

**Site-specific methods for preparation of
glycoconjugates**

Jonathan Peter Dolan

**Submitted in accordance with the requirements for the degree of
Doctor of Philosophy**

The University of Leeds

School of Chemistry

December 2021

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

Chapter 2: The expression constructs & purification methodology for Sortase 7M-His₆, CTB-LPETGA and CTB-His₆ were provided by Dr Zoe Arnott, Dr Tomasz Kaminski, Dr Darren Machin and Dr Gemma Wildsmith.

Chapter 3: Phage display was performed by Dr Anna Tang & Dr Christian Tiede, BioScreening Technology group (BSTG, University of Leeds). I would further like to acknowledge Dr Christian Tiede and Dr Darren Tomlinson for their support and advice in designing the methodology used to screen the targets. The construct for the chitin binding variant of Sortase 7M (CBD-Srt7M) was designed by Dr Michael Webb.

Chapter 4: Azido globotriose (Gb3-N₃) and globotriose oligosaccharide (Gb3os) were prepared enzymatically by Laia Saltor Nunez. The expression construct for CTB(W88E) was prepared by Dr Tom Branson. CTB-HRP and VTB-HRP used in the ELLA and CTB(W88E)-Gb3 inhibitor were prepared by Dr Ryan McBerney. Cyclic Gb3 glycopeptides were synthesised by Dr Vajinder Kumar.

Chapter 5: The construct for Serum Amyloid P component was designed by Prof. Bruce Turnbull.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The right of Jonathan Peter Dolan to be identified as Author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

© 2021 The University of Leeds and Jonathan Peter Dolan

In loving memory of Joan and Zygmunt Ashley

You left fingerprints of grace on our lives. You will never be forgotten.

Acknowledgements

Firstly I must begin with a huge thank you to my family in particular my parents, without their never-ending support and guidance throughout my education and life I would never have had the confidence or drive to pursue a PhD.

To Kieron, Charity and Michelle, despite us all being terrible at keeping contact (mainly me), thank you for without a moment's hesitation opening up your homes whenever I needed an escape or time away. Our friendship and ability to pick up where we left off months before has been invaluable in moments of need.

To the many changing faces of Lab1.49, thank you for making the lab the fun and joyous place it was to work and be a part of. Starting at Leeds would never not have been as easy if it wasn't for Jack C, Sarah and Ryan who from day one at the Astbury Retreat ensured the lab was welcoming to everyone. To Jack C I must thank for his tireless organisation of lab socials, none of the memories would be complete if not for all the Friday pints, weekend lab outings, the annual Dragonboat competition or that one weekend in Amsterdam. From the late nights in the lab with a large cafetiere with Ryan, Jack W and Pablo, to the lonely early mornings driving from Lincolnshire; you all made it worth it!

I must thank Zoe, Gemma and Kristian for their patience and help in teaching a chemist with minimal biochemistry knowledge all things biochemistry. To Ryan and Kristian, despite most conversations about carbohydrates finishing "I don't know if that was particularly useful", just know they were always useful.

To everyone at Icen Glycoscience (formerly Icen Diagnostics) and the John Innes Centre thank you for welcoming me for three months. Although nothing really worked as intended, the experience and time spent away from Leeds was invaluable in showing me an alternative side to research. Finally, a massive thank you to Prof. Bruce Turnbull, Dr Mike Webb and Prof. Robert Field for their support guidance and patience. Your open doors, honest opinions and constant enthusiasm to help I could not be more thankful for.

Abstract

Synthetic natural glycoproteins and neoglycoconjugates have multiple applications ranging from the study of glycobiology to potential therapeutics and detection of pathogens. However, the preparation of natural homogeneous glycoproteins or synthetic neoglycoconjugates is a challenging synthetic task. The aim of the research described in this thesis is to develop site-specific chemical and enzymatic methods for the preparation of well-defined glycoconjugates. Firstly, the use of sortase-mediated C-terminal ligation was investigated as a method for the preparation of a series of CTB-MUC1 glycoconjugates with the aim to identify antibody-like binding proteins with affinity for the ligated glycopeptide.

Secondly, chemical ligation methods have been used to make multivalent neoglycoconjugate inhibitors of the bacterial enterotoxins, cholera toxin and verotoxin. Site-specific ligation of GM1 or Gb3 to a pentameric protein scaffold using a minimal length, bifunctional, and biorthogonal linker, resulted in inhibitors which display sub-nanomolar and micromolar IC_{50} values against CTB or VTB respectively, determined by ELLA inhibition studies.

Abbreviations

AEEAc	8-amino-3,6-dioxaoctanoic acid
Aha	Azidohomoalanine
AIM	Auto induction media
ALBA	Affimer lectin binding assay
APCs	Antigen presenting cells
BCN	Bicyclo[6.1.0]non-4-yne
BSTG	Biostructure Technology Group, University of Leeds
CAYE	Casamino yeast extract
CBD	Chitin binding domain
COSMC	core-1 β 3GalT specific molecular chaperone
CRP	C-Reactive protein
CT	Cholera toxin
CTB	Cholera toxin B-subunit
CuAAC	Copper catalysed azide-alkyne cycloaddition
DBU	1,8-Diazabicyclo(5.4.0)undec-7-ene
DCM	Dichloromethane
Dha	Dehydroalanine
DIAD	Diisoproylazodicarboxylate
DLS	Dynamic light scattering
DMC	2-Chloro-1,3-dimethylimidazolium chloride
DMF	Dimethylformamide
<i>E. coli</i>	<i>Escherichia coli</i>
ELISA	Enzyme-linked immunosorbent assay
ELLA	Enzyme-linked lectin assay
ER	Endoplasmic reticulum
ES-MS	Electrospray mass spectrometry
FDA	U.S. Food and Drug Administration
Fmoc	Fluorenylmethyloxycarbonyl
FPLC	Fast protein liquid chromatography
Gal	Galactose
GalNAc	N-Acetyl galactosamine
Gb3	Gal(α 1-4)Gal(β 1-4)Glc β
Gb3Cer	Gal(α 1-4)Gal(β 1-4)Glc β -ceramide

GM1	Gal(β 1-3)GalNAc(β 1-4)[Neu5Ac(α 2-3)]Gal(β 1-4)Glc β
GPI	Glycosylphosphatidylinositol
HIV	Human immunodeficiency virus
HMDS	Hexamethyldisilazane
HPLC	High-performance liquid chromatography
HRMS	High resolution mass spectrometry
HRP	Horseradish peroxidase
HUS	Haemolytic uremic syndrome
IPTG	Isopropyl- β -thiogalactopyranoside
ITC	Isothermal titration calorimetry
KLH	Keyhole limpet hemocyanin
Lac	Lactose
LB	Luria Broth
LC-MS	Liquid chromatography mass spectrometry
LT	Heat-labile toxin
mAb	Monoclonal antibody
MALDI-TOF	Matrix-assisted laser desorption/ionization time-of-flight spectrometry
MUC1	Mucin 1
NHS	N-Hydroxy succinimide
Ni-NTA	Nickel nitrilotriacetic acid resin
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PhoA	Alkaline phosphatase
SAP	Serum amyloid P component
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
SPAAC	Strain promoted azide-alkyne cycloaddition
SPANC	Strain promoted azide-nitrone cycloaddition
SPPS	Solid phase peptide synthesis
SPR	Surface plasmon resonance

Srt5M	Sortase A pentamutant (P94R, D160N, D165A, K190E and K196T)
Srt7M	Sortase A heptamutant (P94R, D160N, D165A, K190E and K196T, E105K and E108A/Q)
SrtA	Sortase A
TACAs	Tumour associated carbohydrate antigen
TAE	Tris, Acetic acid, EDTA
TCEP	Tris(2-carboxyethyl)phosphine
TEA	Triethylamine
TEMED	Tetramethylethylenediamine
TFA	Trifluoro acetic acid
THF	Tetrahydrofuran
TIS	Triisopropylsilane
TLC	Thin layer chromatography
TMS	Trimethylsilyl
TMSOTf	Trimethylsilyl trifluoromethanesulfonate
Tris	Tris(hydroxymethyl)aminomethane
VNTR	Variable number of tandem repeats
VT	Verotoxin
VTB	Verotoxin B-subunit
WHO	World Health Organisation
WT	Wild type

Table of Contents

Chapter 1: Background	1
1.1 The Glycocalyx	1
1.2 Tumour Associated Carbohydrate Antigens	2
1.2.1 Mucins	4
1.2.2 Structure & Function of Mucin 1	4
1.2.3 Mucin <i>O</i> -linked Glycosylation	6
1.2.4 Glycoconjugate Vaccines	8
1.3 Glycocalyx-Pathogen Interactions	9
1.3.1 Glycolipid Biosynthesis.....	10
1.3.2 Bacterial Toxins.....	10
1.3.3 Multivalent Glycoconjugate Inhibitors.....	13
1.4 Synthetic Glycoproteins & Glycoconjugates.....	18
1.4.1 Unnatural Amino Acid Incorporation.....	19
1.4.2 Post-translational Glycosylation.....	21
1.4.3 Modification of N-Terminal Cysteines.....	25
1.4.4 N-terminal Modification of Threonine & Serine.....	27
1.4.5 Transpeptidases	29
1.5 Project Outline	34
1.5.1 Detection of Novel Carbohydrate Binding Proteins.....	34
1.5.2 Multivalent Neoglycoconjugate Inhibitors.....	36
Chapter 2: Preparation of tumour-associated MUC1 cholera toxin B-subunit conjugates	38
2.1 Glycopeptide Preparation.....	38
2.1.1 Synthesis of GalNAc-Ser/Thr Amino Acid Building Blocks.....	38
2.1.2 Solid Phase Glycopeptide Synthesis of (Tn)MUC1 Fragments	41
2.2 Preparation of Protein Labelling Stocks	44
2.2.1 Preparation of CTB with C-terminal Sortase Sequence	44
2.2.2 Preparation of Sortase 7M.....	49
2.3 Conjugation of Glycopeptides to C-terminus of CTB	50
2.3.1 Effect of Adding Chaotropic Agent on Conjugational Efficiency ..	52
2.3.2 Investigation into Secondary Structure of MUC1 Epitopes	54
2.3.3 Effect of Increased Linker Length on Conjugational Efficiency	57
2.4 Is the Refolded CTB better for C-terminal Ligation with Sortase 7M?.....	61

2.5	Chapter Conclusions	63
Chapter 3: Biotinylation of Multivalent Proteins for Phage Display		64
3.1	Background	64
3.2	Dual N- and C-terminal Labelling of CTB-LPETGA	65
3.2.1	N-terminal Biotinylation.....	65
3.2.2	C-terminal Sortase Ligation of Biotinylated CTB-LPETGA	71
3.2.3	Phage Display Screening of Biotinylated CTB-glycoconjugates	76
3.3	Multivalent Biotinylated (Tn)MUC1 Glycopeptides for Phage Display ...	76
3.3.1	Preparation of Biotinylated Scaffold	76
3.4	Identifying VTB-binding Affimers	80
3.4.1	VTB Expression and Purification.....	81
3.4.2	VTB Biotinylation	82
3.4.3	Phage Display Screening of Verotoxin B-subunit.....	85
3.5	Chapter Conclusion.....	86
Chapter 4: Multivalent Glycoconjugate Inhibitors Based on Non-binding CTB Scaffold.		88
4.1	Background	88
4.2	Preparation of a BCN-functionalised CTB Scaffold	90
4.2.1	Expression of Non-binding CTB(W88E) Mutant	90
4.2.2	Synthesis of Aminoxy BCN-functionalised Linker.....	91
4.2.3	Site-specific BCN-functionalisation of CTB(W88E) by Oxime Ligation.....	92
4.3	Preparation of Glycosyl Azides and Functionalization of BCN-CTB(W88E) Through SPAAC	96
4.3.1	Synthesis of Lactosyl azide	96
4.3.2	Synthesis of GM1 azide.....	97
4.3.3	SPAAC of Glycosyl azides onto BCN-CTB(W88E)	100
4.4	Inhibition Studies by Enzyme-Linked Lectin Assay (ELLA)	103
4.4.1	Inhibition of CTB Binding to GM1	104
4.4.2	Inhibition of VTB Binding to Gb3	108
4.5	Chapter Conclusions	115
Chapter 5: Moving Toward Non-bacterial Pentameric Protein Scaffolds for Multivalent Neoglycoprotein Inhibitors		118
5.1	Background	118
5.2	Design and Expression of a Serum Amyloid P Scaffold	119
5.2.1	The Issues of Protein Glycosylation	121

5.3	C-Reactive Protein	122
5.3.1	Expression Testing & Plasmid Optimisation.....	123
5.3.2	Refolding C-Reactive Protein.....	125
5.4	Chapter Conclusions	126
Chapter 6: Conclusions & Future Work		128
6.1	C-terminal Sortase Mediated Ligation of Glycopeptides	128
6.2	Affimer Phage Display	129
6.2.1	Orthogonal N- & C-terminal Labelling of CTB	129
6.2.2	Multivalent MUC1(Tn) Glycopeptides for Phage Display	131
6.2.3	Working Towards developing Affimers against Verotoxin	131
6.3	Neoglycoconjugate Inhibitors of Bacterial Enterotoxins.....	133
6.3.1	Engineering a Non-bacterial Pentameric Protein Scaffold.....	135
Chapter 7: Chemistry Experimental.....		138
7.1	Synthesis	138
7.1.1	General Reagents and Equipment.....	138
7.1.2	Small Molecule Synthesis	139
7.2	Solid Phase Peptide Synthesis	153
7.2.1	General Reagents and Equipment.....	153
7.2.2	Standard Procedures for Solid Phase Peptide Synthesis	153
7.2.3	Synthesised Peptides & Glycopeptides	155
Chapter 8: Biochemical Experimental.....		170
8.1	General Methods and Equipment.....	170
8.2	Media and Buffers.....	170
8.2.1	Growth Media.....	171
8.2.2	General Buffers for Working with Proteins	171
8.2.3	General Buffers for Working with DNA.....	173
8.3	Standard Protocols	173
8.3.1	Cell Growth and Protein Overexpression	173
8.3.2	Protein Purification.....	175
8.3.3	Protein Analysis.....	180
8.3.4	Protein Modification.....	182
8.3.5	DNA Manipulation, Purification and Analysis	184
8.3.6	ELLA protocols	187

Chapter 9: References.....	189
Chapter 10: Appendix	203
10.1 Peptide & Glycopeptide HPLC Traces	203
10.2 Plasmid Sequences	213
10.3 Protein Sequences	228
10.4 Protein Purification	232
10.5 Primers	239

vaccines. Section 1.3 will introduce glycoalyx-pathogen interactions, focussing on interactions of bacterial toxins and the development of multivalent inhibitors of toxin adhesion. In section 1.4 methods for the site-specific modification of proteins that can be applied to the preparation of structurally defined artificial glycoproteins will be discussed. Finally in section 1.5, the aims of my research will be outlined including how synthetic glycoproteins can be produced for applications ranging from the detection of novel carbohydrate binding proteins to potential therapeutics as multivalent inhibitors of bacterial enterotoxins, cholera toxin & verotoxin.

1.2 Tumour Associated Carbohydrate Antigens

Cell surface glycosylation is a dynamic process, which changes at the onset of carcinogenesis and throughout cancer leading to the display of structures either in higher abundance than in normal tissues, or structures that are not present on normal healthy cells. Understanding the wider expression profiles of tumour associated carbohydrate antigens for malignant tumours is essential for identification of targets for cancer immunotherapy and the development of platforms for rapid screening of samples for cancer (Table 1-1).⁴⁻⁷ Globo-H is an unnatural tumour-specific hexasaccharide antigen which was first discovered covalently attached to lipids of human breast cancer cells MCF-7.⁸ Since its discovery it has been observed in other epithelial tumours such as lung, colon, ovarian and prostate cancers.⁹ Gangliosides are a group of natural cell-surface sialylated glycolipids which are over-expressed in numerous neuroectodermal cancers such as melanoma, neuroblastoma, sarcoma and small-cell lung cancer.¹⁰ Expression of the ganglioside GM2 has also been reported for several epithelial cancers including breast, prostate, ovarian and colon cancers.¹¹⁻¹³ To date Globo-H, GD2, GD3 and GM2 have all been investigated as targets for cancer vaccines as a method of activating the immune system.^{8,14-16}

Table 1-1: Expression profile of Tumour Associated Carbohydrate Antigens for malignant tumours.⁴⁻⁷

		Tumour Associated Carbohydrate Antigens										
Tumours	Tn	STn	ST	Le ^a	SLe ^a	Le ^x	SLe ^x	Le ^y	Globo-H	GD2	GM2	GD3
<i>B-cell Lymphoma</i>										+	+	
<i>Bladder</i>	+	+	+			+						
<i>Breast</i>	+	+	+		+	+	+	+	+		+	
<i>Cervical</i>	+	+										
<i>Colon</i>	+	+	+	+	+	+	+	+	+		+	
<i>Endometrial</i>		+	+		+			+	+		+	
<i>Gastric</i>		+		+	+	+		+	+		+	
<i>Liver</i>			+				+					
<i>Lung</i>	+	+					+	+	+		+	
<i>Melanoma</i>					+		+			+	+	+
<i>Neuroblastoma</i>										+	+	+
<i>Ovarian</i>	+	+	+			+	+	+	+		+	
<i>Pancreas</i>				+	+				+		+	
<i>Prostate</i>	+	+	+				+	+	+		+	
<i>Sarcoma</i>										+	+	+
<i>Small-cell Lung</i>					+				+		+	
<i>Stomach</i>	+	+	+			+					+	

Although the glycolipids in Table 1-1 are some of the more widely investigated tumour-associated antigens it is important to note that changes in glycosylation of glycoproteins, and especially mucins widely occur in cancer. Understanding of the changes in the type and levels of mucin *O*-glycosylation are vital in efforts to produce useful synthetic glycoproteins/glycoconjugates which can be used as potential therapeutics or used for the discovery of novel carbohydrate binding antibody mimetics. The changes to Mucin 1 glycosylation, and as a result function, which occur in cancer will be discussed in detail in section 1.2.2. Studies on the synthesis of neoglycoproteins bearing mucin glycopeptides will be described in chapter 2 and studies toward the development of probes for detecting mucin TACAs will be described in chapter 3.

1.2.1 Mucins

For pathogens or small molecules to access the receptors on the host cells, they must first pass through a protective network of mucins.^{2,17} Mucins are classified into two types; secretory mucins produced by goblet cells and membrane bound mucins produced by most epithelial cells.¹⁸ Membrane bound mucins play a more dynamic role than as a defence mechanism, they also have a fundamental role in tissue homeostasis and signal transduction.¹⁸⁻²⁰ As most epithelial carcinomas develop, the expression and glycosylation of membrane bound mucins are hijacked to allow for control of tissue homeostasis and allow for the processes which contribute towards metastasis.²¹ Mucin 1 was the first mucin to be structurally characterised and its aberrant overexpression, glycosylation and role in cancer is the most well studied.²²

1.2.2 Structure & Function of Mucin 1

Mucin 1 (MUC1) is encoded by the *MUC1* gene, found on the long arm of chromosome 1 at position 22 in humans.²³ MUC1 is an extensively *O*-glycosylated phosphoprotein with a core protein mass between 120-225 kDa. Once glycosylated, the total molecular weight of MUC1 can range from 250-500 kDa.²⁴⁻²⁶ During its expression, MUC1 is cleaved into two domains in the endoplasmic reticulum (ER), the extracellular domain and cytoplasmic-transmembrane domain. The cytoplasmic-transmembrane domain non-covalently anchors the extracellular domain to the apical surface of the epithelia (Figure 1-2).^{21,25,26} The extracellular domain is composed of a highly conserved tandem repeat domain of 20 amino acids which is repeated between 20-125 times depending on the individual.^{25,27} This tandem repeat domain is rich in serine and threonine permitting heavy *O*-glycosylation and proline which contributes to the rod-like structure of the extracellular domain which extends up to 500 nm from the cell surface.^{21,24,25,27}

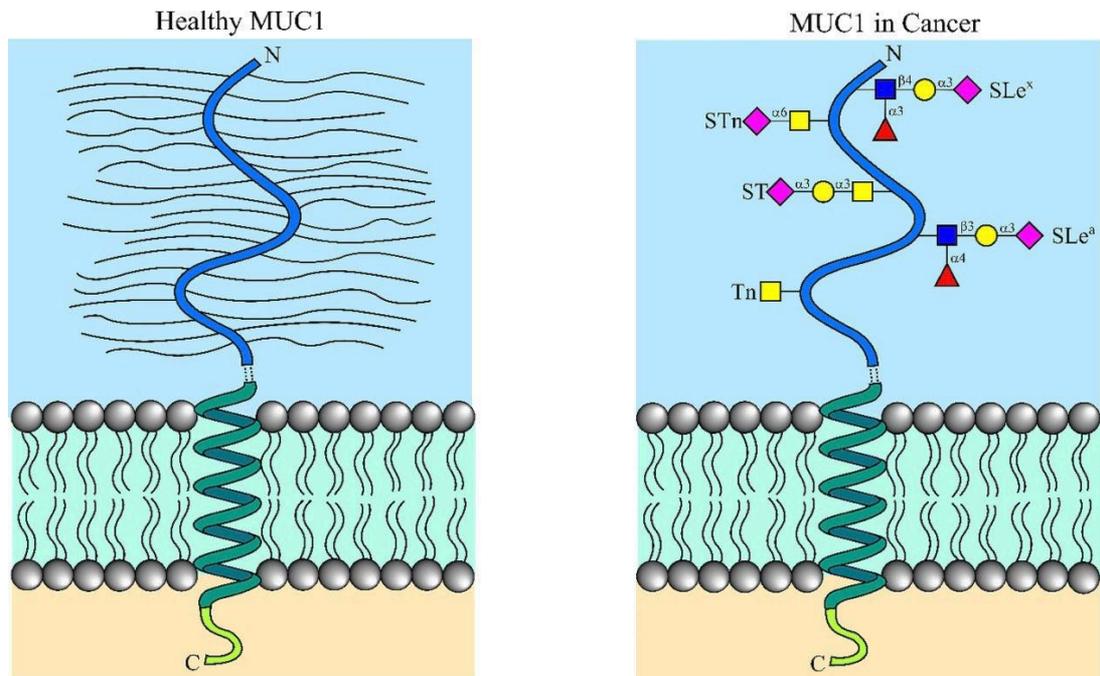


Figure 1-2: Structure of MUC1 in health cells and cancerous epithelial cells. Left: The extracellular domain (blue) in healthy epithelial cells is heavily O-glycosylated with complex glycans. This extracellular domain is non-covalently anchored to transmembrane domain (green). Right: In cancerous epithelial cells aberrant under glycosylation and the appearance of truncated tumour associated carbohydrate glycan structures (Tn, STn, ST, SLe^x & SLe^a).

In healthy human epithelial cells, tethered (non-covalently associated) cell surface mucins are O-glycosylated with complex glycans which protect cells from infection by providing a steric hinderance and disrupting the adhesion of cells and pathogens.^{24,27,28} The extracellular domain of tethered mucins act as releasable decoys upon pathogen binding.^{27,28} These cell surface mucins also play a vital role as modulators of pathogen-induced inflammation.^{28–30} The dense crowding of mucins controls the diffusion of soluble factors from the microenvironment to the cell surface and hence the activation of receptors on the cell surface.^{31,32} The overexpression, under-glycosylation and aberrant intracellular localization has led to the association of MUC1 with various carcinomas.³³ Overexpression of MUC1 by cancer cells is advantageous as its heavy glycosylation permits the binding of various growth factors near their receptors, promoting abnormal proliferation of the cells.²¹ Furthermore, the highly glycosylated extracellular domain prevents the binding of cell surface anoikis-initiating molecules providing a resistance to anoikis, the apoptotic process which occurs in response to a loss of cellular adhesion.³⁴ The cytoplasmic tail of MUC1

binds directly to the p53 regulatory domain resulting in inhibition of the p53-mediated apoptosis pathway.³³ The loss of these apoptotic pathways are prerequisites to allow for the epithelial-mesenchymal transition and metastasis.³⁴

1.2.3 Mucin *O*-linked Glycosylation

Mucin-type *O*-glycans are synthesised in the Golgi apparatus by sequential glycosyltransferase reactions which are well characterised, however little is known about the regulation of mucin-type *O*-glycosylation.^{35,36} Mucin *O*-glycosylation initiates *via* the synthesis of a common precursor GalNAc α -Ser/Thr (Tn antigen). The addition of *N*-acetylgalactosamine (GalNAc) to serine or threonine residues of the mucin backbone is catalysed by a large family of polypeptide GalNAc-transferases.³⁶ From this common precursor, further extensions result in the formation of the Core 1, 2, 3 or 4 structures which comprise the primary glycan structures observed in humans (Figure 1-3).³⁶ Synthesis of the Core 1 structure is catalysed by T-synthase. The active expression of T-synthase is dependent on the co-expression of COSMC (core-1 β 3GalT specific molecular chaperone). COSMC is localised to the endoplasmic reticulum responsible for preventing aggregation and proteasomal degradation of nascent T-synthase and directly interacts with denatured T-synthase.³⁷⁻³⁹ Core 1 structures can be either extended or core 2 structures generated through the addition of *N*-acetylglucosamine by core-2 β 6-GlcNAc transferase.⁴⁰⁻⁴³ Core 3 and 4 structures can be produced from the Tn epitope. Core 3 structures can be generated through the addition of *N*-acetylglucosamine through a β 1-3 linkage to the Tn epitope by core-3 β 3-GlcNAc transferase,⁴⁴ further extension through the addition of GlcNAc through a β 1-6 linkage to the GalNAc of core 3 by core-2/4 β 6-GlcNAc transferase resulting in the core 4 structure.^{42,43} In mammals three isoforms of the β 1,6-*N*-acetylglucosaminyltransferase enzyme exist. Two of these isoforms are only capable of catalysing the formation of the core 2 structure; the third isoform is capable of catalysing the formation of both the core 2 and core 4 structures.⁴³

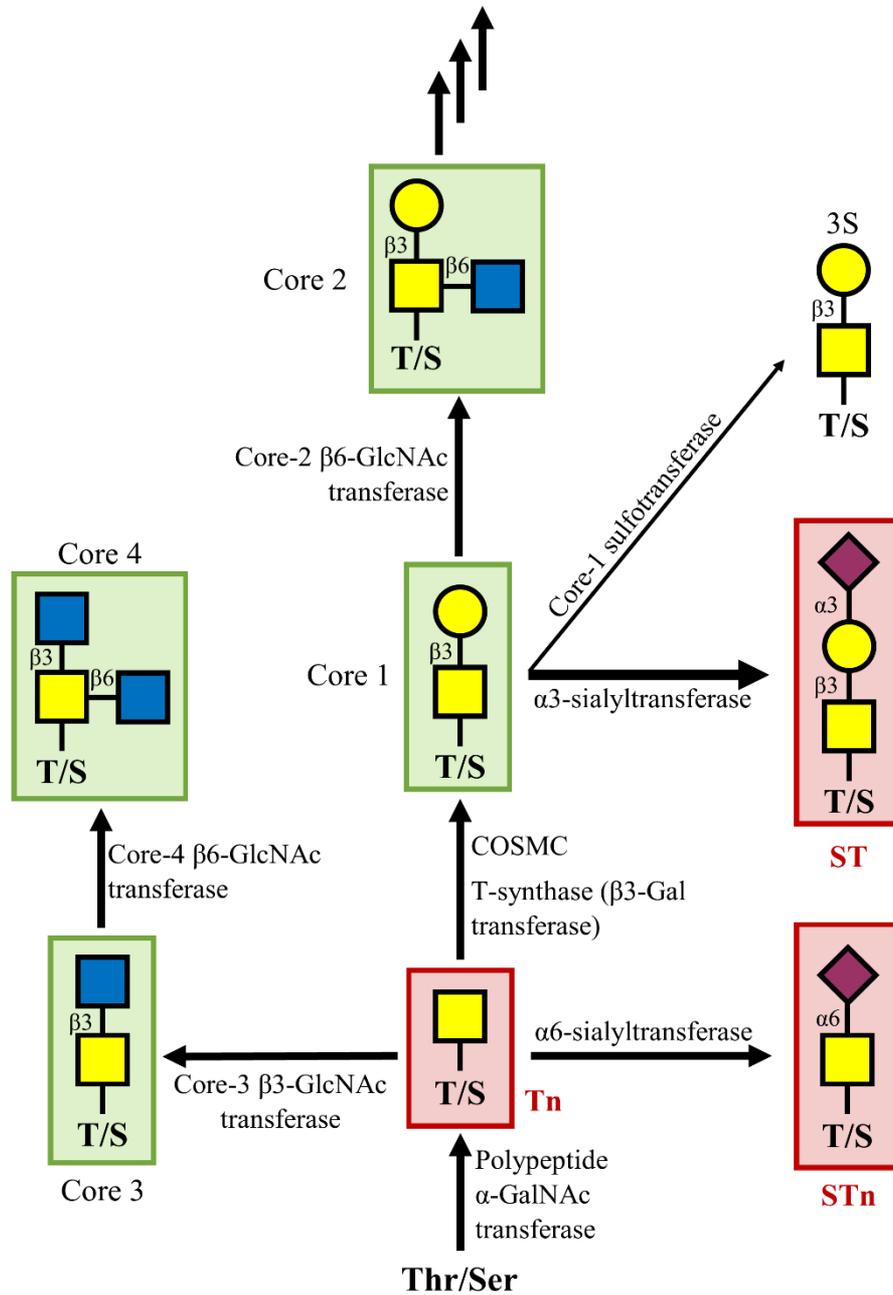


Figure 1-3: Schematic representation of mucin O-type glycosylation. Initiation occurs via the addition of GalNAc to threonine or serine. The resulting Tn epitope is extended into the Core 1, 2, 3, and 4 structures shown in green boxes. The cancer associated epitopes Tn, STn and ST are highlighted in within red boxes.

In many malignant epithelial carcinomas, the expression of mucins becomes deregulated leading to elevated expression. Elevated expression of Mucin 1 is common in pancreatic, breast, colon, lung and prostate cancer.^{33,45-47} Alterations in the glycobiology of Mucin 1 typically occurs through two mechanisms: neo-synthesis or incomplete synthesis.⁴⁸ Alterations typically are characterised by the expression of

truncated core 1 based structures such as Tn, STn and ST (Figure 1-3).⁴⁹ In healthy tissues these structures are typically absent however in many instances of cancer the expression of these truncated structures is driven by alterations to the expression of enzymes involved in the glycosylation process. A significant proportion of cancers exhibit hypermethylation of the *Cosmc* gene resulting in decreased expression of the molecular chaperone required for the correct folding of T-synthase; this results in increased formation of the Tn and STn epitope.^{50,51} Delocalisation of GalNAc transferases from the Golgi to the endoplasmic reticulum (ER) results in an increase in density of GalNAc modification to the Mucin 1 VNTR domain.^{52,53} High density of GalNAc modification is associated with increased aggressiveness in breast cancer.⁵⁴

1.2.4 Glycoconjugate Vaccines

Synthetic oligosaccharide epitopes and neoglycoconjugates offer promising possibilities for the development of vaccines, such as those against *Neisseria meningitidis*, *Streptococcus pneumoniae*, *Salmonella typhi*, HIV, *Plasmodium falciparum*, *Leishmania*.⁵⁵⁻⁶⁰ Therapeutic vaccines against cancer have shown a limited degree of promise with antibodies against carbohydrate-associated tumour antigens able to eliminate circulating tumour cells/micro-metastases in patients.^{61,62} Tumour associated antigens are well tolerated by the immune system and any resulting immune response induced by vaccination very weak.^{12,61,62} As tumours grow, antigens are shed, reinforcing the tolerance. Further, the foreign carrier proteins and linker which attaches the carbohydrate-antigen(s) to the protein typically elicit a strong B-cell response leading to suppression of the response against the carbohydrate epitope.^{61,63,64} The attachment of antigens to an adjuvant scaffold in a well-defined repeatable manner with the minimal length linker results in higher titers of high-affinity anti-glycan antibodies *in vivo*. These high titers of high affinity antibodies are required for long lasting protection and the development of herd immunity.⁶⁵

One of the most common methods for the attachment of glycans in glycoconjugate vaccines is by either reductive amination or NHS labelling of surface exposed lysine residues. Although examples such as KLH-globo H, GM2-KLH and Bacteriophage Q β -Tn vaccines elicit strong IgG immune responses through the recruitment of T-

helper cells the response is typically short term, and all suffer from unpredictability of the conjugation reaction leading heterogeneity between batches.^{8,14,66,67}

1.3 Glycocalyx-Pathogen Interactions

In addition to the essential roles of the glycocalyx in mediating cell-cell interactions, maintaining membrane morphology and membrane integrity in response to physical forces and/or stress, protein-carbohydrate interactions also play a fundamental role in host-pathogen recognition. Bacterial, viral, and parasitic infection is often dependent on recognition of glycoproteins or glycolipids for tissue adhesion and invasion of the cell. The binding of various pathogenic species to their respective glycan ligands within the glycocalyx was reviewed by Imberty and Varrot (Figure 1-4).⁶⁸ Interactions are not limited to recognition of human glycans, pathogens can also exploit carbohydrate binding proteins (lectins) on the human cell surface that recognise bacterial polysaccharides.

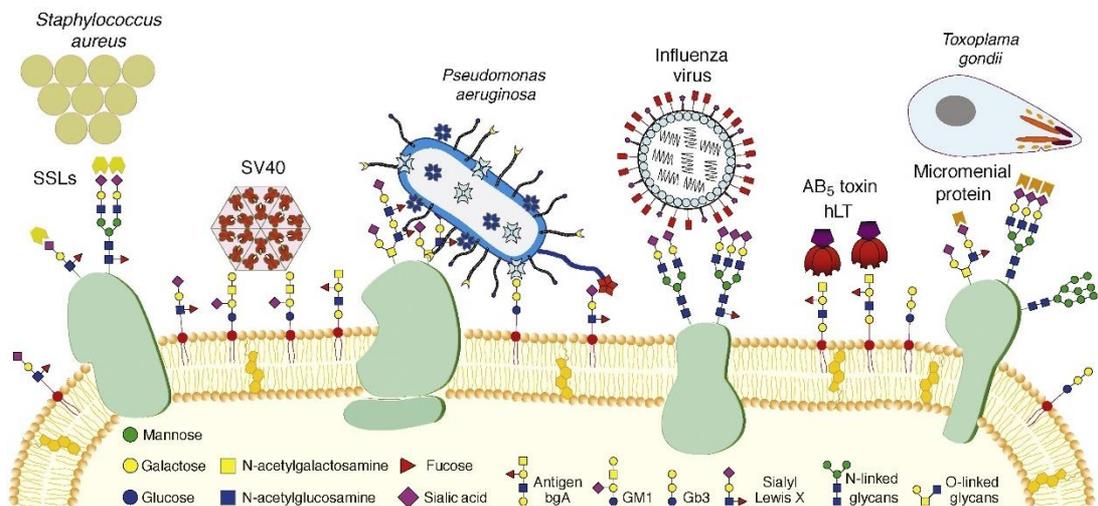


Figure 1-4: Schematic representation of some of the strategies used by pathogens for host glycoconjugates recognition and adhesion. Reproduced from ref 68 with permission from Elsevier.

This section will explore the structure and function of bacterial toxins, cholera toxin and verotoxin. Previous attempts at producing therapeutics to inhibit bacterial toxin binding will be discussed in section 1.3.3. Studies on the synthesis of neoglycoproteins which interrupt the binding of bacterial toxins based will be described in chapters 3 and 4.

1.3.1 Glycolipid Biosynthesis

Glycolipid structures are built upon ceramide or phosphatidylglycerol to create glycosphingolipids and glycophospholipids, respectively, which are embedded within the membrane.⁶⁹ The biosynthesis of glycosphingolipids is known to start in the ER and continue through the Golgi in which glycan extension occurs before being trafficked to the membrane forming lipid rafts.^{70,71} Glycosylphosphatidylinositol (GPI) anchors which are responsible for the anchoring of proteins to the lipid membrane are synthesised from phosphatidylglycerol.^{72,73}

1.3.2 Bacterial Toxins

Some bacteria secrete carbohydrate-binding toxins such as cholera toxin (CT), *E. coli* heat-labile enterotoxin (LT), and tetanus toxin.⁷⁴ Both cholera toxin and heat-labile toxin bind to ganglioside GM1 which is found on the epithelial cells of the gastrointestinal tract; whereas tetanus toxin binds to the gangliosides GT1b and GQ1b.^{75,76} The high avidity binding of their glycan ligands is the first step in endocytosis of the toxin.⁷⁷ The AB₅ family of enterotoxins are composed of two subunits, with a pentameric lectin B-subunit and large toxin A-subunit. The A-subunit has toxic enzymatic activity, whereas the B-subunit has no toxic activity and is only involved in receptor recognition and cellular uptake. Toxin-producing bacteria such as, *Escherichia coli*, *Shigella* species and *Vibrio cholerae*, are major causes of diarrhoeal disease.⁷⁸ In 2016 the World Health Organisation (WHO) ranked diarrhoeal disease as one of the top ten causes of death (9th Globally, 2nd in low-income countries).

1.3.2.1 Cholera Toxin from *Vibrio Cholerae*

The bacterium *V. cholera* is responsible for the disease cholera, a major of gastrointestinal infection in developing countries, which causes diarrhoea, dehydration, and shock; estimated to affect 1.3-4 million people leading to between 21,000 and 143,000 deaths per annum.⁷⁹ Cholera toxin is a prototypical AB₅ toxin composed of a toxic A-subunit associated to a non-toxic pentameric B-unit. The B-subunit is a non-toxic pentameric protein responsible for the delivery of the toxic A-

subunit to the endoplasmic reticulum (ER) (Figure 1-5).^{80,81} Each protomer of the B-subunit has a single binding site for the GM1 oligosaccharide ($K_d \sim 40$ nM, by ITC).⁸² Strong multimeric binding to GM1 on the surface of the cells that line the intestines, allows cholera toxin to enter the cells by endocytosis through the clustering of glycolipids leading to bending of the membrane and the formation of invaginations.^{82–85} A separate binding site in the edge of each B-subunit is capable of recognising the Lewis-y family of oligosaccharides ($K_d \sim 1$ mM (by ITC)).⁸⁶

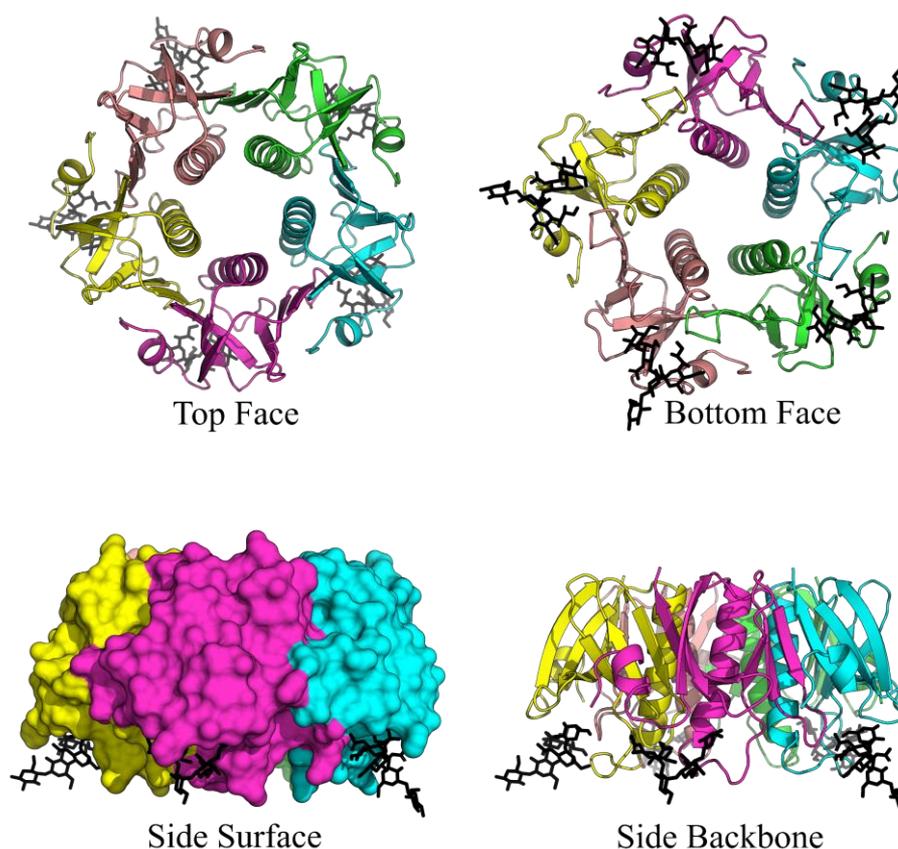


Figure 1-5: Top, bottom and side views of cholera toxin B-subunit with the pentasaccharide GM1 bound. GM1 oligosaccharide shown as black sticks. Figure prepared using PyMOL (version 2.4.0) from Protein Data Bank file 3CHB.pdb

The non-toxic cholera toxin subunit B (CTB) is a strong option for use as a mucosal adjuvant for oral or nasal immunisation.⁸⁰ CTB has increased permeability of the intestinal epithelium leading to enhanced uptake of the co-administered antigen, and enhanced antigen presentation by mucosal antigen-presenting cells (APCs).⁸⁷ Binding of CTB to toll-like receptor 9 (TLR9) on APCs can effectively stimulate both the innate and adaptive immune responses.⁸⁰ As mentioned, CTB has a strong multivalent

interaction with GM1. GM1 is presented on a variety of cells including: epithelial cells of the gut, antigen presenting cells, macrophages, dendritic cells and B cells.⁸¹ This allows CTB to have a strong influence on the immune system, and modulate an immune response against an antigen whether it be coupled by genetic fusion or by chemical manipulation.⁸¹

1.3.2.2 Verotoxin from *Shigella* species

Although cholera is a widely known example of disease caused by toxin-producing bacteria in low-income countries, AB₅ secreting bacteria are not confined to poorer countries. Verotoxin (VT, also known as Shiga-like Toxin) is an AB₅ toxin produced by *E. coli* O157:H7 and various other verotoxin/shiga-like toxin secreting *E. coli* serotypes. Verotoxin causes diarrhoeal disease, with 5-10% going on to develop haemolytic uremic syndrome (HUS); those who develop HUS face a 10% mortality rate.⁸⁷ The highest instances of disease occur in South-East Asia, Europe and American regions through the consumption of contaminated or uncooked food. B-subunit of verotoxin is responsible for binding to the receptor globotriaosylceramide (Gb3Cer, also known as CD77) (Figure 1-6)⁸⁸ Crystal structure analysis has shown that each B-subunit has three Gb3 binding sites, thus each VTB pentamer can bind up to 15 Gb3 simultaneously.^{89,90} VTB has been shown to bind Gb3 oligosaccharide with a dissociation constant of 0.5-1 mM (by ITC).^{91,92} Despite the weak K_d compared to CTB, all the binding sites are arranged on the same face of the protein allowing for multivalent binding which can transform the weak millimolar dissociation constant into sub-nanomolar dissociation constant.⁹² This allows entry to cells by the formation of tubular membrane invaginations, leading to endocytosis.^{84,93} After endocytosis Verotoxin follows either a degradative pathway to lysosomes if it bound to non-lipid raft Gb3 (endosome sorting), or is transported retrograde to the Golgi apparatus or ER if bound to Gb3 associated with lipid rafts.^{94,95}

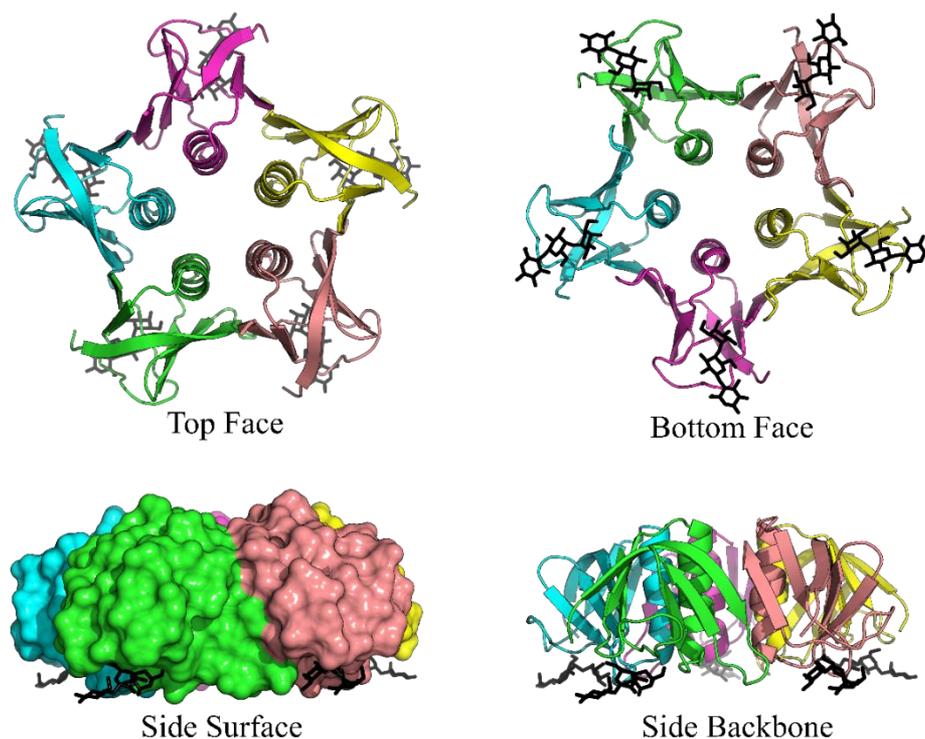


Figure 1-6: Top, bottom and side views of pentameric verotoxin B-subunit (VTB) with trisaccharide Gb3 bound in five of the fifteen binding sites. Gb3 oligosaccharide shown as black sticks. Figure prepared using PyMOL (version 2.4.0) from Protein Data Bank file 1D11.pdb

1.3.3 Multivalent Glycoconjugate Inhibitors

Typically, carbohydrate binding sites are shallow in nature with minimal residues forming contacts resulting in weak individual binding constants in the mM range,⁹⁶ typically measured by isothermal calorimetry (ITC) or surface plasmon resonance (SPR).^{97,98} Despite this protein-carbohydrate interactions play a vital role in cellular function and pathogen adhesion/infection. This is because the majority of interactions are not single events but multiple interactions which combined result in an enhancement of binding affinity to biologically relevant strengths.⁹⁹⁻¹⁰¹ This multivalent binding is driven by the chelate effect, such that the binding of one ligand increases the affinity for the next ligand.^{102,103} The development of neoglycoconjugate inhibitors which take advantage of multivalent binding have been extensively reviewed for various bacterial toxins, lectins and pathogenic agents. This section will investigate some examples of multivalent inhibitors of bacterial AB₅ toxins.

1.3.3.1 Cyclic Multivalent Inhibitors

Cyclic scaffolds which match the valency and symmetry of the target toxin were first reported by Bundle *et al.*¹⁰⁴ The STARFISH inhibitor was created on a functionalised glucose core with ten terminal Gb3 moieties linked at the 2-position of galactose (Figure 1-7). STARFISH was found to have an IC₅₀ of 0.4 nM against Stx-1 and 6 nM against Stx-2, determined by enzyme linked immunosorbent assay (ELISA). The inhibitor was designed to occupy 2 of the 3 Gb3 binding sites per protomer, however crystal studies with the inhibitor bound revealed the inhibitor sandwiched between two different toxin B-subunit pentamers, as opposed to two different binding sites on the same B-subunit.¹⁰⁴

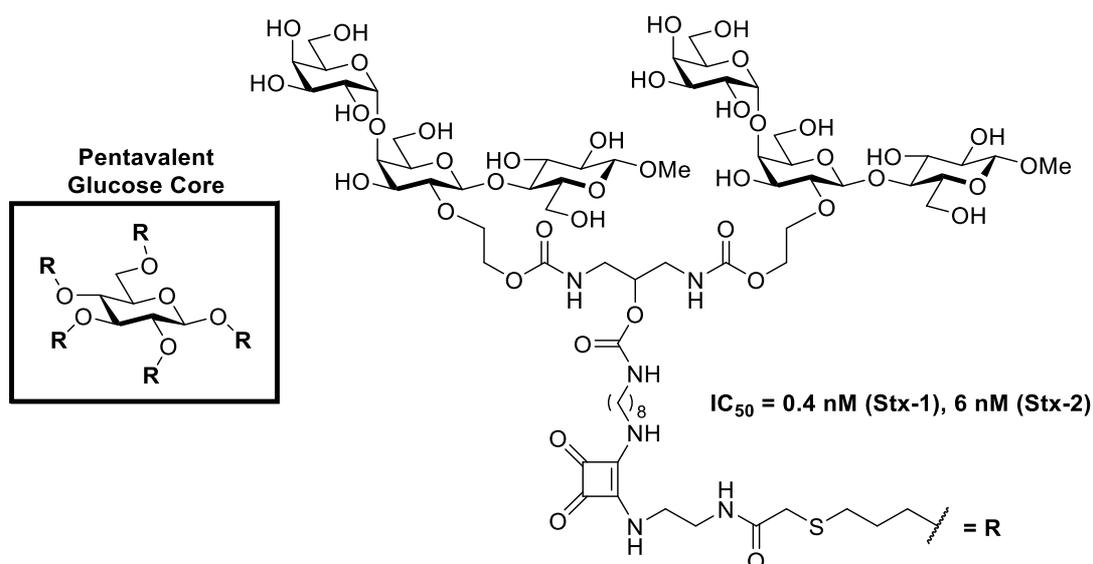


Figure 1-7: Decavalent STARFISH inhibitor of Verotoxin reported by Bundle *et al.*¹⁰⁴

More recently, cyclic pentavalent inhibitors of cholera toxin have been developed using functionalised GM1 oligosaccharide by Siegel *et al.*¹⁰⁵ The symmetrical pentavalent inhibitor was built on a corannulene core functionalized with GM1 with varying length linkers (Figure 1-8). The strongest inhibitor was reported to have an IC₅₀ of 5.1 nM (n = 2), which represents a 3700-fold improvement in potency compared to monovalent GM1os, determined by enzyme linked immunosorbent assay (ELISA).¹⁰⁵

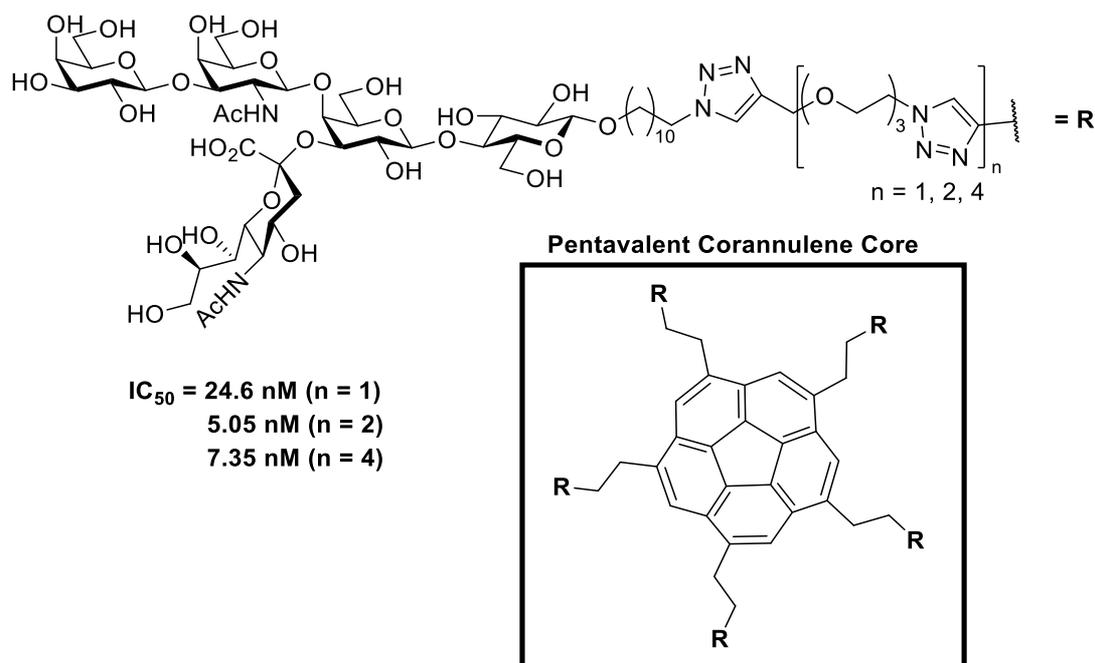


Figure 1-8: Pentavalent GM1os functionalised corannulene cholera toxin inhibitor reported by Siegel *et al.*¹⁰⁵

1.3.3.2 Protein as Scaffolds for Multivalent Inhibitors

Fan *et al.* investigated a divalent bridging approach to inhibitor design, using serum amyloid P component (SAP) and its ability to bind carboxyproline as a method of displaying *m*-nitrophenyl galactoside in the correct symmetry for binding to CTB. Prearranging of the ligands on SAP gave a significant decrease in IC_{50} of $0.98 \mu\text{M}$ against CTB compared to $620 \mu\text{M}$ for the bivalent ligand in the absence of SAP (Figure 1-9a), determined by enzyme linked lectin assay (ELLA).¹⁰⁶ A similar approach was adopted by Bundle *et al.* using cyclic ketal ligands bound to SAP as a method to display Gb3 in an appropriate conformation and spacing to bind to Verotoxin B-subunit (VTB) (Figure 1-9b). In the presence of SAP the heterodivalent cyclic ketal Gb3 ligand had an IC_{50} of $30 \mu\text{M}$ against Stx-1, determined by enzyme linked immunosorbent assay (ELISA).¹⁰⁷

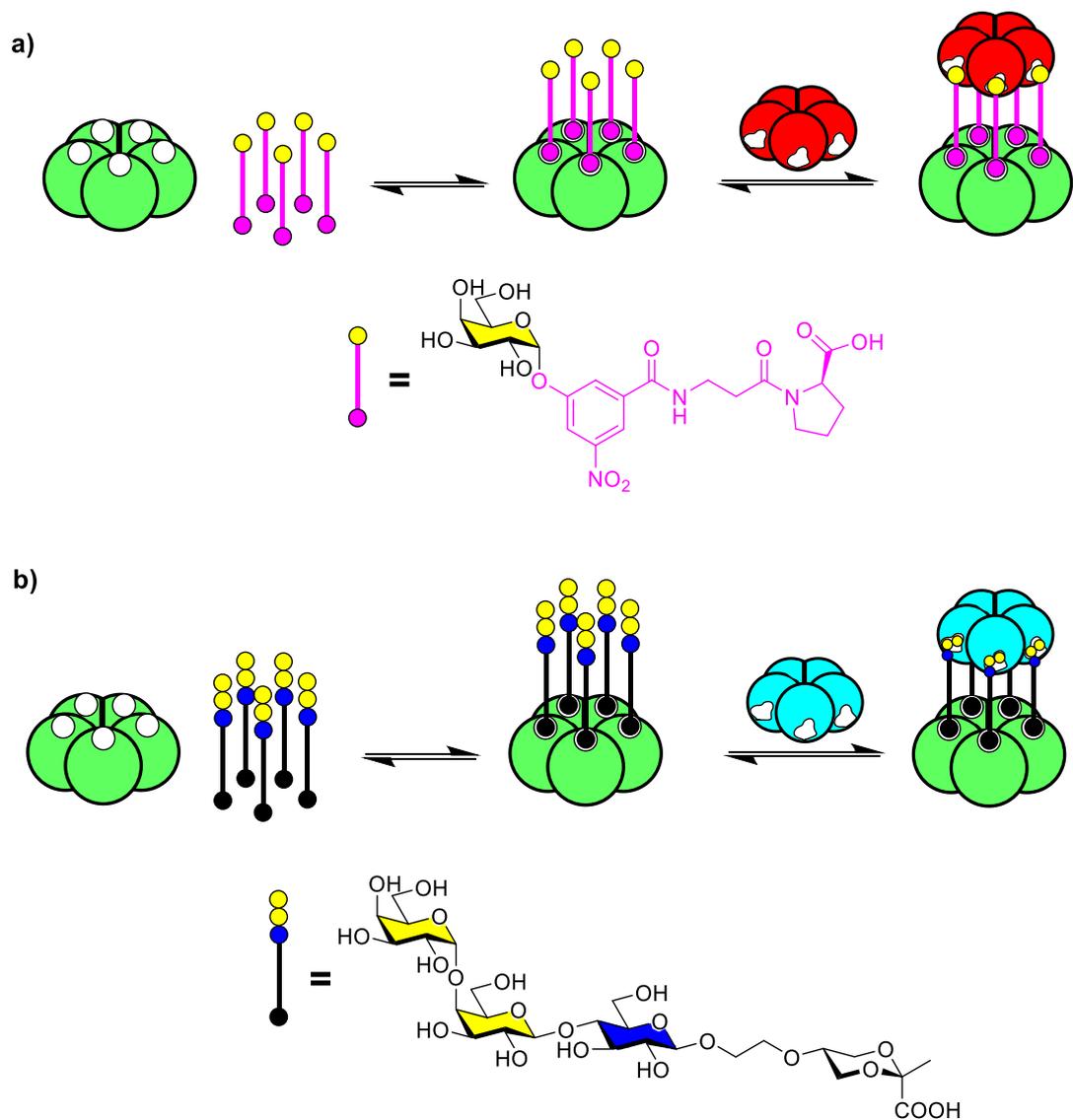


Figure 1-9: Protein templated inhibition of bacterial toxins using heterodivalent ligands capable of binding SAP (green) and bacterial toxin (CTB – red pentamer, VTB/StxB – blue pentamer).^{106,107}

An alternative approach to creating protein-templated design was adopted by Branson *et al.* using neoglycoproteins as multivalent inhibitors.¹⁰⁸ Here, a non-binding variant of the CTB pentamer was covalently derivatised with GM1 *via* N-terminal oxime ligation. Figure 1-10 shows the synthesis of the pentavalent inhibitor from a non-binding variant of CTB-subunit, which demonstrated an IC_{50} of 104 pM against CTB (determined by enzyme-linked lectin assay (ELLA)).¹⁰⁸

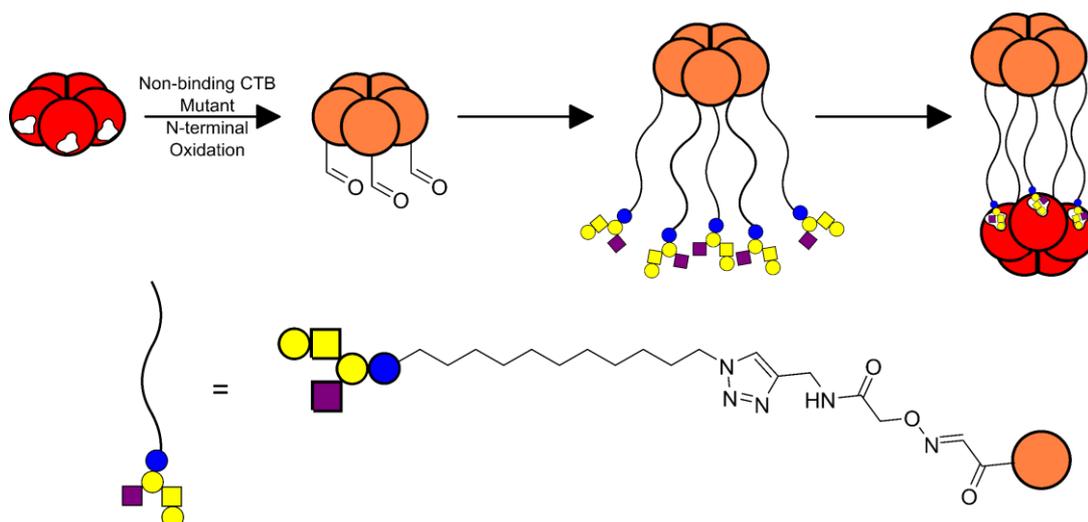


Figure 1-10: Cartoon depiction of novel protein-based CTB inhibitor produced by mutation of wild-type CTB to form a non-binding variant, and attachment by oxime ligation of synthetic carbohydrate derivative created through chemoenzymatic synthesis to form the first neoglycoprotein inhibitor.¹⁰⁸

Chemical modification of the N-terminus of each protomer preorders the glycan in a pentameric arrangement, which closely matches the distance between GM1 binding sites. Dynamic light scattering (DLS) and size exclusion chromatography (SEC) confirmed a 1:1 heterodimer between the inhibitor and WT CTB, suggesting each glycan is occupying the corresponding toxin binding sites.¹⁰⁸ Inhibitors based upon protein scaffolds clearly show the benefit of prearranging glycan in the correct geometric orientation, leading to more potent inhibitors.

1.3.3.3 Glycopolymer & Glycodendrimer-based Inhibitors

Many groups have utilised glycopolymers and glycodendrimers to create highly polyvalent glycosylated structures as inhibitors of bacterial toxins. Recently the Pieters group reported the development of a multiple carbohydrate-based inhibitors of Shiga toxin using galabiose which was attached *via* CuAAC to both a dendrimer scaffold and hyperbranched polymer scaffold (hPG) (Figure 1-11).¹⁰⁹ Inhibition studies showed the monosaccharide building block β -galabiose azide had an IC_{50} of 4.97 mM against StxB-1; increasing the valency by four, the dendrimeric scaffold was shown to have an IC_{50} of 13.5 μ M against StxB-1, a 92-fold improvement in inhibition per galabiose. The best inhibition was demonstrated by the hyperbranched galabiose

glycopolymers with a valency of 20 which exhibited an IC_{50} of 8.3 nM against StxB-1, a 30,000-fold improvement in inhibition per galabiose.¹⁰⁹

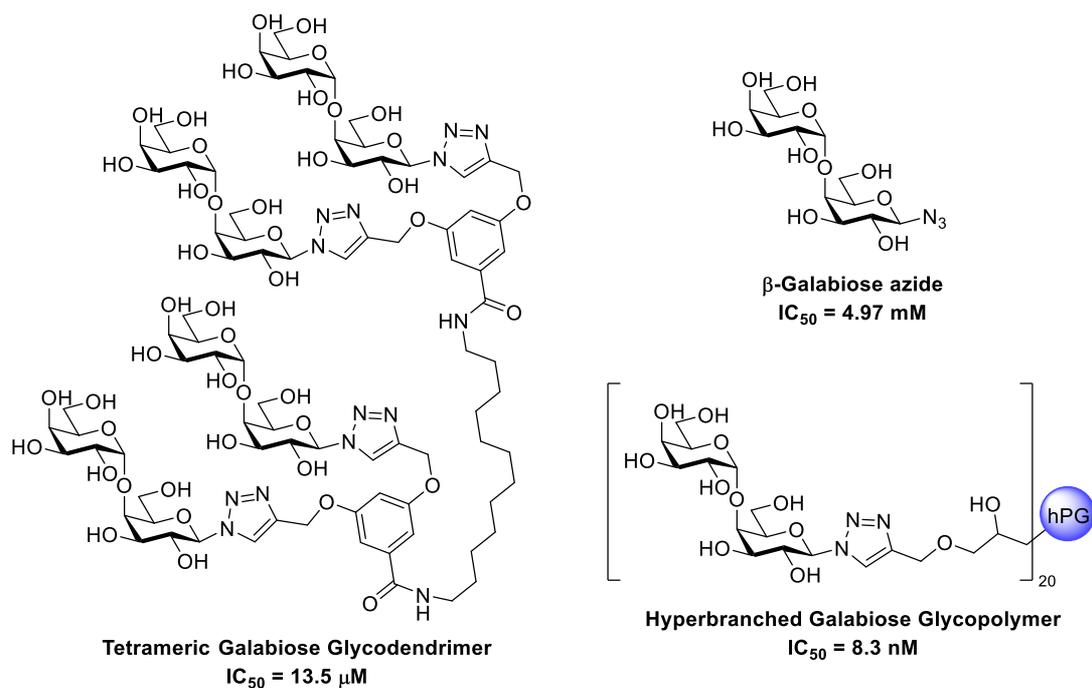


Figure 1-11: Multivalent glycopolymers and glycodendrimer inhibitors of Shiga toxin based on galabiose by Pieters et al.¹⁰⁹

1.4 Synthetic Glycoproteins & Glycoconjugates

The production of synthetic glycoconjugates is a rapidly growing region of interest in terms of prophylactics, therapeutics and as probes in biological research.⁶¹ However, endeavours are complicated by the lack of general methods for routine preparation.⁶¹ Unlike proteins or nucleic acids, glycoproteins produced in biological systems are often found in low concentration and as multiple different glycoforms.^{12,60} As discussed in section 1.2.4, heterogeneity in the attachment of glycans in traditional neoglycoproteins in which glycans are attached to surface exposed lysine residues by either reductive amination or acylation traditionally leads to poor therapeutic output. The use of site-specific methodologies, many of which are already in use within the Turnbull and Webb groups, could be used to present multiple peptide and/or glycopeptide epitopes on to a protein scaffold, removing the issue of microheterogeneity. This section will discuss the variety of methodologies which can

be employed to modify proteins in a site-specific manner and produce well-defined glycoconjugates.

1.4.1 Unnatural Amino Acid Incorporation

The genetic code of all known organisms specify the same 20 amino acid building blocks which have limited functionality.¹¹⁰ Both site-directed and random mutagenesis have been utilised as methods of improving protein stability, catalytic activity and binding affinity/specificity. However, it is clear from nature that sometimes additional post-translational modifications are required for protein functionality/stability.^{111,112} Chemical modification of the common amino acids with reactive side-chains, such as Lys, Cys and Tyr, can lack site-specificity and introduction of a new surface exposed Cys or Lys can be challenging, limiting such methodologies. The ability for researchers to site-specifically introduce unnatural amino acids bearing specific functional groups has greatly expanded our knowledge of protein function and allowed the creation of proteins with bio-orthogonal handles for labelling.¹¹¹⁻¹¹⁴ Over 70 unnatural amino acids have been developed to expand the genetic code of *Escherichia coli*, yeast and mammalian cells (Figure 1-12).¹¹⁵

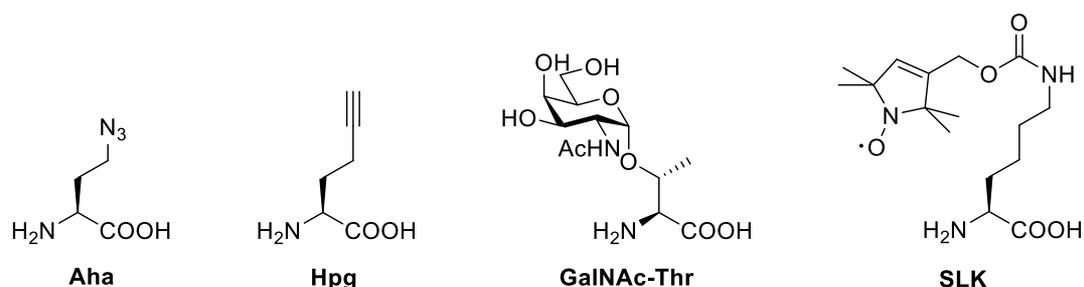
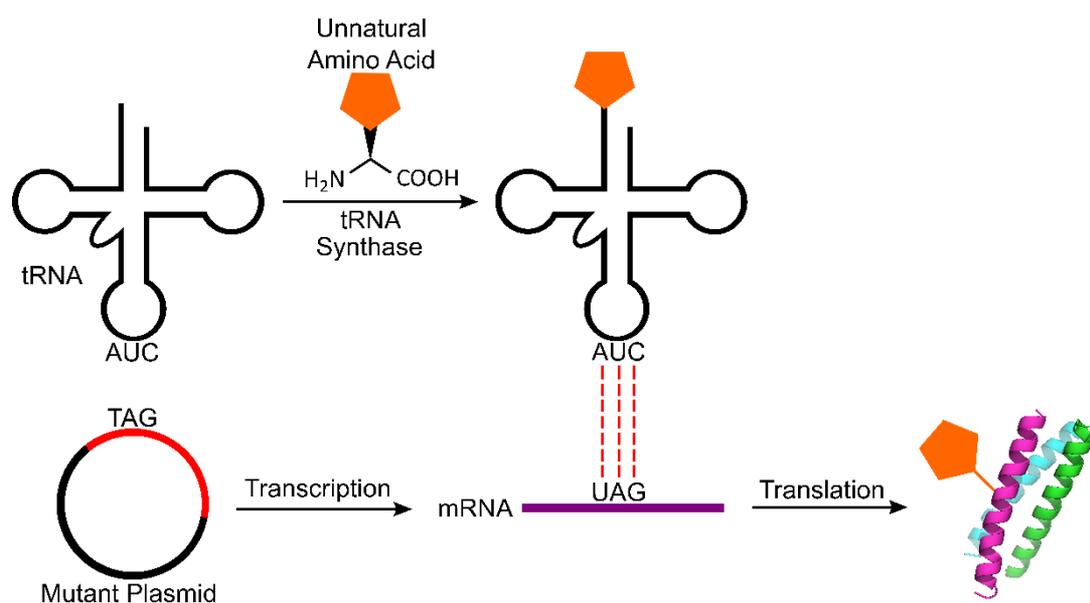


Figure 1-12: Examples of unnatural amino acids incorporated into proteins using amber suppression.

In amber suppression the amber stop codon (UAG) is hijacked and instead used to encode an unnatural amino acid loaded onto a complementary tRNA at one or more sites (Scheme 1-1). The original *in vitro* approach was low yielding and required complex and labour-intensive synthesis of the aminoacylated tRNA_{CUA}.^{116,117} The modern methodology utilises an evolved, orthogonal tRNA_{CUA}/tRNA-synthase pair to synthesise the aminoacylated orthogonal tRNA *in vivo*.^{118,119} The orthogonal tRNA_{CUA} is only a substrate for orthogonal tRNA-synthase and not the endogenous tRNA synthases, allowing for the incorporation of unnatural amino acids which are

also substitutes for the orthogonal tRNA-synthase.¹²⁰ To carry out this methodology, the genes which encode the orthogonal tRNA, orthogonal tRNA-synthase and the modified protein, with amber stop codon mutation, must be inserted into a heterologous host and the cell culture supplemented with the unnatural amino acid.^{117,118} The site-specific incorporation of *N*-acetylgalactosamine- α -*O*-threonine *via* amber suppression has been reported in the literature.^{121,122} However due to difficulty in repeating the experiments and the uncertainty whether the incorporated amino acids were in fact natural in origin casts doubt over these results.^{123–125}



Scheme 1-1: Schematic representation of amber suppression. The aminoacylated-tRNA is synthesised from tRNA_{CUA} using the orthogonal tRNA synthase. The protein with the amber stop codon TAG is transcribed to mRNA. The aminoacylated-tRNA_{CUA} recognises the amber stop codon UAG within the mRNA allowing for site-specific incorporation of the unnatural amino acid into the protein of interest.

A simpler alternative to amber suppression is global residue-specific incorporation using isosteric replacement of canonical amino acids.¹²⁶ Providing the non-canonical amino acid analogue is a substrate for the natural tRNA synthase, the cellular translation machinery can be hijacked allowing for complete replacement of one amino acid at every position it is coded for in the genetic code by its non-canonical analogue.^{126,127} Most commonly methionine mimics are incorporated, such as azidohomoalanine (Aha) or homopropargylglycine (Hpg), however analogues of leucine, tryptophan, phenylalanine, and proline are available.^{126–129} This methodology has been used within the Turnbull & Webb groups to produce an azido-CTB protein

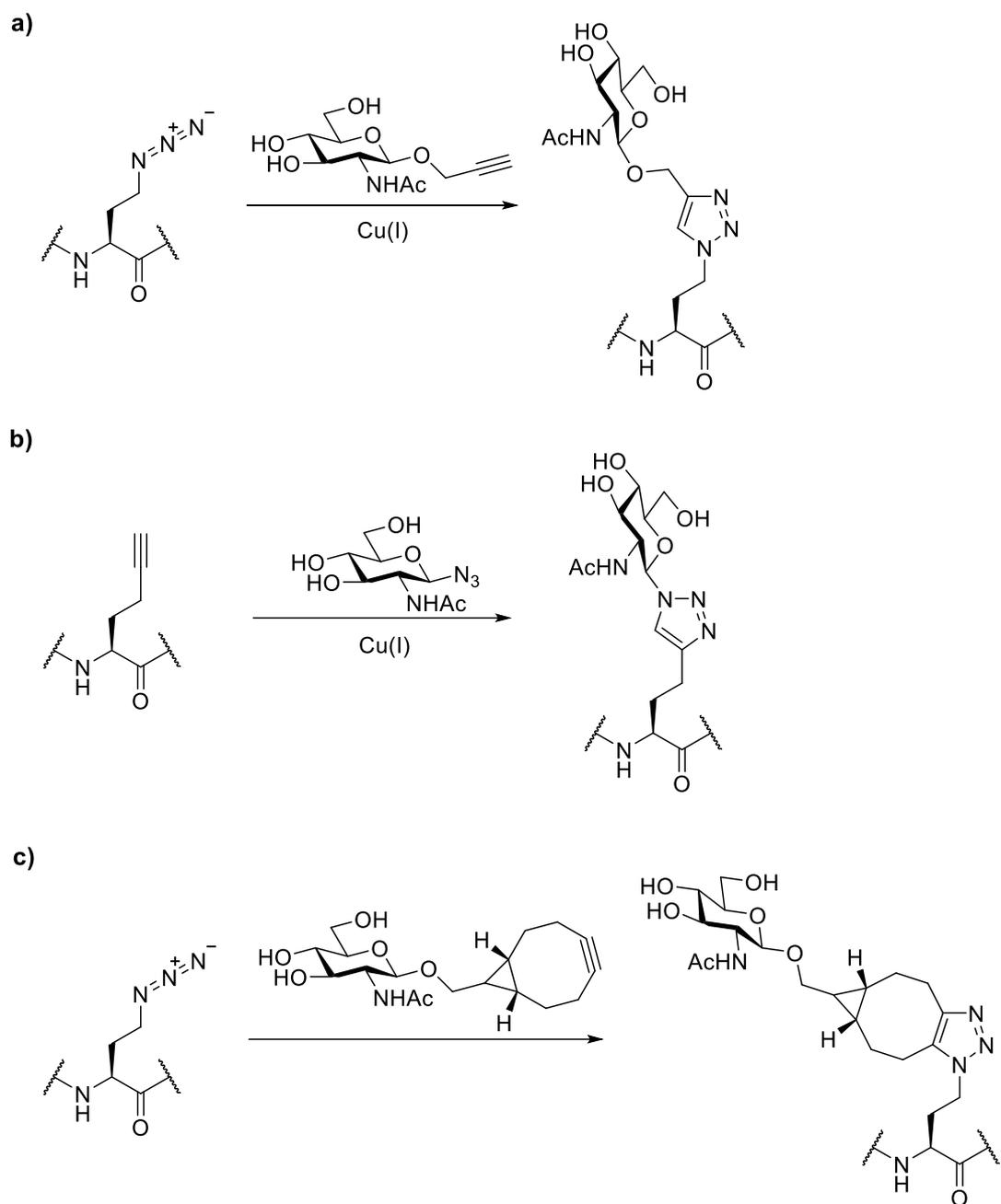
in which a single azidohomoalanine residue by mutating lysine 43 to methionine (K43M).¹³⁰ Sequential site-directed mutagenesis was employed to remove the three native methionine residues in the mature protein.

1.4.2 Post-translational Glycosylation

Following incorporation of a reactive side chain by genetic code expansion using techniques such as amber suppression or global residue-specific incorporation, numerous examples exist of chemical modification of these side chains, termed post-translational.¹³¹

1.4.2.1 Azide-Alkyne Cycloaddition

Incorporation of azidohomoalanine (Aha) or homopropargylglycine (Hpg) provides powerful handles for bioorthogonal modification of proteins.¹²⁶ Copper-catalysed azide-alkyne cycloadditions (CuAAC) and strain-promoted azide-alkyne cycloadditions (SPAAC), also termed ‘click’ reactions for their fast kinetics,¹³² have proven powerful bioorthogonal reactions which have played a pivotal role in chemical biology (Scheme 1-2).¹³³ For true bioorthogonality, the reactants must be biologically inert, no-off target reactivity, and products must not be chemically reactive. Synthetic glycosylation of proteins using CuAAC or SPAAC have been used widely for the preparation of glycoconjugate vaccines,¹³⁴ neoglycoprotein analogues,^{135,136} fluorinated glycoprotein mimics.¹³⁷ Although CuAAC is commonly regarded as bio-orthogonal this can be questioned due to the potential for copper chelation by the protein, the formation of reactive oxygen species leading to oxidation of protein side chains and toxicity of copper in living systems.^{138,139}

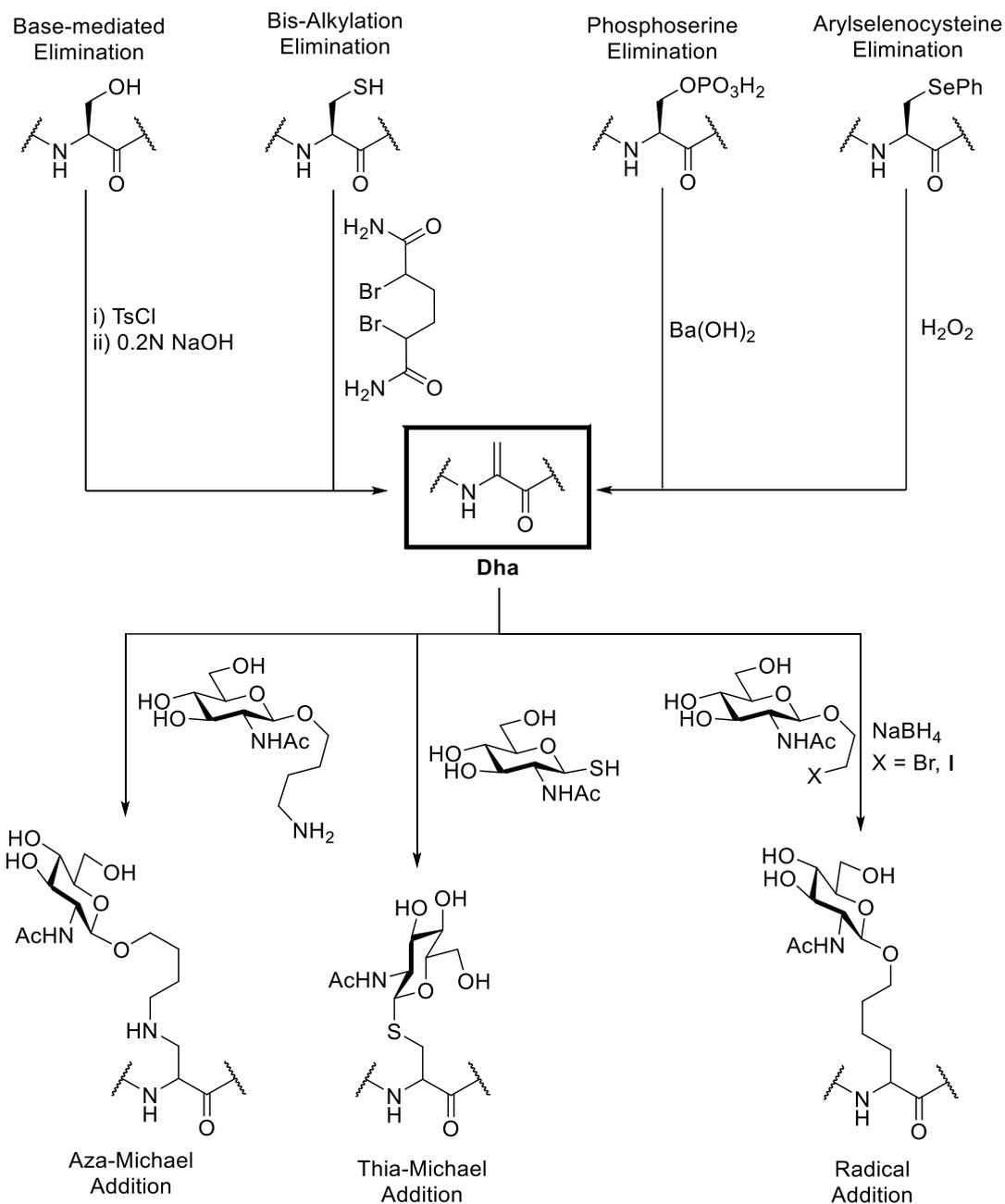


Scheme 1-2: Examples of chemical glycosylation using a) CuAAC to Aha containing protein, b) CuAAC to Hpg containing protein and c) SPAAC ligation to Aha containing protein using BCN functionalised glycan.

1.4.2.2 Dehydroalanine

One of the most widely utilised chemical modifications is the generation of dehydroalanine (Dha) from cysteine, (phospho)serine or arylselenocysteine (Scheme 1-3). In nature dehydroalanine is formed from the enzymatic dehydration of serine in the biosynthesis of lanthionine-containing peptides, such as Nisin.¹⁴⁰ In chemical biology, the generation of Dha was first adopted as a method of transforming a serine protease to a cysteine protease;^{141,142} later it was adopted by Schultz as a useful method to modify protein side-chains and backbone structures.¹⁴³

The α,β -unsaturated carbonyl of Dha has become a tool of chemical glycosylation *via* aza-, thia-Michael or radical additions. The Davis group has utilised the generation of Dha *via* bis-alkylation elimination followed by thia-Michael addition to produce S-linked GlcNAc glycoproteins which can be further extended by Endo-A catalysed glycoylation.^{144,145} However, the use of Dha is not without challenges, conjugation reactions suffer from a lack of stereo-control, difficulty controlling off target side reactions with other nucleophilic side chains and obtaining high levels of conversion for sterically hindered sites.



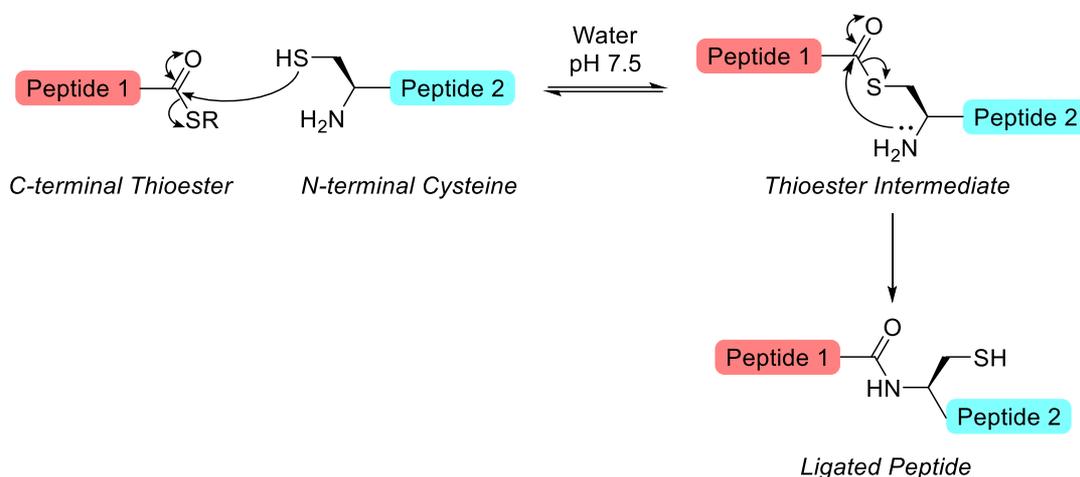
Scheme 1-3: A selected range of methods to introduce dehydroalanine (Dha) into proteins and site-specific glycosylation methods including nucleophilic (aza-, thia-Michael) and radical additions to Dha.¹⁴⁵⁻¹⁴⁷

1.4.3 Modification of N-Terminal Cysteines

Although the incorporation of unnatural amino acids can be used to provide bio-orthogonal handles for protein labelling, the expression of proteins containing modified or unnatural amino acids is not trivial and high levels of optimisation may be required. Protein modification of the N-terminus offers an appealing alternative as it could be used to afford site-specificity but with minimal effect on the levels of protein expression or the function of the protein. The reactivity of N-terminal cysteine residues has given rise to an array of methodologies which will be discussed further in this section.

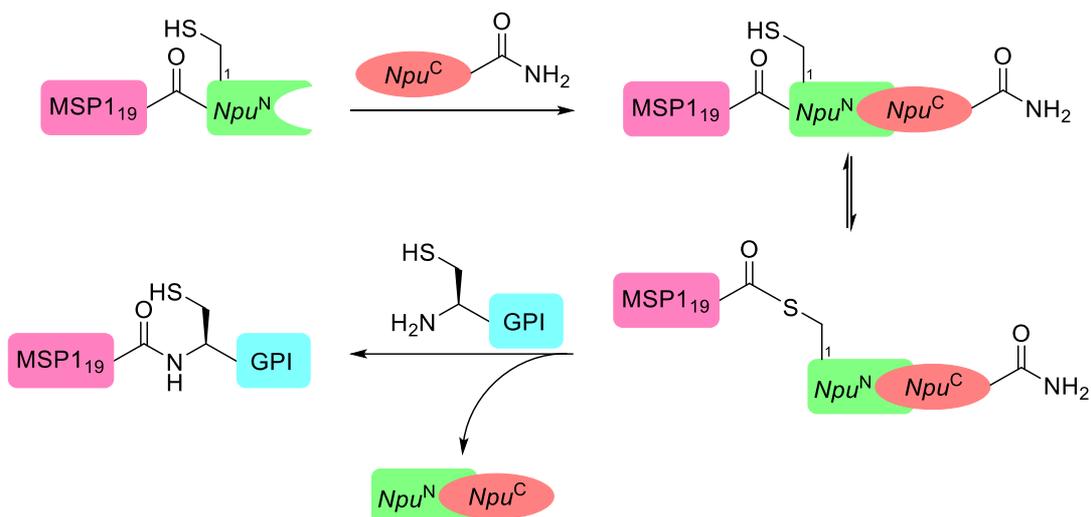
1.4.3.1 Native Chemical Ligation

Introduced by Kent and co-workers in 1994, native chemical ligation involves the coupling of two peptides: one with a C-terminal thioester, the other with a N-terminal cysteine.^{148,149} The key to the regioselectivity and orthogonality of native chemical ligation is the reversibility of the thiol(ate)-thioester exchange in the presence of excess thiol corresponding to the thioester leaving group (Scheme 1-4).¹⁴⁸⁻¹⁵⁰ This excess thiol serves to prevent oxidation of the N-terminal cysteine residues to form disulfide-linked dimers and allow for reversibility in the reaction where internal cysteine residues are present in one or both peptide fragments.¹⁴⁸ The irreversible amide formation (S-to-N acyl shift) occurs spontaneously under the reaction conditions (Scheme 1-4).¹⁴⁸⁻¹⁵⁰ The unique stability of thioesters to hydroxide-catalysed hydrolysis and reactivity towards thiolysis and aminolysis allows for near-quantitative ligation with no loss of peptide-thioester to hydrolysis.¹⁴⁸ In recent years reaction rates have been tuned to allow near-quantitative ligation in few hours at room temperature in the presence of 200 mM (4-carboxymethyl)thiophenol.¹⁵¹ Native chemical ligation has become a key methodology in the preparation of glycoproteins such as GlyCAM-1 and Erythropoietin.¹⁵²⁻¹⁵⁴



Scheme 1-4: The use of native chemical ligation to link two peptide species, one with a C-terminal thioester and the other with an N-terminal cysteine. In the presence of excess thiol (corresponding to the thioester leaving group) the initial step is reversible.

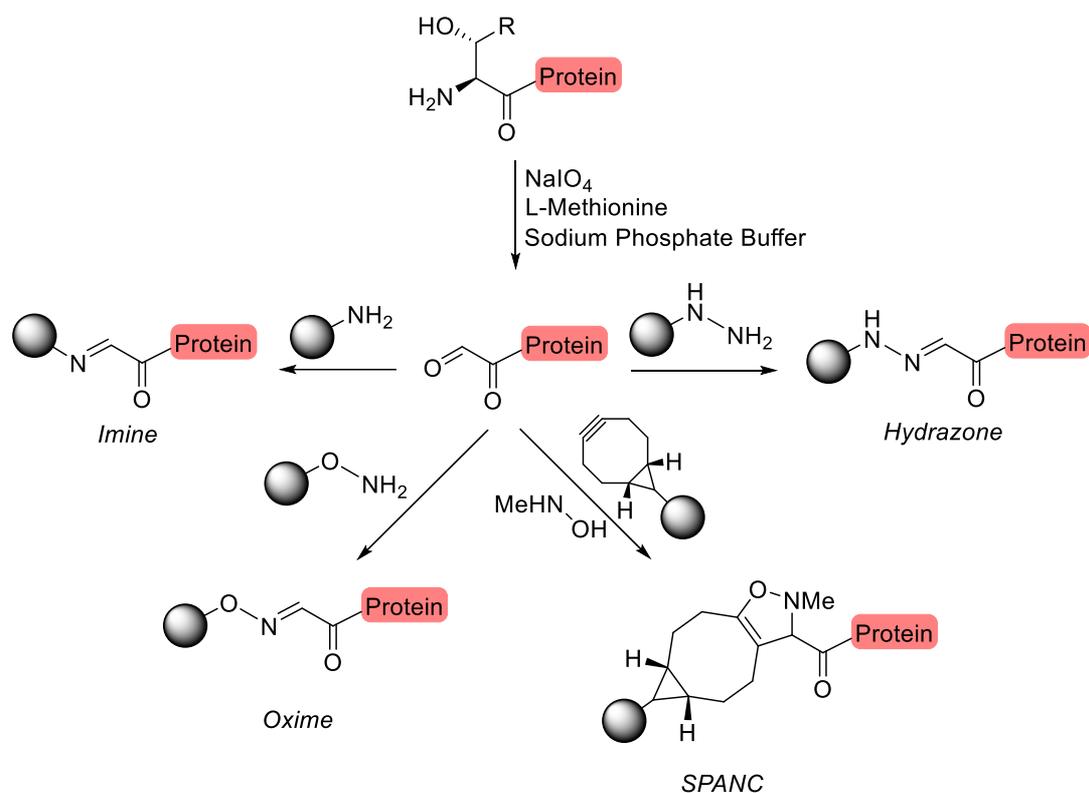
More recently this has been expanded to inteins, naturally occurring auto-processing domains which excise themselves from a polypeptide and join the two remaining portions (exteins) with an amide bond in a mechanism which mimics that of native chemical ligation, which has enhanced and simplified many of the applications of native chemical ligation.^{117,155–158} In 2020, reports in *Angewandte Chemie* described the semisynthesis of functional glycosylphosphatidylinositol-anchored proteins using a non-splicing mutant of the *Nostoc punctiforme* DnaE split intein (Npu^C/Npu^N).¹⁵⁹ MSP1₁₉-GPI found in *Plasmodium berghei* was prepared using two complementary portions of the intein (Int_C and Int_N) which associate noncovalently to produce the active intein capable of protein splicing (Scheme 1-5).^{156,159–161} The semisynthetic MSP1₁₉-GPI was found to induce an enhanced production of pro-inflammatory cytokines IL-12 and TNF- α by bone-marrow-derived dendritic cells.¹⁵⁹ In future this methodology could provide a convenient route to the production of glycolipidated proteins as potential vaccine candidates for a variety of parasitic infections.¹⁵⁹



Scheme 1-5: Semisynthetic pathway to the synthesis of MSP1₁₉-GPI. Adapted from Seeberger and Varón Silva et al.¹⁵⁹

1.4.4 N-terminal Modification of Threonine & Serine

The functionalisation of the highly reactive α -oxo aldehyde to form imine-, hydrazone-, or oxime-bonds have become amongst the most reliable reactions in the bioconjugation toolbox.^{162,163} The α -oxo aldehyde is generated *via* oxidative cleavage of 1,2-amino alcohols, typically exposed serine or threonine residues located at the *N*-terminus or genetically incorporated ϵ -lysine dipeptides harbouring a 1,2-amino alcohol motif (Scheme 1-6).^{108,163,164} Oxime-ligation is ideal for preparing bioconjugates that are stable at biological pH in aqueous media. This is due firstly to higher stability of hydrazones and oximes to hydrolysis due to donation of electron density from the heteroatom adjacent to the sp^2 nitrogen compared to imines.¹⁶⁵ Comparatively in aqueous conditions (pH 7.0) the first-order rate constant for hydrolysis of oximes is 600-fold lower (25 days half-life) than that of hydrazones.^{162,165}



Scheme 1-6: Sodium periodate cleavage of 1,2-amino alcohols to unmask highly reactive α -oxo aldehyde and oxime, imine & hydrazone-bond formation.

