

# **Platforms for *In Vitro* Glycoengineering of proteins**

**Shi Xian Edward Moh**

BSc (Hons) Biomedical Science, Macquarie University

A thesis presented for the degree of

**Doctor of Philosophy**

Under the supervision of

**Prof. Nicolle H. Packer**



**MACQUARIE**  
University  
SYDNEY · AUSTRALIA

Department of Chemistry and Biomolecular Sciences

Macquarie University, Sydney, Australia

Jan 2018



## **Declaration**

I hereby certify that the work presented in this thesis entitled “Platforms for *in vitro* glycoengineering of proteins” has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree to any other university or institution other than Macquarie University. This thesis is an original piece of research and is the result of my own work except where appropriately acknowledged. Biosafety ethics approval for non-GMO (NIP041214BHA, Appendix 1) and GMO (5201600372, Appendix 2) has been duly obtained for research purposes only. I consent to a copy of this thesis being available in the Macquarie University library for relevant consultation, loan and photocopying forthwith.

Shi Xian Edward Moh

May 2017

## Acknowledgements

This work would not have been possible without the help of many and I would like to express my gratitude for all their help.

To my supervisor, Prof. Nicolle Packer, who had firmly believed in my abilities throughout the entire time, giving me the freedom of making project decisions and keeping me reigned in whenever my spontaneous ideas go overboard (or overtime for that matter).

To the post-Doctoral fellows in our Glyco group, Morten, Mathew, Chi Hung, Liisa, Nima and Arun, all of whom had provided me with their technical expertise, practical experience and shown me what it requires to being a qualified researcher and a good post doc.

To the graduates of the Glyco group, Merrina, Vignesh, Kat, Jenny, Ling, Jodie and Zeynep, all of whom had chipped into little things that made my time in the lab a much more enjoyable one, and showing me that finishing is a PhD is not impossible.

To the current members and students of our group, Ian, Chris, Shaz, Wei, Sameera, Harry and Sayatani, for all the experiences in learning, teaching, brainstorming and debating about different scientific aspects revolving around our various projects.

To my iGEM teaching crew, Dr Louise Brown, Prof Robert Willows, Ms Thi Hyunh, Liz and Mike, for all the enjoyable time we spent managing the students in class during iGEM seasons. Crazy, exciting, stressful, rewarding and addictive!

To my fellow PhD colleagues, Aneesh, Zhiping and Mafruha, for all the wonderful discussions about each other's projects and widening my scientific views on things outside of Glyco.

Last but definitely not the least, my family back home. My Mom, who had to deal with both her sons doing PhDs at 2 different continents, and my constant reminders of "just don't ask". My brother, with whom I exchange bickering about whose PhD experience is crazier. My 2 older sisters, who kept me grounded with reality and constantly updating me about news from home. My 3 adorable nephews that I had lots of fun with every time I visit.

My greatest appreciations to you all!

## Abstract

Protein glycosylation is an important post-translational modification that affects the protein structure and function. Glycoengineering is the process where the glycosylation outcome of the protein is modified for a designated purpose such as functionalisation of the glycans or generating a glycan-defined glycoprotein. The biggest advantage of *in vitro* glycoengineering methods is the controllable physical parameters that define the reaction environment.

In this dissertation, three different aspects of *in vitro* glycoengineering approaches were addressed; 1) the design of a DNA-guided protein scaffold for multi-glycosyltransferase reactions (Chapter 2), 2) immobilisation of glycosyltransferases on a stationary column for reusable, on-column glycosylation (Chapter 3), and 3) application of *in vitro* glycoengineering principles to a cancer-specific IgM antibody which was firstly characterised for its site-specific glycosylation (Chapter 4), the knowledge of which was used for multi-purpose conjugation of specific IgM glycans (Chapter 5).

In Chapter 2, the design of a sequence specific DNA-binding protein with a monomeric streptavidin fusion protein adaptor (TALE-mSA) was presented. By tailoring the DNA binding sequence of the adaptor to a predefined custom DNA program, spatial alignment of streptavidin-tagged glycosyltransferases can be achieved. TALE-mSA adaptor proteins were recombinantly expressed and purified from *E. coli* and tested for DNA binding capacity.

In Chapter 3, recombinantly expressed human B4GALT1 containing an *N*-terminal His-tag was used as a model glycosyltransferase to explore the possibility of *in vitro* glycosylation by immobilised of a glycosyltransferase on Ni-NTA resin. Galactosylation of de-sialylated, de-galactosylated bovine fetuin was found to be rapid; approximately 70% of the glycans were optimally galactosylated within 30mins. Re-usability, and adaptability for sequential on-column glycosylation were explored.

In Chapter 4, in-depth site specific glycosylation characterisation was performed on an anti-cancer IgM antibody, PAT-SM6, to establish the baseline glycosylation for the *in vitro* glycoengineering described in Chapter 5. As approximately 50% of the glycans contained a free galactose, a single step sialylation by ST6GAL1 using an azide-labelled CMP-sialic acid was chosen for chemo-enzymatic functionalisation of IgM glycans by click chemistry. In-depth characterisation of the incorporation of the azide-functionalised groups was determined.

In conclusion, development and application of these platforms will provide more tools towards understanding the details of how glycosylation affects protein function.

## List of publications

This thesis contains material that has been published and submitted for publication, as follows:

### Paper I

**E.S.X. Moh**, M. Thaysen-Andersen, N.H. Packer. (2014). *Relative vs Absolute Quantitation in Disease Glycomics*. Proteomics Clinical Applications, 9(3-4), 368-382.

#### *Contributions*

*I wrote the review in collaboration with my supervisor, Prof. Packer, and a Post-doctoral fellow, Dr Thaysen-Andersen.*

### Paper II

**E.S.X. Moh**, C.H. Lin, M. Thaysen-Andersen, N.H. Packer. (2015). *Site-specific N-glycosylation of recombinant pentameric and hexameric human IgM*. Journal of American Society of Mass Spectrometry, 27(7), 1143-1155.

#### *Contributions*

*I performed the experiments and acquired the data with help from Dr. Lin and Dr Thaysen-Andersen. Data discussion and editing of the manuscript was a collaborative effort of all the authors.*

## Related publications

### Paper III

M. Thaysen-Andersen, E. Chertova, C. Bergamaschi, **E.S.X. Moh**, O. Chertov, J. Roser, R. Sowder, J. Bear, J. Lifson, N. H. Packer, B. K. Felber, G. N. Pavlakis (2015). *Recombinant human heterodimeric IL-15 complex displays extensive and reproducible N- and O-linked glycosylation*. Glycoconjugate Journal, 33(3), 417-433.

#### *Contributions*

*I performed the glycan release experiments and the mass spectrometric data acquisition of the released N- and O-glycans.*

### Paper IV

L.Y. Lee, **E.S.X Moh**, B.L. Parker, M. Bern, N.H. Packer, M. Thaysen-Andersen. (2016). *Toward Automated N-Glycopeptide Identification in Glycoproteomics*. J Proteome Res, 15(10), 3904-3915.

#### *Contributions*

*I was involved in the manual characterisation of the basigin glycopeptides (referred to as “expert manual annotation” in the paper) and experimental data discussions*

## **Other publications during the candidature**

J.H.L. Chik, J.Zhou, **E.S.X. Moh**, R.Christopherson, S.J. Clarke, M.P. Molloy, N.H. Packer. (2014). *Comprehensive Glycomics Comparison between Colon Cancer Cell Cultures and Tumours: Implications for Biomarker Studies*. Journal of Proteomics, 108, 146-162.

### *Contributions*

*I performed the purification of MUC2 from LS174T colon cancer cells and the mass spectrometric experimental acquisition of the glycosylation profile of the purified MUC2.*

A.V. Everest-Dass, **E.S.X. Moh**, C.Ashwood, A.M.M. Shathili, N.H. Packer. (2018). *Human disease glycomics: technology advances enabling protein glycosylation analysis – part 1*. Expert Rev Proteomics, (Epub ahead of print). doi: 10.1080/14789450.2018.1421946

Everest-Dass AV, **Moh ESX**, Ashwood C, Shathili AMM, Packer NH. (2018). *Human disease glycomics: technology advances enabling protein glycosylation analysis – part 2*. Expert Rev Proteomics, (under review).

### *Contributions*

*I wrote the sections on the detection of glycoconjugates (Part I) and biological implications of glycans on antibodies (Part II), and was involved in the design and editing of the article.*



## Conference attendance and contributions

**E.S.X. Moh**, M. Thaysen-Andersen, N.H. Packer. *Site-specific N-glycosylation of recombinant pentameric and hexameric human IgM*. Lorne Proteomics, 5<sup>th</sup> – 8<sup>th</sup> Feb 2015, Lorne, Melbourne.

### Poster

**E.S.X. Moh**, M. Thaysen-Andersen, N.H. Packer. *Site-specific N-glycosylation of recombinant pentameric and hexameric human IgM*. 2<sup>nd</sup> Biologics Congress, 1<sup>st</sup> – 2<sup>nd</sup> Feb 2016, Berlin

Germany. **Poster**

Team Advisor for Macquarie\_Australia 2014. *Photophyll, The Green Machine*. iGEM 2014 Giant Jamboree, Oct 30<sup>th</sup> – Nov 3<sup>rd</sup>, Boston, Massachusetts. **Gold award**

Team Advisor for Macquarie\_Australia 2015. *Solar synthesisers*. iGEM 2015 Giant Jamboree, Sept 24<sup>th</sup> – 28<sup>th</sup>, Boston, Massachusetts. **Gold award**

Team Advisor for Macquarie\_Australia 2016. *Chlorophyll II: Return of the Hydrogen*. iGEM 2016 Giant Jamboree, Oct 27<sup>th</sup> – 31<sup>st</sup>, Boston, Massachusetts. **Gold award**

Team Instructor for Macquarie\_Australia 2017. *H<sub>2</sub>ydroGEM - Producers of Pollution-Free Energy*. iGEM 2017 Giant Jamboree, Nov 9<sup>th</sup> – 13<sup>th</sup>, Boston, Massachusetts. **Gold award, Best Energy Project**.

## Awards

- Macquarie University Postgraduate Research Fund 2015
- International Macquarie University Postgraduate Research Excellence Scholarship (iMQRES) 2013

## List of Abbreviations

Asn	Asparagine	IAA	Iodoacetamide
Cys	Cysteine	DTT	Dithiothreitol
Gln	Glutamine	PBS	Phosphate buffered saline
Gly	Glycine	NaBH <sub>4</sub>	Sodium borohydride
Leu	Leucine	ACN	Acetonitrile
Pro	Proline	TFA	Trifluoroacetic
Ser	Serine	SDS	Sodium dodecyl sulphate
Thr	Threonine	PAGE	Polyacrylamide Gel Electrophoresis
Fuc	Fucose	IMER	IMmobilised Enzyme Reactor
Gal	Galactose	Ni-NTA	Nickel-nitrilotriacetic acid
GalNAc	<i>N</i> -acetylgalactosamine	Cy5	Cyanine 5 dye
Glc	Glucose	DMSO	Dimethylsulphoxide
GlcNAc	<i>N</i> -acetylglucosamine	RP	Reversed phase
Man	Mannose	PGC	Porous graphitized carbon
Xyl	Xylose	ZIC	Zwitter ionic
Neu5Ac	<i>N</i> -acetylneuraminic acid	HILIC	Hydrophilic interaction chromatography
Neu5Gc	<i>N</i> -glycolylneuraminic acid	LC	Liquid chromatography
Hex	Hexose	nLC	Nanoflow liquid chromatography
HexNAc	Hexosamine	MS	Mass spectrometry
UDP	Uridine diphosphate	MS/MS	Tandem mass spectrometry
CMP	Cytosine monophosphate	ETD	Electron Transfer Dissociation
GDP	Guanosine diphosphate	CID	Collision Induced Dissociation
SNFG	Symbol Nomenclature For Glycomics	HCD	Higher-energy Collision Dissociation
TALE	Transcription activator-like effector	Da	Dalton
mSA	Monomeric streptavidin	m/z	Mass to charge ratio

## Table of Contents

Declaration.....	I
Acknowledgements.....	II
Abstract.....	III
List of publications .....	V
Conference attendance and contributions .....	VII
List of Abbreviations .....	VIII
Chapter 1 – Introduction .....	1
Protein glycosylation.....	2
<i>N</i> -Glycan biosynthesis .....	4
Nucleotide sugar biosynthesis.....	6
Nucleotide sugar transporters.....	7
Glycosylation and protein function.....	8
Glycoengineering .....	10
<i>In vivo</i> Glycoengineering .....	11
Engineering the target glycoprotein .....	12
Engineering the expression host.....	13
<i>In vitro</i> glycoengineering .....	15
Chemical synthesis of protein glycosylation.....	15
Enzymatic glyco-modification of proteins .....	16
Chemo-enzymatic synthesis of glycoproteins .....	17
Analysis of purified glycoproteins .....	18
Intact glycoprotein analysis .....	19
Glycopeptides analysis .....	20
Deglycosylated glycopeptide analysis.....	21
Released glycan analysis .....	22
References .....	23

Paper I – Relative vs absolute quantitation in disease glycomics .....	31
Thesis Aims .....	47
Chapter 2: <i>In vitro</i> protein glycosylation using scaffold mediated metabolic engineering .....	48
Introduction.....	49
Materials and Methods.....	56
Results.....	59
Discussion .....	66
Conclusion .....	68
References.....	69
Chapter 3 – <i>In vitro</i> protein glycosylation using immobilised glycosyltransferases.....	72
Introduction.....	73
Materials and Methods.....	75
Results.....	78
Discussion .....	84
Conclusion .....	87
References.....	88
Chapter 4 – Characterisation of the glycosylation of Immunoglobulin M.....	90
Introduction.....	91
Paper II – Site specific glycosylation characterisation of IgM.....	98
Additional Supplementary data for Paper II .....	112
References.....	118
Chapter 5: Chemo-enzymatic <i>in vitro</i> glycoengineering for functionalisation of IgM.....	124
Introduction.....	125
Materials and Methods.....	129
Results.....	131
Discussion .....	137
Conclusion .....	140

References .....	141
Chapter 6.....	143
Thesis summary .....	144
Conclusion.....	145
Appendix.....	150
Appendix I – NIP041214BHA.....	151
Appendix II – Biosafety approval 5201600372 .....	157
Appendix III – Sequence information.....	164
Appendix IV – Paper III.....	166
Appendix V – Paper IV .....	183



**Chapter 1 – Introduction**

### Protein glycosylation

Protein glycosylation is a common post translational modification where carbohydrate moieties are added to proteins by enzymatic activity of glycosyltransferases [4]. These carbohydrate moieties can be added as a monosaccharide and subsequently built up (*O*-glycans), or as an oligosaccharide and subsequently remodelled (*N*-glycans). On proteins, the two most common glycosylation observed on proteins are the *N*-linked and *O*-linked glycosylation. *N*-linked glycosylation (*N*-glycan) is characterised by the attachment of the glycan chain to the nitrogen atom of the asparagine amino acid residue of the polypeptide chain [3], while *O*-linked glycosylation (*O*-glycan) is via the oxygen atom on serine or threonine residues [6]. For *N*-glycans, a highly conserved motif of Asn-X-Ser/Thr ( $X \neq \text{Pro}$ ) has been shown to be the consensus sequon for *N*-glycosylation to occur [3], although there are exceptions that occur at low frequencies on Asn-X-Cys (where Cys is in its reduced form), or other sequons [8]. While there is no consensus sequon for *O*-glycans, it has been shown that a Ser/Thr that has Pro at the +3 or -1 position has a higher likelihood to carry an *O*-glycan [9]. Apart from protein glycosylation, glycans can also occur on lipids to form glycolipids [10], GPI anchors [11], and on small molecules to form glycosides [12]. Biomolecules such as peptidoglycans and lipopolysaccharides found in bacterial membranes [13], and proteoglycans and glycosaminoglycans found in the extracellular matrix [14] are also glycosylated with long polysaccharide chains. Although each of these different realms of glycosylation serve different functions in a wide range of research areas and are all biologically important, protein glycosylation is the key interest in this work.

● Glucose (Glc)	■ <i>N</i> -acetylglucosamine (GlcNAc)
● Galactose (Gal)	■ <i>N</i> -acetylgalactosamine (GalNAc)
● Mannose (Man)	◇ <i>N</i> -glycolylneuraminic acid (NeuGc)
▲ Fucose (Fuc)	◆ <i>N</i> -acetylneuraminic acid (NeuAc)

Figure 1. Symbol nomenclature for monosaccharides according to SNFG [1]. This set of 8 monosaccharides make up the most common mammalian *N*- and *O*-glycans.



One of the most distinguishing features of glycans is that glycans form branching structures despite not being template driven [4], while proteins are made from 20 amino acid building blocks that form linear chains based on the corresponding DNA codons. Glycans are built from a small set of monosaccharides; for common mammalian *N*- and *O*-glycans, 8 different monosaccharide building blocks make up the glycan structures (Figure 1). Symbols have often been used to simplify the complexity of the glycan structure, and the standardisation of the symbols has been improved recently under the Symbol Nomenclature for Graphical Representation of Glycans (SNFG)[1]. For *N*-glycans, there are 3 major types, high mannose, hybrid and complex structures (Figure 2), in order of synthesis as the glycans on the proteins are remodelled in the endoplasmic reticulum and the Golgi apparatus [3]. For *O*-glycans, the attachment to the Ser/Thr of the amino acid backbone starts with a GalNAc and is extended into 8 possible core structures for further addition of monosaccharides [6]. These *O*-GalNAc glycans are known as mucin-type *O*-glycans. *O*-GlcNAc glycans are another type of *O*-glycan that are found predominantly intracellularly in the nucleus and cytoplasm, unlike *O*-GalNAc glycans which mostly occur on membrane bound and secretory proteins [15]. *O*-GlcNAc glycans do not get extended and have gained research interest from their impact on transcription factors such as NF- $\kappa$ B and p53, both of which have known implications to cancer, by affecting their interaction with DNA or other associated proteins [15]. Recently, extended *O*-GlcNAc was found to be presented on the human leukocyte antigen class I [16], and was hypothesized to be involved with immune recognition. There are also other types of less common *O*-linked glycans such as *O*-Mannose, *O*-Fucose, and *O*-Glucose which can all be further extended into short glycan chains [17].

## N-Glycan biosynthesis

Biosynthesis of *N*-glycans is more complicated than the biosynthesis of *O*-glycans. For *O*-glycans, individual monosaccharides are put onto the growing glycan chain after the initial GalNAc has been added onto the polypeptide chain [6]. The diversity and complexity is determined by the availability and activity of glycosyltransferases and nucleotide sugar substrates. For *N*-glycans, the glycans are transferred *en bloc* onto the asparagine via the oligosaccharyltransferase protein complex (OST) and subsequently remodelled (Figure 3) [2, 3]. The entire process first begins on the cytosolic face of the endoplasmic reticulum (ER), where the GlcNAc-phosphate is added onto dolichol phosphate (P-Dol) by GlcNAc-1-phosphotransferase (encoded by *ALG7* gene) from UDP-GlcNAc to create GlcNAc-PP-Dol. This is then built up to a  $\text{Man}_5\text{GlcNAc}_2\text{-PP-Dol}$ , where the glycan is then flipped from the cytosolic face into the ER lumen by “flippase”. In humans, the identity of “flippase” has yet to be verified, but in yeast, this was shown to be encoded in the *RFT1p* gene locus [2]. The glycan precursor is subsequently built up to the  $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-PP-Dol}$  by a series of transferases, and transferred onto the asparagine of the *N*-glycosylation sequon on the polypeptide chain by OST. This leaves behind the PP-Dol, which is dephosphorylated back to P-Dol and recycled

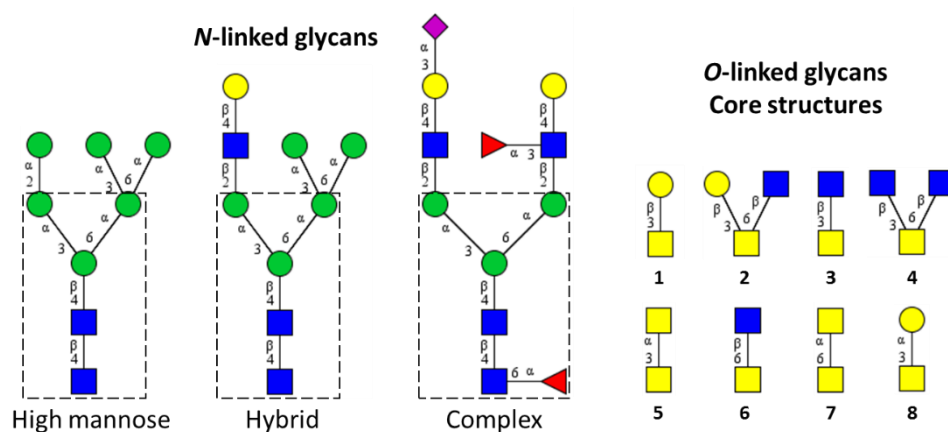


Figure 2. Schematic representation of glycan structures according to SNFG symbol nomenclature [1]. *N*-linked glycans all share a single core structure (boxed), and depending on the antenna composition, are classified into 3 major types, high mannose, hybrid, and complex structures [3]. *O*-linked glycans, unlike *N*-linked structures, branch out from the GalNAc that is linked to the serine/threonine of the amino acid backbone, forming 8 different possible types of core structures [6]. Linkages between monosaccharides are denoted by the anomeric position ( $\alpha$  or  $\beta$ ) and the carbon position (1,2,3,4,6) of the acceptor monosaccharide. Glycan structures were drawn using GlycoWorkBench [7].

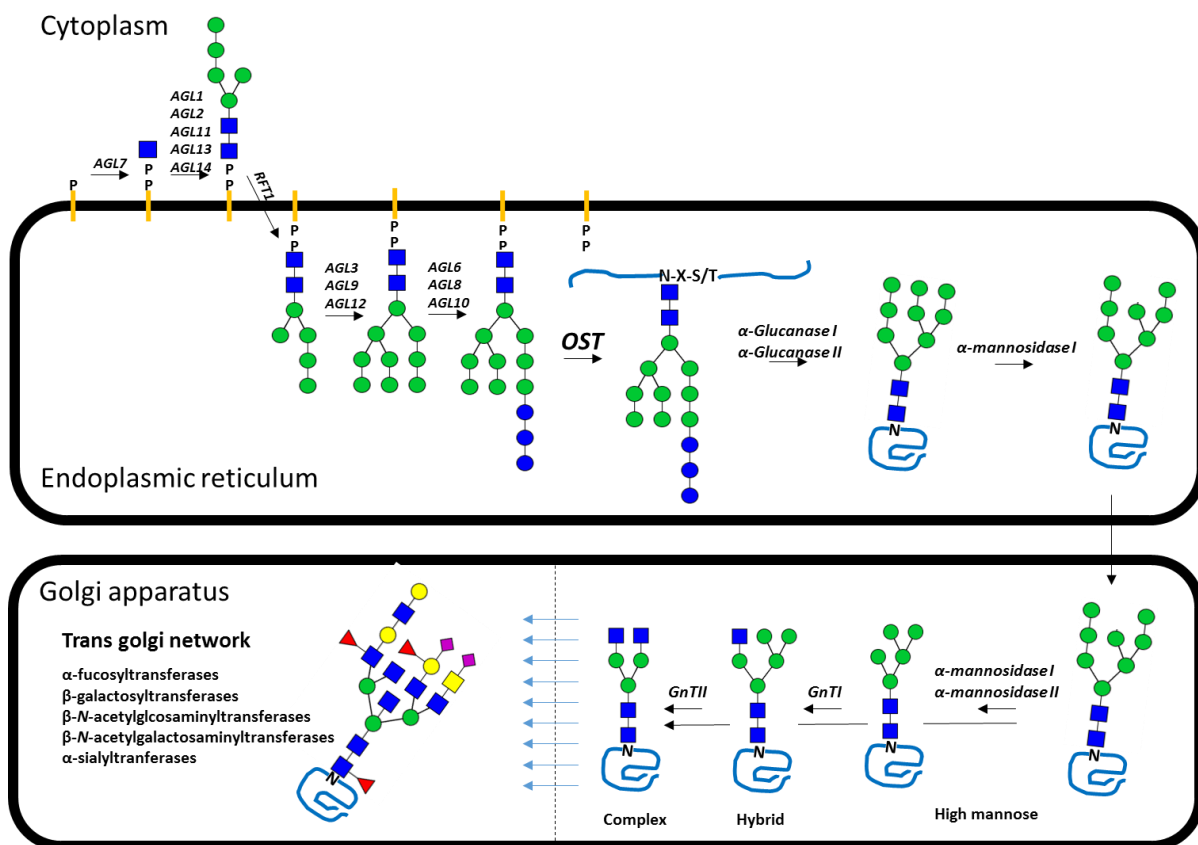


Figure 3. Schematic of the mammalian *N*-glycan biosynthesis pathway (adapted and modified from [2, 3]). The *N*-glycan precursor  $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$  is first synthesized on a dolichol phosphate lipid anchor (Dol-P-P, yellow stroke PP in figure) on the endoplasmic reticulum membrane. As it builds up to the  $\text{Man}_5\text{GlcNAc}_2\text{-Dol-P}$  in the cytoplasmic face of the ER, the glycan gets flipped into the ER lumen by RFT1 (in yeast), and gets build up to the  $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-Dol-P-P}$ . The whole glycan chain is then transferred onto the asparagine of the polypeptide chain by oligosaccharyltransferase (OST). While the protein folds, the 3 Glc residues and the centre mannose get trimmed, and the protein moves to the Golgi. In the Golgi, subsequent trimming of the mannose residues occurs, followed by the addition of the antenna GlcNAc to form the basis of hybrid or complex structures. As the glycoprotein then moves to the trans Golgi network, a wide range of glycosyltransferases act on the glycans to create the final glycosylated protein.

to build the next glycan precursor. As the now glycosylated polypeptide chain folds in the ER, glucosidases and mannosidase act on the  $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$  to trim it down to  $\text{Man}_8\text{GlcNAc}_2$  before it moves into the Golgi apparatus [3]. Once in the Golgi apparatus, the glycans on the glycoprotein can progress to form the 3 different types of common *N*-glycan classes, depending on several factors. Firstly, the glycosylation site accessibility has been shown to correlate with glycan processing [18, 19]; glycosylation sites that are not highly solvent exposed and inaccessible to glyco-enzymes are not able to progress into the more processed states. Secondly, the end-product *N*-glycan is also dependent on the relative abundance/localisation of the different types of glycosyltransferase/glycosidases since they are the direct modifiers of the

## Introduction

glycans. For instance, to progress to a complex structure, the mannosidase would have to completely cleave off all the mannose residues above the core structure, otherwise GlcNAc transferase II would not be about to attach the second GlcNAc to form a complex antenna structure. Similarly, fucosyltransferase VIII (Fut8) needs to act before the galactosyltransferase adds on the galactose onto the GlcNAc to build the core fucosylated glycans. It is this competition of substrates (glycoprotein) amongst the glyco-enzymes that creates the heterogeneous nature of any given glycoprotein. Glycomics, which is the area of research that studies the glycan profile of the glycosylation of any given complex biological material, uses this to reflect upon the state of the glycosylation machinery and to look at the possible correlations between disease states and changes in glycosylation outcomes. This is partly discussed in the review article written as part of this work in Paper I.

## **Nucleotide sugar biosynthesis**

Another important part of the mammalian glycosylation synthesis is the nucleotide sugars and their respective nucleotide sugar transporters. In the Golgi apparatus, every glycosyltransferase reaction catalyses the transfer of the sugar moiety of the nucleotide sugar onto the growing glycan chain, leaving behind the nucleotide diphosphate (except sialic acids which are carried by a monophosphate) [20]. Synthesis of the nucleotide sugars occurs in the cytosol of the cell and involves multiple enzymes for each nucleotide sugar [5] (Figure 4), that are transported into the Golgi lumen via their respective nucleotide sugar transporter [21, 22]. Typically, the monosaccharide gets phosphorylated by a kinase to form a sugar phosphate, and subsequently added to a nucleotide phosphate to form a nucleotide diphosphate. This does not apply to the sialic acids, Neu5Ac and Neu5Gc, which are synthesised via the GlcNAc metabolic pathway to form either ManNAc (from GlcNAc via an isomerase) or ManNAc-6P from UDP-GlcNAc via GNE, a bifunctional UDP-GlcNAc epimerase/ManNAc kinase (blue arrow head with red dash line in Figure 4), that is the substrate for Neu5Ac synthesis. It was suggested that the



## Introduction

corresponding export of CMP out of the Golgi. Multi-specific transporters such as HFRC1 can transport both UDP-GlcNAc and UDP-Glc [21]. With approximately 6-10 transmembrane domains, technical difficulties exist in the structural elucidation of the transporter [22]. Prediction of the substrate specificity by sequence homology is not reliable; in humans, UDP-GlcNAc can be transported by 3 different transporters, each classified into a different subfamily (SLC35A3, SLC35B4 and SLC35D2) that share very low amino acid identity [21]. Non-functional nucleotide sugar transporters have a direct impact on the resulting glycosylation outcome due to the shortage of nucleotide sugar substrates, and have been linked to congenital disorders of glycosylation and several diseases such as cancer, obesity, and dysplasia [21, 22].

## Glycosylation and protein function

The biological implications of glycans on a glycoprotein are not entirely predictable [4]; apart from sialylation contributing to longer serum half-life, having the same glycan on 2 different proteins does not equate to the same functional change, if any, of the non-glycosylated protein. These functional changes can include protein serum half-life [25, 26], protein folding and structure [27-30], protein-protein interaction [31, 32], and glycan-protein interaction [33-35]. For example, for erythropoietin (EPO), the well-known hematopoietic cytokine that is used to treat anaemia, the glycans, specifically the sialic acids on the glycans, are essential for *in vivo* activity affecting the serum half-life, even though they do not interfere with EPO receptor binding [25, 26]. A well-documented example of how glycans affect protein folding was reported for human corticosteroid binding globulin (hCBG). hCBG has six *N*-glycans sites and by testing a series of hCBG glycosylation site mutants, it was found that one of the glycosylation sites, Asn238, was required for cortisol binding [28]. Deglycosylation of the hCBG mutant carrying only the Asn238 glycosylation site did not affect the cortisol binding ability, suggesting that the Asn238 glycan was required only during the synthesis of the protein, affecting the protein folding of the cortisol binding pocket [27]. Recently, it was also shown

that the glycosylation at Asn347 affects the release of cortisol from hCBG by inhibiting the proteolytic activity of elastase [36]. Another well-researched example of how glycans affect protein structure is immunoglobulin G (IgG) glycosylation. The single glycosylation site of IgG, Asn297, is located on the Fc region of the antibody and changes in the glycosylation have been shown to affect the IgG effector functions such as complement activation and immune cell activation [37]. A portion of human IgG (approximately 20%) also contains a glycosylation site in the Fab region, and the presence of the glycan in the Fab region was reported to affect antigen binding [38, 39]. Single monosaccharide differences on the *N*-glycan structures on the Fc glycosylation site at Asn297 that result in glycan structural features such as core-fucosylation, bisecting GlcNAc (Figure 5), galactosylation and sialylation have been reported to alter FcγR binding affinity, which in turn impact the IgG effector functions as mentioned [30, 40-43]. This have been investigated by both computational modelling [30] and experimental data [31, 32]; absence of core fucosylation and presence of bisecting GlcNAc enhances binding affinity between IgG Fc and FcγR.

On the other hand, changes of glycan substructures/epitopes can also cause lectin interactions that were previously not present. Lectins are proteins that bind to carbohydrates, some of which recognise monosaccharides, while others recognise specific glycan epitopes[44]. For example, *Pseudomonas aeruginosa* expresses lectins LecA and LecB that recognise galactose and fucose respectively, and uses them adhere to galactose and fucose containing glycans of proteins in the mucus layer and epithelial cells of the host as part of their colonisation strategy [45, 46]. The biological role of lectins and their involvement in several biological activities such as anti-microbial, anti-cancer and cell signalling have been reviewed extensively [47-53]. These reviews have discussed various aspects of where glycans which bind lectins can have an important biological function/outcome on the relevant biological pathway. For instance, human immune cells have been documented to express a range of lectins such as siglecs, galectins and

## Introduction

selectins, which target sialic acid terminating glycans, galactose terminating glycans, and glycans with sialyl-Lewis<sup>X</sup> structures respectively [50, 51]. These lectin-glycan interactions play a role in recruitment and activation of immune cells such as B-cells and T-cells, and had also been reported to be involved in cancer metastasis [51]. Non-human glycan epitopes such as  $\alpha$ -1,3-Gal [54],  $\alpha$ -1,3 core fucosylation and  $\beta$ -1,2 xylose [55, 56] (Figure 5) on *N*-glycans can trigger an immune response by the host, resulting in production of the respective anti-glycan antibody. Hence, apart from protein-protein interactions, glycan-protein interactions can also play a role in determining protein function/activity.

## Glycoengineering

With implications on serum half-life, *in vivo* activity and immunogenicity, it is no surprise that the bio-pharmaceutical industry has interest in the glycosylation profiles of therapeutic glycoproteins [57-61]. Amongst the top 10 global pharmaceuticals in 2013 (product sales of >5 billion USD), 7 were biopharmaceuticals (6 of which were monoclonal antibodies) [62]. Of note, Rituxan<sup>TM</sup> (Rituximab), which is the monoclonal antibody used in the treatment of CD20-positive non-Hodgkin's lymphoma, was ranked 3<sup>rd</sup>, with global sales of approximately 8 billion USD. Recently, the western world's first glycoengineered monoclonal antibody Gazyva<sup>TM</sup>

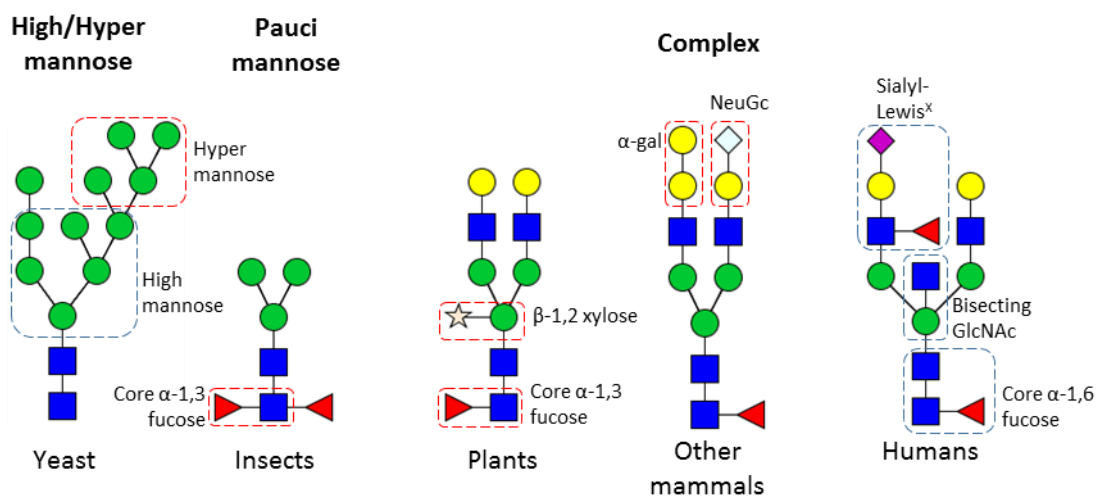


Figure 5. *N*-glycans of commonly used eukaryotic expression systems. Depicted in blue boxes are functional glycan epitopes that have been known to alter glycoprotein function. Depicted in red boxes are known immunogenic glycan epitopes that are not present in human glycans.



(obinutuzumab) that also targets CD20, made its way into clinical trials and has been proving to be more effective than its predecessor [42, 63]. Obinutuzumab has been glycoengineered to contain non-core fucosylated glycans, achieved by over expressing a galactosyltransferase (B4GALT3) to compete with the fucosyltransferase (FUT8) for substrates. Absence of core fucosylation has been demonstrated by many to increase affinity of the IgG Fc towards FcγRIII [29-31, 64-68]. It is important to note that while this is applicable to all IgG1 antibodies, it is not known if the affinity between other glycoproteins and their respective protein receptors would be enhanced by the absence of core-fucosylation.

Glycoengineering is a term that only surfaced in this century, even though much of the earlier work in the 70s looking at effects of glycosylation on protein function qualify as glycoengineering. The techniques can include genetically knocking out [69, 70] and/or adding in glycosylation sites [71, 72], inhibiting the glycosylation synthesis pathway [67, 73, 74], changing expression levels of the enzymes in the glycan synthesis pathway *in vivo* [75-77] or treating the glycoprotein with glycosyltransferases/glycosidases *in vitro* [31, 64, 78, 79], essentially any method that can cause a bias in the glycosylation outcome. Most of such glycoengineering work had been targeted at *N*-glycans, due to knowledge of the consensus sequon, making it easier to identify and manipulate *N*-glycosylation sites than other types of glycosylation. Glycoengineering approaches can be broadly categorised into *in vivo* and *in vitro* glycoengineering.

