

Engineering Synthetic Glycoproteins

Ryan Carl M^cBerney

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

School of Chemistry

University of Leeds

June 2020

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others. 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.

© 2020 University of Leeds and Ryan Carl M^cBerney

To Dad

No longer here to read the pages, but still behind every word.

For Ben

'The beautiful thing about learning is nobody can take it away from you'

Acknowledgements

I must begin with a huge thank you to my family in particular my Mum, who have provided never ending support throughout my education and life. A special thank you to Vikki, for sharing in highs and lows, and for her patience in the months I have spent with my head buried in a computer.

I have many people to thank for making my time in Lab 1.49 the joy it has been. Starting at Leeds would not have been so seamless if not for Jack, it is difficult to summarise in just a few words three years of memories - from day one to the day you packed up - from Leeds to Amsterdam it's been nothing but a ball. To my desk buddy Sarah, our mental arithmetic may have suffered greatly from us forever confusing one another but at least we pioneered the reforming of stationary housing. The weekly trips to Nero for coffee and a vent with Holly and Devon will be sadly missed. From the lonely early starts with a large cafetiere, to late nights spent in the office with Jonny, Jack W and Pablo and everyone in between; who made the final months in the lab before COVID brought an abrupt end my time in Leeds so enjoyable. You all made the drive over the M62 worth it.

I would like to thank Matt, Zoe and Gemma for all their help in learning all things biochemistry, and to Kristian, who I took most of my lab problems from sugars to enzymes he was always on hand; despite many conversations being wound up with "I don't know if that was particularly useful" yet still always was.

Finally, a massive thank you to Bruce and Mike for their support, guidance and patience with all things in and outside of the lab, without which the last three and a half years would have been immensely more difficult. A constant open door, enthusiasm to help and willingness to spend their time (a precious commodity) discussing my work I could not be more thankful for.

Abstract

Post-translational glycosylation of proteins is an extremely complex process, involving large enzymatic cascades leading to many heterogeneous glycoforms. Performing glycosylation using chemical, or chemoenzymatic processes to mimic natural glycoproteins or create synthetic neoglycoproteins, is a challenging synthetic task. Despite the challenges in their production, homogeneous glycoproteins have many potential applications from study of glycobiology to potential therapeutics. This work aims to advance the chemical tools available to produce synthetic homogeneous glycoproteins, using bioorthogonal chemistry. A bifunctional, biorthogonal linker has been developed which combines oxime ligation and strain-promoted azide-alkyne cycloaddition chemistry to functionalise reducing sugars and glycan derivatives for protein attachment. This linker has been used to produce derivatives of the simple and complex oligosaccharides lactose and GM1, and in combination with enzymatic synthetic approaches, the lactose disaccharide has been extended to produce derivatives trisaccharide Gb3.

GM1 and Gb3 are known ligands of the bacterial enterotoxins, cholera toxin and verotoxin, produced by *Vibrio Cholerae* and *Escherichia coli* respectively. The chemically reactive derivatives of these ligands have been used to develop neoglycoprotein inhibitors of these toxins, based upon mutant pentameric protein scaffolds engineered from the cholera toxin B-subunit; excellent inhibition of the cholera toxin B-subunit has been observed for the resulting neoglycoprotein decorated with GM1 ligands with a picomolar IC_{50} , determined by ELLA inhibition studies.

Abbreviations

ALO	Aryl-lesscyclooctyne
Asc	Ascorbate
BARAC	Bisarylazacyclooctyne
BCN	Bicyclononyne
CAYE	Casamino yeast extract
CBD	Carbohydrate binding domain
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
DIBAC	Dibenzoanulatedazacyclooctyne
DIBO	Dibenzoanulatedcyclooctyne
DIFO	Difluorocyclooctyne
DIMAC	Dimethoxyazacyclooctyne
DIPEA	Diisopropylethylamine
DLS	Dynamic light scattering
DMC	2-Chloro-1,3-dimethylimidazolium
DMF	Dimethylformamide
DOPE	1,2-dioleoyl-sn-glycero-3-phosphoethanolamine
EDTA	Ethylenediaminetetraacetic acid
EGCase II	Endoglycoceramidase II
ELLA	Enzyme-linked lectin assay
FGE	Formylglycine-generating enzyme
FPLC	Fast protein liquid chromatography
GAG	Glycosaminoglycans
Gal	Galactose
Glc	Glucose
GPI	Glycosylphosphatidylinositol
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
HILIC	Hydrophilic interaction liquid
HIV	Human Immunodeficiency Virus
HRMS	High resolution mass spectroscopy
HRP	Horseradish peroxidase
HUS	Haemolytic ureamic syndrome
IEDDA	Inverse electron-demand Diels Alder
IPA	Isopropyl alcohol
IPTG	Isopropyl- β -thiogalactopyranoside
ITC	Isothermal titration calorimetry

Lac	Lactose
LB	Luria Broth
LCMS	Liquid chromatography mass spectroscopy
LDA	Lithium diisopropylamine
LG	Leaving group
LT	Heat-labile toxin
LTB	Heat-labile toxin B subunit
Man	Mannose
MOFO	Monofluorinatedcyclooctyne
MS	Mass spectroscopy
MSH	(O-(mesitylsulfonyl)hydroxylamine)
NCAA	Non-canonical amino acid
NHS	N-Hydroxy succinimide
NOFO	Non-fluorocyclooctyne
NMM	New minimal medium
NMO	N-methylmorpholine
NMR	Nuclear magnetic resonance
PAN	<i>para</i> -anisidine
PAN-C	5-methoxyanthranilic acid
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDA	<i>para</i> -phenylenediamine
PTM	Post-translational modification
SAP	Serum amyloid P
SDM	Site directed mutagenesis
SDS	Sodium dodecyl sulfate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SPAAC	Strain-promoted azide-alkyne cycloaddition
SPR	Surface plasmon resonance
ST	Shiga toxin
STEC	Shiga toxin-producing <i>E. coli</i>
SuSy	Sucrose synthase
TAE	Tris, Acetic acid, EDTA
TBDMS	Tert-butyl dimethyl silane
TBS	Tris buffered saline
TEA	Triethylamine
TEMED	Tertramethylethylenediamine
TFIB	Tri(trifluoroacetoxy)iodobenzene
THF	Tetrahydrofuran
TLC	Thin layer chromatography
Tris	Trisaminomethane
TYP	Tryptone yeast phosphate
UDP	Uridine diphosphate

VT	Verotoxin
VTB	Verotoxin B subunit
WHO	World Health Organisation
WT	Wild type

ACKNOWLEDGEMENTS	II
ABSTRACT	III
ABBREVIATIONS.....	IV
CHAPTER 1 AN INTRODUCTION	1
1.1 GLYCOSYLATION	2
1.1.1 <i>N-linked glycoproteins</i>	3
1.1.2 <i>O-glycosylation</i>	4
1.1.3 <i>Other glycosylated structures</i>	5
1.2 ROLES OF GLYCOSYLATION AND LECTINS IN DISEASE.....	5
1.2.1 <i>Pathogens which exploit cell surface carbohydrates</i>	5
1.2.2 <i>Bacterial toxins bind cell surface gangliosides</i>	6
1.3 MULTIVALENT INHIBITORS AGAINST BACTERIAL ADHESION AND INFECTION	10
1.3.1 <i>Glycopolymer-based inhibitors of bacterial lectins</i>	12
1.3.2 <i>Glycodendrimer-based inhibitors</i>	12
1.3.3 <i>Cyclic multivalent inhibitors</i>	13
1.3.4 <i>Protein templated inhibitor design</i>	15
1.4 SYNTHETIC GLYCOPROTEINS	19
1.4.1 <i>Controlling the glycosylation of macromolecules</i>	19
1.4.2 <i>Chemical glycosylation exploiting natural amino acid side chains</i>	19
1.4.3 <i>Chemical mutagenesis and non-canonical amino acid targets for glycosylation</i>	21
1.5 RATIONALE FOR THE PROJECT AND OBJECTIVES.....	29
1.5.1 <i>Crosslinking agents for protein conjugation</i>	31
1.5.2 <i>Amine reactive linkers</i>	32
1.5.3 <i>Thiol-reactive linkers</i>	32
1.5.4 <i>Oxime and hydrazine ligation of reducing sugars</i>	35
1.6 COMBINING FUNCTIONAL GROUPS FOR A NEW BIFUNCTIONAL LINKER	37
1.6.1 <i>Choosing a strained alkyne scaffold</i>	39
1.6.2 <i>Target bi-functional linkers</i>	40
1.7 WHAT FOLLOWS?	ERROR! BOOKMARK NOT DEFINED.
CHAPTER 2 NEW TOOLS FOR SYNTHETIC GLYCOSYLATION.....	43
2.1 LITERATURE SYNTHESIS OF BICYCLONONYNES	44
2.2 DEVELOPING SYN-LEAVING GROUPS TOWARDS A NOVEL STRAINED ALKYNE SYNTHESIS.....	45
2.2.1 <i>Retrosynthetic design of a novel synthesis</i>	46
2.2.2 <i>Synthesis of syn-ditriflate bicyclic ester</i>	46
2.2.3 <i>Screening for new elimination conditions</i>	47
2.2.4 <i>Literature synthesis maintaining the ester</i>	48

2.2.5	<i>Replacement of the ester with a protected alcohol</i>	49
2.3	FUNCTIONAL GROUP INTERCONVERSION OF ALCOHOL TO OXYAMINES	51
2.3.1	<i>Installing protected oxyamines by Mitsunobu chemistry</i>	51
2.3.2	<i>Deprotection to give a primary oxyamine</i>	52
2.3.3	<i>An alternative deprotection strategy for production of alkyl oxyamines</i>	53
2.4	CONCLUSION	56
CHAPTER 3 BIOORTHOGONALLY REACTIVE OLIGOSACCHARIDES		57
CHEMICAL AND ENZYMATIC FUNCTIONALISATION OF REDUCING SUGARS		57
3.1	SPAAC REACTIVE GLYCOSIDES FROM REDUCING SUGARS	58
3.1.1	<i>Optimisation of oxime ligation reactions using disaccharides</i>	58
3.1.2	<i>Purification of BCN derivatised glycans from aqueous media</i>	65
3.2	OXIME LIGATION TO PENTASACCHARIDE GM1OS	67
3.2.1	<i>Expression of EGCase II and enzymatic hydrolysis of GM1 ceramide</i>	67
3.3	SYNTHESIS OF GLYCOSYL AZIDES AND FUNCTIONALISATION THROUGH SPAAC	70
3.3.1	<i>Synthesis of lactosyl azide</i>	71
3.3.2	<i>Synthesis of GM1-N₃</i>	74
3.4	ENZYMATIC EXTENSION OF UN-NATURAL GLYCAN DERIVATIVES.....	77
3.4.1	<i>Expression of α-1,4 galactosyltransferase</i>	78
3.4.2	<i>Testing enzymatic activity of LgtC</i>	79
3.4.3	<i>Enzymatic cycle for continuous glycosyl donor production</i>	80
3.5	CONCLUSION	84
CHAPTER 4 BIOORTHOGONAL SYNTHETIC GLYCOSYLATION		86
4.1	DESIGN AND EXPRESSION OF BIOORTHOGONALLY REACTIVE CTB MUTANTS	89
4.1.1	<i>Non-binding CTB mutant for oxime ligations</i>	89
4.1.2	<i>Expression of the non-binding CTB W88E mutant</i>	89
4.1.3	<i>Generation of an azido-CTB protein scaffold for use in SPAAC ligation</i>	91
4.1.4	<i>SDS-PAGE comparison of CTB mutants</i>	96
4.2	SITE-SPECIFIC GLYCOSYLATION OF CTB-W88E BY OXIME LIGATION	97
4.2.1	<i>N-terminal oxidation of CTB-W88E</i>	98
4.2.2	<i>Synthesis of lactose neoglycoproteins by N-terminal oxime ligation to W88E</i>	100
4.3	SPAAC AS A NOVEL METHOD OF SITE-SPECIFIC GLYCOSYLATION	101
4.3.1	<i>Optimisation of SPAAC glycosylation using lactosyl BCN</i>	101
4.4	A NOVEL NEOGLYCOPROTEIN INHIBITOR OF CTB CREATED BY SPAAC GLYCOSYLATION	106
4.5	NEOGLYCOPROTEIN INHIBITION STUDIES BY ENZYME-LINKED LECTIN ASSAY (ELLA).....	109
4.6	CONCLUSION	113
CHAPTER 5 MULTIVALENT NEOGLYCOPROTEINS INHIBITORS OF THE VEROTOXIN.....		115
5.1	GENERATION OF NEOGLYCOPROTEIN INHIBITORS AGAINST THE VEROTOXIN	116

5.1.1	<i>Generation of CTB-Gb3 neo-glycoproteins by oxime ligation</i>	117
5.1.2	<i>Generation of CTB-Gb3 neo-glycoproteins by SPAAC glycosylation</i>	119
5.2	EXPRESSION OF RECOMBINANT VTB	120
5.2.1	<i>Expression tests</i>	120
5.2.2	<i>VTB purification</i>	122
5.2.3	<i>Summary of VTB expression</i>	128
5.3	DEVELOPMENT OF ELLA ASSAY FOR TESTING OF INHIBITORS AGAINST THE VEROTOXIN	129
5.3.1	<i>VTB-HRP conjugation</i>	129
5.3.2	<i>Synthesis of Gb3 glycolipids</i>	129
5.4	VTB CAPTURE ON MICROTITER PLATES	131
5.5	CONCLUSION	133
CHAPTER 6 TOWARDS NON-BACTERIAL PENTAMERIC PROTEIN SCAFFOLDS FOR MULTIVALENT PROTEIN-BASED THERAPEUTICS		135
6.1	SERUM AMYLOID P PROTEIN A HUMAN PROTEIN SCAFFOLD	137
6.2	DESIGN OF A SAP-BASED SCAFFOLD FOR NEOGLYCOPROTEIN SYNTHESIS	138
6.3	EXPRESSION OF RECOMBINANT SAP	139
6.4	CONCLUSION	142
CHAPTER 7 CONCLUSIONS AND FUTURE STUDIES		143
7.1	BIFUNCTIONAL BIOORTHOGONAL LINKERS	144
7.2	SPAAC AND OXIME REACTIVE SYNTHETIC GLYCOCONJUGATES	145
7.3	NOVEL SYNTHETIC GLYCOSYLATION OF PROTEIN SCAFFOLDS	147
7.4	BACTERIAL TOXIN NEOGLYCOPROTEIN INHIBITORS	148
7.4.1	<i>Engineering a non-bacterial pentameric protein scaffold</i>	149
7.5	FUTURE STUDIES	150
7.5.1	<i>Synthesis and attachment of a GM1-oxamine derivative for attachment to W88E</i>	150
7.5.2	<i>Inhibition studies of neoglycoprotein-based inhibitors of the verotoxin</i>	152
7.5.3	<i>Expression of a mutant human pentameric scaffold</i>	153
CHAPTER 8 EXPERIMENTAL		155
8.1	LINKER SYNTHESIS	156
8.2	SYNTHESIS OF GLYCAN DERIVATIVES	169
8.2.1	<i>Enzymatic synthesis</i>	177
8.3	SYNTHESIS OF NON-CANONICAL AMINO ACID AZIDOHOMOALANINE	182
8.4	SYNTHESIS OF CLICKABLE I-TAG	184
8.5	GENERATION OF Gb3 AFFINITY RESIN	185
8.6	SYNTHESIS OF Gb3 PHOSPHOLIPIDS	186
8.7	MOLECULAR BIOLOGY	189
8.7.1	<i>General methods and instrumentation</i>	189

8.7.2	<i>Buffer list and recipes</i>	189
8.7.3	<i>Bacterial Growth Media</i>	190
8.7.4	<i>General DNA manipulation procedures</i>	191
8.7.5	<i>General protein manipulation</i>	193
8.7.6	<i>Expression and purification of non-binding mutant CTB W88E</i>	195
8.7.7	<i>Expression of MET-W88E</i>	195
8.7.8	<i>Expression of N₃-W88E</i>	196
8.7.9	<i>Expression and purification of LgtC</i>	196
8.7.10	<i>Expression of EGCase II</i>	197
8.7.11	<i>Expression of VTB</i>	199
8.7.12	<i>Expression of Serum Amyloid P</i>	200
8.8	PROTEIN MODIFICATION.....	200
8.8.1	<i>General procedure for oxidation of W88E</i>	200
8.8.2	<i>General procedure of oxime ligation to W88E</i>	201
8.8.3	<i>General procedure for SPAAC ligation to N₃-W88E</i>	201
8.8.4	<i>Lightning-Link[®] HRP conjugation to CTB</i>	201
8.8.1	<i>Lightning-Link[®] HRP conjugation to VTB</i>	201
8.9	ELLA PROTOCOLS	202
8.9.1	<i>CTB inhibition assay</i>	202
8.9.2	<i>VTB inhibition assay</i>	203
8.9.3	<i>Data processing</i>	203
CHAPTER 9 REFERENCES		204
CHAPTER 10 APPENDIX		215
10.1	CTB W88E PLASMID MAP AND PROTEIN SEQUENCE	216
10.1.1	<i>Vector sequence</i>	216
10.1.2	<i>Primer sequences for the introduction of W88E to pSAB 2.2</i>	219
10.1.3	<i>CTB W88E protein sequence</i>	219
10.2	PSAB2.3_AZIDO PLASMID MAP AND PRIMERS FOR W88E, FOR THE EXPRESSION OF MUTANT CTB MET/N ₃ -W88E 220	
10.2.1	<i>Vector sequence</i>	220
10.2.1	<i>Primer sequences for the introduction of W88E to pSAB 2.3(azido)</i>	223
10.2.2	<i>Protein sequence</i>	223
10.3	EGCASE II EXPRESSION VECTOR AND PROTEIN SEQUENCE.....	224
10.3.1	<i>Vector sequence</i>	224
10.3.2	<i>Protein sequence</i>	228
10.4	LGT C IN PET-28A PLASMID MAP AND PROTEIN SEQUENCE	229
10.4.1	<i>Vector sequence</i>	229
10.4.2	<i>Protein sequence</i>	233

10.5	VTB PROTEIN SEQUENCE	233
10.5.1	<i>Reverse compliment of VTB with signalling peptide</i>	233
10.5.2	<i>Forward compliment of VTB with signalling peptide</i>	233
10.5.3	<i>Protein sequence</i>	233
10.6	SERUM AMYLOID P (E167Q) LPETG IN PET11A IN PROTEIN SEQUENCE	234
10.6.1	<i>Vector Sequence</i>	234
10.6.2	<i>SAP-LPETG Protein Sequence</i>	238
10.7	SUPPLEMENTARY FIGURES	238
10.7.1	<i>Results for the inhibition assay of lactose and 3.3(W88E) on CTB-HRP</i>	238
10.7.2	<i>Raw graphical data for VTB-HRP capture on Gb3 microtiter plates at three concentrations of synthetic glycolipid Gb3-DOPE</i>	239

Chapter 1 An Introduction

**The role of glycosylated structures in bacterial toxin infection and
methods of creating synthetic glycoproteins**

1.1 Glycosylation

Glycosylation is ubiquitous throughout nature and essential for complex life. Large oligo- and polysaccharide-decorated structures play major roles in biological systems, with approximately 700 proteins responsible for synthesis of the mammalian glycan library.¹ Glycoconjugates are grouped into three major classifications: N-linked glycoproteins, O-linked glycoproteins and glycolipids; many examples of these are shown on the cell surface in figure 1.1.

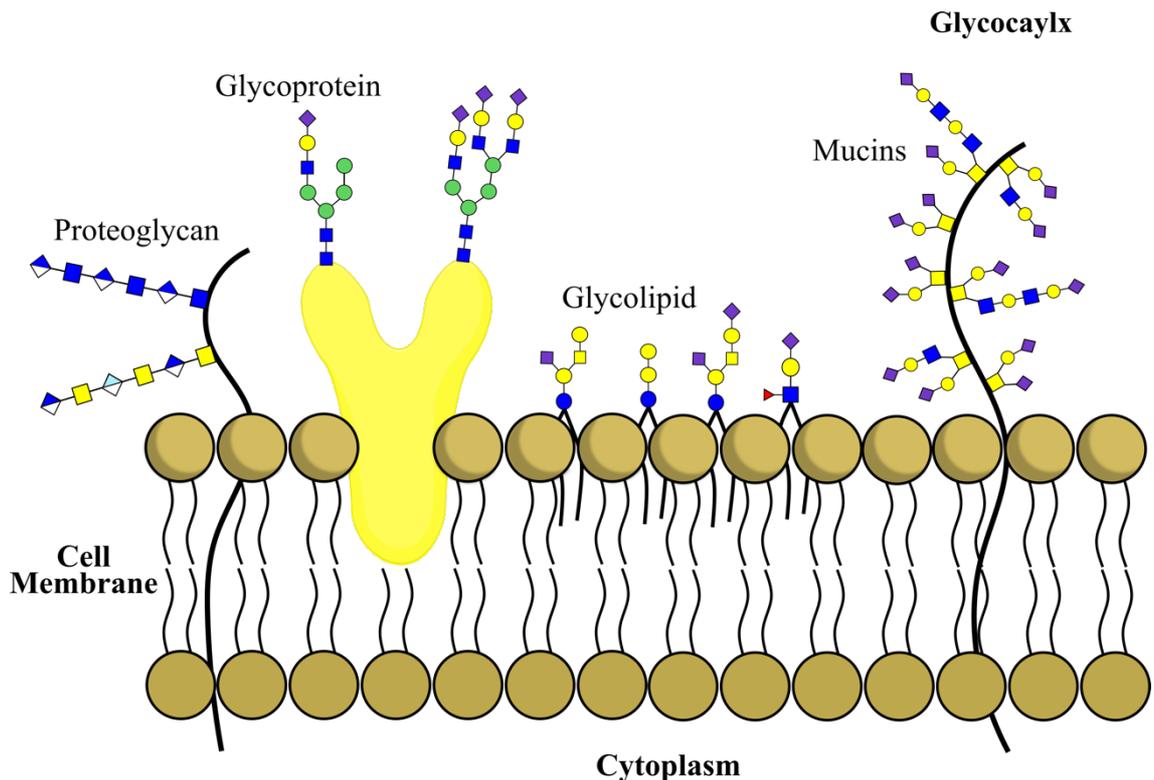


Figure 1.1: Representation of a section of cell membrane showing: glycosylated membrane proteins; glycolipids embedded in the phospholipid bilayer; and the larger mucin and proteoglycans structures, which extend high above the cell membrane. All of these make up the dense glycosylated extracellular structure known as the glycocalyx.

Proteins undergo a plethora of different post-translational modifications (PTM), which lead to greater diversity in the cellular proteome. An analysis of PTM statistics in 2011 quantified data from PTM databases and showed that glycosylation (specifically N-linked glycosylation) is the third most common experimentally-observed PTM, topping the list of putative PTMs.² A more recent study by Duan *et al.*, compiled data on PTMs in the context of profiling potential protein interactions and their possible functions, concluded that potentially disease causing proteins or protein that could be of interest for further research, could be identified based on their protein interaction network.³

1.1.1 N-linked glycoproteins

The process of N-glycosylation begins with the formation of a large lipid-bound donor glycan $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$, on the cytosolic surface of the endoplasmic reticulum membrane, which is then flipped into the ER lumen where it remains lipid-bound on the luminal membrane (figure 1.2).⁴⁻⁶ The large donor glycan is transferred to the polypeptide chain, as it is being translocated through the ER membrane from the ribosome. An enzyme known as oligosaccharyltransferase is responsible for the transfer of the glycan donor to asparagine residues within the peptide sequence, at glycosylation sites characterised by an Asn-X-Ser/Thr sequence.⁷⁻¹⁴

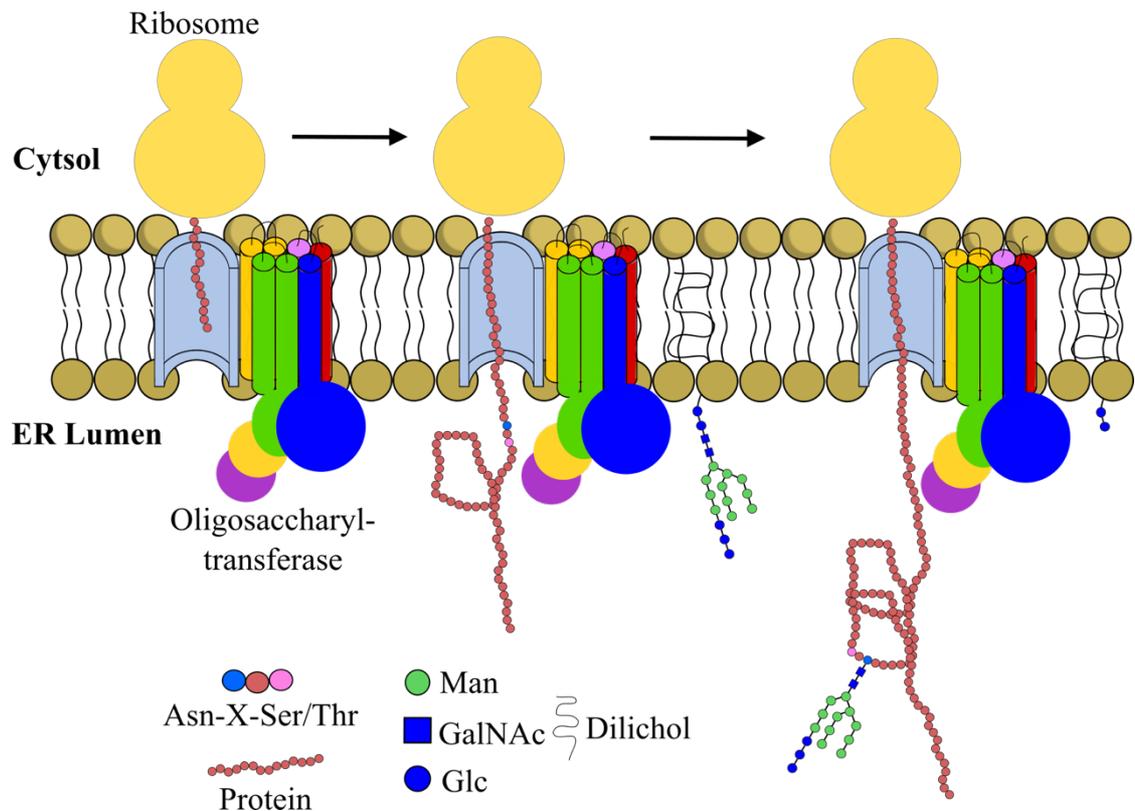


Figure 1.2: A representation of the mechanism for the production of N-linked glycoproteins showing the transfer of an oligosaccharide precursor onto the asparagine residue as the protein is translated into the lumen of the ER.

Once glycosylated, the glycoprotein can fold into its tertiary structure, leaving the glycans surface-exposed; the addition of such glycans profoundly influences the tertiary structure and subsequent cellular trafficking of the mature folded protein.¹⁵⁻¹⁹ The glycoprotein then undergoes a complex sequence of glycan-processing events – this reshaping of the carbohydrate structure has been extensively studied and reviewed, but is yet to be fully understood.²⁰⁻²³ A combination of glycosidase and glycosyltransferase enzymes trim and add different monosaccharides, dependent on the oligosaccharide composition, to create highly complex and diverse heterogeneous glycoproteins.^{20, 24-26}

1.1.2 O-glycosylation

The O-glycosylation of proteins occurs through a very different process to N-glycosylation. O-glycosylation is initiated by the linking of a GalNAc to a Ser/Thr residue. This first GalNAc monosaccharide is then extended to form common core structures (figure 1.3) for example by addition of Gal β -1,3 to form core 1, or by addition of this and GalNAc β 1-6, to form core 2.²⁷ These structures are synthesised by sequential addition of monosaccharides as the polypeptide chain travels through the Golgi. Attachment of the initial GalNAc residue occurs post-translationally and can be carried out by at least 21 different N-acetylgalactosaminyltransferases, that have different peptide recognition sequences.²⁸ Glycosyltransferases within the Golgi then recognise the terminal saccharide and extend the chain using nucleotide sugar donors, in a process that is compartmentally regulated within the cell, as different enzymes are found in different regions of the Golgi. O-glycosylation is abundant in structures that are present on the cell membrane and structures extending above the membrane that make up the glycocalyx. Mucins and proteoglycans are the predominant O-glycosylated protein structures on the cell surface and are vital for maintaining extracellular matrix structure and binding water, and also play roles in regulating cell adhesion.²⁹

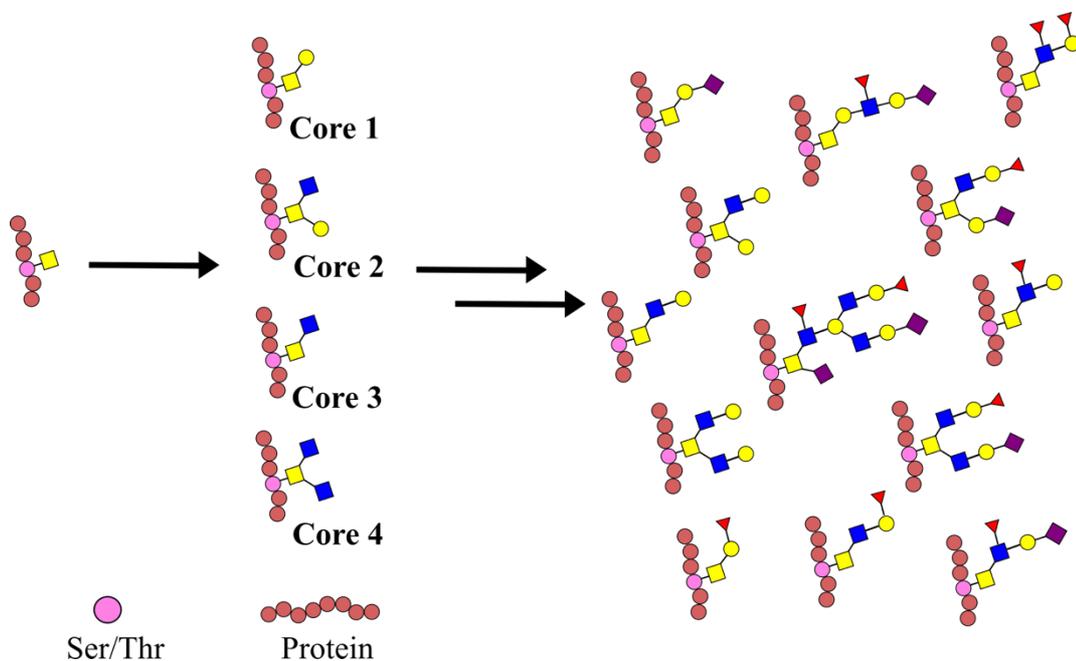


Figure 1.3: O-glycosylation of a protein backbone showing the starting GalNAc-Ser/Thr undergoing glycan extension through the golgi, highlighting four core structures and examples of mature extended structures.

1.1.3 Other glycosylated structures

Glycosylation is not limited to intra- and extracellular proteins but also prevalent in modification of lipids and lipid-linked proteins. Glycolipid structures are built upon either ceramide or phosphatidylglycerol to create glycosphingolipids and glycosphospholipids, respectively, which are embedded within the membrane.³⁰ Glycosphingolipid biosynthesis is known to begin in the ER, and continue through the Golgi in which glycan extension occurs before being trafficked to the membrane forming lipid rafts.^{31, 32} Alternatively, phosphatidylglycerol is used in the formation of glycosylphosphatidylinositol (GPI) anchors – such structures are responsible for the anchoring of proteins to the lipid membrane.^{33, 34}

1.2 Roles of glycosylation and lectins in disease

As has been outlined, glycosylation is abundant across many structures and the glycosylated structures are essential for normal cellular function. Changes to glycan structure and composition through alterations in the glycosylation pathway, can disrupt normal cellular function. ‘Congenital disorders of glycosylation’ is the term used for diseases relating to errors in glycosylation, arising from inherited genetic mutations and cause disease from birth and early development. Developmental diseases of this nature have been associated with both errors in both N- and O-linked glycoproteins, as well as glycolipids.³⁵⁻³⁷ Other disorders of a similar nature, such as lysosomal storage disorders, are caused by an inability to degrade glycosylated structures. In such cases, the build-up of glycosylated molecules has severe degenerative effects leading to disorders such as: Tay-Sachs, Fabry, Gaucher and Hunter syndrome affecting the breakdown of glycolipids and glycosaminoglycans (GAGs).^{38, 39} Defective glycosylation in adults has also been linked to cancer, autoimmune disorders, and diseases of the blood, gastrointestinal tract, liver and kidneys.

1.2.1 Pathogens which exploit cell surface carbohydrates

Beyond diseases related to genetic and acquired mutations, pathogens also exploit the carbohydrate-containing structures on their host cells, in order to infect and disrupt host physiology, leading to illness and disease. Bacteria, virus and parasite infection is often dependent on the recognition of glycoproteins or glycolipids for tissue adhesion and invasion of the cell. Various pathogenic species and their glycan ligands (figure 1.4) within the glycocalyx on the host cell surface were reviewed by Imberty *et al.*⁴⁰ Such

interactions are not limited to recognition of human glycans, pathogens can also exploit lectins (carbohydrate binding proteins) on the cell surface that can recognise bacterial polysaccharides.

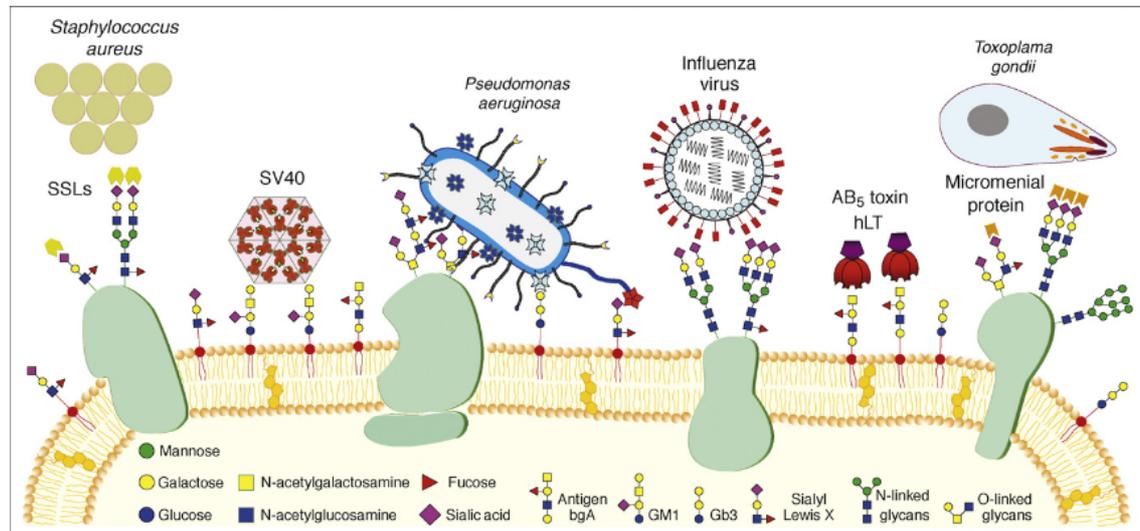


Figure 1.4: Image representing the strategies employed by pathogens for host glycan recognition and adhesion, reproduced from Current opinion in structural biology, 18, A. Imberty, A. Varrot, Microbial recognition of the human cell surface glyconjugates, Pg 567-576., copyright[®] (2008). Reproduced with permission from Elsevier.⁴⁰

Viruses are well known to interact with glycoproteins on the surface of cells through lectins present on the viral envelope. A well understood example is the influenza virus; the lectin hemagglutinin is a viral envelope protein which binds sialic acid-containing glycans on the host cell surface, to initiate membrane fusion.^{41, 42} In the case of HIV, it is the viral glycoprotein gp120 which binds to the human lectin DC-SIGN which recognises mannose,^{43, 44} and ultimately infects T-cells by binding to CD4 through glycoproteins gp120 and gp41 on the exterior of the virion.⁴⁵ Bacteria have a different mechanism for adhering to cells and colonising tissues, expressing various adhesins on the cell surface that allow the bacteria to stick to various tissue types and colonise.^{46, 47}

1.2.2 Bacterial toxins bind cell surface gangliosides

Some bacteria secrete toxins which are often lectins, capable of binding cell surface carbohydrates in a multivalent manner, due to multiple carbohydrate binding domains (CBDs). The high avidity binding to their glycan ligands is the first step in endocytosis of the toxin.⁴⁸ The AB₅ family of enterotoxins are comprised of a two subunit structure with a pentameric lectin B-subunit, and a large A-subunit. The A-subunit has toxic enzymatic activity, whereas the B-subunit lectin has no toxic activity and is only involved in receptor recognition and cellular uptake. These types of toxins exploit the host cellular machinery to release the toxic subunit into the cytosol and elicit their toxic effect on the cell. Some well-known examples of these toxin-producing bacteria include *Escherichia*.

coli, *Shigella* species and *Vibrio cholerae*, which are major causes of diarrhoeal disease. Data from the World Health Organisation (WHO) in 2016 ranked diarrhoeal disease 9th in the top 10 causes of deaths globally, and 2nd in the top 10 causes of death in low income countries.⁴⁹ WHO figures show that diarrhoeal disease is a major threat to children with a total 1.7 billion cases worldwide, killing 525,000 children under five per annum.

1.2.2.1 Cholera toxin from *Vibrio cholerae*

The bacterium *V. cholerae* is the responsible for the disease cholera, a major cause of the gastrointestinal infection in developing countries, leading to diarrhoea, dehydration and shock, estimated to affect 1.3-4 million people leading to between 21,000 and 143,000 deaths per annum.⁵⁰ The cholera holotoxin (CT), is a prototypical AB₅ toxin comprised of a toxic A-subunit (figure 1.5, CTA, blue and green) and the non-toxic pentameric B-subunit (CTB, red).⁵¹ The five B-subunit binding sites are located on the same face of the protein,⁵² with each individual structural unit of the pentamer termed protomer, containing a single binding domain.⁵³ This is a significant structural since each protomer can bind the cell surface ganglioside GM1 (monosialotetrahexosylganglioside), multiple binding sites gives rise to multivalent interactions formed through binding a total of five GM1 structures on the cell surface.⁵⁴

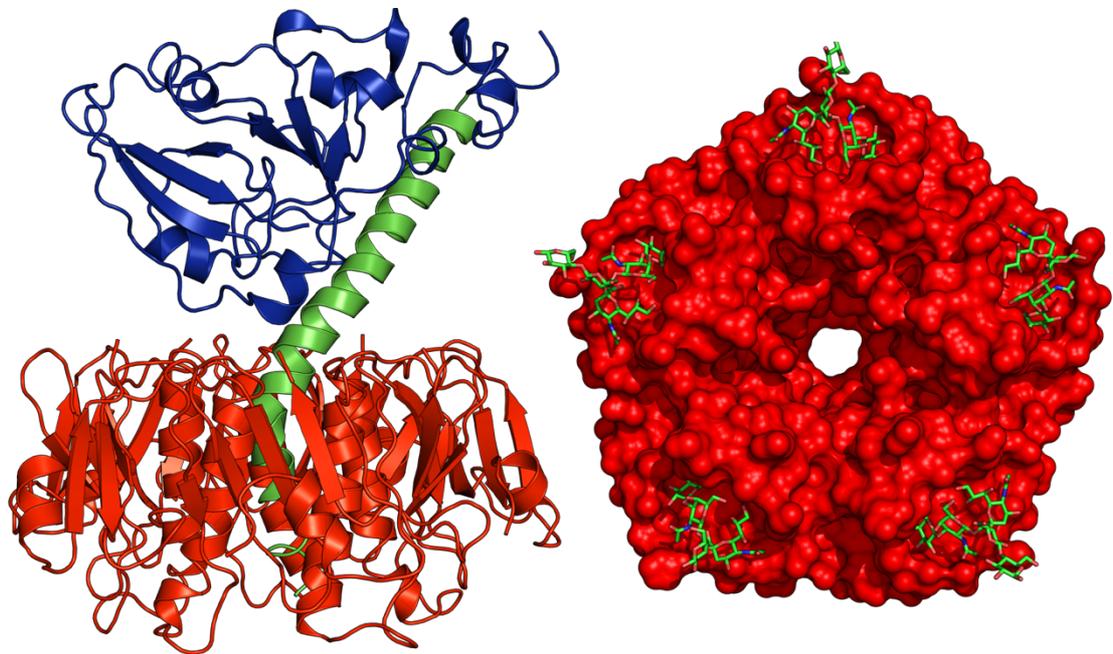


Figure 1.5: Crystal structures of cholera holotoxin generated from the atomic co-ordinates 1S5E.pdb, and surface representation of the binding face of the CTB-subunit bound to GM1 pentasaccharide bound, generated from atomic co-ordinates 3CHB.pdb. The B-subunit (CTB) consists of five identical 11.5 kDa protomers which assemble to form B₅ pentameric subunit, around the A2 polypeptide chain (green).

The binding of GM1 to the B-subunit has been extensively studied through a variety of biophysical techniques. Detailed isothermal titration calorimetry (ITC) experiments

determined that GM1 binds with a K_d of 40 nM, making this one of the strongest known protein-glycan interactions.⁵⁵ Further studies accounting for multivalent interactions to the native GM1 ganglioside (figure 1.6) using surface plasmon resonance determined the apparent K_d to lie in the picomolar range.⁵⁶ The internalisation of the toxin proceeds through various modes of endocytosis; it is understood that the formation of the multivalent interaction aids this internalisation.⁵⁷

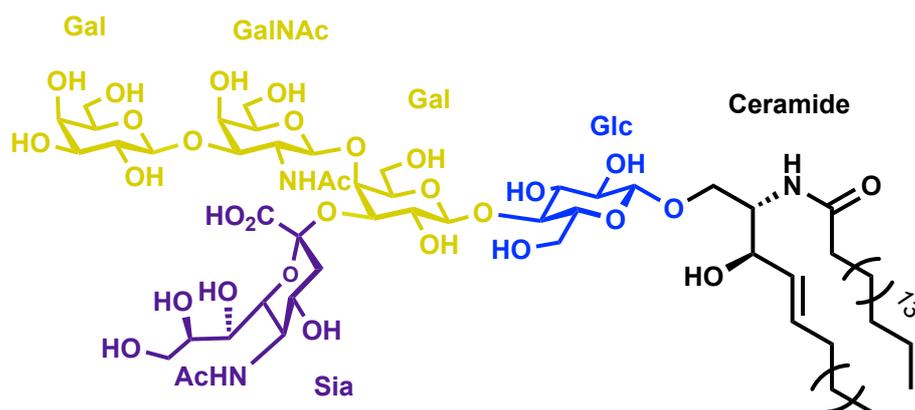


Figure 1.6: Structure of the GM1 ganglioside, displaying the pentasaccharide binding head. (Galactose (Gal) – yellow, N-acetyl galactosamine (GalNAc) – yellow, Sialic acid (Sia) – purple, Glucose (Glc)– blue, ceramide – black.

1.2.2.2 Verotoxin or shiga-like toxin from *E. coli*

Although cholera is a widely known example of disease caused by toxin-producing bacteria prevalent in the developing world, AB₅ toxin-secreting bacteria are not confined to the poorer countries. The verotoxin (VT), also known as the shiga-like toxin, shown in figure 1.7, is an AB₅ toxin produced by *E. coli* O157:H7 and other various verotoxin/shiga-like toxin secreting *E. coli* serotypes (VTEC or STEC), which causes diarrhoeal disease. 5-10% of infected patients then go on to develop haemolytic uremic syndrome (HUS), and those who develop HUS face a 10% mortality rate.⁵⁸ STEC highest incidences of disease in South-East Asia, Europe and American regions, obtained through the consumption of contaminated, undercooked food. VT has no sequence homology to the CT; however, a distinctly similar AB₅ structure is conserved (figure 1.7). Isoforms VT-1 and 2 share varying amounts of sequence homology to the shiga toxin produced by *Shigella*.⁵⁹ VT-1 is almost completely homologous, whereas the more virulent VT-2 shares only 50% homology.⁶⁰ Co-crystallisation of the toxin B-subunit with its Gb₃ ligands, shows 15 copies of the trisaccharide in three different binding sites on the protomer.⁶¹ Interaction of the verotoxin B-subunit (VTB) with the soluble Gb₃Os ligand has a K_d , typical of that between protein-carbohydrate interactions determined to

be 1 mM from ITC and MS studies.^{62, 63} High avidity binding is observed using the native glycolipid with a K_d of 4.2 nM, for Stx 1, determined by enzyme-linked immunosorbent assay (ELISA).⁶⁴ Stx 2 has been shown to have a lower affinity with reported binding of Gb3 determined to have a K_d of 370 nM using radiolabelled Stx2, and 1040 nM by SPR, the K_d using ELISA could not be determined.^{64, 65} In depth studies comparing the binding interaction of Stx1 and Stx2 revealed that the binding of Stx2 was dependent on the presence of other lipids/glycolipids, such as Gal-ceramide or cholesterol.⁶⁴ These studies highlight the differences in binding avidity when multivalent interactions occur; the potential binding of 15 ligands increases affinity from mM to nM K_d .

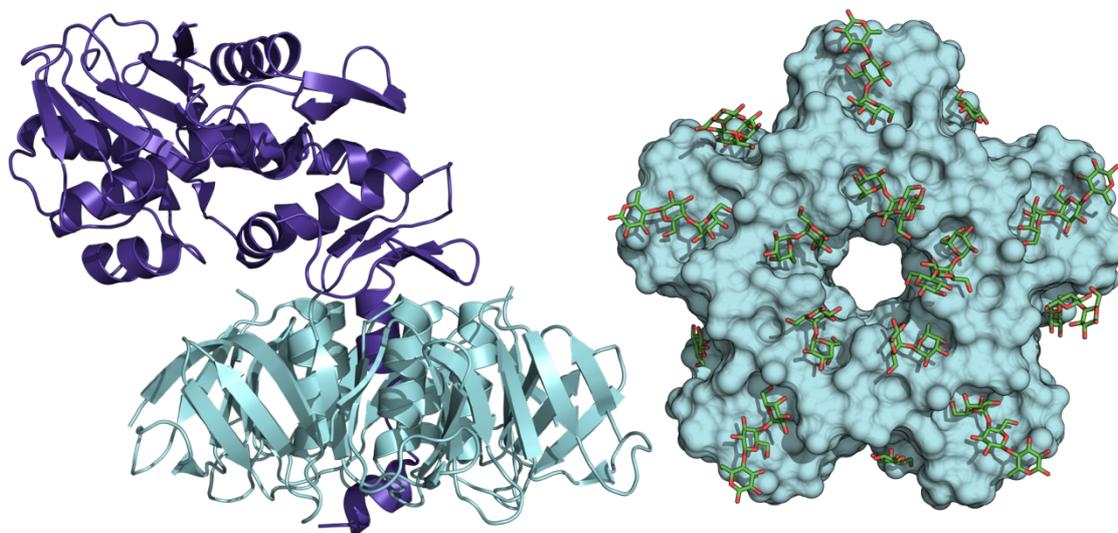


Figure 1.7: Crystal structure of the VT-1 generated from the atomic co-ordinates of 1R4P.pdb, alongside a surface representation of the VTB-subunit bound to the Gb₃ trisaccharide, generated from the atomic co-ordinates of 1BOS.pdb.

The toxin's effect on the intestinal tract is thought to result from damage to the absorptive epithelial cells, resulting in a net secretory effect of ions into the lumen, causing the diarrhoeal symptoms. Many different mechanisms have been proposed for STEC and toxin entry into tissues and bloodstream; however, it is not well understood.⁶⁶ Once in the bloodstream, tissues with high levels of Gb₃ ganglioside (figure 1.8) such as renal tissue are affected.^{67, 68}

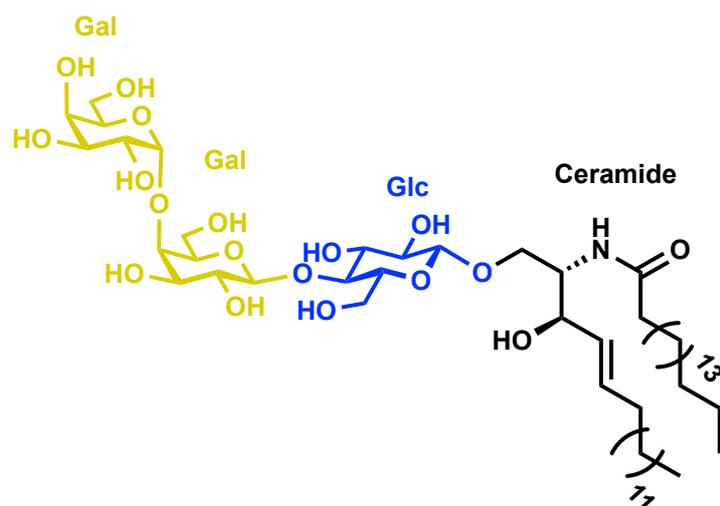


Figure 1.8: Structure of Gb3 ganglioside displaying the trisaccharide binding head. (Galactose (Gal) – yellow, Glucose (Glc) – blue and the ceramide chain (black)).

Much like the cholera toxin, the B-subunit binding leads to internalisation of the A and B-subunits, but in this case, subsequent release of the toxic A-subunit irreversibly inhibits ribosomal function, resulting in cell death.⁶⁹ The toxin is highly cytotoxic, requiring very little of the toxin to bring about fatal disease. A study by Macleod *et al.* showed in a swine model, that a dose of 3 ng/kg was enough to kill 50% of the study population.⁷⁰

1.3 Multivalent inhibitors against bacterial adhesion and infection

Carbohydrate binding sites of proteins are typically shallow in nature, with relatively few residues in the proteins making contact with the saccharide through hydrogen bonds. The studying many protein-carbohydrate interactions through isothermal calorimetry (ITC)⁷¹ and surface plasmon resonance (SPR)⁷² has shown that individual binding interactions are weak in the mM affinity range.^{73, 74} Many studies on various carbohydrate-lectin interactions have shown that it is largely enthalpic contributions that drive the energy of binding.^{71, 75} However, despite the weak affinities protein-carbohydrate interactions play important roles for cellular function. The study of carbohydrate interactions have provided substantial evidence that many protein-carbohydrate interactions are not singular, but in fact, the binding generation of multiple interactions and binding multiple copies of the carbohydrate ligand is what is observed.^{76, 77} This phenomenon is known as multivalent binding, and is responsible for the enhancement of binding affinities to reach biologically relevant strengths.⁷⁸ The observed increases in binding affinities can be rationalised by high avidity (the accumulative binding strength) of multivalent binding between due to the additive effects of binding each ligand. In many cases of multivalent interactions the effect of cooperative binding occurs, in which the binding of one ligand

increases the affinity for the next ligand binding often described as the ‘cluster’ or ‘chelate effect’.^{79, 80} We now understand that multivalent interactions predominate at interfaces such as the cell membrane, and that an effector response can be very much dependent on the binding of more than one ligand. Many cell surface polysaccharides are also known to form clusters, or bundles in the membrane to facilitate multivalent binding. This explains the evolutionary tendency of bacterial lectins to form multivalent interactions with cellular surfaces, binding multiple copies of the ligand creates the strong interactions required for internalisation. Investigation into the thermodynamics of multivalent interactions have been reported on many of the same lectins that have been used to probe monovalent interactions using ITC, which has been used to determine the valency and affinities of binding. Finding affirmed that the increased K_a increased proportionally to the sum of ΔH for the individual interactions.^{71, 81} With multivalent interactions it is also important to consider there is a trade off between the summative enthalpy, and the entropic contributions to binding which a more profound than in case of monovalent binding. In contrast to the summative effects ΔH for multivalent interactions, $T\Delta S$ does not scale proportionally to the number of binding epitopes resulting in more favourable interactions.⁸² The contributions of entropic factors particularly important for the interaction of synthetic multivalent ligands of lectins, due to the penalty due to the loss in conformational flexibility of any tethers linker to the ligands being bound and must be considering when designing such multivalent structures.^{54, 83, 84}

Probing and understanding of multivalent interactions between protein and glycoprotein, has led to interest in the development of multivalent molecules for pathogenic lectins inhibition.⁸⁵ Research into the development of inhibitors towards these types of lectins have frequently exploited the multiple binding sites, as the traditional medicinal chemistry approach of small molecule mimetics provides little improvement in affinity from the natural ligand.⁸⁶⁻⁸⁹ The two approaches in the context of glycomimetics vs glycoconjugates has been reviewed by Ceconi *et al.*,⁹⁰ and explains why a diverse variety of structures have been explored in the design of bacterial toxin inhibitors from the simpler di-, tri- and tetravalent branched structures to larger macromolecular polymers and dendrimers. These inhibitors based upon multivalent scaffolds have been extensively reviewed both for various lectins, pathogenic agents, bacterial toxins,⁹⁰⁻⁹⁴ and in detail for the cholera toxin.⁹⁵

1.3.1 Glycopolymer-based inhibitors of bacterial lectins

Glycopolymers have been utilised to create highly polyvalent glycosylated structures as inhibitors for CT, LT and ST toxins, using various different polymer templates. Early reports by Schengrund *et al.* used simple polylysine chains functionalised with GM1_{os} for the inhibition of CTB⁹⁶; other polypeptide structures have also been reported for the inhibition of toxins for the verotoxin.⁹⁷⁻¹⁰⁰ Other polymer-templated multivalent structures have been developed using functionalised polyacrylamide, polymethacrylamide and dextran backbones.¹⁰¹⁻¹⁰⁴ Glycopolymers like the ones shown in figure 1.9A adopt an extended conformation, and conformation-tuneable polymers like those in 1.9B, are more potent inhibitors than the monovalent ligands. Many studies have shown that the backbone composition and conformation,¹⁰⁴⁻¹⁰⁶ ligand density^{98, 106} and spacing¹⁰⁷ are important factors for the inhibitory potential of glycopolymers.

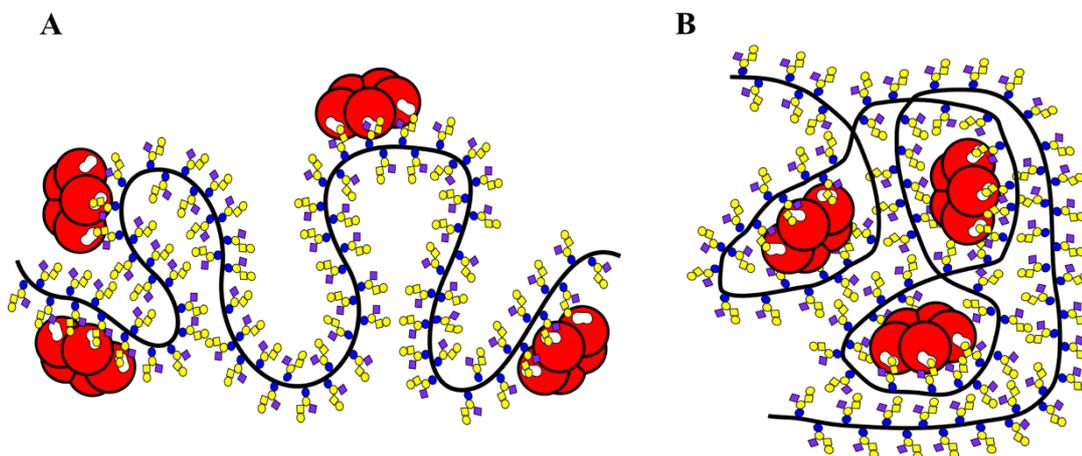


Figure 1.9: Examples of GM1 coated glycopolymers inhibition of cholera toxin. Glycopolymer inhibition shown in image A shows linear glycopolymers binding the CTB-subunits. Figure B shows capture of the toxins by glycopolymers which backbone structure can be tuned.

1.3.2 Glycodendrimer-based inhibitors

Glycodendrimers form complex, branched, tree-like networks, which can be used to create structures with high valencies.¹⁰⁸ Dendrimers have been used to target bacterial toxins using high valence penta(ethylene glycol) galactose-terminated^{109, 110} and polyamidoamine (PANAM) GM1-terminated dendrimeric inhibitors,^{111, 112} reaching low micromolar and low nanomolar affinities, respectively. Pieters and co-workers reported the use of octavalent GM1 dendrimers (figure 1.10), which reached picomolar affinities for the cholera toxin.¹¹³

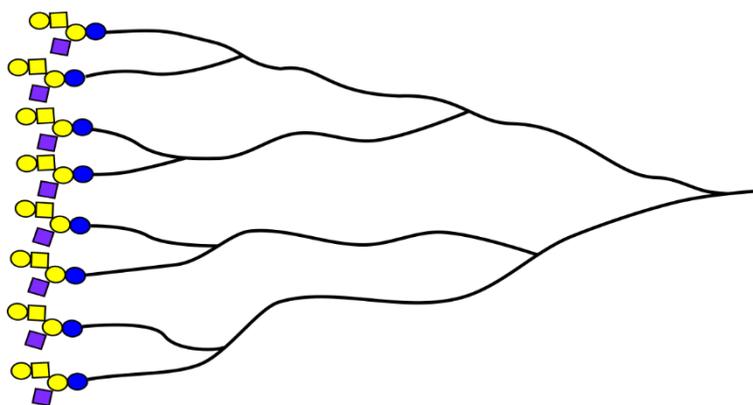


Figure 1.10: A cartoon representation of a GM1 terminated octavalent glycodendrimer produced by Pieters.⁹¹

A silicon-carbon based dendrimer named SUPER-TWIG which has up to 36 copies of the Gb3 trisaccharide, has been found to be effective in neutralisation of circulating shiga-like toxin.^{114, 115} Although some very potent dendrimer-based inhibitors of the bacterial toxins have been produced, which outperform the most potent glycopolymers, the structures do not exploit the symmetry of the toxin lectin subunit, with the dense ligand population not optimal for binding.

1.3.3 Cyclic multivalent inhibitors

Cyclic scaffolds have been used to target each of the binding sites upon bacterial toxins, matching the valency and symmetry of the toxin to give optimal ligand binding. Non-peptide cyclic cores were first reported by Bundle and co-workers, who created the decavalent, Gb3 terminated inhibitors STARFISH (linked at 2-position of Gal) and DAISY (linked at the reducing terminus, figure 1.11) towards the verotoxin based upon a functionalised glucose core, showed over 10^6 fold increase in potency from the monovalent Gb3.^{116, 117} The decavalent structures were originally designed to occupy two of the three different VTB binding sites. Crystal structures of these inhibitors bound to the bacterial toxins revealed that the decavalent inhibitors in fact bound two different toxin B-subunit pentamers, as opposed to two different binding sites of the same B-subunit.

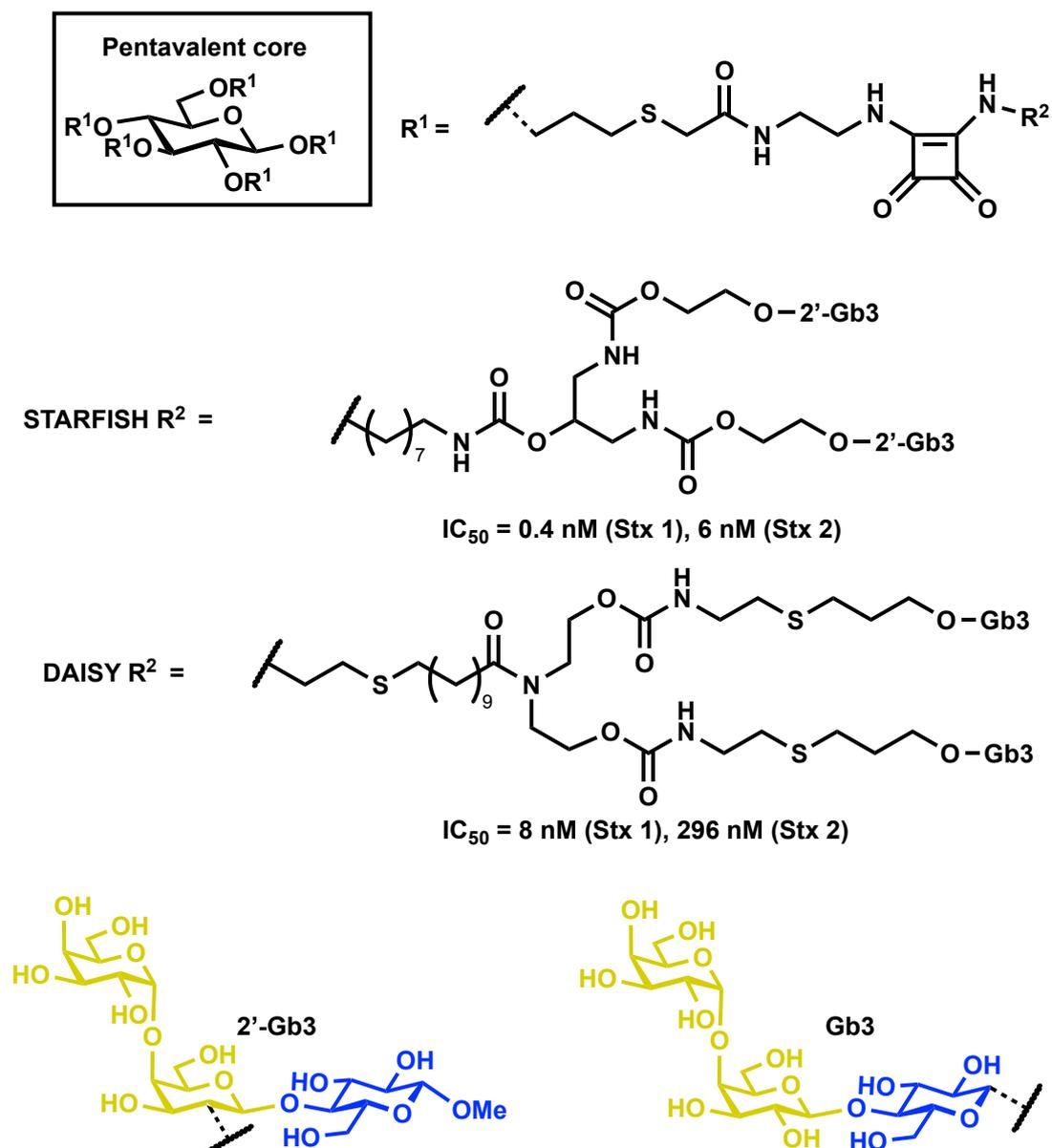


Figure 1.11: Decavalent star-shaped inhibitors of Stx and the relevant IC₅₀ values as reported by Bundle and co-workers.

At the same time, Fan and co-workers targeted the CT and heat-labile toxin (LT), with star-shaped inhibitors based upon a pentacyclin core (figure 1.12), to effectively inhibit both CTB and heat labile toxin B-subunits (LTB).^{118, 119} A decavalent variation of their pentacyclin inhibitor was more potent than the pentavalent inhibitor by a factor of ten, and 10⁶ fold more potent than the monovalent galactose.¹²⁰ Similarly to the decavalent inhibitors described by Bundle, crystal structures of these inhibitors revealed the binding of two pentameric subunits in a face-to-face dimer arrangement.¹¹⁸ A number of cyclic peptides have been reported as scaffolds to create pentavalent structures by Kumar *et al.*¹²¹, and decavalent peptides by Zhang *et al.*¹²⁰

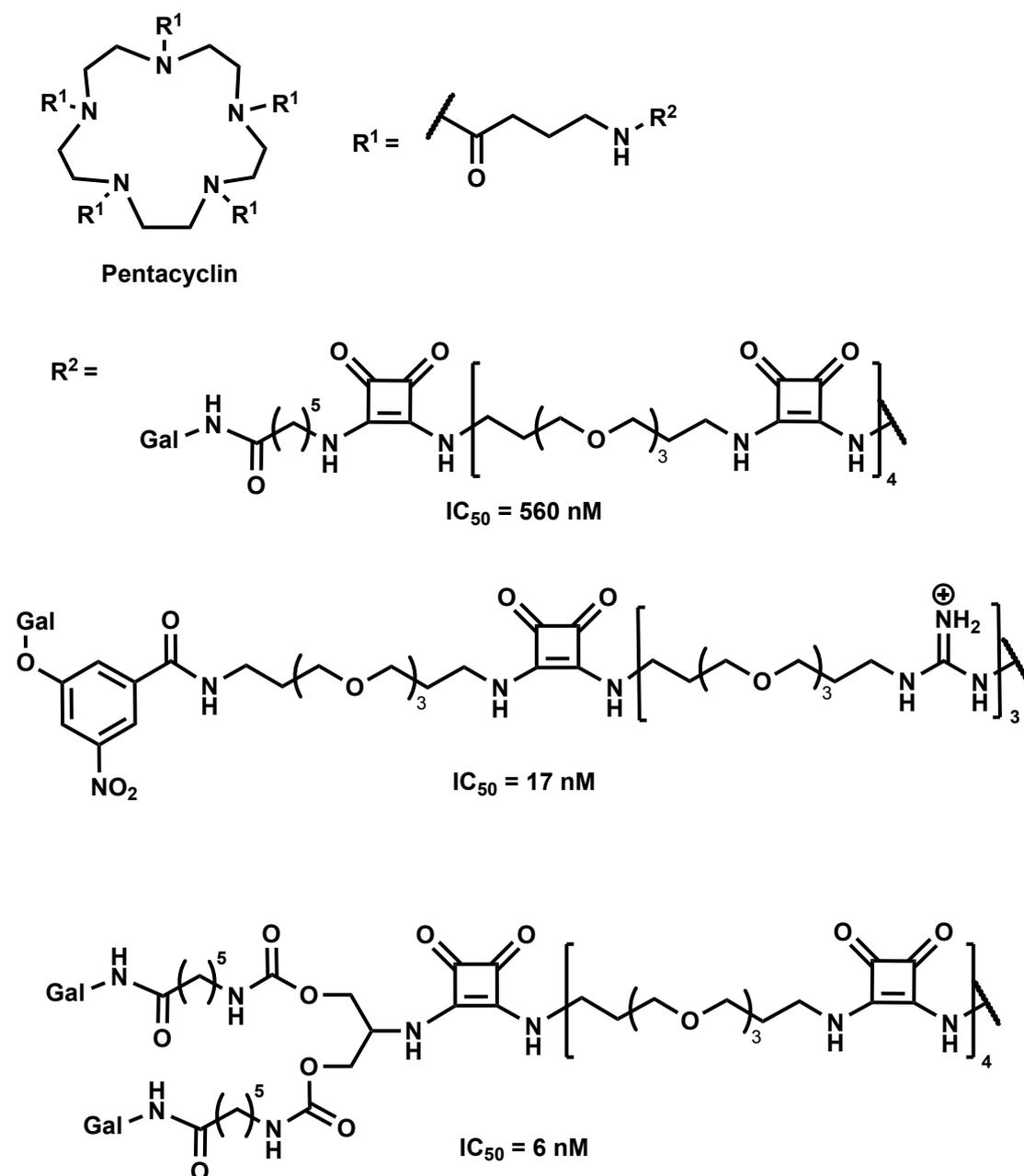


Figure 1.12: Penta- and decavalent star-shaped galactose terminated inhibitors of the CT and LT and the corresponding IC_{50} values reported by Fan and co-workers.

1.3.4 Protein templated inhibitor design

Following on from the success of the penta- and decavalent cyclic inhibitors, both Fan and Bundle approached new inhibitors using the protein dimerisation strategy (figure 1.13) pioneered by Pepys *et al.*; employing homodivalent linkers as bridges between two pentamers.¹²² It was shown that serum amyloid P (SAP, represented in figure 1.13 as a green pentamer) was capable of binding rigid L-carboxyproline derivatives, with an IC_{50} of 0.9 μM , and the homodivalent carboxyproline in figure 1.13 caused the dimerisation of SAP in a face-to-face arrangement. Cyclic ketal ligands (formed from the cyclisation

of pyruvate and glycerol, figure 1.13) were used to construct divalent linkers which are efficient SAP binders, the increased rigidity and decreased rotation proving beneficial.

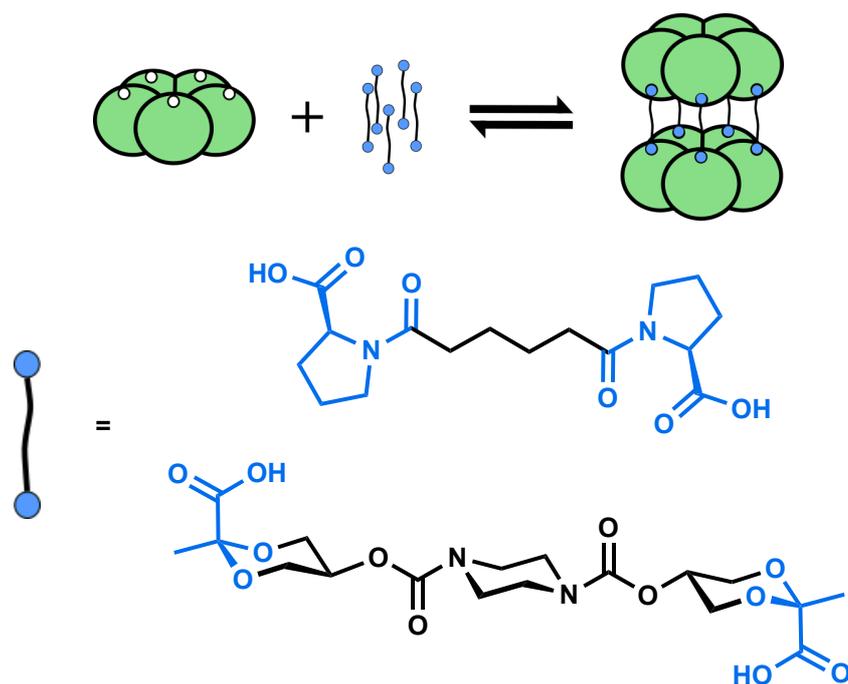
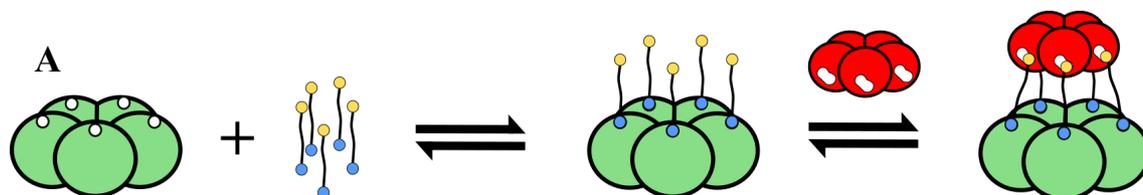


Figure 1.13: Dimerisation of the serum amyloid P protein (SAP, green pentameric protein) pentamer using divalent ligands containing L-carboxyproline or cyclic ketals of pyruvate.

This divalent bridge approach (figure 1.14) was used to bring about the inhibition of bacterial toxins by first prearranging the toxin ligands, in the correct symmetry for toxin binding using the SAP pentamer. Fan and co-workers combined the carboxyproline motif with a *m*-nitrophenyl galactoside, as an inhibition method for the CTB/LTB-subunits. The preordering of the ligands gave a significant increase in inhibition giving an IC_{50} of 0.98 μ M against CTB versus 620 μ M for the bivalent ligand with no SAP present.¹²³ Bundle independently used the same approach for inhibition of the shiga-like toxin, synthesising a cyclic ketal derivative of pyruvate terminated with Gb3 ligands, which was shown to inhibit the adhesion of Stx-1 to Gb3 coated plates with an IC_{50} of 560 nM¹²⁴, a 50-fold improvement on a more flexible analogue they had previously tested.¹²⁵



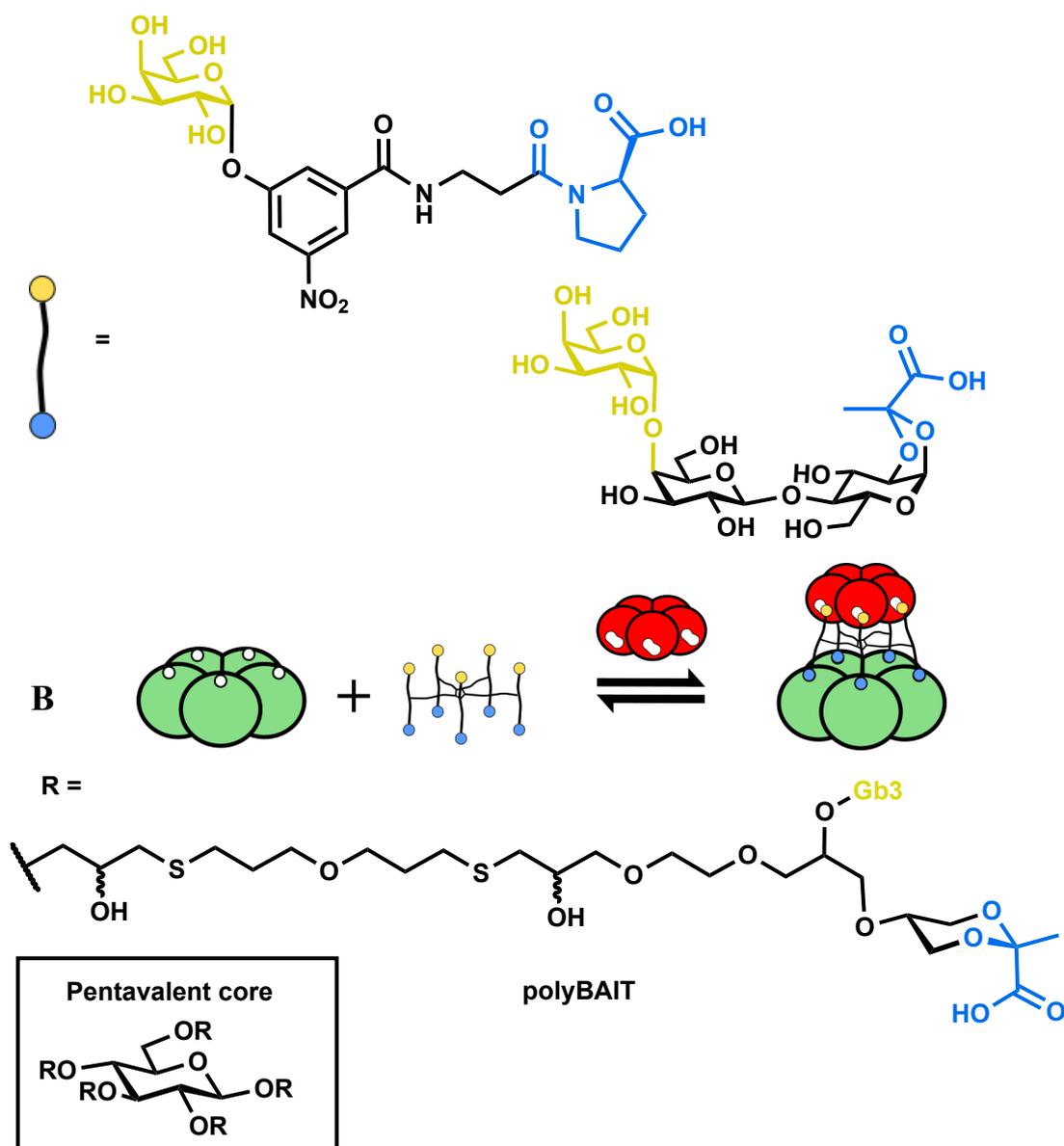


Figure 1.14: Protein templated inhibition of bacterial toxins using heterodivalent ligands capable of binding both SAP (green pentamer) and the bacterial B-subunit (red pentamer) as bridges between template and target proteins. SAP is used as a protein scaffold to preorder the bacterial toxins ligands to better target the binding sites on each protomer unit.

Prior to this work a pentameric divalent ligand nicknamed polyBAIT was developed by Bundle and co-workers.¹²⁶ The structure was based upon their previous STARFISH core, however, applied to a different inhibition strategy shown in figure 1.14B. The pentameric bridging structure was an attempt to combine both the supramolecular and multivalent effects, to inhibit the Stx toxin, however, this achieved sub-micromolar activity proving less successful than the short-rigid linkers reported later.^{125, 127} This can be explained by the entropic benefit when using the shorter ligands, with the larger, more flexible ligands such as polyBAIT experiencing large loss in conformational entropy upon binding.

1.3.4.1 Proteins as scaffold for multivalent inhibitors

An alternative approach to creating protein-templated design was taken by Branson *et al.*¹²⁸, using neoglycoproteins (a glycoprotein with a non-native protein-carbohydrate linkage¹²⁹) as multivalent inhibitors based upon a protein scaffold. Here, a non-binding mutant of the CTB pentamer was used as a protein scaffold and derivatised with a GM1 derivative attached via a linker to an oxidised N-terminal threonine. Figure 1.15 shows the synthesis of the pentavalent neoglycoprotein inhibitor from a non-binding mutant of CTB-subunit, which showed an IC_{50} against CTB of 104 pM.

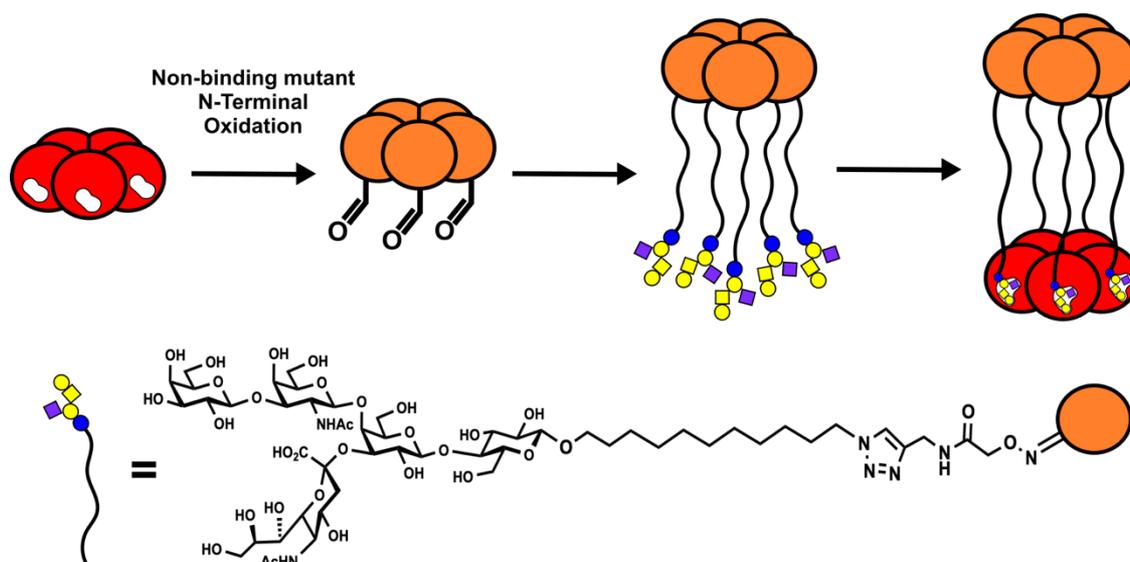


Figure 1.15: Cartoon depiction of novel protein based CTB inhibitor developed mutation of the wild-type CTB-subunit to form a non-binding mutant, and attachment by oxime ligation of synthetic carbohydrate derivatives created through chemoenzymatic synthesis to form the first neoglycoprotein inhibitor.

Chemical modification at the N-terminus of each protomer preorders the glycan in a pentameric arrangement, closely matching the distances between GM1 binding sites. A 1:1 heterodimer between the pentavalent neoglycoprotein and WT CTB, was confirmed by dynamic light scattering (DLS) and size exclusion chromatography (SEC), suggesting each glycan is occupying the corresponding toxin binding site. The inhibitors based upon protein scaffolds clearly show the benefit of prearrangement of the glycan ligands of the inhibitor into the correct geometric orientation, leading to more potent inhibitors.

1.4 Synthetic glycoproteins

With evidence supporting neoglycoproteins as potential therapeutics, the ability to access homogeneous glycosylated proteins through synthetic methods has many potential applications. The ability to produce glycoproteins as a single glycoform, is not only beneficial for developing glycoprotein-based therapeutics, but also in the study of glycobiology. A huge range of chemical methods and tools for building up glycans one monosaccharide at a time are documented, with some elegant techniques for controlling anomeric stereochemistry. Recreating the process of glycosylation on proteins poses many challenges: synthesis of glycans with the correct monosaccharide sequence; stereo- and regio-chemistry; overall abundance of glycans; and sites of glycosylation, each of which represent a significant synthetic problem to control in a single pot.

1.4.1 Controlling the glycosylation of macromolecules

The rapidly expanding field of chemical biology is now armed with a vast toolbox of methods for perturbing biological systems, which has led to greater understanding of biological processes *in vivo*. Development of these tools to design and engineer macromolecules using chemical ligations has crossed over into creating synthetic macromolecules. While many chemical ligations are widely adopted for protein conjugation and modification, there has, however, been growing interest in applying these methods to the production of synthetic glycoproteins.

1.4.2 Chemical glycosylation exploiting natural amino acid side chains

1.4.2.1 Methods of glycosylation at lysine

Lysine is the most abundant nucleophilic side chain in proteins, with its solvent-exposed primary amine functionality.¹³⁰ Amines are known to react with carbohydrates at the anomeric centre through the hemiacetal functionality, forming an imine link with the reducing sugar. Performing a reductive amination in the presence of sodium cyanoborohydride as a mild reducing agent forms a permanent covalent bond between oligosaccharide and the peptide/protein (figure 1.16); but this does result in ring opening of the reducing sugar into the linear chain. This method was first described by Gray in 1974, showing the functionalisation of proteins and amine-functionalised gels with oligosaccharides by reductive amination.¹³¹ The use of reductive amination for the production of glycoconjugates was pioneered by Jennings and co-workers, who developed a wide range conjugate vaccines using meningococcal and streptococcal

bacterial polysaccharides based on the tetanus-toxoid.¹³²⁻¹³⁶ Glycoconjugate vaccines with known tumour-associated antigens have also been produced by reductive amination, following the chemical synthesis of allyl-functionalised glycans that can be ozonolysed to produce an aldehyde.¹³⁷⁻¹³⁹ Advances in the production of such vaccines has been aided by development of the automated synthesis of such allyl containing glycans.¹⁴⁰

Lee *et al.* have reported the used of imido thioglycosides to form links to lysine side chains by amidation.¹⁴¹ These approaches (figure 1.16) have proved widely applicable and were frequently used for the synthesis of neoglycoproteins for many years,^{141, 142} but are defective in the fact that it is difficult to control site-selectivity and homogeneity of the glycoproteins produced.¹⁴³

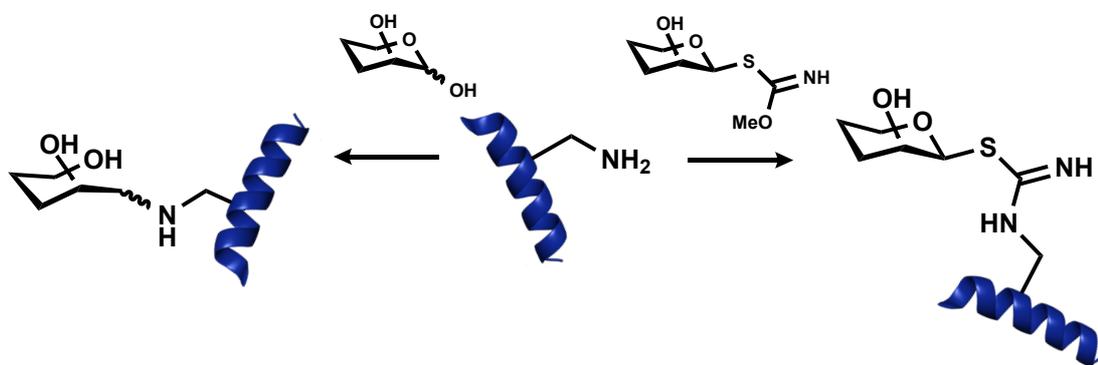


Figure 1.16: Glycosylation at lysine residues by covalent attachment of reducing sugars by reductive amination (left),¹³¹ or reaction with imido thioglycosides derivatives (right) to create neoglycoproteins.¹⁴¹

1.4.2.2 Towards site selectivity by glycosylation of cysteine

Cysteine (Cys) residues have the lowest abundance of any amino acids within a protein at approximately 2% in mammals,¹⁴⁴ and are subsequently rarely found on the surface of proteins, making cysteine particularly useful for site-specific modification. With the aim of mimicking natural glycosylation, iodoacetamide glycoconjugates (figure 1.17) were reported by Flitsch and co-workers for the attachment of N-acetylglucosamine to Bovine serum albumin (BSA).¹⁴⁵ Iodoacetamides were later adopted for the generation of glycopeptides with larger oligosaccharide structures.¹⁴⁶ Alternatively, Davis and co-workers reported a combined site-directed mutagenesis (SDM) and chemical attachment approach to creating disulfide-linked glycoproteins in a single glycoform.¹⁴⁷⁻¹⁵⁰ A mutant protein containing a single cysteine was produced, to which thioglycosides were attached using glycomemethanethiosulfonate (MSH), to form a disulfide bridge following thiol exchange (figure 1.17). In a similar fashion, Boons and co-workers exploited dithiopyridyl glycosides to create disulfide-linked glycoproteins.¹⁵¹ The disulfide link is a reversible conjugation, and can be vulnerable in reducing environments. An alternative,

non-reversible conjugation method is to use maleimido-functionalised sugars for covalent attachment at cysteine.¹⁵²

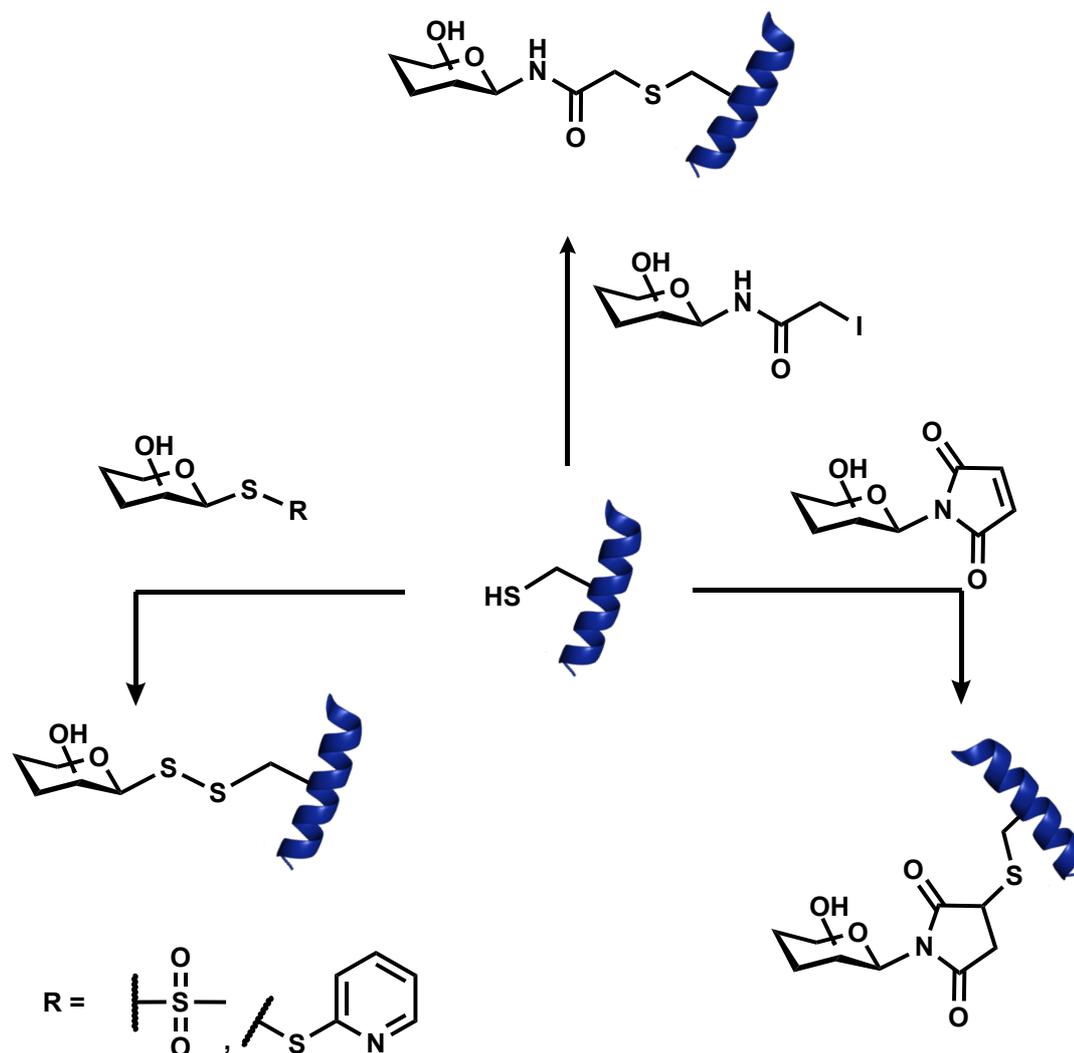


Figure 1.17: Examples of glycosylation at cysteine residues on the surface of protein through reaction with iodoacetamides,¹⁴⁵ maleimides,¹⁵² and methane thiosulfonates.¹⁴⁸

Despite cysteine being a very useful target for site-specific glycosylation and proving a superior method of developing homogenous neoglycoproteins than attachment to lysine residues, the mutagenesis required to introduce surface exposed cysteines is not compatible with all proteins. If careful consideration to avoid incorrect disulfide bond formation is not undertaken, the introduction or removal of cysteines within protein structures can severely affect protein folding, stability and solubility.

