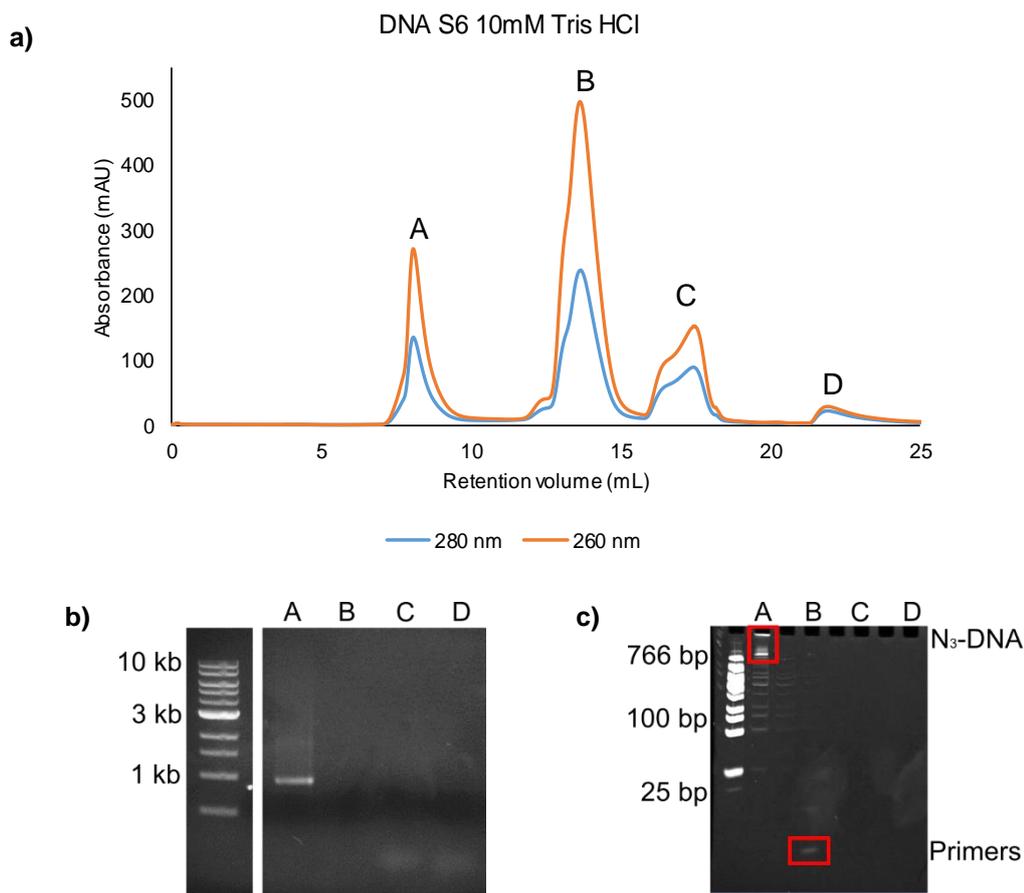
Figure 3.26: a) SEC trace of the purification of the PCR reaction mixture using a using a Superose® 6 10/300 column with 10 mM Tris HCl buffer b) Agarose gel showing only the presence of the 1 kb piece of DNA in the first peak. c) 20% acrylamide-TBE gel shows the 1 kb piece of DNA stuck at the top of the gel and in the well while fraction B appears to contain very small species when stained with EtBr, likely to be the excess primers.

In order to scale up the purification alongside the number of PCR reactions necessary, the Superdex® 200 26/60 was tested as much larger volumes can be injected onto the column. The resulting analysis by agarose gel just showed one single DNA species at 1 kb in the first peak on the trace, and the 20% TBE-acrylamide gel showed that in the second peak there was a very small species as had been seen for the S6 column, again likely to be primers (Figure 3.27a-c). Interestingly, nothing appeared to be present on the gels for the last two peaks on the trace, however by analysing the conductivity it was possible to see two corresponding peaks which indicated that these could be due to the presence of buffer salts and remaining dNTPs (Figure 3.27d). After the SEC run, the fractions containing the N_3-DNA needed to be concentrated.

Initially the plan was to freeze-dry the fractions as it was going to be run in water. However, when using the Tris HCl buffer this was no longer an option, instead the fractions containing the desired DNA were combined and concentrated in a 100 kDa MWCO Amicon centrifugal concentrator which was successful with no DNA species found in the flow through (Figure 3.27e).
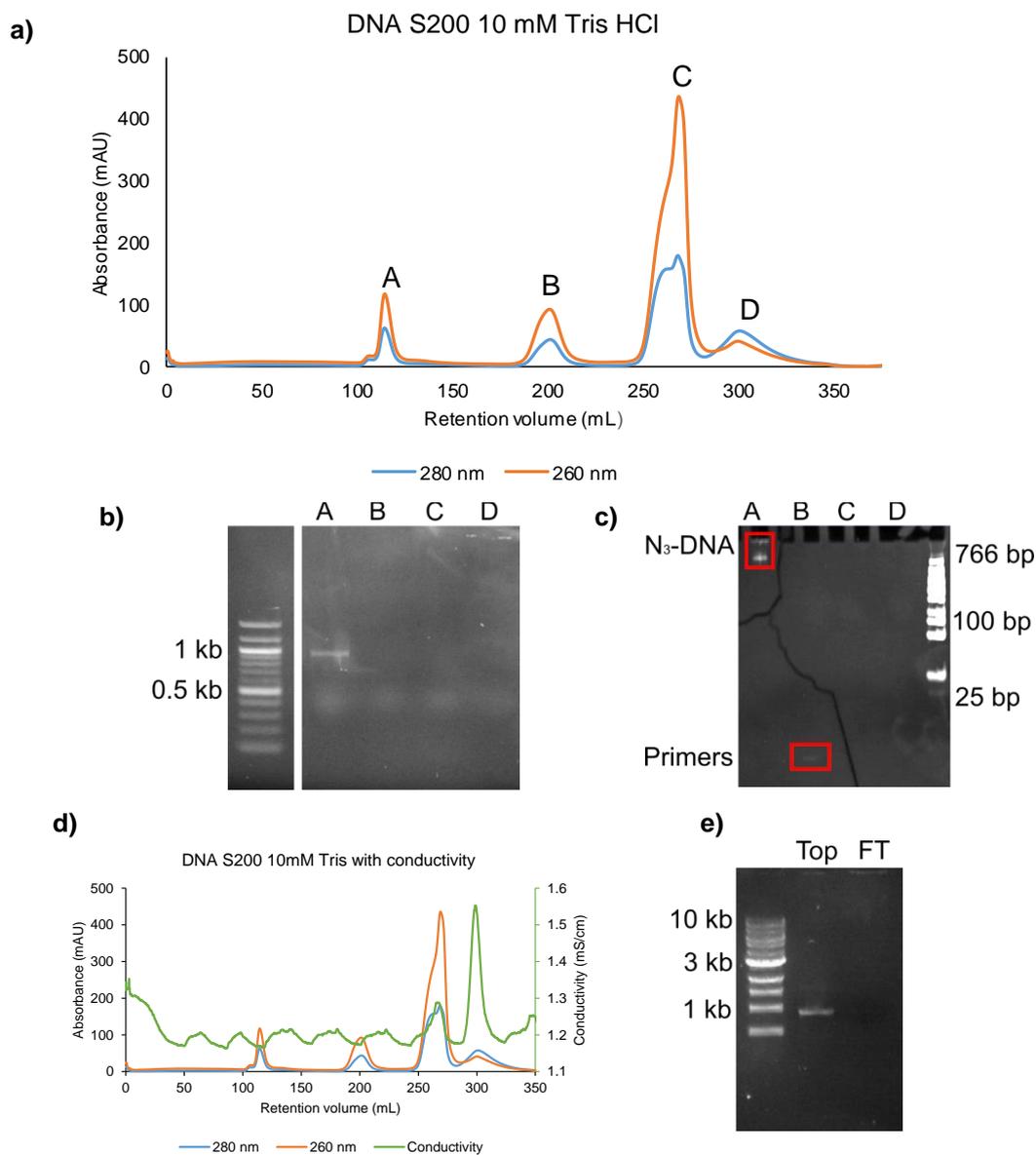

Figure 3.27: a) SEC trace for the larger scale PCR reaction mixture purified using the Superdex® 200 26/60 b) Agarose gel showing only the 1 kb piece of DNA c) 20% acrylamide-TBE gel allowing the small primers found in fraction B to be visualised using EtBr. d) SEC trace showing the changes in conductivity with the absorbance at A_{280} and A_{260} indicating the fractions containing dNTPs and buffer salts. e) Agarose gel showing the successful concentration of SEC purified DNA using a spin concentrator with no bands present for the flow through (FT) sample.

Although the SEC method worked well and allowed the N_3-DNA to be isolated and purified from the rest of the PCR reaction mixture, it was very time consuming and there was a high probability that some of the sample injected was being lost in the loop and syringe. The main problem with this method was the requirement to ensure the column and system was thoroughly cleaned before use to remove any trace amounts of DNase which is routinely used for protein purification. This involved removal of the 20% ethanol column storage buffer through washing with water (1 column volume, CV), followed by a wash step with autoclaved 0.1 M NaOH (1 CV) and autoclaved water (2 CV) and then equilibration with the Tris HCl buffer (1 CV). Due to the column pressure limit, the maximum flow rate able to be used was 2.5 mL min^{-1}, meaning the preparation of the column alone took approximately 11 hours, therefore in order to use this method long time periods needed to be booked on the shared piece of equipment. For small scale PCR reactions Qiagen PCR purification kits are commonly used, these are very quick to carry out and have less points of loss for the DNA. The PCR mixture is flowed through a small anion-exchange column using a binding buffer to allow the DNA to bind to the resin while unwanted substituents of the reaction mixture are washed out prior to the DNA being eluted. The Qiagen kits only allow up to 50 µg of DNA to be purified, however Zymo Research produce a large scale version which claims to allow the isolation of up to 500 µg DNA, a much better scale for this work. By concentrating 16 mL of PCR reaction mixture samples from 160×100 µL reactions in a spin concentrator before application to the column using the binding buffer. Following the protocol provided with the product, purified DNA was able to be collected using water, this could subsequently be freeze-dried and stored as a solid for future use (Figure 3.28). Over several months, 69 mL of PCR reaction mixture was purified using this method, resulting in ~1.49 mg of azide functionalised 1 kb DNA to take on to use as a stock for protein labelling.
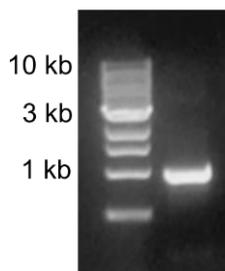
Figure 3.28: Product of the large-scale PCR clean-up of 16 mL of PCR reaction mixture allowing 402 µg of DNA to be isolated.

3.3.2.4 Comparison of both approaches for large scale DNA production

Using this PCR method large amounts of azide-functionalised double stranded DNA of approximately 1 kb in length were able to be obtained. This made it possible to achieve concentrations high enough to attempt reactions with bicyclononyne labelled proteins, for example MBP-BCN. Throughout this process it was essential to continually renew stocks of primers and plasmid every few weeks, since likely contamination with DNase meant the PCR reaction would stop working.

In order to fully understand the financial aspect to both methods, calculations were done to find out how much it would cost to make 1 mg of the 1 kb piece of DNA, all prices quoted below were obtained from Science Warehouse in January 2018. Firstly, for the enzymatic reaction, to make 1 mg of 1 kb of DNA, ~5.4 mg of plasmid DNA would have to be digested by the enzymes, this would require plasmid purification using a Qiagen GIGA prep kit (approx. £60 each) which allow up to 10 mg of plasmid DNA to be isolated. Following this the plasmid would need to be digested using the three different enzymes. In the small scale test reactions discussed in this section, 2.5 units of BstZ17i, 10 units of Nb.BbvCI and 15 units of Nb.BsmI were required to digest 200 ng of plasmid DNA. Therefore, to cut 5.4 mg of plasmid DNA to make 1 mg of 1 kb DNA would mean scaling this up by a factor of 27×10^3. The number of enzyme units required for this along with the prices associated with this route can be seen in Table 3.1 with each enzyme costing

~£0.04 per unit (based on buying a 5000 unit pack). After the plasmid had been digested, the large-scale PCR clean up would allow the isolation of up to 0.5 mg purified DNA.

	Units to cut 200 ng plasmid	Units to cut 5.4 mg plasmid	Cost to produce 1 mg 1 kb DNA
BstZ17i	2.5 units	67,500 units	£27,000
Nb.BbvCI	10 units	270,000 units	£10,800
Nb.BsmI	15 units	405,000 units	£16,200
1 × Giga prep kit			£60
2 × PCR clean up			£14
		TOTAL	**£29,774**

Table 3.1: Prices of the components required for the enzymatic route to make 1 mg of 1 kb DNA with an overhang for annealing to the N$_3$-oligo.

Considering this was just the price to make 1 mg of the desired 1 kb piece of DNA, this seemed a prohibitively large amount of money costing ~£29,800/mg. Thankfully, the PCR reaction was much more reasonable due to the exponential amplification of the desired piece of DNA. The main expenditure for this route was the KOD enzyme kit which comes with 200 µL KOD and 1 mL of 2 mM dNTPs along with all of the necessary buffers. The following calculations are based on the isolation of 409 µg DNA from the PCR clean-up of 16 mL of PCR reaction mixture. One kit provides enough KOD to carry out three 16 mL batches of PCR reaction mixture using half the recommended KOD quantities for 35 cycles (as discussed in the previous section) which would allow ~1.2 mg of DNA to be isolated. However, as only 1 mL of 2 mM dNTPs is provided, an extra 304 µL of 25 mM dNTPs is needed which is obtained from Generon for £31.80/mL. As can be seen in Table 3.2, the PCR route is much more economically viable and the most realistic route for making the azide-functionalised DNA costing ~£340/mg. In practical terms, a higher throughput approach to the PCR such as continuous flow PCR has the potential to make this a practical route.

	Cost to make 1.2 mg 1 kb DNA	Cost to make 1 mg 1 kb DNA
1 × KOD kit	£370	£308
Extra dNTPs	£10	£8
2 × Midi prep kit		£13
2 × PCR clean up		£14
	TOTAL	**£343**

Table 3.2: Prices of the components required to make 1 mg of 1 kb azide-functionalised DNA using the PCR route.

3.3.3 N-terminal labelling of a protein with an azide-functionalised 1 kb piece of DNA

As has already been demonstrated, MBP was able to be efficiently labelled on the N-terminus using sortase 7M and the BCN-depsipeptide (see Section 3.2.5). Using this as a starting point, the MBP-BCN and N_3-DNA were mixed together in a ratio of 1:1.2 and incubated at 37 °C for 24 hours. The resulting mixture was analysed by SDS-PAGE which was stained using ethidium bromide after being used to carry out a Western blot against an MBP antibody. On the SDS-PAGE gel it was possible the see very clear smearing for the MBP-DNA towards a band at ~100 kDa. In contrast, the well containing the N_3-DNA alone remained at the top of the gel. The smearing is indicative of a protein-DNA interaction due to the way in which they move through the gel matrix. In addition to this, a faint band on the Western blot can be seen, also at approximately 100 kDa, implying an MBP species may also be present at that mass (Figure 3.29). However, it is difficult to say whether this is due to a covalent linkage through the successful SPAAC reaction between the MBP-BCN and N_3-DNA, or a non-covalent interaction.
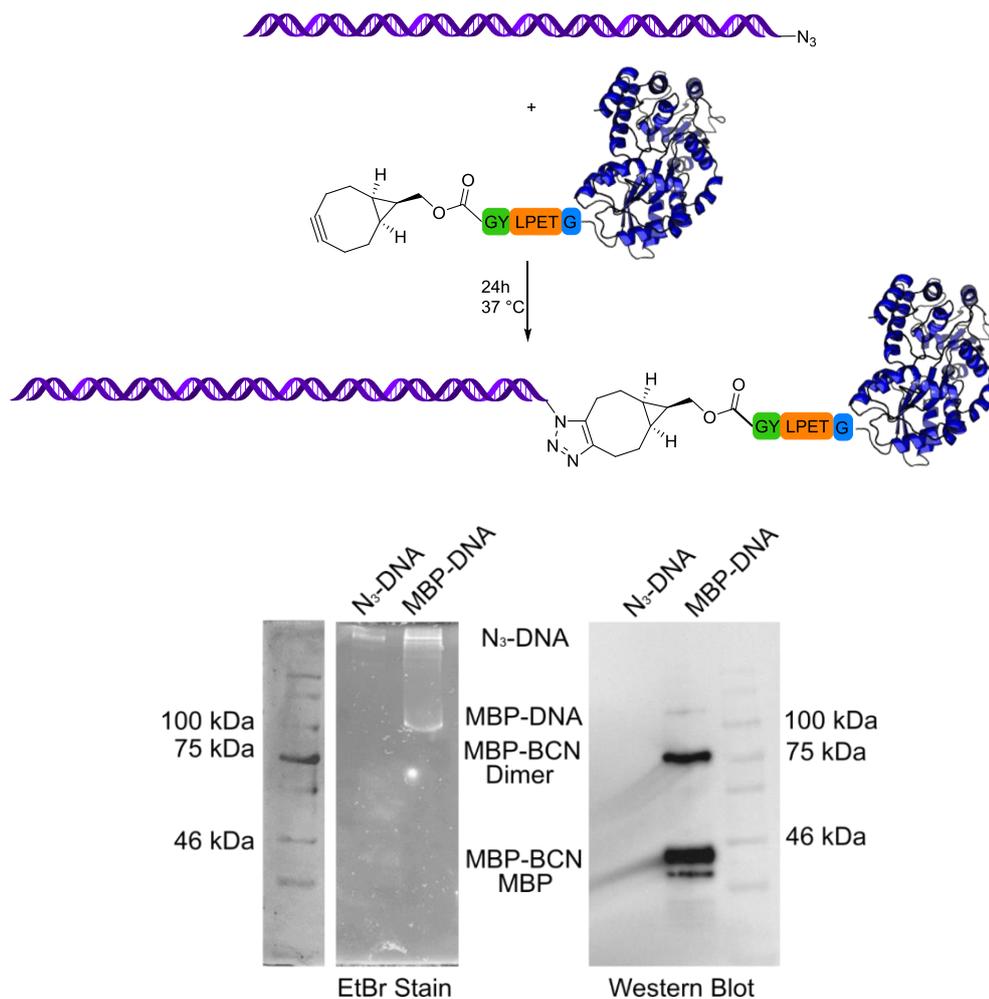
Figure 3.29: Schematic depicting the SPAAC reaction between the N₃-DNA and MBP-BCN alongside the SDS-PAGE gel stained using EtBr and used for a Western blot using an MBP antibody.

3.3.4   Conclusions and future work

In this work, a method has been developed for making and purifying an azide-functionalised 1 kb piece of DNA. This was tested enzymatically using nicking endonucleases and restriction enzymes to create an overhanging piece of DNA and also using PCR with an N_3-oligo acting as a primer in the exponential amplification of the 1 kb strand. Having pursued each method separately to successfully make the desired piece of DNA, the two methods were analysed based on their cost. The enzymatic route was completely unrealistic coming out with a price of ~£30,000 per mg of DNA made, while the PCR route seemed much more appropriately priced at ~£350 per mg.

Once the PCR reactions had been scaled up, a test reaction was successfully carried out between the MBP-BCN and azide through SPAAC, and the gel analysis appeared to show an interaction between the protein and the 1 kb strand of DNA. These results show promise as an approach for creating a protein-DNA fusion using this method and the next steps would proceed towards developing a flow PCR set up enabling further scale up of the N_3-DNA for use in SAS flow experiments.

Since little was known about the ideal length of DNA required to induce anisotropy under the flow conditions, different lengths of DNA needed to be tested with collaborators at the University of Hamburg. This would allow the different DNA species to be analysed with respect to their flow characteristics and help gain a better understanding of the length of DNA required to induce the desired anisotropic effect. As these DNA strands did not need to be functionalised with an azide, a number of EcoRV sites were introduced using site-directed mutagenesis into a pUC18 plasmid (2686 bp) to allow large scale digestion and purification. The location of each site is illustrated in Figure 3.30a with three separate mutated plasmids containing EcoRV sites at A and B, A and C and C and D. Digestion at these locations results in a range of double stranded DNA fragments (~ 500, 2000, 1000 and 1500 bp,). Work so far has successfully isolated large amounts of each mutated plasmid (5-6 mg) which have been digested into the appropriate DNA pieces (Figure 3.30b). However, the purification of the strands has proved to be difficult, failing to result in the separation of even the 1500 and 500 bp mixture using both SEC and anion exchange chromatography. Size-exclusion chromatography using the Superdex® S200 eluted isocratically using 10 mM Tris HCl at pH 8.8 failed to achieve separation. Anion exchange chromatography was attempted using a DEAE Sepharose column which allowed the isolation of the larger 1500 bp species, however the 500 bp was only obtained as a mixture with the larger piece. This work is ongoing to find an optimal purification technique for the separation of the different sized species.
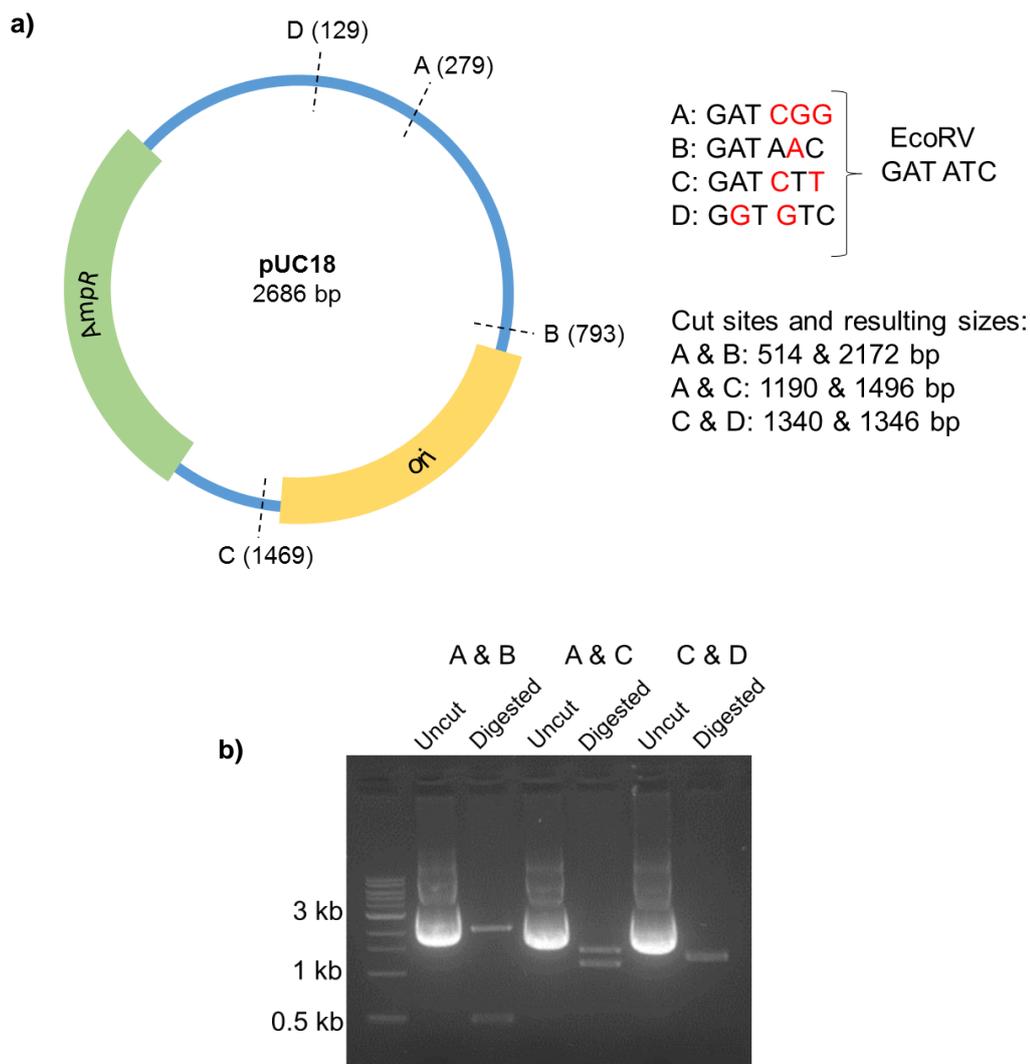
a)

D (129)

A (279)

A: GAT CGG
B: GAT AAC
C: GAT CTT
D: GGT GTC

EcoRV
GAT ATC

pUC18
2686 bp

AmpR

ori

B (793)

Cut sites and resulting sizes:
A & B: 514 & 2172 bp
A & C: 1190 & 1496 bp
C & D: 1340 & 1346 bp

C (1469)

b)

A & B   A & C   C & D

Uncut Digested Uncut Digested Uncut Digested

3 kb

1 kb

0.5 kb

Figure 3.30: a) Mutations for the introduction of EcoRV digestion sites to allow a range of DNA pieces to be produced for behavioural studies in flow systems. b) Agarose gel of successfully digested plasmid DNA containing the appropriately placed EcoRV restriction sites.

3.4   Overall chapter conclusions

Overall in this chapter the initial aim to provide useful applications for the sortase-mediated N-terminal labelling method have been successful. By incorporating the bicyclononyne handle into MBP two very different applications have been studied.

Firstly, the SPAAC between the BCN and a short N_3-oligo and subsequent purification methods allowed MBP to be incorporated into a DNA origami frame. This was then used in HS-AFM to

observe the sortase 7M in action through the recognition of the LPETG incorporated into the system and subsequent cleavage and removal of MBP from the window. By using this method in the future, more enzymatic processes could be monitored in this way. The ability to decorate the DNA origami frame using this approach with different proteins could allow them to be observed for similar processes of displacement and potentially replacement with other species.

The second application is the use in small angle scattering techniques, this work has provided a method for making a long piece of double stranded DNA functionalised with an azide for SPAAC with the MBP-BCN. Although this work has not gone to the proof of concept stages with regards to whether anisotropy is observed using these protein-DNA fusions, it is at a point for scaling up for future applications to the system. By carrying out additional investigations into the length of DNA required to induce alignment in flow, the system can be optimised to try and induce the anisotropic effect desired.

Overall this chapter has discussed the efficiency sortase 7M is able to provide in site-specifically labelling proteins with a useful biorthogonal handle. The reaction can take place within a few hours on the bench but can also take place in the fridge or on ice over longer periods of time. This means unique functionalities like BCN can be incorporated into proteins of variable levels of stability for use in much wider applications. The two discussed in this chapter, HS-AFM and SAXS, are just brief examples of the range of possibilities available using this approach.

4   Biological studies of ADC as a potential antibiotic target

The pantothenamides are a large class of antibiotic compounds, one of them, N-pentylpantothenamide (N5-Pan), has been known to be particularly potent against *Escherichia coli* cells since 1970.[156] In 2002, it was reported that this activity was due to its turnover to ethyldethia coenzyme A (EtdtCoA), an antimetabolite.[157] The work detailed in this chapter investigates the role the PanD/PanZ complex plays in the toxicity of N-pentylpantothenamide, confirming it as a target for EtdtCoA.

4.1   Background

Over the past few years the World Health Organisation (WHO) has described the world as heading towards a "post antibiotic era" with current treatment losing effectiveness as resistance becomes more prominent in the antimicrobial field. This makes any developments greatly encouraged to help avoid a situation where common infections and minor injuries could prove fatal in the near future.[158] In 1970, N-pentylpantothenamide (N5-Pan, **4.01b**) was found to exhibit antibiotic activity against *E. coli* with a minimal inhibitory concentration (MIC) of ~2 µM.[156] Later, Strauss and Begley demonstrated this activity was linked to the metabolism of N5-Pan through the biosynthetic pathway of coenzyme A (CoA, **4.02a**) to form ethyldethia coenzyme A (EtdtCoA, **4.02b**). The enzymes involved in this pathway seemed to prefer the antimetabolite over the natural substrates leading to the EtdtCoA being used as a substrate by phosphopantetheinyl transferases forming inactive acyl carrier proteins known as crypto-ACPs (**4.03b**, Scheme 4.1).[157] Acyl carrier proteins are important components in fatty acid and polyketide biosynthesis and are produced from the coenzyme A biosynthetic pathway. Therefore, by disrupting this process through the formation of the modified crypto-ACPs, fatty acid synthesis is inhibited along with cell growth.[159]
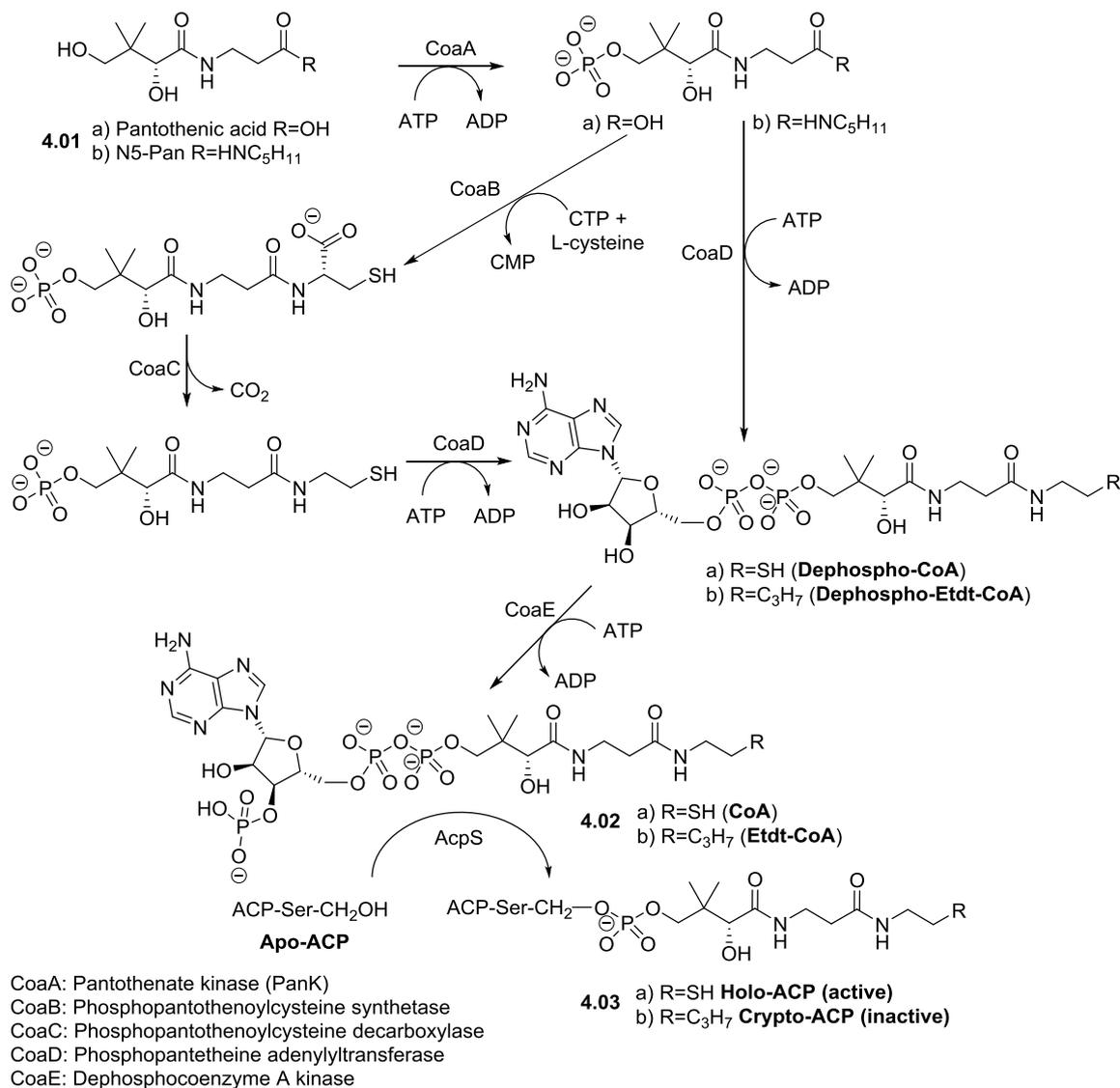
Scheme 4.1: The biosynthetic route for the production of coenzyme A from pantothenic acid and ethyldethia coenzyme A from the precursor N-pentyl pantothenamide, a known antibiotic which can be seen in this schematic to lead to formation of inactive crypto-ACPs.[160,161]

However, it was subsequently shown that crypto-ACPs could be effectively recycled by ACP hydrolase, and in the presence of exogenous pantothenate, cell growth was rescued. This means in the presence of N5-Pan another mechanism must be taking place to deplete the cellular CoA pool.[161] Previous work in the group had suggested that pantothenate biosynthesis in *E. coli* is feedback regulated by CoA. This involves a complex formed between an activated form of PanD, known as ADC (α-aspartate decarboxylase), and PanZ, which limits the supply of β-alanine in

response to the concentration of CoA in a negative feedback loop.[108] In this chapter, the formation of the PanD/PanZ complex will be discussed using isothermal titration calorimetry (ITC) and surface plasmon resonance (SPR) studies and the information they can provide about the role the complex plays in the toxic effects of N5-Pan on *E. coli* cells.

4.2   The PanD/PanZ complex

The final step in the biosynthesis is pantothenate in all organisms involves the condensation of pantoate and β-alanine which is derived from L-aspartate. As has already been briefly mentioned, in bacteria, the turnover of L-aspartate to β-alanine is carried out by the enzyme aspartate α-decarboxylase (ADC).[162] The proenzyme, PanD, is activated to α-aspartate decarboxylase (ADC) through the binding of an accessory protein, PanZ. This activation occurs when the peptide backbone undergoes an N to O acyl shift forming an ester which is cleaved by elimination between Gly24 and Ser25 resulting in a pyruvoyl cofactor which catalyses the decarboxylation of L-aspartate.

The binding of PanZ is essential as it restricts the PanD Thr16-Tyr22 loop placing the Glu23 and Ser25 in a conformation which favours the activation reaction. However, this interaction was also found to be dependent on the presence of CoA with X-ray crystallography indicating the binding site for CoA is very close to the interface between PanD and PanZ. In addition to this, NMR experiments suggested the CoA binding causes PanZ to adopt an "ADC-binding conformation" because it stabilises the PanZ Leu66-Gly76 loop which forms contacts with PanD.[108,163]

Once the PanZ.CoA is bound to PanD it forms a heterooctameric species. The four active sites are located at the interface between PanD and PanZ. Therefore when the PanZ is bound the catalysis of L-aspartate to β-alanine is inhibited. Only when the PanZ falls off is the turnover able to take place (Scheme 4.2).
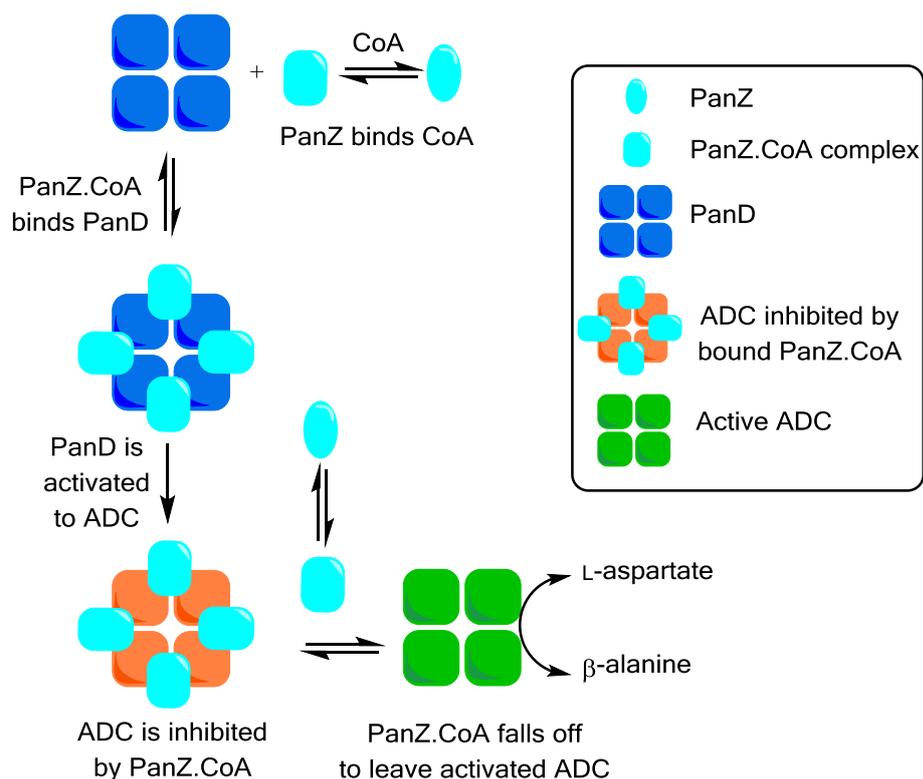
Scheme 4.2: Pathway from PanD and PanZ to activated ADC for β-alanine production. Adapted from Monteiro et al.[108]

Previous work within the group found that this system is regulated by a negative feedback loop; at normal CoA concentrations equilibrium is towards the unbound active state. However, when CoA concentrations are high the PanZ.CoA complex remains bound to the PanD complex inhibiting the β-alanine production and therefore preventing the biosynthesis of more CoA.[108] This response to the changes in cellular concentration of CoA could be similar to the mechanism responsible for the enhanced toxicity of N5-Pan through the build-up of EtdtCoA in this system interrupting the biosynthesis of pantothenate through the PanD/PanZ complex. To gain a better understanding of whether this could be the case within a cellular context, more needed to be done to understand the formation of the PanD/PanZ complex itself.