### ***In vivo* Glycoengineering**

Examples of *in vivo* glycoengineering can be classified into 2 types, one that engineers the glycoprotein, the other engineering the expression host.

## Introduction

### *Engineering the target glycoprotein*

As the glycosylation process is non-template driven, apart from knocking out/adding in glycosylation sites, not much else can be done with regards to engineering of the native glycoprotein. Glycosylation site knock outs are commonly performed by the amino acid mutation of the Asn, or the Ser/Thr at the +2 position, causing the disruption of the *N*-glycosylation sequon [27, 69, 70]. For glycoproteins with multiple glycosylation sites, combinations of different glycosylation site knock outs are required to investigate differences between the functional roles of each glycosylation site. The previously described example of hCBG illustrates this well; amongst the 6 *N*-glycosylation sites, only one was found to be crucial to the protein folding after screening the various *N*-glycan mutants of hCBG [27, 28]. This was also observed in human IgE, where only 1 out of the 7 glycosylation sites was entirely responsible for the binding to the IgE receptor on mast cells [69].

Adding *N*-glycosylation sites into a protein sequence requires more design planning than knocking out sites, as the location of the newly introduced *N*-glycan could potentially affect the overall protein folding. For instance, most tertiary protein structures have an internal hydrophobic structure and a hydrophilic or polar surface; if a glycosylation site is introduced into the amino acids of the hydrophobic region, the folding of the protein would change due to the hydrophilic nature of the added glycan. The rationale behind the addition of *N*-glycosylation sites can also be purposeful. As described earlier, EPO is a glycoprotein where the glycans do not affect the EPO receptor binding, and mainly improve the bioavailability of the protein [25, 26]. A variant of EPO, Aranesp®, is a variant of EPO carrying 2 additional *N*-glycosylation sites, bringing the number of *N*-glycans to five [72]. These extra sites increased the theoretical maximum number of sialic acids from 14 to 22 sialic acids, and were shown to increase the serum half-life of EPO by approximately 3-fold, thus improving *in vivo* activity while requiring a lower dose. In a more recent example, virus-like particles containing the envelop proteins of

hepatitis virus were glycoengineered to have 2 additional *N*-glycosylation sites [71]. The original envelop protein, HBsAgS, had only 1 glycosylation site in the major antigenic loop, that was not always glycosylated. With the increased number of glycosylation sites, the authors showed that the virus-like particles could elicit a higher immunogenic response from the host that generated a better immunisation outcome and led to development of better vaccines [71].

#### *Engineering the expression host*

The alternative of engineering the expression host of a desired glycoprotein can differ in complexity depending on the expression host, and the desired glycosylation outcome. In a straightforward example, to introduce bisecting GlcNAc onto adenylcyclase III, both adenylcyclase III and GntIII, the enzyme required for addition of bisecting GlcNAc, were transfected into a HEK293 variant cell line [80]. This resulted in increased bisecting GlcNAcylated glycans on adenylcyclase III, that was found to enhance adenylcyclase III activity. In another example, to increase the degree of sialylation, instead of over expressing sialyltransferases, the authors over expressed the CMP-sialic acid transporter in CHO cells [81]. By increasing the substrate availability in the Golgi, the amount of sialylation of interferon- $\gamma$  increased by up to 16% and provided an alternative method of enhancing sialylation irrespective of sialic acid linkage. In both examples, the straightforward nature of the glycoengineering process is mainly attributed to the expression host being a mammalian cell culture that has pre-existing mammalian *N*-glycosylation machinery.

However, in the biopharmaceutical industry it is paramount, apart from having a glycoprotein that does not cause adverse reactions by having the wrong glycans, to have a high production yield as it directly impacts sales volume. Mammalian cell cultures grow slowly, require growth media that is more expensive and are more difficult to maintain [43, 57, 58, 82]. From the 2014 list of FDA approved expression hosts for therapeutic proteins, yeast, insect, plant and bacterial

## Introduction

are listed [58]. When non-mammalian eukaryotic hosts such as yeast (*Pichia pastoris*) [75], plant (carrot root cells) [83] and insect cells (*Spodoptera frugiperda*) [84] are used for producing mammalian glycoprotein, multiple aspects of the *N*-glycan biosynthesis pathway need to be engineered to produce glycans that are suitable for humans. Although the expression of *N*-glycosylated proteins in bacteria has improved [85], this host has not yet been used to produce glycosylated biotherapeutics. However, the other 3 hosts, which are cheaper to grow and easier to maintain than mammalian cells, can potentially carry immunogenic glycan epitopes (Figure 5) that could result in an immunogenic response when used in humans; for example, hyper-mannosylation in yeast [86],  $\alpha$ -1,3-core fucosylation in insects and plants, and  $\beta$ -1,2-xylose in plants [56]. Hence, to glycoengineer non-mammalian expression hosts, apart from introducing the necessary enzymes required for the synthesis of mammalian complex glycans, enzymes that synthesise these unwanted glycan epitopes should be removed. For example, in plants, to achieve a sialylated complex glycan, without the  $\alpha$ -1,3-fucose and  $\beta$ -1,2-xylose, involves knocking out the fucosyltransferase and xylosyltransferase, and transfecting the plant with the sialyltransferase, sialic acid transporter, and the necessary enzymes for the synthesis of CMP-sialic acid (Figure 4). This has been successfully performed in *Nicotiana benthamiana*, where functional polymeric IgM with sialic acids but without xylose and core fucosylation was synthesized [76]. In insect cells, *N*-glycans are normally paucimannosidic structures (Figure 5) due to the *N*-acetylglucosaminidase activity present. To circumvent this, a Golgi localised galactosyltransferase was transfected to compete for the substrate as *N*-acetylglucosaminidase cannot act on galactose terminating glycans [87]. The galactosyltransferase activity also impedes the fucosyltransferase activity, which concomitantly reduced the immunogenic  $\alpha$ -1,3-core fucose epitope. Combining this transfection with the necessary enzymes for sialic acid synthesis and sialyltransferase, achieved sialylated glycans without core fucosylation. In yeast, specifically *Pichia pastoris*, this process is even more

complicated as the glycans are naturally terminating with mannose residues only, and UDP-Gal is not naturally present in the yeast [88]. A fusion protein containing a UDP-Glc epimerase,  $\beta$ -1,4-galactosyltransferase and a Golgi localised protein was designed and expressed in a mannosyltransferase deficient (*ALG3*) *P. pastoris* strain that had also been engineered to produce GlcNAc-transferase I and II. This resulted in the efficient formation of galactose terminating glycans in *Pichia*, where UDP-Gal was synthesized and transferred onto the glycan chain sequentially by the fusion protein located in the Golgi [88]. Enzymes required for the synthesis of sialic acids, CMP-Neu5Ac and the sialyltransferase were subsequently transfected into the new strain, enabling the formation of sialylated *N*-glycans [89]. The biotech company GlycoFi was created based on this technology [90], and also researched increasing glycosylation site occupancy by modifying the OST [91], as well as simulating mammalian *O*-GalNAc glycosylation with sialylated *O*-mannose glycans [92].

### ***In vitro* glycoengineering**

Within a living cell, complex networks of biochemical reaction pathways happen continuously and are highly complicated, making targeted *in vivo* control of a desired pathway, in this case the glycosylation pathway, multivariant and complex. The key difference and advantage of *in vitro* over *in vivo* glycoengineering methods are the defined reaction environment, where physical parameters such as concentrations of substrates and co-factors, reaction time and temperature are defined and controllable. *In vitro* glycoengineering methods to produce designed glycan structures on proteins can be broadly grouped into three main streams, chemical synthesis, enzymatic modifications, and the hybrid of both, chemo-enzymatic synthesis.

#### *Chemical synthesis of protein glycosylation*

## Introduction

Synthesising a glycan structure from scratch is not an easy task [93, 94], with multiple reactive hydroxyl groups in each monosaccharide and the need for defined specificity of the linkages between monosaccharides to build the branching structure of an *N*-glycan. Coupled with the requirement for different anomeric positions of certain monosaccharide linkages, multiple regioselective reaction schemes and protective groups have been used [93, 94]. By protecting/deprotecting specific hydroxyl groups of each monosaccharide, specific linkages and anomers can be achieved. To reduce the need for multiple rounds of purification after each monosaccharide addition, solid-phase synthesis methods have also been developed [93, 94]. With the increasing number of available chemical building blocks (di- and tri-saccharides instead of monosaccharides, with various regiospecific protection groups), whole free *N*-glycans with tri-antennae [95] or core fucose [96] have been produced purely by chemical synthesis methods (total chemical synthesis). For the chemical synthesis of glycopeptides, although solid-phase peptide synthesis has been developed [97, 98] producing long peptides (small proteins) of approximately 50 amino acids [99], difficulties still exist with side chain protection of some amino acids such as aspartic acid, arginine and asparagine, giving rise to undesired by-products [100]. By combining peptide synthesis methods with glycan synthesis methods, synthetic glycopeptides have been produced [101]. Progressing even further, through the chemical ligation of multiple glycopeptides, homogenous, chemically defined, small glycoproteins have been successfully synthesized completely from total chemical synthesis [99, 102].

### *Enzymatic glyco-modification of proteins*

Enzymatic modification of purified glycoproteins is a straightforward process that has been done for decades, using purified glyco-enzymes to either trim down or add monosaccharides towards a certain glycan type/structure/feature [103]. Sialylation is a common addition as it is the terminating monosaccharide in common mammalian *N*-glycans, and has been reported to

be involved in glycoprotein serum half-life and protein-receptor binding [25, 68, 78, 79]. De-sialylation by sialidases or sialylation by sialyltransferases are common methods for investigating effects of sialylation on the targeted glycoprotein. Bisecting GlcNAc, fucosylation (core and outer-arm) and galactosylation are also functional epitopes that result in differential protein-protein interaction affinity [29, 66, 67, 78]. While these are often straightforward biochemical reactions, the complexity increases when more than one glyco-enzyme is involved. Although many reports turn to *in vivo* glycoengineering approaches by over-expressing/transfecting the targeted glycosyltransferases to add the desired monosaccharide, others have employed *in vitro* glycoengineering means [31, 33, 64, 68, 78, 104]. Depending on the starting material and experimental setup, the heterogeneous glycoprotein population is sometimes trimmed down to a common base and built back up to the desired glycan structures. For example, to introduce a bisecting GlcNAc into monoclonal IgG, de-galactosylation needs to be performed initially as the bisecting GlcNAc-transferase cannot act on galactosylated glycans [64]. In another example, for the synthesis of sialyl-Lewis<sup>x</sup>, additional sialylation was performed *in vitro* prior to fucosylation to increase the amount of substrate (sialylated lactosamine epitopes) for fucosyltransferase-6 to form sialyl-Lewis<sup>x</sup> [33]. Apart from requiring the right precursor glycan, glycosylation by-products and substrates can also influence the completeness of the reaction. For example, the nucleotide sugar substrates for galactosyltransferase and *N*-acetylglucosaminyltransferase are both UDP-monosaccharides, and the UDP released from the transferase reaction can act as an inhibitor of the activity of either of these enzymes, thus requiring purification procedures between addition of the GlcNAc and the galactose to the precursor *N*-glycan [64]. To achieve reaction completeness, the *in vitro* glycosylation process can take up to several days (dependent on scale of production) and require the addition of more enzymes and nucleotide sugar substrates to the reaction [33, 64, 78].

#### *Chemo-enzymatic synthesis of glycoproteins*

## Introduction

As the term implies, chemo-enzymatic synthesis is the overlap between the total chemical synthesis and enzymatic modification methods. For most cases, the small molecule *N*-glycan precursor is synthesized by total chemical synthesis and is chemically conjugated to the protein made enzymatically within the cell. One of the advantages of using chemo-enzymatic synthesis is simplification of the process, as compared to a totally chemical synthesis or fully biochemical reaction. One example was reported where a homogeneously glycosylated ribonuclease C was produced [105, 106] by chemically ligating a synthetic glycopeptide to the partial amino acid chain of RNaseC made recombinantly. Recently, another chemo-enzymatic glycoengineering method has been developed, using the transglycosylation activity of an endoglycosidase to transfer a synthetic glycan onto the glycoprotein [107, 108]. In this method, a glycosylated IgG1 made in yeast was first deglycosylated using EndoH, leaving behind the GlcNAc attached to the asparagine. Synthetic bisecting *N*-glycan core structures carrying an oxazoline functional group instead of the reducing end GlcNAc were synthesized and added to the remaining GlcNAc on the asparagine by the transglycosylation activity of endoglycosidase A. As the specificity of the glycosidase activity of EndoA is targeted to high mannose glycans, the resulting bisecting glycans on the protein were not affected by the alternate glycosidase activity of EndoA [107, 108]. Subsequently, enzyme mutants were generated to create an endoglycosidase that can only perform the transglycosylation function [109]. This further led to the engineering of an EndoS mutant that can specifically transglycosylate onto the Fc glycan of IgG1, creating an active and homogeneously glycosylated IgG, a target that has been sought by biotherapeutic regulators [110, 111].

### **Analysis of purified glycoproteins**

Glycoprotein analysis, specifically the analysis of the heterogeneous glycoforms of glycoproteins, can currently be broadly classified into four approaches; analysis of the native or intact glycoprotein, analysis of digested glycopeptides, analysis of de-glycosylated peptides,



and analysis of released glycans. Analysis of the native glycoprotein requires the least amount of sample handling, but the outcome is highly dependent on the complexity and purity of the glycoprotein, and does not reveal information on glycan distribution across glycosylation sites. Analysis of glycopeptides provides information on glycosylation site occupancy (de-glycosylated peptides) and micro-heterogeneity (intact glycopeptides), but does not yet deliver information with regards to the detailed glycan structure on the sites. Analysis of released glycans provides details on specific glycan structures, but loses information on the glycan distribution across the glycosylation sites. As such, it is common for complementary techniques to be used for glycoprotein analysis.

#### *Intact glycoprotein analysis*

There are four main methods used to analyse glycoforms of an intact glycoprotein; lectins, isoelectric focusing, anion exchange chromatography and native mass spectrometry. Lectins, as described above, are carbohydrate binding proteins and some are highly selective towards carbohydrate structural targets, making them useful as analytical tools for identifying the presence of the target glycan epitope or linkage [44, 46]. For example, *Maackia amurensis* agglutinin (MAA-I) is specific to  $\alpha$ -2,3-linked sialic acid, while *Sambucus nigra* agglutinin (SNA-I) is specific to  $\alpha$ -2,6-linked sialic acid. Using lectins, or specific anti-glycan antibodies in some cases, the presence of specific glycan epitopes can be identified in a glycoprotein population. Lectin affinity chromatography and lectin arrays are common examples that employ lectins to determine specific glycosylated subsets of the glycoprotein [35, 112, 113]. Isoelectric focusing is a gel based separation technique that resolves proteins by their isoelectric point, and the presence of different glycoforms, especially differentially sialylated glycans, will result in a shift in the protein's isoelectric point [78, 114] that can be seen in an electrophoretic gel or by chromatographic separation. Anion exchange chromatography employs the same principle, using a shift in pH of the mobile phase to gradually elute different glycoforms of the protein

## Introduction

from the stationary phase [115]. Native mass spectrometry is the only way to accurately measure the exact molecular weight of glycoproteins and their glycoforms, although to achieve an interpretable mass spectrum is dependent on the complexity of the glycosylation of the protein. Large proteins and highly glycosylated proteins are currently not compatible with native MS as the deconvolution of the charge state envelop becomes increasingly difficult. Glycoforms of small proteins such as erythropoietin [116], or simple glycosylated proteins with few glycosylation sites such as IgG [117] have been investigated using native MS to look at the distribution of glycoforms of an intact glycoprotein.

### *Glycopeptides analysis*

By digesting the intact glycoprotein into peptide fragments using specific proteases such as trypsin or GluC, the complexity of the analyte decreases (many short peptides instead of an intact protein). Glycopeptide-based analysis is performed predominantly using mass spectrometry based methods, varying in sample preparation and/or fragmentation techniques. It is currently the only methodology that enables characterisation of site-specific glycosylation (with glycan composition) and glycosylation site occupancy, for glycoproteins with more than 1 glycosylation site. One of the biggest hurdles for glycopeptide analysis is the poor ionisation efficiency relative to non-glycosylated peptides [118], often resulting in ion suppression in a complex peptide mixture. Enrichment of glycopeptides is often performed to increase the glycopeptide signals and can be carried out using ZIC-HILIC [119, 120], lectin chromatography [121], or by chemical capturing of metabolically labelled glycans [122].

Characterisation of glycopeptides by mass spectrometry relies on the diagnostic fragmentation ions characteristic of glycopeptides. Different fragmentation techniques result in different fragmentation outcomes [123]; collision induced dissociation (CID) results in partial fragmentation of the glycan keeping mostly the peptide backbone intact, electron transfer

dissociation (ETD) results in fragmentation of the peptide backbone only, leaving the glycan intact, while higher-energy collisional dissociation (HCD) results in fragmentation of both the glycan and peptide backbone. A useful characteristic of glycopeptide fragmentation in HCD (and sometimes CID) mode is the high abundance of the Y<sub>1</sub>-ion comprising the intact peptide backbone and a single GlcNAc, and the presence of monosaccharide ions such as  $m/z$  204 (GlcNAc<sup>+1</sup>). The presence of the  $m/z$  204 ion can be used for product dependent acquisition to trigger fragmentation of the same precursor ion by ETD to obtain the peptide identity [124, 125]. In a newer fragmentation mode, EThcD, which is the combination of ETD and HCD, HCD fragmentation is applied to ETD fragmented ions in a single acquisition to yield a more informative fragment spectrum [126]. While ETD and HCD fragmentation results in a higher confidence of the glycopeptide identity, CID fragmentation provides partial glycan structural features that provide more information than only the glycan composition provided by the MS1 spectra.

#### *Deglycosylated glycopeptide analysis*

Alternatively, the glycopeptides can be de-glycosylated using endoglycosidases such as EndoH and EndoF3 (leaving behind a single GlcNAc) [127] or PNGaseF in presence of <sup>18</sup>O (resulting in a “heavy” aspartic acid residue) [128], which ionise better than their glycosylated precursors [118] and can be identified by standard proteomics/peptide based workflow. This can also be coupled with enrichment strategies where the glycans of glycoproteins were first oxidised to aldehydes for capturing using hydrazide beads. The glycoproteins were then proteolytically digested and the glycosylated peptides released by PNGaseF [129, 130]. These techniques are mostly used for glycosylation site identification analysis as information on glycan composition/structure is lost, unless a prior offline separation of individual glycopeptides before deglycosylation is performed [131].

## Introduction

### *Released glycan analysis*

The main purpose of analysing glycans after removal from the protein is to determine the detailed structural distribution of each glycan within the glycan population. For therapeutic glycoproteins, both structure and abundance have been found to be important. Presence of immunogenic glycan epitopes such as NeuGc and  $\alpha$ -linked galactose can be present when using non-human mammalian cells such as CHO cells [132] or  $\beta$ -1,2-xylose and  $\alpha$ -1,3-fucose present when using plant expression systems [55, 56]. These epitopes can result in a host immunogenic response, the severity of which depends on the relative abundance of these epitopes within the glycoprotein population. Identification of these detailed structural features of a released glycan can be performed by a number of mass spectrometric methods on reduced native glycans [133] or derivatised glycans [134-137] for which the chromatographic separation and detection method differ. For example, reduced native glycans can be separated by porous graphitised carbon chromatography and detected in on-line negative ion electrospray mass spectrometry [133], and the glycan structure elucidated by retention time, mass and presence of diagnostic fragmentation ions [138, 139]. As another example, glycans derivatised with a fluorophore can be separated by weak anion exchange or HILIC and detected using a fluorescence detector [134, 135], and the glycan structure elucidated by retention time and comparative analysis after using specific cocktails of glycosidases [135, 137]. The analytical differences of these methods, and the differences with respect to the final quantitation of individual glycan structures, has been discussed in detail in Paper I.

**References**

1. Varki, A., R.D. Cummings, M. Aebi, N.H. Packer, et al., Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology*, 2015, 25, 1323-1324.
2. Breitling, J. and M. Aebi, *N*-Linked Protein Glycosylation in the Endoplasmic Reticulum. *Cold Spring Harbor Perspectives in Biology*, 2013, 5, a013359.
3. Stanley, P., H. Schachter, and N. Taniguchi, *N*-Glycans, in *Essentials of Glycobiology*, A. Varki, et al. 2009, Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY). p. 101-114.
4. Varki, A. and J.B. Lowe, Biological Roles of Glycans, in *Essentials of Glycobiology*, A. Varki, et al. 2009, Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY). p. 75-88.
5. Bülter, T. and L. Elling, Enzymatic synthesis of nucleotide sugars. *Glycoconj J*, 1999, 16, 147-159.
6. Brockhausen, I., H. Schachter, and P. Stanley, *O*-GalNAc Glycans, in *Essentials of Glycobiology*, A. Varki, et al. 2009, Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY). p. 115-128.
7. Ceroni, A., K. Maass, H. Geyer, R. Geyer, et al., GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res*, 2008, 7, 1650-1659.
8. Valliere-Douglass, J.F., P. Kodama, M. Mujacic, L.J. Brady, et al., Asparagine-linked oligosaccharides present on a non-consensus amino acid sequence in the CH1 domain of human antibodies. *J Biol Chem*, 2009, 284, 32493-32506.
9. Thanka Christlet, T.H. and K. Veluraja, Database analysis of *O*-glycosylation sites in proteins. *Biophys J*, 2001, 80, 952-960.
10. Schnaar, R.L., A. Suzuki, and P. Stanley, Glycosphingolipids, in *Essentials of Glycobiology*, A. Varki, et al. 2009, Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY).
11. Ferguson, M.A.J., T. Kinoshita, and G.W. Hart, Glycosylphosphatidylinositol Anchors, in *Essentials of Glycobiology*, A. Varki, et al. 2009, Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY).
12. Desmet, T., W. Soetaert, P. Bojarová, V. Křen, et al., Enzymatic Glycosylation of Small Molecules: Challenging Substrates Require Tailored Catalysts. *Chemistry – A European Journal*, 2012, 18, 10786-10801.
13. Nizet, V. and J.D. Esko, Bacterial and Viral Infections, in *Essentials of Glycobiology*, A. Varki, et al. 2009, Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY).
14. Esko, J.D., K. Kimata, and U. Lindahl, Proteoglycans and Sulfated Glycosaminoglycans, in *Essentials of Glycobiology*, A. Varki, et al. 2009, Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY).
15. Hart, G.W. and Y. Akimoto, The *O*-GlcNAc Modification, in *Essentials of Glycobiology*, A. Varki, et al. 2009, Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY).
16. Marino, F., M. Bern, G.P. Mommen, A.C. Leney, et al., Extended *O*-GlcNAc on HLA Class-I-Bound Peptides. *J Am Chem Soc*, 2015, 137, 10922-10925.
17. Freeze, H.H. and R.S. Haltiwanger, Other Classes of ER/Golgi-derived Glycans, in *Essentials of Glycobiology*, A. Varki, et al. 2009, Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY).

18. Lee, L.Y., C.H. Lin, S. Fanayan, N.H. Packer, et al., Differential site accessibility mechanistically explains subcellular-specific *N*-glycosylation determinants. *Front Immunol*, 2014, 5, 404.
19. Thaysen-Andersen, M. and N.H. Packer, Site-specific glycoproteomics confirms that protein structure dictates formation of *N*-glycan type, core fucosylation and branching. *Glycobiology*, 2012, 22, 1440-1452.
20. Rini, J., J. Esko, and A. Varki, Glycosyltransferases and Glycan-processing Enzymes, in *Essentials of Glycobiology*, A. Varki, et al. 2009, Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY). p. 63-74.
21. Hadley, B., A. Maggioni, A. Ashikov, C.J. Day, et al., Structure and function of nucleotide sugar transporters: Current progress. *Computational and structural biotechnology journal*, 2014, 10, 23-32.
22. Song, Z., Roles of the nucleotide sugar transporters (SLC35 family) in health and disease. *Molecular Aspects of Medicine*, 2013, 34, 590-600.
23. Varki, A. and R. Schauer, Sialic Acids, in *Essentials of Glycobiology*, A. Varki, et al. 2009, Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY).
24. Samraj, A.N., O.M.T. Pearce, H. Läubli, A.N. Crittenden, et al., A red meat-derived glycan promotes inflammation and cancer progression. *Proceedings of the National Academy of Sciences*, 2015, 112, 542-547.
25. Liu, L., H. Li, S.R. Hamilton, S. Gomathinayagam, et al., The impact of sialic acids on the pharmacokinetics of a PEGylated erythropoietin. *J Pharm Sci*, 2012, 101, 4414-4418.
26. Shahrokh, Z., L. Royle, R. Saldova, J. Bones, et al., Erythropoietin Produced in a Human Cell Line (Dynepo) Has Significant Differences in Glycosylation Compared with Erythropoietins Produced in CHO Cell Lines. *Molecular Pharmaceutics*, 2010, 8, 286-296.
27. Avvakumov, G.V. and G.L. Hammond, Glycosylation of human corticosteroid-binding globulin. Differential processing and significance of carbohydrate chains at individual sites. *Biochemistry*, 1994, 33, 5759-5765.
28. Avvakumov, G.V., S. Warmels-Rodenhisser, and G.L. Hammond, Glycosylation of human corticosteroid-binding globulin at asparagine 238 is necessary for steroid binding. *J Biol Chem*, 1993, 268, 862-866.
29. Mizushima, T., H. Yagi, E. Takemoto, M. Shibata-Koyama, et al., Structural basis for improved efficacy of therapeutic antibodies on defucosylation of their Fc glycans. *Genes Cells*, 2011, 16, 1071-1080.
30. Ferrara, C., S. Grau, C. Jager, P. Sondermann, et al., Unique carbohydrate-carbohydrate interactions are required for high affinity binding between FcγRIII and antibodies lacking core fucose. *Proc Natl Acad Sci U S A*, 2011, 108, 12669-12674.
31. Thomann, M., T. Schlothauer, T. Dashivets, S. Malik, et al., *In Vitro* Glycoengineering of IgG1 and Its Effect on Fc Receptor Binding and ADCC Activity. *PLoS One*, 2015, 10, e0134949.
32. Lin, C.-W., M.-H. Tsai, S.-T. Li, T.-I. Tsai, et al., A common glycan structure on immunoglobulin G for enhancement of effector functions. *Proceedings of the National Academy of Sciences*, 2015.
33. Thomas, L.J., K. Panneerselvam, D.T. Beattie, M.D. Picard, et al., Production of a complement inhibitor possessing sialyl Lewis X moieties by *in vitro* glycosylation technology. *Glycobiology*, 2004, 14, 883-893.
34. Yang, H., Z. Li, X. Wei, R. Huang, et al., Detection and discrimination of alpha-fetoprotein with a label-free electrochemical impedance spectroscopy biosensor array based on lectin functionalized carbon nanotubes. *Talanta*, 2013, 111, 62-68.

35. Stadlmann, J., A. Weber, M. Pabst, H. Anderle, et al., A close look at human IgG sialylation and subclass distribution after lectin fractionation. *Proteomics*, 2009, 9, 4143-4153.
36. Sumer-Bayraktar, Z., O.C. Grant, V. Venkatakrisnan, R.J. Woods, et al., Asn347 Glycosylation of Corticosteroid-binding Globulin Fine-tunes the Host Immune Response by Modulating Proteolysis by *Pseudomonas aeruginosa* and Neutrophil Elastase. *Journal of Biological Chemistry*, 2016, 291, 17727-17742.
37. Jefferis, R., Glycosylation as a strategy to improve antibody-based therapeutics. *Nat Rev Drug Discov*, 2009, 8, 226-234.
38. Bondt, A., Y. Rombouts, M.H. Selman, P.J. Hensbergen, et al., Immunoglobulin G (IgG) Fab glycosylation analysis using a new mass spectrometric high-throughput profiling method reveals pregnancy-associated changes. *Mol Cell Proteomics*, 2014, 13, 3029-3039.
39. van de Bovenkamp, F.S., L. Hafkenscheid, T. Rispens, and Y. Rombouts, The Emerging Importance of IgG Fab Glycosylation in Immunity. *The Journal of Immunology*, 2016, 196, 1435-1441.
40. Vidarsson, G., G. Dekkers, and T. Rispens, IgG Subclasses and Allotypes: From Structure to Effector Functions. *Frontiers in Immunology*, 2014, 5, 520.
41. Hudak, J.E. and C.R. Bertozzi, Glycotherapy: new advances inspire a reemergence of glycans in medicine. *Chem Biol*, 2014, 21, 16-37.
42. Ratner, M., Genentech's glyco-engineered antibody to succeed Rituxan. *Nat Biotechnol*, 2014, 32, 6-7.
43. Loos, A. and H. Steinkellner, IgG-Fc glycoengineering in non-mammalian expression hosts. *Arch Biochem Biophys*, 2012, 526, 167-173.
44. Cummings, R.D. and M.E. Etzler, Antibodies and Lectins in Glycan Analysis, in *Essentials of Glycobiology*, A. Varki, et al. 2009, Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY).
45. Grishin, A.V., M.S. Krivozubov, A.S. Karyagina, and A.L. Gintsburg, *Pseudomonas Aeruginosa* Lectins As Targets for Novel Antibacterials. *Acta Naturae*, 2015, 7, 29-41.
46. Esko, J.D. and N. Sharon, Microbial Lectins: Hemagglutinins, Adhesins, and Toxins, in *Essentials of Glycobiology*, A. Varki, et al. 2009, Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY). p. 489-500.
47. Yau, T., X. Dan, C.C. Ng, and T.B. Ng, Lectins with potential for anti-cancer therapy. *Molecules*, 2015, 20, 3791-3810.
48. Singh, R.S. and A.K. Walia, Microbial lectins and their prospective mitogenic potential. *Critical reviews in microbiology*, 2014, 40, 329-347.
49. Hassan, M.A., R. Rouf, E. Tiralongo, T.W. May, et al., Mushroom lectins: specificity, structure and bioactivity relevant to human disease. *Int J Mol Sci*, 2015, 16, 7802-7838.
50. Loke, I., D. Kolarich, N.H. Packer, and M. Thaysen-Andersen, Emerging roles of protein mannosylation in inflammation and infection. *Molecular Aspects of Medicine*, 2016, 51, 31-55.
51. Cagnoni, A.J., J.M. Pérez Sáez, G.A. Rabinovich, and K.V. Mariño, Turning-Off Signaling by Siglecs, Selectins, and Galectins: Chemical Inhibition of Glycan-Dependent Interactions in Cancer. *Frontiers in Oncology*, 2016, 6.
52. Farrar, C.A., W. Zhou, and S.H. Sacks, Role of the lectin complement pathway in kidney transplantation. *Immunobiology*, 2016, 221, 1068-1072.
53. Nio-Kobayashi, J., Tissue- and cell-specific localization of galectins,  $\beta$ -galactose-binding animal lectins, and their potential functions in health and disease. *Anatomical Science International*, 2016, 1-12.

54. Huai, G., P. Qi, H. Yang, and Y.I. Wang, Characteristics of  $\alpha$ -Gal epitope, anti-Gal antibody,  $\alpha$ 1,3 galactosyltransferase and its clinical exploitation (Review). *International Journal of Molecular Medicine*, 2016, 37, 11-20.
55. Jin, C., F. Altmann, R. Strasser, L. Mach, et al., A plant-derived human monoclonal antibody induces an anti-carbohydrate immune response in rabbits. *Glycobiology*, 2008, 18, 235-241.
56. Bardor, M., C. Faveeuw, A.-C. Fitchette, D. Gilbert, et al., Immunoreactivity in mammals of two typical plant glyco-epitopes, core  $\alpha$ (1,3)-fucose and core xylose. *Glycobiology*, 2003, 13, 427-434.
57. Zhu, J., Mammalian cell protein expression for biopharmaceutical production. *Biotechnol Adv*, 2012, 30, 1158-1170.
58. Dumont, J., D. Euwart, B. Mei, S. Estes, et al., Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Critical Reviews in Biotechnology*, 2016, 36, 1110-1122.
59. Sola, R.J. and K. Griebenow, Effects of glycosylation on the stability of protein pharmaceuticals. *J Pharm Sci*, 2009, 98, 1223-1245.
60. Dotz, V., R. Haselberg, A. Shubhakar, R.P. Kozak, et al., Mass spectrometry for glycosylation analysis of biopharmaceuticals. *TrAC Trends in Analytical Chemistry*, 2015.
61. Bertozzi, C.R., H.H. Freeze, A. Varki, and J.D. Esko, Glycans in Biotechnology and the Pharmaceutical Industry, in *Essentials of Glycobiology*, A. Varki, et al. 2009, Cold Spring Harbor Laboratory Press. : Cold Spring Harbor (NY).
62. Castilho, L., *Biopharmaceutical Products: An Introduction. Current Developments in Biotechnology and Bioengineering: Human and Animal Health Applications*, 2016, 1.
63. Dhillon, S., *Obinutuzumab: A Review in Rituximab-Refractory or -Relapsed Follicular Lymphoma. Targeted Oncology*, 2017, 1-8.
64. Hodoniczky, J., Y.Z. Zheng, and D.C. James, Control of recombinant monoclonal antibody effector functions by Fc *N*-glycan remodeling in vitro. *Biotechnol Prog*, 2005, 21, 1644-1652.
65. Raju, T.S., Terminal sugars of Fc glycans influence antibody effector functions of IgGs. *Curr Opin Immunol*, 2008, 20, 471-478.
66. Takahashi, M., Y. Kuroki, K. Ohtsubo, and N. Taniguchi, Core fucose and bisecting GlcNAc, the direct modifiers of the *N*-glycan core: their functions and target proteins. *Carbohydr Res*, 2009, 344, 1387-1390.
67. Bowden, T.A., K. Baruah, C.H. Coles, D.J. Harvey, et al., Chemical and structural analysis of an antibody folding intermediate trapped during glycan biosynthesis. *J Am Chem Soc*, 2012, 134, 17554-17563.
68. Raju, T.S., J.B. Briggs, S.M. Chamow, M.E. Winkler, et al., Glycoengineering of therapeutic glycoproteins: in vitro galactosylation and sialylation of glycoproteins with terminal *N*-acetylglucosamine and galactose residues. *Biochemistry*, 2001, 40, 8868-8876.
69. Shade, K.T., B. Platzer, N. Washburn, V. Mani, et al., A single glycan on IgE is indispensable for initiation of anaphylaxis. *J Exp Med*, 2015, 212, 457-467.
70. Muraoka, S. and M.J. Shulman, Structural requirements for IgM assembly and cytolytic activity. Effects of mutations in the oligosaccharide acceptor site at Asn402. *J Immunol*, 1989, 142, 695-701.
71. Hyakumura, M., R. Walsh, M. Thaysen-Andersen, N.J. Kingston, et al., Modification of Asparagine-Linked Glycan Density for the Design of Hepatitis B Virus Virus-Like Particles with Enhanced Immunogenicity. *J Virol*, 2015, 89, 11312-11322.