1.4.3 Chemical mutagenesis and non-canonical amino acid targets for glycosylation

As chemical biology has advanced, so too have the tools available to perturb biological systems. A variety of techniques, nicely reviewed by Lang and Chin, display the vast

range of chemical tools presently available to chemical biologists.¹⁵³ Not only can reactive side chains be introduced through expansion of the genetic code, with techniques such as amber suppression or isosteric replacement of amino acids, there have been numerous examples of the chemical modification of amino acids to create new reactive groups termed post-translational or chemical mutagenesis.¹⁵⁴

One of the most commonly utilised chemical modifications is the generation of a dehydroalanine (Dha). First developed as a route to convert a serine protease to a cysteine protease,^{155, 156} it was then reported by Schultz within proteins as a precursor for chemical mutagenesis.¹⁵⁷ From its early reports, dehydroalanine has been generated from a number of precursors, a few of which are shown in figure 1.18. Two examples illustrated are the conversion cysteine by oxidative elimination with *O*-mesitylenesulfonyl hydroxylamine^{158, 159} or elimination of a cyclic sulfonium formed through reaction with methyl dibromovalerate,¹⁶⁰ although a number of other methods have also been developed for the conversion of cysteines to Dha in peptides and proteins, and have been reviewed by Chalker *et al.*¹⁶¹ Alkyl- and arylselenocysteines are capable of undergoing β -elimination upon oxidation to introduce an alkene side chain;^{157, 162} from phosphoserine the elimination of phosphonate under alkali conditions.¹⁶³ The α,β -unsaturated carbonyl of Dha has become a tool for protein modification able to undergo Michael addition reactions with thiols and amines, but also radical addition reactions in the presence of zinc.¹⁶⁴⁻¹⁶⁶

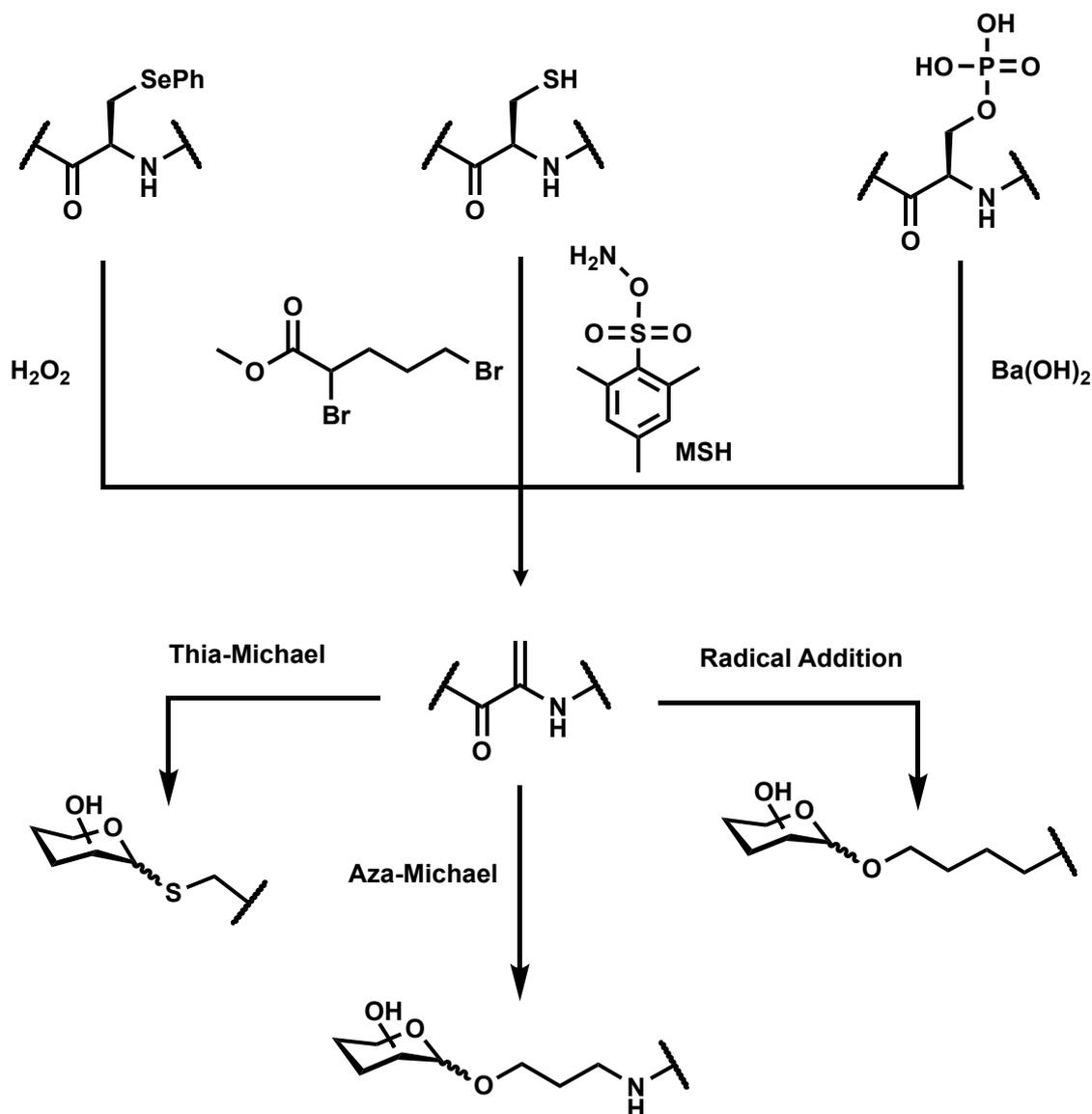


Figure 1.18: Natural and unnatural amino acid precursors to dehydroalanine, and their reactions for chemical mutagenesis to yield dehydroalanine within proteins. Dehydroalanine can then be used as a chemical handle for site selective glycosylation.

The use of dehydroalanine is not without its drawbacks – it is vulnerable to Michael-addition reactions with nucleophiles such as thiols like glutathione,¹⁶⁶ and a lack of stereo-control in conjugation reactions.

1.4.3.1 Expansion of the genetic code and bioorthogonal chemistry

New chemical reactivities can be introduced into biological molecules by global residue specific incorporation, using isosteric replacement of canonical amino acids,¹⁶⁷ and by site specific incorporation through amber suppression.¹⁶⁸ Residue-specific modification has been displayed in a simple pictorial representation in figure 1.19. This allows for the complete replacement of one amino acid at every position it is coded for in the genetic script, by its non-canonical amino acid (NCAA) analogue.¹⁶⁹ A non-canonical amino acid

is a synthetic amino acid derivative which have chemical functionalities not found in nature, for the metabolic labelling of proteins. The desired NCAA must be a substrate for the natural tRNA synthetase in order to hijack cellular translational machinery, with the tRNA synthetase for methionine being the most commonly used; although other modified amino acids e.g. fluorinated analogues have been used to mimic natural substrates such as leucine, tryptophan, phenylalanine and proline.¹⁷⁰⁻¹⁷²

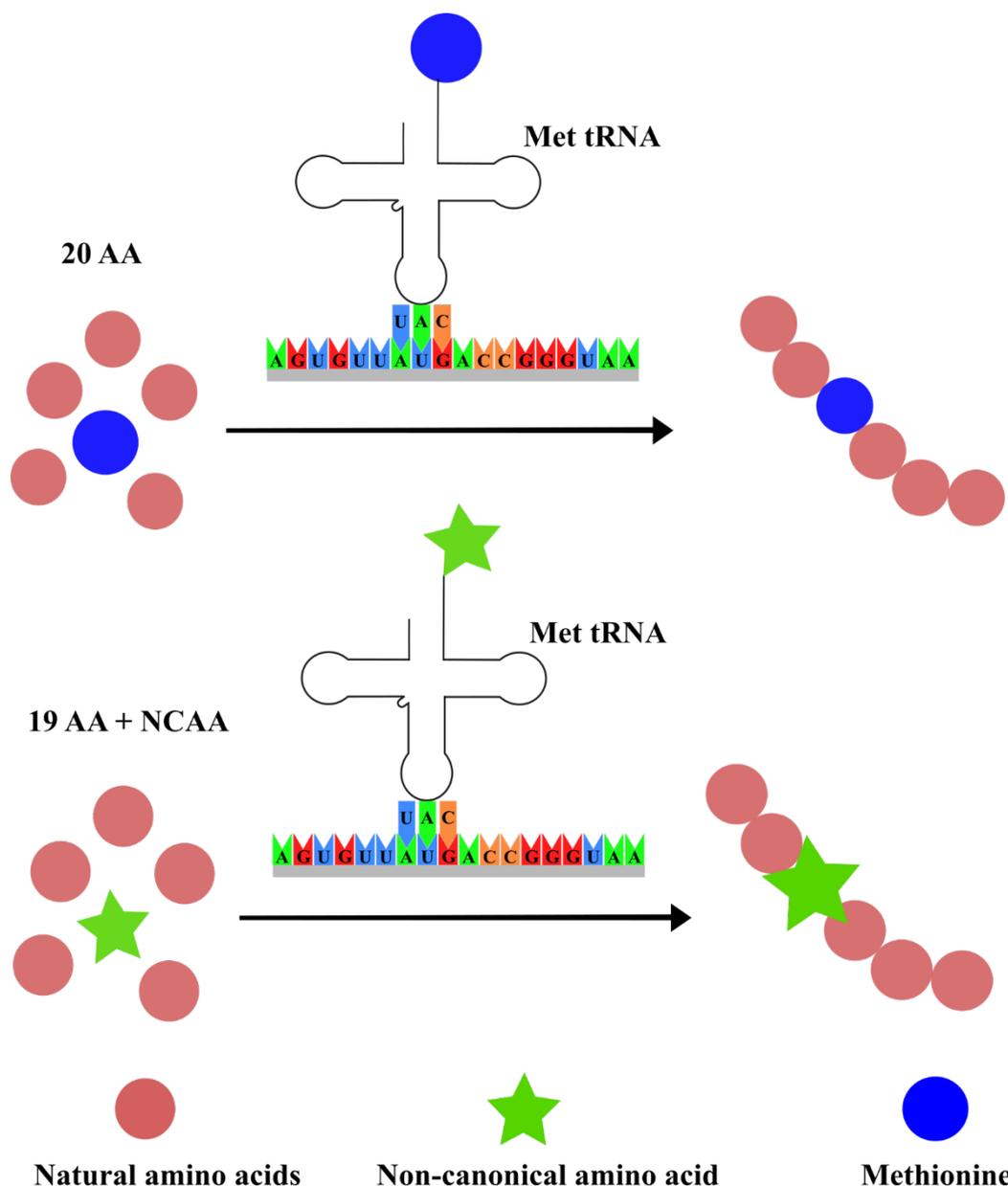


Figure 1.19: The incorporation of an NCAA methionine analogue which is a substrate for the Met tRNA synthetase, allowing it to be incorporated into the polypeptide when methionine is not present in the growth media.

Alternatively, site-specific methods developed by Schultz, employ mutagenic DNA sequences to allow introduce a NCAA in the response to the recognition of the stop codon UAG.¹⁷³ This technique requires expansion of the genetic code, generating a tRNA synthetase that will recognise this stop codon, which would naturally result in termination

of protein synthesis. Mutagenesis techniques can then place this codon at any site within the global protein structure. This methodology was later applied within *E. coli* cells via tRNA/aminoacyl tRNA synthetase pair unrecognised by the bacteria, demonstrating a generic *in vivo* methodology for modification.¹⁷⁴ This technique requires great endeavour in the synthesis of a mutated plasmid and an orthogonal tRNA/tRNA synthetase pair for the NCAA.¹⁷⁵

Using these techniques, a number of bio-isosteric analogues of natural amino acids have been used to introduce different chemical handles into proteins, including azides,^{176, 177} alkynes,^{169, 178-180} aldehydes,^{181, 182} among others.^{153, 183, 184-186}

1.4.3.2 Chemical glycosylation reactions exploiting non-canonical amino acids

The introduction of non-canonical amino acids opened up the use of new bioorthogonal chemistries, which were previously inaccessible for use on proteins. The term bioorthogonal chemistry was first coined by Bertozzi in 2003¹⁸⁷ and is defined as two reactive groups which selectively react with each other but are inert to any off-target reactions with biomolecules.^{187, 188} Bioorthogonal reactions have emerged as very useful methods for perturbing biological systems, particularly by incorporating reactive groups through biosynthetic pathways. Although, certain prerequisites must be fulfilled to be useful in cellular environs: the reactants must be biologically inert, but the products must not be chemically reactive; specific reactive partners with no off-target reactivity; and the kinetics of conjugation must be sufficiently rapid to avoid metabolic clearing.¹⁸⁹

An early example falling into the bioorthogonal category is the Staudinger ligation, a reaction between phosphines and azides.^{190, 191} However, a more commonly known and widely used reaction is the copper-catalysed azide-alkyne cycloaddition (CuAAC),¹⁹² colloquially known as the ‘click’ reaction, which has been a real impetus for chemical biology, and protein modification. The CuAAC is in keeping with the prerequisites of a bioorthogonal reaction, having selective reaction partners, stable products and few off target reactions. The term ‘click’ introduced by K. B. Sharpless is given due to the reaction’s fast kinetics, high atom efficiency, high yields and production of a single stereoisomer.¹⁹³ The simple requirements of a terminal alkyne, an azide and a copper catalyst to carry out the 1,3 dipolar cycloaddition, broadened the scope of substrates, making bioconjugation more accessible. CuAAC quickly became a staple reaction for bioconjugation, and many have used the reaction in the synthesis of neoglycoproteins. Davis and co-workers reported the CuAAC conjugation as part of the ‘tag and modify’ approach for the synthetic glycosylation of proteins, in which an azide or alkyne is first

introduced into the protein as a reactive handle, and CuAAC is used to modify the mutant protein with an azido or alkynyl carbohydrate (figure 1.20).¹⁶⁵ Examples of site-selective CuAAC glycosylation include synthesis of glycoprotein mimetics,¹⁹⁴ fluorinated glycoproteins¹⁹⁵ and virus-like glycodendrons.¹⁹⁶ This approach has also been applied for the synthesis of glycoprotein vaccines to create well defined homogeneous glycoforms.¹⁹⁷

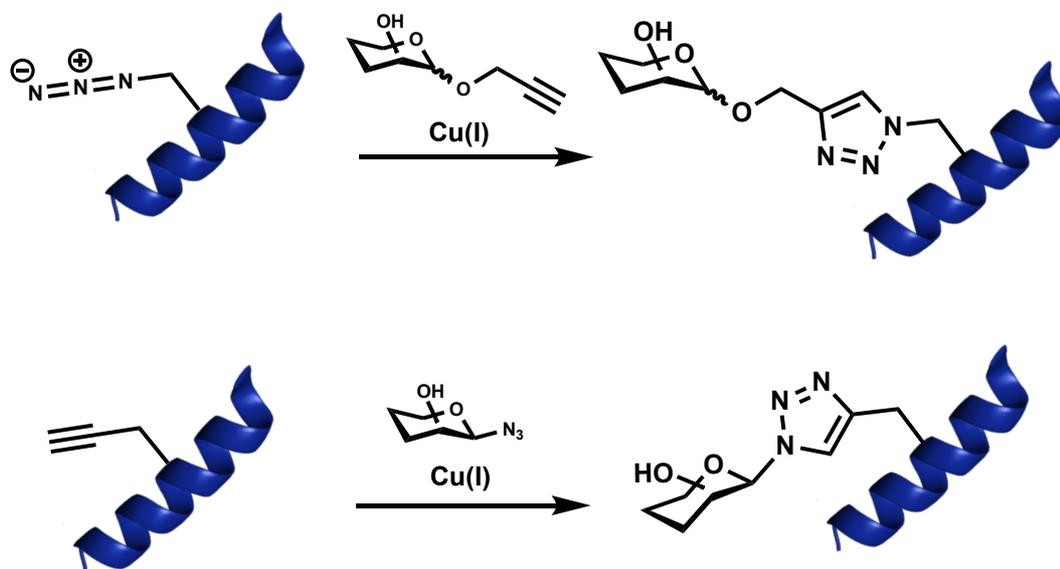


Figure 1.20: Examples of chemical glycosylation through CuAAC mediated bioconjugation on proteins which contain unnatural amino acids introduced through expansion of the genetic code.

Berkland and co-workers have reviewed the key considerations when carrying out azide-alkyne bioconjugation.¹⁹⁸ Particular drawbacks include: metal binding/copper chelation by the protein; formation of reactive oxygen species leading to oxidation of cysteine, methionine, histidine and asparagine residues within proteins¹⁹⁹; and induced cellular toxicity in living systems.

Building upon CuAAC, the development of strain-promoted azide-alkyne cycloaddition (SPAAC) reaction and the inverse electron demand Diels Alder (IEDDA) removed the need for the problematic metal catalyst.²⁰⁰ These reactions have been used in labelling studies, particularly *in vitro*, but have seen surprisingly low use in creation of synthetic biomacromolecules. Machida *et al.* reported a double ligation strategy using CuAAC to ‘click’ on a tetrazine bioorthogonal handle to glycans, which was then used to append the glycans to a strained alkyne and alkene-containing amino acid, by IEDDA for *in cellulo* ligations.²⁰¹

Oxime ligation is another bioorthogonal method adopted in the synthesis of glycoproteins and glycopeptides. Oxime ligation has been reported as a method for the modification of galactose containing glycopeptides with native glycosidic linkages (figure 1.21A), using

galactose oxidase enzymes to introduce an aldehyde at the 6-position of the monosaccharide, which can then react with synthetic oligosaccharides.²⁰² Alternatively, many similar glycopeptide structures have been made by introducing an oxyamine into either the oligosaccharide or the peptide sequence, followed by oxime ligation to an aldehyde/ketone or reducing sugar respectively (figure 1.21 B/C), produced the glycopeptide mimetics containing oxime linkages.²⁰³⁻²⁰⁶

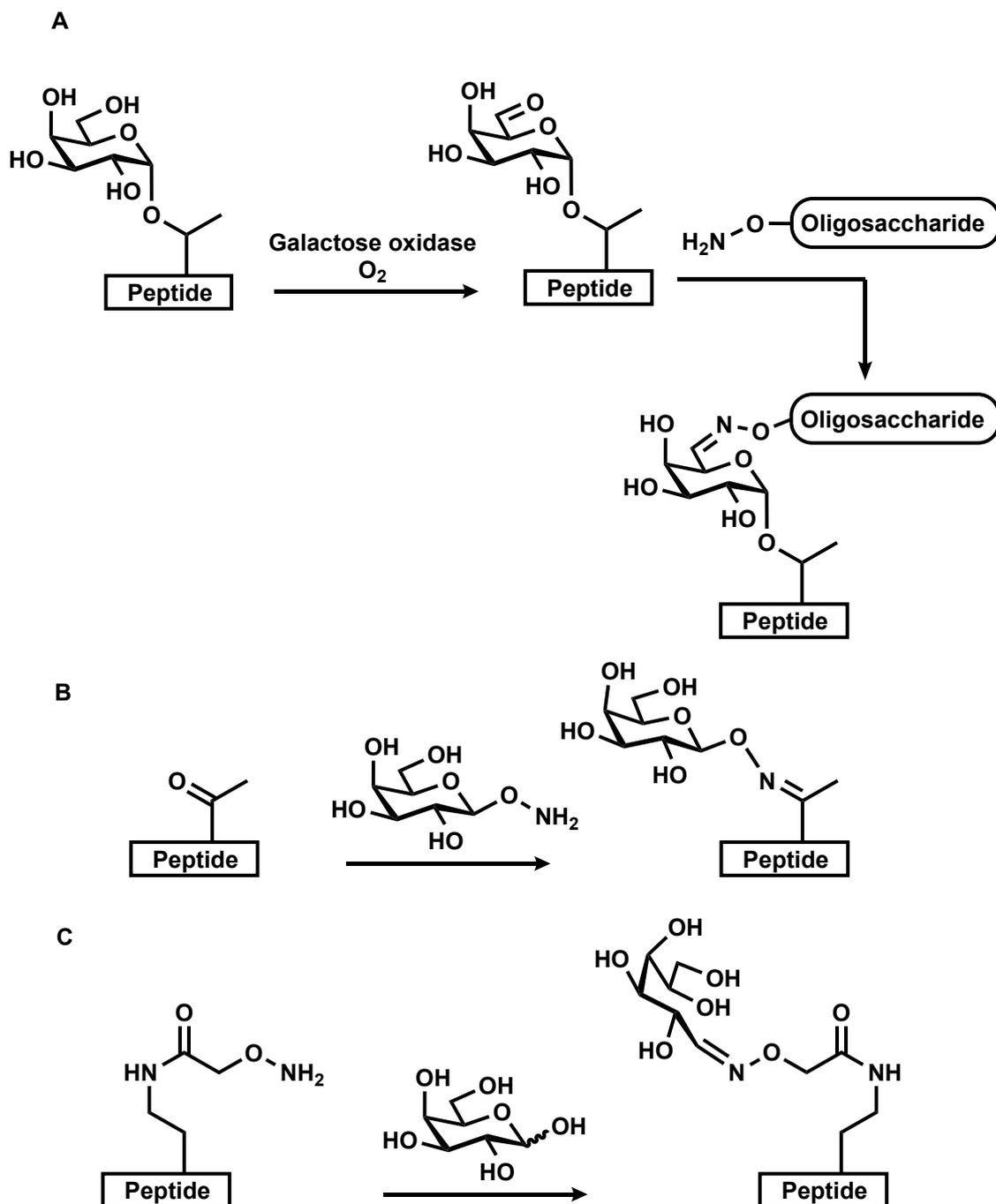


Figure 1.21: Oxime ligation strategies used for the production of glycopeptide mimetics, using chemical and chemoenzymatic approaches.

As previously mentioned, aldehydes have been introduced into proteins and have been exploited for the modification of proteins with glycans. Bertozzi and co-workers

described the conversion of cysteine residues to formylglycine in proteins containing an N-terminal extension using formylglycine-generating enzymes (FGE), allowing the ligation to a variety of tags including amino-oxy glycans to the aldehyde.^{207,208} Using this post-translational conversion, homogeneous glycoforms of IgG antibody were produced using amino-oxy GlcNAc followed by enzymatic glycan extension.²⁰⁹ Bertozzi also outlined versatile chemical and chemoenzymatic methods for the synthesis of amino-oxy glycans such as amino-oxy SLe^x, using α -bromo and trifluoroacetimidate oligosaccharides to introduce a protected oxyamine in the form of *N*-pentenoyl hydroxamic acid, for attachment to human growth hormone.²¹⁰ Furthermore, the genetic incorporation of aldehydes and ketones has led to the use of oximes and hydrazides for the modification of proteins. Schultz demonstrated that ketone functions could be introduced into recombinant proteins in *E. coli* through *p*-acetylphenylalanine, the ketone was then reacted with amino-oxy GlcNAc and a series of transferases used for glycan extension to yield homogenous synthetic glycoprotein mimetics.^{211,212}

1.4.3.3 Enzymatic synthesis of homogeneous glycoproteins

Homogeneous synthetic glycoproteins have also been produced by enzymatic remodelling of glycans on the surface of glycoproteins. The endoglycosidase family of enzymes are responsible for trimming of large glycans off proteins, however, some endoglycosidases have, or have been engineered to have, transglycosidase activity, allowing oxazoline analogues of large glycans to be recognised and attached to proteins. Davis and co-workers reported the use of endoglycosidases with oxazoline disaccharide donors, to construct S-, S-S and triazole linked trisaccharide containing glycoproteins by first introducing a GalNAc synthetically, followed by *en bloc* attachment of the large glycan.²¹³ Fairbanks and co-workers used the same EndoA catalysed approach with larger oxazoline donors that had been chemically synthesised, showing transglycosylation is a robust method for developing complex homogeneous glycoproteins.²¹⁴ Many others have exploited the native glycosidase activity in conjunction with the transglycosidase activity of the endoglycosidase family of enzymes, to trim large glycans from isolated heterogeneous glycoproteins, and remodel them with a single glycan, giving a single glycoform (figure 1.22).²¹⁵⁻²¹⁷ The enzymatic activity of the endoglycosidase family, and their use in glycoprotein synthesis, has recently been summarised in a review by Fairbanks.²¹⁸

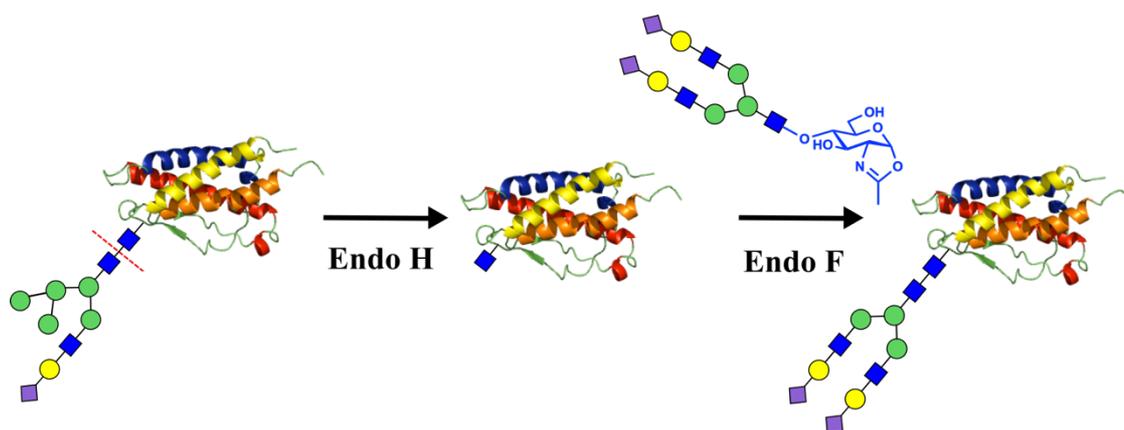


Figure 1.22: Enzymatic remodelling of glycoproteins using ENGases and oxazoline donor glycans to create new homogeneous glycoproteins.

The specificity of glycosyltransferases has also often been exploited for the synthesis of oligosaccharides, with the control of substrate and stereochemistry often simplifying the synthesis of complex oligosaccharides. The diversity that can be generated from the use of glycosyl transferases has been best demonstrated in the synthesis of glycopeptides, ranging from simple monosaccharides to a variety of very complex cores or large repeating glycan units that can mimic both N- and O- glycosylation; examples include such as C34 glycopeptide derived from gp41 known to be involved in HIC infection,²¹⁹ and MUC1 synthetic glycopeptide mimetics.^{219-221, 222}

1.5 Rationale for the project and objectives

This work aims to build upon the generation of neoglycoprotein inhibitors of bacterial toxins reported by Branson *et al.*¹²⁸, by expanding the chemical toolbox for the synthetic glycosylation of proteins. After reviewing many effective methods for the site-selective synthesis of homogeneous glycoproteins, there is still space for new methods which can functionalise reducing sugars, removing the need for chemical synthesis of complex oligosaccharides.

The work of Branson *et al.*¹²⁸ illustrated that site-specific glycosylation of a protein scaffold matching the size and symmetry of the target lectin, is an effective strategy for developing inhibitors. This has been supported by a theoretical model rationalising ligand design for multivalent systems.²²³ The effectiveness of the neoglycoprotein based inhibitors then posed further questions surrounding the linker length in the inhibitor design, the long flexible linker used results in loss of conformational entropy upon binding – therefore could a shorter, more rigid linker affect binding/ inhibitory potential?

Secondly, does the site of ligand attachment have an impact on affinity/inhibitory potential? Could the topography of the protein surface around the site of glycosylation alter the long-range electrostatic interactions, between neoglycoprotein and target lectin? Finally, is it possible to develop a method for synthetic glycosylation that can be used to attach natural underivatized oligosaccharides to specific sites on proteins?

To address these questions, the principal aim of the project was to design a new bifunctional linker for the attachment of carbohydrates to proteins. This strategy for development of neoglycoproteins has been represented as a simple cartoon in figure 1.23, groups A and C signify two different bioorthogonal functional groups. This linker will contain two different bioorthogonally reactive groups which will be used to derivatise both unfunctionalised oligosaccharides and chemically modified oligosaccharides, to produce reactive glycan derivatives. The second bioorthogonal functional group is then left for the site-selective glycosylation of proteins to produce homogenous glycoproteins.

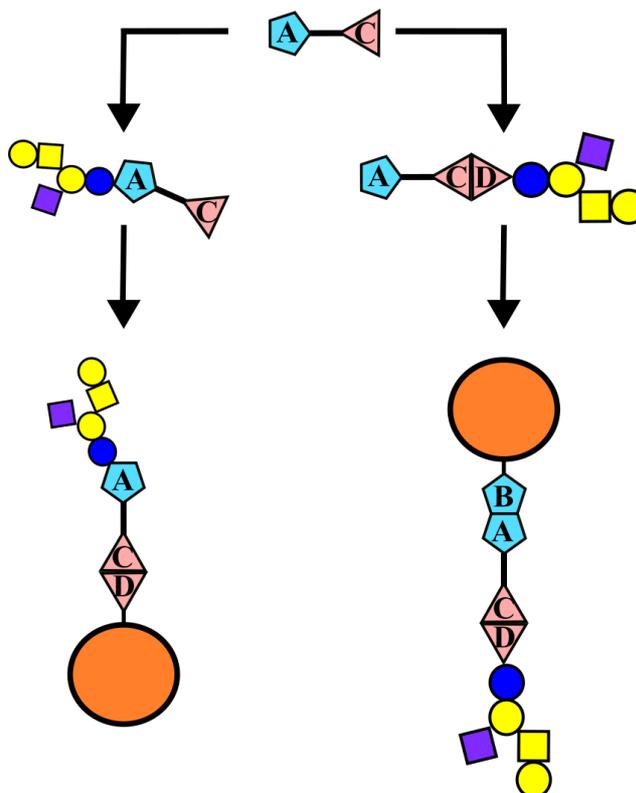


Figure 1.23: Cartoon representation outlining the project aim in creating a bifunctional linker capable of functionalising reducing sugars or chemically modified oligosaccharides, and site-selectively attaching glycans to proteins.

Enzymatic elaboration of simple glycan derivatives will be incorporated to construct complex functionalised oligosaccharides that can be applied to the preparation of synthetic glycoproteins. With this new linker, I will then investigate the effects on linker length and site of glycosylation on the potency of the neoglycoprotein-based inhibitors of the cholera toxin. I will then be able to apply these methods for the development of

different neoglycoprotein inhibitors against various clinically relevant lectins and test their efficacy as inhibitors using enzyme-linked lectin assays. Beyond the development of neoglycoprotein inhibitors, some of these concepts and tools for the development of homogeneous glycoproteins have a place outside of this work in other therapeutic areas such as glycoconjugate vaccines and glycoprotein therapeutics.

1.5.1 Crosslinking agents for protein conjugation

Generally crosslinking reagents consist of two reactive groups, which each undergo reaction to link two molecules together and are termed bifunctional reagents. Bifunctional linkers are themselves categorised into two groups, described as homo- or heterobifunctional having two of the same or two different reactive groups.

It is inherently difficult to control bioconjugate formation when using homobifunctional linkers, as they react towards the same type of reactive group twice. Intramolecular crosslinking is a major drawback, with the possibility of intermolecular polymerisation creating large crosslinked polymeric systems and supramolecular aggregates. This can somewhat be controlled by using two-step conjugation methods i.e. reaction of one molecule with excess crosslinker before reaction with the second molecule. While this offers greater control than single step one-pot conjugation reactions, it requires removal of excess crosslinking reagent and often extensive chromatographic purification to isolate the desired homogeneous protein conjugates. Despite difficulties in generating homogeneous conjugates, homobifunctional reagents are still widely used as a fast and simple technique for conjugation reactions.

Heterobifunctional linkers, by definition, combine two different reactive groups described previously to elicit greater control over conjugation. As heterobifunctional crosslinking reagents cannot react with multiple copies of the same reactive group, polymerisation of their target biomolecules is limited. Heterobifunctional reagents are therefore much better options for obtaining homogeneous conjugates.

Naturally occurring reactive groups present in amino acids side chains, such as amines (Lys), carboxylic acids (Asp, Glu), thiols (Cys) and alcohols (Ser, Thr, Tyr), were used in early conjugation methodologies, before protein engineering was possible, but are still routine targets for conjugation.

1.5.2 Amine reactive linkers

Amines are the most common target for functionalisation due to the high natural abundance on the surface of biological macromolecules. Aldehydes target amine residues such as in reductive amination modifications (section 1.4.2.1), have been used to successfully produce a number of synthetic glycoproteins in the form of glycoconjugate vaccines. The highly reactive nature of aldehydes makes them useful tools for attaching lots of glycans, but not for site-specific conjugation. This reactivity results in a loss of control. Issues surrounding aldehydes as crosslinkers include multiple reductive alkylation on one amine, and a low risk of linking two protein molecules a risk more relevant when using simple aldehydes at high concentrations.

In contrast to aldehyde-based linkers, N-hydroxysuccinimide (NHS) ester crosslinking reagents give greater control in that amide formation can only occur once. The reaction between NHS esters and amines is known to be fast even at physiological pH, and although reaction with other nucleophiles like thiols and alcohols is possible, these (thio)ester products are hydrolysed relatively easily in water. NHS esters can therefore be thought of as chemoselective for amine side chains. Water soluble sulfated NHS analogues, pioneered by Stravos and co-workers, allowed these linkers to be used in buffer without presolubilisation, reflected by the structures of heterobifunctional linkers in figure 1.24.²²⁴ Although NHS esters are powerful chemical tools that are widely used within the literature, many of them have limited half-lives in buffer, and offer very little control in crosslinking except for systems with very few lysine residues.^{225, 226}

1.5.3 Thiol-reactive linkers

As cysteine residues are less abundant on the surface of proteins, targeting cysteine thiols can give greater selectivity for protein modification at specific sites. The unique reactivities of sulfur have thus been exploited to elicit control over protein conjugation, including the propensity of thiols to form disulfides and undergo conjugate addition reactions.

Mixed disulfide exchange with thiol activating groups such as dithiopyridyl have been introduced into heterobifunctional linkers.²²⁷ This technique is, however, limited by its sensitivity to reducing environments (such as in the cytosol), or the presence of other thiol containing molecules. Maleimide structures have been used since the 1950s as non-cleavable crosslinkers that form stable thioether conjugates.²²⁸ Maleimides are recognised

as powerful crosslinking agents due to rapid reactivity towards thiols, at slightly acidic to neutral pH.²²⁹

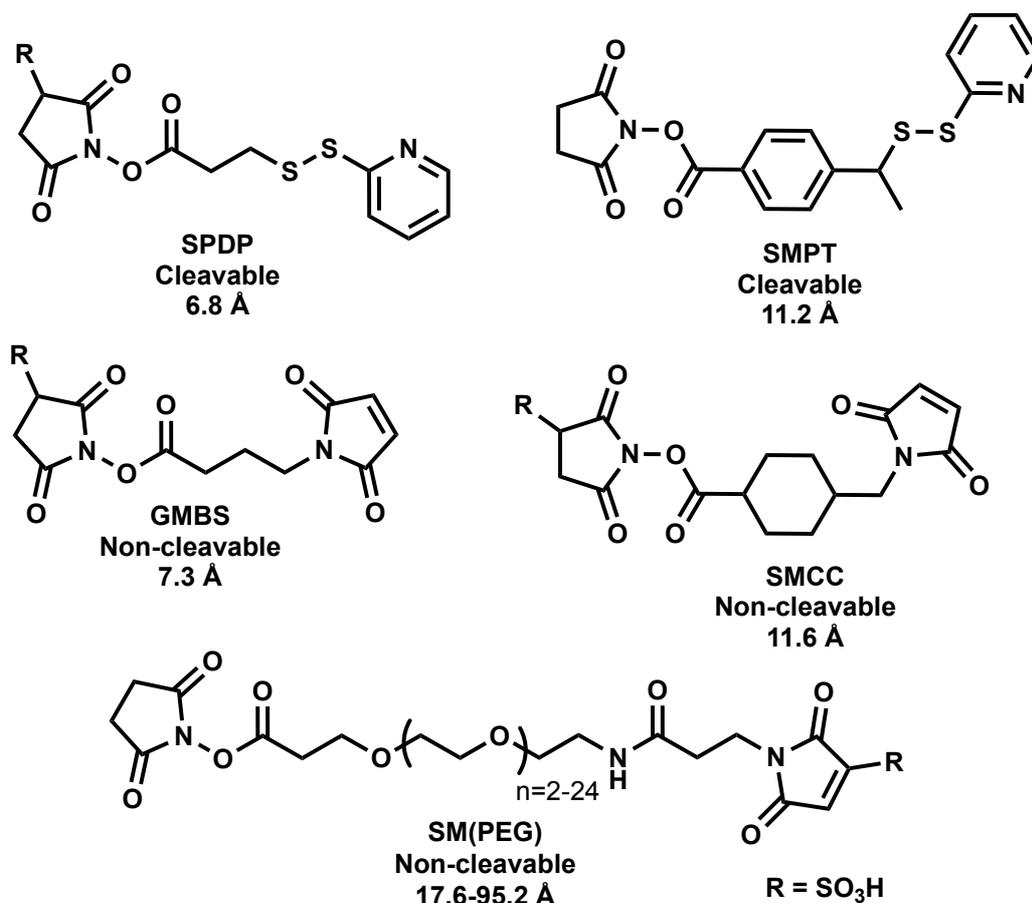


Figure 1.24: Structures of heterobifunctional crosslinking agents combining reactive groups targeting lysine residues (NHS ester) and cysteine residues (thiopyridyl or maleimide) for the crosslinking or labelling of biomolecules. Linkers SPDP (succinimidyl 3-(2-pyridyldithio)propionate) and SMPT ((4-succinimidyl)oxycarbonyl- α -methyl- α -(2-pyridyldithio)toluene) target thiols through disulfide exchange to create cleavable S-S bonds to a biomolecule or labelling reagent. Linkers GMBS (N- γ -maleimidobutyryloxysuccinimide ester), SMCC (Succinimidyl-trans-4-(N-maleimidylmethyl)cyclohexane-1-carboxylate) and SM(PEG) (PEGylated SMCC) are examples of covalent linkers, permanently crosslinking or labelling biomolecules.

The combination of amine and thiol reactive targets has become commonplace for protein conjugation. It is therefore unsurprising that the two most frequently used reactive groups used in homobifunctional linkers are NHS esters and maleimide, although, alternative combinations such as NHS ester and thiopyridyl introduce a cleavable heterobifunctional linker (figure 1.24).²³⁰ Many variations of such crosslinkers are commercially available with wide diversity in spacer structure and length; for example ThermoFisher offers linkers in the range of 4.4 to 95.2 Å. The amalgamation of fast, selective reaction with the target, and aqueous stability of the covalent conjugates, has made these type of linkers particularly useful in a variety of applications from modification of surfaces to the creation of antibody drug conjugates.²³¹⁻²³³

In the context of affixing glycans to the surface of protein scaffolds, as is the objective of this project, the heterobifunctional linkers described are unsuitable, amines are not selective enough and the introduction of additional cysteines is not practical. This reinforces the need for new methods for glycan functionalisation and the introduction of new chemical functionalities through modification of existing amino acids or introduction of unnatural amino acids with bioorthogonal groups. After successful ligation of GM1 ligands to a bacterial scaffold by Branson *et al.*¹²⁸ using N-terminal oxidation and oxime ligation, we decided to build on this previous work continuing to use aldehydes. This bioorthogonal strategy is used sparingly within the literature, despite being highly efficient. A comprehensive review highlighted the various chemical and enzymatic methods for introducing aldehydes into proteins²³⁴: including oxidative cleavage of Ser/Thr (Malaprade reaction)²³⁵; transamination; use of lipoic acid ligase²³⁶ or genetic incorporation of aldehyde tags.^{207, 237} Conjugation methods with aldehydes shown in figure 1.25 include oxime/hydrazone ligation²³⁸, Pictet-Spengler²³⁹ and Wittig²⁴⁰ as a few examples of reactions performed with aldehyde handles in proteins. Oxime/hydrazine ligation has been shown to be effective with all reported aldehydes in proteins, as well as being used extensively for directly functionalising reducing sugars, reacting at the anomeric centre.

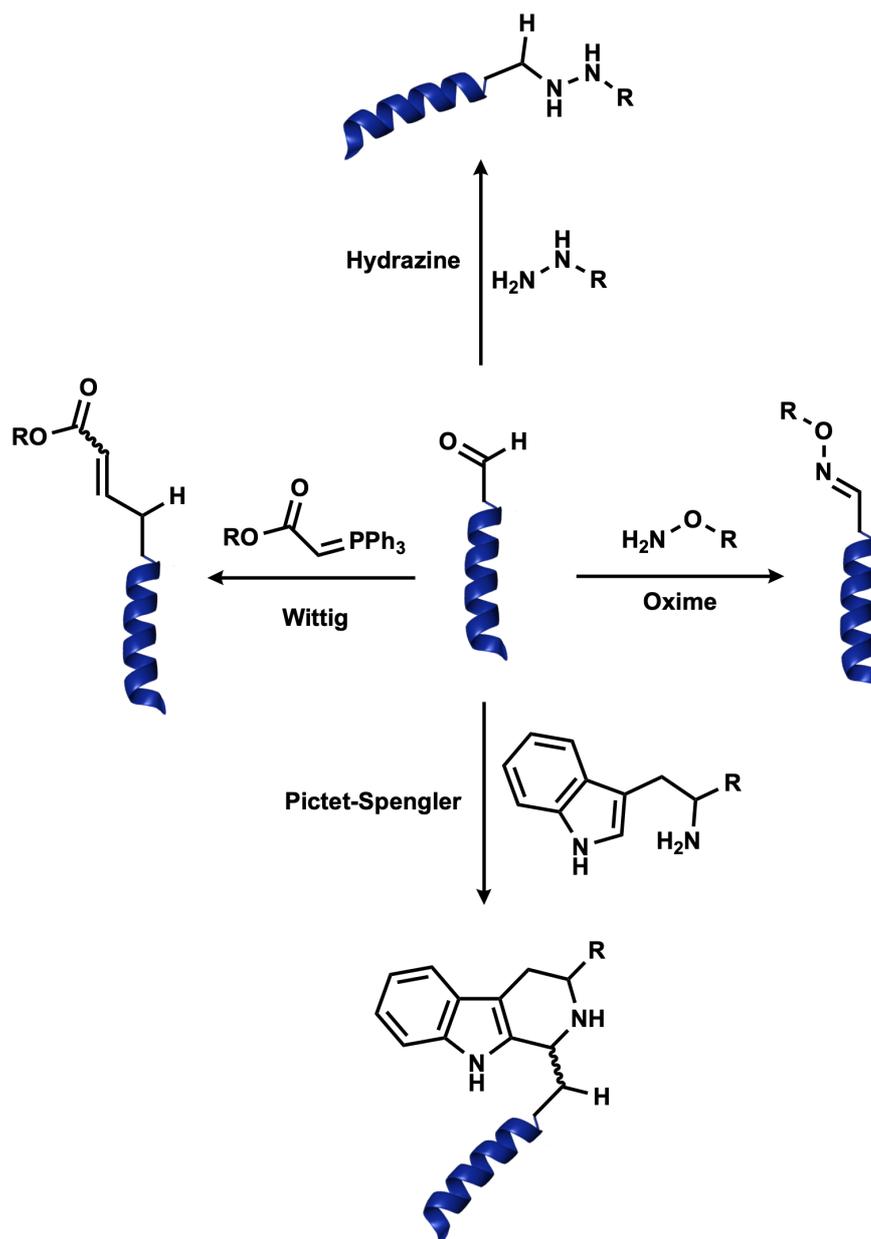
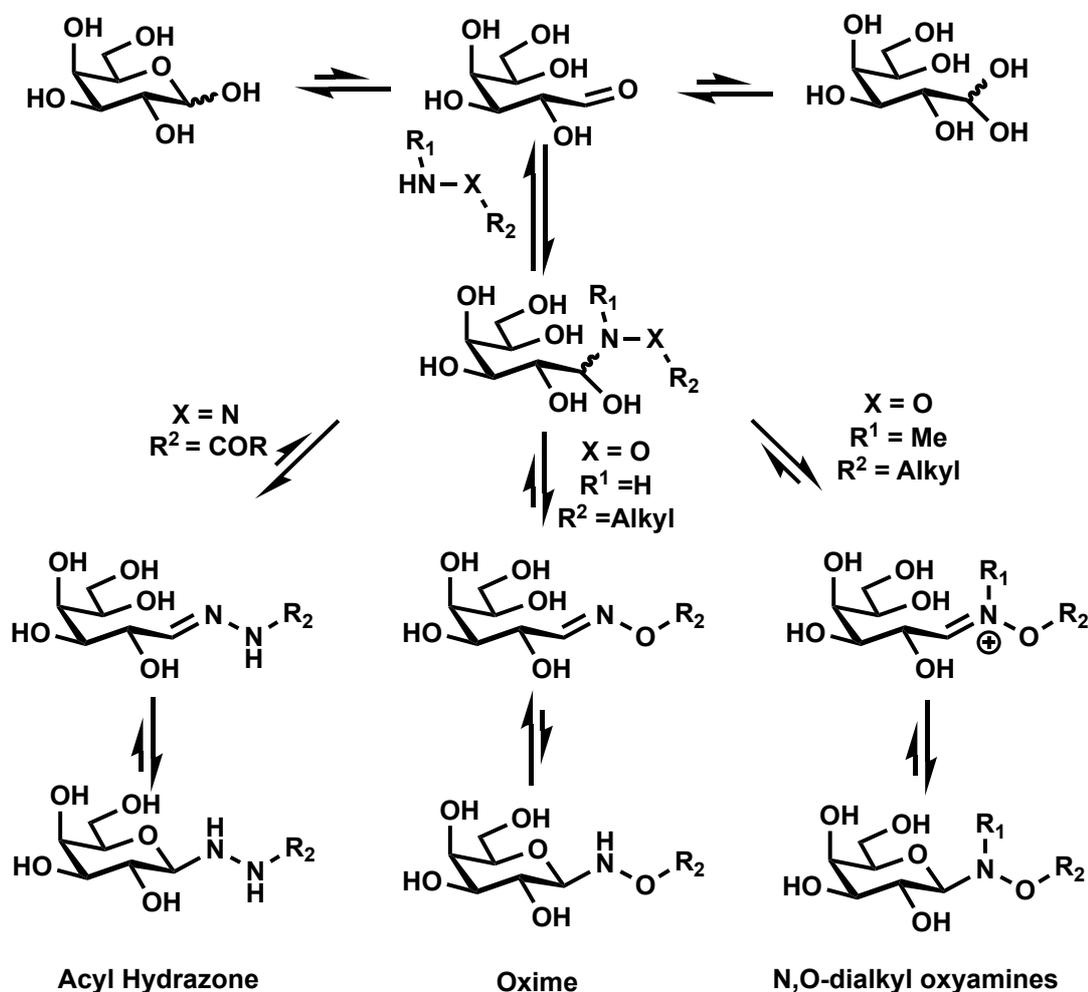


Figure 1.25: Examples of protein conjugation to aldehyde groups introduced into the protein, here they are shown as an N-terminal ligation following oxidation of a Ser/Thr.

1.5.4 Oxime and hydrazone ligation of reducing sugars

The reaction of reducing sugars with oximes/hydrazides has been known since reports by Emil Fischer in 1888²⁴¹, that has been widely adopted owing to its chemoselectivity at the anomeric position. Oximes and hydrazones are highly stable to hydrolysis at neutral pH²⁴², the stability of the C=N bond is enhanced by presence of the α -heteroatom. On this point, different arguments have been proposed; that the pKa of the nitrogen is altered due the charge distribution of the lone pair²⁴³, and is less likely to be protonated, whereas, Raines *et al.* proposed that the α -effect of the heteroatom is primarily responsible for stability.²⁴⁴ Experimental evidence for oximes having greater stability than hydrazones, are in support of the α -effect argument, with the more electronegative oxygen stabilising

the C-N-X bond better than an α -nitrogen. The stability of oxime-ligated glycoconjugates makes them useful for protein conjugation, as they are less susceptible to hydrolysis at biological pH. The reaction pathway for oxime/hydrazide ligation is illustrated in scheme 1.1, the reaction first requires the formation of the open chain aldehyde intermediate, which is intercepted by the nucleophile. The formation of the aldehyde has been identified as the rate-limiting step, as it is dependent on the equilibrium responsible for the carbohydrate undergoing mutarotation in solution which is highly variable between different carbohydrates. Reactions are often performed at pH 4-5 representing the optimal balance between nucleophilic attack, dehydration of the hemiaminal, and ring opening of reducing sugars.²⁴⁵ Upon reaction, the products do exist in different conformations with the oxyamine predominately in the ring-open conformation, whereas the hydrazide/N,O-alkyl oxyamine favour the ring-closed confirmation as the N-linked glycoside.



Scheme 1.1: Reaction pathway of D-galactose forming glycosyl oximes and N-glycosides through oxime/hydrazine ligation.²⁴⁵

The rate of this reaction has been enhanced with the use of an organocatalyst, in which aniline has been the most commonly used organocatalyst for glycan oxime ligation. It was first reported by Jensen and co-workers to give a 20-fold increase in reaction rate

when used at a concentration of 100 mM, by increasing the formation rate of glycosyl imine.²⁴⁶ Since that report, aniline has been routinely used for the oxime ligations of reducing sugars, and also oxidised proteins^{128, 247, 248}, and is particularly useful for reactions employing carbohydrates whose reactivity is low because of their rates of mutarotation.²⁴⁹ The choice of catalysts, reaction media and pH have been analysed systemically and shown to play an important role in the kinetics of ring opening to the aldehyde, and resulting oxime formation through the iminium ion intermediate.²⁵⁰ More recent research has reported many other substituted anilines (figure 1.26), which differ in pKa and have different catalytic efficiencies at specific pH's in the range of 4-7.²⁵¹⁻²⁵⁴

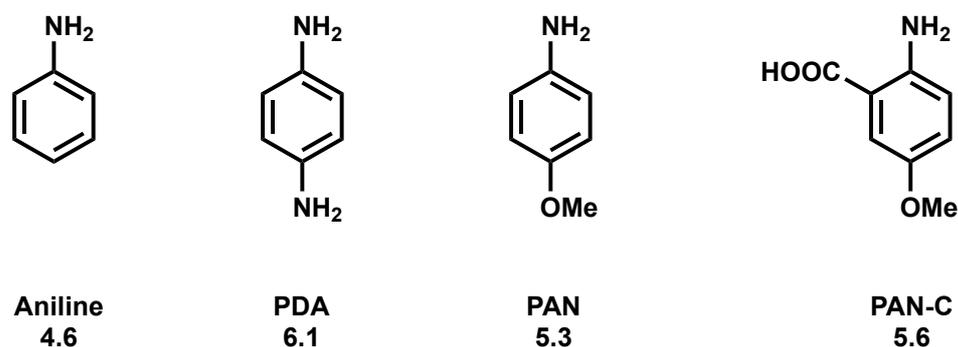


Figure 1.26: Structure of aniline and three substituted anilines reported for the organocatalysis of oxime/hydrazide ligation of carbohydrate and their corresponding pKa values.

1.6 Combining functional groups for a new bifunctional linker

Oxime/hydrazine ligation is a method capable of functionalising both carbohydrates and proteins, ideal for the purposes of synthesising glycoproteins, the foundation of this project. It would not be untoward in thinking bifunctional linkers using this reactivity would have been used for the synthesis of neoglycoproteins, however, bifunctional linkers are more common for applications such as surface modification.²⁵⁵ Commercial bifunctional crosslinkers available that are reactive towards unmodified carbohydrates, a few of which are displayed in figure 1.27, all contain a hydrazide in combination with maleimide, and have not been applied for the purpose of producing neoglycoproteins. The few linkers available for functionalising carbohydrates combine maleimide and thiopyridyl for the attachment of cysteine. As discussed in section 1.5.3 engineering and targeting cysteines in proteins for synthetic purposes is often difficult and requires careful consideration in cysteine positioning. These examples of heterobifunctional linkers (figure 1.27) highlights the need for new chemical tools for synthetic glycosylation, with no linkers currently available which have groups that can react with unfunctionalised

carbohydrates whilst combining the advantages of highly site specific, bioorthogonal reactions.

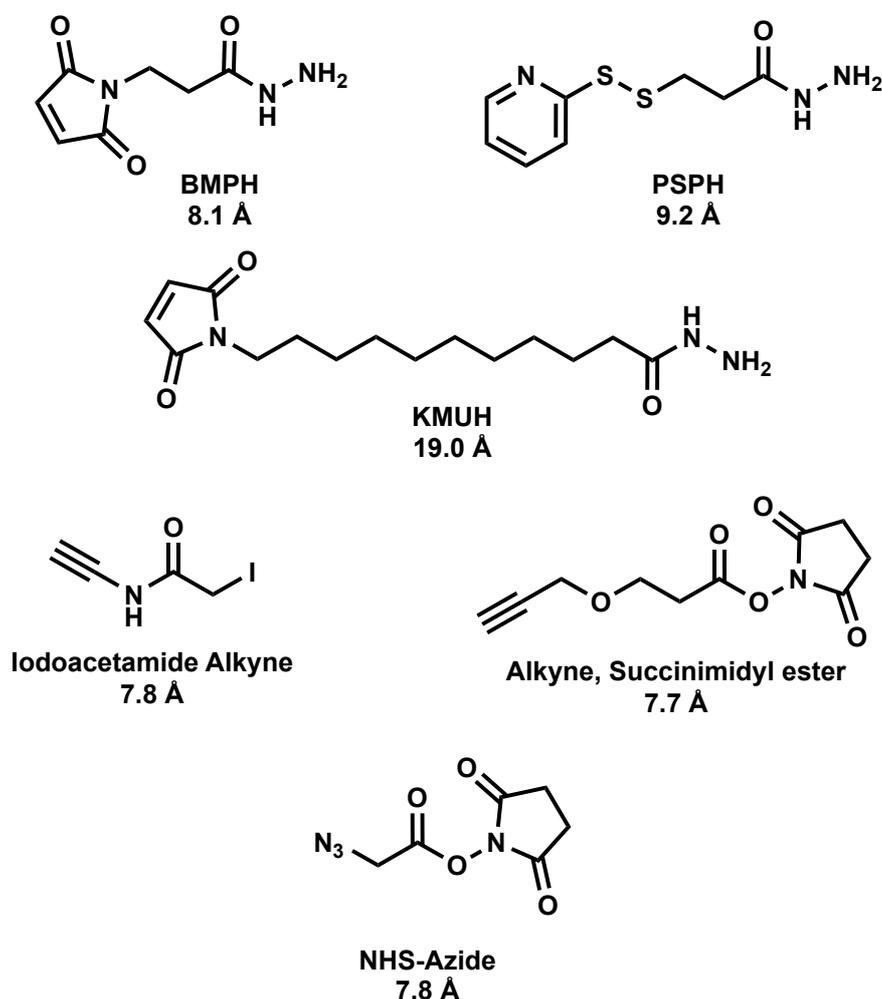


Figure 1.27: Other heterobifunctional linkers that are reactive towards chemical handles that are natural and unnatural in biology, useful for highly selective modification of biomacromolecules with chemical probes and reporters.

After establishing that an oxyamine/hydrazide would be a useful functional group in this setting, a second bioorthogonal group would be needed to create a heterobifunctional, bioorthogonal linker. Azide/alkyne containing heterobifunctional crosslinkers (figure 1.27) are also commercially available and frequently used for protein modification. As discussed azides/alkynes have been routinely introduced into carbohydrates and biological macromolecules as a chemical handle, for derivatisation or conjugation reactions. Combining either an azide/alkyne into a heterobifunctional linker gives a second bioorthogonal handle, proven to have great selectivity in its reactivity. A heterobifunctional linker which can combine the selectivity of a reaction between azides/alkyne, have features of a ‘click’ reaction and be able to derivatise unprotected oligosaccharides would be a very useful chemical tool in the production of synthetic

glycoproteins. Expanding upon the production of the neoglycoprotein based bacterial inhibitor by Branson *et al.*¹²⁸ in our group, the CuAAC was used to attach an oxyamine handle to GM1 azide derivative as is shown in figure 1.15, to allow the glycan derivative to be appended to the protein. Problems were faced in performing the CuAAC, it was determined that the CuAAC had to be performed before oxime ligation. Attempts to carry out oxime ligation of the amino-oxy acetamide followed by CuAAC of the GM1 azide to the protein was unsuccessful, possibly due to the scaffold ability to sequester copper species. Considering this and the problems previously outlined in section 1.4.3.2 with CuAAC, it was decided that use of a strained alkyne in place of a terminal alkyne would maintain the advantages of ‘click’ reactions discussed in section 1.4.3.2 without the downfalls of copper catalysis.

1.6.1 Choosing a strained alkyne scaffold

To develop a SPAAC-reactive linker a strained alkyne motif had to be chosen. Figure 1.28 shows a plot of lipophilicity vs reactivity for many of the reported cyclooctyne structures. It is important to choose a reactive group with an optimal balance of reactivity and correct physiochemical properties compatible with aqueous and buffered solvents. ALO (Aryl-lesscyclooctane), NOFO (Non-fluorocyclooctyne), MOFO (Monofluorinatedcyclooctyne) and DIMAC (Dimethoxyazacyclooctyne) structures were immediately ruled out due their low reactivity, and although DIFO (Difluorocyclooctane) and its analogues have greater reactivity, difficult and low yielding syntheses makes them challenging synthetic targets. The highly strained, benzo-fused structures such as DIBO (Dibenzoanulatedcyclooctyne), DIBAC (Dibenzoanulatedazacyclooctyne) and BARAC (Bisarylazacyclooctane) are some of the most reactive strained alkyne structures within the literature and have been applied in the bioorthogonal labelling of proteins. Similarly, despite their advantages in reactivity, these highly lipophilic structures have many disadvantages including off target hydrophobic interactions with macromolecules; poor aqueous solubility; and long and often low yielding synthesis.

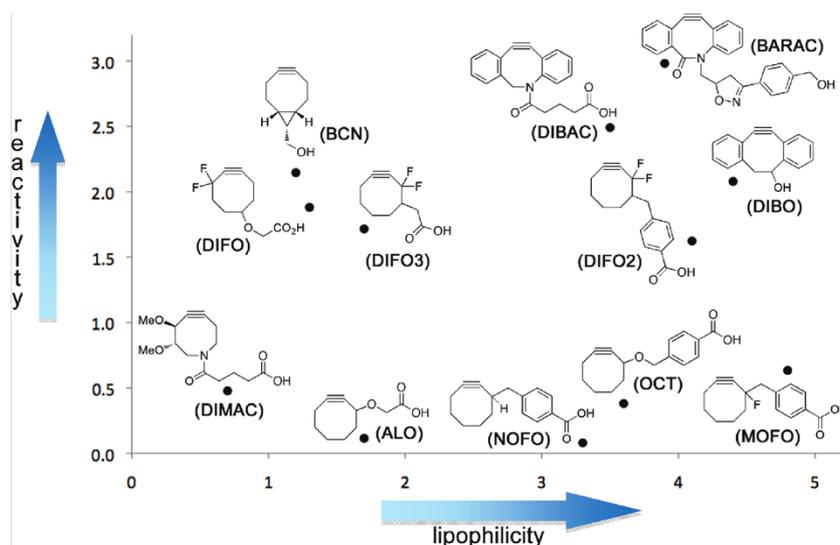


Figure 1.28: Figure showing SPAAC reactivity with benzyl azide versus lipophilicity of cyclooctynes. Reproduced with permission by M. F. Debets, S. S. van Berkel, J. Dommerholt, A. J. Dirks, F. P. J. T. Rutjes and F. L. van Delft, *Accounts of Chemical Research*, 2011, 44, 805-815, copyright © 2011, American Chemical Society.²⁰⁰

Comparing the many possible strained alkyne structures, we determined that the bicyclononyne (BCN) had superior balance of reactivity and lipophilicity to the rest. Other attractive properties included: symmetry, removing the potential for regio-isomers, and free rotation around the C-O bond would allow a limited degree of conformational freedom for the glycan. The structure of the linker gives rise to a carbohydrate attachment point, which would present the carbohydrate perpendicular from the scaffold protein surface and, thus, hopefully more accessible to lectin binding sites. The reported synthetic route to BCN is short, with only three steps to yield the alcohol, which can be functionalised in various ways, commonly through activated carbonates such p-nitrophenol or N-hydroxysuccinimide. The alcohol functional handle within the BCN structure is a versatile functional group and in addition to forming carbamates^{256, 257}, it has also been oxidised and undergone reductive aminations²⁵⁸, and examples of tosylation to form BCN polymers through cationic ring-opening polymerisation.²⁵⁹

1.6.2 Target bi-functional linkers

To combine SPAAC and oxime/hydrazone chemistry in one linker using the BCN scaffold, the objective was to introduce the oxyamine or hydrazone functionality at the site of the BCN alcohol. Three target linkers were proposed, hydrazone (**1.1**), oxyamine (**1.2**) or N-methyl oxyamine (**1.3**) shown in figure 1.29. Each of three BCN strained alkyne derivatives has an eight-atom link between the two reactive groups, therefore, a nine-atom link between glycan and protein.

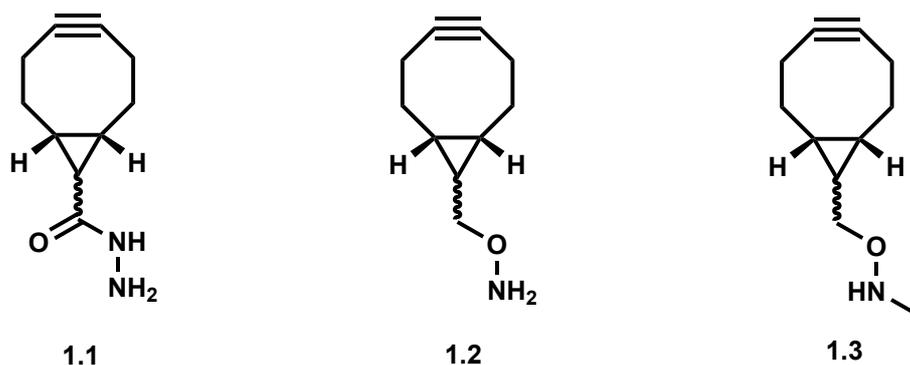


Figure 1.29: Structures of short linker showing three target BCN scaffolds with different attachment functionality, BCN hydrazone (left), BCN oxyamine (middle) and BCN methyl oxyamine(right).

1.7 Synopsis

After concluding on the bioorthogonal groups suitable to develop this type of bifunctional linker and identifying three target structures, the different synthetic glycosylation methods, and the mutant proteins needed to generate the neoglycoproteins are shown in figure 1.30.

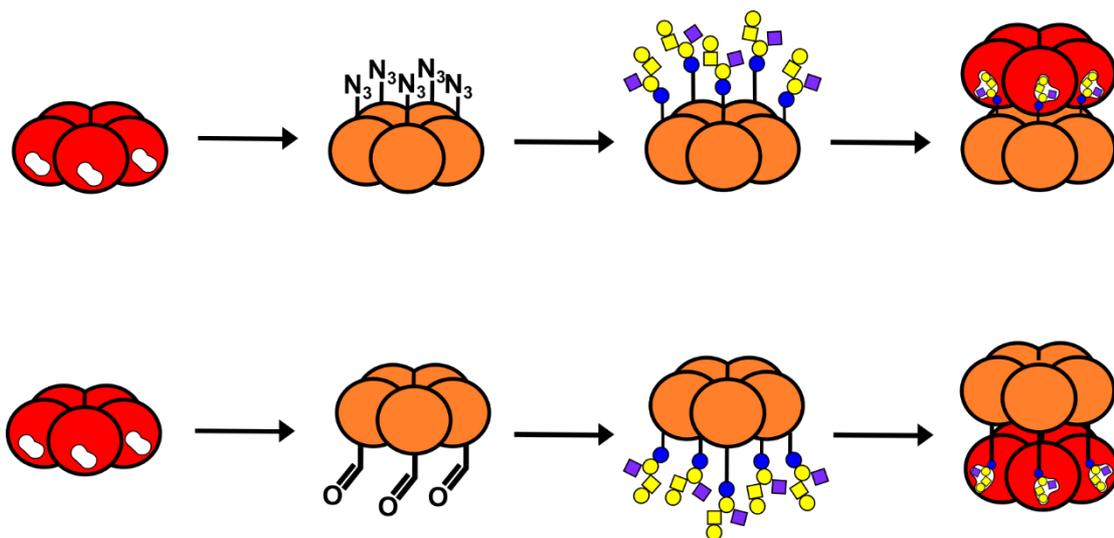


Figure 1.30: A figure showing the desired mutant proteins derived from the WT CTB-subunit, containing an azide on the non-binding face, or aldehyde at the N-terminus and the different GM1 containing neoglycoprotein inhibitors to be produced by attachment of glycans these sites.

The remaining chapters are an account of the approach taken to address the proposed research questions based on neoglycoprotein synthesis, and efforts to develop novel, useful chemical tools to advance the procurement of synthetic glycoproteins. Chapter two addresses the synthetic studies into obtaining each of the target linkers proposed in figure 1.30, chapter three will detail application of the linkers for glycan conjugation and their further enzymatic elaboration. Chapter four outlines the preparation and evaluation of neoglycoprotein inhibitors against the cholera toxin adhesion, and chapter five will discuss the expansion of neoglycoprotein-based inhibitors towards the Shiga-like toxin.

Finally, chapter six will discuss some preliminary efforts to advance neoglycoprotein inhibitors towards a potential therapeutic, by investigating non-bacterial protein scaffolds.

Chapter 2 New tools for synthetic glycosylation

Synthesis of bi-functional, bio-orthogonal linkers for the generation of reactive glycoconjugates

This chapter will cover the chemical synthesis of a new heterobifunctional linker containing two bioorthogonally reactive groups, specifically chosen for their capability to functionalise both carbohydrates and proteins. Figure 2.1 reiterates the three target structures combining a strained alkyne, with a hydrazide **1.1**, oxyamine **1.2** or N-alkyloxime **1.3**. Alongside the synthesis of these linkers, investigation of a potential novel synthetic routes to BCN structures to improve yields and determine if the bicyclic alkyne can be formed under milder reaction conditions, will be discussed.

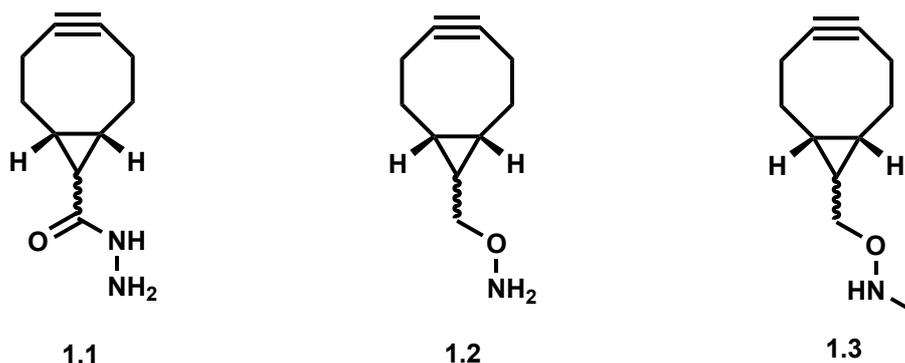
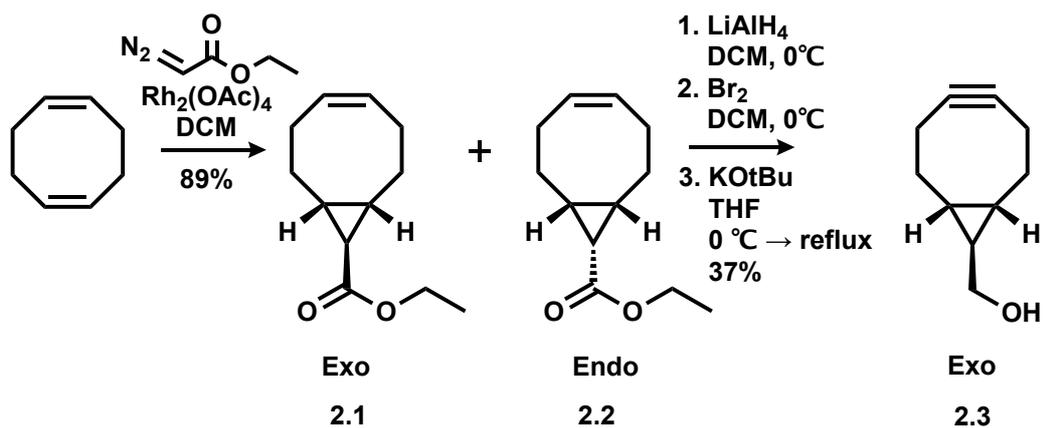


Figure 2.1: Structures of the three target bifunctional linkers based upon the bicyclic strained alkyne BCN.

2.1 Literature synthesis of bicyclononynes

Dommerholt *et al.* first reported the synthesis (Scheme 2.1) of BCN from cyclooctadiene in a four-step synthesis.²⁵⁶ The synthesis proceeds through a rhodium (II) catalysed cyclopropanation of cycloocta-1,5-diene with ethyl diazoacetate, yielding two separable bicyclic diastereoisomers **2.1** and **2.2**. Isolation of the single isomers is performed by silica chromatography, followed by a sequential reduction of the ester, and bromination. Double elimination of the dibromide affords strained alkyne **2.3**. The alcohol is then commonly converted to an activated carbonate, for attachment to specific chemical reporter handles e.g. dyes, biotin. Commercially, NHS or p-nitrophenyl activated carbonate derivatives of BCN can be purchased for reaction with amines, however, rising costs have made them less accessible.²⁵⁶ Scheme 2.1 shows the yields achieved for the route in our lab, ester **2.1** and **2.2** were isolated with a combined yield of 89% in a ratio of *exo* and *endo* isomers 2.2:1, respectively. The reduction and bromination steps were shown to proceed in quantitative yield, requiring no purification post work-up. The highest yield achieved in the elimination to produce strained alkyne **2.3** was 36% after column chromatography. Reproducibility of the elimination reaction was often variable, resulting in low yields and was not readily scalable, underlining the need for improved methods for synthesis of the strained alkynes.

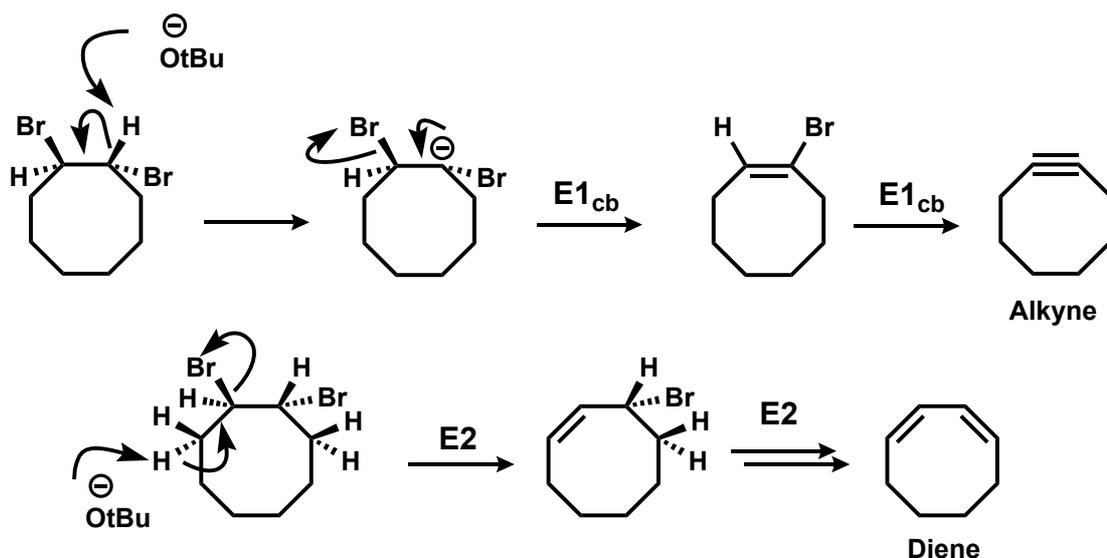


Scheme 2.1: Synthesis of the bicyclonone following the literature synthesis for BCN by Dommerholt et al. a) Ethyl diazoacetate, $\text{Rh}_2(\text{OAc})_4$, DCM, 72 hours, 89%. b) 1. LiAlH_4 , DCM, 0°C , quantitative. 2. Br_2 , DCM, 0°C . KOtBu , THF, 66°C , 37%.

Theoretically, the target hydrazide linker **1.1** could be produced from ester **2.1** by a single step hydrazinolysis with hydrazine hydrate, assuming that the strained alkyne can be produced in the presence of the ester group. Target oxyamine linkers **1.2** and **1.3** could be synthesised from alcohol **2.3** by its activation as a leaving group and subsequent substitution with a protected oxyamine.

2.2 Developing *syn*-leaving groups towards a novel strained alkyne synthesis

Bicyclononyne derivatives have been reported to be produced with yields of 61% as a single *endo* isomer, despite the pathway for the formation of the strained alkyne being an energetically unfavourable one. Halogenation of alkenes using X_2 is known to produce only an *anti*-relationship between halides, as a result of the halonium ion intermediate. This anti-relationship rules out the possibility of the first elimination proceeding through an E2 mechanism due to the lack of an anti-periplanar proton between the substituted carbon atoms shown in scheme 2.2. Therefore, an E1_{cb}-like mechanism better explains the formation of the vinyl bromide intermediate, which in turn explains the potential formation of a diene side product as the formation of the diene can proceed through an E2 mechanism.



Scheme 2.2: Reaction mechanism for double elimination to produce a strained alkyne by E1_{cb} -like elimination and the mechanism for the E2 elimination of antiperiplanar proton to produce the diene by-product.

2.2.1 Retrosynthetic design of a novel synthesis

A novel synthetic route to introduce a *syn*-relationship of leaving groups could potentially lead to more efficient synthesis of strained alkynes under milder conditions. A strategy was devised for generation of an intermediate with a *syn*-leaving group (LG) relationship. In conjunction with changing the stereochemistry of the intermediate compounds, it was an aim to improve the nature of the LG to promote the formation of the enthalpically disfavoured alkyne. From a retrosynthetic analysis of strained alkyne synthesis (figure 2.2), it was proposed that by performing *syn*-dihydroxylation, subsequent conversion to a ditriflate would introduce a very good leaving group with a *syn*-stereochemistry setting the system up to undergo E2 elimination.

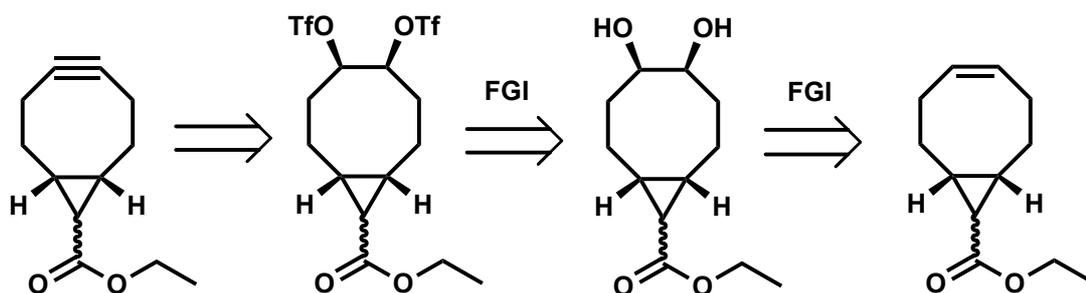
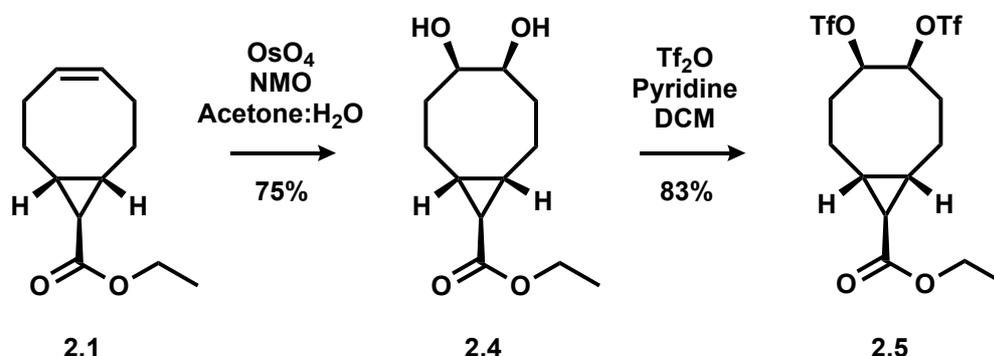


Figure 2.2: Retrosynthetic analysis for the generation of a bicyclononyne ester through *syn*-leaving group intermediate.

2.2.2 Synthesis of *syn*-ditriflate bicyclic ester

Ester **2.1** was prepared by cyclopropanation of cyclooctadiene according to the literature in a yield of 89%. With the desire to produce the strained alkyne while retaining the ester functional group, the isomers were separated and a single isomer was subjected to

dihydroxylation under Upjohn conditions to yield diol **2.4** in 75% yield (Scheme 2.3).²⁶⁰ Efforts to avoid the use of OsO₄ by using Woodward catalytic dihydroxylation²⁶¹ conditions proved to be unsuccessful. Subsequent triflation of the diol yielded **2.5** in good yield, no evidence of elimination in the presence of pyridine was observed, showing the ditriflate was stable in the presence of mild base.



Scheme 2.3: Synthetic route to ditriflate BCN by *syn*-dihydroxylation of **2.1**.

2.2.3 Screening for new elimination conditions

With the 1,2 ditriflate **2.5** synthesised, a base screen was performed to determine if milder conditions could be used to carry out this elimination. With the original protocol using alkoxide bases for elimination, two were chosen for comparison to the amine bases. KOtBu was chosen as a non-nucleophilic, strong alkoxide base, and NaOEt as a milder alkoxide base could be used in the presence of the ester as NaOEt transesterification would result in no change to the starting material. With a pK_a range of 10 to 17 covered, table 2.1 displays reagents and conditions trialled for the double elimination of **2.5**. However, in each case no alkyne formation was observed. Mild amine bases such as DBU, TEA and DIPEA showed no reaction, presumably indicating that the pK_a of the methine proton is not altered dramatically by the presence of a triflate over a bromine. The use of the alkoxide bases led to some reaction as the starting material could no longer be identified, however, there was no formation of alkyne or unwanted alkene observed when performed, either at room temperature or under reflux.

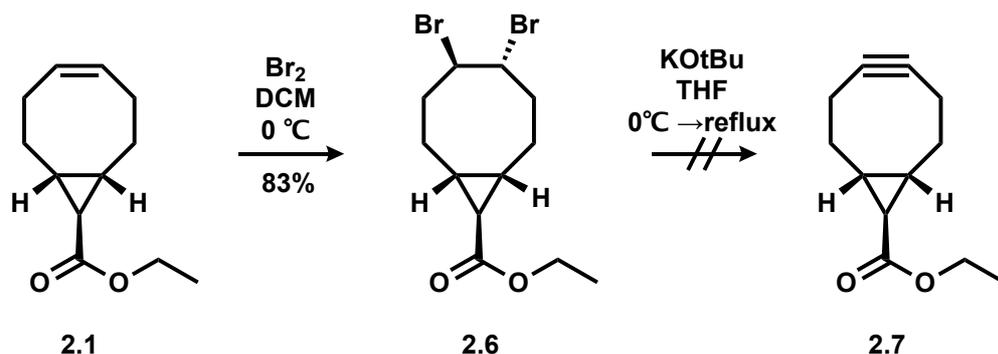
Table 2.1: Screening of organic amine bases and alkoxide bases, their pKa values, reaction conditions, and a summary of the resulting analysis by crude NMR.

<u>Base and conditions</u>	<u>Result</u>	<u>pKa</u>
DBU, DMF, rt.	No reaction	13.5
DIPEA, DCM, rt.	No reaction	10.8
TEA, DMF, rt.	No reaction	11.0
NaOEt, EtOH, rt and Reflux	Loss starting material, no alkyne or diene	15.9
KOt-Bu, THF, rt and Reflux	Loss starting material, no alkyne or diene	17.0

Products obtained after a mild acidic work-up, for eliminations with NaOEt and KOt-Bu could not be fully characterised. NMR spectroscopic analysis of the crude samples showed no evidence of bromine substitution or transesterification in either case. The spectra showed only signals that consisted with fully saturated cyclic compounds, and no distinguishable signals in the aliphatic region. From the data obtained we can only speculate potential reactions that may have occurred, the data alludes to a possible ring opening or possible intramolecular substitutions, through possible deprotonation of the cyclopropane ring proton situated β to the ester. If this is more acidic than the proton α to the bromine, deprotonation could promote: ring opening of the bicyclic system, driven by the relief of ring strain; further cyclisation from generation of an enolate or carbanion, by intramolecular displacement of the triflate creating smaller bicyclic ring structures.

2.2.4 Literature synthesis maintaining the ester

To confirm if deprotonation of the ester was indeed preventing elimination of the triflate, the dibromo analogue **2.6** was synthesised (scheme 2.4) to determine if the same products were produced from elimination of the bromine atoms in the presence of the ester. Bromination of the alkene **2.1** was shown to be successful, and formation of dibromo bicyclic ester **2.6** was confirmed by NMR and MS analysis. When subject to KOtBu, it was found that no alkyne was produced, with the crude NMR showing the same complex aliphatic region observed following base treatment of the BCN ditriflate **2.5**.



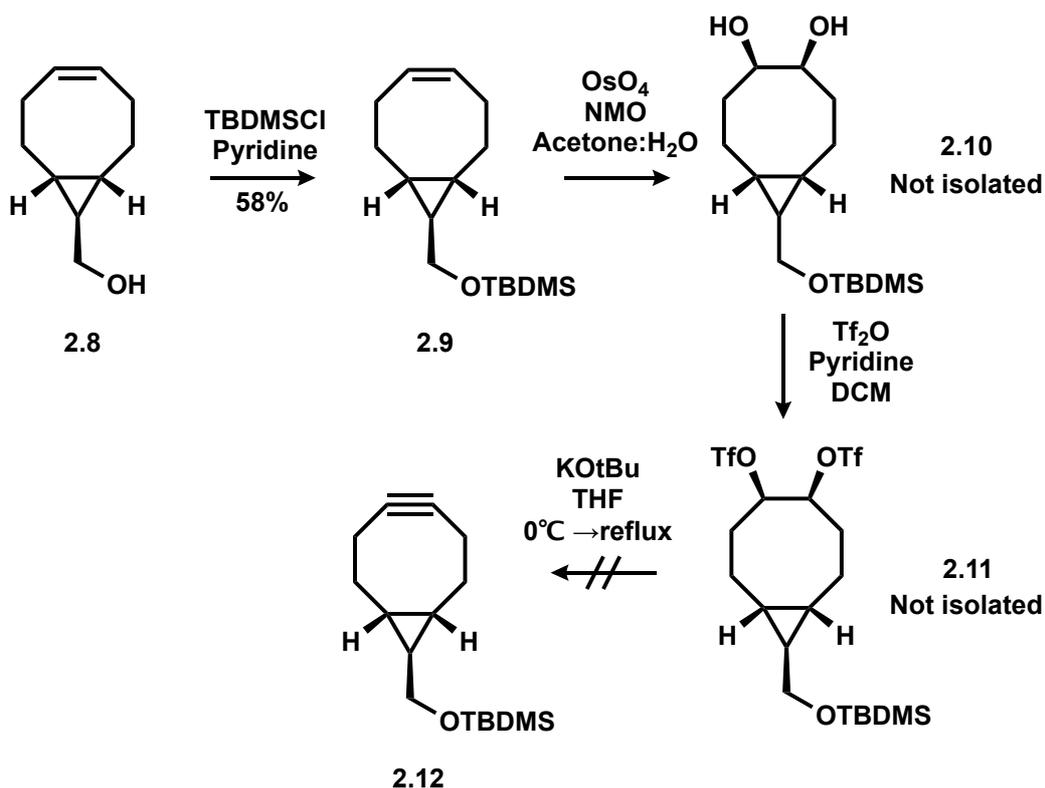
Scheme 2.4: Bromination of **2.1** to give the dibromo BCN ester structure for trialling elimination reaction to the synthesis of the strained alkyne in the presence of the ethyl ester.

The inability to produce a strained alkyne from **2.6** is proof that the ester does interfere with the elimination step. With no evidence of any diene, this result supports a hypothesis that the ester decreases the pK_a of the proton on the cyclopropane ring, resulting in a different reaction pathway that yields none of the expected elimination products.

After determining that the strained alkyne could not be synthesised in the presence of the ester, hydrazinolysis of **2.1** was attempted to investigate if a change to the ester functional group would be tolerated. Hydrazinolysis reactions usually require a small excess of hydrazine, stirred with mild heating in ethanol. However, no conversion of the ester to a hydrazide was observed under these conditions. Successful hydrazinolysis required 50% v/v hydrazine hydrate in ethanol, however, this large excess of hydrazine was also found to reduce the alkene to the fully saturated bicyclononane. NMR spectroscopic analysis of the product supported this conclusion, as the alkene signals at ca. 5 ppm were no longer observed. Synthesis of hydrazide linker **1.1** was subsequently discarded from our target compounds, as neither the hydrazinolysis nor strained alkyne formation were feasible.

2.2.5 Replacement of the ester with a protected alcohol

It was decided to move forward in applying methods for *syn*-triflation to the primary alcohol **2.8** obtained by LiAlH₄ reduction of **2.1** in quantitative yield. In order to avoid triflation of the primary alcohol group, the alcohol was protected with a bulky silyl group to give **2.9**.



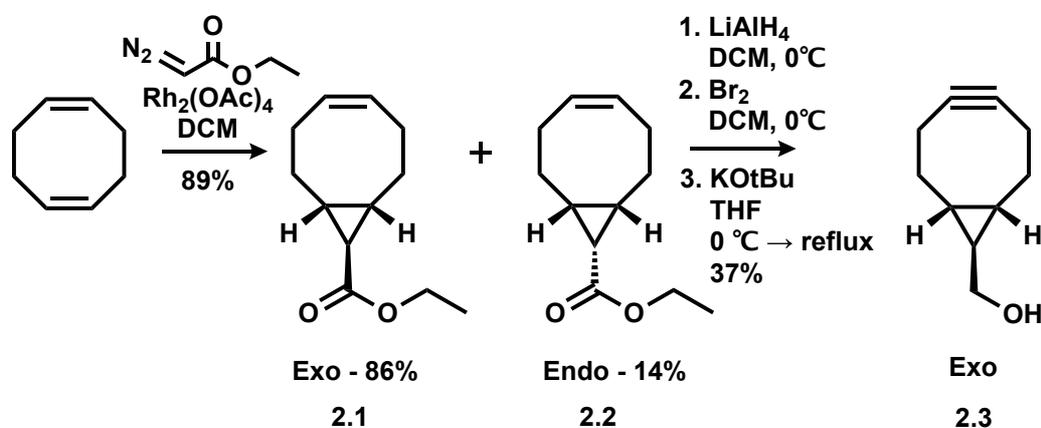
Scheme 2.5: Attempted synthesis of the *syn*-ditriflate with a silyl protected alcohol functionality from the BCN alkene **2.8**, by dihydroxylation and triflation and attempted elimination.

Dihydroxylation and triflation of **2.9** was performed as described previously giving the silyl protected ditriflate **2.11**. As a proof of concept study, intermediate compounds **2.10** and **2.11** were isolated following a mildly acidic aqueous work-up and, following confirmation of successful reaction by NMR spectroscopy, carried through to the next step without further purification. The ditriflate was then subjected to elimination with KOtBu, and the final crude product was analysed for any presence of alkyne by ^1H and ^{13}C NMR spectroscopy. The spectra showed only presence of alkene with signals at 5.56 ppm and 130 ppm, with no trace of any alkyne peak at ca. 99 ppm. Introducing a *syn*-leaving group stereochemistry proved non-beneficial for elimination to the strained alkyne. Any favourable gains by setting up the elimination with a *syn*-leaving group appeared to be negated by the improvement in leaving group from a bromine to a triflate, leading to rapid formation of the diene, making the E2 elimination reaction both the kinetic and thermodynamic product. As treatment of the *syn*-ditriflate **2.11** with KOtBu under reflux did not yield any alkyne, elimination with strong bases LDA and BuLi were investigated at $-78\text{ }^\circ\text{C}$, as the strong bases should remove the most acidic proton. In both cases rapid elimination to the diene was observed, and the diene became the sole product.