Using oxime ligation, the naturally occurring N-terminal Thr residue of cholera toxin B-subunit (CTB) was site-specifically labelled with five aminoxy-peptides. Each of the aminoxy-peptides contained a main chain lysine residue harbouring a 1,2-amino alcohol motif on the ϵ -amino group which could be labelled further by repeated rounds of periodate oxidation and oxime ligation. Using this approach 25 copies of a synthetic aminoxy antigenic peptide based on hemagglutinin from the influenza virus were conjugated onto CTB.¹⁶⁶ However, in order for the oxime ligation to proceed, the pH had to be lowered significantly resulting in dissociation of the pentamer and necessitating refolding at a later stage. In 2006, Dirksen identified that aniline or *p*-methoxyaniline could be used as a catalyst to accelerate the rate of oxime ligation of peptides at pH 4.5 and pH 7.0 (respectively).¹⁶⁷ Aniline catalysis was later used to allow the modification of CTB under neutral conditions (pH 7.0) by Branson *et al.* in 2016, discussed in section 1.3.3.2.¹⁰⁸

1.4.5 Transpeptidases

Transpeptidase enzymes provide an attractive alternative to the previously discussed methodologies to modify proteins. Transpeptidase enzymes are highly selective for a peptide recognition sequence and responsible for linking two peptides together *via* amide bond formation. In particular, the use of Sortase A will be discussed in detail in this section.

1.4.5.1 Sortase A

Sortases are a family of transpeptidase enzymes responsible for sorting and reversibly catalysing the attachment of virulence factors to the cell walls of Gram-positive bacteria.^{168,169} Sortase A (SrtA) is a type II membrane protein native to *Staphylococcus aureus* that ligates proteins carrying the LPXTGX recognition motif (where X is any amino acid) and peptidoglycan substrates carrying an N-terminal oligoglycine motif.^{170,171} The catalytic cysteine (Cys184) in the active site of SrtA cleaves the amide bond between the threonine and glycine forming a thioacyl-enzyme intermediate, which is subsequently attacked by the oligoglycine substrate to form the ligated product (Figure 1-13).^{169,172}

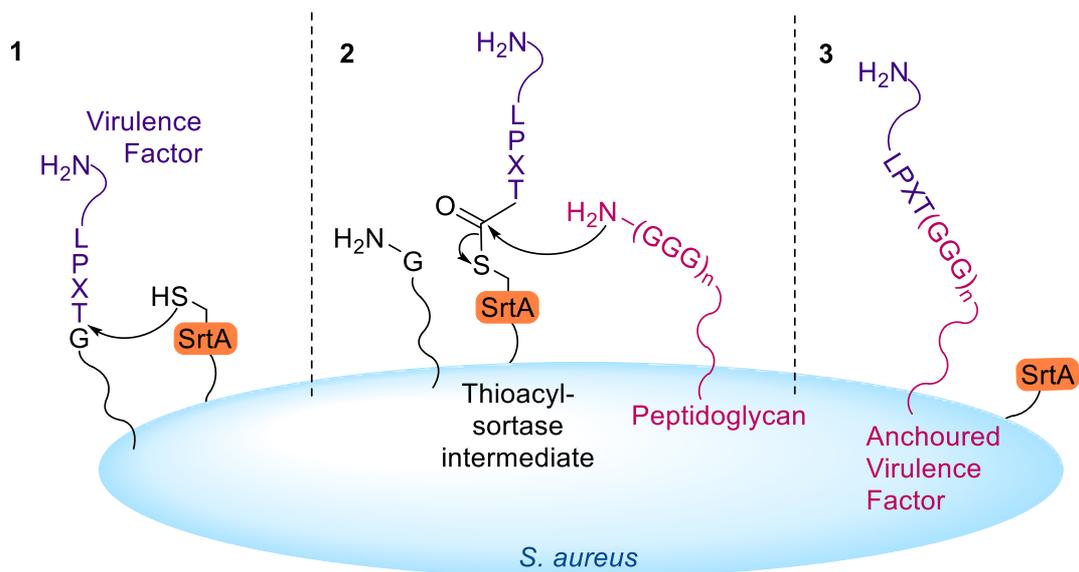


Figure 1-13: Reversible mechanism of SortaseA for attachment of virulence factors to the cell-wall of Gram-positive bacteria. Step 1: The catalytic cysteine (C184) cleaves the amide bond between Thr and Gly within the LPXTG recognition motif. Step 2: The thioacyl-sortase intermediate is attacked by an N-terminal glycine of the peptidoglycan anchored in the cell wall. Step 3: A new amide bond is formed, releasing sortaseA. The virulence factor is covalently attached to the peptidoglycan and anchored to the cell wall. Adapted from Guimares et al.¹⁷²

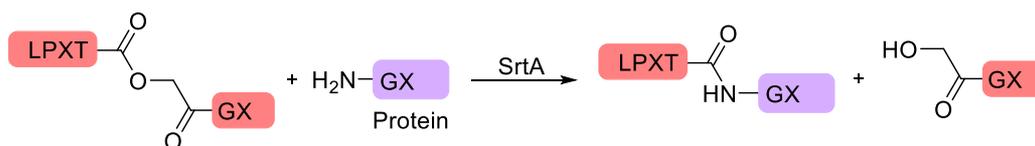
Protein labelling reactions utilising the calcium dependent wild type enzyme require large excesses of labelling peptide, stoichiometric quantity of the SrtA and prolonged incubation times in order to obtain acceptable yields. The discovery of five point mutations near the LPXTGX binding site (P94R, D160N, D156A, K190E, K196T) by directed evolution yielded the pentamutant Srt5M which shows up to 140-fold increase in K_{cat}/K_M , allowing for reactions to be performed at low temperature.^{170,173} Discovery of a further two mutations (E105K, E108A/Q) yielded a calcium independent enzyme.^{174,175} This variant increases the scope of this technique by making it compatible with a wider range of buffers and protein substrates. When combined, these seven mutations yield the heptamutant Srt7M capable of site-specific labelling of a target protein under mild conditions.¹⁷⁰ Reactions utilising Srt7M can be performed using catalytic quantities of the enzyme and fewer equivalents labelling substrate compared to the wild type making it a much more powerful labelling technique for producing well-defined conjugated antigen-adjuvant vaccines.¹⁷⁶

However due to the increased activity of the enzyme an alternative pathway is possible in which the thioacyl intermediate is irreversibly hydrolysed.^{177,178}

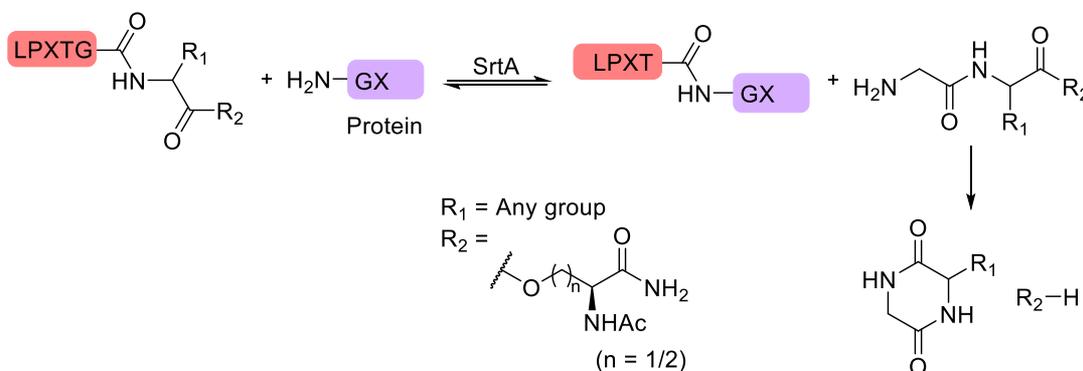
1.4.5.2 Irreversible Sortase Labelling

Multiple groups have developed methods to render sortase labelling reactions irreversible; removing the need to use a large excess of labelling reagent which may be expensive and/or difficult to synthesise. Two such methods for N-terminal sortase labelling of a protein of interest are the use of depsipeptide substrates or the intramolecular rearrangement of the by-product to form diketopiperazine (Scheme 1-7).^{168,169,179} Williamson *et al.* found that introduction of an ester bond between threonine and glycine of the LPXTGX motif results in formation of hydroxyacetyl as by-product which is unable to participate in the reverse reaction, rendering the reaction irreversible (Scheme 1-7a). Using this approach, a range of proteins were quantitatively labelled using 1.5 equivalents of depsipeptide and 10 mol% SrtA within 4 hours at 37 °C.¹⁶⁸ An alternative approach by Liu *et al.* uses the formation of diketopiperazine (DKP) to drive the equilibrium of the sortase reaction to completion (Scheme 1-7b).¹⁷⁹ Two peptide substrates were discovered which were suitable substrates for sortase, LPETGG-isoacyl-serine and LPETGG-isoacyl-homoserine. Following ligation, the by-product undergoes intramolecular rearrangement to form diketopiperazine and isoacyl-(homo)serine, neither of which participate in the reverse reaction.¹⁷⁹

a) Irreversible Depsipeptide Labelling



b) Irreversible Diketopiperazine Formation

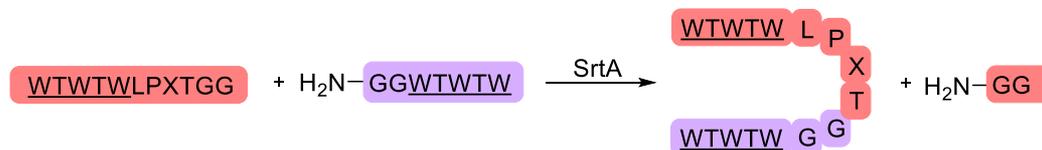


Scheme 1-7: Two methods of by-product deactivation to improve the N-terminal sortase labelling efficiency. A) The introduction of an ester bond between threonine and glycine results in the hydroxyacyl to be formed as by-product, rendering the reaction irreversible. B) Formation of diketopiperazine by-product drives the sortase reaction to completion by removing the reverse reaction nucleophile.

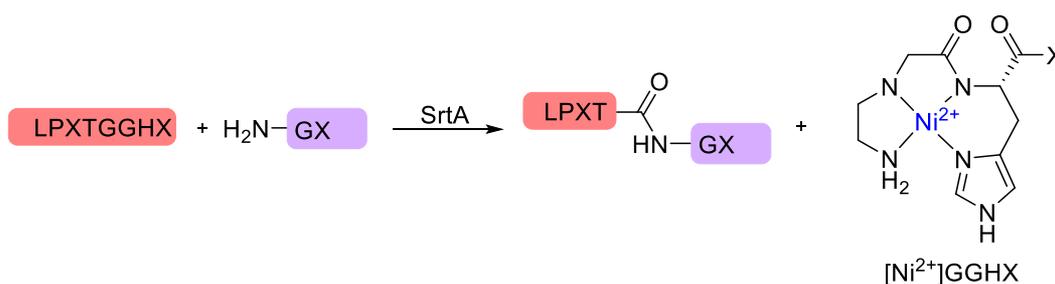
The previous two optimizations only worked for N-terminal labelling of a protein of interest as both required the synthesis of peptides containing an ester linkage. Two further methods have been developed which can be applied to N- or C-terminal labelling of a protein of interest. The first takes advantage of secondary structure to conformationally restrict the ligated product from fitting into the active site of SrtA. Yamamura *et al.* used a Trp zipper-derived sequence to form a stable β -hairpin in the ligated product.¹⁸⁰ Sortase A has a large “L-shaped” groove leading to the active site into which the LPXTG motif binds;¹⁸¹ introduction of a rigid secondary structure around the LPXTG motif inhibits binding, shifting the equilibrium of the reaction to favour formation of the ligated product (Scheme 1-8a).¹⁸⁰ This secondary structure element significantly increases the length of the linker by at least 10 amino acids, a simpler method by the group of John M. Antos enhances the efficiency of sortase reaction by reducing the nucleophilicity of the by-product preventing it from participating in the reverse reaction.^{182,183} Extending the LPXTGX motif to LPXTGGH, results in the release of GGH as by-product which can bind with high

affinity to Ni^{2+} ions forming a square-planar complex (Scheme 1-8b). Addition of 2 equivalent of NiSO_4 resulted in an increase in labelling of 27% from 58% in the absence of NiSO_4 , when using 1 equivalent of the labelling peptide and 10 mol% SrtA.¹⁸²

a) β -Hairpin (Trp Zipper) Formation



b) Nickel-Peptide Complex Formation



Scheme 1-8: Two methods of by-product deactivation to improve the C- or N-terminal sortase labelling efficiency. A) Introduction of Trp zipper inhibits reverse reaction through introduction of secondary structure element which prevents recognition of LPXTG motif. B) Formation of a square planar nickel-peptide complex reduces the nucleophilicity of the by-product rendering the reaction virtually irreversible.

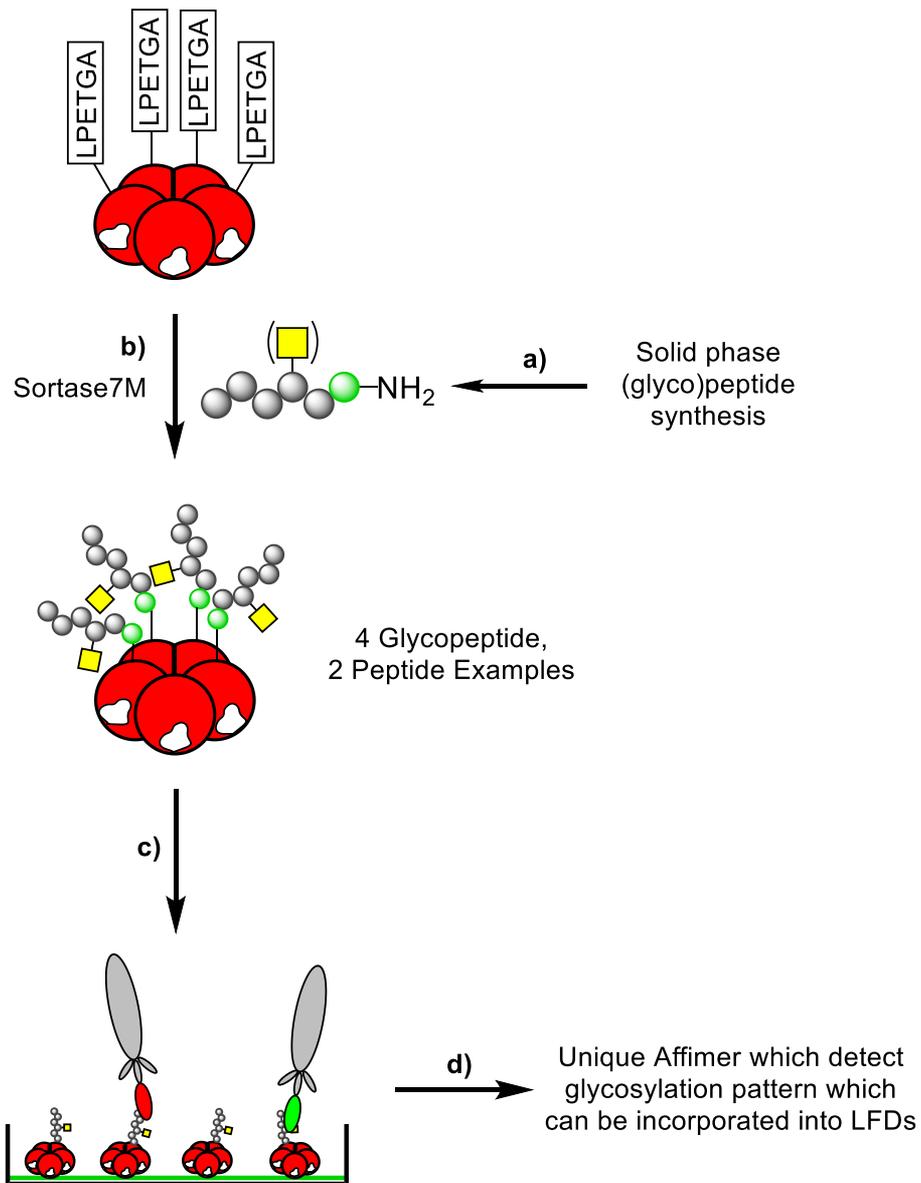
1.5 Project Outline

This work aims to build upon the previous work of the Turnbull and Webb groups using sortase-mediated ligation or a combination of strain-promoted azide-alkyne cycloaddition and oxime ligation to produce well defined glycoconjugates for use in the detection of novel carbohydrate binding proteins or as multivalent inhibitors of bacterial enterotoxins, cholera toxin & verotoxin.

1.5.1 Detection of Novel Carbohydrate Binding Proteins

In recent years lateral flow immunoassays have attracted considerable interest because of the potential to provide near-instantaneous diagnosis directly to patients.¹⁸⁴ Their simple, cheap and easy to use design have made them an attractive option for mass testing; however, the majority of tests rely on polyclonal antibodies conjugated to gold nanoparticles as a method of detection.^{184,185} Whilst cheap and easy to purify, polyclonal antibodies suffer from high heterogeneity within the antibody pool and differences in specificity between batches, both of which can lead to chances of cross reactivity leading to the production of false positives.¹⁸⁶

In this portion of the project, we aimed to investigate if an Affimer, an engineered antibody-like binding protein, could be used as an alternative for polyclonal antibodies. Chapter 2 will discuss the optimisation of sortase ligation to conjugate a series of MUC1 glycopeptide epitopes displaying specific glycosylation patterns to the C-terminus of cholera toxin B-subunit. Then chapter 3 will investigate if the CTB-glycoconjugates can be used to screen for novel carbohydrate binding proteins. Here CTB will be used as a scaffold to preorder the glycopeptides ensuring the correct orientation during phage display allowing for well-defined, high-density display of the glycopeptides to the Affimer-bacteriophage (Scheme 1-9).

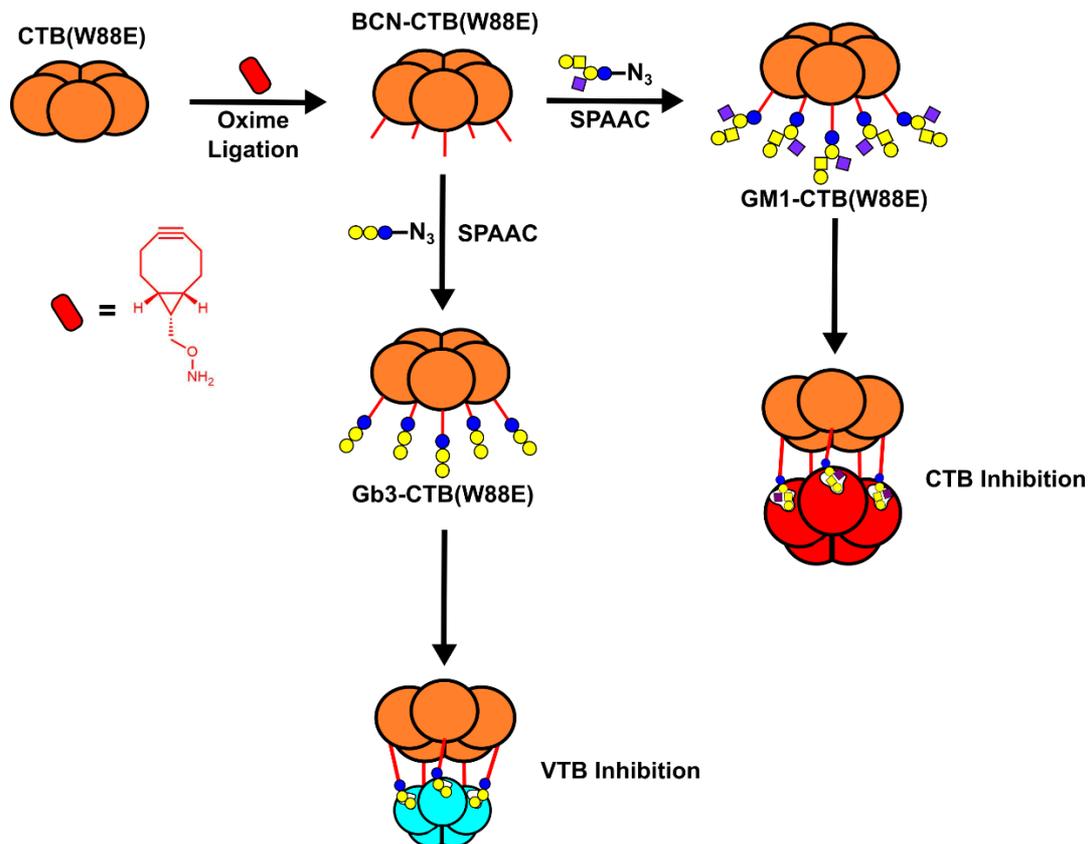


Scheme 1-9: Workflow for discovery of Affimer which bind to carbohydrate antigens. a) Preparation of series of glycopeptide and peptide epitopes by solid phase (glyco)peptide synthesis. b) Sortase-mediated C-terminal labelling of CTB scaffold protein. c) Affimer phage display using GM1-coated microtiter plate to orient CTB, displaying glycopeptides in optimal orientation for presentation to Affimer-bacteriophage. d) Evaluation of Affimer binding affinity and selectivity, followed by incorporation into lateral flow devices (LFDs).

1.5.2 Multivalent Neoglycoconjugate Inhibitors

Previously Branson *et al.*¹⁰⁸ reported the generation of neoglycoconjugate inhibitors of bacterial toxins. This work illustrated that site-specific glycosylation and preordering of the glycan ligands to a scaffold with a similar size and symmetry to that of the target is an effective strategy for developing multivalent inhibitors. Building on this, Dr Ryan McBerney (University of Leeds) developed a minimal length heterodivalent linker that combines a strained alkyne and oxyamine group (Scheme 1-10) to examine the effect of linker length and site of glycosylation had on inhibitory potential. Changing both linker length and site of glycosylation resulted in a 4-fold reduction in inhibition.¹⁸⁷

Chapter 4 will examine the independent impact changing linker length or site of glycosylation has on inhibitory potential, using the same minimal length heterodivalent linker. Following attachment of the heterodivalent linker to the N-terminus *via* oxime ligation, SPAAC will be employed to attach a variety of glycosyl azides (Scheme 1-10). Performing reactions in this order avoids tricky purification and characterisation of potential reactive aminoxy-functionalised glycans. This work will be expanded to investigate the inhibitory potential of Gb3-neoglycoproteins produced using the heterodivalent linker. Finally, chapter 5 will discuss some preliminary work to advance neoglycoprotein inhibitors towards a potential therapeutic by investigating non-bacterial protein scaffolds.



Scheme 1-10: Workflow for preparation of site-specific multivalent glycoconjugate inhibitors of bacterial toxins (CTB – Cholera toxin B-subunit (red) & VTB – Verotoxin B-subunit (blue)). Non-GM1 binding variant of CTB (orange) labelled with BCN group via oxime ligation followed by SPAAC of glycosyl azide.

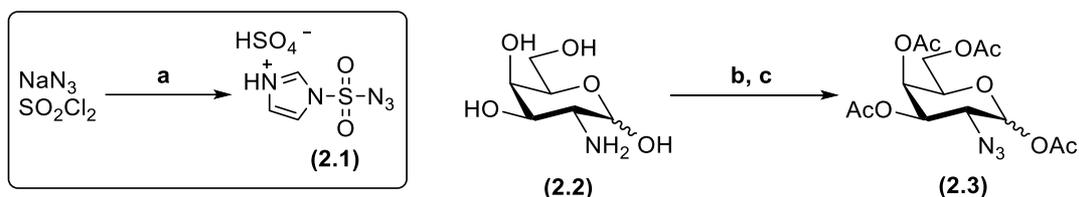
Chapter 2: Preparation of tumour-associated MUC1 cholera toxin B-subunit conjugates

The work within this chapter discusses the generation of a series of well-defined tumour-associated MUC1 cholera toxin B-subunit glycoconjugates using C-terminal sortase 7M-mediated ligation. Using solid phase peptide synthesis, the synthesis of a library of MUC1(Tn) glycopeptide fragments will be described demonstrating all potential glycosylation patterns. Using sortase-mediated ligation these glycopeptide fragments were site-specifically conjugated to the C-terminus of a multimeric scaffold (CTB). The effect of changing sortase conditions and linker length will be discussed as methods for improving conjugational efficiency.

2.1 Glycopeptide Preparation

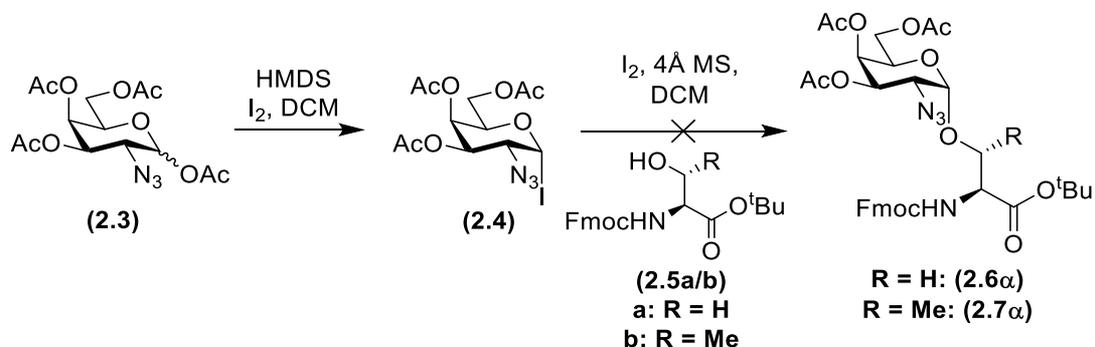
2.1.1 Synthesis of GalNAc-Ser/Thr Amino Acid Building Blocks

As the essential building blocks for solid phase glycopeptide synthesis, the synthesis of the peracetylated α -O-GalNAc-Ser/Thr amino acids (**2.11 α** /**2.12 α**) from D-galactosamine (**2.2**) was performed following a literature procedure.^{188,189} Firstly, the 2-amino group was azido-protected *via* copper(II)-mediated diazotransfer using 1H-imidazole-sulfonyl azide (**2.1**) (Scheme 2-1). Multi-gram quantities of **2.1** were safely isolated *via* precipitation as the hydrogen sulfate salt using a procedure which avoided the isolation or concentration of potentially explosive intermediates such as sulfonyl diazide and hydrazoic acid.^{190,191} Due to the sensitive nature of **2.1**, diazotransfer was performed immediately in a 5:3 mixture of MeOH/H₂O, after which, the hydroxyl groups of 2-azido-galactose were protected by acetylation with acetic anhydride in pyridine. The protected galactosamine (**2.3**) was prepared in a 62% yield with an anomeric ratio of 1:10 α/β .



Scheme 2-1: Protecting group strategy of galactosamine: (a) EtOAc, Imidazole, H₂SO₄ (61%); (b) 2.1, CuSO₄, K₂CO₃, MeOH, H₂O; (c) Ac₂O, pyridine, $\alpha/\beta=1/10$ (62%, 2 steps).

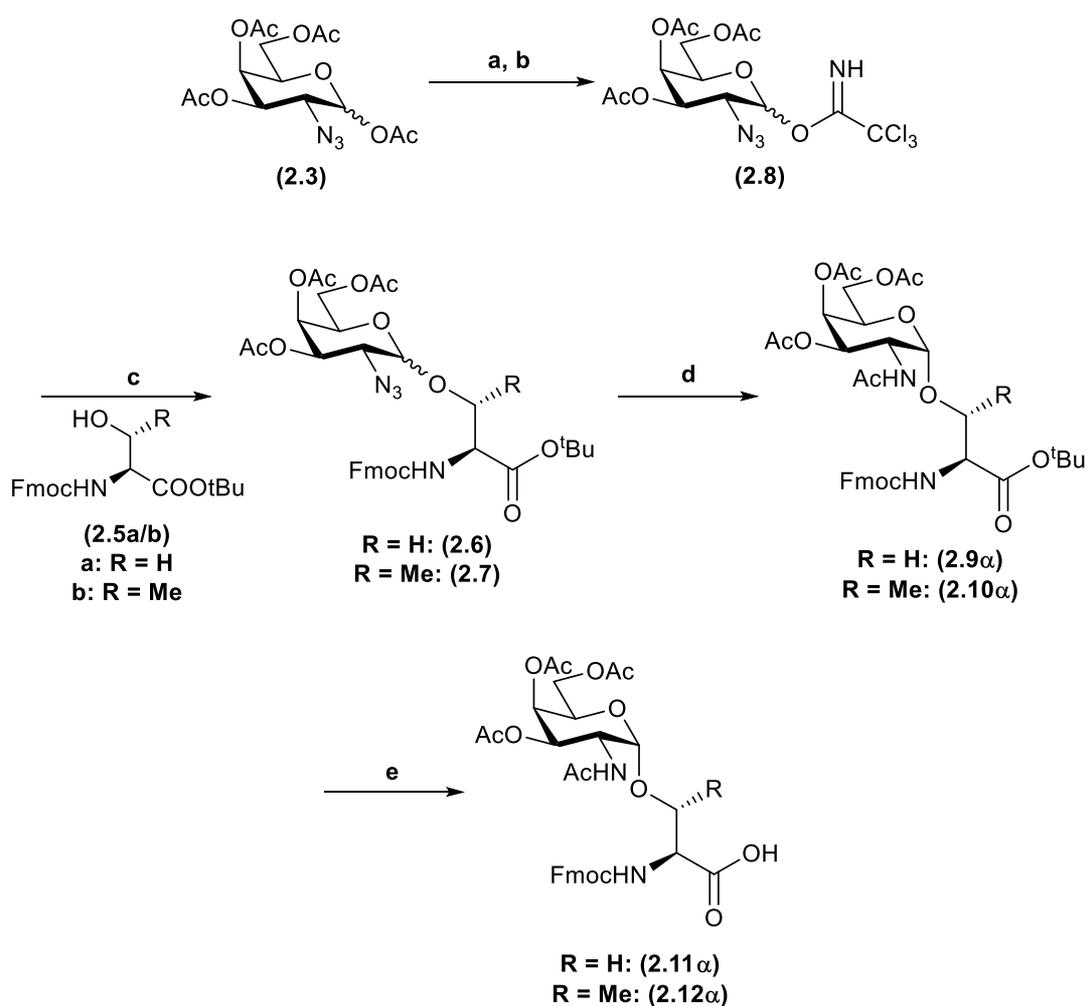
An investigation of the literature suggested the alpha anomer of the Fmoc and tert-butyl protected serine/threonine (**2.6/2.7**) could be synthesised selectively *via* a galactosyl iodide donor (**2.4**) (Scheme 2-2).¹⁹² Field *et al.* propose that in the glycosylation reaction iodine activates the glycosyl iodide. When the resulting triiodide is eliminated, an equilibrium between the alpha and beta forms of the glycosyl iodide is established; as the β -anomer is not stabilised by the anomeric effect, and hence is more reactive, alpha selectivity is observed.¹⁹² By TLC, the reaction of the protected galactosamine (**2.3**) in the presence of iodine and HMDS to produce the galactosyl iodide intermediate (**2.4**) was successful, but attempts to isolate the glycosyl iodide (**2.4**) following the literature procedure were unsuccessful resulting in degradation.¹⁹² Attempts to use the galactosyl iodide without purification by transferring the reaction mixture directly into the glycosylation reaction were also unsuccessful. It is possible that the excess HMDS, not removed from the prior step, protected the hydroxyl group of the amino acid with a TMS group preventing the glycosylation reaction from proceeding.



*Scheme 2-2: Attempted alpha-selective glycosylation route via galactosyl iodide.*¹⁹²

As an alternative approach, glycosylation *via* a trichloroacetimidate was attempted (Scheme 2-3).¹⁹³ Selective anomeric deacetylation using hydrazine acetate in DMF at

60 °C followed by activation of the anomeric hydroxyl using trichloroacetonitrile in the presence of DBU yielded the trichloroacetimidate galactosyl donor (**2.8**) in a yield of 56% (10:1 α/β). Glycosylation of the protected serine/threonine acceptor (**2.6/2.7**) with the trichloroacetimidate donor was performed at -30 °C in a 1:1 mixture of dichloromethane and diethyl ether to favour formation of the desired α anomer. The trichloroacetimidate donor was activated with a catalytic quantity of TMSOTf. At completion of the reaction TMSOTf was quenched with excess *N,N*-diisopropylethylamine rather than excess triethylamine to avoid removal of the Fmoc protecting group.¹⁸⁸



*Scheme 2-3: Synthesis of GalNAc- α -Serine/Threonine building blocks: (a) $H_4N_2 \cdot AcOH$, DMF; (b) Cl_3CCN , DBU, DCM, $\alpha/\beta=10/1$ (56%, 2 steps); (c) TMSOTf, DCM, Et_2O , $-30^\circ C$, $\alpha/\beta=1/1$; (d) Zn, $CuSO_4$, Ac_2O , $AcOH$, THF (**2.9 α** : 39%, **2.10 α** : 40%, 2 steps); (e) 95% TFA in H_2O (**2.11 α** : 74%, **2.12 α** : 92%).*

It was not possible to separate the α/β anomers by flash column chromatography at this point and the mixture of anomers was therefore carried through to the one-pot reduction of the 2-azido group and concomitant *N*-acetylation (Scheme 2-3). Zinc dust suspended in 2% (*w/v*) aqueous copper sulfate solution was added to the crude glycosylation product (**2.6/2.7**) dissolved in a 3:2:1 mixture of THF, AcOH and Ac₂O. Within 2 hours the azido group was reduced, and it was possible to separate the anomers by flash column chromatography. Over two steps the alpha glycosylation products were isolated in a 39% yield for the serine (**2.9 α**) and 40% yield for the threonine derivative (**2.10 α**). The beta anomers were also isolated in a 38% yield for the serine derivative (**2.9 β**) and 33% for the threonine derivative (**2.10 β**).

The tert-butyl ester protecting group was efficiently cleaved with 95% (aq) TFA at room temperature to yield the peracetylated and Fmoc protected α -*O*-GalNAc serine (**2.11 α**) and threonine (**2.12 α**) amino acids in a 74% and 92% yield, respectively, following a final purification through a short plug of silica. The *O*-GalNAc amino acid building blocks were synthesised on a multi-gram scale over seven steps with an overall yield of 10% for the serine derivative (**2.11 α**) and 13% for the threonine derivative (**2.12 α**) from D-galactosamine (**2.2**). The acetyl groups were retained to improve the solubility of the amino acid building blocks in DMF and prevent undesired side coupling during solid phase peptide synthesis.

2.1.2 Solid Phase Glycopeptide Synthesis of (Tn)MUC1 Fragments

Standard Fmoc procedures were followed for solid phase glycopeptide synthesis and a Rink Amide MBHA resin with a low loading capacity (0.33-0.35 mmol/g) was chosen as the solid support to prevent aggregation of the growing glycopeptide chains and improve the coupling efficiency of the bulky GalNAc amino acid building blocks (**2.11 α /2.12 α**). Attempts to synthesise the entire 20 amino acid tandem repeat sequence using an CEM Liberty BlueTM automated peptide synthesiser were unsuccessful, resulting in a complex mixture of peptides no bigger than 10 amino acids in length. The high number of proline residues, and the coupling of two prolines in tandem, is hypothesised to have resulted in poor coupling efficiencies.

A literature search revealed two immunologically relevant regions of the MUC1 tandem repeat domain where antibodies identified from the sera of breast cancer

patients are known to bind to incompletely glycosylated MUC1.^{194–197} The binding domain of these two antibodies (SM3 and H-2D^B) were between six and nine amino acids in length, and more reasonable to produce by manual SPPS (Figure 2-1). During the course of this work, a report by Polonskaya *et al.* noted that higher titers of high-affinity anti-glycan antibodies were raised if the glycans/glycopeptide antigens were conjugated in a well-defined repeatable manner and with the minimal length linker.⁶⁵ To achieve this the peptides (**2.13** & **2.15**) and glycopeptides (**2.14** & **2.16–2.18**) were all synthesised with a minimal length linker composed of an N-terminal glycine-valine to allow for C-terminal sortase mediated ligation.

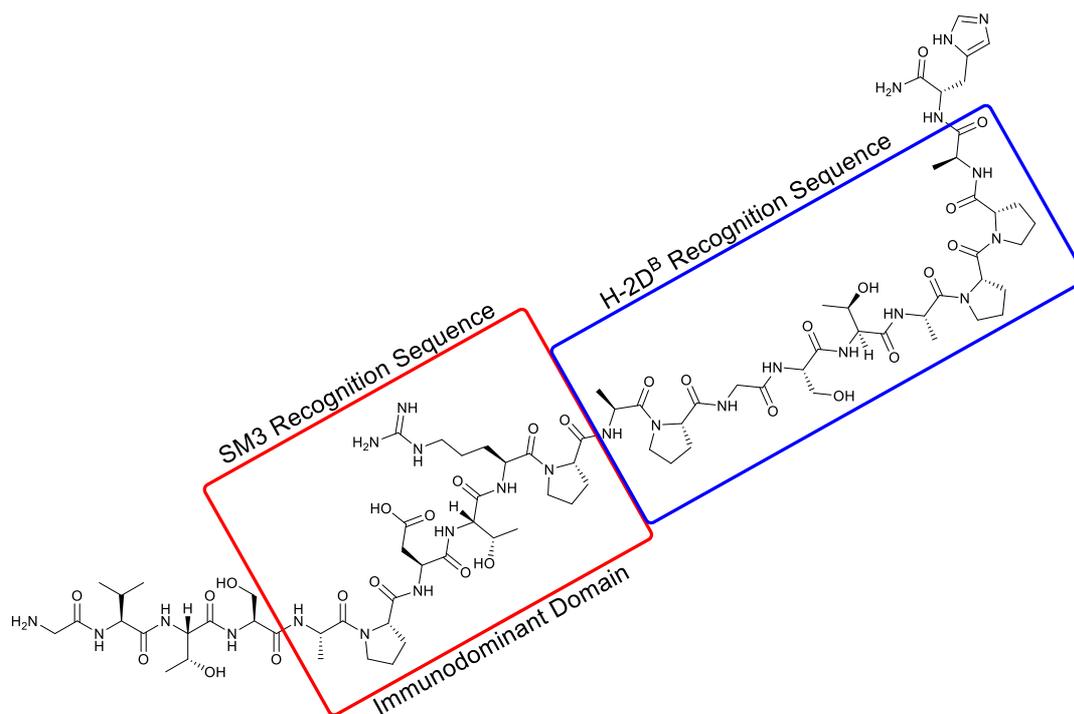
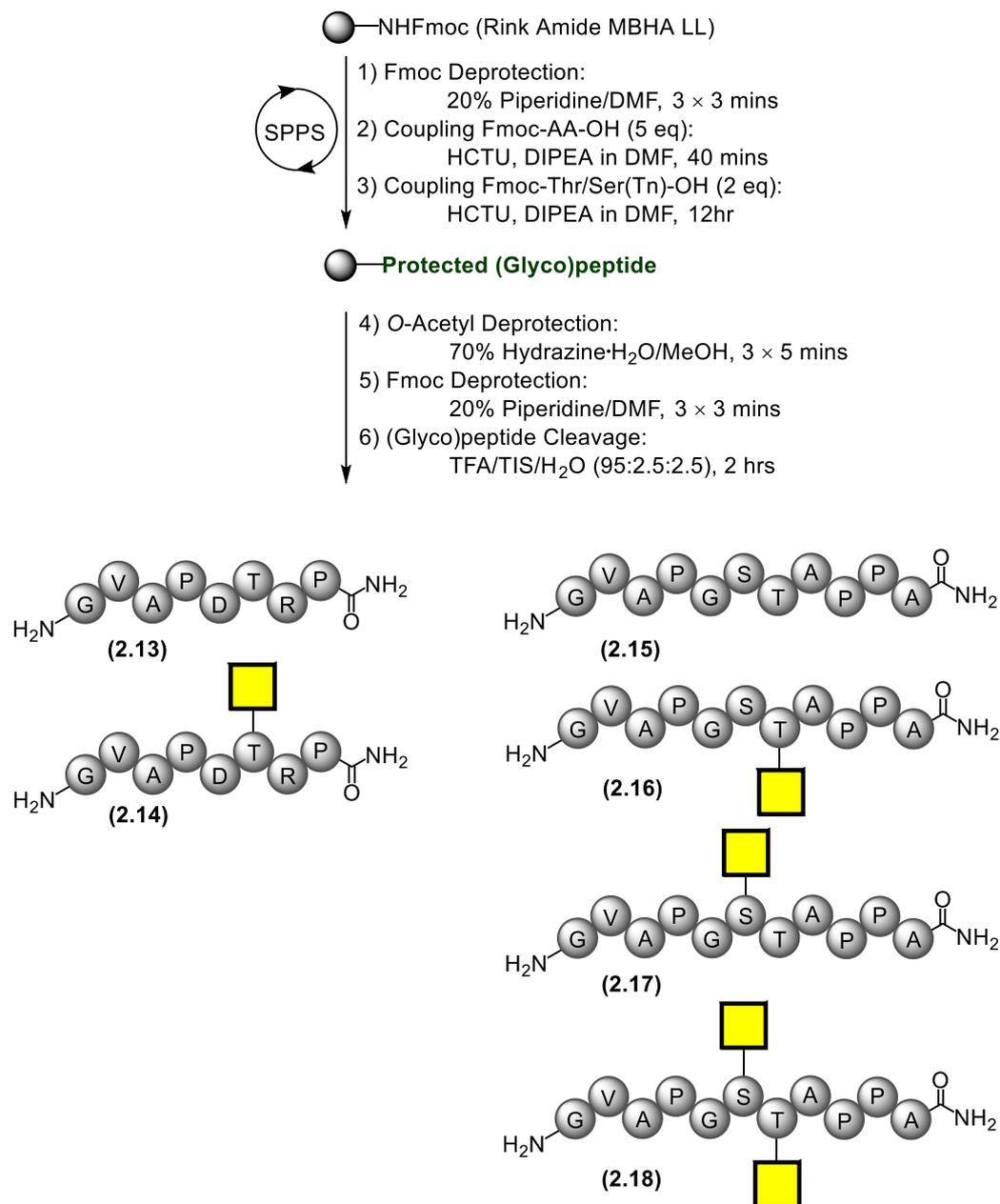


Figure 2-1: 20 amino acid tandem repeat sequence of MUC1 extracellular domain with the SM3 recognition sequence circled in red and the H-2D^B sequence circled in blue.

The couplings of commercial amino acids were performed using 5 equivalents for 40 minutes; couplings of the synthetic GalNAc serine (**2.11 α**) and/or threonine (**2.12 α**) were performed using 2 equivalents for 12 hours (Scheme 2-4). Couplings were repeated to ensure a high coupling efficiency. To reduce off-resin manipulation of the glycopeptides, all non-acid labile protecting groups were removed on resin. The acetyl groups were efficiently removed in 15 minutes using 70% hydrazine hydrate in MeOH followed by Fmoc deprotection using 20% piperidine in DMF, with all the excess reagents and waste products removed by filtration. Cleavage of the

glycopeptides from the resin and global side chain deprotection was achieved using standard TFA cleavage conditions (95% TFA, 2.5% TIS, 2.5% H₂O). Cleavage was not allowed to continue for more than 2 hours, to reduce acid-catalysed cleavage of the glycosidic linkage.



Scheme 2-4: Synthetic route used for all glycopeptides with all deprotection steps performed on resin.

The cleaved glycopeptides were isolated by precipitation with ice-cold ether followed by lyophilisation. All the glycopeptides (**2.14**, **2.16-2.18**) were produced in good yields and analytical HPLC, LC-MS and HRMS revealed that none of the glycopeptides required further purification by prep-HPLC. The non-glycosylated

peptides (2.13, 2.15) were also produced *via* the same procedures (*O*-acetyl deprotection with hydrazine hydrate was not required).

2.2 Preparation of Protein Labelling Stocks

2.2.1 Preparation of CTB with C-terminal Sortase Sequence

Cholera toxin B-subunit (CTB) was chosen as the model scaffold protein for glycopeptide conjugation. CTB glycoconjugates could have potential applications ranging from conjugate vaccines to discovery of antibody-like mimetics by phage display. Previous work in the Turnbull lab by Dr Matthew Balmforth demonstrated that binding of CTB to a GM1 coated microtiter plate allowed effective binding of Affimer phage particles to the top face of the pentamer during phage display (Figure 1-5).^{198,199} Furthermore, CTB has been shown to induce a mucosal immune response following oral/nasal immunisation leading to enhanced antigen presentation by various APCs.^{80,200} The immune response is significantly improved if the antigen is covalently coupled to CTB due to increased uptake across the mucosal barrier and presentation by dendritic cells, macrophages and naïve B cells.^{80,201} CTB is approved by the FDA and has shown promise as a mucosal adjuvant.⁸⁰

2.2.1.1 CTB-LPETGA Prepared by Ammonium Sulfate Precipitation

A plasmid (pSAB2.2-LPETGA) which encodes cholera toxin B-subunit with a C-terminal LPETGA sortase recognition motif (CTB-LPETGA) was prepared by Darren Machin (University of Leeds). CTB-LPETGA was expressed with a periplasmic signalling peptide which directs the protein to the periplasm, where the signal peptide is cleaved and the protein folds into its mature pentameric form, before being secreted into the media.

CTB-LPETGA was expressed in *E. coli* C41. The cells were initially grown in LB media at 37 °C. Once an OD = 0.6 was achieved, protein expression was induced using isopropyl thiogalactoside (IPTG) and left overnight at 25 °C. The protein was isolated from the growth media by ammonium sulfate precipitation and purified by Ni-NTA affinity chromatography. Despite the absence of a His-tag on the protein, the

surface exposed His13 and His94 residues have been found to promote coordination to Ni²⁺ allowing for immobilisation of the pentameric protein on Ni-NTA resin.²⁰² Elution of CTB-LPETGA was confirmed by SDS-PAGE (Figure 2-2a); the protein is sufficiently stable to remain pentameric under the denaturing conditions of SDS-PAGE, unless the sample is first boiled at 95 °C to denature the protein into its protomer form. Due its compact form, pentameric CTB-LPETGA migrates on the gel with an apparent mass lower than its expected MW (~62.8 kDa), however upon boiling of the sample the protomer (denatured) form migrates consistent with its MW (~12.6 kDa). Any remaining impurities following Ni-NTA affinity chromatography were removed by size exclusion chromatography using a Superdex[®] S200 (16/60) column (Figure 2-2b). Approximately 13.5 mg of CTB-LPETGA was isolated per 1 L of cultured cells.

Characterisation of the protein by SDS-PAGE and deconvoluted ES-MS revealed a truncation present in approximately 10% of the CTB protomer units (Figure 2-2c) resulting in eleven amino acids from the C-terminus being absent including the LPETGA sortase motif (Figure 2-3). Investigations to eliminate truncation of the protein using protease inhibitors in both the resuspension buffer and growth media were unsuccessful and so it was concluded that an alternative construct would be required for quantitative conjugation to the pentamer.

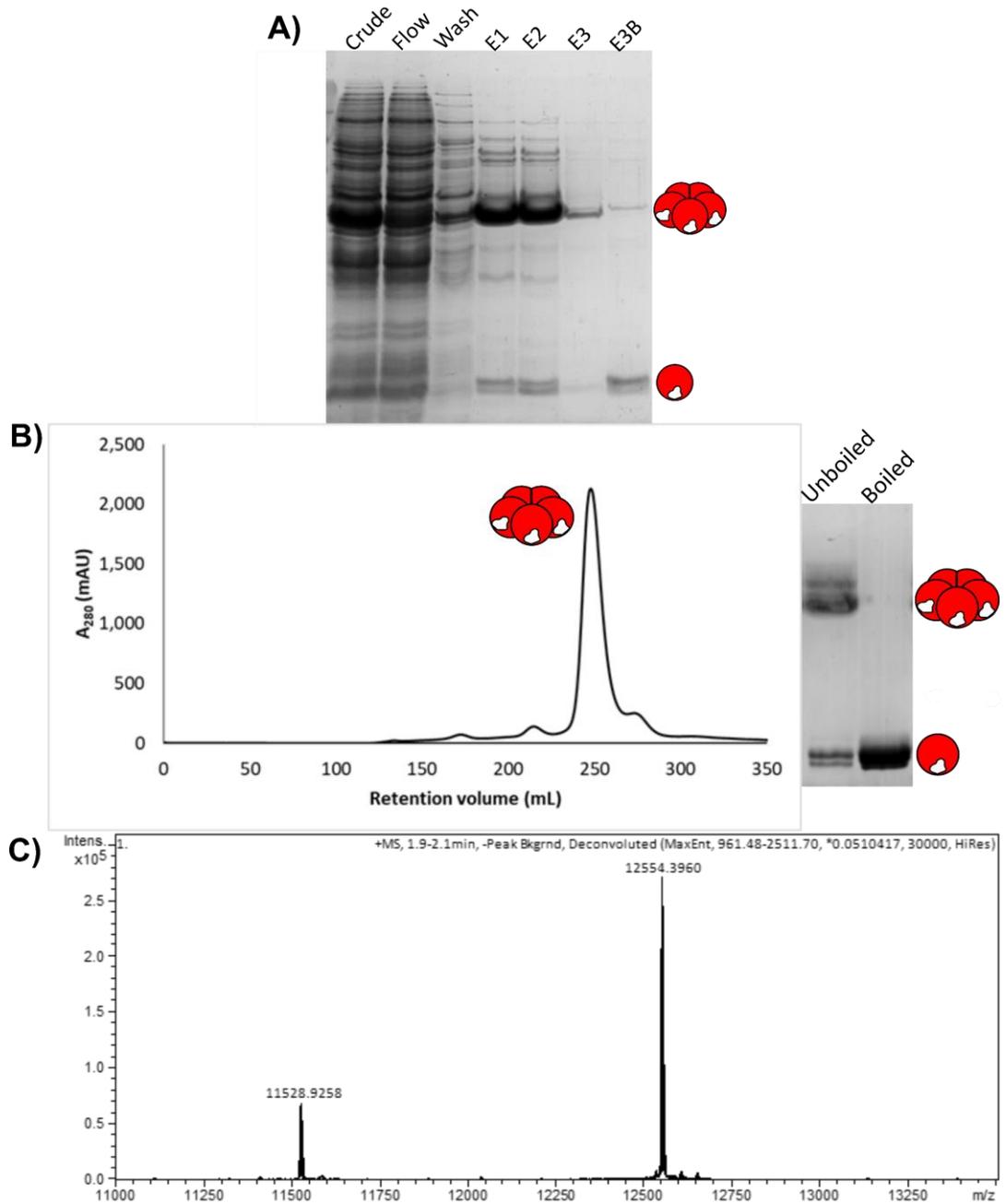


Figure 2-2: Purification of CTB-LPETGA for use in labelling reactions. a) SDS-PAGE gel of Ni-NTA affinity chromatography flow through, wash and elution fractions (E1-3). b) SEC trace for CTB-LPETGA using Superdex[®] S200 16/60 with SDS-PAGE gel showing formation of a stable pentamer which dissociates upon boiling. Two bands appear for protomeric CTB showing a mixture of full length and truncated material. c) Deconvoluted ES-MS of purified CTB-LPETGA protomer (12554.4 Da) and the truncated CTB protomer (11528.9 Da).

2.2.1.2 Refolding of Cholera Toxin B-subunit

An alternative plasmid for cytosolic expression of CTB with a C-terminal sortase tag was prepared by Dr Gemma Wildsmith (University of Leeds). Through three rounds of site directed mutagenesis, the N-terminal periplasmic targeting sequence was deleted, and a methionine ‘start’ codon and a His-tag following the LPETGA sortase motif were inserted to create the pGW-CTB-LPETGAS-H₆ plasmid which encodes CTB-LPETGAS-His₆ (hereafter termed CTB-His₆) (Figure 2-3). As it lacked the periplasmic targeting sequence, CTB-His₆ encoded by this plasmid would not be exported into the periplasm for folding; instead packaged into inclusion bodies (insoluble aggregates of stable protein) which require isolation, denaturing and refolding *in vitro*.

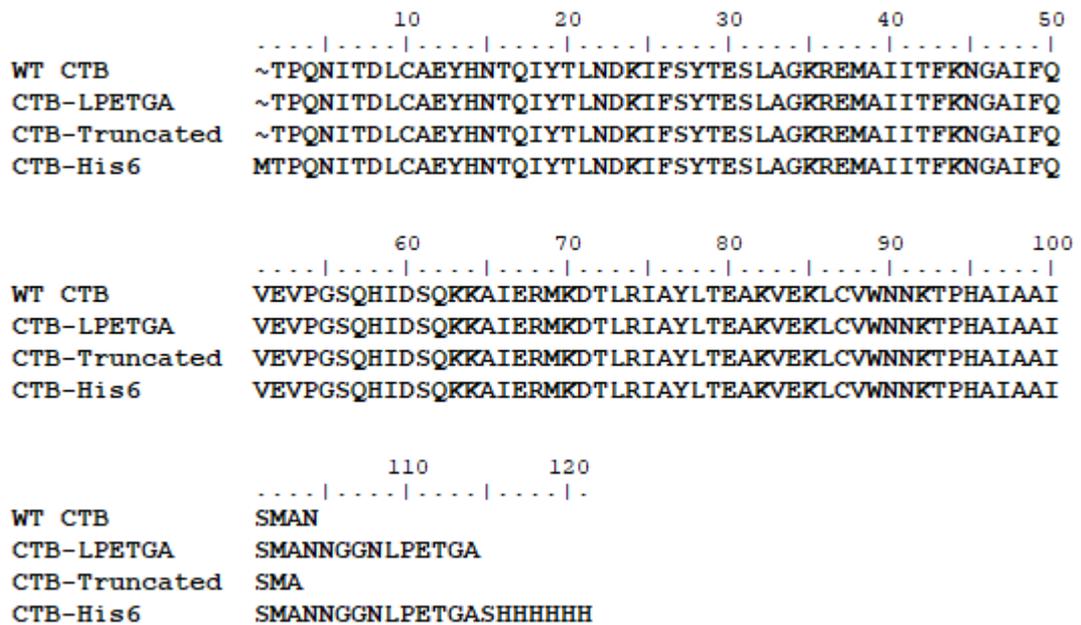


Figure 2-3: Sequence alignment of expressed CTB mutants compared to WT CTB (*El Tor*). CTB-LPETGA contains C-terminal (NGGNLPETGA) ten amino acid extension. CTB-truncated is absent the eleven amino acid extension, resulting in same sequence as WT CTB. CTB-His₆ contains N-terminal methionine and C-terminal SHHHHHH compared to CTB-LPETGA.

CTB-His₆ was expressed in *E. coli* BL21 in LB media. Cells were initially grown at 37 °C, and once an OD = 0.6 was achieved, protein expression was induced using IPTG, and left overnight at 25 °C. Following cell lysis, the inclusion bodies were recovered by centrifugation and resuspended in denaturing buffer (8 M urea, 100 mM NaH₂PO₄, 10 mM Tris, pH 8.0). The inclusion body lysate was purified by Ni-NTA

affinity chromatography, and the eluate was reduced with TCEP (Figure 2-4a). After 2 hours, excess TCEP was removed using a PD-10 desalting column and the eluate was concentrated to at least 10 mg/mL before being added slowly to the chilled arginine refolding buffer containing cysteine and cystine to assist the formation of the intramolecular disulfide bond. CTB-His₆ was left to refold at 4 °C for 3-5 days, and then dialysed against Tris-buffered saline, and purified by size exclusion chromatography using a Superdex[®] S75 (26/60) column to remove mis-folded protein (Figure 2-4b). Characterisation of the refolded CTB-His₆ by ES-MS showed a single species was present (Figure 2-4c). Approximately 12.3 mg of refolded CTB-His₆ was isolated per 1 L of cultured cells.

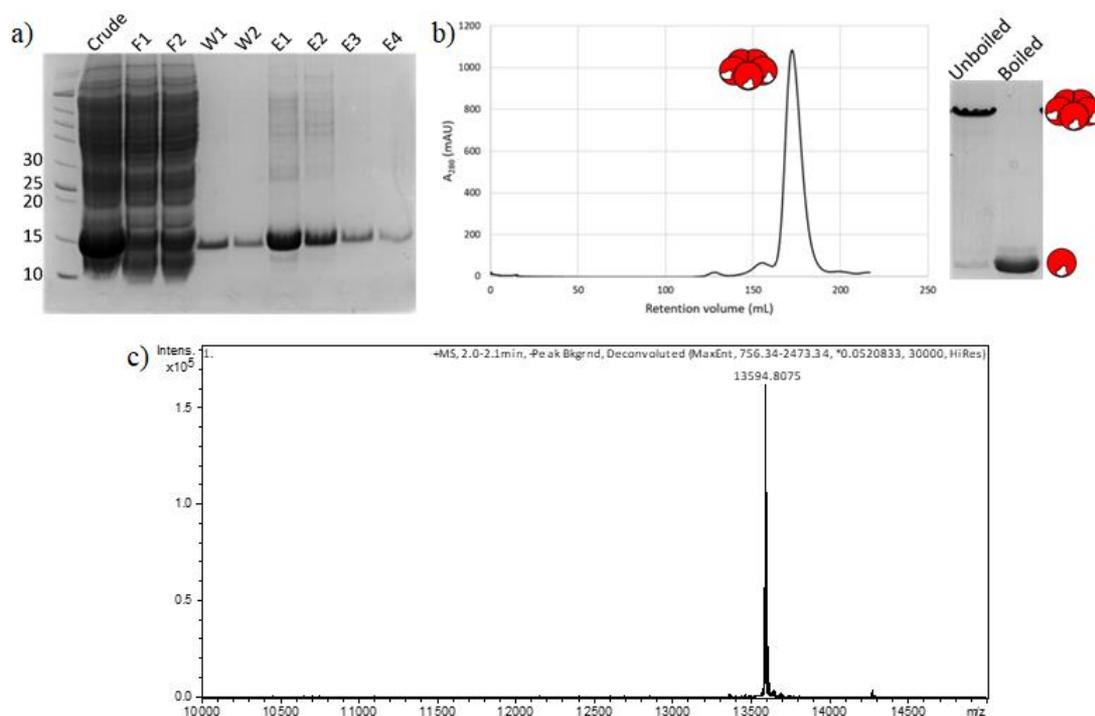


Figure 2-4: Purification of CTB-His₆ for use in labelling reactions. a) SDS-PAGE gel of Ni-NTA affinity chromatography flow through (F1-2), wash (W1-2) and elution fractions (E1-4). b) SEC trace for CTB-His₆ using Superdex[®] S75 26/60 with SDS-PAGE gel showing formation of a stable pentamer which dissociates upon boiling. c) Deconvoluted ES-MS of purified CTB-His₆ (13594.8 Da).

Comparison of the refolded CTB-His₆ and ammonium sulfate precipitated CTB-LPETGA by SDS-PAGE and size exclusion chromatography (Superdex[®] S200 16/60) showed CTB-His₆ had successfully formed a stable pentamer which, as

expected, was slightly larger than CTB-LPETGA due to the addition of the C-terminal his tag (Figure 2-5).

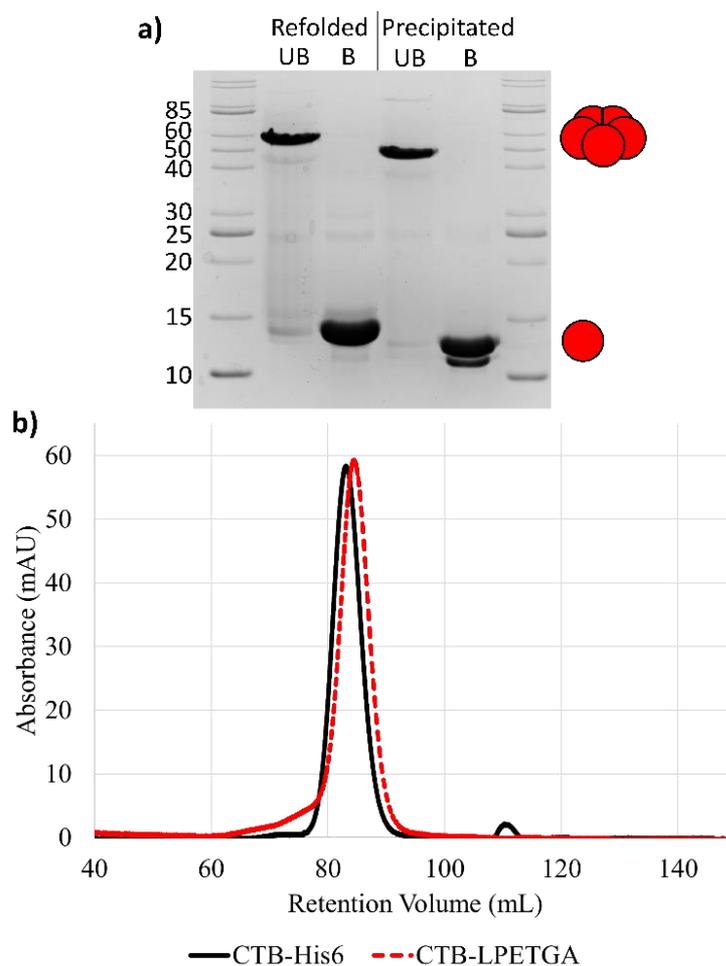


Figure 2-5: a) SDS-PAGE comparison of refolded CTB-His₆ and CTB-LPETGA showing formation of stable pentamers which dissociate upon boiling prior to being loaded onto the gel. (UB: Unboiled, B: Boiled). b) Size exclusion (Superdex S200 16/600) comparison showing shorter retention time of larger CTB-His₆ compared to CTB-LPETGA.

2.2.2 Preparation of Sortase 7M

The pET30b-7MSrtA plasmid for the expression of sortase 7M with a C-terminal His-tag was produced by Dr Tomasz Kaminski (University of Leeds) and transformed into *E. coli* BL21 gold (DE3) competent cells. Sortase 7M was expressed in auto-induction media (AIM) by initially incubating at 30 °C for 5 hours before being left at 25 °C overnight. Following cell lysis, sortase 7M-His₆ was purified first by Ni-NTA affinity

chromatography then by size-exclusion chromatography using a Superdex[®] S75 (26/60) column isocratically eluted with Tris-buffered saline at pH 7.2 (Figure 2-6).

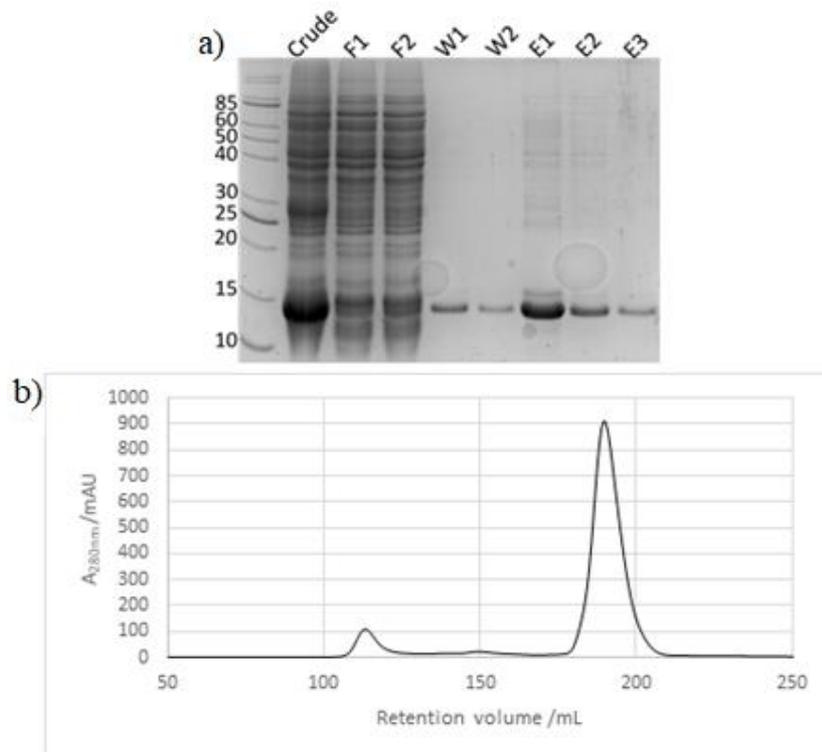


Figure 2-6: Purification of sortase 7M-His₆ for use in labelling reactions. a) SDS-PAGE gel of Ni-NTA affinity chromatography column flow through (F1), wash (W1-2) and elution fractions (E1-3). b) SEC trace for sortase 7M-His₆ using Superdex[®] S75 26/60.

2.3 Conjugation of Glycopeptides to C-terminus of CTB

With all conjugation components in hand, the test sortase ligation of CTB-LPETGA could be performed using 15 equivalents of peptide/glycopeptide (per protomer) and 15 mol% sortase 7M. Reactions were followed using SDS-PAGE by taking timepoint samples at regular intervals and boiling at 95 °C for 5 minutes in SDS loading buffer to stop the ligation reaction and denature the CTB protein. Samples were run on a 15% Tris-Glycine SDS-PAGE gel, stained with Coomassie blue and the relative quantity of conjugated product and unlabelled starting material was quantified using densitometry (Bio-Rad image lab software). Although the ammonium precipitated CTB contained a truncation, the molecular weight difference upon conjugation of a peptide or glycopeptide was sufficient to observe band separation by SDS-PAGE. A

larger peptide is released for the refolded version of CTB (GAS-H₆) and as a result the molecular weight difference upon ligation is too small to result in effective separation by SDS-PAGE.

The initial test ligations were performed using peptide **2.13** and glycopeptide **2.14** which were based on the immunodominant domain of MUC1 and recognised by the mAb SM3 (APDTRP).²⁰³ The first reaction using peptide **2.13** showed a band shift indicating an increase in mass. The conjugated protomer of CTB accounted for 78% of all conjugatable CTB protomers after 2 hours. The reaction involving glycopeptide **2.14** reached a conjugation level of 53%; approximately 25 percentage points lower than that of the non-glycosylated peptide **2.13** (Figure 2-7).

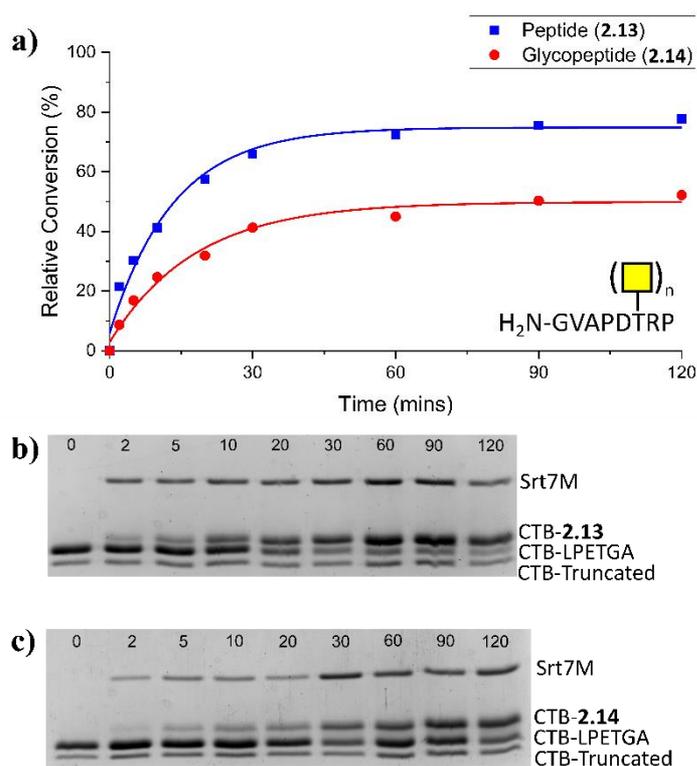


Figure 2-7: a) Time-course of Srt7M-mediated ligation of CTB-LPETGA and peptide (2.13) or glycopeptide (2.14). Percentage conversion calculated based on relative intensity of conjugated CTB and unlabelled CTB-LPETGA. c) SDS-PAGE gel of glycopeptide (2.14) timepoint samples. Data fitted in OriginPro 2019b (version 9.6.5) using an exponential function.

Conjugation of the refolded CTB-His₆ to the H-2D^B MUC1 peptide & glycopeptides (**2.15-2.18**) was performed using identical conditions. After 2 hours the degree of conjugation was determined by ES-MS. The data showed a similar trend to that

already seen during time course studies of the SM3 (glyco)peptides (**2.13** and **2.14**), with the highest level of conjugation seen for peptide **2.15** (Table 2-1). Marginally poorer conjugation efficiencies were seen for the mono-glycosylated glycopeptides **2.16** and **2.17**. The worst conjugation was seen for the doubly glycosylated glycopeptide **2.18**. This trend suggested that the presence of the GalNAc moiety had a dramatic effect on sortase ligation efficiency. It was hypothesised that the GalNAc moiety reinforced an intramolecular hydrogen bond which caused the peptide to adopt a conformation that occluded the N-terminus. In this state the peptide/glycopeptide would be unable to access the active site and act as the nucleophile making the reverse reaction and hydrolysis pathway more favourable. As the number of GalNAc moieties increases, the opportunities for hydrogen bonding also increase and as a result the glycopeptide may spend more of its time in a state which is unable to participate in the ligation reaction resulting in lower level of conjugation.

*Table 2-1: Relative levels of unlabelled, conjugated and hydrolysis for the small-scale sortase mediated ligation with peptides **2.15** and glycopeptides **2.16–2.18** after 2 hours (by ES-MS). Peptide **2.15** sequence: H₂N-GVAPGSTAPPA. The position of the GalNAc moiety is shown in brackets for each glycoconjugate.*

Conjugate	Unlabelled (%)	Conjugated (%)	Hydrolysis (%)
CTB- 2.15	22	50	28
CTB- 2.16 (Thr7)	24	40	36
CTB- 2.17 (Ser6)	27	45	28
CTB- 2.18 (Thr7, Ser6)	0	10	90

2.3.1 Effect of Adding Chaotropic Agent on Conjugational Efficiency

To overcome putative hydrogen bonding within the peptide & glycopeptide substates, the sortase reaction time-courses for peptide **2.13** and glycopeptide **2.14** were repeated with 10% (v/v) DMSO added to the buffer. This concentration of DMSO was expected to be sufficient to reduce the level intramolecular hydrogen bonding without unfolding or significantly reducing the catalytic activity of sortase. The concentration of sortase, CTB and the peptide/glycopeptide were kept constant to compare the effect of adding DMSO as a chaotropic agent. SDS-PAGE of the timepoints showed conjugation with peptide **2.13** in the presence of 10% DMSO resulted in a negligible

improvement of three percentage points, from 78% to 81% (Figure 2-8). While this small improvement could be the result of breaking hydrogen bonds or due to the reduction on concentration of water which would reduce the rate of hydrolysis, this small improvement is more likely within the boundaries of experimental error. Without performing the experiments in triplicate the significance or cause of an improvement this small cannot be predicted. From the SDS-PAGE gels the level of hydrolysis cannot be estimated as the difference in molecular weight between CTB-LPETGA and CTB-LPET is too small to result in two separate bands on an SDS-PAGE gel.

The ligation of glycopeptide **2.14** in the presence of 10% DMSO resulted in an increase of nineteen percentage points in the levels of glycopeptide conjugation from 53% to 72% after 2 hours. This reduced the difference in conjugation between peptide **2.13** and glycopeptide **2.14** from twenty-five percentage points in the absence of DMSO to six percentage points in the presence of 10% DMSO. The level of glycopeptide conjugation was still below what would be optimal for future applications and more work would be required to optimise conjugation to at least 80% especially for the double glycosylated glycopeptide **2.18**.

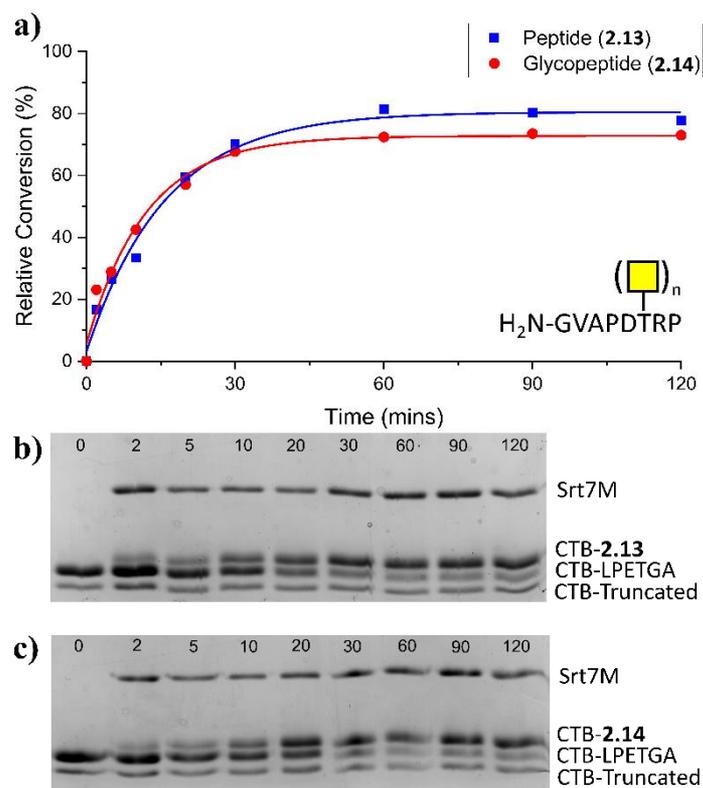


Figure 2-8: a) Time-course of Srt7M-mediated ligation of CTB-LPETGA and peptide (2.13) or glycopeptide (2.14) in buffer spiked with 10% DMSO. Percentage conversion calculated based on relative intensity of conjugated CTB and unlabelled CTB-LPETGA. b) SDS-PAGE gel of peptide (2.13) timepoint samples. c) SDS-PAGE gel of glycopeptide (2.11) timepoint samples. Data fitted in OriginPro 2019b (version 9.6.5) using an exponential function.

2.3.2 Investigation into Secondary Structure of MUC1 Epitopes

It was noted that both sequences shared a similar N-terminal motif of GVAPXS/T (where X is any amino acid). This led us to hypothesise that Pro4 may induce a β -hairpin-like structure in the peptides/glycopeptides which may also impair conjugation by sterically hindering the nucleophilic glycine and preventing it from participating in the reaction. Any secondary structure had to be transient in nature due to the levels of peptide and glycopeptide conjugation observed. As already shown, the conjugation of glycopeptides can be improved through the addition of a chaotropic agent (10% DMSO) further bolstering the idea a transient secondary structure was present which in the glycopeptide substrates was reinforced by an intramolecular hydrogen bonding. To investigate if the proline residue had an influence on

conjugational efficiency, a variant of the SM3 peptide (**2.13**) and glycopeptide (**2.14**) was synthesised in which Pro4 was changed to Ala (Figure 2-9).

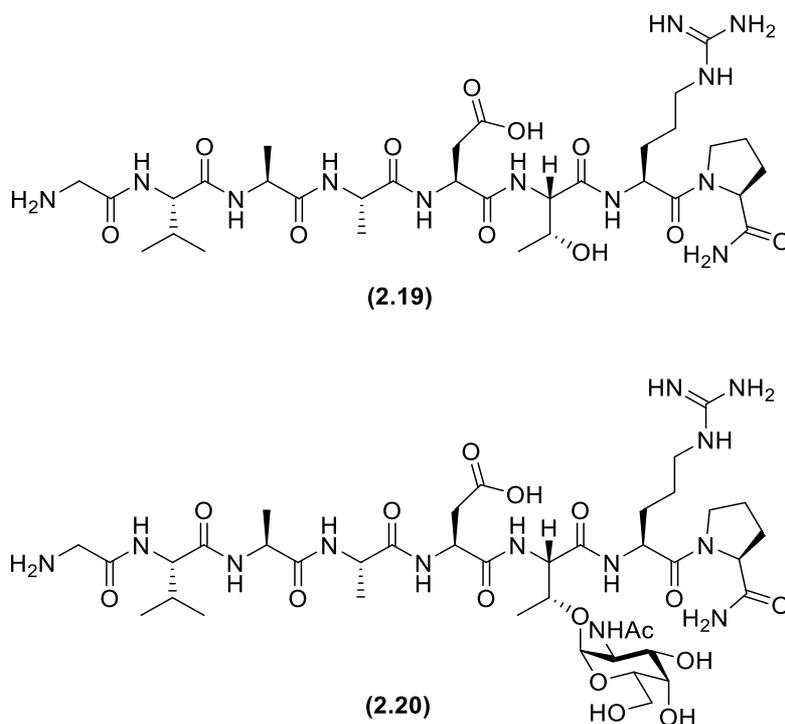


Figure 2-9: Structure of the Pro4Ala peptide (2.19) and glycopeptide (2.20). By changing proline to alanine any backbone induced secondary structure should have been removed.

The Pro4Ala peptide (**2.19**) and glycopeptide (**2.20**) were tested within the sortase ligation reaction using the same conditions as the initial reactions (no DMSO). After 1.5 hours quantitative conjugation was achieved for peptide **2.19** (Figure 2-10). Conjugation of CTB-LPETGA with glycopeptide **2.20** reached a maximum level of 86% at 1.5 hours (Figure 2-10). Both reactions were a significant improvement over that for the natural sequence and confirmed Pro4 had a significant impact on conjugational efficiency. It was believed some level of intramolecular hydrogen bonding was still taking place within glycopeptide **2.20** and ligation with this glycopeptide could be quantitative if the reaction was spiked with 10% DMSO, however an experiment to confirm this was not performed.

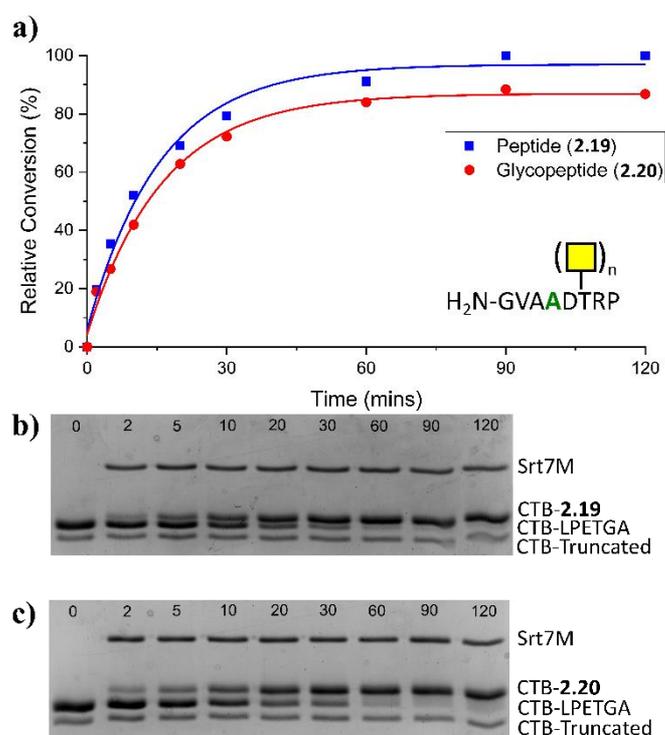


Figure 2-10: a) Time-course of Srt7M-mediated ligation of CTB-LPETGA and P4A mutant peptide (2.19) or glycopeptide (2.20). Percentage conversion calculated based on relative intensity of conjugated CTB and unlabelled CTB-LPETGA. b) SDS-PAGE gel of peptide (2.19) timepoint samples. c) SDS-PAGE gel of glycopeptide (2.20) timepoint samples. Data fitted in OriginPro 2019b (version 9.6.5) using an exponential function.

Having confirmed that Pro4 influenced conjugational efficiency, a combination of 1D and 2D NMR experiments were used in an attempt to characterise the extent of the secondary structure. It was hoped that 2D NOESY NMR would provide insight into specific interactions between the Gly-Val sortase tag and the rest of the peptide allowing the optimal minimal linker length to be predicted. Although significant peak broadening was seen in the proton NMR spectrum of peptide 2.13 at 278 K compared to at 298 K, suggesting a slowing of the rate of exchange between conformations in the peptide at 278 K compared to 298 K; The 2D NOESY NMR spectrum recorded at 278 K showed no significant through space interactions between the Gly-Val sortase tag and the rest of the peptide. All interactions through space interactions observed were within the natural sequence of the immunodominant domain of MUC1 (Figure 2-11).

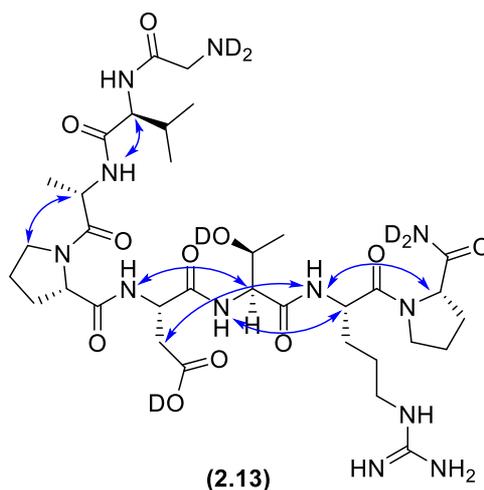


Figure 2-11: Structure of peptide 2.13 showing the strong through space interactions (NOEs) identified at 278 K.

2.3.3 Effect of Increased Linker Length on Conjugational Efficiency

Previously the aim was to keep the length of linker between the glycopeptide and protein to a minimum; however, achieving higher levels of conjugation whilst maintaining the secondary structure of the MUC1(Tn) epitopes would be vital in developing an effective immune response or antibody-like proteins which were capable of distinguishing between healthy MUC1 and under-glycosylated MUC1(Tn) seen in epithelial carcinomas. The effect of introducing linkers of various sizes between the sortase handle (Gly-Val) and the MUC1 sequence of interest was investigated. Increasing the flexibility of the sortase linker was also tested by changing the Gly-Val sortase handle to a traditional triglycine. Glycopeptide **2.14** was resynthesized with these N-terminal modifications to produce the four glycopeptides (Figure 2-12). The PEG linker chosen was 8-amino-3,6-dioxaoctanoic acid (AEEAc-OH), equivalent to the insertion of PEG₃ and is approximately ~12 Å in length and commercially available in the Fmoc protected form.

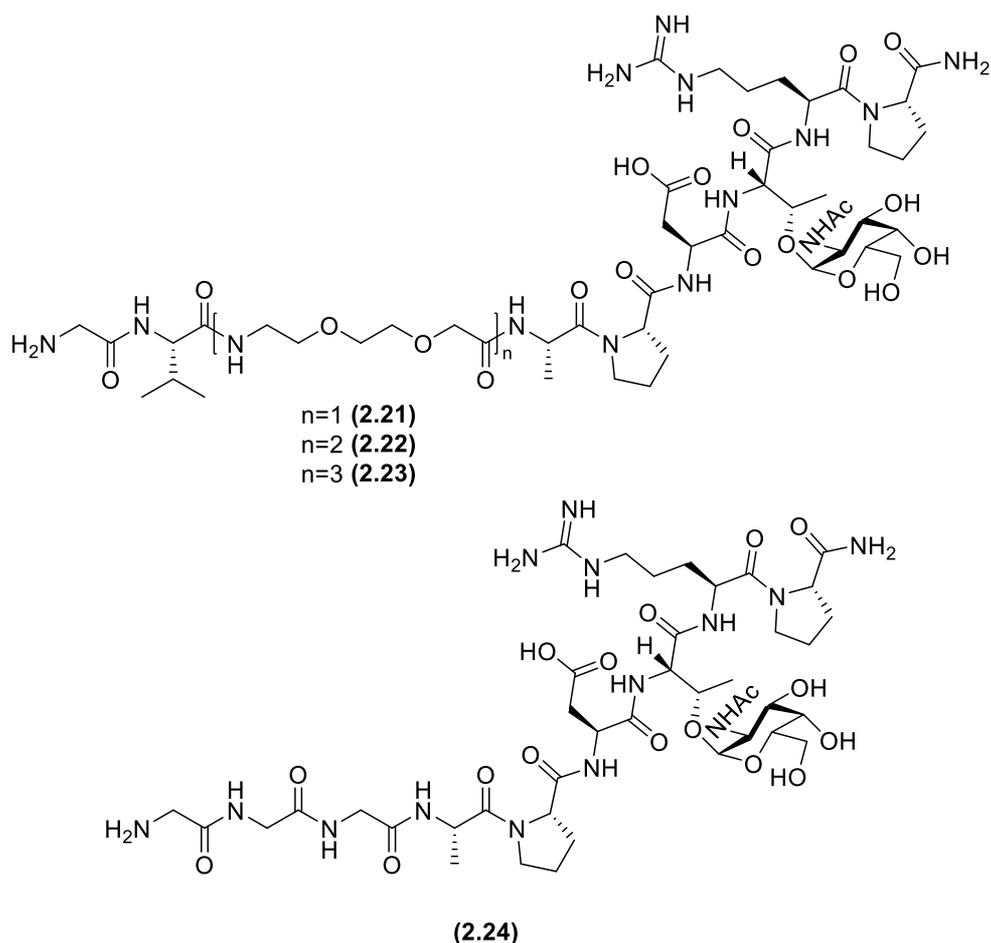


Figure 2-12: Structure of the PEGylated glycopeptides (2.21-2.23) and glycopeptide with traditional N-terminal triglycine linker (2.24).

The test sortase ligation of CTB-LPETGA with the PEGylated glycopeptides (2.21-2.23) and triglycine glycopeptide (2.24) was performed in the presence of 10% DMSO. The buffer was spiked with DMSO having previously shown its addition had a positive effect on the conjugational efficiency. As previously predicted, increasing the flexibility of the N-terminus by adding a flexible linker between the immunodominant domain of MUC1 and the N-terminal sortase tag (Gly-Val) resulted in an increase in the levels of conjugation. Addition of the PEG₃ linker (2.21) resulted in an increase in conjugation from 72% to 87% of the available C-terminal sortase sites (Figure 2-13b). Addition of two or three PEG units (2.22, 2.23) resulted in an increase in conjugation from 72% to 92-93% (Figure 2-13c,d). Although introduction of the PEG linker increases the overall level of conjugation, the initial rate of ligation is slightly reduced as the length of the PEG linker increases. For glycopeptide 2.24 with the N-terminal triglycine, the level of conjugation increased from 72% to 84% after 2 hours (Figure 2-13e). This improvement in conjugation is most likely due to

the increased flexibility of the N-terminus rather than the additional length provided by the addition of a single amino acid (glycine).

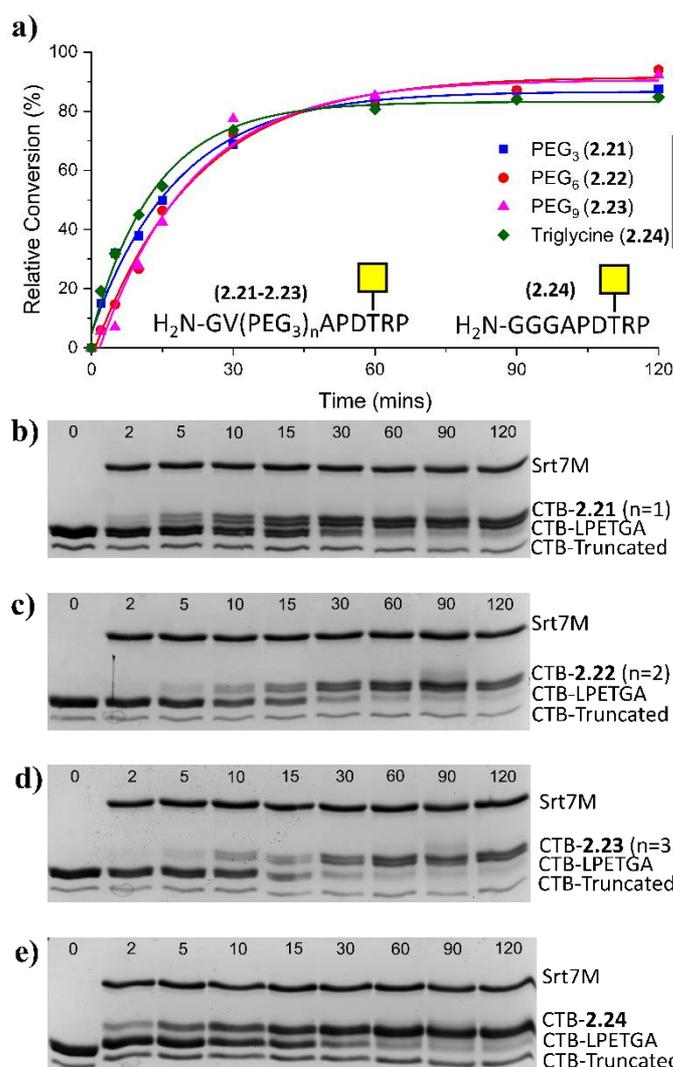


Figure 2-13: a) Time-course of Srt7M-mediated ligation of CTB-LPETGA and PEGylated and triglycine glycopeptide substrates (2.21-2.24). Percentage conversion calculated based on relative intensity of conjugated CTB and unlabelled CTB-LPETGA. b) SDS-PAGE gel of PEG₃ glycopeptide (2.21) timepoint samples. c) SDS-PAGE gel of PEG₆ glycopeptide (2.22) timepoint samples. d) SDS-PAGE gel of PEG₉ glycopeptide (2.23) timepoint samples. e) SDS-PAGE gel of triglycine glycopeptide (2.24) timepoint samples. Data fitted in OriginPro 2019b (version 9.6.5) using an exponential function.

Based on these results, the peptides and glycopeptides based on the mAbs SM3 (APDTRP, immunodominant domain) and H-2D^B (APGSTAPPA) recognition motifs were resynthesized with double AEEAc linker, equivalent to PEG₆ (~24 Å), between

the peptide/glycopeptide and the N-terminal sortase tag (Gly-Val) (Figure 2-14). This length of linker showed the highest levels of SM3 glycopeptide conjugation observed throughout optimisation whilst retaining the proline residue which may have an influence of backbone secondary structure; increasing linker length further, equivalent to PEG₉, showed no additional improvement in conjugational efficiency.

