Sortase-mediated modification and intracellular delivery of cholera toxin analogues

Daniel John Williamson

Submitted with accordance with requirements for the degree of

Doctor of Philosophy

The University of Leeds

School of Chemistry

August 2014

The candidate confirms that the work submitted is his 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 the work of others

1. Chapter 2: Efficient *N*-terminal labelling of proteins using Sortase, includes content from the following publications:

"Efficient N-Terminal Labeling of Proteins by Use of Sortase" Angewandte Chemie International Edition, 2012, volume 51, pages 9377-9380, D. J. Williamson, M. A. Fascione, M. E. Webb and W. B. Turnbull. The manuscript was composed by M. E. Webb and W. B. Turnbull. hMBL was initially expression by M. A. Fascione. All other experimental work was conducted by the candidate.

"Depsipeptide substrates for sortase-mediated N-terminal protein ligation" D. J. Williamson, M. E. Webb, W. B. Turnbull, Nature protocols, 2014, volume 9, pages 253-262. The manuscript was composed by all authors and the experimental work was conducted by the candidate

Acknowledgments

I would like to express my sincere gratitude to all my supervisors Mike Webb, Andrew Macdonald and Bruce Turnbull for their continued support throughout the project. I am especially grateful to Bruce for giving me the opportunity to study for a PhD and for his seemingly unlimited patience over the past 5 years. He has always made the time to discuss any problems I have encountered and has always been encouraging no matter what the results.

It is a pleasure for me to thank the past and present members of lab 1.49 who have made working in Leeds an immensely enjoyable experience. These include Ed, Heather, Jeff, Pintu, James, Tom B, Tom Mc, Martin, Chadamas, Kristian, Kat, Phil, Diana, Darren, Ivona, Matt and Jamie.

I am extremely appreciative to James Ross for providing me with some of the essential proteins required for the project, which allowed me to avoid all forms of molecular biology for the last few years – I cannot thank him enough. I also thank Chris Wasson for teaching me some of his expertise in cell culture.

Special thanks has to go to Martin Fascione who helped me a lot during my masters project and the early days of my PhD when I did not know my arse from my elbow. His enthusiasm for science has been inspirational.

I would like to warmly acknowledge all the technical staff in both the chemistry and biology departments. Martin Huscroft was especially forthcoming in providing assistance with the preparation of HPLC assays that were essential to my project.

I would not have been able to study at University or finish my PhD without the help and continuous support of my family. They have always made every effort to provide me with everything I could possibly need throughout University and my entire life. I cannot express how grateful I am to them.

This thesis would not have been written without the patience and support of Silvia Rodriguez Marin. She has had to endure periods of me being completely stressed out, but she has put up with it and given me much needed perspective when things were tough. I am truly thankful to her.

Last, but by no means least, I would like to thank the “lads” Alun Myden and James Murry for the many trips to the pub.

Abstract

Macromolecule therapeutics, such as oligonucleotides, are playing an increasingly important role in the fight against diseases. Unfortunately, the majority of macromolecules are unable to cross the cell membrane barrier, which severely reduces the number of potential intracellular targets this class of therapeutic can act on. The AB_5 complexes, cholera toxin and heat-labile enterotoxin have evolved the ability to enter mammalian cells effortlessly by using a retrograde transport mechanism. It has been previously reported that non-toxic analogues of these proteins can be used to transport macromolecular cargo into cells. However, the majority of these strategies utilise recombinant technologies and as a result are limited to protein based cargo.

Sortase A (SrtA) is a class of transpeptidase enzyme that attaches virulence factors to the cell wall of gram Gram-positive bacteria. The enzyme recognises proteins carrying a *C*-terminal LPXTG motif and ligates them to peptidoglycans at the *N*-terminus of an oligoglycine sequence. Since its discovery, SrtA has been viewed as a potential new tool for protein modification. However, the enzyme-mediated ligation reaction is relatively inefficient *in vitro*.

In this thesis, I will describe the development of depsipeptide substrates for SrtA that allow the quantitative *N*-terminal labelling of proteins/peptides. In addition, a simple and robust strategy to chemically-modified AB_5 complexes that are able to enter mammalian cells will be reported. Finally, a novel approach to generate oligonucleotide-conjugated AB_5 complexes using a combination of enzymatic and bioorthogonal reactions will also be outlined.

Abbreviations

ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
APS	Ammonium persulfate
ATP	Adenosine triphosphate
BAP	Bacterial alkaline phosphatase
BLA	β-lactamase
BME	β-mercaptoethanol
Boc	*tert*-Butoxylcarbonyl
BSA	Bovine serum albumin
CBD	Chitin-binding domain
CPP	Cell penetrating peptide
CT	Cholera toxin
CTA	Cholera toxin A-subunit
CTB	Cholera toxin B-subunit
CuAAC	Cu-promoted azide-alkyne cycloaddition
DCM	Dichloromethane
DIPEA	N,N-Diisopropylethylamine
DMEM	Dulbecco's modified Eagle media
DMF	N,N-Dimethylformamide
DNA	Deoxyribonucleic acid

DTT	Dithiothreitol
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
ES	Electrospray
ESMS	Electron spray mass spectrometry
FITC	Fluorescein isothiocyanate
Fmoc	Fluorenylmethyloxycarbonyl
GFP	Green fluorescent protein
HCTU	2-(6-Chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium haxafluorophosphate
HEK	Human embryonic kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His-tag	Polyhistidine tag
HPLC	High-performance liquid chromatography
HRMS	High resolution mass spectrometry
HRP	Horseradish peroxidase
hMBL	Human mannose binding lectin
IR	Infrared
LB	Lysogeny broth
LCMS	Liquid chromatography mass spectrometry
LNA	Locked nucleic acid
LT	Heat-labile enterotoxin

LTA	Heat-labile enterotoxin A-subunit
LTB	Heat-labile enterotoxin B-subunit
MBP	Maltose binding protein
MeCN	Acetonitrile
MPS	Mononuclear phagocytic system
NCL	Native chemical ligation
NMR	Nuclear magnetic resonance spectroscopy
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDI	Protein disulfide isomerise
PEG	Polyethylene glycol
PMO	Phosphorodiamidate morpholino oligomer
PNA	Peptide nucleic acids
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNase H	Ribonuclease H
r.t.	Room temperature
SDS	Sodium dodecyl sulfate
siRNA	Small interfering RNA
SPAAC	Strained-promoted alkyne-azide cycloaddition
SPPS	Solid phase peptide synthesis
SrtA	Sortase A

TEV	Tobacco Etch Virus
TBAI	Tetrabutylammonium iodide
TBAF	Tetrabutylammonium fluride
TBDMS	*tert*-Butyldimethylsiyl
TEMED	Tetramethylethylenediamine
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TIS	Triisopropylsilane
Tris	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet

Contents

Acknowledgments ..... **iii**

Abstract ..... **iv**

Abbreviations ..... **v**

Contents ..... **ix**

Chapter 1: Introduction part I **1**

1.1 Biological therapeutics 2

1.1.1 Oligonucleotides therapeutics 2

1.2 Bacterial toxins 11

1.2.1 AB_5 enterotoxins 11

1.2.2 Formation of novel AB_5 complexes 14

1.2.3 Formation of Novel LTB/CTB Proteins 16

1.3 Protein modification 19

1.3.1 Native chemical ligation 19

1.3.2 Sortase-mediated chemical ligations 21

Introduction part II **29**

1.4 Project outline 30

Chapter 2: Efficient *N*-terminal labelling of proteins using Sortase **32**

2.1 Introduction 33

2.2 Investigation into the efficiency of the SrtA-mediated ligation reaction 33

2.2.1 SrtA-mediated ligation with model peptide substrates 33

2.2.2 HPLC time-course analysis of the SrtA-mediated ligation reaction with peptide substrates 36

2.3 Optimization of the SrtA mediated ligation reaction 38

2.3.1 Initial Synthetic route to depsipeptide substrates 39

2.3.2 Improved synthetic route to depsipeptide substrates 41

2.3.3 Synthesis of methyl ester substrates for SrtA 42

2.3.4 Comparison of SrtA reactivity with peptide, depsipeptide and methyl ester substrates 43

2.4 Protein modification using optimised SrtA-mediated ligations 45

2.4.1 Human mannose binding lectin 46

2.4.2 Mouse pumilio-2 puf domain 52

2.4.3 Myoglobin 53

2.4.4 Investigation into the inefficient and failed SrtA-mediated labelling of protein substrates 55

2.5 Conclusion 58

Chapter 3: Modified AB_5 proteins for cellular delivery 59

3.1 Introduction 60

3.2 AB_5 complex modification strategy 61

3.2.1 MBP-AB_5 expression and isolation 61

3.2.2 Preparation of the AB_5 complex for SrtA-mediated ligation 62

3.3 Sortase-mediated modification of the AB_5 protein 66

3.3.1 Dansyl-depsipeptide ligation 66

3.3.2 Fluorescein-depsipeptide ligation 67

3.4 Transportation of AB_5 complex into mammalian cells 70

3.4.1 Studies with various mammalian cell lines 70

3.4.2 Organelle-specific immunostaining 72

3.5 Conclusion 77

Chapter 4: Bioorthogonal depsipeptide substrates for Sortase 78

4.1 Introduction 79

4.2 Bioorthogonal linker selection 79

4.2.1 Hydrazone linker 79

4.2.2 Strained cyclooctyne linker 83

4.3 Strain-promoted cycloadditions with the AB_5 complex 87

4.3.1 Oligonucleotide conjugation to AB_5 protein 89

4.4 Conclusion 93

Chapter 5: Conclusions and future work 94

5.1 Summary 95

5.2 Future work 95

5.2.1 Development and application of RNA-conjugated AB_5 complexes 95

5.2.2 Controlled release of cargo from AB_5 protein conjugates 96

5.2.3 Anticipated problems with bacterial toxin based delivery systems. 98

Chapter 6: Chemical and biological experimental 100

6.1 Chemical experimental 101

6.1.1 General methods 101

6.1.2 Small molecule synthetic data 101

6.1.3 Solid phase peptide synthesis 106

6.1.4 .Spectral data for peptides 108

6.2 Biological material and methods 126

6.2.1 Instrumentation and materials. 126

6.2.2 Buffer solutions and media for biochemistry 127

6.2.3 Molecular biology 129

6.2.4 Protein overexpression, purification and analysis 131

6.2.5 Sortase-mediated chemical ligations 136

6.2.6 Bioconjugation reactions 139

6.2.7 Other enzyme reactions 140

6.2.8 Mammalian cell culture 140

Chapter 7: Appendix 143

7.1 DNA plasmid sequences 144

7.1.1 pET28a I 144

7.1.2 pMAL-c5x I 147

7.1.3 pSAB2.1 151

7.2 Protein sequences 153

Chapter 8: References 157

Chapter 1: Introduction part I

1.1 Biological therapeutics

In 2012, seven of the top ten selling drugs were biological-based therapeutics with the top three being antibodies.[1] This represents a significant paradigm shift in the way diseases are being treated as small molecules were always at the centre of drugs research. Traditionally, small molecule drugs were used to either inhibit or activate proteins, such as enzymes, that are associated with a particular disease. However, this approach has its limitations; it can be difficult to design a drug that specifically targets the protein of interest without producing unwanted side effects. Furthermore, some protein targets are simply "undruggable" using small molecules.[178]

Biopharmaceuticals are highly specific and have a low systemic toxicity compared to their small molecule counterparts. Previously incurable or life debilitating illnesses, such as paroxysmal nocturnal haemoglobinuria,[2] are now being treated with new biological-based therapeutics. Although antibody-based drugs are currently leading worldwide drugs sales, there is also a significant amount of research being invested into the use of therapeutic oligonucleotides for genetic-based therapies.[3] However, there are still significant challenges associated with each different class of biopharmaceutical. These can include high production costs, poor cellular uptake and reduced *in vivo* stability.[2-4]

1.1.1 Oligonucleotides therapeutics

Genetic disorders can cause the production of defective proteins that alter normal cellular function and trigger chronic disease. The RNA and pre-RNA sequences that govern faulty protein expression can be directly targeted by oligonucleotide therapeutics and are either removed or repaired to restore normal bioactivity. Diseases, such as Duchenne muscular dystrophy, that were previously untreatable with conventional medicines are now being fought with a new generation of oligonucleotide therapeutics.[5,6] In principle, this strategy has the potential to treat any genetic disorder, but there are still serious challenges to overcome before oligonucleotides can be used ubiquitously in disease treatment.[3,7]

1.1.1.1 Synthetic oligonucleotides

Natural oligonucleotides are notoriously unstable *in vivo* due to the abundance of enzymes that will hydrolyse them. Consequentially, oligonucleotides therapeutics have to be chemically altered before they can be used as viable drugs (Figure 1.1).[3,7-9] The most common modification is the incorporation of phosphorothioate linkages **1.3** into the backbone of an oligonucleotide. The phosphorothioate linkage reduces the rate of hydrolysis by nuclease enzymes and decreases

the efficiency of renal clearance; therefore, their stability *in vivo* is dramatically increased. RNA based therapeutics can be additionally modified to increase their enzymatic stability. Generally, the 2'-hydroxyl group on the ribose sugar is either alkylated **1.4**, substituted with a fluoride **1.5** or cyclised to form a locked nucleic acid (LNA) **1.6**.[3,7-9] The 2'-modifications also have the additional benefit of improving mRNA binding by forcing the oligonucleotides to adopt a more favourable conformation for base-base pairing. In their structure, oligonucleotide drugs commonly contain a mixture of both 2'-modifications and phosphorothioate linkages to ensure that maximum stability and potency are achieved.

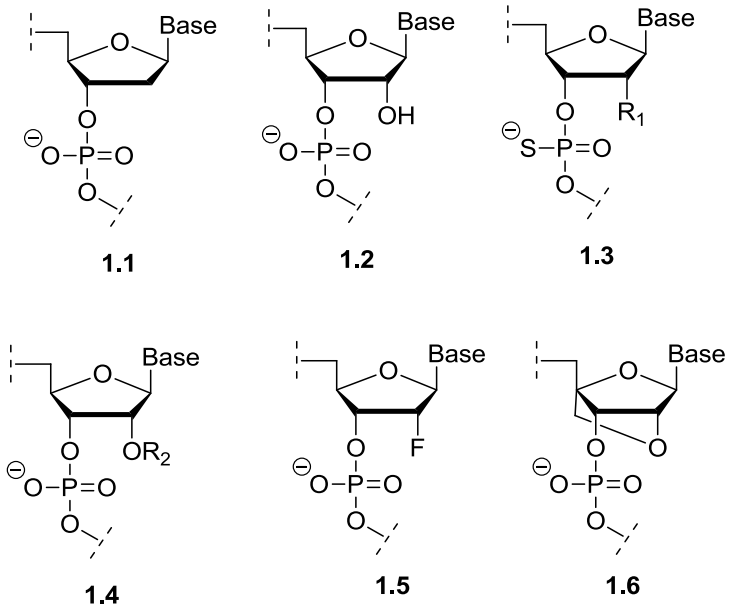


Figure 1.1: Structure of nucleic acids with different chemical modifications. Deoxyribonucleic acid (DNA) **1.1**; ribonucleic acid (RNA) **1.2**; nucleic acid with a phosphorothioate backbone **1.3**, R_1= H, OH; 2'O-alkylated ribonucleic acid **1.4**, R_2 = methyl, methoxyethyl; 2'-fluorinated deoxyribonucleic acid **1.5**; locked nucleic acid (LNA) **1.6** with the 2'O and 4' positions cyclised.
.

Interestingly, it possible to replace the entire backbone structure of oligonucleotides with a range of analogues molecules and still maintain effective base pairing.[3,7] Prime examples are phosphorodiamidate morpholino oligomer (PMO) **1.7** and peptide nucleic acids (PNA) **1.8** (Figure 1.2) .[3,7]

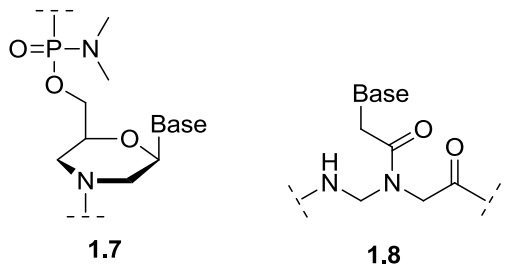


Figure 1.2: Structures of phosphorodiamidate morpholino oligomers (PMO) **1.7** and peptide nucleic acid (PNA) **1.8.**

1.1.1.2 Biological mode of action

There are several classes of oligonucleotide therapeutics that act on different biological targets to either inhibit or fix faulty protein production; four of the most widely studied include:

Antisense oligonucleotide

Antisense oligonucleotides are short, single stranded pieces of RNA that silence gene expression by removing mRNA before it can be translated. The antisense strand binds to target mRNA sequences forming a duplex structure that is recognised and subsequently degraded by ribonuclease H (RNase H) enzymes before protein expression can occur.[3,7,8]

Modern antisense drugs are generally 20mers in length with a central ~10mer binding site.[3,7] The entire RNA backbone is usually modified with phosphorothioate linkages, whereas 2'-modifications are restricted to flanking sequences on the peripheries of the oligonucleotide to ensure RNase H binding is not inhibited.[3,7,8]

Interfering oligonucleotides

Small interfering RNAs (siRNA) are short double stranded pieces of RNA that associate with the RNA-induced silencing complex (RISC) and catalytically silence gene expression.[3,7,8] The RISC binds to the siRNA duplex causing it to dissociate into two separate oligonucleotide sequences; the antisense strand remains bound to the RISC while the sense strand is discarded. The complexed antisense sequence associates with complementary mRNA, which is then subsequently cleaved by the Argonaut-2 component of the RISC complex. This mechanism is repeated continuously allowing many copies of mRNA to be destroyed by a single antisense-RISC enzyme

Typical siRNA therapeutics are ~20 base pairs in length and carry only minimal chemical modifications to guarantee successful RISC binding.[3,7] The guide strand that binds directly to the RISC can only be modified with phosphorothioate linkages and 2'-methyl alkylations at one or two nucleotide positions.[3] However, the complementary sense strand that is discarded during complexation can tolerate a greater number of chemical modifications.

Steric-blocking oligonucleotides

Gene silencing can also be achieved without destroying mRNA if it is neutralised with a complementary binding partner. Heavily modified oligonucleotides, such as PMOs **1.7** (Figure 1.2), bind to mRNA forming a duplex structure that is stable to degradation by RNase H due to

the abundance of chemical modification.[3,7] However, the complexed synthetic oligonucleotide also sterically blocks the ribosome from binding and translating the mRNA into protein.

Splice-switching oligonucleotides

Remarkably, synthetic oligonucleotides are now being used in therapies that elegantly restore the function of defective proteins by reengineering the splicing mechanisms that produce the mRNA.[3,7] The therapeutic oligonucleotides bind to the faulty splice sites or even mutated pre-RNA sequences to form a stable duplex that is unable to be incorporated into the mature mRNA chain, thus restoring full or partial protein function.

1.1.1.3 Delivery vehicles

In most cases, therapeutic oligonucleotides are unable to cross the cell membrane barrier to reach their target as they are large and highly charged species. In addition, they show poor bioavailability *in vivo* and as a result, their efficacy is limited.[10] The use of a delivery vehicle to transport the oligonucleotides directly to a target site is one strategy that could overcome these challenges. In the past, therapeutic nucleic acids were delivered using viral-based systems, but there are now several different classes of non-viral based delivery vehicles; these include liposomes, cell penetrating peptides (CPP) and polymeric carriers

Viral vectors

Viruses have evolved specific mechanisms that enable them to transfect and deliver their genetic material inside mammalian cells. Different types of viruses can gain entry into cells by either receptor-mediated endocytosis, macropinocytosis or phagocytosis.[11,12] However, clathrin-mediated endocytosis is the most common transfection mechanism. This process is triggered when viral particles and other substances bind to receptors on the surface of the cell in areas that are rich in clathrin.[11] The protein is then internalised by the cell in a clathrin-coated vesicle before rapid endosomal formation takes place. Typically, substances internalised by clathrin-mediated endocytosis are trafficked through early endosomes, late endosomes and finally lysosomes. However, viruses avoid the usual intracellular transportation mechanisms and escape into the cytosol.[11,12] This is done in a number of different ways, but in general, it involves the disruption of an endosomal vesicle. For example, adenoviruses use a proton-pump mechanism to reduce the pH of the vesicle and trigger endosomal escape.[11] Whereas, poliovirus is thought to undergo a structural change in the slightly acidic environment of the vesicle to expose a hydrophobic protein domain that disrupts the membrane and aids viral release into the cytosol.[12] The natural endosomal escape mechanisms of viruses make them an attractive vehicle to deliver therapeutic nucleic acids into cells.

The genomes of viral vectors are modified to remove any pathogenic genes and to introduce therapeutic nucleotide sequences.[11,13-15] This delivery approach protects the nucleic acid cargo inside a viral capsid until it is released into the cytosolic interior of a cell. There are currently five commonly used viral vectors based on the adenovirus, adeno-associated virus, herpes simplex virus, lentivirus and retrovirus. Each system has its own associated limitations, such as production difficulties, non-specific cell targeting, insertional mutagenesis and limited insert size.[11,13] However, the major obstacle is safety. There is a possibility that the host will undergo a harsh immunogenic response to the viral vector. Furthermore, viral vectors based on retroviruses have been implicated in the onset of cancer in patients that have taken part in clinical trials.[11,16,17]

Liposomes

Liposomes are biocompatible artificial vesicles composed of a lipid bilayer structure surrounding an aqueous core (Figure 1.3) [18,19] Once absorbed onto the surface of a cell, liposomes are internalised by endocytosis or membrane fusion[20-22] Drug molecules can be encapsulated inside the central core of the structure protecting them from degradation by extra-liposomal reactions. Sequestrated therapeutics can exhibit a significantly altered pharmacokinetic profile to the free drug.[18,19] For example, the anti cancer drug anthracyclines showed prolonged circulation, delayed adsorption and less bio-distribution when injected into mice with colon cancer.[23] In addition, ligands (e.g., carbohydrates,[24] antibodies,[25] etc) can also be incorporated onto the surface of liposomes enabling them to target specific cells and therefore increase the specificity of the treatment.[19,26] However, the successful targeting of receptors is heavily dependent on the vesicle being in circulation for long periods of time.

Liposomal delivery vehicles are systemically removed by the mononuclear phagocytic system (MPS) reducing the effectiveness of the encapsulated therapeutic; a major limitation of the approach.[19,26,27] Smaller unilamellar liposomes with a 50-250 nm diameter are commonly used as drug carriers as they are not readily recognised by the MPS and therefore have an increased circulation time.[18,28] The surface charge of the liposome can be tailored to increase the incorporation of certain substances. However, a non-neutral net charge can significantly increase the rate of MPS clearance,[29] so the final surface composition is usually a compromise between successful encapsulation and optimal liposomal circulation time.[18,19]

Oligonucleotides are only efficiently encapsulated inside or associated to the surface of cationic liposomes.[18,28] Unfortunately, the positively charged vesicles are inherently toxic and are readily cleared by the immune system. However, formulation with hydrophilic polymers, such as poly(ethyleneglycol) (PEG),[30] can help mask them from the MPS, but cytotoxicity is still a considerable problem.[18,22,28]

Liposomes have the potential to be used as universal delivery vehicles for both small and macromolecules, but issues regarding MPS clearance and toxicity of cationic vesicles need to be addressed.

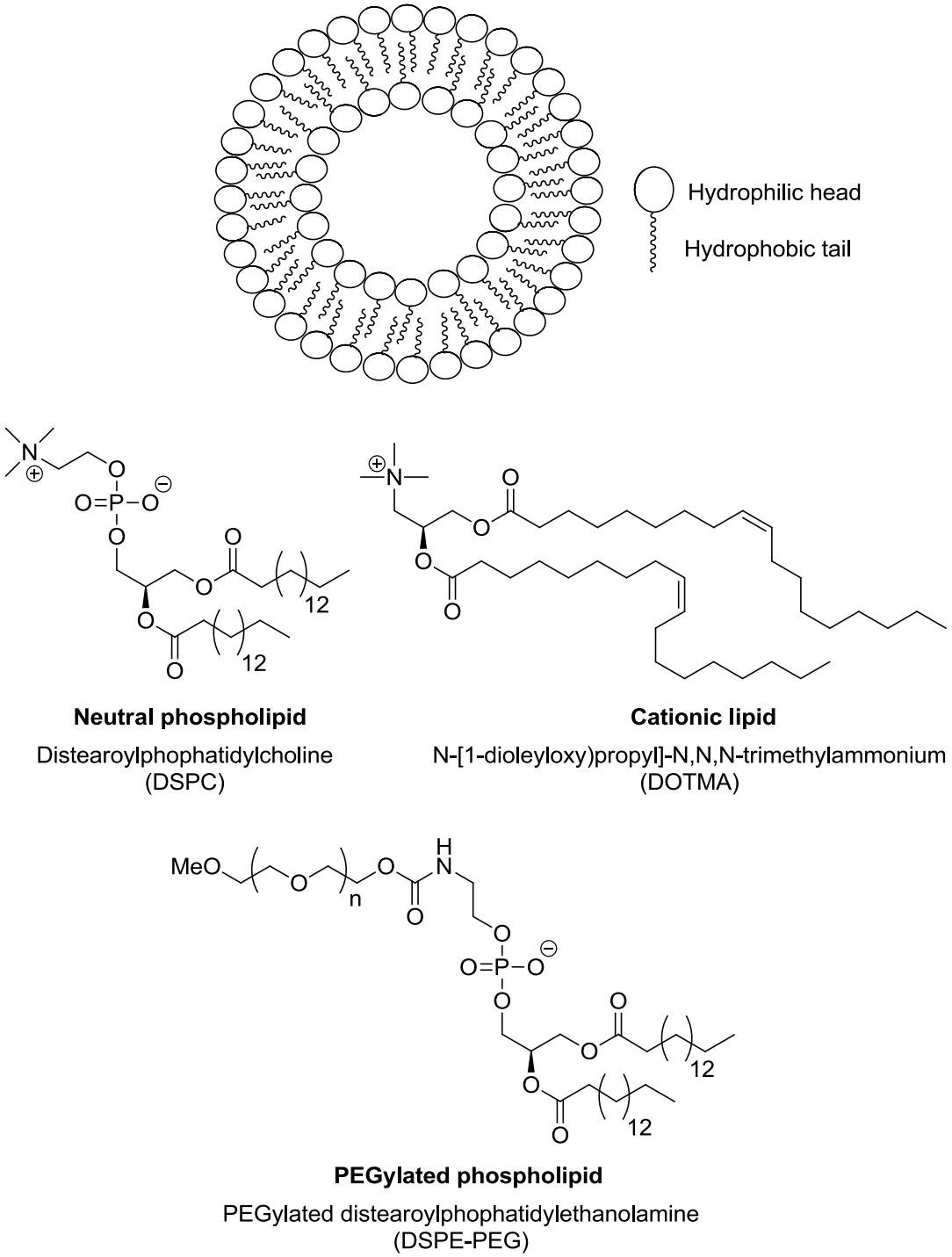


Figure 1.3: Schematic representation of a unilamellar liposome and the structure of some lipids typically used in formulation

Cell penetrating peptides

Cell penetrating peptides (CPP) are short positively charged peptides that can cross the cell membrane barrier.[31] Their structure and sequence can vary substantially, but there is always an abundance of arginine or lysine residues giving them a net positive charge (Table 1.1). The mechanism of internalisation is commonly believed to be endocytosis triggered by electrostatic interactions with negatively charged glycosaminoglycans on the surface of cells.[31] In general, the concentrations of CPPs required for cell entry are low enough to induce a biological effect without causing cytotoxicity, or at the very least they are less toxic than comparable delivery vehicles.

CPP name	Sequence
Tat	RKKRRQRRR
Antp	RQIKIWFQNRRMKWKK
Transportan 10	AGYLLGKINLKALAALAKKIL
Polyarginine	$(R)_n$

Table 1.1: Name and sequence of CPPs peptides: Tat[32], Antp,[33] Transportan 10,[34] and polyarginine[35]

CPPs are powerful delivery vehicles that have been used to transport a range of macromolecules, including oligonucleotides and proteins, inside cells.[36-41] Interestingly, both covalently and non-covalently linked oligonucleotides have been successfully transported using the same technique.[37-41] However, there are still some major limitations associated with CPP mediated delivery.

CPPs frequently become trapped in endosomal compartments after internalisation and this can limit the efficacy of the therapeutic cargo.[31] Endosomal release can be assisted by the incorporation of membrane destabilising sequences, such as KALA[33] or KLLKLLL,[32] into the peptides structure. Additionally, CPPs can be co-administrated with lysosomotropic agents, but they are usually toxic and cannot be used over a prolonged period of time.

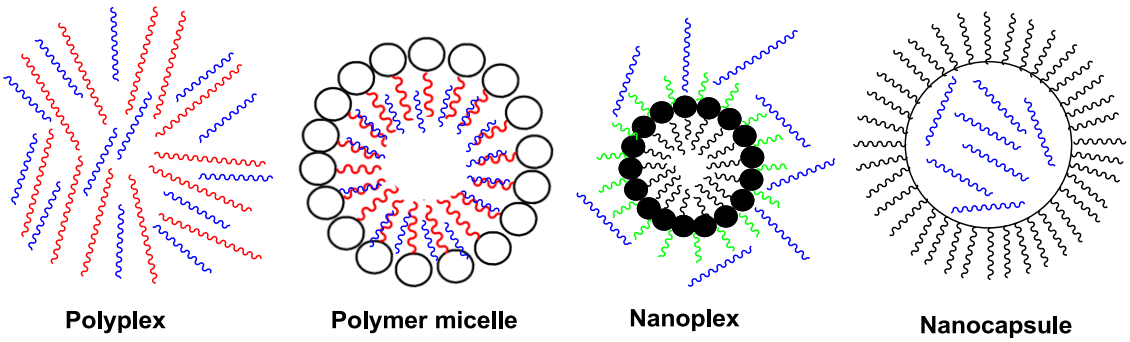
The other main problem of CPP delivery vehicles is their poor stability, *in vivo*.[42] The peptides are subject to degradation by proteases in blood, although chemical modifications to their structure have been shown to increase resistance.[31,43] Furthermore, they are rapidly removed by the MPS of the body and show wide biodistribution indicating low specificity.[31,42] Until these issues are addressed the wider application of CPP delivery systems are limited.

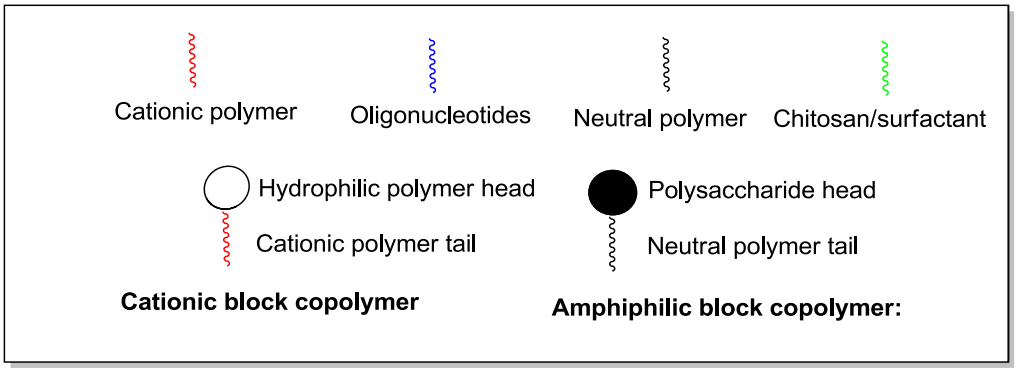
Polymeric carriers

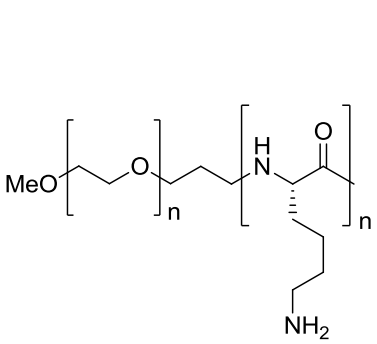
There are several different classes of polymer-based delivery vehicles that can carry oligonucleotides; these include polyplexes, polymer micelle and nanoplexes (Figure 1.4).[44] Polyplexes consist of cationic polymers, such as polyethylenimine, that non-covalently associate with the negatively charged backbone of oligonucleotides forming a polyanionic complex.[44,45] Polymer micelles are produced by the spontaneous assembly of block copolymers, with cationic (e.g. polylysine) and hydrophilic (e.g. PEG) properties, around oligonucleotides.[44,46,47] The resulting complex has a hydrophilic surface surrounding a cationic centre associated with nucleic acids. Nanoplexes are amphiphilic nanoparticles typically produced by the block copolymerisation of poly(alkylcyanoacrylate) with polysaccharides.[44,48] The surface of the nanoparticle is additionally coated with chitosan or surfactants materials to create a positively charged surface that can complex with oligonucleotides to form the nanoplex.

The aforementioned polymeric carriers can be prepared by several different methods, but many produce intrinsically toxic delivery vehicles.[44] The toxicity is either attributed to the materials used in formulation or the high positive charge of the final polymeric carrier. It is often possible to replace the hazardous materials used in the synthesis of the polymeric carrier, but reducing the cation abundance is more problematic. The cationic charge is the basis of oligonucleotide association and has also been demonstrated to facilitate transfection by increasing endosomal escape. However, there is a relatively new class of polymer carrier termed nanocapsules that could be used a viable alternative.

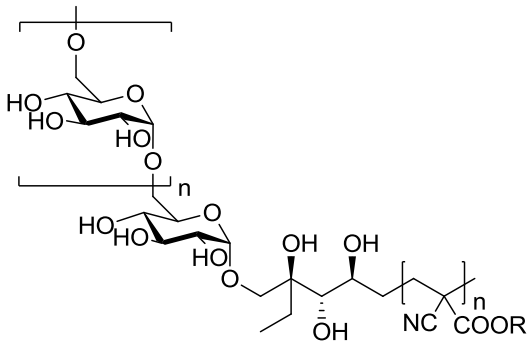
Lambert *et al.* developed a polymerisation technique in an organic solvent/aqueous emulsion that generated polymeric vesicles with an aqueous core surrounded by a polymer membrane, such as poly(iso-butylcyanoacrylate).[49] Oligonucleotides can be encapsulated inside the aqueous interior protecting them from the external environment. Furthermore, oligonucleotide incorporation does not require the use of cationic polymers. However, the efficiency of incorporation is greatly improved when cationic polymers are coencapsulated.[44,50] This class of carrier are considered non-toxic, but their formation typically requires the use of hazardous chlorinated solvents that need to be stringently removed before use *i*n vivo.[44] Further investigation is still required to assess the capability of nanocapsules based delivery vehicles



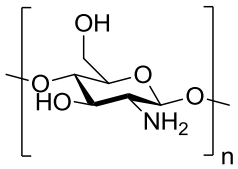




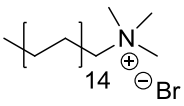
Cationic block copolymer:
Poly(lysine) and PEG



Amphiphilic block copolymer:
Poly(alkylcyanoacrylate) and dextran



Polymer coating:
Chitosan



Surfactant polymer coating:
Cetrimonium bromide

Figure 1.4: Schematic representation of polymer-based delivery vehicles and the structures of their typical polymeric components. [44,51] R= Butyl or isobutyl

1.2 Bacterial toxins

Bacterial toxins, such as cholera and shiga toxins,[52] have evolved specific and efficient mechanisms to breach the cell membrane barrier allowing them to enter mammalian cells effortlessly. If these mechanisms could be manipulated it could lead to the development of new range of cell penetrating biological tools.

1.2.1 AB_5 enterotoxins

Cholera toxin (CT), secreted from *Vibrio cholera,* and the closely related heat-labile enterotoxin (LT), from *Escherichia coli*, share an almost identical protein sequence and structure (Figure 1.5).[53-55] The proteins are composed of a toxic A-subunit (CTA/LTA) surrounded by a pentamer of B-subunits (CTB) that are combined in a hexameric AB_5 structure (Figure 1.6).[53,56] The A-subunit is snipped by a bacterial endoprotease splitting it into two distinct domains that are connected by a single disulfide bond, the A1 and A2-domain (CTA1/LTA1 and CTA2/LTA2).[57]

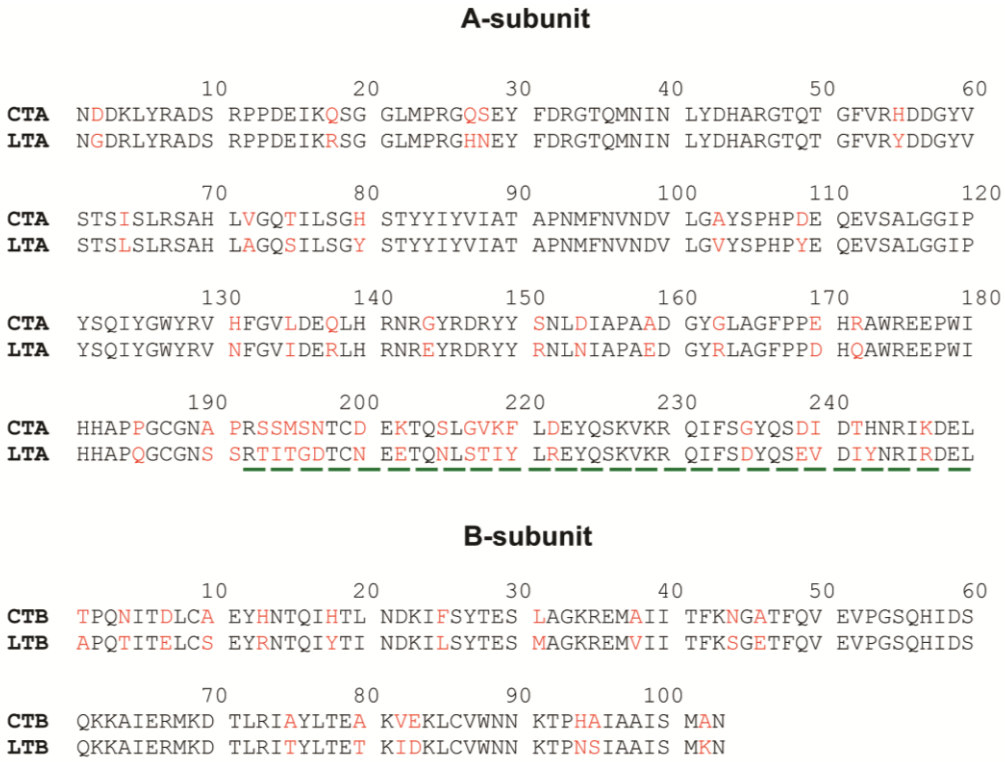


Figure 1.5: Protein sequence of cholera toxin (CT) and heat-labile enterotoxin (LT) taken from the UniProt codes: P01555 (CTA), P01556 (CTB), P06717 (LTA) and P32890 (LTB). Green-dashed line: A2-domain sequence; red: Differences between the CT and LT sequences.

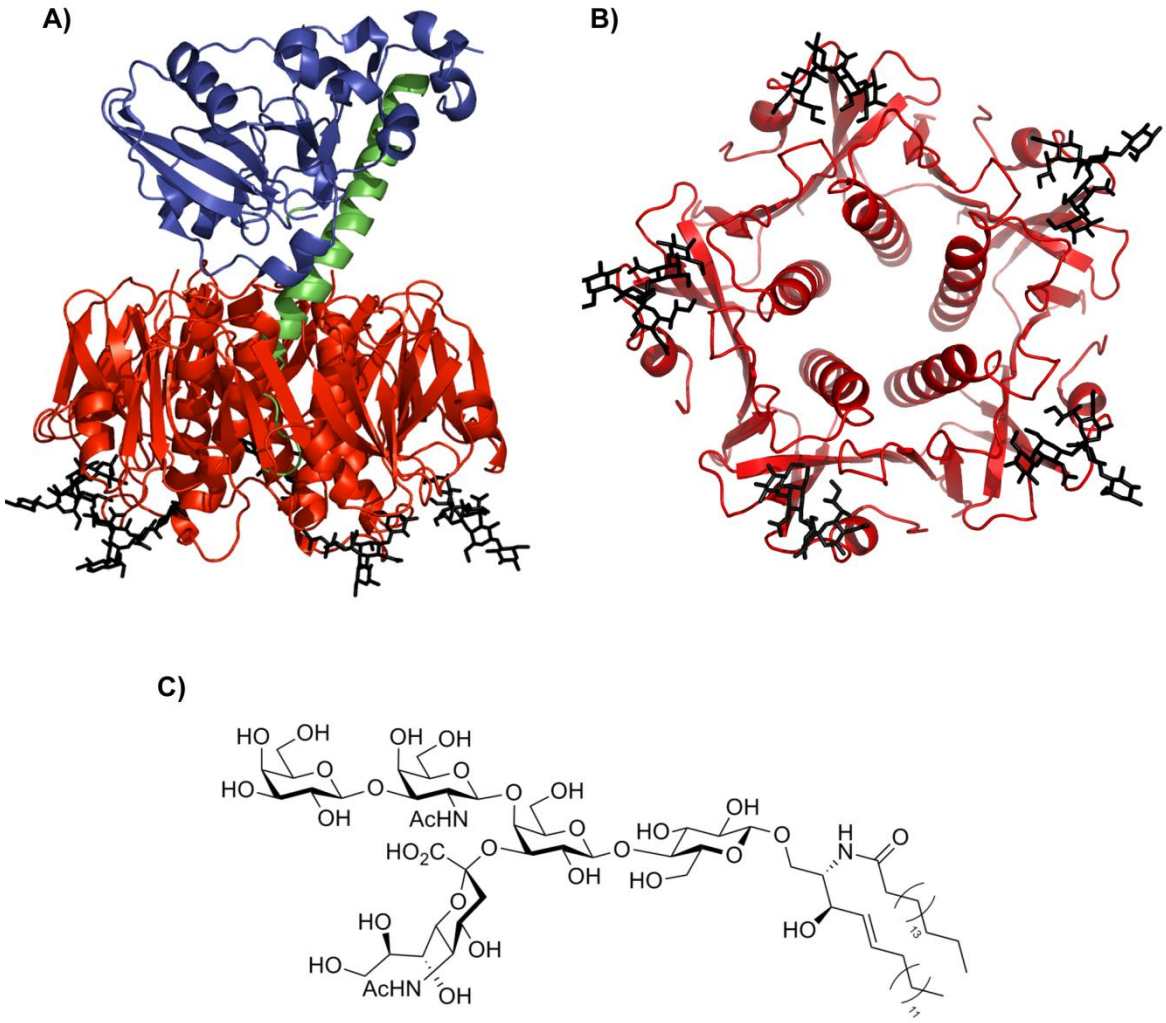


Figure 1.6: A) Overlaid crystal structures of cholera toxin generated from the PDB codes of atomic coordinates: 1S5E and 3CHB (UniProt codes: P01555 and: P01556).[178, 179] A1-domain (blue); A2-domain (green); CTB pentamer (red); GM1 pentasaccharide (black). B) Crystal structure of the GM1 pentasaccharide (black) bound to the bottom surface of the pentameric B-subunit (red) generated from the PDB codes of atomic coordinates: 3CHB C) Structure of the GM1 ganglioside

The holotoxins infect their host using an identical retrograde trafficking mechanism.[53-55] This process is initiated when the toxins bind to the plethora of GM1 ganglioside receptors located at the epithelial cell surface. The five binding sites located on the bottom face of the pentameric B-subunit produce a strong multivalent interaction between the surface of the cell and the toxins.[53,58] This powerful association causes the plasma membrane to engulf the proteins in an endocytic vesicle, which is subsequently trafficked inside the cell.[55,57,59] The membrane phase separates to form regions that are rich in cholesterol and gangliosides and it has been found that these areas, termed lipid rafts, are essential for the successful delivery of the AB_5 toxins to their intracellular target.[55,57,59] The vesicle containing the toxin is delivered to the endoplasmic reticulum (ER) via the trans-Golgi body; however, the mechanisms governing the intracellular transportation process are poorly understood (Figure 1.7).[55,57,59,60]

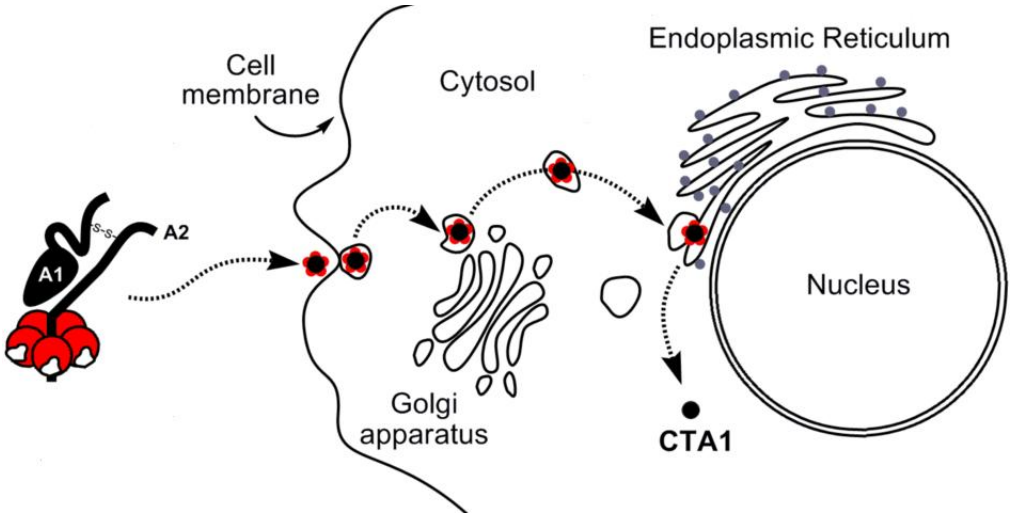


Figure 1.7: Diagram of CT/LT mode of action on mammalian cells.[61]

It has been proposed that when cholera toxin reaches the ER, a protein disulfide isomerise (PDI) enzyme reduces the disulfide bond between the A1 and A2 domains.[55,57,59,60] The free A1-fragment partially unfolds triggering the internal machinery of the ER to eject the protein into the cytosol for proteasomal degradation.[62,63] However, ADP-ribosylation factor 6 binds to the A1-protein in the cytosol and protects it from degradation.[53,63] This process allows the A1-protein to reach its target G-protein and induce toxicity. The A1-protein has ADP-ribosyltransferase activity that modifies the G-protein and inhibits the α-subunit from GTPase activity.[59] This inhibition results in a constant stimulation of adenylate cyclase causing a sharp increase in the concentration of cyclic adenosine monophosphate (AMP) within the cell. As a consequence of this process, ions and fluid are continually lost from the cell triggering watery diarrhoea in the host.

The retrograde trafficking mechanism of the toxins is not cytotoxic;[54] therefore, the transportation process can be hijacked and used to deliver other macromolecules, such as oligonucleotides, into mammalian cells. Several studies have already made significant progress in developing enterotoxin-based delivery vehicles. For example, Tinker *et al*. fused a non-cleavable green fluorescent protein (GFP) to the A2-domain of a truncated cholera toxin.[64] The fluorescent protein was then successfully delivered to the ER of both Y1 adrenal and Vero cells. In a very recent study,[65] Daniell and co-workers bioencapsulated CTB chimeras by expressing the proteins in the chloroplast of plant cells. When orally administered to mice, the bioencapsulated proteins were transported safely to the intestine before eventually crossing the blood brain barrier to produce a therapeutic effect. Impressively, the authors used CTB fused to myelin basic protein to reduce the number of amyloid fibrils in the neurons of genetically altered mice with Alzheimer's disease. In an earlier study, Barrett *et al.* used a CTB-polylysine

conjugate to transport plasmid DNA inside neurons (rat PC12 pheochromocytoma) leading to the transgene expression of GFP within the cells.[66] Furthermore, *in vivo* studies have shown CTB preferentially targets motor neurons when administered to mice and rats with intraperitoneal, intravenous and intramuscular injections.[67,68] This research has established enterotoxins as viable delivery vehicles, but it has also highlighted their rare ability to target motor neurons and cross the blood-brain barrier. In principle, therapeutics linked to the proteins could be used to treat neurological disorders that are currently difficult to target with conventional drugs.

1.2.2 Formation of novel AB_5 complexes

Investigation into the assembly of holotoxins has led to the development of novel AB_5 complexes both *in vivo* and *in vitro.*

1.2.2.1 AB5 formation *in vivo*

Research into the kinetics of cholera toxin assembly, *in vivo* and *in vitro*, was originally pioneered by Hirst and co-workers.[69] The team discovered that CTA does not interact with a preformed pentameric CTB to generate complete cholera toxin, as previously thought.[64,69] On the contrary, it was found that holotoxin assembly only occurred when CTA interacted with CTB monomers during the assembling process.[64] Secondly, the rate of formation of pentameric CTB in *E. coli* cells increased three fold in the presence of CTA.[69] These discoveries were surprising because it had been previously thought that CTA did not play an active role in the formation of pentameric CTB as its formation had been observed without the presence of any CTA.[69,70]

Unsurprisingly, the mechanism of heat-labile enterotoxin assembly seems to be controlled by very similar processes to its closely related cousin, cholera toxin. In another study by Hirst and co-workers,[71] it was found that the total observed yield of pentameric LTB expressed from *E. coli* cells was five-fold lower when it was expressed in the absence of LTA. Moreover, the yield of pentameric LTB even increased when LTA was expressed from a different plasmid within the same *E. coli* cell; this indicated the increase in yield was attributed to a posttranslational process.

The specific interactions between LTA and LTB were explored by Hirst and co workers, using site directed mutagenesis to create LTA mutants with truncated *C*-terminal sequences (Table 1.2).[71] The mutants had their amino acid sequence reduced by four and fourteen residues. Stable holotoxin generation was not observed when LTB was coexpressed with the truncated LTA mutants. Interestingly, both mutants did enhance the yield of pentameric LTB assembly

compared to LTB being expressed on its own. Therefore, it was postulated that interactions distal to the *C*-terminal extremities of LTA are important for the facilitation of LTB expression and/or LTB pentamer assembly. Furthermore, interactions between LTB and the final four *C*-terminal residues of LTA play a key role in anchoring the A-subunit in place after holotoxin generation.[71]

LTA	*C*-terminal sequence
Wild type	YGSEVGIYNRIRDEL
Mutant 1	YGSEVGIYNRI
Mutant 2	Y

Table 1.2 Amino acid sequence of the truncated LTA mutants. [71]

Jobling *et al*. made fusions proteins of CTA2 with maltose binding protein (MBP), β-lactamase (BLA) and bacterial alkaline phosphatase (BAP).[72] The protein fusions were coexpressed with CTB to successfully generate hybrid AB_5 complexes for the first time. This research demonstrated the potential of using chimeric CTA2 proteins to create a range of different AB_5 complexes. Intriguingly, when the fusion proteins were expressed in *V. cholera* they were not transported through the secretory apparatus of the bacteria into the growth media. This is probably due to the large steric bulk of the new hybrid proteins blocking the secretory pathway.

1.2.2.2 AB5 formation *in vitro*

It is possible to reform viable AB_5 complexes from their constituent subunits *in vitro*.[73-75] The basic procedure involves fully denaturing the B-subunits into their monomeric states using acid, or a combination of acid and reducing agent. The acidic-protein solution is then neutralised and the A-subunit is immediately added to allow the AB_5 complex to reform. The exact method used to assemble the holotoxin changes depending on whether CTB or LTB are being used. CTB and LTB share nearly 80% sequence identity, which gives them almost identical properties (Figure 1.5).[76] However, they have different stabilities under acidic conditions. When exposed to pH 1.5 for 2.5 minutes, only 20% of LTB monomers will reassemble into pentamers.[77] In contrast, CTB can remain in acidic condition at pH $\leq$ 1 for up to an hour before its ability to reassemble into pentamers is compromised.[78] The discrepancy between the acid stability of the B-pentamers is probably due to a cis-trans isomerisation of a proline residue (Pro-93) shared by both CTB and LTB. The proline isomerisation alters the configuration of the B-monomer preventing vital interactions with other monomeric proteins and thus inhibiting pentamer assembly.[77,78] Therefore, it is reasonable to assume that CTB must be more resistant to this isomerisation change under acidic conditions compared to LTB.

Takeda *et al.* demonstrated that the A and B subunits of heat-labile enterotoxin and cholera toxin could be assembled inter-changeably *in vitro.*[73] These hybrid holotoxins comprised of LTA/CTB and CTA/LTB proteins showed exactly the same activity as the parent holotoxins from which the A subunits were derived.

It has also been shown that CTA2 fusion proteins can be assembled with CTB to form novel AB_5 structures *in vitro*. Hatic *et al.* generated a fluorescent AB_5 protein by combining CTB monomers with a CTA2-GFP fusion.[74] However, less than 20% of the fluorescent AB_5 protein was recovered. Interestingly, the efficiency of AB_5 complex formation in other *in vitro* experiments has generally gone unreported, which suggests this approach is low yielding

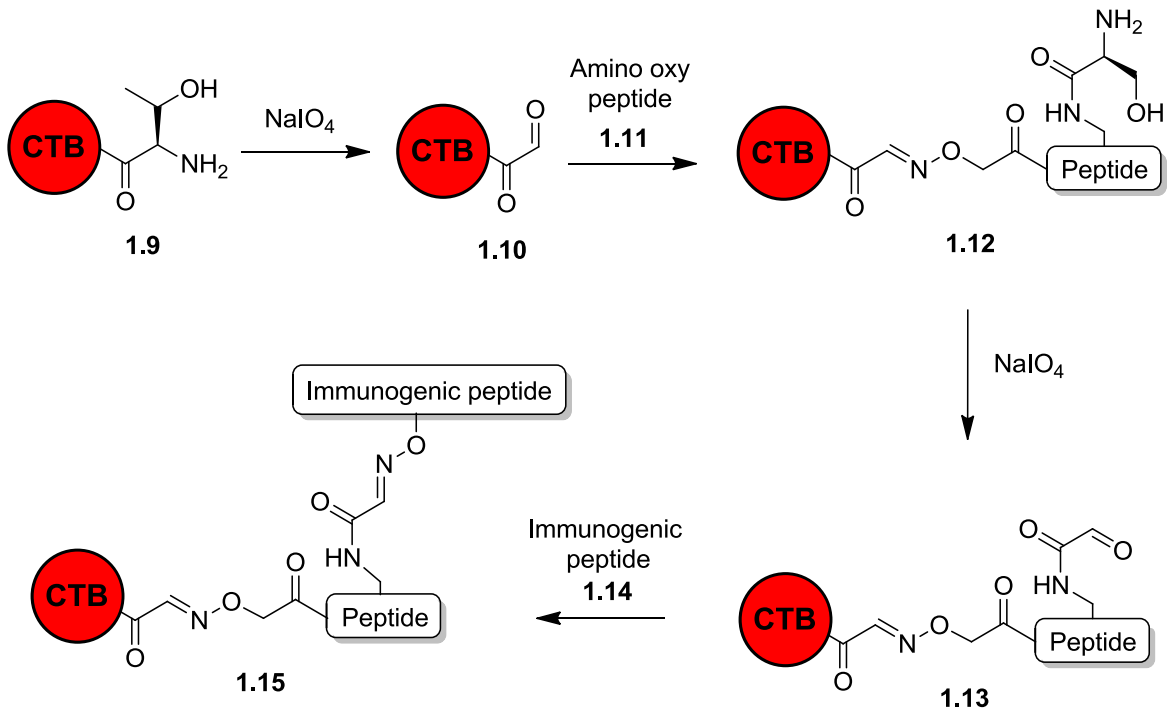
1.2.3 Formation of Novel LTB/CTB Proteins

Modifications to the *N*- or *C*-termini of LTB and CTB have also been reported.

1.2.3.1 N-terminal Modifications

Dertzbaugh *et al.* successfully fused Glucosyltransferase B to the *N*-terminus of CTB via a short peptide linker.[79] The monomeric CTB chimeras formed viable pentamers and also produced an immune response to *Streptococcus mutans*, the bacteria that express Glucosyltransferase B. However, there are limitations to this approach. In another study by Dertzbaugh *et al.*,[80] it was found that the GM1 binding affinity of CTB fusions decreased with increasing *N*-terminal linker length. Interestingly, the ability of the chimeras to induce an immune response also decreased.

Rose *et al*. used a chemical based approach to modify pentameric CTB.[81,82] Peptide extensions were introduced by oxidising the *N*-terminal threonine residue **1.9** to aldehyde **1.10** before peptide **1.11** was coupled via an oximation reaction (Scheme 1.1). Peptide **1.11** contained multiple lysine residues covalently linked, though their amino side chains, to serine residues. The serine residues were subsequently oxidised and reacted with aminooxyacetylated immunogenic peptide **1.14** to generate modified CTB monomers **1.15**.[81,82] The *N*-terminal extensions did not appear to affect the GM1 binding ability of CTB and in one instance, the hybrid protein actually showed an increased binding affinity to GM1.



Scheme 1.1: Chemical modification of CTB by the oxidation and subsequent oximation of the *N*-terminal threonine residue.[81]

1.2.3.2 *C*-terminal Modifications

A number of different research groups have also investigated the effects of *C*-terminal modifications to LTB and CTB. Sandkvist *et al.* extended the *C*-terminus of LTB with random peptide sequences (Table 1.3).[83] The extensions affected the ability of some of the mutants to form stable pentamers and in one case, no LTB protein was detected after expression. Intriguingly, all the mutants were unable to complex with the A-subunit to generate AB_5 proteins.[83] This study highlighted how sensitive the *C*-terminal structure of LTB is to alterations. However, the amino acid extensions were chosen randomly, so it is difficult to ascertain whether it was their length and/or sequence that affected LTB assembly.

No. of amino acid in extension	*C*-terminal sequence	Assembly
0 (Wild type)	EN	Holotoxin
6	EKLGCFGG	Monomeric
16	EKLAWLFWRMRGDFQPDTD	Monomeric
6	EKLFQPDTD	Pentameric
6	EKLAPQKRW	Pentameric
12	ESLAVLADERRFSA	Not detected

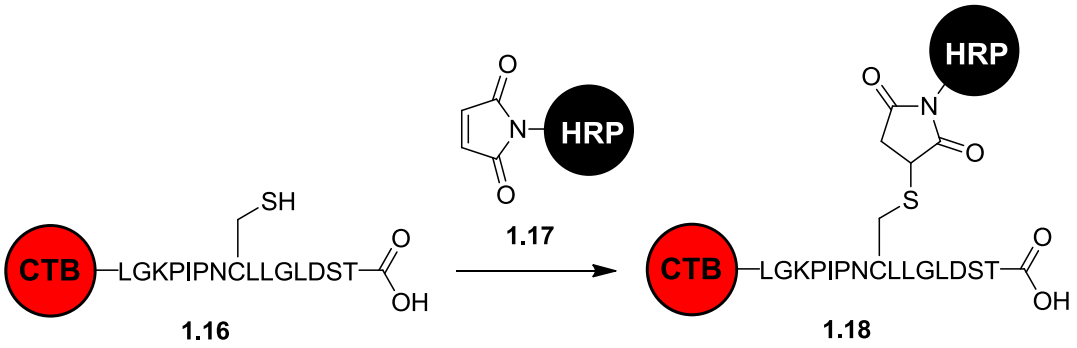
Table 1.3: Amino acid sequence and assembly of LTB wild type and *C*-terminal mutants.[83]

A peptide epitope, designated Pk, was fused onto the *C*-terminus of LTB in a study by Green *et al.* (Table 1.4);[84] the Pk epitope is recognised by the monoclonal antibody SV5-P-k3. The hybrid protein was expressed in *Vibrio sp* 60 cells and exported into the growth media. The expression yield of the hybrid protein was less than wild type LTB; the low yields were probably due to degradation of the peptide extensions by proteases native to *Vibrio sp* 60 cells. Interestingly, the protein fusion was still able to form stable pentamers and raise an immune response to both LTB and the peptide tag. In a separate study,[85] Loregian *et al*. extended the *C*-terminus of LTB with a DNA polymerase fragment (pol) from herpes simplex virus type 1 and observed very similar results concerning protein assembly to the findings reported by Green *et al.* (Table 1.4).

Peptide tag	No. of amino acids in extension	*C*-terminal sequence
Pk	14	GKPIPNPLLGLDST
pol	27	LPTGEASPGAFSAFRGGPQTPYGFGFP

Table 1.4: Amino acid sequence of *C*-terminal LTB fusion tags. [68,69]

The initial research into LTB-Pk fusions was further developed in a study by O'Dowd *et al.*[86] The group introduced a cysteine residue at different positions along the *C*-terminal peptide extension in a series of different LTB-Pk mutants. The expression yield of the fusion proteins decreased with the introduction of cysteine. However, the addition of EDTA to the growth media helped increase the yield by inhibiting native proteases. A Michael addition was used to couple Horseradish peroxidase (HRP) **1.17**, functionalised with a maleimide group, to the cysteine residue in LTB-Pk **1.16** (Scheme 1.2). The LTB-Pk-HRP protein **1.18** was able to bind to GM1 at the surface of Vero cells and induce a mucosal immune response to HRP in mice. This research indicated that chemically-modified LTB could be used as an adjuvant and may be useful in the development of vaccines.



Scheme 1.2: The coupling of CTB to HRP via a Michael addition.[86]

It is clear that extensions to the *C*-terminus of LTB are possible without reducing its ability to form stable pentamers. However, it is less clear what sequence and length of peptide extension will defiantly not interfere with pentamer assembly. Furthermore, no AB_5 complex formation has been observed with *C*-terminal mutants

1.3 Protein modification

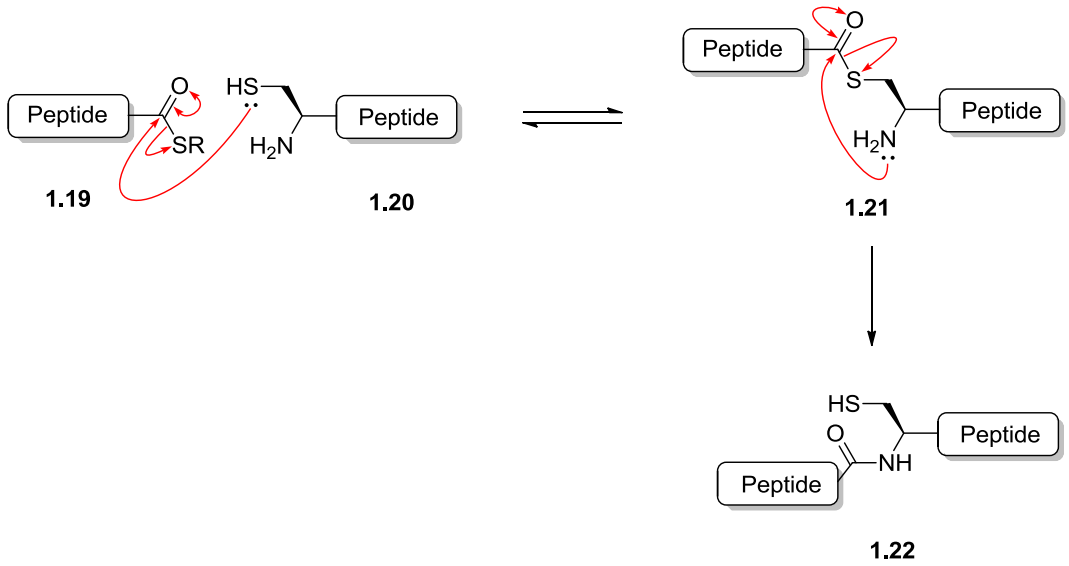
There are many bioorthogonal reactions that can be used to modify proteins, such as strain-promoted cycloadditions and Staudinger ligation.[87,88] However, these methods often necessitate the incorporation of non-natural amino acids into a target proteins.[89] Native chemical ligation (NCL) is one of the earliest and most widely used protein modification techniques; it has been utilised many times in the synthesis of synthetic proteins. In contrast, sortase-mediated chemical ligation is a relevantly new technique that is fast becoming the tool of choice for protein modification. The substrates for both NCL and SrtA are more readily accessible compared to alternative protein labelling techniques.

1.3.1 Native chemical ligation

A landmark paper by Kent and co-workers established a completely new method to connect polypeptides together, which they termed NCL (Scheme 1.3).[90] The authors discovered that substrates bearing an *N*-terminal cysteine residue **1.20** will form a new amide bond with peptides containing a *C*-terminal thioester **1.19**. A reversible transthioesterification reaction between the β-thiol of cysteine **1.20** and the *C*-terminal thioester **1.19** produces a thioester-linked intermediate **1.21**. Finally, in an irreversible intramolecular rearrangement, the intermediate **1.21** forms an amide-linked polypeptide **1.22**.[90] The reaction is usually performed on unprotected peptides under denaturing aqueous conditions with additional free thiol.[91,92] Catalysts, such as (4-carboxymethyl)thiophenol, can greatly increase the reaction speed.[93] The ligation is highly chemospecific even in the presence of other non-terminal cysteine residues.[90,91,94] Cysteine residues contained within the proteins structure are only able to form the transient thioester intermediate **1.21** as they lack the necessary nucleophilic amino group for the final transpeptidation reaction.[92] Furthermore, NCL proceeds with less than 1% racemisation.[91]

One of the main challanges of NCL is the preparation of the thioester peptides. Boc solid phase peptide synthesis (SPPS) coupled with thioester linkers is the most effective way to synthesis thioester peptides.[92,94,95] However, this process routinely requires the use of harsh and toxic reagents, so it is not accessible in many laboratories.[92] An alternative strategy, developed by

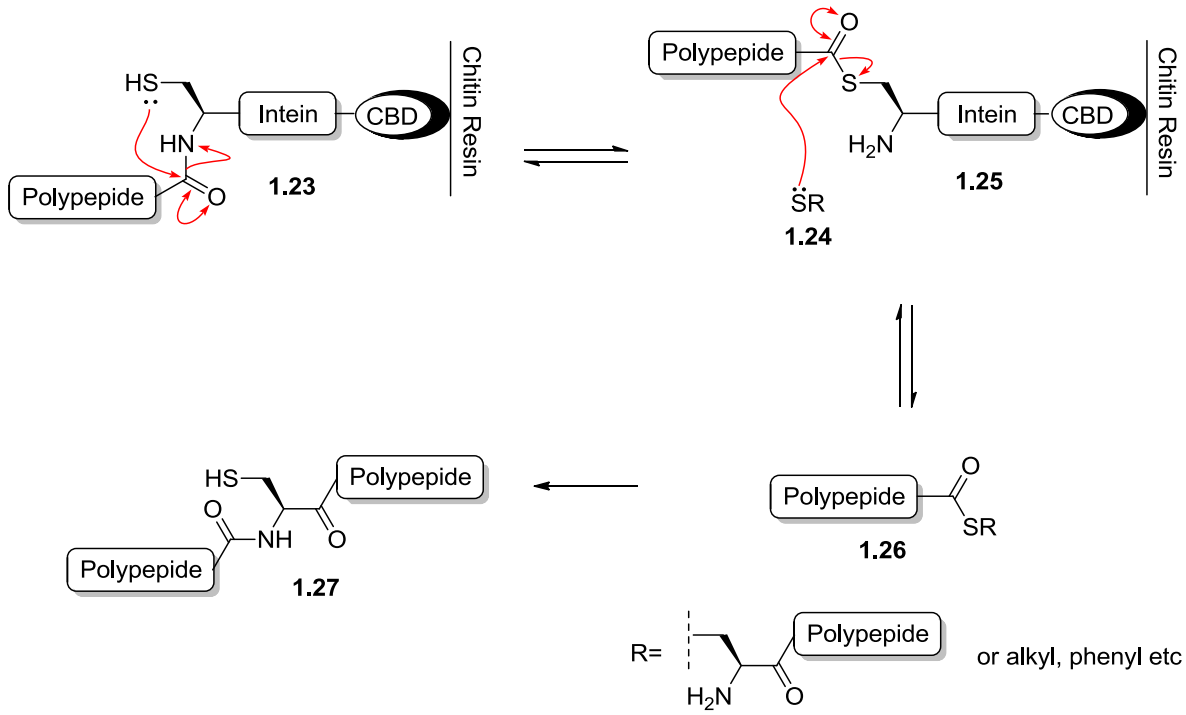
Dawson and co-workers, utilises a diaminobenzoyl linker attached to Rink amide resin to efficiently synthesise peptidyl thioesters using the more conventional Fmoc-SPPS approach.[96]



Scheme 1.3: General mechanism for native chemical ligation.[90]

The synthesis of larger peptidyl thioesters can be achieved recombinantly using an intein fusion.[91,97] Intein chimeras **1.23** have a unique folded structure that distorts the cysteine residue at the junction between the fused polypeptides (Scheme 1.4).[91] The *N*-terminal amide bond of the cysteine residue is twisted, so that the lone pair on the nitrogen is not fully conjugated into the system. The β-thiol on the cysteine residue can attack the distorted amide bond and trigger a reversible intramolecular rearrangement forming a thioester **1.25**.[91,97] The thioester intermediate **1.25** can be intercepted using a suitable thiol **1.24** to generate a peptidyl thioester **1.26**. Exploiting this mechanism, Chong *et al.* developed a technique to purify proteins **1.23** that had been fused to an intein and a chitin-binding domain (CBD). The authors captured the protein chimera **1.23** on chitin resin and used thiols **1.24** to cleave the target protein from the immobilised intein-CBD fusion, forming peptidyl thioesters **1.26**. Ingeniously Muir *et al.* combined this approach with NCL to generate ligated proteins **1.27** directly on solid support by intercepting the thioester intermediate **1.25** with an *N*-terminal cysteine peptide **1.26** instead of a thiol **1.24**.[98]

NCL is a powerful technique that has been used to make numerous synthetic proteins,[73] such as crambin and human lysozyme.[99,100] However, ligations are typically carried out under denaturing conditions, so there is no guarantee that every synthetic protein will fold into the correct higher level structure. Interestingly, other macromolecules, including oligosaccharides and oligonucleotides,[101,102] can be attached onto peptide chains using this approach.