2.3 Functional group interconversion of alcohol to oxyamines

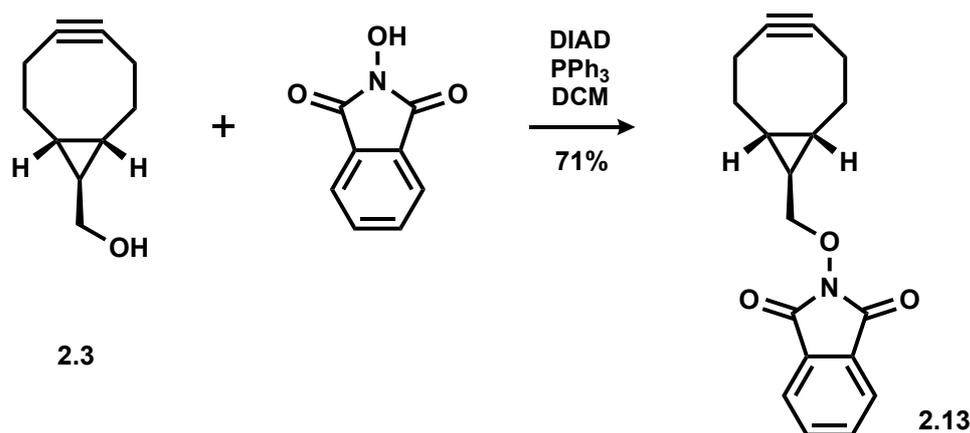
2.3.1 Installing protected oxyamines by Mitsunobu chemistry

With no prospect of improving the elimination chemistry, strained alkyne **2.3** (scheme 2.6) was synthesised following the procedure reported by Dommerholt.²⁵⁶ The literature protocol was maintained for the synthesis and isolation of strained alkyne **2.3**. Yields were identical for the preparation of *exo* and *endo* isomers of **2.3**, however, the *exo* isomer was prepared in larger quantities due to **2.1** being the major product of the cyclopropanation (a ratio 2.2:1 *exo:endo*), the *exo* isomer was carried forward in the linker synthesis.



Scheme 2.6: Synthetic route to showing a single *exo* isomer of BCN alcohol **2.3** of BCN from cyclopentadiene following the reported synthesis by Dommerholt *et al.*²⁵⁶

After isolating strained alkyne **2.3**, the alcohol was then converted to phthalimide **2.13** by Mitsunobu reaction with N-hydroxyphthalimide in 71% isolated yield (scheme 2.7). The phthalimide group was used to introduce an oxyamine group as a protected precursor, allowing the exploration of different deprotection strategies which could be used to synthesise both **1.2** and **1.3** from **2.13**.

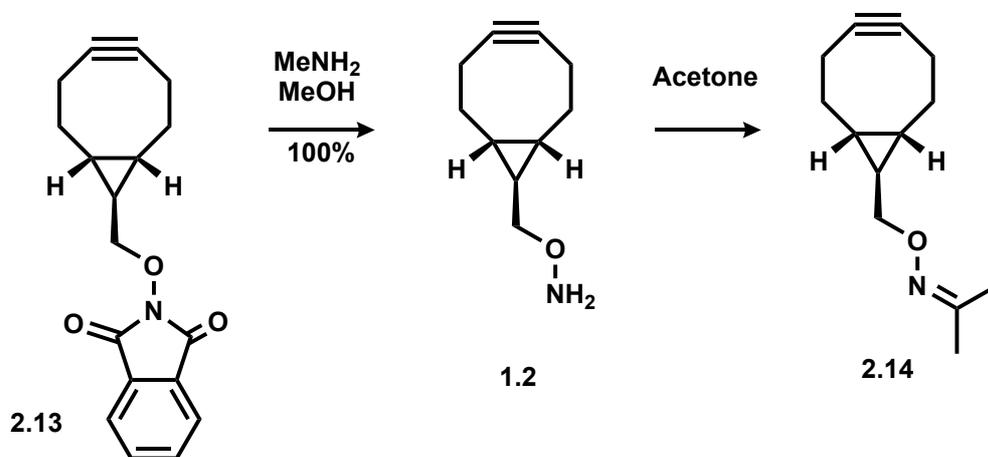


Scheme 2.7: Functional group interconversion from the alcohol **2.13** to phthalimide protected oxyamine using the Mitsunobu reaction on the single exo isomer of BCN.

The formation of strained alkyne **2.3** was also attempted on the alkene analogue of **2.13** carrying out the Mitsunobu reaction before the formation of the alkyne. The bromination of the alkene was successful, however, the elimination to give the strained alkyne led to a complex reaction mixture with only a diene by-product being isolated. This provided further evidence that having the alcohol functionality is significant when forming the alkyne, with limited acceptance of other functional groups.

2.3.2 Deprotection to give a primary oxyamine

The deprotection of phthalimides is routinely performed by ring-opening with hydrazine such as in the Gabriel synthesis of primary alkyl amines. However, it was previously determined that hydrazine cannot be used on the BCN alkyne structures due to its propensity to reduce the strained alkyne. An alternative deprotection method was therefore required for the ring-opening of phthalimide. Aqueous methylamine proved to be a very efficient reagent for the deprotection of **2.13** (scheme 2.8), with the target linker **1.2** achieved in quantitative yield within minutes at room temperature.



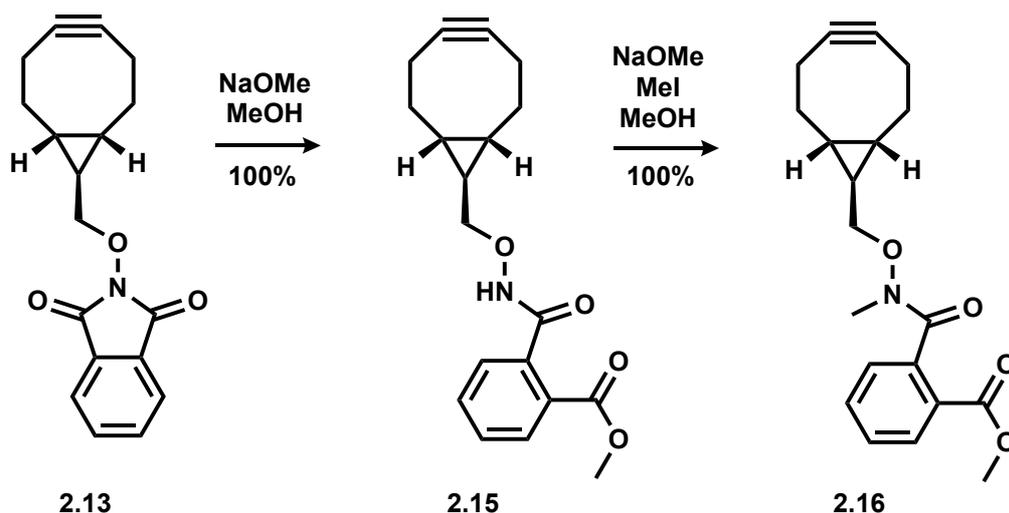
Scheme 2.8: Methylamine deprotection of of the phthalamide **2.13** protecting to group to give free oxyamine.

Following work up and purification, compound **1.2** was found to react readily with acetone in the laboratory atmosphere leading to formation of oxime **2.14**. The isolated product showed the expected signals for **1.2**, however, signals for exchangeable amine protons did not match the expected integration and a further singlet corresponding to the two CH₃ groups of the oxime was present. The formation of the oxime by-product was further confirmed by ¹³C NMR spectroscopy with appearance of a peak for the two methyl groups and imine carbon. The mass spectrum also corroborated the presence of the oxime, as the most abundant species was +40 Da larger than the expected mass for **1.2**.

It was found that the free oxyamine could be regenerated using a nucleophilic alkoxide such as sodium methoxide: upon neutralisation, the oxyamine displaying the correct mass for **1.2** was recovered. Alternatively, it was anticipated that a deprotection method yielding **1.2**, which required no further purification, could be used in subsequent oxime ligation reactions. The by-products of the methylamine deprotection would be inert in the oxime ligation, and any excess methylamine would not interfere in the ligation reaction as hydrolysis of the resulting imine would hydrolyse in water. Therefore, the deprotection strategy was changed to use anhydrous methanolic methylamine, which could then be used directly in subsequent reaction with oligosaccharides. The deprotection conditions were shown by LC-MS to give quantitative conversion to **1.2**, and the methyl phthalamide by-product was sparingly soluble in methanol and precipitated from the crude mixture upon storage at -20 °C.

2.3.3 An alternative deprotection strategy for production of alkyl oxyamines

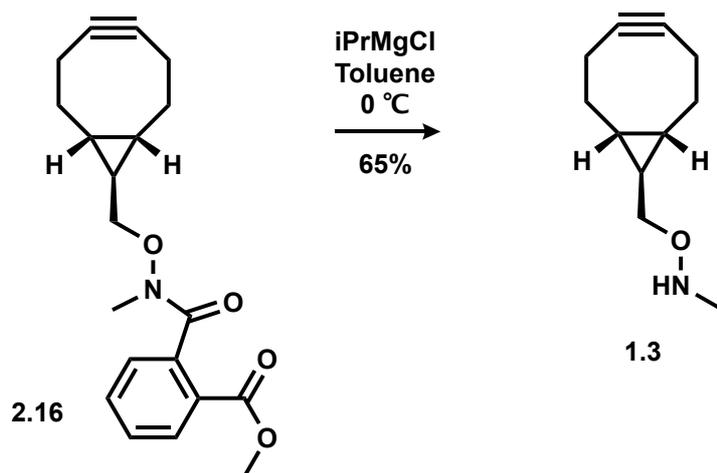
The synthesis of *N*-methyl linker **1.3** was initially attempted using Mitsunobu chemistry with Boc-*N*-methylhydroxylamine, however, this was unsuccessful as no alkyne was isolated. Spectral analysis of the crude mixture by NMR alluded to a 1,3-dipolar cycloaddition product, alongside signals indicating partial substitution of the alcohol. MS data could not be obtained to aid the structure elucidation, due the poor ionisation of these compounds using electrospray ionisation. **1.3** was instead synthesised using an alternative deprotection strategy from intermediate **2.13**, in which an amide group was formed upon ring opening of the phthalimide. Treating phthalimide **2.13** with sodium methoxide (scheme 2.9) resulting in ring opening by attack of methoxide at a single carbonyl to reveal amide **2.15**.



Scheme 2.9: One-pot synthesis of a Weinreb amide **2.16** from the phthalimide **2.13** as an intermediate in the route to alkylated oxyamines.

In a one-pot reaction the amide was methylated using methyl iodide in the presence of excess sodium methoxide as a base. Methylation of the amide was followed by HRMS and was complete after 2 hours; acidic work-up yielded the Weinreb amide **2.16** in quantitative yield in a two-step, one-pot reaction.

Weinreb amides are often used for conversion of carboxylic acids or acyl chlorides into ketones, with use of organometallic alkylating agents. In most cases the alkoxyamine group is often unwanted and only used to create an intermediate for alkylation. Our deprotection approach is an unusual use of a Weinreb amide, in which Weinreb amide intermediate **2.16** was deprotected using Grignard reagents to create an unwanted ketone, and the desired N-methyl-oxyamine. Isopropyl magnesium chloride was chosen as the Grignard reagent, and after nucleophilic attack on the amide and collapse of the tetrahedral intermediate oxyamine **1.3** was isolated upon work-up (scheme 2.10). Further purification by silica gel chromatography was then required to isolate the secondary amine **1.3** in 65% yield.

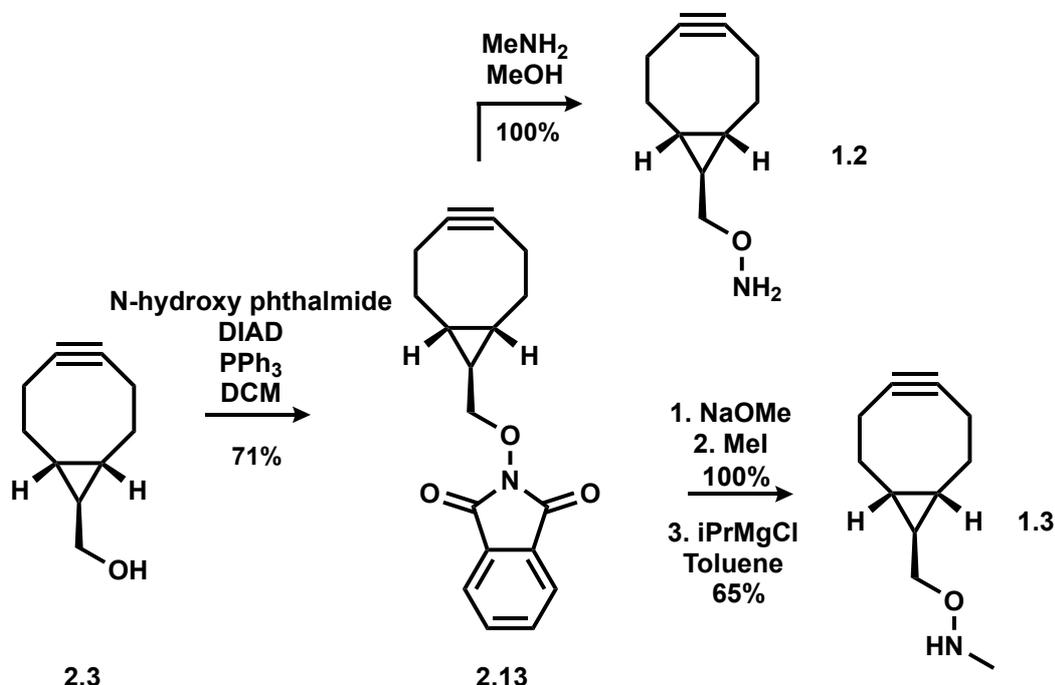


Scheme 2.10: Grignard deprotection of the Weinreb amide intermediate 2.16 isopropyl magnesium chloride to yield the N-Methyl oxyamine 1.3.

This novel method could prove useful for future synthesis of complex N-alkyl oxyamines structures for glycoconjugation. Introducing a phthalimide group is a versatile alternative to using N-Boc hydroxylamine, as a protected N-alkyl oxyamine. Finding such an effective deprotection method for the synthesis of **1.3** was important for this work and the viability of this linker for use in glycoconjugate/glycoprotein synthesis, as a BCN linker which could maintain the cyclic conformation of glycans is highly desirable.

2.4 Conclusion

A synthetic route has been developed (scheme 2.11) to give two novel bifunctional linkers combining two different bioorthogonally reactive groups into a single minimal length linker. The use of Mitsunobu chemistry offers mild conditions to introduce a protected oxyamine, forming a single precursor which can then be converted to the free oxyamines or N-alkyl oxyamines. The divergent deprotection strategy to access N-alkyl and free oxyamines (scheme 2.11), could be applied to a wide range of molecules for bioorthogonal labelling. The use of a phthalimide and conversion to a Weinreb amide for alkylation of the amine, has not previously been reported. The highly reactive strained alkyne functional group has been found to be stable to all reaction conditions, purification methods and storage at $-20\text{ }^{\circ}\text{C}$ in the crude deprotection mixture. These linkers have been designed for the generation of reactive glycoconjugates with different reactivities for the glycosylation of proteins. These linkers are the first to combine the bioorthogonal strained alkyne with a carbohydrate/aldehyde-reactive functional group. The subsequent chapters will describe studies towards their use for glycosylating biomacromolecules in a bioorthogonal fashion, and their potential to meet a need in the field of neoglycoprotein synthesis.



Scheme 2.11: A summary of the novel synthetic route to linkers 1.2 and 1.3 from the reported BCN alcohol through a protected phthalimide and divergent deprotection strategies of a single strained alkyne precursor.

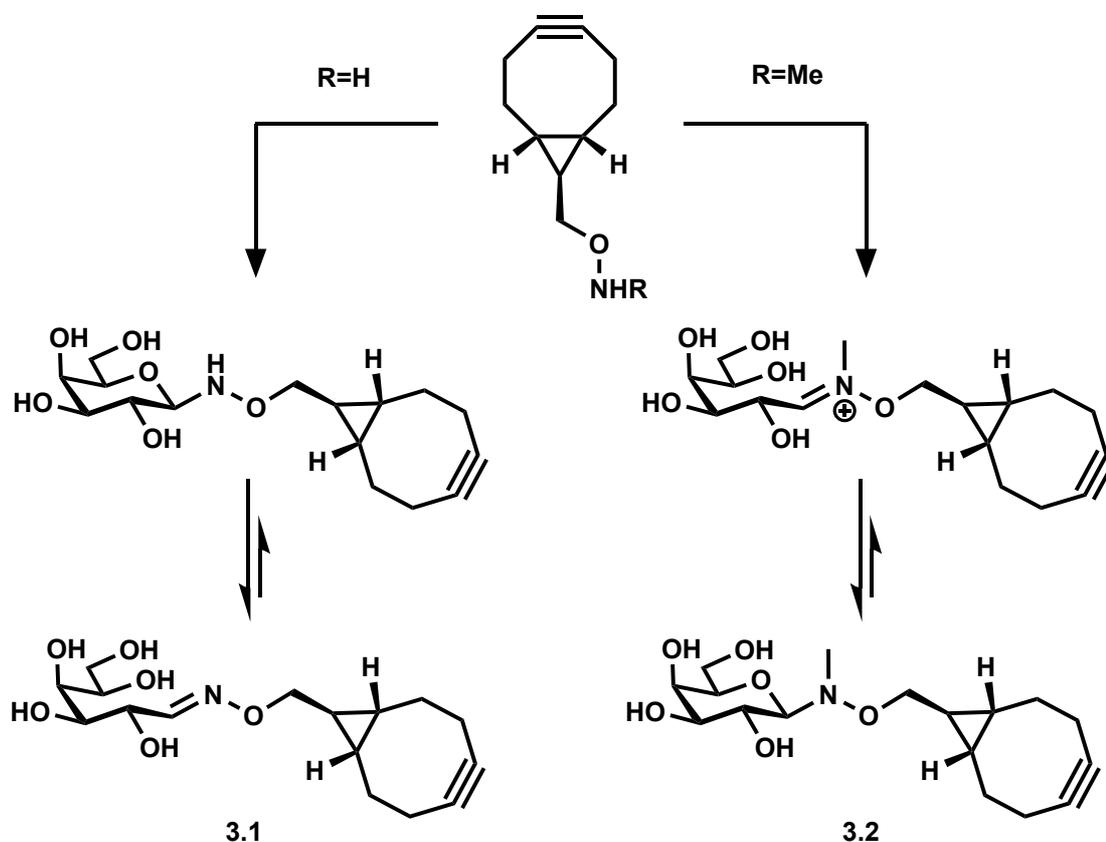
Chapter 3 Bioorthogonally reactive oligosaccharides
Chemical and enzymatic functionalisation of reducing sugars

Following on from the development of a synthetic route to novel linkers **1.2** and **1.3**, our aim was to derivatise oligosaccharides using one reactive group, leaving the other for bioconjugation to proteins. Derivatives of GM1 and Gb3 based on both oxime and SPAAC are targets for the ligation to CTB scaffold, to develop new neoglycoprotein inhibitors with minimal lengths of linkers between glycan and protein. This chapter will discuss the chemical/chemoenzymatic methods used for the functionalisation of unprotected oligosaccharides to BCN linkers **1.2/1.3**, through oxime ligation and SPAAC conjugation.

3.1 SPAAC reactive glycosides from reducing sugars

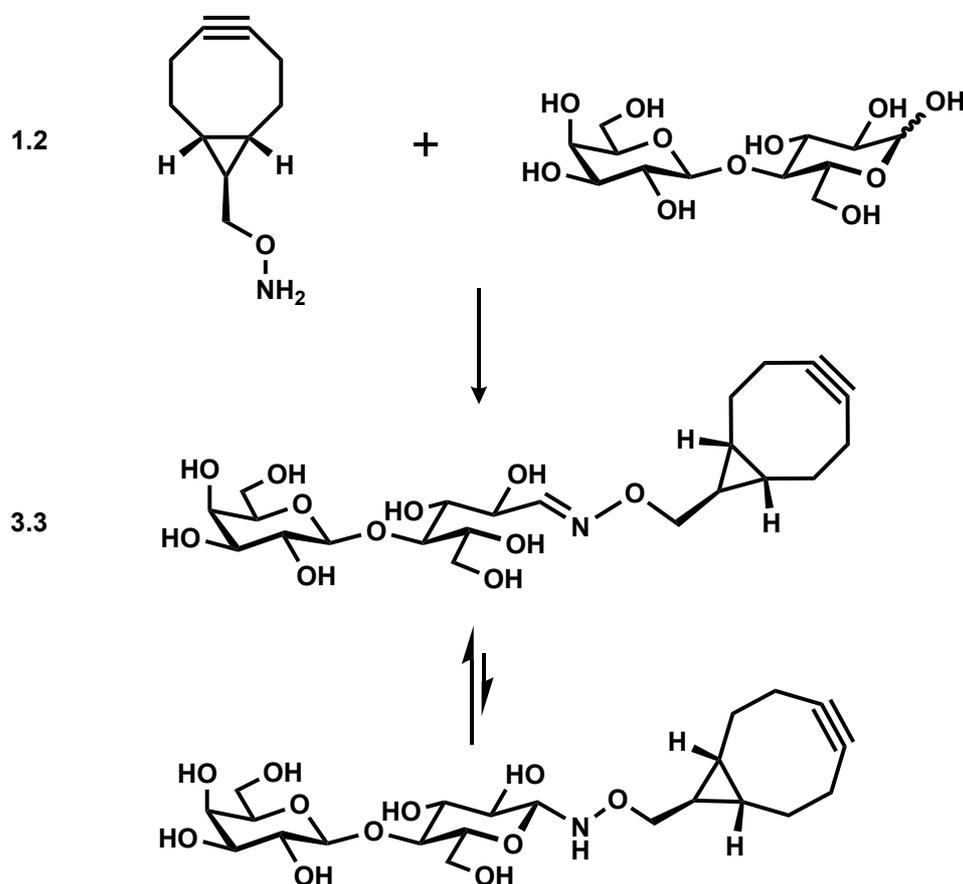
3.1.1 Optimisation of oxime ligation reactions using disaccharides

With the target novel linkers **1.2** and **1.3** synthesised, reaction with unprotected oligosaccharides through oxime ligation would produce two different bioorthogonally reactive oligosaccharides, as described in section 1.5.4. Reaction at the anomeric centre with **1.2** and **1.3**, would produce an oxime or N-linked glycoside respectively, as shown in scheme 3.1 with the ligation of **1.2/1.3** to galactose as an example.



Scheme 3.1: Example monosaccharide derivatives **3.1** and **3.2** formed through oxime ligation at the anomeric position of a reducing sugar to linker **1.2** (left) and **1.3** (right) respectively.

Optimisation of the reaction between **1.2** and reducing sugars was developed first, as it is known that the oxyamines are more reactive than their N-alkyl analogues. Lactose is found at the core of many biologically relevant glycans including both GM1 and Gb3 structures, making it an appropriate model oligosaccharide for method development. Development of efficient ligation was essential, before attempting ligation of larger, more complex and less reactive oligosaccharides, which are often: commercially expensive; isolated from biological material; and usually available in only small quantities. The oxime ligation (scheme 3.2) between linker **1.2** and lactose was initially attempted by combining equimolar quantities at a concentration of 78 mM in 50 mM sodium acetate buffer at pH 4. No reaction was observed despite the reaction being left for 48 hours, appearing to be hindered by the rate of mutarotation/ring-opening of the sugar, as previously outlined as a consequence of an uncatalysed reaction.

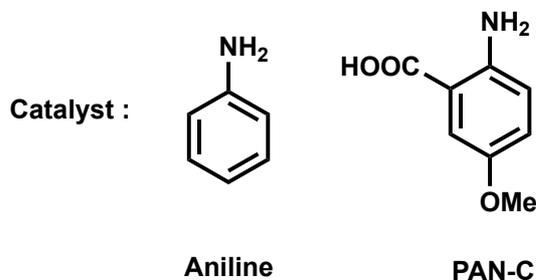
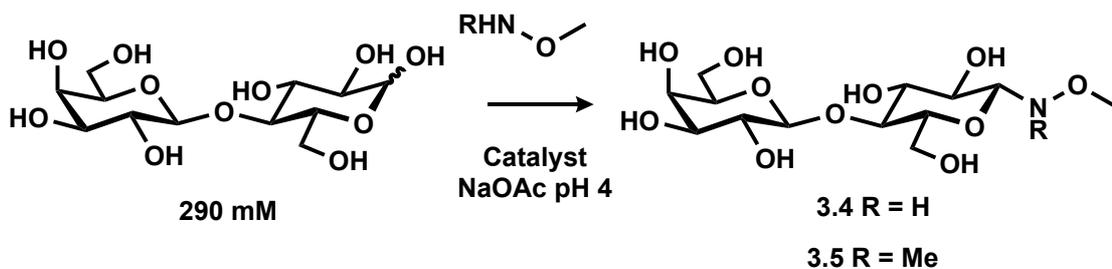


Scheme 3.2: Oxime ligation of lactose with linker **1.2** to give the glycosyl oxime **3.3** to lactose.

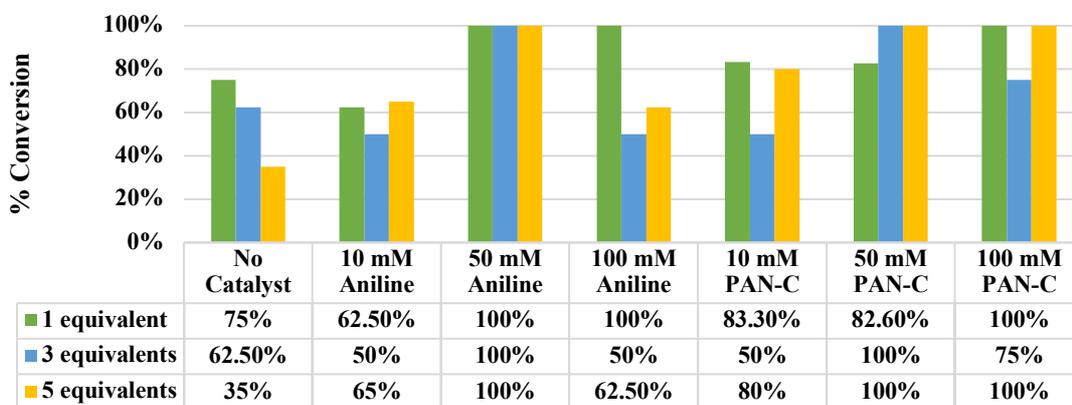
The next step was to investigate the effect of an organocatalyst on this reaction (scheme 3.2). As mentioned, aniline is recognised and widely used as an effective catalyst for oxime ligation with reducing sugars.^{246, 262, 263} Following the literature protocol, the reaction was performed in 100 mM aniline in acetate buffer at pH 4, however, no change in the reaction was observed by TLC, only showing lactose and **1.2**. To ensure the oxime

product was not hydrolysed on silica TLC plates, LC-MS was used to corroborate the findings. A hydrophilic interaction liquid chromatography (HILIC) column was used for LC-MS to analyse the reaction mixture, as it is capable of separating highly polar molecules, and it was anticipated that the addition of the hydrophobic BCN would lead to distinct changes in the chromatogram for the product. Mass spectrometry indicated that in solution, the dominant species was the aniline glycosylamine intermediate. This finding was unexpected, as one would expect the aniline intermediate to quickly react to give the more stable oxime product or be converted back to the reducing sugar. It is possible that this intermediate was the most ionisable species and therefore, detected as the abundant species. Nevertheless, this result did confirm that the aniline was reacting with lactose, but no further reaction with linker **1.2** was observed. It was concluded that using aniline at such a high concentration of 100 mM was hindering the reaction, with the intermediate likely to be intercepted by water or another aniline molecule, in preference to reaction with linker **1.2**. Similar findings had been noted in an initial report by Jensen and co-workers, in which they stated formation of the glycosylamine with aniline can hinder catalytic efficiency.²⁴⁶

Next, experiments were performed using different organocatalysts at varying concentrations to assess the impact of having an organocatalyst present in the reaction. Aniline and another organocatalyst, 5-methoxyanthranilic acid which has been reported to be very efficient at pH 4-5,²⁵¹ were compared after reviewing the study by Jensen which compared a number of reported catalysts.²⁵⁴ Reactions with *O*-methylhydroxylamine and *N,O*-dimethylhydroxylamine as simple oxyamine reactants mimicking linkers **1.2** and **1.3**, respectively with lactose (figure 3.1) were performed in acetate buffer at pH 4, with the type of catalyst, catalyst concentration, the number of equivalents of oxyamine systematically varied; while the concentration of lactose was kept constant at 290 mM. Conversion to the oxime **3.4**/*N*-glycoside **3.5** was then estimated by calculating the relative abundance of each species detected from the HILIC chromatogram. The screen was only performed once, as the intention was to use the data as a guideline to design further test reactions and be used to quantify the reaction yields.



Percentage conversion to 3.4 when R=H

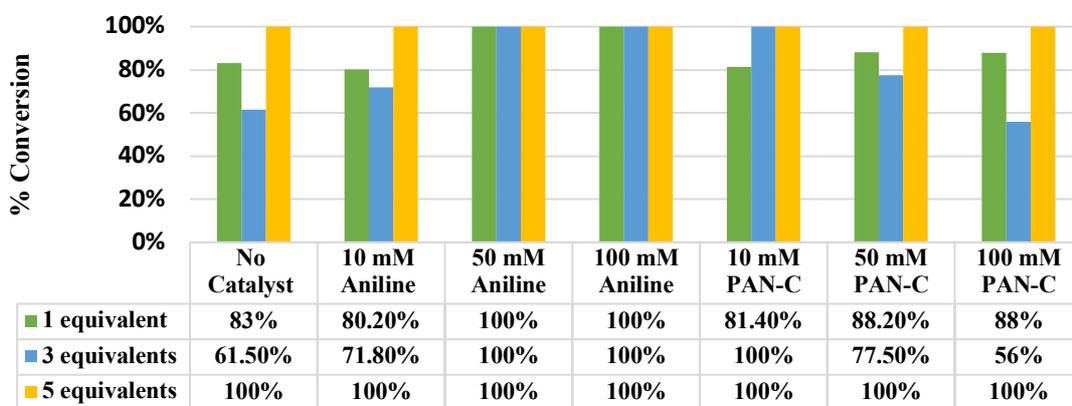


Concentration of Organocatalyst

■ 1 equivalent ■ 3 equivalents ■ 5 equivalents

Equivalents of *O*-methylhydroxylamine

Percentage conversion to 3.5 when R=Me



Concentration of Organocatalyst

■ 1 equivalent ■ 3 equivalents ■ 5 equivalents

Equivalents of *N,O*-dimethylhydroxylamine

Figure 3.1: Investigation of type of organocatalyst and effect of organocatalyst concentration using 1, 3 and 5 equivalents of A) *O*-methylhydroxylamine in an oxime ligation to lactose (290 mM), Conditions: Catalyst-

Acetate buffer pH4, 24 hours at room temperature. B) *N,O*-Dimethylhydroxylamine to lactose (290 mM), Conditions: Catalyst-Acetate buffer pH4, 24 hours at room temperature.

The organocatalyst does appear to have an effect on the ligation of *O*-methylhydroxylamine, giving the highest conversion at 50 mM for both aniline and PAN-C; a finding which is in accordance with results reported by Jensen and co-workers for PAN-C.²⁵⁴

From these model reactions, it was clear that a catalyst of 50 mM outperformed the 100 mM concentration. Following these results, the reaction shown in scheme 3.2 was repeated using 50 mM PAN-C/NaOAc buffer pH 4, and 50 mM aniline/NaOAc buffer at pH 4 carrying out the reaction at 290 mM lactose with one equivalent of linker **1.2** (290 mM). Despite the reactions being left for multiple days and monitored periodically by TLC and HILIC-MS, no formation of the oxime **3.3** was detected. This result was confounding, as the same reaction conditions which had produced 100% conversion in model reactions yielded no reaction with linker **1.2**. Although the model compounds used in the screen do not reflect the physiochemical or reactive properties of linker **1.2**, it would be expected that some conversion to **3.3** would be observed.

With the addition of a catalyst having no effect on the oxime ligation of linker **1.2** to lactose, it was noted that reaction between lactose model hydroxylamine reactants (figure 3.1) was occurring with no catalyst present; with the reactions being performed at a 4-fold increase in concentration than the reaction in scheme 3.2. Therefore, the reaction (scheme 3.2) was repeated increasing the concentration of both lactose and **1.2** to by a factor of 1.5 to 122 mM, without any catalyst present. After 24 hours formation of oxime **3.3** was observed by TLC and isolated in 30% yield. This was a promising result demonstrating the oxime ligation is heavily concentration dependent, a finding supported by a study by Nitz and co-workers on the stability of hydrazide/hydroxylamine derivatives of reducing sugars in aqueous solution, which reported small equilibrium constants for oxime/N-glycoside formation with three monosaccharides.²⁶⁴

Although a 30% isolated yield was promising and manageable when the oligosaccharide is available in sufficient quantities; the reaction with more complex and less reactive glycans, would probably not be achievable. We searched for other methods to use in conjunction with increasing concentration to improve the yield, when designing further reactions between linker **1.2** and lactose. Feizi and co-workers have used oxime ligation for the generation of glycan arrays, with the oxime ligation performed in chloroform/methanol mixtures.²⁶⁵ In other works Godula and Bertozzi have reported examples of high yielding oxime ligation in 1 M NaOAc buffer, with temperatures of up

to 50 °C used.^{266, 267} Jensen and co-workers did acknowledge that some oxime ligations performed at elevated temperatures, had outperformed reactions catalysed by aniline.²⁴⁶ Using this information, two different test reactions were performed using a CHCl₃:MeOH (1:1) mixture and 1 M sodium acetate buffer at pH 5 as solvents where both were heated to 50 °C.

3.1.1.1 Oxime ligation in chloroform:methanol

Performing the reaction in a chloroform:methanol mixture allowed higher concentrations of linker **1.2** (319 mM) to be achieved, as lactose (364 mM) could be added as a solid directly to **1.2** in solution (deprotection described in section 2.3.2 can be performed in CHCl₃:MeOH mixture). A caveat to adding solid lactose is the highly hydrophilic oligosaccharide is only sparingly soluble in this solvent mixture. To prevent evaporation at 50 °C, reactions were carried out in 200 µl Eppendorf tubes, which could then be heated in a (PCR) thermocycler with a heated lid; this provided a constant temperature while preventing evaporation and condensation of the solvent within the lid of the tube. The reaction in CHCl₃:MeOH showed successful reaction between lactose and linker **1.2** with the oxime product isolated in 68.5%. However, there was notable precipitation of lactose from the solution particularly on larger scale reactions, impacting the amount of lactose that could participate in the reaction. Larger oligosaccharides are almost certain to be less soluble than lactose in the CHCl₃:MeOH mixture, increasing the chances of precipitation of the glycan and difficulties in maintain a homogenous reaction mixture at the concentrations required for ligation. This method gives the oxime product in good yield, put there is difficulty in the handling of the reaction and may only be good for small very small-scale reactions.

3.1.1.2 Oxime ligation

Heating in sodium acetate buffer is a more practical approach for larger oligosaccharides, and also for larger scale reactions maintaining high concentrations of both reactants and keeping oligosaccharides in solution. The reaction was also carried out using a (PCR) thermocycler to hold the reaction at 50 °C, as it was found at such small volumes the reactions were prone to drying out with the water condensing within the lid of an Eppendorf tube.

The reaction was set up from the same stock solutions in the reaction in CHCl₃:MeOH (section 3.1.1.1), resulting in lower concentrations of lactose (200 mM) and **1.2** (234 mM) due to 1/5th the reaction volume being taken by a concentrated stock NaOAc solution

(5M). This could have been mitigated by increasing the concentration of the stock solutions, but the concentrations were above that known for successful reaction. After 18 hours at 50 °C, formation of **3.3** was observed by TLC and HILIC-MS analysis confirming the correct mass of 490.28 Da (expected $[M+H] = 490.52$ Da) for the product **3.3** (figure 3.2) and this was isolated in a 73% yield upon purification (outlined in section 3.1.2). NMR analysis of **3.3** revealed that in solution, **3.3** exists as three isomers, an *E/Z* isomer of the oxime and the ring closed N-glycoside in a ratio of 20:5:14 respectively (figure 3.4). This result confirmed that carrying out the reaction in buffer makes the handling of the reagents easier and offers a simpler route to scaling up the reactions, alongside a moderate increase in yield.

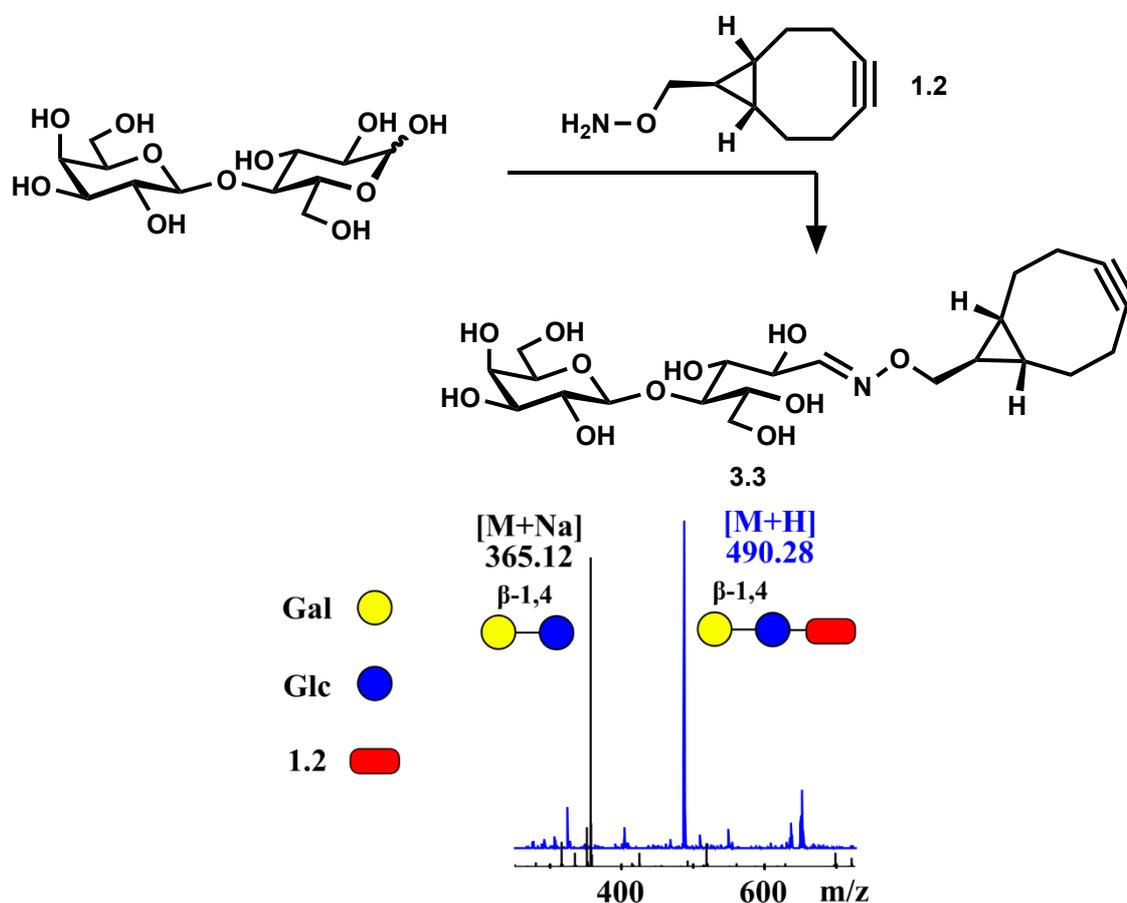
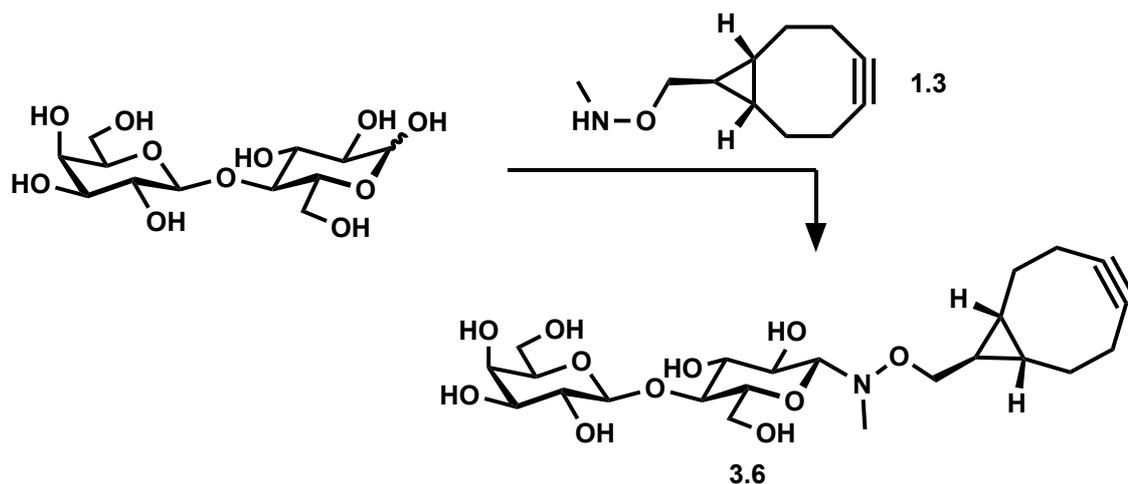


Figure 3.2: Mass spectrum comparison of lactose (black) and a mass spectrum of the reaction with **1.2** in 1 M NaOAc buffer at 50 °C taken after 18 hours, showing full conversion to the BCN ligated lactose **3.3** (blue).

After achieving good yields for the more reactive linker **1.2** with reducing sugars, the reaction was repeated using N-Methyl linker **1.3**, to create the lactose derivatives with a ring-closed structure. Under identical reaction conditions to linker **1.2**, in 1 M NaOAc at 50 °C, formation of the N-glycoside **3.6** was isolated in a yield of 40% as confirmed by HRMS $[M+H] = 504.2446$ Da (expected $[M+H] = 504.2445$ Da).



Scheme 3.3: Functionalisation of lactose with linker 1.3 to give the ring closed BCN derivitised N-glycoside 3.6.

3.1.2 Purification of BCN derivitised glycans from aqueous media

The purification of unprotected carbohydrates is often no trivial task. Many common purification techniques are incapable of purifying unprotected oligosaccharide derivatives, due to their highly polar nature and lack of solubility in solvents other than water. Conveniently, conjugation to linker **1.2/1.3** not only offers a reactive handle, but the hydrophobic properties of the BCN group mean that it can also be exploited as a purification tag for reverse phase chromatography. Adopting a capture-and-release approach to the purification of these BCN functionalised oligosaccharides, using commercial C18 cartridges, allows effective purification from the aqueous reaction mixture. A typical workflow for purification is displayed in figure 3.3: the aqueous reaction mixture is directly loaded following equilibration of the column in water; unreacted carbohydrates and salts are removed during the washing steps, whilst unreacted BCN and product are retained. The product can then be eluted from the column using 20-50% aqueous methanol, leaving any unreacted linker **1.2** stuck to the column.

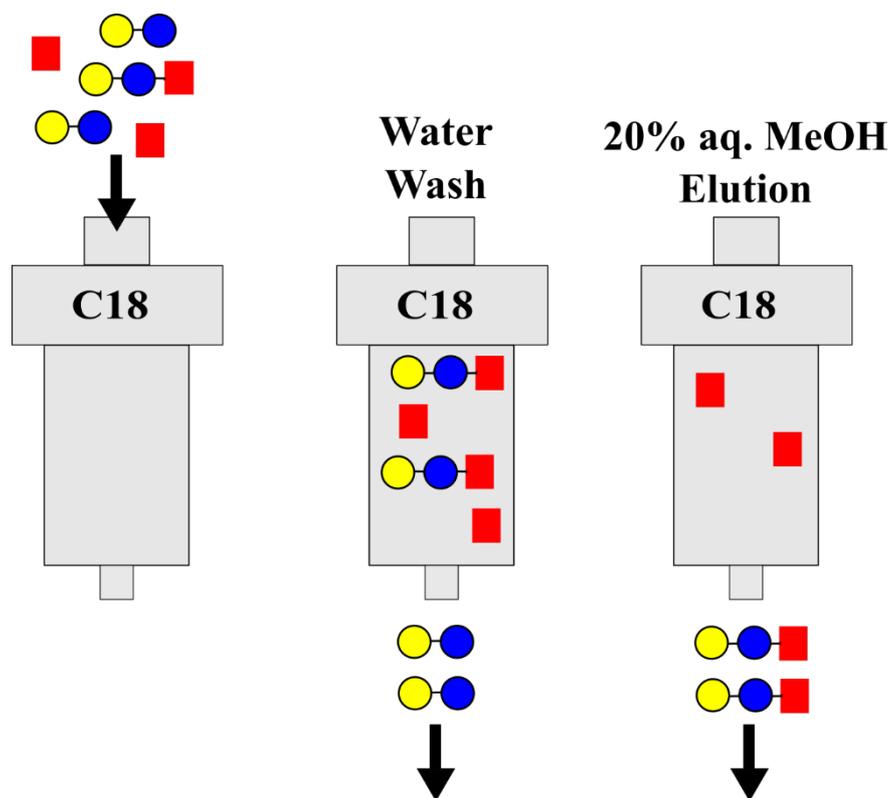


Figure 3.3: Cartoon representation showing the purification of glycan derivatives containing BCN by reverse phase chromatography from aqueous reactions mixtures. A) Crude reaction mixture is loaded in water to the C18 column. B) Unreacted oligosaccharide is washed from the column with aqueous wash. C) BCN functionalised glycans are selectively eluted from the column in 20% aq. methanol.

NMR analysis of the pure **3.3** showed that any unreacted linker **1.2**, did not co-elute with the product. The purified sample was analysed by NMR, with the integrations for signals corresponding to the BCN ring protons, matching those expected from the integrations of the anomeric protons for the three isomers of **3.3** observed in solution confirming the product **3.3** was 1:1 linker to glycan.

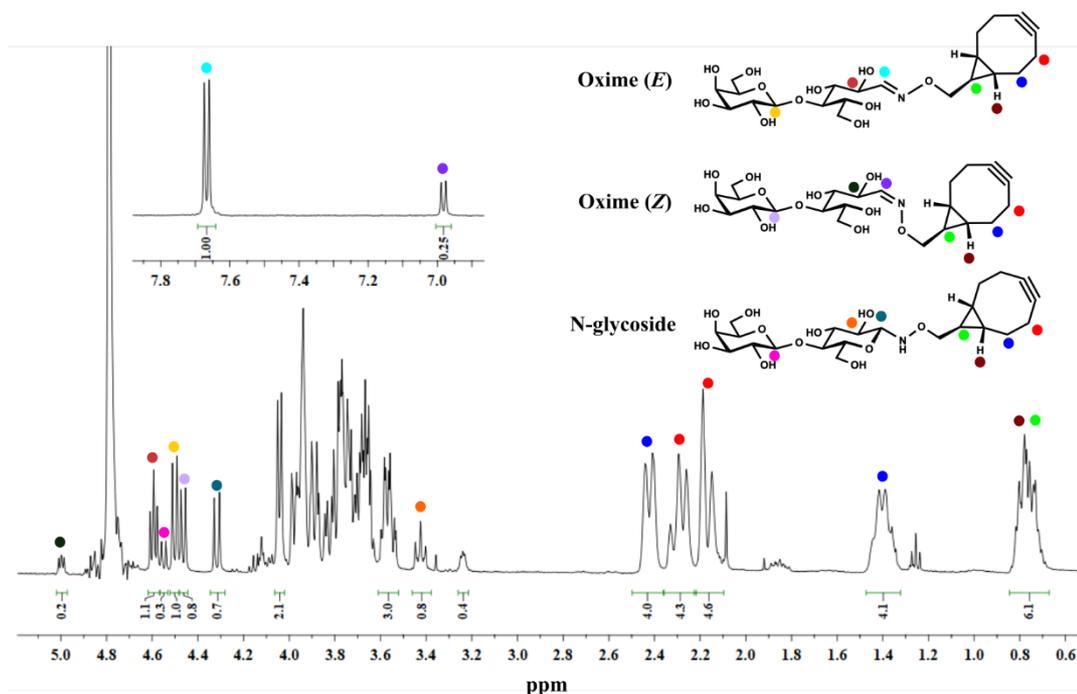


Figure 3.4: NMR spectra with inset showing a purified sample of oxime 3.3 assigning the BCN ring protons and the H₁ and H₂ protons for glucose and galactose residues. Integrations are shown confirming the linker and glycan are in 1:1 ratio for the assigned three observed isomers of 3.3.

3.2 Oxime ligation to pentasaccharide GM1os

After optimising conditions for the ligation of **1.2** to the simple disaccharide lactose, the next objective was to apply these methods to the synthesis of GM1 derivatives, which would be used to develop neoglycoprotein inhibitors of the cholera toxin B subunit. Before trialling the oxime ligation conditions on the GM1, the oligosaccharide first had to be obtained. GM1os is very expensive from commercial sources with 2 mg priced at £853. In contrast, the GM1-ganglioside is a fraction of the price at £325 for 500 mg. Therefore, for our applications it was more sensible to obtain the GM1 oligosaccharide from the ganglioside, through enzymatic cleavage of the glycosphingolipid using the known glycosyl hydrolase, endoglycosylceramidase (EGCase) II from *Rhodococcus sp.* (M-777).²⁶⁸

3.2.1 Expression of EGCase II and enzymatic hydrolysis of GM1 ceramide

Vaughan *et al.* used a codon-optimised gene, for the high yielding expression of EGCase II in *E. coli*.²⁶⁹ Following this work, a codon-optimised gene in the pET-28a expression vector (see appendix, figure 10.4) was purchased from GenScript incorporating an N-terminal His-tag. The plasmid was transformed into *E. coli* BL21 (DE3), which were grown at 37 °C before induction of protein expression at 20 °C by addition of isopropyl-

β -thiogalactopyranoside (IPTG). Any attempts to express the enzyme by cooling the culture slowly from 37 to 20 °C either showed no expression or only insoluble protein found in the cell pellet. Instead, it was found that a cold shock step was essential before induction of expression to enable expression of soluble protein. This was performed by rapidly cooling the cultures on ice, before addition of IPTG. The *E. coli* cold shock response appears to aid expression of stable and soluble protein – an important point that is not highlighted in the literature procedure. The enzyme was successfully purified by Ni-NTA chromatography in a yield of 30 mg/L and SDS-PAGE analysis (figure 3.5) confirmed overexpression with the band at 56 kDa corresponding to EGCase II. HRMS of the combined elution fractions also confirmed the correct mass of 51,942 Da.

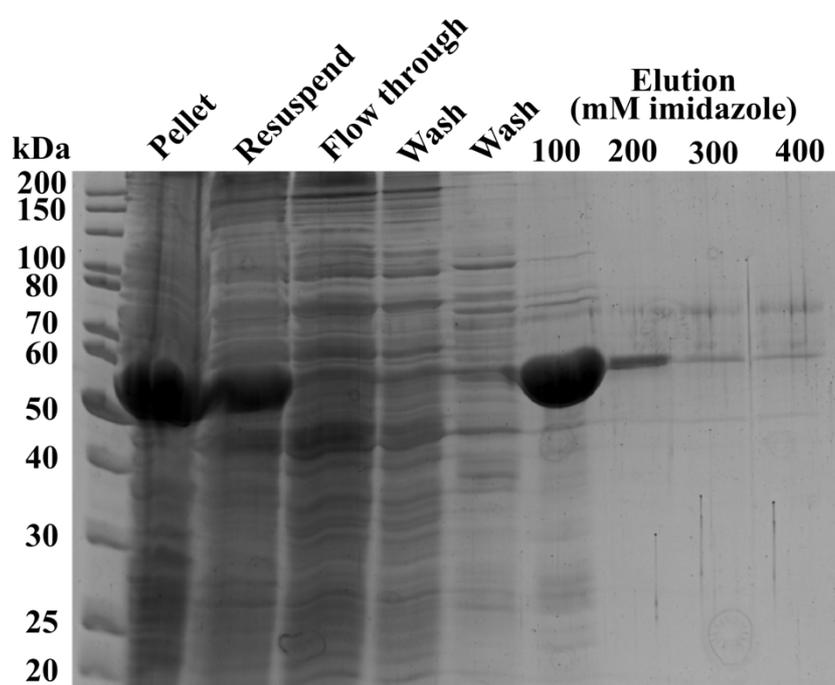
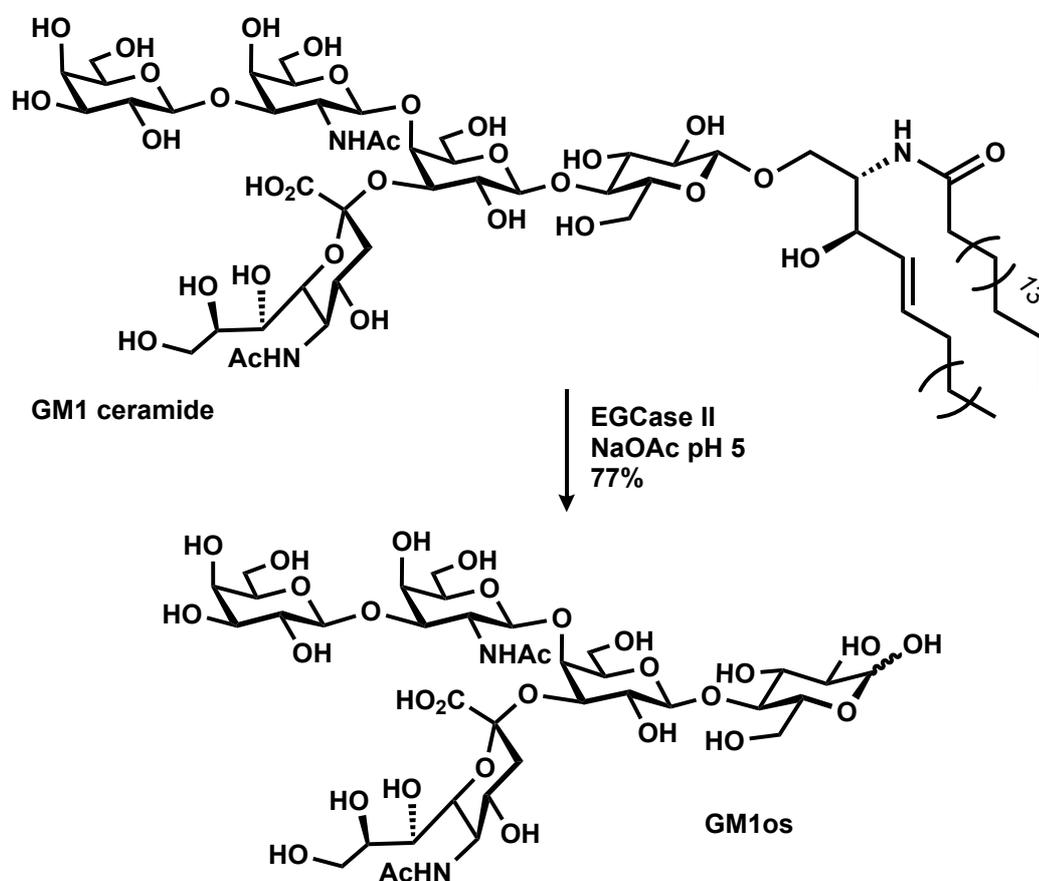


Figure 3.5: SDS-PAGE showing confirming expression of EGCase II, and purification by Ni-NTA column.

Removal of imidazole by dialysis into sodium acetate buffer as outlined in the Withers report²⁶⁹, resulted in precipitation of the enzyme, presumably as the buffer did not contain any salt. It was instead dialysed into a high salt (500 mM NaCl) phosphate buffer before flash freezing in liquid nitrogen and storage at -80 °C.

GM1-ganglioside purchased from Carbosynth as the disodium salt, was treated with the recombinant enzyme (scheme 3.4) as reported in the same paper by Vaughan *et al.*²⁶⁹ Hydrolysis was shown to be complete by TLC after 72 hours. After removal of the enzyme and ceramide by filtration and reverse phase chromatography, respectively, the GM1os was desalted using gel filtration giving an isolated yield of 77%. NMR analysis matched the reported literature values for GM1os.^{270, 271} This enzymatic approach was eventually scaled to produce 200 mg of GM1os, from EGCase II produced in a 200 ml

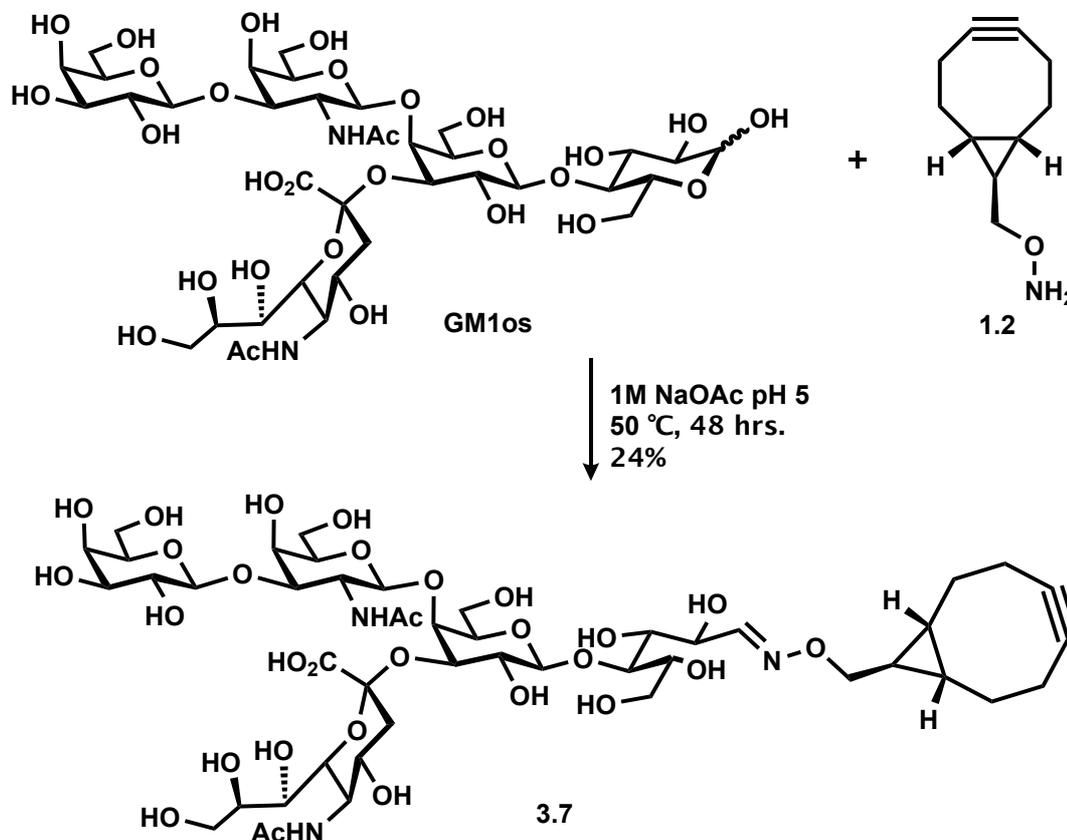
culture. This quantity of GM1os would have a commercial value of £85,000, making this a much more efficient approach than produce GM1os chemically or enzymatically, which would require three glycosyltransferases and the relevant nucleotide donors.



Scheme 3.4: Enzymatic synthesis of of GM1os from GM1 ganglioside, by ceramide cleavage with EGCase II.

The desalted GM1os and **1.2** were combined using the optimised conditions for the oxime ligation (scheme 3.5): 250 mM GM1os and 156 mM **1.2** were heated at 50 °C in 1 M NaOAc buffer, using the PCR thermocycler. After 48 hours the reaction appeared to be complete as TLC showed no further change, and a crude HRMS confirmed the correct mass $[M+H] = 1,146.4571$ Da (expected $[M+H] = 1,145.4486$ Da) for the GM1 oxime **3.7**. The reaction was then diluted and purified by reverse phase chromatography (outlined in section 3.2.2). The BCN group was sufficiently hydrophobic to ensure **3.7** was retained on the C18 column, with no evidence of **3.7** in the flow-through or wash fractions taken. After elution, **3.7** was obtained in an isolated yield of 24%. As **3.7** was produced in 0.28 mg quantities it was not possible to fully assign the NMR spectrum for the pentasaccharide derivative. Spectral analysis was able to confirm the presence of the BCN, a shift of the anomeric signal corresponding from α/β hydroxyl signals, to signals for the *E/Z* isomers of the oxime, observed at 7.68 and 7.00 ppm respectively in correspondence with the signals observed for oxime **3.3**. The integrals for the *E/Z* isomers

of the oxime proton were observed in the same ratio as seen with the oxime product **3.3**, and the integrations of BCN ring protons matched what would be expected for the presence of three isomers. HRMS was taken for the purified sample to corroborate NMR spectral data of **3.7** confirming that the correct product was obtained after purification.



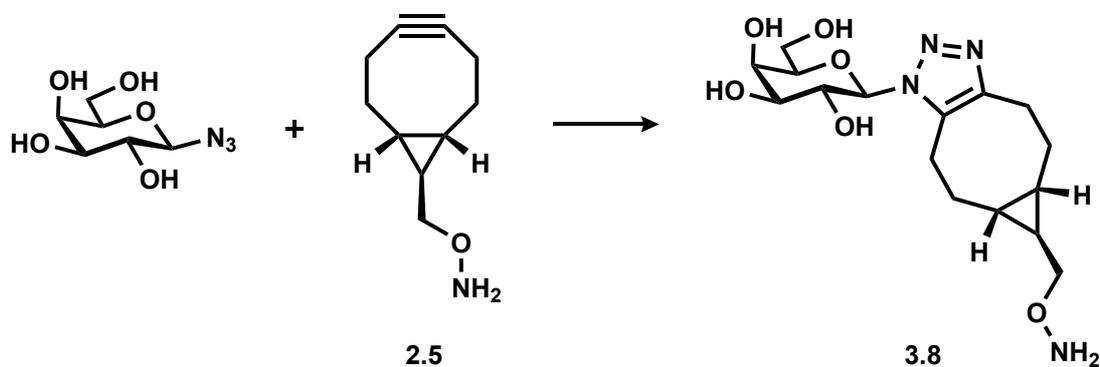
Scheme 3.5: Oxime ligation of linker 1.2 to GM1os to give oxime 3.7.

Isolation of pentasaccharide **3.7** demonstrated the BCN group is suitable as a purification tag for glycans up to five saccharide units including those containing sialic acids, by a catch-and-release on small C18 cartridges. Though there will be a glycan size where the BCN is no longer capable of retaining the oxime product a study on this is required to divulge the length of polysaccharide that cannot be retained by C18 cartridge, and by extension if longer reverse-phase columns or HILIC-HPLC would be better suited for larger polysaccharides.

3.3 Synthesis of glycosyl azides and functionalisation through SPAAC

Having successfully developed oxime ligation reaction conditions for linkers **1.2** and **1.3**, **1.2** was then used to synthesise a second type of glycan derivative through reaction with a glycosyl azide. Use of the SPAAC reaction as shown could provide a highly efficient method of making reactive glycan derivatives with a reactive oxyamine functionality for

oxime ligation to a biological macromolecule (scheme 3.6). This strategy however does require chemical synthesis of glycosyl azides before they can be used in this type of reaction.



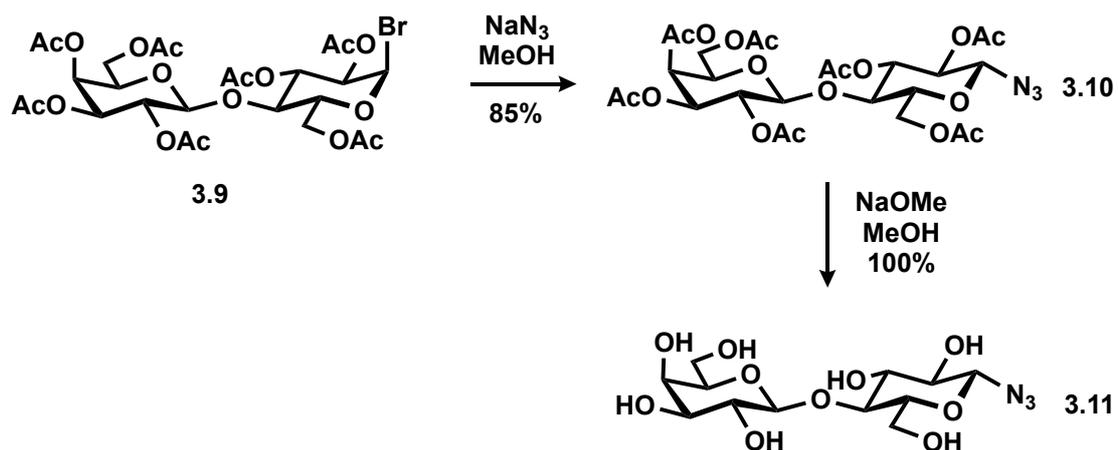
Scheme 3.6: Example of a SPAAC reaction between linker 1.2 and β -galactosyl azide to produce a triazole-linked monosaccharide (3.8) with an oxyamine functional group.

Lactose was once again used as a model oligosaccharide for the development and optimisation of conditions, and as a proof of principle for employing SPAAC to make derivatise glycosyl azides using linker 1.2. Azide analogues of lactose and GM1 were therefore required.

3.3.1 Synthesis of lactosyl azide

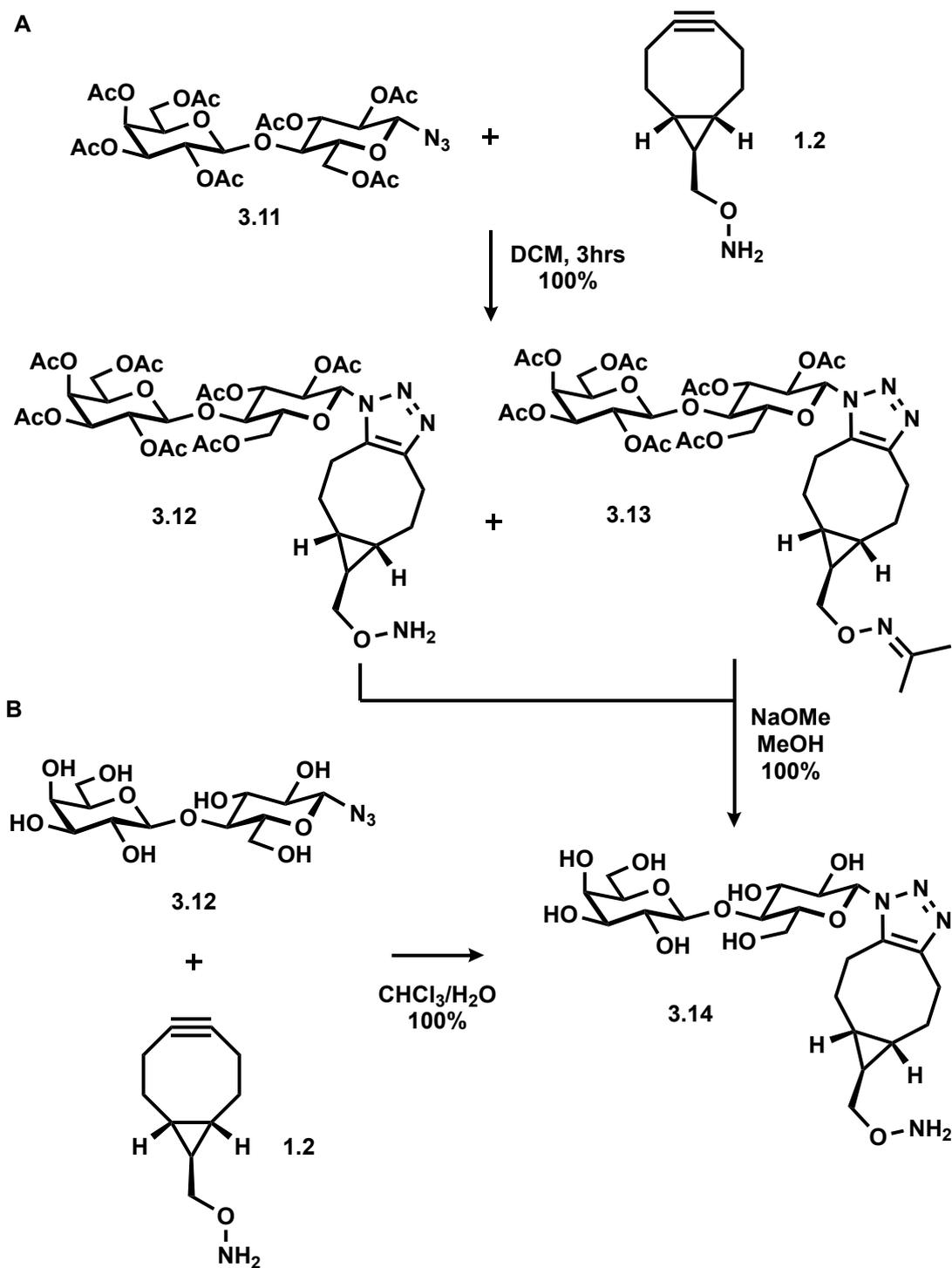
For simple, commercially available mono- and disaccharide products, conversion to glycosyl azides is a straightforward process. One-pot global acetyl protection of the hydroxyl groups and activation of the anomeric centre as a glycosyl bromide, can then be converted to an azide by nucleophilic substitution yielding exclusively the β -glycosyl azide. The synthetic route to azido lactose was therefore undertaken chemically (scheme 3.7) to produce larger scale quantities of β -azidolactose.

Beginning from lactose, a one-pot peracetylation, and bromination was performed with acetic anhydride and HBr to give the α -bromo lactose in a single step with an 89% yield.²⁷² After purification of 3.9, the bromide was displaced by S_N2 substitution using sodium azide, to give peracetylated lactosyl azide in 85% isolated yield, which was then deprotected using sodium methoxide to yield the lactosyl azide 3.11.



Scheme 3.7: Synthesis of of azidolactose from peracetylated α -bromolactose.

SPAAC ligation with **1.2** was attempted on the protected lactosyl azide **3.10** to allow the reaction to be performed in a single-phase solvent system (scheme 3.8A). Using linker **1.2** which had been stored under an inert atmosphere, the reaction was shown to be quantitative after stirring overnight.



Scheme 3.8: Generation of oxyamine-derivatised lactose 3.14 from lactosyl azide and linker 1.2. Conditions: b) DCM, 3 hours, quantitative. c) NaOMe, MeOH, 1 hr, quantitative.

However, as the reaction was not also performed under an inert atmosphere, the oxyamine was found to have reacted with acetone, forming oxime **3.13**. Upon purification, this by-product was found to be the major product with the desired oxyamine **3.12** being converted to **3.13** upon prolonged exposure to the acetone in the laboratory atmosphere. Although this by-product was no longer useful, it did show that the SPAAC reaction was a highly efficient method for making oxyamine functionalised oligosaccharides that can

be used in oxime ligations. Upon attempting to remove the acetyl groups of oxime **3.13**, the isopropylidene oxime could also be removed using the deprotection conditions stirring in sodium methoxide to give **3.14**. However, this required unnecessary deprotection steps and isolation of the free oxyamine still posed a challenge to avoid the purified **3.14** forming the isopropylidene oxime.

Synthesis of the deprotected lactosyl derivative **3.14**, was instead carried out from the crude deprotection mixture of **1.2** vigorously stirred in H₂O with **3.11** in equimolar quantities. After vigorous stirring with a solution of **3.11** in water as shown in scheme 3.8B, reaction was shown by TLC to be complete after three hours, which was corroborated by HILIC-MS. Figure 3.6 shows the overlaid spectrum of **3.11** before the reaction, and a spectrum taken after three hours, showing only the mass of **3.14**. After lyophilisation, **3.14** was isolated without any formation of oxime, minimising the exposure of the product to acetone by carrying out the reactions in the biochemistry laboratory. This model reaction proved that triazole linked glycosides could be produced by SPAAC, whilst maintaining the oxyamine, with no need for further purification.

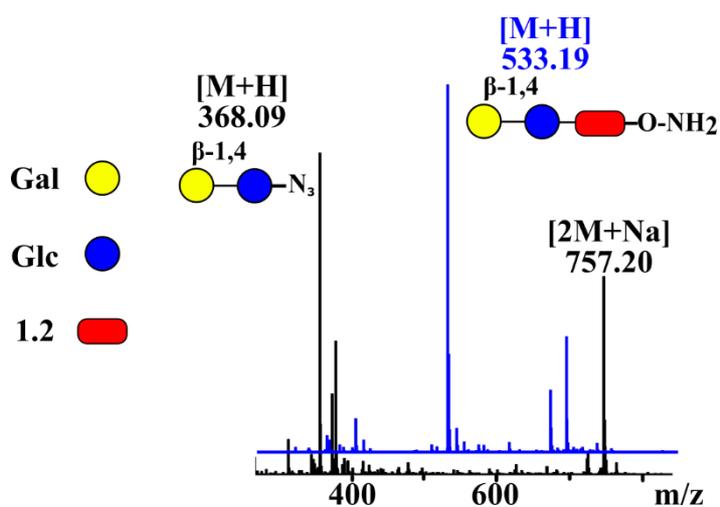


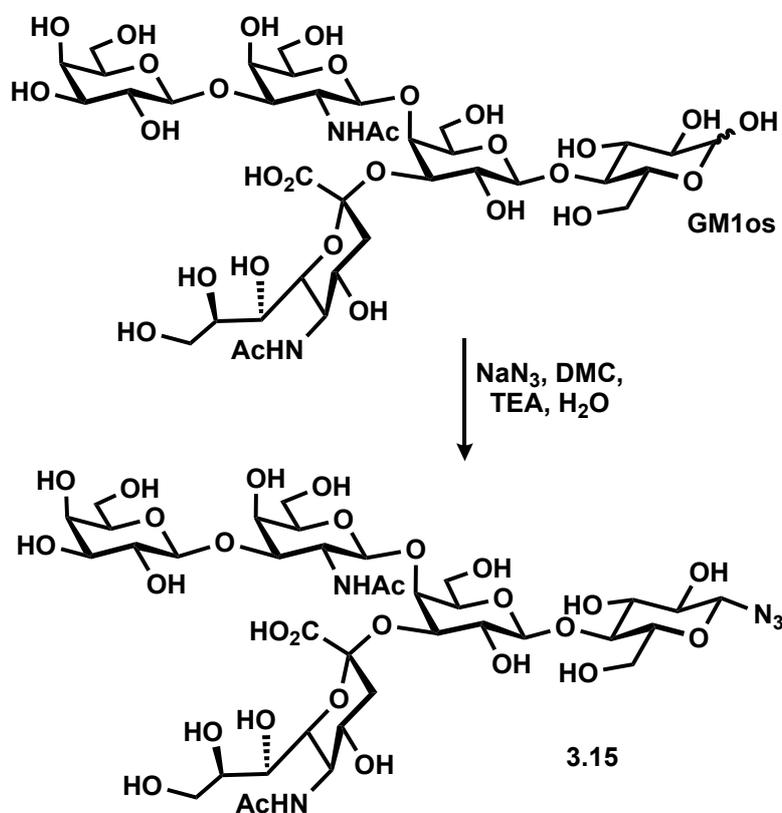
Figure 3.6: Mass spectrum comparison between lactosyl **3.12** (black) and the SPAAC conjugation reaction between **3.16** and **1.2** (blue).

3.3.2 Synthesis of GM1-N₃

For the synthesis of GM1 azide we did not want to follow the same chemical modification used for the synthesis of lactosyl azide **3.11**; these types of chemical modification are not always practical for much larger oligosaccharides and complex glycans, as often the glycans are not available in sufficient quantities to carry through multiple protection/deprotection and purification. A particularly simple approach for the production of glycosyl azides from reducing sugars was reported by Shoda and co-workers, in which the anomeric hydroxyl group was activated using 2-chloro-1,3-

dimethylimidazolium (DMC).²⁷³ Large oligosaccharides up to decasaccharides have been shown to be converted to the corresponding β -azides in the presence of DMC and a base. The simplicity of this reaction to yield a single isomer, makes it an attractive method for the synthesis of large, complex glycosyl azides.

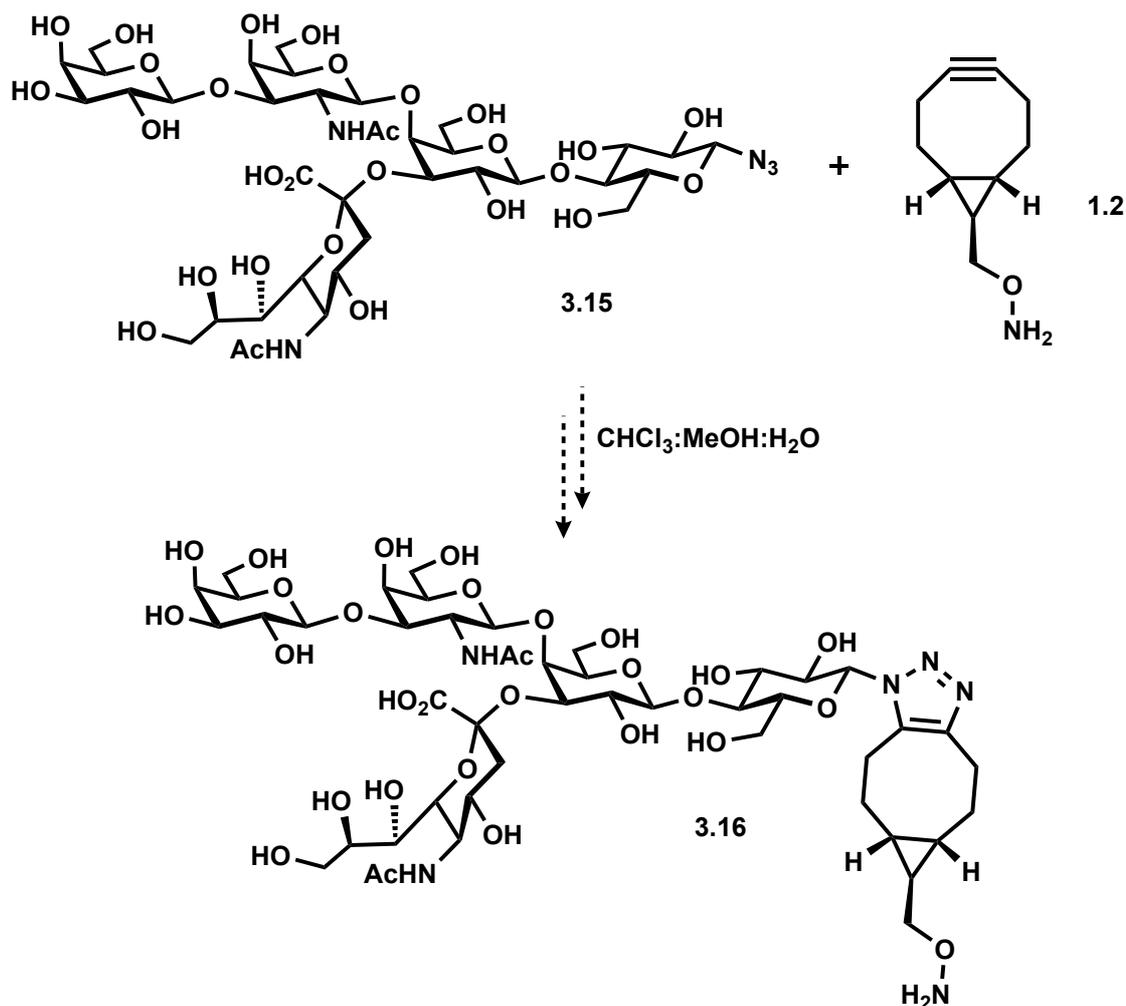
Following the original protocol²⁷³, GM1os was added to a solution containing DMC, sodium azide and 2,6-lutidine as reported as examples of this reaction on sialylated sugars. This required longer reaction times and required much higher equivalents of the base. Other examples by Shoda and co-workers used DIPEA and TEA were used as stronger bases, with TEA being supported by Machida *et al.*²⁰¹ it was decided to attempt the conversion of GM1os to GM1-N₃ **3.15** using nine equivalents of TEA (scheme 3.9) Due to the sub milligram scale of this reaction, no analytical data could be obtained by NMR, and common positive and negative ESI-HRMS methods were not capable of ionising the GM1os or GM1-N₃. Monitoring the reaction progression by TLC proved more convenient, as staining in orcinol (stain specific for carbohydrates), showed a shift in R_f indicating reaction had occurred, however, after purification by gel filtration, NMR spectral analysis was not able to determine if the product isolated was indeed **3.15**, due to the poor signal-to-noise in the obtained spectra.



Scheme 3.9: One-pot synthesis of GM1-N₃ **3.15** using Shoda's reagent and triethylamine in water at room temperature.

The publication by Machida *et al.* used Shoda's reagent for the synthesis of glycosyl azides, but adopted a modified purification protocol.²⁰¹ They reported high isolated yields, by acetylating the carbohydrate upon completion of the azidation, followed reverse phase chromatography, and finally acetyl deprotection in the presence of sodium methoxide to the desired unprotected upon neutralisation. This is a much more universal approach, not requiring specific purification equipment and easily purified by the reverse phase purification outlined in section 1.1.2.

After the large 200 mg batch of GM1os had been produced, this acetylation purification was to be attempted in the final few weeks of this project using 10-20 mg of the GM1os for the azidation. An NMR time course could be performed to confirm the formation of **3.15**, by change in the anomeric signals for the reducing glucose unit, and further plans to pair this spectral analysis with MS data was devised using a 'clickable' MS tag (see chapter 5, section 5.2.2.1) to aid detection of the GM1 azide **3.15** after purification from the reaction mixture. If GM1 azide **3.15** can be obtained with a high purity, the SPAAC reaction in scheme 3.10 could be carried out with no further purification required before ligation to a protein scaffold.



Scheme 3.10: Proposed reaction for the generation GM1 derivative 3.16, by SPAAC reaction between GM1 azide and linker 1.2.

3.4 Enzymatic extension of un-natural glycan derivatives

Following synthesis of four novel glycan derivatives through oxime ligation and SPAAC with bifunctional linkers **1.2/1.3**, thoughts turned to using enzymatic synthesis to build more complex structures that already incorporate the tag.

Once again derivatives of small mono- or disaccharides are good models to explore if the promiscuity of glycosyltransferases towards the unnatural substrates, as with the short distance between the terminal sugar and the linker, they are least likely to be recognised as substrates. If unnatural disaccharide substrates could be accepted by wild type glycosyltransferases, then by analogy larger polysaccharide structures are likely to be accepted.

Using the lactosyl derivatives **3.3** and **3.14** as substrates, enzymatic synthesis would be used to create the Gb3 structure, the ligand for the verotoxin B subunit. Synthesis of the

Gb₃ structure requires the addition of a galactose to **3.3** with an α -1,4 glycosidic link. Enzymatically, this requires an α -1,4 galactosyltransferase and UDP-Gal as a glycosyl donor. Comparatively, chemical synthesis of the glycosyl donors and acceptors to construct the Gb₃os trisaccharide would require 13-17 steps before the oxime ligation reaction could be performed.^{274, 275}

3.4.1 Expression of α -1,4 galactosyltransferase

LgtC is a known α -1,4-galactosyltransferase from *Neisseria meningitidis*²⁷⁶, active WT LgtC has previously been recombinantly expressed in *E. coli* by Wakarchuk *et al.*²⁷⁷ A codon-optimised gene for LgtC with an N-terminal His-tag in a pET-28a expression plasmid was purchased from GenScript (see appendix, figure 10.5). Using a modified expression approach to that reported by Wang and co-workers,²⁷⁸ *E. coli* BL21 (DE3) was transformed with the plasmid, and overexpression of the enzyme was carried out in LB auto-induction media. Figure 3.7 shows successful expression of soluble enzyme upon cell lysis by sonication, and purification by Ni-NTA affinity chromatography, yielding LgtC in 46 mg/L quantities. HRMS confirmed that the observed mass within the isotopic distribution 34,456 Da, was within error at the resolution of the deconvolution of the theoretical mass calculated using bioinformatics tool protparam of 34,452.96.

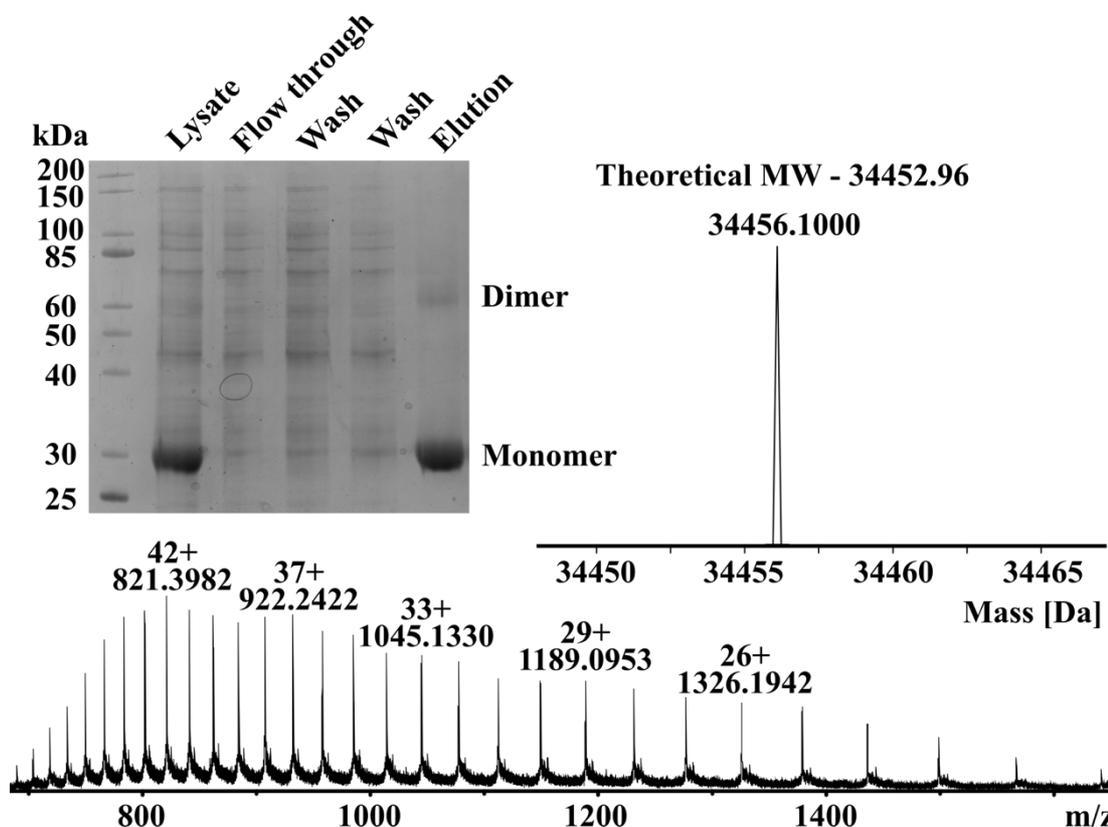
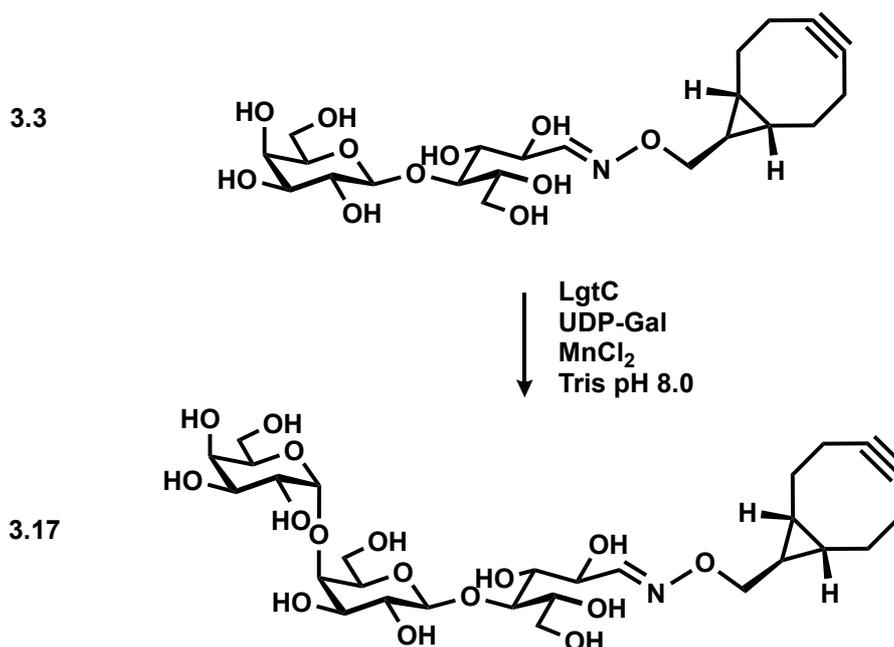


Figure 3.7: Overexpression of LgtC and purification by Ni-NTA chromatography, visualised by SDS-PAGE and high-resolution mass spectrometry of the elution fraction showing both the charge state distribution and deconvolution inset.

