Name: Terence Teo Yung Ling

Degree: Master of Engineering (Chemical)

Dept: Chemical and Biomolecular Engineering

Thesis Title: N-glycosylation analysis and comparative modeling of mouse hybridoma IgM84 & 85

Abstract

The application of human embryonic stem cells (hESCs) in regenerative medicine has remained challenging in the last decade, mainly due to potential teratoma formation of undifferentiated hESCs upon administration *in vivo*. To remove undifferentiated hESCs from the differentiated ones, Bioprocessing Technology Institute (BTI) has generated a mouse hybridoma immunoglobulin M, IgM 84 that exhibits cytotoxic activity *via* oncosis towards undifferentiated hESCs that are not observed in other IgMs such as IgM 85. Previous findings have shown that IgM 84 and 85 bind to the same surface antigen on undifferentiated hESCs, i.e. podocalyxin-like protein-1. Using comparative modeling, we showed that the 3-dimensional (3D) models for the variable regions of IgM 84 and 85 are not significantly different in structure despite major differences within their complementarity determining regions (CDRs). On the other hand, using techniques such as matrix-assisted laser desorption/ionization mass spectrometry, high pH anionic exchange chromatography etc., we found that IgM 84 to be differently N-glycosylated i.e. improper trimming of high mannose type N-glycans in endoplasmic reticulum (ER), and less fucosylation and sialylation of complex type N-glycans in Golgi, as compared to those on IgM 85. We believe that these differences might suggest a differently folded IgM 84 that could shed more light on how multivalent IgM 84 exhibits its cytotoxicity activity.

Keywords: IgM, N-glycosylation, human embryonic stem cells, mouse hybridoma, mass spectrometry, comparative modeling.

N-GLYCOSYLATION ANALYSIS AND COMPARATIVE MODELING OF MOUSE HYBRIDOMA IgM84 & 85

TERENCE TEO YUNG LING

2011

N-GLYCOSYLATION ANALYSIS AND COMPARATIVE MODELING OF MOUSE HYBRIDOMA IgM84 & 85

TERENCE TEO YUNG LING

NATIONAL UNIVERSITY OF SINGAPORE

2012

N-GLYCOSYLATION ANALYSIS AND COMPARATIVE MODELING OF MOUSE HYBRIDOMA IgM84 & 85

TERENCE TEO YUNG LING

(B.Eng (Hons),NUS)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF ENGINEERING

DEPARTMENT OF CHEMICAL AND BIOMOLECULAR ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2012

TABLE OF CONTENTS

ACKNOWLEDGEMENT

All research work described in this thesis was carried out in the Bioprocessing Technology Institute (BTI), under the Agency for Science, Technology and Research (A*STAR). First and foremost, I would like to thank my supervisors Professor Reginald Tan[1] and Associate Professor Muriel Bardor[2] for their astute direction. In particular, I would like to thank Associate Professor Muriel Bardor for her relentless effort and invaluable guidance throughout the progress of my master studies. My deepest gratitude is reserved for Dr. Geoffrey Koh[2] in guiding me through the work related to comparative modeling. Not to forget Dr. Miranda Van Beers[2] for her continuous encouragement and guidance in various aspects of this thesis.

I would also like to acknowledge all my colleagues in Analytics who have rendered their help to me in various experiments throughout this work, and made my stay with the group an enjoyable one. In particular, I would like to highlight three specific individuals who have helped me tremendously in completing this work – Gavin Teo, who works on high pH anionic exchange chromatography (HPAEC); Eddy Tan who works on Size Exclusion Chromatography (SEC)-Static Light Scattering (SLS) on a High Performance Liquid Chromatography (HPLC) system; and Francois Le Mauff who have helped me in analyzing the spectra of Mass Spectrometry/Mass Spectrometry (MS/MS) for peptides and glycopeptides. Last but not least, I would also like to acknowledge the contribution from Downstream Processing group who has purified IgM 84 and 85, and Stem Cell group who has done the full length DNA sequencing for both IgM 84 and 85 that makes completion of this work possible.

[1] Department of Chemical and Biomolecular Engineering, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260
[2] Bioprocessing Technology Institute, 20 Biopolis Way, #06-01 Centros, Singapore 138668

SUMMARY

IgM 84 and 85[1] are immunoglobulins (Ig) M generated by the Stem Cell (SC) group at the Bioprocessing Technology Institute (BTI), that bind to the surface antigen, podocalyxin-like protein-1 (PODXL) on undifferentiated human embryonic stem cells (hESCs). Interestingly, only IgM 84 exhibits cytotoxic activity *via* oncosis towards these undifferentiated hESCs (Choo et al., 2008; Tan et al., 2009). Using antibody fragments[2], it has been shown that binding to antigen sites alone are not sufficient to initiate cytotoxic activity to the same level that was previously observed in pentameric IgM 84 thus suggesting the importance of its multivalency in oncosis (Lim et al., 2011). In this thesis, we were interested (i) to examine if N-glycosylation differences between IgM 84 and 85 could explain the cytotoxic behaviour in multivalent IgM 84; and (ii) to create 3D structural models for variable regions of IgM 84 and 85 and visualize the structural differences on their antigen binding sites.

Using our mouse N-glycan library[3], N-glycans structures were assigned to different mass ions of IgM 84 and 85. We categorized all the N-glycan structures into three main groups – high mannose, biantennary and triantennary complex types and we found several unique complex type N-glycan structures in the respective mass spectra of IgM 84 or 85. In high mannose type, the presence of $Man_9GlcNAc_2$ in IgM 84 but not in IgM 85, indicates the possibility of a differently folded IgM 84 exiting endoplasmic reticulum (ER) because N-glycosylation plays a vital role in ER protein folding mechanism. In addition, IgM 84 seems to be less fucosylated compared to IgM 85 due to the presence of various non-fucosylated complex type N-glycans in the mass spectrum of IgM 84 but not that of IgM 85. A different folded IgM 84 exiting ER may cause these structures to be shielded from fucosylation during

[1] The registered names for commercial use assigned to IgM 84 and 85 are mAb 84 and 85, respectively.
[2] Four antibody fragment formats are generated namely Fab 84, scFv 84, scFv 84-diabody and scFv 84-HTH
[3] Mouse N-glycan library was a consolidation of the most mouse N-glycan profiling from Consortium for Functional Glycomics (CFG) databases

late processing or maturation step of N-glycans in the *trans*-Golgi. Furthermore, sialylation which is another maturation step of N-glycans, was observed to be less in IgM 84. Besides, IgM 85 also possesses two trisialylated complex type N-glycans that was not observed in IgM 84.

IgM 84 and 85 were found to differ mostly in the primary sequences within their variable regions i.e. about 57.8% sequence similarity, especially in their complementarity determining regions (CDRs). Despite these differences, the 3D models for variable regions of IgM 84 and 85 showed only minimal differences between their antigen binding sites, substantiated by the low root mean square difference (RMSD) values i.e. 1.51 Å and 1.27 Å for variable heavy and light chains respectively, upon superimposition. Differences observed around loop or flexible regions between two β-sheets, are not enough to result in a significant structural difference between the antigen binding sites of IgM 84 and 85.

In conclusion, with the lack of evidence that the antigen binding sites of IgM 84 and 85 are different structurally, we propose that a potentially different protein conformation in IgM 84 due to differences in N-glycosylation maturation may help to explain the cytotoxic behaviour of multivalent IgM 84.

NOMENCLATURE

N.1 General Abbreviations and Nomenclature

ADCC	antibody-dependent cell-mediated cytotoxicity
ALG genes	asparagine linked glycosylation genes
BCR	B cell receptor
CDC	complement dependent cytotoxicity
CDP	cytidine diphosphate
CDR	complementarity determining regions
CFG	Consortium for Functional Glycomics
C_H1, C_H2 and C_H3	constant regions 1, 2 and 3 of heavy chains
CID	collision-induced dissociation
CL	constant regions of light chains
CNX	calnexin
CRT	calreticulin
CTP	cytidine triphosphate
CV	column volume
Dol-P-P-GlcNAc	dolichol-P-P-N-acetylglucosamine
ECL	enhanced chemiluminescence
EDEM	ER degradation-enhancing α-mannosidase I–like protein
EMEA	European Medicines Agency
EPO	erythropoietin
ER	endoplasmic reticulum
Erp	endoplasmic reticulum protein
ESI-MS	electrospray ionization-mass spectrometry
F_{ab}	fragment, antigen binding
FAB-MS	fast atom bombardment-mass spectrometry

F_c	fragment, constant
FDA	United States Food and Drug Administration
Fruc	fructose
Fuc	fucose
F_v	fragment, variable
Gal	galactose
GalNAc	*N*-Acetylgalactosamine
GBPs	glycan binding proteins
GDP	guanosine diphosphate
Glc	glucose
GlcNAc	*N*-Acetylglucosamine
GlcNAcT-I	α-1,3-mannosyl-glycoprotein 2-β-*N*-acetylglucosaminyltransferase
GlcNAcT-II	α-1,6-mannosyl- glycoprotein 2-β-*N*-acetylglucosaminyltransferase
GlcNAcT-III	α-1,4-mannosyl-glycoprotein 4-β-*N*-acetylglucosaminyltransferase
GlcNAcT-IV	α-1,3-mannosyl-glycoprotein 4-β-*N*-acetyl-glucosaminyltransferase
GlcNAcT-V	α-1,6-mannosyl-glycoprotein 6-β-*N*-acetylglucosaminyltransferase
GPI anchor	glycosylphosphatidylinositol anchor
GTP	guanosine triphosphate
hESCs	human embroynic stem cells
HPAEC-PAD	high pH anionic exchange chromatography-pulsed amperometric detection
HPLC	high performance liquid chromatography

HTD	hot trypsin digestion
ICH	International Conference on Harmonization
IgA, IgD, IgE, IgG & IgM	immunoglobulin A, D, E, G and M
LacNAc	*N*-acetyllactosamine
LC-MS	liquid chromatography-mass spectrometry
MALDI-MS	matrix-assisted laser desorption/ionization-mass spetrometry
MALDI-TOF	matrix assisted laser desorption/ionization-time of flight
MALDI-TOF-TOF	matrix assisted laser desorption/ionization-time of flight-time of flight
Man	mannose
MEKC	micellar electrokinetic chromatography
mRNA	messenger ribonucleic acid
MS	mass spectrometry
MWCO	molecular weight cutoff
N/A	not applicable
Neu5Ac	*N*-Glycolylneuraminic acid
Neu5Gc	*N*-Acetylneuraminic acid
NMR	nuclear magnetic resonance
OST	oligosaccharyltransferase
PEG	polyethylene glycol
PI	phosphatidylinositol
PODXL	podocalyxin-like protein 1
RMSD	root mean square difference
scFv 84-HTH	single chain variable fragment 84-helix turn helix
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SLS	static light scattering

UDP	uridine diphosphate
UMP	uridine monophosphate
VH	variable regions of heavy chains
VL	variable regions of light chains

N.2 Abbreviations for bioinformatics

PDB	Protein Data Bank
PDB_nr95	Protein Data Bank_non redundance 95%
SCOP	Structural Classification of Proteins
DALI	Distrance matrix alignment
CATH	Class, Architecture, Topology and Homologous
BLAST	Basic Local Alignment Search Tool
PSI-BLAST	Position Specific Iteractive-Basic Local Alignment Search Tool
BLOSUM	BLOcks of Amino Acid SUbstitution Matrix
PAM	Point Accepted Mutation
MD	Molecular Dynamics
PDF	Probability Density Function
DOPE	Discrete Optimized Protein Energy

N.3 List of chemicals

ACN	acetonitrile
α-cyano	α-cyano-4-hydroxycinnamic acid
CaCl2	calcium chloride
ChCl3	chloroform

DHB	2,5-dihydroxy benzoic acid
DMSO	dimethyl sulfoxide
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
GdnHCl	guanidine hydrochloride
HCl	hydrochloride acid
IAA	iodoacetamide
NaCl	Sodium chloride
NaN3	Sodium azide
NaOH	Sodium Hydroxide
NH_4HCO_3	ammonium bicarbonate
PNGase F, A	peptide-N-glycosidase F, A
PVDF	polyvinylidene fluoride
TBS	tris-buffered Saline
TFA	trifluroacetic acid

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Background

1.1.1 Human embryonic stem cells

Human embryonic stem cells (hESCs) are pluripotent stem cells that are derived from the inner cell mass of the blastocyst during the early-stage of an embryo (Thomson et al., 1998). A pluripotent cell is one that is able to differentiate into all derivatives of three primary germ layers - ectoderm, endoderm, and mesoderm, which include more than 220 cell types in the adult body (Thomson et al., 1998). Under defined conditions, hESCs are capable of propagating indefinitely, which makes them a useful tool for regenerative medicine[1] in research and applications include some of the most common neural diseases such as Parkinson's disease, stroke and multiple sclerosis (Lindvall and Kokaia, 2006).

However, one major issue with using hESCs in regenerative medicine is its potential to form teratomas from remnants of undifferentiated hESC upon administration (Knoepfler, 2009). Such safety issue poses a major roadblock to using hESCs as therapeutics. With regards to cell-cell separation methods, there have been major efforts done in the last decade including work by Schriebl and co-workers from our institute, who used stage-specific embryonic antigen 1 (SSEA-1) on undifferentiated mESCs as selection marker to remove them from the pool of differentiated mESCs using highly selective magnetic activated cell sorting method (Schriebl et al., 2010). The work highlighted the limitation of engineering approach to achieve stringent therapeutic requirement[2] of using hESCs that could possibly be done otherwise using specific binding antibodies, better still if these antibodies exert cytotoxicity against them (Schriebl et al., 2012)

[1] Regenerative medicine is a processing of replacing lost or restoring damaged cells, tissues or organs back to normal functions

[2] A log clearance rate of 10 is required to reach a safety margin of 10^{-1} undifferentiated hESCs or a purity of 99.99999999% assuming a single therapeutic dose contains about 10^{7}-10^{9} cells

1.1.2 Discovery of monoclonal antibodies against undifferentiated hESC

At the Bioprocessing Technology Institute (BTI), the Stem Cell group has generated a panel of 10 monoclonal antibodies (mAbs) against surface antigens on undifferentiated hESCs of HES-3[1] cell lines, following immunization of Balb/C mice using the entire HES-3 cells (Choo et al., 2008). Two of these mAbs, licensed as mAb 84 and 85, were found to bind to the same surface antigen on undifferentiated hESCs, i.e. podocalyxin-like protein-1 (POXDL). In this thesis, mAb 84 and mAb 85 will be referred to as IgM 84 and IgM 85, respectively, to emphasize that both antibodies are of the immunoglobulin M isotype.

Interestingly, IgM 84 not only binds but also exhibits cytotoxicity against undifferentiated hESC. The reported cell death caused by IgM 84 is termed oncosis, which is different from apoptosis, triggered by antibody-dependant cell mediated cytotoxicity (ADCC). Oncosis, as described by Tan and co-workers (Tan et al., 2009), is a form of cell death that is preceded by cell aggregation and damage to cell membranes of the undifferentiated hESCs, causing leakage of intracellular Na^+ ions. The proof of such cell killing mechanism is revealed under scanning electron microscope by pore formation in the cell membrane of undifferentiated hESCs due to the clustering of PODXL-1 antigens (Tan et al., 2009).

Lim and co-workers engineered antibody fragments from IgM 84 and showed that only scFv[2] 84-HTH, a fragment that is bivalent and highly flexible, could recapitulate the cytotoxic effect of IgM 84[3], while other fragments, scFv 84, scFv 84 diabody, and Fab 84, that are monovalent or bivalent and more rigid only bound to PODXL (Lim et al., 2011). Moreover, 20 times more of scFv 84-HTH in quantity was required to achieve the same level of cytotoxicity as IgM 84 (Lim et al. 2011). These findings highlights the importance of the *unique structure* of IgM 84 that allows cross-linking of multiple PODXL-1 antigens on the cell surface thus triggering efficiently cell death in hESCs via oncosis.

[1] HES-3: Human embryonic stem cell lines obtained from ES Cell International (ECI, Singapore, http://www.escellinternational.com)
[2] scFv stands for single-chain variable fragment, a fusion protein of variable regions of heavy (V_H) and light chains (V_L) of immunoglobulins
[3] In the article by Lim and co-workers, IgM 84 is referred to as mAb 84 instead, which is the licensed name for this molecule.

N-glycosylation is a biosynthetic process of adding glycans or sugar moieties to the protein backbone of proteins such as immunoglobulins via asparagine linked N-glycosidic linkages. The roles of N-glycosylation in biological activities of immunoglobulins G and M as effectors functions and complement have previously been reported (Wright et al., 1990; Wormald et al., 1991; Mimura et al., 2000; Anthony et al., 2008). Hence, we would like to explore if N-glycosylation of IgM 84 results in a different protein conformation that causes IgM 84 to be cytotoxic against undifferentiated hESCs.

1.2 Thesis Scope

The aim of this thesis is two-fold: a) to study the N-glycosylation of IgM 84 and 85 and to examine if any of the differences in N-glycan types between IgM 84 and 85 could explain the cytotoxic effect of IgM 84, as described in Section 1.2.1; b) to model the variable regions of IgM 84 and 85 and to examine specifically if there is any *structural* difference between the antigen binding sites of IgM 84 and 85, as described in Section 1.2.2.

1.2.1 Comparative N-glycosylation analysis of IgM 84 and 85

IgM 84 and 85 have been previously generated using hybridoma technology, subsequently adapted step-wise and cultured in protein-free, chemically defined media in 5L continuous stirred tank bioreactor (Lee et al., 2009). Cultures in bioreactors were harvested and clarified by centrifugation and depth filtration, before they were purified in two steps – PEG precipitation and Anion-Exchange Chromatography (Tscheliessnig et al., 2009).

Starting from the purified IgM 84 and 85, the N-glycans were released, fractionated and analysed using MALDI-TOF MS. Meanwhile, we built a mouse N-glycan library, from the CFG database to match and assign relevant N-glycan structures to different mass peaks. We performed comparative analysis of the global N-glycan profiling and degree of sialylation of IgM 84 and 85. We also did a preliminary study on the site-specific N-glycan profiling of IgM 84 using a glycopeptide approach.

1.2.2 Visualization of variable binding regions of IgM 84 and 85

We developed and superimposed the 3D structural models for variable regions of IgM 84 and 85 i.e. variable heavy and light chains separately, to visually inspect for any structural differences within the antigen binding sites. Upon superimposition, we also calculated the root mean square difference (RMSD) to quantify the spatial structural differences.

1.2.3 Thesis Organization

Chapter 2 starts with an introduction on immunoglobulin (Ig) in general, and IgM in particular. The chapter then follows with an overview of the types of N-glycan present in mouse hybridoma IgM and the biosynthetic pathway of N-glycans including different terminal structures of N-glycans that are commonly found in mammals. The last part of this chapter will touch on the therapeutic role of Ig and how N-glycosylation plays an important role in this aspect. In Chapter 3, we will discuss our approaches to study the N-glycosylation of IgM 84 and 85 with regards to their macro- and microheterogeneity, the overall percentages of sialylation and distribution, the presence of glyco-epitopes in IgM 84 and 85, and the process to construct 3D structural models for variable regions of IgM 84 and 85 using Discovery Studio software. Chapter 4 then presents the results of comparative analysis of IgM 84 and 85 in terms of global N-glycan profiling and sialylation analysis, and the possible implications will also be discussed. In addition, the constructed 3D structural models are superimposed to visually inspect if there are any differences between IgM 84 and 85 on their antigen binding sites. The concluding chapter, Chapter 5, provides a summary of the main conclusions, and recommendations for future works.

Appendix A shows information regarding the amino acid sequence of heavy and light chains of IgM 84 and 85, the sequence alignment results of the corresponding constant and variable regions, and the respective potential N-glycosylation sites on each chain. Appendix B lists down the resources obtained from Consortium for Functional Glycomics (CFG) to construct our in-house mouse N-glycan library. Appendix C shows a list selected of peptide sequences of trypsin-digested heavy and light chains of IgM 84 and 85 for the analysis of

glycopeptides for site-specific N-glycan profiling studies. Appendix D shows the masses, structures, percentages of relative abundance and distribution of all N-glycan structures observed in IgM 84 and 85.

2 LITERATURE REVIEW

2.1 Immunoglobulins (Ig)

Ig, also known as antibody[1], is based on a single large Y-shaped protein (Figure 2.1), produced by our immune system to identify and neutralize foreign organisms like bacteria and viruses. Such identification is performed through recognition of a unique part on the foreign objects that is called antigen (Janeway, 2001). Antigen-binding site of an antibody is termed paratope, whereas the site on an antigen where the antibody binds is called epitope.

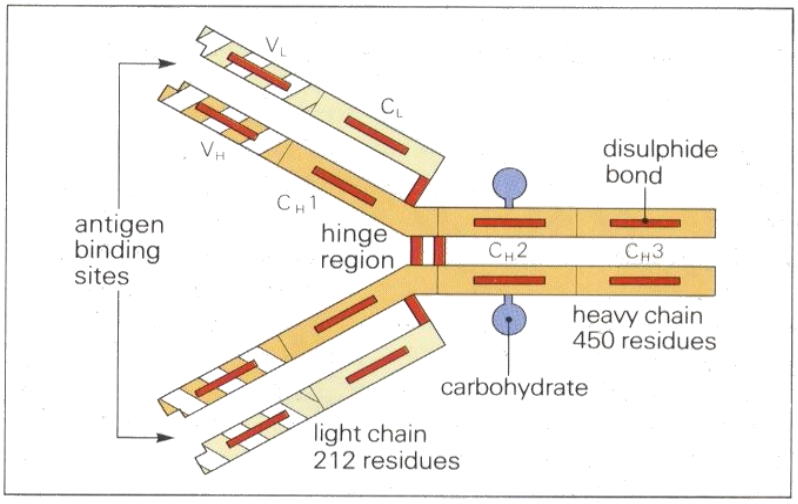

Figure 2.1 Immunoglobulin (Ig) G consists of two heavy chains (V_H, C_H1, C_H2 and C_H3) and two light chains (V_L and C_L) connected by disulphide bonds (red). It has one site for carbohydrate (blue) attachment on each heavy chain

Immunoglobulins (Ig) can be broadly classified into 5 isotypes or classes – IgA, IgD, IgE, IgG and IgM. The prefix Ig stands for immunoglobulin; whereas the capital letter i.e. A, D, E, G and M indicates the type of heavy chain each isotype possesses, as denoted in similar Greek letters α, δ, ε, γ and μ respectively. In mammals, there are two types of light chains across all Ig isotypes i.e. λ and κ. One Ig monomer consists of four polypeptide chains; two heavy chains (H) and two light chains (L) connected by disulfide bridges. Each heavy and light

[1] Antibody can be either monoclonal or polyclonal, which describes its ability to recognize and bind only one, or multiple epitope(s) of an antigen

chain has two regions, the constant region[1] (C) and the variable region[2] (V). The constant region is largely similar for Ig of the same isotype coming from the same source.

In one Ig monomer, there are F_{ab}, F_v and F_c parts that describe the non-covalent association between different domains of heavy and light chains. F_{ab} is the region where domains V_L, C_L, V_H and C_H1 associate; F_c is the region where domains C_H2 and C_H3 from each heavy chain associate; and F_v is the region of V_L and V_H and it is most important region of an antibody for binding to antigens. Near the tip of F_v lie the CDRs which stand for complementarity determining regions. More specifically, they are regions of variable loops of β-strands, three[3] on each of the variable light (V_L) and heavy (V_H) chains that are responsible for epitope recognition a specific antigen.

2.1.1 Immunoglobulin (Ig) M

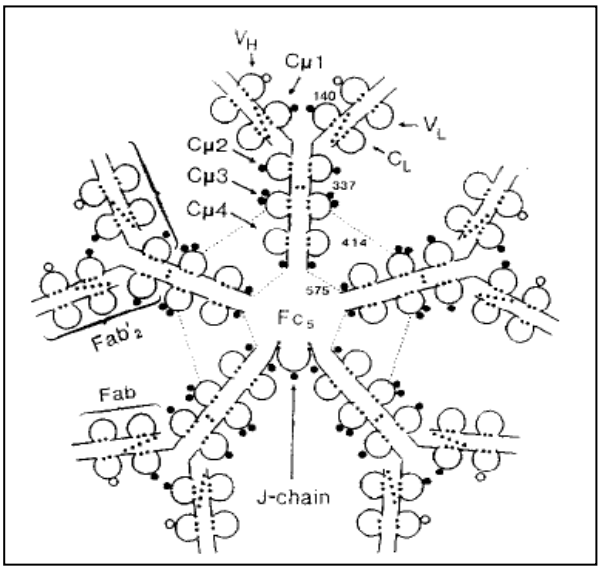

Figure 2.2 Structure of a pentameric mouse IgM (Perkins et al., 1991)

[1] Constant region of heavy chain is made up of three domains i.e. C_H1 C_H2 and $C_H3$for heavy chain; whereas for light chain, it only has one i.e. C_L.
[2] Variable region of heavy and light chains are V_H and V_L respectively
[3] CDR1, CDR2 and CDR3

Antibodies are produced by white blood cells in either soluble form - secreted out of the cell, or membrane-bound - attached to the surface of a B cell or B cell receptor (BCR). These BCR facilitate the activation and subsequent differentiation of B cells into antibody-producing plasma cells or memory B cells that will survive and remain dormant but able to recognize the same antigen faster in future immune response.

Immunoglobulin M or IgM is the first antibody isotype produced in B cells in response to initial immune response to antigen. In our case, IgM 84 and 85 are produced in our mouse hybridoma clones. IgM is the largest immunoglobulin (Ig) among all other isotypes. IgM that is secreted by B cells can exist predominately as pentamer, but also hexamer. A pentameric IgM has a protein size of approximately 900kD and is made up of 5 Ig monomers that are connected by disulphide bridges (Figure 2.2). Besides, a pentameric IgM also has a J-chain that is absent in hexameric IgM. One distinct physical characteristic of an IgM from all other isotypes is the presence of a vast number of N-glycosylation sites. In mouse IgM, there can be between 5 to 6 N-glycosylation sites on the heavy chain, 0 to 1 on the light chain, and 0 to 1 on the J-chain.

2.2 N-glycosylation of Immunoglobulins (Ig)

2.2.1 Carbohydrates and Glycoproteins

Carbohydrates[1] are one major class of molecules that make up a cell, tissue, organ, physiological system, and eventually an intact organism, besides proteins, nucleic acids and lipids. Like these other molecules, carbohydrates also encompass a crucial role in biological activities as intermediates in generating energy and as signalling effectors, recognition markers, and structural components (Varki and Sharon, 2009). Carbohydrates are polymers of

[1] Also commonly known as sugars, oligosaccharides or glycans when they are attached to a protein molecule, or glycoprotein

monosaccharides (Figure 2.3) that are joined together via glycosidic linkages. Therefore, they are sometimes referred to as oligosaccharides.

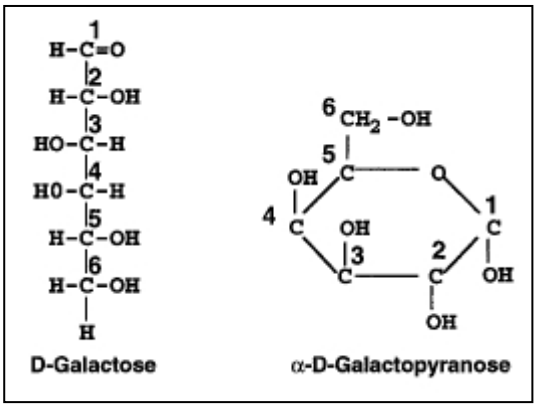

Figure 2.3 Open chain (left) and ring form (right) of D-galactose

In nature, several hundred distinct monosaccharides are known to occur; in mammals, there are only six monosaccharide types that are categorized as follows:

- *Pentoses*: Five-carbon neutral sugars;
- *Hexoses*: Six-carbon neutral sugars;
- *Hexosamines*: Hexoses with an amino group at the 2-position, which can be either free or, more commonly, N-acetylated;
- *Deoxyhexoses*: Six-carbon neutral sugars without the hydroxyl group at the 6-position;
- *Uronic acids*: Hexoses with a negatively charged carboxylate at the 6-position;
- *Sialic acids*: Family of nine-carbon acidic sugars.

Glycoproteins are proteins which contain oligosaccharide chains, or glycans covalently attached to the protein backbone. The glycan is synthesized and attached to the protein either through co- or post-translational modification, of which a process that is known as glycosylation. Most glycans can be attached to side chains of proteins via three types of linkage: glycosylphosphatidylinositol (GPI) anchored, O-linked and N-linked.