Scheme 1.4: Intein catalysed formation of *C*-terminal thioesters on chitin solid support

1.3.2 Sortase-mediated chemical ligations

Sortases are a class of transpeptidase enzyme that covalently sort and link proteins to the cell wall of gram-positive bacteria.[103-105] The first sortase discovered and consequently most studied enzyme is the type II membrane bound protein, Sortase A (SrtA) expressed by *Staphylococcus aureus*[103,105] The enzyme recognises proteins bearing a conserved LPXTG motif and ligates them to peptidoglycans substrates carrying an *N*-terminal oligoglycine sequence.[103,106,107]

$SrtA_{N\Delta59}$ is a truncated variant of SrtA that is catalytically active despite having 59 amino acid residues removed from its *N*-terminus. Zong *et al.* obtained a crystal structure of an inactive $SrtA_{N\Delta59}$ mutant (Cys184Ala) bound to an LPETG peptide substrate (Figure 1.8).[108] However, the substrate conformation in the active site appears to be slightly different to the description reported by the authors. For example, the lipophilic residues proline and leucine are supposed to be sat in a hydrophobic pocket that anchors the substrate in place, but the aliphatic side chain of leucine is not interacting with the surface of the protein at all. Furthermore, the glutamic acid was described as solvated; however, the carboxylate group is pointing towards the protein, which suggests an interaction is taking place. This is unusual as the glutamic acid in the substrate can be replaced with any amino acid; therefore, little or no interaction with the enzyme should be expected. The final threonine and glycine amino acids sit between the arginine (Arg197) and cysteine (Cys184Ala) active-site residues as reported. Although it is clear the

crystal structure does show enzyme-substrate binding, it is unlikely that the substrate is in the active conformation.

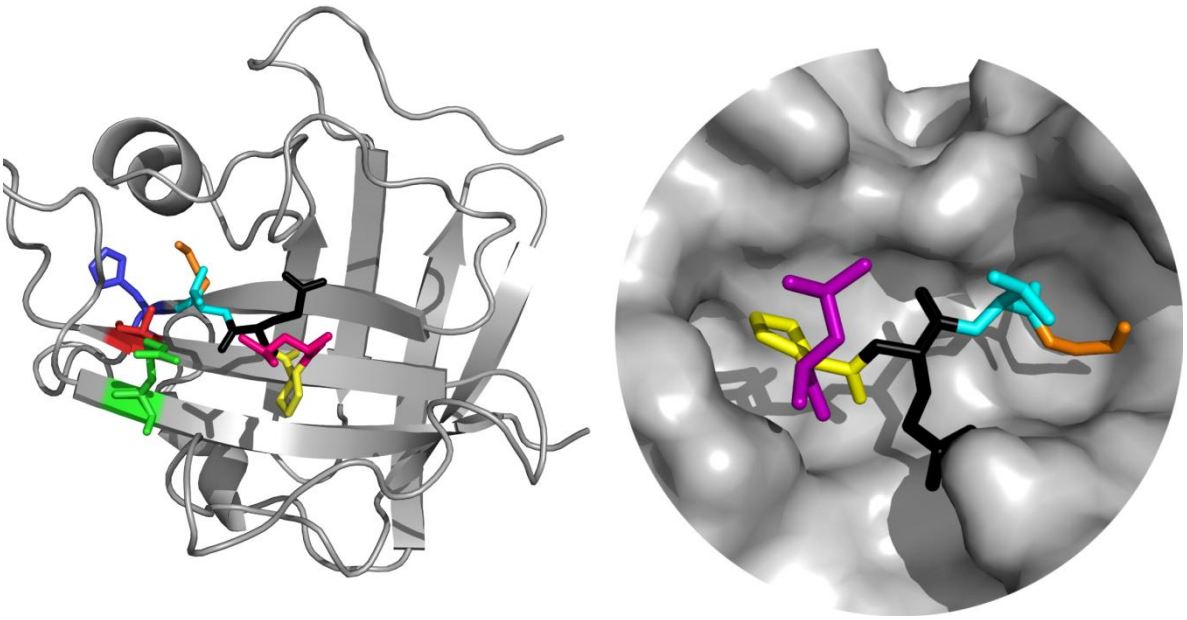
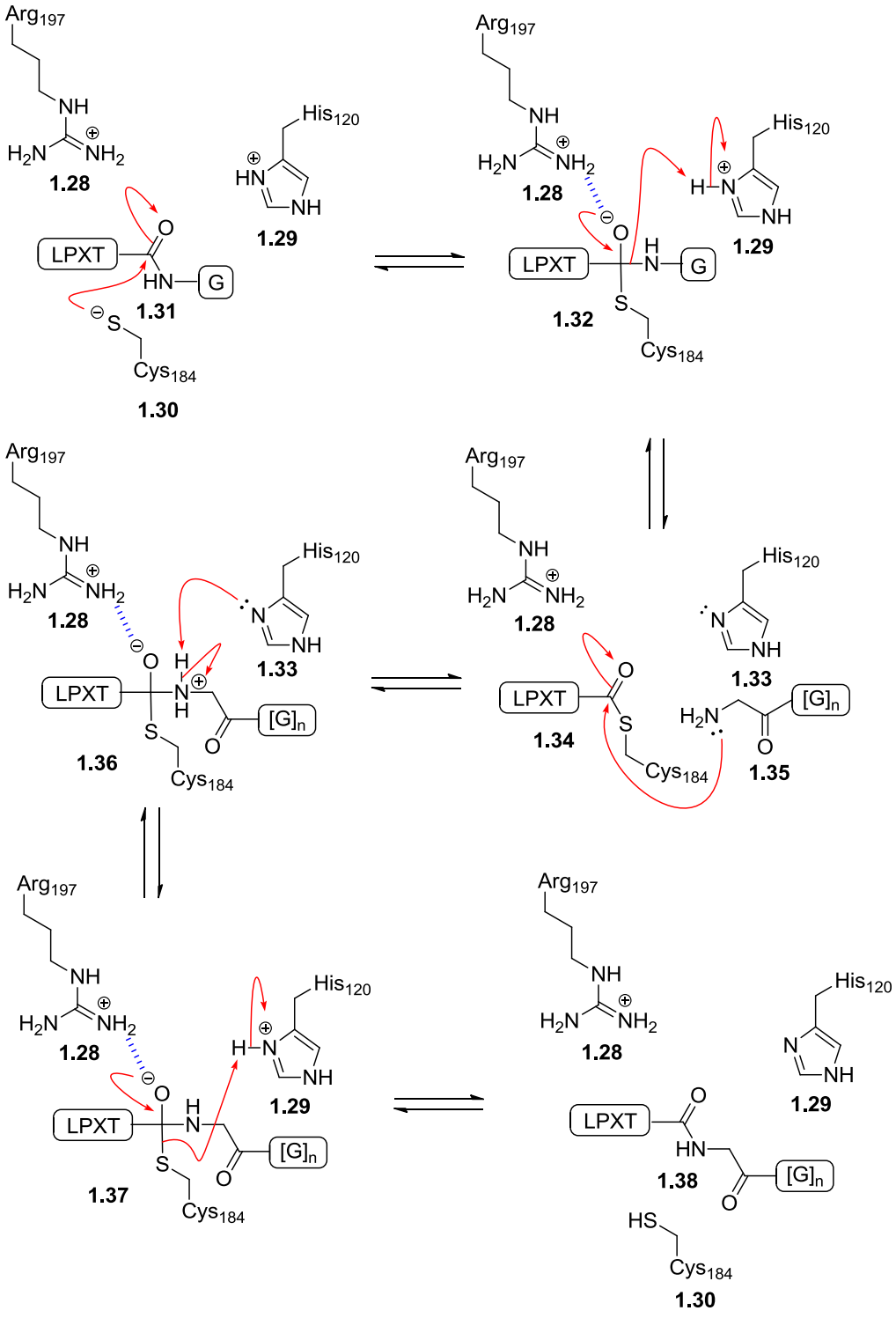


Figure 1.8: Crystal structure of the $SrtA_{N\Delta 59}$ Cys184Ala mutant bound to an LPETG substrate. Active site residues: Blue = His 120; Green = Arg 197; Red = Ala 184. Peptide residues: Orange = Glycine; cyan =: Threonine; black = Glutamic acid; yellow = Proline; and purple = Leucine. Generated from PDB codes of atomic coordinates: 1T2W (UniProt code: Q9S446).[108]

The SrtA/substrate complex is commonly thought to undergo a "reverse proton transfer reaction" (Scheme 1.5).[106,109,110] SrtA is primed when the active site cysteine **1.30** and histidine **1.29** residues exist in a thiolate–imidizolium ion pair; however, under physiological conditions, only o.o6% of the enzyme is in this activated state at any given time.[110,111] The deprotonated form of cysteine attacks the theronine and glycine amide bond of the substrate **1.31**.[106,109-112] This generates a transient oxyanion intermediate **1.32** that is stabilised by hydrogen bonding to the arginine residue **1.28** in the activate site. Subsequently, the intermediate collapses leading to loss of the *C*-terminal glycine amino acid and formation of thioester intermediate **1.34**. The cationic imidizolium group on the histidine residue **1.29** catalyses the cleavage event by protonating the departing glycine amino acid. The incoming oligoglycine substrate **1.35** attacks the enzyme-intermediate thioester to form the ligated peptide **1.38**. It has also been postulated that the imidazole group of the histidine residue **1.33** can act as a base and deprotonate the incoming oligoglycine-substrate **1.35**.[112,113]



Scheme 1.5: Proposed mechanism for the sortase mediated ligation reaction.[106,109,112]

Investigations into the LPXTG sequence has led to a better understating of the optimum recognition site needed for efficient ligations. Kruger *et al*. discovered that any amino acid residue can be incorporated at the X position of the substrate with minimal effect on effect on the total ligation yield;[114] however, cysteine and tryptophan were not tested. A very detailed study into the transpeptidation reaction by Pritz *et al*.,[115] highlighted the importance of amino

acids residues neighbouring the conserved LPXTG motif. The authors varied the amino acid sequence at several key positions after the *C*-terminal LPXTG sequence and measured the ligation efficiency using a model *N*-terminal oligoglycine peptide, GGGWW. They found that a peptide with an optimised LPXTGGG$[X]_n$ motif at the *C*-terminus gave the maximum ligated product yield. Interestingly, a significant increase in yield was also obtained by introducing just one additional glycine residue after the LPXTG motif. Substituting the additional glycine residue with other amino acids led to a decrease in product formation; this change was most dramatic when proline was used. These results support the previous kinetic studies carried out by Huang *et al.*,[116] which suggested the rate determining step of a sortase-transpeptidation reaction is the formation of the thioester intermediate and not attack of the glycine peptide substrate.

1.3.2.1 Substrates for SrtA-mediated modifications of proteins

There are various types of molecules that can react with the thioester enzyme intermediate. *N*-terminal diglycine peptides/proteins are the optimal substrates, but substantial ligation has been observed with substrates carrying only a single *N*-terminal glycine residue.[103,116,117] It has been reported that the thioester intermediate can interact with the ε-amine group of lysine residues; however, the ligation proceeds at an exceptionally reduced rate.[118] Hydroxylamine, a powerful inorganic nucleophile, can also act as a compatible substrate for the SrtA-mediated ligation.[116,119] The thioester intermediate will undergo hydrolysis, albeit very slowly, if no other suitable substrate is available for the transpeptidation reaction.[103,117] Hydrolysis proceeds so slowly that the kinetics of the reaction are inverted; the rate determining step becomes the nucleophilic attack of water instead of the formation of the thioester intermediate.[116]

There is now a huge catalogue of chemical probes that have been site-specifically incorporated into protein substrates using SrtA-mediated ligations.[111] In general, *C*-terminal modification simply requires the introduction of a solvent exposed LPXTG motif into a protein structure.[103,107,111] Protein labelling is then achieved by using peptide substrates carrying an *N*-terminal oligoglycine sequence. This methodology has been used to attach an assortment of chemical probes onto proteins including biotin **1.39**,[120] folate **1.40**, [117] cholesterol **1.41,** [121] and glycosylphosphatidylinositol (GPI) anchor **1.42** [122] (Figure 1.9).

The *C*-terminal labelling strategy can also be reversed. Peptide substrates carrying the LPXTG motif can be ligated onto proteins that present an accessible *N*-terminal glycine residue. Once again, this approach has allowed derivatisation of proteins with a wide range of functional groups, such as azides and fluorophores.[111,123] In principle, any molecule that can be

incorporated into the structure of a peptide can feasibly be attached onto a protein using SrtA. However, to achieve high level of protein modification a large excess of the label is generally required, which could be problematic if the label is precious

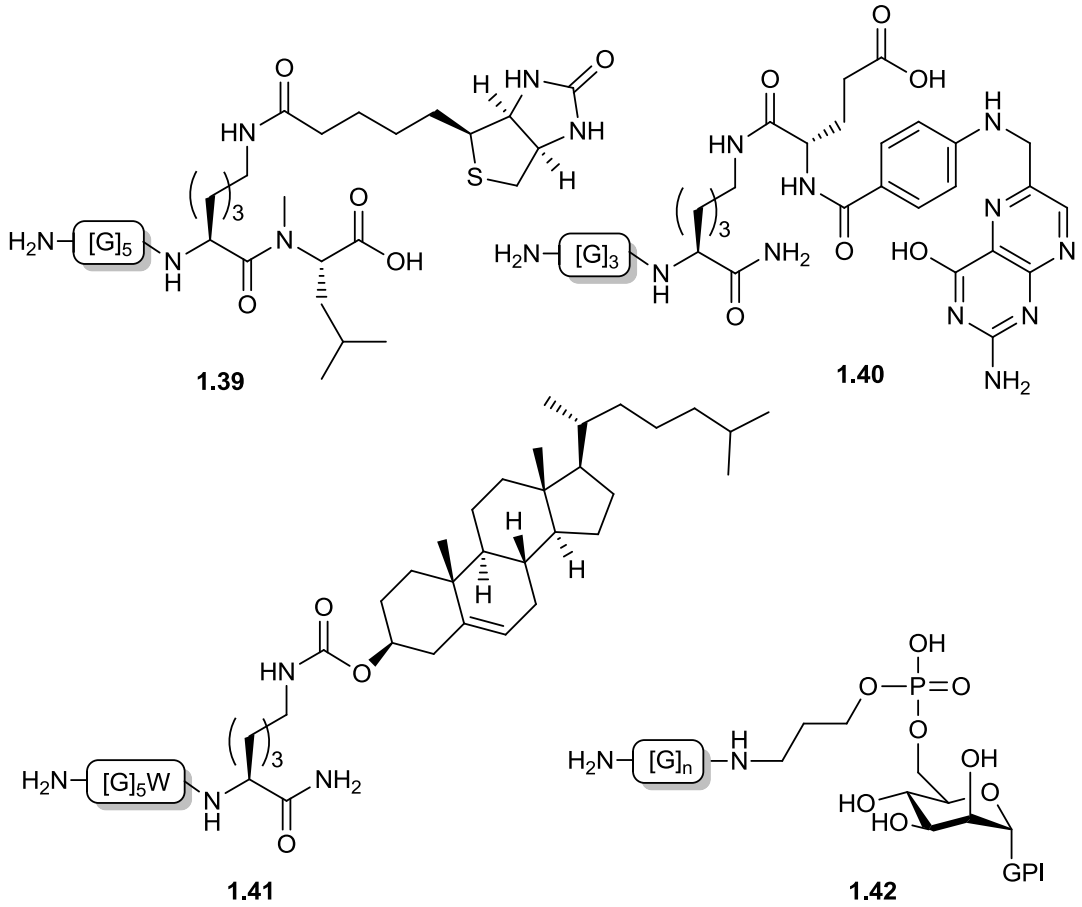
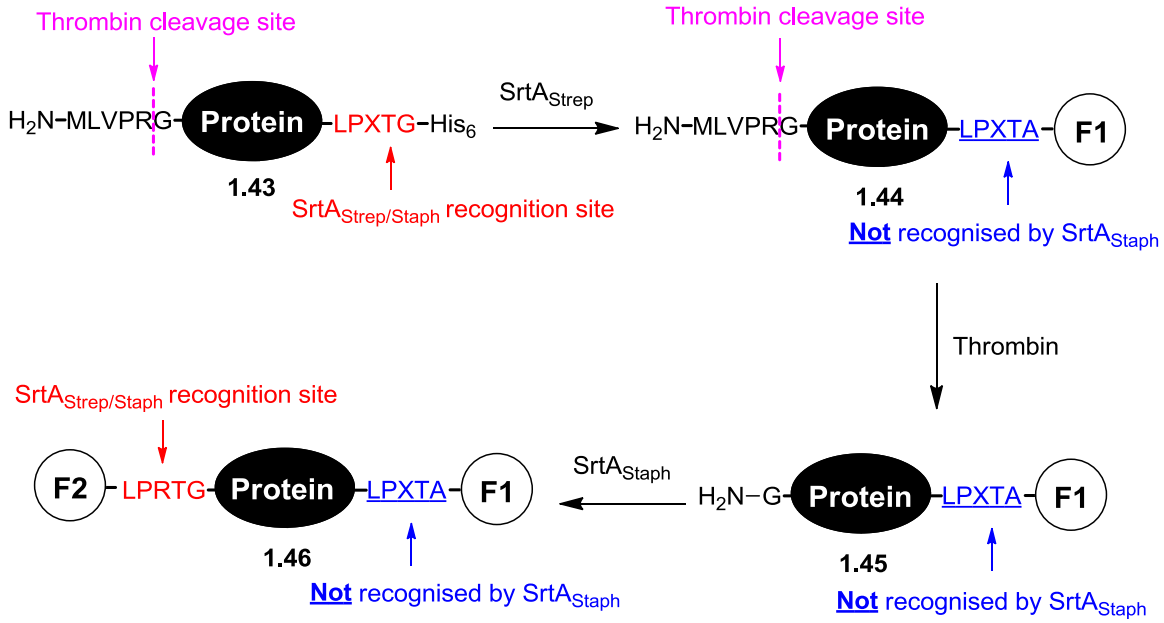


Figure 1.9: Structure of sortase substrates that contain: Biotin **1.39**; Folate **1.40**; Cholesterol **1.41** and a GPI anchor **1.42**

1.3.2.2 Duel SrtA-mediated modifications of both the N- and C-termini of proteins

An ingenious strategy developed by Ploegh and co-workers enabled both the *N*- and *C*-termini of human UCHL3 protein to be labelled using SrtA-mediated ligations (Scheme 1.6).[124] The success of the technique relied on the subtle difference in substrate recognition of SrtA derived from two different species of bacteria, *Streptococcus pyogenes* ($SrtA_{strep}$) and *Staphylococcus aureus* ($SrtA_{staph}$). The ligation mechanism of both enzymes is almost identical except $StrA_{strep}$ can accept *N*-terminal glycine and alanine substrates. Therefore, the ligated products of $StrA_{strep}$ can carry an LPXTA amino acid sequence, which is an extremely poor substrate for $SrtA_{staph}$. The team genetically incorporated a SrtA recognition site at the *C*-terminus, and a thrombin cleavage site at the *N*-terminus of UCHL3 **1.43**. $SrtA_{strep}$ was used to ligate a rhodamine-

conjugated peptide carrying an *N*-terminal alanine residue to the *C*-terminus of UCHL3 **1.44**. The modified protein **1.44** was treated with thrombin to expose an *N*-terminal glycine residue, which was then modified with a fluorescent peptide using $SrtA_{staph.}$ This process generated a duel labelled UCHL3 protein **1.46** at over 90% purity. Furthermore, to demonstrate the versatility of their approach, the authors also labelled the *N*- and *C*- termini of an eGFP protein.



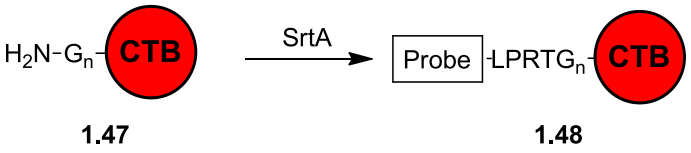
Scheme 1.6: *N*- and *C*- terminal labelling strategy of proteins UCHL3 and eGFP using SrtA-developed by Ploegh[124] F1 = fluorescein; F2= rhodamine.

1.3.2.3 SrtA-mediated modifications of cholera toxin

In a series of elegant studies conducted by Ploegh and co-workers,[60,124] SrtA ligations were used to modify both the A and B subunits of cholera toxin with a variety of chemical and biological probes.

N-terminal modification of CTB

Ploegh and co-workers developed a method to label the *N*-terminus of CTB using SrtA-mediated ligations (Scheme 1.7).[124] The authors introduced one, three or five glycine residues to the *N*-terminus of a series of CTB mutants and labelled the proteins using SrtA-mediated ligations. The mutants with three and five additional glycine residues were labelled successfully, but no modification was observed with the single glycine extension. This suggests the native *N*-terminus of CTB is not accessible to SrtA without an appropriate peptide extension.

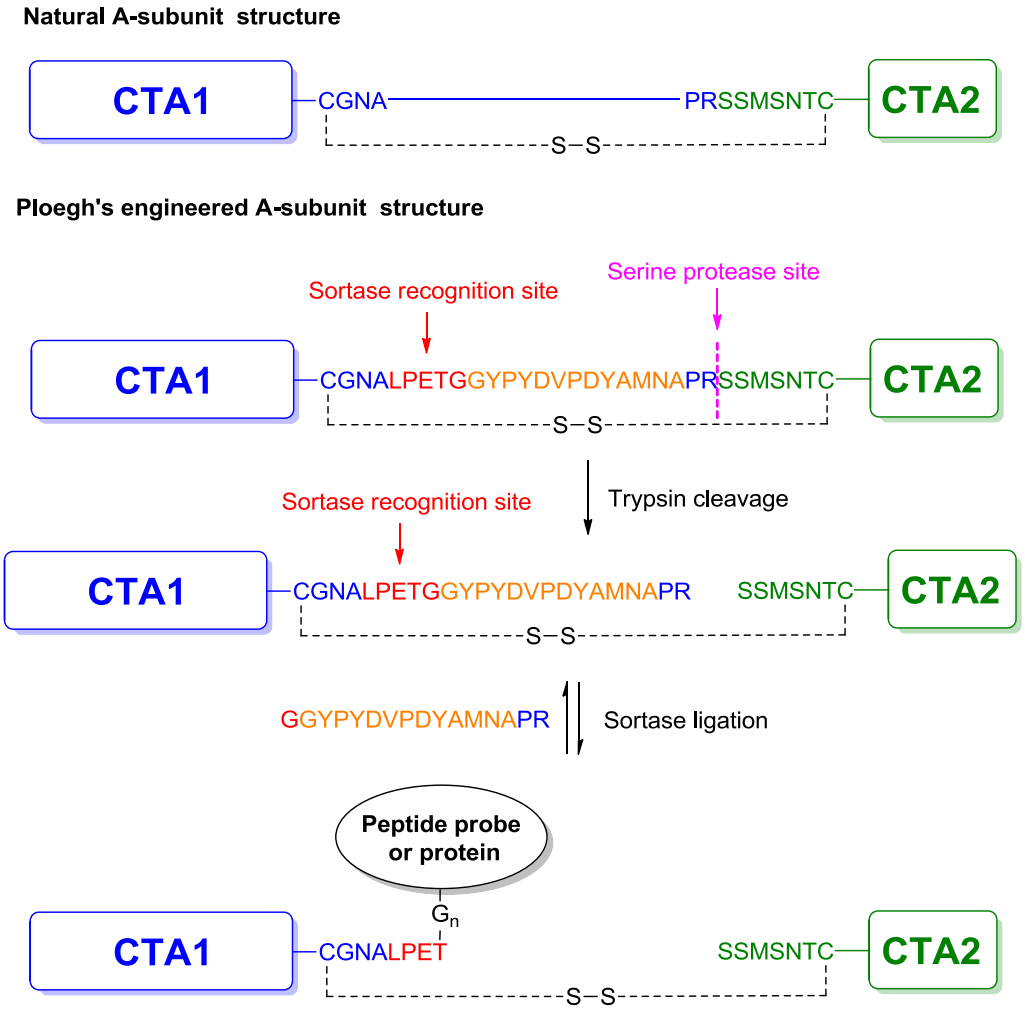


1.47 **1.48**

Scheme 1.7: N-terminal labelling of CTB using SrtA-mediated ligations. n = 3 or 5. Probe = FITC and biotin

Modification of AB_5 proteins

Ploegh and co-workers modified a fully functioning variant of cholera toxin using sortase-mediated ligations.[60] The authors genetically incorporated an extended peptide sequence including a SrtA recognition site at the *C*-terminus of the A1-domain directly before the final proline and serine residues (Scheme 1.8). The AB_5 protein was primed for transpeptidation when the peptide sequence connecting the A1 and A2-domains was cleaved using trypsin. SrtA was then used to chemically modify the *C*-terminus of the A1-domain with a range of biological probes including a fluorescent and biotinylated peptide. Vero cells were transfected with the modified AB_5 proteins and the intracellular fate of the A1-fragment was investigated. The biological probes provided direct evidence of the A1-protein reaching the ER supporting previous studies.[60,125] Using the same approach, the group ligated the catalytic domain of diphtheria toxin onto the A1-protein.[60] The essential genes required for cholera toxin transfection in human cells were screened using the AB_5-fusion protein. However, the most interesting aspect of the study was the observation that the modified A1-protein was still able to escape the ER and deliver the diphtheria toxin fragment to the cytosol of the cell. Theoretically, this methodology could be used to transport an assortment of macromolecules into human cells; however, the cytotoxicity of the A1-fragment is an enormous obstacle in the wider application of this approach. A solution to this problem would be to create a non-toxic A1-protein, but doing so without affecting the ER escape mechanism may be challenging.



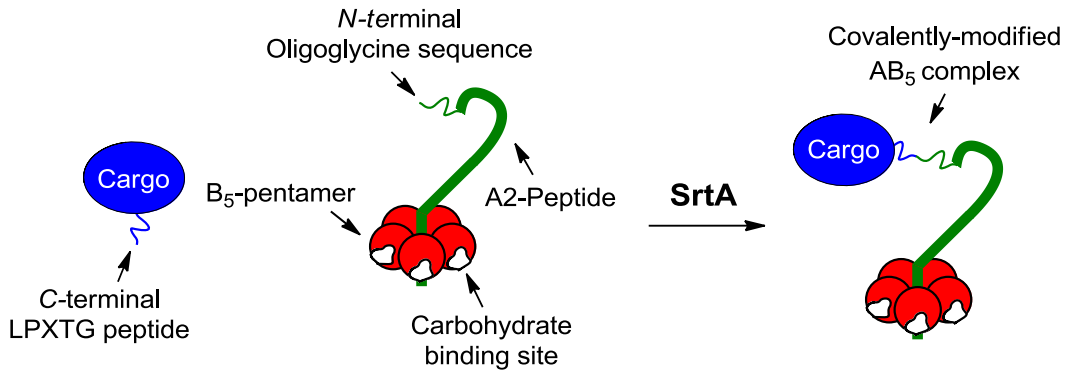
Scheme 1.8: Schematic representation of the engineered CTA variant and modification strategy developed by Ploegh.[60] Blue: natural CTA1 sequence. Red: sortase recognition site. Orange: extended peptide sequence. Green: natural CTA2 sequence.

Introduction part II

1.4 Project outline

The project was an investigation into the formation of bacterial toxin based delivery vehicles that could be used to transport oligonucleotides and other macromolecules inside mammalian cells. Studies discussed in the previous section have already demonstrated the ability of bacterial toxins to deliver chemical probes, proteins and nucleic acids into a range of mammalian cells. However, a simple and robust strategy has not yet been developed to allow the efficient formation of an oligonucleotide-delivery vehicle. The progress towards realising this goal is documented within the work presented here.

The primary objective was to use SrtA-mediated ligations to site-specifically modify a truncated cholera toxin variant at the *N*-terminus of the CTA2-peptide (Scheme 1.9). In general, SrtA ligations need a large excess of labelling regent and stoichiometric quantities of enzyme to achieve high levels of protein modification.[126,127] Therefore, initial research was focused on optimising the efficiency of the SrtA-transpeptidation reaction for *N*-terminal labelling of proteins. To this end, the incorporation of non-natural amino acids and chemical modification of the peptide label was investigated.



Scheme 1.9: AB_5 modification strategy using SrtA-mediated ligations

The second objective was to deliver a modified AB_5 complex into mammalian cells and track its intracellular fate (Figure 1.10). Using a combination of immunostaining and fluorescence microscopy, the pathway of a fluorescently labelled AB_5 complex was studied.

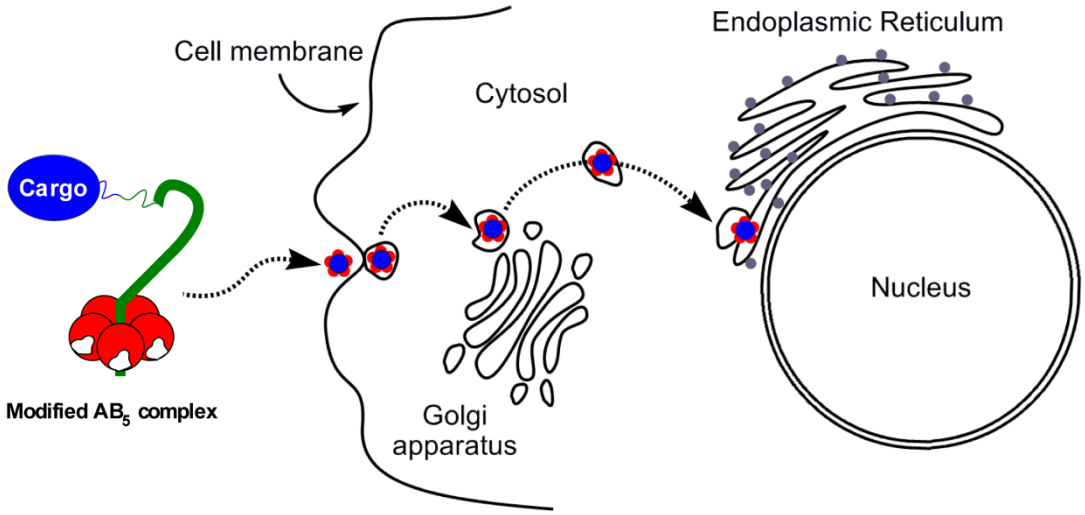
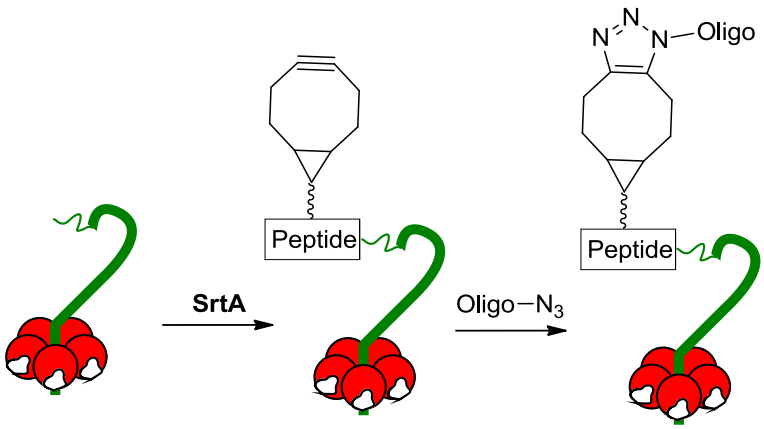


Figure 1.10: Diagram of the predicted intracellular pathway of the modified AB_5 complex

The final objective was to develop a method to attach oligonucleotides site-specifically onto the AB_5 complex (Scheme 1.10). The incorporation of a bioorthogonal linker into peptide substrates for SrtA was investigated. This strategy would allow the site-specific labelling of the AB_5 complex with a functional handle that could be subsequently derivatised with an oligonucleotide macromolecule.



Scheme 1.10: AB_5 modification strategy using SrtA-mediated ligations and a bioorthogonal linker

Chapter 2: Efficient *N*-terminal labelling of proteins using Sortase

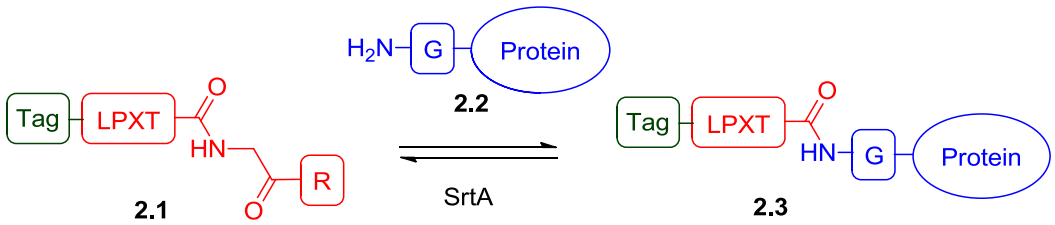
Two papers based on the following section have been published:

"Efficient N-Terminal Labeling of Proteins by Use of Sortase" Angewandte Chemie International Edition, 2012, volume 51, pages 9377-9380, D. J. Williamson, M. A. Fascione, M. E. Webb and W. B. Turnbull.

"Depsipeptide substrates for sortase-mediated N-terminal protein ligation" D. J. Williamson, M. E. Webb, W. B. Turnbull, Nature protocols, 2014, volume 9, pages 253-262.

2.1 Introduction

SrtA can be used to attach a wide variety of peptide labels **2.1**, which carry an LPXTG recognition sequence, onto a protein **2.2** through a sterically accessible *N*-terminal glycine residue (Scheme 2.1).[111,127] In principle, any molecule that can be coupled onto a peptide substrate can be attached to a protein. Furthermore, the conditions of the reaction are extremely mild typically being performed at pH 7.5 in the presence of a buffering agent, such as HEPES, TRIS etc.[111] The elegant simplicity of the SrtA-transpeptidation reaction makes it a powerful protein-modification tool.



Scheme 2.1: General *N*-terminal labelling of proteins using SrtA. X= any amino acid, R= amino acid or NH_2

The reaction is reversible, so many protein modification protocols recommend the use of a large excess of label and stoichiometric amounts of StrA to ensure high ligation yields.[111,124,127] However, if the label is precious, or available in finite quantities, this may not always be possible. In this chapter, the development of a method to increase the efficiency of the transpeptidation reaction using depsipeptide substrates for SrtA will be described.

2.2 Investigation into the efficiency of the SrtA-mediated ligation reaction

There is an abundance of literature that demonstrates the SrtA-mediated transpeptidation reaction as a versatile protein modification technique. Therefore, to allow a comparison to previous studies and gain an insight into the general efficiency of the ligation reaction, the activity of SrtA with model peptide substrates was investigated.

2.2.1 SrtA-mediated ligation with model peptide substrates

A pETMCS-III plasmid harbouring a SrtA gene with an *N*-terminal truncation, which removed the membrane-anchoring peptide sequence, was donated by Dr C. Neylon (University of Southampton). However, the plasmid construct encoded SrtA without a purification tag. To introduce a polyhistidine-tag (His-tag) to the *N*-terminus of the enzyme, the SrtA gene was

subcloned from pETMCS-III into pET28a using restriction enzymes EcoRI and NdeI (Figure 2.1). Then *E. coli* BL21-Gold (DE3) cells were transformed using the subcloned pET28a plasmid to allow SrtA to be overexpressed and subsequently purified by nickel chelation and size-exclusion chromatographies.

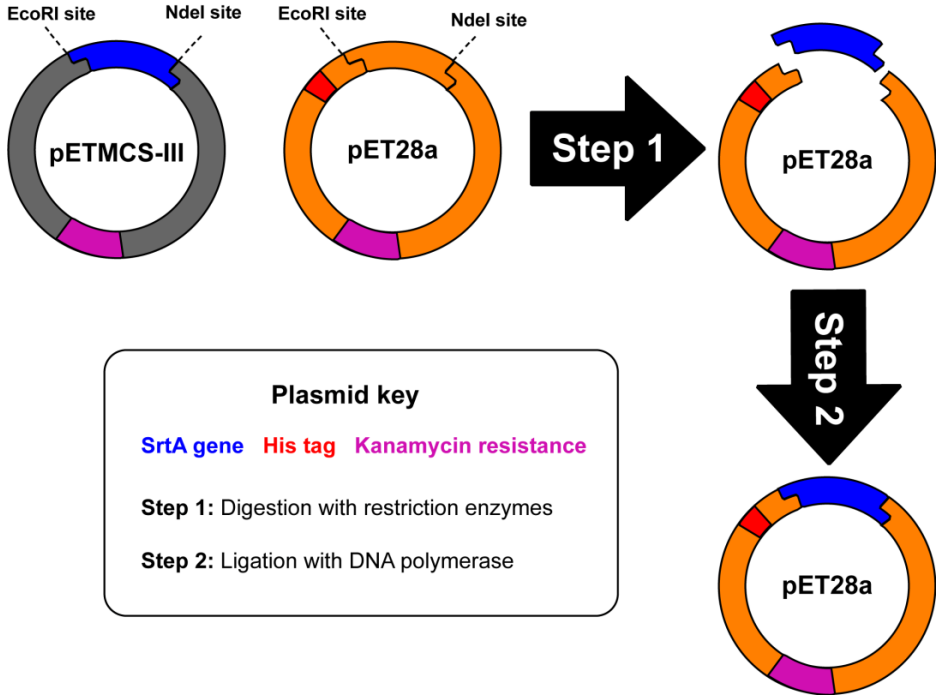
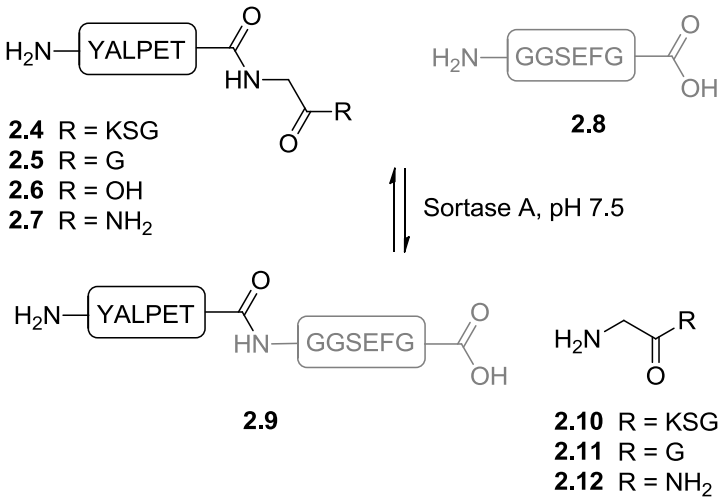


Figure 2.1: Schematic representation of the subcloning of the SrtA gene from pETMCS-III into a pET28a plasmid using restriction digests.

The reactivity of the SrtA mediated ligation reaction was probed with a simple assay; a range of acyl donors **2.4, 2.5** and **2.6** and a single acyl acceptor peptide **2.8** (Scheme 2.1) were synthesised using SPPS in quantitative, or near quantitative yields. The first acyl donor **2.4** was built as a positive control because its structure was based on a compound that had previously shown activity with SrtA.[128] The other peptides **2.5** and **2.6** maintained the common YALPET sequence of the control substrate **2.4**, but the *C*-terminus of both peptides were sequentially truncated to determine the minimum residue length required for SrtA activity. The ability of SrtA to ligate the acyl donor peptides **2.4, 2.5** and **2.6** to a simple acyl acceptor peptide **2.8**, containing an *N*-terminal diglycine motif, was monitored by LCMS (Scheme 2.1).



Scheme 2.1: Sortase mediated ligation of acyl-acceptor peptide **2.8** to acyl-donor peptides **2.4-2.5** forming product **2.9** and by-products **2.10-2.12**

Control peptide **2.4** was successfully ligated onto peptide **2.8** after an overnight incubation with SrtA. LCMS analysis before and after incubation allowed the identification of a new m/z peak with a UV absorbance that corresponded to ligated product **2.9**, thus confirming the reaction was a success. Under the same reaction conditions, peptide **2.5**, which contained a diglycine sequence at its *C*-terminus, showed similar reactivity to control peptide **2.4**. Interestingly, no ligation was observed when the assay was repeated with peptide **2.6**, which had only the minimal LPETG sequence at its *C*-terminus. However, once the pKa and position of the *C*-terminal carboxylic acid were considered, the result was not surprising. Under the conditions of the reaction, the *C*-terminal carboxylic acid will be deprotonated creating an anionic charge at the SrtA recognition site, which could potentially inhibit substrate-enzyme binding. We proposed that replacing the carboxylic acid with a primary amide would remove the anionic charge and allow SrtA to bind to the substrate. Therefore, a fifth peptide **2.7** containing a *C*-terminal amide was synthesised and successfully attached to peptide **2.8** (Scheme 2.1). These observations support previous findings reported by Huang *et al*.[116] The authors measured the rate of ligation between a SrtA-acyl intermediate and a series of acyl acceptors. When glycine was used as a substrate, no reaction was observed. If glycine cannot participate in the reversible SrtA transpeptidation reaction, it is reasonable to assume that corresponding peptides bearing a *C*-terminal glycine residue in the LPXTG motif will also be unsuitable substrates. In addition, the authors observed SrtA activity when glycine amide was used as the acyl acceptor instead of glycine, which vindicates our strategy of replacing the *C*-terminal carboxylic acid of acyl donor **2.8** with an amide bond to restore SrtA reactivity.

2.2.2 HPLC time-course analysis of the SrtA-mediated ligation reaction with peptide substrates

The initial experiments with peptides **2.4-2.7** had confirmed the general substrate selectivity and reactivity of SrtA that had already been widely reported in the literature. However, in each assay where ligation had been successful there was a significant amount of starting material identified in the UV chromatogram and it was impossible to determine if or when the reaction had reached completion. Therefore, an HPLC time-course experiment was used to establish the point at which there was no further reaction. The ligated product **2.9** was synthesised independently and analysed by HPLC to determine its retention time, so that there was no ambiguity about what was being formed in the reaction. Acyl donor **2.5** (500 μM) was incubated with one, two and four equivalents of acyl acceptor **2.8** in three separate assays. The assays were carried out in parallel over an 8 hour period and the amount of product **2.9** formed was plotted as a function of time (Figure 2.2)

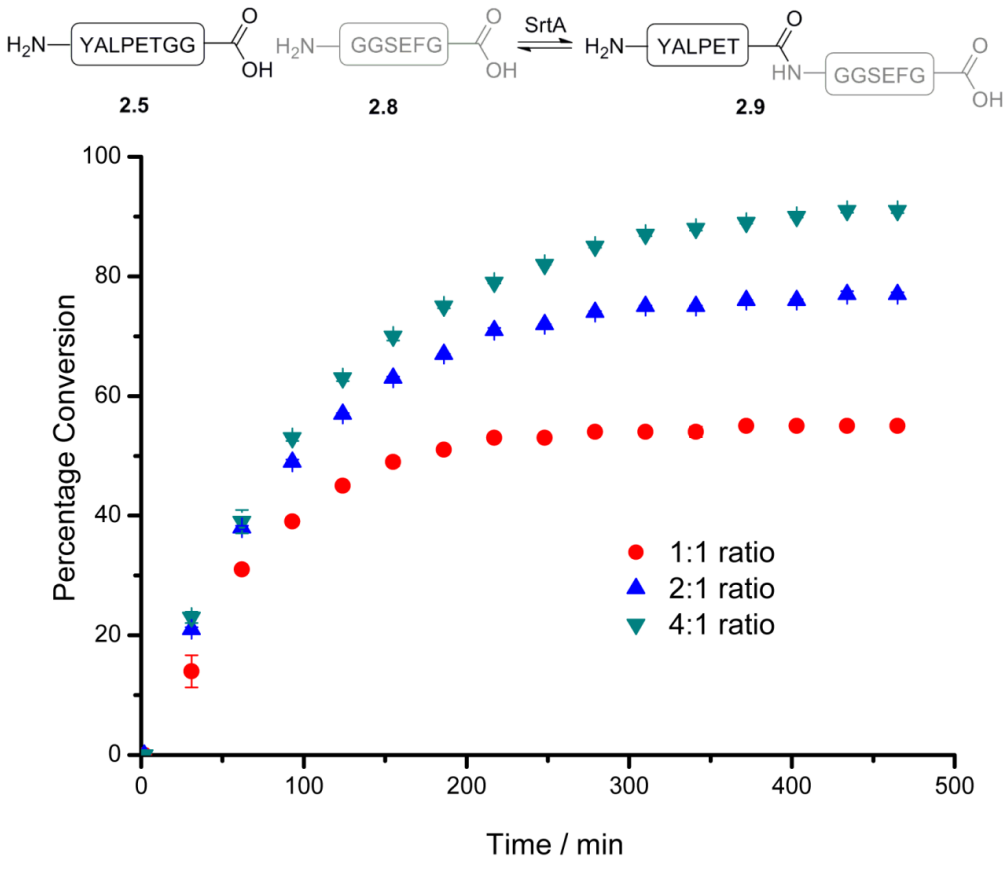


Figure 2.2: Time-course of StrA-mediated ligation of peptides **2.5** (500 μM) and **2.8** (0.5, 1 or 4 mM) forming product **2.9**. The ratio of peptides **2.8** and **2.5** was increased for each experiment, as indicated: Red: 1:1, Blue: 2:1 and Green 4:1 Error bars are present, but in some cases smaller

When peptides **2.5** and **2.8** were mixed in equal concentrations, the time-course showed the reaction reached completion in roughly four hours with ~50% product conversion (Figure 2.2). This is probably due to the reaction reaching equilibrium as it is well documented that the cleaved diglycine residue **2.12** can act as substrate and participate in the reverse reaction.[116] Predictably, increasing the concentration of acyl acceptor **2.8** pushes the equilibrium towards greater product formation, but even with four equivalents there was still some acyl donor **2.5** left at the end of the assay. The time-course experiment was repeated for acyl acceptor **2.8** and acyl donor **2.7**, which had a *C*-terminal amide group (Figure 2.3). The results of the second time-course experiment followed an almost identical pattern to the first experiment: when the substrates were mixed in a 1:1 ratio, the ligation reaction reached about 50% product conversion before levelling off, and increasing the concentration of acyl acceptor **2.8** maximised the product yield. These observations are consistent with the current literature precedent regarding SrtA-mediated ligations.[114,116,117,128] The formation of a 100% ligated product is probably unachievable using SrtA-mediated ligations as the reverse reaction will still be operational even at much higher concentration of one of the substrates.

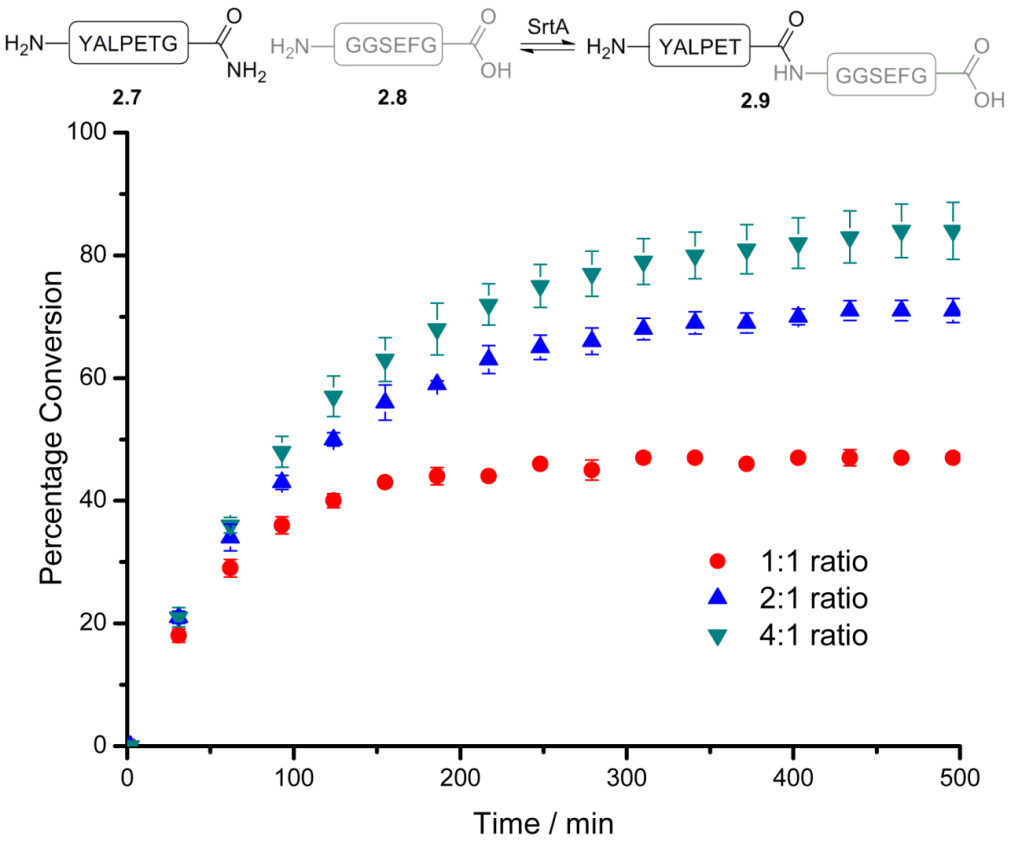
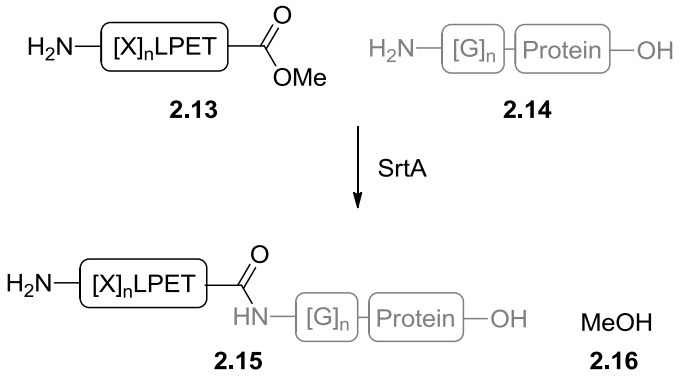


Figure 2.3:Time-course of StrA-mediated ligation of peptides **2.7** (500 μM) and **2.8** (0.5, 1 or 4 mM) forming product **2.9**. The ratio of peptides **2.7** to **2.8** was increased for each experiment, as indicated: Red: 1:1, Blue: 2:1 and Green 4:1

2.3 Optimization of the SrtA mediated ligation reaction

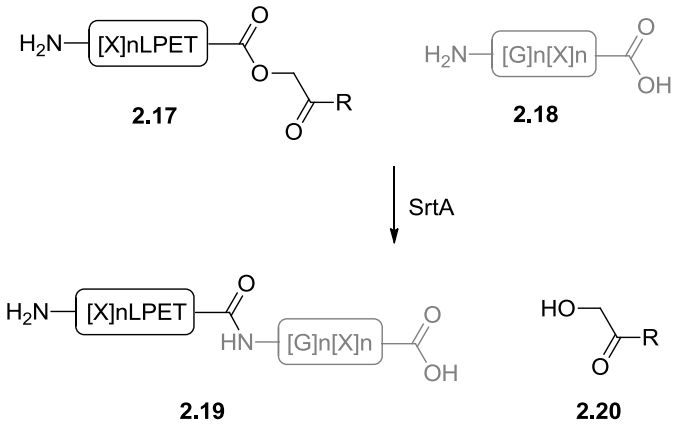
The previous time-course experiments had confirmed the major limitation of the SrtA ligation reaction. High yields can only be achieved when a large excess of one substrate is used, which may not always be possible if only a limited amount of compound is available. Furthermore, future purification steps could be problematic as the ligation reaction did not give complete product formation. An obvious solution to these problems is to try and prevent the reverse ligation reaction, which theoretically could maximise the product yield and obviate the need for excess reagents. Previous work by Ploegh and co workers had endeavoured to tackle the inefficiencies of the SrtA ligation reaction.[124] The authors replaced the *C*-terminal glycine residue in the LPXTG motif with a methyl ester. SrtA continued to recognise the methyl ester **2.13** as a substrate and ligated it to proteins **2.14** bearing an *N*-terminal oligoglycine sequence (Scheme 2.2). The reaction now generated methanol as a byproduct, which is not a substrate for SrtA, so the ligation effectively became irreversible. However, a large excess (in some cases approximately ten equivalents) of methyl ester **2.13** and stoichiometric amounts of StrA were still required for complete conversion to the modified protein. It is probable the authors were forced to use a large excess of reagents and stoichiometric amounts of enzyme to overcome the possible product inhibition of SrtA. The methyl ester **2.13** was probably a poor substrate for SrtA as it did not contain the full LPXTG recognition motif. Therefore, as more ligated product accumulated during the course of the reaction, it is reasonable to assume it would have bound preferentially to SrtA inhibiting the forward ligation reaction.



Scheme 2.2: Sortase ligation of methyl ester **213** and generic protein **2.14** to form ligated product **2.15** and methanol **2.16** byproduct. X = any amino acid

We postulated that depsipeptide substrates **2.17**, which had the scissile amide bond between the threonine and glycine residues replaced with an ester linkage, would also avoid the reverse ligation reaction (Scheme 2.3). If SrtA accepted depsipeptide **2.17** as a substrate then hydroxyacetyl byproduct **2.20** would be generated and, like methanol, it should not be

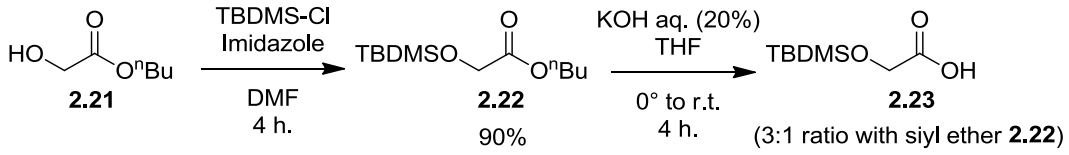
recognised by the enzyme. In principle, this approach should avoid the reverse reaction without reducing the rate of the forward reaction as depsipeptide substrate **2.17** would only contain minimal alterations to the LPXTG recognition sequence. Therefore, only substoichiometric amounts of enzyme and equal concentrations of both substrates should be required for efficient protein modification.



Scheme 2.3: Sortase ligation of depsipeptide **2.17** and generic oligoglycine peptide **2.18** forming ligated product **2.19** and hydroxyacetyl byproduct **2.20**

2.3.1 Initial Synthetic route to depsipeptide substrates

Replacing the glycine residue in the LPXTG sequence with glycolic acid was thought to be the simplest way to create a depsipeptide using SPPS. However, the hydroxyl group of the glycolic acid had to be protected before it could be used as a SPPS building block. The choice of protecting group had to be carefully considered as conventional resins used in Fmoc-SPPS are acid sensitive; this ruled out traditional acid labile projecting groups. After consultation of past literature, a synthetic strategy was found that would lead to silyl-protected glycolic acid **2.23** in two steps (Scheme 2.4).[129] The first step involved capping the hydroxyl group of butyl glycolate **2.21** with TBDMS-Cl in the presence of imidazole to give silyl ether **2.22** in 90% yield. The silyl ether **2.22** was then hydrolysed to form the protected glycolic acid **2.23** as the predominant product (~70% by ^{1}H NMR spectroscopy) in a mixture with starting material **2.22**. The remaining silyl ether **2.22** was not expected to cause complications with the SPPS, so the reaction mixture was used without any further purification.



Scheme 2.4: Synthetic route to silyl protected glycolic acid **2.23**

The silyl-protected glycolic acid **2.23** was now able to be used in the synthesis of a depsipeptide. The previous HPLC-time course investigations into the SrtA-mediated ligation reaction had used acyl donor peptides **2.5** and **2.7**, so it was necessary to synthesise the analogous depsipeptides **2.24** and **2.25** to allow a direct comparison of the two types of substrate (Figure 2.4).

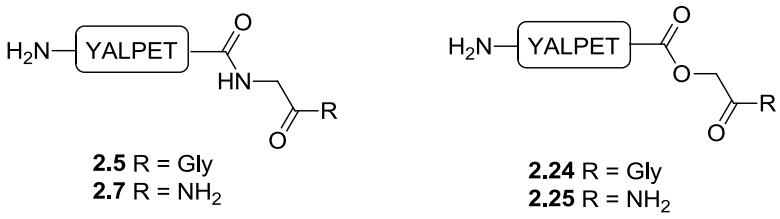


Figure 2.4: Structure of peptides **2. 5** and **2.7**; and depsipeptide analogues **2.24** and **2.25**

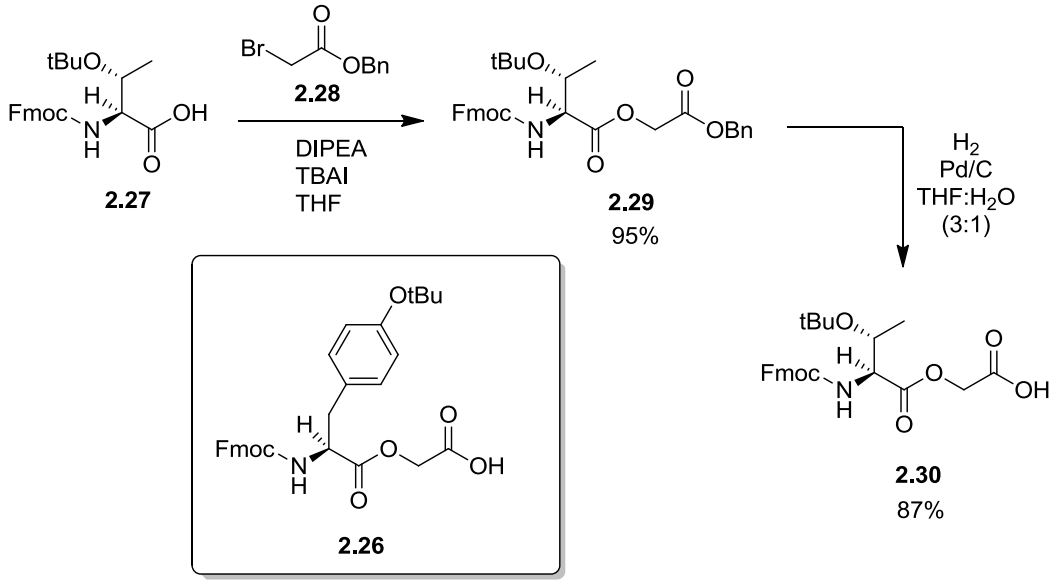
Using standard Fmoc-SPPS conditions, silyl-protected glycolic acid **2.23** was coupled to gly-2-chlorotrityl resin. The resin was treated with TBAF in THF to remove the silyl protecting groups and allow elongation of the depsipeptide through the hydroxyl group of the glycolic acid. Cleavage of the peptide gave the correct product, but with two significant byproducts: an incomplete peptide that did not include the ester linkage, and a depsipeptide that contained the ester linkage, but had not been elongated correctly. These two truncated compounds were indicative of incomplete coupling steps: the initial loading of silyl-protected glycolic acid **2.23** to the resin, and the coupling of the resin-linked glycolic acid to threonine. The SPPS was repeated and in an effort to avoid truncated peptide formation each of the challenging coupling steps were conducted twice to increase the loading efficiency. Regrettably, these extra steps proved ineffective as the reactions always led to a heterogeneous mixture of compounds. In addition, the silyl deprotection step had also been problematic; half of the TBAF treatments had not managed to remove the protecting group. After several failed attempts, it was decided to terminate this approach.

There were two major issues with the direct SPPS of depsipeptide substrates: the deprotection of the silyl-protecting group and the formation of the ester linkage. An obvious solution to the first problem would be to replace the silyl groups with an alternative SPPS protecting group, such as

Fmoc. Seeberger and co-workers use sugars with Fmoc-protected hydroxyl groups in the solid-phase synthesis of oligosaccharides.[130-132] However, this strategy does not address the issue of inefficient ester formation as the solid-phase synthesis of oligosaccharides uses activated glycosyl donors to link the sugars and not the standard SPPS-coupling reagents. It may have been possible to optimise the coupling of the resin-linked glycolic acid to theronine, but it was decided it would be more efficient to connect the residues in solution prior to SPPS. Therefore, an alternative strategy was developed

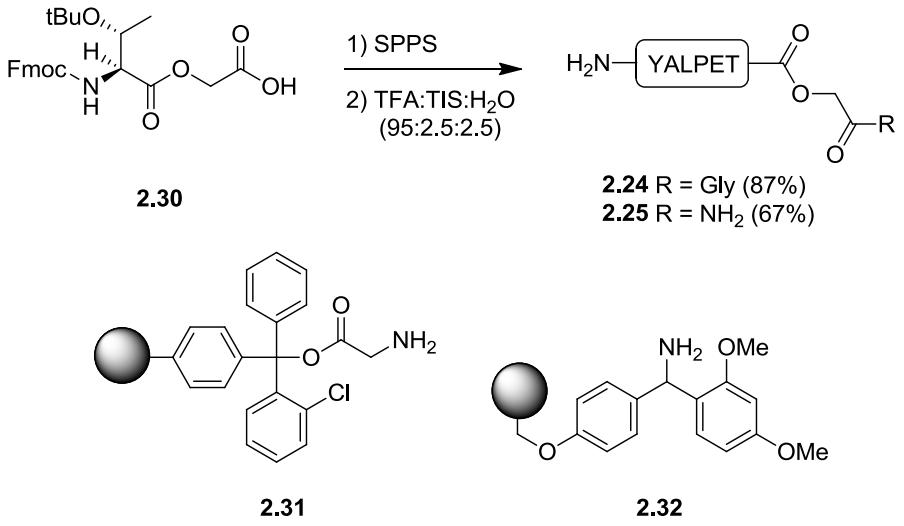
2.3.2 Improved synthetic route to depsipeptide substrates

The solution-phase synthesis of an Fmoc-protected depsipeptide precursor would allow the synthesis of depsipeptide **2.24** and **2.25** on solid support using the standard deprotection/elongation steps of Fmoc-SPPS. Suich *et al.* had already reported a two step synthesis to a protected tyrosine-glycolic acid conjugate **2.26** (Scheme 2.5).[133] Using a modified version of the authors' strategy, depsipeptide precursor **2.30** was synthesised. This was achieved by alkylating Fmoc-threonine **2.27** with 2-bromobenzyl glycolate **2.28** in the presence of tetrabutylammonium iodide (TBAI) to form tri-protected ester **2.29** in 95% yield (Scheme 2.5). After a selective deprotection of the benzyl ester by hydrogenation, depsipeptide precursor **2.30** was generated in 87% yield. This step, although relatively straightforward, had to be monitored carefully as the reaction conditions can remove the Fmoc-protecting group if left for long periods of time. It is also important to note that the crude product could be used without further purification if required.



Scheme 2.5: Synthesis of depsipeptide precursor **2.30.** Adapted from the original synthesis of tyrosine-glycolic acid conjugate **2.26** by Suich *et al.*[133]

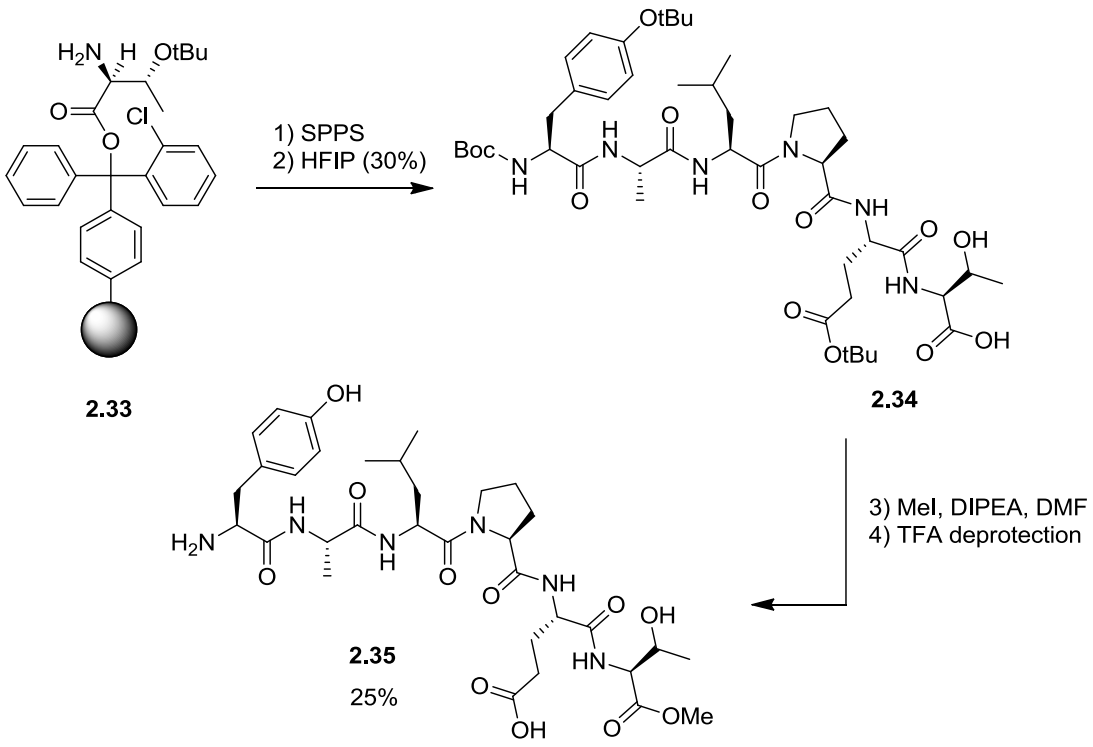
Depsipeptide precursor **2.30** was successfully used to build depsipeptides **2.24** and **2.25** using conventional SPPS (Scheme 2.6). Depsipeptide **2.24** was synthesised, in 87% yield, on preloaded gly-2-chlorotrityl resin **2.31** generating a *C*-terminal carboxylic acid. Depsipeptide **2.25**, which contained the *C*-terminal amide, was constructed on rink amide resin **2.32** with a slightly lower yield of 67%.