72. Egrie, J.C. and J.K. Browne, Development and characterization of novel erythropoiesis stimulating protein (NESP). *Br J Cancer*, 2001, 84 Suppl 1, 3-10.
73. Belcher, J.D., C. Chen, J. Nguyen, F. Abdulla, et al., The Fucosylation Inhibitor, 2-Fluorofucose, Inhibits Vaso-Occlusion, Leukocyte-Endothelium Interactions and NF- $\kappa$ B Activation in Transgenic Sickle Mice. *PLoS One*, 2015, 10, e0117772.
74. Salvini, R., A. Bardoni, M. Valli, and M. Trinchera, beta 1,3-Galactosyltransferase beta 3Gal-T5 acts on the GlcNAc $\beta$ 1- $\rightarrow$ 3Gal $\beta$ 1- $\rightarrow$ 4GlcNAc $\beta$ 1- $\rightarrow$ R sugar chains of carcinoembryonic antigen and other *N*-linked glycoproteins and is down-regulated in colon adenocarcinomas. *J Biol Chem*, 2001, 276, 3564-3573.
75. Hamilton, S.R. and T.U. Gerngross, Glycosylation engineering in yeast: the advent of fully humanized yeast. *Curr Opin Biotechnol*, 2007, 18, 387-392.
76. Loos, A., C. Gruber, F. Altmann, U. Mehofer, et al., Expression and glycoengineering of functionally active heteromultimeric IgM in plants. *Proc Natl Acad Sci U S A*, 2014, 111, 6263-6268.
77. Isaji, T., J. Gu, R. Nishiuchi, Y. Zhao, et al., Introduction of bisecting GlcNAc into integrin  $\alpha$ 5 $\beta$ 1 reduces ligand binding and down-regulates cell adhesion and cell migration. *J Biol Chem*, 2004, 279, 19747-19754.
78. Sohn, Y., J.M. Lee, H.-R. Park, S.-C. Jung, et al., Enhanced sialylation and in vivo efficacy of recombinant human  $\alpha$ -galactosidase through in vitro glycosylation. *BMB reports*, 2013, 46, 157-162.
79. Colucci, M., H. Stockmann, A. Butera, A. Masotti, et al., Sialylation of *N*-linked glycans influences the immunomodulatory effects of IgM on T cells. *J Immunol*, 2015, 194, 151-157.
80. Li, W., M. Takahashi, Y. Shibukawa, S. Yokoe, et al., Introduction of bisecting GlcNAc in *N*-glycans of adenyl cyclase III enhances its activity. *Glycobiology*, 2007, 17, 655-662.
81. Wong, N.S., M.G. Yap, and D.I. Wang, Enhancing recombinant glycoprotein sialylation through CMP-sialic acid transporter over expression in Chinese hamster ovary cells. *Biotechnol Bioeng*, 2006, 93, 1005-1016.
82. Hossler, P., S.F. Khattak, and Z.J. Li, Optimal and consistent protein glycosylation in mammalian cell culture. *Glycobiology*, 2009, 19, 936-949.
83. Xu, J. and N. Zhang, On the way to commercializing plant cell culture platform for biopharmaceuticals: present status and prospect. *Pharmaceutical bioprocessing*, 2014, 2, 499-518.
84. Kost, T.A., J.P. Condreay, and D.L. Jarvis, Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat Biotechnol*, 2005, 23, 567-575.
85. Baker, J.L., E. Çelik, and M.P. DeLisa, Expanding the glycoengineering toolbox: the rise of bacterial *N*-linked protein glycosylation. *Trends in Biotechnology*, 2013, 31, 313-323.
86. Barreto-Bergter, E. and R.T. Figueiredo, Fungal glycans and the innate immune recognition. *Frontiers in Cellular and Infection Microbiology*, 2014, 4, 145.
87. Geisler, C., H. Mabashi-Asazuma, C.W. Kuo, K.H. Khoo, et al., Engineering beta1,4-galactosyltransferase I to reduce secretion and enhance *N*-glycan elongation in insect cells. *J Biotechnol*, 2015, 193, 52-65.
88. Bobrowicz, P., R.C. Davidson, H. Li, T.I. Potgieter, et al., Engineering of an artificial glycosylation pathway blocked in core oligosaccharide assembly in the yeast *Pichia pastoris*: production of complex humanized glycoproteins with terminal galactose. *Glycobiology*, 2004, 14, 757-766.
89. Hamilton, S.R., R.C. Davidson, N. Sethuraman, J.H. Nett, et al., Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science*, 2006, 313, 1441-1443.

90. Beck, A., O. Cochet, and T. Wurch, GlycoFi's technology to control the glycosylation of recombinant therapeutic proteins. *Expert Opin Drug Discov*, 2010, 5, 95-111.
91. Choi, B.K., S. Warburton, H. Lin, R. Patel, et al., Improvement of *N*-glycan site occupancy of therapeutic glycoproteins produced in *Pichia pastoris*. *Appl Microbiol Biotechnol*, 2012, 95, 671-682.
92. Hamilton, S.R., W.J. Cook, S. Gomathinayagam, I. Burnina, et al., Production of sialylated *O*-linked glycans in *Pichia pastoris*. *Glycobiology*, 2013, 23, 1192-1203.
93. Mrazek, H., L. Weignerova, P. Bojarova, P. Novak, et al., Carbohydrate synthesis and biosynthesis technologies for cracking of the glycan code: recent advances. *Biotechnol Adv*, 2013, 31, 17-37.
94. Seeberger, P.H., The Logic of Automated Glycan Assembly. *Accounts of Chemical Research*, 2015, 48, 1450-1463.
95. Walczak, M.A., J. Hayashida, and S.J. Danishefsky, Building Biologics by Chemical Synthesis: Practical Preparation of Di- and Triantennary *N*-linked Glycoconjugates. *J Am Chem Soc*, 2013, 135, 4700-4703.
96. Nagasaki, M., Y. Manabe, N. Minamoto, K. Tanaka, et al., Chemical Synthesis of a Complex-Type *N*-Glycan Containing a Core Fucose. *The Journal of Organic Chemistry*, 2016, 81, 10600-10616.
97. Coin, I., M. Beyermann, and M. Bienert, Solid-phase peptide synthesis: from standard procedures to the synthesis of difficult sequences. *Nat. Protocols*, 2007, 2, 3247-3256.
98. Rush, J. and C.R. Bertozzi, An  $\alpha$ -Formylglycine Building Block For Fmoc-Based Solid-Phase Peptide Synthesis. *Organic letters*, 2006, 8, 131-134.
99. Fernández-Tejada, A., J. Brailsford, Q. Zhang, J.-H. Shieh, et al., Total Synthesis of Glycosylated Proteins. *Topics in current chemistry*, 2015, 362, 1-26.
100. Behrendt, R., P. White, and J. Offer, Advances in Fmoc solid-phase peptide synthesis. *Journal of Peptide Science*, 2016, 22, 4-27.
101. Wang, P., B. Aussedat, Y. Vohra, and S.J. Danishefsky, An Advance in the Chemical Synthesis of Homogeneous *N*-Linked Glycopolypeptides by Convergent Aspartylation. *Angewandte Chemie (International ed. in English)*, 2012, 51, 11571-11575.
102. Wang, P., S. Dong, J.-H. Shieh, E. Peguero, et al., Erythropoietin Derived by Chemical Synthesis. *Science (New York, N.Y.)*, 2013, 342, 1357-1360.
103. Meynial-Salles, I. and D. Combes, In vitro glycosylation of proteins: an enzymatic approach. *J Biotechnol*, 1996, 46, 1-14.
104. Sudo, M., Y. Yamaguchi, P.J. Späth, K. Matsumoto-Morita, et al., Different IVIG Glycoforms Affect *In Vitro* Inhibition of Anti-Ganglioside Antibody-Mediated Complement Deposition. *PLoS One*, 2014, 9, e107772.
105. Piontek, C., P. Ring, O. Harjes, C. Heinlein, et al., Semisynthesis of a Homogeneous Glycoprotein Enzyme: Ribonuclease C: Part 1. *Angewandte Chemie International Edition*, 2009, 48, 1936-1940.
106. Piontek, C., D. Varón Silva, C. Heinlein, C. Pöhner, et al., Semisynthesis of a Homogeneous Glycoprotein Enzyme: Ribonuclease C: Part 2. *Angewandte Chemie International Edition*, 2009, 48, 1941-1945.
107. Zou, G., H. Ochiai, W. Huang, Q. Yang, et al., Chemoenzymatic synthesis and Fc $\gamma$  receptor binding of homogeneous glycoforms of antibody Fc domain. Presence of a bisecting sugar moiety enhances the affinity of Fc to Fc $\gamma$ IIIa receptor. *J Am Chem Soc*, 2011, 133, 18975-18991.
108. Wang, L.-X., The Amazing Transglycosylation Activity of Endo- $\beta$ -N-acetylglucosaminidases. *Trends in glycoscience and glycotechnology : TIGG*, 2011, 23, 33-52.

109. Wang, L.X. and J.V. Lomino, Emerging technologies for making glycan-defined glycoproteins. *ACS Chem Biol*, 2012, 7, 110-122.
110. Huang, W., J. Giddens, S.-Q. Fan, C. Toonstra, et al., Chemoenzymatic Glycoengineering of Intact IgG Antibodies for Gain of Functions. *J Am Chem Soc*, 2012, 134, 12308-12318.
111. Giddens, J.P. and L.X. Wang, Chemoenzymatic Glyco-engineering of Monoclonal Antibodies. *Methods Mol Biol*, 2015, 1321, 375-387.
112. Belický, Š., J. Katrik, and J. Tkáč, Glycan and lectin biosensors. *Essays Biochem*, 2016, 60, 37-47.
113. Dan, X., W. Liu, and T.B. Ng, Development and Applications of Lectins as Biological Tools in Biomedical Research. *Medicinal Research Reviews*, 2016, 36, 221-247.
114. Croset, A., L. Delafosse, J.-P. Gaudry, C. Arod, et al., Differences in the glycosylation of recombinant proteins expressed in HEK and CHO cells. *J Biotechnol*, 2012, 161, 336-348.
115. Rohrer, J.S., L. Basumallick, and D. Hurum, High-performance anion-exchange chromatography with pulsed amperometric detection for carbohydrate analysis of glycoproteins. *Biochemistry. Biokhimiia*, 2013, 78, 697-709.
116. Yang, Y., F. Liu, V. Franc, L.A. Halim, et al., Hybrid mass spectrometry approaches in glycoprotein analysis and their usage in scoring biosimilarity. *Nature Communications*, 2016, 7, 13397.
117. Rosati, S., E.T.J. van den Bremer, J. Schuurman, P.W.H.I. Parren, et al., In-depth qualitative and quantitative analysis of composite glycosylation profiles and other micro-heterogeneity on intact monoclonal antibodies by high-resolution native mass spectrometry using a modified Orbitrap. *mAbs*, 2013, 5, 917-924.
118. Stavenhagen, K., H. Hinneburg, M. Thaysen-Andersen, L. Hartmann, et al., Quantitative mapping of glycoprotein micro-heterogeneity and macro-heterogeneity: an evaluation of mass spectrometry signal strengths using synthetic peptides and glycopeptides. *J Mass Spectrom*, 2013, 48, 627-639.
119. Alagesan, K., S.K. Khilji, and D. Kolarich, It is all about the solvent: on the importance of the mobile phase for ZIC-HILIC glycopeptide enrichment. *Analytical and Bioanalytical Chemistry*, 2016, 1-10.
120. Mysling, S., G. Palmisano, P. Hojrup, and M. Thaysen-Andersen, Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics. *Anal Chem*, 2010, 82, 5598-5609.
121. Kim, J.Y., S.K. Kim, D. Kang, and M.H. Moon, Dual lectin-based size sorting strategy to enrich targeted *N*-glycopeptides by asymmetrical flow field-flow fractionation: profiling lung cancer biomarkers. *Anal Chem*, 2012, 84, 5343-5350.
122. Woo, C.M., A.T. Iavarone, D.R. Spiciarich, K.K. Palaniappan, et al., Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact *N*- and *O*-glycopeptide discovery and analysis. *Nat Meth*, 2015, advance online publication.
123. Wuhler, M., M.I. Catalina, A.M. Deelder, and C.H. Hokke, Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J Chromatogr B Analyt Technol Biomed Life Sci*, 2007, 849, 115-128.
124. Saba, J., S. Dutta, E. Hemenway, and R. Viner, Increasing the productivity of glycopeptides analysis by using higher-energy collision dissociation-accurate mass-product-dependent electron transfer dissociation. *Int J Proteomics*, 2012, 2012, 560391.
125. Wu, S.-W., T.-H. Pu, R. Viner, and K.-H. Khoo, Novel LC-MS2 Product Dependent Parallel Data Acquisition Function and Data Analysis Workflow for Sequencing and Identification of Intact Glycopeptides. *Anal Chem*, 2014, 86, 5478-5486.

126. Frese, C.K., A.F.M. Altelaar, H. van den Toorn, D. Nolting, et al., Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. *Anal Chem*, 2012, 84, 9668-9673.
127. Go, E.P., G.S. Hewawasam, B.J. Ma, H.-X. Liao, et al., Methods development for Analysis of Partially Deglycosylated Proteins and Application to an HIV Envelope Protein Vaccine Candidate. *Int J Mass Spectrom*, 2011, 305, 209-216.
128. Kaji, H. and T. Isobe, Stable isotope labeling of *N*-glycosylated peptides by enzymatic deglycosylation for mass spectrometry-based glycoproteomics. *Methods Mol Biol*, 2013, 951, 217-227.
129. Sun, B., J.A. Ranish, A.G. Utleg, J.T. White, et al., Shotgun glycopeptide capture approach coupled with mass spectrometry for comprehensive glycoproteomics. *Mol Cell Proteomics*, 2007, 6, 141-149.
130. Zhang, H., X.-j. Li, D.B. Martin, and R. Aebersold, Identification and quantification of *N*-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotech*, 2003, 21, 660-666.
131. Parker, B.L., M. Thaysen-Andersen, N. Solis, N.E. Scott, et al., Site-specific glycan-peptide analysis for determination of *N*-glycoproteome heterogeneity. *J Proteome Res*, 2013, 12, 5791-5800.
132. Bosques, C.J., B.E. Collins, J.W. Meador, H. Sarvaiya, et al., Chinese hamster ovary cells can produce galactose- $\alpha$ -1,3-galactose antigens on proteins. *Nat Biotechnol*, 2010, 28, 1153-1156.
133. Jensen, P.H., N.G. Karlsson, D. Kolarich, and N.H. Packer, Structural analysis of *N*- and *O*-glycans released from glycoproteins. *Nat Protoc*, 2012, 7, 1299-1310.
134. Pabst, M., D. Kolarich, G. Poltl, T. Dalik, et al., Comparison of fluorescent labels for oligosaccharides and introduction of a new postlabeling purification method. *Anal Biochem*, 2009, 384, 263-273.
135. Royle, L., M.P. Campbell, C.M. Radcliffe, D.M. White, et al., HPLC-based analysis of serum *N*-glycans on a 96-well plate platform with dedicated database software. *Anal Biochem*, 2008, 376, 1-12.
136. Reiding, K.R., D. Blank, D.M. Kuijper, A.M. Deelder, et al., High-throughput profiling of protein *N*-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification. *Anal Chem*, 2014, 86, 5784-5793.
137. Knezevic, A., J. Bones, S.K. Kracun, O. Gornik, et al., High throughput plasma *N*-glycome profiling using multiplexed labelling and UPLC with fluorescence detection. *Analyst*, 2011, 136, 4670-4673.
138. Pabst, M., J.S. Bondili, J. Stadlmann, L. Mach, et al., Mass + Retention Time = Structure: A Strategy for the Analysis of *N*-Glycans by Carbon LC-ESI-MS and Its Application to Fibrin *N*-Glycans. *Anal Chem*, 2007, 79, 5051-5057.
139. Everest-Dass, A.V., J.L. Abrahams, D. Kolarich, N.H. Packer, et al., Structural features for distinguishing *N*- and *O*-linked glycan isomers by LC-ESI-IT MS/MS. *J Am Soc Mass Spectrom*, 2013, 24, 895-906.

## **Paper I – Relative vs absolute quantitation in disease glycomics**

**E.S.X. Moh**, M. Thaysen-Andersen, N.H. Packer. (2014). *Relative vs Absolute Quantitation in Disease Glycomics*. *Proteomics Clinical Applications*, 9(3-4), 368-382.

Paper I is a review of the current methodologies available with respect to the analysis of glycans released from proteins for glycomics analysis and is particularly focused on the quantitative aspects of the various methods, and the advantages and disadvantages of each method of choice. The relevance of relative and absolute quantitation in glycomics analysis is also discussed with respect to clinically relevant matters in this review.

Pages 32-46 of this thesis have been removed as they contain published material. Please refer to the following citation for details of the article contained in these pages.

Moh, E. S. X., Thaysen-Andersen, M., Packer, N. H. (2015). Relative versus absolute quantitation in disease glycomics. *Proteomics Clinical Applications*, 9(3-4), 368-382.

DOI: [10.1002/prca.201400184](https://doi.org/10.1002/prca.201400184)

## Thesis Aims

Having discussed the importance of protein glycosylation and the usefulness of gaining control of the glycosylation process by glycoengineering, this dissertation develops on three challenges of *in vitro* glycoengineering; the multi-enzymatic nature (aim 1), the expense arising from single use and need for subsequent purification (aim 2), and the application of *in vitro* glycoengineering on IgM therapeutic antibodies (aim 3).

### Thesis aims

1. Using the *in vitro* expression of glycosyltransferases to modify glycans on glycoproteins is a method currently used for glycoengineering, but the use of multiple enzymes in a sequential fashion poses technical challenges. A novel scaffold-mediated *in vitro* metabolic engineering method is designed, aiming to create a self-assembling, DNA guided multi-enzymatic protein scaffold that can spatially align multiple enzymes *in vitro* (such as for the possible creation of an “*artificial Golgi*”). – Chapter 2
2. The single use nature of applying external glycosyltransferases for *in vitro* modelling of protein glycosylation decreases the efficiency of glycoengineering. The possibility of using glycosyltransferases sequentially immobilised on a resin to glycosylate proteins is explored, targeting the reuse of the bound enzymes and the possibility of creating an “*artificial Golgi column*”. – Chapter 3
3. *In vitro* glycoengineering efforts have largely been channelled to the IgG biopharmaceutical research space, while other medically useful antibody classes have been unexplored. In collaboration with a biotech company, Patrys Ltd., an anti-cancer IgM antibody will be the target for *in vitro* glycoengineering to develop a way to use the IgM as a diagnostic antibody for cancer cell detection. – Chapter 4 & 5

**Chapter 2: *In vitro* protein glycosylation using scaffold mediated metabolic engineering**

A novel method of addressing multi-enzymatic reactions *in vitro* is described. By employing principles from protein scaffolding, a self-assembling DNA guided protein scaffold is designed for *in vitro* glycoengineering with multiple glycosyltransferases.



## Introduction

### *Benefits of in vitro metabolic engineering*

Many biological processes within the cell involve multiple enzymes. From a yield-driven perspective of a metabolic engineer, optimising the product formation *in vivo* often involves introducing novel enzymes, disrupting/knocking out competing/conflicting pathways to create a favourable bias towards product formation [1, 2]. Metabolic flux analysis, high-throughput screening methods and large genetic libraries are commonly part of this strategy for finding the one clone that would become the workhorse for the desired product production, whilst being viable enough to replicate continuously and pass on the identical genetic traits.

A simplified alternative to *in vivo* metabolic engineering is to work in a controlled *in vitro* format, where the individual enzymes involved in the product formation are recombinantly expressed and purified [2], and applied sequentially. While this can also be a highly laborious process, it requires less technical expertise to express, purify and characterise individual enzymes, and provides the capacity to regulate the biochemical properties of each enzyme. For this process to become cost effective, the balance between a “low cost” primary substrate and the shortest possible enzymatic pathway is required.

Recent developments in cell-free protein synthesis procedures come as a bridge between the *in vivo* and *in vitro* format. Cell-free protein synthesis bypasses the experimental procedures required for the transformation/transfection of the enzyme containing genetic construct and instead synthesizes the enzymes using the protein production machinery of the cell extracts [4]. Depending on the source of the cell extract (bacterial, yeast, insect, mammalian etc.), the genetic elements of the DNA encoding material (promoter type, DNA/mRNA format), the yield and the post-translational modifications such as glycosylation and disulphide bond formation will result in differing protein products. In addition, starting from a plasmid DNA encoding for

## Chapter 2

the enzyme, it would require at least 2 days for conventional *in vivo* methods (transform/transfect cells, induce protein expression, harvest) to achieve a crude enzyme mixture as compared to cell-free synthesis that only require a few hours [4]. By decreasing the amount of time required to obtain small quantities of enzymes for *in vitro* tests, the selection of the right enzyme combinations for possible subsequent *in vivo* expression can also be streamlined.

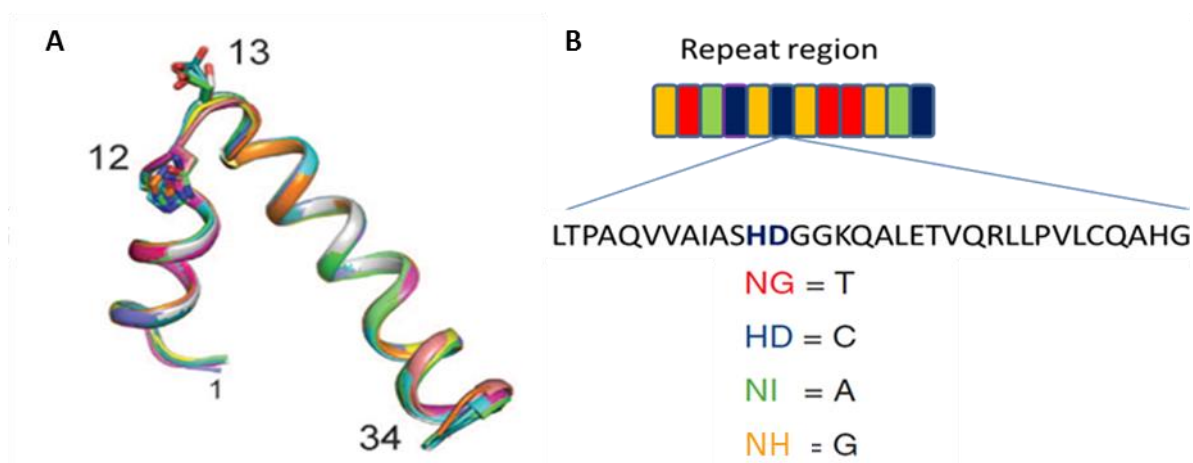
### *Substrate channelling*

Substrate channelling describes the process where sequential enzymes of a pathway pass the intermediate products through a line of enzymes at close spatial proximity, thereby increasing the reaction speed of the whole enzymatic pathway [5-7]. This is commonly achieved through induced close proximity between enzymes via a common interacting protein or location [8]. One naturally occurring example can be found in the MAP-Kinase pathway, where multiple kinases are normally anchored onto a scaffold protein, STE5, and phosphorylation of the first kinase rapidly triggers the phosphorylation of subsequent kinases [9]. For biochemical pathways with multiple enzymes, substrate channelling is a desired outcome that improves overall reaction kinetics which results in faster product formation.

To mimic and recreate micro-environments where substrate channelling can occur, several methods have been developed [5, 10-17]. These include localising the targeted enzymes into a specific location such as a bacterial micro-compartment [18], liposomes and synthetic vesicles [16], or tethering a protein domain that is capable of self-assembly/targeted interaction as a scaffold to bring different enzymes next to each other [5, 8, 9, 11, 15, 19]. Apart from anchoring onto lipids and other vesicles, anchoring onto DNA via sequence-specific DNA binding proteins has also been exploited for these scaffolding purposes [11, 17] and benefits from the high specificity of the sequence-specific DNA binding proteins and the low cost of synthesis of

designer DNA sequences. The sequence-specific binding properties of DNA binding proteins enable spatial alignment of the enzymes based on the corresponding target DNA sequences. One of the more prominent examples of such a sequence-specific DNA binding protein is the transcription activator-like effectors (TALEs). TALEs were originally found in the plant colonising bacteria *Xanthomonas sp.*, for use in modulating the gene transcription levels in their host. The sequence specific binding properties come from the 34 amino acid helix-loop-helix repeats (Figure 1, A) where specific amino acids in the 12<sup>th</sup> and 13<sup>th</sup> position determine the binding nucleotide, and by combining specific TALE monomers together, sequence specific targeting can be achieved (Figure 1, B) [20-24].

The modularity of these systems is highly valued in synthetic biology as it enables streamlined assembly methods of the corresponding DNA sequences to create custom, sequence-specific DNA binding proteins relatively easily [25-27]. Using TALEs as a sequence specific binding domain of a fusion protein has been proven previously as a sequence-specific nuclease by fusing a nuclease (TALEN) against a target gene [20, 25, 28-33]. Similar work was performed using multiple sequence-specific zinc-finger proteins [34], but due to proprietary rights on zinc-fingers [35] and relative ease of customising TALEs, TALEN became the more popular genome



**Figure 1.** Schematic representation of TALE repeat region adapted and modified from [21, 24]. Each TALE monomers comprises of a 34 amino acid sequence that folds into the helix-loop-helix motif, where the 12<sup>th</sup> and 13<sup>th</sup> amino acid (RVD, repeat variable diresidue) at the loop region varies between each monomer (overlay of 20 TALE monomer from [24], dependant on the nucleotide it is specific towards (A)). By discovering the RVD specific for each of the 4 nucleotides (represented by the 4 colours), combining the TALE monomers in a user defined fashion enables the targeting of specific sequences of DNA (B).

editing tool until CRISPR-Cas9 was discovered. Apart from fusing with nucleases, transcription factors and transcription inhibitors have also been coupled with TALEs to work as gene specific activators or inhibitors [20, 22]. The idea of coupling enzymes to sequence-specific DNA binding proteins first surfaced in 2010, at the international Genetically Engineered Machine (iGEM) competition, an international synthetic biology competition for undergraduates (at which I represented Macquarie University, as a member in 2011, and as team advisor in 2012-2016), presented by a team of students from Slovenia (<https://2010.igem.org/Team:Slovenia>). The team created a series of zinc-finger proteins fused to 4 different enzymes to target a specific series of DNA sequences that self-assembled *in vivo* in *E. coli* to produce the coloured compound violacein. This work was subsequently published, showing the production of three other metabolic products, resveratrol, 1,2-propanediol and mevalonate, by this approach [11]. The application of this system was also shown to work *in vitro* by a separate group in Korea, who improved the yield of L-threonine production in *E. coli* [17]. A parallel approach using TALE fusion proteins instead of zinc-finger proteins was reported by a more recent iGEM team, NUDT China 2015, to produce indole-3-acetic acid from L-tryptophan ([http://2015.igem.org/Team:NUDT\\_CHINA](http://2015.igem.org/Team:NUDT_CHINA)).

For the application of this approach to the glycosylation pathway which occurs naturally in the Golgi apparatus, specifically the trans-Golgi network (TGN), a diverse population of glycosyltransferases are necessary for the synthesis of the glycans on each glycoprotein [36]. The majority of these transferases are type II transmembrane proteins, anchored in the TGN via an *N*-terminal transmembrane domain, and the sequential reaction order of the enzymes is relatively well defined. For example, on an *N*-glycan, a galactose can only be attached after the antennary GlcNAc had been added; similar for the terminal sialic acid on the galactose.

Applying the synthetic DNA-protein binding scaffold could possibly improve and control the efficiency of the glycosylation pathway, and has the potential for creating an artificial Golgi system for *in vitro* glycoengineering. For this purpose, in this thesis a synthetic DNA-protein scaffold using TALEs with a monomeric streptavidin (mSA) was designed.

*Experimental approach used for the design of a synthetic DNA-protein scaffold for glycosylation enzyme attachment*

In all the previous work on DNA aligning protein scaffolds [11, 17], the DNA binding domain was fused directly to each enzyme. This means that for every enzyme involved, a new fusion protein would have to be created. To bypass this issue here, mSA was selected as a fusion partner to the TALE protein, thus creating a general fusion protein for any biotinylated or streptavidin-tagged protein to be aligned onto a defined DNA sequence. This TALE-mSA fusion protein will be referred to henceforth as the “adaptor protein”.

mSA is an engineered variant of streptavidin, which is naturally a tetramer [12, 37]. Using streptavidin (tetramer) and rhizavidin (dimer) as a base, several amino acids were replaced around the binding pocket and dimerisation interface, and non-essential variable loops were truncated, resulting in a soluble and stable monomeric streptavidin variant. The monomeric nature of mSA enables a 1:1 stoichiometric binding between a tagged enzyme and the adaptor protein, which is essential for controlling the relative abundance of enzymes in the pathway. For the enzymes attached to the adaptor protein, mono-biotinylation can be achieved by expressing with a BirA ligase tag for site specific biotinylation [38], or by protein *N*-terminal selective chemical conjugations [39]. Alternatively, a streptavidin purification tag (strep-tag) could also be used to attach the enzyme to the mSA of the adaptor proteins.

To aid in the visualisation of the overall design concept, a schematic of the final design is shown in Figure 2. Glycosyltransferases involved in the desired glycosylation outcome will first be

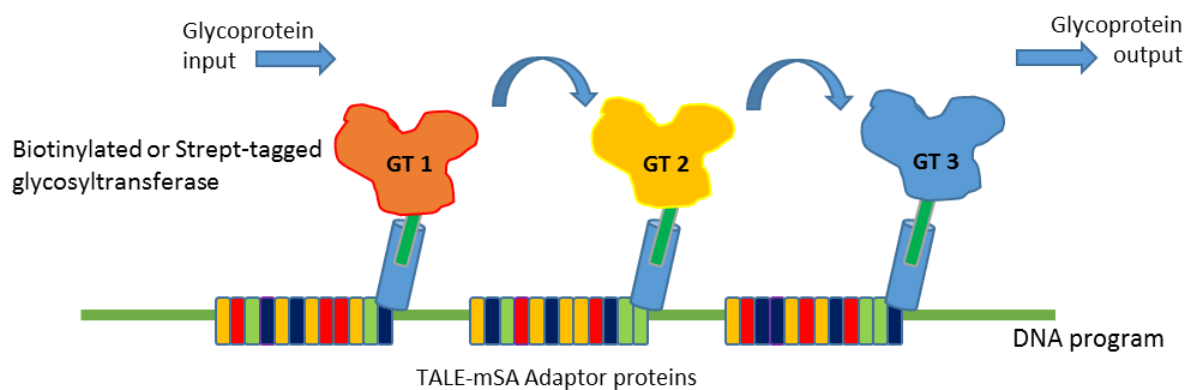


Figure 2. Schematic representation of how the final design would function. Biotinylated or Strep-tagged glycosyltransferases (GT) would be first attached to their individual adaptor protein, and subsequently aligned onto the DNA program. The glycoprotein precursor will then be introduced to the complexed DNA program for *in vitro* glycoengineering.

individually expressed and purified. The Resource for Integrated Glycotechnology has established a repository of glyco-enzyme expression constructs (<http://glycoenzymes.ccruc.uga.edu/>), where individual human glycosyltransferases are cloned into plasmids for transfection purposes. Multiple construct designs are available, one of which is the His-strep-tag purification tag that can be directly applied to this design. By coupling the glycosyltransferase with their individual adaptor protein, the complex can be aligned onto a customised DNA sequence (DNA program) based on reaction order for scaffolding purpose.

A total of four different TALE-mSA were designed, each TALE targeting a unique 9 base pair sequence, CAGAATGAG, ACTTGGCAA, ACGCGTGAA and TTCCAGCGA, that were selected based on the corresponding amino acid codon, QNE (1), TWQ (2), TRE (3) and FQR (4), for naming purposes only. While previous TALE fusion proteins were made to target longer sequences (between 12-25) [11, 20, 21, 24], the adaptor proteins used here were designed for *in vitro* use only on a 125 base pair custom DNA sequence, disregarding the need for high sequence specificity. Each adaptor protein has an approximate molecular weight of 46kDa; the molecular weight was kept low to increase the chance of proper folding. Between the TALE and mSA, three glycine residues (GGG) were added to achieve some structural flexibility based on the zinc-finger protein scaffolds that had a linker of ten amino acids (GGSGGGSGGS) [11],



Figure 3. A theoretical protein 3D model of the adaptor protein using RaptorX, an online freeware that predicts protein 3 dimensional structures by sequence homology from other known structures [3]. The TALE domain is coloured pink and the mSA domain yellow.

which was reduced in this case to decrease the degrees of freedom. A theoretical protein structure was generated using RaptorX (Figure 3), a free online protein structural prediction tool, based on protein threading and homology modelling [3], to approximate how the protein folds. However, as the basis of RaptorX is by modelling against known protein structures, and as both domains of the adaptor protein are known, the predicted output will simply be the addition of both known structures in a single protein chain, and might not truly reflect how the protein finally folds.

On the custom DNA sequence, a number of design features were also incorporated (Figure 4). A 125 base pair custom DNA sequence was designed, and will be referred to as the DNA program. The DNA program was designed with four 9 base pair target DNA binding sites, with an 11 base pair spacer between each binding site. As the double stranded DNA exists in a helical format, the twists in the DNA would be expected to change the axial alignment of the bound adaptor proteins. Since the twists in the double stranded DNA re-align approximately every 10-10.5 base pairs [40], the length of the spacer plus DNA binding site should be in multiples of 10-10.5, to ensure proper axial alignment. Every 10 base pairs of DNA have a length of approximately 34 Å, or 3.4 nm. With the design of the adaptor proteins to target 9 base pair sequences, if a one nucleotide spacer was used, there might be steric hindrance between the

## Chapter 2



Figure 4. Sequence features of the DNA program. The targeting sites for the 4 scaffold adaptors are marked in larger font and bold. In the 2 red boxes are the NsiI and PstI cut sites that can be used to create a longer DNA program. Random sequences at the 3' end was included as the minimum length of double stranded DNA that could be commercially purchased was 125 bp.

adaptor proteins or between the enzymes that attach to the scaffold, hence 11 base pair spacers were selected to create a total 20 base pair gap. This is similar to previous zinc finger scaffolds designed as 12 base pair target with 8bp spacers [17], and 9 base pair targets with 12 base pair spacers [11]. Spacers that were shorter and resulted in axial off alignment (9 base pair target with 4 base pair spacer) [11], or longer (12 base pair targets with 18 or 28 base pair spacers) [17], were shown to have a reduced yield.

Apart from the four DNA binding sites, restriction endonuclease sites were also included in the DNA program to facilitate increasing the number of enzymes that could be anchored on to the DNA program, while keeping the number of adaptor proteins the same. On the 5'-end of the DNA program is a NsiI cut site, while the 3'-end has a PstI cut site (Figure 4, red boxes). The two enzymes create compatible overhangs (TGCA) that ligate to ATGCAG, a site that has no known restriction endonuclease. Primers for the DNA program can potentially also be functionalised for immobilisation of the DNA on a surface to act as a stationary phase for streamlined removal of enzymes from products/substrates.

### Materials and Methods

#### *Insertion of scaffold DNA into pET-15b*

Protein sequences coding for individual TALEs specific for the four different nucleotides [22] and for mSA [37] obtained from published data [22, 37] and were combined to create the four scaffold adaptor proteins, all with a 6x-His tag at the *N*-terminus and a Gly-Gly-Gly linker in-between (sequence information in Appendix III). Codon optimisation for *E. coli* expression was



performed using the vendor's (Genscript) in-house algorithm at the time of purchase. Primers targeting the 6x-His tag and the C-terminus of the protein was used to introduce a 5' NcoI site and a 3' BamHI site for cloning into the pET-15b expression vector. To increase the efficiency of the insertion, an NdeI linearisation (only targeting non-insertions) was performed prior to transformation. Transformation of the *E. coli* BL21 (DE3) was performed via heat shock at 42 °C for 35s. Colony selection was bypassed by screening the total plasmid population from a liquid culture directly from the transformation. Plasmids were isolated using GenElute Miniprep kit (Sigma Aldrich, Australia), and were screened by restriction mapping using EcoRI and XbaI restriction enzymes. If the intensity of the band corresponding to successful insertion was lower than that of a non-insertion plasmid, the plasmids were linearised by NdeI and transformed again. Otherwise, the plasmids were re-transformed and 2-3 single colonies were selected for confirmation of the insertion.

#### *Protein expression and Ni-NTA purification*

Protein induction was performed using IPTG induction (1mM) when OD600nm of the culture was between 0.4-0.6, at room temperature for at least 16 hours. The cells were harvested by centrifugation and resuspended in 10% (v/v) of the original culture volume of 10mM Tris-HCl, pH 7.5, 0.5M NaCl, 20mM imidazole (wash buffer). Cell lysis was carried out by probe sonication (2X 10s at 60% amplitude), and the "soluble fraction" was collected after centrifugation at 5000g for 30 mins. The pellet was resuspended in the wash buffer containing 4M urea (denaturing buffer), and the "insoluble fraction" was collected after clarification by centrifugation at 14,000g for 20 mins. Purification of the adaptor proteins was performed using Ni-NTA magnetic beads (for small scale cultures of 50ml) (Genscript) or His-Gravitrapp (for larger cultures of 200ml) (GE Healthcare). The Ni-NTA beads were pre-equilibrated with the fraction respective buffer prior to sample loading. For the insoluble fraction, after sample loading, the sample was washed with 10 column volumes of the denaturing buffer, followed by

## Chapter 2

a step decrease in urea concentration by passing through 2 column volumes of wash buffer containing decreasing concentrations of urea (3M urea, 2M urea, 1M urea, no urea). This on-column refolding method was adapted from [41], for use on a gravity flow column. The His-tagged proteins were washed with another 5 column volumes of wash buffer before elution with 3 column volumes of wash buffer containing 300mM imidazole, and presence of adaptor proteins was verified by SDS-PAGE and mass spectrometry.

### *Adaptor protein identification by RP-LC-MS*

From the SDS-PAGE gel, the 46kDa band from the His-tag purification eluate corresponding to the molecular weight of the adaptor protein was excised for protein identification by mass spectrometry. The gel pieces were prepared using standard protocols for protein *in-gel* digestion as described in Paper II [42]. In brief, the gel pieces were destained, reduced with 10mM DTT and alkylated with 25mM IAA, and digested with 1µg of trypsin (Promega) overnight at 37°C. Digested peptides were then extracted, dried and reconstituted in 10µl of de-ionised water for loading onto the RP-LC-MS/MS. Theoretical tryptic peptide masses generated *in silico* were used to verify the identity of the adaptor protein.