Table 2.1 Monosaccharides commonly found in mammalian glycoproteins

No	Monosaccharide	Type	Abbreviation	Symbol
1	D-Glucose	Hexose	Glc	
2	D-Galactose	Hexose	Gal	
3	D-Mannose	Hexose	Man	
4	L-Fucose	Deoxyhexose	Fuc	
5	*N*-Acetylgalactosamine	Hexosamine	GalNAc	
6	*N*-Acetylglucosamine	Hexosamine	GlcNAc	
7	*N*-Acetylneuraminic acid	Sialic Acid	Neu5Ac	
8	*N*-Glycolylneuraminic acid	Sialic Acid	Neu5Gc	

2.2.1.1 Glycosylphosphatidylinositol (GPI) anchor

A GPI anchor is a glycolipid that is attached to the C-terminus of a protein and the lipid bilayer of cell membrane via two phosphodiester linkages of phophoethanolamine and phosphatidylinositol (PI), respectively (Figure 2.4). Such structure constitutes the only anchor to the lipid bilayer of cell membrane and it is important for the function of membrane bound protein in the extracellular space. Defects in GPI anchor is linked to various rare diseases such as paroxysmal nocturnal hemoglobinuria and hyperphosphatasia with mental retardation syndrome.

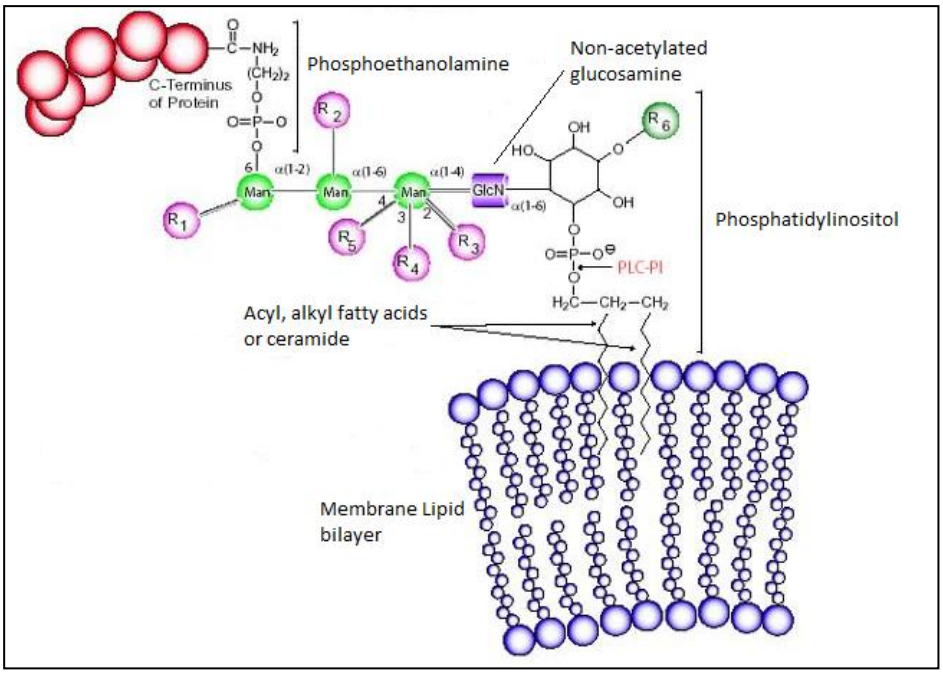

Figure 2.4 GPI anchor connects C-terminus of protein to membrane lipid bilayer via two phosphodiester linkages of phosphoethanolamine and phosphatidylinositol, respectively. R_1=Manα(1-2); R_2,R_3=Phosphoethanolamine; R_4=Gal$_4$; R_5=GalNAcβ(1-4); R_6=Fatty Acid at C_2 or C_3 of inositol (Adapted from GPI Anchor Structure found in www.sigmaaldrich.com)

2.2.1.2 O-linked glycan or O-glycan

An O-linked glycan is an oligosaccharide structure covalently α-linked to a glycoprotein via *N*-acetylgalactosamine (GalNAc). Typically, O-glycan is attached to the hydroxy oxygen of a serine (Ser or S) or threonine (Thr or T) residue of glycoprotein by an O-glycosidic bond that can be extended into a variety of different structural core classes. O-glycans, also called O-GalNAc glycans, are often found in mucins, glycoproteins with high content of serine, theorine, and proline residues. O-glycans of mucins are essential for their ability to hydrate and protect the underlying epithelium by trapping bacteria via specific receptor sites within O-glycans. In addition, these hydrophilic and negatively charged O-glycans also promote binding of water and salts that cause mucus to be viscous, forming a physical barrier between lumen and epithelium.

2.2.1.3 N-linked glycan or N-glycan

A N-glycan is an oligosaccharide structure that is covalently linked to an asparagine (Asn or N) residue of a protein. Such linkage commonly involves a GlcNAc sugar unit of the oligosaccharide and it is mostly found within the consensus peptide sequence of Asn-X-Ser/Thr[1]. Recent reports also suggest N-glycans to be found on Asn-X-Cys sequon in mammals, yeast and plants (Sato et al., 2000; Gil et al., 2009; Matsui et al., 2011). N-Glycans share a common pentasaccharide core ($Man_3GlcNAc_2$) that can be further extended into three main general classes: high-mannose (oligomannose) type, complex type, and hybrid type (Figure 2.5). In reality, a much diverse pool of oligosaccharide structures is found under each N-glycan type than those presented in Figure 2.5. From the perspective of a single protein molecule, the variable site occupancy or variability in location and number of glycosyl attachment sites is called macroheterogeneity; and variability in oligosaccharide structure at specific glycosylation sites is called microheterogeneity. Furthermore, higher number of potential N-glycosylation sites can add into the complexity and heterogeneity of the glycoprotein.

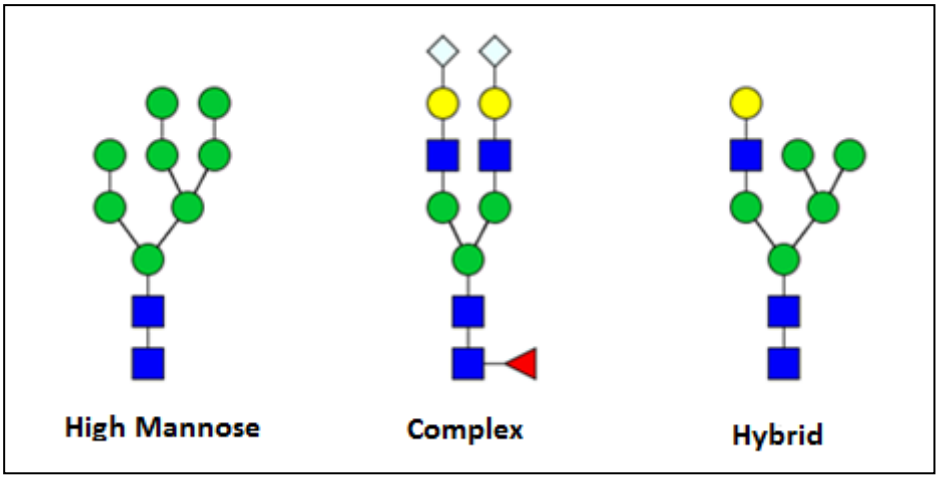

Figure 2.5 High mannose, complex and hybrid types are three typical N-glycan types found in mammals. Each structure here is just the representation that each N-glycan type could have.

[1] X can be any amino acid except proline (Pro) or aspartic acid (Asp)

As previously mentioned, immunoglobulin M (IgM) is highly N-glycosylated protein molecule bearing potential N-glycosylation sites of 5 to 6 on one heavy chain. One pentameric IgM molecule could have between 50 and 60 potential N-glycosylation sites and in certain cases, light chains of IgM were reported to sometimes bear 1 potential N-glycosylation site as well (Perkins et al., 1991). Due to this massive structure, only a handful of literature has successfully demonstrated using chemical cleavage method and nuclear magnetic resonance (NMR) (Chapman and Kornfeld, 1979; Chapman and Kornfeld, 1979; Brenckle and Kornfeld, 1980; Anderson et al., 1985; Monica et al., 1995). However, results that have been shown for mouse IgM in these reports still lacked the comprehensiveness of a full glycan profile that one might desire. While human serum IgM glycosylation has been recently characterized (Arnold et al., 2005), a full N-glycan profiling for mouse IgM has not yet been completely reported.

Because of the presence of such much N-glycans in the IgM and before we proceed to characterize N-glycosylation of our IgM 84 and 85, it would be worthwhile to spend some time to discuss the biosynthesis of N-glycans and the roles of N-glycosylation in therapeutic proteins.

2.3 Biosynthesis of N-Glycans

N-glycans are covalently attached to proteins at asparagine (Asn) residues of glycoprotein backbone by an N-glycosidic bond. There are five different N-glycan linkages[1] that have been reported, of which *N*-acetylglucosamine linkage to asparagines (GlcNAcβ1-Asn) is the most common (Stanley et al., 2009).

2.3.1 Synthesis of Dolichol-P-P-oligosaccharide[2] precursor

The biosynthesis of eukaryotic N-glycans begins with the synthesis of dolichol pyrophosphate *N*-acetylglucosamine (Dol-P-P-GlcNAc) on the cytoplasmic face of membrane of Endoplasmic Reticulum (ER), where GlcNAc-P is first transferred from UDP-GlcNAc to lipid-like precursor dolichol phosphate (Dol-P) (Figure 2.6).

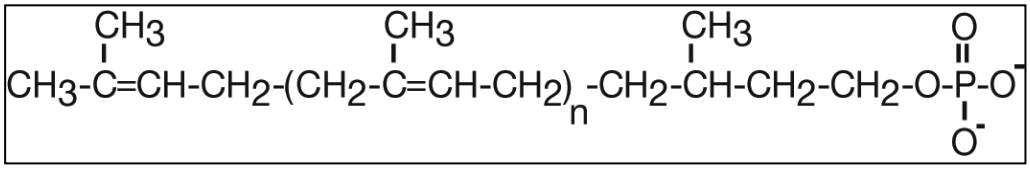

Figure 2.6 Dolichol phosphate (Dol-P) (Adapted from Essentials of Glycobiology, 2nd edition)

Figure 2.6 shows the overall process of how subsequent thirteen sugars[2] are sequentially added by a series of enzymatic reactions to Dol-P-P-GlcNAc to form $Glc_3Man_9GlcNAc_2$-P-P-Dol prior to its *en-bloc* transfer to Asn-X-Ser/Thr sequon of a nascent protein by oligosaccharyltransferase (OST). It is worth noting that the entire reactions do not just take place on the cytoplasmic face of ER. When Dol-P-P-GlcNAc is extended to $Man_5GlcNAc_2$-P-P-Dol, it is being "flipped" across the ER membrane and therefore the rest of the reactions take place inside the ER lumen, including the *en-bloc* transfer. Enzymes that are

[1] Other linkages are gluose, *N*-acetylgalactosamine (GalNAc), rhamnose, and linkage to argine: glucose
[2] Glycan – $Glc_3Man_9GlcNAc_2$ is made up of two sugar units of N-acetylglucosamine (GlcNAc), nine sugar units of mannose (Man) and three sugar units of glucose (Glc).

involve in adding the sugar units are encoded by *ALG*[1] while the sugar units that are being added are transferred directly from UDP-GlcNAc and GDP-Man on the cytoplasmic face; and indirectly via Dol-P-Man and Dol-P-Glc inside the ER lumen (Figure 2.7). Meanwhile, the nascent protein is synthesized in the ribosome and translocated into the ER lumen co-translationally.

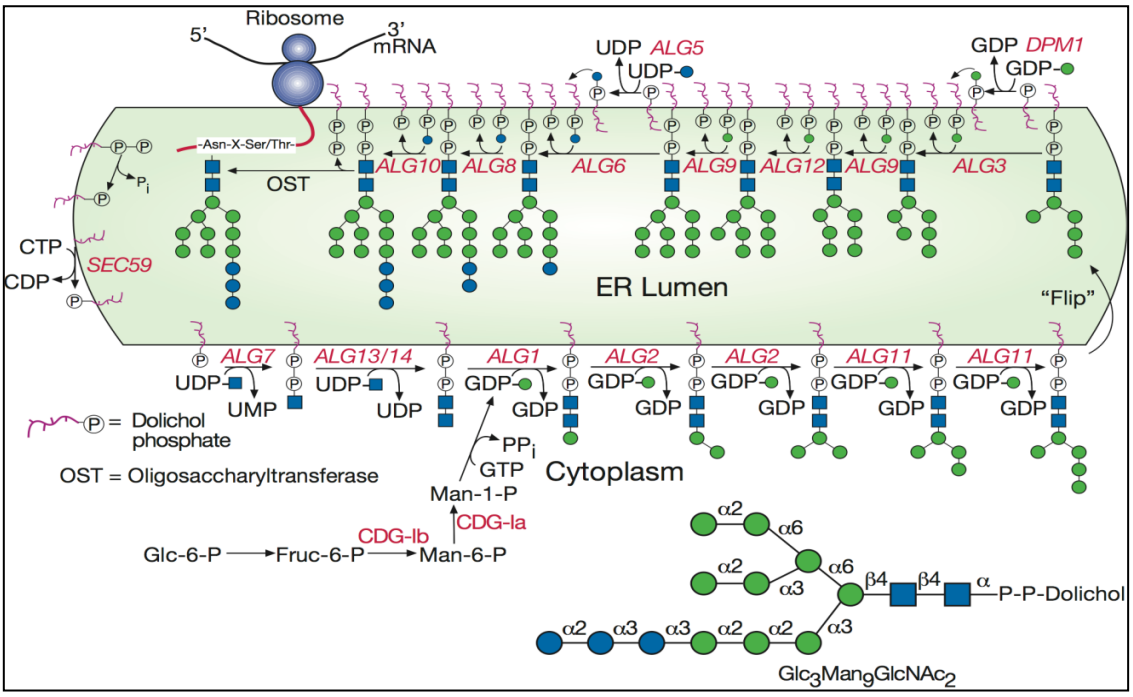

Figure 2.7 Synthesis of $Glc_3Man_9GlcNAc_2$-P-P-Dol starts on the cytoplasmic face of ER where Dol-P-P-GlcNAc is extended to $Man_5GlcNAc_2$-P-P-Dol before it is being "flipped" onto the luminal face of ER. After that, more glucose and mannose sugars are added to form the full 14-sugar N-glycan precursor and attach to a nascent protein. (Adapted from Essentials of Glycobiology, 2nd edition)

[1] *ALG* genes stand for asparagine linked glycosylation genes, identified primarily from the studies of yeast *Saccharomyces cerevisiae*.

2.3.2 Biosynthesis of N-glycan types

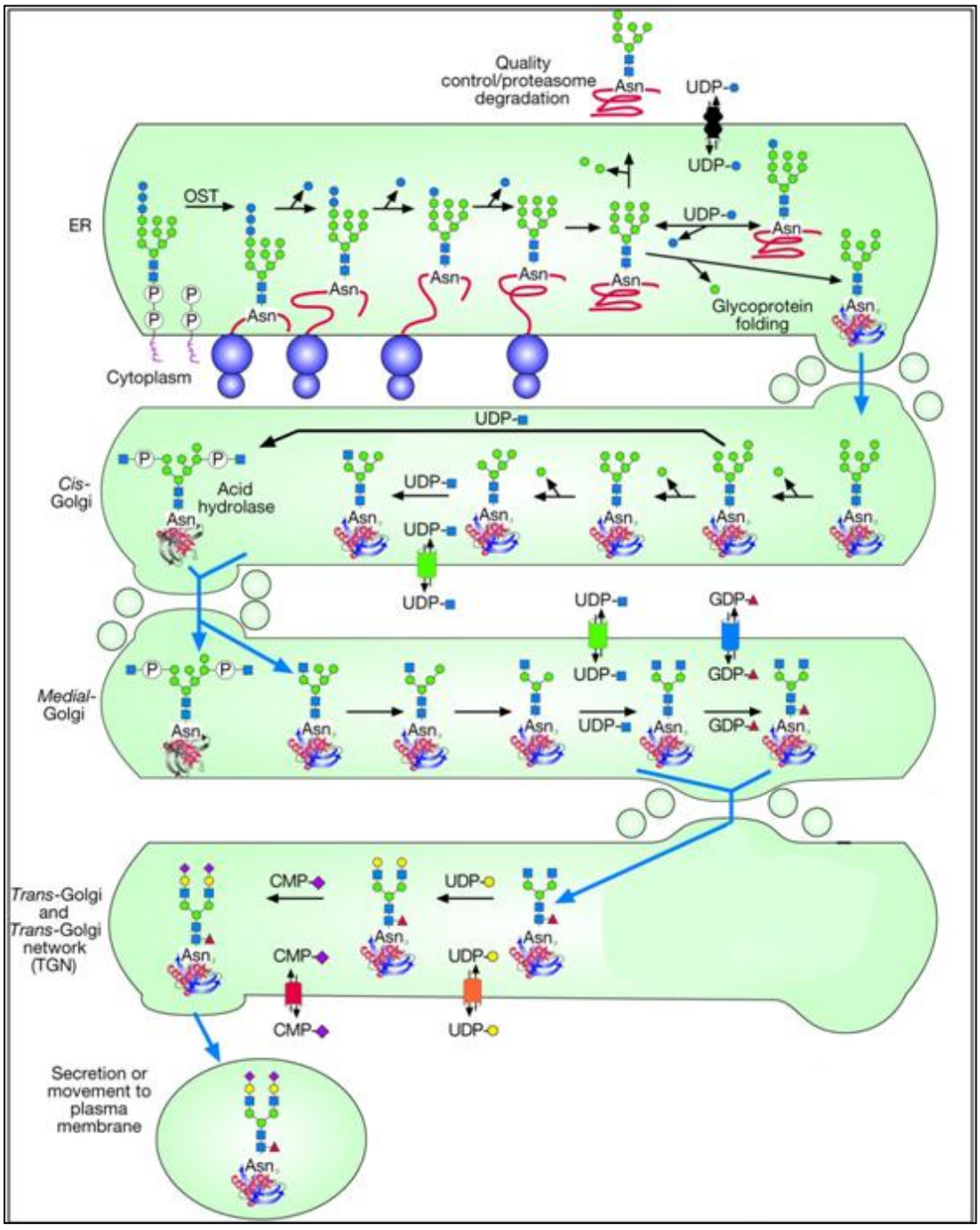

Figure 2.8 Once transferred, the oligosaccharide precursor $Glc_3Man_9GlcNAc_2$ is being trimmed sequentially to $Man_8GlcNAc_2$ in ER lumen prior to the export of the glycoprotein to the Golgi apparatus. In the *cis*-Golgi, trimming process continues until $Man_5GlcNAc_2$ the basic structure for synthesizing hybrid and complex N-glycan types, is formed. If the second trimming process is escaped, high mannose N-glycan types will be present on the secreted mature glycoprotein. In *medial*-Golgi, GlcNAc sugar units are added to the core and two Man sugar units are removed prior to maturation steps like galactosylation, fucosylation and sialylation. of N-glycans in the late *medial*- and *trans*-Golgi (Adapted from Essential of Glycobiology, 2nd edition).

In a nutshell, following the *en-bloc* transfer, the $Glc_3Man_9GlcNAc_2$ N-glycan precursor is initially trimmed to the $Man_5GlcNAc_2$ – basic structure for synthesizing complex and hybrid N-glycans, via a series of enzymatic reactions catalyzed by membrane-bound glycosidases in ER and *cis*-Golgi followed by the subsequent addition of other sugar units by glycosyltransferases in *cis*-, *medial*- and *trans*-Golgi (Figure 2.8). The expression of these trimming glycosidases has been quite conserved across eukaryotes and is known to interact with ER chaperones that recognize specific features of the trimmed N-glycan, that result in different protein folding in the ER. Details on this process will be discussed in following Section 2.3.4.

In the first stage of the trimming process, three glucoses are removed from $Glc_3Man_9GlcNAc_2$ sequentially by α-glucosidases I and II, which act specifically to remove one α1-2Glc and two α1-3Glc residues, respectively. Majority of glycoproteins exit ER en route to the *cis*-Golgi, carrying $Man_{8-9}GlcNAc_2$ depending if they have been acted on by ER α-mannosidase I which specifically cleaves off terminal α1-2Man. A second α-mannosidase I–like protein, also called **EDEM** (**E**R **d**egradation-**e**nhancing α-**m**annosidase I–like protein), is important in the recognition of misfolded glycoproteins, thereby targeting them for ER degradation (Freeze et al., 2009).

Further trimming of α1-2Man residues continues with the action of α1–2 mannosidase I in the *cis*-Golgi to give $Man_5GlcNAc_2$ (Figure 2.8). However, part of the $Man_{8-9}GlcNAc_2$ may escape modifications by the mannosidase I that results in a range of high-mannose type N-glycans i.e. $Man_{5-9}GlcNAc_2$ on the mature secreted glycoproteins. Biosynthesis of hybrid and complex type N-glycans begins in the core $Man_5GlcNAc_2$ with the addition of GlcNAc residue to C-2 of the mannose α1-3, initiated by a *N*-acetylglucosaminyltransferase I, also called GlcNAcT-I, to form $GlcNAcMan_5GlcNAc_2$. Following this step, two mannose residues i.e. α1-3Man and α1-6Man can then be removed by α-mannosidase II, inside *medial*-Golgi to form $GlcNAcMan_3GlcNAc_2$ (Figure2.8). Afterwards, a second GlcNAc is added to C-2 of the mannose α1–6 in the core by the action of GlcNAcT-II to yield the precursor for all complex

type N-glycans. However, if the two mannose residues are not removed, no further modification could occur in that mannose α1–6 branch leading to the formation of hybrid type N-glycans instead. These hybrid type N-glycans may occasionally carry "bisecting" GlcNAc (as indicated by red arrow in Figure 2.9). Further modification in the other branch by adding different terminal structures is still possible and will be discussed in Section 2.3.3.

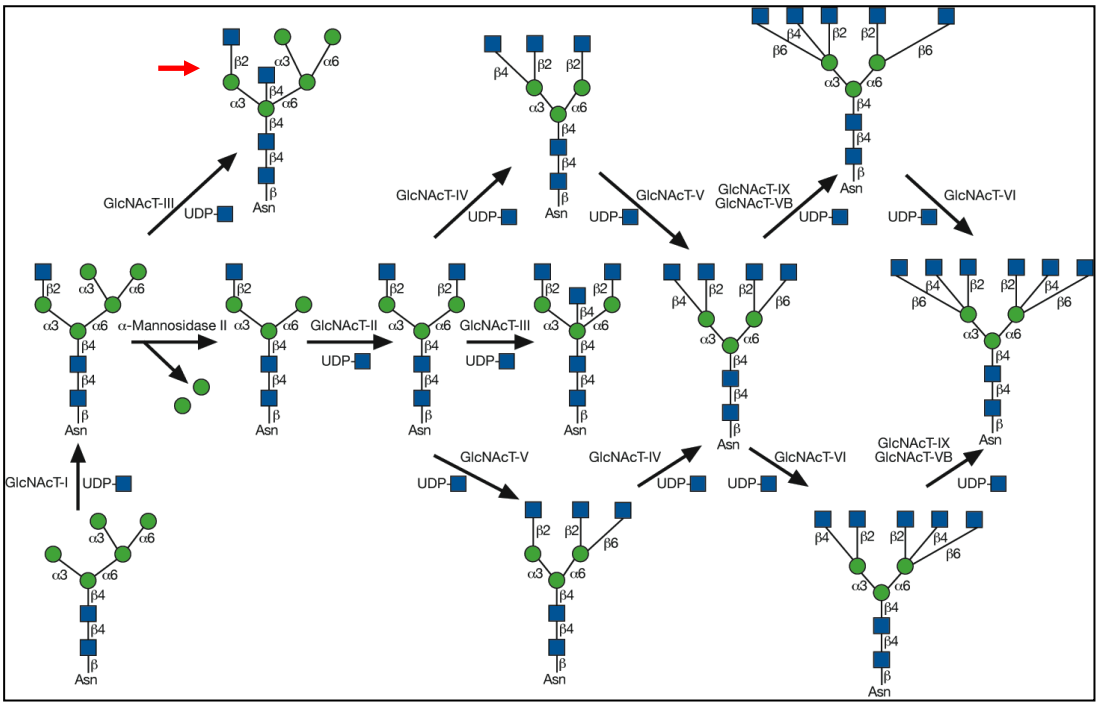

Figure 2.9: Branching of complex type N-glycans (Adapted from Essentials of Glycobiology, 2nd edition)

In complex type N-glycans, additional branches[1] can be extended at C-4 of the core mannose α1-3 (by GlcNAcT-IV) and C-6 of the core mannose α1-6 (by GlcNAcT-V) to yield tri- and tetra-antennary ones. Further branching reactions to form highly branched hepta-antennary structures by other enzymes such as GlcNAcT-IX, GlcNAcT-VB and GlcNAcT-XI (Figure2.9) are also possible in birds and fish, but not mammals. Besides hybrid type, complex type N-glycans may also carry a "bisecting" GlcNAc that is attached to the β-mannose of the core structure by GlcNAcT-III after the actions of adding second GlcNAc residue to the core by GlcNAcT-II (Figure 2.9). The presence of this "bisecting" GlcNAc could therefore inhibit

[1] Additional branches are extended by adding more GlcNAc units in the mannose α1–3 and mannose α1–6 arms

further actions of GlcNAcT-IV and GlcNAcT-V to create more branches from the core (Figure2.9).

2.3.3 Maturation of N-Glycans

Final maturation of the N-glycans occurs in the *trans*-Golgi (Figure 2.8), converting the limited repertoire of hybrid and branched N-glycans into extensive array of mature, complex N-glycans. The first step of maturation is typically β 1-4 galactosylation – adding one galactose residue linked in β 1-4 to each of the existing branching GlcNAc of the antenna. Following this step, four major modifications – fucosylation, sialylation, 1,3 galactosylation and elongation of LacNAc tandem repeats, are widely observed in mammals.

Fucosylation: Addition of fucose residue via a) α1-6 linkage to Asn-linked *N*-acetylglucosamine (GlcNAc) of the core structure; or b) α1-3 linkage to branching *N*-acetylglucosamines of $GlcNAc_{1\text{-}4}Man_3GlcNAc_2$ (Figure 2.10). The latter modification is part of the terminal "capping" or "decorating" reactions[1] to branches.

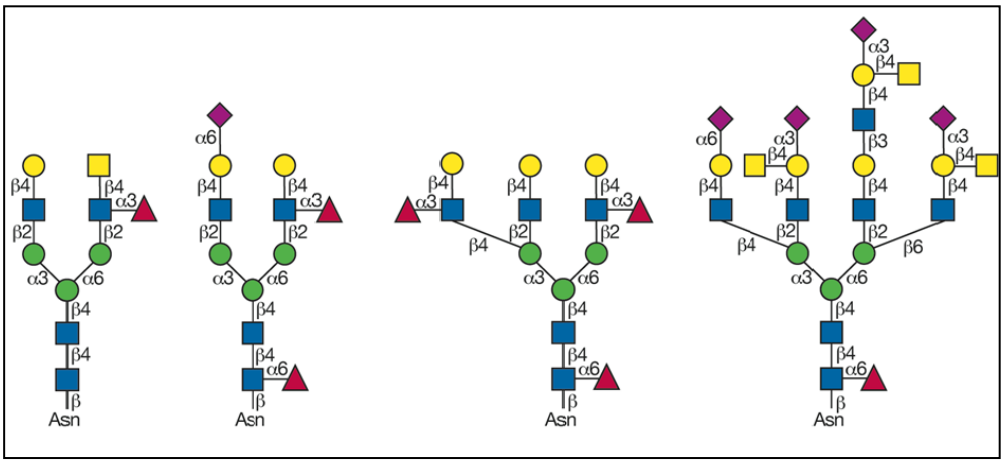

Figure 2.10 Typical complex N-glycan structures found on mature glycoproteins (Adapted from Essentials of Glycobiology, 2nd edition)

[1] Other "capping" and "decorating" reactions that are not mentioned in the main text include addition of *N*-acetylgalactosamine (GalNAc) (yellow square) to branching galactose or *N*-acetylglucosamine (GlcNAc) ((Figure 2.10)

The Lewis blood group[1] antigens are a related set of glycans that carry α1–3 or α1–4 fucose residues, resulted from fucosylation on polyLacNAc chains (discussed later in this section). There are two types of Lewis epitopes: Lewis^a (Le^a) and Lewis^b (Le^b). The structures of Le^a and Le^b epitopes can be shown below (Figure 2.11).

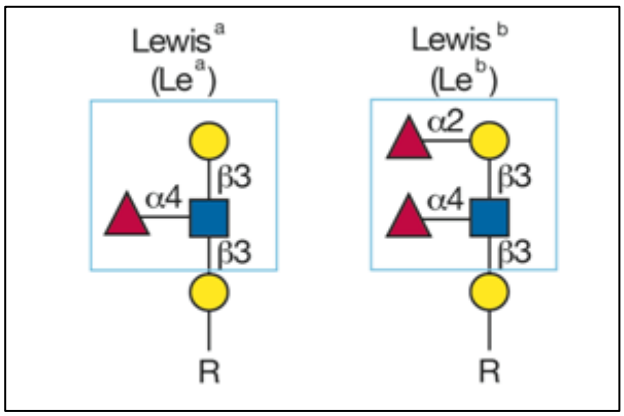

Figure 2.11 Structures of Lewis^a (left) and Lewis^b (right) epitopes (Essentials of Glycobiology, 2nd edition)

Sialylation: Addition of sialic acids i.e *N*-acetylneuraminic acid (Neu5Ac) or *N*-glycolylneuraminic acid (Neu5Gc) is one of the "capping" or "decorating" reactions following addition of β1-4 galactose to branching GlcNAc (Figure 2.10).