3.4.2 Testing enzymatic activity of LgtC

Enzymatic activity was initially confirmed using the natural substrate lactose (10 mM) and donor sugar UDP-Gal (10 mM), in the presence of 10 mM MnCl₂ as reported by Wang and co-workers.²⁷⁸ After 15 hours at 37 °C the reaction was analysed by HILIC-MS confirming the formation of Gb3os detecting [M+Na] = 527.45 Da (expected Gb3os+Na = 527.17). After confirming the enzyme was indeed active, initial enzymatic reactions carried out at 58 μM LgtC for the conversion of **3.3** to Gb3 oxime **3.17** (scheme 3.11) did not show full conversion when using two equivalents of UDP-Gal.



Scheme 3.11: Conditions for the enzymatic galactosylation of **3.3** using the galactosyltransferase LgtC and the galactosyl donor UDP-Gal.

Conversion of **3.3** to **3.17** was then tested at various concentrations of LgtC to determine the optimal conditions for enzyme activity. Table 3.1 shows the percentage conversion of **3.3** to **3.17** at 10, 50 and 100 μM of LgtC at 37 °C, measured from the relative abundance observed in the HILIC chromatogram.

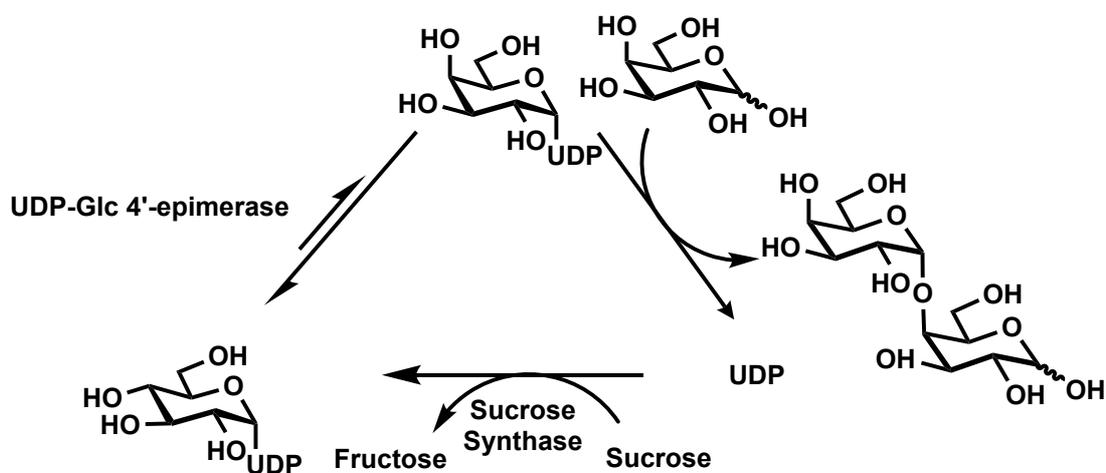
Table 3.1: Percentage conversion of α -1,4-galactosylation of lactose in a pH screen for optimal galactosylation using LgtC and UDP-galactose. Conditions: 1 mM **3.3**, 2 mM UDP-Gal, 1 mM MnCl₂ 1 mg/ml BSA, 50 mM Tris, 100 mM NaCl at 37 °C for 18 hours.

[LgtC]	Conversion to 3.17 (%)
10 μM	0
50 μM	50
100 μM	60

As would be expected the increase in enzyme concentration leads to an increase in conversion, as the rate of reaction is proportional to enzyme concentration but full conversion of **3.3** was not observed. It was proposed that the limited conversion could be explained by several possible factors: the rate of enzyme-catalysed hydrolysis of the UDP-galactose being greater than rate of glycosylation of **3.3**; poor binding because of saccharide conformation or steric clash; or that the pH is not optimal for the catalytic residues within the active site of LgtC.²⁷⁶ Conversion may be improved by changing buffer, as it is known that LgtC exhibits optimal activity in HEPES, however, wanting to combine this enzyme into a catalytic cycle the reactions needed to be carried out in Tris to be compatible with other enzymes.^{279, 280}

3.4.3 Enzymatic cycle for continuous glycosyl donor production

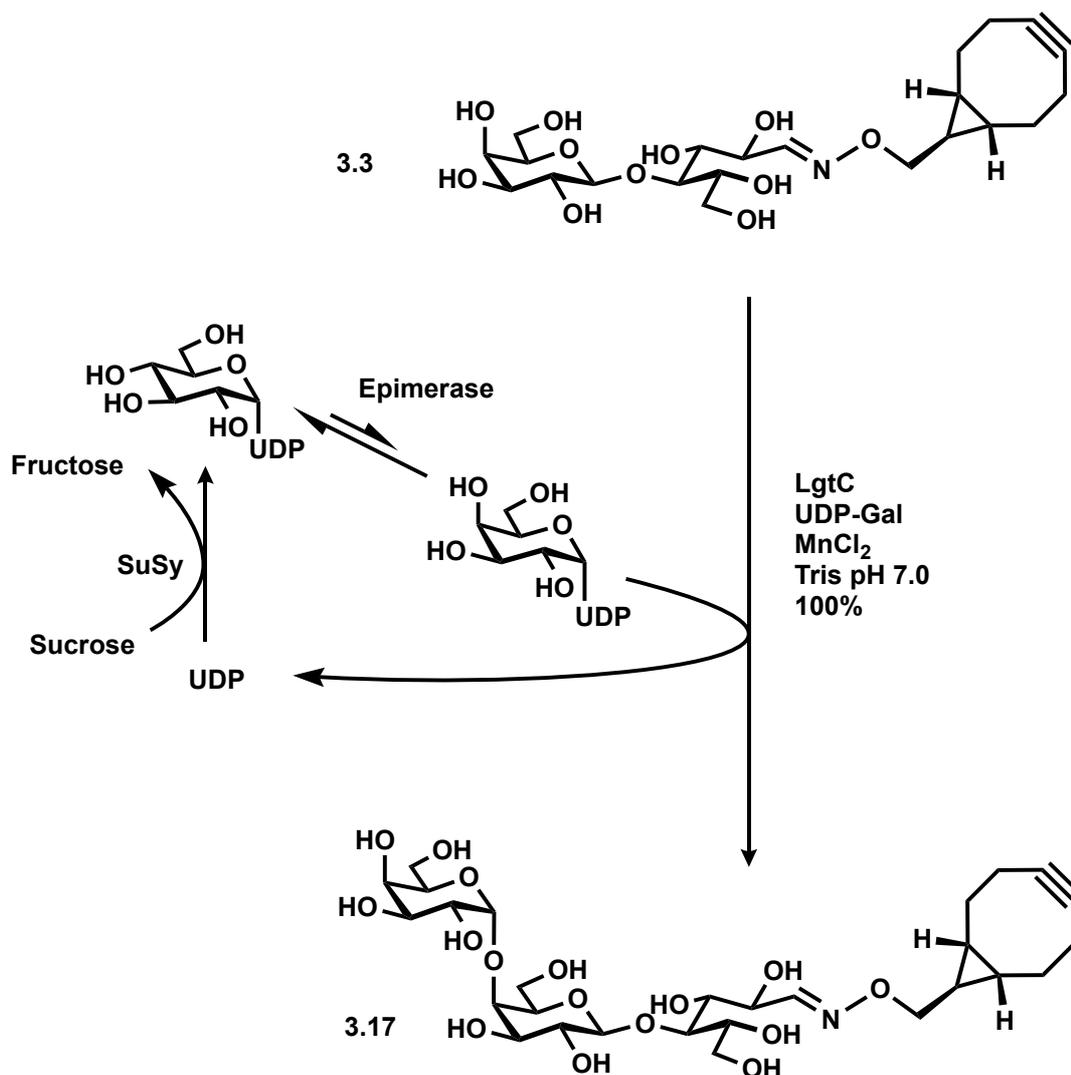
In showing that LgtC could turn over the lactosyl BCN derivative **3.3** to yield the Gb3 BCN **3.17**, we had also developed a route to the reactive glycan ligands for the verotoxin. To enable scale-up of the enzymatic synthesis, an enzymatic cycle in which the UDP-donor sugar is continuously regenerated *in situ*, is useful to reduce the number of equivalents of the donor needed for full conversion to the trisaccharide. Such a cycle, similar to that described by Elling and co-workers,²⁸¹ is shown in scheme 3.12 in which UDP-galactose is generated *in situ* from cheaper commercially available materials UDP or UDP-glucose. In combination with α -1,4-galactosyltransferase LgtC, this would be a viable method for scaling-up the synthesis of Gb3os and Gb3 derivatives without requiring large amounts of expensive UDP-galactose.



Scheme 3.12: Scheme showing the three-enzyme cycle of Sucrose synthase, UDP-Glc 4' epimerase and a galactosyl transferase for the regeneration of UDP-Gal and α -1,4 galactosylation developed by Elling and co-workers.²⁸¹

The enzymatic cycle can begin with UDP, UDP-Glc or UDP-Gal and is driven by the large excess of sucrose which is digested to UDP-Glc and fructose, by the backward reaction of the enzyme sucrose synthase. UDP-Glc is then converted to the UDP-Gal donor by epimerisation of the 4-hydroxyl using a UDP-Glc-4-epimerase. Finally, the UDP-Gal is recognised by the α -1,4-galactosyltransferase to create Gal- α -1,4-Gal product.

UDP-glucose purchased from Carbosynth was used as the feed for this cycle, and enzymes glucose-4-epimerase and sucrose synthase (SuSy) were obtained from Prozomix as ammonium sulfate precipitates as a part of a BBSRC collaborative project within the group. The cycle was set up as shown in scheme 3.13 using 10 mM **3.3**, 0.2 mol% of the UDP-Glc, in the presence of a large excess of sucrose (0.5 M), with MnCl₂ (1 mM) and 1 mg/ml BSA in Tris-buffered saline (TBS, 50 mM Tris 100 mM NaCl). A higher pH was not used for the enzymatic cycle, although it does improve the activity of LgtC on the lactosyl BCN **3.3**, it is likely not to be optimal for the activity of the other enzymes (Glc-4-epimerase and SuSy); therefore the reactions were performed at pH 7 adhering to literature reports. SuSy and the epimerase were collected as precipitates by centrifugation, resuspended, in ddH₂O, and added directly to the buffered reaction mixture. LgtC was used at 100 μ M according to previous findings showing greater conversion, and the high yield expression meant the enzyme could be produced in sufficient quantities to support its use at high concentrations.



Scheme 3.13: A three enzyme cycle for the continuous regeneration of UDP-Gal donor from UDP, and the galactosylation of oxime 3.3 to trisaccharide 3.17.

After incubation at 37 °C for 18 hours, MS analysis of the enzymatic cycle shown in figure 3.8 showed complete conversion to **3.17**. The continuous regeneration of UDP-Gal removes the effect of enzyme-catalysed hydrolysis of UDP-Gal, driving the conversion of unnatural glycosyl acceptor substrates towards completion through constant supply of the glycosyl donor. Furthermore, the recycling of UDP removes the potential inhibition of the GalT by UDP occupation of the active site. Investigation into the amount of sucrose required to drive the cycle, revealed that full conversion can be achieved using as little as a five-fold excess of sucrose, in conjunction with a 0.1 mol% ratio of the UDP-Glc. Purification and isolation of Gb3 derivatives by reverse phase chromatography was very effective, despite the complex composition of the enzymatic cycle. Any precipitated enzyme was removed by centrifugation, followed by the purification protocol outlined in section 3.1.2.

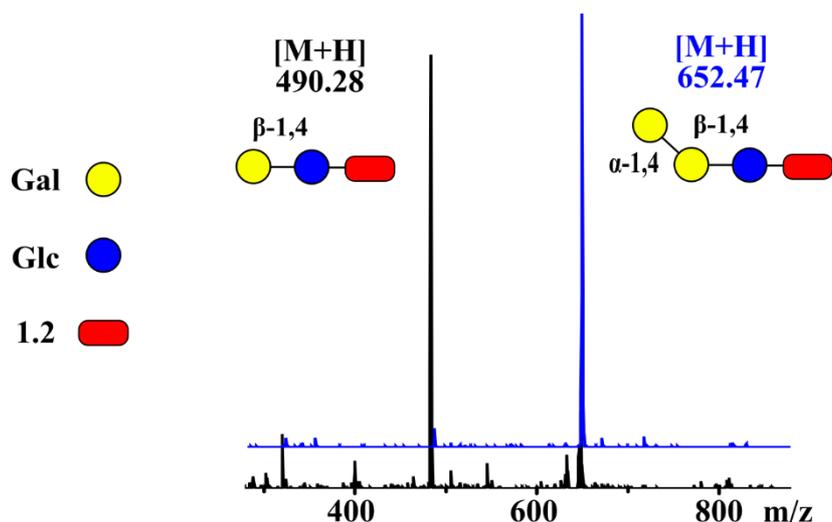
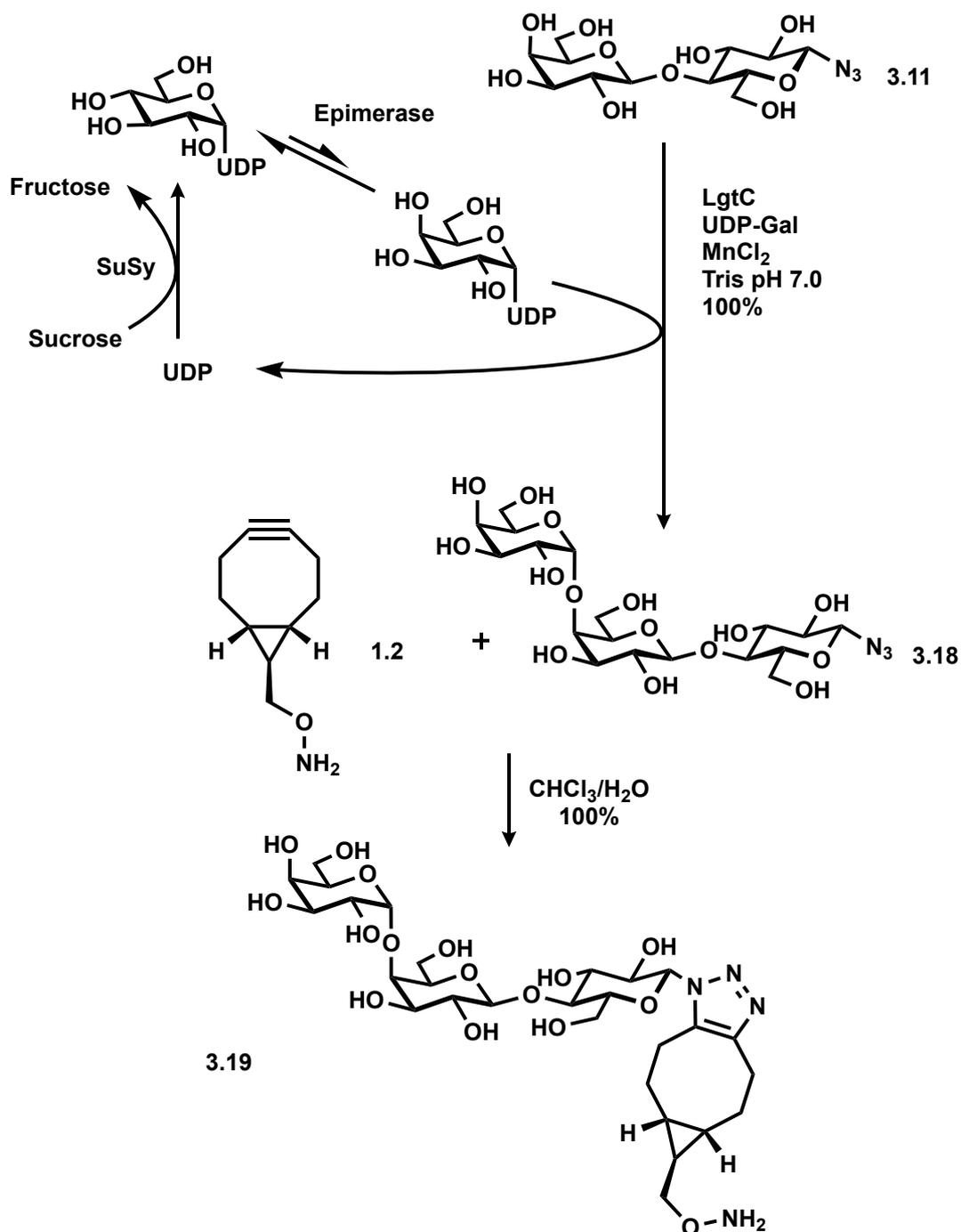


Figure 3.8: Mass spectra of the enzymatic glycosylation of 3.3 to 3.17 with 100 μ M LgtC after 18 hours run on HILIC-MS. Black – Compound 3.3. Blue – Compound 3.17.

With the SPAAC reactive Gb3 derivative **3.17** synthesis optimised, the enzymatic cycle was also used to produce the oxime reactive Gb3 derivatives. With the concern that the oxyamine could form unintended oximes, the glycosylation reaction was used to first create β -azidoGb3 from β -azidolactose **3.11**. Using the optimised cycle conditions **3.11** was converted to trisaccharide **3.18** as shown in scheme 3.14, before excess sucrose in the reaction mixture was hydrolysed by the addition of the enzyme invertase purchased from Sigma-Aldrich. TLC analysis was used to determine that degradation of the sucrose was complete, after which the **3.17** was isolated by size exclusion chromatography (SEC) using P2 gel, which was able to separate the trisaccharide from the monosaccharides and salts. After isolating **3.18**, the SPAAC conjugation to **1.2** gave the oxyamine **3.19**. HRMS confirmed the mass of the trisaccharide **3.19** at $[M+H] = 695.2956$ Da (expected $[M+H] = 695.2909$ Da).



Scheme 3.14: Enzymatic cascade for the production of 3.18 from 3.11, using UDP-Glac as the cycle feed at 0.1 mol% and 5-fold excess of sucrose; and reaction of the trisaccharide 3.18 with linker 1.2 to yield 3.19 in quantitative yield.

3.5 Conclusion

Linkers 1.2 and 1.3 have demonstrated versatility in conjugation methods, with each functional group capable of reacting with reducing sugars or glycosyl azides to produce simple and complex bioorthogonally reactive oligosaccharides. Oxime ligation was used to obtain lactosyl derivatised 3.3 and the ring closed analogue 3.4 in very good yield after

optimisation, which were later applied to the larger oligosaccharide GM1 isolating **3.7** respectable isolated yields.

The SPAAC reaction proved a very effective method for the synthesis of functionalised oligosaccharides containing an oxyamine group. The reactions were rapid and required little purification. Although time constraints did not allow for the isolation of GM1 azide **3.15** in sufficient quantities to be taken forward for the synthesis of GM1 derivative **3.16**, a synthetic plan and purification method from the reports by Machida *et al.*²⁰¹ holds promise for the isolation of **3.16**.

Enzymatic carbohydrate synthesis has been a very useful tool in expanding the complexity from simple, readily available oligosaccharides. These unnatural substrates can be recognised by glycosyltransferases, and the use of enzymatic cycles also provides a practical route to larger scale production, if access to the glycosyl donors and glycosyltransferases is available. Throughout the synthesis we have also exploited the reactive handle as an effective purification tag, demonstrating the BCN group is a highly effective for isolation from aqueous reaction mixtures, in a very efficient and reproducible manner. The combination of direct carbohydrate functionalisation, chemical modification and enzymatic glycosylation has produced one reactive GM1 derivative and two Gb3 derivatives with different reactivities, which can then be used for generation of homogeneous synthetic glycoproteins.

Chapter 4 Bioorthogonal synthetic glycosylation

Having discussed the use of two different reactive groups of bioorthogonal linkers outlined in chapter two to derivatise oligosaccharides. This chapter will explore the use of the same reactive groups to functionalise proteins with the glycan derivatives described in chapter three, to address the question in how best to design neoglycoprotein inhibitors. Having previously outlined some approaches devised for engineering protein scaffolds for use as inhibitors of bacterial toxins section 1.3.3, it is evident that covalent attachment and prearrangement of the glycan ligands is beneficial for inhibition of the WT cholera toxin. Figure 1.15 displays the oxime ligation of a GM1 glycoconjugate to the mutated CTB scaffold to produce a neoglycoprotein inhibitor.¹²⁸ As proposed in the project aims we wanted to explore the effects of linker length and position of the glycans, on the inhibitory potential of CTB-based neoglycoprotein inhibitors. To investigate this the short-length BCN linkers **1.2** and **1.3** were synthesised with two different reactive groups, to carry out glycosylation on the two opposing faces of mutant CTB proteins which can no longer bind to the native GM1 ligand; therefore, can be used as protein scaffolds which can be built upon with carbohydrates to target bacterial toxins (figure 4.1). Ligation on top face (non-binding face) of the protein, involves the introduction of non-canonical amino acids to introduce an azide capable of reaction with glycan derivatives with a strained alkyne. The other follows in the footsteps of Branson *et al.*¹²⁸ using the already established oxidation of β -aminoalcohols with NaIO_4 (Malaprade oxidation) for the oxidation of followed by oxime ligation.

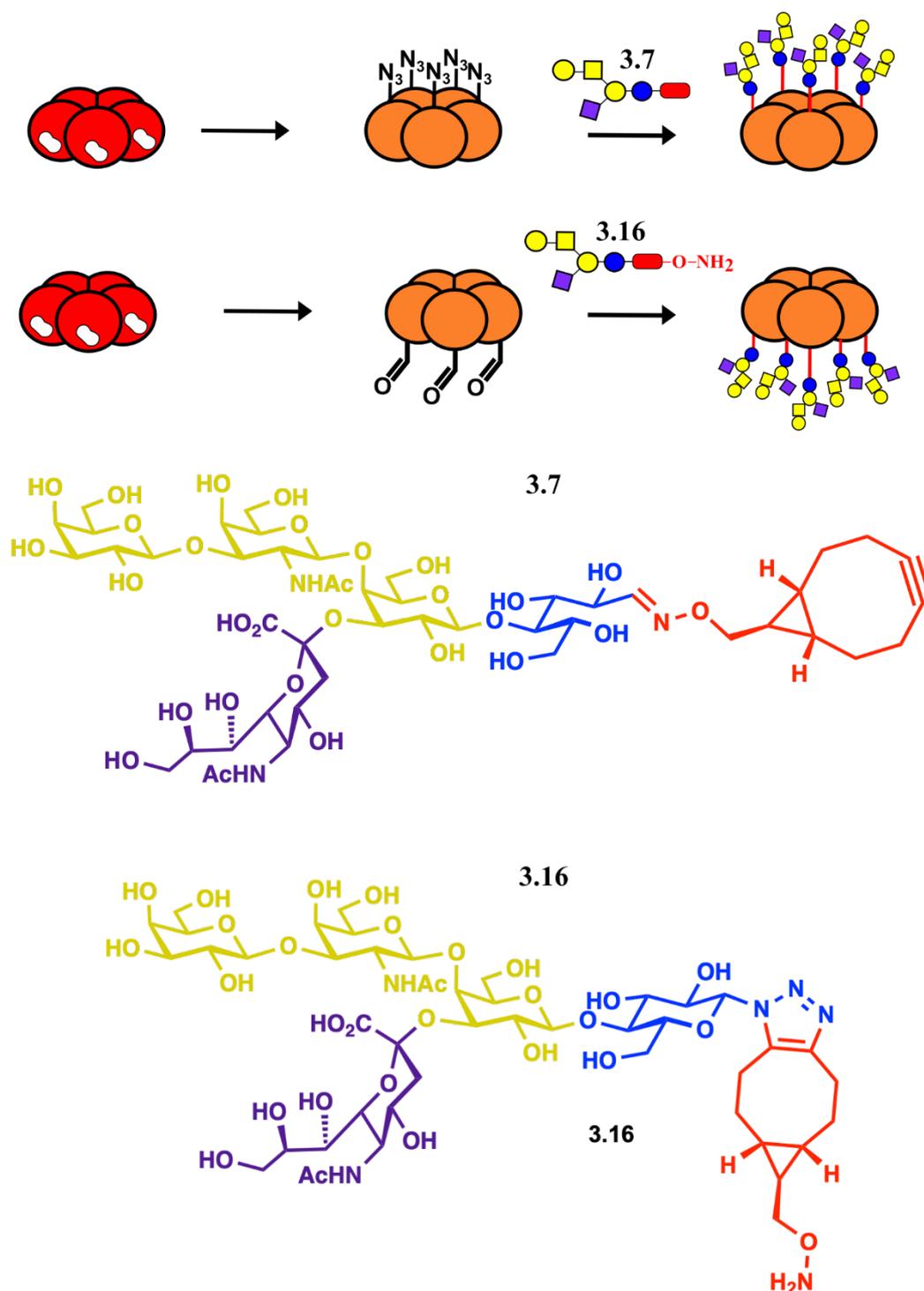


Figure 4.1: Proposed methods for the synthesis of two GM1-based neoglycoprotein inhibitors with a minimal-linker length GM1 derivatives; by top face functionalisation of a non-binding azido-CTB mutant with the 3.3, and bottom face N-terminal face functionalisation of a non-binding CTB mutant with yet to be synthesised 3.15. These reactions will demonstrate two methods of bioorthogonally and site-specifically glycosylating the CTB pentamer, using glycan derivatives from the same heterobifunctional linker.

This chapter will describe the design strategy for the mutant proteins and recombinant expression, before presenting studies for the conjugation of glycan derivatives and finally results of inhibitor binding assays against the adhesion of WT CTB-subunit.

4.1 Design and Expression of bioorthogonally reactive CTB mutants

4.1.1 Non-binding CTB mutant for oxime ligations

The binding of CTB to GM1 can be removed by mutation of the tryptophan residue 88 (shown in blue, figure 4.2) to a glutamate, which abolishes GM1 binding as reported by Jobling and Holmes.¹³ The CTB subunit naturally has an threonine at the N-terminus (highlighted in green, figure 4.2), which can be oxidised to give an aldehyde using sodium periodate (Malaprade reaction). It was shown previously¹²⁸ that the oxime ligation to the oxidised N-terminal threonine of CTB gave 100% conjugation to the glycoprotein in the presence of aniline, making it a highly efficient method of protein conjugation which is routinely used in the Turnbull group.

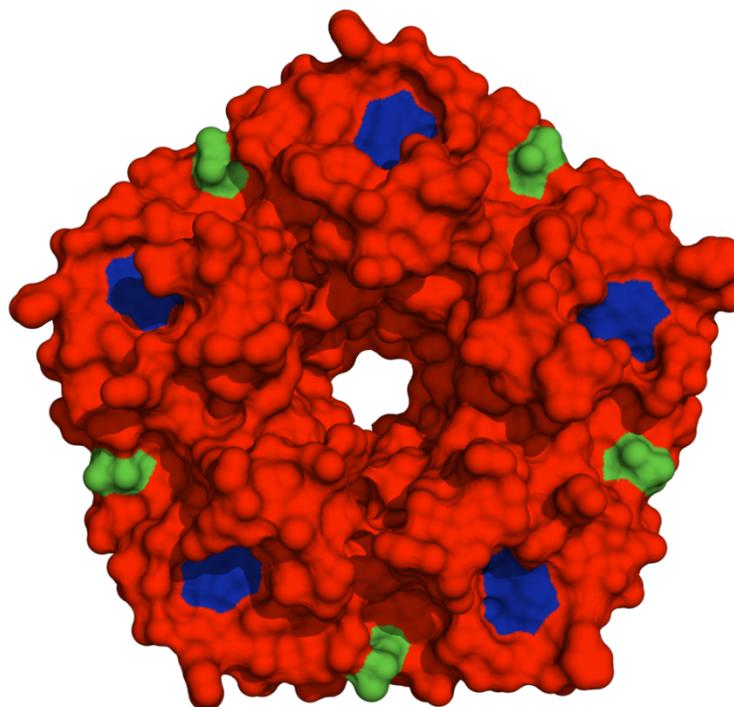


Figure 4.2: Surface representation of CTB pentameric subunit showing the GM1 binding face. Specific residues highlighted: N-terminal threonine residue shown in green; tryptophan 88 within the GM1 binding pocket, the site of mutation for removing GM1 binding, generated from co-ordinates 1CHB.pdb

4.1.2 Expression of the non-binding CTB W88E mutant

For the overexpression of **CTB W88E** a plasmid based on the pMAL-c5x expression vector containing the sequence for the non-binding mutant of CTB (created by Dr Tom Branson, University of Leeds, see appendix figure 10.1) was available, and the expression procedure is well defined within the group.¹²⁸ A glycerol stock of *E. coli* C41 expression cells (generated by Dr Tom Branson) harbouring the plasmid for **CTB W88E** was used to express the recombinant **CTB W88E** (procedure found in section 7.4.6) The gene

coding for CTB **W88E** is preceded by a periplasmic targeting sequence which is responsible for directing export of CTB to the periplasm. Following removal of the targeting sequence, the protein folds and assembles into the pentamer which is ultimately secreted from the cell into the media. The protein was harvested from the growth medium by ammonium sulfate precipitation. After solubilisation the protein was further purified by Ni-NTA affinity chromatography. Expression of the pentameric subunit was confirmed by SDS-PAGE analysis (figure 4.3); due to the highly stable nature of the pentamer, it does not denature in the SDS-loading buffer if the sample is not boiled prior to loading onto the gel, and can thus be observed in the pentameric state. Due to the large number of proteins precipitated from the growth medium, nickel affinity chromatography was insufficient to isolate pure CTB **W88E**, and required further purification by size exclusion chromatography (SEC). HRMS then corroborated the SDS-PAGE results, showing the correct mass of the monomeric CTB **W88E** at 11585 kDa (figure 4.3, theoretical mass = 11587.). The recombinant CTB **W88E** was expressed in a yield of 9.5 mg/L.

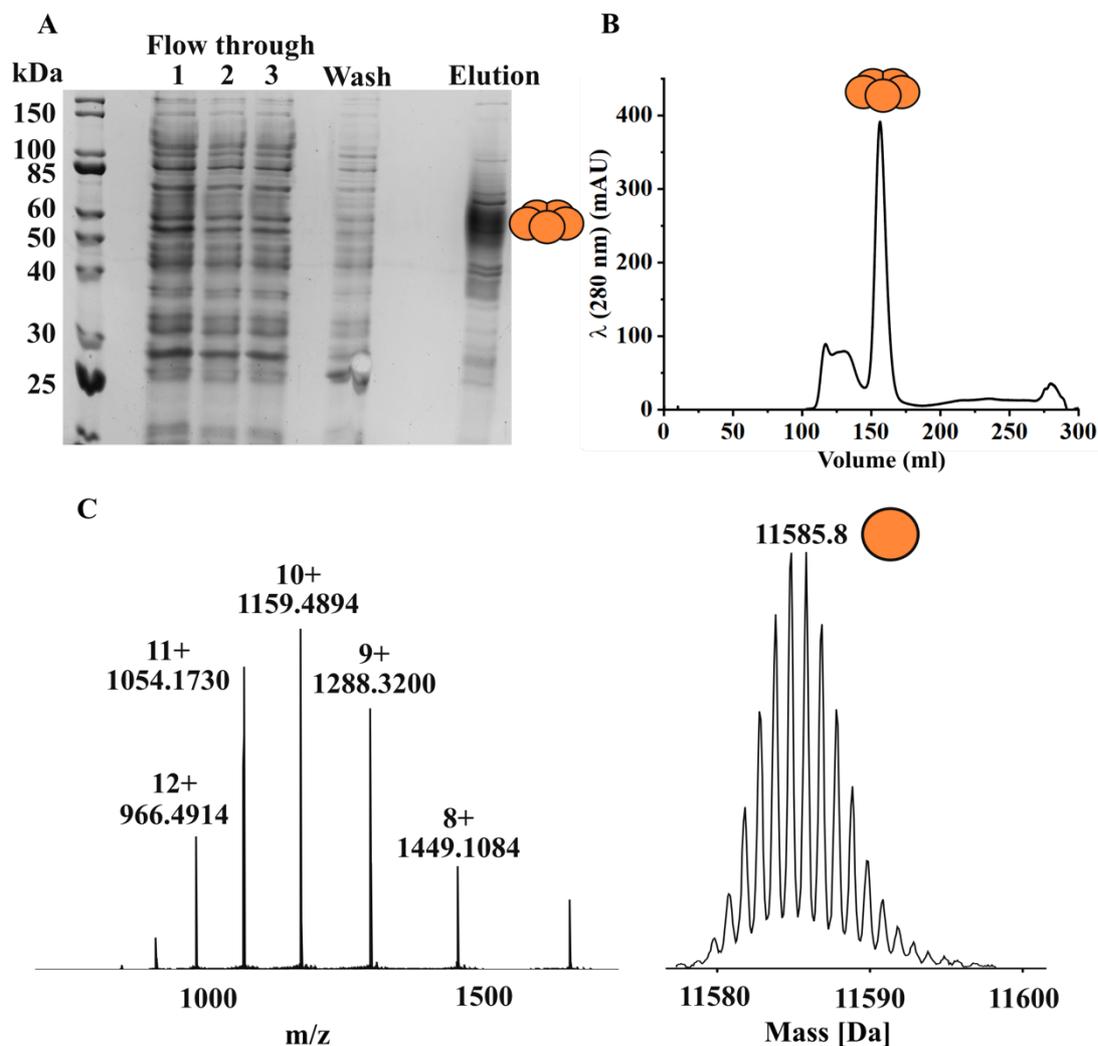


Figure 4.3: A) SDS-PAGE analysis of the Ni-NTA purification of W88E from the ammonium sulfate precipitation of the growth medium. W88E (orange pentamer) shown in the elution fraction. B) HRMS of W88E following secondary purification by size exclusion chromatography with both the charge state and deconvoluted spectrums.

4.1.3 Generation of an azido-CTB protein scaffold for use in SPAAC ligation

With the CTB W88E protein scaffold successfully expressed, attention turned to the generation of a non-binding CTB which contained a surface exposed azide for developing a new chemical glycosylation method using the SPAAC reaction.

4.1.3.1 Positioning of the azide handle

For SPAAC-based glycosylation, the azide group was to be placed on the non-binding face of CTB, to investigate the effect of the glycosylation site on the inhibitory potential of resulting neoglycoproteins. The motive was that the non-binding face surface is flatter than the binding face which has loops protruding down below the N-terminus, as shown in figure 4.4 in which the crystal structure is viewed from an angle perpendicular to the binding face. Attaching the GM1 ligands on the top face may allow their better

presentation towards the toxin's binding sites. To exploit the multivalent binding of CTB and target the five binding sites, five reactive groups were required per pentamer; the sites must retain symmetry and be equally spaced, the importance of this was shown by the high-affinity binding and low IC₅₀ values of the neoglycoprotein inhibitors produced through glycosylation of the N-terminus.

With the CTB monomeric sequence only containing three methionine residues, by performing four mutations the sequence would have a single methionine at a specific site. Using global residue-specific incorporation (see section 1.4.3.1) could be used in a site-selective way, to replace the natural amino acid methionine with the bioisosteric NCAA azidohomoalanine which is recognised by the aminoacyl-tRNA synthetase. The tRNA synthetase is then able to attach azidohomoalanine to the native methionine tRNA in *E. coli*, resulting in the incorporation of azidohomoalanine into proteins produced by the *E. coli* which are grown in growth medium which is methionine deficient.

An azido-CTB protein has recently been described in the group, used for the coating of nanoparticles.²⁸² This CTB expression mutant was designed to position the azide group in place of lysine 43, by engineering a single methionine in the sequence at position 43. The expression vector named pSAB 2.3(azido) was created by Dr Daniel Williamson by sequential site-directed mutagenesis removing three native methionine residues by introducing the mutations M37L, M68L, M101L. A fourth mutation K43M introduced the single methionine at position 43, creating a construct for the expression of an azido-CTB mutant.²⁸² This construct was applicable for our purpose of producing neoglycoprotein inhibitors. Distances between lysine residues at position 43 (figure 4.4, highlighted in cyan), closely match the distances between the N-terminal threonine residues (figure 4.4, highlight in green). PyMol was used to compare the distances between: two terminal galactose bound in GM1 binding site (29 Å); two N-terminal threonine (31 Å); and two lysines (30 ± 3 Å depending on side chain conformation), introducing a glycosylation site at position 43 would therefore produce a neoglycoprotein in which the ligand spacing match the distance between binding sites whilst maintaining pentagonal symmetry. The crystal structure shows these lysine residues are solvent exposed and the side chain lies perpendicular to the protein surface, which makes it an ideal location for a SPAAC reactive chemical handle.



Figure 4.4: Cartoon representation of the CTB pentameric subunit, from a side viewpoint. Specific residues highlighted: tryptophan 88 in blue; N-terminal threonine in green; and Lysine 43 on the non-binding face, the site identified for the introduction of azide, shown in Cyan. Image was generated from the co-ordinates 1CHB.pdb.

4.1.3.2 Creating and expression testing a construct for azido-CTB-W88E

The azido-CTB construct (created by Dr Daniel Williamson) was not completely fit for purpose for this project, as it still retained the native GM1 binding capabilities making it unsuitable as a protein scaffold. Taking the plasmid pSAB 2.3(azido), another mutagenesis reaction was performed using primers designed to introduce the W88E mutation (see appendix, section 10.2) to remove the GM1 binding capability, producing an expression vector pSAB 2.4 coding for **N₃-W88E**. Successful mutagenesis was confirmed DNA electrophoreses (see appendix, figure 10.3) with the mutated construct not digested by the digestion enzyme Dpn1. *E. coli* BL21 (DE3) were transformed with the pSAB2.4 plasmid. To confirm CTB over-expression could be achieved using the mutated plasmid, the CTB-M37L/M68L/M101L/K43M/W88E mutant was expressed using a methionine rich growth medium, to first determine if the protein needed to be isolated from the growth medium or from cell lysis.

The *E. coli* clone harbouring plasmid pSAB 2.4 was grown in Lysogeny Broth (LB) and over-expression was induced by addition of IPTG. After incubation at 37 °C overnight, the protein was successfully isolated from the soluble fraction upon lysis of cells. This was a peculiar discovery for the **MET/N₃ W88E** mutants; previous attempts to express other CTB mutants' proteins including **W88E** considering in *E. coli* BL21 cells had yielded no expression. Expression was only successful for other CTB mutants was observed in the C41 cell expression line, and the CTB subunit was exported into the growth medium. SDS-PAGE analysis (figure 4.5) confirmed the presence of pentameric

CTB in the elution fraction, the introduction of five mutations does not perturb the structure enough to cause dissociation during the running of SDS-PAGE. Melting temperature studies were not performed to confirm if the stability of the N₃-W88E mutant differed from WT CTB, however, previous work carried out by Dr James Ross (University of Leeds) confirmed that the ability to maintain a pentameric structures correlates with melting temperatures of 70 °C. The protein was further purified by SEC and the correct protomeric mass of 11,535 Da was confirmed by HRMS (figure 4.5).

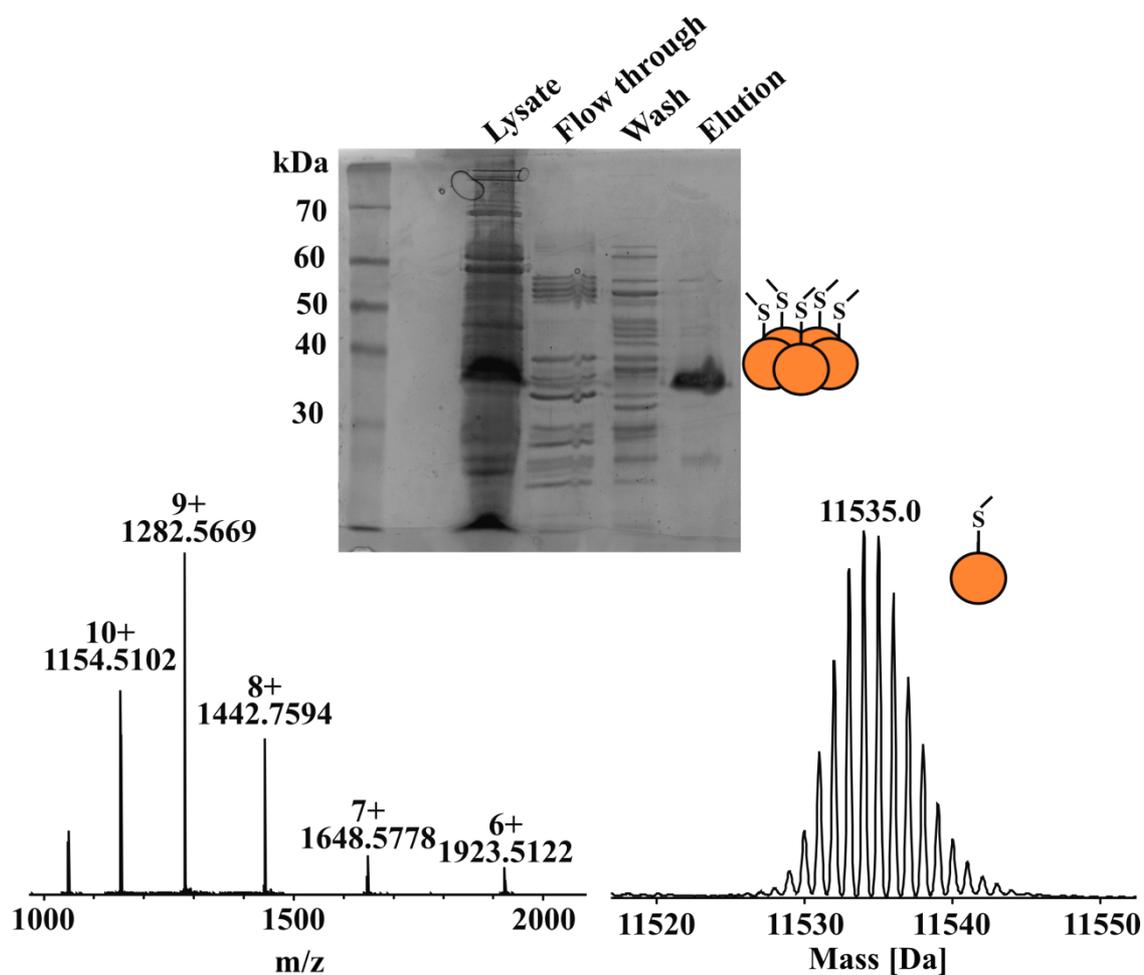


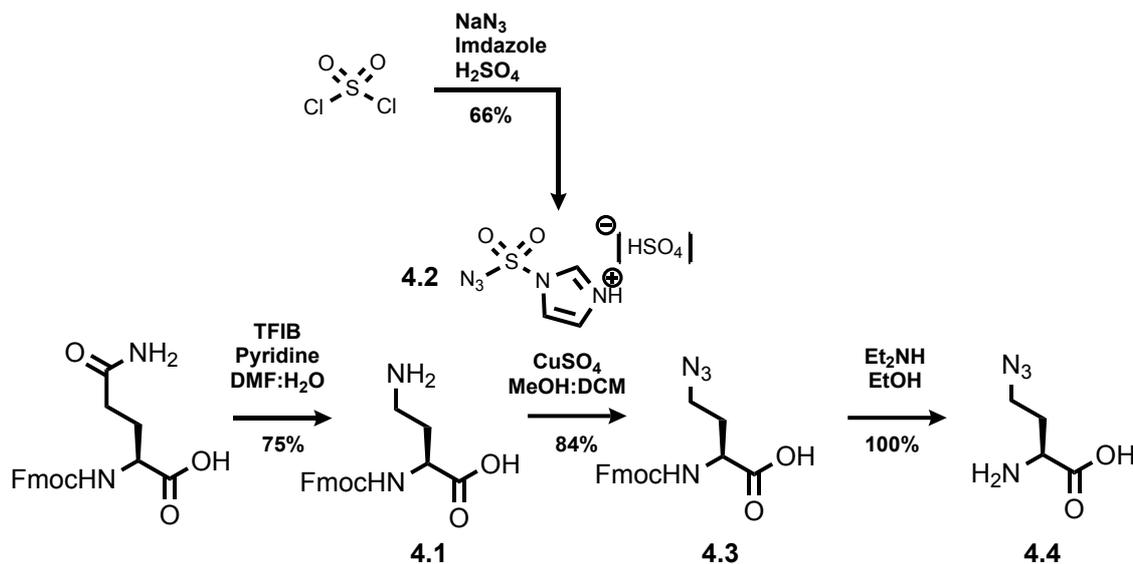
Figure 4.5: SDS-PAGE analysis from the expression test of the non-binding MET-CTB mutant in the lysate and successful purification by Ni-NTA affinity chromatography. HRMS of the elution fraction are shown including the charge state and deconvoluted spectra.

4.1.3.3 Expression of non-binding azido-CTB

The expression test confirmed the protein could be expressed in the *E. coli* BL21 cells, isolated from the periplasm and that the recombinant protein retains its correct pentameric structure and stability. The next step was to use the pSAB2.4 plasmid to express the CTB mutant incorporating azidohomoalanine.

Azidohomoalanine was synthesised from L-glutamine following a procedure reported by Lau *et al.*²⁸³ which reported the synthesis of Fmoc-protected L-azidohomoalanine in two

steps. The synthesis proceeds by Hoffmann rearrangement of Fmoc-glutamine to give **4.1**, a diazotransfer transfer then gives the azide (**4.2**). The original reports used the HCl salt of sulfonyl azide diazotransfer reagent **4.2**, however, the H₂SO₄ salt is known to be more stable and therefore safer.²⁸⁴ The Fmoc protecting group of **4.3** (synthesised by Dr Daniel Williamson) was then removed under basic conditions using diethylamine, to yield azidohomoalanine.



Scheme 4.1: Synthesis of azidohomoalanine from Fmoc-glutamine, by Hoffman rearrangement and diazotransfer with reagent **4.2**.

Methionine auxotrophic *E. coli* B834 (DE3) was transformed with the pSAB2.4 plasmid. Cells were then grown in “new minimal media” (NMM) following procedures outlined by Wiltschi and co-workers.²⁸⁵ Cells were supplemented with enough methionine to grow to ‘lag phase’, at which point over-expression was induced by addition of IPTG alongside supplementation of the media with azidohomoalanine. Upon cell lysis with BugBuster® the protein was purified by Ni-NTA affinity chromatography. The results of SDS-PAGE are shown in figure 4.6. Affinity chromatography proved sufficient to purify the N₃-**W88E**, and so SEC was not required. Once again pentameric protein was observed at an apparent molecular weight of 35 kDa, confirmed to be the pentameric CTB by the dissociation into the monomeric CTB upon boiling the sample.

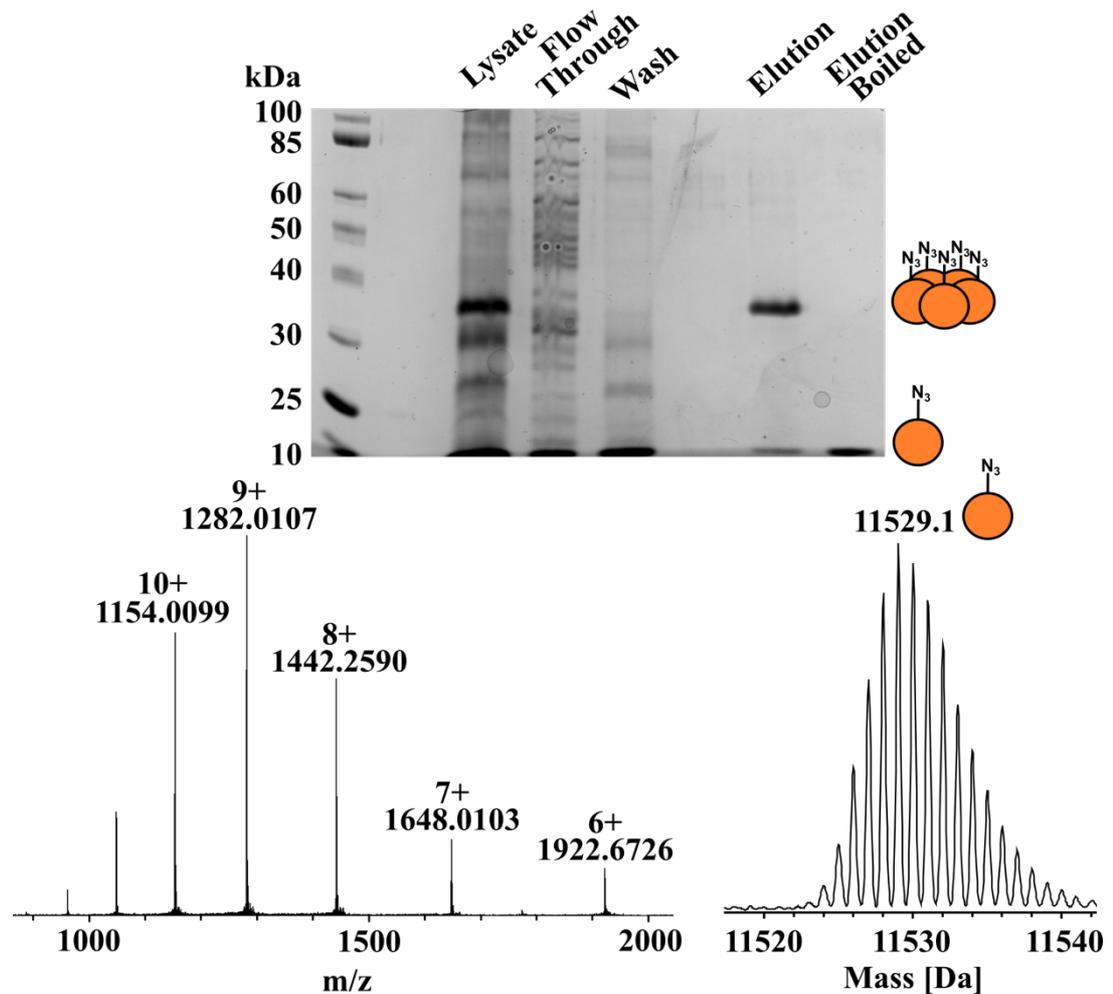


Figure 4.6: SDS-PAGE analysis of the expression of N_3 -W88E. showing the cell lysate and fractions taking from Ni-NTA, with the pentameric N_3 -W88E in the elution fraction and the monomer after boiling. B) shows HRMS of the elution fraction from Ni-NTA column with the charge state and the deconvoluted spectrums.

4.1.4 SDS-PAGE comparison of CTB mutants

As shown in the expression of WT CTB and each of the mutants, the pentamer is highly stable and not susceptible to denaturation by SDS at room temperature and can therefore be visualised by SDS-PAGE. Dissociation of the pentamer requires heating at 95 °C, before the monomeric CTB can be observed by SDS-PAGE. In each case, the CTB pentamer strangely migrates faster in SDS-PAGE compared to what would be expected for an extended chain conformation of a protein with the same mass as the CTB pentamer (58 kDa). Figure 4.7 shows an SDS-PAGE gel comparing WT CTB to all three CTB mutants expressed, highlighting the differences in observed mass as an effect of mutagenesis. The WT CTB subunit is observed approximately 10 kDa from its expected size, and that the single mutant W88E is similar to the WT. The compact folded shape of CTB, results in a lower surface exposure to SDS compared to an unfolded protein; and the pentamer also has a smaller hydrodynamic radius for its molecular weight. This results

in the pentamer structures migrating through the porous gel more rapidly and appearing at a smaller molecular mass than would be expected when compared to protein standards. The pentamers of **MET-W88E** and **N₃-W88E**, however, appear to have an apparent molecular weight of 37 kDa, a 20 kDa difference from the expected molecular weight. The removal of three methionine residues within the proteins folded structure is unlikely to affect the shape, size or interactions with SDS. This exaggerated shift in the observed mass away from the known mass of the pentamer can be explained by mutation of the surface lysine residues resulted in a change surface charge on the protein, removing five positive charges from the protein surface allowing the pentamer to migrate more quickly through the electric field. In combination with the small size and compact shape of the pentamer, the protein is able to move through the gel more rapidly than the unfolded protein would.

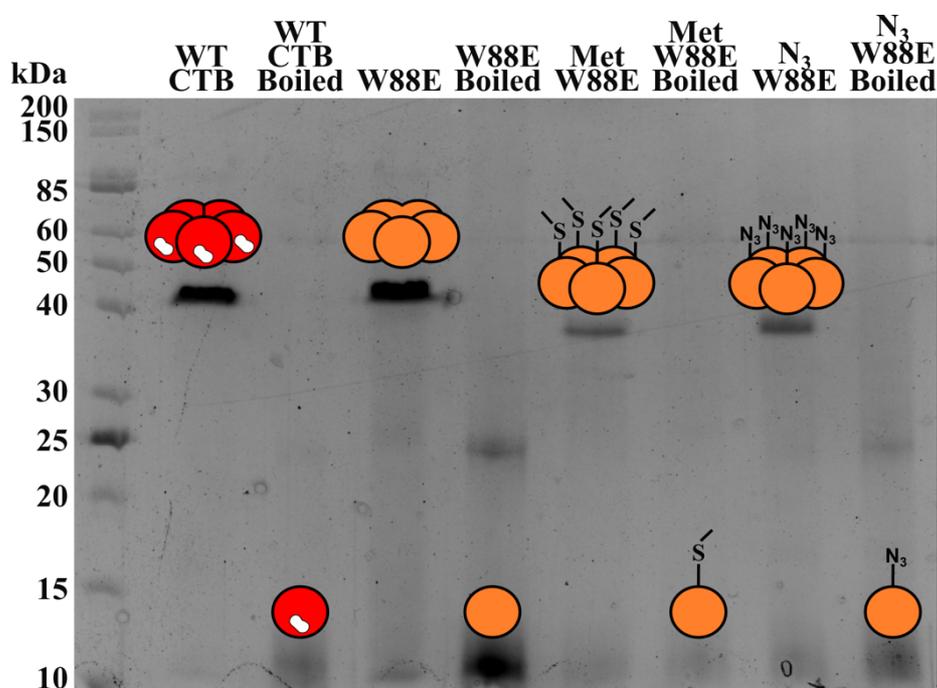


Figure 4.7: Comparison of WT CTB (red, pentamer) to the single mutant, W88E (orange, pentamer), and the 5-point mutations MET-W88E and N₃-W88E alongside the boiled samples to visualise both pentameric and monomeric B subunit by SDS-PAGE.

4.2 Site-specific glycosylation of CTB-W88E by oxime ligation

After successful expression of the CTB mutant protein scaffolds which each comprise of one half of the orthogonal reactive pair to the derivatised glycans **3.3**, **3.7** and **3.14** produced in chapter 3. Firstly, the oxime ligation to the **W88E** scaffold to confirm the same chemistry used to produce the previous neoglycoprotein inhibitor, could still be used for glycosylation reactions with oxyamine-derivatised lactose **3.14**.

4.2.1 N-terminal oxidation of W88E

The oxidation of CTB was first described by Offord and Rose.²⁸⁶ CTB was exposed to NaIO₄ in the presence of methionine, which was used as a preventative measure to avoid oxidation of methionine residues within the protein structure. We adapted this procedure as it was determined that the presence of methionine was not required, as no mass corresponding to the oxidation of any of the three methionine residues in the protein was ever observed when only NaIO₄ was added. The addition of methionine in the buffer appeared to slow the rate of oxidation, as reactions took around 20 minutes to reach completion; whereas with no methionine the reaction was complete before a sample could be analysed by mass spectrometry.

Figure 4.8 shows the HRMS of the oxidation reaction following the additions of NaIO₄ (5 eq, 2.5-3.5 mM), showing two resulting products. The desired **W88Eox** is not observed as the aldehyde in solution, but instead as the hydrate at 11,557 Da. A second product at the mass of 11,540 Da is also observed, which would appear to have the correct mass of the aldehyde; but instead corresponds to a cyclised by-product in which rotation around a proline residue at the N-terminus brings the amide into close proximity with the aldehyde, resulting in cyclisation by intramolecular attack. The formation of this product is minimised by performing the oxidation at high concentrations of **W88E** in region of 500-700 μM CTB (protomer concentration).

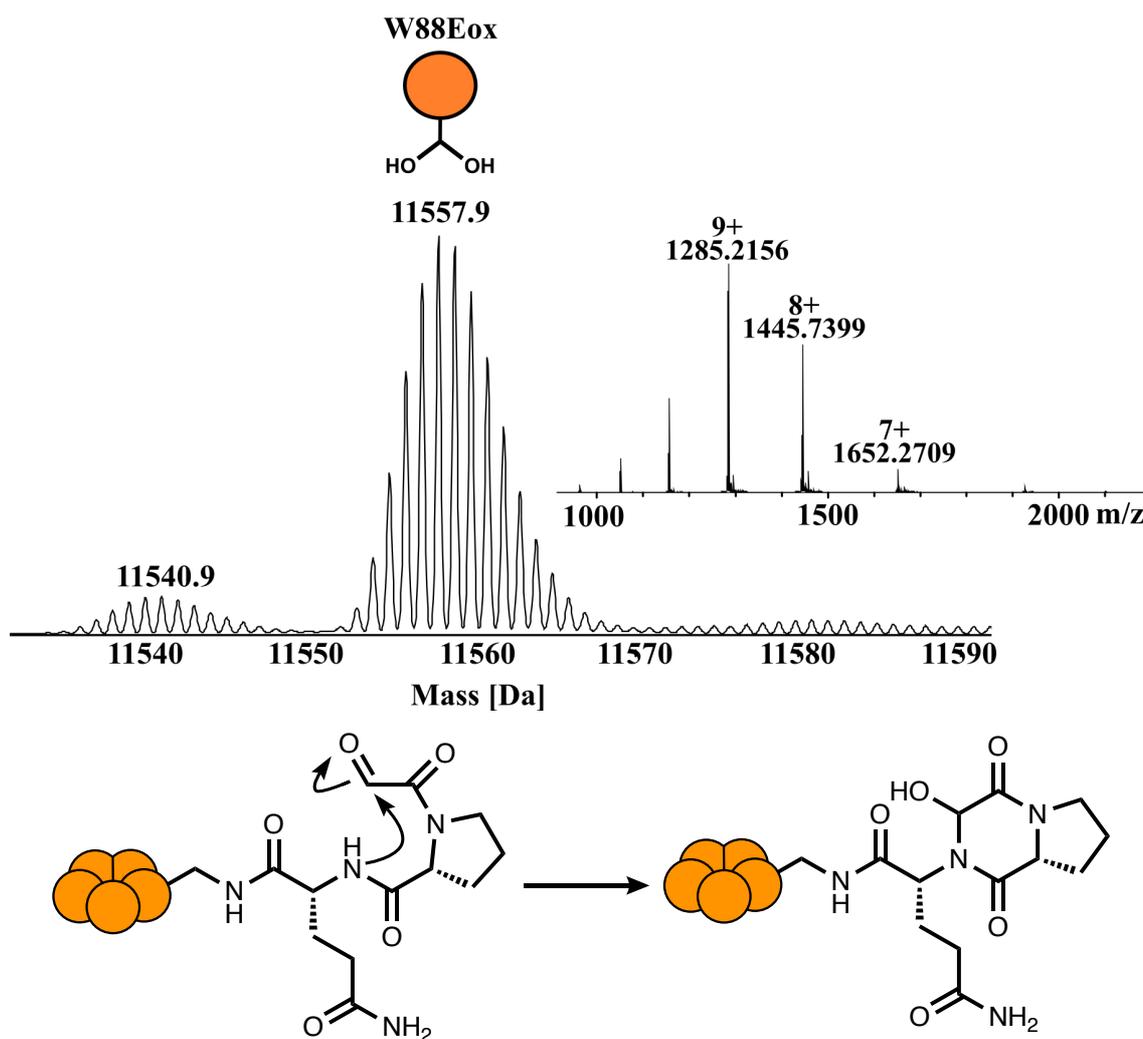


Figure 4.8: Charge state and deconvoluted spectra following oxidation W88E, and mechanism showing cyclisation of the CTB N-terminus following the oxidation of threonine which gives the mass of 11540 Da.

Curiously, on some occasions when performing these oxidation reactions, the conversion only reached 50%, whereas previous samples from a different expression batch had achieved full conversion to the oxidised product. After a process of trial and error in which the number of equivalents of methionine, periodate and the concentration of protein were tested without any effect on the total conversion, it was noticed that the only difference between the two reactions was the buffer. Reactions which reached full oxidation were performed in sodium phosphate buffer, whereas, only limited oxidation was observed when the protein was in PBS. When the protein in PBS was dialysed into phosphate buffer and then treated with sodium periodate, full oxidation was once again achieved. The presence of potassium in the buffer results in counterion exchange of sodium periodate to give potassium periodate which has poor aqueous solubility. This lack of solubility, therefore, limits the amount of available periodate to participate in the oxidation and oxidation of the protein is limited. This important observation has now been published in collaboration with the Fascione group at University of York.²⁸⁷

4.2.2 Synthesis of lactose neoglycoproteins by N-terminal oxime ligation to CTB W88E

With the oxidised protein in hand, the oxime ligation to glycoconjugate **3.14** was used as a test for the oxime ligation with BCN glycoconjugates. The oxime ligation was carried out using identical conditions used in the synthesis of the neoglycoprotein by Branson *et al.*¹²⁸ To **W88Eox** at 450 μ M (monomeric concentration) in phosphate buffer, glycoconjugate **3.14** was added in excess of 10 eq (4.5 mM), with 100 mM aniline required for the reaction to proceed at pH 6.8.

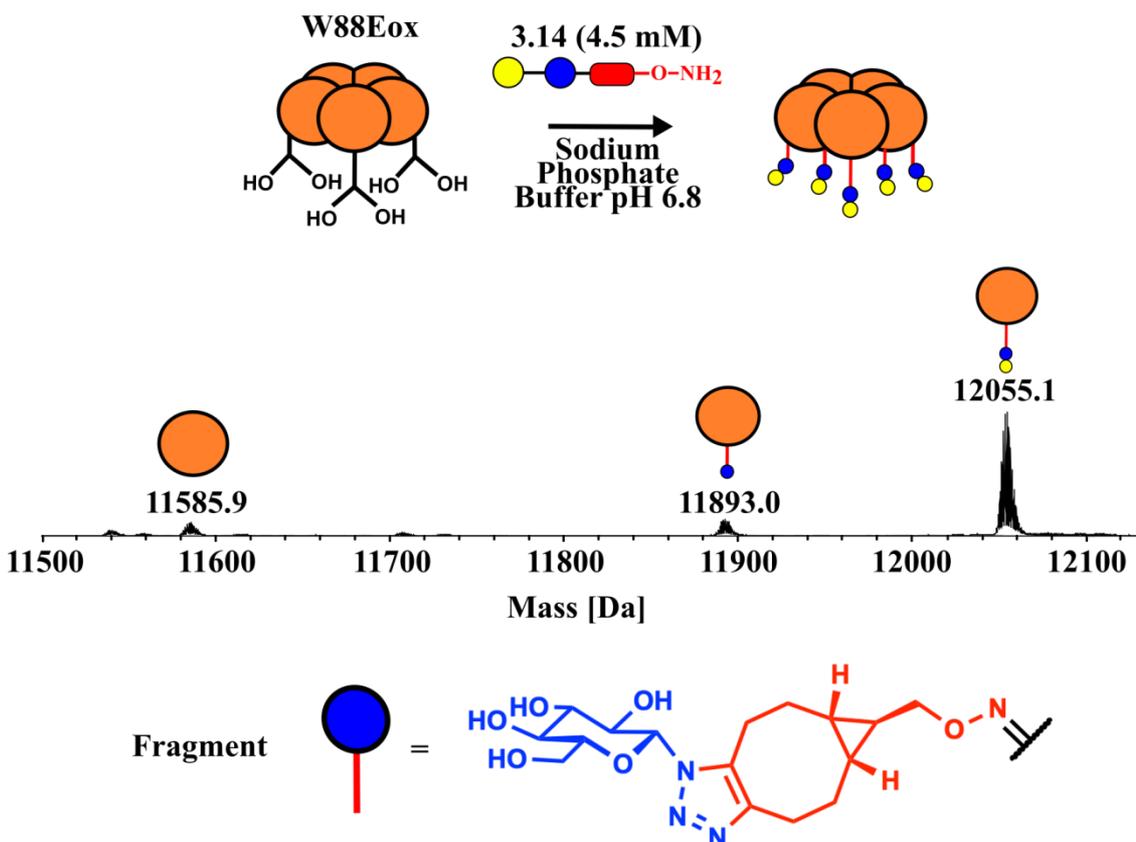


Figure 4.9: Deconvoluted mass spectrum showing the ligation of **3.14** (4.5 mM) to **W88Eox** (450 μ M).

The ligation was left agitating overnight at room temperature, and the reaction progression was examined by HRMS. Figure 4.9 shows the acquired spectrum after purification using a PD-10 desalting column. The major species detected is the glycoprotein product **W88E(3.14)**, with two other species relating to fragmentation of the glycoconjugate within the mass spectrometer determined to be the loss of the galactose giving 11,893 Da. The areas of the deconvoluted peaks in the mass spectrum indicated 100% ligation and no oxidised **W88Eox** was detected, however around 3% unoxidized CTB **W88E** was also present. The successful ligation to **3.14** showed that N-terminal ligation is still possible with this minimal-linker length between carbohydrate and reactive oxyamine.

4.3 SPAAC as a novel method of site-specific glycosylation

4.3.1 Optimisation of SPAAC glycosylation using lactosyl BCN

After successful expression of the N_3 -W88E, the SPAAC method for protein glycosylation could be investigated. To our knowledge synthetic glycosylation using the SPAAC reaction has not previously been reported, therefore, there is no previous knowledge of conditions to provide the best reaction conversion. BCN-Lac oxime **3.3** was the easiest to produce synthetically, and having been prepared on a larger scale, it was used for reaction optimisation. The glycosylation was first trialled at low concentrations of N_3 -W88E (50 μ M protomer concentration) with 1.25 eq of cyclooctyne **3.3**. The reactions were incubated at 37 °C overnight, after which time the deconvoluted mass spectrum (figure 4.10) showed a 1:1 ratio of N_3 -W88E and (**3.3**)W88E indicating only 50% glycosylation.

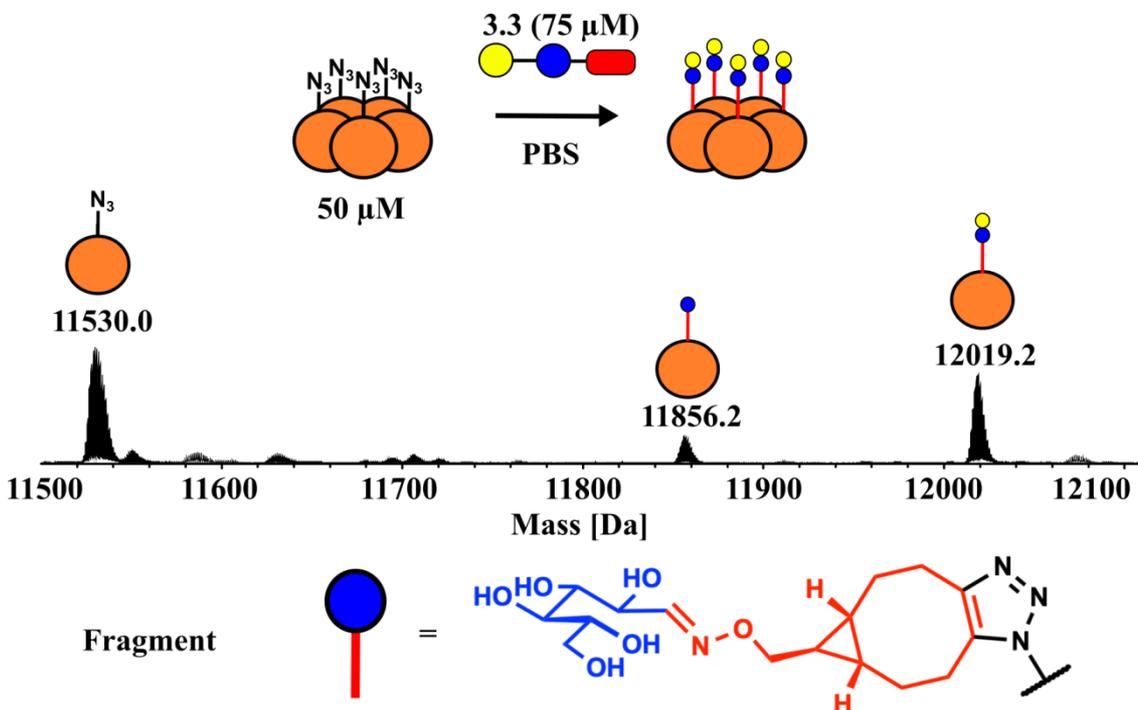


Figure 4.10: Deconvoluted HRMS of the SPAAC glycosylation with N_3 -W88E (50 μ M) and cyclooctyne **3.3** (75 μ M, 1.25 eq) at 37 °C overnight.

A number of conditions were investigated to improve overall percentage glycosylation. Reactions were performed using N_3 -W88E at concentrations of 50, 100 and 200 μ M, and at each concentration, 1.5, 3 and 5 eq of lactose derivative **3.3** was added. Each reaction was incubated at 37 °C for 16 hours before evaluation by HRMS. The resulting deconvoluted spectra presented in figure 4.11 all show that the glycoprotein was produced.

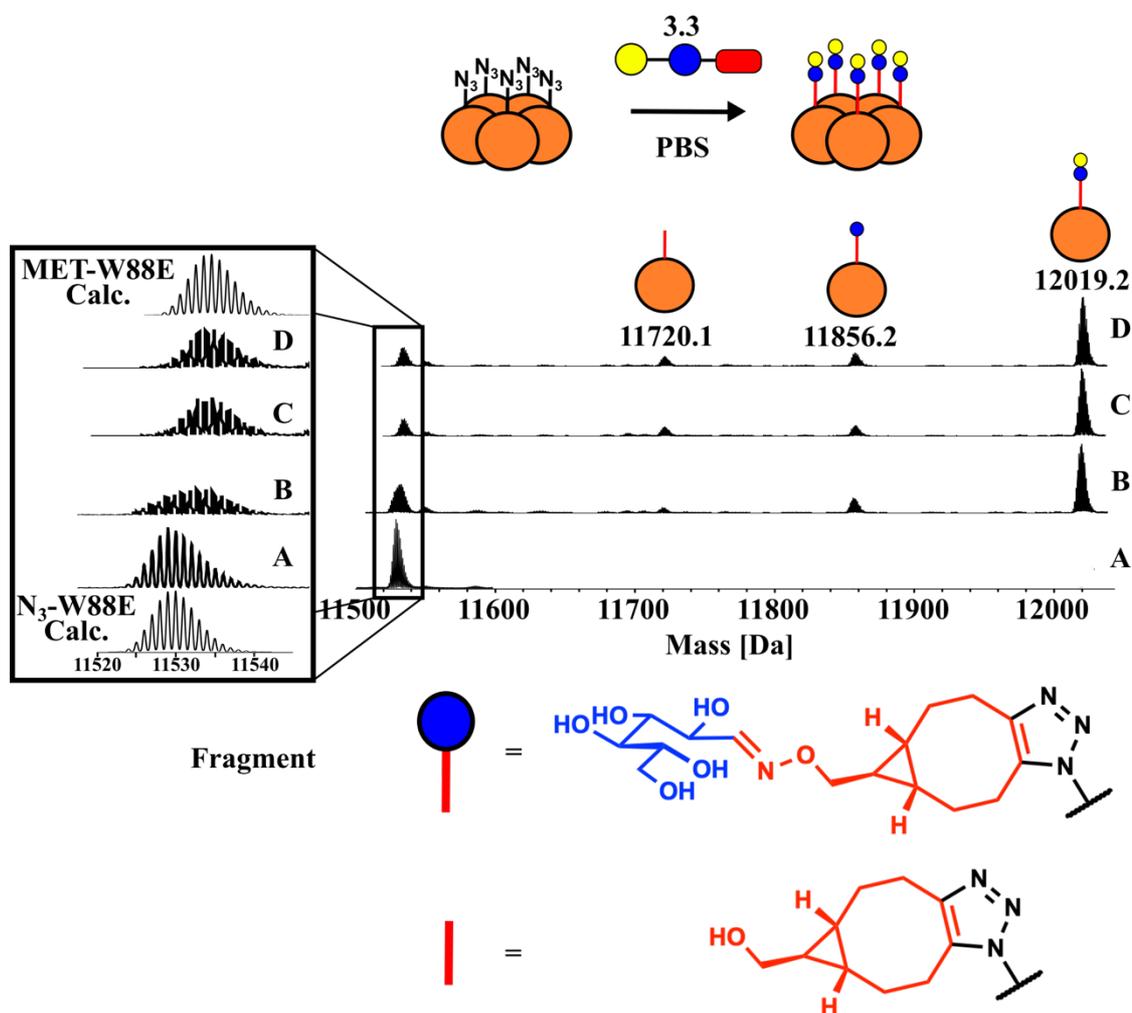


Figure 4.11: Stacked deconvoluted mass spectrums for the starting N_3 -W88E and results of SPAAC glycosylation reactions at $100 \mu\text{M}$ N_3 -W88E varying the excess of 3.3 at 37°C for 16 hours, and inset stacking reactions spectrums with the simulated isotope patterns for N_3 -W88E and MET-W88E. A) N_3 -W88E starting protein. B) $100 \mu\text{M}$ N_3 -W88E and 1.5 eq 3.3. C) $100 \mu\text{M}$ N_3 -W88E and 3 eq 3.3. D) $100 \mu\text{M}$ N_3 -W88E and 5 eq 3.3. Simulated deconvoluted mass spectrums ('calc') are also shown for N_3 -W88E (95:5 N_3 -W88E:MET-W88E).

At first glance the spectra do not appear to show full conversion to the glycoprotein, with a peak remaining mass region of azido starting material, first thought to be unreacted N_3 -W88E. However, close inspection of the overlaid spectra show a shift in the mass of the peak from 11,530 to 11,535 kDa (figure 4.11 inset), when these peaks were overlaid with the simulated isotope patterns for both N_3 -W88E and MET-W88E, it became clear that conditions shown in spectra C and D contained only MET-W88E concluding that all the available N_3 -W88E had been converted to the glycosylated product. The peak at 11,535 Da could not be seen in the starting N_3 -W88E as the less abundant MET-W88E sits underneath the isotopic distribution pattern of N_3 -W88E. The abundance of MET-W88E was calculated to be approximately 20% after glycosylation, meaning the incorporation of azidohomoalanine was only 80% successful. This was then confirmed by simulating the isotopic pattern for an 80:20 mixture of N_3 -W88E:MET-W88E, which then matched the isotopic distribution for expressed N_3 -W88E spectrum (figure 4.12)

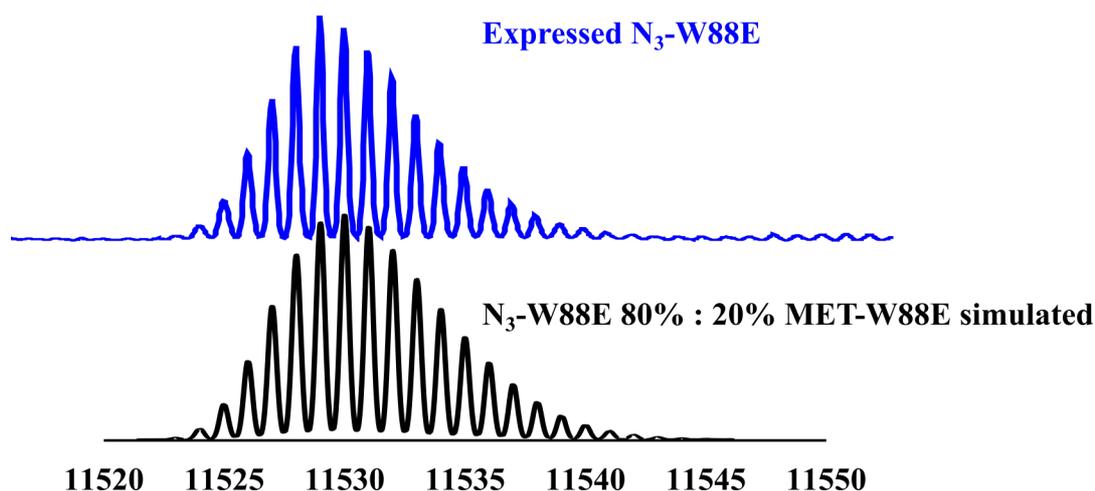


Figure 4.12: Comparison between the simulated isotopic distribution pattern for a sample containing 80:20 N_3 -W88E:MET-CTB and the deconvoluted mass spectrum for expressed N_3 -W88E.

The poor degree of methionine replacement prompted re-expression of N_3 -W88E. Taking caution to ensure no methionine was carried forward into the expression cultures, cells were harvested from the LB starter culture by centrifugation at $4500 \times g$ for 25 minutes. The cells were resuspended in a small volume of NMM media before inoculation of the larger cultures. Each culture was grown until cell growth reached ‘lag phase’, at which point expression was induced with IPTG. Following isolation and purification, the resulting N_3 -W88E protein was used in time-course reactions to follow the kinetics of the SPAAC glycosylation reactions.

Following the SPAAC bioconjugation by ESI-HRMS time-courses were performed by taking samples of the reaction at specific time points during the reaction. Following dilution of each sample, the reaction progress was halted and the ratio of products to starting material could be analysed. Taking snapshots of the reaction at specific points allowed accurate conclusion of reaction completion and could be used determine differences in the reaction kinetics under different reaction conditions. Each time point was analysed by comparing determining the abundance of neoglycoproteins products including fragmented species to the starting N_3 -W88E, to determine reaction progression. Deconvolution of the data was performed before the determining the species abundance, as the mass shift charge states in the ESI spectrum was more difficult to interpret than the larger mass shift observed for the single charge state products in the deconvoluted spectrums.

HRMS time courses of three different glycosylation conditions are displayed in figure 4.14. Figure 4.13A shows the reaction with $250 \mu\text{M}$ N_3 -W88E and 5 mM cyclooctyne 3.3. The glycosylated product appears within 30 minutes and increases in abundance at

each subsequent time point but is not complete after 6 hours showing some **N₃-W88E** remaining, and after 24 hours the low mass species shows a clear shift to being only **MET-W88E**. When the concentration of cyclooctyne **3.3** was doubled to 10 mM at the same protein concentration (figure 4.13B), all of the **N₃-W88E** depleted and completion was reached by the six hour time point but without time points between six and 24 hours a comment on the extent of the rate increase from 5 mM cannot be made. When the concentration of **N₃-W88E** was increased to 500 μM, with 10 mM **3.3** a seven-hour time course showed full SPAAC ligation by the four-hour time point (figure 4.13C). The ratio of glycosylated product to **MET-W88E** was calculated to be 95:5 from the relative abundance detected in the mass spectrometer. This represents 5% of the CTB monomers containing a methionine residue, thus azidohomoalanine was incorporated with a 95% efficiency.

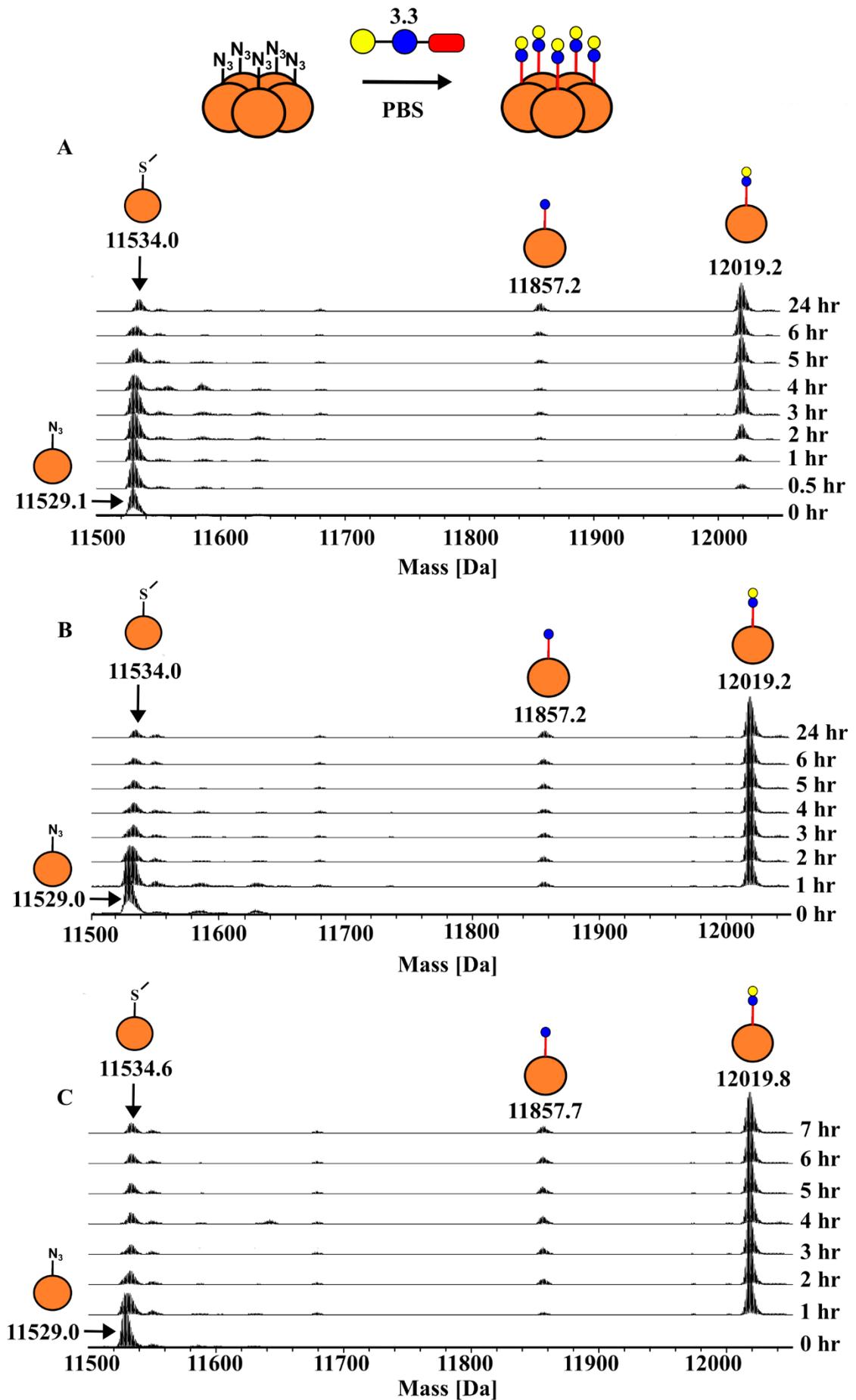


Figure 4.13: A) Cartoon representation of the SPAAC glycosylation reaction between N₃-W88E and neoglycoconjugate 3.3 B) Deconvoluted HRMS time course following the SPAAC glycosylation at 250 μM N₃-W88E, 5 mM 3.3 performed in PBS at 37 °C. C) Deconvoluted HRMS time course following the SPAAC glycosylation with 250 μM N₃-W88E and 10 mM 3.3 performed in PBS at 37 °C. D) Deconvoluted HRMS time course following the SPAAC glycosylation with 500 μM N₃-W88E and 10 mM 3.3 performed in PBS at 37 °C

Further analysis of reaction product by SDS-PAGE (figure 4.14) shows no N₃-W88E pentamer and a shift to higher mass for the glycosylated (3.3)W88E pentamer. Upon boiling the sample, the glycosylated pentamer dissociates to give the monomeric glycoprotein as a single band.

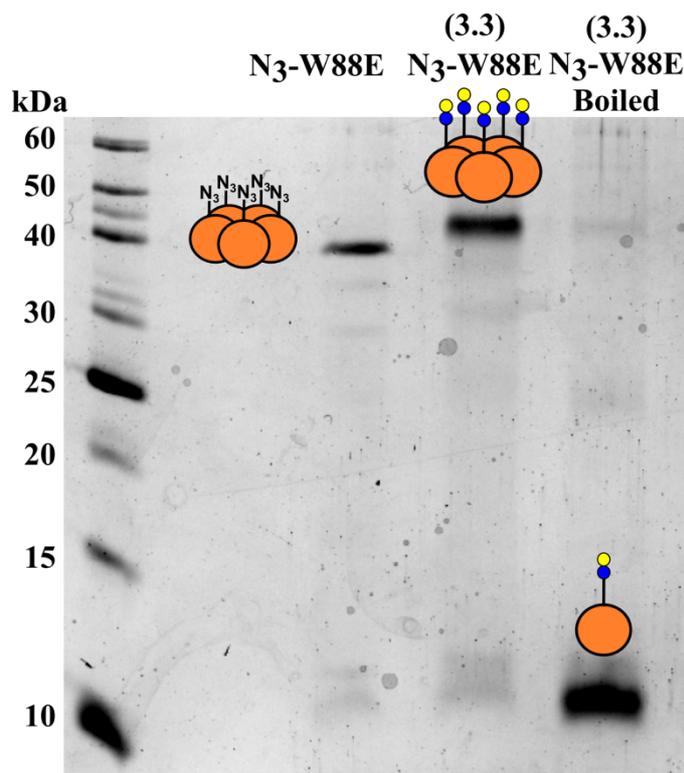


Figure 4.14: SDS-PAGE comparison of the N₃-W88E scaffold before and after the SPAAC glycosylation to glycoconjugate 3.3. A sample of unboiled and boiled glycoprotein was run to confirm that glycosylation did not change the pentameric structure or affect the thermal decomposition of the pentamer.

4.4 A novel neoglycoprotein inhibitor of CTB created by SPAAC glycosylation

With conditions optimised for fast and quantitative SPAAC glycosylation of the N₃-W88E scaffold to give the lactose neoglycoprotein (3.3)W88E, synthesis of an analogous GM1 neoglycoprotein was performed by SPAAC glycosylation using cyclooctyne 3.7. The glycoconjugate 3.7 was not available in the same quantities as 3.3, therefore, could not be used at the same concentration of 10 mM which had given such fast SPAAC ligation. The previous optimisation experiments had determined that such high

concentrations were only beneficial for the reaction kinetics, and that full conjugation could still be achieved at 5 mM cyclooctyne.

Figure 4.15 shows the result of the reaction between N_3 -**W88E** (500 μ M) and GM1-BCN **3.7** (5 mM) performed in PBS at 37 °C. After incubation for five hours, HRMS analysis determined the reaction to be complete and the product was purified by PD-10 desalting column. The spectrum shows formation of the glycoprotein (**3.7**)**W88E** at 12,675 Da and, as before, a small amount of the **MET-W88E** at 11,536 Da, in less than 5% abundance. Small peaks corresponding to fragmentation products were observed but were not assigned, due to the complexity of the different fragmentation patterns for the pentasaccharide. Surprisingly, the abundance of fragment products observed for the GM1 neoglycoprotein was much less than that for the lactose neoglycoprotein, despite the presence of a sialic acid which would be expected to be more labile. This result suggests a greater stability of the glycan could be attributed to the stability of the pentasaccharide on the protein but would require more in-depth mass spectroscopy studies which we do not have the expertise to perform.

