Investigating the *N*-linked glycan specificity of bacterial endo-β-*N*-acetylglucosaminidase S



By

Jonathan Goodfellow

Oriel College, Oxford

SUBMITTED FOR THE DEGREE OF MASTER OF SCIENCE BY RESEARCH

AT THE

UNIVERSITY OF OXFORD

TRINITY 2011

Supervised By

Professor Ben G. Davis

For Dad.....

2nd October 1956-1st September 2010



“Goodfellow by name, good fellow by nature”

"The time has come, "the Walrus said,

"To talk of many things".

-The Walrus and The Carpenter, Lewis Carroll, (from *Through the Looking-Glass and What Alice Found There, 1872)*




Endo
glycans
mL
protein
N-linked
glycan
reaction
Fc
IgG
incubated
using
mixture
hours
cells
mmol
mass
analysis
glycosylation
solution
aqueous
buffer
stirred
water
column
purified
digestion
expressed
glycoproteins
gp120BAL
transglycosylation
SDS-PAGE
released
benzyl
reactions
yield
minutes
DNA
core
Glycans
pH
Lane
NEB
GlcNAc
CDCl3
Pet
mM
µL
cHarom
type
sample
glycoprotein
synthesis
used
overnight
Ether/EtOAc
kDa
MALDI-MS
Buffer
minute
organic
starting
filtered
RT
Na
concentrated
enzymatic
CH
3H
2H
LRMS
centrifuged
HEK293T
NMR
MHz
CH3
Mw
tube
hour
human
peak
activity
complex
endoglycosidase
sodium
dd
m/z
specificity
washed
purification
gel
anhydrous
PNGase
resultant
shown
coli
acid
glycoproteins
PBS
SWN
app
Chem
H-3
H-5
I3C
C-2
C-1
J6
J5
J1
H-1
H-2
H-4
H-6
WT
one
solid
Biochemistry
containing
cleavage
group
enzyme
Table
Formula
argon
Milli-Q
Protein
material
mol-I
wild
Phth
expression
vacuo
showed
spectra
found
system
benzylidene
Rf
resulting
combined
complete
3:2
SCH2CH3

Table of Contents

Abstract iv

Acknowledgements v

Abbreviations vi

Amino Acid Information xi

Explanatory Note On Carbohydrate Numbering xii

Chapter 1

Introduction

1.1. Glycoproteins....1

1.2. Recombinant expression systems....2

1.3. The Chemoenzymatic Approach....3

1.4. Endo-β-*N*-acetylglucosaminidase enzymes....4

1.5. Streptococcus pyogenes....9

1.6. Endo S and IgG activity....10

1.7. Aims of the project....13

1.8. References....14

Chapter 2

Materials and Methods

2.1. General Considerations for Biology....20

2.2. Protein Expression....28

2.3. Glycan Cleavage reactions....31

2.4. MALDI Mass Spectrometric Analysis....32

2.5. References....35

Chapter 3

N-linked glycan specificity of bacterical endo-β-*N*-acetylglucosaminidase S

3.1. Expression and purification of wild type Endo S....36

3.2. Probing the activity of Endo S....43

3.3. Quantification of the *N*-linked glycan released on IgG by Endo S....45

3.4. Glycan cleavage enzymatic reactions on IgG using other Endo enzymes....48

3.5. Quantitative analysis of Endo A and Endo H glycan release from IgG....50

3.6. Glycan specificity of WT Endo S....52

3.7. Expression and purification of WT Fc and SWN Fc glycoproteins....59

3.8. Conclusion....67

3.9. References 71

Chapter 4

Studies towards a transglycosylation model

4.1. Transglycosylation model 75

4.2. Quantitative analysis of Endo A on $gp120_{BAL}$ glycans 76

4.3. Towards the synthesis of a glycosyl donor for Endo A glycan extension 80

4.4. General Considerations for Chemistry 83

4.5. Chemical Compounds 85

4.6. ^{1}H NMR Spectra 103

4.7. References 109

Chapter 5

Appendix

5.1. *N*-linked Glycan types with susceptibility to Endo H 112

5.2. *N*-linked Glycan types with susceptibility to Endo S 114

Abstract

Endoglycosidase (Endo) S, secreted by bacterial pathogen *Streptococcus pyogenes*, is an important immunomodulatory enzyme which specifically hydrolyses the functionally important *N*-linked complex biantennary glycan found at Asn297 on the C_H2 domain of fragment crystallisable (Fc) on human immunoglobulin G (IgG). Deglycosylation of IgG Fc leads to complete loss of binding to cellular Fcγ receptors (FcγRs) thus abolishing antibody dependent cellular cytoxicity (ADCC) and prolonging bacterial survival. Here, the *N*-linked glycan specificity of Endo S was further investigated using unnatural *N*-linked glycans. Recombinant Endo S was expressed and purified and its specificity for different *N*-linked oligo-, high-mannose, and complex type glycans was studied using MALDI-TOF analysis. Furthermore, Endo S specificity for core fucosylated *N*-linked complex glycans on human IgG was also investigated. These studies revealed that *N*-linked oligo- and high-mannose glycans are resistant to Endo S hydrolysis and that the presence of α-1,6 linked fucose in the chitobiose core is not essential for Endo S activity. These studies further revealed that the *N*-linked glycan specificity of Endo S is reciprocal to that of other known endoglycosidases namely Endo A and Endo H. Taken together, the resistance of these engineered *N*-linked oligo- and high-mannose glycans on IgG to Endo S digestion opens up the possibility of using these antibodies as a potential therapeutic for autoimmune diseases.

Acknowledgements

Firstly, I'd like to thank Ben for giving me the amazing opportunity to work in the BGD group and for his constant motivation and enthusiasm on a fantastic project.

Secondly, I'd really like to thank everyone in the BGD group for their many sources of help, advice, tea, chats and of course, the odd pint or two at the infamous PT. To all the fellow lab members of LG10 for their 'fine' music choices - what WILL we do when Glee finishes? There are far too many people to individually thank for their invaluable friendship during my time in the group, you know who you are! Chris and Pieter, you guys really made my time in the group fantastic and I'll miss you immensely.

To my adopted group: Camille, Dan, Kavitha, Chris and Max. I'll always be grateful for you guys putting me up with me. Many thanks to Ben Davis, Chris Scanlan, Smita and Kavitha for their proof reading and my constant questions and drafts - much appreciated!

To the amazing members of 23 Reagent Street: Cat, Charley, Laura and Matt, I'll miss you immensely! Who will keep me entertained outside the lab?! Never forget your mews. To Scott, Natalie, Josh, Simi and Dan – you really are all JBs.

Lastly, to my loving family and Greg who are always there to support me, listen to my rants, feed me pizza and who constantly help me to achieve my dreams - it means the world to me. Dad, hope you're proud.

Love,



Jonathan Goodfellow

Michaelmas 2009 - Trinity 2011

Abbreviations

-ve	negative
+ve	positive
Ab	antibody
Ac	acetyl
Anh.	anhydrous
aq.	aqueous
Asn	asparagine
Bn	benzyl
BDMA	benzaldehyde dimethylacetal
BSA	bovine serum albumin
^{o}C	degrees Celsius
calc.	calculated
cat.	catalytic
$CDCl_3$	deuterated chloroform
CD_3OD	deuterated methanol
conc.	concentrated
COSY	proton-proton correlation spectroscopy (NMR)
δ	chemical shift (ppm)
D_2O	deuterated water
Da	Dalton
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
DCM	dichloromethane
dm^{-3}	per decimetre (litre)

DNA	deoxyribose nucleic acid
Endo A	endo-β-*N*-acetylglucosaminidase from *Arthrobacter protophormaie*
Endo H	endo-β-*N*-acetylglucosaminidase from *Streptomyces plicatus*
Endo M	endo-β-*N*-acetylglucosaminidase from *Mucor hiemalis*
Endo S	endo-β-*N*-acetylglucosaminidase from *Streptococcus pyogenes*
Et	ethyl
Et_2O	diethyl ether
et al.	*et alia* (Latin: and others)
EtOAc	ethyl acetate
eq.	equivalent
ES	electrospray
Fab	fragment (antibody binding) of Igs
Fc	fragment (crystallisable) of Igs
FC	flash chromatography
g	grams
G	guanine
GlcNAc	*N*-acetylglucosamine
GST	*Glutathione S-Transferase*
h	hour(s)
HMQC	heteronuclear multiple-quantum correlation
Hz	hertz
Igs	immunoglobulins
IgG	immunoglobulin G
IPTG	isopropyl-1-thio-β-D-galactopyranoside

J	NMR coupling constant (Hz)
kb	kilo base pairs
LB	Luria-Bertani
LC-MS	liquid chromatography-mass spectrometry
LRMS	low resolution mass spectrometry
m	milli
M	molar (mol dm^{-3})
M	molecular peak (mass spectrum)
µ	micro
MALDI	matrix-assisted laser desorption ionization
MeOH	methanol
MHz	megahertz
Man	mannose
min	minute
$MgSO_4$	magnesium sulfate
mL	millilitre
mol	mole
m.p.	melting point
MS	mass spectrometry
m/z	mass to charge ratio
N.B.	*nota bene* (Latin: note well)
NEB	New England BioLabs®
neg.	negative
nm	nanometre
N_2	nitrogen

NMR	nuclear magnetic resonance	
NMR data:	app.	apparent
	br.	broad
	s	singlet
	d	doublet
	dd	double doublet
	ddd	double double doublet
	t	triplet
	q	quartet
	m	multiplet
OD	optical density	
ox	oxazoline	
PBS	phosphate buffered saline	
PCR	polymerase chain reaction	
pH	-log [H^+ ions]	
Ph	phenyl	
Phth	phthalimido	
pos.	positive	
ppm	parts per million	
PTM	posttranslational modification	
py	pyridine	
quant.	quantitative	
R	unspecified group	
R_f	retention factor	
RT	room temperature	

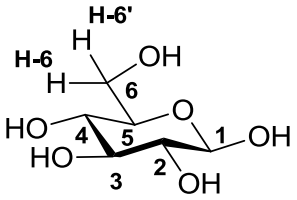
sat.	saturated
SBL	subtilisin *Bacillus lentus*
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
soln.	solution
T	thymine
TBS	*tert*-butyldimethylsilyl
tert-	tertiary
TEA	triethylamine
Tf	triflate
TFA	trifluoroacetic acid
TLC	thin layer chromatography
TMS	tetramethylsilane

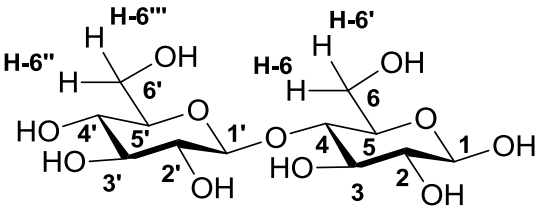
Amino Acid Information

Amino Acid	Three-Letter Abbreviation	One-Letter Symbol	Molecular Weight (Da)
Alanine	Ala	A	89
Arginine	Arg	R	174
Asparagine	Asn	N	132
Aspartic Acid	Asp	D	133
Asparagine or aspartic acid	Asx	B	133
Cysteine	Cys	C	121
Glutamine	Gln	Q	146
Glutamic acid	Glu	E	147
Glutamine or glutamic acid	Glx	Z	147
Glycine	Gly	G	75
Histidine	His	H	155
Isoleucine	Ile	I	131
Leucine	Leu	L	131
Lysine	Lys	K	146
Methionine	Met	M	149
Phenylalanine	Phe	F	165
Proline	Pro	P	115
Serine	Ser	S	105
Threonine	Thr	T	119
Tryptophan	Trp	W	204
Tyrosine	Tyr	Y	181
Valine	Val	V	117

Explanatory Note On Carbohydrate Numbering Convention

Carbohydrates are numbered according to the standard carbohydrate convention:





Chapter 1

Introduction

1.1. Glycoproteins

Glycoproteins are an essential class of biomolecules found within a living system. They are proteins that contain glycans covalently attached to the surface of the protein through a polypeptide side-chain and naturally can be found to be either *C*-linked, *O*-linked or *N*-linked. Another type of covalent attachment is the attachment of the glycolipid, glycosylphosphatidylinositol (GPI anchor) to the C-terminus of a protein during posttranslational modification.[1] The *N*-linked oligosaccharide moieties of these glycoproteins are added to the proteins through protein glycosylation, a distinctive and common form of co-translational modification that occurs in the endoplasmic reticulum and Golgi apparatus of eukaryotes. Further examples of other modifications are shown below (Figure 1.1).[2]

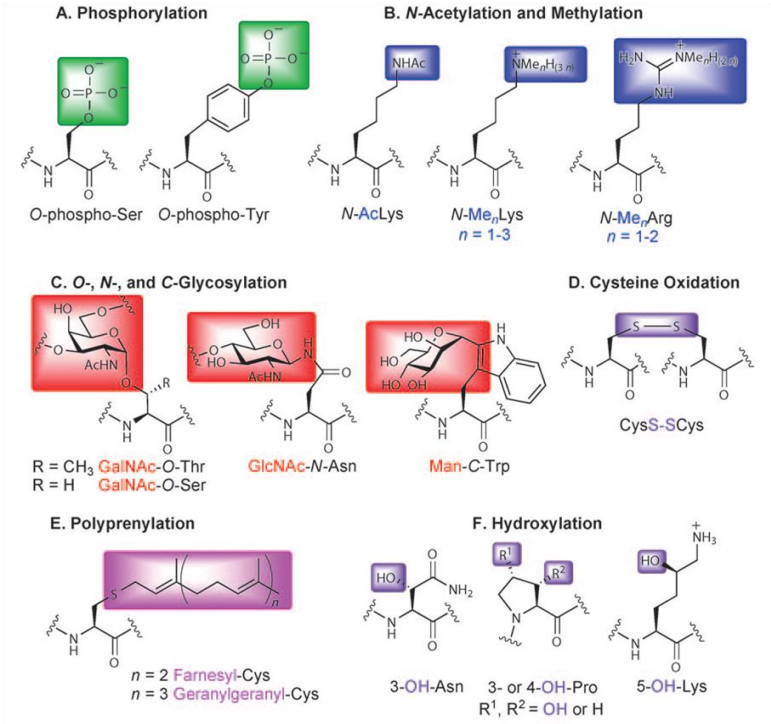


Figure 1.1: Common examples of posttranslational modifications (PTMs) that occur in eukaryotes. Reproduced from Chalker *et. al.*[2]

Glycans are fundamental for correct protein folding, cellular recognition[3] and play important roles in other biological processes[4] such as viral infection, lectin binding,[5] tumor metastasis[4] and the immune response.[6-8] Currently, 70% of the total therapeutic proteins that are in clinical or preclinical trials are glycoprotein based and the global market for biopharmaceuticals, including protein based products is estimated to be more than $30 billion.[9] Natural glycoproteins usually exist as mixtures of glycoforms (glycoproteins containing the same peptide backbone but possess glycans that have different glycosylation patterns and at different sites). This microheterogeneity can make the isolation of well-defined, pure glycoproteins problematic.[10] Since it has been found that certain glycoforms are more active than others, the access to specific uniform glycoforms can confer significant therapeutic advantages. Therefore, there is a need to resolve the supply issues of homogeneous glycoprotein production for the development of effective glycoprotein-based therapeutics.[11] Recombinant expression systems and chemoenzymatic methods have emerged as techniques that can resolve this supply issue.

1.2. Recombinant expression systems

The use of purely biological methods for the synthesis of glycoproteins using recombinant mammalian, yeast, plant and fungal expression systems is a viable method for producing glycoproteins.[12] The yeast *Pichia pastoris* is used extensively as the host cell for large-scale production of secreted recombinant proteins.[13, 14] This is achieved by the genetic modification (or engineering) of the *N*-glycosylation pathway in *P. pastoris,* a eukaryotic cell capable of producing soluble, correctly folded recombinant proteins, to yield *N*-linked glycans that are the same as the intermediates of mammalian-protein *N*-glycosylation.[15] This ‘humanised’ approach has been successful in the production of

glycosylated antibodies.[16] The advantages of these recombinant expression approaches are that the desired protein can be expressed in high titers and the relative inexpensive of the growth media needed.[13]

1.3. The Chemoenzymatic Approach

One of the most recent and exciting developing methods of protein modification is to use a chemoenzymatic approach[17], which combines the enzymatic elaboration of sugar chains with chemical polypeptide synthesis to remodel mixtures of glycoforms.[18] The heterogeneous mixture of glycoforms is treated with an endoglycosidase enzyme (Endo) to yield a single protein glycoform with an innermost *N*-acetyl glucosamine (GlcNAc) residue at an *N*-linked glycosylation site. Endo mediated glycosylation of the single GlcNAc residues can then go on to produce the desired homogeneous sample of the desired glycoprotein[17] (Figure 1.2, Route A).

Alternatively, another chemoenzymatic approach that has been developed consists of a site-selective convergent chemical glycosylation of a protein scaffold with a single GlcNAc residue followed by Endo mediated glycosylation which could also give access to homogeneous proteins.[11] This process requires the preparation of only monosaccharide-tagged GlcNAc polypeptides from solid-phase peptide synthesis (SPPS) and the glycan chain is then extended by the attachment of an intact oligosaccharide in one step to give the desired target glycoprotein. This enzymatic reaction is carried out in aqueous solution with free polypeptides, without the need for protecting groups.[19] This approach is known as the 'Tag-and-Modify' approach (Figure 1.2, Route B).

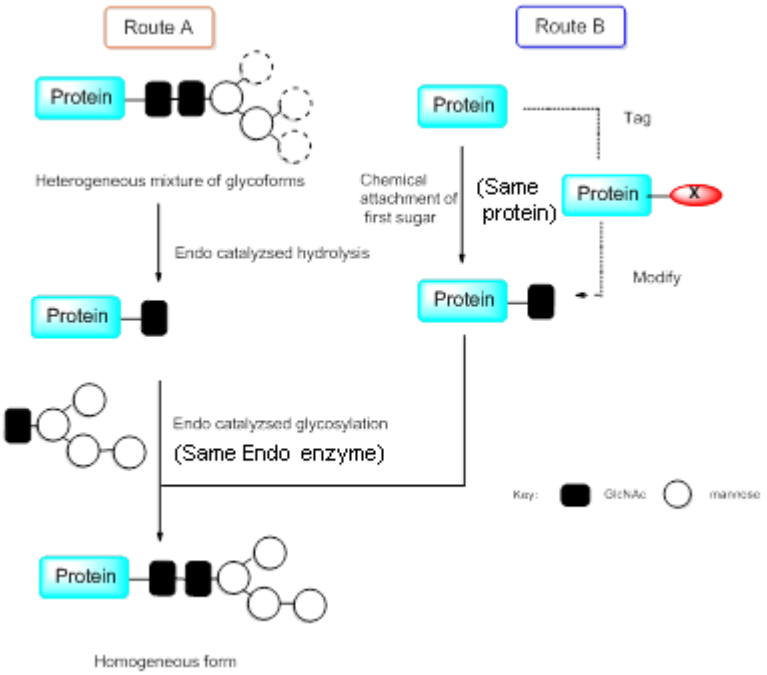


Figure 1.2: Simple schematic of current chemoenzymatic strategies for endoglycosidase catalyzed protein glycosylation. (A) Endoglycosidase 'trimming' of heterogeneous mixture. (B) 'Tag-and-modify' approach.

An advantage of this approach is the ability to transfer synthetic glycans onto the protein. The main disadvantage of this approach is that site of glycosylation (i.e. the site of GlcNAc-Asn residues) is already well established and cannot be easily manipulated. Another disadvantage is that the enzymatic reaction may not go to completion in either the trimming of the intrinsic glycans or the extension using synthetic glycans giving rise to more microheterogeneity in the sample and another level of uncertainty.

1.4. Endo-β-*N*-acetylglucosaminidase enzymes

Endo-β-*N*-acetylglucosaminidases (Endo) [ENGases, E.C. 3.2.1.96] are a class of enzymes that hydrolyse the GlcNAc-β(1→4)-GlcNAc glycosidic bond (Figure 1.3) in the *N,N'*-diacetylchitobiose core present in *N*-linked (asparagines, Asn) glycans of numerous glycoproteins and glycopeptides.[20]

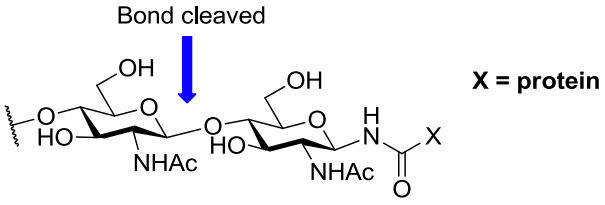


Figure 1.3: Site of bond cleavage due to endo-β-N-acetylglucosaminidases.

Besides hydrolytic activity, some ENGases of the GH85 family processes significant transglycosylation activity (the ability to transfer the released oligosaccharide moiety to a suitable acceptor other than water).[21] It has been found that endo-β-*N*-acetylglucosaminidases from bacterial *Arthrobacter protophormiae*, denoted Endo A[22, 23] and fungal *Mucor hiemalis*, which is denoted Endo M[24, 25] have shown strong *in vitro* transglycosylation activity unlike other endo-β-*N*-acetylglucosaminidases in their family.

Endo A has been shown to effectively transfer high-mannose type oligosaccharides to various glycosides and glycoproteins whereas Endo M is unique in that it exhibits transglycosylation and hydrolysis activity on all three types of *N*-glycans*:* complex type oligosaccharides, high-mannose type and hybrid type (Figure 1.4).[26] This transglycosylation activity is useful for the glycosylation remodelling of glycoproteins.[11]

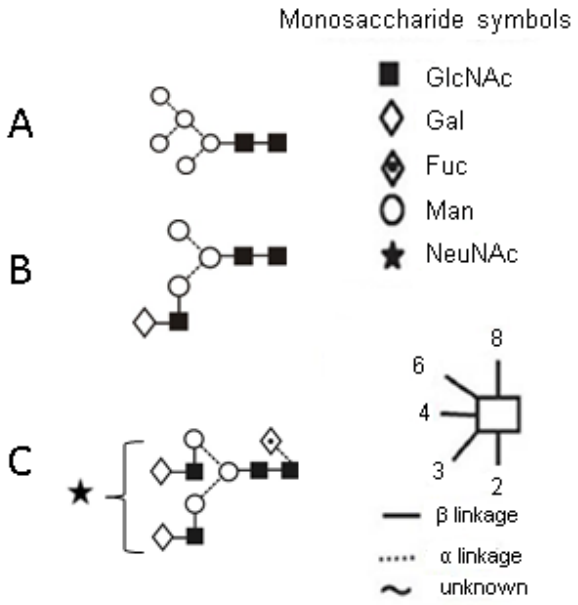


Figure 1.4: Different types of *N*-linked oligosaccharides. (A) High-mannose type glycan. (B) Hybrid type glycan. (C) Complex type glycan.

The effectiveness of these enzymes for the manipulation of glycopeptides and glycoproteins is due to the unusual 'neighbouring group' participation reaction that is harnessed by this enzyme class.[27] This participation involves a glycosidic bond cleavage that proceeds with overall net retention of the anomeric configuration *via* a cyclic oxazoline/oxazolinium ion intermediate. This reaction is aided by catalytic acid and base residues situated in the active site of the enzyme (Figure 1.5).

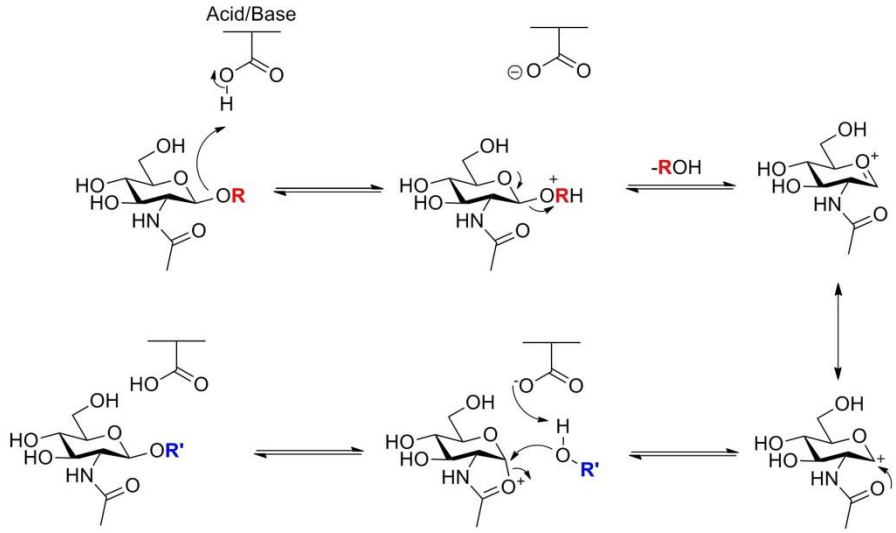


Figure 1.5: 'Neighbouring group' participation reaction utilised in transglycosylation reactions resulting in an overall net retention in stereochemistry at the anomeric centre.

However, this chemoenzymatic transglycosylation using Endo enzymes has some disadvantages. Firstly, ENGases are intrinsically a class of glycosyl hydrolases which means that product hydrolysis is a major problem from this approach as when the desired product formed accumulates it is the ideal substrate of the enzyme. This attributes to the low yields (usually ranging from 5-20%) that have been reported due to the fast competitive hydrolysis of the product. Lastly, this chemoenzymatic approach is limited in its usefulness as it is restricted to only natural *N*-linked glycans.[19]

To overcome some of these issues there have been several possible solutions proposed. Firstly, organic co-solvents have been incorporated into the reaction which can enhance the transglycosylation activity and therefore decrease the undesired product hydrolysis.[28] Secondly, it has been postulated that the ENGase-catalyzed hydrolysis retaining mechanism proceeds via an oxazolinium ion intermediate[26, 29] so the use of synthetic oligosaccharide intermediates as glycosyl donors can increase yields of synthesis and broaden substrate availability. Lastly, the active site of Endo A has been

recently shown to contain two essential tryptophan ‘gates’ which open and close to allow substrate binding and help with the catalytic turnover of the enzyme (W215, W244) (Figure 1.6). Carrying out mutagenesis on these gates can also enhance transglycosylation activities and broaden tolerance towards other linkages too.[30, 31]

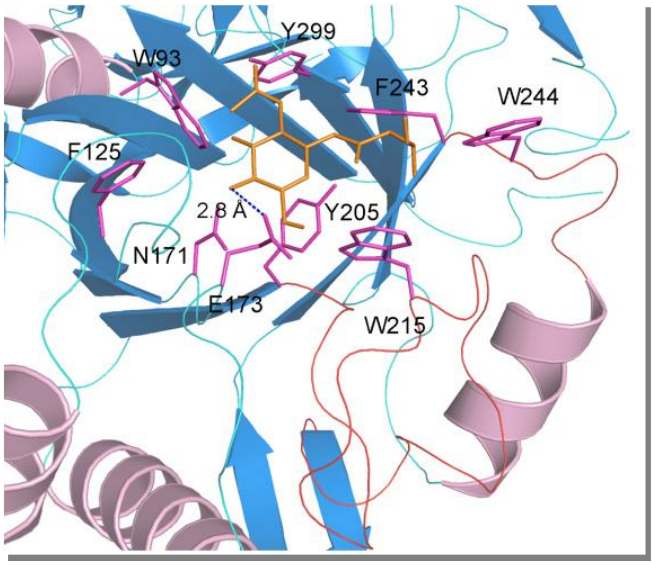


Figure 1.6: Crystal structure of Endo A active site, showing catalytic residues and important tryptophan ‘gates’. PDB3FHQ.[26]

1.5. Streptococcus pyogenes

S. pyogenes is a spherical, Gram-positive, β-haemolytic bacterium that is responsible for streptococcal illness[32] (Figure 1.7).

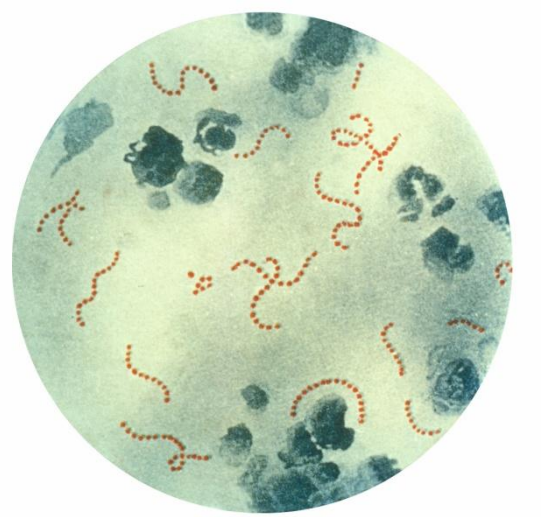


Figure 1.7: Pappenheim's stain of *Streptococcus pyogenes* bacteria at 900× magnification. Image from Centers for Disease Control and Prevention's Public Health Image Library (PHIL), with identification number #2110.[33]

S. pyogenes is a significant human pathogen that selectively interacts with antibody proteins involved in the humoral defense system, such as immunoglobulins (Igs).[34] *S. pyogenes* causes many human diseases, primarily infecting the upper respiratory tract and skin resulting in mild superficial skin disorders such as impetigo or when in the throat, strep throat.[35, 36] However, occasionally it disseminates and can be known to cause severe invasive diseases such as necrotising fasciitis and sepsis, diseases that have a high mortality rate.[37]

This pathogen was found to secrete a novel 108 kDa protein, Endo S which has shown endoglycosidase activity on human immunoglobulin G (IgG), contributing to the complex pathogenicity of this microorganism.[34] Interestingly, neither human IgA nor IgM are affected by Endo S.[34]

1.6. Endo S and IgG activity

Endo S is an important immunomodulatory enzyme which targets the adaptive branch of the immune system, hydrolyzing the functionally important defined *N*-linked complex biantennary oligosaccharide containing a bisecting *N*-acetylglucosamine (GlcNAc) and a core fucose at Asn297 on the γ-chains contained in this opsonising IgG (Figure 1.8). Endo S cleaves the *N*-linked glycans attached to all four subclasses of IgG (IgG1, IgG2, IgG3, and IgG4).[38] This protein seems to be expressed as a putative immune evasion factor.

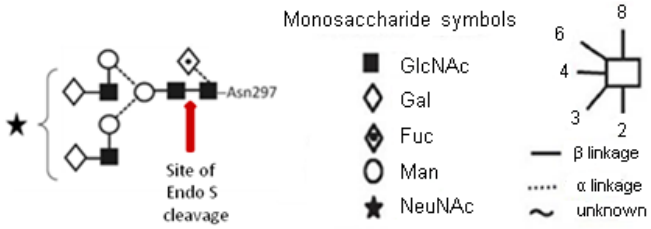


Figure 1.8: Site of Endo S hydrolysis on the complex *N*-linked glycan found natively on IgG

IgG is a 150 kDa glycoprotein that is comprised of two identical light chains and two identical heavy chains (Figure 1.9).[39]

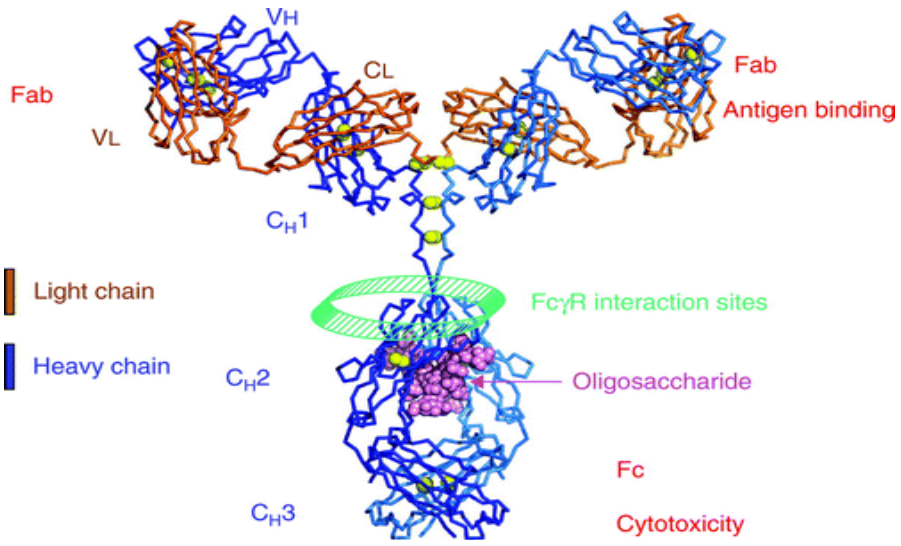


Figure 1.9: Structural representation of Human IgG. Reproduced from Jefferis.[39]

The upper fragments of IgG are known as the fragment antigen-binding (Fab) regions which are identical in structure. The lower fragment is known as fragment crystallisable (Fc) region (Figure 1.9) and forms the interaction sites for ligands that activate clearance and transport mechanisms.[39] The IgG Fc region is a homodimer comprised of covalent inter-heavy chain disulfide-bonded hinge regions and non-covalently paired C_H3 domains. The Fc region of IgG mediates the effector functions of antibodies by binding to cellular Fcγ receptors (FcγRs). The C_H2 domains are glycosylated through the covalent attachment of an *N*-linked oligosaccharide at Asn297. It is this oligosaccharide that Endo S selectively targets.

Human IgGs are known to vary in the *N*-linked glycosylation at the conserved *N*-glycosylation site of the Fc part, this means that IgG found in human serum comprises of a number of distinct glycoforms. In general, IgG Fc *N*-linked glycans are mostly biantennary complex-type structures, most of them being core-fucosylated, Some of the glycans carry a bisecting *N*-acetylglucosamine, varying addition of galactose, and a small

portion may carry a sialic acid residue (less than 10% of the total IgG Fc molecules) on the antennae.[39,40] The *N*-linked glycan moiety found at the Asn297 residue terminates with various different carbohydrate structures, and these differing forms are important in the development of disease. Fc glycans comprise of various biantennary, core fucosylated structures which potentially leads to the generation of at least 128 unique glycoforms being observed in serum.[39] The *N*-linked glycan present on Fc glycoproteins contributes to their solubility and stability.[41] Usually, less than 10% of the total normal human IgG have an IgG Fc region containing sialic acid residues[42, 43]. Differential sialylation affects whether IgG has pro- or anti-inflammatory properties. [44] Defucosylation has been shown to increase antibody dependent cellular cytotoxicity (ADCC).[45] A summary of the roles of the various carbohydrate moieties of the *N*-linked glycan found at Asn297 is summarised in Figure 1.10.[46]

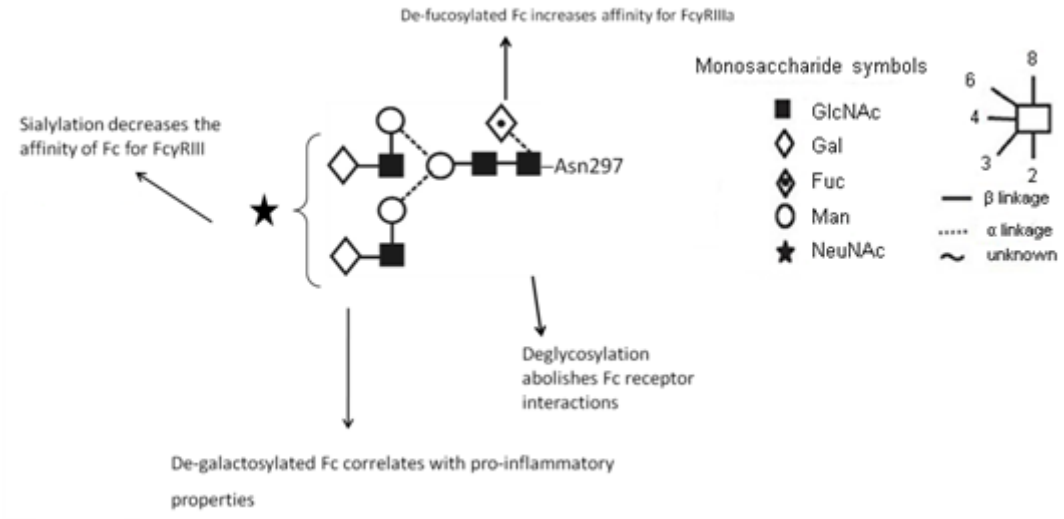


Figure 1.10: Roles of the various carbohydrate moieties of the *N*-linked glycan found at Asn297. Reproduced from Scanlan *et. al*[46]

The presence or absence of IgG Fc oligosaccharides does not directly affect antigen binding but has an important effect on the biological mechanisms that are

activated by the immune complexes formed.[39] Secreted Endo S seems to 'fine tune' the immune system by hydrolysing and neutralising IgG molecules. Deglycosylation of Fc by Endo S abolishes Fc-mediated effector function, such as Fc:FcγR binding, complement binding and activation, and opsonisation.[38] This ultimately leads to an increased survival of the *S. pyogenes* pathogen in human blood resulting in an increased likelihood of infection.[34]

This *N*-linked glycan specificity of Endo S for human IgG and the resultant deactivation of effector functions makes Endo S an important immunomodulatory molecule. This is being currently studied as a potential therapeutic anti-inflammatory for autoimmune diseases such as arthritis and lethal IgG-driven immune (or idiopathic) thrombocytopenic purpura (ITP) in mouse models.[47, 48]

1.7. Aims of the project

The aims of this project are to express and purify recombinant Endo S protein and determine its specificity for different *N*-linked oligo-, high-mannose, and complex type glycans using MALDI-TOF analysis. Furthermore, the specificity of Endo S for core fucosylated *N*-linked complex glycans on human IgG will be investigated. Lastly, the specificity of Endo S will be compared to that of other known endoglycosidases, namely Endo A and Endo H.