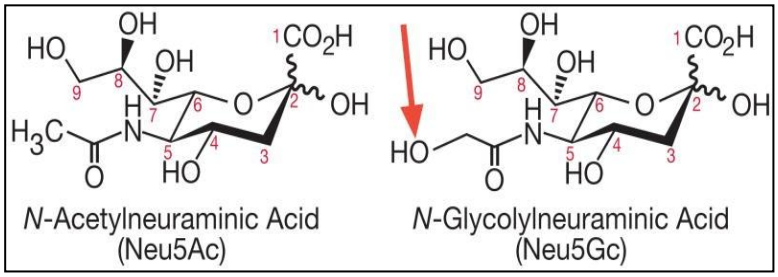

Figure 2.12 Two main types of sialic acids found in mammals – Neu5Ac (left) and Neu5Gc (right)

Sialic acids are terminating monosaccharide units typically found on branches of complex N-glycans, O-glycans, and glycosphingolipids (gangliosides). There are two main types of sialic acids: *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc) (Figure

[1] The term Lewis has its name derived from the family who suffered from a red blood cell incompatibility that also helped in the discovery of this blood group

2.12). The main difference between Neu5Ac and Neu5Gc lies in the extra oxygen atom, in between carbon and hydrogen atoms (indicated by the red arrow in Figure 2.12). Neu5Ac, or NANA is sialic acid terminal sugar exclusive to human as a result of mutation in an enzyme that inserts oxygen atom into Neu5Ac to Neu5Gc (Lieberman, 2008). Neu5Gc is therefore absent in humans but not in other mammals. In therapeutic glycoproteins, animal-derived products used in culture media provide a metabolic source of Neu5Gc (Bardor et al., 2005).

α 1,3-galactosylation: Galα1-3Gal epitope is one that carries two consecutive galactose residues joined via an α1-3 linkage (Figure 2.13). It is immunogenic to human because of the presence of anti-Galα1–3Gal antibodies in human serum. The existence of "α-Gal" epitope could be attributed to the use of non-human cell line that expresses α1-3 galactosyl transferase (Figure 2.13) to produce therapeutic products.

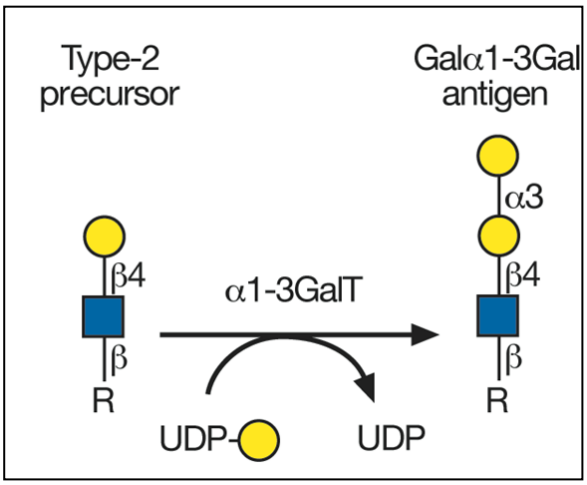

Figure 2.13 Structure of "α-Gal" epitope (right) (Essentials of Glycobiology, 2nd edition)

As earlier mentioned, since humans have circulating antibodies against immunogenic glyco-epitopes such as α-Gal or Neu5Gc, a potential for antigen–antibody responses exists that could be detrimental and/or affect the efficiency of therapeutic proteins.

Elongation of LacNAc tandem repeats (-3Galβ1-4GlcNAcβ1-) to form poly-N-acetyllactosamine or polyLacNAc (Figure 2.14). Poly-*N*-acetyllactosamine biosynthesis is directed by the alternating actions of β1–4 galactosyltransferases and β1–3 *N*-acetylglucosaminyltransferases to add galactose and *N*-acetylglucosamine (GlcNAc), respectively (Figure 2.14). PolyLacNAc chains serve as acceptors for subsequent glycosylations, including fucosylation and sialylation. The linear nature and hydrophilic character allows it to be extended and serve as scaffolds presenting specific terminal glycans for recognition by mammalian galectins.

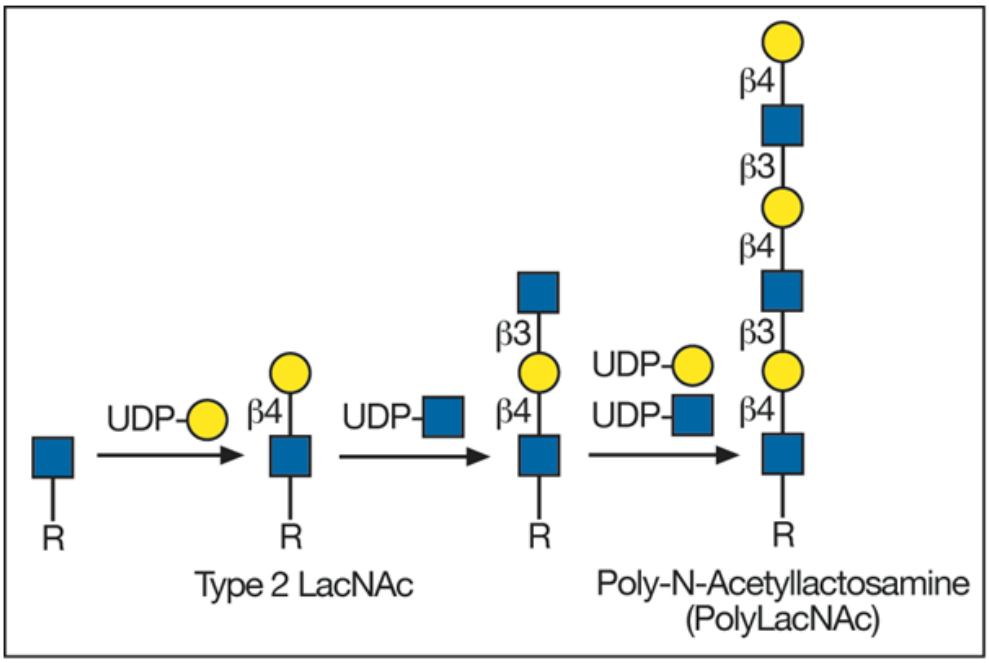

Figure 2.14 Elongation of branch N-acetylglucosamine residues of N-glycans (Adapted from Essentials of Glycobiology, 2nd edition)

2.3.4 Roles of N-glycans in protein folding

Besides the primary sequence, early N-glycan processing in the endoplasmic reticulum (ER) plays an important role in proper folding of membrane proteins and proteins destined to be secreted, and thus their three-dimensional protein conformation. Proper folding of glycoprotein involves formation of secondary structures like α-helices and β-strands, burying of hydrophobic residues in the interior of the protein, formation of disulphide bonds, and

quaternary associations via oligomerization and multimerization. The ER lumen has a highly specialized environment for proper protein folding due to its oxidizing environment that promotes disulfide bond formation and reservoir of Ca^{++} required for binding activities of chaperones like calnexin (CNX) and calreticulin (CRT). These lectin-like chaperones recognize and bind to monoglucosylated forms of the N-glycan i.e. $Glc_1Man_9GlcNAc_2$ on glycoprotein backbone to ensure correct protein folding prior to exit from ER.

In mammals, N-glycan precursors i.e. $Glc_3Man_9GlcNAc_2$-P-P-Dol, are added to the potential N-glycosylation sites of an incompletely folded protein backbone. The non-charged, bulky and hydrophilic nature of N-glycan precursors keep the glycoproteins soluble during folding while modulating protein conformation by forcing amino acids near the N-glycan precursors into a hydrophilic environment. Meanwhile, two glucose sugar units are removed by α-glucosidases I and II to yield $Glc_1Man_9GlcNAc_2$, where it is then bound to CNX/CRT complex. Other chaperones[1], on the other hand, bind to hydrophobic patches exposed on misfolded proteins and maintain their solubility as the proteins acquire their final conformation. In addition, enzymes such as protein disulfide isomerases (PDI) and endoplasmic reticulum proteins ERp59, ERp72 and ERp57 promote proline *cis–trans* isomerization and protein disulfide bond formation that are essential to proper protein folding. Therefore, a difference in N-glycosylation between two proteins in ER could result in a different protein conformation en route to Golgi thus causing further differences in the late processing or maturation of N-glycans such as fucosylation, sialylation etc.

[1] BiP/Grp78, a glucose-regulated protein and member of the hsp70 family of chaperones, Grp94, and Grp170

2.4 Roles of N-glycans in therapeutic proteins

2.4.1 Glycans in Biotechnology and the Pharmaceutical Industry

Glycans are important components of many therapeutic agents, from natural products[1] and small molecules[2] based on rational design, to recombinant glycoproteins[3] because they can have important effects on biosynthesis, biologicial activities, and therapeutic efficacy of the glycoproteins (Bertozzi et al., 2009; Hossler et al., 2009). Glycobiology and carbohydrate chemistry have become increasingly important in modern biotechnology. In 1996, US Food and Drug Administration (FDA) requires for patent application that the glycoform profile of a therapeutic glycoprotein be extensively characterized. Glycoproteins, which include monoclonal antibodies, enzymes, and hormones, are fast growing in the biotechnology industry, with sales exceeding billions of dollars annually.(Varki and Sharon, 2009).

2.4.2 Therapeutic glycoproteins

2.4.2.1 Sialylated glycans improve circulating half-life of Erythropoietin (EPO)

Erythropoietin (EPO) is perhaps the most successful biotechnology therapeutics to date. It is a cytokine that circulates and binds to the erythropoietin receptor, inducing proliferation and differentiation of erythroid progenitors in the bone marrow thereby promoting erythropoiesis – red blood cell production. As a therapeutic, it is used to treat anaemia caused by lack of erythropoietin or by bone marrow suppression[4].

EPO is a recombinant glycoprotein that carries three sialylated complex N-glycans and one sialylated O-glycan. Though only marginal difference in activity is observed for glycosylated and deglycosylated EPO in vitro, glycosylation is crucial for the circulating half-

[1] Natural products that possess glycan structures are antibiotics such as streptomycin, erythromycin A, chemotherapeutic drug such as doxorubicin, and digoxin used in cardiovascular disease
[2] Examples are synthetic influenza neuraminidase inhibitors Relenza™ and Tamiflu™
[3] Glycoproteins are monoclonal antibodies, hormone, enzymes etc.
[4] Lack of erythropoietin can be caused by renal failure; bone marrow suppression can be result after chemotherapy

life of EPO. It is found that the activity of deglycosylated EPO reduced by about 90% as they are rapidly cleared from the body before EPO can act on the receptors (Cummings and McEver, 2009). To reduce the rapid clearance effect for EPO, one can do so by having fully sialylated N-glycan chains; by increasing the amount of tetra-antennary branching[1]; by deliberately adding a N-glycosylation site; and by covalently linking polyethylene glycol (PEG) to the glycoprotein.

2.4.2.2 Effector functions of immunoglobulin (Ig) F_c is glycan-dependant

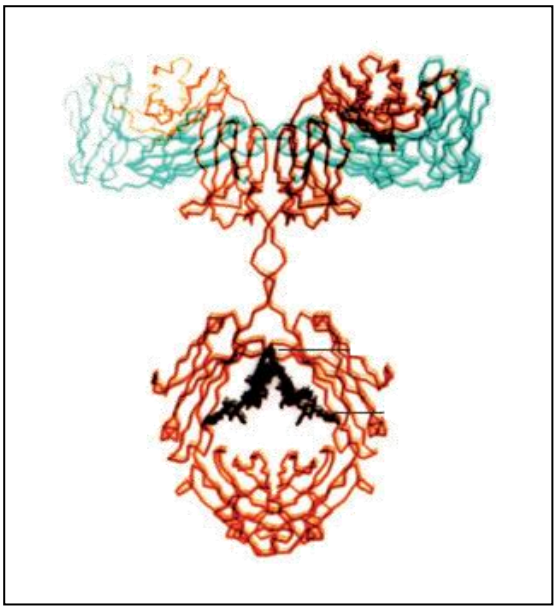

Figure 2.15 The α-carbon structure of the immunoglobulin (Ig) G (Adapted from Jefferis 2009)

While an immunoglobulin uses its F_{ab} to bind and recognize surface antigens on target, it relies on F_c domains to activate the complement and effector functions[2]. To activate the effector functions, F_c of immunoglobulins bind to limited set of effector molecules and F_c receptors of natural killer cells, neutrophils and eosinophils (Siberil et al., 2007; Nimmerjahn and Ravetch, 2008) that trigger inflammatory response and antigen elimination via mechanism such as antibody-dependent cell-mediated cytotoxicity (ADCC) and complement dependent cytotoxicity (CDC) (Raju, 2008; Chan and Carter, 2010).

[1] These approaches increase activity of EPO by nearly ten-fold

[2] Complement system is a process of marking the antigen bearing targets and ingestion by phagocytes, which is also called opsonisation; effector function is triggered when F_c receptors of NK cells neutrophils and eosinophils bind to F_c of immunoglobulins that eventually cause ADCC to occur.

The primary amino acid sequences of different F_c domains[1] play the primary role in effector functions in general. Recent studies have shown that F_c glycosylation of glycoproteins are also essential for the activation of effector functions and complement necessary for ADCC (Burton and Dwek, 2006; Kaneko et al., 2006; Anthony et al., 2008). In particular, the impact of F_c glycosylation on antibody structure and therefore therapeutic efficacy is also evident (Mimura et al., 2000; Ha et al., 2011). In another study, the removal of the fucose from N-glycan core structure attached to the F_c domains of IgG has also been shown to dramatically enhance the effector functions (Shibata-Koyama et al., 2009). Besides immunoglobulin G, effect of *N*-glycosylation to the conformation of immunoglobulin M (IgM) has also been reported (Wormald et al., 1991). Abnormality of glycosylation at Asn-402 due to amino acid exchange at position 406[2], on the heavy μ chain of IgM has shown to cause defect in complement-dependant cytolysis. It was believed that the point mutation causes a conformation change of Cμ3 domain folding that affects glycosylation at the Asn-402 (Wright et al., 1990).

Hence, it is believed that N-glycosylation has a role to play in protein conformation and also biological functions of an antibody that would allow strategic optimization of glycosylation of therapeutic glycoprotein to achieve maximum efficacy (Jefferis, 2009; Jiang et al., 2011).

[1] Different F_c portions give rise to different classes or subclasses of immunoglobulin (Ig). Examples of subclasses for IgG are IgG1, IgG2, IgG3 and IgG4

[2] Mutation has occurred that cause serine as position to be exchanged for asparagine

2.5 Characterization of immunoglobulins (Ig)

2.5.1 Glycomics

Glycomics, which belongs to one of the "omics" science[1], is the systematic and methodological elucidation of the glycome[2], the complete spectrum of glycans and their biological relationship in a given cell type or organism. Compared to the genome or proteome, elucidating the glycome of a cell type is no less than a daunting task, simply because of the vast structural diversity in glycans –heterogeneity of glycosylation sites and glycan linkages; heterogeneity across different dynamic changes[3]; and of intraspecies and interspecies. Therefore, such extensive works require collaborative efforts like the Consortium for Functional Glycomics (CFG), which allows selected participating investigators to contribute and reveal structures, functions of glycans and glycan-binding proteins (GBPs) that have impact on human health and diseases. Within CFG, comprehensive databases on glycan array screening, glycogene microarray screening, mouse phenotyping, and glycan profiling are available on http://www.functionalglycomics.org.

Diversity of glycans has proven to be vital in almost every biological activity, from intracellular signalling to organ development and tumor growth. Glycomics complements other "omics" sciences in providing a better picture of the physiology of a cell or organism. Despite its importance, progress in glycomics has always lagged behind that of proteomics and genomics[4] until the 1980s, when the development of new technologies for exploring the structures and functions of glycans became available.

[1] Genomics, Transcriptomics and Proteomics
[2] The totality of glycan structures
[3] Dynamic changes in the course of development, differentiation, metabolic changes, malignancy, inflammation, or infection of cell type

2.5.2 Characterization of glycosylated immunoglobuins

2.5.2.1 Detection of glycosylated proteins

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) separates proteins based on size. Glycosylated proteins usually show one or more diffuse bands due to heterogeneity in the glycans. One could further investigate the presence of N-glycans by treating the glycoproteins with endoglycosidases such as peptide N-glycosidase F (PNGase F) or A (PNGase A). The result of such treatment is a change in mobility of protein bands such as collapse of diffuse bands or reduced molecular weight due to release of glycans from the protein backbone. The common detection methods for SDS-PAGE gels are performed using silver-staining (www.invitrogen.com) or coomassie blue staining depending on the amount of protein available and the level of sensitivity required.

2.5.2.2 Detection of terminal glycan structures or glyco-epitopes

Lectins are sugar binding proteins that are specific to terminal fucose, galactose, *N*-acetylgalactosamine and terminal sugars such as α2,6-linked sialic acid. Usually, lectins are used as primary antibodies in Western blot to detect the presence of specific glycan types. One example of a commercially available lectin is concanavalin A that binds to non-reducing terminal α-D-mannosyl and α-D-glucosyl groups (Goldstein and Poretz 1986).

Besides lectins, there are other antibodies that are specific to certain glyco-epitopes such as α-Gal, Neu5Gc or Lewis^a that can be used to detect their presence on proteins, especially immunoglobulins. Other methods such as radiolabelling and metabolic labelling are also used to detect the presence of glycans on glycoproteins (Mulloy et al., 2009).

2.5.2.3 Detection of glycoforms

Micellar electrokinetic chromatography (MEKC) is a technique that is used for evaluation of the percentage of site-occupancy analysis or fractionation and profiling of different glycoforms of immunoglobulins (James et al., 1994; Hooker and James, 2000). In MEKC, glycoform migration time is inversely proportional to the amount of glycans attached

to the protein. In other words, the *M* number of *N*-glycosylation sites of a glycosylated protein may be occupied at varying degree thus generating the possibility of *M+1* peaks including one that has no site being occupied at all.

2.5.3 Characterization of N-glycans

2.5.3.1 Release and fractionation of N-glycans

The release of N-glycans is often performed using enzymatic digestion by endoglycosidases. Peptide-N glycosidase F or PNGase F is mostly used (Tarentino and Plummer, 1994), because it cleaves all N-linked glycans except those with fucose attached to α1-3 of the proximal GlcNAc (Tretter et al., 1991). Such fucose linkage is usually present in plant-produced glycoprotein and it can be cleaved using PNGase A instead. Released N-glycans are then purified by chromatography followed by either fluorescent labelling or permethylation to increase sensitivity towards ionization. Then, labelled or permethylated N-glycan preparations are cleaned and desalted before being analyzed using mass spectrometry (MS), which will be discussed in Section 3.2.2.5.

Depending on the conformation of the glycoprotein, N-glycans may sometimes be hidden in the core structure and therefore be inaccessible to the digestive action of endoglycosidases. In such cases, glycoproteins are trypsinized[1] to yield peptides and glycopeptides with N-glycans that are accessible to endoglycosidases. Instead of endoglycosidase treatment, peptides and glycopeptides resulting from trypsinization can also be injected into a liquid chromatochraphy (LC) column followed by MS analysis to obtain site-specific information of N-glycans (Sumer-Bayraktar et al., 2011). Besides using trypsin, it has

[1] Trypsinization – enzymatic action by trypsin, a serine protease that cleaves peptide chains at the carboxyl side of amino acids lysine or arginine, except when either is followed by proline

been reported that chemical cleavage methods like cyanogen bromide[1] and mercaptan-induced fragmentation were used to elucidate the site-occupancy of multiple N-glycan sites present in murine IgM (Anderson and Grimes, 1982).

2.5.3.2 Profiling of released N-glycans using Mass Spectrometry (MS)

Mass spectrometry (MS) is an analytical technique that ionizes and separates the resulting charged particles according to their mass-to-charge ratio (*m/z*). Mass spectrometry is well-known for its high sensitivity in detecting a wide range of masses and thus allows user to obtain the entire N-glycan profile of a glycoprotein. Information that can be obtained from mass spectrometry includes molecular mass, composition, sequence or branching of a glycan or glycopeptide chain. An MS instrument typically consists of a vaporization/ionization unit, an analyzer, and a detector. There are three types of vaporization/ionization units which are as follows:

1. Fast atom bombardment (FAB);
2. Matrix-assisted laser desorption/ionization (MALDI);
3. Electrospray ionization (ESI).

These three technologies allow direct ionization of non-volatile intact glycans, peptides and glycopeptides fragments.

In FAB-MS, samples are dissolved in a liquid matrix and ionization/desorption is effected by a high-energy beam of particles fired from an atom or ion gun. High field magnets are the most powerful analyzers for this type of mass spectrometry. The strong ionisation nature of FAB-MS that generates unnecessary in source fragment ions, makes it unsuitable for analysis of biological compounds as intact molecules. As a result, ESI-MS and MALDI-MS, soft ionization methods which do not induce in source ion fragmentation, are more popular in the analysis of N-glycans on biological compounds.

[1] Mouse myeloma immunoglobulin IgM heavy chains were cleaved with cyanogen bromide into nine peptide fragments, four of which contain asparagine-linked sites of glycosylation. There are five potential N-glycosylation sites, two of them were found in 1 peptide fragment and other three sites were found in separate peptide fragments.

In MALDI-MS, glycan samples and matrix are spotted, and air-dried to form crystals on the surface of a metal target. Matrix molecules, such as 2,5-dihydroxy benzoic acid (DHB), absorb energy from laser pulses and assist the energy transfer and ionization of glycan or glycopeptide samples. Ionized molecules would then travel through a time-of-flight (TOF) analyzer, where samples are separated according to the *m/z* ratio i.e. those with lower *m/z* would travel faster through the analyzer than those with higher *m/z*. MALDI-TOF itself does not generate in source fragments like FAB-MS does. This shortcoming is usually overcome by using two analyzers in tandem i.e. MALDI-TOF-TOF (MS/MS) for structural analysis which will be discussed in Section 2.5.4. However, MALDI-TOF has the advantage of being highly sensitive and able to produce singly charged ions.

In ESI-MS, a stream of liquid containing the samples is stripped of solvent as it enters the ionization chamber, producing multiply charged particles. ESI-MS experiments are often performed in tandem with a quadrupole analyzer, and/or micro- or nanobore liquid chromatography (LC) permitting on-line chromatographic separation and mass spectrometry of different glycopeptides for example (LC/ESI-MS) (Morelle and Michalski, 2007).

2.5.3.3 Sialic acid profiling of N-glycans

Released N-glycans can be fractionated based on sialic acid content using HPAEC-PAD (High pH anionic exchange chromatography with pulsed amperometric detection). Sialic acids are negatively charged terminal sugar entities due to the presence of carboxylic acid (COO^-) groups within the structures. Under high pH conditions and anionic exchange column, more sialylated N-glycans are more strongly retained by the stationary phase than those that are less or non-sialylated. This results in a separation profile based on the sialic acid content of the released N-glycans. Besides total sialic acid distribution, a high-throughput method to quantify the sialic acid content of glycoproteins has also been recently described (Markely et al., 2010).

2.5.4 Structural analysis of N-glycans

Mass spectrometry (MS) generates spectra of mass ions that must be substantiated by the glycan database associated with the host cell producing that glycoprotein. However, in certain cases, more than one glycan structure can be attributed to a specific mass peak, which requires further analysis to confirm the claim of one specific structure to that peak. There are several ways to do so, e.g. by MS/MS, NMR or exoglycosidase digestion.

In MALDI-TOF-TOF (MS/MS), selected intact masses from the first analyzer are subjected to collision in an environment filled with inert gases such as air, helium (He), argon (Ar) in the chamber between two TOF analyzers and therefore fragmented. The generated profile of fragmented masses associated to the selected mass peak would provide information on the sequence, linkage or branching information of a glycan chain. Analyzing the profile of mass peaks by MALDI-TOF-TOF (MS/MS) or MALDI-TOF (MS) can either be performed using certain commercially available SimGlycan® software which analyzes the mass spectrum of MS/MS and provides suggested structures with different scores and ranks (discussed later), or manually through mass differences of consecutive mass peaks which will be discussed in Section 4.2.2.

Nuclear magnetic resonance (NMR) spectroscopy, on the other hand, can provide anomericity, sequence and linkages of a particular monosaccharide residue in a glycan using ^{1}H-NMR but also requires large quantities of materials, i.e. typically at least one milligram. Furthermore, removal of specific sugars by exoglycosidases such as sialidase or b-galactosidase, from the terminal ends of glycans may result in a mobility change depending on the residues removed. Sometimes, glycans that are subjected to actions of these enzymes may also be reanalyzed using mass spectra by looking at the shift of certain mass peaks to substantiate the claim of specific structures to those peaks.

2.6 Comparative modeling of protein 3D structures

Protein 3-dimensional (3D) structures can provide insights into many biochemical functions at a near atomic-level resolution. However, the high costs involved and immense efforts required to experimentally determine protein structures limits the use of such approach. Computational prediction of protein structures provides a low-cost alternative to obtain such structures. Comparative modeling is a template-based approach to predict the 3D structure of a target protein, primarily based on its sequence similarity to existing homologous protein structures. Though comparative modeling is predictive in nature, it is the most accurate computational method to date. In fact, a model structure can be quite close to that of a real protein if its sequence similarity is relatively high between them.

One reason for the viability of computational protein structure prediction techniques is that protein structures are more dependent on their amino acid sequences and less on the species that produces them (Bajaj and Blundell, 1984; Chothia and Lesk, 1986; Chothia and Lesk, 1987). It has been recently reported that there are about 78,477 protein structures deposited on the Protein Data Bank (PDB) as of today when this thesis was written, and automatic prediction has generated approximately 1.9 million models that have not been determined experimentally (Liu et al., 2011).

For comparative modeling, sequence identity between the target protein and its template can be as low as 30% for relatively reliable structures to be predicted (Ginalski, 2006). Below this level, one could resort to fold recognition algorithms i.e. piecewise assembly of smaller peptides to model a protein structure from a target sequence. Comparative modeling, or homology modeling, requires a full template protein structure to be present and identified. Fold recognition techniques, on the other hand, does not require the full template, and thus has broader applicability especially for uncharacterized proteins with no existing templates in PDB. There are two algorithms for fold recognition: *ab initio* and *de novo* methods. *Ab initio* methods only rely on physicochemical principles for atom simulation, whereas *de novo* methods also include information from known protein structures. There are

approximately 1,400 unique folds in the current Structural Classification of Proteins (SCOP) database (Andreeva et al., 2008).

To augment the existing PDB and SCOP databases, structural genomics initiatives (SGIs) have been launched to explore different regions of protein structural space by selecting targets from novel, structurally uncharacterized protein families (Chandonia and Brenner, 2006). To date, these initiatives have added 9,600 new structures to PDB.

2.6.1 Methods for comparative modeling

Comparative modeling is best viewed as a strategy, rather than a single technique, for assembling information from various component methods (including assembly and associative techniques) toward a 3D structure prediction (Lushington, 2008). A flowchart of comparative modeling is shown (Figure 2.16), which comprises four sequential steps that are shown as follows.