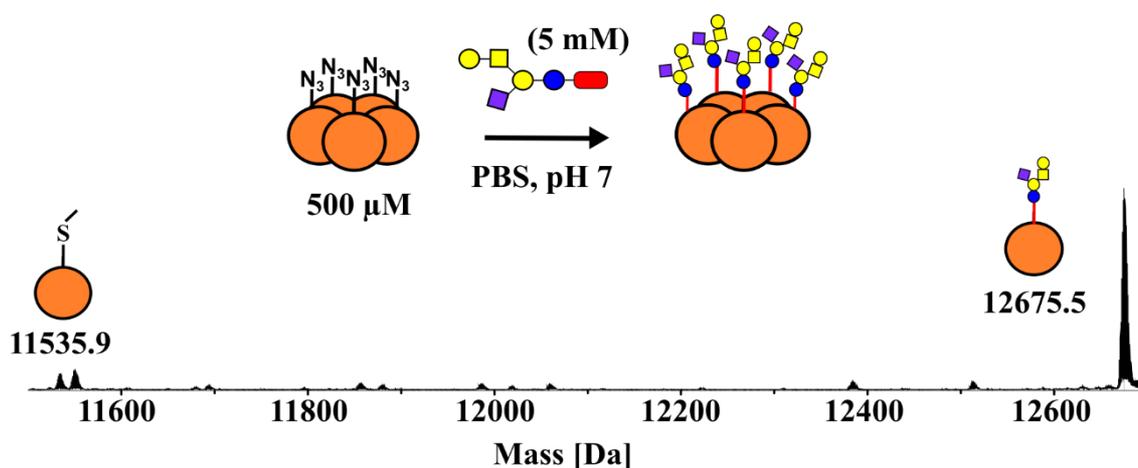


Figure 4.15: Deconvoluted HRMS following the purification of the product of SPAAC glycosylation between N_3 -**W88E** (500 μ M) and cyclooctyne **3.7** (5 mM).

The two neoglycoproteins were then compared on a single SDS-PAGE gel (figure 4.16). A shift in mass from N_3 -**W88E** was observed for both GM1- and lactosyl-neoglycoproteins consistent with addition of the carbohydrates. Both of the unboiled neoglycoproteins appeared above 40 kDa as expected for a pentamer, with the GM1 glycoprotein (**3.7**)**W88E** appearing more diffuse than the lactosyl glycoprotein (**3.3**)**W88E**. Interestingly, when the samples were boiled, the **GM1-W88E** monomer appears as two bands, one band with a similar mass to that of the (**3.3**)**W88E** which could correspond to partial hydrolysis of the pentasaccharide, which cleavage of the sialic would be most likely; and a band at a slightly higher mass which presumably would

correspond to the GM1 monomeric glycoprotein. The samples were subject to heating at 95 °C for the same length of time (eight minutes), the gel shows some (3.7)W88E pentamer remains, whereas, glycoprotein (3.3)W88E seems to fully dissociate. This could be a result of poor thermal contact in the heating block, or be a result of greater thermal stability, but this would need confirmation by performing melting temperature studies.

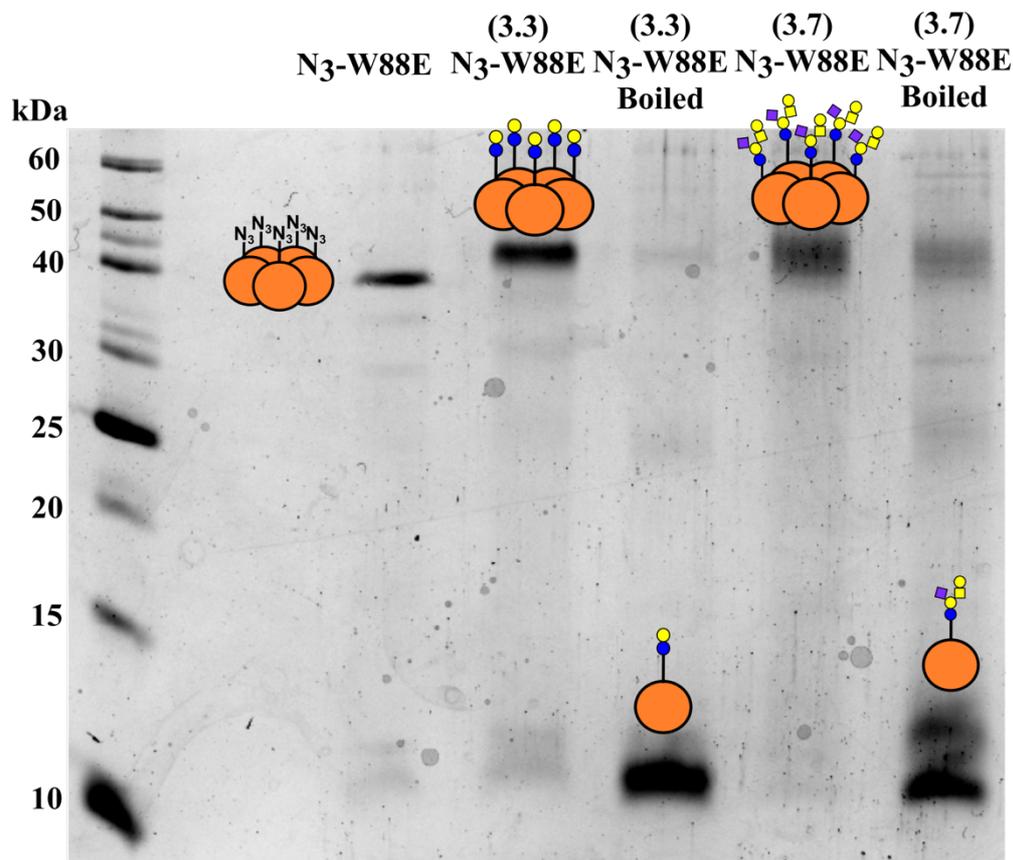


Figure 4.16: SDS-PAGE analysis comparing both (3.3)W88E and (3.7)W88E native and boiled samples with N₃-W88E.

Here, a novel method of site-specific glycosylation has been developed, to synthesise neoglycoproteins with simple and complex glycoconjugates. This is a highly efficient and rapid method of synthetic glycosylation, producing homogeneous glycoproteins, with quantitative glycosylation of the azido protein. The use of strained alkynes and the SPAAC reaction has not previously been used for synthesis of neoglycoproteins, has now been demonstrated in the development two neoglycoproteins based on an azido mutant of the CTB protein. This method of creating neoglycoproteins in conjugation with the generation of the glycan derivatives from any reducing sugar using linkers 1.2 and 1.3, would be an efficient method for the generation of homogeneous glycoproteins, provided that an azide group can be introduced into the desired protein.

4.5 Neoglycoprotein inhibition studies by enzyme-linked lectin assay (ELLA)

Having prepared the desired neoglycoprotein with GM1 attached on the non-binding face of the CTB-W88E mutant protein, it was tested as an inhibitor of WT CTB binding to GM1 coated microtiter plates using an enzyme-linked lectin assay (ELLA; figure 4.17). This assay detects the amount of CTB adsorbed to a GM1 microtiter plate, by creating a protein conjugate with horseradish peroxidase (HRP) enzyme; the amount of CTB-HRP bound can be determined by highly sensitive fluorescence measurement of the peroxidase activity, using the substrate Amplex Red[®] which is converted to resorufin in the presence of H₂O₂.

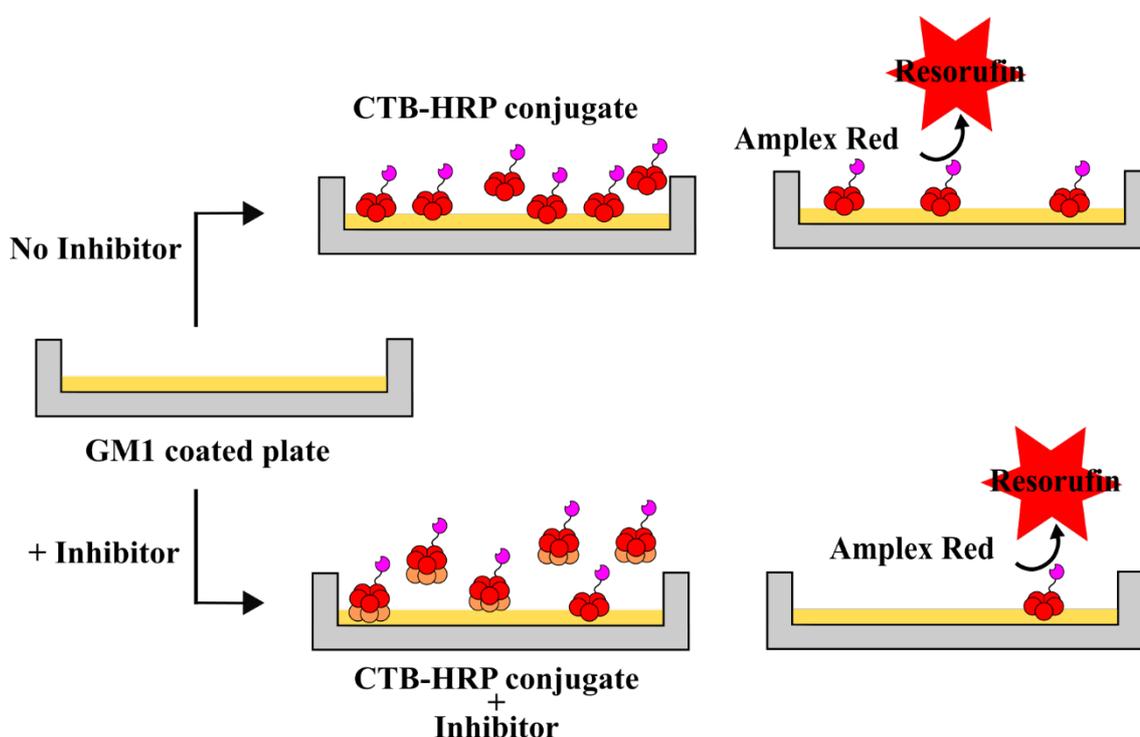


Figure 4.17: Cartoon representation of the ELLA assay for determined inhibition of CTB binding to GM1 coated plates.

For this assay, a horseradish peroxidase (HRP)-CTB protein conjugate was produced using Lightning-Link[®] horseradish peroxidase kit purchased from Expedeon, mixing WT CTB with the supplied activator solution with the freeze dried HRP. The HRP conjugation reaction was left for three hours before the addition of the supplied quenching solution. An assay to determine the optimal concentration of CTB-HRP required was then performed, carrying out a serial dilution starting at 500 ng/ml CTB-HRP (figure 4.18). Concentrations above 10 ng/ml reach a plateau saturating the detector limit in just a few data points. The previous ELLA assays were performed at 2.5 ng/ml for CTB-HRP purchased commercially from Sigma, however, this assay showed the CTB-HRP

conjugate produced using the lightning kit could be used at concentrations of 1-2 ng/ml and produce a measurable difference in rate from baseline fluorescence as calculated from the gradient of the linear trendline. Reducing the concentration of CTB-HRP allows the sensitivity of the assay to be increased.

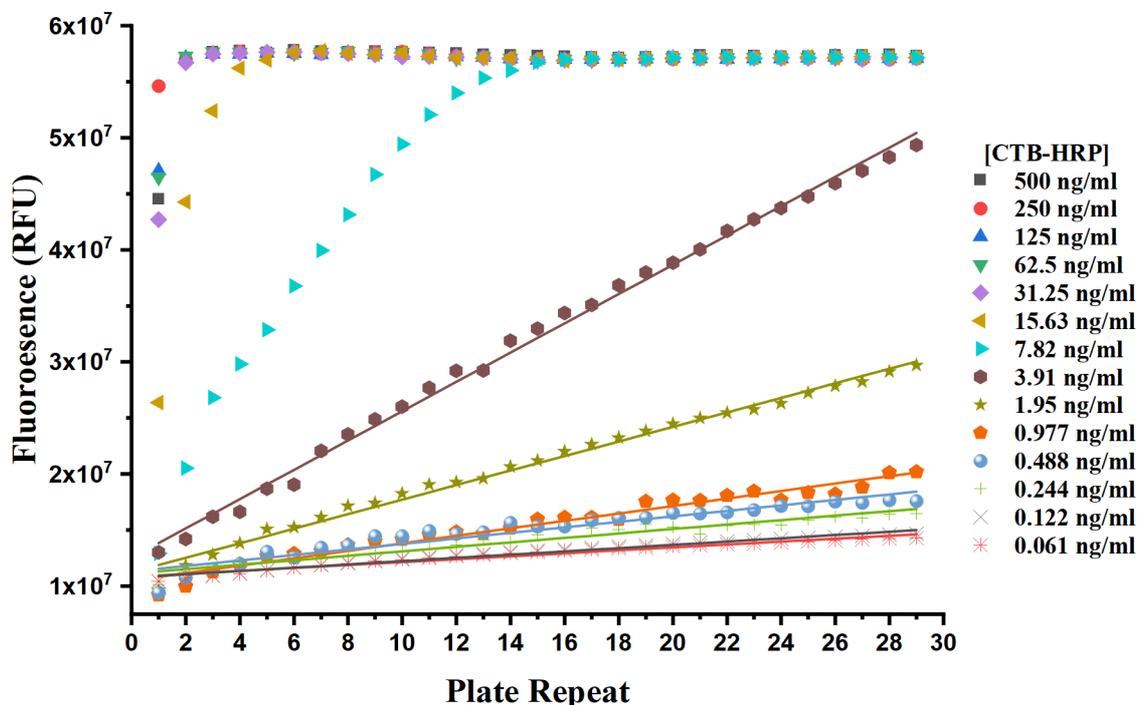


Figure 4.18: A graph comparing the results of the fluorescent measurement for the concentration gradient assay of CTB-HRP following a 2-fold serial dilution from 500 ng/ml and the addition of Amplex Red/H₂O₂.

Once the CTB-HRP conjugate was confirmed to bind to the GM1-coated microtitre plates it was then used to study the inhibition of CTB with the three neoglycoprotein inhibitors (figure 4.19). Monovalent lactose is known to be a weak inhibitor of CTB and when tested using the ELLA showed no inhibition even at 500 mM (data not shown). The aim was to investigate if the same increase in inhibition observed by attachment of GM1 to the CTB-subunit, could be achieved for lactose. Any inhibition achieved would further highlight the effect of having the pentavalent scaffold. One the research questions to be addressed by performing these inhibition assays was to study the effect of linker length between oligosaccharide and protein scaffold, and attachment point of the oligosaccharides on inhibitory potential.

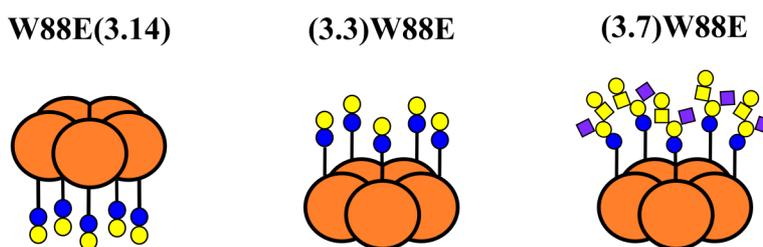


Figure 4.19: Cartoon structures of two lactose decorated and one GM1 decorated neoglycoprotein inhibitors to be tested for inhibition of CTB-subunit using the established ELLA.

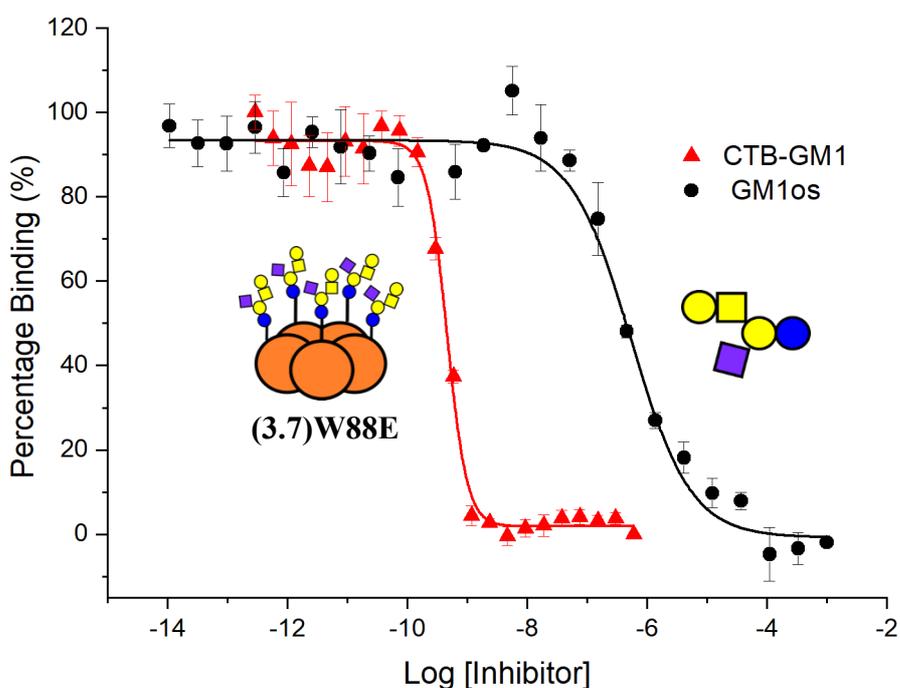
To prevent any non-specific binding of CTB-HRP and neoglycoprotein inhibitor to the wells which could produce false positive results, the wells were blocked with BSA, prior to addition of any CTB:HRP/inhibitor to the wells. The assay included positive control well (containing CTB:HRP and no inhibitor), and negative control wells (containing no CTB:HRP or inhibitor) and used to maximum and minimum fluorescent values for each assay plate. To keep the assay design as close to the previous assay, CTB-HRP (2.5 ng/ml) and was premixed with inhibitor **(3.7)W88E** before being added to a GM1-ganglioside coated plate. After washing the plate, the IC₅₀ of the inhibitor was determined by detecting how much of the CTB-HRP was bound. Each inhibition was performed in triplicate and the average across each repeated used in analysis; in some cases, anomalous data points were removed leaving duplicate results, to avoid misrepresentation in inhibition values and errors. Initial rates were calculated in cases where the data plots deviated from linearity and data processing was performed as described in section 8.9.3.

Neoglycoprotein inhibitor **(3.7)W88E** was found to be a potent inhibitor of CTB adhesion with an IC₅₀ of 457 pM (figure 4.19). In contrast, both lactose and ligated neoglycoproteins **(3.3)W88E** showed no inhibition of WT CTB (see appendix, figure 10.7 and 10.8), therefore, inhibitor **W88E(3.14)** was not tested. When compared to the monovalent ligand GM1os measured by Branson *et al.*¹²⁸ in an assay of the same design, the neoglycoprotein inhibitor **(3.7)W88E** was 231-fold greater in potency per GM1os than the monovalent ligand. The **N₃-W88E** scaffold was not tested, as the mutation is known to abolish GM1 binding and the **W88E** CTB mutant had already been shown not bind to GM1 coated microtiter plates.¹²⁸

The IC₅₀ value of 457 pM is comparable to that reported for the analogous **W88E(GM1)** inhibitor (104 pM) reported by Branson *et al.*¹²⁸ with only 4-fold difference in potency. The two results could be more comparable if tested by the same person at the same time, and the discrepancy could be down to nuances when performing the assay. It is also important to note that the measured values are close to the measurable limit of the assay using the concentration of CTB-HRP of 2.5 ng/ml. With the assay shown in figure 4.19 showing the CTB-HRP concentration could be reduced to increase the assay sensitivity, testing the inhibitors with a more sensitive assay would confirm the differences, if any, in the inhibitory potential of the two neoglycoproteins with greater certainty. Conclusions drawn from this data show that reduction in linker length between scaffold and glycan, and moving the site of glycosylation gave little change in potency with as the

neoglycoprotein **(3.7)W88E** having an IC_{50} value in the picomolar range. However, from this result it cannot be determined the individual contributions of the glycosylation site and linker length. Hypothetically one could envisage a decrease in IC_{50} by **(3.7)W88E** being a result of numerous reasons: an enthalpy penalty upon bringing the two pentameric proteins together in such close proximity; increased electrostatic repulsions for the same reasons; or that the shorter, more rigid linker results in strained conformations for the glycan to reach its binding site.

With the synthesis of GM1 derivative **3.16**, attachment of **3.16** to the **W88E** mutant, and inhibition data for the resulting neoglycoprotein inhibitor, will produce a more detailed picture of the effects of linker length, and site of glycosylation. Only with this data can there be a rational attempt to state the what is the optimal site/position for glycosylation and linker length when designing future neoglycoprotein inhibitors towards bacterial toxins.



Inhibitor	Valency	Log (IC ₅₀)	IC ₅₀ (nM)	Relative potency (per GM1) ^c
W88E	-	- ^a	- ^a	- ^a
GM1os	1	-6.27±0.04	530	1 (1)
W88E(GM1)^b	5	-9.98±0.08	0.104	5096 (1019)
(3.7)W88E	5	-9.34±0.02	0.457	1160 (231)

Figure 4.20: Potency of neoglycoprotein inhibitors determined by ELLA for the inhibition of WT CTB to a GM1 coated 96-well plate by the neoglycoprotein GM1-W88E and GM1os. A table comparing the IC₅₀ values for GM1os, (3.7)W88E, W88E(GM1) their valences and relative potency per GM1. a-no inhibition detected for W88E scaffold by Branson *et al.*¹²⁸ b-W88E-GM1 inhibitor reported by Branson *et al.*¹²⁸ (GM1 ligated by N-terminal oxime ligation). c-Potency measured relative to monovalent GM1os.

A point raised by Branson *et al.*¹²⁸ argued that the use of a neoglycoprotein approach to toxin inhibitors was advantageous as the synthesis of neoglycoprotein **W88E(GM1)** was easier than the octavalent dendrimer reported by the Pieters group which had comparable potency with an IC₅₀ of 50 pM.¹¹³ This point also stands true for the neoglycoprotein inhibitor **(3.7)W88E**, and comparatively the synthesis of this neoglycoprotein inhibitor has been simplified further, as the development of oxime and SPAAC reactive glycoconjugates does not require the extensive chemoenzymatic synthesis which was performed to produce the ligand shown in figure 1.15, making this a more accessible and versatile approach to neoglycoprotein based inhibitors.

4.6 Conclusion

An azide containing, non-binding mutant of CTB with five-point mutations has been used successfully as a scaffold for making a neoglycoprotein-based inhibitor of CTB adhesion. Residue-specific incorporation in a methionine auxotrophic *E. coli* expression strain was used to replace methionine at position 43 in the CTB sequence with the bioisostere azidohomoalanine with ca. 95% efficiency. After purification of the recombinant protein, the azide group was used as a handle to glycosylate the mutant CTB using cyclooctyne derivatives **3.3** and **3.7**, demonstrating the first use of the SPAAC reaction as a method for synthetic glycosylation. SPAAC proved to be a very efficient and rapid method of covalently modifying proteins, at concentrations from 50 to 500 μM of the N₃- **W88E** and as little as three equivalents of the glycan derivative. The low propensity of CTB to aggregate and its high stability allowed reactions to be performed at high concentrations for fast conversion to the glycoprotein in a few hours. Quantitative modification was also

possible at much lower concentrations making this synthetic glycosylation reaction potentially applicable to a wide variety of proteins, which do not have the same stability as the CTB pentameric subunit.

The GM1-ligated neoglycoprotein **(3.7)W88E** exhibited a very low IC_{50} of 457 pM, showing a only a 4-fold difference in comparison with the previously reported **W88E(GM1)** which had a longer linker, and different site of glycosylation. The direct comparison between linker lengths and site of glycosylation is yet to be obtained requiring the synthesis of minimal length GM1 derivative **3.15** and its attachment to **W88E** by oxime ligation analogous to work of Branson *et al.*¹²⁸ in the absence of this data it can be concluded that this data further supports previous findings that the preordering and presentation of the GM1 ligands has a great enhancement on potency, with the minimal length linker still sufficient to present the GM1 groups to the binding site the resulting in multivalent binding, and inhibition was not perturbed by any penalties resulting bringing two pentamers in very close proximity.

Chapter 5 Multivalent neoglycoproteins inhibitors of the Verotoxin

Work in the preceding chapter, and the previous work by Branson *et al.*¹²⁸ provides experimental evidence in support of multivalent neoglycoproteins being effective agents to combat CTB adhesion. After developing a simplified and generalised approach to the synthesis of neoglycoproteins, this methodology could be applied to the synthesis of neoglycoprotein inhibitors of the clinically relevant verotoxin. This chapter will outline the synthesis of neoglycoproteins towards this therapeutic target which brings the complex challenge of targeting up to 15 potential ligand binding sites and a lower affinity interaction as described in section 1.2.2.2. Alongside this, the recombinant expression of the verotoxin and approaches taken to develop an ELLA inhibition assay by which to test the neoglycoprotein inhibitors will be discussed.

5.1 Generation of neoglycoprotein inhibitors against the Verotoxin

As a proof of concept of a neoglycoprotein-based inhibitor of the VTB-subunit, pentavalent neoglycoproteins based on the mutant CTB scaffold were proposed that could target any of the three observed binding sites from the VTB crystal structure. The approach to developing neoglycoproteins for VTB shown in figure 5.1 was identical to that described in the previous chapter, however, the scaffolds would be decorated with Gb3 ligands. Two synthetic glycosylation methods were established in chapter four (sections 4.2 and 4.3); expression of the relevant mutant CTB scaffolds were described in section 4.1; and the oxyamine and strained alkyne-containing Gb3 glycoconjugate partners to the mutant CTB scaffolds were described in section 3.4.

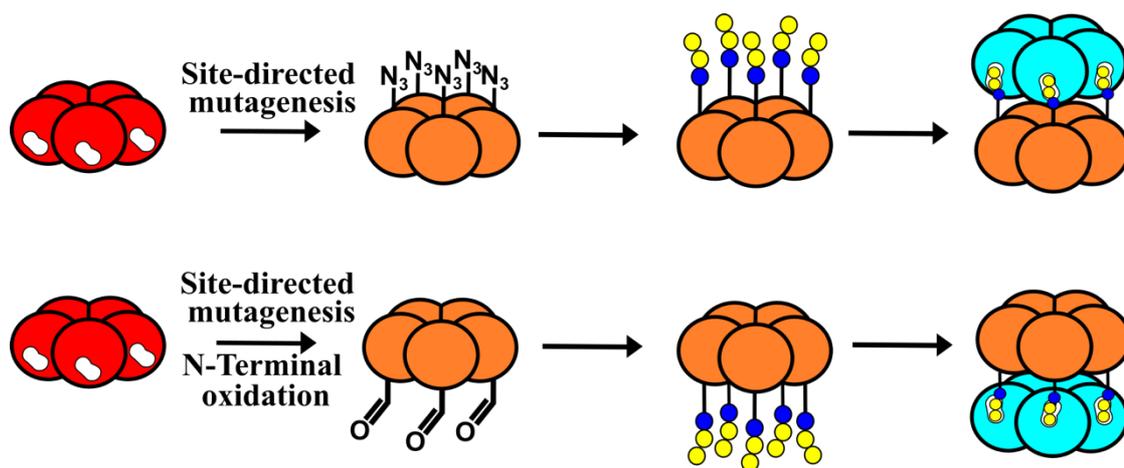


Figure 5.1: Site-selective bioorthogonally glycosylation strategies for pentameric scaffold, for the generation of pentavalent glycoprotein inhibitors of bacterial toxin B-subunit.

5.1.1 Generation of CTB-Gb3 neo-glycoproteins by oxime ligation

Having successfully attached the lactose glycoconjugate **3.13** by oxime ligation to the W88E scaffold, it was first thought that this could be enzymatically converted to the Gb3 trisaccharide on the protein. The addition of an α -galactosyl residue using the LgtC galactosyltransferase was attempted but showed no conversion of the lactosyl- to Gb3-glycoprotein when the experiment was followed by ESMS. It is possible that the LgtC active site is unable to accommodate the terminal galactosyl residue when it is attached directly to the protein surface, either because of steric hindrance or if electrostatic repulsion of the two protein could contribute to the lack of enzymatic activity observed. Therefore, oxime ligation of the glycoconjugate **3.19** was performed (figure 5.2). The deconvoluted high-resolution mass spectra are presented for **W88E**, overlaid with the spectra after oxidation, and following oxime ligation to **3.19**. The formation of **W88Eox** was near quantitative showing a small amount of cyclised by-product, with 2% unoxidised **W88E** remaining. Upon ligation, all **W88Eox** is converted to the ligated **W88E(3.19)** corresponding to the mass 12,217 kDa. The major species detected appears to be the lactose product at 12,055 kDa. Additional experiments confirmed that the sample of **3.19** was indeed the correct and had not been contaminated with **3.13**, therefore we could conclude that the terminal α -1,4-Gal is more susceptible to fragmentation in the mass spectrometer than the Gal- α -1,4-Glc core. Summation of the signals for **W88E(3.19)** and its corresponding fragments gave a percentage conversion of 95% from the starting **W88E**.

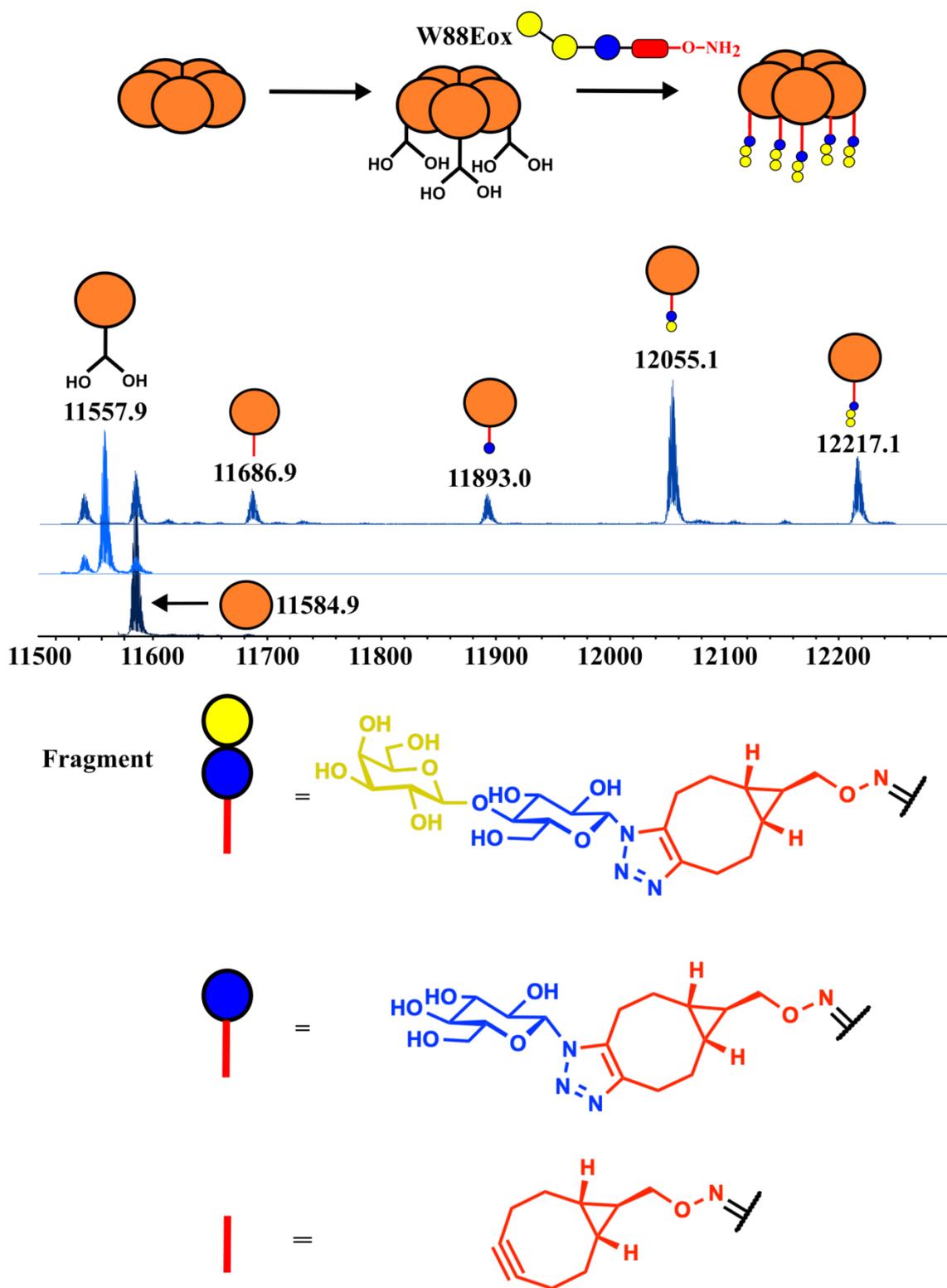


Figure 5.2: Overlaid deconvoluted HRMS showing the starting W88E protein, the spectrum following oxidation (392 μM N_3 -W88E, 4 mM in phosphate buffer pH 7) and the spectrum after the oxime ligation (636 μM N_3 -W88E, 6.46 mM 3.19, 100 mM aniline in phosphate buffer pH 6.8, 18hours at room temperature).

5.1.2 Generation of CTB-Gb3 neo-glycoproteins by SPAAC glycosylation

A neoglycoprotein where Gb3 is ligated to the non-binding face of CTB, was synthesised to allow comparison of the glycosylation site on the inhibitory potential towards VTB. The SPAAC glycosylation between the trisaccharide **3.17** and **N₃-W88E** was carried out using the optimised reaction conditions described in section 4.3. The reaction was determined to be complete after four hours, following desalting by PD-10 column (figure 5.3).

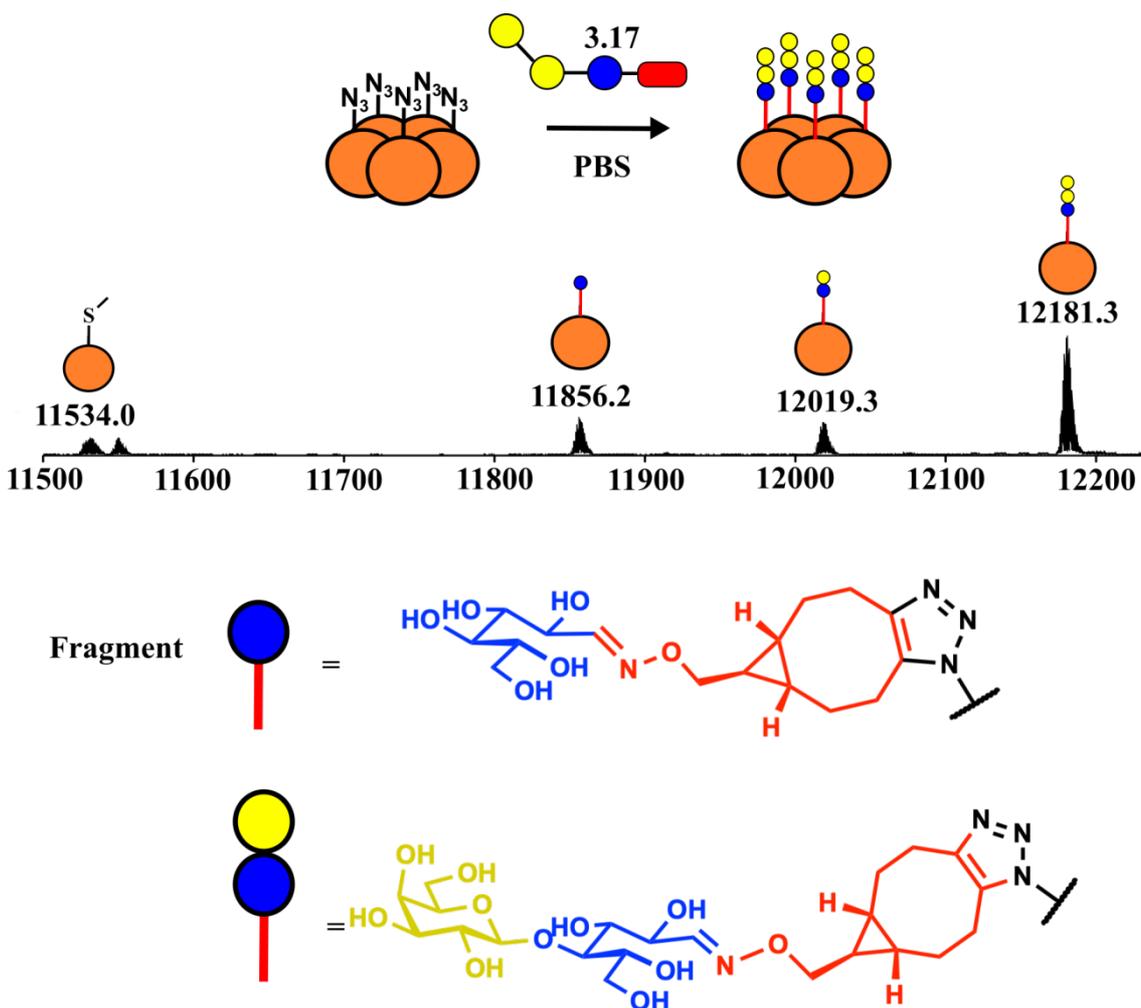


Figure 5.3: Deconvoluted HRMS showing the results of SPAAC glycosylation between **N₃-W88E** and glycoconjugate **3.17** (500 μ M **N₃-W88E**, 10 mM **3.17**, 4hours at room temperature).

5.2 Expression of recombinant VTB

After the synthesis of two neoglycoproteins towards VTB, the next task was to express the VTB-subunit to perform inhibition studies. VTB had not previously been expressed within our group, however, cell stocks of *E. coli* XL10 and expression cell lines of WT VTB and various VTB mutants were inherited as a gift from Prof. Steve Homans (University of Leeds). Although information on the expression vector, promoter system or expression protocol was not available for these clones, publications from Homans and co-workers indicated that the gene had been cloned a pUC-18 vector or a derivative of the pUC-18 vector. A primer designed to anneal in the middle of the VTB gene, was used to confirm the vector was pUC-18 and that the gene was under the control of the *LacZa* promoter system. Knowing the vector, the full gene was then sequenced using the M13 reverse primer, confirming the full VTB sequence and periplasmic target sequence (appendix, section 10.5). Expression tests of this plasmid were carried out with the assistance of MChem student Charlie Stevenson.

5.2.1 Expression tests

Expression tests were carried out using the glycerol stocks of the inherited expression *E. coli* cell lines. Cultures were grown in LB and LB autoinduction media at temperatures of 37 °C and 25 °C. For each media and temperature, the growth media, cell lysate and cell pellet were all examined for the presence of VTB-subunit and no expression of the VTB subunit was observed (data not shown). Given the age of the glycerol stocks, the first approach to addressing this problem was to make new glycerol stocks. The VTB plasmid was isolated by mini-prep from culture of *E. coli* XL10 cell stocks harbouring the VTB plasmid. Chemically competent *E. coli* BL21 expression cells were then transformed with the VTB plasmid and used in a second round of expression tests. Disappointingly no expression was observed under any of the conditions (data not shown).

A literature search of recombinant expression/purification of VTB indicated that high-yielding expression of bacterial toxins, and particularly VTB, had been observed in *E. coli* grown in CAYE (casamino acid-yeast extract) growth medium modified according to Evans.²⁸⁸⁻²⁹¹ Another protocol with repeated literature precedent was the extraction of VTB by periplasmic extraction^{292,293}, as the toxin is targeted to periplasmic space where it folds. Periplasmic extraction involves the breaking apart the outer membrane isolating

proteins transported into the periplasm, without extraction of the whole cellular proteome leaving the inner membrane intact.

Small scale expression of VTB was then attempted in 5 ml cultures with the newly prepared BL21 glycerol stock. After the cultures were induced by addition of IPTG, the incubation time was varied from 1.5 hours to 24 hours to determine optimal expression time. Once the cells had been harvested, the periplasmic extraction was performed, and rest of the cellular components discarded. The SDS-PAGE gel (figure 5.4) shows the successful expression of VTB, confirmed by the appearance of monomeric VTB protein at 7.7 kDa, as the pentamer is not stable in the presence of SDS. This gel highlights the benefit of a periplasmic extraction leaving the samples are relatively clean, compared to a whole cell lysis.

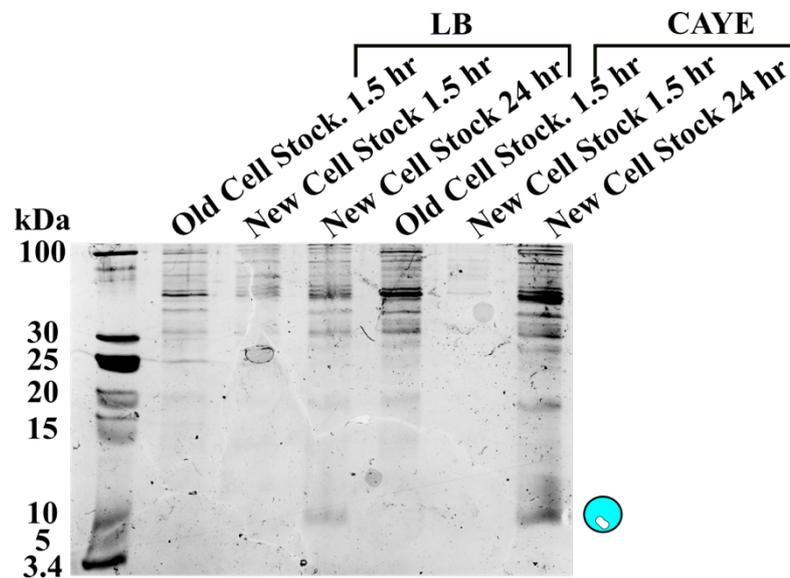


Figure 5.4: SDS-PAGE showing the results of an expression test (5 ml culture) of old and new cell stocks in LB and CAYE with various times after IPTG induction and following periplasmic extraction of the harvested cells.

Interestingly, the newly prepared cell stock showed expression in both LB and CAYE media, however, expression was markedly better in CAYE broth (figure 5.4). After confirming the expression in CAYE broth, larger 1 L cultures were grown and expression times of 18 and 42 hours were trialled. Figure 5.5 shows the levels of overexpression loading the same volume of periplasmic extract, from the same volume of cell culture. Expression overnight showed much higher concentrations of VTB than expression over two days; longer expression time could result in protein degradation or cell death releasing the contents of the cell into the growth medium, and decreasing protein yields. To confirm any possible export/release of the toxin into the growth medium, expression would have to be revisited and analysis of the growth medium performed by ammonium sulfate precipitation.

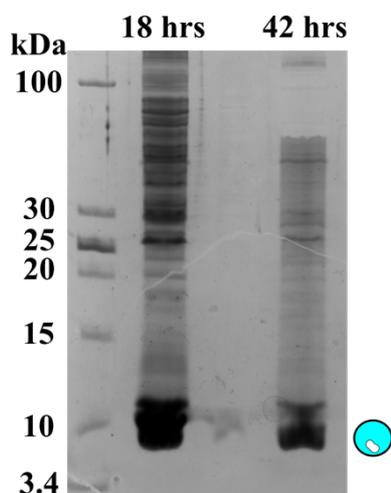


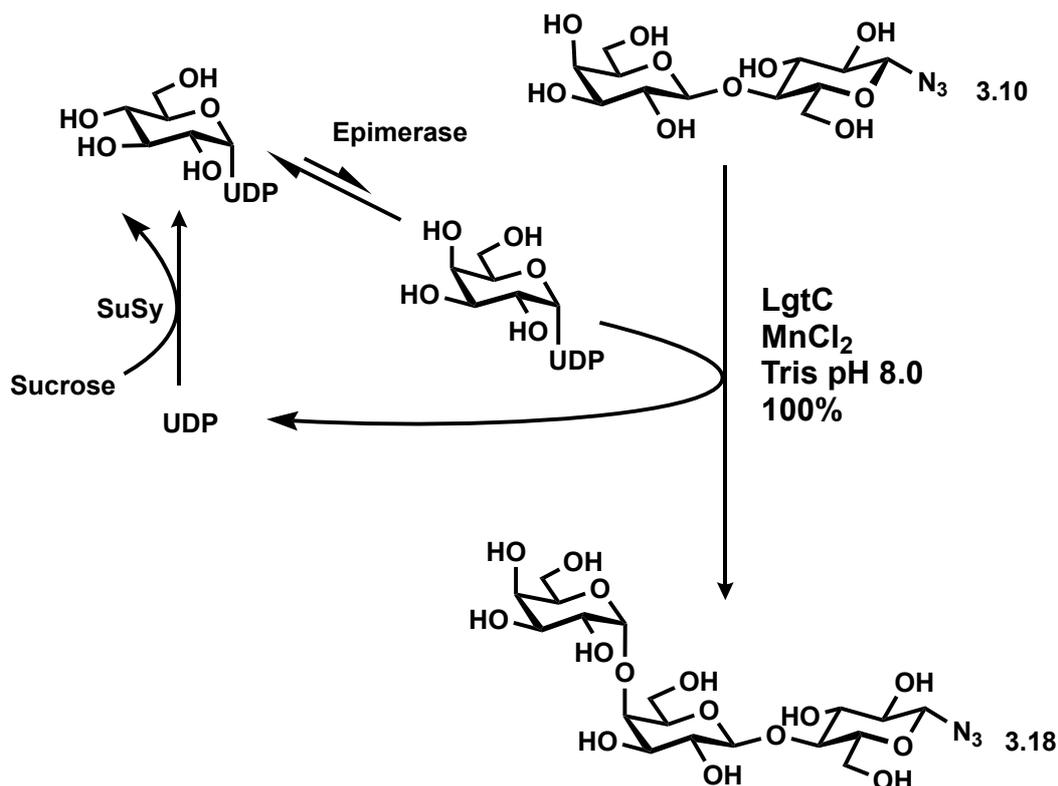
Figure 5.5: Larger scale growth cultures expressing VTB in CAYE broth in 500 ml, the results of expression over 18 and 42 hours at 25 °C shown by SDS-PAGE of the crude protein after periplasmic extraction from harvested cells.

5.2.2 VTB purification

Upon successful expression, a purification method needed to be identified. Even though no purification tag was encoded in the gene sequence, periplasmic extraction produced a relatively clean lysis mixture, which made purification a simpler task. Chromatofocusing was a potential purification method known to be successful,^{294, 295} however, we did not have the required equipment available. Some affinity resins have been reported for VTB affinity purification, such as sepharose resins functionalised with galactose,²⁹⁰ galabiose,²⁹⁶ or Gb₃,²⁹⁷ the latter two of which contain the Gal- α -1,4-Gal I ligand. Despite one report of a commercial resin,²⁹⁶ we could not find such resin from a commercial source; therefore, an affinity resin would have to be made from scratch. By generating a Gb₃ affinity column, we can exploit the VTB's native Gb₃ binding with confidence that purification can be achieved by strong interaction due to the 15 binding sites of the protein. No other proteins expressed by *E. coli* are expected to have any innate binding to Gb₃, therefore, this should act as a one-step purification method.

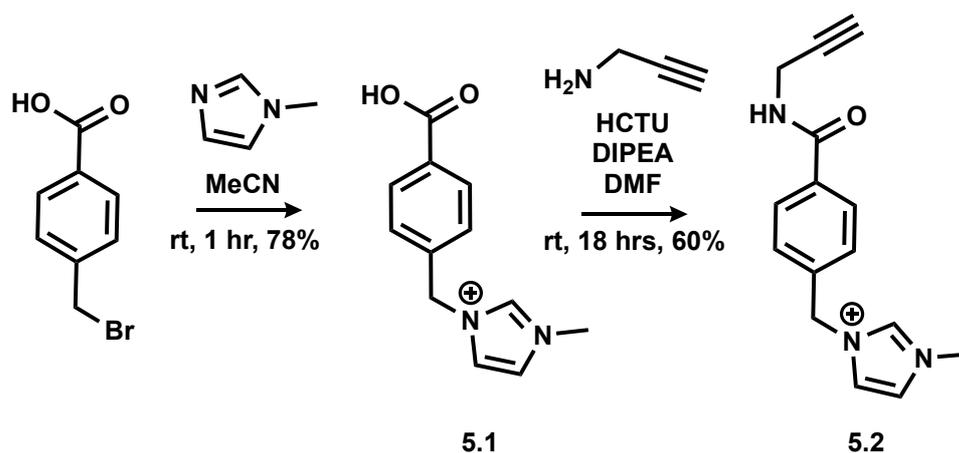
5.2.2.1 Enzymatic synthesis of chemically reactive Gb₃

As scheme 5.1 shows, the optimised enzymatic cycle for Gb₃ azide **3.18** was increased to a 500 mg scale in a 10 ml volume to produce enough of **3.18** to functionalise 5-10 ml of a resin. The enzymatic cascade was effective on a large scale, maintaining quantitative conversion of the lactosyl azide **3.11** when left for 16 hours.



Scheme 5.1: Enzymatic cascade for the large scale synthesis of Gb3-N₃ 3.18. Conditions: 3.11 (368 mg, 100 mM), LgtC (100 μM), UDP-Glc (20 mM) in 10 ml.

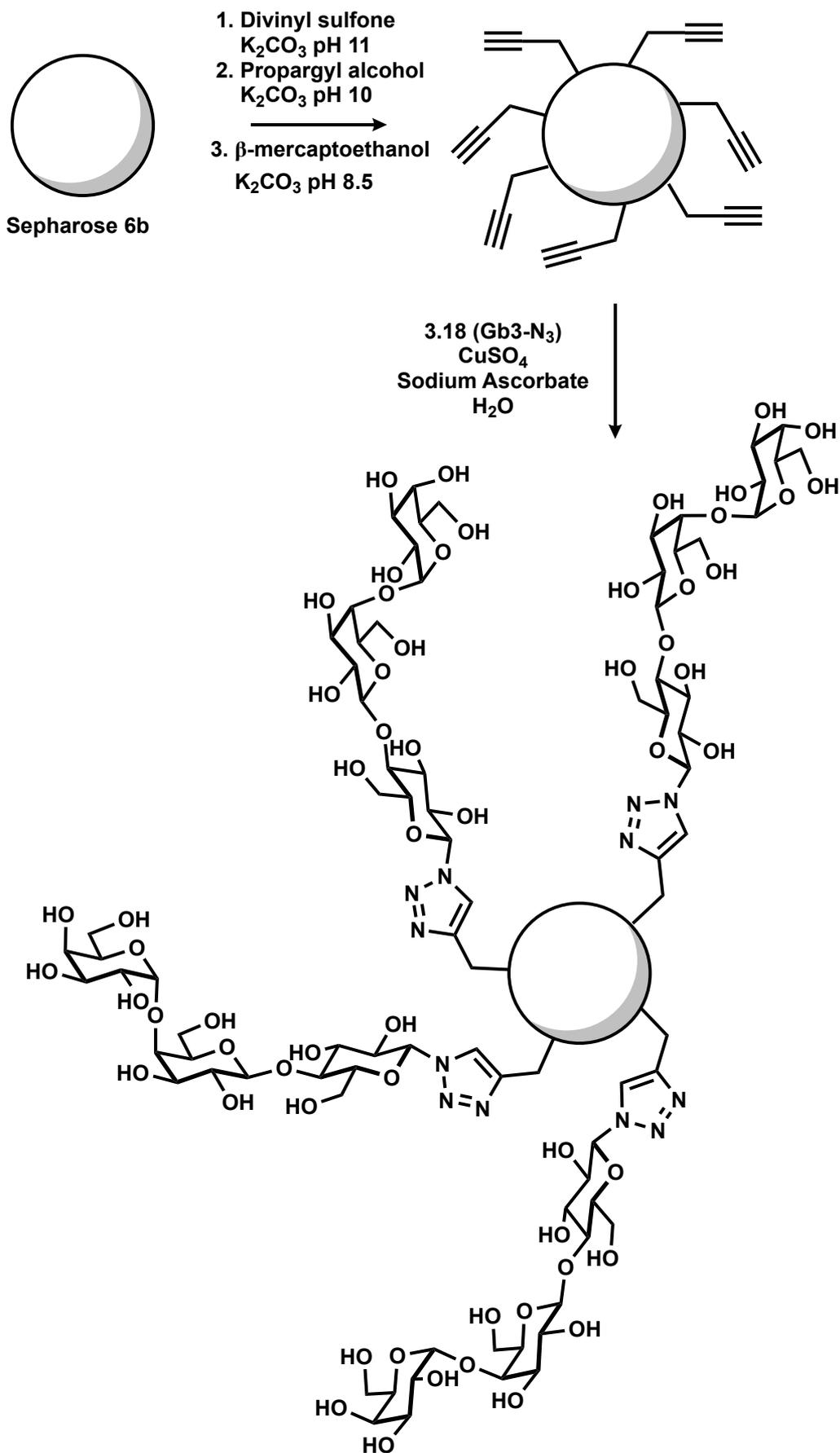
Full conversion was confirmed by TLC, but in support of this a mass spectroscopy tag was developed to give a highly sensitive tool to confirm the sole glycosyl azide in the reaction was the Gb3 product. This tag was based upon the work of Galan and co-workers, who developed ‘I-tag’ for purification and analysis of carbohydrate reactions.¹⁰¹ These structures have a permanent cation, making them useful mass spectroscopy tools for oligosaccharides which are difficult to detect by electrospray MS. A short length ‘clickable’ I-tag analogue was obtained in two-steps (scheme 5.2), starting from 4-(bromomethyl)benzyl alcohol. The terminal alkyne of **5.2** was then used for attachment to glycosyl azides using CuAAC chemistry with a sample of crude reaction mixture. After five minutes, the sample was examined by LC-MS with the mass of the Itag conjugate was observed $[M]^+ = 783.18$ Da (expected Gb3-Itag $[M]^+ = 783.30$ Da), with no mass for the I-tag conjugate of lactosyl azide **3.11** detected. Having detected only a single glycosyl azide, the crude reaction mixture could be used to functionalise the resin directly, following precipitation and removal of the enzymes.



Scheme 5.2: Synthesis of a novel mass spectrum tag from the 4-(bromomethyl) benzoic acid for monitoring glycosylations by LC-MS.

5.2.2.2 Functionalisation of sepharose 6b for the creation of a Gb₃ affinity resin

After producing **3.18** on a scale large enough to functionalise a significant amount of resin, a resin then needed to be functionalised with the orthogonal reactive handle. A sepharose 6b resin was decorated with alkyne groups, by activation with divinyl sulfone following a procedure reported by Portath *et al.*²⁹⁸ (scheme 5.3). As it proved challenging to quantify the degree of modification of the resin and thus determine number of alkynes present on the resin, instead a fluorescent reporter was used to confirm that alkyne functionalisation on the resin was successful.



Scheme 5.3: Preparation of Gb3 affinity resin for the purification by VTB. Hydroxyl groups on the sepharose 6b resin were derivatised with propargyl groups, through activation with divinyl sulfone and reaction with propargyl alcohol. The affinity resin is then produced by CuAAC reaction of the resin to Gb3-N₃.

CuAAC-catalysed attachment of TAMRA-N₃ to the resin was performed, mixing the resin with Cu(SO₄), sodium ascorbate and TAMRA-N₃ was gently agitated for two hours. The resin was collected by centrifugation and the supernatant removed, followed by repeated washing steps until the supernatant remained colourless. Figure 5.6 shows an image of a control (left) reaction in which no Cu(SO₄) was added vs the CuAAC reaction with TAMRA-N₃ (right): only the reaction with Cu(SO₄) retained a visible red colour. The presence of TAMRA was verified by measurement of fluorescence using Bio-rad Chemidoc exciting at 555 nm and measuring the fluorescence at 580 nm using. Only the reaction contain Cu(SO₄) showed fluorescence, confirming successful functionalisation of the resin.

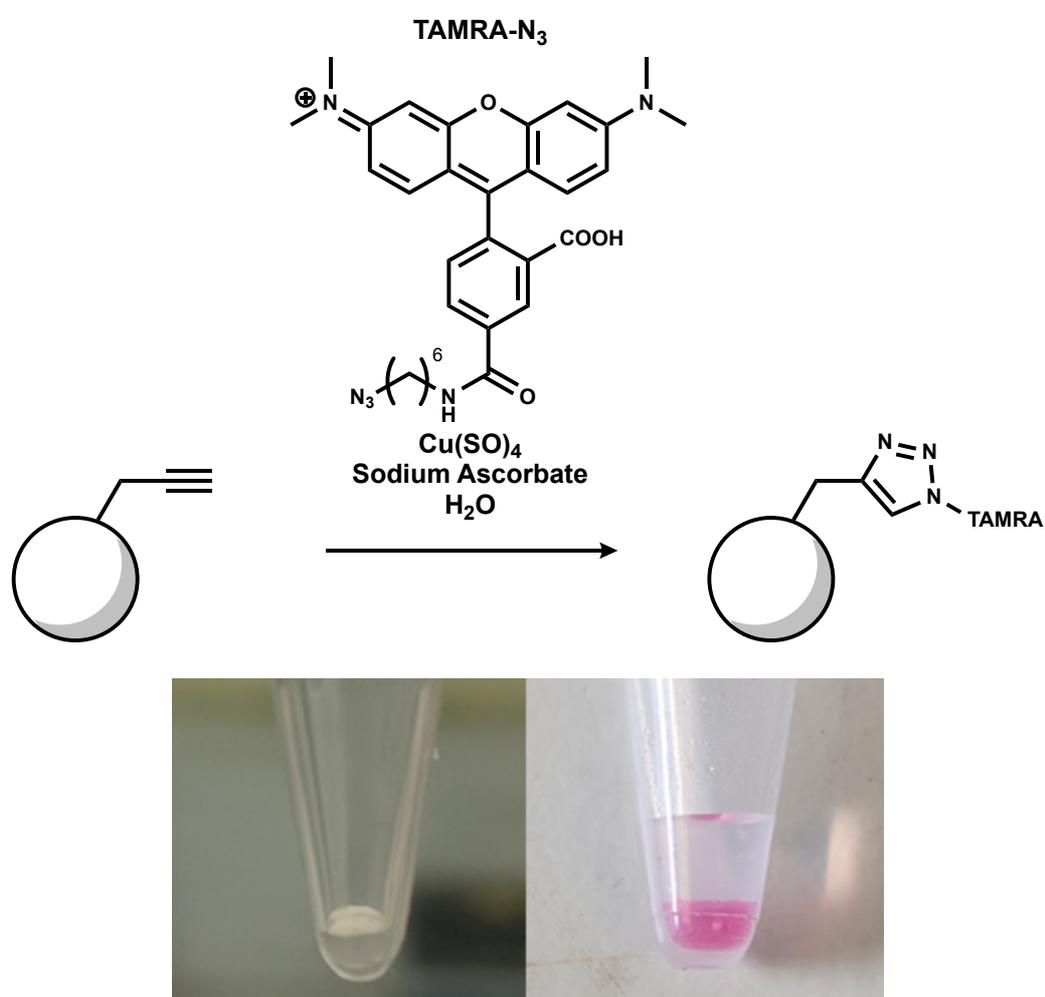


Figure 5.6: Images showing the TAMRA labelling of alkyne functionalised sepharose to confirm the presence of alkynes on the surface of the resin. Control: Resin, TAMRA-N₃ Na-ascorbate, H₂O (left). TAMRA label: TAMRA-N₃, Resin, Cu(SO₄), Sodium ascorbate, H₂O (right)

The CuAAC-catalysed attachment of **3.18** to the alkyne-sepharose was carried out using the crude enzymatic reaction mixture in a large excess, ensuring there was sufficient **3.18** to react with all alkynes on the resin. Following incubation of the reaction at 37 °C the

resin was purified by sequential centrifugation and washing, removing the reactants from both enzymatic and CuAAC reactions.

Taking the periplasmic extract from the 42-hr expression, was flowed through the Gb3-affinity column, before washing the column to remove any unwanted proteins. VTB was then eluted from the column in 4.5 M MgCl₂ in PBS, disrupting the interaction of the Gb3 ligand in the VTB binding sites with the high ionic strength solution.^{290, 297, 299, 300} Successful elution of VTB was observed in the elution fractions, the flow-through following spin concentration was analysed to confirm the VTB pentamer had not degraded in 4.5 MgCl₂ as the monomer would pass through the cellulose membrane (figure 5.7). No VTB was detected in the flow-through, indicating that the pentamer remained intact during the purification. The affinity resin was a successful method of purification of the VTB subunit, and similarly in the synthesis of glycoconjugates, the enzymatic cascades allow large scale synthesis of the Gb3 oligosaccharides required to make such purification methods possible. Exploiting the affinity of VTB for the Gb3 makes this a highly selective method of purification from the native *E. coli* periplasmic proteins.

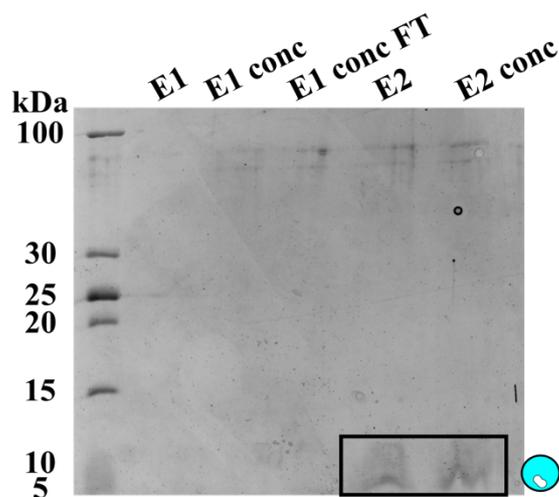


Figure 5.7: SDS-PAGE following the purification of VTB expression purified by Gb3 affinity chromatography.

5.2.2.3 Alternative purification methods

Alongside the development of an affinity resin other purification methods were investigated. The periplasmic extraction was so successful in minimising the amount of unwanted proteins before purification, and with the VTB-subunit overexpressed, it was also possible to use SEC on Superdex 75 column (figure 5.8, carried out by MChem student Charlie Stevenson), to isolate the VTB pentamer directly following periplasmic lysis with great simplicity. The purified sample of VTB was then analysed by ES-HRMS confirming the mass of the VTB monomer at 7687.8 Da (expected mass = 7688.6 Da).

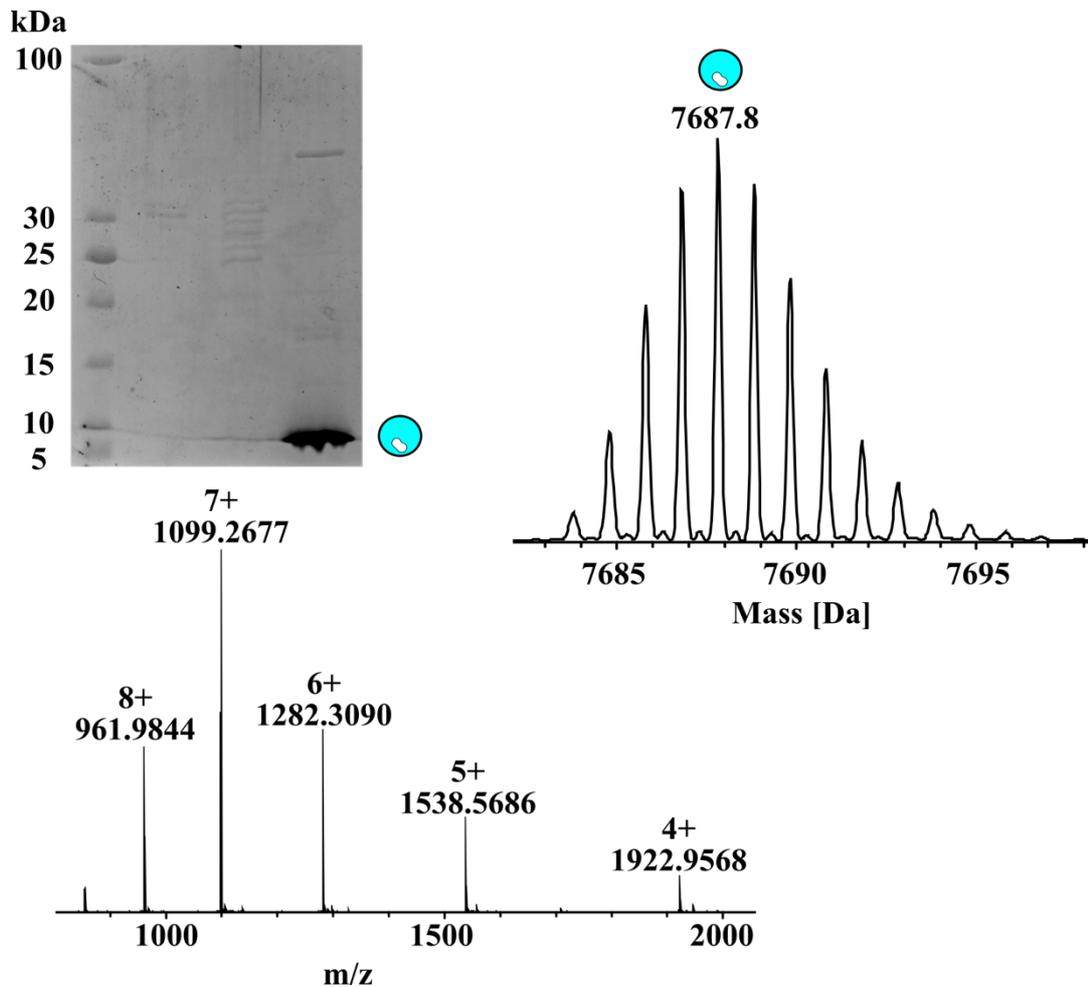


Figure 5.8: SDS-PAGE analysis of fractions showing an absorbance at 280 nm from purification of VTB by size exclusion on Superdex S75 column and ES-HRMS of the purified sample.

5.2.3 Summary of VTB expression

Although the overexpression of VTB in *E. coli* is not novel within the literature, it was a novel task within the Turnbull group. An expression vector inherited from the Homans lab was sequenced, and CAYE broth was found to greatly improve the expression of VTB. It was determined that performing a periplasmic extraction is effective in isolating the protein without then having to separate the target protein from the rest of the cellular proteome, which proved vital for detection and purification of the VTB pentamer which did not contain a purification tag. Two methods were used for successful final purification of the VTB pentamer, a Gb3 affinity resin and SEC were both effective for isolating pure VTB protein, although the SEC method was more practical and accessible.

5.3 Development of ELLA assay for testing of inhibitors against the Verotoxin

5.3.1 VTB-HRP conjugation

The next component of the ELLA to be prepared was the HRP protein conjugate for the fluorescent assessment of VTB binding. As with the previous assay for CTB, the protein conjugation was carried out using the commercial Lightning-link[®] HRP conjugation kit from Expedeon. In attempts to control the amount of HRP conjugated per pentamer the conjugation was performed using a 1:1 stoichiometric amount of VTB pentamer to HRP. HRMS was used to confirm the formation of the HRP conjugate, and the ratio of VTB pentamer:HRP that was present after conjugation. Figure 5.9 shows the deconvoluted mass spectrum following the quenching step of the conjugation protocol. As a wide mass range to be observed a simple deconvolution was applied which showed a mass of 48,816 Da corresponding to the VTB monomer conjugated to HRP. HRP is a glycoprotein with eight glycosylation sites, taking the mass of the peptide sequence taken from Uniprot was calculated at 38,825. The addition of VTB monomer at 7.68 kDa and taking into consideration the known presence of a haeme group and the commercial linker this data suggests the conjugation is to a deglycosylated isoform of HRP.

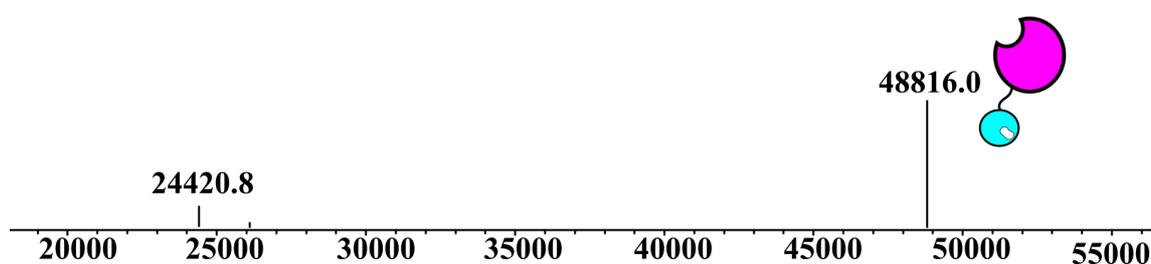


Figure 5.9: A deconvoluted ES-HRMS spectrum showing the high mass corresponding to a VTB monomer conjugated a deglycosylated HRP

5.3.2 Synthesis of Gb3 glycolipids

The design of the assay to assess the potency of these inhibitors was to follow that of the ELLA used for the CT inhibitors. Therefore, a Gb3 glycolipid was required with which to coat the microtiter plates. For the previous assay, microtiter plates were coated with the natural ganglioside, which is commercially available at £310 for 500 mg. Purchasing globotriaoceramide or Gb₃ ceramide was considered, but was determined to be too expensive at a cost of £517 per mg.

Two possible routes were investigated to produce a Gb3 glycolipid at a lower cost and in larger quantities. The first was the enzymatic conversion of the commercial lactosyl

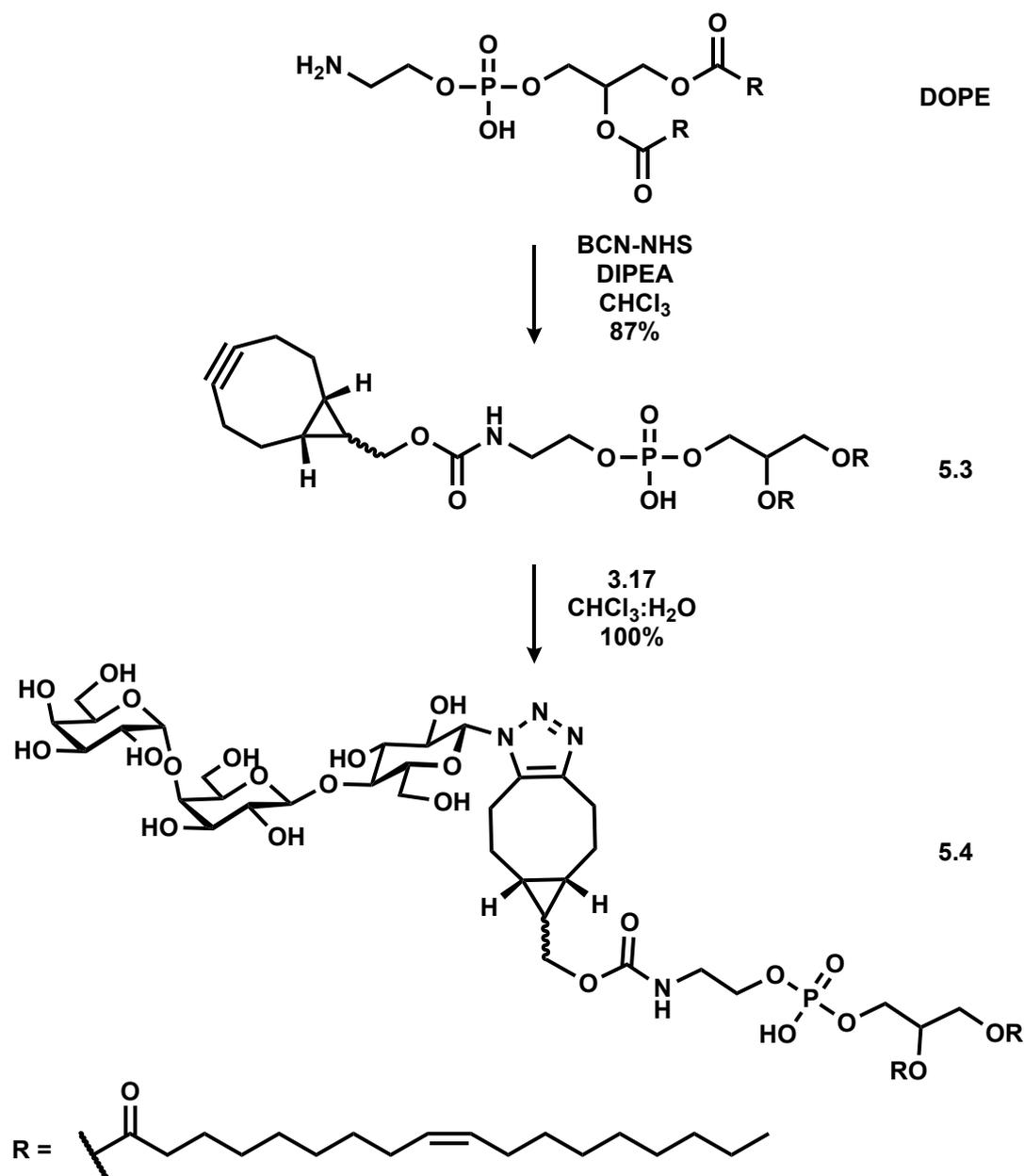
ceramide using the LgtC enzyme available within the lab. Alternatively, a commercial phospholipid 1,2-Dioleoyl-sn-glycero-3-phosphoethanolamine(DOPE) could be derivatised with an alkyne group, to which the Gb3-N₃ **3.18** could be attached to create a synthetic glycolipid.

5.3.2.1 Enzymatic synthesis of Gb3 ceramide

The enzymatic conversion of lactosyl ceramide (Carbosynth) was investigated with MChem student Charlie Stevenson. Palcic and co-workers had previously reported the synthesis of Gb3 ceramide analogues, and noted that the solubilisation of the glycolipid was required for reaction in aqueous solvent.³⁰¹ Therefore, a cyclodextrin derivative was required to aid the solubilisation of the lipid, the enzymatic galactosylation was carried out using a large excess of UDP-Gal to account for any possible hydrolysis of the donor sugar. In this case the enzymatic cycle (section 3.4.3) was not used as it would complicate the purification of the glycolipid significantly. The reaction was incubated at 37 °C and monitored by TLC with further addition of UDP-Gal until the reaction was determined to be complete. Preparative TLC allowed isolation of three separate compounds, however, analysis by mass spectrometry and NMR spectroscopy was unsuccessful as all attempts to get the products into homogenous solution failed. As a functional test for whether the reaction had been successful, each of the three compounds isolated from the prep-TLC was dissolved in methanol to a concentration of 1.3 μM and adsorbed onto a microtiter plate, and the ability of each compound to capture the VTB-HRP conjugate was determined. After incubation of the VTB-HRP conjugate at room temperature for 1 hour, the plate was washed and Amplex Red® added, however no fluorescence was detected, indicating that none of the compounds isolated were capable of binding VTB.

5.3.2.2 Synthesis of a Gb3 phospholipid

The second option was to make a Gb3 neoglycolipid based on DOPE, which bears a free amine (scheme 5.3), and BCN group for attachment of Gb3-azide **3.18** with the SPAAC reaction. A commercial NHS-activated BCN (Galchima) reacted efficiently with DOPE (scheme 5.3). The increased hydrophobicity of the product facilitated purification of the BCN phospholipid **5.3**, which was isolated 87 % yield. SPAAC ligation to **3.18** yielding the synthetic glycopospholipid in quantitative yield with no purification required.



Scheme 5.3: Synthesis of a strained alkyne containing phospholipid from the commercial lipid DOPE, through amide coupling of commercial BCN-NHS.

5.4 VTB capture on microtiter plates

The Gb3 phospholipid **5.4**, was adsorbed onto the wells of a microtiter plate in an identical protocol to the ELLA for CTB. A preliminary assay was performed to determine what concentration of lipid needed to be adsorbed on the surface, to capture VTB on the plate and what concentration of VT-HRP gave a good level of signal-to-noise with Amplex Red[®]. The experiment showed that the synthetic Gb3-phospholipid **5.4** was capable of binding the VTB-HRP conjugate and the HRP enzyme retained its enzyme activity. From the raw graphical data (see appendix, figure 10.9) produced by the plate reader it was clear that the VTB-HRP conjugate has been captured on the plate at all concentrations of

Gb3-DOPE, the fact that the data is very similar at each concentration shows that the well is saturated with the glycolipid even lowest concentration tested. Plotting the data obtained for the 2-fold dilution of the VTB-HRP conjugate at the lowest concentration of 1.25 μM DOPE (figure 5.10) it was concluded that a good signal-to-noise was obtained in the region of 3-6 ng/ml of VTB-HRP. Therefore 5 ng/ml was adopted for subsequent assays.

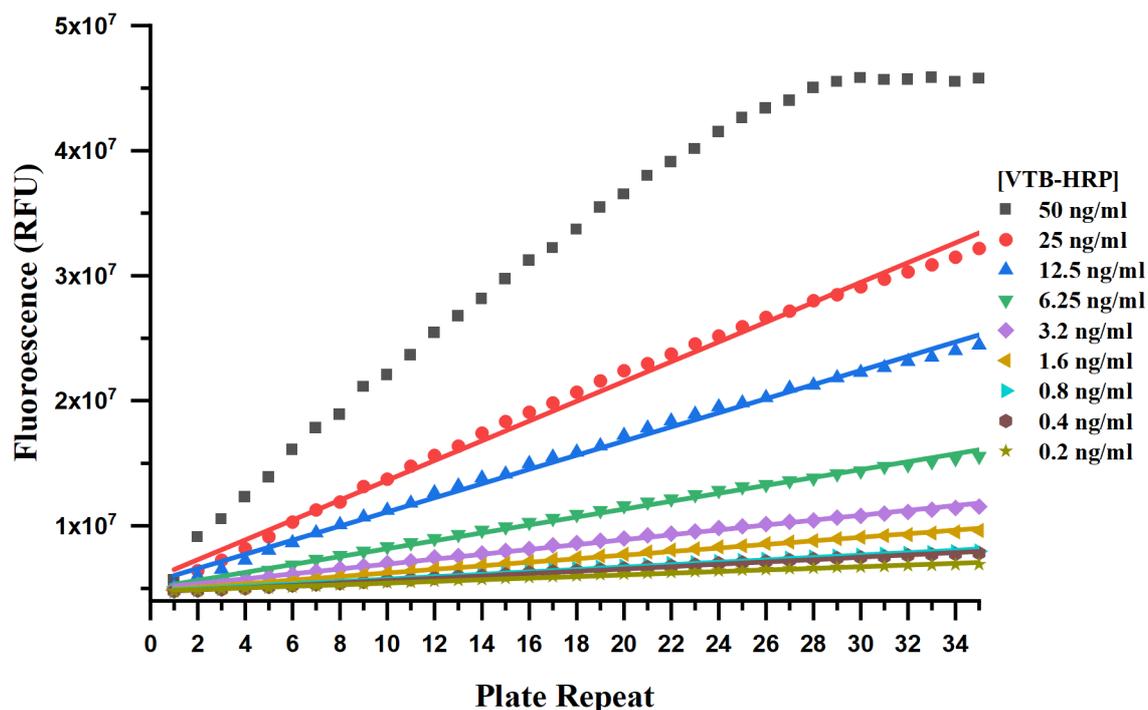


Figure 5.10: Preliminary assay to determine the concentration of 5.4 (Gb3-DOPE) required to capture VTB-HRP on microtiter plate and the concentration of VTB-HRP required for good signal-to-noise. The raw graphical data at each glycolipid concentration is shown and the fluorescence measurements at 1.25 μM 5.4 compared following a 2-fold serial dilution of VTB-HRP from 50 ng/ml.

An inhibition test for Gb3os was carried out following this at the VTB-HRP concentration of 5ng/ml one month after the preliminary experiment. Surprisingly, the assay produced no data with even the positive control wells showing no evidence presence of VTB-HRP. A preliminary assay which produced the data in figure 5.10 was repeated to ensure the null result was not caused addition of VTB protein, but it was found that data (figure 5.10) could not be reproduced. With the batch of VTB-HRP conjugate produced a few months previously, a new kit was purchased and a new sample of VTB-HRP conjugate prepared. No binding was observed once again leading to the conclusion that there was a problem with the phospholipid.

The successful binding assay showing VTB capture was obtained immediately after the dissolution and adsorption of Gb3-DOPE to the microtiter plate. The lipid was then stored

in methanol at -20 °C for one month, between the successful assay and the failed inhibition study. Gb3-DOPE was stored under the same conditions as ganglioside GM1, which was stable for long periods of time even at low concentration. Potential hydrolysis/methanolysis of the lipid would result in the Gb3 ligands being washed away. The phospholipid was to be remade but required the synthesis of more Gb3-N₃ **3.18**. However, the final few months of the project were cut short by the coronavirus lockdown and this could not be carried out. To conclude this work and develop an assay for assessment of neoglycoprotein-based inhibitors of VTB, Gb3-DOPE must be resynthesised, and effective adsorption/storage methods are required to produce a functional assay. Once reproducible, capture and detection of VTB-HRP can be achieved, the inhibition studies comparing both lactose and Gb3-based neoglycoproteins ligated through SPAAC and oxime ligation can be assessed along with the natural Gb3os.

5.5 Conclusion

Two pentavalent neoglycoproteins bearing the Gb3 ligand have been prepared using a combination of the synthetic glycosylation methods described in chapter 4, and the mutant CTB protein as a scaffold. A protocol for the successful overexpression and purification of VTB was developed, including a method using a homemade affinity resin generated through large scale enzymatic synthesis of Gb3-N₃ **3.18**. After purification of VTB, a protein conjugate with the enzyme HRP was successfully prepared using a commercial lightning-link conjugation kit. In order to establish an ELLA to test these neoglycoproteins as inhibitors of VTB adhesion, a method of adsorbing Gb3 to a microtiter plate surface was required. Attempts to prepare Gb3-ceramide by galactosylation of lactosyl-ceramide were unsuccessful; instead a synthetic Gb3 neoglycolipid **4.6** was synthesised from the commercial lipid DOPE. A trial assay using this synthetic glycolipid to capture VTB-HRP onto a microtiter plate gave a successful result, however, this could not be reproduced after the lipid had been stored in the freezer for one month. We concluded that the glycolipid is labile and potentially susceptible to hydrolysis/methanolysis when stored in methanol. To confirm this hypothesis, neoglycolipid **4.6** needed to be resynthesised, however due to the project being cut short by the coronavirus lockdown, this could not be done. Without successful reproduction of the assay, the inhibitory potential of the neoglycoproteins could not be assessed at this time. Once the capture VTB-HRP conjugate can be validated, the neoglycoprotein and Gb3os will be tested as inhibitors of the VTB adhesion.

**Chapter 6 Towards non-bacterial pentameric protein scaffolds for
multivalent protein-based therapeutics**

The CTB scaffold is convenient to use as it is easy to express and manipulate; highly stable; has pentagonal symmetry; it is useful to demonstrate to the potential of neoglycoprotein inhibitors. However, despite the B subunit having no toxic activity, as a bacterial toxin there is a risk of evoking a host immune response if used *in vivo*. Therefore, to move this class of neoglycoprotein inhibitors towards a viable therapy, the scaffold protein would ideally need to be changed. One possibility would be to use a class of pentameric human proteins that are pentameric known as the pentraxin family. Figure 6.1 shows two closely related members of the pentraxin family; C-reactive protein (CRP) and serum amyloid-P (SAP). These proteins are homopentamers that share 51% sequence homology have many different binding partners and functions.³⁰² CRP is involved in the innate immune response and has a major role in inflammatory pathways. It is also clinical diagnostic tool used as a marker for inflammation.³⁰³ SAP, on the other hand, is responsible for the recruitment of macrophages and is particularly known for its binding to amyloid and fibrils, to recruit cells responsible for the clearance of cellular debris and protein aggregates.^{304 305}

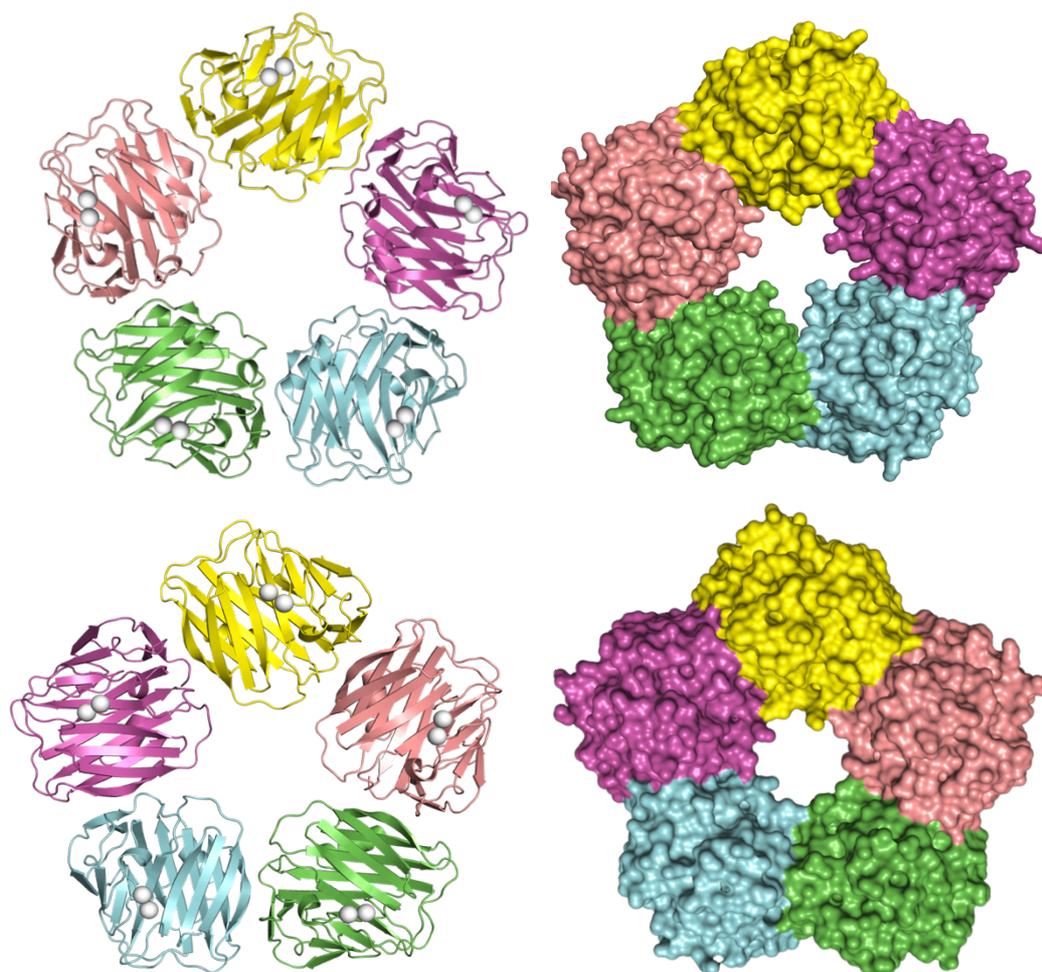


Figure 6.1: Crystal structures showing two members of the pentraxin family CRP (top) and SAP (bottom) in both cartoon and surface representations with each monomeric unit coloured by chain.

6.1 Serum amyloid P protein a human protein scaffold

The manipulation of SAP has been thought of as potential targets for therapeutic treatments as an antifibrotic for example³⁰⁵ as it is understood from mouse models a lack of SAP results in increased pulmonary fibrosis, but it could be used for glycoconjugate based therapeutics if exploited as a scaffold for synthetic glycosylation. SAP's homopentameric structure and pentagonal symmetry make it an ideal candidate for creating neoglycoprotein based therapeutics, which target AB₅ enterotoxins using the same multivalent inhibition strategy which was so effective against CTB.

Overlaying the structures of the SAP and VTB pentamers and viewing along their common five-fold symmetry axis (figure 6.2), highlights the difference in size between the two lectins. The SAP protein was examined for potential glycosylation sites which would be optimal for presenting Gb3 to each the VTB binding sites. Figure 6.2 also shows three theoretical models based upon the SAP and VTB crystal structures, but in which a C-terminal extension has been generated *in silico* and the linker **1.2** modelled and connected to the bound Gb3 ligands. This model confirms that the extension is capable of reaching each of the three VTB binding sites, remaining in a linear conformation. This also confirmed that modification of the C-terminus would be suitable as the distances between the C-termini closely match that of the VTB pentamer and the Gb3 binding sites.