1.8. References

1. Low, M. G.; Saltiel, A. R., Structural and Functional Roles of Glycosyl-Phosphatidylinositol in Membranes. *Science* **1988,** 239, (4837), 268-275.

2. Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G., Chemical Modification of Proteins at Cysteine: Opportunities in Chemistry and Biology. *Chemistry-an Asian Journal* **2009,** 4, (5), 630-640.

3. Parodi, A. J., Protein glucosylation and its role in protein folding. *Annual Review of Biochemistry* **2000,** 69, 69-93.

4. Ohtsubo, K.; Marth, J. D., Glycosylation in cellular mechanisms of health and disease. *Cell* **2006,** 126, (5), 855-867.

5. Varki, A. C., R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds.; , Essentials of Glycobiology. *Cold Spring Harbor Laboratory Press* **2009**.

6. Rudd, P. M.; Elliott, T.; Cresswell, P.; Wilson, I. A.; Dwek, R. A., Glycosylation and the immune system. *Science* **2001,** 291, (5512), 2370-2376.

7. Varki, A., Biological Roles of Oligosaccharides - All of the Theories Are Correct. *Glycobiology* **1993,** 3, (2), 97-130.

8. Dwek, R. A., Glycobiology - Towards Understanding the Function of Sugars. *Biochemical Society Transactions* **1995,** 23, (1), 1-25.

9. Sethuraman, N.; Stadheim, T. A., Challenges in therapeutic glycoprotein production. *Current Opinion in Biotechnology* **2006,** 17, (4), 341-346.

10. Roth, J., Protein N-glycosylation along the secretory pathway: Relationship to organelle topography and function, protein quality control, and cell interactions. *Chemical Reviews* **2002,** 102, (2), 285-303.

11. Davis, B. G., Synthesis of glycoproteins. *Chemical Reviews* **2002,** 102, (2), 579-601.

12. Betenbaugh, M. J.; Tomiya, N.; Narang, S.; Hsu, J. T. A.; Lee, Y. C., Biosynthesis of human-type N-glycans in heterollogous systems. *Current Opinion in Structural Biology* **2004,** 14, (5), 601-606.

13. Koury, M. J., Genetic engineering of Pichia pastoris to humanize N-glycosylation of proteins. *Trends in Biotechnology* **2003,** 21, (11), 459-462.

14. Hamilton, S. R.; Bobrowicz, P.; Bobrowicz, B.; Davidson, R. C.; Li, H. J.; Mitchell, T.; Nett, J. H.; Rausch, S.; Stadheim, T. A.; Wischnewski, H.; Wildt, S.; Gerngross, T. U., Production of complex human glycoproteins in yeast. *Science* **2003,** 301, (5637), 1244-1246.

15. Daly, R.; Hearn, M. T. W., Expression of heterologous proteins in Pichia pastoris: a useful experimental tool in protein engineering and production. *Journal of Molecular Recognition* **2005,** 18, (2), 119-138.

16. Li, H. J.; Sethuraman, N.; Stadheim, T. A.; Zha, D. X.; Prinz, B.; Ballew, N.; Bobrowicz, P.; Choi, B. K.; Cook, W. J.; Cukan, M.; Houston-Cummings, N. R.; Davidson, R.; Gong, B.; Hamilton, S. R.; Hoopes, J. P.; Jiang, Y. W.; Kim, N.; Mansfield, R.; Nett, J. H.; Rios, S.; Strawbridge, R.; Wildt, S.; Gerngross, T. U., Optimization of humanized IgGs in glycoengineered Pichia pastoris. *Nature Biotechnology* **2006,** 24, (2), 210-215.

17. Bennett, C. S.; Wong, C.-H., Chemoenzymatic approaches to glycoprotein synthesis. *Chemical Society Reviews* **2007,** 36, (8), 1227-1238.

18. Huang, W.; Ochiai, H.; Zhang, X. Y.; Wang, L. X., Introducing N-glycans into natural products through a chemoenzymatic approach. *Carbohydrate Research* **2008,** 343, (17), 2903-2913.

19. Wang, L. X., Chemoenzymatic synthesis of glycopeptides and glycoproteins through endoglycosidase-catalyzed transglycosylation. *Carbohydrate Research* **2008,** 343, (10-11), 1509-1522.

20. Li, B.; Song, H. J.; Hauser, S.; Wang, L. X., A highly efficient chemoenzymatic approach toward glycoprotein synthesis. *Organic Letters* **2006,** 8, (14), 3081-3084.

21. Fujita, K.; Takegawa, K., Chemoenzymatic synthesis of neoglycoproteins using transglycosylation with endo-beta-N-acetylglucosaminidase A. *Biochemical and Biophysical Research Communications* **2001,** 282, (3), 678-682.

22. Takegawa, K.; Yamabe, K.; Fujita, K.; Tabuchi, M.; Mita, M.; Izu, H.; Watanabe, A.; Asada, Y.; Sano, M.; Kondo, A.; Kato, I.; Iwahara, S., Cloning, sequencing, and expression of Arthrobacter protaphormiae endo-beta-N-acetylglucosaminidase in Escherichia coli. *Archives of Biochemistry and Biophysics* **1997,** 338, (1), 22-28.

23. Takegawa, K.; Yamaguchi, S.; Kondo, A.; Iwamoto, H.; Nakoshi, M.; Kato, I.; Iwahara, S., Transglycosylation Activity of Endo-Beta-N-Acetylglucosaminidase from Arthrobacter-Protophormiae. *Biochemistry International* **1991,** 24, (5), 849-855.

24. Yamamoto, K. J.; Kadowaki, S.; Watanabe, J.; Kumagai, H., Transglycosylation Activity of Mucor Hiemalis Endo-Beta-N-Acetylglucosaminidase Which Transfers Complex Oligosaccharides to the N-Acetylglucosamine Moieties of Peptides. *Biochemical and Biophysical Research Communications* **1994,** 203, (1), 244-252.

25. Yamanoi, T.; Tsutsumida, M.; Oda, Y.; Akaike, E.; Osumi, K.; Yamamoto, K.; Fujita, K., Transglycosylation reaction of Mucor hiemalis endo-beta-N-acetylglucosaminidase using sugar derivatives modified at C-1 or C-2 as oligosaccharide acceptors. *Carbohydrate Research* **2004,** 339, (7), 1403-1406.

26. Yin, J.; Li, L.; Shaw, N.; Li, Y.; Song, J. K.; Zhang, W. P.; Xia, C. F.; Zhang, R. G.; Joachimiak, A.; Zhang, H. C.; Wang, L. X.; Liu, Z. J.; Wang, P., Structural Basis and

Catalytic Mechanism for the Dual Functional Endo-beta-N-Acetylglucosaminidase A. *Plos One* **2009,** 4, (3), -.

27. Fairbanks, A. J.; Heidecke, C. D.; Parsons, T. B., Endohexosaminidase-catalysed glycosylation with oxazoline donors: effects of organic co-solvent and pH on reactions catalysed by Endo A and Endo M. *Carbohydrate Research* **2009,** 344, (18), 2433-2438.

28. Akaike, E.; Tsutsumida, M.; Osumi, K.; Fujita, M.; Yamanoi, T.; Yamamoto, K.; Fujita, K., High efficiency of transferring a native sugar chain from a glycopeptide by a microbial endoglycosidase in organic solvents. *Carbohydrate Research* **2004,** 339, (3), 719-722.

29. Williams, S. J.; Mark, B. L.; Vocadlo, D. J.; James, M. N. G.; Withers, S. G., Aspartate 313 in the Streptomyces plicatus hexosaminidase plays a critical role in substrate-assisted catalysis by orienting the 2-acetamido group and stabilizing the transition state. *Journal of Biological Chemistry* **2002,** 277, (42), 40055-40065.

30. Umekawa, M.; Huang, W.; Li, B.; Fujita, K.; Ashida, H.; Wang, L. X.; Yamamoto, K., Mutants of Mucor hiemalis endo-beta-N-acetylglucosaminidase show enhanced transglycosylation and glycosynthase-like activities. *Journal of Biological Chemistry* **2008,** 283, (8), 4469-4479.

31. Ochiai, H.; Huang, W.; Wang, L. X., Expeditious chemoenzymatic synthesis of homogeneous N-glycoproteins carrying defined oligosaccharide ligands. *Journal of the American Chemical Society* **2008,** 130, (41), 13790-13803.

32. Ray, C. G. K., J. R., Sherris Medical Microbiology - 4th Ed. **2004**.

33. This media comes from the Centers for Disease Control and Prevention's Public Health Image Library (PHIL), with identification number #2110.

34. Collin, M.; Olsen, A., EndoS, a novel secreted protein from Streptococcus pyogenes with endoglycosidase activity on human IgG. *Embo Journal* **2001,** 20, (12), 3046-3055.

35. Bisno, A. L.; Stevens, D. L., Streptococcal skin infections. *New England Journal of Medicine* **1996,** 334, (22), 1478-1478.

36. Bisno, A. L.; Stevens, D. L., Current concepts - Streptococcal infections of skin and soft tissues. *New England Journal of Medicine* **1996,** 334, (4), 240-245.

37. Collin, M.; Svensson, M. D.; Sjoholm, A. G.; Jensenius, J. C.; Sjobring, U.; Olsen, A., EndoS and SpeB from Streptococcus pyogenes inhibit immunoglobulin-mediated opsonophagocytosis. *Infection and Immunity* **2002,** 70, (12), 6646-6651.

38. Allhorn, M.; Briceno, J. G.; Baudino, L.; Lood, C.; Olsson, M. L.; Izui, S.; Collin, M., The IgG-specific endoglycosidase EndoS inhibits both cellular and complement-mediated autoimmune hemolysis. *Blood* **2010,** 115, (24), 5080-5088.

39. Jefferis, R., Glycosylation as a strategy to improve antibody-based therapeutics. *Nature Reviews Drug Discovery* **2009,** 8, (3), 226-234.

40. Huhn, C.; Selman, M. H. J.; Ruhaak, L. R.; Deelder, A. M.; Wuhrer, M., IgG glycosylation analysis. *Proteomics* **2009,** 9, (4), 882-913.

41. Mimura, Y.; Church, S.; Ghirlando, R.; Ashton, P. R.; Dong, S.; Goodall, M.; Lund, J.; Jefferis, R., The influence of glycosylation on the thermal stability and effector function expression of human IgG1-Fc: properties of a series of truncated glycoforms. *Molecular Immunology* **2000,** 37, (12-13), 697-706.

42. Arnold, J. N.; Wormald, M. R.; Sim, R. B.; Rudd, P. M.; Dwek, R. A., The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annual Review of Immunology* **2007,** 25, 21-50.

43. Mizuochi, T.; Taniguchi, T.; Shimizu, A.; Kobata, A., Structural and Numerical Variations of the Carbohydrate Moiety of Immunoglobulin-G. *Journal of Immunology* **1982,** 129, (5), 2016-2020.

44. Kaneko, Y.; Nimmerjahn, F.; Ravetch, E. V., Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. *Science* **2006,** 313, (5787), 670-673.

45. Kanda, Y.; Yamane-Ohnuki, N.; Sakai, N.; Yamano, K.; Nakano, R.; Inoue, M.; Misaka, H.; Iida, S.; Wakitani, M.; Konno, Y.; Yano, K.; Shitara, K.; Hosoi, S.; Satoh, M., Comparison of cell lines for stable production of fucose-negative antibodies with enhanced ADCC. *Biotechnology and Bioengineering* **2006,** 94, (4), 680-688.

46. Scanlan, C. N.; Burton, D. R.; Dwek, R. A., Making autoantibodies safe. *Proceedings of the National Academy of Sciences of the United States of America* **2008,** 105, (11), 4081-4082.

47. Nandakumar, K. S.; Collin, M.; Olsen, A.; Nimmerjahn, F.; Blom, A. M.; Ravetch, J. V.; Holmdahl, R., Endoglycosidase treatment abrogates IgG arthritogenicity: Importance of IgG glycosylation in arthritis. *European Journal of Immunology* **2007,** 37, (10), 2973-2982.

48. Collin, M.; Shannon, O.; Bjorck, L., IgG glycan hydrolysis by a bacterial enzyme as a therapy against autoimmune conditions. *Proceedings of the National Academy of Sciences of the United States of America* **2008,** 105, (11), 4265-4270.

Chapter 2

Materials and Methods

Molecular biology techniques are based upon procedures described in Sambrook and Russell (2001) Molecular Cloning: A Laboratory Manual (3rd Edition).[1]

2.1. General Considerations for Biology

2.1.1. Sterile conditions

Standard sterile techniques were used for experimental work where appropriate. All growth media and applicable equipment was autoclaved at 121°C for 45 minutes. For procedures requiring a sterile atmosphere, a Heraeus HERAsafe sterile cabinet was used. Water was purified using a Milli-Q purification system.

2.1.2. Antibiotics

All plasmids used confer ampicillin resistance. Stock solutions of 100 mg/mL of ampicillin solution (Sigma-Aldrich, CAS 69-52-3) were prepared with Milli-Q H_2O, filtered through a 0.22 µm membrane filter and stored at 4°C until use.

2.1.3. Protein concentration

Protein concentration was determined using the bicinchoninic acid (BCA) assay at 280 nm using known concentrations of bovine serum albuminute (BSA) as standards.

2.1.4. Cells

Unless otherwise stated the cells [BL21(DE3) (NEB), TOP10 OneShot (Invitrogen), NovaBlue GigaSingles (Novagen) and XL10-Gold (Stratagene)] were used following the suppliers' instructions.

2.1.5. Preparation of growth media

25 g of LB media [tryptone (10 g), NaCl (10 g), yeast extract (5 g)] (Melford Laboratories Ltd.) was dissolved into 1 L of deionised water. The flask was covered with a foam stopper, tinfoil and placed into the autoclave for 45 minutes. Media was allowed to cool before use.

2.1.6. Preparation of Agar Plates

6.25 g of LB media was dissolved in 250 mL of deionised water in a 500 mL Duran bottle. 3.75 g of biological agar (Melford Laboratories Ltd.) was added to this mixture and the solution media was autoclaved at 121°C for 45 minutes before use. Ampicillin (final concentration = 100 µg/mL, 250 µL) was added upon cooling, and solution poured into Petri dishes (~20 mL per plate) and allowed to set at 4°C.

2.1.7. Transformation of pUC57:*ndoS* plasmid into *E. coli* BL21 (DE3) competent cells

All transformations were carried out using protocols recommended by the manufacturer [BL21 (DE3) competent cells, NEB].

1 µL of pUC57:*ndoS* plasmid was added to 50 µL of competent cells BL21 (DE3) in a polypropylene round-bottomed tube and was incubated on ice for 30 minutes. Then, the tube was incubated for exactly 30 seconds in a 42°C water bath. After this, they were placed back on ice for 2 minutes. Next, 250 µL pre-warmed (37°C) SOC medium was added to the tube, shaken for 1 hour at 37°C before 20 µL of the cells were plated directly to the previously prepared pre-warmed LB agar plates (ampicillin). A concentrated plate containing 200 µL of the cells was also set up as a comparison. The plates were inverted and incubated at 37^{o}C overnight.

2.1.8. Overnight bacterial culture

A colony from the normal plate [*i.e.* not the concentrated plate] was selected by pipette tip and was added to 20 mL LB media containing 20 µL of ampicillin (final concentration = 100 µg/mL). This was grown up in a shaking incubator (37°C, 225 rpm) overnight.

2.1.9. Purification of plasmid DNA from overnight *E. coli* culture in LB medium

The overnight culture described previously was centrifuged [4,500 × g, 10 minutes, 4°C], and the resulting supernatant decanted off. The bacterial cell pellet produced was resuspended in 250 µL of Buffer P1 (including RNase A) and transferred to a 1.5 mL microcentrifuge tube. Next, 250 µL of Buffer P2 was added and mixed by inverting the tube 4-6 times followed by the addition of 350 µL of Buffer N3 again mixed by inverting the tube 4-6 times. The mixture was centrifuged [15,000 × g, 20 minutes, 4°C] and the resultant supernatant was added to the QIAprep spin column where it was centrifuged [15,000 × g, 1 minute, 4°C]. Flow through was discarded and the QIAprep spin column was washed by the addition of 0.5 mL Buffer PB and again centrifuged [15,000 × g, 1 minute, 4°C]. The flow through was discarded and the column was washed by the addition of 0.75 mL Buffer PE, centrifuged [15,000 × g, 1 minute, 4°C], flow through discarded and centrifuged again [15,000 × g, 1 minute, 4°C]. The tube was left open for 2 minutes at room temperature. To elute the DNA, the QIAprep spin column was placed in a new 1.5 mL microcentrifuge tube and 30 µL of Milli-Q H_2O was added to the column, left to stand for 5 minutes and then centrifuged [15,000 × g, 1 minute, 4°C]. The flow through containing the DNA was stored at -20°C until needed.

2.1.10. 1% DNA Agarose gel analysis

0.3 g Agarose was dissolved in 50 mL of TAE (Tris acetate EDTA) buffer by heating for 1 minute in the microwave. Upon cooling, 4 μL ethidium bromide (Thermo Scientific, Northumberland, UK) was added and poured into gel a holder. The gel was left for 20 minutes to set. For each sample, 10 μL of reaction mixture with 2 μL of 5 x loading dye (NEB) were loaded to the wells, and the gel was placed in tank containing TAE buffer and ran at 80 V for 60 minutes. Samples were run alongside a 1 kb DNA ladder (NEB). The resulting DNA bands were visualised by illumination with an ultraviolet lamp (λ = 366 nm).

2.1.11. QIAquick Gel extraction protocol

The desired DNA fragment was excised from the agarose DNA gel with a clean, sharp scalpel. The fragment was weighed and 3 volumes of Buffer QG was added to the tube. The tube was incubated at 50°C for 10 minutes, mixed by vortexing every 2 to 3 minutes as needed. After the gel fragment had completely dissolved, 1 volume of isopropanol was added to the sample and mixed by inverting. The QIAquick colum was placed into a 2 mL collection tube and sample was applied to the column. The tube was centrifuged for 1 minute [15,000 × g, 4°C] to bind the DNA. To wash, 0.75 mL of Buffer PE was added to QIAquick column and centrifuged for 1 minute [15,000 × g, 4°C], with the flow through being discarded. To elute the DNA, 30 μL of Milli-Q H_2O was added to the centre of the column and left to stand at RT for 5 minutes. The column was placed into a clean 1.5 mL tube and centrifuged for 1 minute [15,000 × g, 4°C].

2.1.12. Transformation of plasmids pUC57:*ndoS* and pGEX-4T-1 into *E.coli* BL21 (DE3) competent cells

All transformations were carried out using protocols recommended by the manufacturer [BL21 (DE3) competent cells, NEB].

To 2 x 150 µL sample of BL21 (DES) competent cells, 1 µL pUC57:*ndoS* plasmid and 0.5 µL pGEX-4T-1 respectively were added directly to polypropylene round-bottomed tubes and these were incubated on ice for 30 minutes. The tubes were 'heat shocked' for exactly 30 seconds in a 42°C water bath. After this, they were replaced back on ice for 2 minutes. Next, 500 µL of pre-warmed LB medium (37°C) was added to the tubes, shaken for 1 hour at 37°C and 20 µL of the cells were plated directly to the previously prepared pre-warmed LB agar plates (ampicillin). A concentrated plate containing 200 µL of the cells was also set up as a comparison. The plates were inverted and incubated at 37°C overnight.

2.1.13. QIAprep Spin DNA miniprep

The overnight sample of small starter culture was spun down [4,500 × g, 4°C, 10 minutes], and the supernatant was discarded. Pelleted cells were resuspended in 250 µL Buffer P1 and the solution transferred to microcentrifuge tube. 250 µL of Buffer P2 was added and the tube was mixed by inverting it 4-6 times. Next, 350 µL of Buffer N3 was added to the solution and mixed by inverting it 4-6 times. The tube was then centrifuged for 10 minutes [15,000 × g, 4°C] and resulting supernatant was applied to the QIAprep spin column by pipetting. The tube was centrifuged for 1 minute [15,000 × g, 4°C], and the flow through was discarded. Next, 0.75 mL Buffer PE was added to the tube and subsequently centrifuged for 1 minute [15,000 × g, 4°C]. Flow through was discarded and tube was centrifuged for an additional 1 minute [15,000 × g, 4°C]. To elute DNA, the

QIAprep column was placed into a clean 1.5 mL microcentrifuge tube and 30 µL of Milli-Q water was added to the centre of the column. After standing at room temperature for 5 minutes, the tube was centrifuged for 1 minute [15,000 × g, 4°C].

2.1.14. DNA sequencing

The concentration of eluted DNA was measured using a Nanodrop spectrophotometer. DNA sequencing was carried out by Geneservice (Department of Biochemistry, Oxford University) on an ABI Prism 3730xl DNA analyzer. T7R, T7R primers (pUC57) and PGEX5', PGEX3' (pGEX-4T-1) primers were used, as listed in Table 2.1. Sequenced files were viewed by FinchTV v1.4 software (Geospiza)

Primer	Sequence	Length	% G+C	Tm (°C)
PGEX 3'	CCGGGAGCTGCATGTGTCAGAGG	23	65	62
PGEX 5'	GGGCTGGCAAGCCACGTTTGGTG	23	65	62
T7F	TAATACGACTCACTATAGGG	20	40	48
T7R	GCTAGTTATTGCTCAGCGG	19	53	51

Table 2.1. Sequencing Primers for pUC57 and pGEX-4T-1 vectors

2.1.15. DNA Amplification by PCR

The One*Taq*™ DNA Polymerase kit (NEB) was used for polymerase chain reactions (PCR). Reaction mixtures for PCR amplification were typically composed of cDNA template (100 pg), forward primer (33 pmol, 5'→3'), reverse primer (33 pmol, 3'→5'), dNTP mixture (deoxyribonucleotides triphosphate; 200 µM of each dNTP), 1.5 mM magnesium chloride buffer, One*Taq* DNA polymerase (2.5 U), Milli-Q H_2O (to give total volume of 50 µL). PCR mixtures were vortexed, centrifuged and placed a thermal cycler. Fragments were amplified using recommended One*Taq*™ thermal cycling parameters, as shown in Table 2.2. PCR products were separated and visualised using 1.0% TAE-agarose gel electrophoresis.

No. of cycles	Temperature (°C)	Time
1	94	30 seconds
25	94	15 seconds
	58	30 seconds
	68	45 seconds
1	68	5 minutes
1	4	∞

Table 2.2. Thermal cycling parameters for an example PCR reaction

2.1.16. PCR Purification kit protocol

5 Volumes of Buffer PB was added to 1 volume of the PCR sample and was mixed by inverting. A QIAquick spin column was placed into a 2 mL collection tube. The sample prepared was applied to the column and centrifuged for 1 minute [15,000 × g, 4°C]. The flow through was discarded and column replaced back into the same tube. To wash, 0.75 mL Buffer PE was added to the column and centrifuged for 1 minute [15,000 × g, 4°C]. The flow through was discarded and column replaced back into the same tube. The tube was centrifuged for an additional 1 minute [15,000 × g, 4°C]. To elute the DNA, 30 µL of Milli-Q H_2O was added to the centre of the column and left to stand at RT for 5 minutes. The column was placed into a clean 1.5 mL tube and centrifuged for 1 minute [15,000 × g, 4°C].

2.1.17. Restriction Enzyme Digest

The PCR fragment of the Endo S gene and pGEX-4T-1 vector were digested at 37°C for two hours separately with *BamHI* and *NotI* restriction endonucleases using NEB recommended general rection mixtures. Following digestion, enzyme digested products were then purified using QIAprep Gel purification kit™ (Qiagen, Crawley UK).

2.1.18. Ligation reactions

Ligation reactions were carried out using the Quick Ligation™ Kit (NEB). Ligation of the restriction enzyme digested fragment with the pGEX-4T-1 vector was performed using Quick T4 DNA ligase (1 U), DNA ligase buffer, the PCR fragment (100 ng), the vector (50 ng) and Milli-Q H_2O. Reactions were incubated at 25°C overnight before transformation into *E. coli.*

2.1.19. Large scale DNA purification

To isolate milligram quantities of Fc, pHLsec:Fc DNA for transient expression in Human Embryonic Kidney (HEK) cells (CRL-1573), bacterial colonies transformed with target DNA were inoculated in 0.5 L LB media containing ampicillin (100 µg/mL) and were grown overnight in a rotary shaker (225 rpm, 37°C). Plasmid DNA was isolated from the resulting spun down cell pellets (8,000 × g, 10 minutes, 4°C,) using the Endotoxin-Free Plasmid Mega Kit (Qiagen, Crawley, UK).

2.2. Protein Expression

2.2.1. Overexpression of GST-Endo S fusion protein from *E. coli* BL21(DE3) cells

To 1 L of autoclaved LB media containing ampicillin (100 μg/mL), a 10 mL overnight culture of *E. coli* BL21(DE3) cells was added after the removal of a blank sample of LB (1.2 mL) and the LB was incubated (37°C, 225 rpm). The optical density ($OD_{600\ nm}$) was monitored using the blank sample as a reference until it reached 0.6-0.7. After 2.75 hours, this value was reached, and expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) (3.5 mL, 100 mM) and the LB was further incubated overnight. The cells were centrifuged (10,000 × g, 4°C, 8 minutes) and the cell pellet was collected and stored at -80°C by 'snap freezing' with liquid nitrogen.

2.2.2. Purification of GST-Endo S fusion protein by ÄKTA FPLC™

Cells were resuspended in 15 mL of binding buffer (PBS buffer, pH 7.4, 10 mM). 10 mg of lysozyme and 10 μg of DNase A were then added, and the mixtures stirred at 4°C for 1 hour. The cells were then lysed by sonication (3 x 30 seconds) with a wait time of one minute in between, and the cell debris collected by centrifugation (20,000 × g, 60 minutes). The supernatant was then purified using a glutathione S-transferase (GST) affinity column (GSTrap™ HP, GE Healthcare) washing first with binding buffer (~ 30 column volumes) and then eluting with elution buffer (50 mM Tris-HCl, 10 mM reduced L-glutathione, pH 8.0). Eluent fractions were analysed by SDS-PAGE [4%-12% gel, MOPS buffer, 200 V, 60 minutes] and pure fractions combined. After the combined fractions were dialysed overnight (30,000 M_w membrane, 1 L PBS, pH 7.4, 10 mM), they were concentrated by Viva Spin 15R (30,000 M_w, 3,000 x *g*, 60 minutes). Protein concentration was then determined by BCA assay (1.14 mg/mL). [as seen on SDS-PAGE as M_w GST-Endo S = 134 kDa].

2.2.3. Cleavage of the GST-tag by reaction with thrombin

One unit of thrombin (NEB) will cleave >90% of 100 µg of GST-Endo S protein in PBS buffer (10 mM, pH 7.3) at 37°C for 16 hours. [N.B. M_w Endo S = 108 kDa, GST-tag = 26 kDa]. In general, 1 µL of thrombin = one unit. One unit is approximately equivalent to 0.2 NIH units. 1.4 mg/mL of purified GST-Endo S fusion protein was incubated with 14 µL thrombin in 1 mL PBS buffer (10 mM, pH 7.3) at 37°C for 16 hours. Protein sample was stored at -20°C until it was further purified by ÄKTA FPLC™.

2.2.4. Purification of Endo S fusion protein by ÄKTA FPLC™

The cleaved protein was then purified by GST affinity chromatography (GSTrap™ HP, GE Healthcare), washing first with binding buffer (PBS, pH 7.4, 10 mM, ~30 column volumes) and then eluting with (50 mM Tris-HCl, 10 mM reduced L-glutathione, pH 8.0). Eluent fractions were analysed by SDS-PAGE [4%-12% gel, MOPS buffer, 200 V, 60 minutes] and pure fractions were combined. After the combined fractions were submitted to overnight dialysis (30,000 M_w membrane, 1 L PBS, pH 7.4), they were concentrated by Viva Spin 15R (30,000 M_w, 3,000 x *g*, 60 minutes). Protein concentration was then determinuteed by BCA assay (3.72 mg/mL). The mass of the purified Endo S protein (M_w = 108 kDa) was confirmed by MALDI analysis.

2.2.5. SDS-PAGE analysis

Recombinantly expressed proteins were visualised on 4%-12% Bis-Tris gradient gels (Invitrogen, Paisley, UK). 10 µL of protein were lysed by 2 µL of 5× reducing sample buffer and subsequently placed on the heat block at 100°C for 3-4 minutes. 15 µL of sample supernatant was loaded into each well. The gel was placed in the electrophoresis tank containing 1× running buffer (MES or MOPS) and ran at 200 V for 60 minutes. Depending on the molecular mass of the protein, protein gels were run in SDS running buffer containing MES (2-(*N*-morpholino)ethanesulfonic acid) or MOPS (3-(*N*-

morpholino)propanesulfonic acid). Protein molecular weights were approximated by comparison to a protein marker Perfect Protein Markers 15-150 kDa or 10-225 kDa (Novagen). The resultant gel was visualised by staining in Coomassie Brilliant Blue R-250 stain solution for at least 60 minutes.

2.2.6. Large Scale Transfection of HEK293T with Fc DNA

The transfection of HEK293 cells was carried as published previously.[2] For large-scale transfections, polyethyleneimine (PEI) was used as the transfection reagent. For all transfections, a DNA to PEI mass ratio 1:1.5 was used. Cells were transiently transfected in tissue culture flasks T225 with 2 mg purified DNA per 1 L of 90% confluent cells. To generate the hybrid glycans, 20 µM swainsonine (Sigma-Aldrich, CAS 72741-87-8), a class II golgi α-mannosidase inhibitor, was added to the medium. Cells were incubated for 10 days at 37°C in the presence of 5% CO_2.

2.2.7. Purification of Wild Type (WT) Fc protein and Swainsonine (SWN) Fc protein

Cell supernatant was harvested ten days following transfection. Cell debris was spun down [10,000 × g, 4°C, 60 minutes], the media sterile filtered through a 0.22 µm membrane filter, and incubated with Chelating Sepharose Fast Flow Ni^{2+}-agarose beads (GE Healthcare, Buckinghamshire, UK) for 2 hours at 16°C in PBS buffer (10 mM, pH 7.4). After incubation, nickel beads were washed with PBS (20 column volumes), followed by PBS with 5 mM imidazole buffer (5 column volumes). His_6-tagged protein was eluted with 500 mM imidazole buffer plus 150 mM NaCl. All buffered contained Tris-HCl (10 mM, pH 8.0). Protein was concentrated with M_w 10 kDa cut-off concentrators (Vivascience, Gloucestershire, UK) as described before and filtered with a 0.22 µm spin-filter unit (Millipore, Watford, UK). Protein purity was assessed by SDS-

PAGE analysis. The yield of pure WT Fc protein was 58 mg/mL and the yield of pure SWN Fc protein was 38 mg/mL.

2.3. Glycan Cleavage reactions

2.3.1. Human Immunoglobulin G

Human Immunoglobulin G (IgG) was obtained from Sigma-Aldrich [CAS I4506]. Protein was ≥95% purity (SDS-PAGE), essentially salt-free, and received as a lyophilized powder. IgG was stored in PBS buffer (50 µg samples in 50 µL solutions, 10 mM, pH 7.4) at -20 °C until use. Samples were incubated in PBS buffer (10 mM, pH 7.4) for one hour at 37°C.

2.3.2 Endo A digestion

10 µL (10 µg) of IgG was denatured with 1 µL of 10× denaturing buffer (0.5% SDS, 40 mM DTT) by heating at 100°C for 10 minutes. The reaction mixture was made up to a total volume of 20 µL by adding 2 µL of Endo A (10 milliunits, BGD group) and 7 µL of Milli-Q filtered water. The reaction mixture was incubated at 37°C for one hour.

2.3.3. Endo H digestion

10 µL (10 µg) of IgG was denatured with 1 µL of 10× denaturing buffer (0.5% SDS, 40 mM DTT) by heating at 100°C for 10 minutes. The reaction mixture was made up to a total volume of 20 µL by adding 2 µL of Endo H (250 NEB units), 2 µL of 10× G5 reaction buffer (50 mM sodium citrate, pH 5.5) and 5 µL Milli-Q filtered water. The reaction mixture was incubated at 37°C for one hour.

2.3.4. Endo S digestion

10 µL (100 µg, 33.3 µmol) of IgG was denatured with 1 µL of 10× denaturing buffer (0.5% SDS, 40 mM DTT) by heating reaction at 100°C for 10 minutes. The reaction mixture was made up to a total volume of 20 µL by adding 2 µL of Endo S (1 µg, 0.46

μmol) and 7 μL of Milli-Q filtered water. The reaction mixture was incubated at 37°C for one hour.

2.3.5. PNGase F digestion

10 μL (10 μg) of IgG was denatured with 1 μL of 10× denaturing buffer (0.5% SDS, 40 mM DTT) by heating at 100°C for 10 minutes. The reaction mixture was made up to a total volume of 20 μL by adding 2 μL of PNGase F (250 NEB units), 2 μL of 10× G7 reaction buffer (50 mM sodium phosphate, pH 7.5), 2 μL 10 % NP40 solution and 3 μL Milli-Q filtered water. The reaction mixture was incubated at 37°C for one hour.

All reaction mixtures were analysed by SDS-PAGE analysis [4%-12% gel, MOPS buffer, 200 V, 60 minutes].

2.4. MALDI mass spectrometric analysis

2.4.1. For Endo S and IgG proteins

For each analysis, 2 μL of the reaction mixture of IgG:Endo S was diluted to 20 μL with 0.1 % trifluoroacetic acid (TFA) in water. The matrix solution was prepared using a solution of mass spectrometry grade sinapinic acid (10 mg/mL in 1:1 water:acetonitrile with 0.1% TFA). From the sample solution, 0.5 μL was spotted onto a stainless steel target and allowed to co-crystallize at room temperature with 2 μL of the matrix over 30 minutes.

MALDI analysis was conducted on a Water MALDI Micro MX spectrometer with TOF detection, in positive (+ve) linear mode. Unmodified IgG protein was used as a lock mass calibrant (M_w = 150 kDa). For each comparative set of experiments, laser energy and

pulse width were optimised for a modified protein sample, and kept identical for the remainder of the sample set. In general, multiple laser shots were combined to produce the final spectrum. Data were further processed using Mass Lynx 4.1 (Water-Micromass, Manchester, UK). Mass spectra were smoothed with mean smoothing.

2.4.2. For *N*-linked glycans released from $gp120_{BAL}$ and Fc proteins

Oligosaccharides were released from target glycoproteins with PNGase F (NEB) from Coomassie blue-stained NuPAGE.[3] Excised bands were washed three times alternatively with acetonitrile and deionised water, and rehydrated with a 3000 Units/mL of PNGase F water solution. After incubation for 12 hours at 37°C, the enzymatically released *N*-linked glycans were eluted with water.

For enzymatic digest, *N*-linked glycans (released from 20-50 µg protein, 29.17 µmol) were incubated with Endo S (1 µg, 1 µL of a 1 mg/mL solution, 0.93 µmol), made up to a total volume of 10 µL with Milli-Q H_2O and incubated at 37°C for two hours. Samples were analysed by positive ion matrix-assisted laser desorption/ionization (MALDI) time-of-flight (TOF) mass spectra with a Shimazu AXIMA TOF^2 MALDI TOF/TOF mass spectrometer (Kratos Analytical, Manchester, UK) fitted with delayed extraction and a nitrogen laser (337 nm) as previously described.[4]

Samples were cleaned on a Nafion 117 membrane (Sigma-Aldrich, CAS 274674), and then prepared for mass spectrometry by adding 0.5 µL of an aqueous solution of the glycans to the matrix solution (0.3 µL of a solution of 2,5-dihydroxybenzoic acid (DHB) in acetonitrile:water (1:1, v:v)) on the stainless steel target plate and allowing it to dry at room temperature (20 minutes). The sample/matrix mixture was then recrystallized from

0.2 µl ethanol. Further samples were examined after removal of any potential sialic acids by heating at 80°C for 1 hour with 1.0% acetic acid solution.

2.5. References

1. Sambrook J, M. F., E.F and Maniatis, T., *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, **1989**.

2. Jones, E. Y.; Aricescu, A. R.; Lu, W. X., A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallographica Section D-Biological Crystallography* **2006,** *62,* 1243-1250.