4.2.1 Biophysical studies of the PanD/PanZ.CoA interaction

To be able to explore this system further it was necessary to understand more about the interactions taking place to activate PanD to ADC through the binding of PanZ.CoA. Collaborators in Japan at the National Institute of Genetics have previously studied a range of PanD and PanZ mutants. The overexpression of PanZ has been seen to inhibit bacterial growth by inducing β-alanine auxotrophy.[108] They found two mutants, PanD(K119A) and PanZ(R73A), were not susceptible to this overexpression-induced growth inhibition (Figure 4.1a). This implied they were both able to produce a catalytically activated PanD but the bound PanZ did not inhibit β-alanine production when PanZ was overexpressed. They also found that in analytical size-exclusion chromatography (SEC), the wild-type PanD/PanZ elutes as a heterooctameric complex. However, no complex was observed for PanZ(R73A) and partial disassembly was seen for the complex with PanD(K119A) (Figure 4.1b). As both mutants allow activated ADC to be produced, these mutated residues must be involved in the PanD/PanZ interaction. The results of the SEC for PanD(K119A) imply weakened protein-protein interactions. However, PanZ(R73A) didn't appear to be able to form a stable complex at all. This indicated that the mutated residues are involved in the protein-protein interface though neither play a direct role in the interaction. PanD Lys119 is thought to be required for the stabilisation of the hinge region of the complex, whereas PanZ Arg73 is thought to be involved in the binding of CoA, the δ-guanidino group locks the pantetheine in the binding pocket by forming a salt bridge with the Glu103 residue.[164]
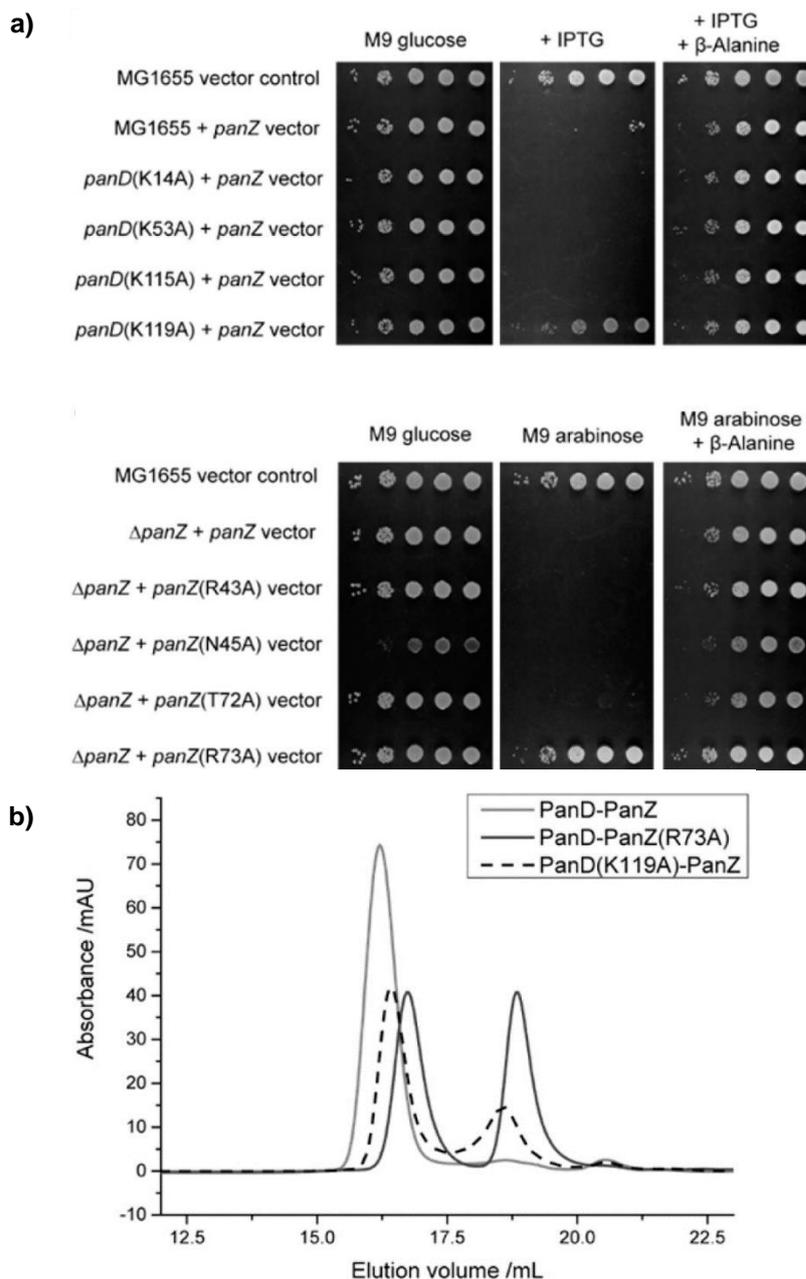
Figure 4.1: This work was carried out by collaborators in Japan at the National Institute of Genetics a) Growth inhibition assays of a number of PanD and PanZ mutants resulting in the discovery that PanD (K119A) and PanZ (R73A) both formed catalytically active PanD and no inhibition in growth was observed when PanZ was overexpressed. b) SEC trace comparing the level of complexation observed between different PanD/PanZ species. The wild-type proteins complex and elute as one species at ~16 mL whereas the PanD(K119A) and WTPanZ run as a mixture of complex and PanD(K119A) tetramer. The WTPanD and PanZ(R73A) elute from the column as tetrameric and monomeric species, without any heterooctamer present.[164]

It was predicted that these mutations meant the binding between the mutant PanD and PanZ.CoA (Figure 4.2) is transient and much weaker in comparison to the wild type system which is seen by *E. coli* cell growth when PanZ is overexpressed. To demonstrate the involvement of the Arg73 residue in CoA binding, isothermal titration calorimetry (ITC) experiments were carried out on the mutant PanZ(R73A) to see if there is an observable difference between the binding of the mutant and WTPanZ with acetyl CoA and an inactivatable PanD(S25A) mutant. By not being able to activate to ADC, this mutant allows the PanD/PanZ interactions to be monitored without complications associated with the activation reaction. In addition to this, acetyl CoA was used instead of CoA because in the past CoA has been found to provide less reliable data in ITC experiments than AcCoA. The hypothesis for this was that CoA could potentially form disulphide-linked dimeric species which may disrupt the interactions being monitored. However, AcCoA is unable to form these dimers and making this swap appeared to allow more reproducible data to be obtained.[165]
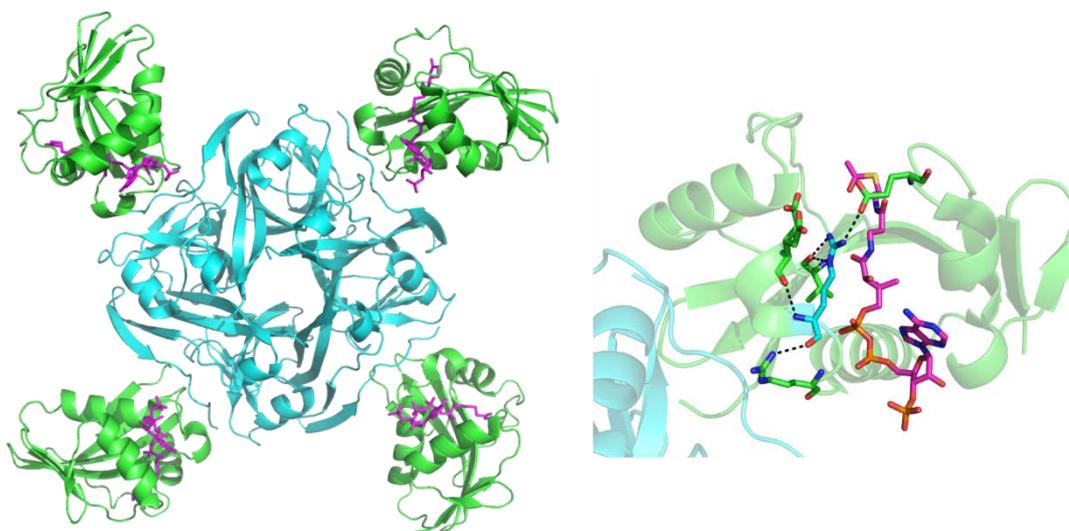
Figure 4.2: Crystal structure of ADC (blue tetramer) bound to PanZ.CoA showing the interactions taking place between the residues of the helices of PanZ (shown in green) and a molecule of CoA (shown in pink). Image was made using PYMOL, PDB ID: 4CRZ.[108]

Isothermal titration calorimetry (ITC) allows thermodynamic properties of biomolecular interactions to be obtained including binding constants (K_d), stoichiometry (n) and enthalpy (ΔH).

It is achieved by titrating one binding partner (the ligand) into another (the receptor) while maintaining a constant temperature (usually ~25 °C). The ligand is contained within a stirring syringe while the receptor is placed in a cell and as the ligand is added to the receptor the subsequent heat change as a result of binding can be found. This is done by comparing the power required to keep the cell at the constant temperature with the power required to maintain the reference cell, containing only water, at that temperature too. Small injections are made over a set time period, and as the amount of ligand in the cell increases and the receptor binding sites fill up, the heat change experienced by the cell decreases. For example, in the case of exothermic binding a sharp negative peak will be observed followed by a return to baseline before the next injection. These peaks will gradually decrease in size as the point of saturation is reached, providing the raw ITC data. This data is then further processed and modelled using computer software to provide the thermodynamic parameters of the binding.

The ITC experiments investigated whether there was a change in the K_d value obtained from experiments using WTPanZ and PanZ(R73A) when binding to acetyl CoA (AcCoA) and PanD(S25A). All three proteins were overexpressed in lysogeny broth (LB), inducing using arabinose for the PanZ proteins or IPTG for PanD. The cells were lysed, and the proteins were purified firstly by Ni-NTA affinity chromatography followed by size-exclusion chromatography (50 mM Tris, 100 mM NaCl and 0.1 mM DTT, pH 7.4). The binding studies between PanZ(R73A) and AcCoA showed a dramatic loss in the affinity of the PanZ for AcCoA by approximately 250-fold with a dissociation constant of $221 \pm 48\,\mu M$ (Figure 4.3a) while for WTPanZ, $K_d = 0.99 \pm 0.05\,\mu M$ (Figure 4.3b).

However, only a ~5-fold decrease was seen for the affinity for PanD(S25A) in the presence of high concentrations of AcCoA with $K_d = 734 \pm 60\,nM$ (Figure 4.3c) and previously reported WTPanZ and PanD(S25A) interaction having K_d ~0.15 μM.[108] The dramatic decrease in affinity for AcCoA illustrated in these results explains the lack of complex formation seen by SEC between PanD and PanZ(R73A). The PanZ(R73A) and PanD(S25A) complex was still able to form; however, it required extremely high concentrations of AcCoA.

a) 50 μM WTPanZ and 400 μM AcCoA
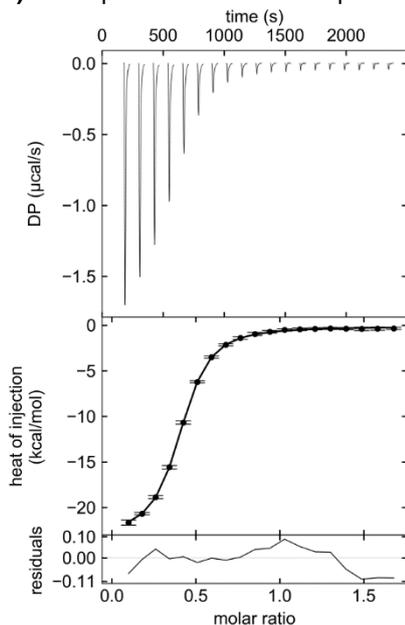
	a	b	c
Kd (μM)	0.99 ± 0.05	221 ± 48	0.73 ± 0.6
ΔG (kcal/mol)	-8.18 ± 0.03	-4.98 ± 0.34	-8.36 ± 0.05
ΔH (kcal/mol)	-22.96 ± 0.24	-16.53 ± 0.04	-12.51 ± 0.22
ΔS (cal/mol*K)	-49.59	-38.8	-13.9
n	0.39	0.99	0.96

b) 50 μM PanZ(R73A) and 5mM AcCoA
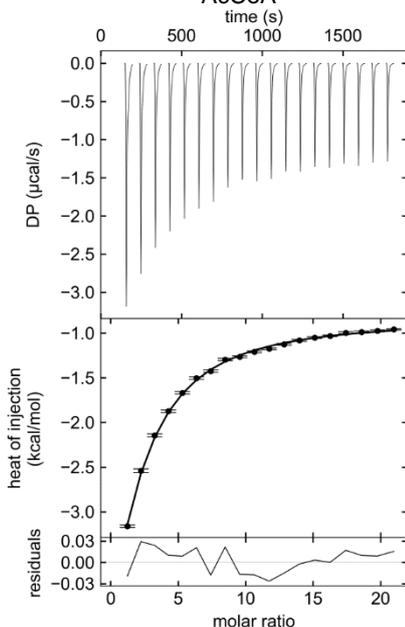
c) 26 μM PanZ(R73A), 260 μM PanD(S25A) & 1 mM AcCoA
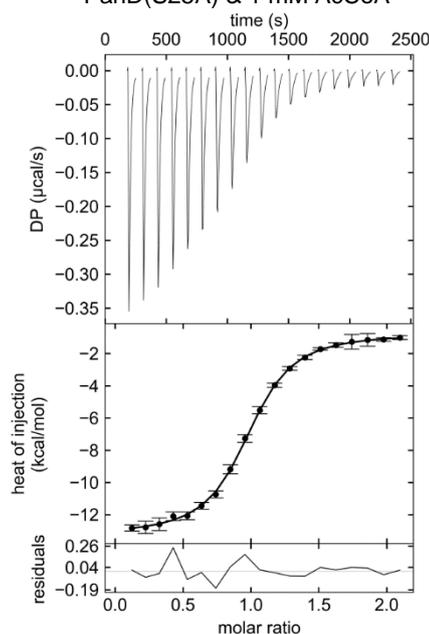
Figure 4.3 ITC traces and thermodynamic data for the following experiments: a) 50 μM WTPanZ and 400 μM AcCoA , Kd= 0.99 ± 0.05 μM. b) 50 μM PanZ(R73A) and 5mM AcCoA, Kd= 221 ± 48 μM, loss of the arginine 73 residue resulted in ~250-fold decrease in affinity. c) 26 μM PanZ(R73A) and 260 μM PanD(S25A) in the presence of 1 mM AcCoA, Kd= 734 ± 60 nM, only a 5-fold loss in affinity compared to WTPanZ (0.15 μM).[164]

4.3 The PanD/PanZ complex as a target of EtdtCoA

Wang *et al.* performed a high-throughput proteomic abundance screening for *E. coli* and found the concentrations to be approximately 500 nM for PanD and 150 nM for PanZ. This means *in vivo*, it is possible that the PanZ.CoA complex concentrations could be too low to inhibit PanD.[166] To explore this further, N5-Pan was used as a small molecule probe to look at the behaviours of PanD and PanZ at these lower concentrations.

As was discussed earlier, N5-Pan has been known to show antibiotic activity for several years with particular potency against *E. coli* in comparison to other bacteria. PanD is responsible for the synthesis of β-alanine (**4.05**) from L-aspartate (**4.04**) which then undergoes a condensation reaction with pantoate to make pantothenic acid (**4.01a**). N5-Pan (**4.01b**) is a very similar structure to pantothenic acid and has been shown by Strauss *et al.* to compete against it for the CoA metabolic pathway. The antimetabolite acts as an alternative and preferential substrate for the CoA biosynthetic pathway (Scheme 4.3) which leads to the synthesis of ethyldethia coenzyme A (EtdtCoA, **4.02b**) over coenzyme A (CoASH, **4.02a**) resulting in a toxic effect.[157]
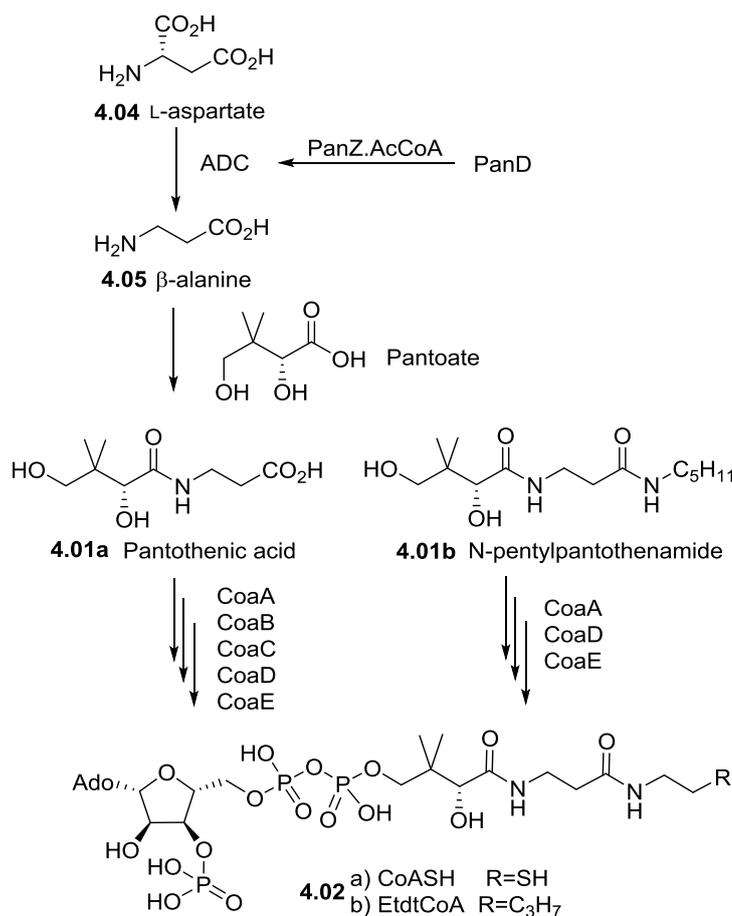
Scheme 4.3: Comparison between the biosynthetic pathway from β-alanine to CoA and the preferential metabolism of N-pentylpantothenamide to EtdtCoA.[164]

The reduction of CoA biosynthesis as a result of EtdtCoA formation could be due to an interaction between PanZ and EtdtCoA leading to the inhibition of PanD. It is known that PanZ binds to CoA and AcCoA with equal affinity, illustrating a lack of sensitivity for the ligand in this region and suggesting it will also bind to EtdtCoA. Work done alongside Holly Morgan involved the synthesis of EtdtCoA through chemical and enzymatic methods for use in ITC studies to understand whether this hypothesis could be true.

4.3.1 Synthesis of EtdtCoA

The synthesis of EtdtCoA involves multiple synthetic steps to form 4'-phospho-N-pentylpantothenamide which was carried out by Holly Morgan before being used

as a substrate for the enzymatic turnover to form the final product using CoaD and CoaE. The first step is N5-Pan (**4.01b**) from pantothenic acid (**4.01a**), generated through ion-exchange chromatography of calcium pantothenate, through an amide coupling reaction between the carboxylic acid of the pantothenic acid and the amine of amylamine (**4.06**). Following this, N-pentyl-4'-O-dibenzylphosphate (**4.07**) was synthesised using tetrazole-mediated phosphorylation involving the coupling of a phosphoramidite group to the N5-Pan followed by an oxidation to form the phosphate. The final step of the chemical synthesis of 4'-phospho-N-pentylpantothenamide (**4.08**) was a benzyl deprotection through palladium catalysed hydrogenation (Scheme 4.4 of synthetic route).[157]
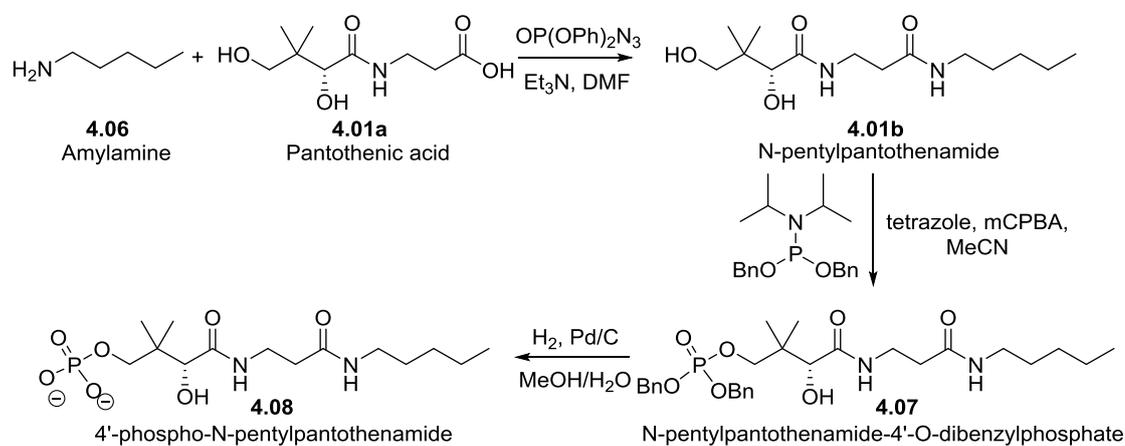
Scheme 4.4: Synthesis of the chemical precursor to EtdtCoA, 4'-phospho-N-pentylpantothenamide done by Holly Morgan following the protocol detailed by Strauss et al.[157]

The enzymes involved in the transformation of 4'-phospho-N-pentylpantothenamide (**4.08**) to EtdtCoA (**4.02b**) are phosphopantetheine adenylyltransferase (CoaD) and dephosphocoenzyme A kinase (CoaE). The two enzymes were expressed in auto induction media (AIM) by Reem Al-Shidani. The pellets were later lysed and purified using Ni-NTA affinity chromatography by Holly Morgan. The enzymes were used in a "one-pot" approach described by Strauss and Begley in which both enzymes and 4'-phospho-N-pentylpantothenamide (**4.08**) were added in the presence of magnesium chloride and ATP (Scheme 4.5).[157] The reaction mixture was incubated at 37 °C for 4 hours before removing the enzymes through denaturation and subsequent

precipitation by heating at 95 °C. This enzymatic turnover works by CoaD first converting

4'-phospho-N-pentylpantothenamide (**4.08**) to the intermediate, dephosphoethyldethia

coenzyme A (**4.09**) using ATP (which must be bound to Mg^+ ion to make it biologically active,

hence the presence of $MgCl_2$). This adds AMP to the phosphate group of the starting compound

and PP_i is released. CoaE then uses ATP to add a final phosphate group resulting in the formation

of EtdtCoA (**4.02b**) and releasing ADP as a side product.
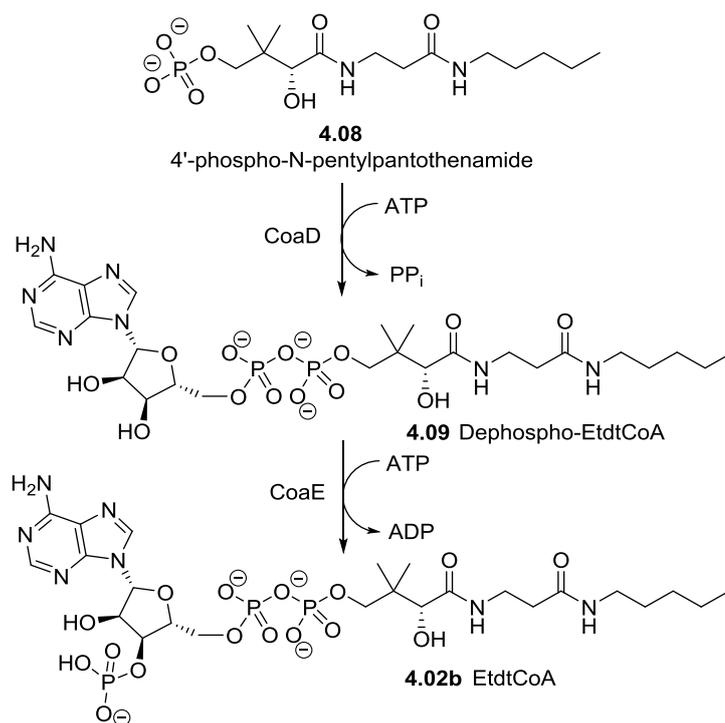
Scheme 4.5: Enzymatic turnover of 4'-phospho-N-pentylpantothenamide using CoaD and CoaE in the presence of ATP leading to the formation of EtdtCoA.

The literature suggested purification of the EtdtCoA using ion-exchange chromatography.

However, after attempting this with Q Sepharose and DEAE-cellulose resins using a gradient

elution with NH_4HCO_3 (50-300 mM, over 100-600 mL) there was little success with only very

small quantities of the expected m/z (777.55) observed by LCMS. Corkey *et al.* demonstrated the

separation of a number of CoA intermediates using HPLC purification using isocratic elution with

50 mM KH_2PO_4 (pH 5.3) and methanol in a ratio of 9:1.[167] This worked well allowing the ADP

and ATP to be separated from the reaction mixture leaving the purified EtdtCoA which was

subsequently desalted a number of times to try and reduce salt content to enable it to be used for ITC experiments. The resulting concentration of EtdtCoA was calculated using the A_{257} and an extinction coefficient of 16,840 M^{-1} cm^{-1} obtained from Dawson *et al.*[168]

4.3.2 Isothermal titration calorimetry to compare the binding of AcCoA and EtdtCoA

Purified wild-type PanZ was tested by ITC against both AcCoA and EtdtCoA to see if PanZ would bind to EtdtCoA in a similar manner. The experiments revealed that AcCoA bound tightly, as expected, with $K_d = 985 \pm 25$ nM. However, it is important to note that when purifying PanZ, some CoA co-purifies, meaning there is an inherent CoA presence leading to substoichiometric binding to be observed. Interestingly, the EtdtCoA experiment showed a very similar binding profile with a K_d value of 795 ± 190 nM. This value was obtained by globally fitting four separate experiments using varied concentrations of PanZ (Figure 4.4). Due to the sensitivity of ITC for differences in buffers between the substrate and ligand, this was a promising result. However, it needed to be explored using a technique where the salt content of EtdtCoA would not be an issue, for example surface plasmon resonance.

a) 18 µM PanZ and 85 µM AcCoA
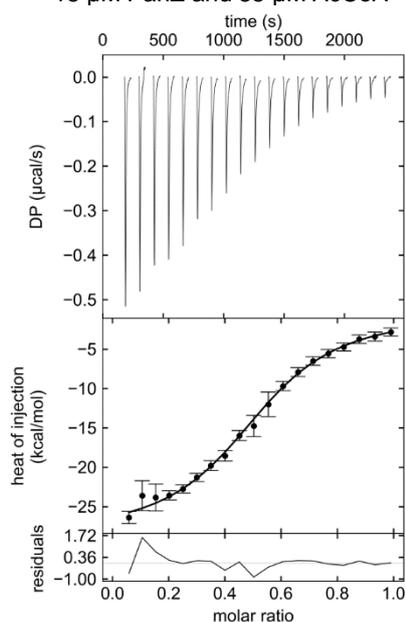
	AcCoA	EtdtCoA
Kd (µM)	0.99 ± 0.03	0.80 ± 0.19
ΔG (kcal/mol)	-8.19 ± 0.14	-8.31 ± 0.14
ΔH (kcal/mol)	-28.96 ± 1.05	-39.34 ± 2.78
ΔS (cal/mol*K)	-69.71	-104.12
n	0.52	0.53

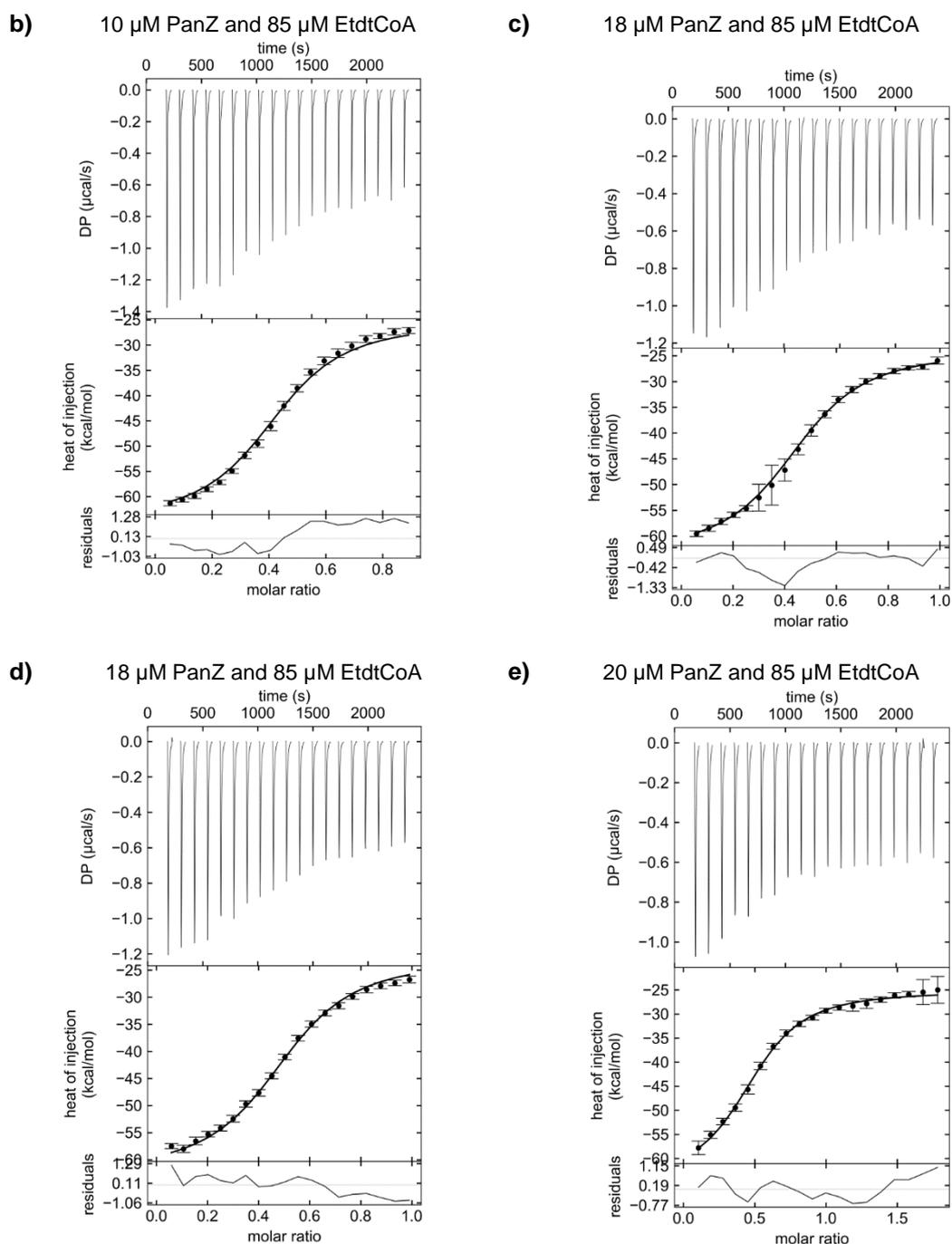*Figure 4.4: ITC traces and thermodynamic data for the following reactions where b-c were globally fitted: a) 18 μM PanZ and 85 μM AcCoA, K_d= 985 ± 25 nM. b) 10 μM PanZ and 85 μM EtdtCoA, c) & d) 18 μM PanZ and 85 μM EtdtCoA titrations and e) 20 μM PanZ and 85 μM EtdtCoA).[164]*

4.3.3 Surface plasmon resonance of AcCoA and EtdtCoA binding to PanD/Z

Surface plasmon resonance is a technique which allows the kinetics and dynamics of a binding pair to be studied. This is achieved by immobilising one of the binding partners to a gold chip which can be achieved in a variety of ways; for this work a biotinylated-PanD was used with a streptavidin coated chip. A light is shone at the chip from a light emitting diode (LED) which is reflected to a detector. As the sample is flowed over the chip any binding interactions cause a change in the characteristics of the reflected light which is picked up by the detector allowing an optical response to be obtained. In the past, work had been attempted by another member of the group using biotinylated-PanZ for SPR while flowing PanD and CoA over the chip, however there was no significant signal observed when carrying out the experiment in that order, therefore this work focussed on attaching the PanD to the chip instead.

Using this technique, the PanD(T57V) was N-terminally biotinylated using sortase 7M and a biotin-depsipeptide synthesised by Jack Caudwell (Figure 4.5a). The T57V mutant of PanD is similar to the S25A mutant in the fact that it is also unable to undergo the post translational modifications which lead to PanD activation to ADC. The PanD(T57V) was expressed and purified in the same manner as described in Section 2.2.3, followed by treatment with TEV protease (20 mol%) to cleave the His$_6$-tag before being labelled with 3 equivalents of biotin-depsipeptide and 5 mol% sortase 7M. The labelling reaction didn't achieve anywhere near quantitative labelling (Figure 4.5b). However, the complexity of the steric hindrance of the N-termini of the PanD system has already been discussed in Chapter 2. In addition to this, the biotin-depsipeptide was found to be only ~40% pure when analysed by HPLC (Figure 4.5c), this is likely to be the main reason for such poor labelling efficiency. Regardless of the ratio of labelled to unlabelled protein there was enough biotin-PanD present to take forward to SPR as only the biotinylated product would become attached to the streptavidin chip. The WTPanZ was then flowed over the chip surface and the on/off rates of the protein were monitored either with or without the addition of AcCoA/EtdtCoA.
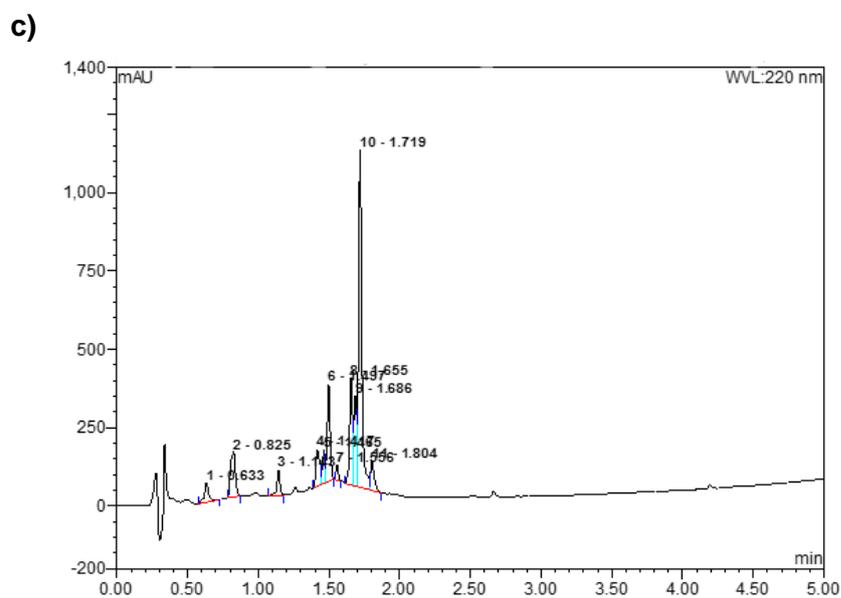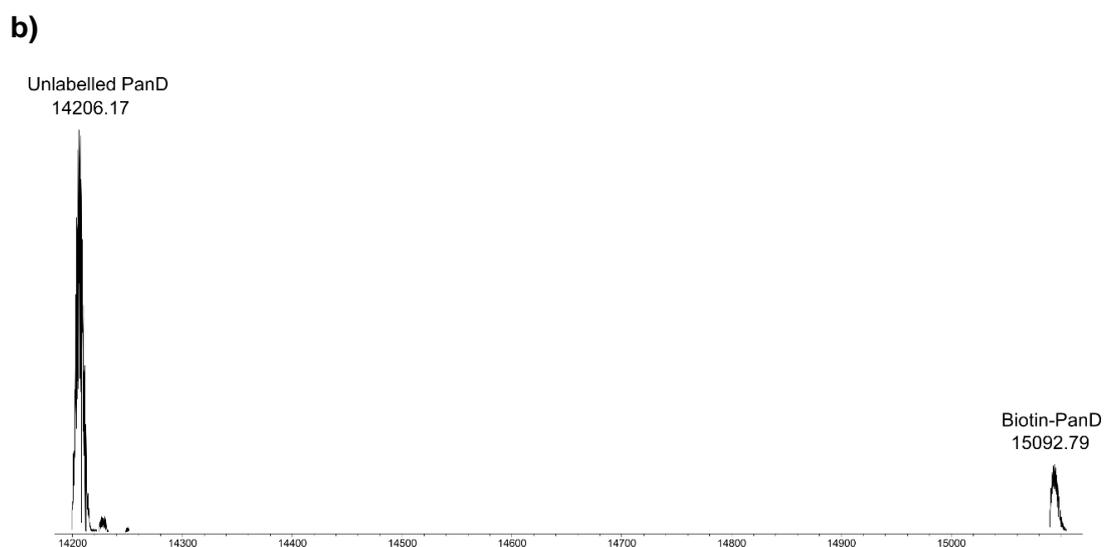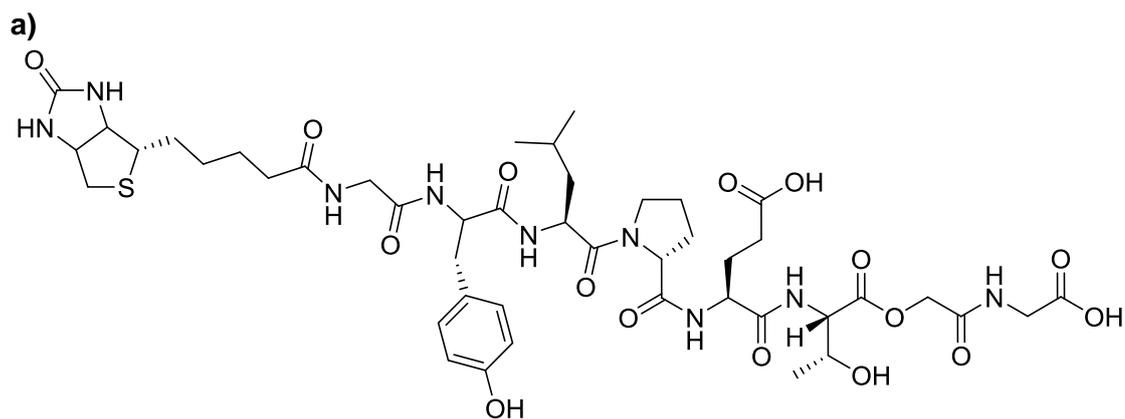
a)

b)

c)

Figure 4.5: a) Biotinylated depsipeptide species synthesised by Jack Caudwell b) HRMS analysis of PanD labelling with biotin-depsipeptide. Majority species is unlabelled PanD, even after 3 hours. c) Analytical HPLC trace of biotin-depsipeptide showing it to be only ~40% pure.

Without the addition of a CoA species PanZ appeared to bind once the concentration exceeded 30 nM. This is because the inherent levels of CoA when PanZ is expressed and purified is usually half that of the PanZ, so when the PanZ concentration reached 30 nM, the CoA concentration was ~15 nM. Above these concentrations enough CoA happened to be present to allow binding to take place between PanD and PanZ (Figure 4.6a). The maximum points of the curves were plotted against the corresponding concentration of PanZ. This was then fitted in Origin using the one-site binding equation y= Bmax*(x/K$_d$+x). However, as can be seen in Figure 4.6b, this fitting was not very good and resulted in a very high error in the predicted K$_d$ value of 981 ± 368 nM. However it serves a tool for comparing to experiments where an excess of AcCoA is added.

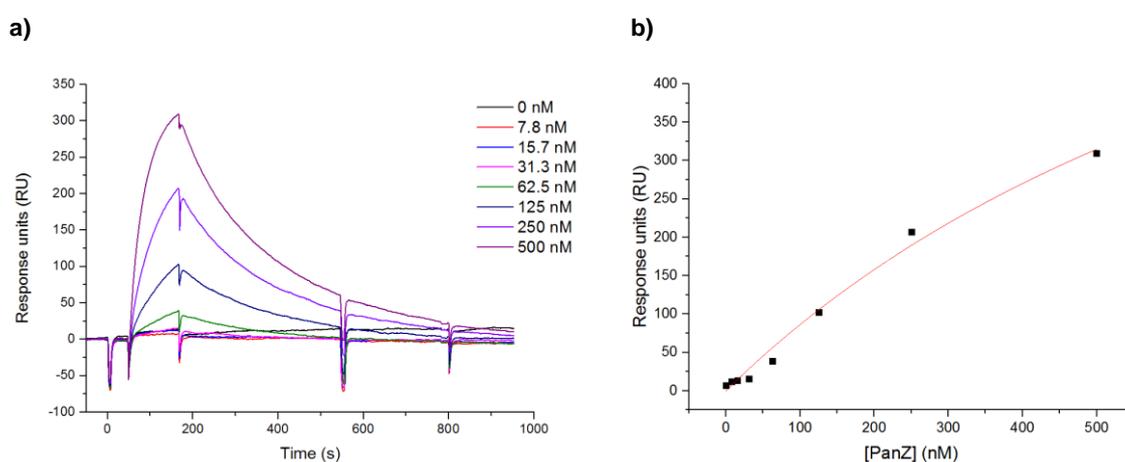a)   b)

Figure 4.6: a) SPR data obtained from experiments using varied concentrations of PanZ (0-500 nM) without the addition of any CoA species. Binding does appear to take place when [PanZ] exceeds 30 nM as CoA is co-purified with the protein and is generally at a concentration of 0.5[PanZ]. b) Highest points of the SPR curves were plotted against concentration, resulting in a K$_d$ value for PanZ-PanD(T57V) of 981 ± 368 nM.