1. Fold recognition and template identification
2. Target-template sequence alignment
3. Model building and refinement
4. Model evaluation or validation

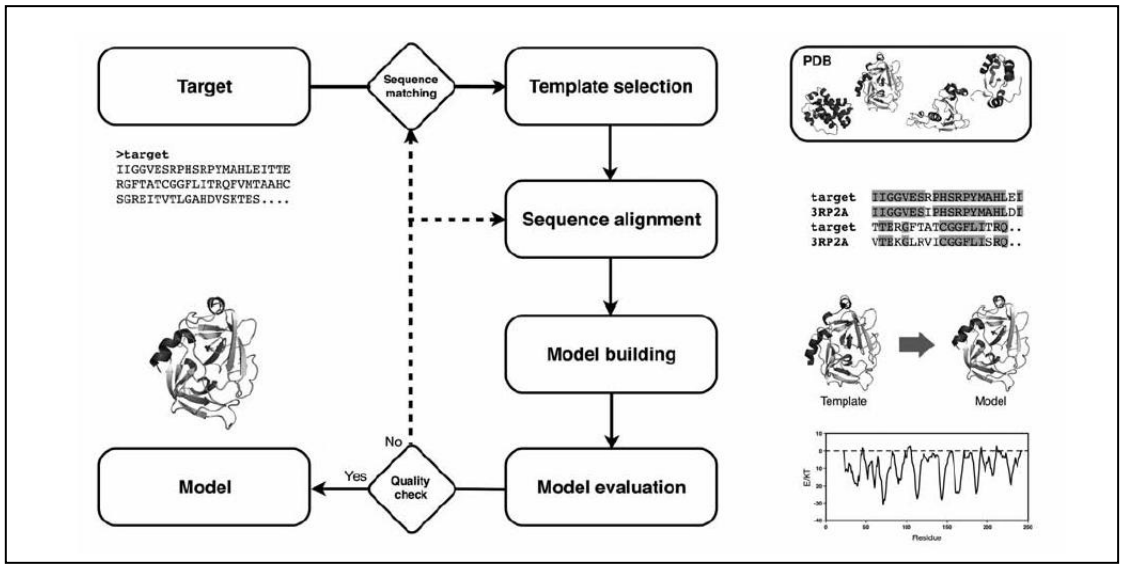

Figure 2.16 Flowchart of comparative modeling method (Liu et al. 2011)

2.6.1.1 Fold recognition and template identification

The first step of the comparative modeling strategy involves the identification of template protein structures. These structures can be obtained from PDB database (Berman et al., 2007). Other databases such as SCOP (Andreeva et al., 2008), DALI[1] (Holm and Sander, 1998), and CATH[2] (Cuff et al., 2009), can be used to narrow the search. To detect homologous protein structures from databases, searching algorithms such as BLAST (Altschul et al., 1990) and FASTA (Pearson, 1990) based on pairwise sequence comparison of target and template sequence, are used. In cases when sequence identity is low, PSI-BLAST (Altschul et al., 1997) algorithm can be used to improve the detection of homologous protein sequences for a specific target protein sequence. Besides BLAST, FASTA and PSI-BLAST, other profile-based algorithms, which were believed to perform better in the implementation of alignment procedure have also been described (Liu et al., 2011).

Generally, BLAST search within the databases will return more than one suggested template structures that are sorted in order of their bits scores (highest first) and expectation values or E-values generated by the algorithms. Bits score indicates the quality of the best

[1] DALI is an acronym that stands for Distance mAtrix aLIgnment
[2] CATH is an acronym of four main levels of classification – C, A, T and H stand for Class, Architecture, Topology and Homologous superfamily, respectively.

alignment between target sequence and found template; whereas E-value tells of the biological significance of the search result, and the likelihood of common ancestry between target and template. Selection of which template to use for model building depends on what the final model is used for - whether it is for the study of protein-protein/ligand interactions, or conformation of target's active-site. Templates that contain similar types of interactions as the target are important if it is for interaction study, whereas in the case of the conformation study, high resolution template i.e. one with highest bit score, is more desirable. In addition, to improve the quality of final predicted structure, use of multiple templates has also been attempted (Larsson et al., 2008).

2.6.1.2 Target-template sequence alignment

After the templates are found, they must be aligned to the target protein sequence. Standard sequence alignment methods used are Needleman-Wunsch (Needleman and Wunsch, 1970) and Smith-Waterman (Smith and Waterman, 1981). These methods are based on dynamic programming algorithms that calculate scoring matrices such as BLOSUM (Henikoff and Henikoff, 1992) and PAM (Dayhoff et al. 1978). If sequence identity between target and template is high, these methods produce similar alignment; however, if sequence identity is low (usually less than 40%), multiple sequence alignment of homologous proteins has been used to improve alignment results (Jones et al., 1999; Rychlewski et al., 2000; Marsden et al., 2002; Capriotti et al., 2004). Upon alignment, the molecular model can then be constructed.

2.6.1.3 Model building and refinement

There are three main approaches towards model building – rigid body assembly (Sutcliffe et al., 1987), segment matching (Levitt, 1992) and satisfaction of spatial restraints (Sali and Blundell, 1993). In rigid body assembly, atomic coordinates of the conserved regions are used to construct the main chain of conserved residues, core of the target protein, where loops[1] and side chain atoms that fit the core protein conformation are added. On the other

[1] Loops are selected by scanning a database of structural peptide fragments. Loops are usually added before side chain atoms to create the entire protein backbone structure first

hand, segment matching uses 100 six-residue peptides that account for 76% of protein conformational space to build the target core using Cα atoms from conserved residues (Unger et al., 1989). Lastly, satisfaction of spatial restraints uses the similarity of structural features of conserved residues to build 3D model subjected to restraints such as generic stereochemistry from molecular mechanics force fields, distance and angles between equivalent residues based on entire target-template alignment. Finally, an optimization is performed to search for global low energy conformations that minimize the restraint violations.

The most difficult tasks during model building are the prediction of loop regions and side-chain conformations, which are often performed by tedious methods and protocols[1] via trial-and-error. Model refinement, on the other hand, is done using molecular dynamics (MD) techniques. With the improvement in automated tools, the accuracy of the model built automatically is comparable to manually curated multiple template models (Venclovas and Margelevicius, 2009).

2.6.1.4 Model evaluation/validation

The final step in comparative modeling is the evaluation or validation of predicted 3D model structures. Algorithms that can be used to perform such task include PROCHECK, AQUA, SFCHECK, Squid, Molprobity (Oldfield, 1992; Laskowski et al., 1996; Vaguine et al., 1999; Chen et al., 2010). These algorithms check the stereochemical properties[2] of the predicted model structures to access their reliability.

There is also another class of programs that evaluates the predicted model structures based on statistical potential mean force, of which its theoretical basis is highly debated (Finkelstein et al., 1995; Rooman and Wodak, 1995; Thomas and Dill, 1996).

[1] Methods for predictions are molecular graphics through database searching and *ab initio* methods. Protocols for side chain building are Minimum Perturbation and Coupled Perturbation described in Liu et al. 2011

[2] Bond lengths and angles, peptide bond and side-chain ring planarities, chirality, main-chain and side-chain torsion angles, and clashes between non-bonded pairs of atoms.

Other algorithms using structure-bases scoring functions and physics-based energy functions have also been used to perform model assessment (Benkert et al., 2008; Benkert et al., 2009).

3 MATERIALS AND METHODS

3.1 Materials

3.1.1 Purified IgM 84 and 85

mAb 84 and 85 are monoclonal antibodies, or immunoglobulin (Ig) M produced by mouse hybridoma clones, which in this thesis are referred to as IgM 84 and 85, respectively. They were generated by the Stem Cell group at Bioprocessing Technology Institute (BTI) as previously described (Choo et al., 2008). The clones of mAb 84 and 85 were first adapted and cultured in protein-free media by the Animal Cell Technology group at BTI in 5L bioreactor. Cell culture supernatants of mAb 84 and 85 were then clarified, captured and purified in two steps[1] by the Downstream Processing group at BTI to achieve product purity above 95% in final storage buffer of 30mM sodium phosphate, 100mM NaCl, 5mM EDTA, 0.05% Tween 80 and 73mM or 2.5% trehalose pH 7.5 (Tscheliessnig et al., 2009).

3.2 Methods

3.2.1 Construction of mouse N-glycans library

Using Glycan Profiling database from Consortium for Functional Glycomics (CFG), mouse N-glycans library was constructed. The coverage of this library includes all cell types and spleen tissues of mouse species from various participating investigator as listed under Appendix C. All the structures are drawn using GlycoWorkbench software and saved as *.gws* file for analysis of mass spectra for N-glycan profiling of IgM 84 and 85.

[1] Previously discussed, 2-step purification strategy consists of protein precipitation, followed by anionic exchange chromatography (AEX)

3.2.2 Release and Fractionation of free N-glycans from IgM 84 & 85

3.2.2.1 Fragmentation of IgM 84 and 85

Fragmentation of IgM 84 and 85 was performed using Pierce® IgM Fragmentation Kit. To do so, 400µg of purified IgM 84 (or *approx.* 174µl of 2.3 mg/mL) and 85 (or *approx.* 191µl of 2.1 mg/mL) were first buffer-exchanged into digestion buffer (50mM Tris, 150mM NaCl, 10mM $CaCl_2$, 0.05% NaN_3 pH 8.0) that is provided in the kit, 3 times using centrifugal concentrators (10,000 MWCO, *Amicon Ultra, Millipore*) and final concentration was adjusted to 1.0 mg/mL assuming no loss of samples during this step. Samples of IgM were loaded onto immobilized typsin column at 60^0C for 40mins before fragments of IgM were eluted and collected in fractions. Details of this method, which is also known as hot trypsin digestion (HTD) in the kit, can be found in the datasheet of Pierce® IgM Fragmentation Kit (*Thermo Scientific).* Two major fragments for IgM of mouse origins, as suggested by the protocol, are believed to be $F(ab')_2$ (150kDa) and "IgG" type-M (200kDa) (Figure 3.1).

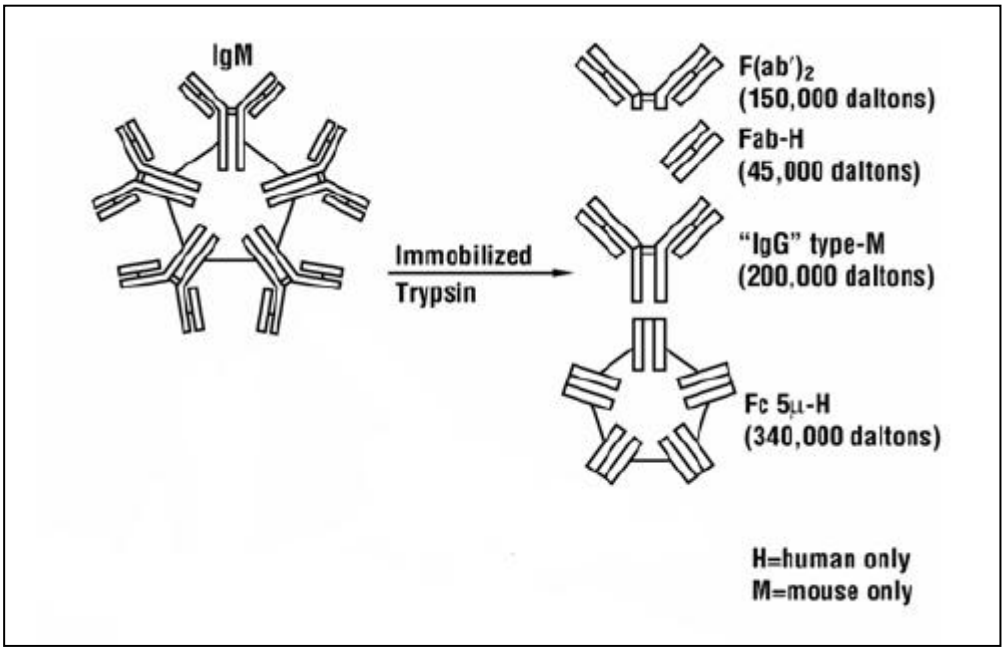

Figure 3.1 IgM fragments generated using trypsin (Adapated from Pierce® IgM Fragmentation Kit (Thermo Scientific) datasheet

3.2.2.2 Trypsin digestion of IgM 84 and 85

Eluted fractions containing fragments of IgM84 and IgM85 were pooled, and equal volume of trypsin digestion buffer (50mM NH_4HCO_3, pH 8.2) was added to each pool of IgM sample. These samples were then further digested by adding 400µl of 0.2µg/µl Sequencing Grade Modified Trypsin (*Promega*), which had previously been reconstituted in 1mM HCl solution, and incubated at 37^0C with end-over-end mixing overnight. The same amount or volume of trypsin was added to each sample after 24 hours, and incubated once again at 37^0C for a total of 48 hours to ensure complete digestion. This step was verified using SDS-PAGE technique, similar to that was described in Section 3.2.5.2. (Data not shown)

Finally, N-Glycanase® (GKE-5006B, *Prozyme*) or Peptide-N-Glycosidase F (PNGase F) was diluted 25 times in N-Glycanase Reaction Buffer (100mM sodium phosphate, 0.1% NaN_3, pH 7.5) before adding 100ul of 100mU/ml PNGase F into the mixture of completely digested IgM. Resultant mixtures were incubated overnight at 37^oC. PNGase F from *Prozyme* cleaves all asparagine-linked complex, hybrid and high mannose oligosaccharides.

3.2.2.3 Reversed-phase capture of free N-glycans using Hypercarb column

Hypercarb SPE cartridges (200mg sorbent bed weight, *Thermo Scientific*) were used to capture free N-glycans from pool of digested mixture of peptides, enzymes and etc. Cartridges were pre-washed sequentially in the following order: 1CV[1] of 1M filtered NaOH, 2CVs of water[2], 1CV of 30% acetic acid (v/v), 1CV of water, 1CV of 50% ACN/0.1% TFA (v/v), and 2CVs of 5% ACN/0.1%TFA (v/v) and 2CVs of water. Prior to sample loading, half the volume of each sample was kept for HPAEC-PAD in Section 3.2.4.2, which will be discussed later.

After sample loading, each sample vial was washed thoroughly with 2mL of water, and the washes were also loaded onto the cartridge. This step is repeated three times to allow

[1] CV stands for column volume; 200mg sorbent bed weight has a 3mL column volume. Hence, 1CV is equivalent to 3mL etc.

[2] Water is always referred to as ultra pure water (18 M Ω .cm, *Sartorius)* unless stated otherwise

for maximum product recovery. Each cartridge was washed with 3mL of water, followed by 3mL of 5% ACN/0.1% TFA (v/v) before free N-glycans were eluted stepwise with 50%ACN/0.1%TFA (v/v) i.e. 500μL of elution buffer was loaded onto each cartridge to reach a final total volume fraction of 2mL.

All eluted samples were collected in different glass vials, and then dried under constant blowing of N_2 gas at room temperature until the entire elution buffer vaporized.

3.2.2.4 Permethylation

Four sodium hydroxide (NaOH) pellets (*Merck*) were ground in *approx.* 3mL of dry dimethyl sulfoxide (DMSO) (*Merck*) that had been added into a dry mortar to form slurry. About 0.6mL of the resulting DMSO/NaOH slurry was added into the glass vial containing dried samples, followed by 1.0mL[1] of iodomethane (*Merck*) to arrive at a volumetric ratio of *approx.* 2:1. Note that iodomethane should be added sufficiently to avoid under-permethylation. Reaction mixture was then left to react at room temperature for at least 1 hour under constant end-over-end mixing. At the end of reaction, water was added *dropwise* to the reaction mixture to quench any excess iodomethane.

To extract permethylated samples from aqueous phase, approximately 1-2 ml of chloroform ($CHCl_3$) (*Merck*) or organic phase was added. The resulting two-phase solution was vigorously mixed and allowed to settle under gravity. Aqueous phase, which would settle on top of organic phase, was aspirated under vacuum. Following this extraction step, organic phase (which now contains the sample) was washed 5-8 times before they were dried under constant blowing of N_2 gas at room temperature until the entire organic phase vaporized. Permethylation[2] increases the hydrophobicity of free N-glycans.

[1] In the protocol that was used in-house, 0.5mL of iodomethane was used. This was because the starting amount of protein used in the protocol was half the amount i.e. 200μg that I used in this experiment.
[2] Permethylation is a chemical reaction that converts all –OH groups which are free in the molecule to –OCH_3 groups

3.2.2.5 Desalting step using Sep-Pak® column

Sample in each glass vial was reconstituted in 200ul of 50% methanol (v/v) aqueous solution. Sep-Pak cartridges (C_{18}, 200mg sorbent bed weight, *Waters*) were pre-conditioned in the following order: 5mL of 100% methanol, 5mL of water, 5mL of 100% ACN and finally 5mL of water prior to sample loading. Upon sample loading, glass vials were washed with 2mL of water and loaded onto the respective cartridges to allow maximum product recovery. Cartridge was washed with 5mL of water to dissociate any potential non-specific binding, and also serve as a clean-up or desalting step. Finally, bounded permethylated N-glycans were eluted in four fractions - 15% ACN, 35% ACN, 50% ACN and 75% ACN aqueous solution. Similar to previous elution steps done in Hypercarb cartridge, each fraction was eluted in four steps i.e. 0.5mL x4 to arrive at a final total volume of 2mL. Eluted samples were freeze-dried before they were characterized by MALDI-TOF mass spectrometry (MS).

3.2.2.6 MALDI-TOF MS

MALDI-TOF MS stands for matrix-assisted laser desorption/ionization (MALDI) time-of-flight (TOF) mass spectrometry (MS). It is one of the three most commonly used mass spectrometry methods to characterize oligosaccharides, peptides or glycopeptides in glycobiology. In our case, permethylated N-glycans from IgM 84 and 85 were reconstituted in 30ul of 80% methanol (v/v). Tubes were thoroughly vortexed to ensure all N-glycans dissolve in the aqueous solution. 0.5ul of sample solution was spotted on a MALDI plate, specially designed for Applied Biosystems 4800 Plus MALDI-TOF/TOF (*Applied Biosystems*), followed by 0.5ul of 2,5-dihydrobenzoic acid (DHB) matrix solution. Additional 0.5ul of 100% ACN was spotted on top of the sample and matrix solutions to allow sample crystallization to take place. To use the Applied Biosystems 4800 Plus for MALDI-TOF MS, the following parameters were set as shown below (Table 3.1).

Table 3.1 Parameters that were set on TOF/TOF™ Series Explorer™ Software

Parameters	Range
Calibration:	
Acquisition methods	
a) Instrument →Mass range	800-5000 Da
b) Automatic control →Laser intensity	5000
Processing methods	
a) Calibration	Internal
Min Signal/Noise (S/N)	70
Min Peaks to match	5
Outlier error	6 ppm
Sample:	
Acquisition methods	
a) Instrument →Mass range	800-5000
b) Automatic control →Laser intensity	5000
c) Spectrum → Acquisition mode	
(shots per sub spectrum/total shots per spectrum)	200/2000[1]
Processing methods	
a) Calibration →Default	Default

[1] 200 shots that pass the acceptance criteria defined by calibration were accepted, and accumulation of shots stopped when 10 sets of 200 shots i.e. 2000 total shots per spectrum were achieved

3.2.2.7 MALDI-TOF-TOF/MS-MS

Selected mass peaks of IgM 84 and 85 samples that had been identified to have few suggested glycan structures were subjected for further fragmentation using MALDI-TOF-TOF or MS-MS. Fragmentation of selected ionized masses passing through collision-induced dissociation (CID) chamber were controlled 1) by increasing potential difference of across analyzers from 0, 1 or 2kV and 2) using heavier colliding inert gases such as argon (Ar) compared to air. Prior to CID fragmentation, CID was purged with inert gases. The accuracy of masses that are entered into preset *MSMS Acquisition Methods* were up to 2 decimal places.

3.2.3 Site specific N-glycan profiling of IgM 84 & 85

3.2.3.1 Reduction and alkylation of IgM 84 & 85

Purified IgM 84 and 85 (100μg) in storage buffer (as described in Section 3.1.1.) were denatured by adding appropriate volume of 8M guanidine hydrochloride (GdnHCl) in 0.3M Tris-HCl pH 8.4, such that the final concentration of GdnHCl is 6M. The resulting mixture was incubated at 37^{o}C for 1 hour. Following denaturation, 1mM dithiothreitol (DTT) aqueous solution was added according to molar ratio of DTT: IgM = 300:1, before incubation at 37^{o}C for 1 hour. After the previous reduction step, IgM samples were alkylated by adding 2.5mM iodoacetamide (IAA) according to molar ratio of IAA: DTT = 2:1. The final mixture was incubated at 37^{o}C for 1 hour in the dark. After final incubation, the mixture was dialyzed twice for 3 hours in large quantity of water i.e. at least 1L using Pierce dialysis cassettes (10kDa MWCO, *Thermo Scientific*) before the samples were left overnight for dialysis at 4^{o}C.

3.2.3.2 In-gel trypsin digestion

Reduced and alkylated samples of IgM 84 and 85 in water were loaded onto a 4-12% Bis-Tris NuPAGE gel (*Invitrogen*) and run for 35 mins at 200V in 1X MES running buffer[1]. Gel was removed from the cassette housing and stained in Coomassie Blue for 5 mins

[1] 20X MES buffer has a formulation of 50mM MES, 50mM Tris Base, 0.1% SDS, 1mM EDTA pH 7.3

before destaining in 10% ethanol, 10% acetic acid for 10 mins in multiple intervals. The destaining buffer was continually changed until protein bands became visible. Protein bands of interest[1] were excised and cut into smaller gel cubes with a sterile scalpel before transferring them into clean Eppendorf tubes. Gel cubes were washed with 200µl 0.1M NH_4HCO_3/ACN (1:1 in volume) by vortexing and incubation under constant end-over-end mixing on a rotary shaker for 15 mins at room temperature. Supernatants of were removed and dried under vacuum. The washings and drying steps were repeated twice. Prior to in-gel trypsin digestion, all protein bands in gel cubes are reduced and alkylated again in by incubation in the dark in 200µl of 0.1M DTT (45 mins, 55^oC) and 200µl of 55mM IAA (30min, room temperature) solutions, respectively. After incubation, samples were cooled to room temperature before they were washed again with 0.1M NH_4HCO_3/ACN (1:1 in volume) before drying under vacuum for 3min. This washing and drying steps were repeated once only.

Lypholized sequencing grade modified trypsin (*Promega*) was reconstituted in 100ul of water to form trypsin solution. 100ul of 20% trypsin solution in 0.1M NH_4HCO_3 (v/v) was added to each sample and incubated for 45 mins at 4^oC before incubating samples at 37^oC overnight. In the next day, gel cubes were washed in the following order: 100µl of 100% ACN, 100µl 0.1M NH4HCO3, 100µl ACN and finally 100µl 5% formic acid. All supernatants were recovered during all these washing steps, and transferred into a glass vial for lyophilisation.

3.2.3.3 Fractionation of glycopeptides/peptides using Sep-Pak® column

Lyophilised samples of IgM 84 and 85 glycopeptides and peptides were reconstituted in 200µl of 50% methanol aqueous solution (v/v). Steps to pre-condition Sep-Pak® cartridges, sample loading and elution, were done according to those have been described in Section 3.2.2.5, except that fractions were eluted at 5 different ACN concentrations instead: 5%, 10%, 15% and 20% ACN and 100% ACN (v/v) before drying.

[1] Protein bands of interest were determined by identifying their molecular weight using a molecular weight marker (Novex Sharp unstained or prestained, *Invitrogen*) loaded alongside with samples

3.2.3.4 MALDI-TOF MS and MALDI-TOF-TOF (MS/MS)

After drying, 30ul of 80%v/v methanol (v/v) was added to each sample before spotting them on MALDI metal plate with DHB matrix and ACN for crystallization of samples to take place. Detail steps were previously described in Section 3.2.2.6. One difference here was the use of α-cyano-4-hydroxycinnamic acid (α-Cyano) over DHB because of its ability to ionize peptides or peptide portion of glycopeptides better. Set of parameters used for TOF/TOF™ Series Explorer™ Software can be found under Table 3.1. For selected mass peaks that needed to be fragmented further, procedure was done according to that had been described in Section 3.2.2.7

3.2.3.5 Amino acid sequence analysis

Full DNA sequencing for IgM 84 and 85 was performed by Stem Cell group and the respective primary amino acid sequences were then derived. We determined all the potential N-glycosylation sites on the heavy and light chains of IgM 84 and 85 besides analyzing the peptide and glycopeptides sequences from trypsin digestion, which was required to determine the site-specific N-glycosylation information of IgM 84 and 85. Only small segment of these sequences are shown in Appendix A due to confidentiality of information.

Furthermore, sequence alignment was performed by Dr. Miranda Van Beers to determine the sequence similarities between the variable and constant regions of heavy and light chains of IgM 84 and 85 (Website: http://pir.georgetown.edu/). The sequence similarities results of different variable regions of IgM 84 and 85 were cross-compared with their respective 3D structural models that we generated in this thesis.

3.2.4 Sialylation of IgM 84 & 85

3.2.4.1 Sialic Acid (SA) quantification using high throughput method (HTM)

The total amount of sialic acids on purified IgM 84, and IgM 85 was quantified by performing a high throughput method as previously described (Markely et al., 2010). To obtain

a calibration range, 1M sialic acid stock solution (*Sigma Aldrich*) was serially diluted to obtain standards at five different concentrations i.e. 0, 10, 20, 40 and 60µM. On the other hand, IgM samples were diluted in storage buffer, as previously described in Section 3.1.1, to a concentration of 100µg/ml.

A volume of 30µl of each standard or sample was diluted in 30µl of 0.2M acetate buffer in 200µl tubes (*Axygen*). 0.2M acetate buffer pH 5.0 was pre-adjusted with 1.21M HCl such that the resulting acetate buffer would give a pH 5.2 to the final mixture of samples/standards that is optimal for neuraminidase activities.For each sample tube, 2µl of water, 1.25µl of neuraminidase[1] (5U/100µl, *Roche*) and 3.75µl of 50mM acetate buffer pH 5.2 were added; whereas for each standard, 2µl of water and 5µl of 50mM acetate buffer pH 5.2 were added without neuraminidase, and the resulting mixtures were incubated at 37^{o}C for 5 min. Following incubation, 90µl of 0.15M borate buffer pH 9.4 was added to each tube, followed by 12µl of in-house prepared malononitrile before the final mixtures were incubated at 80^{o}C for 5 mins. Reaction was then stopped by incubating on ice for 1 min. Finally, mixtures were transferred to 96-well plate to measure fluorescence emission at a wavelength of 430nm after excitation at 357nm.

3.2.4.2 Relative percentage quantification of sialylated N-glycans using HPAEC-PAD

Set of samples that were kept from Section 3.2.2.3, were spiked with 10ul of Raffinose before they were being loaded onto different set of pre-conditioned Hypercarb SPE cartridges (200mg, sorbent bed weight, *Thermo Scientific*). Pre-conditioning of the cartridges and sample elution were done as described in Section 3.2.2.3. HPAEC-PAD, which stands for high pH anionic exchange chromatography with pulsed amperometric detection, is used to fractionate the pool of sialylated N-glycans according to their differential surface charges. After drying, samples were reconstituted in 150µl of water before a volume of 30µl of each sample was

[1] Isolated from *Clostridium perfringens*, an acylneuraminyl hydrolase, EC 3.2.1.18; it cleaves terminal sialic acids linked via α(2-3), α(2-6), or α(2-8), and it is supplied in lyophilized form and is reconstituted in 100µl of water before use

injected onto CarboPac® PA200 analytical column (3x250mm, *Dionex*) of the BioLC system (*Dionex*) at flow rate of 0.3ml/min. Mobile phases are 500mM acetic acid and 500mM NaOH with varying degree of mixing between 0 and 100%.

3.2.4.3 Relative percentage quantification of sialic acid types using HPAEC-PAD

There are two types of sialic acids i.e. Neu5Ac and Neu5Gc and the main difference between them is previously described in Section 2.3.3 and Figure 2.12. These two types of sialic acids were released from samples of IgM 84 and 85 using acidic treatment and incubated in 2M acetic acid for 3 hours at 80^{o}C. Released Neu5Ac and Neu5Gc remain in the solution and were collected in the filtrate after spinning down using 10kDa MWCO centrifugal filters (Amicon® Ultra, *Millipore*) at 14,000rpm for 20 min on a benchtop centrifuge (*Beckman Coulter*). Samples were then dried under vacuum and reconstituted in 150µl of water before injecting samples onto the CarboPac® PA20 analytical column (0.4x150mm, *Dionex*) of the BioLC system (*Dionex*) at flow rate of 0.5ml/min, mobile phases are 500mM acetic acid and 500mM NaOH with varying degree of mixing between 0 and 100%.. Standards of Neu5Ac and Neu5Gc were also injected in separated runs to identify the elution times of these sialic acids in our samples.

3.2.5 Gel electrophoresis and Western blot analysis of glyco-epitopes

3.2.5.1 Protein extraction from mouse heart

Mouse hearts were first harvested, followed by protein extraction using Novex® Tris-Glycine Native Sample Buffer[1] (2X concentrated, *Invitrogen*) diluted with water. Sufficient volume of tris-glycine native sample buffer (diluted to 1X) was used for sonication, which was performed intermittently to avoid excessive heat build-up during this process. Protein extracts were then spun down at 14,000 rpm for 20 mins using a benchtop microcentrifuge (*Beckman Coulter*) and supernatant was removed and used for gel loading in the next step. This sample is

[1] 100mM Tris HCl, 10% Glycerol, 0.0025% Bromophenol Blue, water, pH 8.6

used as the positive control for Western blot detection of gal α(1,3) gal terminal epitope of N-glycans released from IgM 84 and 85.

3.2.5.2 SDS-PAGE Gel Electrophoresis

IgM (between 0.5 – 2.0μg) samples were pre-treated by adding LDS Sample Buffer[1] (4X concentrated, *Invitrogen*) and 500mM dithiothreitol (DTT) (10X concentrated, *Invitrogen)* at 70^0C for 10 mins before loading onto a pre-cast gel of 4-12% Bis-Tris SDS-PAGE[2] (*Invitrogen*). The gel was run at 200V for 35 mins according to protocol by Invitrogen. However, to identify the J-chain of IgM 84 and 85, we reduced the run time to 30 mins due to its relatively smaller molecular weight i.e. *approx.* 20kDa of this IgM domain.