H-2D^B Recognition Sequence

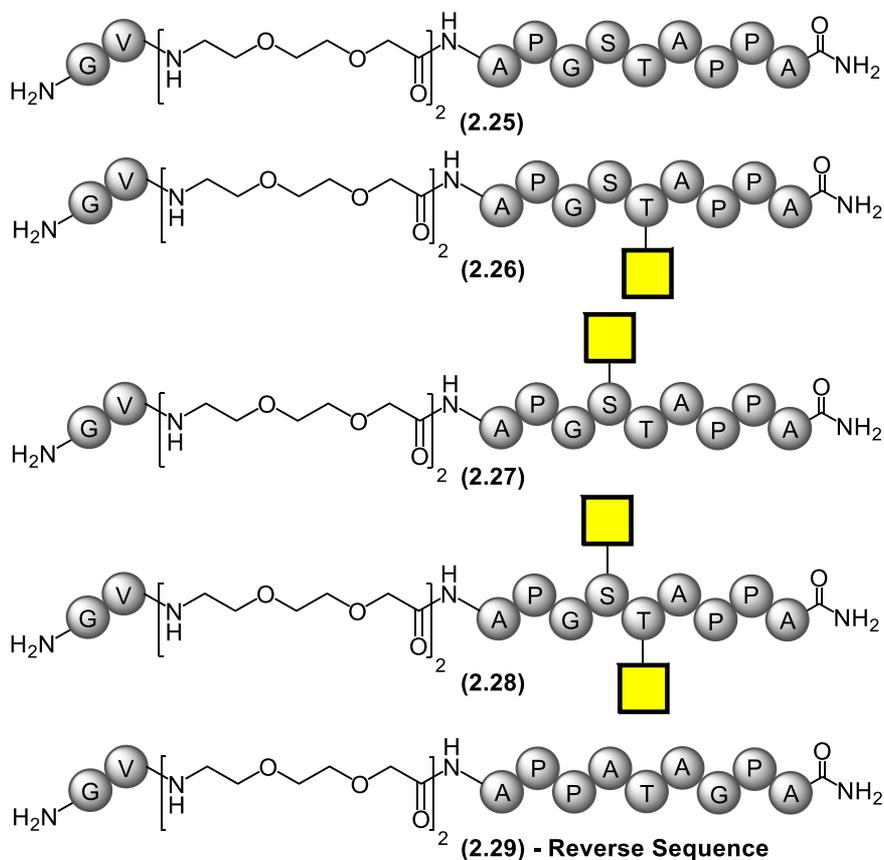


Figure 2-14: Structure of the optimised PEGylated peptides & glycopeptides (2.22, 2.25-2.30) contained PEG₆ equivalent linker.

2.4 Is the Refolded CTB better for C-terminal Ligation with Sortase 7M?

With optimised conditions and peptide/glycopeptide constructs in hand, test ligations of CTB-LPETGA using the optimised sortase conditions were performed. All peptides (**2.25, 2.29, 2.30**) and glycopeptides (**2.22, 2.26-2.28**) showed high levels of C-terminal ligation between 79% and 86% (of available sortase sites) (Figure 2-15, Table 2-2). As discussed previously approximately 10% of the C-terminal sortase extensions in CTB-LPETGA are missing due to a truncation.

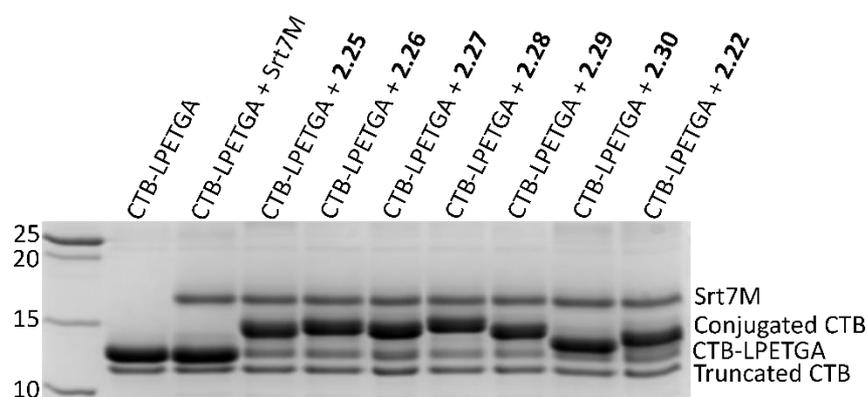


Figure 2-15: SDS-PAGE gel showing sortase ligation after 2 hours using PEG₆ peptides (**2.23, 2.29, 2.30**) and glycopeptides (**2.22, 2.26-2.28**). Levels of conjugation listed in Table 2-2.

Table 2-2: Percentage of conjugated and unlabelled sortase sites on CTB-LPETGA after 2 hours. Densitometry performed by comparing intensity of unlabelled (CTB-LPETGA) and conjugated CTB bands. CTB without the LPETGA tag (CTB) accounts for ~10% of each sample.

Conjugate	Unlabelled (%)	Conjugated (%)
CTB- 2.25	19	81
CTB- 2.26 (Thr7)	21	79
CTB- 2.27 (Ser6)	21	79
CTB- 2.28 (Thr7, Ser6)	18	82
CTB- 2.29 (Reverse)	15	85
CTB- 2.30	14	86
CTB- 2.22 (Thr6)	15	85

Ligation was repeated using the refolded CTB-His₆ which did not have any C-terminal truncation. Reactions were performed using identical conditions as previously used, and conjugational efficiency was determined from the deconvoluted ES-MS. Unlike for ligation of the precipitated CTB-LPETGA, little or no consistency was seen in the percentage conjugation obtained from the deconvoluted ES-MS (Table 2-3). Deconvoluted ES-MS of all the reactions showed minimal (<5%) or no hydrolysis product. Inconsistency in the obtained percentage conjugation could be due to the method of analysis. Mass spectrometry is not inherently quantitative and differences in the unlabelled, conjugated, and hydrolysed CTB may mean each has a different propensity to fly or be ionised within the spectrometer. SDS-PAGE & densitometry could not be used to determine level of protein conjugation as a larger peptide is released for the refolded version of CTB (GAS-H₆); as a result, the molecular weight difference between unlabelled and conjugated protein was too similar to result in effective separation on an SDS-PAGE gel.

Table 2-3: Percentage conjugation of CTB-LPETGA-His₆ after 2 hours, determined from deconvoluted ES-MS.

Conjugate	Conjugation (%)
CTB- 2.25	20
CTB- 2.26 (Thr7)	90
CTB- 2.27 (Ser6)	50
CTB- 2.28 (Thr7, Ser6)	45
CTB- 2.30	55
CTB- 2.22 (Thr6)	65

Although the conjugation appears inconsistent, compared to the non-PEGylated H-2D^B glycopeptides (**2.16-2.18**), an improvement in conjugation between 18 and 66 percentage points was observed using the optimised linker and conditions. It was hoped the additional 10% of available sortase sites in the refolded CTB-His₆ would make up for the lower level of conjugation.

2.5 Chapter Conclusions

Following literature procedures, two peracetylated *O*-GalNAc amino acids (**2.11a/2.12a**) were synthesised and incorporated into solid phase peptide synthesis to generate a library of Mucin1 glycopeptides based on two known antibody recognition sequences. The peptides/glycopeptides synthesised exhibited all potential combinations of *O*-GalNAc glycosylation. Two cholera toxin B-subunit mutants were expressed from *E. coli* both harbouring a C-terminal LPETGA tag for sortase ligation. The first expressed into the media *via* a periplasmic targeting sequence, was found to have a C-terminal extension resulting in the absence of the sortase recognition sequence on 10% of protomers. An alternative mutant was expressed in the absence of the periplasmic targeting sequence and refolded from inclusion bodies. Both proteins were shown to form stable pentamers which dissociate upon boiling.

Sortase ligation was initially found to be poor for glycopeptide substrates, with higher numbers of GalNAc moieties having a more negative impact on conjugational efficiency. The addition of DMSO as chaotropic agent resulted in an improvement in glycopeptide conjugation which, within margin of error, was comparable to the conjugation with the equivalent peptide substrate. Furthermore, proline at position 4 has been shown to induce conformational changes within the peptide/glycopeptide which has a negative effect on conjugational efficiency. Addition of a PEG₆ equivalent linker between the Gly-Val sortase handle and the glycopeptide sequence was found to improve conjugation whilst retaining the proline residue in the sequence. Using the optimised conditions and linker length, high levels of C-terminal conjugation of CTB-LPETGA was demonstrated for all the peptides (**2.25, 2.29, 2.30**) and glycopeptides (**2.22, 2.26-2.28**) substrates with conjugation between 79% and 86% of the available sortase sites observed by SDS-PAGE and densitometry.

Chapter 3: Biotinylation of Multivalent Proteins for Phage Display

Having discussed the optimisation of the C-terminal sortase ligation of CTB with a series of biologically relevant MUC1(Tn) glycopeptides in chapter 2; chapter 3 will explore if such glycoconjugates can be used to discover novel antibody-like binding proteins, Affimers, capable of selectively recognising MUC1(Tn) and distinguishing between the different glycosylation patterns. This chapter will also explore preliminary work towards detecting antibody-like binding proteins that bind to verotoxin B-subunit and could be used for targeted drug delivery.

3.1 Background

Antibodies are widely used in diagnostic devices for the detection of specific biological molecules; however, they can have limited stability. As a result, there has been huge interest in development of smaller proteins that mimic antibody function. The Affimer scaffold, developed in Leeds, is one of many artificial binding protein scaffolds generated through the alignment of 57 phycocystatin sequences.²⁰⁴ The Affimer scaffold displays all the desirable characteristics for a peptide presenting scaffold; it's small, monomeric, highly soluble and exhibits high stability with regard to temperature and pH.^{205,206} Affimers consist of a β -sheet composed of four anti-parallel strands which curve around an α -helix forming a half-closed palm (Figure 3-1). For the Affimer phage display library, each of the two loops in the parent scaffold were replaced with a randomised group of nine amino acids to generate a library containing approximately 1.3×10^{10} different coding sequence variants.^{205–207} Manipulation of the triplet codons in the loops ensured no cysteines or stop codons were introduced. Affimers are versatile reagents with uses ranging from the detection of small molecules or proteins, modulation of protein function *in vitro* and *in vivo*, *in vivo* tumour labelling reagents, drug delivery conjugates and co-crystallisation chaperones.^{206,208–210}

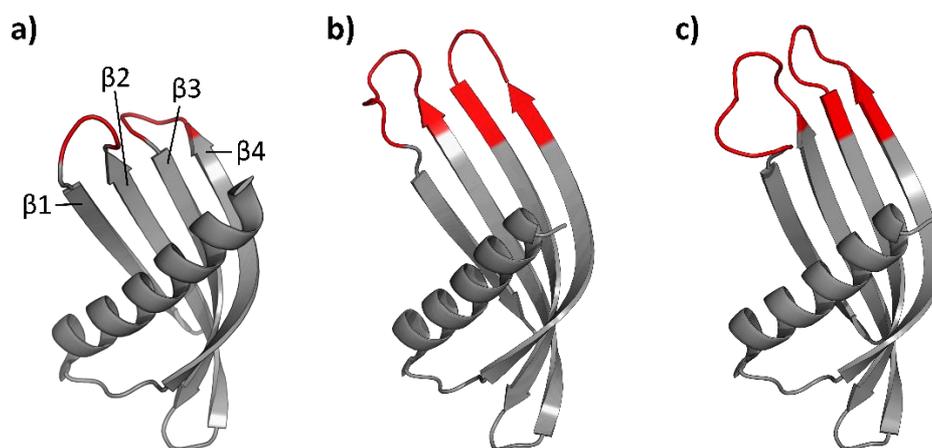


Figure 3-1: Crystal structures of three Affimer, demonstrating common cystatin scaffold. Variable residues shown in red. a) Adhiron/Affimer Scaffold (PDB: 4N6T).²⁰⁵ b) Affimer raised against BCL-2 (PDB: 6ST2).²¹¹ c) Affimer raised against Alpha-actinin-2 (PDB: 6SWT).²¹²

The initial plan to identify binders against the (Tn)MUC1 glycopeptide epitopes, was to immobilise the CTB-glycoconjugates described in the previous chapter onto GM1-coated wells of a microtitre plate for phage display screening. Previously this approach was used successfully by Dr Matthew Balmforth (University of Leeds) to identify a series of Affimers which are capable of binding selectively with the top face of cholera toxin B-subunit.¹⁹⁹ However, limits on laboratory access as a result of the Covid-19 pandemic meant that the glycoconjugates had instead to be biotinylated so they could be screened using the standard high-throughput methodology employed within the BioScreening Technology Group (BSTG) at the University of Leeds.

3.2 Dual N- and C-terminal Labelling of CTB-LPETGA

3.2.1 N-terminal Biotinylation

Random biotinylation using NHS-biotin is often the preferred method of biotinylation of proteins for phage display as it enables screening of the protein in an array of orientations. For our screening experiments random biotinylation was not preferable; CTB has nine lysine residues per protomer distributed across the protein (Figure 3-2), and thus random biotinylation would result in a distribution of biotin molecules across the two faces, including the top face where the C-terminus was to be conjugated with the (Tn)MUC1 glycopeptides. In contrast, the N-termini of the CTB protomers are

found in between the GM1 binding sites and point towards the opposite face of the protein from the C-termini (Figure 3-2). Site-specific modification of the N-termini should place CTB in a favourable orientation on the microtitre plate to present the glycopeptides conjugated the C-terminus to the Affimer-phage particles.

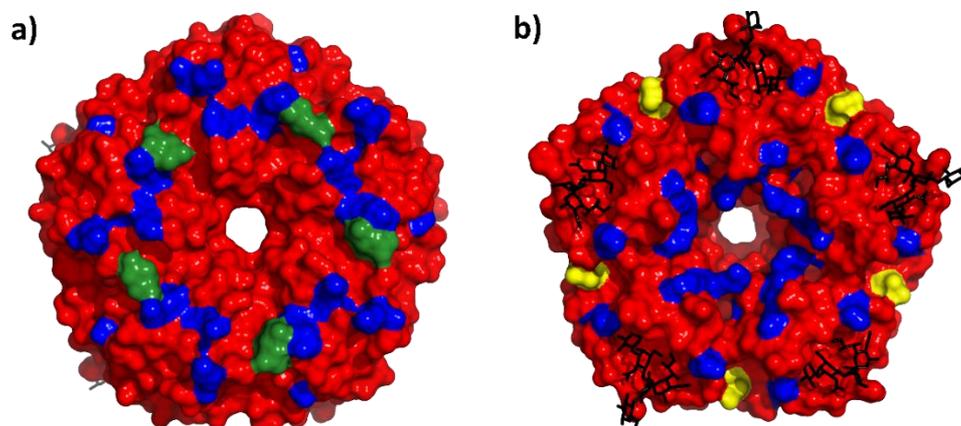
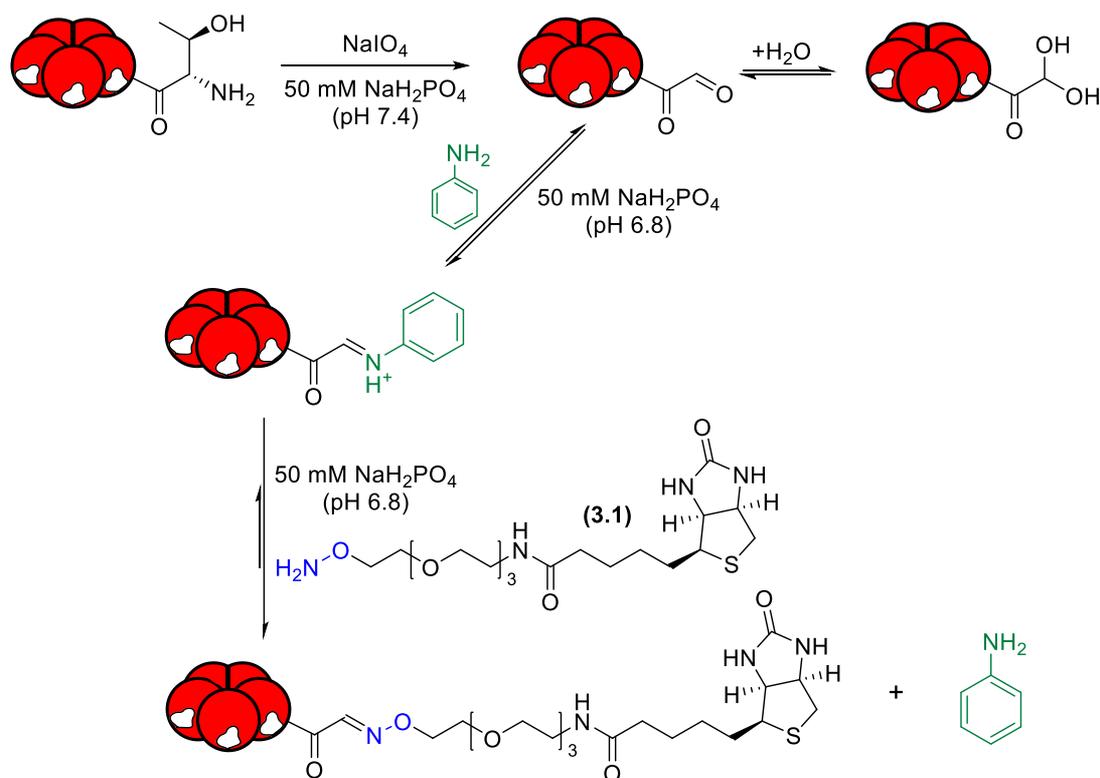


Figure 3-2: Surface representation of CTB showing GM1os (black) bound in the GM1 binding sites on the bottom face with N-termini (yellow) located between GM1 binding sites on the bottom face. C-terminus coloured green. Surface exposed lysine residues coloured blue. a) Top face. b) Bottom face. (PDB: 3CHB).

Oxime ligation to the N-terminal of CTB was first described by Rose *et al.* for the construction of well-defined CTB-viral peptide immunogens.¹⁶⁶ Matthew Balmforth (University of Leeds) previously optimised the oxime-mediated N-terminal biotinylation of El Tor CTB with an aminooxy PEG₄ biotin (**3.1**) and showed biotinylation of the N-termini with this linker does not lead to aggregation of the protein. Furthermore, he demonstrated that the oxime linker was not broken during the stringent washing conditions which occur during phage display screening. CTB-LPETGA was used for dual N- and C-terminal labelling despite 10% of the sample containing a C-terminal truncation. Although the refolded CTB could be used instead, the N-terminal methionine would first need to be removed using a methionyl aminopeptidase to uncover an N-terminal threonine which in turn could be oxidized and used for oxime ligation. As shown in chapter 2, it is easier to quantify the progress of conjugation reactions using the non-refolded CTB due to the greater increase in molecular weight upon conjugation which allows SDS-PAGE and densitometry to be used instead of mass spectroscopy.



Scheme 3-1: Oxidation of N-terminal Thr of CTB using sodium periodate to form an aldehyde which reversibly hydrates in aqueous media. The aldehyde reacts with aniline to form an iminium intermediate which is rapidly displaced by the aminooxy functionalised PEG₄ biotin (3.1) to form an oxime linkage.

Oxidation of CTB-LPETGA was performed using five equivalents of sodium periodate in the presence of ten equivalents of L-methionine. The reaction was monitored by ES-MS and typically found to be complete within 10 minutes, with both species showing a mass loss of 27 Da (Figure 3-3b). In aqueous solution aldehydes are prone to hydration and the 27 Da mass loss can be attributed to its hydrated form (Scheme 3-1). It is essential that phosphate-buffered saline is not used for this oxidation reaction. During the course of this work, researchers in our lab, and collaborators from the University of York, noticed this oxidation reaction would fail to reach completion if performed in the presence of potassium ions,¹⁶³ which are part of the standard recipe for phosphate-buffered saline. This is likely due to the poor solubility of potassium periodate in water.¹⁶³

The oxidised CTB-LPETGA was separated from the excess periodate, L-methionine, and the acetaldehyde byproduct using G-25 mini-trap desalting column, during simultaneous buffer exchange into sodium phosphate buffer at pH 6.8. The desalted

oxidised CTB-LPETGA was combined with the aniline catalyst,¹⁰⁸ ten equivalents of aminoxy PEG₄ biotin (**3.1**), and left overnight to react at 37°C. Due to the immiscibility of aniline with the aqueous buffer, the mixture must be mixed by vortex for aniline to become dispersed in solution, mixing by pipette or inverting the tube is not sufficient to disperse aniline into solution. Deconvoluted ES-MS confirmed both CTB-LPETGA and its truncation had been quantitatively N-terminally biotinylated with a mass increase of 398 Da from the oxidised CTB-LPETGA (hydrate) or oxidised CTB (hydrate) (Figure 3-3c). The mass of biotin-CTB-LPETGA was calculated as 12924.53 Da, and mass of biotin-CTB was calculated as 11900.26 Da. The biotinylated CTB-LPETGA was purified by PD-10 desalting column and size-exclusion chromatography used to ensure the pentameric structure had been retained (Figure 3-4).

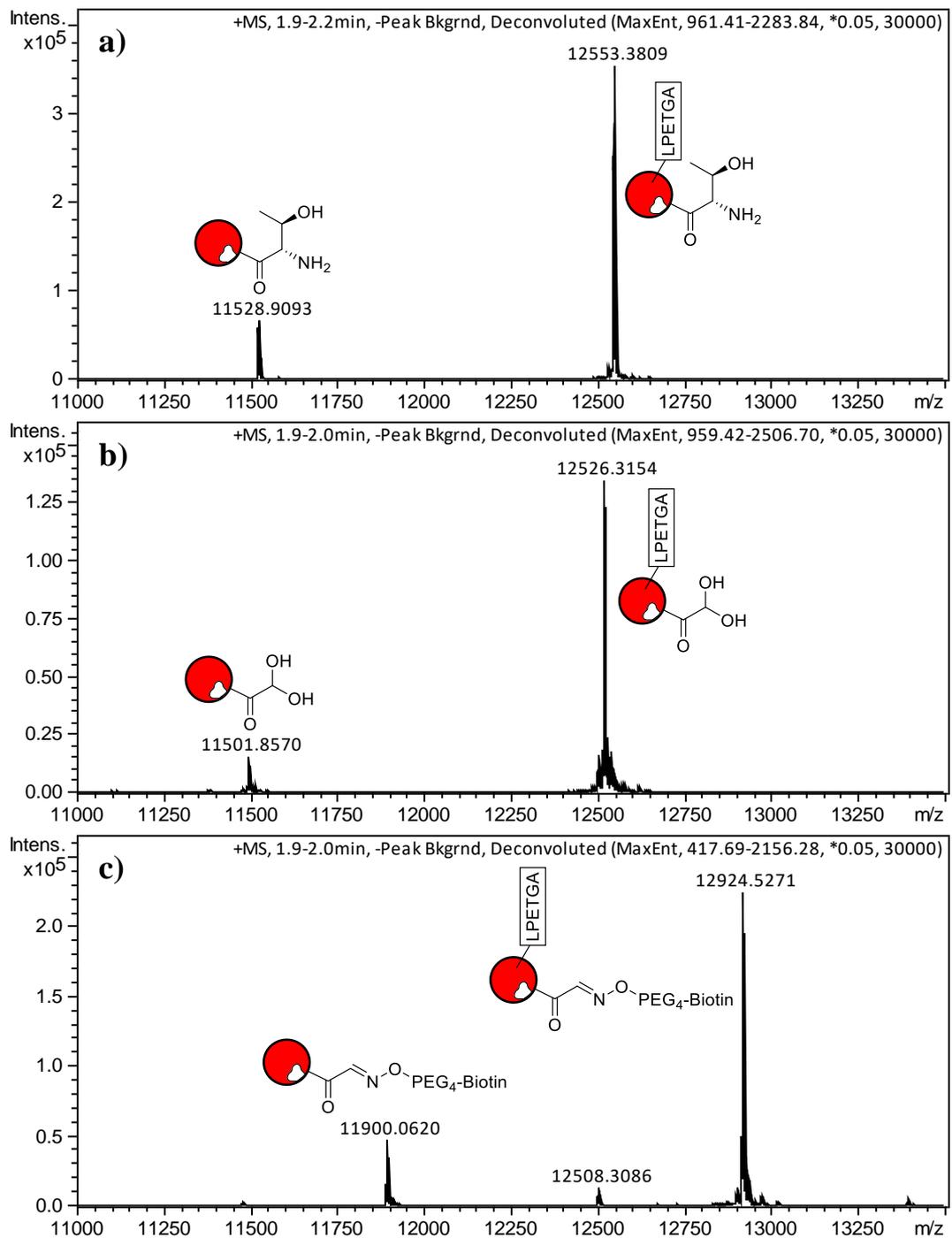


Figure 3-3: Deconvoluted high resolution protein mass spectrometry of: a) CTB-LPETGA before oxidation with NaIO₄. b) CTB-LPETGA oxidation after 10 minutes showing quantitative oxidation as hydrate. c) Oxidised-CTB-LPETGA biotinylation after 16hrs, showing near quantitative N-terminal biotinylation. A small amount of oxidised-CTB-LPETGA that has cyclised to a diketopiperidine is present at 12508.3086 Da. CTB without the LPETGA tag (CTB) accounts for ~10% of each sample.

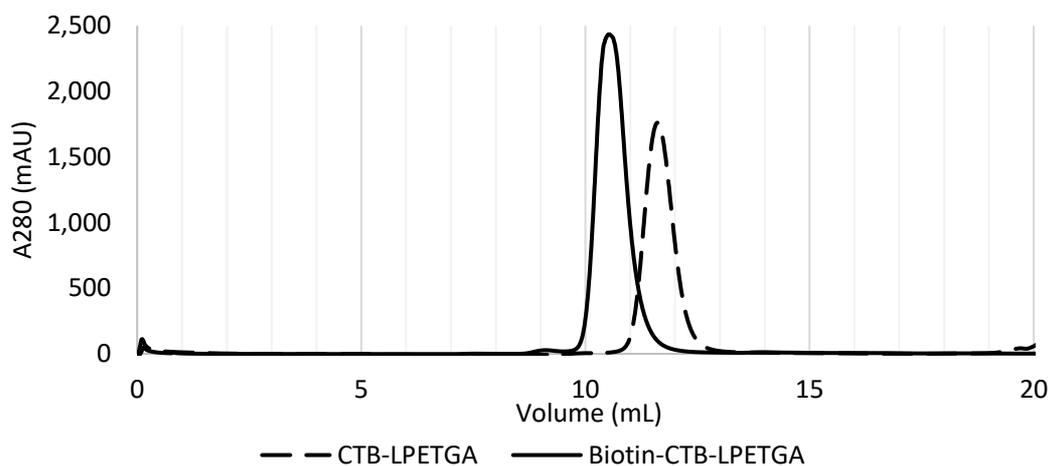
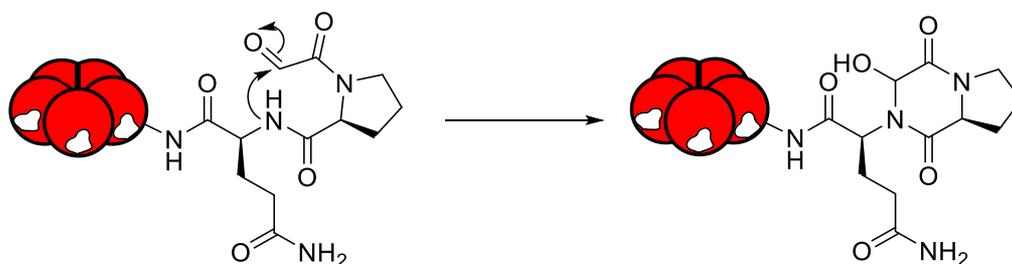


Figure 3-4: Size exclusion chromatography analysis (Superdex 75 10/300GL) of biotinylated CTB-LPETGA showing a shift in retention volume of Biotin-CTB-LPETGA compared to CTB-LPETGA.

An inert cyclisation product can form if the oxidised protein is not used rapidly in a subsequent oxime ligation reaction. The formation of this cyclised product is thought to occur due to kinking of the polypeptide backbone induced by the proline preceding the oxidised residue.²¹³ The kinking places the terminal aldehyde in close proximity to the amide nitrogen of the glutamine promoting the nucleophilic attack and cyclisation (Scheme 3-2). This cyclised diketopiperidine product has the same molecular weight as the aldehyde and hence any ES-MS peak that appears to correspond to the aldehyde is instead assumed to be this inert cyclised product. Provided the oxidised protein is used rapidly in subsequent oxime ligation, this cyclisation can be minimised if not eliminated. A small quantity of this cyclised product for CTB-LPETGA can be observed following biotinylation (Figure 3-3c).



Scheme 3-2: Cyclisation of the oxidised N-terminus of CTB. Kinking of the polypeptide chain by Pro2 brings the terminal aldehyde into close proximity with amide nitrogen of Gln3, leading to nucleophilic attack and cyclisation.

Following successful biotinylation of CTB-LPETGA, the protein was ready to be C-terminally ligated to the (Tn)MUC1 glycopeptides using the sortase conditions optimised in Chapter 2.

3.2.2 C-terminal Sortase Ligation of Biotinylated CTB-LPETGA

For preparative scale sortase reactions it would be vital to remove the sortase enzyme from the ligation reactions efficiently, as prolonged exposure to sortase will eventually lead to hydrolysis of the ligated product. The standard methodology for removing sortase from ligation reactions by exploiting the affinity of its His-tag for Ni-NTA resin was not possible due to the natural affinity of CTB for Ni-NTA resin. Attempts to exploit the difference in affinity of these two proteins to the Ni-NTA resin were unsuccessful, and further attempts to isolate the ligation product by size-exclusion chromatography (SEC) led to increased levels of hydrolysis compared to before being loaded on the SEC column. As a result, an alternative construct of sortase 7M was investigated which could be selectively removed from the reaction products.

3.2.2.1 Preparation of Chitin-binding Sortase 7M Variant

An N-terminal chitin binding domain (CBD) was selected as an alternative affinity tag due to its irreversible binding to chitin resin and its small size which was not expected to interfere with the sortase enzymatic activity. The synthetic pET28a-CBD-Srt7M plasmid, designed by Dr Michael Webb (University of Leeds) and ordered from Genscript, was transformed into commercial *E. coli* BL21 gold (DE3) competent cells. The protein was expressed in auto-induction media (AIM) initially at 30 °C for 5 hours before being left at 25 °C overnight. Following cell lysis, chitin binding variant of sortase 7M was purified first by Ni-NTA affinity chromatography then by size-exclusion chromatography using a Superdex[®] S75 (26/60) (Figure 3-5). The yield of the protein was significantly improved compared to the non-chitin binding variant with a yield of 181 mg of pure protein isolated per 1 L of cultured cells.

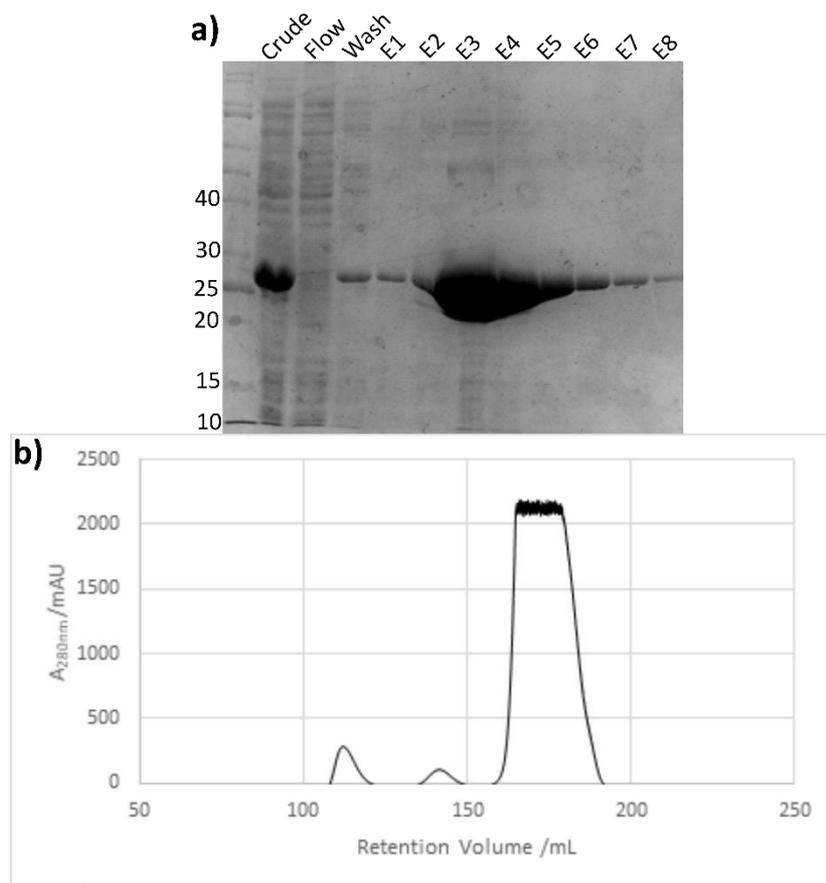


Figure 3-5: Purification of Chitin-binding Sortase 7M for use in preparative scale ligation reactions. a) SDS-PAGE gel of Ni-NTA affinity chromatography column flow through, wash and elution fractions (E1-8). b) SEC trace for Chitin-binding Sortase 7M using Superdex[®] S75 26/60.

3.2.2.2 Activity and Optimisation of Chitin-binding Sortase 7M Removal Protocol

The activity of this variant was probed to ensure the chitin binding domain was not interfering with the activity of the enzyme. The conjugation of CTB-LPETGA with glycopeptide **2.22** was repeated using both the regular sortase 7M and the chitin binding variant (Figure 3-6). Both constructs were tested side-by-side under identical conditions and prepared using the same stock solutions of CTB-LPETGA and glycopeptide **2.22**. The chitin binding variant of sortase 7M was found to ligate slightly faster over the first 30 minutes compared to the non-chitin binding variant. After 30 minutes conjugation had reached 68% using chitin binding sortase 7M, compared to 55% when using the non-chitin binding variant. However, at timepoints

beyond 1 hour the difference in conjugation between the two enzymes was negligible with conjugation reaching 90-91% for both enzymes after 2 hours.

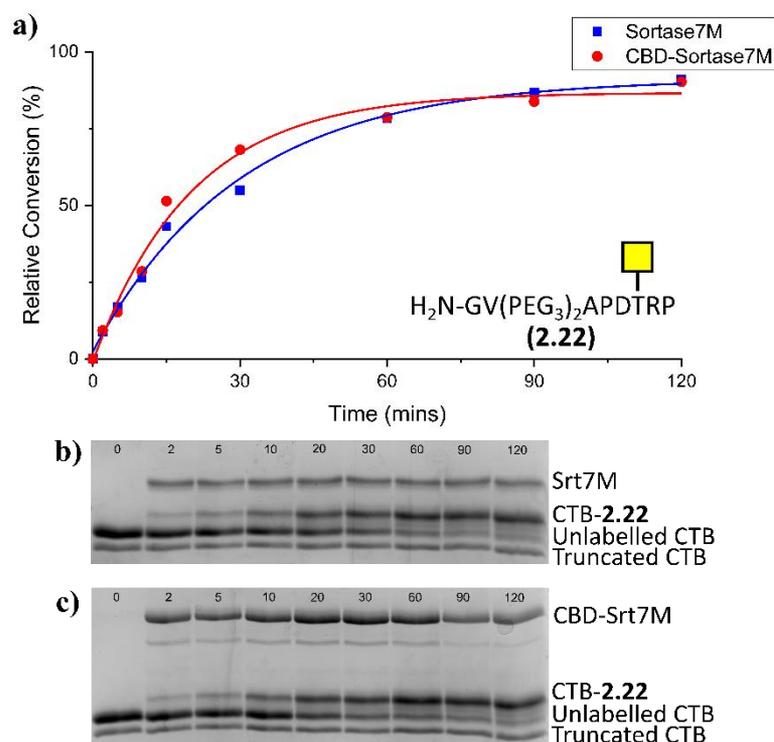


Figure 3-6: a) Fitted densitometry data for test sortase ligation using Sortase 7M and Chitin-binding domain Sortase 7M (CBD-Srt7M). Both reactions performed using CTB-LPETGA and the PEGylated glycopeptide **2.22** in 10% DMSO. b) SDS-PAGE gel for reaction using sortase 7M. c) SDS-PAGE gel for reaction using CBD-Sortase 7M. Data fitted in OriginPro 2019b (version 9.6.5) using an exponential function.

Chitin resin from New England Biolabs (NEB) is listed to have a binding capacity of 47 nmol per 1 mL of resin; it was believed 1.5 equivalents of chitin resin would be sufficient to bind and remove the CBD-sortase 7M enzyme. To test this hypothesis conjugation of glycopeptide **2.22** or peptide **2.30** to CTB-LPETGA was performed using the optimised conditions. After 2 hours, 1.5 equivalents of chitin resin was added and the mixture was incubated on ice to minimise any further reaction past the deemed endpoint and to improve the binding of the chitin binding domain to the chitin resin as recommended by the NEB application notes. After 15 minutes, the chitin resin was removed by spin filtration. Analysis of samples before and after the addition of the chitin resin showed complete removal of the CBD-Srt7M (Figure 3-7).

Densitometry showed no significant difference in the level of conjugation before and after removal of chitin resin.

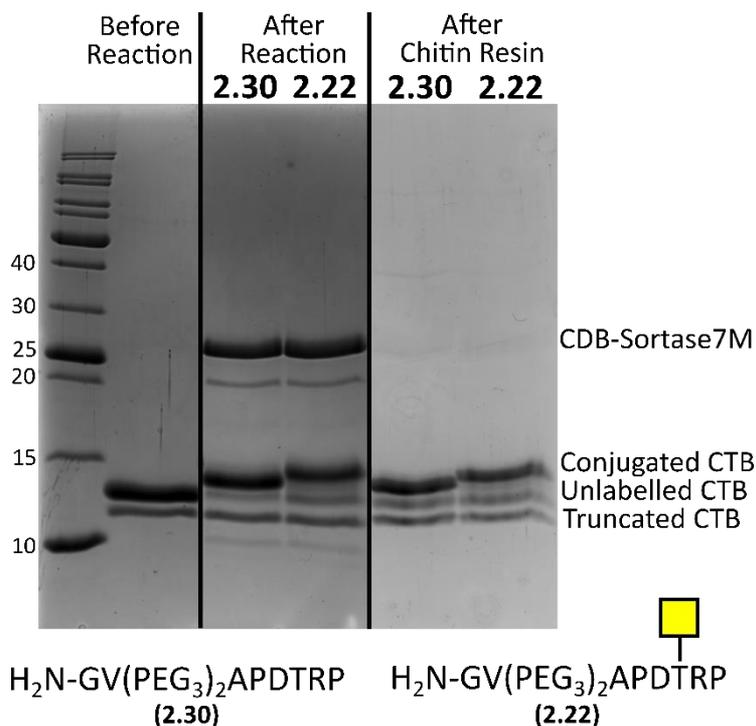


Figure 3-7: SDS-PAGE gel demonstrating removal of CBD-Sortase 7M from reaction mixture using 1.5 eq of chitin resin.

3.2.2.3 Production, Purification, and Evaluation of Biotinylated CTB-MUC1 Conjugates

Preparative scale sortase ligation of the biotinylated CTB-LPETGA with the PEG₆ peptides & glycopeptides (**2.22** & **2.25-2.30**) was performed using the optimised conditions. After 2 hours each reaction was placed on ice and CBD-Srt7M removed using chitin resin. The biotinylated glycoconjugates were further purified by size exclusion chromatography (Superdex[®] 75 10/300 increase) to remove the excess labelling substrate. After size-exclusion, the level of conjugation was evaluated by SDS-PAGE (Figure 3-8). No significant difference in sortase ligation efficiency was observed between CTB-LPETGA with and without the N-terminal biotin tag.

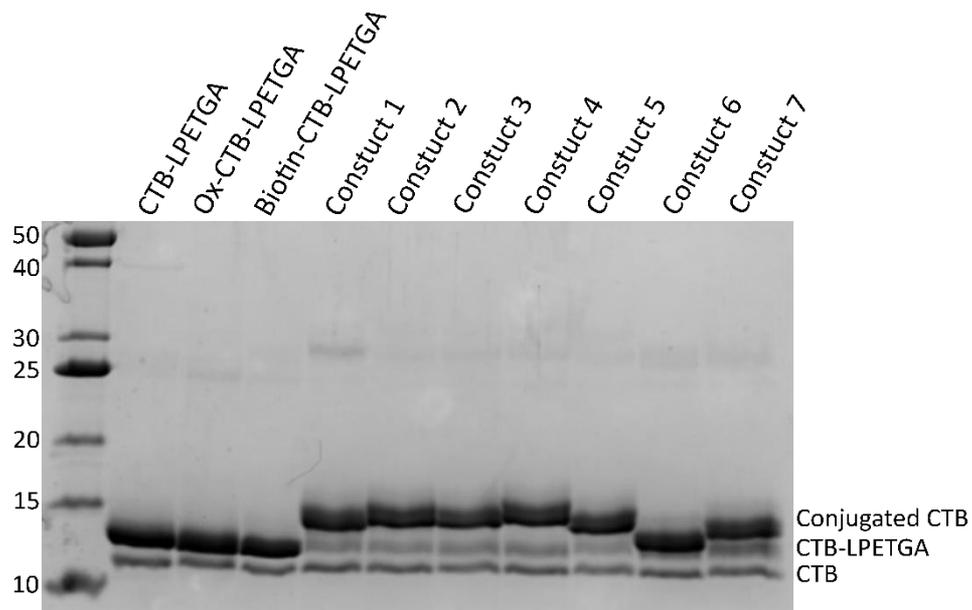


Figure 3-8: 15% SDS-PAGE gel of CTB-LPETGA, oxidised-CTB-LPETGA, biotinylated CTB-LPETGA and C-terminally conjugated & biotinylated CTB constructs 1-7 listed in Table 3-1 after purification by size-exclusion chromatography.

Table 3-1: Percentage C-terminal conjugation calculated using densitometry as a comparison of CTB-LPETGA and Conjugated CTB bands. CTB without the LPETGA tag (CTB) accounts for ~10% of each sample.

Construct #	Labelling Substrate	Conjugation (%)
1	2.25 - (GV(PEG ₃) ₂ APGSTAPPA	86
2	2.27 - (GV(PEG ₃) ₂ APGS(Tn)TAPPA	85
3	2.26 - (GV(PEG ₃) ₂ APGST(Tn)APPA	85
4	2.28 - (GV(PEG ₃) ₂ APGS(Tn)T(Tn)APPA	87
5	2.29 - (GV(PEG ₃) ₂ APPATSGPA	89
6	2.30 - (GV(PEG ₃) ₂ APDTRP	83
7	2.22 - (GV(PEG) ₆ APDT(Tn)RP	85

Following successful N-terminal oxime biotinylation, C-terminal sortase ligation and purification, the biotinylated CTB (Tn)MUC1-glycoconjugates (constructs 2, 3, 4 & 7) were handed over to BSTG who ran the Affimer phage display screening.

3.2.3 Phage Display Screening of Biotinylated CTB-glycoconjugates

Phage display screening was performed by Anna Tang in the BioScreening Technology Group (BSTG) at the University of Leeds with the support of Christian Tiede. Commercially available streptavidin-coated 96 well plates and streptavidin coated beads were used for immobilisation of the CTB-glycoconjugates.

After three independent screens, Dr Anna Tang and Dr Christian Tiede (University of Leeds) concluded that it was too difficult to selectively amplify Affimers which were binding to the glycopeptides. Increasing the stringency of the pre-panning, washing and competitive selection conditions facilitated a reduction in background binding to CTB significantly, however, it also resulted in Affimers with weak affinity for the glycopeptides being stripped from the screen. As a result, no Affimer could be amplified or identified that would bind effectively to the glycopeptides.

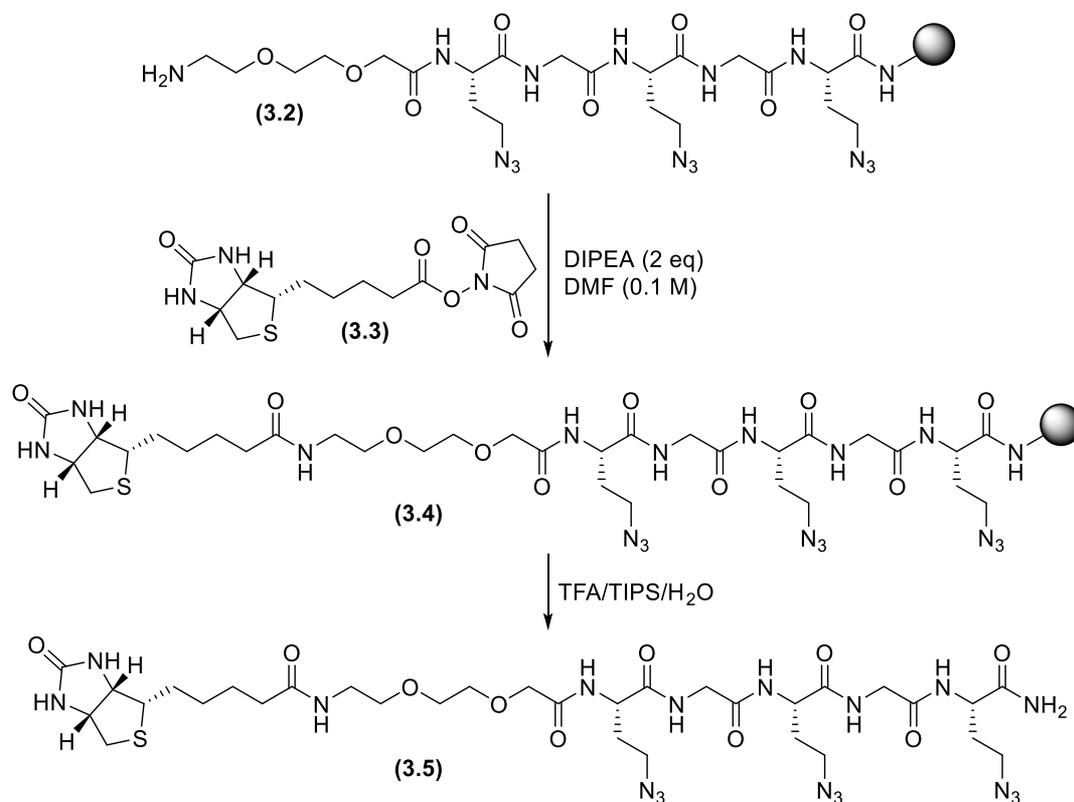
3.3 Multivalent Biotinylated (Tn)MUC1 Glycopeptides for Phage Display

Using CTB as a multivalent scaffold likely caused a selectivity issue despite the majority of Affimer identified in previous screens by Dr Matthew Balmforth (University of Leeds) having relatively weak binding affinities.¹⁹⁸ To avoid the use of a protein as a multivalent scaffold a small simple peptide was designed to which the glycopeptides could be easily attached. The peptide scaffold included an N-terminal biotin for immobilisation to a streptavidin-coated plate, a short PEG linker to improve aqueous solubility of the peptide and provide a flexibility linker between the biotin and glycopeptides, and three azidohomoalanine (Aha) residues interspaced with glycine (**3.5**). The azidohomoalanine residues could easily be labelled with the BCN-functionalised glycopeptides *via* SPAAC.

3.3.1 Preparation of Biotinylated Scaffold

The multivalent biotinylated scaffold was synthesised by solid phase peptide synthesis (SPPS) using Fmoc-Aha-OH provided by Dr Ryan McBerney (University of Leeds).¹⁸⁷ Following incorporation of the PEG spacer (Fmoc-AEEAc-OH) and

Fmoc deprotection, the N-terminus was capped with biotin on resin using 2 equivalents of Biotin-N-hydroxysuccinimide ester (**3.3**, Scheme 3-3). LCMS of a small portion of the peptide cleaved from the resin showed complete N-terminal biotinylation of the peptide, and the peptide was cleaved from the resin using standard TFA cleavage conditions for 2 hours. The biotinylated trivalent azido peptide scaffold (**3.5**) was isolated *via* ether precipitation and lyophilisation.

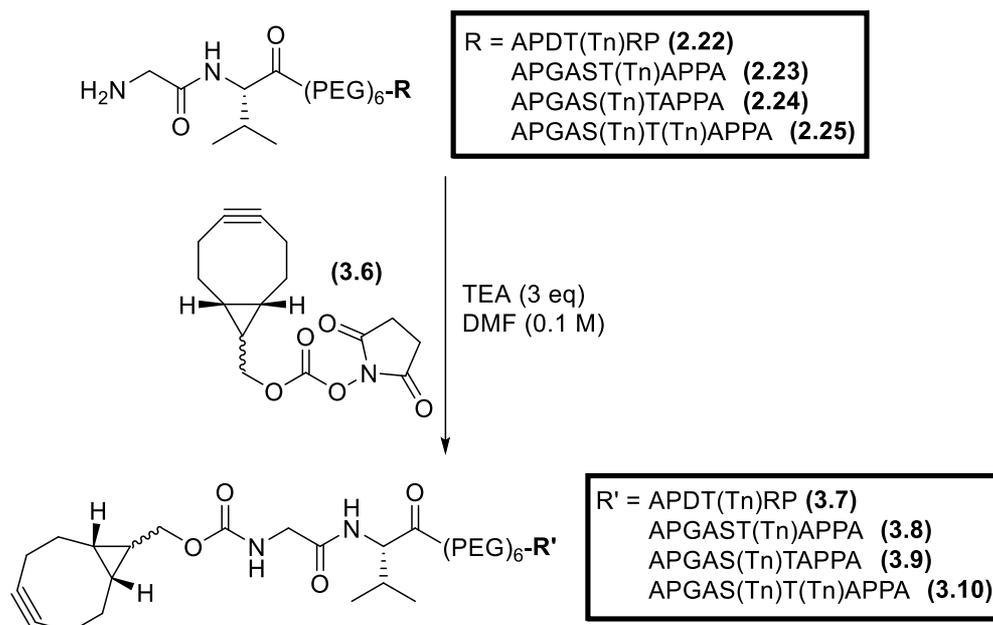


Scheme 3-3: Preparation of trivalent biotinylated peptide scaffold prepared via SPPS on Rink amide MBHA low loading resin using Fmoc-azidohomoalanine (Aha), Fmoc-glycine, Fmoc-AEEAc-OH and biotin-OSu (3.3). Cleaved from resin using 95:2.5:2.5 TFA/TIS/H₂O.

3.3.1.1 BCN labelling of Glycopeptides and Strain-Promoted Azide-Alkyne Cycloaddition

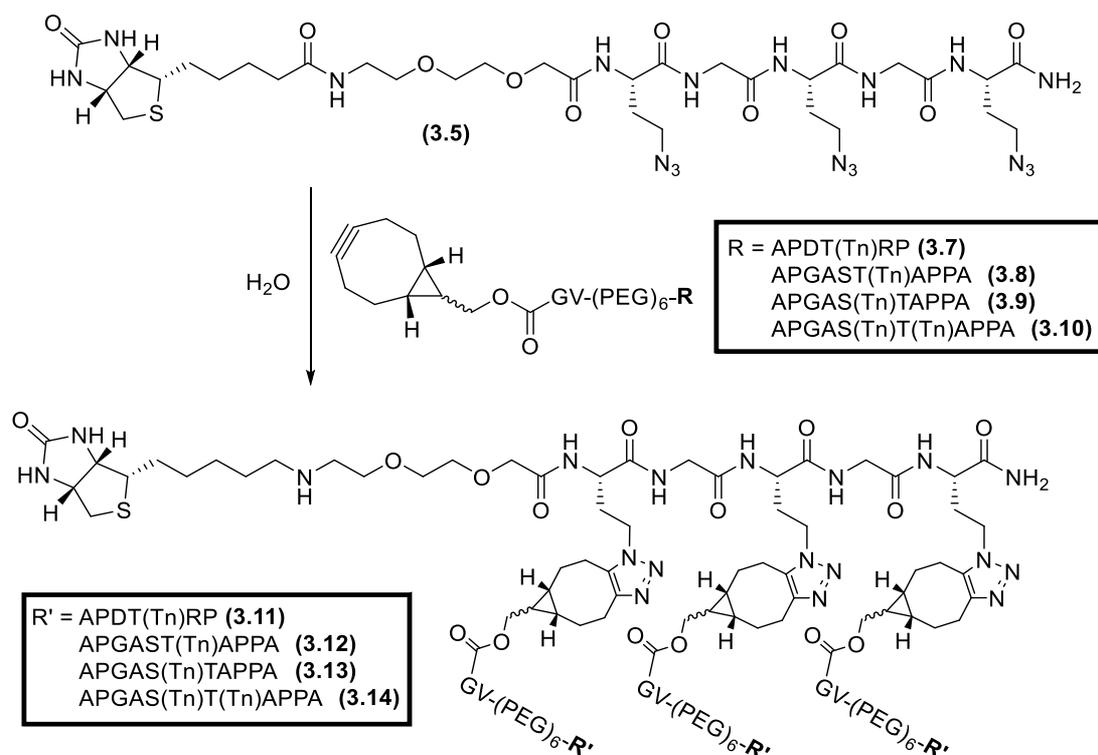
The four MUC1(Tn) glycopeptides of interest (**2.22**, **2.26**, **2.27** and **2.28**) were N-terminally labelled using 1.2 equivalents of BCN-NHS in DMF overnight (Scheme 3-4). Once LC-MS confirmed the reactions were complete, the BCN-glycopeptides

were purified by size-exclusion using a Biogel P2 column. Four BCN-labelled MUC1(Tn) glycopeptides (**3.7**, **3.8**, **3.9**, **3.10**) were prepared in 62–81% yields.



Scheme 3-4: BCN functionalisation of the N-terminus of glycopeptides (2.22-2.25) using BCN-NHS (3.6).

Each of the four BCN glycopeptides were attached to the scaffold (**3.5**) via SPAAC using 3.3 equivalents of the BCN labelled glycopeptide (Scheme 3-5). The reactions were monitored by MALDI-TOF mass spectrometry; if required, additional BCN-functionalised glycopeptide was added to ensure all three triazole linkages were formed. Once deemed to have reach completion, the biotinylated trimeric glycopeptides were purified by size-exclusion chromatography using Superdex 30 10/300 eluting with 20 mM ammonium formate to remove the excess BCN-labelled glycopeptide.



Scheme 3-5: SPAAC of BCN functionalised glycopeptides (3.7-3.10) and trivalent azido peptide scaffold (3.5).

3.3.1.2 Phage Display Screening of Multivalent Biotinylated (Tn)MUC1 Glycopeptides

The four biotinylated trimeric glycopeptides (**3.11-3.14**) were handed over to BSTG who ran the Affimer phage display screening. Phage display screening was performed by Dr Christian Tiede (BSTG, University of Leeds) using the standard high throughput methodology used previously for the CTB glycoconjugates. Screening was repeated three times and no Affimer proteins could be raised against the MUC1(Tn) glycopeptides. This may suggest the Affimer is not an appropriate scaffold for developing binders against glycans or glycopeptides. Alternative phage display libraries should be considered such as those for genetically coded cyclic peptides or other antibody-like scaffolds such as Nanobody, Trans-body or Affibody.²¹⁴⁻²¹⁶

3.4 Identifying VTB-binding Affimers

While neither of the previous strategies led to Affimers that bound to the glycopeptides, the first strategy principally failed because of high background binding to the CTB scaffold protein. Previously within our lab, Matthew Balmforth began work to develop a method of targeted drug delivery of proteins using Affimers that can bind to the cholera toxin B-subunit.¹⁹⁹ The Affimer-CTB complex was shown to be internalised by GM1-expressing cell line (Vero cells). When injected into the tongue of a mouse, the complex was trafficked from the neuromuscular junction to the cell bodies of the hypoglossal nerve in the brain stem (Figure 3-9).^{198,199}

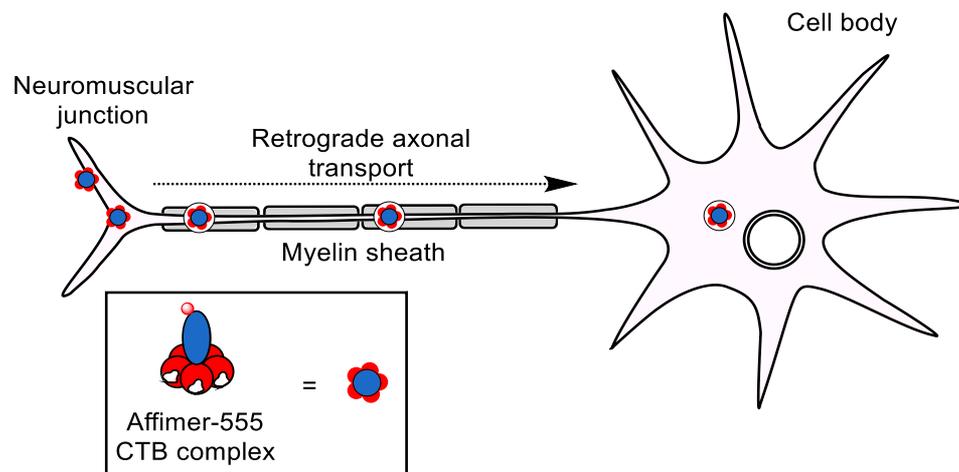


Figure 3-9: Internalisation of fluorescently labelled Affimer complexed to CTB and trafficking from neuromuscular junction to the cell bodies of hypoglossal nerve.

Adapted from Balmforth.¹⁹⁸

We postulated that other bacterial toxin B-subunits with different glycan-binding specificities could also be used to safely deliver drugs and biologics to other cell types, such as verotoxin B-subunit (VTB). Studies have shown 62% of pancreatic and 81% of colon adenocarcinomas showed elevated Gb3Cer/CD77 expression, indicating an association of this marker with neoplastic transformation.^{217,218} Therefore, Affimer-VTB complexes could potentially be used for targeted drug delivery in pancreatic and colon cancer. Initially Affimers would be screened against biotinylated VTB. To identify Affimers which exclusively bound to the non-Gb3 binding face of VTB a plate-based Affimer-lectin binding assay (ALBA) would be used to characterise binding specificity and rank the binders.

3.4.1 VTB Expression and Purification

The XL10 and expression cell lines of WT Verotoxin B-subunit was inherited as a gift from Prof. Steve Homans (University of Leeds).^{219,220} Expression of WT VTB and purification by periplasmic extraction was previously optimised by Dr Ryan McBerney.¹⁸⁷ Wild type VTB with a periplasmic targeting sequence was expressed recombinantly from *E. coli* grown in CAYE (casamino acid-yeast extract) growth medium modified according to Evans.²²¹ Unlike CTB, following folding in the periplasm and cleavage of the signal peptide, the protein is not exported into the media. Instead VTB was harvested by periplasmic extraction *via* osmotic shock. Periplasmic extraction involves breaking apart the outer membrane leaving the inner membrane intact, allowing for isolation of proteins transported to the periplasm without extraction of the whole cellular proteome. Even though the protein contained no purification tag, SDS-PAGE analysis showed the periplasmic extraction produced a relatively clean lysate and it was possible to use size-exclusion chromatography to remove any remaining impurities and isolate the VTB pentamer (Figure 3-10b). ES-MS of the protein confirmed successful expression, showing the correct mass for the monomeric VTB at 7687.80 Da (Figure 3-10c, theoretical mass = 7688.66 Da).

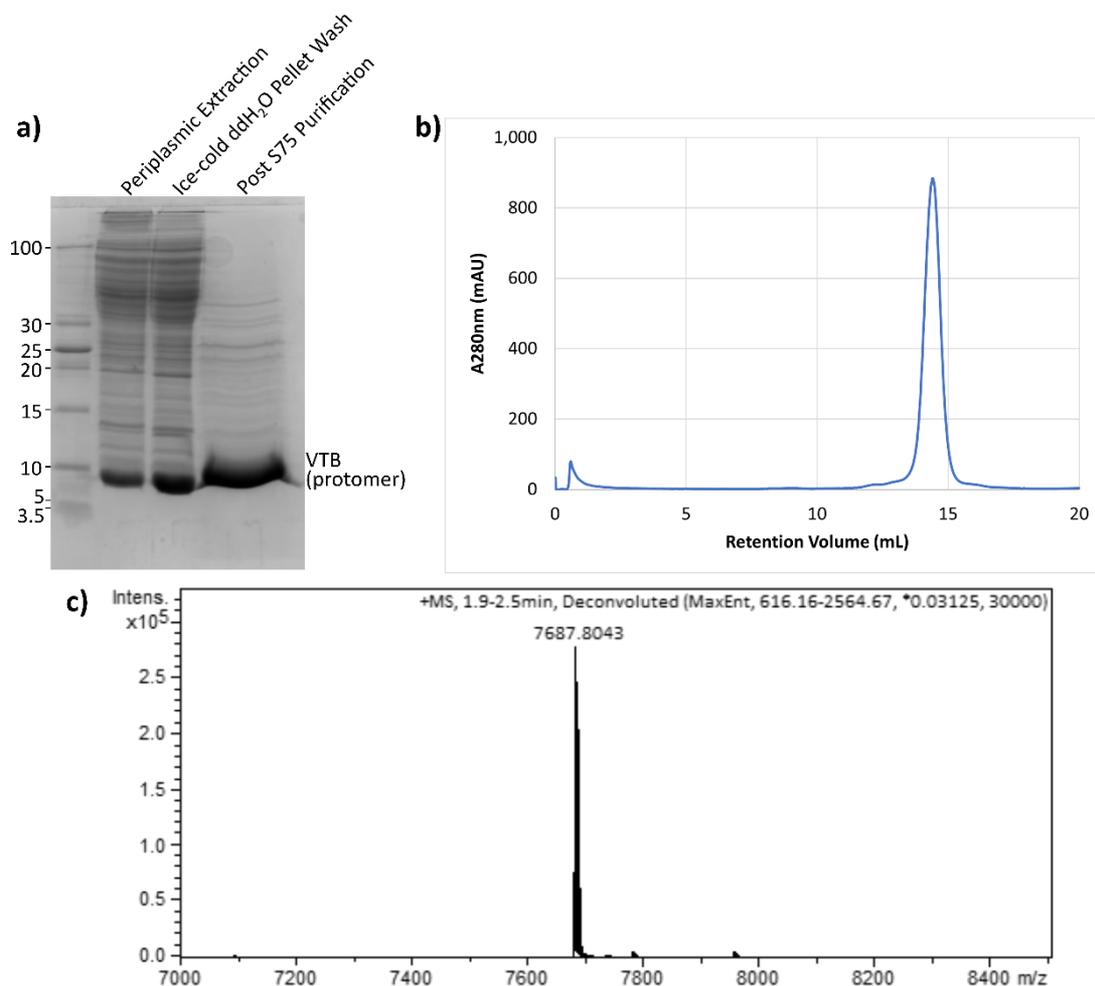
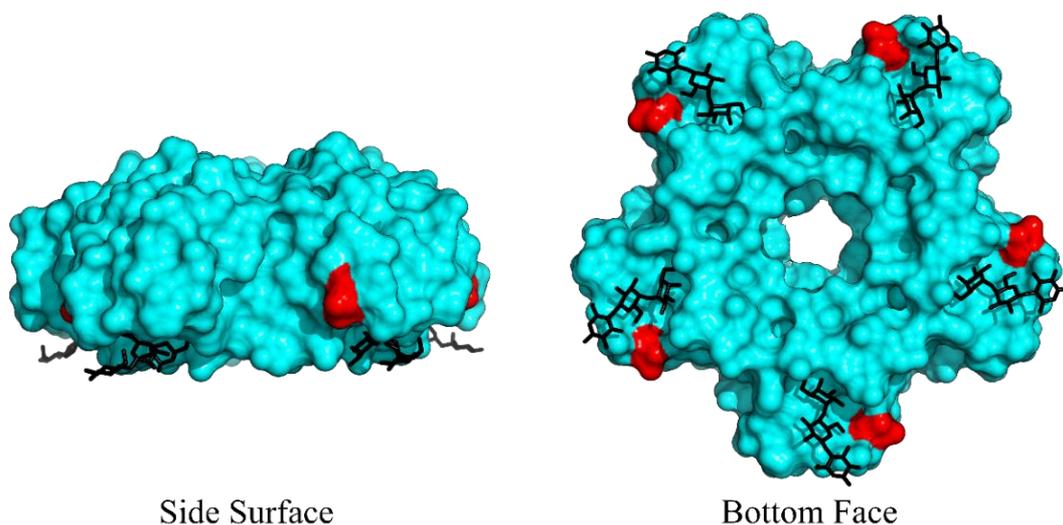


Figure 3-10: Purification of VTB. a) SDS-PAGE gel of periplasmic extraction buffer, osmotic shock pellet wash & VTB after purification by size-exclusion chromatography. b) Size-exclusion chromatography of VTB using Superdex® S75 10/300. c) Deconvoluted ES-MS of purified VTB (7687.8 Da).

3.4.2 VTB Biotinylation

For phage display VTB was biotinylated following the same protocol previously employed for CTB-LPETGA. The N-terminal threonine residue of VTB is located on the bottom face of the protein close to one of the Gb3 binding sites therefore site-specific modification of the N-termini should place VTB in a favourable orientation to present the non-Gb3 binding face to the Affimer-phage particles, resulting in fewer undesired binders to the bottom face.



Side Surface Bottom Face

Figure 3-11: Surface representation of pentameric Verotoxin B-subunit showing five Gb3 oligosaccharides (black) bound to the binding site of the bottom face of the pentamer. N-terminal threonine residue coloured red. Figure prepared using PyMOL (version 2.4.0) from Protein Data Bank file 1D11.pdb

Oxidation of VTB was performed using five equivalents of sodium periodate in the presence of ten equivalents of L-methionine in sodium phosphate buffer (pH 7.4). The reaction was monitored by ES-MS and typically found to be complete within 10 minutes, with an observed mass loss of 27 Da (Figure 3-12b). The oxidised VTB was purified by PD-10 desalting column with buffer exchange into sodium phosphate buffer at pH 6.8. The oxidised VTB was combined with aniline, ten equivalents of the aminoxy PEG₄ biotin (**3.1**) and left overnight to react at 37 °C. Deconvoluted ES-MS confirmed VTB had near-quantitatively been N-terminally biotinylated with a mass increase of 398 Da from the oxidised VTB (Figure 3-12c).

Deconvoluted ES-MS after oxime biotinylation shows a small quantity of inert N-terminally cyclised VTB could be seen (7642.7 Da), which has the same mass as the oxidised VTB aldehyde (Figure 3-12c). As discussed previously, the formation of this cyclised product is thought to occur due to kinking of the polypeptide backbone induced by the proline preceding the oxidised residue. Backbone kinking places the terminal aldehyde in close proximity to the amide nitrogen of the aspartic acid promoting the nucleophilic attack and cyclisation.

The biotinylated VTB was purified by size-exclusion chromatography. Unlike CTB, VTB does not run as a pentamer on an SDS-PAGE gel. Using size-exclusion chromatography allowed for confirmation the pentameric structure of VTB had been retained following biotinylation (Figure 3-13).

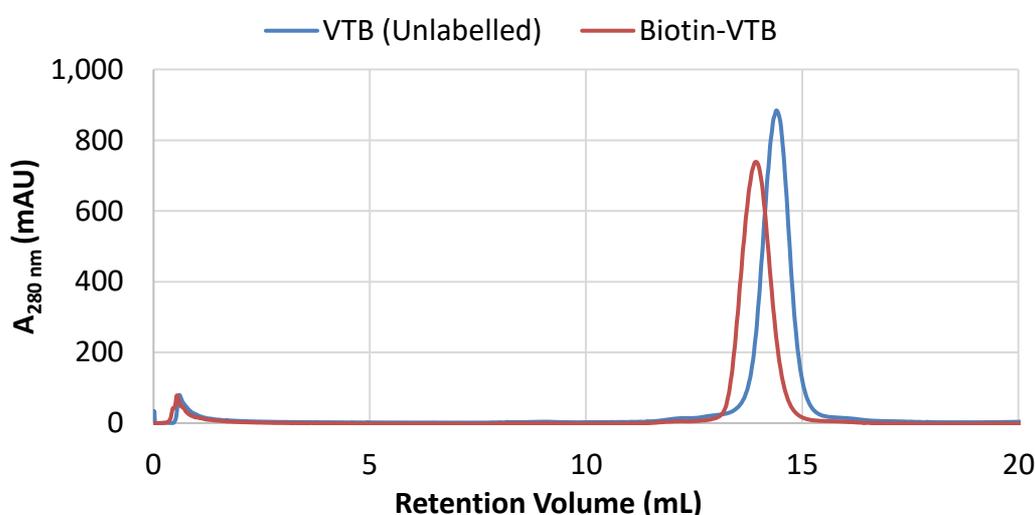


Figure 3-13: Size-exclusion chromatography analysis (Superdex 75 10/300GL) of biotinylated VTB, showing a slight shift in retention volume of Biotin-VTB compared to unlabelled VTB.

3.4.3 Phage Display Screening of Verotoxin B-subunit

The biotinylated VTB was handed over to the University of Leeds Bio-Screening Technology group (BSTG) for Affimer phage display screening. Firstly the naïve library was pre-panned against streptavidin-coated plates to remove Affimer that bound to streptavidin or any competent of the plate. The phage library was then enriched, screening the biotin-VTB immobilised on both streptavidin-coated plates

and beads, allowing for a change in the mode of presentation. No competitive selection was performed using unbiotinylated VTB. Due to the weak binding interaction between CTB and the bound Affimer, Matthew Balmforth found competitive selection was not possible and when performed no Affimer against CTB could be identified.¹⁹⁸ From the enriched library, twenty-four randomly picked clones were tested by phage ELISA (Figure 3-14) and the eleven positive clones sequenced; this gave rise to nine unique Affimer sequences. Further work is underway within the Turnbull group to characterise their binding mode, specificity, and affinity for VTB. Binders could have potential uses ranging from drug-delivery conjugates, development of VTB-inhibitors or fusogenic lectins.

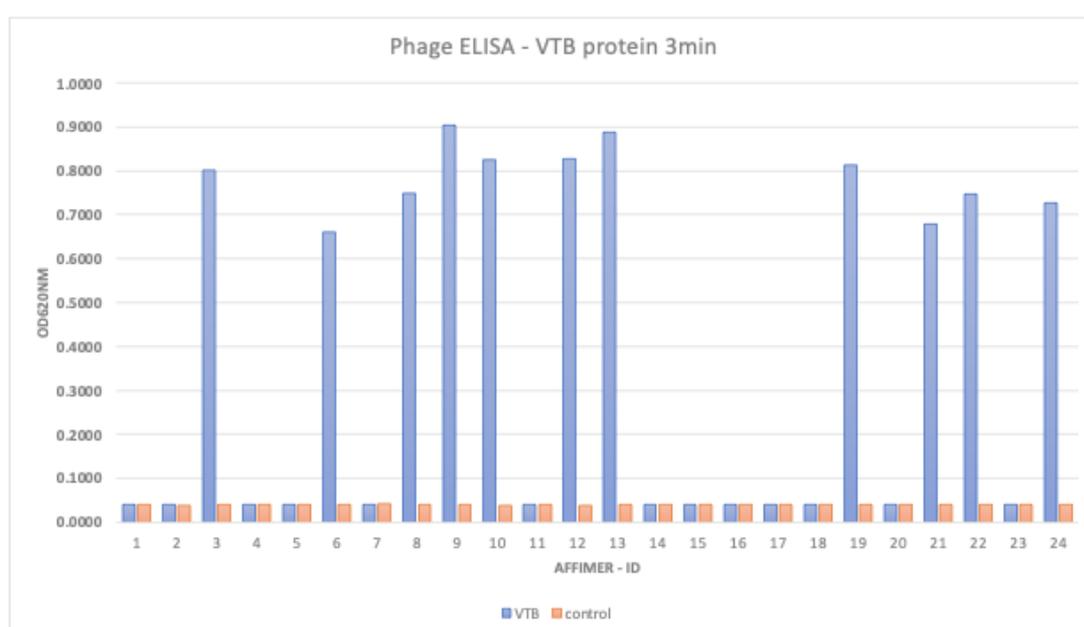


Figure 3-14: Analysis by phage ELISA of Affimers isolated following panning against VTB-biotin performed by Dr Christian Tiede (University of Leeds). A single species of Affimer is tested against either VTB-biotin bound to streptavidin or streptavidin alone. In the event of binding, a chromophore is produced resulting in an increase in absorbance at 620 nm. Twenty-four Affimers were tested, with eleven producing a large increase in absorbance in the presence of VTB-biotin.

3.5 Chapter Conclusion

A CTB scaffold has been orthogonally labelled successfully at both the N- and C-terminus using two orthogonal site-specific methods. The N-terminus of CTB-

LPETGA was quantitatively labelled with an aminoxy-functionalised biotin (**3.1**) *via* oxime ligation. Sortase-mediated ligation can be used to ligate a series of MUC1 peptide and glycopeptide epitopes to the C-terminus of this biotinylated scaffold site-specifically with a ligation efficiency of 83-89%. Phage display screening of the biotinylated CTB-glycoconjugates was unsuccessful. Although there was a chance Affimers with affinity for the glycan or glycopeptides were present, eliminating undesired binders against the CTB scaffold was not possible. To avoid the use of CTB as scaffold, a biotinylated peptide scaffold (**3.5**) was synthesised containing three azidohomoalanine residues. This scaffold was labelled with bicyclononyne functionalised glycopeptides *via* SPAAC. Four examples of trimeric-MUC1(Tn) biotinylated glycopeptides (**3.11-3.14**) were synthesised. Affimer phage display screening was unsuccessful in identifying any unique binders.

Verotoxin B-subunit was successfully overexpressed in *E. coli* and isolated from the periplasm *via* osmotic shock (periplasmic extraction) followed by purification by size-exclusion chromatography. Using oxime ligation, VTB was near-quantitatively N-terminally biotinylated with aminoxy-PEG₄-Biotin (**3.1**) and shown to remain pentameric following biotinylation. Phage display screening of the biotinylated-VTB identified nine unique binders.

Chapter 4: Multivalent Glycoconjugate Inhibitors Based on Non-binding CTB Scaffold.

4.1 Background

As discussed in section 1.3, toxin-secreting bacteria are a major cause of diarrhoeal disease, ranked in the top ten causes of death globally. Previously Dr Tom Branson (University of Leeds) reported first neoglycoprotein inhibitor against CTB (section 1.3.3.2).¹⁰⁸ As proposed in the project aims, Ryan and I wanted to explore the effects of reducing linker length and changing position of the glycans has on inhibitory potential of CTB-based neoglycoprotein inhibitors.

Previously Dr Ryan McBerney (University of Leeds) developed two minimal length, bifunctional linkers (Figure 4-1).¹⁸⁷ GM1-BCN (**4.3**) was synthesised by oxime ligation using bifunctional linker (**4.1**) and subsequently ligated to azido-CTB(W88E) via SPAAC ligation (Figure 4-2). The azido-CTB(W88E) scaffold is based on the previously reported azido-CTB which contains four mutations compared to wild type El Tor CTB (M37L, K43M, M68L, M101L) which allows introduction of an azidohomoalanine residue at position 43 by isosteric replacement of methionine.^{130,187} The W88E mutation was also introduced to remove GM1 binding capability so that the protein could be used as a neoglycoprotein scaffold. This inhibitor was found to have an IC₅₀ of 457 pM against CTB binding to a GM1-coated microtitre plate.¹⁸⁷

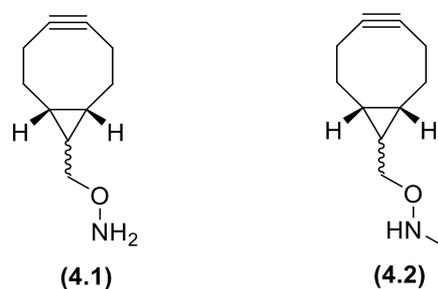


Figure 4-1: Structure of two minimal length, bifunctional BCN-linkers developed by Ryan McBerney.¹⁸⁷

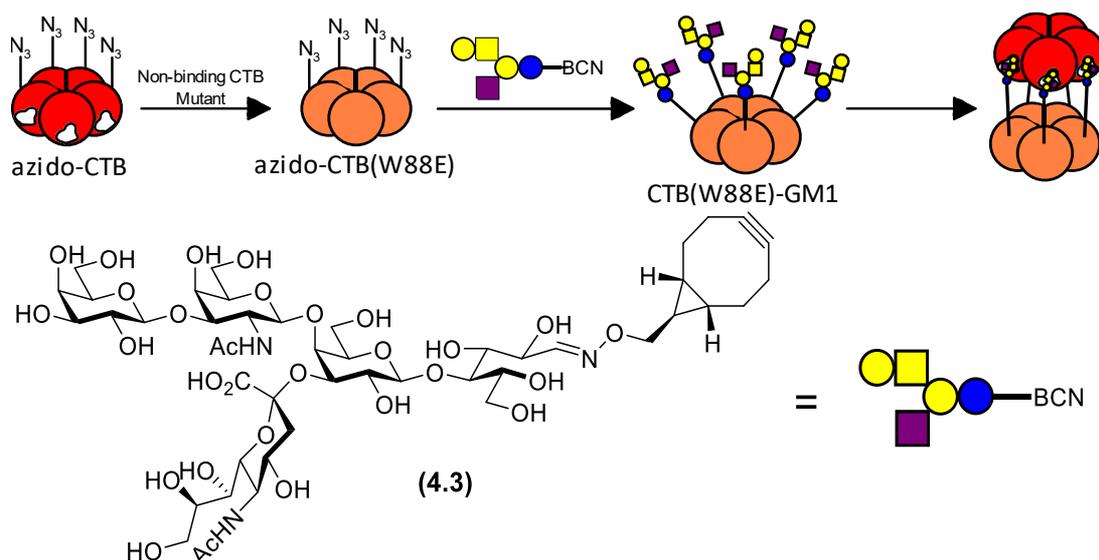


Figure 4-2: Cartoon depiction of protein based CTB inhibitor developed using minimal length linker. Mutation of wild type azido-CTB to form a non-binding variant, and attachment of BCN functionalised glycan (4.3) by SPAAC.¹⁸⁷

This chapter will discuss the development of a second neoglycoprotein-based pentavalent inhibitor of CTB prepared *via* N-terminal oxime ligation of BCN-ONH₂ (4.1) and strain-promoted azide alkyne cycloaddition (SPAAC) of GM1 azide (Figure 4-3). This approach should overcome issues of undesired side-reactions between the aminoxy-BCN compounds (4.1, 4.2) and acetone present within the chemistry lab atmosphere, that were encountered previously by Ryan McBerney. Furthermore, this approach also presents an advantage of allowing a much smaller quantity of azido-functionalised glycan to be used in the SPAAC ligation. Given the impact neoglycoconjugate inhibitors have shown against cholera toxin, this work will also be expanded to produce pentavalent inhibitors against verotoxin, which can cause food poisoning and haemolytic uremic syndrome.

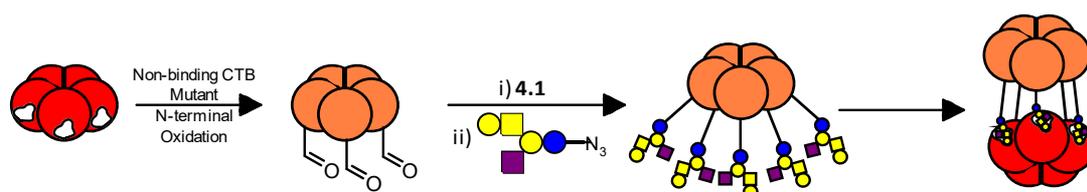


Figure 4-3: Cartoon depiction of alternative route to protein based CTB inhibitor. N-terminal oxidation and oxime ligation of BCN-ONH₂ (4.1) followed by SPAAC of GM1 azide.

4.2 Preparation of a BCN-functionalised CTB Scaffold

4.2.1 Expression of Non-binding CTB(W88E) Mutant

Cholera toxin B-subunit binds GM1 oligosaccharide (GM1os) very tightly with a dissociation constant (K_d) of ~40 nM (measured by isothermal titration calorimetry, ITC), while simple galactosides have millimolar dissociation constants.^{82,86} CTB-GM1 binding can be eliminated by mutation of tryptophan 88 (shown in blue, Figure 4-4) to glutamate.^{222,223} As demonstrated in Chapter 3, CTB naturally has a threonine residue at the N-terminus (shown in yellow, Figure 4-4), which can easily be oxidised to give an aldehyde and then labelled by oxime ligation in the presence of aniline.^{108,166}

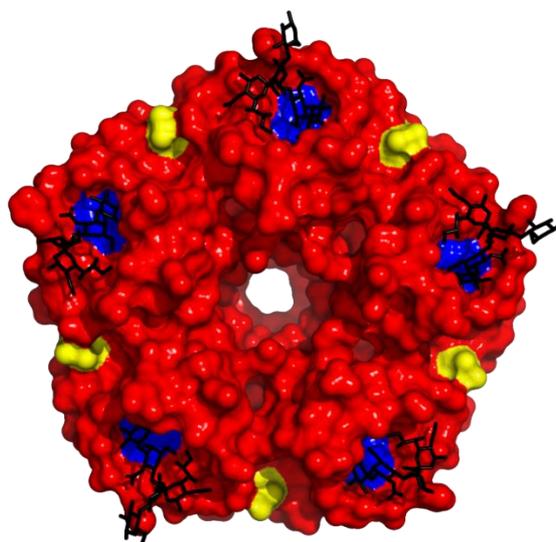


Figure 4-4: Surface representation of pentameric Cholera Toxin B-subunit showing the GM1 binding face. N-terminal threonine residues coloured yellow; Tryptophan 88 required for GM1 binding coloured blue. GM1 oligosaccharide shown as black sticks. Figure prepared using PyMOL (version 2.4.0) from Protein Data Bank file 3CHB.pdb

The pTRB-W88E plasmid which encodes CTB(W88E), a non-GM1 binding mutant of CTB was prepared by Dr Tom Branson (University of Leeds).¹⁰⁸ CTB(W88E) was expressed with a periplasmic signalling peptide which directs the protein to the periplasm, where the signal peptide is cleaved and the protein folds into its mature pentameric form, before being secreted into the media. The protein was harvested from the growth media by ammonium sulfate precipitation and purified by Ni-NTA

affinity chromatography. Expression and purification was confirmed by SDS-PAGE analysis (Figure 4-5a). Any remaining impurities following Ni-NTA affinity chromatography were removed by size-exclusion chromatography using a Superdex S75 column (Figure 4-5b). ES-MS of the protein confirmed successful expression, showing the correct mass for the monomeric CTB(W88E) at 11584.95 Da (Figure 4-5c, theoretical mass = 11584.27 Da).

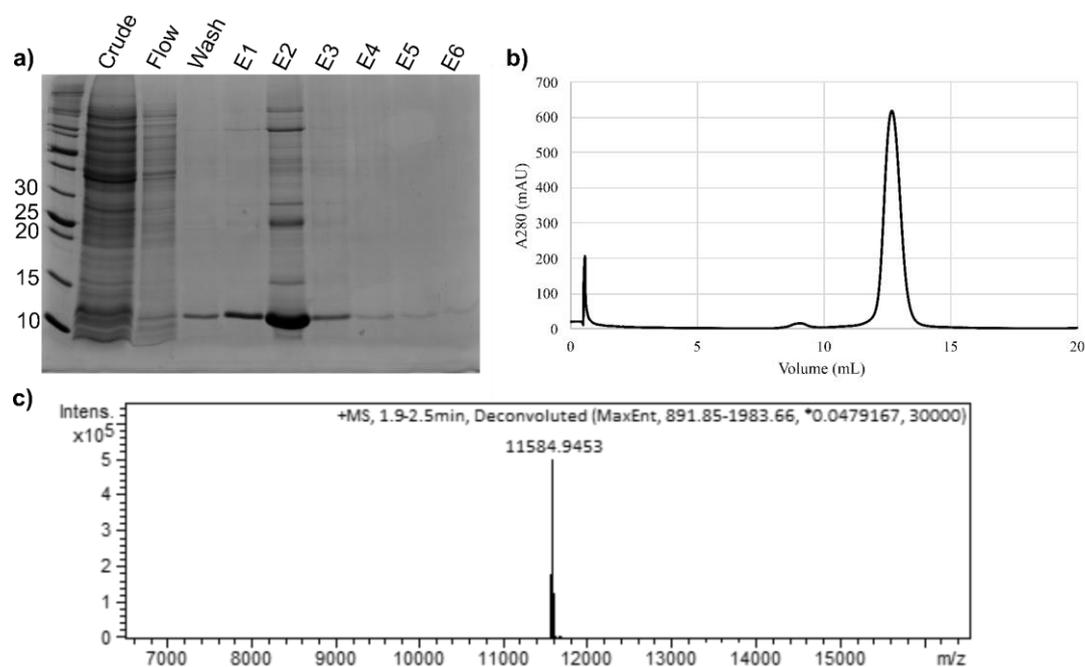
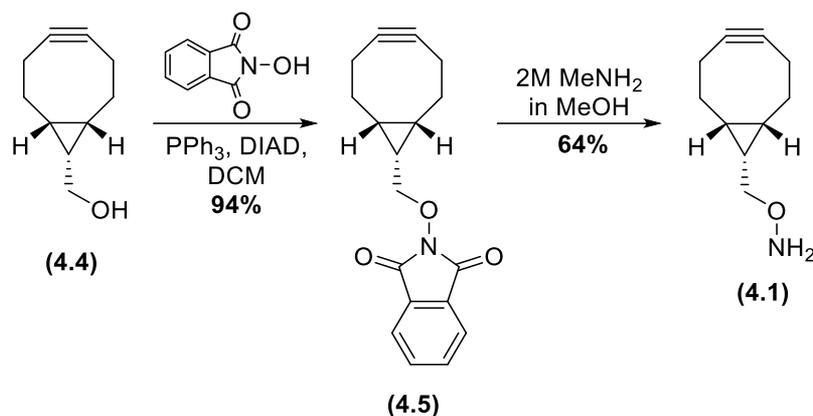


Figure 4-5: Purification of CTB(W88E) a non GM1 binding mutant of CTB. a) SDS-PAGE gel of Ni-affinity chromatography (elution fractions: E1-6). Samples boiled prior to loading. b) Size-exclusion chromatography of CTB(W88E) using Superdex® S75 10/300. c) Deconvoluted ES-MS of purified CTB(W88E) (11584.9 Da).

4.2.2 Synthesis of Aminoxy BCN-functionalised Linker

Endo-BCN-OH (**4.4**) was purchased commercially to avoid its challenging and low yielding synthesis.^{187,224} A protected oxyamine (**4.5**) was installed *via* Mitsunobu reaction with N-hydroxyphthalimide in 94% yield.¹⁸⁷ From this point BCN-ONH₂ (**4.1**) was synthesised *via* ring opening of the phthalimide (**4.5**) using 2M methanolic methylamine in an acetone free environment (Scheme 4-1).¹⁸⁷



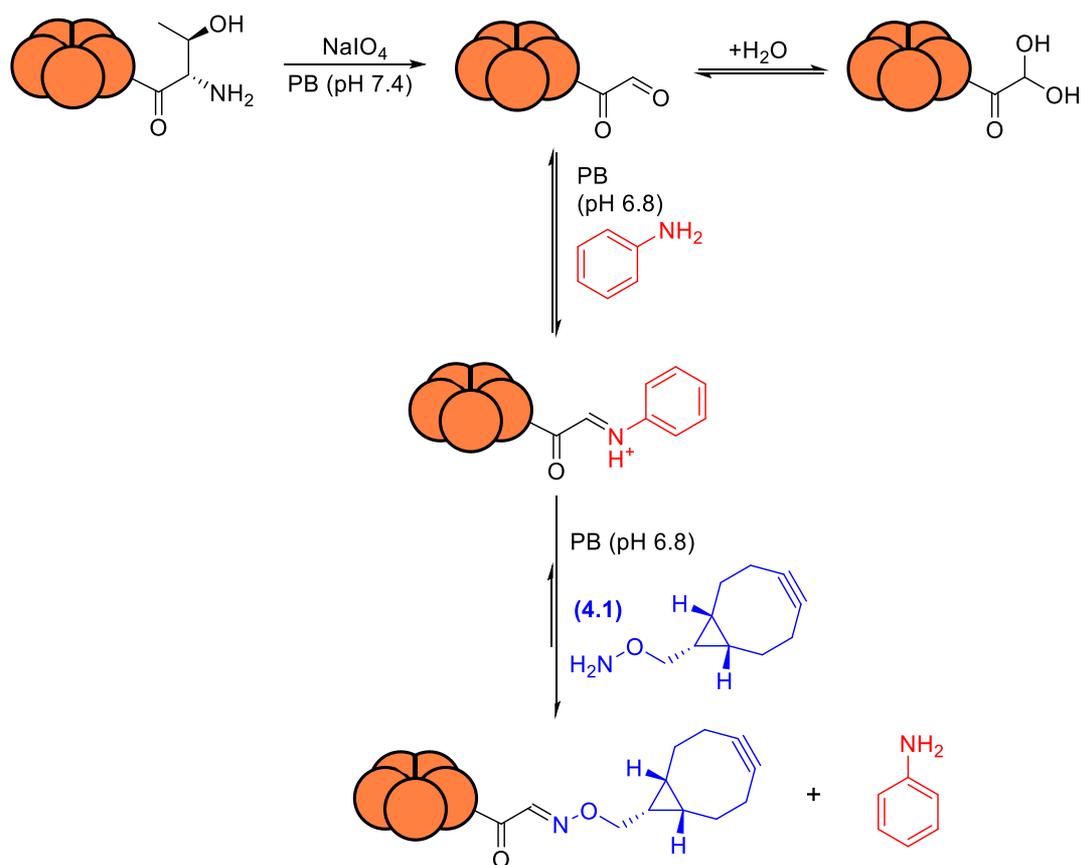
Scheme 4-1: Synthetic route to aminoxy-endo-BCN linkers 4.1.

4.2.3 Site-specific BCN-functionalisation of CTB(W88E) by Oxime

Ligation

In preliminary work, Ryan McBerney had performed SPAAC ligation between the aminoxy-BCN (4.1) and lactosyl azide, before performing oxime ligation to oxidised CTB(W88E). Although his lactosyl reaction products were easy to purify using the BCN group as a hydrophobic tag for reverse phase chromatography, GM1 azide could not be produced in sufficient quantity for him to make the target neoglycoprotein inhibitor. Therefore, I chose to investigate an alternative synthetic route that would be more amenable to smaller quantities of azido-functionalised glycans. The synthesis route was rearranged to functionalise CTB(W88E) with five BCN groups per pentamer first, which could then be labelled with any glycosyl azide of interest. This would open the route to a more toolbox-style approach allowing, in the future, an array of glycans to be displayed on a multivalent scaffold and then screened for inhibitory activity.

Having successfully expressed the non-GM1 binding CTB scaffold (CTB(W88E)) and synthesised the aminoxy-BCN (4.1) handle oxime ligation to prepare the BCN-CTB(W88E) construct could take place.



Scheme 4-2: Oxidation of N-terminal Thr of CTB(W88E) using sodium periodate to form an aldehyde which becomes hydrated in aqueous media. The aldehyde reacts with aniline to form an iminium intermediate which is rapidly displaced by the aminoxy functionalised endo-BCN to form an oxime linkage.

Oxidation of CTB(W88E) was performed as described in section 3.2.1 (Scheme 3-1). The reaction was monitored by ES-MS and typically found to be complete within 10 minutes, with an observed mass loss of 27 Da (Figure 4-7b). The desalted oxidised CTB(W88E) was combined with the aniline catalyst (1% w/v), ten equivalents of aminoxy functionalised endo-BCN (**4.1**) and left overnight to react. The BCN-functionalised CTB(W88E) was purified by PD-10 desalting column and the protein characterised by size-exclusion chromatography to ensure the pentameric structure had been retained (Figure 4-6).