Scheme 2.6: Synthetic route to depsipeptide **2.24** and **2.25.** Structure of gly-2-chlorotrityl **2.31** and rink amide **2.32** used in their SPPS

2.3.3 Synthesis of methyl ester substrates for SrtA

The reactivities of depsipeptides **2.24** and **2.25** could now be compared to their peptide analogues **2.5** and **2.7**. However, it was also important to evaluate the reactivity of methyl ester variant **2.35**, which was analogous to that reported by Ploegh,[124] as it would allow a direct comparison of both alternative ester-linked substrates under the same reaction conditions (Scheme 2.7). The synthesis of methyl ester **2.35** was achieved by constructing a partially protected peptide **2.34** on Thr-2-chlorotrityl resin **2.33** before cleaving the peptide with the very mild acid, hexafluoroisopropanol (HFIP). The use of HFIP allows peptide cleavage from 2-chlorotrityl resins, but without the loss of protecting groups from the amino acid side chains. The partially protected peptide **2.34** was then methylated with methyl iodide in the presence of DIPEA. The crude methyl ester intimidate was treated with a standard TFA cleavage cocktail to generate methyl ester **2.35** in a modest 25% yield.



Scheme 2.7: Synthetic route of methyl ester **2.35**

2.3.4 Comparison of SrtA reactivity with peptide, depsipeptide and methyl ester substrates

The SrtA mediated ligation of depsipeptide **2.24** to acyl acceptor peptide **2.8** was monitored by a HPLC time-course assay (Figure 2.5). When one equivalent of depsipeptide **2.24** was used, peptide **2.8** was almost completely transformed into ligated product **2.9** in approximately 4 hours**.** This was a huge improvement on the amide-linked substrate **2.5**, which only reached ~50% product conversion in the same time period. In contrast, methyl ester **2.35** reacted significantly slower than both depsipeptide **2.24** and peptide **2.5** to give only ~30% of ligated product **2.9** during the entire assay.

Depsipeptide **2.24** appeared to slowly hydrolyse under the conditions of the reaction, which probably prevented complete product formation. Furthermore, as the product would still be in equilibrium with the SrtA-linked thioester intermediate, the 2 mol% of enzyme would also reduce the maximum product yield. Therefore, the experiment was repeated with 1.5 equivalents of depsipeptide **2.24**; the small increase in the number of equivalents allowed the reaction to reach completion rapidly. However, some enzymatic hydrolysis of the product was observed if the ligation was left for long periods of time; e.g. overnight. This is expected as it is

well documented that SrtA will catalyse the hydrolysis of acyl donor substrates if there is no acyl acceptor present.[103,116]

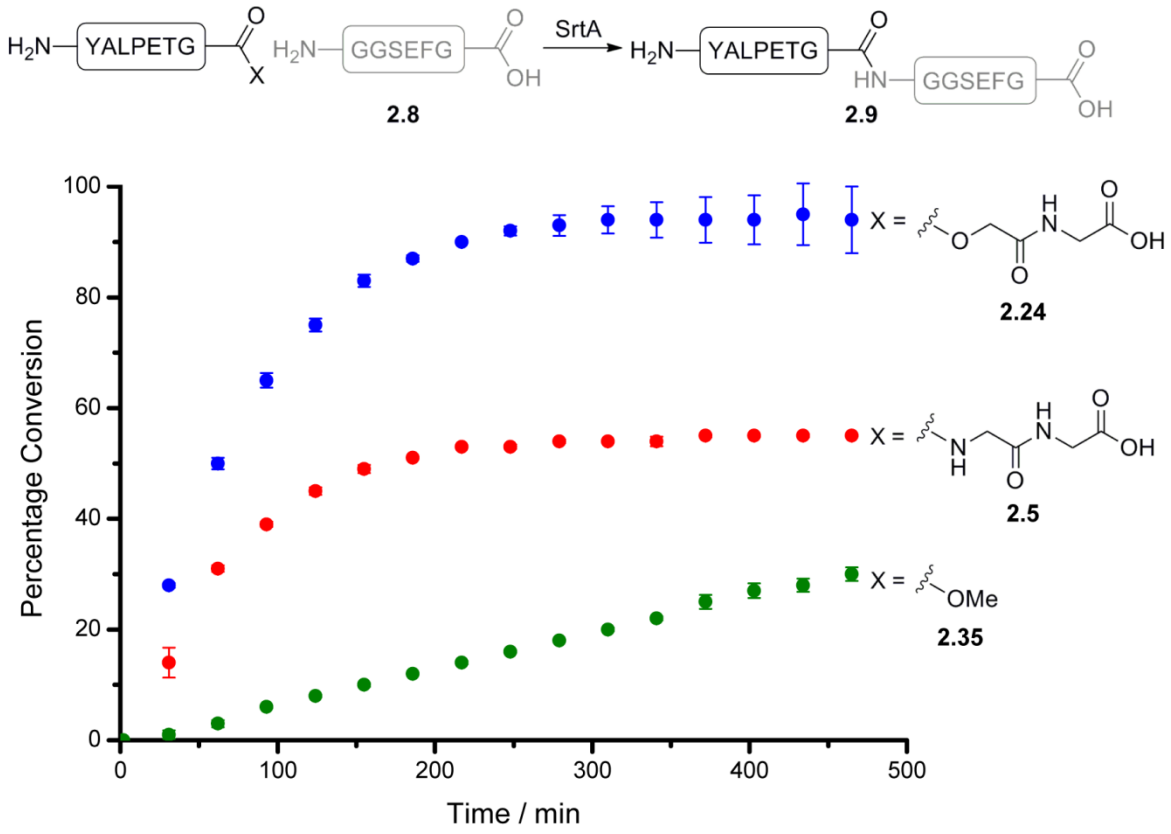


Figure 2.5: Time-course of the SrtA-mediated ligation of peptide **2.5**, depsipeptide **2.24** and methyl ester **2.35** to peptide **2.8** forming ligated product **2.9**. The acyl donor substrates **2.5**, **2.24** and **2.36** (500 μM) and acyl acceptor peptide **2.8** (500 μM) were mixed in a 1:1 ratio and monitored over an 8 period.

We propose that depsipeptide **2.24** was a much more efficient substrate for the SrtA-mediated ligation reaction than methyl ester **2.35** because it was able to bind efficiently to the enzyme and rapidly form ligated product **2.9**. In contrast, the truncated SrtA-recognition site in methyl ester **2.35** was predicted to cause inefficient enzyme-substrate binding and as a consequence the rate of the ligation was significantly reduced. Therefore, even though both ester-linked substrates produce non-reactive alcohol byproducts that cannot participate in the reverse reaction, only depsipeptide **2.24** was turned over efficiently to produce a high level of ligated product **2.9**. Furthermore, it is unlikely that methyl ester **2.35** would have been completely converted to ligated product **2.9** if the reaction had been left for a longer period of time. It was more likely, however, that the reaction would have started to slow down before eventually stopping as a result of product inhibition. This hypothesis seems plausible because ligated product **2.9** carried the full recognition sequence, so it would have presumably bound more tightly to SrtA than methyl ester **2.35** and as a result the forward reaction would be inhibited. This could explain

why Ploegh had to use stoichiometric amounts of SrtA and a large excess of methyl ester to achieve complete protein modification.[124]

The ligation reaction was repeated using amide-terminated depsipeptide **2.25**, which was almost quantitatively converted to product **2.9** (Figure 2.6). Once again, the reaction could be pushed to completion with the addition of 1.5 equivalents of depsipeptide **2.25**. Intriguingly, the formation of ligated product **2.9** appeared to be more rapid than the ligation with depsipeptide **2.24**; ~95% yield was achieved in just over 2 hours compared to the 4 hours for depsipeptide **2.24**. However, once the maximum yield had been obtained the amount of observed product declined to ~90% after 8 hours. The reasons for this are unclear, but SrtA mediated hydrolysis of the product is one possibility.

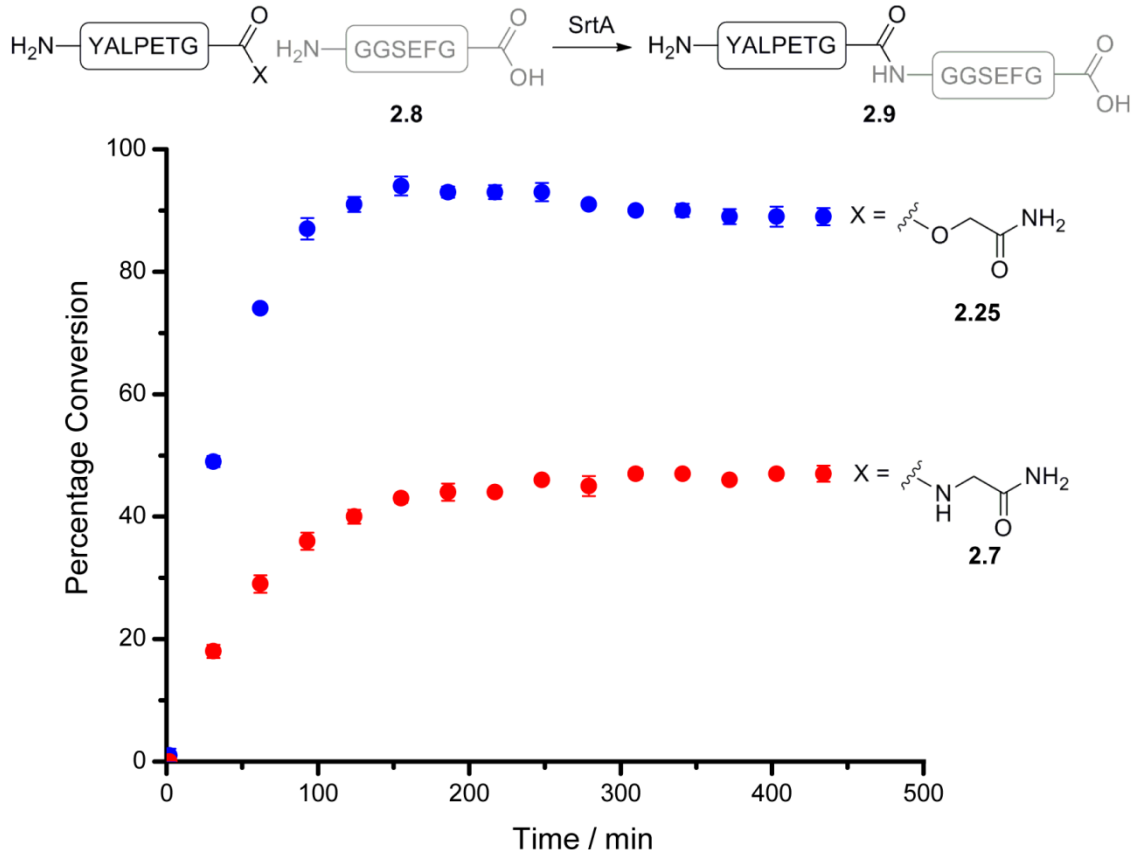


Figure 2.6: Time-course of the SrtA-mediated ligation of peptide **2.7** (500 μM) and depsipeptide **2.25** (500 μM) to peptide **2.8** (500 μM) forming ligated product **2.9**.

2.4 Protein modification using optimised SrtA-mediated ligations

The initial time-course investigations had proved conclusively that depsipeptide substrates can be used to transform a model acyl donor peptide into the corresponding ligated product in near

quantitative yields. However, it was less clear how the new system would perform when a protein was used as the acyl acceptor rather than a model peptide. This was a more challenging task as proteins are larger molecules with less flexible *N*-termini that may not be freely accessible to SrtA. Furthermore, it was unlikely that protein substrates would be available in the same quantities as the model peptide substrates, so the transpeptidation reaction would have to be performed at much lower concentrations. Therefore, the rate of the reaction may be significantly reduced, which could allow the non-enzymatic hydrolysis of the depsipeptide substrate to start competing with the ligation step.

Three proteins bearing an *N*-terminal glycine residue were identified as potential substrates. A construct for a variant of human mannose binding lectin (hMBL) was kindly provided by Prof. K. Drickamer (Imperial Collage London). Myoglobin, from equine heart, was purchased from Sigma. The puf domain of mouse pumilio-2 was donated by Dr. T.A. Edwards (University of Leeds). The pumilio protein had been treated with tobacco etch virus (TEV) protease to remove its purification tag and leave an *N*-terminal glycine residue.

It is important to note that each of the three proteins selected for the ligation assays contained only a single *N*-terminal glycine residue. This is significant because it has been previously demonstrated that the rate of SrtA-mediated ligation is maximised when there are two or more glycine residues present.[116,117] Therefore, the limits of the optimised ligation reaction were thoroughly tested with the choice of proteins.

2.4.1 Human mannose binding lectin

hMBL plays an important role in the innate immune system defence against bacterial pathogens.[134-136] hMBL is found in blood serum where it binds to complementary glycoproteins on the surface of yeast or bacteria inducing an immune response. [135,136] The hMBL is a trimeric protein that has a triple helical coiled-coil central structure; the *N*-terminal amino acid residues are located at the apex of the coiled-coil (Figure 2.7). The construct provided by Prof. K. Drickamer encoded a variant of hMBL, which has an active site mutation that changes the binding selectivity of the protein from mannose to galactose. However, the parent structure of the protein is essentially the same as native hMBL apart from a short peptide extension, GDSSL, at the *N*-terminus.

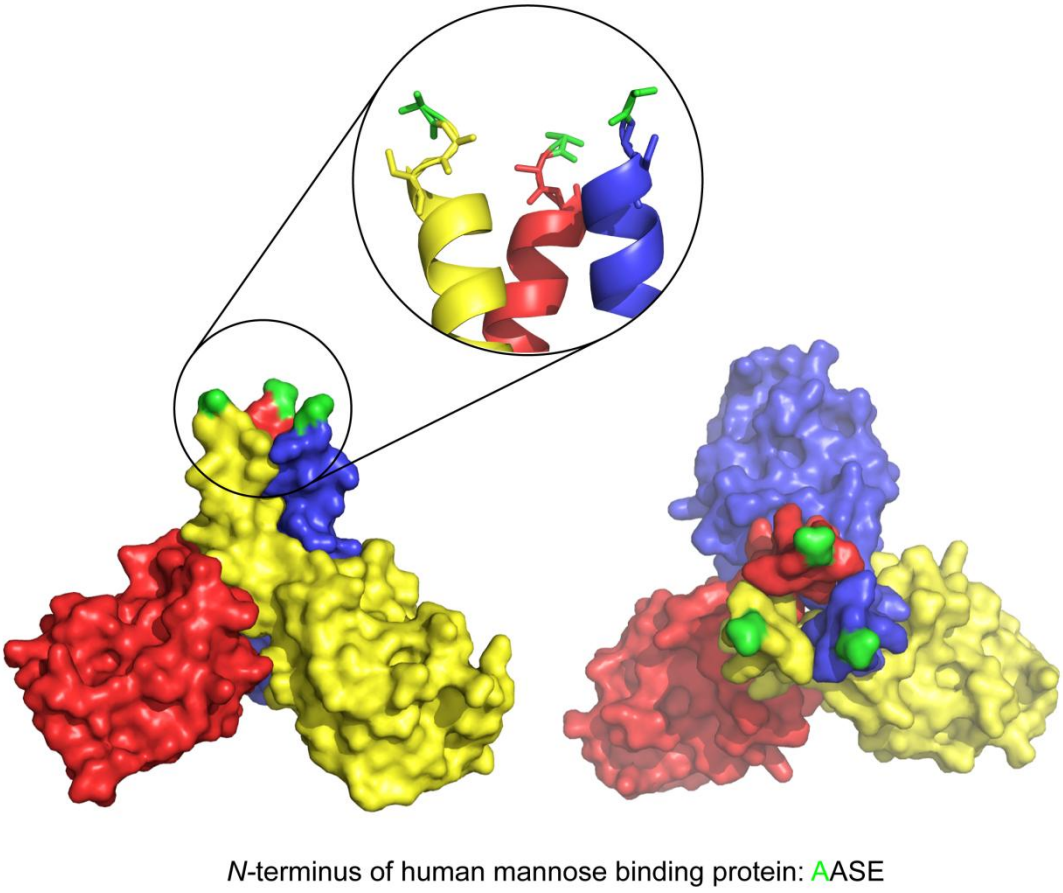


Figure 2.7: Crystal structure of human mannose binding lectin. Red, blue and yellow = monomeric protein. Green = *N*-terminus of protein. Generated from PDB code of atomic coordinates: 1HUP (UniProt code: P11226).[180]

The hMBL variant was isolated using the overexpression and purification methods described in the experimental chapter. The initial protein modification assays were performed using fluorescent depsipeptide **2.36**, which was synthesised in an 89% yield (Figure 2.8). The *C*-terminal glycine depsipeptide **2.24** was chosen as the template for dansyl-depsipeptide **2.36** because product degradation had not been observed during the HLPC time course experiment (Figure 2.4).

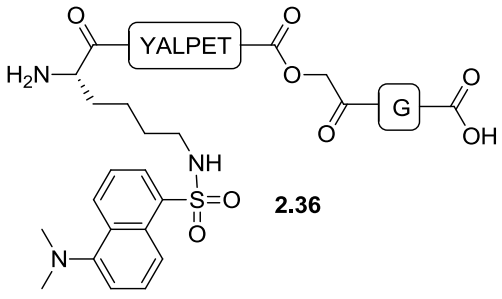


Figure 2.8: Structure of dansyl-depsipeptide **2.36**

The *N*-termini of hMBL are located in close proximity to each other (Figure 2.7) and even though the variant protein has an *N*-terminal peptide extension it was not clear if steric effects would reduce the efficiency of the SrtA-mediated ligation reaction. In addition, each monomeric protein only contained a single *N*-terminal glycine residue, which, as previously discussed, is not ideal for SrtA recognition. Therefore, the initial modification experiments screened hMBL at 60 μM with increasing concentrations of dansyl-depsipeptide **2.36** to establish an optimum depsipeptide ratio for effective protein ligation. The total concentration of substrates used in the ligation reaction was significantly lower than that used in the previous model peptide studies and as a consequence the rate of the reaction was predicted to be slower. Thus, the concentration of SrtA was increased from 2 mol% to 10 mol% to maintain the relative reactivity of the transpeptidation reaction. After 4 hours incubation, SDS-PAGE was used to analyse the reaction mixtures. When the polyacrylamide gel was irradiated with ultraviolet light, the labelled hMBL could be seen as a discrete fluorescent band (Figure 2.9). The depsipeptide **2.36** starting material and more surprisingly the thioester-SrtA intermediate were also visible as separate fluorescent bands. The modified hMBL could also be seen as a slightly elevated band once the gel had been stained with Coomassie blue and the proteins visualised. The preliminary assays had identified 1.5 equivalents of depsipeptide as the minimal substrate concentration for efficient protein ligation. However, the modified and unmodified hMBL bands ran very close to one another on the polyacrylamide gel, so it was difficult to determine the level of protein labelling. Therefore, the experiments were repeated and mass spectrometry was used to follow the progress of the reaction as the analytical technique is highly sensitive. Mass spectrometry identified modified protein and could not detect the presence of any unlabelled protein when 1.5, 3 and 5 equivalents of dansyl-depsipeptide **2.36** were used the ligation reaction (Figure 2.9). These results suggest that a high level of protein modification was achieved with only a minimal excess of depsipeptide label. However, product hydrolysis was observed when the ligation reactions were left overnight. The level of protein hydrolysis was less apparent in the samples that had a greater concentration of depsipeptide label as it would have taken longer for SrtA to turn over the excess substrate.

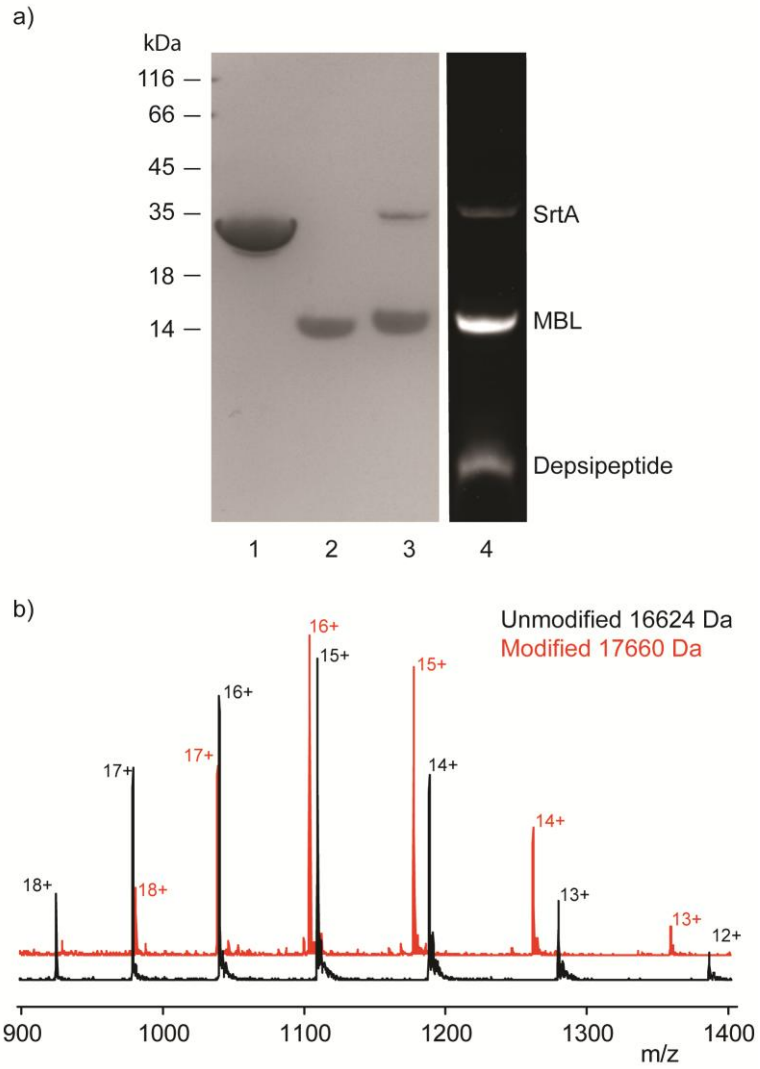


Figure 2.9: SrtA-mediated ligation of dansyl-depsipeptide **2.36** (90 μM, 1.5 equivalents) to hMBL (60 μM) after 4 hours incubation with SrtA (10 mol %) at 37ºC. a) Polyacrylamide gel (15%) of the reaction mixture: Lane 1: SrtA, lane 2: hMBL, lane 3: Reaction mixture, lane 4: UV irradiation of reaction mixture. b) Overlaid mass spectra of modified and unmodified hMBL. No starting material was detected by mass spectrum analysis after 4 hours incubation. Calculated mass change after modification, 1036 Da - observed mass change, 1035 Da.

The previous ligation experiments had used depsipeptide substrates that all contained a common YALPETG sequence; however, only the LPETG motif is essential for SrtA activity. Therefore, a second fluorescein-depsipeptide **2.37** (Figure 2.10) with a minimal LPETG sequence linked to a fluorescein molecule was synthesised; the compound was isolated in a modest 52% yield.

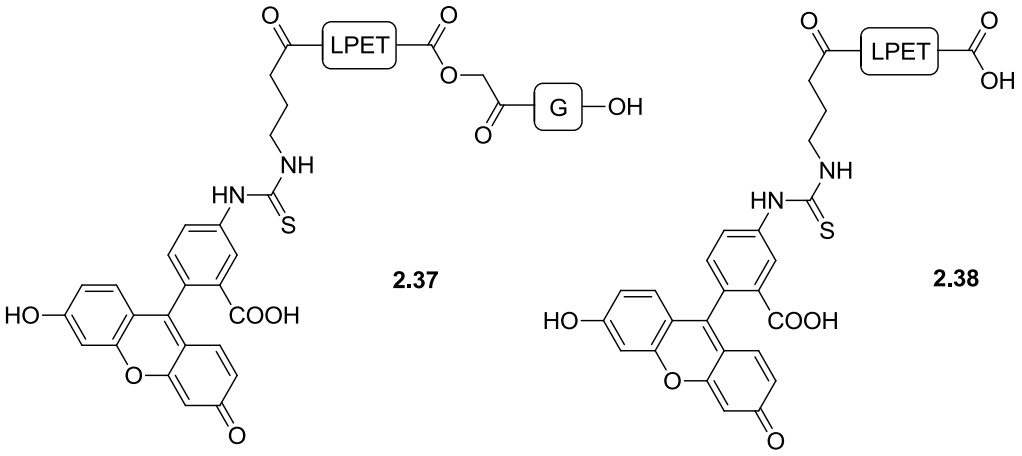


Figure 2.10: Structure of fluorescein-depsipeptide **2.37** and hydrolysed peptide **2.38**

It had already been demonstrated that hMBL could be labelled in high yields using just 1.5 equivalents of dansyl-depsipeptide **2.36**, so the protein-modification procedure was repeated with truncated fluorescein-depsipeptide **2.38**. Unfortunately, the same results could not be reproduced and only partial labelling of the protein was observed. Mass spectrometry analysis showed that fluorescein-depsipeptide **2.37** had hydrolysed to form peptide **2.38** before complete protein modification was detected. Although, depsipeptide hydrolysis had been observed in previous experiments it had not been this rapid. Consequently, the ligation assay was repeated with two equivalents of fluorescein-depsipeptide **2.37** to ensure there was enough available substrate to allow the transpeptidation reaction to proceed efficiently. However, after 4 hours incubation, fluorescein-depsipeptide **2.37** had hydrolysed again before hMBL had been sufficiently labelled, but extensive protein modification was observed when an extra equivalent of depsipeptide was added to the ligation mixture and the incubation time was extended by 1.5 hours (Figure 2.11a). There were still a few small signals in the mass spectrum that corresponded to unmodified protein; therefore, SDS-PAGE analysis was performed to elucidate the extent of the protein labelling (Figure 2.11b). The modified hMBL could be seen as a discrete band that had shifted up the gel from the position of the unlabelled protein. There was also a small quantity of unlabelled protein (approx. <10%) in the reaction mixture, but it was quite clear that the majority of hMBL had been modified, which is in good agreement with the mass spectrometry analysis.

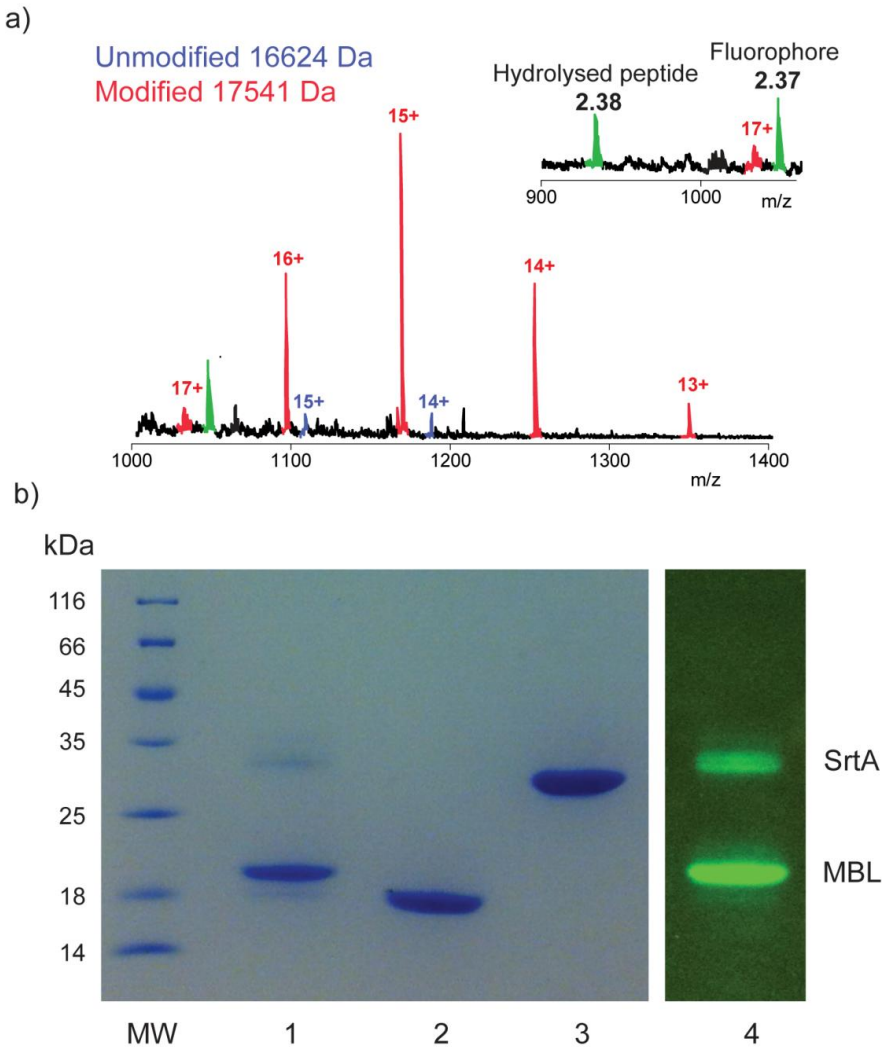


Figure 2.11: SrtA-mediated ligation of fluorescein-depsipeptide **2.37** (180 μM, 3 equivalents) to hMBL (60 μM) after 5 hours incubation with SrtA (15 mol %) at 37ºC. a) Mass spectrum of reaction mixture, b) Polyacrylamide gel (15%) of reaction mixture. Lane 1: reaction mixture, lane 2: hMBL, lane 3: SrtA, lane 4: UV irradiation of reaction mixture. Calculated mass change after modification, 915 Da - observed mass change, 917 Da.

The successful modification of hMBL using fluorophores **2.36** and **2.37** demonstrated the versatility of the depsipeptide substrate in SrtA-mediated ligation reaction. However, fluorescein-depsipeptide **2.37** appeared to be more susceptible to hydrolysis under the conditions of the reaction compared to the previous depsipeptide substrates. The reason for this is unclear, but it is unlikely that removal of the tyrosine and alanine residues from the depsipeptide structure would cause such a drastic change in stability. The more likely explanation is that the fluorescein molecule is catalysing the rate of background hydrolysis by some unknown mechanism.

2.4.2 Mouse pumilio-2 puf domain

Pumilio is a class of protein that controls eukaryotic gene expression by regulating mRNA transcription.[137] For example, during the early embryonic development of *Drosophila*, pumilio proteins bind to maternal mRNA sequences, termed *hunchback* mRNA, and recruit other proteins to repress their transcription, which results in anterior/posterior patterning of the embryo.[137-140] The mouse pumilio-2 homologue has not been as extensively studied, but it is believed to play a similar role in embryonic development.[137] However, all pumilio proteins carry a conserved RNA-binding domain or puf domain that consists of a series of eight alpha-helical repeats. The mouse pumilio-2 puf domain, donated by Thomas Edwards, has an *N*-terminal glycine residue slightly removed from the globular protein (Figure 2.12)

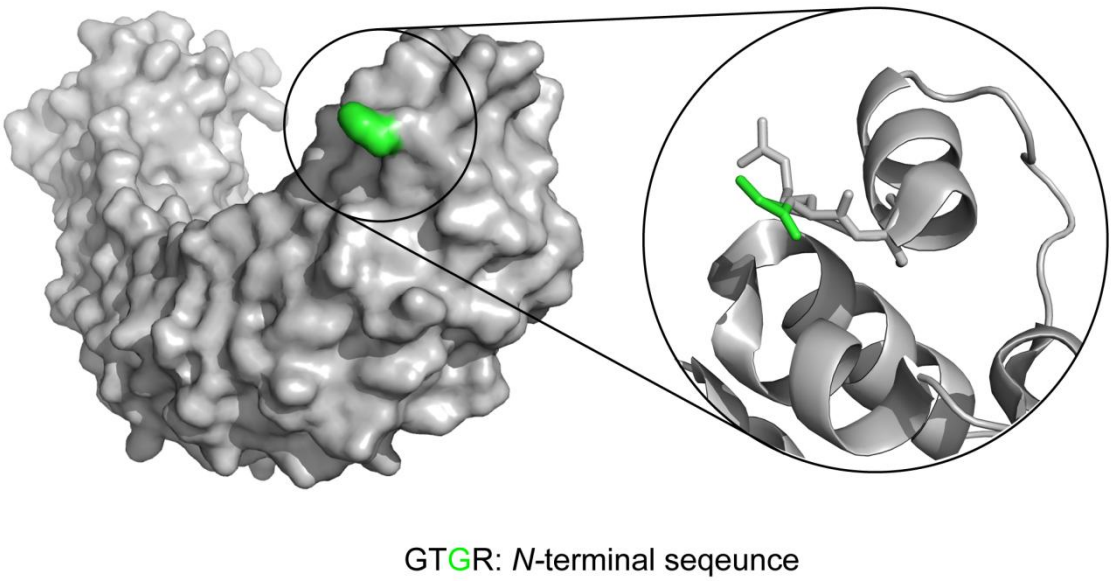


Figure 2.12: Crystal structure of mouse pumilio-2 puf domain generated from PDB code of atomic coordinates: 3GVO (UniProt code: Q80U58).[137]Green = *N*-terminus of protein in crystal structure. The final *N*-terminal amino acid residues, GT, are not resolved.

Following the successful labelling of hMBL with just 1.5 equivalents of dansyl-depsipeptide **2.36** and 10 mol% SrtA, the same protein-modification procedure was repeated with the mouse pumilio-2 puf domain at 20 μM. SDS-PAGE analysis of the transpeptidation reaction after 4 hours incubation indicated the protein had been modified (Figure 2.13a); however, mass spectrometry only showed partial labelling. The incubation time of the reaction was extended, but this did not increase the total ligation yield. It was proposed that the SrtA-mediated transpeptidation reaction was proceeding less efficiently with the puf domain when compared to hMBL because the protein substrate concentration was too low; i.e., the concentration of hMBL (60 μM) was three fold higher than the amount of puf domain (20 μM). Therefore, the rate of

the ligation reaction was sufficiently reduced to allow the non-enzymatic hydrolysis of dansyl-depsipeptide **2.36** to compete and prevent high levels of protein modification.

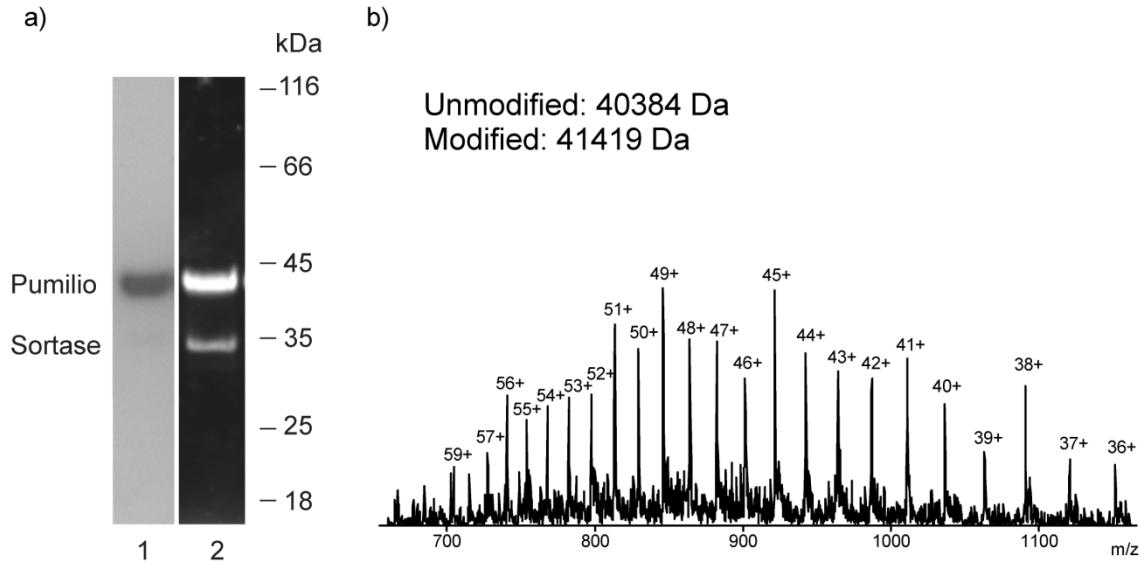


Figure 2.13: a) Polyacrylamide gel (12%) of the SrtA-mediated ligation of dansyl-depsipeptide **2.36** (30 μM, 1.5 equivalents) to mouse pumilio-2 puf domain (20 μM) after 4 hours incubation with SrtA (10 mol %) at 37ºC. Lane 1: reaction mixture, lane 2: UV irradiation of reaction mixture. b) Mass spectrum of the modified mouse pumilio after treatment with 3 equivalents of dansyl-depsipeptide **2.36** and 36 mol% SrtA. Calculated mass change after modification, 1036 Da - observed mass change, 1035 Da.

The reaction could not be repeated at a higher protein concentration as there was only a finite supply of pumilio-2 puf domain. However, effective protein labelling was observed when the reaction was repeated using 3 equivalents of dansyl-depsipeptide **2.36** and 36 mol% SrtA; mass spectrometry analysis of the reaction mixture showed no unmodified protein after 5 hours incubation (Figure 2.13b). It was postulated that the increase in concentration of both SrtA and depsipeptide allowed a build up of the enzyme-thioester intermediate. This prevented significant loss of the depsipeptide through non-enzymatic hydrolysis and gave enough time for the slower protein ligation to take place.

2.4.3 Myoglobin

Myoglobin is a cytoplasmic hemoprotein that is found in the cardiac tissue and oxidative muscle fibres of vertebrate organisms.[141,142] The protein transports oxygen from the sarcolemma to the mitochondria of muscle cells during periods of increased metabolic activity. In addition, it also acts as an oxygen store in muscle tissue to allow aquatic mammals, such as seals, to function normally throughout prolonged periods of apnoea. The protein is formed from eight alpha-helical structures assembled around a central heme subunit. The *N*-terminal glycine residue is

one amino acid removed from an alpha-helix that makes up the globular structure of the protein (Figure 2.14).

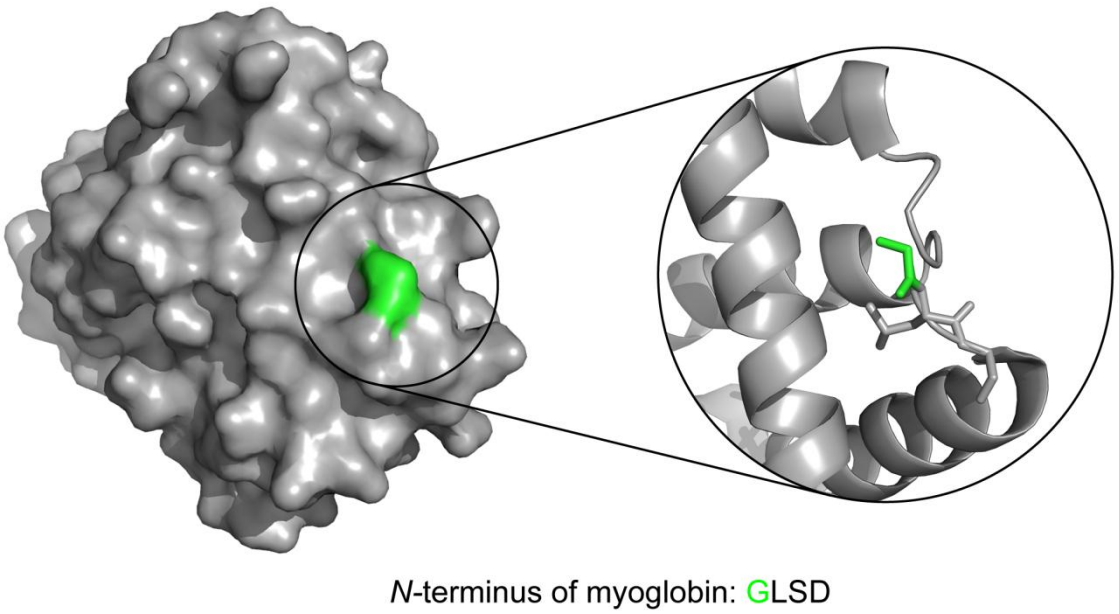


Figure 2.14 Crystal structure of myoglobin from *equine* heat. Green = *N*-terminus of protein. Generated from PDB code of atomic coordinates: 1DWR (UniProt code: P68082).[181]

Following the mixed success of the previous protein-labelling experiments, myoglobin was incubated with increasing concentrations of dansyl-depsipeptide **2.36** and SrtA. In addition, the protein substrate concentration was raised to 119 μM to reduce the possibility that it would limit the rate of the transpeptidation reaction. The assays were monitored at 4 hour intervals by mass spectrometry, but modified protein was not detected under any of the conditions of the ligation reaction. Interestingly, SDS-PAGE analysis did reveal an extremely faint fluorescent band on a gel at the position corresponding to myoglobin after a 12 hour incubation with 3 equivalents of depsipeptide substrate and 10 mol% SrtA (Figure 2.15). This suggests that the protein was being labelled by SrtA, but the level was negligible as the fluorescent band was barely visible and mass spectrometry could not detect any modified protein. There is the possibility that the modified protein was not ionising as readily as the starting material and as a result was not being detected. However, if significant labelling was taking place, a more intense fluorescent band would be visible on a gel after SDS-PAGE analysis, but this was not observed. A final ligation experiment using 5 equivalents of dansyl-depsipeptide **2.36** failed to improve the efficiency of the ligation reaction. Therefore, it likely that myoglobin was simply not a favourable substrate for the SrtA-transpeptidation reaction.

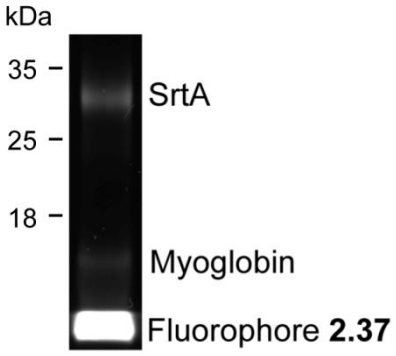


Figure 2.15:UV-irradiated polyacrylamide gel (10%) of the SrtA-mediated ligation of dansyl-depsipeptide **2.36** (356 μM, 3 equivalents) to myoglobin (119 μM) after 12 hours incubation with SrtA (10 mol %) at 37ºC,

2.4.4 Investigation into the inefficient and failed SrtA-mediated labelling of protein substrates

There were two possible explanations as to why SrtA had failed to label myoglobin. One reason could be that the *N*-terminal peptide sequence was not recognised by SrtA, but this seems unlikely as there is plenty literature precedent demonstrating the ability of SrtA to accept a wide variety of substrates, including non-peptides, as long as there is glycine residue present at the *N*-terminus.[111] The second more plausible explanation could be that the steric bulk of the globular protein in the vicinity of the *N*-terminal glycine residue was preventing SrtA-substrate binding. The *N*-terminal glycine residue of myoglobin is only one residue removed from an alpha helical structure. This suggests that there must be a minimal unstructured region at the *N*-terminus of a protein to allow SrtA binding. In addition to the failed myoglobin labelling, modification of the mouse pumilio-2 puf domain had been inefficient. We proposed that the low concentration of the protein substrate was reducing the rate of the transpeptidation reaction. However, further investigation into the effectiveness of the SrtA-transpeptidation reaction at low concentration was required to confirm this hypothesis.

An HPLC time-course experiment was performed using depsipeptide **2.24** and two short acyl acceptor peptides that contained the first four amino residues of myoglobin **2.39** and mouse pumilio-2 puf domain **2.40** (Figure 2.16). The peptides also contained a *C*-terminal tyrosine residue to allow detection by HPLC and a glycine residue because they were built on preloaded gly-2-chlorotrityl resin. The concentrations used in the time-course experiment were identical to the initial mouse pumilio-2 puf labelling experiment; i.e., acyl acceptor **2.39** or **2.40** (20 μM), depsipeptide **2.24** (30 μM, 1.5 equivalents) and SrtA (2 μM, 10 mol%). This experiment would not only help determine if the SrtA reaction was sequence or structure dependent, but it would also give an idea of the rate of the reaction when at low concentration of substrate. The results

were as expected; both acyl acceptor peptides **2.39** and **2.40** were compatible with SrtA and shared almost identical initial rates of reactivity. This indicated that the *N*-terminal amino acid sequence of the protein was not inhibiting the reaction. Furthermore, it also showed that the rate of the ligation is severally reduced when working at lower concentrations, which is probably due to the concentration of the substrate being below its K_m. This explains why 36 mol % SrtA and 3 equivalents of fluorophore **2.47** was required to drive the mouse pumilio modification to a high level of protein labelling.

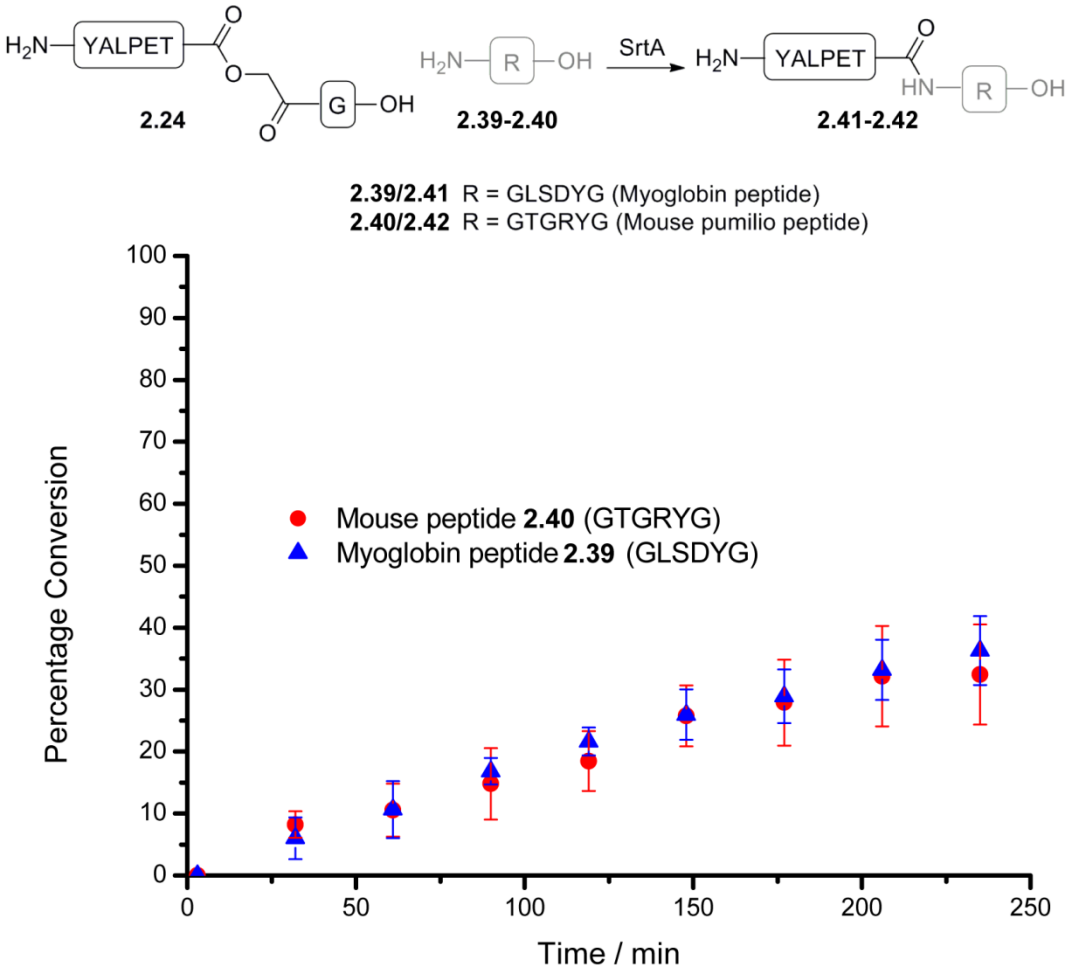


Figure 2.16: Time-course of the SrtA-mediated ligation of depsipeptide **2.24** (30 μM) and peptides **2.39** and **2.40** (20 μM)

A second HPLC time-course was performed with the short peptide sequences, but this time the concentration of the reactants was increased to match the original depsipeptide **2.24** and diglycine peptide **2.8** ligation experiment (Figure 2.17); i.e., acyl acceptor peptides **2.39** or **2.40** (500 μM), depsipeptide **2.34** (500 μM, 1 equivalent) and SrtA (11 μM, 2 mol%) This allowed direct comparison of the ligation rates of the diglycine and single glycine substrates. Unsurprisingly, the rate of product formation was faster when the diglycine acyl donor **2.8** was

used. After 100 minutes, 70% conversion to product was observed compared to the 50% when the single *N*-terminal glycine peptides **2.39** and **2.40** were used (Figure 2.5). This is consistent with previously reported observations that the rate of the SrtA ligation is dependent on having two or more *N*-terminal glycine residues.[116,117] In addition, peptides **2.39** and **2.40** were converted to a similar quantity of ligated product; however, total product formation was 5-10% lower than what was observed when diglycine peptide **2.8** was the acyl acceptor. This can probably be attributed to the difference in the rate of the transpeptidation reaction between the single and diglycine peptides. The ligation rates of the single glycine peptides **2.39** and **2.40** were slower, so more background hydrolysis of the depsipeptide **2.24** could occur before it was able to react with SrtA; therefore, the total available depsipeptide **2.24** was reduced limiting the amount of product formation.

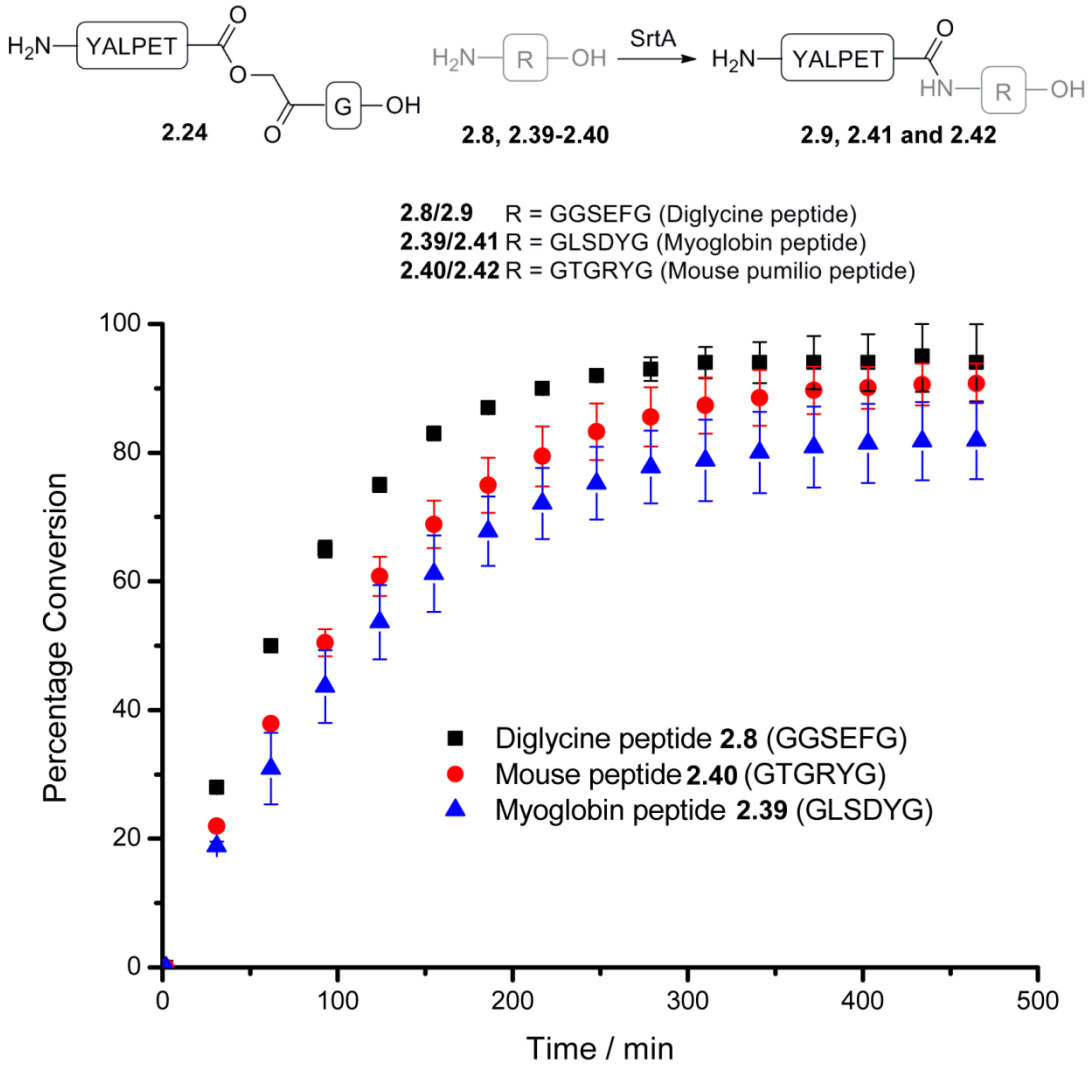


Figure 2.17: Time-course of the SrtA-mediated ligation of depsipeptide **2.24** (500 μM) and peptides **2.8**, **2.39** and **2.40** (500 μM).

2.5 Conclusion

In conclusion, the use of one equivalent of depsipeptide substrate has allowed the near quantitative labelling of acyl donor peptides and proteins that carry a single *N*-terminal glycine. This is a huge advancement on current techniques that require a large excess of either the acyl donor or acceptor for efficient SrtA-mediated ligation. However, there are two caveats that have to be considered for protein labelling to proceed successfully. The concentration of the protein has to be >60 μM to ensure efficient transpeptidation takes place or the non-enzymatic hydrolysis of the depsipeptide substrate will compete to reduce the level of protein labelling. If the protein concentration cannot be altered, then efficient modification can be achieved by increasing the concentration of both enzyme and depsipeptide. Furthermore, the *N*-terminal glycine residue has to be located in a solvent exposed area of the substrate protein or SrtA will not be able to access the intended modification site. However, the space between the globular protein mass and the *N*-terminus does not have to be vast as demonstrated with the successful labelling of the pumilio-2 puf domain. SrtA was able to access the *N*-terminal glycine residue of the protein even though there were only three amino acid residues separating it from the globular puf domain.

Chapter 3: Modified AB_5 proteins for cellular delivery

3.1 Introduction

The AB_5 proteins, cholera toxin and the heat-labile enterotoxin use a retrograde trafficking mechanism to enter mammalian cells and deliver their toxic cargo. Modified analogues of these proteins have already been used to transport biological macromolecules and smaller chemical probes into cells.[60] Several strategies have been developed to modify both the AB_5 complex and the pentameric B-subunit in a site-specific manner (Figure 3.1) - see **Chapter 1: Introduction** for an in-depth discussion of the literature. However, many of the previously reported methodologies have limitations. For example, modification to the *N*-terminus of the pentameric B-subunit has often resulted in reduced GM1 binding,[80] while pentameric instability and inhibited AB_5 complex formation has been observed with *C*-terminal modifications.[83] A more robust approach was developed by Ploegh and co-workers.[60] The authors employed SrtA-mediated ligation to attach a range of biological probes to the A1-subunit of an AB_5 complex before delivering them into cells. However, this method utilised a variant of cholera toxin that still induced toxicity, so its application as a therapeutic delivery vehicle is limited. Chimeric AB_5 complexes have also been exploited as delivery vehicles,[64] but this approach is restricted to protein macromolecules. In addition, polylysine-conjugated B-subunits have been used to transport DNA into mammalian cells;[66] however, this method relies on non-specific modification.

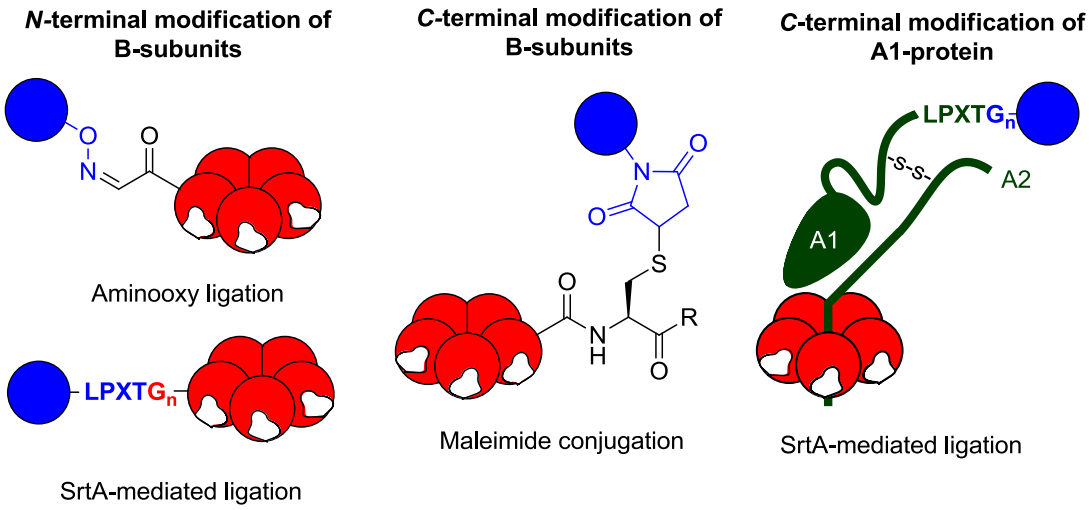


Figure 3.1: Previously reported bioconjugation strategies used to site-specifically modify cholera toxin variants. X = any amino acid R = peptide.

In this chapter I will outline a simple and robust strategy to chemically-modified AB_5 proteins that are capable of entering mammalian cell. This was achieved using a combination of chimeric AB_5 complexes, enzymatic manipulations and SrtA-mediated ligations coupled with depsipeptide substrates.

3.2 AB_5 complex modification strategy

A chimeric AB_5 complex, in which the toxic A1-domain of the A-subunit had been replaced with a maltose binding protein domain (MBP), was readily accessible using a technique developed by Holmes *et al.*,[72] and adapted by James Ross at the University of Leeds (Figure 3.2).[143] The MBP-AB_5 complex had a Tobacco Etch Virus (TEV) protease-recognition site incorporated between the MBP tag and the A2-peptide. Following enzymatic cleavage of MBP, the resulting N-terminus of the A2-peptide will display a triglycine sequence that can be labelled using SrtA-mediated ligation. This approach could potentially allow quick access to a huge variety of modified AB_5 proteins by simply varying the depsipeptide substrate for SrtA. Furthermore, it was postulated that any modification to the A2-peptide would not interfere with the ability of the protein to transfect mammalian cells as it is distal to the binding sites of the B-subunits.

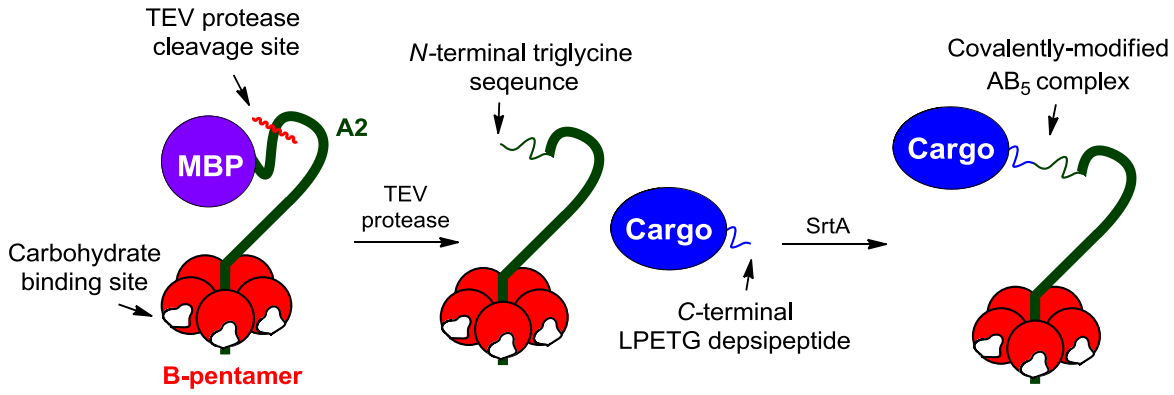


Figure 3.2: AB_5 modification strategy of an MBP-CTA2 fusion complexed with pentameric CTB. Step 1: Cleavage of MBP tag using TEV protease. Step 2: SrtA-modification of the A2-peptide

3.2.1 MBP-AB_5 expression and isolation

The MBP-CTA2 fusion protein was co-expressed with CTB in *E. coli* C41 cells harbouring the pSAB 2.1 plasmid (Appendix). Co-expressing the proteins triggered self-assembly of an MBP-CTA2/B_5 (MBP-AB_5) complex in the periplasm of the cells. A combination of amylose and nickel affinity chromatography was used to isolate the MBP-AB_5 protein complex. Initially, amylose affinity chromatography captured any species containing MBP, including the MBP-AB_5 and uncomplexed MBP-CTA2 proteins. A nickel affinity column was used to separate out the resulting mixture of proteins; the pentameric B-subunit of the MBP-AB_5 complex bound to the nickel resin while the remaining MBP-CTA2 protein passed through the column. After purification the protein can be seen as a discrete band on an SDS-PAGE gel (Figure 3.3, Lane

2). However, the hexamer appears to be unstable under the detergent conditions of SDS-PAGE analysis and it can be seen to partially dissociate on the gel into pentameric CTB and MBP-CTA2 proteins. If the MBP-AB_5 protein is completely denatured by heating before applying to the gels, it falls apart into its constituent CTB monomer and MBP-CTA2 proteins (Figure 3.3, Lane 1).

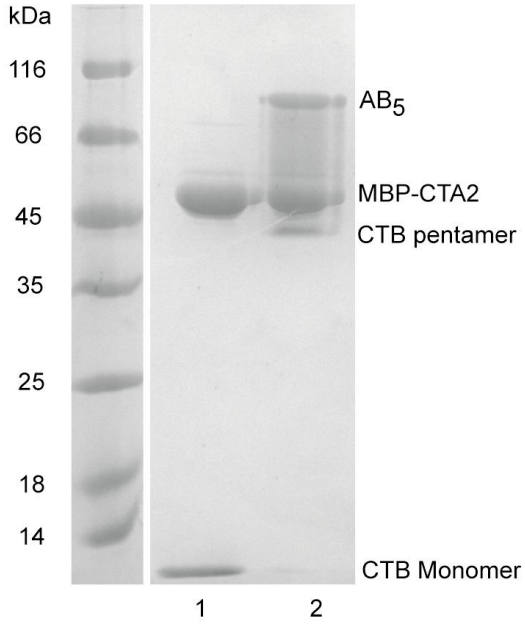


Figure 3.3: Polyacrylamide gel (12%) showing isolated MBP-AB_5 protein. Lane 1: Boiled; Lane 2: Not boiled

3.2.2 Preparation of the AB_5 complex for SrtA-mediated ligation

The MBP-purification tag had to be removed from the AB_5 complex in preparation for SrtA-mediated modification. The MBP-CTA2 construct had been designed to include an optimum TEV protease recognition sequence (ENLYFQG) to allow enzymatic cleavage and separation of the two proteins. TEV protease will specifically hydrolyse the amide bond between the QG amino acid residues to expose a triglycine sequence at the *N*-terminus of the CTA2-peptide.

TEV protease expression and purification

E. coli BL21-Gold (DE3) cells containing a pMAL-C2 expression vector for TEV protease was obtained from Dr. T. Edwards at the University of Leeds. The enzyme construct is designed to include a His-tag at the *N*-terminus of the protein followed by a TEV protease recognition sequence and MBP (Figure 3.4). Therefore, TEV protease is initially expressed as a fusion protein, but the MBP is cleaved *in vivo.* The protein was overexpressed following induction with IPTG and purification was attempted using nickel affinity and size exclusion

chromatographies. However, concentration of the protein prior to size exclusion chromatography triggered precipitation and complete loss of the sample. The use of the partially purified enzyme after the initial nickel affinity column also proved problematic. The protein precipitated during dialysis at 4 °C and it was difficult to determine if any remained in solution. After several failed attempts to increase the stability of the protein, the strategy was abandoned.

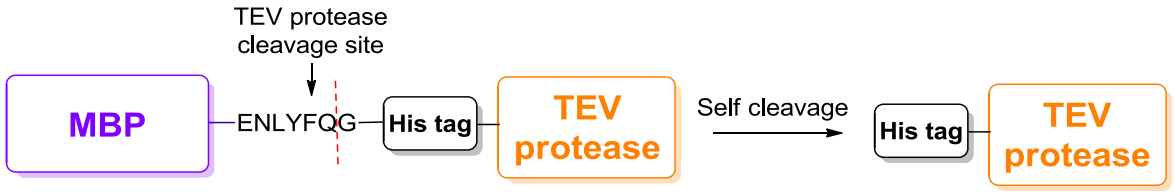


Figure 3.4: Construct of the TEV protease expressed from pMAL-c2

MBP-TEV chimera expression and purification

The poor solubility of TEV protease made it extremely difficult to handle. It was postulated that the stability of the protein could be increased by permanently fusing it to MBP. Site-directed mutagenesis could have been used to mutate the TEV protease recognition sequence in the protein construct to inhibit MBP cleavage during expression. However, this would have required two or more rounds of mutagenesis. It was, therefore, simpler to introduce the TEV protease gene into a different plasmid to form a new MBP protein fusion. This was accomplished by subcloning the gene into a pMAL-c5x plasmid as follows (Figure 3.5.): A polymerase chain reaction (PCR) was used to copy and amplify the gene from the original pMAL-C2 plasmid. In addition, the primers used in the PCR were designed to introduce Bam-HI and PST restriction sites at the termini of the amplified gene sequence to allow DNA ligation. Following digestion of the pMAL-c5x plasmid/genomic DNA with Bam-HI and PST restriction enzymes, the DNA parts were stitched together to generate the hybrid pMAL-c5x I expression vector (the aforementioned molecular biology required to create the protein fusion was conducted by Dr James Ross at the University of Leeds).

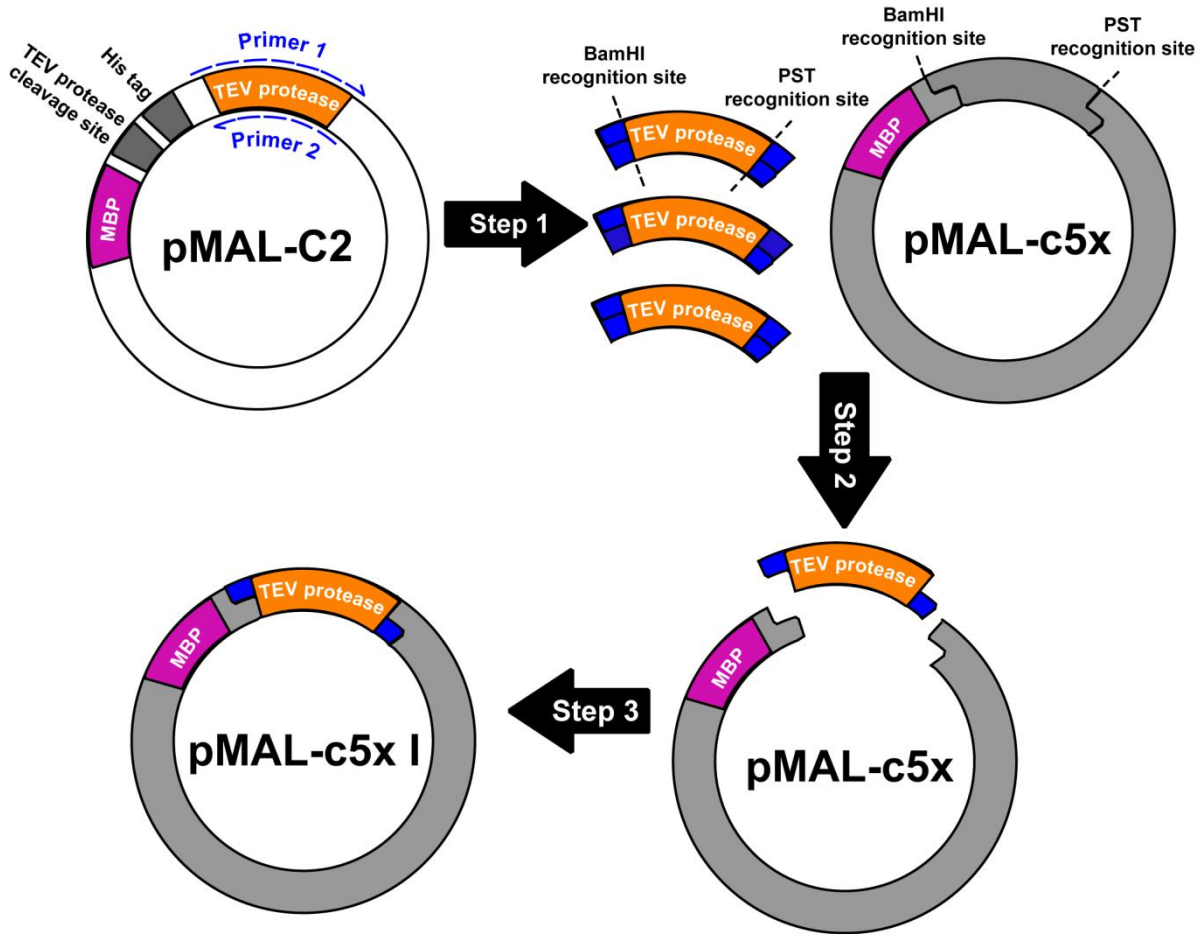


Figure 3.5: Formation of the pMAL-c5x I plasmid for MBP-TEV chimeric protein expression. Step 1: PCR amplification of the TEV gene. Step 2: Digestion of pMAL-c5x vector and gene insert using restriction enzymes BamHI and PST. Step 3: DNA ligation of digested DNA to form the pMAL-c5x I expression vector.