3.2.5.3 Silver Staining

Gels were removed from the plastic cassette and stained using reagents supplied by SilverQuest™ Silver Staining Kit, according to the Basic Protocol as described under the instruction menu. More details can be found on website: www.invitrogen.com.

3.2.5.4 Western Blot

Protein samples were also transferred and detected using Western blot. To perform this experiment, gels were removed and transferred onto PVDF membrane using iBlot™ (*Invitrogen*) for 7 mins using iBlot® Gel Transfer Stacks PVDF. Depending on the epitopes or immunoglobulin chains, different primary, secondary antibodies and blocking agents were used as shown in Table 3.2. After blocking overnight at 4^0C, the membrane blots were washed and incubated for 5 mins with Tris-buffered Saline (TBS), 0.1% Tween 20 or TBST (20mM Tris, 500mM NaCl, 1mM $CaCl_2$, 0.1% Tween 20 pH 7.0) and this washing step was repeated four more times before incubation of primary antibodies. Again, the membranes were washed as described 5 more time before incubation of secondary antibodies, and another 5 times

[1] LDS stands for lithium dodecyl sulfate and the buffer composition for 4x LDS sample buffer – 10% Glycerol, 141mM Tris Base, 106mM Tris HCl, 2% LDS, 0.51mM EDTA, 0.22mM SERVA® Blue G250, 0.175mM Phenol Red, pH 8.5

[2] Pre-cast gel is available off-the-shelf at different gradient and loading well; choice of gel used depends on number of samples and loading volume of sample has to be adjusted appropriately. More details can be found on website: www.invitrogen.com

before adding detection agent. ECL Plus (*GE Healthcare*) was used as substrate for chemiluminescent detection of protein domains and different epitopes of N-glycans from IgM 84 and 85 using X-ray film in the dark room.

Table 3.2 Primary and secondary antibodies used in different western blots

Target	Blocking	Primary	Secondary
α-Gal	1% BA[1] in TBST (v/v) (*Sialix*)	Anti-Gal(1,3)gal (*Millipore*)	Goat anti-human IgG-HRP (*Millipore*)
Neu5Gc	1% BA in TBST (v/v) (*Sialix*)	Anti-Neu5Gc (*Sialix*)	Donkey anti-chicken IgY-HRP (*Jackson ImmunoResearch*)
J-chain	5% NFM[1] in TBST (v/v) (*Sialix*)	Anti mouse J-chain (*Santa Cruz*)	Goat anti-rabbit IgG-HRP (*Santa Cruz*)

3.2.6 Molecular weight and monomer fraction determination using SEC

100ul of IgM 84 and 85 samples (concentration: 1mg/ml) were injected onto Tosoh TSK G4000 SWXL (7.8 mm x 30 cm) at 25^{o}C on a High Performance Liquid Chromatography (HPLC) system (*Shimadzu*). Samples were run under constant flow rate of 0.6ml/min using 0.2M sodium phosphate, 0.1M potassium sulfate buffer pH 6.0. Static Light Scattering (SLS) is used in tandem to UV_{280nm} detector to measure molecular weight (MW) and hydrodynamic radius (r_H) of both IgM 84 and 85.

[1] BA stands for Blocking Agents, supplied by Sialix, Inc; NFM, on the other hand, stands for non-fat milk of Anlene

3.2.7 Mass spectrum analysis

3.2.7.1 Data Explorer

Mass spectrum data files were generated by AB SCIEX Voyager Instruments using MALDI-TOF MS or MALDI-TOF-TOF MS/MS. These data files were viewed in Data Explorer® as shown in Figure 3.2. Every mass ion displayed on a mass spectrum shows a distinct set of four peaks due to the isotopic nature of hydrogen i.e. ^{1}H and ^{2}H, and carbon ^{12}C and ^{13}C. Each of these four peaks was separated from each other by 1Da.

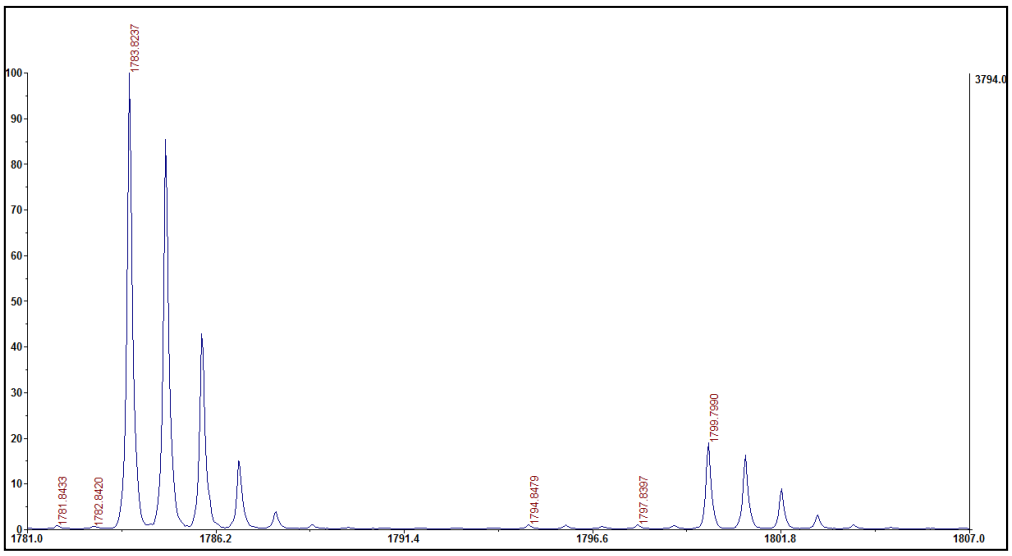

Figure 3.2 Section of mass spectrum generated by MALDI TOF MS was displayed. Y- and X-axes represent the intensity of mass ion and absolute mass (Da) respectively. Each of the two mass ions above display a distinct set of four peaks due to the presence of isotopes like ^{1}H, ^{2}H ^{12}C and ^{13}C with ^{1}H and ^{12}C being the most abundant species

Besides identifying the presence of specific mass ion, Data explorer® also provides us with information of absolute intensity of each individual peak. In identifying the absolute intensity of each mass peak, only the first or lowest mass peak was chosen but not necessarily the one with highest intensity (Figure 3.2). With absolute intensity, we then calculated the percentage relative abundance (%*RA*) of each mass ion as follows.

$$\%RA = \frac{Absolute\ intensity\ of\ specific\ mass\ peak}{Total\ absolute\ intensity\ of\ all\ mass\ peaks} x100\%$$

Besides, we categorized all the peaks in four groups – high mannose, biantennary and triantennary complex type, and hybrid. With this, we also calculated the percentage distribution (%*D*) of individual mass ion within a group as follows.

$$\%D = \frac{Absolute\ intensity\ of\ specific\ mass\ peak}{Total\ absolute\ intensity\ of\ all\ mass\ peaks\ in\ a\ group} x100\%$$

The results of the *%RA* and *%D* were tabulated in Appendix D, as shown in Table D1 and D2. However, prior to that, individual mass ion of interest need to be first identified by matching the masses of mass spectra with the theoretical masses of all N-glycans in mouse N-glycan library, which can be constructed using GlycoWorkbench.

3.2.7.2 GlycoWorkbench

Mouse N-glycan profiling library was constructed using GlycoWorkbench software. Source for these N-glycan structures were obtained from the public domain - Functional Glycomics Gateway, led by Consortium for Functional Glycomic (CFG) (http://functionalglycomics.org). CFG is a large international initiative that enables Participating Investigators to share functions of glycans and glycan-binding proteins (GBPs) that impact human health and disease. At the same time, this also serves as a knowledge base to the scientific community where one can access to the latest findings and development in study of glycans.

List of mass peaks that satisfy certain signal-to-noise ratio was exported from Data Explorer® and used to match with the theoretical masses of corresponding N-glycan structures in the library. For masses that match multiple suggested structures, MALDI-TOF-TOF MS/MS that fragmented intact mass ion further was used to further elucidate the identity of mass ion. Mass spectra of these fragment mass ions were analyzed using SimGlycan® software.

3.2.7.3 SimGlycan Enterprise Client 2.92

SimGlycan® predicts the structure of a glycan from MALDI-TOF-TOF MS/MS data. SimGlycan® matches MS/MS data generated by mass spectrometry against its own database of theoretical fragmentation of over 8,000 glycans and generates a list of probable glycan structures. The SimGlycan® has a robust database that consists of theoretical fragments of known glycan structures made up of 62 different monosaccharides. Every glycan in the database is fragmented for each of the possible fifty one reaction conditions using an intensive fragmentation algorithm. The extensive and comprehensive nature of this database ensure the high fidelity of the probable glycan structures.

The search mechanism used in SimGlycan software is based on matching algorithm. It compares the experimentally determined fragment masses with a reference set of theoretical fragment masses.

Composition score calculates how fully a suggested structure is supported by the experimental masses regardless of other suggested structures. Higher score is given to those candidate structures whose theoretical glycosidic fragment masses match maximum of the experimental mass. Composition score consists of the following percentage matches:

a) **% Glycosidic Match**: percentage match of single glycosidic and glycosidic fragments against the experimental masses
b) **% Cross ring Match:** percentage match of single cross ring and cross ring/glycosidic fragments against the experimental masses

Branching Pattern score determines the degree of closeness of one suggested structure to the real glycan relative to other suggested structures. Higher score is given to the suggested structure whose theoretical mass fragments match those of experimental masses with higher intensity or relative abundance. This score is also based on a weighted scoring system, where greatest weight is assigned to average match of glycosidic intensity and

sequentially followed by average match cross ring intensity and average match overall intensity.

Suggested structure with the highest composition score and highest branching pattern score was given the highest glycan rank, and therefore the most proximate structure for the unknown glycan subjected to MS/MS. The list of masses that were subjected for MALDI-TOF-TOF can be found in Appendix D.

3.2.8 Discovery Studio – software for homology modeling

Discovery studio is a comprehensive software suite for molecular modeling and drug discovery. We used this software to create 3D model structures for variable regions of IgM 84 and 85 using homology modeling. In Discovery Studio, creating homology models from a protein sequence uses a number of protocols which can be summarized into the following four steps:

Template identification: Potential templates that can be used for model construction can be identified using *Sequence Analysis* protocols such as *BLAST Search (DS Server)* or PSI-*BLAST Search*. If the sequence identity of target sequence is between either 25%-60% or above 60%, BLAST can identify correct templates effectively. However, if the sequence identity is below 25%, iterative searching method (PSI-BLAST) can be used instead. There are two sequence databases, PDB and PDB_nr95 available. It is most common to search templates against the non-redundant sequence database PDB_nr95[1,2].

Scoring matrix used is BLOSUM62. BLAST then compares the homology between two sequences using the following equation (Eddy, 2004)

$$s(a,b) = \frac{1}{\lambda} log \frac{P_{ab}}{f_a f_b}$$

where

P_{ab} = likelihood of test hypothesis, i.e. two residues *a* and *b* are correlated because they are homologous

f_a, f_b = likehood of null hypothesis, i.e. two residues *a* and *b* are uncorrelated and unrelated, occurring independently

[1] nr95 = 95% non-redundancy
[2] Two or more homologous proteins with sequence identities that are larger than 95% would be considered as the same, is the criterion to define different protein under PDB_nr95.

Aligning model sequence to templates: *Align Sequence to Templates* protocol aligns model sequence with selected templates based on structure similarities between the two. Better alignment can be obtained by creating a sequence profile and aligning the sequence profile to the pre-aligned structures. A good sequence profile is a sequence alignment that contains a large set of homologous, but non-redundant set of sequences.

Building 3D model using MODELER: The *Align Structures (MODELER)* protocol creates model protein structures of a target sequence based on the target-template alignment. *MODELER* treats ligands of template as rigid bodies and they are copied to the model structures as BLOCK residues. Loops are segment of sequences on the model that can be further refined within the protocol itself. This protocol calculates and returns two scores - PDF Total Energy[1] or Physical Energy[2], and DOPE (Discrete Optimized Protein Energy) for each model that it builds for evaluating the quality of the models. A model is better optimized against the homology restraints when it has lower PDF Total Energy and models with the least violation to homology restraints are preferred. However, if the models all have similar PDF Total Energy, you can use the DOPE score which is based on statistical potential as a measure of the model quality i.e. the lower the score, the better the model.

Assessing validity of the 3D structure: The *Verify Protein (Profiles-3D)* protocol allows you to evaluate the fitness of a protein sequence in its current 3D environment. It can be applied to assess the quality of a theoretical model or to examine the characteristics of an experimental structure. For example, it can be used to find hydrophobic patches on the surface of a structure. More hydrophobic patches on the surface of a protein, except for membrane proteins, would result in lower score as they tend to reside within the core in tertiary or quaternary structures. The protocol returns Verify Score[3] for the protein, together with

[1] PDF stands for probability density function; PDF Total Energy is the sum of the scoring function value of all homology-derived pseudo-energy terms and stereochemical pseudo-energy terms
[2] Sum of energies of the stereochemical pseudo-energy terms which consist of valence bonds, valence angles and torsion angles, improper torsion angles, and soft-sphere repulsion, as well as knowledge based non-bonded potentials used only for loop and mutant modeling
[3] Sum of the scores of all residues in the protein

Expected High Score[1] and Expected Low Score[2]. If the model structure has a Verify score higher than the expected high score, the structure is likely to be correct. If the overall quality score is between the reference values, then some or all of the structure may be incorrect, and it requires closer scrutiny. If the overall quality is lower than the expected low score, then the structure is almost certainly misfolded.

Last but not least, using *Align and Superimpose Proteins* protocol, we also performed a structural superimposition of the 3D model structures for variable regions of IgM 84 and 85 and calculated the root mean square difference (RMSD) to quantify their spatial differences.

[1] Statistical analysis of high-resolution structures in the Protein Data Bank (PDB)
[2] 45 percent of the high score and is typical of grossly misfolded structures having this sequence length

4 RESULTS AND DISCUSSION

4.1 Characterization of protein IgM 84 and 85

4.1.1 Physical properties of IgM 84 and 85 using SEC-HPLC/SLS

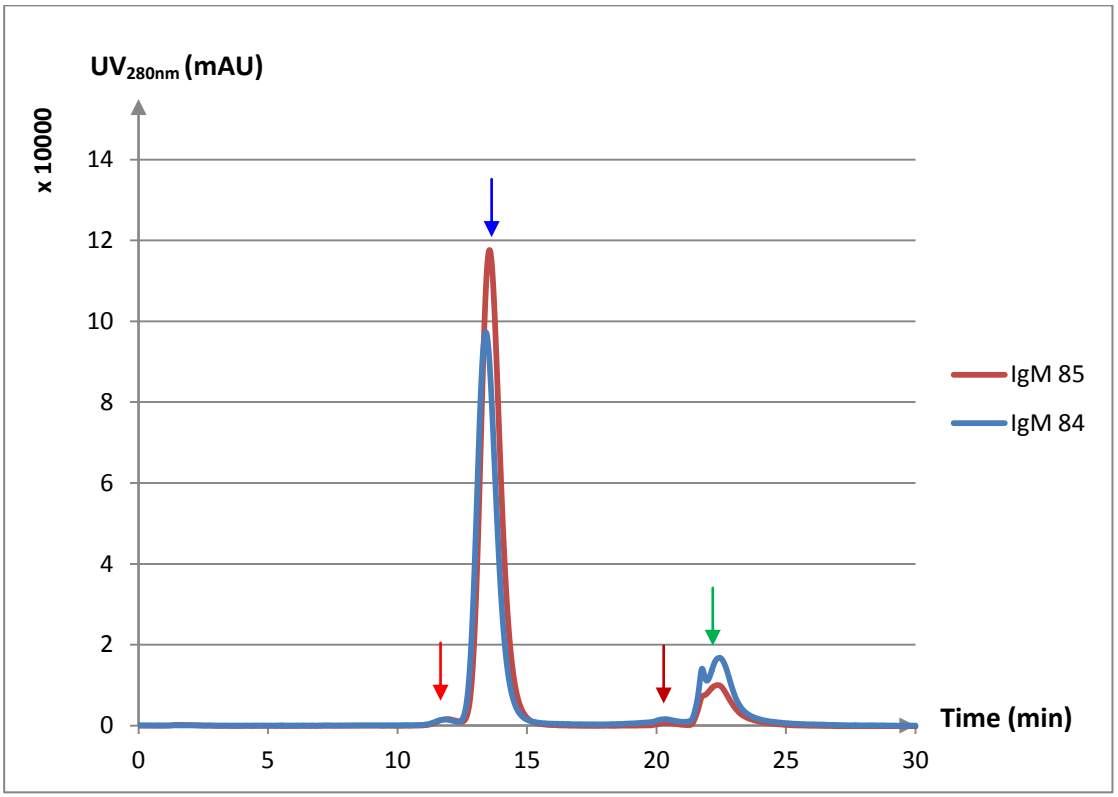

Figure 4.1 SEC-HPLC UV_{280nm} of IgM 84 and 85

Size exclusion chromatography – high performance liquid chromatography (SEC-HPLC) was used in tandem with static light scattering (SLS) detector to characterize a few physical properties of protein IgM 84 and 85 i.e. hydrodynamic radius (r_H), molecular weight (MW) percentage population of IgM aggregates, pentamers and fragments. This analysis is particularly crucial to determine the quality of the purified IgM because presence of large amount of IgM aggregates may render glycosylation analysis more difficult to interpret and more complex during preparation steps mentioned in Chapter 3 thus affecting the quality of our results in these experiments.

SEC separates protein molecules primarily on r_H or size of protein, which in this case are IgM aggregates, pentamers and fragments. IgM aggregates, being the largest in size, were excluded from the pores of chromatographic column and therefore eluted first at 11.82 min ± 0.12%[1] (IgM 84) and 11.93 min ± 0.77% (IgM 85) (as indicated by red arrow in Figure 1); while IgM pentamers eluted at 13.40 min ± 0.00% (IgM 84) and 13.55 min (IgM 85) (as indicated by blue arrow in Figure 1). Fragments of IgM 84 and 85 were eluted close to but before the buffer/salt peaks (as indicated by green arrow Figure 4.1) at 20.33 min ± 0.00% and 20.39 min ± 0.21%, respectively (as indicated by brown arrow Figure 4.1). Samples of purified IgM 84 and 85 in final storage buffer contained high percentage of IgM pentamers, and low levels of aggregates (Table 4.1). The percentages of IgM 84 and 85 fragments could not be reliably determined due to their closeness and/or overlapping with buffer/salt peak intensities (Figure 4.1). However, they account for not more than 1% of the total amount of IgM. Using Static Light Scattering (SLS), MW and r_H of IgM 84 and 85 determined and they were found to be rather similar (Table 4.1).

Table 4.1 Physical properties of IgM 84 and 85 determined using SEC-HPLC/SLS

IgM	IgM aggregate (%±RSD)	IgM pentamer (%±RSD)	Molecular weight (kDa±RSD)	r_H by SLS[1] (nm)
IgM 84	1.30 ± 4.35%	97.83 ± 0.22%	889.00 ± 0.10%	14.05 ± 0.50%
IgM 85	1.37 ± 2.59%	98.25 ± 0.03%	888.65 ± 0.22%	14.45 ± 1.47%

4.1.2 Sequence analaysis of IgM 84 and 85

4.1.2.1 Identifying N-glycosylation sites on IgM 84 and 85

Part of the amino acid sequences of IgM 84 and 85 can be found in Appendix A. Six potential N-glycosylation sites were identified in heavy μ chain constant regions (Cμ1 to Cμ4) of both IgM 84 and 85 (Table 4.2) . Interestingly, one additional potential N-glycosylation site

[1] In the representation of our results, standard deviation (SD) was expressed as %RSD or percentage of relative standard deviation instead of absolute numbers because we find that reproducibility of our results can be more easily seen using this approach.

was also identified in the light chain constant regions of both antibodies. This site was found at the tail-end of the light chain bearing the consensus amino acid sequence of Asn-Xaa-Cys (N-X-C), which has not been reported for IgM so far (Table 4.2). Comparing the relative positions of these N-glycosylation sites, there is a side shift of 2 amino acids for N-glycosylation sites on the heavy chain; and 1 amino acid for that on the light chain between IgM 84 and 85. This is because IgM 84 has 2 amino acids less in the heavy chain variable regions and 1 amino acid less in the light chain variable regions than that of IgM 85. This observation is evident by the presence of gaps in the sequence alignment results shown in Appendix A.

Table 4.2 Potential N-glycosylation sites of IgM 84 and 85

IgM	N-glycosylation sites
IgM 84 • Heavy chain • Light chain	Asn-160, Asn-254, Asn-325, Asn-357,Asn-372, Asn-395 Asn-211
IgM 85 • Heavy chain • Light chain	Asn-158, Asn-252, Asn-323, Asn-355,Asn-370, Asn-393 Asn-212

4.1.2.2 Sequence alignment of IgM 84 and 85

Table 4.3 shows the percentage sequence similarity resulting from sequence alignment of IgM 84 and 85. The constant regions of IgM 84 and 85 were found to be 99.80% identical, except that IgM 84 has a threonine at position 291 whereas IgM 85 has a serine at position 289 on their respective heavy chains. The full sequence alignment results also showed that there are 3 different gaps observed in IgM 84 due to the different lengths of IgM 84 and 85. Major differences between the two sequences lie in the variable regions of both heavy and light chains that is, sequence similarities are 51.30% (heavy), 64.82% (light) and 57.80% (both heavy and light chains) (Table 4.3).

Table 4.3 Sequence similarities between IgM 84 and 85 constant and variable regions

Domains of IgM	Sequence similarity
Full length IgM (monomer)	**87.60%**
Heavy (μ) chain	89.70%
• Constant regions	99.77%
• Variable regions	51.30%
Light (κ) chain	82.20%
• Constant regions	100.00%
• Variable regions	64.82%
Constant regions	**99.80%**
Variable regions	**57.80%**

4.2 Characterization of the N-glycans of IgM 84 and 85

4.2.1 Global N-glycan profiling

In general, we found three main types of N-glycan on IgM 84 and 85 – high mannose, complex and hybrid types that we categorize them in four groups as shown in Table 4.4. Table 4.4 summarized the percentage relative abundance (%RA) of high mannose, biantennary complex, triantennary complex, and hybrid types and showed only the top most abundant N-glycan species within each group in terms of percentage distribution.

As an overview, both IgM 84 and 85 share some similarities in terms of percentage relative intensities – high mannose type N-glycans being the most abundant N-glycan types i.e. from 67.3% to 82.5%, followed by complex type i.e. from 11.6% to 27.7%[1] and hybrid the least i.e. from 5.3% - 7.3% (Table 4.4). In addition to that, according to percentage distribution, $GlcNAc_2Man_6$, $GlcNAc_2A_2G_3S_1$, $GlcNAc_2A_3G4S_1$ and $GlcNAc_2A_1Man_5G_1S'_1$ (see note of Table 4.4 for nomenclature) are among the most abundant N-glycan species within the respective N-glycan groups of both IgM 84 and 85. On the other hand, $GlcNAc_2A_2G_2S'_1$ was found to be the most abundant species in IgM 84 only; whereas $FcGlcNAc_2A_2$, $FcGlcNAc_2A_2G_3S'_1$, $GlcNAc_2A_3G_5S_1$ and $FcGlcNAc_2A_3G_3S'_2$ were observed to be the top most abundant species in IgM 85 only within the respective groups. With this information, we then explored further to elucidate the differences between IgM 84 and 85 in terms of the N-glycan types present in each of them.

[1] These figures are calculated by adding %relative intensities of biantennary and triantennary complex types together to reflect the total for complex type N-glycans

Table 4.4 Summary of differences between IgM 84 and 85 in terms of percentage relative abaundance (%RA) of four main groups of N-glycan and their percentage distributions (%D) within each group. Full list of N-glycan masses, structures, %RA and %D can be found in Appendix D.

		IgM 84		IgM 85	
N-glycan masses	**N-glycan structures**	**Run 1**	**Run 2**	**Run 1**	**Run 2**
1. High mannose					
	%RA	**70.4%**	**82.5%**	**67.3%**	**66.9%**
	%D:				
1783.88	$GlcNAc_2Man_6$	67.0%	69.6%	71.1%	79.2%
-	Others	0.3 - 16.8%	0.3 - 15.8%	1.6 - 20.7%	1.1 - 16.0%
2a. Biantennary complex type					
	%RA	**13.5%**	**8.4%**	**17.6%**	**20.5%**
	%D:				
1835.93	$FcGlcNAc_2A_2$			14.1%	17.9%
2461.22	$GlcNAc_2A_2G_2S'1$	12.3%	14.7%		
2635.31	$GlcNAc_2A_2G_3S_1$	15.7%	17.0%	21.2%	25.8%
2839.41	$FcGlcNAc_2A_2G_3S'_1$			13.4%	15.4%
-	Others	0.7 - 6.6%	0.6 - 6.8%	0.9 - 6.4%	0.6 - 4.1%
2b. Triantennary complex type					
	%RA	**8.8%**	**3.2%**	**8.9%**	**7.2%**
	%D:				
3084.54	$GlcNAc_2A_3G4S_1$	13.3%	16.4%	13.3%	13.2%
3288.64	$GlcNAc_2A_3G_5S_1$			13.7%	21.0%
3475.72	$FcGlcNAc_2A_3G_3S'_2$			9.2%	10.5%
-	Others	1.0 - 8.8%	0.0 - 13.2%	1.0 - 6.2%	1.2 - 9.3%
3. Hybrid					
	%RA	**7.3%**	**5.9%**	**6.2%**	**5.3%**
	%D:				
2420.19	$GlcNAc_2A_1Man_5G_1S'_1$	46.7%	54.5%	55.3%	51.8%
-	Others	2.3 - 21.3%	0.9 - 19.8%	2.3 - 21.3%	0.9 - 19.8%
	Total %RA	**100.0%**	**100.0%**	**100.0%**	**100.0%**

Note: A_1, A_2 and A_3 represent trimannosyl core with one, two and three GlcNAc sugar units (or antenna), respectively; whereas Fc, G, S, S' represent fucose, galactose, Neu5Ac and Neu5Gc sugar units, respectively; and B represents a bisecting GlcNAc sugar that is attached β1-4 to A trimannosyl core. Subscript of each sugar shows the number of sugar units that are attached.

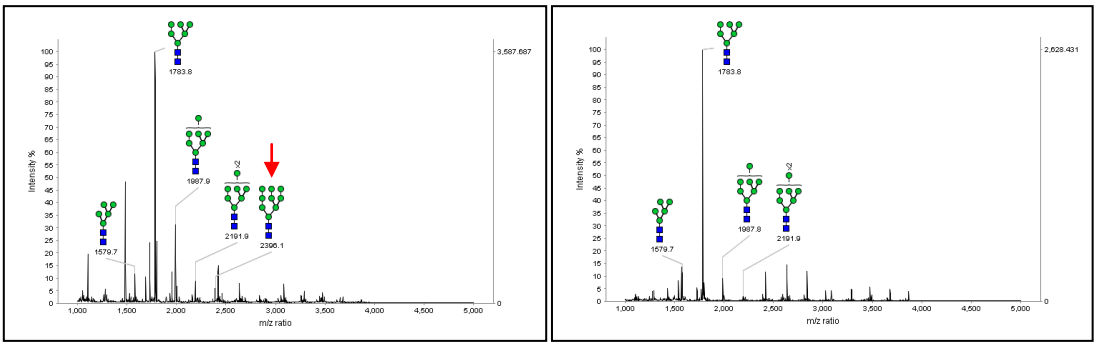

Figure 4.2A High mannose N-glycan types on IgM 84 (left) and IgM 85 (right)

Figure 4.2A shows the presence of high mannose type N-glycans on IgM 84 and IgM 85. One difference between IgM 84 and 85 is the presence of $Man_9GlcNAc_2$ in IgM 84 (indicated by red arrow), which is absent in IgM 85. In the biosynthetic pathway of N-glycans, high mannose N-glycan types are typically trimmed down to $Man_5GlcNAc_2$ in *cis*-Golgi before a GlcNAc sugar residue is added to form complex N-glycan types. Hence, the presence of $Man_9GlcNAc_2$ suggests that the trimming processes of high mannose type N-glycans in IgM 84 are less mature than that in IgM 85. This may therefore result in a different protein conformation of IgM 84 compared to IgM 85 because the trimming process of $Man_9GlcNAc_2$ to $Man_8GlcNAc_2$ occurs inside the Endoplasmic reticulum (ER) and N-glycosylation in ER plays an important role in how a protein is folded prior to the exit (Stanley et al., 2009).