3. Kuster, B.; Wheeler, S. F.; Hunter, A. P.; Dwek, R. A.; Harvey, D. J., Sequencing of N-linked oligosaccharides directly from protein gels: In-gel deglycosylation followed by matrix-assisted laser desorption/ionization mass spectrometry and normal-phase high-performance liquid chromatography. *Analytical Biochemistry* **1997,** *250* (1), 82-101.

4. Harvey, D. J.; Royle, L.; Radcliffe, C. M.; Rudd, P. M.; Dwek, R. A., Structural and quantitative analysis of N-linked glycans by matrix-assisted laser desorption ionization and negative ion nanospray mass spectrometry. *Analytical Biochemistry* **2008,** *376* (1), 44-60.

Chapter 3

N-linked glycan specificity of bacterial endo-β-*N*-acetylglucosaminidase S

3.1. Expression and purification of recombinant Endo S

Work by Collins *et. al*[1] has shown that bacterial endo-β-*N*-acetylglucosaminidase (Endo) S secreted by *Streptococcus pyogenes* hydrolyses the functionally important *N*-linked complex biantennary, core fucosylated oligosaccharide that is found on IgG. This sugar is attached at the Asn-297 position in the CH_2 domain on each heavy chain found on the Fc portion on IgG (Figure 3.1).[2] Recombinant Endo S was expressed and purified in order to determine its mode of action. Firstly, it was necessary to test its specificity against a range of mammalian *N*-linked glycans.

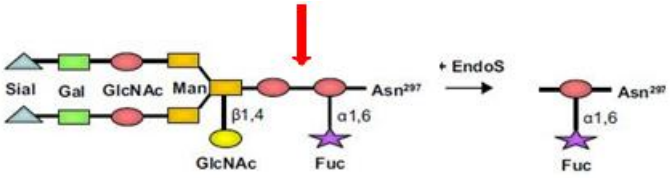


Figure 3.1: Site of Endo S hydrolysis in the CH_2 domain on IgG.

The nucleotide sequence encoding for Endo S (M_w = 108 kDa) was obtained from GenBank under accession No. AF296340.[1b] The DNA sequence of the desired *ndoS* gene was codon optimised for expression in *Escherichia coli* and was synthesised by GenScript. The gene was further subcloned from cloning vector pUC57 into expression vector pGEX-4T-1 using *BamHI* and *NotI* restriction endonucleases (NEB), and T4 DNA ligase (Invitrogen), which encodes an *N*-terminal glutathione-S-transferase (GST) tag, a thrombin cleavage site and confers ampicillin resistance (Figure 3.2). The resulting protein is predicted to have 996 residues, an average M_w of 111,806 Da and a pI of 6.68.

The nucleotide (Table 3.1.) and translated protein sequence (Table 3.2.) are shown below.

DNA sequence of *ndoS* gene (2988 Nucleotides):

```
   1 ATGGATAAAC ATTTGTTGGT AAAAAGAACA CTAGGGTGTG TTTGTGCTGC AACGTTGATG
  61 GGAGCTGCCT TAGCGACCCA CCATGATTCA CTCAATACTG TAAAAGCGGA GGAGAAGACT
 121 GTTCAGGTTC AGAAAGGATT ACCTTCTATC GATAGCTTGC ATTATCTGTC AGAGAATAGC
 181 AAAAAAGAAT TTAAAGAAGA ACTCTCAAAA GCGGGGCAAG AATCTCAAAA GGTCAAAGAG
 241 ATATTAGCAA AAGCTCAGCA GGCAGATAAA CAAGCTCAAG AACTTGCCAA AATGAAAATT
 301 CCTGAGAAAA TACCGATGAA ACCGTTACAT GGTCCTCTCT ACGGTGGTTA CTTTAGAACT
 361 TGGCATGACA AAACATCAGA TCCAACAGAA AAAGACAAAG TTAACTCGAT GGGAGAGCTT
 421 CCTAAAGAAG TAGATCTAGC CTTTATTTTC CACGATTGGA CAAAAGATTA TAGCCTTTTT
 481 TGGAAAGAAT TGGCCACCAA ACATGTGCCA AAGTTAAACA AGCAAGGGAC ACGTGTCATT
 541 CGTACCATTC CATGGCGTTT CCTAGCTGGG GGTGATAACA GTGGTATTGC AGAAGATACC
 601 AGTAAATACC CAAATACACC AGAGGGAAAT AAAGCTTTAG CCAAAGCTAT TGTTGATGAA
 661 TATGTTTATA AATACAACCT TGATGGCTTA GATGTGGATG TTGCACATGA TAGTATTCCA
 721 AAAGTTGACA AAAAAGAAGA TACAGCAGGC GTAGAACGCT CTATTCAAGT GTTTGAAGAA
 781 ATTGGGAAAT TAATTGGACC AAAAGGTGTT GATAAATCGC GGTTATTTAT TATGGATAGC
 841 ACCTACATGG CTGATAAAAA CCCATTGATT GAGCGAGGAG CTCCTTATAT TAATTTATTA
 901 CTGGTACAGG TCTATGGTTC ACAAGGAGAG AAAGGTGGTT GGGAGCCTGT TTCTAATCGA
 961 CCTGAAAAAA CAATGGAAGA ACGATGGCAA GGTTATAGCA AGTATATTCG TCCTGAACAA
1021 TACATGATTG GTTTTTCTTT CTATGAGGAA AATGCTCAAG AAGGGAATCT TTGGTATGAT
1081 ATTAATTCTC GCAAGGACGA GGACAAAGCA AATGGAATTA ACACTGACAT AACTGGAACG
1141 CGTGCCGAAC GGTATGCAAG GTGGCAACCT AAGACAGGTG GGGTTAAGGG AGGTATCTTC
1201 TCCTACGCTA TTGACCGAGA TGGTGTAGCT CATCAACCTA AAAAATATGC TAAACAGAAA
1261 GAGTTTAAGG ACGCAACTGA TAACATCTTC CACTCAGATT ATAGTGTCTC CAAGGCATTA
1321 AAGACAGTTA TGCTAAAAGA TAAGTCGTAT GATCTGATTG ATGAGAAAGA TTTCCCAGAT
1381 AAGGCTTTGC GAGAAGCTGT GATGGCGCAG GTTGGAACCA GAAAAGGTGA TTTGGAACGT
1441 TTCAATGGCA CATTACGATT GGATAATCCA GCGATTCAAA GTTTAGAAGG TCTAAATAAA
1501 TTTAAAAAAT TAGCTCAATT AGACTTGATT GGCTTATCTC GCATTACAAA GCTCGACCGT
1561 TCTGTTTTAC CCGCTAATAT GAAGCCAGGC AAAGATACCT TGGAAACAGT TCTTGAAACC
1621 TATAAAAAGG ATAACAAAGA AGAACCTGCT ACTATCCCAC CAGTATCTTT GAAGGTTTCT
1681 GGTTTAACTG GTCTGAAAGA ATTAGATTTG TCAGGTTTTG ACCGTGAAAC CTTGGCTGGT
1741 CTTGATGCCG CTACTCTAAC GTCTTTAGAA AAAGTTGATA TTTCTGGCAA CAAACTTGAT
1801 TTGGCTCCAG GAACAGAAAA TCGACAAATT TTTGATACTA TGCTATCAAC TATCAGCAAT
1861 CATGTTGGAA GCAATGAACA AACAGTGAAA TTTGACAAGC AAAAACCAAC TGGGCATTAC
1921 CCAGATACCT ATGGGAAAAC TAGTCTGCGC TTACCAGTGG CAAATGAAAA AGTTGATTTG
1981 CAAAGCCAGC TTTTGTTTGG GACTGTGACA AATCAAGGAA CCCTAATCAA TAGCGAAGCA
2041 GACTATAAGG CTTACCAAAA TCATAAAATT GCTGGACGTA GCTTTGTTGA TTCAAACTAT
2101 CATTACAATA ACTTTAAAGT TTCTTATGAG AACTATACCG TTAAAGTAAC TGATTCCACA
2161 TTGGGAACCA CTACTGACAA AACGCTAGCA ACTGATAAAG AAGAGACCTA TAAGGTTGAC
2221 TTCTTTAGCC CAGCAGATAA GACAAAAGCT GTTCATACTG CTAAAGTGAT TGTTGGTGAC
2281 GAAAAAACCA TGATGGTTAA TTTGGCAGAA GGCGCAACAG TTATTGGAGG AAGTGCTGAT
2341 CCTGTAAATG CAAGAAAGGT ATTTGATGGG CAACTGGGCA GTGAGACTGA TAATATCTCT
2401 TTAGGATGGG ATTCTAAGCA AAGTATTATA TTTAAATTGA AAGAAGATGG ATTAATAAAG
2461 CATTGGCGTT TCTTCAATGA TTCAGCCCGA AATCCTGAGA CAACCAATAA ACCTATTCAG
2521 GAAGCAAGTC TACAAATTTT TAATATCAAA GATTATAATC TAGATAATTT GTTGGAAAAT
2581 CCCAATAAAT TTGATGATGA AAAATATTGG ATTACTGTAG ATACTTACAG TGCACAAGGA
2641 GAGAGAGCTA CTGCATTCAG TAATACATTA AATAATATTA CTAGTAAATA TTGGCGAGTT
2701 GTCTTTGATA CTAAAGGAGA TAGATATAGT TCGCCAGTAG TCCCTGAACT CCAAATTTTA
2761 GGTTATCCGT TACCTAACGC CGACACTATC ATGAAAACAG TAACTACTGC TAAAGAGTTA
2821 TCTCAACAAA AAGATAAGTT TTCTCAAAAG ATGCTTGATG AGTTAAAAAT AAAAGAGATG
2881 GCTTTAGAAA CTTCTTTGAA CAGTAAGATT TTTGATGTAA CTGCTATTAA TGCTAATGCT
2941 GGAGTTTTGA AAGATTGTAT TGAGAAAAGG CAGCTGCTAA AAAAATAA
```

Table 3.1: Nucleotide sequence of *ndoS* gene. Accession No. AF296340.

Amino acid sequence of Endo S protein (996 amino acids):

```
1    MDKHLLVKRT LGCVCAATLM GAALATHHDS LNTVKAEEKT
41   VQVQKGLPSI DSLHYLSENS KKEFKEELSK AGQESQKVKE
81   ILAKAQQADK QAQELAKMKI PEKIPMKPLH GPLYGGYFRT
121  WHDKTSDPTE KDKVNSMGEL PKEVDLAFIF HDWTKDYSLF
161  WKELATKHVP KLNKQGTRVI RTIPWRFLAG GDNSGIAEDT
201  SKYPNTPEGN KALAKAIVDE YVYKYNLDGL DVDVAHDSIP
241  KVDKKEDTAG VERSIQVFEE IGKLIGPKGV DKSRLFIMDS
281  TYMADKNPLI ERGAPYINLL LVQVYGSQGE KGGWEPVSNR
321  PEKTMEERWQ GYSKYIRPEQ YMIGFSFYEE NAQEGNLWYD
361  INSRKDEDKA NGINTDITGT RAERYARWQP KTGGVKGGIF
401  SYAIDRDGVA HQPKKYAKQK EFKDATDNIF HSDYSVSKAL
441  KTVMLKDKSY DLIDEKDFPD KALREAVMAQ VGTRKGDLER
481  FNGTLRLDNP AIQSLEGLNK FKKLAQLDLI GLSRITKLDR
521  SVLPANMKPG KDTLETVLET YKKDNKEEPA TIPPVSLKVS
561  GLTGLKELDL SGFDRETLAG LDAATLTSLE KVDISGNKLD
601  LAPGTENRQI FDTMLSTISN HVGSNEQTVK FDKQKPTGHY
641  PDTYGKTSLR LPVANEKVDL QSQLLFGTVT NQGTLINSEA
681  DYKAYQNHKI AGRSFVDSNY HYNNFKVSYE NYTVKVTDST
721  LGTTTDKTLA TDKEETYKVD FFSPADKTKA VHTAKVIVGD
761  EKTMMVNLAE GATVIGGSAD PVNARKVFDG QLGSETDNIS
801  LGWDSKQSII FKLKEDGLIK HWRFFNDSAR NPETTNKPIQ
841  EASLQIFNIK DYNLDNLLEN PNKFDDEKYW ITVDTYSAQG
881  ERATAFSNTL NNITSKYWRV VFDTKGDRYS SPVVPELQIL
921  GYPLPNADTI MKTVTTAKEL SQQKDKFSQK MLDELKIKEM
961  ALETSLNSKI FDVTAINANA GVLKDCIEKR QLLKK
```

Table 3.2: Amino acid sequence of Endo S protein. Accession No. AF296340.

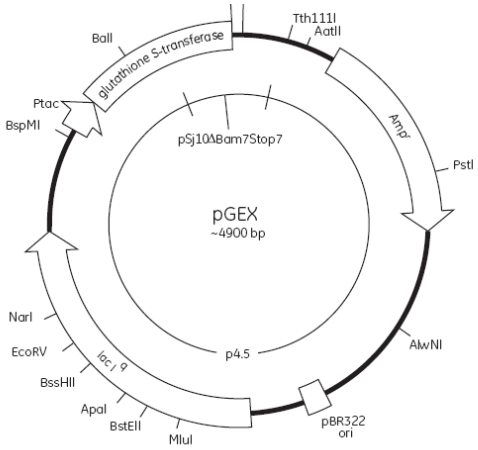


Figure 3.2: Gene map of pGEX-4T, Glutathione S-transferase fusion vector.

The plasmid, which we named pGEX-41-1:*ndoS*, was transformed into *E. coli* strain BL21(DE3) using standard protocols given by the supplier (Novagen).[3] Expression of

recombinant Endo S protein in *E. coli* cultures was induced with 1 mM isopropyl β-D-1 thiogalactopyranoside (ITPG) at 37°C overnight and the resulting cell pellet was harvested from 1 L of induced culture by centrifugation. The cell pellet was lysed by treatment with lysozyme and DNAse I for 1 hour at 4°C followed by sonication to yield the *N*-terminally GST-tagged construct (M_w = 134 kDa). Gene expression, as judged by SDS-PAGE analysis appeared high.

The protein was then purified using a glutathione sepharose high performance column (GSTrap™ HP, GE Healthcare) on an ÄKTA FPLC™ system according to the manufacturer's instructions (Figure 3.3A and 3.3B).

Binding Buffer: 10 mM phosphate buffered saline (PBS) pH 7.4

Elution Buffer: 50 mM Tris-HCl, 10 mM reduced L-glutathione pH 8.0

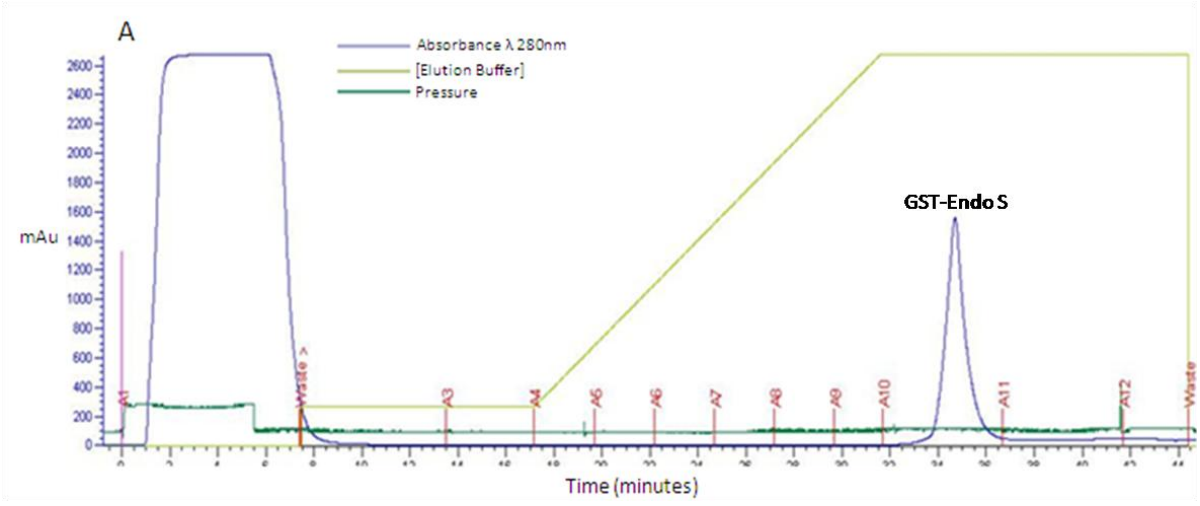
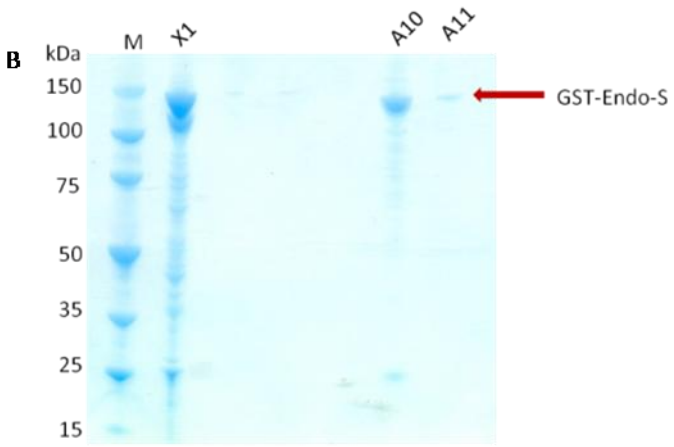


Figure 3.3A: UV trace of GST-Endo S purification by ÄKTA FPLC™.

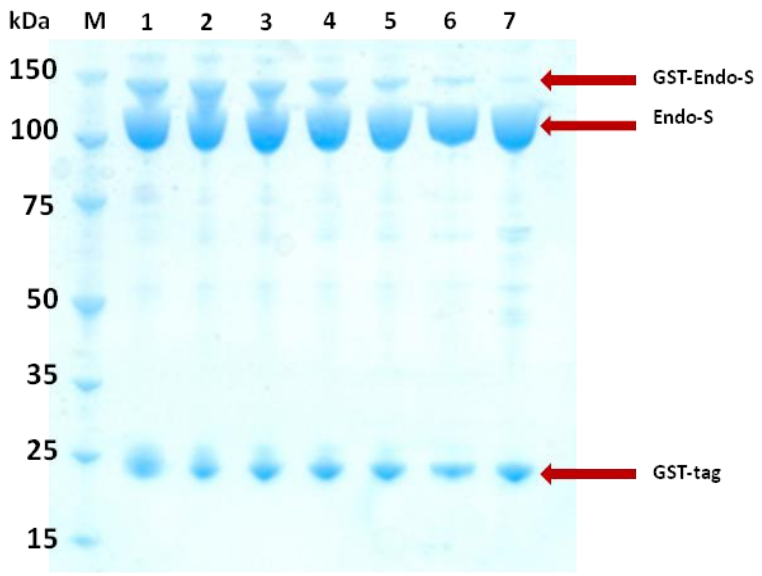


Lane X1: Wash fractions from GSTrap™ purification; Lane A10-A11: Elution fractions containing GST-tagged Endo S.

Figure 3.3B: GSTrap™ purification showing protein in fraction A10 at M_w = 134 kDa.

Fractions containing purified GST-Endo S were pooled and dialysed with PBS (10 mM, pH 7.4) for 4 hours at 4°C before the removal of the GST-tag (M_w = 26 kDa) using thrombin protease (M_w = 36 kDa) (NEB).[4] One unit of thrombin will cleave >90% of 100 µg of GST-Endo S protein in PBS buffer (10 mM, pH 7.4) at 37 ^{o}C for 16 hours. In general, 1 µL of thrombin = one unit. One unit is approximately equivalent to 0.2 National Institute of Health (NIH) units.

For the cleavage reaction, 14 mg/mL of purified GST-Endo S fusion protein was incubated with 140 µL thrombin in 1 mL PBS buffer (10 mM, pH 7.4) at 37 °C for 16 hours. A time course assay was carried out and analysed by SDS-PAGE showing that overnight incubation cleaved the GST-tag (Figure 3.4).



Samples taken from thrombin reaction mixture after certain time periods. Lane 1: 30 minutes; Lane 2: 1 hour; Lane 3: 2 hours; Lane 4: 3 hours; Lane 5: 5 hours; Lane 6: 8 hours; Lane 7: 16 hours.

Figure 3.4: Time course assay of GST-tag cleavage by thrombin protease.

The cleaved fractions were then purified by ÄKTA FPLC™ (Figure 3.5A), using the same method as described above to yield pure Endo S in a typical yield of 10 mg per 1 L culture. A gel band corresponding to roughly M_w = 108 kDa was observed by SDS-PAGE (Figure 3.5B).

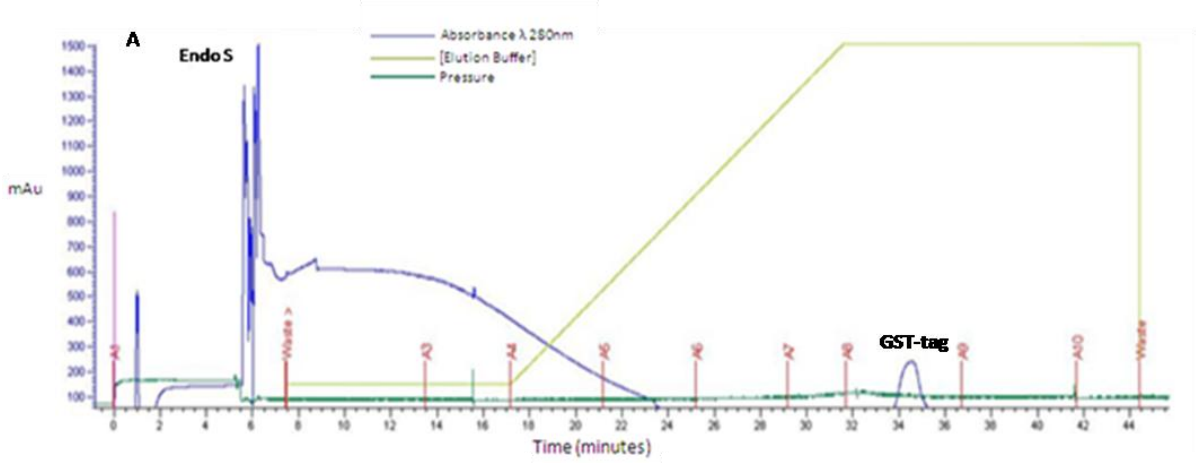
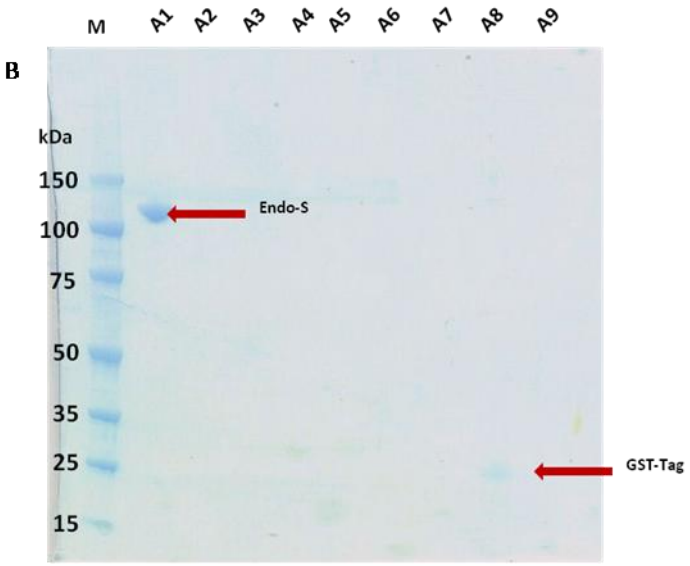


Figure 3.5A: UV trace of Endo S purification by ÄKTA FPLC™.



Lane A1: Wash fractions from GSTrap^{TM} purification containing Endo S; Lane A2-A9: Elution fractions with GST-tag visible in lane A9.

Figure 3.5B: GSTrap^{TM} purification showing protein in fraction A1 at roughly M_w = 108 kDa.

The molecular weight of Endo S was confirmed by matrix-assisted laser desorption ionization (MALDI) analysis (Figure 3.6).

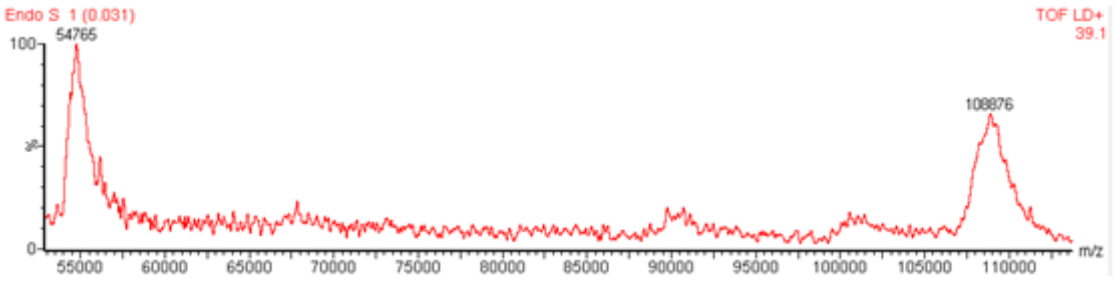


Figure 3.6: MALDI-MS spectrum of purified Endo S with total mass [M_w = 108 kDa] and half mass [M_w = 54 kDa] peak shown.

3.2. Probing the activity of Endo S

To test the activity of the Endo S protein, commercially available human IgG was incubated in PBS buffer (10 mM, pH 7.4, 20 µL) with purified Endo S (1 µg, 0.46 µmol) at 37°C for 1 hour at differing molar ratios of IgG:Endo S (717:1, 360:1, 145:1, 72:1). The reaction mixtures were analysed by both MALDI-MS (Figure 3.7) and SDS-PAGE (Figure 3.8).

Under the reducing conditions used for SDS-PAGE analysis (presence of reducing agent, such as dithiothreitol (DTT)) all IgGs consist predominantly of two bands comprising heavy (M_w = 50 kDa) and light chains (M_w = 22 kDa), the glycan that Endo S hydrolyses being present on the heavy chain (Figure 3.9). The MALDI-MS show a reduction in mass *(*observed glycan release ~2 kDa, calculated glycan release = 2.2 kDa) of the IgG molecule even at the lowest concentration of Endo S.

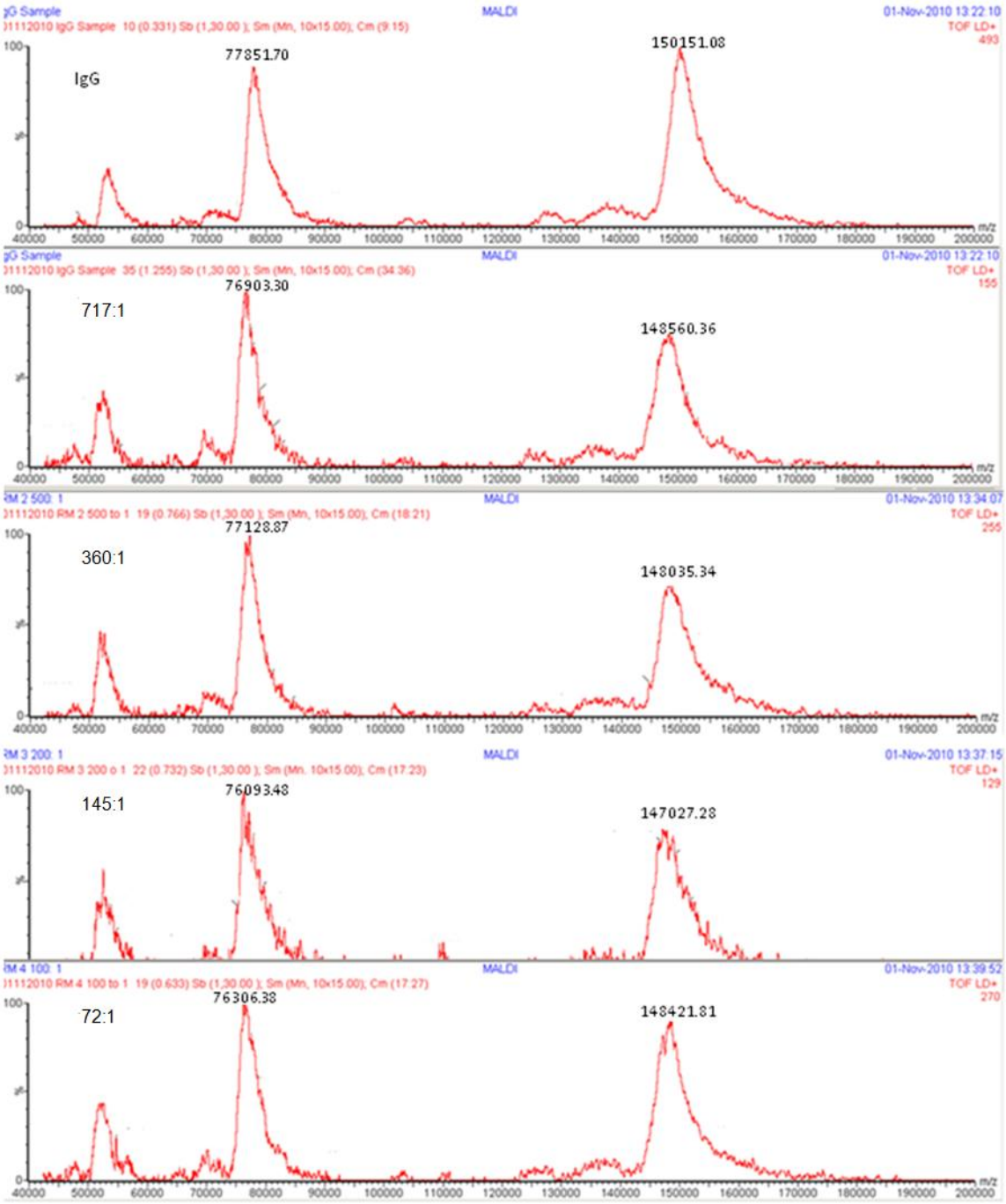
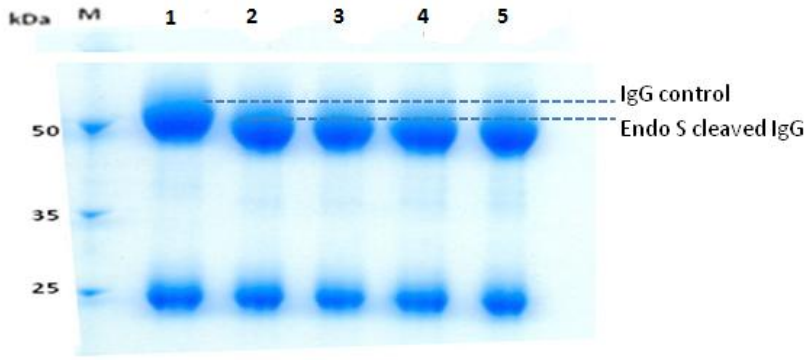


Figure 3.7: Endo S cleaves IgG glycans at molar ratio of IgG:Endo S 717:1. MALDI-MS analysis of Endo S catalysed hydrolysis of human IgG at varying molar ratios of IgG:Endo S showing the hydrolysed IgG at M_w = 148 kDa.



Lane 1: IgG under reducing conditions; Lane 2: IgG:Endo S (717:1); Lane 3: IgG:Endo S (360:1); Lane 4: IgG:Endo S (145:1); Lane 5: IgG:Endo S (72:1).

Figure 3.8: Apparent molecular weight shifts of Endo S treated IgG. Enzymatic reactions of IgG with Endo S at 37°C for 1 hour at varying molar ratios of IgG:Endo S.

This glycan-hydrolysing activity of immunomodulating Endo S has been attributed to an essential glutamic acid residue found at position 235[5] but to date, no X-ray crystal structure of wild type Endo S has been elucidated. Work is currently on-going in the BGD group to achieve this goal.

3.3. Quantification of the *N*-linked glycan released on IgG by Endo S

In order to quantitatively analyse the *N*-linked glycans that are released from IgG by Endo S incubation, MALDI-MS was used.[6] The enzymatic reactions were all carried out using at least 100 µg of IgG to ensure that there would be enough glycans released for analysis.

Oligosaccharides were released from IgG (100 µg, 33.3 µmol) by in solution Endo S (1 µg, 0.46 µmol) digestion at 37°C for 1 hour. The resulting glycan samples were filtered through a MultiScreen® 96-Well Filtration Plate (Millipore), dried by SpeedVac and treated with a 1% AcOH solution at 90°C for 1 hour to remove terminal sialic acid residues. Finally, samples were cleaned using a Nafion membrane (Sigma Aldrich). Mass spectra were recorded using a Shimazu AXIMA TOF^2 MALDI TOF/TOF mass spectrometer (Kratos Analytical) as previously described.[6] All glycans are calculated as the sodium adduct – $[M + Na]^+$. Samples ran in positive reflective mode. The structure of a glycan that can be removed by Endo S is shown below (Figure 3.9).

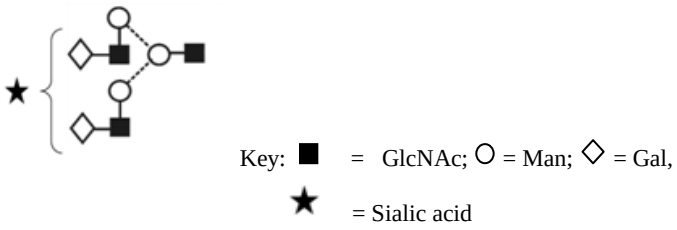


Figure 3.9: Structure of a glycan removed by Endo S from IgG, leaving a core fucose and GlcNAc intact at Asn297.

The MALDI-MS spectra obtained are shown below in Figure 3.10A and Figure 3.10B.

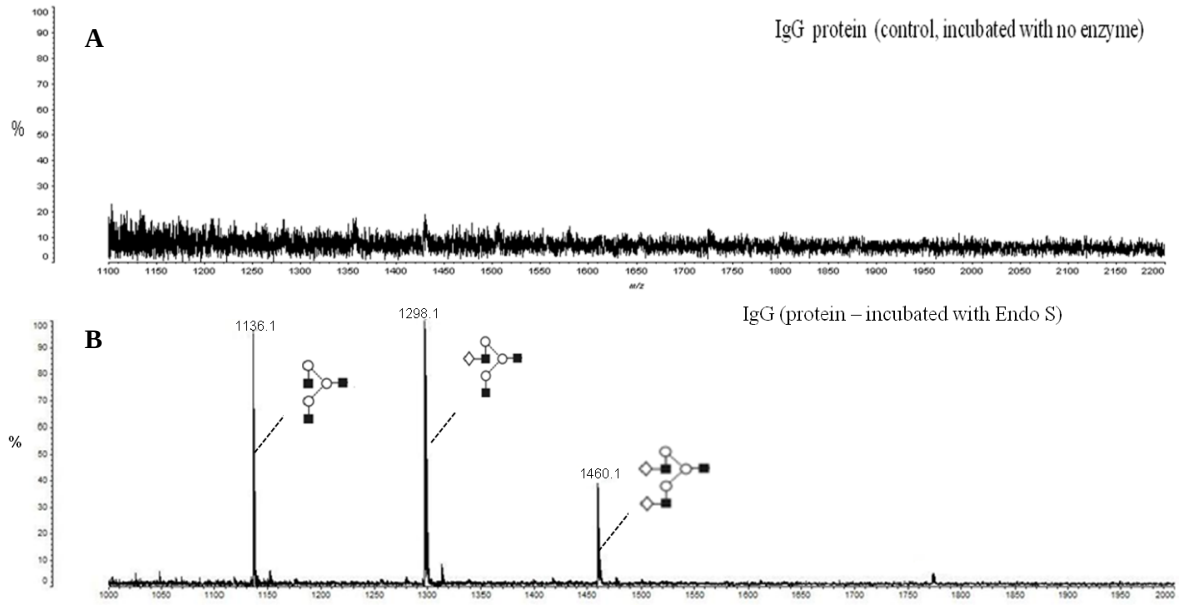
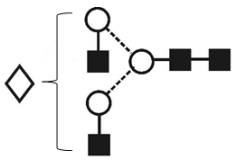


Figure 3.10: Hydrolysis of IgG glycans by Endo S. MALDI-MS of the released glycans obtained when IgG. **A** Control of IgG incubated without Endo S. **B** IgG incubated with Endo S for 1 hour.

N.B. Whilst the structures drawn show a particular isomer of the glycan removed, it should be noted that the exact isomer of the glycan removed is unknown. It should be considered more like structure shown below.



Endo S (1 µg, 0.46 µmol) was incubated with IgG (100 µg, 33.3 µmol) in PBS buffer (10 mM, pH 7.4, 20 µL) at 37°C and a time course assay was performed. Samples were taken over a period of twenty four hours at 30 minutes, 1 hour, 2 hours, 4 hours and 24 hours to see the rate of glycan cleavage. From the results obtained (not shown), we

can conclude that Endo S cleaves the glycans just after being incubated at 37 °C for 30 minutes.