### *Size exclusion chromatography*

The His-tagged adaptor proteins were then concentrated 3 fold by evaporation in a Speedvac before loading onto a SuperDex 200 10/300GL size exclusion column (GE Healthcare) for size separation. The column was equilibrated with phosphate buffered saline (PBS tablets, Amresco) prior to sample loading. Fractions (1ml) were collected during the elution window of 0.3 to 1 column volume, monitoring the absorbance at 215nm, 260nm and 280nm, and fractions corresponding to peak signals were concentrated using a 10 kDa cut-off centrifugal filter (Merck Millipore) to approximately 100 µl volume. An SDS-PAGE (of 10% of the sample) was performed to determine the purity of the fraction.

### *Circular Dichroism*

Circular dichroism (CD) was performed using a Jasco 810 circular dichroism spectrometer, flushed with nitrogen at 10L/min. The protein sample (approximately 0.15mg/ml, dissolved in PBS), was loaded into a quartz cuvette with a 1mm path length. The CD spectrum was measured at wavelengths from 190-260nm, with a spectral bandwidth of 1nm, at 100nm/min, at room temperature. The spectrum acquired was averaged over 4 measurements.

### *Electrophoretic mobility shift assay (EMSA)*

The 125 base pair DNA program was synthesized commercially (IDT Australia) and amplified by PCR in a 25µl reaction volume. 10µl of the adaptor protein containing fraction, verified by SDS-PAGE, was added to 1µl of PCR product and incubated at room temperature for 10 mins. The samples were mixed with 2µl of 6x gel loading dye (NEB) and loaded into a 2% (w/v) agarose gel for gel electrophoresis at 60V for 1.5 h. The gel was then post-stained with GelRed for 30 mins for visualisation under UV.

## **Results**

### *Selection by linearisation of plasmid DNA*

Colony selection without the use of reporter markers can be a very tedious and chance based process. While reporters greatly simplify the selection process, it can create unnecessary production of fluorogenic or chromogenic products that are not required for other downstream activities. pET-15b contains a NcoI restriction site at the 5' end of the intrinsic His-tag, an NdeI restriction site at the start of the multi-cloning site, and a BamHI site at the 3'-end of the cloning site. By insertion of the target gene using NcoI/BamHI restriction sites, successful insertions would abolish the NdeI site. Selection by linearisation is not a common selection method, but

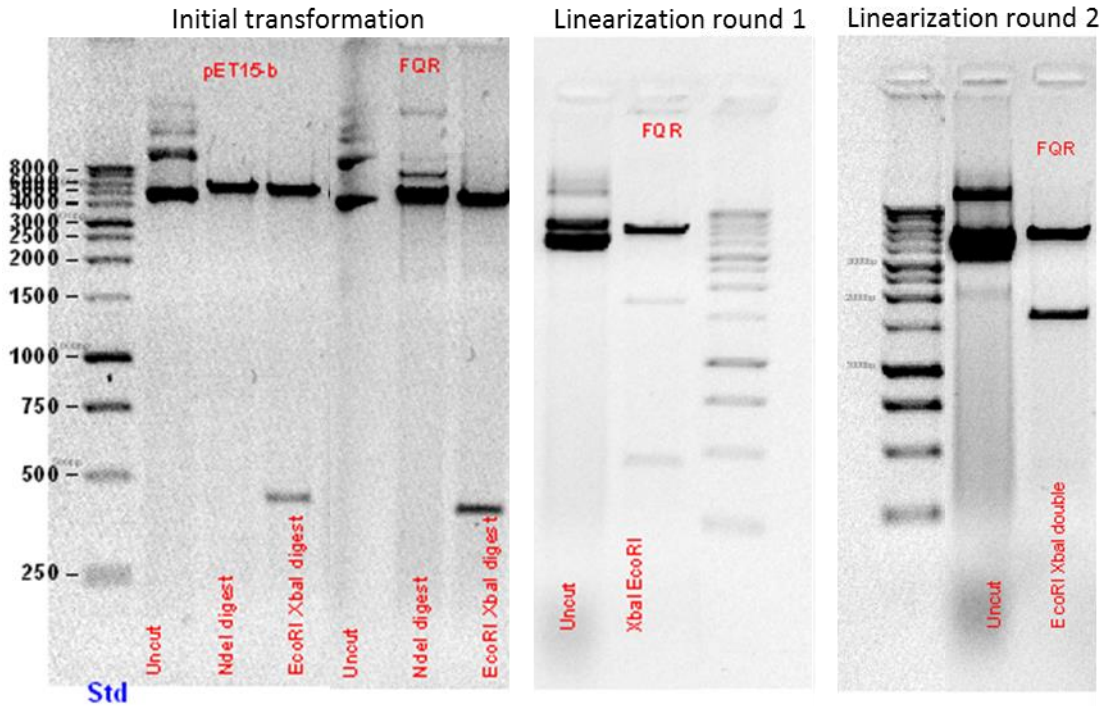


Figure 5. Agarose gel electrophoresis of the plasmids undergoing verification by restriction mapping. A non-insertion plasmid would have a 430 base pair fragment, and a successful insertion would have a 1.7 k bp fragment.

it worked well in this case. By screening the whole plasmid population at once instead of screening multiple individual clones, the chance based process of screening colonies is eliminated. *E. coli* does not take up linear exogenous DNA as efficiently as circular plasmid DNA [43], hence, linearisation of the non-insertion plasmids creates a bias in the transformation process (Figure 5). From the initial screening of the total plasmid population, while the desired band size (approx. 1.7kbp  $\rightarrow$  1.3kbp gene product + 430bp flanking sequence to the EcoRI/XbaI site) could not be seen with the EcoRI/XbaI digest, the presence of multiple bands in the NdeI digest indicated the presence of the inserted plasmid. Following 2 rounds of transformation with linearisation, the band intensity showing the insertion of the target gene was greatly increased, reducing the number of colonies that would be required for screening to select for the successful insertion.

### Identification of scaffold adaptor proteins in the inclusion bodies

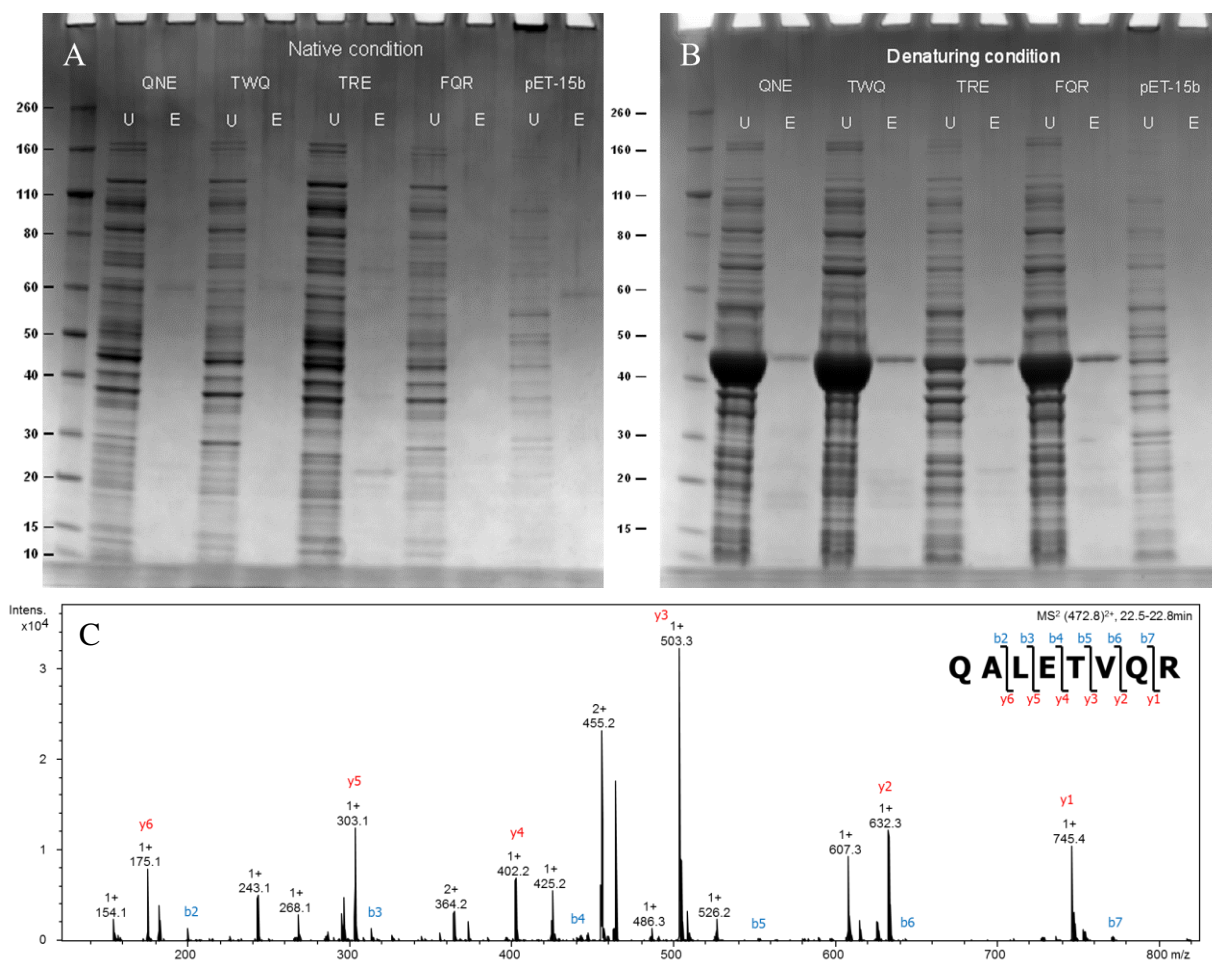


Figure 6. His-tag purification of the 4 scaffold adaptor proteins (QNE, TWQ, TRE, FQR, pET-15b as empty vector) in native (soluble fraction, **A**) and denaturing (insoluble fraction, **B**) conditions, using Ni-NTA magnetic beads (U – unbound proteins, E – elution fraction). The single band found in the elution fraction of the denaturing condition corresponded to the correct size of the TALE-mSA adaptor protein of approximately 46kDa. While a lot of the adaptor proteins were found in the unbound fraction of the denaturing condition, it was due to the saturation of the Ni-NTA magnetic beads (for 100 $\mu$ l of magnetic beads used, it had a capacity of approx. 60-200 $\mu$ g of protein). (**C**) Unique peptide from the TALE domain of the adaptor protein was identified by RP-LC-MS/MS, verifying that the purified protein was indeed the adaptor protein.

Based on a previous report of the developers of the mSA approach, it was found that expression of mSA as an individual protein in *E. Coli* forms inclusion bodies [44]. This problem was not present when mSA was expressed as a fusion protein with soluble proteins such as maltose binding protein and glutathione-S-transferase [45]. Expression of TALEs in *E. coli* has been previously reported and the TALEs were found as soluble proteins and not in inclusion bodies [21, 24]. It thus was surprising that, when expressed together as TALE-mSA, the resulting fusion protein was expressed as inclusion bodies in *E. Coli* as evidenced by the appearance of the correct molecular mass only after denaturation (Figure 6, B); the pure protein band at 46kDa

was only observed in the eluate under denaturing conditions. This was found to occur in both small (50ml) and large (200ml) scale inductions. The identity of the 46kDa band was further verified by RP-LC-MS/MS (Figure 6, C) where the unique peptide QALETVQR (residue 17-24 of the 34 amino acid TALE monomer) was found by manually searching against theoretical tryptic peptides of the adaptor proteins generated *in silico*.

#### *On-column re-folding of the adaptor proteins verified by circular dichroism*

On-column refolding is a procedure where the environment of a denatured protein is gradually changed from a denaturing to a non-denaturing environment, as opposed to the traditional refolding via dialysis under low protein concentrations [41]. The main aim of these methods relies on removing the solubilised denatured protein from an environment where aggregation, that promotes mis-folding, could occur. On-column refolding uses the immobilisation of each solubilised denatured protein to prevent interactions between each other, which in turn prevents aggregation. By gradually decreasing the concentration of the denaturant (urea in this case) in the environment, the immobilised proteins refold individually without interference from each other [41]. Analysis of protein secondary structure by circular dichroism is a quick way of

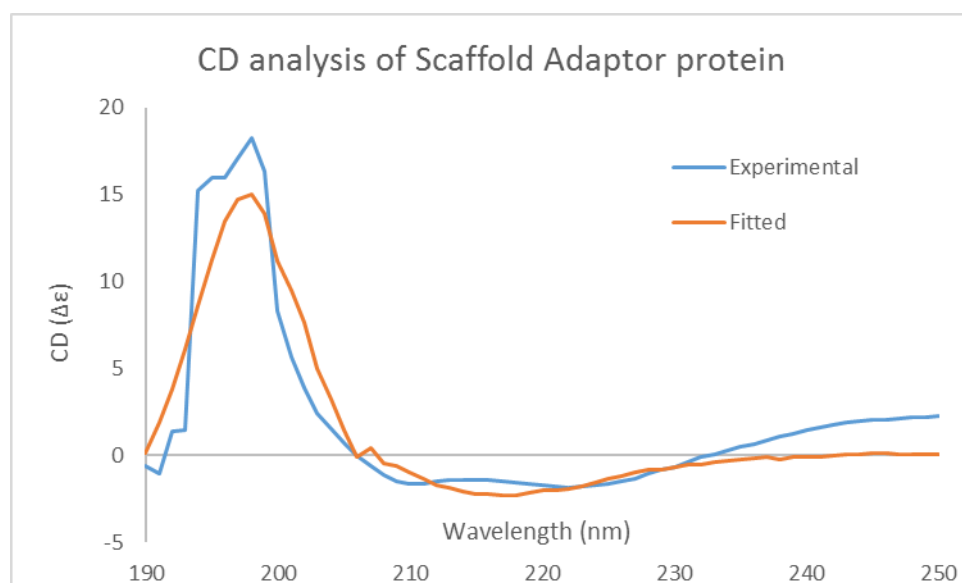


Figure 7. CD spectrum after fitting using the online webtool BeStSel. The experimental data (blue) is a representative trace of scaffold adaptor protein FQR. The fitted CD spectrum estimated that the protein was 49.3%  $\alpha$ -helix, 20.2% anti-parallel and 12.2% parallel (which comes to 32.4%  $\beta$ -sheet structures).

estimating the relative proportions of  $\alpha$ -helix,  $\beta$ -sheet and random coils in a refolded protein [46, 47]. While the general CD spectra for  $\alpha$ -helices are highly consistent, CD spectra of  $\beta$ -sheet structures vary quite significantly, due to the presence of parallel and anti-parallel  $\beta$ -sheets, and are hard to distinguish by previous CD spectral fitting algorithms. A recent CD fitting tool that was developed 2 years ago (mid 2015), used pre-existing protein 3D structures obtained from X-ray crystallography and NMR to develop a new algorithm to solve this issue [47]. As of early 2017, it had already been cited 80 times (Google Scholar) and the fitting tool is freely available on the web (BeStSel, <http://bestsel.elte.hu/index.php>). Based on the CD spectrum fit using this tool (Figure 7), it was estimated that the scaffold adaptor protein had 49.3%  $\alpha$ -helix, 20.2% anti-parallel and 12.2% parallel  $\beta$ -sheet structure. For the 46kDa scaffold protein, the TALE domain which is pre-dominantly  $\alpha$ -helix makes up approximately 33kDa and the mSA domain which is a  $\beta$ -barrel makes up 13kDa. Within the TALE domain, approximately 75% of the polypeptide chain is in an  $\alpha$ -helix secondary structure, accounting for about 54%  $\alpha$ -helix of the TALE-mSA adaptor protein, which is quite similar to the CD spectrum fitting analysis. Correspondingly, the  $\beta$ -barrel structure of mSA accounts for approximately 28% of the whole protein, which is in relatively close agreement with the CD spectrum fitting result. This suggested that the protein was refolded by the on-column refolding procedure.

#### *Testing DNA binding capabilities of the expressed scaffold adaptors*

Electrophoretic mobility shift assay (EMSA) is a common method for testing of whether a protein has the capacity to bind to DNA [48]. While it is common to perform EMSA using polyacrylamide gels, higher percentage agarose gels (2%) have been used for resolving the mobility shift after protein binding. In the previous work by the 2010 Slovenia iGEM team, the binding of a 9 base pair targeting zinc-finger protein produced a significant mobility shift of a 70 base pair DNA (<https://2010.igem.org/Team:Slovenia/PROJECT/proof/studies/emsa>). Therefore, for the scaffold adaptor proteins produced here that target a 9 base pair sequence of

the 125 base pair DNA program, it could be expected that the mobility shift after binding should be resolvable on a 2% agarose gel. However, a shift in electrophoretic mobility was not observed, suggesting that the scaffold adaptors were not binding to the DNA program (Figure 8); migration of the 125bp DNA program was unchanged despite presence of the adaptor proteins, indicating that no DNA-protein interaction occurred. While changing the binding conditions of temperature, buffer ionic strength and salt additives to stabilise the complex have been used for binding optimisation [48], this was not attempted here as the final use of the scaffold adaptor was designed for anchoring of enzymes which require physiological conditions for activity.

An alternate approach was attempted to verify the DNA binding capability of the adaptor proteins by separating the DNA-protein complex using size exclusion chromatography. Five major peaks were observed from the chromatographic separation of the DNA-protein mixture, and the fractions were collected, dried down and ran on both an agarose gel and SDS-PAGE to identify the elution of the DNA program and the scaffold adaptor proteins. Theoretically, if the adaptor proteins could complex with the DNA program, the adaptor proteins would be present in the same fraction as the DNA (Figure 9, peak 1). Unfortunately, the scaffold adaptors eluted

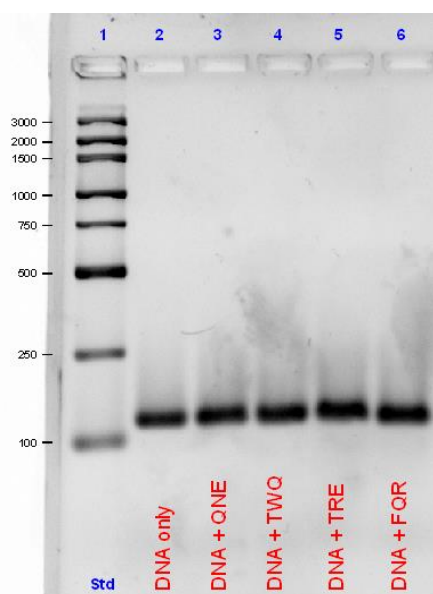


Figure 8. An electrophoretic mobility shift assay of each scaffold adaptor protein on the DNA program. No significant mobility retardation was observed for any of the scaffold proteins, suggesting that there was no DNA-protein interaction between the adaptors and the DNA program.



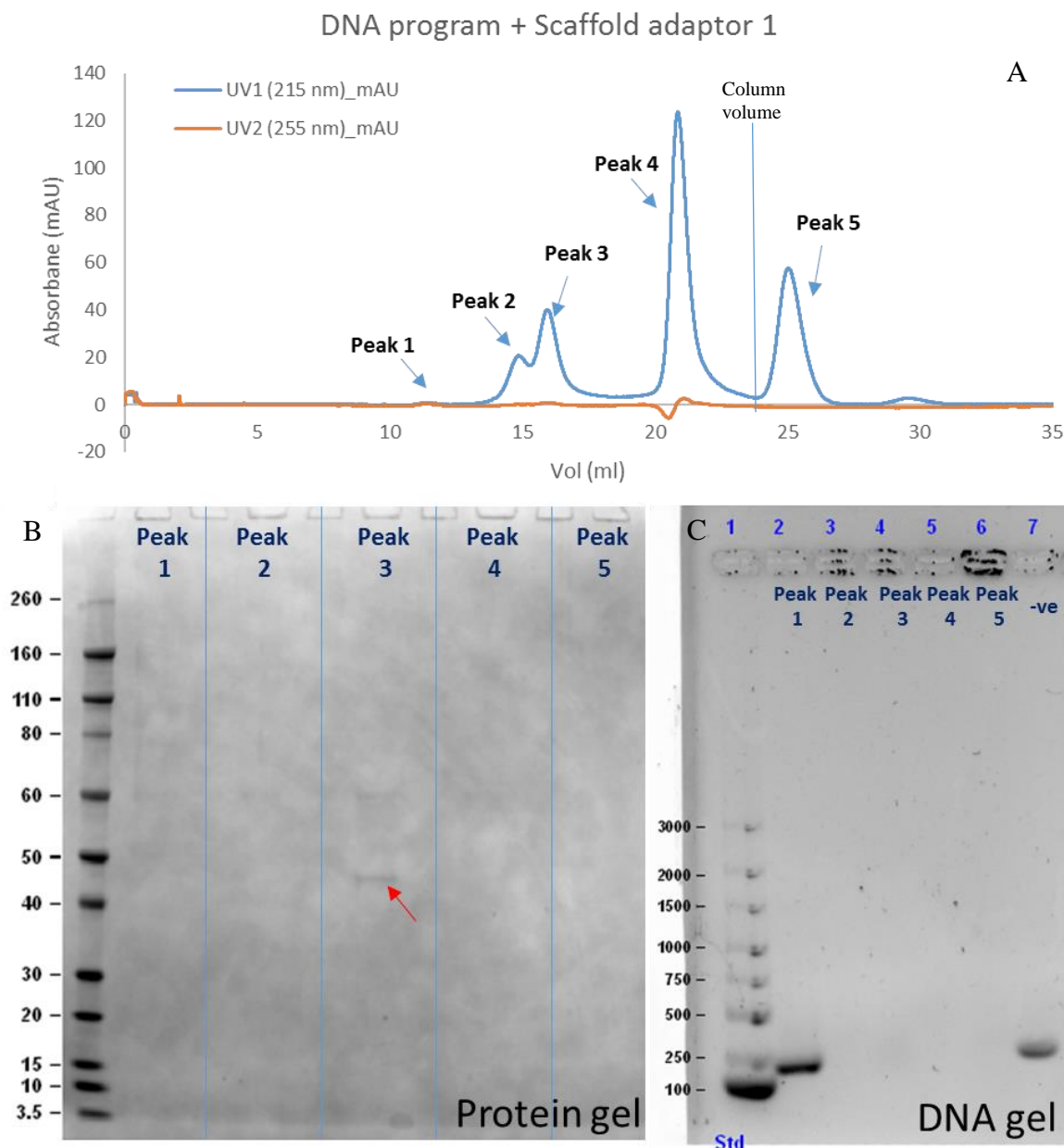


Figure 9. DNA-protein complex purification using Superdex200 size exclusion column chromatography using PBS as mobile phase. A: The protein signal was monitored at 215nm as the number of aromatic residues in the scaffold adaptors were very low for measurement at 280nm. B: From the observed peaks in A, the fractions were dried and run on an SDS-PAGE and an agarose gel for protein and DNA visualisation respectively. The red arrow marks the presence of the scaffold adaptor protein in peak 3, which was consistent with the size exclusion elution carried out during purification (data not shown). The absence of scaffold adaptor proteins in the peak 1, where the DNA elutes, suggests that there is no interaction between the two.

at the same retention volume in physiological PBS buffer as they did for their purification using the same column (data not shown) (Figure 9, peak 3), and no signs of the adaptor proteins were found in the same fraction as the DNA program (Figure 9, peak 1). Based on this data, it appeared that while the adaptor proteins were likely to have refolded, the TALE-mSA fusion

## Chapter 2

protein designed to target the 9 base pair DNA unfortunately did not have DNA binding capability.

### **Discussion**

From this entire process of designing gene to protein to function, many lessons were learnt; the most obvious being the incorrect assumption that it would be possible to create a gain/combination of function by simply expressing two functional protein domains together. Protein folding is still a process that baffles many; even with the vast amount of knowledge on how to express and purify a protein, the actual process of how a protein folds is still not fully understood [49, 50]. For the scaffold adaptors, many improvements can be made to the entire process in the future to increase the chance of creating a functional adaptor protein for anchoring and aligning enzymes onto a DNA program. These include:

In house genetic construction

One of the major lessons learnt was purchasing of the TALE-mSA genetic construct as a fully synthesized gene product (4x TALE-mSA), instead of purchasing the genes for individual TALE monomers (4 monomers targeting the 4 different nucleotide bases) and mSA to combine them in house via genetic engineering. The original idea of purchasing the synthesized construct was to save time and resources for the construction of the adaptor protein gene from scratch. One of the benefits of construction from scratch would be to allow for design changes such as increasing the number of TALE repeats to target a longer sequence, having a different linker between the TALE and mSA, changing the mSA to other high affinity binder pairs such as protein G/IgG-Fc, or adding localisation tags for periplasmic expression, many of which could have changed the resulting unsuccessful outcome.

*Avidity based binding of TALE to DNA*

The helix-loop-helix repeat motif of TALEs is not the same as the typical helix-turn-helix DNA binding motif. For the helix-turn-helix motif, the DNA interaction occurs inside the turn and between the helices [51], while the helix-loop-helix motif of TALEs interacts with the DNA in the loop region [21, 24]. Due to the larger interaction interface, helix-turn-helix motifs can target a longer sequence of DNA than the individual TALE helix-loop-helix motif. This raises the possibility that TALE-DNA interaction is avidity driven, and a certain level of avidity has to be achieved for TALE to strongly attach to the target the DNA sequence. The shortest recombinant TALE reported in literature targets 12 base pairs [21, 22], longer than the 9 base pairs designed in this work. Apart from being 3 TALE repeats longer, the 12 base pair targeting also represents that the TALE could completely wind around the double helix, which has twist repeats of approximately 10-11 base pairs [40]. This difference could have been the reason for the inability of the adaptor proteins to not bind the DNA strongly enough during the interaction experiments.

### *Folding of mSA*

The presence of TALE-mSA in the insoluble fraction which suggests the formation of inclusion bodies was an unexpected hurdle, as described previously. From the published work on mSA, it was reported to be soluble when expressed in yeast and humans [44, 45], even when expressed as a single protein. In *E. coli*, it required a soluble fusion partner to be expressed as a soluble protein. As part of the monomerisation process of mSA from streptavidin and rhizavidin, a single disulphide bridge was introduced to maintain the mSA structural integrity. The disulphide bridge, which could not be formed in the intracellular space of *E. coli*, could have resulted in the aggregation process in which the mSA domain failed to fold, thus forming inclusion bodies. The developers of this technique had to supplement with biotin during the refolding process of the individual mSA to aid in protein refolding [44], a procedure which would be counter-productive in this case as the bound biotin would have to be subsequently

## Chapter 2

removed to allow the binding of the tagged enzymes to occur. Periplasmic localisation could be a way to solve this issue in the future. Signal peptides that result in periplasmic localisation are known and the periplasm environment allows disulphide bond formation. A previously insoluble enzyme when expressed intracellularly was found to be produced at high yields and in a soluble format when expressed with the periplasmic localisation signal peptide [52].

### **Conclusion**

As described above, there are many ways to improve the design of the proposed scaffold adaptor proteins, some of which were attempted here, in order to design a sequential *in vitro* synthetic machine for the glycosylation of proteins i.e. a potential artificial Golgi. The scaffold adaptor protein idea and design of a self-assembling, multi-enzymatic protein scaffold by metabolic engineering for *in vitro* protein glycosylation as described here was not successful, although it shows future promise given more time and resources.

The following Chapter 3 describes an alternate approach of a simplified system where individual expressed glycosyltransferases are immobilised on a stationary phase resin to create the potential of an immobilised enzyme reactor (IMER) for *in vitro* glycoengineering.

## References

1. Yadav, V.G., M. De Mey, C.G. Lim, P.K. Ajikumar, et al., The Future of Metabolic Engineering and Synthetic Biology: Towards a Systematic Practice. *Metab Eng*, 2012, *14*, 233-241.
2. Guo, W., J. Sheng, and X. Feng, Mini-review: In vitro Metabolic Engineering for Biomanufacturing of High-value Products. *Computational and structural biotechnology journal*, 2017, *15*, 161-167.
3. Peng, J. and J. Xu, RaptorX: exploiting structure information for protein alignment by statistical inference. *Proteins*, 2011, *79*, 161-171.
4. Chong, S., Overview of Cell-Free Protein Synthesis: Historic Landmarks, Commercial Systems, and Expanding Applications. *Curr Protoc Mol Biol*, 2014, *108*, 16.30.11-16.30.11.
5. Dueber, J.E., G.C. Wu, G.R. Malmirchegini, T.S. Moon, et al., Synthetic protein scaffolds provide modular control over metabolic flux. *Nat Biotechnol*, 2009, *27*, 753-759.
6. Spivey, H.O. and J. Ovadi, Substrate channeling. *Methods*, 1999, *19*, 306-321.
7. Wheeldon, I., S.D. Minter, S. Banta, S.C. Barton, et al., Substrate channelling as an approach to cascade reactions. *Nat Chem*, 2016, *8*, 299-309.
8. Schoffelen, S. and J.C.M. van Hest, Multi-enzyme systems: bringing enzymes together in vitro. *Soft Matter*, 2012, *8*, 1736-1746.
9. Good, M.C., J.G. Zalatan, and W.A. Lim, Scaffold Proteins: Hubs for Controlling the Flow of Cellular Information. *Science*, 2011, *332*, 680-686.
10. Blackstock, D., Q. Sun, and W. Chen, Fluorescent protein-based molecular beacons by zinc finger protein-guided assembly. *Biotechnol Bioeng*, 2014, n/a-n/a.
11. Conrado, R.J., G.C. Wu, J.T. Boock, H. Xu, et al., DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. *Nucleic Acids Res*, 2012, *40*, 1879-1889.
12. Demonte, D., E.J. Drake, K.H. Lim, A.M. Gulick, et al., Structure-based engineering of streptavidin monomer with a reduced biotin dissociation rate. *Proteins*, 2013.
13. Mukai, C., M. Bergkvist, J.L. Nelson, and A.J. Travis, Sequential reactions of surface-tethered glycolytic enzymes. *Chem Biol*, 2009, *16*, 1013-1020.
14. Ono, Y., M. Kitajima, S. Daikoku, T. Shiroya, et al., Sequential enzymatic glycosyltransfer reactions on a microfluidic device: Synthesis of a glycosaminoglycan linkage region tetrasaccharide. *Lab Chip*, 2008, *8*, 2168-2173.
15. Zhang, Y., S.Z. Li, J. Li, X. Pan, et al., Using unnatural protein fusions to engineer resveratrol biosynthesis in yeast and Mammalian cells. *J Am Chem Soc*, 2006, *128*, 13030-13031.
16. Rother, C., B. Nidetzky, and M.C. Flickinger, Enzyme Immobilization by Microencapsulation: Methods, Materials, and Technological Applications, in *Encyclopedia of Industrial Biotechnology* 2009, John Wiley & Sons, Inc.
17. Lee, J.H., S.C. Jung, M. Bui le, K.H. Kang, et al., Improved production of L-threonine in *Escherichia coli* by use of a DNA scaffold system. *Appl Environ Microbiol*, 2013, *79*, 774-782.
18. Lawrence, A.D., S. Frank, S. Newnham, M.J. Lee, et al., Solution Structure of a Bacterial Microcompartment Targeting Peptide and Its Application in the Construction of an Ethanol Bioreactor. *ACS Synthetic Biology*, 2014, *3*, 454-465.

19. Conrado, R.J., J.D. Varner, and M.P. DeLisa, Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. *Curr Opin Biotechnol*, 2008, 19, 492-499.
20. Sanjana, N.E., L. Cong, Y. Zhou, M.M. Cunniff, et al., A transcription activator-like effector toolbox for genome engineering. *Nat Protoc*, 2012, 7, 171-192.
21. Mak, A.N., P. Bradley, A.J. Bogdanove, and B.L. Stoddard, TAL effectors: function, structure, engineering and applications. *Curr Opin Struct Biol*, 2013, 23, 93-99.
22. Cong, L., R. Zhou, Y.C. Kuo, M. Cunniff, et al., Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun*, 2012, 3, 968.
23. Boch, J., H. Scholze, S. Schornack, A. Landgraf, et al., Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, 2009, 326, 1509-1512.
24. Deng, D., C. Yan, X. Pan, M. Mahfouz, et al., Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science*, 2012, 335, 720-723.
25. Reyon, D., S.Q. Tsai, C. Khayter, J.A. Foden, et al., FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol*, 2012, 30, 460-465.
26. Uhde-Stone, C., N. Gor, T. Chin, J. Huang, et al., A do-it-yourself protocol for simple transcription activator-like effector assembly. *Biol Proced Online*, 2013, 15, 3.
27. Weber, E., R. Gruetzner, S. Werner, C. Engler, et al., Assembly of designer TAL effectors by Golden Gate cloning. *PLoS One*, 2011, 6, e19722.
28. Li, T., B. Liu, M.H. Spalding, D.P. Weeks, et al., High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat Biotech*, 2012, 30, 390-392.
29. Meng, X., M.B. Noyes, L.J. Zhu, N.D. Lawson, et al., Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol*, 2008, 26, 695-701.
30. Neville, E.S., C. Le, Z. Yang, M.C. Margaret, et al., A transcription activator-like effector toolbox for genome engineering. *Nat Protoc*, 2012, 7, 171-192.
31. Sun, N. and H. Zhao, Transcription activator-like effector nucleases (TALENs): A highly efficient and versatile tool for genome editing. *Biotechnol Bioeng*, 2013.
32. Zu, Y., X. Tong, Z. Wang, D. Liu, et al., TALEN-mediated precise genome modification by homologous recombination in zebrafish. *Nat Methods*, 2013, 10, 329-331.
33. Mussolino, C., R. Morbitzer, F. Lutge, N. Dannemann, et al., A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res*, 2011, 39, 9283-9293.
34. Durai, S., M. Mani, K. Kandavelou, J. Wu, et al., Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic Acids Res*, 2005, 33, 5978-5990.
35. Chandrasekharan, S., S. Kumar, C.M. Valley, and A. Rai, Proprietary science, open science and the role of patent disclosure: the case of zinc-finger proteins. *Nat Biotechnol*, 2009, 27, 140-144.
36. Varki, A., J.D. Esko, and K.J. Colley, *Cellular Organization of Glycosylation*, in *Essentials of Glycobiology*, A. Varki, et al. 2009: Cold Spring Harbor (NY).
37. Lim, K.H., H. Huang, A. Pralle, and S. Park, Stable, high-affinity streptavidin monomer for protein labeling and monovalent biotin detection. *Biotechnol Bioeng*, 2013, 110, 57-67.
38. Fairhead, M. and M. Howarth, Site-specific biotinylation of purified proteins using BirA. *Methods in molecular biology (Clifton, N.J.)*, 2015, 1266, 171-184.
39. Chan, A.O., C.M. Ho, H.C. Chong, Y.C. Leung, et al., Modification of N-terminal alpha-amino groups of peptides and proteins using ketenes. *J Am Chem Soc*, 2012, 134, 2589-2598.

40. Wang, J.C., Helical repeat of DNA in solution. *Proceedings of the National Academy of Sciences*, 1979, *76*, 200-203.
41. Lemercier, G., N. Bakalara, and X. Santarelli, On-column refolding of an insoluble histidine tag recombinant exopolyphosphatase from *Trypanosoma brucei* overexpressed in *Escherichia coli*. *J Chromatogr B*, 2003, *786*, 305-309.
42. Moh, E.S., C.H. Lin, M. Thaysen-Andersen, and N.H. Packer, Site-Specific *N*-Glycosylation of Recombinant Pentameric and Hexameric Human IgM. *J Am Soc Mass Spectrom*, 2016, *27*, 1143-1155.
43. El Karoui, M., P. Dabert, A. Gruss, and S.K. Amundsen, Gene replacement with linear DNA in electroporated wild-type *Escherichia coli*. *Nucleic Acids Res*, 1999, *27*, 1296-1299.
44. Lim, K.H., H. Huang, A. Pralle, and S. Park, Engineered streptavidin monomer and dimer with improved stability and function. *Biochemistry*, 2011, *50*, 8682-8691.
45. Demonte, D., C.M. Dundas, and S. Park, Expression and purification of soluble monomeric streptavidin in *Escherichia coli*. *Appl Microbiol Biotechnol*, 2014, *98*, 6285-6295.
46. Greenfield, N.J., Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc*, 2006, *1*, 2876-2890.
47. Micsonai, A., F. Wien, L. Kernya, Y.-H. Lee, et al., Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proceedings of the National Academy of Sciences*, 2015, *112*, E3095-E3103.
48. Hellman, L.M. and M.G. Fried, Electrophoretic Mobility Shift Assay (EMSA) for Detecting Protein-Nucleic Acid Interactions. *Nat Protoc*, 2007, *2*, 1849-1861.
49. Shakhnovich, E., Protein Folding Thermodynamics and Dynamics: Where Physics, Chemistry and Biology Meet. *Chem Rev*, 2006, *106*, 1559-1588.
50. Deng, N.-j., W. Dai, and R.M. Levy, How Kinetics within the Unfolded State Affects Protein Folding: an Analysis Based on Markov State Models and an Ultra-Long MD Trajectory. *J Phys Chem B*, 2013, *117*, 10.1021/jp401962k.
51. Struhl, K., Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins. *Trends in Biochemical Sciences*, 1989, *14*, 137-140.
52. Loo, T., M.L. Patchett, G.E. Norris, and J.S. Lott, Using secretion to solve a solubility problem: high-yield expression in *Escherichia coli* and purification of the bacterial glycoamidase PNGase F. *Protein Expr Purif*, 2002, *24*, 90-98.