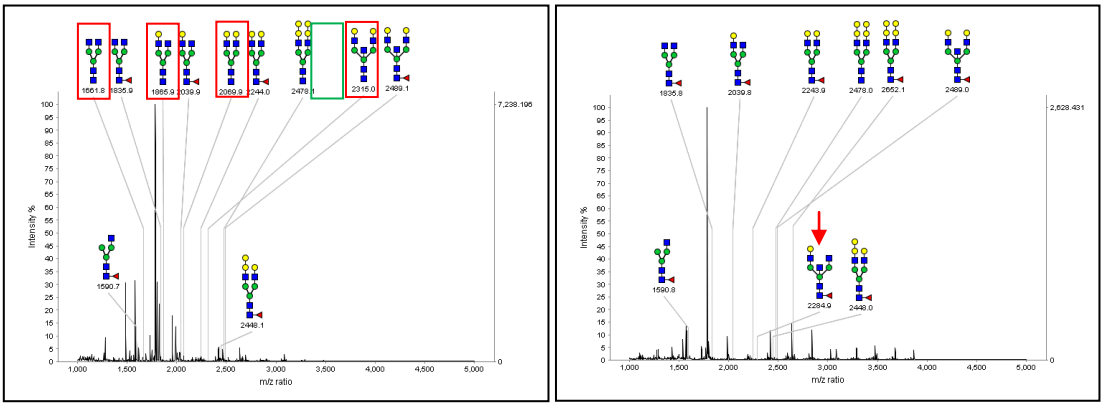

Figure 4.2B Asialylated biantennary complex N-glycan types on IgM 84 (left) IgM 85 (right)

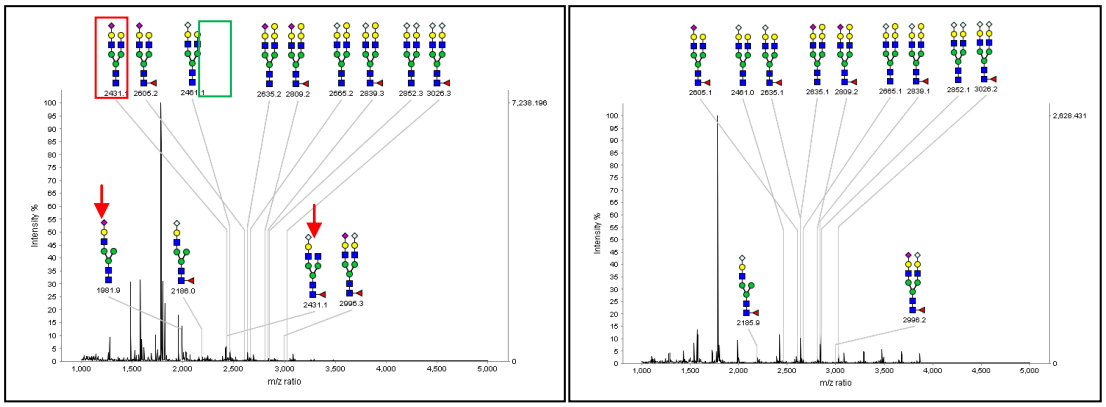

Figure 4.2C Sialylated biantennary complex N-glycan types (biantennary) on IgM 84 (left) IgM 85 (right)

Biantennary complex N-glycan types were categorized into asialylated groups in Figure 4.2B, and sialylated groups in Figure 4.2C. For ease of comparison, we paired up any two complex N-glycan types that differ by only one fucose that is attached to the proximal GlcNAc of the core structure via α(1-6) linkage, which resulted in 10 structure pairs i.e. 5 pairs in Figure 4.2B and 5 pairs in Figure 4.2C. These 10 structure pairs were presented at the top of both figures, while other structures were presented below, or nearer to the x-axes of mass spectra. Once this pairing comparison was done, it became clear that there is a major difference between IgM 84 and IgM 85 in fucosylation, which is one of the maturation steps in the biosynthesis of N-glycans. Overall, the N-glycans of IgM 84 are less fucosylated than those of IgM 85 as observed by the presence of five non-fucosylated biantennary complex N-glycan types in IgM 84 (red boxes in Figures 4.2B and 4.2C), which are not present in IgM 85. Moreover, IgM 85 has two fucosylated biantennary complex N-glycan types (green boxes in Figures 4.2B and 4.2C), which are absent in IgM 84. The incomplete fucosylation in IgM 84 could be explained by a difference in protein conformation of IgM 84 entering the Golgi apparatus that causes the N-glycosylation sites on IgM 84 bearing these complex N-glycan types to be shielded from enzymes that add fucose in the *trans*-Golgi. Besides, we observed three unique N-glycan types – one bisecting complex N-glycan type in IgM 85 only, and two sialylated biantennary complex N-glycan types in IgM 84 only (as indicated by red arrows in Figures 4.2B and 4.2C).

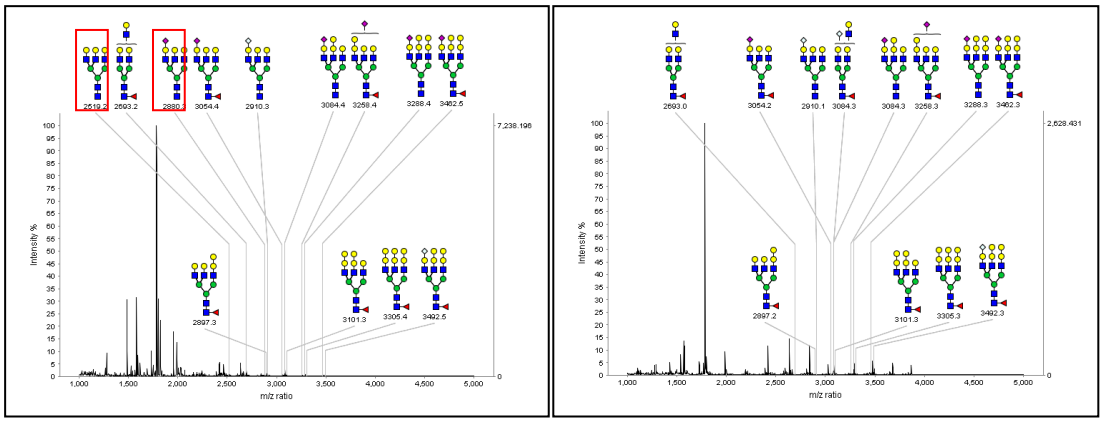

Figure 4.2D Asialylated and monosialylated triannttennary complex N-glycan types on IgM 84 (left) and IgM 85 (right)

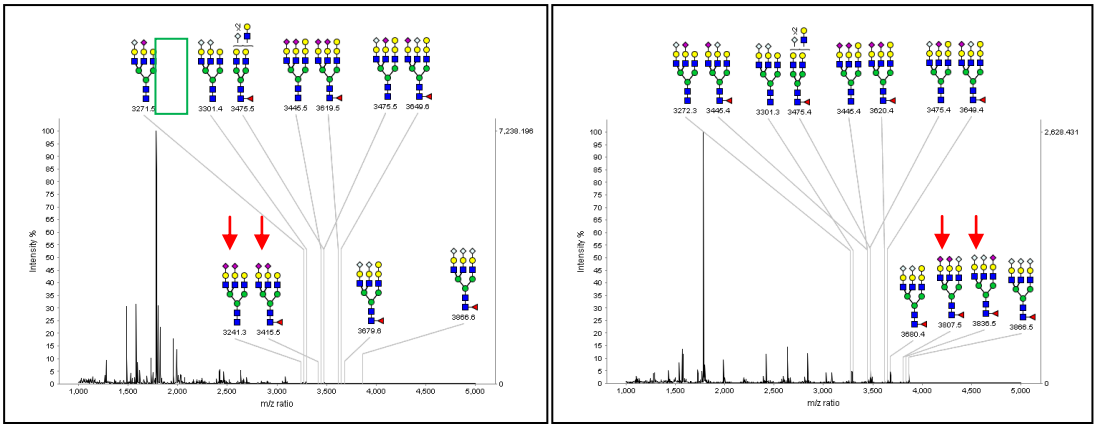

Figure 4.2E Disialylated and trisialylated triantennary complex N-glycan types on IgM 84 (left) and IgM 85 (right)

Figures 4.2D and 4.2E show the asialylated, mono-, di- and trisialylated triantennary complex N-glycan types of IgM 84 and 85.The triantennary complex N-glycans of IgM 84 appear to be less fucosylated compared to those of IgM 85, similar to what was observed for biantennary complex N-glycan types. This claim was substantiated by the presence of two non-fucosylated triantennary complex N-glycan types (red boxes in Figure 4.2D) and absence of one fucosylated triantennary complex N-glycan type (green box in Figure 4.2E) in IgM 84, when compared with IgM 85. Besides, there are three unique triantennary complex N-glycan types – one disialylated triantennary complex N-glycan type in IgM 84 only, and two trisialylated triantennary complex N-glycan types in IgM 85only (as indicated by three red arrows in Figures 4.2D and 4.2E).

4.2.1.1 Detection of immunogenic glyco-epitopes

As shown in Figures 4.2B – 4.2E, two glyco-epitopes, α-Gal and Neu5Gc that are immunogenic to human (Galili, 2005; Padler-Karavani et al., 2008) were observed to be present in some N-glycans on IgM 84 and 85. We performed Western blot and confirmed the presence of these epitopes (Figure 4.3). Protein bands (between 80kDa and 100kDa) most likely correspond to the full heavy μ chains of IgM 84 and 85, while bands (between 60kDa and 80kDa) are probably degraded heavy μ chain fragments. Also, light chains of IgM 84 and 85 were not detected in the following Western blot which indicates the absence of alpha gal and Neu5Gc epitopes on the light chains of the IgM. Interestingly, we also discovered an extra band (red arrow) that seems to suggest the presence of complex type N-glycan bearing Neu5Gc on the J-chain as confirmed on the other western blot using anti-J chain antibody (Figure 4.3 (right)).

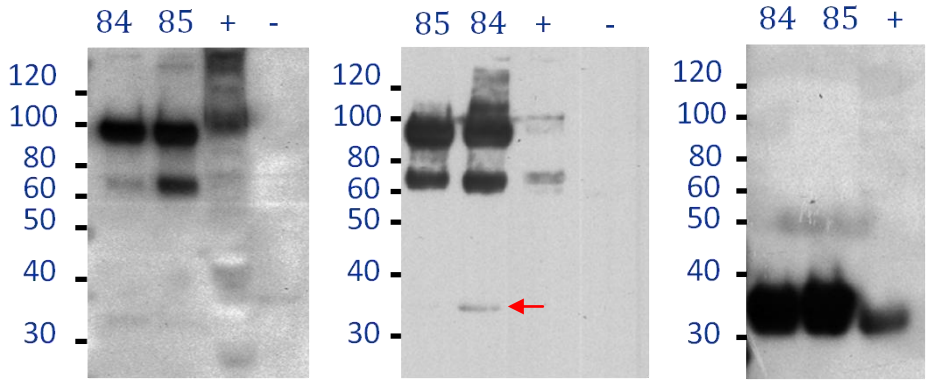

Figure 4.3 Western blots that detect presence of α-Gal (left), Neu5Gc (middle) and J-chain (right) in both IgM 84 and 85. Numbers on the left indicate the molecular weights of the standards (in kDa). Positive and negative controls are listed in Table 4.5 below

Table 4.5 Positive and negative controls used in Western blot to detect glyco-epitopes of IgM 84 and 85

	Positive control (+)	Negative control (-)
α-Gal	Protein extract from mouse heart	Ribonuclease B (*Sigma Aldrich*)
Neu5Gc	Fetuin (*Sigma Aldrich*)	Ribonuclease B (*Sigma Aldrich*)
J-chain	Protein extract from mouse spleen (*Santa Cruz*)	-

4.2.1.2 Sialylation of IgM84 and 85

From the results of N-glycan global profiling (Figure 4.2B – 4.2E), we observed the presence of asialylated, monosialylated, disialylated and trisialylated complex type N-glycans. In order to compare their relative amounts, we separated the complete pool of released N-glycans by HPAEC-PAD. Thus, we obtained several population of N-glycans based on their extent of sialylation.

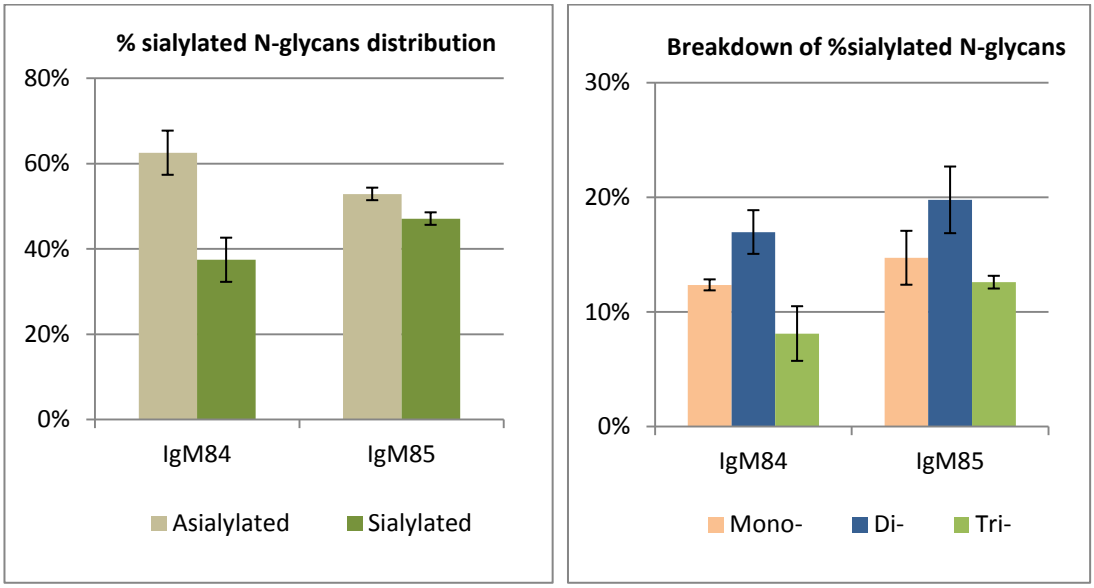

Figure 4.4 Percentage of asialylated and sialylated N-glycans (left) and distribution of mono-, di-, and trisialylated N-glycans within the sialylated N-glycans pool (right) of IgM 84 and 85

Figure 4.4 (left) showed that the percentage of sialylated N-glycans on IgM84 is lower than that on IgM 85 i.e. 37.5 ± 5.2% (IgM 84) versus 47. 1 ± 1.5% (IgM 85). To view the

contribution of different sialylated N-glycans, we plotted another bar chart (Figure 4.4 (right)) which showed that disialylated complex N-glycan types are the main contributors in both IgM 84 and 85, followed by monosialylated and trisialylated N-glycans. These results were based on the same amount of protein (i.e. 100μg) rather than the total sialylated N-glycans. In other words, to better cross-compare the relative contribution of each type of sialylated N-glycans, we normalized the individual types of %sialylated N-glycans in Figure 4.4 (right) by the total %sialylated N-glycans in Figure 4.4 (left). It was therefore observed that, for the same amount of total sialylated N-glycans, IgM 84 has more disialylated and less trisialylated than that of the IgM 85. Monosialylated N-glycans are almost the same in either sample (Figure 4.5).

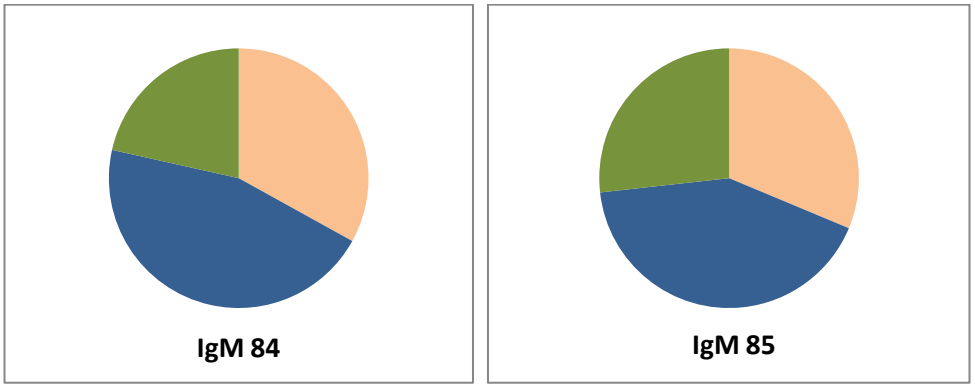

Figure 4.5 Breakdown of %sialylated N-glycans distribution of IgM 84 and 85. Color codes: green – trisialylated, blue – disialylated, peach – monosialylated

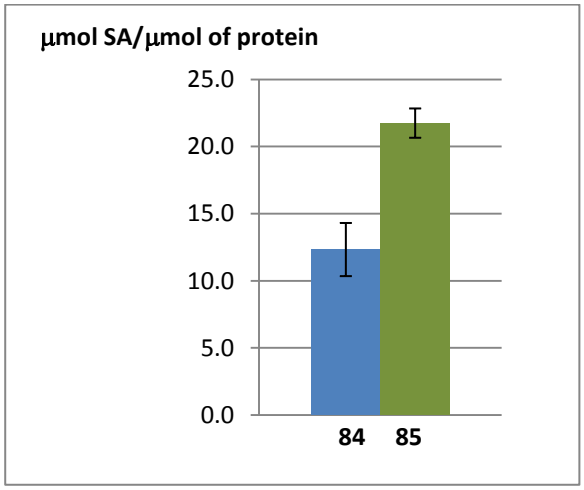

Figure 4.6 Total sialic acid content [μmol SA/μmol of protein] of IgM 84 and 85

The total amount of sialic acid were quantified using a high throughput method recently developed by researcher at MIT in collaboration with our group (Raga et al., 2010). Results obtained in Figure 4.6 are average of three independant runs. In Figure 4.6, it is observed that total sialic acid content of IgM 85 is higher than that of IgM 84 by almost twice in absolute amount i.e. 21.74 ± 1.09 μmol/μmol IgM 85 and 12.32 ± 1.98 μmol/μmol for IgM 84 (means of 3 independant experiments).It is also worth noting that the relative contribution of sialic acid types i.e. Neu5Gc and Neu5Ac is quite different in each of IgM 84 and 85. However, these results require further examination to conclude.

These results showed that the sialylation – one of the late maturation steps in biosynthetic pathway of N-glycans occuring in the *trans*-Golgi cisternae of IgM 84 may be less mature compared to IgM 85 due to lower percentage of total sialic acids (Figure 4.6) and relatively lower percentage of sialylated N-glycans (Figure 4.4 (left)). In addition, relatively more disialylated N-glycans (blue pie in Figure 4.5) and lesser trisialylated N-glycans (green pie in Figure 4.5) of IgM 84 may also indicate less maturation processing of N-glycans on IgM 84.

4.2.2 Microheterogeneity

4.2.2.1 Site-specific N-glycan profiling

To identify the site-occupancy and type of N-glycans present at the different N-glycosylation sites of IgM 84 and 85, IgM were digested into peptides and glycopeptides mixtures and separated using Sep-Pak® column. The interesting glycopeptides were further identified using MALDI-TOF/MS and MALDI-TOF-TOF/MS-MS. The full profiles of glycopeptides, peptides and amino acids after trypsin digestion can be found in Appendix D. In our first trial, we were able to identify one glycopeptide ion, T36 with sequence of IMESHFN$_{395}$GTFSAK^{+} for IgM 84 bearing high mannose N-glycan types. To verify this, we identified four main mass peaks (red dotted lines) with a mass difference of 162Da, which corresponds to the mass of an hexose. The masses of these four main peaks (Table 4.6) match the calculated masses of the suggested structures.

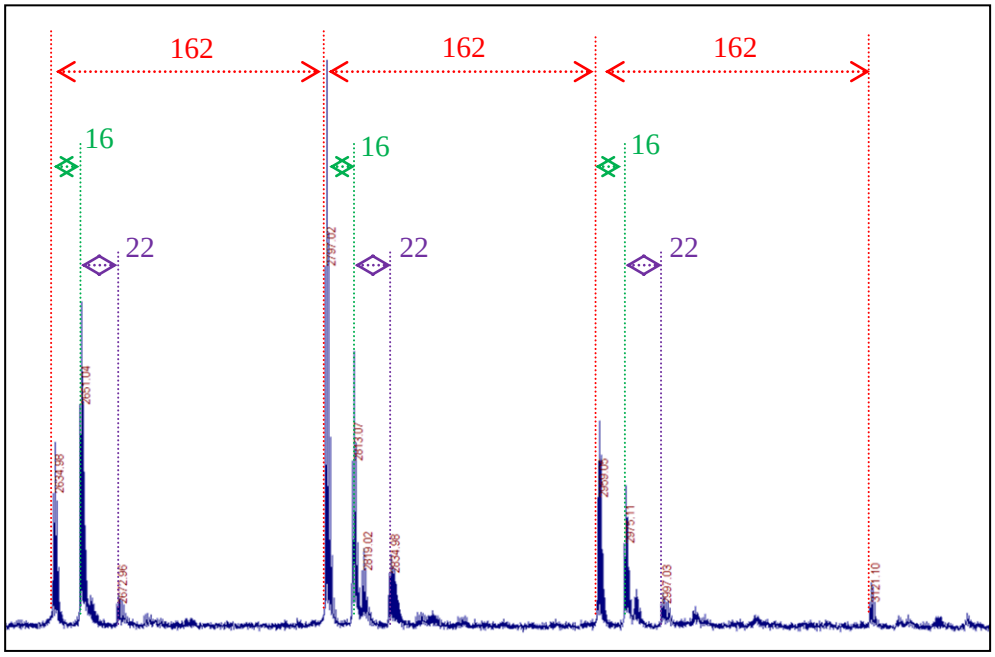

Figure 4.7A MALDI-TOF (MS) of T36 glycopeptides of IgM 84

Table 4.6 Masses of four main peaks (red dotted lines)

Mass Peak	Mass (Da)	Structure
1	2634.98	T36-$Man_5GlcNAc_2$
2	2796.83	T36-$Man_6GlcNAc_2$
3	2958.85	T36-$Man_7GlcNAc_2$
4	3120.88	T36-$Man_8GlcNAc_2$

Other subpeaks adjacent to the main peaks serve as a confirmation that these main mass peaks are indeed glycopeptide ions. For instance, the mass difference of 16Da indicates the extra mass added to peptide T36 due to oxidation of methionine, and 22Da indicates adduct Na^+ ion of glycopeptides. Further fragmentation of these peaks yielded mass peaks of the fragments of the suggested structure. When mass peak #2 (highest intensity in Figure 4.7A) was subjected to MS-MS, we indeed observed mass peaks that correspond to the fragments of the suggested structures (brown dotted lines in Figure 4.7B). Thus we could confirm the presence of high mannose N-glycan types $Man_{5-8}GlcNAc\text{-}N_{395}$ on glycopeptide T36 of IgM 84.

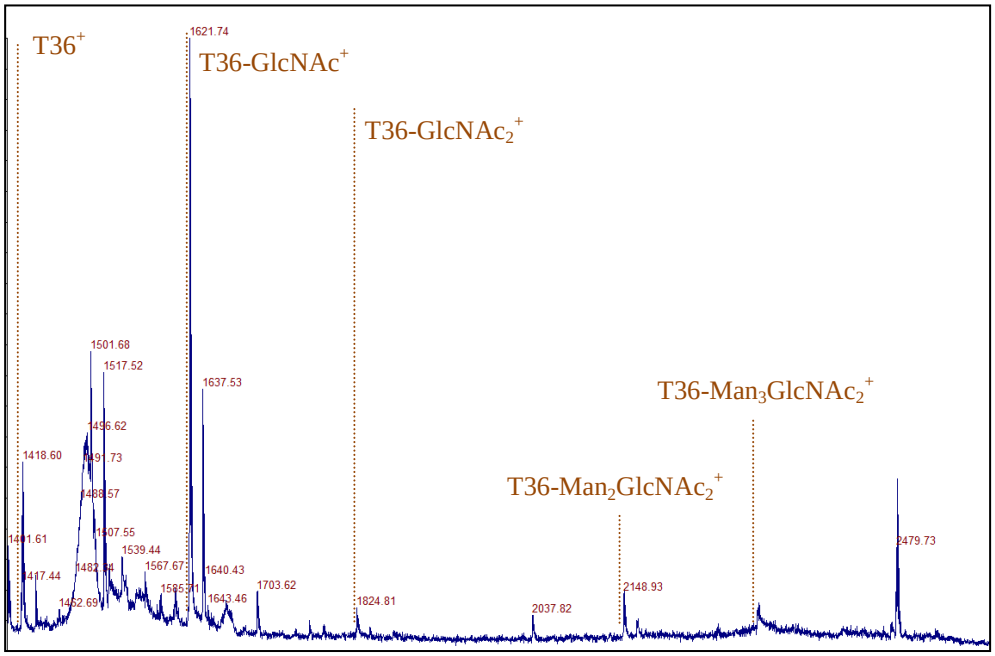

Figure 4.7B MALDI-TOF-TOF (MS/MS) of T36 glycopeptides of IgM 84

4.3 Comparative modeling of IgM84 and 85

4.3.1 Template identification

Using Discovery Studio, we use BLAST against DS Server to search for templates of variable regions of IgM 84 and 85 for individual heavy and light chains. Templates with highest bit scores and lowest E-values (Table 4.7) were selected for sequence alignment. The selected protein templates were found to originate from mouse. Although IgM 84 and 85 are produced in mouse hybridoma, these results were simply coincident and not part of the selection criteria, as earlier mentioned in Section 2.6.

Table 4.7 Template identified with highest bit score and lowest E-value for each of the variable regions of IgM 84 and 85

No	Target	Template	Organism	Bit score	E-value
1	IgM 84_V_H	1SM3_H	House mouse	193.4	$6.5x10^{-51}$
2	IgM 85_V_H	1NLD_H	House mouse	204.1	$3.8x10^{-54}$
3	IgM 84_V_L	1AY1_L	House mouse	186.0	$9.5x10^{-49}$
4	IgM 85_V_L	2FGB_A	House mouse	202.2	$1.3x10^{-53}$

4.3.2 Sequence alignment

We next re-aligned the protein sequences with their corresponding templates. Table 4.8 shows that the target-template sequence alignment has a high percentage of sequence identity and similarity between individual target and template.

Table 4.8 Target-template sequence alignment results for each of the variable regions of IgM 84 and 85

No	Target	Template	Sequence identity	Sequence similarity
1	IgM 84_V_H	1SM3_H	94.0%	94.9%
2	IgM 85_V_H	1NLD_H	86.1%	91.3%
3	IgM 84_V_L	1AY1_L	91.6%	97.2%
4	IgM 85_V_L	2FGB_A	93.5%	97.2%

4.3.3 Model building

Using MODELER, five models were created for each target-template aligned sequence. Each of these models returned a different value for PDF Total and Physical Energy and DOPE. The model with the lowest PDF energies and DOPE Score (most negative) suggests the best created model for a particular target (Table 4.9).

Table 4.9 Best models for variable regions of IgM 84 and 85 based on lowest PDF energies and DOPE Score

No	Target	Template	PDF Total Energy	PDF Physical Energy	DOPE Score
1	IgM 84_V_H	1SM3_H	518.1	315.9	-11257.4
2	IgM 85_V_H	1NLD_H	569.7	328.8	-11605.5
3	IgM 84_V_L	1AY1_L	554.9	318.2	-9522.7
4	IgM 85_V_L	2FGB_A	563.0	322.32	-10136.2

Heavy variable regions (V_H) of IgM84 and 85

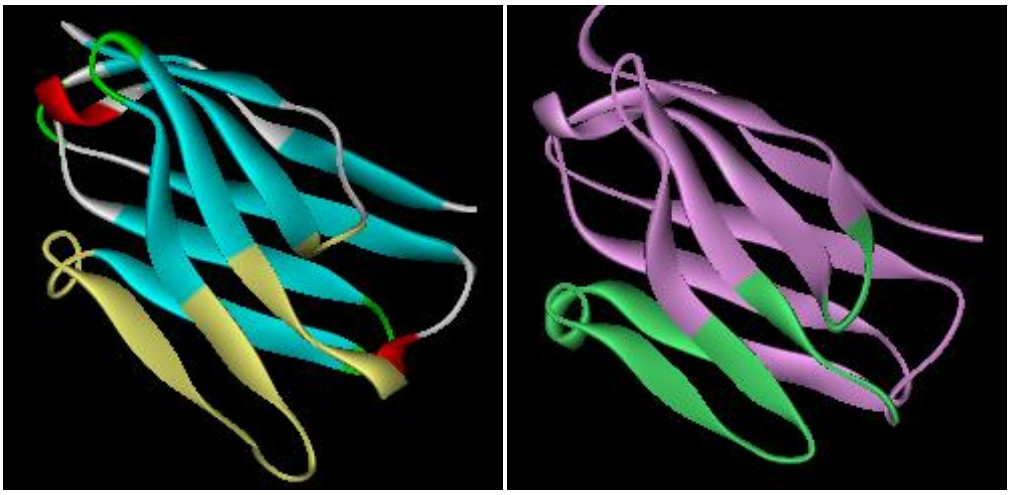

Figure 4.8 IgM 84_V_H (left) and IgM 85_V_H (right)

Light variable regions (V_L)of IgM84 and 85

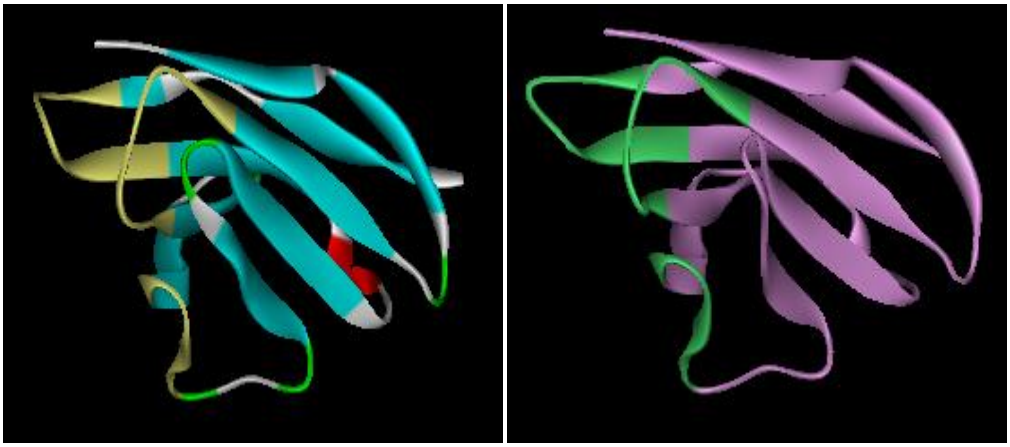

Figure 4.9 IgM 84_V_L (left) and IgM 85_V_L (right)

Figure 4.8 shows the best models for IgM 84 and 85 variable heavy chain domains. The portions, which are highlighted in yellow (left of Figure 4.8) and in green (right of Figure 4.8), depict the Complementarity Determining Regions (CDRs) of IgM 84 and 85 variable heavy chains, respectively. Similarly, we have identified the CDRs for IgM 84 and 85 variable light chains using the same color codes, as shown in Figure 4.9.