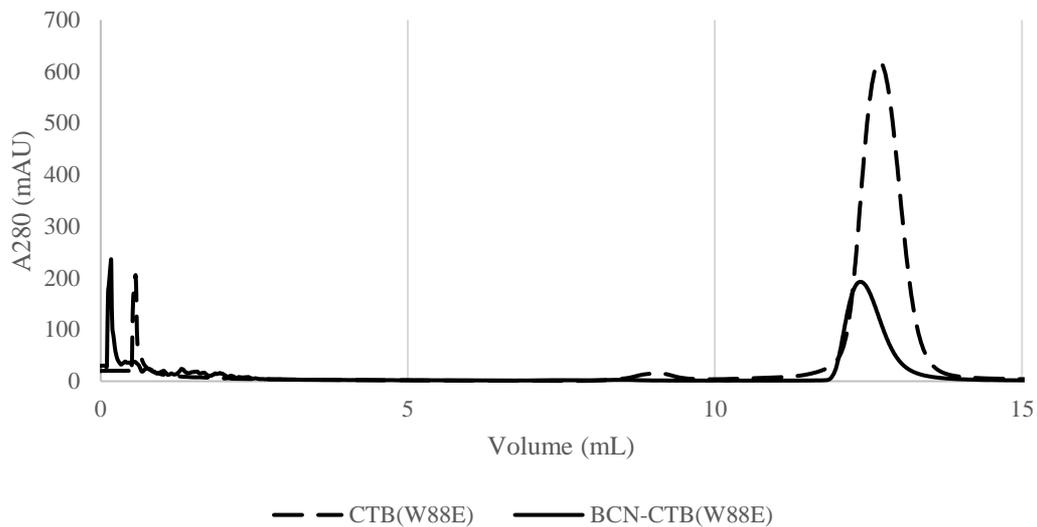


Figure 4-6: Size-exclusion chromatography analysis (Superdex S75 10/300GL) of BCN functionalised CTB(W88E), showing a slight shift in retention volume of BCN-CTB(W88E) compared to CTB(W88E).

Following purification, ES-MS confirmed CTB(W88E) had been quantitatively N-terminally functionalised with a mass increase of 129 Da from the oxidised CTB(W88E) (hydrate) (Figure 4-7c). The mass of BCN-CTB(W88E) was calculated to be 11686.99 Da. Following successful functionalisation of CTB(W88E), the protein was ready to be further functionalised with glycosyl azides *via* strain-promoted azide-alkyne cycloaddition (SPAAC).

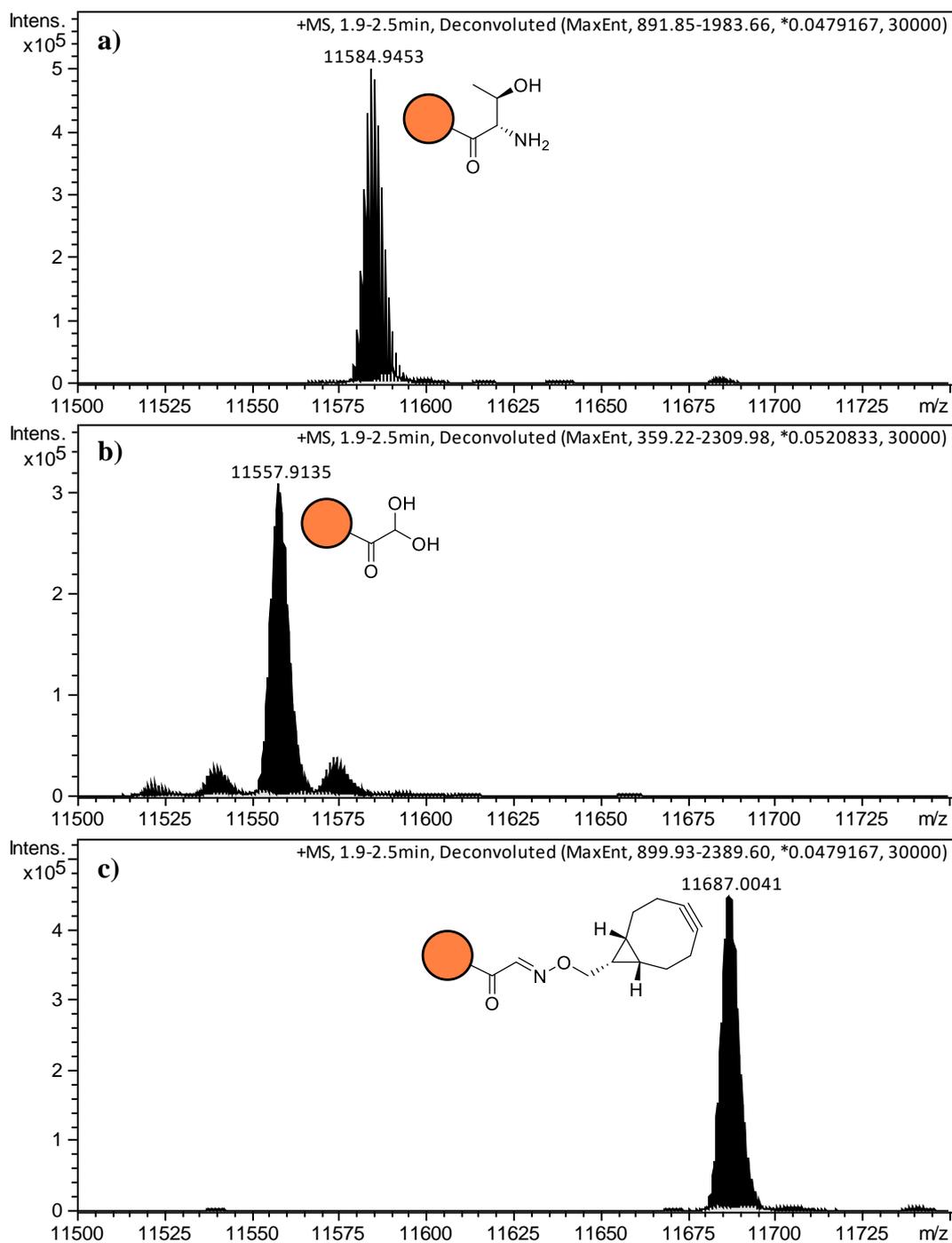


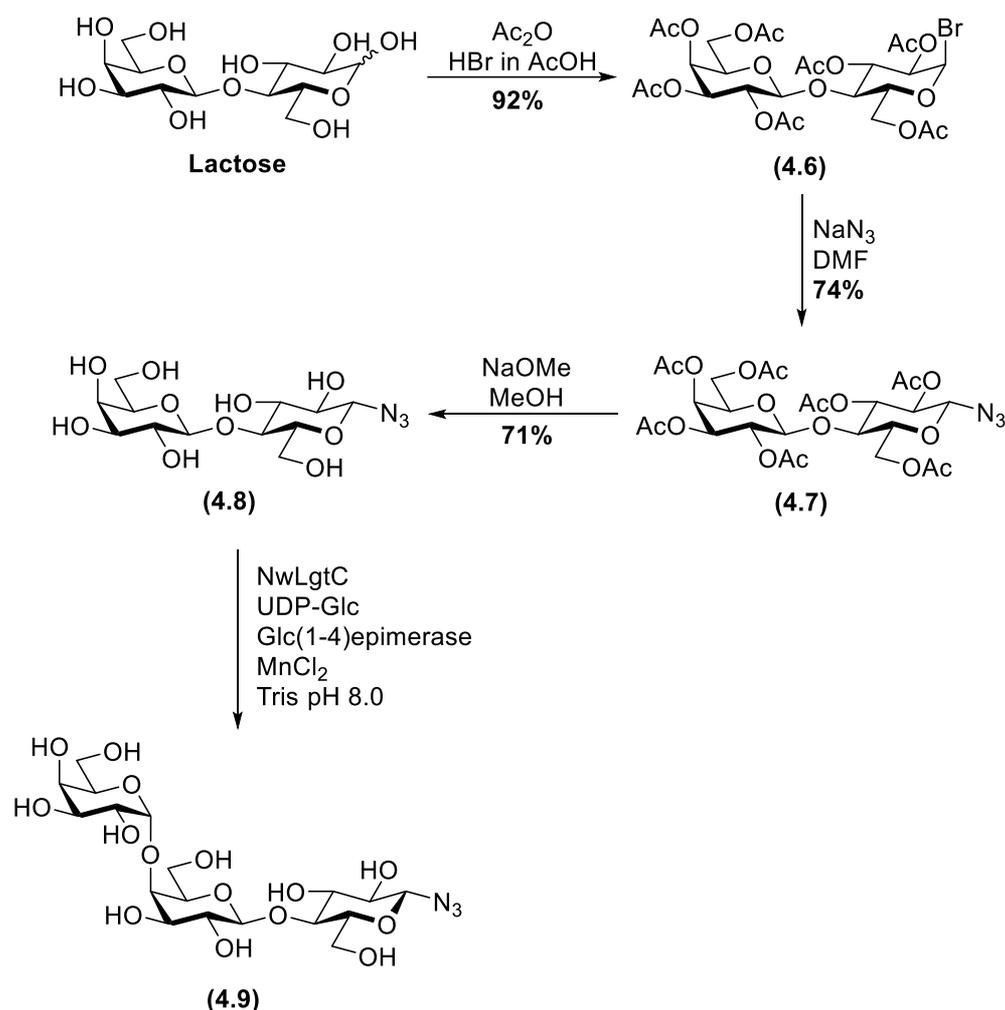
Figure 4-7: Deconvoluted ES-MS of: a) CTB(W88E) before oxidation with NaIO₄. b) CTB(W88E) oxidation reaction after 10 minutes showing quantitative oxidation as hydrate. c) Oxidised-CTB(W88E) BCN oxime ligation after 16hrs, showing quantitative N-terminal labelling with aminoxy-BCN (4.1).

4.3 Preparation of Glycosyl Azides and Functionalization of BCN-CTB(W88E) Through SPAAC

Following on from the preparation of a site-specifically BCN labelled CTB(W88E), a series of glycosyl azides were prepared which could be used for strain-promoted azide-alkyne cycloaddition (SPAAC) onto the CTB scaffold. Three carbohydrates were chosen: the first lactosyl azide, a simple disaccharide and the core to many biologically relevant glycans. The synthetic procedure to prepare lactosyl azide is well reported and would be used for testing of SPAAC ligation reaction.²²⁵ The second was Gb3 azide which could be synthesised enzymatically from lactosyl azide using a Gal α (1-4)Gal transferase. The third was GM1 azide, which would be prepared *via* enzymatic hydrolysis of GM1 ganglioside followed by Shoda's DMC-mediated glycosyl azide formation.²²⁶

4.3.1 Synthesis of Lactosyl azide

Large quantities of commercially available mono- and disaccharide glycans can be simply converted to the corresponding glycosyl azide *via* a straightforward three step process. β -D-Lactosyl azide was synthesised *via* three steps from D-lactose (Scheme 4-3). Firstly, a one-pot acetylation and anomeric bromination of lactose using acetic anhydride and HBr in acetic acid yielded peracetylated lactosyl bromide (**4.6**) in a single step with a yield of 92%.²²⁷ After aqueous workup of **4.6**, the bromide was displaced *via* S_N2 substitution with sodium azide, yielding peracetylated lactosyl azide (**4.7**) in 74% yield after purification by flash column chromatography. Acetyl deprotection using sodium methoxide yielded β -D-lactosyl azide **4.8** in 71% yield after purification by size exclusion using a biogel P2 column eluting with ammonium formate. This provided material of sufficient purity and quantity to be used for testing SPAAC ligation to CTB-BCN(W88E) and for the enzymatic synthesis of β -Gb3 azide (**4.9**) performed by Laia Saltor Nunez (University of Leeds).



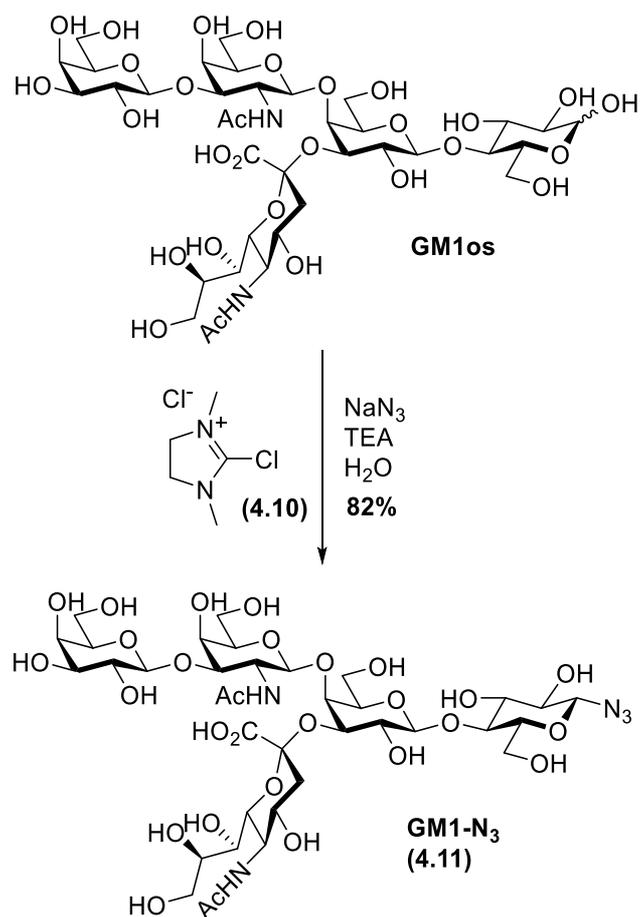
Scheme 4-3: Synthetic route to β -lactosyl azide (4.8) from lactose monohydrate via α -lactosyl bromide (4.6). Enzymatic galactosylation to synthesise Gb3 azide (4.9) performed by Laia Saltor Nunez.

4.3.2 Synthesis of GM1 azide

The three-step preparation of glycosyl azides presented in Section 4.3.1 is typically not appropriate for larger and more complex oligosaccharides as the limited quantity of these glycans typically available is not amenable to multiple protection/deprotection and purification steps. As a result, a simple one step approach was chosen for the preparation of GM1 azide (4.10) from GM1 oligosaccharide (GM1os) using 2-chloro-1,3-dimethylimidazolium chloride (DMC, 4.9) to activate the anomeric hydroxyl group as presented by Shoda and coworkers.²²⁶ Although literature provides a precedence of this methodology being used on complex sialylated glycans, it has not previously been used on ganglioside-derived oligosaccharides. The

original protocol by Shoda requires prolonged reaction times and a high number of equivalents of 2,6-lutidine (base), sodium azide and DMC (**4.10**) to get the reaction to proceed to completion for sialylated glycans.²²⁶ Using alternative conditions (fewer equivalents of the stronger base TEA) proposed by Machida *et al.*,²²⁸ Ryan McBerney had been able to observe a shift to a higher R_f by TLC which indicated the reaction had reached completion within 5 hours.¹⁸⁷ However due to the sub-milligram quantity of the reaction performed, no analytical data could be obtained to confirm conversion of GM1os to GM1 azide.

A 200 mg batch of crude GM1os was available in the lab which had been produced by Dr Ryan McBerney *via* enzymatic hydrolysis of GM1 ganglioside using endoglycosylceramidase (EGCase) II from *Rhodococcus Sp.* (M-777).²²⁹ The GM1os was purified in two steps by reverse phase chromatography to remove any unreacted GM1 ganglioside, and then desalting by size exclusion chromatography using a Biogel P2 column. A milligram scale test reaction was performed following the procedure used by Ryan McBerney, and after 5 hours the products were isolated by size exclusion chromatography. NMR spectroscopic analysis of the resulting GM1 azide (**4.11**) showed the reaction was only ~60% complete based on reduction in intensity of GM1os anomeric signals. Doubling the number of equivalents of TEA and increasing reaction time from 5 hours to 2 days at 37°C resulted in a ~25% increase in conversion (estimated by TLC). Leaving the reaction for more than 2 days or adding further equivalents of sodium azide had no effect to the progress of the reaction.



Scheme 4-4: Synthetic route to β -GM1- N_3 (4.11) from GM1os using Shoda's reagent (DMC, 4.10) and sodium azide.

Following purification by size exclusion chromatography, ^1H NMR spectroscopy confirmed the reduction in intensity of the GM1os Glc anomeric signals. Using variable temperature NMR at 278 K, the residual water signal was shifted to 5.0 ppm and eliminated from the spectrum using presaturation. This allowed for measurement of the chemical shift and coupling constant of GM1-azide Glc anomeric centre confirming the anomeric configuration as beta ($J = 8.7$ Hz) (Figure 4-8). The level of conversion could not be quantified from the VT NMR as presaturation results in alteration of the signal intensities for protons which exchange with water or have a chemical shift near the residual water signal. TLC and ES-MS suggest the reaction had reached ~90% conversion.

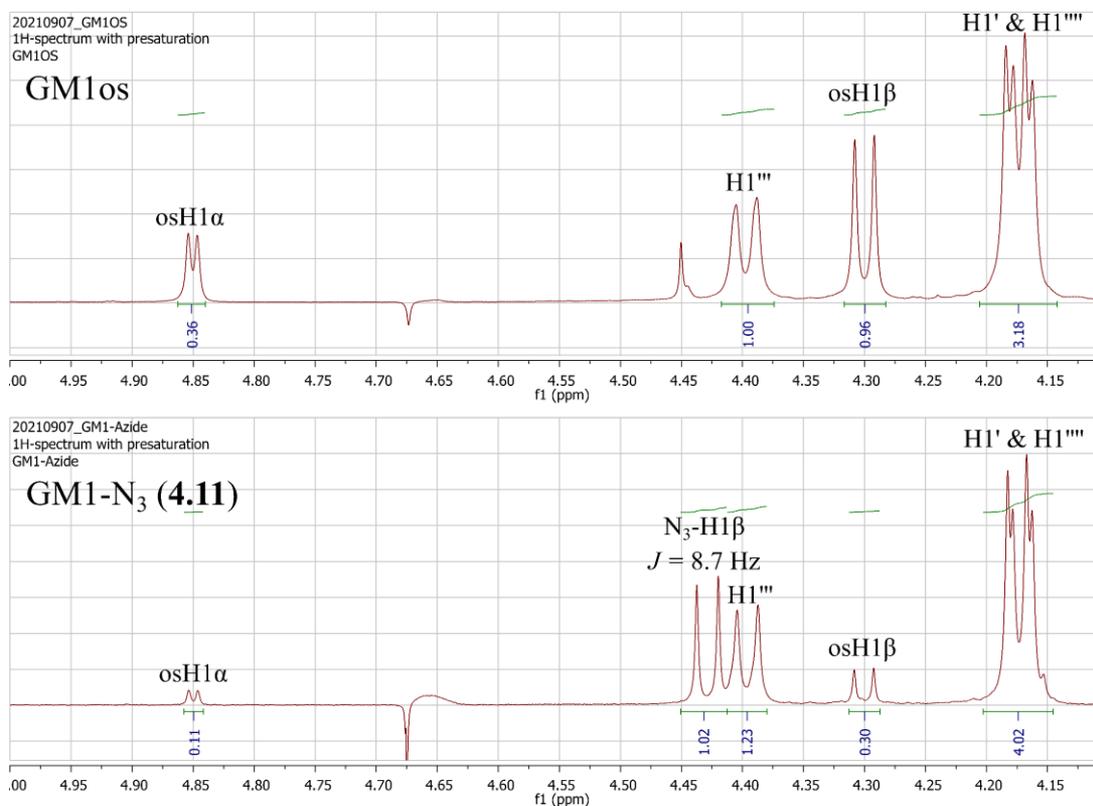
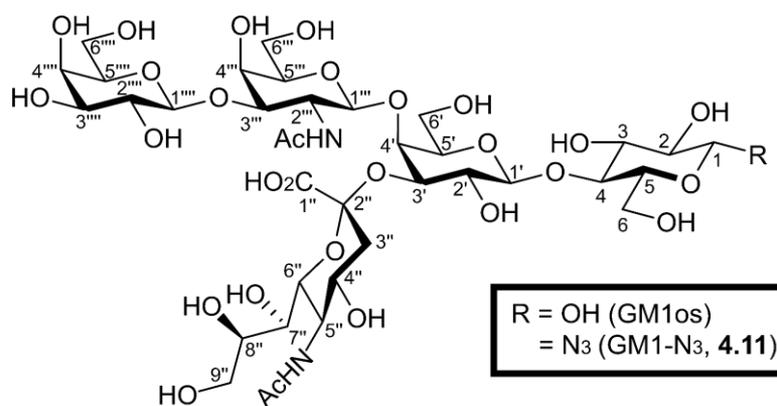


Figure 4-8: Expansion of the anomeric region of NMR spectra for GM1os (top) and GM1 azide (bottom, 4.11). Spectra show appearance of new peak for Glu N₃-H1 β anomeric centre and overall reduction in signal intensity for peaks corresponding to osH1 α/β compared to other anomeric centres (H1', H1'', & H1''').

4.3.3 SPAAC of Glycosyl azides onto BCN-CTB(W88E)

Lactosyl azide (4.8) was used to test the strain-promoted azide-alkyne cycloaddition (SPAAC) with the BCN-functionalised CTB(W88E) and optimise the ligation conditions. Ten equivalents of lactosyl azide was added to BCN-CTB(W88E) and left at room temperature. After 16 hrs, ES-MS showed the absence of the BCN-

CTB(W88E), with lactose-CTB(W88E) being the major species (Figure 4-9). Two other species were present in the spectrum, both of which were believed to be due to fragmentation within the mass spectrometer; the first due to the loss of galactose giving 11893.08 Da, and the second due to the loss of lactose through breakage of anomeric C-N bond to leave CTB-triazole giving a mass of 11730.03 Da.

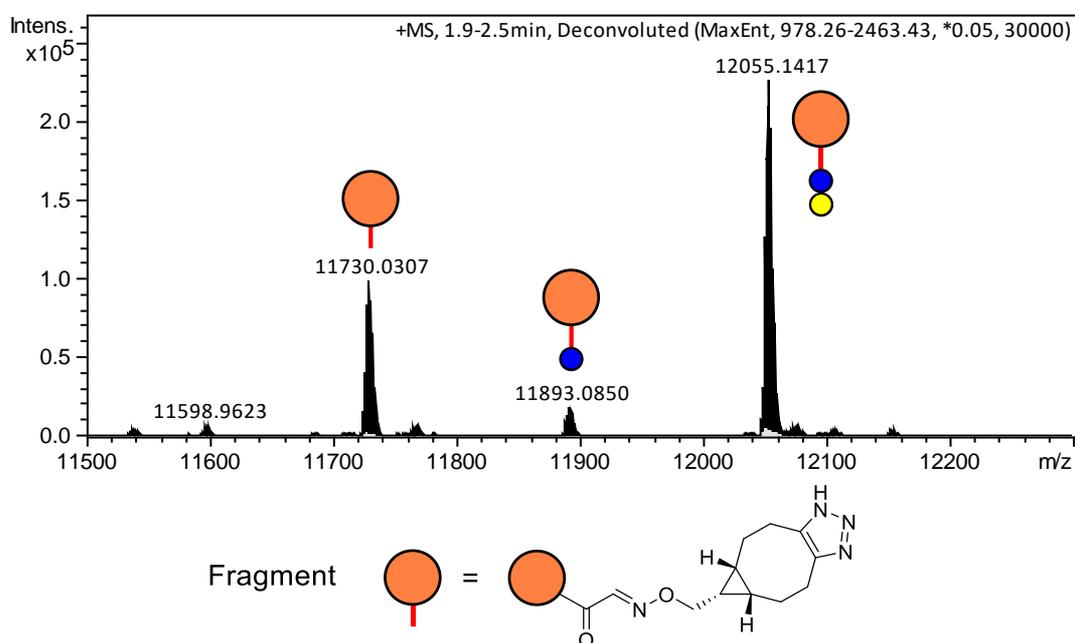


Figure 4-9: Deconvoluted ES-MS for SPAAC ligation of lactosyl azide (**4.8**, 10 eq) to BCN-CTB(W88E) after 16 hrs. The lactosyl triazole group fragments within the spectrometer to lose terminal galactose (11893.09 Da) and lactose (11730.03 Da).

The second peak within the spectrum could also be attributed to the BCN group reacting with any sodium azide which had not been removed during the size-exclusion chromatography/desalting purification or from hydrolysis of the glycosyl azide. To test if this was a possibility BCN-CTB(W88E) was incubated with ten equivalents of sodium azide and ES-MS timepoints taken at 30 mins, 1 hr, 2 hrs, and 4 hrs. The reaction of BCN-CTB(W88E) with lactosyl azide (10 equivalents) was repeated with ES-MS timepoints also taken at 30 mins, 1 hr, 2 hrs, and 4 hrs (Figure 4-10a). As expected, no reaction was observed when BCN-CTB(W88E) (11687.0 Da) was treated with sodium azide (Figure 4-10b). When BCN-CTB(W88E) was treated with lactosyl azide (**4.8**), the appearance of a signal for triazole-CTB (11730.0 Da) and target product lactosyl-CTB (12055.1 Da) in Figure 4-10c indicates that the triazole species is more likely formed by fragmentation of the expected product within the mass spectrometer.

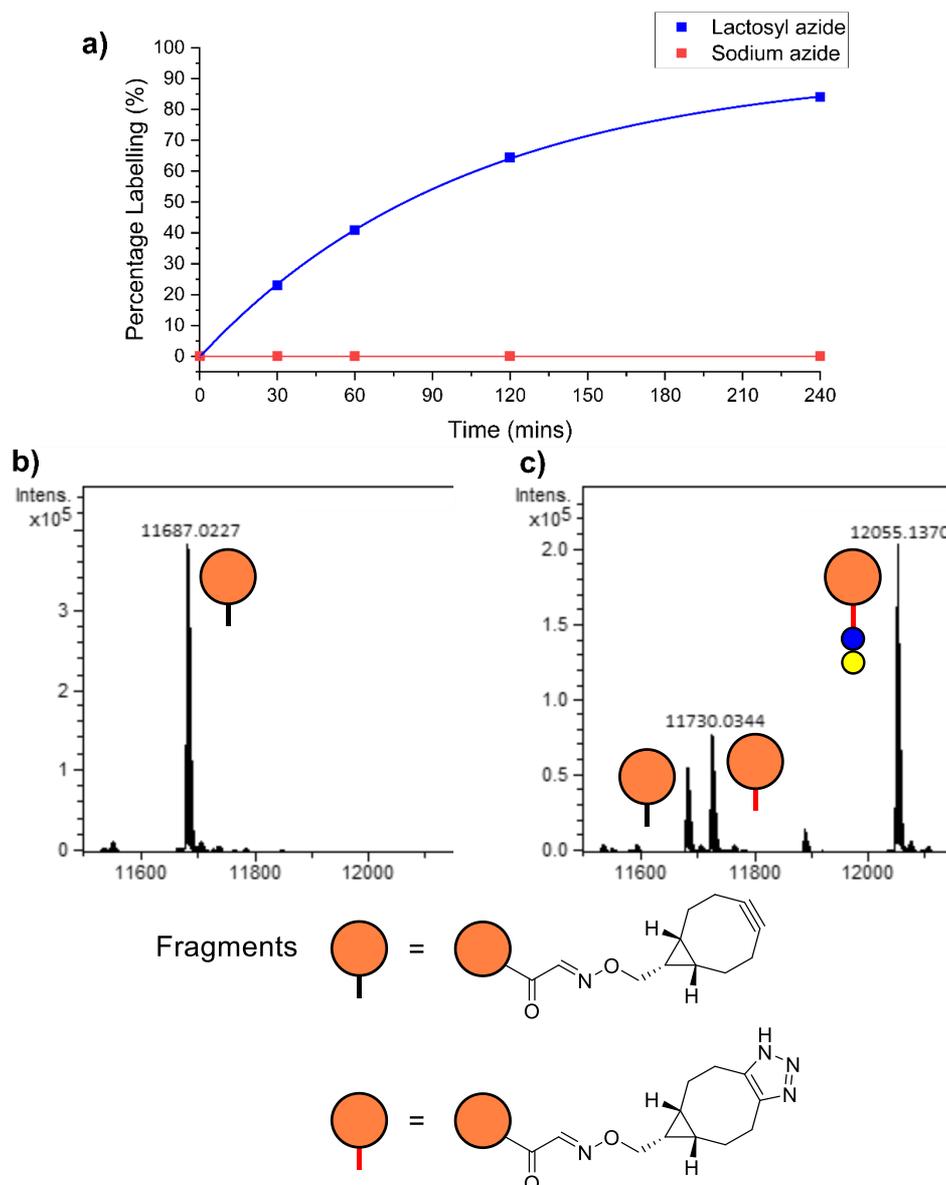


Figure 4-10: a) SPAAC labelling reaction progress over 4 hours for sodium azide and lactosyl azide (4.8). b) Deconvolved ES-MS of BCN-CTB treated with sodium azide (10 eq) for 4 hours. c) Deconvolved ES-MS of BCN-CTB treated with lactosyl azide (10 eq) for 4 hours.

Confident that the triazole-CTB(W88E) was due to fragmentation within the mass spectrometer, attention shifted to labelling CTB(W88E) with Gb3 azide (4.9) and GM1 azide (4.11). SPAAC ligation was performed in an identical manner to the lactosyl azide reaction: BCN-CTB(W88E) was treated with 10 equivalents of the glycosyl azide and incubated at room temperature overnight. ES-MS showed both Gb3 and GM1 could be ligated quantitatively to the modified N-terminus of BCN-CTB(W88E). The mass spectrum of Gb3-CTB(W88E) had fragmentation peaks with

the loss of the terminal galactose, Gal- α (1,4)-Gal and Gb3 (Figure 4-11). The spectrum of GM1-CTB(W88E) showed no fragmentation of the major species, which we concluded was due to the additional stability of the glycan imparted by the sialic acid.

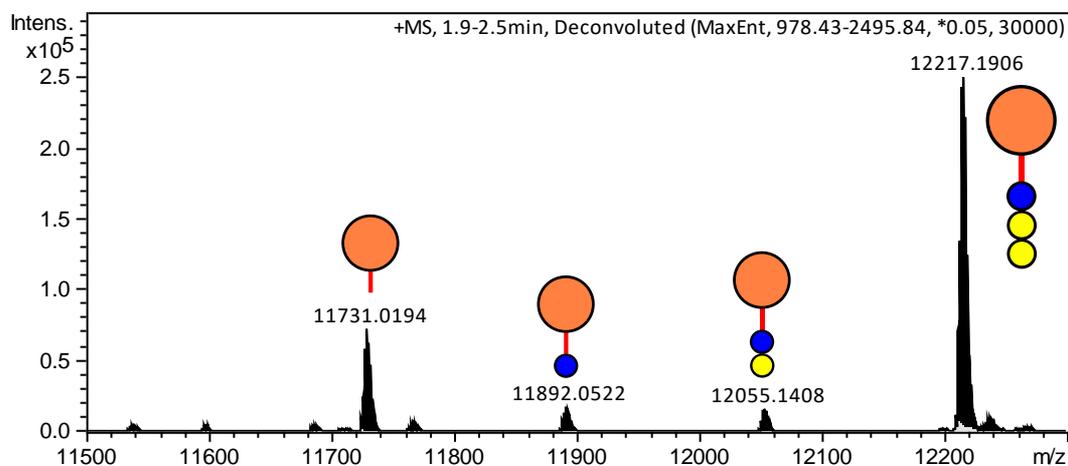


Figure 4-11: Deconvolved ES-MS for SPAAC ligation of Gb3 azide (**4.9**, 10eq) to BCN-CTB(W88E) after 16 hrs. Gb3 triazole moiety fragments within the spectrometer to lose terminal galactose (12055.14 Da), Gal- α (1,4)-Gal (11892.05 Da) and Gb3 (11731.02 Da).

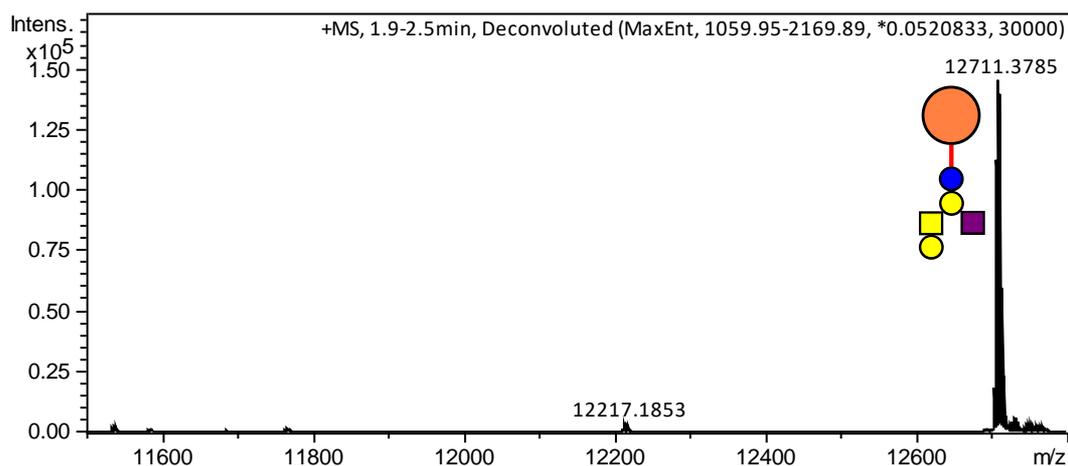


Figure 4-12: Deconvolved ES-MS for SPAAC ligation of GM1 azide (**4.11**, 10eq) to BCN-CTB(W88E) after 16hrs.

4.4 Inhibition Studies by Enzyme-Linked Lectin Assay (ELLA)

An enzyme-linked lectin assay (ELLA) was used to test the inhibitory potency of the prepared neoglycoconjugates against their respective targets. The ELLA protocol is

widely used in the field of glycoconjugate inhibitors for detecting the quantity of uninhibited enterotoxins bound to a microtiter plate.^{105,108,109} Although other methods of determining bacterial toxin inhibition using a functional human intestinal organoid assay are arguably more realistic, the quantity of toxin required changes from assay to assay depending on the cells used.^{230,231} Horseradish peroxidase conjugates of CTB and VTB (CTB-HRP and VTB-HRP) were prepared previously by Dr Ryan McBerney using Lightning-Link[®] horseradish peroxidase kit from Expedeon.¹⁸⁷ The quantity of CTB-HRP/VTB-HRP bound to the microtiter plate can be determined by measuring the fluorescence of resorufin produced from turnover of Amplex Red[®] by HRP in the presence of H₂O₂ (Figure 4-13).

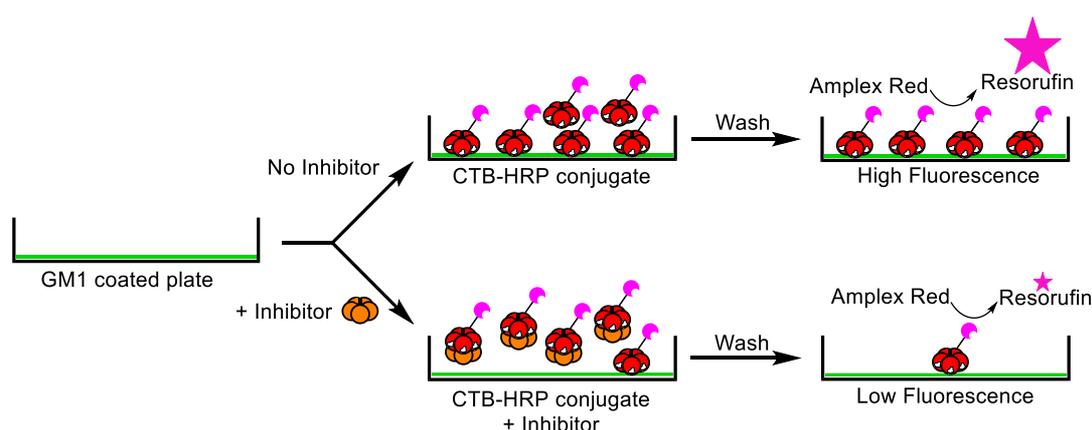


Figure 4-13: Cartoon representation of the enzyme-linked lectin assay (ELLA) for determining inhibition of CTB binding to GM1 coated plates. Addition of inhibitor reduces quantity of CTB-HRP bound to GM1 coated plate, resulting in lower fluorescence.

4.4.1 Inhibition of CTB Binding to GM1

The optimal concentration of CTB-HRP for the ELLA was determined by preparing a serial dilution of CTB-HRP from 50 ng/mL to 0.1 ng/mL, which was added to the wells of a ganglioside GM1-coated plate. After incubating for 30 minutes and washing, a solution of Amplex Red[®] and H₂O₂ was added, and the fluorescence recorded every 30 seconds for 30 minutes (Figure 4-14). The assay showed CTB-HRP prepared previously by Dr Ryan McBerney remained active and could be used at concentrations much lower than those used in previous studies,^{108,187} and still produce a measurable difference in rate from baseline fluorescence as calculated from the

linear trendline. Concentrations higher than 10 ng/mL reach a plateau, indicating a depletion of the Amplex Red®/H₂O₂, within the 30 minutes plate reading. The data for experiments with less than 1 ng/mL CTB-HRP could be fitted to a straight line (R-squared > 0.9998).

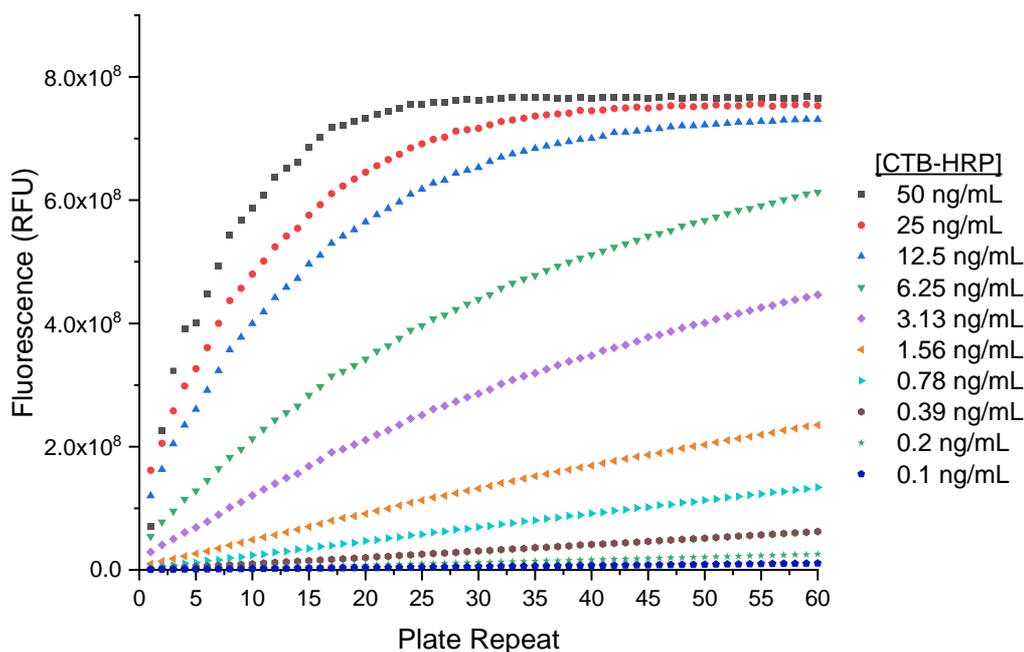


Figure 4-14: Comparing the rate of increase in fluorescence intensity following the addition of Amplex Red®/H₂O₂ to CTB-HRP. CTB-HRP prepared following a 2-fold serial dilution from 50 ng/mL. Data points collected at 30 second intervals.

Excitation 531 nm; Emission 595 nm.

Having confirmed the CTB-HRP conjugate was capable of binding to ganglioside GM1-coated microtiter plates and remained active after long term storage, it was then used to study the inhibition of CTB with the GM1 based neoglycoconjugate inhibitor, GM1-CTB(W88E). The assay included wells for positive control (containing only CTB-HRP) used to measure the maximum expected fluorescence in the absence of inhibitors, and wells for negative control (containing no CTB-HRP or inhibitor) used to measure the background fluorescence to allow normalisation of the data. CTB-HRP was used at a final concentration of 0.5 ng/mL, resulting in a 5-fold increase in sensitivity compared to previous assays performed by Ryan McBerney and Tom Branson.^{108,187} A 2-fold serial dilution of the GM1-CTB(W88E) inhibitor was premixed with CTB-HRP before being added to a ganglioside GM1-coated plate. After washing the plate, Amplex Red and hydrogen peroxide were added, and the rate of fluorescence increase was recorded over 5 minutes (10 measurements).

Measurements were performed in triplicate and averaged for analysis (three technical repeats); in some cases, anomalous data points were removed leaving duplicate results. Data was fitted using OriginPro logistic fitting equation and used to determine IC₅₀ values (Equation 4-1).

$$y = \frac{A_1 - A_2}{1 + (x/x_0)^p} + A_2$$

Equation 4-1: Equation for logistic curve fitting in OriginPro 2019b. A_1 is the curve's maximum, A_2 is the curve's minimum, x_0 is equal to the $\log(\text{IC}_{50})$, x is the $\log(\text{inhibitor concentration})$, and p is the Hill slope parameter.

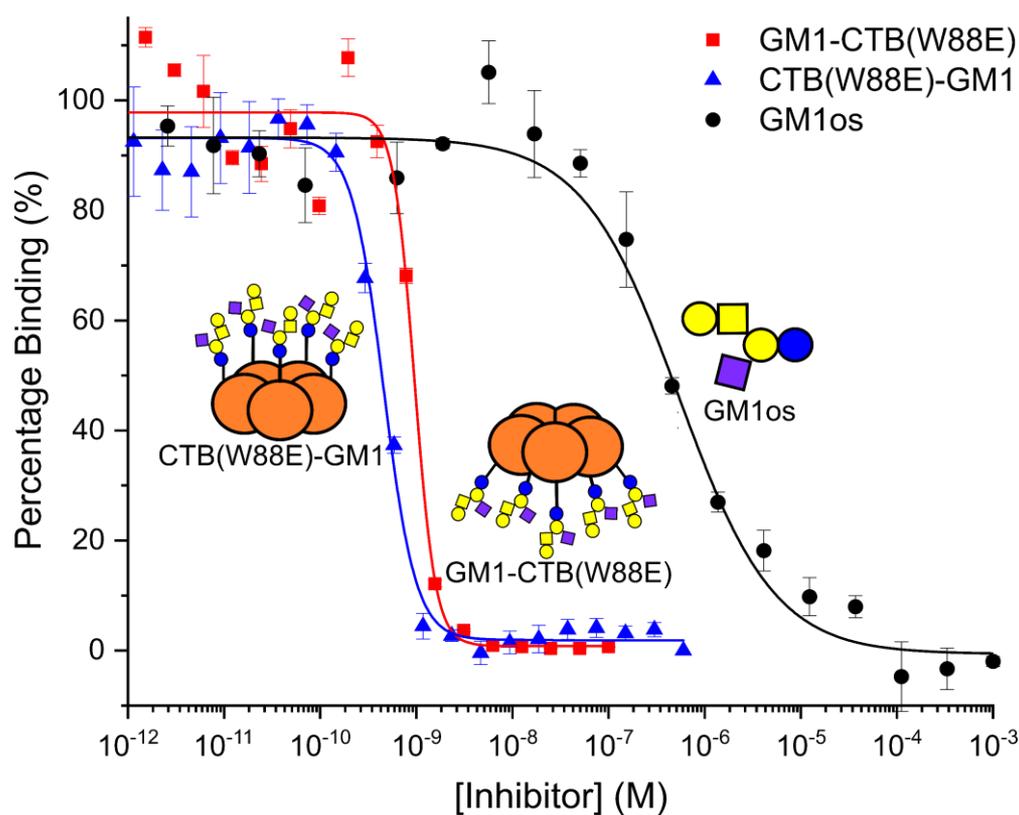


Figure 4-15: Enzyme-linked lectin assay (ELLA) indicates inhibitory potential of multivalent GM1-CTB(W88E). ELLA inhibition data for CTB(W88E)-GM1 reported by Ryan McBerney.¹⁸⁷ ELLA inhibition data for GM1os reported by Branson et al.¹⁰⁸ Error bars indicate the standard error of the percentage binding. No fixing of parameters was required for fitting of the curves to reach convergence.

Table 4-1: Comparison of the IC₅₀ values for GM1os, GM1(CH₂)₁₁CTB(W88E), CTB(W88E)-GM1, GM1-CTB(W88E), their valences and relative potency per GM1.

Inhibitor	Valency	Log (IC ₅₀)	IC ₅₀ (nM)	Relative Potency (per GM1) ^a
CTB(W88E) ^b	-	-	-	-
GM1os ^c	1	-6.27±0.04	530	1 (1)
GM1(CH ₂) ₁₁ CTB(W88E) ^b	5	-9.98±0.08	0.104	5096 (1019)
CTB(GM1, W88E) ^d	5	-9.34±0.02	0.457	1160 (232)
GM1-CTB(W88E)	5	-9.03±0.07	0.924	574 (115)

^a Potency measured relative to monovalent GM1os. ^b No inhibition detected for CTB(W88E) scaffold by Branson *et al.*¹⁰⁸ ^c Inhibitor reported by Branson *et al.*¹⁰⁸ GM1 ligated to N-terminus *via* flexible linker and oxime ligation. ^d Inhibitor reported by Ryan McBerney; GM1 ligated to top face of azido CTB(W88E).¹⁸⁷

Neoglycoprotein inhibitor GM1-CTB(W88E) was found to be a potent inhibitor of CTB adhesion with an IC₅₀ of 924 pM (Figure 4-15, Table 4-1). The inhibitor GM1-CTB(W88E) has 115-fold greater potency per GM1 when compared to monovalent GM1os measured by Branson *et al.* using an assay of the same design.¹⁰⁸ The IC₅₀ value of 924 pM is a 9-fold loss of potency compared to the inhibitor with a longer linker produced by Tom Branson (104 pM), and 2-fold reduction compared to the analogous CTB(W88E)-GM1 inhibitor produced by Ryan McBerney (457 pM) where the GM1 was attached to the top face of CTB rather than to its N-termini.

Conclusions drawn from this data shown that reduction in linker length between the scaffold and glycan and moving the site of glycosylation has only a small impact on potency. Hypothetically one could envisage a decrease in IC₅₀ of GM1-CTB(W88E) and CTB(W88E)-GM1 being a result of numerous reasons: an enthalpic penalty upon bringing two pentameric proteins into close proximity; increased electrostatic repulsions for the same reasons; or that the shorter, more rigid linker results in strained conformation for the glycan to reach the binding site. The use of a neoglycoprotein approach to toxin inhibition in addition to the development of oxime and SPAAC reactive glycoconjugates which do not require extensive chemoenzymatic synthesis, making this a more accessible and versatile approach.

4.4.2 Inhibition of VTB Binding to Gb3

Unlike the fairly standardised CTB ELLA where CTB is captured onto GM1-ganglioside coated plate, there appears to be little constancy for VTB ELLA. Bundle *et al.* utilised two different assays to measure inhibitor potency.¹⁰⁴ In the first, the inhibitor and P^k trisaccharide-BSA-biotin conjugate compete for VTB adsorbed onto a microtiter plate. The quantity of P^k trisaccharide bound is measured using a streptavidin-HRP conjugate and 3,3',5,5'-tetramethylbenzidine (TMB). For the second assay, the inhibitor and VTB are mixed and the quantity of uninhibited VTB bound to synthetic C₁₆ P^k trisaccharide coated plate measured using rabbit anti-VTB antibody and HRP-labelled anti-rabbit antibody. Recently Pieters *et al.* reported the use of a commercially available synthetic glycolipid in an ELLA inhibition assay involving VTB.¹⁰⁹ FSL-Gb3, **4.12**, is composed of Gb3 conjugated to dioleoylphosphatidylethanolamine (DOPE) *via* an O(CH₂)₃NH linker (Figure 4-16). The inclusion of the linker would provide additional flexibility between the glycan and the plate which would be helpful in facilitating a favourable conformation for binding to VTB. The assay used an HisProbe-HRP conjugate to detect the quantity of uninhibited his-tagged VTB bound to FSL-Gb3.¹⁰⁹ To try standardise the CTB and VTB ELLA protocols used within the Turnbull group, FSL-Gb3 was used to coat microtiter plates and a VTB-HRP conjugate used instead of VTB and a secondary HRP-antibody conjugate.

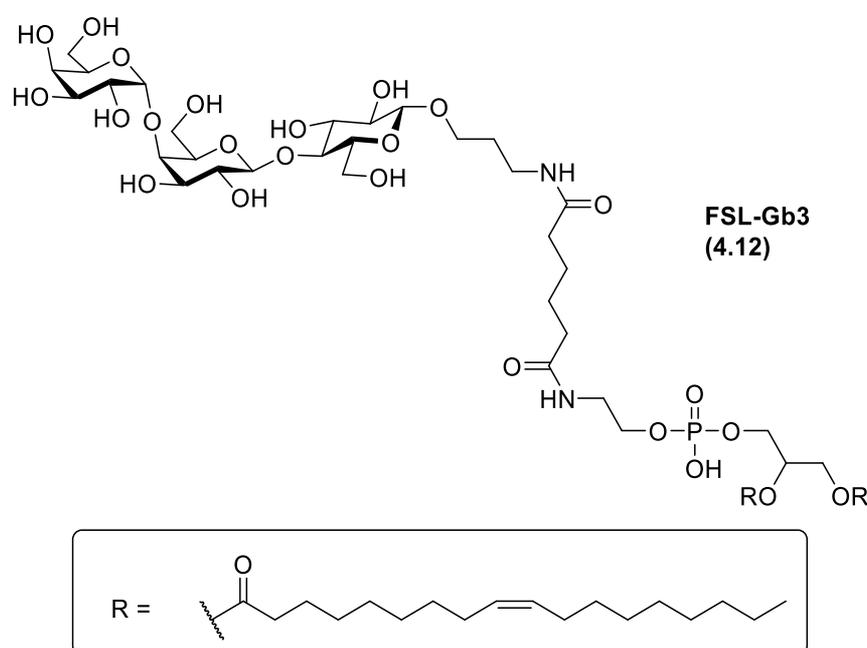


Figure 4-16: Structure of FSL-Gb3 (**4.12**) from Merck.

4.4.2.1 Optimisation of VTB-HRP Capture on Microtiter Plates

FSL-Gb3 was adsorbed onto microtiter plates as described by Pieters *et al.*¹⁰⁹ using a 2 µg/mL (1.39 µM) solution of FSL-Gb3 in PBS at room temperature for 3 hours. Plates were washed with PBS, blocked with BSA, and washed again. The ability of the adsorbed Gb3 to capture VTB was tested by applying a 2-fold serial dilution of VTB-HRP (500 ng/mL maximum concentration) to the plate. After incubating for 30 minutes and washing, a solution of Amplex Red[®] and H₂O₂ was added, and the fluorescence was recorded every 30 seconds (Figure 4-17). This preliminary assay confirmed FSL-Gb3 was successfully adsorbed onto microtiter plates and was capable of capturing VTB-HRP on the plate. From plotting and fitting of the data it was concluded that a good level of signal-to-noise was obtained in the region of 4-15 ng/mL of VTB-HRP. A VTB-HRP concentration of 10 ng/mL was adopted for the subsequent ELLA inhibition assays.

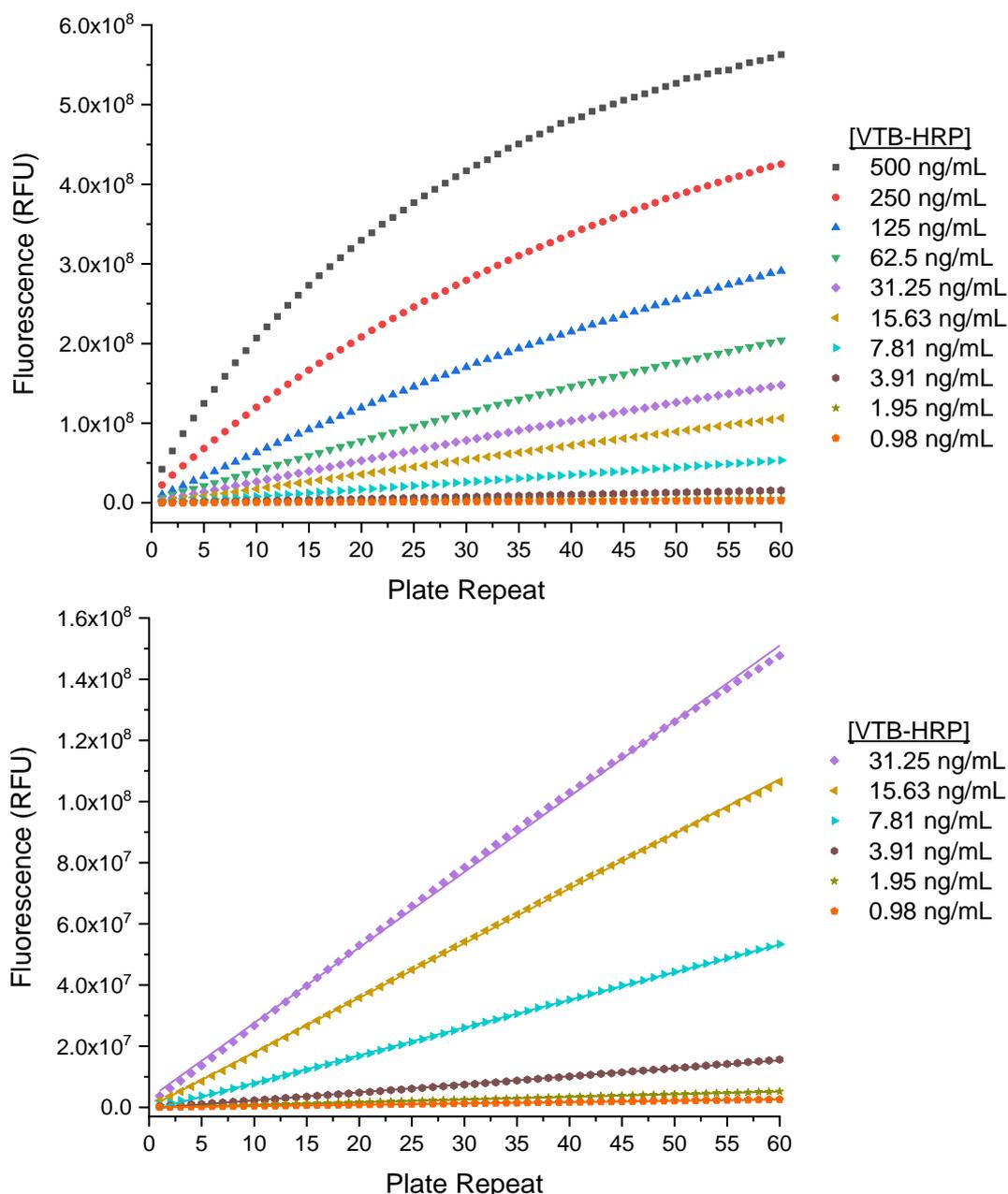


Figure 4-17: Fluorescence intensity for the concentration gradient assay of VTB-HRP following a 2-fold serial dilution from 500 ng/mL, binding to a plate coated with FSL-Gb3 4.12, and the addition of Amplex Red[®]/H₂O₂. Data points collected at 30 second intervals. Excitation 531 nm; Emission 595 nm.

4.4.2.2 Enzyme-Linked Lectin Assay of Pentavalent VTB Inhibitors

Having confirmed FSL-Gb3 coated microtiter plates could bind VTB-HRP and the VTB-HRP prepared by Ryan McBerney remained active after long-term storage, measurement of the VTB inhibitor potency was performed using a similar assay

protocol previously optimised for CTB inhibitors. The assay included wells for positive control (containing only VTB-HRP) used to measure the maximum expected fluorescence and wells for negative control (containing no VTB-HRP or inhibitor) used to measure the background fluorescence and normalisation of the data. VTB-HRP was used at a final concentration of 10 ng/mL.

Three compounds were to be tested for the initial screening, the first was Gb3 oligosaccharide which was to be used for comparison of inhibitor potency, prepared by Laia Saltor Nunez (University of Leeds). The second was Gb3-CTB(W88E) where Gb3 azide (**4.9**) was ligated to the N-terminus of CTB(W88E), as described in section 4.3.3. The third construct was Gb3-BCN ligated to azido-CTB(W88E), produced by Ryan McBerney.¹⁸⁷ A 2-fold serial dilution of Gb3os, Gb3-CTB(W88E) and CTB(W88E)-Gb3 were premixed with VTB-HRP (10 ng/mL) before being added to FSL-Gb3 coated plates. After washing the plate, Amplex Red and hydrogen peroxide were added, and the rate of fluorescence increase was recorded over 5 minutes (10 measurements). Measurements were performed in triplicate and averaged for analysis (three technical replicates); in some cases, anomalous data points were removed leaving duplicate results. Data were fitted using the OriginPro logistic fitting equation (Equation 4-1) and used to determine IC₅₀ values.

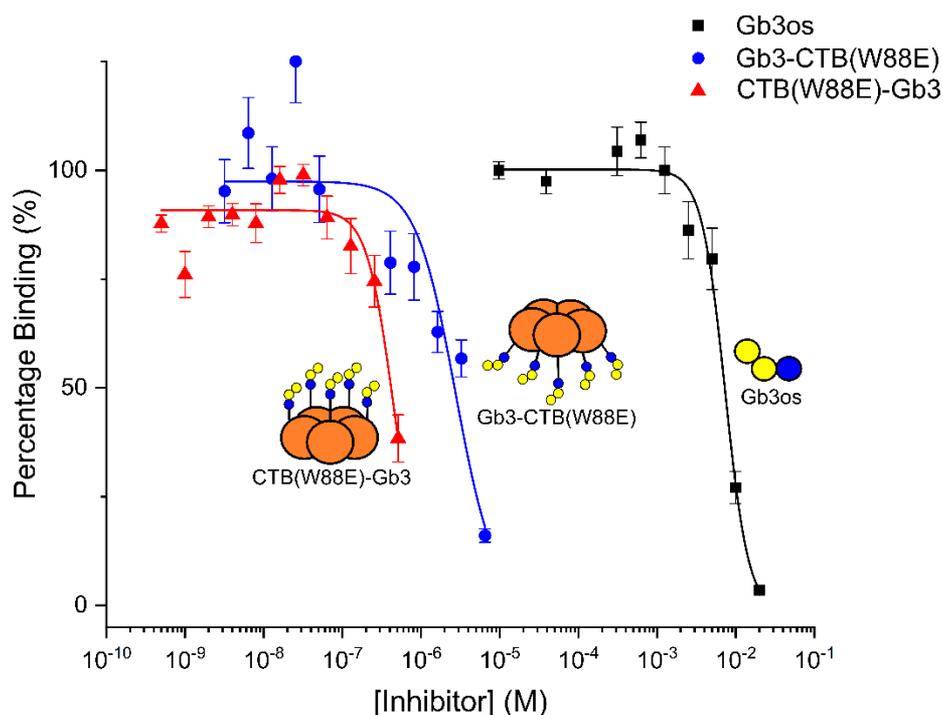


Figure 4-18: Enzyme-linked lectin assay (ELLA) indicates inhibitory potential of multivalent Gb3-CTB(W88E) and CTB(W88E)-Gb3 inhibitors compared to monovalent Gb3os. Error bars indicate the standard error of the percentage binding. Fitting of the curve for Gb3-CTB(W88E) required fixing of the bottom of the asymptote to 0 (zero) for fitting to reach convergence. No parameters required fixing for Gb3os or CTB(W88E)-Gb3 to reach convergence.

Table 4-2: Comparison of the IC_{50} values for Gb3os, CTB(W88E)-Gb3, Gb3-CTB(W88E), their valences and relative potency per Gb3.

Inhibitor	Valency	Log (IC_{50})	IC_{50} (μ M)	Relative Potency (per Gb3) ^a	Hill coefficient (p)
Gb3os ^b	1	-2.13±0.04	7340	1 (1)	2.90±0.75
Gb3-CTB(W88E)	5	-5.57±0.18	2.67	2749 (550)	1.72±0.37
CTB(W88E)-Gb3 ^c	5	-6.38±0.50	0.417	17,602 (3520)	2.56±0.1.22

^a Potency measured relative to monovalent Gb3os. ^b Gb3os prepared by Laia Saltor Nunez. ^c Inhibitor prepared by Ryan McBerney; Gb3-BCN ligated to top face of azido CTB(W88E).¹⁸⁷

Gb3 oligosaccharide was found to have an IC_{50} of 7.34 mM (Figure 4-18, Table 4-2). The ELLA assay suggests both the neoglycoprotein inhibitors are potent inhibitors of

VTB binding to Gb3. Attachment of Gb3 azide to the N-terminus of CTB (Gb3-CTB(W88E)) results in an inhibitor with an IC_{50} of 2.67 μ M, a 550-fold greater potency per Gb3 compared to monovalent Gb3os. The inhibitor synthesised by Ryan McBerney in which Gb3 was attached to the top face of azido-CTB(W88E), CTB(W88E)-Gb3 is estimated to have an IC_{50} of 417 nM; a 3520-fold improvement in potency per Gb3os compared to monovalent Gb3. Due to the limited quantity of this inhibitor produced, evaluation of inhibition at higher concentrations was not possible. Fitting of the curve for the CTB(W88E)-Gb3 did not require fixing of any parameters to reach convergence; fitting of the curve for Gb3-CTB(W88E) required fixing of the bottom of the asymptote to 0 (zero) for fitting to reach convergence.

Previously a 2-fold improvement in inhibition was seen if GM1 was attached to the top face of CTB compared to the N-terminus. It could therefore be expected that the inhibitor where Gb3 was attached to the top-face of azido-CTB might be between 2 and 10-fold greater compared to Gb3 attached to the N-terminus. In fact, a 6-fold improvement in inhibition is seen if Gb3 is attached to the top-face of azido-CTB compared to the N-terminus of CTB(W88E). This is likely due to closer spacing of Gb3 ligands which better matches the spacing between Gb3 binding sites on the VTB pentamer, resulting in a less strained conformation.

Although the expected improvement in IC_{50} for CTB(W88E)-Gb3 compared to Gb3-CTB(W88E) is within the range that could be anticipated in analogy with the CTB inhibition data, and both inhibitors show significant improvements in IC_{50} compared to monovalent Gb3os, further work is required before more concrete conclusions can be drawn. The production of both pentavalent inhibitors at higher concentration is required for evaluation of inhibition at higher concentrations.

4.4.2.3 Enzyme-Linked Lectin Assay of Cyclic Glycopeptide VTB Inhibitors

Neoglycoprotein inhibitors described so far have only had five glycans attached, but the crystal structure of the VTB-Gb3os complex shows 15 potential binding sites.⁸⁹ To investigate if binding affinity could be improved further the number of Gb3 groups was increased. Cyclic glycopeptides displaying three Gb3 glycans were designed and prepared by Dr Vajinder Kumar (University of Leeds). Attachment to the CTB scaffold would result in decapentameric display of Gb3, equating to one Gb3 per VTB

pentamer binding site. Two functionalised glycopeptides were provided by Dr Vajinder Kumar: one with a BCN group (**4.13**) for SPAAC to azido-CTB(W88E) and a second with an aminoxy functional group (**4.14**) for oxime ligation to CTB(W88E). However, attempts to attach each of the glycopeptides to their respective scaffolds was found to result in precipitation of the protein. As a result, the cyclic glycopeptide (**4.14**) was tested without being attached to the CTB scaffold as an inhibitor of VTB adhesion.

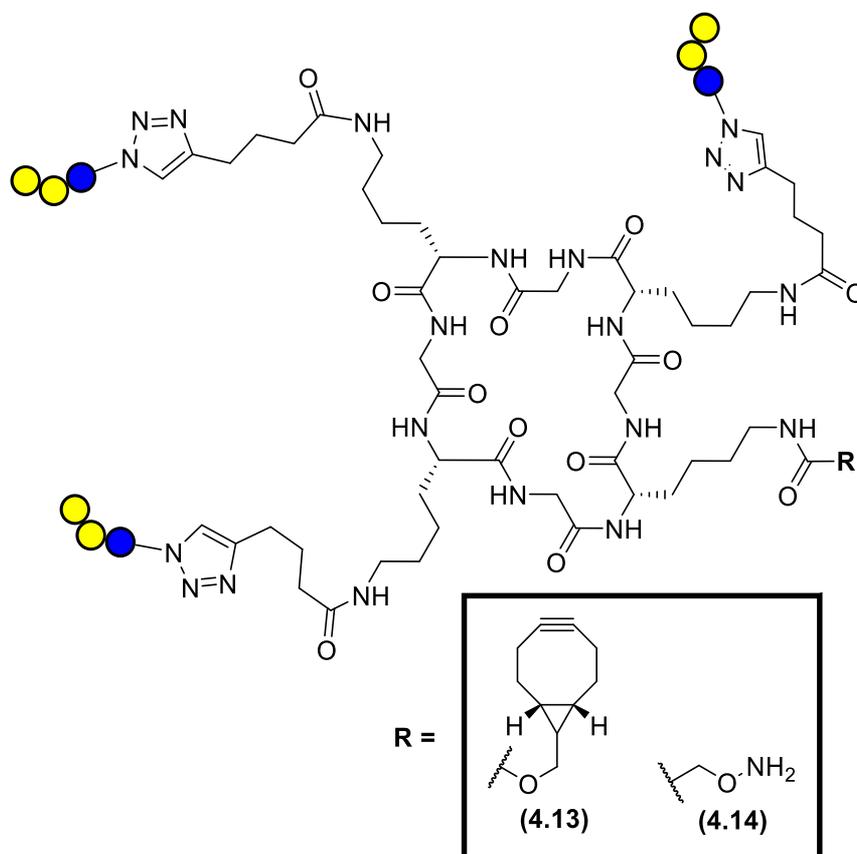


Figure 4-19: Structure of BCN (**4.13**) and aminoxy (**4.14**) functionalised cyclic Gb3 glycopeptide inhibitors prepared by Dr Vajinder Kumar.

Testing of inhibitor **4.14** found it had an IC₅₀ of 3.67 μ M, a 667-fold improvement in potency per Gb3os compared to monovalent Gb3. The inhibition is equivalent to that observed for the pentameric display of Gb3 ligated to the N-terminus of CTB. Further work is underway within the Turnbull group to elucidate the binding mode of this inhibitor. One potential binding mode which may explain the strong inhibition observed may be the glycopeptide bridging two VTB pentamers in a similar fashion to the STARFISH inhibitor developed by Bundle *et al.*¹⁰⁴ (section 1.3.3.1).

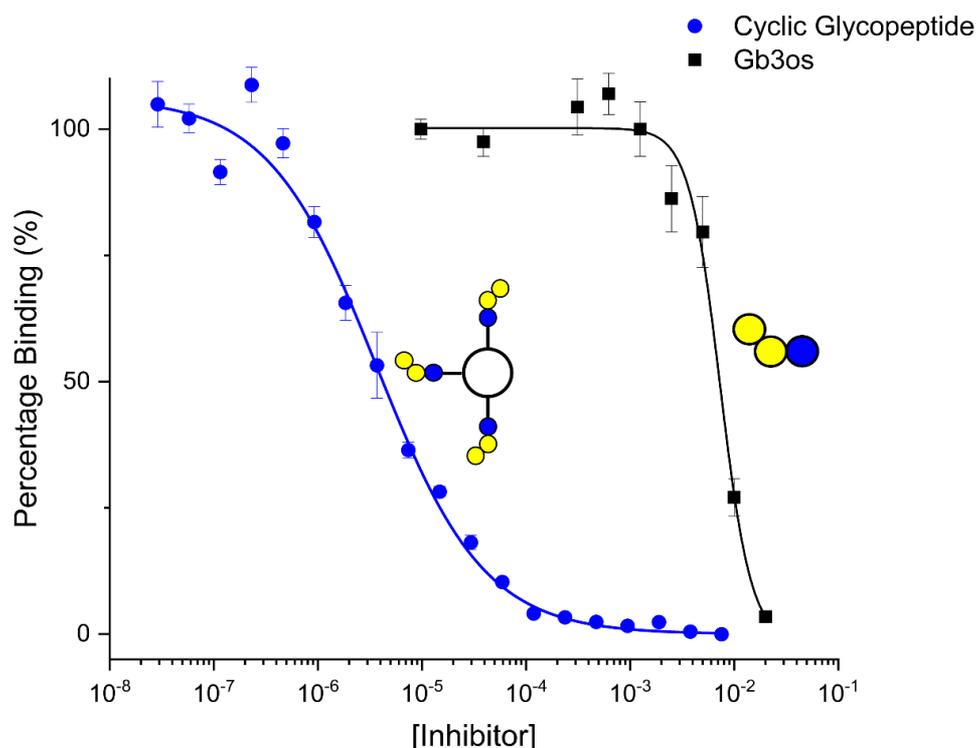


Figure 4-20: Enzyme-linked lectin assay (ELLA) indicates inhibitory potential of multivalent Gb3 cyclic glycopeptide inhibitor compared to monovalent Gb3os. Error bars indicate the standard error of the percentage binding. No parameters required fixing for Gb3os or cyclic glycopeptide inhibitor to reach convergence.

Table 4-3: Comparison of the IC_{50} values for Gb3os, Cyclic Gb3 glycopeptide (4.14), their valences and relative potency per Gb3.

Inhibitor	Valency	Log (IC_{50})	IC_{50} (μ M)	Relative Potency (per Gb3) ^a	Hill coefficient (p)
Gb3os	1	-2.13±0.04	7340	1 (1)	2.90±0.75
Cyclic Gb3 glycopeptide (4.14) ^b	3	-5.44±0.30	3.67	2000 (667)	0.82±0.08

^a Potency measured relative to monovalent Gb3os. ^b Inhibitor 4.14 prepared by Vajinder Kumar.

4.5 Chapter Conclusions

A bicyclononyne-functionalised, non-GM1-binding mutant of CTB has been successfully synthesised and used to produce a series of multivalent neoglycoprotein inhibitors of CTB and VTB adhesion. Following expression and purification of the

CTB mutant, the N-terminus was quantitatively labelled with an aminoxy functionalised BCN linker (**4.2**) *via* oxime ligation. From this common BCN functionalised scaffold, SPAAC ligation proved a very efficient and rapid method of synthetic glycosylation using a series of glycosyl azides synthesised chemically and enzymatically. Quantitative SPAAC modification of the scaffold with each glycosyl azide allowed for easy access to well-defined site-specifically labelled neoglycoconjugates.

The GM1-ligated neoglycoprotein (GM1-CTB(W88E)) exhibited sub-nanomolar inhibition of CTB adhesion (924 pM), showing only a 2-fold difference in comparison to the previously reported CTB(W88E)-GM1 inhibitor produced by Ryan McBerney using the same linker but a different site of glycosylation.¹⁸⁷ Compared to the inhibitor produced by Branson *et al.*¹⁰⁸ which has a longer linker and the same site of glycosylation only a 9-fold loss of inhibition was observed. It can be concluded that preordering and the location of the GM1 ligands has a modest impact on potency, with the minimal length heterodivalent linker devised by Ryan McBerney still sufficient to present the GM1 ligands to the binding site. Bringing the two pentamers into closer proximity may have greater impact on observed inhibitory potency than the site of glycosylation.

Further pentavalent neoglycoproteins bearing Gb3 ligands have also shown the potential to be very potent inhibitors of VTB adhesion. A VTB ELLA protocol analogous to the one previously used for CTB to test these neoglycoprotein inhibitors has been established, tested and optimised using a commercially available FSL-Gb3. The Gb3-ligated neoglycoprotein (Gb3-CTB(W88E)) exhibited micromolar inhibition of VTB adhesion (2.67 μ M), a 550-fold improvement in inhibition compared to monovalent Gb3. The neoglycoprotein inhibitor produced by Ryan McBerney (CTB(W88E)-Gb3) was estimated to have sub-micromolar inhibition, a 6-fold improvement compared to the inhibitor produced using the same linker but a different glycosylation site. However, only limited amounts of this inhibitor were available for testing, so further work is required to acquire a full inhibition curve with datapoints at higher concentrations. Although the cyclic glycopeptide inhibitors produced by Vajinder Kumar were found to result in protein precipitation upon conjugation, inhibitor **4.14** was found to exhibit micromolar inhibition of VTB adhesion of 3.67 μ M. This inhibition is comparable to pentameric display of Gb3

ligated to the N-terminus of CTB. Depending on the binding mode of this inhibitor, this may suggest if the cyclic glycopeptides could be ligated to a pentameric scaffold they may prove to be very potent inhibitors of VTB, offering superior inhibition compared to the pentameric inhibitors. In their current state, the advantages of the pentameric neoglycoproteins over cyclic glycopeptides is clear in terms of ease of synthesis and more potent inhibition.

Chapter 5: Moving Toward Non-bacterial Pentameric Protein Scaffolds for Multivalent Neoglycoprotein Inhibitors

As previously demonstrated, cholera toxin B-subunit is a convenient, easy to express/manipulate, highly stable scaffold for pentameric neoglycoconjugate inhibitors of bacterial toxins. However due to the risk of invoking a host immune response using bacterial toxin proteins, an alternative scaffold would be desirable to move this class of neoglycoconjugates to a more viable strategy. This chapter will discuss the ongoing work looking at using a class of pentameric human proteins known as pentraxins as scaffolds for multivalent neoglycoprotein inhibitors of bacterial toxins.

5.1 Background

Pentraxins are an evolutionary conserved family of proteins involved in acute immunological responses.^{232,233} All members of the family are characterised by pentameric radial symmetry, a high degree of sequence homology and calcium-dependent binding to a wide array of ligands including phosphate esters, polysaccharides, and polyelectrolytes.^{233,234} Initially serum amyloid P (SAP) component was investigated as a potential scaffold for synthetic glycosylation. SAP's homopentameric structure and pentagonal symmetry make it an ideal candidate for multivalent neoglycoprotein based therapeutics which target AB₅ toxins. Models produced *in silico* by Ryan McBerney (University of Leeds) showed insertion of a C-terminal sortase motif and modification with Gb3-containing glycopeptide would allow binding to all three VTB binding sites (Figure 5-1). Further modelling confirmed modification of the C-terminus closely match the spacing and symmetry of the VTB pentamer and Gb3 binding sites.

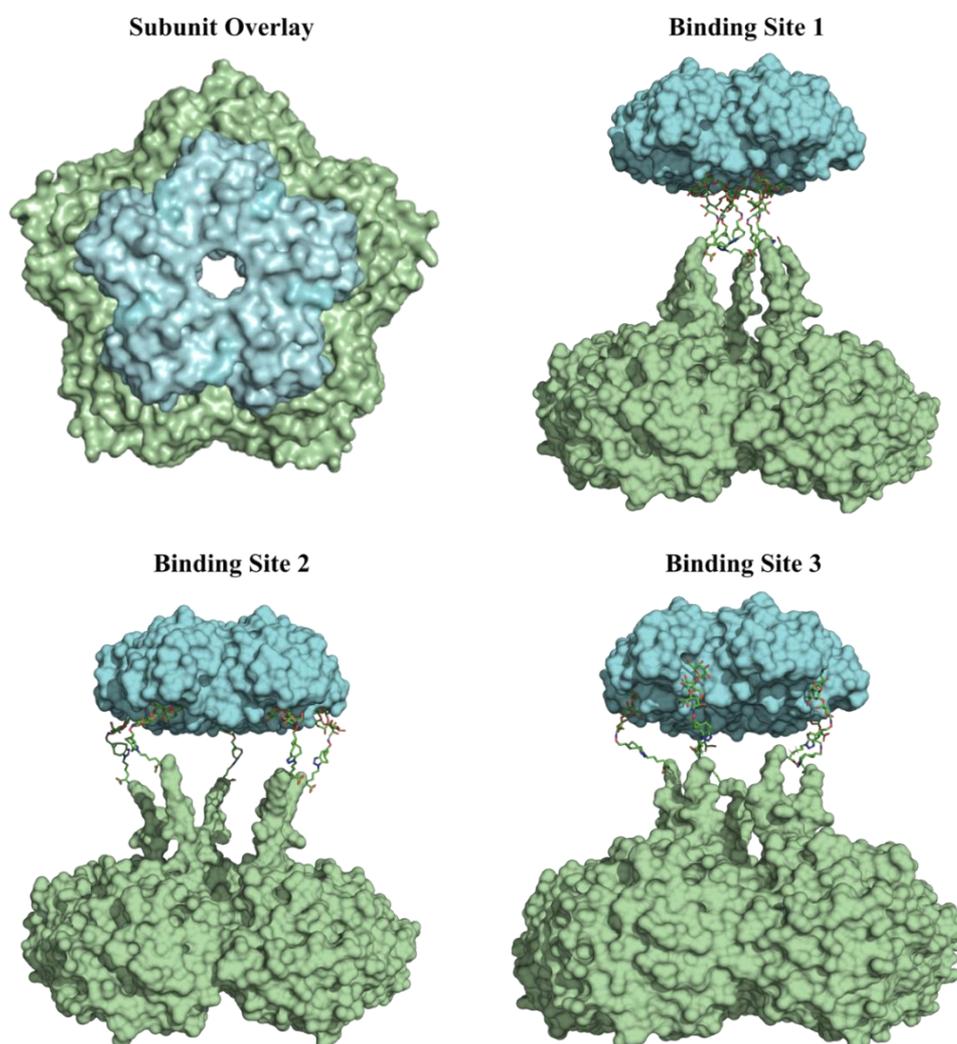


Figure 5-1: Comparison of size and geometry of serum amyloid P (green) and VTB (blue). in silico modelling of SAP-LPET-GGGK-Gb3 binding to each of the three VTB binding sites produced by Ryan McBerney in Pymol (version 2.4.0) using Protein Data Bank files 1SAC.pdb (SAP) & 1D11.pdb (VTB).

5.2 Design and Expression of a Serum Amyloid P Scaffold

The native SAP protein sequence had to be reengineered for use as a protein scaffold (Figure 5-2). Firstly, SAP's propensity to self-aggregate in a calcium-dependent manner was eliminated by removal of the calcium binding site through an E167Q mutation.²³⁴ Secondly, a C-terminal LPETGA sortase recognition sequence and hexahistidine tag was inserted, which would allow a peptide containing an azide or BCN to be ligated. Finally, an N-terminal serine residue was introduced into the sequence to allow for N-terminal oxime ligation to be performed in the future.

```

          10      20      30      40      50
WT SAP      |-----|-----|-----|-----|-----|
--HTDLSGKV FVFPRESVTD HVNLITPLEK PLQNFTLCFR AYSDLSRAYS
SAP (E167Q)-LPETGA MS.....

          60      70      80      90     100
WT SAP      |-----|-----|-----|-----|-----|
LFSYNTQGRD NELLVYKERV GEYSLYIGRH KVTSKVIEKF PAPVHICVSW
SAP (E167Q)-LPETGA .....

          110     120     130     140     150
WT SAP      |-----|-----|-----|-----|-----|
ESSSGIAEFW INGTPLVKKG LRQGYFVEAQ PKIVLGQEQD SYGGKFDRSQ
SAP (E167Q)-LPETGA .....

          160     170     180     190     200
WT SAP      |-----|-----|-----|-----|-----|
SFVGEIGDLY MWDSVLPPEN ILSAYQGTPL PANILDWQAL NYEIRGYVII
SAP (E167Q)-LPETGA .....Q.....

          210     220
WT SAP      |-----|-----|-----|-----|-----|
KPLVWV-----
SAP (E167Q)-LPETGA .....GSGL PETGAHHHHH H

```

Figure 5-2: Sequence alignment of WT SAP and SAP(E167Q)-LPETGAH₆.

A pET11a plasmid encoding SAP(E167Q)-LPETGAH₆ was ordered from Genscript and transformed into two different *E. coli* expression cell lines: C41(DE3) and BL21(DE3). Expression trials and initial attempts at protein refolding were performed with the assistance of Chris Field (undergraduate summer student). Overexpression of SAP into inclusion bodies was observed for both cell lines, with more intense bands were observed for expression in C41(DE3) compared to BL21(DE3) when analysed by SDS-PAGE (data not shown). SAP was expressed in auto-induction media (AIM) initially at 30 °C for 5 hours before being left at 25 °C overnight. The inclusion bodies were recovered, solubilised in 8M urea and purified by Ni-NTA affinity chromatography. SDS-PAGE confirmed high levels of overexpression and purification of the denatured protein (Figure 5-3). The protein was refolded following two methods, the first using the rapid dilution into arginine refolding buffer and the second by stepwise reduction in the concentration of urea by dialysis. Although no aggregation was observed using either refolding protocol, the folded protein was found to have a propensity to aggregate upon concentration by centrifugal ultrafiltration.

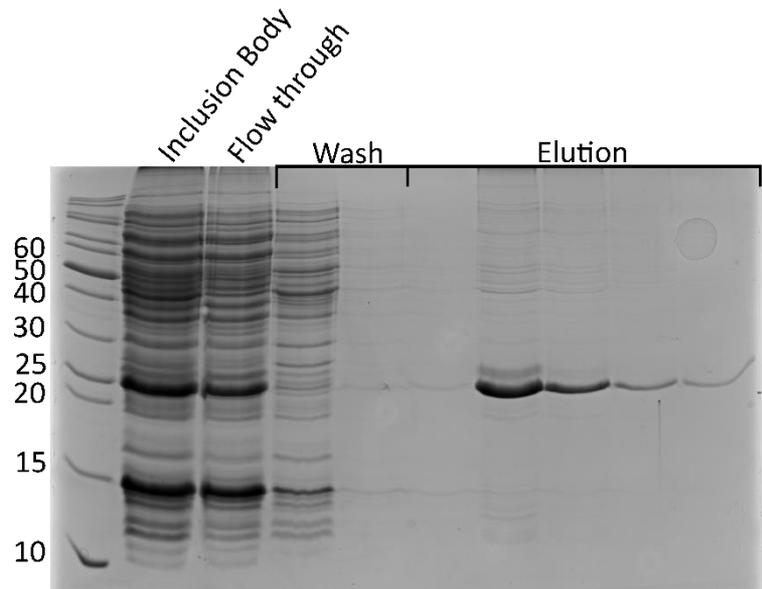


Figure 5-3: SDS-PAGE gel showing isolation of SAP(E167A)-LPETGASH₆ by Ni-NTA affinity chromatography from solubilised inclusion bodies.

5.2.1 The Issues of Protein Glycosylation

There are many reasonable explanations for the mutant SAP's propensity to aggregate. Mapping of the surface exposed hydrophilic and hydrophobic residues reveals large hydrophobic regions on the top face of SAP (Figure 5-4). Furthermore, in humans SAP is modified with a large N-linked glycan on each protomer.²³⁵ Modelling of the location of N-linked glycosylation using the GLYCAM web server suggests the glycans covers these regions. In the absence of glycosylation these hydrophobic regions remain solvent exposed and upon concentration may lead to aggregation.

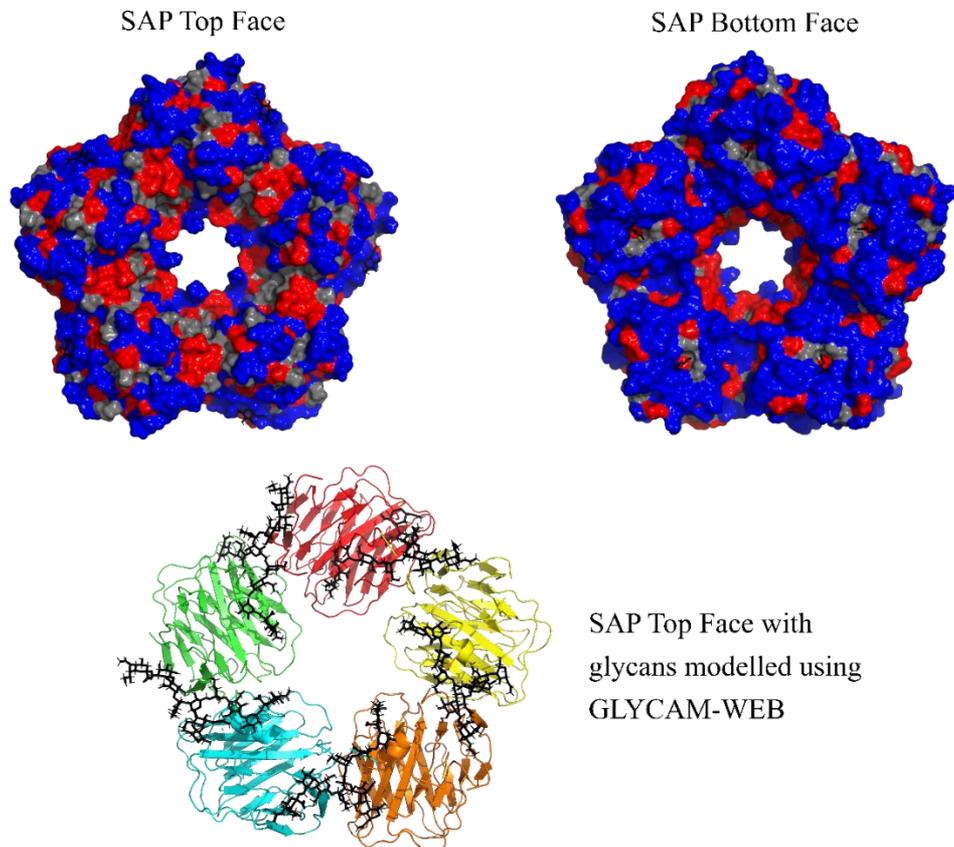


Figure 5-4: Mapping of hydrophobic (red) and hydrophilic (blue) residues on the surface of wt SAP (PDB: 3KQR). Energy minimalised model of wt SAP showing N-glycosylation of Asn32 (modelled using www.glycam.org glycoprotein builder).

5.3 C-Reactive Protein

C-Reactive protein (CRP) was considered as a potential alternative to serum amyloid P component (SAP). CRP shares 51% sequence homology with SAP; is not known to be glycosylated in humans; and expression of the wild type protein has been reported in *E. coli*.^{236,237} Mapping of the surface-exposed hydrophobic and hydrophilic residues showed no large solvent-exposed hydrophobic regions (Figure 5-5). Together this suggests CRP could be a better candidate for neoglycoprotein synthesis.

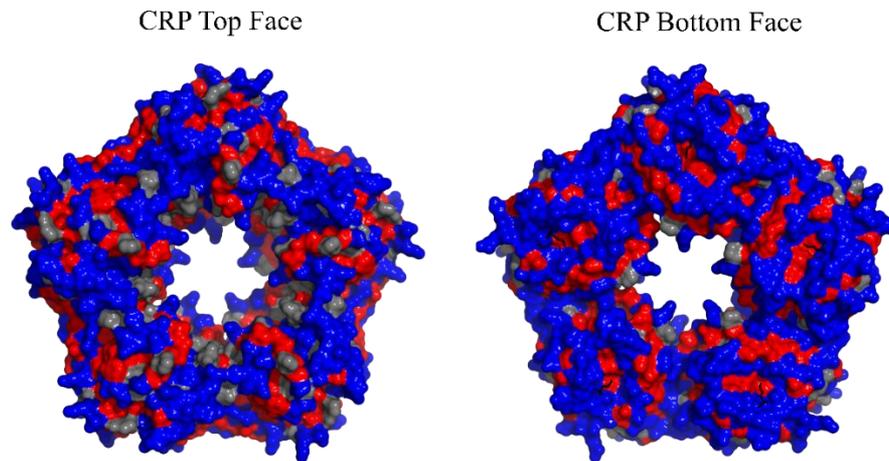


Figure 5-5: Mapping of hydrophobic (red) and hydrophilic (blue) residues on the surface of WT CRP (PDB: 1B09).

5.3.1 Expression Testing & Plasmid Optimisation

Recombinant expression of soluble wild type CRP into the culture supernatant has been reported, providing CRP was periplasmically targeted and co-expressed with a downstream Shine Dalgarno sequence and *kil* gene in *E. coli*.²³⁷ The *kil* gene encodes a colicin E1 release-lysis protein from *E. coli*.²³⁸ In the absence of the *kil* gene, CRP is expressed into periplasmic inclusion bodies.²³⁹ The reported plasmid design was replicated, using a signal peptide derived from *E. coli* alkaline phosphatase (PhoA) to direct the protein to the periplasm for folding.²³⁷ A C-terminal LPETGASHHHHHH sortase recognition sequence and hexa-histidine tag was inserted at the end of the gene for CRP (Figure 5-6). The resulting pET19b plasmid encoding CRP-LPETGASH₆ with a downstream *kil* gene was ordered from Twist Bioscience and transformed into *E. coli* C41(DE3) and BL21(DE3)pLysS competent cells. Despite extensive expression trails no soluble protein was recovered from either cell line, with protein only found in the insoluble fraction as inclusion bodies (Figure 5-7a). Initial cell growth was found to be slow, and the resulting pellets were found to be gloopy upon resuspension, leading to the conclusion the *kil* gene was too strongly expressed leading to cell lysis. Attempts to reduce *kil* gene expression through mutation of the downstream Shine Dalgarno ribosomal binding site also resulted in insoluble protein and suspected cell lysis (Figure 5-7b). Having had some positive results in previous attempts to refold SAP, the downstream Shine Dalgarno sequence & *kil* gene were

deleted *via* site-directed mutagenesis along with the PhoA periplasmic targeting sequence (Figure 5-6); this should result in CRP-LPETGASH₆ expressed solely into cytoplasmic inclusion bodies.

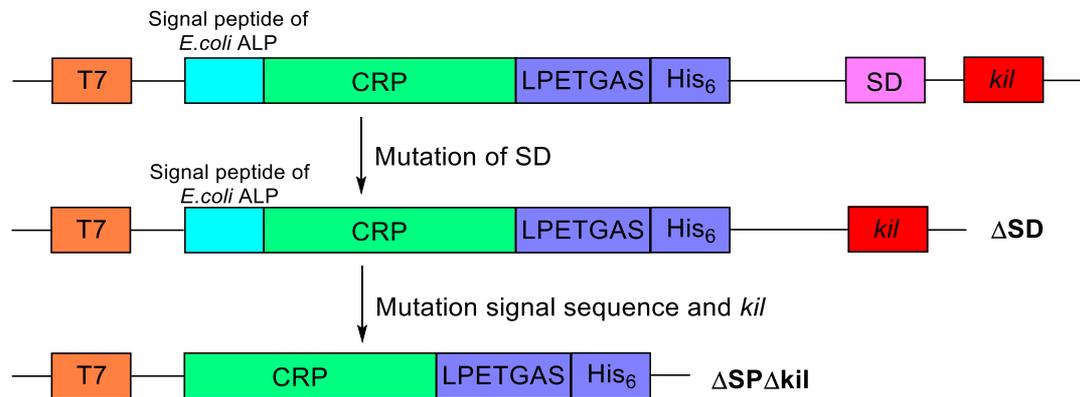


Figure 5-6: Gene construction for the production of CRP-LPETGASH₆. The gene constructed was composed of T7 promoter & ribosomal binding site, signal peptide of *E. coli* ALP (alkaline phosphatase), CRP gene, C-terminal LPETGAS sortase recognition sequence and hexa-histidine tag, SD (Shine-Dalgarno) sequence, and *kil* gene. Two further generations produced first via mutation of downstream SD, then mutation of signal peptide and *kil* gene.

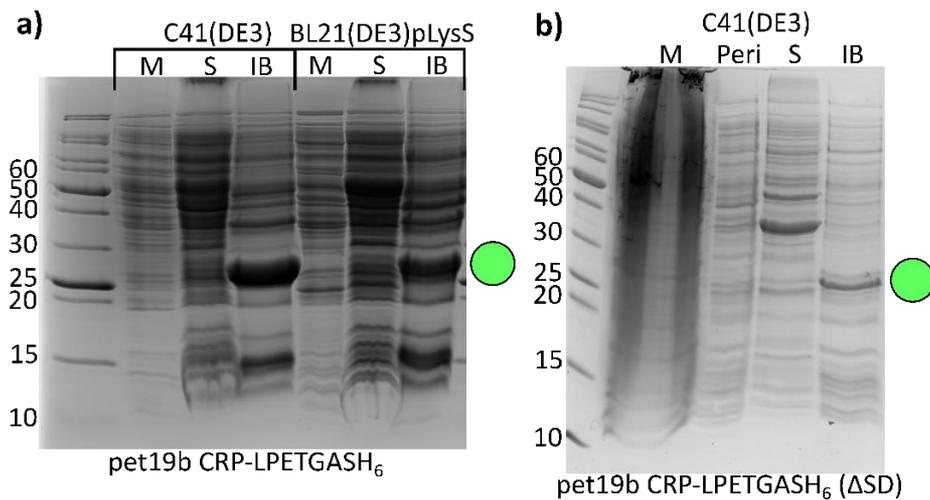


Figure 5-7: SDS-PAGE gels of CRP-LPETGA expression tests showing CRP-LPETGASH₆ expressed only into inclusion bodies. Expected migration for CRP-LPETGASH₆ indicated by green circle. a) Shine Dalgarno sequence before *kil* gene expressed from *E. coli* C41(DE3) and BL21(DE3)pLysS. b) No Shine Dalgarno sequence before *kil* gene expressed from *E. coli* C41(DE3). M – media precipitate, Peri – Periplasmic extract, S – soluble lysate, IB – Inclusion bodies.