The pMAL-c5x I expression vector was transformed into *E. coli* BL21-Gold (DE3) cells and the MBP-TEV fusion was overexpressed using IPTG induction. The protein chimera was purified using a combination of amylose affinity and size exclusion chromatographies (Figure 3.6). SDS-PAGE analysis identified an overexpressed protein at ~70 kDa on the gel (Figure 3.6, Lane A). Further analysis by ESMS confirmed the mass of the protein to be 71665 Da, which is in good agreement with the calculated molecular weight for the MBP-TEV chimera (71633 Da). A second overexpressed protein that ran at ~40 kDa on the gel (Figure 3.6, Lane B) was also isolated; this was probably native MBP that had been expressed naturally by the *E. coli* cells.

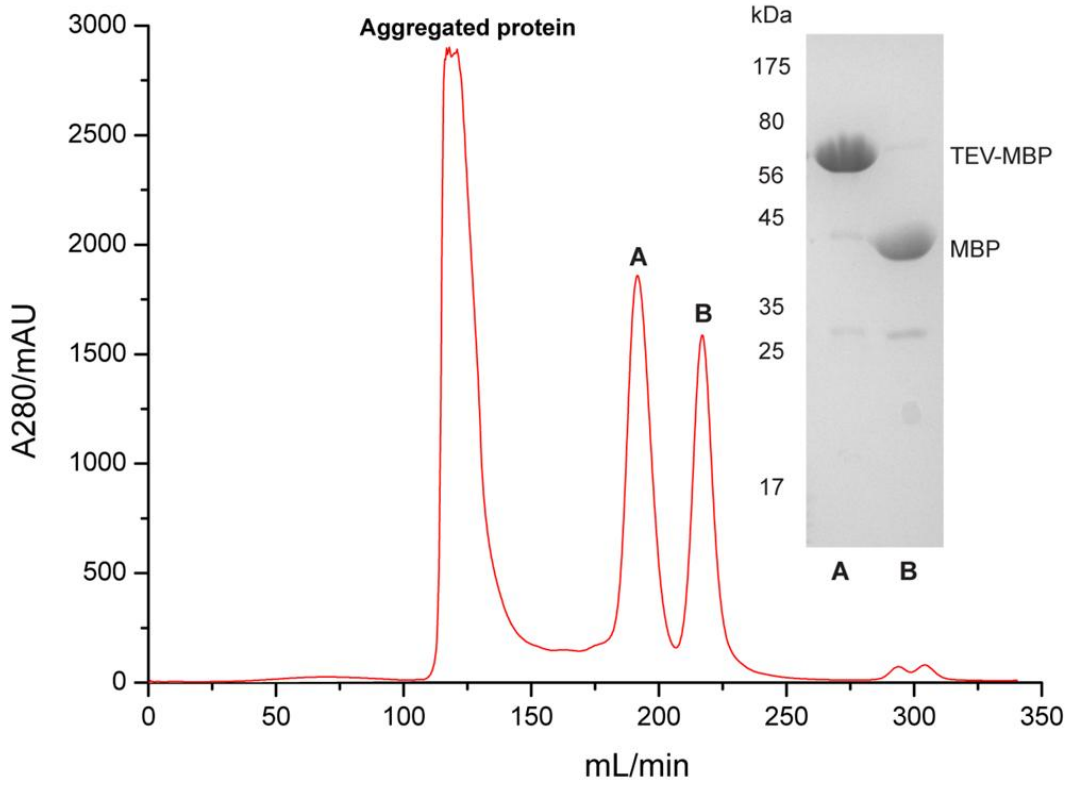


Figure 3.6: Size exclusion chromatogram and polyacrylamide gel (12%) gel of MBP-TEV purification.

TEV protease-mediated MBP cleavage

Treating the MBP-AB_5 complex (45 μM) with 10 mol% MBP-TEV (4.5 μM) for 1 hour at room temperature successfully generated a truncated CTA2-AB_5 complex. The reaction was monitored by ESMS and the molecular weight of the MBP-CTA2 protein decreased by the mass of the CTA2-peptide, 5055 Da. However, at higher MBP-TEV concentrations or longer incubation periods, indiscriminate proteolysis of the protein was observed. The truncated AB_5 complex was isolated from the cleavage mixture using nickel affinity chromatography, and two sets of signals corresponding to the CTA2 and CTB chains were identified by ESMS (Figure 3.7).

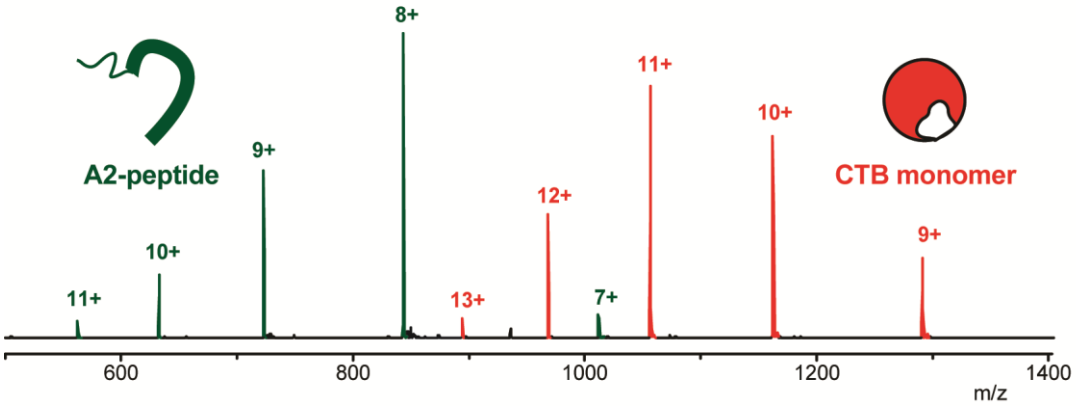


Figure 3.7: Mass spectrum of the truncated CTA2-AB_5 complex. Green: CTA2 peptide (5055 Da); Red: CTB monomer (11611 Da).

3.3 Sortase-mediated modification of the AB_5 protein

The CTA2-AB_5 complex was now primed for use with SrtA-mediated modification. To compare the efficiency of the ligation reaction to previous protein substrates for SrA, fluorescent-depsipeptides **2.36** and **2.37** were used to label the complex (Figure 3.8).

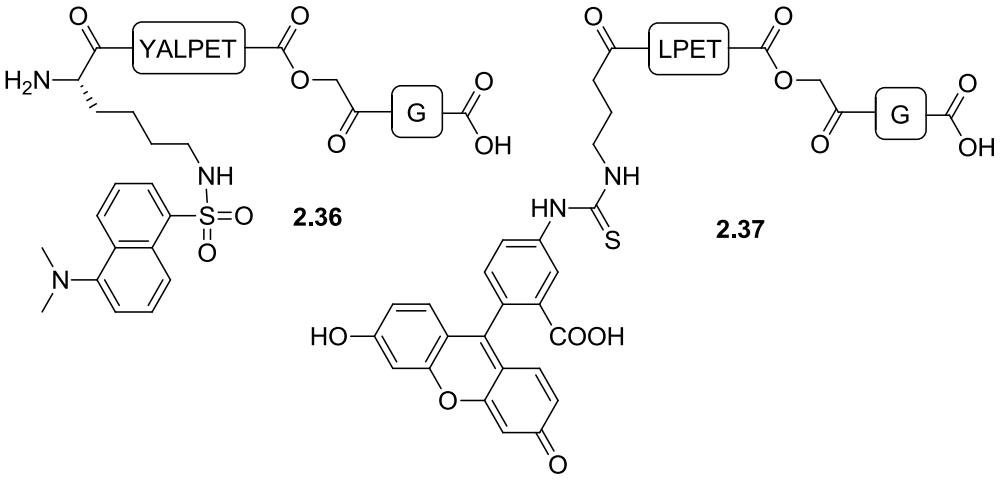


Figure 3.8: Structures of dansyl-depsipeptide **2.36** and fluorescein-depsipeptide **2.37**

3.3.1 Dansyl-depsipeptide ligation

The CTA2-AB_5 complex (44 μM) was modified with 1.5 equivalents of dansyl-depsipeptide **2.36** (71 μM) and 10 mol% SrtA (4.4 μM) at 37 °C in just 3 hours (Figure 3.9). The ligation reaction was faster than the previous protein-labelling experiments reported in Chapter 2. It was postulated that the triglycine sequence at the *N*-terminus of the A2-peptide was a more favourable substrate for SrtA than the single glycine residues displayed by the earlier protein

substrates. This observation was in good agreement with the suggestion that SrtA has a diglycine binding site.[116]

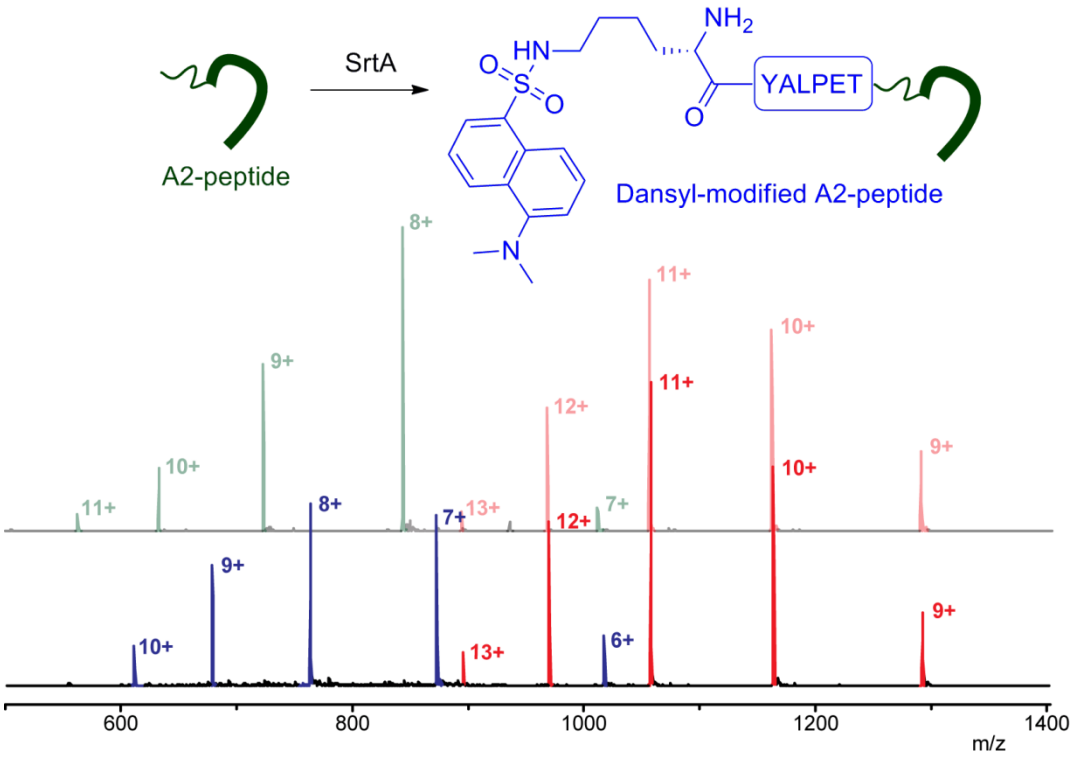


Figure 3.9 Mass spectra of the dansyl-modified-AB_5 complex overlaid on top of the unmodified starting material. Blue: CTA2 labelled with dansyl-depsipeptide **2.36** (6091 Da); opaque green: CTA2 (5055 Da) and red/opaque red: CTB monomer (11611 Da). Modification was achieved using 10 mol% SrtA and 1.5 equivalents of depsipeptide **2.36.** Calculated mass change after SrtA-mediated modification, 1036 Da - observed mass change, 1036 Da.

The A2-domain of the AB_5 complex was the only protein species labelled by SrtA, which demonstrated the high selectivity of the SrtA-mediated ligation reaction. However, the pentameric B-subunit is a very stable and compact protein, so it may be more resistant to non-specific labelling compared to more flexible proteins.

3.3.2 Fluorescein-depsipeptide ligation

The CTA2-AB_5 complex (43 μM) was also labelled in 3 hours with 2 equivalents of fluorescein-depsipeptide **2.37** (90 μM) and 12 mol% SrtA (Figure 3.10). However, the efficiency of the SrtA-mediated ligation has been inconsistent; in some instances, it has taken between 2-5 equivalents of depsipeptide, 10-20 mol% SrtA and 3-6 hours to achieve complete protein modification. There are several factors that could have contributed to the variation in the results. These include the hydrolytic stability of the depsipeptide, enzyme activity and protein substrate concentration. Rapid hydrolysis of the fluorescein-depsipeptide **2.37** has been

observed in previous protein labelling experiments with MBP (Chapter 2). The rate of hydrolysis may be sensitive to fluctuations in the pH of the buffer and as a result the amount of available substrate may vary in different experiments. It is more likely, however, that the protein substrate concentration in each assay was the main contributing factor to the difference in the SrtA-mediated transpeptidation reactions. It has been demonstrated that the rate of the ligation is slower at lower concentrations (Chapter 2). Therefore, non-enzymatic hydrolysis of the depsipeptide substrate will be more significant preventing efficient modification of the protein.

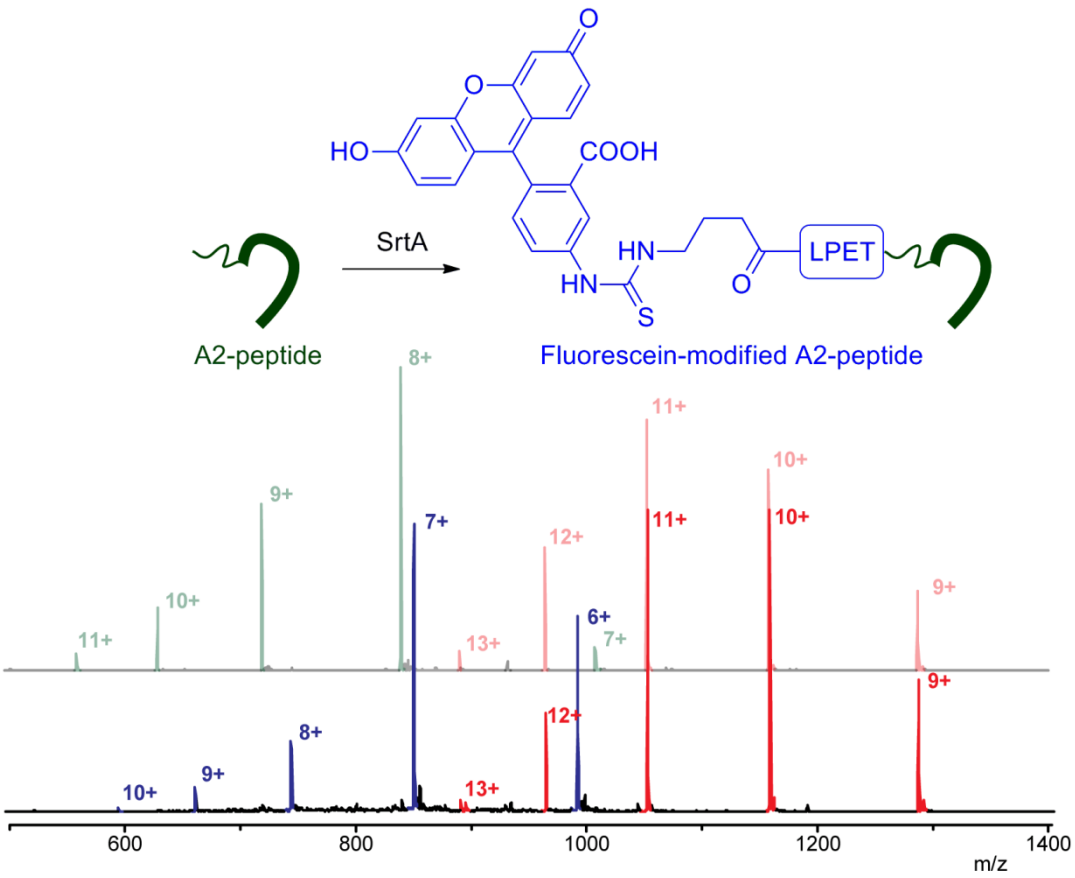


Figure 3.10 Mass spectra of the fluorescein -modified-AB_5 complex overlaid on top of the unmodified starting material. Blue: CTA2 labelled with fluorescein-depsipeptide **2.37** (5970 Da); opaque green: CTA2 (5055 Da) and red/opaque red: CTB monomer (11611 Da). Modification was achieved using 10 mol% SrtA and 2 equivalents of depsipeptide **2.37.** Calculated mass change after SrtA-mediated modification, 915 Da - observed mass change, 915 Da.

3.3.2.1 Isolation and analysis of fluorescein-AB_5 complex

Nickel affinity chromatography was used to isolate the fluorescein-AB_5 complex from the SrtA-ligation mixture. Although both proteins are able to bind the nickel resin, SrtA binds with much higher affinity because it carries an *N*-terminal His-tag. Conversely, the AB_5 complex binds through a series of surface exposed histidine residues on the B-pentamer. Therefore, only a 50

mM imidazole wash was required to elute the modified AB_5 protein and thus separate it from the majority of SrtA. A combination of SDS-PAGE analysis and UV irradiation indentified the fluorescein-AB_5 complex as a fluorescent protein band on the gel (Figure 3.11). In addition, the pentameric B-subunit and fluorescein labelled CTA2-peptide were also observed due to partial dissociation of the AB_5 complex. Two extremely faint fluorescent protein bands were also detected; the highest weight band could belong to a multimeric AB_5 species because it disappears once the sample had been boiled, which suggests it existed as a non-covalently linked complex. The other fluorescent band is probably a small amount of the SrtA-acyl intermediate that was not removed by nickel affinity purification.

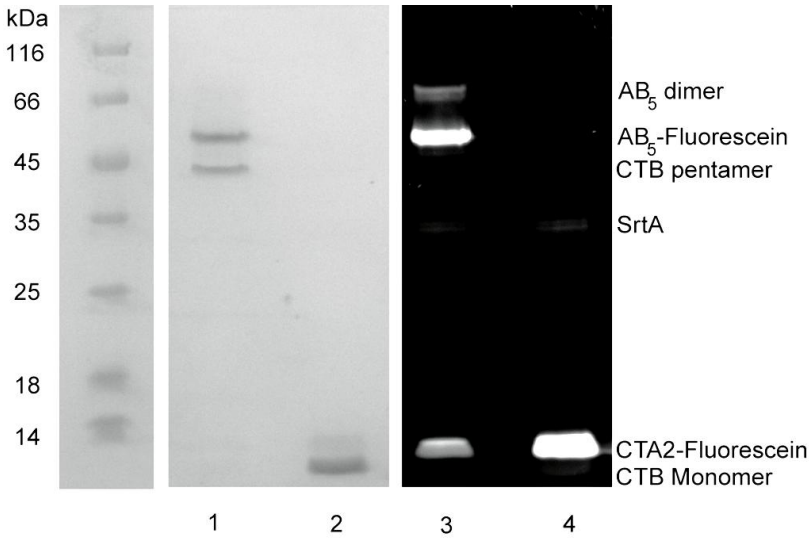


Figure 3.11: Polyacrylamide gel (12%) showing isolated fluorescein-AB_5 complex. Lane 1: Not boiled; lane 2: boiled; lane 3: UV irradiation of lane 1; lane 4: UV irradiation of lane 2.

3.4 Transportation of AB_5 complex into mammalian cells

AB_5 complexes had been successfully labelled using SrtA-mediated ligations, but their ability to enter mammalian cells had yet to be determined. A selection of mammalian cells were treated with the fluorescein-AB_5 complex and fluorescence microscopy was used to determine protein translocation.

3.4.1 Studies with various mammalian cell lines

The fluorescein-AB_5 complex was initially screened against HeLa cells, which are derived from human cervical cancer tissue.[144] Unfortunately, no fluorescence was detected around or inside the cells after several experiments. This was surprising as the ability of cholera toxin to enter HeLa cells has been documented.[145] However, a recent study by Fujitani *et al*. into the glycome of common human cell lines could explain why no protein internalisation was observed.[146] The authors discovered that glycolipids, including the GM1 receptor required for cholera toxin binding, were only present in tiny quantities in HeLa cells (Figure 3.12). Therefore, it is likely that the AB_5 complex was binding to the cells in such small quantities that the fluorescence was not detected.

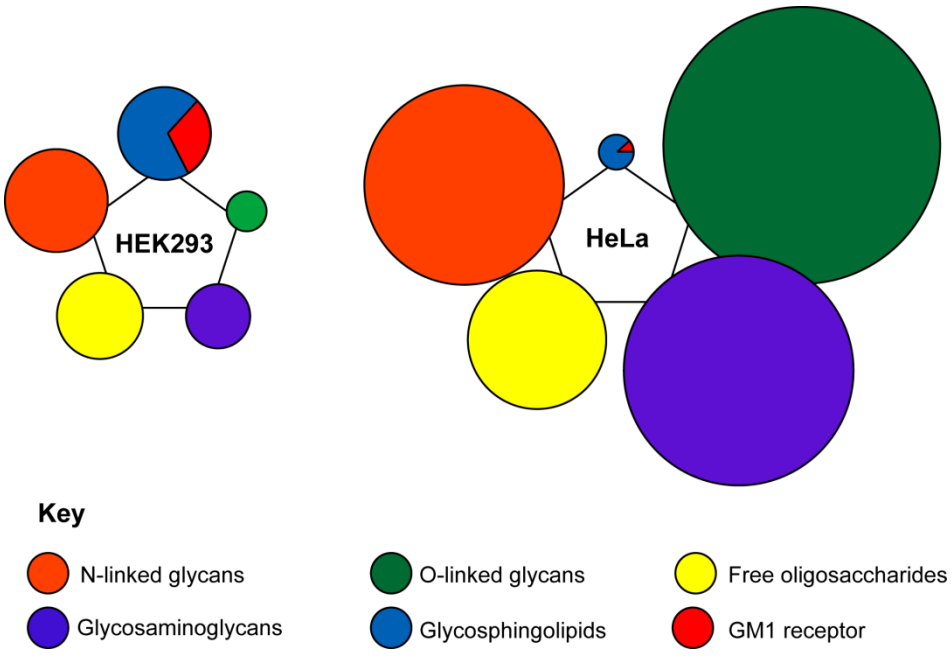


Figure 3.12: Glycome profile of HEK293 and HeLa cells. The size of each circle and constituent fraction represents the absolute quantity of the different classes of glycan (pmol/ μg protein). The size of the free oligosaccharide circle is increased by 40-fold.[146]

In the same study by Fujitani *et al*., the GM1 receptor was found to be present in much higher concentrations in human embryonic kidney 293 (Hek-293) cells than HeLa cells. Therefore, it

was decided to repeat the protein transportation experiments with Hek-293 cells. Human embryonic kidney 293 (Hek-293) cells were treated with the AB_5 complex for a one hour incubation period before being imaged with a fluorescent microscope (Figure 3.13). Translocation of the Hek-293 cells was a success and a large amount of fluorescence was detected. Interestingly, there appeared to be small punctate structures in some of the cells, which suggests the protein had been internalised and sequestered inside an intracellular compartment. However, a large proportion of the fluorescence was localised at the membrane of the cell, so translocation was either slow or the cell was being flooded with a large excess of protein. Further investigation was required to establish the intracellular fate of the AB_5 complex.

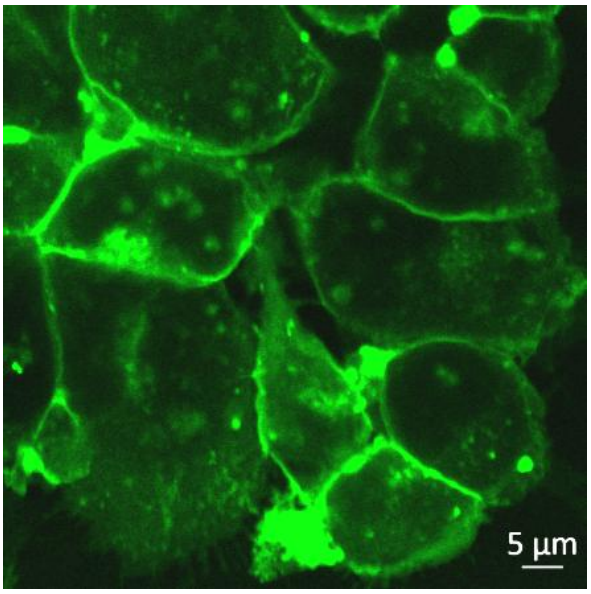


Figure 3.13: Hek-293 cells transfected incubated fluorescence-AB_5 protein (~40 µg / mL) in PBS for 1 hour at 37 °C. Fluorescein-AB_5 (green)

Immunostaining coupled with fluorescence microscopy allows the detection of specific organelles within a cell. Differentiation of intracellular targets can be difficult in Hek-293 cells, so it was decided to switch cell lines. Vero cells, derived from the epithelial cells of African green monkeys, were selected as an alternative because they typically give clear immunostaining images. Although the absolute quantity of GM1 in Vero cells was not known, several studies have successfully used these cells to study cholera toxin translocation.[64,147] After incubation with the fluorescein-AB_5 complex, translocation was successfully detected in a number of Vero cells, but it appeared to be less homogenous than for the Hek-293 cells. Majoul *et al.* discovered that the life-cycle of Vero cells governs the biosynthesis of surface exposed GM1 gangliosides and as a result directly affects the ability of cholera toxin to bind to the cells.[147] Therefore, it is likely that fluorescein-AB_5 complex internalisation was only observed in

Vero cells that were at the correct life-cycle stage to display an abundance of GM1 gangliosides at their surface.

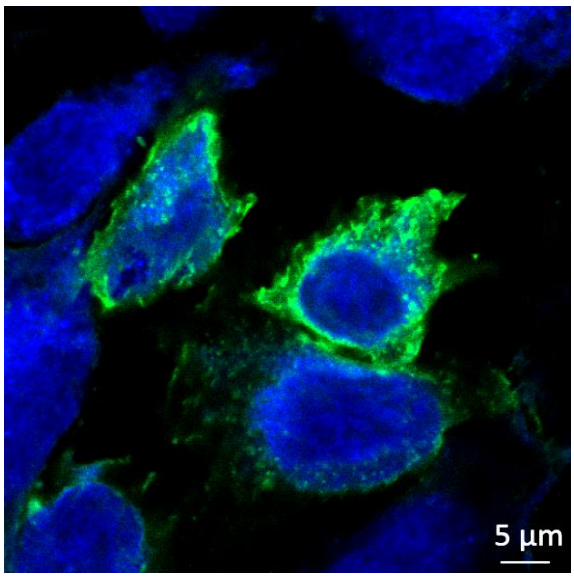


Figure 3.14: Vero cells transfected incubated fluorescence-AB_5 protein (~40 µg / mL) in PBS for 1 hour at 37 °C. Merged image: nuclear stain (blue, DAPI), AB_5-fluorescein (green)

3.4.2 Organelle-specific immunostaining

Native cholera toxin is thought to be transported in an endosomal vesicle to the endoplasmic reticulum (ER) via the trans-Golgi network. Therefore, it was important to identify all of the organelles involved in the trafficking mechanism as it was likely that the modified AB_5 complex would follow the same intracellular pathway. Furthermore, the majority of unwanted proteins are trafficked to lysosomes for degradation, so it was prudent to tag this compartment too. To track the intracellular progress of the protein at different time points, Vero cells were incubated with the AB_5 complex at 37 °C for 15, 30 and 60 minutes before being chemically fixed. Additionally, the cells were incubated at 4°C for 15 minutes in an effort to inhibit internalisation and observe the protein binding to the cell surface.

A series of control experiments was conducted to try to avoid the number of false positive results. Vero cells that had not been immunostained or treated with fluorescein-AB_5 were exposed to the blue (408 nm), green (488 nm) and red (555 nm) lasers to establish the level of background fluorescence (Figure 3.15). Excitation with the blue laser identified the DAPI fluorophore, which is present in the cell mounting agent, but no fluorescence was detected after exposure with the green and red laser. Therefore, any fluorescence detected at these wavelengths after incubation with the protein could not be attributed to the cell.

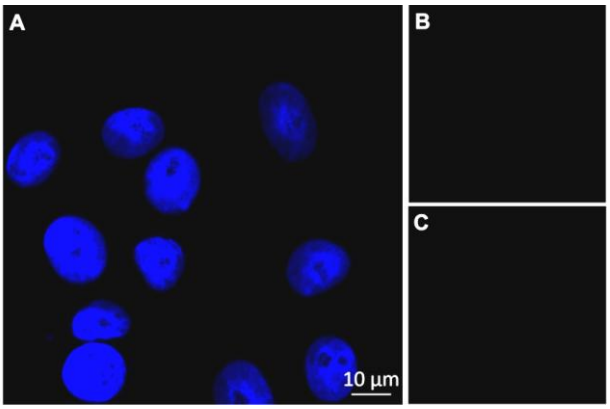


Figure 3.15: Vero cells. A) Blue image, nuclear stain (DAPI). B) Green image. C) Red image. There is no signal in B and C; therefore, there is no autofluorscence from the Vero cells in the green and red channels.

A secondary antibody, derived from either goat or hamster, linked to the same AlexaFluor® 555 dye was used in every immunostain experiment. It was important to establish if there was any significant emission from the dye after excitation with the green laser, as the fluorescein molecule linked to the AB_5 protein absorbs in this region (Figure 3.16). Only residual fluorescence was detected even when the green laser was set to maximum gain (Figure 3.16 B). Therefore, any fluorescence detected after excitation with the green laser can be attributed to the presence of the protein and not the AlexaFluor® 555 dye.

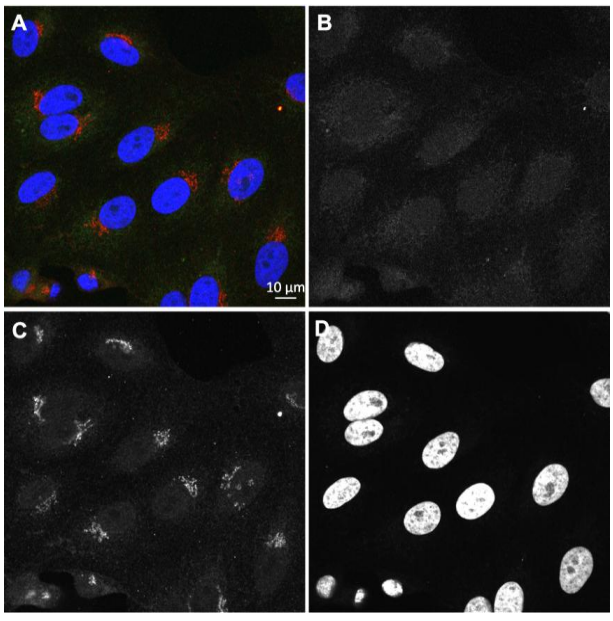


Figure 3.16 Immunostained Vero cells treated with AlexaFluor® 555 dye linked secondary antibody. A) Merged image: nuclear stain (blue, DAPI), Golgi stain (red, AlexaFluor® 555 dye). B) Green image: Maximum gain C) Red Image. D) Blue image

The immunostaining was successful for the late endosome and Golgi organelles, but only non-specific staining was observed when the ER, lysosome and early endosome antibodies were used. It was postulated that the fixation process might be disrupting antibody recognition by altering the structure of the target antigen on the organelles. In an effort to limit antigen denaturing, the Vero cells were fixed under different conditions: methanol, 4% para-formaldehyde in PBS or 4% formaldehyde in PBS. However, this did not improve the specificity of the immunostaining, so further optimisation is required to visualise the remaining organelles.

The surface of the Vero cells was surrounded by the fluorescein-AB_5 complex and little, or no, internalisation was detected when the cells were incubated at 4 ˚C for 15 minutes (Figure 3.17). This suggests the protein was binding to the GM1 receptors on the cell surface, but endocytosis was extremely slow or simply inactive; these observations are consistent with findings previously reported in the literature.[64]

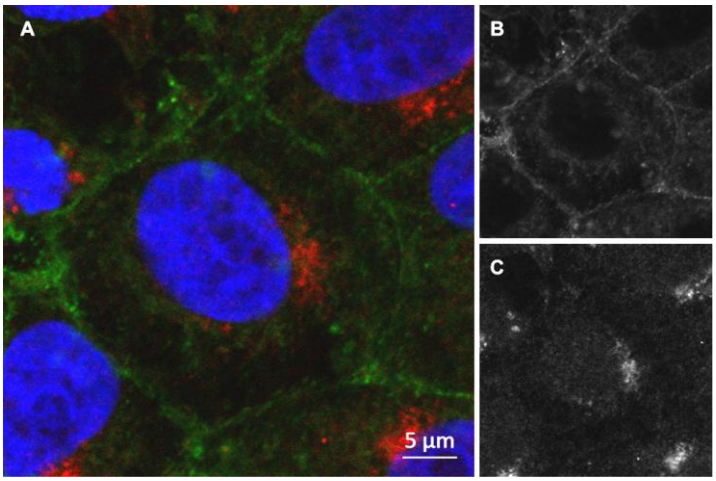


Figure 3.17 Immunostained Vero cells incubated with fluorescence-AB_5 protein (~40 µg / mL) in PBS for 15 minutes at 4 °C. A) Merged image: nuclear stain (blue, DAPI), AB_5-fluorescein (Green) and Golgi stain (red, AlexaFluor® 555 dye). B) Green image C) Red Image

Internalisation of the AB_5 complex was observed after 15 minutes incubation at 37 ˚C, but there was still a significant amount of protein on the surface of the cell (Figure 3.18). A small amount of yellow florescence was detected at the Golgi suggesting some of the protein was starting to localise within the organelle. However, most of the flurescein-AB_5 complex was spread out within the cell in other punctate structures, probably early endosomal vesicles, but it is difficult to determine the exact locations without the other immunostaining markers.

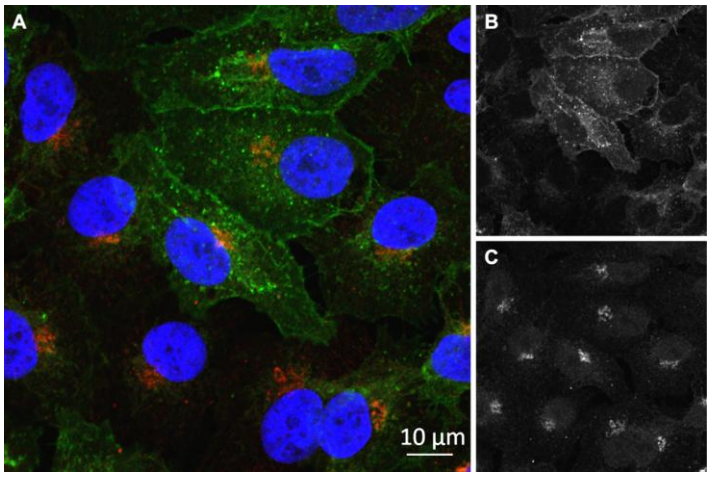


Figure 3.18 Immunostained Vero cells incubated with fluorescence-AB_5 protein (~40 µg / mL) in PBS for 15 minutes at 37 °C. A) Merged image: nuclear stain (blue, DAPI), AB_5-fluorescein (Green), Golgi stain (red, AlexaFluor® 555 dye) and co-localisation (yellow). B) Green image C) Red Image

After 30 minutes incubation at 37 ˚C, most of the protein was completely internalised and a small proportion appeared to be localised within the late endosome (Figure 3.19). However, once again most of the protein florescence did not appear to be concentrated at any specific location, but simply extended throughout the cell. In a study by Tinker *et al*.,[64] a GFP-AB_5 protein was transported to the Golgi after 30 minutes incubation at 37 °C, but unfortunately immunostaining of the Golgi failed at this time point making a direct comparison impossible.

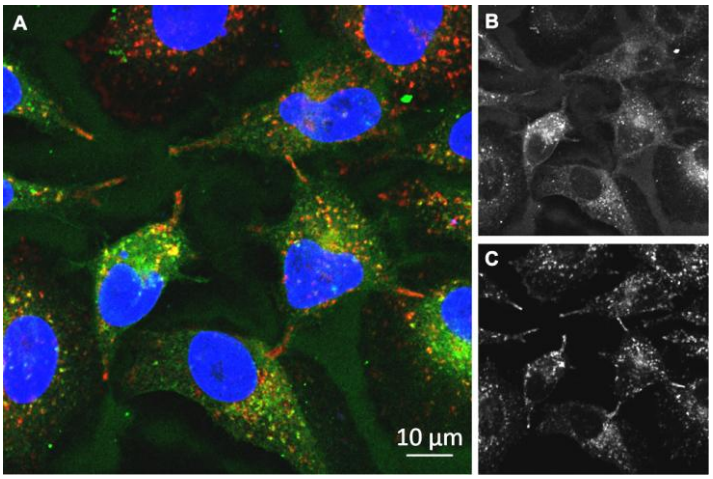


Figure 3.19:Immunostained Vero cells incubated with fluorescence-AB_5 protein (~40 µg / mL) in PBS for 30 minutes at 37 °C. A) Merged image: nuclear stain (blue, DAPI), AB_5-fluorescein (Green), late endosome stain (red, AlexaFluor® 555 dye) and co-localisation (yellow). B) Green image C) Red Image

The fluorescein-AB_5 complex appeared to be sequestered in larger punctate structures within the cell after 60 minutes incubation at 37 °C (Figure 3.20). There was still a small amount of co-localisation with the late endosomal markers, but the majority of the fluorescence was situated in a compartment directly next to the nucleus; this is most clearly seen when the green florescence image is viewed on its own (Figure 3.20 B and E). The punctate structure resembled the size and shape of the Golgi, but direct comparison to the immunostained Golgi was not conclusive as protein translocation appeared less prevalent in the cells (Figure 3.20). However, there was a small quantity of yellow fluorescence, which indicates at least some localisation in that compartment.

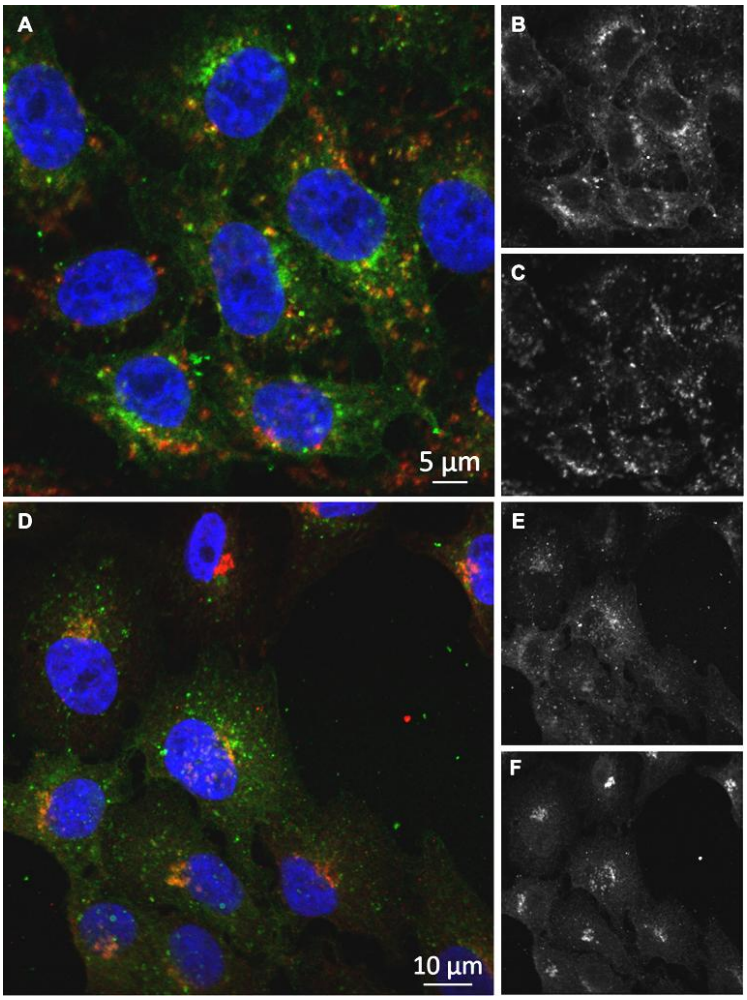


Figure 3.20: Immunostained Vero cells incubated with fluorescence-AB_5 protein (~40 µg / mL) in PBS for 60 minutes at 37 °C. A) Merged image of B and C: nuclear stain (blue, DAPI), AB_5-fluorescein (Green), late endosome stain (red, AlexaFluor® 555 dye) and co-localisation (yellow). B) Green image of late endosome immunostain. C) Red image of late endosome immunostain.. C) Merged image of E and F: nuclear stain (blue, DAPI), AB_5-fluorescein (Green), Golgi stain (red, AlexaFluor® 555) and co-localisation (yellow). E) Green image of Golgi immunostain. F) Red image of Golgi immunostain.

It is quite clear from the preliminary protein translocation assays that the fluorescein labelled AB_5 complex is easily internalised by different classes of mammalian cells. However, further investigation is required to establish the exact intracellular pathway of the modified protein; this information will be vital in deciding how best to utilise AB_5 delivery vehicles.

3.5 Conclusion

Using a series of enzymatic reactions, it has been possible to create a chemically modified AB_5 complex that is capable of transfecting mammalian cells. In theory, this strategy could be used to easily create a wide range of modified AB_5 proteins that could deliver a variety of different therapeutic cargo inside cells. However, more work needs to be done to determine the intracellular fate of the modified proteins.

Chapter 4: Bioorthogonal depsipeptide substrates for Sortase

4.1 Introduction

SrtA-mediated ligations allow the efficient labelling of proteins with a range of different peptide substrates. In general, most substrates are synthesised using SPPS; therefore, nucleic acids and other non-amino acid-based molecules that are not compatible with this process cannot be attached to proteins using SrtA. Alternative bioorthogonal chemistry such as 1,3-dipolar cycloadditions and Diels Alder reactions ligations can be used to attach non-peptidic molecules onto proteins.[148] However, these methods often require non-natural amino acids to be introduced into the protein of interest before conjugation can take place. In contrast, only a simple *N*-terminal oligoglycine sequence or a *C*-terminal LPXTG recognition motif needs to be present in a protein for SrtA-mediated ligations to proceed. Therefore, introducing a bioorthogonal handle into a depsipeptide substrate for SrtA would allow efficient site-specific labelling of proteins with the prospect of further modifications with non-peptidic molecules.

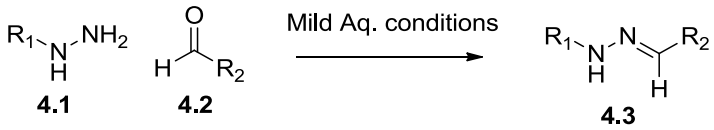
Building on our previous work into the SrtA-mediated modification and transportation of AB_5 complexes into mammalian cells, it was postulated that the proteins could be used as oligonucleotide delivery vehicles. This chapter describes the development of a method that uses a combination of depsipeptide substrates for SrtA and bioorthogonal linkers to attach oligonucleotide macromolecules onto an AB_5 protein.

4.2 Bioorthogonal linker selection

The ultimate aim was to conjugate oligonucleotides onto the AB_5 complex; therefore, the bioorthogonal linker had to be complementary to the commercially available linkers that are routinely attached to oligonucleotides. In addition, it had to be easily incorporated into depsipeptide substrates for SrtA and stable throughout the initial protein ligation step. An investigation to find a suitable bioorthogonal linker that satisfied the above criteria was launched.

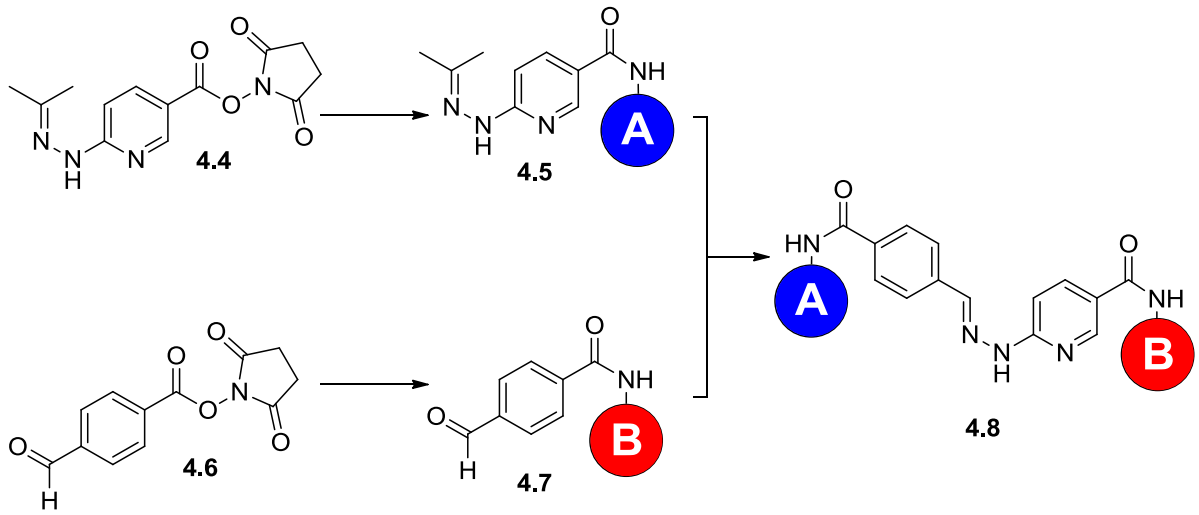
4.2.1 Hydrazone linker

Hydrazine **4.1** and aldehyde-linked molecules **4.2** can condense under mild aqueous conditions to form hydrazones **4.3** (Scheme 4.1)[149-152] This approach has been used extensively to attach both chemical and biological probes to biological macromolecules.[151,153,154] The structure of the hydrazone can directly affect its hydrolytic stability. For example, acyl hydrazones are susceptible to hydrolysis under physiological conditions and rapidly hydrolyse below pH 6.[149,155] In contrast, bis-aryl hydrazones are stable to hydrolysis at neutral pH and have been widely used in the formation of biological conjugates.[149,151,153,154]



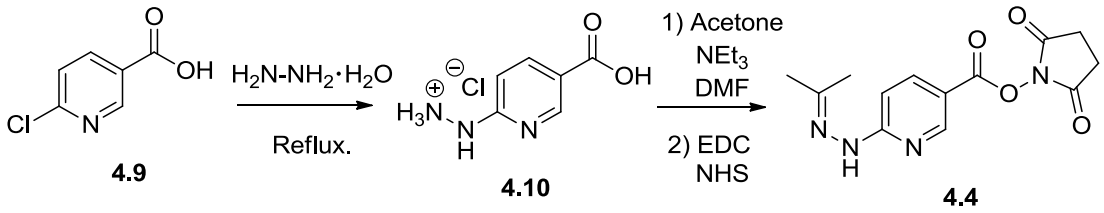
Scheme 4.1: Hydrazonation reaction between hydrazine **4.1** and aldehyde **4.2** to form hydrazone **4.3**. R_1 = acyl or aryl group. R_2 = aryl, alkyl or hydrogen.

Solulink manufacture a protein cross-linking kit that is commonly used to conjugate antibodies to oligonucleotides/other proteins (Scheme 4.2). Conjugation is achieved by using a trans-hydrazonation reaction to form a bis-aryl hydrazone bond between biological macromolecules that have been pre-functionalised with either an aryl aldehyde **4.7** or an aryl hydrazone **4.5**. The aryl aldehyde **4.4** and aryl hydrazone cross-linkers **4.6** are introduced into the biomolecules of interest non-specifically. Therefore, the number of cross-linkers per biomolecule is not homogenous, which in turn produces a heterogeneous mix of protein-conjugates during the trans-hydrazonation reaction. Taking inspiration from this approach, a strategy was developed to attach the aryl hydrazone **4.4** onto a depsipeptide substrate for SrtA to allow the site-specific incorporation of the linker onto the AB_5 complex. This will allow oligonucleotides functionalised with aryl aldehyde to be conjugated homogeneously onto the protein.



Scheme 4.2: General protein labelling and conjugation procedure using cross-linkers **4.4** and **4.6** produced by Solulink. A and B = biological macromolecules.

Aryl hydrazone **4.4** was synthesised using a slightly modified strategy developed by Grotzky *et. al.*[156] Initially, chloropyridine **4.9** was treated with hydrazine monohydrate to give hydrazinium salt **4.10** in 52% yield (Scheme 4.3). Without further purification, the hydrazinium salt **4.10** was exposed to acetone under basic conditions before the addition of N-hydroxysuccinimide formed aryl hydrazone **4.4** in a modest 46 % yield.



Scheme 4.3: Synthesis of hydrazone cross-linker **5.4**.

The aryl hydrazone **4.4** could have been coupled onto a depsipeptide substrate during SPPS; however, it was predicted that the aryl hydrazone would hydrolyse under the strong acidic conditions used in the peptide cleavage step to form an aryl hydrazine **4.11** (Figure 4.1). In principle, the aryl hydrazine **4.11** could attack the ester linkage in the depsipeptide to form polymerised byproducts **4.12**. Therefore, it was decided to couple the aryl hydrazone **4.4** to a depsipeptide in solution after SPPS.

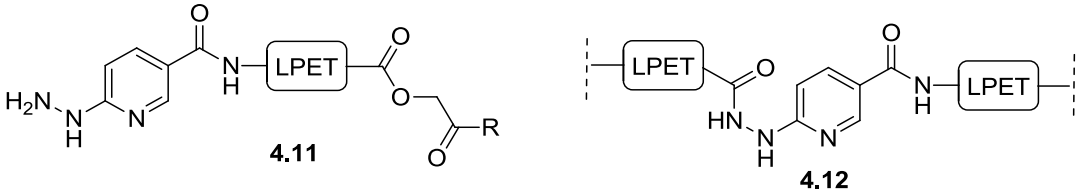
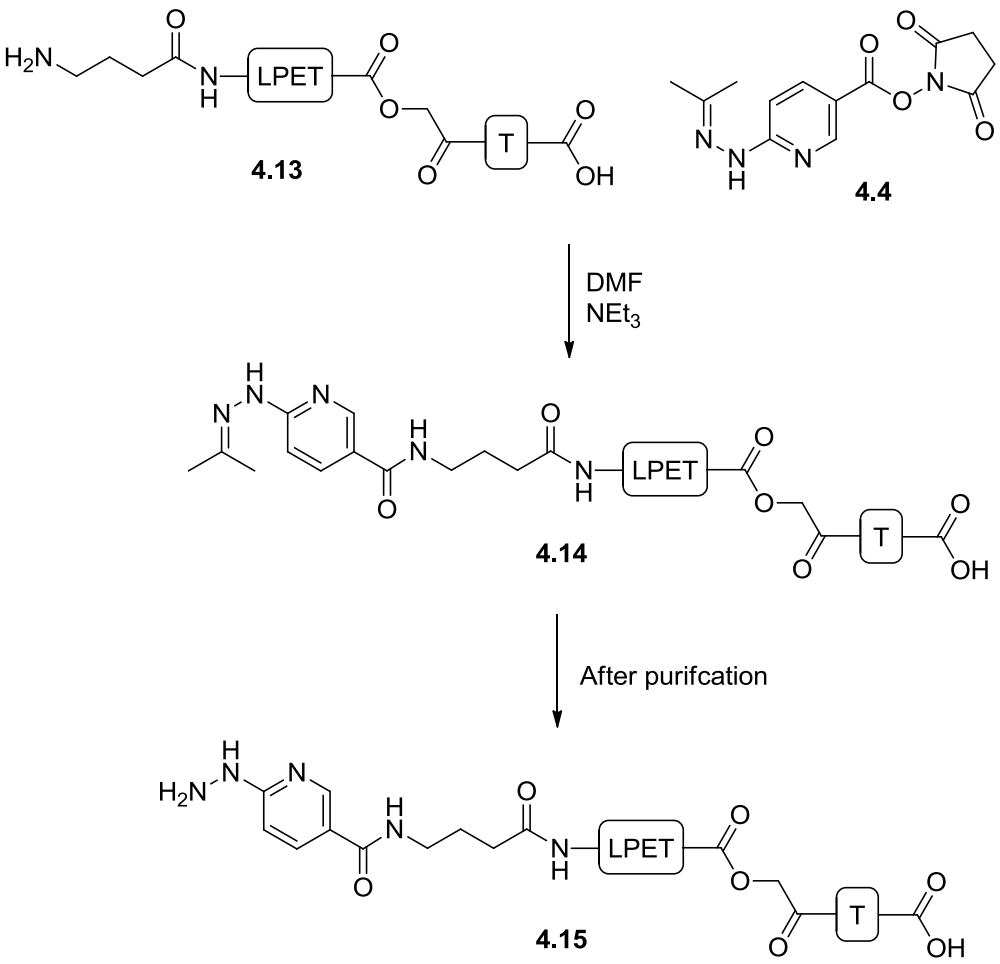


Figure 4.1: Structure of aryl hydrazine **4.11** and polymerised byproducts **4.12** that could form during the SPPS of depsipeptides containing aryl hydrazone **4.4**.

Depsipeptide **4.13** carrying the minimal SrtA recognition sequence was conjugated to aryl hydrazone **4.4** in DMF under basic conditions (Scheme 4.4). The reaction was monitored by LCMS and was typically complete within 2-4 hours. Hydrazone depsipeptide **4.14** was successfully isolated using mass-directed HPLC, but after the fractions were combined and lyophilised, the product had hydrolysed to form hydrazine depsipeptide **4.15**. The ester linkage in the depsipeptide appeared to be stable to the aryl hydrazine functionality and it was predicted to be more reactive to aldehyde compounds than the original hydrazone depsipeptide **4.1**. Therefore, it was used in subsequent experiments.



Scheme 4.4: Synthesis of hydrazine depsipeptide **4.15**.

4.2.1.1 Conjugation of hydrazine onto the protein

The CTA2-AB_5 protein was treated with hydrazine depsipeptide **4.15** and SrtA in a similar manner to previous transpeptidation experiments. The ligation was monitored by ESMS and modification of the protein was observed (Figure 4.2). However, after a prolonged incubation time, the signal of the ligated product decreased by approximately 14 Da. It was postulated that the hydrazine functionality on the pyridine ring was being hydrolysed to generate a 2-hydroxypyridine as this agreed with the change in mass. To further investigate the stability of the hydrazine functionality, the hydrazine depsipeptide **4.15** was incubated in HEPES buffer at 37 °C and LCMS was used at intermittent intervals to monitor any change in its structure. The hydrazine functionality appeared to be stable over a 4 hour period, but after an overnight incubation a small UV absorbance signal corresponding to the 2-hydroxypyridine hydrolysed-depsipeptide was observed. In conclusion, the hydrazine functionality would be sufficiently stable during the time period required for an average SrtA-ligation reaction to reach completion, but it is also likely that the hydrazine linkage would hydrolyse during the subsequent protein

purification steps. Therefore, this class of linkage was deemed unsuitable for further use and an alternative was found.

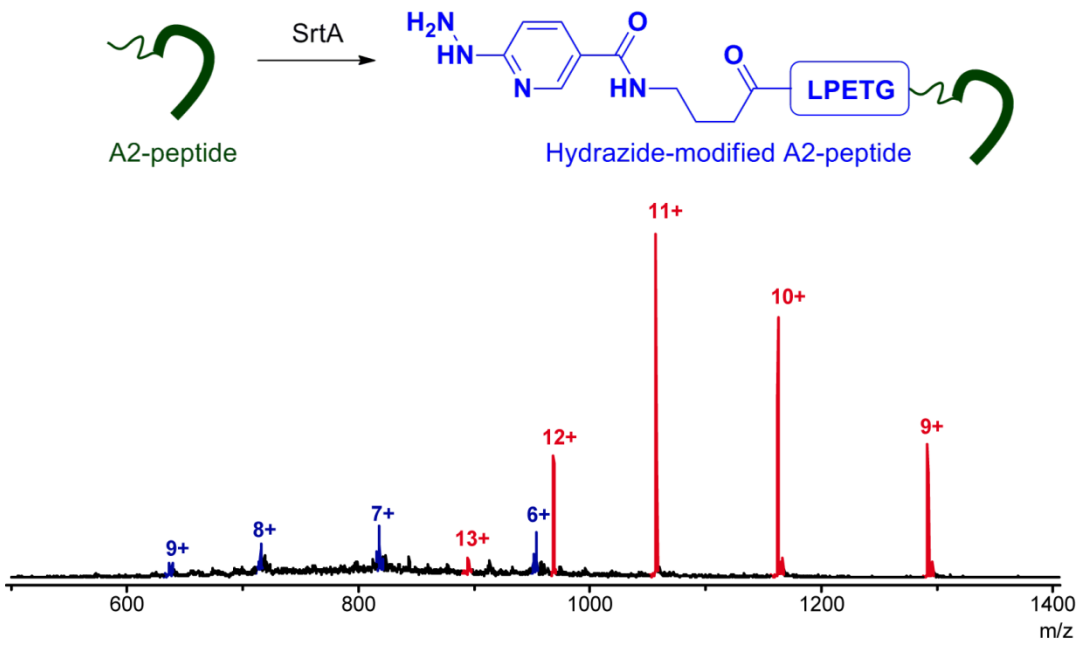
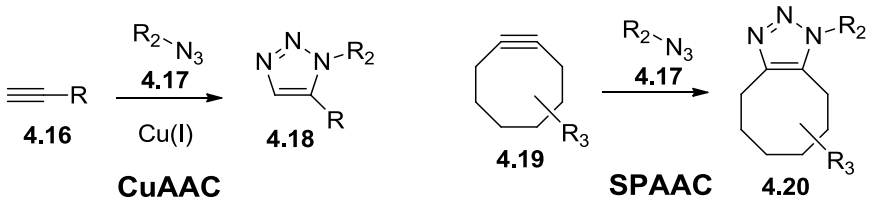


Figure 4.2: Mass spectrum of the hydrazine-linked AB_5 complex. Red: CTB (11611 Da) Blue: CTA2 modified with hydrazine depsipeptide **4.15** (5717 Da). Calculated mass change after SrtA-mediated modification, 661 Da - observed mass change, 662 Da.

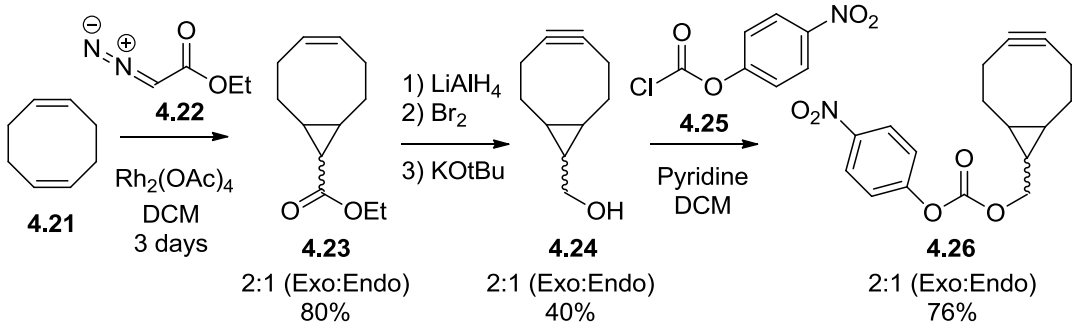
4.2.2 Strained cyclooctyne linker

One of the most important bioorthogonal reactions is the 1,3-dipolar cycloaddition of an azide and alkyne to from a triazole (Scheme 4.5).[157,158] Sharpless and co-workers coined the term "click chemistry" to describe this type of cycloaddition and pioneered the use of the reaction as an effective method to combine a variety of different molecules under relatively mild conditions.[159] However, a toxic Cu(I) catalyst is required in the reaction, which limits its application in many biological systems. An alternative strategy was devised by Bertozzi and co-workers.[160] The authors' approach utilised the properties of a strained cyclooctyne to promote the cycloaddition of an azide without the use of a Cu(I) catalyst; this 1,3-dipolar cycloaddition is commonly referred to as a strained-promoted alkyne-azide cycloaddition (SPAAC) (Scheme 4.5).



Scheme 4.5: Cu-promoted azide-alkyne cycloaddition (CuAAC) between terminal alkyne **4.16** and azide **4.17** forming triazole **4.18.** Strained-promoted alkyne-azide cycloaddition (SPAAC) between strained cyclooctyne **4.19** and azide **4.17** forming triazole **4.20**

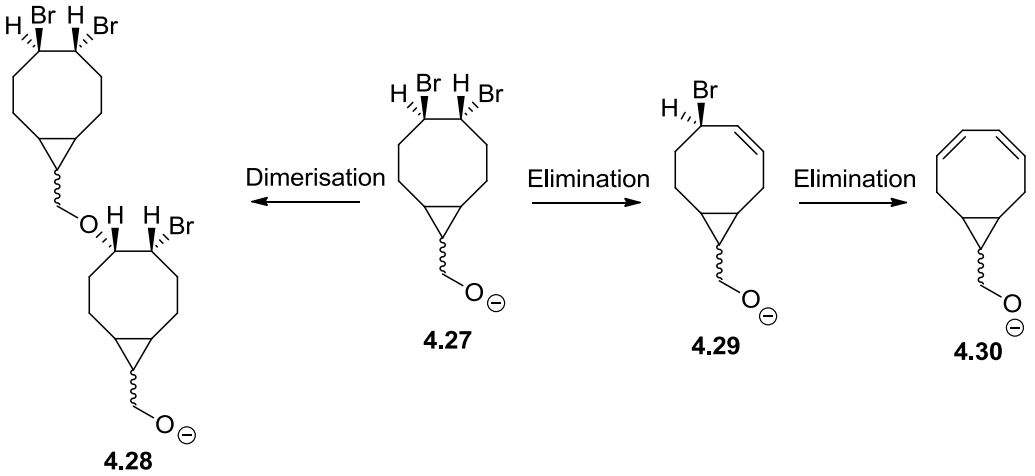
The Cu-promoted azide-alkyne cycloaddition (CuAAC) and SPAAC could both have been used to link oligonucleotides onto the AB_5 protein. However, the CuAAC was avoided as the toxic catalyst might have caused complications *in vivo* if it was not completely removed. Therefore, it was decided to attach a strained cyclooctyne onto the AB_5 protein via a SrtA-mediated ligation. This strategy could have been reversed by incorporating the azide into a depsipeptide substrate, but oligonucleotides functionalized with azides are more common. While these studies were in progress Ploegh and co-workers reported a similar strategy to introduce a strained cyclooctyne into a protein using SrtA-mediated ligations.[161] The authors used SPAACs to create unnatural chimeric proteins linked through their *N*-termini. Several synthetic strategies have now been reported to a range of different strained cyclooctynes;[162] however, these procedures can be non-trivial and often convoluted. For example, the aza-dibenzocyclooctyne used by Ploegh and co-workers is reported to take seven steps to synthesise.[161,163] In contrast, a relatively simple synthetic route to activated cyclooctyne **4.26** (Scheme 4.6) has been developed by Dommerholt *et al.*[164] It was postulated that the activated cyclooctyne **4.26** could be coupled through its carbonate linkage onto a growing depsipeptide during SPPS.



Scheme 4.6: Synthetic route to cyclooctyne **4.26.**

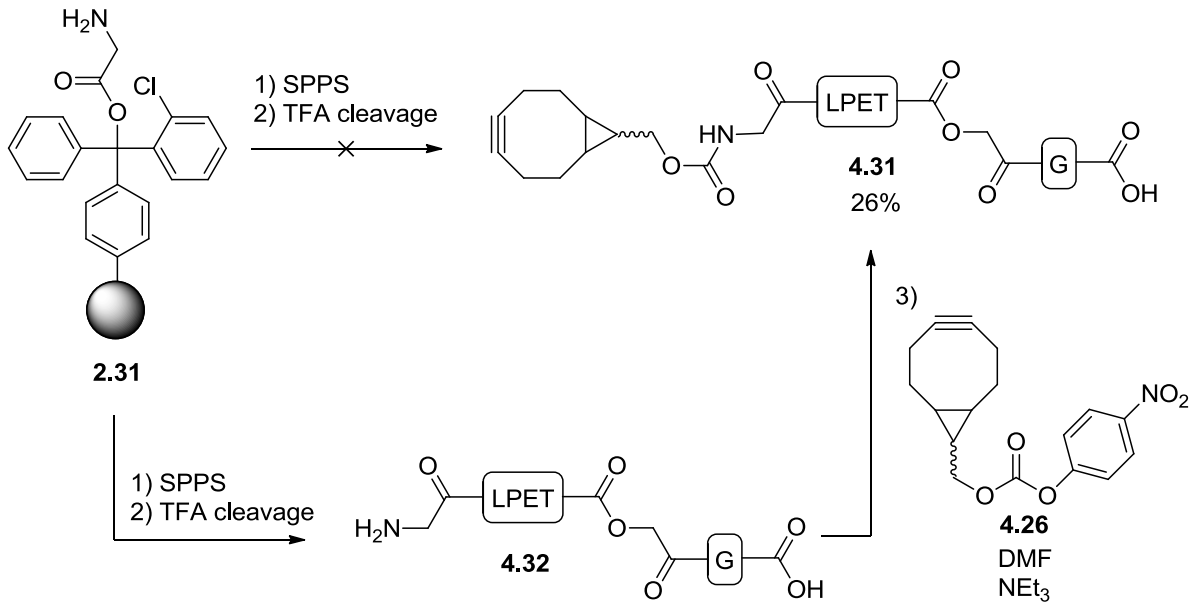
Cyclooctadiene **4.21** was treated with ethyl diazoacetate **4.22** in the presence of a Rhodium(II) catalyst to form bicyclic alkene **4.23** in an 80% yield and a 1:2 endo:exo diastereomeric ratio. Although it was possible to separate the two diastereoisomers, it was more convenient to carry

the mixture through to the next step. The bicyclic alkene **4.23** was reduced, dibrominated and exposed to strong basic conditions to form cyclooctyne alcohol **4.24** in 40% yield across the three steps. The first two steps were practically quantitative, but the yield of the final double elimination reaction varied quite dramatically. Interestingly, decreasing the concentration of the reactants seemed to increase the yield. This suggests that the side reactions taking place were concentration dependent. The hydroxyl group on cyclooctane dibromide **4.27** would be deprotonated under the basic conditions of the reaction, so the corresponding nucleophilic alkoxide could have displaced a bromine atom on another molecule of cyclooctane dibromine **4.27** to form a dimeric species **4.28** (Scheme 4.7). Additionally, elimination could have occurred at a different location on the cyclooctane ring forming an allyl bromine species **4.29**, which is primed to undergo further elimination or nucleophilic substitution. Therefore, lowering the concentration of the reaction reduced the chance of unwanted nucleophilic substitutions, thus increasing the yield. Finally, activated cyclooctyne **4.26** was synthesised in 76% yield by conjugating chloroformate **4.25** onto cyclooctyne alcohol **4.24**.