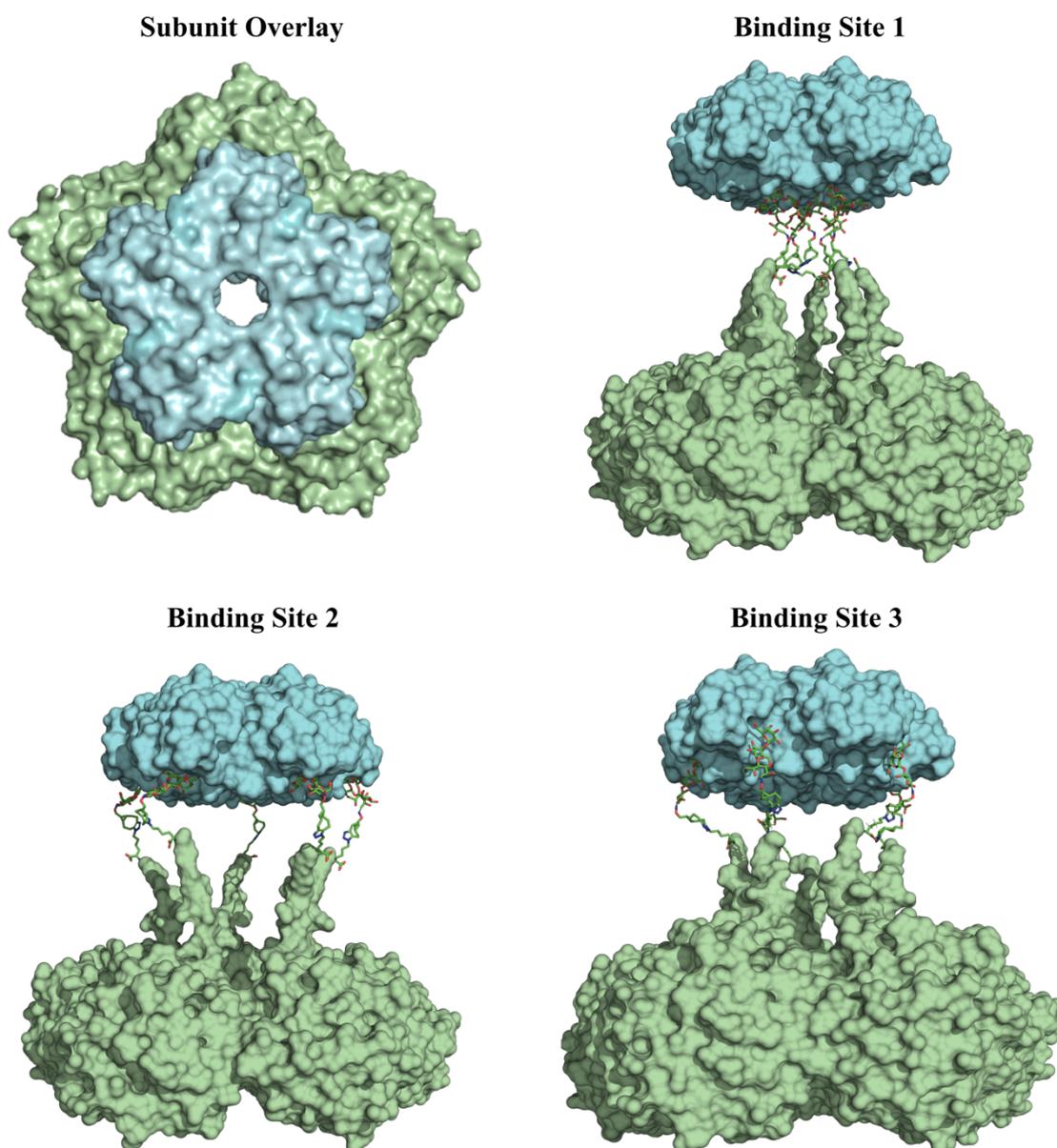
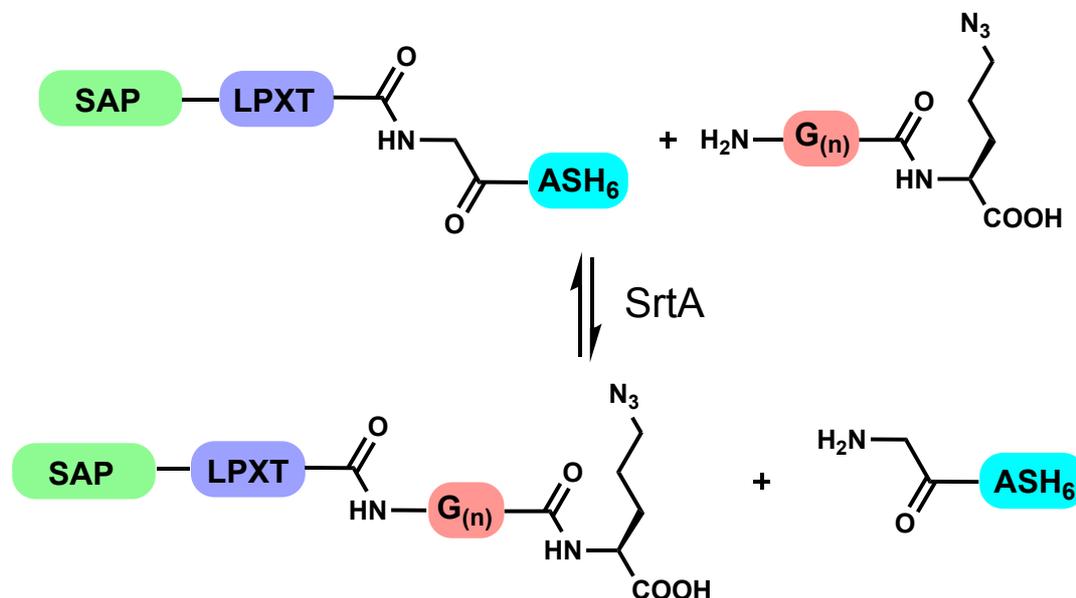


Figure 6.2: Overlay of the human pentraxin protein SAP (green) and the pentameric verotoxin B subunit (blue), comparing the size and geometry of the pentameric proteins. Three structures showing an *in silico* generated SAP with a LPETGGGK extension connected to Gb3-BCN bound to VTB-subunit in each of the three binding sites.

6.2 Design of a SAP-based scaffold for neoglycoprotein synthesis

The native SAP had to be reengineered for use as a protein scaffold. Most importantly a handle had to be introduced which would allow the attachment of the glycoconjugate ligands. With the unknown stability of the SAP mutants and particularly an unknown territory of expressing this protein in bacteria rather than the isolation from human serum, the same manipulations used introduce an azide into the CTB pentamer may be much more difficult. With this in mind a different method was proposed for ligations using a recognition sequence for the transpeptidase enzyme sortase.³⁰⁶ Sortase is an enzyme responsible for the anchoring of proteins to cell wall of gram-negative bacteria, and it has

been utilised *in vitro* as a technology for protein modification.^{307, 308} Sortase has been used for many applications in our lab³⁰⁹⁻³¹², here we wanted to apply this method to introduce an C-terminal azide by ligation of a short peptide to a five amino acid C-terminal sortase recognition sequence containing LPETG (scheme 6.1).³¹³



Scheme 6.1: Proposed modification of the LPETGASHHHHHH C-terminal extension of the mutant SAP protein using sortase A enzyme. Transpeptidation of the mutant SAP removes the His-tag and adds a short peptide to leave a C-terminal azidolysine.

When redesigning the protein, we also had to consider the calcium binding and the proteins propensity to self-aggregate in a calcium-dependent manner.³¹⁴ With this in mind the protein sequence was altered to remove the ability of SAP to bind calcium through a E167A mutation.³¹⁵ A final thought was to introduce an N-terminal threonine residue, to allow N-terminal modification to be performed should this be a future requirement. A pET-11a vector was ordered from Genscript harbouring the gene for the reengineered SAP, the gene coded for SAP with a C-terminal LPETGASHHHHHH extension giving a cleavable histidine purification tag and sortase recognition sequence. A mutation was introduced into the genetic of the Glu 167 residue to Ala, removing the calcium dependent self-aggregation.

6.3 Expression of recombinant SAP

The Genscript plasmid was transformed into two different *E. coli* expression cell lines C41 (DE3) and BL21(DE3). Overexpression of SAP into inclusion bodies was observed in both cell lines, although C41 showed slightly better expression when the bands were analysed by densitometry (data not shown). The SAP protein was isolated from the inclusion bodies^{316, 317} following a general *in vitro* refolding protocol for proteins with

disulfide bonds. The inclusion bodies were solubilised in 8 M urea, prior to purification by Ni-NTA affinity chromatography exploiting the C-terminal his-tag on each monomer. SDS-PAGE gel following expression and purification of SAP from a 800 ml culture, showed a very high levels of overexpression (figure 6.3).

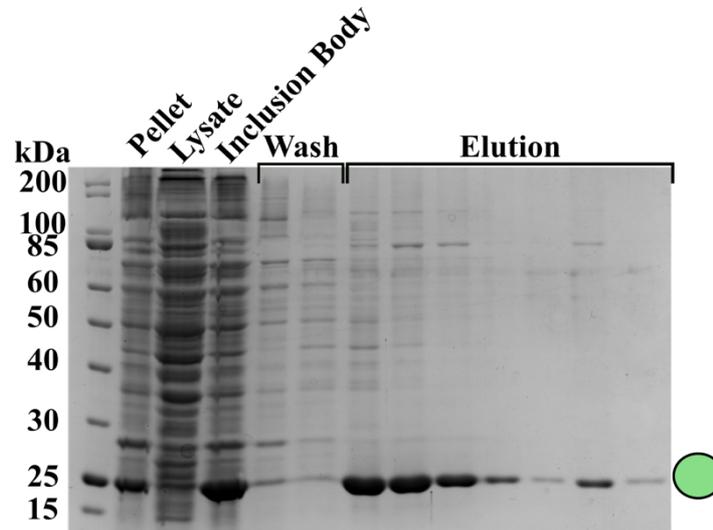


Figure 6.3: SDS-PAGE gel showing the expression of SAP isolating insoluble protein by solubilisation of inclusion bodies and purification by Ni-NTA.

The refolding protocol works by breaking the disulfide bonds and allowing them to reform with the protein refold around the formation of the correct disulfide bridges. With SAP only having a single disulfide per monomer, it is likely that the disulfide bond forms upon translation of the protein and aggregation is due to the bacteria not having the correct cellular machinery to aid in successful folding.

To investigate the need for disulfide exchange in the refolding of SAP, two refolding methods were trialled concurrently. The first followed the disulfide exchange refolding protocol and the second involved sequential dialysis in order to gradually decrease the denaturant concentration. Refolding involving disulfide exchange was a long, intricate process involving many days stirring in refolding buffer and it was important that the unfolded protein was rapidly diluted from a very high concentration into the refolding buffer. This rapid dilution left the refolded protein in a very large volume taking many hours to concentrate the folded protein back to workable concentrations. Sequential dialysis slowly removing urea from buffer until no denaturant remained in the buffer, proved a much simpler method. Despite this being a long process, the smaller volume made for easier handling of the protein across fewer concentration steps.

Both methods showed no aggregation of the protein during refolding, although the sequential dialysis required the total volume to be above 20 ml to prevent aggregation. This was promising as no aggregation of misfolded monomeric SAP shows the

recombinant protein is capable of refolding, with or without reformation of the disulfide bridges. One would then expect the pentamer to quickly form upon folding of the monomer units, with the monomeric proteins not known to be stable in solution. Trouble arose when concentrating the refolded protein. Despite taking great care it was found the protein had a propensity to aggregate particularly when concentrated by centrifugal ultrafiltration.

There are many reasonable explanations for this self-aggregation. In humans, SAP is modified with a large N-linked glycan on each protomer, however, recombinant expression in *E. coli* results in no glycosylation and the lack of these glycans could significantly alter the solubility of the protein. Figure 6.4 reveals the amino acid composition of SAP shown as hydrophilic (blue) or hydrophobic (red) residues, the SAP surface is largely hydrophobic and dimerisation of SAP pentamers would mask these large hydrophobic regions, improving the protein solubility. After removing the calcium binding capabilities and therefore the induced dimerisation, these hydrophobic regions remain solvent exposed and upon concentration self-aggregation may be favoured. These two features are disadvantages of expressing mutant and recombinant mammalian proteins.

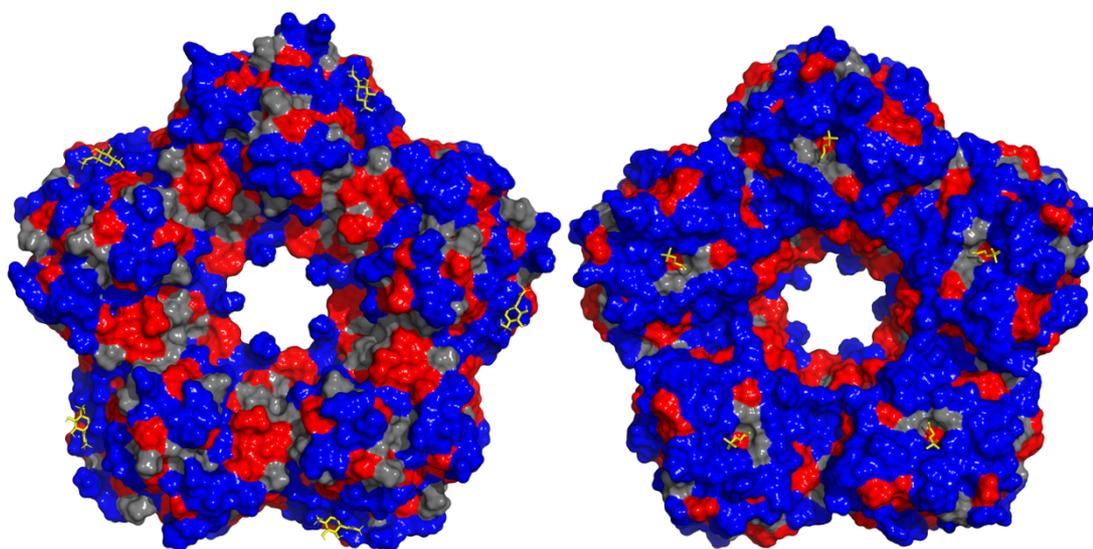


Figure 6.4: A comparison between the hydrophilic (blue) and hydrophobic (red) residues on the surface of the two faces of the SAP pentamer. Images rendered from the atomic co-ordinates 3KQR in the PDB.

This preliminary data for the expression of SAP shows promise with a small amount of soluble protein obtained. With further optimisation of the refolding protocol the yield of soluble protein may be improved, avoiding aggregations at higher concentrations needed for further modifications with sortase and SPAAC. CRP should also be investigated as a potential scaffold as expression of WT CRP has been carried out in *E.*

coli and the protein is not known to be glycosylated, therefore, could pose a better candidate for purpose of neoglycoprotein synthesis.

6.4 Conclusion

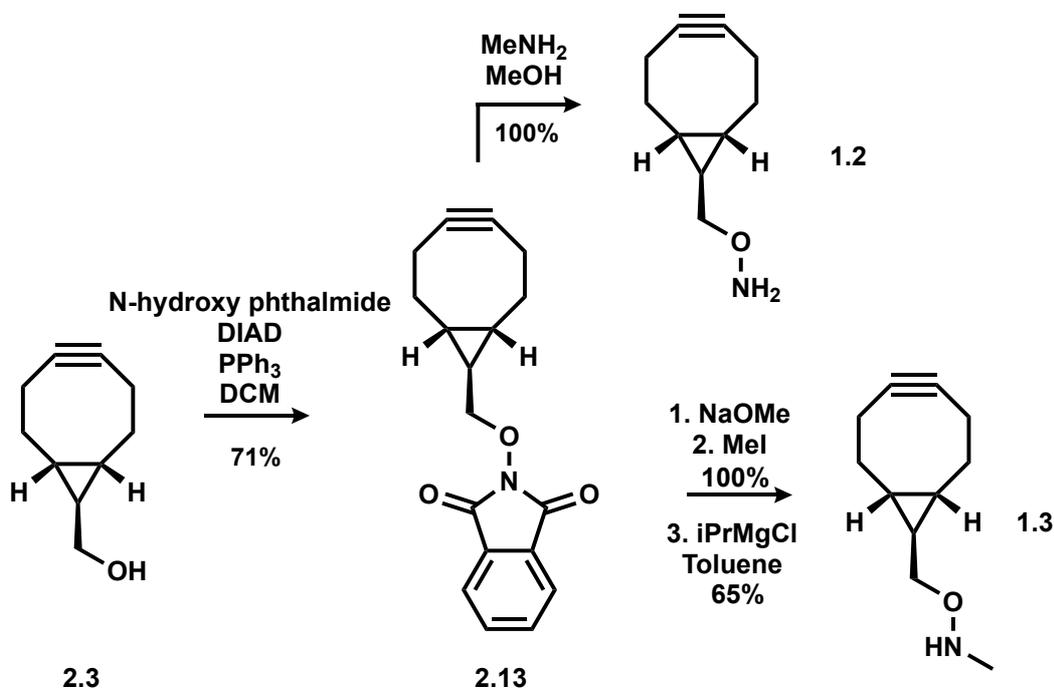
A final objective for this project, is to move away from pentameric bacterial proteins as scaffolds for the neoglycoprotein-based inhibitors, to a less immunogenic pentameric scaffold. As discussed, the pentraxin family of proteins were identified as potential candidates, with CRP and SAP identified as two scaffolds of interest. SAP was chosen as the first scaffold to look at, with modelling showing that a C-terminal extension and sortase ligation the resulting neoglycoproteins are capable of extending the Gb3 ligands to each of the three binding sites of the VTB-subunit. After designing a plasmid for the expression of a mutant SAP, we attempted to express the protein recombinantly, denaturing the protein and following a refolding protocol. The mutant SAP protein was successfully overexpressing in *E. coli* and isolated from inclusion bodies denaturing the protein before purification by affinity chromatography. Refolding of the protein was then performed using a simple refolding procedure involving the removal of the denaturant by sequential dialysis. The denaturant was successfully removed without any precipitation of the SAP, however, upon concentration the protein aggregated and was no longer soluble. Further investigation into the refolding procedure and in particular the methods for concentrating the protein without causing aggregation

Chapter 7 Conclusions and future studies

The primary aim of project was to expand method of producing synthetic neoglycoprotein to investigate how the inhibitory potential of multivalent neoglycoprotein-based inhibitors of bacterial toxins, is affected by changes in linker length and the position of glycans, Within this aim, a specific objective was to develop new chemicals tools to expand upon the current methods for synthetic glycosylation in which an unprotected oligosaccharide can be functionalised, and that glycoconjugate then be used to attach the glycan to a protein scaffold.

7.1 Bifunctional bioorthogonal linkers

It was proposed that a heterobifunctional linker could be used to achieve this aim, using one reactive group to derivatise reducing sugars, and the second reactive group for glycan attachment to the protein. Scheme 7.1 summarises the novel synthetic route that has been developed, to install a second bioorthogonally reactive functional group into the bicyclononyne strained alkyne scaffold. After following the literature synthesis to BCN alcohol **2.16**, Mitsunobu chemistry was used to install a protected oxyamine reactive group in the form of an N-hydroxyphthalimide. Different deprotection strategies were designed to produce a primary and secondary oxyamine, linkers **1.2** and **1.3** which were synthesised by complete deprotection or partial deprotection of the imide, respectively.



Scheme 7.1: A summary of the novel synthetic route to linkers **1.2** and **1.3** from the reported BCN alcohol through a protected phthalimide and divergent deprotection strategies of a single strained alkyne precursor.

The heterobifunctional linkers **1.2** and **1.3** have been designed specifically for the functionalisation of unprotected oligosaccharides and glycosyl azides, leaving a second reactive group for attachment to a protein. The result combined the benefits of bioorthogonal chemistry, site-selective protein modification, with fast reaction kinetics for effective covalent attachment of carbohydrates to proteins. The synthesis of this linker and its use in glycosylation, is one method of addressing the compounded endeavour in accessing homogeneous glycoproteins for biological study and practical applications. The variations of -ONH₂ and -ONHMe analogues allow for open and closed ring forms. Extension of the linker length could easily be carried by PEGylation of the BCN alcohol and conversion to the oxyamine using the same synthetic steps, allowing this linker to be used in the design of neoglycoproteins of broader applications.

7.2 SPAAC and oxime reactive synthetic glycoconjugates

With linkers **1.2** and **1.3** we addressed the synthesis of reactive glycoconjugates from unfunctionalised oligosaccharides. Glycoconjugates containing a strained alkyne shown for reaction with azides (figure 7.2) were produced through oxime ligation to lactose (**3.3**) and GM1os (**3.7**), demonstrating the linkers' ability to functionalise simple and complex oligosaccharides. The addition of a BCN to the oligosaccharides, not only introduced a reactive group, but also acted as a purification tag allowing for efficient purification of the glycoconjugates using a 'capture-and-release' method of reverse phase chromatography. The versatility of having a second reactive group was shown by the reaction of linker **1.2** with lactosyl azide **3.11**, producing an oxyamine containing glycoconjugate **3.14**. With further optimisation of the methods described by Shoda *et al.*²⁷³, this could be applied to GM1os and for other complex glycans, as a direct method to glycosyl azides to produce oxime reactive glycoconjugates.

Glycoconjugates **3.3** and **3.14** were shown to be suitable substrates for enzymatic synthesis. Using the galactosyltransferase LgtC, a galactose was added to the lactose core of **3.3** and **3.14** creating the Gb3 structure. The enzymatic reactions were successful using the donor sugar UDP-Gal in a single enzyme reaction, and in a one-pot enzymatic cycle in which UDP-Gal was produced *in situ* from sucrose and UDP, presenting an option for a cheap and scalable synthesis. Both methods gave excellent conversion from the disaccharide to the trisaccharide product and products were obtained in very high isolated yields. This is an exciting prospect for the production of reactive glycoconjugates, if the

unnatural substrates can be used with a variety of glycosyltransferases this method could be capable of building up large and diverse glycan structures.

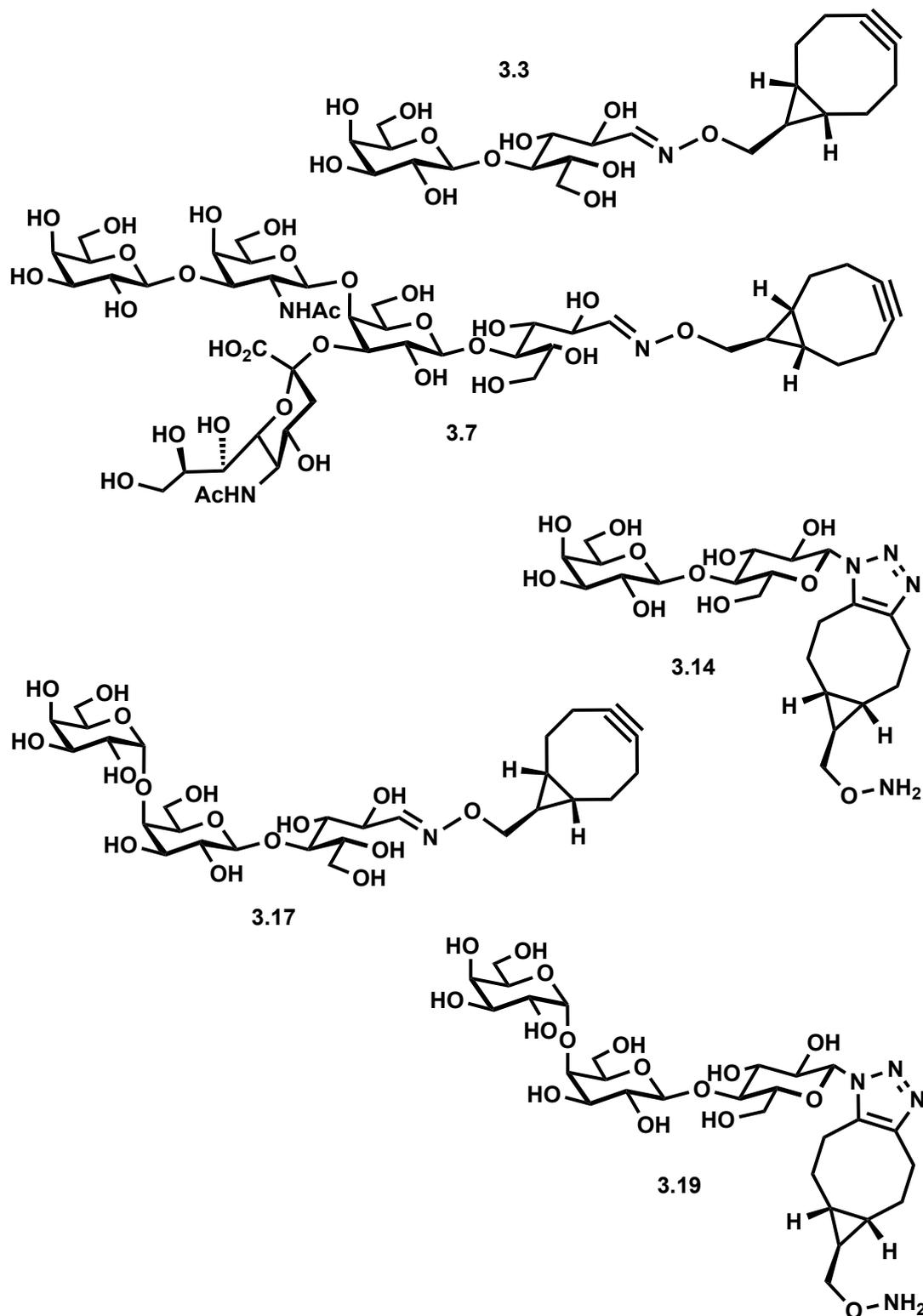


Figure 7.2: Structures of reactive glycoconjugates by reaction of linker 1.2. Glycoconjugates 3.3 and 3.7 synthesised by oxime ligation (highlighted in blue), 3.14 by SPAAC conjugation to glycosyl azide 3.11 (green) and 3.17, 3.18 produced by enzymatic extension of 3.3 and 3.14 respectively (brown).

7.3 Novel synthetic glycosylation of protein scaffolds

A pentameric protein scaffold was created which had five mutations in the WT CTB protein sequence, generating a mutant protein with no GM1 binding capabilities and a single methionine residue which became the site of an azide chemical handle. An azide group was then introduced on the non-binding face of CTB at position 43, by residue-specific incorporation of azidohomoalanine. The mutant protein was modified with carbohydrates **3.3**, **3.7**, and **3.17** using strain-promoted azide-alkyne cycloaddition as a new method for rapid, efficient and site-specific synthetic glycosylation. Figure 7.3A shows the synthetic route to three neoglycoproteins using SPAAC glycosylation, to produce lactose, Gb3 and GM1-based neoglycoproteins, each as a single glycoform. This method of synthetic glycosylation was unaffected by the complexity of the oligosaccharide and optimised conditions achieved full glycosylation of the protein in four hours.

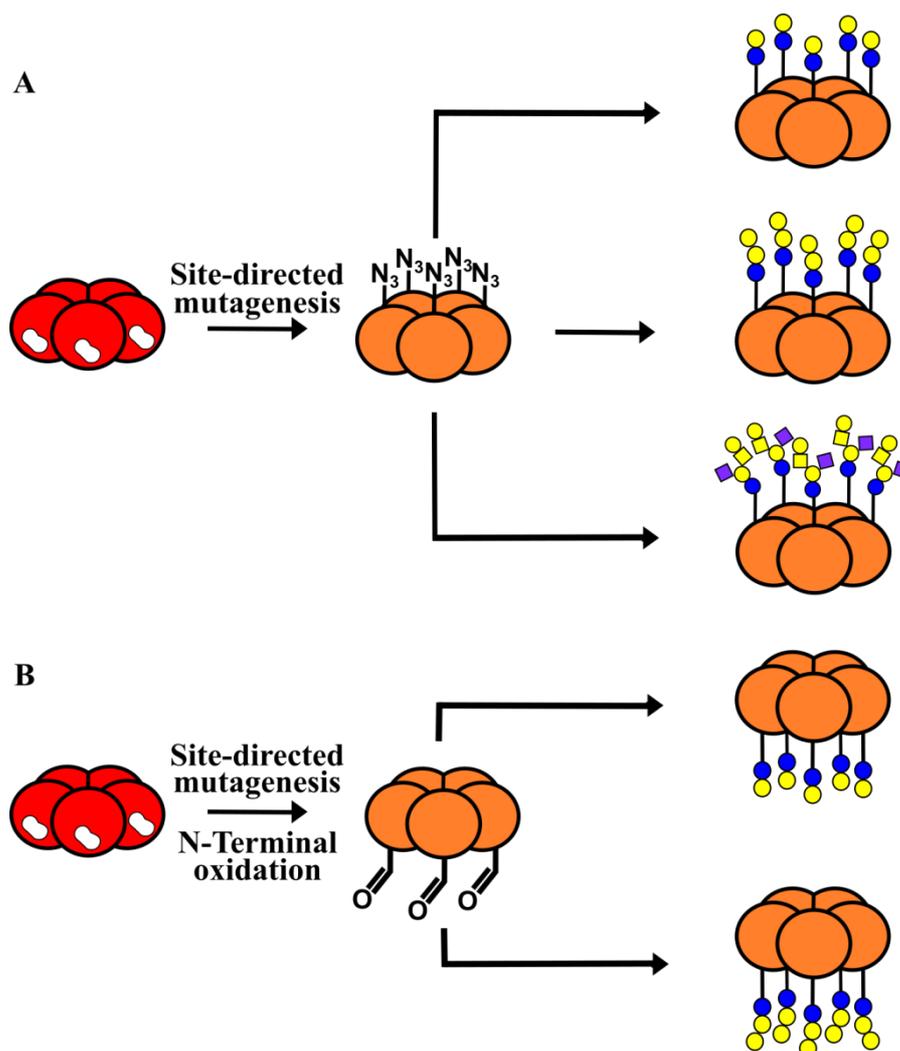


Figure 7.3: Summary of both synthetic glycosylation methods based upon the mutant CTB scaffolds W88E- N_3 , and to produce lactose, Gb3 and GM1 containing neoglycoproteins through A) SPAAC glycosylation,

functionalised on the non-binding face of CTB. B) Through oxime ligation of the N-terminus on binding face of CTB produced lactose and Gb3 containing neoglycoproteins.

A non-binding mutant of CTB produced by a single point mutation as outlined by in the work of Branson *et al.*¹²⁸ was also used to attach glycoconjugates **3.14** and **3.19** through oxime ligation producing lactose and Gb3-based neoglycoproteins appended at the N-terminus, as single glycoforms (figure 7.3B).

7.4 Bacterial toxin neoglycoprotein inhibitors

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 project. By combining the minimal length linker **1.2** and two different glycosylation methods, the carbohydrate ligands could be positioned on the non-binding and binding face of the CTB scaffold (figure 7.4). The five neoglycoproteins, based upon two different mutant CTB-subunit scaffolds, were designed as potential neoglycoprotein inhibitors for the CTB and VTB-subunits.

A functional ELLA inhibition assay was established for CTB, and the inhibitory potential of top-face modified GM1 neoglycoprotein was tested. **(GM1)W88E** was found to have a very low IC_{50} of 457 pM, which was a 231-fold increase on the monovalent GM1 oligosaccharide. We can only speculate that the modest decrease in potency from the inhibitor reported by Branson *et al.*¹²⁸ reported inhibitor (114 pM) is due to the much shorter linker length. Without a direct comparison to the neoglycoprotein with a short linker length and GM1 attached at the N-terminus, it is difficult to comment on the effect of the glycosylation site. The production of neoglycoproteins from direct functionalisation of isolated complex oligosaccharides and SPAAC/oxime synthetic glycosylation, is a much simpler route to neoglycoprotein inhibitors with little loss in potency.

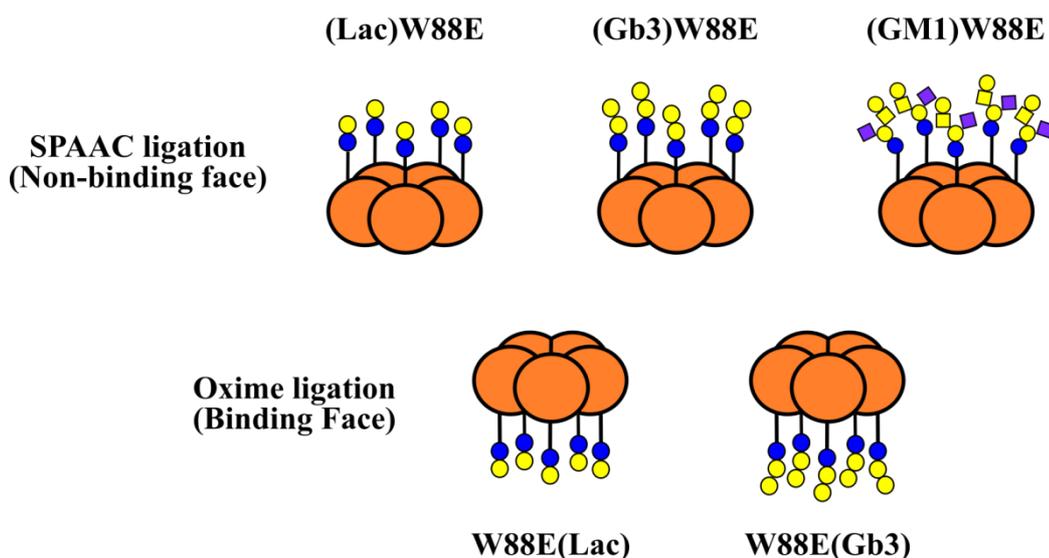


Figure 7.4: Cartoon representations of the lactose, Gb3 and GM1 neoglycoproteins produced as inhibitors for verotoxin and cholera toxin. Highlighted in red is the GM1 bottom face modified neoglycoprotein, which could not be synthesised for inhibition studies.

An ELLA inhibition assay for the verotoxin was not operational before the conclusion of the project, therefore, the four neoglycoprotein inhibitors modified with lactose and Gb3 could not be tested for inhibition of the verotoxin.

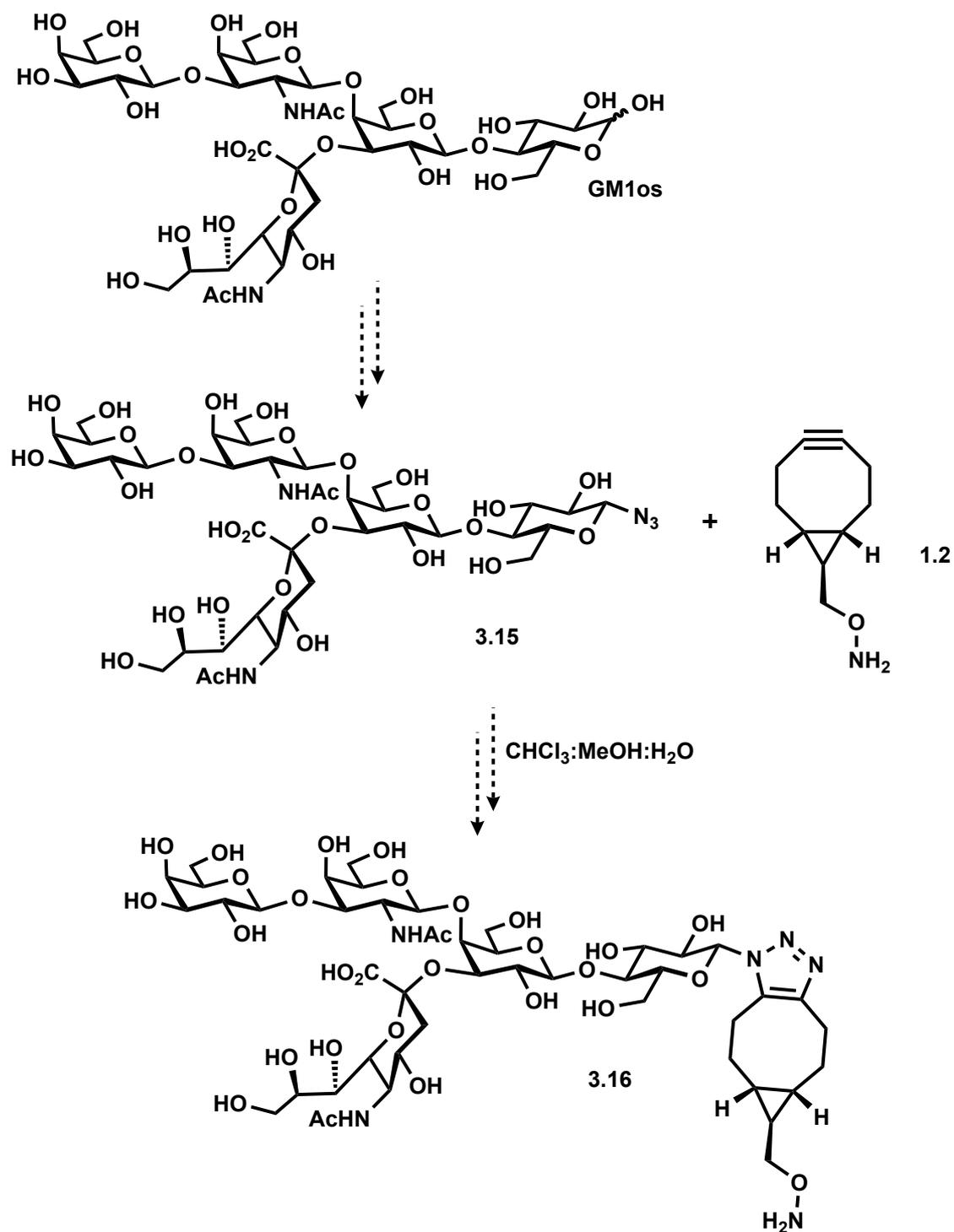
7.4.1 Engineering a non-bacterial pentameric protein scaffold

Assuming that the Gb3-neoglycoproteins targeting the verotoxin show successful inhibition of VTB binding, the final aim of this project was to develop non-bacterial-based protein scaffolds for use in the synthesis of neoglycoproteins. After identifying SAP as a potential candidate for a protein scaffold and *in silico* modelling a neoglycoprotein inhibitor based on SAP could target the three different Gb3 binding sites of VTB, a plasmid was engineered for a mutant SAP containing a C-terminal extension for further modification. Initial attempts to express this recombinantly proved successful, however, the problems could not be isolated due to problems with aggregation upon refolding of the protein. Potential explanations for the aggregation have been identified including the lack of glycosylation, and SAP often existing as a dimer due to its highly hydrophobic surface. With good overexpression achieved, optimisation of the refolding and protein concentration steps may yet yield soluble SAP for further protein modification. However, CRP could prove a more viable candidate in place of SAP, as it is near identical in size and structure, and more importantly has been expressed recombinantly.

7.5 Future studies

7.5.1 Synthesis and attachment of a GM1-oxyamine derivative for attachment to W88E

After obtaining inhibition data for one neoglycoprotein inhibitor of the cholera toxin, the data showed the reduction in linker length and changing the position of the GM1 ligands does not affect inhibitory potential; a conclusion on the optimal linker length and site of glycosylation could not be drawn. For this a second neoglycoprotein inhibitor with the reduced linker length and GM1 attached at the N-terminus is required. For the synthesis of this neoglycoprotein a reproducible synthesis of GM1-N₃ **3.15** (scheme 7.2) must be completed, and upon isolation of **3.15** and reaction with linker **1.2**, the GM1 derivative with a free oxyamine **3.16** can be used to glycosylate the W88E scaffold through oxime ligation. Once the neoglycoprotein is tested for inhibition of CTB adhesion, a direct comparison to the inhibitor reported by Branson *et al.*¹²⁸ can be drawn; from this data conclusions on the importance of linker length, and position of the glycan for optimal inhibition can be made. This will then feedback into the rationale for the design of such neoglycoprotein-based inhibitors for future targets and if the BCN linkers used in this project should be extended in length for better inhibition.



Scheme 7.2: Proposed synthesis of a GM1-oxyamine derivative 3.16 from GM1os, involving the conversion of GM1os to GM1-N₃ 3.15, followed by SPAAC reaction with linker 1.2.

7.5.2 Inhibition studies of neoglycoprotein-based inhibitors of the verotoxin.

To assess if the Gb3-neoglycoproteins produced in this project can act as inhibitors of the verotoxin, the inhibitory potential of the synthesised neoglycoproteins inhibitors (see section 7.4, figure 7.4) need to be assessed. To be able to test the lactose and Gb3 neoglycoproteins against VTB binding, a working ELLA inhibition has to be produced. The components of this assay have each been individually produced and although VTB capture on microtiter plates was confirmed only once, this must become consistently reproducible. Once this can be achieved the testing of neoglycoprotein inhibitors against VT can be performed (figure 7.5).

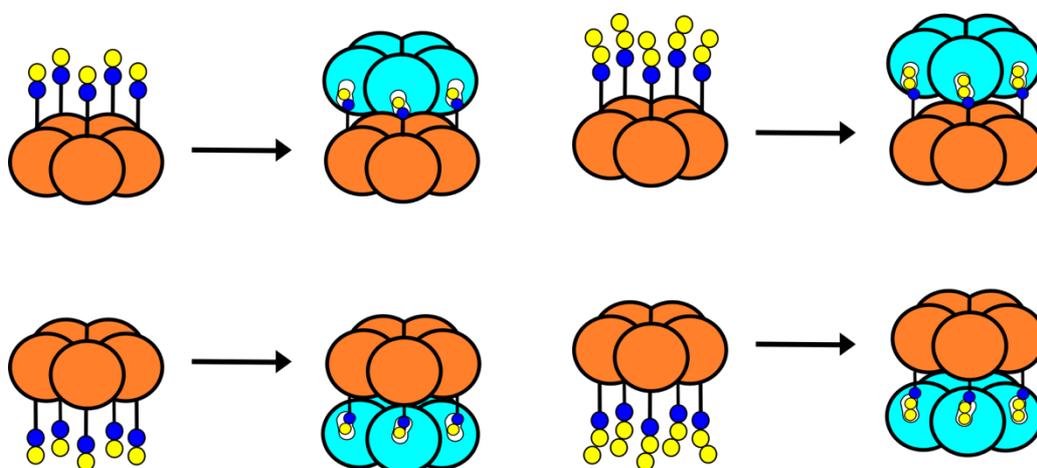


Figure 7.5: Cartoon representation of the four neoglycoproteins to be tested as inhibitors of the verotoxin B-subunit.

If these neoglycoproteins show inhibition of VTB binding to Gb3 coated plates, a further extension to this part of the project would be to target all potential Gb3 binding sites of the VTB-subunit by increasing the valency of a neoglycoprotein-based inhibitor to 15 Gb3 ligands on the pentameric scaffold (figure 7.6). This work has been ongoing within the group by developing trivalent glycopeptides which has been carried out by Dr Vajinder Kumar, a post-doctoral researcher within the lab. These trivalent glycopeptides have recently been appended to produce CTB-based neoglycoprotein through oxime ligation at the N-terminus. A derivative of the trivalent glycoconjugates containing a BCN has been produced and is awaiting attachment to N₃-W88E CTB scaffold using the SPAAC glycosylation method.

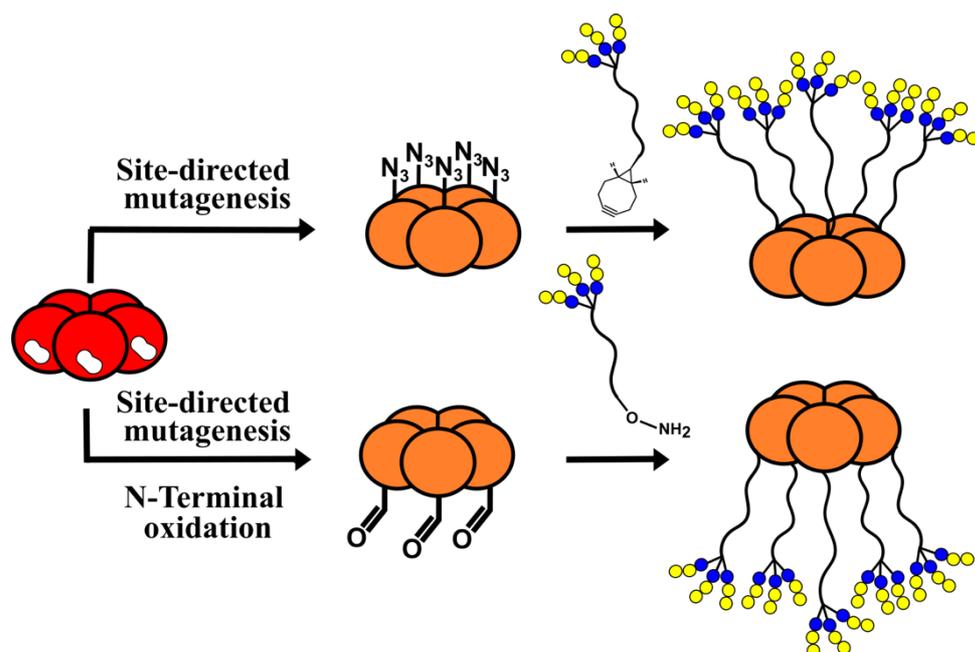


Figure 7.6: Cartoon representation showing the synthesis of two multivalent neoglycoprotein inhibitors with 15 copies of the Gb3 trisaccharide to target all potential binding sites of the verotoxin B-subunit.

These neoglycoproteins will then be tested as inhibitors of the verotoxin, with the intention of targeting all known Gb3 binding sites on the VTB subunit. Inhibition data for neoglycoprotein inhibitors would be interesting to observe, if the pre-ordered design combined with high valency boosts inhibitory potential against the verotoxin. It is possible that a more sensitive assay may be required to obtain good inhibition data for such inhibitors, if the inhibitors show dramatic increases in potency, the likes of which have been observed for the neoglycoprotein-based inhibitors of the cholera toxin.

7.5.3 Expression of a mutant human pentameric scaffold

To move towards a human pentameric scaffold for neoglycoprotein-based inhibitors as discussed in chapter six, further investigation is needed to determine which pentraxin, CRP or SAP, is the most suitable for expression in *E. coli* and furthermore, which of these proteins is the easiest to handle for further reactions with sortase. This will involve attempts to optimise the refolding procedure currently used in the SAP expression, with strict control over the concentration process to prevent aggregation. With CRP having already been expressed in *E. coli* and available from some commercial sources, a gene for a mutant CRP with the same C-terminal extension as SAP will be purchased and following known procedures for CRP expression, the mutant CRP will then be expressed in *E. coli*.

After determining the most suitable protein for modification with glycan derivatives, optimisation of the sortase ligation reactions will occur. During this the optimal

concentration of pentameric scaffold and peptide will be determined, to avoid large amount of hydrolysis products which arise from nucleophilic attack of water on the thioester intermediate generated when the protein is attached to the enzyme. This reaction will then be used to attach an azide to the protein by addition of a short peptide containing azidolysine.

Finally, once a human pentameric scaffold bearing a C-terminal azide is in hand, SPAAC can then be used to append Gb3 ligands to the scaffold as outlined in chapters four and five. The resulting neoglycoproteins can be tested for inhibition of the verotoxin both *in vivo*, then compared to the bacterial protein-based analogues. If successful inhibition is observed, then further studies could be carried out *in vitro* to determine the efficacy and if this approach is a viable therapeutic option for treatment of verotoxin infection. The neoglycoproteins both bacterial- and non-bacterial-based should be subjected to immunogenicity testing, and any effect these neoglycoproteins would have on the immune system, and if moving to a non-bacterial protein scaffold could be used a therapeutic option.

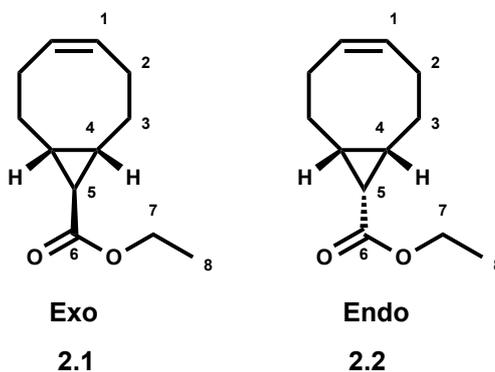
Chapter 8 Experimental

Chemicals were purchased from commercial suppliers Sigma Aldrich, Merck, Alfa Aeser, Fisher, Fluorochem. All solvents were HPLC grade purchased from Fisher scientific or Sigma Aldrich. Anhydrous solvents were obtained using an Innovative technology solvent purification system. Distilled water was purified by ELGA PURLAB classic. InertSep® SLIM C18-C reverse SPE phase cartridges were purchased from BGB analytik. NMR analysis was performed using Bruker Avance 500 MHz or 500 MHz cryoprobe. LC-MS analysis was performed on Bruker AmaZon X series LC-MS spectrometer. HILIC-LC-MS was performed on the same spectrometer using a Kinetex® 2.6 µm HILIC 100 Å, LC column 30 x 2.1 mm purchased from Phenomenex. HRMS was performed on a Bruker maXisimpact spectrometer.

NMR assignment abbreviations: s – singlet, d – doublet, t – triplet, q – quartet, m – multiplet, dd – doublet of doublets, dt – doublet of triplets, b – broad.

8.1 Linker synthesis

Compound 2.1/2.2: Ethyl (Z)-bicyclo[6.1.0]non-4-ene-9-carboxylate



1,5-cyclooctadiene (19.6 ml, 160 mmol, 8 eq) and rhodium (II) acetate (380 mg, 0.86 mmol, 0.043 eq) were dissolved in DCM (10 ml) and stirred under nitrogen for 15 minutes. A solution of ethyl diazoacetate (2.4 ml, 20 mmol, 1 eq) in DCM (10 ml) was added dropwise, over the course 3 hours. The mixture was allowed to stir for three days, after which the mixture was concentrated *in vacuo* to yield a blue oil. The oil was passed through a silica plug eluting in neat hexane, followed by 10% ethyl acetate to yield a mixture of diastereoisomers as a colourless oil. The two stereoisomers were separated by flash column chromatography eluting in neat toluene to give the *exo* isomer (1.95 g) and *endo* isomer (0.71 g) in a ratio of 2.2:1, respectively and an overall combined yield of 2.266 g, 89%.²⁵⁶

Exo

$R_f = 0.06$ (1% Ethyl acetate in hexane).

$^1\text{H NMR}$ (500 MHz; CDCl_3) δ : 1.18 (1H, H_5 t, $J = 4.6$ Hz), 1.25 (3H, t, $J = 7.2$ Hz, H_8), 1.44-1.52 (2H, m, H_3), 1.54-1.59 (2H, m, H_4), 2.05-2.12 (2H, m, $\text{H}_{2,3'}$), 2.16-2.23 (2H, m, H_2), 2.27-2.34 (2H, m, $\text{H}_{2',3'}$), 4.04 (2H, q, $J = 7.2$ Hz, H_7), 5.64 (2H, bdt, $J = 4.12$, H_1).

$^{13}\text{C NMR}$ (125 MHz; CDCl_3) δ 14.40 (C_8), 26.81 (C_2), 27.89 (C_4), 28.04 (C_5), 28.42 (C_3), 60.3 (C_7), 129.9 (C_1), 174.5 (C_6).

Endo

$R_f = 0.24$ (1% Ethyl acetate in hexane).

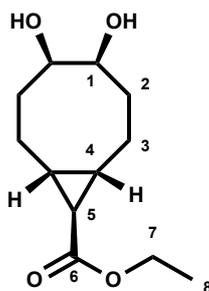
$^1\text{H NMR}$ (500MHz; CDCl_3) δ 1.26 (3H, t, $J = 7.2$ Hz, H_8), 1.39 (2H, m, H_3), 1.71 (1H, t, $J = 8.8$ Hz, H_5), 1.83 (2H, m, H_4), 2.06 (2H, m, H_2), 2.20 (2H, m, H_3), 2.51 (2H, m, H_2), 4.12 (2H, q, $J = 7.2$ Hz, H_7), 5.61 (2H, bdt, $J = 4.3$ Hz, H_1).

$^{13}\text{C NMR}$ (125 MHz; CDCl_3) δ 14.2 (C_8), 21.26 (C_5), 22.67 (C_3), 24.19 (C_4), 27.09 (C_2), 59.7 (C_7), 129.5 (C_1), 172.1 (C_6).

HRMS - $\text{C}_{12}\text{H}_{18}\text{O}_2 + \text{Na}$ requires 217.1204 Da. Measured m/z $[\text{M} + \text{Na}]^+ = 217.1198$ Da. err = 0.3 ppm.

IR ($\nu_{\text{max}} / \text{cm}^{-1}$) – 1151 (C-O), 1181 (C=C, bend), 1718 (C=O), 2979 (C-H).

Compound 2.4: Ethyl (1R,4R,5S,8S,9r)-4,5-dihydroxybicyclo[6.1.0]nonane-9-carboxylate



2.4

Compound **2.1** (500 mg, 2.58 mmol, 1 eq) was dissolved in 8:1 acetone/water mixture (4.6 ml:0.57 ml, 0.5 M), to which *N*-methylmorpholine,*N*-oxide (NMO) (604 mg, 5.16 mmol, 2 eq) was added. Once NMO was completely suspended, OsO₄ in H₂O (0.82 ml, 0.129 mmol, 0.05 eq) was added slowly to the reaction vessel. The reaction was followed by TLC and was shown to be complete after ≈3 hours, showing the starting material had dissipated. The reaction was then quenched with saturated aq. sodium metabisulphite, and the resulting precipitate was removed by filtration. The pH of the filtrate was then adjusted to pH 5 using 1 M HCl, and the product extracted with ethyl acetate (3×20 ml). The combined organic layers were concentrated *in vacuo* to yield a white solid (439 mg, 75%). The mixture of diastereoisomers could not be separated.

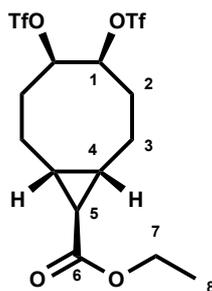
R_f = 0.06 (10% EtOAc in hexane).

¹H NMR (500 MHz; DMSO-d₆) δ 1.02-1.07 (2H, m, **H_{4,5}**), 1.17 (3H, t, *J* = 7.1 Hz, **H₈**), 1.23-1.33 (1H, m, **H₄**), 1.33-1.38 (2H, m, **H_{2,3}**), 1.53-1.61 (2H, m **H_{2,3}**), 1.66-1.72 (2H, m, **H_{2'}**), 1.73-1.83 (3H, m, **H_{2',3'}**), 1.88 (1H, dd, *J* = 14.49, 3.29, **H_{3'}**), 3.68 (1H, b dd, major isomer, **H₁**), 3.87 (1H, br dd, minor isomer, **H₁**), 4.01 (2H, **H₇**, q, *J* = 7.2 Hz), 4.31 (1H, br dd, major isomer, **H₁**), 4.39 (1H, br dd, minor isomer, **H₁**).

¹³C NMR (125 MHz, DMSO-d₆) δ 14.3 (**C₈**), 21.6 (**C₄** minor isomer), 23.4 (**C₄** major isomer), 26.8 (**C₂** minor isomer), 27.0 (**C₂** major isomer), 27.9 (**C₅**), 32.9 (**C₃** major isomer), 33.5 (**C₃** minor isomer), 60.5 (**C₇**), 75.5 (**C₁** minor isomer), 76.4 (**C₁** major isomer), 172.6 (**C₆**).

HRMS - C₁₂H₂₀O₄+Na requires 229.1434 Da. Measured *m/z* [M+Na]⁺ = 229.1433 Da. err = 0.5 ppm.

Compound 2.5: Ethyl (1*R*,4*R*,5*S*,8*S*,9*r*)-4,5-bis(((trifluoromethyl)sulfonyl)oxy)bicyclo[6.1.0]nonane-9-carboxylate



2.5

Compound **2.1** (100 mg, 0.515 mmol, 1 eq) and pyridine (0.1 ml, 1.545 mmol, 3 eq) were dissolved in DCM (5 ml, 0.1 M), placed under a nitrogen atmosphere and cooled to 0 °C. Distilled triflic anhydride (0.17 ml, 1.03 mmol, 2 eq) was then added slowly to the solution an instant colour change from colourless to yellow was observed. The reaction was left to stir for 10 minutes and monitored by TLC and the product was extracted with EtOAc (3 x 10 ml) and concentrated *in vacuo* to yield a yellow oil (209 mg, 83%).

R_f = 0.36 and 0.24 (10% EtOAc in hexane).

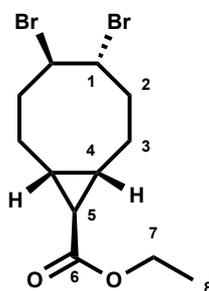
$^1\text{H NMR}$ (500MHz; CDCl_3) δ 0.98-1.10 (2H, m, H_3), 1.20 (1H, t, $J = 4.4$ Hz, H_5), 1.25 (4H, t, $J = 7.3$ Hz, $\text{H}_{4,8}$), 1.35-1.59 (2H, m, $\text{H}_{3',4}$), 1.93-2.06 (1H, m, H_2), 2.06-2.16 (2H, m, $\text{H}_{2,3'}$), 2.16-2.30 (3H, m, $\text{H}_{2,2'}$), 4.11 (2H, q, $J = 7.2$ Hz, H_7), 5.21 (2H, br dd, H_1), 5.42 (2H, dd, $J = 8.56, 4.1$ Hz, H_1).

$^{13}\text{C NMR}$ (125 MHz; CDCl_3) δ 14.22 (C_8), 25.43 (C_4), 26.53 (C_5), 27.44 (C_2), 29.70 (C_3), 60.66 (C_7), 89.85 (C_1), 117.16 (C_9) (q, $J = 319.1\text{Hz}$), 172.60 (C_6).

HRMS - $\text{C}_{14}\text{H}_{19}\text{F}_6\text{O}_8\text{S}_2+\text{Na}$ requires 515.0240 Da. Measured m/z $[\text{M}+\text{Na}]^+ = 515.02408$ Da. err = 1.2 ppm.

IR - ($\nu_{\text{max}} / \text{cm}^{-1}$)-890 (S-O-C), 1138 (S=O), 1200 (S=O), 1242 (C-O) 1412 (C-F) , 1718 (C=O), 2935 (C-H).

Compound 2.6: Ethyl (1R,4R,5R,8S,9R)-4,5-dibromobicyclo[6.1.0]nonane-9-carboxylate



2.6

Compound **2.1** (600 mg, 3.13 mmol, 1 eq) in anhydrous DCM (11.6 ml) was added to a flask under a nitrogen atmosphere, and cooled to 0 °C. A solution of bromine (575 mg, 3.76 mmol, 1.2 eq) in DCM (4 ml) was added slowly, resulting in the solution becoming brown. The reaction was then left to stir for 10 minutes and reaction progress was

observed by TLC. The reaction was then quenched with saturated aq. sodium thiosulfate (50 ml), and the product was extracted into DCM (3×20 ml). The combined organic layers were washed with distilled H₂O, dried over MgSO₄ and concentrated *in vacuo* to yield a white solid (804 mg, 73% yield).

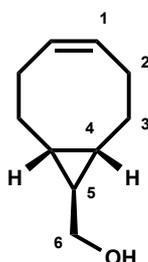
R_f = 0.09 (5% EtOAc in hexane).

¹H NMR (500MHz; CDCl₃) δ 1.20 (1H, t, *J* = 4.2 Hz, **H₅**), 1.26 (3H, t, *J* = 6.9 Hz, **H₈**), 1.37-1.53 (2H, m, **H₄**) 1.62-1.68 (2H, m, **H₃**), 1.69-1.76 (2H, m **H₃**), 2.07-2.15 (3H, m, **H_{2,3'}**), 2.26-2.34 (1H, m, **H_{2'}**), 2.60-2.67 (1H, m, **H_{3'}**), 2.69-2.77 (1H, m, **H_{2'}**), 4.12 (2H, q, *J* = 7.2 Hz, **H₇**), 4.76-4.85 (2H, m, **H₁**).

¹³C NMR (125 MHz; CDCl₃) δ minor diastereoisomer :14.3 (**C₈**), 23.78 (**C₄**), 22.98 (**C_{4'}**), 25.66 (**C₅**), 27.75 (**C₃**), 27.99 (**C_{3'}**), 34.30 (**C₂**), 34.80 (**C_{2'}**), 52.24 (**C₁**), 55.7 (**C_{1'}**), 60.38 (**C₇**), 173.79 (**C₆**).

HRMS - C₁₂H₁₈Br₂O₂+NH₄ requires 370.0011 Da. Measured *m/z* [M+NH₄]⁺ = 370.0005 Da. err = 0.6 ppm

Compound 2.8: ((1*R*,8*S*,9*r*,*Z*)-bicyclo[6.1.0]non-4-en-9-yl)methanol



Exo

2.8

Compound **2.1** (1.094 g, 5.63 mmol, 1 eq) was placed in a round bottom flask with anhydrous DCM (28 ml, 0.2 M) under a nitrogen atmosphere. The solution was cooled to 0 °C and 1 M lithium aluminium hydride (LiAlH₄) in THF (11.3 ml, 11.3 mmol, 2 eq) was added dropwise to the solution. The reaction was monitored by TLC until completion (≈1 hour) and the reaction was then quenched with sodium sulphate decahydrate. The mixture was acidified with 1 M HCl, and the product was extracted with DCM (3 × 20 ml), and the combined organic layers were dried over MgSO₄ and concentrated *in vacuo* to yield a clear oil (857 mg, quantitative).²⁵⁶

$R_f = 0.71$ in 20% EtOAc in hexane.

Endo

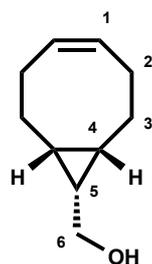
$^1\text{H NMR}$ (500MHz; CDCl_3) δ 0.97-1.05 (2H, m, H_4), 1.09-1.16 (1H, m, H_5), 1.31 (1H, bs, OH), 1.37-1.45 (2H, m, H_3), 1.94-2.02 (2H, m, H_3'), 2.06-2.14 (2H, m, H_2), 2.32-2.40 (2H, m, H_2'), 3.73 (2H, d, H_6 , $J = 7.6$ Hz), 5.63 (2H, t, H_1 , $J = 4.0$ Hz).

$^{13}\text{C NMR}$ (125 MHz; CDCl_3) δ 19.15 (C_4), 20.93 (C_5), 24.03 (C_3), 27.81 (C_2), 60.45 (C_6), 129.88 (C_1).

HRMS - Could not be obtained.

IR ($\nu_{\text{max}}/\text{cm}^{-1}$) - 1015 (C-O), 1094 (C=C, bend), 2911 (C-H), 3316 (O-H).

The *endo* isomer was prepared using the same procedure from ethyl (1*R*,8*S*,9*S*,*Z*)-bicyclo[6.1.0]non-4-ene-9-carboxylate (compound **2.2**) in quantitative yield.



Endo

Endo

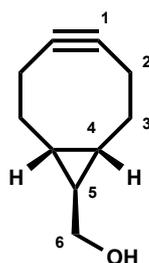
$^1\text{H NMR}$ (500MHz; CDCl_3) δ 0.97-1.05 (2H, m, H_4), 1.09-1.16 (1H, m, H_5), 1.31 (1H, bs, OH), 1.37-1.45 (2H, m, H_3), 1.94-2.02 (2H, m, H_3'), 2.06-2.14 (2H, m, H_2), 2.32-2.40 (2H, m, H_2'), 3.73 (2H, d, $J = 7.6$ Hz, H_6), 5.63 (2H, t, $J = 4.0$ Hz, H_1).²⁵⁶

$^{13}\text{C NMR}$ (125 MHz; CDCl_3) δ 19.15 (C_4), 20.93 (C_5), 24.03 (C_3), 27.81 (C_2), 60.45 (C_6), 129.88 (C_1).

HRMS - Could not be obtained.

IR ($\nu_{\text{max}}/\text{cm}^{-1}$) - 1015 (C-O), 1094 (C=C, bend), 2911 (C-H), 3316 (O-H).

Compound 2.3: ((1R,8S,9r)-bicyclo[6.1.0]non-4-yn-9-yl)methanol



Exo

2.3

Compound **2.8** (500 mg, 3.28 mmol, 1 eq) was dissolved in anhydrous DCM (16 ml, 0.2 M) under a nitrogen atmosphere. The solution was cooled to 0 °C and bromine (0.2 ml, 3.94 mmol 1.2 eq) was added dropwise until a brown colour persisted. The reaction was then allowed to stir at room temperature for 30 minutes. After this time the reaction was quenched with sat. aq. Na₂S₂O₃ (30 ml) and the organics extracted with DCM (3×20 ml). The combined organic layers were washed with H₂O (25 ml) and brine (25 ml), dried over MgSO₄ and concentrated *in vacuo* to yield an oil. The oil was then dissolved in anhydrous THF (33 ml) under nitrogen and cooled to 0 °C before adding 1 M KOt-Bu in THF (9.84 ml, 9.84 mmol, 3 eq). The solution was then heated to reflux for five hours. After this time the reaction was quenched with sat. aq. NH₄Cl (50 ml) and the organics extracted with DCM (3 x 50 ml). The combined organic layers were concentrated *in vacuo* to yield a brown oil. The crude product was purified by flash column chromatography eluting in 10% EtOAc in hexane to 20% EtOAc in hexane, to yield a clear oil (181 mg, 37%).²⁵⁶

R_f- 0.24 (30% EtOAc in hexane).

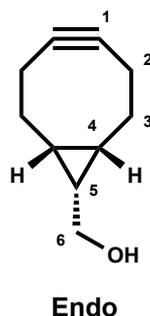
Exo

¹H NMR (500 MHz; CDCl₃) δ 0.59-0.65 (1H, m, **H₅**), 0.72-0.79 (2H, m, **H₄**), 1.35-1.43 (2H, m, **H₃**), 1.61 (1H, bs, **H_{OH}**), 2.00-2.08 (2H, m, **H_{2'}**), 2.11-2.19 (2H, m, **H₂'**), 2.23-2.32 (2H, m, **H_{3'}**), 3.45 (2H, d, *J* = 7.0 Hz, **H₆**).

¹³C NMR (125 MHz; CDCl₃) δ 21.47 (**C₄**), 22.59 (**C₂**), 27.32 (**C₅**), 33.43 (**C₃**), 67.15 (**C₆**), 98.80 (**C₁**).

HRMS - Could not be obtained

The *endo* isomer of **2.3** was prepared using the same procedure from ((1*R*,8*S*,9*S*,*Z*)-bicyclo[6.1.0]non-4-en-9-yl)methanol (*endo* isomer of compound **2.8**) in 37% yield.²⁵⁶



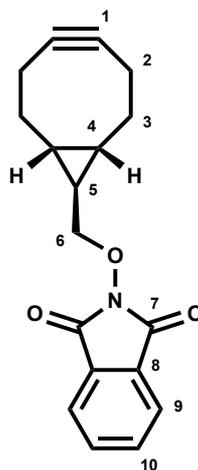
Endo

¹H NMR (500 MHz; CDCl₃) δ 0.81-0.91 (2H, m, **H**₄), 1.22-1.31 (1H, m, **H**₅), 0.44 (1H, bs, **OH**), 1.48-1.58 (2H, m, **H**₃), 2.10-2.28 (3H, m, **H**_{2,2',3'}), 3.65 (2H, d, *J* = 7.8, **H**₆).

¹³C NMR (125 MHz; CDCl₃) δ 20.00 (**C**₄), 21.37 (**C**₅), 21.50 (**C**₃), 29.03 (**C**₂) 59.88 (**C**₆), 98.89 (**C**₁).

HRMS - Could not be obtained.

Compound 2.13: 2-(((1*R*,8*S*,9*r*)-bicyclo[6.1.0]non-4-yn-9-yl)methoxy)isoindoline-1,3-dione



Compound **2.3** (80 mg, 1.6 mmol, 1eq), triphenylphosphine (154 mg, 0.59 mmol, 1.1 eq) and *N*-hydroxyphthalamide (96 mg, 0.59 mmol, 1.1 eq) were dissolved in anhydrous DCM (5 ml, reaction concentration 0.1 M) and placed under a nitrogen atmosphere. The mixture was cooled to 0 °C, at which point DIAD (0.11 ml, 0.59mmol, 1.1 eq) was added.

The reaction was left stirring and allowed to warm to room temperature. After four hours the mixture was concentrated *in vacuo* to yield a crude residue, which was purified by flash column chromatography eluting in 10% EtOAc in hexane to 20% EtOAc in hexane to yield a white crystalline solid (111 mg, 71% yield).

R_f - 0.42 (30% EtOAc in hexane).

Exo

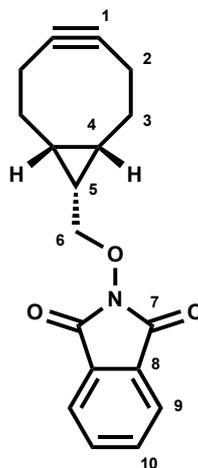
¹H NMR (500 MHz; CDCl₃) δ 0.78-0.88 (3H, m, **H_{4,5}**), 1.38-1.45 (2H, m, **H₃**), 2.14-2.20 (2H, m, **H₂**), 2.26-2.30 (2H, m, **H_{2'}**), 2.39-2.42 (2H, dd, **H_{3'}**, *J* = 13.8, 2.19 Hz), 4.15 (2H, d, **H₆**, *J* = 7.0), 7.73-7.78 (2H, m, **H₁₀**) 7.82-7.87 (2H, m, **H₉**).

¹³C NMR (125 MHz, CDCl₃) δ 21.28 (**C₂**), 22.66 (**C₅**), 23.03 (**C₄**), 33.15 (**C₃**), 82.34 (**C₆**), 98.73 (**C₁**), 123.46 (**C₁₀**), 128.98 (**C₈**), 134.46 (**C₉**), 163.74 (**C₇**).

HRMS – C₁₈H₁₇NO₃+Na requires 318.1101 Da. Measured *m/z* [M+Na]⁺ = 318.1101 Da. err = -0.2 ppm.

IR - (ν_{max}/cm⁻¹) – 1124 (C-O), 1720 (C=O).

The *endo* isomer of **2.13** was prepared from ((1*R*,8*S*,9*S*)-bicyclo[6.1.0]non-4-yn-9-yl)methanol using the same procedure in



Endo
2.13

Endo

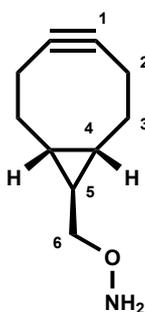
¹H NMR (500MHz; CDCl₃) δ 0.76-0.89 (2H, m, **H₄**), 1.25 (1H, t, **H₅**, $J = 7.0$) 1.34-1.42 (2H, m, **H₃**) 2.09-2.17 (2H, m, **H_{3,3'}**) 2.35 (6H, m, **H_{2,2'},H_{3'}**), 4.31 (2H, d, **H₆**, $J = 8.0$ Hz), 7.73-7.77 (2H, m, **H₁₀**) 7.82-7.86 (2H, m, **H₉**).

¹³C NMR (125 MHz, CDCl₃) δ 16.79 (**C₅**), 19.87 (**C₄**), 24.14 (**C₃**), 27.64 (**C₂**), 77.73 (**C₆**), 99.04 (**C₁**), 123.55 (**C₁₀**), 129.14 (**C₈**), 134.39 (**C₉**), 163.81 (**C₇**).

HRMS – C₁₈H₁₇NO₃+Na requires 318.1101 Da. Measured m/z [M+Na]⁺ = 318.1101 Da. err = -0.2 ppm.

IR - ($\nu_{\max}/\text{cm}^{-1}$) – 1124 (C-O), 1720 (C=O).

Compound 1.2: *O*-(((1*R*,8*S*,9*r*)-bicyclo[6.1.0]non-4-yn-9-yl)methyl)hydroxylamine



1.2

Compound **2.13** (*exo*) (100 mg, 0.34 mmol, 1 eq) was added to a anhydrous 2 M methanolic methylamine (0.15 ml, 1.7 mmol, 5 eq) was added. The reaction was shown

to be complete by TLC within 2 minutes, the product was used without further purification. (65 mg, quantitative).

$R_f = 0.64$ (50% EtOAc in hexane).

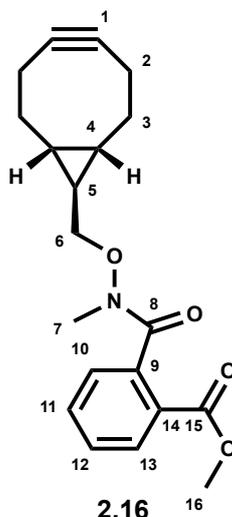
Exo

¹H NMR (500 MHz; CDCl₃) δ 0.50-0.70 (3H, m, **H_{4,5}**), 1.27-1.43 (2H, m, **H₃**), 2.07-2.43 (6H, m, **H_{2,2',3'}**), 3.56 (2H, d, $J = 7.0$, **H₆**), 5.22 (2H, bs, **NH₂**).

¹³C NMR (125 MHz, CDCl₃) δ 21.45 (**C₄**), 22.78 (**C₃**), 23.03 (**C₄**), 23.43 (**C₅**), 33.35 (**C₂**), 80.08 (**C₆**), 98.82 (**C₁**).

HRMS - C₁₀H₁₅NO+H requires 166.1223 Da. Measured m/z [M+H]⁺ = 166.1226 Da. err = 1.6 ppm.

Compound 2.16: Methyl 2-((((1R,8S,9r)-bicyclo[6.1.0]non-4-yn-9-yl)methoxy)carbamoyl)benzoate



Compound **2.13** (111 mg, 0.38 mmol, 1 eq) was dissolved in MeOH (1 ml). To the solution NaOMe (41 mg, 1.14 mmol, 3 eq) was added and the reaction was stirred for 30 minutes at room temperature until all starting material was consumed when observed by TLC (30% EtOAc in hexane). Methyl iodide (100 μ l, 1.52 mmol, 4 eq) was added to the solution, and the reaction was monitored by HRMS. Once the methylation reaction was deemed to be complete (\approx 1 hour) the reaction was quenched with H₂O (5 ml) and the product was extracted with EtOAc (3 x 10 ml). The combined organic layers were dried over MgSO₄ and concentrated *in vacuo* to yield a yellow residue (129 mg, 100%).

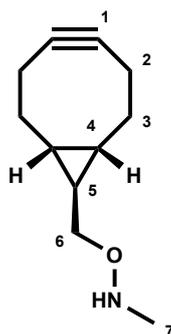
Exo

¹H NMR (500 MHz; CDCl₃) δ 0.74 (2H, bt, **H₄**), 0.89-1.09 (3H, m, **H_{3,5}**), 1.81 (2H, bd, **H_{3'}**), 2.000-2.07 (2H, m, **H₂**), 2.10-2.20 (2H, m, **H_{2'}**), 3.37 (3H, bs, **H₇**), 3.60 (2H, bd, $J = 7.8$ Hz, **H₆**), 3.90 (3H, s, **H₁₆**), 7.40-7.47 (2H, m, **H_{11,12}**), 7.57 (1H, dt, $J = 7.6, 1.0$ Hz, **H₁₀**), 7.97 (1H, d, $J = 7.6$ Hz, **H₁₃**).

¹³C NMR (125 MHz; CDCl₃) δ 17.05 (**C₅**), 20.42 (**C₄**), 21.82 (**C₂**), 28.98 (**C₃**), 34.65 (**C₇**), 52.97 (**C₁₆**), 71.74 (**C₆**), 99.05 (**C₁**), 127.85 (**C₁₁**), 128.76 (**C₉**), 129.26 (**C₁₀**), 130.13 (**C₁₃**), 132.68 (**C₁₂**), 137.93 (**C₁₅**), 166.62 (**C₈**), 172.10 (**C₁₅**).

HRMS - C₂₀H₂₃NO₄+H requires 342.1699 Da. Measured m/z [M+H]⁺ = 342.1697 Da. err = 0.8 ppm.

Compound 1.3: O-(bicyclo[6.1.0]non-4-yn-9-ylmethyl)hydroxylamine



1.3

Compound **2.16** (50 mg, 0.17 mmol, 1 eq) was dissolved in anhydrous toluene (1 ml) into an oven dried flask under a nitrogen atmosphere. The solution was cooled to $-78\text{ }^{\circ}\text{C}$ and a solution of 3M isopropyl magnesium bromide in diethyl ether (0.057 ml, 0.18 mmol, 3 eq) was added dropwise. The reaction was left to stir for 2 hours and allowed to warm to room temperature. The reaction was quenched with H_2O (10 ml) and the product extracted with EtOAc (3×10 ml). The combined organic layers were dried over MgSO_4 and concentrated *in vacuo* to yield a crude oil. The crude oil was purified by flash column chromatography eluting in 10% EtOAc in hexane to yield a clear residue (18 mg, 60%).

R_f - 0.15 (20% EtOAc in hexane).

Endo

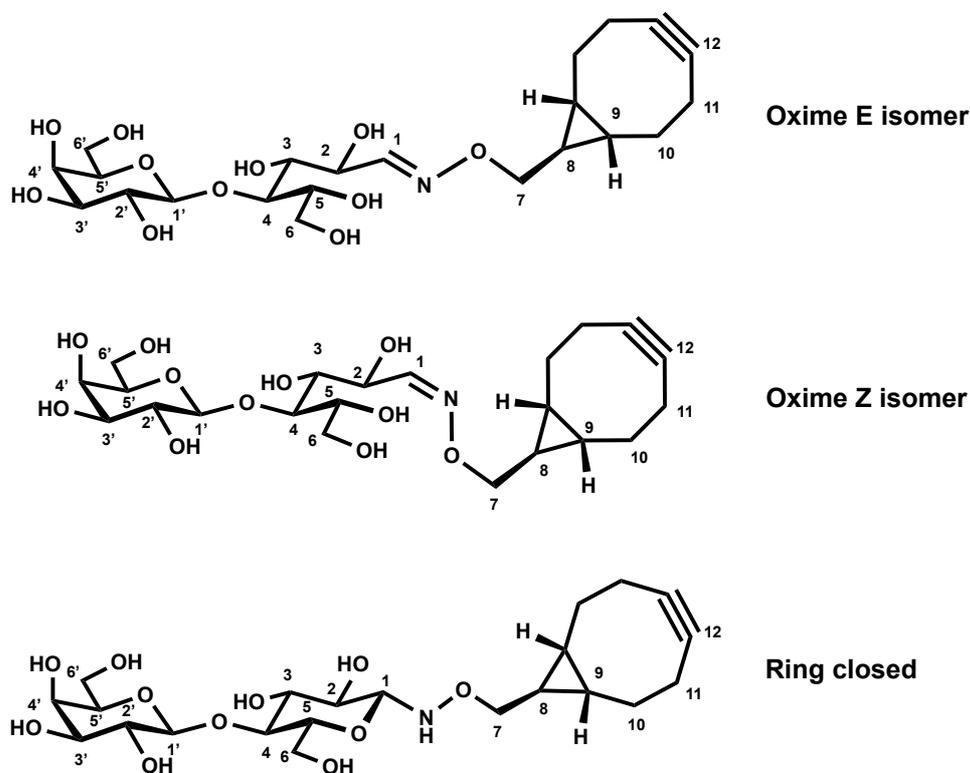
$^1\text{H NMR}$ (500 MHz; CDCl_3) δ 0.75-0.89 (2H, m, H_4), 1.14-1.28 (1H, m, H_5), 1.43-1.63 (2H, m, H_3), 2.08-2.30 (6H, m, $\text{H}_{2,2',3'}$), 2.67 (3H, s, H_7), 3.56 (2H, d, $J = 7.6$ Hz, H_6).

$^{13}\text{C NMR}$ (125 MHz; CDCl_3) δ 17.65 (C_5), 19.84 (C_4), 21.50 (C_3), 29.18 (C_2), 39.34 (C_7), 70.35 (C_6), 98.94 (C_1).

HRMS - $\text{C}_{11}\text{H}_{17}\text{NO}+\text{H}$ requires 180.1382 Da. $[\text{M}+\text{H}]^+$ measured $m/z = 180.1388$ Da. err = -3.0 ppm.

8.2 Synthesis of glycan derivatives

Compound 3.3: Lac-BCN



3.3

Method 1: Oxime ligation in CHCl₃:MeOH

Lactose (20 μ l of 1.5 M stock in H₂O, final concentration 150 mM) was added to 180 μ l of 390 mM **1.2** in 1:1 CHCl₃:MeOH (final concentration 351 mM) in a PCR tube. The reaction was heated at 50 °C using a lid heated PCR machine and held at 50 °C for 24 hours. The organic solvent was allowed to evaporate and the crude product was then diluted to 2 ml μ l in ddH₂O and loaded onto a C18 reverse phase cartridge, the cartridge was then washed with ddH₂O until no carbohydrate derivatives were detected by TLC staining in orcinol. **3.3** was then eluted from the cartridge in 20% aq. methanol until there was no product observed by TLC staining in orcinol. Fractions in which **3.3** was observed were combined and concentrated by lyophilisation to yield **3.3** as a white solid. (10 mg, 69%)

Method 2: Oxime ligation in 1 M NaOAc pH 5

Lactose (80 μ l of 1.5M stock in H₂O, final concentration 300 mM) was added to 80 μ l of 5 M NaOAc pH 5 in a PCR tube, to the solution 240 μ l of 390 mM **1.2** in 1:1

CHCl₃:MeOH was added. The reaction was heated at 50 °C using a lid heated PCR machine and held at 50 °C for 24 hours. The crude mixture was then diluted to 2 ml in ddH₂O and loaded onto a C18 reverse phase cartridge, the cartridge was then washed with ddH₂O until no carbohydrate derivatives were detected by TLC staining in orcinol. **3.3** was then eluted from the cartridge in 20% aq. methanol until there was no product observed by TLC staining in orcinol. Fractions in which **3.3** was observed were combined and concentrated by lyophilisation to yield **3.3** as a white solid. (42.9 mg, 69%) in a ratio of 20:5:14 Oxime E:Oxime Z:Ring closed.

¹H NMR (500 MHz; D₂O) Oxime *E* isomer - δ 0.84-0.96 (3H, m, **H_{8,9}**), 1.48-1.60 (2H, m, **H₁₀**), 2.32 (2H, bd, **H₁₁**), 2.43 (2H, bt, **H_{11'}**), 2.57 (2H, bd, **H_{10'}**), 3.52-4.13(12H, m, **H_{2',3,3',4,4',5,5',6,6',7}**), 4.50 (1H, d, , *J* = 7.8 Hz, **H_{1'}**), 4.59 (1H, t, *J* = 6.4 Hz, **H₂**), 7.66 (1H, d, *J* = 5.8 Hz, **H₁**).

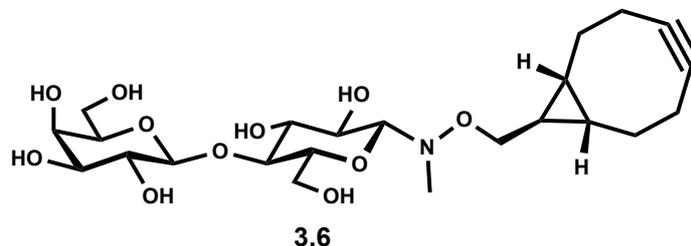
Oxime *Z* isomer - δ 0.84-0.96 (3H, m, **H_{8,9}**), 1.48-1.60 (2H, m, **H₁₀**), 2.28-2.35 (2H, bd, **H₁₁**), 2.43 (2H, bt, **H_{11'}**), 2.56 (2H, bd, **H_{10'}**), 3.52-4.13(12H, m, **H_{2',3,3',4,4',5,5',6,6',7}**), 4.55 (1H, d, *J* = 7.8 Hz, **H_{1'}**) 4.99 (1H, dd, *J* = 5.3, 1.3 Hz, **H₂**), 6.99 (1H, d, *J* = 5.4 Hz, **H₁**).

Ring closed - δ 0.84-0.96 (3H, m, **H_{8,9}**), 1.48-1.60 (2H, m, **H₁₀**), 2.28-2.35 (2H, bd, **H₁₁**), 2.43 (2H, bt, **H_{11'}**), 2.56 (2H, bd, **H_{10'}**), 3.53 (1H, t, *J* = 9.0 Hz, **H₂**), 3.52-4.13 (12H, m, **H_{2',3,3',4,4',5,5',6,6',7}**), 4.46 (1H, d, *J* = 9.2 Hz, **H_{1'}**), 4.60 (1H, d, *J* = 7.8 Hz, **H₁**).

¹³C NMR (125 MHz; D₂O) δ 20.69, 22.35, 22.48, 22.53, 22.58, 22.57, 23.15, (**C_{11,10,9,8}**) 32.73 (**C₇**), 57.83, 60.86, 61.04, 62.09, 68.48, 68.48, 68.57, 69.31, 69.34, 70.97, 71.05, 71.16. 72.53, 72.56, 75.07, 75.36, 75.39, 75.95, 78.10, 78.27, 78.38, 79.33 (**C_{2,2',3,3',4,4',5,5',6,6'}**) 90.06 (**C_{1(C)}**), 103.8 (**C₁₂**), 102.89 (**C_{1'(A)}**), 130.00 (**C_{1'(B,C)}**), 151.70 (**C_{1(A,B)}**).

HRMS - C₂₂H₃₅NO₁₁+H requires 490.2282 Da. Measured m/z [M+H]⁺ = 490.2282 Da. err = 0.1 ppm.

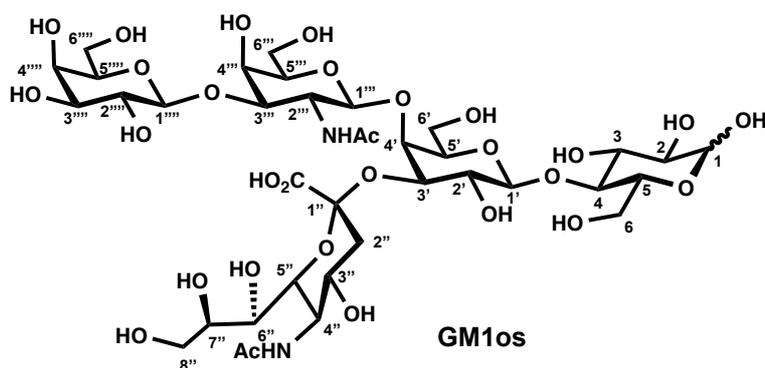
Compound 3.6: Lac-NMeBCN



Lactose (20 μ l of a 1 M stock in H₂O, final concentration 400 mM) was added to 10 μ l of 5 M NaOAc buffer pH 5 in a PCR tube, followed by 20 μ l of a 1M solution of **1.3** in CHCl₃ (final concentration 400 mM). The reaction was heated at 50 °C using a lid heated PCR machine and held at 50 °C for 48 hours. The crude mixture was then diluted to 500 μ l in ddH₂O and loaded onto a C18 reverse phase cartridge, the cartridge was then washed with ddH₂O until no carbohydrate derivatives were detected by TLC staining in orcinol. **3.3** was then eluted from the cartridge in 50% aq. methanol until there was no product observed by TLC staining in orcinol. Fractions in which **3.6** was observed were combined and concentrated by lyophilisation to yield **3.6** as a white solid. (1.6 mg, 40%).

HRMS - C₂₃H₃₇NO₁₁+H requires 504.2445 Da. Measured m/z [M+H]⁺ = 504.2446 Da. err = 0.19 ppm.

GM1os



GM1-ceramide (5 mg, 5.01 μ mol), and dissolved in 450 μ l 50 mM NaOAc buffer, 0.2% v/v Triton X-100 pH 5 in an Eppendorf tube. To the solution EGCCase II (50 μ l of 68 μ M) the reaction was left incubating at 37 °C for between 3-5 days until the no GM1-ceramide was observed by TLC (3:2:1 CHCl₃:MeOH:H₂O). Once the reaction was complete the

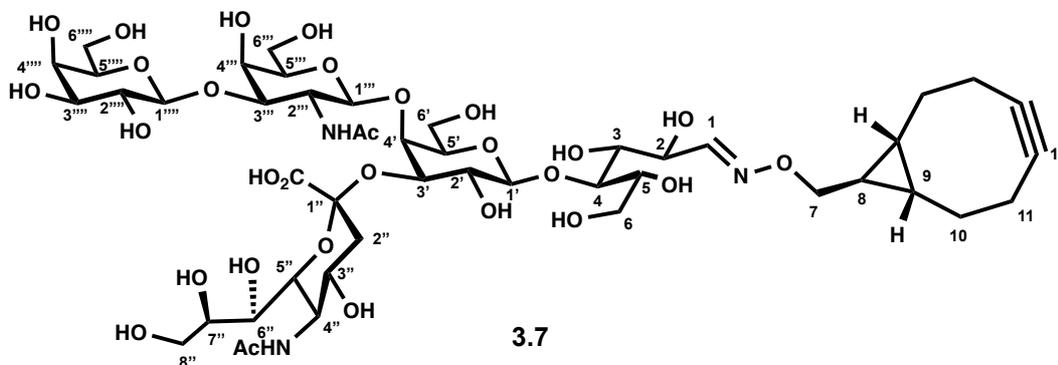
solution was washed with diethyl ether (3 x 30 ml). The aqueous layer was then filtered through 0.22 μm filter, the filtrate was then loaded onto a C18 reverse phase cartridge, GM1os was obtained by eluting in H_2O . Fractions containing GM1os were combined and concentrated by lyophilisation to obtain a white solid. Further purification was by gel filtration (LH20 or P2 bio-gel) was required to remove any salts. Fractions containing GM1os were once again combined and concentrated by lyophilisation to yield GM1os as white solid (3.85 mg, 77%).

$R_f = 0.06$ (60:40:8 CHCl_3 :MeOH:H $_2\text{O}$).