5.3.2 Refolding C-Reactive Protein

The pET29b plasmid which encoded CRP-LPETGASH₆ (Δ SP Δ kil) was transformed into BL21(DE3) competent cells and expressed in auto-induction media (AIM) at 30 °C for 5 hours before being left overnight at 25 °C to maximise cell density whilst ensuring overexpression. The inclusion bodies were recovered, solubilised in 8 M urea, and purified by Ni-NTA affinity chromatography. SDS-PAGE confirmed high levels of overexpression and purification of the denatured protein (Figure 5-8). The protein was refolded by sequential dialysis in order to reduce the concentration of urea in the buffer until no denaturant remained. No aggregation was observed during the refolding protocol until the buffer was changed from Tris-buffered saline containing 1M urea to Tris-buffered saline containing no urea, suggesting CRP was still partially unfolded in 1M urea. The remaining CRP in solution was purified by size exclusion chromatography using a Superdex[®] S200 (16/600) column to remove any aggregates. Elution from the SEC column suggest a mixture of folded pentameric protein (125 kDa) eluting at ~74 mL and protomer (25 kDa) eluting at ~90 mL (Figure 5-9). For comparison CTB-LPETGA (62.8 kDa) run on the same column under the same conditions is known to elute at ~84 mL (Figure 5-9). Due to the low yield of CRP no high-resolution mass spectrum could be obtained.

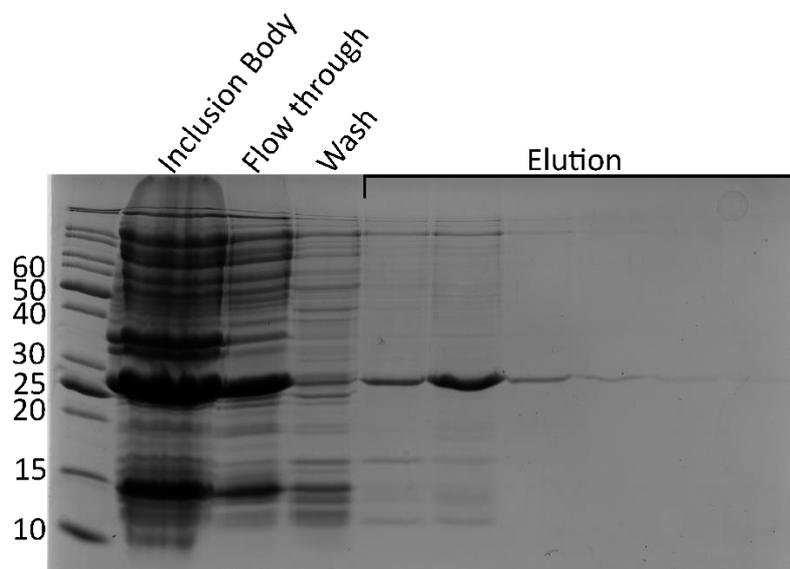


Figure 5-8: SDS-PAGE gel showing isolation CRP-LPETGASH₆ from inclusion bodies solubilised in 8 M urea and purification by Ni-NTA affinity chromatography.

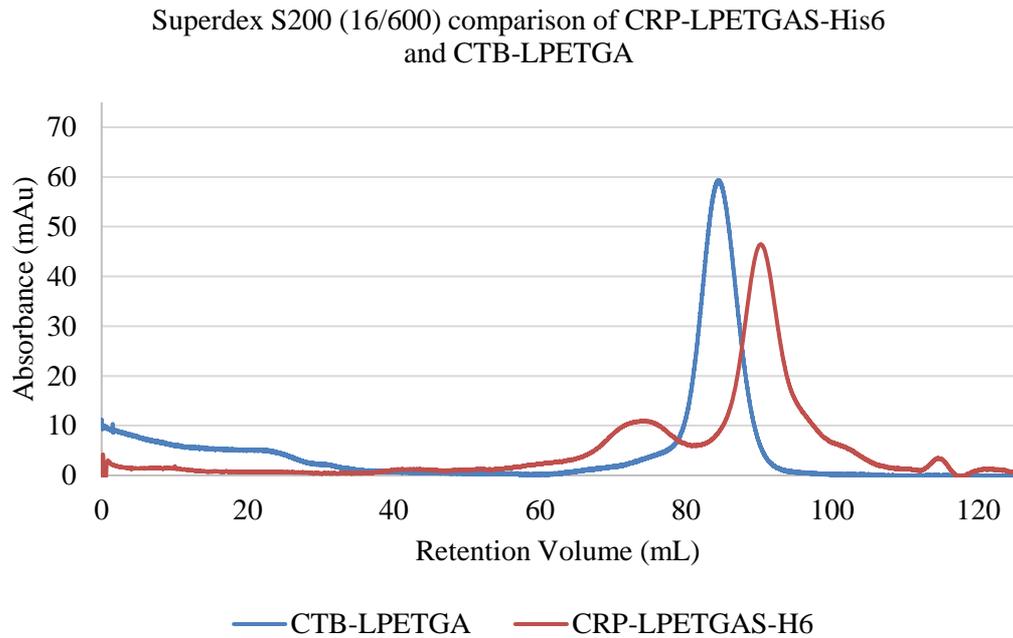


Figure 5-9: Size exclusion comparison of CTB-LPETGA (62.8 kDa) and CRP-LPETGA-H₆ (124.3 kDa). Both proteins were run on Superdex S200 (16/600) using same flow rate at 4 °C. Trace suggests CRP is a mixture of pentamer and protomer, both elute in line with the suggested retention volume for their molecular weight as reported in the Cytiva Gel filtration calibration application note.

5.4 Chapter Conclusions

A final aim for this project is to move away from pentameric bacterial protein scaffolds to a less immunogenic pentameric scaffold such as SAP or CRP for neoglycoprotein-based inhibitors. Initially SAP was chosen as the first scaffold of interest, with modelling by Ryan McBerney showing that a neoglycoprotein produced from addition of Gb3 ligands to a C-terminal extension of SAP *via* sortase should be capable of binding to each of the three VTB binding sites. The mutant SAP protein was successfully overexpressed in *E. coli* and refolded by stepwise reduction in the concentration of urea by dialysis or rapid dilution. However, upon concentration the mutant SAP was found to aggregate. Mapping of surface-exposed hydrophobic residues and N-glycosylation showed large hydrophobic regions which could be covered by the large N-glycan on each protomer.

A second member of the pentraxin family, CRP, was proposed as a better alternative. Wild type CRP is reported to have been produced recombinantly from *E. coli* and is available commercially. Literature suggested a route to allow expression of CRP into the growth media, however attempts to replicate this approach with a CRP construct suitable for sortase ligation, resulted in expression into inclusion bodies and cell lysis. The inclusion bodies were solubilised in 8 M urea, and no aggregation of the mutant CRP was observed during refolding by sequential dialysis until the urea concentration was reduced from 1M to no urea, potentially suggesting the protein was misfolded. Size-exclusion chromatography of the refolded CRP corroborate this conclusion with two peaks being observed; one with an expected retention time for the pentamer and a second with the expected retention time for the protomer. To evaluate if the mutant CRP is capable of being refolded further optimisation is required with additional steps of dialysis required below 1M urea. If refolding is not possible expression of the mutant SAP or CRP could be performed in mammalian cells where folding and glycosylation would be performed naturally.

Chapter 6: Conclusions & Future Work

The primary aims of this project were to develop and evaluate site-specific methodologies to prepare glycoconjugates which could be used to detect novel carbohydrate binding proteins or as multivalent inhibitors of bacterial enterotoxins, cholera toxin and verotoxin. One strategy investigated using enzymatic methods sought to prepare a series of well-defined MUC1-glycoconjugates which could be used in phage display to detect novel carbohydrate binding proteins. The second strategy using a minimal length, bifunctional and bioorthogonal handle sought to expand methods for synthetic chemical glycosylation of proteins and investigate the impact of linker length and site of glycosylation has on inhibitor potency.

6.1 C-terminal Sortase Mediated Ligation of Glycopeptides

For the first strategy using enzymatic methods to prepare a series of well-defined MUC1-glycoconjugates, a library of MUC1(Tn) glycopeptides based on two known antibody recognition sequences were synthesised which displayed all possible patterns of α -O-GalNAc (Tn) glycosylation. For C-terminal sortase mediated ligation two cholera toxin B-subunit mutants containing a C-terminal LPETGA motif for sortase-mediated ligation were expressed recombinantly from *E. coli*. The first construct had a periplasmic targeting sequence and was found to contain a C-terminal truncation affecting ~10% of protomers. To combat the truncation problem, an alternative construct was expressed which lacked the periplasmic targeting sequence and hence had to be isolated and refolded from inclusion bodies. However, as this construct also had a C-terminal his-tag, the full length protein could be purified under denaturing conditions.

Sortase-mediated ligation of glycopeptides was initially found to be poor, with increasing numbers of GalNAc moieties having a negative impact on conjugational efficiency. It was speculated that conformational changes in the labelling substrates potentially reduce their nucleophilicity, leading to poor levels of conjugation. The addition of a PEG₆-equivalent linker was found to be optimal to improve ligation efficiency whilst retaining secondary structure. Using the optimised conditions and linker length, high levels of C-terminal labelling were demonstrated for all the

peptides (2.22, 2.26, 2.27) and glycopeptides (2.19, 2.23-2.25) substrates with conjugation between 79% and 86% of the available sortase sites.

Although multiple groups have investigated CTB with C- and N-terminal peptide extensions to evaluate them as potential vaccine candidates, most were produced as fusion proteins.^{240,241} The optimisation described here could provide new methods for site-specific glycopeptide conjugation using sortase. As discussed in section 1.3.2.1, CTB is a strong candidate for use as a mucosal adjuvant.²⁰⁰ Further the attachment of antigens to an adjuvant scaffold in a well-defined repeatable manner with the minimal length linker results in higher titers of high-affinity anti-glycan antibodies *in vivo*.⁶⁵ This may suggest well-defined CTB-glycoconjugates have broader applications as promising mucosal vaccine candidates.

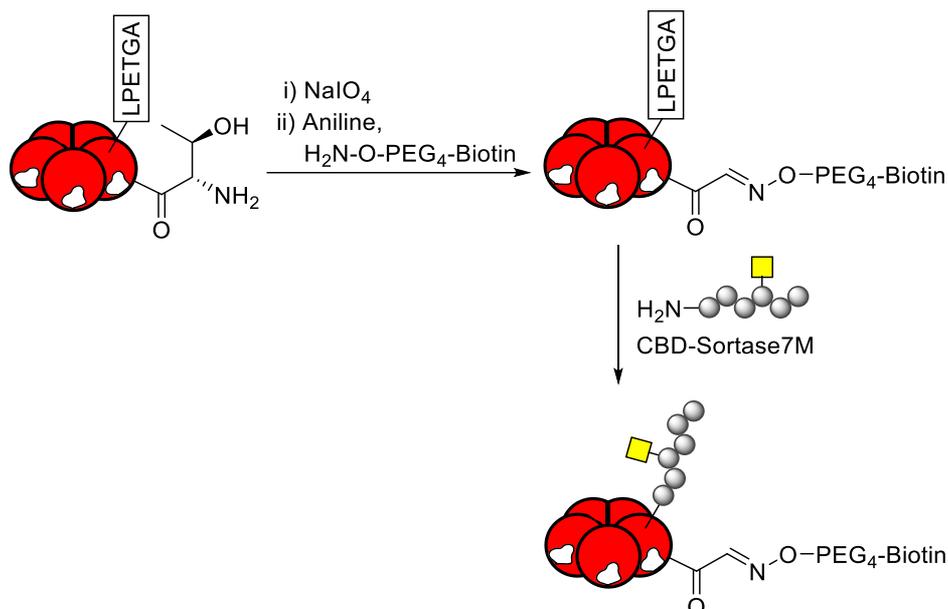
6.2 Affimer Phage Display

Previously Matthew Balmforth had shown Affimers with a propensity to bind to the top face of CTB could be selected during phage display by orientating CTB on a surface using its GM1 binding sites.¹⁹⁹ We originally sought to use this methodology to discover Affimers with affinity for the glycopeptide attached to the C-terminus of CTB (distal to the GM1 binding site). However, due to complications arising from the Covid-19 pandemic phage-display screening had to be performed using the standard high-throughput methodology routinely used in the Bio-Structure Technology Group (BSTG), University of Leeds, which meant the CTB scaffold required biotinylation.

6.2.1 Orthogonal N- & C-terminal Labelling of CTB

Site-specific N-terminal biotinylation *via* oxime ligation was selected over random NHS-biotin labelling of surface exposed lysine residues to ensure the biotinylation did not interfere with accessibility of the glycopeptides. To avoid potential oxidation of GalNAc moiety, N-terminal oxime ligation was performed first followed by C-terminal sortase ligation of the peptide/glycopeptide substrates (Scheme 6-1). To allow removal of sortase7M from the reaction once the desired reaction end point was reached, a chitin-binding variant of sortase7M was designed and expressed. Testing

of this variant showed the addition of the N-terminal chitin binding domain had no negative effect on enzyme activity.



Scheme 6-1: Synthetic route to orthogonally N- and C-terminally labelled CTB glycoconjugates.

The N-terminus of CTB-LPETGA was quantitatively labelled with the aminoxy-functionalised biotin (**3.1**) via oxime ligation. C-Terminal sortase labelling with the optimised MUC1 peptide and glycopeptide epitopes resulted in labelling levels of 83-89% following removal of the chitin binding variant of sortase7M. Dual N- and C-terminal labelling of a protein is not unprecedented;²⁴² however, the combination of sortase-mediated ligation and oxime ligation presented here is unique to the literature. Although presented here with a biotin, if used to produce glycoconjugate vaccines oxime ligation could be used to introduce other functional peptide/glycopeptides, such as universal T-helper epitope PADRE or additional glycopeptides demonstrating an alternative glycoform.²⁴³

Phage display screening of the biotinylated CTB-glycoconjugates was unsuccessful. Although there was a chance Affimers with affinity for the glycan or glycopeptides were present, eliminating undesired binders against the CTB scaffold was not possible. Affimers are known to prefer larger regions of secondary/tertiary structure, such as grooves or crevices, into which the loops can bind.²⁰⁶ It is therefore unsurprising the Affimers preferred to bind to the protein scaffold rather than the glycopeptides which lack a defined secondary structure.

6.2.2 Multivalent MUC1(Tn) Glycopeptides for Phage Display

To avoid the use of CTB as scaffold, a biotinylated peptide scaffold (**3.5**) was synthesised by solid phase peptide synthesis and the BCN labelled MUC1(Tn) glycopeptides attached by strain-promoted azide-alkyne cycloaddition. This provided facile access to four trimeric-MUC1(Tn) biotinylated glycopeptides. As with the previous screens, no binders could be identified. This may suggest the Affimer scaffold is not an appropriate for developing binders against glycans or glycopeptides. This is very likely due to the lack of secondary structure onto which the loops can bind.

Alternative phage display libraries should be considered such as those suggested in section 3.3.1. The difficulty with many of these phage display libraries is their focus on small, easy to produce scaffolds which do not take advantage of multivalent binding or have large binding interfaces ideally suited for glycan binding.²⁴⁴ Alternative scaffolds such as DARPin, whose parent protein is known to have glycan binding pockets,²⁴⁵ could be a potential future option.^{246,247} These proteins could be ideal for glycan and glycopeptide binding due to their modular design which allows for multiple binding faces, taking advantage of multivalent binding leading to a higher avidity interaction.

6.2.3 Working Towards developing Affimers against Verotoxin

Previously Matthew Balmforth developed a method of targeted drug delivery of proteins using cholera toxin B-subunit.¹⁹⁹ Given its structural similarity and entry mechanism into cells, VTB could also be used to safely deliver drugs and biologics in pancreatic and colon cancer. Verotoxin B-subunit was successfully expressed and isolated from the periplasm *via* osmotic shock (periplasmic extraction). To control the orientation of VTB during phage display, site-specific N-terminal biotinylation was performed *via* oxime ligation. Although this chemistry is well known for CTB,^{108,166,199} it is unclear if this has been performed on VTB. Despite the VTB pentamer being less stable than CTB, the protein remains pentameric upon labelling. The ability to further functionalise VTB without loss of pentameric structure is an

important advancement. Affimer phage display screening of VTB identified nine unique binders.

The binding affinity and specificity for VTB of these hits requires further characterisation potentially through the use of a plate-based Affimer-lectin binding assay (ALBA) developed by Matthew Balmforth,¹⁹⁹ SPR and/or ITC (Figure 6-1). Further understanding of the fate of the Affimer-VTB complexes upon immunisation is also required before deciding the best use cases of each complex. Previously Matthew Balmforth showed his CTB-Affimer complexes had different fates once inside the neurone. The first complex was shown to readily dissociate, with the Affimer trafficked into lysosomes; the other appeared to remain associated, and after a week the complex appeared to be trafficked to lysosomes.¹⁹⁹

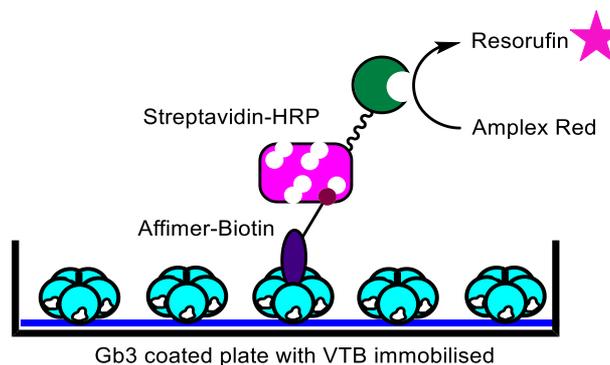
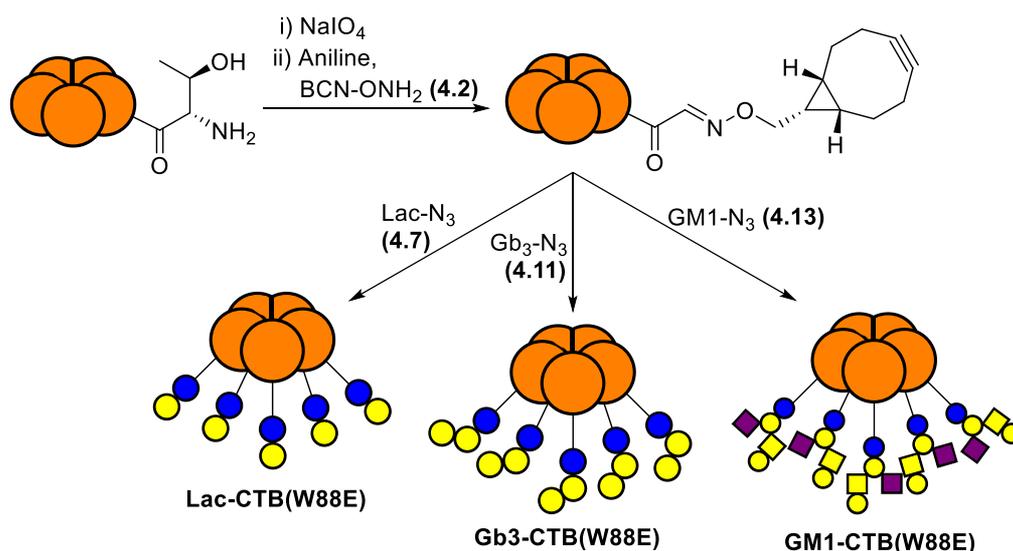


Figure 6-1: The setup for an Affimer-lectin binding assay. Biotinylated Affimer is tested against immobilised glycoconjugate/VTB. Affimer complex is detected through the use of streptavidin-HRP conjugate and a solution of Amplex red[®] and hydrogen peroxide.

As discussed in section 3.1, Affimers are versatile reagents ideally suited as an alternative to antibodies in lateral flow devices.²⁰⁶ Depending on the binding site of the Affimers, a lateral flow device could be used to detect verotoxin rapidly & selectively in blood, urinal or faecal samples, allowing for rapid treatment of the underlying infection. Rapid treatment is vital in increasing survivability rates through the reduction in the chance of onset haemolytic uremic syndrome. A lateral flow device which take advantage of Affimers which bind to SARS-CoV-2 spike protein is commercially available and approved by the UK Medicines and Healthcare products Regulatory Agency (MHRA).^{248,249}

6.3 Neoglycoconjugate Inhibitors of Bacterial Enterotoxins

Oxime ligation of an aminoxy-functionalised bicyclononyne to the N-terminus of CTB followed by strain-promoted azide-alkyne cycloaddition of a series of biologically relevant glycosyl azides, provides a simple and rapid method of the preparation of site-specific neoglycoconjugates as single glycoforms (Scheme 6-2). Although the site-specific incorporation of strained alkynes has been performed using genetic code expansion,²²⁸ it has not been performed previously *via* oxime-ligation. The simplicity of this method for the site-specific incorporation of BCN provides an accessible route to an easily derivatisable scaffold.



Scheme 6-2: Synthetic route to multivalent neoglycoprotein inhibitors of enterotoxins, cholera toxin and verotoxin.

The effect of linker length and site of glycosylation on inhibitory potential of a bacterial toxin-derived neoglycoproteins was a primary research question for this section of the project. Using the ELLA inhibition assay established for CTB, the inhibitory potential of bottom-face modified GM1 neoglycoprotein inhibitor (**GM1-CTB(W88E)**) was tested and found to have a very low IC₅₀ of 924 pM, a 115-fold increase on monovalent GM1 oligosaccharide. Compared to previous inhibitors developed within the group, only a modest decrease in potency is observed. Reduction in linker length results in only a 9-fold loss of potency compared to that reported by Branson *et al.*¹⁰⁸ and changing the location of glycosylation but using the same linker results in only a 2-fold loss of potency.¹⁸⁷ Both minimal length linkers observe at least a 4-fold decrease in potency suggesting bringing the two pentamers into closer

proximity has a greater impact on the loss of inhibition than changing the site of glycosylation. This is likely due to a combination of increased electrostatic repulsion and an enthalpic penalty encountered as a result of bringing the two pentameric proteins into close proximity. The 2-fold improvement in inhibition observed for the top-face modified neoglycoconjugate compared to the bottom-face is likely caused by closer spacing of the GM1 ligands which better matches the spacing between binding sites. The production of neoglycoconjugates with minimal derivatisation from complex oligosaccharides, is a much simpler route to neoglycoprotein inhibitors with minimal loss of inhibition.

Given the impact neoglycoconjugate inhibitors have shown against cholera toxin, we sought to investigate if the same improvement in inhibitor potency would be observed against verotoxin. Using an analogous ELLA inhibition assay for VTB, the inhibitory potential of top-face and bottom-face modified Gb3 neoglycoproteins was tested. Both pentavalent neoglycoproteins were shown to be very potent inhibitors of VTB adhesion, with at least a 550-fold increase in potency compared to monovalent Gb3 oligosaccharide. Although further work is required to confirm inhibition at higher concentrations, this work shows that previous methods applied to CTB can also be applied to VTB and Gb3-neoglycoproteins have the potential to be potent inhibitors of VTB. Simian virus 40 (SV40), a non-enveloped DNA virus of the polyomavirus family, also binds GM1 with a similar pentameric arrangement to CTB.²⁵⁰ Therefore, this methodology could be expanded to produce the first neoglycoprotein inhibitors of SV40 or other members of that family that infect humans.

Although attempts to conjugate a Gb3-functionalised cyclic glycopeptide to the non-binding CTB scaffold were unsuccessful; the unconjugated cyclic glycopeptide (**4.16**) was found to exhibit micromolar inhibition of VTB, comparable to pentameric display of Gb3 ligated to the N-terminus of CTB. Given Gb3-neoglycoconjugates are easier to synthesise and appear to be more potent inhibitors may suggest cyclic glycopeptides on their own are not the perfect option. However, further work is required to determine the mode of inhibition. If the hypothesis is correct that cyclic glycopeptide bridges two VTB pentamers in a similar fashion seen for the STARFISH inhibitor developed by Bundle et al.,¹⁰⁴ it may suggest a dual sided design may further improve inhibition for neoglycoconjugates. In chapter 3, a novel combination of orthogonal oxime and sortase-ligation was presented. Combining oxime and sortase-

ligation could provide an easy route to such glycoconjugates, whilst maintaining the benefits obtained from prearranging glycans in the correct spacing and geometric orientation. Alternatively, this method could be used to present two different glycans on the top and bottom-faces of CTB, resulting in an inhibitor of two different toxins.

6.3.1 Engineering a Non-bacterial Pentameric Protein Scaffold

Having shown the Gb3-neoglycoprotein targeting verotoxin have the potential to be potent inhibitors of VTB adhesion, a final aim of the project was to develop non-bacterial-based protein scaffold for use in the synthesis of neoglycoproteins. Initial attempts to express and refold a mutant serum amyloid P (SAP) proved successful; however, it was found to have a propensity to aggregate upon centrifugal concentration. Mapping of the surface exposed hydrophobic & hydrophilic residues showed large hydrophobic regions; mapping the location of N-glycosylation suggests these regions are masked by the N-glycan.

A second member of the pentraxin family, C-reactive protein (CRP), was proposed as a better alternative. Wild type CRP was available commercially recombinantly expressed from *E. coli* and a method reported for the expression of soluble WT CRP directly into the expression medium. Attempts to replicate the expression platform with the mutant CRP resulted in expression only into inclusion bodies and cell lysis. The method of CRP expression and purification requires further optimisation; however initial results suggest the protein is capable of being refolded into a pentamer. If issues with refolding persist, expression of the mutant SAP or CRP could be performed in mammalian cells where folding and glycosylation would be performed naturally. If expression in a mammalian system is still not possible the issue maybe caused by the C-terminal extension. The location of the C-terminus is close the interface between two protomers and the central pore, as a result the addition of the C-terminal extension may prevent assembly of the pentamer (Figure 6-2). Given the length of the C-terminal extension and its flexibility (Figure 5.1), the addition of a similar length extension to the N-terminus should be sufficient to reach all three VTB binding sites.

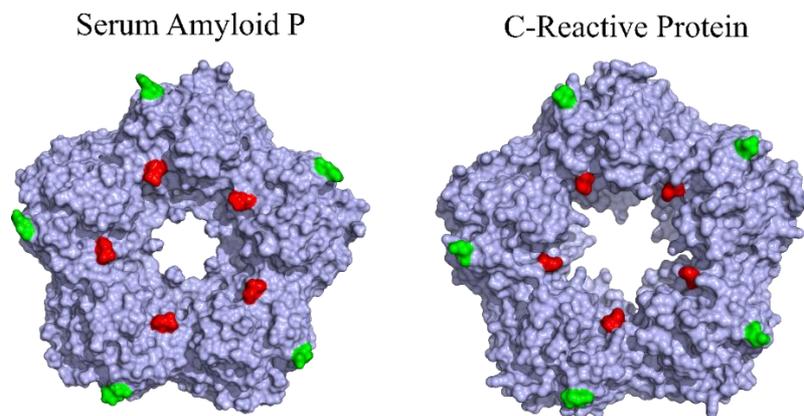


Figure 6-2: Surface representations of Serum amyloid P (SAP) and C-Reactive Protein (CRP). N-terminus coloured green; c-terminus coloured red. Produced in Pymol (version 2.4.0) using Protein Data Bank files 3KQR.pdb (SAP) & 1B09.pdb (CRP).

After determining the most suitable expression protocol, optimisation of sortase ligation conditions can be performed allowing the installation of either a terminal azide or BCN functional group. Strain-promoted azide-alkyne cycloaddition can be used to append Gb3 ligands to the scaffold (Figure 6-3). This will use either the glycosyl azides as outlined in chapter four or use the BCN-functionalised Gb3 described by Ryan McBerney.¹⁸⁷ The resulting neoglycoproteins can be tested for inhibition of their respective bacterial toxins *in vitro*, comparing to the CTB-scaffolded analogues. If successful inhibition is observed, further studies could be carried out *in vivo* to determine the efficacy of the approach as a viable therapeutic option for treatment of bacterial toxin infection. Further the neoglycoproteins should be subjected to immunogenicity testing to determine the effect, if any, they have on the immune system.

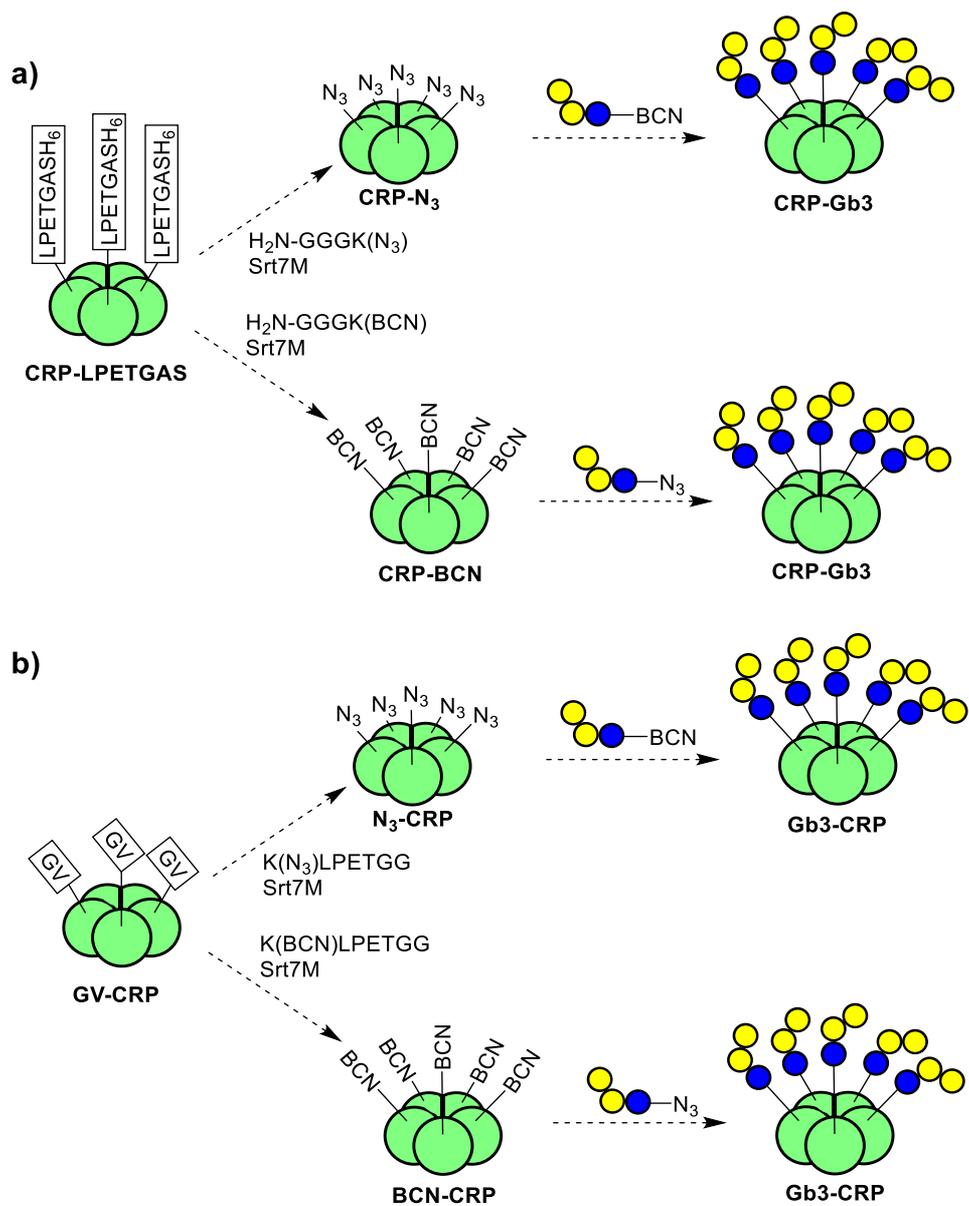


Figure 6-3: Functionalization of a) CRP-LPETGAS or b) GV-CRP with Gb3 ligands, first by sortase ligation of either BCN or azido functionalised peptides, followed by SPAAC with respective BCN-functionalised Gb3 or Gb3 azide.

Chapter 7: Chemistry Experimental

7.1 Synthesis

7.1.1 General Reagents and Equipment

Unless stated otherwise, all starting materials and reagents were purchased from commercial suppliers and used without further purification. All solvents used were dried prior to use, according to standard methods, unless otherwise stated. Reactions were performed under an N₂ atmosphere and within glassware which was oven dried. Completion of reactions was initially determined by TLC and visualized using shortwave ultraviolet light (254 nm) and/or charring with 5% H₂SO₄/MeOH. TLC plates used were Merck Silica-Gel 60 F²⁵⁴ Aluminium backed. Silica chromatography columns prepared using Fisher 60Å 43-60 micron silica gel. Lyophilisation carried out using Virtis Benchtop K freeze dryer. Centrifugation was performed using an Eppendorf Centrifuge 5810. Size exclusion chromatography was performed using a Biogel P2 column attached to a GE Pharmacia ÄKTA Prime FPLC system.

NMR spectroscopy was recorded using Bruker AV3HD-400 (400 MHz, BBO Probe), Bruker AV4 NEO (500 MHz, BBO, TXI & TBO Probe) and Bruker AV4 NEO (500 MHz, C/H cryoprobe) spectrometers. NMR data is reported in parts per million (ppm) referenced to residual solvent signal at room temperature.²⁵¹ The following abbreviations are used in ¹H NMR analysis: Ar = aromatic, s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, dd = double doublet, dt = doublet of triplets, td = triplet of doublets and ddd = double double doublet.

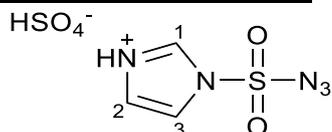
HRMS was performed using Bruker Daltonics MicroTOF mass spectrometer employing electrospray (ES+) ionisation. LC-MS analysis performed on Bruker AmaZon X series LC-MS spectrometer. MALDI-TOF performed on Shimadzu AXIMA Performance using α -Cyano-4-hydroxycinnamic acid (CHCA) matrix. IR spectroscopy was recorded using a Bruker Platinum ATR spectrometer. Optical rotations were measured with a Schmidt and Haensch Polartronic H 532 at the sodium D-line with the $[\alpha]_D^{20}$ values given in the units 10⁻¹ deg cm² g⁻¹.

Standard numbering conventions for carbohydrates in pyranoside and furanoside systems is followed (i.e numbering starts from the anomeric centre). Numbering schemes of more complex structures will be demonstrated on the structure provided.

7.1.2 Small Molecule Synthesis

7.1.2.1 Fmoc-Thr(Tn)-OH & Fmoc-Ser(Tn)-OH Synthesis

Imidazole-1-sulfonyl azide hydrogen sulfate salt (**2.1**)¹⁹⁰



Imidazole-1-sulfonyl azide is highly explosive in its neutral form. Reaction should only be performed behind Perspex blast shield and no attempt to concentrate solutions of imidazole-1-sulfonyl azide should be made. Sodium azide and imidazole were dried overnight (12-14 hrs) in a vacuum desiccator over phosphorous pentoxide prior to use. Extra dry EtOAc over molecular sieves was purchased from Acros Organics; a fresh bottle was used for each reaction.

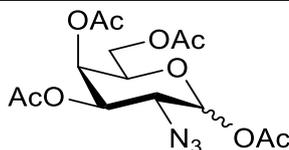
Sodium azide (5.00 g, 77 mmol, 1.0 eq) was placed in a dry 500 mL three neck RBF with a dry stirrer bead. Extra dry EtOAc (77 mL) was added to the flask and the resulting suspension cooled to 0 °C. Sulfuryl chloride (6.2 mL, 77 mmol, 1.0 eq) was added dropwise over 5 minutes and the mixture allowed to warm to room temperature and stirred for 24 hrs. The suspension was recooled to 0 °C and imidazole (10.00 g, 146 mmol, 1.9 eq) was added continuously over 5 minutes. The suspension was stirred at RT for 3 hrs. The mixture was basified by the addition of sat. aq. NaHCO₃ solution (150 mL). Once bubbling had ceased, the mixture was separated, the organic portion washed with water (150 mL), dried over MgSO₄ and filtered. Filtrate was recooled to 0 °C and placed under N₂ atmosphere. Conc. H₂SO₄ (4.1 mL, 77 mmol, 1.0 eq) was added dropwise over the course of 5 minutes and gradually warmed to room temperature with vigorous stirring. Over the course of 30 minutes, a colourless precipitate formed and was collected by vacuum filtration. Precipitate was washed with a small amount of ice-cold EtOAc and the crystals were dried under high vacuum to yield imidazole-1-sulfonyl azide hydrogen sulfate salt, **2.1** (12.71 g, 46.8 mmol, 61%).

NMR data is in agreement with reported data.¹⁹⁰

¹H NMR (400 MHz, DMSO): δ 14.29 (1 H, br S, NH⁺), 12.09 (1 H, br S, HSO₄⁻), 9.07 (1 H, s, H₁), 7.67 (2 H, s, H₂, H₃); ¹³C NMR (100 MHz, DMSO): δ 134.5 (C₁), 119.4 (C₂, C₃); IR (ν_{max}/cm⁻¹): 3083 (C-H stretch), 2860 (C-H stretch), 2177 (N=N=N)

stretch), 1586 (N-H bend), 1460 (S=O stretch); **HRMS** [ES⁺] found [M-HSO₄⁻]⁺ 174.0077, C₃H₆N₅O₆S requires 174.0080.

1,3,4,6-Tetra-O-acetyl-2-azido-2-deoxy-D-galactopyranose (**2.3**)^{252,253}



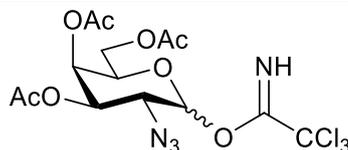
D-Galactosamine hydrochloride (1.00 g, 5.58 mmol, 1.0 eq). K₂CO₃ (2.16 g, 15.62 mmol, 2.8 eq) and CuSO₄·5H₂O (14 mg, 56 μmol, 1 mol%) were dissolved MeOH/H₂O (5:3, 29 mL). To this mixture imidazole-1-sulfonyl azide hydrogen sulfate, **2.1**, (1.817 g, 6.70 mmol, 1.2 eq) was slowly added and left to stir at room temperature. After 3.5 hrs, the mixture was filtered through a pad of celite and co-evaporated with toluene. The water bath was not allowed to exceed 20 °C. Acetic anhydride (2.7 mL, 28.56 mmol, 5.1 eq) was added dropwise to the crude azido-galactose dissolved in pyridine (30 mL) and stirred overnight. Pyridine was removed *in vacuo*, co-evaporating with toluene twice. The residue dissolved in EtOAc (50 mL) and washed with sat. aq. NaHCO₃ solution (3 × 25 mL) and brine (1 × 25 mL). The organic extract was dried over MgSO₄, filtered, and concentrated to dryness. Crude product was purified using flash column chromatography (2:1 Hexane–EtOAc), to yield 1,3,4,6-tetra-*O*-acetyl-2-azido-2-deoxy-D-galactose, **2.3** (1.30 g, 3.48 mmol, 62%, 1:10 α/β) as a colourless oil.

NMR data is in agreement with reported data.^{252,253}

R_f 0.41 (2:1 Hexane–EtOAc); ¹H NMR (300 MHz, CDCl₃): δ 6.29 (1 H, d, *J* = 3.5 Hz, **H**_{1α}), 5.54 (1 H, d, *J* = 8.5 Hz, **H**_{1β}), 5.44 (1 H, d, *J* = 2.4 Hz, **H**_{4α}), 5.35 (1 H, d, *J* = 3.1 Hz, **H**_{4β}), 5.29 (1 H, dd, *J* = 11.0, 2.9 Hz, **H**_{3α}), 4.89 (1 H, dd, *J* = 10.8, 3.3 Hz, **H**_{3β}), 4.13–3.95 (3 H, m, **H**₅, **H**_{6a}, **H**_{6b}), 3.91 (1 H, dd, *J* = 11.0, 3.6 Hz, **H**_{2α}), 3.81 (1 H, dd, *J* = 10.3, 9.0 Hz, **H**_{2β}), 2.17 (3 H, s, Me), 2.13 (3 H, s, Me), 2.03 (3 H, s, Me), 2.00 (3 H, s, Me); ¹³C NMR (75 MHz, CDCl₃): δ 170.2 (C=O), 169.8 (C=O), 169.5 (C=O), 168.4 (C=O), 92.7 (C_{1β}), 90.3 (C_{1α}), 71.6 (C₆), 71.2 (C_{3β}), 68.6 (C_{3α}), 66.8 (C_{4α}), 66.1 (C_{4β}), 61.0 (C_{5α}), 60.9 (C_{5β}), 59.6 (C_{2β}), 56.7 (C_{2α}), 20.9 (Me), 20.7 (Me), 20.5 (Me), 20.5 (Me); **IR** (ν_{max}/cm⁻¹): 2939 (C-H stretch), 2112 (N=N=N stretch) and

1743 (C=O stretch); **HRMS** [ES⁺] found [M+NH₄]⁺ 391.1458, C₁₄H₂₃N₄O₉ requires 391.1460.

3,4,6-Tri-*O*-acetyl-2-azido-2-deoxy-*D*-galactopyranosyl trichloroacetimidate (**2.8**)¹⁸⁸

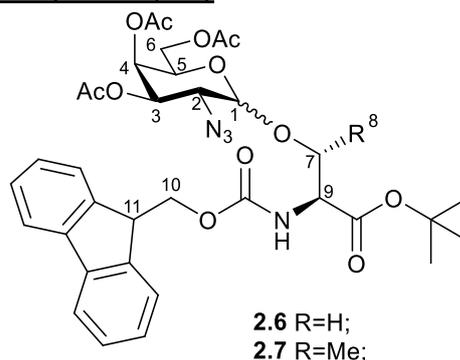


1,3,4,6-Tetra-*O*-acetyl-2-azido-2-deoxy-galactopyranose, **2.3** (1.299 g, 3.48 mmol, 1.0 eq) and hydrazine acetate (385 mg, 4.18 mmol, 1.2 eq) were stirred in dry DMF (6 mL) and heated to 60 °C. After 1 hour, TLC showed reaction had reached completion and reaction was cooled to room temperature. Reaction was diluted with EtOAc (50 mL) and washed with water (2 × 50 mL) and brine (1 × 50 mL). Organic portion was dried over MgSO₄ and concentrated to dryness to yield the crude hemiacetal. Crude hemiacetal was dissolved in dry DCM (12 mL) and cooled to 0 °C. Trichloroacetonitrile (7.00 mL, 69.81 mmol, 20.0 eq) and DBU (260 μL, 1.74 mmol, 0.5 eq) were added and mixture stirred at 0 °C. After 2 hours, TLC showed no further reaction and reaction was diluted with DCM (40 mL) and washed with sat. aq. NH₄Cl solution (3 × 50 mL) and brine (1 × 50 mL), dried over MgSO₄, filtered and concentrated *in vacuo*. Crude trichloroacetimidate was purified by flash column chromatography (2:3 EtOAc–Hexane) to yield pure 3,4,6-tri-*O*-acetyl-2-azido-2-deoxy-*D*-galactopyranosyl trichloroacetimidate, **2.8** (927 mg, 1.95 mmol, 56%, 10:1 α/β) as a yellow foam.

NMR data is in agreement with reported data for the α-anomer.¹⁸⁸

R_f 0.38 (2:3 EtOAc–Hexane); **¹H NMR** (400 MHz, CDCl₃): δ 8.79 (1 H, br. s, NH), 6.78 (1 H, d, *J* = 3.6 Hz, **H₁**), 5.50 (1 H, d, *J* = 2.3 Hz, **H₄**), 5.34 (1 H, dd, *J* = 3.2, 11.1 Hz, **H₃**), 4.38 (1 H, t, *J* = 6.7, 6.7 Hz, **H₅**), 4.15–3.98 (3 H, m, **H₂**, **H_{6a}**, **H_{6b}**), 2.13 (3 H, s, Me), 2.04 (3 H, s, Me), 1.97 (3 H, s, Me); **¹³C NMR** (100 MHz, CDCl₃): δ 170.3 (C=O), 170.0 (C=O), 170.0 (C=O), 160.7 (C=N), 94.5 (C₁), 90.7 (CCl₃), 69.2 (C₅), 68.7 (C₃), 67.0 (C₄), 61.2 (C₆), 57.1 (C₂), 20.7 (Me), 20.6 (Me), 20.6 (Me); **IR** (ν_{max}/cm⁻¹): 3337 (C-H stretch), 3136 (C-H stretch), 2968 (N-H stretch), 2112 (N=N=N stretch), 1744 (C=O stretch) and 1676 (C=N stretch); **HRMS** [ES⁺] found [M+Na]⁺ 497.0012, C₁₄H₁₇Cl₃N₄O₈Na requires 497.0004.

N-Fmoc-*O*-(3,4,6-tri-*O*-acetyl-2-azido-2-deoxy- α/β -*D*-galactopyranosyl)-*L*-serine (**2.6**) and threonine tert-butyl ester (**2.7**)



Trichloroacetimidate, **2.8**, (927 mg, 1.95 mmol, 1.1 eq) and Fmoc-Ser-*O*^tBu, **2.5a**, (679 mg, 1.77 mmol, 1.0 eq) or Fmoc-Thr-*O*^tBu threonine derivative **2.5b** (704 mg, 1.77 mmol, 1.0 eq) were combined and dried overnight under vacuum, then dissolved in a mixture of dry DCM:Et₂O (1:1, 10 mL). After cooling to -30 °C, TMSOTf (42 μ L, 0.23 mmol, 0.13 eq) was added dropwise. The mixture was stirred at -30 °C for 3 hours. The reaction mixture was quenched with DIPEA (53 μ L, 0.30 mmol, 0.17 eq). The reaction mixture was diluted with DCM (40 mL) and washed with water (2 \times 50 mL) and brine (1 \times 50 mL), the organic portion was dried over MgSO₄, filtered and concentrated *in vacuo* to yield *N*-Fmoc-*O*-(3,4,6-tri-*O*-acetyl-2-azido-2-deoxy- α -*D*-galactopyranosyl)-*L*-serine tert-butyl ester, **2.6** (1.564 g, 1:1 α/β) or threonine derivative, **2.7** (1.537 g, 1:1 α/β), both as yellow oils. Both **2.6** and **2.7** were carried through to the next step without further purification.

NMR data is in agreement with reported data.¹⁸⁸

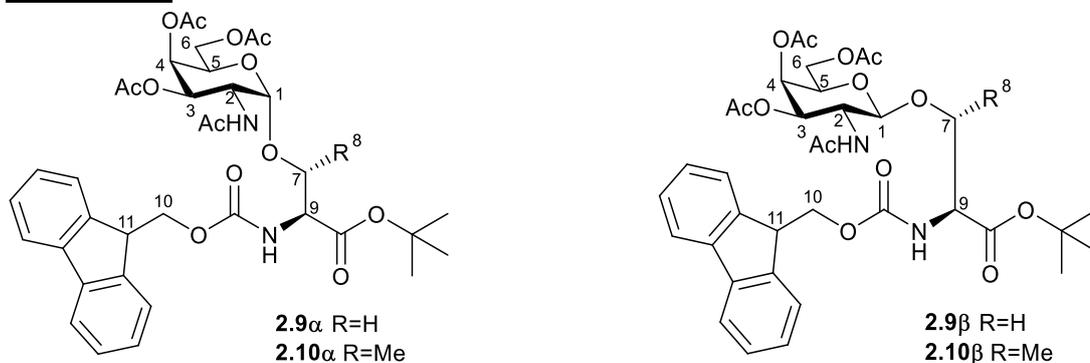
2.6:

R_f: 0.26 (3:7 EtOAc–Hexane); **¹H NMR** (400 MHz, CDCl₃): δ 7.70 (4 H, d, J = 7.6 Hz, ArH), 7.52 (4 H, d, J = 7.2 Hz, ArH), 7.34 (4 H, t, J = 7.5 Hz, ArH), 7.25 (4 H, t, J = 7.5 Hz, ArH), 5.60 (2 H, d, J = 6.3 Hz, NH(Thr)), 5.47 (1 H, d, J = 8.5 Hz, **H₁**), 5.31 (1 H, d, J = 3.4 Hz, **H₄**), 4.82 (1 H, dd, J = 10.7, 4.1 Hz, **H₃**), 4.35 (4 H, d, J = 7.3 Hz, **H₆**), 4.29–4.22 (2 H, m, **H₉**), 4.16 (2 H, t, J = 7.3 Hz, **H₅**), 4.11–4.00 (4 H, m, **H₁₀**), 3.93 (1 H, t, J = 6.6 Hz, **H₁₁**), 3.90–3.82 (4 H, m, **H₇**), 3.77 (1 H, dd, J = 11.6, 8.1 Hz, **H₂**), 2.29 (6 H, s, Me), 2.12 (6 H, s, Me), 1.99 (6 H, s, Me), 1.42 (18 H, s, Me); **HRMS** [ES⁺] found [M+H]⁺ 697.2712, C₃₅H₄₃N₄O₁₂ requires 697.2715.

2.7:

R_f 0.51 (2:3 EtOAc–Hexane); $^1\text{H NMR}$ (400 MHz, CDCl_3): δ 7.69 (4 H, dd, $J = 7.7$, 3.9 Hz, ArH), 7.55 (4 H, dt, $J = 8.0$, 2.7 Hz, ArH), 7.31 (4 H, dt, $J = 7.2$, 2.4 Hz, ArH), 7.10 (4 H, dd, $J = 7.7$, 3.2 Hz, ArH), 5.61 (1 H, d, $J = 9.8$ Hz, NH(Thr) $_{\alpha}$), 5.50 (1 H, d, $J = 9.8$ Hz, NH(Thr) $_{\beta}$), 5.40 (1 H, d, $J = 3.0$ Hz, $H_{4\alpha}$), 5.27 (1 H, dd, $J = 10.4$, 3.3 Hz, $H_{3\alpha}$), 5.24 (1 H, d, $J = 3.0$, $H_{4\beta}$), 5.04 (1 H, d, $J = 3.8$ Hz, $H_{1\alpha}$), 4.69 (1 H, dd, $J = 10.7$, 3.3 Hz, $H_{3\beta}$), 4.51 (1 H, qd, $J = 12.4$, 5.9, 2.1 Hz, $H_{7\alpha}$), 4.41 (1 H, d, $J = 8.0$, $H_{1\beta}$), 4.38 (1 H, dd, $J = 6.8$, 1.4 Hz, $H_{7\beta}$), 4.34 (1 H, dd, $J = 7.7$, 2.7 Hz, $H_{9\beta}$), 4.34–4.27 (3 H, m, $H_{5\alpha}$, $H_{5\beta}$), 4.27–4.13 (5 H, m, $H_{9\alpha}$, $H_{6\alpha}$, $H_{6\beta}$), 4.04 (4 H, dd, $J = 6.6$, 2.7 Hz, $H_{10\alpha}$, $H_{10\beta}$), 3.78 (1 H, t, $J = 6.8$ Hz, H_{11}), 3.57 (2 H, dd, $J = 11.1$, 3.4 Hz, $H_{2\alpha}$, $H_{2\beta}$), 2.28 (6 H, s, Me), 2.08 (6 H, s, Me), 2.00 (6 H, s, Me), 1.43 (18 H, s, t-Bu(Me)), 1.27 (6 H, dd, $J = 12.8$, 6.2 Hz, Me-8); **IR** ($\nu_{\text{max}}/\text{cm}^{-1}$): 3354 (C-H stretch), 2979 (N-H stretch), 2111 (N=N=N stretch), 1721 (C=O stretch); **HRMS** [ES $^+$] found [M+Na] $^+$ 733.2704, $\text{C}_{35}\text{H}_{42}\text{N}_4\text{O}_{12}\text{Na}$ requires 733.2691.

N-Fmoc-*O*-(2-acetimido-2-deoxy-3,4,6-Tri-*O*-acetyl- α -*D*-galactopyranosyl)-*L*-serine (2.9 α) and threonine tert-butyl ester (2.10 α)¹⁸⁸ & *N*-Fmoc-*O*-(2-acetimido-2-deoxy-3,4,6-Tri-*O*-acetyl- β -*D*-galactopyranosyl)-*L*-serine (2.9 β) and threonine tert-butyl ester (2.10 β)



Zinc dust (15.3 g, 234 mmol, 17.5 eq) was activated in 2% aq. CuSO_4 (120 mL). After 10 minutes, the zinc/copper mixture was added to crude **2.6** (9.34 g, 13.4 mmol, 1.0 eq) or crude **2.7** (9.51 g, 13.4 mmol, 1.0 eq) dissolved in a mixture of THF/AcOH/Ac $_2$ O (3:2:1, 180 ml). The mixture was stirred at room temperature for 4 hours. Once evolution of gas had ceased and TLC showed the reaction had reached completion; the mixture was filtered through celite and the pad was washed with THF (60 mL). The filtrate was concentrated in vacuo and residue dissolved in DCM (300

mL) and washed with 0.1 M HCl (2 × 400 mL). The organic portion was dried over MgSO₄, filtered and concentrated *in vacuo*. Crude acetamide was purified by flash column chromatography (1% MeOH in DCM) to yield pure **2.9α** (2.626 g, 3.68 mmol, 39%) and **2.9β** (2.098 g, 2.94 mmol, 31%) or **2.10α** (3.113 g, 4.28 mmol, 40%) and **2.10β** (2.537 g, 3.49 mmol, 33%) as glassy white foams.

NMR data is in agreement with reported data.¹⁸⁸

2.9α:

[α]_D²⁵ +59 (c 1, chloroform); *R*_f 0.65 (15:1 DCM–MeOH); ¹H NMR (400 MHz, DMSO-*d*₆): δ 7.76 (2 H, d, *J* = 7.5 Hz, ArH), 7.61 (2 H, d, *J* = 7.1 Hz, ArH), 7.40 (2 H, t, *J* = 7.4 Hz, ArH), 7.32 (2 H, t, *J* = 7.4 Hz, ArH), 5.80 (2 H, d, *J* = 7.9 Hz, NHAc, NH(Ser)), 5.37 (1 H, d, *J* = 2.8 Hz, **H**₄), 5.12 (1 H, dd, *J* = 11.2, 2.1 Hz, **H**₃), 4.83 (1 H, d, *J* = 3.3 Hz, **H**₁), 4.59 (1 H, td, *J* = 11.1, 3.3 Hz, **H**₂), 4.46–4.39 (3 H, m, **H**₉, **H**₁₀), 4.24 (1 H, t, *J* = 6.9 Hz, **H**₁₁), 4.15–3.99 (3 H, m, **H**₅, **H**₆), 3.96 (1 H, d, *J* = 9.5 Hz, **H**_{7a}), 3.84 (1 H, d, *J* = 8.5 Hz, **H**_{7b}), 2.15 (3 H, s, Me), 1.99 (3 H, s, Me), 1.99 (3 H, s, Me), 1.93 (3 H, s, Me), 1.48 (9 H, s, t-Bu(Me)); ¹³C NMR (100 MHz, DMSO-*d*₆): δ 170.9 (C=O), 170.4 (C=O), 170.3 (C=O), 170.1 (C=O), 169.0 (C=O), 155.9 (C=O), 143.8 (Ar), 141.3 (Ar), 127.8 (Ar), 127.1 (Ar), 125.1 (Ar), 120.1 (Ar), 98.8 (C₁), 83.1 (tBu(quat. C)), 69.3 (C₇), 68.4 (C₃), 67.3 (C₄), 67.3 (C₅), 62.0 (C₆), 54.8 (C₉), 53.5 (C₁₀), 47.6 (C₂), 47.1 (C₁₁), 28.1 (tBu(Me)), 23.2 (Me), 20.8 (Me), 20.8 (Me), 20.6 (Me); HRMS [ES⁺] found [M+H]⁺ 713.2925, C₃₆H₄₅N₂O₁₃ requires 713.2916.

2.10α:

[α]_D²⁵ +62 (c 1, chloroform); *R*_f 0.83 (15:1 DCM–MeOH); ¹H NMR (400 MHz, CDCl₃): δ 7.70 (2 H, d, *J* = 7.1 Hz, ArH), 7.56 (2 H, d, *J* = 7.2 Hz, ArH), 7.33 (2 H, t, *J* = 6.5, 6.5 Hz, ArH), 7.25 (2 H, t, *J* = 7.2, 7.2 Hz, ArH), 6.01 (1 H, d, *J* = 9.9 Hz, NH(Thr)), 5.62 (1 H, d, *J* = 9.4 Hz, NHAc), 5.31 (1 H, d, *J* = 2.5 Hz, **H**₄), 5.02 (1 H, dd, *J* = 11.3, 2.5 Hz, **H**₃), 4.82 (1 H, d, *J* = 3.1 Hz, **H**₁), 4.55 (1 H, td, *J* = 10.5, 10.5, 3.1 Hz, **H**₂), 4.45–4.31 (2 H, m, **H**_{6a}, **H**_{6b}), 4.22–4.12 (4 H, m, **H**₁₁, **H**₉, **H**₇, **H**₅), 4.04–3.97 (2 H, m, **H**₁₀), 2.08 (3 H, s, Me), 1.96 (3 H, s, Me), 1.92 (6 H, s, Me), 1.32 (9 H, s, OtBu), 1.24 (3 H, d, *J* = 6.1 Hz, **H**₈); ¹³C NMR (100 MHz, CDCl₃): 170.8 (C=O), 170.3 (C=O), 170.2 (C=O), 169.9 (C=O), 156.4 (C=O), 143.6 (Ar), 141.2 (Ar), 127.7 (Ar), 127.0 (Ar), 125.0 (Ar), 119.9 (Ar), 99.8 (C₁), 83.1 (tBu(quat. C)), 76.8 (C₁₁),

68.6 (C₃), 67.3 (C₄), 67.2 (C₅), 67.1 (C₆), 62.1 (C₁₀), 58.9 (C₉), 47.2 (C₂), 47.1 (C₇), 28.0 (tBu(Me)), 23.1 (Me), 20.6 (Me), 20.5 (Me), 18.5 (C₈); **IR** ($\nu_{\max}/\text{cm}^{-1}$): 3256 (s, C-H stretch), 2980 (s, C-H stretch), 1748 (s, C=O (ester) stretch), 1651 (s, C=O (amide) stretch); **HRMS** [ES⁺] found [M+H]⁺ 727.3072, C₃₇H₄₇N₂O₁₃ requires 727.3073).

2.9 β :

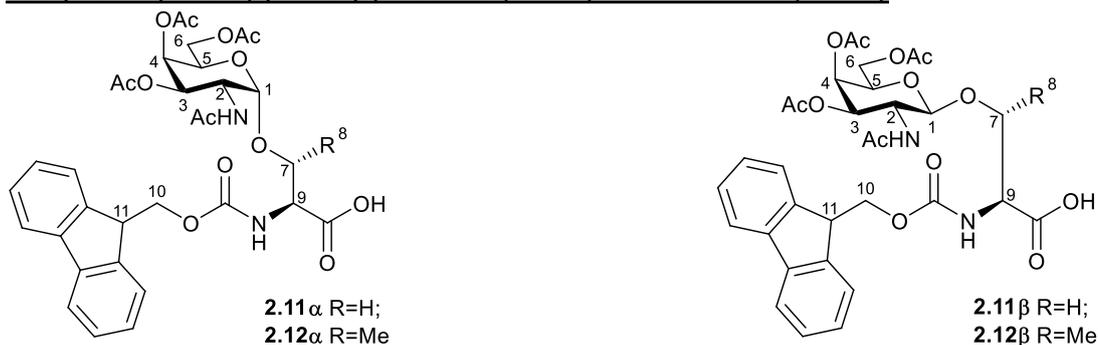
[α]_D²⁵ +4 (*c* 1, chloroform); **R_f** 0.57 (15:1 DCM–MeOH); **¹H NMR** (400 MHz, CDCl₃): δ 7.76 (2 H, d, *J* = 7.4 Hz, ArH), 7.65 (2 H, d, *J* = 7.3 Hz, ArH), 7.39 (2 H, t, *J* = 7.4 Hz, ArH), 7.32 (2 H, t, *J* = 7.3 Hz, ArH), 5.77 (1 H, d, *J* = 7.9 Hz, NH(Thr)), 5.51 (1 H, d, *J* = 8.3 Hz, NHAc), 5.34 (1 H, d, *J* = 3.1 Hz, H₄), 5.22 (1 H, dd, *J* = 11.2, 3.1 Hz, H₃), 4.65 (1 H, d, *J* = 8.3 Hz, H₁), 4.49–4.32 (3 H, m, H₉, H₁₀), 4.26–4.15 (2 H, m, H₁₁, H_{7a}), 4.11 (2 H, d, *J* = 6.6 Hz, H₆), 3.96–3.80 (3 H, m, H₂, H₅, H_{7b}), 2.12 (3 H, s, Me), 2.02 (3 H, s, Me), 1.99 (3 H, s, Me), 1.79 (3 H, s, Me), 1.46 (2 H, s, tBu(Me)); **¹³C NMR** (100 MHz, CDCl₃): δ 170.6 (C=O), 170.4 (C=O), 170.3 (C=O), 170.1 (C=O), 168.5 (C=O), 156.0 (C=O), 143.8 (Ar), 143.7 (Ar), 141.3 (Ar), 141.2 (Ar), 127.7 (Ar), 127.1 (Ar), 125.1 (Ar), 119.9 (Ar), 100.8 (C₁), 82.6 (tBu(quat.C)), 70.7 (C₂), 69.8 (C₃), 69.1 (C₇), 66.7 (C₁₀), 66.6 (C₄), 61.4 (C₆), 54.7 (C₉), 51.4 (C₅), 47.2 (C₁₁), 27.8 (tBu(Me)), 23.3 (Me), 20.6 (Me); **HRMS** [ES⁺] found [M+H]⁺ 713.2927, C₃₆H₄₅N₂O₁₃ requires 713.2916.

2.10 β :

[α]_D²⁵ –2 (*c* 1, chloroform); **R_f** 0.76 (15:1 DCM–MeOH); **¹H NMR** (400 MHz, CDCl₃): δ 7.69 (2 H, d, *J* = 7.6 Hz, ArH), 7.60 (2 H, t, *J* = 7.8 Hz, ArH), 7.32 (2 H, t, *J* = 7.4 Hz, ArH), 7.24 (2 H, t, *J* = 7.3 Hz, ArH), 5.70 (1 H, d, *J* = 9.1 Hz, NH(Thr)), 5.64 (1 H, d, *J* = 8.3 Hz, NHAc), 5.34–5.21 (2 H, m, H₃, H₄), 4.67 (1 H, d, *J* = 8.3 Hz, H₁), 4.36 (2 H, dd, *J* = 10.0, 7.3 Hz, H_{6a}, H₇), 4.30 (1 H, dd, *J* = 10.5, 7.3 Hz, H_{6b}), 4.19 (2 H, dd, *J* = 6.0, 3.5 Hz, H₉, H₅), 4.03 (2 H, d, *J* = 6.1 Hz, H₁₀), 3.84 (1 H, t, *J* = 6.6 Hz, H₁₁), 3.74 (1 H, dt, *J* = 14.8, 6.1 Hz, H₂), 2.02 (3 H, s, Me), 1.97 (3 H, s, Me), 1.93 (3 H, s, Me), 1.87 (3 H, s, Me), 1.42 (9 H, s, t-Bu(Me)), 1.10 (1 H, d, *J* = 6.1 Hz, H₈); **¹³C NMR** (100 MHz, CDCl₃): δ 170.6 (C=O), 170.5 (C=O), 170.4 (C=O), 170.2 (C=O), 169.1 (C=O), 156.9 (C=O), 143.8 (Ar), 141.3 (Ar), 127.7 (Ar), 127.1 (Ar), 125.3 (Ar), 119.9 (Ar), 97.5 (C₁), 82.2 (tBu(quat. C)), 73.2 (C₇), 70.5 (C₁₁), 69.5 (C₄), 67.0 (C₆), 66.7 (C₃), 61.4 (C₁₀), 59.0 (C₉), 52.0 (C₂), 47.2 (C₅), 27.9

(tBu(Me)), 23.5 (Me), 20.7 (Me), 20.6 (Me), 16.0 (C8); **IR** ($\nu_{\max}/\text{cm}^{-1}$): 3305 (s, C-H stretch), 2977 (s, C-H stretch), 1742 (s, C=O (ester) stretch), 1666 (s, C=O (amide) stretch); **HRMS** [ES⁺] found [M+H]⁺ 727.3071, C₃₇H₄₇N₂O₁₃ requires 727.3073.

N-Fmoc-*O*-(2-acetimido-2-deoxy-3,4,6-Tri-*O*-acetyl- α -*D*-galactopyranosyl)-*L*-serine (**2.11 α**) and threonine (**2.12 α**)¹⁸⁸ or *N*-Fmoc-*O*-(2-acetimido-2-deoxy-3,4,6-Tri-*O*-acetyl- β -*D*-galactopyranosyl)-*L*-serine (**2.11 β**) and threonine (**2.12 β**)



2.9 α (1.00 g, 1.38 mmol, 1.0 eq), **2.10 α** (1.00 g, 1.40 mmol, 1.0 eq), **2.9 β** (1.00 g, 1.38 mmol, 1.0 eq) or **2.10 β** (1.00 g, 1.40 mmol, 1.0 eq) was dissolved in 95% TFA (3.0 mL) and stirred at room temperature for 1 hour. Solvent was removed *in vacuo* and co-evaporated with toluene. The crude product purified by flash column chromatography (5:1 Toluene–EtOH) to yield **2.11 α** (671 mg, 1.02 mmol, 74%), **2.12 α** (854 mg, 1.27 mmol, 92%), **2.11 β** (662 mg, 1.01 mmol, 73%) or **2.12 β** (901 mg, 1.34 mmol, 96%) as a glassy off-white foams.

NMR data is in agreement with reported data.¹⁸⁸

2.11 α :

[α]_D²⁵ +88 (*c* 1, chloroform); **R_f** 0.47 (5:1 Toluene–EtOH); **¹H NMR** (500 MHz, DMSO-*d*₆): δ 7.90 (2 H, d, *J* = 7.5 Hz, ArH), 7.81 (1 H, d, *J* = 8.4 Hz, NHAc), 7.73 (1 H, d, *J* = 8.4 Hz, NH(Ser)), 7.72 (2 H, d, *J* = 7.1, ArH), 7.42 (2 H, t, *J* = 7.7 Hz, ArH), 7.33 (2 H, t, *J* = 7.8 Hz, ArH), 5.31 (1 H, d, *J* = 2.6 Hz, **H₄**), 5.07 (1 H, dd, *J* = 11.6, 3.2 Hz, **H₃**), 4.86 (1 H, d, *J* = 3.6 Hz, **H_I**), 4.38 (2 H, dd, *J* = 7.4, 3.6 Hz, **H₆**), 4.31–4.23 (3 H, m, **H₅**, **H₉**, **H₁₁**), 4.20 (1 H, ddd, *J* = 11.7, 8.4, 3.6 Hz, **H₂**), 4.05 (1 H, dd, *J* = 11.3, 5.8 Hz, **H_{10a}**), 3.98 (1 H, dd, *J* = 10.6, 7.0 Hz, **H_{10b}**), 3.83 (1 H, dd, *J* = 10.9, 4.1 Hz, **H_{7a}**), 3.77 (1 H, dd, *J* = 10.9, 4.8 Hz, **H_{7b}**), 2.10 (3H, s, Me), 1.95 (3H, s, Me), 1.90 (3H, s, Me), 1.81 (3H, s, Me); **¹³C NMR** (125 MHz, DMSO-*d*₆): δ 171.5 (C=O), 170.1 (C=O), 169.9 (C=O), 169.8 (C=O), 169.7 (C=O), 156.2 (C=O), 143.9

(Ar), 143.8 (Ar), 140.8 (Ar), 137.4 (Ar), 128.9 (Ar), 128.2 (Ar), 125.3 (Ar), 125.2 , 120.2 (Ar), 97.9 (C₁), 67.6 (C₇), 67.5 (C₃), 67.1 (C₄), 66.3 (C₅), 65.7 (C₆), 61.6 (C₁₀), 54.3 (C₉), 47.1 (C₂), 46.7 (C₁₁), 22.5 (Me), 21.1 (Me), 20.6 (Me), 20.5 (Me); **HRMS** [ES+] found [M+H]⁺ 657.2291, C₃₂H₃₇N₂O₁₃ requires 657.2290.

2.12 α :

[α]_D²⁵ +72 (*c* 1, chloroform); **R_f** 0.31 (5:1 Toluene–EtOH); **¹H NMR** (400 MHz, DMSO-*d*₆): δ 7.30 (2 H, d, *J* = 7.5 Hz, ArH), 7.74 (2 H, dd, *J* = 7.2, 3.6 Hz, ArH), 7.67 (1 H, d, *J* = 9.4 Hz, NHAc), 7.59 (1 H, d, *J* = 9.8 Hz, NH(Thr)), 7.42 (2 H, t, *J* = 7.4 Hz, ArH), 7.33 (2 H, t, *J* = 7.2 Hz, ArH), 5.32 (1 H, d, *J* = 2.8 Hz, H₄), 5.05 (1 H, dd, *J* = 11.6, 3.2 Hz, H₃), 4.81 (1 H, d, *J* = 3.8 Hz, H₁), 4.45 (2 H, ddd, *J* = 25.8, 10.8, 7.0 Hz, H₆), 4.34–4.27 (2 H, m, H₇, H₅), 4.26–4.19 (2 H, m, H₁₁, H₂), 4.15 (1 H, dd, *J* = 9.9, 1.2 Hz, H₉), 4.03 (2 H, d, *J* = 6.1 Hz, H₁₀), 2.11 (3H, s, Me), 1.99 (3H, s, Me), 1.91 (3H, s, Me), 1.84 (3H, s, Me), 1.18 (3 H, d, *J* = 6.4 Hz, H₈); **¹³C NMR** (100 MHz, DMSO-*d*₆): δ 170.7 (C=O), 170.1 (C=O), 170.0 (C=O), 169.9 (C=O), 169.5 (C=O), 156.9 (C=O), 143.8 (Ar), 140.8 (Ar), 127.7 (Ar), 127.1 (Ar), 125.2 (Ar), 120.2 (Ar), 98.8 (C₁), 74.9 (C₇), 67.6 (C₃), 67.2 (C₄), 66.4 (C₁₁), 65.6 (C₆), 62.0 (C₁₀), 58.4 (C₉), 46.8 (C₅), 46.6 (C₂), 22.7 (Me), 20.5 (Me), 18.6 (C₈); **IR** (ν_{\max} /cm⁻¹): 3329 (s, C-H stretch), 2941 (s, C-H stretch), 1744 (s, C=O (ester) stretch), 1657 (s, C=O (amide) stretch); **HRMS** [ES+] found [M+H]⁺ 671.2450, C₃₃H₃₉N₂O₁₃ requires 671.2447.

2.11 β :

[α]_D²⁰ +2 (*c* 1, chloroform); **R_f** 0.37 (5:1 Toluene–EtOH); **¹H NMR** (500 MHz, DMSO-*d*₆): δ 7.89 (2 H, d, *J* = 7.7 Hz, ArH), 7.84 (1 H, d, *J* = 9.5 Hz, NHAc), 7.74 (2 H, t, *J* = 6.8 Hz, ArH), 7.42 (2 H, t, *J* = 7.7 Hz, ArH), 7.34 (2 H, dd, *J* = 13.2, 5.5 Hz, ArH), 7.24 (1 H, d, *J* = 7.7 Hz, NH(Ser)), 5.23 (1 H, d, *J* = 3.6 Hz, H₄), 5.00 (1 H, dd, *J* = 11.4, 3.6 Hz, H₃), 4.59 (1 H, d, *J* = 8.7 Hz, H₁), 4.31 (2 H, d, *J* = 6.8 Hz, H₁₀), 4.26–4.19 (3 H, m, H₅, H₉, H₁₁), 4.04 (2 H, s, H₆), 4.01 (1 H, dd, *J* = 10.9, 5.0 Hz, H_{7a}), 3.88 (1 H, dt, *J* = 19.5, 9.1 Hz, H₂), 3.77 (1 H, dd, *J* = 10.4, 4.1 Hz, H_{7b}), 2.10 (3 H, s, Me), 1.99 (3 H, s, Me), 1.90 (3 H, s, Me), 1.75 (3 H, s, Me); **¹³C NMR** (125 MHz, DMSO-*d*₆): δ 171.4 (C=O), 170.0 (C=O), 169.9 (C=O), 169.8 (C=O), 169.6 (C=O), 156.0 (C=O), 143.8 (Ar), 143.7 (Ar), 140.8 (Ar), 140.7 (Ar), 137.4 (Ar), 128.9 (Ar), 128.2 (Ar), 127.7 (Ar), 127.1 (Ar), 125.3 (Ar), 120.1 (Ar), 101.0 (C₁), 70.4 (C₃), 70.0 (C₅), 68.5 (C₇), 66.6 (C₄), 65.9 (C₁₀), 61.4 (C₆), 54.0 (C₉), 49.3 (C₂),

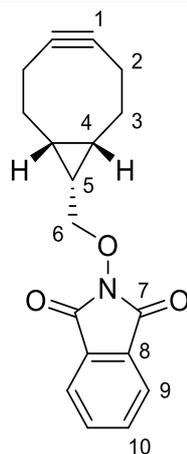
46.6 (C₁₁), 22.8 (Me), 21.1 (Me), 20.5 (Me), 20.5 (Me); **HRMS** [ES⁺] found [M+H]⁺ 657.2290, C₃₂H₃₇N₂O₁₃ requires 657.2290.

2.12β:

[α]_D²⁰ -4 (c 1, chloroform); **R_f** 0.34 (5:1 Toluene–EtOH); **¹H NMR** (400 MHz, DMSO-*d*₆): δ 7.94 (1 H, d, *J* = 8.8 Hz, NHAc), 7.88 (2 H, d, *J* = 7.3 Hz, ArH), 7.77 (2 H, dd, *J* = 7.3, 3.4 Hz, ArH), 7.41 (2 H, t, *J* = 7.3 Hz, ArH), 7.33 (2 H, t, *J* = 7.3 Hz, ArH), 6.46 (1 H, d, *J* = 8.8 Hz, NH(Thr)), 5.25 (1 H, d, *J* = 3.2 Hz, H₄), 5.04 (1 H, dd, *J* = 11.3, 3.2 Hz, H₃), 4.61 (1 H, d, *J* = 8.6 Hz, H₁), 4.32–4.21 (4 H, m, H₅, H₆, H₁₁), 4.09 (1 H, d, *J* = 8.8, 2.9 Hz, H₉), 4.06–4.00 (3 H, m, H₇, H₁₀), 3.85 (1 H, dt, *J* = 11.3, 8.6 Hz, H₂), 2.09 (3 H, s, Me), 1.98 (3 H, s, Me), 1.91 (3 H, s, Me), 1.82 (3 H, s, Me), 1.15 (3 H, d, *J* = 6.4 Hz, H₈); **¹³C NMR** (100 MHz, DMSO-*d*₆): δ 171.5 (C=O), 170.1 (C=O), 170.0 (C=O), 169.7 (C=O), 156.3 (C=O), 143.8 (Ar), 140.8 (Ar), 137.4 (Ar), 127.7 (Ar), 127.2 (Ar), 125.4 (Ar), 120.1 (Ar), 99.3 (C₁), 74.2(C₁₁), 70.2 (C₃), 69.8 (C₇), 66.4 (C₄), 66.2 (C₆), 61.1 (C₁₀), 58.5 (C₉), 49.7 (C₂), 46.7 (C₅), 21.1 (Me), 20.5 (Me), 16.7 (C₈); **HRMS** [ES⁺] found [M+H]⁺ 671.2460, C₃₃H₃₉N₂O₁₃ requires 671.2447.

7.1.2.2 Multivalent Inhibitors of Bacterial Toxins

2-(((1*R*,8*S*,9*S*)-Bicyclo[6.1.0]non-4-yn-9-yl)methoxy)isoindoline-1,3-dione (4.5)



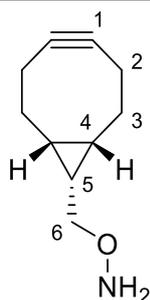
Endo-BCN-OH^a (20 mg, 0.13 mmol, 1.0 eq), triphenylphosphine (38 mg, 0.15 mmol, 1.1 eq) and N-hydroxyphthalimide (24 mg, 0.15 mmol, 1.1 eq) were dissolved in

^a Purchased from Merck

anhydrous DCM (1.2 mL) under a N₂ atmosphere and chilled to 0 °C. DIAD (29 μL, 0.15 mmol, 1.1 eq) was added dropwise and the reaction left stirring at 0 °C for 10 minutes, after which the reaction was allowed to warm to room temperature. After 4 hours the mixture was concentrated *in vacuo* to yield a crude yellow residue and purified by flash column chromatography (1:9 EtOAc/Hexane) to yield **4.5** as a white crystalline solid (37 mg, 94%).

R_f 0.58 (3:7 EtOAc/Hexane); ¹H NMR (500 MHz; CDCl₃) δ 7.83 (2 H, dd, *J* = 5.4, 3.1, Hz, **H₉**), 7.75 (2 H, dd, *J* = 5.4, 3.1 Hz, **H₁₀**), 4.31 (2 H, d, *J* = 8.0 Hz, **H₆**), 2.34–2.26 (4 H, m, **H₂**), 2.25–2.18 (2 H, m, **H_{3'}**), 1.69–1.60 (2 H, m, **H₃**), 1.60–1.52 (2 H, m, **H₅**), 1.09–1.02 (2 H, m, **H₄**); ¹³C NMR (125 MHz; CDCl₃) δ 163.8 (C₇), 134.6 (C₉), 129.1 (C₈), 123.6 (C₁₀), 99.0 (C₁), 76.4 (C₆), 29.3 (C₂), 21.5 (C₃), 20.8 (C₄), 17.3 (C₅); HRMS [ES⁺] found [M+Na]⁺ 318.1099, C₁₈H₁₇NO₃Na requires 318.1101.

O-(((1*R*,8*S*,9*S*)-Bicyclo[6.1.0]non-4-yn-9-yl)methyl)hydroxylamine (**4.1**)

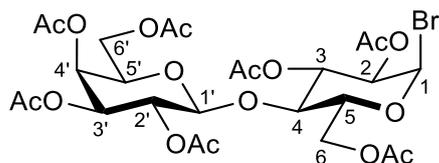


This reaction and purification was performed using oven dried, nitrogen flushed glassware in an acetone free atmosphere. The target compound has been previously found to sequester acetone from the atmosphere.¹⁸⁷

Compound **4.5** (20 mg, 68 μmol, 1.0 eq) was added to anhydrous 2 M methanolic methylamine (169 μL, 339 μmol, 5.0 eq). The reaction was shown to be complete within 2 minutes. The product was diluted in 1:9 EtOAc/Hexane and purified by flash column chromatography to yield **4.1** as a colourless residue (7.2 mg, 64%).

R_f 0.25 (3:7 EtOAc/Hexane); ¹H NMR (500 MHz; CDCl₃) δ 5.41 (2 H, br. s, NH₂), 3.75 (2 H, d, *J* = 7.7 Hz, **H₆**), 2.34–2.16 (6 H, m, **H₂**, **H_{3'}**), 1.62–1.52 (2 H, m, **H₃**), 1.35–1.25 (1 H, m, **H₅**), 0.96–0.86 (2 H, m, **H₄**); ¹³C NMR (100 MHz; CDCl₃) δ 99.1 (C₁), 73.2 (C₆), 29.4 (C₃), 21.6 (C₂), 20.0 (C₄), 17.6 (C₅); HRMS [ES⁺] found [M+H]⁺ 166.1225, C₁₀H₁₆NO requires 166.1226.

2,3,6-Tri-O-acetyl-4-(tetra-O-acetyl- β -D-galactopyranosyl)- α -D-glucopyranosyl bromide (4.6)

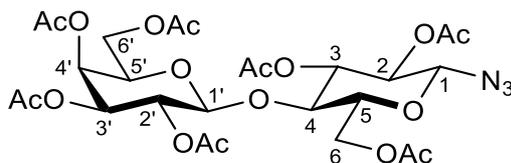


Lactose (5.00 g, 14.6 mmol, 1.0 eq) was added to acetic anhydride (17.0 mL, 175 mmol, 12.0 eq) under N_2 , followed by 33% (w/v) HBr in AcOH (3.3 mL, 18.4 mmol, 1.2 eq) and stirred at room temperature. After 1 hours, HBr in AcOH (17.0 mL, 94.8 mmol, 6.2 eq) was added and the reaction stirred at room temperature for a further 3 hours. The solvent was removed *in vacuo* whilst ensuring the mixture did not exceed 30 °C. The resulting orange solid was redissolved in DCM (100 mL) and washed with H_2O (50 mL), sat. aq. $NaHCO_3$ (50 mL) and brine (50 mL). The organic layer was dried over $MgSO_4$, filtered and concentrated *in vacuo* to yield crude **4.6** as a white solid (9.39 g, 92%).

NMR data is in agreement with reported data.²²⁷

R_f 0.70 (3:7 EtOAc/DCM); 1H NMR (500 MHz; $CDCl_3$) δ 6.51 (1 H, d, $J = 4.1$ Hz, H_1), 5.54 (1 H, t, $J = 9.7$ Hz, H_3), 5.34 (1 H, dd, $J = 3.5, 0.9$ Hz, $H_{4'}$), 5.11 (1 H, dd, $J = 10.4, 8.0$ Hz, $H_{2'}$), 4.94 (1 H, dd, $J = 10.4, 3.5$ Hz, $H_{3'}$), 4.74 (1 H, dd, $J = 9.9, 4.1$ Hz, H_2), 4.50 (1 H, d, $J = 7.9$ Hz, $H_{1'}$), 4.48 (1 H, d, $J = 10.0$ Hz, H_{6a}), 4.21–4.04 (4 H, m, $H_5, H_{6b}, H_{6a'b'}$), 3.90–3.82 (2 H, m, H_4, H_5), 2.15 (3 H, s, Me), 2.12 (3 H, s, Me), 2.08 (3 H, s, Me), 2.05 (3 H, s, Me), 2.05 (3 H, s, Me), 2.04 (3 H, s, Me), 1.94 (3 H, s, Me); ^{13}C NMR (125 MHz; $CDCl_3$) δ 170.5 (C=O), 170.3 (C=O), 170.3 (C=O), 170.2 (C=O), 170.1 (C=O), 169.3 (C=O), 169.1 (C=O), 100.9 ($C_{1'}$), 86.5 (C_1), 75.1 (C_4), 73.1 (C_5), 71.1 ($C_{3'}$), 71.0 (C_2), 70.9 ($C_{5'}$), 69.7 (C_3), 69.1 ($C_{2'}$), 66.7 (C_4), 61.2 (C_6), 61.0 ($C_{6'}$), 20.9 (Me), 20.9 (Me), 20.8 (Me), 20.8 (Me), 20.8 (Me), 20.8 (Me), 20.6 (Me); HRMS [ES+] found $[M+Na]^+$ 721.0958, $C_{26}H_{35}BrO_{17}Na$ requires 721.0950.

2,3,6-Tri-O-acetyl-4-(tetra-O-acetyl- β -D-galactopyranosyl)- β -D-glucopyranosyl azide (4.7)

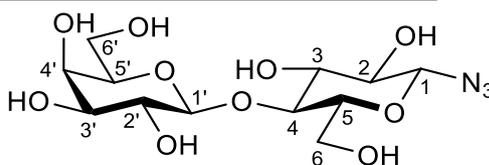


Peracetylated lactosyl bromide **4.6** (9.2 g, 13.2 mmol, 1.0 eq) was dissolved in anhydrous DMF (120 mL). To the solution NaN₃ (0.94 g, 14.5 mmol, 1.1 eq) was added and the reaction stirred at room temperature for 16 hours. The reaction was diluted with DCM (400 mL) and washed with NaHCO₃ (3 × 200 mL), H₂O (200 mL), then brine (200 mL). The organic layer was dried over MgSO₄, filtered, concentrated *in vacuo* and purified by flash column chromatography (1:1 EtOAc/Hexane) to yield **4.7** as a white crystalline solid (6.42 g, 74%).

NMR data is in agreement with reported data.²⁵⁴

R_f 0.24 (1:1 EtOAc/Hex); ¹H NMR (500 MHz; CDCl₃) δ 5.32 (1 H, dd, *J* = 3.5, 1.1 Hz, **H**_{4'}), 5.17 (1 H, t, *J* = 9.4 Hz, **H**₃), 5.06 (1 H, dd, *J* = 10.4, 7.9 Hz, **H**_{3'}), 4.93 (1 H, dd, *J* = 10.4, 3.5 Hz, **H**_{2'}), 4.82 (1 H, t, *J* = 9.0 Hz, **H**₂), 4.61 (1 H, d, *J* = 8.8 Hz, **H**₁), 4.47 (1 H, dd, *J* = 12.1, 2.1 Hz, **H**_{6a'}), 4.46 (1 H, d, *J* = 8.0 Hz, **H**_{1'}), 4.12–4.02 (3 H, m, **H**_{6ab}, **H**_{6b'}), 3.85 (1 H, td, *J* = 6.5, 1.0 Hz, **H**_{5'}), 3.79 (1 H, t, *J* = 9.5 Hz, **H**₄), 3.68 (1 H, ddd, *J* = 10.0, 5.0, 2.0 Hz, **H**₅), 2.12 (3H, s, Me), 2.10 (3H, s, Me), 2.04 (3H, s, Me), 2.03 (3H, s, Me), 2.01 (3H, s, Me), 2.01 (3H, s, Me), 1.93 (3H, s, Me); ¹³C NMR (125 MHz; CDCl₃) δ 170.4 (C=O), 170.4 (C=O), 170.2 (C=O), 170.1 (C=O), 169.7 (C=O), 169.6 (C=O), 169.1 (C=O), 101.2 (C_{1'}), 87.7 (C₁), 75.8 (C₄), 74.9 (C₅), 72.6 (C₃), 71.0 (C_{2'}), 71.0 (C₂), 70.8 (C_{5'}), 69.1 (C_{3'}), 66.7 (C_{4'}), 61.8 (C_{6'}), 60.9 (C₆), 20.9 (Me), 20.8 (Me), 20.7 (Me), 20.7 (Me), 20.7 (Me), 20.7 (Me), 20.6 (Me); HRMS [ES⁺] found [M+Na]⁺ 684.1879, C₂₆H₃₅N₃O₁₇Na requires 684.1859.

4-(β -D-galactopyranosyl)- β -D-glucopyranosyl azide (4.8)

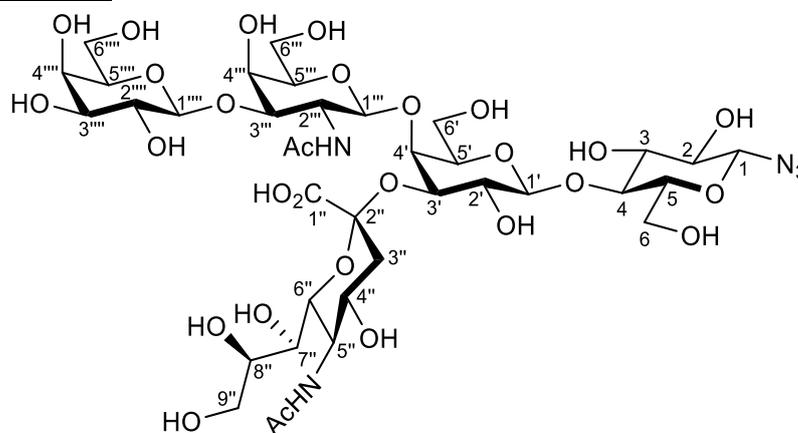


Peracetylated lactosyl azide **4.7** (5.82 g, 8.8 mmol, 1.0 eq) was dissolved in MeOH (40 mL) and NaOMe (3.8 g, 70.4 mmol, 8.0 eq) was added. Once TLC confirmed the

reaction had reached completion (~ 1 hour), washed DOWEX-50 H⁺ was added until the reaction was neutral. The resin was removed by vacuum filtration through celite and the solution concentrated *in vacuo*. The resulting oil was redissolved in H₂O (20 mL) and lyophilised to yield crude **4.8**. The crude lactosyl azide **4.8** was purified by size exclusion chromatography using Biogel P2 column equilibrated with 20 mM ammonium formate. β-lactosyl azide **4.8** was obtained as a white foam (2.29 g, 71%). NMR data is in agreement with reported data.²⁵⁴

R_f 0.20 (6:4:0.8 CHCl₃/MeOH/H₂O); ¹H NMR (500 MHz; CDCl₃) δ 4.77 (1 H, d, *J* = 9.2 Hz, **H₁**), 4.45 (1 H, d, *J* = 7.7 Hz, **H_{1'}**), 4.00 (1 H, dd, *J* = 12.4, 1.6 Hz, **H_{6a}**), 3.92 (1 H, dd, *J* = 3.4, 0.6 Hz, **H_{4'}**), 3.86–3.64 (8 H, m, **H₃**, **H_{3'}**, **H₄**, **H₅**, **H_{5'}**, **H_{6b}**, **H_{6a'b'}**), 3.54 (1 H, dd, *J* = 10.0, 7.7 Hz, **H_{2'}**), 3.35–3.29 (1 H, m, **H₂**); ¹³C NMR (125 MHz; CDCl₃) δ 102.9 (C₁), 89.9 (C₁), 77.7 (C₄), 76.7 (C₅), 75.3 (C_{5'}), 74.3 (C₃), 72.5 (C₂), 72.5 (C_{3'}), 70.9 (C_{2'}), 68.5 (C_{4'}), 61.0 (C₆), 59.8 (C_{6'}); HRMS [ES⁺] found [M+Na]⁺ 390.1123, C₁₂H₂₁N₃O₁₀Na requires 390.1119.

β-GM1 azide (4.11)



GM1os⁸² (9 mg, 9 μmol, 1 eq), 2-Chloro-1,3-dimethylimidazolium chloride (DMC, 15 mg, 90 μmol, 10 eq) and NaN₃ (25 mg, 400 μmol, 44 eq) were dissolved in H₂O (0.1 mL) in an Eppendorf tube. TEA (22.5 μL, 161 μmol, 18 eq) was added and mixed by vortex and left at 37 °C for 48 hours. The reaction mixture was purified by size exclusion chromatography using Biogel P2 column equilibrated with 20 mM ammonium formate to yield β-GM1 azide **4.11** as a white foam (8.6 mg, 93%).

R_f 0.58 (2:2:1 n-BuOH/MeOH/H₂O); ¹H NMR (500 MHz; D₂O; 275K) δ 4.78 (1 H, d, *J* = 8.8 Hz, **H_{1β}**), 4.75 (1 H, d, *J* = 8.8 Hz, **H_{1'''}**), 4.53 (1 H, d, *J* = 7.7 Hz, **H_{1'}**),

4.52 (1 H, d, $J = 7.8$ Hz, $H_{1''''}$), 4.19–4.11 (3 H, m, $H_4, H_{2''}, H_{4''}$), 4.01 (1 H, dd, $J = 10.8, 8.7$ Hz, $H_{3'}$), 3.99–3.54 (24 H, m, *not assigned*), 3.53–3.45 (2 H, m, $H_{3''}, H_4$), 3.34 (1 H, dd, $J = 9.3, 8.2$ Hz, H_2), 2.63 (1 H, dd, $J = 12.3, 4.5$ Hz, $H_{3''equatorial}$), 2.01 (3H, s, *Ac*), 1.99 (3 H, s, *Ac*), 1.94 (1 H, t, $J = 12.3$ Hz, $H_{3''axial}$); HRMS [ES+] found $[M+Na]^+$ 1046.3417, $C_{37}H_{61}N_5O_{28}Na$ requires 1046.3401.