While maintaining a high concentration of AcCoA (20 µM), the PanZ concentration was then varied from 0-500 nM to observe the impact this had on the binding by SPR. When comparing these results to those seen in Figure 4.6, it is obvious that the binding is greatly improved by adding the CoA to the system, with the lowest concentration tested, 7.8 nM, causing binding to the surface (Figure 4.7a). The maximum points of the curves were plotted against the corresponding concentration of PanZ. This was also fitted using the one-site binding equation

y= Bmax*(x/K_d+x). The result of this fitting was a K_d value of 35 ± 1 nM, almost identical to the

results using ITC, where the K_d value for the PanZ-PanD(T57V) interaction with excess AcCoA

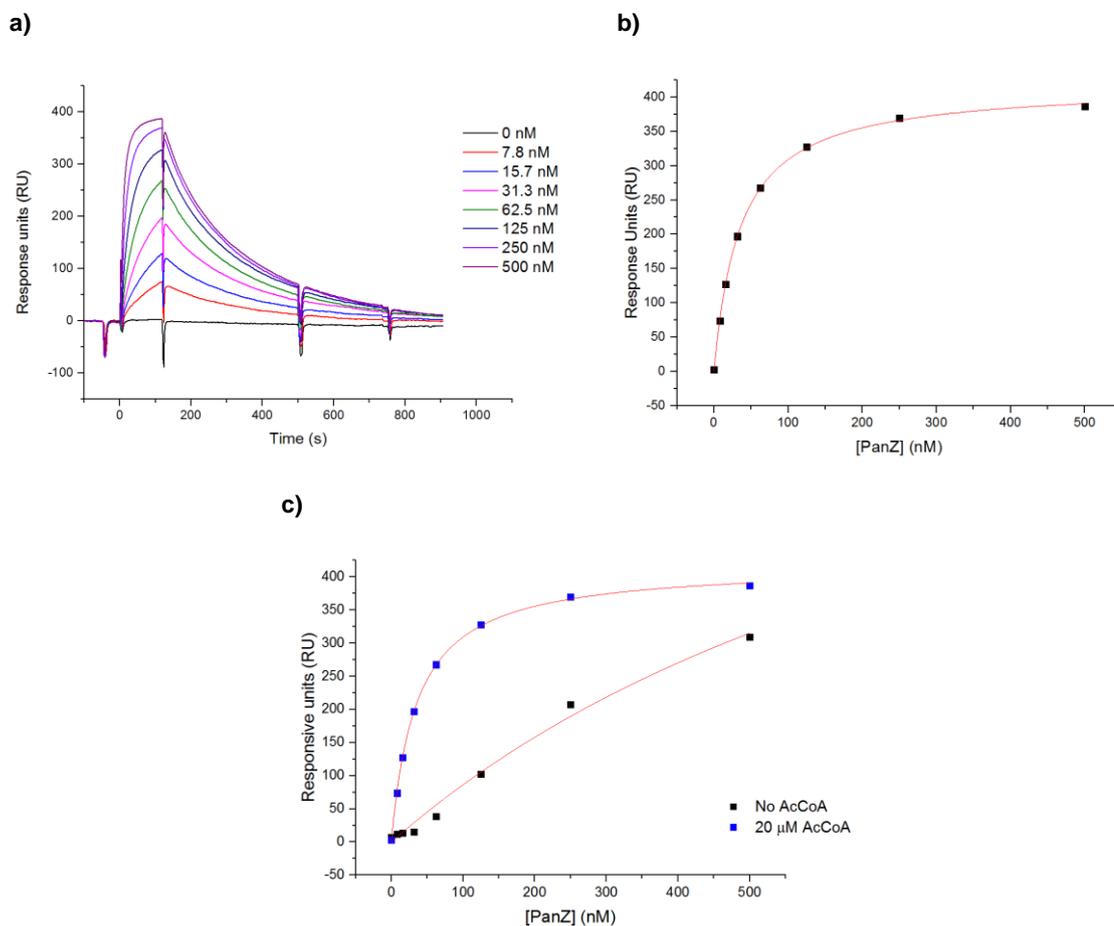was found to be $\sim 35 \pm 9$ nM when studied by Diana Monteiro.

Figure 4.7: a) SPR traces for varied concentrations of PanZ (0-500 nM) in the presence of a large excess of AcCoA (20 µM). b) Plot of the highest points of the SPR curves against the corresponding PanZ concentration, K_d= 35 ± 1 nM. c) Comparison of the highest points obtained with and without AcCoA added.

Experiments were then performed at a much lower constant level of PanZ (50 nM) with varied

concentrations of AcCoA or EtdtCoA (0-2 µM) to observe the effect they have on the system as

a whole. The data collected from these experiments indicated that the EtdtCoA may infact bind

the system with a higher affinity than the AcCoA. The AcCoA curves show a gradual increase

from 50-750 nM with the change in response units decreasing as the concentration reaches the

higher values. However, the EtdtCoA curves jump massively with the lowest concentration of

50 nM showing response units at a similar level as if 1000 nM AcCoA had been added. To analyse the approximate K_d values for this data the highest points were again plotted against the corresponding CoA concentration. However, this time a different equation was used to find out the constants. This is because the inherent CoA within the reaction from the co-purification with PanZ needed to be taken into account, therefore the equation used was $y=A*(x+C)/(K_d+x+C)$ where C is the background concentration of CoA. Using this equation the approximate K_d values were 27 ± 2 nM for EtdtCoA and 584 ± 50 nM for AcCoA (Figure 4.8). This data implies a dramatic increase in affinity between PanD and PanZ.EtdtCoA in comparison to PanZ.AcCoA, especially considering the ITC data for PanZ and the CoA species appeared to be very similar.
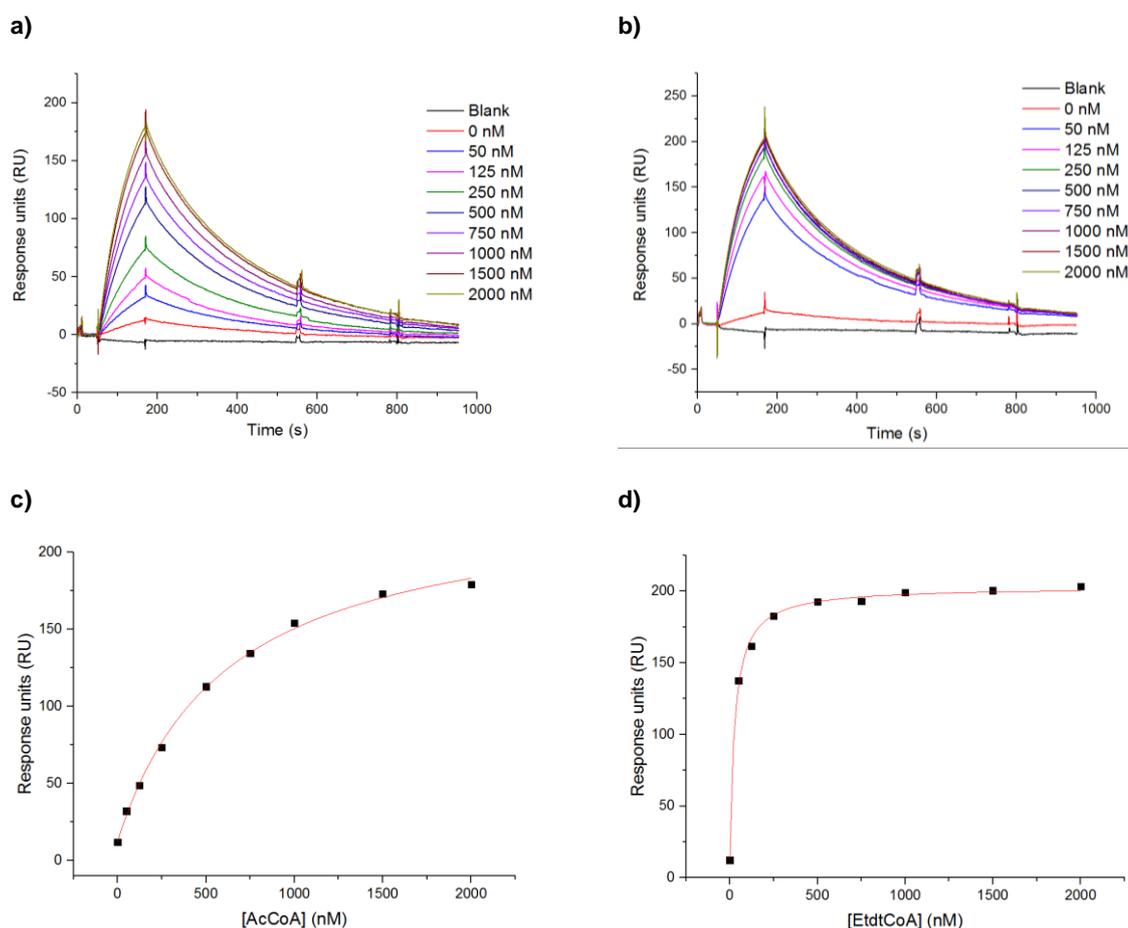

Figure 4.8:a) SPR trace for 50 nM PanZ and varied concentrations of AcCoA (0-2 µM) flowed over a PanD(T57V) functionalised chip b) SPR trace for 50 nM PanZ and varied concentrations of EtdtCoA (0-2 µM) flowed over a PanD(T57V) functionalised chip c) Highest points of AcCoA SPR trace plotted against concentration, $K_d= 584 \pm 50$ nM. d) Highest points of EtdtCoA SPR trace plotted against concentration, $K_d= 27 \pm 2$ nM.

4.3.4 Fluorescence resonance energy transfer between PanD and PanZ

Following the successful FRET experiments in Chapter 2, it was chosen as a potential method of observing the binding between PanD and PanZ in the presence of CoA. If successful, this would have the potential to be a useful high-throughput experimental set up as an assay for sensing CoA. This was done alongside Darren Machin who was working on the sortase-mediated C-terminal labelling of PanZ with the Rhodamine peptide, H_2N-GVSK-Rhod-OH. Using FITC-labelled PanD alongside the Rhod-PanZ it was possible to see whether any FRET would take place as the two came into close proximity with each other.

The Rhod-PanZ and FITC-PanD were mixed together with and without the addition of acetyl CoA and placing in the cuvette for a single reading. FITC-PanD was used at a concentration of 2 µM to which 2 or 4 µM Rhod-PanZ was added alongside 0, 40 or 100 µM AcCoA. Only a small FRET signal was apparent, the signals were normalised to the FITC emission maximum to see if any changes in the curve shape could be observed between the FITC-PanD alone and the PanD/PanZ mixtures (Figure 4.9).
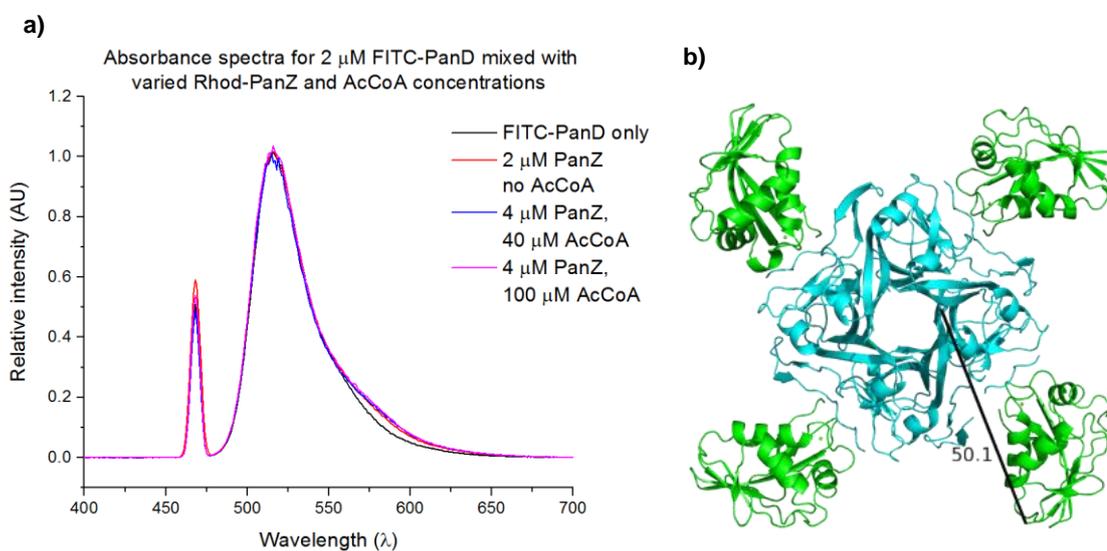

Figure 4.9: a) Absorbance spectra for FRET experiments between FITC-PanD and Rhod-PanZ with or without the addition of AcCoA. b) PanD/PanZ complex with indication of the distance between the N-terminus of PanD and C-terminus of PanZ.

Although the curves do not show a big difference in shape, there is definitely a change associated with the addition of the Rhod-PanZ at a wavelength of ~580 nm. When analysing the crystal structure of the PanD/PanZ complex, it is possible to see why there is only a small amount of FRET being observed for this system. The distance between the N-terminus of PanD and C-terminus of PanZ is ~50 Å, which would be increased further in the mutant PanD being used due to the addition of the short GGSSS tail and the required C-terminal extension on PanZ to facilitate C-terminal labelling. The Förster radius for FITC and rhodamine is ~50 Å, so even without the short extensions on both proteins this system would be at the limit of being able to observe the effects of FRET.[169] However, this shows promise for FRET being used in this way in the future, by adapting the fluorescent molecules being used.

4.4 Conclusions and future work

In this piece of work the aim was to investigate the hypothesis that the known antibiotic, N5-Pan, has enhanced bacteriostatic activity in *E. coli* cells through two mechanisms. The first was as a consequence of the formation of crypto-ACPs due to the preferential metabolism of N5-Pan to form EtdtCoA. However, it was known that this could not be the only mode of action as cell growth has the potential to be rescued from this state by the addition of exogenous pantothenate.[157] This work was focussed on exploring the second mechanism, thought to be causing the downregulation of pantothenate biosynthesis, preventing the cells affected by crypto-ACPs from being rescued. Having already found that PanZ regulates the production of pantothenate in a negative feedback loop within the PanD/PanZ system, the hypothesis was that this could be one of the targets for EtdtCoA, binding to PanZ and leading to the formation of the PanD/PanZ complex.[108]

It was already known that the CoA/PanZ binding interface is able to accept structural changes between CoA and AcCoA so it was expected that EtdtCoA would be able to take part in the same interaction. To investigate this EtdtCoA was synthesised in the lab alongside Holly Morgan through a number of chemical and biochemical steps to use in two types of binding studies. Initial

work using ITC showed EtdtCoA and AcCoA to have almost indistinguishable binding profiles. However, the purification process of EtdtCoA lead to the presence of a lot of salt in the product. This lead to the SPR experiments which are less sensitive to differences in buffer composition and require much smaller quantities of protein and substrate. Using SPR also allowed the binding to be observed between the three components involved; PanD, PanZ and the CoA species of interest. To do this, previous work on N-terminal labelling using sortase 7M was incorporated into this experiment by using a biotin-depsipeptide to label the PanD for attachment to a streptavidin functionalised chip. By exploring the PanD/PanZ interaction using this technique it was possible to see the effect AcCoA and EtdtCoA had on the binding taking place and implies that PanZ.EtdtCoA actually binds to PanD with significantly greater affinity than when AcCoA is used. In addition to this, the SPR provided a more hands off approach to gathering a lot of information about the interaction with rapid on/off rates visible in the shapes of the curves, making it a promising tool for future use when looking into other CoA analogues and this system.

To support this information within a cellular context, studies carried out in Japan found that a strain of *E. coli* in which PanD and PanZ cannot interact was resistant to the N5-Pan. This SN218 strain of *E. coli* had the *E. coli* ADC removed and replaced with ADC from *Bacillus subtilis* (Δ*panD::BspanD*) as it has previously been seen that *B. subtilis* ADC does not require PanZ for activation.[164] Using this strain of *E. coli* supported the hypothesis that the metabolism of N5-Pan forms EtdtCoA within the cellular environment and provides a source of ligand for PanZ which is able to become part of the PanD/PanZ regulatory pathway.

Looking to the future, the work described in this chapter provides a potential platform for advances towards the development of antibacterial therapeutics by targeting this regulatory mechanism. It is already known that N-butylpantothenamide and N-heptylpantothenamide have much lower levels of activity compared to N-pentylpantothenamide.[164] Structurally, the acetyl binding pocket of PanZ does not take part in any specific hydrogen bonding interactions with the carbonyl of AcCoA. This means there are no penalties affecting the binding affinity if it is replaced with EtdtCoA. In addition to this, the binding pocket in which the acetyl group sits is

also able to allow the terminal methyl group of EtdtCoA to fit. However, the difference in activity for the longer analogues implies that the binding pocket is unable to accommodate the likes of N-heptylpantothenamide. The fact that shorter analogues don't seem as effective is unexpected as their metabolites, e.g. methyldethia coenzyme A, should be able to fit into the binding pocket. This indicates that the problem here may be the metabolism of the compounds within the cell by CoaD and CoaE which could be an area to explore in the future. The increase in affinity observed when studying the system by SPR could be due to tighter binding between PanZ.EtdtCoA and PanD. However, this seems unlikely as the CoA species is not directly involved in the PanD/PanZ interface. Therefore, it is more likely to be related to the difference in binding between PanZ and the CoA species. The problems associated with the high concentration of buffer salts following purification by HPLC could have prevented the difference in binding from being observed when analysed in ITC experiments and further work needs to be done to investigate the reason for the higher affinity of EtdtCoA. Overall, this work has allowed the second mechanism of action to be validated for the known antibiotic N-pentylpantothenamide and confirms that this is an area to be explored for potential inhibitors of this biosynthetic pathway for future antibacterial development.

5   Summary and conclusions

Through the development and optimisation of sortase-mediated N-terminal labelling a number of proteins have been labelled for different uses within chemical and structural biology.

The combination of depsipeptide labelling substrates, known for enabling reversible labelling, and a more active sortase variant, Srt 7M, was investigated as an improved method of protein modification. Using MBP as a model system enabled the lower limits of protein and enzyme concentration to be found. An investigation was also carried out to analyse the effect of enzyme concentration and temperature on the rate of labelling and product hydrolysis. This was also observed by the use of a FRET assay which offered a continuous method of analysis while also providing evidence that the use of depsipeptides over peptides as labelling reagents enabled much higher levels of labelling to be achieved.

By using the information gained from these optimisation experiments, MBP was labelled with a bio-orthogonal handle, bicyclononyne. This was reacted through SPAAC with an azide-functionalised oligonucleotide, forming an MBP-oligonucleotide conjugate which was then incorporated into a DNA origami frame. This was used in HS-AFM, which provided images of the sortase 7M hydrolysis reaction cleaving the MBP from the DNA frame. A much longer strand of azide-functionalised double stranded DNA was made using a PCR approach to attach to the protein. This piece of work is ongoing for a potential application in improving the resolution of SAXS experiments.

The PanD/PanZ complex has been a system of interest within the Webb group for a number of years now. The sortase labelling method was able to be applied to this system to functionalise PanD with biotin using a biotin-depsipeptide. This enabled SPR experiments to be carried out to gain more information about the interactions between PanD, PanZ and the antimetabolite of known antibiotic N-pentyl panthothenamide known as EtdtCoA. A significant difference was observed in the interactions between PanD, PanZ and the two different CoA species, AcCoA and EtdtCoA, confirming the system as a target for N-pentyl pantothenamide.

Ultimately throughout this work the sortase-mediated N-terminal labelling approach developed has been used for different applications. This has allowed more information to be gained about peptide *versus* depsipeptide labelling substrates and hydrolysis of labelled product when using the heptamutant enzyme. It has also been used to introduce a bio-orthogonal handle to incorporate MBP within a DNA origami frame and used as a method of biotinylation for SPR. This provided proof that one of the mechanisms of action for the known antibiotic N-pentyl pantothenamide involves the PanD/PanZ system and inhibition of the pantothenic acid biosynthetic pathway.

6 Chemistry Experimental

6.1   Synthesis

6.1.1   General reagents and equipment

Reagents were obtained from commercial suppliers and used without further purification unless otherwise stated. Chromatography columns were prepared using Fisher Chemicals 60Å 35–70 micron silica gel. Nuclear magnetic resonance spectra were recorded using Bruker DPX400 and DPX500 MHz spectrometers. Chemical shifts are reported in parts per million (δ) downfield relative to the internal reference tetramethylsilane. NMR spectra were recorded in deuterochloroform at room temperature. Abbreviations used: Ar= aromatic, d= doublet, dd= doublet of doublets, m= multiplet, q= quartet, s= singlet, t= triplet (br)= broad signal. Mass spectra were recorded using a micromass ZMD 2000 spectrometer employing the electrospray (ES+) ionisation technique. Accurate molecular masses were obtained using a Bruker Daltonics MicroTOF mass spectrometer. Liquid chromatography mass spectrometry was obtained using Bruker HCTUltra mass spectrometer. Infra-red spectra were recorded using a Bruker Platinum ATR spectrometer. Lyophilisation was carried out using a Virtis Benchtop K freeze dryer.

6.1.2 Small molecule synthesis

2.03: Fmoc-Thr(OtBu)-Gc-OBn

Benzyloxycarbonylmethyl (2S, 3R)-3-*tert*butyloxy-2-(((9H-fluoren-9-ylmethoxy)carbonyl) aminobutanoate[81]
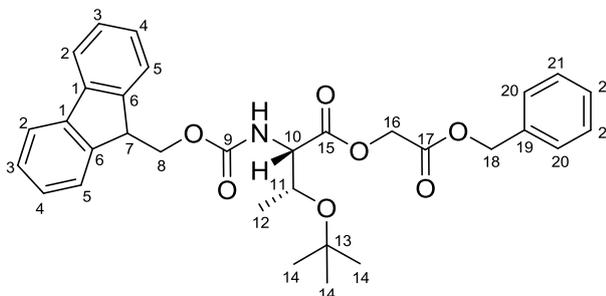
DIPEA (1.9 mL, 11 mmol) was added to a stirred solution of Fmoc-Thr(O^{t}Bu)-OH (4.0 g, 10 mmol), benzyl 2-bromoacetate (2.4 mL, 15 mmol) and tetrabutylammonium iodide (1.4 g, 4 mmol) in THF (10 mL). After a few minutes the reaction mixture turned yellow and a precipitate formed. The reaction mixture was stirred at room temperature for 2 hours before being diluted with H_2O (200 mL) and extracted with EtOAc (2 × 200 mL). The organic layers were combined and washed with 10 % sodium thiosulphate solution (400 mL) and H_2O (400 mL), dried (MgSO$_4$) and concentrated *in vacuo* to leave a yellow oil. The crude product was purified using flash chromatography (silica; 4:1 (*v/v* hexane-ethyl acetate) to afford **2.03** (4.83 g, 89 %) as a colourless solid.

¹H NMR (500 MHz, CDCl$_3$): δ 7.80 (2H, d, *J*= 7.2 Hz, *H$_2$*), 7.67 (2H, t, *J*= 8.1 Hz, *H$_5$*), 7.45-7.41 (2H, m, *H$_3$*), 7.41-7.38 (2H, m, *H$_4$*), 7.37-7.31 (5H, m, *H$_{20-22}$*), 5.66 (1H, d, *J*= 9.4 Hz, N*H*), 5.24 (2H, s, *H$_{18}$*), 4.79 (1H, d, *J*= 15.9 Hz, *H$_{16}$*), 4.69 (1H, d, *J*= 15.9 Hz, *H$_{16'}$*), 4.47-4.38 (1H, m, *H$_{10}$*), 4.47-4.38 (2H, m, *H$_8$*), 4.35-4.25 (1H, m, *H$_7$*), 4.35-4.25 (1H, m, *H$_{11}$*), 1.29 (3H, d, *J*= 6.1 Hz, *H$_{12}$*), 1.18 (9H, s, *H$_{14}$*). **¹³C NMR** (125 MHz, CDCl$_3$): δ 170.6 (CO), 167.0 (CO), 156.7 (CO), 144.1 (C$_{Ar}$), 141.3 (C$_{Ar}$), 128.7 (C$_{Ar}$), 128.6 (C$_{Ar}$), 128.4 (C$_{Ar}$), 127.6 (C$_{Ar}$), 127.1 (C$_{Ar}$), 125.2 (C$_{Ar}$), 120.0 (C$_{Ar}$), 74.3 (C$_{13}$), 67.3 (C$_8$), 67.2 (C$_{11}$), 61.3 (C$_{18}$), 59.8 (C$_{10}$), 47.2 (C$_7$), 28.4 (C$_{14}$), 20.9 (C$_{12}$). **IR** (v_{max}/cm^{-1}): 3437 (NH), 3014 (C-H), 1725 (C=O). **HRMS** [ES+] found [M+Na]$^+$ 568.2320, C$_{32}$H$_{39}$N$_2$O$_7$Na requires 568.2306. **R$_f$**: 0.63 (2:1 (*v/v*) hexane-ethyl acetate).

2.04: Fmoc-Thr-(OtBu)-Gc-OH

((2S, 3R)-3-*tert*butyloxy-2-(((9H-fluoren-9-ylmethoxy)carbonyl)amino)butanoyloxy) acetic acid[81]
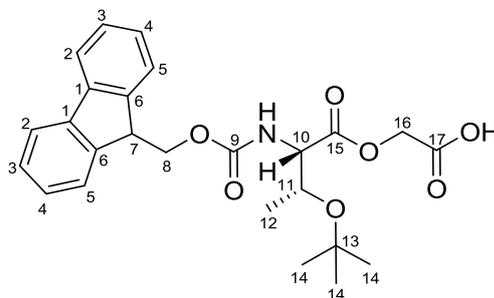
Fmoc-Thr(OtBu)-Gc-OBn (0.75 g, 1.30 mmol) was dissolved in MeOH/CH$_2$Cl$_2$ (20 mL, 9:1 (*v/v*)) and Pd/C (112 mg, 20 % (*w/w*)) was added to the reaction mixture which was then stirred for 10 minutes under H$_2$ atmosphere. The reaction mixture was then filtered through a celite pad which was washed with MeOH and the filtrate was concentrated *in vacuo* to leave **2.04** a colourless oil (0.56 g, 95 %).

^{1}H NMR (500 MHz, CDCl$_3$): δ 8.29-7.83 (1H, s(br), O*H*), 7.79 (2H, d, *J*= 7.6 Hz, *H$_2$*), 7.65 (2H, t, *J*= 7.4 Hz, *H$_5$*), 7.45-7.38 (2H, m, *H$_3$*), 7.37-7.32 (2H, m, *H$_4$*), 5.75 (1H, t(br), *J*= 7.5, N*H*), 4.77 (1H, d, *J*= 16.3 Hz, *H$_{16}$*), 4.65 (2H, d, *J*= 14.5, *H$_{16'}$*), 4.43 (2H, d, *J*= 6.5 Hz, *H$_8$*), 4.34-4.23 (1H, m, *H$_{10}$*), 4.34-4.23 (1H, m, *H$_7$*), 4.34-4.23 (1H, m, *H$_{11}$*), 1.26 (3H, d, *J*= 6.1 Hz, *H$_{12}$*), 1.18 (9H, s, *H$_{14}$*). **^{13}C NMR** (125 MHz, CDCl$_3$): δ 170.4 (CO), 156.9 (CO), 149.0 (CO), 143.8 (C$_{Ar}$), 141.3 (C$_{Ar}$), 127.7 (C$_{Ar}$), 126.9 (C$_{Ar}$), 125.2 (C$_{Ar}$), 120.0 (C$_{Ar}$), 74.6 (C$_{13}$), 67.4 (C$_{16}$), 67.3 (C$_{11}$), 61.3 (C$_8$), 59.8 (C$_{10}$), 47.2 (C$_7$), 28.4 (C$_{14}$), 20.7 (C$_{12}$). **IR** (v_{max}/cm^{-1}): 3432 (N-H), 3400-2600 (OH, broad), 1723 (C=O). **HRMS** [ES+] found [M+Na]$^+$ 478.1841, C$_{25}$H$_{30}$NO$_7$Na requires 478.1836. **R$_f$**: 0.19 (94:5:1 (*v/v*) CH$_2$Cl$_2$/MeOH/AcOH).

3.02: (1R, 8S, 9R, Z)-Ethyl bicyclo[6.1.0]non-4-ene-9-carboxylate (*anti*) & (1R, 8S, 9S, Z)-Ethyl bicyclo[6.1.0]non-4-ene-9-carboxylate (*syn*)[28]
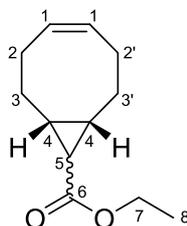
Rhodium (II) acetate (47 mg, 0.16 mmol) was added to a dry round-bottomed flask, containing a magnetic stirrer, and equipped with a N_2 inlet adaptor. The flask was evacuated and refilled with N_2 three times to ensure a N_2 atmosphere. 1, 5-cyclooctadiene (20 mL, 163 mmol) was added to the flask *via* syringe and stirred vigorously. Ethyl diazoacetate (2.4 mL, 22 mmol) was added *via* syringe pump over 12 hours. The reaction mixture was concentrated *in vacuo* and purified using flash chromatography (silica, 0-1.5 % ether-hexane) to yield **3.02 (*anti*)** (1.67 g, 39 %) and **3.02 (*syn*)** (1.94 g, 45 %) as colourless oils.

3.02 (*anti*): ¹H NMR (500 MHz, CDCl$_3$): δ 5.72-5.59 (2H, m, H_1), 4.13 (2H, q, *J* = 7.1 Hz, H_7), 2.37-2.29 (2H, m, H_2), 2.26-2.18 (2H, m, H_3), 2.16-2.07 (2H, m, $H_{2'}$), 1.62-1.56 (2H, m, H_4), 1.55-1.47 (2H, m, $H_{3'}$), 1.28 (3H, t, *J* = 7.1, H_8), 1.21 (1H, t, *J* = 4.6 Hz, H_5). **¹³C NMR** (125 MHz, CDCl$_3$): δ 174.4 (C$_6$), 129.9 (C$_1$), 60.3 (C$_7$), 28.3 (C$_2$), 27.9 (C$_5$), 27.7 (C$_4$), 26.7 (C$_3$), 14.3 (C$_8$). **IR** (ν$_{max}$/cm^{-1}): 2934 (C-H), 1718 (C=O). **R$_f$**: 0.42 (10 % ether-hexane).

3.02 (*syn*) : ¹H NMR (500 MHz, CDCl$_3$): δ 5.68-5.61 (2H, m, H_1), 4.15 (2H, q, *J* = 7.1 Hz, H_7), 2.58-2.49 (2H, m, H_2), 2.28-2.19 (2H, m, H_3), 2.14-2.04 (2H, m, $H_{2'}$), 1.96-1.82 (2H, m, $H_{3'}$), 1.74 (1H, t, *J* = 8.8 Hz, H_5), 1.47-1.38 (2H, m, H_4), 1.29 (3H, t, *J* = 7.1 Hz, H_8). **¹³C NMR** (125 MHz, CDCl$_3$): δ 172.1 (C$_6$), 129.5 (C$_1$), 59.7 (C$_7$), 27.1 (C$_2$), 24.2 (C$_4$), 22.7 (C$_3$), 21.3 (C$_5$), 14.4 (C$_8$). **IR** (ν$_{max}$/cm^{-1}): 2936 (C-H), 1719 (C=O). **R$_f$**: 0.56 (10 % ether-hexane).

3.04 (*anti*): (1R, 8S, 9R)- bicyclo[6.1.0]non-4-yn-9-ylmethanol[28]
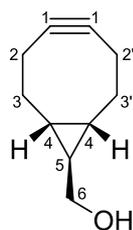
3.02 (*anti*) (0.74 g, 3.81 mmol) was dissolved in dry THF (15 mL) with stirring and cooled to -78 °C using a dry ice bath. A solution of LiAlH$_4$ (3.81 mL, 1M in THF, 3.81 mmol) was added dropwise and stirred at RT for 1 hour. The reaction mixture was quenched at 0 °C with the cautious, dropwise addition of H$_2$O (0.15 mL), 15 % solution NaOH (0.15 mL) and finally H$_2$O (0.45 mL). A solid precipitate formed which was removed under reduced pressure and washed with THF. The resulting filtrate was concentrated *in vacuo*, dissolved in CH$_2$Cl$_2$ and washed with brine (10 mL), dried (MgSO$_4$) and concentrated again *in vacuo* to yield **3.03 (*anti*)** as a colourless oil. Under an N$_2$ atmosphere, the oil was dissolved in dry CH$_2$Cl$_2$ (7 mL) and stirred at 0 °C to which a solution of Br$_2$ (153 μL) in dry CH$_2$Cl$_2$ (2 mL) was added dropwise until a persistent yellow/orange colour was observed. The reaction mixture was quenched with 10 % sodium thiosulphate solution (5 mL) and the organic layer was separated and washed with H$_2$O (10 mL), brine (10 mL), dried (MgSO$_4$) and concentrated *in vacuo* to leave the dibromide intermediate as a colourless solid in oil. Without further purification, the dibromide was dissolved in THF (25 mL) with stirring and cooled to 0 °C. A solution of KO^{t}Bu (6.90 mL, 1M in THF, 6.90 mmol) was added dropwise. The reaction mixture was then heated under reflux at 67 °C for 2 hours. After being left to cool to RT, the mixture was quenched with saturated NH$_4$Cl-solution (25 mL) and extracted with CH$_2$Cl$_2$ (3 × 50 mL). The organic layers were combined, dried (MgSO$_4$) and concentrated *in vacuo* to leave the crude product as a yellow oil. Purification by flash chromatography (silica; 2:1 (*v/v* hexane-ethyl acetate) yielded **3.04 (*anti*)** as a colourless oil (68 mg, 20 %).

^{1}H NMR (500 MHz, CDCl$_3$): δ 3.55 (2H, d, *J*= 6.3 Hz, *H$_6$*), 2.42 (2H, dd, *J*= 13.3, 2.5 Hz, *H$_2$*), 2.34-2.25 (2H, m, *H$_3$*), 2.16-2.14 (2H, m, *H$_{2'}$*), 1.67 (1H, s(br), O*H*), 1.44-1.34 (2H, m, *H$_{3'}$*), 0.74-0.63 (3H, m, *H$_4$, H$_5$*). **^{13}C NMR** (125 MHz, CDCl$_3$): δ 98.8 (C$_1$), 67.2 (C$_6$), 33.5 (C$_2$), 27.4

(C$_5$), 22.6 (C$_4$), 21.5 (C$_3$). **IR** (ν_{max}/cm^{-1}): 3359 (OH), 2972-2857 (C-H). **R$_f$**: 0.13 (2:1 (*v/v*)

hexane/ethyl acetate).

3.04 (*syn*): (1R, 8S, 9S)- bicyclo[6.1.0]non-4-yn-9-ylmethanol[28]
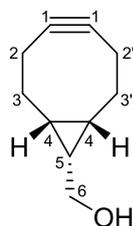
This was done using the same protocol described for **3.04 (*anti*)** and yielded **3.04 (*syn*)** as a

colourless solid (34 mg, 10 %).

^{1}H NMR (500 MHz, CDCl$_3$): 3.67 (2H, d, *J*= 8.0 Hz, **H$_6$**), 2.28-2.09 (6H, m, **H$_2$, H$_{2'}$, H$_3$**),

1.59-1.51 (2H, m, **H$_{3'}$**), 1.32-1.24 (1H, m, **H$_5$**), 1.08 (1H, s(br), O**H**), 0.90-0.85 (2H, m, **H$_4$**).

^{13}C NMR (125 MHz, CDCl$_3$): δ 98.9 (C$_1$), 60.0 (C$_6$), 29.1 (C$_2$), 21.5 (C$_4$), 21.4 (C$_5$), 20.1 (C$_3$).

R$_f$: 0.13 (2:1 (*v/v*) hexane/ethyl acetate).

3.05 (*syn*): (1R, 8S, 9S)-bicyclo[6.1.0]non-4-yn-9-yl-methyl(4-nitrophenyl)carbonate[28]
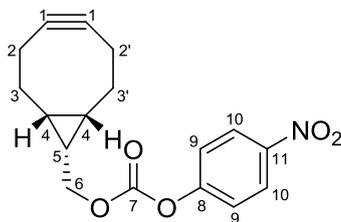

3.04 (*syn*) (34 mg, 0.23 mmol) was dissolved in dry CH$_2$Cl$_2$ (5 mL). Pyridine (46 μL, 0.58 mmol)

and *p*-nitrophenyl chloroformate (56 mg, 0.28 mmol) were added and the mixture was stirred at

RT for 20 minutes. Once complete, the mixture was quenched with saturated NH$_4$Cl-solution

(5 mL) and extracted with CH$_2$Cl$_2$ (3 × 5 mL). The combined organic extracts were dried

(MgSO$_4$) and concentrated *in vacuo*. The residue was purified using flash chromatography (silica;

8:1 (*v/v* hexane-ethyl acetate) to yield **3.05 (*syn*)** as a colourless solid (50 mg, 57 %).

^{1}H NMR (500 MHz, CDCl$_3$): δ 8.29 (2H, d, *J*= 9.1 Hz, **H$_{10}$**), 7.40 (2H, d, *J*= 9.3 Hz, **H$_9$**), 4.41

(2H, m, **H$_6$**), 2.38-2.22 (6H, m, **H$_2$, H$_{2'}$, H$_3$**), 1.66-1.57 (2H, m, **H$_{3'}$**), 1.53-1.48 (1H, m, **H$_5$**),

1.11-1.03 (2H, m, H_4). ^{13}**C NMR** (125 MHz, CDCl$_3$): δ 155.6 (C$_7$), 152.6 (C$_{11}$), 145.4 (C$_8$), 125.3 (C$_{10}$), 121.8 (C$_9$), 98.7 (C$_1$), 68.0 (C$_6$), 29.7 (C$_2$), 21.4 (C$_4$), 20.5 (C$_5$), 17.3 (C$_3$). **IR** (ν_{max}/cm^{-1}): 2977-2915 (C-H), 1757 (C=O). **R$_f$**: 0.30 (9:1 *v/v* hexane/ethyl acetate).

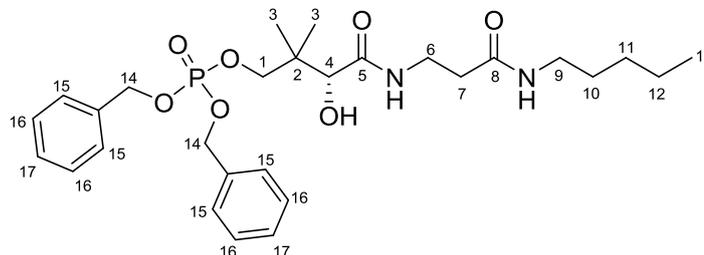### 4.01b: *N*-pentylpantothenamide[157]
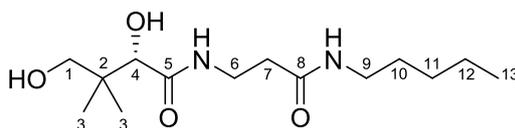
Calcium pantothenate (2.0 g, 8.3 mmol) was dissolved in H$_2$O (~20 mL) and passed through an Amberlite IR-120 (H$^+$) column. This was washed with 2 × 20 mL H$_2$O and eluents were combined and lyophilised overnight to leave pantothenic acid as a colourless oil (1.4 g). The oil was dissolved in DMF (10 mL) and pentylamine (1.16 mL, 10 mmol) and diphenylphospharyl azide (2.24 mL, 10 mmol) were added. The mixture was cooled to 0 °C and triethylamine (1.39 mL, 1.2 mmol) was added followed by stirring for 2 hours at 0 °C and a further 5 hours at RT. The DMF was removed under reduced pressure to leave a colourless oil which was purified by flash chromatography three times (silica; 5 % *v/v* methanol-CH$_2$Cl$_2$ and gradient elution 1-10 % *v/v* methanol-CH$_2$Cl$_2$) due to co-elution in the presence of traces of DMF. Fractions were combined and evaporated under reduced pressure to yield a colourless oil which was lyophilised overnight to leave a colourless solid on standing (0.84 g, 35 %).