$^1\text{H NMR}$ (D_2O) δ 2.02 (3H, s, Ac), 2.05 (3H, s, Ac), 2.65-2.70 (1H, m, $\text{H}_{3''}$ equatorial), 3.27-3.31 (1H, m, H_2), 3.35-3.41 (2H, m, $\text{H}_{2'',4}$), 3.49-3.99 (46H, m, not assigned), 4.03-4.08 (2H, m, $\text{H}_{2''''}$), 4.12-4.19 (5H, m, $\text{H}_{3',4',4''}$), 4.53 (1H, d, $J = 7.9$ Hz, $\text{H}_{1''''}$), 4.56 (1H, d, $J = 7.9$ Hz, $\text{H}_{1'}$), 4.69 (1H, d, $J = 8.0$ Hz, $\text{H}_{1\beta}$), 4.79 ($\text{H}_{1''}$, lies under D_2O peak) 5.23 (1H, d, $J = 5.2$ Hz, $\text{H}_{1\alpha}$).

Signals conformed with reported literature values.^{270, 271}

Compound 3.7 - GM1-BCN



Oxime ligation in CHCl_3 :MeOH

GM1os (1 mg, 1 μmol , 1 eq) was suspended in 2 μl of 1:1 CHCl_3 :MeOH to a concentration of 250 mM in a PCR tube. To the suspension 3 μl of 390 mM solution of linker **3** (234 mM) was added and the reaction was heated to 50 $^\circ\text{C}$ using a lid heated PCR machine and held at 50 $^\circ\text{C}$ for 48 hours. Completion of the reaction was confirmed by TLC (2:2:1 BuOH:MeOH:H $_2\text{O}$). The crude mixture was then diluted to 100 μl in ddH $_2\text{O}$ and loaded onto a C18 reverse phase cartridge, and the cartridge was then washed with

ddH₂O until no carbohydrate derivatives were detected by TLC staining in orcinol. Compound **11** was then eluted from the cartridge in 20% aq. methanol until there was no product observed by TLC staining in orcinol. Fractions in which **11** was observed were combined and concentrated by lyophilisation to yield compound **11** as a white solid. (0.38 mg, 33%)

Oxime ligation in 1 M NaOAc pH 5

GM1os (1 mg, 1 μ mol, 1 eq) was dissolved in 2 μ l to a concentration of 250 mM in a PCR tube to which 1 μ l of 5M NaOAc buffer pH 5 (final concentration 1 M) was added. To the solution, 2 μ l of 390 mM linker **3** was added and the reaction was heated to 50 $^{\circ}$ C using a lid heated PCR machine and held at 50 $^{\circ}$ C for 48 hours. Completion of the reaction was confirmed by TLC (2:2:1 BuOH:MeOH:H₂O). The crude mixture was diluted to 100 μ l in ddH₂O and loaded onto a C18 reverse phase cartridge, the cartridge was then washed with ddH₂O until no carbohydrate derivatives were detected by TLC staining in orcinol. Compound **11** was eluted from the cartridge in 20% aq. methanol until there was no product observed by TLC staining in orcinol. Fractions in which compound **11** was observed were combined and concentrated by lyophilisation to yield **11** as a white solid. (0.28 mg, 24%).

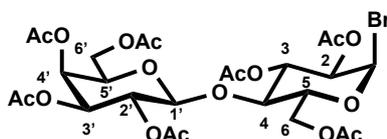
R_f- 0.63 (2:2:1 BuOH:MeOH:H₂O)

¹H NMR (500 MHz; D₂O) δ 0.76-0.82 (3H, m, **H**_{8,9}), 1.36-1.46 (2H, m, **H**₃), 2.00-2.5 (6H, m, **Ac**), 2.12-2.21 (2H, m, **H**₂), 2.25-2.35 (2H, m, **H**_{2'}), 2.40-2.47 (2H, m, **H**₃), 3.29-4.19 (70 H, **not assigned**, integration corresponds to signals for all three isomers), 4.30-4.32 (1H, m, **H**₁ ring closed), 4.50 (1H, t, **H**₂ oxime), 4.54 (2H, m, **H**_{1'',1'''}), 7.49 (1H, d, **H**₁ oxime).

¹³C NMR – Poor resolution could not be assigned.

HRMS – C₄₇H₇₅N₃O₂₉+H requires 1456.4564 Da. Measured m/z [M+H]⁺ = 1456.4571 Da. err. 0.4 ppm

Compound 3.9: Hepta-O-acetyl- α -D-lactosyl bromide



3.9

174

Lactose (15g, 43.8 mmol, 1 eq) was added to acetic anhydride (50 ml, 528 mmol, 12 eq) under N₂, followed by 10 ml HBr in AcOH and stirred at room temperature for 1 hour. A further 50 ml of HBr in AcOH was added and the reaction stirred for another 3 hours. The reaction was concentrated *in vacuo* heating below 30 °C, to yield an orange solid. The product was redissolved in DCM (250 ml), and washed with H₂O (100ml), then sat. aq. NaHCO₃ (100ml) and H₂O (100 ml). The organic layer was dried over MgSO₄ and concentrated *in vacuo*, co-evaporating with toluene, yielding a white solid (25.17 g, 82%) **R_f** = 0.7 (30% EtOAc in DCM).²⁷²

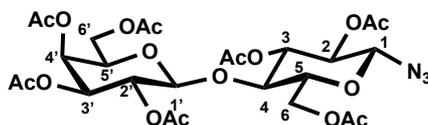
¹H NMR (500 MHz; CDCl₃) δ 1.97, 2.06, 2.07, 2.08, 2.10, 2.14, 2.16 (3H, s, Ac), 3.81 (2H, m, H_{4,6b}), 4.04 (4H, m, H_{5, 6b, 6'a, 6'b}), 4.48 (1H, d, *J* = 9.7, H_{5'}), 4.50 (1H, d, *J* = 7.8, H_{1'}), 4.74 (1H, dd, *J* = 9.8, 4.0, H₂), 4.95 (1H, dd, *J* = 10.4, 3.5, H_{3'}), 5.11 (1H, dd, *J* = 10.4, 8.0, H_{2'}), 5.34 (1H, d, *J* = 3.0, H_{4'}), 5.53 (1H, t, *J* = 9.8, H₃), 6.51 (1H, d, *J* = 4.0, H₁).

¹³C NMR (125 MHz; CDCl₃) δ 20.51, 20.56, 20.66, 20.68, 20.69, 20.81, 20.82 (Ac), 60.92 (C₆), 61.10 (C_{6'}), 66.66 (C_{4'}), 69.09 (C₂), 69.66 (C₃), 70.85 (C_{5'}), 70.91 (C_{2'}), 71.05 (C_{3'}), 73.02 (C₅), 75.01 (C₄), 86.41 (C₁), 100.84 (C_{1'}), 168.9, 169.25, 170.01, 170.10, 170.17, 170.21, 170.37 (Ac).

HMRS - C₂₆H₃₅BrO₁₇+Na requires 721.0949 Da. Measured *m/z* [M+Na]⁺ = 721.0949 Da. err = 0.1 ppm.

IR (ν_{max}/cm⁻¹) - 1014 (C-O), 1369 (C-O, ester), 1743 (C=O).

Compound 3.10: Hepta-*O*-acetyl-β-D-lactosyl azide



3.10

Compound **3.9** (1g, 1.43 mmol, 1 eq) was dissolved in DMF (14 ml, 0.1 M). To the solution sodium azide (102 mg, 1.57 mmol, 1.1 eq) was added, and the reaction was left to stir overnight at room temperature. After this time the reaction was diluted in DCM (40 ml) and washed with NaHCO₃ (3×50 ml), H₂O (50 ml) and brine (50 ml). The organic layer was dried over MgSO₄ and concentrated *in vacuo* and placed on high vacuum to yield a white crystalline solid (801 mg, 85%).³¹⁸

$R_f = 0.57$ (30% EtOAc in DCM).

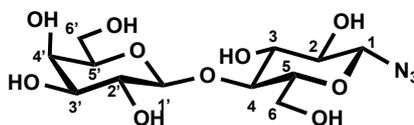
$^1\text{H NMR}$ (500 MHz; CDCl_3) δ 1.96 (3H, s, Ac), 2.04 (6H, s, Ac), 2.06 (3H, s, Ac), 2.07 (3H, s, Ac), 2.13 (3H, s, Ac), 2.15 (3H, s, Ac), 3.66-3.72 (1H, m, H_5), 3.77-3.89 (2H, m, $\text{H}_{4,6'a}$), 4.04-4.19 (3H, m, $\text{H}_{6a,b,5'}$), 4.50 (2H, d, $J = 7.9$ Hz, $\text{H}_{1',6'b}$), 4.62 (1H, d, H_1 , $J = 8.8$), 4.86 (1H, t, H_2 , $J = 9.2$ Hz), 4.95 (1H, dd, H_2' , $J = 10.4, 3.4$ Hz), 5.10 (1H, dd, H_3' , $J = 10.4, 2.4$ Hz), 5.20 (1H, t, H_3 , $J = 9.2$ Hz), 5.35 (1H, d, H_4' , $J = 2.7$ Hz).

$^{13}\text{C NMR}$ (125 MHz; CDCl_3) δ 20.49 (Ac), 20.63 (4×Ac), 20.73 (Ac), 20.80 (Ac), 61.87 (C_6), 61.82 ($\text{C}_{6'}$), 66.68 (C_4'), 69.19 ($\text{C}_{2'}$), 70.88 ($\text{C}_{5'}$), 71.02 (C_2), 71.12 ($\text{C}_{3'}$), 72.64 (C_3), 74.92 (C_5), 75.86 (C_4), 87.79 (C_1), 101.19 ($\text{C}_{1'}$), 169.05 (Ac), 169.47 (Ac), 169.61 (Ac), 170.03 (Ac), 170.09 (Ac), 170.29 (Ac).

HMRS - $\text{C}_{26}\text{H}_{39}\text{NO}_{17} + \text{NH}_4$ requires 679.2305 Da. Measured m/z $[\text{M} + \text{Na}]^+ = 679.2295$ Da err = 1.4 ppm.

IR ($\nu_{\text{max}}/\text{cm}^{-1}$) - 1040 (C-O), 1209 (C-O, ester), 1376 (C-O), 1738 (C=O), 2118 (N=N=N).

Compound 3.11: β -D-lactosyl azide



3.11

Compound **3.10** (801 mg, 1.21 mmol, 1 eq) was dissolved in MeOH (5 ml); to the solution excess NaOMe was added. The reaction was left to stir for 30 minutes. To the mixture DOWEX-50 H^+ was added until neutral. The resin was removed by vacuum filtration, and the solution was concentrated *in vacuo*. The resulting oil was redissolved in H_2O (1 ml) and lyophilised to yield a white solid (44 mg, quantitative yield).³¹⁸

$R_f = 0.2$ (60:40:8 CHCl_3 :MeOH: H_2O).

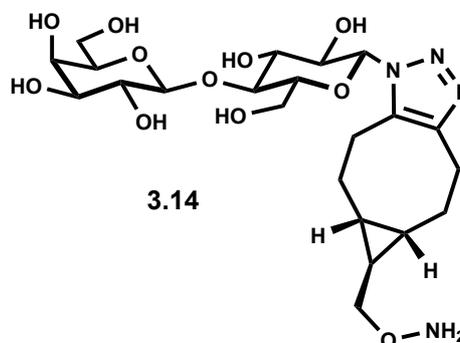
$^1\text{H NMR}$ (500 MHz; D_2O) δ 3.20-3.26 (1H, m, H_2), 3.42-3.46 (1H, dd, $J = 7.8, 1.9$ Hz, H_2'), 3.55-3.76 (8H, m, $\text{H}_{3,3',4,5,5',6,6'}$), 3.82-3.84 (1H, d $J = 3.4$ Hz, H_4'), 3.87-3.90 (1H, dd, $J = 12.0, 1.5$ Hz, H_6), 4.34-4.37 (1H, d, $J = 7.8$ Hz, $\text{H}_{1'}$), 4.68 (1H, lies under D_2O peak, H_1).

$^{13}\text{C NMR}$ (125 MHz; D_2O) δ 59.95 ($\text{C}_{6'}$), 61.12 (C_6), 68.84 (C_4'), 71.03 (C_3), 72.59 ($\text{C}_{2'}$), 72.60 (C_2), 74.44, 75.44, 76.79, 77.83 ($\text{C}_{4,3',5',5}$), 90.01 (C_1), 102.97 ($\text{C}_{1'}$).

HMRS - $C_{12}H_{21}N_3O_{10} + Na$ requires 390.1119 Da. Measured $m/z [M+Na]^+ = 390.1125$
Da. err = -1.7 ppm.

IR (ν_{max}/cm^{-1}) - 3287 (broad, OH), 2117 (N=N=N), 1354 (C-O).

Compound 3.14 – Lac-BCN-OH₂



A solution of linker **1.2** (390 mM in $CHCl_3:MeOH$, 15 μ l, 0.585 mmol, 1 eq) was added to a solution of lactosyl azide **3.11** (100 mM in H_2O , 0.5 ml, 10 eq). The reaction was stirred vigorously for eight hours, before being determined to be complete by HILIC-MS. The reaction mixture was then diluted in H_2O (0.5 ml) loaded to C18 reverse phase cartridge and the excess **3.11** was washed off cartridge with H_2O (2 ml). The product was then eluted from the column in 30% aq. MeOH. The solvent was removed in vacuo followed by lyophilisation to yield a white powder (3.11 mg, 100%).

HRMS – $C_{22}H_{36}N_4O_{11}+H$ requires 533.2459. Measured $m/z [M+H]^+ = 533.2454$. err = 0.94 ppm.

8.2.1 Enzymatic synthesis

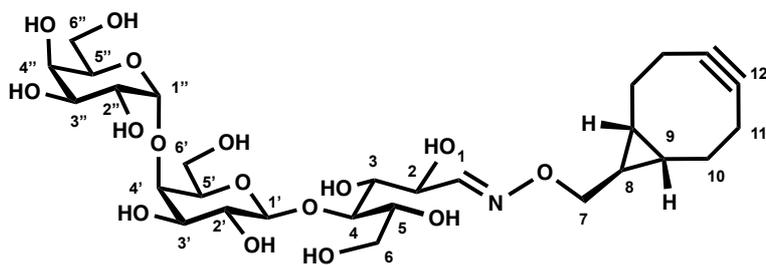
General procedures for the enzymatic galactosylation of oligosaccharides and oligosaccharide derivatives using LgtC.

8.2.1.1 General procedure for enzymatic α 1-4 galactosylation with UDP-Gal

To an Eppendorf tube the lactosyl substrate (final concentration 1 mM), UDP-Gal (final concentration 2 mM), $MgCl_2$ (final concentration 1 mM), LgtC (final concentration 100 μ M) and NaCl (final concentration 100 mM) were added to 50 mM Tris at pH 8.0 and the reaction was incubated at 37 $^{\circ}C$ overnight. The reaction was monitored by LC-MS on analytical HILIC column.

8.2.1.2 Synthesis of glycan derivatives using enzymatic cycle

Compound 3.3: Gb3-BCN



3.17

To an Eppendorf tube the following were added together: 1 μ l of 100 mM compound **3.3** (final concentration 10 mM), 0.2 μ l of 100 mM UDP-Glc (final concentration 2 mM) 0.1 μ l of 2.5 M sucrose (final concentration 250 mM), 0.1 μ l of 100 mM MnCl_2 (final concentration 1 mM), 0.1 μ l of 100 mg/ml BSA (final concentration 1 mg/ml) and 0.1 μ l of 5 M NaCl (final concentration 50 mM) was added to 1 μ l of 500 mM Tris pH 7, followed by 0.9 μ l of ddH₂O. 10 μ l of SuSy and epimerase were taken from an ammonium sulfate precipitate and pelleted in separate tubes by centrifugation at $17,000 \times g$. The supernatant was removed, and each pellet resuspended in 10 μ l ddH₂O. 3 μ l of SuSy (final concentration 235 ng/ml) and 3 μ l of Glc-epimerase (final concentration 3.87 mg/ml) was then added to the reaction tube, along with 1.8 μ l of 572 μ M LgtC (final concentration 100 μ M). The reaction was incubated at 37 $^\circ\text{C}$ for 18 hours.¹ The reaction was monitored by LC-MS on analytical HILIC column. Once the mass of trisaccharide **3.17** was the sole species detected, any precipitated enzyme was removed by centrifugation at $17,000 \times g$ for 10 minutes and the supernatant was removed. The crude mixture was then diluted to 100 μ l in ddH₂O and loaded onto a C18 reverse phase cartridge, the cartridge was then washed with ddH₂O until no carbohydrate derivatives were detected by TLC staining in orcinol. Trisaccharide **3.17** was then eluted from the cartridge in 20% aq. methanol until there was no product observed by TLC staining in orcinol. Fractions in which **3.17** was observed were combined and concentrated by lyophilisation to yield **3.17** as a white solid. (65 μ g, 100%).

¹H NMR (500 MHz; D₂O) δ 0.71-0.86 (3H, m, **H**_{8,9}), 1.36-1.46 (2H, m, **H**₃), 2.14-2.20 (2H, m, **H**₂), 2.25-2.34 (2H, m, **H**_{2'}), 2.39-2.45 (2H, m, **H**₃), 3.36-3.45 (1H, m, **H**_{2'}), 3.54-4.15 (not assigned)

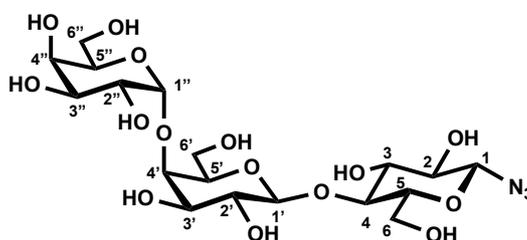
3.29-4.19 (**not assigned**), 4.30-4.32 (1H, m, **H₁** ring closed), 4.28-4.35 (not assigned) (1H, t, **H₂** oxime), 4.54-4.56 (2H, m, **H_{1'}**), 4.94-5.05 (1H, m, **H_{1''}**), 6.91-6.98 (1H, m, **H₁** oxime *Z* isomer), 7.64 (1H, d, *J* = 5.9 Hz, **H₁** oxime *E* isomer).

¹³C NMR – Poor resolution could not be assigned.

HRMS – C₂₈H₄₅NO₁₆+H requires 652.2817 Da. Measured m/z [M+H]⁺ = 652.2815 Da. err. = -0.3 ppm.

Note 1 – Reaction was scaled by a factor of 200 keeping the same final concentrations.

Compound 3.18 - β-D-azido Gb3



3.18

To an Eppendorf the following were added together: 500 μl of 250 mM **3.11** (final concentration 125 mM), 50 μl of 500 mM UDP-Glc (final concentration 25 mM), 100 μl of 2.5 mM sucrose (final concentration 250 mM), 10 μl of 100 mM MnCl₂ (final concentration 1 mM), 10 μl of 100 mg/ml BSA (final concentration 1 mg/ml) and 10 μl of 5 M NaCl (final concentration 50 mM) were added to 100 μl of 500 mM Tris pH 7 (final concentration 50 mM). 100 μl of SuSy and epimerase were taken from an ammonium sulfate precipitate and pelleted in separate tubes by centrifugation at 17,000 × *g*. The supernatant was removed, and the pellet resuspended in 100 μl ddH₂O. 100 μl of SuSy (final concentration 470 ng/ml) and 100 μl of Glc-epimerase (final concentration 1.2 mg/ml) was then added to the reaction tube, along with 100 μl of LgtC (final concentration 100 μM). The reaction was incubated at 37 °C for 18 hours. To confirm completion of the reaction 10 μl of the reaction was taken and added to an Eppendorf containing 1 μl of 1M **5.2** (final concentration 10 mM), 0.5 μl of 1M CuSO₄ (final concentration 5 mM), 1.5 μl of 1M sodium ascorbate (final concentration 15 mM) in 90 μl of ddH₂O. This reaction was left for five minutes before being analysed by HILIC-MS. Once only trisaccharide Itag conjugate product was detected, any precipitated enzyme

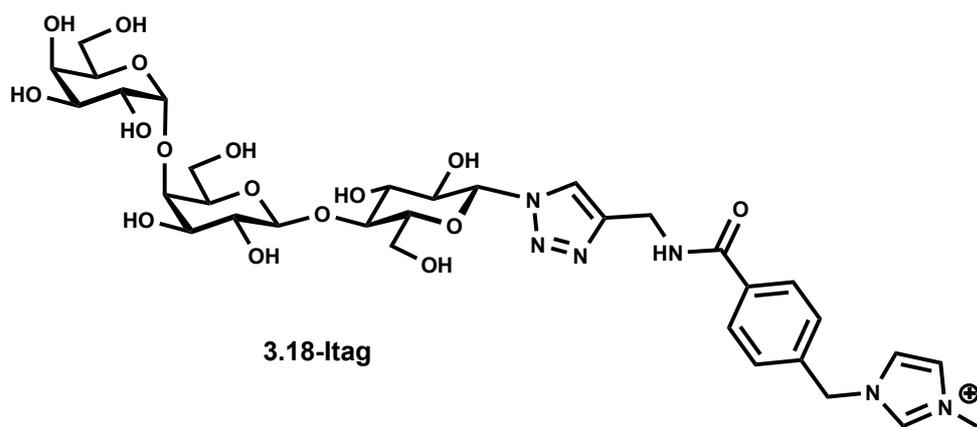
from the mother reaction was removed by centrifugation at $17,000 \times g$ for 10 minutes. To the supernatant Invertase from baker's yeast (purchased from sigma) was added and the reaction was incubated at $37 \text{ }^\circ\text{C}$ for 18 hours. The digestion of sucrose was monitored by TLC until all sucrose was digested. The reaction was then dry-loaded onto silica and the nucleotide donors and UDP by-products were removed by flash column chromatography eluting in 1:1 EtOAc:MeOH. Fractions containing carbohydrate derivatives as observed by TLC (1:1 EtOAc:MeOH) were combined and concentrated *in vacuo* to yield a white residue. The crude product was then further purified by gel filtration on P2 bio-gel eluting in H_2O .

$R_f = 0.1$ (60:40:8 CHCl_3 :MeOH: H_2O).

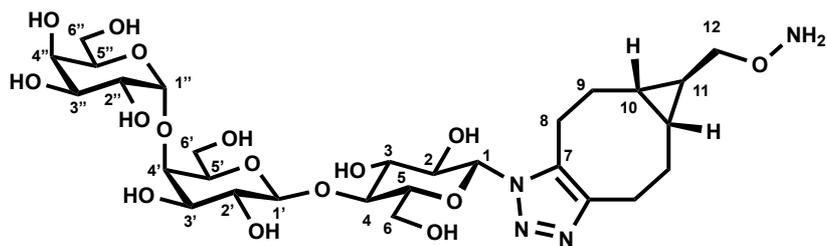
$^1\text{H NMR}$ (500 MHz; D_2O) δ 3.23-3.27 (1H, dd, $\text{H}_{2'}$, $J = 8.0, 1.2$ Hz), 3.56-3.61 (1H, dd, H_2 , $J = 10.2, 2.5$ Hz), 3.63-4.05 (H, m, $\text{H}_{2'',3,3'',3''}, 4, 4', 5, 5', 6, 6', 6''$), 4.13 (1H, d, $\text{H}_{4''}$), 4.36 (1H, t, $J = 6.6$ Hz, $\text{H}_{5''}$), 4.53 (1H, d, $J = 7.9$ Hz, $\text{H}_{1'}$), 4.77-4.79 (1H, H_1 , lies under D_2O peak), 5.23 (1H, d, $J = 3.8$ Hz, $\text{H}_{1''}$).

$^{13}\text{C NMR}$ (125 MHz; D_2O) δ 60.35, 60.48, 62.58 ($\text{C}_6, 6', 6''$), 68.84 (C_4'), 68.54, 70.79, 70.91, 71.14, 72.59, 72.53, 74.03, 74.37, 75.42, 76.72, 77.95 ($\text{C}_2, 2', 2'', 3, 3', 3'', 4, 4'', 5, 5', 5''$), 89.92 (C_1), 100.23 ($\text{C}_{1''}$), 102.97 (C_1).

LC-MS – 3.18-Itag conjugate $\text{C}_{33}\text{H}_{47}\text{N}_6\text{O}_{16}^+$ requires 783.30 Da. Measured m/z $[\text{M}]^+ = 783.18$ Da. err = 153.2 ppm.



Compound 3.19 – Gb3-BCN-OH₂



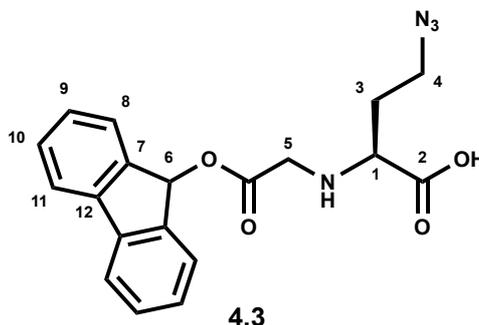
3.19

To a solution of trisaccharide **3.18** (0.5 ml, 0.17 M) in H₂O, 22 μ l of 390 mM **1.2** in CHCl₃:MeOH was added, the reaction was stirred vigorously for 18 hours. The crude mixture was then loaded onto a C18 reverse phase cartridge, the cartridge was then washed with ddH₂O until no carbohydrate derivatives were detected by TLC staining in orcinol. Compound **3.19** was then eluted from the cartridge in 50% aq. methanol until there was no product observed by TLC staining in orcinol. Fractions in which compound **3.19** was observed were combined and concentrated by lyophilisation to yield **3.19** as a white solid. (3 mg, 100%).

HRMS – C₂₈H₄₆N₄O₁₆+H requires 695.2987 Da. Measured m/z [M+H]⁺ = 695.2987 Da. err = -0.7 ppm.

8.3 Synthesis of non-canonical amino acid azidohomoalanine

Fmoc-azidohomoalanine



Fmoc-Dab-OH (300 mg, 0.88 mmol, 1 eq), imidazole-1-sulfonyl azide · H₂SO₄ (662 mg, 2.45 mmol, 2.8 eq) and CuSO₄ hydrate (1.31 mg, 5 μmol) were dissolved in a solution of MeOH:DCM:H₂O (10:8:5, 23 mL). The pH of the solution was adjusted to pH 9 with the addition of sat. K₂CO₃ and the reaction was stirred at r.t. overnight. The reaction mixture was diluted with DCM (50 mL) and extracted with H₂O (30 mL), followed by sat NaHCO₃ (2 × 50 mL). The combined aqueous extracts were washed with Et₂O (2 × 50 mL), acidified to pH 2 with 1M HCl and extracted with Et₂O (3 × 100 mL). The organic extracts were dried over MgSO₄ and concentrated to leave as an amorphous white solid (268 mg, 84%).²⁸³

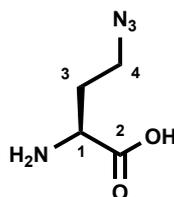
¹H NMR (500 MHz; MeOD-d₄) δ 1.95-1.84 (1H, m, H₃), 2.17-2.06 (1H, m, H₃), 3.40-3.32 (1H, m, H₄), 3.47-3.40 (1H, m, H₄), 4.29-4.21 (1H, m, H₁), 4.29-4.21 (1H, m, H₇), 4.39 (2H, t, *J* = 6.9 Hz, H₆), 7.31 (2H, td, *J* = 7.5, 1.2 Hz, H₁₀), 7.39 (2H, t, *J* = 7.5 Hz, H₁₁), 7.68 (2H, t, *J* = 6.8 Hz, H₉), 7.80 (2H, d, *J* = 7.6 Hz, H₁₂).

¹³C NMR (125 MHz; MeOD-d₄): δ 31.9 (C₃), 52.8 (C₁), 67.90 (C₅), 129.09, 126.30, 128.20, 128.80, 142.60, 145.20, 145.40 (C_{Ar}), 158.70 (C₅), 175.20 (C₂). C₄ and C₆ hidden by NMR solvent signal.

NMR assignments matched literature values.²⁸³

HRMS – C₁₉H₁₈N₄O₄+Na requires 389.1220 Da. Measured *m/z* [M+Na]⁺ = 389.1223 Da. err = 0.8 ppm.

Azidohomoalanine



4.4

Fmoc-Azidohomoalanine (600 mg, 1.64 mmol, 1 eq) was dissolved in EtOH (2 ml), to the solution excess diethylamine (1 ml) was added. The reaction was left to stir at room temperature overnight. The reaction was observed to be complete by MS and the reaction was diluted in H₂O (20 ml), the aqueous layer was subsequently washed with Et₂O (5×20 ml) and concentrated by lyophilisation to yield a white solid. The product contained ≈25% diethylamine.²⁸³

¹H NMR (500 MHz; D₂O) δ 1.93-2.14 (1H, m, H₃), 3.48 (2H, dt, *J* = 6.4, 2.3, H₄), 3.73 (1H, t, *J* = 6.4, H₁).

¹³C NMR (125 MHz; D₂O) δ 29.50 (C₃), 47.57 (C₄), 52.89 (C₁), 173.80 (C₂).

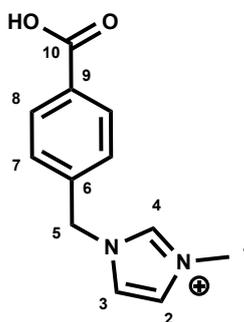
NMR assignments matched literature values.²⁸³

HMRS – C₄H₈N₄O₂+H requires 145.1420 Da. Measured *m/z* [M+H]⁺ = 145.0708 Da. err = 1.4 ppm.

IR (ν_{max}/cm⁻¹) – 1409 (C-N, amine), 1557 (C-O, acid), 1584 (C=O), 2091 (N=N=N), 2948 (O-H, acid), 3197 (N-H).

8.4 Synthesis of clickable I-tag

3-methyl-1-(4-(prop-2-yn-1-ylcarbamoyl)benzyl)-1H-imidazol-3-ium



5.1

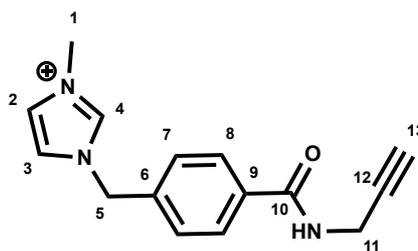
4-(bromomethyl) benzoic acid (500 mg, 2.33 mmol, 1 eq) was added to MeCN (5ml, 0.5 M), to which 1-Methyl imidazole (1.5 ml, 18.64 mmol, 8 eq). A white precipitate formed after 15 minutes. The white precipitate was collected by filtration washing with MeCN (393 mg, 78%).

$^1\text{H NMR}$ (500 MHz; MeOD- d_4) δ 3.94 (3H, s, H_1), 5.49 (2H, H_5 , s), 7.45 (2H, d, $J = 8.5$ Hz, H_6), 7.63 (2H, dt, $J = 21.7, 1.89$ Hz, $\text{H}_{2,3}$), 8.00 (2H, d, $J = 8.5$ Hz, H_7), 9.09 (1H, s, H_4).

$^{13}\text{C NMR}$ (125 MHz; MeOD- d_4) δ 36.66 (C_1), 53.64 (C_5), 123.79 (C_4), 125.36 (C_3), 129.27 (C_7), 131.35 (C_8), 136.48 (C_6), 138.24 (C_9), 138.56 (C_2), 171.57 (C_{10}).

HRMS- $\text{C}_{12}\text{H}_{13}\text{N}_2\text{O}_2^+$ requires 217.0972 Da. Measured m/z [M^+] = 217.0982 Da. err = 4.6 ppm.

3-methyl-1-(4-(prop-2-yn-1-ylcarbamoyl)benzyl)-1H-imidazol-3-ium



5.2

Compound **5.1** (100mg, 0.46 mmol, 1 eq) was dissolved in DMF (2 ml, 0.2 M) with HCTU (248 mg, 0.58 mmol, 1.3 eq) and DIPEA (0.16 ml, 0.92 mmol, 2.5 eq). The reaction mixture was stirred for 30 minutes; propargylamine was added to the reaction dropwise. The reaction was stirred for 18 hours. The product was then purified by cation exchange chromatography eluting in a gradient of 0.1-1M (NH₄)HCO₃, the fractions between 0.1 M and 0.5 M were combined and the product was isolated by lyophilisation to yield an off white solid (60%).

¹H NMR (500 MHz; MeOD-d₄) δ 3.01 (1H, s, H₁₃), 3.93 (3H, s, H₁), 4.15 (2H, s, H₁₁), 5.49 (2H, H₅, s), 7.48 (2H, d, J = 8.4, H₇), 7.63 (2H, dd, J = 14.86, 2.0, H_{3,4}), 7.90 (2H, d, J = 8.4, H₈).

¹³C NMR (125 MHz; MeOD-d₄) δ 29.95 (C₁), 36.52 (C₁₂), 36.56 (C₁₁), 40.05 (C₁₃), 52.07 (C₅), 123.78 (C₄), 125.39 (C₃), 129.39 (C₇), 129.65 (C₈), 135.97 (C₆), 138.92 (C₉), 161.47 (C₂), 168.78 (C₁₀).

HRMS - C₁₅H₁₆N₃O⁺ requires 254.1288 Da. Measured m/z [M]⁺ = 254.1302 Da. err = 5.5 ppm.

8.5 Generation of Gb3 affinity resin

Chemical activation of sepharose 6b

30 ml of sepharose 6b resin (Sigma) was washed on a sinter under vacuum with H₂O (3 x 30 ml). The resin was resuspended in 0.5 M aq. NaHCO₃ pH 11. To the suspension 3 ml of divinyl sulfone was added, and the suspension left on a Stuart rotor for 70 minutes. The resin was then washed with H₂O (3 x 30 ml) on a sinter under vacuum. The activated resin was then resuspended in 0.5 M aq. NaHCO₃ pH 10 with 20% w/v propargyl alcohol.

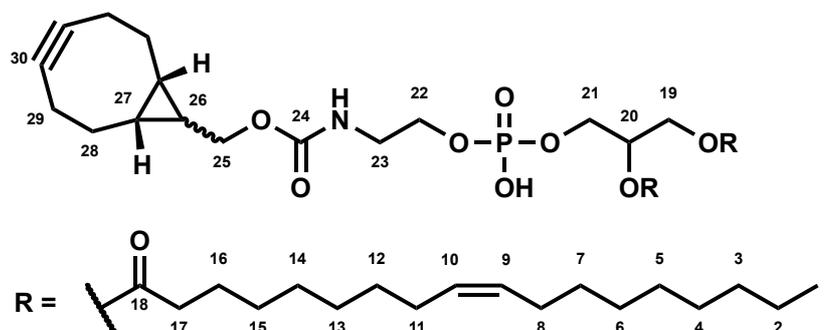
The suspension was then left on Stuart rotor for 24 hours. After this time the resin was washed with H₂O (3 x 30 ml) on a sinter under vacuum and resuspended in 0.5 M aq NaHCO₃ pH 8.5 with 600 µl β-mercaptoethanol. The suspension was left on Stuart rotor for 2 hours, then washed with H₂O (5 x 30 ml) on a sinter under vacuum. The resin was then stored in 50% EtOH in water.

Attachment of 3.18 to alkyne resin

Sepharose resin (6.35 ml) functionalised with alkyne residues and suspended in 50% EtOH was concentrated by centrifugation at 4500 rpm for 10 minutes. The 50% EtOH was removed and the resin washed with 7 x ddH₂O (4 ml). In a Falcon tube, Gb3-N₃ (100 mg), sodium ascorbate (381 µl of 1 M stock, 20 mmol) and ddH₂O (5.84 ml) were added to the resin and finally, CuSO₄ (127 µl of 1 M stock, 20 mmol) was added. The resin was agitated at RT, 600 rpm for 1.5 hours. The resin was added to a Bio-rad Econo-Pac® chromatography column and the water it was suspended in was allowed to run through. The resin was washed with 3 x ddH₂O (10 ml) to remove residual reagents from the 1,3-dipolar cycloaddition.

8.6 Synthesis of Gb3 phospholipids

Compound 5.3-DOPE-BCN



5.3

DOPE (50 mg, 71 µmol, 1 eq) was dissolved in CHCl₃ (2.5 ml), to the solution DIPEA (60 µl, 168 µmol, 2.5 eq) was added and the solution was stirred for 15 minutes. To the solution BCN-NHS (21.5 mg, 73.9 µmol, 1.1 eq) was added and the solution was stirred for 18 hours at room temperature. After this time the reaction was shown to be complete by TLC, the reaction was diluted in CHCl₃ (20 ml) and washed with 0.1 M HCl (30 ml), followed by H₂O (30 ml). The organic layer was dried over MgSO₄ and concentrated *in vacuo* to yield a yellow residue. The crude product was then purified by flash column

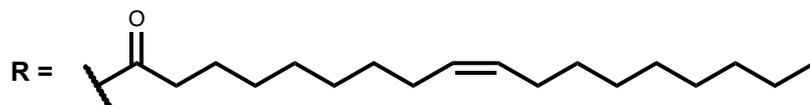
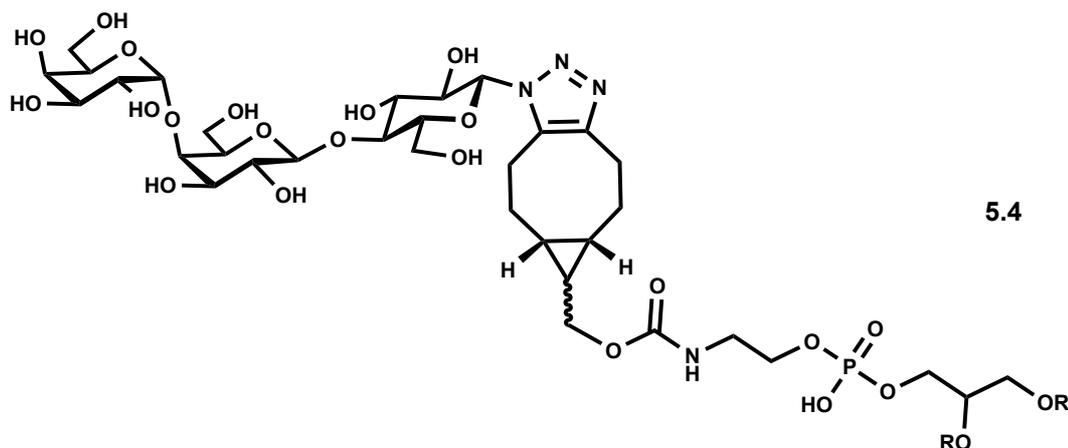
chromatography eluting in 10 % MeOH in CHCl₃, after concentration *in vacuo* the purified product **5.3** was obtained as a clear solid (54 mg, 87%).

R_f = 0.62 (30% EtOAc in CHCl₃).

¹H NMR (500 MHz; CDCl₃) (Spectrum is very broad in comparison to DOPE, not all protons could be assigned). Many assignments based upon assignments obtained from DOPE and BCN starting material NMR spectra) δ 0.72-0.88 (9H, m), 1.11-1.31 (58 H, bm, not assigned), 1.45-1.55 (7H, bm, H_{16,28}), 1.89-1.97 (8H, bm, H_{8,11}), 2.09-2.27 (12H, bm, H_{17, 28, 29, 29'}), 3.30-3.38 (2H, bs, H₂₂), 3.76-3.94 (5H, bs, H_{18, 25}), 3.99-4.16 (3H, bs, H_{21,23}), 4.27-4.37 (1H, bs, H₂₁), 5.08-5.18 (1H, bs, H₂₀), 5.21-5.31 (4H, m, H_{9,10}).

¹³C NMR (125 MHz; CDCl₃) δ 14.13 (C₁), 21.45, 22.70, 24.68, 24.95, 27.25, 29.26, 29.34, 29.36, 29.40, 29.58, 29.72, 29.79, 29.62, 32.06, 34.21, 34.49, 63.20, 63.63, 63.89, 70.82, (could not be assigned) 98.77 (C₃₀), 129.65, 130.04 (C_{9,10}).

Compound 5.4 – Gb3-DOPE



Compound **5.3** (18 mg, 19.6 μmol, 1 eq) was dissolved in a 1:1 CHCl₃:MeOH (250 μl) mixture to which **3.18** (10.35 mg, 19.6 μmol, 1 eq) was added in H₂O (250 μl) the biphasic mixture was stirred vigorously for 18 hours. The reaction was confirmed to be complete by TLC, after which the CHCl₃ was removed *in vacuo*, and the remaining solvent removed by lyophilisation to yield the product **5.4** as a white solid (28 mg, 100%*)

*Yield quoted is based upon isolated material in relation to theoretical yield, no spectral data could be obtained to ascertain if there were any solvent or other impurities present.

Reaction completion was based upon the shift from two species to one as observed by TLC.

$R_f = 5.3$ (30% EtOAc in CHCl_3).

NMR data could not be obtained as sample dissolution was not obtained to gain sufficient signal-to-noise in which the trisaccharide signals could be assigned in relation to the lipid, there is potential of the glycosphospholipids forming micelle structures in solution so further investigation would be required to gain good NMR spectral data.

HRMS – Could not be obtained using ESI-HRMS, mass data could be obtained in future using a MALDI spectrometer.

Attachment of the Gb3 trisaccharide to the lipid was determined by capture of VTB on microtiter plates (see chapter 5, section 5.4, figure 5.10 and figure 10.9).

8.7 Molecular Biology

8.7.1 General methods and instrumentation

All reagents were supplied by commercial sources Sigma-Merck, Fisher, Novogen, NEB, Thermo. Commercial competent cells were obtained from Agilant Technologies. Growth media were purchased from Fisher Scientific and Formamedium.

All SDS-PAGE gels were run using Bio-rad electrophoresis chambers and were stained in Coomassie Blue or InstantBlue™ and imaged on Bio-rad Chemidoc Protein molecular weight markers used for SDS-PAGE were obtained from New England BioLabs or Thermofisher. Protein and DNA concentration were analysed using a Nanodrop 2000 spectrophotometer (Thermoscientific).

8.7.2 Buffer list and recipes

All solutions were made 15 mΩ water (ELGA PURLAB classic), adjusting pH with 5M NaOH and/or 5M HCl.

PBS buffer (pH 7.4)-10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 137 mM NaCl, 2.7 mM KCl.

PBS-T (pH 7.4)-10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 137 mM NaCl, 2.7 mM KCl, 0.1% (v/v) Tween-20, 0.1% (v/v) Bovine serum album.

Phosphate buffer (pH 7.2) – 100 mM Sodium phosphate, 100 mM NaCl

Phosphate buffer high-salt (pH 7.2) – 50 mM Sodium phosphate, 500 mM NaCl

Tris buffer (pH 7.0) – 50 mM Tris base, 150 mM NaCl.

8.7.2.1 Protein unfolding and refolding buffers

Cell resuspension buffer – 50 mM sodium phosphate, 300 mM NaCl pH 8.0.

Inclusion body solubilisation buffer – 0.1 M sodium phosphate, 0.1 M Tris-base, 8 M urea pH 8.0.

Ni-NTA column wash buffer – 0.1 M sodium phosphate, 0.1 M Tri-base, 25 mM imidazole, 8 M urea pH 8.0.

Refolding dialysis buffer – 100 mM Tris, 300 mM NaCl pH 8.5.

8.7.2.2 Protein and DNA analysis buffers

6x DNA loading buffer - 5 mM Tris-HCl pH 8.0, 0.15% w/v orange G (sigma), 2.5% w/v Ficoll 400 (sigma), 10 mM EDTA, dissolved in sterile H₂O.

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

SDS-PAGE running buffer - 25 mM Tris base, 192 mM Glycine. 0.1% SDS.

5x SDS-PAGE loading buffer - 250 mM Tris-HCl, 1M β -mercaptoethanol, 50% w/v glycerol, 10 w/v SDS, 0.01% w/v Bromophenol blue.

8.7.3 Bacterial Growth Media

Lysogeny broth (LB) – 1% w/v tryptone, 0.5 w/v yeast extract, 1% w/v NaCl.

LB agar – 25g LB/L, 15g agar/L.

LB autoinduction (formamedium) – 34.85 g/L.

2TY – 1.6% w/v tryptone, 1.0% w/v yeast extract, 0.5% w/v NaCl.

TYP broth-1.6 g/L tryptone, 1.6 g/L yeast, 5 g/L NaCl, 2.5 g/L K_2HPO_4

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

Table 8.1 – NMM recipe to make 0.4 L culture. All stocks were stored at r.t in the dark with the exception of amino acid solution – 4 °C.

NMM (New Minimal Media)²⁸⁵ –

Component	Final Concentration	Volume	Stock Concentration	Sterilisation
(NH ₄)SO ₄	7.5 mM	20 ml	150 mM	Autoclave
NaCl	8.5 mM		170 mM	
K ₂ HPO ₄	22 mM		440 mM	
KH ₂ PO ₄	50 mM		1 M	
MgSO ₄	1 mM	400 μ l	1 M	
CaCl ₂	1 mg/L	400 μ l	1 mg/ml	0.22 μ m Filter sterilised
FeCl ₂	1 mg/L	400 μ l	1 mg/ml	0.22 μ m Filter sterilised
Canonical amino acids (excluding methionine) – note 1	50 mg/L each	40 ml	0.5 g/L each	0.22 μ m Filter sterilised
Glucose	20 mM	8 ml	1 M	Autoclave
Methionine	70 μ M	280 μ l	100 mM	0.22 μ m Filter sterilised

Trace elements (CuSO ₄ , ZnCl ₂ , MnCl ₂ , (NH ₄)MoO ₄)	10 µg/L	400 µl	10 mg/L each	Autoclave
Thiamine	10 mg/L	400 µl	10 mg/ml	0.22 µm Filter sterilised
Biotin (note 2)	10 mg/L	400 µl	10 mg/ml	0.22 µm Filter sterilised

Note: 1 – To solubilise all amino acids, the solution was warmed to 40 °C. 2 – To solubilise the biotin, the pH was increased with 1M NaOH until dissolved.²⁸⁵ 3 - Methionine, Biotin and Thiamine were stored at -20 °C

8.7.4 General DNA manipulation procedures

8.7.4.1 Site-directed Mutagenesis PCR

Novogen enzymes and buffers were used, and all products were analysed by nucleic acid electrophoresis). Standard PCR conditions were set up in 50 µl reactions on ice, containing the following:

- 1 x KOD hot start buffer
- 140 nM forward primer
- 140 nM reverse primer
- 200 µM dNTPs
- 50 ng DNA template
- 1 unit KOD polymerase
- 1.5 µM MgSO₄

The solutions were made up to 50 µl with dd H₂O and using in the following program cycles:

2 minutes – 95 °C
20 seconds – 95 °
20 seconds – 55 °C
2.5 minutes – 70 °C
10 minutes – 72 °C

} 20 cycles

Restriction digests were carried out using Dpn1 restriction enzymes from New England Biolabs (NEB) and manufacturer buffer. Digests for DNA analysis were set up to a final volume of 20 µl, with approximately 1 unit of enzyme per µg of DNA. The digests were analysed by nucleic acid electrophoresis, following incubation for two hours at 37 °C. The DNA product was then transformed (7.7.4.3) into chemically competent XL10 *E. coli* cells.

8.7.4.2 Nucleic acid electrophoresis

DNA separation by size was carried using 1% w/v agarose gels for DNA greater than 500 bp. The gels were prepared by dissolution of 0.5 g agarose in 50 ml TAE buffer by heating in a microwave until fully dissolved, the solution was allowed to cool slightly before the addition of SYBAsafe (0.5 µl) for visualisation of the DNA by UV exposure. The solution was poured into mould and allowed to cool until set. DNA samples were added to 6x loading buffer to a total volume of 10 µl and loaded onto the gel. Gels were electrophoresed at 80 V for 30 minutes in TAE buffer in Bio-rad mini sub-cell gel tanks.

8.7.4.3 Heat shock transformation into chemically competent *E. coli*

The plasmid (2-5 µl) of interest was added to chemically competent cells (20-50 µl) freshly thawed in an Eppendorf tube and placed on ice for 15 minutes. The solution was heat shocked for 30 seconds at 42 °C before being placed back on ice for 2 minutes. LB or 2TY (500 µl) was then added, and the cells were incubated at 37 °C for 45 minutes. The cells were then concentrated by centrifugation 5000 × g for 2 minutes, the supernatant was discarded, and the pellet was resuspended in 100-200 µl of LB and all the cells were plated onto LB agar containing an appropriate antibiotic.

8.7.4.4 Small scale plasmid DNA purification

XL10 colonies were chosen from LB-agar plates and used to inoculate 5 ml 2TY media. 5 µl of an appropriate antibiotic was added to selectively grow cells containing the desired plasmid. The mini culture was shaken at 160 rpm for 18 hours, at 37 °C.

8.7.5 General protein manipulation

All antibiotics used were used from a 1000x concentrated stock. Plasmid coding for the desired protein was transformed (7.5.3) into an *E. coli* expression cell line. Induction of protein expression was done with 2000x IPTG (240 mg/ml, 1 M) to a concentration of 0.1 mM. Cell lysis was performed by sonication, cell disruption or chemical lysis using BugBuster (Novagen). Centrifugation of cultures and cell lysis was carried out on Beckman Coulter Avanti J30-I centrifuge.

8.7.5.1 Purification of proteins with His-tag

Ni-NTA resin slurry was transferred to a Bio-rad Econo-Pac® chromatography column and equilibrated with 5x resin volume of lysis buffer. The cell lysate was flowed through the Ni-NTA resin two times, to ensure complete protein binding. The resin was then washed with 4x resin volume of wash buffer, then the desired protein eluted off with elution buffer until no protein is observed when added to Bradford Reagent®.

8.7.5.2 Purification of proteins by size exclusion chromatography

Proteins that required further purification post-affinity chromatography were purified by gel filtration. Gel filtration separates proteins based upon shape and size, by flowing them through porous resins usually dextran or agarose based. Small molecules have longer retention times as they are trapped within the pores, whereas, as larger molecules elute faster as they take shorter paths through the resin. Proteins are then detected by UV absorbance, plotted against the retention time and can be used for to identify collected fractions which contain protein.

Large scale purification was carried out on a HiLoad™ 16/60 Superdex™ 75 prep-grade column (GE Healthcare) was attached to Äkta purifier FPLC system flowed at a rate of 1 ml min⁻¹. This column was used for sample sizes of 500 µl to 5 ml, of proteins above 10 kDa.

Analytical purification was carried out on Superdex™ 75 10/300 attached to the FPLC system and flowed at a rate of 0.5 ml min⁻¹. This was used for volumes of 50 µl to 1 ml, for proteins above 10 kDa. All columns were stored in 20% EtOH, and any solutions applied to column were first filtered through 0.22 µm Sartorius Ministart filters. Columns were initially equilibrated with H₂O, followed by 1.5 column volumes of elution buffer.

8.7.5.3 Protein concentration and concentration analysis

Following purification, protein samples were concentrated using Amicon Ultra-15 centrifugal filtration unit with a cut off range of 10 kDa. Fractions were transferred to the filtration units and centrifuged at 3,500 rpm, until the appropriate volume was achieved.

Protein concentrations were measured using UV absorption at 280 nm. The concentration was determined using the Beer-Lambert law (Equation 8.2).

$$A = \epsilon cl$$

Equation 7.1: Equation for the Beer-Lambert law where: **A** is the absorbance; ϵ is the extinction coefficient in mol dm cm; **c** is the concentration in mol dm⁻³ and **l** is the pathlength in cm.

8.7.5.4 SDS Polyacrylamide Gel Electrophoresis

Analysis of proteins by separation by molecular weight was carried out by loading 10 μ l of protein in SDS loading buffer; protein samples may have been boiled depending on the stability of the protein in SDS. A resolving polyacrylamide gel was cast in Bio-Rad 1 mm spacer plates using the recipe shown in Table 7.2 following the addition of TEMED. The gel was left to set with a surface layer of isopropyl alcohol (IPA) to ensure the gel was level and prevent dehydration. Once set the IPA was removed and a stacking gel cast with a BioRad 1 mm comb. After the gel had set the comb was carefully removed and the gel transferred to a Bio-Rad electrophoresis tank. The tank was filled with 0.75 L of SDS-PAGE running buffer ensuring there was no leaking by first filling the gel tank before the surrounding area of the tank. 10 μ l of the protein samples was loaded onto the gel, and the gel was electrophoresed at 180 V for an appropriate time.

Table 8.2 – Recipes for SDS-PAGE gel.

	12%	15% Resolving	Stacking Gel
ddH ₂ O	2.5 ml	3.76 ml	3.13 ml
1.5 M Tris pH 8.8	1.25 ml	2.5 ml	-
1 M Tris pH 6.8	-	-	1.25 ml
10% SDS	50 μ l	50 μ l	50 μ l
10% APS	50 μ l	50 μ l	50 μ l
40% Acrylamide sol	1.25 ml	1.24 ml	0.62 ml
TEMED	5 μ l	5 μ l	5 μ l

8.7.5.5 Protein mass spectrometry

Purified protein samples were diluted to 20 – 40 μM in a 500 μl glass tapered MS vial in appropriate buffer. The protein was ionised on by electrospray on a Bruker MaXis impact instrument. The fragment pattern was analysed and using maximum entropy deconvolution using Bruker DataAnalysis software the correct mass of the protein was acquired.

8.7.6 Expression and purification of non-binding mutant CTB W88E

A glycerol stock of BL21 Gold (DE3) cells containing the CTB W88E mutant plasmid (gifted from Dr Matt Balmforth), were used to inoculate 2 x 5 ml of LB media containing 100 $\mu\text{g/ml}$ ampicillin. The cultures were incubated at 30 $^{\circ}\text{C}$ overnight with shaking at 160 rpm and then used inoculate 4 x 1 L of LB media containing 100 $\mu\text{g/ml}$ of ampicillin. The culture was then grown at 37 $^{\circ}\text{C}$ until an OD_{600} 0.6-0.8 was reached. Protein expression was induced by the addition of IPTG (final concentration 0.5 mM), and the cells were incubated for a further 20 hours at 30 $^{\circ}\text{C}$. Cells were pelleted by centrifugation 13,000 $\times g$ for 20 minutes, after which the supernatant was retained, and the pellet discarded.

Proteins from the supernatant were isolated by precipitation by the addition of solid ammonium sulfate to saturation (57% w/v), the saturated solution was stirred for a further 3 hours at room temperature. The solution was then centrifuged at 13,000 $\times g$ for 30 minutes, the supernatant was then discarded, and the pellet was re-suspended in PBS pH 7.2. The suspension was centrifuged at 13,000 $\times g$ for 10 minutes, and sequentially filtered through 0.8 and 0.22 μm filters. The protein was purified by Ni-NTA affinity chromatography (8.7.5.1) washing with PBS (no imidazole). CTB has native surface histidine residue meaning the pentamer is able to bind to Ni-NTA resin without the presence of a His-tag. The protein was then eluted off the column by washing with PBS (200 mM imidazole), then further purified by size exclusion chromatography on Superdex s75 16/60 in PBS (8.7.5.2).

8.7.7 Expression of MET-W88E

The W88E mutation was introduced to pSAB2.3-N₃ derived from pMAL-p5X (provided by Dr Daniel Williamson)²⁸² by a single round of site directed mutagenesis (8.7.4.2) to generate the plasmid pSAB2.4-N₃. *E.coli* BL21 (DE3) cells were then transformed (8.7.4.3) with the plasmid pSAB2.4 containing the gene for N₃-W88E, and were used to inoculate 2 x 5 ml LB containing 100 $\mu\text{g/ml}$ ampicillin, the starter cultures were incubated at 37 $^{\circ}\text{C}$ with shaking at 200 rpm overnight. The starter cultures were then used to

inoculate 2 x 100 ml of LB media containing 100 µg/ml of ampicillin. The cultures were incubated at 37 °C until an OD₆₀₀ of 0.6 was reached (approximately 2 hours). Expression was induced by addition of IPTG (final concentration 0.5 mM), and the cultures were left at 30 °C overnight. The cells were harvested by centrifugation at 13,000 × g for 30 minutes. The cells were resuspended in 4 ml PBS pH 7.2 and lysed using 0.4 ml BugBuster (10x lysis reagent). The suspension was agitated on rocker at room temperature for 40 minutes. The cell fragments were pelleted by centrifugation at 13,000 × g for 10 minutes, and the pellet discarded. The supernatant was purified using Ni-NTA chromatography (8.7.5.1), washing in PBS pH 7.2 and eluting in PBS pH 7.2 containing 200 mM imidazole. The protein was further purified by size exclusion chromatography on a Superdex s75 16/60 column (8.7.5.2).

8.7.8 Expression of N₃-W88E

Methionine auxotrophic *E.coli* B834 Gold (DE3) cells were transformed (8.7.4.3) with the plasmid pSAB2.4 was then used to inoculate 2 x 5 ml LB starter culture containing 100 µg/ml ampicillin, these starter cultures were incubated at 37 °C with shaking at 200 rpm overnight. The cells of the starter culture were harvested by centrifugation at 4500 × g for 10 minutes, and the supernatant was discarded. The pellet was then resuspended in NMM media (5 ml) and 1 ml of the starter culture was used to inoculate 4 x 400 ml of NMM media containing 100 µg/ml of ampicillin. The cultures were incubated at 37 °C until the OD₆₀₀ had stabilised and cell growth had halted (due to depletion of methionine in NMM) which was reached at around 3 hours. The medium was supplemented with azidohomoalanine (final concentration 0.6 mM) and protein expression was induced by addition of IPTG (final concentration 0.5 mM), and the cultures were left at 30 °C overnight. The cells were harvested by centrifugation at 13,000 × g for 30 minutes. The cells were resuspended in PBS pH 7.2 (10 ml) and lysed by addition of 1 ml BugBuster (10x lysis reagent). The suspension was agitated on rocker at room temperature for 40 minutes. The cell fragments were pelleted by centrifugation at 13,000 × g for 10 minutes, and the pellet discarded. The supernatant was purified using Ni-NTA chromatography (8.7.5.1), washing in PBS pH 7.2 and eluting in PBS pH 7.2 containing 200 mM imidazole. The protein was further purified by gel filtration of Superdex s75 16/60 in PBS (8.7.5.2).

8.7.9 Expression and purification of LgtC

A pET 28 plasmid harbouring the gene for LgtC (Uniprot – Q93EK7) was obtained from GenScript was transformed (8.7.4.3) into BL21 DE3 cells, the cells were grown overnight on LB-agar plates containing Kanamycin. A single colony was used to inoculate 5 ml of LB media containing 50 µg/ml Kanamycin and the culture was grown for 18 hours. The starter culture was shaken at 160 rpm for 18 hours, at 30 °C. The culture was used to inoculate 2 x 500 ml of sterile auto-induction LB media containing kanamycin. The cultures were incubated at 37 °C for 2 hours, then 25 °C overnight. The cells were harvested by centrifugation at 13,000 × g for 40 minutes. The pellet was then resuspended in PBS and the cells were lysed by sonication. Cell fragments were pelleted by centrifugation at 13,000 × g for 30 minutes, and the pellet discarded. The supernatant was then purified by Ni-NTA chromatography (8.7.5.1), washing with Tris buffer containing 20 mM imidazole and eluting in Tris buffer containing 200 mM imidazole. The protein was then dialysed with Tris pH 8.0. The enzyme was concentrated and stored as a 40% glycerol stock, to prevent degradation.

8.7.10 Expression of EGCCase II

A pET-28a plasmid purchased from Genscript containing the gene for EGCCase II (Uniprot - O33853)^{269, 319} was transformed (8.7.4.3) into *E. coli* BL21 Gold cells, and a single colony was used to inoculate 5 ml LB containing 50 µg/ml kanamycin (5 µl of 50 mg/ml stock). The culture was incubated at 37 °C with shaking at 200 rpm overnight and used to prepare a glycerol cell stock. The glycerol cell stock was then used to inoculate 2 x 100 ml TYP broth containing 50 µg/ml kanamycin (100 µl of 50 mg/ml stock). The culture was incubated at 37 °C with shaking at 200 rpm until saturation was reached, at which point the cultures were removed from the incubator and placed in ice for 10 minutes. The incubator was cooled to 20 °C, and protein expression was induced by addition of IPTG (final concentration 0.1 mM) and the culture was left incubating at 20 °C with shaking at 200 rpm for 8 hours. The cells were then harvested by centrifugation at 17000 × g, and the cell pellet was stored at -20 °C. The pellet was then resuspended in 2 ml phosphate buffer (50 mM NaH₂PO₄, 500 mM NaCl, pH 7) and the cells were lysed by addition of BugBuster. The cell debris was removed by centrifugation at 30,000 × g, and the supernatant was retained, the cell debris was resuspended in the same buffer to ensure the protein had been solubilised. The supernatants were combined (totalling 5 ml) and flowed through a Ni-NTA affinity column (8.7.5.1) which had been equilibrated in the same buffer. The column was washed with phosphate buffer (50 mM NaH₂PO₄, 500 mM NaCl, pH 7) containing 20 mM imidazole (3 x CV) and the protein was eluted from the column

using an elution gradient of imidazole of 50, 100, 200, 300, 400, 500 mM (2 x CV)
fractions of each elution.

8.7.11 Expression of VTB

A stab of an XL10 glycerol stock harbouring the plasmid for VTB (Uniprot – Q8X4M7)²⁹⁵ was used to inoculate 5 ml of LB containing 100 µg/ml ampicillin (5 µl of 100 mg/ml stock). The culture was grown at 37 °C with shaking at 200 rpm overnight. The DNA was harvested from the cells using QIAprep® spin Miniprep Kit following the high yield protocol provided by manufacturer. The plasmid was then transformed into BL21 expression cell line, and a single colony was used to generate a glycerol stock. A stab of BL21 glycerol stock harbouring the plasmid for VTB was used to inoculate 5 ml of LB media containing 100 µg/ml ampicillin (5 µl of 100 mg/ml stock). The starter culture was grown at 37 °C overnight with shaking at 200 rpm. 1 ml of the culture was used to inoculate 400 ml CAYE media containing 100 µg/ml ampicillin (400 µl of 100 mg/ml stock), the culture was then incubated at 37 °C with shaking at 200 rpm until an OD₆₀₀ of 0.6-0.8 was reached. Protein expression was then induced by addition of IPTG (final concentration 0.5 mM) the cultures were then incubated at 37 °C with shaking at 200 rpm overnight. After this time the cells were harvested by centrifugation at 17,000 × g for 25 minutes. The pellet was then resuspended in periplasmic extraction buffer and incubated at 37 °C with shaking at 200 rpm for 10 minutes. The cells were then pelleted by centrifugation at 4500 × g for 20 minutes. The supernatant was discarded, and the pellet was resuspended in ice cold ddH₂O. The suspension was shaken on ice for 10 minutes, before the cell debris was harvested by centrifugation at 4500 × g for 20 minutes. The supernatant was retained to which concentrated PBS was added to buffer the solution. The supernatant was then purified by size exclusion chromatography on Superdex s75 16/60 column (8.7.5.2).

8.7.12 Expression of Serum Amyloid P

A pET-11a plasmid containing the gene for mutant SAP component (based upon the Uniprot – P02743, containing mutations E167Q, N-terminal serine, C-terminal LPETG extension, figure 10.6)³⁰⁴ purchased from Genscript was transformed (**8.7.4.3**) into C41 (DE3) expression cell line. A single colony was used to inoculate 5 ml of LB containing 100 µg/ml ampicillin (5 µl of 100 mg/ml stock). The starter culture was incubated at 37 °C with shaking at 200 rpm overnight, 1 ml of the starter culture was then used to inoculate 400 ml of LB media containing 100 µg/ml ampicillin (400 µl of 100 mg/ml stock). The cultures were then incubated at 37 °C with shaking at 200 rpm until an OD₆₀₀ of 0.6-0.8 was reached, expression was then induced by addition of IPTG (final concentration 0.5 M) and the culture was then incubated at 25 °C with shaking at 200 rpm overnight. The cells were then harvested by centrifugation at 17000 × g for 25 minutes, the supernatant was discarded, and the cell pellet was resuspended in phosphate buffer pH 8.0 (50 mM phosphate, 300 mM NaCl) and the cells were lysed by cell high pressure cell disruption. The cell debris was harvested, and the supernatant was discarded. The pellet was then resuspended in an inclusion body resolubilisation buffer (0.1M Tris, 0.1 M NaH₂PO₄ with 8M urea). The suspension was then centrifuged at 4500 × g for 20 minutes to remove any insoluble cellular components. The suspension was then flowed through a Ni-NTA chromatography (**8.7.5.1**) which had been equilibrated in the same buffer, then column was then washed in 50 ml Tris buffer pH 8.0 (50 mM Tris, 300 mM NaCl 20 mM imidazole, 0.1% Triton-X100) The protein was then eluted from the column in Tris buffer pH 8.0 (50 mM Tris, 300 mM NaCl 200 mM imidazole) taking 10 ml fractions until no more protein was detected by Bradford assay. The protein was concentrated to a 10 ml volume using Amicon® ultra spin concentrator with a 10,000 kDa at 4500 × g, the urea was then removed incrementally dialysing into Tris buffer containing 4 M, 2 M, 1 M and finally just Tris buffer. Each dialysis step was performed at 4 °C leaving the dialysis overnight.

8.8 Protein modification

8.8.1 General procedure for oxidation of W88E

To a solution of W88E CTB (100 µl of 703 µM stock, final concentration) in phosphate buffer pH 7.2, NaIO₄ (7 µl of 50 mM stock, final concentration 4.68 mM, 5 eq) was added and the reaction made up 150 µl with sodium phosphate buffer pH 7.2 (note 1). The

reaction was observed to be complete after 15 minutes. The protein was purified using a PD-G25 desalting column eluting in PBS pH 6.8.

Note 1 – It is important that the buffer does not contain potassium, reactions do not reach completion in PBS or potassium containing buffers.²⁸⁷

8.8.2 General procedure of oxime ligation to W88E

W88Eox from the elution of PD-G25 column using Amicon spin concentrator (MW = 10,000 Da cut off) to a concentration above 400 μ M and then used directly in ligation by adding 100 mM aniline, and oxyamine substrate (10 eq per protomer). The reaction was left agitating at room temperature until HRMS showed completion of reaction generally between 8-16 hours. The reaction was then purified using a PD-G25 desalting column, eluting the glycoprotein in which had been eluting in sodium phosphate buffer pH 7.2.

8.8.3 General procedure for SPAAC ligation to N₃-W88E

N₃-W88E in PBS pH 7.0 was concentrated to 500 μ M (protomeric concentration). In an Eppendorf tube BCN glycan substrate (1-20 eq) in H₂O, to which concentrated PBS (20 x) was added. The concentrated protein solution was then added to the reaction tube and the reaction was incubated at 37 °C for between 4-8 hours, until the reaction was observed to be complete by HRMS. Once SPAAC conjugate was complete the reaction was purified using a PD-G25 desalting column eluting the protein in PBS pH 7.2.

8.8.4 Lightning-Link[®] HRP conjugation to CTB

Using a 10-40 μ g vial of the Lightning-Link[®] HRP conjugation kit, 10 μ l of WT CTB (461 μ M, protomeric concentration, 1:1 pentamer:HRP) was added to the supplied vial and the solution was mixed with to resuspend the lyophilised HRP. 1 μ l of the LL-modifier reagent was added, and the reaction was left for 3 hours at room temperature. The reaction was then quenched with 1 μ l of LL-quencher reagent, the quenching step was left for 30 minutes before use or storage. CTB-HRP was stored as 1 ml aliquots at 90 ng/ml in PBS which were flash frozen and stored at -20 °C.

8.8.1 Lightning-Link[®] HRP conjugation to VTB

Using a 100-400 μ g vial of the Lightning-Link[®] HRP conjugation kit, 100 μ l of WT VTB-1 (411 μ M, protomeric concentration, 1:1 pentamer:HRP) was added to the supplied vial with 10 μ l of LL-modifier reagent. The solution was mixed to resuspend the lyophilised HRP. The reaction was left for 3 hours at room temperature before the reaction was

quenched by addition of 10 μ l LL-quencher reagent. VTB-HRP was then stored in 1 ml aliquots of 1 μ g/ml in PBS with 50% glycerol. The aliquots were then flash frozen before being stored at -20 °C.

8.9 ELLA protocols

CTB-HRP and VTB-HRP prepared using Expedeon Lightning-Link HRP conjugation kit were used in the ELLA inhibition assays. Assays were performed using Amplex red® to detect the presence of HRP conjugates. A preliminary assay was performed to determine a suitable concentration of toxin-HRP conjugates to produce linear data plots when bound to glycolipid coated microtiter plates.

8.9.1 CTB inhibition assay

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) removing any remaining GM1, any remaining binding sites on the well were blocked with BSA by incubating with a PBS solution containing 1% (w/v) BSA (100 μ l) for 30 minutes at 37 °C. The wells were then washed again with PBS (3 x 200 μ l). 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 in the same buffer to give a final concentration of 2.5 ng/ml. The mixture of inhibitor and toxin were incubated for 2 hours at room temperature before 100 μ l was transferred to the GM1 coated well by the Hamilton microlab STAR liquid handling system. The limits of detection were analysed by control samples containing just CTB-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. 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-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.

8.9.2 VTB inhibition assay

The ELLA was performed as described above with plates being coated synthetic glycophospholipid 5.4 (100 μ l, 1.3 μ l in methanol). Inhibitors were premixed with VTB-HRP conjugate to a final concentration of 5 ng/ml).

8.9.3 Data processing

All samples were analysed in triplicate. The error of each sample was calculated as standard error (equation 1) 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 1: 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 2), where A_1 is the curve's maximum, A_2 is the curve's minimum, x_0 is equal to the IC_{50} , 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 2: Equation for logistic curve fitting.

Chapter 9 References

1. A. Nairn and K. Moremen, *Handbook of Glycomics*, Academic Press, 2009.
2. G. A. Khoury, R. C. Baliban and C. A. Floudas, *Scientific Reports*, 2011, **1**, 90.
3. G. Duan and D. Walther, *PLoS Computational Biology*, 2015, **11**, e1004049.
4. S. C. Hubbard and R. J. Ivatt, *Annual Review of Biochemistry*, 1981, **50**, 555-583.
5. S. C. Hubbard and P. W. Robbins, *Journal of Biological Chemistry*, 1980, **255**, 11782-11793.
6. R. J. Staneloni and L. F. Leloir, *Critical Reviews in Biochemistry*, 1982, **12**, 289-326.
7. R. D. Marshall, *Annual Review of Biochemistry*, 1972, **41**, 673-702.
8. R. D. Marshall, *Biochem Soc Symp*, 1974, 17-26.
9. D. D. Pless and W. J. Lennarz, *Proc Natl Acad Sci U S A*, 1977, **74**, 134-138.
10. K. E. Kronquist and W. J. Lennarz, *Journal of Supramolecular Structure*, 1978, **8**, 51-65.
11. C. B. Sharma, L. Lehle and W. Tanner, *European journal of biochemistry*, 1981, **116**, 101-108.
12. S. Silberstein and R. Gilmore, *The FASEB Journal*, 1996, **10**, 849-858.
13. M. Aebi, *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 2013, **1833**, 2430-2437.
14. S. Shrimal, N. A. Cherepanova and R. Gilmore, *Seminars in Cell & Developmental Biology*, 2015, **41**, 71-78.
15. K. Olden, J. B. Parent and S. L. White, *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes*, 1982, **650**, 209-232.
16. S. M. Hurtley and A. Helenius, *Annual Review of Cell Biology*, 1989, **5**, 277-307.
17. R. D. Klausner and R. Sitia, *Cell*, 1990, **62**, 611-614.
18. A. Helenius and M. Aebi, *Annual Review of Biochemistry*, 2004, **73**, 1019-1049.
19. L. W. Ruddock and M. Molinari, *Journal of Cell Science*, 2006, **119**, 4373.
20. R. Kornfeld and S. Kornfeld, *Annual Review of Biochemistry*, 1985, **54**, 631-664.
21. P. Hossler, B. C. Mulukutla and W.-S. Hu, *PLOS ONE*, 2007, **2**, e713.
22. H. Schachter, *Glycoconjugate Journal*, 2000, **17**, 465-483.
23. P. Stanley, N. Taniguchi and M. Aebi, in *Essentials of Glycobiology [Internet]. 3rd edition*, ed. A. Varki, Cold Spring Harbor Laboratory Press, 2017, ch. 9.
24. A. Herscovics, *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1999, **1473**, 96-107.
25. L. L. Lairson, B. Henrissat, G. J. Davies and S. G. Withers, *Annual Review of Biochemistry*, 2008, **77**, 521-555.
26. J. M. Rini and J. D. Esko, in *Essentials of Glycobiology [Internet]. 3rd edition*, ed. A. Varki, Cold Spring Harbor Laboratory Press, 2017, ch. 6.
27. R. G. Spiro, *Glycobiology*, 2002, **12**, 43R-56R.
28. P. V. d. Steen, P. M. Rudd, R. A. Dwek and G. Opdenakker, *Critical Reviews in Biochemistry and Molecular Biology*, 1998, **33**, 151-208.
29. H. C. Hang and C. R. Bertozzi, *Bioorganic & Medicinal Chemistry*, 2005, **13**, 5021-5034.
30. G. D'Angelo, S. Capasso, L. Sticco and D. Russo, *The FEBS Journal*, 2013, **280**, 6338-6353.
31. A. H. Merrill, *Chemical Reviews*, 2011, **111**, 6387-6422.
32. K. Simons and M. J. Gerl, *Nature Reviews Molecular Cell Biology*, 2010, **11**, 688-699.
33. M. A. J. Ferguson and A. F. Williams, *Annual Review of Biochemistry*, 1988, **57**, 285-320.
34. I. Levental, M. Grzybek and K. Simons, *Biochemistry*, 2010, **49**, 6305-6316.
35. J. G. Leroy, *Pediatric Research*, 2006, **60**, 643-656.
36. J. Jaeken and G. Matthijs, *Annual Review of Genomics and Human Genetics*, 2007, **8**, 261-278.
37. I. J. Chang, M. He and C. T. Lam, *Annals of Translational Medicine*, 2018, **6**, 12.

38. H. Freeze, T. Kinoshita and R. Schnaar, in *Essentials of Glycobiology [Internet]. 3rd edition.*, Cold Spring Harbor (NY), Cold Spring Harbor Laboratory Press. , 2017, ch. 44.
39. F. M. Platt, A. d'Azzo, B. L. Davidson, E. F. Neufeld and C. J. Tiffit, *Nature Reviews Disease Primers*, 2018, **4**, 27.
40. A. Imberty and A. Varrot, *Current Opinion in Structural Biology*, 2008, **18**, 567-576.
41. Q. Wang, X. Tian, X. Chen and J. Ma, *Proc Natl Acad Sci U S A*, 2007, **104**, 16874.
42. J. J. Skehel and D. C. Wiley, *Annual Review of Biochemistry*, 2000, **69**, 531-569.
43. T. B. H. Geijtenbeek, R. Torensma, S. J. van Vliet, G. C. F. van Duijnhoven, G. J. Adema, Y. van Kooyk and C. G. Figdor, *Cell*, 2000, **100**, 575-585.
44. T. B. H. Geijtenbeek, D. S. Kwon, R. Torensma, S. J. van Vliet, G. C. F. van Duijnhoven, J. Middel, I. L. M. H. A. Cornelissen, H. S. L. M. Nottet, V. N. KewalRamani, D. R. Littman, C. G. Figdor and Y. van Kooyk, *Cell*, 2000, **100**, 587-597.
45. C. B. Wilen, J. C. Tilton and R. W. Doms, *Cold Spring Harb Perspect Med*, 2012, **2**, a006866.
46. C. Berne, C. K. Ellison, A. Ducret and Y. V. Brun, *Nature Reviews Microbiology*, 2018, **16**, 616-627.
47. K. A. Kline, S. Fälker, S. Dahlberg, S. Normark and B. Henriques-Normark, *Cell Host & Microbe*, 2009, **5**, 580-592.
48. E. A. Merritt and W. G. J. Hol, *Current Opinion in Structural Biology*, 1995, **5**, 165-171.
49. W. H. O. Geneva, 2018.
50. M. Ali, A. R. Nelson, A. L. Lopez and D. A. Sack, *PLOS Neglected Tropical Diseases*, 2015, **9**, e0003832.
51. B. D. Spangler, *Microbiological Reviews*, 1992, **56**, 622.
52. S. Fukuta, J. L. Magnani, E. M. Twiddy, R. K. Holmes and V. Ginsburg, *Infect Immun*, 1988, **56**, 1748-1753.
53. E. A. Merritt, P. Kuhn, S. Sarfaty, J. L. Erbe, R. K. Holmes and W. G. J. Hol, *Journal of Molecular Biology*, 1998, **282**, 1043-1059.
54. M. Mammen, S.-K. Choi and G. M. Whitesides, *Angewandte Chemie International Edition*, 1998, **37**, 2754-2794.
55. W. B. Turnbull, B. L. Precious and S. W. Homans, *Journal of the American Chemical Society*, 2004, **126**, 1047-1054.
56. G. M. Kuziemko, M. Stroh and R. C. Stevens, *Biochemistry*, 1996, **35**, 6375-6384.
57. D. J.-F. Chinnapen, H. Chinnapen, D. Saslowsky and W. I. Lencer, *FEMS Microbiology Letters*, 2007, **266**, 129-137.
58. C. M. Thorpe, *Clinical Infectious Diseases*, 2004, **38**, 1298-1303.
59. A. R. Melton-Celsa, *Microbiology Spectrum*, 2014, **2**, doi:10.1128/microbiolspec.EHEC-0024-2013.
60. A. Donohue-Rolfe, I. Kondova, S. Oswald, D. Hutto and S. Tzipori, *The Journal of Infectious Diseases*, 2000, **181**, 1825-1829.
61. H. Ling, A. Boodhoo, B. Hazes, M. D. Cummings, G. D. Armstrong, J. L. Brunton and R. J. Read, *Biochemistry*, 1998, **37**, 1777-1788.
62. P. M. St. Hilaire, M. K. Boyd and E. J. Toone, *Biochemistry*, 1994, **33**, 14452-14463.
63. E. N. Kitova, P. I. Kitov, E. Paszkiewicz, J. Kim, G. L. Mulvey, G. D. Armstrong, D. R. Bundle and J. S. Klassen, *Glycobiology*, 2007, **17**, 1127-1137.
64. K. M. Gallegos, D. G. Conrady, S. S. Karve, T. S. Gunasekera, A. B. Herr and A. A. Weiss, *PLOS ONE*, 2012, **7**, e30368.
65. H. Nakajima, N. Kiyokawa, Y. U. Katagiri, T. Taguchi, T. Suzuki, T. Sekino, K. Mimori, T. Ebata, M. Saito, H. Nakao, T. Takeda and J. Fujimoto, *Journal of Biological Chemistry*, 2001, **276**, 42915-42922.
66. J. C. Paton and A. W. Paton, *Clinical Microbiology Reviews*, 1998, **11**, 450.
67. B. Binnington and C. Lingwood, *Nephron*, 1989, **51**, 207-210.
68. I. Arvidsson, A.-I. Ståhl, M. Manea Hedström, A.-C. Kristofferson, C. Rylander, J. S. Westman, J. R. Storry, M. L. Olsson and D. Karpman, *The Journal of Immunology*, 2015, **194**, 2309.
69. Y. Endo, K. Tsurugi, T. Yutsudo, Y. Takeda, T. Ogasawara and K. Igarashi, *European Journal of Biochemistry*, 1988, **171**, 45-50.

70. D. L. MacLeod, C. L. Gyles and B. P. Wilcock, *Veterinary Pathology*, 1991, **28**, 66-73.
71. T. K. Dam and C. F. Brewer, *Chemical Reviews*, 2002, **102**, 387-430.
72. G. Safina, *Analytica Chimica Acta*, 2012, **712**, 9-29.
73. Y. C. Lee and R. T. Lee, *Accounts of Chemical Research*, 1995, **28**, 321-327.
74. S. Pérez and I. Tvaroška, in *Advances in Carbohydrate Chemistry and Biochemistry*, ed. D. Horton, Academic Press, 2014, vol. 71, pp. 9-136.
75. B. A. Williams, M. C. Chervenak and E. J. Toone, *Journal of Biological Chemistry*, 1992, **267**, 22907-22911.
76. J. C. Sacchetti, L. G. Baum and C. F. Brewer, *Biochemistry*, 2001, **40**, 3009-3015.
77. L. L. Kiessling, T. Young, T. D. Gruber and K. H. Mortell, in *Glycoscience: Chemistry and Chemical Biology*, eds. B. O. Fraser-Reid, K. Tatsuta and J. Thiem, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, DOI: 10.1007/978-3-540-30429-6_64, pp. 2483-2523.
78. R. T. Lee and Y. C. Lee, *Glycoconjugate Journal*, 2000, **17**, 543-551.
79. W. P. Jencks, *Proceedings of the National Academy of Sciences*, 1981, **78**, 4046-4050.
80. M. I. Page and W. P. Jencks, *Proceedings of the National Academy of Sciences*, 1971, **68**, 1678-1683.
81. T. K. Dam, R. Roy, S. K. Das, S. Oscarson and C. F. Brewer, *Journal of Biological Chemistry*, 2000, **275**, 14223-14230.
82. T. K. Dam and C. F. Brewer, in *Advances in Carbohydrate Chemistry and Biochemistry*, ed. D. Horton, Academic Press, 2010, vol. 63, pp. 139-164.
83. R. S. Kane, *Langmuir*, 2010, **26**, 8636-8640.
84. M. Reynolds and S. Pérez, *Comptes Rendus Chimie*, 2011, **14**, 74-95.
85. Y. M. Chabre, D. Giguère, B. Blanchard, J. Rodrigue, S. Rocheleau, M. Neault, S. Rauthu, A. Papadopoulos, A. A. Arnold, A. Imberty and R. Roy, *Chemistry – A European Journal*, 2011, **17**, 6545-6562.
86. I. Vrasidas, J. Kemmink, R. M. J. Liskamp and R. J. Pieters, *Organic Letters*, 2002, **4**, 1807-1808.
87. D. D. Mitchell, J. C. Pickens, K. Korotkov, E. Fan and W. G. J. Hol, *Bioorganic & Medicinal Chemistry*, 2004, **12**, 907-920.
88. E. A. Merritt, S. Sarfaty, I. K. Feil and W. G. J. Hol, *Structure*, 1997, **5**, 1485-1499.
89. P. I. Kitov and D. R. Bundle, *Journal of the Chemical Society, Perkin Transactions 1*, 2001, 838-853.
90. S. Cecioni, A. Imberty and S. Vidal, *Chemical Reviews*, 2015, **115**, 525-561.
91. T. R. Branson and W. B. Turnbull, *Chemical Society Reviews*, 2013, **42**, 4613-4622.
92. A. Bernardi, J. Jiménez-Barbero, A. Casnati, C. De Castro, T. Darbre, F. Fieschi, J. Finne, H. Funken, K.-E. Jaeger, M. Lahmann, T. K. Lindhorst, M. Marradi, P. Messner, A. Molinaro, P. V. Murphy, C. Nativi, S. Oscarson, S. Penadés, F. Peri, R. J. Pieters, O. Renaudet, J.-L. Reymond, B. Richichi, J. Rojo, F. Sansone, C. Schäffer, W. B. Turnbull, T. Velasco-Torrijos, S. Vidal, S. Vincent, T. Wennekes, H. Zuilhof and A. Imberty, *Chemical Society Reviews*, 2013, **42**, 4709-4727.
93. N. P. Pera and R. J. Pieters, *MedChemComm*, 2014, **5**, 1027-1035.
94. R. J. Pieters, *Medicinal Research Reviews*, 2007, **27**, 796-816.
95. V. Kumar and W. B. Turnbull, *Beilstein Journal of Organic Chemistry*, 2018, **14**, 484-498.
96. C. L. Schengrund and N. J. Ringler, *Journal of Biological Chemistry*, 1989, **264**, 13233-13237.
97. M. Kato, M. Watanabe-Takahashi, E. Shimizu and K. Nishikawa, *Applied and Environmental Microbiology*, 2015, **81**, 1092-1100.
98. T. Mitsui, M. Watanabe-Takahashi, E. Shimizu, B. Zhang, S. Funamoto, S. Yamasaki and K. Nishikawa, *Infect Immun*, 2016, **84**, 2653-2661.
99. D. J. Stearns-Kurosawa, V. Collins, S. Freeman, D. Debord, K. Nishikawa, S.-Y. Oh, C. S. Leibowitz and S. Kurosawa, *Pediatric Nephrology*, 2011, **26**, 2031-2039.
100. K. Tsutsuki, M. Watanabe-Takahashi, Y. Takenaka, E. Kita and K. Nishikawa, *Infect Immun*, 2013, **81**, 2133-2138.

101. H.-A. Tran, P. I. Kitov, E. Paszkiewicz, J. M. Sadowska and D. R. Bundle, *Organic & Biomolecular Chemistry*, 2011, **9**, 3658-3671.
102. C. S. Mahon, M. A. Fascione, C. Sakonsinsiri, T. E. McAllister, W. Bruce Turnbull and D. A. Fulton, *Organic & Biomolecular Chemistry*, 2015, **13**, 2756-2761.
103. D. Haksar, L. Quarles van Ufford and R. J. Pieters, *Organic & Biomolecular Chemistry*, 2020, **18**, 52-55.
104. C. S. Mahon, G. C. Wildsmith, D. Haksar, E. de Poel, J. M. Beekman, R. J. Pieters, M. E. Webb and W. B. Turnbull, *Faraday Discussions*, 2019, **219**, 112-127.
105. S. Liu and K. L. Kiick, *Macromolecules*, 2008, **41**, 764-772.
106. B. D. Polizzotti and K. L. Kiick, *Biomacromolecules*, 2006, **7**, 483-490.
107. S.-J. Richards, M. W. Jones, M. Hunaban, D. M. Haddleton and M. I. Gibson, *Angewandte Chemie International Edition*, 2012, **51**, 7812-7816.
108. W. B. Turnbull and J. F. Stoddart, *Reviews in Molecular Biotechnology*, 2002, **90**, 231-255.
109. H. M. Branderhorst, R. M. J. Liskamp, G. M. Visser and R. J. Pieters, *Chemical Communications*, 2007, 5043-5045.
110. I. Vrasidas, Nico J. de Mol, Rob M. J. Liskamp and Roland J. Pieters, *European Journal of Organic Chemistry*, 2001, **2001**, 4685-4692.
111. J. P. Thompson and C.-L. Schengrund, *Glycoconjugate Journal*, 1997, **14**, 837-845.
112. J. P. Thompson and C.-L. Schengrund, *Biochemical Pharmacology*, 1998, **56**, 591-597.
113. A. V. Pukin, H. M. Branderhorst, C. Sisu, C. A. G. M. Weijers, M. Gilbert, R. M. J. Liskamp, G. M. Visser, H. Zuilhof and R. J. Pieters, *ChemBioChem*, 2007, **8**, 1500-1503.
114. K. Nishikawa, K. Matsuoka, E. Kita, N. Okabe, M. Mizuguchi, K. Hino, S. Miyazawa, C. Yamasaki, J. Aoki, S. Takashima, Y. Yamakawa, M. Nishijima, D. Terunuma, H. Kuzuhara and Y. Natori, *Proceedings of the National Academy of Sciences*, 2002, **99**, 7669-7674.
115. K. Nishikawa, K. Matsuoka, M. Watanabe, K. Igai, K. Hino, K. Hatano, A. Yamada, N. Abe, D. Terunuma, H. Kuzuhara and Y. Natori, *The Journal of Infectious Diseases*, 2005, **191**, 2097-2105.
116. P. I. Kitov, J. M. Sadowska, G. Mulvey, G. D. Armstrong, H. Ling, N. S. Pannu, R. J. Read and D. R. Bundle, *Nature*, 2000, **403**, 669-672.
117. G. L. Mulvey, P. Marcato, P. I. Kitov, J. Sadowska, D. R. Bundle and G. D. Armstrong, *The Journal of Infectious Diseases*, 2003, **187**, 640-649.
118. E. A. Merritt, Z. Zhang, J. C. Pickens, M. Ahn, W. G. J. Hol and E. Fan, *Journal of the American Chemical Society*, 2002, **124**, 8818-8824.
119. E. Fan, Z. Zhang, W. E. Minke, Z. Hou, C. L. M. J. Verlinde and W. G. J. Hol, *Journal of the American Chemical Society*, 2000, **122**, 2663-2664.
120. Z. Zhang, J. Liu, C. L. M. J. Verlinde, W. G. J. Hol and E. Fan, *The Journal of Organic Chemistry*, 2004, **69**, 7737-7740.
121. V. Kumar, N. Yadav and K. P. R. Kartha, *Carbohydrate Research*, 2016, **431**, 47-55.
122. M. B. Pepys, J. Herbert, W. L. Hutchinson, G. A. Tennent, H. J. Lachmann, J. R. Gallimore, L. B. Lovat, T. Bartfai, A. Alanine, C. Hertel, T. Hoffmann, R. Jakob-Roetne, R. D. Norcross, J. A. Kemp, K. Yamamura, M. Suzuki, G. W. Taylor, S. Murray, D. Thompson, A. Purvis, S. Kolstoe, S. P. Wood and P. N. Hawkins, *Nature*, 2002, **417**, 254-259.
123. J. Liu, Z. Zhang, X. Tan, W. G. J. Hol, C. L. M. J. Verlinde and E. Fan, *Journal of the American Chemical Society*, 2005, **127**, 2044-2045.
124. P. I. Kitov, T. Lipinski, E. Paszkiewicz, D. Solomon, J. M. Sadowska, G. A. Grant, G. L. Mulvey, E. N. Kitova, J. S. Klassen, G. D. Armstrong and D. R. Bundle, *Angewandte Chemie International Edition*, 2008, **47**, 672-676.
125. D. Solomon, P. I. Kitov, E. Paszkiewicz, G. A. Grant, J. M. Sadowska and D. R. Bundle, *Organic Letters*, 2005, **7**, 4369-4372.
126. P. I. Kitov, G. L. Mulvey, T. P. Griener, T. Lipinski, D. Solomon, E. Paszkiewicz, J. M. Jacobson, J. M. Sadowska, M. Suzuki and K.-i. Yamamura, *Proceedings of the National Academy of Sciences*, 2008, **105**, 16837-16842.