3.4. Glycan cleavage enzymatic reactions on IgG using other Endo enzymes

Next, we wanted to compare the ability of Endo S to hydrolyse the complex glycan with other endoglycosidase enzymes in the same class [ENGases, E.C. 3.2.1.96]; Endo A from *Arthrobacter Protophormaie*[7] (M_w = 80 kDa, BGD lab), Endo H from *Streptomyces plicatus*[8]*(*M_w = 29 kDa, NEB*)* which both belong to Glycoside Hydrolase (GH) family 85 and with deglycosidase enzyme [E.C. 3.5.1.52] peptide: *N*-Glycosidase F from *Flavobacterium meningosepticum*[9](M_w = 36 kDa, NEB).

Endo A is a recombinant glycosidase enzyme that cleaves within the chitobiose core of high mannose oligosaccharides from *N*-linked glycoproteins (Figure 3.11).[10] It is cloned from *Arthrobacter protophormaie* and overexpressed in *E. coli.*[11]

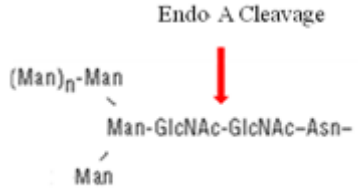


Figure 3.11: Site of Endo A cleavage.

Endo H (NEB) is a recombinant glycosidase enzyme that cleaves the chitobiose core of high mannose and some hybrid oligosaccharides from *N*-linked glycoproteins (Figure 3.12).[10] It is cloned from *Streptomyces plicatus*[8] and overexpressed in *E. coli.*[12]

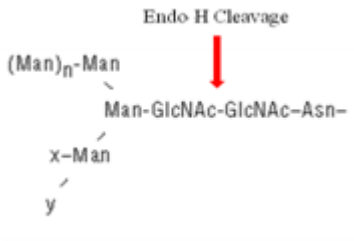


Figure 3.12: Site of Endo H cleavage.

Peptide: *N*-Glycosidase F (PNGase F, NEB) is an amidase which cleaves between the innermost GlcNAc and asparagines residues of high mannose, hybrid and complex type oligosaccharides from *N*-linked glycoproteins and is therefore referred to as a deglycosidase (Figure 3.13).[10, 13] It is purified from *Flavobacterium meningosepticum.*[14]

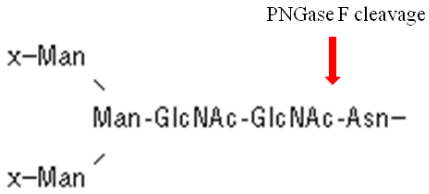
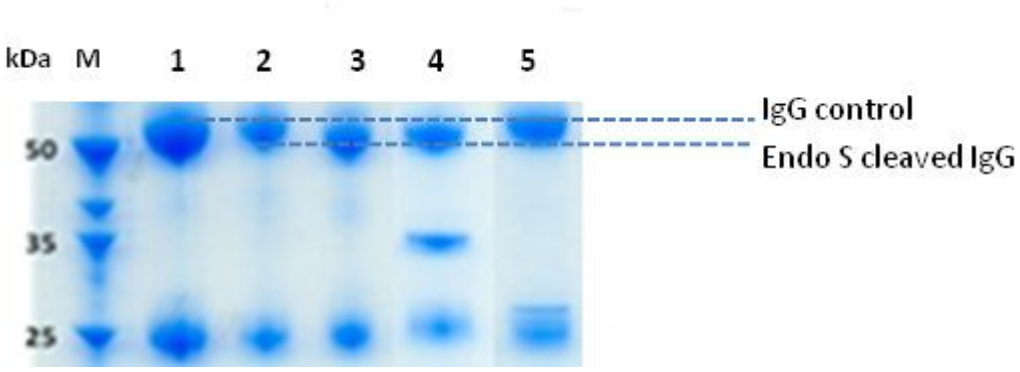


Figure 3.13: Site of PNGase F cleavage.

Human IgG (100 µg per reaction) was incubated in PBS buffer (10 mM, pH 7.4, 20 µL) with the corresponding enzyme (IgG:Enzyme molar ratio 100:1) according to the manufacturer's instructions at 37°C for 1 hour. The reaction mixtures were analysed by SDS-PAGE (Figure 3.14).



Enzymatic reactions with IgG at 37°C for 1 hour. Lane 1: IgG; Lane 2: IgG:Endo A (100:1); Lane 3: IgG:Endo S (100:1) Lane 4: IgG:PNGase F (100:1) with PNGase F protein at M_w 36 kDa; Lane 5: IgG:Endo H (100:1) with Endo H protein at M_w 29 kDa.

Figure 3.14: Enzymatic reactions of IgG with other enzymes at 37°C for 1 hour.

The SDS-PAGE shows a qualitative reduction in mass of IgG when incubated with Endo S and PNGase F enzymes which are capable of hydrolysing the complex type glycan present on the heavy chains and no shift of mass on the gel when incubated with Endo A and Endo H.

3.5. Quantitative analysis of Endo A and Endo H glycan release from IgG

Next, we wanted to compare the ability of Endo S to hydrolyse the complex glycan attached to IgG with other Endo enzymes in the same class. IgG (100 µg, 33.3 µmol) in PBS buffer (10 mM, pH 7.4) was incubated with Endo A and Endo H according to the manufacturer's (NEB) instructions at 37°C for 1 hour to see the extent of glycan cleavage from IgG (Figure 3.15A-C). Samples were analysed by same technique as described above. The structure of the glycan released is shown in Figure 3.9.

.

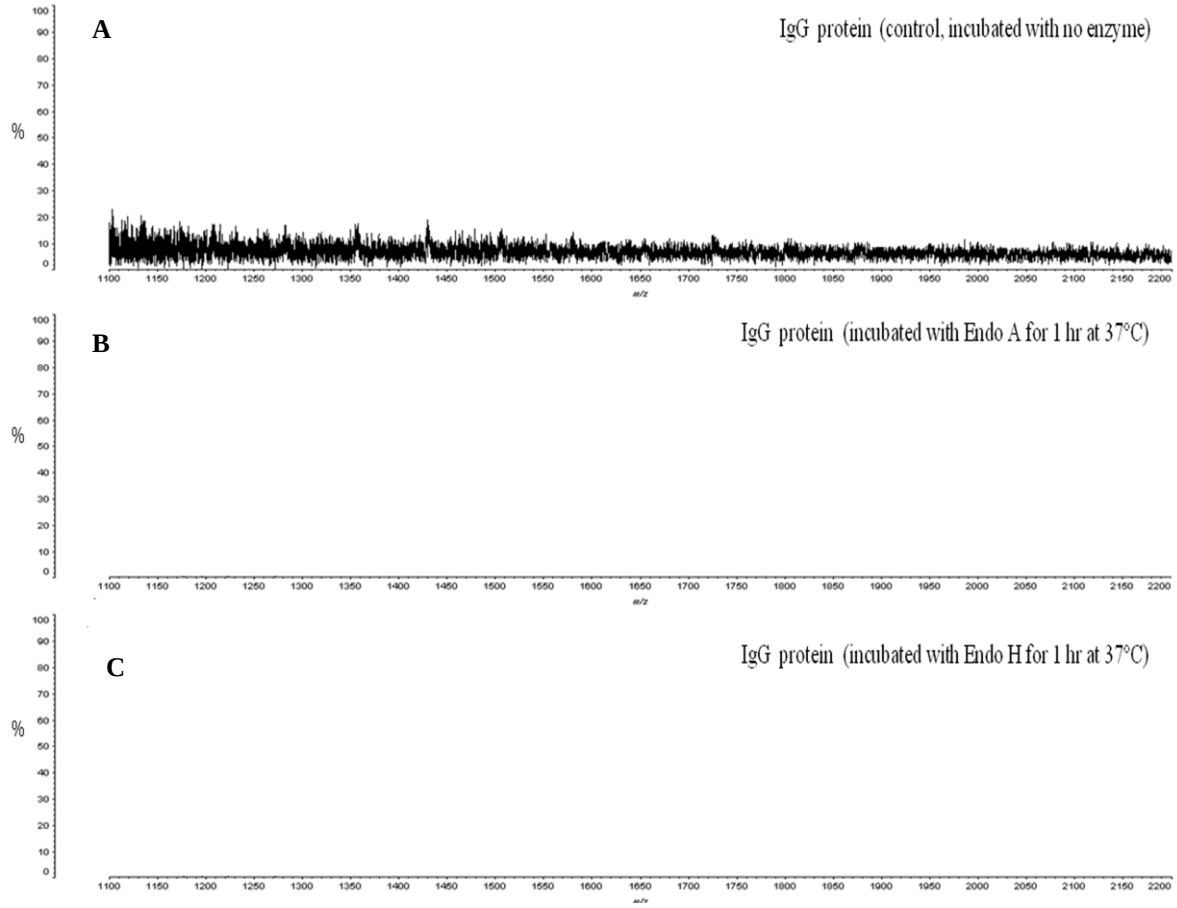
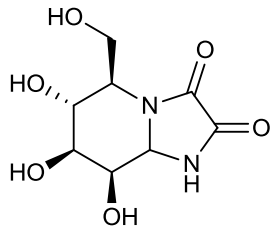


Figure 3.15: Endo A and Endo H digestion with IgG results in no glycans being released. MALDI-MS spectra of glycan release from IgG. **A** Control experiment of IgG with no enzymes incubated. **B** IgG incubated with Endo A. **C** IgG incubated with Endo H.

glycans being detected by MALDI analysis. This observation is as predicted due to both these Endo enzymes not being able to hydrolyse complex-type glycans that are naturally found on IgG glycoproteins. This result had previously been shown by lack of band shift on the SDS-PAGE (Figure 3.14).

3.6. Glycan specificity of WT Endo S

In order to work out the *N*-linked glycan specificity of WT Endo S, experiments were designed to test whether Endo S would cleave various *N*-linked glycans released from recombinant $gp120_{BAL}$ protein – a variant gp120 protein purified from a macrophage-tropic HIV-1 strain, that is called 'bal' . It is known to contain a range of complex and mannose *N*-linked glycans by in-gel PNGase F digestion which are ideal for exploring the specificity of Endo S.[15] The main advantage of the mammalian *N*-linked glycosylation pathway is that it leads to many chemically diverse glycans compared to that of lower eukaryotes. This pathway can be manipulated using inhibitors, which can be readily applied to an expression system. Three different expression systems for recombinant $gp120_{BAL}$ were tested.



Kifunensine

Figure 3.16: Inhibitor of the *N*-linked glycosylation pathway. Kifunensine (Kif) inhibits ER α-mannosidase I and Golgi α-mannosidase IA-C.

1) $gp120_{BAL}$ protein expressed in HEK293T (Human Embryonic Kidney) cells

   - Normal mammalian *N*-linked glycosylation pathway.

2) $gp120_{BAL}$ protein expressed in HEK293T cells with kifunensine (Kif) inhibitor (Figure 3.16.)[16]

- Kifunensine inhibits ER α-mannosidase I and Golgi α-mannosidase IA-C enzyme in mammalian *N*-linked glycan pathway so $gp120_{BAL}$ protein has mainly $Man_9GlcNAc_2$ sugars present with small amounts of Man_{5}-$_{8}GlcNAc_2$.

3) $gp120_{BAL}$ protein expressed in GnT-I HEK293S cells

- This cell line is deficient in GnT-I, therefore expressed glycoproteins contain exclusively $Man_5GlcNAc_2$ glycans with some core fucosylation present.

The mammalian *N*-linked glycosylation pathway is summarised in Figure 3.17 with the point of inhibition in the glycosylation pathway by kifunensine and swainsonine (SWN) (Figure 3.23) highlighted.[15]

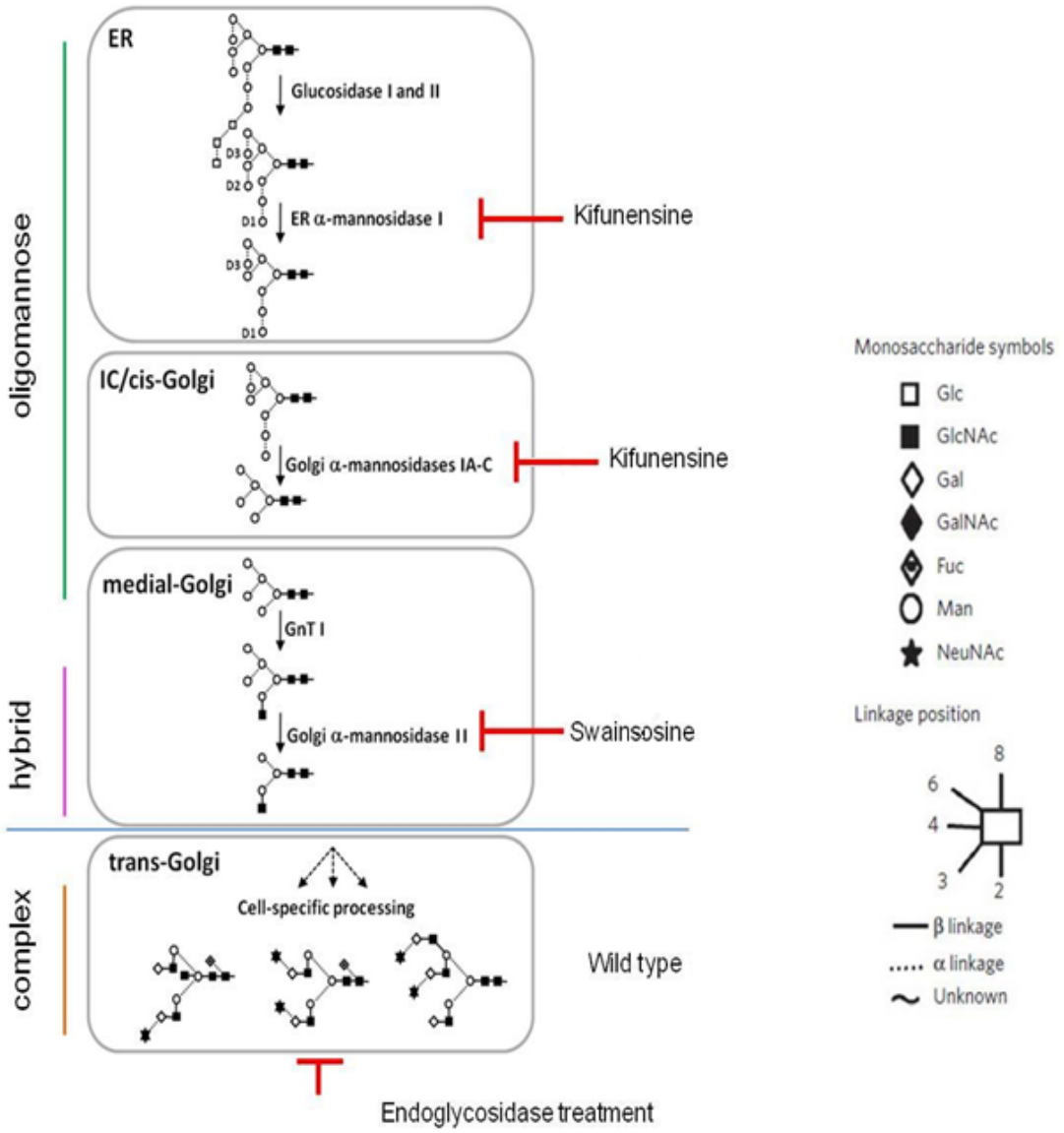


Figure 3.17: Mammalian *N*-linked glycosylation pathway. Reproduced from Scanlan *et. al.*[15]

In all of the recombinant $gp120_{BAL}$ experiments, three conditions were run:

1) A control experiment of the free *N*-linked glycans released from $gp120_{BAL}$ protein by in-gel PNGase F digestion.

2) The free glycans incubated with Endo S for 2 hours at 37°C

3) The free glycans incubated with Endo H for 2 hours at 37°C

All reaction samples were filtered through a MultiScreen® 96-Well Filtration Plate (Millipore), dried by SpeedVac and cleaned using a Nafion membrane (Sigma Aldrich). Mass spectra were recorded using a Shimazu AXIMA TOF^2 MALDI TOF/TOF mass spectrometer (Kratos Analytical). Recombinant $gp120_{BAL}$ glycans expressed in HEK293T cells were desialyated with 1.0% AcOH at 80°C for 1 hour prior to analysis.

The resultant MALDI-MS spectra of the enzymatic glycan digestion reactions using glycans released from recombinant $gp120_{BAL}$ expressed in HEK293T cells with kifunensine and glycans expressed in HEK293S cells are shown in Figure 3.18A-F.

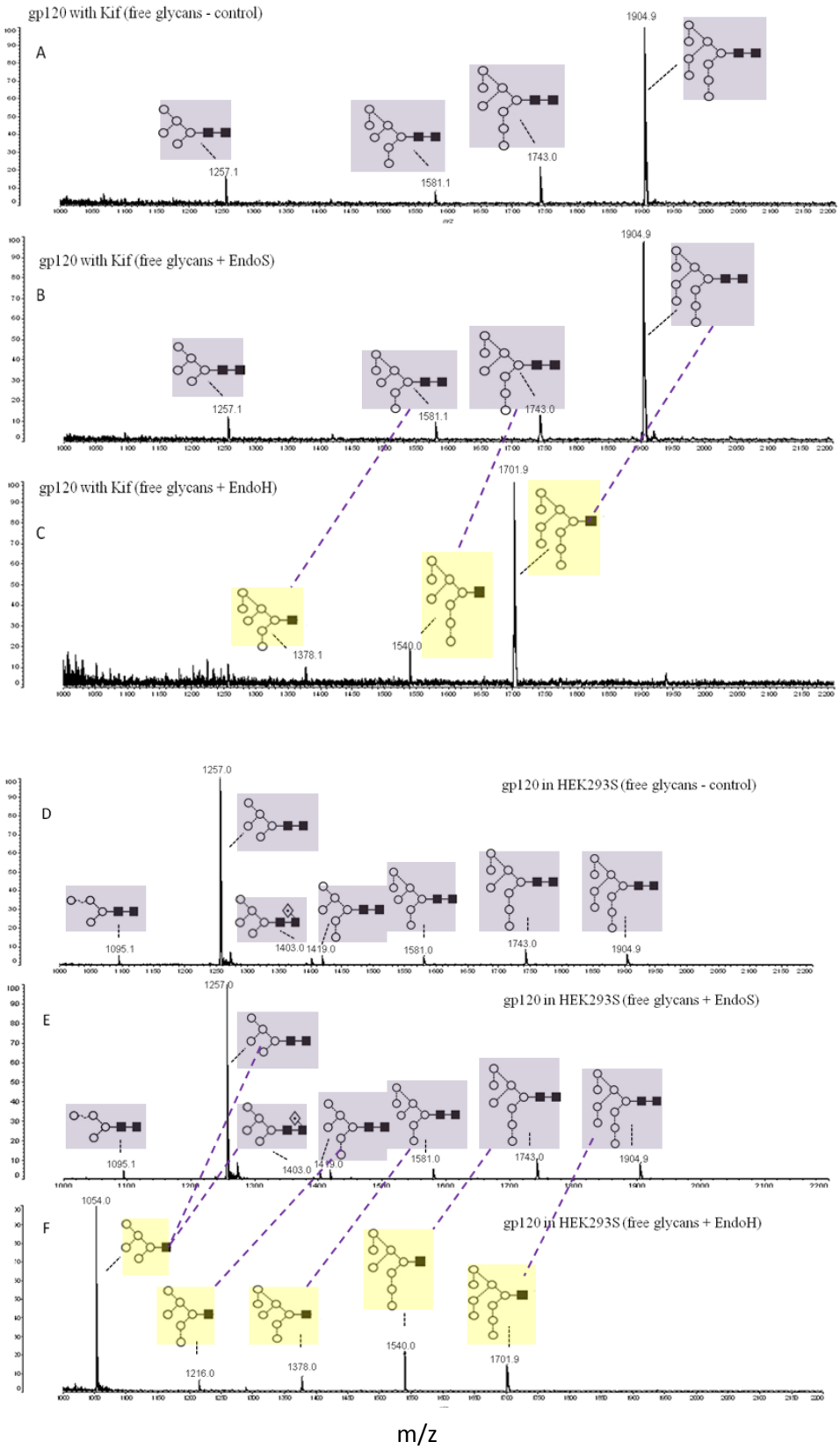


Figure 3.18: Oligo-mannose and high-mannose *N*-linked glycans are resistant to Endo S digestion. MALDI-MS spectra of enzymatic glycan reactions. **A** Glycans released from $gp120_{BAL}$ expressed in HEK293T cells with kifunensine. **B** Glycans

incubated with Endo S. **C** Glycans incubated with Endo H. **D** Glycans released from $gp120_{BAL}$ expressed in HEK293S cells. **E** Glycans incubated with Endo S. **F** Glycans incubated with Endo H.

When Endo S is incubated with free glycans released from recombinant $gp120_{BAL}$ in HEK293T cells incubated with kifunensine and $gp120_{BAL}$ expressed in HEK293S cells, we can see there is no change in the glycan profile observed in comparison to the control digests. However, when compared to the results obtained from digestion with Endo H, we can see a mass shift (203 $gmol^{-1}$) for all of the high-mannose and oligo-mannose *N*-linked glycan peaks, corresponding to the loss of a single GlcNAc sugar within the chitobiose core. This data suggests that Endo S is not specific for high-mannose or oligo-mannose *N*-linked glycans.

The resultant MALDI-MS spectra of the enzymatic glycan digestion reactions using glycans released from recombinant $gp120_{BAL}$ expressed in HEK293T cells are shown in Figure 3.19A-C.

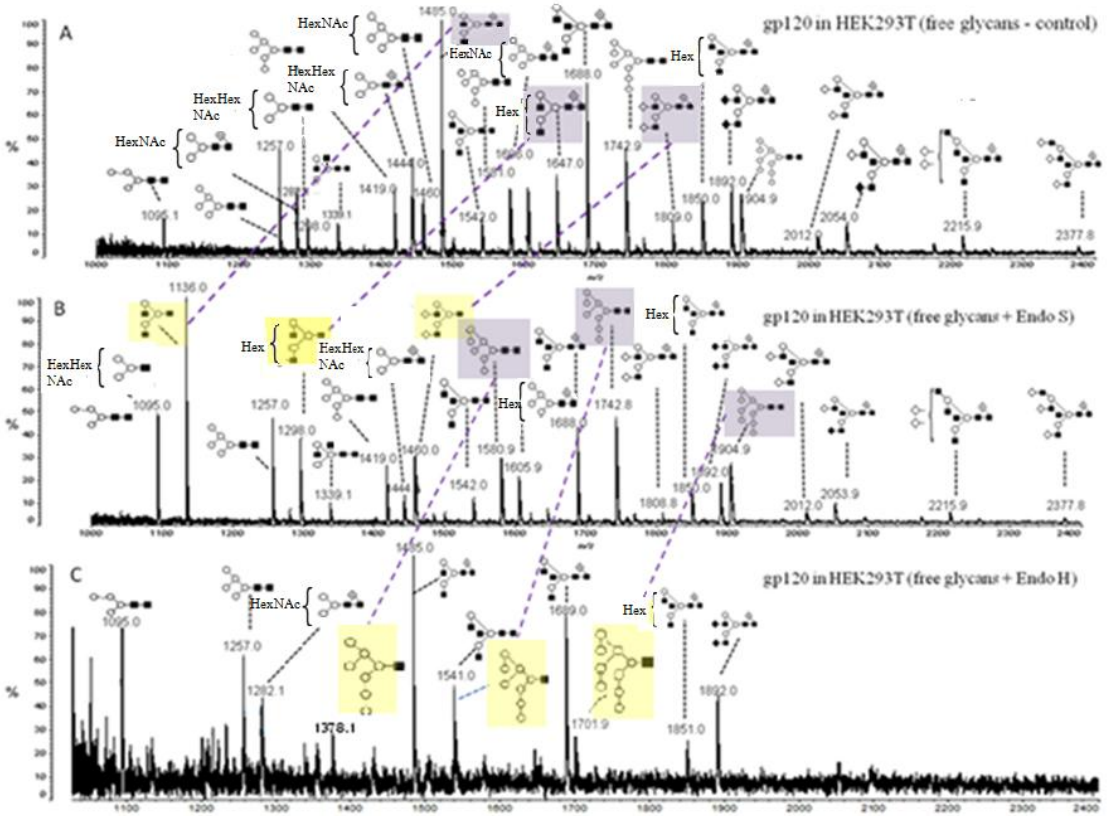


Figure 3.19: Complex biantennary *N*-linked glycans are susceptible to Endo S digestion. MALDI-MS spectra of enzymatic glycan reactions. **A** Glycans released from $gp120_{BAL}$ expressed in HEK293T cells. **B** Glycans (A) incubated with Endo S. **C** Glycans (A) incubated with Endo H.

Conversely, when Endo S was incubated with *N*-linked glycans released from recombinant $gp120_{BAL}$ expressed in HEK293T cells which undergo full mammalian *N*-linked glycosylation, we observe a reduction in percentage abundance of complex type glycans containing a core fucose. This observed mass reduction (349 $gmol^{-1}$) corresponds to the removal of the core fucosylated GlcNAc sugar in the chitobiose core. Lack of digestion at m/z 1339.1 $gmol^{-1}$ (Figure 3.19B) suggests that Endo S might not be capable of digesting bisecting GlcNAc sugars but further work would be need to prove this hypothesis. Complete enzymatic digestion is observed at m/z 1485.0 $gmol^{-1}$ (Figure 3.19A). When these results are compared to digestion with Endo H, we can see that only

the high-mannose type glycans are hydrolysed. This data shows that Endo S is specific for complex type glycans that have a core fucose group in this glycan set and that the selectively observed is reciprocal to that of Endo H.

The observation that high-mannose and oligo-mannose *N*-linked glycans are resistant to Endo S mediated cleavage led us to further investigate whether or not this result is due to the lack of core fucosylation.

3.7. Expression and purification of WT Fc and SWN Fc glycoproteins

Endo S has evolved to be highly specific for the native complex type *N*-linked glycans found on the Fc fragment of IgG.[1b] This is consistent with the observed specificity against various *N*-linked $gp120_{BAL}$ glycans. In order to test whether or not the lack of core fucosylation affects the glycan-hydrolysing ability of Endo S wild type Fc glycoprotein was expressed and purified.

The plasmid, which we named pHLsec:Fc, was successfully transformed into *E. coli* strain XL1-Blue using standard protocols given by the supplier (Agilent).[3] Expression of recombinant Fc protein in mammalian HEK293T cells was achieved by transfection of endotoxin free DNA along with polyethyleneimine (PEI) transfection reagent. DNA:PEI ratio of 1:1.5 was used and a total of 125 µg of Fc DNA was added to 90% confluent HEK293T cells grown in T225 cm^2 tissue culture flasks. Transfected cells were incubated at 37°C and 5% CO_2 for 5-10 days. Supernatant was collected to yield the His_6-tagged construct (M_w = 57 kDa). The protein was then purified using Ni-NTA sepharose beads (GE Healthcare) to give a typical yield of 54 mg/mL. Gene expression, as judged by SDS-PAGE analysis appeared high (Figure 3.20A).

Binding Buffer: 10 mM phosphate buffered saline (PBS), 10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.4

Elution Buffer: 250 mM imidazole, pH 8.0

The expression and purification method for wild type (WT) Fc protein is summarised in Figure 3.20B.

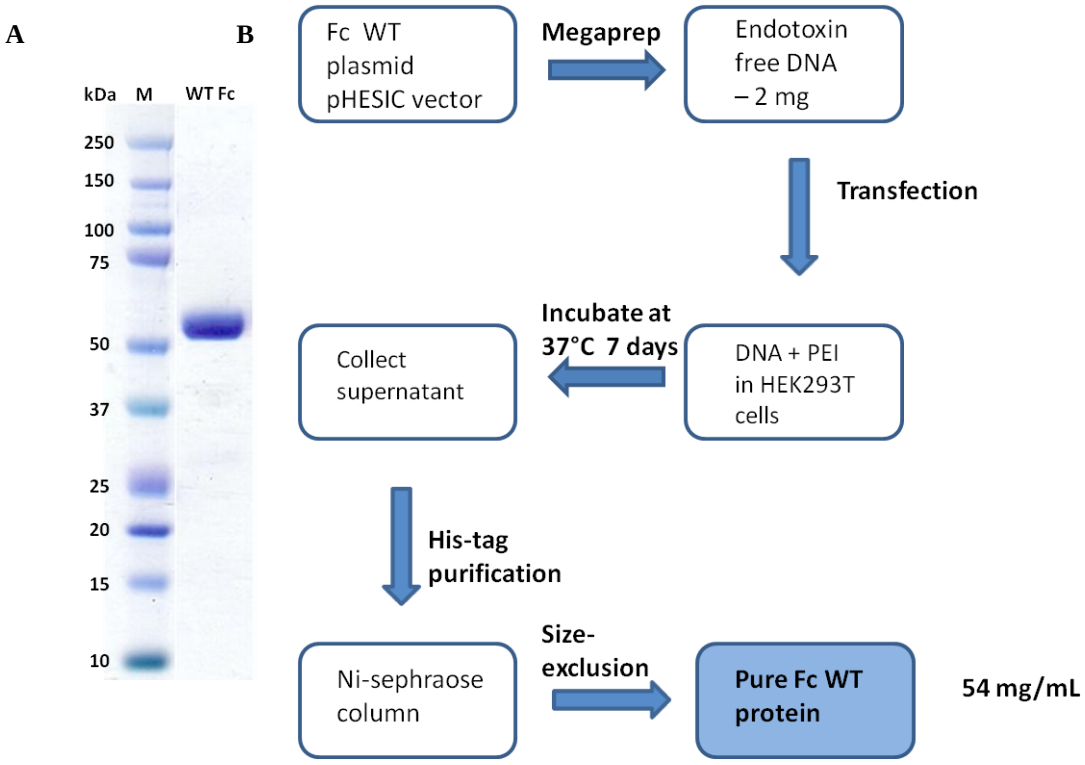


Figure 3.20: A Wild type Fc glycoprotein M_w 57 kDa. **B** Flow diagram of wild type Fc expression and purification.

The resulting Fc protein is predicted to have 230 residues, an average M_w of 57,102 Da and a pI of 5.07. The sequenced nucleotide (Table 3.3.) and translated protein sequence (Table 3.4.) are shown below.[17,18]

DNA sequence of Fc gene (690 Nucleotides):

```
1      AAATCTTGTG  ACAAAACTCA  CACATGCCCA  CCGTGCCCAG  CACCTGAACT
51     CCTGGGGGGA  CCGTCAGTCT  TCCTCTTCCC  CCCAAAACCC  AAGGACACCC
101    TCATGATCTC  CCGGACCCCT  GAGGTCACAT  GCGTGGTGGT  GGACGTGAGC
```

151	CACGAAGACC	CTGAGGTCAA	GTTCAACTGG	TACGTGGACG	GCGTGGAGGT
201	GCATAATGCC	AAGACAAAGC	CGCGGGAGGA	GCAGTACAAC	AGCACGTACC
251	GTGTGGTCAG	CGTCCTCACC	GTCCTGCACC	AGGACTGGCT	GAATGGCAAG
301	GAGTACAAGT	GCAAGGTCTC	CAACAAAGCC	CTCCCAGCCC	CCATCGAGAA
351	AACCATCTCC	AAAGCCAAAG	GGCAGCCCCG	AGAACCACAG	GTGTACACCC
401	TGCCCCCATC	CCGGGATGAG	CTGACCAAGA	ACCAGGTCAG	CCTGACCTGC
451	CTGGTCAAAG	GCTTCTATCC	CAGCGACATC	GCCGTGGAGT	GGGAGAGCAA
501	TGGGCAGCCG	GAGAACAACT	ACAAGGCCAC	GCCTCCCGTG	CTGGACTCCG
551	ACGGCTCCTT	CTTCCTCTAC	AGCAAGCTCA	CCGTGGACAA	GAGCAGGTGG
601	CAGCAGGGGA	ACGTCTTCTC	ATGCTCCGTG	ATGCATGAGG	CTCTGCACAA
651	CCACTACACG	CAGAAGAGCC	TCTCCCTGTC	TCCGGGTAAA	

Table 3.3: Nucleotide sequence of Fc gene.[18]

Amino acid sequence of Fc protein (230 amino acids):

1	KSCDKTHTCP	PCPAPELLGG	PSVFLFPPKP	KDTLMISRTP	EVTCVVVDVS
51	HEDPEVKFNW	YVDGVEVHNA	KTKPREEQYN	STYRVVSVLT	VLHQDWLNGK
101	EYKCKVSNKA	LPAPIEKTIS	KAKGQPREPQ	VYTLPPSRDE	LTKNQVSLTC
151	LVKGFYPSDI	AVEWESNGQP	ENNYKATPPV	LDSDGSFFLY	SKLTVDKSRW
201	QQGNVFSCSV	MHEALHNHYT	QKSLSLSPGK		

Table 3.4: Amino acid sequence of Fc protein.[18]

Wild type Fc *N*-linked glycans were released by in-gel PNGase F digestion as described before. For comparison, defucosylated Fc *N*-linked glycans were obtained by incubating wild type Fc glycoprotein with α-L-fucosidase (Sigma-Aldrich, CAS 9037-65-4) following the supplier's instructions. This fucosidase has broad substrate specificity cleaving α-1→(2,3,4,6) linked fucose from *N*- and *O*-linked glycans, with the highest efficiency on the α-1→6 linked fucose on the trimannosyl core of *N*-linked glycans. The resultant MALDI-MS spectra of the Endo S glycan digestion reactions using glycans released from wild type Fc glycoprotein and wild type Fc glycoprotein treated with α-L-fucosidase are shown in Figure 3.21A-D.

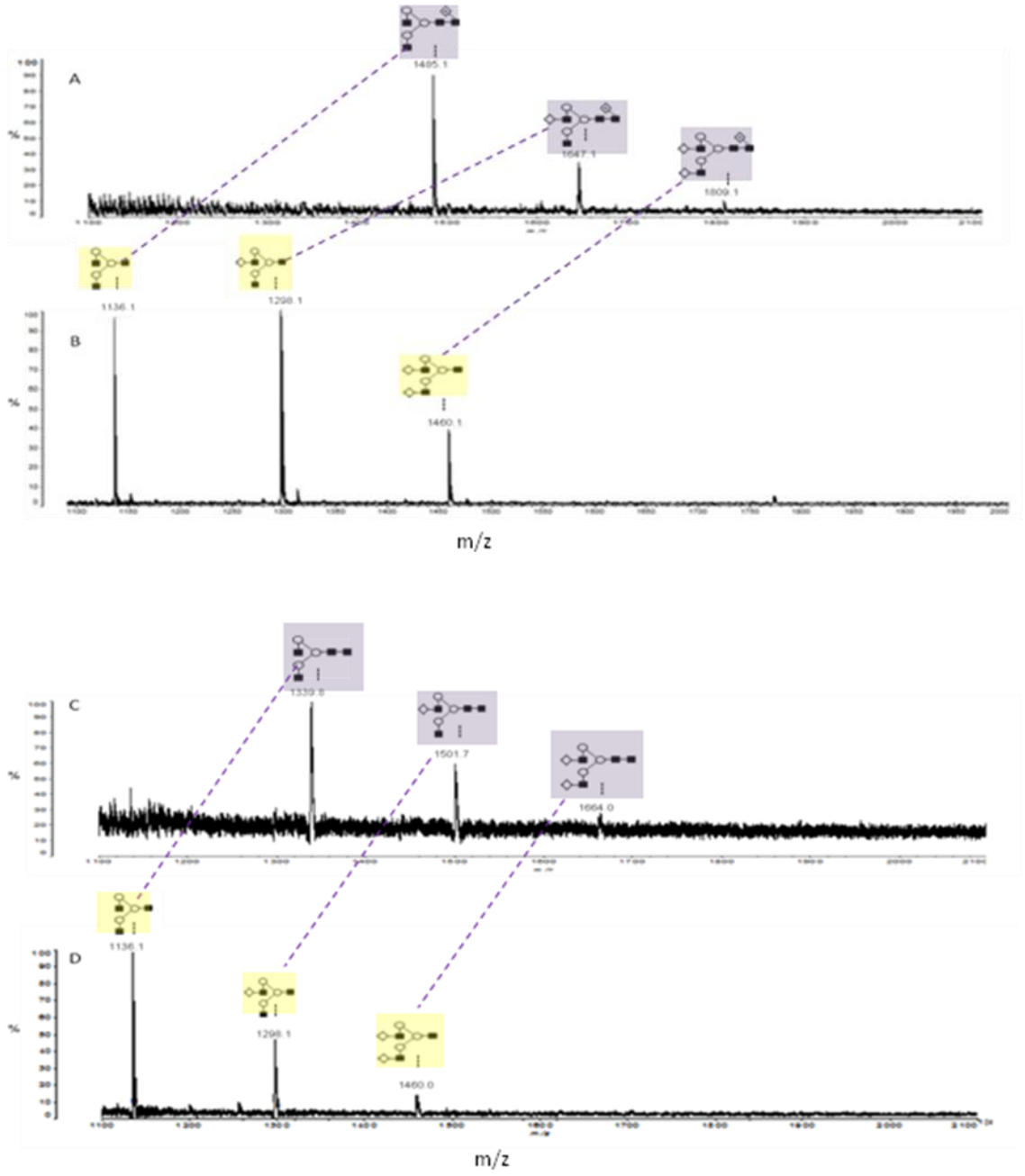


Figure 3.21: Core fucosylation does not affect Endo S digestion. MALDI-MS spectra of enzymatic glycan reactions. **A** Glycans released from wild type Fc expressed in HEK293T cells. **B** Glycans (A) incubated with Endo S. **C** Glycans released from wild type Fc expressed in HEK293T cells and treated with α-L-fucosidase. **D** Glycans (C) incubated with Endo S.