^{1}**H NMR** (500 MHz, CDCl$_3$): δ 7.45 (1H, t(br), *J*= 6.3 Hz, N*H*), 6.09 (1H, t(br), *J*= 5.4 Hz, N*H*), 4.26 (1H, d, *J*= 5.5 Hz, O*H*), 3.99 (1H, d, *J*= 5.3 Hz, O*H*), 3.73 (1H, s(br), *H$_4$*), 3.56 (2H, m, *H$_1$*), 3.49 (2H, d, *J*= 5.5 Hz, *H$_6$*), 3.21 (2H, q, *J*= 7.2, *H$_9$*), 2.43 (2H, t, *J*= 6.0 Hz, *H$_7$*), 1.49 (2H, m, *H$_{10}$*), 1.30 (4H, m, *H$_{11}$, H$_{12}$*), 1.00 (3H, s, C*H$_3$*), 0.93 (3H, s, C*H$_{3'}$*), 0.90 (3H, t, *J*= 6.7, C*H$_{13}$*). ^{13}**C NMR** (125 MHz, CDCl$_3$): δ 173.4 (C$_{5/8}$), 171.2 (C$_{5/8}$), 77.7 (C$_4$), 70.9 (C$_1$), 39.7 (C$_7$), 39.4 (C$_2$), 35.8 (C$_9$), 35.4 (C$_6$), 29.1 (C$_{10}$), 29.0 (C$_{11/12}$), 22.2 (C$_{11/12}$), 21.4 (C$_3$), 20.4 (C$_{3'}$), 13.8 (C$_{13}$). **IR** (ν_{max}/cm^{-1}): 3280 (N-H), 3000-2800 (OH, broad), 1 642 (C=O). **HRMS** [ES+] found [M+Na]$^+$ 311.1945, C$_{25}$H$_{30}$NO$_7$Na requires 311.1941. **R$_f$**: 0.15 (10 % MeOH in CH$_2$Cl$_2$).

The following synthesis was carried out by Holly Morgan for the preparation of 4'-phospho-N-pentylpantothenamide:

**4.07: N-pentylpantothenamide-4'-O-dibenzylphosphate**[157]

N-pentylpantothenamide (300 mg, 1.04 mmol) was dissolved in dry acetonitrile (12 mL) under a N_2 atmosphere. A solution of tetrazole in acetonitrile (0.45 M, 2.31 mL, 1.04 mmol) and dibenzyl-N, N-diisopropyl-phosphoramidite (0.35 mL, 1.04 mmol) were added and the reaction mixture was stirred for 5 min at room temperature. mCPBA (269.2 mg, 1.56 mmol) was then added and the reaction mixture was stirred for 2h at room temperature. The product was concentrated *in vacuo*, extracted with ethyl acetate (3 × 20 mL) and then washed with 1M H_2SO_4 (2 × 5 mL) and 1M $NaHCO_3$ (2 × 5 mL). The resulting solution was dried with $MgSO_4$ and concentrated *in vacuo*. The product was purified by flash column chromatography on silica gel (7 % (*v/v*) methanol in CH_2Cl_2) to yield **4.07** as a colourless oil (415.3 mg, 73 %)

¹H NMR (500 MHz, CDCl₃): δ 7.57 (1H, t(br), J = 6.2 Hz, N*H*), 7.51 (1H, s(br), O*H*), 7.49 (1H, s(br), O*H*), 7.42-7.24 (10H, m, H_{15}, H_{16}, H_{17}), 6.87 (1*H*, t(br), J = 5.7 Hz, N*H*), 5.16-4.98 (4H, m, H_{14}), 4.04 (1H, dd, J = 9.9, 6.1 Hz, H_1), 3.97 (1H, s, H_4), 3.72 (1*H*, dd, J = 9.9, 6.1 Hz, H_1) 3.57 (2H, q, J = 6.2 Hz, H_6), 3.2 (2H, q, J = 6.5 Hz, H_7), 2.56-2.53 (2H, m, H_9), 1.48 (2H, p, J = 7.3 Hz, H_{10}), 1.33-1.22 (4H, m, H_{11}, H_{12}), 1.05 (3H, s, H_3), 0.91-0.83 (6H, m, H_3, H_{13}). **IR** (v_{max}/cm⁻¹): 3300 (N-H), 3095-2873 (OH, broad), 1638 (C=O). **LCMS** (ES+): RT = 0.6-0.7 min, m/z = 549.35 (M+H)⁺.

4.08: 4'-Phospho-N-pentylpantothenamide[157]
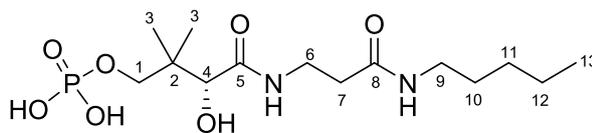
N-pentylpantothenamide-4'-O-dibenzylphosphate (415.3 mg, 0.75 mmol) was dissolved in 9:1 CH_3OH/H_2O (20 mL) and stirred under an N_2 atmosphere. The catalyst, Pd/C (10 % (*wt/wt*), 41.5 mg) was added and the suspension was stirred at atmospheric pressure and room temperature for 6h. More Pd/C (10 % (*wt/wt*), 41.5 mg) was then added and the reaction mixture was stirred for a further 2h. The solution was filtered through a celite pad before being concentrated *in vacuo*. The product was then extracted with water (3 × 20 mL) and lyophilised to yield **4.08** as an orange solid (48 mg, 17 %).

^{1}H NMR (500 MHz, D_2O): δ 3.95 (1H, s, H_4), 3.78 (1H, dd, J = 9.7, 4.8 Hz, H_1), 3.59 (1H, dd, J = 9.7, 4.8 Hz, H_1) 3.43 (2H, m, H_6), 3.07 (2H, m, H_7), 2.4 (2H, t, J = 6.4 Hz, H_9), 1.41 (2H, p, J = 7.0 Hz, H_{10}), 1.21 (4H, m, H_{11}, H_{12}), 0.9 (3H, s, H_3), 0.83 (3H, s, H_3), 0.79 (3H, t, J = 6.8 Hz, H_{13}). ^{13}C NMR (100 MHz, D_2O): δ 174.79, 173.60, 133.80, 129.61, 128.70, 74.40, 71.59, 39.51, 38.33, 28.33, 27.90, 21.68, 20.58, 18.60, 13.24; **^{31}P NMR** (400 MHz, D_2O): δ 0.12. **IR** (v_{max}/cm^{-1}): 3310 (N-H), 3097-2871 (OH, broad), 1637 (C=O). **LCMS** (ES+): RT = 0.4-0.5 min, m/z = 369.26 (M+H)$^+$, 737.35 (2M+1)$^+$, 367.25 (M-1)$^-$, 735.47 (M-1)$^-$.

4.02b: Ethyldethia Coenzyme A[157,164]
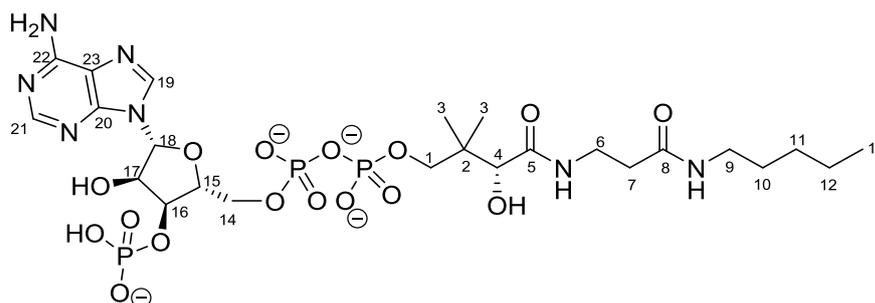
4'-Phospho-N-pentylpantothenamide (10.5 mM, **4.08** provided by Holly Morgan) was added to a mixture of both CoaD (9 µM) and CoaE (11 µM) in 50 mM Tris HCl buffer (pH 7.6) in the presence of $MgCl_2$ (5 mM) and ATP (50 mM) in a reaction volume of 900 µL. They were mixed thoroughly before being incubated at 37 °C for 2-4 hours while production of the EtdtCoA was analysed by LCMS. The CoaD and CoaE were precipitated from the mixture by heating at 95 °C for 15 minutes and removed by centrifugation. The supernatant was purified by reverse phase HPLC (C_{18}-silica column) using a 5-95 % 50 mM KH_2PO_4 (pH 5.3)/MeOH gradient to allow separation of the ATP/AMP from the EtdtCoA. The MeOH was removed *in vacuo* and the samples were adjusted to pH 7.6 using 5 M NaOH before being lyophilized. The product was resuspended in H_2O and put through Sephadex® G-10 desalting columns to remove the KH_2PO_4 from the mixture, the elution of EtdtCoA was monitored by absorbance at 257 nm. Fractions containing the EtdtCoA were then lyophilized again to be dissolved in an appropriate amount of buffer for use in ITC/SPR. Carrying out this reaction on a 900 µL scale afforded 1.54 mg (21 %) EtdtCoA after HPLC purification and desalting columns. The concentration of EtdtCoA was calculated using A_{257} and extinction coefficient of 16,840 M^{-1} cm^{-1} obtained from Dawson *et al.*[168]

^{1}H NMR (500 MHz, D_2O): δ 8.45 (1H, s, H_{21}), 8.17 (1H, s, H_{19}), 6.07 (1H, d J= 6.9, H_{18}), 4.72 (1H, br s, H_{17}), 4.47 (1H br s, H_{16}), 4.13 (2H, br s, H_{14}), 3.90 (1H, s, H_{15}), 3.71 (1H, dd, J= 9.5,4.9 Hz, $H_{1'}$), 3.43 (1H, dd, J= 9.5,4.9 Hz, $H_{1''}$), 3.37 (1H, dt, J= 14, 6.7 Hz, $H_{6'}$), 3.33 (1H, dt, J= 14, 6.7 Hz, $H_{6''}$), 2.98 (2H, dt J= 6.9, 3.0 Hz, H_9), 2.33 (2H, t, J= 6.6 Hz, H_7), 1.31 (2H, p, J= 7.0 Hz, H_{10}) 1.08 (4H, m, $H_{11,12}$), 0.76 (3H, s, $H_{3'}$), 0.72 (3H, t, J= 6.9 Hz, H_{13}), 0.62 (3H, s, $H_{3''}$);

^{13}C NMR (150 MHz, D2O) δ 174.8, 173.6, 155.8, 153.1, 149.5, 140.0, 118.7, 86.6, 84.0, 74.4,

73.8, 72.0, 62.6, 39.6, 38.4, 38.3, 35.6, 35.5, 28.4, 28.0, 21.7, 20.9, 18.1, 13.3; **^{31}P NMR** (162 MHz, D2O) δ 3.49 (s), -10.68 (d, *J*= 20.1 Hz), -11.21 (d, *J*= 20.1 Hz) **HRMS** [ES+] found [M+H]$^+$ 778.1961, $C_{24}H_{43}N_7O_{16}P_3$ requires 778.1934.

6.2   Solid Phase Peptide Synthesis

6.2.1   General reagents and equipment

All amino acids, resins and coupling reagents were purchased from Sigma Aldrich, VWR International and Cambridge Bioscience. All chemical compounds were used without further purification. Fritted polypropylene tubes (10 mL) purchased from Grace and Co were used as a vessel for all solid phase reactions. Agitation of the solid phase reaction mixture was achieved by rotation on a Stuart rotator. H-Gly-2-Cl-Trt (Sigma, loading: 1.1 mmol/g) was used for the solid phase peptide synthesis.

Peptides were analysed using an Agilent 1290 affinity LC system equipped with an Ascentis Peptide ES C18, 100 x 2.1mm, 2.7um particle size (0.5 mL min^{-1}) and diode array detection. Gradient from 0.1 % TFA/5 % MeCN (*v/v*) in H$_2$O to 0.1 % TFA/95 % MeCN (*v/v*) in H$_2$O over 10 mins.

6.2.2   General methods for solid phase peptide synthesis

Method One: Preparation of resin and amino acid coupling

H-Gly-2-chloro-trityl resin was swollen in DMF (6 mL) for 30 minutes. Fmoc-amino acid (5 equiv.), OxymaPure (5 equiv.) and DIC (5 equiv.) were dissolved in DMF (6 mL) and added to the swollen resin. The mixture was spun for 1 hour at room temperature. After the coupling reaction the solution was removed by filtration and the resin was washed with DMF (3×6 mL $\times$ 2 min spins).

Method Two: Analysis of coupling

A small sample of resin was isolated and exposed to TFA (100 μL) for 2 minutes before being quenched with methanol (1 mL). The solution was filtered and analysed by LCMS, if any starting material was seen, method one was repeated.

Method Three: Fmoc deprotection

After the DMF wash, the Fmoc-amino acid-resin was washed with piperidine solution (20 % (*v/v*) in DMF, 6 mL) and was spun for 2 minutes which was repeated 5 times. The resin was then washed again with DMF (5 × 6 mL × 2 min spins). Subsequent couplings were done in the same manner as mentioned in steps 1-3 (excluding the resin swelling) until the desired peptide was synthesised.

Method Four: Resin capping (optional)

If the Fmoc-amino acid coupling step is inefficient, the resin was capped with a solution containing acetic anhydride (5 equiv.), DIPEA (5 equiv.) in DMF (6 mL) which was spun for 1 hour. The solution was removed by filtration and the resin was washed with DMF (3 × 6 mL × 2 min spins). This step would be carried out between methods two and three.

Method Five: Cleavage from resin and peptide isolation

After synthesising the desired peptide on the resin, it was washed with DMF (3 × 6 mL × 2 min spins), CH_2Cl_2 (3 × 6 mL × 2 min spins), and methanol (3 × 6 mL × 2 min spins). The resin was isolated by filtration before being dried under high vacuum for 3 hours. A cleavage cocktail containing TFA : H_2O : TIS (95 : 2.5 : 2.5) (3 mL) was made, 2 mL was added to the dried resin and the mixture was spun for 2 hours. The washings from the resin were filtered into cold diethyl ether (40 mL) to precipitate the peptide. The remaining cleavage cocktail was added and spun for another 5 minutes and filtered into the cold ether. The precipitate was pelleted by centrifugation (4,000 × *g*, 10 minutes) and the cold diethyl ether was decanted. The pellet was re-suspended in cold diethyl ether (40 mL), this process was repeated three times and the residual ether was

removed under a stream of nitrogen. The peptide was analysed by LCMS before being dissolved

in a minimal amount of water and freeze dried overnight.

6.2.3 Synthesised peptides

BCN-GYLPEToGG-COOH
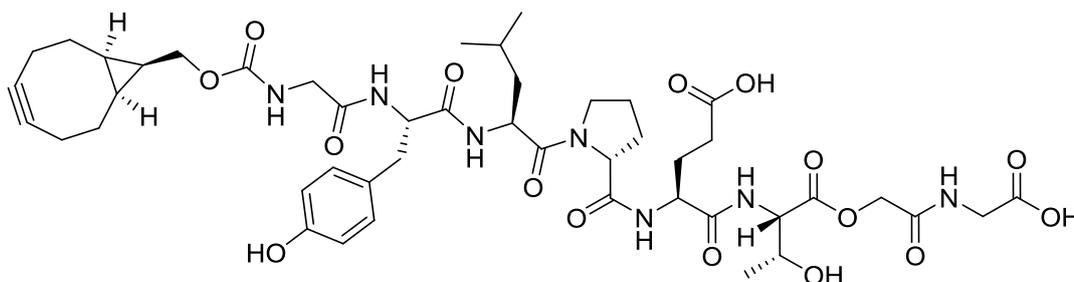
Depsipeptide was synthesised using methods 1-5. 20 mg (0.03 mmol) was dissolved in DMF

(~3 mL) and *para*-nitrophenyl bicyclo[6.1.0]nonyne (12 mg, 0.04 mmol) and triethylamine

(5 μL, 0.04 mmol) was added. The mixture was stirred at room temperature overnight. The DMF

was removed under reduced pressure and the crude peptide was dissolved in dioxane and

lyophilized to leave a fluffy white solid (20 mg, 74 %). LCMS [ES+] found [M+Na]$^+$ 992.4238,

$C_{46}H_{63}N_7O_{16}Na$ requires 992.4224.
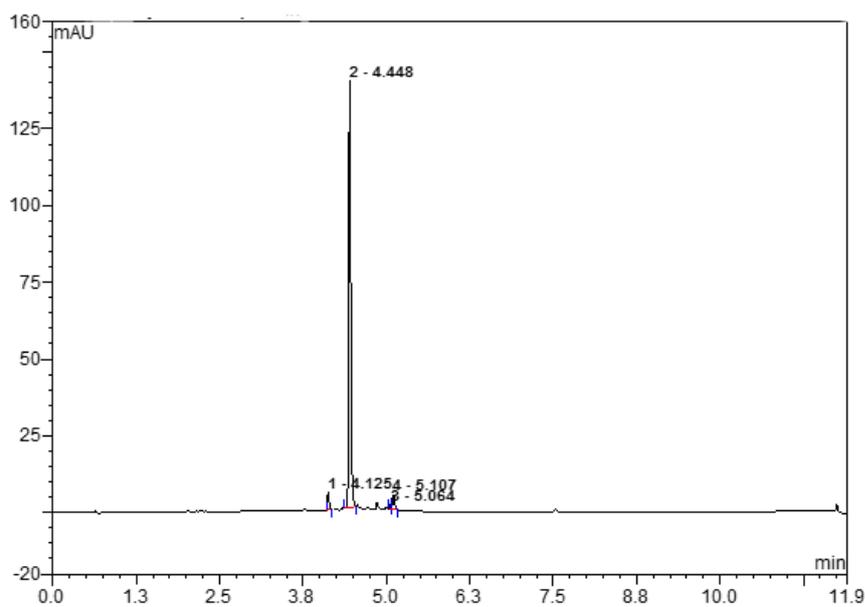
FITC-GABA-LPEToGG-COOH
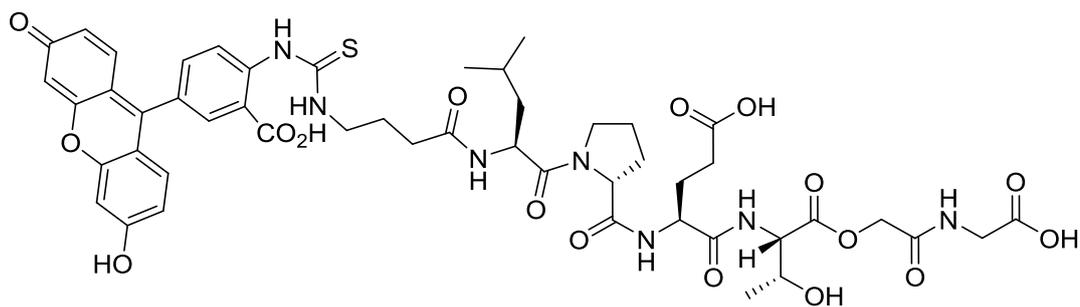
Once the peptide had been synthesised using methods 1-4 and a final coupling with a GABA residue, the peptide was treated with a solution of FITC (6 equiv.) and DIPEA (14 equiv.) in DMF. The resin was spun overnight at room temperature in darkness. The resin was washed with DMF (4×10 mL $\times$ 2 min spins) before method five was carried out. Yield: (40 mg, 53 %) LCMS [ES+] found [M+2H]$^{2+}$ 525.05, expected [M+2H]$^{2+}$ requires 525.69 and $C_{49}H_{59}N_7O_{17}S$ requires 1047.35.
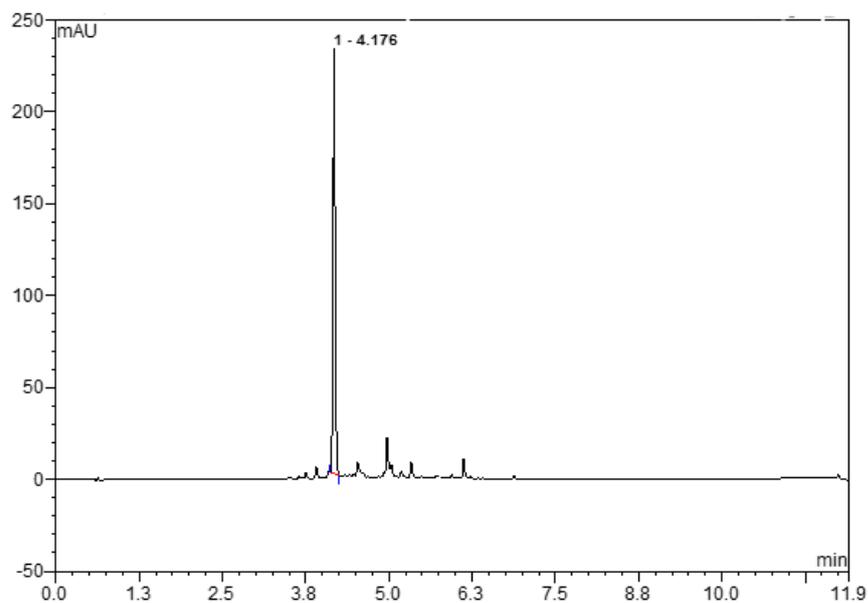
H₂N-LPEToGG-COOH
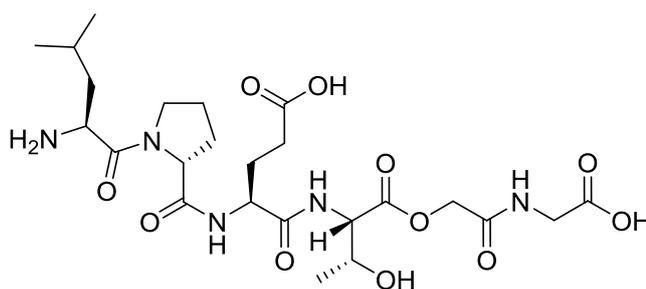
Depsipeptide was synthesised using methods 1-5. Yield: 11 mg, 52 %. LCMS [ES+] found [M+H]⁺ 574.63, $C_{24}H_{40}N_5O_{11}$ requires 574.27.
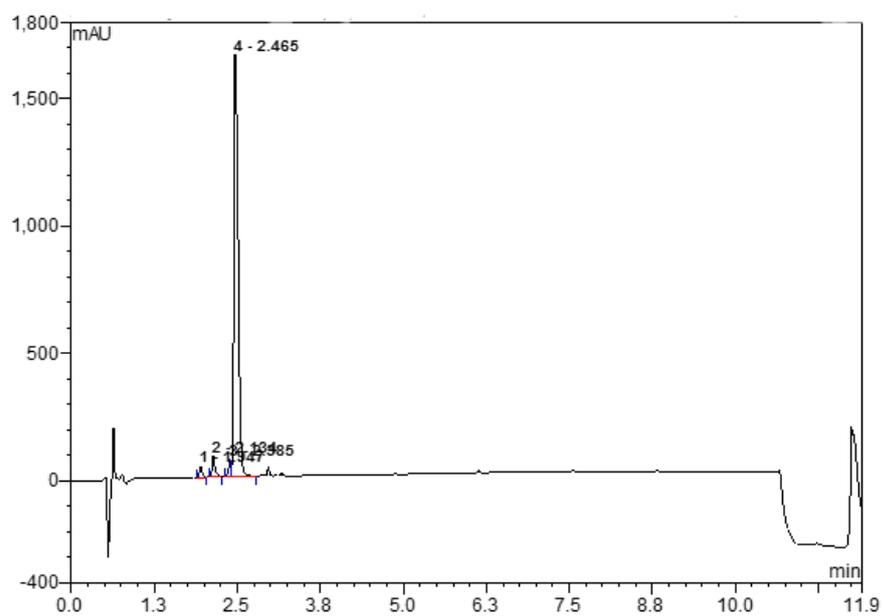
FITC-GABA-LPETGG-COOH
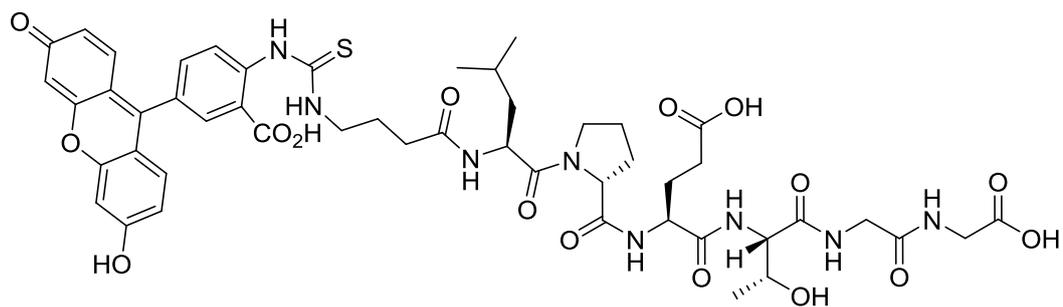
Once the peptide had been synthesised using methods 1-4 and a final coupling with a GABA residue, the peptide was treated with a solution of FITC (6 equiv.) and DIPEA (14 equiv.) in DMF. The resin was spun overnight at room temperature in darkness. The resin was washed with DMF (4×10 mL $\times 2$ min spins) before method five was carried out. Yield: (76 mg, 66 %) LCMS [ES+] found $[M+2H]^{2+}$ 524.54, expected $[M+2H]^{2+}$ requires 525.70 and $C_{49}H_{60}N_8O_{16}S$ requires 1046.37.
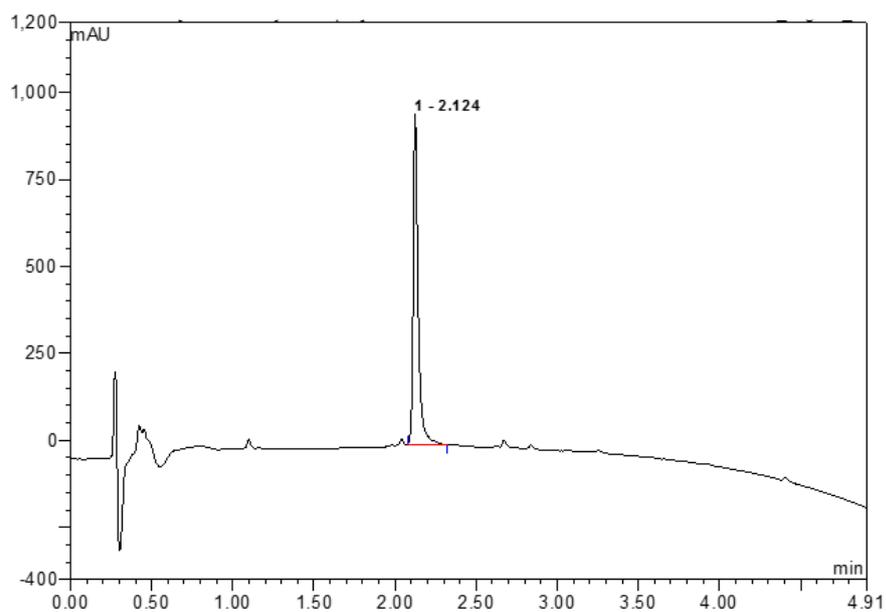
H₂N-LPETGG-COOH
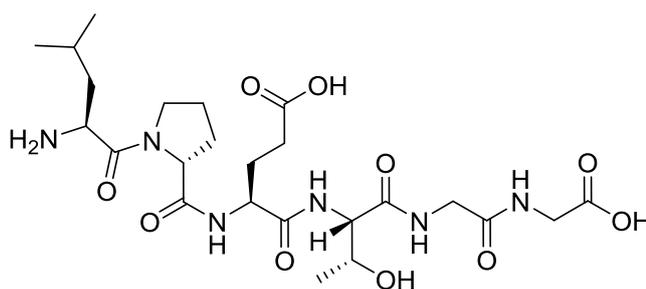
This peptide was synthesised using methods 1-5. Yield: 38 mg, 89 %. LCMS [ES+] found [M+H]$^+$ 573.65, $C_{24}H_{41}N_6O_{10}$ requires 573.62.

Biotin-GYLPEToGG-COOH (Sythesised by Jack Caudwell)

Once the peptide had been synthesised using methods 1-4 and a final coupling with a solution of Biotin-NHS ester (5 equiv.) and DIPEA (14 equiv.) in DMF. The resin was spun overnight at room temperature. The resin was washed with DMF (4×10 mL $\times$ 2 min spins) before cleavage was carried out. The cleavage cocktail was collected in a flask and toluene (2 mL) was added before being concentrated *in vacuo* to leave a brown solid which was dissolved in H_2O and lyophilised. LCMS [ES+] found [M+H]$^+$ 1020.56, $C_{45}H_{65}N_9O_{16}S$ requires 1019.43.

H₂N-GVSK-Rhod-COOH (Synthesised by Kristian Hollingsworth)
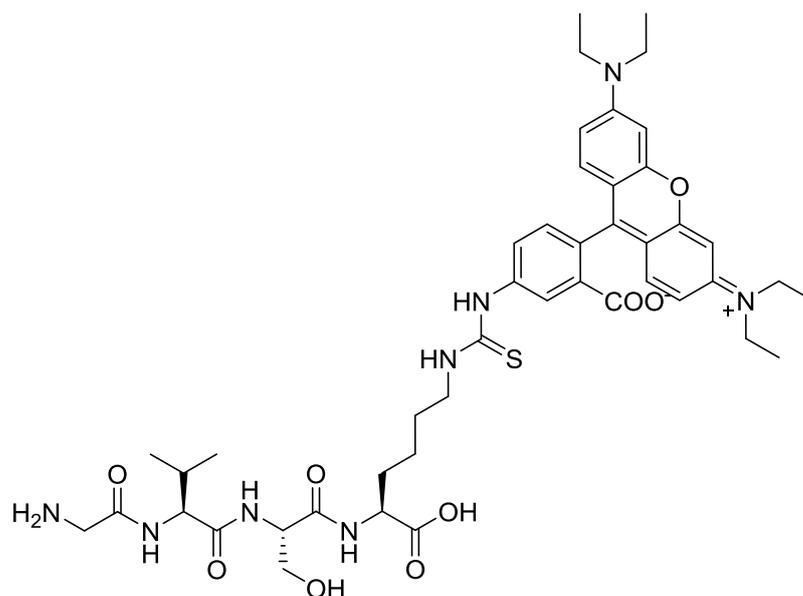
Once the peptide had been synthesised using methods 1-4. The 1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl (Dde) protecting group on lysine was removed using 2 % hydrazine in DMF followed by a final coupling a solution of Rhodamine isothiocyanate (5 equiv.) and DIPEA (14 equiv.) in DMF. The resin was spun overnight at room temperature in darkness. The resin was washed with DMF (4 × 10 mL × 2 min spins) before method five was carried out. LCMS [ES+] found [M+2H]$^{2+}$ 445.50, expected [M+2H]$^{2+}$ is 446.23 and $C_{45}H_{60}N_8O_9S$ requires 888.42.

7 Biochemical Experimental

7.1 General methods and equipment

Sterilisation of media, buffers and appropriate equipment was achieved using either a Prestige Medical bench top autoclave or a LTE Touchclave-R autoclave. Thermo Electron Corporation SAFE 2010 Class II laminar flow cabinet or a bench-top bunsen burner were used to maintain a sterile environment. Bacterial cultures were incubated in a Kuhner ShakerX ISF1-X or Stuart Orbital Incubator. LB-agar plates and small-scale assays were incubated in a Binder BD23 incubator.

Centrifugation was performed using either a Beckman Coulter™ Avanti™ JXN-30 centrifuge, a Heraeus multifuge 3 S-R centrifuge or a Heraeus pico centrifuge. Gel filtration chromatography was achieved using a GE Pharmacia ÄKTA FPLC system or Bio-Rad NGC Chromatography system. Spectrophotometric readings were measured using either a ThermoFisher Scientific Nanodrop 2000. SDS-PAGE was carried out using a Bio-Rad mini protean 3 apparatus. Bio-Rad molecular imagerR Gel DocTM XR was used to visualise polyacrylamide gels using a combination of UV and white light. Protein concentrations were conducted using a 10k MWCO Amicon® Ultra-15 Centrifugal Filter Device. Mass spectrometry analysis of proteins was done using a using a Bruker Daltonics MicroTOF mass spectrometer. Protein samples were loaded directly at a concentration of 50 µM (made up with H_2O) into the instrument before they were automatically diluted into 0.1 % TFA/50 % MeCN (*v/v*) in H_2O prior to mass spectrometry analysis. Analytical grade reagents were supplied by Sigma-Aldrich, Fisher Scientific, Melford laboratories and VWR International. All recipe components were dissolved in 18.2 MΩ H_2O to the final volume stated. The pH of the solutions was adjusted using 1 M NaOH or 5 M HCl.

7.2   Media and buffers

Buffers and media were made in-house unless otherwise stated. Gel filtration buffers were filtered through a 0.22 µm membrane under reduced pressure. Media was sterilised by autoclave at 120 °C for 20 minutes, buffers were used without sterilisation, unless otherwise stated.

7.2.1   Growth media

LB media: 25 g L^{-1} LB freeze-dried powder (Fisher) in H_2O

LB-agar media: 25 g L^{-1} LB powder (Fisher) and 15 g L^{-1} of Agar powder (Fisher) in H_2O

Auto induction media: 10 g L^{-1} tryptone, 5 g L^{-1} yeast extract, 4 % (*v/v*) 25 × salts, 1000 × metals and 4.3 mM $MgCl_2$ in H_2O. The solution was autoclaved at 120 °C for 20 minutes before 1 mL L^{-1} 50 % (*w/v*) glucose, 10 mL L^{-1} 20 % (*w/v*) lactose and 10 mL L^{-1} 50 % (*v/v*) glycerol were added by filter sterilisation (0.2 µm Sartorius Minisart)

25 × salts: 1.25 mM Na_2HPO_4, 1.25 mM KH_2PO_4, 2.5 mM NH_4Cl, 250 mM Na_2SO_4

1000 × metals: 50 mM $FeCl_3.6H_2O$, 20 mM $CaCl_2.6H_2O$, 100 mM $MnCl_2.4H_2O$, 100 mM $ZnSO_4.7H_2O$, 1.7 mM $CoCl_2.6H_2O$, 2 mM $CuCl_2.2H_2O$, 4.1 mM $Na_2MoO_4.2H_2O$, 2 mM Na_2SeO_3, 2 mM H_3BO_3, 2 mM $NiSO_4.6H_2O$, 50 % (*v/v*) 0.1 M HCl

7.2.2   General buffers for working with proteins

Tris-buffered saline (TBS) (pH 7.4): 50 mM Tris, 150 mM NaCl

HEPES buffer (pH 7.5): 50 mM HEPES, 200 mM NaCl

Phosphate-buffered saline (PBS) (pH 7.4): 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4

WT-Sortase buffer (pH 7.4): 50 mM Tris, 150 mM NaCl, 5 mM $CaCl_2$

Ni-NTA 10 × stock buffer (pH 7.4): 0.5 M KH_2PO_4 and 3 M NaCl

Ni-NTA lysis buffer (pH 7.4): 10 × stock buffer and 10 mM imidazole

Ni-NTA wash buffer (pH 7.4): 10 × stock buffer and 50 mM imidazole

Ni-NTA elution buffer (pH 7.4): 10 × stock buffer and 250 mM imidazole

Sortase 7M Ni-NTA lysis buffer (pH 7.2): 50 mM Tris, 150 mM NaCl

Sortase 7M Ni-NTA wash buffer (pH 7.2): Srt 7M lysis buffer and 50 mM imidazole

Sortase 7M Ni-NTA elution buffer (pH 7.2): Srt 7M lysis buffer and 500 mM imidazole

Amylose elution buffer (pH 7.4): PBS and 20 % (*wt/v*) glucose

ÄKTA Tris buffer for PanZ and PanD (pH 7.4): 50 mM Tris, 100 mM NaCl, 0.1 mM DTT

ÄKTA Tris buffer for Sortase 7M and mCherry (pH 7.4): 50 mM Tris, 150 mM NaCl

ÄKTA Tris buffer for WT-Sortase (pH 7.4): 20 mM Tris, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT

ÄKTA Tris buffer for MupB (pH 7.4): 50 mM Tris, 150 mM NaCl

ÄKTA Phosphate buffer for MBP (pH 7.4): 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl.

Tris-tricine gel buffer (pH 8.45): 3 M Tris HCl, 0.3 % SDS

Tris-tricine 5 × cathode running buffer (pH 8.25): 0.1 M Tris, 0.1 M Tricine, 0.1 % (*w/v*) SDS

Tris-tricine 5 × anode running buffer (pH 8.90): 0.2 M Tris

Tris-glycine separating gel buffer (pH 8.8): 1.5 M Tris

Tris-glycine stacking buffer (pH 6.8): 0.5 M Tris

5 × Tris-glycine running buffer: 25 mM Tris, 192 mM glycine, 0.1 % (*w/v*) SDS

Loading buffer (pH 6.8): 50 mM Tris HCl, 2 % (*w/v*) SDS, 200 mM DTT, 10 % (*v/v*) glycerol, bromophenol blue and water

Coomassie stain: Coomassie G250, 40 % (*v/v*) methanol and 10 % (*v/v*) acetic acid.

Coomassie destain: 40 % (*v/v*) methanol and 10 % (*v/v*) acetic acid

Bjerrum Schafer-Nielsen Buffer (pH 9.2): 48 mM Tris, 39 mM glycine, 20 % (*v/v*) methanol

Tris-buffered saline with Tween-20 (pH 7.5): 10 mM Tris HCl, 150 mM NaCl, 0.5 % (*v/v*) Tween-20

7.2.3 General buffers for working with DNA

Pfu buffer (pH 8.8): 20 mM Tris HCl, 10 mM KCl, 10 mM (NH4)$_2$SO$_4$, 0.1-4 mM MgSO$_4$, 1 % (*v/v*) Triton X-100, 0.1 % (*w/v*) BSA

10 × KOD buffer (pH 8.8): 1.2 M Tris, 60 mM (NH$_4$)$_2$SO$_4$, 100 mM KCl, 1 % (*v/v*) Triton and 0.01 % (*w/v*) BSA

NEB buffer 3.1 (pH 7.9): 100 mM NaCl, 50 mM Tris HCl, 10 mM MgCl$_2$, 0.01 % (*w/v*) BSA

T4 PNK ligase buffer (pH 7.6): 70 mM Tris HCl, 10 mM MgCl$_2$, 5 mM DTT

DNA ÄKTA Buffer (pH 8.0): 10 mM Tris HCl

Anion exchange low salt buffer (pH 5.0): 10 mM Tris HCl, 1 mM EDTA, 0.2 M NaCl

Anion exchange high salt buffer (pH 5.0): 10 mM Tris HCl, 1 mM EDTA, 2 M NaCl

50 × Tris-acetate-EDTA (TAE) buffer: 2 M Tris HCl, 20 mM acetic acid and 1 mM EDTA

5 × Tris-borate-EDTA (TBE) buffer (pH 8.3): 445 mM Tris, 445 mM Boric acid, 10 mM EDTA

6 × Bromophenol blue loading buffer (pH 8.0): 30 % (*w/v*) glycerol, 0.25 % (*w/v*) bromophenol blue

Buffer P1 (Qiagen): 50 mM Tris HCl (pH 8.0), 10 mM EDTA, 0.01 % (*w/v*) RNase A, 0.1 % (*v/v*) LyseBlue

Buffer P2 (Qiagen): 200 mM NaOH, 1 % (*w/v*) SDS

Buffer PB (Qiagen): 5 M guanidine HCl, 30 % (*v/v*) isopropanol

Buffer PE (Qiagen): 10 mM Tris HCl (pH 7.5), 80 % (*v/v*) ethanol

Buffer BB (Qiagen): Composition confidential

Buffer ETR (Qiagen): Composition confidential

Buffer N3 (Qiagen): 4.2 M guanidine HCl, 0.9 M KAc

Buffer QG (Qiagen): 20 mM Tris HCl (pH 6.6), 5.5 M guanidine thiocyanate

DNA binding buffer (Zymo research): Composition confidential

DNA wash buffer (Zymo research): Composition confidential

DNA elution buffer (Zymo research): 10 mM Tris HCl (pH 8.5), 0.1 mM EDTA

7.3 Standard Protocols

7.3.1 Cell growth and protein overexpression

7.3.1.1 Transformation of competent E. coli cells

10 μL of *E. coli* competent cells and 1 μL of plasmid were mixed together in a sterile Eppendorf tube on ice. The cells were incubated for 10 minutes to allow diffusion of the plasmid, followed by a heat shock for 45 s at 42 °C and another incubation on ice for 10 minutes to allow plasmid uptake. LB media (1 mL) was added to the cells and they were incubated at 37 °C with shaking (200 rpm) for 1 hour. A series of 10 μL, 100 μL and concentrated cell aliquots were used to inoculate sterilised agar plates made using either ampicillin (100 μg mL^{-1}) or kanamycin (50 μg mL^{-1}). The concentrated cell aliquots were taken from the remaining cell culture which was centrifuged at 13,000 × *g* for 30s, the supernatant was removed and the pelleted cells were re-suspended. The plates were incubated at 37 °C overnight and removed in the morning to stop growth.