7.2 Solid Phase Peptide Synthesis

7.2.1 General Reagents and Equipment

All amino acids, resins and coupling reagents were purchased from Novabiochem and used without further purification. Fritted polypropylene tubes (2.5 mL) were purchased from Supelco (Merck) and were used for all solid phase reactions. Agitation of solid phase reaction mixture was achieved by rotation on a Stuart blood rotator at room temperature. Peptides and glycopeptides were synthesised as C-terminal amides using Rink Amide MBHA LL resin (50 μ M, loading 0.35 mmol/g).

Peptides were analysed by HPLC using an Agilent 1290 affinity LC system equipped with an Ascentis Express 10 cm \times 2.1 mm, 2.7 μ m ES-C18 peptide column (0.5 ml min^{-1}) and ultraviolet (UV) detection at 220–280 nm. The peptide column was run with a gradient from 0.1% TFA/5% MeCN (v/v) in H_2O to 0.1% TFA/95% MeCN (v/v) in H_2O over 10 minutes.

7.2.2 Standard Procedures for Solid Phase Peptide Synthesis

7.2.2.1 Preparation of resin

The resin was added to fritted polypropylene tube and swollen in DMF (2 mL) for 1 hour. Resin was washed with DMF (2×2 mL \times 2 min spins). Swollen resin was Fmoc protected and required Fmoc deprotection prior to use.

7.2.2.2 Fmoc deprotection and peptide elongation

After DMF wash, resin was treated with piperidine solution (20% (v/v) in DMF) (3×2 mL \times 3 min spins). Resin was washed with DMF (3×2 mL \times 3 min spins). Peptide

elongation was carried out using Fmoc-amino acid (5 equiv), HCTU (4.9 equiv) dissolved in DMF (2 mL); Mixture transferred to swollen resin and DIPEA (10 equiv) added. Mixture left to spin for 40 minutes. Resin isolated by through filtration and washed with DMF ($3 \times 2 \text{ mL} \times 3 \text{ min spins}$).

For the glycosylated amino acids, Fmoc-(GalNAc)-amino acids (**2.11a** & **2.12a**, 2 equiv), HCTU (4.9 equiv) were dissolved in DMF (2 mL) and added to the swollen resin. DIPEA (10 equiv) were added and spin for 12 hours. Resin was isolated by filtration and washed with DMF ($3 \times 2 \text{ mL} \times 3 \text{ min spins}$).

7.2.2.3 Coupling reaction analysis

A small number of resin beads were isolated prior to Fmoc-deprotection and exposed to TFA (100 μL) for 2 mins before quenching with methanol (1 mL). The solution was filtered and analysed by LCMS. If any starting material was seen, the coupling was repeated.

7.2.2.4 Deprotection of *O*-acetyl groups

Following the coupling of the final amino acid, the resin was washed with DMF ($2 \times 2 \text{ mL} \times 3 \text{ min spins}$), DCM ($2 \times 2 \text{ mL} \times 3 \text{ min spins}$) and methanol ($2 \times 2 \text{ mL} \times 3 \text{ min spins}$). The resin was treated with 70% hydrazine hydrate in MeOH ($3 \times 2 \text{ mL} \times 5 \text{ min spins}$). Resin was washed with methanol ($2 \times 2 \text{ mL} \times 3 \text{ min spins}$). The efficiency of the *O*-acetyl deprotection was checked following the methodology set out in section 7.2.2.3. If required a further treatment with hydrazine was performed until all *O*-acetyl groups had been removed. Once all *O*-acetyl groups had been removed, the resin was washed with methanol ($2 \times 2 \text{ mL} \times 3 \text{ min spins}$), DCM ($2 \times 2 \text{ mL} \times 3 \text{ min spins}$) and DMF ($2 \times 2 \text{ mL} \times 3 \text{ min spins}$).

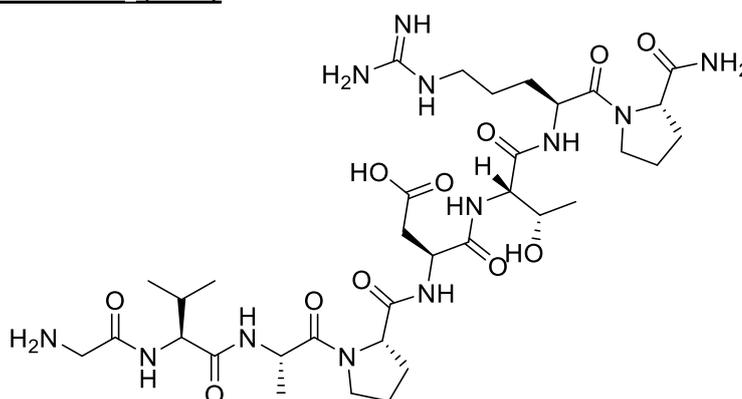
7.2.2.5 Cleavage and isolation of peptide

Following the *O*-Acetyl and/or Fmoc deprotection, the amino acid-resin was washed with DMF ($2 \times 2 \text{ mL} \times 3 \text{ min spins}$), DCM ($2 \times 2 \text{ mL} \times 3 \text{ min spins}$) and methanol ($2 \times 2 \text{ mL} \times 3 \text{ min spins}$). The resin was isolated by filtration before drying on high

vacuum for 3 hours. A cleavage cocktail of TFA (95%), H₂O (2.5%) and TIS (2.5%) (total 2 mL) was added to the resin and the mixture was left to spin for 2 hours. The resin was filtered into falcon tube and washed with TFA (3 × 1 mL × 1 min spins). The volume of TFA solution was reduced under a stream of N₂ to approx. 1 mL. The peptide was precipitated with ice cold Et₂O (10 mL) and pelleted by centrifugation at 3150 x g for 10 mins. The Et₂O was decanted and peptide pellet washed with ice cold Et₂O (10 mL) and re-pelleted as previously. Et₂O was decanted, the pellet was dried under stream of nitrogen and dissolved in the minimum volume of water before being lyophilised overnight.

7.2.3 Synthesised Peptides & Glycopeptides

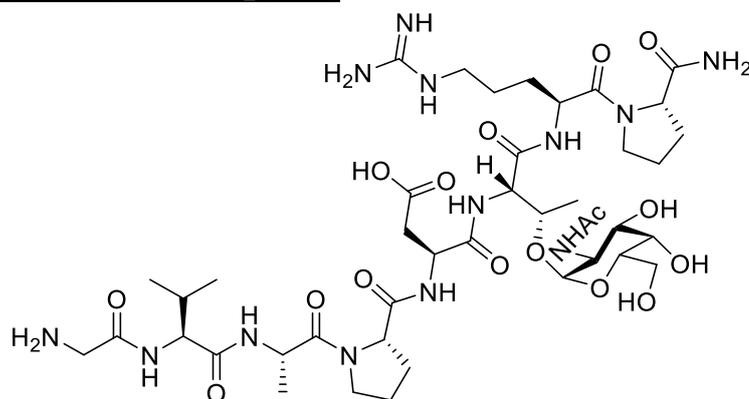
H₂N-GVAPDTRP-CONH₂ (2.13)



Peptide was synthesised using methodology described in section 7.2.2.1 to 7.2.2.5.
Yield: 30 mg, 58%.

HRMS [ES+] found [M+H]⁺ 811.4433, C₃₄H₅₉N₁₂O₁₁ requires 811.4421; **LCMS** [ES+] found [M+H]⁺ 811.55, C₃₄H₅₉N₁₂O₁₁ requires 811.44.

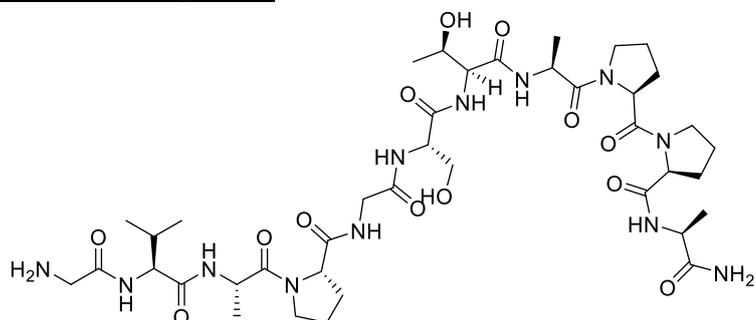
H₂N-GVAPDT(GalNAc)RP-CONH₂ (**2.14**)



Glycopeptide was synthesised using methodology described in sections 7.2.2.1 to 7.2.2.5. Yield: 28 mg, 41%.

HRMS [ES+] found [M+H]⁺ 1014.5224, C₄₀H₇₁N₁₃O₁₆ requires 1014.5214; **LCMS** [ES+] found [M+H]⁺ 1014.63, C₄₀H₇₁N₁₃O₁₆ requires 1014.52.

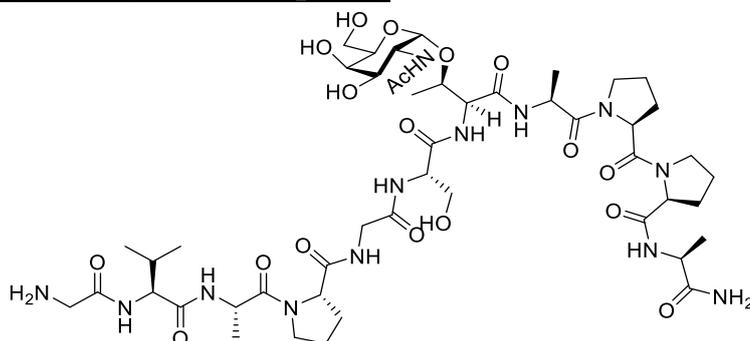
H₂N-GVAPGSTAPPA-CONH₂ (**2.15**)



Peptide was synthesised using methodology described in sections 7.2.2.1 to 7.2.2.5. Yield: 30 mg, 65%.

HRMS [ES+] found [M+H]⁺ 923.4960, C₄₀H₆₆N₁₂O₁₃ requires 923.4945; **LCMS** [ES+] found [M+H]⁺ 923.61, C₄₀H₆₆N₁₂O₁₃ requires 923.49.

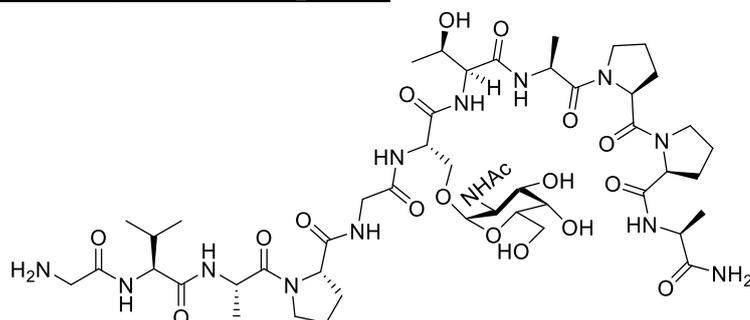
H₂N-GVAPGST(GalNAc)APPA-CONH₂ (2.16)



Glycopeptide was synthesised using methodology described in sections 7.2.2.1 to 7.2.2.5. Yield: 45 mg, 80%.

HRMS [ES+] found [M+H]⁺ 1126.5759, C₄₈H₇₉N₁₃O₁₈ requires 1126.5739; **LCMS** [ES+] found [M+H]⁺ 1126.69, C₄₈H₇₉N₁₃O₁₈ requires 1126.57.

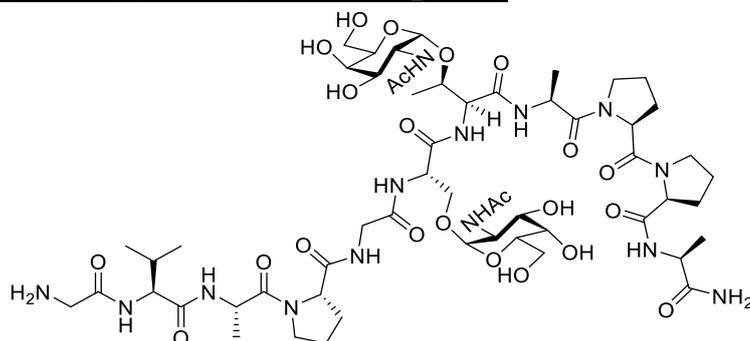
H₂N-GVAPGS(GalNAc)TAPPA-CONH₂ (2.17)



Glycopeptide was synthesised using methodology described in sections 7.2.2.1 to 7.2.2.5. Yield: 47 mg, 75%.

HRMS [ES+] found [M+H]⁺ 1126.5765, C₄₈H₇₉N₁₃O₁₈ requires 1126.5739; **LCMS** [ES+] found [M+H]⁺ 1126.70, C₄₈H₇₉N₁₃O₁₈ requires 1126.57.

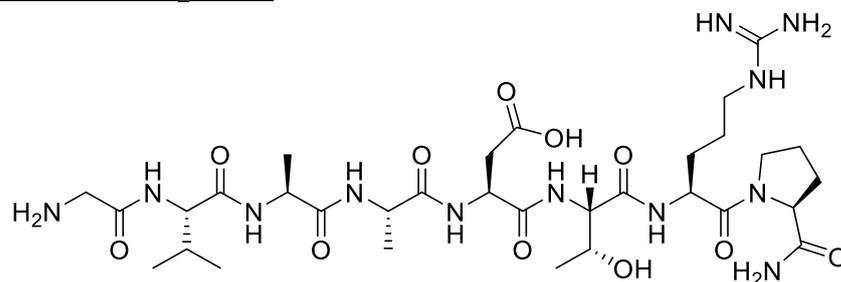
H₂N-GVAPGS(GalNAc)T(GalNAc)APPA-CONH₂ (2.18)



Glycopeptide was synthesised using methodology described in sections 7.2.2.1 to 7.2.2.5. Yield: 40 mg, 55%.

HRMS [ES+] found [M+H]⁺ 1329.6536, C₅₆H₉₂N₁₄O₂₃ requires 1329.6533; **LCMS** [ES+] found [M+H]⁺ 1329.78, C₅₆H₉₂N₁₄O₂₃ requires 1329.65.

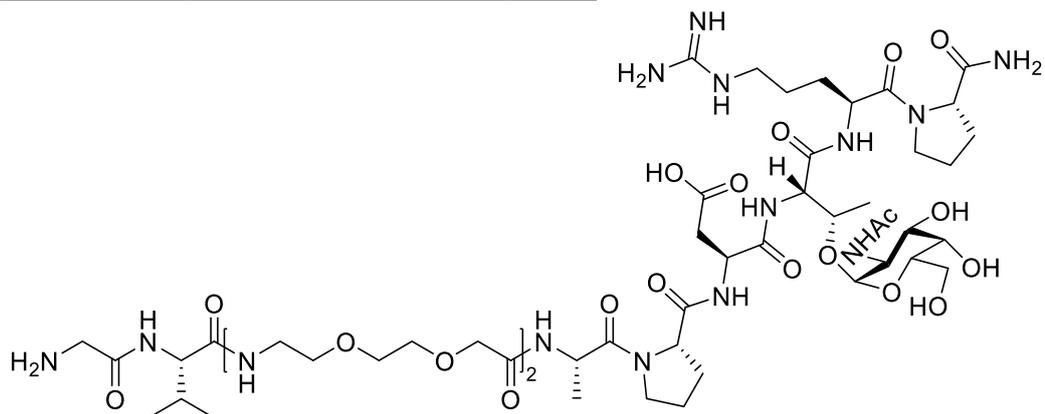
H₂N-GVAADTRP-CONH₂ (2.19)



Glycopeptide was synthesised using methodology described in sections 7.2.2.1 to 7.2.2.5. Yield: 32 mg, 82%.

HRMS [ES+] found [M+H]⁺ 785.4274, C₃₂H₅₇N₁₂O₁₁ requires 785.4264; **LCMS** [ES+] found [M+H]⁺ 785.41, C₃₂H₅₇N₁₂O₁₁ requires 785.43.

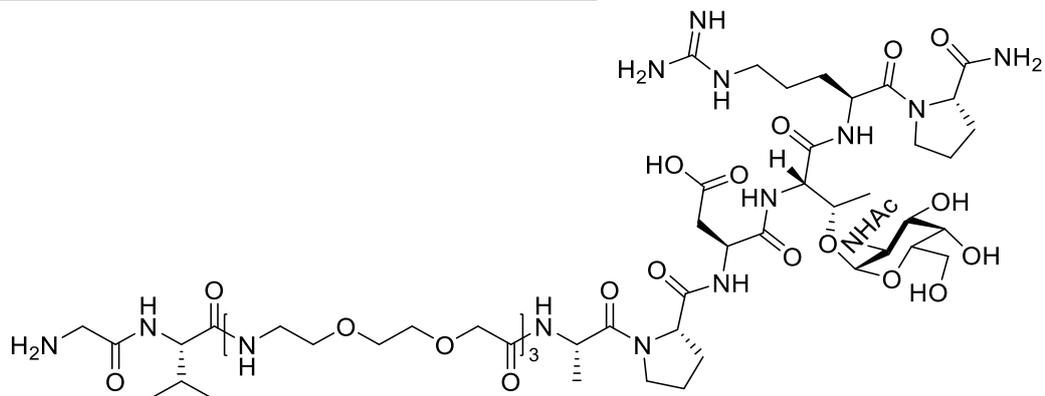
H₂N-GV(AEEAc)₂APDT(GalNAc)RP-CONH₂ (2.22)



Glycopeptide was synthesised using methodology described in sections 7.2.2.1 to 7.2.2.5. Yield: 42 mg, 64%.

HRMS [ES⁺] found [M+H]⁺ 1304.6702, C₅₄H₉₄N₁₅O₂₂ requires 1304.6692; **LCMS** [ES⁺] found [M+H]⁺ 1305.24, C₅₄H₉₄N₁₅O₂₂ requires 1304.67.

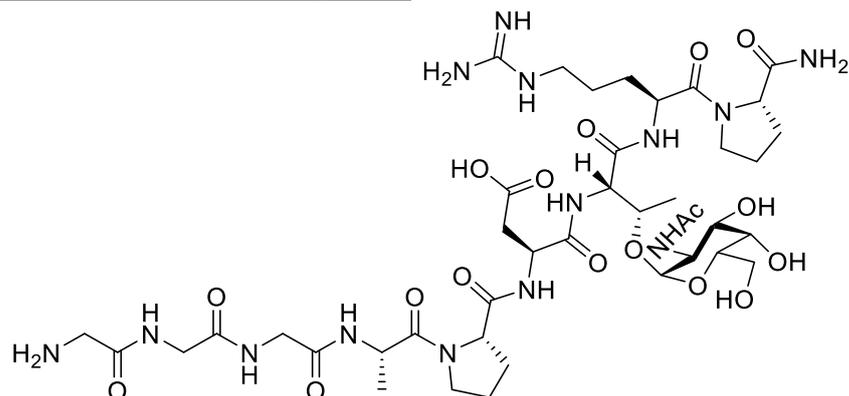
H₂N-GV(AEEAc)₃APDT(GalNAc)RP-CONH₂ (2.23)



Glycopeptide was synthesised using methodology described in sections 7.2.2.1 to 7.2.2.5. Yield: 50 mg, 69%.

HRMS [ES⁺] found [M+2H]²⁺ 725.3766, C₆₀H₁₀₆N₁₆O₂₅ requires 1450.7504; **LCMS** [ES⁺] found [M+2H]²⁺ 725.76, C₆₀H₁₀₆N₁₆O₂₅ requires 1450.75.

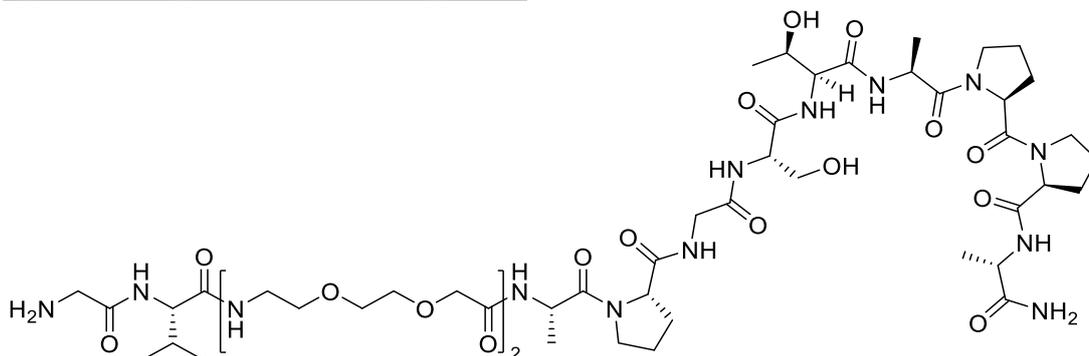
H₂N-GGGAPDT(GalNAc)RP-CONH₂ (2.24)



Glycopeptide was synthesised using methodology described in sections 7.2.2.1 to 7.2.2.5. Yield: 48 mg, 93%.

HRMS [ES+] found [M+H]⁺ 1029.4956, C₄₁H₆₉N₁₄O₁₇ requires 1029.4960; **LCMS** [ES+] found [M+H]⁺ 1030.04, C₄₁H₆₉N₁₄O₁₇ requires 1029.50.

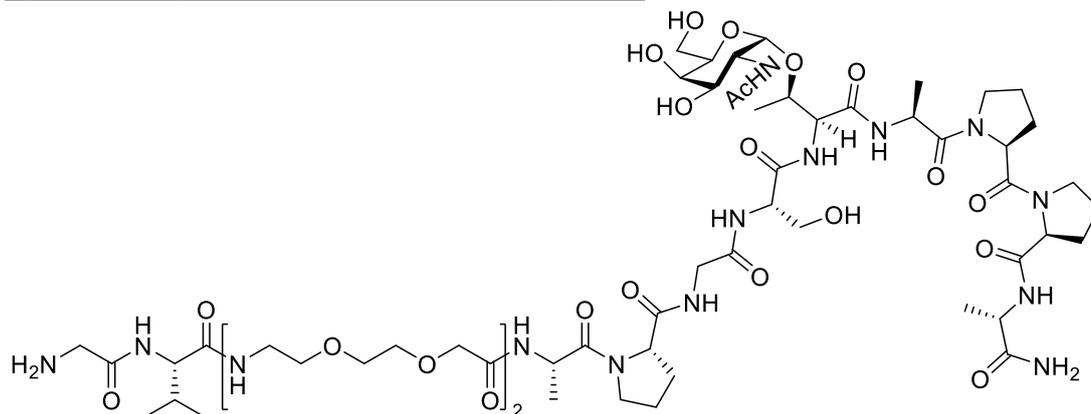
H₂N-GV(PEG)₆APGSTAPPA-CONH₂ (2.25)



Peptide was synthesised using methodology described in sections 7.2.2.1 to 7.2.2.5. Yield: 42 mg, 84%.

HRMS [ES+] found [M+H]⁺ 1213.6423, C₅₂H₈₉N₁₄O₁₉ requires 1213.6423; **LCMS** [ES+] found [M+H]⁺ 1214.29, C₅₂H₈₉N₁₄O₁₉ requires 1213.64.

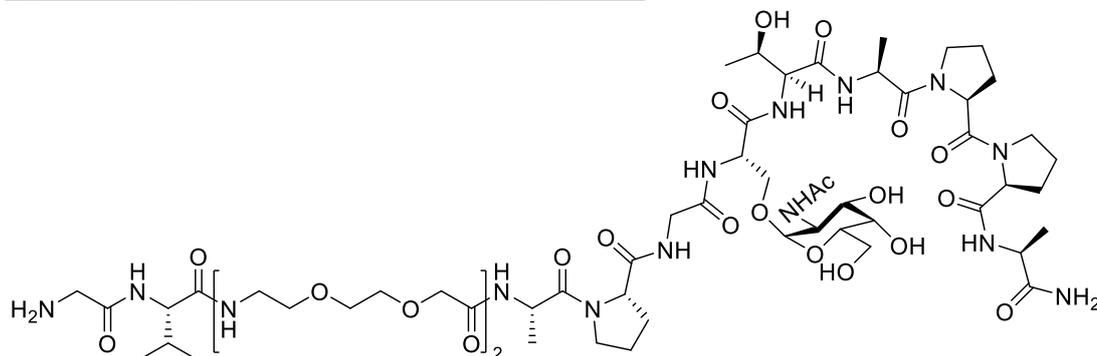
H₂N-GV(PEG)₆APGST(GalNAc)APPA-CONH₂ (2.26)



Glycopeptide was synthesised using methodology described in sections 7.2.2.1 to 7.2.2.5. Yield: 48 mg, 76%.

HRMS [ES⁺] found [M+H]⁺ 1416.7214, C₆₀H₁₀₂N₁₅O₂₄ requires 1416.7217; **LCMS** [ES⁺] found [M+H]⁺ 1417.40, C₆₀H₁₀₂N₁₅O₂₄ requires 1416.72.

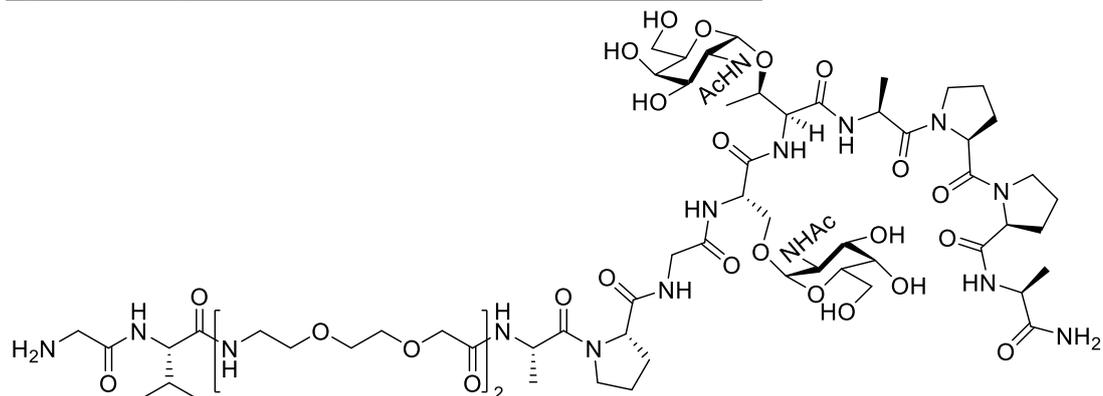
H₂N-GV(PEG)₆APGS(GalNAc)TAPPA-CONH₂ (2.27)



Glycopeptide was synthesised using methodology described in sections 7.2.2.1 to 7.2.2.5. Yield: 43 mg, 68%.

HRMS [ES⁺] found [M+H]⁺ 1416.7222, C₆₀H₁₀₂N₁₅O₂₄ requires 1416.7217; **LCMS** [ES⁺] found [M+H]⁺ 1417.42, C₆₀H₁₀₂N₁₅O₂₄ requires 1416.72.

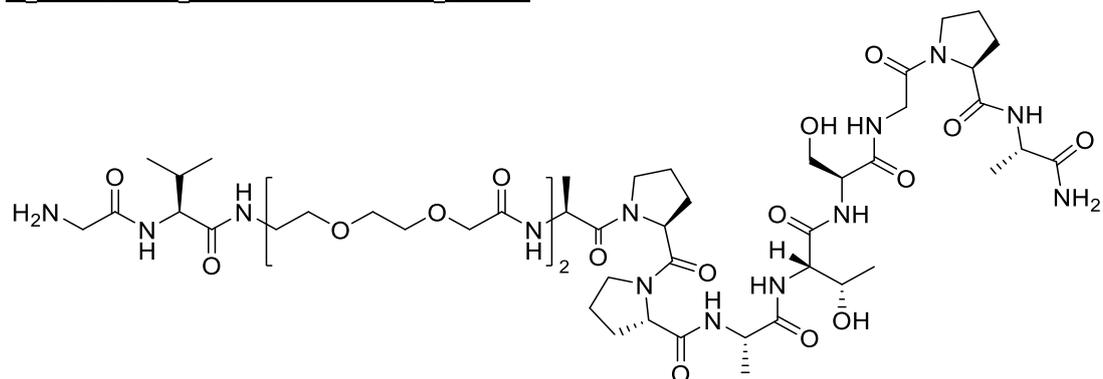
H₂N-GV(PEG)₆APGS(GalNAc)T(GalNAc)APPA-CONH₂ (2.28)



Glycopeptide was synthesised using methodology described in sections 7.2.2.1 to 7.2.2.5. Yield: 55 mg, 75%.

HRMS [ES⁺] found [M+2H]²⁺ 810.4036, C₆₈H₁₁₆N₁₆O₂₉ requires 1620.8094; **LCMS** [ES⁺] found [M+H]⁺ 1620.57, C₆₈H₁₁₅N₁₆O₂₉ requires 1619.80.

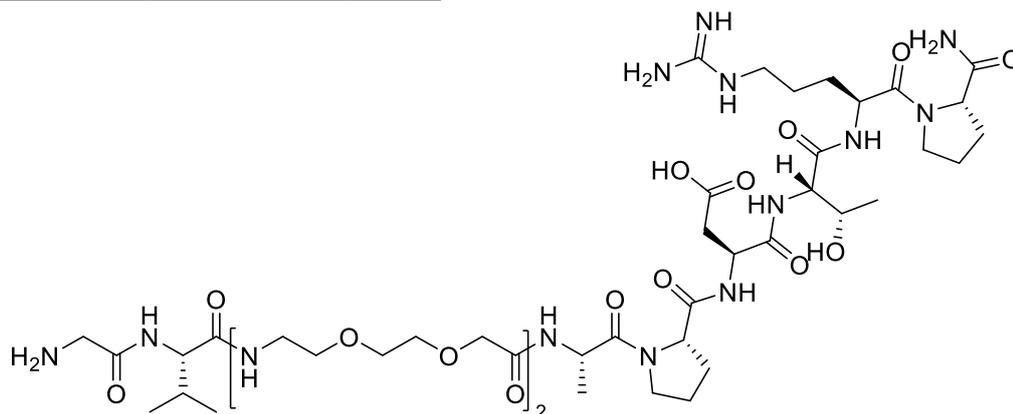
H₂N-GV(PEG)₆APPATSGPA-CONH₂ (2.29)



Peptide was synthesised using methodology described in sections 7.2.2.1 to 7.2.2.5. Yield: 41 mg, 77%.

HRMS [ES⁺] found [M+H]⁺ 1213.6418, C₅₂H₈₉N₁₄O₁₉ requires 1213.6423; **LCMS** [ES⁺] found [M+H]⁺ 1214.31, C₅₂H₈₉N₁₄O₁₉ requires 1213.64.

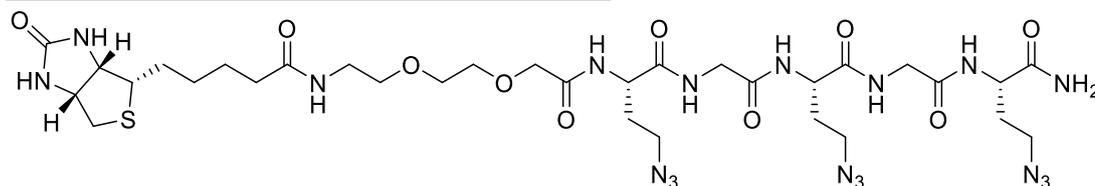
H₂N-GV(PEG)₆APDTRP-CONH₂ (2.30)



Peptide was synthesised using methodology described in sections 7.2.2.1 to 7.2.2.5.
Yield: 37 mg, 77%.

HRMS [ES⁺] found [M+H]⁺ 1101.5903, C₄₆H₈₁N₁₄O₁₇ requires 1101.5899; **LCMS** [ES⁺] found [M+H]⁺ 1102.20, C₄₆H₈₁N₁₄O₁₇ requires 1101.59.

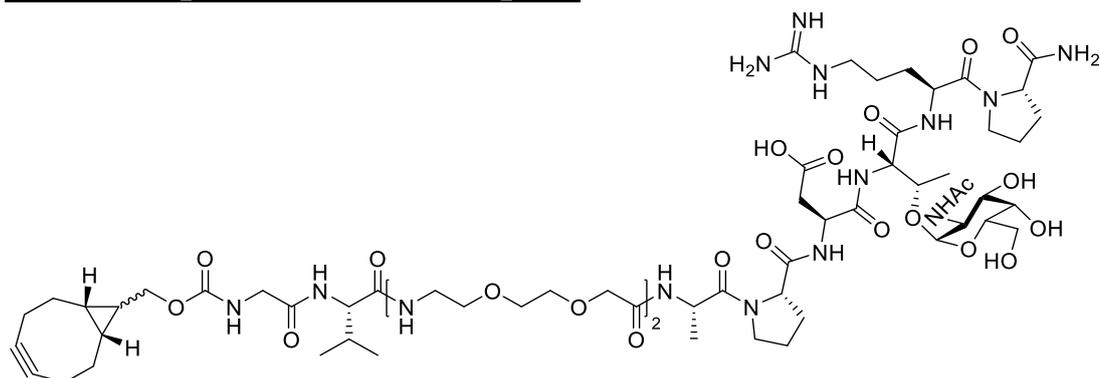
Biotin-PEG₃-Aha-Gly-Aha-Gly-Aha-CONH₂ (3.5)



Peptide synthesised on 100 μ M scale. Once the peptide had been synthesised using methodology described in sections 7.2.2.1 to 7.2.2.3, the peptide was treated with Biotin-OSu (2 equiv.) and DIPEA (10 equiv.) in DMF. The resin was spun overnight, then washed with DMF (3 \times 5 mL) before the methodology described in section 7.2.2.5 was carried out. Yield: 78 mg, 89%.

HRMS [ES⁺] found [M+NH₄]⁺ 897.3873, C₃₂H₅₆N₁₉O₁₀S requires 898.4173; **LCMS** [ES⁺] found [M+NH₄]⁺ 897.29, C₃₂H₅₆N₁₉O₁₀S requires 898.42.

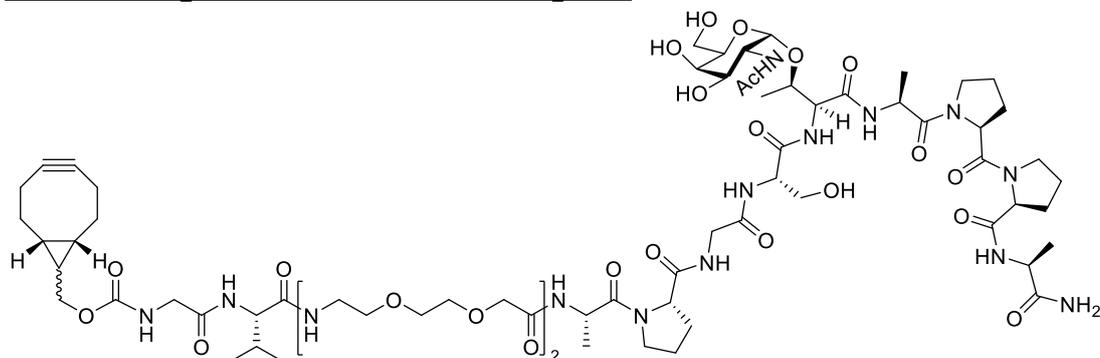
BCN-GV(PEG)₆APDT(GalNAc)RP-CONH₂ (3.7)



Glycopeptide **2.22** (20 mg, 15 μ mol, 1.0 eq) and BCN-NHS (5.4 mg, 18 μ mol, 1.2 eq) were dissolved in dry DMF (1 mL). TEA (6.4 μ L, 46 μ mol, 3.0 eq) was added and the mixture stirred overnight at room temperature. Reaction concentrated *in vacuo* and residue dissolved in ddH₂O (1 mL) and lyophilised to dryness. Residue purified by size-exclusion using Biogel P2 column eluting in 20 mM ammonium formate. Yield: 15.7 mg, 70%.

HRMS [ES⁺] found [M+H+Na]²⁺ 751.8718, C₆₅H₁₀₆N₁₅NaO₂₄ requires 1503.7422;
LCMS [ES⁺] found [M+2H]²⁺ 740.80, C₆₅H₁₀₇N₁₅O₂₄ requires 1481.76.

BCN-GV(PEG)₆APGST(GalNAc)APPA-CONH₂ (3.8)

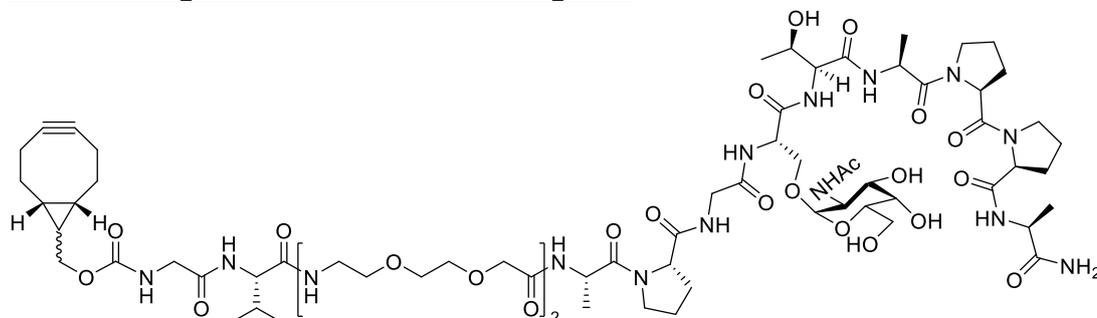


Glycopeptide **2.26** (20 mg, 16 μ mol, 1.0 eq) and BCN-NHS (5.5 mg, 19 μ mol, 1.2 eq) were dissolved in dry DMF (1 mL). TEA (6.6 μ L, 47 μ mol, 3.0 eq) was added and the mixture stirred overnight at room temperature. Reaction concentrated *in vacuo* and residue dissolved in ddH₂O (1 mL) and lyophilised to dryness. Residue purified by size-exclusion using Biogel P2 column eluting in 20 mM ammonium formate. Yield: 20.6 mg, 81%.

HRMS [ES+] found [M+2Na]²⁺ 818.8889, C₇₁H₁₁₃N₁₅O₂₆Na₂ requires 1637.7766;

LCMS [ES+] found [M+2H]²⁺ 796.92, C₇₁H₁₁₅N₁₅O₂₆ requires 1593.81.

BCN-GV(PEG)₆APGS(GalINAc)TAPPA-CONH₂ (3.9)

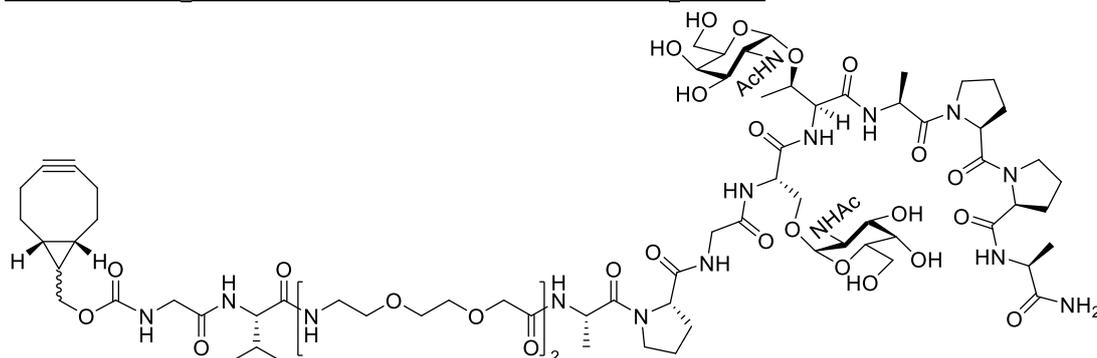


Glycopeptide **2.27** (20 mg, 16 μmol, 1.0 eq) and BCN-NHS (5.5 mg, 19 μmol, 1.2 eq) were dissolved in dry DMF (1 mL). TEA (6.6 μL, 47 μmol, 3.0 eq) was added and the mixture stirred overnight at room temperature. Reaction concentrated *in vacuo* and residue dissolved in ddH₂O (1 mL) and lyophilised to dryness. Residue purified by size-exclusion using Biogel P2 column eluting in 20 mM ammonium formate. Yield: 18.9 mg, 74%.

HRMS [ES+] found [M+2Na]²⁺ 818.8897, C₇₁H₁₁₃N₁₅O₂₆Na₂ requires 1637.7766;

LCMS [ES+] found [M+2H]²⁺ 796.82, C₇₁H₁₁₅N₁₅O₂₆ requires 1593.81.

BCN-GV(PEG)₆APGS(GalINAc)T(GalINAc)APPA-CONH₂ (3.10)



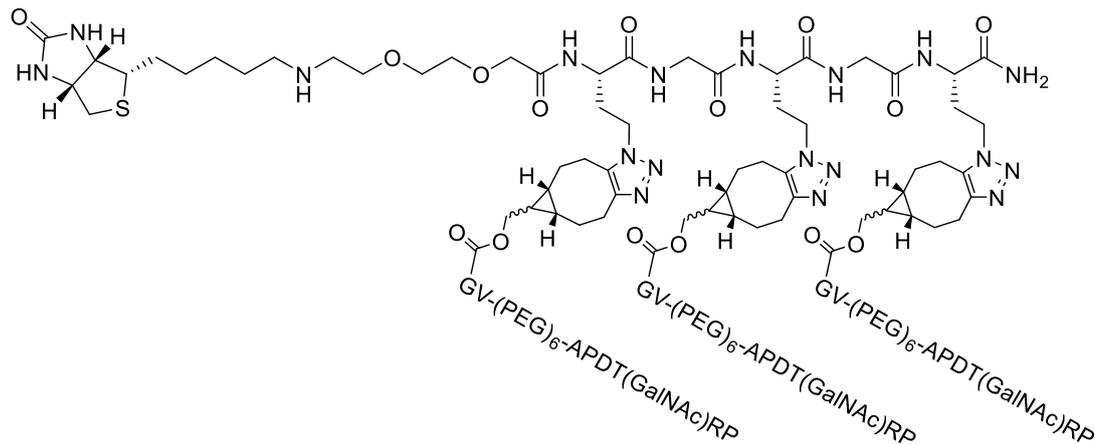
Glycopeptide **2.28** (20 mg, 14 μmol, 1.0 eq) and BCN-NHS (4.7 mg, 16 μmol, 1.2 eq) were dissolved in dry DMF (1 mL). TEA (5.7 μL, 41 μmol, 3.0 eq) was added and the mixture stirred overnight at room temperature. Reaction concentrated *in vacuo* and residue dissolved in ddH₂O (1 mL) and lyophilised to dryness. Residue purified by

size-exclusion using Biogel P2 column eluting in 20 mM ammonium formate. Yield: 15.7 mg, 62%.

HRMS [ES+] found $[M+2Na]^{2+}$ 920.4289, $C_{79}H_{126}N_{16}O_{31}Na_2$ requires 1840.8559;

LCMS [ES+] found $[M+2H]^{2+}$ 898.37, $C_{79}H_{128}N_{16}O_{31}$ requires 1796.89.

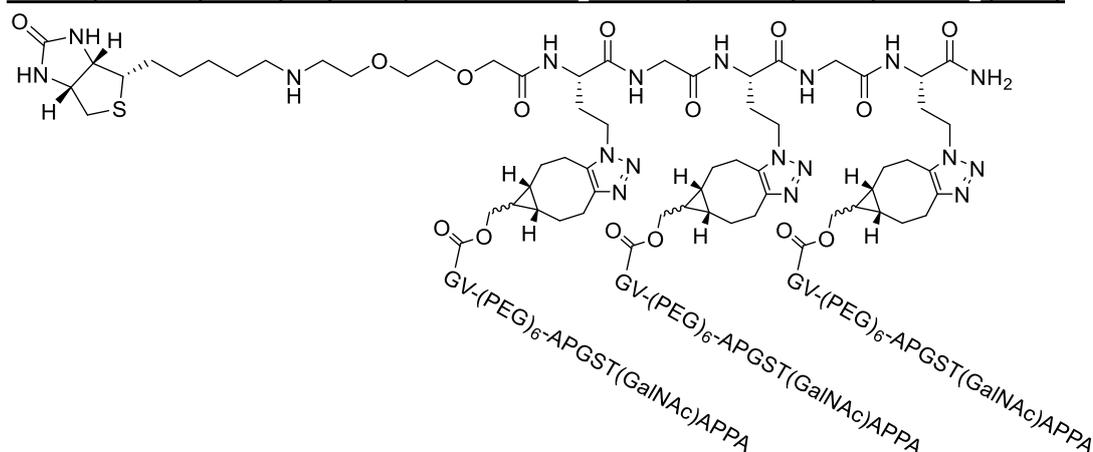
Biotin-PEG₃-Aha(BCN-GV-PEG₆-APDT(GalNAc)RP)-Gly-Aha(BCN-GV-PEG₆-APDT(GalNAc)RP)-Gly-Aha(BCN-GV-PEG₆-APDT(GalNAc)RP)-CONH₂ (3.11)



Biotin scaffold peptide **3.5** (0.9 mg, 1.0 μ mol, 1.0 eq) and glycopeptide **3.7** (5.0 mg, 3.3 μ mol, 3.3 eq) dissolved in ddH₂O (200 μ L) and mixture shaken overnight at room temperature. Reaction mixture purified by size-exclusion using Superdex S30 (10/300) column eluting in 20 mM ammonium formate. Yield: 4.5 mg, 83%.

HRMS [MALDI+] found $[M+H]^+$ 5306.4243, $C_{227}H_{370}N_{63}O_{81}S$ requires 5306.6485.

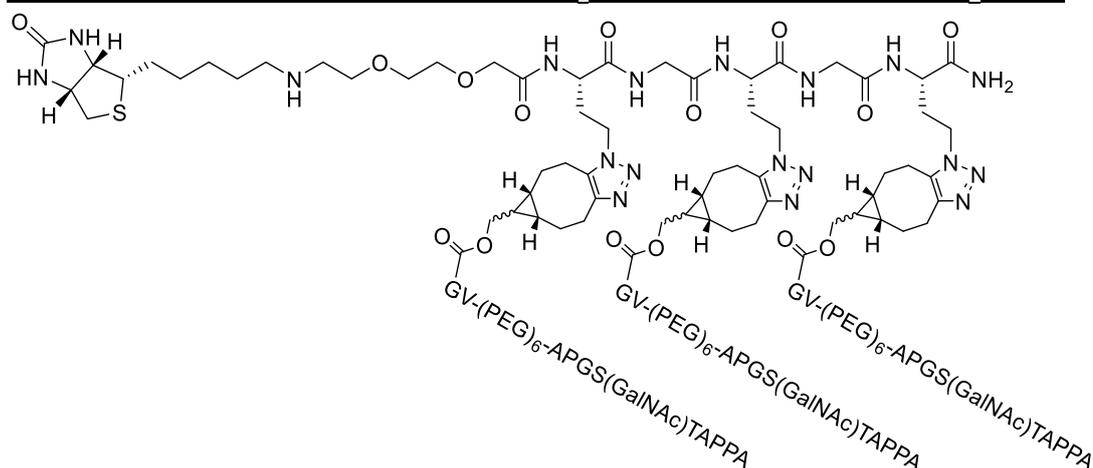
Biotin-PEG₃-Aha(BCN-GV-PEG₆-APGST(GalNAc)APPA)-Gly-Aha(BCN-GV-PEG₆-APGST(GalNAc)APPA)-Gly-Aha(BCN-GV-PEG₆-APGST(GalNAc)APPA)-CONH₂ (3.12)



Biotin scaffold peptide **3.5** (0.8 mg, 0.9 μmol , 1.0 eq) and glycopeptide **3.8** (5.1 mg, 3.2 μmol , 3.5 eq) dissolved in ddH₂O (200 μL) and mixture shaken overnight at room temperature. Reaction mixture purified by size-exclusion using Superdex S30 (10/300) column eluting in 20 mM ammonium formate. Yield: 4.6 mg, 90%.

HRMS [MALDI+] found $[\text{M}+\text{H}]^+$ 5643.3186, C₂₄₅H₃₉₄N₆₃O₈₇S requires 5642.8058.

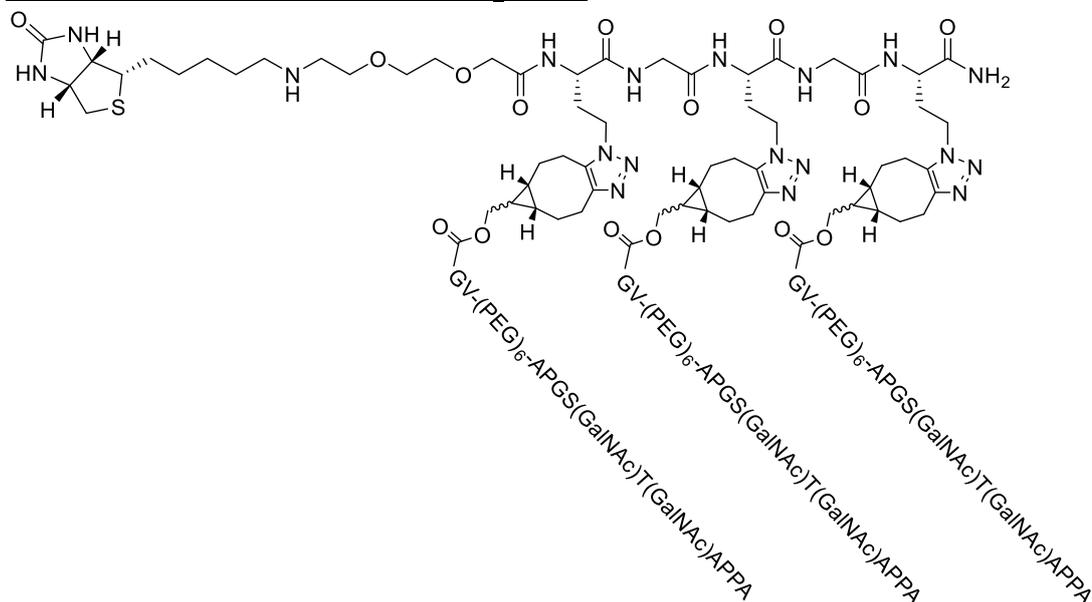
Biotin-PEG₃-Aha(BCN-GV-PEG₆-APGS(GalNAc)TAPPA)-Gly-Aha(BCN-GV-PEG₆-APGS(GalNAc)TAPPA)-Gly-Aha(BCN-GV-PEG₆-APGS(GalNAc)TAPPA)-CONH₂ (3.13)



Biotin scaffold peptide **3.5** (0.8 mg, 0.9 μmol , 1.0 eq) and glycopeptide **3.9** (5.1 mg, 3.2 μmol , 3.5 eq) dissolved in ddH₂O (200 μL) and mixture shaken overnight at room temperature. Reaction mixture purified by size-exclusion using Superdex S30 (10/300) column eluting in 20 mM ammonium formate. Yield: 4.7 mg, 91%.

HRMS [MALDI+] found $[\text{M}+\text{H}]^+$ 5643.0763, C₂₄₅H₃₉₄N₆₃O₈₇S requires 5642.8058.

Biotin-PEG₃-Aha(BCN-GV-PEG₆-APGS(GalNAc)T(GalNAc)APPA)-Gly-Aha(BCN-GV-PEG₆-APGS(GalNAc)T(GalNAc)APPA)-Gly-Aha(BCN-GV-PEG₆-APGS(GalNAc)T(GalNAc)APPA)-CONH₂ (3.14)



Biotin scaffold peptide **3.5** (0.7 mg, 0.8 μmol , 1.0 eq) and glycopeptide **3.10** (4.7 mg, 2.6 μmol , 3.3 eq) dissolved in ddH₂O (200 μL) and mixture shaken overnight at room temperature. Reaction mixture purified by size-exclusion using Superdex S30 (10/300) column eluting in 20 mM ammonium formate. Yield: 3.9 mg, 79%.

HRMS [MALDI+] found $[\text{M}+\text{H}]^+$ 6252.8049, C₂₆₉H₄₃₃N₆₆O₁₀₂S requires 6252.0439.

Chapter 8: Biochemical Experimental

8.1 General Methods and Equipment

Sterilisation of media and equipment was carried out in either a Prestige Medical bench top autoclave or a LTE Touchclave-R autoclave at 121 °C for 15 minutes. Thermo Electron Corporation SAFE 2010 Class II laminar flow cabinet or bench top bunsen burner were using to maintain a sterile environment. Bacterial cultures were incubated in a Kuhner ShakerX ISF1-X or Stuart Orbital incubator. LB-agar plates and enzymatic reactions were incubated in a Binder BD23 incubator.

Centrifugation was performed using either a Beckman Coulter™ Avanti™ JXN-30 centrifuge, Heraeus multifuge 3S-R centrifuge or Heraeus pico centrifuge. Gel filtration chromatography was achieved using a GE Pharmacia ÄKTA FPLC system or BioRad NGC FPLC system. Spectrophotometric reading were using a ThermoScientific Nanodrop 2000. SDS-PAGE was carried out using Bio-Rad mini protean 3 apparatus, and a Bio-Rad imager Gel Doc™ XR was used to visualise polyacrylamide gels. Proteins were concentrated either using 10k or 30k MWCO Amicon® Ultra-15 centrifugal filter device or Amicon® Stirred Ultrafiltration Cell (10k MWCO membrane). High resolution mass spectrometry of protein samples was performed using a Bruker Daltonics MicroTOF mass spectrometer. Protein samples were loaded at a concentration of 20-40 µM (made up with H₂O) into the instrument before being automatically diluted into 0.1% TFA/50% MeCN (v/v) in H₂O prior to analysis. Analytical grade reagents were supplied by commercial suppliers.

8.2 Media and Buffers

Buffers and media were made in-house unless otherwise stated using analytical grade reagents supplied by commercial suppliers. All common buffers and media were prepared with 18.2 MΩ water to the required volume. The pH of the solutions was adjusted using 5 M NaOH or 5 M HCl. Gel filtration buffers were filtered through a 0.22 µm membrane under reduced pressure. Media was sterilised by autoclave at 121 °C for 20 minutes.

8.2.1 Growth Media

Millers lysogeny broth (LB) medium: 1.0% (w/v) Tryptone, 0.5% (w/v) Yeast Extract, 1.0% (w/v) NaCl.

Millers lysogeny broth (LB) agar: 1.0% (w/v) Tryptone, 0.5% (w/v) Yeast Extract, 1.0% (w/v) NaCl, 1.5% (w/v) Agar.

2TY medium: 1.6% (w/v) Tryptone, 1.0% (w/v) Yeast Extract, 0.5% (w/v) NaCl.

TYP medium: 1.6% (w/v) Tryptone, 1.6% (w/v) Yeast Extract, 0.5% (w/v) NaCl, 0.25% (w/v) K₂HPO₄.

CAYE broth modified acc. to Evans (purchased from Merck): 39.5 g/L

Super Optimal Culture (SOC): 2.0% (w/v) Tryptone, 0.5% (w/v) Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄. Solution autoclaved at 120 °C for 20 mins before addition of 20 mL L⁻¹ 1 M glucose.

Auto induction medium: 1.0% (w/v) Tryptone, 0.5% (w/v) Yeast Extract, 4% (v/v) 25 × salts, 1% (v/v) 1000 × metals, 4.3 mM MgCl₂. Solution autoclaved at 120 °C for 20 mins before addition of 1 mL L⁻¹ 50% (w/v) glucose, 10 mL L⁻¹ 20% (w/v) lactose and 10 mL L⁻¹ 50% (v/v) glycerol by filter sterilisation (0.2 µm Sartorius Minisart).

25 × salts: 1.25 mM Na₂HPO₄, 1.25 mM KH₂PO₄, 2.5 mM NH₄Cl, 250 mM Na₂SO₄

1000 × metals: 50 mM FeCl₃·6H₂O, 20 mM CaCl₂·6H₂O, 100 mM MnCl₂·4H₂O, 100 mM ZnSO₄·7H₂O, 1.7 mM CoCl₂·6H₂O, 2 mM CuCl₂·2H₂O, 4.1 mM Na₂MoO₄·2H₂O, 2 mM Na₂SeO₃, 2 mM H₃BO₄, 2 mM NiSO₄·6H₂O, 50% (v/v) 0.1 M HCl.

8.2.2 General Buffers for Working with Proteins

Cholera Toxin B-subunit (CTB) Buffers (Ammonium Sulfate Precipitated)

Sodium Phosphate Buffered Saline: 50 mM NaH₂PO₄, 100 mM NaCl, pH 7.4.

CTB Ni-NTA wash buffer: 25 mM Imidazole, made up with Sodium Phosphate Buffered Saline.

CTB Ni-NTA elution buffer: 250 mM Imidazole, made up with Sodium Phosphate Buffered Saline.

CTB SEC buffer: Sodium Phosphate Buffered Saline 0.22 μm filtered.

Insoluble His-Tagged Protein Buffers

Phosphate lysis buffer: 50 mM NaH_2PO_4 , 300 mM NaCl, pH 8.0.

Denaturing buffer: 8 M Urea, 0.1 M NaH_2PO_4 , 0.01 M Tris, pH 8.0.

Ni-NTA urea wash buffer: 25 mM Imidazole, made up with denaturing buffer.

Ni-NTA urea elution buffer: 250 mM Imidazole, made up with denaturing buffer.

In-vitro folding buffer: 0.1 M Tris, 0.6 M Arginine Hydrochloride, 5 mM Cysteine, 0.5 mM Cystine, pH 8.5 (Degassed with Nitrogen).

Tris dialysis buffer: 50 mM Tris, 200 mM NaCl, pH 7.5.

Refolded protein SEC buffer: Tris dialysis buffer 0.22 μm filtered.

Sortase Buffers

Sortase 7M lysis buffer: 50 mM Tris, 150 mM NaCl, pH 7.2.

Sortase 7M Ni-NTA wash buffer: 25 mM Imidazole, made up Sortase 7M lysis buffer.

Sortase 7M Ni-NTA elution buffer: 250 mM Imidazole, made up Sortase 7M lysis buffer.

Sortase 7M SEC buffer: Sortase 7M lysis buffer, 0.22 μm filtered.

Verotoxin (VTB) & Oxime Ligation Buffers

Periplasmic extraction buffer: 30 mM Tris, 20% (w/v) Sucrose, 1 mM EDTA, pH 8.0.

Sodium Phosphate Buffered Saline: 50 mM NaH_2PO_4 , 100 mM NaCl, pH7.4

VTB SEC buffer: Sodium Phosphate Buffered Saline buffer, 0.22 μm filtered.

SDS-PAGE Buffers

5 \times SDS-PAGE loading buffer: 250 mM Tris, 10% (w/v) SDS, 1.0 M DTT, 50% (v/v) glycerol, 0.01% (w/v) bromophenol blue.

SDS-PAGE running buffer: 25 mM Tris-base, 192 mM Glycine, 0.1% (w/v) SDS.

Coomassie stain: Coomassie G250 (0.625% (w/v), 40% (v/v) methanol, 10% (v/v) acetic acid.

Coomassie destain: 40% (v/v) methanol, 10% (v/v) acetic acid.

8.2.3 General Buffers for Working with DNA

6 × DNA loading buffer, purple (no SDS, from NEB): 15% (w/v) Ficoll® -400, 60 mM EDTA, 140 mM Tris-HCl, 0.12% (w/v) Dye 1, 0.006% (w/v) Dye 2, pH 8.0.

TAE buffer: 40 mM Tris base, 20 mM acetic acid, 1 mM EDTA.

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

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

Buffer N3 (Qiagen): 4.2 M Guanidine HCl, 0.9 M potassium acetate, pH 4.8

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

8.3 Standard Protocols

8.3.1 Cell Growth and Protein Overexpression

8.3.1.1 Transformation of Chemically-Competent *E. coli* cells

10 μ L of *E. coli* competent cells and 1 μ L of plasmid were mixed together in a sterile Eppendorf tube on ice. The cells were incubated for 10 minutes to allow diffusion of the plasmid, followed by a heat shock for 45 seconds at 42 °C and another incubation on ice for 10 minutes to allow plasmid uptake. SOC media (1 mL) was added to the cells and they were incubated at 37°C with shaking (200 rpm) for 1 hour. A 20 μ L and concentrated cell aliquots were used to inoculate sterilised agar plates containing either ampicillin (100 μ g mL⁻¹) or kanamycin (50 μ g mL⁻¹). Concentrated cell aliquots were taken from the remaining cell culture which was centrifuged (13,000 \times g, 30 seconds). 800 μ L of supernatant was removed and the pelleted cells resuspended in the remaining media, a 20 μ L aliquot of this concentrate was used inoculate a sterilised agar selection plate. The plates were incubated at 37 °C overnight and removed in the morning to stop growth. Plates were stored at 4 °C until use, maximum of 1 week.

8.3.1.2 Preparation of Mini-culture

A single colony from the transformed *E. coli* cells grown on the sterilised agar plates overnight or glycerol stock stab was picked and incubated with shaking (200 rpm) in ~5 mL of LB media containing the necessary antibiotic (ampicillin ($100 \mu\text{g mL}^{-1}$) or kanamycin ($50 \mu\text{g mL}^{-1}$)) at 37°C overnight. This mini-culture was used to inoculate 1 L of overexpression media and/or glycerol stocks were made using a 1:1 mixture of the cells and 80% sterilised glycerol (in water). Glycerol stocks were left at room temperature for 15 minutes then flash frozen in liquid nitrogen and stored at -80°C for future use.

8.3.1.3 Overexpression in LB media

1 L of sterilised LB media containing the desired antibiotic (ampicillin ($100 \mu\text{g mL}^{-1}$) or kanamycin ($50 \mu\text{g mL}^{-1}$)) in a 2 L flask was inoculated with 2 mL of cells from the overnight culture. The culture was incubated at 37°C with shaking (200 rpm) and overexpression by the addition of IPTG (0.5 mM) when $\text{OD}_{600} = 0.6\text{--}0.8$ was reached. The cells were incubated overnight at a 25°C before being collected by centrifugation ($10,000 \times g$, 20 minutes, 4°C). Cells were either used immediately or stored at -80°C for future lysis and purification.

8.3.1.4 Overexpression in Auto-induction media

1 L of sterilised auto-induction media containing the desired antibiotic (ampicillin ($100 \mu\text{g mL}^{-1}$) or kanamycin ($50 \mu\text{g mL}^{-1}$)) in a 2 L flask was inoculated with 2 mL of cells from the overnight culture. The culture was incubated at 30°C with shaking (200 rpm) for 5 hours then the incubator temperature was reduced to 25°C . The cells were incubated overnight before being collected by centrifugation ($10,000 \times g$, 20 minutes, 4°C). Cells were either used immediately or stored at -80°C for future lysis and purification.

8.3.1.5 Overexpression of Proteins Extracted from Growth Media

1 L of sterilised LB media containing the desired antibiotic (ampicillin ($100 \mu\text{g mL}^{-1}$) or kanamycin ($50 \mu\text{g mL}^{-1}$)) in a 2 L flask was inoculated with 2 mL of cells from the overnight culture. The culture was incubated at 37°C with shaking (200 rpm) and overexpression by the addition of IPTG (0.5 mM) when $\text{OD}_{600} = 0.6\text{--}0.8$ was reached. The cells were incubated overnight at a 25°C before being collected by centrifugation ($10,000 \times g$, 20 minutes, 4°C). The *E. coli* cells are discarded and media retained. Solid ammonium sulfate was added and dissolved to a final concentration of 57% (*w/v*) to precipitate the protein of interest. The saturated solution was stirred overnight at 4°C . The solution was centrifuged ($17,600 \times g$, 1 hour, 4°C), and the supernatant discarded. The protein pellet was resuspended in Sodium Phosphate Buffered Saline (10 mL per 1 L LB media) and centrifuged ($17,600 \times g$, 10 minutes, 4°C) to remove insoluble material and to allow the suspension to be filtered through a $0.8 \mu\text{m}$ Sartorius Minisart filter. The filtered suspension was purified by Ni-NTA affinity column.

8.3.2 Protein Purification

8.3.2.1 Lysis of *E. coli* containing soluble proteins

E. coli cells were resuspended in the corresponding lysis buffer at 4°C . The cell mixture was lysed mechanically using a Constant Systems cell disruptor (20 kpsi). The cell debris was pelleted by centrifugation ($30,000 \times g$, 45 mins, 4°C) and the cell lysate was decanted off and treated with DNase. Purification of the clarified lysate performed by Ni-NTA affinity chromatography.

8.3.2.2 Purification of His₆-tagged proteins using Ni-NTA chromatography

Ni-NTA agarose resin (Qiagen, 10 mL) was transferred to a Bio-Rad Econo-Pac® chromatography column and equilibrated with 5 column volumes of lysis buffer. The solution containing the His-tagged protein was applied to the column, and flow-through reloaded. The resin was washed with 5 column volumes of Ni-NTA wash buffer. The protein was eluted in 10 mL fractions using 5 column volumes of Ni-NTA

elution buffer. The protein eluates were monitored using Bradford reagent. The protein size, expression level and purity were assessed by SDS-PAGE.

8.3.2.3 Regeneration of Ni-NTA column

After being used for protein purification, the Ni-NTA agarose resin was regenerated for future use. Firstly, the resin was washed with 0.1% (w/v) sodium dodecyl sulphate (SDS, 3 column volumes), then 0.1 M NaOH (3 column volumes). The Ni²⁺ was removed by chelation with 10 mM ethylenediaminetetraacetic acid (EDTA, 3 column volumes), the resin was washed with water (10 column volumes) and regenerated with 500 mM NiSO₄ (2 column volumes). A final wash with H₂O was done to remove any excess Ni²⁺ and the column was flushed and stored in 20% EtOH.

8.3.2.4 Size-exclusion chromatography

For large scale purification of proteins a Superdex® 75 or 200 prep grade column (HiLoad® 16/60 or 26/60, GE Healthcare Life Sciences) was attached to a ÄKTApurifier FPLC system (GE Healthcare Life Sciences) or Bio-Rad NGC FPLC system. The column was stored in 20% ethanol and prior to use initially equilibrated with 1.5 column volumes of water (0.22 µm filtered) at a flow rate of 0.5 mL min⁻¹, followed by 1.5 column volumes of the appropriate ÄKTA buffer at a flow rate of 1 mL min⁻¹.

Concentrated protein samples following affinity column purification were concentrated to 500 µL to 5 mL. Once an appropriate volume for the injection loop in use was reached, the sample was manually injected into the injection loop of the FPLC system. The automated system pumps the injection loop with buffer (1.5 × the loop volume) to apply the sample to the resin. The column was run isocratically in the appropriate SEC buffer. Once the void volume of the column had been reached, 5 mL fractions were collected and the protein eluate was monitored using absorption at 280 nm.

8.3.2.5 Purification and refolding of insoluble proteins

Rapid Dilution Method

The bacterial cell pellet was resuspended on ice in phosphate lysis buffer (10 mL per 1 L media). The cell suspension was mechanically lysed using a constant systems cell disruptor and treated with DNase. Cell debris pelleted by centrifugation ($30,000 \times g$, 40 minutes, 4 °C). The supernatant was discarded and the cell debris was resuspended and homogenised on ice in denaturing buffer (15 mL per 1 L media). Suspension was left at 4 °C overnight then sonicated on ice (output 50%, 50% time pulses, 1 minute), and the remaining debris was pelleted by centrifugation (40 minutes at $30,000 \times g$, 4 °C). The cell debris was discarded, whilst the supernatant retained and sonicated on ice (output 60%, 50% time pulses, 1 minute) to reduce the viscosity.

Ni-NTA agarose resin (Qiagen, 3 mL per 1 L LB media) was transferred to a Bio-Rad Econo-Pac® chromatography column and equilibrated with 5 column volumes of denaturing buffer. The supernatant containing the denatured His-tagged protein was applied to the column, and flow-through reloaded. The resin was washed with 5 column volumes of Ni-NTA urea wash buffer. The protein was eluted in 5 mL fractions using 6 column volumes of Ni-NTA urea elution buffer. The protein elution's were monitored using Bradford reagent. The protein size, expression level and purity were assessed by SDS-PAGE. Fractions containing unfolded protein were combined and concentrated to ~2 mL (per 1 L media) using an Amicon® Stirred Ultrafiltration Cell fitted with a Ultracel 10K MWCO ultrafiltration membrane.

0.5 M TCEP was added to the concentrated fractions to a final concentration of 50 mM and the mixture incubated at room temperature for 2 hours. TCEP was removed using a PD-10 desalting column (GE Scientific) equilibrated with denaturing buffer and eluted following the manufacturers recommended protocol. The reduced PD-10 eluate was concentrated as much as possible with a 10k MWCO Amicon® Ultra-15 centrifugal filter device to at least 10 mg/mL and less than 1 mL (per 1 L media).

The *in vitro* folding buffer (100 mL per 1 L media) was freshly prepared, degassed with nitrogen and equilibrated to 4 °C. With gently stirring, 200 µL aliquots of the concentrated reduced unfolded protein were added every minute. The bottle was sealed inbetween additions to minimize the introduction of air. If the solution began to turn cloudy a fresh batch of *in vitro* folding buffer was prepared. The mixture was

incubated at 4 °C with gentle stirring for 3-5 days to allow the protein to refold. After refolding at 4 °C, any precipitate was removed first by centrifugation ($4600 \times g$) and then by filtration through a 0.8 μm Sartorius Minisart filter. The folded protein was concentrated to ~10 mL (per 1 L media) and dialysed overnight against Tris buffered saline (pH 7.2) at 4 °C. Once dialysed into Tris buffer, the protein was further concentrated using 10k MWCO Amicon® Ultra-15 centrifugal filter device and purified by size exclusion chromatography.

Sequential Dialysis Method

The bacterial cell pellet was resuspended on ice in phosphate lysis buffer (10 mL per 1 L media). The cell suspension was mechanically lysed using a constant systems cell disruptor and treated with DNase. Cell debris pelleted by centrifugation ($30,000 \times g$, 40 minutes, 4 °C). The supernatant was discarded and the cell debris was resuspended and homogenised on ice in denaturing buffer (15 mL per 1 L media). Suspension was left at 4 °C overnight then sonicated on ice (output 50%, 50% time pulses, 1 minute), and the remaining debris was pelleted by centrifugation (40 minutes at $30,000 \times g$, 4 °C). The cell debris was discarded, whilst the supernatant retained and sonicated on ice (output 60%, 50% time pulses, 1 minute) to reduce the viscosity.

Ni-NTA agarose resin (Qiagen, 3 mL per 1 L LB media) was transferred to a Bio-Rad Econo-Pac® chromatography column and equilibrated with 5 column volumes of denaturing buffer. The supernatant containing the denatured His-tagged protein was applied to the column, and flow-through reloaded. The resin was washed with 5 column volumes of Ni-NTA urea wash buffer. The protein was eluted in 5 mL fractions using 6 column volumes of Ni-NTA urea elution buffer. The protein eluates were monitored using Bradford reagent. The protein size, expression level and purity were assessed by SDS-PAGE. Fractions containing unfolded protein were combined and placed in dialysis tubing. Protein dialysed against denaturing buffer overnight. The concentration of urea was slowly reduced by dialysing into Tris Buffer containing 4 M, 2 M, 1 M urea and finally Tris Buffer containing no urea. Each dialysis step was performed at 4 °C overnight. After refolding any precipitate was removed first by centrifugation ($4600 \times g$) and then by filtration through a 0.8 μm Sartorius Minisart

filter. The folded protein was concentrated using 10k MWCO Amicon® Ultra-15 centrifugal filter device and purified by size exclusion chromatography.

8.3.2.6 Expression & Purification of Verotoxin (VTB)

1 mL of VTB starter culture (in LB) was used to inoculate 400 mL of sterilised CAYE media containing ampicillin ($100 \mu\text{g mL}^{-1}$) in a 2 L baffled flask. The culture was incubated at 37°C with shaking (200 rpm) and overexpression induced through the addition of IPTG (0.5 mM) when $\text{OD}_{600} = 0.6-0.8$ was reached. The cells were incubated overnight at 37°C before being collected by centrifugation ($10,000 \times g$, 20 minutes, 4°C).

The bacterial cell pellet was resuspended on ice in periplasmic extraction buffer (2 mL per 400 mL media) and incubated with shaking at 37°C (200 rpm) for 10 minutes. The cell pellet was recovered by centrifugation ($4500 \times g$, 20 mins, 4°C). The supernatant was decanted, and the cell pellet resuspended in ice cold $15 \text{ M}\Omega \text{ H}_2\text{O}$ (2 mL per 300 mL media). The suspension was shaken on ice for 10 minutes, before the cell debris was harvested by centrifugation ($4500 \times g$, 20 mins, 4°C). The supernatant was retained and $20 \times \text{NaH}_2\text{PO}_4$ buffered saline (0.1 mL) was added to buffer the solution. The supernatant was purified by size-exclusion chromatography using Superdex S75 16/60 column on a Biorad FPLC system.

8.3.2.7 Concentration of protein samples

Protein containing fractions from Ni-NTA column were combined and concentrated to an appropriate volume either using the 10k or 30k MWCO Amicon® Ultra-15 centrifugal filter device ($4000 \times g$) or Amicon® Stirred Ultrafiltration Cell. A P200 Gilson pipette was used to resuspend the protein every 15 minutes, if required, as the protein collects on the centrifugal filter.

8.3.2.8 Dialysis

Dialysis was frequently used to remove small molecules such as imidazole from protein samples.

An appropriate length of Spectrum Laboratories Spectra/Por® molecularporous membrane tubing (3500 MWCO) was placed in buffer and left for 10 mins to hydrate. One end of the tubing was folded over and sealed with a tubing clip. To protein solution to be dialysed was placed within the tubing, and the tubing was sealed with a second clip ensuring no air bubbles were present. The tubing was placed within 2 to 4 L of buffer, with a magnetic follower, and left stirring at 4 °C overnight.

8.3.3 Protein Analysis

8.3.3.1 SDS Polyacrylamide Gel Electrophoresis

To separate the protein on the basis of their mass, 20 µL of the protein samples were added to 5 µL of 5 × SDS-PAGE loading buffer. If required, the samples were boiled for 5 mins at 95 °C to denature the protein. DTT reduces any potential disulfide bonds within the protein and SDS coats the protein in a negative charge.

The resolving gel was prepared using the recipe in Table 8-1 (adding TEMED last), the solution was poured immediately into a fixed Bio-Rad 1.00 mm gel cast, and left to set with a layer isopropanol was added to the surface of resolving gel to produce a flat surface. Once the gel had set the isopropanol was removed and top of the gel washed with water. Once the water was removed, the stacking gel was added followed by a 1.00 mm Bio-Rad comb to create wells. Once the gel had set, the comb was carefully removed and gel transferred to a Bio-Rad electrophoresis tank. The inside of the tank, where the gel was fixed, was filled with SDS-PAGE running buffer to ensure no leaked and finally the surrounding section was half filled with running buffer. 10 µL aliquots of each sample was loaded into the sample wells and gel electrophoresis was run at 180 V in SDS loading buffer for 45 minutes, or until the protein bands can be seen to have reached the bottom of the gel.

Table 8-1: Materials and quantities for preparing SDS-PAGE gels

	15% Resolving Gel	12% Resolving Gel	Stacking Gel
H₂O	3.50 mL	4.25 mL	3.32 mL
1.5 M Tris pH 8.8	2.53 mL	2.53 mL	-
0.5 M Tris pH 6.8	-	-	94.5 μ L
40% (w/v) Acrylamide	3.75 mL	3.00 mL	625 μ L
10% (w/v) SDS	100 μ L	100 μ L	50 μ L
10% (w/v) APS	150 μ L	150 μ L	50 μ L
TEMED	10 μ L	10 μ L	5 μ L

8.3.3.2 Gel imaging

Once electrophoresis had finished, the gels were removed from the casts and stained with either InstantBlue (Expedeon) or Coomassie Blue G-250. Gels were stained with InstantBlue (Expedeon) for 1.5 hrs in the dark on a rocker at room temperature. Once bands became visible, the InstantBlue was decanted and the gel washed with water for 30 minutes. Gels stained in Coomassie Blue G-250 were stained for 12 hours on a rocker at room temperature. The stain was decanted and replaced with Coomassie destain and incubated for a further 3 hours with rocking, then washed with water. Gels were photographed using Bio-Rad Gel Doc XR system.

8.3.3.3 Protein Quantification

Protein concentration (M) were measured using a Thermo Scientific NanoDrop 2000 spectrophotometer. The instrument was blanked with the flow through of either the stirred ultrafiltration cell, centrifugal filter unit or dialysis buffer. Absorbance at 280 nm of a 5 μ L sample of protein was measured and using the molar extinction coefficient (ϵ) provided by ProtParam sequence analysis (ExpASY; <http://web.expasy.org/protparam/>), the protein concentration was calculated. Protein concentration was calculated using the following equation.

$$[\text{Protein}] = \frac{A_{280}}{\epsilon}$$

Equation 8-1: Beer-Lambert Law: Abs = absorbance, ϵ = molar extinction coefficient ($M^{-1} \text{ cm}^{-1}$), c = concentration (M) and l = pathlength (cm)

Table 8-2: Extinction coefficients for proteins at $\lambda = 280 \text{ nm}$

Protein	Extinction Coefficient, ϵ_{280}
Sortase7M-His ₆	14440 $M^{-1} \text{ cm}^{-1}$
CBD-Sortase7M	35535 $M^{-1} \text{ cm}^{-1}$
CTB-LPETGA	11585 $M^{-1} \text{ cm}^{-1}$
CTB-LPETGAS-His ₆	11585 $M^{-1} \text{ cm}^{-1}$
CTB(W88E)	6085 $M^{-1} \text{ cm}^{-1}$
VTB	8605 $M^{-1} \text{ cm}^{-1}$
SAP-LPETGA-His ₆ (E167Q)	45505 $M^{-1} \text{ cm}^{-1}$
CRP-LPETGA-His ₆	45045 $M^{-1} \text{ cm}^{-1}$

8.3.4 Protein Modification

8.3.4.1 Typical sortase A 7M mediated labelling (analytical scale)

For a total volume of 100 μL , the protein to be labelled (200 μM) and peptide/glycopeptide (3000 μM , 15 eq.) were mixed and the reaction mixture was made up to volume with the appropriate buffer before the addition of sortase A 7M (30 μM , 15 mol%). The reaction mixture was incubated for 2 hours at 25 $^{\circ}\text{C}$ and product formation followed by mass spectrometry or SDS-PAGE. For time point studies by SDS-PAGE, samples (5 μL) were taken and diluted with the appropriate buffer (15 μL) and mixed with 5 \times SDS loading buffer (5 μL). Samples were heated to 95 $^{\circ}\text{C}$ for 5 minutes before being frozen and stored at -20°C for future analysis by SDS-PAGE.

8.3.4.2 Typical His₆-CBD-Srt7M mediated labelling (preparative scale)

For a total volume of 200 μ L, the protein to be labelled (200 μ M) and peptide/glycopeptide (3000 μ M, 15 eq.) were mixed and the reaction mixture was made up to volume with the appropriate buffer before the addition of His₆-CBD-Srt7M (30 μ M, 15 mol%). The reaction mixture was incubated for 2 hours at 25°C. Once the reaction has reached completion by mass spectrometry, 200 μ L of chitin resin (50% slurry in TBS (pH 7.5)) was added and the mixture incubated on ice for 15 mins. The solution was filtered through a centrifugal cellulose acetate filter (4000 \times g, 1 minute). The excess labelling substrate was removed by either diafiltration, dialysis or size-exclusion chromatography (Superdex 75 10/300, Tris Buffered Saline (pH7.5)).

8.3.4.3 General procedure for Thr/Ser Oxidation and Oxime Ligation

Protein to be labelled was buffer exchanged into sodium phosphate buffer (pH 7.4); it is very important the buffer does not contain potassium. Performing reactions in the presence of K⁺ ions will result in oxidation failing to reach completion.

The protein to be labelled (500 μ L, 200 μ M) in sodium phosphate buffer was mixed with L-methionine (4 μ L of 250 mM stock, final concentration 2.0 mM, 10 eq) and NaIO₄ (2 μ L of 250 mM stock, final concentration 1.0 mM, 5 eq). Reaction was followed by ES-MS and typically complete within 10 minutes. The protein was purified using G-25 mini-trap desalting column equilibrated and eluted with sodium phosphate buffer (pH 6.8).

To the oxidised protein from G25-minitrapp desalting column (1 mL, ~100 μ M) was added aniline (18 μ L, final concentration 200 mM, 1000 equiv.), and the oxyamine substrate (8 μ L of 250 mM stock in DMSO, final concentration 2.0 mM, 10 equiv.). The mixture was mixed by vortex and incubated at 37 °C for 8-16 hrs. The oxime labelled protein was purified using PD-10 desalting column equilibrated and eluted with sodium phosphate buffer (pH 6.8).

8.3.4.4 General procedure for SPAAC ligation

To the BCN labelled protein in sodium phosphate buffer (50 μ L, 138 μ M) was added the azido functionalised glycan (0.67 μ L of 20 mM stock, final concentration 276 mM, 2 eq). The reaction was incubated at 16 °C until reaction was observed to be complete by ES-MS (typically 4-8 hours). The SPAAC labelled protein was purified using G-25 mini-trap desalting column equilibrated and eluted with sodium phosphate buffer (pH 6.8).

8.3.5 DNA Manipulation, Purification and Analysis

8.3.5.1 Small scale DNA isolation and sequence analysis (mini-prep)

XL10 *E.coli* cells, harbouring the desired plasmid, were collected under sterile conditions from a glycerol stock or LB-agar plate, and added to a falcon tube with 5 mL of LB media which had been inoculated with the appropriate antibiotic, ampicillin (100 μ g/mL) or kanamycin (50 μ g/mL). The culture was incubated at 37 °C with shaking (200 rpm) for 18hrs. The bacterial cell pellet was isolated from the mini cultures by centrifugation (4500 \times g, 15 minutes) at room temperature and the DNA was extracted using a QIAprep Spin Mini Prep Kit following the manufacturers recommended protocol. All centrifugation steps were carried out at 17,000 \times g in a microfuge at room temperature.

The cell pellet was resuspended in buffer P1 (250 μ L) and mixed thoroughly with buffer P2 (250 μ L) until the solution had turned blue (maximum of 4 minutes). Buffer N3 (350 μ L) was added and the Eppendorf tube inverted until the solution turned colourless. Insoluble debris was pelleted by centrifugation for 10 minutes. The supernatant was applied to a QIAprep Spin column and centrifuged for 1 minute to bind the DNA to the column. The flow through was discarded and column washed with buffer PE (750 μ L) and centrifuged for 2 \times 1 minute. The flow through was discarded and spin filter transferred to a sterile Eppendorf tube. The DNA sample was eluted with H₂O (50 μ L) and centrifuged for 1 minute.

Sequencing was performed by GATC Biotech (Eurofins) and the results were analysed using BioEdit sequence alignment editor or SnapGene® viewer version 5..

8.3.5.2 Medium scale DNA isolation

XL10 *E. coli* cells, harbouring the desired plasmid, were collected under sterile conditions from a glycerol stock or LB-agar plate, and added to a falcon tube with 35 mL of LB media which had been inoculated with the appropriate antibiotic, ampicillin (100 µg/mL) or kanamycin (50 µg/mL). The culture was incubated at 37 °C with shaking (200 rpm) for 18 hrs. The bacterial cell pellet was isolated from the mini cultures by centrifugation (4500 × g, 15 minutes) at room temperature and the DNA was extracted using a QIAGEN Plasmid Plus Midi Kit following the manufacturers high-yield protocol. All centrifugation steps were carried out at 17,000 × g in a microfuge at room temperature for 1 minute.

The cell pellet was resuspended in buffer P1 (4 mL) before buffer P2 was added (4 mL) and the mixture was incubated for 3 minutes at room temperature. Buffer S3 (4 mL) was added to the lysate, and mixture transferred to a QIA filter cartridge and incubated at room temperature for 10 mins. Precipitate was removed by filtration and buffer BB (2 mL) was added to the cleared lysate. The mixture was passed through the QUIAGEN plasmid plus midi spin column using a vacuum manifold, binding the DNA to the silica. The column was washed (by centrifugation) with buffer ETR (700 µL), followed by buffer PE (700 µL). Residual buffer was removed by centrifugation for 1 minute and spin filter transferred to a sterile Eppendorf tube. The DNA sample was eluted with H₂O (200 µL) and centrifuged for 1 minute.

8.3.5.3 DNA quantification

DNA concentrations were measured using a Thermo Scientific NanoDrop 2000 spectrophotometer. The instrument was blanked with H₂O and absorbance at 260 nm of a 2 µL sample was measured. Concentration was estimated using Equation 8-2.

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

Equation 8-2: DNA concentration calculation

8.3.5.4 Site-directed mutagenesis

Protocol used was based on the NEB Q5 hot start protocol. Forward and reverse primers were designed with the help of NEBaseChanger. The PCR reaction mixture (50 μL) was made up of 20 ng of the template DNA plasmid, 500 nM of the forward and reverse primers and Q5 Hot Start High-Fidelity 2 \times Master Mix (25 μL). The final concentration of Q5 Master Mix contains 200 μM dNTPs, 2 mM Mg^{2+} and 1 unit of Q5 hot start DNA polymerase in a proprietary buffer. The thermocycler program used is described in Table 8-3. Once completed, a DpnI digestion was performed by adding 1 μL DpnI (20 U μL^{-1}) and incubating at 37 $^{\circ}\text{C}$ for 1 hour, followed by heat inactivation at 80 $^{\circ}\text{C}$ for 20 mins. Following this 16.5 μL of the PCR reaction mixture was taken and combined with T4 PNK (0.5 μL , 10 U μL^{-1}) in T4 ligase buffer (2 μL) and incubated at 37 $^{\circ}\text{C}$ for 1 hour. To this T4 DNA ligase (1 μL , 400 U μL^{-1}) was added and incubated at room temperature for 30 mins before being transformed into fresh *E. coli* XL10-gold ultra-competent cells.

Table 8-3: Q5 Hot Start polymerase PCR program for site-directed mutagenesis

Step		Temperature	Time
Initial Denature		98 $^{\circ}\text{C}$	30 secs
30 Cycles	Denature	98 $^{\circ}\text{C}$	10 secs
	Anneal	*50–72 $^{\circ}\text{C}$	30 secs
	Elongation	72 $^{\circ}\text{C}$	2.5–3.0 mins
Final Elongation		72 $^{\circ}\text{C}$	2.0 mins
Hold		4–10 $^{\circ}\text{C}$	--

* Annealing temperature predicted using the NEB Tm Calculator

8.3.5.5 Agarose gel electrophoresis

Agarose gel electrophoresis was used to analyse DNA following reactions. A 1% agarose gel was prepared by dissolving agarose (0.4 g) in TAE buffer (40 mL) using a microwave for 1 minute. The solution was allowed to cool for a few minutes before SYBR®Safe (2 μL of 10,000 \times in DMSO) was added and the mixture was poured into a mould with a comb and left to set. The DNA samples to be loaded were prepared by mixing the DNA sample (5 μL) with 6 \times gel loading Dye, purple, no SDS (1 μL).

Once set, the gel was placed into the tank and filled with TAE buffer. The samples were loaded into the well and gel was run at 80 V for 35 minutes. The gels were imaged using BioRad Gel Doc XR system.

8.3.6 ELLA protocols

8.3.6.1 Preparation of ganglioside GM1-coated microtiter plates

96-well flat bottomed, high binding, polystyrene microtiter plates (Grenier 655077) were coated with ganglioside GM1 (100 μ l, 1.3 M in methanol) and the solvent was allowed to evaporate at room temperature. The plate was washed with PBS (3 x 200 μ l) to remove any remaining unbound GM1, and any remaining binding sites on the well were blocked with BSA by incubating with a PBS solution containing 1% (*w/v*) bovine serum albumin (BSA, 100 μ l) for 30 minutes at 37 °C. The wells were then washed again with PBS (3 x 200 μ l).

8.3.6.2 Preparation FSL-Gb3 coated microtiter plates

96-well flat bottomed, high binding, polystyrene microtiter plates (Grenier 655077) were coated with FSL-Gb3 (100 μ l, 2 μ g/mL, 1.4 μ M) and incubated at room temperature for 3 hours. The plate was washed with PBS (3 x 200 μ l) to remove any remaining unbound FSL-Gb3, and any remaining binding sites on the well were blocked with BSA by incubating with a PBS solution containing 1% (*w/v*) bovine serum albumin (BSA, 100 μ l) for 30 minutes at 37 °C. The wells were then washed again with PBS (3 x 200 μ l).

8.3.6.3 ELLA Assay

Inhibitor samples were prepared performing a 2-fold serial dilution in PBS containing 0.1% (*w/v*) BSA, 0.05% (*v/v*) Tween-20 (PBS-T) in V-bottomed propylene 96-well plates (Grenier 651201) using a program developed on the Hamilton microlab STAR liquid handling system analysing each sample in triplicate. The samples were mixed with CTB-HRP (final conc: 0.5 ng/mL) or VTB-HRP (final conc: 10 ng/mL) in the same buffer. The mixture of inhibitor and toxin were incubated for 2 hours at room

temperature before 100 μ l was transferred to the GM1 or Gb3 coated well by the Hamilton microlab STAR liquid handling system. The limits of detection were analysed by control samples containing just CTB-HRP or VTB-HRP and no inhibitor, and a sample containing just buffer, used to obtain the maximum and minimum optical density values for CTB-HRP binding to the GM1 coated wells or VTB-HRP binding to the Gb3 coated wells. The inhibitor toxin mixture was incubated for 30 minutes at room temperature before the plate was washed with PBS-T (3 x 200 μ l) to remove any unbound CTB/VTB-HRP-inhibitor complex. After washing the plate, a solution of Amplex red® (100 μ l, 5 μ M Amplex Red®, 5 μ M H₂O₂ in PBS) was added and fluorescence was measured at 25 °C for 30 minutes in a Perkin Elmer EnVision plate reader (excitation 531 nM; emission 595 nm). Initial rates were calculated if the data plots deviated from linearity.

8.3.6.4 Data processing

All samples were analysed in triplicate. The error of each sample was calculated as standard error (Equation 8-3) where n equals the samples size, x is the observed initial rate value for each sample and \bar{x} is the mean rate value for each sample.

$$\text{standard error} = \frac{\sqrt{\frac{\sum(x - \bar{x})^2}{(n - 1)}}}{\sqrt{n}}$$

Equation 8-3: Standard error in initial rate obtained for each sample

The fluorescence was then converted to percentage binding by comparison with the maximum and minimum values obtained from the positive and negative controls and errors were propagated accordingly. This data was then plotted against log (inhibitor concentration) for each sample in origin and the curve fitting was performed using the non-linear curve fit, using the Origin logistic function (Equation 8-4), where A_1 is the curve's maximum, A_2 is the curve's minimum, x_0 is equal to the IC₅₀, x is the log (inhibitor concentration), and p is the Hill slope parameter.

$$y = \frac{A_1 - A_2}{1 + (x/x_0)^p} + A_2$$

Equation 8-4: Equation for logistic curve fitting.