Incubation of wild type Fc glycans and defucosylated wild type Fc glycans individually with Endo S for one hour at 37°C resulted in a mass shift corresponding to the removal of the core fucosylated GlcNAc sugar (349 $gmol^{-1}$) in the wild type Fc glycan profile (Figure 3.21B) and a mass shift corresponding to the removal of the innermost GlcNAc sugar (203 $gmol^{-1}$) in the defucosylated wild type Fc glycan profile (Figure 3.21D). We then proceeded to digest a mixture of wild type and defucosylated Fc *N*-linked glycans with Endo S as before.

The MALDI-MS spectra for the mixed enzymatic digestion are shown below (Figure 3.22A-B).

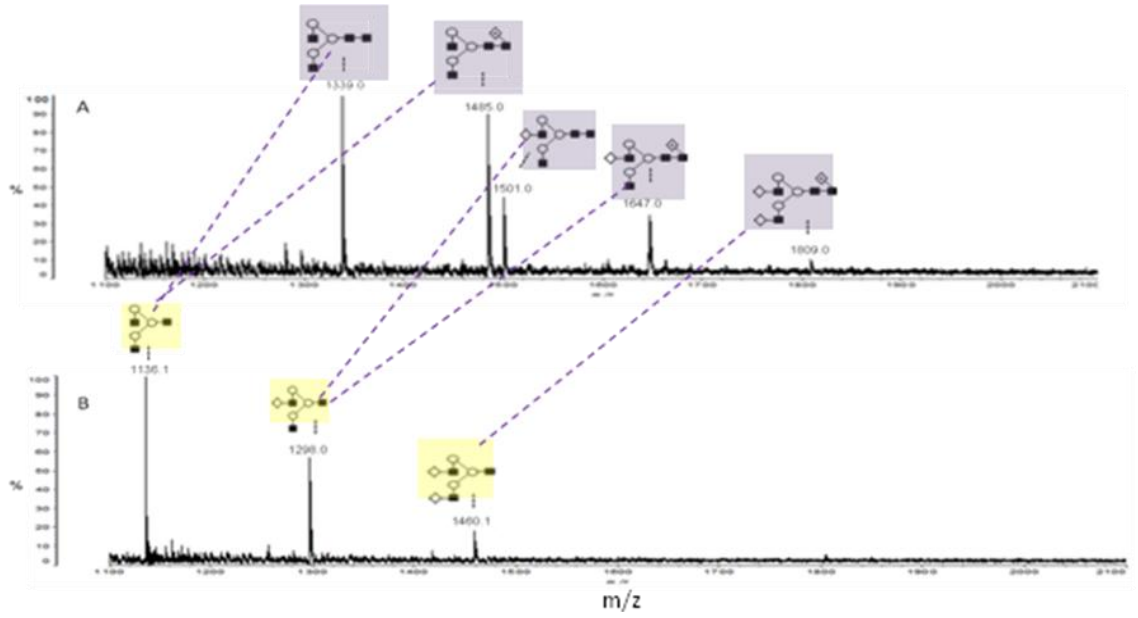


Figure 3.22: Core fucosylation does not affect Endo S digestion. MALDI-MS spectra of mixed enzymatic glycan reactions. **A** Glycans released from wild type and defucosylated Fc glycoprotein. **B** Glycans (A) incubated with Endo S.

We again observed a mass shift corresponding to the removal of the core fucosylated GlcNAc sugar (349 $gmol^{-1}$) and a mass shift corresponding to the removal of the innermost GlcNAc sugar (203 $gmol^{-1}$) in the defucosylated wild type Fc glycan

profile (Figure 3.22B). These results indicate that Endo S is able to cleave core fucosylated and defucosylated complex type *N*-linked glycans with similar efficiency.

As well as expressing the wild type Fc glycoprotein used for the specificity experiments, we wanted to obtain unnatural hydrid *N*-linked glycans to further probe the specificity of Endo S against this glycan set. These glycans were obtained by the manipulation of the mammalian *N*-linked glycosylation pathway using the indolizine alkaloid, swainsonine (SWN) (Figure 3.23).[19] Swainsonine inhibits Golgi α-mannosidase II inducing the biosynthesis of hybrid-type *N*-linked glycans (Figure 3.17).

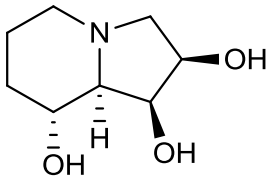


Figure 3.23: Golgi α-mannosidase II inhibitor, swainsonine.

To obtain SWN Fc glycoprotein, 10 μM swainosonine was added to the HEK293T cells during the transfection stage of protein expression (Figure 3.24B). SWN Fc was purified in a typical yield of 38 mg/mL. Gene expression, as judged by SDS-PAGE analysis appeared high (Figure 3.24A)

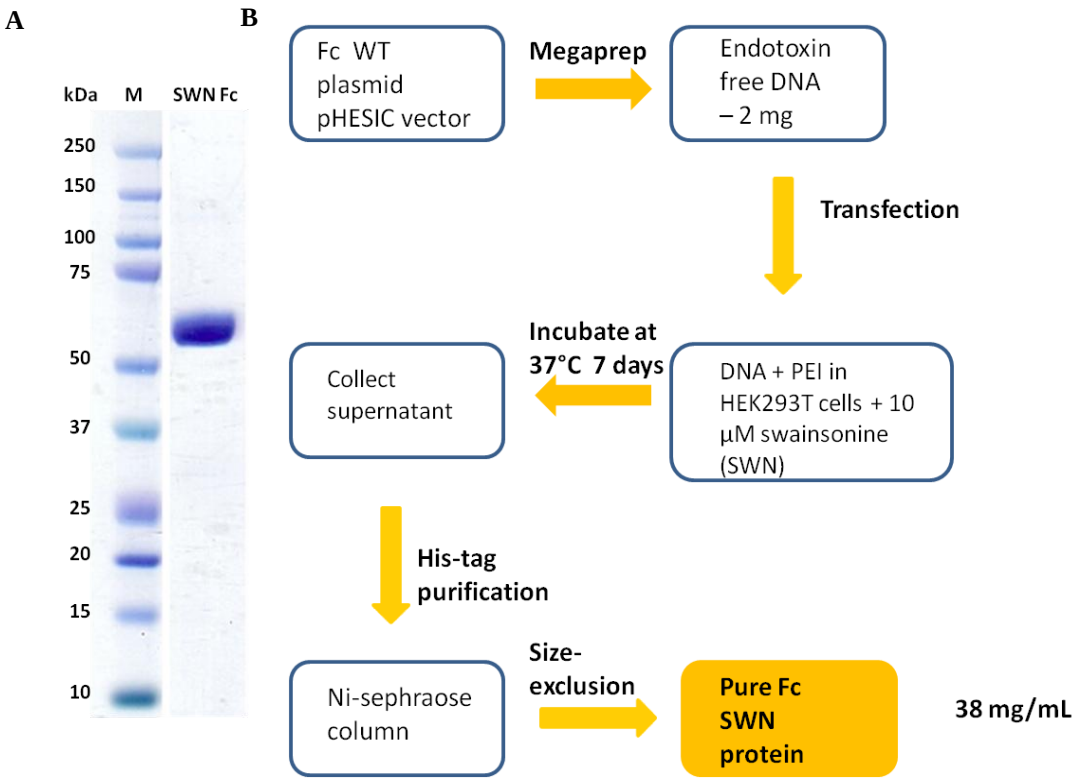
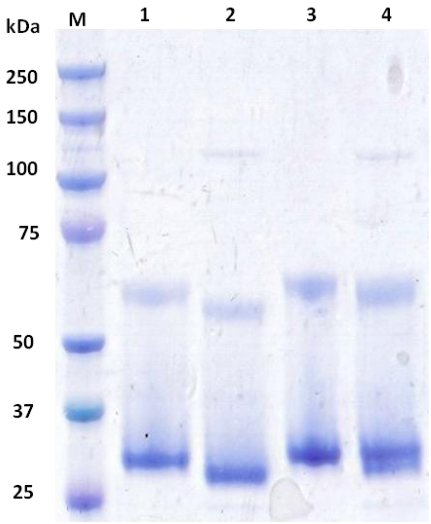


Figure 3.24: A SWN Fc glycoprotein M_w 57 kDa. **B** Flow diagram of SWN Fc expression and purification.

Fc WT glycoprotein and Fc SWN glycoprotein were incubated with Endo S (100:1 molar ratio) at 37°C for two hours and the reaction mixture was analysed by SDS-PAGE (Figure 3.25). Endo S treated WT and SWN Fc glycoproteins show a decrease in apparent molecular weight on SDS-PAGE, further confirming the glycan-hydrolysing activity of Endo S.



Enzymatic reactions of Endo S with WT Fc and SWN Fc at 37°C for 2 hours. Lane 1:WT Fc; Lane 2: Endo S treated WT Fc; Lane 3: SWN Fc; Lane 4:Endo S treated SWN Fc. Both reduced and non-reduced bands can be observed in the gel due to incomplete denaturing of Fc glycoprotein.

Figure 3.25: Endo S digestion of WT and SWN Fc glycoproteins results in apparent decrease in mass. Enzymatic reactions of WT Fc and SWN Fc with Endo S.

Endo S digestions of PNGase F released hydrid *N*-linked glycans from SWN Fc using MALDI analysis showed a complete absence of the typical mass peaks corresponding to hydrid core fucosylated *N*-linked Fc glycans. However, mass peaks for the resultant Endo S hydrolysed glycans lacking the core fucosylated GlcNAc were not observed. Although these results are inconclusive, further studies are ongoing to elucidate the specificity of Endo S against hydrid type *N*-linked glycans.

3.8. Conclusion

In this study we sought to determine the activity of Endo S against a range of natural human *N*-linked glycans and showed that it readily hydrolysed core-fucosylated bi-antennary complex-type glycans typically found on wild type mammalian antibody Fc. Beyond this we found that Endo S is active against a surprisingly restricted range of complex-type *N*-linked glycans and shows no detectable activity against oligomannose-type glycans. The catalytic specificity of Endo S is therefore broadly reciprocal to the well-studied Endo H (Table 3.5) which hydrolyses oligomannose and hybrid but not complex-type glycans, and which therefore only deactivates Fc with engineered oligomannose-type glycans.

	Endo H	Endo S
High Mannose	+++	-
Oligomannose	+++	-
Hybrid	+++	-
+ core fucose	++	++
Biantennary	-	+++
+ core fucose	-	+++
Triantennary	-	-

Table 3.5: *N*-linked glycan di-*N*-acetylchitobiose core hydrolysing endoglycosidase enzyme H and S specificity.

Bacterially expressed Endo S deactivates serum IgG through cleavage of the Fc glycans. The molecular basis for this endoglycosidase mediated antibody deactivation was determined by our collaborators in the Glycobiology institute (Kavitha Baruah, Dr. Chris Scanlan) and the Division of Structural Biology, Wellcome Trust Centre for

Human Genetics (Thomas A. Bowden). The crystal structure of the deglycosylated Fc by Endo H revealed a collapsed quaternary architecture with a shift in CH2 domain orientation, as determined by X-ray crystallography to 2.3 Å, incompatible with any known Fc:FcR interactions (Figure 3.26). This finding was further confirmed by surface plasma resonance (SPR) binding assays of deglycosylated Fc to the FcRγIIIa (Figure 3.27).

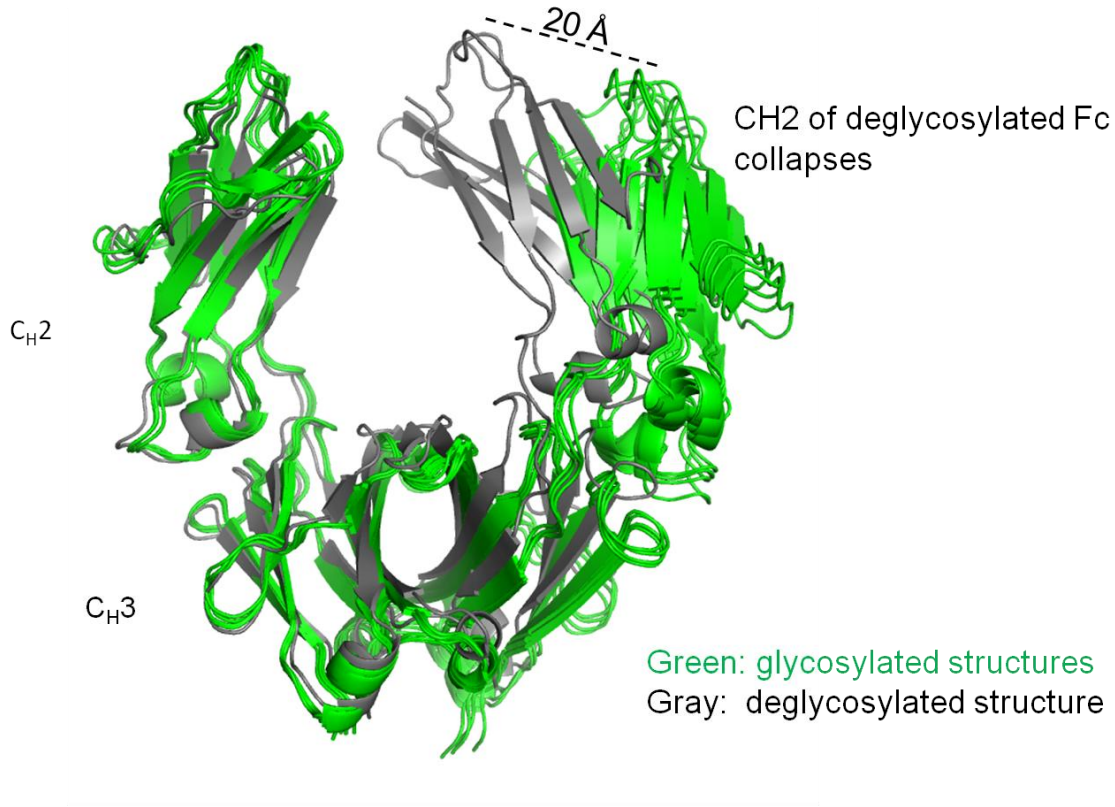


Figure 3.26: Enzymatic deglycosylation causes collapse of Fc C_H2 domains. X-ray crystal structure overlays of the glycosylated and deglycosylated Fc with the collapsed CH2 region of the Fc highlighted in grey. *Bowden et. al, unpublished data.*

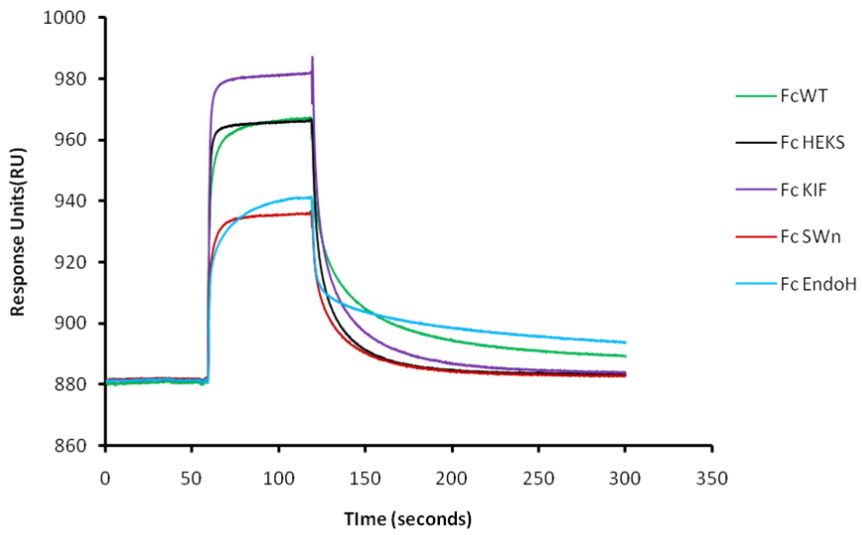


Figure 3.27: Enzymatic deglycosylation causes inactivation of Fc. SPR binding assays of deglycosylated Fc to the FcrγIIIa. Data provided by Kavitha Baruah. *Baruah et. al, unpublished data.*

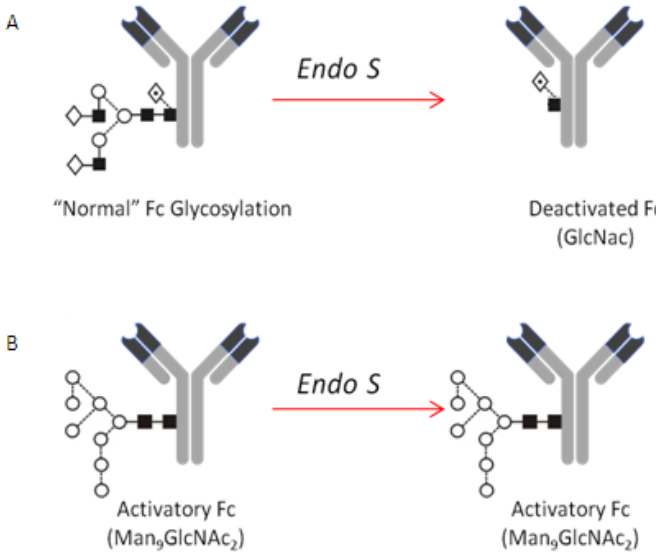


Figure 3.28: Endo S reduces the affinity of human serum antibodies but not of engineered (oligomannose) monoclonals. Schematic diagram showing the action of Endo S against 'normal' glyccosylated Fc (A) and against engineered glycosylation Fc (B). *Bowden et. al, unpublished data.*

Enzymatic deglycosylation of serum antibody has emerged as a therapeutic approach to abolish the Fc:FcR interaction used by pathogens responsible for a number of autoimmune diseases.[20] The reciprocal specificities discussed in this chapter of Endo S and Endo H hint at the possibility of controlling the activity of natural and glycoengineered antibodies where the cellular immune system could be refocused towards a single monoclonal antibody. Work is currently ongoing to determine whether the affinity of unnatural oligomannose-type Fc to FcrγIIIa in human serum is unaffected by the hydrolysing ability of Endo S, and whether the amount of unnatural antibody bound to the receptor increases as the natural glycoforms, susceptible to Endo S hydrolysis, are displaced (Figure 3.28).

A manuscript for publication reporting these findings is currently being prepared (See Appendix).

References

1. (a) Collin, M.; Olsen, A., Effect of SpeB and EndoS from Streptococcus pyogenes on human immunoglobulins. *Infect Immun* **2001,** *69* (11), 7187-7189; (b) Collin, M.; Olsen, A., EndoS, a novel secreted protein from Streptococcus pyogenes with endoglycosidase activity on human IgG. *Embo Journal* **2001,** *20* (12), 3046-3055.

2. Sutton, B. J.; Phillips, D. C., The 3-Dimensional Structure of the Carbohydrate within the Fc-Fragment of Immunoglobulin-G. *Biochem Soc T* **1983,** *11* (2), 130-132.

3. Sambrook J, M. F., E.F and Maniatis, T., *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, **1989**.

4. (a) Chang, J. Y., Thrombin Specificity - Requirement for Apolar Amino-Acids Adjacent to the Thrombin Cleavage Site of Polypeptide Substrate. *European Journal of Biochemistry* **1985,** *151* (2), 217-224; (b) Chang, J. Y.; Alkan, S. S.; Hilschmann, N.; Braun, D. G., Thrombin Specificity - Selective Cleavage of Antibody Light-Chains at the Joints of Variable with Joining Regions and Joining with Constant Regions. *European Journal of Biochemistry* **1985,** *151* (2), 225-230.

5. Allhorn, M.; Olsen, A.; Collin, M., EndoS from Streptococcus pyogenes is hydrolyzed by the cysteine proteinase SpeB and requires glutamic acid 235 and tryptophans for IgG glycan-hydrolyzing activity. *Bmc Microbiol* **2008,** *8*.

6. Harvey, D. J.; Royle, L.; Radcliffe, C. M.; Rudd, P. M.; Dwek, R. A., Structural and quantitative analysis of N-linked glycans by matrix-assisted laser desorption ionization and negative ion nanospray mass spectrometry. *Analytical Biochemistry* **2008,** *376* (1), 44-60.

7. Takegawa, K.; Yamaguchi, S.; Kondo, A.; Iwamoto, H.; Nakoshi, M.; Kato, I.; Iwahara, S., Transglycosylation Activity of Endo-Beta-N-Acetylglucosaminidase from Arthrobacter-Protophormiae. *Biochemistry International* **1991,** *24* (5), 849-855.

8. Robbins, P. W.; Trimble, R. B.; Wirth, D. F.; Hering, C.; Maley, F.; Maley, G. F.; Das, R.; Gibson, B. W.; Royal, N.; Biemann, K., Primary Structure of the Streptomyces Enzyme "Endo-Beta-N-Acetylglucosaminidase-H. *J Biol Chem* **1984,** *259* (12), 7577-7583.

9. Trimble, R. B.; Tarentino, A. L., Identification of Distinct Endoglycosidase (Endo) Activities in Flavobacterium-Meningosepticum - Endo-F1, Endo-F2, and Endo-F3 - Endo-F1 and Endo-H Hydrolyze Only High Mannose and Hybrid Glycans. *J Biol Chem* **1991,** *266* (3), 1646-1651.

10. Maley, F.; Trimble, R. B.; Tarentino, A. L.; Plummer, T. H., Characterization of Glycoproteins and Their Associated Oligosaccharides through the Use of Endoglycosidases. *Analytical Biochemistry* **1989,** *180* (2), 195-204.

11. (a) Fujita, K.; Takegawa, K., Chemoenzymatic synthesis of neoglycoproteins using transglycosylation with endo-beta-N-acetylglucosaminidase A. *Biochem Bioph Res Co* **2001,** *282* (3), 678-682; (b) Takegawa, K.; Yamabe, K.; Fujita, K.; Tabuchi, M.; Mita, M.; Izu, H.; Watanabe, A.; Asada, Y.; Sano, M.; Kondo, A.; Kato, I.; Iwahara, S., Cloning, sequencing, and expression of Arthrobacter protaphormiae endo-beta-N-acetylglucosaminidase in Escherichia coli. *Archives of Biochemistry and Biophysics* **1997,** *338* (1), 22-28.

12. (a) Guan, C. a. W., S., unpublished results. *New England Biolabs Inc*; (b) Robbins, P. W.; Wirth, D. F.; Hering, C., Expression of the Streptomyces Enzyme Endoglycosidase-H in Escherichia-Coli. *J Biol Chem* **1981,** *256* (20), 640-644.

13. Tarentino, A. L.; Gomez, C. M.; Plummer, T. H., Deglycosylation of asparagine-linked glycans by peptide:N-glycosidase F. *Biochemistry-Us* **1985,** *24* (17), 4665-4671.

14. Plummer, T. H.; Tarentino, A. L., Purification of the oligosaccharide-cleaving enzymes of. *Glycobiology* **1991,** *1* (3), 257-263.

15. Doores, K. J.; Bonomelli, C.; Harvey, D. J.; Vasiljevic, S.; Dwek, R. A.; Burton, D. R.; Crispin, M.; Scanlan, C. N., Envelope glycans of immunodeficiency virions are almost entirely oligomannose antigens. *P Natl Acad Sci USA* **2010,** *107* (31), 13800-13805.

16. (a) Owens, R. J.; Chang, V. T.; Crispin, M.; Aricescu, A. R.; Harvey, D. J.; Nettleship, J. E.; Fennelly, J. A.; Yu, C.; Boles, K. S.; Evans, E. J.; Stuart, D. I.; Dwek, R. A.; Jones, E. Y.; Davis, S. J., Glycoprotein structural genomics: Solving the glycosylation problem. *Structure* **2007,** *15* (3), 267-273; (b) Elbein, A. D.; Tropea, J. E.; Mitchell, M.; Kaushal, G. P., Kifunensine, a Potent Inhibitor of the Glycoprotein Processing Mannosidase-I. *J Biol Chem* **1990,** *265* (26), 15599-15605.

17. Jones, E. Y.; Aricescu, A. R.; Lu, W. X., A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallographica Section D-Biological Crystallography* **2006,** *62*, 1243-1250.

18. Khamlichi, A. A.; Aucouturier, P.; Preudhomme, J. L.; Cogne, M., Structure of Abnormal Heavy-Chains in Human Heavy-Chain-Deposition Disease. *Eur J Biochem* **1995,** *229* (1), 54-60.

19. Gross, V.; Tranthi, T. A.; Vosbeck, K.; Heinrich, P. C., Effect of Swainsonine on the Processing of the Asparagine-Linked Carbohydrate Chains of Alpha1-Antitrypsin in Rat Hepatocytes - Evidence for the Formation of Hybrid Oligosaccharides. *J Biol Chem* **1983,** *258* (6), 4032-4036.

20. (a) Allhorn, M.; Briceno, J. G.; Baudino, L.; Lood, C.; Olsson, M. L.; Izui, S.; Collin, M., The IgG-specific endoglycosidase EndoS inhibits both cellular and complement-mediated autoimmune hemolysis. *Blood* **2010,** *115* (24), 5080-5088; (b) Ravetch, J. V.; Albert, H.; Collin, M.; Dudziak, D.; Nimmerjahn, F., In vivo enzymatic modulation of IgG glycosylation inhibits autoimmune disease in an IgG subclass-

dependent manner. *P Natl Acad Sci USA* **2008,** *105* (39), 15005-15009; (c) Nandakumar, K. S.; Collin, M.; Olsen, A.; Nimmerjahn, F.; Blom, A. M.; Ravetch, J. V.; Holmdahl, R., Endoglycosidase treatment abrogates IgG arthritogenicity: Importance of IgG glycosylation in arthritis. *Eur J Immunol* **2007,** *37* (10), 2973-2982.

Chapter 4

Studies towards a transglycosylation model

4.1. Transglycosylation model

The use of endoglycosidase enzymes to chemoenzymatically remodel glycoproteins has become a successful method for homogeneous glycoprotein synthesis.[1] Endo-β-*N*-acetylglucosaminidase (Endo) A isolated from *Arthrobacter protophormiae*[2] demonstrates the ability to transglycosylate (the transfer of a glycosyl group to an acceptor) *N*-linked glycans.[3-7] The transglycosylation ability of Endo A has been recently explored in the BGD group[8] which focused on the Endo A catalysed transglycosylation of the glycoprotein mutant subtilisin *bacillus lentus* (SBL S156C) bearing 'unnatural linkages' (Figure 4.1). This work utilised the knowledge that Endo A catalyses the transglycosylation of oligo-mannose oxazoline donors.[9, 10]

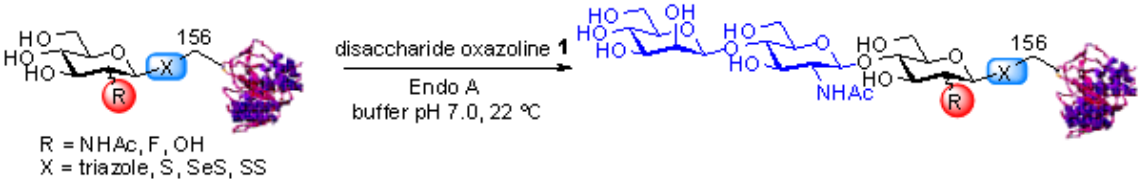


Figure 4.1: Endo A catalyzed transglycosylations of SBL protein with various linkers using disaccharide oxazoline **1** as a glycosyl donor.[8]

In this thesis, the glycan-hydrolysing ability of Endo S was explored and it was found to be resistant to oligo-mannose *N*-linked glycans. Using WT Fc glycoprotein (expressed and purified as described in Chapter 3) as a model protein for this study, we were able to enzymatically cleave the native complex type *N*-linked glycan on the Fc fragment with Endo S leaving a core fucosylated GlcNAc intact on the protein. Following core defucosylation using α-L-fucosidase, Endo A can catalyse

transglycosylation using a synthetic disaccharide oxazoline to generate synthetic oligo-mannose Fc glycoprotein which is resistant to Endo S activity (Figure 4.2).

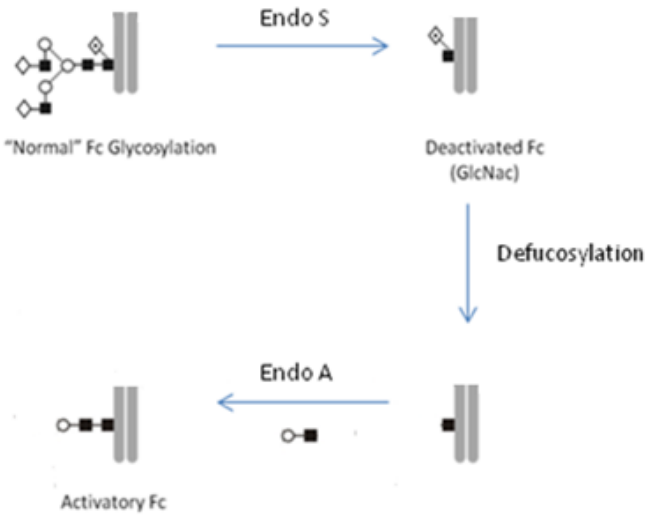


Figure 4.2: Transglycosylation model to make synthetic oligo-mannose Fc glycoproteins.

4.2. Quantitative analysis of Endo A on gp120$_{BAL}$ glycans

We wanted to compare the glycan-hydrolysing ability of Endo A to Endo S using the same glycan set used in the Endo S experiments (Chapter 3) to confirm that the specificity of Endo A is reciprocal to that of Endo S.

Glycans released from recombinant gp120$_{BAL}$ protein by in-gel PNGase F digestion from three different types of expression systems were tested:

1) gp120$_{BAL}$ protein expressed in HEK293T cells

   - Normal mammalian *N*-linked glycosylation pathway.

2) $gp120_{BAL}$ protein expressed in HEK293T cells with kifunensine inhibitor (Figure 3.16).[11, 12]

   - Kifunensine inhibits ER α-mannosidase I and Golgi α-mannosidase IA-C protein in glycan pathway so $gp120_{BAL}$ protein has mainly $Man_9GlcNAc_2$ sugars present with small amounts of $Man_{5\text{-}8}GlcNAc_2$.

3) $gp120_{BAL}$ protein expressed in GnT-I HEK293S cells

   - This cell line is deficient in GnT-I, therefore expressed glycoproteins contain exclusively $Man_5GlcNAc_2$ glycans with some core fucosylated present.

In the $gp120_{BAL}$ experiments, two conditions were ran:

1) A control experiment of just the free *N*-linked glycans released from $gp120_{BAL}$ protein by in-gel PNGase F digestion.

2) The free *N*-linked glycans incubated with Endo A for 2 hours at 37°C

All reaction samples were filtered through a MultiScreen® 96-Well Filtration Plate (Millipore), dried by SpeedVac and cleaned using a Nafion membrane (Sigma Aldrich).

The resultant MALDI-MS of the Endo A enzymatic glycan digestion reactions are shown below in Figure 4.3A-D.

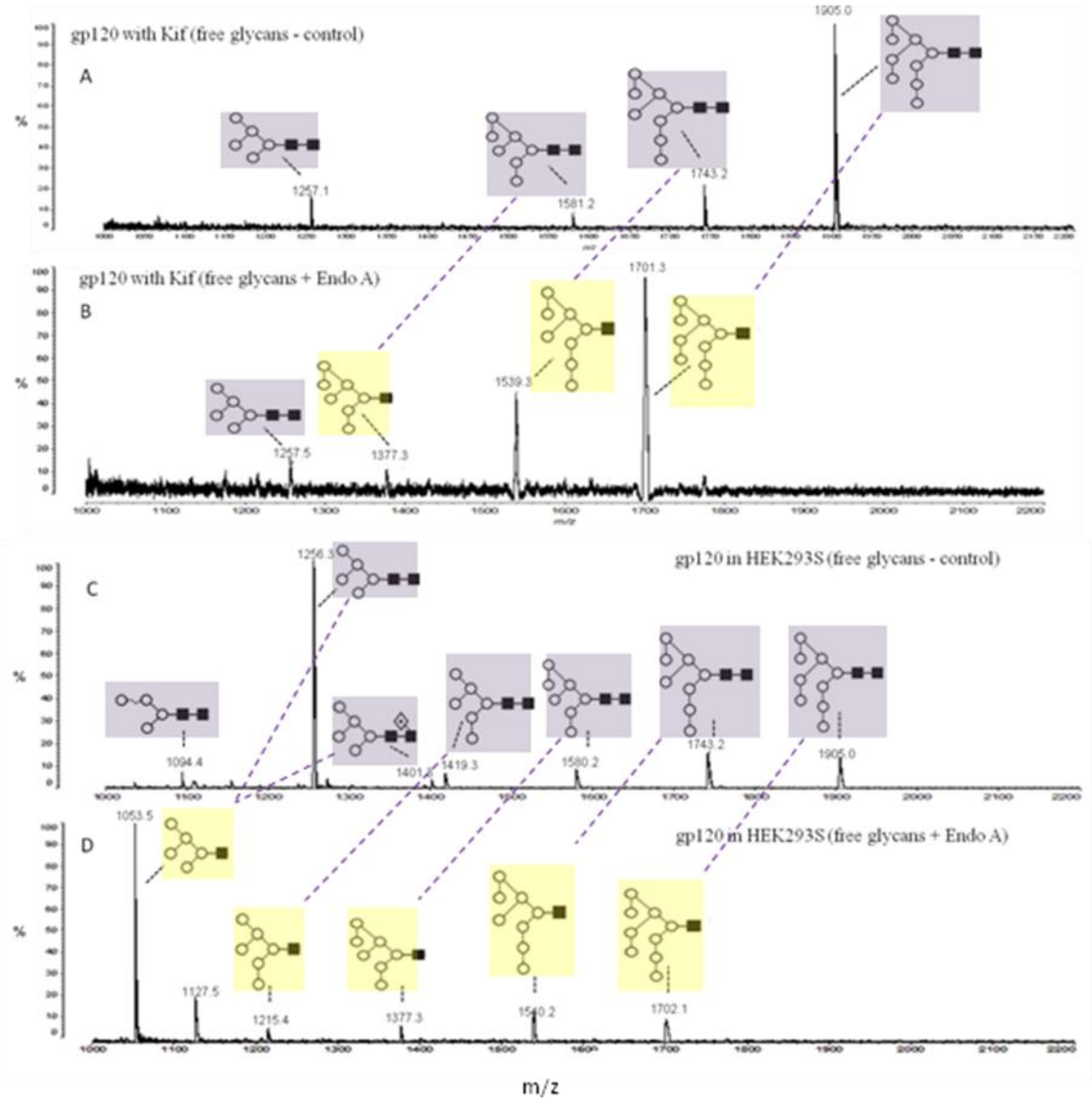


Figure 4.3: Oligo-mannose and high-mannose *N*-linked glycans are susceptible to Endo A hydrolysis. MALDI-MS of Endo A enzymatic glycan reactions. **A** Glycans released from $gp120_{BAL}$ expressed in HEK293T cells with kifunensine. **B** Glycans (A) incubated with Endo A. **C** Glycans released from $gp120_{BAL}$ expressed in HEK293S cells. **D** Glycans (C) incubated with Endo A.

When Endo A is incubated with free glycans released from $gp120_{BAL}$ with Kif inhibitor and gp120 from HEK293S cell line we can see a mass shift for all of the high-

mannose glycan peaks, corresponding to the removal of a core GlcNAc sugar (203 gmol^{-1}). This result is as expected and the same as the result obtained from Endo H.

N-linked glycans released from $gp120_{BAL}$ expressed in HEK293T cells, which undergo the normal *N*-linked glycosylation pathway, were desialyated with 1.0% AcOH at 80°C for 1 hour prior to MALDI analysis (Figure 4.4).