7.3.1.2 Preparation mini-culture

A single colony from the transformed *E. coli* cells grown on the sterilised agar plates overnight was picked and incubated with shaking in ~5 mL of LB media containing the necessary antibiotic (either ampicillin (100 μg mL^{-1}) or kanamycin (50 μg mL^{-1})) at 37 °C overnight. This mini-culture was then used to inoculate 1 L of overexpression media, or glycerol stocks were made using a mixture of the cells and 80 % sterilised glycerol (in H$_2$O) in a 1:1 ratio which were flash frozen in liquid nitrogen and stored at -80 °C for future use.

7.3.1.3 Overexpression in LB media

1 L of sterilised LB media containing the desired antibiotic (1 mL of either ampicillin (100 μg mL^{-1}) or kanamycin (50 μg mL^{-1})) in a 2 L flask was inoculated with 2 mL of cells from the overnight mini-culture (MupB also required the addition of 0.5 mL 1000 × metals). The culture was incubated at 37 °C with shaking (200 rpm) and overexpression was induced by the

addition of IPTG (0.5 mM) or arabinose (PanZ only, 5 g L^{-1}) when OD$_{600}$= 0.6-0.8. The cells were incubated for 24 hours at a specified temperature (Table 7.1) before being collected by centrifugation (10 minutes at 10,000 × g). The cell pellet was stored at -80 °C for future cell lysis and protein purification.

Protein	Temperature after induction
PanD	30 ºC
PanZ	30 ºC
Sortase	30 °C
MBP	37 °C
mCherry	37 °C
MupB	25 °C

Table 7.1: Temperatures after induction using IPTG for expression using LB

7.3.1.4   Overexpression in auto induction media

1 L of sterilised AI media containing the desired antibiotic (1 mL of either ampicillin (100 μg mL^{-1}) or kanamycin (50 μg mL^{-1})) in a 2 L flask was inoculated with 2 mL of cells from the overnight mini culture. The culture was incubated at a 25 °C for 24 hours and the cell pellet was collected by centrifugation (10 minutes at 10,000 × g) and stored at -80 °C for future lysis and protein purification.

7.3.2   Protein purification

7.3.2.1   Purification of His$_6$-tagged proteins using Ni-NTA chromatography

E. coli cells were collected from overexpression cultures and re-suspended in Ni-NTA lysis buffer (10 mM imidazole) with protease inhibitor cocktail (cOmplete EDTA-free, Roche), unless purifying sortase, at 4 °C. The cell mixture was lysed mechanically using a Constant Systems cell disruptor (20 kpsi). The cell debris was pelleted by centrifugation (30,000 × g, 45 min) and the cell lysate was decanted off and treated with DNase. The Ni-NTA agarose resin (Qiagen, 10 mL) was washed under gravity with 30 mL of H$_2$O and equilibrated with 50 mL of Ni-NTA lysis buffer

(10 mM imidazole). The cell lysate was loaded to the Ni-NTA agarose column and low affinity binders removed by washing the resin with 50 mL of Ni-NTA wash buffer (50 mM imidazole). The protein was eluted in 5 mL fractions of Ni-NTA elution buffer (250 mM imidazole). The protein elution was monitored using Bradford reagent. The protein size, expression levels and purity were then assessed by SDS-PAGE.

7.3.2.2   Regeneration of Ni-NTA column

After being used for protein purification, the Ni-NTA agarose resin was regenerated for future use. Firstly, the resin was washed with 0.1 % sodium dodecyl sulphate (SDS, 30 mL), then 0.1 M NaOH (30 mL). The Ni^{2+} was removed by chelation with 10 mM ethylenediaminetetraacetic acid (EDTA, 30 mL), the resin was washed with 100 mL of H_2O and regenerated with 500 mM Ni_2SO_4 (20 mL). A final wash with H_2O was done to remove any excess Ni^{2+} and the column was flushed and stored in 20 % EtOH.

7.3.2.3   Amylose affinity chromatography

E. coli cells were collected from overexpression cultures and re-suspended in PBS with protease inhibitor cocktail (cOmplete EDTA-free, Roche) at 4 °C. The cell mixture was lysed mechanically using a Constant Systems cell disruptor (20 kpsi). The cell debris was pelleted by centrifugation ($30,000 \times g$, 45 min) and the cell lysate was treated with DNase. The amylose resin (Qiagen, 10 mL) was washed under gravity with 30 mL of H_2O and equilibrated with 50 mL of PBS. The cell lysate was loaded onto the amylose column and low affinity binders removed by washing the resin with 50 mL of PBS. The protein was eluted in 5 mL fractions of amylose elution buffer (10 mM maltose). The protein elution was monitored using Bradford reagent. The protein size, expression levels and purity were then assessed by SDS-PAGE.

7.3.2.4   *Washing of amylose column*

After being used for protein purification, the amylose resin was prepared for future use. Firstly, the resin was washed with 0.1 % sodium dodecyl sulphate (SDS, 30 mL), then 0.1 M NaOH (30 mL), followed by 100 mL of H_2O and was flushed and stored in 20 % EtOH.

7.3.2.5   *Size-exclusion chromatography*

Protein-containing fractions eluted from the Ni-NTA column were combined and concentrated to 1-1.5 mL (Amicon centrifugal concentrator 10 kDa MWCO, 4,000 *g*). The concentrated protein was loaded on to a Superdex® 75 or 200 column (HiLoad® 16/60 or 26/60, GE Healthcare Life Sciences) and eluted isocratically with an appropriate gel filtration buffer in 5 mL fractions which are stored at 4 °C. Protein elution was monitored using 280 nm absorption.

7.3.3   Protein analysis

7.3.3.1   *Tris-Tricine gel running procedure*

20 μL of purified protein in buffer was mixed with 20 μL of loading buffer. The mixtures were heated at 95-100 °C for 10 minutes to ensure the protein was fully denatured.

The separating gel was prepared using all of the components listed in the table below (Table 7.2). Everything but the ammonium persulfate (APS) and tetramethylethylenediamine (TEMED) were mixed by inversion. The APS and TEMED were added and gently mixed just before pouring the gel between the plates and a layer of *iso*-propanol was added to ensure a level front. Once set, the *iso*-propanol was washed off with H_2O ready for the stacking gel which was made using the same method as above with a comb insert placed in the top. Once set, the gel was loaded using the samples in loading buffer, the anode and cathode buffers were added and the gels were run at a constant current of 35 mA per gel for approximately 3.5 hours, or until the protein bands can be seen to have reached the bottom of the gel.

2 × 1 mm gels	Separating	Stacking
Glycerol	2.00 g	-
Tris-Tricine gel buffer	5.00 mL	3.10 mL
40 % (*w/v*) acrylamide	3.75 mL	1.25 mL
H_2O	5.25 mL	8.20 mL
20 % APS	100 µL	100 µL
TEMED	10 µL	10 µL

Table 7.2: Tris-Tricine SDS-PAGE gel receipe

7.3.3.2 Tris-glycine running procedure

The samples and gels were prepared as noted for the tris-tricine gel procedure above, however different components were used for the gel itself (Table 7.3). Once set, the gel was loaded using the samples in loading buffer, the running buffer was added and the gels were run at a constant voltage of 120 V for approximately 45 minutes, or until the protein bands can be seen to have reached the bottom of the gel.

	Stacking	Separating			
2 × 1 mm gel	5 %	7 %	10 %	12 %	15 %
1.5 M Tris	-	2.53 mL	2.53 mL	2.53 mL	2.53 mL
0.5 M Tris	0.95 mL	-	-	-	-
40 % (*w/v*) acrylamide	0.62 mL	1.75 mL	2.53 mL	3.00 mL	3.75 mL
10 % (*w/v*) SDS	50 µL	100 µL	100 µL	100 µL	100 µL
H_2O	3.30 mL	5.50 mL	4.68 mL	4.25 mL	3.50 mL
10 % (*w/v*) APS	75 µL	150 µL	150 µL	150 µL	150 µL
TEMED	10 µL	10 µL	10 µL	10 µL	10 µL

Table 7.3: Tris-glycine SDS-PAGE gel recipes

7.3.3.3 Gel imaging

The gels were stained with either Coomassie blue or Instant Blue. For the Coomassie blue, stain was added to the gel, the mixture was gently warmed and then incubated at room temperature for

2 hours with rocking. The stain was removed and replaced with Coomassie destain which was then incubated for another 3 hours with rocking before being washed with H_2O. For the instant blue, the gel was incubated on the rocker at room temperature until protein bands were visible. Gels were photographed using Gel Doc XR or MP system, Bio-Rad. **If fluorescent labels were present, the gels were imaged using UV illumination before staining.**

7.3.3.4 *Protein concentration determination*

UV absorbance of the protein samples were taken using the Nanodrop 2000 at 280 nm. 2 μL of sample were loaded on to the instrument after a blank buffer measurement had been taken. The absorbance was measured at 280 nm and the concentration was calculated using the Beer-Lambert law (Equation 7.1) and extinction coefficients (Table 7.4). These were obtained empirically using the ProtParam online platform.[170]

$$A = \varepsilon c l$$

Equation 7.1: Beer-Lambert Law: A = absorbance, ε = molar extinction coefficient ($M^{-1} cm^{-1}$), c = concentration (M) and l = path length (cm)

Protein	Extinction Coeff., ε_{280}	Protein	Extinction Coeff., ε_{280}
PanD	$15470\ M^{-1} cm^{-1}$	GVGK-MCherry	$35870\ M^{-1} cm^{-1}$
PanZ	$26470\ M^{-1} cm^{-1}$	GVGK-MBP	$66350\ M^{-1} cm^{-1}$
TEV-GG-PanD	$16960\ M^{-1} cm^{-1}$	CoaD	$8480\ M^{-1} cm^{-1}$
GG-PanD	$15470\ M^{-1} cm^{-1}$	CoaE	$16760\ M^{-1} cm^{-1}$
SrtA (7M)	$14440\ M^{-1} cm^{-1}$	TEV-G-MupB	$178540\ M^{-1} cm^{-1}$
GGG-CTB	$11585\ M^{-1} cm^{-1}$	G-MupB	$177800\ M^{-1} cm^{-1}$

Table 7.4: Extinction coefficients for proteins at λ= 280 nm.[170]

7.3.4 Enzymatic reactions

7.3.4.1 Typical TEV cleavage procedure of TEV-GG-PanD

For a total volume of 250 µL, TEV-GG-PanD (50 µM) and His$_6$-TEV (10 µM, 20 mol%, provided by Gemma Wildsmith) were made up to the total volume with TBS. The reaction mixture was incubated at 37 °C until seen to be complete using mass spectrometry analysis (approx. 2 hours). The reaction mixture was then put down a PD-10 desalting column (GE Healthcare Life Sciences) pre-equilibrated with TBS to remove any traces of DTT from the purification of the TEV-GG-PanD. Batch binding with Ni-NTA resin was used to remove any His$_6$-TEV and the His$_6$-tag cut from GG-PanD. A spin filter column was used to remove the resin and the flow through was concentrated (Amicon centrifugal concentrator 10 kDa MWCO, $10,000 \times g$) to leave cleaved GG-PanD. The same procedure was used for the cleavage of TEV-G-MupB.

7.3.4.2 Typical sortase 7M mediated labelling

For a total volume of 150 µL, the protein to be labelled (50 µM) and depsipeptide (150 µM, 3 eq.) were mixed and the reaction mixture was made up to volume with an appropriate buffer before the addition of sortase A 7M (2.50 µM, 5 mol%). The reaction mixture was incubated for 1-4 hours at room temperature and product formation was monitored by mass spectrometry where possible, or time point samples were taken. To remove the His$_6$-sortase, the sample was mixed with Ni-NTA resin incubated for 15 minutes on ice before the resin was removed using a spin filter. The excess depsipeptide labelling reagent was removed using a PD-10 desalting column (GE Healthcare Life Sciences) pre-equilibrated with the appropriate buffer. Finally, the labelled protein sample was concentrated (Amicon centrifugal concentrator 10 kDa MWCO, $10,000 \times g$). For time point studies, samples (5 µL) were taken at the appropriate time, mixed with loading buffer (5 µL) and heated at 95 °C for 5 minutes before being frozen and stored at -20 °C for future analysis by SDS-PAGE.

7.3.4.3   *Typical WT-sortase mediated labelling*

For a total volume of 150 µL, the protein to be labelled (50 µM) and depsipeptide (150 µM, 3 eq.) were mixed and the reaction mixture was made up to volume with WT-sortase buffer before the addition of WT sortase A (10 µM, 20 mol%). The reaction mixture was incubated for 1-4 hours at 37 °C and time point samples were taken (5 µL), mixed with loading buffer (5 µL) and heated at 95 °C for 5 minutes before being frozen and stored at -20 °C for future analysis by SDS-PAGE.

7.3.5   DNA mutagenesis, digestion and amplification

7.3.5.1   *Site-directed mutagenesis*[50,51]

This was based on the KOD hot start protocol.[154] Forward and reverse primers were designed based on the QuikChange protocol guidelines which state the primers designed must be 25-45 bp in length, contain $\geq$ 40 % GC, with each primer having one or more G/C at the end and $T_m \geq$ 78 °C. The primer T_m was calculated using Equation 7.2.[152]

$$T_m = 81.5 + 0.41(\% \, GC) - \frac{675}{N} - \% \, mismatch$$

Equation 7.2: Calculation of T_m where N is the number of bases and % GC and % mismatch are given as integers.[152]

The PCR reaction mixture (50 µL) was made up of 50 ng of the template DNA plasmid, 0.14 µM of each of the forward and reverse primers, 200 µM dNTPs, 1.5 µM MgSO$_4$ and 1 unit of KOD hot start DNA polymerase (1 U µL^{-1}) in 1 × KOD Buffer. The thermocycler program used is described in Table 7.5. Once completed, a DpnI digestion was done by adding 1 µL DpnI (20 U µL^{-1}) and incubating at 37 °C for 1 hour. This DNA was then used to carry out a transformation into fresh *E. coli* XL10-gold ultracompetent cells (Section 7.3.1.1). If this transformation is unsuccessful then the DNA was purified from the reaction mixture using a using a QIAquick PCR purification method (see small scale PCR reaction purification). Following this some DNA was taken (16.5 µL) and T4 PNK (0.5 µL, 10 U µL^{-1}) in T4 ligase buffer (2 µL) was

added and the mixture was incubated at 37 °C for 1 hour. To that T4 DNA ligase (1 µL, 400 U µL^{-1}) was added and incubated at room temperature for 30 minutes before being used to carry out another transformation into fresh *E. coli* XL10-gold ultracompetent cells.

Step	Temperature	Time	Number of cycles
1. Initial denaturing	95 ℃	2 min.	1
2. Denaturing	95 ℃	20 s	
3. Annealing	55 ℃	20 s	20
4. Elongation	72 ℃	2.5 min.	
5. Final elongation	72 ℃	10 min.	1

Table 7.5: KOD hot-start polymerase PCR program for general site-directed mutagenesis.[152,155]

7.3.5.2   Restriction and nicking endonuclease enzymatic reaction

Plasmid DNA (200 ng) was mixed with 2.5 units of BstZ17i, 15 units of Nb.BsmI and 10 units of Nb.BbvCI in $1 \times$ NEB Buffer 3.1 (total volume, 30 µL). The number of units of each enzyme used is based on the number of sites which will be cut in the plasmid by each enzyme. The mixture was heated at 65 °C for 1 hour, 37 °C for 1 hour and finally 95 °C for 2 minutes to denature the enzymes. The sample was then analysed on an agarose gel.

7.3.5.3   PCR for preparation of azide-DNA using Pfu

The PCR reaction mixture (50 µL) was made up of 50 ng of the template DNA plasmid, 0.5 µM of each of the forward and reverse primers, 20 µM dNTPs (25 mM stock mixture of each) and 0.5 unit of *Pfu* polymerase (3 U µL^{-1}) in $1 \times$ *Pfu* buffer (15 mM MgSO$_4$) (Table 7.6a). The thermocycler program used is described in Table 7.6b.

a)

Components	Volume
Template DNA	50 ng
10 µM F	5 µL
10 µM R	5 µL
25 mM dNTPs	1 µL
PWO	0.8 µL
Buffer	5 µL
H$_2$O	Up to 50 µL

b)

Step	Temperature	Time	Number of cycles
1. Initial denaturing	95 ℃	2 min.	1
2. Denaturing	95 ℃	30 s	
3. Annealing	60 ℃	30 s	30
4. Elongation	72 ℃	2 min.	
5. Final elongation	72 ℃	5 min.	1

Table 7.6:a) Reaction components for PCR using Pfu. b) PCR program for preparation of azide-DNA using Pfu.[153]

7.3.5.4   PCR for preparation of azide-DNA using KOD

The PCR reaction mixture (50 µL) was made up of 50 ng of the template DNA plasmid, 0.3 µM of each of the forward and reverse primers, 200 µM dNTPs (25 mM stock mixture of each), 1.5 µM MgSO$_4$, and 0.5 unit of KOD polymerase (1 U µL^{-1}) in 1 × KOD buffer. These reaction components and the thermocycler program are displayed in Table 7.7 a&b.

a)

Components	Volume
Template DNA	50 ng
10 µM F	1.5 µL
10 µM R	1.5 µL
25 mM dNTPs	0.4 µL
KOD	0.5 µL
MgSO$_4$ Buffers	3 µL
KOD Buffer	5 µL
H$_2$O	Up to 50 µL

b)

Step	Temperature	Time	Number of cycles
1. Initial denaturing	95 ℃	2 min.	1
2. Denaturing	95 ℃	20 s	
3. Annealing	62 ℃	15 s	30
4. Elongation	72 ℃	15 s	
5. Final elongation	72 ℃	5 min.	1

Table 7.7:a) PCR program for preparation of azide-DNA using KOD. b) PCR program for preparation of azide-DNA using KOD [155]

7.3.6   DNA purification and analysis

7.3.6.1   Small scale DNA isolation and sequence analysis (Mini-prep)

The bacterial cell pellet was isolated from mini cultures (5 mL) by centrifugation (15 minutes at $4,500 \times g$) at room temperature in a bench top centrifuge and the DNA was extracted using a QIAprep Spin Mini Prep Kit by following the manufacturer's protocol. All of the following centrifugation steps were carried out at $17,000 \times g$ in a micro centrifuge at room temperature. The cell pellet was resuspended in buffer P1 (250 µL) and mixed thoroughly with buffer P2 (250 µL) before being incubated for a maximum of 4 minutes at room temperature. Buffer N3 (350 µL) was added and mixed by inversion of the Eppendorf before being centrifuged for 10 minutes. The supernatant was then applied to the QIAprep Spin column and centrifuged for

1 minute to bind the DNA to the column. The flow through was discarded and the column was washed with buffer PE (750 µL) and centrifuged for 2×1 minute. The DNA sample was eluted in H_2O (30 µL) from the column by centrifugation for 1 minute. Sequencing was performed by GATC Biotech and the results were analysed using BioEdit sequence alignment editor.[171]

7.3.6.2   Medium scale DNA isolation (Midi-prep)

The bacterial cell pellet was isolated from medium scale cultures (35 mL) by centrifugation (20 minutes at $4500 \times g$) at room temperature in a bench top centrifuge. The DNA was extracted using a Plasmid Plus Midi Kit (Qiagen) by following the manufacturer's protocol and a vacuum manifold was used at approximately 300 mbar where necessary. The cells pellet was resuspended in buffer P1 (4 mL) before buffer P2 was added (4 mL) and the mixture was incubated for 3 minutes at room temperature. Buffer S3 (4 mL) was added at the lysate was applied to the QIA filter cartridge and incubated at room temperature for 10 minutes before the precipitate was removed by filtration. Buffer BB (2 mL) was added to the cleared lysate and the mixture was passed through the QIAGEN plasmid plus midi spin column using the vacuum manifold, binding the DNA to the silica. The column was washed with buffer ETR (700 µL), followed by buffer PE (700 µL). Residual buffer was removed by centrifugation for 1 minute at $17,000 \times g$ in a micro centrifuge before the DNA was eluted with H_2O (200 µL).

7.3.6.3   Large scale DNA isolation (Giga-prep)

The bacterial cell pellet was isolated from 3 L of culture by centrifugation (10 minutes at $10,000 \times g$). The DNA was extracted using a Plasmid Plus Giga Kit (Qiagen) by following the manufacturer's protocol. All centrifugation steps involved were carried out at $4,000 \times g$ in a benchtop centrifuge at room temperature unless otherwise stated and a vacuum manifold was used at approximately 300 mbar where necessary. The pellet was resuspended in buffer P1 (100 mL) before buffer P2 was added (100 mL) and the mixture was incubated for 4 minutes at room temperature. Buffer S3 was added and the debris was pelleted using centrifugation (10 minutes at $10,000 \times g$ using an Avanti JXN-30 centrifuge) before being applied to the QIA filter cartridge

on the vacuum manifold. Buffer BB (100 mL) was added to the cleared lysate and applied to the QIAGEN plasmid plus giga spin column, binding the DNA to the silica. The column was then washed using centrifugation with Buffer ETR (4 × 20 mL) and Buffer PE (2 × 25 mL). Residual buffer was removed by centrifugation for 10 minutes before the DNA was eluted with H_2O (5 mL).

7.3.6.4 Concentration analysis

UV absorbance of the DNA samples were taken using the Nanodrop 2000 at 260 nm and 320 nm. 2 µL of sample were loaded on to the instrument platform after a blank buffer measurement had been taken. The concentration is then estimated using Equation 7.3.

$$Concentration\ (\mu g\ mL^{-1}) = A_{260} \times 50\ \mu g\ mL^{-1}$$

Equation 7.3: DNA concentration calculation

7.3.6.5 Small scale PCR reaction mixture purification (PCR clean-up)

The PCR reaction mixtures were purified using a QIAquick PCR purification method using a microcentrifuge (17,000 × g) at room temperature. Buffer PB (5 volumes) was mixed with the PCR reaction mixture and passed through the QIAquick Spin Column by centrifugation for 1 minute. The flow through was discarded and the silica was washed with buffer PE (750 µL) by centrifugation for 1 minute to remove salts. The flow through was discarded and the column was centrifuged for a further minute to remove any residual buffer. The DNA was finally eluted using H_2O (30 µL), after being left to stand for 1 minute, the DNA was collected in a clean Eppendorf by centrifugation for 1 minute.

7.3.6.6 Large scale PCR reaction mixture purification (Large scale PCR clean-up)

The PCR reaction mixtures were purified using a DNA Clean & Concetrator-500 (Zymo research) using a benchtop centrifuge (3,000 × g) at room temperature. The DNA was mixed with 5

volumes of DNA binding buffer and applied to the Zymo-Spin VI column inside a 50 mL falcon tube which was centrifuged for 5 minutes. The flow through was discarded and DNA wash buffer (10 mL) was applied to the column to wash the bound DNA by centrifugation for 5 minutes. The column was then transferred to a new tube and 2-3 mL of DNA elution buffer or H_2O was added to the column and incubated at room temperature for 1 minute before isolating the DNA by centrifugation for 3 minutes.

7.3.6.7 *Size-exclusion chromatography purification of DNA*

Prior to the DNA sample being applied to the column it was necessary to ensure no DNase was present from previous use with protein samples. To do this it was washed with filtered 0.1 M NaOH followed by filtered and autoclaved H_2O and then equilibrated with filtered and autoclaved DNA ÄKTA buffer. The fractions containing DNA were analysed by agarose gel electrophoresis and concentrated (Amicon centrifugal concentrator 10 kDa MWCO, $4,000 \times g$).

7.3.6.8 *Anion exchange chromatography*

The DNA sample was applied to a 5 mL DEAE Sepharose fast-flow column (GE Healthcare) and eluted using either a gradient or stepwise increase of salt concentration (0.2-2 M). The A_{260} was monitored throughout elution to determine the DNA-containing fractions before being analysed by agarose gel electrophoresis.

7.3.6.9 *Agarose gel electrophoresis*

To analyse the DNA after reactions 1 % agarose gel electrophoresis was used. They were prepared by dissolving agarose (0.40 g) in TAE buffer (40 mL) using a microwave for 1 minute. The solution was allowed to cool for a few minutes before ethidium bromide (1.33 μL of 10 μg mL^{-1}) or SYBR®Safe (2 μL of $10,000 \times$ in DMSO) was added and the mixture was poured into a mould with a comb added and left to set. The DNA samples (5 μL) were prepared by mixing with $6 \times$ bromophenol blue loading buffer (1 μL). Once set, the gel was put into the tank and filled

with TAE buffer, the samples were then loaded into the wells and run at 100 V for 25 minutes. The gels were then imaged using Gel Doc XR or MP system, Bio-Rad.

7.3.6.10   Bis-acrylamide TBE SDS-PAGE gel

This gel was made as one mixture (Table 7.8) with the comb placed straight into the top between the plates and allowed to set. The comb was then removed while submerged in the TBE running buffer in the tank and was loaded using the samples in loading buffer and the gels were run at a constant voltage of 200 V for approximately 50 minutes, or until the bands can be seen to have reached the bottom of the gel. Once complete, the gels were places in an ethidium bromide bath (1 µg mL^{-1}) for 15 minutes with rocking and then imaged using Gel Doc XR or MP system, Bio-Rad.

1 × 1.5 mm gels	Volume
5 × TBE	2 mL
bisacrylamide mix (37:5:1)	5 mL
H$_2$O	3 mL
20 % (*w/v*) APS	100 µL
TEMED	10 µL

Table 7.8: DNA acrylamide gel recipe

7.3.7   Protein-DNA conjugation

Following the successful sortase 7M mediated labelling of MBP with the bicyclononyne functionalised depsipeptide (5 mol% Srt7M, 3 eq. BCN-depsipeptide, RT for 1.5 hours), the MBP-BCN was used in two strain-promoted azide alkyne cycloaddition reactions using the azide-oligonucleotide or azide-DNA synthesised using PCR.

7.3.7.1 MBP-oligonucleotide preparation

Firstly, a 4 mM stock solution of N_3-oligo was made up using H_2O from the solid provided by Darren Machin. This was used to react with MBP (50 µM) in an excess of 1.2 equivalents and a total volume of 200 µL. The reactions were incubated at room temperature for 24 hours before being analysed by SDS-PAGE. Following successful generation of the MBP-oligo, excess oligonucleotide was removed using an amylose column, the sample was applied to the resin, excess oligo was washed out using PBS and the MBP-oligo was eluted using 20 % glucose in PBS. The MBP-BCN and MBP-oligo mixture was subsequently purified using size-exclusion chromatography. This was carried out using the Superdex® S75 increase column having been thoroughly cleaned (Section 6.3.6.7) and monitored by A_{260} and A_{280}. The resulting protein-DNA containing fractions were analysed by both tris-glycine and TBE SDS-PAGE by mixing the samples with 1 µL of 50 µM complementary oligonucleotide before being run on a gel stained using EtBr (5 % solution in H_2O). Fractions containing MBP-oligo were combined and concentrated (Amicon centrifugal concentrator 30 kDa MWCO, $10,000 \times g$) before calculations were performed to find the concentration.

7.3.7.2 Concentration determination for MBP-oligonucelotide

An approximate concentration was determined using the ThermoFisher Scientific Nanodrop 2000 "Protein & Labels" setting, presuming the absorbance at 280 nm is mostly a contribution from the protein and the absorbance at 260 nm is mostly a contribution from the oligonucleotide. First MBP-BCN was analysed, taking three measurements at 260 and then 280 nm before working out the average and calculating a value for 260/280 to be 0.5. Then the azide-oligo was analysed in the same way, also calculating a value for 260/280 to be 1.84. Following this a series of 260 and 280 nm measurements were taken for the MBP-oligo which were averaged to give two numbers. From here two simultaneous equations were made where *x* are contributions from MBP-BCN and *y* are contributions from the azide-DNA:

$$A_{280} \qquad x + y = A_{280(\text{MBP-oligo})}$$

$$A_{260} \qquad 0.5\,x + 1.84\,y = A_{260(\text{MBP-oligo})}$$

Working through the equations provides a specific value for x and y which can then be used as the absorbance for the protein and oligonucleotide. To find the concentrations, they are then divided by their appropriate extinction coefficients (MBP: 66350 $M^{-1}cm^{-1}$, N_3-oligo: 194200 $M^{-1}cm^{-1}$). Extinction coefficient for the oligonucleotide were obtained from IDT oligo analyser.[172]

7.3.7.3   MBP-DNA preparation

The MBP-DNA was formed on small scale using an 8.9 µM stock of N_3- DNA prepared from purified successful PCR reactions in H_2O. This was mixed with MBP-BCN (5 µM) in a 1:1 ratio and a total volume of 15 µL before being incubated at room temperature overnight. The resulting mixture was then run on a 10 % tris-glycine SDS-PAGE gel and stained using EtBr (5 % solution in H_2O).

7.3.7.4   Western blotting of MBP-DNA conjugate

A 1.0 mm 10 % acrylamide gel was run as described in section 3.3.3 using a pre-stained ladder (New England Biosciences). The gel was then washed with H_2O before being equilibrated in Bjerrum Schafer-Nielsen buffer for 15 minutes. In this time, a piece of PVDF membrane was submerged in methanol for one minute before being equilibrated in Bjerrum Schafer-Nielsen buffer along with two pieces of extra thick blotting paper for 10 minutes. The transfer stack was then assembled on the anode with one piece of blotting paper on the base, followed by the membrane, gel and then another piece of blotting paper. It is advised to cut a small nick out of one corner of the membrane to keep track of the orientation. Air bubbles were carefully removed by rolling a pipette over the stack before being left to transfer for 1 hour at a constant voltage of 15 V. Following transfer, the gel was stained using ethidium bromide (5 % solution in H_2O), and

the membrane was placed in methanol for one minute before being transferred to a blocking bath of 5 % (*w/v*) skimmed milk powder in tris-buffered saline with Tween-20 (TBST buffer) for at least 2 hours at room temperature or in the fridge overnight. The blocking solution was then removed, and the antibody solution was added, this consisted of the anti-MBP antibody in a ratio of 1:30,000 along with 2 % (*w/v*) skimmed milk powder in TBST buffer. Following incubation for one hour at room temperature, the membrane was washed with 3×20 mL TBST before being imaged. To visualise the antibody, 2 mL of pre-mixed reagents in a ratio of 1:1 from the BioRad Clarity Max Western ECL kit was carefully added to the membrane. Imaging was carried out using the GelDoc MP in chemilluminescence mode, followed by colorimetric mode in order to allow the ladder to be seen.

7.3.8 Typical FRET sortase experiment

The concentrations of the peptides used in these experiments were calculated using the ThermoFisher Scientific Nanodrop 2000 and their respective extinction coefficients. For FITC this was found in literature to be 68,000 M cm^{-1} at λ= 494 nm.[173] However, for the non-fluorescent species the extinction coefficients at 214 nm had to be calculated using Equation 7.4. This basically means the number of peptide bonds present are multiplied by the extinction coefficient of one peptide bond (923 M^{-1} cm^{-1}), which is then added to the extinction coefficient of each individual amino acid, found in the literature. These were calculated to be 7496 M cm^{-1} for the NF-peptide, and 6630 M cm^{-1} for the NF-depsipeptide (this was calculated by removing a peptide bond from the NF-peptide and then adding 57 which is the difference between glycine and glutamic acid, therefore presumed to be that of the carbonyl in the ester bond).[174] All FRET experiments were performed using the Cary eclipse fluorescent spectrophotometer.

$$\varepsilon_{pept} = \varepsilon_{peptbond} \times n_{\text{peptbond}} + \sum_{i=1}^{20} EAA(i) \times n_{AA}$$

Equation 7.4: Equation used to calculate the extinction coefficient of peptides at 214 nm containing no fluorescent species.[174]

7.3.8.1   *mCherry and FITC-depsipeptide labelling*

For a total volume of 400 µL, varied amounts of mCherry and the non-fluorescent depsipeptide "spiked" with FITC-depsipeptide (in a ratio of 4:1 or 9:1) were mixed before the addition of the sortase 7M. This was gently mixed and transferred to the cuvette, ensuring no bubbles were present the cuvette was then placed in the fluorimeter as quickly as possible. The FITC was excited at 470 nm with emission measurements taken at 518 nm to monitor FITC and 610 nm to monitor mCherry fluorescence. The data was normalised to yield a change in intensity which was then plotted against time in Origin 2016.

7.3.8.2   *Rhodamine-peptide labelling with peptides and depsipeptides*

The rhodamine-peptide was provided by Kristian Hollingsworth at a concentration of 10 mM. This was used in the experiments at 25 µM and mixed with the appropriate combination of NF-depsipeptide/peptide and FITC-depsipeptide/peptide (in a ratio of 4:1 or 9:1). The sortase 7M was then added and mixed before the reaction mixture was transferred to the cuvette ensuring no bubbles were present. This was then placed in the fluorimeter as quickly as possible. Measurements were taken by exciting the FITC at 468 nm and measuring the emission at 516 nm for the FITC and 580 nm for the rhodamine. In addition to this, the rhodamine was separately analysed through excitation at 562 nm to account for any photo-bleaching experienced by the rhodamine molecule when processing the data using excel and Origin 2016.

7.3.9   Atomic force microscopy

Experiments by Andrew Lee, a member of Christoph Wälti's group in Bioelectronics at the University of Leeds. All samples were imaged in tapping mode (amplitude modulation), in aqueous buffer, with a Bruker Dimension Fastscan AFM (Bruker Nanosurfaces, Santa Barbara, USA), using Fastscan D etched Si_3N_4 cantilevers (nominal spring constant = 0.25 N/m, resonant frequency approx. 110 kHz in liquid) containing a Si tip (nominal radius of curvature = 5 nm). Cantilevers were driven close to resonance under liquid and images were typically acquired with

a pixel density of 512×512, a scan speed of 38 Hz and a tapping amplitude of $4 - 6$ nm. Images were flattened by plane-fitting using the associated Nanoscope analysis software (Bruker Nanosurfaces, Santa Barbara, USA).

7.3.9.1   DNA origami frame preparation

The 224 oligonucleotides used to form the nanostructure were synthesised by Integrated DNA Technologies and the M13mp18 single-stranded DNA used as the scaffold was purchased from New England Biolabs.

To form the frame, the oligonucleotides were mixed with the M13mp18 scaffold in a 5:1 ratio in 10 mM Tris Acetate (pH 7.4), 10 mM $Mg(OAc)_2$ and 1 mM EDTA. The DNA was folded using a cooling gradient of -1 °C min^{-1} from 95 °C to 15 °C. The bridging oligo was included at this point also in a 5:1 ratio with the scaffold. A typical reaction setup can be found in Table 7.9.

Component	Volume	Stock conc.
Oligo mix	20 µL	100 nM
M13mp18	4 µL	100 nM
Bridge oligo	2 µL	1 µM
Tris acetate	2 µL	100 mM
Magnesium acetate	2 µL	100 mM
EDTA	2 µL	10 mM
H$_2$O	8 µL	

Table 7.9: Typical reaction set up for forming the DNA origami frames carried out by Andrew Lee.

Following this the folded nanostructures were purified using size-exclusion spin columns containing Sephacryl® 400 (S400). The DNA origami passes through the void volume and was eluted in a 1 min spin at 750 rpm. The excess oligonucleotides are retained in the column and discarded.

The MBP-oligo fusion was then introduced in a ~5:1 ratio, note here the ratio is likely to be in a larger excess than 5:1 as the efficiency of the recovery from the S400 columns is unlikely to be 100 %. However, provided there is an excess of MBP-oligo to frame this isn't an issue. Incorporation is conducted using a cooling gradient of -0.5 °C min^{-1} from 30 °C to 4 °C. This achieves a >90 % incorporation yield, assessed by surveying resultant structures with AFM. A final purification was conducted as before using S400 columns.

The final concentration was estimated using UV-Vis spec, giving typical yield of ~4 nM. Note, the calculations are based on the coefficients for double stranded DNA, which for these structures are not accurate, hence this only provides a guide. Structures are aliquoted and stored at -20 °C for future use.

7.3.9.2   HS-AFM imaging

A total of ~0.1 pmol of complete DNA frames were deposited upon freshly cleaved Muscovite Mica (cleavage is when the material is split along a structural plane), which had been pre-incubated with 10 mM NiCl$_2$. The sample was incubated for 15 mins and rinsed off with deionised water before being immersed in a buffer containing 10 mM Tris acetate (pH 7.4) and 10 mM Mg(OAc)$_2$. Samples were imaged with the AFM prior to introduction of sortase to characterize the surface population.

Sortase was introduced following the guide 95:5 mol ratio as specified by previous labelling experiments. In practical terms, this means 1 µL of a 0.005 pmol/ µL solution of sortase 7M was introduced into the imaging buffer. The imaging volume was ~ 200 µL which equates to ~25 amol/ µL sortase. It is important to note that although the amount of DNA structures added to the surface is roughly known, it is hard to quantify the amount of DNA structures bound to the surface for the experiment, therefore the interaction ratio is likely to differ from the 95:5 mol ratio (i.e. there is likely to be more sortase to frame than planned). Imaging was conducted for 1 hour at room temperature.

7.3.10 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) was used to study the biophysical characteristics of PanD and PanZ and the binding affinity between PanZ and AcCoA in comparison to the antimetabolite EtdtCoA to investigate this interaction as a mechanism for the antibiotic properties of N-pentylpantothenamide.