Scheme 4.7: Theoretical byproducts formed during the synthesis of cyclooctyne alcohol **4.24** from cyclooctane dibromine **4.27**

Activated cyclooctyne **4.26** was used as building block in the SPPS of alkyne depsipeptide **4.31** (Scheme 4.8). However, unconjugated depsipeptide **4.32** was the only product isolated after peptide cleavage, which suggests the activated cyclooctyne **4.26** did not couple onto the growing peptide chain correctly or the carbonate linkage did not survive the cleavage conditions. Therefore, it was decided to change strategy and couple the activated cyclooctyne **4.26** to depsipeptide **4.32** in solution. This approach led to the formation of alkyne depsipeptide **4.31** in a modest 26% yield.



Scheme 4.8: Synthesis of alkyne depsipeptide **4.31**

4.2.2.1 Conjugation of a cyclooctyne onto the AB_5 protein.

The CTA2-AB_5 complex (56 μM) was incubated with alkyne depsipeptide **4.31** (140 μM, 2.5 equivalents) in the presence of SrtA (11 μM, 20 mol%) at 37 °C. After 3 hours, ESMS identified the presence of a significant amount of ligated product and a small signal for the +6 charged state of the unmodified protein (Figure 4.3). In addition, the ESMS also detected the presence of alkyne depsipeptide **4.31** and hydrolysed depsipeptide **4.33**. An additional equivalent of alkyne depsipeptide **4.31** was added to the ligation mixture and incubation was continued for another hour before the protein was purified using Ni-affinity chromatography. SDS-PAGE analysis identified the protein as a band at ~58 kDa on a gel that dissociated upon boiling (Figure 4.4, lanes 4 and 5). This suggests the protein was still stable as an AB_5 complex after the strained cyclooctyne had been attached. Further analysis by ESMS detected a small signal corresponding to unlabelled starting material, but quantifying this by SDS-PAGE was difficult. The AB_5 complex was poorly defined on a gel as a result of it partially dissociating into the pentameric B-subunit and A2-peptide. Therefore, it is not possible to state with complete certainty if it was a single band that belonged to the modified AB_5 complex or two overlapping bands that corresponded to modified and unmodified protein. Although it was not clear how much of the AB_5 complex had been labelled with alkyne depsipeptide **4.31**, the protein was still functional and was deemed suitable for use in subsequent SPAAC experiments

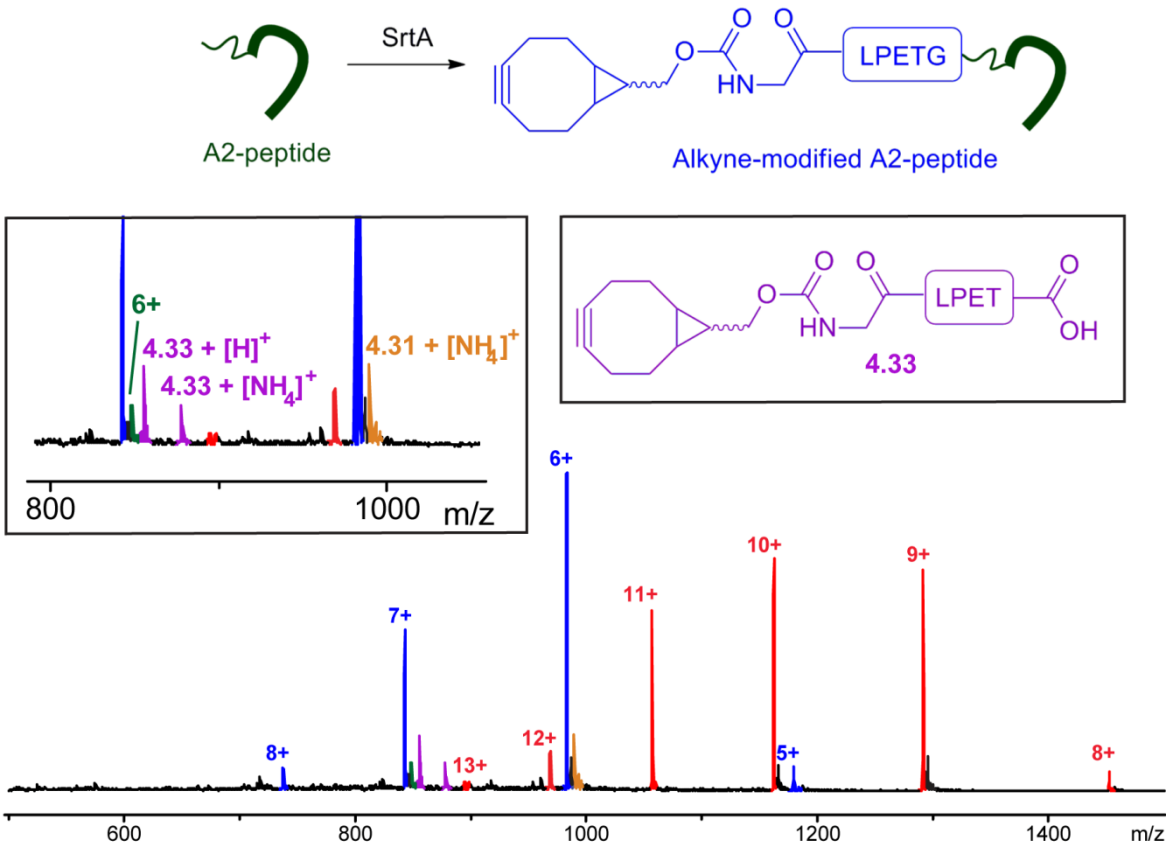


Figure 4.3: Mass spectrum of the alkyne-linked AB_5 complex. Red: CTB (11611 Da); blue: CTA2 modified with alkyne depsipeptide **4.31** (5892 Da); green: Unlabelled CTA2 protein (5055 Da); purple: Adducts of hydrolysed depsipeptide **4.33**; orange: Adduct of alkyne depsipeptide **4.31.** Calculated mass change after SrtA-mediated modification, 865 Da - observed mass change, 837 Da.

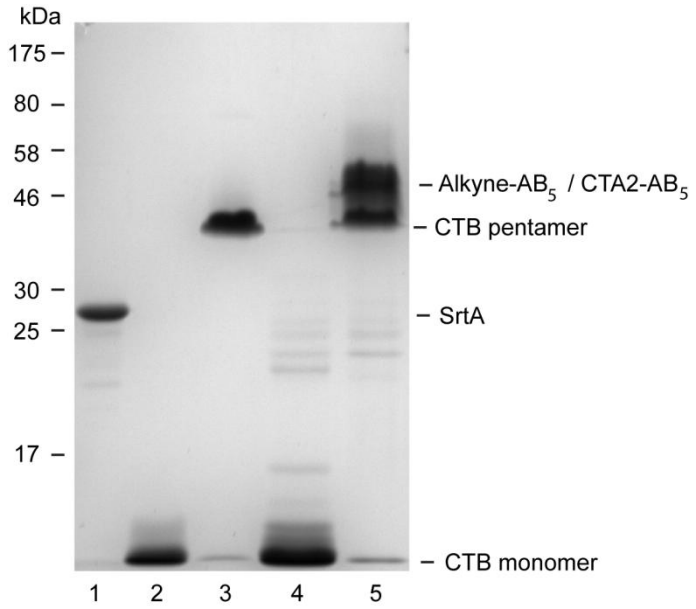
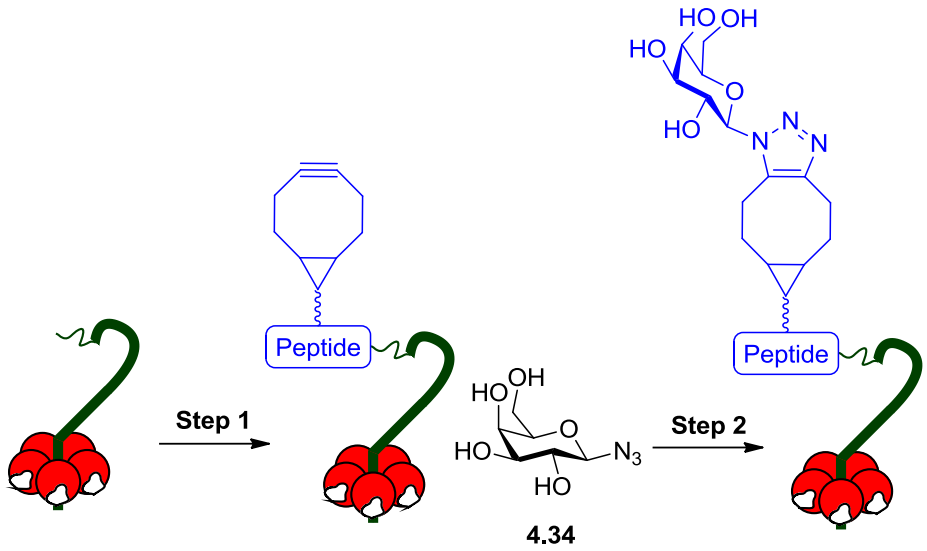


Figure 4.4: Polyacrylamide gel (12%) showing the isolated alkyne-AB_5 complex. Lane 1: SrtA; Lane 2: CTB (boiled); Lane 3: CTB; Lane 4: Alkyne-AB_5 (boiled); Lane 5: Alkyne-AB_5

4.2.3 Strain-promoted cycloadditions with the AB_5 complex

A strategy now existed to site-specifically incorporate a strained cyclooctyne onto the AB_5 complex, but the SPAAC had yet to be tested. Additionally, it was also important to establish if the reaction could be performed directly after the initial SrtA-mediated modification of the AB_5 complex as this could potentially reduce the number of purification steps. Therefore, the efficiency of a sequential SrtA ligation and SPAAC was probed using a simple galactosyl azide **4.34** (Scheme 4.9) acquired from Dr. Tom McAllister (University of Leeds).



Scheme 4.9: Sequential modification of the AB_5 complex. Step 1: SrtA-mediated modification of CTA2-AB_5 complex with alkyne depsipeptide **4.32**. Step 2: SPAAC between alkyne-AB_5 complex and galactosyl azide **4.34**

The CTA2-AB_5 complex (56 μM) and alkyne depsipeptide (138 μM, 2.5 equivalents) were incubated at 37 °C with SrtA (11 μM, 20 mol%). ESMS was used to monitor the ligation and after 3 hours, only a trace amount of unlabelled protein was detected. At this point, galactosyl azide **4.34** (2.25 mM) and an extra equivalent of depsipeptide **4.34** were added to the reaction mixture before incubation was continued at room temperature overnight. The extra depsipeptide substrate was added to reduce the possibility of SrtA hydrolysing the labelled protein during the extended incubation period. After the overnight incubation, ESMS analysis identified the triazole product without any obvious alkyne starting material (Figure 4.5). There also appeared to be a small signal that corresponded to the +6 charge state of the CTA2-fragment, which was either the result of incomplete protein labelling or enzymatic hydrolysis of the ligated product. These results demonstrated that the SPAAC can be successfully performed in HEPES buffer

without having to isolate the alkyne-AB_5 complex from the SrtA-ligation mixture. However, the reaction still needed to be optimised as there was a 9-fold excess of azide to alkyne.

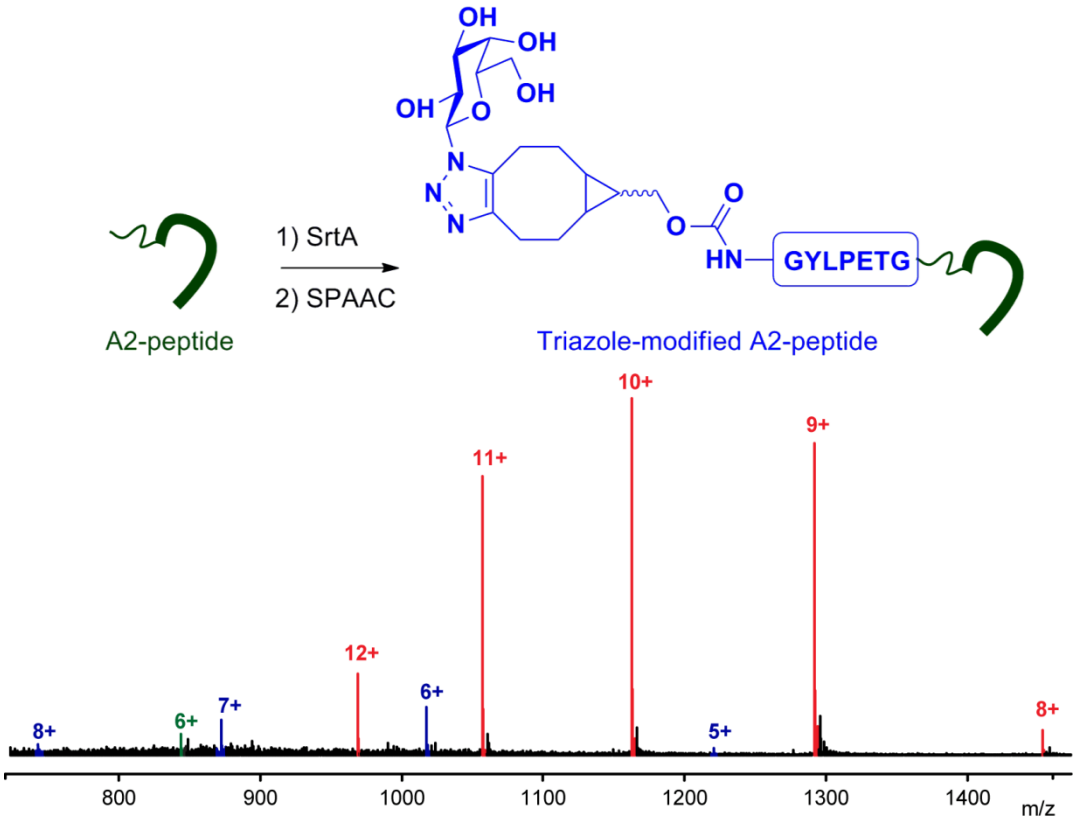


Figure 4.5: Mass spectrum of the triazole-linked AB_5 complex. Red: CTB (11611 Da); blue: CTA2 modified with galactosyl azide **4.34** (6098 Da); green: CTA2 protein (5055 Da). Calculated mass change after SrtA-mediated modification and SPAAC, 1042 Da - observed mass change, 1043 Da.

4.2.4 Oligonucleotide conjugation to AB_5 protein

RNA oligonucleotides containing 2′-methoxy groups, phosphorothioate linkages and a 5′-azido functional group or a 5′-fluorescein fluorophore (Figure 4.6) had been synthesised by Darren Machin (University of Leeds). The 5′-fluorescein oligonucleotide **4.36** is a variant of an intron skipper drug used by GlaxoSmithKline to treat Duchenne Muscular Dystrophy and 5′-azido oligonucleotide **4.35** is its complementary strand.[165] Combining the two oligonucleotides leads to the formation of a duplex structure with an azide functional group and a fluorescein probe. It was postulated that the alkyne-AB_5 protein could be linked to the duplex and the formation of the protein-conjugate could be directly visualised as a fluorescent band on an SDS-PAGE gel.

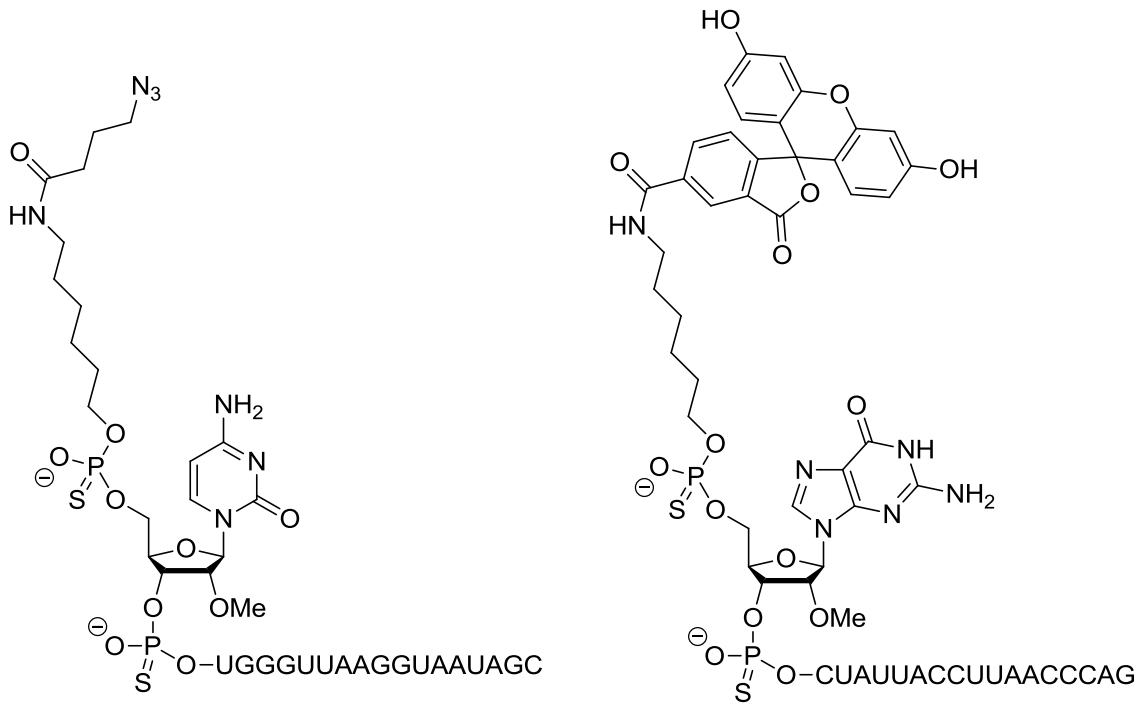


Figure 4.6: Sequence and structure of the 5′-azido complementary and 5′-fluorescein intron skipper oligonucleotide strands **4.35** and **4.36** respectively.

Oligonucleotides **4.35** and **4.36** were combined in equal concentrations, heated to 95 °C and cooled to room temperature to allow formation of the duplex structure. The duplex (100 μM, 3.33 equivalents) was incubated with the alkyne-AB_5 complex (27 μM) at room temperature before SDS-PAGE analysis was performed (Figure 4.7). It is important to note that the initial SrtA-ligation reaction to generate the alkyne-AB_5 complex had not gone to completion (See previous section), so the large band at ~56 kDa on the gel (Lane 3) could actually be two overlapping bands that had not resolved. Therefore, it is possible that the two discrete bands that can be detected at the same position on the gel after the SPAAC (Lane 5) are simply labelled and unlabelled protein from the previous reaction rather than oligonucleotide-conjugated protein. Interestingly, a new faint protein band was identified on the gel at ~24 kDa when the samples were boiled (Lane 6), which was not present in the unboiled sample or starting material (lanes 3 and 5). This suggests the oligonucleotide duplex had been successfully linked to the alkyne-CTA2 fragment; although, the mass of the protein conjugate is ~21 kDa, which is slightly lower than the mass of the new species on the gel. However, as SDS-PAGE is designed primarily to analyse proteins and not oligonucleotides, the observed mass was probably not an accurate representation of the species in solution. The new band was also fluorescent after it was irradiated with UV light (Lane 12), which is compelling evidence that it was duplexed

oligonucleotide conjugated to the alkyne-CTA2 peptide. Furthermore, the fluorescent band shifted up the gel when the sample was not boiled (Lane 11) indicating it was linked to the higher weight AB_5 complex.

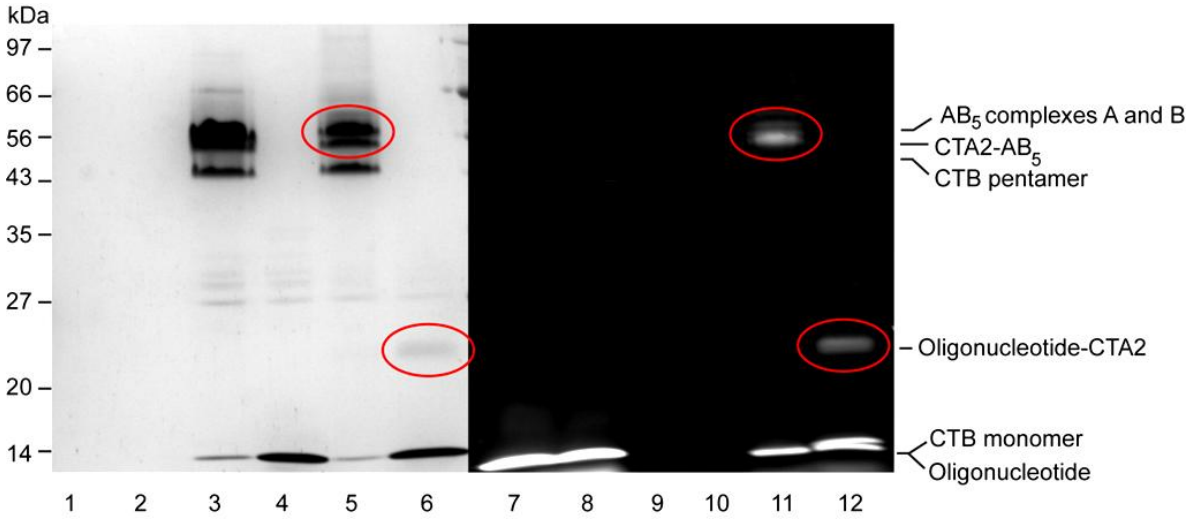


Figure 4.7: Polyacrylamide gel (12%) showing the duplex oligonucleotide conjugation to alkyne-AB_5 protein. Red circle = new protein band. Lane 1: Duplex Oligonucleotide; lane 2: Duplex oligonucleotide (boiled); lane 3: Alkyne-AB_5 lane 4: Alkyne-AB_5 (boiled); lane 5: Reaction mixture; lane 6: Reaction mixture boiled; lane 7-12: UV exposed lanes 1-6. AB_5 complex A = Alkyne-AB_5 B = Oligonucleotide-conjugated AB_5.

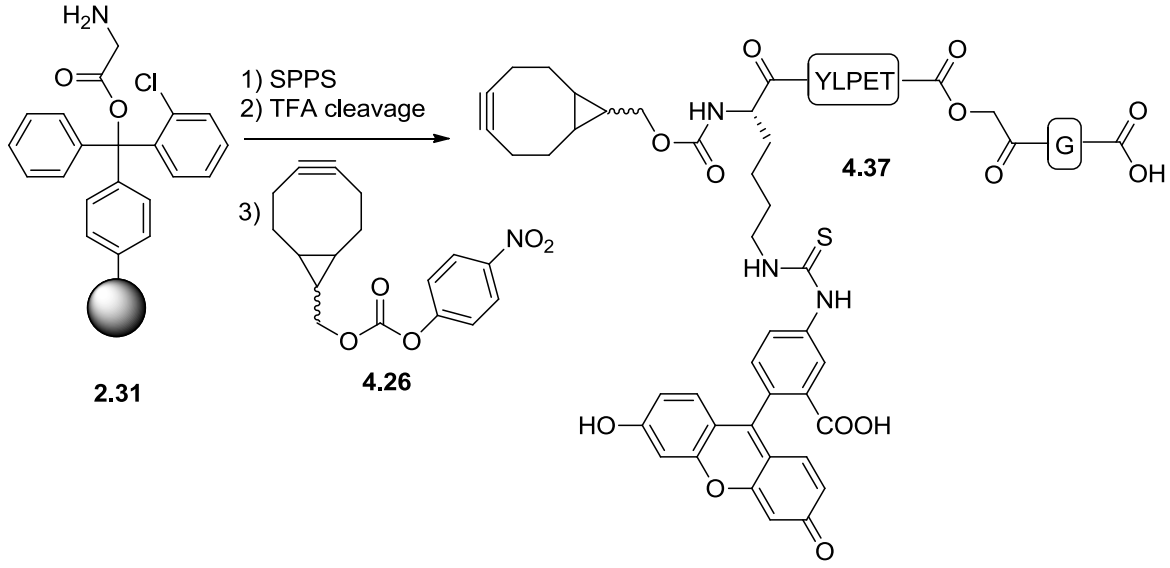
The preliminary results were encouraging, but it is extremely difficult to draw too many conclusions about the efficiency of the SPAAC for a number of reasons. The purity of the alkyne-AB_5 starting material was not known at the beginning of the experiment, so there may have been a significant amount of unlabelled CTA2-AB_5 complex in solution reducing the effective concentration of the alkyne. When the samples were boiled, the alkyne-CTA2 band was either not visible or indistinguishable from the monomeric B-subunit on the gel, so it could not be directly compared to the oligonucleotide-CTA2 band (Lanes 4 and 6). Furthermore, the unboiled AB_5 complex looked practically identical before and after conjugation; once again, making comparison difficult (Lane 3 and 5). However, a very approximate assessment can be made by comparing the intensity of the fluorescent bands. There was roughly three times more oligonucleotide duplex than the AB_5 complex; therefore, if the protein was completely labelled, it should fluoresce at roughly 1/3 the intensity of the excess oligonucleotide duplex, but it seems considerably less intense.

There are several factors that could have reduced the efficiency of the oligonucleotide protein conjugation, but the low concentration of the reactants was probably the primary cause. The SPAAC is a bimolecular reaction, so if the concentration of the azide and alkyne was too low, the rate of the reaction would have been reduced decreasing product formation. Therefore,

increasing the temperature and concentration of the reaction in future experiments may aid the formation of the conjugated product.

4.2.4.1 Redesigning the alkyne-depsipeptide linker

An alternative strategy to identify oligonucleotide conjugation was also being developed. Incorporating a fluorescent group into an alkyne-depsipeptide substrate would allow direct visualisation of the alkyne-labelled protein on an SDS-PAGE gel making any subsequent oligonucleotide conjugation more obvious. Therefore, it was decided to synthesise bifunctional depsipeptide **4.37**, which had both a fluorescein probe and an alkyne group in its structure (Scheme 4.10). The backbone of the depsipeptide, including the fluorescein group, was built using SPPS before it was cleaved and subsequential coupled to activated cyclooctyne **4.26** to produce octyne/fluorescent bifunctional depsipeptide **4.37** in 33% yield.



Scheme 4.10: Synthesis of octyne/fluorescent bifunctional depsipeptide **4.37**.

The CTA2-AB_5 complex (37 μM) was treated with bifunctional depsipeptide **4.37** (110 μM, 3 equivalents) in the presence of SrtA (7.7 μM, 20 mol %). However, modification of the protein proved problematic. ESMS showed the formation of ligated product, but the unmodified protein was the dominant signal even after an extended incubation time with an extra 2 equivalents of bifunctional depsipeptide **4.37** and 30 mol% SrtA. The protein was isolated using lactose-affinity chromatography and its absorbance at 495 nm and 280 nm was measured to determine the fluorescein:protein ratio. The absorbance was extremely weak at 495 nm suggesting the labelling had been negligible. It is quite likely that the bulky substituents at the *N*-terminus of the bifunctional depsipeptide **4.37** were interfering with SrtA-substrate binding and slowing or inhibiting the ligation reaction. If the rate of the ligation reaction was significantly reduced, the

depsipeptide substrate could undergo non-enzymatic hydrolysis before it could interact with SrtA. A simple solution to reduce SrtA-substrate interference would be to include additional amino acid residues between the LPXTG recognition site and the *N*-terminal chemical probes.

4.3 Conclusion

A synthetic route to a depsipeptide substrate with cyclooctyne functionality has been developed for use with SrtA to allow the introduction of a bioorthogonal handle site-specifically into a protein. The SPAAC expands the range of modifications that can be performed on a protein by simply using SrtA-mediated ligations. Furthermore, using SrtA to incorporate the cyclooctyne means only a simple LPXTG recognition site or an N-terminal glycine sequence needs to be incorporated into the protein of interest, unlike other methods that require prior chemical or biological modification to attach the linker.

The alkyne depsipeptide has been successfully coupled onto an AB_5 protein and then conjugated to an azide linked oligonucleotide. However, the SrtA ligation needs to be optimised as the AB_5 complex was not completely modified with the depsipeptide substrate. Further investigation is also required into the SPAAC as currently only a very approximate qualitative assessment of the oligonucleotide-protein bioconjugation has been performed. These issues need to be addressed before this bioorthogonal approach can be used as a viable method to attach oligonucleotides onto the AB_5 complex.

Chapter 5: Conclusions and future work

5.1 Summary

The SrtA-mediated ligation reaction has been optimised with the use of depsipeptide substrates to allow the efficient *N*-terminal labelling of proteins/peptides. A range of depsipeptides has been synthesised by conventional SPPS, using an Fmoc-protected amino acid precursor that was obtained in two simple solution phase steps. The proteins hMBL, mouse pumilio and the CTA2-AB_5 complex have been modified with 1.5-5 equivalents of depsipeptide demonstrating the versatility of this approach. However, the efficiency of the ligation is dependent on the concentration of the protein substrate as non-enzymatic hydrolysis of the depsipeptide can compete with the reaction at low concentrations.

A new methodology has been developed to generate an AB_5 complex that is primed for SrtA-mediated modification. This was achieved by overexpressing a chimeric MBP-AB_5 complex that was subsequently truncated with TEV protease to leave an activated protein for SrtA ligation. The AB_5 complex has been derivatised at the *N*-terminus of the CTA2 domain with fluorescent probes and bioorthogonal functionalities. Crucially, when mammalian cells were treated with a fluorescein-labelled AB_5 protein, translocation of the complex across the cell membrane barrier to intracellular targets was observed. It is reasonable to assume that further modifications to the protein at the *N*-terminus of the CTA2 domain will not interfere with its ability to be endocytosed. Therefore, a wide selection of biological probes could potentially be transported inside cell using this approach.

The SrtA-mediated modification of the AB_5 complex is limited to molecules that can be coupled to depsipeptide substrates. To increase the number of potential modifications that can performed, a strained alkyne was linked to a depsipeptide and site-specifically incorporated onto the protein. This allowed an azide-linked RNA therapeutic to be attached onto the AB_5 complex using a SPAAC. Although the efficiency of the conjugation reaction needs to be optimised, it has successfully generated an oligonucleotide-protein conjugate under mild conditions.

5.2 Future work

Now that a simple and robust SrtA-mediated labelling strategy has been developed to label AB_5 complexes there are several avenues of research that can be investigated.

5.2.1 Development and application of RNA-conjugated AB_5 complexes

The immediate priority is to optimise the RNA conjugation as only a small amount of the RNA-AB_5 conjugate appears to have been formed; although, this has not been quantified. There are

several variables that could be altered to improve the yield of the bioconjugation reaction. These include temperature, buffer identity and reactant concentration; as the SPAAC is a bimolecular reaction, increasing the reactant concentration should immediately boost the efficiency of the bioconjugation.

The ability of the RNA-protein conjugates to enter mammalian cells needs to be assessed using cell culture and immunofluorescence. In addition, an investigation into the efficacy of the therapeutic RNA *in vivo* compared to that of the unconjugated RNA is also required. Issues regarding immune response to the oligonucleotide-protein conjugates could be a potential stumbling block in the development of this technology. If *in vivo* testing is successful, this methodology could be applied to other oligonucleotide therapeutics that are unable to enter mammalian cells.

5.2.2 Controlled release of cargo from AB_5 protein conjugates

Currently the AB_5 delivery vehicles are permanently linked to their molecular cargo. Introducing a labile linker into the CTA2 peptide either synthetically or genetically would allow the controlled release of the cargo within a mammalian cell. This could alter how the cargo is trafficked and may allow it to reach different intracellular targets to the parent protein. There are several strategies, such as photocleavable and pH sensitive linkers, that could be used to achieve this, but incorporating a disulfide bond is probably the most practical.

5.2.2.1 Disulfide linkers

The A1-domain and A2-domain of cholera toxin are linked by a single disulfide bride that is reduced by a PDI in the ER of mammalian cells. The A1-domain is then liberated from the rest of the protein enabling it to induce toxicity. In principle, this mechanism could be exploited to release covalently-linked cargo from AB_5 delivery vehicles. This could be achieved by simply introducing a disulfide bond and an endoprotease cleavage site, such as Factor Xa, into the CTA2-peptide sequence (Figure 5.1). The original SrtA-mediated modification strategy could then be used to label the AB_5 complex, but prior to cell internalisation the extended A2-domain would be treated with Factor Xa. The enzyme will cleave the CTA2-domain into two peptide fragments that are held together by a single disulfide bond. After the AB_5 complex has been transported inside a mammalian cell, the PDI will reduce the disulfide bond releasing the cargo from the parent protein.

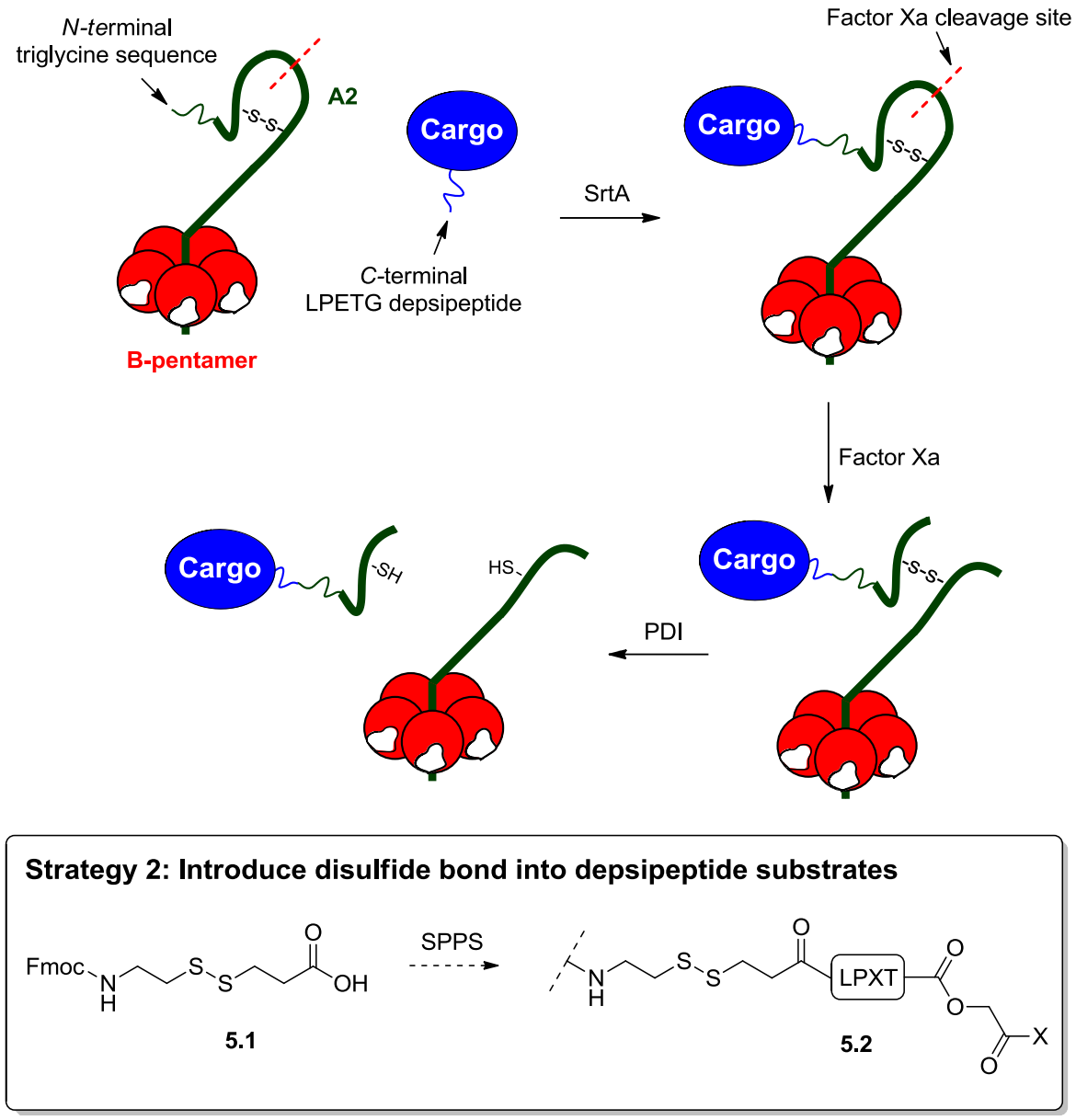


Figure 5.1: Introduction of a disulfide bridge into the AB_5 complex to PDI-mediated cargo release. Strategy 1: Reengineer protein. Strategy 2: Synthesis of depsipeptide substrate **5.2** from unnatural amino acid **5.1**. X = any amino acid.

Alternatively, a disulfide bond could be synthetically incorporated into an AB_5 complex without having to reengineer the protein. The synthesis of an Fmoc-protected disulfide amino acid **5.1** (Figure 5.1) has been reported in previous scientific literature.[166] Depsipeptide substrates **5.2** containing the unnatural amino acid **5.1** could be readily synthesised using SPPS and ligated onto the AB_5 complex. This would introduce a synthetic disulfide bond into the protein that would still be reduced by PDI allowing the conjugated cargo to be released inside mammalian cells.

5.2.3 Anticipated problems with bacterial toxin based delivery systems.

Although this thesis has demonstrated that modified AB_5 complexes are able to enter mammalian cells, there may be problems regarding their use as therapeutic delivery agents.

5.2.3.1 Endosomal entrapment of the delivery vehicle

Once cholera toxin has entered a target cell, the catalytic A1-domain is able to reach the cytosol using the sec61p channel in the ER;[167] this channel is used by cells to remove damaged and unfolded proteins from the ER.[168] Non-protein cargo that has been delivered and subsequently released from modified AB_5 complexes (see above) will be unable to use this mechanism to enter the cytosol. Consequently, the efficacy of the cargo could be severally limited if it is unable to reach its intracellular target.

Other classes of delivery vehicle also suffer similar limitations. For example, CPPs frequently become trapped in endosomal compartments after internalisation;[31,33] however, the introduction of repeating membrane destabilising peptide sequences, such as KALA, can help trigger endosomal escape. Therefore, it may be possible to use this approach to disrupt the endosomes that transport the modified AB_5 complexes inside a cell. The peptide sequences could be installed at the *N*- and/or *C*-termini of the pentameric B-subunits to create a network of membrane destabilising motifs that could disrupt a vesicle membrane, thus allowing the cargo to enter the cytosol.

5.2.3.2 Development of stealth delivery vehicle

The immunogenicity of AB_5 delivery vehicles may also be a potential hurdle in its development as a clinical tool. It has been widely reported that cholera toxin can act as a powerful adjuvant.[169,170] Therefore, there is a strong possibility that modified AB_5 complexes could only be administered to a subject a limited number of times without inducing a massive immune response. Gene delivery agents, such as adenoviral vectors,[171] have been modified with PEG groups to mask them from the immune system;[172] this approach could be used to reduce the immunogenicity of the AB_5 delivery vehicles. Succinimidyl PEG groups could be non-specifically conjugated to the surface exposed lysine residues on the pentameric B-subunit (Figure 5.2). This would suppress the antigenic features present on the surface of the AB_5 complex and as a result make detection by the immune system more difficult.

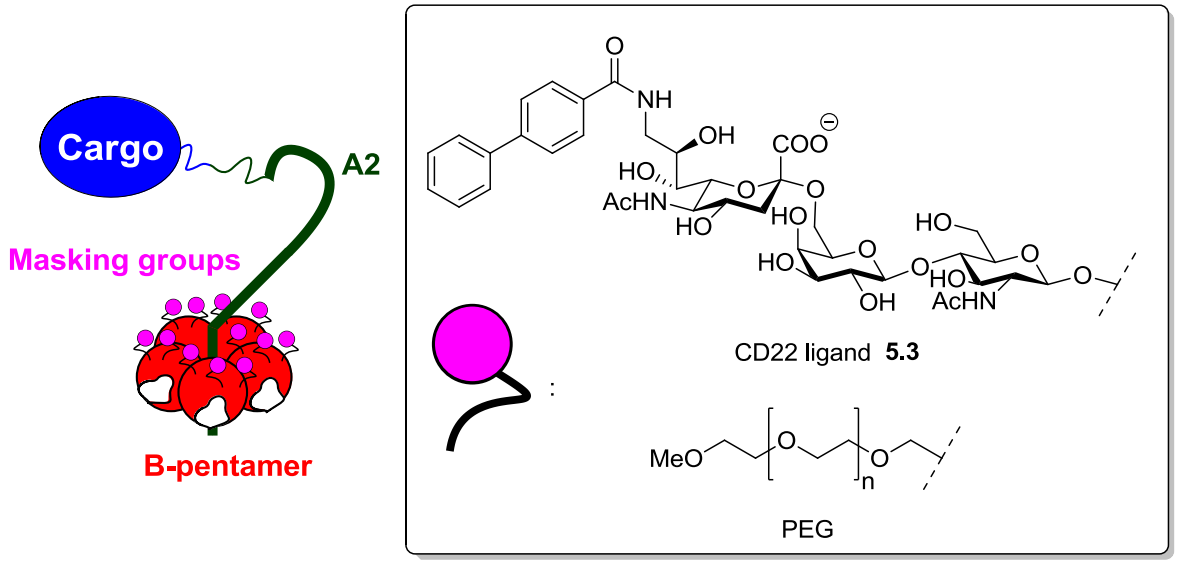


Figure 5.2: Modification of AB_5 complex with functional groups to mask the immune system. CD22 ligand **5.3** and PEG.

Paulson and co-workers devised a more elegant solution to mask liposomal delivery vehicles from the immune system.[173] The authors decorated the surface of their liposomes with various protein antigens, and synthetic carbohydrate ligands **5.3** that are recognised by the CD22 receptors on B-cells. The CD22 receptors regulate the activity of B-cells by inducing apoptosis upon activation. The B-cells are brought into close proximity with the carbohydrate ligands **5.3** when they detect the antigens on the surface of the liposomes. Once in close enough proximity, the carbohydrate ligands **5.3** stimulate the CD22 receptors to trigger apoptosis of the cell. The authors' ingenious approach specifically reduces the number of antigen-active B-cells that recognise the functionalised liposomes and it has been demonstrated both *in vitro* and *in vivo*. Once again, this technology could be adopted for use with the AB_5 delivery vehicles.

Chapter 6: Chemical and biological experimental

6.1 Chemical experimental

6.1.1 General methods

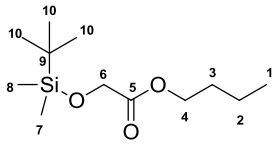
Commercial reagents were used without purification, unless otherwise stated. Analytical grade reagents were supplied by Sigma-Aldrich, Fisher Scientific, Melford laboratories and VWR International. All concentrations were performed *in vacuo*, unless otherwise stated. All reactions were carried out at room temperature unless stated otherwise

^{1}H NMR spectra were recorded using a Bruker Avance 500 instrument at 500 MHz or on a Bruker Avance 400 instrument at 400 MHz. ^{13}C NMR spectra were recorded on a Bruker Avance 300 instument at 75 MHz or a Bruker Avance 500 instrument at 125 MHz. Coupling constants (J) are given in MHz. Chemical shifts are given in ppm downfield from tetramethylsilane (TMS). The following abbreviations are used in 1H NMR analysis: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, dd = double doublet, dt = double triplet, ddd= double double doublet etc. All peptides are assigned according to IUPAC recommendations.[174]

Low resolution electrospray (ES+) ionisation mass spectra were obtained on a Bruker HCT Ultra mass spectrometer, and high resolution (ES+) were performed on a Bruker Daltonics MicroTOF mass spectrometer. HPLC analysis was done using the Agilent technologies 1290 infinity. Freeze drying of compounds was done using Virtis Benchtop K freeze dryer. Centrifugation was achieved using the Eppendorf Centrifuge 5810.

6.1.2 Small molecule synthetic data

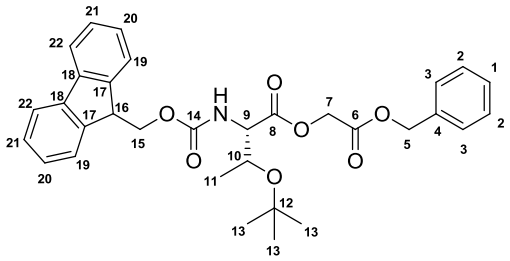
_n_-Butyl [(tert-Butyldimethylsilyl)oxy]acetate 2.22



TBSMCl (7.9 g, 53 mmol) was added to a stirred solution of butyl glycolate (5.8 mL, 45 mmol) and imidazole (8.9 g, 132 mmol) in DMF (40 mL) at 0 °C. The reaction mixture was stirred for 15 minutes at 0 °C and then a further 2.5 h at r.t. before being diluted with EtOAc (250 mL) and washed with 1 N HCl (2 × 200 mL), water (200mL) and brine (200 mL). The organic layer was concentrated and the crude product purified by flash column chromatography (silica; 95:5 (*v*/*v*) hexane- EtOAc) to afford **2.22** as a clear oil (8.95 g, 78%). R_F 0.79 (1:1 (*v*/*v*) hexane-EtOAc); ^{1}H NMR (500 MHz,CD_3OD) δ = 4.21 (2H, s, H_6), 4.12 (2H, t, $J_{H4\text{-}H3}$ 6.7 Hz, H_4), 1.66-1.56 (2H, m, H_3), 1.36 (2H, app hex, $J_{H2\text{-}H3/1}$ 7.4 Hz, H_2), 0.91 (3H, t, $J_{H1\text{-}H2}$ 7.4 Hz, H_1), 0.90 (9H, s,

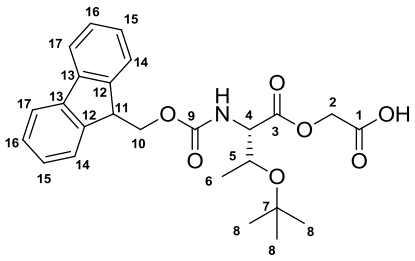
H_{10}), 0.08 (6H, s, $H_{7/8}$). ^{13}C NMR (125 MHz, CD_3OD): δ = 171.9 (CO), 64.7 (C_4), 61.9 (C_6) 30.8 (C_3), 25.8 (C_{10}), 19.2 (C_2), 18.5 (C_9), 13.8 (C_1), -5.4 ($C_{8/7}$). The NMR spectra were in good agreement with the previously reported literature values.[129]

Benzyloxycarbonylmethyl-(2S,3R)-3-*tert*butyloxy-2-(((9H-fluoren-9-ylmethoxy)carbonyl)amino)butanoate 2.29



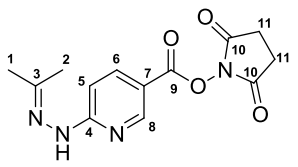
DIPEA (1.9 mL, 11 mmol) was added to a stirred solution of Fmoc-Thr(OtBu)-OH (4.0 g, 10 mmol), benzyl 2-bromoacetate (2.4 mL, 15 mmol) and TBAI (1.4 g, 4 mmol) in THF (10 mL). The reaction mixture was stirred at r.t. for a few minutes before the solution turned yellow and a precipitate started to form. The reaction mixture was left stirring overnight, before it was diluted with H_2O (200 mL) and extracted with EtOAc (2 × 200 mL). The organic layers were combined, washed with 10% sodium thiosulfate (400 mL), H_2O (400 mL) and concentrated to leave an oil. The crude product was purified by flash column chromatography (silica; 4:1 (*v*/*v*) hexane- EtOAc) to afford **2.29** (5.2 g, 95%) as a glassy solid, R_F 0.53 (2:1 (*v*/*v*) hexane-EtOAc); $[\alpha]_D^{22.7}$ -0.4 (*c* 1, DCM); ^{1}H NMR (500 MHz,$CDCl_3$) δ = 7.77 (2H, d, $J_{H19\text{-}H20}$ 6.2 Hz, H_{22}), 7.65-7.63 (2H, m, H_{19}), 7.41-7.39 (2H, m, H_{21}), 7.38-7.29 (2H, m, H_{20}), 7.38-7.29 (5H, m, $H_{1\text{-}3}$), 5.21 (2H, s, $H_{5/5'}$), 4.76 (1H, d, $J_{H7\text{-}H7'}$ 15.9 Hz, $H_{7'}$), 4.66 (1H, d, $J_{H7\text{-}H7'}$ 15.9 Hz, H_7), 4.44-4.35 (1H, m, H_9), 4.44-4.35 (2H, m, $H_{15/15'}$), 4.32-4.25 (1H, m, H_{16}), 4.32-4.25 (1H, m, H_{10}), 1.26 (3H, d, $J_{H11\text{-}H10}$ 6.2 Hz, H_{11}), 1.15 (9H, s, H_{13}). ^{13}C NMR (125 MHz, $CDCl_3$): δ = 170.6 (CO), 167.1(CO), 156. 8 (CO), 144.2 (ArC), 141.4 (ArC), 128.8 (ArC), 128.7(ArC), 128.5(ArC), 127.8 (ArC), 127.2 (ArC), 125.2 (ArC), 120.1 (ArC), 79.4 (C_{12}), 67.4 (C_{15}), 67.4 (C_5), 67.3 (C_{10}), 61.4 (C_7), 59.9 (C_9), 47.2 (C_{16}) 28.5(C_{13}), 21.0 (C_{11}). **IR** (ν_{max}/ cm $^{-1}$) 3440 (NH); 2976 (CH); 1755, 1725, 1648 (C=O). **HRMS:** Found $[M+Na]^+$ 568.2321, $C_{32}H_{35}NO_7Na$ requires 568.2306.

Benzyloxycarbonylmethyl-((2S,3R)-3-*tert*butyloxy-2-(((9H-fluoren-9-ylmethoxy)carbonyl) amino)butanoyloxy)acetic acid 2.30



Benzyloxycarbonylmethyl-(2S,3R)-3-*tert*butyloxy-2-(((9H-fluoren-9-ylmethoxy)carbonyl)amin o)b utanoate **2.29** (1 g, 1.8 mmol) was dissolved in THF and H_2O (3:1, 21 mL), before Pd/C (10%, 50 mg) was added. Under a H_2 atmosphere, the reaction mixture was stirred at r.t for 40 minutes before being filtered through celite and concentrated to leave a clear oil. The crude product was purified by flash column chromatography (silica; 9:1 (v/v) DCM-methanol) to afford **2.30** (726 mg, 87%) as a white lyophilisate, R_F 0.63 (9:1 (v/v) DCM- EtOAc and 1% acetic acid); $[\alpha]_D^{22.5}$ -0.5 (*c* 1, DCM); ^{1}H NMR (500 MHz, $CDCl_3$) δ = 7.65 (2H, d, $J_{H17\text{-}H16}$ 7.6 Hz, H_{17}), 7.57-7.54 (2H, m, H_{14}), 7.29-7.26 (2H, m, H_{16}), 7.21-7.19 (2H, m, H_{15}), 4.55 (1H, d, $J_{H2\text{-}H2'}$ 14.4 Hz, H_2), 4.44-4.29 (1H, m, $H_{2'}$), 4.44-4.29 (2H, m, $H_{10/10'}$), 4.30-4.22 (1H, m, H_4), 4.22-4.11 (1H, m, H_{11}), 4.22-4.11 (1H, m, H_5), 1.09 (3H, d, $J_{H6\text{-}H5}$ 5.9 Hz, H_6), 1.04 (9H, s, H_6) ^{13}C NMR (125 MHz, $CDCl_3$): δ = 174.1 (CO), 170.5 (CO), 157.6 (CO), 144.3 (ArC), 143.9 (ArC) 141.3 (ArC), 127.6 (ArC), 127.0 (ArC), 125.3 (ArC), 119.9 (ArC), 74.2 (C_7), 67.8 (C_{10}), 67.0 (C_5), 63.3 (C_2), 60.4 (C_4), 47.3 (C_{11}), 28.5 (C_8), 21.0 (C_6) **IR** (ν_{max}/ cm $^{-1}$) 3436 (NH); 3436 (OH); 2977 (CH); 1749, 1720, (C=O). **HRMS:** Found $[M+Na]^+$ 478.1837, $C_{25}H_{29}NO_7Na$ requires 478.1836.

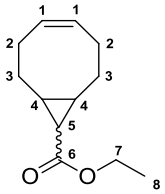
2,5-Dioxo-1-pyrrolidinyl 6-(isopropylidenehydrazino)nicotinate 4.4



6-Choronicotinic acid (1.9 g, 12.1 mmol) was dissolved in hydrazine monohydrate (10 mL). The reaction mixture was stirred under reflux for 6 h and allowed to cool to r.t. before being concentrated under vacuo to leave a yellow solid. The crude product was dissolved in H_2O (30 mL) and the pH of the stirred solution was adjusted to pH 5 triggering precipitate formation. The precipitate was filtered and washed with EtOH (2 × 10 mL) to afford hydrazine salt **4.10** as a yellow solid (1.2 g, 52%). Without further purification hydrazine salt **4.10** (1.05 g, 5.5 mmol) was added at r.t. to stirred solution of acetone (0.564 mL) and Et_3N (4 mL) in anhydrous DMF (20 mL). After 3.5 h EDC (1.27 g, 6.6 mmol) and N-hydroxysuccinimide (759 mg, 6.6 mmol)

were added to the reaction mixture and it was stirred at r.t. overnight. The reaction mixture was diluted with H_2O (50 mL) and extracted with CH_2Cl_2 (3 × 50 mL). The organic layers were combined, washed with brine (150 mL) and concentrated under vacuo to leave a yellow solid. The crude product was precipitated from MeCN to afford **4.4** as an off white amorphous solid (0.740 g, 46%). δH (500 MHz, DMSO-d6) δ = 10.38 (1H, s, NH), 8.79 (1H, dd, $J_{H8\text{-}H6}$ 2.5 Hz, $J_{H8\text{-}H5}$ 0.7 Hz, H_8), 8.14 (1H, dd, $J_{H6\text{-}H5}$ 9.0 Hz, $J_{H6\text{-}H8}$ 2.5 Hz, H_6), 7.20 (1H, dd, $J_{H5\text{-}H6}$ 9.0 Hz, $J_{H5\text{-}H8}$ 0.7 Hz, H_5), 2.90 (4H, s, H_{11}), 2.03 (3H, s, $H_{1/2}$), 2.01 (3H, s, $H_{1/2}$). ^{13}C NMR (125 MHz, DMSO-d6): δ = 170.5 (C_{10}), 161.1 (C_4), 169.9 (C_9), 152.3 (C_3), 151.6 (C_8), 138.8 (C_6), 110.0 (C_7), 106.0 (C_5), 25.5 (C_{11}), 25.2 ($C_{1/2}$), 17.4 ($C_{1/2}$). **HRMS:** Found $[M+H]^+$ 291.1075, $C_{13}H_{15}N_4O_4$ requires 291.1088. The compound spectra were in good agreement with the previously reported literature values.[156]

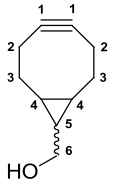
(1R,8S,9r,Z)-Ethyl bicyclo[6.1.0]non-4-ene-9-carboxylate (*exo*) and (1R,8S,9s,Z)-ethyl bicyclo[6.1.0]non-4-ene-9-carboxylate (*endo*) 4.23



1,5-Cyclooctoadiene (19.6 mL, 160 mmol), rhodium(II) acetate (380 mg. 0.86 mmol) and DCM (10 mL) were stirred under an atmosphere of N_2 before ethyl diazoacetate (85% in DCM, 2.4 mL, 20 mmol) as a solution in DCM (10 mL) was added dropwise at r.t over a 3 h period. The reaction mixture was left stirring for 3.5 days at r.t. before being concentrated *in vacuo* and passed through a plug of silica (1:1 (*v*/*v*) hexane-EtOAc) to remove the excess 1,5-cyclohexadiene. The crude product was purified by flash column chromatography (silica; 20:0.5 (v/v) hexane-EtOAc) to afford **4.23** as clear oil (3.1g, 80%, 2:1 exo:endo ratio). **Exo**: R_F 0.32 (20:1 (v/v) hexane-EtOAc). ^{1}H NMR (500 MHz,$CDCl_3$) δ = 5.59-5.58 (2H, m, H_1), 4.10 (2H, q, $J_{H7\text{-}H8}$ 7.1 Hz, H_7), 2.37-2.35 (1H, m, H_2), 2.24-2.14 (2H, m, H_3), 2.13-2.02 (2H, m, $H_{2'}$), 1.60-1.52 (2H, m, H_4), 1.54-1.42 (2H, m, $H_{3'}$) 1.25 (3H, t, $J_{H7\text{-}H8}$ 7.1 Hz, H_8), 1.18 (1H,d, $J_{H5\text{-}H4}$ 4.2 Hz, H_5) ^{13}C NMR (125 MHz,$CDCl_3$): δ = 174.6 (C_6), 130.1 (C_1), 60.4 (C_7), 28.4 (C_3), 28.0 (C_5), 27.9 (C_4), 26.8 (C_2), 14.4 (C_8). **Endo**: R_F 0.42 (20:1 (v/v) hexane-EtOAc). ^{1}H NMR (500 MHz,$CDCl_3$) δ = 5.63-5.54 (2H, m, H_1), 4.10 (2H, q, $J_{H7\text{-}H8}$ 7.1 Hz H_7), 2.55-2.45 (2H, m, H_2), 2.26-2.12 (2H, m, H_3), 2.09-2.00 (2H, m, $H_{2'}$), 1.60-1.52 (2H, m, $H_{3'}$), 1.69 (1H, t, $J_{H5\text{-}H4}$ 8.8 Hz H_5), 1.44-1.32 (2H, m, H_4) 1.25 (3H, t, $J_{H8\text{-}H7}$ 7.1 Hz, H_8) ^{13}C NMR (125 MHz,$CDCl_3$): δ = 172.4 (C_6), 129.6 (C_1), 59.8 (C_7), 27.2 (C_2), 22.8 (C_4), 21.4 (C_3), 26.8 (C_5), 14.5 (C_8). **IR** (v_{max}/ cm $^{-1}$) 2979 (CH); 1719.12 (C=O). **HRMS:** Found $[M+H]^+$ 195.1371, $C_{12}H_{19}O_2$ requires

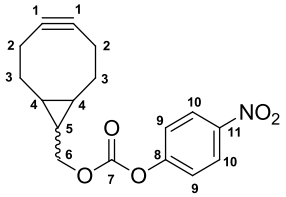
195.1380. The compound spectra were in good agreement with the previously reported literature values.[164]

(1R,8S,9s)-Bicyclo[6.1.0]non-4-yn-9-ylmethanol (*exo*) and (1R,8S,9r)-Bicyclo[6.1.0]non-4-yn-9-ylmethanol (*endo*) 4.24



Under a N_2 atmosphere, a solution of **4.23** (800 mg, 4.1 mmol) in DCM (15 mL) was added dropwise to a stirred suspension of lithium aluminium hydride (135 mg, 3.5 mmol) in DCM (15 mL) at 0 °C. The suspension was stirred at r.t. for 25 minutes before being cooled to 0 ˚C and quenched with the addition of sodium sulfate hexahydrate (3 g). The reaction mixture was filtered through celite, concentrated and azeotroped with toluene to leave a clear oil (660 mg). Under a N_2 atmosphere, the crude product was dissolved in DCM (15 mL) with stirring at 0 ˚C before a solution of Bromine (265 μL, 5.2 mmol) in DCM (2.5 mL) was added dropwise until a persistent red colour was observed. The reaction was quenched with the addition of 10% sodium thiosulfate (40 mL) before the organic layer was separated; washed with H_2O (100 mL) and brine (100 mL); concentrated and azeotroped with toluene to leave a white solid. Under a N_2 atmosphere, the crude product was dissolved in THF (50 mL) with stirring at 0 ˚C before a solution 1 M potassium tert-butoxide in THF (12 mL) was added dropwise. After the addition the reaction mixture was heated to reflux for 5.5 h before being cooled to r.t., quenched with saturated ammonium chloride solution (50 mL) and extracted with DCM (3 × 100 mL). The organic layers were combined and concentrated to leave a crude residue that was purified using flash chromatography (silica; 1.5:1 (v/v) hexane-EtOAc) to afford **4.24** as clear oil. (242 mg, 40%, 2:1 exo: endo ratio). **Exo** R_F 0.33 (1:1 (v/v) hexane-EtOAc).^{1}H NMR (500 MHz,$CDCl_3$) δ = 3.55 (2H, d, $J_{H6\text{-}H5}$ 6.3 Hz, H_6), 2.45-2.36 (2H, m, H_3), 2.34-2.12 (4H, m, $H_{2/2'}$), 1.45-1.29 (2H, m, $H_{3'}$), 0.74-0.63 (1H, m, H_5), 0.74-0.63 (2H, m, H_4).^{13}C NMR (125 MHz,$CDCl_3$): δ = 98.8 (C_1), 67.2 (C_6), 33.4 (C_3), 27.3 ($C_{4/5}$), 22.6 ($C_{4/5}$), 21.5 (C_2). **Endo**: R_F 0.33 (1:1 (v/v) hexane-EtOAc). ^{1}H NMR (500 MHz,$CDCl_3$) δ = 3.73 (2H, d, $J_{H6\text{-}H5}$ 7.9 Hz, H_6), 2.34-2.12 (2H, m, H_3), 2.34-2.12 (4H, m, $H_{2/2'}$), 1.66-1.54 (2H, m, $H_{3'}$), 1.45-1.29 (1H, m, H_5), 0.98-0.89 (2H, m, H_4). ^{13}C NMR (125 MHz,$CDCl_3$): δ = 98.9 (C_1), 60.0 (C_6), 29.1 (C_3), 21.5 (C_2), 21.5 ($C_{4/5}$), 20.0 ($C_{4/5}$). **IR** (ν_{max}/ cm $^{-1}$) 3320 (OH). 2988 (CH); **HRMS:** Found $[M+H]^+$ 150.1051, $C_{10}H_{14}O$ requires 150.1045. The compound spectra were in good agreement with the previously reported literature values.[164]

(1R,8S,9s)-Bicyclo[6.1.0]non-4-yn-9-ylmethyl (4-nitrophenyl) carbonate and (1R,8S,9r)-Bicyclo[6.1.0]non-4-yn-9-ylmethyl (4-nitrophenyl) carbonate 4.26



4-Nitrophenyl chloroformate (202 mg, 1 mmol) and pyridine (109 μl, 1.35 mmol) were added to a stirred solution of **4.24** (100 mg, 0.67 mmol) in DCM (10 mL). The solution was stirred at r.t. for 20 minutes before saturated ammonium chloride solution (10 mL) was added and the crude product extracted with DCM (2 × 10mL). The organic layers were combined and concentrated to leave yellow oil that was purified using flash chromatography (silica; 6:1 (v/v) hexane-EtOAc) to afford **4.26** as an amorphous solid (160 mg, 76%, 2:1 exo: endo ratio). **Exo**: R_F 0.37 (6:1 (v/v) hexane-EtOAc). ^{1}H NMR (500 MHz,$CDCl_3$) δ = 8.25 (2H, d, $J_{H10\text{-}9}$ 9.1 Hz, H_{10}), 7.37 (2H, d, $J_{H9\text{-}H10}$ 9.1 Hz, H_9), 4.22 (2H, app d, $J_{H6\text{-}H5}$ 6.7 Hz, H_6) 2.47-2.39 (2H, m H_3), 2.36-2.10 (4H, m, $H_{2/2'}$), 1.45-1.34 (2H, m, $H_{3'}$), 0.88-0.76 (2H, m, H_4), 0.88-0.76 (1H, m, H_5). ^{13}C NMR (125 MHz, $CDCl_3$): δ = X (C_7), X (C_{11}), X (C_9), 125.4 (C_{10}), 124.8 (C_9), 98.7 (C_1), 74.0 (C_6), 33.2 (C_3), 23.3 ($C_{4/5}$), 23.0 ($C_{4/5}$), 21.3 (C_2). **Endo**: R_F 0.37 (6:1 (v/v) hexane-EtOAc). ^{1}H NMR (500 MHz,$CDCl_3$) δ = 8.26 (2H, d, $J_{H10\text{-}9}$ 9.4 Hz, H_{10}), 7.37 (2H, d, $J_{H9\text{-}H10}$ 9.4 Hz, H_9), 4.38 (2H, app d, $J_{H6\text{-}H5}$ 8.2 Hz, H_6) 2.36-2.10 (2H, m H_3), 2.36-2.10 (4H, m, $H_{2/2'}$), 1.66-1.53 (2H, m, $H_{3'}$), 1.55-1.44 (1H, m, H_5), 1.09-0.99 (2H, m, H_4). ^{13}C NMR (125 MHz,$CDCl_3$): δ = X (C_7), X (C_{11}), X (C_9), 125.4 (C_{10}), 124.8 (C_9), 98.8 (C_1), 68.1 (C_6), 29.1 (C_3), 21.3 (C_2), 20.6 ($C_{4/5}$), 17.3 ($C_{4/5}$). **IR** (v_{max}/ cm $^{-1}$) 3118, 2974 (CH), 1750 (CO); **HRMS:** Found $[M+Na]^+$ 338.0985, $C_{17}H_{17}NO_5Na$ requires 338.0999. The compound spectra were in good agreement with the previously reported literature values.[164]

6.1.3 Solid phase peptide synthesis

6.1.3.1 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 as received. 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 which allowed multiple fritted polypropylene tubes to be spun simultaneously. H-Gly-2-ClTrt (Sigma, loading: 1.1 mmol/g), H-Thr-2-ClTrt (Sigma, loading: 1.1 mmol/g) and

Rink Amide NovaGel (Novabiochem, loading: 0.66 mmol/g) were used for solid phase peptide synthesis.

Peptides having a C-terminal glycine residue were synthesised on H-Gly-2-ClTrt resin (Sigma). Peptides having a C-terminal amide group were synthesised on Rink Amide NovaGel resin (Novabiochem). Depsipeptide building block 12 was incorporated into the peptides using the standard coupling and deprotection methods described below.

Peptides were analysed using an Agilent 1290 affinity LC system equipped with an Ascentis Express 10 cm × 2.1 mm, 2.7 μm ES-C18 peptide column (0.5 mL min−1) and ultraviolet (UV) detection at 220 nm - 280 nm. Gradient from 0.1% TFA/5% MeCN (*vol*/*vol*) in H_2O to 0.1% TFA/95% MeCN in H_2O over 5 min 40 s.

6.1.3.2 General procedure for solid phase peptide synthesis

Method One: Coupling amino acids onto solid phase resin.

The resin was swollen in DMF for 30 minutes. HCTU (4.9 equiv) and an Fmoc-amino acid (5 equiv) were dissolved in DMF (6 mL), followed by the addition of DIPEA (10 equiv). The solution was transferred to the swelled resin and the mixture was left to spin for 1 h at room temperature. The resin was isolated by filtration and washed with DMF (3 × 6 mL × 2 min spins).

Method Two: Coupling reaction analysis

A small sample of beads from the isolated Fmoc-amino acid-resin were exposed to TFA (100 μl) for 2 minutes before the solution was quenched with methanol (1 mL). The solution was filtered and analysed by LCMS. If any starting material was detected method one was repeated.

Method Three: Fmoc deprotection and peptide elongation

The Fmoc-amino acid-resin was treated with piperidine (20% in DMF) (10 mL) and left to spin for 2 min at rt. This process was repeated five times before the amino acid resin was washed with DMF (4 × 10 mL × 2 min spins). The peptide sequence was elongated by repeating methods 1-3 (excluding the resin swelling step) until the correct peptide was synthesised.

Method Four: Fluorescein labelling (optional)

After elongation of the peptide to the desired length, a final coupling step using a Boc-Lys-OH or GABA residue was performed. After this coupling step, the resin was treated with a solution of f (FITC, 6 equiv) and DIPEA (14 equiv) dissolved in a minimal amount of DMF. The

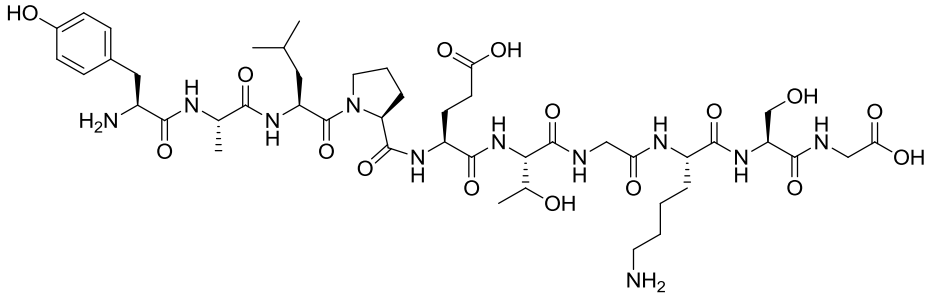
reaction was spun at r.t., overnight in the dark. The resin was drained and washed with DMF (4 × 10 mL × 2 min spins), before method five was initiated.