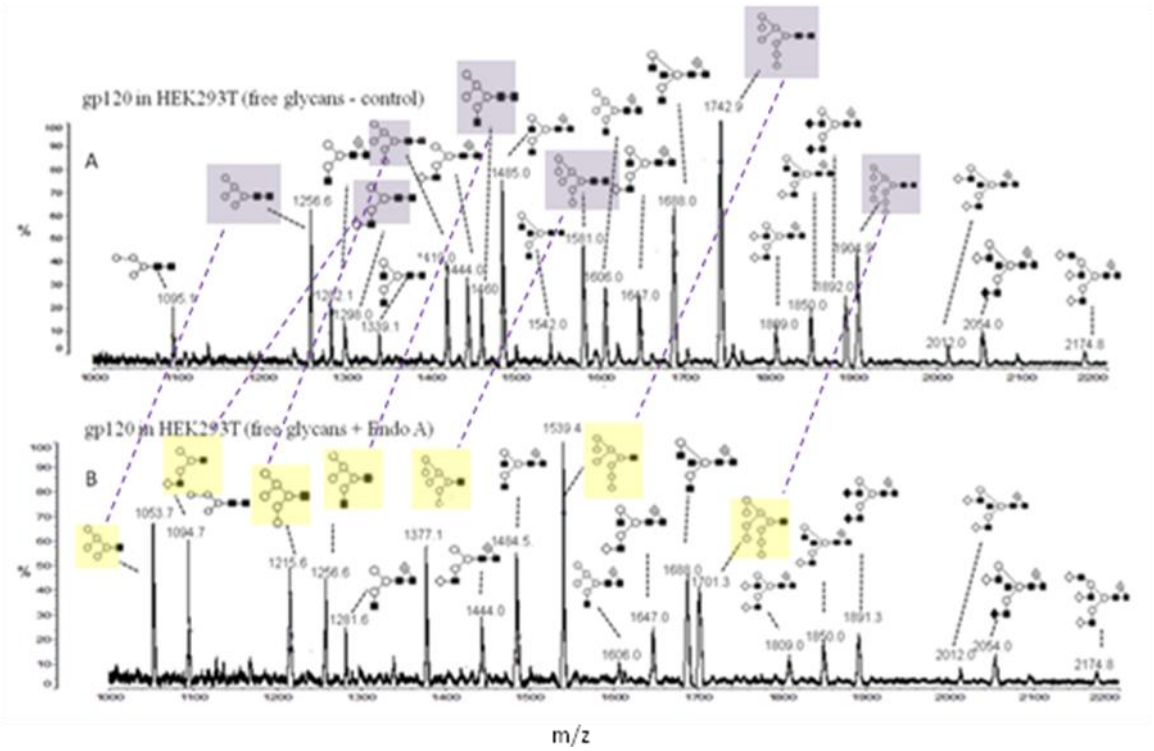


Figure 4.4: Complex type *N*-linked glycans are resistant to Endo A hydrolysis. MALDI-MS of Endo A enzymatic glycan reactions. **A** Glycans released from $gp120_{BAL}$ expressed in HEK293T cells. **B** Glycans (A) incubated with Endo A.

When Endo A was incubated with *N*-linked glycans released from $gp120_{BAL}$ in HEK293T cells, we observe a shift in the high-mannose and oligo-mannose peaks (Figure 4.4B). The observed mass shift (203 gmol^{-1}) corresponds to the loss of a core GlcNAc sugar. The complex type *N*-linked glycans in this set have not been used as

substrates for hydrolysis by Endo. This result shows that Endo A, like Endo H is reciprocal in its substrate specificity to Endo S (Chapter 3).

These results also suggest that the usual disaccharide oxazoline substrate[13] which was used for extension of the glycan chain to form homogeneous glycoproteins with Endo A[8] would not be an ideal substrate for the potential transglycosylation using Endo S.

4.3. Towards the synthesis of a glycosyl donor for Endo A glycan extension

Previous studies have shown that the oxazoline of disaccharide β-Man-(1→4)-GlcNAc **1** is a good substrate for a variety of endoglycosidase catalysed glycosylations.[8] The dissacharide oxazoline **1** has been synthesised by different routes involving the starting materials **6** and **11**.[9]

The retrosynthetic disconnection of disaccharide β-Man-(1→4)-GlcNAc oxazoline **1** is shown below (Figure 4.5).

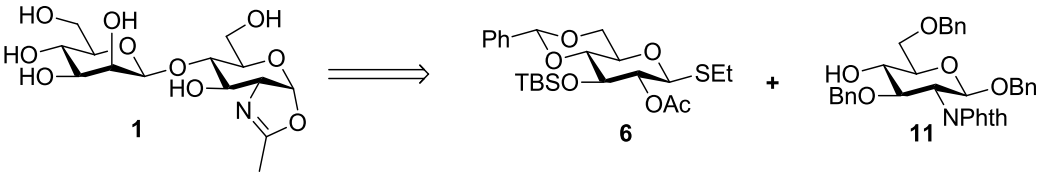
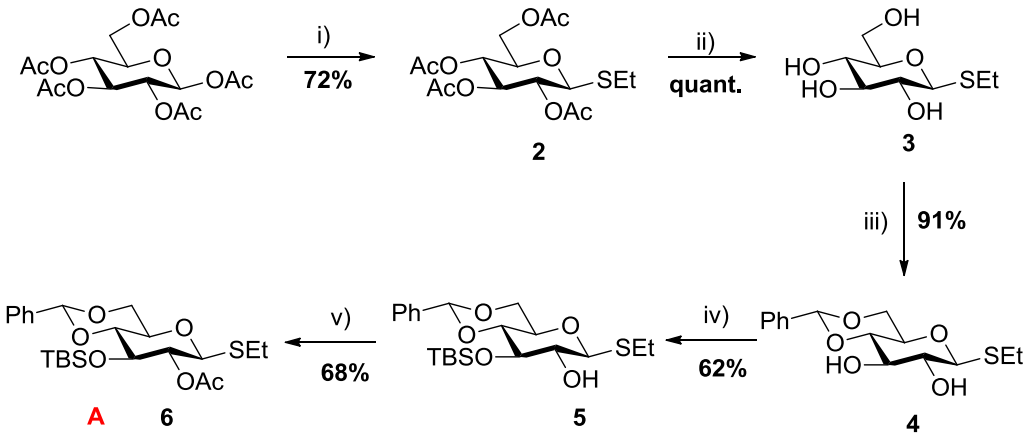
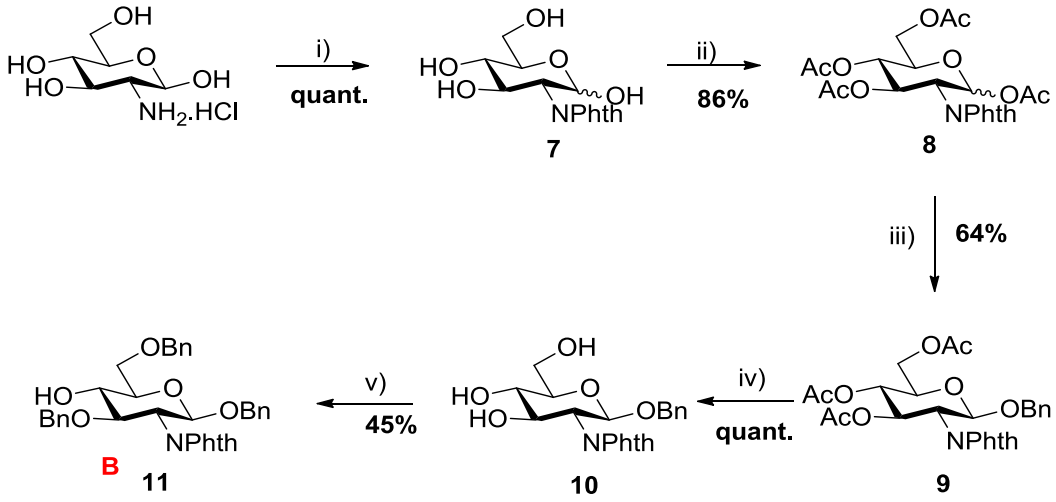


Figure 4.5: Retrosynthetic disconnection of disaccharide β-Man-(1→4)-GlcNAc oxazoline.

The synthetic routes to the monosaccharide glycosyl donor **6** and glycosyl acceptor **11** are summarised in Scheme 1 and Scheme 2 respectively.

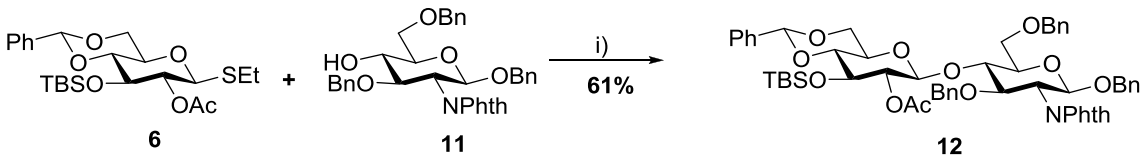


Scheme 1. Conditions i) EtSH, $SnCl_4$, DCM; ii) NaOMe, MeOH; iii) BDMA, TsOH, MeCN; iv) TBSCl, imidazole, DMF; v) Ac_2O, py.



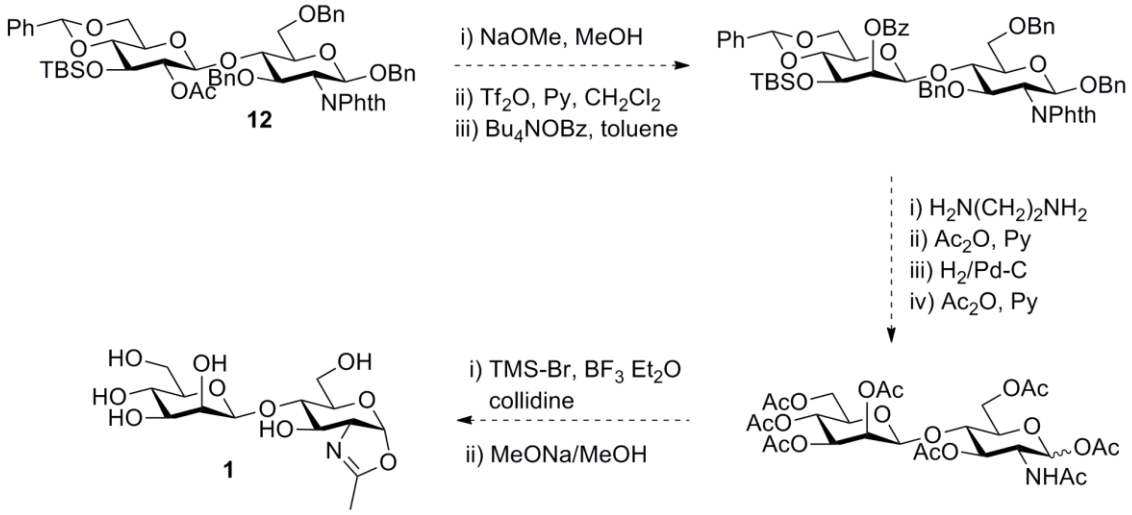
Scheme 2. Conditions i) $Phth_2O$, NaOAc, Acetone:Water (4:1), $NaHCO_3$; ii) Ac_2O, py; iii)BnOH, TMSOTf, DCM; iv) NaOMe, MeOH; v) a) Bu_2SnO, PhMe; b) TBAI, BnBr,PhMe

One of the key steps in the synthesis is the carefully stereocontrolled β-glycosylation of the thioglycoside **6** and the alcohol **11** using *N*-iodosuccinimide (NIS) and triflic acid.[14] The synthetic route is summarised in Scheme 3 and yields the desirable dissacharide core in 61% yield with control of anomeric stereochemistry due to the neighbouring group participation of the acetate group on the thioglycoside (Figure 1.5).



Scheme 3. Conditions i) NIS, TfOH, DCM, -40°C, 4 hrs

Due to time constraints, the synthesis of the disaccharide oxazoline **1** could unfortunately not be completed. However, the synthesis would have been continued as indicated in Scheme 4 to achieve the desired substrate for glycan extension.[9]



Scheme 4. Planned synthesis of disaccharide oxazoline **1**

Work is currently on-going in the BGD group to complete this synthesis and explore its use using Endo A for enzymatic glycan extension of WT Fc glycoprotein.

4.4. General Considerations for Chemistry

4.4.1. NMR data

Proton nuclear magnetic resonance spectra (^{1}H) were recorded on a Bruker DPX 200 (200 MHz), Bruker DPX 400 (400 MHz), Bruker DQX 400 (400 MHz), Bruker AC 500 (500 MHz) spectrometer. Carbon nuclear magnetic resonance spectra (^{13}C) were recorded on a Bruker DQX 400 (100 MHz) spectrometer or on a Bruker AC 500 (125 MHz) with a ^{13}C cryoprobe (125 MHz). Spectra were fully assigned based on chemical shifts using a combination of COSY, HMQC, DEPT 135 data and comparison with spectra of related compounds. All chemical shifts are referenced to the relevant residual solvent peak and were quoted in ppm with respect to the internal standard of tetramethylsilane (TMS). Resonances are described as s (singlet), d (doublet), t (triplet), q (quartet), quin (quintet), sxt (sextet), sept (septet), oct (octet), multiplet (m), app. (apparent) and br (broad). Coupling constants (*J*) are given in hertz (Hz).

4.4.2. Mass Spectra

Mass spectra were recorded on a Fisons Platform VG Spectrometer (ESI) using electrospray ionisation in a positive (ESI+) or negative (ESI-) scan mode. *m/z* are reported with their percentage abundance.

4.4.3. Chromatography

Thin layer chromatography (TLC) was carried out on Merk Kieselgel $60F_{254}$ 0.2 mm precoated aluminium backed plates. Product spots were visualized with a combination of the following: 254 nm UV lamp, aqueous phosphomolybdic acid and $Ce^{(IV)}$ (2.5% phosphomolybdic acid hydrate, 1% cerium(IV) sulfate hydrate, and 6% H_2SO_4); ammonium molybdate (5% in 2M H_2SO_4), and ninhydrin (0.2% in ethanol). Retention

factors (R_f) are reported with the solvent system used in parentheses. Flash column chromatography (FC) was carried out with Fluka Kiegselgel 60 220-440 mesh silica gel, with the solvent system used in parentheses; graduated solvent systems are indicated with an arrow.

4.4.4. Solvents and reagents

Unless specified, reagents and starting materials were obtained from standard commercial sources and were used without further purification. DCM was dried by distillation over alumina. Remaining anhydrous solvents were purchased from Aldrich or Fluka. All other solvents were used as supplied (analytical or HPLC grade), without further purification. 'Petrol' refers to the fraction of petroleum ether boiling in the range 40-60 °C. Brine refers to a saturated aqueous solution of sodium chloride.

4.4.5. Reactions

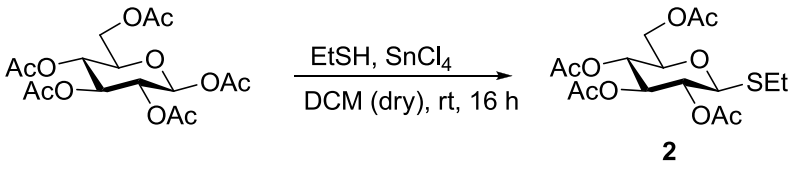
All non-aqueous reactions were carried out in oven-dried glassware under an inert atmosphere of argon where appropriate. Dowex ion exchange resins were activated by washing with methanol, water, 1 M HCl, and water until the filtrate was pH neutral.

4.5. Chemical Compounds

4.5.1. Ethyl 2,3,4,6-tetra-*O*-acetyl-1-thio-β-D-glucopyranoside (2)[15]

Formula: $C_{16}H_{24}O_9S$

M_r: 392.42 g mol^{-1}



2

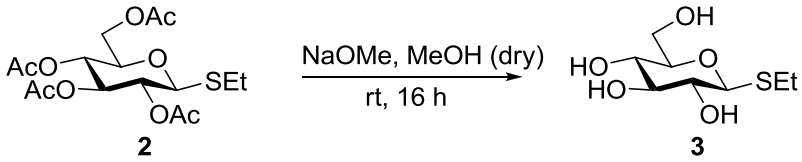
To a solution of 1,2,3,4,6-penta-*O*-acetyl-β-D-glucopyranoside (25.0 g, 64.1 mmol, 1 eq) in anhydrous dichloromethane (130 mL), ethanethiol (8.63 mL, 115 mmol, 1.8 eq) was added followed by the careful dropwise addition of tin (IV) chloride (1.12 mL, 9.61 mmol, 0.15 eq). The reaction mixture was stirred at RT under argon overnight. After 16 hours, TLC analysis [Pet. Ether/EtOAc (3:2)] showed complete consumption of starting material. HCl (2 M, 130 mL) was added to the reaction mixture and stirred for 30 mins. The reaction mixture was then poured into a separating funnel and the organic layer was extracted with DCM (2 x 200 mL). Combined organic extracts were washed with water (200 mL), saturated aqueous sodium hydrogen carbonate (200 mL) and brine (200 mL). After drying over $MgSO_4$, filtering and concentrating *in vacuo*, the residue was recrystallised from ethanol to yield the desired product **2** (18.62 g, 72%) as a white crystalline solid. The material was spectroscopically identical to that reported in the literature[16]. R_f 0.53 [Pet. Ether/EtOAc (3:2)]; ^{1}H NMR (400 MHz, $CDCl_3$) δ = 5.23 (1H, t, $J_{3,4}$ = 9.4 Hz, H-3), 5.12-5.02 (2H, m, H-4, H-2), 4.50 (1H, d, $J_{1,2}$ = 10.1 Hz, H-1), 4.26 (1H, dd, $J_{5,6}$ = 12.4, $J_{6,6'}$ =5.1 Hz, H-6), 4.15 (1H, dd, $J_{5,6'}$ = 12.4 Hz, $J_{6,6'}$ = 2.3 Hz, H-6'), 3.74-3.70 (1H, m, H-5), 2.74-2.70 (2H, m, S$CH_2$$CH_3$), 2.09 (3H, s, CH_3), 2.07 (3H, s,

C<u>H</u>*)*, 2.04 (3H, s, C<u>H</u>*)*, 2.02 (3H, s, C<u>H</u>), 1.28 (3H, t, *J* = 7.5 Hz, $SCH_2C\underline{H}$*);* ^{13}C NMR (100 MHz, $CDCl_3$) δ = 83.5 (C-1), 75.9 (C-5), 73.9 (C-3), 69.8 (C-2), 68.3 (C-4), 62.1 (C-6), 24.1 ($S\underline{C}H_2CH_3$), 20.7 (4 OAc), 14.8 ($SCH_2\underline{C}H_3$); LRMS m/z (ES+) $[M + Na]^+$ peak 415 (100%).

4.5.2. Ethyl 1-thio-β-D-glucopyranoside (3)[17]

Formula: $C_8H_{16}O_5S$

M_r: 224.27 g mol^{-1}

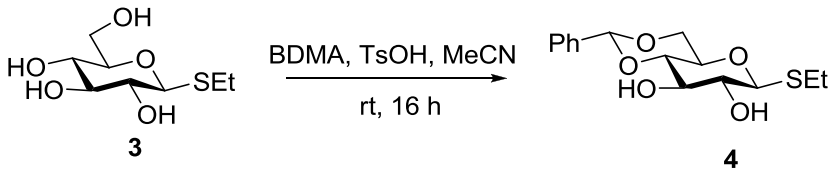


To a solution of ethyl 2,3,4,6-tetra-*O*-acetyl-1-thio-β-D-glucopyranoside **2** (11.4 g, 29.1 mmol, 1 eq) in anhydrous methanol (180 mL), a catalytic amount of sodium methoxide (0.16 g, 2.91 mmol, 0.1 eq) was added. The reaction mixture was stirred at RT under argon overnight. After 16 hours, TLC analysis [DCM/MeOH (9:1)] showed complete consumption of starting material. The reaction was neutralised using acidified Dowex, filtered and concentrated *in vacuo* to give product **3** (6.53 g, quant.) as a white crystalline solid. R_f 0.21 [DCM/MeOH (9:1)]; ^{1}H NMR (400 MHz, MeOD) δ = 4.36 (1H, d, $J_{1,2}$ = 9.6 Hz, H-1), 3.85 (1H, app. d, $J_{5,6}$ = 10.4 Hz, H-6), 3.64 (1H, dd, $J_{5,6'}$ = 11.8 Hz, $J_{6,6'}$ = 5.3 Hz, H-6'), 3.37-3.26 (3H, m, H-3, H-4, H-5), 3.19 (1H, app. t, $J_{1,2}$ =8.6 Hz, H-2), 2.80-2.68 (2H, m, $SC\underline{H_2}CH_3$), 1.28 (3H, t, *J* = 7.5 Hz, $SCH_2C\underline{H_3}$*);* ^{13}C NMR (100 MHz, MeOD) δ = 85.9 (C-1), 81.0 (C-5), 78.6 (C-3), 73.3 (C-2), 70.5 (C-4), 61.9 (C-6), 23.8 ($S\underline{C}H_2CH_3$), 14.4 ($SCH_2\underline{C}H_3$); LRMS m/z (ES+) $[M + Na]^+$ peak 245 (100%).

4.5.3. Ethyl 4,6-*O*-benzylidene-1-thio-β-D-glucopyranoside (4)[16]

Formula: $C_{15}H_{20}O_5S$

M_r: 312.38 g mol^{-1}



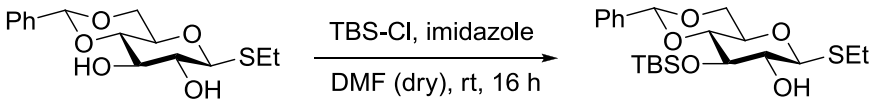
To a solution of ethyl 1-thio-β-D-glucopyranoside **3** (8.57 g, 38.2 mmol, 1 eq) dissolved in anhydrous acetonitrile (200 mL), benzaldehyde dimethylacetal (6.88 mL, 45.9 mmol, 1.2 eq) and *para*-toluenesulfonic acid monohydrate (218 mg, 1.15 mmol, 0.03 eq) were added. The reaction mixture was stirred at RT under argon overnight. After 16 hours, TLC analysis [DCM/MeOH (9:1)] showed complete consumption of starting material. The reaction was neutralised with the addition of triethylamine (0.15 mL), stirred for 30 mins and concentrated *in vacuo*. The resultant oil was partitioned between EtOAc (250 mL) and water (250 mL). The aqueous layer was extracted with EtOAc (2 x 200 mL). Combined organic extracts were washed with water (200 mL), saturated aqueous sodium hydrogen carbonate (200 mL) and brine (200 mL). After drying over $MgSO_4$, filtering and concentrating *in vacuo*, the crude residue was recrystallised from toluene to yield the desired product **4** (10.83 g, 91%) as a white powder. R_f 0.69 [DCM/MeOH (9:1)]; ^{1}H NMR (400 MHz, $CDCl_3$) δ = 7.51-7.49 (2H, m, C*H*$_{arom.}$ benzylidene), 7.40-7.37 (3H, m, C*H*$_{arom.}$ benzylidene), 5.55 (1H, s, C*H* benzylidene), 4.47 (1H, d, $J_{1,2}$ = 9.8 Hz, H-1), 4.36 (1H, dd, , $J_{5,6}$ = 10.6 Hz, $J_{6,6'}$ = 4.8 Hz, H-6), 3.86 (1H, app. t, $J_{2,3}$ = 7.1 Hz, H-3), 3.79 (1H, app.t, $J_{5,6'}$ = 10.2 Hz, H-6'), 3.58 (1H, app.t, $J_{3,4}$ = 9.1 Hz, H-4), 3.54-3.49 (2H, m, H-5, H-2), 2.80-2.75 (2H, m, SC*H*$_2$$CH_3$), 1.34 (3H, t, *J* = 7.5 Hz, SCH_2C*H*$_3$); ^{13}C NMR

(100 MHz, $CDCl_3$) δ = 128.4 (*C*H$_{arom.}$ benzylidene), 126.3 (*C*H$_{arom.}$ benzylidene), 101.9 (*C*H benzylidene), 86.7 (C-1), 80.3 (C-4), 74.6 (C-3), 73.2 (C-2), 70.6 (C-5) 68.6 (C-6), 24.8 (S*C*H$_2$CH$_3$), 15.3 (SCH$_2$*C*H$_3$) ; LRMS m/z (ES+) $[Na + Na]^+$ 335 (100%)

4.5.4. Ethyl 4,6-*O*-benzylidene-3-*O*-(*tert*-butyldimethylsilyl)-1-thio-β-D-glucopyranoside (5)[18]

Formula: $C_{21}H_{34}O_5SSi$

M_r: 426.64 g mol^{-1}



4 **5**

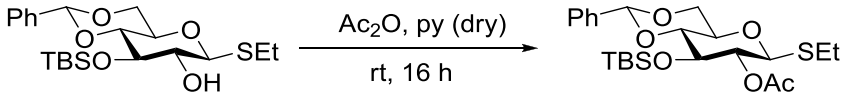
To a stirred solution of ethyl 4,6-*O*-benzylidene-1-thio-β-D-glucopyranoside **4** (5.50 g, 17.6 mmol, 1 eq) and imidazole (1.80 g, 26.4 mmol, 1.1 eq) in anhydrous dimethylformamide (50 mL), *tert*-butyldimethylsilyl chloride (2.92 g, 19.4 mmol, 1.5 eq) was added in small portions at 0 ^oC. The reaction mixture was stirred at RT under argon overnight. After 16 hours, TLC analysis [Pet. Ether/EtOAc (4:1)] showed complete consumption of starting material. The reaction was concentrated *in vacuo*, and the resultant oil was partitioned between EtOAc (150 mL) and aqueous NH_4Cl (150 mL). Then the aqueous layer was extracted with EtOAc (2 x 250 mL) and the combined organic extracts were washed with 1 M aqueous HCl (100 mL), saturated aqueous sodium hydrogen carbonate (100 mL) and brine (100 mL). After drying over $MgSO_4$, filtering and concentrating *in vacuo*, the resulting clear oil was purified by column chromatography [Pet. Ether/EtOAc (4:1)] to yield the title compound **5** (4.60 g, 62%) as a clear colourless oil. R_f 0.67 [Pet. Ether/EtOAc (4:1)]; ^1H NMR (400 MHz, $CDCl_3$) δ =

7.50-7.48 (2H, m, C*H*$_{arom.}$ benzylidene), 7.38-7.36 (3H, m, C*H*$_{arom.}$ benzylidene), 5.53 (1H, s, C*H* benzylidene), 4.47 (1H, d, $J_{1,2}$ = 9.8 Hz, H-1), 4.34 (1H, dd, $J_{5,6}$ = 11.4 Hz, $J_{6,6'}$ = 4.8 Hz, H-6), 3.79-3.74 (2H, m H-6', H-3), 3.54-3.44 (3H, m, H-4, H-5, H-2), 2.80-2.74 (2H, m, SC*H*$_2$CH$_3$), 1.34 (3H, t, J = 7.5 Hz, SCH$_2$C*H*$_3$), 0.88 (9H, s, (C*H*$_3$)$_3$), 0.12 (3H, s, SiC*H*$_3$), 0.05 (3H, s, SiC*H*$_3$) ; LRMS m/z (ES+) $[M + Na]^+$ peak 449 (100%).

4.5.5. Ethyl 2-*O*-acetyl-4,6-*O*-benzylidene-3-*O*-(*tert*-butyldimethylsilyl)-1-thio-β-D-glucopyranoside (6)

Formula: $C_{23}H_{36}O_6SSi$

M_r: 468.68 g mol^{-1}



5 **6**

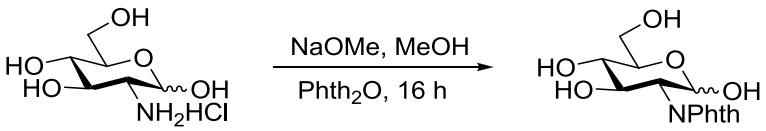
To a stirred solution of ethyl 4,6-*O*-benzylidene-3-*O*-(*tert*-butyldimethylsilyl)-1-thio-β-D-glucopyranoside **5** (4.48 g, 10.5 mmol, 1 eq) in anhydrous dichloromethane (100 mL), anhydrous pyridine (4.23 mL, 52.5 mmol, 5 eq) was added. Upon cooling the solution down to 0 ^{o}C, acetic anhydride (2.97 mL, 31.5 mmol, 3 eq) and 4-dimethylaminopyridine (0.13 g, 1.05 mmol, 0.1 eq) were added. The reaction mixture was stirred at RT under argon overnight. After 16 hours, TLC analysis [Pet. Ether/EtOAc (4:1)] showed complete consumption of starting material. The reaction was concentrated *in vacuo*, by co-evaporation with toluene. The resultant oil was partitioned between DCM (150 mL) and water (150 mL). Then the aqueous layer was extracted with DCM (2 x 250 mL) and the combined organic extracts were washed with 1 M aqueous HCl (100 mL), saturated aqueous sodium hydrogen carbonate (100 mL) and brine (100 mL). After drying over $MgSO_4$, filtering and concentrating *in vacuo*, the title compound **6** was obtained (3.38 g,

64 %) as a white solid. R_f 0.80 [Pet. Ether/EtOAc (4:1)]; ^{1}H NMR (400 MHz, $CDCl_3$) δ = 7.50-7.47 (2H, m, C*H*$_{arom.}$ benzylidene), 7.38-7.36 (3H, m, C*H*$_{arom.}$ benzylidene), 5.53 (1H, s, C*H* benzylidene), 5.01 (1H, app. t, *J* = 8.7 Hz, H-2), 4.46 (1H, d, $J_{1,2}$ = 10.1 Hz, H-1), 4.36 (1H, dd, , $J_{5,6}$ = 10.6 Hz, $J_{6,6'}$ = 4.8 Hz, H-6), 3.89 (1H, app. t, *J* = 8.7 Hz, H-3), 3.76 (1H, app. t, *J* = 10.1 Hz, H-6'), 3.55 (1H, app. t, *J* = 9.1 Hz, H-4), 3.51-3.45 (1H, m, H-5), 2.76-2.68 (2H, m, SC*H*$_2$CH$_3$), 2.12 (3H, s, OC*H*$_3$), 1.27 (3H, t, *J* = 7.5 Hz, SCH$_2$C*H*$_3$), 0.82 (9H, s, (C*H*$_3$)$_3$), 0.03 (3H, s, SiC*H*$_3$), 0.01 (3H, s, SiC*H*$_3$); ^{13}C NMR (100 MHz, $CDCl_3$) δ = 169.5, 137.0, 129.0 (C*H* benzylidene), 128.1 (*C*H$_{arom.}$ benzylidene), 126.3 (*C*H$_{arom.}$ benzylidene), 101.8 (*C*H benzylidene), 84.2 (C-1), 81.3 (C-4), 74.0 (C-3), 72.7 (C-2), 70.8 (C-5), 68.6 (C-6), 25.6 ((C*H*$_3$)$_3$), 23.9 (S*C*H$_2$CH$_3$), 21.2 (O*C*H$_3$), 14.8 (SCH$_2$*C*H$_3$), -4.2 (Si*C*H$_3$), -5 (Si*C*H$_3$).; LRMS m/z (ES+) $[M+Na]^+$ peak 491 (100%)

4.5.6. 2-Deoxy-2-*N*-phthalimido-D-glucopyranose (7)[19]

Formula: $C_{14}H_{15}NO_7$

M_r: 309.27 g mol^{-1}



7

To a stirred solution of D-glucosamine hydrochloride (50.0 g, 0.23 moles, 1 eq) in a 1 M solution of NaOH (11.2 g in 285 mL water, 0.28 moles, 1.2 eq), phthalic anhydride (34.4 g, 0.23 moles, 1 eq) was added. The reaction mixture was stirred at RT under argon overnight. After 16 hours, TLC analysis [WIpE (Water/Isopropanol/EtOAc) (1:2:2)] showed complete consumption of starting material. The reaction was then poured into a

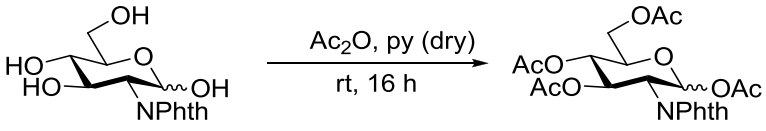
separating funnel, and partitioned with Et_2O. Then the aqueous layer was extracted with Et_2O (2 x 150 mL) and the combined organic extracts were washed with brine (100 mL). After drying over $MgSO_4$, filtering and concentrating *in vacuo*, to give 2-deoxy-2-phthalimido-D-glucopyranose **7** (71 g, 100%) as afforded as white solid. R_f 0.29 [WIpE (1:2:2)]; ^{1}H NMR (400 MHz, MeOD) δ = 7.70-7.63 (2H, m, C$\underline{H}_{arom.}$ Phth (α, β)), 7.51-7.43 (6H, m, C$\underline{H}_{arom.}$ Phth (α, β)), 5.25 (1H, d, $J_{1,2}$ = 3.8 Hz, H-1 (α)), 4.86 (1H, d, $J_{1,2}$ = 8.5 Hz, H-1 (β)), 4.01 (1H, dd, *J* = 10.9 Hz, *J* = 3.8 Hz, H-2 ((α)), 3.87-3.60 (10H, m), 3.51-3.37 (4H, m), 3.21 (1H, dd, *J* =10.6 Hz, *J* = 3.41 Hz, H-4 (α)).

NMR data: For the purposes of calculating multiplicities of individual proton shifts from integration values, the two anomers were treated independently of one another and the quoted multiplicities do not reflect the α/β ratio.

4.5.7. 1,3,4,6-Tetra-*O*-acetyl-2-deoxy-2-*N*-phthalimido-D-glucopyranoside (8)[19]

Formula: $C_{22}H_{23}NO_{11}$

M_r: 477.42 g mol^{-1}



7 **8**

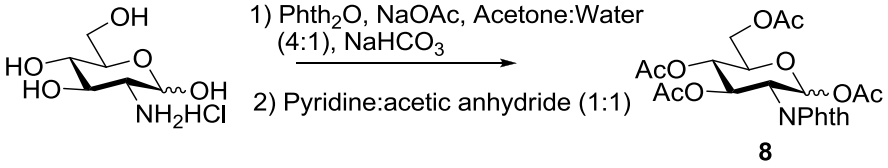
To a stirred solution of 2-deoxy-2-phthalimido-D-glucopyranoside **7** (71.0 g, 0.23 moles, 1 eq) in anhydrous pyridine (300 mL), acetic anhydride (130 mL, 1.61 moles, 7 eq) was added to the solution slowly at 0 ^oC. The reaction mixture was stirred at RT under argon overnight. After 16 hours, TLC analysis [Pet. Ether/EtOAc (3:2)] showed complete

consumption of starting material. The reaction was then concentrated *in vacuo*, by co-evaporation with toluene. The resultant brown oil was then partitioned between DCM and water. Then the aqueous layer was extracted with DCM (2 x 250 mL) and the combined organic extracts were washed with 1 M aqueous HCl (100 mL), saturated aqueous sodium hydrogen carbonate (100 mL) and brine (100 mL). After drying over $MgSO_4$, filtering and concentrating *in vacuo*, the crude material was purified by column chromatography [Pet. Ether/EtOAc (1:3)] to yield the title compound **8** (94.43 g, 86%) as a white solid. R_f 0.57 [Pet. Ether/EtOAc (3:2)]; ^{1}H NMR (400 MHz, $CDCl_3$) δ = 7.88-7.84 (m, 4H, C*H*$_{arom.}$ Phth (α+β)), 7.78-7.74 (m, 4H, C*H*$_{arom.}$ Phth (α+β)), 6.57 (1H, app. t, *J* = 9.1 Hz, H-3 (α)), 6.52 (1H, d, $J_{1,2}$ = 8.8 Hz, H-1 (β)), 6.30 (1H, d, $J_{1,2}$ = 3.3 Hz, H-1 (α)), 5.89 (1H, app. t, *J* = 9.3 Hz, H-3 (β)), 5.24-5.15 (2H, m, H-4 (α+β)), 4.74 (1H, dd, *J* = 11.6, *J* = 3.5 Hz, H-2 (α)), 4.47 (1H, app. t, *J* = 8.8 Hz, H-2 (β)), 4.37 (2H, dd, $J_{5,6}$ = 12.4, $J_{6,6'}$ = 4.3 Hz, H-6 (α+β)), 4.34-4.30 (1H, m, H-5 (α)), 4.17-4.12 (2H, m, H-6' (α+β)), 4.05-4.01 (m, H-5 (β)), 2.13 (3H, s, C*H*$_3$), 2.09 (3H, s, C*H*$_3$), 2.07 (3H, s, C*H*$_3$), 2.05 (3H, s, C*H*$_3$), 2.00 (3H, s, C*H*$_3$), 1.87 (3H, s, C*H*$_3$); ^{13}C NMR (100 MHz, $CDCl_3$) δ = 134.4, 123.8, 123.7, 90.5 (C-1 (α)), 89.8 (C-1 (β)), 72.6 (C-5 (β)), 70.5 (C-3 (β)), 70.2 (C-5 (α)), 69.4 (C-4 (α)), 68.3 (C-4 (β)), 67.0 (C-3 (α)), 61.5 (C-6 (α+β)), 53.5 (C-2 (β)), 52.8 (C-2 (α)), 20.9, 20.8, 20.7, 20.6, 20.4 (6 x OAc); LRMS m/z (ES+) $[M + Na]^+$ peak 500 (100%).