All ITC experiments were performed using a MicroCal iTC200 (GE Healthcare Life Sciences Life Sciences). Beforehand, out both the cell and syringe were washed with a surfactant followed by plenty of water and finally the buffer being used for the interaction analysis. The reference cell was filled with degassed water and the sample cell was filled with a total of 203 µL of protein sample. Both cells were filled very carefully to ensure no bubbles were left behind and by moving the tip of the syringe to the ledge, any excess liquid was removed from the cell to ensure they contain equal volumes. The syringe was then filled with 39 µL of ligand sample at a concentration of 7-10 × that of the protein in the cell, again ensuring no bubbles were present in the sample. The cells were heated to 25 °C before gently placing the syringe into the sample cell. The ligand was then added to the cell over 20 × 4 s injections (1 × 0.5 µL sacrificial injection followed by 19 × 2 µL injections) with a 120 s recovery time between each injection. The cell temperature was maintained at 25 °C with DP measured over this time. All buffers were matched prior to ITC through gel filtration or dialysis overnight. Data was fitted and analysed using NITPIC and SEDPHAT and displayed using GUSSI (NIH).[175]

7.3.11 Surface plasmon resonance

Surface plasmon resonance was used as an additional method to ITC for studying the binding of the PanZ-AcCoA/EtdtCoA using immobilised PanD(T57V), functionalised with a biotin using N-terminal sortase-mediated labelling.

All SPR experiments were performed using a Biocore 3000 instrument alongside streptavidin-coated chips (Sensor Chip SA, GE Healthcare). Firstly, the system was equilibrated

using filtered and de-gassed SPR buffer (Tris-buffered saline + 0.1 % Tween-20). All four flow cells of the sensor chip were conditioned using $3 \times 40\ \mu L$ injections of conditioning solution (1 M NaCl and 50 mM NaOH) at $40\ \mu L/min$. Biotinylated-PanD was diluted to a concentration of 5 nM and injected over three of the four flow cells individually using $20\ \mu L$ at differing rates in order to achieve two cells at a similar population, and one at approximately half of that, while keeping one empty as a control. The experiments were then carried out using a series of concentrations of WTPanZ with or without different amounts of AcCoA or EtdtCoA. The flow rate used was $50\ \mu L/min$ for $100\ \mu L$ injections with a 10 minute dissociation time. The experimental data was then processed using BIAevaluation software and Origin 2016.

8 Appendix

8.1 Final compound NMR spectra

8.1.1 2.04: Fmoc-Thr-(OtBu)-Gc-OH
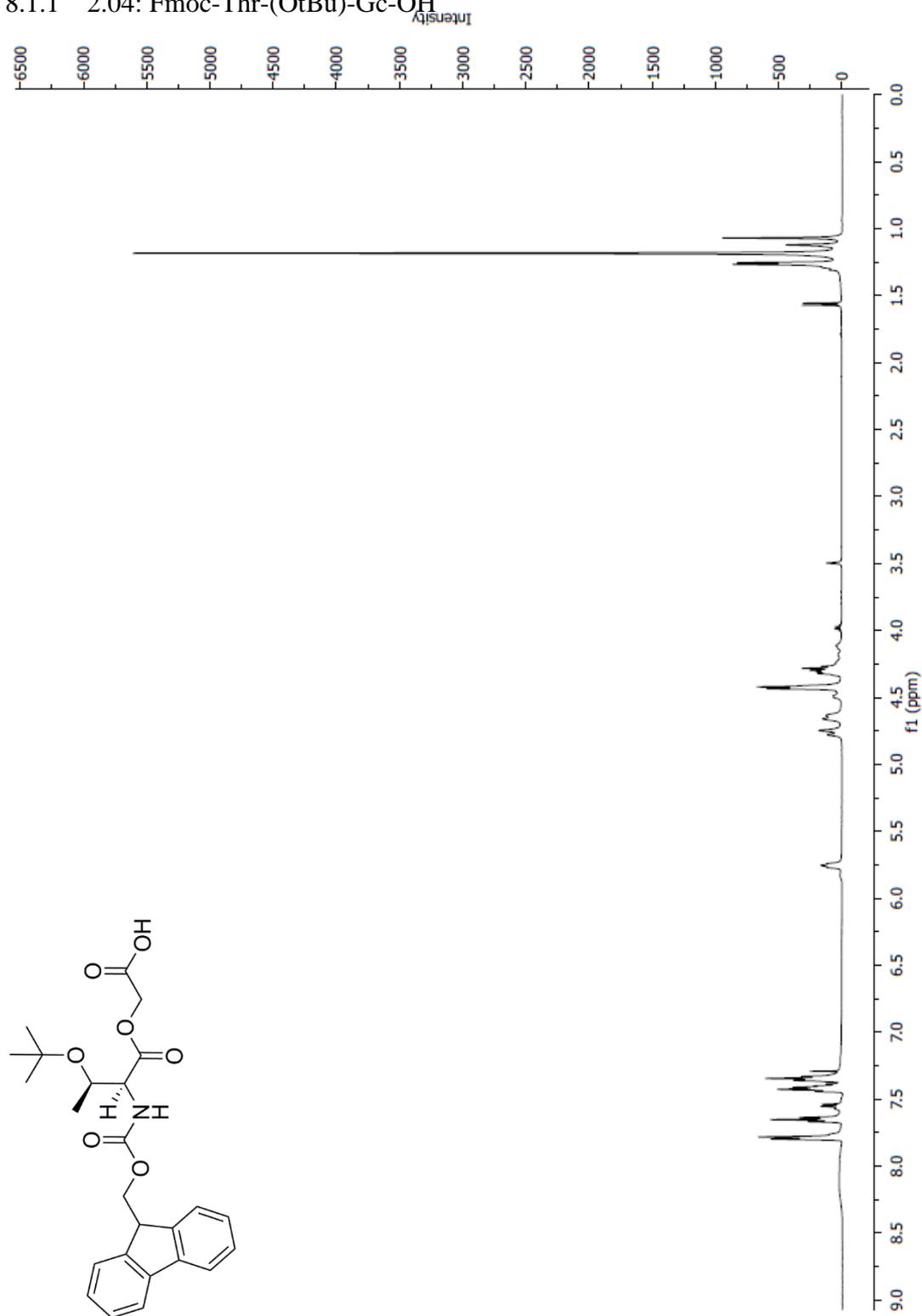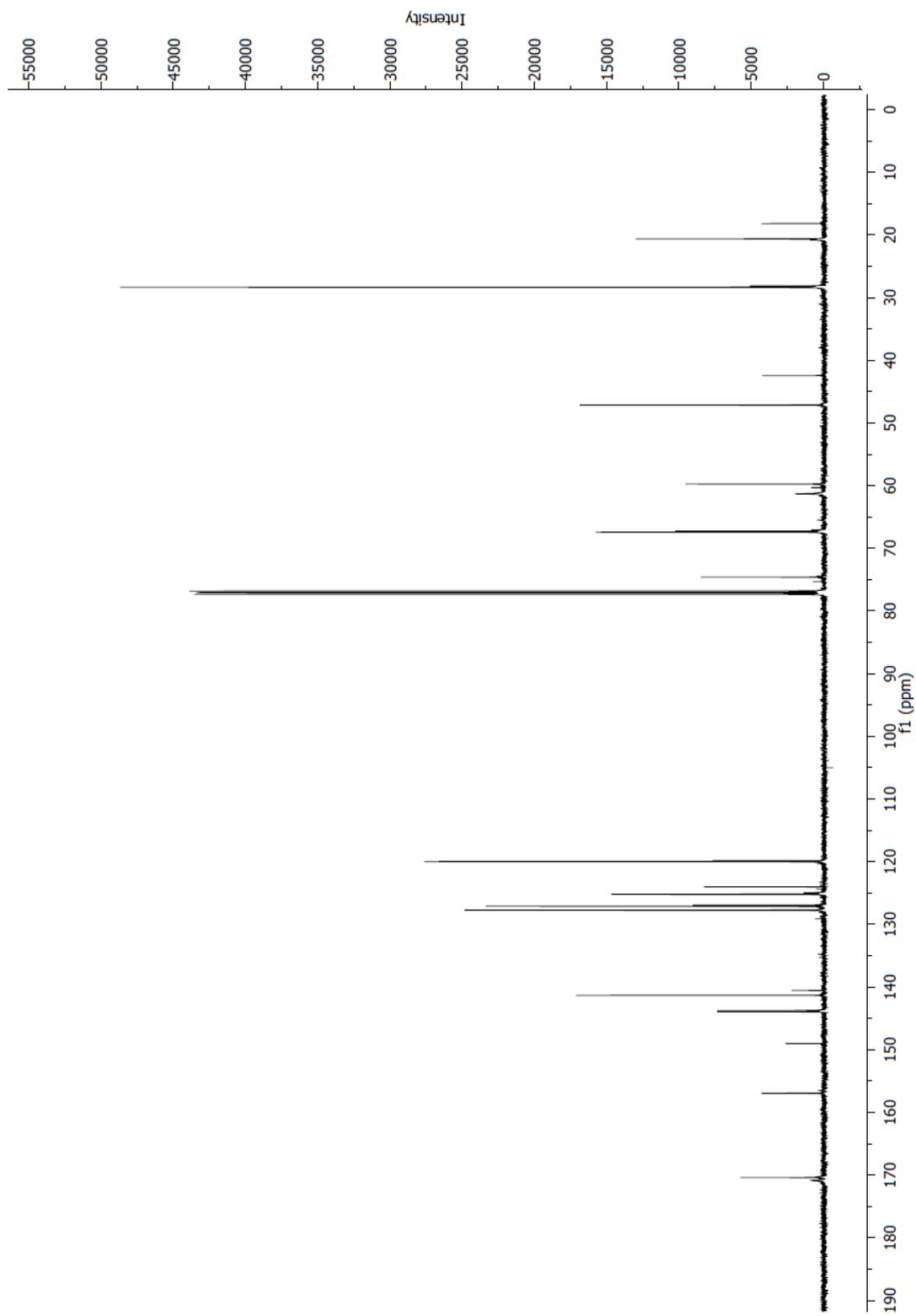
8.1.2 3.05(syn): (1R, 8S, 9S)-bicyclo[6.1.0]non-4-yn-9-yl-methyl(4-nitrophenyl) carbonate
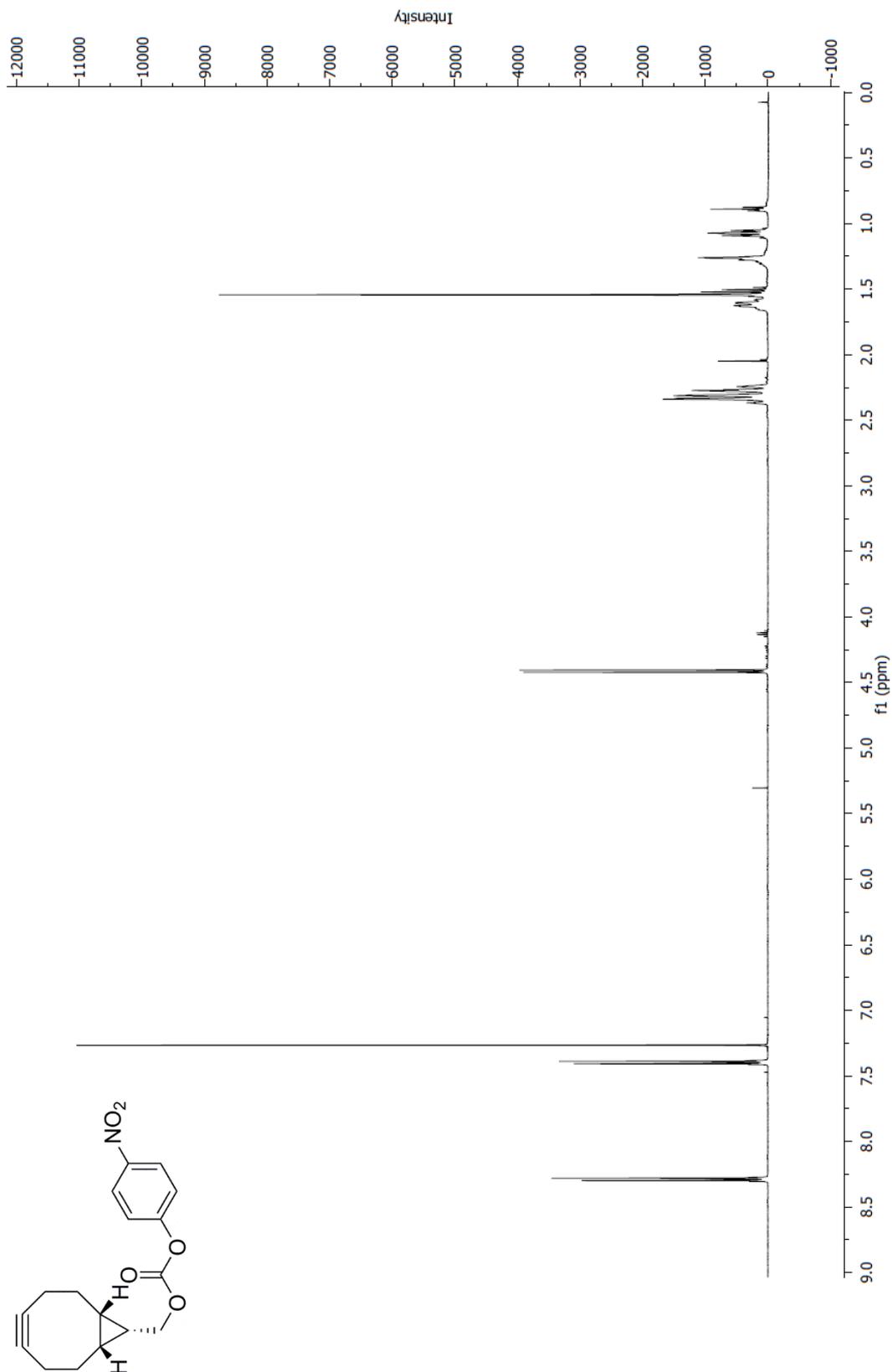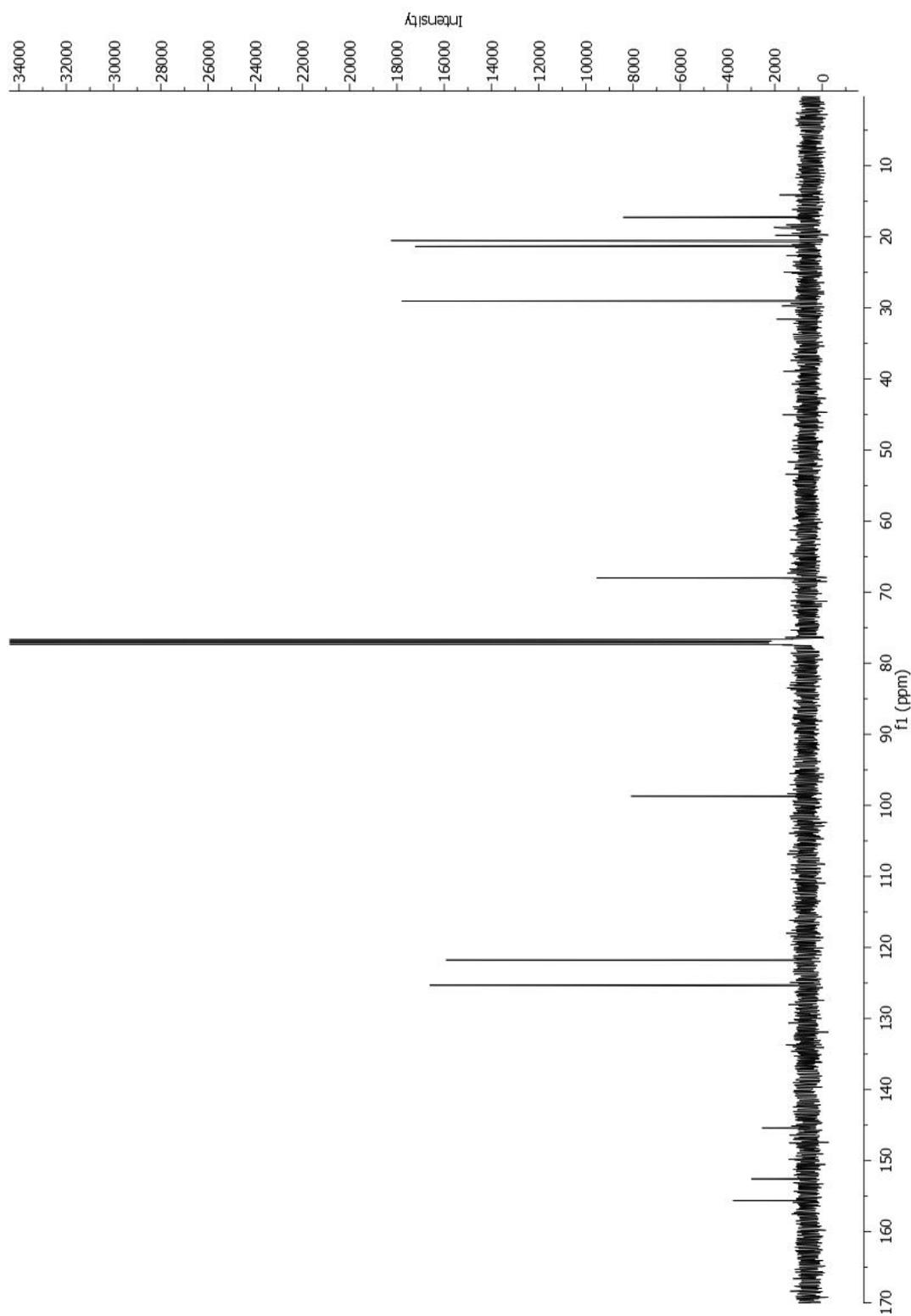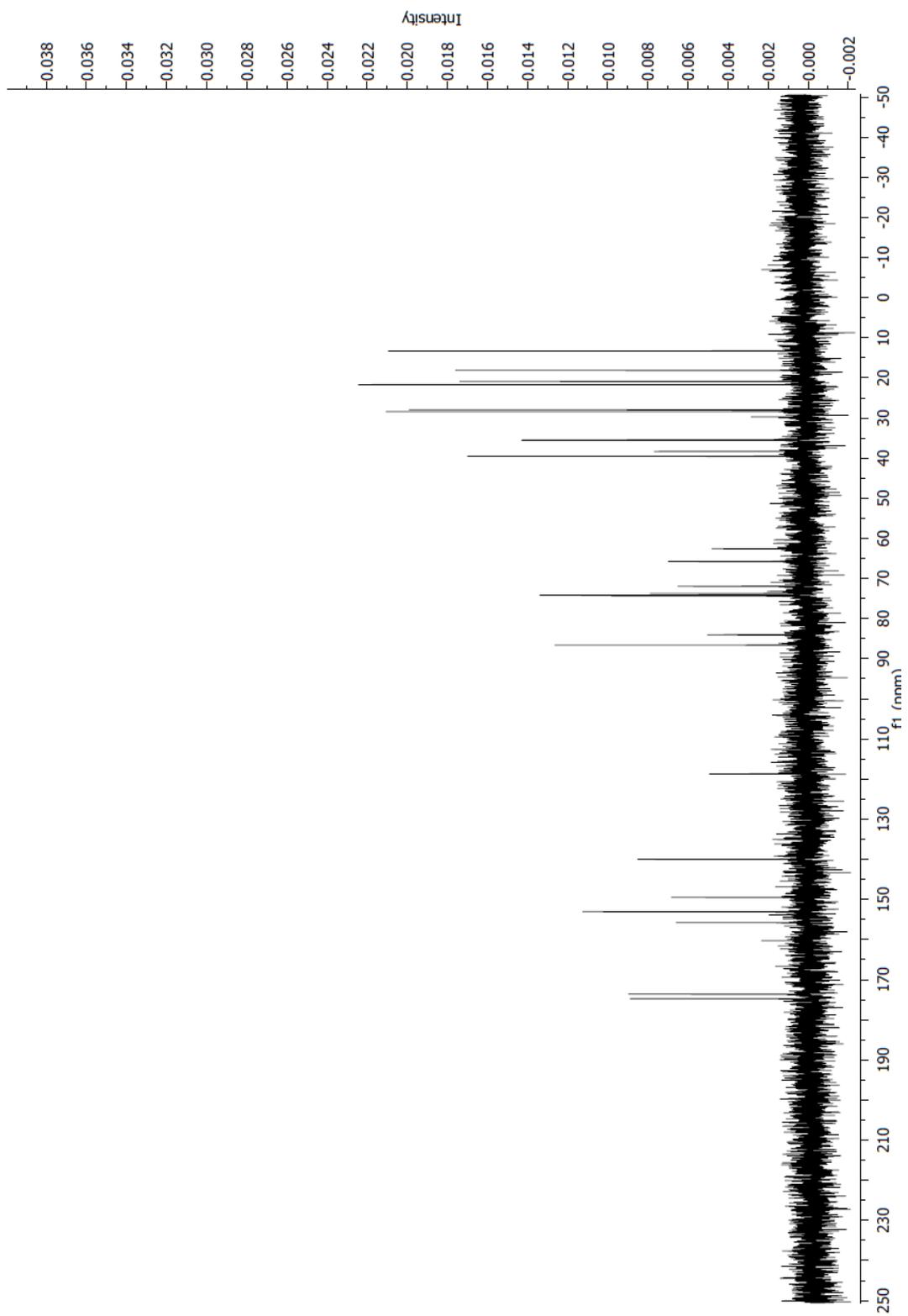
8.1.3   4.01b: *N*-pentylpantothenamide
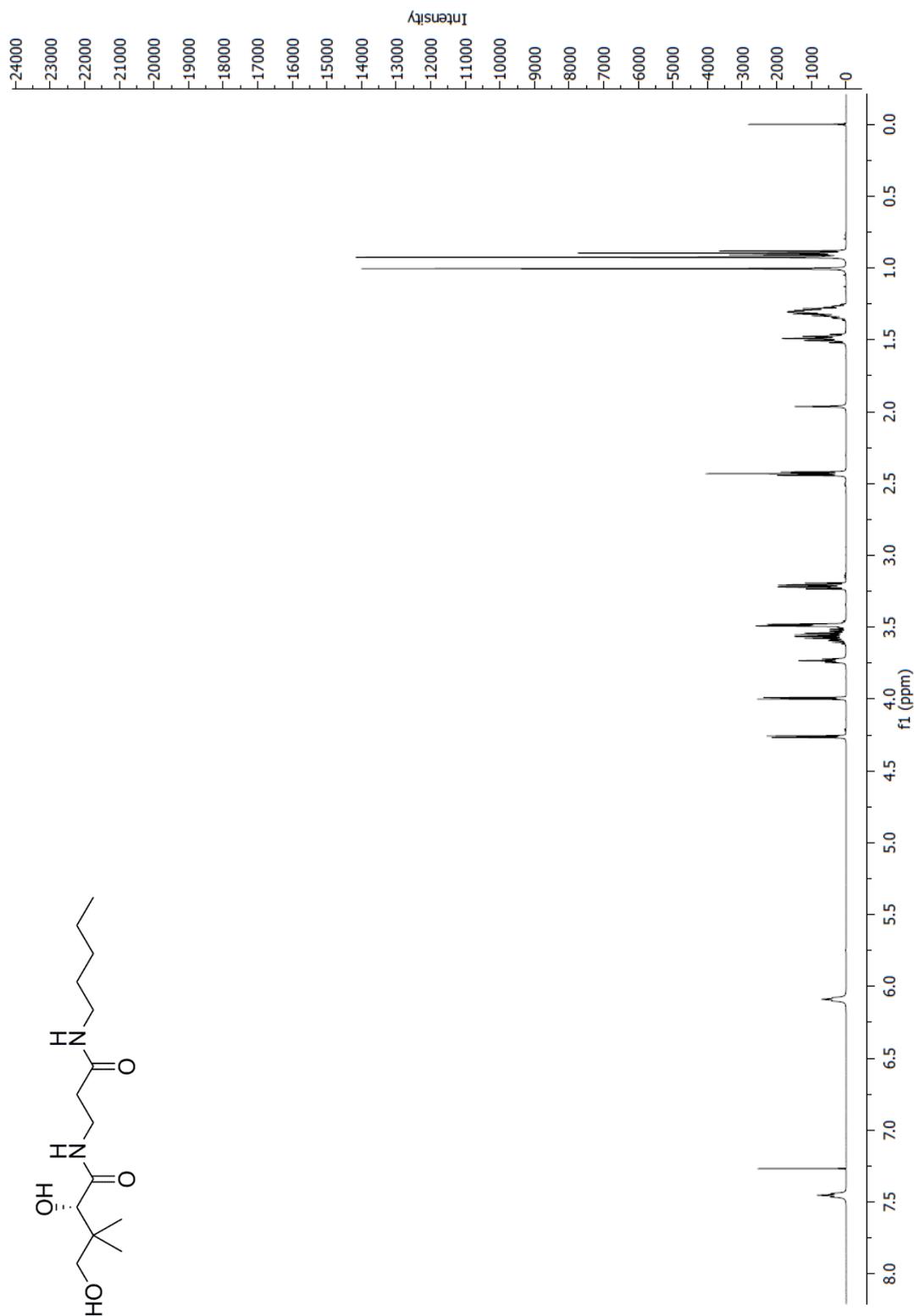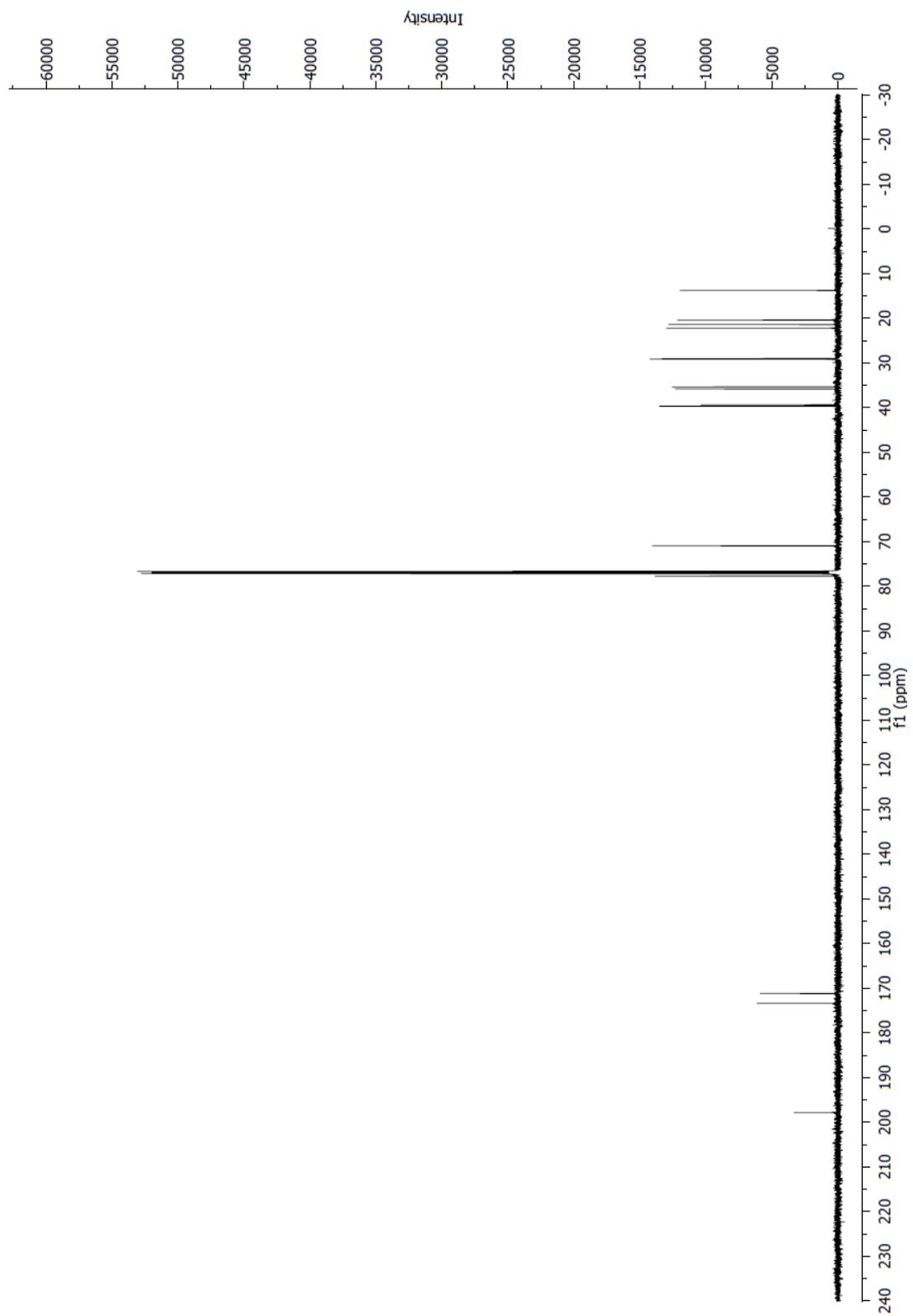
8.1.4   4.02b: Ethyldethia Coenzyme A
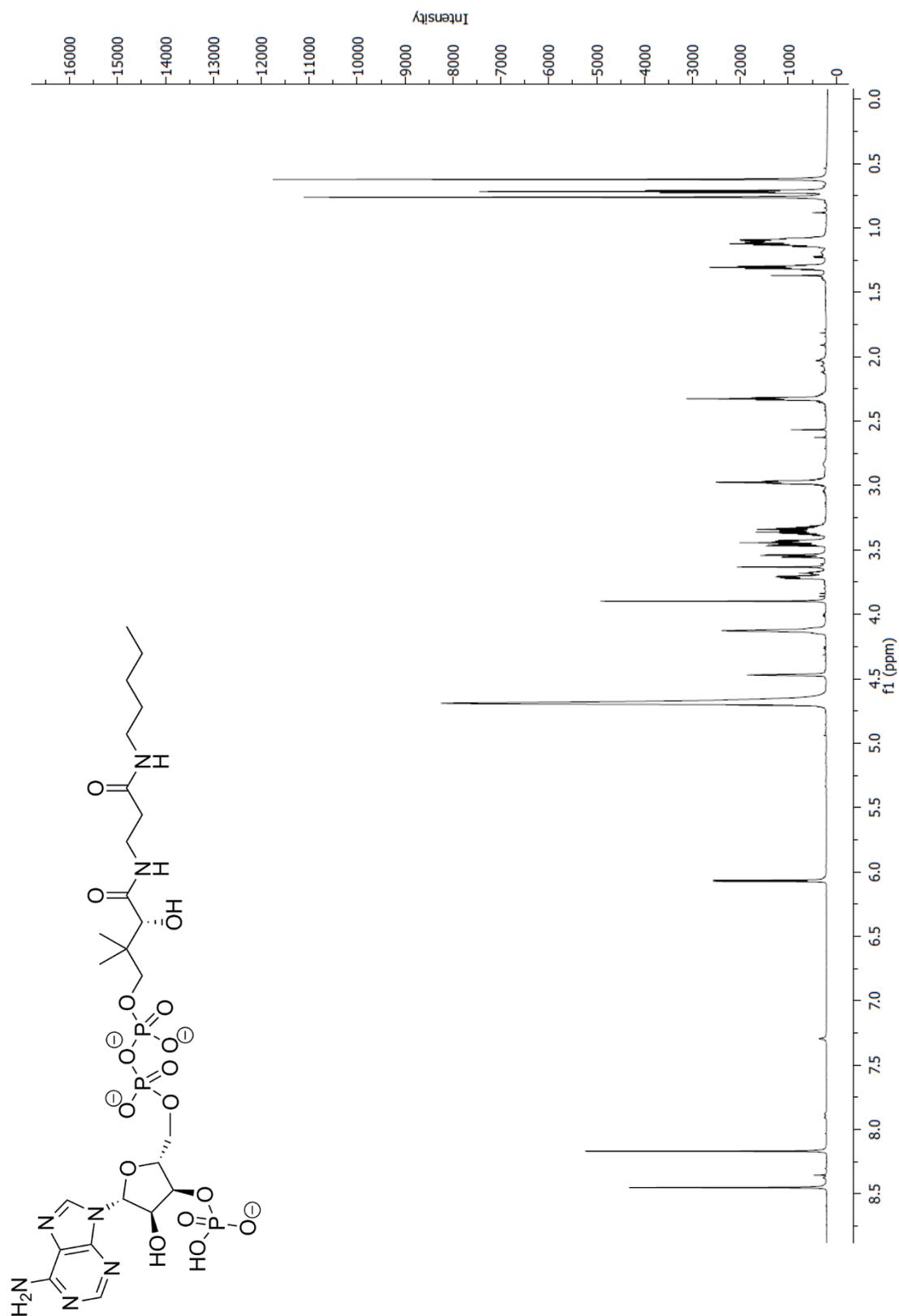
## 8.2 Raw mass spectrometry data from labelling experiments

8.2.1 Labelling of CTB with FITC-depsipeptide

The following spectra show the overlaid raw data for time points 0h (black) and 2h (red) which was deconvoluted (using the maximum entropy setting) to make Figure 2.12. The green label indicates excess FITC-depsipeptide species within the sample.
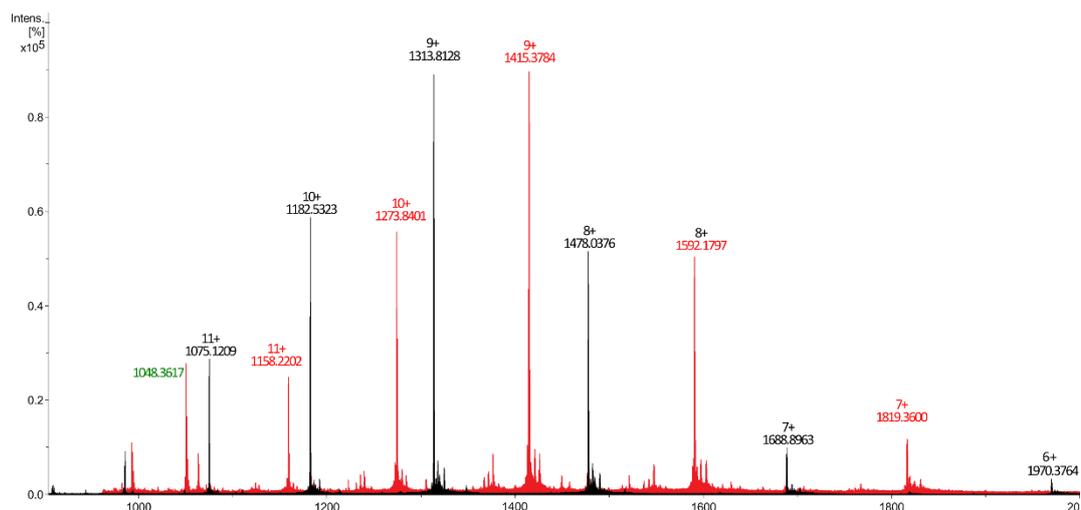
8.2.2 Labelling of MBP with FITC-depsipeptide

The following spectra show the overlaid raw data for time points 0h (black) and 4h (red) which was deconvoluted (using the maximum entropy setting) to make Figure 2.11.
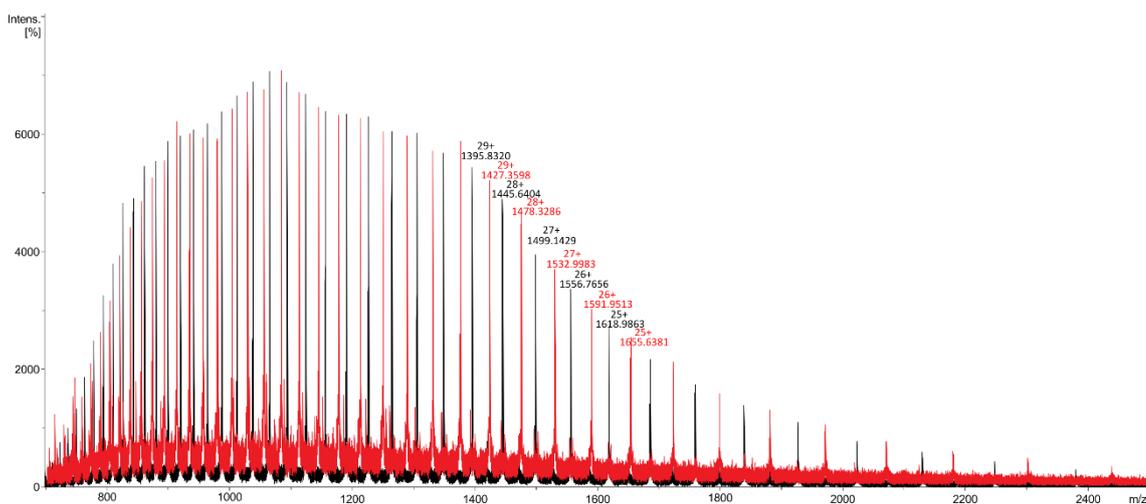
8.3 Plasmids

Plasmid	Use	Cell line used	Origin
pET30b	SrtA 7M	BL21 (DE3)	Addgene, Hidde Ploegh[80,176]
pET28a	WT SrtA	BL21 (DE3)	Williamson et al.[81]
	DNA for digestion	XL10	Section 3.3.2.1
	DNA for PCR	XL10	Section 3.3.2.2
	G-MupB	BL21 (DE3)	Jennifer Tomlinson
pET11a	PanD (T57V)	BL21 ΔDZ (DE3)	Webb et al.[177]
	GGSSS-PanD (T57V)	BL21 ΔDZ (DE3)	Section 2.2.3
	PanD (S25A)	BL21 ΔDZ (DE3)	Schmitzberger et al.[178]
pBAD24	WT PanZ	BL21 ΔDZ (DE3)	Nozaki et al.[179]
	PanZ (R73A)	BL21 ΔDZ (DE3)	Arnott & Nozaki et al.[164]
pRSETa	GV-mCherry	BL21 (DE3)	Matthew Balmforth
pMal-c5x	GVGK-MBP	BL21 (DE3)	Section 2.2.1
pCA24N	CoaD	AG1(ME5305)	Aska library, Kitagawa et al.[180]
	CoaE	AG1(ME5305)	Aska library, Kitagawa et al.[180]
pUC18	EcoRV digestion	XL10	Section 3.3.4

8.4   Primers

All the following primer sequences are written from 5' to 3'.

8.4.1   Protein site-directed mutagenesis

8.4.1.1   PanD(T57V): TEV site and GGSSS introduction on N-terminus

Forward: ATGGTGAGAATCTGTACTTCCAAGGCGGTAGCTCTAGCATGATTCG
             CACGATGCTGCA

Reverse: TGCAGCATCGTGCGAATCATGCTAGAGCTACCGCCTTGGAAGTACA
             GATTCTCACCAT

Sequencing: TAATACGACTCACTATAGGG (t7)

8.4.1.2   MBP: GVGK extension introduced to N-terminus

Forward: CCAACAAGGACCATAGATTATGGGCGTGGGTAAAATCGAAGAAG
             GTAAACTGG

Reverse: CCAGTTTACCTTCTTCGATTTTACCCACGCCCATAATCTATGGTCC
             TTGTTGG

Sequencing: GGCAAATATTCTGAAATGAGC (t12-1-1)

8.4.1.3   MBP: Introduction of stop codon at C-terminus

Forward: CTGAAAGACGCGCAGACTTAATCGAGCTCGAACAACAAC

Reverse: GTTGTTGTTCGAGCTCGATTAAGTCTGCGCGTCTTTCAG

Sequencing: CAGCAAAGTGAATTATGGTGTAACGGTAC

8.4.2 Azide-functionalised DNA

The azide-oligonucleotide has the sequence from 5' to 3':

N$_3$-CGTTAACAGCTGAGGAGCGC-OH

8.4.2.1 Inserting

Forward:	CCTCAGCGCTCCTCAGCAGCTGTTAACGCATTCCTTGCATGCACCA TTCCTTGC
Reverse:	GAATGCGTTAACAGCTGCTGAGGAGCGCTGAGGCGAAACAAGCGC TCATGAGC
Sequencing:	GGTTATGCTAGTTATTGCTC

8.4.2.2 Adding and removing restriction sites for the enzymatic route

Removal of BstZ17i site at ~3 kbp:

Forward:	GTAGCGATAGCGGAGTGTTTACTGGCTTAACTATGCGG
Reverse:	CCGCATAGTTAAGCCAGTAAACACTCCGCTATCGCTAC
Sequencing:	GAAAACCTCTGACACATGCA

Changing PsiI site to BstZ17i site:

Forward:	CGAAATCGGCAAAATCCCGTATACATCAAAAGAATAGACCG
Reverse:	CGGTCTATTCTTTTGATGTATACGGGATTTTGCCGATTTCG
Sequencing:	TAATACGACTCACTATAGGG (t7)

8.4.2.3 PCR primers for amplification of 1 kb strand

The test reactions employed a reverse primer which was not functionalised with the azide. The forward primer was used for all reactions and designed to be 1 kb upstream of the forward primer.