### **Chapter 3 – *In vitro* protein glycosylation using immobilised glycosyltransferases**

A different approach towards *in vitro* glycoengineering of proteins by immobilising glycosyltransferases onto a stationary phase resin is presented with the aim of enzyme re-use and creation of an artificial Golgi column



## Introduction

Galactosylation is an important part of mammalian *N*-glycan biosynthesis as it forms the precursor monosaccharide substrate for sialyltransferases to add terminal sialic acid to glycan structures [1]. Sialic acids on glycoproteins have been shown to affect glycoprotein serum half-life, inflammatory response, and receptor binding [2-6]. Galactosylation of IgG1 was also shown to be involved in anti-inflammatory activity by promoting association of FcγRIIB with C-type lectin dectin-1 [7], and thereby enhancing FcγRIIIa binding affinity [8].

$\beta$ -1,4-galactosyltransferase 1 (B4GALT1) is one of the galactosyltransferase family of enzymes that transfers the galactose from UDP-galactose onto the GlcNAc of a growing glycan structure. B4GALT1 is a type II transmembrane protein that is localised in the trans-Golgi network (TGN), and over-expression of B4GALT1 results in increased galactosylation and subsequent sialylation of proteins expressed by CHO cells [9]. Enzymatic galactosylation of glycans *in vitro* using B4GALT1 had been reported by many [7, 8, 10-13], although no-one had considered the recycling of the B4GALT1 enzyme for re-use after purification of the product glycoprotein. The once-off use of glycosyltransferases in *in vitro* glycoengineering methods is one of the disadvantages that limits this approach in practice from a commercial perspective.

Immobilisation of enzymes onto a stationary phase such as micro- or nano-particles, resins or functionalised surfaces can enable the recovery of the enzyme for re-use after the reaction [14, 15] and potentially allows the construction of a reusable enzyme machine. These enzymes conjugated stationary phases are collectively termed immobilised enzyme reactors (IMER) and different methods such as passive physical adsorption, chemical conjugation and engineered affinity tags have been used for the immobilisation of enzymes [14-17]. By being able to re-use the conjugated enzymes, use of excess enzyme also becomes possible, increasing the overall reaction kinetics and lowering the reaction time. Immobilisation also provides an easy

purification procedure to remove the enzyme from the product, unlike conventional single pot reactions. It also enables the aligning of sequential enzymes of a reaction pathway with multiple IMER, either on- or off-line, to create multi-enzyme reactors [17, 18].

An attempt at using this type of enzyme immobilisation specifically for *in vitro* glycoengineering has been performed using a microfluidic device, creating a single function “Golgi-on-a-chip” that was able to add the sulphates onto heparin sulphate chains by D-glucosaminyl 3-*O*-sulfotransferase isoform-1 enzyme activity [19]. The heparin sulphate was first biotinylated and attached to a streptavidin conjugated magnetic particle. Oil droplets containing the enzyme were “moved” to merge with another droplet containing the heparin sulphate chains by the fluidics device for the reaction to occur. After the reaction, the heparin sulphate chains were held in place by the magnet and wash extensively to remove the enzyme [19]. This approach at creating an artificial Golgi apparatus is not applicable if the modified protein is desired as a standalone product; immobilising the target protein facilitates the analysis of the modification, but the product remains conjugated to the nanoparticle. Recently, a theoretical design for a multi-glycosyltransferase large scale reactor was published, detailing the possible effects of different physical parameters on the overall glycosylation “completeness” starting from a  $\text{Man}_5\text{GlcNAc}_2$  glycan to produce a bi-antennary di-galactosylated glycan [18]. While there is no experimental evidence on the actual feasibility of the design, it has provided models containing physical parameters that can be experimentally optimised to gain control of the glycosylation process, which is currently not possible in an *in vivo* setting.

As mentioned in Chapter 2, the Resource for Integrated Glycotechnology has set up a repository of glyco-enzyme expression constructs (<http://glycoenzymes.crcr.uga.edu/>), where plasmids containing individual human glycosyltransferases can be purchased from the DNA repository DNASU (<https://dnasu.org/DNASU/Home.do>). These constructs are available in different formats, varying by the purification tag (8x-His-Strep tag or 8x-His-AviTag), location of

purification tag (protein *N*- or *C*-terminus), and fusion domain (no fusion, GFP fusion, or GFP, IgG-Fc fusion). The construct containing the human B4GALT1 with *N*-terminal 8x-His-Strep tag was obtained for this work.

In the preliminary work presented here, we trialled the possibility of galactosylating a glycoprotein *in vitro* using a recombinantly expressed His-tagged human B4GALT1 immobilised onto Ni-NTA resin, as a proof of concept of creating a mini-Golgi column for *in vitro* designer protein glycosylation.

### **Materials and Methods**

B4GALT1 containing plasmid (HsCD00522338) was purchased from the DNASU plasmid repository, deposited by the Repository of Glyco-enzymes Expression constructs (<http://glycoenzymes.ccruc.uga.edu/>). Culture media and HEK293FT cells were purchased from ThermoFisher Scientific, Australia. Plasmid Midiprep kit was obtained from Promega, Australia. Ni-NTA Gravitrap purification columns were purchased from GE Healthcare. All other chemicals, otherwise specified, were obtained from Sigma Aldrich, Australia.

#### *Transient transfection of HEK293FT cells with B4GALT1 plasmid*

B4GALT1 plasmid (pGEN1-DEST) was cloned into competent *E. coli* cells for plasmid amplification, and purified using a midiprep kit. Transient transfection protocol was based on that supplied by the Glyco-enzymes repository [20], with modifications to adapt for adhesion of HEK293FT cells used for expression. HEK293FT cells were cultured in DMEM GlutaMAX supplemented with 10% (v/v) foetal calf serum with penicillin/streptomycin (10000 U/ml) at 37°C with 5% CO<sub>2</sub>, and passaged when confluent by trypsinisation. For the transient transfection, the cells were cultured in 150mm cell culture dishes until approximately 50-60% confluency. The cells were washed once with PBS, before introducing 15ml of transfection medium containing 3ug/ml of plasmid DNA, 9ug/ml of 25 kDa polyethylenimine (Polysciences)

## Chapter 3

in Freestyle293 medium, pre-incubated at 37 °C for 15mins. After 4 hours of transfection, the culture medium was added to 25ml of Freestyle293 medium containing valproic acid (final concentration of 2.2mM), and left to incubate for 6 days before harvesting. When harvesting, culture medium was collected, clarified of any cells by centrifugation, and stored at -30°C until used.

### *B4GALT1 purification by immobilised metal affinity chromatography*

Culture media (harvested from 1 culture dish) containing the secreted B4GALT1 was spiked with imidazole (final concentration 10mM) to reduce non-specific binding to the Ni-NTA resin before adding to 400µl Ni-NTA resin (90µm 6% agarose beads, Sepharose 6 fast flow, unpacked from a His-Gravitrapp column and loaded into a disposable 6ml capacity polypropylene column, Qiagen). The resin was pre-equilibrated with 10mM HEPES buffer, pH 7.5, containing 0.5M potassium chloride, 20mM imidazole and 5% glycerol (His-tag wash buffer). After sample loading, the resin was washed with 10 column volumes of His-tag wash buffer, and stored in 50mM HEPES buffer, pH 7, containing 0.25M potassium chloride, 20mM manganese chloride, 2.5% (v/v) glycerol (reaction buffer) at 4°C until use. An aliquot of resin (20ul) was mixed with 50µl of His-tag wash buffer containing 300mM imidazole and 20µl of the supernatant run on an SDS-PAGE gel to verify B4GALT1 expression and purity.

### *B4GALT1 activity assay after immobilisation on resin in Eppendorf tube*

To test B4GALT1 activity, de-sialylated-de-galactosylated bovine fetuin substrate was produced by incubation of bovine fetuin (100µg) with 5mU of sialidase A (Prozyme) and 5mU of β(1-3,4) galactosidase (Prozyme) in 50 mM sodium phosphate, pH 6.0, incubated at 37°C overnight. De-sialylation-de-galactosylation of the bovine fetuin was verified by SDS-PAGE and mass spectrometric analysis of the released glycans, and is referred to as Fet-SG (Figure 2, B and C). To immobilise the recombinantly expressed enzyme, 30µl of Ni-NTA resin bound

B4GALT1 was equilibrated with 50 $\mu$ l of reaction buffer in a 500 $\mu$ l Eppendorf tube, and the resin was collected by centrifugation using a benchtop centrifuge at 10000 rpm for 30 seconds. 20 $\mu$ l of reaction buffer was removed and replaced with 20 $\mu$ l of reaction buffer containing 5 $\mu$ g of Fet-SG and UDP-galactose (final concentration of 1mM, reaction volume of 50 $\mu$ l). The reaction tubes were placed on a weighing boat and left on an orbital shaker at 200rpm in a 37 $^{\circ}$ C room for 2 hours. At various collection time points, the supernatant was collected by centrifugation, and pooled with 3 washes (250 $\mu$ l each) of His-tag wash buffer. To test for resin reusability, the resin after the 30min reaction time point was re-used another 3 times (total of 4 rounds) by re-equilibrating with the two substrates by the same procedure. To test the enzyme reaction by analysis of the released product glycans, the pooled supernatant was slot blotted onto a PVDF membrane using a 96-well vacuum manifold, and the glycans released and analysed using the method described in Paper III.

*B4GALT1 activity assay after immobilisation on resin in column format*

For the “on-column” galactosylation, 100 $\mu$ l of Ni-NTA bound B4GALT1 resin bed was loaded into a 6ml polypropylene column (bed dimension ~ 7mm diameter X 2mm height) with volumes estimated by markings made with known amounts of liquid, or onto a 1ml syringe (bed dimension ~ 4mm diameter X 5mm height), blocked with a 4mm diameter filter disc (Figure 1). The measured flow rate (time taken to flow through 1ml of buffer), assisted by capillary action (i.e. tip of column and syringe was in contact with the walls of the collection tube), was approximately 40s/ml for the column and 100s/ml for the syringe. As flow is directed by gravity,



Figure 1. Photo of resin bound B4GALT1 packed onto a column (left) and a syringe (right). 100 $\mu$ l of bed volume was estimated by markings made using known volumes. The inner diameter of the syringe (bed dimension ~ 4mm diameter X 5mm height) is approximately half that of the column (bed dimension ~ 7mm diameter X 2mm height), resulting in a longer resin bed. The measured flow rate, assisted by capillary action (i.e. tip of column and syringe was in contact with the walls of the collection tube) was approximately 40s/ml for the column and 100s/ml for the syringe.

the flowrate increases with the amount of liquid in the column due to increased pressure and slows down as liquid volume decreases. All materials, including columns and pipette tips, were pre-warmed in the 37°C room for 30mins before the assay. The stationary phases were equilibrated with 4 column volumes of reaction buffer, followed by 1 column volume of reaction buffer containing 0.1mM of UDP-galactose. One column volume of reaction buffer containing 10µg of Fet-SG or IgG and 1mM UDP-galactose was added to the stationary phases and the flow through collected and pooled with a subsequent wash of 4 column volumes of His-tag wash buffer. As before, the pooled flowthrough samples were blotted onto PVDF for glycan release with PNGaseF as described in Paper III.

## Results

### SDS-PAGE of enzyme B4GALT1 and substrate Fet-SG

Human B4GALT1 is a type II transmembrane protein that is anchored in the Golgi apparatus of the cell. The B4GALT1 construct used in this work was designed such that the synthesized

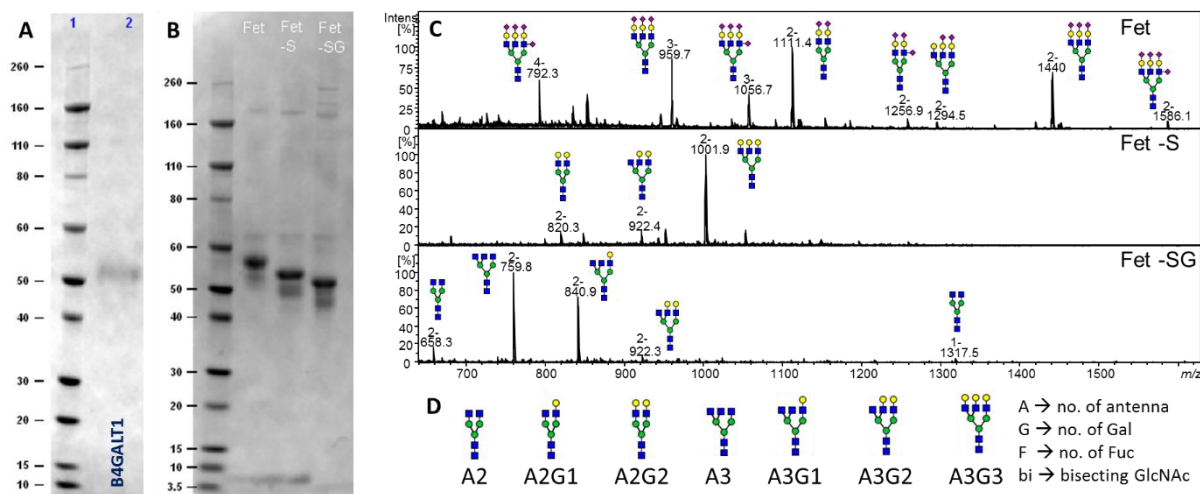


Figure 2. SDS-PAGE gels of (A) purified His-tagged B4GALT1 and (B) de-sialylated-de-galactosylated bovine fetuin. (A) The amount of recombinantly expressed B4GALT1 loaded into the gel was equivalent to approximately 2% of the total amount of resin, and is seen to be of high purity. The theoretical molecular weight of the His-tag B4GALT1 construct is approximately 44kDa, and contains 1 *N*-glycosylation site, resulting in the slower migration. (B) Migration differences between native fetuin, de-sialylated fetuin (Fet-S) and de-sialylated-de-galactosylated fetuin (Fet-SG) (5ug each) was visible by SDS-PAGE separation. The extent of glycan trimming was also confirmed by mass spectrometric analysis of released glycans (C). Averaged mass spectrum of the released glycans clearly showing the loss of sialic acid after sialidase treatment, and the majority of the galactose after galactosidase treatment. (D) Nomenclature of the glycans based on SNFG and Oxford notation for *N*-glycans, the latter of which describes only the features apart from the core structure, that is used in the text for subsequent glycan structure descriptions.

enzyme is secreted into the culture medium. This was done by replacing the protein *N*-terminal transmembrane domain with a secretion signal peptide, along with purification tags (8x His-tag and Strep-tag), and a TEV protease cut site, should the purification tags be undesired downstream. Secreted B4GALT1 was enriched by IMAC affinity chromatography and the enzyme was isolated at high purity as estimated by band intensity as shown in the SDS-PAGE gel (Figure 2, A). It was also observed that His-tag wash buffer supplemented with 0.5M potassium chloride resulted in a better purity compared to the traditional 0.5M sodium chloride used in His-tag purifications (data not shown). Resin bound B4GALT1 was then used for all subsequent experiments.

Concurrent de-glycosylation of bovine fetuin by sialidase and galactosidase was evident, shown by clear migration differences between untreated fetuin, de-sialylated fetuin and de-sialylated-de-galactosylated fetuin (Figure 2, B). Released glycan analysis showed galactosidase activity was incomplete, with approximately 40% of the fetuin glycans still containing 1 galactose (A3G1) (Figure 2, C and D). As A3G1 could still be galactosylated with 2 more galactoses and remains as a possible substrate for B4GALT1, complete degalactosylation was not pursued and this batch of Fet-SG was used for all subsequent experiments.

Due to the limited amount of resin bound B4GALT1 per purification and de-sialylated-de-galactosylated fetuin substrate, time point measurements and experimental format differences was pursued using an experimental setup of  $n=1$  as a proof-of-concept experiment.

#### *Resin bound B4GALT1 activity on Fet-SG*

His-tagged B4GALT1 bound to Ni-NTA was tested for activity by incubating the resin with Fet-SG and UDP-galactose in an Eppendorf tube, incubated in a 37°C warm room, lying flat in a weighing boat, on an orbital shaker. From experimental observations, the 90µm resin particles were prone to “sticking” on the walls of the tubes, resulting in temporary drying of the resin if

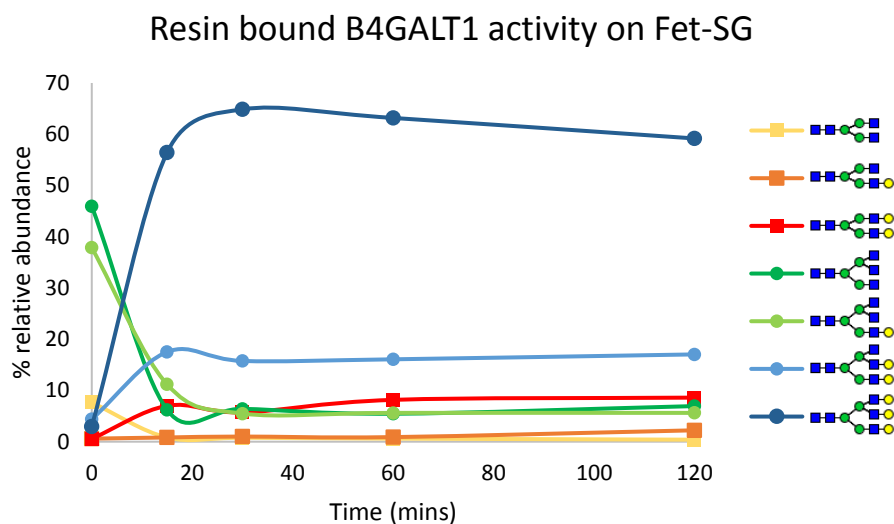


Figure 3. Time course measurement of resin bound B4GALT1 activity on Fet-SG. Each time point measurement, 15min, 30min, 1h, 2h, was an individual reaction tube. The abundance of the 6 major glycoforms were calculated as a relative abundance of the total glycan structure intensities.

the tubes were placed on a rotary or inverting shaker. The gentle rocking of the tubes in the weighing boat provided gentle stirring (the small resin particles were very buoyant) and the surface tension of the liquid kept the body of liquid intact. A time course measurement was set up for four time points, 15min, 30min, 1h and 2h, in individual reaction tubes, and the resulting relative abundance of the precursor and product glycans compared (Figure 3). The appearance of the fully galactosylated tri-antennary glycan (A3G3) from the de-sialylated and de-galactosylated precursor Fet-SG was observed in high abundance after 15min, with a corresponding decrease in the precursor tri-antennary glycans (A3 and A3G1). An identical trend was observed for the bi-antennary glycans, where the precursor glycan A2 was quickly depleted and the final product A2G2 formed within 15 mins, and showed that the resin bound B4GALT1 enzyme was active and fast acting in this experimental setup. Approximately 68% of the glycans were fully galactosylated (A2G2 and A3G3), but approximately 20% of the glycans could have been further galactosylated with one more Gal (A2G1 and A3G2), 6% with 2 more Gal (A2 and A3G1) and 6% with 3 more Gal (A3). This reduced final galactosylation may possibly be caused by product inhibition as the UDP released from UDP-galactose is a known inhibitor of B4GALT1 activity.



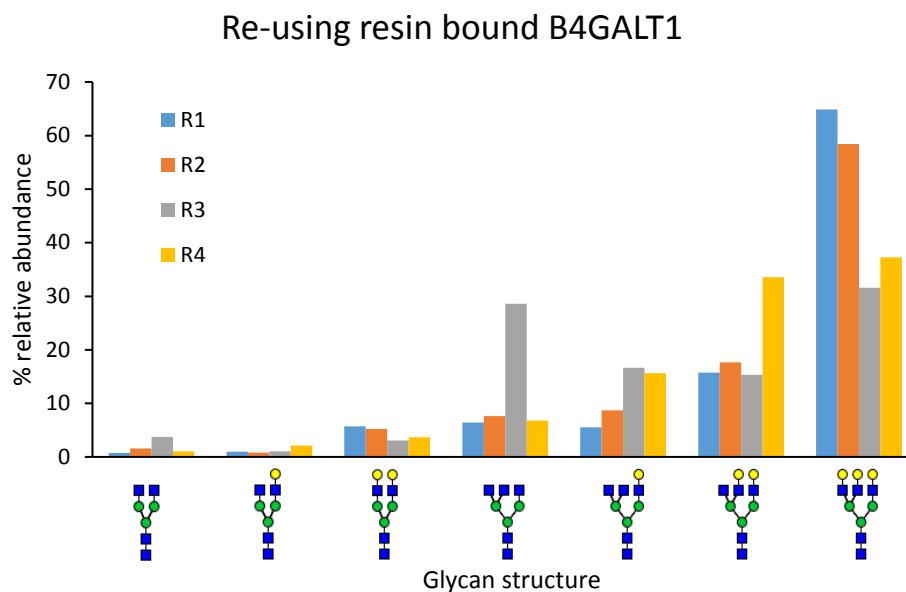


Figure 4. Relative abundance of released glycans, after repeated galactosylation reaction with the same resin bound B4GALT1. R1-R4 refer to the re-use number of reactions, and data point R1 is from the resin activity at the 30min time point in Figure 3. The subsequent reactions were also carried out for 30mins.

#### *Re-using resin bound B4GALT1*

One of the advantages of using immobilised enzymes is the ability to re-use the enzymes after the separation of the product. The resin was centrifuged after the 30min time point in Figure 3 and was tested for re-use activity for another 3 rounds of reaction after washing and re-equilibration as detailed in the Methods (Figure 4). B4GALT1 activity was overall decreased after round 2 with less fully galactosylated tri-antennary glycan A3G3 observed. It is possible that the loss of activity was not due to the loss of enzyme functionality, but rather due to the leeching and loss of the bound enzyme into the solution. Non-fetuin glycans, evident by the presence of core fucosylation, were detected in the second round of reaction, and is characteristic of glycoproteins produced in the HEK293 recombinant expression system (data not shown). This reinforces the suggestion that some bound B4GALT1 could have detached from the resin, resulting in less activity in reaction round 3 and 4. The loss of enzyme from the resin could have been caused by shear forces during the washes and centrifugation steps, or by

exposure to excess manganese during re-equilibration that may have interfered with the interaction between the immobilised nickel ion and the His-tag.

*B4GALT1 activity in a stationary phase column format*

Considering the fast reaction outcome of resin bound B4GALT1 in solution, the ability to “instantaneously” galactosylate proteins was investigated by packing the resin bound B4GALT1 as a stationary phase in either the purification column (Figure 5, Col flow) or in a 1

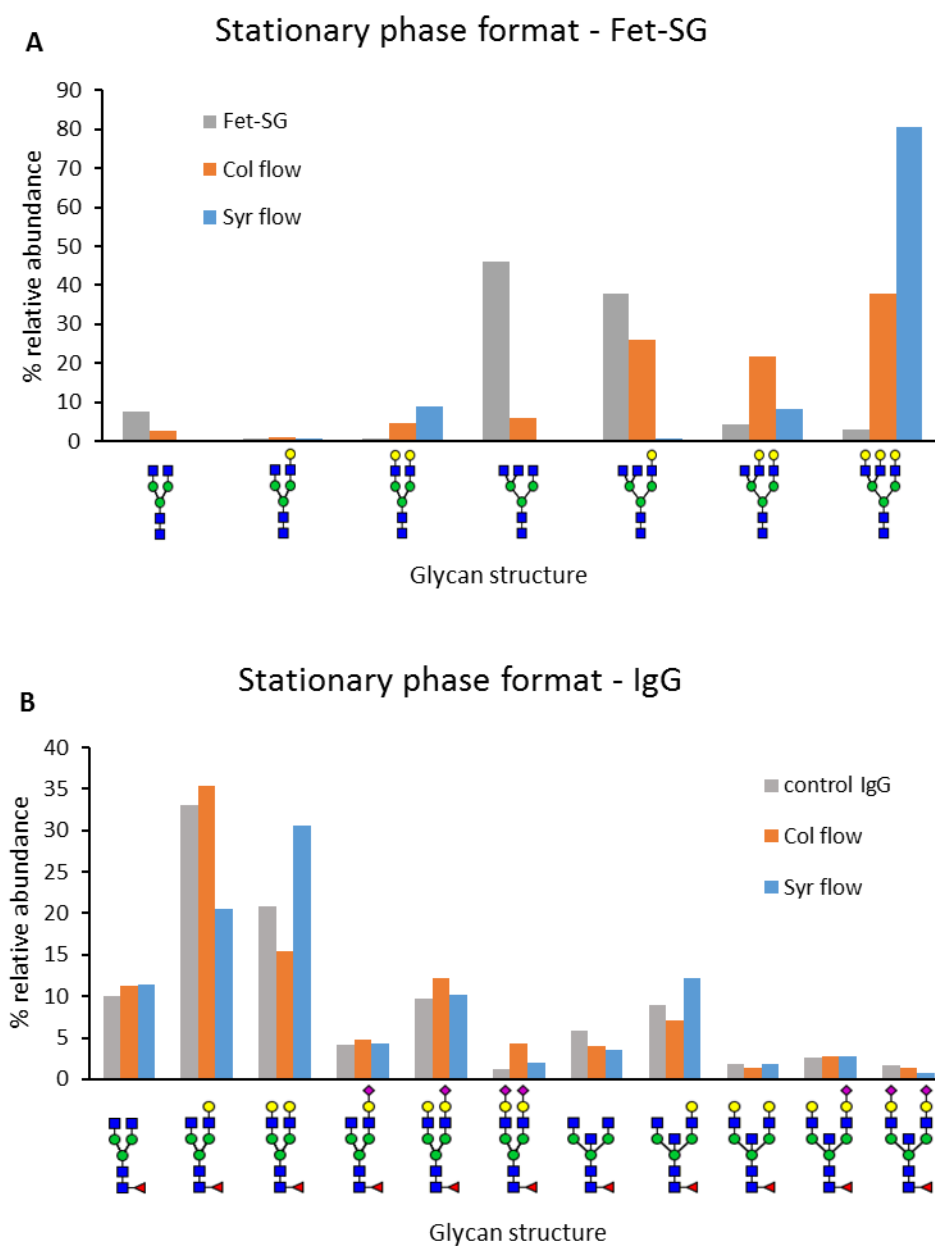


Figure 5. Relative abundance of released glycans from (A) Fet-SG and (B) IgG, after reacting with bound B4GALT1 in a packed resin format. “Col” denotes column, and “Syr” denotes syringe, as shown in the Figure 1 in the methods.

ml syringe (Figure 5, Syr flow) (Figure 1 in Methods). As the flow rates were not controlled via peristaltic pumps, but by gravity, the physical parameters of the tubes containing the enzyme conjugated resin directly affects the flow rate. The syringe has a smaller inner diameter (bed dimension ~ 4mm diameter X 5mm height) than the column (bed dimension ~ 7mm diameter X 2mm height), resulting in a longer resin bed (for approximately 100 $\mu$ l resin). The main forces affecting the flow rate by gravity are the weight (pressure) of the liquid, surface tension and friction between the liquid and the walls of the receptacle. The flowrate decreases in relation to the loading volume, thus the weight of the liquid decreases as the liquid flows through. With the pressure exerted by the liquid in the syringe higher than that of the column due to the difference in receptacle diameter, the interacting surface area between the liquid and the walls of the receptacle in the syringe is higher than the column, resulting in higher friction and surface tension. This resulted in a slower flow rate of liquids in the syringe (~100s/ml) as compared to the column (~40s/ml), assisted by capillary action (by placing the end of the stationary phase in contact with the walls of the collection tube). It was observed that without capillary action, flow rates at low volumes became inconsistent, as the droplet formation at the outlet lacked the weight to overcome surface tension and ambient air pressure. In testing the ability of the stationary phase of the column to galactosylate proteins “instantaneously”, Fet-SG and human serum IgG (10 $\mu$ g in 100 $\mu$ l volume) were flowed through the stationary phase sequentially, with a wash and re-equilibration step in-between. Human serum IgG is used here as an example of a glycoprotein with a less exposed glycosylation site than fetuin with approximately 60% of IgG glycans as possible substrates (FA2, FA2G1, FbiA2, and FbiA2G1) for galactosylation.

Based on released glycan analysis of the product glycoproteins, it was shown that the packed syringe was highly efficient in galactosylating fetuin, with the precursor glycans (A2 and A3) completely depleted (Figure 5, A), forming predominantly fully galactosylated glycans (A2G2 and A3G3). The packed column was less efficient, but galactosylation activity was also present,

## Chapter 3

with approximately 43% of the glycans fully galactosylated (A2G2 and A3G3). With human serum IgG, galactosylation was only observed using the packed syringe (FA2G1 to FA2G2, and FbiA2 to FbiA2G1). For both protein substrates, the packed syringe galactosylated the proteins better than the packed column. This could be attributed to a slower flow rate of liquids in the syringe, and a higher resin bed height resulting in a higher number of chromatographic theoretical plates. Leeching of immobilised B4GALT1 was not observed in these column configurations compared to the solution assay as no non-fetuin glycans were detected in the flowthrough of the Fet-SG samples, and the relative abundance of sialylated IgG glycans were consistent with the control IgG. Although this current data set was derived from an experimental acquisition of  $n=1$ , the observed trend of the higher galactosylating capacity of the immobilised B4GALT1 syringe configuration for galactosylating both Fet-SG and human serum IgG provided an interesting proof-of-concept that IMER in column format could be applied to future rapid *in vitro* glycoengineering applications.

### Discussion

Herein, the galactosylation of protein substrates by an immobilised glycosyltransferase demonstrates that this type of *in vitro* glycoengineering of proteins is controllable. Time and format have been tested showing that these parameters can modulate outcomes. Likely, other physical parameters such as enzyme-substrate ratio, temperature, incubation time, stirring speed, concentrations of co-factors and chemical additives can also be modulated to optimise reactions in this format. Use of non-endogenous enzymes is also readily optimised this type of *in vitro* glycoengineering, as there is little risk of problems that might occur from mis-folding, degradation or aggregation arising from expression of non-endogenous enzymes by metabolic engineering as described in Chapter 2. In addition, bacterial glycosyltransferases, that carry out the identical linkage transfer of nucleotide sugars onto glycan terminal monosaccharides, can

be used this type of *in vitro* glycoengineering, thus allowing an easier bacterial expression system to be used to express the transferase enzymes instead of a mammalian expression system.

Comparing the galactosylation outcome between the resin bound B4GALT1 in solution (Figure 2, T=15min) and the stationary phase packed B4GALT1 in the syringe (Figure 4, A), showed that the galactosylation reaction was more efficient in the packed resin syringe than using a solution based reaction in an Eppendorf tube. Despite a shorter contact time of 15 mins for the free resin bound enzyme compared to approximately 2 mins (including washing steps) for the syringe based reaction, approximately 20% more fully galactosylated glycans (A2G2 and A3G3) of Fet-SG was observed for the syringe format than the free resin. Although the resin:substrate was not identical (30 $\mu$ l resin:5 $\mu$ g Fet-SG for Eppendorf tube reaction, 100 $\mu$ l resin:10 $\mu$ g Fet-SG for syringe reaction), it is likely that product inhibition was less apparent in the stationary phase packed format. With the free resin, the buoyancy of the beads and shaking speed affects the mixing efficiency of the reaction that in turn may affect the enzyme exposure to the substrates. From experimental observations, in this small-scale Eppendorf tube reaction, the mixing efficiency was not optimal as much of the resin remained aggregated at the bottom of the Eppendorf tube or stuck to the wall of the tube which would reduce the exposure of the enzyme to the substrates. In a large-scale format, this might be optimised in reactors with proper stirring mechanisms; it was reported that 1kg of human IgG was galactosylated in a 40L bioreactor, achieving approximately >90% fully galactosylated IgG in 24h, using approximately 4g of purified B4GALT1 [13]. By packing the resin in a column, the active flow of the elution liquid brings the substrate to the enzyme, creating a local environment of high enzyme concentration [17]. The flow of the liquid also elutes the reacted products and released by-products from the enzyme reaction to reduce the interference of product inhibition. Regulation of the flow rate is an important factor for micro-fluidic based IMER as it directly impacts the contact time (reaction time) of the substrate with the enzyme [17, 18].

The lower efficiency of the galactosylation of human serum IgG as compared to Fet-SG was expected. Bovine fetuin is normally synthesised with highly sialylated glycans, with highly accessible glycosylation sites. Human serum IgG is commonly glycosylated with FA2 (also known as G0F) and FA2G1 (also known as G1F) glycans, and the glycosylation site is less exposed compared to the attachment sites of glycans on fetuin. Glycosylation site accessibility was not considered in the recently published theoretical model [18], but is a known factor affecting the degree of glycan processing of glycoproteins [21, 22].

Re-usability of the enzymes is one of the key advantages of enzyme immobilisation [14, 15, 17, 23], but is affected by enzyme stability on the immobilisation platform and the enzyme half-life. Leeching of bound enzymes was observed when the Eppendorf tube resin was re-used (Figure 4). Immobilisation by IMAC of the His-tagged recombinant enzyme might thus not be the best choice for glycosyltransferases as metal ions such as manganese are often required as a co-factor for glycosyltransferase activity and might interfere with the nickel binding of the His-tag. The B4GALT1 construct designed by the Repository of Glyco-enzyme constructs (<http://glycoenzymes.ccruc.uga.edu/>) also contained a Strep-tag, enabling immobilisation of the transferase on streptavidin conjugated resins. However, streptavidin resins are protein based which will degrade over time and are costlier than Ni-NTA resins. Hence, optimisation with regards to the choice of immobilisation methods and co-factor (manganese) concentration should be considered for subsequent trials of using B4GALT1 on IMER stationary phases. The half-life of B4GALT1 is not known, and the activity of bound B4GALT1 was only tested up to 3 days post purification. While free enzymes can be stored in freezing conditions to prolong storage, immobilised enzymes, especially agarose resin bound enzymes, cannot be frozen as freezing results in the breakage of the resin (manufacturer guidelines). Ni-NTA magnetic beads could be used as an alternative, but comes at a higher cost. Storage buffers are another parameter

required for optimisation to prolong the lifespan and re-usability of the enzyme immobilised on resins.

### **Conclusion**

In this work, the possibility of using an IMER stationary phase for rapid *in vitro* glycosylation of glycoproteins was shown using galactosylation as an example. Each step in the process required optimisation to promote higher enzyme activity and cost-effectiveness but its success demonstrates the feasibility of using such an *in vitro* system for sequentially derived glycosylation of glycoproteins using a packed resin column enzymatic format. Larger scale tests, and subsequent configurations involving other transferases are required in the future to truly create an *in vitro* “Golgi column” where the glycosylation processes can be regulated to produce designed glycans on proteins.

## References

1. Geisler, C., H. Mabashi-Asazuma, C.W. Kuo, K.H. Khoo, et al., Engineering beta1,4-galactosyltransferase I to reduce secretion and enhance *N*-glycan elongation in insect cells. *J Biotechnol*, 2015, *193*, 52-65.
2. Varki, A. and R. Schauer, *Sialic Acids*, in *Essentials of Glycobiology*, A. Varki, et al. 2009, Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY).
3. Sola, R.J. and K. Griebenow, Effects of glycosylation on the stability of protein pharmaceuticals. *J Pharm Sci*, 2009, *98*, 1223-1245.
4. Bertozzi, C.R., H.H. Freeze, A. Varki, and J.D. Esko, *Glycans in Biotechnology and the Pharmaceutical Industry*, in *Essentials of Glycobiology*, A. Varki, et al. 2009, Cold Spring Harbor Laboratory Press. : Cold Spring Harbor (NY).
5. Shade, K.-T. and R. Anthony, Antibody Glycosylation and Inflammation. *Antibodies*, 2013, *2*, 392.
6. Liu, L., H. Li, S.R. Hamilton, S. Gomathinayagam, et al., The impact of sialic acids on the pharmacokinetics of a PEGylated erythropoietin. *J Pharm Sci*, 2012, *101*, 4414-4418.
7. Karsten, C.M., M.K. Pandey, J. Figge, R. Kilchenstein, et al., Anti-inflammatory activity of IgG1 mediated by Fc galactosylation and association of Fc[gamma]RIIB and dectin-1. *Nat Med*, 2012, *18*, 1401-1406.
8. Kurogochi, M., M. Mori, K. Osumi, M. Tojino, et al., Glycoengineered Monoclonal Antibodies with Homogeneous Glycan (M3, G0, G2, and A2) Using a Chemoenzymatic Approach Have Different Affinities for FcγRIIIa and Variable Antibody-Dependent Cellular Cytotoxicity Activities. *PLoS One*, 2015, *10*, e0132848.
9. Weikert, S., D. Papac, J. Briggs, D. Cowfer, et al., Engineering Chinese hamster ovary cells to maximize sialic acid content of recombinant glycoproteins. *Nat Biotech*, 1999, *17*, 1116-1121.
10. Hodoniczky, J., Y.Z. Zheng, and D.C. James, Control of recombinant monoclonal antibody effector functions by Fc *N*-glycan remodeling in vitro. *Biotechnol Prog*, 2005, *21*, 1644-1652.
11. Thomann, M., T. Schlothauer, T. Dashivets, S. Malik, et al., In Vitro Glycoengineering of IgG1 and Its Effect on Fc Receptor Binding and ADCC Activity. *PLoS One*, 2015, *10*, e0134949.
12. Raju, T.S., J.B. Briggs, S.M. Chamow, M.E. Winkler, et al., Glycoengineering of therapeutic glycoproteins: in vitro galactosylation and sialylation of glycoproteins with terminal *N*-acetylglucosamine and galactose residues. *Biochemistry*, 2001, *40*, 8868-8876.
13. Warnock, D., X. Bai, K. Autote, J. Gonzales, et al., In vitro galactosylation of human IgG at 1 kg scale using recombinant galactosyltransferase. *Biotechnol Bioeng*, 2005, *92*, 831-842.
14. Care, A., P.L. Bergquist, and A. Sunna, Solid-binding peptides: smart tools for nanobiotechnology. *Trends Biotechnol*, 2015, *33*, 259-268.
15. Puri, M., C.J. Barrow, and M.L. Verma, Enzyme immobilization on nanomaterials for biofuel production. *Trends in Biotechnology*, 2013, *31*, 215-216.
16. Jia, F., B. Narasimhan, and S. Mallapragada, Materials-based strategies for multi-enzyme immobilization and co-localization: A review. *Biotechnol Bioeng*, 2014, *111*, 209-222.