4.5.8. 1,3,4,6-Tetra-*O*-acetyl-2-deoxy-2-*N*-phthalimido-D-glucopyranoside (8)[20]

Formula: $C_{22}H_{23}NO_{11}$

M_r: 477.42 g mol^{-1}



8

A solution of D-glucosamine hydrochloride (8.63 g, 40.0 mmol, 1 eq) in deionised water (80 mL) was added dropwise to a mixture of phthalic anhydride (7.11 g, 48.0 mmol, 1.2 eq) and anhydrous NaOAc.3H_2O (6.53 g, 48.0 mmol, 1.2 eq) in 4:1 acetone: deionised water mixture (180 mL). The reaction mixture was stirred at RT for 15 mins, followed by the addition of $NaHCO_3$ (5.04 g, 60.0 mmol, 1.5 eq) in small portions over 10 mins. The reaction mixture was stirred at RT under argon overnight. After 16 hours, the reaction mixture was concentrated *in vacuo* to give a crude orange solid which was used directly in the next step without purification. The residue was treated with 1:1 pyridine: acetic anhydride mixture (120 mL), heated to 60 ^oC under reflux for 1 hour until the entire residue had solubilised and the mixture was then left to stir at RT under argon overnight. After 16 hours, the mixture was quenched with EtOH (60 mL) at 0 ^oC for 20 mins, before being concentrated *in vacuo*, by co-evaporation with toluene. The resultant orange oil was and partitioned between EtOAc and water. Then the aqueous layer was extracted with EtOAc (2 x 250 mL) and the combined organic extracts were washed with 1 M aqueous HCl (2 x 100 mL), saturated aqueous sodium hydrogen carbonate (2 x 100 mL) and brine (100 mL). After drying over $MgSO_4$, filtering and concentrating *in vacuo*, the crude material was purified by column chromatography [Pet. Ether/EtOAc (3:2)] to yield the title compound **8** (16.20 g, 86%) as a white solid in the ratio of α/β = 1:1.5. R_f 0.39

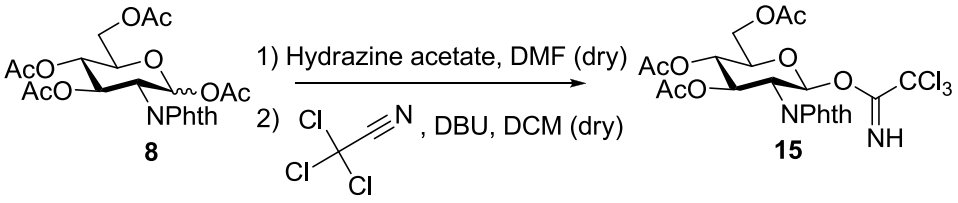
[Pet. Ether/EtOAc (1:1)]; ^{1}H NMR (400 MHz, $CDCl_3$) δ = 7.88-7.84 (m, 4H, C$\underline{H}_{arom.}$ Phth (α+β)), 7.78-7.74 (m, 4H, C$\underline{H}_{arom.}$ Phth (α+β)), 6.57 (1H, app. t, *J* = 9.1 Hz, H-3 (α)), 6.52 (1H, d, $J_{1,2}$ = 8.8 Hz, H-1 (β)), 6.30 (1H, d, $J_{1,2}$ = 3.3 Hz, H-1 (α)), 5.89 (1H, app. t, *J* = 9.3 Hz, H-3 (β)), 5.24-5.15 (2H, m, H-4 (α+β)), 4.74 (1H, dd, $J_{2,3}$ = 11.6, $J_{1,2}$ = 3.5 Hz, H-2 (α)), 4.47 (1H, app. t, *J* = 8.8 Hz, H-2 (β)), 4.37 (2H, dd, $J_{5,6}$ = 12.4, $J_{6,6'}$ = 4.3 Hz, H-6 (α+β)), 4.34-4.30 (1H, m, H-5 (α)), 4.17-4.12 (2H, m, H-6' (α+β)), 4.05-4.01 (m, H-5 (β)), 2.13 (3H, s, C$\underline{H}_3$), 2.09 (3H, s, C$\underline{H}_3$), 2.07 (3H, s, C$\underline{H}_3$), 2.05 (3H, s, C$\underline{H}_3$), 2.00 (3H, s, C$\underline{H}_3$), 1.87 (3H, s, C$\underline{H}_3$); ^{13}C NMR (100 MHz, $CDCl_3$) δ = 134.4, 123.8, 123.7, 90.5 (C-1 (α)), 89.8 (C-1 (β)), 72.6 (C-5 (β)), 70.5 (C-3 (β)), 70.2 (C-5 (α)), 69.4 (C-4 (α)), 68.3 (C-4 (β)), 67.0 (C-3 (α)), 61.5 (C-6 (α+β)), 53.5 (C-2 (β)), 52.8 (C-2 (α)), 20.9, 20.8, 20.7, 20.6, 20.4 (6 x OAc); LRMS m/z (ES+) $[M + Na]^+$ peak 500 (100%).

NMR data: For the purposes of calculating multiplicities of individual proton shifts from integration values, the two anomers were treated independently of one another and the quoted multiplicities do not reflect the α/β ratio.

4.5.9. *O*-[3,4,6-Tri-*O*-acetyl-2-deoxy-2-*N*-phthalimido-β-D-glucopyranosyl] trichloroacetimidate (15)[21]

Formula: $C_{22}H_{21}Cl_3N_2O_{10}$

M_r: 579.77 g mol^{-1}



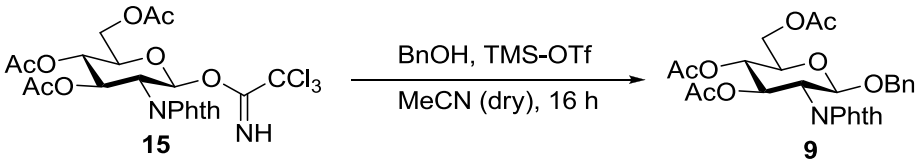
To a stirred solution of 1,3,4,6-tetra-*O*-acetyl-2-deoxy-2-phthalimido-α/β-D-glucopyranoside (19.0 g, 40.0 mmol, 1 eq) **8** in anhydrous dimethylformamide (100 mL), hydrazine acetate (5.38 g, 60.0 mmol, 1.5 eq) was added. The reaction mixture was stirred at RT under argon for 1.5 hours after which the reaction mixture was then poured into a separating funnel, and partitioned between EtOAc and water. Then the aqueous layer was extracted with EtOAc (2 x 250 mL) and the combined organic extracts were washed with 1 M aqueous HCl (200 mL), saturated aqueous sodium hydrogen carbonate (200 mL) and brine (200 mL). After drying over $MgSO_4$, filtering and concentrating *in vacuo*, the crude material was used directly in the next step. To a stirred solution of the residue (15.0 g, 34.0 mmol, 1 eq) in anhydrous DCM (100 mL), trichloroacetonitrile (16.14 mL, 159.2 mmol, 4 eq) and DBU (1.19 mL, 7.96 mmol, 0.2 eq) were added. The reaction mixture was stirred at RT under argon overnight. After 16 hours, TLC analysis [Pet. Ether/EtOAc (2:1)] showed complete consumption of starting material. The reaction was then concentrated *in* vacuo and the resultant yellow foam was purified by column chromatography [Pet. Ether/EtOAc (3:2) + 1% Et_3N] to yield the title compound **15** (8.74 g, 38%) as a yellow solid. R_f 0.46 [Pet. Ether/EtOAc (3:2)]; ^1H NMR (400 MHz, $CDCl_3$)

δ = 8.67 (1H, s, NH), 7.84-7.82 (2H, dd, *J* = 5.3 Hz, *J* = 3.0 Hz, C*H*$_{arom.}$ Phth), 7.73-7.71 (2H, dd, *J* = 5.5 Hz, *J* = 3.0 Hz, C*H*$_{arom.}$ Phth), 6.62 (1H, d, $J_{1,2}$ = 8.8 Hz, H-1), 5.91 (1H, app. t, *J* = 9.4 Hz, H-3), 5.27 (1H, app. t, *J* = 9.6 Hz, H-4), 4.62 (1H, app. t, *J* = 9.1 Hz, H-2), 4.39 (1H, dd, $J_{5,6}$ = 12.4 Hz, $J_{6,6'}$ = 4.3 Hz, H-6), 4.19 (1H, dd, $J_{5,6'}$ = 12.4 Hz, $J_{6,6'}$ = 1.8 Hz, H-6'), 4.09-4.05 (1H, m, H-5), 2.11 (3H, s, C*H*$_3$), 2.06 (3H, s, C*H*$_3$), 1.90 (3H, s, C*H*$_3$); ^{13}C NMR (100 MHz, $CDCl_3$) δ = 134.4 (*C*H$_{arom.}$ Phth), 123.7 (*C*H$_{arom.}$ Phth), 93.6 (C-1), 72.8 (C-5), 70.4 (C-3), 68.4 (C-4), 61.5 (C-6), 55.5 (C-2), 20.7, 20.6, 20.4 (OAc); LRMS m/z (ES+) $[M + Na]^+$ peak 602 (100%).

4.5.10. Benzyl 3,4,6-tri-*O*-acetyl-2-deoxy-2-*N*-phthalimido-β-D-glucopyranoside (9)

Formula: $C_{27}H_{27}NO_{10}$

M_r: 525.50 g mol^{-1}



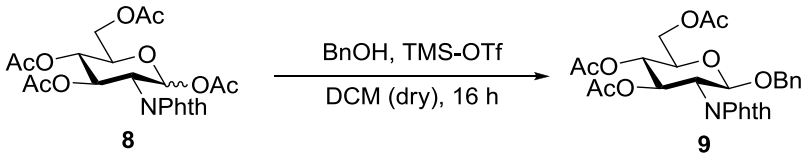
To a stirred solution of 3,4,6-tri-*O*-acetyl-2-deoxy-2-*N*-phthalimido-β-D-glucopyranosyl trichloroacetimidate **15** (7.93 g, 13.7 mmol, 1 eq) in anhydrous acetonitrile (100 mL), benzyl alcohol (2.85 mL, 27.4 mmol, 2 eq) was added followed by the dropwise addition of trimethylsilyl trifluoromethanesulfonate (0.25 mL, 1.37 mmol, 0.1 eq). The reaction mixture was stirred at RT under argon overnight. After 16 hours, TLC analysis [Pet. Ether/EtOAc (3:2)] showed complete consumption of starting material. The reaction mixture was quenched with Et_3N (0.40 mL, 0.2 eq), and concentrated *in vacuo*. The resultant yellow foam was purified by column chromatography [Pet. Ether/EtOAc (3:2)]

to yield the title compound **9** (5.25 g, 74%) as a white solid. R_f 0.42 [Pet. Ether/EtOAc (3:2)]; ^{1}H NMR (400 MHz, $CDCl_3$) δ = 7.78-7.71 (4H, m, C*H*$_{arom.}$ Phth), 7.13-7.06 (5H, m, C*H*$_{arom.}$ benzyl), 5.79 (1H, app. t, *J* = 9.4 Hz, H-3), 5.37 (1H, d, $J_{1,2}$ = 8.6 Hz, H-1), 5.19 (1H, t, *J* = 9.6 Hz, H-4), 4.65 (1H, d, *J* = 12.1 Hz, C*H* benzyl), 4.53 (1H, d, *J* = 12.1 Hz, C*H* benzyl), 4.39-4.33 (2H, m, H-2, H-6'), 4.19 (1H, dd, $J_{5,6'}$ = 12.1 Hz, $J_{6,6'}$ = 2.0 Hz, H-6'), 3.89-3.85 (1H, m, H-5), 2.12 (3H, s, C*H*$_3$), 2.02 (3H, s, C*H*$_3$), 1.85 (3H, s, C*H*$_3$); ^{13}C NMR (100 MHz, $CDCl_3$) δ = 136.6, 134.2 (*C*H$_{arom.}$ Phth), 128.2, 127.7 (*C*H$_{arom.}$ benzyl), 123.5 (*C*H$_{arom.}$ Phth), 97.2 (C-1), 77.1, 76.7, 71.8 (C-5), 71.3 (*C*H benzyl), 70.6 (C-3), 68.9 (C-4), 62.0 (C-6), 54.6 (C-2), 20.8, 20.6, 20.5 (3 x OAc); LRMS m/z (ES+) $[M + Na]^+$ peak 548 (100%).

4.5.11. Benzyl 3,4,6-tri-*O*-acetyl-2-deoxy-2-*N*-phthalimido-β-D-glucopyranoside (9)[22]

Formula: $C_{27}H_{27}NO_{10}$

M_r: 525.50 g mol^{-1}



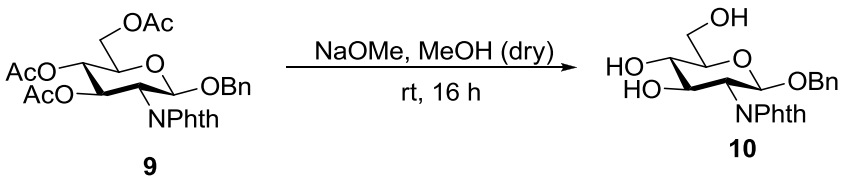
To a stirred solution of 1,3,4,6-tetra-*O*-acetyl-2-deoxy-2-*N*-phthalimido-D-glucopyranoside **8** (14.0 g, 29.7 mmol, 1 eq) and benzyl alcohol (6.19 mL, 59.7 mmol, 2 eq) in anhydrous dichloromethane (100 mL), trimethylsilyl trifluoromethanesulfonate (5.31 mL, 29.7 mmol, 1 eq) was added by careful dropwise addition. The reaction mixture was stirred at RT under argon overnight. After 16 hours, TLC analysis [Pet. Ether/EtOAc (3:2)] showed complete consumption of starting material. The reaction mixture was quenched with Et_3N (4.14 mL, 1 eq) and concentrated *in vacuo*. The resultant orange oil was then partitioned between DCM and water. Then the aqueous

layer was extracted with DCM (2 x 250 mL) and the combined organic extracts were washed with 1 M aqueous HCl (2 x 100 mL), saturated aqueous sodium hydrogen carbonate (2 x 100 mL) and brine (100 mL). After drying over $MgSO_4$, filtering and concentrating *in vacuo*, the crude material was purified by column chromatography [Pet. Ether/EtOAc (3:2)] to yield the title compound **9** (9.94 g, 64%) as a white solid. R_f 0.41 [Pet. Ether/EtOAc (3:2)]; ^{1}H NMR (400 MHz, $CDCl_3$) δ = 7.78-7.72 (4H, m, C*H*$_{arom.}$ Phth), 7.13-7.08 (5H, m, C*H*$_{arom.}$ benzyl), 5.80 (1H, app.t, *J* = 9.1 Hz, H-3), 5.38 (1H, d, $J_{1,2}$ = 8.3 Hz, H-1), 5.20 (1H, t, *J* = 9.7 Hz, H-4), 4.86 (1H, d, *J* = 12.1 Hz, C*H* benzyl), 4.55 (1H, d, *J* = 12.1 Hz, C*H* benzyl), 4.39-4.34 (2H, m, H-2, H-6), 4.21 (1H, dd, $J_{5,6'}$ = 12.1 Hz, $J_{6,6'}$ = 2.3 Hz, H-6'), 3.90-3.85 (1H, m, H-5), 2.14 (3H, s, C*H*$_3$), 2.03 (3H, s, C*H*$_3$), 1.86 (3H, s, C*H*$_3$); ^{13}C NMR (100 MHz, $CDCl_3$) δ = 170.1, 169.5, 136.6, 134.1 (*C*H$_{arom.}$ Phth), 128.2, 127.8 (*C*H$_{arom.}$ benzyl), 123.5 (*C*H$_{arom.}$ Phth), 97.2 (C-1), 77.1, 76.7, 71.8 (C-5), 71.3 (*C*H benzyl), 70.7 (C-3), 69.0 (C-4), 62.0 (C-6), 54.6 (C-2), 20.8, 20.6, 20.5 (3 x OAc); LRMS m/z (ES+) $[M + Na]^+$ peak 548 (100%).

4.5.12. Benzyl 2-deoxy-2-*N*-phthalimido-β-D-glucopyranoside (10)

Formula: $C_{21}H_{21}NO_7$

M_r: 399.39 g mol^{-1}



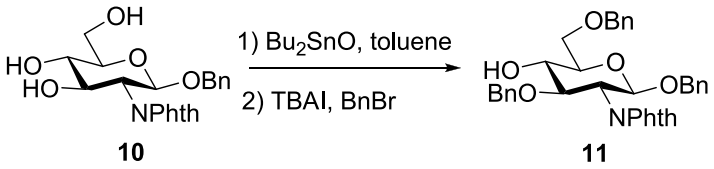
9 **10**

To a stirred solution of benzyl 3,4,6-tri-*O*-acetyl-2-deoxy-2-*N*-phthalimido-β-D-glucopyranoside **9** (4.86 g, 92.5 mmol, 1 eq) in anhydrous methanol (100 mL), sodium methoxide (0.05 g, 0.93 mmol, 0.1 eq) was added. The reaction mixture was stirred at RT under argon overnight. After 16 hours, TLC analysis [DCM/MeOH (9:1)] showed complete consumption of starting material. The reaction was then neutralised with acidified Dowex, filtered and concentrated *in vacuo* to give product **10** (3.62 g, 98%) as a white crystalline solid. R_f 0.27 [[DCM/MeOH (9:1)]; ^{1}H NMR (400 MHz, MeOD) δ = 5.17 (1H, d, $J_{1,2}$ = 8.3 Hz, H-1), 4.80 (1H, d, *J* = 12.3 Hz, C*H* benzyl), 4.55 (1H, d, *J* = 12.1 Hz, C*H* benzyl), 4.27 (1H, dd, *J* = 10.6 Hz, *J* = 8.1 Hz, H-3), 4.04 (1H, dd, *J* = 10.6 Hz, *J* = 8.6 Hz, H-2), 3.98 (1H, dd, $J_{5,6}$ = 11.6 Hz, $J_{6,6'}$ = 1.4 Hz, H-6), 3.78 (1H, dd, $J_{5,6'}$ = 11.8 Hz, $J_{6,6'}$ = 5.3 Hz, H-6'), 3.47-3.41 (2H, m, H-4, H-5); ^{13}C NMR (100 MHz, $CDCl_3$) δ = 166.7, 137.7 (*C*$H_{arom.}$ Phth), 134.4, 128.2, 127.9, 127.7 (*C*$H_{arom.}$ benzyl), 123.2, 97.7 (C-1), 77.3 (C-4), 71.6 (C-5), 71.4 (C-3), 70.9 (*C*H benzyl), 61.8 (C-6), 57.6 (C-2); LRMS m/z (ES+) $[M + Na]^+$ peak 422 (100%).

4.5.13. 1,3,6-*O*-Benzyl-2-deoxy-2-*N*-phthalimido-β-D-glucopyranoside (11)[23]

Formula: $C_{35}H_{33}NO_7$

M_r: 579.64 g mol^{-1}



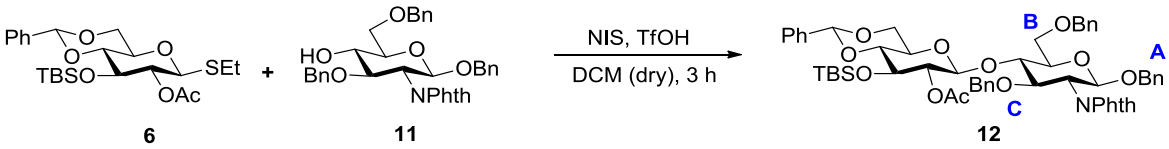
10 **11**

To a stirred solution of benzyl 2-deoxy-2-*N*-phthalimido-β-D-glucopyranoside **10** (3.30 g, 8.26 mmol, 1 eq) in toluene (70 mL), dibutyltinoxide (4.52 g, 18.2 mmol, 2.2 eq) was added and the resulting reaction mixture was heated under reflux (Dean-Stark apparatus). After 16 hours, tetrabutylammonium iodide (6.71 g, 18.2 mmol, 2.2 eq) and benzyl bromide (2.16 mL, 18.2 mmol, 2.2 eq) were added and the mixture was gently refluxed. After 4 hours, the resulting mixture was cooled, filtered and concentrated *in vacuo*. The resulting residue was purified by column chromatography [Pet. Ether/EtOAc (4:1)] to yield the title compound **11** (2.10 g, 45%) as a white solid. R_f 0.51 [Pet. Ether/EtOAc (3:2)]; ^{1}H NMR (400 MHz, $CDCl_3$) δ = 7.81-7.70 (4H, m, C$\underline{H}_{arom.}$ Phth), 7.38-7.28 (5H, m, C$\underline{H}_{arom.}$ benzyl), 7.12-6.94 (10H, m, C$\underline{H}_{arom.}$ benzyl), 5.16 (1H, d, $J_{1,2}$ = 7.8 Hz, H-1), 4.87 (1H, d, *J* = 12.4 Hz, C$\underline{H}$ benzyl B), 4.79 (1H, d, *J* = 12.4 Hz, C$\underline{H}$ benzyl A), 4.74 (1H, d, *J* = 12.6 Hz, C$\underline{H}$ benzyl C), 4.67 (1H, d, *J* = 12.1 Hz, C$\underline{H}$ benzyl B), 4.61 (1H, d, *J* = 11.9 Hz, C$\underline{H}$ benzyl A), 4.49 (1H, d, *J* = 12.4 Hz, C$\underline{H}$ benzyl C), 4.25-4.23 (2H, m, H-2, H-3), 3.88-3.80 (3H, m, H-4, H-6, H-6'), 3.66-3.63 (1H, m, H-5), 2.91 (1H, d, *J* = 2.5 Hz, H-4(OH)); LRMS m/z (ES+) $[M + Na]^+$ peak 602 (100%).

4.5.14. Benzyl *O*-(2-*O*-acetyl-4,6-*O*-benzylidene-3-*O*-*tert*-butyldimethylsilyl-β-D-glucopyranosyl)-(1→4)-3,6-di-*O*-benzyl-2-deoxy-2-phthalamido-β-D-glucopyranoside (12)[9]

Formula: $C_{56}H_{63}NO_{13}Si$

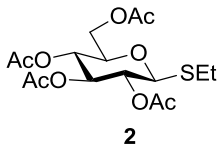
M_r: 986.18 g mol^{-1}



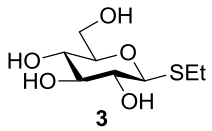
A suspension of ethyl 2-O-acetyl-4,6-O-benzylidene-3-O-(tert-butyldimethylsilyl)-1-thio-β-D-glucopyranoside **6** (178 mg, 0.38 mmol, 1.1 eq) and 1,3,6-O-benzyl-2-deoxy-2-*N*-phthalimido-β-D-glucopyranoside **11** (200 mg, 0.35 mmol, 1 eq) containing activated 4 Å molecular sieves (200 mg) in anhydrous dichloromethane (5 mL) was stirred under argon for 1 hour at RT. After cooling to -40 °C, *N*-iodosuccinimide (116 mg, 0.52 mmol, 1.5 eq) and trifluoromethanesulfonic *acid* (3 µL, 0.04 mmol, 0.1 eq) were added and the mixture was stirred at -40 °C to RT for 3 hours. The resulting purple reaction mixture was quenched with Et_3N, filtered through Celite and then washed with $Na_2S_2O_3$. After drying over $MgSO_4$, filtering and concentrating in vacuo the crude orange oil was purified by column chromatography [Pet. Ether/EtOAc (4:1)] to yield the title compound **12** (207 mg, 61%) as a white solid. R_f 0.68 [Pet. Ether/EtOAc (3:2)]; ^{1}H NMR (500 MHz, $CDCl_3$) δ = 7.78-7.71 (4H, m, C$\underline{H}_{arom.}$ Phth), 7.13-7.06 (5H, m, C$\underline{H}_{arom.}$ benzyl), 5.41 (1H, s, C$\underline{H}$ benzylidene), 5.11 (1H, app. d, $J_{1,2}$ = 8.5 Hz, H-1), 4.93 (1H, app.t, J = 8.8 Hz, H-2'), 4.85 (1H, d, J = 12.3 Hz, C$\underline{H}$ benzyl B), 4.79 (1H, d, J = 12.3 Hz, C$\underline{H}$ benzyl A), 4.74 (1H, d, J = 12.6 Hz, C$\underline{H}$ benzyl C), 4.57 (1H, d, $J_{1',2'}$ = 8.2 Hz, H-1'), 4.51 (1H, d, J = 12.0 Hz, C$\underline{H}$ benzyl B), 4.47 (1H, d, J = 12.3 Hz, C$\underline{H}$ benzyl A), 4.42 (1H, d, J = 12.3

Hz, C<u>H</u> benzyl C), 4.26 (1H, dd, $J_{5',6''}$ = 10.4 Hz, $J_{6',6''}$ = 5.0 Hz, H-6_a''), 4.20 (2H, app. d, J = 8.5 Hz, H-2, H-3), 4.10-4.06 (1H, m, H-4), 3.83 (1H, dd, $J_{5,6'}$ = 10.4 Hz, $J_{6,6'}$ = 5.0 Hz, H-6_{a+b}'), 3.69 (1H, t, J = 8.8 Hz, H-3'), 3.51-3.40 (3H, m, H-5, H-6_b', H-4'), 3.22-3.17 (1H, m, H-5'), 2.07 (3H, s, OC<u>H</u>$_3$), 0.81 (9H, s, (C<u>H</u>$_3$)$_3$), 0.03 (3H, s, SiC<u>H</u>$_3$), -0.03 (3H, s, SiC<u>H</u>$_3$); ^{13}C NMR (125 MHz, $CDCl_3$) δ = 169.0, 138.5, 137.9, 137.2, 137.1, 133.5, 131.6, 129.0, 128.5, 128.1, 128.0, 127.9, 127.8, 127.6, 127.5, 127.0, 126.2, 123.2, 101.8, 100.6, 97.3, 81.5, 77.2, 74.9, 74.8, 74.5, 73.7, 72.8, 70.7, 68.6, 67.6, 65.9, 55.6, 53.4, 25.5, 21.1, 17.9, 0.9, -4.2, -5.0; LRMS m/z (ES+) $[M + Na]^+$ peak 1008 (100%), $[2M + Na]^+$ peak 1995 (50%),

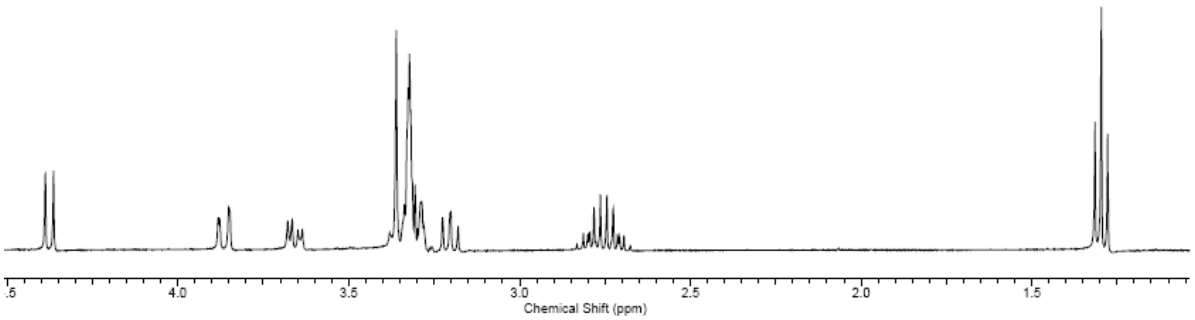
4.6. ^{1}H NMR Spectra

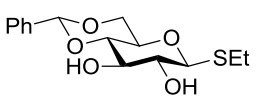


2

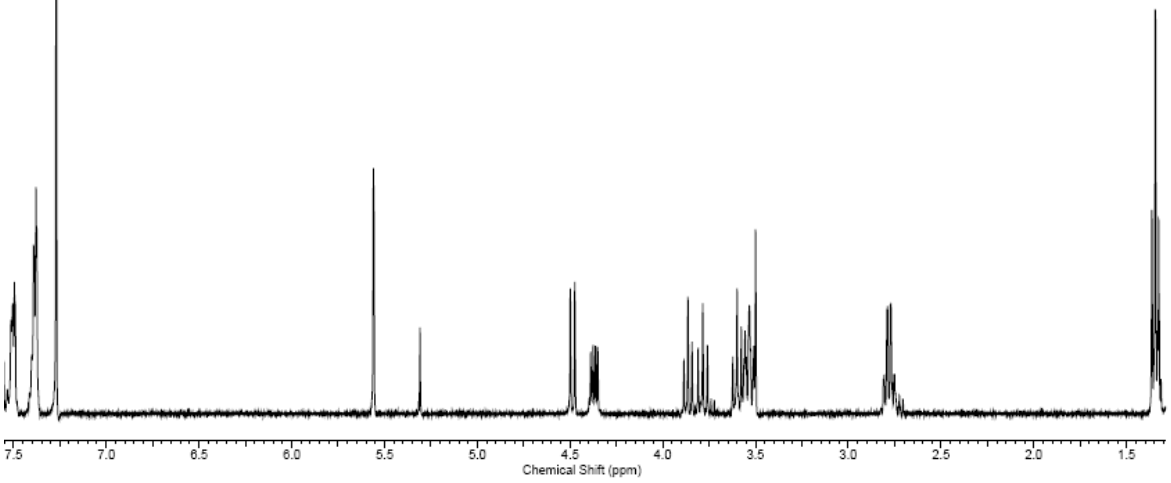


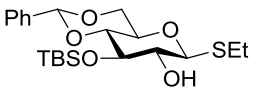
3




Ph
O
O
O
HO
SEt
OH

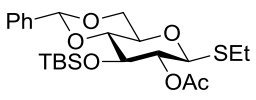
4


7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
Chemical Shift (ppm)


Ph
O
O
O
TBSO
SEt
OH

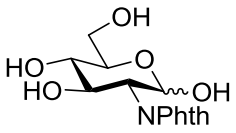
5

7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0


Ph
O
O
O
TBSO
SEt
OAc

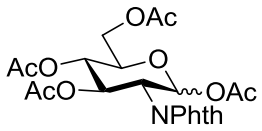
6

7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0
Chemical Shift (ppm)


OH
HO
O
HO
OH
NPhth

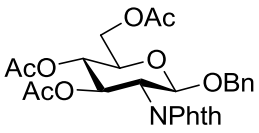
7

7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
Chemical Shift (ppm)


OAc
O
AcO
AcO
OAc
NPhth

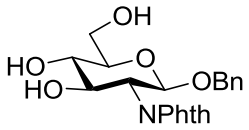
8

7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
Chemical Shift (ppm)


OAc
O
AcO
AcO
OBn
NPhth

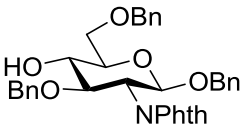
9

7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
Chemical Shift (ppm)


OH
HO
O
HO
OBn
NPhth

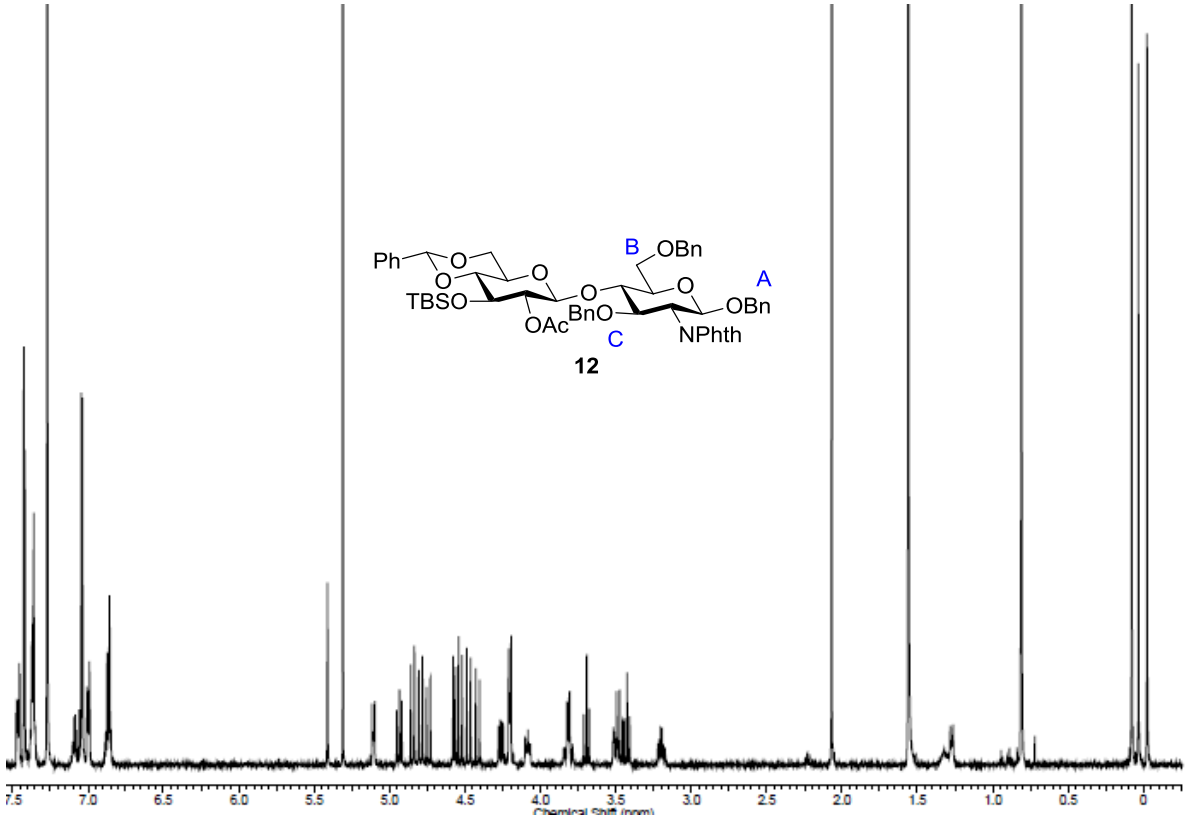
10

8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
Chemical Shift (ppm)


OBn
HO
O
BnO
OBn
NPhth

11

8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5


Ph
O
O
O
TBSO
OAc
O
BnO
B
OBn
O
A
OBn
C
NPhth
12
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0

4.7. References

1. Schwarz, F.; Huang, W.; Li, C. S.; Schulz, B. L.; Lizak, C.; Palumbo, A.; Numao, S.; Neri, D.; Aebi, M.; Wang, L. X., A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. *Nature Chemical Biology* **2010,** 6, (4), 264-266.

2. Takegawa, K.; Yamabe, K.; Fujita, K.; Tabuchi, M.; Mita, M.; Izu, H.; Watanabe, A.; Asada, Y.; Sano, M.; Kondo, A.; Kato, I.; Iwahara, S., Cloning, sequencing, and expression of Arthrobacter protaphormiae endo-beta-N-acetylglucosaminidase in Escherichia coli. *Archives of Biochemistry and Biophysics* **1997,** 338, (1), 22-28.

3. Fujita, K.; Takegawa, K., Chemoenzymatic synthesis of neoglycoproteins using transglycosylation with endo-beta-N-acetylglucosaminidase A. *Biochemical and Biophysical Research Communications* **2001,** 282, (3), 678-682.

4. Takegawa, K.; Tabuchi, M.; Yamaguchi, S.; Kondo, A.; Kato, I.; Iwahara, S., Synthesis of Neoglycoproteins Using Oligosaccharide-Transfer Activity with Endo-Beta-N-Acetylglucosaminidase. *Journal of Biological Chemistry* **1995,** 270, (7), 3094-3099.

5. Wang, L. X.; Huang, W., Enzymatic transglycosylation for glycoconjugate synthesis. *Current Opinion in Chemical Biology* **2009,** 13, (5-6), 592-600.

6. Wang, L. X., Chemoenzymatic synthesis of glycopeptides and glycoproteins through endoglycosidase-catalyzed transglycosylation. *Carbohydrate Research* **2008,** 343, (10-11), 1509-1522.

7. Takegawa, K.; Yamaguchi, S.; Kondo, A.; Iwamoto, H.; Nakoshi, M.; Kato, I.; Iwahara, S., Transglycosylation Activity of Endo-Beta-N-Acetylglucosaminidase from Arthrobacter-Protophormiae. *Biochemistry International* **1991,** 24, (5), 849-855.

8. Fernandez-Gonzalez, M.; Boutureira, O.; Bernardes, G. J. L.; Chalker, J. M.; Young, M. A.; Errey, J. C.; Davis, B. G., Site-selective chemoenzymatic construction of synthetic glycoproteins using endoglycosidases. *Chemical Science* **2010**.

9. Li, B.; Zeng, Y.; Hauser, S.; Song, H. J.; Wang, L. X., Highly efficient endoglycosidase-catalyzed synthesis of glycopeptides using oligosaccharide oxazolines as donor substrates. *Journal of the American Chemical Society* **2005,** 127, (27), 9692-9693.