Forward: GGCAAAATCCCGTATACATCAAAAG

Reverse: CGTTAACAGCTGAGGAGCGC

8.4.3   Introduction of EcoRV restriction sites in pUC18

8.4.3.1   Mutation A at 279 bp

Forward: CTGTTGGGAAGGGCGATATCTGCGGGCCTCTTCGC

Reverse: GCGAAGAGGCCCGCAGATATCGCCCTTCCCAACAG

Sequencing: GAGTGCACCATATGCGGTGTG

8.4.3.2   Mutation B at 793 bp

Forward: CCACAGAATCAGGGGATATCGCAGGAAAGAACATG

Reverse: CATGTTCTTTCCTGCGATATCCCCTGATTCTGTGG

Sequencing: GTTTGCGTATTGGGCGCTCTTC

8.4.3.3   Mutation C at 1469 bp

Forward: GGATCTCAAGAAGATCCTTTGATATCTTCTACGGGGTCTGACGC

Reverse: GCGTCAGACCCCGTAGAAGATATCAAAGGATCTTCTTGAGATCC

Sequencing: GATCCGGCAAACAAACCACCGC

8.4.3.4   Mutation D at 129 bp

Forward: CAG CGG GTG TTG GCG GAT ATC GGG GCT GGC TTA ACT ATG

Reverse: CAT AGT TAA GCC AGC CCC GAT ATC CGC CAA CAC CCG CTG

Sequencing: GTGCCACCTGACGTCTAAGA

8.5 Protein sequences

8.5.1 Sortase 7M

```
          10         20         30         40         50         60
MQAKPQIPKD KSKVAGYIEI PDADIKEPVY PGPATREQLN RGVSFAKENQ SLDDQNISIA

          70         80         90        100        110        120
GHTFIDRPNY QFTNLKAAKK GSMVYFKVGN ETRKYKMTSI RNVKPTAVEV LDEQKGKDKQ

         130        140        150
LTLITCDDYN EETGVWETRK IFVATEVKLE HHHHHH
```

8.5.2 Wild-type Sortase A

```
          10         20         30         40         50         60
MGSSHHHHHH SSGLVPRGSH MKPHIDNYLH DKDKDEKIEQ YDKNVKEQAS KDKKQQAKPQ

          70         80         90        100        110        120
IPKDKSKVAG YIEIPDADIK EPVYPGPATP EQLNRGVSFA EENESLDDQN ISIAGHTFID

         130        140        150        160        170        180
RPNYQFTNLK AAKKGSMVYF KVGNETRKYK MTSIRDVKPT DVGVLDEQKG KDKQLTLITC

         190        200
DDYNEKTGVW EKRKIFVATE VK
```

8.5.3 GGSSS-PanD (T57V)

```
          10         20         30         40         50         60
MRGSHHHHHH GENLYFQGGS SSMIRTMLQG KLHRVKVTHA DLHYEGSCAI DQDFLDAAGI

          70         80         90        100        110        120
LENEAIDIWN VTNGKRFSVY AIAAERGSRI ISVNGAAAHC ASVGDIVIIA SFVTMPDEEA

         130        140
RTWRPNVAYF EGDNEMKRTA KAIPVQVA
```

8.5.4 PanD (S25A)

```
          10         20         30         40         50         60
MIRTMLQGKL HRVKVTHADL HYEGACAIDQ DFLDAAGILE NEAIDIWNVT NGKRFSTYAI

          70         80         90        100        110        120
AAERGSRIIS VNGAAAHCAS VGDIVIIASF VTMPDEEART WRPNVAYFEG DNEMKRTAKA

IPVQVA
```

8.5.5 Wild-type PanZ

```
          10         20         30         40         50         60
MKLTIIRLEK FSDQDRIDLQ KIWPEYSPSS LQVDDNHRIY AARFNERLLA AVRVTLSGTE

          70         80         90        100        110        120
GALDSLRVRE VTRRRGVGQY LLEEVLRNNP GVSCWWMADA GVEDRGVMTA FMQALGFTAQ

         130
QGGWEKCSGS GHHHHH
```

8.5.6 PanZ (R73A)

```
          10         20         30         40         50         60
MKLTIIRLEK FSDQDRIDLQ KIWPEYSPSS LQVDDNHRIY AARFNERLLA AVRVTLSGTE

          70         80         90        100        110        120
GALDSLRVRE VTARRGVGQY LLEEVLRNNP GVSCWWMADA GVEDRGVMTA FMQALGFTAQ

         130
QGGWEKCSGS GHHHHH
```

8.5.7 GVGK-MBP

```
          10         20         30         40         50         60
GVGKIEEGKL VIWINGDKGY NGLAEVGKKF EKDTGIKVTV EHPDKLEEKF PQVAATGDGP

          70         80         90        100        110        120
DIIFWAHDRF GGYAQSGLLA EITPDKAFQD KLYPFTWDAV RYNGKLIAYP IAVEALSLIY

         130        140        150        160        170        180
NKDLLPNPPK TWEEIPALDK ELKAKGKSAL MFNLQEPYFT WPLIAADGGY AFKYENGKYD

         190        200        210        220        230        240
IKDVGVDNAG AKAGLTFLVD LIKNKHMNAD TDYSIAEAAF NKGETAMTIN GPWAWSNIDT

         250        260        270        280        290        300
SKVNYGVTVL PTFKGQPSKP FVGVLSAGIN AASPNKELAK EFLENYLLTD EGLEAVNKDK

         310        320        330        340        350        360
PLGAVALKSY EEELVKDPRI AATMENAQKG EIMPNIPQMS AFWYAVRTAV INAASGRQTV

DEALKDAQT
```

8.5.8 GGG-CTB

```
        10         20         30         40         50         60
GGGTPQNITD LCAEYHNTQI YTLNDKIFSY TESLAGKREM AIITFKNGAI FQVEVPGSQH

        70         80         90        100
IDSQKKAIER MKDTLRIAYL TEAKVEKLCV WNNKTPHAIA AISMAN
```

8.5.9 GV-mCherry

```
        10         20         30         40         50         60
GVRGSHHHHH HGMASMTGGQ QMGRDLYDDD DKDPATMVSK GEEDNMAIIK EFMRFKVHME

        70         80         90        100        110        120
GSVNGHEFEI EGEGEGRPYE GTQTAKLKVT KGGPLPFAWD ILSPQFMYGS KAYVKHPADI

       130        140        150        160        170        180
PDYLKLSFPE GFKWERVMNF EDGGVVTVTQ DSSLQDGEFI YKVKLRGTNF PSDGPVMQKK

       190        200        210        220        230        240
TMGWEASSER MYPEDGALKG EIKQRLKLKD GGHYDAEVKT TYKAKKPVQL PGAYNVNIKL

       250        260        270
DITSHNEDYT IVEQYERAEG RHSTGGMDEL YK
```

8.5.10 G-MupB

```
        10         20         30         40         50         60
HHHHHHHHHH SSGHENLYFQ GSLENENIIE EQKILNFWKE ENIFKKSIDN RKNDNPFVFY

        70         80         90        100        110        120
DGPPTANGLP HTGHVLGRVI KDLFARYKTM QGFYVERKAG WDTHGLPVEL GVEKKLGIKD

       130        140        150        160        170        180
KNEIEKYGIE KFINECKNSV FMYEKQWREF SELIGYWVDM EKPYKTMDNT YIESIWYILS

       190        200        210        220        230        240
DFHKKGLLYK GHKVTPYCPS CETSLSSHEV AQGYKEVKDI SVILKFPILD SDENFLVWTT

       250        260        270        280        290        300
TPWSLPGNIA LAINAEEIYV KVNYDNEIFI IMESLLQSVF KDEDNIDIVS KHKGKEFVGK

       310        320        330        340        350        360
EYLAPFPNKS LMNNENSYKV LPADFVTNKD GTGIVHIAPA YGEDDYKLVQ ENNIPFINVI

       370        380        390        400        410        420
DSRGKYNQDS PIFKGELAKE SDINIIKELT HLNLLFKKEK YEHSYPFCWR CDNPLIYYAM

       430        440        450        460        470        480
EGWFIKTTAY KNEIKENNQK IEWYPDHIKN GRFGNFLDNM IDWNIGRKRY WGTPLNIWKC
```

```
        490        500        510        520        530        540
STCSHEFSPK SINDLIQHSI EDIPSDIELH RPYIDNVKCK CQNCGGDMCR EEEVIDVWFD

        550        560        570        580        590        600
SGSMPFAQNH YPFSGPIQNS YPADFIAEGV DQTRGWFYSL LVISTIFKGE APYKNALSLG

        610        620        630        640        650        660
HILDSNGQKM SKSKGNVIDP ISMIKTYGAD SLRWTLVSDS VPWTNKRFSE NMVAQSKSRV

        670        680        690        700        710        720
IDTLKNIFNF YNMYQKIDNY DYTRDTPKQL NLLDNWAISR MNSVIKEVEL HLEKYNPTNA

        730        740        750        760        770        780
SRAIGEFINE ISNWYIRRSR SRFWSSEMNE DKKSAYFTLR LILINTCKII APFTPFTSEE

        790        800        810        820        830        840
IHLNLTKKSV HLEDFPQAKE EYINLKLEED MNKVLDIVEK SRSIRNNINI KTKQPLSNMY

        850        860        870        880        890        900
IYDNNNLDNE FLRKYKDIIK DEINVKKINI VSDLDNFLEY DVKPNFSTLG PKLGKDMKQF

        910        920        930        940        950        960
QILFKNIKKE EMNKLINDFD KLQKVFDSLG VTIEEKDFII SKIPKKGFSL SSNDSDRLII

        970        980        990       1000       1010       1020
LDTNLTQELI REGFVRELIR VIQQLRKQQN FNIEERINVV IDIDSDGLLS IKNNINILKE

       1030       1040       1050
NVLINNLKFE KRETMKYFKI NQKEIGIQLM SSFTN
```

8.5.11 CoaD

```
         10         20         30         40         50         60
MQKRAIYPGT FDPITNGHID IVTRATQMFD HVILAIAASP SKKPMFTLEE RVALAQQATA

         70         80         90        100        110        120
HLGNVEVVGF SDLMANFARN QHATVLIRGL RAVADFEYEM QLAHMNRHLM PELESVFLMP

        130        140        150
SKEWSFISSS LVKEVARHQG DVTHFLPENV HQALMAKLA
```

8.5.12 CoaE

```
         10         20         30         40         50         60
MRYIVALTGG IGSGKSTVAN AFADLGINVI DADIIARQVV EPGAPALHAI ADHFGANMIA

         70         80         90        100        110        120
ADGTLQRRAL RERIFANPEE KNWLNALLHP LIQQETQHQI QQATSPYVLW VVPLLVENSL
```
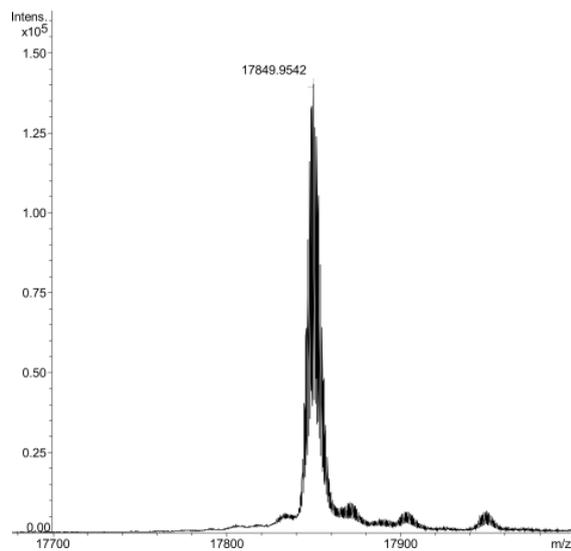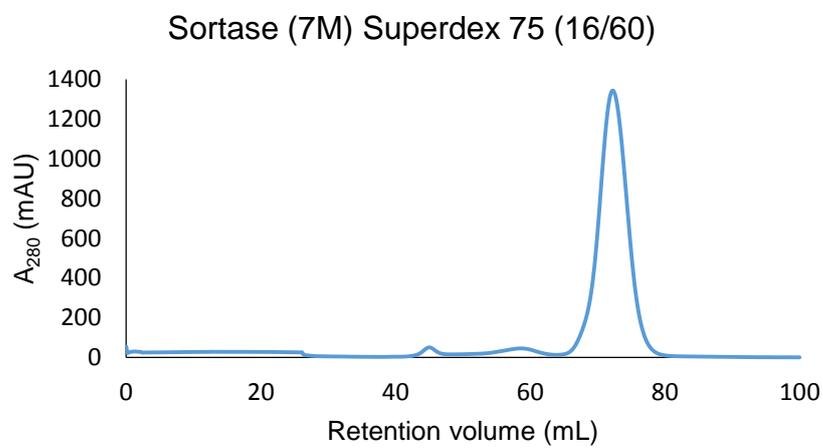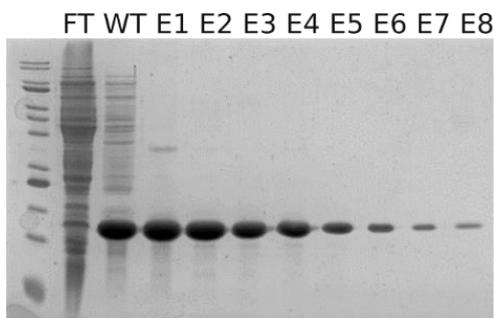
```
          130         140         150         160         170         180
YKKANRVLVV DVSPETQLKR TMQRDDVTRE HVEQILAAQA TREARLAVAD DVIDNNGAPD

          190         200
AIASDVARLH AHYLQLASQF VSQEKP
```

8.6 Protein purification

8.6.1 Sortase 7M

FT WT E1 E2 E3 E4 E5 E6 E7 E8

Sortase (7M) Superdex 75 (16/60)

8.6.2   Wild-type sortase

WT Sortase Superdex 75 (26/60)
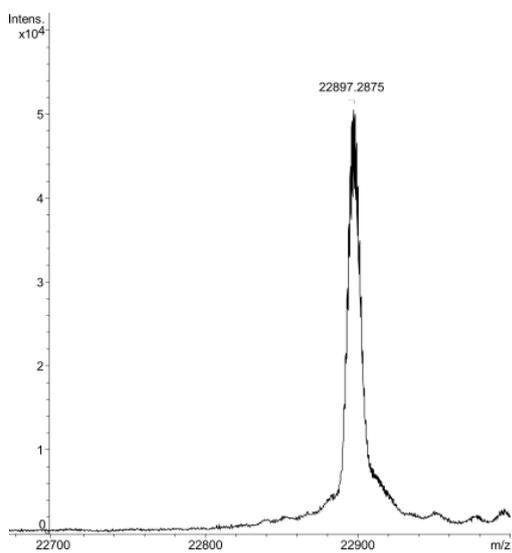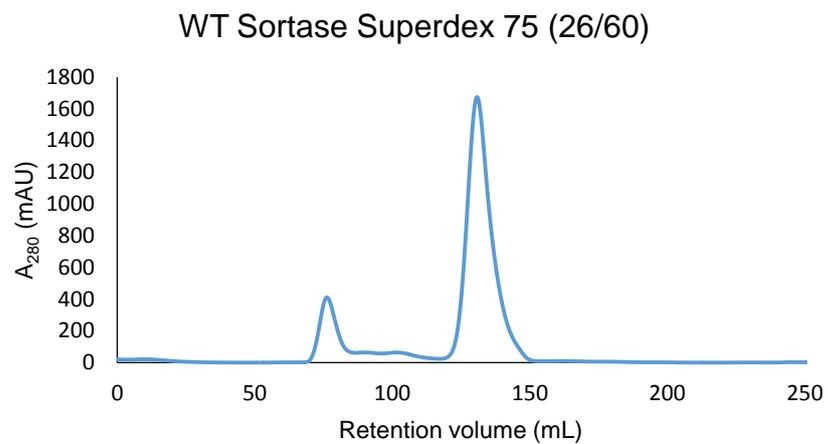
8.6.3 GGSSS-PanD (T57V)

FT FT WT E1 E2 E3 E4 E5 E6
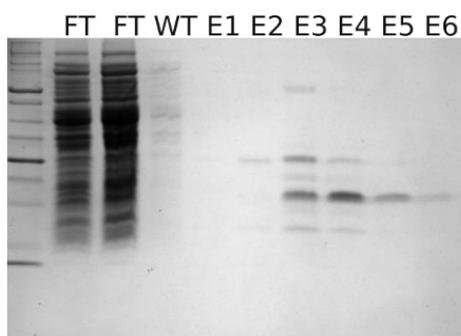
GGSSS-PanD(T57V) Superdex 75 (16/60)
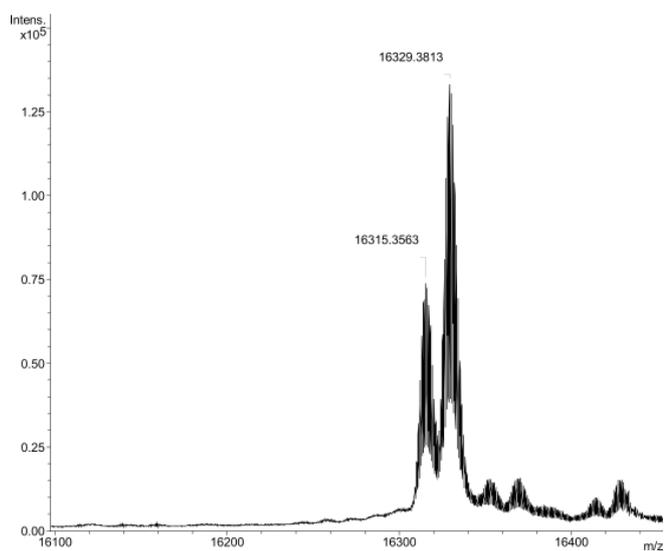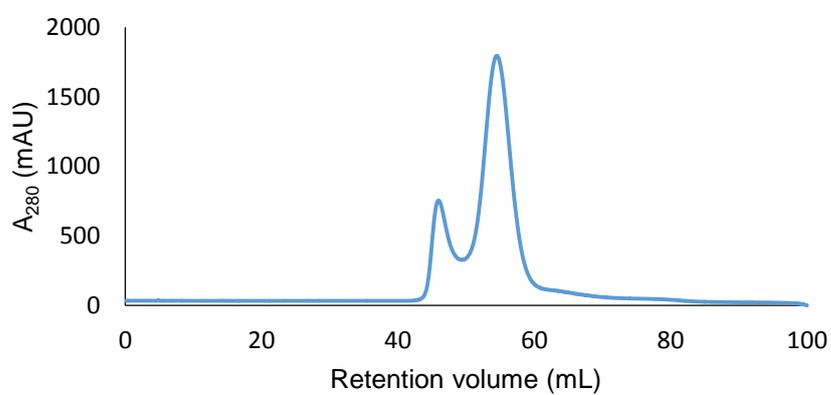
8.6.4 PanD (S25A)
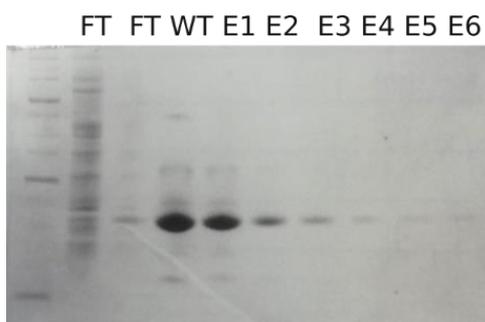
PanD(S25A) Superdex S75 (16/60)

8.6.5 Wild-type PanZ

WTPanZ Superdex 75 (16/60)
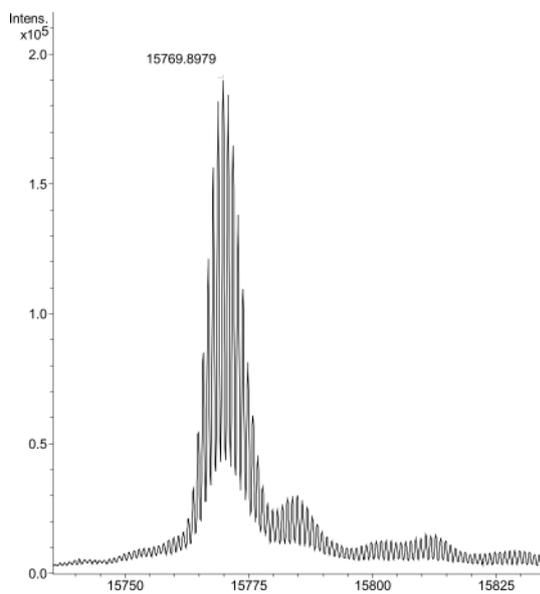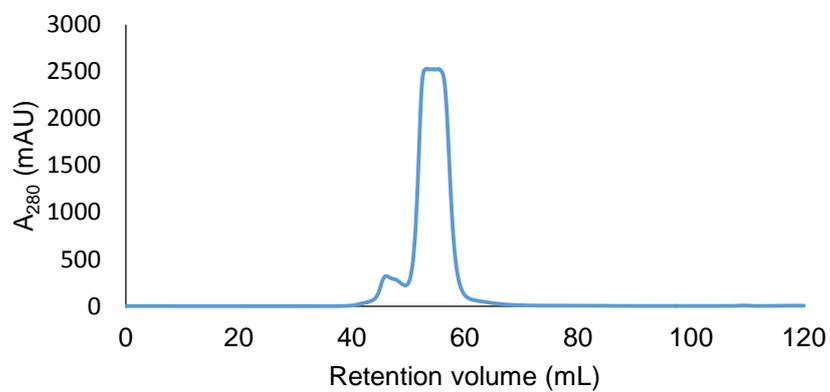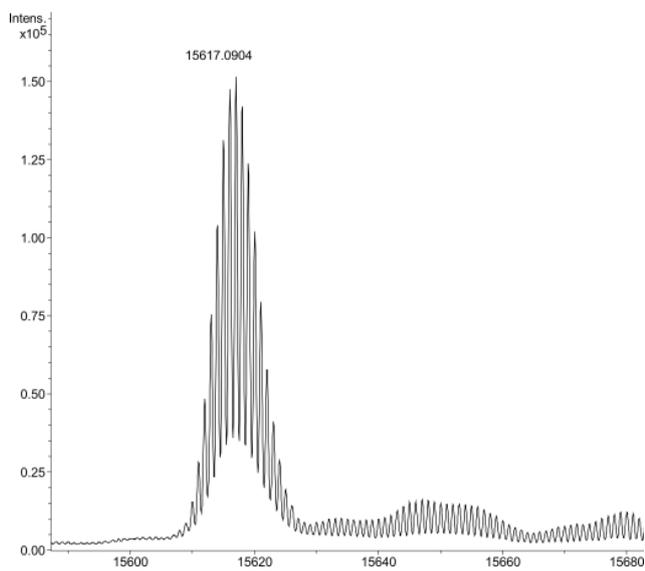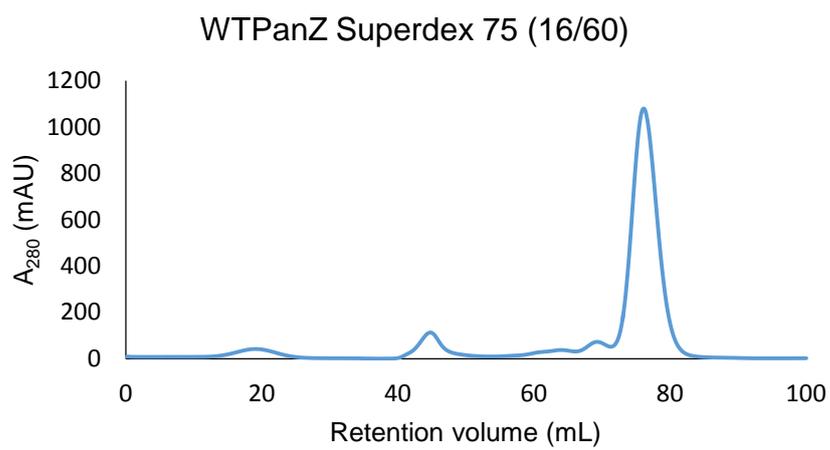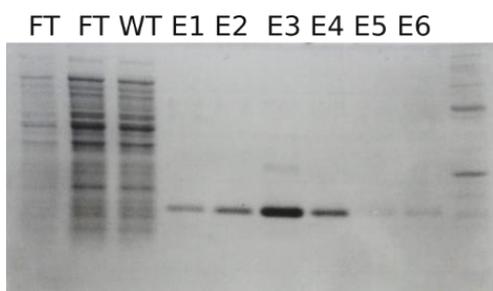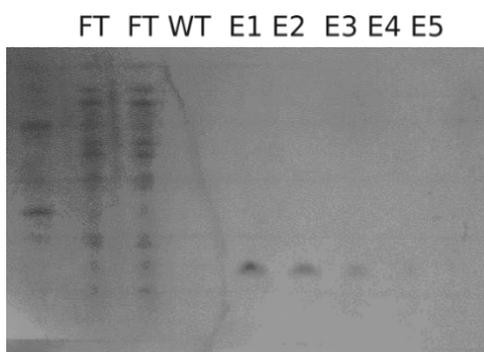
8.6.6   PanZ (R73A)

PanZ(R73A) Superdex S75 (16/60)
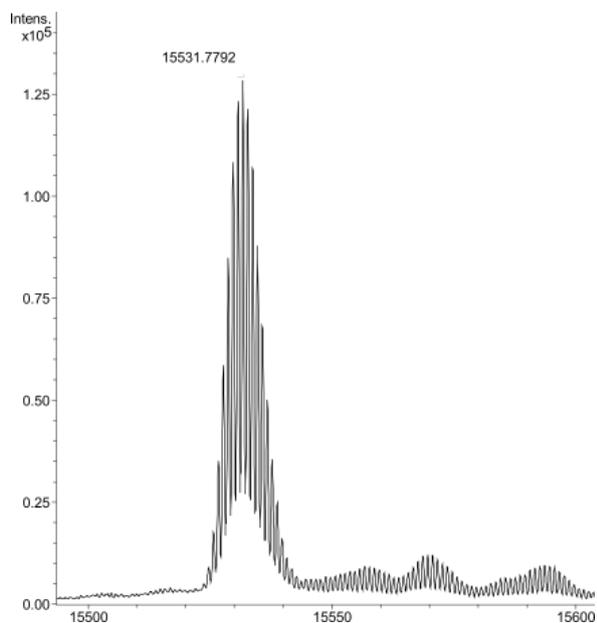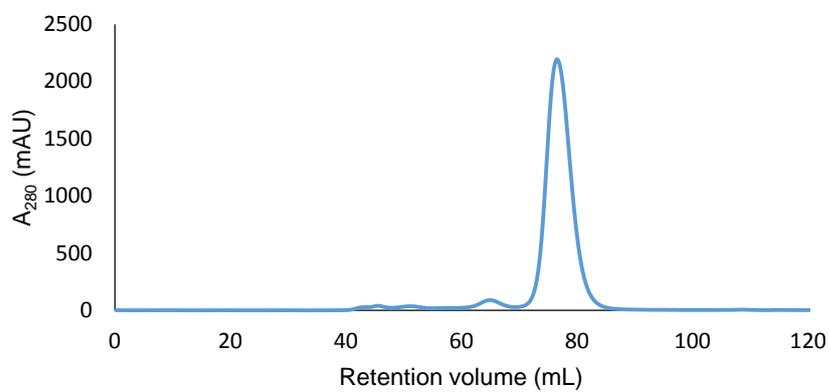
8.6.7 GVGK-MBP
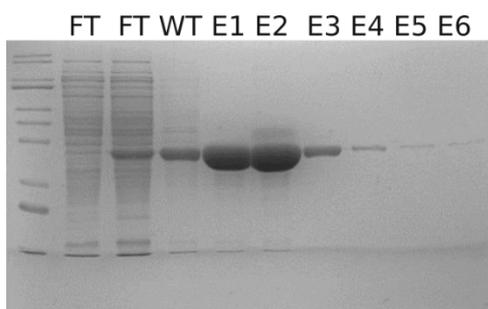
GVGK-MBP Superdex S200 (26/60)
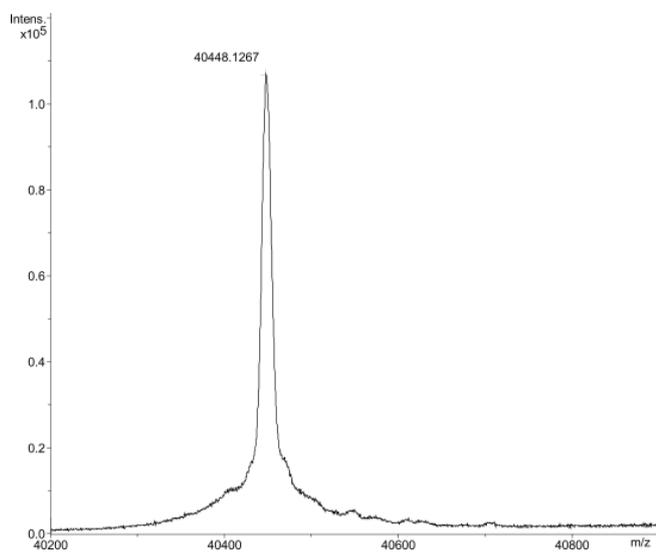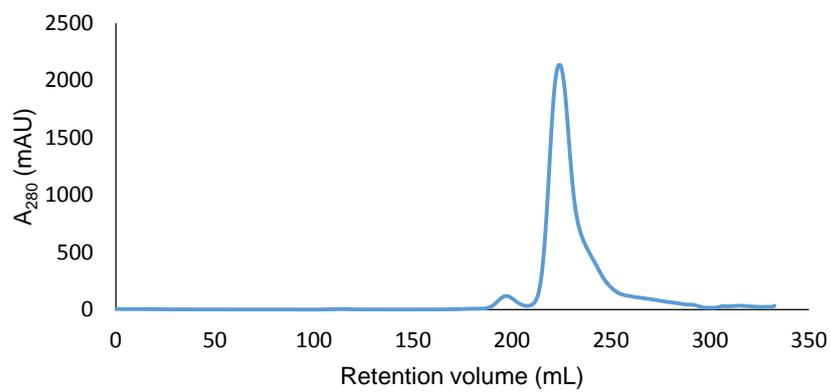
8.6.8 GV-mCherry
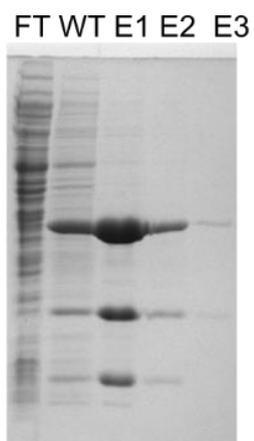
GV-mCherry Superdex S200 (26/60)

8.6.9 G-MupB

G-MBP Superdex 200 (26/60)
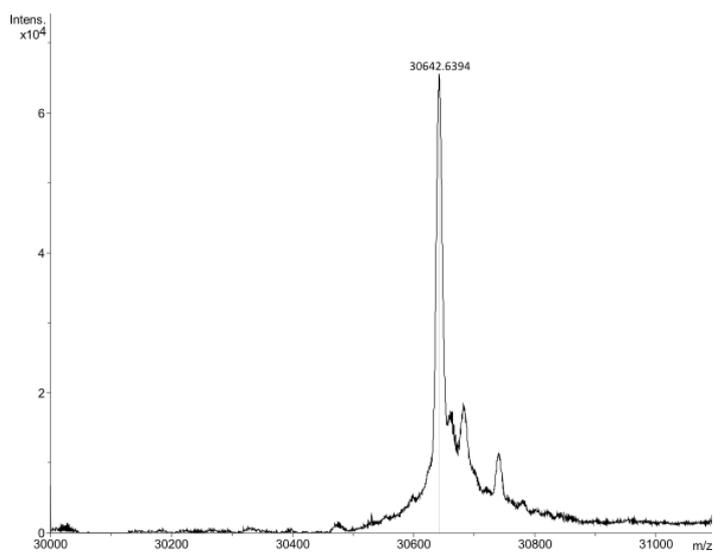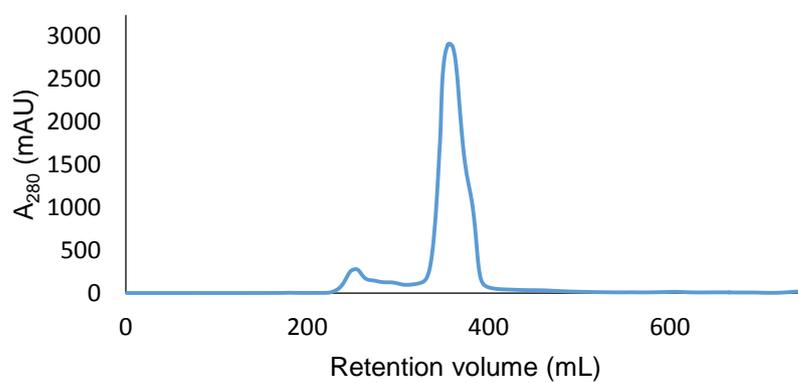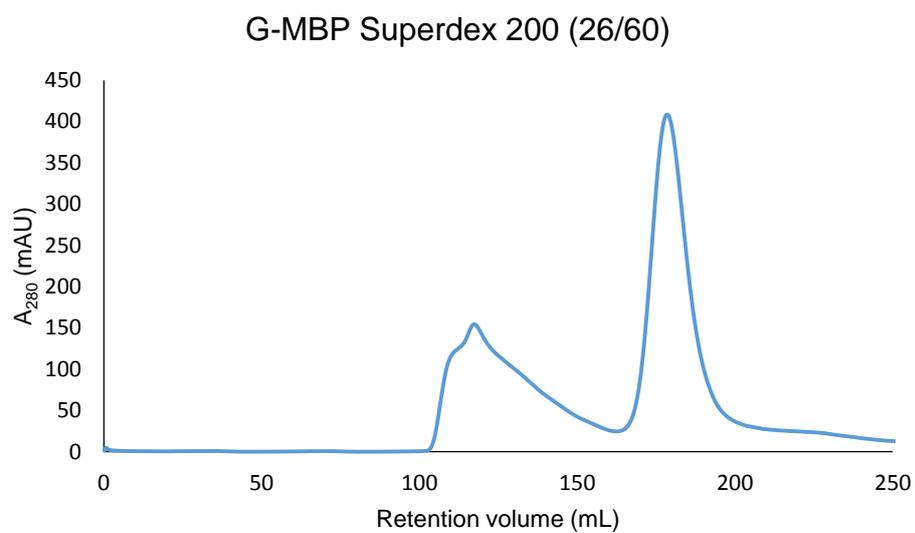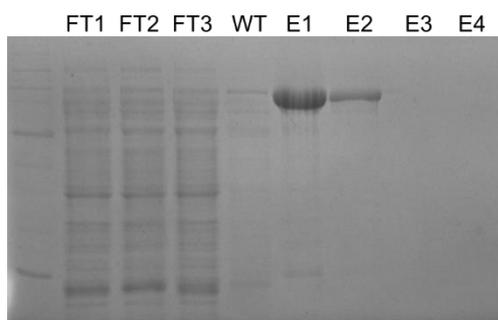
8.6.10 CoaD
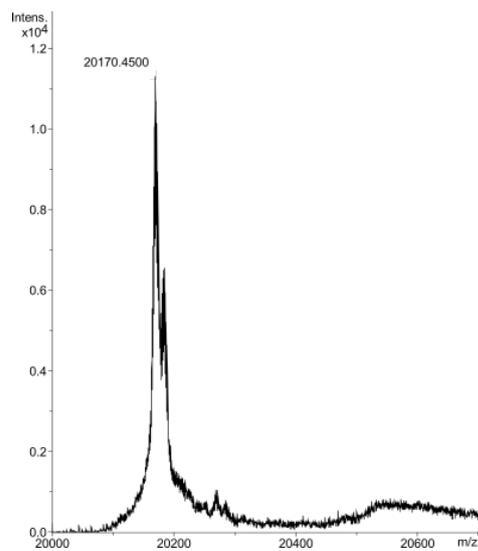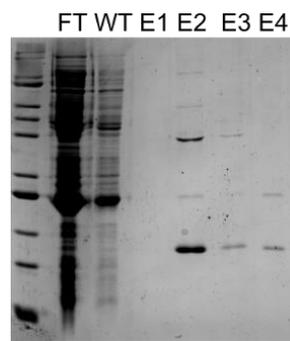
8.6.11 CoaE
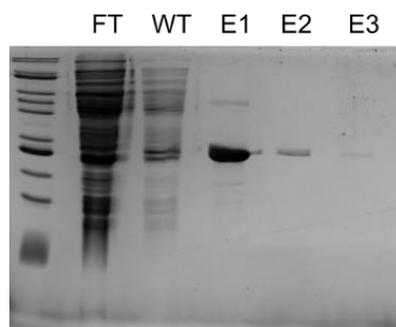
9   References

1   J. M. Chalker and B. G. Davis, Chemical mutagenesis: selective post-expression interconversion of protein amino acid residues, *Curr. Opin. Chem. Biol.*, 2010, **14**, 781–789.

2   D. P. Gamblin, P. Garnier, S. van Kasteren, N. J. Oldham, A. J. Fairbanks and B. G. Davis, Glyco-SeS: Selenenylsulfide-Mediated Protein Glycoconjugation—A New Strategy in Post-Translational Modification, *Angew. Chem. Int. Ed.*, 2004, **43**, 828–833.

3   H. C. Hang, C. Yu, D. L. Kato and C. R. Bertozzi, A metabolic labeling approach toward proteomic analysis of mucin-type O-linked glycosylation, *Proc. Natl. Acad. Sci.*, 2003, **100**, 14846–14851.

4   M. Köhn and R. Breinbauer, The Staudinger ligation-a gift to chemical biology., *Angew. Chem. Int. Ed.* , 2004, **43**, 3106–16.

5   E. Saxon and C. R. Bertozzi, Cell surface engineering by a modified Staudinger reaction., *Science* , 2000, **287**, 2007–10.

6   E. Saxon, J. I. Armstrong and C. R. Bertozzi, A "Traceless" Staudinger Ligation for the Chemoselective Synthesis of Amide Bonds, *Org. Lett.*, 2000, **2**, 2141–2143.

7   F. L. Lin, H. M. Hoyt, H. Van Halbeek, R. G. Bergman and C. R. Bertozzi, Mechanistic Investigation of the Staudinger Ligation, *J. Am. Chem. Soc.*, 2005, **127**, 2686–2695.

8   T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto and S. Yokoyama, Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N(epsilon)-(o-azidobenzyloxycarbonyl) lysine for site-specific protein modification., *Chem. Biol.*, 2008, **15**, 1187–97.

9   K. Lang and J. W. Chin, Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins., *Chem. Rev.*, 2014, **114**, 4764–806.