17. Meller, K., M. Szumski, and B. Buszewski, Microfluidic reactors with immobilized enzymes—Characterization, dividing, perspectives. *Sensors and Actuators B: Chemical*, 2017, *244*, 84-106.
18. Klymenko, O.V., N. Shah, C. Kontoravdi, K.E. Royle, et al., Designing an Artificial Golgi reactor to achieve targeted glycosylation of monoclonal antibodies. *AIChE Journal*, 2016, *62*, 2959-2973.
19. Martin, J.G., M. Gupta, Y. Xu, S. Akella, et al., Toward an artificial Golgi: redesigning the biological activities of heparan sulfate on a digital microfluidic chip. *J Am Chem Soc*, 2009, *131*, 11041-11048.
20. Subedi, G.P., R.W. Johnson, H.A. Moniz, K.W. Moremen, et al., High Yield Expression of Recombinant Human Proteins with the Transient Transfection of HEK293 Cells in Suspension. *J Vis Exp*, 2015, e53568.
21. Thaysen-Andersen, M. and N.H. Packer, Site-specific glycoproteomics confirms that protein structure dictates formation of *N*-glycan type, core fucosylation and branching. *Glycobiology*, 2012, *22*, 1440-1452.
22. Lee, L.Y., C.H. Lin, S. Fanayan, N.H. Packer, et al., Differential site accessibility mechanistically explains subcellular-specific *N*-glycosylation determinants. *Front Immunol*, 2014, *5*, 404.
23. Care, A., K. Petroll, E.S. Gibson, P.L. Bergquist, et al., Solid-binding peptides for immobilisation of thermostable enzymes to hydrolyse biomass polysaccharides. *Biotechnology for biofuels*, 2017, *10*, 29.

## **Chapter 4 – Characterisation of the glycosylation of Immunoglobulin M**

For the *in vitro* glycoengineering of the IgM antibody, PAT-SM6, site-specific glycosylation analysis of the immunoglobulin M needed to be performed for the first time in order to understand the underlying glycosylation profiles at each glycosylation site of the heavily glycosylated antibody. This provided the potential glycan targets that can be modified and functionalised for the development of PAT-SM6 as a diagnostic and possible therapeutic tool for the treatment of cancer.

## Introduction

Antibodies play an important role in immunity, as part of the adaptive immune response that clears foreign entities from the host. In humans, there are 5 major types of antibodies, IgG, IgA, IgM, IgE and IgD, in order of natural abundance. Typically, each antibody monomer comprises of two heavy and 2 light chains ( $\kappa$ -light chain or  $\lambda$ -light chain), where the 1<sup>st</sup> *N*-terminal domain makes up the F<sub>ab</sub> region of the antibody, binding to the antigen, and the remaining the F<sub>c</sub> region that binds the specific antibody receptor (Figure 1). For IgA and IgM, a J-chain joins the monomers (2 for IgA, and 5 for IgM) to form a polymeric antibody. All antibodies are glycosylated proteins, with different numbers of glycosylation sites on the heavy chain of each antibody type. Glycosylated  $\kappa$ -light chains have been reported in isolated cases of multiple myeloma [1] and amyloidosis [2]. With IgG, glycoengineering approaches have shown that core-fucosylation, bisecting GlcNAc, sialylation and galactosylation all affect IgG Fc $\gamma$ R binding to various degrees [4-10]. With IgM and IgE, it has been demonstrated that the naturally high mannose glycosylation site at Asn402 on IgM [11, 12] and Asn394 on IgE [13] is crucial for antibody Fc effector function. For this thesis, as part of a collaboration project with Patrys

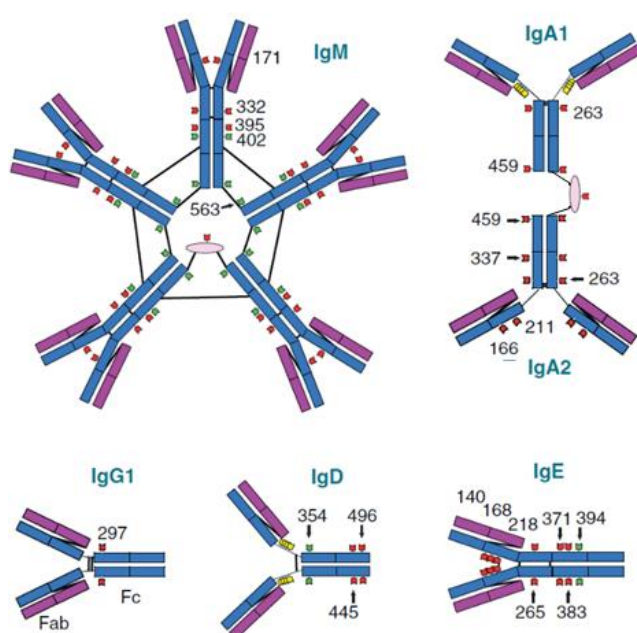


Figure 1. Schematic of the 5 types of human antibodies and their *N*-glycosylation sites adapted and modified from [3]. The light chain is coloured purple, and the heavy chain blue. *N*-glycosylation sites are marked red (predominant hybrid/complex glycans) and green (predominant high mannose glycans). *O*-glycans are present on the hinge region of IgA1 and IgD, and are marked yellow. Secretory IgA is naturally found as a dimer, joined by a J-chain, which is also seen on IgM (oval). For simplicity purposes, the two IgA types that are differentially glycosylated are shown together as a dimer; biologically, IgA dimers are homodimers of either IgA1 only or IgA2 only. While there are 4 subclasses of IgG, they are all glycosylated in the same way, differing only in the sequence of the Fc region [14].

Ltd in Melbourne, Australia, who have developed 3 different IgMs for anti-cancer diagnosis and therapeutic targeting, IgM will be the main antibody of interest.

### *Biological role of IgM*

In humans, IgM not only plays a role in adaptive immunity, but also in innate immunity [15]. The IgMs involved in innate immunity are termed natural IgM and are germline encoded antibodies that exist without any antigenic stimuli from a foreign biological substance. These natural IgMs have been reported to target self-antigens to facilitate clearance of apoptotic cells and cell debris, and pathogen-associated molecular pattern (PAMP) molecules such as bacterial lipopolysaccharides, fungal cell wall and viral DNA/RNA to strengthen the mucosal defence against pathogenic micro-organisms [16-18]. The counterpart, adaptive immune IgM, is the first antibody that B-cells produce against a foreign antigen. This leads into the subsequent isotype switching and affinity maturation of the corresponding IgG in memory B-cells.

Prior to the elucidation of a polymeric IgM specific Fc-receptor (Fc $\mu$ R) [19], C1q was thought to be the main interacting partner of IgM based on the fact that upon binding of the antigen, IgM can activate the complement pathway, either via the classical pathway by C1q binding [20] or via the lectin pathway by mannose binding lectin (MBL) binding [21, 22]. Due to its polymeric nature, IgM is more efficient at C1q activation than IgG; one molecule of IgM can interact with one molecule of C1q, whereas at least 2 molecules of IgG would be required for C1q binding. Polymeric IgM is most often found as pentameric forms (5 IgM monomer + 1 J-chain, but in certain situations such as lipopolysaccharide stimulation [23] or cold agglutinin disease [24], hexameric IgM (6 IgM monomer without J-chain) can also be found in increased abundance. Hexameric IgM is also a better binder of C1q than pentameric IgM presumably because C1q is also a polymeric protein with 6 arms, resulting in a better binding symmetry [25]. However, since the identification of Fc $\mu$ R as an IgM receptor, there has been growing

evidence to suggest IgM plays a role in B-cell development and immune homeostasis [26-29]. The binding and internalisation of IgM by Fc $\mu$ R have been exploited as a target for treatment of chronic lymphocytic leukaemia, which over expresses the Fc $\mu$ R [30]. A cytotoxic drug chemically conjugated to an IgM Fc scaffold reduced the tumour size by >60% over one week in a xenograft model.

Unlike IgG, which binds to its target antigen with high affinity, the affinity of IgM towards its antigen is low. This is compensated by a high avidity interaction, where polymeric IgM binds to multiple copies of its target antigen on a cell/pathogen, resulting in an overall high binding strength. While this means that IgM requires a high concentration of the target antigen within a locality, it also serves as a threshold for normal cells; based on tumour biomarker studies such as HER2 in breast cancer [31], where it is the over expression and not the unique presence of a certain biomarker that distinguishes between a cancerous and normal cell. This can be seen in examples of natural IgM being able to distinguish between normal and cancerous cells/tissues [32-34], as part of their immune surveillance and quality control capabilities. There have also been examples where an IgG has been engineered into an IgM isotype to result in high affinity and high avidity binding, thereby increasing the therapeutic potential of the antibody. This engineered hybrid has been demonstrated with the anti-HIV antibody, 2G12 to 2G12IgM [35, 36], and an anti-TRAIL receptor 1, TR-1 IgG to TR-1 IgM [37], where the resulting IgM variant was more efficient at neutralising the HIV virus (2G12IgM) or killing the cancer cells (TR-1 IgM). It is also of interest to note that majority of the known anti-carbohydrate (anti-glycan) antibodies are of IgM isotype [38-40]. For instance, monoclonal IgM L612, an IgM being used in Phase I clinical trials for treatment of metastatic melanoma, targets the ganglioside GM3 [41]. Another monoclonal IgM, HIgM12, was shown to have positive effects on treatment of multiple sclerosis and central nervous system injuries, and has been identified as binding to polysialic acid [42].

### *Structure-function relationship of IgM*

Secretory IgM predominantly exists as a pentamer with a single J-chain, without which it forms into a hexamer [43], making it the largest known antibody isotype so far. The structure of polymeric IgM is held together by 3 disulphide bridges at Cys337, Cys414 and Cys575 with the corresponding Cys in the neighbouring IgM heavy chain [44]. Cys337 forms the disulphide bridge within the IgM monomer, Cys414 and Cys575 forms disulphide bridges between the IgM monomers. Two of the Cys575 in the pentameric IgM forms the disulphide bridge with the J-chain; substitution of Cys575 results in the formation of predominantly hexameric IgM [43, 44]. From mutagenic studies, it was shown that substitution of Cys414 with Ser, dramatically decreases complement activation, suggesting that the disulphide bridge at Cys414 is critical to maintaining structural integrity of the polymeric IgM. IgM is also the most glycosylated antibody known, with five *N*-glycosylation sites per heavy chain, totally to 50 *N*-glycosylation sites (+1 on the J-chain) on a pentameric IgM and 60 *N*-glycosylation sites on the hexameric IgM. The effects of glycosylation on IgM structure and function has been discussed in my Paper II [45].

### *Anti-cancer IgM PAT-SAM6*

PAT-SAM6 (PAT-SM6) is a proprietary monoclonal natural IgM licensed by Patrys Ltd for the treatment of multiple myeloma [46, 47] (Patent no. – WO2005047332A1). PAT-SM6 was originally isolated from a patient suffering from stomach cancer using hybridoma technology, where the spleen cells from the patient are immortalised and screened for antibody production against a target cancer cell/tissue [34]. The reactivity of PAT-SM6 was found to recognise a wide range of cancer tissues, showing positive immunohistological staining of cancerous breast, liver, oesophageal, colon, stomach, bone marrow (multiple myeloma) tissue/biopsy sections [48-52], while having non-detectable staining against normal cells. A cancer variant of the

78kDa glucose regulator protein (GRP78) was identified to be an antigen of PAT-SM6, and the binding epitope was found to be *O*-glycosylated as treatment of cells with *O*-glycanase, an enzyme that removes Gal-GalNAc *O*-glycans, removed the binding affinity of PAT-SM6 to the pancreatic carcinoma cell BXPC-3 [49, 50]. Cell surface localisation of GRP78 is known to be a sign of cellular ER stress [53, 54], which is a characteristic of cancerous cells. Apart from activating complement dependent cytotoxicity, PAT-SM6 was reported to be able to simultaneously bind to the target GRP78 antigen and human LDL, resulting in the internalisation of LDL into the cell [52], which has been reported to cause the accumulation of lipids within the cell, resulting in apoptosis [48].

To produce PAT-SM6 at large scale for use as a potential therapeutic antibody, PAT-SM6 was recombinantly expressed in the FDA approved cell line, PER.C6, which was derived from human embryonic retinal cells [55]. Expression in PER.C6 cells enabled compatible human *N*-glycosylation, and produced yields of up to 2 g/L of functional polymeric IgM. PAT-SM6 has been subsequently tested in Phase I clinical trials and shown promising diagnostic results for multiple myeloma patients [46]. However, the Phase II clinical trials have now been halted due to production problems.

#### *Technical challenges of use of IgM for disease diagnostics and therapeutics*

One of the reasons IgG is so widely researched and utilised is largely due to the availability of well tested and streamlined protocols with regards to expression [9, 56-58], purification [59] and characterisation [60-62]. An intact monomeric IgG is approximately 150kDa, usually has two *N*-glycosylation sites, and is relatively stable in a range of buffer conditions. Characterisation of IgG and the resulting glycoforms using mass spectrometric methods can be performed on intact IgG [63], IgG glycopeptides, or released glycans [60-62], all of which correlate with the single glycosylation site at Asn297, with the exception of some IgGs with

Fab [64, 65] or  $\kappa$ -light [1, 2] chain glycosylation. An intact pentameric IgM has a molecular mass of approximately 950kDa, has 51 *N*-glycosylation sites, and is quite sensitive to the buffer environment; low salt concentrations (<50mM, experimental observation), low temperature (4°C, cryoglobulin IgM, [66]), low pH (< pH 5) [67] can result in IgM aggregation and precipitation. The hydrophilic glycans contribute to the IgM solubility; de-sialylation by sialidase or deglycosylation by EndoF3 results in precipitation of IgM (personal experimental observation). It was also reported that purification of IgM after desialylation of human serum IgM only recovered 27% of the original starting amount [68]. For small scale purifications, IgM can be purified by size exclusion chromatography following a prior IgM enrichment by ammonium sulphate precipitation [66] or by affinity pull downs using anti-IgM antibodies [68]. For large scale, IgM purification has been performed using ion exchange chromatography on hydroxyapatite monolithic columns [69]. In the case of PAT-SM6, clinical trials were halted due to a failure in the PAT-SM6 purification after the chromatographic column formulation was changed (personal communication from Patrys Ltd). New chromatographic methods have been developed in recent years, such as steric exclusion chromatography [70] and affinity purification using protein L, which targets the  $\lambda$ -light chain (but only works on IgMs with a  $\lambda$ -light chain) or using a llama expressed antibody which targets the IgM heavy chain [71, 72].

The characterisation of IgM is thus also more complex than of IgG. Even though IgM is most commonly found as a pentamer, monomeric and hexameric IgM also exist, and different levels of polymerisation have been found to affect IgM complement activation (hexamer > pentamer, monomer does not activate complement) [73, 74]. To characterise the different levels of polymerisation of IgM is also problematic. Visualisation by electron microscopy [43, 75] or atomic force microscopy [76] has been successful but is not readily available, and separation by non-reducing gel electrophoresis requires low salt concentrations that can cause IgM to precipitate (personal experimental observation). The overall glycosylation profile is contributed



to by 51 glycosylation sites in the pentamer plus J chain and the glycans have been shown to be unevenly distributed across the 5 glycosylation sites on each heavy chain [77, 78] (further discussed in Chapter 4 [45]). This means that released glycan profiles lose information related to the site-specific glycosylation which may have functional implications on IgM activity. Analysis of intact IgM by mass spectrometry is still an unavailable challenge. An unsuccessful attempt was made to achieve the mass spectrometric measure of intact PAT-SM6 when Professor Albert Heck kindly invited myself to visit his lab at Utrecht University, in the Netherlands, to use his custom Q-TOF [63]. With native MS, the importance of being able to resolve the charge state envelope of the protein is key to determining the molecular mass of the protein [79]. Heterogeneous glycosylation increases the heterogeneity of the protein population, thereby increasing the difficulty of resolving different charge states of different glycoforms, and the difficulty increases as the number of glycosylation sites increase. Unfortunately, we were unable to resolve the charge state envelope of intact PAT-SM6, although we could show that the aliquot of PAT-SM6 we analysed contained fragments corresponding to the Fab region of PAT-SM6, suggesting some level of degradation of the intact PAT-SM6 (data not shown).

## **Paper II – Site specific glycosylation characterisation of IgM**

**E.S.X. Moh**, C.H. Lin, M. Thaysen-Andersen, N.H. Packer. (2015). *Site-specific N-glycosylation of recombinant pentameric and hexameric human IgM*. Journal of American Society of Mass Spectrometry, 27(7), 1143-1155.

Paper II comprises the detailed characterisation of the IgM used in this work, PAT-SM6. Site-specific glycosylation was performed on intact glycopeptides of tryptic and GluC digested PAT-SM6, and the data analysis was performed manually with high stringency on glycopeptide identity. Glycosylation site occupancy and the site-specific glycosylation profile differences between pentameric and hexameric variants of PAT-SM6 was compared and this work had been published in Journal of American Society of Mass Spectrometry [45]. The glycopeptide data have also been submitted to a developing database project on immunoglobulin glycosylation, IgCarbKB [80]. The possible glycan isoforms of the identified glycopeptides for the pentameric PAT-SM6 were also characterised and presented in the additional Supplementary section.

Pages 99-111 of this thesis have been removed as they contain published material. Please refer to the following citation for details of the article contained in these pages.

Moh, E. S. X., Lin, C.-H., Thaysen-Andersen, M., & Packer, N. H. (2016). Site-specific N-glycosylation of recombinant pentameric and hexameric human IgM. *Journal of American Society of Mass Spectrometry*, 27, 1143-3155.

DOI: [10.1007/s13361-016-1378-0](https://doi.org/10.1007/s13361-016-1378-0)

## **Additional Supplementary data for Paper II**



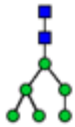
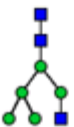
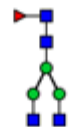

### *Determination of the glycan structural isoforms on the IgM glycopeptides*

The data shown in the above Paper II [45], because of current technology limitations, shows the monosaccharide compositions of the glycans attached to the glycopeptides of the pentamer and hexamer forms of IgM. This section describes the determination of the possible structural glycan isoforms that correspond to these compositions on the glycopeptides. The detailed linkages and sequences of the glycan structures can best be performed after the release of the glycans from the protein and subsequent analysis by PGC-LC-ESI-MS/MS by the method as described in my Paper III. The results are presented as the relative abundance of each glycan isoform, if any, for each glycan composition to directly complement the glycopeptide data in Paper II.

The assignments of the glycan structures are based on porous graphitised carbon chromatographic retention time rules and tandem MS diagnostic fragment ions [81-83]. Some of these retention time rules include  $\alpha$ -2,6 sialic acid modified glycans eluting earlier than those terminating in  $\alpha$ -2,3 sialic acid, core fucosylation eluting later than outer-arm fucosylation, and sialylated glycans eluting later than neutral glycans. Some of the tandem MS fragment ions used as diagnostic for certain structural glycan features include  $m/z$  350 and 368 for core fucose,  $m/z$  290 and 655 for sialylated glycans and  $m/z$  510 for outer arm fucose [82].

In an overview, the glycan structures identified were similar to the glycosylation profile of other proteins synthesized by the Per.C6 cell line [84, 85]. As a cell line derived from embryonic human retinal cells, the glycosylation profile was human-like and has abundant outer-arm fucosylation (Lewis epitopes), similar to proteins found in secretory fluids such as tears [86] and saliva [82]. Both  $\alpha$ -2,3 and  $\alpha$ -2,6 sialylation was present, similar to previous reports of other Per.C6 expressed proteins [84, 85].

Table. List of glycan structural isoforms of same mass characterised for each glycan composition and relative abundance corresponding to the glycopeptide analysis.

Glycan Composition	no. of isoforms	Structures verified (and abundance distribution)
(Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	nd	
(Hex)1 + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	nd	
(Hex)2 + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	2	 1.56 ± 0.37                      98.44 ± 0.37
(Hex)2(HexNAc)1(D eoxyhexose)1 + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	nd	
(Hex)1(HexNAc)1 + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	1	
(Hex)3 + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	1	
(Hex)2(HexNAc)1 + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	1	
(HexNAc)2(D eoxyhexose)1 + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	1	
(Hex)4 + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	2	 30.57 ± 2.22                      69.43 ± 2.22
(Hex)1(HexNAc)1(NeuAc)1 + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	nd	
(Hex)1(HexNAc)1(D eoxyhexose)2 + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	nd	

(Hex) <sub>2</sub> (HexNAc) <sub>1</sub> (Deoxyhexose) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	3				6.69 ± 1.15	46.83 ± 4.89	46.48 ± 4.52
(Hex) <sub>3</sub> (HexNAc) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	1						
(Hex) <sub>1</sub> (HexNAc) <sub>2</sub> (Deoxyhexose) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	2				34.35 ± 5.20	65.65 ± 5.20	
(Hex) <sub>2</sub> (HexNAc) <sub>2</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	1						
(HexNAc) <sub>3</sub> (Deoxyhexose) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	1						
(Hex) <sub>1</sub> (HexNAc) <sub>1</sub> (Deoxyhexose) <sub>1</sub> (NeuAc) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	1						
(Hex) <sub>5</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	2				74.70 ± 2.90	25.30 ± 2.90	
(Hex) <sub>2</sub> (HexNAc) <sub>1</sub> (NeuAc) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	1						
(Hex) <sub>3</sub> (HexNAc) <sub>1</sub> (Deoxyhexose) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	3				39.17 ± 5.87	52.79 ± 6.40	8.04 ± 1.34

(Hex) <sub>1</sub> (HexNAc) <sub>2</sub> (Deoxyhexose) <sub>2</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	1		
(Hex) <sub>2</sub> (HexNAc) <sub>2</sub> (Deoxyhexose) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	1		
(Hex) <sub>1</sub> (HexNAc) <sub>3</sub> (Deoxyhexose) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	2		41.04 ± 1.79
(Hex) <sub>2</sub> (HexNAc) <sub>1</sub> (Deoxyhexose) <sub>1</sub> (NeuAc) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	2		9.87 ± 2.91
(Hex) <sub>6</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	1		90.13 ± 2.91
(Hex) <sub>3</sub> (HexNAc) <sub>1</sub> (NeuAc) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	2		16.79 ± 4.56
(Hex) <sub>1</sub> (HexNAc) <sub>2</sub> (Deoxyhexose) <sub>1</sub> (NeuAc) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	2		83.21 ± 4.56
(Hex) <sub>2</sub> (HexNAc) <sub>2</sub> (NeuAc) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	3		18.94 ± 3.17
(Hex) <sub>2</sub> (HexNAc) <sub>2</sub> (NeuAc) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	2		81.06 ± 3.17
(Hex) <sub>2</sub> (HexNAc) <sub>2</sub> (NeuAc) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	3		7.83 ± 1.17
(Hex) <sub>2</sub> (HexNAc) <sub>2</sub> (Deoxyhexose) <sub>2</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	2		24.11 ± 1.68
(Hex) <sub>2</sub> (HexNAc) <sub>2</sub> (Deoxyhexose) <sub>2</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	2		59.72 ± 4.57
			68.06 ± 2.39

(Hex) <sub>2</sub> (HexNAc) <sub>3</sub> (Deoxyhexose) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	3				38.18 ± 2.64	21.72 ± 4.46	40.10 ± 3.17
(Hex) <sub>3</sub> (HexNAc) <sub>1</sub> (Deoxyhexose) <sub>1</sub> (NeuAc) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	1						
(Hex) <sub>2</sub> (HexNAc) <sub>2</sub> (Deoxyhexose) <sub>1</sub> (NeuAc) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	2				68.45 ± 5.68	31.55 ± 5.68	
(Hex) <sub>1</sub> (HexNAc) <sub>3</sub> (Deoxyhexose) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	nd						
(Hex) <sub>2</sub> (HexNAc) <sub>3</sub> (Deoxyhexose) <sub>2</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	nd						
(Hex) <sub>3</sub> (HexNAc) <sub>3</sub> (Deoxyhexose) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	1						
(Hex) <sub>2</sub> (HexNAc) <sub>2</sub> (NeuAc) <sub>2</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	1						
(Hex) <sub>2</sub> (HexNAc) <sub>2</sub> (Deoxyhexose) <sub>2</sub> (NeuAc) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	2				79.51 ± 3.03	20.49 ± 3.03	
(Hex) <sub>2</sub> (HexNAc) <sub>3</sub> (Deoxyhexose) <sub>1</sub> (NeuAc) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	2				44.85 ± 3.16	55.15 ± 3.16	
(Hex) <sub>3</sub> (HexNAc) <sub>3</sub> (Deoxyhexose) <sub>2</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	nd						
(Hex) <sub>2</sub> (HexNAc) <sub>2</sub> (Deoxyhexose) <sub>1</sub> (NeuAc) <sub>2</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	3				65.35 ± 4.84	29.88 ± 4.64	4.77 ± 2.46



(Hex)2(HexNAc)3(Deoxyhexose)2(NeuAc)1 + (Man) <sub>2</sub> (GlcNAc) <sub>2</sub>	nd	
(Hex)3(HexNAc)3(Deoxyhexose)1(NeuAc)1 + (Man) <sub>2</sub> (GlcNAc) <sub>2</sub>	2	 53.94 ± 6.83      46.06 ± 6.83
(Hex)3(HexNAc)3(NeuAc)2 + (Man) <sub>2</sub> (GlcNAc) <sub>2</sub>	4	 28.26 ± 6.18      15.43 ± 4.74      39.59 ± 6.51      16.72 ± 4.26
(Hex)3(HexNAc)3(Deoxyhexose)1(NeuAc)2 + (Man) <sub>2</sub> (GlcNAc) <sub>2</sub>	nd	

Glycans were only considered for quantitation if a corresponding tandem MS fragment profile was acquired, if MS2 was not possible the structures are labelled as isoforms “nd”. Relative abundance was only compared for glycan compositions with more than one structural isoform ( $n=4$ ). The glycan structures were solved based on retention time, diagnostic tandem MS fragment ions and *N*-glycan synthesis rules of mammalian glycosylation. For sialic acids linkages,  $\alpha$ -2,3 and  $\alpha$ -2,6 linkages that could be identified based on retention time comparison are shown with angled lines as per SNFG/Oxford combined annotation [87].

## References

1. Hashimoto, R., T. Toda, H. Tsutsumi, M. Ohta, et al., Abnormal N-glycosylation of the immunoglobulin G kappa chain in a multiple myeloma patient with crystalglobulinemia: case report. *International journal of hematology*, 2007, 85, 203-206.
2. Omtvedt, L.A., D. Bailey, D.V. Renouf, M.J. Davies, et al., Glycosylation of immunoglobulin light chains associated with amyloidosis. *Amyloid : the international journal of experimental and clinical investigation : the official journal of the International Society of Amyloidosis*, 2000, 7, 227-244.
3. Arnold, J.N., M.R. Wormald, R.B. Sim, P.M. Rudd, et al., The Impact of Glycosylation on the Biological Function and Structure of Human Immunoglobulins. *Annual Review of Immunology*, 2007, 25, 21-50.
4. Lin, C.-W., M.-H. Tsai, S.-T. Li, T.-I. Tsai, et al., A common glycan structure on immunoglobulin G for enhancement of effector functions. *Proceedings of the National Academy of Sciences*, 2015.
5. Mizushima, T., H. Yagi, E. Takemoto, M. Shibata-Koyama, et al., Structural basis for improved efficacy of therapeutic antibodies on defucosylation of their Fc glycans. *Genes Cells*, 2011, 16, 1071-1080.
6. Raju, T.S., Terminal sugars of Fc glycans influence antibody effector functions of IgGs. *Curr Opin Immunol*, 2008, 20, 471-478.
7. Bowden, T.A., K. Baruah, C.H. Coles, D.J. Harvey, et al., Chemical and structural analysis of an antibody folding intermediate trapped during glycan biosynthesis. *J Am Chem Soc*, 2012, 134, 17554-17563.
8. Huang, W., J. Giddens, S.-Q. Fan, C. Toonstra, et al., Chemoenzymatic Glycoengineering of Intact IgG Antibodies for Gain of Functions. *J Am Chem Soc*, 2012, 134, 12308-12318.
9. Loos, A. and H. Steinkellner, IgG-Fc glycoengineering in non-mammalian expression hosts. *Arch Biochem Biophys*, 2012, 526, 167-173.
10. Zou, G., H. Ochiai, W. Huang, Q. Yang, et al., Chemoenzymatic synthesis and Fc gamma receptor binding of homogeneous glycoforms of antibody Fc domain. Presence of a bisecting sugar moiety enhances the affinity of Fc to Fc gamma IIIa receptor. *J Am Chem Soc*, 2011, 133, 18975-18991.
11. Muraoka, S. and M.J. Shulman, Structural requirements for IgM assembly and cytolytic activity. Effects of mutations in the oligosaccharide acceptor site at Asn402. *J Immunol*, 1989, 142, 695-701.
12. Wright, J.F., M.J. Shulman, D.E. Isenman, and R.H. Painter, C1 binding by mouse IgM. The effect of abnormal glycosylation at position 402 resulting from a serine to asparagine exchange at residue 406 of the mu-chain. *J Biol Chem*, 1990, 265, 10506-10513.
13. Shade, K.T., B. Platzer, N. Washburn, V. Mani, et al., A single glycan on IgE is indispensable for initiation of anaphylaxis. *J Exp Med*, 2015, 212, 457-467.
14. Wuhler, M., J.C. Stam, F.E. van de Geijn, C.A. Koeleman, et al., Glycosylation profiling of immunoglobulin G (IgG) subclasses from human serum. *Proteomics*, 2007, 7, 4070-4081.
15. Gronwall, C., J. Vas, and G.J. Silverman, Protective Roles of Natural IgM Antibodies. *Front Immunol*, 2012, 3, 66.
16. Gronwall, C. and G.J. Silverman, Natural IgM: beneficial autoantibodies for the control of inflammatory and autoimmune disease. *J Clin Immunol*, 2014, 34 Suppl 1, S12-21.

17. Kaveri, S.V., G.J. Silverman, and J. Bayry, Natural IgM in immune equilibrium and harnessing their therapeutic potential. *J Immunol*, 2012, 188, 939-945.
18. Mahla, R.S., M.C. Reddy, D.V.R. Prasad, and H. Kumar, Sweeten PAMPs: Role of Sugar Complexed PAMPs in Innate Immunity and Vaccine Biology. *Frontiers in Immunology*, 2013, 4, 248.
19. Kubagawa, H., S. Oka, Y. Kubagawa, I. Torii, et al., Identity of the elusive IgM Fc receptor (FcmuR) in humans. *J Exp Med*, 2009, 206, 2779-2793.
20. Shulman, M.J., C. Collins, N. Pennell, and N. Hozumi, Complement activation by IgM: evidence for the importance of the third constant domain of the mu heavy chain. *Eur J Immunol*, 1987, 17, 549-554.
21. McMullen, M.E., M.L. Hart, M.C. Walsh, J. Buras, et al., Mannose-binding lectin binds IgM to activate the lectin complement pathway in vitro and in vivo. *Immunobiology*, 2006, 211, 759-766.
22. Arnold, J.N., M.R. Wormald, D.M. Suter, C.M. Radcliffe, et al., Human serum IgM glycosylation: identification of glycoforms that can bind to mannan-binding lectin. *J Biol Chem*, 2005, 280, 29080-29087.
23. Randall, T.D., R.M.E. Parkhouse, and R.B. Corley, J-Chain Synthesis and Secretion of Hexameric Igm Is Differentially Regulated by Lipopolysaccharide and Interleukin-5. *Proc Natl Acad Sci U S A*, 1992, 89, 962-966.
24. Hughey, C.T., J.W. Brewer, A.D. Colosia, W.F. Rosse, et al., Production of IgM hexamers by normal and autoimmune B cells: implications for the physiologic role of hexameric IgM. *J Immunol*, 1998, 161, 4091-4097.
25. Randall, T.D., L.B. King, and R.B. Corley, The biological effects of IgM hexamer formation. *European Journal of Immunology*, 1990, 20, 1971-1979.
26. Nguyen, T.T.T., K. Klasener, C. Zurn, P.A. Castillo, et al., The IgM receptor Fc[ $\mu$ ]R limits tonic BCR signaling by regulating expression of the IgM BCR. *Nat Immunol*, 2017, *advance online publication*.
27. Wang, H., J.E. Coligan, and H.C. Morse, Emerging Functions of Natural IgM and Its Fc Receptor FCMR in Immune Homeostasis. *Frontiers in Immunology*, 2016, 7.
28. Ouchida, R., Q. Lu, J. Liu, Y. Li, et al., FcmuR interacts and cooperates with the B cell receptor To promote B cell survival. *J Immunol*, 2015, 194, 3096-3101.
29. Kubagawa, H., S. Oka, Y. Kubagawa, I. Torii, et al., The Long Elusive IgM Fc Receptor, Fc $\mu$ R. *Journal of Clinical Immunology*, 2014, 1-11.
30. Vire, B., M. Skarzynski, J.D. Thomas, C.G. Nelson, et al., Harnessing the Fc $\mu$  Receptor for Potent and Selective Cytotoxic Therapy of Chronic Lymphocytic Leukemia. *Cancer Res*, 2014.
31. Yardley, D.A., P.A. Kaufman, W. Huang, L. Krekow, et al., Quantitative measurement of HER2 expression in breast cancers: comparison with 'real-world' routine HER2 testing in a multicenter Collaborative Biomarker Study and correlation with overall survival. *Breast Cancer Research : BCR*, 2015, 17, 41.
32. Azuma, Y., Y. Ishikawa, S. Kawai, T. Tsunenari, et al., Recombinant human hexamer-dominant IgM monoclonal antibody to ganglioside GM3 for treatment of melanoma. *Clin Cancer Res*, 2007, 13, 2745-2750.
33. Brandlein, S., N. Rauschert, L. Rasche, A. Dreykluft, et al., The human IgM antibody SAM-6 induces tumor-specific apoptosis with oxidized low-density lipoprotein. *Mol Cancer Ther*, 2007, 6, 326-333.
34. Peter Vollmers, H., R. O'Connor, J. Müller, T. Kirchner, et al., SC-1, a Functional Human Monoclonal Antibody against Autologous Stomach Carcinoma Cells. *Cancer Res*, 1989, 49, 2471-2476.