10. Li, B.; Song, H. J.; Hauser, S.; Wang, L. X., A highly efficient chemoenzymatic approach toward glycoprotein synthesis. *Organic Letters* **2006,** 8, (14), 3081-3084.

11. Owens, R. J.; Chang, V. T.; Crispin, M.; Aricescu, A. R.; Harvey, D. J.; Nettleship, J. E.; Fennelly, J. A.; Yu, C.; Boles, K. S.; Evans, E. J.; Stuart, D. I.; Dwek, R. A.; Jones, E. Y.; Davis, S. J., Glycoprotein structural genomics: Solving the glycosylation problem. *Structure* **2007,** 15, (3), 267-273.

12. Elbein, A. D.; Tropea, J. E.; Mitchell, M.; Kaushal, G. P., Kifunensine, a Potent Inhibitor of the Glycoprotein Processing Mannosidase-I. *Journal of Biological Chemistry* **1990,** 265, (26), 15599-15605.

13. Fairbanks, A. J.; Heidecke, C. D.; Parsons, T. B., Endohexosaminidase-catalysed glycosylation with oxazoline donors: effects of organic co-solvent and pH on reactions catalysed by Endo A and Endo M. *Carbohydrate Research* **2009,** 344, (18), 2433-2438.

14. Autar, R.; Khan, A. S.; Schad, M.; Hacker, J.; Liskamp, R. M. J.; Pieters, R. J., Adhesion inhibition of F1C-fimbriated Escherichia coli and Pseudomonas aeruginosa PAK and PAO by multivalent carbohydrate ligands. *Chembiochem* **2003,** 4, (12), 1317-1325.

15. Houdier, S.; Vottero, P. J. A., Synthesis of Benzylated Cycloisomaltotetraose. *Carbohydrate Research* **1993,** 248, 377-384.

16. Verduyn, R.; Douwes, M.; Vanderklein, P. A. M.; Mosinger, E. M.; Vandermarel, G. A.; Vanboom, J. H., Synthesis of a Methyl Heptaglucoside - Analog of the Phytoalexin Elicitor from Phytophtora-Megasperma. *Tetrahedron* **1993,** 49, (33), 7301-7316.

17. Wang, Z., Zemplén Deacetylation. *Comprehensive Organic Name Reactions and Reagents.* **2010**, 3123-3128.

18. Mo, K. F.; Li, H. Q.; Mague, J. T.; Ensley, H. E., Synthesis of the beta-1,3-glucan, laminarahexaose: NMR and conformational studies. *Carbohydrate Research* **2009,** 344, (4), 439-447.

19. Davis, B. G., Fairbanks, A. J., Preparation of thiosugars and their use. *United States Patent Application Publication* **2009**, US 2009/0176970.

20. Issaree, A., Synthesis of Hetero-chitooligosaccharides. *DPhil Thesis* **2008**, 63-64.

21. Grundler, G.; Schmidt, R. R., Application of Trichloracetamide Processes to 2-Deoxy-2-Phthalimido-D-Glucose Derivatives - Synthesis of Oligosaccharides of the Core Region of Ortho-Glycoproteins of Mucine Type. *Carbohydrate Research* **1985,** 135, (2), 203-218.

22. Lafont, D.; Boullanger, P.; Fenet, B., Comparative Reactivity of Various D-Glucosamine Acceptors during Synthesis of Chitobiose Precursors. *Journal of Carbohydrate Chemistry* **1994,** 13, (4), 565-583.

23. Aly, M. R. E.; Castro-Palomino, J. C.; Ibrahim, E. S. I.; El-Ashry, E. S. H.; Schmidt, R. R., The dimethylmaleoyl group as amino protective group - Application to the synthesis of glucosamine-containing oligosaccharides. *European Journal of Organic Chemistry* **1998**, (11), 2305-2316.

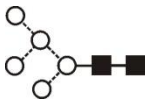
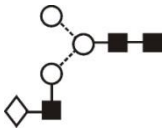
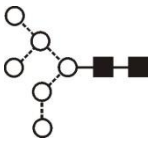
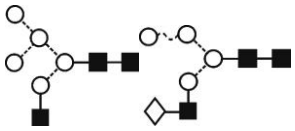
Chapter 5

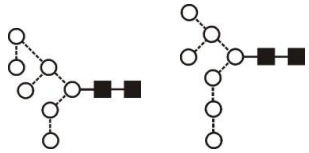
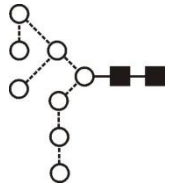
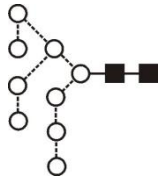
Appendix

5.1. *N*-linked Glycan types with susceptibility to Endo H

Mass and composition of the *N*-linked glycans susceptible to Endo H, as obtained after desialylation with 1% AcOH solution at 90°C for one hour:

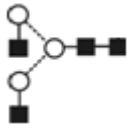
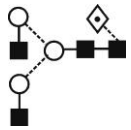
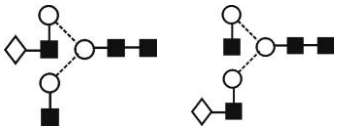
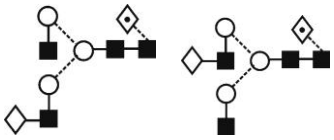
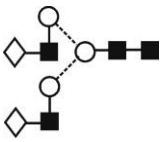
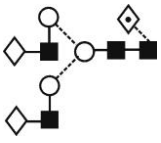
Key: ■ = GLcNAc; ○ = Man; ◇ = Gal; ◆ = GalNAc ; ⟐ = Fuc

MALDI m/z ([M+Na]+)	Composition			Proposed structure
	Hex	HexNAc	dHex	
1095.1	4	2	0	
1257.1	5	2	0	
1298.0	4	3	0	
1419.0	6	2	0	
1460.0	5	3	0	

1581.0	7	2	0	
1743.0	8	2	0	
1904.9	9	2	0	

5.2. *N*-linked Glycan types with susceptibility to Endo S

Mass and composition of the *N*-linked glycans susceptible to Endo S, as obtained after desialylation with 1% AcOH solution at 90°C for one hour:

MALDI m/z ([M+Na]+)	Composition			Proposed structure
	Hex	HexNAc	dHex	
1339.0	3	4	0	
1485.0	3	4	1	
1501.0	4	4	0	
1647.0	4	4	1	
1664.0	5	4	0	
1809.0	5	4	1	

An Endoglycosidase with Reciprocal Glycan Specificity allows Broadened Glycoprotein Remodelling

Jonathan J. Goodfellow,‡ Keisuke Yamamoto,‡ Kavitha XXXX∫, Camille XXXX∫, Max XXXX∫, Christoper N. Scanlan∫ and Benjamin G. Davis‡*

‡Department of Chemistry, University of Oxford, Chemistry Research Laboratory, Mansfield Road, Oxford OX1 3TA, UK, Ben. Davis @chem.ox.ac.uk

∫Department of Biochemistry, University of Oxford

Endo-S, glycoproteins, N-linked glycans, glycosidase

Supporting Information Placeholder

ABSTRACT: Protein endoglycosidases ('Endos') are useful tools in the biocatalytic alteration of glycans on protein surfaces. However, the currently limited selectivity of Endos has prevented effective manipulation of certain *N*-linked glycans widely found in nature. Here, we reveal that a bacterial endoglycosidase from *Streptococcus pyogenes*, EndoS, is reciprocal to that of other known endoglycosidases (EndoA, EndoH) used for current protein remodeling. It allows processing of complex type *N*-linked glycans +/- core-fucosylation but does not process oligo- and high-mannose glycans. This biocatalytic activity now allows processing of previously refractory glycoprotein samples including eukaryotic proteins and from human sources.

Glycan display is critical for development and physiology of many living systems.[1] Protein glycosylation, in particular is a diverse form of posttranslational modification comprising 50% of cellular proteome and 90% of secreted proteome.[2,3] 70% of the therapeutic proteins that are currently in clinical trials are glycoproteins,[4] typically with the presence of the so-called 'core fucosylation' motif (Figure 1). Natural glycoproteins exist as mixtures of glycoforms – same peptide backbone but different glycosylation pattern and site – making isolation of well-defined glycoproteins complicated.[5] This results in a lack of detailed structural and functional studies on the specific role of glycans. Moreover, since certain glycoforms are more active than others,[6] access to specific uniform glycoforms can confer significant therapeutic advantages. In the case of antibodies, the alteration of IgG-Fc glycosylation leads to modulation of binding to cellular Fcγ receptors (FcγRs) thus tuning antibody-dependant cellular cytoxicity (ADCC)[7] and determines anti-inflammatory activity.[8]

Recombinant expression systems, chemoenzymatic and chemical methods have all emerged as techniques that can resolve the supply issues for some homogeneous synthetic glycoproteins.[9-14] One powerful approach is to remodel a glycoprotein enzymatically: the initial heterogeneous glycoform mixture is treated with an endoglycosidase ("Endo") to trim off the variable portions of the attached oligosaccharides attached to the OH-4 of the first *N*-linked residue, *N*-acetyl glucosamine (GlcNAc), found at the *N*-linked glycosylation site(s) (Figure 1b), thereby yielding a reduced number of protein glycoforms. Subsequent enzyme-mediated glycosylation of the resulting exposed GlcNAc residue can then produce a more homogenous sample of a desired glycoprotein.[12,15-18] However, the Endo enzymes currently used to attach glycans to proteins in this way show restricted sugar selectivity that prevents such combined removal and attachment of the complex-type glycans normally found on mammalian proteins or glycans that contain 'core-fucose'.[19,20,21]

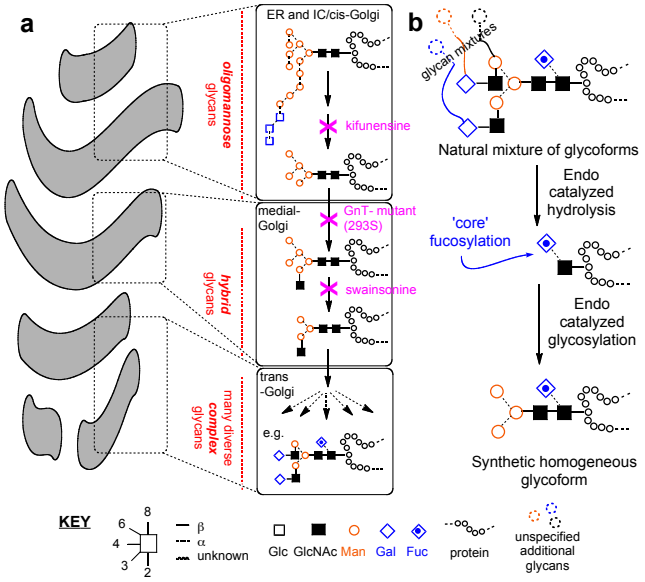


Figure 1. (a) Intracellular *N*-linked protein glycan alteration pathway following attachment to the protein and corresponding methods (magenta) to access truncated glycan structures through inhibited pathways. (b) Glycan remodeling using Endo-type enzymes (shown here for the idealized remodeling of complex-type glycoproteins).

In an attempt to identify new activities we surveyed all known and putative Endo (E.C. 3.2.1.96) orfs (putative *ndo* genes). Whilst those in the CAZy[22] glycoside hydrolase family GH85 are most commonly used in current methods, other important potentially useful orfs exist in other families, implying possibly different substrate specificities. The *ndoS* gene from *Streptococcus pyogenes* belongs to the GH18 family and codes for the production of an apparently immunomodulatory evasion factor that prolongs bacterial

survival and acts upon four subclasses of IgG.[23,24] Codon-optimized *ndoS* gene was generated (see SI) and subligated into a pGEX-4T-1 vector which allowed efficient (~13 mg/L) heterologous expression of EndoS enzyme in an *N*-terminally-GST-tagged form using *E. coli* BL21(DE3). The introduction of a thrombin-cleavable site between tag and the rest of the protein sequence allowed EndoS to be produced in both tagged and untagged form (see SI). Both forms were highly active (*vide infra*).

To evaluate the comparative ability of Endo enzymes to process important and relevant targets with relevant glycosylation patterns, we restricted our analyses to glycoprotein substrates produced by mammalian (largely human) expression. First, human IgG (see SI, Figure SXX- GEL SHIFT AND MS) was incubated with EndoS as well as with widely-used Endo enzymes EndoA (family GH85)[25] and EndoH (family GH85)[26] at 37°C for 1h in PBS (pH 7.4, 10 mM). Electrophoretic and spectrometric analysis (reducing SDS-PAGE and MALDI-MS) revealed a complete absence of activity by EndoA and EndoH towards IgG. In striking contrast, highly efficient processing of IgG was seen using EndoS; glycans were removed from the heavy chain even at very low concentration (< 50 nM, $k_{app} > 0.8$ s^{-1}). The IgG glycans released by all the Endo enzymes were analysed in detail (Figure 2) and revealed clean release of the so-called G0, G1 and G2 glycans[27] by EndoS and confirmed a complete lack of glycan cleavage by EndoA and EndoH (Figure SXX).

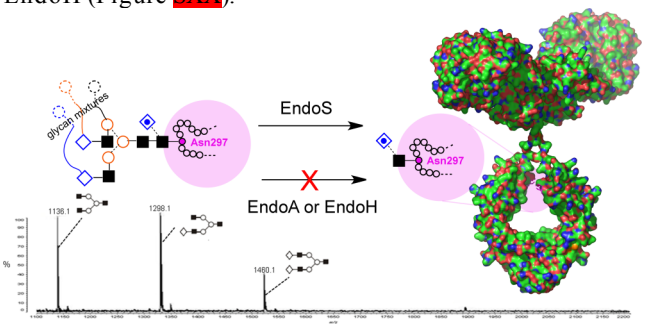


Figure 2. Trimming of glycans on human IgG by EndoS but not by EndoA or EndoH. MALDI-TOF MS analysis (lower pane) of released glycans confirms unique processing of complex-type (see also Figure SXX, SI).

Next, having confirmed EndoS's unique ability to process IgG glycans, we tested it in the remodeling of sugars from other important recombinant proteins. We used the viral protein gp120 (found in the envelope of HIV) as its Bal sequence variant ($gp120_{BAL}$).{ref TO PRIMARY REF FOR SEQUENCE OR GENBANK ACCESSION} $gp120_{BAL}$ and its glycosylation has been the focus of intensive research associated with elucidation of the mechanism of HIV and with the development of associated vaccines.[28] Indeed, remodeling of gp120 has been proposed as a suggested method of vaccine development but has not yet been achieved.[16] To create diversity in displayed glycans we expressed this protein in human (HEK293) cells under a variety of conditions using both wild-type (WT) cell lines, mutant cell lines deficient in certain glycosyltransferases and also cell-permeable inhibitors to prevent glycan processing (see inhibited pathways shown in Figure 1a).[29]

With a focus on glycan patterns that have not yet been amenable to glycoprotein remodeling we created proteins displaying the highly heterogenous glycans found in the normal mammalian *N*-linked glycosylation pathway (using

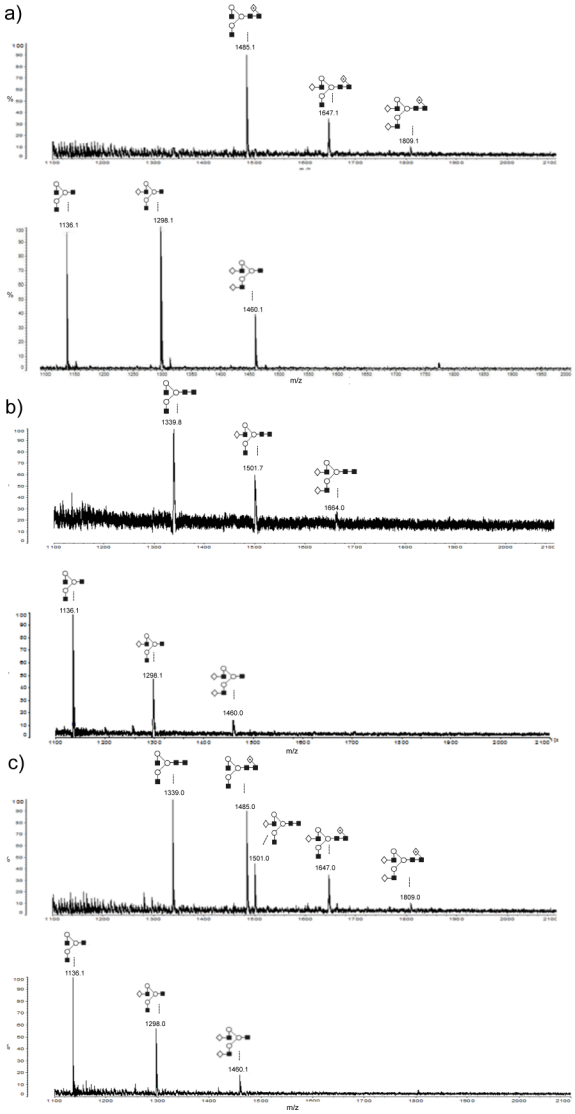


Figure 4. Plastic processing of both (a) fucosylated and (b) non-fucosylated glycans with similar efficiency; mixed samples showed cleavage at similar rates and high efficiencies (c).

HEK293T, Figure 3c), those displaying primarily $Man_9GlcNAc_2$ glycans (see SI, using HEK293T treated with ER and Golgi α-mannosidase inhibitor kifunensine,[30] Figure 3a), and those displaying $Man_5GlcNAc_2$ (using HEK293S cells that lack the GnT-I glycosyltransferase, Figure 3b). Protein glycans (see Figure 3 and SI) displayed strikingly different behaviours when treated with either EndoH or EndoS: EndoH was capable of cleaving only high- and oligo-mannose glycans $Man_{5-9}GlcNAc_2$ within all three glycan distributions (Figure 3a-c), which were leftuntouched by EndoS. Conversely, within the heterogeneous glycans produced in the wild-type glycoprotein pathway (Figure 3c) EndoS was uniquely capable of processing complex-type glycans.

Next, we tested the ability of EndoS to process other glycan types that are not processed by known Endo enzymes. Core fucosylation can play a key role in modulating effector function[7] and is prevalent in both human and therapeutic antibodies. IgG-Fc of Cetuximab, for example contains primarily fucosylated biantennary complex type *N*-glycans.[31] IgG-Fc was transiently expressed (see SI) in wild-type human (HEK293T) cells and the processing of their glycans

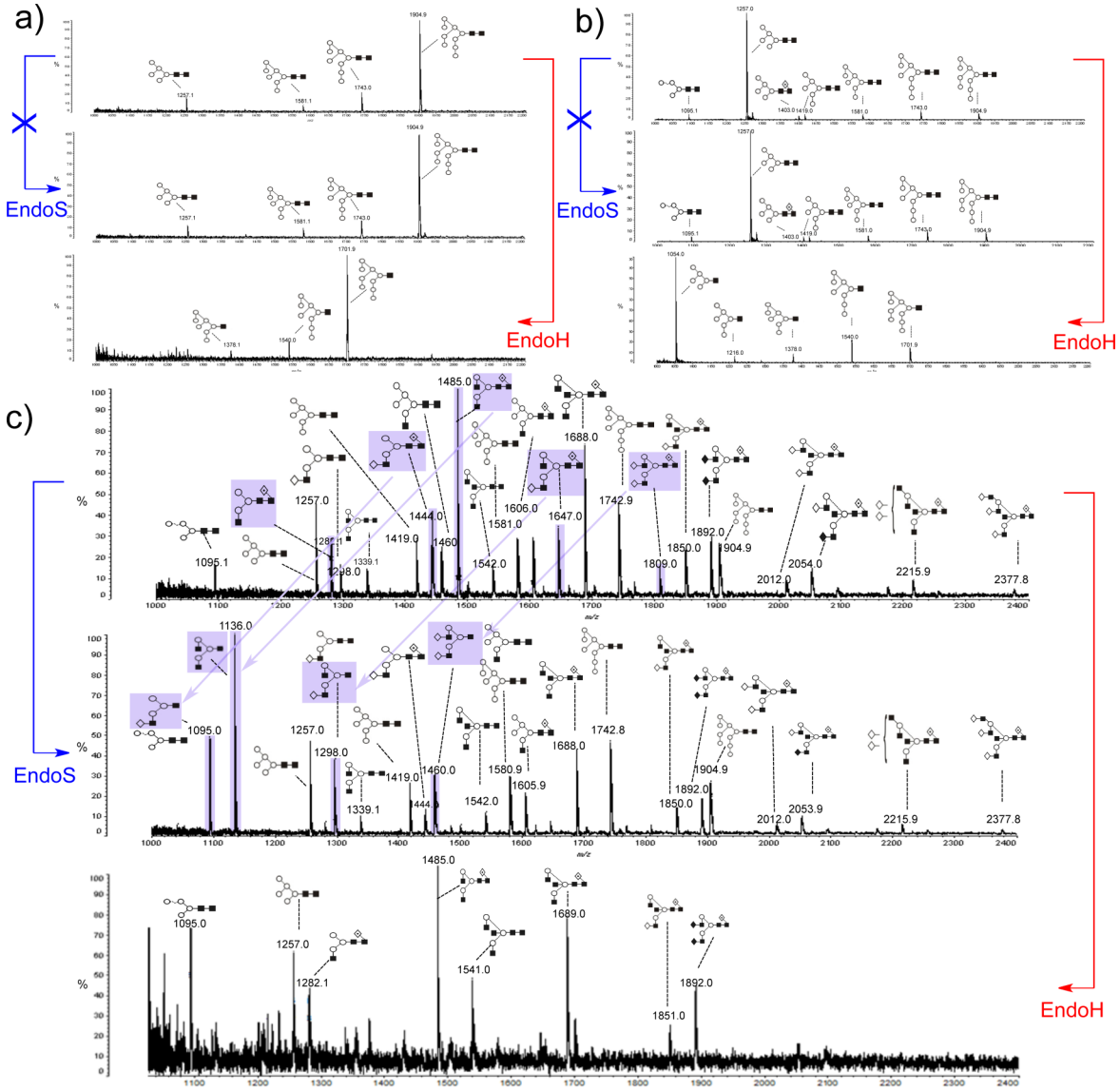


Figure 3. Reciprocal processing of protein glycans with distributions from different protein glycosylation pathways (see Fig 1a for pathway manipulation strategy). (a) Oligomannose $Man_9GlcNAc_2$ (some $Man_{5\text{-}8}GlcNAc_2$) processed by EndoH but not EndoS. (b) $Man_5GlcNAc_2$ (some $Man_{6\text{-}9}GlcNAc_2$) processed by EndoH but not EndoS. (c) In heterogeneous wild-type pathway protein glycans EndoS processes complex-type whilst EndoH processes oligomannose (see also Table SXX in SI for summary).

in fucosylated (Figure 4a), non-fucosylated (Figure 4b) and mixed fucosylated states (Figure 4c) by EndoS was explored. These revealed that unlike previously known Endo enzymes, EndoS shows complete plasticity for the presence of the 'core' α(1,6)-linked Fuc residue. These are the first examples of such usefully broad sugar-processing activity.

In this way, we demonstrated here that the EndoS system has, by virtue of its unique selectivity, the potential to remodel (Figure 1b) those proteins produced in mammalian eukaryotic systems and hence current therapeutics. To demonstrate this unique potential, we remodeled the heterogenous glycoforms (Fig 5b) of a human IgG (Fig 5).

Endo-β-*N*-acetylglucosaminidases hydrolyze the glycosidic bond in the *N*,*N*'-diacetylchitobiose core of *N*-linked glycans. Some of them, such as Endo-A[25] and Endo-M[32] also possess significant transglycosylation activity and have been used, for example, to remodel the *N*-glycoprotein RNaseB.[33,34] Importantly, the efficiency of glycosylation has been valuably improved[35-40] by the use of sugar oxazolines as donor substrates and mutant Endo enzymes. Recently, the re-engineering of bacterial glycosylation[18] or use of chemical attachment[41] has allowed the repositioning of the first sugar residue required for subsequent Endo-mediated elaboration, thereby allowing additional control of glycosylation site.

Notably, previous remodeling of yeast produced Fcs has only proved possible by virtue of their non-fucosylated, high/oligo-mannose glycoform content.[31] Moreover, other approaches using Endos have proven difficult; using an alternative system based on highly innovative prokaryotic protein production only yields of less than 5% were reported.[18] Activated glycan donor tetrasaccharide oxazoline **1** was synthesized (see SI) and after EndoS trimming was used in EndoS-catalyzed glycosylation of IgG to yield essentially a single glycoform (>95%). Remodeling of glycoproteins directly from mammalian eukaryotic sources has not been possible until now.

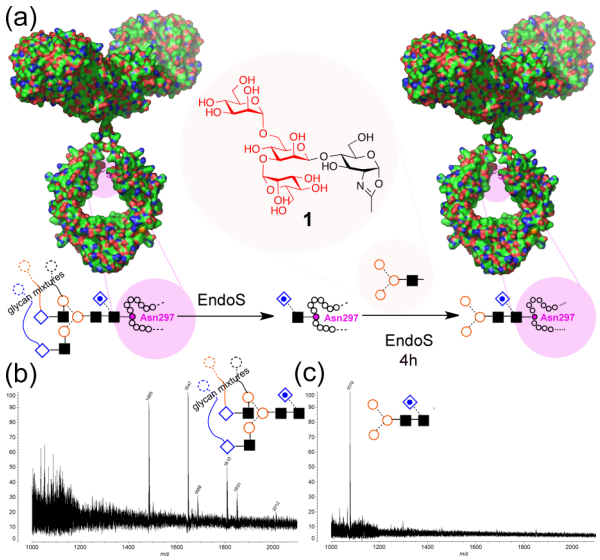


Figure 5. Remodeling of glycans on human IgG by EndoS. (a) By combining the unique trimming activity of EndoS (step 1) with subsequent EndoS-catalyzed transglycosylation using oxazoline **1** the first examples of human IgG remodeling were achieved to produce a single unnatural IgG glycoform bearing simply the core-pentasaccharide in a core-fucosylated state. (b),(c) analysis of glycans on IgG by MALDI-TOF following PNGase release from native (b) and remodeled (c) human IgG.

In summary, EndoS shows exciting breadth of glycan processing applicable to relevant remodeling of a wider range of glycoproteins than was possible through previous methods. The biocatalytic complementarity of EndoS to other Endo enzymes also raises the possibility of differential processing of mixed glycoform populations (Figure 3c and Table SXX). For example, it has been suggested that differently processed antibodies could be used as potential therapeutics for autoimmune diseases.[42] It would appear that EndoS possesses many traits (both glycan selectivity and protein efficiency) that will make it an attractive tool for synthesizing pure relevant glycoforms (e.g., complex-type) of important eukaryotic proteins (e.g. IgGs). The molecular basis of this remarkable activity and its use as a tool for IgG manipulation is currently under further study.

Notably, also, *S. pyogenes*, the genetic origin of *ndoS*, is a significant human pathogen that causes many human diseases,[43,44] including severe invasive diseases that have a high mortality rate such as necrotising fasciitis and sepsis.[45] Our studies may also address such pathogenicity, where EndoS appears to play a key role.[23]

We thank the BBSRC and Glycoform Ltd (JJG), VTU (KY) for funding and Drs Chris Urch, Thomas Purkarthofer, Roland Weis for helpful discussions. XXXXany othersXXX

Supporting Information. Experimental protocols, electrophoresis, spectra, additional analysis. This material is available free of charge via the Internet at http://pubs.acs.org.

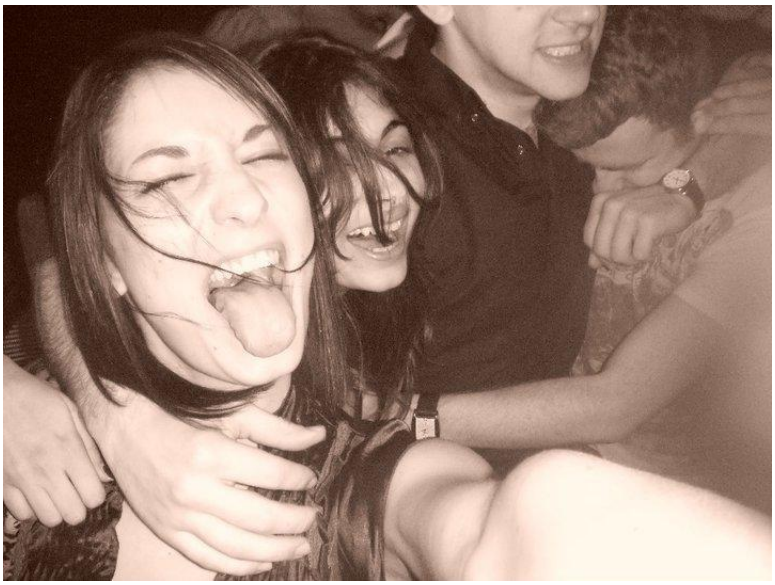
Author Contributions: JJG, KY, MC, CNS, BGD designed the experiments. JJG, KY performed experiments. CXX, KXX kindly provided gp120 & first Fc substrate proteins, respectively. CXX, KXX, JJG, KY collected MS; JJG, KY collected all other data. JJG, KY, BGD analyzed the data. BGD, KY, BGD wrote the manuscript.

(1) Ohtsubo, K.; Marth, J. D. *Cell* **2006**, *126*, 855-867.
(2) Apweiler, R.; Hermjakob, H.; Sharon, N. *Biochim. Biophys. Acta* **1999**, *1473*, 4-8.
(3) Jung, E.; Veuthey, A.-L.; Gasteiger, E.; Bairoch, A. *Proteomics* **2001**, *1*, 262–268.
(4) Sethuraman, N.; Stadheim, T. A. *Curr. Opin. Biotechnol.* **2006**, *17*, 341-346.
(5) Rudd, P. M.; Joao, H. C.; Coghill, E.; Fiten, P.; Saunders, M. R.; Opdenakker, G.; Dwek, R. A. *Biochemistry* **1994**, *33*, 17-22.
(6) Suzuki, T.; Kitajima, K.; Inoue, S.; Inoue, Y. *Glycoconjugate J.* **1995**, *12*, 183-193.
(7) Jefferis, R. *Nature Reviews Drug Discovery* **2009**, *8*, 226-234.
(8) Anthony, R. M.; Nimmerjahn, F.; Ashline, D. J.; Reinhold, V. N.; Paulson, J. C.; Ravetch, J. V. *Science* **2008**, *320*, 373-376.
(9) Gerngross, T. U. *Nat. Biotechnol.* **2004**, *22*, 1409-1414.
(10) Betenbaugh, M. J.; Tomiya, N.; Narang, S.; Hsu, J. T. A.; Lee, Y. C. *Curr. Opin. Struct. Biol.* **2004**, *14*, 601-606.
(11) Kowarik, M.; Numao, S.; Feldman, M. F.; Schulz, B. L.; Callewaert, N.; Kiermaier, E.; Catrein, I.; Aebi, M. *Science* **2006**, *314*, 1148 -1150.
(12) Bennett, C. S.; Wong, C.-H. *Chem. Soc. Rev.* **2007**, *36*, 1227.
(13) Gamblin, D. P.; Scanlan, E. M.; Davis, B. G. *Chem. Rev.* **2009**, *109*, 131-163.
(14) Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G. *Chem. Asian J.* **2009**, *4*, 630-640.
(15) Witte, K.; Sears, P.; Martin, R.; Wong, C.-H. *J. Am. Chem. Soc.* **1997**, *119*, 2114-2118.
(16) Wang, L.-X. *Carbohydr. Res.* **2008**, *343*, 1509-1522.
(17) Rich, J. R.; Withers, S. G. *Nat. Chem. Biol.* **2009**, *5*, 206-215.
(18) Schwarz, F.; Huang, W.; Li, C.; Schulz, B. L.; Lizak, C.; Palumbo, A.; Numao, S.; Neri, D.; Aebi, M.; Wang, L.-X. *Nat. Chem. Biol.* **2010**, *6*, 264-266.
(19) Elegant synthetic strategies have allowed the use of Endos that do not cleave complex or core-fucosylated glycans to instead attach them using unnatural substrates under forcing conditions. The processing of core-fucosylation has recently been shown on short peptide fragments. See the following references.
(20) Huang, W.; Yang, Q.; Umekawa, M.; Yamamoto, K.; Wang, L.-X. *ChemBioChem* **2010**, *11*, 1350-1355.
(21) Huang, W.; Li, J.; Wang, L.-X. *ChemBioChem* **2011**, *12*, 932.
(22) http://www.cazy.org/.
(23) Collin, M.; Olsen, A. *EMBO J.* **2001**, *20*, 3046-3055.
(24) Allhorn, M.; Briceno, J. G.; Baudino, L.; Lood, C.; Olsson, M. L.; Izui, S.; Collin, M. *Blood* **2010**, *115*, 5080-5088.
(25) Takegawa, K.; Yamaguchi, S.; Kondo, A.; Iwamoto, H.; Nakoshi, M.; Kato, I.; Iwahara, S. *Biochem. Int.* **1991**, *24*, 849-855.
(26) Robbins, P. W.; Trimble, R. B.; Wirth, D. F.; Hering, C.; Maley, F.; Maley, G. F.; Das, R.; Gibson, B. W.; Royal, N.; Biemann, K. *J. Biol. Chem.* **1984**, *259*, 7577-7583.
(27) Jefferis, R. *Biotechnol. Prog.* **2005**, *21*, 11-16.
(28) Gallo, R. C. *Retrovirology* **2009**, *6(suppl 2)*, I3.
(29) Doores, K. J.; Bonomelli, C.; Harvey, D. J.; Vasiljevic, S.; Dwek, R. A.; Burton, D. R.; Crispin, M.; Scanlan, C. N. *Proc. Natl Acad. Sci, U.S.A.* **2010**, *107*, 13800-13805.
(30) Elbein, A. D.; Tropea, J. E.; Mitchell, M.; Kaushal, G. P. *J. Biol. Chem.* **1990**, *265*, 15599-15605.
(31) Wang, L. X.; Wei, Y. D.; Li, C. S.; Huang, W.; Li, B.; Strome, S. *Biochemistry* **2008**, *47*, 10294-10304.
(32) Yamamoto, K. J.; Kadowaki, S.; Watanabe, J.; Kumagai, H. *Biochem. Biophys. Res. Commun.* **1994**, *203*, 244-252.
(33) Fujita, K.; Yamamoto, K. *Biochim. Biophys. Acta* **2006**, *1760*, 1631-1635.
(34) Takegawa, K.; Tabuchi, M.; Yamaguchi, S.; Kondo, A.; Kato, I.; Iwahara, S. *J. Biol. Chem.* **1995**, *270*, 3094-3099.
(35) Li, B.; Song, H.; Hauser, S.; Wang, L-X. *Org Lett* **2006**, *8*, 3081.
(36) Heidecke, C. D.; Ling, Z.; Bruce, N. C.; Moir, J. W. B.; Parsons, T. B.; Fairbanks, A. J. *ChemBioChem* **2008**, *9*, 2045-2051.
(37) Rising, T. W.; Heidecke, C. D.; Moir, J. W.; Ling, Z.; Fairbanks, A. J. *Chem. Eur. J.* **2008**, *14*, 6444-6464.
(38) Ochiai, H.; Huang, W.; Wang, L.-X. *J. Am. Chem. Soc.* **2008**, *130*, 13790-13803.
(39) Huang, W.; Ochiai, H.; Zhang, X.; Wang, L.-X. *Carbohydr. Res.* **2008**, *343*, 2903-2913.
(40) Huang, W.; Li, C.; Li, B.; Umekawa, M.; Yamamoto, K.; Zhang, X.; Wang, L.-X. *J. Am. Chem. Soc.* **2009**, *131*, 2214-2223.
(41) Fernandez-Gonzalez, M.; Boutureira, O.; Bernardes, G.J.L.; Chalker, J.M.; Young, M.; Errey, J.C.; Davis, B.G. *Chem. Sci.* **2010**.

(42) Allhorn, M.; Collin, M. *Ann. NY Acad. Sci.* **2009**, *1173*, 664.
(43) Bisno, A. L.; Stevens, D. L. *New Eng. J. Med.* **1996**, *334*, 240.
(44) Bisno, A. L.; Stevens, D. L. *New Engl. J. Med.* **1996**, *334*, 1478.
(45) Collin, M.; Svensson, M. D.; Sjoholm, A. G.; Jensenius, J. C.; Sjobring, U.; Olsen, A. *Infect. Immun.* **2002**, *70*, 6646-6651.

JACS Authors note: Authors are required to submit a graphic entry for the Table of Contents (TOC) that, in conjunction with the manuscript title, should give the reader a representative idea of one of the following: A key structure, reaction, equation, concept, or theorem, etc., that is discussed in the manuscript. The TOC graphic may be ***no wider than 9.0 cm*** and ***no taller than 3.5 cm*** as the graphic will be reproduced at 100% of the submission size. A surrounding margin will be added to this width and height during *Journal* production.

Insert Table of Contents artwork here



Pictures speak a thousand words, but this picture just speaks one: friendship.

I’ll miss you guys loads.