10   V. V. Rostovtsev, L. G. Green, V. V. Fokin and K. B. Sharpless, A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes, *Angew. Chem. Int. Ed.*, 2002, **41**, 2596–2599.

11   F. Himo, T. Lovell, R. Hilgraf, V. V. Rostovtsev, L. Noodleman, K. B. Sharpless and V. V. Fokin, Copper(I)-Catalyzed Synthesis of Azoles. DFT Study Predicts Unprecedented Reactivity and Intermediates, *J. Am. Chem. Soc.*, 2005, **127**, 210–216.

12   A. Makarem, R. Berg, F. Rank Rominger and B. F. Straub, A Fluxional Copper Acetylide Cluster in CuAAC Catalysis, *Angew. Chem. Int. Ed.* , 2015, **54**, 7431–7435.

13   Q. Wang, T. R. Chan, R. Hilgraf, V. V Fokin, K. B. Sharpless and M. G. Finn, Bioconjugation by copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition., *J. Am. Chem. Soc.*, 2003, **125**, 3192–3.

14   C. C. Liu and P. G. Schultz, Adding new chemistries to the genetic code., *Annu. Rev. Biochem.*, 2010, **79**, 413–44.

15   J. T. Ngo and D. A. Tirrell, Noncanonical amino acids in the interrogation of cellular protein synthesis., *Acc. Chem. Res.*, 2011, **44**, 677–85.

16   A. Deiters, T. A. Cropp, M. Mukherji, J. W. Chin, J. C. Anderson and P. G. Schultz, Adding amino acids with novel reactivity to the genetic code of Saccharomyces cerevisiae., *J. Am. Chem. Soc.*, 2003, **125**, 11782–3.

17   S. C. Fry, Oxidative scission of plant cell wall polysaccharides by ascorbate-induced hydroxyl radicals., *Biochem. J.*, 1998, **332**, 507–15.

18   P.-Y. Liu, N. Jiang, J. Zhang, X. Wei, H.-H. Lin and X.-Q. Yu, The Oxidative Damage of Plasmid DNA by Ascorbic Acid Derivativesin vitro: The First Research on the Relationship between the Structure of Ascorbic Acid and the Oxidative Damage of Plasmid DNA, *Chem. Biodivers.*, 2006, **3**, 958–966.

19   M. R. Levengood, C. C. Kerwood, C. Chatterjee and W. A. van der Donk, Investigation

of the Substrate Specificity of Lacticin 481 Synthetase by Using Nonproteinogenic Amino Acids, *ChemBioChem*, 2009, **10**, 911–919.

20   P. J. Thornalley, Glutathione-dependent detoxification of h-oxoaldehydes by the glyoxalase system: involvement in disease mechanisms and antiproliferative activity of glyoxalase I inhibitors, *Chem. Biol. Interact.*, 1998, **111**, 137–151.

21   C. Besanceney-Webler, H. Jiang, T. Zheng, L. Feng, D. Soriano del Amo, W. Wang, L. M. Klivansky, F. L. Marlow, Y. Liu and P. Wu, Increasing the efficacy of bioorthogonal click reactions for bioconjugation: a comparative study., *Angew. Chem. Int. Ed.*, 2011, **50**, 8051–6.

22   V. Hong, S. I. Presolski, C. Ma and M. G. Finn, Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation., *Angew. Chem. Int. Ed.*, 2009, **48**, 9879–83.

23   J. A. Codelli, J. M. Baskin, N. J. Agard and C. R. Bertozzi, Second-generation difluorinated cyclooctynes for copper-free click chemistry., *J. Am. Chem. Soc.*, 2008, **130**, 11486–93.

24   J. M. Baskin, J. A. Prescher, S. T. Laughlin, N. J. Agard, P. V Chang, I. A. Miller, A. Lo, J. A. Codelli and C. R. Bertozzi, Copper-free click chemistry for dynamic in vivo imaging., *Proc. Natl. Acad. Sci.*, 2007, **104**, 16793–7.

25   N. J. Agard, J. A. Prescher and C. R. Bertozzi, A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems., *J. Am. Chem. Soc.*, 2004, **126**, 15046–7.

26   X. Ning, J. Guo, M. A. Wolfert and G.-J. Boons, Visualizing metabolically labeled glycoconjugates of living cells by copper-free and fast huisgen cycloadditions., *Angew. Chem. Int. Ed.*, 2008, **47**, 2253–5.

27   A. Kuzmin, A. Poloukhtine, M. A. Wolfert and V. V. Popik, Surface Functionalization

Using Catalyst-Free Azide−Alkyne Cycloaddition, *Bioconjug. Chem.*, 2010, **21**, 2076–2085.

28   J. Dommerholt, S. Schmidt, R. Temming, L. J. A. Hendriks, F. P. J. T. Rutjes, J. C. M. van Hest, D. J. Lefeber, P. Friedl and F. L. van Delft, Readily Accessible Bicyclononynes for Bioorthogonal Labeling and Three-Dimensional Imaging of Living Cells, *Angew. Chem. Int. Ed.*, 2010, **49**, 9422–9425.

29   D. Bang and S. B. H. Kent, A One-Pot Total Synthesis of Crambin, *Angew. Chem. Int. Ed.*, 2004, **43**, 2534–2538.

30   J. Dommerholt, O. van Rooijen, A. Borrmann, C. F. Guerra, F. M. Bickelhaupt and F. L. van Delft, Highly accelerated inverse electron-demand cycloaddition of electron-deficient azides with aliphatic cyclooctynes., *Nat. Commun.*, 2014, **5**, 5378.

31   D. L. Boger, Diels-Alder Reactions of Heterocyclic Azadienes: Scope and Applications, *Chem. Rev*, 1986, **86**, 781–793.

32   M. L. Blackman, M. Royzen and J. M. Fox, Tetrazine ligation: fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity., *J. Am. Chem. Soc.*, 2008, **130**, 13518–9.

33   T. S. Elliott, F. M. Townsley, A. Bianco, R. J. Ernst, A. Sachdeva, S. J. Elsässer, L. Davis, K. Lang, R. Pisa, S. Greiss, K. S. Lilley and J. W. Chin, Proteome labeling and protein identification in specific tissues and at specific developmental stages in an animal, *Nat. Biotechnol.*, 2014, **32**, 465–472.

34   K. Lang, L. Davis, J. Torres-Kolbus, C. Chou, A. Deiters and J. W. Chin, Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction., *Nat. Chem.*, 2012, **4**, 298–304.

35   M. T. Taylor, M. L. Blackman, O. Dmitrenko and J. M. Fox, Design and Synthesis of Highly Reactive Dienophiles for the, *J. Am. Chem. Soc.*, 2011, **133**, 9646–9649.

36 W. Chen, D. Wang, C. Dai, D. Hamelberg and B. Wang, Clicking 1,2,4,5-tetrazine and cyclooctynes with tunable reaction rates., *Chem. Commun.* , 2012, **48**, 1736–8.

37 N. K. Devaraj and R. Weissleder, Biomedical applications of tetrazine cycloadditions., *Acc. Chem. Res.*, 2011, **44**, 816–27.

38 G. Clavier and P. Audebert, s-Tetrazines as Building Blocks for New Functional Molecules and Molecular Materials, *Chem. Rev.*, 2010, **110**, 3299–3314.

39 M. R. Karver, R. Weissleder and S. A. Hilderbrand, Synthesis and Evaluation of a Series of 1,2,4,5-Tetrazines for Bioorthogonal Conjugation, *Bioconjug. Chem.*, 2011, **22**, 2263–2270.

40 D. S. Liu, A. Tangpeerachaikul, R. Selvaraj, M. T. Taylor, J. M. Fox and A. Y. Ting, Diels-Alder cycloaddition for fluorophore targeting to specific proteins inside living cells., *J. Am. Chem. Soc.*, 2012, **134**, 792–5.

41 J. A. Johnson, Y. Y. Lu, J. A. Van Deventer and D. A. Tirrell, Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications., *Curr. Opin. Chem. Biol.*, 2010, **14**, 774–80.

42 P. E. Dawson and S. B. Kent, Synthesis of native proteins by chemical ligation., *Annu. Rev. Biochem.*, 2000, **69**, 923–60.

43 M. Rashidian, J. K. Dozier and M. D. Distefano, Enzymatic labeling of proteins: Techniques and approaches, *Bioconjug. Chem.*, 2013, **24**, 1277–1294.

44 L. Davis and J. W. Chin, Designer proteins: applications of genetic code expansion in cell biology., *Nat. Rev. Mol. Cell Biol.*, 2012, **13**, 168–82.

45 K. A. Daggett and T. P. Sakmar, Site-specific in vitro and in vivo incorporation of molecular probes to study G-protein-coupled receptors., *Curr. Opin. Chem. Biol.*, 2011, **15**, 392–8.

46 Q. Wang, A. R. Parrish and L. Wang, Expanding the genetic code for biological studies.,

Chem. Biol., 2009, **16**, 323–36.

47   T. S. Young and P. G. Schultz, Beyond the canonical 20 amino acids: expanding the genetic lexicon., *J. Biol. Chem.*, 2010, **285**, 11039–44.

48   K. Lang and J. W. Chin, Bioorthogonal reactions for labeling proteins., *ACS Chem. Biol.*, 2014, **9**, 16–20.

49   L. Wang and P. G. Schultz, Expanding the genetic code., *Angew. Chem. Int. Ed.* , 2004, **44**, 34–66.

50   M. P. Weiner, G. L. Costa, W. Schoettlin, J. Cline, E. Mathur and J. C. Bauer, Site-directed mutagenesis of double-stranded DNA by the polymerase chain reaction., *Gene*, 1994, **151**, 119–23.

51   J. Braman, C. Papworth and A. Greener, in *In Vitro Mutagenesis Protocols*, Humana Press, New Jersey, 1996, pp. 31–44.

52   E. M. Sletten and C. R. Bertozzi, Bioorthogonal chemistry: fishing for selectivity in a sea of functionality., *Angew. Chem. Int. Ed.* , 2009, **48**, 6974–98.

53   C. Noren, S. Anthony-Cahill, M. Griffith and P. Schultz, A general method for site-specific incorporation of unnatural amino acids into proteins, *Science* , 1989, **244**, 182–188.

54   J. D. Bain, E. S. Diala, C. G. Glabe, T. A. Dix and A. R. Chamberlin, Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide, *J. Am. Chem. Soc.*, 1989, **111**, 8013–8014.

55   J. Ellman, D. Mendel, S. Anthony-Cahill, C. J. Noren and P. G. Schultz, Biosynthetic method for introducing unnatural amino acids site-specifically into proteins, *Methods Enzymol.*, 1991, **202**, 301–336.

56   L. Wang and P. G. Schultz, Expanding the genetic code, *Chem. Commun.*, 2002, **0**, 1–11.

57   J. Xie and P. G. Schultz, An expanding genetic code., *Methods*, 2005, **36**, 227–38.

58   J. W. Chin, T. A. Cropp, J. C. Anderson, M. Mukherji, Z. Zhang and P. G. Schultz, An Expanded Eukaryotic Genetic Code, *Science* , 2003, **301**, 964–967.

59   Ning Wu, Alexander Deiters, T. Ashton Cropp, A. David King and P. G. Schultz*, A Genetically Encoded Photocaged Amino Acid, *J. Am. Chem. Soc.*, 2004, **126**, 14306–14307.

60   H. Neumann, A. L. Slusarczyk and J. W. Chin, De novo generation of mutually orthogonal aminoacyl-tRNA synthetase/tRNA pairs., *J. Am. Chem. Soc.*, 2010, **132**, 2142–4.

61   S. M. Hancock, R. Uprety, A. Deiters and J. W. Chin, Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair., *J. Am. Chem. Soc.*, 2010, **132**, 14819–24.

62   P. R. Chen, D. Groff, J. Guo, W. Ou, S. Cellitti, B. H. Geierstanger and P. G. Schultz, A facile system for encoding unnatural amino acids in mammalian cells., *Angew. Chem. Int. Ed.* , 2009, **48**, 4052–5.

63   S. Greiss and J. W. Chin, Expanding the genetic code of an animal., *J. Am. Chem. Soc.*, 2011, **133**, 14196–9.

64   A. Bianco, F. M. Townsley, S. Greiss, K. Lang and J. W. Chin, Expanding the genetic code of Drosophila melanogaster., *Nat. Chem. Biol.*, 2012, **8**, 748–50.

65   P. E. Dawson, T. W. Muir and S. B. H. Kent, Synthesis of Proteins by Native Chemical Ligation, *Science* , 1994, **266**, 776–779.

66   E. C. B. Johnson and S. B. H. Kent, Insights into the mechanism and catalysis of the native chemical ligation reaction, *J. Am. Chem. Soc.*, 2006, **128**, 6640–6646.

67   T. M. Hackeng, J. H. Griffin and P. E. Dawson, Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology., *Proc. Natl. Acad. Sci.* , 1999, **96**, 10068–73.

68   M. Schnölzer, P. Alewood, A. Jones, D. Alewood and S. B. H. Kent, In Situ Neutralization in Boc-chemistry Solid Phase Peptide Synthesis, *Int. J. Pept. Res. Ther.*, 2007, **13**, 31–44.

69   J. B. Blanco-Canosa and P. E. Dawson, An Efficient Fmoc-SPPS Approach for the Generation of Thioester Peptide Precursors for Use in Native Chemical Ligation, *Angew. Chem. Int. Ed.*, 2008, **47**, 6851–6855.

70   S. K. Mazmanian, G. Liu, H. Ton-That and O. Schneewind, Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall., *Science*, 1999, **285**, 760–3.

71   H. Ton-That, S. K. Mazmanian, L. Alksne and O. Schneewind, Anchoring of surface proteins to the cell wall of Staphylococcus aureus. Cysteine 184 and histidine 120 of sortase form a thiolate-imidazolium ion pair for catalysis., *J. Biol. Chem.*, 2002, **277**, 7447–52.

72   H. Ton-That, S. K. Mazmanian, K. F. Faull and O. Schneewind, Anchoring of surface proteins to the cell wall of Staphylococcus aureus. Sortase catalyzed in vitro transpeptidation reaction using LPXTG peptide and NH(2)-Gly(3) substrates., *J. Biol. Chem.*, 2000, **275**, 9876–81.

73   H. Mao, S. A. Hart, A. Schink and B. A. Pollok, Sortase-Mediated Protein Ligation: A New Method for Protein Engineering, *J. Am. Chem. Soc.*, 2004, **126**, 2670–2671.

74   J. M. Antos, G.-L. Chew, C. P. Guimaraes, N. C. Yoder, G. M. Grotenbreg, M. W.-L. Popp and H. L. Ploegh, Site-Specific N- and C-Terminal Labeling of a Single Polypeptide Using Sortases of Different Specificity, *J. Am. Chem. Soc.*, 2009, **131**, 10800–10801.

75   M. W. Popp, K. Artavanis-Tsakonas and H. L. Ploegh, Substrate filtering by the active site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations., *J. Biol. Chem.*, 2009, **284**, 3593–602.

76   C. P. Guimaraes, M. D. Witte, C. S. Theile, G. Bozkurt, L. Kundrat, A. E. M. Blom and

H. L. Ploegh, Site-specific C-terminal and internal loop labeling of proteins using sortase-mediated reactions., *Nat. Protoc.*, 2013, **8**, 1787–99.

77　C. S. Theile, M. D. Witte, A. E. M. Blom, L. Kundrat, H. L. Ploegh and C. P. Guimaraes, Site-specific N-terminal labeling of proteins using sortase-mediated reactions., *Nat. Protoc.*, 2013, **8**, 1800–7.

78　S. Pritz, Y. Wolf, O. Kraetke, J. Klose, M. Bienert and M. Beyermann, Synthesis of biologically active peptide nucleic acid-peptide conjugates by sortase-mediated ligation, *J. Org. Chem.*, 2007, **72**, 3909–3912.

79　R. L. Policarpo, H. Kang, X. Liao, A. E. Rabideau, M. D. Simon and B. L. Pentelute, Flow-Based Enzymatic Ligation by Sortase A, *Angew. Chem. Int. Ed.*, 2014, **53**, 9203–9208.

80　M. D. Witte, T. Wu, C. P. Guimaraes, C. S. Theile, A. E. M. Blom, J. R. Ingram, Z. Li, L. Kundrat, S. D. Goldberg and H. L. Ploegh, Site-specific protein modification using immobilized sortase in batch and continuous-flow systems, *Nat. Protoc.*, 2015, **10**, 508–516.

81　D. J. Williamson, M. A. Fascione, M. E. Webb and W. B. Turnbull, Efficient N-terminal labeling of proteins by use of sortase., *Angew. Chem. Int. Ed.* , 2012, **51**, 9377–80.

82　K. W. Clancy, J. A. Melvin and D. G. McCafferty, Sortase transpeptidases: Insights into mechanism, substrate specificity, and inhibition, *Biopolymers*, 2010, **94**, 385–396.

83　B. A. Frankel, Y. Tong, M. L. Bentley, M. C. Fitzgerald and D. G. McCafferty, Mutational analysis of active site residues in the Staphylococcus aureus transpeptidase SrtA, *Biochemistry*, 2007, **46**, 7269–7278.

84　I. Chen, B. M. Dorr and D. R. Liu, A general strategy for the evolution of bond-forming enzymes using yeast display, *Proc. Natl. Acad. Sci.*, 2011, **108**, 11399–11404.

85　H. Hirakawa, S. Ishikawa and T. Nagamune, Design of Ca2+ -independent

Staphylococcus aureus sortase A mutants, *Biotechnol. Bioeng.*, 2012, **109**, 2955–2961.

86   M. T. Naik, N. Suree, U. Ilangovan, C. K. Liew, W. Thieu, D. O. Campbell, J. J. Clemens, M. E. Jung and R. T. Clubb, Staphylococcus aureus Sortase A transpeptidase. Calcium promotes sorting signal binding by altering the mobility and structure of an active site loop., *J. Biol. Chem.*, 2006, **281**, 1817–26.

87   L. K. Swee, S. Lourido, G. W. Bell, J. R. Ingram and H. L. Ploegh, One-Step Enzymatic Modification of the Cell Surface Redirects Cellular Cytotoxicity and Parasite Tropism, *ACS Chem. Biol.*, 2015, **10**, 460–465.

88   H. Hirakawa, S. Ishikawa and T. Nagamune, Ca2+ -independent sortase-A exhibits high selective protein ligation activity in the cytoplasm of Escherichia coli, *Biotechnol. J.*, 2015, **10**, 1487–1492.

89   B. M. Dorr, H. O. Ham, C. An, E. L. Chaikof and D. R. Liu, Reprogramming the specificity of sortase enzymes, *Proc. Natl. Acad. Sci.*, 2014, **111**, 13343–13348.

90   K. Piotukh, B. Geltinger, N. Heinrich, F. Gerth, M. Beyermann, C. Freund and D. Schwarzer, Directed Evolution of Sortase A Mutants with Altered Substrate Selectivity Profiles, *J. Am. Chem. Soc.*, 2011, **133**, 17536–17539.

91   L. Schmohl, J. Bierlmeier, F. Gerth, C. Freund and D. Schwarzer, Engineering sortase A by screening a second-generation library using phage display, *J. Pept. Sci.*, 2017, **23**, 631–635.

92   L. Freiburger, M. Sonntag, J. Hennig, J. Li, P. Zou and M. Sattler, Efficient segmental isotope labeling of multi-domain proteins using Sortase A, *J. Biomol. NMR*, 2015, **63**, 1–8.

93   Y. Kobashigawa, H. Kumeta, K. Ogura and F. Inagaki, Attachment of an NMR-invisible solubility enhancement tag using a sortase-mediated protein ligation method, *J. Biomol. NMR*, 2009, **43**, 145–150.

94   Y. Yamamura, H. Hirakawa, S. Yamaguchi and T. Nagamune, Enhancement of sortase A-mediated protein ligation by inducing a β-hairpin structure around the ligation site, *Chem. Commun.*, 2011, **47**, 4742.

95   R. David Row, T. J. Roark, M. C. Philip, L. L. Perkins and J. M. Antos, Enhancing the efficiency of sortase–mediated ligations through nickel–peptide complex formation, *Chem. Commun.*, 2015, **51**, 12548–12551.

96   D. J. Williamson, M. E. Webb and W. B. Turnbull, Depsipeptide substrates for sortase-mediated N-terminal protein ligation, *Nat. Protoc.*, 2014, **9**, 253–62.

97   F. Liu, E. Y. Luo, D. B. Flora and A. R. Mezo, Irreversible sortase a-mediated ligation driven by diketopiperazine formation, *J. Org. Chem.*, 2014, **79**, 487–492.

98   X. Huang, A. Aulabaugh, W. Ding, B. Kapoor, L. Alksne, K. Tabei and G. Ellestad, Kinetic Mechanism of Staphylococcus aureus Sortase SrtA, *Biochemistry*, 2003, **42**, 11307–11315.

99   W. W. Cleland, The kinetics of enzyme-catalyzed reactions with two or more substrates or products. I. Nomenclature and rate equations., *Biochim. Biophys. Acta*, 1963, **67**, 104–37.

100   W. J. Bradshaw, A. H. Davies, C. J. Chambers, A. K. Roberts, C. C. Shone and K. R. Acharya, Molecular features of the sortase enzyme family, *FEBS J.*, 2015, **282**, 2097–2114.

101   B. A. Frankel, R. G. Kruger, D. E. Robinson, N. L. Kelleher and D. G. McCafferty, Staphylococcus aureus Sortase Transpeptidase SrtA: Insight into the Kinetic Mechanism and Evidence for a Reverse Protonation Catalytic Mechanism, *Biochemistry*, 2005, **44**, 11188–11200.

102   G. K. T. Nguyen, S. Wang, Y. Qiu, X. Hemu, Y. Lian and J. P. Tam, Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis, *Nat. Chem. Biol.*, 2014,

10, 732–738.

103   G. K. T. Nguyen, Y. Qiu, Y. Cao, X. Hemu, C.-F. Liu and J. P. Tam, Butelase-mediated cyclization and ligation of peptides and proteins, *Nat. Protoc.*, 2016, **11**, 1977–1988.

104   G. K. T. Nguyen, Y. Cao, W. Wang, C. F. Liu and J. P. Tam, Site-Specific N-Terminal Labeling of Peptides and Proteins using Butelase 1 and Thiodepsipeptide, *Angew. Chem. Int. Ed.*, 2015, **54**, 15694–15698.

105   R. Yang, Y. H. Wong, G. K. T. Nguyen, J. P. Tam, J. Lescar and B. Wu, Engineering a Catalytically Efficient Recombinant Protein Ligase, *J. Am. Chem. Soc.*, 2017, **139**, 5351–5358.

106   D. C. Machin, University of Leeds, 2016.

107   E. A. Merritt, P. Kuhn, S. Sarfaty, J. L. Erbe, R. K. Holmes and W. G. Hol, The 1.25 A resolution refinement of the cholera toxin B-pentamer: evidence of peptide backbone strain at the receptor-binding site., *J. Mol. Biol.*, 1998, **282**, 1043–59.

108   D. C. F. Monteiro, V. Patel, C. P. Bartlett, S. Nozaki, T. D. Grant, J. A. Gowdy, G. S. Thompson, A. P. Kalverda, E. H. Snell, H. Niki, A. R. Pearson and M. E. Webb, The Structure of the PanD/PanZ Protein Complex Reveals Negative Feedback Regulation of Pantothenate Biosynthesis by Coenzyme A, *Chem. Biol.*, 2015, **22**, 492–503.

109   H. Sahoo, Förster resonance energy transfer – A spectroscopic nanoruler: Principle and applications, *J. Photochem. Photobiol. C Photochem. Rev.*, 2011, **12**, 20–30.

110   T. Förster, Zwischenmolekulare Energiewanderung und Fluoreszenz, *Ann. Phys.*, 1948, **437**, 55–75.

111   P. R. Selvin, The renaissance of fluorescence resonance energy transfer, *Nat. Struct. Biol.*, 2000, **7**, 730–734.

112   M. Suzuki, Y. Ito, I. Sakata, T. Sakai, Y. Husimi and K. T. Douglas, Caspase-3 sensitive signaling in vivo in apoptotic HeLa cells by chemically engineered intramolecular

fluorescence resonance energy transfer mutants of green fluorescent protein, *Biochem. Biophys. Res. Commun.*, 2005, **330**, 454–460.

113   O. Coban, D. C. Lamb, E. Zaychikov, H. Heumann and G. U. Nienhaus, Conformational Heterogeneity in RNA Polymerase Observed by Single-Pair FRET Microscopy, *Biophys. J.*, 2006, **90**, 4605–4617.

114   J. Yin, A. J. Lin, P. D. Buckett, M. Wessling-Resnick, D. E. Golan and C. T. Walsh, Single-Cell FRET Imaging of Transferrin Receptor Trafficking Dynamics by Sfp-Catalyzed, Site-Specific Protein Labeling, *Chem. Biol.*, 2005, **12**, 999–1006.

115   D. Kajihara, R. Abe, I. Iijima, C. Komiyama, M. Sisido and T. Hohsaka, FRET analysis of protein conformational change through position-specific incorporation of fluorescent amino acids, *Nat. Methods*, 2006, **3**, 923–929.

116   ThermoScientific, Fluorescence SpectraViewer, https://www.thermofisher.com/uk/en/home/life-science/cell-analysis/labeling-chemistry/fluorescence-spectraviewer.html, (accessed 24 February 2018).

117   C. Deka, B. E. Lehnert, N. M. Lehnert, G. M. Jones, L. A. Sklar and J. A. Steinkamp, Analysis of fluorescence lifetime and quenching of FITC-conjugated antibodies on cells by phase-sensitive flow cytometry, *Cytometry*, 1996, **25**, 271–279.

118   J. Hendrix, C. Flors, P. Dedecker, J. Hofkens and Y. Engelborghs, Dark states in monomeric red fluorescent proteins studied by fluorescence correlation and single molecule spectroscopy., *Biophys. J.*, 2008, **94**, 4103–13.

119   K. M. Dean, J. L. Lubbeck, J. K. Binder, L. R. Schwall, R. Jimenez and A. E. Palmer, Analysis of red-fluorescent proteins provides insight into dark-state conversion and photodegradation., *Biophys. J.*, 2011, **101**, 961–9.

120   M. Royzen, G. Yap and J. Fox, A photochemical synthesis of functionalized trans-cyclooctenes driven by metal complexation, *J. Am. Chem. Soc.*, 2008, **1**, 3760–3761.

121   K. Lang, L. Davis, S. Wallace, M. Mahesh, D. J. Cox, M. L. Blackman, J. M. Fox and J. W. Chin, Genetic Encoding of bicyclononynes and trans-cyclooctenes for site-specific protein labeling in vitro and in live mammalian cells via rapid fluorogenic Diels-Alder reactions., *J. Am. Chem. Soc.*, 2012, **134**, 10317–20.

122   Sigma Aldrich, Rhodium (II) acetate catalyst, https://www.sigmaaldrich.com/catalog/product/aldrich/482285?lang=en&region=GB, (accessed 21 February 2018).

123   G. Binnig, C. F. Quate and C. Gerber, Atomic Force Microscope, *Phys. Rev. Lett.*, 1986, **56**, 930.

124   T. Ando, High-speed atomic force microscopy coming of age, *Nanotechnology*, 2012, **23**, 1–27.

125   C. Joo, H. Balci, Y. Ishitsuka, C. Buranachai and T. Ha, Advances in Single-Molecule Fluorescence Methods for Molecular Biology, *Annu. Rev. Biochem.*, 2008, **77**, 51–76.

126   E. Betzig and J. K. Trautman, Near-Field Optics: Microscopy, Spectroscopy, and Surface Modification Beyond the Diffraction Limit, *Science* , 1992, **257**, 189–195.

127   T. Funatsu, Y. Harada, M. Tokunaga, K. Saito and T. Yanagida, Imaging of single fluorescent molecules and individual ATP turnovers by single myosin molecules in aqueous solution, *Nature*, 1995, **374**, 555–559.

128   M. B. Viani, L. I. Pietrasanta, J. B. Thompson, A. Chand, I. C. Gebeshuber, J. H. Kindt, M. Richter, H. G. Hansma and P. K. Hansma, Probing protein–protein interactions in real time, *Nat. Struct. Biol.*, 2000, **7**, 644–647.

129   T. Ando, N. Kodera, Y. Naito, T. Kinoshita, K. Furuta and Y. Y. Toyoshima, A High-speed Atomic Force Microscope for Studying Biological Macromolecules in Action, *ChemPhysChem*, 2003, **4**, 1196–1202.

130   T. Ando, T. Uchihashi and N. Kodera, High-Speed AFM and Applications to

Biomolecular Systems, *Annu. Rev. Biophys.*, 2013, **42**, 393–414.

131   Q. Zhong, D. Inniss, K. Kjoller and V. B. Elings, Fractured polymer/silica fiber surface studied by tapping mode atomic force microscopy, *Surf. Sci. Lett.*, 1993, **290**, L688–L692.

132   N. Kodera, D. Yamamoto, R. Ishikawa and T. Ando, Video imaging of walking myosin v by high-speed atomic force microscopy, *Nature*, 2010, **468**, 72–76.

133   M. Shibata, T. Uchihashi, H. Yamashita, H. Kandori and T. Ando, Structural changes in bacteriorhodopsin in response to alternate illumination observed by high-speed atomic force microscopy, *Angew. Chem. Int. Ed.*, 2011, **50**, 4410–4413.

134   S. Furuike, M. D. Hossain, Y. Maki, K. Adachi, T. Suzuki, A. Kohori, H. Itoh, M. Yoshida and K. Kinosita, Axle-less F1-ATPase rotates in the correct direction, *Science* , 2008, **319**, 955–958.

135   P. W. K. Rothemund, Folding DNA to create nanoscale shapes and patterns, *Nature*, 2006, **440**, 297–302.

136   H. G. Hansma and D. E. Laney, DNA binding to mica correlates with cationic radius: assay by atomic force microscopy., *Biophys. J.*, 1996, **70**, 1933–9.

137   G. L. Gaines and D. Tabor, Surface Adhesion and Elastic Properties of Mica, *Nature*, 1956, **178**, 1304–1305.

138   C. Seah, D. C. Alexander, L. Louie, A. Simor, D. E. Low, J. Longtin and R. G. Melano, MupB, a new high-level mupirocin resistance mechanism in Staphylococcus aureus., *Antimicrob. Agents Chemother.*, 2012, **56**, 1916–20.

139   S. Skou, R. E. Gillilan and N. Ando, Synchrotron-based small-angle X-ray scattering of proteins in solution., *Nat. Protoc.*, 2014, **9**, 1727–39.

140   P. V. Konarev, V. V. Volkov, A. V. Sokolova, M. H. J. Koch and D. I. Svergun, PRIMUS : a Windows PC-based system for small-angle scattering data analysis, *J. Appl. Crystallogr.*, 2003, **36**, 1277–1282.

141   P. V. Konarev, M. V. Petoukhov, V. V. Volkov and D. I. Svergun, ATSAS 2.1, a program package for small-angle scattering data analysis, *J. Appl. Crystallogr.*, 2006, **39**, 277–286.

142   M. V. Petoukhov and D. I. Svergun, Global rigid body modeling of macromolecular complexes against small-angle scattering data., *Biophys. J.*, 2005, **89**, 1237–50.

143   D. I. Svergun, Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing., *Biophys. J.*, 1999, **76**, 2879–86.

144   D. I. Svergun and H. B. Stuhrmann, New developments in direct shape determination from small-angle scattering. 1. Theory and model calculations, *Acta Crystallogr. Sect. A Found. Crystallogr.*, 1991, **47**, 736–744.

145   D. I. Svergun, M. V. Petoukhov and M. H. Koch, Determination of domain structure of proteins from X-ray solution scattering., *Biophys. J.*, 2001, **80**, 2946–53.

146   D. I. Svergun, V. V. Volkov, M. B. Kozin and H. B. Stuhrmann, New Developments in Direct Shape Determination from Small-Angle Scattering. 2. Uniqueness, *Acta Crystallogr. Sect. A Found. Crystallogr.*, 1996, **52**, 419–426.

147   D. I. Svergun and M. H. J. Koch, Small-angle scattering studies of biological macromolecules in solution, *Reports Prog. Phys.*, 2003, **66**, 1735–1782.

148   K.-H. Tang, V. S. Urban, J. Wen, Y. Xin and R. E. Blankenship, SANS investigation of the photosynthetic machinery of Chloroflexus aurantiacus., *Biophys. J.*, 2010, **99**, 2398–407.

149   D. I. Svergun, M. H. J. Koch, P. A. Timmins and R. P. May, *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules*, OUP Oxford, 2013.

150   B. Jacrot, The study of biological structures by neutron scattering from solution, *Reports Prog. Phys.*, 1976, **39**, 911–953.

151   S. Maric, N. Skar-Gislinge, S. Midtgaard, M. B. Thygesen, J. Schiller, H. Frielinghaus, M. Moulin, M. Haertlein, V. T. Forsyth, T. G. Pomorski and L. Arleth, Stealth carriers for

low-resolution structure determination of membrane proteins in solution., *Acta Crystallogr. D. Biol. Crystallogr.*, 2014, **70**, 317–28.

152   Stratagene, *QuikChange ® II Site-Directed Mutagenesis Kit*, 2007.

153   Roche, *Pwo DNA Polymerase*, 2006.

154   M. Takagi, M. Nishioka, H. Kakihara, M. Kitabayashi, H. Inoue, B. Kawakami, M. Oka and T. Imanaka, Characterization of DNA polymerase from *Pyrococcus* sp. strain KOD1 and its application to PCR, *Appl. Environ. Microbiol.*, 1997, **63**, 4504–4510.

155   Merck, *KOD Hot Start DNA Polymerase*, 2011.

156   G. Clifton, S. R. Bryant and C. G. Skinner, N1-(substituted) pantothenamides, antimetabolites of pantothenic acid, *Arch. Biochem. Biophys.*, 1970, **137**, 523–528.

157   E. Strauss and T. P. Begley, The antibiotic activity of N-pentylpantothenamide results from its conversion to ethyldethia-coenzyme A, a coenzyme A antimetabolite, *J. Biol. Chem.*, 2002, **277**, 48205–48209.

158   World Health Organisation, *Antimicrobial resistance: global report on surveillance 2014*, 2014.

159   Y. Zhang, M. W. Frank, K. G. Virga, R. E. Lee, C. O. Rock and S. Jackowski, Acyl carrier protein is a cellular target for the antibacterial action of the pantothenamide class of pantothenate antimetabolites., *J. Biol. Chem.*, 2004, **279**, 50969–75.

160   C. Spry, K. Kirk and K. J. Saliba, Coenzyme A biosynthesis: An antimicrobial drug target, *FEMS Microbiol. Rev.*, 2008, **32**, 56–106.

161   J. Thomas and J. E. Cronan, Antibacterial Activity of N-Pentylpantothenamide Is Due to Inhibition of Coenzyme A Synthesis, *Antimicrob. Agents Chemother.*, 2010, **54**, 1374–1377.

162   J. M. Williamson and G. M. Brown, Purification and properties of L-aspartate-a-

decarboxylase, an enzyme that catalyzes the formation of B-alanine in Escherichia coli., *J. Biol. Chem.*, 1979, **254**, 8074–8082.

163 D. C. F. Monteiro, M. D. Rugen, D. Shepherd, S. Nozaki, H. Niki and M. E. Webb, Formation of a heterooctameric complex between aspartate α-decarboxylase and its cognate activating factor, PanZ, is CoA-dependent, *Biochem. Biophys. Res. Commun.*, 2012, **426**, 350–355.

164 Z. L. P. Arnott, S. Nozaki, D. C. F. Monteiro, H. E. Morgan, A. R. Pearson, H. Niki and M. E. Webb, The Mechanism of Regulation of Pantothenate Biosynthesis by the PanD–PanZ·AcCoA Complex Reveals an Additional Mode of Action for the Antimetabolite N-Pentyl Pantothenamide (N5-Pan), *Biochemistry*, 2017, **56**, 4931–4939.

165 D. C. F. Monteiro, University of Leeds, 2015.

166 M. Wang, M. Weiss, M. Simonovic, G. Haertinger, S. P. Schrimpf, M. O. Hengartner and C. von Mering, PaxDb, a database of protein abundance averages across all three domains of life., *Mol. Cell. Proteomics*, 2012, **11**, 492–500.

167 B. E. Corkey, M. Brandt, R. J. Williams and J. R. Williamson, Assay of short-chain acyl coenzyme a intermediates in tissue extracts by high-pressure liquid chromatography, *Anal. Biochem.*, 1981, **118**, 30–41.

168 R. M. C. Dawson, D. C. Elliott, W. H. Elliott and K. M. Jones, *Data for Biochemical Research*, 1968.

169 D. A. Hiller, J. M. Fogg, A. M. Martin, J. M. Beechem, N. O. Reich and J. J. Perona, Simultaneous DNA Binding and Bending by EcoRV Endonuclease Observed by Real-Time Fluorescence, *Biochemistry*, 2003, **42**, 14375–85.

170 E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel and A. Bairoch, Protein Identification and Analysis Tools on the ExPASy Server, *Proteomics Protoc. Handb.*, 2005, 571–607.

171   T. A. Hall, *Nucleic Acids Symp. Ser.*, 1999, 41, 95–98.

172   Integrated DNA Technologies, OligoAnalyzer 3.1, http://www.idtdna.com/calc/analyzer, (accessed 27 March 2018).

173   ThermoScientific, *Calculate dye : protein (F / P) molar ratios*, 2011.

174   B. J. H. Kuipers and H. Gruppen, Prediction of molar extinction coefficients of proteins and peptides using UV absorption of the constituent amino acids at 214 nm to enable quantitative reverse phase high-performance liquid chromatography-mass spectrometry analysis, *J. Agric. Food Chem.*, 2007, **55**, 5445–5451.

175   S. Keller, C. Vargas, H. Zhao, G. Piszczek, C. A. Brautigam and P. Schuck, High-precision isothermal titration calorimetry with automated peak-shape analysis., *Anal. Chem.*, 2012, **84**, 5066–73.

176   H.-J. Jeong, G. C. Abhiraman, C. M. Story, J. R. Ingram and S. K. Dougan, Generation of Ca2+-independent sortase A mutants with enhanced activity for protein and cell surface labeling, *PLoS One*, 2017, **12**, e0189068.

177   M. E. Webb, B. A. Yorke, T. Kershaw, S. Lovelock, C. M. C. Lobley, M. L. Kilkenny, A. G. Smith, T. L. Blundell, A. R. Pearson, C. Abell and IUCr, Threonine 57 is required for the post-translational activation of Escherichia coli aspartate α-decarboxylase, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2014, **70**, 1166–1172.

178   F. Schmitzberger, M. L. Kilkenny, C. M. C. Lobley, M. E. Webb, M. Vinkovic, D. Matak-Vinkovic, M. Witty, D. Y. Chirgadze, A. G. Smith, C. Abell and T. L. Blundell, Structural constraints on protein self-processing in L-aspartate-a-decarboxylase, *EMBO J.*, 2003, **22**, 6193–6204.

179   S. Nozaki, M. E. Webb and H. Niki, An activator for pyruvoyl-dependent l -aspartate α-decarboxylase is conserved in a small group of the γ-proteobacteria including Escherichia coli, *Microbiologyopen*, 2012, **1**, 298–310.

180   M. Kitagawa, T. Ara, M. Arifuzzaman, T. Ioka-Nakamichi, E. Inamoto, H. Toyonaga and H. Mori, Complete set of ORF clones of Escherichia coli ASKA library (A Complete Set of E. coli K-12 ORF Archive): Unique Resources for Biological Research, *DNA Res.*, 2006, **12**, 291–299.