Method Five: Cleavage and isolation

The amino acid-resin was washed with DMF (3× 6 mL × 2 min spins), DCM (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 h. A cleavage cocktail consisting of H_2O: TIS: TFA (2.5: 2.5: 95) (3 mL) was added to the resin and the mixture was left to spin for 2 h at rt. The resin was filtered and washed with TFA (2 × 3 mL). The washings were combined and added to cold diethyl ether (40 mL) to precipitate the peptide. The precipitate was pelleted by centrifugation (6,000 × g, 10 min) and the cold diethyl ether discarded. The peptide pellet was re-suspended in cold diethyl ether (40 mL) and the process repeated three times. The peptide was analysed by LCMS before being dissolved in the minimal amount of water and freeze dried.

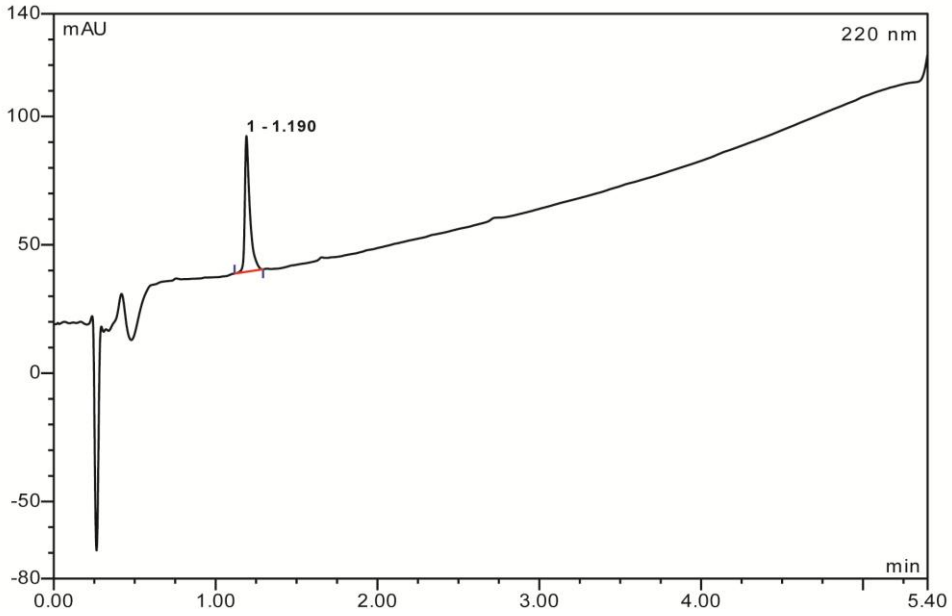
6.1.4 Spectral data for peptides

H2N-YALPETGKSG-COOH 2.4

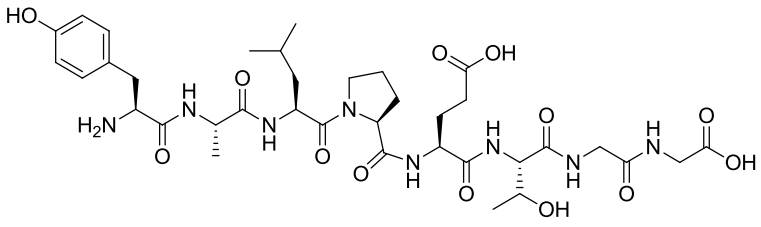


Yield: 120 mg, >99% ^{1}H NMR (500 MHz, CD_3OD): δ = **Y1:** 4.06 (1H, dd, $J_{\beta3\text{-}\alpha}$ 8.4 Hz, $J_{\beta2\text{-}\alpha2}$ 5.5 Hz, H_α), 3.18 (1H, dd, $J_{\beta2\text{-}\alpha2}$ 5.5 Hz, $J_{\beta2\text{-}\beta3}$ 14.4 Hz, $H_{\beta2}$), 2.94 (1H, dd, $J_{\beta3\text{-}\alpha2}$ 8.4 Hz, $J_{\beta3\text{-}\beta2}$ 14.4 Hz, $H_{\beta3}$), 7.11 (2H, d, $J_{\delta1/2\text{-}\varepsilon1/2}$ 8.5 Hz, $H_{\delta1/2}$), 6.77 (2H, d, $J_{\varepsilon1/2\text{-}\delta1/2}$ 8.5 Hz, $H_{\varepsilon1/2}$). **A2:** 4.45 (1H, q, $J_{\alpha\text{-}\beta2}$ 7.1 Hz, H_α), 1.39 (3H, d, $J_{\beta2\text{-}\alpha}$ 7.1 Hz, $H_{\beta2}$). **L3:** 4.69 (1H, dd, $J_{\alpha\text{-}\beta2}$ 8.7 Hz, $J_{\alpha\text{-}\beta3}$ 5.3 Hz, H_α), 1.72-1.59 (2H, m, $H_{\beta2/\beta3}$), 1.82-1.72 (1H, m, H_γ), 0.99 (6H, app d, $J_{\delta1/\delta2\text{-}\gamma}$ 6.5 Hz, $H_{\delta1/\delta2}$). **P4:** 4.42 (1H, dd, $J_{\alpha\text{-}\beta2}$ 8.1 Hz, $J_{\alpha\text{-}\beta3}$ 5.5 Hz, H_α), 2.30-2.20 (1H, m, $H_{\beta2}$) 2.06-1.94 (1H, m, $H_{\beta3}$), 2.15-2.06 (1H, m, $H_{\gamma2}$), 2.06-1.95 (1H, m, $H_{\gamma3}$), 3.94-3.75 (1H, m, $H_{\delta2}$), 3.75-3.67 (1H, m, $H_{\delta3}$). **E5**: 4.35 (1H, dd, $J_{\beta3\text{-}\alpha}$ 5.0 Hz, $J_{\beta2\text{-}\alpha}$ 9.3 Hz, H_α), 2.20-2.12 (1H, m, $H_{\beta2}$), 2.06-1.95 (1H, m, $H_{\beta3}$), 2.52-2.41 (2H, m, $H_{\gamma2/3}$). **T6:** 4.26 (1H, d, $J_{\alpha\text{-}\beta2}$ 4.2 Hz, H_α), 4.24 (1H, qd, $J_{\beta\text{-}\alpha}$ 4.2 Hz, $J_{\beta\text{-}\gamma2}$ 6.3 Hz, H_β), 1.22 (3H, d, $J_{\gamma2\text{-}\beta}$ 6.2 Hz, $H_{\gamma2}$). **G7:** 4.02-3.75 (2H, m, $H_{\alpha2/\alpha3}$). **K8:** 4.49 (1H, dd, $J_{\alpha\text{-}\beta2}$ 4.1 Hz, $J_{\alpha\text{-}\beta3}$ 5.6 Hz, H^α), 2.30-2.20 (1H, m, $H_{\beta2}$) 1.96-1.85 (1H, m, $H^{\beta3}$), 1.53-1.40 (2H, m, $H_{\gamma2/\gamma3}$), 1.72-1.59 (2H, m, $H_{\delta2/\delta3}$), 1.97-1.90 (2H, m, $H_{\delta2/\delta3}$). **S9:** 4.48-4.45 (1H, m, H_α), 3.90-3.77

(2H, 1H, m, $H_{\beta2/\beta3}$), **G10:** 4.02-3.75 (2H, m, $H_{\alpha2/\alpha3}$).^{13}C NMR (75 MHz, CD_3OD): δ = **Y1:** 169.7 (CO), 55.0 (C_α), 37.8 (C_β), 126.0 (C_γ), 116.9 (C_δ), 131.7 (C_ε), 158.2 (C_ζ). **A2:** 174.0 (CO), 50.3 (C_α), 18.3 (C_β). **L3:** 173.4 (CO), 51.4 (C_α), 40.9 (C_β), 25.9 (C_γ), 23.7 ($C_{\delta1}$), 22.0 ($C_{\delta2}$). **P4:** 175.0 (CO), 62.2 (C_α), 26.1 (C_β), 30.4 (C_γ), 49.7 (C_δ). **E5:** 174.7 (CO), 55.3 (C_α), 27.6 (C_β), 31.9 (C_γ), 177.6 (CO_δ). **T6**: 174.4 (CO), 60.1 (C_α), 68.4 (C_β), 20.2 (C_γ). **G7:** 173.6 (CO), 43.5 (C_α). **K8:** 173.7 (CO), 54.4 (C_α), 32.4 (C_β), 23.5 (C_γ), 27.7 (C_δ), 40.6 (C_ε). 172.2 **S9:** (CO), 56.9 (C_α), 63.0 (C_β). **G10:** 171.9 (CO), 44.0 (C_α). **Analytical HPLC** (below). **HRMS** of peak eluting at 1.19 min: Found [M + H^+] 1022.5158, $C_{45}H_{72}N_{11}O_{16}$ requires 1022.5153.

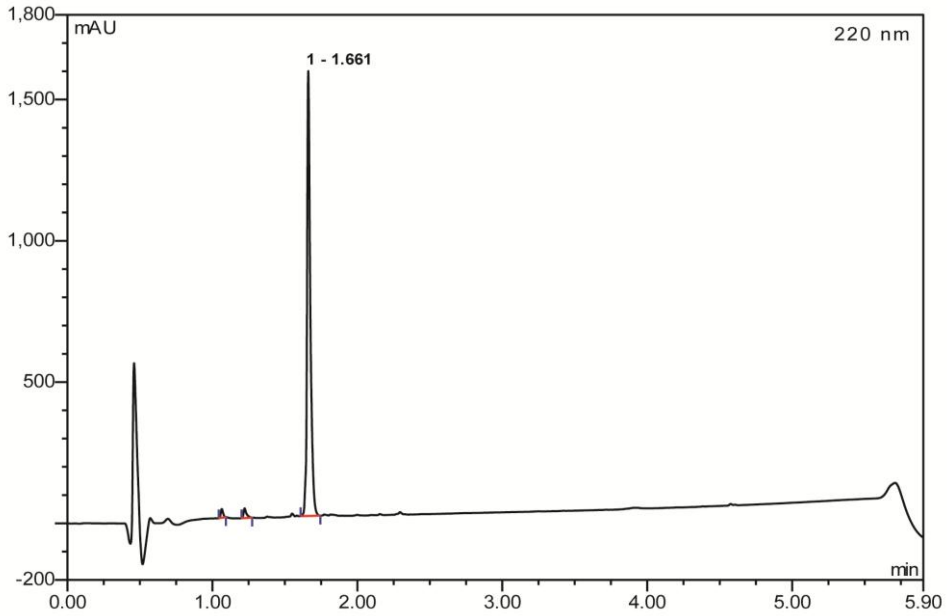


H_2N-Y-A-L-P-E-T-G-G-COOH 2.5

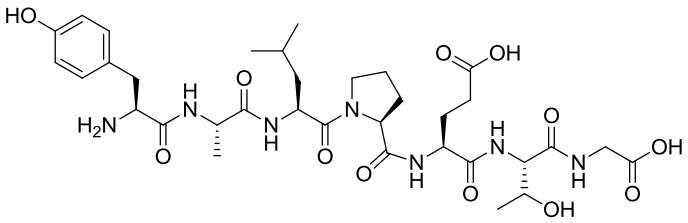


Yield: 140 mg, 98%^{1}H NMR (500 MHz, CD_3OD): δ = **Y1:** 4.07 (1H, dd, $J_{\alpha\text{-}\beta3}$ 8.6 Hz, $J_{\alpha\text{-}\beta2}$ 5.3 Hz, H_α), 3.21 (1H, dd, $J_{\beta2\text{-}\alpha}$ 5.3 Hz, $J_{\beta2\text{-}\beta3}$ 14.4 Hz, $H_{\beta2}$), 2.94 (1H, dd, $J_{\beta3\text{-}\alpha}$ 8.6 Hz, $J_{\beta3\text{-}\beta2}$ 14.4 Hz, $H_{\beta3}$), 7.14 (2H, d, $J_{\delta1/2\text{-}\varepsilon1/2}$ 8.6 Hz, $H_{\delta1/2}$), 6.80 (2H, d, $J_{\varepsilon1/2\text{-}\delta1/2}$ 8.6 Hz, $H_{\varepsilon1/2}$). **A2:** 4.45 (1H, q, $J_{\alpha\text{-}\beta2}$ 7.1 Hz, H_α), 1.39 (3H, d, $J_{\beta2\text{-}\alpha}$ 7.1 Hz, $H_{\beta2}$). **L3:** 4.67 (1H, dd, $J_{\alpha\text{-}\beta2}$ 8.4 Hz, $J_{\alpha\text{-}\beta3}$ 6.0 Hz, H_α), 1.67-1.59 (2H, m, $H_{\beta2/\beta3}$), 1.82-1.73 (1H, m, H_γ), 0.99 (3H, d, $J_{\delta1\text{-}\gamma}$ 6.6 Hz, $H_{\delta1}$), 0.99 (3H, d, $J_{\delta2\text{-}\gamma}$ 6.6 Hz, $H_{\delta2}$). **P4:** 4.45-4.41 (1H, m, H_α), 2.30-2.09 (2H, m, $H_{\beta2}$) 2.06-1.94 (1H, m, $H_{\beta3}$), 2.30-2.09 (1H, m, $H_{\gamma2}$), 2.06-1.94 (1H, m, $H_{\gamma3}$), 4.00-3.84 (1H, m, $H_{\delta2}$), 3.95-3.76 (1H, m, $H_{\delta3}$). **E5**: 4.36 (1H, dd, $J_{\alpha\text{-}\beta3}$ 5.0 Hz, $J_{\alpha\text{-}\beta2}$ 9.2 Hz, H_α), 2.30-2.09 (1H, m, $H_{\beta2}$), 2.06-1.94 (1H, m, $H_{\beta3}$),

2.57-2.44 (2H, m, $H^{\gamma 2/3}$). **T6:** 4.31 (1H, d, $J_{\alpha\text{-}\beta}$ 3.7 Hz, H_α), 4.25 (1H, qd, $J_{\beta\text{-}\alpha}$ 3.7 Hz, $J_{\beta\text{-}\gamma 2}$ 6.2 Hz, H_β), 1.18 (3H, d, $J_{\gamma 2\text{-}\beta}$ 6.2 Hz, $H_{\gamma 2}$). **G7:** 4.91 (1H, d, $J_{\alpha 2\text{-}\alpha 3}$ 17.8 Hz, $H_{\alpha 2}$), 3.94 (1H, d, $J_{\alpha 3\text{-}\alpha 2}$ 17.8 Hz, $H_{\alpha 3}$). **G8:** 4.13 (1H, d, $J_{\alpha 2\text{-}\alpha 3}$ 16.6 Hz, $H_{\alpha 2}$), 3.93 (1H, d, $J_{\alpha 3\text{-}\alpha 2}$ 16.6 Hz, $H_{\alpha 3}$).^{13}C NMR (125 MHz, CD_3OD): δ = **Y1:** 55.7 (C_α), 37.8 (C_β), 125.9 (C_γ), 131.6 (C_δ), 116.9 (C_ε), 158.3 (C_ζ). **A2:** 50.2 (C_α), 18.2 (C_β). **L3:** 51.2 (C_α), 41.1 (C_β), 25.8 (C_γ), 23.7 ($C_{\delta 1}$), 21.9 ($C_{\delta 2}$). **P4:** 61.9 (C_α), 26.1 (C_β), 30.4 (C_γ), 48.6 (C_δ). **E5:** 55.0 (C_α), 27.7 (C_β), 31.3 (C_γ). **T6**: 60.5 (C_α), 68.2 (C_β), 20.1 (C^γ). **G7:** 41.8 (C_α). **G8:** 43.5 (C_α). **CO (× 9):** 176.9, 174.9, 174.3, 174.1, 173.6, 173.0, 172.9, 172.0, 169.5. **Analytical HPLC** (below). **HRMS** of peak eluting at 1.66 min: Found [M + H^+] 807.3901, $C_{36}H_{55}N_8O_{13}$ requires 807.3883.

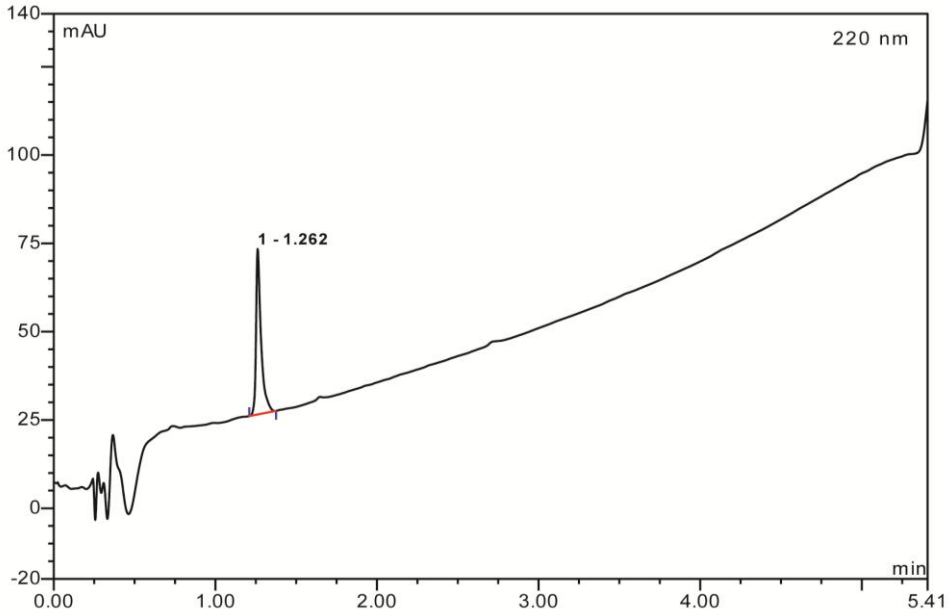


H_2N-Y-A-L-P-E-T-G-COOH 2.6

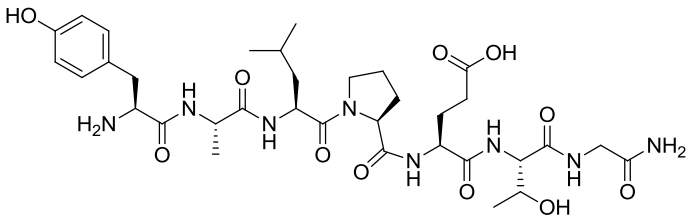


Yield: 80 mg, 98% ^{1}H NMR (500 MHz, CD_3OD): δ = **Y1:** 4.04 (1H, dd, $J_{\beta 3\text{-}\alpha}$ 8.5 Hz, $J_{\beta 2\text{-}\alpha 2}$ 5.3 Hz, H_α), 3.18 (1H, dd, $J_{\beta 2\text{-}\alpha 2}$ 5.3 Hz, $J_{\beta 2\text{-}\beta 3}$ 14.6 Hz, $H_{\beta 2}$), 2.92 (1H, dd, $J_{\beta 3\text{-}\alpha 2}$ 8.5 Hz, $J_{\beta 3\text{-}\beta 2}$ 14.6 Hz, $H_{\beta 3}$), 7.11 (2H, d, $J_{\delta 1/2\text{-}\varepsilon 1/2}$ 8.5 Hz, $H_{\delta 1/2}$), 6.77 (2H, d, $J_{\varepsilon 1/2\text{-}\delta 1/2}$ 8.6 Hz, $H_{\varepsilon 1/2}$). **A2:** 4.45 (1H, q, $J_{\alpha\text{-}\beta 2}$ 7.1 Hz, H_α), 1.37 (3H, d, $J_{\beta 2\text{-}\alpha}$ 7.1 Hz, $H_{\beta 2}$). **L3:** 4.67 (1H, dd, $J_{\alpha\text{-}\beta 2}$ 8.9 Hz, $J_{\alpha\text{-}\beta 3}$ 5.6 Hz, H_α), 1.66-1.60 (2H, m, $H_{\beta 2/\beta 3}$), 1.82-1.73 (1H, m, H_γ), 1.00 (3H, d, $J_{\delta 1\text{-}\gamma}$ 6.6 Hz, $H_{\delta 1}$), 0.99 (3H, d, $J_{\delta 1\text{-}\gamma}$ 6.6 Hz, $H_{\delta 2}$). **P4:** 4.46-4.42 (1H, m, H_α), 2.30-2.18 (1H, m, $H_{\beta 2}$) 2.06-1.95 (1H, m, $H_{\beta 3}$),

2.15-2.06 (1H, m, $H_{\gamma 2}$), 2.06-1.95 (1H, m, $H_{\gamma 3}$), 3.91-3.77 (1H, m, $H_{\delta 2}$), 3.72-3.59 (1H, m, $H_{\delta 3}$). **E5**: 4.37 (1H, dd, $J_{\beta 3\text{-}\alpha}$ 5.1 Hz, $J_{\beta 2\text{-}\alpha}$ 9.2 Hz, H_α), 2.20-2.15 (1H, m, $H_{\beta 2}$), 2.06-1.95 (1H, m, $H_{\beta 3}$), 2.53-2.41 (2H, m, $H_{\gamma 2/3}$). **T6:** 4.35 (1H, d, $J_{\alpha\text{-}\beta 2}$ 3.8 Hz, H_α), 4.22 (1H, qd, $J_{\beta\text{-}\alpha}$ 3.8 Hz, $J_{\beta\text{-}\gamma 2}$ 6.4 Hz, H_β), 1.20 (3H, d, $J_{\gamma 2\text{-}\beta}$ 6.4 Hz, $H_{\gamma 2}$). **G7:** 4.00 (1H, d, $J_{\alpha 2\text{-}\alpha 3}$ 17.8 Hz, $H_{\alpha 2}$), 3.94 (1H, d, $J_{\alpha 3\text{-}\alpha 2}$ 17.8 Hz, $H_{\alpha 3}$). ^{13}C NMR (125 MHz, CD_3OD): δ = **Y1:** 169.4 (CO), 54.7 (C_α), 37.8 (C_β), 125.9 (C_γ), 116.8 (C_δ), 131.6 (C_ε), 158.3 (C_ζ). **A2:** 174.3 (CO), 50.2 (C_α), 18.2 (C_β). **L3:** 173.4 (CO), 51.1 (C_α), 41.1 (C_β), 25.8 (C_γ), 23.7 ($C_{\delta 1}$), 21.9 ($C_{\delta 2}$). **P4:** 174.8 (CO), 61.7 (C_α), 26.1 (C_β), 30.4 (C_γ), 48.6 (C_δ), **E5:** 173.8 (CO), 55.7 (C_α), 27.8 (C_β), 31.3 (C_γ), 176.8 (CO_δ). **T6**: 172.8 (CO), 60.1 (C_α), 68.3 (C_β), 20.0 (C_γ). **G7:** 172.9 (CO), 42.0 (C_α). **Analytical HPLC** (below). **HRMS** of peak eluting at 1.26 min: Found [M + H^+] 750.3659, $C_{34}H_{52}N_7O_{12}$ requires 750.365.

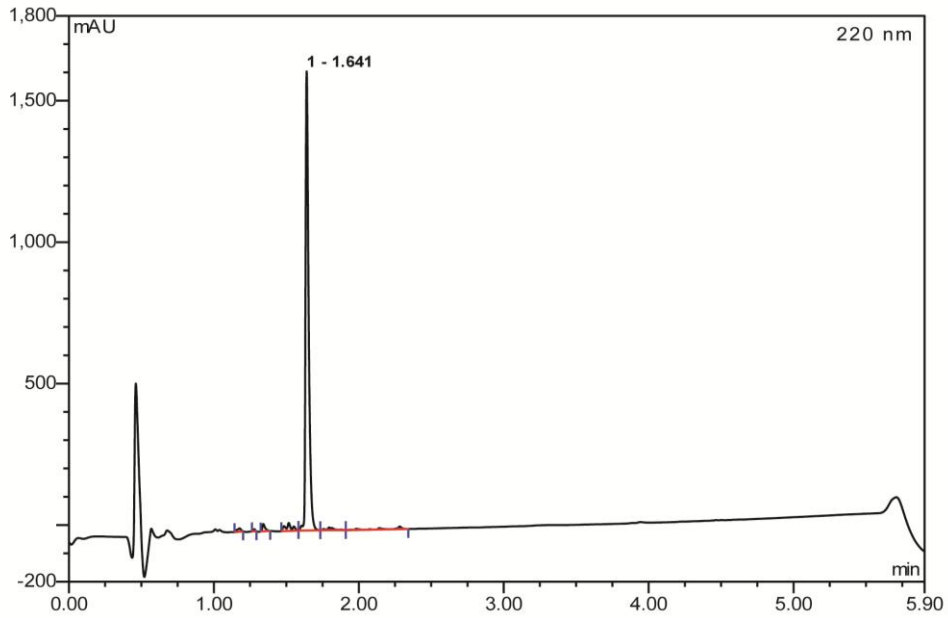


H_2N-Y-A-L-P-E-T-G-$CONH_2$ 2.7

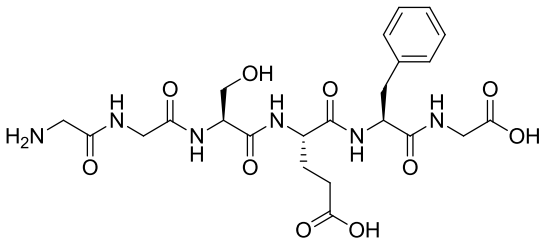


Yield: 94 mg, 98% ^{1}H NMR (500 MHz, CD_3OD): δ = **Y1:** 4.04 (1H, dd, $J_{\alpha\text{-}\beta 3}$ 8.4 Hz, $J_{\alpha\text{-}\beta 2}$ 5.4 Hz, H_α), 3.18 (1H, dd, $J_{\beta 2\text{-}\alpha}$ 5.4 Hz, $J_{\beta 2\text{-}\beta 3}$ 14.5 Hz, $H_{\beta 2}$), 2.91 (1H, dd, $J_{\beta 3\text{-}\alpha}$ 8.4 Hz, $J_{\beta 3\text{-}\beta 2}$ 14.5 Hz, $H_{\beta 3}$), 7.11 (2H, d, $J_{\delta 1/2\text{-}\varepsilon 1/2}$ 8.5 Hz, $H_{\delta 1/2}$), 6.77 (2H, d, $J_{\varepsilon 1/2\text{-}\delta 1/2}$ 8.5 Hz, $H_{\varepsilon 1/2}$). **A2:** 4.44 (1H, q, $J_{\alpha\text{-}\beta 2}$ 7.1 Hz, H_α), 1.36 (3H, d, $J_{\beta 2\text{-}\alpha}$ 7.1 Hz, $H_{\beta 2}$). **L3:** 4.67 (1H, dd, $J_{\alpha\text{-}\beta 2}$ 8.1 Hz, $J_{\alpha\text{-}\beta 3}$ 6.3 Hz, H_α), 1.69-1.58 (2H, m, $H_{\beta 2/3}$), 1.84-1.71 (1H, m, H_γ), 0.99 (6H, app d, $J_{\delta 1\text{-}\gamma}$ 6.4 Hz, $H_{\delta 1}$), 0.99

(6H, app d, $J_{\delta 2\text{-}\gamma}$ 6.4 Hz, $H_{\delta 2}$). **P4:** 4.42 (1H, dd, $J_{\alpha\text{-}\beta 2}$ 8.3 Hz, $J_{\alpha\text{-}\beta 3}$ 5.3 Hz H_α), 2.26-2.06 (1H, m, $H_{\beta 2}$) 2.07-1.95 (1H, m, $H_{\beta 3}$), 2.26-2.06 (1H, m, $H_{\gamma 2}$), 2.07-1.95 (1H, m, $H_{\gamma 3}$), 3.92-3.77 (1H, m, $H_{\delta 2}$), 3.72-3.60 (1H, m, $H_{\delta 3}$). **E5**: 4.37 (1H, dd, $J_{\alpha\text{-}\beta 3}$ 5.1 Hz, $J_{\alpha\text{-}\beta 2}$ 9.2 Hz, H_α), 2.26-2.06 (1H, m, $H_{\beta 2}$), 2.07-1.95 (1H, m, $H_{\beta 3}$), 2.56-2.38 (2H, m, $H_{\gamma 2/3}$). **T6:** 4.28 (1H, d, $J_{\alpha\text{-}\beta 2}$ 3.7 Hz, H_α), 4.22 (1H, qd, $J_{\beta\text{-}\alpha}$ 3.7 Hz, $J_{\beta\text{-}\gamma 2}$ 6.4 Hz, H_β), 1.20 (3H, d, $J_{\gamma 2\text{-}\beta}$ 6.4 Hz, $H_{\gamma 2}$). **G7:** 3.94 (1H, d, $J_{\alpha 2\text{-}\alpha 3}$ 17.1 Hz, $H_{\alpha 2}$), 3.84 (1H, d, $J_{\alpha 3\text{-}\alpha 2}$ 17.1 Hz, $H_{\alpha 3}$).^{13}C NMR (125 MHz, CD_3OD): δ = **Y1:** 55.8 (C_α), 37.8 (C_β), 125.9 (C_γ), 131.6 (C_δ), 116.9 (C_ε), 158.3 (C_ζ). **A2:** 50.2 (C_α), 18.2 (C_β). **L3:** 51.2 (C_α), 41.0 (C_β), 25.8 (C_γ), 23.7 ($C_{\delta 1}$), 21.9 ($C_{\delta 2}$). **P4:** 61.9 (C_α), 26.1 (C_β), 30.4 (C_γ), 48.7 (C_δ), **E5:** 55.0 (C_α), 27.6 (C_β), 31.3 (C_γ). **T6**: 60.6 (C_α), 68.3 (C_β), 20.1 (C_γ). **G7:** 43.3 (C_α). **CO (× 8):** 176.8, 174.9, 174.4, 174.3, 174.1, 173.6, 173.0, 169.5. **Analytical HPLC** (below). **HRMS** of peak eluting at 1.64 min: Found [M + H^+] 749.3844, $C_{34}H_{53}N_8O_{11}$ requires 749.3828.

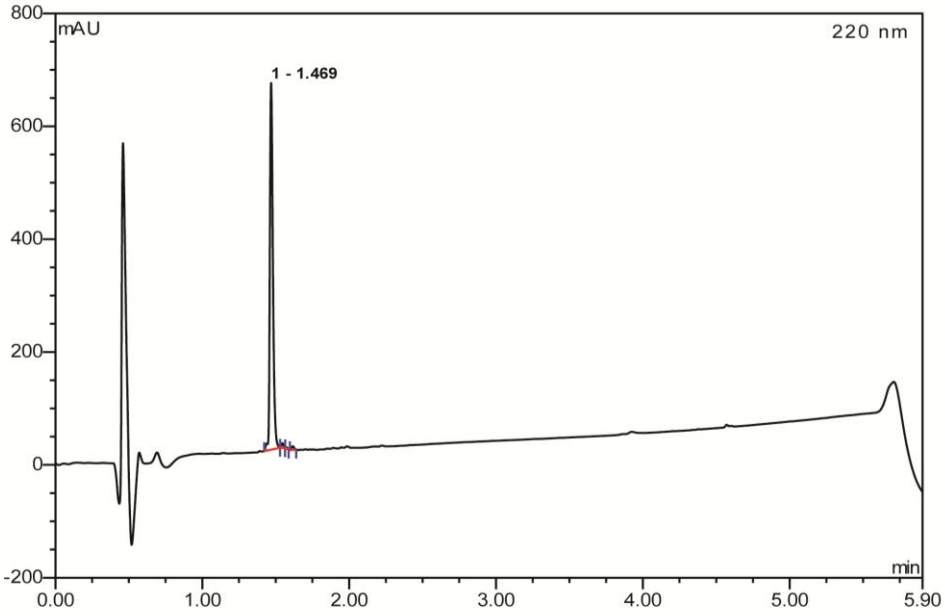


H_2N-G-G-S-E-F-G-COOH 2.8

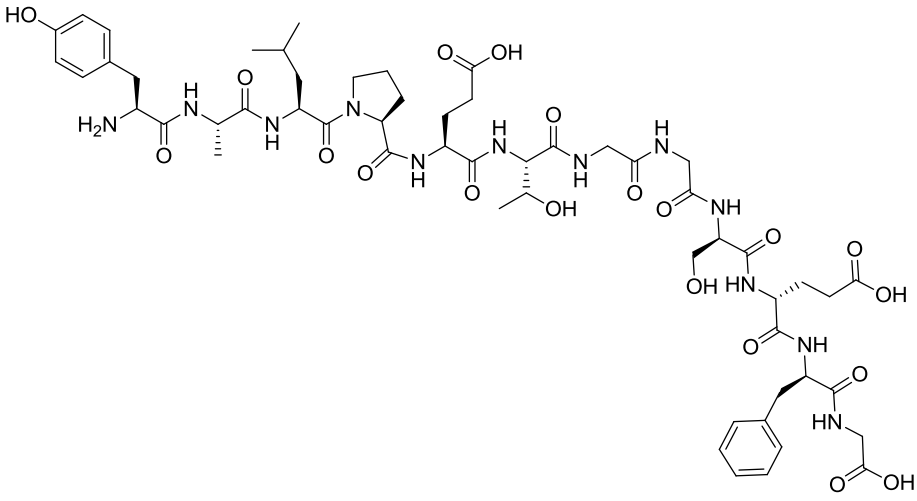


Yield: 120 mg, 97%. ^{1}H NMR (500 MHz, D_2O): δ = **G1:** 3.85 (1H, s, $H_{\alpha 2/3}$) **G2:** 4.01 (1H, d, $J_{\alpha 2\text{-}\alpha 3}$ 17.8 Hz, $H_{\alpha 2}$), 3.96 (1H, d, $J_{\alpha 3\text{-}\alpha 2}$ 17.8 Hz, $H_{\alpha 3}$) **S3:** 4.66 (1H, dd, $J_{\alpha\text{-}\beta 2}$ 5.9 Hz, $J_{\alpha\text{-}\beta 3}$ 9.4 Hz, H_α), 3.81 (1H, dd, $J_{\beta 2\text{-}\alpha}$ 6.0 Hz, $J_{\beta 2\text{-}\beta 3}$ 11.5 Hz, $H_{\beta 2}$), 3.75 (1H, dd, $J_{\beta 3\text{-}\alpha}$ 5.2 Hz, $J_{\beta 3\text{-}\beta 2}$ 11.5 Hz, $H_{\beta 3}$) **E4:** 4.24 (1H, dd, $J_{\alpha\text{-}\beta 3}$ 6.1 Hz, $J_{\alpha\text{-}\beta 2}$ 8.5 Hz, H_α), 1.78-1.94 (1H, m, $H_{\beta 2/3}$), 2.31-2.21 (1H, m,

$H_{\gamma 2}$), 2.19-2.10 (1H, m, $H_{\gamma 3}$) **F5**: 4.41 (1H, app q, $J_{\alpha\text{-}\beta 3/2}$ 5.7 Hz, H_α), 2.06 (1H, dd, $J_{\beta 2\text{-}\alpha}$ 6.0 Hz, $J_{\beta 2\text{-}\beta 3}$ 13.5 Hz, $H_{\beta 2}$), 1.79 (1H, dd, $J_{\beta 3\text{-}\alpha}$ 9.4 Hz, $J_{\beta 3\text{-}\beta 2}$ 13.5 Hz, $H_{\beta 3}$), 7.34-7.39 (2H, m, $H_{\delta 1/2}$), 7.28-7.21 (2H, m, $H_{\varepsilon 1/2}$), 7.28-7.21 (1H, m, $H_{\zeta 1}$). **G6:** 4.09 (1H, s, $H_{\alpha 2/3}$).^{13}C NMR (125 MHz, CD_3OD): δ = **G1:** 36.8 (C_α). **G2:** 41.2 (C_α). **S3:** 55.3 (C_α), 60.9 (C_β). **E4:** 53.3 (C_α), 25.8 (C_β), 29.6 (C_γ). **F5:** 54.5 (C_α), 36.8 (C_β), 136.3 (C_γ), 128.6 (C_δ), 129.0 (C_ε), 128.0 (C_ζ). **G6**: 42.2 (C_α). **CO (× 7):** 176.9, 173.1, 173.1, 172.7, 171.8, 171.3, 167.7. **Analytical HPLC** (below). **HRMS** of peak eluting at 1.47 min: Found [M - H^+] 551.2117, $C_{23}H_{31}N_6O_{10}$ requires 551.2107

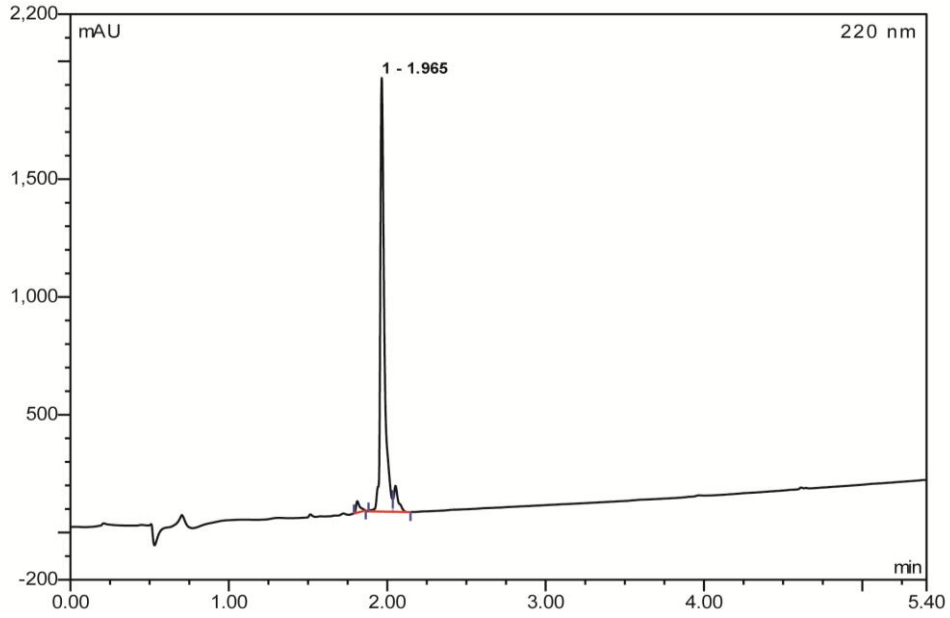


H_2N-Y-A-L-P-E-T-G-G-S-E-F-G-COOH 2.9

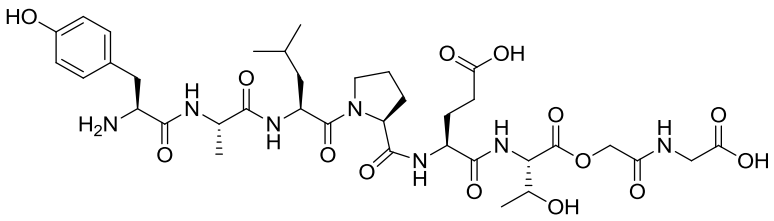


Purified by HPLC: Yield: 80 mg, 40% ^{1}H NMR (500 MHz, CD_3OD): δ = **Y1:** 4.05 (1H, dd, $J_{\alpha\text{-}\beta 3}$ 8.6 Hz, $J_{\alpha\text{-}\beta 2}$ 5.3 Hz, H_α), 3.18 (1H, dd, $J_{\beta 2\text{-}\alpha}$ 5.3 Hz, $J_{\beta 2\text{-}\beta 3}$ 14.2 Hz, $H_{\beta 2}$), 2.91 (1H, dd, $J_{\beta 3\text{-}\alpha}$ 8.6 Hz, $J_{\beta 3\text{-}\beta 2}$ 14.2 Hz, $H_{\beta 3}$), 7.11 (2H, d, $J_{\delta 1/2\text{-}\varepsilon 1/2}$ 8.6 Hz, $H_{\delta 1/2}$), 6.77 (2H, d, $J_{\varepsilon 1/2\text{-}\delta 1/2}$ 8.6 Hz,

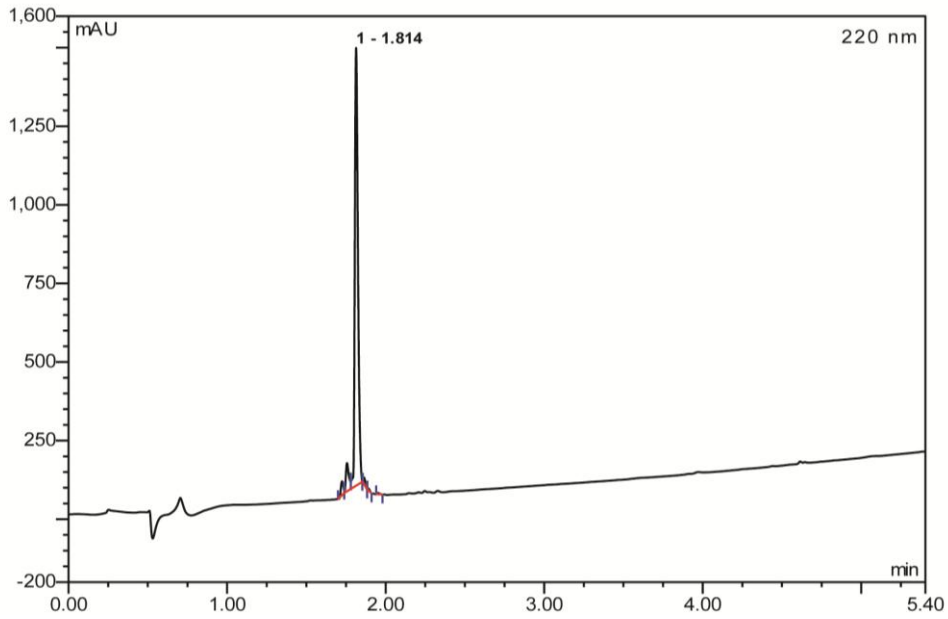
$H_{\varepsilon 1/2}$). **A2:** 4.48-4.44 (1H, m, H_{α}), 1.35 (3H, d, $J_{\beta 2\text{-}\alpha}$ 7.1 Hz, $H_{\beta 2}$). **L3:** 4.71-4.64 (1H, m, H_{α}), 1.67-1.61 (2H, m, $H_{\beta 2/3}$), 1.81-1.71 (1H, m, H_{γ}), 0.99 (3H, d, $J_{\delta 1\text{-}\gamma}$ 6.5 Hz, $H_{\delta 1}$), 0.98 (3H, d, $J_{\delta 2\text{-}\gamma}$ 6.6 Hz, $H_{\delta 2}$). **P4:** 4.42 (1H, dd, $J_{\alpha\text{-}\beta 2}$ 8.5 Hz, $J_{\alpha\text{-}\beta 3}$ 5.2 Hz H_{α}), 2.28-2.07 (1H, m, $H_{\beta 2}$) 2.05-1.95 (1H, m, $H_{\beta 3}$), 2.28-2.07 (1H, m, $H_{\gamma 2}$), 2.05-1.95 (1H, m, $H_{\gamma 3}$), 4.00-3.84 (1H, m, $H_{\delta 2}$), 3.72-3.65 (1H, m, $H_{\delta 3}$).**E5**: 4.37 (1H, dd, $J_{\alpha\text{-}\beta 2}$ 5.0 Hz, $J_{\alpha\text{-}\beta 3}$ 9.2 Hz, H_{α}), 2.28-2.07 (1H, m, $H_{\beta 2}$), 2.05-1.95 (1H, m, $H_{\beta 3}$), 2.47 (2H, app t, $J_{\gamma 2/3\text{-}\beta 2/3}$ 7.2 Hz $H_{\gamma 2/3}$). **T6:** 4.33 (1H, d, $J_{\alpha\text{-}\beta}$ 3.9 Hz, H_{α}), 4.29-4.24 (1H, m, H_{β}), 1.20 (3H, d, $J_{\gamma 2\text{-}\beta}$ 6.4 Hz, $H_{\gamma 2}$). **G7:** 4.00-3.84 (2H, m, $H_{\alpha 2/3}$). **G8:** 4.00-3.84 (2H, m, $H_{\alpha 2/3}$). **S9:** 4.48-4.44 (1H, m, H_{α}), 4.00-3.84 (2H, m, $H_{\beta 2}$), 3.75-3.65 (1H, m, $H_{\beta 3}$). **E10:** 4.29-4.24 (1H, m, H_{α}), 1.94-1.87 (1H, m, $H_{\beta 2}$), 1.81-1.79 (1H, m, $H_{\beta 3}$), 2.28-2.07 (1H, m, $H_{\gamma 2/3}$). **F11**: 4.71-4.64 (1H, m, H_{α}), 3.28 (1H, dd, $J_{\beta 2\text{-}\alpha}$ 4.9 Hz, $J_{\beta 2\text{-}\beta 3}$ 14.0 Hz, $H_{\beta 2}$), 2.88 (1H, dd, $J_{\beta 3\text{-}\alpha}$ 10.1 Hz, $J_{\beta 3\text{-}\beta 2}$ 14.0 Hz, $H_{\beta 3}$), 7.29-7.23 (2H, m, $H_{\delta 1/2}$), 7.29-7.23 (2H, m, $H_{\varepsilon 1/2}$), 7.21-7.16 (1H, m, $H_{\zeta 1}$). **G12:** 4.00-3.84 (2H, m, $H_{\alpha 2/3}$).^{13}C NMR (125 MHz, CD_3OD): δ = **Y1:** 55.8 (C_{α}), 38.5 (C_{β}), 125.9 (C_{γ}), 131.6 (C_{δ}), 116.9 (C_{ε}), 158.3 (C_{ζ}). **A2:** 50.2 (C_{α}), 18.3 (C_{β}). **L3:** 51.2 (C_{α}), 41.0 (C_{β}), 25.9 (C_{γ}), 23.7 ($C_{\delta 1}$), 21.9 ($C_{\delta 2}$). **P4:** 62.1 (C_{α}), 26.1 (C_{β}), 30.4 (C_{γ}), 49.7 (C_{δ}), **E5:** 55.2 (C_{α}), 27.6 (C_{β}), 31.3 (C_{γ}), 176.9 (CO_{δ}). **T6**: 60.6 (C_{α}), 68.3 (C_{β}), 20.1 (C_{γ}). **G7:** 43.8 (C_{α}), **G8:** 41.9 (C_{α}), **S9:** 56.0 (C_{α}), 62.9 (C_{β}). **E10:** 55.0 (C_{α}), 27.6 (C_{β}), 31.0 (C_{γ}). **F11:** 56.9 (C_{α}), 37.9 (C_{β}), 138.6 (C_{γ}), 130.3 (C_{δ}), 129.5 (C_{ε}), 128.0 (C_{ζ}). **G12**: 44.0 (C_{α}). **CO (× 14):** 177.0, 176.6, 174.9, 174.3, 174.2, 173.8, 173.7, 173.6, 173.5, 173.2, 172.7, 172.5, 171.9, 169.4. **Analytical HPLC** (below). **HRMS** of peak eluting at 1.96 min: Found [M - H^+] 1225.5391, $C_{55}H_{77}N_{12}O_{20}$ requires 1225.5383



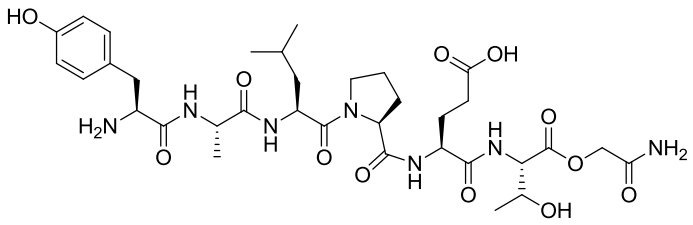
H_2N-Y-A-L-P-E-T-[G]-G-COOH 2.24



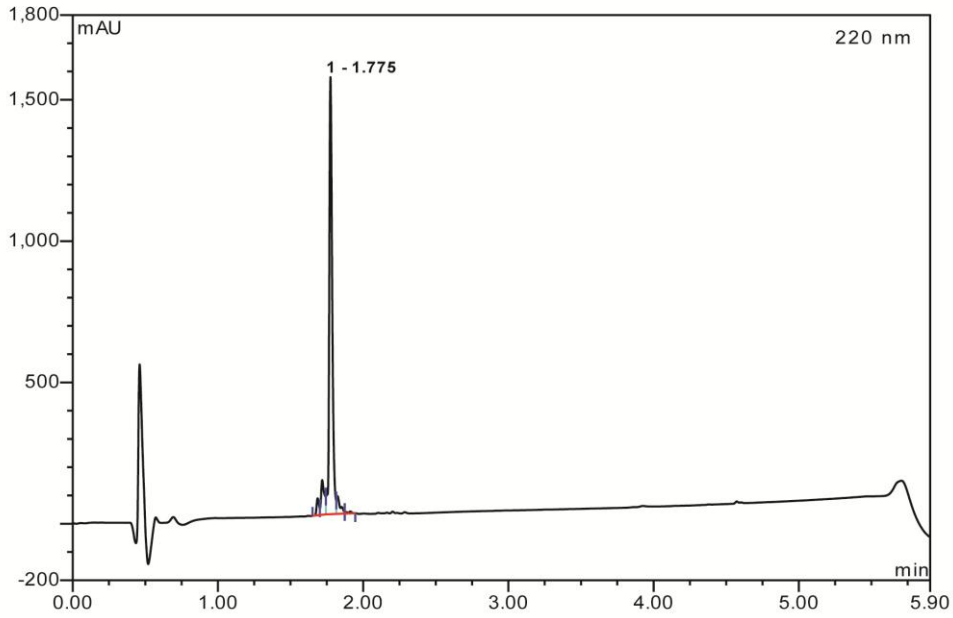
Purified by mass directed HPLC: Yield: 60 mg, 67% ^{1}H NMR (500 MHz, CD_3OD): δ = **Y1:** 4.00 (1H, dd, $J_{\alpha\text{-}\beta2}$ 5.5 Hz, $J_{\alpha\text{-}\beta3}$ 8.5 Hz, H_α), 3.17 (1H, dd, $J_{\beta2\text{-}\alpha}$ 5.5 Hz, $J_{\beta2\text{-}\beta3}$ 14.3 Hz, $H_{\beta2}$), 2.90 (1H, dd, $J_{\beta3\text{-}\alpha}$ 8.5 Hz, $J_{\beta3\text{-}\beta2}$ 14.3 Hz, $H_{\beta3}$), 7.10 (2H, d, $J_{\delta1/2\text{-}\varepsilon1/2}$ 8.4 Hz, $H_{\delta1/2}$), 6.76 (2H, d, $J_{\varepsilon1/2\text{-}\delta1/2}$ 8.4 Hz, $H_{\varepsilon1/2}$). **A2:** 4.47-4.35 (1H, m, H_α), 1.36 (3H, d, $J_{\beta2\text{-}\alpha}$ 7.1 Hz, $H_{\beta2}$). **L3:** 4.73-4.64 (1H, m, H_α), 1.67-1.57 (2H, m, $H_{\beta2/3}$), 1.82-1.70 (1H, m, H_γ), 0.99 (3H, d, $J_{\delta1\text{-}\gamma}$ 6.5 Hz, $H_{\delta1}$), 0.98 (3H, d, $J_{\delta2\text{-}\gamma}$ 6.5 Hz, $H_{\delta2}$). **P4:** 4.47-4.35 (1H, m, H_α), 2.29-2.05 (1H, m, $H_{\beta2}$) 2.05-1.95 (1H, m, $H_{\beta3}$), 2.29-2.05 (1H, m, $H_{\gamma2}$), 2.05-1.95 (1H, m, $H_{\gamma3}$), 3.89-3.79 (1H, m, $H_{\delta2}$), 3.72-3.61 (1H, m, $H_{\delta3}$). **E5**: 4.47-4.35 (1H, m, H_α), 2.29-2.05 (1H, m, $H_{\beta2}$), 2.05-1.95 (1H, m, $H_{\beta3}$), 2.48-2.43 (2H, m, $H_{\gamma2/3}$). **T6:** 4.60 (1H, d, $J_{\alpha\text{-}\beta}$ 3.2 Hz, H_α), 4.47-4.35 (1H, m, H_β), 1.20 (3H, d, $J_{\gamma2\text{-}\beta}$ 6.4 Hz, $H_{\gamma2}$). **G7:** 4.73-4.64 (2H, m, $H_{\alpha2/3}$), **G8:** 3.89-3.79 (2H, m, $H_{\alpha2/3}$).^{13}C NMR (125 MHz, CD_3OD): δ = **Y1:** 56.0 (C_α), 38.1 (C_β), 126.3 (C_γ), 131.6 (C_δ), 116.8 (C_ε), 158.3 (C_ζ). **A2:** 50.5 (C_α), 18.2 (C_β). **L3:** 51.2 (C_α), 41.1 (C_β), 25.9 (C_γ), 23.7 ($C_{\delta1}$), 21.9 ($C_{\delta2}$). **P4:** 62.0 (C_α), 26.1 (C_β), 30.4 (C_γ), 48.8 (C_δ), **E5:** 54.6 (C_α), 28.1 (C_β), 31.9 (C_γ). **T6**: 59.6 (C_α), 68.6 (C_β), 20.1 (C_γ). **G7:** 64.0 (C_α). **G8:** 43.5 (C_α). **CO (× 9):** 175.3, 174.3, 174.4, 174.2, 173.7, 173.5, 170.8, 170.1, 169.5. . **Analytical HPLC** (below). **HRMS** of peak eluting at 1.81 min: Found [M + H^+] 808.3698, $C_{36}H_{54}N_7O_{14}$ requires 808.3723



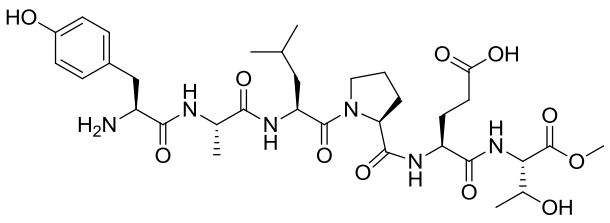
H_2N-Y-A-L-P-E-T-[G]-$CONH_2$ 2.25



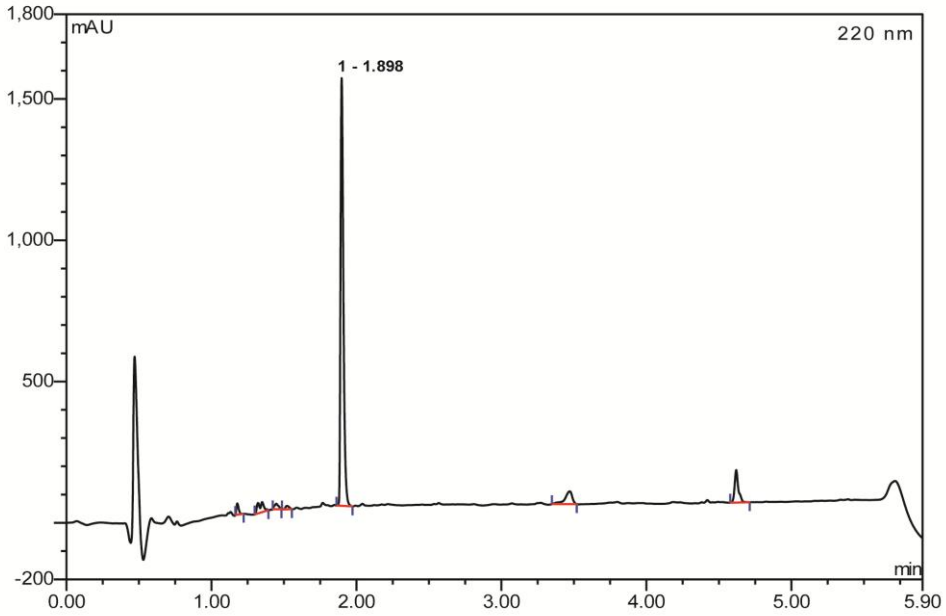
Yield: 90 mg, 87% ^{1}H NMR (500 MHz, CD_3OD): δ = **Y1:** 4.03 (1H, dd, $J_{\alpha\text{-}\beta3}$ 8.6 Hz, $J_{\alpha\text{-}\beta2}$ 5.3 Hz, H_α), 3.18 (1H, dd, $J_{\beta2\text{-}\alpha}$ 5.3 Hz, $J_{\beta2\text{-}\beta3}$ 14.6 Hz, $H_{\beta2}$), 2.91 (1H, dd, $J_{\beta3\text{-}\alpha}$ 8.6 Hz, $J_{\beta3\text{-}\beta2}$ 14.6 Hz, $H_{\beta3}$), 7.11 (2H, d, $J_{\delta1/2\text{-}\varepsilon1/2}$ 8.6 Hz, $H_{\delta1/2}$), 6.77 (2H, d, $J_{\varepsilon1/2\text{-}\delta1/2}$ 8.6 Hz, $H_{\varepsilon1/2}$). **A2:** 4.44-4.41 (1H, m, H_α), 1.37 (3H, d, $J_{\beta2\text{-}\alpha}$ 7.1 Hz, $H_{\beta2}$). **L3:** 4.68-4.63 (1H, m, H_α), 1.68-1.59 (2H, m, $H_{\beta2/3}$), 1.82-1.73 (1H, m, H_γ), 0.99 (3H, d, $J_{\delta1\text{-}\gamma}$ 6.6 Hz, $H_{\delta1}$), 0.99 (3H, d, $J_{\delta2\text{-}\gamma}$ 6.6 Hz, $H_{\delta2}$). **P4:** 4.44-4.41 (1H, m, H_α), 2.25-2.06 (1H, m, $H_{\beta2}$) 2.04-1.94 (1H, m, $H_{\beta3}$), 2.25-2.06 (1H, m, $H_{\gamma2}$), 2.04-1.94 (1H, m, $H_{\gamma3}$), 3.90-3.85 (1H, m, $H_{\delta2}$), 3.70-3.61 (1H, m, $H_{\delta3}$). **E5**: 4.44-4.41 (1H, m, H_α), 2.25-2.06 (1H, m, $H_{\beta2}$), 2.04-1.94 (1H, m, $H_{\beta3}$), 2.49-2.44 (2H, m, $H_{\gamma2/3}$). **T6:** 4.58 (1H, d, $J_{\alpha\text{-}\beta}$ 3.2 Hz, H_α), 4.40 (1H, qd, $J_{\beta\text{-}\alpha}$ 3.2 Hz, $J_{\beta\text{-}\gamma2}$ 6.6 Hz, H_β), 1.21 (3H, d, $J_{\gamma2\text{-}\beta}$ 6.6 Hz, $H_{\gamma2}$). **G7:** 4.68-4.65 (2H, m, $H_{\alpha2/3}$).^{13}C NMR (125 MHz, CD_3OD): δ = **Y1:** 55.7 (C_α), 37.8 (C_β), 125.9 (C_γ), 131.6 (C_δ), 116.9 (C_ε), 158.3 (C_ζ). **A2:** 50.2 (C_α), 18.1 (C_β). **L3:** 51.2 (C_α), 41.1 (C_β), 25.9 (C_γ), 23.7 ($C_{\delta1}$), 21.9 ($C_{\delta2}$). **P4:** 61.6 (C_α), 26.0 (C_β), 30.4 (C_γ), 48.6 (C_δ), **E5:** 54.3 (C_α), 28.1 (C_β), 31.2 (C_γ). **T6**: 59.7 (C_α), 68.6 (C_β), 19.9 (C_γ). **G7:** 63.6 (C_α). **CO (× 8):** 176.8, 174.3, 174.3, 174.2, 173.4, 172.5, 171.2, 169.5. **Analytical HPLC** (below). **HRMS** of peak eluting at 1.78 min: Found [M + H^+] 750.3667, $C_{34}H_{52}N_7O_{12}$ requires 750.3668:



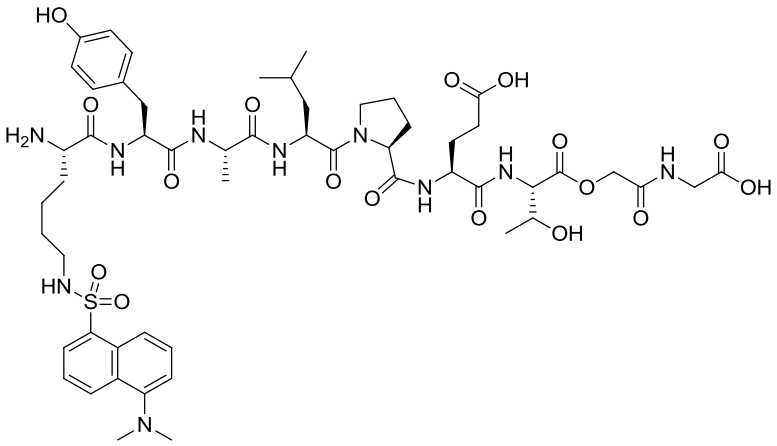
H_2N-Y-A-L-P-E-T-OMe 2.35



BocHN-Y(OtBu)ALPE(OtBu)T-OH was synthesised on H-Thr-2-chloro-trityl resin (Sigma) following standard methods 1-3 (above) and cleaved from the resin using 30% hexafluoroisopropanol in DCM (2 × 3 mL). Following precipitation in cold ether (50 mL) and lyophilisation, the partially protected peptide (61 mg, 0.06 mmol) was converted to the methyl ester using iodomethane (18 mg, 0.127 mmol) and cesium carbonate (19.5 mg, 0.06 mmol) in dry DMF (0.5 mL). The reaction mixture was stirred at r.t., overnight under N_2. The solvent was removed by evaporation to leave a yellow solid. The solid was dissolved in DCM (10 mL), washed with H_2O (2 × 25 mL) and brine (2 × 25 mL), before being concentred to leave a clear oil. The oil was treated with a solution of TFA:H_2O:TIS (95:2.5:2.5, 3 mL) for 1 h with stirring. The cleavage mixture was precipitated into cold ether (50 mL) and freeze dried to leave the titled white fluffy peptide (11.2 mg, 26%) ^{1}H NMR (500 MHz, CD_3OD): δ = **Y1:** 4.03 (1H, dd, $J_{\alpha\text{-}\beta3}$ 8.6 Hz, $J_{\alpha\text{-}\beta2}$ 5.3 Hz, H_α), 3.18 (1H, dd, $J_{\beta2\text{-}\alpha}$ 5.3 Hz, $J_{\beta2\text{-}\beta3}$ 14.6 Hz, $H_{\beta2}$), 2.91 (1H, dd, $J_{\beta3\text{-}\alpha}$ 8.6 Hz, $J_{\beta3\text{-}\beta2}$ 14.6 Hz, $H_{\beta3}$), 7.11 (2H, d, $J_{\delta1/2\text{-}\varepsilon1/2}$ 8.5 Hz, $H_{\delta1/2}$), 6.77 (2H, d, $J_{\varepsilon1/2\text{-}\delta1/2}$ 8.5 Hz, $H_{\varepsilon1/2}$). **A2:** 4.47-4.42 (1H, m, H_α), 1.37 (3H, d, $J_{\beta2\text{-}\alpha}$ 7.1 Hz, $H_{\beta2}$). **L3:** 4.66 (1H, dd, $J_{\beta3\text{-}\alpha}$ 8.4 Hz, $J_{\beta2\text{-}\alpha2}$ 6.0 Hz, H_α), 1.66-1.58 (2H, m, $H_{\beta2/3}$), 1.81-1.73 (1H, m, H_γ), 0.99 (3H, d, $J_{\delta1\text{-}\gamma}$ 6.6 Hz, $H_{\delta1}$), 0.99 (3H, d, $J_{\delta2\text{-}\gamma}$ 6.6 Hz, $H_{\delta2}$). **P4:** 4.74-4.42 (1H, m, H_α), 2.25-2.06 (1H, m, $H_{\beta2}$) 2.05-1.94 (1H, m, $H_{\beta3}$), 2.25-2.06 (1H, m, $H_{\gamma2}$), 2.05-1.94 (1H, m, $H_{\gamma3}$), 3.89-3.82 (1H, m, $H_{\delta2}$), 3.70-3.61 (1H, m, $H_{\delta3}$). **E5**: 4.47-4.42 (1H, m, H_α), 2.25-2.06 (1H, m, $H_{\beta2}$), 2.05-1.94 (1H, m, $H_{\beta3}$), 2.49-2.44 (2H, m, $H_{\gamma2/3}$). **T6:** 4.47-4.42 (1H, m, H_α), 4.40 (1H, qd, $J_{\beta\text{-}\alpha}$ 3.1 Hz, $J_{\beta\text{-}\gamma2}$ 6.4 Hz, H_β), 1.18 (3H, d, $J_{\gamma2\text{-}\beta}$ 6.4 Hz, $H_{\gamma2}$). **OMe:** 3.74 (1H, s, H_α). ^{13}C NMR (125 MHz, CD_3OD): δ = **Y1:** 55.8 (C_α), 37.8 (C_β), 126.0 (C_γ), 131.6 (C_δ), 116.9 (C_ε), 158.3 (C_ζ). **A2:** 50.2 (C_α), 18.2 (C_β). **L3:** 51.1 (C_α), 41.0 (C_β), 25.9 (C_γ), 23.7 ($C_{\delta1}$), 21.9 ($C_{\delta2}$). **P4:** 61.5 (C_α), 26.0 (C_β), 30.4 (C_γ), 48.7 (C_δ), **E5:** 54.2 (C_α), 28.2 (C_β), 31.1 (C_γ). **T6**: 59.3 (C_α), 68.4 (C_β), 20.3 (C_γ). **OMe:** 52.7 (C_α). **CO (X 7):** 176.7, 174.3, 174.3, 174.0, 173.4, 172.3, 169.5. **Analytical HPLC** (below). **HRMS** of peak eluting at 1.90 min: Found [M + H^+] 707.3621, $C_{33}H_{51}N_6O_{11}$ requires 707.3610.