4.3.4 Model Validation

To verify the reliability of these models, we performed Verify Protein (Profiles-3D) protocol under Discovery Studio, and Molprobity, open and web-based validation software (http://molprobity.biochem.duke.edu/).

4.3.4.1 Verify Protein (Profiles-3D)

Table 4.10 Verify scores for the best model of each target sequence of IgM 84 and 85

No	Model	Verify Expected Low Score	Verify Score	Verify Expected High Score
1	IgM84_V_H	23.2	53.6	51.5
2	IgM85_V_H	23.2	55.4	51.5
3	IgM84_V_L	21.7	52.6	48.3
4	IgM85_V_L	21.7	48.4	48.3

In order for a model to be reliable, the calculated *verify score* for each model has to be higher than the *verify expected high score*, and much higher than the *verify expected low score* that are generated when calculating the *verify score*. According to these specifications, our models were reliable (Table 4.10).

4.3.4.2 Molprobity (Ramachandran Plot)

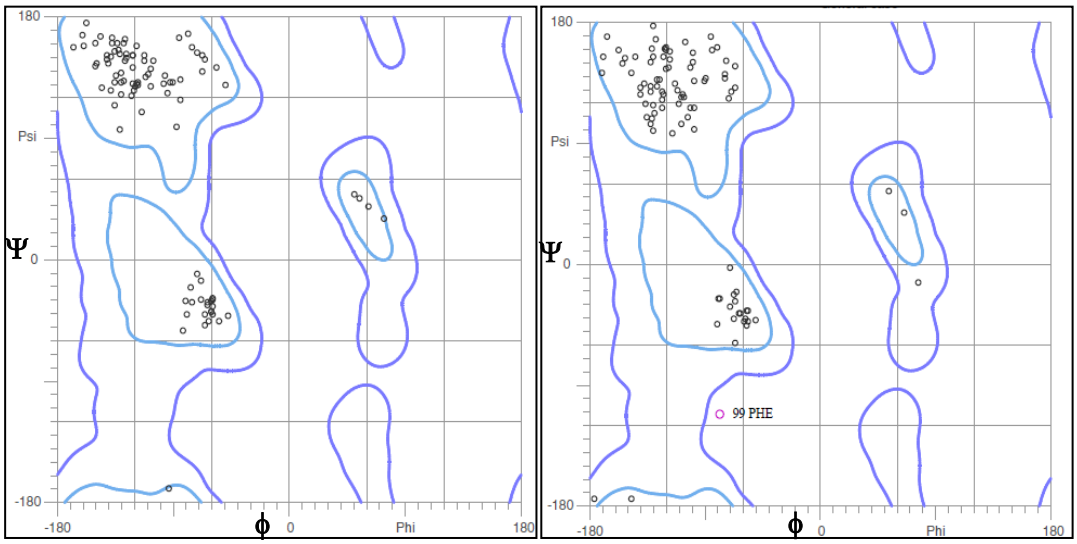

Figure 4.10 Ramachandran Plots for IgM 84_V_H (left) and IgM 85_V_H (right)

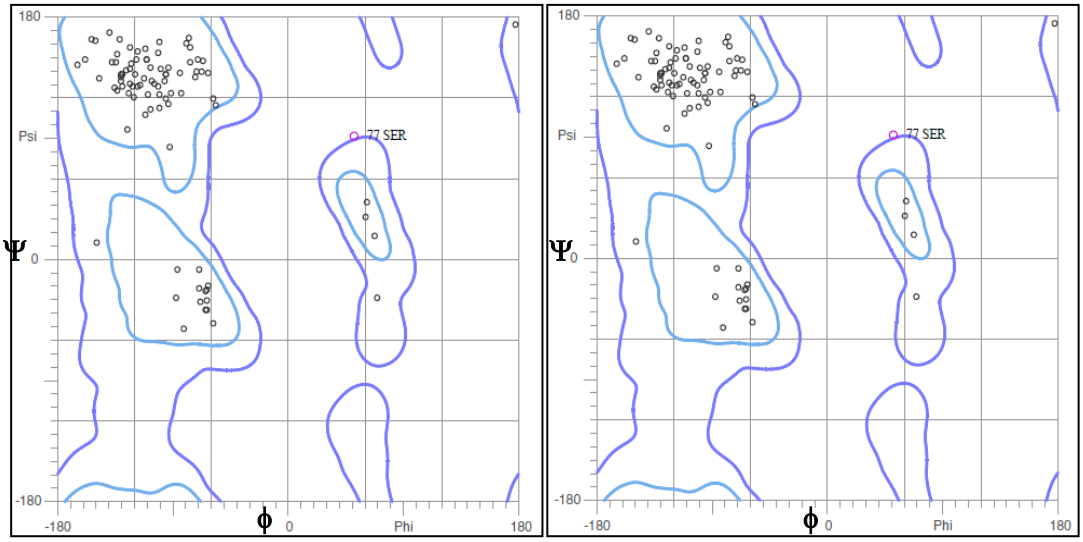

Figure 4.11 Ramachandran Plots for IgM 84_V_L (left) and IgM 85_V_L (right)

Ramachandran plots measure and visualize the dihedral angles Ψ and ϕ of amino acid residues in protein backbones. In order for a protein model structure to be considered reliable, most of the amino acid residues (indicated by open circles in figures 4.10 and 4.11) are required to fall within the favored regions (outlined by light blue lines) and few within the allowed regions (outlined in purple lines). However, if too many of the amino acid residues are found outside of these two regions, they might be under too much stress as they are forced to follow the template 3D structure. In our case, all except one single amino acid Phe_{99} of IgM

85_V_H, Ser_{77} of IgM 84_V_L and IgM 85_V_L, fall within the two regions we just mentioned, which is consistent with the results obtained from Verify Protein (Profiles-3D) protocol.

4.3.5 Model superimposition

To compare the structural differences between variable regions of IgM 84 and 85, we superimposed their structures using *Align and Superimpose Proteins* protocol and results are shown in Figure 4.12. The main differences lie in the four loop regions – three on the CDRs (white circles) and one between two β-sheets (orange circle) in both model superimpositions. Other parts of the variable domains are structurally similar. Results of RMSD of variable heavy and light regions are below 3.5 Å (Table 4.11), which is indicative of significant fold similarity and possible structural homology (Cuff, A et al. 2008). Hence, we concluded that there is a lack of evidence to suggest any structural differences between the variable regions of IgM 84 and 85.

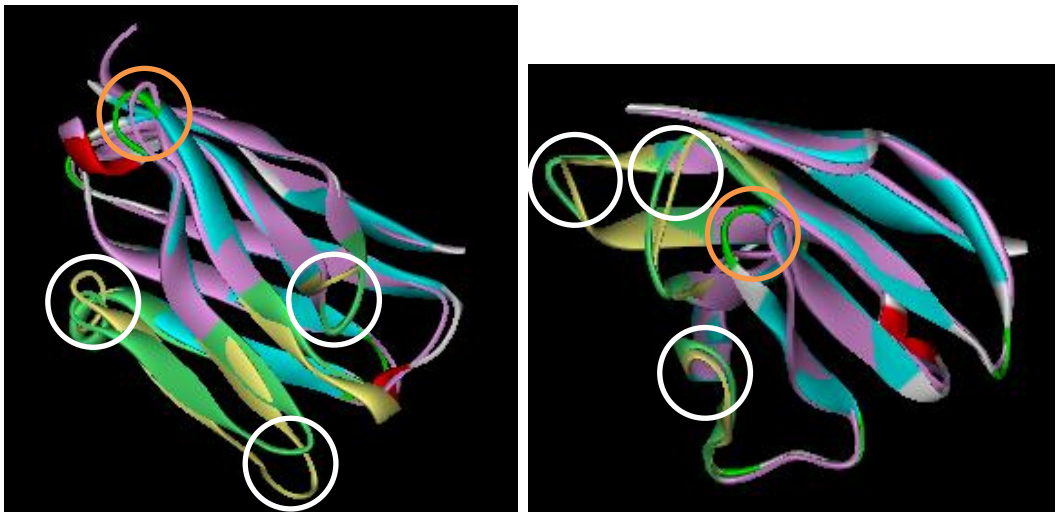

Figure 4.12 Model superimposition of two variable regions – heavy chains (left) and light chains (right) of IgM 84 and 85

Table 4.11 RMSD of model superimposition of the heavy chain and light chain variable regions of IgM 84 and 85

Protein domains	RMSD (Å)
Heavy variable regions (V_H)	1.51
Light variable regions (V_L)	1.27

5 CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The N-glycan analysis described in this thesis showed that the N-glycans of IgM 84 are less mature than those of IgM 85. This was evident from the presence of the immature high mannose type N-glycan $Man_9GlcNAc_2$, more non-fucosylated complex type N-glycans, lower overall degree of sialylation in IgM 84. The incomplete trimming of high mannose type N-glycans during the early stage of post-translational modifications causes IgM 84 to be differently folded as compared to IgM 85 because N-glycosylation plays an important role in protein folding in the endoplasmic reticulum (ER). Furthermore, the presence of various bi- and triantennary complex type N-glycan structures in IgM 84 that are non-fucosylated could also point to a less mature fucosylation in IgM 84 compared to that of IgM 85. This could then be due to a difference in protein folding between IgM 84 and 85 exiting the ER, i.e. shielding the structures at certain N-glycosylation sites from fucosylation during the late stage of post-translational modification. We also observed that the N-glycans of IgM 84 are generally less sialylated, besides the presence of two trisialylated complex type N-glycans that are unique to IgM 84, which lead to the possibility that sialylation of IgM 84 is also less mature compared to that of IgM 85.In summary, we suggest that the difference in mannose trimming and in fucosylation and sialylation are the effect resulted from the differently folded IgM 84 and 85 proteins. We propose that this difference in protein conformation of IgM 84 may attribute to the cytotoxic nature of IgM against undifferentiated human embryonic stem cells (hESCs) as it was previously suggested that multivalency of pentameric IgM 84 plays an important role in cytotoxicity effect (Lim et al., 2011).

Our 3D models of the variable binding of IgM 84 and 85 have demonstrated that the structural differences between IgM 84 and 85 in their antigen binding sites are subtle, although differences in their primary sequences had led us to think otherwise. This finding further substantiates the earlier claim that oncosis against undifferentiated hESCs was not observed

with single-valence binding of antigen sites (Lim et al., 2011), possibly due to the structural similarities between the variable binding sites of IgM 84 and 85 as suggested by our models. Though four loop regions are identified to be different upon superimposition, we are sceptical that such difference around the connecting loop could cause such a big difference in terms of cytotoxic functionality of the IgM 84.

In conclusion, our 3D models first discounted the possibility that any explanation to the cytotoxicity of IgM 84 was due to its antigen binding sites alone. In addition, our findings about the differences in N-glycosylation maturation between IgM 84 and 85 further suggest a differently folded IgM 84 that may attribute to its cytotoxic nature against undifferentiated hESCs.

5.2 Recommendations for Future Work

Following the work from this thesis, we recommend proceeding to identify the site specific differences of IgM 84 and 85 in terms of N-glycosylation. Though we have already identified a single site on IgM 84 that possesses high mannose type N-glycans, more work is required to elucidate the types of N-glycans that are present on each potential N-glycosylation site of both IgM 84 and 85. To do so, we propose to fractionate and analyze the individual glycopeptides (Appendix C) directly using liquid chromatography-mass spectrometry (LC-MS) to reduce sample loss during sample preparation steps and obtain enough resolution for having a good separation of all the glycopeptides, before we could complete this work. With such information on site-occupancy, it would allow us to shed more light on the different protein conformations of IgM 84 and 85. In addition, we would also continue our work on comparative modeling of the full pentameric IgM 84 and 85, as well as some experimental studies on their respective protein conformations, to demonstrate our hypothesis that the defect of maturation in glycosylation indeed results in a different protein conformation of IgM 84.

REFERENCES

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman. Basic local alignment search tool. Journal of molecular biology *215*(3):pp.403-410.1990

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research *25*(17):pp.3389-3402.1997

Anderson, D. R., P. H. Atkinson and W. J. Grimes. Major carbohydrate structures at five glycosylation sites on murine IgM determined by high resolution 1H-NMR spectroscopy. Archives of biochemistry and biophysics *243*(2):pp.605-618.1985

Anderson, D. R. and W. J. Grimes. Heterogeneity of asparagine-linked oligosaccharides of five glycosylation sites on immunoglobulin M heavy chain from mineral oil plasmacytoma 104E. The Journal of biological chemistry *257*(24):pp.14858-14864.1982

Andreeva, A., D. Howorth, J. M. Chandonia, S. E. Brenner, T. J. Hubbard, C. Chothia and A. G. Murzin. Data growth and its impact on the SCOP database: new developments. Nucleic acids research *36*(Database issue):pp.D419-425.2008

Anthony, R. M., F. Nimmerjahn, D. J. Ashline, V. N. Reinhold, J. C. Paulson and J. V. Ravetch. Recapitulation of IVIG anti-inflammatory activity with a recombinant IgG Fc. Science *320*(5874):pp.373-376.2008

Arnold, J. N., M. R. Wormald, D. M. Suter, C. M. Radcliffe, D. J. Harvey, R. A. Dwek, P. M. Rudd and R. B. Sim. Human serum IgM glycosylation: identification of glycoforms that can bind to mannan-binding lectin. The Journal of biological chemistry *280*(32):pp.29080-29087.2005

Bajaj, M. and T. Blundell. Evolution and the tertiary structure of proteins. Annual review of biophysics and bioengineering *13*:pp.453-492.1984

Bardor, M., D. H. Nguyen, S. Diaz and A. Varki. Mechanism of uptake and incorporation of the non-human sialic acid N-glycolylneuraminic acid into human cells. J Biol Chem *280*(6):pp.4228-4237.2005

Benkert, P., S. C. Tosatto and D. Schomburg. QMEAN: A comprehensive scoring function for model quality assessment. Proteins *71*(1):pp.261-277.2008

Benkert, P., S. C. Tosatto and T. Schwede. Global and local model quality estimation at CASP8 using the scoring functions QMEAN and QMEANclust. Proteins *77 Suppl 9*:pp.173-180.2009

Berman, H., K. Henrick, H. Nakamura and J. L. Markley. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. Nucleic acids research *35*(Database issue):pp.D301-303.2007

Bertozzi, C. R., H. H. Freeze, A. Varki and J. D. Esko. Glycans in Biotechnology and the Pharmaceutical Industry. In Essentials of Glycobiology. 2nd, ed by A. Varki, R. D. Cummings, J. D. Eskoet al, Cold Spring Harbor (NY).2009

Brenckle, R. and R. Kornfeld. Structure of the oligosaccharides of mouse immunoglobulin M secreted by the MOPC 104E plasmacytoma. Archives of biochemistry and biophysics *201*(1):pp.160-173.1980

Burton, D. R. and R. A. Dwek. Immunology. Sugar determines antibody activity. Science *313*(5787):pp.627-628.2006

Capriotti, E., P. Fariselli, I. Rossi and R. Casadio. A Shannon entropy-based filter detects high- quality profile-profile alignments in searches for remote homologues. Proteins *54*(2):pp.351-360.2004

Chan, A. C. and P. J. Carter. Therapeutic antibodies for autoimmunity and inflammation. Nature reviews. Immunology *10*(5):pp.301-316.2010

Chandonia, J. M. and S. E. Brenner. The impact of structural genomics: expectations and outcomes. Science *311*(5759):pp.347-351.2006

Chapman, A. and R. Kornfeld. Structure of the high mannose oligosaccharides of a human IgM myeloma protein. I. The major oligosaccharides of the two high mannose glycopeptides. The Journal of biological chemistry *254*(3):pp.816-823.1979

Chapman, A. and R. Kornfeld. Structure of the high mannose oligosaccharides of a human IgM myeloma protein. II. The minor oligosaccharides of high mannose glycopeptide. The Journal of biological chemistry *254*(3):pp.824-828.1979

Chen, V. B., W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson. MolProbity: all-atom structure validation for macromolecular crystallography. Acta crystallographica. Section D, Biological crystallography *66*(Pt 1):pp.12-21.2010

Choo, A. B., H. L. Tan, S. N. Ang, W. J. Fong, A. Chin, J. Lo, L. Zheng, H. Hentze, R. J. Philp, S. K. Oh and M. Yap. Selection against undifferentiated human embryonic stem cells by a cytotoxic antibody recognizing podocalyxin-like protein-1. Stem cells *26*(6):pp.1454-1463.2008

Chothia, C. and A. M. Lesk. The relation between the divergence of sequence and structure in proteins. The EMBO journal *5*(4):pp.823-826.1986

Chothia, C. and A. M. Lesk. The evolution of protein structures. Cold Spring Harbor symposia on quantitative biology *52*:pp.399-405.1987

Cuff, A., Redfern, O, Orengo, C. Classification of Protein Structures. In Computational structural biology: methods and applications. ed by T. Schwede, and M.C. Peitsch, pp. 153-188. World Scientific. 2008.

Cuff, A. L., I. Sillitoe, T. Lewis, O. C. Redfern, R. Garratt, J. Thornton and C. A. Orengo. The CATH classification revisited--architectures reviewed and new ways to characterize structural divergence in superfamilies. Nucleic acids research *37*(Database issue):pp.D310-314.2009

Cummings, R. D. and R. P. McEver. C-type Lectins. In Essentials of Glycobiology. 2nd, ed by A. Varki, R. D. Cummings, J. D. Eskoet al, Cold Spring Harbor (NY).2009

Dayhoff, M.O., Schwartz, R. and Orcutt, B.C. A model of Evolutionary Change in Proteins. Atlas of protein sequence and structure, Vol 5: Supp 3, ed byM.O. Dayhoff, pp.345–358. Nat. Biomed. Res. Found. 1978.

Eddy, S. R. Where did the BLOSUM62 alignment score matrix come from? Nature biotechnology *22*(8):pp.1035-1036.2004

Finkelstein, A. V., A. Badretdinov and A. M. Gutin. Why do protein architectures have Boltzmann-like statistics? Proteins *23*(2):pp.142-150.1995

Freeze, H. H., J. D. Esko and A. J. Parodi. Glycans in Glycoprotein Quality Control. In Essentials of Glycobiology. 2nd, ed by A. Varki, R. D. Cummings, J. D. Eskoet al, Cold Spring Harbor (NY).2009

Galili, U. The alpha-gal epitope and the anti-Gal antibody in xenotransplantation and in cancer immunotherapy. Immunology and cell biology *83*(6):pp.674-686.2005

Goldstein, I.J. and Poretz, R.D. Isolation, physicochemical characterization, and carbohydrate-binding specificity of lectins. In.The Lectins Properties, Functions and Applications in Biology and Medicine, ed by I.E. Liener, N. Sharon and I.J. Goldstein, pp.233-247. Orlando: Academic Press. 1986.

Gil, G. C., W. H. Velander and K. E. Van Cott. N-glycosylation microheterogeneity and site occupancy of an Asn-X-Cys sequon in plasma-derived and recombinant protein C. Proteomics *9*(9):pp.2555-2567.2009

Ginalski, K. Comparative modeling for protein structure prediction. Current opinion in structural biology *16*(2):pp.172-177.2006

Ha, S., Y. Ou, J. Vlasak, Y. Li, S. Wang, K. Vo, Y. Du, A. Mach, Y. Fang and N. Zhang. Isolation and characterization of IgG1 with asymmetrical Fc glycosylation. Glycobiology *21*(8):pp.1087-1096.2011

Henikoff, S. and J. G. Henikoff. Amino acid substitution matrices from protein blocks. Proceedings of the National Academy of Sciences of the United States of America *89*(22):pp.10915-10919.1992

Holm, L. and C. Sander. Touring protein fold space with Dali/FSSP. Nucleic acids research *26*(1):pp.316-319.1998

Hooker, A. D. and D. C. James. Analysis of glycoprotein heterogeneity by capillary electrophoresis and mass spectrometry. Molecular biotechnology *14*(3):pp.241-249.2000

Hossler, P., S. F. Khattak and Z. J. Li. Optimal and consistent protein glycosylation in mammalian cell culture. Glycobiology *19*(9):pp.936-949.2009

James, D. C., R. B. Freedman, M. Hoare and N. Jenkins. High-resolution separation of recombinant human interferon-gamma glycoforms by micellar electrokinetic capillary chromatography. Analytical biochemistry *222*(2):pp.315-322.1994

Janeway, C. (2001). Immunobiology the immune system in health and disease. NCBI bookshelf. New York, Garland Pub.

Jefferis, R. Glycosylation as a strategy to improve antibody-based therapeutics. Nature reviews. Drug discovery *8*(3):pp.226-234.2009

Jiang, X. R., A. Song, S. Bergelson, T. Arroll, B. Parekh, K. May, S. Chung, R. Strouse, A. Mire-Sluis and M. Schenerman. Advances in the assessment and control of

the effector functions of therapeutic antibodies. Nature reviews. Drug discovery *10*(2):pp.101-111.2011

Jones, D. T., M. Tress, K. Bryson and C. Hadley. Successful recognition of protein folds using threading methods biased by sequence similarity and predicted secondary structure. Proteins *Suppl 3*:pp.104-111.1999

Kaneko, Y., F. Nimmerjahn and J. V. Ravetch. Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. Science *313*(5787):pp.670-673.2006

Knoepfler, P. S. Deconstructing stem cell tumorigenicity: a roadmap to safe regenerative medicine. Stem cells *27*(5):pp.1050-1056.2009

Larsson, P., B. Wallner, E. Lindahl and A. Elofsson. Using multiple templates to improve quality of homology models in automated homology modeling. Protein science : a publication of the Protein Society *17*(6):pp.990-1002.2008

Laskowski, R. A., J. A. Rullmannn, M. W. MacArthur, R. Kaptein and J. M. Thornton. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. Journal of biomolecular NMR *8*(4):pp.477-486.1996

Lee, J., A. Tscheliessnig, A. Chen, Y. Y. Lee, G. Adduci, A. Choo and A. Jungbauer. Adaptation of hybridomas to protein-free media results in a simplified two-step immunoglobulin M purification process. Journal of chromatography. A *1216*(13):pp.2683-2688.2009

Levitt, M. Accurate modeling of protein conformation by automatic segment matching. Journal of molecular biology *226*(2):pp.507-533.1992

Lieberman, B. Human evolution: details of being human. Nature *454*(7200):pp.21-23.2008

Lim, D. Y., Y. H. Ng, J. Lee, M. Mueller, A. B. Choo and V. V. Wong. Cytotoxic antibody fragments for eliminating undifferentiated human embryonic stem cells. Journal of biotechnology *153*(3-4):pp.77-85.2011

Lindvall, O. and Z. Kokaia. Stem cells for the treatment of neurological disorders. Nature *441*(7097):pp.1094-1096.2006

Liu, T., G. W. Tang and E. Capriotti. Comparative modeling: the state of the art and protein drug target structure prediction. Combinatorial chemistry & high throughput screening *14*(6):pp.532-547.2011

Lushington, G. H. Comparative modeling of proteins. Methods in molecular biology *443*:pp.199-212.2008

Markely, L. R., B. T. Ong, K. M. Hoi, G. Teo, M. Y. Lu and D. I. Wang. A high-throughput method for quantification of glycoprotein sialylation. Analytical biochemistry *407*(1):pp.128-133.2010

Marsden, R. L., L. J. McGuffin and D. T. Jones. Rapid protein domain assignment from amino acid sequence using predicted secondary structure. Protein science : a publication of the Protein Society *11*(12):pp.2814-2824.2002

Matsui, T., E. Takita, T. Sato, S. Kinjo, M. Aizawa, Y. Sugiura, T. Hamabata, K. Sawada and K. Kato. N-glycosylation at noncanonical Asn-X-Cys sequences in plant cells. Glycobiology *21*(8):pp.994-999.2011

Mimura, Y., S. Church, R. Ghirlando, P. R. Ashton, S. Dong, M. Goodall, J. Lund and R. Jefferis. The influence of glycosylation on the thermal stability and effector function expression of human IgG1-Fc: properties of a series of truncated glycoforms. Molecular immunology *37*(12-13):pp.697-706.2000

Monica, T. J., S. B. Williams, C. F. Goochee and B. L. Maiorella. Characterization of the glycosylation of a human IgM produced by a human-mouse hybridoma. Glycobiology *5*(2):pp.175-185.1995

Morelle, W. and J. C. Michalski. Analysis of protein glycosylation by mass spectrometry. Nature protocols *2*(7):pp.1585-1602.2007

Mulloy, B., G. W. Hart and P. Stanley. Structural Analysis of Glycans. In Essentials of Glycobiology. 2nd, ed by A. Varki, R. D. Cummings, J. D. Eskoet al, Cold Spring Harbor (NY).2009

Needleman, S. B. and C. D. Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of molecular biology *48*(3):pp.443-453.1970

Nimmerjahn, F. and J. V. Ravetch. Fcgamma receptors as regulators of immune responses. Nature reviews. Immunology *8*(1):pp.34-47.2008

Oldfield, T. J. SQUID: a program for the analysis and display of data from crystallography and molecular dynamics. Journal of molecular graphics *10*(4):pp.247-252.1992

Padler-Karavani, V., H. Yu, H. Cao, H. Chokhawala, F. Karp, N. Varki, X. Chen and A. Varki. Diversity in specificity, abundance, and composition of anti-Neu5Gc antibodies in normal humans: potential implications for disease. Glycobiology *18*(10):pp.818-830.2008

Pearson, W. R. Rapid and sensitive sequence comparison with FASTP and FASTA. Methods in enzymology *183*:pp.63-98.1990

Perkins, S. J., A. S. Nealis, B. J. Sutton and A. Feinstein. Solution structure of human and mouse immunoglobulin M by synchrotron X-ray scattering and molecular graphics modelling. A possible mechanism for complement activation. Journal of molecular biology *221*(4):pp.1345-1366.1991

Raju, T. S. Terminal sugars of Fc glycans influence antibody effector functions of IgGs. Current opinion in immunology *20*(4):pp.471-478.2008

Rooman, M. J. and S. J. Wodak. Are database-derived potentials valid for scoring both forward and inverted protein folding? Protein engineering *8*(9):pp.849-858.1995

Rychlewski, L., L. Jaroszewski, W. Li and A. Godzik. Comparison of sequence profiles. Strategies for structural predictions using sequence information. Protein science : a publication of the Protein Society *9*(2):pp.232-241.2000

Sali, A. and T. L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. Journal of molecular biology *234*(3):pp.779-815.1993

Sato, C., J. H. Kim, Y. Abe, K. Saito, S. Yokoyama and D. Kohda. Characterization of the N-oligosaccharides attached to the atypical Asn-X-Cys sequence of recombinant human epidermal growth factor receptor. Journal of biochemistry *127*(1):pp.65-72.2000

Schriebl, K., S. Lim, A. Choo, A. Tscheliessnig and A. Jungbauer. Stem cell separation: a bottleneck in stem cell therapy. Biotechnology journal *5*(1):pp.50-61.2010