35. Wolbank, S., R. Kunert, G. Stiegler, and H. Katinger, Characterization of Human Class-Switched Polymeric (Immunoglobulin M [IgM] and IgA) Anti-Human Immunodeficiency Virus Type 1 Antibodies 2F5 and 2G12. *J Virol*, 2003, 77, 4095-4103.
36. Chromikova, V., A. Mader, S. Hofbauer, C. Gobl, et al., Introduction of germline residues improves the stability of anti-HIV mAb 2G12-IgM. *Biochim Biophys Acta*, 2015, 1854, 1536-1544.
37. Piao, X., T. Ozawa, H. Hamana, K. Shitaoka, et al., TRAIL-receptor 1 IgM antibodies strongly induce apoptosis in human cancer cells in vitro and in vivo. *OncoImmunology*, 2016, e1131380.
38. Muthana, S.M. and J.C. Gildersleeve, Factors Affecting Anti-Glycan IgG and IgM Repertoires in Human Serum. *Scientific Reports*, 2016, 6, 19509.
39. Muthana, S.M., L. Xia, C.T. Campbell, Y. Zhang, et al., Competition between serum IgG, IgM, and IgA anti-glycan antibodies. *PLoS One*, 2015, 10, e0119298.
40. Sterner, E., N. Flanagan, and J.C. Gildersleeve, Perspectives on Anti-Glycan Antibodies Gleaned from Development of a Community Resource Database. *ACS Chem Biol*, 2016, 11, 1773-1783.
41. Irie, R.F., D.W. Ollila, S. O'Day, and D.L. Morton, Phase I pilot clinical trial of human IgM monoclonal antibody to ganglioside GM3 in patients with metastatic melanoma. *Cancer Immunology, Immunotherapy*, 2004, 53, 110-117.
42. Watzlawik, J.O., R.J. Kahoud, S. Ng, M.M. Painter, et al., Polysialic acid as an antigen for monoclonal antibody HIgM12 to treat multiple sclerosis and other neurodegenerative disorders. *Journal of neurochemistry*, 2015, 134, 865-878.
43. Davis, A.C., K.H. Roux, and M.J. Shulman, On the structure of polymeric IgM. *Eur J Immunol*, 1988, 18, 1001-1008.
44. Davis, A.C., K.H. Roux, J. Pursey, and M.J. Shulman, Intermolecular disulfide bonding in IgM: effects of replacing cysteine residues in the mu heavy chain. *Embo J*, 1989, 8, 2519-2526.
45. Moh, E.S., C.H. Lin, M. Thaysen-Andersen, and N.H. Packer, Site-Specific N-Glycosylation of Recombinant Pentameric and Hexameric Human IgM. *J Am Soc Mass Spectrom*, 2016, 27, 1143-1155.
46. Rasche, L., J. Duell, I.C. Castro, V. Dubljevic, et al., GRP78-directed immunotherapy in relapsed or refractory multiple myeloma - results from a phase 1 trial with the monoclonal immunoglobulin M antibody PAT-PAT-SM6. *Haematologica*, 2015, 100, 377-384.
47. Hensel, F., M. Eckstein, A. Rosenwald, and S. Brandlein, Early development of PAT-PAT-SM6 for the treatment of melanoma. *Melanoma Res*, 2013, 23, 264-275.
48. Pohle, T., S. Brandlein, N. Ruoff, H.K. Muller-Hermelink, et al., Lipoptosis: tumor-specific cell death by antibody-induced intracellular lipid accumulation. *Cancer Res*, 2004, 64, 3900-3906.
49. Rauschert, N., S. Brandlein, E. Holzinger, F. Hensel, et al., A new tumor-specific variant of GRP78 as target for antibody-based therapy. *Lab Invest*, 2008, 88, 375-386.
50. Rosenes, Z., T.D. Mulhern, D.M. Hatters, L.L. Ilag, et al., The anti-cancer IgM monoclonal antibody PAT-PAT-SM6 binds with high avidity to the unfolded protein response regulator GRP78. *PLoS One*, 2012, 7, e44927.
51. Rasche, L., J. Duell, C. Morgner, M. Chatterjee, et al., The natural human IgM antibody PAT-PAT-SM6 induces apoptosis in primary human multiple myeloma cells by targeting heat shock protein GRP78. *PLoS One*, 2013, 8, e63414.

52. Rosenes, Z., Y.F. Mok, S. Yang, M.D. Griffin, et al., Simultaneous binding of the anti-cancer IgM monoclonal antibody PAT-PAT-SM6 to low density lipoproteins and GRP78. *PLoS One*, 2013, 8, e61239.
53. Gonzalez-Gronow, M., M.A. Selim, J. Papalas, and S.V. Pizzo, GRP78: a multifunctional receptor on the cell surface. *Antioxid Redox Signal*, 2009, 11, 2299-2306.
54. Zhang, Y., R. Liu, M. Ni, P. Gill, et al., Cell Surface Relocalization of the Endoplasmic Reticulum Chaperone and Unfolded Protein Response Regulator GRP78/BiP. *Journal of Biological Chemistry*, 2010, 285, 15065-15075.
55. Tchoudakova, A., F. Hensel, A. Murillo, B. Eng, et al., High level expression of functional human IgMs in human PER.C6 cells. *MAbs*, 2009, 1, 163-171.
56. Dodev, T.S., P. Karagiannis, A.E. Gilbert, D.H. Josephs, et al., A tool kit for rapid cloning and expression of recombinant antibodies. *Scientific Reports*, 2014, 4, 5885.
57. Zha, D., Glycoengineered Pichia-based expression of monoclonal antibodies. *Methods Mol Biol*, 2013, 988, 31-43.
58. Frenzel, A., M. Hust, and T. Schirrmann, Expression of Recombinant Antibodies. *Frontiers in Immunology*, 2013, 4, 217.
59. Girard, V., N.-J. Hilbold, C.K.S. Ng, L. Pegon, et al., Large-scale monoclonal antibody purification by continuous chromatography, from process design to scale-up. *J Biotechnol*, 2015, 213, 65-73.
60. Wooding, K.M., W. Peng, and Y. Mechref, Characterization of Pharmaceutical IgG and Biosimilars Using Miniaturized Platforms and LC-MS/MS. *Curr Pharm Biotechnol*, 2016, 17, 788-801.
61. He, X., N. Washburn, E. Arevalo, and J.H. Robblee, Analytical characterization of IgG Fc subclass variants through high-resolution separation combined with multiple LC-MS identification. *Analytical and Bioanalytical Chemistry*, 2015, 407, 7055-7066.
62. Zhang, H., W. Cui, and M.L. Gross, Mass spectrometry for the biophysical characterization of therapeutic monoclonal antibodies. *FEBS Lett*, 2014, 588, 308-317.
63. Rosati, S., E.T.J. van den Bremer, J. Schuurman, P.W.H.I. Parren, et al., In-depth qualitative and quantitative analysis of composite glycosylation profiles and other micro-heterogeneity on intact monoclonal antibodies by high-resolution native mass spectrometry using a modified Orbitrap. *mAbs*, 2013, 5, 917-924.
64. Bondt, A., Y. Rombouts, M.H. Selman, P.J. Hensbergen, et al., Immunoglobulin G (IgG) Fab glycosylation analysis using a new mass spectrometric high-throughput profiling method reveals pregnancy-associated changes. *Mol Cell Proteomics*, 2014, 13, 3029-3039.
65. van de Bovenkamp, F.S., L. Hafkenscheid, T. Rispens, and Y. Rombouts, The Emerging Importance of IgG Fab Glycosylation in Immunity. *The Journal of Immunology*, 2016, 196, 1435-1441.
66. Ramsland, P.A., S.S. Terzyan, G. Cloud, C.R. Bourne, et al., Crystal structure of a glycosylated Fab from an IgM cryoglobulin with properties of a natural proteolytic antibody. *Biochem J*, 2006, 395, 473-481.
67. Sharma, L., J. Baker, A.M. Brooks, and A. Sharma, Study of IgM aggregation in serum of patients with macroglobulinemia. *Clinical chemistry and laboratory medicine*, 2000, 38, 759-764.
68. Colucci, M., H. Stockmann, A. Butera, A. Masotti, et al., Sialylation of N-linked glycans influences the immunomodulatory effects of IgM on T cells. *J Immunol*, 2015, 194, 151-157.
69. Valasek, C., J. Cole, F. Hensel, P. Ye, et al., Production and purification of a PER. C6-expressed IgM antibody therapeutic. *BioProcess Int*, 2011, 9, 28-37.

70. Lee, J., H.T. Gan, S.M. Latiff, C. Chuah, et al., Principles and applications of steric exclusion chromatography. *J Chromatogr A*, 2012, *1270*, 162-170.
71. Mader, A., V. Chromikova, and R. Kunert, Recombinant IgM expression in mammalian cells: A target protein challenging biotechnological production. *Advances in Bioscience and Biotechnology*, 2013, *04*, 38-43.
72. Hanala, S., The new ParaDIgM: IgM from bench to clinic: November 15-16, 2011, Frankfurt, Germany. *MAbs*, 2012, *4*, 555-561.
73. Collins, C., F.W. Tsui, and M.J. Shulman, Differential activation of human and guinea pig complement by pentameric and hexameric IgM. *Eur J Immunol*, 2002, *32*, 1802-1810.
74. Taylor, B., J.F. Wright, S. Arya, D.E. Isenman, et al., C1q binding properties of monomer and polymer forms of mouse IgM mu-chain variants. Pro544Gly and Pro434Ala. *The Journal of Immunology*, 1994, *153*, 5303-5313.
75. Parkhouse, R.M., B.A. Askonas, and R.R. Dourmashkin, Electron microscopic studies of mouse immunoglobulin M; structure and reconstitution following reduction. *Immunology*, 1970, *18*, 575-584.
76. Czajkowsky, D.M. and Z. Shao, The human IgM pentamer is a mushroom-shaped molecule with a flexural bias. *Proc Natl Acad Sci U S A*, 2009, *106*, 14960-14965.
77. Pabst, M., S.K. Kuster, F. Wahl, J. Krismer, et al., A Microarray-Matrix-assisted Laser Desorption/Ionization-Mass Spectrometry Approach for Site-specific Protein N-glycosylation Analysis, as Demonstrated for Human Serum Immunoglobulin M (IgM). *Mol Cell Proteomics*, 2015, *14*, 1645-1656.
78. Loos, A., C. Gruber, F. Altmann, U. Mehofer, et al., Expression and glycoengineering of functionally active heteromultimeric IgM in plants. *Proc Natl Acad Sci U S A*, 2014, *111*, 6263-6268.
79. Lu, J., M.J. Trnka, S.H. Roh, P.J. Robinson, et al., Improved Peak Detection and Deconvolution of Native Electrospray Mass Spectra from Large Protein Complexes. *J Am Soc Mass Spectrom*, 2015, *26*, 2141-2151.
80. Alagesan, K., M. Campbell, R. Hennig, M. Hoffmann, et al., GLYCO 23 XXIII International Symposium on Glycoconjugates - IgCarbKB: a glycomics and glycoproteomics focused immunoglobulin knowledge base *Glycoconj J*, 2015, *32*, 173-342.
81. Everest-Dass, A.V., D. Kolarich, M.P. Campbell, and N.H. Packer, Tandem mass spectra of glycan substructures enable the multistage mass spectrometric identification of determinants on oligosaccharides. *Rapid Commun Mass Spectrom*, 2013, *27*, 931-939.
82. Everest-Dass, A.V., J.L. Abrahams, D. Kolarich, N.H. Packer, et al., Structural feature ions for distinguishing N- and O-linked glycan isomers by LC-ESI-IT MS/MS. *J Am Soc Mass Spectrom*, 2013, *24*, 895-906.
83. Pabst, M., J.S. Bondili, J. Stadlmann, L. Mach, et al., Mass + Retention Time = Structure: A Strategy for the Analysis of N-Glycans by Carbon LC-ESI-MS and Its Application to Fibrin N-Glycans. *Anal Chem*, 2007, *79*, 5051-5057.
84. Jones, D., N. Kroos, R. Anema, B. van Montfort, et al., High-level expression of recombinant IgG in the human cell line per.c6. *Biotechnol Prog*, 2003, *19*, 163-168.
85. Pau, M.G., C. Ophorst, M.H. Koldijk, G. Schouten, et al., The human cell line PER.C6 provides a new manufacturing system for the production of influenza vaccines. *Vaccine*, 2001, *19*, 2716-2721.
86. Nguyen-Khuong, T., A.V. Everest-Dass, L. Kautto, Z. Zhao, et al., Glycomic characterization of basal tears and changes with diabetes and diabetic retinopathy. *Glycobiology*, 2015, *25*, 269-283.

87. Varki, A., R.D. Cummings, M. Aebi, N.H. Packer, et al., Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology*, 2015, 25, 1323-1324.

## **Chapter 5: Chemo-enzymatic *in vitro* glycoengineering for functionalisation of IgM**

*In vitro* glycoengineering of PAT-SM6 IgM anti-cancer antibody was carried out, following the characterisation of the site-specific glycosylation profile of PAT-SM6 in Chapter 4. A chemo-enzymatic method was used to modify PAT-SM6 IgM specific glycans with an azide group followed by the click chemistry addition of a functional group, thus enabling site-specific, potentially quantitative, labelling of IgM whilst not affecting the antigen binding domain.



## Introduction

### *Antibodies as research tools*

As discussed in the previous chapters, antibodies are biologically important proteins that have high specificity and high affinity towards their target antigen. Apart from using antibodies in biological settings as therapeutic biopharmaceuticals, they are highly sought after as a research tool. Many biomolecular methods exploit the specificity of antibodies towards their target; ELISA, western blotting, immunoprecipitation, immunohistochemistry and flow cytometry are some of the common molecular tools that use antibodies. The common feature of these experimental setups relies on the antibody binding to the target antigen for detection or visualisation to correlate with the presence of the antigen in question. This can be achieved by either detection of the bound antibody with a labelled secondary antibody, or by using a fluorescently labelled primary antibody for direct detection.

Direct detection of the antigen is a more straightforward process than detection using a secondary antibody, requiring less handling and experimental steps; for secondary detection, another set of controls showing the specificity of the secondary detection antibody towards the primary antibody is required. The advantage of secondary antibody detection lies in keeping the primary antibody in its native state, reducing complications in which the label used for direct detection might alter the primary antibody's specificity towards the target. Secondary antibody labelling on the other hand has general applicability since it targets the Fc region of any primary antibody. There are many ways to directly label an antibody, and many have been applied to IgG type antibodies [1-3]. These include direct chemical conjugation to the reactive functional groups on the protein backbone such as primary amines, carboxylic acids and thiols, or by genetically engineering functional domains such as green fluorescence protein, phycoerythrin and horseradish peroxidase onto the IgG. The genetically engineered functional domains are

more commonly used for secondary detection antibodies, as the approach labels the detection antibody during synthesis, rather than post purification.

### *Conjugation strategies for antibodies*

Chemical conjugation of IgG antibodies has been performed with a wide range of compounds, from small fluorescent molecules such as FITC and Cy5, to surfaces of nanoparticles such as gold nanoparticles [4] and quantum dots [5]. Amine coupling is the most commonly used strategy for conjugating IgG antibodies to the desired probe as at physiological pH, primary amines on lysine and arginine of the protein backbone can be reacted with many functional groups such as carboxyl, aldehydes and sulphonyl chloride to name a few. However, as the coupling reaction is non-discriminatory, the reacted primary amine could come from any lysine or arginine residue in the protein, and could result in disruption of the antibody binding to the antigen if the Fab region was affected.

Site specific labelling of IgG antibodies has thus been developed, targeting cysteine residues that are specifically introduced into the antibody [2, 6, 7], or targeting the glycans in the Fc region [8-10]. Site-specific cysteine labelling is performed by introducing cysteine residues onto the exposed surface of the IgG Fc region, increasing free, non-disulphide bonded cysteine residues that can be employed for thiol reactive conjugation chemistries [2, 6, 7]. However, the chemical conjugation step requires the partial reduction of the intermolecular disulphide bonds of the IgG, which would become problematic for the conjugation of polymeric IgM as it would reduce into monomeric IgM and result in loss of avidity. While the glycans on IgG have been shown to affect Fc effector functions, the antigen binding ability has not been found to be affected (except for a subset of antibodies that contains a glycosylated Fab region). IgG glycan functionalisation has been performed either by chemical means or by glycosyltransferase enzymes transferring functionalised nucleotide sugars to the glycan chains [8-10]. Chemical

modification of glycans involving the oxidation of the vicinal diols of the component monosaccharides into aldehyde groups can subsequently be ligated with amines or hydrazides following the formation of a Schiff's base [11]. By adjusting the periodate concentration, selective oxidation of the sialic acids can be achieved. This conjugation approach has been demonstrated in IgG, where sialic acid was introduced onto the IgG glycans by sialyltransferase activity before selective oxidation in which the aldehyde groups were conjugated with an amine functionalised anti-cancer drug [8]. However, the oxidation under the required mild acidic conditions (pH 5.6) might cause IgM denaturation and result in loss of function [12]. An alternative glycoconjugation approach using glycosyltransferases to add a functionalised nucleotide sugar substrate such as azide labelled CMP-sialic acid [13] or azide labelled UDP-GalNAc [9, 10] provides an easily reacted target for subsequent Click chemistry addition of any alkyne. The transfer of the azide-labelled nucleotide sugar by its respective glycosyltransferase is biologically compatible, but is dependent on the terminal monosaccharide presenting on the pre-existing glycan substrate on the antibody. This approach has been used with a modified B4GALT1 enzyme (Tyr289Leu) that transferred azide labelled UDP-GalNAc onto GlcNAc terminating glycans on IgG [9, 10]. To achieve higher incorporation efficiency, the IgG was first de-galactosylated to increase the available GlcNAc moieties prior to the addition of the azide-labelled GalNAc.

To date, published applications of direct labelling of IgM have mostly been performed via chemical conjugation methods [14-16], as the polymeric nature of IgM makes engineering genetically incorporated tags difficult. Like IgG labelling, primary amines have mainly been targeted for conjugation, as pentameric IgM contains approximately 400 lysines and arginines in total (41 lysine and arginine per IgM  $\mu$  chain, UniProt P01871). However, the non-discriminatory nature of chemical conjugation to amines may result in antigen binding issues as described before. Site-specific functionalisation of IgM via the glycans was first reported in

1984 [17], by reacting hydrazide-biotin with aldehydes generated from periodate oxidation of monosaccharide vicinal diols. Attempts at applying this approach to PAT-SM6 were met with excessive sample loss due to protein precipitation during the periodate oxidation and buffer exchange (Dr Nima Sayyadi, experimental observation). Unlike IgG which predominantly contains galactose or GlcNAc terminating glycans on a single *N*-glycosylation site, IgM is heavily glycosylated and we have shown that PAT-SM6 contains approximately 35% high mannose terminating glycans, 45% sialic acid terminating glycans and 20% galactose terminating glycans (Paper II, and partly in Table 1 below). To apply the modified

Identified glycan composition that are potential substrates for ST6GAL1	No. of free galactose for ST6GAL1	Glycopeptide site				
		IgM heavy chain				J-chain
		Asn171	Asn332	Asn395	Asn563	Asn71
man5 A1G1	1	2.71			0.83	
F man5 A1G1	1	6.13				
A1G1	1				0.72	
FA1G1F1	1	1.08				
A2G2S1	1				2.16	24.18
FA2G1	1		4.34	7.87		
FA2G1F1	1		2.25	2.47		
FA2G2F1S1	1	16.08	15.62	10.86		
FA2G2S1	1	19.15	34.81	16.84	1.25	9.97
A3G3S2	1				0.51	
FA3G2	1	0.90	2.92	6.84		
FA3G2F1	1		1.05	1.50		
FA3G2F1S1	1	0.44				
FA3G2S1	1		1.60	1.74		
FA3G3S2	1		2.69	2.06		
A2G2	2				0.89	20.02
FA2G2	2	4.74	5.04	13.25		20.02
FA2G2F1	2	5.35	5.39	9.99		
A3G3S1	2					4.77
A3G3S1F1	2				0.10	
FA3G1	2		3.21	5.04		
FA3G3S1	2		7.75	2.46		
A3G3	3					13.08
FA3G3	3		1.28	1.12		1.22
FA3G3F1	3		2.52			
A4G4	4					1.95
% non sialylated glycans		20.91	28.01	48.08	2.43	43.22
% Total at each glycosylation site		56.58	90.48	82.04	6.45	95.21
% contribution in Pentameric IgM		11.1	17.7	16.1	1.3	1.9

Table 1. Summary of glycan compositions and their relative abundance on the respective glycopeptides that are potential substrates for ST6GAL1 based on glycopeptide quantitation data from Paper II (Supplementary S5). Glycan compositions are listed in Oxford nomenclature as described in Paper II (Supplementary S5); prefix F → core fucose, man → no. of mannose, A → no. of antenna GlcNAc, G → no. of galactose, F → no. of outer arm fucose, S → no. of sialic acid. % contribution in pentameric IgM was calculated by scaling the summed abundance against the total number of glycosylations sites in pentameric IgM.

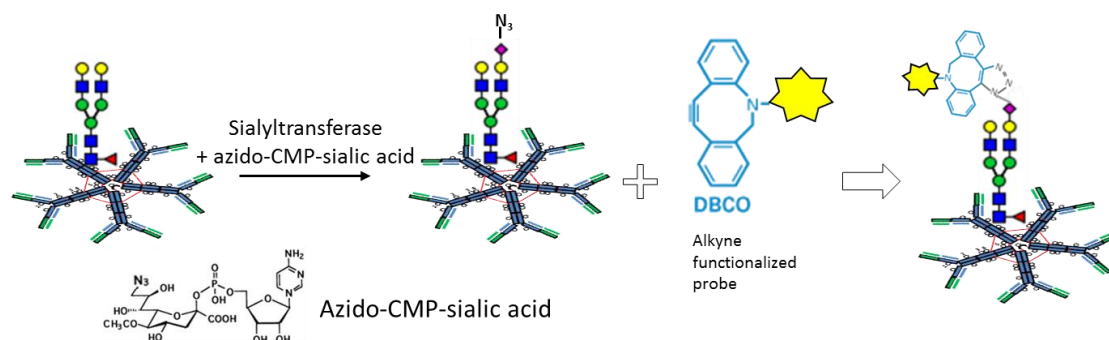


Figure 1. Schematic representation of the overall functionalisation procedures. The azide functional group is first transferred onto the IgM glycans by sialyltransferase activity, and can be specifically ligated with the cyclic-octyne functionalised probe by copper-free click chemistry

galactosyltransferase B4GALT1 enzyme method as used on IgG to add azide labelled UDP-GalNAc onto GlcNAc terminating glycans on IgM would require an additional sialidase treatment to first expose the galactose for further galactosidase activity, which has been seen to result in PAT-SM6 precipitation (personal experimental observation).

Building from the glycopeptide analysis of PAT-SM6 in Paper II (Supplementary figure S5), approximately 48% of the glycans of the attached *N*-glycans are potential candidates (i.e. contain a terminal galactose substrate) for sialylation (Table 1). The aim of the next part of the thesis was to use the detailed knowledge gained on the glycosylation of PAT-SM6 to specifically engineer these glycans in a single step by using sialyltransferase ST6GAL1 to incorporate azide labelled CMP-NeuAc onto the galactose terminating glycans. The azide functional group can then be conjugated with any alkyne terminating label, such as a fluorescent probe or anti-cancer drug, using the efficient Click chemistry (Figure 1).

## Materials and Methods

### *Azide incorporation into IgM glycans using Azido-CMP-sialic acid and ST6GAL1 enzyme*

50µg of PAT-SM6 (Patrys Ltd, produced by Patheon), was mixed with 2.5µg of recombinant human ST6GAL1 (R&D systems) in a 50µl reaction mixture containing 10mM HEPES pH7.5, 10mM MnCl<sub>2</sub>, 5mM CaCl<sub>2</sub>, 5mM MgSO<sub>4</sub> and 1mM of azido-CMP-sialic acid (R&D systems),

and incubated overnight at 37°C. An aliquot of the reaction was used for confirmation of the azide-sialic incorporation on to the IgM glycans.

*Click reaction of azide-sialic labelled glycans with Cy5-DBCO*

The excess azido-CMP-sialic acid was removed by buffer exchange using a 10kDa molecular weight cut-off centrifugal filter (Millipore) with 3 rounds of washing with 10mM HEPES, pH 7.5, 150mM NaCl and 10% (v/v) glycerol, resulting in a final volume of approximately 100µl. Cy5-Dibenzocyclooctyne (DBCO)(Lumiprobe) was resuspended in 1ml DMSO to a concentration of 0.93mM, and 10µl was added to the PAT-SM6 solution and incubated overnight at room temperature in the dark. Excess Cy5-DBCO was removed by a 10kDa molecular weight cut-off centrifugal filter. An aliquot corresponding to approximately 5µg of IgM was run on an SDS-PAGE gel for fluorescent visualisation of the conjugation. The SDS-PAGE gel was first fixed in 50% (v/v) ethanol, 10% (v/v) acetic acid for 15 mins, washed twice with distilled water and visualised using the Typhoon Trio imager (GE Healthcare) at 633nm with the photomultiplier tube set at 250V. Subsequently, the same gel was stained with Coomassie Blue for detection of protein corresponding to the Cy5 fluorescence.

*Confirmation of azide-sialic incorporation into IgM glycans*

Approximately 15µg of labelled PAT-SM6 was reduced in 10mM DTT at 50°C for 1h, and alkylated with 25mM IAA in the dark for 1h. The mixture was split into 3 (5µg each) for 1) glycan analysis by release with PNGaseF (Promega), and glycopeptide analysis using 2) Trypsin (Promega) or 3) GluC (Promega). Glycan analysis was carried out as described in my Paper III. In brief, the reduced and alkylated IgM was dot-blotted onto a PVDF membrane and glycans released with 5U of PNGaseF overnight at 37°C. The glycans were then collected, reduced and desalted for analysis by PGC-LC-MS. Glycopeptide analysis was carried out as described in my Paper II. Briefly, the reduced and alkylated IgM was added to 0.1µg of trypsin

or 0.1 $\mu$ g of GluC constituted in 100mM ammonium bicarbonate, pH 8, in a reaction volume of 19 $\mu$ l. Digested glycopeptides were added to a solution of 80 $\mu$ l of 100% ACN and 1 $\mu$ l of 100% TFA to create a final concentration of 80% ACN, 1% TFA for glycopeptide enrichment using ZIC-HILIC as described in Paper II[18]. The enriched glycopeptides were then dried and reconstituted in 0.1% (v/v) formic acid for analysis by nLC-ESI-MS/MS on the Thermo QExactive as described in Paper IV, with slight modifications to the acquisition mass range and separation gradient; acquisition  $m/z$  600-1800 and a gradient of 0-30% (v/v) ACN in 0.1% (v/v) formic acid over 75min. The acquired MS/MS spectra were searched using Byonic (Protein Metrics), using a protein database containing human Ig  $\mu$  chain (Uniprot P01871), human Ig J-chain (Uniprot P01591) and human ST6GAL1 (Uniprot P15907) with peptide modifications of Cys carbamidomethylation (fixed), variable modifications of Met oxidation, Met dethiomethylation, Asn and Gln deamidation, and the corresponding proteolytic cleavages of trypsin and GluC. As the azide-labelled sialic acid used in the labelling has an additional mass of 25Da compared to non-derivatised sialic acid (the azide group of 42Da replaces the hydroxyl group of 17Da on the 9<sup>th</sup> carbon on the sialic acid [13]), a custom glycan database containing the modification mass was created for the Byonic search. The Byonic output of identified glycopeptides was then manually curated with multiple criteria as described in Paper IV such as Byonic score (>150), presence of glycopeptide oxonium ions, and peptide retention window.

## Results

The non-discriminatory nature of chemical conjugation usually results in random chemical conjugation of the targeted functional group on the protein (excluding targeted thiol labelling of engineered cysteines [6]), that may cause problems in labelling reproducibility. Targeting an exact location of a modification provides the potential of quantitation and better understanding of interactions. In this work, the glycans of PAT-SM6 IgM were functionalised by the enzymatic addition of an azide-labelled sialic acid for multipurpose labelling by alkyne

functionalised compounds using the well-known specific chemical ligation strategy of the “click” reaction between an azide and alkyne. As shown in Table 1, approximately 48% of the glycans on PAT-SM6 are potential substrates for ST6GAL1, which could lead to a theoretical maximum of 25 azide additions per PAT-SM6 pentamer (assuming 1 azide per potential glycan substrate, regardless of number of free galactose). The potential precursor glycans and glycopeptides and the final labelled products were characterised to confirm the reactions.

#### *Characterisation of released glycans from IgM before and after azide-sialic acid labelling*

After the reaction of PAT-SM6 IgM with ST6GAL1 and azido-CMP-sialic acid, the *N*-glycans released from the IgM by PNGaseF showed a significant decrease in the relative abundance of the digalactosylated, biantennary glycan structure FA2G2 (approximately 7.5% to 0.77%), which is a substrate for ST6GAL1 (Figure 2). An observable decrease in relative abundance was also observed for other non-sialylated glycans such as FA2G1, FA2G2F1 and FA2G2F2. Interestingly, the relative abundance of the mono-sialylated, bi-antennary glycans were unchanged, suggesting that either the ST6GAL1: PAT-SM6 ratio of 1:20 was insufficient to catalyse the transfer of azido-CMP-sialic acid to glycans with a pre-existing sialic acid and a

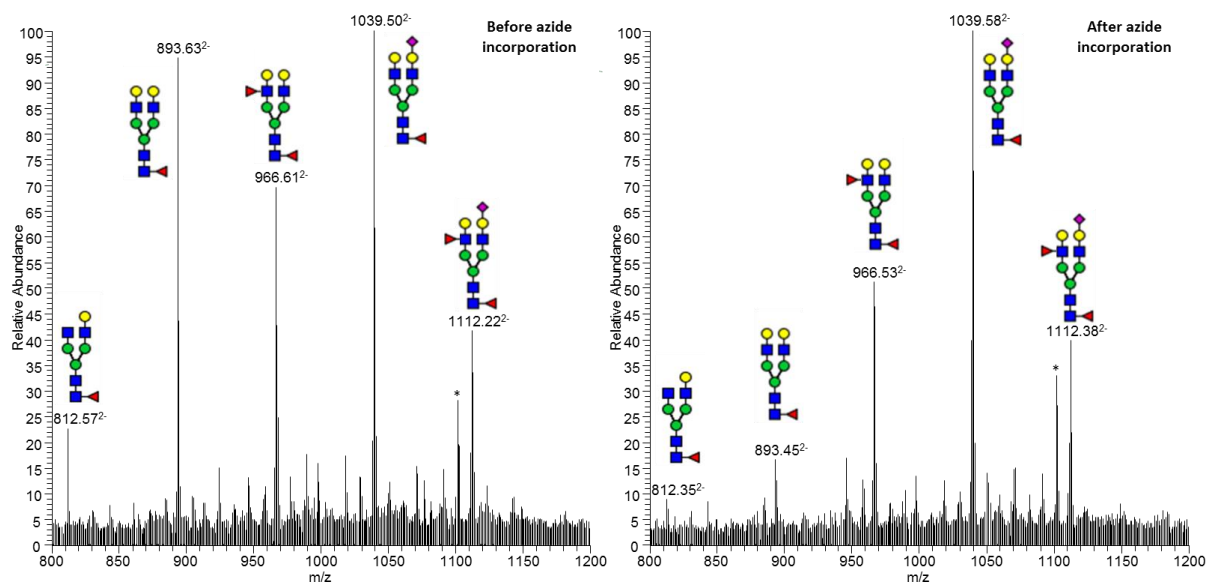


Figure 2. Averaged MS spectra showing relative abundance of released glycans, across the retention time where bi-antennary complex glycans eluted, before and after the reaction of PAT-SM6 IgM with ST6GAL1 and azido-CMP-sialic acid. Only the glycans with high abundance were annotated, showing a decrease in the non-sialylated precursor structures after the azide incorporation. \* - contaminant



free galactose, or that ST6GAL1 could not act on the monosialylated biantennary glycan as a substrate. A recent publication on large scale *in vitro* glycoengineering of IgG by enzymatic additions reported the use of a sialyltransferase (in development by Merck) that preferentially forms di-sialylated glycans [19], suggesting that the latter is more likely to be the case for ST6GAL1. Surprisingly, the azide labelled product glycan (additional +25Da, resulting from the substitution of the -OH group on the 9<sup>th</sup> carbon to N<sub>3</sub> [13]) was not detected in the mass spectrum after the reaction of the IgM with ST6GAL1 and azido-CMP-sialic acid. Initially, it was hypothesized that the azide labelled glycan was lost during the clean-up procedures (strong cation exchange and enrichment on carbon resin), or that it was retained in the PGC column. It was discovered, in retrospect, that the azide functional group had been lost during the reduction step of *N*-glycans by sodium borohydride [20] during the preparation of the sample for PGC separation, resulting in the formation of an amine group and resulting in an overall loss of 1Da

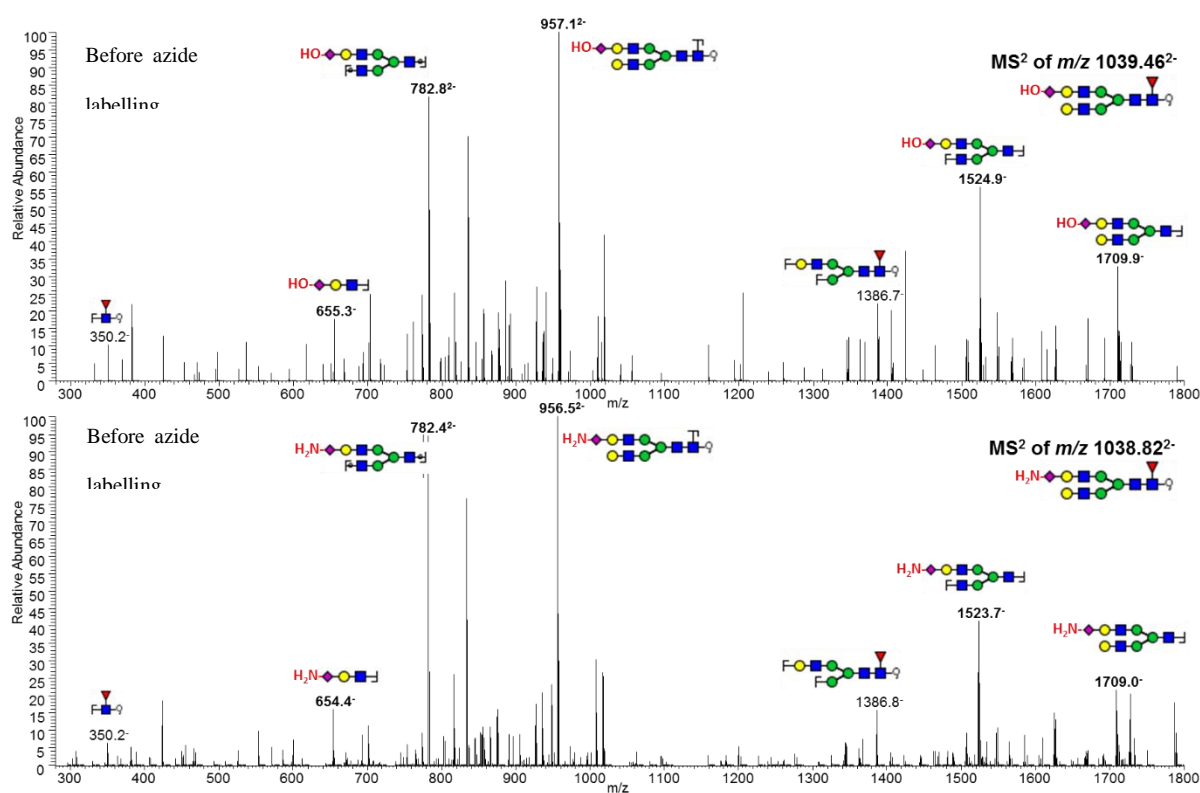


Figure 3. Fragment ions of the FA2G2S1 glycan before and after azide labelling. Reduction of glycans by sodium borohydride reduces the azide group into an amine group, resulting in a mass change of -1Da. The -1Da difference was observed in several fragment ions of the theoretical product of the azide labelling (FA2G2 → FA2G2S1-N<sub>3</sub> → FA2G2S1-NH<sub>2</sub>). Symbols denoting the cleavage position were based on GlycoWorkBench.

compared to the precursor reduced glycan structures. Evidence for this is based on the tandem mass spectrum of the postulated reduced product FA2G2S1-NH<sub>2</sub> (observed  $m/z$  1038.8<sup>2-</sup>), where several of the fragment ions containing sialic acid were 1  $m/z$  less ( $\sim 0.5$   $m/z$  for doubly charged ions) than the corresponding ion of the sialylated equivalent FA2G2S1-OH (Figure 3). However, the relative abundance of the product FA2G2S1-NH<sub>2</sub> (approximately 1%) was lower than the decrease of the precursor FA2G2 (decreased by approximately 7%), suggesting that there might be differences between the two in negative mode ionisation. We hypothesize that the -NH<sub>2</sub> group resulting from the reduction of the azide group may remain protonated under the separation conditions of pH 8.0, resulting in some of the ions gaining less charge in negative ionisation mode, thereby remaining as singly charged ions, outside of the acquisition window of 600-2000  $m/z$  (singly charged FA2G2S1-NH<sub>2</sub> is 2078.8  $m/z$ ). For accurate quantitation of glycan relative abundance, quantitation of all identified charge states of the same glycan is important, and had been discussed in Paper I.