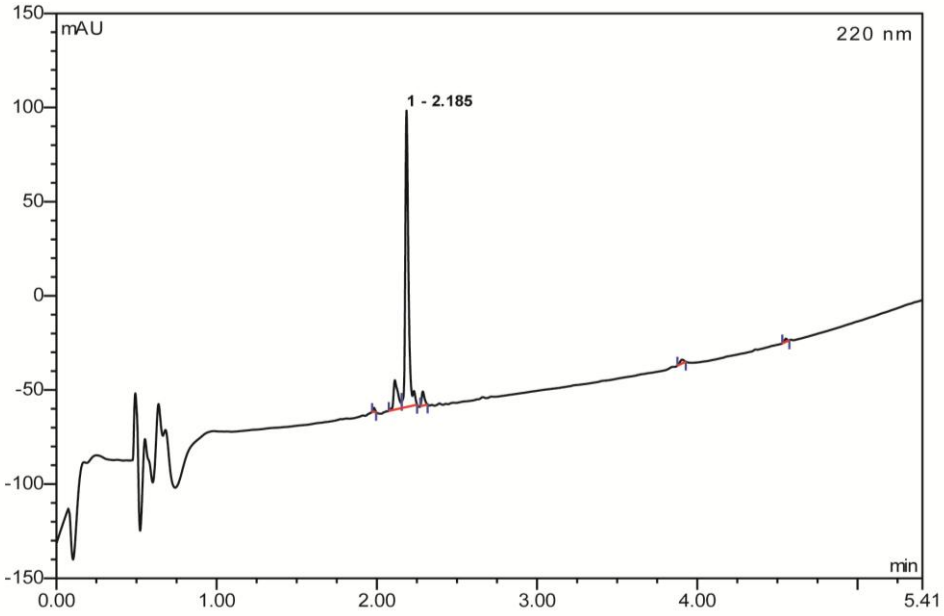


H_2N-K(Dansyl)-Y-A-L-P-E-T-[G]-G-COOH 2.36

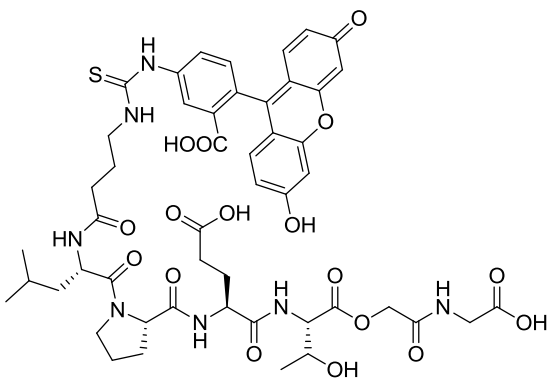


Yield: 120 mg, 89% ^{1}H NMR (500 MHz, CD_3OD): δ = **Dansyl:** 8.56 (1H, d, 8.5 Hz, Ar-H), 8.37 (1H, d, 8.7 Hz, Ar-H), 8.20 (1H, d, 7.3 Hz, Ar-H), 7.64-7.57 (2H, m, Ar-H), 7.32 (1H, d, 7.5 Hz, Ar-H), 2.92 (6H, s, $N(C\underline{H}_3)_2$) **K1:** 3.76 (1H, app t, $J_{\alpha\text{-}\beta2/3}$ 6.3 Hz, H_α), 1.78-1.69 (2H, m, $H_{\beta2/3}$), 1.78-1.69 (2H, m, $H_{\gamma2/3}$), 1.78-1.69 (2H, m, $H_{\delta2/3}$), 2.91-2.77 (2H, m, $H_{\varepsilon2/3}$). **Y2:** 4.67-4.60 (1H, m, H_α), 3.10 (1H, dd, $J_{\beta2\text{-}\alpha}$ 4.9 Hz, $J_{\beta2\text{-}\beta3}$ 14.2 Hz, $H_{\beta2}$), 2.91-2.77 (1H, m, $H_{\beta3}$), 7.10 (2H, d, $J_{\delta1/2\text{-}\varepsilon1/2}$ 8.5 Hz, $H_{\delta1/2}$), 6.70 (2H, d, $J_{\varepsilon1/2\text{-}\delta1/2}$ 8.5 Hz, $H_{\varepsilon1/2}$). **A3:** 4.42-4.37 (1H, m, H_α), 1.33 (3H, d, $J_{\beta2\text{-}\alpha}$ 7.1 Hz, $H_{\beta2}$). **L4:** 4.63-4.60 (1H, m, H_α), 1.64-1.54 (2H, m, $H_{\beta2/3}$), 1.80-1.68 (1H, m, H_γ), 0.99 (6H, d, $J_{\delta1\text{-}\gamma}$ 6.4 Hz, $H_{\delta1}$), 0.99 (6H, d, $J_{\delta2\text{-}\gamma}$ 6.4 Hz, $H_{\delta2}$), **P5:** 4.42-4.37 (1H, m, H_α), 2.23-2.03 (1H, m, $H_{\beta2}$) 2.02-1.92 (1H, m, $H_{\beta3}$), 2.23-2.03 (1H, m, $H_{\gamma2}$), 2.02-1.92 (1H, m, $H_{\gamma3}$), 3.86-3.79 (1H, m, $H_{\delta2}$), 3.69-3.59 (1H, m, $H_{\delta3}$). **E6**: 4.42-4.37 (1H, m, H_α), 2.23-2.03 (1H, m, $H_{\beta2}$), 2.02-1.92 (1H, m, $H_{\beta3}$), 2.51-2.44 (2H, m, $H_{\gamma2/3}$). **T7:** 4.58 (1H, d, $J_{\alpha\text{-}\beta2}$ 3.2 Hz, H_α), 4.42-

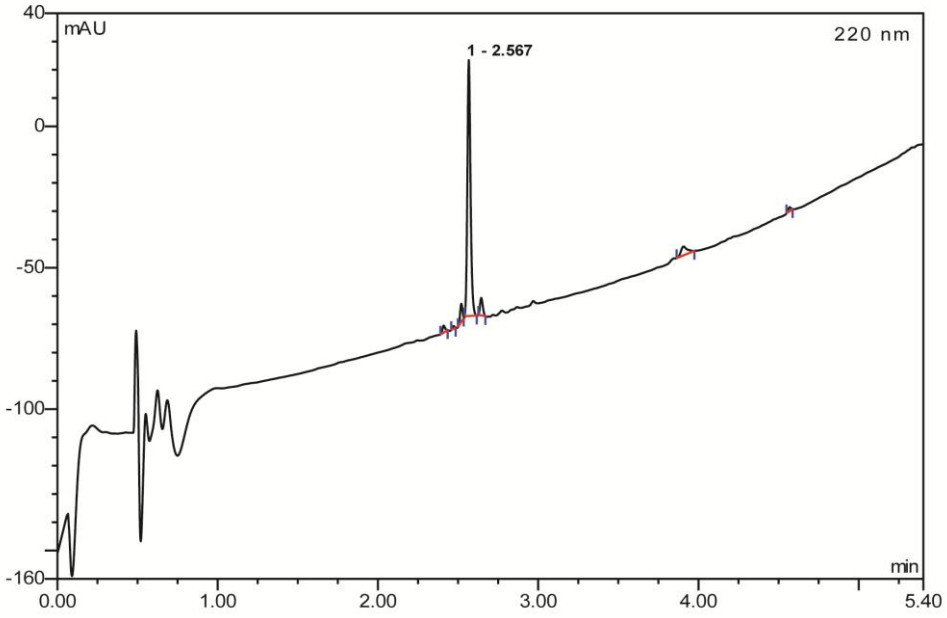
4.37 (1H, m, H_β), 1.20 (3H, d, $J_{\gamma2\text{-}\beta}$ 6.4 Hz, $H_{\gamma2}$). **G8:** 4.74-4.67 (2H, m, $H_{\alpha2/3}$), **G9:** 3.94 (1H, d, $J_{\alpha2\text{-}\alpha3}$ 17.7 Hz, $H_{\alpha2}$), 3.99 (1H, d, $J_{\alpha3\text{-}\alpha2}$ 17.7 Hz, $H_{\alpha3}$).^{13}C NMR (125 MHz, CD_3OD): δ = **Dansyl:** 152.5, 137.0, 131.0, 131.0, 130.9, 130.3, 129.2, 124.6, 121.0 (Ar-C), 46.0 ($N(CH_3)_2$). **K1:** 54.1 (C_α), 31.2 (C_β), 22.5 (C_γ), 30.4 (C_δ), 43.3 (C_ε), 158.3. **Y2:** 56.5 (C_α), 37.9 (C_β), 128.8 (C_γ), 131.3 (C_δ), 116.3 (C_ε), 157.4 (C_ζ). **A3:** 50.2 (C_α), 18.1 (C_β). **L4:** 51.3 (C_α), 41.1 (C_β), 25.8 (C_γ), 23.7 ($C_{\delta1}$), 22.0 ($C_{\delta2}$). **P5:** 61.7 (C_α), 26.0 (C_β), 30.3 (C_γ), 48.7 (C_δ), **E6:** 54.3 (C_α), 28.0 (C_β), 32.1 (C_γ). **T7**: 59.5 (C_α), 68.5 (C_β), 20.0 (C_γ). **G8:** 63.9 (C_α). **G9:** 41.5 (C_α). **CO (× 10):** 176.8, 174.6, 174.4, 174.2, 173.5, 173.3, 173.1, 172.6, 170.9, 170.1. **Analytical HPLC** (below). **HRMS** of peak eluting at 2.18 min: Found [M + Na^+] 1191.4994, $C_{54}H_{76}N_{10}NaO_{17}S$ requires 1191.5003.



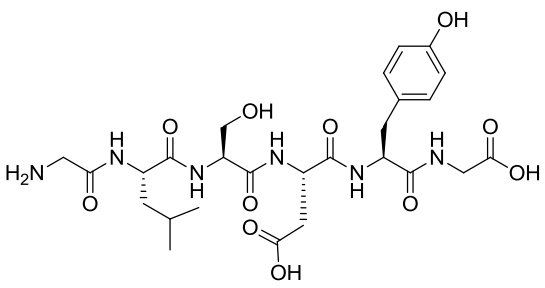
Fluorescein-GABA-L-P-E-T-[G]-G-COOH 2.37



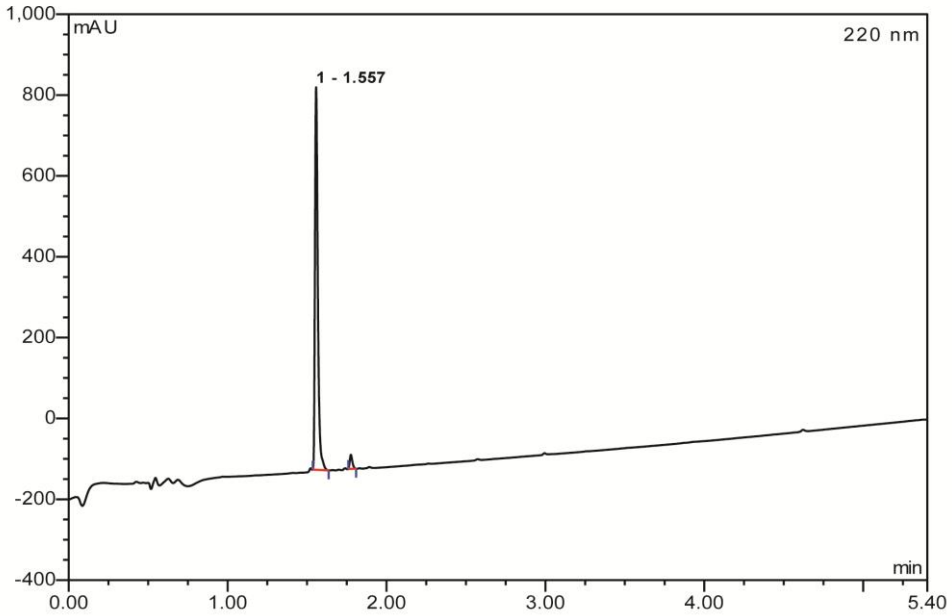
Yield: 60 mg, 52% ^{1}H NMR (500 MHz, CD_3OD): δ = **Fluorescein:** 8.24 (1H, s, Ar-H), 7.93-7.78 (1H, m, Ar-H), 7.22 (1H, d, 8.2 Hz, Ar-H), 6.99-6.87 (2H, m, Ar-H), 6.85 (2H, s, Ar-H), 6.75-6.60 (2H, m, Ar-H) **GABA1:** 2.37 (2H, app t, $J_{\alpha\text{-}\beta2/3}$ 6.9 Hz, H_α), 2.11-1.91 (2H, m, $H_{\beta2/3}$), 3.78-3.58 (2H, m, $H_{\gamma2/3}$) **L2:** 4.66 (1H, dd, $J_{\alpha\text{-}\beta2}$ 9.8 Hz, $J_{\alpha\text{-}\beta3}$ 4.7 Hz H_α), 1.67-1.56 (2H, m, $H_{\beta2/3}$), 1.82-1.69 (2H, m, $H_{\gamma2/3}$), 0.98 (3H, d, $J_{\delta1\text{-}\gamma}$ 6.4 Hz, $H_{\delta1}$), 0.97 (3H, d, $J_{\delta2\text{-}\gamma}$ 6.4 Hz, $H_{\delta2}$). **P3:** 4.49-4.38 (1H, m, H_α), 2.11-1.91 (2H, m, $H_{\beta2/3}$), 2.26-2.11 (1H, m, $H_{\gamma2}$), 2.11-1.91 (1H, m, $H_{\gamma3}$) 3.90-3.83 (1H, m, $H_{\delta2}$), 3.72-3.58 (1H, m, $H_{\delta3}$), **E4:** 4.49-4.38 (1H, m, H_α), 2.26-2.11 (1H, m, $H_{\beta2}$), 2.26-2.11 (1H, m, $H_{\beta3}$), 2.45 (2H, app t, $J_{\gamma\text{-}\beta2/3}$ 7.5 Hz $H_{\beta2}$) **T5:** 4.58 (1H, d, $J_{\alpha\text{-}\beta1}$ 7.5 Hz H_α), 4.49-4.38 (1H, m, $H_{\beta1}$), 1.21 (3H, d, $J_{\gamma\text{-}\beta1}$ 7.5 Hz, H_γ) **G8**: 4.70 (2H, s, $H_{\alpha2/3}$) **G9**: 3.99 (1H, d, $J_{\alpha2\text{-}\alpha3}$ 17.7 Hz $H_{\alpha2}$), 3.94 (1H, d, $J_{\alpha3\text{-}\alpha2}$ 17.7 Hz $H_{\alpha3}$),^{13}C NMR (125 MHz, CD_3OD): δ = **Fluorescein:** 155-103 broad signals (Ar-C) **GABA1:** 44.0 (C_α), 26.1 (C_β) 45.4 (C_γ), **L2:** 51.5 (C_α), 41.5 (C_β), 26.0 (C_γ), 23.6 ($C_{\delta1}$), 21.8 ($C_{\delta2}$) **P3:** 61.8 (C_α), 26.1 (C_β), 30.3 (C_γ), 48.4 (C_δ), **E4:** 54.3 (C_α), 28.0 (C_β), 31.2 (C_γ). **T5:** 59.5 (C_α), 68.6 (C_β), 20.0 (C_γ), **G6:** 63.9 (C_α), **G7:** 40.8 (C_α). **CO and CS:** 183-170 broad signals. **Analytical HPLC** (below). **HRMS** of peak eluting at 2.57 min: Found [M + H^+] 1048.3594, $C_{49}H_{57}N_7O_{17}S$ requires 1048.3604.



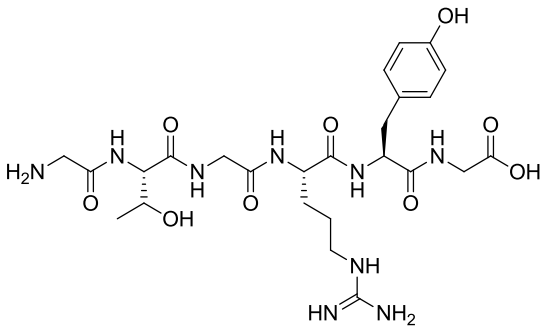
H_2N-G-L-S-D-Y-G-COOH 2.39



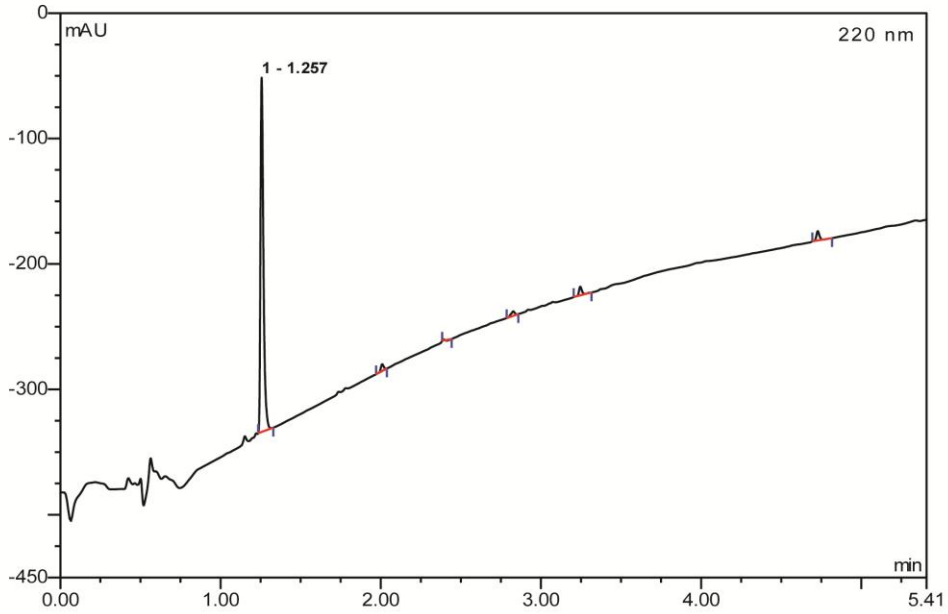
Yield: 40 mg, 60% ^{1}H NMR (500 MHz, D_2O): δ = **G1:** 3.95-3.79 (2H, m, $H_{\alpha 2/3}$), **L2:** 4.41 (1H, dd, $J_{\alpha\text{-}\beta 2}$ 5.5 Hz, $J_{\alpha\text{-}\beta 3}$ 9.2 Hz, H_α), 1.75-1.59 (2H, m, $H_{\beta 2/3}$), 1.75-1.59 (1H, m, H_γ) 0.96 (3H, d, $J_{\delta 1\text{-}\gamma}$ 6.2 Hz, $H_{\delta 1}$), 0.92 (3H, d, $J_{\delta 2\text{-}\gamma}$ 6.2 Hz, $H_{\delta 2}$) **S3:** 4.48 (1H, app t, $J_{\alpha\text{-}\beta 2/3}$ 5.6 Hz, H_α) 3.95-3.79 (2H, m, $H_{\beta 2/3}$), **D4**: 4.67 (1H, dd, $J_{\alpha\text{-}\beta 2}$ 5.6 Hz, $J_{\alpha\text{-}\beta 3}$ 7.8 Hz, H_α), 2.78 (1H, dd, $J_{\beta 2\text{-}\alpha}$ 5.6 Hz, $J_{\beta 2\text{-}\beta 3}$ 16.7 Hz, $H_{\beta 2}$), 2.69 (1H, dd, $J_{\beta 3\text{-}\alpha}$ 7.8 Hz, $J_{\beta 3\text{-}\beta 2}$ 16.7 Hz, H_α), **Y5:** 4.64 (1H, dd, $J_{\alpha\text{-}\beta 2}$ 5.9 Hz, $J_{\alpha\text{-}\beta 3}$ 8.9 Hz, H_α), 3.16 (1H, dd, $J_{\beta 2\text{-}\alpha}$ 5.9 Hz, $J_{\beta 2\text{-}\beta 3}$ 14.1 Hz, $H_{\beta 2}$), 2.95 (1H, dd, $J_{\beta 3\text{-}\alpha}$ 8.9 Hz, $J_{\beta 3\text{-}\beta 2\beta 2}$ 14.1 Hz, $H_{\beta 3}$), 7.17 (2H, d, $J_{\delta 1/2\text{-}\varepsilon 1/2}$ 8.6 Hz, $H_{\delta 1/2}$), 6.88 (2H, d, $J_{\varepsilon 1/2\text{-}\delta 1/2}$ 8.6 Hz, $H_{\varepsilon 1/2}$). **G6:** 3.95-3.79 (2H, m, $H_{\alpha 2/3}$).^{13}C NMR (125 MHz, D_2O): δ = **G1:** 42.2 (C_α). **L2:** 52.9 (C_α), 39.8 (C_β) 24.3 (C_γ) 22.0 ($C_{\delta 1}$) 20.7 ($C_{\delta 2}$) **S3:** 55.3 (C_α) 60.9 (C_β) **D4:** 50.7 (C_α), 36.1 (C_β), **Y5:** 55.0 (C_α), 36.3 (C_β), 128.3 (C_γ), 130.5 (C_δ), 115.5 (C_ε), 154.4 (C_ζ). **G6**: 40.4 (C_α). **CO (× 7):** 175.2, 174.8, 174.5, 172.9, 172.1, 171.3, 167.3. **Analytical HPLC** (below). **HRMS** of peak eluting at 1.56 min: Found [M + Na^+] 633.2463, $C_{26}H_{38}N_6O_{11}Na$ requires 633.2491.



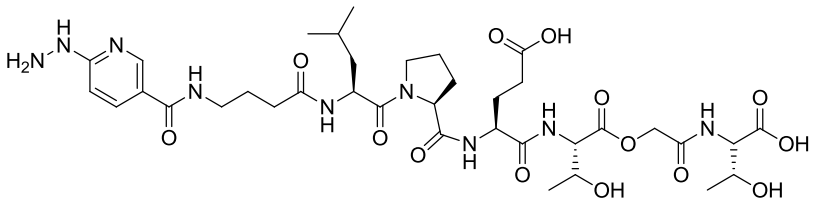
H_2N-G-T-G-R-Y-G-COOH 2.40



Yield: 36 mg, 53% ^{1}H NMR (500 MHz, D_2O): δ = **G1:** 4.01-3.91 (2H, m, $H_{\alpha 2/3}$), **T2:** 4.46 (1H, d, $J_{\alpha\text{-}\beta 2}$ 3.9 Hz, H_α), 4.30 (1H, qd, $J_{\beta 1\text{-}\alpha}$ 3.9 Hz, $J_{\beta 2\text{-}\gamma}$ 6.4 Hz, $H_{\beta 1}$), 1.27 (1H, dd, $J_{\gamma\text{-}\beta}$ 6.4 Hz, H_γ) **G3:** 4.01-3.91 (2H, m, $H_{\alpha 2/3}$) **R4**: 4.44 (1H, app t, $J_{\alpha\text{-}\beta 2/3}$ 7.2 Hz, H_α), 1.70-1.63 (2H, m, $H_{\beta 2/3}$), 1.47-1.39 (2H, m, $H_{\gamma 2/3}$), 3.13 (2H, app t, $J_{\delta\text{-}\gamma 2/3}$ 7.2 Hz H_δ), **Y5:** 4.70 (1H, dd, $J_{\alpha\text{-}\beta 3}$ 9.3 Hz, $J_{\alpha\text{-}\beta 2}$ 5.9 Hz, H_α), 3.20 (1H, dd, $J_{\beta 2\text{-}\alpha}$ 5.9 Hz, $J_{\beta 2\text{-}\beta 3}$ 14.0 Hz, $H_{\beta 2}$), 2.94 (1H, dd, $J_{\beta 3\text{-}\alpha}$ 9.3 Hz, $J_{\beta 3\text{-}\beta 2}$ 14.0 Hz, $H_{\beta 3}$), 7.19 (2H, d, $J_{\delta 1/2\text{-}\varepsilon 1/2}$ 8.1 Hz, $H_{\delta 1/2}$), 6.87 (2H, d, $J_{\varepsilon 1/2\text{-}\delta 1/2}$ 8.1 Hz, $H_{\varepsilon 1/2}$). **G6:** 4.01-3.91 (2H, m, $H_{\alpha 2/3}$) ^{13}C NMR (125 MHz, D_2O): δ = **G1:** 42.4 (C_α). **T2:** 59.1 (C_α), 67.1 (C_β) 18.6 (C_γ). **G3:** 41.4 (C_α) **R4:** 53.6 (C_α), 28.9 (C_β), 24.0 (C_γ) 40.4 (C_ε). **Y5:** 54.6 (C_α), 36.0 (C_β), 128.3 (C_γ), 130.6 (C_δ), 115.4 (C_ε), 154.4 (C_ζ). **G6**: 40.5 (C_α). **CO (× 6):** 173.4, 173.3, 173.3, 172.4, 171.3, 167.3 (**CNH**): 156.7. **Analytical HPLC** (below). **HRMS** of peak eluting at 1.26 min: Found [M + H^+] 610.2958, $C_{25}H_{40}N_9O_9$ requires 610.2944.

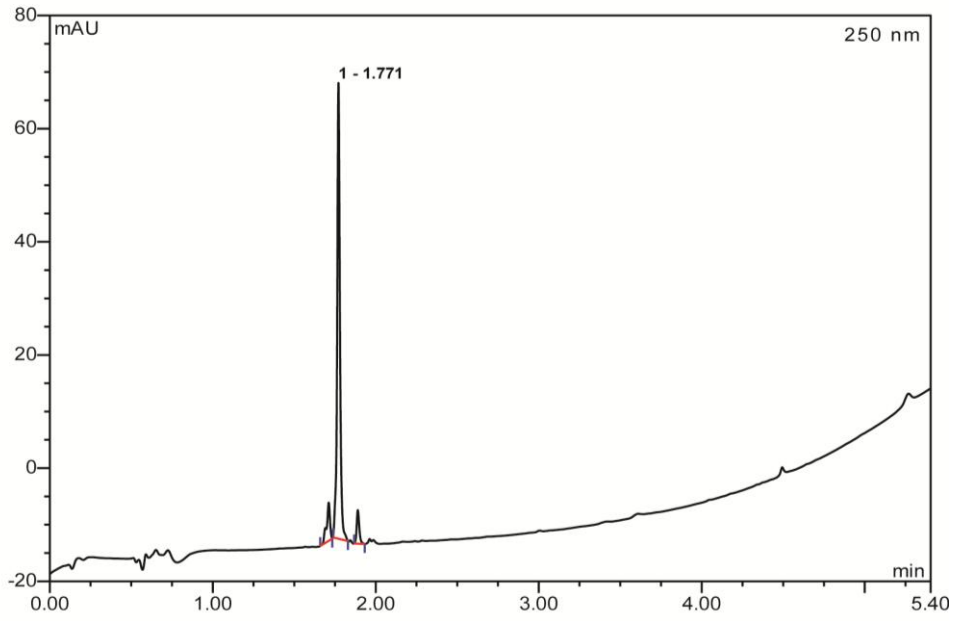


Hydrazine-GABA-L-P-E-T-[G]-T-OH 4.15

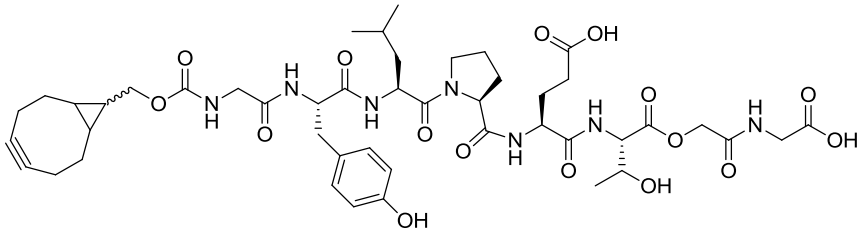


H_2N-(GABA)-LPET[G]T-OH was synthesised on H-Thr-2-chloro-trityl resin and isolated (Sigma) using standard SPPS methods (listed above). The lyophilised peptide (9.5 mg, 0.014 mmol) was dissolved in DMF (0.5 mL) before NEt_3 (3.9 μL, 0.028 mmol) and 2,5-Dioxo-1-pyrrolidinyl 6-(isopropylidenehydrazino)nicotinate **4.4** (7.5 mg, 0.025 mmol) were added. The

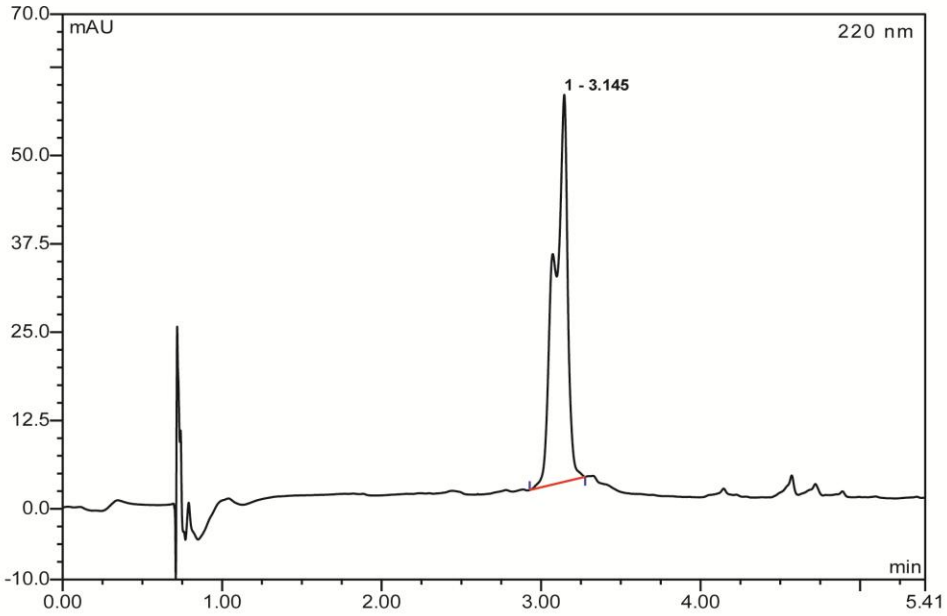
reaction mixture was stirred at r.t. for 4 hours and concentrated before the hydrazone-linked product was isolated using mass-directed HPLC. The HPLC fractions were combined and lyophilised to leave **4.15** as a white fluffy peptide (5 mg, 43%). **Analytical HPLC** (below). **HRMS** of peak eluting at 1.77 min: Found [M + H^+] 838.3920, $C_{36}H_{56}N_9O_{14}$ requires 838.3941.



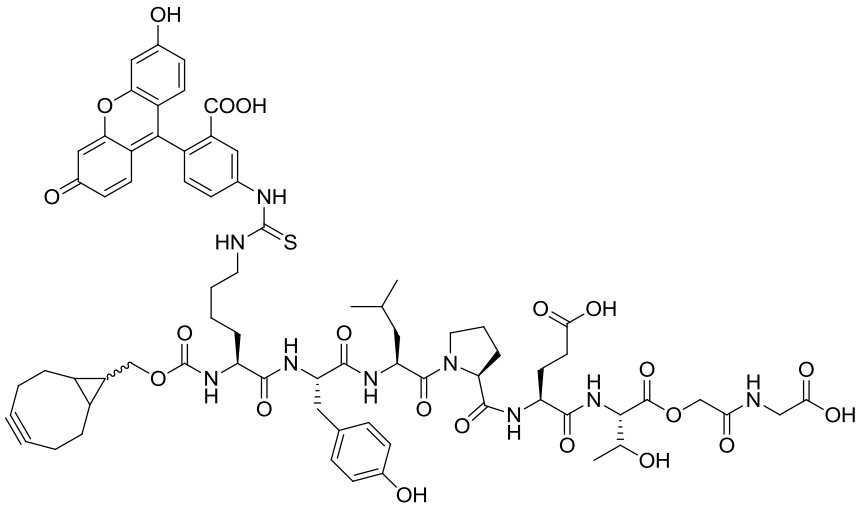
Alkyne-G-Y-L-P-E-T-[G]-G-OH 4.31



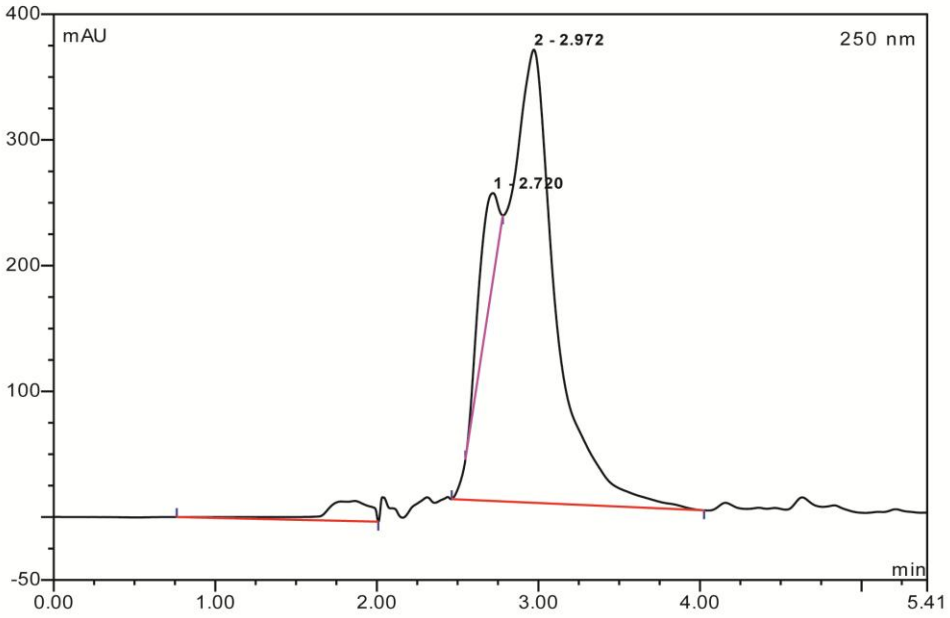
H_2N-GLPET[G]G-OH was synthesised on H-Gly-2-chloro-trityl resin and isolated (Sigma) using standard SPPS methods (listed above). The lyophilised peptide (21 mg, 0.028 mmol) and a diastereomeric mixture of (1R,8S,9r)-Bicyclo[6.1.0]non-4-yn-9-ylmethyl (4-nitrophenyl) carbonate 4.26 (14 mg, 0.043 mmol) were dissolved in DMF (2 mL) before NEt_3 (6 μL, 0.043 mmol) was added. The reaction mixture was stirred at r.t. overnight before being concentrated to leave an oil that was purified using mass-directed HPLC. The HPLC fractions were combined and lyophilised to leave **4.31** as a white fluffy peptide (7 mg, 26%). **Analytical HPLC** (below). **HRMS** of peak eluting at 3.14 min: Found [M + H^+] 970.4407, $C_{46}H_{64}N_7O_{16}$ requires 970.4404.



Alkyne-K(FITC)-Y-L-P-E-T-[G]-G-OH 4.37



H_2N-K(FITC)YLPET[G]G-OH was synthesised on H-Gly-2-chloro-trityl resin and isolated (Sigma) using standard SPPS methods (listed above). The lyophilised peptide (10 mg, 0.008 mmol) and a diastereomeric mixture of (1R,8S,9r)-Bicyclo[6.1.0]non-4-yn-9-ylmethyl (4-nitrophenyl) carbonate **4.26** (3.5 mg, 0.0011 mmol) were dissolved in DMF (2 mL) before NEt_3 (1.6 μL, 0.011 mmol) was added. The reaction mixture was stirred at r.t. for 6.5 h before an additional equivalent of both NEt_3 and the diastereomeric mixture of (1R,8S,9r)-Bicyclo[6.1.0]non-4-yn-9-ylmethyl (4-nitrophenyl) carbonate **4.26** were added. After stirring overnight at r.t., the reaction mixture was concentrated and the resulting oil purified by mass-directed HPLC. The HPLC fractions were combined and lyophilised to leave **4.37** as a white fluffy peptide (3.8 mg, 33%). **Analytical HPLC** (below). **HRMS** of peak eluting at 2.72 and 2.97 min: Found [M + $2H^+$] 715.7786, $C_{71}H_{85}N_9O_{21}S$ requires 715.7785.


mAU
250 nm
2 - 2.972
1 - 2.720
min
400
300
200
100
-50
0.00
1.00
2.00
3.00
4.00
5.41

6.2 Biological material and methods

6.2.1 Instrumentation and materials.

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 sterilise 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™ J-30I Centrifuge, a Heraeus multifuge 3 S-R centrifuge or a Heraeus pico centrifuge. Lactose affinity and size exclusion chromatography was achieved using a GE Pharmacia ÄKTA FPLC system. Spectrophotometric readings were measured using either a Thermo scientific Nanodrop 2000 or WPA Biowave II spectrometer. SDS-PAGE and agarose electrophoresis was performed using a Bio-Rad mini protean 3 apparatus and SCIE-PLAS Midi Horizontal Unit respectively. BIO RAD molecular imager^{R} Gel Doc^{TM} XR was used to visualise both the agarose and polyacrylamide gels using a combination of UV and white light. Protein concentrations were conducted using either a 10k or 30k MWCO Amicon® Ultra-15Centrifugal Filter Device. Dialysis was accomplished using SnakeSkin® pleated dialysis tubing (Thermo Scientific, 7k MWCO).

Mass spectrometry analysis of proteins was done using a using a Bruker HCT Ultra MS system equipped with an Agilent 1200 series autosampler. Protein samples were loaded directly into the instrument before they were automatically diluted into 0.1% TFA/50% MeCN (*v/v*) in H_2O prior mass spectrometry analysis.

Analytical grade reagents were supplied by Sigma-Aldrich, Fisher Scientific, Melford laboratories and VWR International.

Mammalian cell culture manipulations were performed inside a Thermo Scientific MSC-Advantage Class II Biological Safety Cabinet. A Sanyo CO_2 cell culture incubator was used to grow the cells Florescence microscopy analysis of mammalian cell culture slides was performed using a Zeiss LSM700 inverted confocal microscope coupled with ZEN imaging software. Wavelengths at 408 nm, 488 nm and 555 nm were used to excite the DAPI, Alexa fluor 555 and fluorescein fluorophores respectively. Primary and secondary antibodies for immunofluorescence were supplied by either Millipore or Invitrogen

.

6.2.2 Buffer solutions and media for biochemistry

All solutions were made up to 1 L, unless otherwise stated, with 18.2 MΩ purified water. The pH of the solutions were adjusted using 1M NaOH and/or 4M HCl

Standard buffers

TRIS buffer I (pH 7.2): 150 mM NaCl (8.7 g), 50 mM tris(hydroxymethyl)aminomethane (Tris, 6.1 g) and 10 % (*v/v*) glycerol

TRIS buffer II (pH 7.8): 1.25 M NaCl (72.5 g), 25 mM Tris (3.1 g) and 25 mM $CaCl_2$ (2.8 g)

TRIS buffer III (pH 7.8): 1.25 M NaCl (72.5 g), 25 mM Tris (3.1 g) and 2.5 mM Ethylenediaminetetraacetic acid (EDTA, 0.73 g)

TRIS buffer IV (pH 8.0): 50 mM Tris (6.1 g) and 1 mM β-mercaptoethanol (BME, 68 μL)

TRIS buffer V (pH 8.0): 150 mM NaCl (8.7 g), 50 mM Tris (6.1 g) and 1 mM BME (68 μL)

HEPES buffer (pH 7.5): 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, 11.9 g) 200 mM NaCl (11.6 g), and 5mM $CaCl_2$ (0.55 g)

Bacteria growth media components

20 × NPSC (100 mL, pH 6.8): 443 mM NH_4Cl (2.35 g), 227 mM Na_2SO_4 (3.22 g), 500 mM KH_2PO_4 (6.8 g) and 500 mM Na_2HPO_4 (7.1 g). The solution was autoclaved for 20 minutes at 121 °C.

50 × 5052 (100 mL): 2.7 M Glycerol (25 g), 138 mM glucose (2.5 g) and 292 mM α-lactose (10 g). The solution was heated in microwave until all the solids had dissolved before being filtered (0.20 μm) into a sterile container.

1000 × Metals Stock (100 mL): 50 mM $FeCl_3.6H_2O$ (1.35 g), 20 mM $CaCl_2.6H_2O$ (438 mg), 100 mM $MnCl_2.4H_2O$ (198 mg), 100 mM $ZnSO_4.7H_2O$ (288 mg), 1.7 mM $CoCl_2.6H_2O$ (47.6 mg), 2 mM $CuCl_2.2H_2O$ (34.1 mg), 4.1 mM $Na_2MoO_4.2H_2O$ (48.8 mg), 2 mM Na_2SeO_3 (34.6 mg), 2 mM H_3BO_3 (12.4 mg), 2 mM $NiSO_4.6H_2O$ (52.6 mg) and 60 μM HCl.

50 × Salts Stock (500 mL): 1.25 mM Na_2HPO_4 (88.7 mg), 1.25 mM KH_2PO_4 (85 mg), 2.5 mM NH_4Cl (67 mg) and 250 mM Na_2SO_4 (17.8 g)

Bacterial growth media

Auto Inducing media: Tryptone (10 g), yeast extract (5 g), 50 × salts (20 mL), 1000 × metals (200 μl), 2 mM $MgCl_2$ (48 mg) and 0.5% (*v/v*) glycerol (10 ml from a 50% (*v/v*) glycerol stock). The solution was autoclaved for 20 minutes at 121 °C before the remaining components of 0.05% (*w/v*) glucose (1 ml from a 50% (*v/v*) glucose stock) and 0.2% (*w/v*) lactose (4 ml from a 50% (*v/v*) lactose stock) solutions (sterile filtered, 0.2 μm Sartorius Minisart) and ampicillin (100 μg/mL) were added.

Lysogeny broth (LB)*:* Tryptone (10 g), 172 mM NaCl (10 g) and Yeast Extract (5 g). The solution was autoclaved for 20 minutes at 121 °C.

Buffers for DNA analysis and manipulations

50 × TAE Buffer: 2 M Tris-HCl (314 g), concentrated acetic acid (57.1 mL) and EDTA (0.5 M at pH 8.0, 100 mL)

6 × Agarose electrophoresis *loading buffer (10 ml, pH 7.6):* 15% (*w/v*) Ficoll 400, 66 mM EDTA (193 mg), 19.8 mM Tris-HCl (31 mg), 0.12% (*w/v*) SDS and 0.9% (*w/v*) Orange G.

10 × EcoRI buffer (10 mL, pH 7.5): 1 M Tris-HCl (1.57 g), 500 mM NaCl (292 mg), 100 mM $MgCl_2$ (95 mg) and 0.25% (*v/v*) Triton X-100

10 × Ligase Buffer (10 mL pH 7.5): 500 mM Tris-HCl (7.85 g), 100 mM $MgCl_2$ (95 mg) 10 mM adenosine triphosphate (ATP, 55 mg) and 100 mM dithiothreitol (DTT, 154 mg)

Buffers and solutions for protein analysis

SDS-PAGE loading buffer (10 mL, pH 6.8*):* 50 mM Tris-HCl (61 mg), 2% (*w/v*) SDS, 200 mM DTT (0.31 mg), 10% (*v/v*) glycerol, bromophenol blue (2 mg) and water (7.2 mL).

5× SDS-PAGE running buffer: 125 mM tris base (15.15 g), 960 mM glycine (72.1 g) and 0.5% (*w/v*) SDS

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

6.2.3 Molecular biology

6.2.3.1 Subcloning of the SrtA gene into Pet28A

The SrtA gene was subcloned from the pETMCS-III plasmid donated by Dr. C. Neylon (University of Southampton) into a pET28a plasmid using the molecular biology techniques described below.

Restriction digest

Insert preparation - A restriction digest was used to extract the SrtA gene from the pETMCS-III plasmid. The plasmid DNA (5 μL) was added to a solution of EcoRI (1 μL, 20 U), NdeI (1 μL, 20 U), 10 × EcoRI buffer (2 μL) and sterile water (11 μL). The reaction mixture was incubated at 37 °C for 1.5 h before the digested DNA was isolated by agarose gel electrophoresis and purified using a QIAquick Gel Extraction Kit.

Vector preparation - The pET28a vector was prepared for DNA ligation using a slightly modified restriction digest procedure. The plasmid DNA (5 μL) was added to a solution of EcoRI (1 μL, 20 U), NdeI (1 μL, 20 U), 10 × EcoRI buffer (2 μL) and sterile water (11 μL). The reaction mixture was incubated at 37 °C for 1 h before calf-intestinal phosphatase (1 μL, 10 U) was added. The incubation was continued at 37 °C for an additional hour before the linearised plasmid was isolated and purified as previously described.

DNA ligation

The SrtA gene insert and linear pET28a vector were ligated together to form a pET28a-I hybrid plasmid using the following procedure. A ligation mixture was prepared over ice using the linear plasmid vector (50 ng), gene insert (3:1 molar ratio of insert to vector), 10 × ligase buffer (1 μL), T4 ligase (1 μL, 20 U) and H_2O to make a total volume of 10 mL. The reaction mixture was incubated at 4 °C overnight before being transformed into chemically competent *E. coli* XL10 cells.

DNA transformation

Freshly thawed chemically competent *E. coli* XL10 cells (5 mL) were held on ice for 10 minutes before the ligation mixture (2 mL) was added and thoroughly mixed. The transformation mixture was incubated on ice for 20 minutes before being heated to 42 °C for 45 seconds and immediately returned to the ice for a further 10 minutes. LB media (900 μl) was added and the solution was incubated with shaking at 37 °C for 1 h. A sample of the culture (100 μl) was used to inoculate an LB-agar plate containing kanamcyin (50 mg/ml). The

remaining culture was centrifuged at 13 × kg for 1 minute and the excess supernatant was discarded. The cell pellet was resuspended in the residual supernatant (100 μL) and applied to another LB-agar plate containing kanamcyin (50 mg/ml). The LB-agar plates were incubated at 37 °C overnight.

A single colony from the LB-agar plates was used to inoculate LB media (5 ml) containing kanamcyin (50 μg/ml). The mini culture was incubated at 37 °C overnight before an aliquot (0.5 ml) was mixed with 80% (v/v) glycerol and stored at -80 °C as a cell stock. The remaining culture was centrifuged at 10 × kg for 10 minutes, the supernatant discarded and the cell pellet retained for DNA analysis.

DNA purification and sequence determination

The DNA was extracted from the bacterial cell pellet using a QIAprep Spin Mini Prep Kit and sequenced by GATC Biotech. The results were analysed by BioEdit sequence alignment editor ClustalW,[175] which confirmed the transformed pET28a-I plasmid contained the SrtA gene. The pET28a-I plasmid was extracted from the frozen stock of E. coli XL10 cells and transformed into *E. coli* BL21-Gold (DE3) cells for protein overexpression.

6.2.3.2 DNA analysis

DNA concentrations

The DNA concentration and yield were estimated using Equation 6.1 and the absorbance measurements at 260 nm and 320 nm.

$$Concentration\ (\mu g\, mL^{-1}) = (A_{260} - A_{320}) \times 50\ \mu g\, mL^{-1}$$

$$Yield\ (\mu g) = Concentration\ (\mu g\, mL^{-1})\ \times Volume\ (mL)$$

Equation 6.1: DNA concentration and yield calculations. A_{260} and A_{320} is the absorbances at 260 nm and 320 nm respectively

Agarose gel electrophoresis

The purity of plasmid and digested DNA was assessed using agarose gel electrophoresis. Agarose gels (1.2% *w/v*) were prepared by dissolving agarose (380 mg) in TEA buffer (40 mL) using microwave assisted heating for 1-2 minutes. The solution was allowed to cool for 5-10 minutes before ethidium bromide (1.2 μL, 0.3 μg / ml) was added and mixed. The solution was then immediately poured into a mould and a comb added before the gel was left to set. DNA (5 μL) was mixed with 6 × loading buffer (1 μL) and loaded onto the gel. Electrophoresis was

performed at 100 V for approximately 20 min in TAE buffer. Gels were then visualised under UV light using the Gel Doc system.

6.2.4 Protein overexpression, purification and analysis

6.2.4.1 Sortase overexpression from *E. coli*

A single colony of *E. coli* BL21-Gold (DE3) cells harbouring the pET28a-SrtA plasmid was used to inoculate LB media (5 mL, 50 μg/mL kanamycin). The starter culture was incubated at 37 °C for 18 h before 1 mL was added to *auto-inducing media* (1 L, 50 μg/mL kanamycin). The culture was grown at 37 °C for 24 h and the cells isolated by centrifugation (10 × kg, 10 min), the supernatant discarded and the bacterial pellet stored at -80 °C.

The frozen bacteria pellet was resuspended in *Tris buffer I (*10 mL) containing DNase I (5,000 units/mL) at 0 °C. The suspension was mechanically disrupted using a Constant Systems Cell Disruptor (20 kpsi, 10 mL injections, 4 °C), the lysate cleared by centrifugation (30 × kg, 45 minutes), the pellet discarded and the supernatant passed down a nickel affinity column (Qiagen, 10 ml) pre-equilibrated in *Tris buffer I* . The column was washed sequentially with *Tris buffer I* (5 × CV) and *Tris buffer I* supplemented with 50 mM imidazole (5 × CV) before the protein was eluted with *Tris buffer I* containing 500 mM imidazole (5 × CV). Protein-containing fractions were identified by SDS-PAGE and concentrated to a total volume of 1 mL by centrifugal concentration at 4 °C. The protein solution was applied to a Superdex 200 26/60 size-exclusion column and isocratically eluted into *HEPES buffer* before the purity of the isolated SrtA protein was analysed by SDS-PAGE (Figure 6.1). The protein concentration was measured by UV spectroscopy at 280 nm using a theoretical extinction molar coefficient of 17420 M^{-1} cm^{-1}.

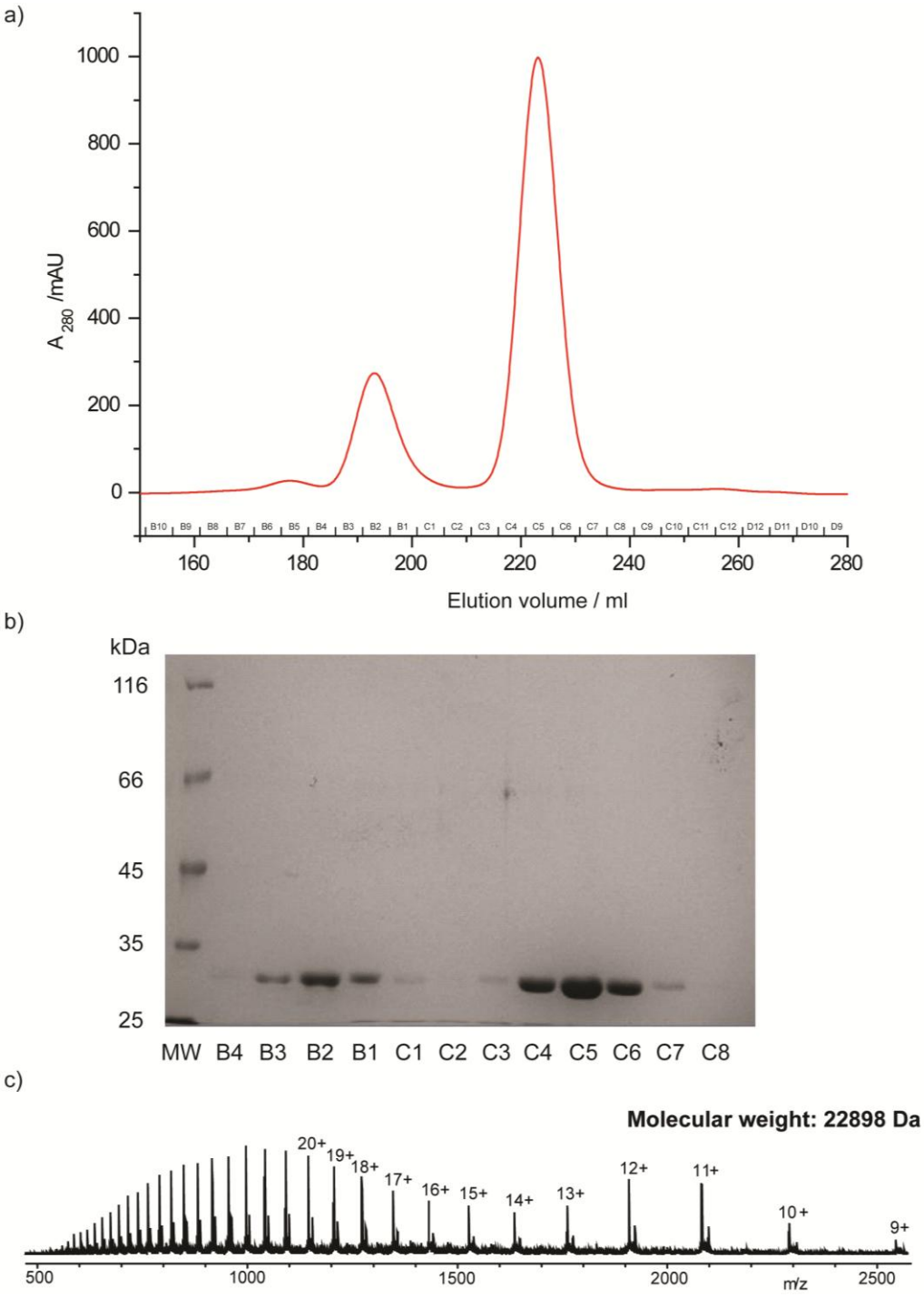


Figure 6.1: a) Chromatogram from the gel filtration of SrtA. b) Polyacrylamide gel (10%) of fractions recovered from the gel filtration indicates that both peaks are SrtA. c) ESMS spectrum of SrtA (Charge-states only partially labelled for clarity). Calculated molecular weight (23029 Da) is 131 Da higher than observed mass due to loss of *N*-terminal methionine residue during *E. coli* protein expression.

6.2.4.2 hMBL overexpression from *E. coli*

A frozen glycerol stock of E. *coli* Gold (DE3) cells harbouring the hMBP302 expression vector was used to inoculate LB media (5 mL, 100 μg/mL ampicillin). The starter culture was incubated at 37 °C for 18 h before 1 mL was added to LB media (100 mL, 100 μg/mL ampicillin), which was then grown at 30 ºC overnight before 20 mL was added to LB media (1 L, 100 μg/mL ampicillin). After 3.5 h at 24.5 ºC, expression was induced by addition of IPTG (10 mg, 40 μM) and 1M $CaCl_2$ (100 mL). After incubation overnight at 24.5 ºC, the cells were isolated by centrifugation at (10 k*g*, 15 min) and stored at −80 ºC.

The frozen cell pellet was resuspended in *Tris buffer II* (10 mL) and mechanically disrupted using a Constant Systems Cell Disruptor (25 kpsi, 4 ºC). The lysate was cleared by centrifugation (35 × kg, 45 min) and the supernatant filtered (0.8 μm filter, Sartorius Minisart) before loading onto a lactose-sepharose 6B affinity column (10 mL) equilibrated with *Tris buffer II*. The column was washed with *Tris buffer II* (5 × CV) before the protein was eluted with *Tris buffer III* (5 × CV). Protein-containing fractions were identified by SDS-PAGE, combined, concentrated by centrifugal concentration at 4 °C and extensively dialysed into *Tris buffer II*. The protein concentration was measured by UV spectroscopy at 280 nm using a theoretical extinction molar coefficient of 25230 M^{-1} cm^{-1}

6.2.4.3 MBP-AB_5 overexpression from *E. coli*

A frozen glycerol stock of E. *coli* BL21 Gold cells harboring the pSAB2.1 plasmid was used to inoculate LB media (5 mL, 100 μg/mL ampicillin). The starter culture was incubated at 37 °C for 18 h before 2.5 ml was added to LB media (1 L, 100 μg/mL ampicillin), which was then grown at 37 °C. Protein overexpression was induced once the OD_{600} had reached ~ 0.6 with the addition IPTG (1ml, 300 mM). Incubation was continued for 20 h at 30 °C before the cells were isolated by centrifugation (10 × kg, 10 min), the cell pellet discarded and the supernatant retained.

The supernatant was filtered (0.8 μm filter, Sartorius Minisart) and passed down an amylose column to capture MBP-AB_5 and uncomplexed MBP-CTA_2 proteins. The column was washed with *HEPES buffer* (5 × CV) and the proteins eluted with *HEPES buffer* supplemented with 10 mM maltose (5 × CV). Protein-containing fractions were identified by SDS-PAGE and applied to either a Nickel (Qiagen) or lactose (resin supplied by Sigma) affinity columns to isolate the MBP-AB_5 complex.

Nickel affinity purification: The column was washed sequentially with *HEPES buffer* (5 × CV) and *HEPES buffer* supplemented with 25 mM imidazole (5 × CV) before the protein was eluted with *HEPES buffer* containing 500 mM imidazole (5 × CV).

Lactose affinity purification: The column was washed with *HEPES buffer* (5 × CV) and the protein eluted with *HEPES buffer* containing 300 mM lactose (5 × CV).

The protein solution was concentrated to a total volume of 1 mL by centrifugal concentration at 4 °C before it was extensively dialysed into *HEPES buffer* at 4 °C. The purity of the isolated protein was determined by SDS-PAGE and the concentration measured by UV spectroscopy at 280 nm using a theoretical extinction molar coefficient of 128120 M^{-1} cm^{-1}

6.2.4.4 MBP-TEV overexpression from *E. coli*

A single colony of *E. coli* BL21-Gold (DE3) cells harbouring the pMAL-MBP-TEV plasmid was used to inoculate LB media (5 mL, 100 μg/mL ampicillin). The starter culture was incubated at 37 °C for 18 h before 3 ml was added to LB media (1 L, 100 μg/mL ampicillin), which was then grown at 37 °C. Protein overexpression was induced once the OD_{600} had reached ~ 0.7 with the addition IPTG (1ml, 500 mM). Incubation was continued for 20 h at 20 °C before the cells were isolated by centrifugation (10 × kg, 10 min), the supernatant discarded and the cell pellet retained.

The frozen bacteria pellet was resuspended in *Tris buffer IV (*10 mL) at 0 °C. The suspension was mechanically disrupted using a Constant Systems Cell Disruptor (20 kpsi, 10 mL injections, 4 °C), the lysate cleared by centrifugation (30 × kg, 45 minutes), the pellet discarded and the supernatant passed down a Amylose affinity column (Qiagen, 20 ml) equilibrated in *Tris buffer IV* . The column was washed with *Tris buffer IV* (5 × CV) and the protein eluted with *Tris buffer V* containing 10 mM Maltose (5 × CV). Protein-containing fractions were identified by SDS-PAGE and concentrated to a total volume of 1 mL by centrifugal concentration at 4 °C. The protein solution was applied to a size-exclusion column and isocratically eluted into *Tris buffer V* before the purity of the isolated SrtA was analysed by SDS-PAGE. The protein concentration was measured by UV spectroscopy at 280 nm using a theoretical extinction molar coefficient of 98329 M^{-1} cm^{-1}

6.2.4.5 Protein analysis

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

BioRad tetragel apparatus was used to determine the purity and size of proteins. A resolving gel of appropriate percentage was prepared using the materials listed in Table 6.1. After the

TEMED was added the solution was thoroughly mixed and immediately added to the apparatus before n-butanol (0.5 mL) was loaded onto the top of the gel. Once the resolving gel had set the n-butanol was carefully removed with H_2O and a stacking gel (5%) was prepared using the materials listed in Table 6.1. After the TEMED was added the solution was thoroughly mixed and immediately applied to the top of the resolving gel. A comb with a suitable number of lanes was inserted into the stacking gel layer before it set

	Resolving gel Volume / mL			Stacking gel Volume / mL
Components	10%	12%	15%	5%
H_2O	4	3.2	2.3	2.45
30 % acrylamide/bisacrylamide mix (37:5:1)	3.3	4	5	0.67
1.5 M Tris (pH 8.8)	2.5	2.5	2.5	N/A
1 M Tris (pH 6.8)	N/A	N/A	N/A	0.75
10% Sodium dodecyl sulfate (SDS)	0.1	0.1	0.1	0.04
20% or 10% Ammonium persulfate (APS)	0.1	0.1	0.1	0.04
Tetramethylethylenediamine (TEMED)	0.005	0.005	0.005	0.005

Table 6.1: Stacking and resolving gel components for SDS-PAGE analysis.

Once the gel had set it was sequestered inside an electrophoresis tank, submerged in an optimum volume of SDS running buffer and the comb was removed. Protein samples (10 μl) were mixed with an equal volume of loading buffer, heated to 95 °C for 10 minutes and applied to the lanes on the gel. However, unless AB_5 proteins are specifically referred to as "boiled" in gel figures the samples were not heated prior to application, but instead immediately loaded onto the gel. Electrophoresis was performed between 100-180 V until the blue loading buffer could be seen to reach the bottom of the gel. The gel was submerged in Coomassie Blue stain overnight before being treated with Coomassie destain for 2-4 h to allow visualisation of the protein bands. Alternatively, the gel was incubated directly with Instant Blue Stain (TripleRed) for 30 minutes to visualise the protein bands. The stained gel was either imaged straight away using or stored for longer periods in H_2O. **Gels containing fluorescently labelled protein samples were imaged prior to staining.**

Protein concentrations

Protein concentrations were determined using the Beer-Lambert law (Equation 6.2) and the absorbance measurement at 280 nm. Expasy ProtParam was used to calculate the theoretical molar extinction coefficient for all individual proteins

$$A = \epsilon c l$$

Equation 6.2: Beer-Lambert Law. A = absorbance, ε = molar extinction coefficient (mol^{-1} cm^{1}), c = concentration (mol) and l = path length (cm).

6.2.5 Sortase-mediated chemical ligations

6.2.5.1 Model peptide ligations

General procedure for HPLC assay

All stock solutions were prepared in *HEPES buffer* at the following concentrations: Acyl acceptor **2.8** (8 mM), acyl donor **2.5, 2.7, 2.24, 2.25** or **2.35** (2 mM), and SrtA (304 μM).

Typical procedure for HPLC assay (1:1 ratio)

For a total volume of 200 μL in the ratio 1:1, acyl acceptor (25 μL, 500 μM), acyl donor (50 μL, 500 μM), SrtA (7.5 μL, 11.5 μM) and *HEPES buffer* (117.5 μL) were added to a tapered HPLC vial. The solution was mixed thoroughly and capped before being immediately incubated at 37 °C, in the auto loader of a Agilent technologies 1290 HPLC, fitted with Ascentis es-C18 (10 cm × 2.1 mm, 2.7 μm) column.

Each 5 μL sample was injected and analysed using a linear gradient (5% H_2O, 0.5% TFA in MeCN to 95% H_2O, 0.5%TFA over 6.5 min. Typically 5 reactions were assayed in parallel: each reaction mixture was prepared in the final 3 minutes of the 1st analysis of the previous reaction mixture. After every five analyses, 20 μL of water was injected onto the column and the column washed with 95% H_2O, 5% MeCN, 0.5% TFA for 2.5 min to remove any residual peptides prior to the subsequent analyses. The HPLC analysis was carried out over a 9 h period (16 analyses per sample). Integrated traces were normalised using the peak intensity for SrtA (retention time: 2.2 min).

Sample HPLC for typical ligation assay

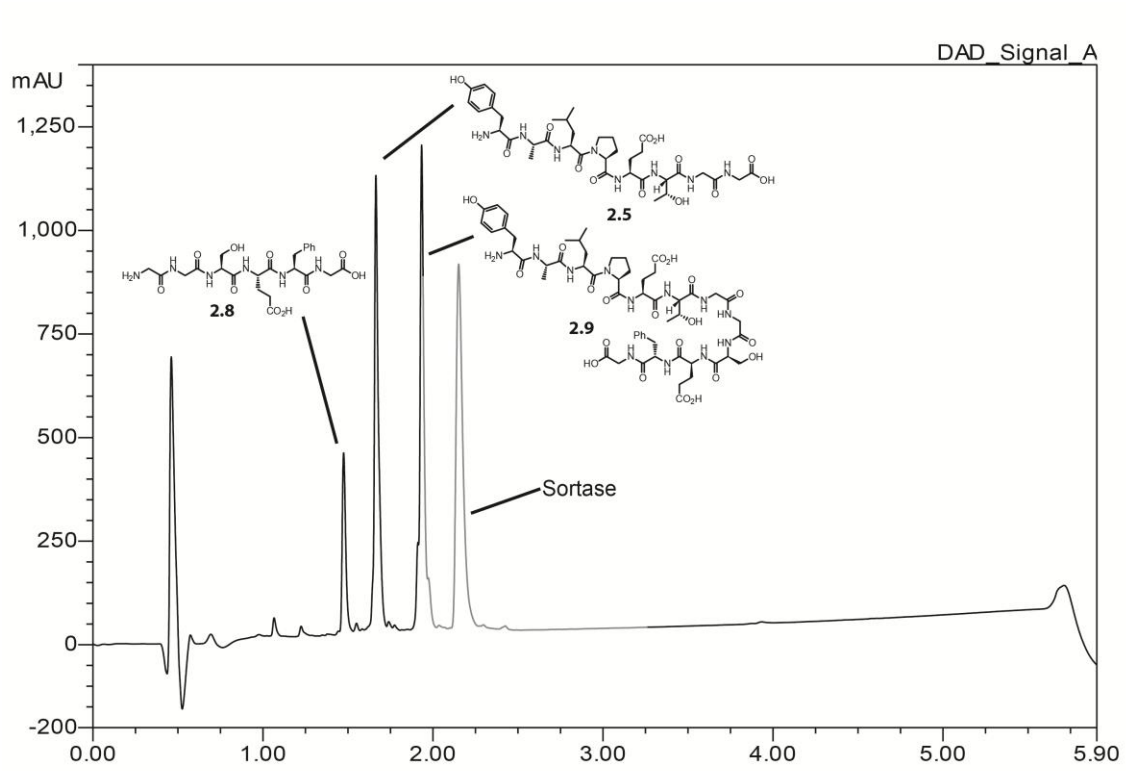


Figure 6.2: Sample HPLC for typical SrtA ligation assay between acyl donor **2.5** and acyl acceptor **2.8** to form ligated product **2.9.**

6.2.5.2 Protein labelling and purification

General procedure

SrtA, proteins and depsipeptides stocks were prepared in *HEPES buffer*.

SrtA (10-30 mol%), protein (1 equivalent) and depsipeptide substrate (1.5-5 equivalents) were mixed together in the desired volume of *HEPES buffer*, typically 50-500 μL. The ligation mixture was incubated with agitation at 37°C for 4-6 h or until ligation was complete. The reaction progress was monitored every 2-3 h by ESMS and if necessary additional depsipeptide substrate (1.5-5) was added to the reaction mixture to ensure complete modification of proteins.

Typical procedure for hMBL labelling

*Labelling with dansyl-depsipeptide **2.36**:* For a total volume of 100 μL, hMBL (50 μL, 60 μM), depsipeptide **2.36** (9 μL, 90 μM, 1.5 equivalents), SrtA (2 μL, 6 μM, 10 mol%). and *HEPES*

buffer (39 μL) were together mixed thoroughly and incubated at 37 °C for 4 h. The reaction was monitored by ESMS and SDS-PAGE.

*Labelling with fluorescein-depsipeptide **2.37**:* For a total volume of 20 μL, hMBL (17.5 μL, 60 μM), depsipeptide **2.37** (1.25 μL, 120 μM, 2 equivalents) and SrtA (1.25 μL, 9 μM, 15 mol%.) were together mixed thoroughly and incubated at 37 °C. After 4 h an additional 0.63 μL of depsipeptide **2.37** (1 mol equivalent) was added and the incubation continued for a further 1.5 h. The reaction was monitored by ESMS and SDS-PAGE.

Typical procedure for Mouse pumilio labelling

*Labelling with dansyl-depsipeptide **2.36***: For a total volume of 15 μL, Mouse pumilio (11 μL, 19 μM), depsipeptide **2.36** (1.8 μL, 60 μM, 3 equivalents), SrtA (0.36 μL, 7.2 μM, 36 mol%) and *HEPES buffer* (1.9 μL) were together mixed thoroughly and incubated at 37 °C for 4 h. The reaction was monitored by ESMS and SDS-PAGE.

Typical procedure for CTA2-AB_5 labelling

*Labelling with dansyl-depsipeptide **2.36:*** For a total volume of 20 μL, CTA_2-AB_5 (17.6 μL, 44 μM), depsipeptide **2.36** (1.4 μL, 71 μM, 1.6 equivalents) and SrtA (1 μL, 3.8 μM, 9 mol%) were mixed together thoroughly and incubated at 37 °C for 3 h. The reaction was monitored by ESMS.

*Labelling with fluorescein-depsipeptide **2.37**:* For a total volume of 250 μL, CTA_2-AB_5 (200 μL, 43 μM), depsipeptide **2.37** (11.2 μL, 90 μM, 2 equivalents), SrtA (4.2 μL, 5 μM, 12 mol%) and *HEPES buffer* (34.5 μL) were mixed together thoroughly and incubated at 37 °C for 2 h. The reaction was monitored by ESMS. **When this reaction has been repeated it has taken anywhere between 2-5 equivalents of depsipeptide, 10-30 mol% of SrtA and 3-6 h to achieve complete protein modification.**

*Labelling with alkyne depsipeptide **4.31:*** For a total volume of 126 μL, CTA_2-AB_5 (100 μL, 56 μM), depsipeptide **4.31** (17.4 μL, 138 μM, 2.5 equivalents) and SrtA (8.5 μL, 11 μM, 20 mol%) were mixed together thoroughly and incubated at 37 °C for 3 h. The reaction was monitored by ESMS. **When the reaction has been repeated it has taken anywhere between 2-5 equivalents of depsipeptide, 10-20 mol% of SrtA and 2-5 h to achieve complete protein modification.**

Purification of labelled CTA2-AB5 proteins

Once the protein had been completely modified it was immediately purified using either Nickel or lactose affinity chromatography.

Nickel affinity purification: The column was washed sequentially with *HEPES buffer* (5 × CV) and *HEPES buffer* supplemented with 25 mM imidazole (5 × CV) before the protein was eluted with *HEPES buffer* containing 50 mM imidazole (5 × CV).

Lactose affinity purification: The column was washed with *HEPES buffer* (5 × CV) and the protein eluted with *HEPES buffer* containing 300 mM lactose (5 × CV).

The modified-protein solution was concentrated to a total volume of 1 mL by centrifugal concentration at 4 °C before being extensively dialysed into *HEPES buffer*. The purity of the isolated protein was determined by SDS-PAGE.