Chapter 9: References

1. Martinez-Seara Monne, H., Danne, R., Róg, T., Ilpo, V. & Gurtovenko, A. Structure of Glycocalyx. *Biophys. J.* **104**, 251a (2013).
2. Möckl, L. The Emerging Role of the Mammalian Glycocalyx in Functional Membrane Organization and Immune System Regulation. *Front. Cell Dev. Biol.* **8**, (2020).
3. Akintayo, A. & Stanley, P. Roles for Golgi Glycans in Oogenesis and Spermatogenesis. *Front. Cell Dev. Biol.* **7**, (2019).
4. Zhang, S. *et al.* Selection of tumor antigens as targets for immune attack using immunohistochemistry: I. Focus on gangliosides. *Int. J. Cancer* **73**, 42–49 (1997).
5. Zhang, S. *et al.* Selection of tumor antigens as targets for immune attack using immunohistochemistry: II. Blood group-related antigens. *Int. J. Cancer* **73**, 50–56 (1997).
6. Julien, S., Videira, P. A. & Delannoy, P. Sialyl-Tn in Cancer: (How) Did We Miss the Target? *Biomolecules* **2**, 435–466 (2012).
7. Heimburg-Molinaro, J. *et al.* Cancer vaccines and carbohydrate epitopes. *Vaccine* **29**, 8802–8826 (2011).
8. Zhou, Z., Liao, G., Mandal, S. S., Suryawanshi, S. & Guo, Z. A fully synthetic self-adjuvanting globo H-Based vaccine elicited strong T cell-mediated antitumor immunity. *Chem. Sci.* **6**, 7112–7121 (2015).
9. Danishefsky, S. J., Shue, Y.-K., Chang, M. N. & Wong, C.-H. Development of Globo-H Cancer Vaccine. *Acc. Chem. Res.* **48**, 643–652 (2015).
10. Krenzel, U. & Bousquet, P. A. Molecular Recognition of Gangliosides and Their Potential for Cancer Immunotherapies. *Front. Immunol.* **5**, (2014).
11. Hakomori, S. Tumor malignancy defined by aberrant glycosylation and sphingo(glyco)lipid metabolism. *Cancer Res.* **56**, 5309–5318 (1996).
12. Galonić, D. P. & Gin, D. Y. Chemical glycosylation in the synthesis of glycoconjugate antitumour vaccines. *Nature* **446**, 1000–1007 (2007).
13. Hakomori, S. Aberrant Glycosylation In Tumors And Tumor-Associated Carbohydrate Antigens. in *Advances in Cancer Research Vol. 52* 257–331 (1989). doi:10.1016/S0065-230X(08)60215-8.
14. Helling, F. *et al.* GM2-KLH conjugate vaccine: increased immunogenicity in melanoma patients after administration with immunological adjuvant QS-21. *Cancer Res.* **55**, 2783–2788 (1995).
15. Helling, F. *et al.* GD3 vaccines for melanoma: superior immunogenicity of keyhole limpet hemocyanin conjugate vaccines. *Cancer Res.* **54**, 197–203 (1994).
16. O’Cearbhaill, R. *et al.* A Phase I Study of Unimolecular Pentavalent (Globo-H-GM2-sTn-TF-Tn) Immunization of Patients with Epithelial Ovarian, Fallopian Tube, or Peritoneal Cancer in First Remission. *Cancers (Basel)*. **8**, 46 (2016).
17. Linden, S. K., Sutton, P., Karlsson, N. G., Korolik, V. & McGuckin, M. A. Mucins in the mucosal barrier to infection. *Mucosal Immunol.* **1**, 183–197 (2008).
18. Pinzón Martín, S., Seeberger, P. H. & Varón Silva, D. Mucins and Pathogenic Mucin-Like Molecules Are Immunomodulators During Infection and Targets for Diagnostics and Vaccines. *Front. Chem.* **7**, (2019).

19. Dhanisha, S. S., Guruvayoorappan, C., Drishya, S. & Abeesh, P. Mucins: Structural diversity, biosynthesis, its role in pathogenesis and as possible therapeutic targets. *Crit. Rev. Oncol. Hematol.* **122**, 98–122 (2018).
20. van Putten, J. P. M. & Strijbis, K. Transmembrane Mucins: Signaling Receptors at the Intersection of Inflammation and Cancer. *J. Innate Immun.* **9**, 281–299 (2017).
21. Hollingsworth, M. A. & Swanson, B. J. Mucins in cancer: protection and control of the cell surface. *Nat. Rev. Cancer* **4**, 45–60 (2004).
22. Johansson, M. E. V. & Hansson, G. C. The Mucins. in *Encyclopedia of Immunobiology* 381–388 (Elsevier, 2016). doi:10.1016/B978-0-12-374279-7.02019-1.
23. Swallow, D. M. *et al.* The hypervariable gene locus PUM, which codes for the tumour associated epithelial mucins, is located on chromosome 1, within the region 1q21-24. *Ann. Hum. Genet.* **51**, 289–94 (1987).
24. Brayman, M., Thathiah, A. & Carson, D. D. MUC1: a multifunctional cell surface component of reproductive tissue epithelia. *Reprod. Biol. Endocrinol.* **2**, 4 (2004).
25. Hattrup, C. L. & Gendler, S. J. Structure and Function of the Cell Surface (Tethered) Mucins. *Annu. Rev. Physiol.* **70**, 431–457 (2008).
26. Nath, S. & Mukherjee, P. MUC1: a multifaceted oncoprotein with a key role in cancer progression. *Trends Mol. Med.* **20**, 332–342 (2014).
27. Baldus, S. E., Engelmann, K. & Hanisch, F.-G. MUC1 and the MUCs: a family of human mucins with impact in cancer biology. *Crit. Rev. Clin. Lab. Sci.* **41**, 189–231 (2004).
28. Dhar, P. & McAuley, J. The Role of the Cell Surface Mucin MUC1 as a Barrier to Infection and Regulator of Inflammation. *Front. Cell. Infect. Microbiol.* **9**, e117 (2019).
29. Ng, G. Z. & Sutton, P. The MUC1 mucin specifically inhibits activation of the NLRP3 inflammasome. *Genes Immun.* **17**, 203–206 (2016).
30. Guang, W. *et al.* Muc1 Cell Surface Mucin Attenuates Epithelial Inflammation in Response to a Common Mucosal Pathogen. *J. Biol. Chem.* **285**, 20547–20557 (2010).
31. Pan, H., Colville, M. J., Supekar, N. T., Azadi, P. & Paszek, M. J. Sequence-Specific Mucins for Glycocalyx Engineering. *ACS Synth. Biol.* **8**, 2315–2326 (2019).
32. Polefka, T. G., Garrick, R. A., Redwood, W. R., Swislocki, N. I. & Chinard, F. P. Solute-excluded volumes near the Novikoff cell surface. *Am. J. Physiol. Physiol.* **247**, C350–C356 (1984).
33. Gendler, S. J. MUC1, The Renaissance Molecule. *J. Mammary Gland Biol. Neoplasia* **6**, 339–353 (2001).
34. Zhao, Q. *et al.* MUC1 extracellular domain confers resistance of epithelial cancer cells to anoikis. *Cell Death Dis.* **5**, e1438–e1438 (2014).
35. Wang, Y. *et al.* Cosmc is an essential chaperone for correct protein O-glycosylation. *Proc. Natl. Acad. Sci.* **107**, 9228–9233 (2010).
36. Hanson, R. & Hollingsworth, M. Functional Consequences of Differential O-glycosylation of MUC1, MUC4, and MUC16 (Downstream Effects on Signaling). *Biomolecules* **6**, 34 (2016).
37. Ju, T., Aryal, R. P., Stowell, C. J. & Cummings, R. D. Regulation of protein O-glycosylation by the endoplasmic reticulum-localized molecular chaperone Cosmc. *J. Cell Biol.* **182**, 531–542 (2008).

38. Aryal, R. P., Ju, T. & Cummings, R. D. The Endoplasmic Reticulum Chaperone Cosmc Directly Promotes in Vitro Folding of T-synthase. *J. Biol. Chem.* **285**, 2456–2462 (2010).
39. Beatson, R. *et al.* The Breast Cancer-Associated Glycoforms of MUC1, MUC1-Tn and sialyl-Tn, Are Expressed in COSMC Wild-Type Cells and Bind the C-Type Lectin MGL. *PLoS One* **10**, e0125994 (2015).
40. Bierhuizen, M. F. & Fukuda, M. Expression cloning of a cDNA encoding UDP-GlcNAc:Gal beta 1-3-GalNAc-R (GlcNAc to GalNAc) beta 1-6GlcNAc transferase by gene transfer into CHO cells expressing polyoma large tumor antigen. *Proc. Natl. Acad. Sci.* **89**, 9326–9330 (1992).
41. Schwientek, T. *et al.* Control of O -Glycan Branch Formation. *J. Biol. Chem.* **275**, 11106–11113 (2000).
42. Tarp, M. A. & Clausen, H. Mucin-type O-glycosylation and its potential use in drug and vaccine development. *Biochim. Biophys. Acta - Gen. Subj.* **1780**, 546–563 (2008).
43. Tran, D. T. & Ten Hagen, K. G. Mucin-type O -Glycosylation during Development. *J. Biol. Chem.* **288**, 6921–6929 (2013).
44. Iwai, T. *et al.* Molecular cloning and characterization of a novel UDP-GlcNAc:GalNAc-peptide beta1,3-N-acetylglucosaminyltransferase (beta 3Gn-T6), an enzyme synthesizing the core 3 structure of O-glycans. *J. Biol. Chem.* **277**, 12802–9 (2002).
45. Krishn, S. R. *et al.* Mucins and associated glycan signatures in colon adenoma–carcinoma sequence: Prospective pathological implication(s) for early diagnosis of colon cancer. *Cancer Lett.* **374**, 304–314 (2016).
46. Awaya, H., Takeshima, Y., Yamasaki, M. & Inai, K. Expression of MUC1, MUC2, MUC5AC, and MUC6 in Atypical Adenomatous Hyperplasia, Bronchioloalveolar Carcinoma, Adenocarcinoma With Mixed Subtypes, and Mucinous Bronchioloalveolar Carcinoma of the Lung. *Am. J. Clin. Pathol.* **121**, 644–653 (2004).
47. Singh, A. P. *et al.* Aberrant expression of transmembrane mucins, MUC1 and MUC4, in human prostate carcinomas. *Prostate* **66**, 421–429 (2006).
48. Hakomori, S. Aberrant glycosylation in cancer cell membranes as focused on glycolipids: overview and perspectives. *Cancer Res.* **45**, 2405–14 (1985).
49. Springer, G. T and Tn, general carcinoma autoantigens. *Science (80-.)*. **224**, 1198–1206 (1984).
50. Radhakrishnan, P. *et al.* Immature truncated O-glycophenotype of cancer directly induces oncogenic features. *Proc. Natl. Acad. Sci.* **111**, E4066–E4075 (2014).
51. Mi, R. *et al.* Epigenetic Silencing of the Chaperone Cosmc in Human Leukocytes Expressing Tn Antigen. *J. Biol. Chem.* **287**, 41523–41533 (2012).
52. Gill, D. J., Chia, J., Senewiratne, J. & Bard, F. Regulation of O-glycosylation through Golgi-to-ER relocation of initiation enzymes. *J. Cell Biol.* **189**, 843–858 (2010).
53. Gill, D. J., Clausen, H. & Bard, F. Location, location, location: new insights into O-GalNAc protein glycosylation. *Trends Cell Biol.* **21**, 149–158 (2011).
54. Springer, G. F. Tn epitope (N-acetyl-d-galactosamine α -O -serine/threonine) density in primary breast carcinoma: A functional predictor of aggressiveness. *Mol. Immunol.* **26**, 1–5 (1989).
55. Seeberger, P. H. & Werz, D. B. Synthesis and medical applications of oligosaccharides. *Nature* **446**, 1046–1051 (2007).

56. Ada, G. & Isaacs, D. Carbohydrate—protein conjugate vaccines. *Clin. Microbiol. Infect.* **9**, 79–85 (2003).
57. Wang, L.-X. Toward oligosaccharide- and glycopeptide-based HIV vaccines. *Curr. Opin. Drug Discov. Devel.* **9**, 194–206 (2006).
58. Liu, X. *et al.* Enhancement of the Immunogenicity of Synthetic Carbohydrates by Conjugation to Virosomes: A Leishmaniasis Vaccine Candidate. *ACS Chem. Biol.* **1**, 161–164 (2006).
59. Schofield, L., Hewitt, M. C., Evans, K., Siomos, M.-A. & Seeberger, P. H. Synthetic GPI as a candidate anti-toxic vaccine in a model of malaria. *Nature* **418**, 785–789 (2002).
60. Colombo, C., Pitirollo, O. & Lay, L. Recent Advances in the Synthesis of Glycoconjugates for Vaccine Development. *Molecules* **23**, 1712 (2018).
61. Boltje, T. J., Buskas, T. & Boons, G.-J. Opportunities and challenges in synthetic oligosaccharide and glycoconjugate research. *Nat. Chem.* **1**, 611–622 (2009).
62. Slovin, S. F., Keding, S. J. & Ragupathi, G. Carbohydrate vaccines as immunotherapy for cancer. *Immunol. Cell Biol.* **83**, 418–428 (2005).
63. Cipolla, L., Peri, F. & Airoidi, C. Glycoconjugates in Cancer Therapy. *Anticancer. Agents Med. Chem.* **8**, 92–121 (2008).
64. Buskas, T., Thompson, P. & Boons, G.-J. Immunotherapy for cancer: synthetic carbohydrate-based vaccines. *Chem. Commun.* 5335 (2009) doi:10.1039/b908664c.
65. Polonskaya, Z., Savage, P. B., Finn, M. G. & Teyton, L. High-affinity anti-glycan antibodies: challenges and strategies. *Curr. Opin. Immunol.* **59**, 65–71 (2019).
66. Gilewski, T. *et al.* Immunization of metastatic breast cancer patients with a fully synthetic globo H conjugate: A phase I trial. *Proc. Natl. Acad. Sci.* **98**, 3270–3275 (2001).
67. Yin, Z. *et al.* Boosting Immunity to Small Tumor-Associated Carbohydrates with Bacteriophage Q β Capsids. *ACS Chem. Biol.* **8**, 1253–1262 (2013).
68. Imberty, A. & Varrot, A. Microbial recognition of human cell surface glycoconjugates. *Curr. Opin. Struct. Biol.* **18**, 567–576 (2008).
69. D’Angelo, G., Capasso, S., Sticco, L. & Russo, D. Glycosphingolipids: synthesis and functions. *FEBS J.* **280**, 6338–6353 (2013).
70. Merrill, A. H. Sphingolipid and Glycosphingolipid Metabolic Pathways in the Era of Sphingolipidomics. *Chem. Rev.* **111**, 6387–6422 (2011).
71. Simons, K. & Gerl, M. J. Revitalizing membrane rafts: new tools and insights. *Nat. Rev. Mol. Cell Biol.* **11**, 688–699 (2010).
72. Ferguson, M. A. J. & Williams, A. F. CELL-SURFACE ANCHORING OF PROTEINS VIA GLYCOSYL-PHOSPHATIDYLINOSITOL STRUCTURES. *Annu. Rev. Biochem.* **57**, 285–320 (1988).
73. Levental, I., Grzybek, M. & Simons, K. Greasing Their Way: Lipid Modifications Determine Protein Association with Membrane Rafts. *Biochemistry* **49**, 6305–6316 (2010).
74. Lehmann, F., Tiralongo, E. & Tiralongo, J. Sialic acid-specific lectins: occurrence, specificity and function. *Cell. Mol. Life Sci.* **63**, 1331–1354 (2006).
75. MacKenzie, C. R., Hirama, T., Lee, K. K., Altman, E. & Young, N. M. Quantitative Analysis of Bacterial Toxin Affinity and Specificity for Glycolipid Receptors by Surface Plasmon Resonance. *J. Biol. Chem.* **272**, 5533–5538 (1997).

76. Lencer, W. I., Hirst, T. R. & Holmes, R. K. Membrane traffic and the cellular uptake of cholera toxin. *Biochim. Biophys. Acta - Mol. Cell Res.* **1450**, 177–190 (1999).
77. Merritt, E. A. & Hol, W. G. AB5 toxins. *Curr. Opin. Struct. Biol.* **5**, 165–171 (1995).
78. World Health Organization (WHO). *Cholera Annual Report 2019*. (2019).
79. Ali, M., Nelson, A. R., Lopez, A. L. & Sack, D. A. Updated Global Burden of Cholera in Endemic Countries. *PLoS Negl. Trop. Dis.* **9**, e0003832 (2015).
80. Holmgren, J. *et al.* Mucosal adjuvants and anti-infection and anti-immunopathology vaccines based on cholera toxin, cholera toxin B subunit and CpG DNA. *Immunol. Lett.* **97**, 181–188 (2005).
81. Stratmann, T. Cholera Toxin Subunit B as Adjuvant—An Accelerator in Protective Immunity and a Break in Autoimmunity. *Vaccines* **3**, 579–596 (2015).
82. Turnbull, W. B., Precious, B. L. & Homans, S. W. Dissecting the Cholera Toxin–Ganglioside GM1 Interaction by Isothermal Titration Calorimetry. *J. Am. Chem. Soc.* **126**, 1047–1054 (2004).
83. Branson, T. R. & Turnbull, W. B. Bacterial toxininhibitors based on multivalent scaffolds. *Chem. Soc. Rev.* **42**, 4613–4622 (2013).
84. Römer, W. *et al.* Shiga toxin induces tubular membrane invaginations for its uptake into cells. *Nature* **450**, 670–675 (2007).
85. Johannes, L., Parton, R. G., Bassereau, P. & Mayor, S. Building endocytic pits without clathrin. *Nat. Rev. Mol. Cell Biol.* **16**, 311–321 (2015).
86. Kumar, V. & Turnbull, W. B. Carbohydrate inhibitors of cholera toxin. *Beilstein J. Org. Chem.* **14**, 484–498 (2018).
87. Thorpe, C. M. Shiga Toxin--Producing Escherichia coli Infection. *Clin. Infect. Dis.* **38**, 1298–1303 (2004).
88. Lindberg, A. A. *et al.* Identification of the carbohydrate receptor for Shiga toxin produced by Shigella dysenteriae type 1. *J. Biol. Chem.* **262**, 1779–85 (1987).
89. Ling, H. *et al.* Structure of the Shiga-like Toxin I B-Pentamer Complexed with an Analogue of Its Receptor Gb 3 ,. *Biochemistry* **37**, 1777–1788 (1998).
90. Siukstaite, L., Imberty, A. & Römer, W. Structural Diversities of Lectins Binding to the Glycosphingolipid Gb3. *Front. Mol. Biosci.* **8**, (2021).
91. St. Hilaire, P. M., Boyd, M. K. & Toone, E. J. Interaction of the Shiga-like Toxin Type 1 B-Subunit with Its Carbohydrate Receptor. *Biochemistry* **33**, 14452–14463 (1994).
92. Turnbull, W. B. Biophysical Studies on the Interactions of Bacterial Toxins. in *Glyco-bioinformatics* (eds. Hick, M. G. & Kettner, C.) 55–68 (Beilstein Institute, 2012).
93. Römer, W. *et al.* Actin Dynamics Drive Membrane Reorganization and Scission in Clathrin-Independent Endocytosis. *Cell* **140**, 540–553 (2010).
94. Sandvig, K., Bergan, J., Dyve, A.-B., Skotland, T. & Torgersen, M. L. Endocytosis and retrograde transport of Shiga toxin. *Toxicon* **56**, 1181–1185 (2010).
95. Sandvig, K. *et al.* Retrograde transport of endocytosed Shiga toxin to the endoplasmic reticulum. *Nature* **358**, 510–512 (1992).
96. Lee, Y. C. & Lee, R. T. Carbohydrate-Protein Interactions: Basis of Glycobiology. *Acc. Chem. Res.* **28**, 321–327 (1995).
97. Dam, T. K. & Brewer, C. F. Thermodynamic Studies of Lectin–Carbohydrate

- Interactions by Isothermal Titration Calorimetry. *Chem. Rev.* **102**, 387–430 (2002).
98. Safina, G. Application of surface plasmon resonance for the detection of carbohydrates, glycoconjugates, and measurement of the carbohydrate-specific interactions: A comparison with conventional analytical techniques. A critical review. *Anal. Chim. Acta* **712**, 9–29 (2012).
 99. Sacchettini, J. C., Baum, L. G. & Brewer, C. F. Multivalent Protein–Carbohydrate Interactions. A New Paradigm for Supramolecular Assembly and Signal Transduction. *Biochemistry* **40**, 3009–3015 (2001).
 100. Kiessling, L. L., Young, T., Gruber, T. D. & Mortell, K. H. Multivalency in Protein–Carbohydrate Recognition. in *Glycoscience* 2483–2523 (Springer Berlin Heidelberg, 2008). doi:10.1007/978-3-540-30429-6_64.
 101. Lee, R. T. & Lee, Y. C. Affinity enhancement by multivalent lectin–carbohydrate interaction. *Glycoconj. J.* **17**, 543–551 (2000).
 102. Jencks, W. P. On the attribution and additivity of binding energies. *Proc. Natl. Acad. Sci.* **78**, 4046–4050 (1981).
 103. Page, M. I. & Jencks, W. P. Entropic Contributions to Rate Accelerations in Enzymic and Intramolecular Reactions and the Chelate Effect. *Proc. Natl. Acad. Sci.* **68**, 1678–1683 (1971).
 104. Kitov, P. I. *et al.* Shiga-like toxins are neutralized by tailored multivalent carbohydrate ligands. *Nature* **403**, 669–672 (2000).
 105. Mattarella, M., Garcia-Hartjes, J., Wennekes, T., Zuilhof, H. & Siegel, J. S. Nanomolar cholera toxininhibitors based on symmetrical pentavalent ganglioside GM1os- sym -corannulenes. *Org. Biomol. Chem.* **11**, 4333–4339 (2013).
 106. Liu, J. *et al.* Protein Heterodimerization through Ligand-Bridged Multivalent Pre-organization: Enhancing Ligand Binding toward Both Protein Targets. *J. Am. Chem. Soc.* **127**, 2044–2045 (2005).
 107. Kitov, P. I. *et al.* An Entropically Efficient Supramolecular Inhibition Strategy for Shiga Toxins. *Angew. Chemie Int. Ed.* **47**, 672–676 (2008).
 108. Branson, T. R. *et al.* A Protein-Based Pentavalent Inhibitor of the Cholera Toxin B-Subunit. *Angew. Chemie Int. Ed.* **53**, 8323–8327 (2014).
 109. Haksar, D. *et al.* Fighting Shigella by Blocking Its Disease-Causing Toxin. *J. Med. Chem.* **64**, 6059–6069 (2021).
 110. Wang, L. & Schultz, P. G. Expanding the genetic code. *Angew. Chemie Int. Ed.* **44**, 1–11 (2005).
 111. Wang, L. Expanding the Genetic Code. *Science (80-.)*. **302**, 584–585 (2003).
 112. Wang, L. & Schultz, P. G. Expanding the genetic code. *Chem. Commun.* 1–11 (2002) doi:10.1039/b108185n.
 113. Schmidt, M. J., Borbas, J., Drescher, M. & Summerer, D. A Genetically Encoded Spin Label for Electron Paramagnetic Resonance Distance Measurements. *J. Am. Chem. Soc.* **136**, 1238–1241 (2014).
 114. Fleissner, M. R. *et al.* Site-directed spin labeling of a genetically encoded unnatural amino acid. *Proc. Natl. Acad. Sci.* **106**, 21637–21642 (2009).
 115. Liu, C. C. & Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **79**, 413–444 (2010).
 116. Ellman, J., Mendel, D., Anthony-Cahill, S., Noren, C. J. & Schultz, P. G. Biosynthetic method for introducing unnatural amino acids site-specifically into proteins. in *Methods in Enzymology Vol. 202* 301–336 (1991). doi:10.1016/0076-6879(91)02017-4.

117. Sletten, E. M. & Bertozzi, C. R. Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angew. Chemie Int. Ed.* **48**, 6974–6998 (2009).
118. Brabham, R. & Fascione, M. A. Pyrrolysine Amber Stop-Codon Suppression: Development and Applications. *ChemBioChem* **18**, 1973–1983 (2017).
119. Blight, S. K. *et al.* Direct charging of tRNACUA with pyrrolysine in vitro and in vivo. *Nature* **431**, 333–335 (2004).
120. Bain, J. D., Diala, E. S., Glabe, C. G., Dix, T. A. & Chamberlin, A. R. Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide. *J. Am. Chem. Soc.* **111**, 8013–8014 (1989).
121. Xu, R. *et al.* Site-Specific Incorporation of the Mucin-Type N - Acetylgalactosamine- α - O -threonine into Protein in Escherichia coli. *J. Am. Chem. Soc.* **131**, 15654–15655 (2004).
122. Matsubara, T., Iijima, K., Watanabe, T., Hohsaka, T. & Sato, T. Incorporation of glycosylated amino acid into protein by an in vitro translation system. *Bioorg. Med. Chem. Lett.* **23**, 5634–5636 (2013).
123. Xu, R. *et al.* Site-Specific Incorporation of the Mucin-Type N - Acetylgalactosamine- α - O -threonine into Protein in Escherichia coli. *J. Am. Chem. Soc.* **131**, 15654–15655 (2009).
124. Check Hayden, E. ‘Killer application’ for protein synthesis is retracted. *Nature* **462**, 707–707 (2009).
125. Antonczak, A. K., Simova, Z. & Tippmann, E. M. A Critical Examination of Escherichia coli Esterase Activity. *J. Biol. Chem.* **284**, 28795–28800 (2009).
126. Singh-Blom, A., Hughes, R. A. & Ellington, A. D. Residue-Specific Incorporation of Unnatural Amino Acids into Proteins In Vitro and In Vivo. in 93–114 (2013). doi:10.1007/978-1-62703-293-3_7.
127. van Hest, J. C. M., Kiick, K. L. & Tirrell, D. A. Efficient Incorporation of Unsaturated Methionine Analogues into Proteins in Vivo. *J. Am. Chem. Soc.* **122**, 1282–1288 (2000).
128. Tang, Y. *et al.* Fluorinated Coiled-Coil Proteins Prepared In Vivo Display Enhanced Thermal and Chemical Stability. *Angew. Chemie Int. Ed.* **40**, 1494–1496 (2001).
129. Saleh, A. M., Wilding, K. M., Calve, S., Bundy, B. C. & Kinzer-Ursem, T. L. Non-canonical amino acid labeling in proteomics and biotechnology. *J. Biol. Eng.* **13**, 43 (2019).
130. Haigh, J. L. *et al.* A versatile cholera toxin conjugate for neuronal targeting and tracing. *Chem. Commun.* **56**, 6098–6101 (2020).
131. Chalker, J. M. & Davis, B. G. Chemical mutagenesis: selective post-expression interconversion of protein amino acid residues. *Curr. Opin. Chem. Biol.* **14**, 781–789 (2010).
132. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chemie Int. Ed.* **40**, 2004–2021 (2001).
133. Sletten, E. M. & Bertozzi, C. R. From Mechanism to Mouse: A Tale of Two Bioorthogonal Reactions. *Acc. Chem. Res.* **44**, 666–676 (2011).
134. Hu, Q.-Y. *et al.* Synthesis of a well-defined glycoconjugate vaccine by a tyrosine-selective conjugation strategy. *Chem. Sci.* **4**, 3827 (2013).
135. Lee, D. J., Cameron, A. J., Wright, T. H., Harris, P. W. R. & Brimble, M. A. A synthetic approach to ‘click’ neoglycoprotein analogues of EPO employing one-pot native chemical ligation and CuAAC chemistry. *Chem. Sci.* **10**, 815–

- 828 (2019).
136. van Kasteren, S. I. *et al.* Expanding the diversity of chemical protein modification allows post-translational mimicry. *Nature* **446**, 1105–1109 (2007).
 137. Boutureira, O. *et al.* Fluoroglycoproteins: ready chemical site-selective incorporation of fluorosugars into proteins. *Chem. Commun.* **46**, 8142 (2010).
 138. Li, L. & Zhang, Z. Development and Applications of the Copper-Catalyzed Azide-Alkyne Cycloaddition (CuAAC) as a Bioorthogonal Reaction. *Molecules* **21**, 1393 (2016).
 139. Hong, V., Presolski, S. I., Ma, C. & Finn, M. â. G. Analysis and Optimization of Copper-Catalyzed Azide-Alkyne Cycloaddition for Bioconjugation. *Angew. Chemie Int. Ed.* **48**, 9879–9883 (2009).
 140. Jung, G. Lantibiotics—Ribosomally Synthesized Biologically Active Polypeptides containing Sulfide Bridges and α,β -Didehydroamino Acids. *Angew. Chemie Int. Ed. English* **30**, 1051–1068 (1991).
 141. Strumeyer, D. H., White, W. N. & Koshland, D. E. ROLE OF SERINE IN CHYMOTRYPSIN ACTION. CONVERSION OF THE ACTIVE SERINE TO DEHYDROALANINE. *Proc. Natl. Acad. Sci.* **50**, 931–935 (1963).
 142. Neet, K. E. & Koshland, D. E. The conversion of serine at the active site of subtilisin to cysteine: a ‘chemical mutation’. *Proc. Natl. Acad. Sci.* **56**, 1606–1611 (1966).
 143. Wang, J., Schiller, S. M. & Schultz, P. G. A Biosynthetic Route to Dehydroalanine-Containing Proteins. *Angew. Chemie Int. Ed.* **46**, 6849–6851 (2007).
 144. Fernández-González, M. *et al.* Site-selective chemoenzymatic construction of synthetic glycoproteins using endoglycosidases. *Chem. Sci.* **1**, 709 (2010).
 145. Chalker, J. M. *et al.* Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chem. Sci.* **2**, 1666 (2011).
 146. Dadová, J., Galan, S. R. & Davis, B. G. Synthesis of modified proteins via functionalization of dehydroalanine. *Curr. Opin. Chem. Biol.* **46**, 71–81 (2018).
 147. Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science (80-.)*. **354**, (2016).
 148. Kent, S. B. H. Total chemical synthesis of proteins. *Chem. Soc. Rev.* **38**, 338–351 (2009).
 149. Dawson, P., Muir, T., Clark-Lewis, I. & Kent, S. Synthesis of proteins by native chemical ligation. *Science (80-.)*. **266**, 776–779 (1994).
 150. Agouridas, V. *et al.* Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations. *Chem. Rev.* **119**, 7328–7443 (2019).
 151. Johnson, E. C. B. & Kent, S. B. H. Insights into the Mechanism and Catalysis of the Native Chemical Ligation Reaction. *J. Am. Chem. Soc.* **128**, 6640–6646 (2006).
 152. Wang, P. *et al.* At Last: Erythropoietin as a Single Glycoform. *Angew. Chemie Int. Ed.* **51**, 11576–11584 (2012).
 153. Murakami, M., Okamoto, R., Izumi, M. & Kajihara, Y. Chemical Synthesis of an Erythropoietin Glycoform Containing a Complex-type Disialyloligosaccharide. *Angew. Chemie Int. Ed.* **51**, 3567–3572 (2012).
 154. Macmillan, D. & Bertozzi, C. R. Modular Assembly of Glycoproteins: Towards the Synthesis of GlyCAM-1 by Using Expressed Protein Ligation.

- Angew. Chemie Int. Ed.* **43**, 1355–1359 (2004).
155. Anraku, Y., Mizutani, R. & Satow, Y. Protein Splicing: Its Discovery and Structural Insight into Novel Chemical Mechanisms. *IUBMB Life* **57**, 563–574 (2005).
 156. Muralidharan, V. & Muir, T. W. Protein ligation: an enabling technology for the biophysical analysis of proteins. *Nat. Methods* **3**, 429–438 (2006).
 157. Muir, T. W. Semisynthesis of Proteins by Expressed Protein Ligation. *Annu. Rev. Biochem.* **72**, 249–289 (2003).
 158. Muir, T. W., Sondhi, D. & Cole, P. A. Expressed protein ligation: A general method for protein engineering. *Proc. Natl. Acad. Sci.* **95**, 6705–6710 (1998).
 159. Roller, R. F. *et al.* Semisynthesis of Functional Glycosylphosphatidylinositol-Anchored Proteins. *Angew. Chemie Int. Ed.* **59**, 12035–12040 (2020).
 160. Ozawa, T., Nogami, S., Sato, M., Ohya, Y. & Umezawa, Y. A Fluorescent Indicator for Detecting Protein–Protein Interactions in Vivo Based on Protein Splicing. *Anal. Chem.* **72**, 5151–5157 (2000).
 161. Gariat, I. & Muir, T. W. Protein Semi-Synthesis in Living Cells. *J. Am. Chem. Soc.* **125**, 7180–7181 (2003).
 162. Ulrich, S., Boturyn, D., Marra, A., Renaudet, O. & Dumy, P. Oxime Ligation: A Chemoselective Click-Type Reaction for Accessing Multifunctional Biomolecular Constructs. *Chem. - A Eur. J.* **20**, 34–41 (2014).
 163. Brabham, R. L. *et al.* Rapid sodium periodate cleavage of an unnatural amino acid enables unmasking of a highly reactive α -oxo aldehyde for protein bioconjugation. *Org. Biomol. Chem.* **18**, 4000–4003 (2020).
 164. Spears, R. J. & Fascione, M. A. Site-selective incorporation and ligation of protein aldehydes. *Org. Biomol. Chem.* **14**, 7622–7638 (2016).
 165. Kalia, J. & Raines, R. T. Hydrolytic Stability of Hydrazones and Oximes. *Angew. Chemie Int. Ed.* **47**, 7523–7526 (2008).
 166. Chen, J., Zeng, W., Offord, R. & Rose, K. A Novel Method for the Rational Construction of Well-Defined Immunogens: The Use of Oximation To Conjugate Cholera Toxin B Subunit to a Peptide–Polyoxime Complex. *Bioconjug. Chem.* **14**, 614–618 (2003).
 167. Dirksen, A., Hackeng, T. M. & Dawson, P. E. Nucleophilic Catalysis of Oxime Ligation. *Angew. Chemie Int. Ed.* **45**, 7581–7584 (2006).
 168. Williamson, D. J., Fascione, M. A., Webb, M. E. & Turnbull, W. B. Efficient N-Terminal Labeling of Proteins by Use of Sortase. *Angew. Chemie Int. Ed.* **51**, 9377–9380 (2012).
 169. Williamson, D. J., Webb, M. E. & Turnbull, W. B. Depsipeptide substrates for sortase-mediated N-terminal protein ligation. *Nat. Protoc.* **9**, 253–262 (2014).
 170. Jeong, H.-J., Abhiraman, G. C., Story, C. M., Ingram, J. R. & Dougan, S. K. Generation of Ca²⁺-independent sortase A mutants with enhanced activity for protein and cell surface labeling. *PLoS One* **12**, e0189068 (2017).
 171. Zong, Y., Bice, T. W., Ton-That, H., Schneewind, O. & Narayana, S. V. L. Crystal Structures of *Staphylococcus aureus* Sortase A and Its Substrate Complex. *J. Biol. Chem.* **279**, 31383–31389 (2004).
 172. Guimaraes, C. P. *et al.* Site-specific C-terminal and internal loop labeling of proteins using sortase-mediated reactions. *Nat. Protoc.* **8**, 1787–1799 (2013).
 173. Chen, I., Dorr, B. M. & Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci.* **108**, 11399–11404 (2011).
 174. Hirakawa, H., Ishikawa, S. & Nagamune, T. Design of Ca²⁺-independent

- Staphylococcus aureus sortase A mutants. *Biotechnol. Bioeng.* **109**, 2955–2961 (2012).
175. Naik, M. T. *et al.* Staphylococcus aureus Sortase A Transpeptidase. *J. Biol. Chem.* **281**, 1817–1826 (2006).
 176. Popp, M. W.-L., Antos, J. M. & Ploegh, H. L. Site-Specific Protein Labeling via Sortase-Mediated Transpeptidation. *Curr. Protoc. Protein Sci.* **Chapter 15**, 15.3.1-15.3.9 (2009).
 177. Witte, M. D. *et al.* Site-specific protein modification using immobilized sortase in batch and continuous-flow systems. *Nat. Protoc.* **10**, 508–516 (2015).
 178. Heck, T., Pham, P.-H., Yerlikaya, A., Thöny-Meyer, L. & Richter, M. Sortase A catalyzed reaction pathways: a comparative study with six SrtA variants. *Catal. Sci. Technol.* **4**, 2946–2956 (2014).
 179. Liu, F., Luo, E. Y., Flora, D. B. & Mezo, A. R. Irreversible Sortase A-Mediated Ligation Driven by Diketopiperazine Formation. *J. Org. Chem.* **79**, 487–492 (2014).
 180. Yamamura, Y., Hirakawa, H., Yamaguchi, S. & Nagamune, T. Enhancement of sortase A-mediated protein ligation by inducing a β -hairpin structure around the ligation site. *Chem. Commun.* **47**, 4742 (2011).
 181. Suree, N. *et al.* The Structure of the Staphylococcus aureus Sortase-Substrate Complex Reveals How the Universally Conserved LP X TG Sorting Signal Is Recognized. *J. Biol. Chem.* **284**, 24465–24477 (2009).
 182. David Row, R., Roark, T. J., Philip, M. C., Perkins, L. L. & Antos, J. M. Enhancing the efficiency of sortase-mediated ligations through nickel-peptide complex formation. *Chem. Commun.* **51**, 12548–12551 (2015).
 183. Reed, S. A., Brzovic, D. A., Takasaki, S. S., Boyko, K. V. & Antos, J. M. Efficient Sortase-Mediated Ligation Using a Common C-Terminal Fusion Tag. *Bioconjug. Chem.* **31**, 1463–1473 (2020).
 184. Koczula, K. M. & Gallotta, A. Lateral flow assays. *Essays Biochem.* **60**, 111–120 (2016).
 185. Baker, A. N. *et al.* Glycan-Based Flow-Through Device for the Detection of SARS-COV-2. *ACS Sensors* acssensors.1c01470 (2021) doi:10.1021/acssensors.1c01470.
 186. Actor, J. K. Assessment of Immune Parameters and Immunodiagnostics. in *Introductory Immunology* 135–152 (Elsevier, 2014). doi:10.1016/B978-0-12-420030-2.00019-6.
 187. McBerney, R. Engineering Synthetic Glycoproteins. (University of Leeds, 2020).
 188. Koeller, K. M., Smith, M. E. B. & Wong, C. H. Chemoenzymatic synthesis of PSGL-1 glycopeptides: Sulfation on tyrosine affects glycosyltransferase-catalyzed synthesis of the O-glycan. *Bioorganic Med. Chem.* **8**, 1017–1025 (2000).
 189. Xu, R. *et al.* Site-Specific Incorporation of the Mucin-Type N - Acetylgalactosamine- α -O-threonine into Protein in Escherichia coli. *J. Am. Chem. Soc.* **126**, 15654–15655 (2004).
 190. Potter, G. T., Jayson, G. C., Miller, G. J. & Gardiner, J. M. An Updated Synthesis of the Diazo-Transfer Reagent Imidazole-1-sulfonyl Azide Hydrogen Sulfate. *J. Org. Chem.* **81**, 3443–3446 (2016).
 191. Fischer, N. *et al.* Sensitivities of Some Imidazole-1-sulfonyl Azide Salts. *J. Org. Chem.* **77**, 1760–1764 (2012).
 192. van Well, R., Ravindranathan Kartha, K. & Field, R. Iodine Promoted

- Glycosylation with Glycosyl Iodides: α -Glycoside Synthesis. *J. Carbohydr. Chem.* **24**, 463–474 (2005).
193. Grundler, G. & Schmidt, R. R. Glycosylimidate, 13. Anwendung des Trichloracetimidat-Verfahrens auf 2-Azidoglucose- und 2-Azidogalactose-Derivate. *Liebigs Ann. der Chemie* **1984**, 1826–1847 (1984).
 194. Dokurno, P. *et al.* Crystal structure at 1.95 Å resolution of the breast tumour-specific antibody SM3 complexed with its peptide epitope reveals novel hypervariable loop recognition. *J. Mol. Biol.* **284**, 713–728 (1998).
 195. Burchell, J. M., Mungul, A. & Taylor-Papadimitriou, J. O-linked glycosylation in the mammary gland: changes that occur during malignancy. *J. Mammary Gland Biol. Neoplasia* **6**, 355–64 (2001).
 196. Martínez-Sáez, N. *et al.* Deciphering the Non-Equivalence of Serine and Threonine O -Glycosylation Points: Implications for Molecular Recognition of the Tn Antigen by an anti-MUC1 Antibody. *Angew. Chemie Int. Ed.* **54**, 9830–9834 (2015).
 197. Apostolopoulos, V., Chelvanayagam, G., Xing, P.-X. & McKenzie, I. F. C. Anti-MUC1 Antibodies React Directly with MUC1 Peptides Presented by Class I H2 and HLA Molecules. *J. Immunol.* **161**, 767 LP – 775 (1998).
 198. Balmforth, M. R. Piggybacking on the cholera toxin: Using cholera toxin B chain for the targeted delivery of proteins to motor neurones. (University of Leeds, 2017).
 199. Balmforth, M. R. *et al.* Piggybacking on the Cholera Toxin: Identification of a CTB-Binding Protein as an Approach for Targeted Delivery of Proteins to Motor Neurons. *Bioconjug. Chem.* [acs.bioconjchem.1c00373](https://doi.org/10.1021/acs.bioconjchem.1c00373) (2021) doi:10.1021/acs.bioconjchem.1c00373.
 200. Holmgren, J., Lycke, N. & Czerkinsky, C. Cholera toxin and cholera B subunit as oral-mucosal adjuvant and antigen vector systems. *Vaccine* **11**, 1179–84 (1993).
 201. Pizza, M. *et al.* Mucosal vaccines: non toxic derivatives of LT and CT as mucosal adjuvants. *Vaccine* **19**, 2534–41 (2001).
 202. Dertzbaugh, M. T. & Cox, L. M. The affinity of cholera toxin for Ni²⁺ ion. *Protein Eng. Des. Sel.* **11**, 577–581 (1998).
 203. Rangappa, S. *et al.* Effects of the multiple O-glycosylation states on antibody recognition of the immunodominant motif in MUC1 extracellular tandem repeats. *Medchemcomm* **7**, 1102–1122 (2016).
 204. Arai, S., Abe, K. & Emori, Y. Phytocystatins and Their Target Enzymes-Molecular Cloning, Expression and Possible Functions. in 73–78 (1996). doi:10.1007/978-1-4613-0335-0_8.
 205. Tiede, C. *et al.* Adhiron: a stable and versatile peptide display scaffold for molecular recognition applications. *Protein Eng. Des. Sel.* **27**, 145–155 (2014).
 206. Tiede, C. *et al.* Affimer proteins are versatile and renewable affinity reagents. *Elife* **6**, (2017).
 207. Arrata, I., Barnard, A., Tomlinson, D. C. & Wilson, A. J. Interfacing native and non-native peptides: using Affimers to recognise α -helix mimicking foldamers. *Chem. Commun.* **53**, 2834–2837 (2017).
 208. Robinson, J. I. *et al.* Affimer proteins inhibit immune complex binding to Fc γ RIIIa with high specificity through competitive and allosteric modes of action. *Proc. Natl. Acad. Sci.* **115**, E72–E81 (2018).
 209. Fisher, M. J. *et al.* Trivalent Gd-DOTA reagents for modification of proteins. *RSC Adv.* **5**, 96194–96200 (2015).

210. Heidelberger, J. B. *et al.* Proteomic profiling of VCP substrates links VCP to K6-linked ubiquitylation and c-Myc function. *EMBO Rep.* **19**, (2018).
211. Miles, J. A. *et al.* Selective Affimers Recognise the BCL-2 Family Proteins BCL-x L and MCL-1 through Noncanonical Structural Motifs**. *ChemBioChem* **22**, 232–240 (2021).
212. Curd, A. P. *et al.* Nanoscale Pattern Extraction from Relative Positions of Sparse 3D Localizations. *Nano Lett.* **21**, 1213–1220 (2021).
213. Rose, K. *et al.* New Cyclization Reaction at the Amino Terminus of Peptides and Proteins. *Bioconjug. Chem.* **10**, 1038–1043 (1999).
214. Simonetti, L. & Ivarsson, Y. Genetically Encoded Cyclic Peptide Phage Display Libraries. *ACS Cent. Sci.* **6**, 336–338 (2020).
215. Hey, T., Fiedler, E., Rudolph, R. & Fiedler, M. Artificial, non-antibody binding proteins for pharmaceutical and industrial applications. *Trends Biotechnol.* **23**, 514–522 (2005).
216. Muyldermans, S. A guide to: generation and design of nanobodies. *FEBS J.* **288**, 2084–2102 (2021).
217. Robert, A. & Wiels, J. Shiga Toxins as Antitumor Tools. *Toxins (Basel)*. **13**, 690 (2021).
218. Distler, U. *et al.* Shiga Toxin Receptor Gb3Cer/CD77: Tumor-Association and Promising Therapeutic Target in Pancreas and Colon Cancer. *PLoS One* **4**, e6813 (2009).
219. Thompson, G. S., Shimizu, H., Homans, S. W. & Donohue-Rolfe, A. Localization of the Binding Site for the Oligosaccharide Moiety of Gb3 on Verotoxin 1 Using NMR Residual Dipolar Coupling Measurements. *Biochemistry* **39**, 13153–13156 (2000).
220. Calderwood, S. B., Acheson, D. W., Goldberg, M. B., Boyko, S. A. & Donohue-Rolfe, A. A system for production and rapid purification of large amounts of the Shiga toxin/Shiga-like toxin I B subunit. *Infect. Immun.* **58**, 2977–2982 (1990).
221. MacLeod, D. L. & Gyles, C. L. Effects of culture conditions on yield of Shiga-like toxin-IIv from Escherichia coli. *Can. J. Microbiol.* **35**, 623–629 (1989).
222. Jobling, M. G. & Holmes, R. K. Analysis of structure and function of the B subunit of cholera toxin by the use of site-directed mutagenesis. *Mol. Microbiol.* **5**, 1755–1767 (1991).
223. Jobling, M. G. & Holmes, R. K. Mutational Analysis of Ganglioside GM 1 - Binding Ability, Pentamer Formation, and Epitopes of Cholera Toxin B (CTB) Subunits and CTB/Heat-Labile Enterotoxin B Subunit Chimeras. *Infect. Immun.* **70**, 1260–1271 (2002).
224. Dommerholt, J. *et al.* Readily Accessible Bicyclononynes for Bioorthogonal Labeling and Three-Dimensional Imaging of Living Cells. *Angew. Chemie Int. Ed.* **49**, 9422–9425 (2010).
225. Shiao, Tze Chieh; Giguere, Denis; Galanos, Nicolas; Roy, R. Efficient synthesis of hepta-O-acetyl- β -lactosyl azide via phase transfer catalysis. in *Carbohydrate Chemistry: Proven Synthetic Methods Volume 2* (eds. van derMarcel, G. & Codee, J.) 257–262 (CRC Press, 2014).
226. Tanaka, T., Nagai, H., Noguchi, M., Kobayashi, A. & Shoda, S. One-step conversion of unprotected sugars to β -glycosyl azides using 2-chloroimidazolium salt in aqueous solution. *Chem. Commun.* 3378 (2009) doi:10.1039/b905761g.
227. Kartha, K. P. R. & Jennings, H. J. A Simplified, One-Pot Preparation of

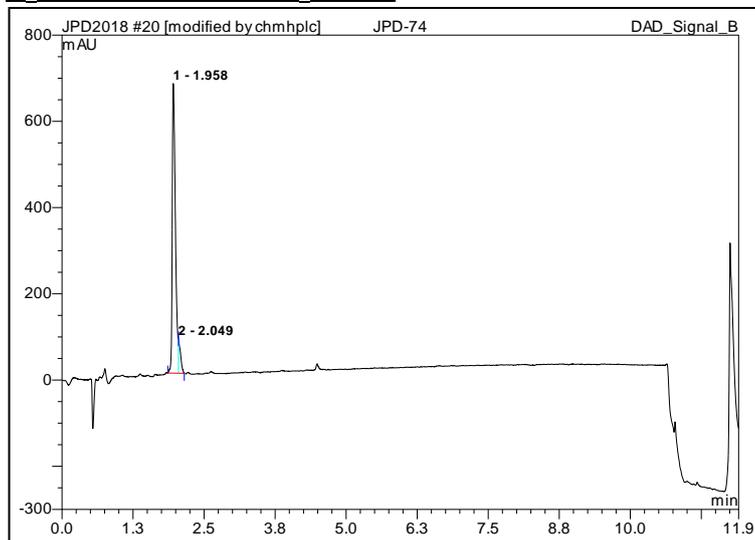
- Acetobromosugars from Reducing Sugars. *J. Carbohydr. Chem.* **9**, 777–781 (1990).
228. Machida, T., Lang, K., Xue, L., Chin, J. W. & Winssinger, N. Site-Specific Glycoconjugation of Protein via Bioorthogonal Tetrazine Cycloaddition with a Genetically Encoded trans -Cyclooctene or Bicyclononyne. *Bioconjug. Chem.* **26**, 802–806 (2015).
 229. Vaughan, M. D. *et al.* Glycosynthase-Mediated Synthesis of Glycosphingolipids. *J. Am. Chem. Soc.* **128**, 6300–6301 (2006).
 230. Mahon, C. S. *et al.* A ‘catch-and-release’ receptor for the cholera toxin. *Faraday Discuss.* **219**, 112–127 (2019).
 231. Zomer-van Ommen, D. D. *et al.* Functional Characterization of Cholera Toxin Inhibitors Using Human Intestinal Organoids. *J. Med. Chem.* **59**, 6968–6972 (2016).
 232. Wang, Z. *et al.* The Basic Characteristics of the Pentraxin Family and Their Functions in Tumor Progression. *Front. Immunol.* **11**, (2020).
 233. Gewurz, H., Zhang, X.-H. & Lint, T. F. Structure and function of the pentraxins. *Curr. Opin. Immunol.* **7**, 54–64 (1995).
 234. Pepys, M. B. *et al.* Amyloid P component. A critical review. *Amyloid* **4**, 274–295 (1997).
 235. Tennent, G. A. & Pepys, M. B. Glycobiology of the pentraxins. *Biochem. Soc. Trans.* **22**, 74–79 (1994).
 236. Srinivasan, N. *et al.* Comparative analyses of pentraxins: implications for protomer assembly and ligand binding. *Structure* **2**, 1017–1027 (1994).
 237. Tanaka, T., Horio, T. & Matuo, Y. Secretory production of recombinant human C-reactive protein in *Escherichia coli*, capable of binding with phosphorylcholine, and its characterization. *Biochem. Biophys. Res. Commun.* **295**, 163–166 (2002).
 238. Cavard, D. Synthesis and functioning of the colicin E1 lysis protein: comparison with the colicin A lysis protein. *J. Bacteriol.* **173**, 191–196 (1991).
 239. Dortay, H., Schmöckel, S. M., Fettke, J. & Mueller-Roeber, B. Expression of human c-reactive protein in different systems and its purification from *Leishmania tarentolae*. *Protein Expr. Purif.* **78**, 55–60 (2011).
 240. Lee, H.-H. *et al.* Expression, purification and crystallization of CTB-MPR, a candidate mucosal vaccine component against HIV-1. *IUCrJ* **1**, 305–317 (2014).
 241. Lebens, M. & Holmgren, J. Mucosal vaccines based on the use of cholera toxin B subunit as immunogen and antigen carrier. *Dev. Biol. Stand.* **82**, 215–27 (1994).
 242. Rehm, F. B. H. *et al.* Site-Specific Sequential Protein Labeling Catalyzed by a Single Recombinant Ligase. *J. Am. Chem. Soc.* **141**, 17388–17393 (2019).
 243. Cremer, G.-A. *et al.* Synthesis of Branched Oxime-Linked Peptide Mimetics of the MUC1 Containing a Universal T-Helper Epitope. *Chem. - A Eur. J.* **10**, 6353–6360 (2004).
 244. Warkentin, R. & Kwan, D. H. Resources and Methods for Engineering “Designer” Glycan-Binding Proteins. *Molecules* **26**, 380 (2021).
 245. Hofmeister, D. L., Thoden, J. B. & Holden, H. M. Investigation of a sugar N -formyltransferase from the plant pathogen *Pantoea ananatis*. *Protein Sci.* **28**, 707–716 (2019).
 246. Shilova, O. N. & Deyev, S. M. DARPinS: Promising Scaffolds for Theranostics. *Acta Naturae* **11**, 42–53 (2019).

247. Stumpp, M. T., Binz, H. K. & Amstutz, P. DARPinS: A new generation of protein therapeutics. *Drug Discov. Today* **13**, 695–701 (2008).
248. Avacta. *Validation of AffiDX® SARS-CoV-2 Antigen Lateral Flow Test with Variants of Concern. Validation of AffiDX® SARS-CoV-2 Antigen Lateral Flow Test with Variants of Concern* <https://avacta.com/diagnostics/products/validation-of-affidx-sars-cov-2-antigen-lft-with-variants-of-concern/> (2021).
249. Morgan, F. *Avacta Group receives UK approval for AffiDX® COVID-19 test. Avacta Group receives UK approval for AffiDX® COVID-19 test in Vox Markets* <https://www.voxmarkets.co.uk/articles/avacta-group-receives-uk-approval-for-affidx-covid-19-test-0fa651e/> (2021).
250. Ewers, H. *et al.* GM1 structure determines SV40-induced membrane invagination and infection. *Nat. Cell Biol.* **12**, 11–18 (2010).
251. Fulmer, G. R. *et al.* NMR Chemical Shifts of Trace Impurities: Common Laboratory Solvents, Organics, and Gases in Deuterated Solvents Relevant to the Organometallic Chemist. *Organometallics* **29**, 2176–2179 (2010).
252. Alper, P. B., Hung, S.-C. & Wong, C.-H. Metal catalyzed diazo transfer for the synthesis of azides from amines. *Tetrahedron Lett.* **37**, 6029–6032 (1996).
253. De Silva, R. A., Wang, Q., Chidley, T., Appulage, D. K. & Andreato, P. R. Immunological response from an entirely carbohydrate antigen: Design of synthetic vaccines based on Tn-PS A1 conjugates. *J. Am. Chem. Soc.* **131**, 9622–9623 (2009).
254. Seifried, B. M. *et al.* Glycoprotein Mimics with Tunable Functionalization through Global Amino Acid Substitution and Copper Click Chemistry. *Bioconjug. Chem.* **31**, 554–566 (2020).

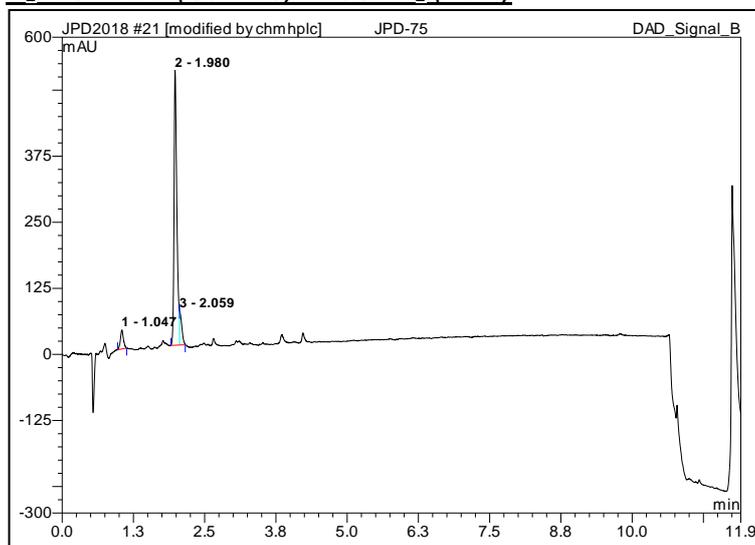
Chapter 10: Appendix

10.1 Peptide & Glycopeptide HPLC Traces

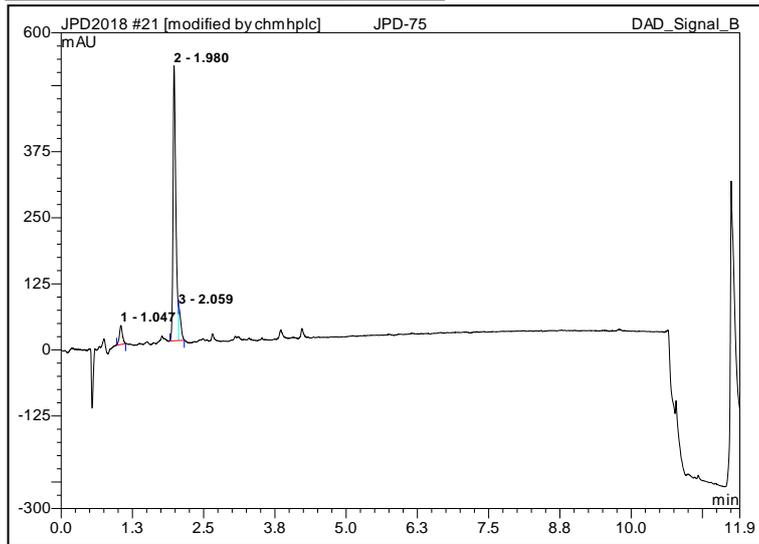
H₂N-GVAPDTRP-CONH₂ (2.13)



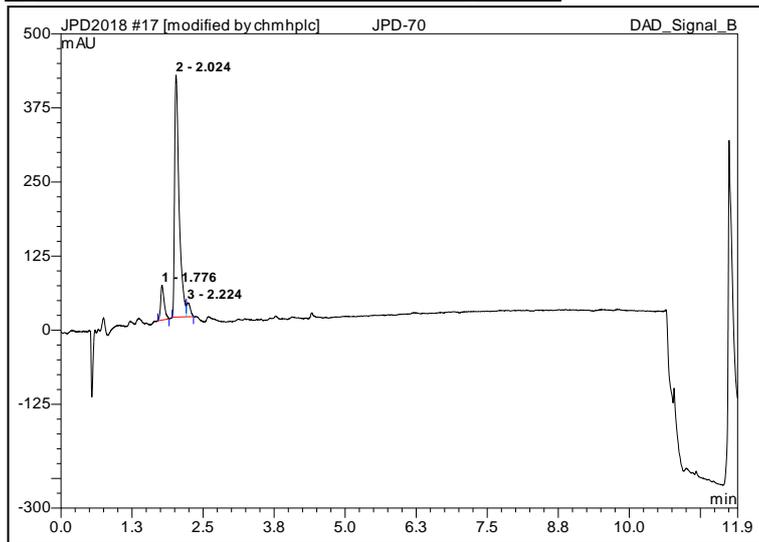
H₂N-GVAPDT(GalNAc)RP-CONH₂ (2.14)



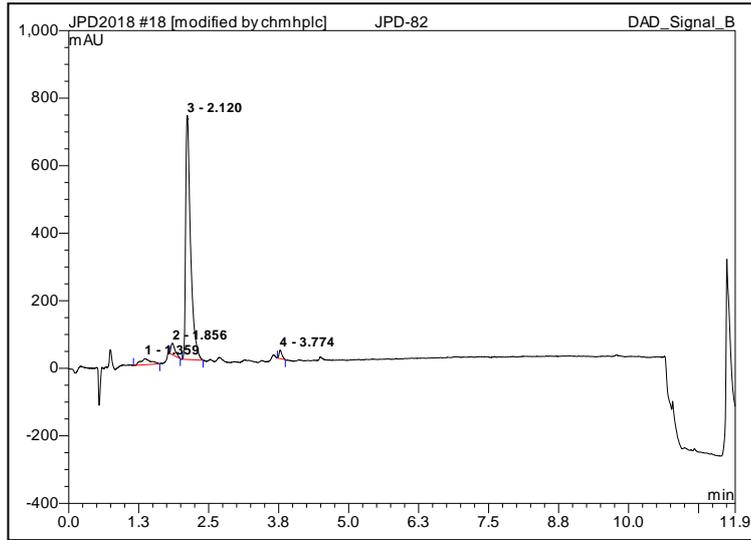
H₂N-GVAPGSTAPPA-CONH₂ (2.15)



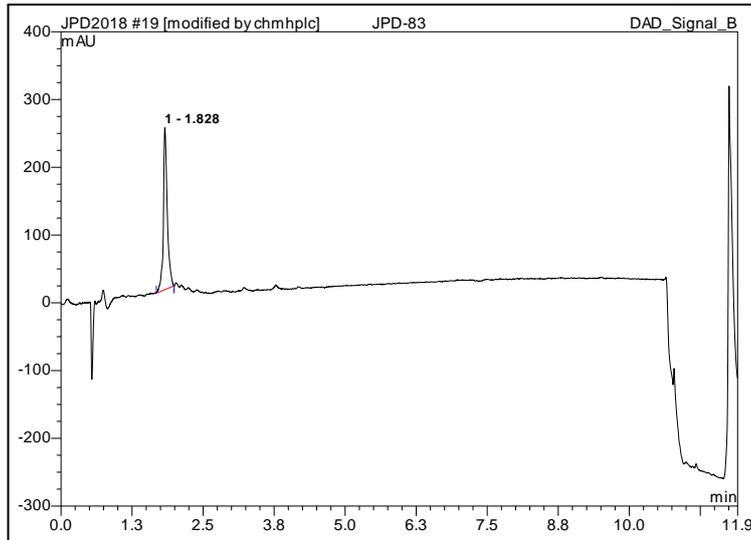
H₂N-GVAPGST(GalNAc)APPA-CONH₂ (2.16)



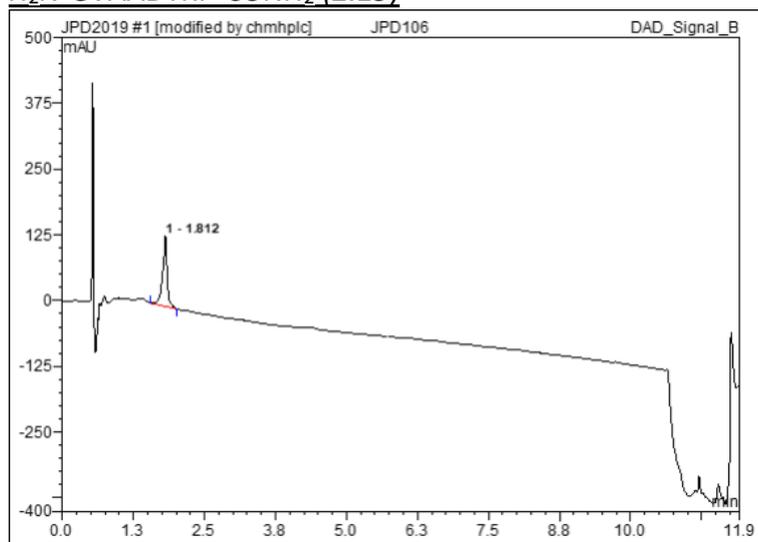
H₂N-GVAPGS(GalNAc)TAPPA-CONH₂ (2.17)



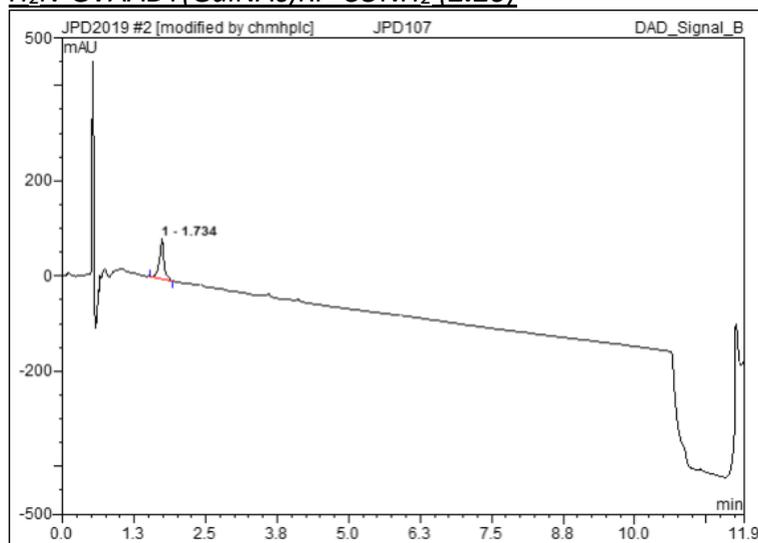
H₂N-GVAPGS(GalNAc)T(GalNAc)APPA-CONH₂ (2.18)



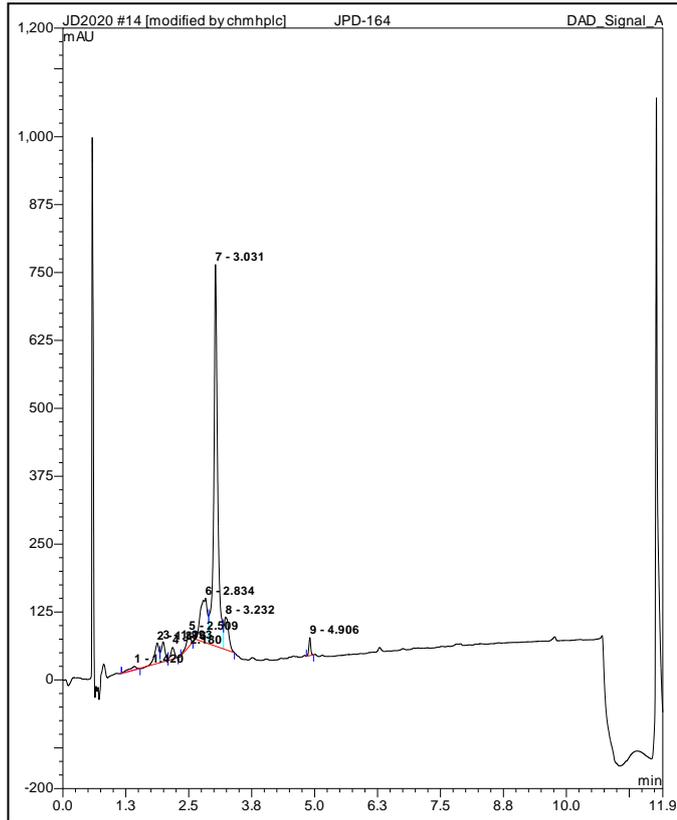
H₂N-GVAADTRP-CONH₂ (2.19)



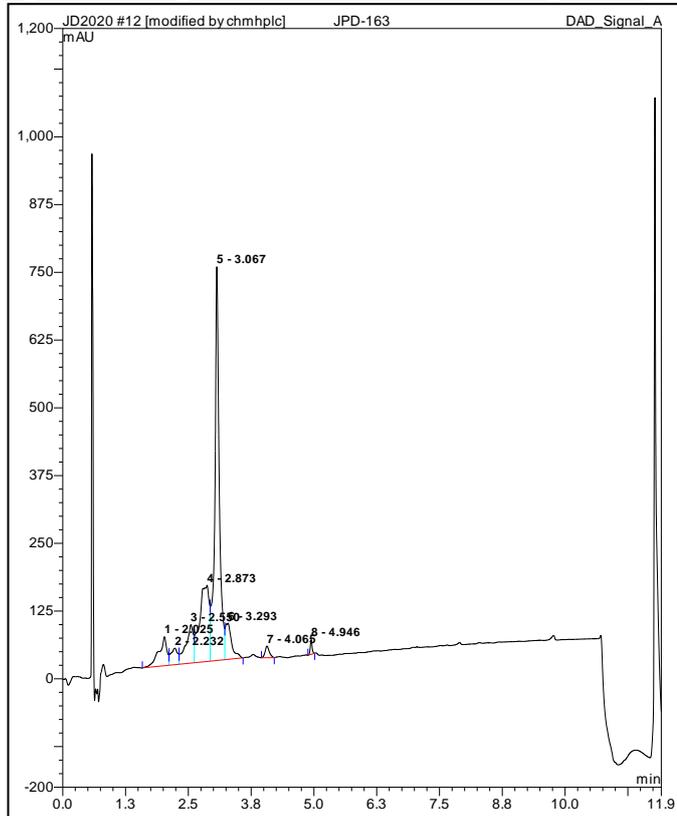
H₂N-GVAADT(GalNAc)RP-CONH₂ (2.20)



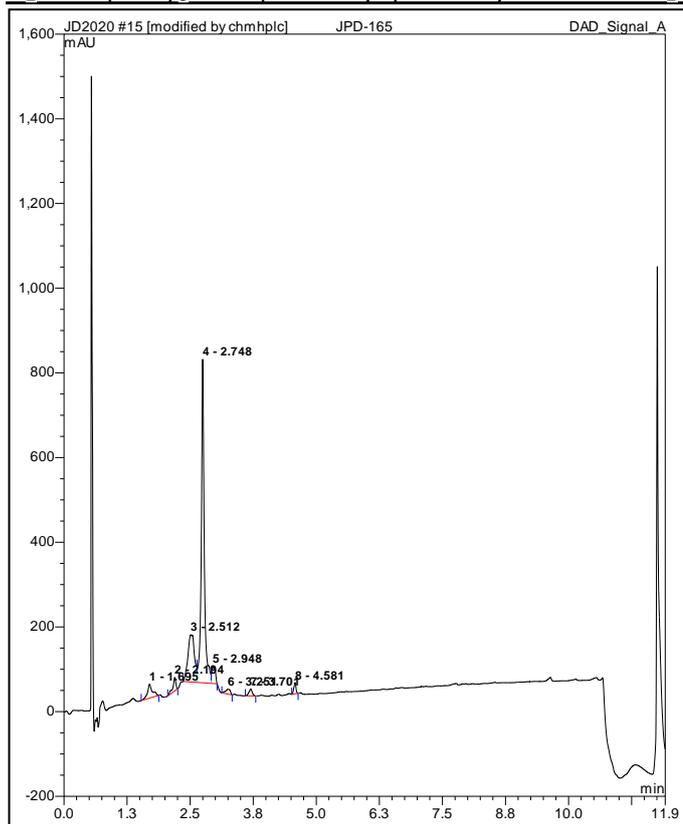
H₂N-GV(PEG)₆APGST(GalNAc)APPA-CONH₂ (2.26)



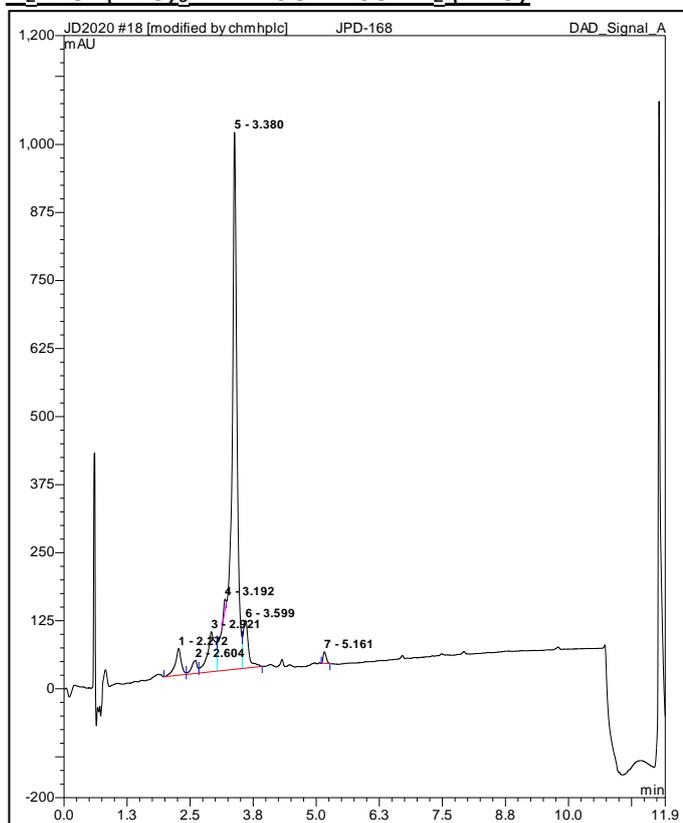
H₂N-GV(PEG)₆APGS(GalNAc)TAPPA-CONH₂ (2.27)



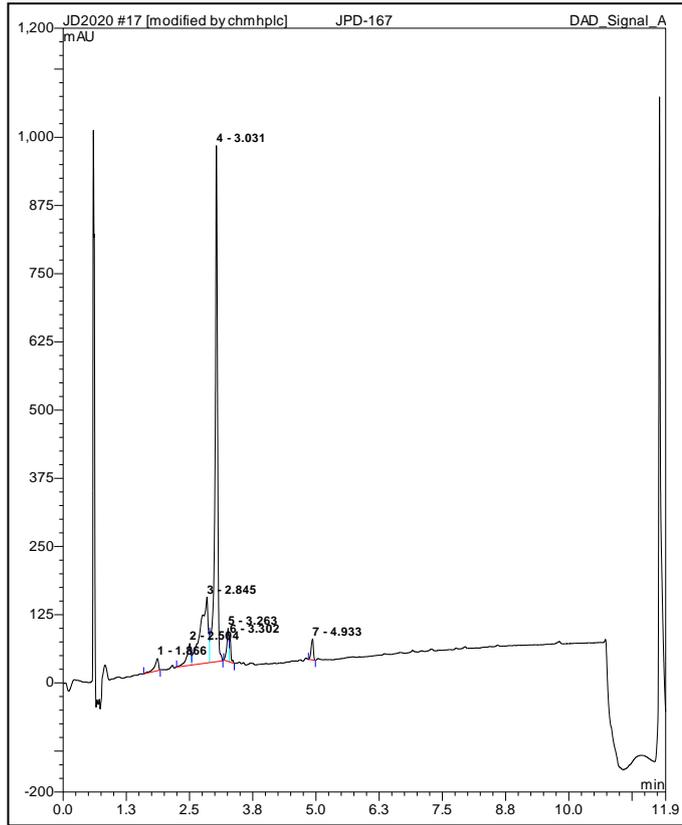
$H_2N-GV(PEG)_6APGS(GalNAc)T(GalNAc)APPA-CONH_2$ (2.28)



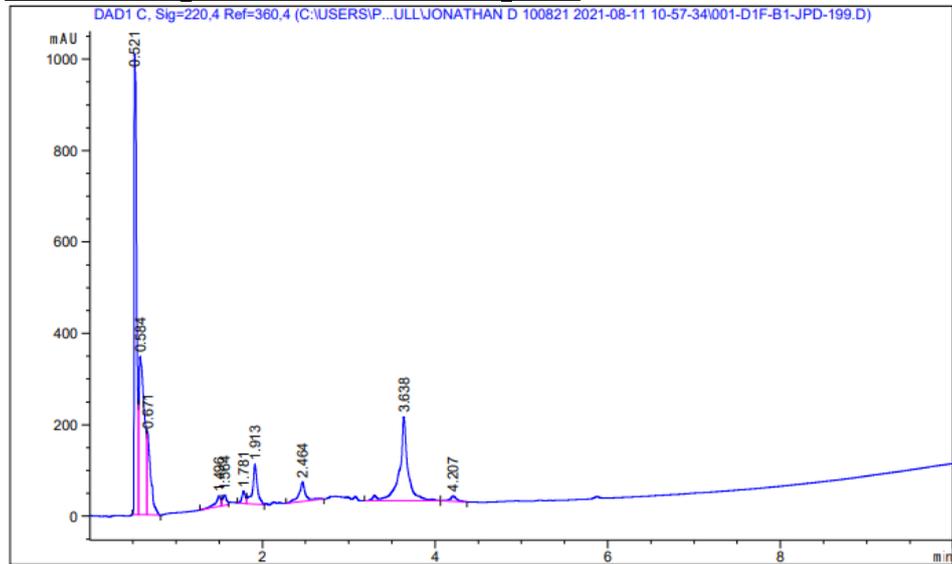
$H_2N-GV(PEG)_6APPATSGPA-CONH_2$ (2.29)



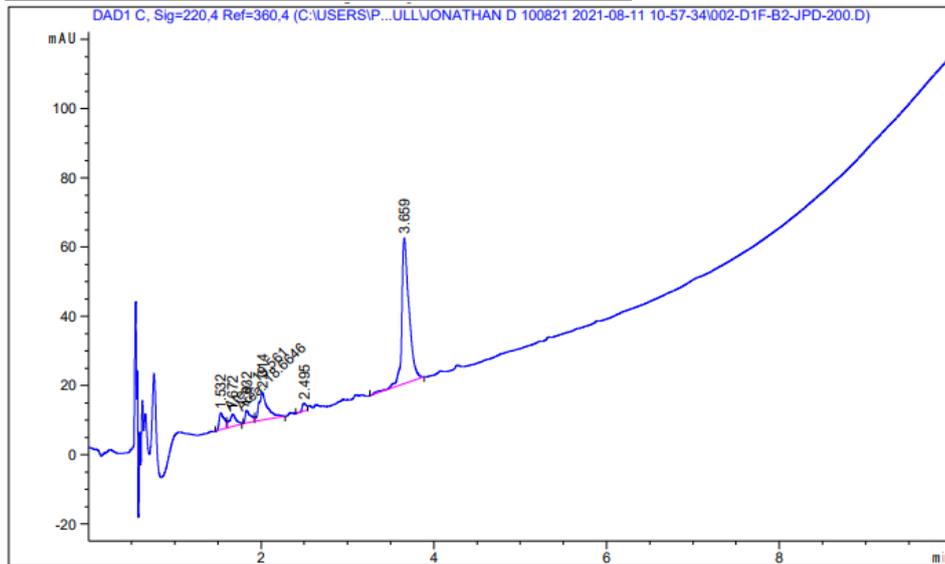
H₂N-GV(PEG)₆APDTRP-CONH₂ (2.30)



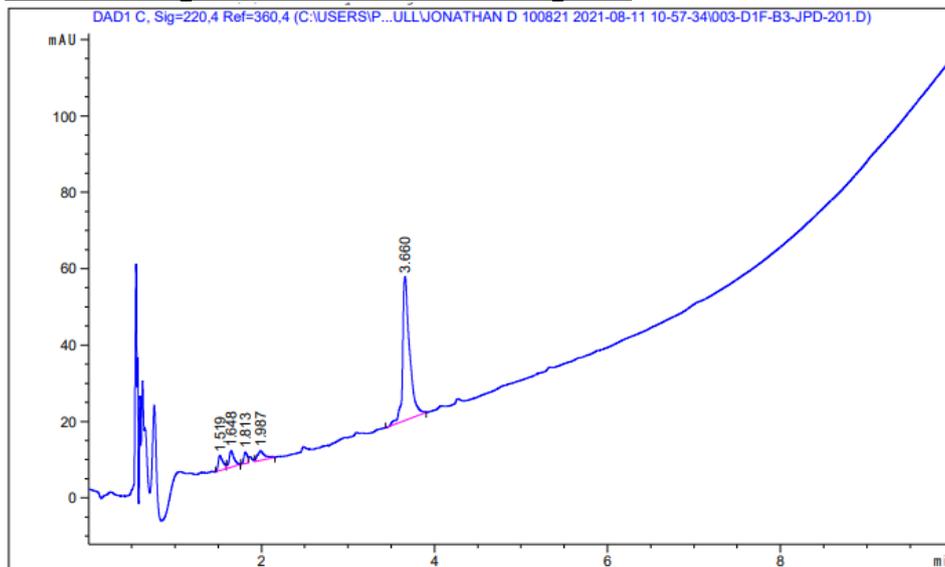
BCN-GV(PEG)₆APDT(GalNAc)RP-CONH₂ (3.7)



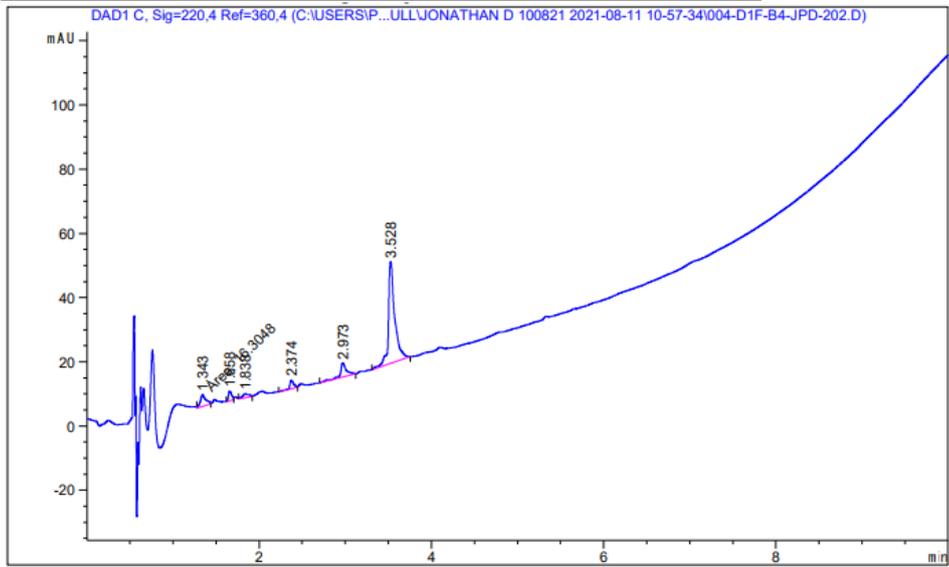
BCN-GV(PEG)₆APGST(GalNAc)APPA-CONH₂ (3.8)



BCN-GV(PEG)₆APGS(GalNAc)TAPPA-CONH₂ (3.9)



BCN-GV(PEG)₆APGS(GalINAc)₇(GalINAc)APPA-CONH₂ (3.10)



10.2 Plasmid Sequences

pSAB2.2-LPETGA

The pSAB2.2-LPETGA plasmid containing CTB-LPETGA was derived from pMAL-p5X plasmid available from NEB. This plasmid was used to express CTB-LPETGA as prepared by Darren Machin (University of Leeds). CTB-LPETGA is shown in red text, the periplasmic targeting sequence purple text, and the SphI and PstI restriction sites used for sub-cloning by green and blue highlighting respectively.

```
CCGACACCATCGAATGGTGCAAACCTTTTCGCGGTATGGCATGATAGCGCCCAGGAAGAG
AGTCAATTCAGGGTGGTGAATGTGAAACCAGTAACGTTATACGATGTCGCAGAGTATGC
CGGTGTCTCTTATCAGACCGTTTCCCGCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAA
AACGCGGGAAAAAGTGGAAGCGGCGATGGCGGAGCTGAATTACATTCCCAACCGCGTG
GCACAACAACCTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCCAGTCTGGC
CCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGGTG
CCAGCGTGGTGGTGTGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGT
GCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCA
GGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGCGGTTATTTCTTGATGTCTC
TGACCAGACACCCATCAACAGTATTATTTCTCCCATGAAGACGGTACGCGACTGGGCG
TGGAGCATCTGGTCGATTGGGTACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAAGT
TCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAATCAAATT
CAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCAT
GCAAATGCTGAATGAGGGCATCGTTCCTCACTGCGATGCTGGTTGCCAACGATCAGATGG
CGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGGGATATTTCCG
GTAGTGGGATACGACGATACCGAAGACAGCTCATGTTATATCCCGCCGTTAACCCACCAT
CAAACAGGATTTTCGCCTGCTGGGGCAAACCCAGCGTGGACCGCTTGTGCAACTCTCTC
AGGGCCAGGGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAAC
CACCCTGGCGCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCA
GCTGGCACGACAGGTTTTCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTA
AGTTAGCTCACTCATTAGGCACAATTCTCATGTTTGACAGCTTATCATCGACTGCACGGT
GCACCAATGCTTCTGGCGTCAGGCAGCCATCGGAAGCTGTGGTATGGCTGTGCAGGTCG
TAAATCACTGCATAATTCGTGTCGCTCAAGGCGCACTCCCGTTCTGGATAATGTTTTTTG
CGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTGACAATTAATCATC
GGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTACACAGGAAACAGCCAGT
CCGTTTAGGTGTTTTACGAGCAATTGACCAACAAGGACCATAGATTATGAGCTTTAAG
AAAATTATCAAGGCATTTGTTATCATGGCTGCTTTGGTATCTGTTTCAGGCCTCAAACT
CCTCAAAATATTACTGATTTGTGCGCAGAATACCACAACACACAATATATACGCTAAA
TGATAAGATCTTTTCGTATACAGAATCGCTAGCGGGAAAAAGAGAGATGGCTATCATT
CTTTTAAGAATGGTGCAATTTTTCAAGTAGAGGTACCAGGTAGTCAACATATAGATTAC
AAAAAAGGCAATCGAACGTATGAAGGATACCCTGAGGATTGCATATCTTACTGAAGCT
AAAGTCGAAAAGTTATGTGTATGGAATAATAAAACGCCCTCATGCGATCGCCGCAATTAG
TATGGCAAACAATGGCGGTAACCTGCCGGAACCGGTGCGTAAAGTTTTCCCTGCAGGTA
ATTAATAAGCTTCAAATAAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTA
TCTGTTGTTTGTGCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGA
ACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCCGATAAACTGCCAG
GCATCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCT
TTCGGTCCGTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAA
CCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGATGATGAGTATCAACATTTCCG
TGTCGCCCTTATCCCTTTTTGCGGCATTTTGCCTTCTGCTTACTCACCCAGAAACG
CTGGTGAAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACT
GGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTCCCAATGA
TGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTGACGCCGGGCAAG
AGCAACTCGGTGCGCCGATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCA
CAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACC
ATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCT
```

AACCGCTTTTTTGCACAACATGGGGGATCATGTAACCTCGCCTTGATCGTTGGGAACCGGA
GCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCA
ACAACGTTGCGCAAACCTATTAAGTGGCGAACTACTTACTCTAGCTTCCCAGCAACAATTA
ATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCCTTCTGCGCTCGGCCCTTCCGGC
TGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGC
AGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTC
AGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAG
CATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTTCCTTAGGACT
GAGCGTCAACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGT
AATCTGCTGCTTGCAAACAAAAAACCACCGCTACCAGCGGTGGTTTTGTTGCCGGATC
AAGAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCGCAGATAACCAAT
ACTGTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCCGCT
ACATACTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCTGTGT
CTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTGGGGTGAAC
GGGGGTTCTGTCACACAGCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACC
TACAGCTTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTA
TCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAAC
GCCTGGTATCTTTATAGTCCTGTGCGGTTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGT
GATGCTCGTCAGGGGGGCGGAGCCTATGGAAAACGCCAGCAACGCGGCCTTTTTACGG
TTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCTGCGTTATCCCCTGATTCTGT
GGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGACGCCGAACGACCG
AGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTT
ACGCATCTGTGCGGTATTTACACCCGCATATAAGGTGCACTGTGACTGGGTGATGGCTGC
GCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCAT
CCGTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTACCGT
CATCACCGAAACGCGCGAGGCAGCTGCGGTAAGCTCATCAGCGTGGTCTGTCAGCGAT
TCACAGATGTCTGCCTGTTTATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAAT
GTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCTGTTTGGTCACTGAT
GCCTCCGTGTAAGGGGGATTTCTGTTTATGGGGTAATGATACCGATGAAACGAGAGAG
GATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGG
GTAAACAACCTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATG
CCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCTGCGA
TGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTTCCAGACTTTACGAA
ACACGGAAACCGAAGACCATTGTTGTTGCTCAGGTGCGCAGACGTTTTGTCAGCAGCA
GTCGTTTACGTTTCGCTCGCGTATCGGTGATTGCTGCTAACCAGTAAGGCAACCCCG
CCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGCCAGGACC
CAACGCTGCCCCGAAATT

pGW-CTB-LPETGAS-H₆

The pGW-CTB-LPETGAS-H₆ plasmid containing CTB-LPETGAS-His₆ was derived from the pSAB2.2-LPETGA plasmid containing CTB-LPETGA made by Darren Machin (University of Leeds). This plasmid was used to express CTB-LPETGAS-His₆ as prepared by Gemma Wildsmith (University of Leeds) by first removing the periplasmic targeting sequence then inserting a His₆ tag and reintroducing Methionine(1). CTB-LPETGAS-His₆ is shown in red text and the MfeI and PstI restriction sites used for sub-cloning by green and blue highlighting respectively.

CCGACACCATCGAATGGTGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCCGGAAGAG
AGTCAATTCAGGGTGGTGAATGTGAAACCAGTAACGTTATACGATGTGCGCAGAGTATGC
CGGTGTCTTATCAGACCGTTTTCCGCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAA
AACGCGGGAAAAAGTGGAAAGCGGCGATGGCGGAGCTGAATTACATTCCCAACCGCGTG
GCACAACAACCTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCCAGTCTGGC
CCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGGTG

CCAGCGTGGTGGTGTTCGATGGTAGAACGAAGCGGGCGTCGAAGCCTGTAAAGCGGGCGT
GCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCA
GGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTC
TGACCAGACACCCATCAACAGTATTATTTTCTCCCATGAAGACGGTACGCGACTGGGCG
TGGAGCATCTGGTCGATTGGGTACCAGCAAATCGCGCTGTTAGCGGGGCCATTAAGT
TCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAATCAAATT
CAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCAT
GCAATGCTGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAGATGG
CGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGATATTTG
GTAGTGGGATACGACGATACCGAAGACAGCTCATGTTATATCCCGCCGTTAACCCAT
CAAACAGGATTTTCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGTGCAACTCTCTC
AGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAAC
CACCTGGCGCCAATACGCAAACCGCTCTCCCCGCGCGTTGGCCGATTCATTAATGCA
GCTGGCACGACAGGTTTCCCGACTGGAAGCGGGCAGTGAGCGCAACGCAATTAATGTA
AGTTAGCTACTCATTAGGCACAATTCTCATGTTTGACAGCTTATCATCGACTGACGGT
GCACCAATGCTTCTGGCGTCAGGCAGCCATCGGAAGCTGTGGTATGGCTGTGCAGGTCG
TAAATCACTGCATAATTCGTGTGCTCAAGGCGCACTCCCCTTCTGGATAATGTTTTTTG
CGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTGACAATTAATCATC
GGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTACACAGGAAACAGCCAGT
CCGTTTAGGTGTTTTACGAGCAATTGACCAACAAGGACCATAGATGACTCCTCAAAAT
ATTACTGATTTGTGCGCAGAATACCACAACACACAATATATACGCTAAATGATAAGAT
CTTTTCGTATACAGAATCGCTAGCGGGAAAAAGAGAGATGGCTATCATTACTTTAAGA
ATGGTGAATTTTTCAAGTAGAGGTACCAGGTAGTCAACATATAGATTCACAAAAAAG
GCAATCGAACGTATGAAGGATACCCTGAGGATTGCATATCTTACTGAAGCTAAAGTCGA
AAAGTTATGTGTATGGAATAATAAACGCCTCATGCGATCGCCGAATTAGTATGGCAA
ACAATGGCGGTAACCTGCCGAAACCGGTGCGAGCCACCATCATCATCATTAATAA
GTTTTCCCTGCGAGTAATTAATAAGCTTCAAATAAAAACGAAAGGCTCAGTCGAAAGAC
TGGGCTTTTCGTTTTATCTGTTGTTGTGCGGTGAACGCTCTCCTGAGTAGGACAAATCCG
CCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCC
CGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTT
GCGTTTCTACAACTCTTTCGGTCCGTTGTTTATTTTTCTAAATACATTCAAATATGTATC
CGCTCATGAGACAATAACCCTGATAAATGCTTCAAATAATATTGAAAAAGGAAGAGTATG
AGTATTCAACATTTCCGTGTGCGCCTTATCCCTTTTTTTCGGGCATTTTGCCTTCTGTTTT
TGCTACCCAGAAACGCTGGTGAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAG
TGGGTTACATCGAACTGGATCTAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAA
GAACGTTTCCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGT
GTTGACGCCGGCAAGAGCAACTCGGTGCGCCGATACACTATTCTCAGAATGACTTGGT
TGAGTACTACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGAAGAGAATTAT
GCAGTCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATC
GGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACCTCGCCT
TGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACG
ATGCCTGTAGCAATGGCAACAACGTTGCGCAAATACTTAACTGGCGAACTACTTACTCT
AGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTC
TGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTG
GGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTA
TCTACACGACGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGAT
AGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTA
GATTGATTTCTTAGGACTGAGCGTCAACCCCGTAGAAAAGATCAAAGGATCTTCTTGA
GATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCAAAACAAAAAACCACCGCTACCAGCG
GTGGTTTGTGTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACCTGGCTTCAGC
AGAGCGCAGATACCAAATACTGTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAA
GAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGC
CAGTGGCGATAAGTTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGG
CGCAGCGGTGCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGAC
CTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAG
GGAGAAAGGCGGACAGGTATCCGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGA
GGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCTGTCGGGTTTTGCCACCTCT
GACTTGAGCGTGCATTTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGGAAAAACGCC
AGCAACGCGGCCTTTTACGGTTCCTGGCTTTTGTGCTGGCTTTTGTCTCATATGTTCTTC
CTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCTTTGAGTGAGCTGATACCG

CTCGCCGAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCG
CCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTACACCCGCATATAAGGTGC
ACTGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGA
CGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGC
ATGTGTCAGAGGTTTTACCGTCATACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTC
ATCAGCGTGGTCGTGCAGCGATTACAGATGTCTGCCTGTTTCATCCGCGTCCAGCTCGTT
GAGTTTCTCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGG
TTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTCTGTTTCATGGGGGTAAT
GATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCC
GGTTACTGGAACGTTGTGAGGGTAAACAACCTGGCGGTATGGATGCGGCGGGACCAGAG
AAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGG
GTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTC
CGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTTCATGTTGTTGCTCAGGTC
GCAGACGTTTTGCAGCAGCAGTCGCTTACGTTTCGCTCGCGTATCGGTGATTTCATTCTGC
TAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCTCAACGACAGGAGCACGATCAT
GCGCACCCGTGGCCAGGACCCAACGCTGCCCGAAATT

pET30b (including Sortase7M-His₆)

The pET30b plasmid containing Sortase7M with a C-terminal His tag was produced by Tomasz Kaminski (University of Leeds). Sortase7M-His₆ is shown with red text and the NdeI and XhoI restriction sites used for sub-cloning by green and blue highlighting respectively.

TGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAACCACCGCTACCA
GCGGTGGTTTGGTTTCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGGCTTC
AGCAGAGCGCAGATACCAAATACTGTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTT
CAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGC
TGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATA
AGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCAGCTTGGAGCGAAC
GACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCG
AAGGGAGAAAGGCGGACAGGTATCCGGTAAAGCGGCAGGGTCGGAACAGGAGAGCGCA
CGAGGCGACTTCCAGGGGAAACGCCTGGTATCTTTATAGTCTGTCGGGTTTCGCCACC
TCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGAGCCTATGAAAAAC
GCCAGCAACGCGGCCTTTTACGGTTCCTGGCCTTTTGTGCGCCTTTTGTCTACATGTTCT
TTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATA
CCGCTCGCCGACCGGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGA
GCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTACACCCGCAATGGTGC
ACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACTACTCCGCTATCGC
TACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGA
CGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGC
ATGTGTCAGAGGTTTTACCGTCATACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTC
ATCAGCGTGGTCGTGAAGCGATTACAGATGTCTGCCTGTTTCATCCGCGTCCAGCTCGTT
GAGTTTCTCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGG
TTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTCTGTTTCATGGGGGTAAT
GATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCC
GGTTACTGGAACGTTGTGAGGGTAAACAACCTGGCGGTATGGATGCGGCGGGACCAGAG
AAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGG
GTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTC
CGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTTCATGTTGTTGCTCAGGTC
GCAGACGTTTTGCAGCAGCAGTCGCTTACGTTTCGCTCGCGTATCGGTGATTTCATTCTGC
TAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCTCAACGACAGGAGCACGATCAT
GCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTTGG
TGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATCCGAATACCGCAAG
CGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCTCGCCGAAAATGACCCAGA
GCGCTGCCGGCACCTGTCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCG

ACGATAGTCATGCCCCGCGCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGG
CATCGGTGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTC
ACTGCCCGCTTTCCAGTCGGGAAACCTGTCTGTGCCAGCTGCATTAATGAATCGGCCAAC
GCGCGGGGAGAGGCGGTTTTCGTATTGGGCGCCAGGGTGGTTTTTCTTTTACCAGTGA
GACGGGCAACAGCTGATTGCCCTTACC GCCTGGCCCTGAGAGAGTTGCAGCAAGCGGT
CCACGCTGGTTTTGCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATA
TAACATGAGCTGTCTTCGGTATCGTTCGTATCCCACTACCGAGATGTCCGCACCAACGCG
AGCCCCGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAG
CATCGCAGTGGGAACGATGCCCTCATTACGATTTTGCATGGTTTTGTTGAAAACCGGACAT
GGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTT
ATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGA ACTTAATGGGCCCGCTAACAGC
GCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCAGTCGCGTACCGTCTTCA
TGGGAGAAAATAACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAAATAACGCCG
AACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAA
TGATCAGCCACTGACGCTTTCGCGGAGAAGATTGTGCACCGCCGCTTACAGGCTTCG
ACGCCGTTTCGTTCTACCATCGACACCACCGCTGGCACCCAGTTGATCGGCGCGGAGA
TTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGC
CAATCAGCAACGACTGTTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCA
GCTCCGCCATCGCCGTTTCCACTTTTTCCCGCGTTTTTCGCAGAAACGTGGCTGGCCTGGT
TCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAAC
GTTACTGGTTTTACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATA
CCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGA
CTCCTGCATTAGGAAGCAGCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCCGCCGCAA
GGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACC
ATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATC
GGTGTATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGTATGCCGGCA
CGATGCGTCCGGCGTAGAGGATCGAGATCGATCTCGATCCCGCGAAATTAATACGACTC
ACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTGTTTAACTTT
AAGAAGGAGATATA **CATATGCAAGCTAAACCTCAAATTCCGAAAGATAAATCAAAAAGT**
GGCAGGCTATATTGAAATTCCAGATGCTGATATTAAGAACCAGTATATCCAGGACCAG
CAACACGCGAACAATTAATAGAGGTGAAGCTTTGCAAAAGAAAATCAATCACTAGAT
GATCAAAAATTTTCTATTGCAGGACACACTTTTATTGACCGTCCGAACTATCAATTTACA
AATCTTAAAGCAGCCAAAAAAGGTAGTATGGTGTACTTTAAAGTTGGTAATGAAACACG
TAAGTATAAAATGACAAGTATAAGAAACGTTAAGCCAACAGCTGTAGAAGTTCTGGATG
AACAAAAAGGTAAAGATAAACAATTAACATTAATTAATTGTTGATGATTACAATGAAGAG
ACAGGCGTTTTGGGAAACACGTAAAATCTTTGTAGCTACAGAAGTCAAACCTCGAG**CACCA**
CCACCACCACACTGAGATCCGGCTGCTAACAAAGCCCCGAAAGGAAGCTGAGTTGGCTG
CTGCCACCGCTGAGCAATAACTAGCATAAACCCTTGGGGCCTTAAACGGGTCTTGAGG
GGTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCTGTAGC
GGCGCATTAAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAG
CGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCTTCTTCTCGCCACGTTTCGCCGGCTTT
CCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCAC
CTCGACCCCCAAAAA ACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATA
GACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAA
ACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTATTATAAGGGATTTTGCCG
ATTTCCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAA
AAAATATTAACGCTTACAATTTAGGTGGCACTTTTTCGGGAAATGTGCGCGGAACCCCT
ATTTGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATCTTAGAAA
AACTCATCGAGCATCAAATGAACTGCAATTTATTATCATATCAGGATTATCAATACCATAT
TTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAA ACTCACCGAGGCAGTTCCATAGGAT
GGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAA
TTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAAT
CCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTC AACAGGCCAGCCAT
TACGCTCGTCATCAAAAATCACTCGCATCAACCAACCGTTATTCAATTCGTGATTGCGCCT
GAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATG
CAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTACCTGAATCAGGATATT
CTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCAT
CAGGAGTACGGATAAAAATGCTTGATGGTTCGGAAGAGGCATAAATTCGTCAGCCAGTTT
AGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAAC
AACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTGCACCTGATTGCCCGACA

TTATCGCGAGCCCATTTATACCCATATAAAATCAGCATCCATGTTGGAATTTAATCGCGGC
CTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATG
TAAGCAGACAGTTTTATTGTTTCATGACCAAATCCCTTAACGTGAGTTTTCGTCCACTG
AGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCT

pET28a (including CBD-Sortase7M)

The pET28a plasmid containing Sortase7M with a N-terminal His₆-Chitin binding domain was designed by Michael Webb (University of Leeds) and ordered from/synthesised by genscript. This plasmid was used to express CBD-Sortase7M. Sortase7M is shown with red text, the chitin binding domain is shown in blue text and the XbaI and NdeI restriction sites used for sub-cloning by green and blue highlighting respectively.

TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAAGCGCGGCGGGTGTGGTGGTTACGC
GCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCCTCCCTTC
CTTCTCGCCACGTTTCGCGGCTTTCCTCGTCAAGCTCTAAATCGGGGGCTCCCTTAGG
GTTCCGATTTAGTGCTTACGGCACCTCGACCCCAAAAACCTTGATTAGGGTGATGGTTC
ACGTAGTGGGCATCGCCCTGATAGACGGTTTTTCGCCCTTGGACGTTGGAGTCCACGTT
CTTTAATAGTGGACTCTGTTCAAAACCTGGAACAACACTCAACCCTATCTCGGTCTATTC
TTTTGATTTATAAGGGATTTTGCCGATTTTCGCGCTATTGGTTAAAAAATGAGCTGATTTA
ACAAAAATTTAACGCGAATTTTAAACAAAATATTAACGTTTACAATTTACAGGTGGCACTTT
TCGGGGAATGTGCGCGGAACCCCTATTTGTTTATTTTCTAAATACATTCAAATATGTA
TCCGCTCATGAATTAATTCTTAGAAAACTCATCGAGCATCAAATGAAACTGCAATTTAT
TCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAA
ACTCACCGAGGCAGTTCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACT
CGTCCAACATCAATACAACCTATTAATTTCCCTCGTCAAAAATAAGGTTATCAAGTGAG
AAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTT
CCAGACTTGTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCA
AACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAA
GGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAA
CAATATTTTACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGA
TCGAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCCGGA
AGAGGCATAAATCCGTCAGCCAGTTAGTCTGACCATCTCATCTGTAACATCATTGGCA
ACGCTACCTTTGCCATGTTTCAGAAACAACCTCTGGCGCATCGGGCTTCCCATACAATCGA
TAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCATTTATACCCATATAAAATCA
GCATCCATGTTGGAATTTAATCGCGGCTAGAGCAAGACGTTTCCCGTTGAATATGGCTC
ATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTTCATGACCAAAT
CCCTTAACGTGAGTTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGAT
CTTCTTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGC
TACCAGCGGTGGTTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACG
GCTTCAGCAGAGCGCAGATACCAAATACTGTCTTCTAGTGTAGCCGTAGTTAGGCCAC
CACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTG
GCTGCTGCCAGTGCGGATAAGTCTGTCTTACCAGGTTGGACTCAAGACGATAGTTACC
GGATAAGGCGCAGCGGTCCGGCTGAACGGGGGGTTCGTGCACACAGCCAGCTTGGAG
CGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCT
TCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAG
CGCACGAGGGAGCTTCCAGGGGAAACGCCTGGTATCTTTATAGTCTGTCCGGTTTTCG
CCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAA
AAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACAT
GTTCTTTCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCT
GATACCGCTCGCCGACCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGG
AAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTACACCGCATAT
ATGGTCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCG

CTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGC
GCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGG
GAGCTGCATGTGTCAGAGGTTTTACCGTCATACCGAAACGCGCGAGGCAGCTGCGGT
AAAGTTCATCAGCGTGGTTCGTGAAGCGATTACAGATGTCTGCCTGTTTCATCCGCGTCCA
GCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAA
GGGCGGTTTTTCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTTCATGG
GGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGA
ACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACCTGGCGGTATGGATGCGGGCGG
ACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTT
CCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGC
TGACTTCCGCGTTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTTCATGTTGTTGC
TCAGGTGCGAGACGTTTTGCAGCAGCAGTCGCTTACGTTTCGCTCGCGTATCGGTGATTC
ATTCTGTAACAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCTCAACGACAGGAGCA
CGATCATGCGCACCCGTGGGGCCGCTATGCCGGCGATAATGGCCTGCTTCTCGCGGAAA
CGTTTTGGTGGCGGACCAGTGACGAAGGCTTGAGCGAGGGCGTCAAGATTCCGAATAC
CGAAGCGACAGGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCTCCGCGAAAATGA
CCCAGAGCGCTGCCGGCACCTGTCTACGAGTTGCATGATAAAGAAGACAGTCATAAGT
GCGGGACGATAGTCATGCCCGCGCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCT
CAAGGGCATCGGTGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGT
TGCGCTCACTGCCCGTTTTCCAGTCGGGAAACCTGTGCGTGCAGCTGCATTAATGAATCG
GCCAACGCGCGGGGAGAGGGCGTTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTACC
AGTGAGACGGGCAACAGCTGATTGCCCTTACCAGCCTGGCCCTGAGAGAGTTGCAGCAA
GCGGTCCACGCTGGTTTTGCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCG
GGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCACTACCGAGATATCCGCACCAA
CGCGCAGCCCGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCA
ACCAGCATCGCAGTGGGAACGATGCCCTCATTACAGCATTTCATGGTTTTGTTGAAAACC
GGACATGGCACTCCAGTCGCCTTCCCGTTCGGCTATCGGCTGAATTTGATTGCGAGTGAG
ATATTTATGCCAGCCAGCCAGACGCGAGACGCGCCGAGACAGAACTTAATGGGCCCGCTA
ACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCAGTCGCGTACCG
TCTTCATGGGAGAAAATAAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAA
CGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCCTCCTGGTCATCCAGCGGAT
AGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAG
GCTTCGACGCCGCTTCGTTCTACCATCGACACCACCGCTGGCACCCAGTTGATCGGGC
CGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGC
AACGCCAATCAGCAACGACTGTTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGT
AATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTTCGCAGAAACGTGGCTGG
CCTGGTTCACCACGCGGGAAACGGTCTGATAAAGAGACACCGGCATACTCTGCGACATCG
TATAACGTTACTGGTTTCACATTCACCACCTGAATTGACTCTCTCCGGGCGTATCAT
GCCATACCGCGAAAGGTTTTGCGCCATTGATGGTGTCCGGGATCTCGACGCTCTCCCTT
ATGCGACTCCTGCATTAGGAAGCAGCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCG
CCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCC
TGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCGAAAGTGGCGAGCCCGATCTT
CCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATG
CCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGA
CTCACTATAGGGGAATTGTGAGCGGATAACAATTCCTCTCTAGAAATAATTTGTTAAC
TTTAAGAAGGAGATATACCATGCAACATCACCATCACCATCACGGCGAATTCACGACAA
ATCCTGGTGTATCCGCTTGGCAGGTCAACACAGCTTATACTGCGGGACAGTTGGTCACAT
ATAACGGCAAGACGTATAAATGTTTGCAGCCCCACACCTCCTTGGCAGGATGGGAACCA
TCCAACGTTCCCTGCCTTGTGGCAGCTTGGTTCTCAATTGTCTCAGGCTAAACCTCAAATT
CCGAAAGATAAATCAAAGTGGCAGGCTATATTGAAATTCCAGATGCTGATATTAAGA
ACCAGTATATCCAGGACCAGCAACACGCGAACAATTAATAGAGGTGTAAGCTTTGCAA
AAGAAAATCAATCACTAGATGATCAAAATATTTCTATTGACGGACACACTTTTCATTGACC
GTCCGAACTATCAATTTACAAATCTTAAAGCAGCCAAAAAAGGTAGTATGGTGTACTTT
AAAGTTGGTAATGAAACACGTAAGTATAAAATGACAAGTATAAGAAACGTTAAGCCAA
CAGCTGTAGAAGTTCTGGATGAACAAAAGGTAAAGATAAACAATTAACATTAATFACT
TGTGATGATTACAATGAAGAGACAGGCGTTTTGGGAAACACGTAATAATCTTTGTAGCTAC
AGAAGTCAAAGGATCTGGTTCAATATGCTATCCATGGTGGATCCTGACTCGAGCACC
ACCACCACCACCTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCT
GCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAG
GGGTTTTTGTGTAAGGAGGAACTATATCCGGAT

pTRB-W88E (including CTB(W88E))

The pTRB-W88E plasmid containing CTB(W88E) was derived from pSAB2.2 plasmid available produced by Dr James Ross (University of Leeds). This plasmid was used to express CTB(W88E) as prepared by Thomas Branson (University of Leeds). CTB(W88E) is shown in red text, the periplasmic targeting sequence purple text, and the SphI and PstI restriction sites used for sub-cloning by green and blue highlighting respectively.

CCGACACCATCGAATGGTGCAAAACCTTTTCGCGGTATGGCATGATAGCGCCCCGGAAGAG
AGTCAATTCAGGGTGGTGAATGTGAAACCAGTAACGTTATACGATGTCGCAGAGTATGC
CGGTGTCTCTTATCAGACCGTTTCCC GCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAA
AACGCGGGAAAAAGTGGAAGCGGCGATGGCGGAGCTGAATTACATTCCCAACCGCGTG
GCACAACAACCTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCCAGTCTGGC
CCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGGTG
CCAGCGTGGTGGTGTGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGGCGGT
GCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCA
GGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTC
TGACCAGACACCCATCAACAGTATTATTTCTCCCATGAAGACGGTACGCGACTGGGCG
TGGAGCATCTGGTCGATTGGGTACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAGT
TCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAATCAAATT
CAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCAT
GCAAATGCTGAATGAGGGCATCGTTCCTACTGCGATGCTGGTTGCCAACGATCAGATGG
CGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGGGATATTTTCG
GTAGTGGGATACGACGATACCGAAGACAGTCTATGTTATATCCCGCCGTTAACCCACAT
CAAACAGGATTTTCGCTGTGGGGCAAACCAGCGTGGACCGTTGCTGCAACTCTCTC
AGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAAC
CACCCTGGCGCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCA
GCTGGCAGCAGAGTTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTA
AGTTAGCTCACTCATTAGGCACAATTCTCATGTTTGACAGCTTATCATCGACTGCACGGT
GCACCAATGCTTCTGGCGTCAGGCAGCCATCGGAAGCTGTGGTATGGCTGTGCAGGTCG
TAAATCACTGCATAATTCGTGTGCTCAAGGCGCACTCCC GTTCTGGATAATGTTTTTTG
CGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTGACAATTAATCATC
GGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTACACAGGAAACAGCCAGT
CCGTTTAGGTGTTTTACGAGCAATTGACCAACAAGGACCATAGATTATGAGCTTTAAG
AAAATTATCAAGGCATTTGTTATCATGGCTGCTTTGGTATCTGTT CAGGCCTCACTCAACT
CCTCAAAATATTACTGATTTGTGCGCAGAATACCACAACACACAATATATACGCTAAA
TGATAAGATCTTTTCGTATACAGAATCGCTAGCGGGAAAAAGAGAGATGGCTATCATT
CTTTTAAAGATGGTGCAATTTTTCAAGTAGAGGTACCAGGTAGTCAACATATAGATTAC
AAAAAAAGCGATTGAAAGGATGAAGGATACCCTGAGGATTGCATATCTTACTGAAGCT
AAAGTCGAAAAGTTATGTGTAGAGAATAATAAAACGCCTCATGCGATCGCCGCAATTAG
TATGGCAAATAAGTTTTCCCTGCGAGGTAATTAATAAGCTTCAAATAAAACGAAAGGC
TCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTGCGGTGAACGCTCTCCTGAGT
AGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCCGGAGGGTGGC
GGG CAGGACGCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGAC
GGATGGCCTTTTTGCGTTTCTACAAACTCTTTCGGTCCGTTGTTTATTTTTCTAAATACAT
TCAAATATGATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAAATATTGAAA
AAGGAAGATAGTATTCAACATTTCCGTGTCGCCCTTATTCCTTTTTTTCGCGCAATT
TGCTTCTGTTTTGCTCACCCAGAAACGCTGGTGAAGTAAAAGATGCTGAAGATCA
GTTGGGTGCACGAGTGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGA
GTTTTCGCCCCGAAGAACGTTTCCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCG
CGGTATTATCCCGTGTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTC
AGAATGACTTGGTTGAGTACTACCAGTCACAGAAAAGCATCTTACGGATGGCATGACA
GTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAAACA CTGCGGCCAACTTACT
TCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTTCACAACATGGGGGATC
ATGTAACCTCGCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAG

CGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACCTATTAACCTGGCGA
 ACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTG
 CAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAG
 CCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCC
 CGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACA
 GATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACCTGTCAGACCAAGTTTACTC
 ATATATACTTTAGATTGATTTCCCTTAGGACTGAGCGTCAACCCCGTAGAAAAAGATCAAA
 GGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAACCA
 CCGCTACCAGCGGTGGTTTGTGTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTA
 ACTGGCTTCAGCAGAGCGCAGATAACAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGG
 CCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACC
 AGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGT
 TACCGGATAAAGGCGCAGCGGTCCGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTG
 GAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCA
 CGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCCGTAAGCGGCAGGGTTCGGAAGCAAG
 AGAGCGCACGAGGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCTGTCCGGT
 TTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTAT
 GGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTC
 ACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTG
 AGCTGATACCGCTCGCCGAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAA
 GCGGAAGAGCGCCTGATGCGGTATTTCTCCTTACGCATCTGTGCGGTATTTACACCCG
 ATATAAGGTGCACTGTGACTGGGTGATGGCTGCGCCCCGACACCCGCCAACACCCGCTG
 ACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCT
 CCGGGAGCTGCATGTGTCAGAGGTTTTACCGTCATCACCGAAACGCGCGAGGCAGCTG
 CGGTAAAGCTCATCAGCGTGGTCTGTCAGCGATTACAGATGTCTGCCTGTTTATCCGCG
 TCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATG
 TTAAGGGCGGTTTTTCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCA
 TGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTACTGATGAT
 GAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACCTGGCGGTATGGATGCGGCG
 GGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTG
 TTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGC
 GCTGACTTCCGCGTTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTGTTGTT
 GCTCAGGTGCGAGACGTTTTGACAGCAGCAGTCGCTTACGTTGCTCGCGTATCGGTGAT
 TCATTCTGTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGTCTCAACGACAGGAG
 CACGATCATGCGCACCCGTGGCCAGGACCCAACGCTGCCCCGAAATT

pET11a (including SAP-LPETGA (E167Q))

The pET11a plasmid containing SAP-LPETGA (E167Q) was designed by Bruce Turnbull (University of Leeds) and ordered from/synthesised by genscript. SAP-LPETGA (E167Q) is shown in red text and the NdeI and BamHI restriction sites used for sub-cloning by green and blue highlighting respectively.

TTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAAT
 AATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTG
 TTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAAT
 GCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTGCCCCCTAT
 TCCCTTTTTTGCGGCATTTTGCCTTCCCTGTTTTTGTCTACCCAGAAACGCTGGTGAAGTA
 AAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAG
 CGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAA
 AGTTCTGCTATGTGGCGCGGTATTATCCCGTGTGACGCCGGGCAAGAGCAACTCGGTC
 GCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTACCAGTACACAGAAAAGCAT
 CTTACGGATGGCATGACAGTAAGAGAATTATGTCAGTGTGCCATAACCATGAGTGATAA
 CACTCGGGCCACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTT
 TGCACAACATGGGGGATCATGTAACCTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAA
 GCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAACGTTGCG

CAAAC TATTA ACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGAT
GGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTAT
TGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGC
CAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATG
GATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACT
GTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACCTTCATTTTTAATTTAAA
AGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAAATCCCTTAACGTGAGTTT
TCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTT
TTTCTGCGCGTAATCTGCTGCTTGC AAAACAAAAAACCACCGCTACCAGCGGTGGTTTGT
TTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTA ACTGGCTTCAGCAGAGCGCA
GATACCAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGA ACTCTGT
AGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGA
TAAGTCGTGTCTTACCGGGTTGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGT
CGGGCTGAACGGGGGGTTCGTGCACACAGCCAGCTTGGAGCGAACGACCTACACCGA
ACTGAGATACCTACAGCGTGAGCTATGAGAAAAGCGCCACGCTTCCCGAAGGGGAGAAG
GCGGACAGGTATCCGGTAAGCGGCAGGGTTCGGAACAGGAGAGCGCACGAGGGAGCTTC
CAGGGGGAAACGCCTGGTATCTTTATAGTCTGTCGGGTTTCGCCACCTCTGACTTGAGC
GTTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCG
GCCTTTTTACGGTTCCTGGCCTTTTTGCTGGCCTTTTTGCTCACATGTTCTTTCCTGCGTTATC
CCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGAG
CCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGG
TATTTTCTCCTTACGCATCTGTGCGGTATTTACACCGCATATATGGTGC ACTCTCAGTAC
AATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACTACTCCGCTATCGCTACGTGACTGG
GTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCT
GCTCCCGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGA
GGTTTTACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAGCTCATCAGCGTGG
TCGTGAAGCGATTACAGATGTCTGCCTGTTATCCGCGTCCAGCTCGTTGAGTTTCTCC
AGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCTGT
TTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTTCATGGGGGTAATGATACCGATG
AAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGA
ACGTTGTGAGGGTAAACA ACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACT
CAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCA
GCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCA
GACTTTACGAAACACGGAAACCGAAGACCATTTCATGTTGTTGCTCAGGTTCGACAGCGTT
TTGCAGCAGCAGTCGCTTACGTTTCGCTCGCGTATCGGTGATTTCATTCTGCTAACCAGTA
AGGCAACCCCGCCAGCCTAGCCGGTTCCTCAACGACAGGAGCACGATGCGCACCCG
TGGCCAGGATCCCAACGCTGCCCGAGATGCGCCCGCTGCGGCTGCTGGAGATGGCGGACG
CGATCGGATATGTTCTGCAAGGGTTGGTTTGGCGCATTACAGTTCTCCGCAAGAAATTGAT
TGGCTCCAATTCTTGGAGTGGTGAATCCGTTAGCGAGGTGCCCGCGGCTTCCATTGAGGT
CGAGGTGGCCCGGCTCCATGCACCCGCGACGCAACGCGGGGAGGCAGACAAGGTATAGG
GCGGCGCCTACAATCCATGCCAACCCGTTCCATGTGCTCGCCGAGGCGGCATAAATCGC
CGTGACGATCAGCGGTCCAGTGATCGAAGTTAGGCTGGTAAGAGCCGCGAGCGATCCTT
GAAGCTGTCCCTGATGGTCGTCATCTACCTGCCTGGACAGCATGGCCTGCAACGCGGGC
ATCCCGATGCCGCCGGAAGCGAGAAGAATCATAATGGGGAAGGCCATCCAGCCTCGCGT
CGCGAACGCCAGCAAGACGTAGCCAGCGCGTCCGCCGCCATGCCGGCGATAATGGCCT
GCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGC
AAGATCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTC
CTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCTTACGAGTTGCATGATAAAGA
AGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCACCGGAAGGAGCTGACT
GGGTTGAAGGCTCTCAAGGGCATCGGTGAGATCCCGTGCCTAATGAGTGAGCTAACT
TACATTAATTGCGTTGCGCTCACTGCCCGCTTCCAGTCCGGAAACCTGTGCTGCCAGCT
GCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTTGCGTATTGGGCGCCAGGGTG
GTTTTTCTTTTACCAGTGAGACGGGCAACAGCTGATTGCCCTTACCAGCCTGGCCCTGA
GAGAGTTGCAGCAAGCGGTCCACGCTGGTTTTGCCCCAGCAGGCGAAAATCCTGTTTGAT
GGTGGTTAACGGCGGGATATAACATGAGCTGCTTCCGGTATCGTCGTATCCCACTACCGA
GATATCCGCACCAACGCGCAGCCCGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCA
TCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTACGATTTGCATGG
TTTGTGAAAACCGGACATGGCACTCCAGTGCCTTCCCGTTCCGCTATCGGCTGAATTT
GATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGACGCGCCGAGACAGA ACTT
AATGGGCCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCC

CAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGA
 CATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGG
 TCATCCAGCGGATAGTTAATGATCAGCCACTGACGCGTTGCGCGAGAAGATTGTGCAC
 CGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACC
 CAGTTGATCGGCGGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCA
 GACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACG
 CGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTTCGCA
 GAAACGTGGCTGGCCTGGTTCACCACGCGGAAACGGTCTGATAAGAGACACCGGCATA
 CTCTGCGACATCGTATAACGTTACTGGTTTTACATTACCACCCTGAATTGACTCTCTTCC
 GGGCGCTATCATGCCATAACCGCGAAAGTTTTGCGCCATTGATGGTGTCCGGGATCTCG
 ACGTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTT
 GAGCACCGCCCGCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCG
 GCCACGGGGCCTGCCACCATAACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCG
 AGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCCGACCTGTGG
 CGCCGGTATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGA
 AATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATA
 ATTTTGTTTAACTTTAAGAAGGAGATATACATATCAGCCACACCGACCTGAGCGGTAAA
 GTGTTTCGTTTTCCGCGTGAGAGCGTGACCGATCACGTTAACCTGATCACCCCGCTGGAA
 AAGCCGCTGCAGAACTTACCCTGTGCTTTCGTGCGTACAGCGACCTGAGCCGTGCGTAC
 AGCCTGTTACAGCTATAACACCCAAGGCCGTGATAACGAGCTGCTGGTGTATAAAGAGCG
 TGTTGGTGAATACAGCCTGTATATTGGCCGTCACAAGGTGACCAGCAAAGTTATCGAAA
 AGTTTTCCGCGCCGGTGCACATTTGCGTTAGCTGGGAGAGCAGCAGCGGTATCGCGGAA
 TTCTGGATTAACGGCACCCCGCTGGTGAAGAAAGGTCTGCGTCAGGGCTACTTTGTGGA
 GCGCAACCGAAAATCGTTCTGGGTCAGGAACAAGACAGCTATGGTGGCAAGTTTCGATC
 GTAGCCAGAGCTTTGTGGGCGAGATCGGCGACCTGTACATGTGGGATAGCGTTCTGCCG
 CCGCAGAACATTCTGAGCGCGTATCAAGGTACCCCGCTGCCGGCGAACATCCTGGACTG
 GCAAGCGCTGAACTACGAGATTCGTGGTTATGTTATCATTAAACCGCTGGTGTGGGTTGG
 TAGCGGTCTGCCGAAACCGGTGCGCACACCACCACCACCCTAAAGGATCCGGCTGCT
 AACAAAGCCCAGAAAGGAAGCTGAGTTGGCTGCTGCCACCCTGAGCAATAACTAGCATA
 ACCCTTGGGGCCTCTAAACGGGTCTTGGGGGTTTTTTGCTGAAAGGAGGAACTATATC
 CGGATATCCCGCAAGAGGCCCGGCAGTACCGGCATAACCAAGCCTATGCCTACAGCATC
 CAGGGTGACGGTGCCGAGGATGACGATGAGCGCATTTGTTAGATTTACATACCGGTGCCT
 GACTGCGTTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCTTATCGATGATAAGCT
 GTCAAACATGAGAA

pET-29b (including CRP-LPETGA & downstream kil peptide)

The pET-29b plasmid containing CRP-LPETGA was designed by Jonathan Dolan (University of Leeds) and ordered from/synthesised by Twist Bioscience. CRP-LPETGA is shown in red text, *E. coli* alkaline phosphatase signal peptide is shown in purple text. The NdeI and EcoRI restriction sites used for sub-cloning is highlighted in green and blue respectively. The sequence for the downstream *kil* peptide is shown in blue text. The Shine-Dalgarno sequence for the *kil* peptide is highlighted in yellow. The PstI and HindIII restriction sites used for sub-cloning is highlighted in pink and red respectively. The *kil* peptide was not isolated/purified and only used for the export of CRP-LPETGA from the periplasm to the media.

TAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATT
 TTGTTTAACTTTAAGAAGGAGATATACATATCAAGCAGTCAACCATCGCGTTAGCTTTAT
 TGCCTCTGCTTTTCACTCCCGTAACCAAGGCTCAAACCGATATGTCTCGCAAAGCCTTCG
 TTTTCCCGAAGGAATCAGACACCAGTTACGTCAGTCTTAAGGCCCTCTGACCAAACCAT
 TAAAGGCATTTACGGTTTGTTTACATTTCTATAACAGAGCTCAGCAGCACACGCGGATATT

CGATCTTTAGCTACGCGACTAAACGCCAGGATAACGAAATCTTGATCTTCTGGTCCAAA
GACATCGGCTATTCGTTACGGTAGGCGGCAGTGAGATTCTTTTCGAAGTCCCAGAGGTT
ACGGTTGCGCCGGTGCATATCTGCTACTAGTTGGGAATCTGCTTCTGGCATTGTTGAATTT
TGGGTTGACGGCAAACCGCGTGTACGCAAATCACTTAAGAAAGGTTATACCGTTGGCGC
CGAGGCGAGTATTATTTTAGGTCAAGAACAAGACAGTTTTGGCGGCAATTTTCGAGGGCT
CACAGAGCTTAGTCGGTGATATCGGGAACGTAAATATGTGGGATTTTCGTATTGAGTCCG
GACGAAATCAATACGATTTACTTGGGTGGTCCGTTTTCTCCAAACGTGTTGAATTGGCGT
GCCTTGAATACGAGGTTACGGGCGAGGTGTTTACAAAGCCGCAATTATGGCCGGGCAG
TGGCTTGCAGAGACAGGAGCCTCTCATACCACCACCACCCTAA**GAATTC**TGAAGGT
TACGA**CTGCAGAGGAGG**TTATACGTATGCGTAAGCGTTTCTTCGTTGGCATCTTTGCTAT
CAATCTGCTGGTAGGGTGCCAAGCCAATTACATCCCGGACGTGCAAGGCGGCACACTATTG
CGCCGAGTAGCAGTTCGAAACTTACCGGCATTGCAGTCCAATAG**AAGCTT**CTCGAGCAC
CACCACCACCACCCTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGC
TGCTGCCACCCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGA
GGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCTGTGA
GCGGGCATTAAAGCGCGGGGTGGTGGTTACGCGCAGCGTACCCGCTACACTTGC
AGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTTCGCCGGCT
TTCCCCGTCAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGC
ACCTCGACCCCAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGA
TAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCC
AACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGC
CGATTTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTA
ACAAAATATTAACGCTTACAATTTAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCC
CTATTTGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGA
AAAATCATCGAGCATCAAATGAACTGCAATTTATTCATATCAGGATTATCAATACCAT
ATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAATCACCGAGGCAGTTCATAGG
ATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATT
AATTTCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGA
ATCCGGTGAGAATGGCAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCC
ATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGC
CTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAA
TGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATA
TTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATC
ATCAGGAGTACGGATAAAAATGCTTGATGGTTCGGAAGAGGCATAAATTCGTCAGCCAGT
TAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAA
ACAATCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGA
CATTATCGCAGAGCCATTTATACCCATATAAAATCAGCATCCATGTTGGAATTTAACTCGG
GCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGATTAATCTGTTA
TGTAAGCAGACAGTTTTATTGTTTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCAC
TGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGC
GTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTGTTGCCGGA
TCAAGAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTACGACAGCGCAGATACCAA
ATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCCG
CTACATACTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGT
GTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCCGGGCTGA
ACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATA
CCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGG
TATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAA
ACGCTGGTATCTTTATAGTCCTGTGCGGTTTTCGCCACCTCTGACTTGAGCGTCGATTTTT
GTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAACGCCAGCAACCGGCCTTTTTAC
GGTTCCTGGCCTTTTGTGCTGCTTTTGTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCT
GTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGACCCGAACGAC
CGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTCTCC
TTACGCATCTGTGCGGTATTTACACCCGCAATGGTGCACCTCTCAGTACAATCTGCTCTGA
TGCCGCATAGTTAAGCCAGTATACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCG
CCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCC
GCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTGAGAGGTTTTACCCGTCA
TCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATT
ACAGATGTCTGCCTGTTTATCCGCGTCCAGCTCGTTGAGTTTTCTCCAGAAGCGTTAATGT
CTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCTGTTTGGTCACTGATGC

CTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGA
TGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGT
AAACAACCTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCC
AGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATG
CAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAAC
ACGGAACCGAAGACCATTTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGT
CGCTTACGTTTCGCTCGCGTATCGGTGATTTCATTCTGCTAACCCAGTAAGGCAACCCCGCC
AGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATG
CCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTTGGTGGCGGGACCAGTGACGAAGGC
TTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGC
TCCAGCGAAAGCGGTCTCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACG
AGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCA
CCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTGAGATCCCGGTGCCT
AATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGTTTTCCAGTCGGGAA
ACCTGTCGTGCCAGTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGTTTTGCGT
ATTGGGCGCCAGGGTGGTTTTCTTTTTACCAGTGAGACGGGCAACAGCTGATTGCCCTT
CACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTTGCCCCAGCAGGC
GAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGT
CGTATCCACTACCGAGATGTCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGC
ATTGCGCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTC
ATTCAGCATTTGCATGGTTTTGTTGAAAACCGGACATGGCACTCCAGTCGCTTCCCCTTC
CGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGAC
GCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGGATTTGCTGGTGACCCAATGCG
ACCAGATGCTCCACGCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGAT
GGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCA
CAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCACTGACGCGTTGC
GCGAGAAGATTGTGCACCGCCGCTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGAC
ACCACCAGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGA
CGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCG
CCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTT
TTTCCC CGTTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGAT
AAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTTACATTCACCACCC
TGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTGCA
TGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCAGT
AGTAGGTTGAGGCCGTTGAGCACCGCCGCGCAAGGAATGGTGCATGCAAGGAGATGG
CGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCACGCCGAAACAAGCGCTC
ATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCC
AGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCG
AGATCGATCTCGATCCCGCGAAAT

pET-29b (CRP-LPETGAS)

The pET-29b plasmid described in section 5.3.1 containing CRP-LPETGA was modified through 2 rounds of mutagenesis. The first round removed the downstream Shine Dalgarno ribosomal binding site and *kil* gene; the second removed the *E.coli* alkaline phosphatase periplasmic signal peptide. CRP-LPETGA is shown in red text, the NdeI and EcoRI restriction sites used for sub-cloning is highlighted in green and blue respectively.

TAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCTCTAGAAATAATT
TTGTTAACTTTAAGAAGGAGATATA **CATATG**CAAACCGATATGTCTCGCAAAGCCTTCG
TTTTCCCGAAGGAATCAGACACCCAGTTACGTCACTTAAAGGCCCTCTGACCAAACCAT
TAAAGGCATTTACGGTTTTGTTTACATTTCTATAACAGAGCTCAGCAGCACACGCGGATATT
CGATCTTTAGCTACGCGACTAAACGCCAGGATAACGAAATCTTGATCTTCTGGTCCAAA
GACATCGGCTATTCGTTACGGTAGGCGGCAGTGAGATTCTTTTTCGAAGTCCCAGAGGTT

ACGGTTGCGCCGGTGCATATCTGCACTAGTTGGGAATCTGCTTCTGGCATTGTTGAATTT
TGGGTTGACGGCAAACCGCGTGTACGCAAATCACTTAAGAAAGGTTATACCGTTGGCGC
CGAGGCGAGTATTATTTTAGGTCAAGAACAAGACAGTTTTGGCGGCAATTTTCGAGGGCT
CACAGAGCTTAGTCGGTATATCGGGAACGTAAATATGTGGGATTTTCGTATTGAGTCCG
GACGAAATCAATACGATTTACTTGGGTGGTCCGTTTTCTCAAACGTGTTGAATTGGCGT
GCCTTGAATACGAGGTTTCAGGGCGAGGTGTTTACAAAGCCGCAATTTAGGCCGGGCAG
TGGCTTGCCAGAGACAGGAGCCTCTCATACCACCACCACCCTAA**GAATTC**TGAAGGT
TACGACTGCAGAAGCTTCTCGAGCACCACCACCACCACCCTGAGATCCGGCTGCTAAC
AAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCCTGAGCAATAACTAGCATAACC
CCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAECTATATCCG
GATTGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAAGCGCGGCGGGTGTGGTGGTTA
CGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCC
CTTCTTTCTCGCCACGTTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTT
AGGGTTCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAACTTGATTAGGGTATG
GTTCCAGTAGTGGCCATCGCCCTGATAGACGGTTTTTTTCGCCCTTTGACGTTGGAGTCCA
GTTCTTTAATAGTGGACTCTTGTTCAAACTGGAACAACACTCAACCCTATCTCGGTCT
ATTCTTTGATTTATAAGGGATTTTGCCTATTTCGGCCTATTGGTTAAAAAATGAGCTGA
TTTAAACAAAAATTTAACGCGAATTTTAAACAAAATATTAACGCTTACAATTTAGGTGGCAC
TTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATAT
GTATCCGCTCATGAATTAATTCTTAGAAAACTCATCGAGCATCAAATGAAACTGCAATT
TATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTTCTGTAATGAAGGAG
AAAACCTACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCG
ACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAAATAAGGTTATCAAGT
GAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCT
TTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAAC
CAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAA
AAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATC
ACAATATTTTACCTGAATCAGGATATCTTCTAATACCTGGAATGCTGTTTTCCCGGG
GATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAAATGCTTGATGGTCG
GAAGAGGCATAAATCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGG
CAACGCTACCTTTGCCATGTTTCAGAAACAACCTCTGGCGCATCGGGCTTCCCATACAATC
GATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCATTTATACCCATATAAAT
CAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTTCCCGTTGAATATGGC
TCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTTCATGACCAAA
ATCCCTTAACGTGAGTTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGG
ATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAACAAAAAAACCACC
GCTACCAGCGGTGGTTTTGTTTGCCTGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAC
TGGCTTACAGAGCGCAGATACCAAAATACTGTCTTCTAGTGTAGTTAGTTAGGCC
ACCATTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAG
TGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTA
CCGGATAAGGCGCAGCGGTTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGG
AGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCAC
GCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTTCGGAACAGGA
GAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCTGTCGGGTT
TCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATG
GAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGTGCTGGCCTTTTGTCA
CATGTTCTTCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGA
GCTGATACCGCTCGCCGACGCCGAACGACCGAGCGCAGCGAGTCACTGAGCGAGGAAG
CGGAAGAGCGCCTGATGCGGTATTTCTCCTTACGCATCTGTGCGGTATTTACACCCGCA
ATGGTGCATCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCG
CTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGC
GCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCCTCTCCGG
GAGCTGCATGTGTCAGAGGTTTTACCGTCATACCGAAACGCGCGAGGCAGCTGCGGT
AAAGCTCATCAGCGTGGTCTGTAAGCGATTACAGATGTCTGCCTGTTTCATCCGCGTCCA
GCTCGTTGAGTTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAA
GGGCGTTTTTTCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTTCATGG
GGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGA
ACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACCTGGCGGTATGGATGCGGCGGG
ACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTT
CCACAGGGTAGCCAGCAGCATCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGC

TGACTTCCGCGTTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTTCATGTTGTTGC
TCAGGTTCGACAGCGTTTTGCAGCAGCAGTCGCTTCACGTTTCGCTCGCGTATCGGTGATTC
ATTCTGCTAACAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCA
CGATCATGCGCACCCGTGGGGCCGCCATGCCGCGGATAATGGCCTGCTTCTCGCCGAAA
CGTTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCGGAATAC
CGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCCTCGCCGAAAATGA
CCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGT
GCGGGCAGCATAGTCATGCCCCGCGCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCT
CAAGGGCATCGGTTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGT
TGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCG
GCCAACGCGCGGGGAGAGGGCGGTTTTGCGTATTGGGCGCCAGGGTGGTTTTTTCTTTTACC
AGTGAGACGGGCAACAGCTGATTGCCCTTACCAGCCTGGCCCTGAGAGAGTTGCAGCAA
GCGGTCCACGCTGGTTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCG
GGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATGTCCGACCAA
CGCGCAGCCCGACTCGGTAATGGCGCCATTGCGCCAGCGCCACTCTGATCGTTCGCA
ACCAGCATCGCAGTGGGAACGATGCGCCCTCATTAGCATTTCATGGTTTGTGAAAACC
GGACATGGCACTCCAGTCGCCTTCCCGTTCGCTATCGGCTGAATTTGATTGCGAGTGAG
ATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTA
ACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCAGTCGCGTACCG
TCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAA
CGCCGGAACATTAGTGACAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGAT
AGTTAATGATCAGCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAG
GCTTCGACGCCGCTTCGTTCTACCATCGACACCACCGCTGGCACCCAGTTGATCGGCG
CGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGC
AACGCCAATCAGCAACGACTGTTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGT
AATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTTCGACAGAAACGTGGCTGG
CCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCG
TATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTTCCGGGCGCTATCAT
GCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTT
ATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCG
CCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCC
TGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCCGATCTT
CCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATG
CCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCGATCTCGATCCCGCGAAAT

10.3 Protein Sequences

All DNA sequences are given in the 5' – 3' direction, beginning with the 5' restriction site and ending with the 3' restriction site. The coding sequence for the protein is shown with red text. All amino acid sequences are given in the N-terminal to C-terminal direction showing the expressed protein sequence without the periplasmic signal peptide (if appropriate).

CTB-LPETGA

DNA insert sequence

GCATGCAACTCCTCAAAATATTACTGATTTGTGCGCAGAATACCACAACACACAAATAT
ATACGCTAAATGATAAGATCTTTTCGTATACAGAATCGCTAGCGGGAAAAAGAGAGATG
GCTATCATTACTTTTAAGAATGGTGCAATTTTTCAAGTAGAGGTACCAGGTAGTCAACAT
ATAGATTCACAAAAAAGGCAATCGAACGTATGAAGGATACCCTGAGGATTGCATATCT
TACTGAAGCTAAAGTCGAAAAGTTATGTGTATGGAATAATAAAACGCCTCATGCGATCG
CCGCAATTAGTATGGCAAACAATGGCGGTAACCTGCCGAAACCGGTGCGTAAGTTTTTC
CCTGCAG

Expressed amino acid sequence

TPQNITDLCAEYHNTQIYTLNDKIFSYTESLAGKREMAIITFKNGAIFQVEVPGSQHIDSQKKA
IERMKDTRLRIAYLTEAKVEKLCVWNNKTPHAIAAISMANNGGNLPETGA

CTB-His₆

DNA insert sequence

CAATTGACCAACAAGGACCATAGATGACTCCTCAAAATATTACTGATTTGTGCGCAGAA
TACCACAACACACAAATATATACGCTAAATGATAAGATCTTTTCGTATACAGAATCGCT
AGCGGGAAAAAGAGAGATGGCTATCATTACTTTTAAGAATGGTGCAATTTTTCAAGTAG
AGGTACCAGGTAGTCAACATATAGATTCACAAAAAAGGCAATCGAACGTATGAAGGA
TACCCTGAGGATTGCATATCTTACTGAAGCTAAAGTCGAAAAGTTATGTGTATGGAATA
ATAAAACGCCTCATGCGATCGCCGCAATTAGTATGGCAAACAATGGCGGTAACCTGCCG
GAAACCGGTGCGAGCCACCATCATCATCATTAATAAGTTTTCCCTGCAG

Expressed amino acid sequence

MTPQNITDLCAEYHNTQIYTLNDKIFSYTESLAGKREMAIITFKNGAIFQVEVPGSQHIDSQKK
AIERMKDTRLRIAYLTEAKVEKLCVWNNKTPHAIAAISMANNGGNLPETGASHHHHHH

Sortase7M-His₆

DNA insert sequence

CATATGCAAGCTAAACCTCAAATTCGAAAGATAAATCAAAGTGCCAGGCTATATTGA
AATTCCAGATGCTGATATTAAGAACCAGTATATCCAGGACCAGCAACACGCGAACAAAT
TAAATAGAGGTGTAAGCTTTGCAAAAAGAAAATCAATCACTAGATGATCAAAATATTTCT
ATTGCAGGACACACTTTCATTGACCGTCCGAACTATCAATTTACAAATCTTAAAGCAGCC
AAAAAAGGTAGTATGGTGTACTTTAAAGTTGGTAATGAAACACGTAAGTATAAAATGAC
AAGTATAAGAAACGTTAAGCCAACAGCTGTAGAAGTTCTGGATGAACAAAAAGGTAAA
GATAAACAATTAACATTAATTACTTGTGATGATTACAATGAAGAGACAGGCGTTTGGGA
AACACGTAATCTTTGTAGCTACAGAAGTCAAACCTCGAGCACCACCACCACCAC

Expressed amino acid sequence

MQAKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATREQLNRGVSFAKENQSLDDQNISIAGHT
FIDRPNYQFTNLKAAKKGSMVYFKVGNETRKYKMTSIRNVKPTAVEVLDEQKGGKDKQLTLI
TCDDYNEETGVWETRKIFVATEVKLEHHHHHH

CBD-Sortase7M

DNA insert sequence

TCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGCAACATCACCATCACC
ATCACGGCGAATTCACGACAAATCCTGGTGTATCCGCTTGGCAGGTCAACACAGCTTAT
ACTGCGGGACAGTTGGTCACATATAACGGCAAGACGTATAAATGTTTGCAGCCCCACAC
CTCCTTGGCAGGATGGGAACCATCCAACGTTCTGCCTTGTGGCAGCTTGGTTCTCAATT
GTCTCAGGCTAAACCTCAAATCCGAAAGATAAATCAAAAAGTGGCAGGCTATATTGAAA
TTCCAGATGCTGATATTAAGAACCAGTATATCCAGGACCAGCAACACGCGAACAATTA
AATAGAGGTGTAAGCTTTGCAAAGAAAATCAATCACTAGATGATCAAAAATTTTCTAT
TGCAGGACACACTTTCATTGACCGTCCGAACTATCAATTTACAAATCTTAAAGCAGCCAA
AAAAGGTAGTATGGTGTACTTTAAAGTTGGTAATGAAACACGTAAGTATAAAATGACAA
GTATAAGAAACGTTAAGCCAACAGCTGTAGAAGTTCTGGATGAACAAAAAGGTAAAGA
TAAACAATTAACATTAATTACTTGTGATGATTACAATGAAGAGACAGGCGTTTGGGAAA
CACGTAAAATCTTTGTAGCTACAGAAGTCAAAGGATCTGGTTCACATATGGCTATCCATG
GTGGATCC

Expressed amino acid sequence

MQHHHHHHGEFTTNPGVSAWQVNTAYTAGQLVTYNGKTYKCLQPHTSLAGWEPSNVPAL
WQLGSQLSQAKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATREQLNRGVSFAKENQSLDDQ
NISIAGHTFIDRPNYQFTNLKAAKKGSMVYFKVGNETRKYKMTSIRNVKPTAVEVLDEQK
KDKQLTLITCDDYNEETGVWETRKIFVATEVKGSGSHMAIHGGS

Verotoxin (VTB)

DNA insert sequence

ATGAAAAACATTTATTAATAGCTGCATCGCTTTCATTTTTTTCAGCAAGTGCCTGGCG
ACGCTGATTGTGTAAGTGGAAAGGTGGAGTATACAAAATATAATGATGACGATACCTT
TACAGTTAAAGTGGGTGATAAAGAATTATTTACCAACAGATGGAATCTTCAGTCTCTTCT
TCTCAGTGCGCAAAATACGGGGATGACTGTAACCATTAATAAATAATGCCTGTCATAATG
GAGGGGGATTCAGCGAAGTTATTTTTCGTTGA

Expressed amino acid sequence

TPDCVTGKVEYTKYNDDDTFTVKVGDKELFTNRWNLQSLLSAQITGMTVTIKTNACHNGG
GFSEVIFR

CTB(W88E)

DNA insert sequence

GCATGCAACTCCTCAAATATTACTGATTTGTGCGCAGAATACCACAACACACAAATAT
ATACGCTAAATGATAAGATCTTTTCGTATACAGAATCGCTAGCGGGAAAAAGAGAGATG
GCTATCATTACTTTTAAAGAATGGTGCAATTTTTCAAGTAGAGGTACCAGGTAGTCAACAT
ATAGATTCACAAAAAAGCGATTGAAAGGATGAAGGATACCCTGAGGATTGCATATCT
TACTGAAGCTAAAGTCGAAAAGTTATGTGTAGAGAATAATAAAACGCCTCATGCGATCG
CCGCAATTAGTATGGCAAACCTAAGTTTTCCCTGCAG

Expressed amino acid sequence

TPQNITDLCAEYHNTQIYTLNDKIFSYTESLAGKREMAITFKNGAIFQVEVPGSQHIDSQKKA
IERMKDTRLRIAYLTEAKVEKLCVENNKTPHAI A ISMAN

SAP-LPETGA (E167Q)

DNA insert sequence

CATATGAGCCACACCGACCTGAGCGGTAAAGTGTTTCGTTTTTCCGCGTGAGAGCGTGAC
CGATCACGTAAACCTGATCACCCCGCTGAAAAAGCCGCTGCAGAACTTCACCCTGTGCTT
TCGTGCGTACAGCGACCTGAGCCGTGCGTACAGCCTGTTTTCAGCTATAACACCCAAGGCC
GTGATAACGAGCTGCTGGTGTATAAAGAGCGTGTGGTGAATACAGCCTGTATATTGGC
CGTCACAAGGTGACCAGCAAAGTTATCGAAAAGTTTCCGGCGCCGGTGCACATTTGCGT
TAGCTGGGAGAGCAGCAGCGGTATCGCGGAATTCTGGATTAACGGCACCCCGCTGGTGA
AGAAAGGTCTGCGTCAGGGCTACTTTGTGGAGGCGCAACCGAAAATCGTTCTGGGTGAG
GAACAAGACAGCTATGGTGGCAAGTTCGATCGTAGCCAGAGCTTTGTGGGCGAGATCGG
CGACCTGTACATGTGGGATAGCGTTCTGCCGCCGAGAACATTCTGAGCGCGTATCAAG
GTACCCCGCTGCCGGCGAACATCCTGGACTGGCAAGCGCTGAACTACGAGATTCTGGT
TATGTTATCATTAAACCGCTGGTGTGGGTTGGTAGCGGTCTGCCGAAACCGGTGCGCA
CCACCACCACCACCTAAGGATCC

Expressed amino acid sequence

MSHTDLSGKVFVFPRESVTDHVNLTIPLEKPLQNFTLCFRAYS DLSRAYSLFSYNTQGRDNEL
LVYKERVGEYSLYIGRHKVTSKVIKFPAPVHICVSWESSGIAEFWINGTPLVKKGLRQGYF
VEAQPKIVLGQEQDSYGGKFDQRSQSFVGEIGDLYMWDSVLPPQNILSAYQGTPLPANILDWQ
ALNYEIRGYVIIKPLVWVGSGLPETGAHHHHHH

CRP-LPETGAS (with PhoA periplasmic signal peptide)

DNA insert sequence

CATATGAAGCAGTCAACCATCGCGTTAGCTTTATTGCCTCTGCTTTTCACTCCCGTAACC
AAGGCTCAAACCGATATGTCTCGCAAAGCCTTCGTTTTCCCGAAGGAATCAGACACCAG
TTACGTCAAGTCTTAAGGCCCTCTGACCAAACCATTAAGGCATTTACGGTTTGTTTACA
TTTCTATACAGAGCTCAGCAGCACACGCGGATATTCGATCTTTAGCTACGCGACTAAACG
CCAGGATAACGAAATCTTGATCTTCTGGTCAAAGACATCGGCTATTCGTTACGGTAGG
CGGCAGTGAGATCTTTTTCGAAGTCCAGAGGTTACGGTTGCGCCGGTGCATATCTGCAC
TAGTTGGGAATCTGCTTCTGGCATTGTTGAATTTGGGTTGACGGCAAACCGCGTGTACG
CAAATCACTTAAGAAAGGTTATACCGTTGGCGCCGAGGGCGAGTATTATTTAGGTCAAG
AACAAAGACAGTTTGGCGGCAATTTTCGAGGGCTCACAGAGCTTAGTCGGTGATATCGGG
AACGTAAATATGTGGGATTTTCGTATTGAGTCCGGACGAAATCAATACGATTTACTTGGT
GGTCCGTTTTCTCAAACGTGTTGAATTGGCGTGCCTTGAAATACGAGGTTACGGGCGAG
GTGTTTACAAAGCCGCAATTATGGCCGGGAGTGGCTTCCAGAGACAGGAGCCTCTCA
TCACCACCACCACCTAAGAATTC

Expressed amino acid sequence

QTDMSRKAFVFPKESDTSYVSLKAPLTKPLKAFTVCLHFYTELSSTRGYSIFS YATK
RQDNEILIFWSKDIGYSFTVGGSEILFEVPEVTVAPVHICTSWESASGIVEFWVDGKP
RVRKSLKKG YTVGAEASIILGQEQDSFGGNFEQS QSLVGDIGNVNMWDFVLSPEI
NTIYLGGPFSPNVLNWRALKYEVQGEVFTKPQLWPGSGLPETGASHHHHHH

ColE1 release-lysis protein (Kil gene)

The ColE1 was not isolated/purified and only used for the export of CRP-LPETGA from the periplasm to the media. This was deleted from the plasmid used to express CRP-LPETGAS as it resulted in cell lysis and no soluble protein was recovered during its use. Its DNA and amino acid sequences are included here for future reference.

DNA insert sequence

CTGCAGAGGAGGTTATACGTATGCGTAAGCGTTTCTTCGTTGGCATCTTTGCTATCAATC
TGCTGGTAGGGTGCCAAGCCAATTACATCCCGGACGTGCAAGGCGGCACTATTGCGCCG
AGTAGCAGTTCGAAACTTACCGGCATTGCAGTCCAATAGAAGCTT

Expressed amino acid sequence

MRKRFFVGIFAINLLVGCQANYIPDVQGGTIAPSSSSKLTGIAVQ

CRP-LPETGAS (without PhoA periplasmic signal peptide)

DNA insert sequence

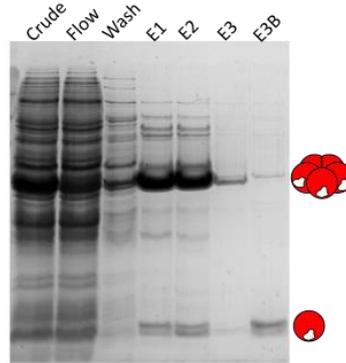
CATATGCAAACCGATATGTCTCGCAAAGCCTTCGTTTTCCCGAAGGAATCAGACACCAG
TTACGTCAGTCTTAAGGCCCTCTGACCAAACCATTAAGGCATTTACGGTTTGTTTACA
TTTCTATACAGAGCTCAGCAGCACACGCGGATATTCGATCTTTAGCTACGCGACTAAACG
CCAGGATAACGAAATCTTGATCTTCTGGTCCAAAGACATCGGCTATTCGTTACGGTAGG
CGGCAGTGAGATTCTTTTCGAAGTCCAGAGGTTACGGTTGCGCCGGTGCATATCTGCAC
TAGTTGGGAATCTGCTTCTGGCATTGTTGAATTTGGGTTGACGGCAAACCGCGTGTACG
CAAATCACTTAAGAAAGGTTATACCGTTGGCGCCGAGGCGAGTATTATTTTAGGTCAAG
ACAAGACAGTTTGGCGGCAATTTGAGGGCTCACAGAGCTTAGTCGGTGATATCGGG
AACGTAAATATGTGGGATTTTCGTATTGAGTCCGGACGAAATCAATACGATTTACTTGGGT
GGTCCGTTTTCTCAAACGTGTTGAATTGGCGTGCCTTGAAATACGAGGTTACGGGCGAG
GTGTTTACAAAGCCGCAATTATGGCCGGCAGTGGCTTGCCAGAGACAGGAGCCTCTCA
TCACCACCACCACCTAAGAATTC

Expressed amino acid sequence

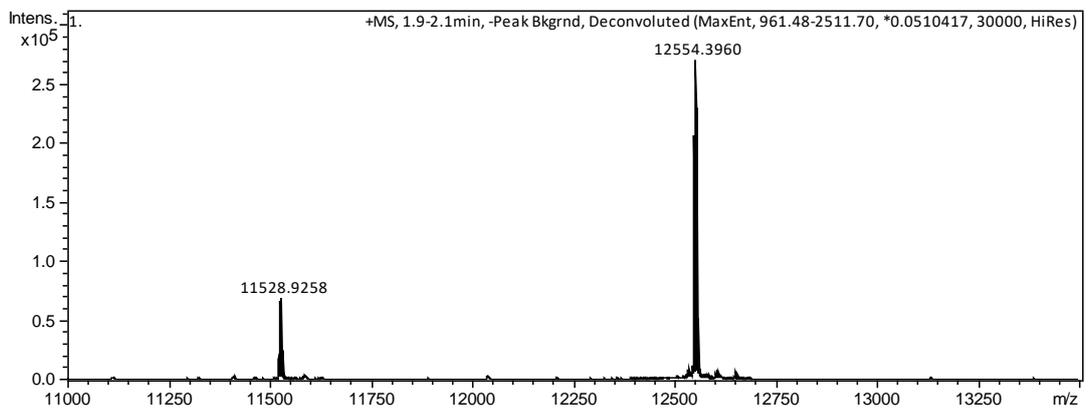
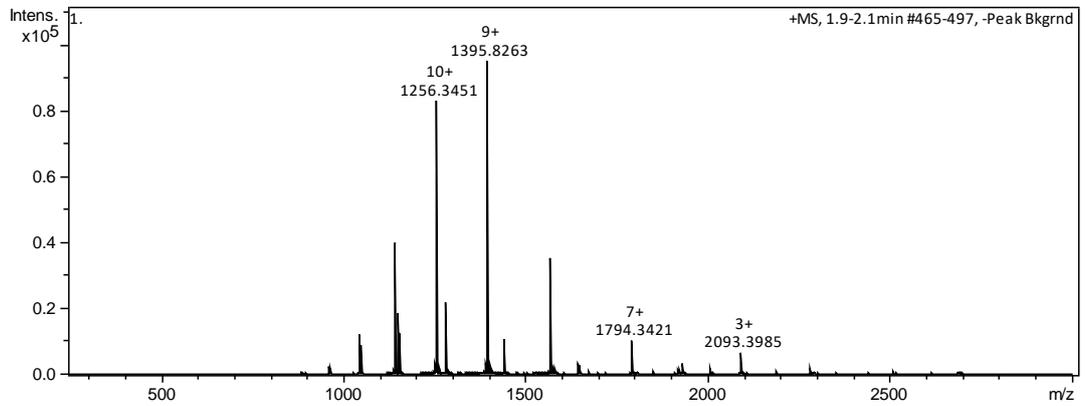
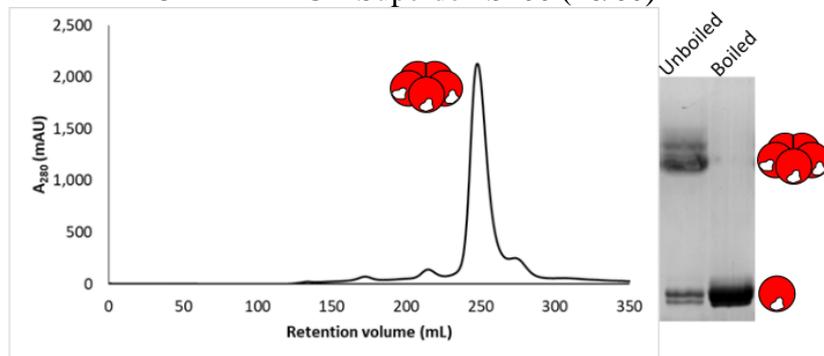
MQTDMSRKAFVFPKESDTSYVSLKAPLTKPLKAFTVCLHFYTELSSTRGYSIFS YATKRQDN
EILIFWSKDIGYSFTVGGSEILFEVPEVTVAPVHICTSWESASGIVEFWVDGKPRVRKSLKKG
TVGAEASIILGQEQDSFGGNFEGSQSLVGDIGNVMWDFVLSPEINTIYLGPFSPNVLNWR
ALKYEVQGEVFTKPQLWPGSGLPETGASHHHHHH

10.4 Protein Purification

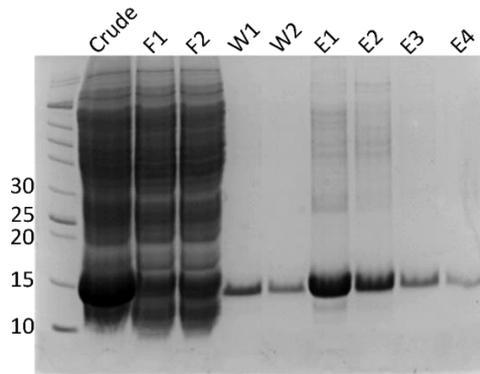
CTB-LPETGA



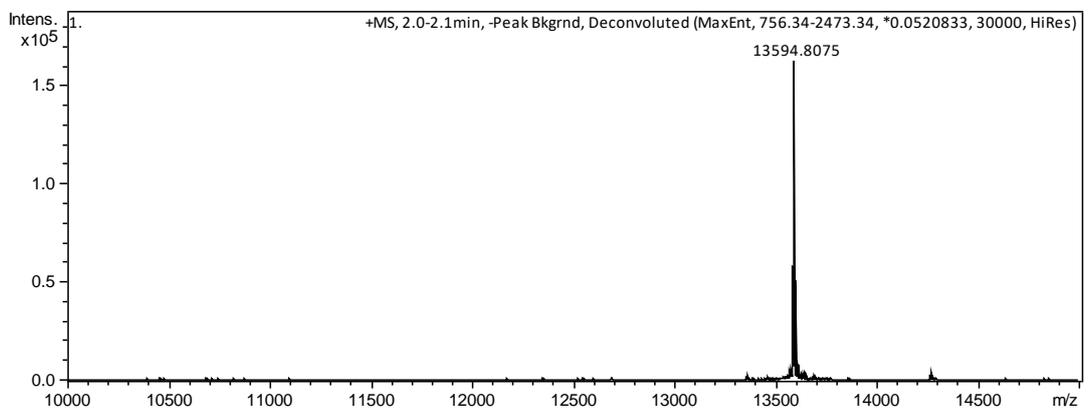
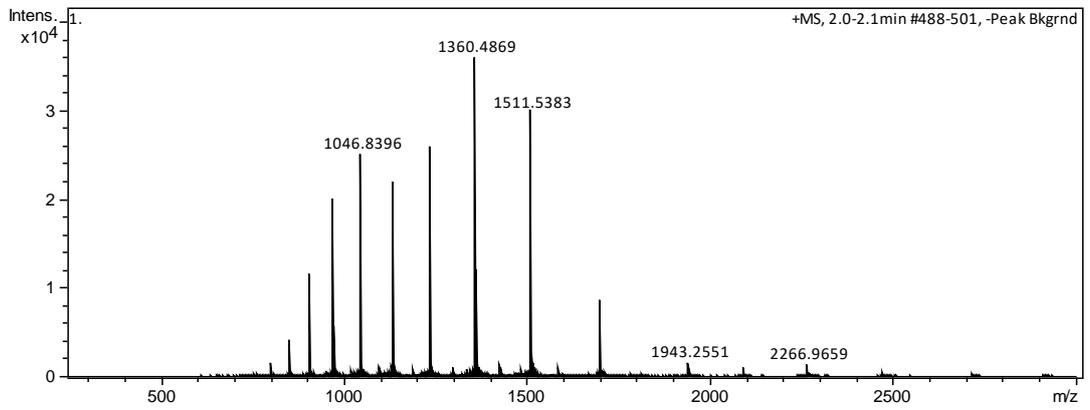
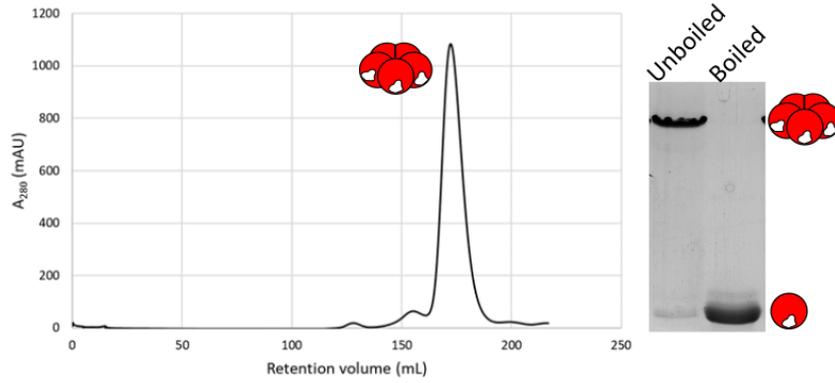
CTB-LPETGA Superdex S200 (26/60)



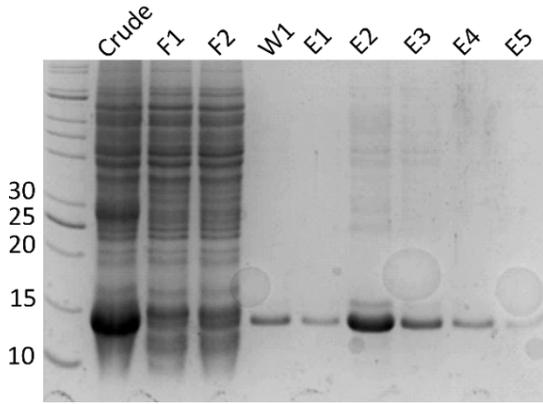
CTB-His₆



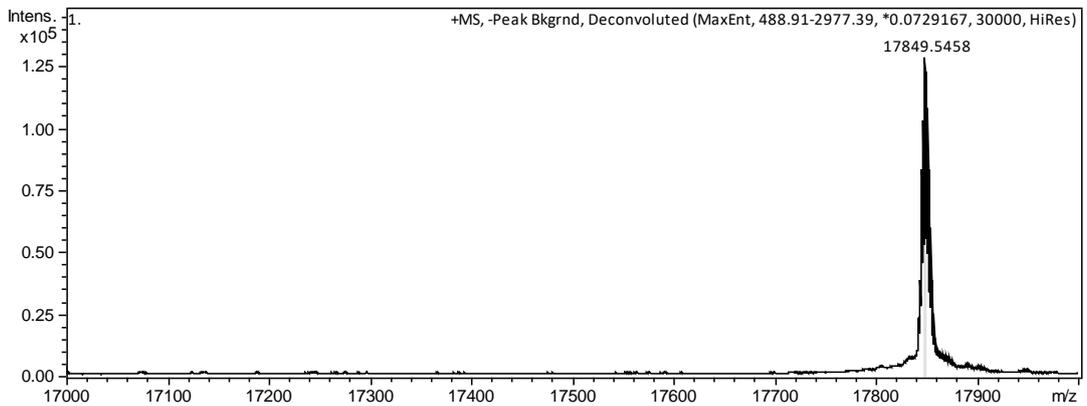
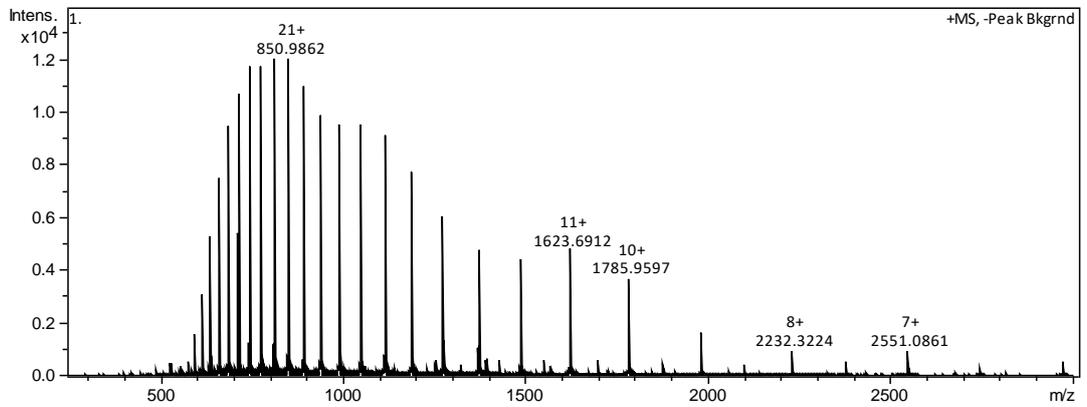
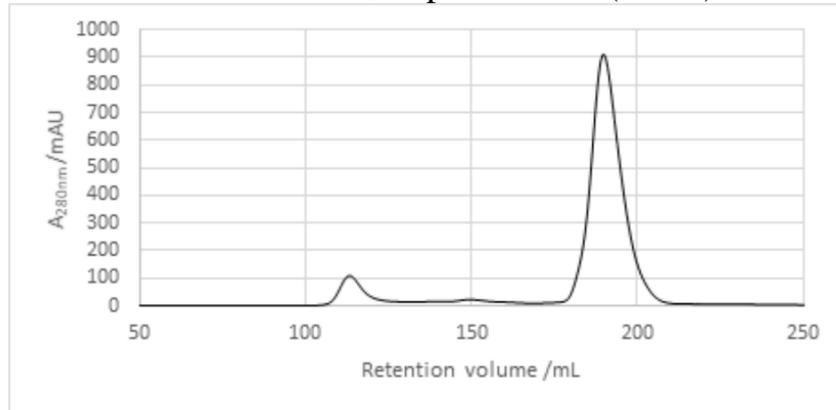
CTB-His₆ Superdex S75 (26/60)



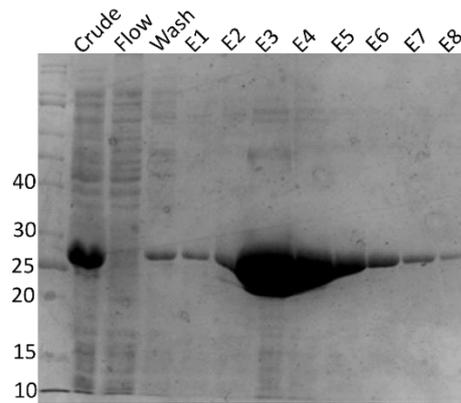
Sortase7M-His₆



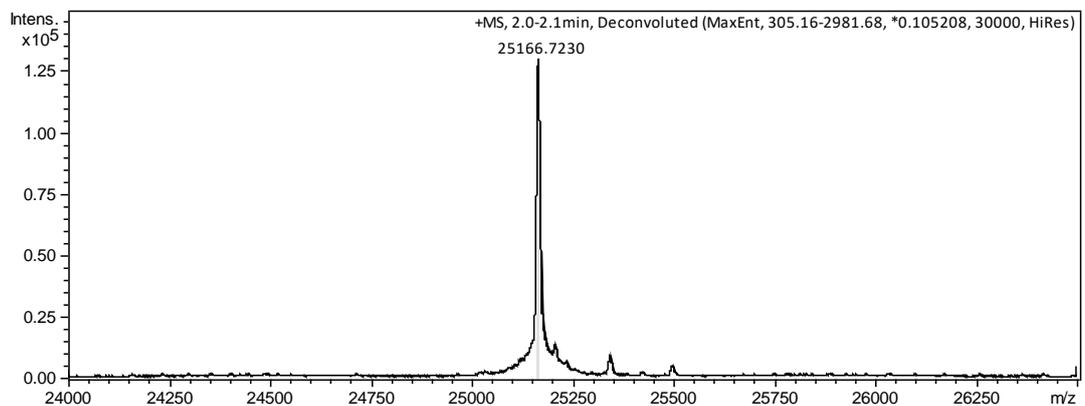
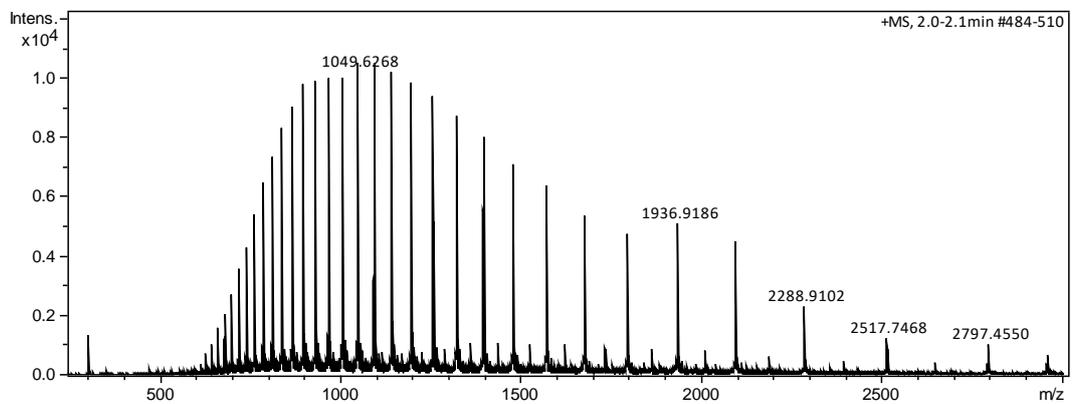
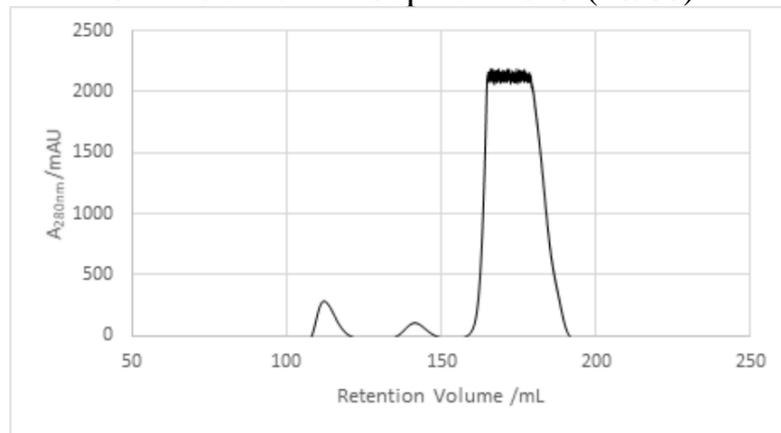
Sortase7M-His₆ Superdex S75 (26/60)



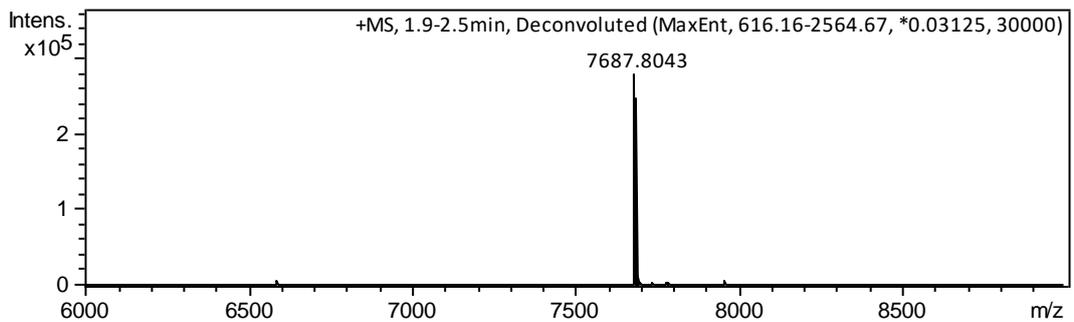
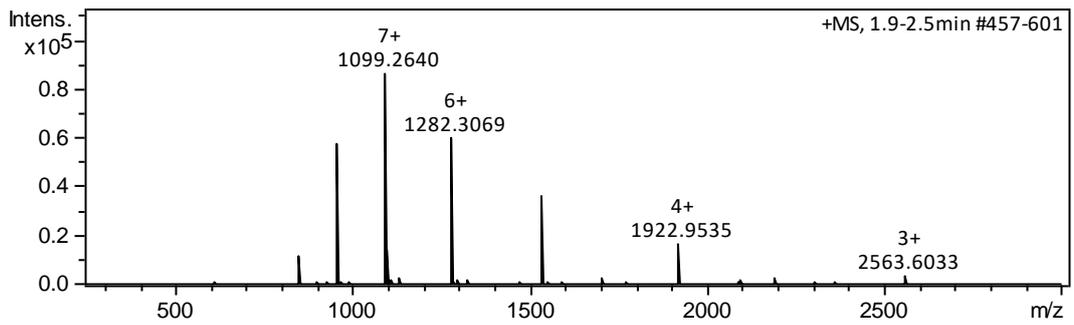
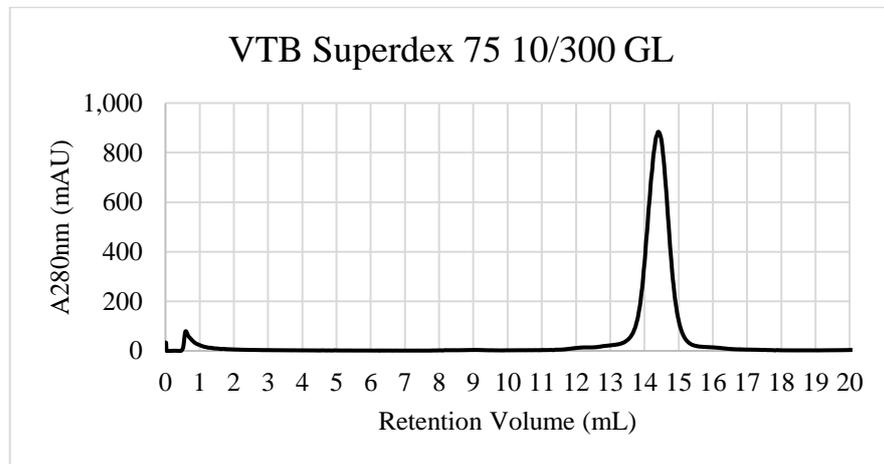
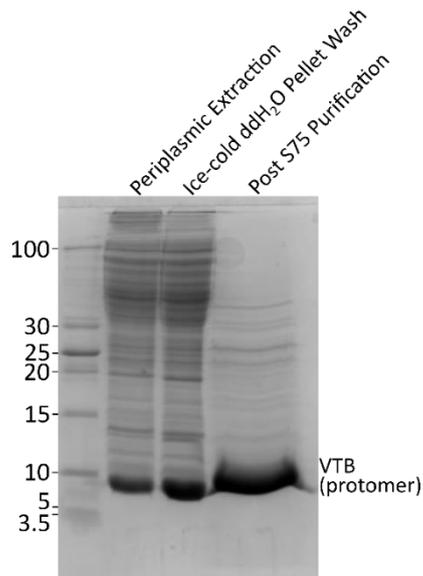
CBD-Sortase7M



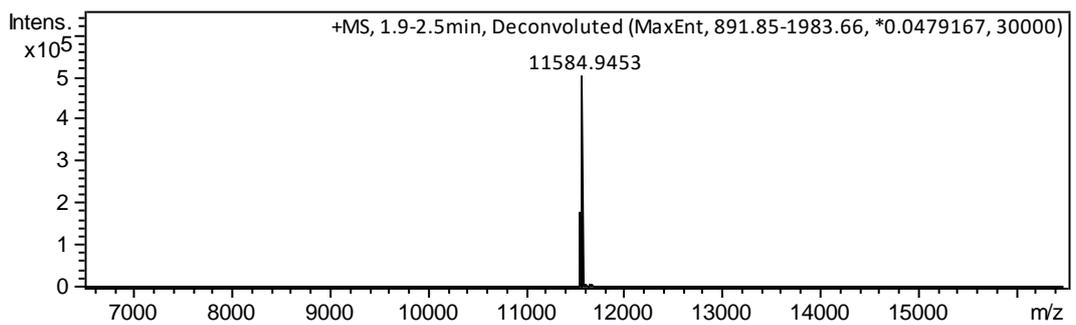
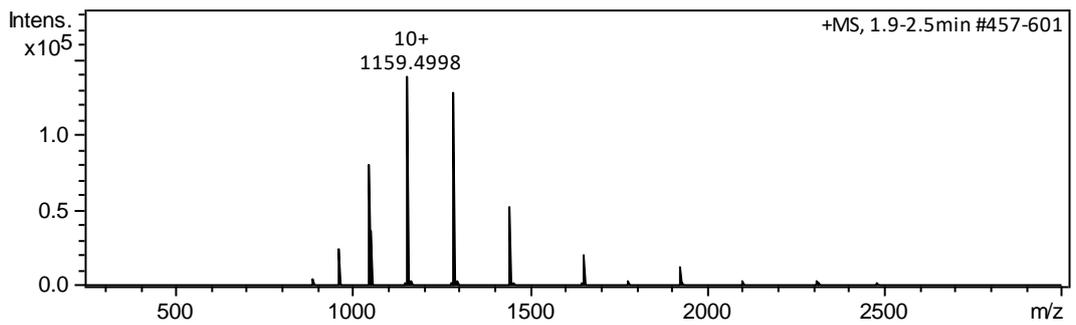
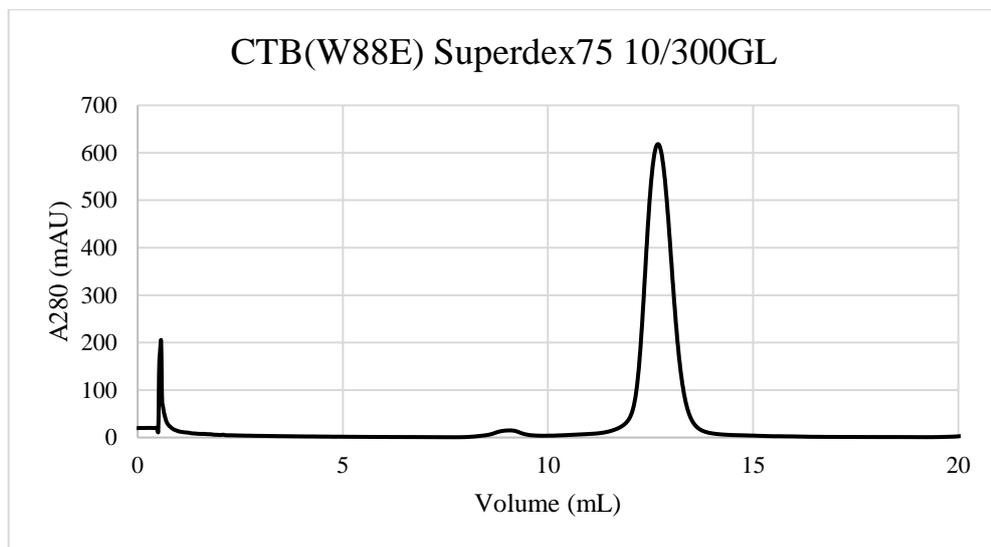
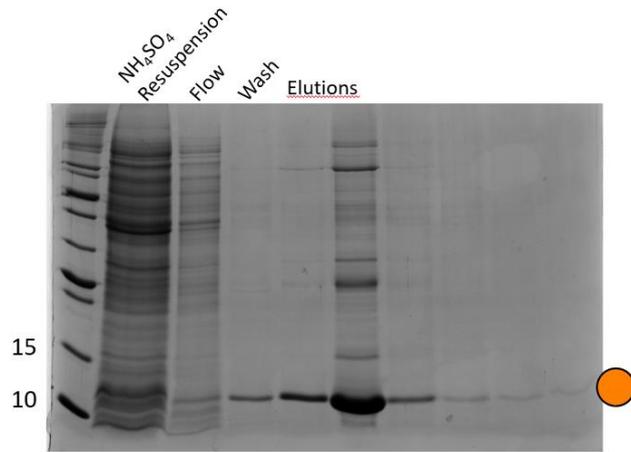
CBD-Sortase7M Superdex S75 (26/60)



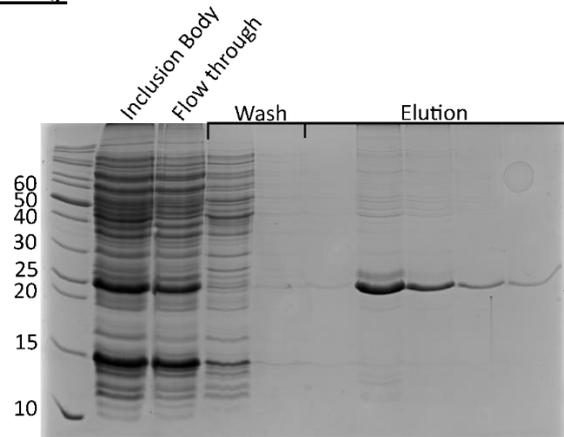
Verotoxin (VTB)



CTB(W88E)

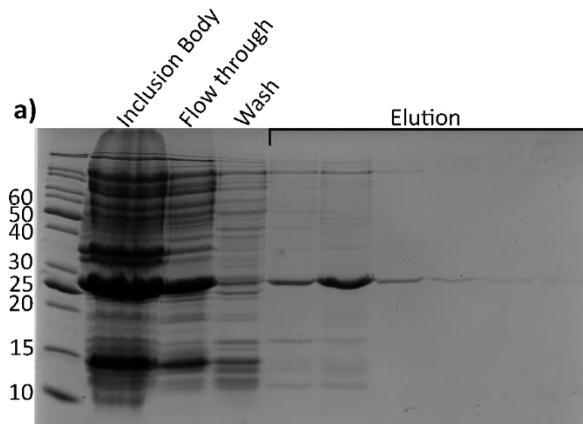


SAP-LPETGA (E167Q)

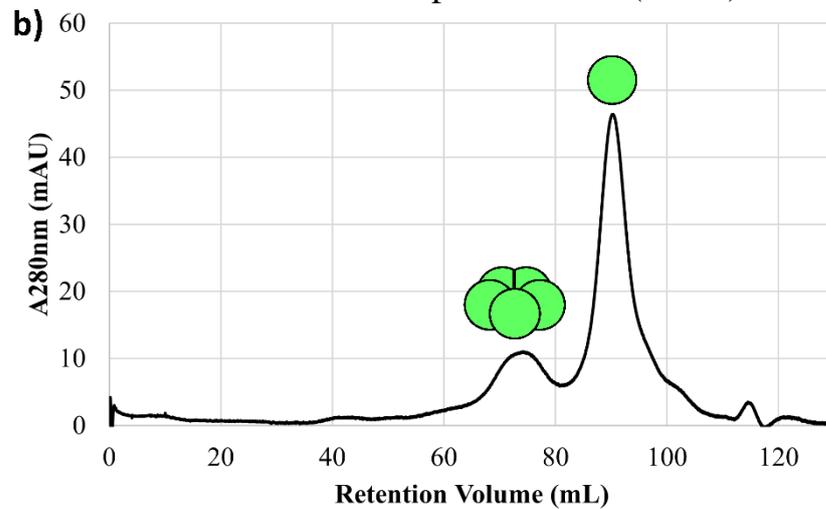


*It was not possible to refold Serum amyloid P (SAP)-LPETGAS in the course of this project.

CRP-LPETGAS



CRP-LPETGAS Superdex S200 (26/60)



10.5 Primers

Deletion of downstream Shine Dalgarno from CRP-LPETGAS

Primer Name	Sequence	Annealing Temperature
pet29b Del-SD-F1	5' - TTA TAC GTA TGC GTA AGC G - 3'	60 °C
pet29b Del-SD/Kil-R1	5' - CTG CAG TCG TAA CCT TCA G - 3'	

Deletion of downstream Shine Dalgarno & Kil gene from CRP-LPETGAS

Primer Name	Sequence	Annealing Temperature
pet29b Del-Kil-F1	5' - AAG CTT CTC GAG CAC CAC - 3'	64 °C
pet29b Del-SD/Kil-R1	5' - CTG CAG TCG TAA CCT TCA G - 3'	

Deletion of E.Coli alkaline phosphatase periplasmic signal sequence from CRP-LPETGAS

Primer Name	Sequence	Annealing Temperature
pet29b Del-PeriSignal-F	5' - CAA ACC GAT ATG TCT CGC - 3'	56 °C
pet29b Del-PeriSignal-R	5' - CAT ATG TAT ATC TCC TTC TTA AAG - 3'	