### *Characterisation of glycopeptides of IgM before and after azide-sialic acid labelling*

Based on the site-specific glycosylation analysis on PAT-SM6 described in Paper II, it is known that the incorporation of the azido-CMP-sialic acid could only predominantly occur at three *N*-glycosylation sites, Asn171, Asn332 and Asn395, which carry complex-type *N*-glycans as well as on a small proportion of Asn563 and the J-chain (Table 1). From the released glycan analysis, the greatest decrease in precursor relative abundance was found to be the bi-antennary di-galactosylated glycan (FA2G2). The relative abundance of FA2G2 was highest at site Asn395 (Table 1) and indeed, the mass of the Asn395 glycopeptide with the azide-sialic acid incorporated onto FA2G2 was found (Figure 4) where a decrease in the precursor Asn395 with FA2G2, and appearance of the azide labelled FA2G2S1 at a delayed retention time was observed (Figure 4, A). Azide-sialic labelled glycopeptides of other glycan galactosylated structures such as FA2G2F1 and FA2G2F2 were also observed at lower abundance with a

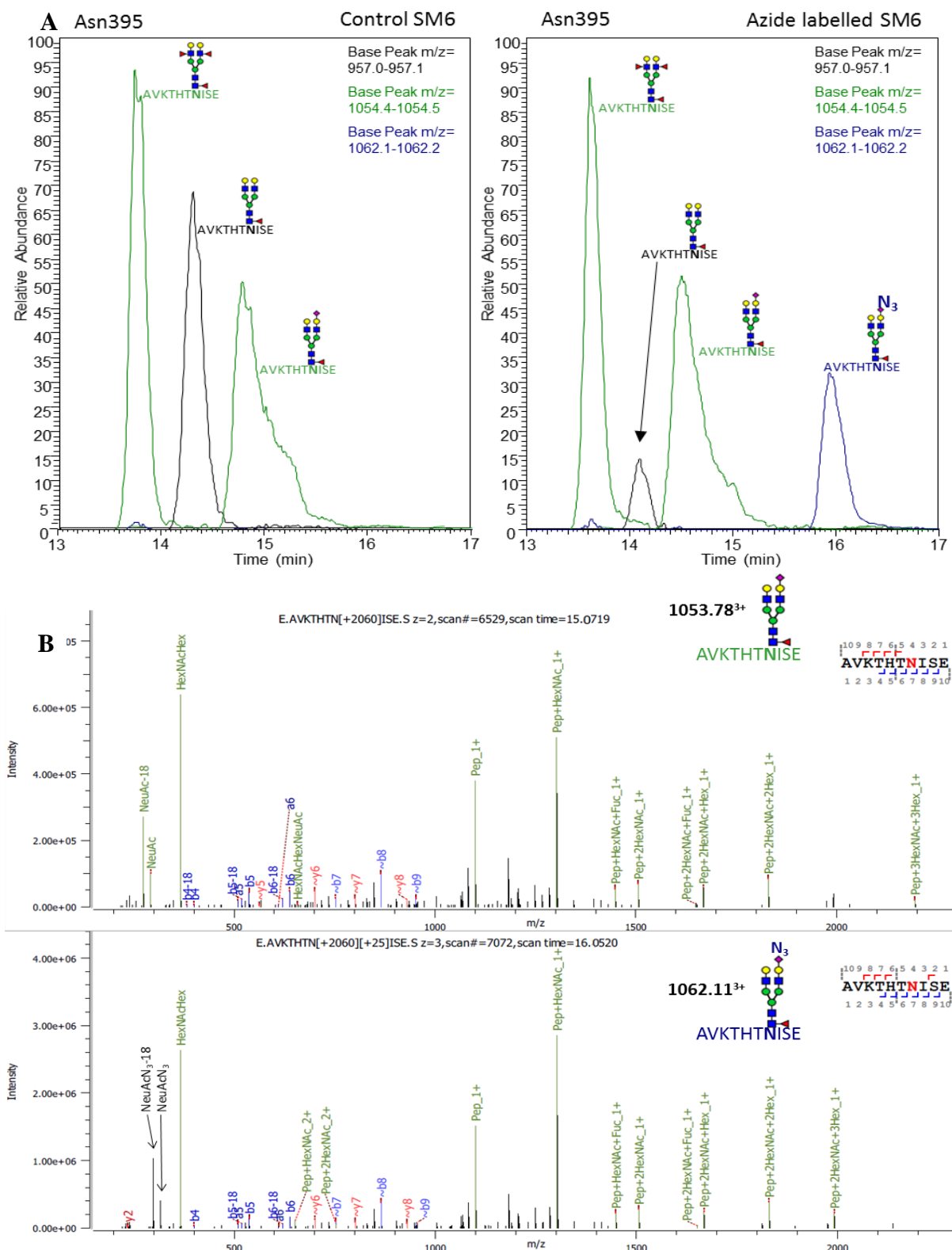


Figure 4. (A) Representative extracted ion chromatograms of triply charged GluC glycopeptides, Asn395 peptide, showing the successful incorporation of the azide from the transfer of azido-CMP-sialic acid by ST6GAL1, using Bionic assisted identification. The presence of 2 peaks at 1054.4<sup>3+</sup> is due to the small mass difference between 1 sialic acid and 2 fucoses (1 sialic acid = 291Da, 2 fucose = 292 Da). (B) Bionic identification of the azide labelled Asn395 glycopeptide, with identical b- and y-ions, differing by the precursor mass and the presence of azide labelled sialic fragments (arrow).

similar delayed retention relative to the non-azide containing sialylated analogue (data not shown). HCD fragmentation of the azide labelled glycopeptides showed similar fragmentation

profiles to the sialylated, non-azide containing glycopeptides. This was expected as HCD fragmentation yields mostly peptide backbone ions (b-/y-ions, Y-ions) that do not contain the terminal glycan fragments (Figure 4, B). The azide containing sialic acid fragments could also be detected as  $m/z$  317 (NeuAcN<sub>3</sub>) and  $m/z$  299 (NeuAcN<sub>3</sub>-18) in the MS/MS of the azide labelled glycopeptides. By area under curve quantitation of Asn395 glycopeptides, approximately 15% of the Asn395 glycopeptides had the azide-sialic acid group incorporated, corresponding to 1-2 labelled Asn395 per pentameric IgM. Presence of the azide label was also found on Asn171, at approximately 5% incorporation (data not shown), that corresponds to 0-1 labelled Asn171 per pentameric IgM. No azide labelled glycopeptides were identified on Asn332, Asn563 and Asn71 (J-chain). For Asn332, it could be because Asn332 contained a higher percentage of already sialylated glycans (approximately 70%) than Asn395 (approximately 40%) (Paper II), and had less possible substrates for ST6GAL1 (28% non sialylated glycans for Asn332, compared to 48% for Asn395, Table 1). For Asn563 and J-chain Asn71, low abundance of the glycopeptides could have hindered the detection of azide labelled glycopeptides; Asn563 only had 2.4% non-sialylated glycans, while J-chain is ten-fold intrinsically less abundant than other glycopeptides (1 J-chain per 10 IgM heavy chain). The incorporation of azide-sialic acids thus sums to be in the order of 1-3 azide functional groups per pentameric IgM.

### *Fluorescence visualisation of labelled IgM-Cy5 by SDS-PAGE*

To verify that the azide functionalised PAT-SM6 was labelled by the “click” chemistry addition of an alkyne to the azide-sialic acid, Cy5-Dibenzocyclooctyne (DBCO) was used to ligate a fluorescent small molecule by copper-free click chemistry, for visualisation by SDS-PAGE. Cy5 label is commonly used in two-dimensional fluorescence difference gel electrophoresis (2D-DIGE) and is not sensitive to conditions used in SDS-PAGE (boiling temperatures and

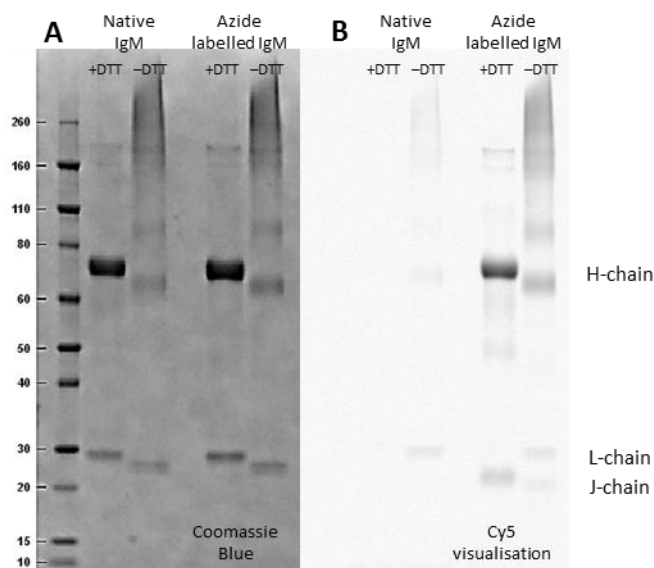


Figure 5. SDS-PAGE of azide labelled PAT-SM6 under reducing and non-reducing conditions, visualised by Coomassie blue (A) or Cy5 fluorescence (633nm) (B). “Click” reaction between DBCO-Cy5 and azide labelled sialic acid was performed on azide-sialic labelled and unlabelled IgM. Cy5 fluorescence was observed for the labelled IgM on both the heavy chain and the J-chain.

reducing agents) [21]. PAT-SM6 was run in both reducing and non-reducing conditions to verify that the polymeric nature of IgM was preserved during the labelling procedures (Figure 5, A, – DTT). Both the intact IgM and the IgM heavy chain were highly fluorescent under detection at 633nm (Figure 5, B). The J-chain, which was not visible under Coomassie stain due to the much lower abundance (1 J-chain for every 10 IgM heavy/light chain), was also observed as being fluorescently labelled, suggesting that although azide-sialic incorporated into the J-chain glycopeptide was not detected in the mass spectrometric glycopeptide analysis, azide incorporation did occur on the galactose-terminated J-chain glycans. As there is only 1 glycosylation site on the J-chain, and non-azide labelled J-chain Asn71 glycopeptide in the labelled sample was identified by Byonic (data not shown), it suggests that the incorporation of azide onto J-chain would be 0-1. This brings the theoretical number of labels to 1-4 per IgM pentamer (1-3 on IgM heavy chains, 0-1 on J-chain).

## Discussion

Functionalisation of proteins with azide groups serves as a useful tool for small molecule targeting by chemical specific ligation [13, 22-24]. In the natural biological system, azide functional groups are not present thus reducing background interference as compared with

lysine or thiol targeting methods. Development of copper-free click chemistry reactions such as azide coupling to the small molecule alkynes, cyclic-octynes and phosphine probes, have enabled biologically compatible chemical ligation environments suitable for cell labelling of metabolically labelled glycoproteins *in vivo* [13, 23, 25].

In this work, the possibility of specific-site glycoengineering for labelling of IgM by chemo-enzymatic means has been demonstrated. Specific site functionalisation of the PAT-SM6 IgM glycans was enabled by transfer of azido-CMP- sialic acid using ST6GAL1 under biologically compatible reaction conditions. Labelling of IgM on these Fc-glycans should not affect the antigen targeting capacity, as has been shown with IgG, making this labelling method generally useful for targeting any cellular IgM antigens. Based on the analysis of azide labelled PAT-SM6 glycopeptides, a range of 1-4 azide functional groups was introduced onto the pentameric IgM glycans. By knowing the number of azide functional groups on the IgM, provides a theoretical maximum of the number of subsequent conjugates, which could either be a fluorescent molecule as shown in this work, or an anti-cancer drug, for example, to create IgM-antibody-drug conjugates. Knowing the number of conjugates for antibody drug conjugates is important as it affects the drug concentration and dosage and in turn the efficacy of the treatment [2]. Validation of the final antibody-drug conjugate still poses a technical challenge for IgM analysis; native mass spectrometry can be employed to measure conjugation outcomes of intact IgG [26], a technique still unavailable for pentameric IgM. For fluorescent probes or small molecule drugs with distinct spectroscopic properties, UV-Vis/fluorescence spectrometry methods could be used to quantitate the concentration of conjugates based on known standards as a ratio against the IgM amount offers a verification and quantitation method.

While the final estimated number of labelled molecules per IgM was quite low for such a large antibody (ranging from 1-4), more labels could potentially be incorporated by addition of a different monosaccharide using azide-labelled GDP-fucose and fucosyltransferase III or VI,

resulting in glycan outer-arm azide-fucosylation. Fucosyltransferase VI can accept sialylated glycans as substrates [27], and could result in the functionalisation of the Asn332 glycans, which was unable to be labelled by ST6GAL1. Combination of a sialyltransferase with a fucosyltransferase could form Sialyl-Lewis<sup>X</sup> epitopes which may contain up to 2 azide functional groups per epitope (NeuAcN<sub>3</sub>Fuc<sub>1</sub>Gal<sub>1</sub>GlcNAc<sub>1</sub>, NeuAc<sub>1</sub>FucN<sub>3</sub>Gal<sub>1</sub>GlcNAc<sub>1</sub> or NeuAcN<sub>3</sub>FucN<sub>3</sub>Gal<sub>1</sub>GlcNAc<sub>1</sub>). Another approach would be to engineer the azide label into the glycan using metabolic labelling of the cells; producing the IgM by feeding with ManNAz, an azide labelled ManNAc, which would be converted into CMP-Neu5Ac-azide *in vivo*, and subsequently transferred on to proteins as NeuAc-azide [28]. This approach is currently used as a cell based labelling technique to visualise differences in the glycome and has been used on live cells [25] and even zebrafishes [29]. As this approach is non-discriminatory of the protein; any glycoprotein that normally carries the sialic acid would be labelled in this manner.

It is important to note that while labelling antibodies at the glycans on the Fc region of IgM should not affect antigen binding, interaction between IgM Fc glycans and cell surface receptors of white cells has been shown to be affected by sialylation of the IgM glycans [30] as desialylation of human serum IgM removed the internalisation response to IgM by T-cells. In that study, by testing human serum IgM, it was not known which receptor on the T-cell mediated the IgM-T-cell interaction. The human serum IgM was also not complexed with an antigen. Hence, it is unclear if this difference only impacts free IgM or if it is T-cell and white cell specific. Nonetheless, the potential of this site-specific labelling of IgM in medical applications is very high in such uses as quantitative discriminating diagnostic tools and as a means of forming targeted drug conjugates.

## **Conclusion**

In conclusion, the possibility of functionalising and conjugating small molecules site specifically onto IgM glycans under biologically compatible conditions was shown. By functionalising the IgM via the glycans, which are away from the Fab region, the specificity of IgM should not be altered, and further experiments would be carried out to verify this aspect.



## References

1. Beck, A., Review of Antibody-Drug Conjugates, Methods in Molecular Biology series. *mAbs*, 2013, 6, 30-33.
2. Panowski, S., S. Bhakta, H. Raab, P. Polakis, et al., Site-specific antibody drug conjugates for cancer therapy. *MAbs*, 2014, 6, 34-45.
3. Teicher, B.A., Antibody drug conjugates. *Curr Opin Oncol*, 2014, 26, 476-483.
4. Jazayeri, M.H., H. Amani, A.A. Pourfatollah, H. Pazoki-Toroudi, et al., Various methods of gold nanoparticles (GNPs) conjugation to antibodies. *Sensing and Bio-Sensing Research*, 2016, 9, 17-22.
5. Wang, Y. and L. Chen, Quantum dots, lighting up the research and development of nanomedicine. *Nanomedicine*, 2011, 7, 385-402.
6. Junutula, J.R., H. Raab, S. Clark, S. Bhakta, et al., Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nat Biotechnol*, 2008, 26, 925-932.
7. Agarwal, P. and C.R. Bertozzi, Site-Specific Antibody–Drug Conjugates: The Nexus of Bioorthogonal Chemistry, Protein Engineering, and Drug Development. *Bioconjug Chem*, 2015, 26, 176-192.
8. Zhou, Q., J.E. Stefano, C. Manning, J. Kyazike, et al., Site-specific antibody-drug conjugation through glycoengineering. *Bioconjug Chem*, 2014, 25, 510-520.
9. Zeglis, B.M., C.B. Davis, R. Aggeler, H.C. Kang, et al., Enzyme-mediated methodology for the site-specific radiolabeling of antibodies based on catalyst-free click chemistry. *Bioconjug Chem*, 2013, 24, 1057-1067.
10. Qasba, P.K., E. Boeggeman, and B. Ramakrishnan, Site-specific linking of biomolecules via glycan residues using glycosyltransferases. *Biotechnol Prog*, 2008, 24, 520-526.
11. Hermanson, G.T., *Chapter 2 - Functional Targets for Bioconjugation*, in *Bioconjugate Techniques (Third edition)* 2013, Academic Press: Boston. p. 127-228.
12. Sharma, L., J. Baker, A.M. Brooks, and A. Sharma, Study of IgM aggregation in serum of patients with macroglobulinemia. *Clinical chemistry and laboratory medicine*, 2000, 38, 759-764.
13. Wu, Z.L., X. Huang, A.J. Burton, and K.A.D. Swift, Glycoprotein labeling with click chemistry (GLCC) and carbohydrate detection. *Carbohydrate Research*, 2015, 412, 1-6.
14. Watanabe, M., K. Fujioka, N. Akiyama, H. Takeyama, et al., Conjugation of Quantum Dots and JT95 IgM Monoclonal Antibody for Thyroid Carcinoma Without Abolishing the Specificity and Activity of the Antibody. *IEEE Transactions on NanoBioscience*, 2011, 10, 30-35.
15. Kinsey, B.M., R.M. Macklis, J.M. Ferrara, W.W. Layne, et al., Efficient conjugation of DTPA to an IgM monoclonal antibody in ascites fluid. *International Journal of Radiation Applications and Instrumentation. Part B. Nuclear Medicine and Biology*, 1988, 15, 285-292.
16. Brandlein, S., N. Rauschert, L. Rasche, A. Dreykluft, et al., The human IgM antibody SAM-6 induces tumor-specific apoptosis with oxidized low-density lipoprotein. *Mol Cancer Ther*, 2007, 6, 326-333.
17. O'Shannessy, D.J., M.J. Dobersen, and R.H. Quarles, A novel procedure for labeling immunoglobulins by conjugation to oligosaccharide moieties. *Immunol Lett*, 1984, 8, 273-277.

18. Moh, E.S., C.H. Lin, M. Thaysen-Andersen, and N.H. Packer, Site-Specific *N*-Glycosylation of Recombinant Pentameric and Hexameric Human IgM. *J Am Soc Mass Spectrom*, 2016, 27, 1143-1155.
19. Thomann, M., T. Schlothauer, T. Dashivets, S. Malik, et al., In Vitro Glycoengineering of IgG1 and Its Effect on Fc Receptor Binding and ADCC Activity. *PLoS One*, 2015, 10, e0134949.
20. Ali, Y. and A.C. Richardson, The reduction of azides with sodium borohydride: a convenient synthesis of methyl 2-acetamido-4,6-*O*-benzylidene-2-deoxy- $\alpha$ -d-allopyranoside. *Carbohydrate Research*, 1967, 5, 441-448.
21. Tannu, N.S. and S.E. Hemby, Two-dimensional fluorescence difference gel electrophoresis for comparative proteomics profiling. *Nat Protoc*, 2006, 1, 1732-1742.
22. Baskin, J.M., J.A. Prescher, S.T. Laughlin, N.J. Agard, et al., Copper-free click chemistry for dynamic in vivo imaging. *Proceedings of the National Academy of Sciences*, 2007, 104, 16793-16797.
23. Chang, P.V., J.A. Prescher, E.M. Sletten, J.M. Baskin, et al., Copper-free click chemistry in living animals. *Proceedings of the National Academy of Sciences*, 2010, 107, 1821-1826.
24. Belardi, B., A. de la Zerda, D.R. Spiciarich, S.L. Maund, et al., Imaging the Glycosylation State of Cell Surface Glycoproteins by Two-Photon Fluorescence Lifetime Imaging Microscopy. *Angewandte Chemie International Edition*, 2013, 52, 14045-14049.
25. Hubbard, S.C., M. Boyce, C.T. McVaugh, D.M. Peehl, et al., Cell surface glycoproteomic analysis of prostate cancer-derived PC-3 cells. *Bioorganic & Medicinal Chemistry Letters*, 2011, 21, 4945-4950.
26. Rosati, S., E.T.J. van den Bremer, J. Schuurman, P.W.H.I. Parren, et al., In-depth qualitative and quantitative analysis of composite glycosylation profiles and other micro-heterogeneity on intact monoclonal antibodies by high-resolution native mass spectrometry using a modified Orbitrap. *mAbs*, 2013, 5, 917-924.
27. Thomas, L.J., K. Panneerselvam, D.T. Beattie, M.D. Picard, et al., Production of a complement inhibitor possessing sialyl Lewis X moieties by in vitro glycosylation technology. *Glycobiology*, 2004, 14, 883-893.
28. Saxon, E., S.J. Luchansky, H.C. Hang, C. Yu, et al., Investigating cellular metabolism of synthetic azidosugars with the Staudinger ligation. *J Am Chem Soc*, 2002, 124, 14893-14902.
29. Laughlin, S.T., J.M. Baskin, S.L. Amacher, and C.R. Bertozzi, In Vivo Imaging of Membrane-Associated Glycans in Developing Zebrafish. *Science*, 2008, 320, 664-667.
30. Colucci, M., H. Stockmann, A. Butera, A. Masotti, et al., Sialylation of *N*-linked glycans influences the immunomodulatory effects of IgM on T cells. *J Immunol*, 2015, 194, 151-157.

## **Chapter 6**

Thesis summary and conclusion

## Thesis summary

In this dissertation, 3 different aspects of *in vitro* glycoengineering were addressed.

To deal with the problem of requiring multiple enzymes to remodel glycoproteins *in vitro*, a self-assembling, DNA guided protein scaffold, aimed to spatially align multiple glycosyltransferases was designed to mimic the sequential enzymatic reactions of glycosyltransferases in the Golgi apparatus (Chapter 2). Although it was not proven to work in the current design, the design flaws identified from this exercise highlight potential improvements for the next attempt at creating these adaptor proteins that could conceivably be adapted for any multi-enzymatic pathway reactions *in vitro*.

To deal with the problem of the “once off” use of glycosyltransferases in *in vitro* glycoengineering a method using purified enzymes, applicability of an immobilised glycosyltransferase in a resin-based column stationary phase format was designed (Chapter 3). Based on the preliminary data, a functional proof-of-concept was created, showing the ability to galactosylate glycoproteins by simply applying the glycoprotein onto the immobilised enzyme column and harvesting the column output. To further this work, multiple physical parameters of the reaction conditions need to be controlled and optimised. This approach provides a basis for joining multiple immobilised glycosyltransferases sequentially to create a synthetic Golgi apparatus.

With current *in vitro* glycoengineering efforts, the biomedical potential of IgG has been improved upon and developed as a research tool and for advanced treatment drugs. IgM, as described, has been largely overlooked due to the technical challenges that come with its size and complexity. With the advancements in mass spectrometric techniques, site-specific glycosylation characterisation of IgM is now possible (Chapter 4), and served as a basis for an *in vitro* glycoengineering application in this thesis. PAT-SM6, a natural IgM that has been

commercially developed for detection of cells in multiple cancer types, can, with the work described here, now become potentially useful as a diagnostic tool for direct cancer imaging, cancer cell tracking and drug delivery using the functionalised IgM. Targeted addition to IgM glycans, away from the antigen binding site, of azide functional groups by chemo-enzymatic means (Chapter 5), provides the potential for many applications such as site-specific addition of small molecule probes that are MRI reactive or have anti-cancer properties.

## **Conclusion**

*It has always been about the dollars*

While it might not be plainly evident, cost effectiveness is a major underlying theme of this dissertation, and is a driver of the research ideas that were put forward. Immobilisation of glycosyltransferases onto the DNA mediated multi-enzymatic scaffold is designed with cost effectiveness in mind. Without the presence of a protein scaffold, to glycosylate a protein *in vitro* can only be performed by either sequentially introducing glycosyltransferases and their nucleotide sugar substrates into the reaction, or by mixing all the transferases and substrates together in a single pot reaction. The former takes a long time to complete and requires purification steps between reactions, and the latter results in a population of “under-glycosylated” proteins caused by reaction kinetic differences between enzymes. Implementation of the TALE-mSA scaffold can produce the substrate channelling effect and enhance the overall reaction kinetics by disregarding diffusion of the reaction “intermediate” into the solution environment for the subsequent enzyme to use as a substrate (Chapter 2). The scaffold orientation would reduce the overall reaction time, and “force” the reactions to completion. The use of DNA as the foundation is not just for the spatial alignment of the individual scaffold adaptors, but also to lower the cost of the design. Custom double stranded DNA and custom DNA primers with functional groups such as 5'-biotin or 5'-NH<sub>2</sub> can now be

readily purchased at low cost, as compared to protein based scaffolds that require expression and purification procedures. Amplification of DNA by polymerase chain reaction using the functionalised primers produces a steady supply of DNA programs that can be immobilised onto a stationary phase such as streptavidin resins or carboxyl functionalised surfaces and particles. The combination of these would lead to a single pot or column, re-usable, multi-enzymatic immobilised enzyme reactor that is theoretically able to perform like an artificial Golgi apparatus. Even though the current scaffold design had flaws that require revisiting, a simpler alternative, where the individual glycosyltransferases could be separately immobilised (Chapter 3) and lined up as consecutive IMERs, could be more readily implemented for simple glycan modifications. From a practical perspective, consecutive IMERs would likely to be more cost effective in implementation when only 2-3 transferases are involved.

As a counter argument, it is likely that an *in vivo* system which would continuously express and supply the transferases, substrates and the desired glycoprotein may be a more viable option, reducing the requirement for purification of the individual transferases required. This is contrasted with the *in vitro* systems where cost increases proportionately with the number of glycosyltransferases required. However, *in vivo* systems come with their own set of complexities. The choice of expression system matters greatly to produce a homogeneously glycosylated glycoprotein; mammalian systems are more expensive to maintain, yield less product, and contain multiple “unwanted” glycosyltransferases that requires knocking out, while non-mammalian systems such as yeast or insect require the knocking in of transferases, nucleotide sugar metabolic enzymes, and nucleotide sugar transporters, which compete with the synthesis of the desired glycoprotein for cellular resources. In these *in vivo* systems, “under-glycosylated” proteins can still be present and require subsequent *in vitro* glycosylation to achieve homogeneity. Furthermore, for every glycan structure desired, a separate strain of the cells would be required for the entire production process.

In the *in vitro* systems described, by selecting the right precursor glycoform of the glycoprotein, the number of glycosyltransferases required can be optimised. For example, by using an insect expression system with the  $\alpha$ -1,3-fucosyltransferase knocked out, to make the precursor glycoprotein, the resulting glycans would only be the *N*-glycan core structure of  $\text{Man}_3\text{GlcNAc}_2$ . To then achieve a bisecting, biantennary glycan with sialyl-Lewis<sup>X</sup> glycan would require six transferases to be synthesised and immobilised; GlcNAc transferase I, II, III,  $\beta$ -galactosyltransferase, fucosyltransferase (except FUT8) and sialyltransferase. By only knocking out the one enzyme in the insect cell, instead of the *in vivo* synthetic approach that needs to introduce/remove many enzymes, more cellular resources can be directed towards the over expression of the desired precursor glycoprotein. The glycosyltransferases required for immobilisation could then be expressed in other systems separately. Non-mammalian glycosyltransferases from bacteria or fungi, which can be produced more efficiently, or mutant/directed evolved enzymes that are faster or more stable, can be more readily adapted into the *in vitro* systems described here. Furthermore, different glycan structures can be built by simply switching the glycosyltransferases, which when produced individually, also allows optimisation of the overall reaction kinetics for each enzyme. Taken together as a whole, it is my belief that the development of the “Golgi-like” sequential *in vitro* glycosylation systems could become a cost-effective process while gaining valuable basic scientific data that would benefit glyco-related science.

With regards to the IgM glycan *in vitro* functionalisation work, cost effectiveness is also an important factor. For IgM, an underexplored antibody that has tremendous potential, recent technological developments in IgM production and purification have started to increase the research interest for applications in medicine and therapeutics. Although antibodies are bifunctional, binding the target antigen at the Fab region, and triggering immune related responses in the Fc region, it is the former that gives it its greatest therapeutic value. The ability to target

specific biological molecules is one of the most sought after uses in medicine as it reduces undesired implications from off target activities. Structure-based drug design can achieve this to a certain extent, but the pre-requisite of needing to have a known molecular structure of the target makes it difficult to become as powerful as an antibody. Exploiting the antibody's targeting ability while successfully using it as a biological tracker or carrier is every antibody related biotechnology company's most prized possession. By using the chemo-enzymatic method described in this work (Chapter 5) to functionalise IgM on the glycans, three major aims were achieved. Firstly, by performing the functionalisation in a biocompatible environment loss of IgM was avoided; loss of antibody during production is monetary loss for a biotech company. Secondly, functionalisation of IgM with a biologically "transparent" azide functional group enables the use of targeted ligation reactions with any alkyne as the click chemistry used is well known, efficient and reproducible. Most importantly, by functionalisation on the glycans, the Fab antigen binding region is undisturbed, reducing the probability of the antibody losing its targeting abilities. Similarly, this functionalisation method could be performed *in vivo* through metabolic engineering by feeding azide labelled ManNAc, but would result in a lot of waste as the azide labelled sialic acid would be present on all sialylated proteins produced by the cell and not just the IgM. Keeping cost effectiveness in mind, this method would increase the cost of production. The knowledge of the number of azide functional groups on the IgM is also important as an optimal amount of small molecule alkyne can be used, thus reducing wastage. In the same vein, for small molecule drugs, it also enables quantitative calculations of the directed chemical load, which in turn affects the dosage of the IgM required. The potential applications of this simply and specifically functionalised IgM are numerous and will provide researchers with more tools for IgM related biological studies.

Although it is evident from this work that a personal preference for *in vitro* systems of protein glycosylation is present, the point was never about developing a completely *in vitro*



glycosylation process, but rather at finding the least complicated pathway, such as has been evidenced by the IgM glycoengineering example.

## Appendix

### **Appendix**

Appendix I Biosafety approval form for NIP041214BHA

Appendix II Biosafety approval form for 5201600372

Appendix III Sequence information for Chapter 2

Appendix IV Paper III

Appendix V Paper IV

Appendix I & II of this thesis have been removed as they may contain sensitive/confidential content

**Appendix III – Sequence information**

TALE monomers

Nucleotide specificity	Protein sequence
A	LTPEQVVAIAS <b>N</b> IGGKQALETVQRLLPVLCQAHG
T	LTPEQVVAIAS <b>N</b> GGGKQALETVQRLLPVLCQAHG
C	LTPEQVVAIAS <b>H</b> DGGKQALETVQRLLPVLCQAHG
G	LTPEQVVAIAS <b>N</b> HGGKQALETVQRLLPVLCQAHG

mSA protein sequence

AEAGITGTWYNQSGSTFTVTAGADGNLTGQYENRAQGTGCQNSPYTLTGRYNGTKLEWRVE  
WNNSTENCHSRTEWRGQYQGGAEARINTQWNLTYEGGSGPATEQGQDTFTKVKPSAAS

Protein sequence used for reverse translation and codon optimisation

Scaffold QNE → target sequence CAGAATGAG

MGHHHHHHLTPEQVVAIAS**H**DGGKQALETVQRLLPVLCQAHGGLTPEQVVAIAS**N**IGGKQAL  
ETVQRLLPVLCQAHGGLTPEQVVAIAS**N**HGGKQALETVQRLLPVLCQAHGGLTPEQVVAIAS**N**  
GGKQALETVQRLLPVLCQAHGGLTPEQVVAIAS**N**IGGKQALETVQRLLPVLCQAHGGLTPEQVV  
AIAS**N**GGGKQALETVQRLLPVLCQAHGGLTPEQVVAIAS**N**HGGKQALETVQRLLPVLCQAHG  
LTPEQVVAIAS**N**IGGKQALETVQRLLPVLCQAHGGLTPEQVVAIAS**N**HGGKQALETVQRLLPV  
LCQAHG**G**GGGAEAGITGTWYNQSGSTFTVTAGADGNLTGQYENRAQGTGCQNSPYTLTGRY  
NGTKLEWRVEWNNSTENCHSRTEWRGQYQGGAEARINTQWNLTYEGGSGPATEQGQDTFT  
KVKPSAAS\*

Scaffold TWQ → target sequence ACTTGCAA

MGHHHHHHLTPEQVVAIAS**N**IGGKQALETVQRLLPVLCQAHGGLTPEQVVAIAS**H**DGGKQAL  
ETVQRLLPVLCQAHGGLTPEQVVAIAS**N**GGGKQALETVQRLLPVLCQAHGGLTPEQVVAIAS**N**  
GGKQALETVQRLLPVLCQAHGGLTPEQVVAIAS**N**HGGKQALETVQRLLPVLCQAHGGLTPEQV  
VAIAS**N**HGGKQALETVQRLLPVLCQAHGGLTPEQVVAIAS**H**DGGKQALETVQRLLPVLCQA  
GLTPEQVVAIAS**N**IGGKQALETVQRLLPVLCQAHGGLTPEQVVAIAS**N**IGGKQALETVQRLLPV  
LCQAHG**G**GGGAEAGITGTWYNQSGSTFTVTAGADGNLTGQYENRAQGTGCQNSPYTLTGRY  
NGTKLEWRVEWNNSTENCHSRTEWRGQYQGGAEARINTQWNLTYEGGSGPATEQGQDTFT  
KVKPSAAS\*

Scaffold TRE → target sequence ACGCGTGAA

MGHHHHHHLTPEQVVAIAS**N**IGGKQALETVQRLLPVLCQAHGGLTPEQVVAIAS**H**DGGKQAL  
ETVQRLLPVLCQAHGGLTPEQVVAIAS**N**HGGKQALETVQRLLPVLCQAHGGLTPEQVVAIAS**H**  
GGKQALETVQRLLPVLCQAHGGLTPEQVVAIAS**N**HGGKQALETVQRLLPVLCQAHGGLTPEQV  
VAIAS**N**GGGKQALETVQRLLPVLCQAHGGLTPEQVVAIAS**N**HGGKQALETVQRLLPVLCQA  
GLTPEQVVAIAS**N**IGGKQALETVQRLLPVLCQAHGGLTPEQVVAIAS**N**IGGKQALETVQRLLPV  
LCQAHG**G**GGGAEAGITGTWYNQSGSTFTVTAGADGNLTGQYENRAQGTGCQNSPYTLTGRY  
NGTKLEWRVEWNNSTENCHSRTEWRGQYQGGAEARINTQWNLTYEGGSGPATEQGQDTFT  
KVKPSAAS\*

Scaffold FQR → target sequence TTCCAGCGA

MGHHHHHHLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNGGGKQAL  
 ETVQRLLPVLCQAHGLTPEQVVAIAS**HD**GGKQALETVQRLLPVLCQAHGLTPEQVVAIAS**HD**  
 GGKQALETVQRLLPVLCQAHGLTPEQVVAIAS**N**IGGKQALETVQRLLPVLCQAHGLTPEQVVA  
 AIAS**NH**GGKQALETVQRLLPVLCQAHGLTPEQVVAIAS**HD**GGKQALETVQRLLPVLCQAHG  
 LTPEQVVAIAS**NH**GGKQALETVQRLLPVLCQAHGLTPEQVVAIAS**N**IGGKQALETVQRLLPV  
 LCQAHGGGGAEAGITGTWYNQSGSTFTVTAGADGNLTGQYENRAQGTGCQNSPYTLTGRY  
 NGTKLEWRVEWNNSTENCHSRTEWRGQYQGGAEARINTQWNLTYEGGSGPATEQQGQDTFT  
 KVKPSAAS\*

\* denotes stop codon

DNA program sequence

GCCATGCATAAT**CAGAA**T**GAGG**CCGGATCCAATA**CTTGGCA**AGCCGAATTCAAT**ACGCGTGAAG**  
 CCGCTAGCAAT**TTCCAGCGA**GCCCTGCAGAATGAGTTCCAAGTTGAGTACAGTGAAGCCTC

Location of adaptor protein binding sites are marked out in **bold**.

Primers used

Purpose	Primer	Sequence	Length	T <sub>m</sub> used for PCR [°C]
Cloning into pET-15b plasmid	Scaffold-F	GTAACCATGGGTCACCATCACCATCACC	28	52
	Scaffold-R	AAGGATCCTTAGGAGGCCGCACTC	24	
DNA program amplification	Scaffold-F	GCCATGCATAAT <b>CAGAA</b> T <b>GAGG</b>	19	52
	Scaffold-R	<b>GAGG</b> CTTCACTGTACTCAAC	20	

Pages 166-194 of this thesis have been removed as they contain published material. Please refer to the following citation for details of the articles contained in these pages.

Thaysen-Andersen, M., Chertova, E., Bergamaschi, C., Moh, E. S. X., Chertov, O., Roser, J., Sowder, R., Bear, J., Lifson, J., Packer, N. H., Felber, B. K., & Pavlakis, G. N. (2016). Recombinant human heterodimeric IL-15 complex displays extensive and reproducible N- and O-linked glycosylation. *Glycoconjugate Journal*, 33(3), 417-433.  
doi: 10.1007/s10719-015-9627-1

Lee, L. Y., Moh, E. S. X., Parker, B. L., Bern, M., Packer, N. H., & Thaysen-Andersen, M. (2016). Toward automated N-glycopeptide identification in glycoproteomics. *Journal of Proteome Research*, 15(10), 3904-3915.  
doi: 10.1021/acs.jproteome.6b00438