6.2.6 Bioconjugation reactions

Proteins and azide stocks were prepared in *HEPES buffer*

Typical procedure for "one pot" CTA_2-AB_5 labelling with galactosyl azide

For a total volume of 283 μL, CTA_2-AB_5 (200 μL, 30 μM), alkyne-depsipeptide **4.31** (38 μL, 150 μM, 5 equivalents), SrtA (37 μL, 6 μM, 20 mol%) and *HEPES buffer* (8 μL) were thoroughly mixed together and incubated at 37 °C for 3 h. A further 6.9 μL of alkyne depsipeptide (30 μM, 1 equivalent) and galactosyl azide **4.34** (14.6 μL, 2.1 mM) were added giving a 1:12 alkyne:zide ratio. The incubation was continued at room temperature overnight and extent of the azide /protein bioconjugation was monitored by ESMS.

Typical procedure for Alkyne-AB_5 labelling with duplex-RNA azide

For a total volume of 20 μL, Alkyne-AB_5 (10 μL, 27.5 μM, 1 equivalent), duplex-RNA azide (3.2 μL, 100 μM, 3 equivalents) and *HEPES buffer* (6.8 μL) were thoroughly mixed together and incubated at 37 °C overnight. The extent of the RNA/protein bioconjugation was analysed by SDS-PAGE

6.2.7 Other enzyme reactions

Tobacco Etch Virus protease cleavage of MBP-AB_5

Typical procedure: MBP-AB_5 (500 μL, 90 μM) was incubated in the presence of TEV protease (40 μL, 4.4 μM, 4 mol%) at room temperature for 60-120 minutes. The reaction was monitored by ESMS until only cleavage product could be observed. The reaction mixture was immediately passed down a Nickel affinity column (Qiagen) pre-equilibrated in *HEPES buffer*. The column was washed with *HEPES buffer* (5 × CV) and the protein eluted with *HEPES buffer* supplemented with 500 mM imidazole (5 × CV). Protein-containing fractions were combined, centrifugally concentrated at 4 °C and dialysed extensively into *HEPES buffer* at 4 °C.

6.2.8 Mammalian cell culture

Media, buffers and reagents

PBS (100 mL): A single phosphate buffered saline (PBS) tablet was dissolved H_2O. The solution was autoclaved for 20 minutes at 121 °C.

1X trypsin solution (50 mL): Trypsin-EDTA Solution 10X (Sigma, 5 mL) was diluted (0.2 μm) into to *PBS* and filter-sterilised (0.2 μm).

DMEM (500 mL): Dulbecco's modified Eagle media with high glucose and L-glutamine (GE healthcare) additional supplemented with penicillin (100 unit /ml), streptomycin (100 μg/ml), 10% (*v/v*) foetal calf serum and glutamine (4 mM).

1% or 4% BSA (50 mL): 1% or 4% (w/v) bovine serum albumin was dissolved in *PBS*

0.1 % Triton (50 mL): 0.1% (v/v) triton (50 μL) was dissolved in PBS

4% *PFA* (100 mL)*:* 4% (w/v) paraformaldehyde was dissolved in PBS with heating (70 °C), filtered (0.2 μm) and left to cool to room temperature before use.

General passage procedure

Once the mammalian cells had reached the optimal confluency (~ 80%) the growth media was removed and the cells carefully washed with *PBS* (10 ml) before *1X trypsin solution* (4 ml) was added. The cells were incubated at 37 °C for 5 minutes before the cleavage reaction was quenched with the addition of *DMEM* (5 ml). The excess cell suspension was removed and either retained for further experiments or disposed off. *DMEM* was added to the remaining cell

suspension leaving a total volume of 12 ml with a 1:3 cell suspension: media ratio. The cell culture was then incubated at 37 °C (5% CO_2) for roughly 2 days before the process was repeated.

Determining cell number and viability

Cell suspension (10 μL) was mixed with an equal volume of Trypan Blue and applied to a haemocytometer. The cells were analysed using a microscope at 100 × magnification and the clear viable cells counted – any cells stained blue were dead and not included in the count. Using Equation 6.3 the average number of cells per mL was calculated.

$$Average\ no.\ of\ cells\ per\ mL = Cell\ count\ \times\ dilution\ factor\ \times\ CF$$

Equation 6.3: Average number of cells per mL in cell suspension. CF = Correction factor (1×10^4)

Cell preparation for transfection assay

Cover slips were sequestered in a 12-well plate, sterilised with 70% ethanol (1 mL) and washed with *PBS* (1 mL). If HEK293 cells were used the cover slips were treated with poly-L-lysine (1 ml) for 15 minutes at room temperature and extensively washed with *PBS* (5 × 1 mL). The cell suspension was diluted into *DMEM* (Equation 6.4) and 1 mL applied to each cover slip. The culture was incubated at 37 °C (5% CO_2) for 1-2 days until ~80% confluency was achieved

$$Seed\ cells\ (mL\ per\ mL)\ = \frac{\text{Average no. of cells per mL}}{1\ \times\ 10^5}\ \times CF$$

Equation 6.4: No. of μL of cell suspension required per mL of media for optimal seed cell concentration. CF = correction factor for optimum cell growth: 1 for HeLa and HEK293 cells; 2 for Vero Cells

Typical protein transfection procedure

Mammalian cells were grown in the desired number of well plates and once they had reached ~80% confluency the growth media was removed and the cells washed with PBS (1 ml) before FITC-AB_5 in *PBS* (500 μL, 20-40 μg per mL) was added. The cells were completely covered and incubated at 37°C (5% CO_2) or 4 °C for the required time period. The transfection reaction was quenched by the chemical fixation of the cells to the cover slips.

Cell fixation

The growth media or transfection solution was removed from each well plate and the cells were washed with PBS (2 × 1 mL) before being treated with 4% *PFA* (1 ml) for 10 minutes at room temperature. The cells were washed and stored in *PBS* (2 × 1 mL) at 4 °C for no longer than a week. Transfection experiments were stored in the dark to prevent photo-beaching of the fluorophore.

Cell permeabilisation, immunofluorescence staining and slide immobilisation

The *PBS* was removed from the fixed cells before they were permeabilised with 0.1 % Triton (1 mL) for 10 minutes at room temperature. The cells were washed with *PBS* (1 mL) and incubated with *4% BSA* (1 mL) for 1 h at room temperature. The *BSA* was removed and the cells treated with the required primary antibody (**Error! Reference source not found.**) in *1% BSA* (500 μL) or 1 h at room temperature. The cells were extensively washed with *PBS* (5 × 1 mL) before the secondary antibody (Table 6.3) in *1% BSA* (500 μL) was added. After 1 h incubation at room temperature the cells were washed thoroughly with *PBS* (5 × 1 mL) before being mounted onto a glass slide with ProLong® Gold Antifade Reagent with DAPI (10 μL) - DAPI reagent allows visualisation of the nucleus. The slides were left overnight in the dark before being sealed with nail varnish to preserve the cells. The cells were stored at 4 °C in the dark for weeks to months without degradation

Primary polyclonal antibody	Derivation from	Target	Recommended dilution
EEA1	Rabbit	Early endosome	1/250
CD63	Mouse	Late Endosome	1/500
Lamp-1	Rabbit	Lysosome	1/500
Calreticulin	Rabbit	Enoplasmic reticulin	1/500
Calnexin	Rabbit	Enoplasmic reticulin	1/500
TGN46	Rabbit	Trans-golgi	1/500

Table 6.2: List of primary antibodies used in immunofluorescence assay

Secondary antibody	Derivation	Target	Recommended dilution
Alexa Fluorphore 555	Goat	Rabbit	1/300
Alexa Fluorphore 555	Hamster	Mouse	1/300

Table 6.3: List of secondary antibodies used in immunofluorescence assay

Chapter 7: Appendix

7.1 DNA plasmid sequences

The genes encoding SrtA, TEV-MBP, CTA2-MBP and CTB are highlighted in blue, orange, green and red respectively.

7.1.1 pET28a I

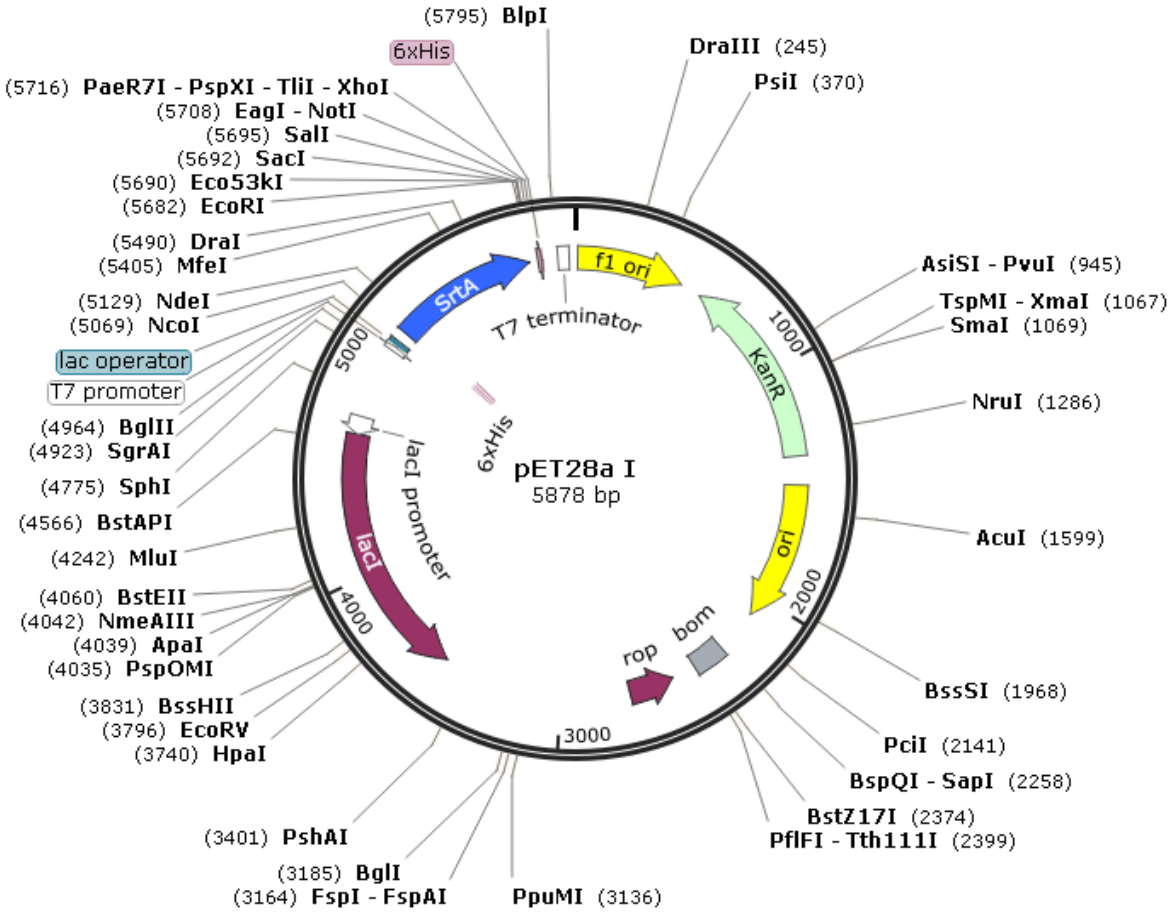


Figure 7.1: Plasmid map for pET28a I

The pET28a I plasmid was used to express SrtA in *E. coli* cells (Figure 7.1). It is derived from a pET28a plasmid obtained from Dr Jeff Hollins (University of Leeds). The SrtA gene was subcloned into the plasmid to form the resulting pET28a I expression vector

Full sequence 5'-3':

TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGC
GTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCG
CCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAG
TGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCG

CCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGT
TCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCC
GATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAA
ATATTAACGTTTACAATTTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTT
TATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAACTCATCGA
GCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCG
TTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGG
TCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGT
TATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCAT
TTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACC
AAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGAC
AATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTC
ACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGT
AACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCA
GCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAG
AAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACA
TTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAG
AGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGA
CAGTTTTATTGTTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCC
CGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAA
ACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTC
CGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTT
AGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCA
GTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGG
ATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGAC
CTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGA
AAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAG
GGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATT
TTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGG
TTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGG
ATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAG
CGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGC
GGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAG
CCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACAC
CCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGT
CTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGG
TAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCT
CGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGT

TTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATA
CCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGG
AACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGG
TCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCG
ATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACAC
GGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCA
CGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGG
GTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGCCT
GCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGAT
TCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAA
ATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTG
CGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGG
CATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTG
CCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGA
GAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCT
GATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAG
CAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCG
TCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTG
CGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCAT
TTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGA
ATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTA
ATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCG
CGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAAT
AACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGT
TAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGAC
GCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATC
GCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACG
ACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGC
TTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTC
TGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCC
TGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGT
GTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTT
GAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCC
CCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAG
CCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGT
GATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGAC
TCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAA
GAAGGAGATATACC**ATGGGCAGCAGCCATCATCATCATCACAGCAGCGGCCTGGTGCCGCG**

CGGCAGCCATATGAAACCACATATCGATAATTATCTTCACGATAAAGATAAAGATGAAAAGATT
GAACAATATGATAAAAATGTAAAAGAACAGGCGAGTAAAGATAAAAAGCAGCAAGCTAAACCTC
AAATTCCGAAAGATAAATCGAAAGTGGCAGGCTATATTGAAATTCCAGATGCTGATATTAAAGA
ACCAGTATATCCAGGACCAGCAACACCTGAACAATTAAATAGAGGTGTAAGCTTTGCAGAAGAA
AATGAATCACTAGATGATCAAAATATTTCAATTGCAGGACACACTTTCATTGACCGTCCGAACT
ATCAATTTACAAATCTTAAAGCAGCCAAAAAAGGTAGTATGGTGTACTTTAAAGTTGGTAATGA
AACACGTAAGTATAAAATGACAAGTATAAGAGATGTTAAGCCTACAGATGTAGGAGTTCTAGAT
GAACAAAAAGGTAAAGATAAACAATTAACATTAATTACTTGTGATGATTACAATGAAAAGACAG
GCGTTTGGGAAAAACGTAAAATCTTTGTAGCTACAGAAGTCAAATAATCGAATTCGAGCTCCGT
CGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAA
GCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGG
CCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT

7.1.2 pMAL-c5x I

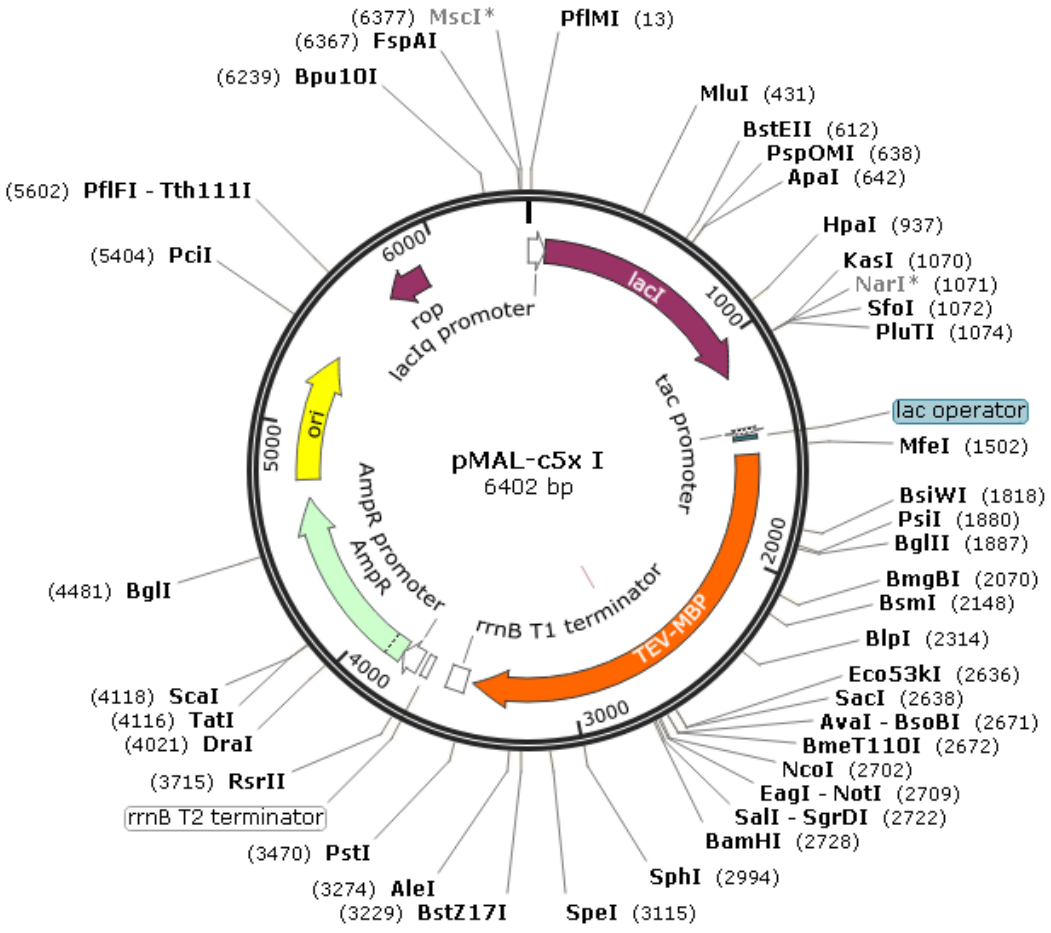


Figure 7.2: Plasmid map for pMAL-c5x I

The pMAL-c5x I plasmid was used to express TEV-MBP in *E. coli* cells (Figure 7.2). It is derived from a pMAL-c5x plasmid obtained from New England Biolabs. The TEV gene was introduced by Dr James Ross (University of Leeds) to form the pMAL-c5x I expression vector for the TEV-MBP fusion.

Full sequence 5'-3':

CCGACACCATCGAATGGTGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCA
ATTCAGGGTGGTGAATGTGAAACCAGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCT
TATCAGACCGTTTCCCGCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAG
TGGAAGCGGCGATGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAA
ACAGTCGTTGCTGATTGGCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTC
GCGGCGATTAAATCTCGCGCCGATCAACTGGGTGCCAGCGTGGTGGTGTCGATGGTAGAACGAA
GCGGCGTCGAAGCCTGTAAAGCGGCGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGAT
CATTAACTATCCGCTGGATGACCAGGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCG
GCGTTATTTCTTGATGTCTCTGACCAGACACCCATCAACAGTATTATTTTCTCCCATGAAGACG
GTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGCGGG
CCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAAT
CAAATTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCA
TGCAAATGCTGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAGATGGCGCT
GGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGATATTTCGGTAGTGGGA
TACGACGATACCGAAGACAGCTCATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTC
GCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGAAGGG
CAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAACC
GCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAA
GCGGGCAGTGAGCGCAACGCAATTAATGTAAGTTAGCTCACTCATTAGGCACAATTCTCATGTT
TGACAGCTTATCATCGACTGCACGGTGCACCAATGCTTCTGGCGTCAGGCAGCCATCGGAAGCT
GTGGTATGGCTGTGCAGGTCGTAAATCACTGCATAATTCGTGTCGCTCAAGGCGCACTCCCGTT
CTGGATAATGTTTTTTGCGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTG
ACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAA
CAGCCAGTCCGTTTAGGTGTTTTCACGAGCAATTGACCAACAAGGACCATAGATT**ATGAAAATC**
GAAGAAGGTAAACTGGTAATCTGGATTAACGGCGATAAAGGCTATAACGGTCTCGCTGAAGTCG
GTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTGAGCATCCGGATAAACTGGAAGA
GAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTCTGGGCACACGACCGC
TTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTTCCAGGACA
AGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCGC
TGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAA
GAGATCCCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGC
AAGAACCGTACTTCACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAA

CGGCAAGTACGACATTAAAGACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTC
CTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCGATTACTCCATCGCAGAAGCTG
CCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACAC
CAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCAAGGGTCAACCATCCAAACCGTTC
GTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAGCTGGCAAAAGAGTTCC
TCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGTGC
CGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGTGAAAGATCCGCGTATTGCCGCCACTATGGAA
AACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGC
GTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCA
GACTAATTCGAGCTCGAACAACAACAACAATAACAATAACAACAACCTCGGGATCGAGGGAAGG
ATTTCACATATGTCCATGGGCGGCCGCGATATCGTCGACGGATCCGGAAAAAGCTTGTTTAAGG
GGCCGCGTGATTACAACCCGATATCGAGCACCATTTGTCATTTGACGAATGAATCTGATGGGCA
CACAACATCGTTGTATGGTATTGGATTTGGTCCCTTCATCATTACAAACAAGCACTTGTTTAGA
AGAAATAATGGAACACTGTTGGTCCAATCACTACATGGTGTATTCAAGGTCAAGAACACCACGA
CTTTGCAACAACACCTCATTGATGGGAGGGACATGATAATTATTCGCATGCCTAAGGATTTCCC
ACCATTTCCTCAAAAGCTGAAATTTAGAGAGCCACAAAGGGAAGAACGCATATGTCTTGTTACA
ACCAACTTCCAAACTAAGAGCATGTCTAGCATGGTGTCAGACACTAGTTGCACATTCCCTTCAT
CTGATGGCATATTCTGGAAGCATTGGATTCAAACCAAGGATGGGCAGTGTGGCAGTCCATTAGT
ATCAACTAGAGATGGGTTCATTGTTGGTATACACTCAGCATCGAATTTCACCAACACAAACAAT
TATTTCACAAGCGTGCCGAAAAACTTCATGGAATTGTTGACAAATCAGGAGGCGCAGCAGTGGG
TTAGTGGTTGGCGATTAAATGCTGACTCAGTATTGTGGGGGGGCCATAAAGTTTTCATGCCGAA
ACCTGAAGAGCCTTTTCAGCCAGTTAAGGAAGCGACTCAACTCATGAATCGTCGTCGCCGTCGC
TAATAAGGACTGCAGGTAATTAAATAAGCTTCAAATAAAACGAAAGGCTCAGTCGAAAGACTGG
GCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAG
CGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGC
CAGGCATCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTT
TCGGTCCGTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCT
GATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCT
TATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTA
AAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTA
AGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTCCCAATGATGAGCACTTTTAAAGTTCTGCT
ATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTAT
TCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAG
TAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGAC
AACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGC
CTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGC
CTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCG
GCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTT

CCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTG
CAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGC
AACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAA
CTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTCCTTAGGACTGAGCGTCAACCCC
GTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAA
CAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCC
GAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTA
GGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAG
TGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGA
TAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACC
TACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAA
AGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGG
GGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTT
TTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGT
TCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGA
TAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGC
GAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCG
GTATTTCACACCGCATATAAGGTGCACTGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAA
CACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGAC
CGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTG
CGGTAAAGCTCATCAGCGTGGTCGTGCAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCA
GCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGC
GGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATG
ATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTAC
TGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCA
GGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCT
GCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAA
CACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCT
TCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGC
CGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGCCAGGACCCAACGCTGCCCGAAA
TT

7.1.3 pSAB2.1

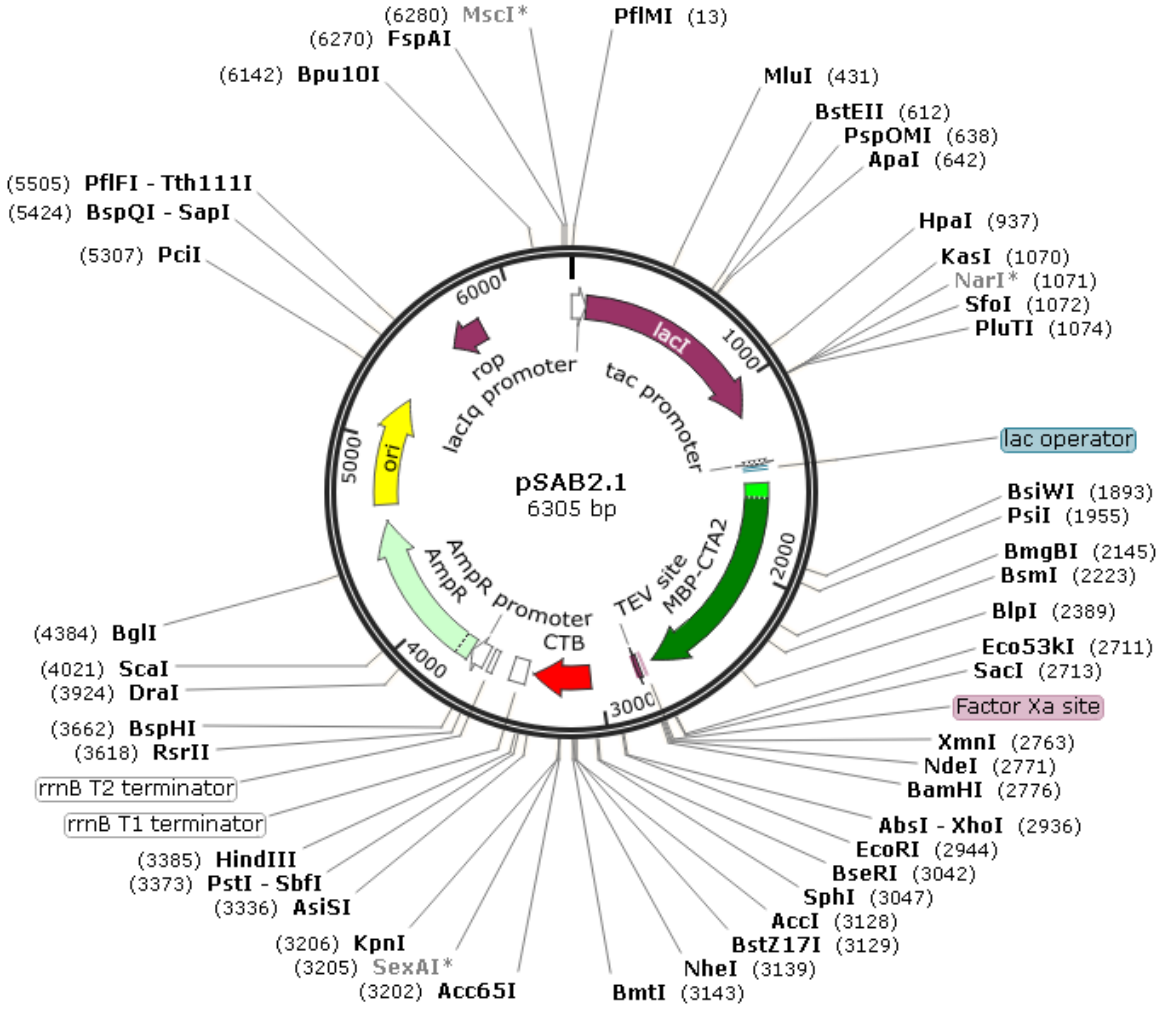


Figure 7.3 Plasmid map for pSAB2.1

The pSAB2.1 plasmid was used to express MBP-AB_5 in *E. coli* cells (Figure 7.3). The plasmid is derived from pMAL-p5x and was prepared by Dr James Ross (University of Leeds).[176*]

Full sequence 5'-3':

CCGACACCATCGAATGGTGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCA
ATTCAGGGTGGTGAATGTGAAACCAGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCT
TATCAGACCGTTTCCCGCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAG
TGGAAGCGGCGATGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAA
ACAGTCGTTGCTGATTGGCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTC

[*] The plasmid sequence in J.F. Ross, PhD thesis is incorrect. Correct sequence given here.

GCGGCGATTAAATCTCGCGCCGATCAACTGGGTGCCAGCGTGGTGGTGTCGATGGTAGAACGAA
GCGGCGTCGAAGCCTGTAAAGCGGCGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGAT
CATTAACTATCCGCTGGATGACCAGGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCG
GCGTTATTTCTTGATGTCTCTGACCAGACACCCATCAACAGTATTATTTTCTCCCATGAAGACG
GTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGCGGG
CCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAAT
CAAATTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCA
TGCAAATGCTGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAGATGGCGCT
GGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGATATTTCGGTAGTGGGA
TACGACGATACCGAAGACAGCTCATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTC
GCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGAAGGG
CAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAACC
GCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAA
GCGGGCAGTGAGCGCAACGCAATTAATGTAAGTTAGCTCACTCATTAGGCACAATTCTCATGTT
TGACAGCTTATCATCGACTGCACGGTGCACCAATGCTTCTGGCGTCAGGCAGCCATCGGAAGCT
GTGGTATGGCTGTGCAGGTCGTAAATCACTGCATAATTCGTGTCGCTCAAGGCGCACTCCCGTT
CTGGATAATGTTTTTTGCGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTG
ACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAA
CAGCCAGTCCGTTTAGGTGTTTTCACGAGCAATTGACCAACAAGGACCATAGATTATGAAAATA
AAAACAGGTGCACGCATCCTCGCATTATCCGCATTAACGACGATGATGTTTTCCGCCTCGGCTC
TCGCC**AAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGATAAAGGCTATAACGGTCT**
CGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTGAGCATCCGGAT
AAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTCTGGG
CACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGC
GTTCCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCT
TACCCGATCGCTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAA
AAACCTGGGAAGAGATCCCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGAT
GTTCAACCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTC
AAGTATGAAAACGGCAAGTACGACATTAAAGACGTGGGCGTGGATAACGCTGGCGCGAAAGCGG
GTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCGATTACTCCAT
CGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCC
AACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCAAGGGTCAACCAT
CCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAGCTGGC
AAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAA
CCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGTGAAAGATCCGCGTATTGCCG
CCACTATGGAAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTG
GTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTG
AAAGACGCGCAGACTAATTCGAGCTCGAACAACAACAACAATAACAATAACAACAACCTCGGGA
TCGAGGGAAGGATTTCACATATGGGATCCGAAAACCTGTACTTTCAGGGTGGCGGTGATGAAAA
AACCCAAAGTCATGGTGTAAAATTCCTTGACGAATACCAATCTAAAGTTAAAAGACAAATATTT
TCAGGCTATCAATCTGATATTGATACACATAATAGAATTAAGGATGAATTATGACCTCGAGGTG
AATTCACGAGCAATTGACCAACAAGGACCATAGATTATGAGCTTTAAGAAAATTATCAAGGCAT
TTGTTATCATGGCTGCTTTGGTATCTGTTCAGGCGCATGCA**GCTCCTCAAAATATTACTGATTT**
GTGCGCAGAATACCACAACACACAAATATATACGCTAAATGATAAGATCTTTTCGTATACAGAA
TCGCTAGCGGGAAAAAGAGAGATGGCTATCATTACTTTTAAGAATGGTGCAATTTTTCAAGTAG
AGGTACCAGGTAGTCAACATATAGATTCACAAAAAAAAGCGATTGAAAGGATGAAGGATACCCT
GAGGATTGCATATCTTACTGAAGCTAAAGTCGAAAAGTTATGTGTATGGAATAATAAAACGCCT
CATGCGATCGCCGCAATTAGTATGGCAAACTAAGTTTTCCCTGCAGGTAATTAAATAAGCTTCA
AATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAAC
GCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAG
GGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGAC
GGATGGCCTTTTTGCGTTTCTACAAACTCTTTCGGTCCGTTGTTTATTTTTCTAAATACATTCA
AATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGA
GTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGT
TTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTG
GGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTT

TCCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGG
GCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTC
ACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGA
GTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTT
TTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCC
ATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTAT
TAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAA
AGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGA
GCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTA
TCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGA
GATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAG
ATTGATTTCCTTAGGACTGAGCGTCAACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCT
TTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTT
TGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACC
AAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCT
ACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTA
CCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTC
GTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTA
TGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCG
GAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGG
GTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGG
AAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGT
TCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATAC
CGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTG
ATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATAAGGTGCACTGTGAC
TGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGC
TCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTC
ACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGCAGCGAT
TCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCT
GGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGT
GTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATA
CGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTAT
GGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGT
AGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGC
GCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTC
AGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTG
CTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGC
ACCCGTGGCCAGGACCCAACGCTGCCCGAAATT

7.2 Protein sequences

SrtA:

Protein sequence obtained from pET28a I plasmid

```
        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
```

TEV-MBP:

Protein sequence obtained from pMAL-c5x I plasmid

```
        10         20         30         40         50         60
MKIEEGKLVI WINGDKGYNG LAEVGKKFEK DTGIKVTVEH PDKLEEKFPQ VAATGDGPDI

        70         80         90        100        110        120
IFWAHDRFGG YAQSGLLAEI TPDKAFQDKL YPFTWDAVRY NGKLIAYPIA VEALSLIYNK

       130        140        150        160        170        180
DLLPNPPKTW EEIPALDKEL KAKGKSALMF NLQEPYFTWP LIAADGGYAF KYENGKYDIK

       190        200        210        220        230        240
DVGVDNAGAK AGLTFLVDLI KNKHMNADTD YSIAEAAFNK GETAMTINGP WAWSNIDTSK

       250        260        270        280        290        300
VNYGVTVLPT FKGQPSKPFV GVLSAGINAA SPNKELAKEF LENYLLTDEG LEAVNKDKPL

       310        320        330        340        350        360
GAVALKSYEE ELVKDPRIAA TMENAQKGEI MPNIPQMSAF WYAVRTAVIN AASGRQTVDE

       370        380        390        400        410        420
ALKDAQTNSS SNNNNNNNNN NLGIEGRISH MSMGGRDIVD GSGKSLFKGP RDYNPISSTI

       430        440        450        460        470        480
CHLTNESDGH TTSLYGIGFG PFIITNKHLF RRNNGTLLVQ SLHGVFKVKN TTTLQQHLID

       490        500        510        520        530        540
GRDMIIIRMP KDFPPFPQKL KFREPQREER ICLVTTNFQT KSMSSMVSDT SCTFPSSDGI

       550        560        570        580        590        600
FWKHWIQTKD GQCGSPLVST RDGFIVGIHS ASNFTNTNNY FTSVPKNFME LLTNQEAQQW

       610        620        630        640
VSGWRLNADS VLWGGHKVFM PKPEEPFQPV KEATQLMNRR RRR
```

CTB:

Protein sequence obtained from pSAB2.1 plasmid

```
        10         20         30         40         50         60
APQNITDLCA EYHNTQIYTL NDKIFSYTES LAGKREMAII TFKNGAIFQV EVPGSQHIDS

        70         80         90        100
QKKAIERMKD TLRIAYLTEA KVEKLCVWNN KTPHAIAAIS MAN
```

CTA2-MBP:

Protein sequence obtained from pSAB2.1 plasmid.The TEV protease cleavage site in the MBP-CTA2 sequence is marked in purple. The sequence for CTA2, highlighted in green, begins at G399 in the MBP-CTA2 sequence.

```
        10         20         30         40         50         60
KIEEGKLVIW INGDKGYNGL AEVGKKFEKD TGIKVTVEHP DKLEEKFPQV AATGDGPDII

        70         80         90        100        110        120
FWAHDRFGGY AQSGLLAEIT PDKAFQDKLY PFTWDAVRYN GKLIAYPIAV EALSLIYNKD

       130        140        150        160        170        180
LLPNPPKTWE EIPALDKELK AKGKSALMFN LQEPYFTWPL IAADGGYAFK YENGKYDIKD

       190        200        210        220        230        240
VGVDNAGAKA GLTFLVDLIK NKHMNADTDY SIAEAAFNKG ETAMTINGPW AWSNIDTSKV

       250        260        270        280        290        300
NYGVTVLPTF KGQPSKPFVG VLSAGINAAS PNKELAKEFL ENYLLTDEGL EAVNKDKPLG

       310        320        330        340        350        360
AVALKSYEEE LVKDPRIAAT MENAQKGEIM PNIPQMSAFW YAVRTAVINA ASGRQTVDEA

       370        380        390        400        410        420
LKDAQTNSSS NNNNNNNNNN LGIEGRISHM GSENLYFQGG GDEKTQSHGV KFLDEYQSKV

       430        440
KRQIFSGYQS DIDTHNRIKD EL
```

Mouse pumilio-2 puf domain:

Protein sequence obtained from UniProt code: Q80U58

```
        10         20         30         40         50         60
GTGRSRLLED FRNNRFPNLQ LRDLIGHIVE FSQDQHGSRF IQQKLERATP AERQIVFNEI

        70         80         90        100        110        120
LQAAYQLMTD VFGNYVIQKF FEFGSLDQKL ALATRIRGHV LPLALQMYGC RVIQKALESI

       130        140        150        160        170        180
SSDQQSEMVK ELDGHVLKCV KDQNGNHVVQ KCIECVQPQS LQFIIDAFKG QVFVLSTHPY

       190        200        210        220        230        240
GCRVIQRILE HCTAEQTLPI LEELHQHTEQ LVQDQYGNYV IQHVLEHGRP EDKSKIVSEI

       250        260        270        280        290        300
RGKVLALSQH KFASNVVEKC VTHASRAERA LLIDEVCCQN DGPHSALYTM MKDQYANYVV

       310        320        330        340        350
QKMIDMAEPA QRKIIMHKIR PHITTLRKYT YGKHILAKLE KYYLKNSPDL G
```

Human mannose binding lectin (galactose-binding variant):

Protein sequence obtained from Prof. K. Drickamer (Imperial College London).

```
        10         20         30         40         50         60
GDSSLAASER KALQTEMARI KKWLTFSLGK QVGNKFFLTN GEIMTFEKVK ALCVKFQASV

        70         80         90        100        110        120
ATPRNAAENG AIQNLIKEEA FLGITDEKTE GQFVDLTGNR LTYTNWNEGQ PDDWYGHGLG

       130        140        150        160        170        180
GGEDCVLLLK NGQWNDVPCS TSHLAVCEFP IPHITTLRKY TYGKHILAKL EKYYLKNSPD


LG
```

Equine heart myoglobin:

Protein sequence obtained from UniProt code: P68082

```
        10         20         30         40         50         60
GLSDGEWQQV LNVWGKVEAD IAGHGQEVLI RLFTGHPETL EKFDKFKHLK TEAEMKASED

        70         80         90        100        110        120
LKKHGTVVLT ALGGILKKKG HHEAELKPLA QSHATKHKIP IKYLEFISDA IIHVLHSKHP

       130        140        150
GDFGADAQGA MTKALELFRN DIAAKYKELG FQG
```

Chapter 8: References

(1) *Nat. Med.* **2012**, *18*, 636.

(2) Beck, A.; Wurch, T.; Bailly, C.; Corvaia, N. *Nat. Rev. Immunol.* **2010**, *10*, 345.

(3) Kole, R.; Krainer, A. R.; Altman, S. *Nat. Rev. Drug Discov.* **2012**, *11*, 125.

(4) Cohen, M. G.; Purdy, D. A.; Rossi, J. S.; Grinfeld, L. R.; Myles, S. K.; Aberle, L. H.; Greenbaum, A. B.; Fry, E.; Chan, M. Y.; Tonkens, R. M.; Zelenkofske, S.; Alexander, J. H.; Harrington, R. A.; Rusconi, C. P.; Becker, R. C. *Circulation* **2010**, *122*, 614.

(5) Goemans, N. M.; Tulinius, M.; van den Akker, J. T.; Burm, B. E.; Ekhart, P. F.; Heuvelmans, N.; Holling, T.; Janson, A. A.; Platenburg, G. J.; Sipkens, J. A.; Sitsen, J. M. A.; Aartsma-Rus, A.; van Ommen, G.-J. B.; Buyse, G.; Darin, N.; Verschuuren, J. J.; Campion, G. V.; de Kimpe, S. J.; van Deutekom, J. C. *New. Engl. J. Med.* **2011**, *364*, 1513.

(6) Cirak, S.; Arechavala-Gomeza, V.; Guglieri, M.; Feng, L.; Torelli, S.; Anthony, K.; Abbs, S.; Garralda, M. E.; Bourke, J.; Wells, D. J.; Dickson, G.; Wood, M. J. A.; Wilton, S. D.; Straub, V.; Kole, R.; Shrewsbury, S. B.; Sewry, C.; Morgan, J. E.; Bushby, K.; Muntoni, F. *Lancet* **2011**, *378*, 595.

(7) Eckstein, F. *Expert Opin. Biol. Ther.* **2007**, *7*, 1021.

(8) Burnett, John C.; Rossi, John J. *Chem. Biol.* **2012**, *19*, 60.

(9) Shukla, S.; Sumaria, C. S.; Pradeepkumar, P. I. *ChemMedChem* **2010**, *5*, 328.

(10) Debart, F.; Abes, S.; Deglane, G.; Moulton, H. M.; Clair, P.; Gait, M. J.; Vasseur, J.-J.; Lebleu, B. *Curr. Top. Med. Chem.* **2007**, *7*, 727.

(11) Ziello, J. E.; Huang, Y.; Jovin, I. S. *Mol. Med.* **2010**, *16*, 222.

(12) Thorley, J. A.; McKeating, J. A.; Rappoport, J. Z. *Protoplasma* **2010**, *244*, 15.

(13) Ibraheem, D.; Elaissari, A.; Fessi, H. *Int. J. Pharm.* **2014**, *459*, 70.

(14) Kay, M. A.; Glorioso, J. C.; Naldini, L. *Nat. Med.* **2001**, *7*, 33.

(15) Giacca, M.; Zacchigna, S. *J. Control. Release* **2012**, *161*, 377.

(16) Kohn, D. B.; Sadelain, M.; Glorioso, J. C. *Nat. Rev. Cancer.* **2003**, *3*, 477.

(17) Hacein-Bey-Abina, S.; Garrigue, A.; Wang, G. P.; Soulier, J.; Lim, A.; Morillon, E.; Clappier, E.; Caccavelli, L.; Delabesse, E.; Beldjord, K.; Asnafi, V.; MacIntyre, E.; Dal Cortivo, L.; Radford, I.; Brousse, N.; Sigaux, F.; Moshous, D.; Hauer, J.; Borkhardt, A.; Belohradsky, B. H.; Wintergerst, U.; Velez, M. C.; Leiva, L.; Sorensen, R.; Wulffraat, N.; Blanche, S.; Bushman, F. D.; Fischer, A.; Cavazzana-Calvo, M. *J. Clin. Invest.* **2008**, *118*, 3132.

(18) Fenske, D. B.; Chonn, A.; Cullis, P. R. *Toxicol. Path.* **2008**, *36*, 21.

(19) Torchilin, V. P. *Annu. Rev. Biomed. Eng.* **2006**, *8*, 343.

(20) Margolis, L. B.; Baibakov, B. A.; Neyfakh Jr, A. A.; Victorov, A. V.; Galkina, S. I.; Bergelson, L. D. *Biochim. Biophys. Acta.* **1984**, *804*, 23.

(21) Blumenthal, R.; Ralston, E.; Dragsten, P.; Leserman, L. D.; Weinstein, J. N. *Mol. Membr. Biol.* **1982**, *4*, 283.

(22) Lasic, D. D. *Trends Biotechnol.* **1998**, *16*, 307.

(23) Allen, T. M.; Newman, M. S.; Woodle, M. C.; Mayhew, E.; Uster, P. S. *Int. J. Cancer.* **1995**, *62*, 199.

(24) *Expert. Rev. Vaccines* **2010**, *9*, 1251.

(25) Torchilin, V. P. *Nat. Rev. Drug Discov.* **2005**, *4*, 145.

(26) Lian, T.; Ho, R. J. Y. *J. Pharmacol. Sci.* **2001**, *90*, 667.

(27) Senior, J. H. *Crit. Rev. Ther. Drug Carrier Syst.* **1987**, *3*, 123.

(28) Immordino, M. L.; Dosio, F.; Cattel, L. *Int. J. Nanomedicine* **2006**, *1*, 297.

(29) Tong, A. W.; Jay, C. M.; Senzer, N.; Maples, P. B.; Nemunaitis, J. *Curr. Gene. Ther.* **2009**, *9*, 45.

(30) Papahadjopoulos, D.; Allen, T. M.; Gabizon, A.; Mayhew, E.; Matthay, K.; Huang, S. K.; Lee, K. D.; Woodle, M. C.; Lasic, D. D.; Redemann, C. *Proc. Natl. Acad. Sci.USA* **1991**, *88*, 11460.

(31) Hoyer, J.; Neundorf, I. *Acc. Chem. Res.* **2012**, *45*, 1048.

(32) Saleh, A. F.; Aojula, H.; Arthanari, Y.; Offerman, S.; Alkotaji, M.; Pluen, A. *J. Control. Release* **2010**, *143*, 233.

(33) Min, S. H.; Kim, D. M.; Kim, M. N.; Ge, J.; Lee, D. C.; Park, I. Y.; Park, K. C.; Hwang, J. S.; Cho, C. W.; Yeom, Y. I. *Biomaterials* **2010**, *31*, 1858.

(34) Langel, K.; Lindberg, S.; Copolovici, D.; Arukuusk, P.; Sillard, R.; Langel, Ű. *Int. J. Pept. Res. Ther.* **2010**, *16*, 247.

(35) Ifediba, M. A.; Medarova, Z.; Ng, S.-w.; Yang, J.; Moore, A. *Bioconjug. Chem.* **2010**, *21*, 803.

(36) Schwarze, S. R.; Ho, A.; Vocero-Akbani, A.; Dowdy, S. F. *Science* **1999**, *285*, 1569.

(37) Andaloussi, S. E.; Lehto, T.; Lundin, P.; Langel, U. *Methods Mol. Biol.* **2011**, *683*, 361.

(38) Trabulo, S.; Cardoso, A. L.; Mano, M.; de Lima, M. C. P. *Pharmaceutical* **2010**, *3*, 961.

(39) Lundberg, P.; Langel, Ü. *J. Mol. Recognit.* **2003**, *16*, 227.

(40) Joliot, A.; Prochiantz, A. *Nat. Cell. Biol.* **2004**, *6*, 189.

(41) Heitz, F.; Morris, M. C.; Divita, G. *Br. J. Pharmacol.* **2009**, *157*, 195.

(42) Sarko, D.; Beijer, B.; Boy, R. G.; Nothelfer, E.-M.; Leotta, K.; Eisenhut, M.; Altmann, A.; Haberkorn, U.; Mier, W. *Mol. Pharmaceutics.* **2010**, *7*, 2224.

(43) Gentilucci, L.; De Marco, R.; Cerisoli, L. *Curr. Pharm. Des.* **2010**, *16*, 3185.

(44) de Martimprey, H.; Vauthier, C.; Malvy, C.; Couvreur, P. *Eur. J. Pharm. Biopharm.* **2009**, *71*, 490.

(45) Boussif, O.; Lezoualc'h, F.; Zanta, M. A.; Mergny, M. D.; Scherman, D.; Demeneix, B.; Behr, J. P. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 7297.

(46) Harada, A.; Togawa, H.; Kataoka, K. *Eur. J. Pharm. Sci.* **2001**, *13*, 35.

(47) Sato, A.; Choi, S. W.; Hirai, M.; Yamayoshi, A.; Moriyama, R.; Yamano, T.; Takagi, M.; Kano, A.; Shimamoto, A.; Maruyama, A. *J. Control. Release* **2007**, *122*, 209.

(48) Maksimenko, A.; Polard, V.; Villemeur, M.; Elhamess, H.; Couvreur, P.; Bertrand, J. R.; Aboubakar, M.; Gottikh, M.; Malvy, C. *Ann. N. Y. Acad. Sci.* **2005**, *1058*, 52.

(49) Lambert, G.; Fattal, E.; Pinto-Alphandary, H.; Gulik, A.; Couvreur, P. *Pharmaceut. Res.* **2000**, *17*, 707.

(50) Hillaireau, H.; Le Doan, T.; Chacun, H.; Janin, J.; Couvreur, P. *Int. J. Pharm* **2007**, *331*, 148.

(51) Bertholon, I.; Lesieur, S.; Labarre, D.; Besnard, M.; Vauthier, C. *Macromolecules* **2006**, *39*, 3559.

(52) Sandvig, K.; van Deurs, B. *Physiol. Rev.* **1996**, *76*, 949.

(53) Merritt, E. A.; Hol, W. G. J. *Curr. Opin. Struc. Biol.* **1995**, *5*, 165.

(54) Spangler, B. D. *Microbiol. Rev.* **1992**, *56*, 622.

(55) Lencer, W. I.; Tsai, B. *Trends. Biochem. Sc.* **2003**, *28*, 639.

(56) Zhang, R.-G.; Scott, D. L.; Westbrook, M. L.; Nance, S.; Spangler, B. D.; Shipley, G. G.; Westbrook, E. M. *J. Mol. Biol.* **1995**, *251*, 563.

(57) Fujinaga, Y. *J. Biochem.* **2006**, *140*, 155.

(58) Merritt, E. A.; Sarfaty, S.; Vandenakker, F.; Lhoir, C.; Martial, J. A.; Hol, W. G. J. *Protein. Sci.* **1994**, *3*, 166.

(59) Haan, L. d.; Hirst, T. R. *Mol. Membr. Biol.* **2004**, *21*, 77.

(60) Guimaraes, C. P.; Carette, J. E.; Varadarajan, M.; Antos, J.; Popp, M. W.; Spooner, E.; Brummelkamp, T. R.; Ploegh, H. L. *J. Cell Biol.* **2011**, *195*, 751.

(61) Turnbull, W. B.; Hayes, E. D. Monovalent and multivalent inhibitors of bacterial toxins, in "*Synthesis and Biological Applications of Multivalent Glycoconjugates*", Bentham Science Publishers, 2011, pp 78-91

(62) Orlandi, P. A. *J. Biol. Chem.* **1997**, *272*, 4591.

(63) Ampapathi, R. S.; Creath, A. L.; Lou, D. I.; Craft Jr, J. W.; Blanke, S. R.; Legge, G. B. *J. Mol. Biol.* **2008**, *377*, 748.

(64) Tinker, J. K.; Erbe, J. L.; Holmes, R. K. *Infect. Immun.* **2005**, *73*, 3627.

(65) Kohli, N.; Westerveld, D. R.; Ayache, A. C.; Verma, A.; Shil, P.; Prasad, T.; Zhu, P.; Chan, S. L.; Li, Q.; Daniell, H. *Mol. Ther.* **2014**, *22*, 535.

(66) Barrett, L. B.; Berry, M.; Ying, W. B.; Hodgkin, M. N.; Seymour, L. W.; Gonzalez, A. M.; Read, M. L.; Baird, A.; Logan, A. *J. Gene. Med.* **2004**, *6*, 429.

(67) Alisky, J. M.; van de Wetering, C. I.; Davidson, B. L. *Exp. Neurol.* **2002**, *178*, 139.

(68) Havton, L. A.; Broman, J. *J. Neurosci. Meth.* **2005**, *149*, 101.

(69) Hardy, S. J.; Holmgren, J.; Johansson, S.; Sanchez, J.; Hirst, T. R. *Proc. Natl. Acad. Sci USA* **1988**, *85*, 7109.

(70) Hirst, T. R.; Sanchez, J.; Kaper, J. B.; Hardy, S. J.; Holmgren, J. *Proc. Natl. Acad. Sci USA* **1984**, *81*, 7752.

(71) Streatfield, S. J.; Sandkvist, M.; Sixma, T. K.; Bagdasarian, M.; Hol, W. G.; Hirst, T. R. *Proc. Natl. Acad. Sci USA* **1992**, *89*, 12140.

(72) Jobling, M. G.; Holmes, R. K. *Infect. Immun.* **1992**, *60*, 4915.

(73) Takeda, Y.; Honda, T.; Taga, S.; Miwatani, T. *Infect. Immun.* **1981**, *34*, 341.

(74) Hatic, S. O.; McCann, J. A.; Picking, W. D. *Anal. Biochem.* **2001**, *292*, 171.

(75) Tinker, J. K.; Erbe, J. L.; Hol, W. G. J.; Holmes, R. K. *Infect. Immun.* **2003**, *71*, 4093.

(76) Mekalanos, J. J.; Swartz, D. J.; Pearson, G. D. N.; Harford, N.; Groyne, F.; de Wilde, M. *Nature* **1983**, *306*, 551.

(77) Ruddock, L. W.; Coen, J. J. F.; Cheesman, C.; Freedman, R. B.; Hirst, T. R. *J. Biol. Chem.* **1996**, *271*, 19118.

(78) Lesieur, C.; Cliff, M. J.; Carter, R.; James, R. F. L.; Clarke, A. R.; Hirst, T. R. *J. Biol. Chem.* **2002**, *277*, 16697.

(79) Dertzbaugh, M. T.; Peterson, D. L.; Macrina, F. L. *Infect. Immun.* **1990**, *58*, 70.

(80) Dertzbaugh, M. T.; Elson, C. O. *Infect. Immun.* **1993**, *61*, 384.

(81) Chen, J.; Zeng, W.; Offord, R.; Rose, K. *Bioconjug. Chem.* **2003**, *14*, 614.

(82) Rose, K.; Chen, J.; Dragovic, M.; Zeng, W.; Jeannerat, D.; Kamalaprija, P.; Burger, U. *Bioconjug. Chem.* **1999**, *10*, 1038.

(83) Sandkvist, M.; Hirst, T. R.; Bagdasarian, M. *J. Bacteriol.* **1987**, *169*, 4570.

(84) Green, E. A.; Botting, C.; Webb, H. M.; Hirst, T. R.; Randall, R. E. *Vaccine* **1996**, *14*, 949.

(85) Loregian, A.; Hirst, T. R.; Marsden, H. S.; Palù, G. *Protein Expres. Purif.* **1996**, *8*, 381.

(86) O'Dowd, A. M.; Botting, C. H.; Precious, B.; Shawcross, R.; Randall, R. E. *Vaccine* **1999**, *17*, 1442.

(87) van Berkel, S. S.; van Eldijk, M. B.; van Hest, J. C. M. *Angew. Chem. Int. Ed.* **2011**, *50*, 8806.

(88) Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2004**, *126*, 15046.

(89) de Graaf, A. J.; Kooijman, M.; Hennink, W. E.; Mastrobattista, E. *Bioconjug. Chem.* **2009**, *20*, 1281.

(90) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 776.

(91) Dawson, P. E.; Kent, S. B. H. *Annu. Rev. Biochem.* **2000**, *69*, 923.

(92) Kent, S. B. H. *Chem. Soc. Rev.* **2009**, *38*, 338.

(93) Johnson, E. C. B.; Kent, S. B. H. *J. Ame. Chem. Soc.* **2006**, *128*, 6640.

(94) Hackeng, T. M.; Mounier, C. M.; Bon, C.; Dawson, P. E.; Griffin, J. H.; Kent, S. B. H. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 7845.

(95) Schnölzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. H. *Int. J. Pept. Res. Ther.* **2007**, *13*, 31.

(96) Blanco-Canosa, J. B.; Dawson, P. E. *Angew. Chem. Int. Ed.* **2008**, *47*, 6851.

(97) Chong, S.; Mersha, F. B.; Comb, D. G.; Scott, M. E.; Landry, D.; Vence, L. M.; Perler, F. B.; Benner, J.; Kucera, R. B.; Hirvonen, C. A.; Pelletier, J. J.; Paulus, H.; Xu, M.-Q. *Gene* **1997**, *192*, 271.

(98) Muir, T. W.; Sondhi, D.; Cole, P. A. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705.

(99) Bang, D.; Kent, S. B. *Angew. Chem. Int. Ed.* **2004**, *43*, 2534.

(100) Durek, T.; Torbeev, V. Y.; Kent, S. B. H. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 4846.

(101) Izumi, M.; Murakami, M.; Okamoto, R.; Kajihara, Y. *J. Pept. Sci.* **2014**, *20*, 98.

(102) Diezmann, F.; Eberhard, H.; Seitz, O. *Pep. Sci.* **2010**, *94*, 397.

(103) Tsukiji, S.; Nagamune, T. *ChemBioChem* **2009**, *10*, 787.

(104) Paterson, G. K.; Mitchell, T. J. *Trends Microbiol.* **2004**, *12*, 89.

(105) Mazmanian, S. K.; Liu, G.; Ton-That, H.; Schneewind, O. *Science* **1999**, *285*, 760.

(106) Frankel, B. A.; Tong, Y.; Bentley, M. L.; Fitzgerald, M. C.; McCafferty, D. G. *Biochemistry* **2007**, *46*, 7269.

(107) Proft, T. *Biotechnol. Lett.* **2010**, *32*, 1.

(108) Zong, Y. N.; Bice, T. W.; Ton-That, H.; Schneewind, O.; Narayana, S. V. L. *J. Biol. Chem.* **2004**, *279*, 31383.

(109) Clancy, K. W.; Melvin, J. A.; McCafferty, D. G. *Biopolymers **2010***, *94*, 385.

(110) Frankel, B. A.; Kruger, R. G.; Robinson, D. E.; Kelleher, N. L.; McCafferty, D. G. *Biochemistry* **2005**, *44*, 11188.

(111) Popp, M. W.; Ploegh, H. L. *Angew. Chem. Int. Ed.* **2011**, *50*, 5024.

(112) Ton-That, H.; Mazmanian, S. K.; Alksne, L.; Schneewind, O. *J. Biol. Chem.* **2002**, *277*, 7447.

(113) Ilangovan, U.; Ton-That, H.; Iwahara, J.; Schneewind, O.; Clubb, R. T. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 6056.

(114) Kruger, R. G.; Otvos, B.; Frankel, B. A.; Bentley, M.; Dostal, P.; McCafferty, D. G. *Biochemistry* **2004**, *43*, 1541.

(115) Pritz, S.; Wolf, Y.; Kraetke, O.; Klose, J.; Bienert, M.; Beyermann, M. *J. Org. Chem.* **2007**, *72*, 3909.

(116) Huang, X.; Aulabaugh, A.; Ding, W.; Kapoor, B.; Alksne, L.; Tabei, K.; Ellestad, G. *Biochemistry* **2003**, *42*, 11307.

(117) Mao, H. Y.; Hart, S. A.; Schink, A.; Pollok, B. A. *J. Am. Chem. Soc* **2004**, *126*, 2670.

(118) Mohlmann, S.; Mahlert, C.; Greven, S.; Scholz, P.; Harrenga, A. *ChemBioChem* **2011**, *12*, 1774.

(119) Ton-That, H.; Liu, G.; Mazmanian, S. K.; Faull, K. F.; Schneewind, O. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 12424.

(120) Popp, M. W.; Antos, J. M.; Grotenbreg, G. M.; Spooner, E.; Ploegh, H. L. *Nat. Chem. Biol.* **2007**, *3*, 707.

(121) Antos, J. M.; Miller, G. M.; Grotenbreg, G. M.; Ploegh, H. L. *J. Am. Chem. Soc* **2008**, *130*, 16338.

(122) Wu, Z.; Guo, X.; Wang, Q.; Swarts, B. M.; Guo, Z. *J. Am. Chem. Soc* **2010**, *132*, 1567.

(123) Nelson, J. W.; Chamessian, A. G.; McEnaney, P. J.; Murelli, R. P.; Kazmiercak, B. I.; Spiegel, D. A. *ACS Chem. Biol.* **2010**, *5*, 1147.

(124) Antos, J. M.; Chew, G.-L.; Guimaraes, C. P.; Yoder, N. C.; Grotenbreg, G. M.; Popp, M. W.-L.; Ploegh, H. L. *J. Am. Chem. Soc.* **2009**, *131*, 10800.

(125) Fujinaga, Y.; Wolf, A. A.; Rodighiero, C.; Wheeler, H.; Tsai, B.; Allen, L.; Jobling, M. G.; Rapoport, T.; Holmes, R. K.; Lencer, W. I. *Mol. Biol. Cell* **2003**, *14*, 4783.

(126) Guimaraes, C. P.; Witte, M. D.; Theile, C. S.; Bozkurt, G.; Kundrat, L.; Blom, A. E.; Ploegh, H. L, *Nat. Protoc.* **2013**, *8*, 1787.

(127) Theile, C. S.; Witte, M. D.; Blom, A. E.; Kundrat, L.; Ploegh, H. L.; Guimaraes, C. P. *Nat. Protoc.* **2013**, *8*, 1800.

(128) Guo, X.; Wang, Q.; Swarts, B. M.; Guo, Z. *J. Am. Chem. Soc.* **2009**, *131*, 9878.

(129) Bischofberger, N.; Waldmann, H.; Saito, T.; Simon, E. S.; Lees, W.; Bednarski, M. D.; Whitesides, G. M. *J. Org.Chem.* **1988**, *53*, 3457.

(130) Eller, S.; Collot, M.; Yin, J.; Hahm, H. S.; Seeberger, P. H. *Angew. Chem. Int. Ed.* **2013**, *52*, 5858.

(131) Krock, L.; Esposito, D.; Castagner, B.; Wang, C.-C.; Bindschadler, P.; Seeberger, P. H. *Chem. Sci.* **2012**, *3*, 1617.

(132) Routenberg Love, K.; Seeberger, P. H. *Angew. Chem. Int. Ed.* **2004**, *43*, 602.

(133) Suich, D. J.; Ballinger, M. D.; Wells, J. A.; DeGrado, W. F. *Tetrahedron Lett.* **1996**, *37*, 6653.

(134) Turner, M. W. *Immun. Today* **1996**, *17*, 532.

(135) Drickamer, K.; Taylor, M. E. *Annu. Rev. Cell Biol.* **1993**, *9*, 237.

(136) Bouwman, L. H.; Roep, B. O.; Roos, A. *Hum. Immun.* **2006**, *67*, 247.

(137) Jenkins, H. T.; Baker-Wilding, R.; Edwards, T. A. *J. Struc. Biol.* **2009**, *167*, 271.

(138) Edwards, T. A.; Wilkinson, B. D.; Wharton, R. P.; Aggarwal, A. K. *Genes Dev.* **2003**, *17*, 2508.

(139) Sonoda, J.; Wharton, R. P. *Gene. Dev.* **1999**, *13*, 2704.

(140) Struhl, G.; Johnston, P.; Lawrence, P. A. *Cell* **1992**, *69*, 237.

(141) Gros, G.; Wittenberg, B. A.; Jue, T. *J. Exp. Biol.* **2010**, *213*, 2713.

(142) Ordway, G. A.; Garry, D. J. *J. Exp. Biol.* **2004**, *207*, 3441.

(143) Ross, J. *Unpublished work*

(144) Gey, G. O.; Coffman, W. D.; Kubicek, M. T. *Cancer Res.* **1952**, *12*, 264.

(145) Torgersen, M. L.; Skretting, G.; van Deurs, B.; Sandvig, K. *J. Cell Sci.* **2001**, *114*, 3737.

(146) Fujitani, N.; Furukawa, J.-i.; Araki, K.; Fujioka, T.; Takegawa, Y.; Piao, J.; Nishioka, T.; Tamura, T.; Nikaido, T.; Ito, M.; Nakamura, Y.; Shinohara, Y. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 2105.

(147) Majoul, I.; Schmidt, T.; Pomasanova, M.; Boutkevich, E.; Kozlov, Y.; Soling, H. D. *J. Cell Sci.* **2002**, *115*, 817.

(148) Lang, K.; Chin, J. W. *ACS Chem. Biol.* **2014**, *9*, 16.

(149) Dirksen, A.; Yegneswaran, S.; Dawson, P. E. *Angew. Chem. Int. Ed.* **2010**, *49*, 2023.

(150) Park, K. D.; Liu, R.; Kohn, H. *Chem. Biol.* **2009**, *16*, 763.

(151) Dirksen, A.; Dawson, P. E. *Bioconjug. Chem.* **2008**, *19*, 2543.

(152) Sander, E. G.; Jencks, W. P. *J. Am. Chem. Soc.* **1968**, *90*, 6154.

(153) Kattah, M. G.; Coller, J.; Cheung, R. K.; Oshidary, N.; Utz, P. J. *Nat. Med.* **2008**, *14*, 1284.

(154) Walls, Z. F.; Gambhir, S. S. *Bioconjug. Chem.* **2007**, *19*, 178.

(155) Kalia, J.; Raines, R. T. *Angew. Chem. Int. Ed.* **2008**, *120*, 7633.

(156) Grotzky, A.; Nauser, T.; Erdogan, H.; Schlüter, A. D.; Walde, P. *J. Am. Chem. Soc.* **2012**, *134*, 11392.

(157) Tornøe, C. W.; Christensen, C.; Meldal, M. *J. Org. Chem.* **2002**, *67*, 3057.

(158) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2002**, *41*, 2596.

(159) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2001**, *40*, 2004.

(160) Baskin, J. M.; Prescher, J. A.; Laughlin, S. T.; Agard, N. J.; Chang, P. V.; Miller, I. A.; Lo, A.; Codelli, J. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 16793.

(161) Witte, M. D.; Theile, C. S.; Wu, T.; Guimaraes, C. P.; Blom, A. E.; Ploegh, H. L. *Nat. Protoc.* **2013**, *8*, 1808.

(162) Debets, M. F.; van Berkel, S. S.; Dommerholt, J.; Dirks, A. J.; Rutjes, F. P. J. T.; van Delft, F. L. *Acc. Chem. Res.* **2011**, *44*, 805.

(163) Debets, M. F.; van Berkel, S. S.; Schoffelen, S.; Rutjes, F. P. J. T.; van Hest, J. C. M.; van Delft, F. L. *Chem. Commun.* **2010**, *46*, 97.

(164) Dommerholt, J.; Schmidt, S.; Temming, R.; Hendriks, L. J. A.; Rutjes, F. P. J. T.; van Hest, J. C. M.; Lefeber, D. J.; Friedl, P.; van Delft, F. L. *Angew. Chem. Int. Ed.* **2010**, *49*, 9422.

(165) Sazani, P.; Gemignani, F.; Kang, S. H.; Maier, M. A.; Manoharan, M.; Persmark, M.; Bortner, D.; Kole, R. *Nat. Biotechnol.* **2002**, *20*, 1228.

(166) Sun, Q.; Cai, S.; Peterson, B. R. *J. Ame. Chem. Soc.* **2008**, *130*, 10064.

(167) Schmitz, A.; Herrgen, H.; Winkeler, A.; Herzog, V. *J. Cell Biol.* **2000**, *148*, 1203.

(168) Wiertz, E. J. H. J.; Tortorella, D.; Bogyo, M.; Yu, J.; Mothes, W.; Jones, T. R.; Rapoport, T. A.; Ploegh, H. L. *Nature* **1996**, *384*, 432.

(169) Lycke, N. *Res. Immunol.* **1997**, *148*, 504.

(170) Marinaro, M.; Staats, H. F.; Hiroi, T.; Jackson, R. J.; Coste, M.; Boyaka, P. N.; Okahashi, N.; Yamamoto, M.; Kiyono, H.; Bluethmann, H.; Fujihashi, K.; McGhee, J. R. *J. Immunol.* **1995**, *155*, 4621.

(171) Mok, H.; Palmer, D. J.; Ng, P.; Barry, M. A. *Mol. Ther.* **2005**, *11*, 66.

(172) Croyle, M. A.; Chirmule, N.; Zhang, Y.; Wilson, J. M. *J. Virol.* **2001**, *75*, 4792.

(173) Macauley, M. S.; Pfrengle, F.; Rademacher, C.; Nycholat, C. M.; Gale, A. J.; von Drygalski, A.; 2Paulson, J. C. *J. Clin. Invest.* **2013**, *123*, 3074.

(174) Markley, J. L.; Bax, A.; Arata, Y.; Hilbers, C. W.; Kaptein, R.; Sykes, B. D.; Wright, P. E.; Wüthrich, K. *Pure Appl. Chem.* **1998**, *70*, 117.

(175) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. *Nucleic. Acids Res.* **1994**, *22*, 4673.

(176) Ross, J. F. PhD thesis, University of Leeds, 2013.

(177) Hopkins, A. L., Groom, C. R. *Nat. Rev.* **2002**, *1*, 727.

(178) O'Neal C. J., Amaya, E. I., Jobling, M. G., Holmes, R. K., Hol, W. G. J., *Biochemistry* **2004**, 43, 3773.

(179) Merritt, E. A., Kuhn, P., Sarfaty, S., Erbe, J. E., Holmes, R. K., Hol, W. G. J., *J. Mol. Biol*. **1998**, 282, 1043.

(180) Sheriff, S., Chang, C. Y., Ezekowitz, R. A., *Nat. Struct. Biol*. **1996**, 1, 103.

(181) Chu. K., Vojtchovsky, J., McMahon, B. H., Sweetk, R. M., Berendzen, J., Schlichting, I., Nature **2000**, 403, 921.