127. P. I. Kitov, E. Paszkiewicz, J. M. Sadowska, Z. Deng, M. Ahmed, R. Narain, T. P. Griener, G. L. Mulvey, G. D. Armstrong and D. R. Bundle, *Toxins*, 2011, **3**, 1065-1088.
128. T. R. Branson, T. E. McAllister, J. Garcia-Hartjes, M. A. Fascione, J. F. Ross, S. L. Warriner, T. Wennekes, H. Zuilhof and W. B. Turnbull, *Angewandte Chemie International Edition*, 2014, **53**, 8323-8327.
129. M. J. Krantz, N. A. Holtzman, C. P. Stowell, Y. C. Lee, J. W. Weiner and H. H. Liu, *Biochemistry*, 1976, **15**, 3963-3968.
130. M. E. Webb, R. S. Bon and M. H. Wright, in *Chemical and Biological Synthesis: Enabling Approaches for Understanding Biology*, The Royal Society of Chemistry, 2018, ch. 12, pp. 313-356.
131. G. R. Gray, *Archives of Biochemistry and Biophysics*, 1974, **163**, 426-428.
132. L. C. Paoletti, M. R. Wessels, F. Michon, J. Difabio, H. J. Jennings and D. L. Kasper, *Infect Immun*, 1992, **60**, 4009-4014.
133. L. C. Paoletti, M. R. Wessels, A. K. Rodewald, A. A. Shroff, H. J. Jennings and D. L. Kasper, *Infect Immun*, 1994, **62**, 3236-3243.
134. M. R. Wessels, L. C. Paoletti, D. L. Kasper, J. L. DiFabio, F. Michon, K. Holme and H. J. Jennings, *The Journal of Clinical Investigation*, 1990, **86**, 1428-1433.
135. H. J. Jennings and C. Lugowski, *The Journal of Immunology*, 1981, **127**, 1011.
136. L. J. Rubinstein, P. A. García-Ojeda, F. Michon, H. J. Jennings and K. E. Stein, *Infect Immun*, 1998, **66**, 5450-5456.
137. S. F. Slovin, G. Ragupathi, S. Adluri, G. Ungers, K. Terry, S. Kim, M. Spassova, W. G. Bornmann, M. Fazzari, L. Dantis, K. Olkiewicz, K. O. Lloyd, P. O. Livingston, S. J. Danishefsky and H. I. Scher, *Proc Natl Acad Sci U S A*, 1999, **96**, 5710.
138. T. Gilewski, G. Ragupathi, S. Bhuta, L. J. Williams, C. Musselli, X.-F. Zhang, K. P. Bencsath, K. S. Panageas, J. Chin, C. A. Hudis, L. Norton, A. N. Houghton, P. O. Livingston and S. J. Danishefsky, *Proc Natl Acad Sci U S A*, 2001, **98**, 3270.
139. M. T. Bilodeau, T. K. Park, S. Hu, J. T. Randolph, S. J. Danishefsky, P. O. Livingston and S. Zhang, *Journal of the American Chemical Society*, 1995, **117**, 7840-7841.
140. D. B. Werz, B. Castagner and P. H. Seeberger, *Journal of the American Chemical Society*, 2007, **129**, 2770-2771.
141. Y. C. Lee, C. P. Stowell and M. J. Krantz, *Biochemistry*, 1976, **15**, 3956-3963.
142. J. H. Pazur, B. Liu, N. Q. Li and Y. C. Lee, *Journal of Protein Chemistry*, 1990, **9**, 143-150.
143. F. Berti and R. Adamo, *Chemical Society Reviews*, 2018, **47**, 9015-9025.
144. A. Miseta and P. Csutora, *Molecular Biology and Evolution*, 2000, **17**, 1232-1239.
145. N. J. Davis and S. L. Flitsch, *Tetrahedron Letters*, 1991, **32**, 6793-6796.
146. S. Y. Wong, G. R. Guile, R. A. Dwek and G. Arsequell, *Biochemical Journal*, 1994, **300**, 843-850.
147. B. G. Davis, R. C. Lloyd and J. B. Jones, *The Journal of Organic Chemistry*, 1998, **63**, 9614-9615.
148. B. G. Davis, M. A. T. Maughan, M. P. Green, A. Ullman and J. B. Jones, *Tetrahedron: Asymmetry*, 2000, **11**, 245-262.
149. B. G. Davis, R. C. Lloyd and J. B. Jones, *Bioorganic & Medicinal Chemistry*, 2000, **8**, 1527-1535.
150. B. G. Davis, *Chemical Communications*, 2001, 351-352.
151. W. M. Macindoe, A. H. van Oijen and G.-J. Boons, *Chemical Communications*, 1998, DOI: 10.1039/A708701B, 847-848.
152. I. Shin, H.-j. Jung and M.-r. Lee, *Tetrahedron Letters*, 2001, **42**, 1325-1328.
153. K. Lang and J. W. Chin, *Chemical Reviews*, 2014, **114**, 4764-4806.
154. J. M. Chalker and B. G. Davis, *Current Opinion in Chemical Biology*, 2010, **14**, 781-789.
155. K. E. Neet and D. E. Koshland, Jr., *Proc Natl Acad Sci U S A*, 1966, **56**, 1606-1611.
156. D. H. Strumeyer, W. N. White and D. E. Koshland, Jr., *Proc Natl Acad Sci U S A*, 1963, **50**, 931-935.
157. J. Wang, S. M. Schiller and P. G. Schultz, *Angewandte Chemie International Edition*, 2007, **46**, 6849-6851.

158. D. H. Rich, J. Tam, P. Mathiaparanam, J. A. Grant and C. Mabuni, *Chemical Communications*, 1974, 897-898.
159. G. J. L. Bernardes, J. M. Chalker, J. C. Errey and B. G. Davis, *Journal of the American Chemical Society*, 2008, **130**, 5052-5053.
160. P. M. Morrison, P. J. Foley, S. L. Warriner and M. E. Webb, *Chemical Communications*, 2015, **51**, 13470-13473.
161. J. M. Chalker, S. B. Gunnoo, O. Boutureira, S. C. Gerstberger, M. Fernández-González, G. J. L. Bernardes, L. Griffin, H. Hailu, C. J. Schofield and B. G. Davis, *Chemical Science*, 2011, **2**, 1666-1676.
162. K. Hashimoto, M. Sakai, T. Okuno and H. Shirahama, *Chemical Communications*, 1996, DOI: 10.1039/CC9960001139, 1139-1140.
163. A. Yang, S. Ha, J. Ahn, R. Kim, S. Kim, Y. Lee, J. Kim, D. Söll, H.-Y. Lee and H.-S. Park, *Science*, 2016, **354**, 623.
164. T. H. Wright, B. J. Bower, J. M. Chalker, G. J. L. Bernardes, R. Wiewiora, W.-L. Ng, R. Raj, S. Faulkner, M. R. J. Vallée, A. Phantumrathiwath, O. D. Coleman, M.-L. Thézénas, M. Khan, S. R. G. Galan, L. Lercher, M. W. Schombs, S. Gerstberger, M. E. Palm-Espling, A. J. Baldwin, B. M. Kessler, T. D. W. Claridge, S. Mohammed and B. G. Davis, *Science*, 2016, **354**, aag1465.
165. J. M. Chalker, G. J. L. Bernardes and B. G. Davis, *Accounts of Chemical Research*, 2011, **44**, 730-741.
166. J. Dadová, S. R. G. Galan and B. G. Davis, *Current Opinion in Chemical Biology*, 2018, **46**, 71-81.
167. G. Cohen and D. Cowie, *Comptes rendus hebdomadaires des seances de l'Academie des sciences*, 1957, **244**, 680-683.
168. C. C. Liu and P. G. Schultz, *Annual Review of Biochemistry*, 2010, **79**, 413-444.
169. J. C. M. van Hest, K. L. Kiick and D. A. Tirrell, *Journal of the American Chemical Society*, 2000, **122**, 1282-1288.
170. Y. Tang, G. Ghirlanda, W. A. Petka, T. Nakajima, W. F. DeGrado and D. A. Tirrell, *Angewandte Chemie International Edition*, 2001, **40**, 1494-1496.
171. E. Yoshikawa, M. J. Fournier, T. L. Mason and D. A. Tirrell, *Macromolecules*, 1994, **27**, 5471-5475.
172. C. Renner, S. Alefelder, J. H. Bae, N. Budisa, R. Huber and L. Moroder, *Angewandte Chemie International Edition*, 2001, **40**, 923-925.
173. C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith and P. G. Schultz, *Science*, 1989, **244**, 182.
174. L. Wang, A. Brock, B. Herberich and P. G. Schultz, *Science*, 2001, **292**, 498-500.
175. J. W. Chin, *Annual Review of Biochemistry*, 2014, **83**, 379-408.
176. K. L. Kiick, E. Saxon, D. A. Tirrell and C. R. Bertozzi, *Proc Natl Acad Sci U S A*, 2002, **99**, 19.
177. J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang and P. G. Schultz, *Journal of the American Chemical Society*, 2002, **124**, 9026-9027.
178. A. Deiters and P. G. Schultz, *Bioorganic & Medicinal Chemistry Letters*, 2005, **15**, 1521-1524.
179. D. P. Nguyen, H. Lusic, H. Neumann, P. B. Kapadnis, A. Deiters and J. W. Chin, *Journal of the American Chemical Society*, 2009, **131**, 8720-8721.
180. S. Wallace and J. W. Chin, *Chemical Science*, 2014, **5**, 1742-1744.
181. S. Ye, C. Köhrer, T. Huber, M. Kazmi, P. Sachdev, E. C. Y. Yan, A. Bhagat, U. L. RajBhandary and T. P. Sakmar, *Journal of Biological Chemistry*, 2008, **283**, 1525-1533.
182. J. W. Chin, T. A. Cropp, J. C. Anderson, M. Mukherji, Z. Zhang and P. G. Schultz, *Science*, 2003, **301**, 964.
183. C. D. Spicer and B. G. Davis, *Nature Communications*, 2014, **5**, 4740.
184. J. C. Jackson, S. P. Duffy, K. R. Hess and R. A. Mehl, *Journal of the American Chemical Society*, 2006, **128**, 11124-11127.
185. R. J. Blizzard, D. R. Backus, W. Brown, C. G. Bazewicz, Y. Li and R. A. Mehl, *Journal of the American Chemical Society*, 2015, **137**, 10044-10047.

186. H. S. Jang, X. Gu, R. B. Cooley, J. J. Porter, R. L. Henson, T. Willi, J. A. DiDonato, S. L. Hazen and R. A. Mehl, *ACS Chemical Biology*, 2020, **15**, 562-574.
187. H. C. Hang, C. Yu, D. L. Kato and C. R. Bertozzi, *Proc Natl Acad Sci U S A*, 2003, **100**, 14846.
188. J. A. Prescher and C. R. Bertozzi, *Nature Chemical Biology*, 2005, **1**, 13-21.
189. E. M. Sletten and C. R. Bertozzi, *Accounts of Chemical Research*, 2011, **44**, 666-676.
190. R. S. Loka, C. M. Sadek, N. A. Romaniuk and C. W. Cairo, *Bioconjugate Chemistry*, 2010, **21**, 1842-1849.
191. G. J. L. Bernardes, L. Linderoth, K. J. Doores, O. Boutureira and B. G. Davis, *ChemBioChem*, 2011, **12**, 1383-1386.
192. V. V. Rostovtsev, L. G. Green, V. V. Fokin and K. B. Sharpless, *Angewandte Chemie International Edition*, 2002, **41**, 2596-2599.
193. H. C. Kolb, M. G. Finn and K. B. Sharpless, *Angewandte Chemie International Edition*, 2001, **40**, 2004-2021.
194. S. I. van Kasteren, H. B. Kramer, H. H. Jensen, S. J. Campbell, J. Kirkpatrick, N. J. Oldham, D. C. Anthony and B. G. Davis, *Nature*, 2007, **446**, 1105-1109.
195. O. Boutureira, F. D'Hooge, M. Fernández-González, G. J. L. Bernardes, M. Sánchez-Navarro, J. R. Koeppe and B. G. Davis, *Chemical Communications*, 2010, **46**, 8142-8144.
196. R. Ribeiro-Viana, M. Sánchez-Navarro, J. Luczkowiak, J. R. Koeppe, R. Delgado, J. Rojo and B. G. Davis, *Nature Communications*, 2012, **3**, 1303.
197. Q.-Y. Hu, M. Allan, R. Adamo, D. Quinn, H. Zhai, G. Wu, K. Clark, J. Zhou, S. Ortiz, B. Wang, E. Danieli, S. Crotti, M. Tontini, G. Brogioni and F. Berti, *Chemical Science*, 2013, **4**, 3827-3832.
198. C. J. Pickens, S. N. Johnson, M. M. Pressnall, M. A. Leon and C. J. Berkland, *Bioconjugate Chemistry*, 2018, **29**, 686-701.
199. V. Hong, S. I. Presolski, C. Ma and M. G. Finn, *Angewandte Chemie International Edition*, 2009, **48**, 9879-9883.
200. M. F. Debets, S. S. van Berkel, J. Dommerholt, A. J. Dirks, F. P. J. T. Rutjes and F. L. van Delft, *Accounts of Chemical Research*, 2011, **44**, 805-815.
201. T. Machida, K. Lang, L. Xue, J. W. Chin and N. Winssinger, *Bioconjugate Chemistry*, 2015, **26**, 802-806.
202. E. C. Rodriguez, K. A. Winans, D. S. King and C. R. Bertozzi, *Journal of the American Chemical Society*, 1997, **119**, 9905-9906.
203. P. Durieux, J. Fernandez-Carneado and G. Tuchscherer, *Tetrahedron Letters*, 2001, **42**, 2297-2299.
204. F. Peri, P. Dumy and M. Mutter, *Tetrahedron*, 1998, **54**, 12269-12278.
205. L. A. Marcaurelle, E. C. Rodriguez and C. R. Bertozzi, *Tetrahedron Letters*, 1998, **39**, 8417-8420.
206. S. Peluso and B. Imperiali, *Tetrahedron Letters*, 2001, **42**, 2085-2087.
207. I. S. Carrico, B. L. Carlson and C. R. Bertozzi, *Nature Chemical Biology*, 2007, **3**, 321-322.
208. D. Rabuka, J. S. Rush, G. W. deHart, P. Wu and C. R. Bertozzi, *Nat Protoc*, 2012, **7**, 1052-1067.
209. E. L. Smith, J. P. Giddens, A. T. Iavarone, K. Godula, L.-X. Wang and C. R. Bertozzi, *Bioconjugate Chemistry*, 2014, **25**, 788-795.
210. J. E. Hudak, H. H. Yu and C. R. Bertozzi, *Journal of the American Chemical Society*, 2011, **133**, 16127-16135.
211. L. Wang, Z. Zhang, A. Brock and P. G. Schultz, *Proc Natl Acad Sci U S A*, 2003, **100**, 56.
212. H. Liu, L. Wang, A. Brock, C.-H. Wong and P. G. Schultz, *Journal of the American Chemical Society*, 2003, **125**, 1702-1703.
213. M. Fernández-González, O. Boutureira, G. J. L. Bernardes, J. M. Chalker, M. A. Young, J. C. Errey and B. G. Davis, *Chemical Science*, 2010, **1**, 709-715.
214. T. W. D. F. Rising, C. D. Heidecke, J. W. B. Moir, Z. Ling and A. J. Fairbanks, *Chemistry – A European Journal*, 2008, **14**, 6444-6464.

215. F. Schwarz, W. Huang, C. Li, B. L. Schulz, C. Lizak, A. Palumbo, S. Numao, D. Neri, M. Aebi and L.-X. Wang, *Nature Chemical Biology*, 2010, **6**, 264-266.
216. B. Li, H. Song, S. Hauser and L.-X. Wang, *Organic Letters*, 2006, **8**, 3081-3084.
217. C. D. Heidecke, Z. Ling, N. C. Bruce, J. W. B. Moir, T. B. Parsons and A. J. Fairbanks, *ChemBioChem*, 2008, **9**, 2045-2051.
218. A. J. Fairbanks, *Chemical Society Reviews*, 2017, **46**, 5128-5146.
219. L.-X. Wang, H. Song, S. Liu, H. Lu, S. Jiang, J. Ni and H. Li, *ChemBiochem : a European Journal of Chemical Biology*, 2005, **6**, 1068-1074.
220. K. Naruchi, T. Hamamoto, M. Kurogochi, H. Hinou, H. Shimizu, T. Matsushita, N. Fujitani, H. Kondo and S.-I. Nishimura, *The Journal of Organic Chemistry*, 2006, **71**, 9609-9621.
221. T. Matsushita, H. Hinou, M. Fumoto, M. Kurogochi, N. Fujitani, H. Shimizu and S.-I. Nishimura, *The Journal of Organic Chemistry*, 2006, **71**, 3051-3063.
222. K. M. Koeller, M. E. B. Smith and C.-H. Wong, *Bioorganic & Medicinal Chemistry*, 2000, **8**, 1017-1025.
223. S. Liese and R. R. Netz, *ACS Nano*, 2018, **12**, 4140-4147.
224. J. V. Staros, *Biochemistry*, 1982, **21**, 3950-3955.
225. A. J. Lomant and G. Fairbanks, *Journal of Molecular Biology*, 1976, **104**, 243-261.
226. P. S. R. Anjaneyulu and J. V. Staros, *International Journal of Peptide and Protein Research*, 1987, **30**, 117-124.
227. G. N. Zecherle, A. Oleinikov and R. R. Traut, *Biochemistry*, 1992, **31**, 9526-9532.
228. E. Friedmann, D. H. Marrian and I. Simonreuss, *Br J Pharmacol Chemother*, 1949, **4**, 105-108.
229. M. D. Partis, D. G. Griffiths, G. C. Roberts and R. B. Beechey, *Journal of Protein Chemistry*, 1983, **2**, 263-277.
230. J. Carlsson, H. Drevin and R. Axén, *Biochem J*, 1978, **173**, 723-737.
231. G. Wu, R. F. Barth, W. Yang, S. Kawabata, L. Zhang and K. Green-Church, *Molecular Cancer Therapeutics*, 2006, **5**, 52.
232. W. Yang, R. F. Barth, G. Wu, S. Kawabata, T. J. Sferra, A. K. Bandyopadhyaya, W. Tjarks, A. K. Ferketich, M. L. Moeschberger, P. J. Binns, K. J. Riley, J. A. Coderre, M. J. Ciesielski, R. A. Fenstermaker and C. J. Wikstrand, *Clinical Cancer Research*, 2006, **12**, 3792.
233. S. Chirayil, R. Chirayil and K. J. Luebke, *Nucleic Acids Research*, 2009, **37**, 5486-5497.
234. R. J. Spears and M. A. Fascione, *Organic & Biomolecular Chemistry*, 2016, **14**, 7622-7638.
235. R. Fields and H. Dixon, *Biochemical Journal*, 1968, **108**, 883-887.
236. J. D. Cohen, P. Zou and A. Y. Ting, *ChemBioChem*, 2012, **13**, 888-894.
237. L. Wang and P. G. Schultz, *Angewandte Chemie International Edition*, 2005, **44**, 34-66.
238. K. F. Geoghegan and J. G. Stroh, *Bioconjugate Chemistry*, 1992, **3**, 138-146.
239. T. Sasaki, K. Kodama, H. Suzuki, S. Fukuzawa and K. Tachibana, *Bioorganic & Medicinal Chemistry Letters*, 2008, **18**, 4550-4553.
240. M.-J. Han, D.-C. Xiong and X.-S. Ye, *Chemical Communications*, 2012, **48**, 11079-11081.
241. E. Fischer, *Berichte der deutschen chemischen Gesellschaft*, 1888, **21**, 2631-2634.
242. A. V. Gudmundsdottir and M. Nitz, *Carbohydrate Research*, 2007, **342**, 749-752.
243. B. Bendiak, *Carbohydrate Research*, 1997, **304**, 85-90.
244. J. Kalia and R. T. Raines, *Angewandte Chemie International Edition*, 2008, **47**, 7523-7526.
245. J. W. Haas and R. E. Kadunce, *Journal of the American Chemical Society*, 1962, **84**, 4910-4913.
246. M. B. Thygesen, H. Munch, J. Sauer, E. Cló, M. R. Jørgensen, O. Hindsgaul and K. J. Jensen, *The Journal of Organic Chemistry*, 2010, **75**, 1752-1755.
247. K. Rose, *Journal of the American Chemical Society*, 1994, **116**, 30-33.
248. H. F. Gaertner, K. Rose, R. Cotton, D. Timms, R. Camble and R. E. Offord, *Bioconjugate Chemistry*, 1992, **3**, 262-268.

249. W. Pigman, E. A. Cleveland, D. H. Couch and J. H. Cleveland, *Journal of the American Chemical Society*, 1951, **73**, 1976-1979.
250. D. K. Kölmel and E. T. Kool, *Chemical Reviews*, 2017, **117**, 10358-10376.
251. P. Crisalli and E. T. Kool, *The Journal of Organic Chemistry*, 2013, **78**, 1184-1189.
252. W. P. Jencks, *Journal of the American Chemical Society*, 1959, **81**, 475-481.
253. M. Wendeler, L. Grinberg, X. Wang, P. E. Dawson and M. Baca, *Bioconjugate Chemistry*, 2014, **25**, 93-101.
254. M. Østergaard, N. J. Christensen, C. T. Hjuler, K. J. Jensen and M. B. Thygesen, *Bioconjugate Chemistry*, 2018, **29**, 1219-1230.
255. E. Cló, O. Blixt and K. J. Jensen, *European Journal of Organic Chemistry*, 2010, **2010**, 540-554.
256. J. Dommerholt, S. Schmidt, R. Temming, L. J. Hendriks, F. P. Rutjes, J. C. van Hest, D. J. Lefeber, P. Friedl and F. L. van Delft, *Angewandte Chemie International Edition*, 2010, **49**, 9422-9425.
257. S. A. Meeuwissen, M. F. Debets and J. C. M. van Hest, *Polymer Chemistry*, 2012, **3**, 1783-1795.
258. E. H. P. Leunissen, M. H. L. Meuleners, J. M. M. Verkade, J. Dommerholt, J. G. J. Hoenderop and F. L. van Delft, *ChemBioChem*, 2014, **15**, 1446-1451.
259. M. Glassner, S. Maji, V. R. de la Rosa, N. Vanparijs, K. Ryskulova, B. G. De Geest and R. Hoogenboom, *Polymer Chemistry*, 2015, **6**, 8354-8359.
260. V. VanRheenen, R. C. Kelly and D. Y. Cha, *Tetrahedron Letters*, 1976, **17**, 1973-1976.
261. R. B. Woodward and F. V. Brutcher, *Journal of the American Chemical Society*, 1958, **80**, 209-211.
262. K. McReynolds, D. Dimas, G. Floyd and K. Zeman, *Pharmaceuticals (Basel)*, 2019, **12**, 39.
263. G.-P. Wang, *Chemical Research in Chinese Universities*, 2011, 41.
264. A. V. Gudmundsdottir, C. E. Paul and M. Nitz, *Carbohydrate Research*, 2009, **344**, 278-284.
265. Y. Liu, T. Feizi, M. A. Campanero-Rhodes, R. A. Childs, Y. Zhang, B. Mulloy, P. G. Evans, H. M. I. Osborn, D. Otto, P. R. Crocker and W. Chai, *Chemistry & Biology*, 2007, **14**, 847-859.
266. L. A. Marcaurelle, Y. Shin, S. Goon and C. R. Bertozzi, *Organic Letters*, 2001, **3**, 3691-3694.
267. K. Godula, D. Rabuka, K. T. Nam and C. R. Bertozzi, *Angewandte Chemie International Edition*, 2009, **48**, 4973-4976.
268. M. Ito and T. Yamagata, *Journal of Biological Chemistry*, 1986, **261**, 14278-14282.
269. M. D. Vaughan, K. Johnson, S. DeFrees, X. Tang, R. A. J. Warren and S. G. Withers, *Journal of the American Chemical Society*, 2006, **128**, 6300-6301.
270. S. Sabesan, K. Bock and R. U. Lemieux, *Canadian Journal of Chemistry*, 1984, **62**, 1034-1045.
271. R. L. Ong and R. K. Yu, *Archives of Biochemistry and Biophysics*, 1986, **245**, 157-166.
272. K. P. R. Kartha and H. J. Jennings, *Journal of Carbohydrate Chemistry*, 1990, **9**, 777-781.
273. T. Tanaka, H. Nagai, M. Noguchi, A. Kobayashi and S.-i. Shoda, *Chemical Communications*, 2009, 3378-3379.
274. D. Shapiro and A. J. Acher, *Chemistry and Physics of Lipids*, 1978, **22**, 197-206.
275. C. Wang, Q. Li, H. Wang, L.-H. Zhang and X.-S. Ye, *Tetrahedron*, 2006, **62**, 11657-11662.
276. H. D. L. Karina Persson, Manuela Dieckelmann, Warren W. Wakarchuk, Stephen G. Withers and a. N. C. J. Strynadka, *Nat. Struc. Biol.*, 2001, **8**, 166-175.
277. A. C. Warren W. Wakarchuk, David C. Watson, N. Martin Young, *Protein Engineering*, 1998, **11**, 295-302.
278. J. Zhang, P. Kowal, J. Fang, P. Andreana and P. G. Wang, *Carbohydrate Research*, 2002, **337**, 969-976.
279. M. Ema, Y. Xu, S. Gehrke and G. K. Wagner, *MedChemComm*, 2018, **9**, 131-137.

280. R. Shen, S. Wang, X. Ma, J. Xian, J. Li, L. Zhang and P. Wang, *Biochemistry (Moscow)*, 2010, **75**, 944-950.
281. A. Zervosen and L. Elling, *Journal of the American Chemical Society* 1996, **118**, 1836-1840.
282. J. L. Haigh, D. J. Williamson, E. Poole, Y. Guo, D. Zhou, M. E. Webb, S. A. Deuchars, J. Deuchars and W. B. Turnbull, *Chemical Communications*, 2020, **56**, 6098-6101.
283. Y. H. Lau and D. R. Spring, *Synlett*, 2011, **2011**, 1917-1919.
284. N. Fischer, E. D. Goddard-Borger, R. Greiner, T. M. Klapötke, B. W. Skelton and J. Stierstorfer, *The Journal of Organic Chemistry*, 2012, **77**, 1760-1764.
285. B. Wiltschi, in *Synthetic Gene Networks: Methods and Protocols*, eds. W. Weber and M. Fussenegger, Humana Press, Totowa, NJ, 2012, vol. 813, ch. 13, pp. 211-225.
286. J. Chen, W. Zeng, R. Offord and K. Rose, *Bioconjugate Chemistry*, 2003, **14**, 614-618.
287. R. L. Brabham, T. Keenan, A. Husken, J. Bilsborrow, R. McBerney, V. Kumar, W. B. Turnbull and M. A. Fascione, *Organic & Biomolecular Chemistry*, 2020, **18**, 4000-4003.
288. D. L. MacLeod and C. L. Gyles, *Canadian Journal of Microbiology*, 1989, **35**, 623-629.
289. D. J. Evans, D. G. Evans and S. L. Gorbach, *Infect Immun*, 1973, **8**, 725.
290. H. Arimitsu, K. Sasaki, H. Kojima, T. Yanaka and T. Tsuji, *PLOS ONE*, 2013, **8**, e83577.
291. D. H. Mundell, C. R. Anselmo and R. M. Wishnow, *Infect Immun*, 1976, **14**, 383-388.
292. T. M. Zollman, P. R. Austin, P. E. Jablonski and C. J. Hovde, *Protein Expression and Purification*, 1994, **5**, 291-295.
293. P. R. Austin and C. J. Hovde, *Protein Expression and Purification*, 1995, **6**, 771-779.
294. D. W. Acheson, S. B. Calderwood, S. A. Boyko, L. L. Lincicome, A. V. Kane, A. Donohue-Rolfe and G. T. Keusch, *Infect Immun*, 1993, **61**, 1098.
295. G. S. Thompson, H. Shimizu, S. W. Homans and A. Donohue-Rolfe, *Biochemistry*, 2000, **39**, 13153-13156.
296. M. T. Tarragó-Trani and B. Storrie, *Protein Expression and Purification*, 2004, **38**, 170-176.
297. H. Nakajima, Y. U. Katagiri, N. Kiyokawa, T. Taguchi, T. Suzuki, T. Sekino, K. Mimori, M. Saito, H. Nakao, T. Takeda and J. Fujimoto, *Protein expression and purification*, 2001, **22**, 267-275.
298. J. Porath, T. L. Sjöström and J.-C. Janson, *Journal of Chromatography A*, 1975, **103**, 49-62.
299. H. Tomoda, M. Arai, N. Koyama, H. Matsui, S. Ōmura, R. Obata and Y. C. Lee, *Analytical Biochemistry*, 2002, **311**, 50-56.
300. S. B. Calderwood, D. W. Acheson, M. B. Goldberg, S. A. Boyko and A. Donohue-Rolfe, *Infect Immun*, 1990, **58**, 2977.
301. D. Adlercreutz, J. T. Weadge, B. O. Petersen, J. Ø. Duus, N. J. Dovichi and M. M. Palcic, *Carbohydrate research*, 2010, **345**, 1384-1388.
302. N. Srinivasan, H. E. White, J. Emsley, S. P. Wood, M. B. Pepys and T. L. Blundell, *Structure*, 1994, **2**, 1017-1027.
303. N. R. Sproston and J. J. Ashworth, *Front Immunol*, 2018, **9**, 754-754.
304. M. B. Pepys, D. R. Booth, W. L. Hutchinson, J. R. Gallimore, I. M. Collins and E. Hohenester, *Amyloid*, 1997, **4**, 274-295.
305. D. Pilling and R. H. Gomer, *Front Immunol*, 2018, **9**, 2328.
306. U. Ilangovan, H. Ton-That, J. Iwahara, O. Schneewind and R. T. Clubb, *Proc Natl Acad Sci U S A*, 2001, **98**, 6056-6061.
307. S. K. Mazmanian, G. Liu, H. Ton-That and O. Schneewind, *Science*, 1999, **285**, 760.
308. H. Ton-That, G. Liu, S. K. Mazmanian, K. F. Faull and O. Schneewind, *Proc Natl Acad Sci U S A*, 1999, **96**, 12424.
309. D. J. Williamson, M. A. Fascione, M. E. Webb and W. B. Turnbull, *Angewandte Chemie International Edition*, 2012, **51**, 9377-9380.
310. D. Machin, D. Williamson, P. Fisher, V. Miller, G. Wildsmith, J. F. Ross, C. Wasson, A. MacFonald, B. I. Andrews, D. Ungar, W. B. Turnbull and M. Webb, *Sortase-Modified Cholera Toxoids Show Specific Golgi Localization*, 2019.
311. D. J. Williamson, M. E. Webb and W. B. Turnbull, *Nat Protoc*, 2014, **9**, 253-262.

312. P. M. Morrison, M. R. Balmforth, S. W. Ness, D. J. Williamson, M. D. Rugen, W. B. Turnbull and M. E. Webb, *Chembiochem : a European Journal of Chemical Biology*, 2016, **17**, 753-758.
313. C. P. Guimaraes, M. D. Witte, C. S. Theile, G. Bozkurt, L. Kundrat, A. E. Blom and H. L. Ploegh, *Nat Protoc*, 2013, **8**, 1787-1799.
314. R. A. Schwalbe, B. Dahlbäck and G. L. Nelsestuen, *Journal of Biological Chemistry*, 1990, **265**, 21749-21757.
315. J. Emsley, H. E. White, B. P. O'Hara, G. Oliva, N. Srinivasan, I. J. Tickle, T. L. Blundell, M. B. Pepys and S. P. Wood, *Nature*, 1994, **367**, 338-345.
316. H. Yamaguchi and M. Miyazaki, *Biomolecules*, 2014, **4**, 235-251.
317. L. F. Vallejo and U. Rinas, *Microbial Cell Factories*, 2004, **3**, 11.
318. B. M. Seifried, W. Qi, Y. J. Yang, D. J. Mai, W. B. Puryear, J. A. Runstadler, G. Chen and B. D. Olsen, *Bioconjugate Chemistry*, 2020, **31**, 554-566.
319. M. E. C. Caines, M. D. Vaughan, C. A. Tarling, S. M. Hancock, R. A. J. Warren, S. G. Withers and N. C. J. Strynadka, *Journal of Biological Chemistry*, 2007, **282**, 14300-14308.
320. J. F. Ross, G. C. Wildsmith, M. Johnson, D. L. Hurdiss, K. Hollingsworth, R. F. Thompson, M. Mosayebi, C. H. Trinh, E. Paci, A. R. Pearson, M. E. Webb and W. B. Turnbull, *Journal of the American Chemical Society*, 2019, **141**, 5211-5219.

Chapter 10 Appendix

10.1 CTB W88E plasmid map and protein sequence

Created with SnapGene®

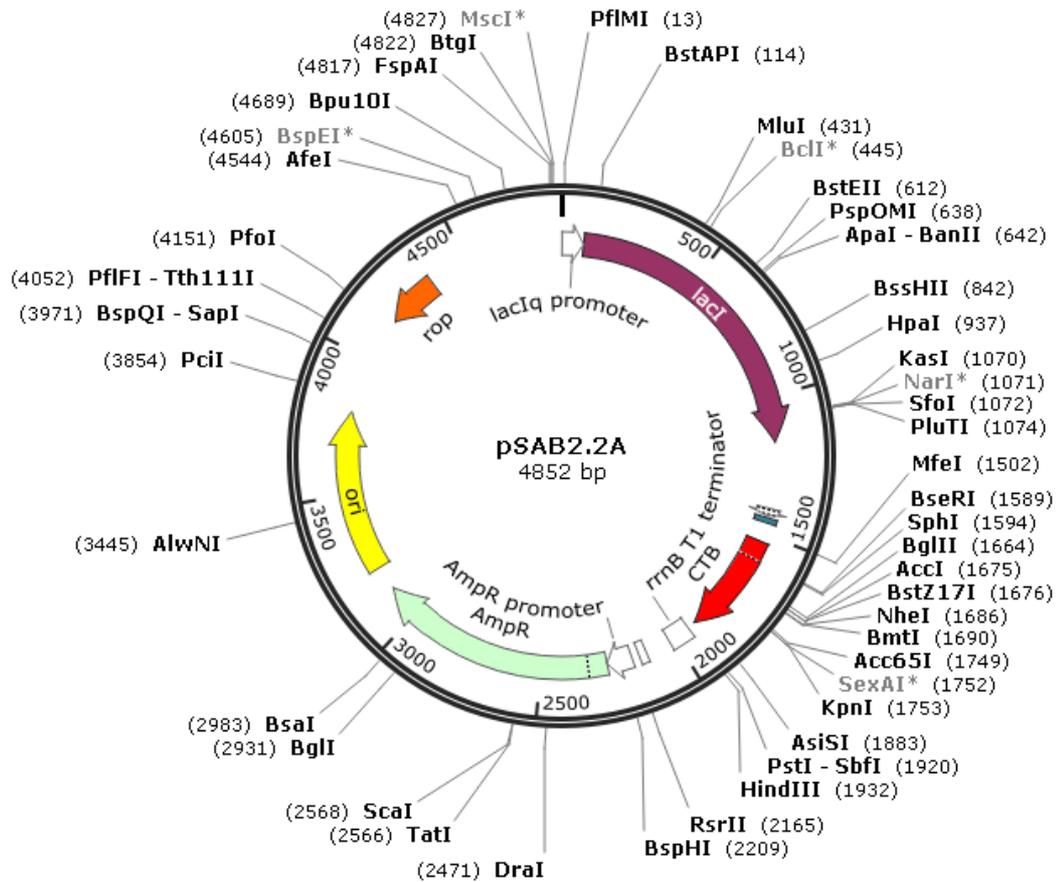


Figure 10.1 – Plasmid map for CTB W88E in the pMAL-p5x expression vector.

The pSAB2.2³²⁰ (figure 10.1) plasmid for expression of CTB-subunit in *E. coli*. It is derived from the pMAL-p5X obtained from Genescript. The plasmid is derived from the plasmid pSAB2.1A³¹⁰ containing the gene for MBP-CTA2 fusion as well as the gene for CTB-subunit (for expressing MBP-CTB), by removal of the gene for MBP-CTA2 fusion protein.

Derivates of this plasmid include pSAB2.2-W88E.¹²⁸

10.1.1 Vector sequence

LTIIb periplasmic targeting sequence

CTB

```
CCGACACCATCGAATGGTGCAAACCTTTTCGCGGTATGGCATGATAGCGCCCGGAAGAG
AGTCAATTCAGGGTGGTGAATGTGAAACCAGTAACGTTATACGATGTCGCAGAGTATGC
CGGTGTCTCTTATCAGACCGTTTCCC GCGTGGTGAACCAGGCCAGCCACGTTTCTGCGA
AAACGCGGGAAAAAGTGGAAGCGGCGATGGCGGAGCTGAATTACATTCCCAACCGCGTG
GCACAACAACCTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCCAGTCTGGC
```

CCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGGTG
CCAGCGTGGTGGTGTGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGTG
CACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTA ACTATCCGCTGGATGACCA
GGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCT
CTGACCAGACACCCATCAACAGTATTATTTTCTCCCATGAAGACGGTACGCGACTGGGC
GTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAG
TTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAATCAAA
TTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACC
ATGCAAATGCTGAATGAGGGCATCGTTC CCACTGCGATGCTGGTTGCCAACGATCAGAT
GGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGATATTT
CGGTAGTGGGATACGACGATACCGAAGACAGCTCATGTTATATCCCGCCGTTAACCACC
ATCAAACAGGATTTTTCGCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTC
TCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAA
CCACCCTGGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATG
CAGCTGGCACGACAGGTTTTCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATG
TAAGTTAGCTCACTCATTAGGCACAATTCTCATGTTTGACAGCTTATCATCGACTGCAC
GGTGCACCAATGCTTCTGGCGTCAGGCAGCCATCGGAAGCTGTGGTATGGCTGTGCAGG
TCGTAAATCACTGCATAATTCGTGTCGCTCAAGGCGCACTCCCGTTCTGGATAATGTTT
TTTGCGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTGACAATTAA
TCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTACACAGGAAACAG
CCAGTCCGTTT TAGGTGTTTTACGAGCAATTGACCAACAAGGACCATAGATTAATGAGCT
TTAAGAAAATTATCAAGGCATTTGTTATCATGGCTGCTTTGGTATCTGTTTCAGGCGCAT
GCAGCTCCTCAAATATTA CTGATTTGTGCGCAGAATACCACAACACAAATATATAC
GCTAAATGATAAGATCTTTTCGTATACAGAATCGCTAGCGGGAAAAGAGAGATGGCTA
TCATTACTTTTAAGAA TGGTGCAATTTTTCAAGTAGAGGTACCAGGTAGTCAACATATA
GATTCACAAAAAAAAGCGATTGAAAGGATGAAGGATACCCTGAGGATTGCATATCTTAC
TGAAGCTAAAGTCGAAAAGTTATGTGTATGGAATAATAAACGCCTCATGCGATCGCCG
CAATTAGTATGGCAA ACTAAGTTTTCCCTGCAGGTAATTAAATAAGCTTCAAATAAAAC
GAAAGGCTCAGTCGAAAGACTGGGCCTTTTCGTTTTATCTGTTGTTTGTGCGGTGAACGCT
CTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCCG
AGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCA
TCCTGACGGATGGCCTTTTTGCGTTTTCTACAAACTCTTTCGGTCCGTTGTTTATTTTTT
TAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATA
ATATTGAAAAGGAAGAGTATGAGTATTC AACATTTCCGTGTCGCCCTTATTCCTTTT
TTGCGGCATTTTGCCTTCCTGTTTTTGTCTACCCAGAAACGCTGGTGAAAGTAAAAGAT

GCTGAAGATCAGTTGGGTGCACGAGTGGGTACATCGAACTGGATCTCAACAGCGGTAA
GATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTCCCAATGATGAGCACTTTTAAAGTTC
TGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTGCGCCG
ATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTAC
GGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAAAGCTG
CGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCAC
AACATGGGGGATCATGTAAGTGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCAT
ACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAAC
TATTAAGTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAG
GCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTTATTGC
TGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAG
ATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGAGTCAGGCAACTATGGAT
GAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTTGGTAACTGTC
AGACCAAGTTTACTCATATATACTTTAGATTGATTTCCCTTAGGACTGAGCGTCAACCCC
GTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGGTAATCTGCTGCTT
GCAAACAAAAAACCACCGCTACCAGCGGTGGTTTTGTTTGCCGGATCAAGAGCTACCAA
CTTTTTTCCGAAGGTAAGTGGCTTACGACAGAGCGCAGATACCAAATACTGTCTTCTA
GTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGC
TCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGT
TGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCCGGCTGAACGGGGGGTTCG
TGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGA
GCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCG
GCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTT
TATAGTCCTGTGGGTTTTGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTC
AGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCT
TTTGCTGGCCTTTTTGCTCACATGTTCTTTTCTGCGTTATCCCCTGATTCTGTGGATAAC
CGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGACCGAACGACCGAGCGCAG
CGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATC
TGTGCGGTATTTACACCCGCATATAAGGTGCACTGTGACTGGGTCATGGCTGCGCCCCG
ACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTT
ACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTACCGTCATCA
CCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGCAGCGATTACA
GATGTCTGCCTGTTTATCCGCGTCCAGCTCGTTGAGTTTTCTCCAGAAGCGTTAATGTCT
GGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCTGTTTGGTCACTGATGCC
TCCGTGTAAGGGGGATTTCTGTTTATGGGGTAATGATACCGATGAAACGAGAGAGGAT

GCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTA
AACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAG
CGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCA
GATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACAC
GGAAACCGAAGACCATTTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCG
CTTCACGTTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCGTAAGGCAACCCCGCCA
GCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGCCAGGACCCAA
CGCTGCCCGAAATT

10.1.2 Primer sequences for the introduction of W88E to pSAB 2.2

5' – GTCGAAAAGTTATGTGTAGAGAATAATAAAACGCCTCATGCGATTGCCG –3'

3' – CAGCTTTTCAATACACATCTTATTATTTTTCGGGAGTACGCTAACGGC –5'

10.1.3 CTB W88E protein sequence

TPQNITDLCAEYHNTQIYTLNDKIFSYTESLAGKREMAIITFKNGAIFQVEVPGSQHID
SQKKAIERMKDTRLRIAYLTEAKVEKLCVENNKTPHAIAAISMAN

10.2 pSAB2.3_azido plasmid map and primers for W88E, for the expression of mutant CTB MET/N₃-W88E

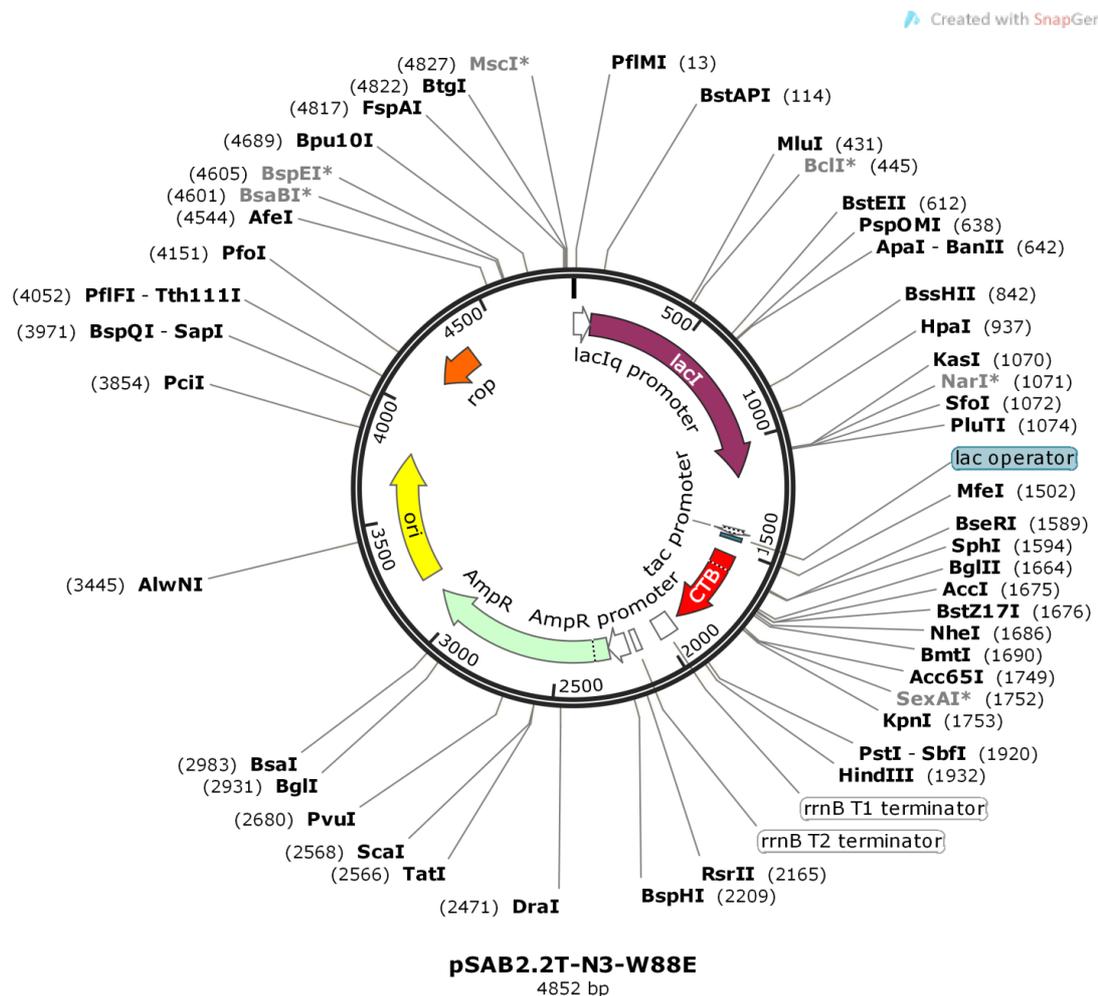


Figure 10.2 – Plasmid map for pSAB2.3(azido) based upon the pMAL-p5x expression vector

The CTB plasmid for expression in *E. coli* named pSAB2.3 (figure 10.2). The plasmid is based upon the pMAL-p5X obtained from Genescript. This plasmid is derived from the plasmid pSAB2.2³²⁰ (prepared by Dr James Ross, University of Leeds) for the expression of CTB and CTB mutants by introducing the mutations M37L, M68L, M101L, K43M (performed by Dr Daniel Williamson, University of Leeds)²⁸².

pSAB2.4 is a derivative of this plasmid which includes the W88E mutation to remove GM1 binding capabilities.

10.2.1 Vector sequence

```
CCGACACCATCGAATGGTGCAAAACCTTTTCGCGGTATGGCATGATAGCGCCCGGAAGAG
AGTCAATTCAGGGTGGTGAATGTGAAACCAGTAACGTTATACGATGTCGCAGAGTATGC
CGGTGTCTCTTATCAGACCGTTTCCCGCGTGGTGAACCAGGCCAGCCACGTTTCTGCGA
```

AAACGCGGGAAAAAGTGGAAGCGGCGATGGCGGAGCTGAATTACATTCCCAACCGCGTG
GCACAACAACCTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCCAGTCTGGC
CCTGCACGCGCCGTTCGCAAATTGTCGCGGGGATTAATCTCGCGCCGATCAACTGGGTG
CCAGCGTGGTGGTGTTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGTG
CACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCA
GGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCT
CTGACCAGACACCCATCAACAGTATTATTTCTCCCATGAAGACGGTACGCGACTGGGC
GTGGAGCATCTGGTTCGATTGGGTACCCAGCAAATCGCGCTGTTAGCGGGCCCATTAAG
TTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAATCAA
TTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACC
ATGCAAATGCTGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAGAT
GGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGC GGATATTT
CGGTAGTGGGATACGACGATACCGAAGACAGCTCATGTTATATCCCGCCGTTAACCACC
ATCAAACAGGATTTTCGCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTC
TCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAA
CCACCCTGGCGCCCAATACGCAAACCGCCTCTCCCCGCGGTTGGCCGATTCATTAATG
CAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATG
TAAGTTAGCTCACTCATTAGGCACAATTCTCATGTTTGACAGCTTATCATCGACTGCAC
GGTGCACCAATGCTTCTGGCGTCAGGCAGCCATCGGAAGCTGTGGTATGGCTGTGCAGG
TCGTAAATCACTGCATAATTCGTGTGCTCAAGGCGCACTCCCGTTCTGGATAATGTTT
TTTGCGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTGACAATTAA
TCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTACACAGGAAACAG
CCAGTCCGTTTAGGTGTTTTACGAGCAATTGACCAACAAGGACCATAGATTATGAGCT
TTAAGAAAATTATCAAGGCATTTGTTATCATGGCTGCTTTGGTATCTGTTTCAGGCGCAT
GCAACTCCTCAAATATTAAGTATTTGTGCGCAGAATACCACAACACAAATATATAC
GCTAAATGATAAGATCTTTTCGTATACAGAATCGCTAGCGGGAAAAAGAGAGCTGGCTA
TCATTACTTTTATGAATGGTGCAATTTTTCAAGTAGAGGTACCAGGTAGTCAACATATA
GATTCACAAAAAAGCGATTGAAAGGCTGAAGGATACCCTGAGGATTGCATATCTTAC
TGAAGCTAAAGTCGAAAAGTTATGTGTAGAGAATAATAAAACGCCTCATGCGATTGCCG
CAATTAGTCTGGCAAAC TAAGTTTTCCCTGCAGGTAATTAAATAAGCTTCAAATAAAAC
GAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTTCGGTGAACGCT
CTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGG
AGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCA
TCCTGACGGATGGCCTTTTTGCGTTTCTACAACTCTTTCGGTCCGTTGTTTATTTTTC
TAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATA

ATATTGAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTT
TTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGAT
GCTGAAGATCAGTTGGGTGCACGAGTGGGTACATCGAACTGGATCTCAACAGCGGTAA
GATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTCCAATGATGAGCACTTTTAAAGTTC
TGCTATGTGGCGCGGTATTATCCCGTGTGACGCCGGGCAAGAGCAACTCGGTGCGCCG
ATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTAC
GGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAAACTG
CGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCAC
AACATGGGGGATCATGTAACCTGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCAT
ACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAAC
TATTAACCTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAG
GCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGTTTTATTGC
TGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAG
ATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGAT
GAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTC
AGACCAAGTTTACTCATATATACTTTAGATTGATTTCCCTTAGGACTGAGCGTCAACCCC
GTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTT
GCAAACAAAAAACCACCGCTACCAGCGGTGGTTTTGTTTGCCGGATCAAGAGCTACCAA
CTTTTTTCCGAAGGTAACCTGGCTTACGACAGAGCGCAGATACCAAATACTGTCTTCTA
GTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGC
TCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGT
TGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTGGGCTGAACGGGGGGTTTCG
TGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGA
GCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCG
GCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTT
TATAGTCCTGTCGGGTTTTCGCCACCTCTGACTTGAGCGTCGATTTTTTGATGCTCGTC
AGGGGGCGGAGCCTATGGAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCT
TTTGCTGGCCTTTTTGCTCACATGTTCTTTCTGCGTTATCCCCTGATTCTGTGGATAAC
CGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAG
CGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATC
TGTGCGGTATTTACACCCGCATATAAGGTGCACTGTGACTGGGTCATGGCTGCGCCCCG
ACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTT
ACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTGTCAGAGGTTTTACCGTCATCA
CCGAAACGCGGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGCAGCGATTACA
GATGTCTGCCTGTTTATCCCGTCCAGCTCGTTGAGTTTTCTCCAGAAGCGTTAATGTCT

GGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTTCCTGTTTGGTCACTGATGCC
TCCGTGTAAGGGGGATTTCTGTTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGAT
GCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTA
AACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAG
CGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCA
GATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTTCCAGACTTTACGAAACAC
GGAAACCGAAGACCATTCATGTTGTTGCTCAGGTGCGCAGACGTTTTGCAGCAGCAGTCG
CTTCACGTTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCCAGTAAGGCAACCCCGCCA
GCCTAGCCGGGTCCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGCCAGGACCCAA
CGCTGCCCGAAATT

10.2.1 Primer sequences for the introduction of W88E to pSAB

2.3(azido)

5' – GTCGAAAAGTTATGTGTAG**GA**GAAATAATAAAACGCCTCATGCGATTGCCG –3'
3' – CAGCTTTTCAATACACAT**CT**CTTATTATTTTGC GGAGTACGCTAACGGC –5'

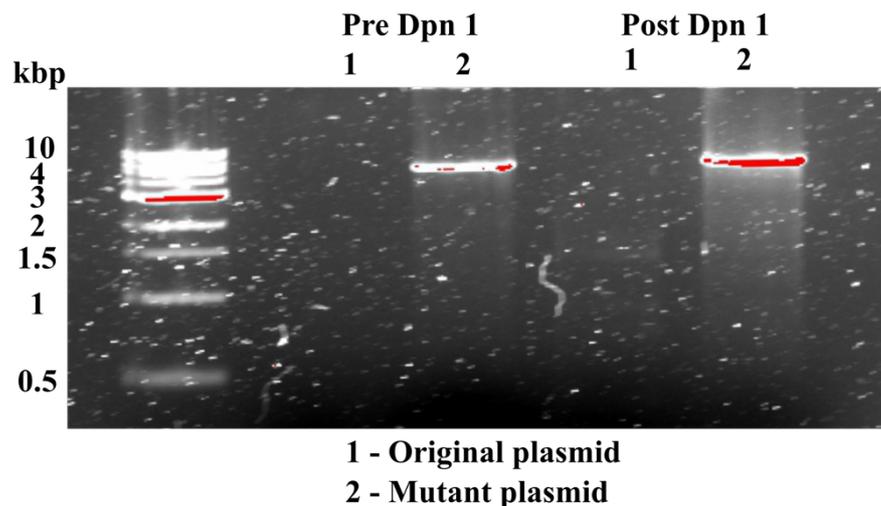


Figure 10.3: Nucleic acid PAGE gel confirming the successful mutagenesis of the pSAB2.3(azido) plasmid to introduce the W88E mutation giving pSAB2.4

10.2.2 Protein sequence

TPQNIITDLCAEYHNTQIYTLNDKIFSYTESLAGKRELAIITFMNGAIFQVEVPGSQHID
SQQKAIERLKDTRLRIAYLLEAKVEKLCVENNKTPHAIAAISLAN

TATGCATCAGGACGTGTATAGTGGCGCCATCACCCCGGAGGGTAACTCCGGCAATGGTG
CCGGTGCAATTGGTAATGGTGCACCGGCTTGGGCGACCTACATGGATGGTCTGCCGGTG
GAACCGCAGCCGCGTTGGGAACTGTATTACATTCAACCGGGCGTTATGCGCGCCTTTGA
TAACTTCTGGAATACCACGGGTAAACATCCGGAAC TGGTTGAACACTATGCCAAAGCAT
GGCGTGCTGTCGCGGATCGCTTTGCAGATAACGACGCGGTGGTTGCCTACGACCTGATG
AATGAACCGTTTGGCGGTTCTCTGCAGGGTCCGGCATT TGAAGCTGGTCCGCTGGCGGC
CATGTATCAACGTACCACGGATGCGATCCGCCAGGTGGATCAAGATACCTGGGTTTGTG
TTGCGCCGCAGGCAATTGGTGTTAACCAAGGCCTGCCGTCAGGTCTGACGAAAATCGAT
GATCCGCGTGCGGGTCAGCAACGTATTGCATATTGTCCGCATCTGTACCCGCTGCCGCT
GGATATCGGTGACGGCCACGAAGGTCTGGCGCGTACCCTGACGGATGTGACCATTGACG
CATGGCGTGCTAATACCGCCCATAACCGCCCGCGTTCTGGGTGATGTCCCGATTATCCTG
GGTGAATTTGGTCTGGATAACCACCTGCCGGGTGCCCGTGA CTATATCGAACGCGTGTA
CGGTACCGCACGCGAAAATGGGTGCTGGCGTTTCATATTGGAGCAGCGATCCGGGTCCGT
GGGGTCCGTACCTGCCGGATGGCACCCAGACGCTGCTGGTGCACACCCTGAATAAACCG
TATCCGCGTGCCGTTGCAGGTACCCCGACGGAATGGTCATCGACCTCGGATCGCCTGCA
GCTGACGATTGAACCGGACGCAGCTATCACCGCACCGACGAAAATTTACCTGCCGGAAG
CTGGTTTCCCGGGCGATGTGCACGTTGAAGGCGCCGATGTCGTGGGTTGGGACCGTCAG
AGCCGCCTGCTGACGGTTTCGCACCCCGGCTGATAGCGGTAATGTCACGGTCACGGTCA
CCCGGCGGCATAACTCGAGCACCACCACCACCACCTGAGATCCGGCTGCTAACAAAG
CCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCCTGAGCAATAACTAGCATAACCCCTT
GGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTTGTCTGAAAGGAGGAACTATATCCGGATT
GGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCG
CAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCTT
CCTTTCTCGCCACGTTTCGCCGGCTTTCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTA
GGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAACTTGATTAGGGTGATGG
TTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTTCGCCCTTTGACGTTGGAGTCCA
CGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTC
TATTCTTTTGATTTATAAGGGATTTTGCCGATTTTCGGCCTATTGGTTAAAAAATGAGCT
GATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTTCAGGTG
GCACTTTTCGGGGAAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCA
AATATGTATCCGCTCATGAATTAATTCTTAGAAAACTCATCGAGCATCAAATGAAACT
GCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAGCCGTTTCTGTAAAT
GAAGGAGAAAAC TACCGAGGCAGTCCATAGGATGGCAAGATCCTGGTATCGGTCTGC
GATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCTCGTCAAAAATAAGGT
TATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTA

TGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACT
CGCATCAACCAAACCGTTATTCATTTCGTGATTGCGCCTGAGCGAGACGAAATACGCGAT
CGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCC
AGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGT
TTTCCCAGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCT
TGATGGTCGGAAGAGGCATAAATTCCGTGAGCCAGTTTAGTCTGACCATCTCATCTGTA
ACATCATTTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACCTCTGGCGCATCGGGCTT
CCCATAACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTAT
ACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCC
CGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTAT
TGTTTCATGACCAAATCCCTTAACGTGAGTTTTTCGTTCCACTGAGCGTCAGACCCCGTA
GAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCA
AACAAAAAACCACCGCTACCAGCGGTGGTTTGTGGCCGGATCAAGAGCTACCAACTC
TTTTTCCGAAGGTAACCTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTG
TAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCT
GCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGG
ACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTTCGTGC
ACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCT
ATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCA
GGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGAAACGCCTGGTATCTTTAT
AGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTTGTGATGCTCGTCAGG
GGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTT
GCTGGCCTTTTGCTCACATGTTCTTTCTGCGTTATCCCCTGATTCTGTGGATAACCGT
ATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGA
GTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGT
GCGGTATTTACACCCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCAT
AGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGAC
ACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGCTGCTCCCGGCATCCGCTTAC
AGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTACCGTCATCACC
GAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTACAGA
TGTCTGCCTGTTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGG
CTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCTGTTTGGTCACTGATGCCTC
CGTGTAAGGGGGATTTCTGTTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGC
TCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAA
CAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCG

CTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGA
TCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGG
AAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTTGCAGCAGCAGTCGCT
TCACGTTGCTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGC
CTAGCCGGGTCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCG
GCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTG
AGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCC
AGCGAAAGCGGTCTCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGT
TGCATGATAAAGAAGACAGTCATAAGTGGCGGACGATAGTCATGCCCCGCGCCACCG
GAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTGAGATCCCGGTGCCTAAT
GAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCAGTCGGGAAAC
CTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTTGCGTAT
TGGGCGCCAGGGTGGTTTTTCTTTTACCAGTGAGACGGGCAACAGCTGATTGCCCTTC
ACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCG
AAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGT
CGTATCCCCTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGC
ATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTC
ATTCAGCATTGTCATGGTTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTT
CCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGA
CGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGC
GACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGA
TGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCC
ACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCCTGACGCGTTG
CGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCG
ACACCACCACGCTGGCACCCAGTTGATCGGCGGAGATTTAATCGCCGCGACAATTTGC
GACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCC
CGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCA
CTTTTTCCCGCGTTTTTCGCAGAAACGTGGCTGGCCTGGTTTACCACGCGGGAAACGGTC
TGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTACATTCAC
CACCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATAACCGCAAAGGTTTTGCGCC
ATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCA
GCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGAAGGAATGGTGCATGCAAGG
AGATGGCGCCCAACAGTCCCCGGCCACGGGGCCTGCCACCATAACCACGCCGAAACAA
GCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCATCGGTGATGTCGGCGATATA

GGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGA
GGATCGAGATCT

10.3.2 Protein sequence

MGSSHHHHHSSGLVPRGSHMSGSGSGSGTALTPSYLKDDDGRSLILRGFNTASSAKSA
PDGMPQFTEADLAREYADMGTNFVRFLISWRSVEPAPGVYDQQYLDRVEDRVGWYAERG
YKVMLDMHQDVYSGAITPEGNSGNGAGAI GNGAPAWATYMDGLPVEPQPRWELYIIPG
VMRAFDNFWNTTGKHPELVEHYAKAWRAVADRFADNDVVAYDLMNEPFGGSLQGPAFE
AGPLAAMYQRTTDAIRQVDQDTWVCVAPQAIGVNQGLPSGLTKIDDPRAGQQR IAYCPH
LYPLPLDIGDGHEGLARTLTDVTIDAWRANTAHTARVLGDVPIILGEFGLD TTLPGARD
YIERVYGTAREMGAGVSYWSSDPGPWGPYLPDGTQTL LVDTLNKPYPRAVAGTPTEWSS
TSDRLQLTIEPDAAITAPTEIYLPEAGFP GDVHVEGADVVGWDRQSRLLTVRTPADSGN
VTVTVTPAA

GGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTCTAAATACATTC
AAATATGTATCCGCTCATGAATTAATTCTTAGAAAACTCATCGAGCATCAAATGAAAC
TGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAA
TGAAGGAGAAAACTCACCGAGGCAGTTCATAGGATGGCAAGATCCTGGTATCGGTCTG
CGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCTCGTCAAAAAATAAGG
TTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAGTTT
ATGCATTTCTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAATCAC
TCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGA
TCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGC
CAGCGCATCAACAATATTTTACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTG
TTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGC
TTGATGGTTCGGAAGAGGCATAAATCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGT
AACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACCTCTGGCGCATCGGGCT
TCCATAACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTA
TACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTC
CCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTA
TTGTTTATGACCAAAATCCCTTAAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGT
AGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGC
AAACAAAAAACACCGCTACCAGCGGTGGTTTTGTTTGCCGGATCAAGAGCTACCAACT
CTTTTTCCGAAGGTAACCTGGCTTACAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGT
GTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTC
TGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTG
GACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCCGGCTGAACGGGGGGTTCGTG
CACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGC
TATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGC
AGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGAAACGCCTGGTATCTTTA
TAGTCCTGTCCGGTTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAG
GGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTT
TGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCG
TATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGACCGGAACGACCGAGCGCAGCG
AGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTG
TGCGGTATTTACACCGCATATATGGTGCACCTCTCAGTACAATCTGCTCTGATGCCGCA
TAGTTAAGCCAGTATACTCCGCTATCGCTACGTGACTGGGTTCATGGCTGCGCCCCGA
CACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTA
CAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTACCGTCATCAC

CGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTACAG
ATGTCTGCCTGTTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTG
GCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTTCTGTTTGGTCACTGATGCCT
CCGTGTAAGGGGGATTTCTGTTCATGGGGTAATGATACCGATGAAACGAGAGAGGATG
CTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAA
ACAACCTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGC
GCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAG
ATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACG
GAAACCGAAGACCATTTCATGTTGTTGCTCAGGTGCGAGACGTTTTGCAGCAGCAGTCGC
TTCACGTTGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAG
CCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCC
GGCGATAATGGCCTGCTTCTCGCCGAAACGTTTTGGTGGCGGGACCAGTGACGAAGGCTT
GAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTC
CAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAG
TTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACC
GGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTGCGAGATCCCGGTGCCTAA
TGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAA
CCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTTGCGTA
TTGGGCGCCAGGGTGGTTTTTTCTTTTACCAGTGAGACGGGCAACAGCTGATTGCCCTT
CACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTTGCCCCAGCAGGC
GAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCG
TCGTATCCCCTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCG
CATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCT
CATTGAGCATTTCATGGTTTTGTTGAAAACCGGACATGGCACTCCAGTCGCTTCCCCT
TCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAG
ACGCGCCGAGACAGAACTTAATGGGCCCCGCTAACAGCGCGATTTGCTGGTGACCCAATG
CGACCAGATGCTCCACGCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTG
ATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTC
CACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTT
GCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATC
GACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTG
CGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGC
CCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCC
ACTTTTTCCCGGTTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGT
CTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTTACATTCA

CCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGC
CATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGC
AGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAG
GAGATGGCGCCCAACAGTCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACA
AGCGCTCATGAGCCGAAGTGGCGAGCCCGATCTTCCCATCGGTGATGTCGGCGATAT
AGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAG
AGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCG
GATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGG
GCAGCAGCCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGC**CATATG**
ATGGACATCGTGTTTTGCCGCAGATGATAACTATGCGGCTTACCTCTGTGTTGCCGCGAA
ATCCGTGAAGCGGCGCACCCGGACACGGAAATTCGTTTTTCATGTTTTGGACGCGGGCA
TCTCAGAAGCGAACCGTGCAGCGGTTGCAGCGAACTTGCGTGCGGTGGCGGTAATATC
CGTTTCATTGACGTGAACCCGGAGGATTTTGTGTTTTCCCCTGAACATTCGTCATAT
CTCGATTACGACGTATGCCCGCCTTAACTGGGCGAATATATTGCGGATTGCGATAAAG
TGCTGTATTTGGATATTGATGTTCTGGTCCGTGATTCTTTGACTCCTCTGTGGGACACG
GACTTGGGCGATAACTGGCTGGGCGCATGCATTGATTTGTTTGTGAGCGACAGGAAGG
GTACAAACAGAAAATCGGTATGGCCGACGGTGAATACTACTTTAATGCAGGCGTCCTGC
TGATTAATTTGAAAAAATGGCGGCGTCACGATATCTTTAAAATGAGCTGCGAGTGGGTA
GAGCAGTATAAAGACGTAATGCAATATCAAGATCAGGATATTCTCAATGGTTTATTCAA
AGGCGGAGTTTGCTACGCGAATAGTCGTTTTCAACTTCATGCCTACCAATTACGCTTTCA
TGGCTAACCGTTTCGCATCGCGTCATACCGACCCGCTGTACCGCGACCGCACTAACACG
GTGATGCCGGTTGCAGTGTCACTATTGCGGTCCGGCTAAACCTTGGCATCGTACTG
TACGGCCTGGGGCGCCGAGCGCTTTACTGAGCTGGCCGGAAGTCTGACCACGGTACCCG
AAGAGTGGCGCGCAAACCTGGCTGTTCCACACTAAGGATCCGAATTCGAGCTCCGTCTGA
CAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCCTGAGATCCGGCTGCTAACA
AAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCC
CTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGG
AT

10.4.2 Protein sequence

GSSHHHHHSSGLVPRGSHMMDIVFAADDNYAAYLCVAAKSVEAAHPDTEIRFHVLDAG
LSEANRAAVAANLRGGGGNIRFIDVNPEDFAGFPLNIRHISITTYARLKLGEYIADCCK
VLYLDIDVLRDSLTPPLWDTDLGDNWLGACIDLFVERQEGYKQKIGMADGEYYFNAGVL
LINLKKWRRHDIKMSCEWVEQYKDVMOYQDQDILNGLFKGGVCYANSRNFNMPNTYAF
MANRFASRHTDPLYRDRNTVMPVAVSHYCGPAKPWHRDCTAWGAERFTELAGSLTTVP
EEWRGKLAVP

10.5 VTB protein sequence

From sequencing data obtained the plasmid was determined to be in the commercially available PUC-18 vector (full plasmid was not sequenced) cloned into the M13 restriction sites. The M-13R primer was used to sequence the full VTB gene (Uniprot – Q8X4M7) within the plasmid.

10.5.1 Reverse compliment of VTB with signalling peptide

TCAACGAAAAATAACTTCGCTGAATCCCCCTCCATTATGACAGGCATTAGTTTTAATGG
TTACAGTCATCCCCGTAATTTGCGCACTGAGAAGAAGAGACTGAAGATTCCATCTGTTG
GTAAATAATTCTTTATCACCCACTTTAACTGTAAAGGTATCGTCATCATTATATTTTGT
ATACTCCACCTTTCAGTTACACAATCAGGCGTCGCCAGCGCACTTGCTGAAAAAATG
AAAGCGATGCAGCTATTAATAAATGTTTTTTTCAT

10.5.2 Forward compliment of VTB with signalling peptide

ATGAAAAACATTTATTAATAGCTGCATCGCTTTCATTTTTTTCAGCAAGTGCGCTGGC
GACGCTGATTGTGTAAGTGGAAAGGTGGAGTATACAAAATATAATGATGACGATACCT
TTACAGTTAAAGTGGGTGATAAAGAATTATTTACCAACAGATGGAATCTTCAGTCTCTT
CTTCTCAGTGCGCAAATTACGGGGATGACTGTAACCATTAAACTAATGCCTGTCATAA
TGGAGGGGGATTACAGCGAAGTATTTTTTCGTTGA

10.5.3 Protein sequence

Periplasmic targeting sequence

MKKTLLIAASLSFFSASALATPDCVTGKVEYTKYNDDDTFTVKVGDKELFTNRWNLQSL
LLSAQITGMTVTIKTNACHNGGGFSEVIFR

10.6 Serum Amyloid P (E167Q) LPETG in pET11a in protein sequence

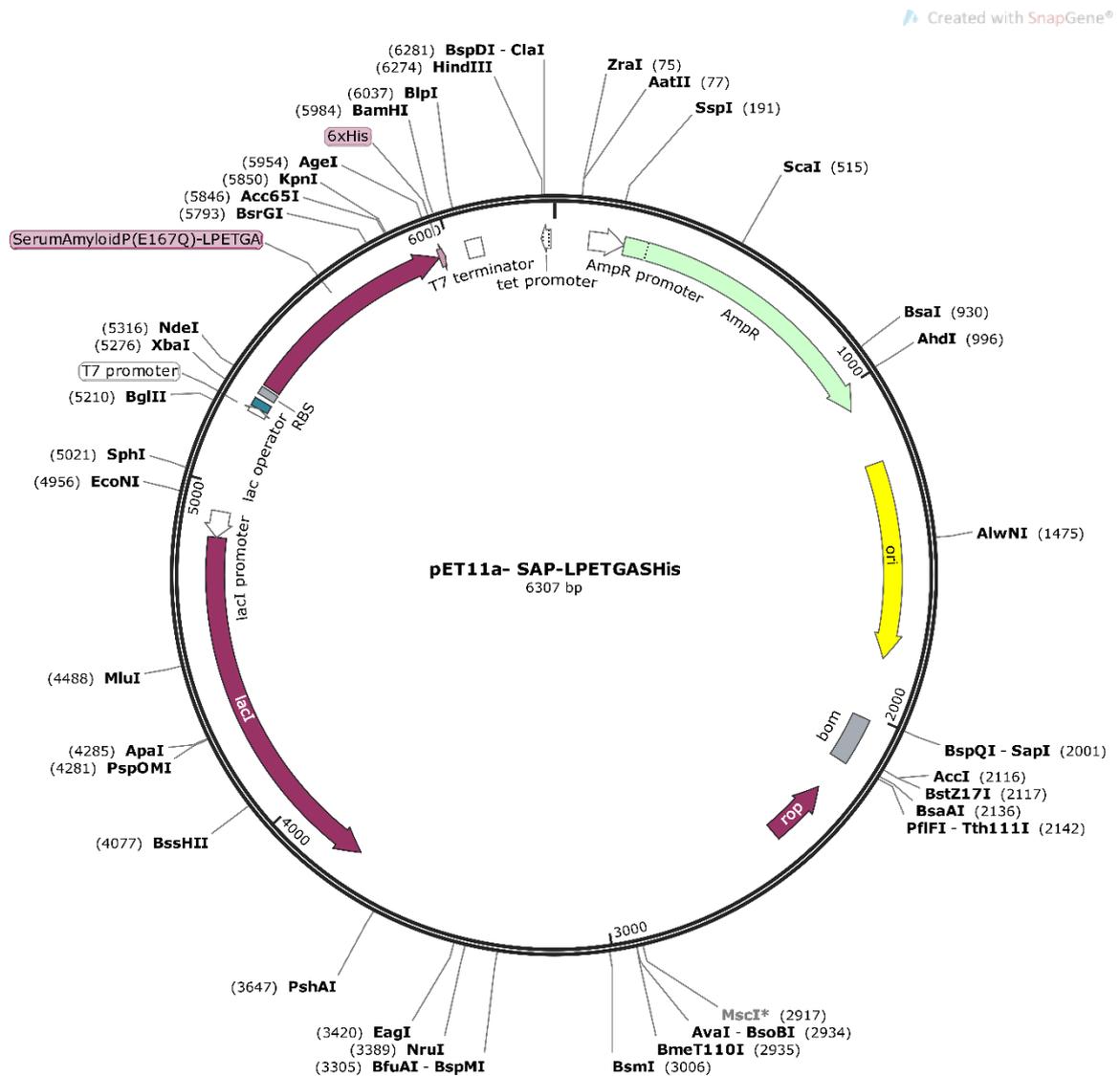


Figure 10.6: Plasmid map for SAP-LPETGASHis in the pET11a expression vector

The gene for SAP (Uniprot – P02743) was engineered to prevent calcium induced aggregation using introducing E167Q mutation based on reports by Pepys *et al.*³⁰⁴, a C-terminal extension was introduced to include sortase recognition sequence with His-tag, and the N-terminus was redesigned to contain an terminal serine. pET-11a expression vector harbouring the gene for mutant SAP was purchased from GenScript.

10.6.1 Vector Sequence

TTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAA
 TAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATT
 TGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATA

AATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTGCCCC
TTATTCCCTTTTTTTCGCGCATTTTTGCCTTCCTGTTTTTGGCTCACCCAGAAACGCTGGTG
AAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCT
CAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCA
CTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCCTGTTGACGCCGGGCAAGAGCAA
CTCGGTGCGCCGATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGA
AAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGA
GTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACC
GCTTTTTTGCACAACATGGGGGATCATGTAACCTCGCCTTGATCGTTGGGAACCGGAGCT
GAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAA
CGTTGCGCAAACCTATTAACCTGGCGAACTACTTACTCTAGCTTCCC GGCAACAATTAATA
GACTGGATGGAGGCGGATAAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGG
CTGGTTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAG
CACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAG
GCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCA
TTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACCTTCATT
TTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCT
TAACGTGAGTTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTC
TTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAACACCGCTAC
CAGCGGTGGTTTTGTTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACCTGGC
TTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCA
CTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGG
CTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCG
GATAAGGCGCAGCGGTGCGGCTGAACGGGGGGTTCGTGCACACAGCCAGCTTGGAGCG
AACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTC
CCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGC
ACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCTGTGCGGTTTTCGCCA
CCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAA
ACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTTGCTGGCCTTTTGCTCACATG
TTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGC
TGATAACCGCTCGCCGACCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGG
AAGAGCGCCTGATGCGGTATTTCTCCTTACGCATCTGTGCGGTATTTACACCGCATA
TATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACTC
CGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGAC
GCGCCCTGACGGGCTTGTCTGCTCCC GG CATCCGCTTACAGACAAGCTGTGACCGTCTC

CGGGAGCTGCATGTGTCAGAGGTTTTACCGTCATCACCGAAACGCGCGAGGCAGCTGC
GGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTACAGATGTCTGCCTGTTTCATCCGCG
TCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCAT
GTTAAGGGCGGTTTTTTTCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTCTGT
TCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGAT
GATGAACATGCCCCGTTACTGGAACGTTGTGAGGGTAAACAACCTGGCGGTATGGATGCG
GCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAG
GTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAG
GGCGTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTTCATGT
TGTTGCTCAGGTCGCAGACGTTTTTGCAGCAGCAGTCGCTTCACGTTTCGCTCGCGTATCG
GTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGAC
AGGAGCACGATCATGCGCACCCGTGGCCAGGACCCAACGCTGCCCGAGATGCGCCGCGT
GCGGCTGCTGGAGATGGCGGACGCGATGGATATGTTCTGCCAAGGGTTGGTTTGCGCAT
TCACAGTTCTCCGCAAGAATTGATTGGCTCCAATTCTTGGAGTGGTGAATCCGTTAGCG
AGGTGCCGCCGGCTTCCATTCAGGTCGAGGTGGCCCGGCTCCATGCACCCGCGACGCAAC
GCGGGGAGGCAGACAAGGTATAGGGCGGCGCTACAATCCATGCCAACCCGTTCCATGT
GCTCGCCGAGGCGGCATAAATCGCCGTGACGATCAGCGGTCCAGTGATCGAAGTTAGGC
TGGTAAAGAGCCGCGAGCGATCCTTGAAGCTGTCCCTGATGGTCGTCATCTACCTGCCTG
GACAGCATGGCCTGCAACGCGGGCATCCCGATGCCGCCGGAAGCGAGAAGAATCATAAT
GGGGAAGGCCATCCAGCCTCGCGTCGCGAACGCCAGCAAGACGTAGCCCAGCGCGTCCG
CCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTG
ACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCAT
CGTCGCGCTCCAGCGAAAGCGGTCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCT
GTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCC
CGCGCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTGAGATCC
CGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC
AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGGGGAGAGGC
GGTTTGCATATTGGGCGCCAGGGTGGTTTTTCTTTTACCAGTGAGACGGGCAACAGCT
GATTGCCCTTACC GCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGC
CCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTC
TTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGG
TAATGGCGCGCATTGCGCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGA
ACGATGCCCTCATTTCAGCATTTCATGGTTTGTGAAAACCGGACATGGCACTCCAGTC
GCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAG
CCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGG

TGACCCAATGCGACCAGATGCTCCACGCCAGTCGCGTACCGTCTTCATGGGAGAAAAT
AATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGC
AGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCA
CTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCG
TTCTACCATCGACACCACCACGCTGGCACCAGTTGATCGGCGCGAGATTTAATCGCCG
CGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAAC
GACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCAT
CGCCGCTTCCACTTTTTCCCGCGTTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGC
GGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGT
TTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAA
GGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGC
ATTAGGAAGCAGCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGG
TGCATGCAAGGAGATGGCGCCCAACAGTCCCCGGCCACGGGGCCTGCCACCATAACCCA
CGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCATCGGTGATG
TCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCG
TCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGA
ATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGA
TATACA **TATGAGCCACACCGACCTGAGCGGTAAAGTGTTTCGTTTTTCCGCGTGAGAGCG**
TGACCGATCACGTTAACCTGATCACCCCGCTGGAAAAGCCGCTGCAGAACTTCACCCTG
TGCTTTCGTGCGTACAGCGACCTGAGCCGTGCGTACAGCCTGTTTCAGCTATAACACCCA
AGGCCGTGATAACGAGCTGCTGGTGTATAAAGAGCGTGTTGGTGAATACAGCCTGTATA
TTGGCCGTCACAAGGTGACCAGCAAAGTTATCGAAAAGTTTCCGGCGCCGGTGACATT
TGCGTTAGCTGGGAGAGCAGCAGCGGTATCGCGGAATTCTGGATTAACGGCACCCCGCT
GGTGAAGAAAGGTCTGCGTCAGGGCTACTTTGTGGAGGCGCAACCGAAAATCGTTCTGG
GTCAGGAACAAGACAGCTATGGTGGCAAGTTCGATCGTAGCCAGAGCTTTGTGGGCGAG
ATCGGCGACCTGTACATGTGGGATAGCGTTCTGCCGCCGAGAACATTCTGAGCGCGTA
TCAAGGTACCCCGCTGCCGGCGAACATCCTGGACTGGCAAGCGCTGAACTACGAGATTC
GTGGTTATGTTATCATTAACCGCTGGTGTGGGTTGGTAGCGGTCTGCCGAAACCGGT
GCGCACACCACCACCACCTAAGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTG
AGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGG
GTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATATCCCGCAAGAGGCC
GGCAGTACCGGCATAACCAAGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGAT
GACGATGAGCGCATTGTTAGATTTTCATACACGGTGCCTGACTGCGTTAGCAATTTAACT
GTGATAAACTACCGCATTAAAGCTTATCGATGATAAGCTGTCAAACATGAGAA

10.6.2 SAP-LPETG Protein Sequence

```
MSHTDLSGKVFVFPRESVTDHVNLITPLEKPLQNFTLCFRAYSDDLRSAYSLFSYNTQGR  
DNELLVYKERVGEYSLYIGRHKVTSKVIEKFPAPVHICVSWESSSGIAEFWINGTPLVK  
KGLRQGYFVEAQPKIVLGQEQDSYGGKFDRSQSFVGEIGDLYMWDSVLPQNILSAYQG  
TPLPANILDWQALNVEIRGYVVIKPLVWVGSGLPETGAHHHHHH
```

10.7 Supplementary Figures

10.7.1 Results for the inhibition assay of lactose and 3.3(W88E) on CTB-HRP

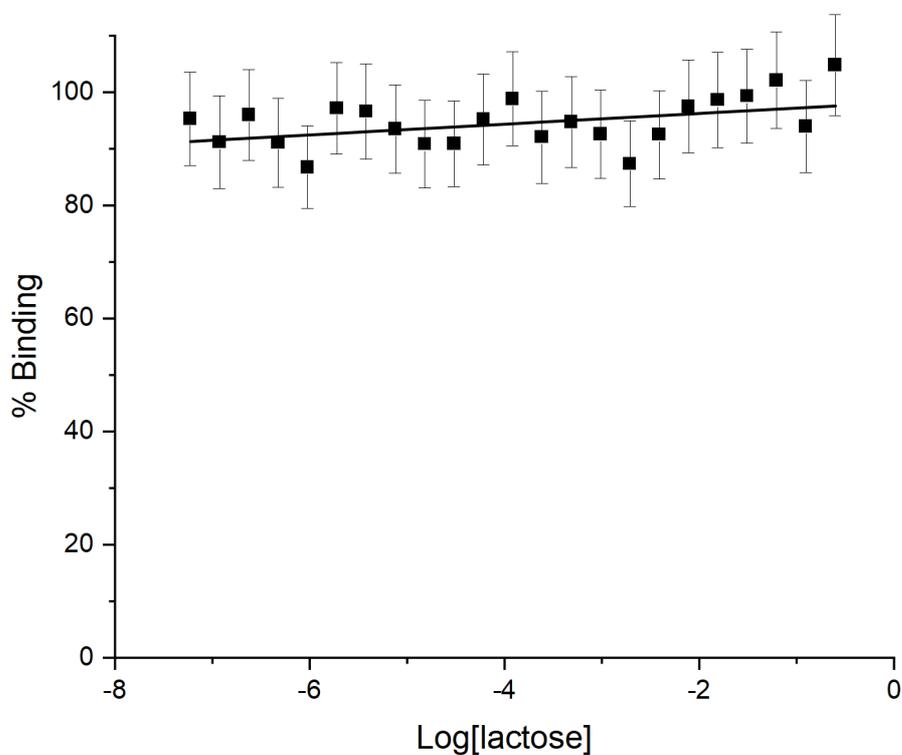


Figure 10.7: Results for the ELLA inhibition assay of monovalent lactose against CTB-HRP adhesion to GM1 coated microtiter plates.

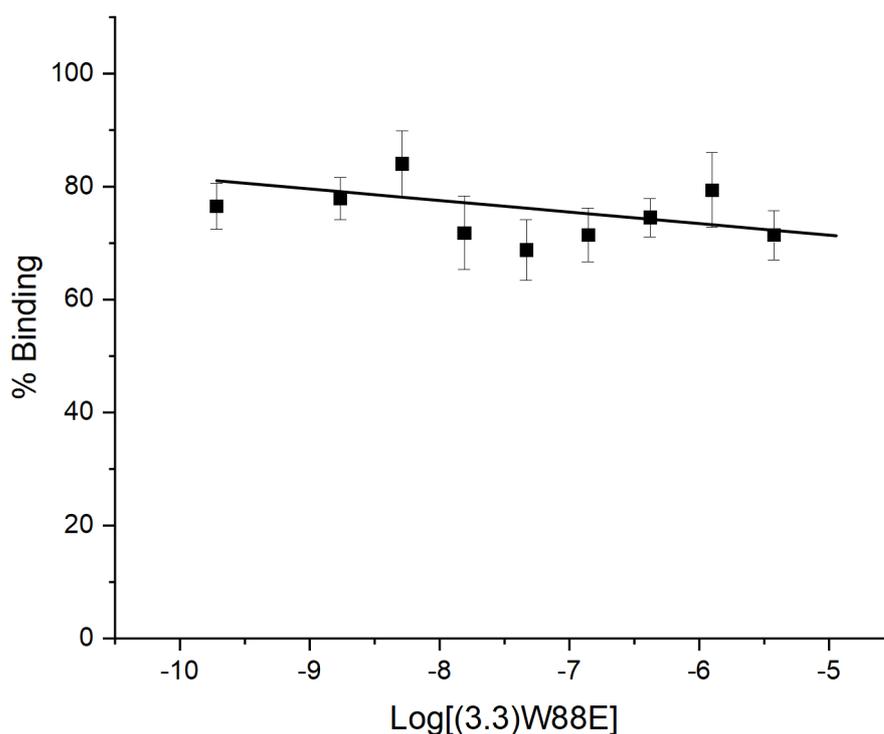


Figure 10.8: Results for the ELLA inhibition assay of pentavalent neoglycoprotein (3.3)W88E against CTB-HRP adhesion to GM1 coated microtiter plates.

10.7.2 Raw graphical data for VTB-HRP capture on Gb3 microtiter plates at three concentrations of synthetic glycolipid Gb3-DOPE

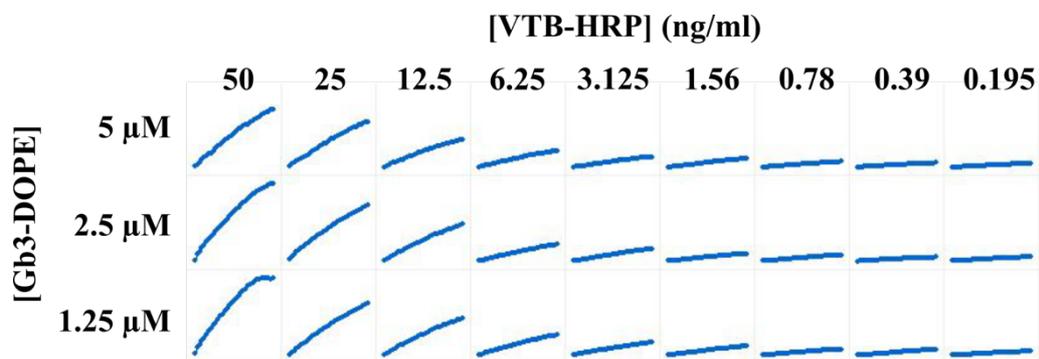


Figure 10.9: Raw graphical data showing capture of VTB-HRP on Gb3 coated microtiter plates. The data shows serial 2-fold dilution of VTB-HRP at three concentrations of Gb3-DOPE to determine the optimal concentration of VTB-HRP that gave good signal-to-noise.