Schriebl, K., G. Satianegara, A. Hwang, H. L. Tan, W. J. Fong, H. H. Yang, A. Jungbauer and A. Choo. Selective Removal of Undifferentiated Human Embryonic Stem Cells Using Magnetic Activated Cell Sorting Followed by a Cytotoxic Antibody. Tissue Eng Part A.2012

Shibata-Koyama, M., S. Iida, A. Okazaki, K. Mori, K. Kitajima-Miyama, S. Saitou, S. Kakita, Y. Kanda, K. Shitara, K. Kato and M. Satoh. The N-linked oligosaccharide at Fc gamma RIIIa Asn-45: an inhibitory element for high Fc gamma RIIIa binding affinity to IgG glycoforms lacking core fucosylation. Glycobiology *19*(2):pp.126-134.2009

Siberil, S., C. A. Dutertre, W. H. Fridman and J. L. Teillaud. FcgammaR: The key to optimize therapeutic antibodies? Critical reviews in oncology/hematology *62*(1):pp.26-33.2007

Smith, T. F. and M. S. Waterman. Identification of common molecular subsequences. Journal of molecular biology *147*(1):pp.195-197.1981

Stanley, P., H. Schachter and N. Taniguchi. N-Glycans. In Essentials of Glycobiology. 2nd, ed by A. Varki, R. D. Cummings, J. D. Eskoet al, Cold Spring Harbor (NY).2009

Sumer-Bayraktar, Z., D. Kolarich, M. P. Campbell, S. Ali, N. H. Packer and M. Thaysen-Andersen. N-glycans modulate the function of human corticosteroid-binding globulin. Mol Cell Proteomics *10*(8):pp.M111 009100.2011

Sutcliffe, M. J., I. Haneef, D. Carney and T. L. Blundell. Knowledge based modelling of homologous proteins, Part I: Three-dimensional frameworks derived from the simultaneous superposition of multiple structures. Protein engineering *1*(5):pp.377-384.1987

Tan, H. L., W. J. Fong, E. H. Lee, M. Yap and A. Choo. mAb 84, a cytotoxic antibody that kills undifferentiated human embryonic stem cells via oncosis. Stem cells *27*(8):pp.1792-1801.2009

Tarentino, A. L. and T. H. Plummer, Jr. Enzymatic deglycosylation of asparagine-linked glycans: purification, properties, and specificity of oligosaccharide-cleaving enzymes from Flavobacterium meningosepticum. Methods in enzymology *230*:pp.44-57.1994

Thomas, P. D. and K. A. Dill. Statistical potentials extracted from protein structures: how accurate are they? Journal of molecular biology *257*(2):pp.457-469.1996

Thomson, J. A., J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall and J. M. Jones. Embryonic stem cell lines derived from human blastocysts. Science *282*(5391):pp.1145-1147.1998

Tretter, V., F. Altmann and L. Marz. Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase F cannot release glycans with fucose attached alpha 1-3 to the asparagine-linked N-acetylglucosamine residue. European journal of biochemistry / FEBS *199*(3):pp.647-652.1991

Tscheliessnig, A., D. Ong, J. Lee, S. Pan, G. Satianegara, K. Schriebl, A. Choo and A. Jungbauer. Engineering of a two-step purification strategy for a panel of monoclonal immunoglobulin M directed against undifferentiated human embryonic stem cells. Journal of chromatography. A *1216*(45):pp.7851-7864.2009

Unger, R., D. Harel, S. Wherland and J. L. Sussman. A 3D building blocks approach to analyzing and predicting structure of proteins. Proteins *5*(4):pp.355-373.1989

Vaguine, A. A., J. Richelle and S. J. Wodak. SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. Acta crystallographica. Section D, Biological crystallography *55*(Pt 1):pp.191-205.1999

Varki, A. and N. Sharon. Historical Background and Overview. In Essentials of Glycobiology. 2nd, ed by A. Varki, R. D. Cummings, J. D. Eskoet al, Cold Spring Harbor (NY).2009

Venclovas, C. and M. Margelevicius. The use of automatic tools and human expertise in template-based modeling of CASP8 target proteins. Proteins *77 Suppl 9*:pp.81-88.2009

Wormald, M. R., E. W. Wooten, R. Bazzo, C. J. Edge, A. Feinstein, T. W. Rademacher and R. A. Dwek. The conformational effects of N-glycosylation on the tailpiece from serum IgM. European journal of biochemistry / FEBS *198*(1):pp.131-139.1991

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. The Journal of biological chemistry *265*(18):pp.10506-10513.1990

APPENDIX A: Sequence analysis of IgM 84 and 85

Partial amino acid sequences of heavy and light chains of IgM 84 and 85 are shown as follows i.e. Seq1, Seq2, Seq3 and Seq4. Amino acids are represented by the 1-letter code. For each sequence, variable regions are underlined; potential N-glycosylation sites are highlighted in **red i.e. NNT, NFT, NVS, NLT, NIS, NGT** and **NEC**; and complementarity determining regions (CDRs) are highlighted in **green**. Heavy chains of IgM 84 and 85 have 551 and 549 amino acids, respectively, whereas light chains of IgM 84 and 85 have 213 and 214 amino acids, respectively.

Seq1. Partial heavy chain amino acid sequence of IgM 84

1 QVQLQQSGGGLVQPGGSMKLSCVASGFTFS**NYWMN**WVRQSPEKGLEWVA**EIRLKSNNYAT**

61 **HYAESVKGR**FTISRDDSKSSVYLQMNNLRAEDTGIYYCTG**ERA**WGQGTTVTVSAESQSFP

121 NVFPLVSCESPLSDKNLVAMGCLARDFLPSTISFTWNYQ**NNT**E... EAT**NFT**PKPI...

321 TFLK**NVS**ST...FADIFLSKSA**NLT**CLVSNLATYETL**NIS**WASQ...TKIKIMESHP**NGT**

541 TERTVDKSTGK

Seq2. Partial light chain amino acid sequence of IgM 84

1 DIELTQSPAIMSASPGEKVTMTC**SASSSVNYMY**WYQQKPGSSPRLLIY**DTSNLAS**GVPVR

61 FSGSGSGTSYSLTISRMEAEDAATYYC**QQWSSYPYT**FGGGTKLEIKRADAAPTVSIF...

181 TK...FNR**NEC**

Seq3. Partial heavy chain sequence of IgM 85

1 QVKLQESGPGLVQPSQSLSITCTVSGFSLT**GYGLH**WVRQSPGKGLEWLG**VIWRGGNTDYN**

61 **AAFMSR**LSITKDNSKSQVFFKMNSLQADDTAIYYCAR**DFDY**WGQGTTVTVSSESQSFPNV

121 FPLVSCESPLSDKNLVAMGCLARDFLPSTISFTWNYQ**NNT**E... EAT**NFT**PKPITV...

321 TFLK**NVS**ST...FADIFLSKSA**NLT**CLVSNLATYETL**NIS**WASQ...TKIKIMESHP**NGT**

541 TERTVDKSTGK

Seq4. Partial heavy chain sequence of IgM 85

1 DIELTQSPSSLSASLGERVSLTC**RASQEISDYL**SWLQQKPDGTIKRLIY**AASTLDS**GVPK

61 RFSGSRSGSDYSLTISSLESEDFADYYC**LQYSSHPYT**FGGGTKLEIKRADAAPTVSI...

181 LTK...FNR**NEC**

Sequence alignment between constant and variable regions of IgM 84 and 85 are shown as follows i.e. SA1 and SA2 respectively. We highlighted the portion light chains of these regions in **brown** and the CDRs in **green**. The constant regions are almost identical i.e. 99.82% sequence identity except for one amino acid at position 177 (highlighted in red) i.e. IgM 84 has a theorine (Thr or T) and IgM 85 has a serine (Ser or S). The main difference between primary amino acid sequences of IgM 84 and 85 therefore, comes from variable regions i.e. 57.85% sequence identity, in particular the CDRs.

SA1. Segment of alignment between constant regions of IgM 84 and 85

```
               10        20        30        40        50        60
IgM84  ESQSFPNVFPLVSCESPLSDKNLVAMGCLARDFLPSTISFTWNYQNNTEVIQGIRTFPTL
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
IgM85  ESQSFPNVFPLVSCESPLSDKNLVAMGCLARDFLPSTISFTWNYQNNTEVIQGIRTFPTL
               10        20        30        40        50        60

              130       140       150       160       170       180
IgM84  RDGFSGPAPRKSKLICEATNFTPKPITVSWLKDGKLVESGFTTDPVTIENKGSTPQTYKV
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::.:::
IgM85  RDGFSGPAPRKSKLICEATNFTPKPITVSWLKDGKLVESGFTTDPVTIENKGSTPQSYKV
              130       140       150       160       170       180

              430       440       450       460       470       480
IgM84  ALPHLVTERTVDKSTGKADAAPTVSIFPPSSEQLTSGGASVVCFLNNFYPKDINVKWKID
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
IgM85  ALPHLVTERTVDKSTGKADAAPTVSIFPPSSEQLTSGGASVVCFLNNFYPKDINVKWKID
              430       440       450       460       470       480


IgM84  NEC
       :::
IgM85  NEC
```

Note: There are three symbols represented by sequence alignment results i.e ‘ : ’ means two aligned amino acids are identical; ‘ . ’ means two aligned amino acids are similar but not identical; ‘ ’ means two aligned amino acids are neither identical nor similar.

SA2. Segment of alignment between variable regions of IgM 84 and 85 variable regions

```
               10        20        30        40        50        60
IgM84  QVQLQQSGGGLVQPGGSMKLSCVASGFTFSNYWMNWVRQSPEKGLEWVAEIRLKSNNYAT
       ::.::.:: :::::. :....:..:::....: ..::::::: :::::.. :    .:   :
IgM85  QVKLQESGPGLVQPSQSLSITCTVSGFSLTGYGLHWVRQSPGKGLEWLGVIWRGGN---T
               10        20        30        40        50
```

```
     120       130       140       150       160       170
IgM84  QSPAIMSASPGEKVTMTCSASSSV-NYMYWYQQKPGSSPRLLIYDTSNLASGVPVRF...
       :::. .::: ::.:..:: ::. . .:. : :::: .. . ::: .:.: :::: ::
IgM85  QSPSSLSASLGERVSLTCRASQEISDYLSWLQQKPDGTIKRLIYAASTLDSGVPKRF...
       120       130       140       150       160       170

          200       210       220
IgM84  ...YYCQQWSSYPYTFGGGTKLEIKR
          ::: :.::.:::::::::::::::
IgM85  ...YYCLQYSSHPYTFGGGTKLEIKR
          200       210       220
```

Note: There are three symbols represented by sequence alignment results i.e ‘ : ’ means two aligned amino acids are identical; ‘ . ’ means two aligned amino acids are similar but not identical; ‘ ’ means two aligned amino acids are neither identical nor similar.

APPENDIX B. N-linked glycan profiling resources

Table B1. N-linked glycan profiling data for all mouse cell types available from Consortium for Functional Glycomics (CFG) resources

Mouse Cell Type	Comments	Participating Investigator
A9 fibroblasts	A9 fibroblasts	Agbandje-McKenna, Mavis
SV40 transformed fibroblast cell line, NB324K	SV40 transformed fibroblast cell line, NB324K, NG	
EL4 T lymphocytes	EL4 T lymphocytes, NG	
A9 fibroblasts	A9 fibroblasts, NG	
SV40 transformed fibroblast cell line, NB324K	SV40 transformed fibroblast cell line, NB324K, NG	
Neutrophils	Mouse Neutrophils	Cummings,Richard D.
WEHI-3	Murine WEHI-3	
Murine mammary carcinoma	4T07 p.20 7-21-05	Rittenhouse-Olson, Kate
Murine mammary carcinoma	67NR p.22 7-25-05	
Murine mammary carcinoma	4TI p.31 7-21-05	
RAW	RAW Cells	Merrill, Alfred
Macrophages	Non-treated thioglycolate macrophages	Gordon, Siamon
Macrophages	IL-4 treated thioglycolate macrophages	
WEHI 231 B Cells	WEHI 231 B Cells, Control	Cook-Mills, Joan
WEHI 231 B Cells	WEHI 231 B Cells, Test (apoptotic)	
WEHI 231 B Cells	WEHI 231 B Plasma Membranes, Control	
WEHI 231 B Cells	WEHI 231 B Plasma Membranes, Test (apoptotic)	

Table B2. Mouse N-linked glycan profiling data for all mouse cell types available from Consortium for Functional Glycomics (CFG) resources

Cell Type	Comments	Participating Investigator
Cytokine-induced killer (CIK) cells	Cytokine-induced killer (CIK) cells, Unsorted	Contag, Christopher
Cytokine-induced killer (CIK) cells	Cytokine-induced killer (CIK) cells, NKG2D Treated	
Eosinophils	Eosinophils (EOS), IL5 Tg mice (NJ1638) E782	Paulson, James
Eosinophils	Eosinophils (EOS), IL5 Tg mice (NJ1638) E793	
Osteoclasts	Osteoclast Cells, Mouse 2, Control, NG	Wu, Hui
Osteoclasts	Osteoclast Cells, Mouse 1, RANKL, NG	
Osteoclasts	Osteoclast Cells, Mouse 1, Tunicamycin, NG	
Osteoclasts	Osteoclast Cells, Mouse 1, RANKL+Tunicamycin, NG	
Osteoclasts	Osteoclast Cells, Mouse 2, Control, NG	
Osteoclasts	Osteoclast Cells, Mouse 2, RANKL, NG	
Osteoclasts	Osteoclast Cells, Mouse 2, Tunicamycin, NG	
Osteoclasts	Osteoclast Cells, Mouse 2, RANKL+Tunicamycin, NG	
B1 Cells	Murine B1 cells (peritoneal cavity), mouse A (MB1/A), NG	Nitschke, Lars
B1 Cells	Murine B1 cells (peritoneal cavity), mouse B (MB1/B), NG	
B1 Cells	Murine B2 cells (spleen), mouse A (MB2/A), NG	
B1 Cells	Murine B2 cells (spleen), mouse B (MB2/B), NG	
Antibodies	Murine antibodies, LN 2-4G2, NG	
Antibodies	Murine antibodies, RA36B2, NG	

Table B3. Mouse N-linked glycan profiling data for all mouse spleen tissues available from Consortium for Functional Glycomics (CFG) resources

Mouse Strain	Mice Colony Code	Mouse Type
Wild type	N.A.	C57BL/6
Fuct IV	FA	C57BL/6
Fuct VII	FB	C57BL/6
Fuct IV + VII	FC	C57BL/6
Gal 3	GT	C57BL/6
ST3 Gal 1	ST	C57BL/6
CT GalNAcT	N.A.	C57BL/6
Wild Type	N.A.	129x1SvJ

APPENDIX C. Glycopeptide sequences of digested IgM 84 and 85

IgM 84 and 85 that are completely digested with trypsin result in pools of glycopeptides, peptides and amino acids. Table C1, C2, C3 and C4 show the theoretical sequences of these glycopeptides, peptides and amino acids analyzed from the enzymatic action of trypsin, which lyzes the C-terminal of lysine (Lys or K) or arginine (Arg or R) except for those that are immediately followed by proline (Pro or P). The sequences are presented in 1-letter code of amino acids. Again, potential N-glycosylation sites are highlighted in **red** i.e. in T12, T23, T31, T32 and T34 of IgM 85 heavy chain; T21 of IgM 85 light chain; T14, T25, T33, T34 and T36 of IgM 84 heavy chain; and T18 of IgM 84 light chain.

Table C1. Selected list of amino acid sequences of glycopeptides, peptides and amino acids of IgM 85 full heavy chain after trypsin digestion

Label	Glycopeptides, Peptides sequences or Amino Acid
T1	QVK
T2	LQESGPGLVQPSQSLSITCTVSGFSLTGYGLHWVR
T3	QSPGK
...	...
T12	DFLPSTISFTWNYQ**NNT**EVIQGIR
T13	TFPTLR
...	...
T23	LICEAT**NFT**PK
T24	PITVSWLK
...	...
T31	**NVS**STCAASPSTDILTFTIPPSFADIFLSK
T32	SA**NLT**CLVSNLATYETL**NIS**WASQSGEPLETK
T33	IK
T34	IMESHP**NGT**FSAK
T35	GVASVCVEDWNNR
...	...
T49	STGK

Table C2. Selected list of amino acid sequences of glycopeptides, peptides and amino acids of IgM 85 full light chain after trypsin digestion

Label	Glycopeptides, Peptides sequences or Amino Acid
T1	DIELTQSPSSLSASLGER
T2	VSLTCR
T3	ASQEISDYLSWLQQKPDGTIK
...	...
T21	**NEC**

Table C3. Selected list of amino acid sequences of glycopeptides, peptides and amino acids of IgM 84 full heavy chain after trypsin digestion

Label	Glycopeptide, Peptide or Amino acid sequence(s)
T1	QVQLQQSGGGLVQPGGSMK
T2	LSCVASGFTFSNYWMNWVR
T3	QSPEK
...	...
T14	DFLPSTISFTWNYQ**NNT**EVIQGIR
T15	TFPTLR
...	...
T25	LICEAT**NFT**PK
T26	PITVSWLK
...	...
T33	**NVS**STCAASPSTDILTFTIPPSFADIFLSK
T34	SA**NLT**CLVSNLATYETL**NIS**WASQSGEPLETK
T35	IK
T36	IMESHP**NGT**FSAK
T37	GVASVCVEDWNNR
...	...
T51	STGK

Table C4. Selected list of amino acid sequences of glycopeptides, peptides and amino acids of IgM 84 full light chain after trypsin digestion

No.	Glycopeptide, Peptide or Amino acid sequence(s)
T1	DIELTQSPAIMSASPGEK
T2	VTMTCSASSSVNYMYWYQQKPGSSPR
T3	LLIYDTSNLASGVPVR
...	...
T18	**NEC**

APPENDIX D. Masses, structures, percentages of relative abundance and distribution of all N-glycans on IgM 84 and 85

We categorized all the N-glycan structures in four groups – high mannose, biantennary, triantennary complex and hybrid structures as shown in Table D1 and D2. Mass peaks obtained experimentally from MALDI-TOF MS are of mass accuracies[1] within range of 0.0022 – 0.0134% (IgM 84) and 0.0038 – 0.0203%. (IgM 85). The absolute intensities of the mass peaks were chosen to calculate the percentage relative abundance (%RA) and percentage distributions (%D) according to formulae described in Section 3.2.7.1.

Table D1. Percentage relative abundance (%RA) and percentage distribution (%D) of all N-glycans of IgM 84 measured by MALDI-TOF MS of two independent runs i.e. Run 1 and Run 2

		Run 1		Run 2	
N-glycan masses	**N-glycan structures**	**%RA**	**%D**	**%RA**	**%D**
High mannose					
1579.78	$GlcNAc_2Man_5$	11.8%	16.8%	13.1%	15.8%
1783.88	$GlcNAc_2Man_6$	47.2%	67.0%	57.4%	69.6%
1987.98^	$GlcNAc_2Man_7$	9.2%	13.1%	10.1%	12.2%
2192.08^	$GlcNAc_2Man_8$	2.0%	2.8%	1.9%	2.3%
2396.18	$GlcNAc_2Man_9$	0.2%	0.3%	0.1%	0.2%
	Subtotal	**70.4%**		**82.5%**	
Biantennary complex type					
1590.80	$FcGlcNAc_2A_1$	0.3%	2.2%	0.2%	2.9%
1661.84	$GlcNAc_2A_2$	0.2%	1.6%	0.1%	1.2%
1835.93	$FcGlcNAc_2A_2$	0.8%	5.9%	0.5%	6.1%
1865.94	$GlcNAc2A_2G_1$	0.4%	3.0%	0.2%	2.8%
1981.98	$GlcNAc2A_1G_1S_1$	0.1%	1.0%	0.1%	1.2%
2040.03	$FcGlcNAc_2A_2G_1$	0.4%	3.3%	0.3%	3.1%
2070.04	$GlcNAc_2A_2G_2$	0.9%	6.6%	0.5%	6.3%
2186.08	$FcGlcNAc_2A_1G_1S'_1$	0.4%	2.9%	0.3%	3.5%
2244.13	$FcGlcNAc_2A_2G_2$	0.8%	6.3%	0.6%	6.8%
2315.16	$GlcNAc_2A_2BG_2$	0.3%	2.0%	0.1%	1.1%
2431.21^	$FcGlcNAc_2A_2G_1S'_1$	0.6%	4.2%	0.4%	4.9%
2448.22^	$FcGlcNAc_2A_2G_3$	0.6%	4.8%	0.4%	4.4%
2461.22	$GlcNAc_2A_2G_2S'_1$	1.7%	12.3%	1.2%	14.7%
2478.24^	$GlcNAc_2A_2G_4$	0.6%	4.7%	0.2%	2.0%

[1] Mass accuracy = [(Mass $peak_{mass\ spectrum}$ – Theoretical mass)/Theoretical mass] x 100%

2489.25	$FcGlcNAc_2A_2BG_2$	0.4%	2.8%	0.2%	2.0%
2605.30	$FcGlcNAc_2A_2G_2S_1$	0.5%	3.6%	0.3%	4.0%
2635.31^	$GlcNAc_2A_2G_3S_1$	2.1%	15.7%	1.4%	17.0%
2665.32	$GlcNAc_2A_2G_3S'_1$	0.7%	5.3%	0.5%	6.1%
2809.40^	$FcGlcNAc_2A_2G_3S_1$	0.3%	2.3%	0.2%	1.8%
2839.41^	$FcGlcNAc_2A_2G_3S'_1$	0.6%	4.5%	0.3%	3.8%
2852.40	$GlcNAc_2A_2G_2S'_2$	0.3%	2.0%	0.1%	1.8%
2996.48	$FcGlcNAc_2A_2G_2S_1S'_1$	0.1%	0.7%	0.0%	0.6%
3026.49	$FcGlcNAc_2A_2G_2S'_2$	0.3%	2.2%	0.2%	1.8%
	Subtotal	13.5%		8.4%	
Triantennary complex type					
2519.26	$GlcNAc_2A_3G_3$	0.5%	5.7%	0.2%	7.1%
2693.35^	$FcGlcNAc_2A_3G_3$	0.8%	8.8%	0.4%	13.2%
2880.44	$GlcNAc_2A_3G_3S_1$	0.3%	3.2%	0.1%	4.2%
2897.45^	$FcGlcNAc_2A_3G_4$	0.5%	6.0%	0.2%	7.6%
2910.45	$GlcNAc_2A_3G_3S'_1$	0.4%	4.2%	0.2%	6.1%
3054.52^	$FcGlcNAc_2A_3G_3S_1$	0.5%	6.0%	0.2%	6.6%
3084.54^	$GlcNAc_2A_3G_4S_1$	1.2%	13.3%	0.5%	16.4%
3101.55^	$FcGlcNAc_2A_3G_5$	0.6%	6.7%	0.2%	4.8%
3241.61^	$GlcNAc_2A3G_3S_2$	0.1%	1.0%	0.0%	0.0%
3258.62^	$FcGlcNAc_2A_3G_4S_1$	0.5%	5.3%	0.2%	5.3%
3271.62	$GlcNAc_2A_3G_3S_1S'_1$	0.2%	1.8%	0.0%	0.0%
3288.64^	$GlcNAc_2A_3G_5S_1$	0.6%	6.5%	0.2%	6.9%
3301.63	$GlcNAc_2A_3G_3S'_2$	0.2%	1.9%	0.0%	1.6%
3305.65^	$FcGlcNAc_2A_3G_6$	0.4%	4.0%	0.0%	0.0%
3415.70	$FcGlcNAc_2A_3G_3S_2$	0.1%	1.5%	0.1%	1.9%
3445.71^	$GlcNAc_2A_3G_4S_2$	0.3%	3.5%	0.1%	2.7%
3462.72^	$FcGlcNAc_2A_3G_5S_1$	0.4%	4.4%	0.1%	2.5%
3475.72^	$FcGlcNAc_2A_3G_3S'_2$	0.4%	4.8%	0.1%	3.4%
3492.73^	$FcGlcNAc_2A_3G_5S'_1$	0.3%	4.0%	0.1%	2.3%
3619.80^	$FcGlcNAc_2A_3G_4S_2$	0.1%	1.5%	0.0%	1.5%
3649.81	$FcGlcNAc_2A_3G_4S'_1S_1$	0.2%	2.2%	0.1%	2.1%
3679.82	$FcGlcNAc_2A_3G_4S'_2$	0.2%	2.2%	0.1%	2.2%
3866.90	$FcGlcNAc_2A_3G_3S'_3$	0.1%	1.3%	0.1%	1.7%
	Subtotal	**8.8%**		**3.2%**	
Hybrid and others					
1824.91	$GlcNAc_2A_1Man_5$	1.4%	19.8%	0.8%	14.2%
2029.01	$GlcNAc_2A_1Man_5G_1$	1.6%	21.3%	1.2%	19.8%
2216.09	$GlcNAc_2A_1Man_4G_1S'_1$	0.7%	9.9%	0.6%	10.6%
2420.19	$GlcNAc_2A_1Man_5G_1S'_1$	3.4%	46.7%	3.2%	54.5%
3041.53*	$FcGlcNAcA_3Fc_2G_3$	0.2%	2.3%	0.1%	0.9%
	Subtotal	**7.3%**		**5.9%**	
	Total	**100.0%**		**100.0%**	

Note: A_1, A_2 and A_3 represent trimannosyl core with one, two and three GlcNAc sugar units (or antenna), respectively; whereas Fc, G, S, S' represent fucose galactose, Neu5Ac and Neu5Gc sugar units, respectively; and B represents a bisecting GlcNAc sugar that is attached β1-4 to A trimannosyl core. Subscript of each sugar shows the number of sugar units that are attached. *Mass peak of 3041.53 is a biantennary complex type N-glycan with a Le^a epitope, however the presence of this structure requires further confirmation by western blot. ^N-glycan masses require further fragmentation by MALDI-TOF-TOF MS/MS to elucidate the structure.

Table D2. Percentage relative abundance (%RA) and percentage distribution (%D) of all N-glycans of IgM 85 measured by MALDI-TOF MS of two independent runs i.e. Run 1 and Run 2

		Run 1		Run 2	
N-glycan masses	**Structure**	**%RA**	**%D**	**%RA**	**%D**
High mannose					
1579.78	$GlcNAc_2Man_5$	13.9%	20.7%	10.7%	16.0%
1783.88	$GlcNAc_2Man_6$	47.8%	71.1%	53.1%	79.2%
1987.98^	$GlcNAc_2Man_7$	4.5%	6.6%	2.5%	3.7%
2192.08^	$GlcNAc_2Man_8$	1.1%	1.6%	0.7%	1.1%
	Subtotal	**67.3%**		**66.9%**	
Biantennary complex type					
1590.80	$GlcNAc_2A_1$	0.6%	3.4%	0.5%	2.5%
1835.93	$FcGlcNAc_2A_2$	2.5%	14.1%	3.7%	17.9%
2040.03	$FcGlcNAc_2A_2G$	0.7%	3.7%	0.8%	3.9%
2186.08	$FcGlcNAc_2A_1G_1S'_1$	0.3%	1.7%	0.3%	1.4%
2244.13	$FcGlcNAc_2A_2G_2$	0.5%	3.1%	0.5%	2.2%
2285.15	$FcGlcNAcA_2BG_1$	0.4%	2.1%	0.4%	1.8%
2448.22^	$FcGlcNAc_2A_2G_3$	0.8%	4.5%	0.7%	3.3%
2461.22^	$GlcNAc_2A_2G_2S$*1	0.6%	3.1%	0.5%	2.6%
2478.24^	$GlcNAc_2A_2G_4$	0.3%	1.8%	0.2%	1.1%
2489.25	$FcGlcNAc_2A_2BG_2$	0.5%	2.9%	0.5%	2.6%
2605.30	$FcGlcNAc_2A_2G_2S_1$	0.5%	2.7%	0.5%	2.3%
2635.31^	$GlcNAc_2A_2G_3S_1$	3.7%	21.2%	5.3%	25.8%
2652.32^	$FcGlcNAc_2A_2G_4$	1.1%	6.4%	0.8%	4.1%
2665.32	$GlcNAc_2A_2G_3S'_1$	0.8%	4.4%	0.8%	3.9%
2809.40^	$FcGlcNAc_2A_2G_3S_1$	0.8%	4.4%	0.7%	3.4%
2839.41^	$FcGlcNAc_2A_2G_3S'1$	2.4%	13.4%	3.2%	15.4%
2852.40	$GlcNAc_2A_2G2S'_2$	0.3%	1.7%	0.3%	1.3%
2996.48	$FcGlcNAc_2A_2G_2S_1S'_1$	0.2%	0.9%	0.1%	0.6%
3026.49	$FcGlcNAc_2A_2G_2S'_2$	0.8%	4.4%	0.8%	3.7%
	Subtotal	**17.6%**		**20.5%**	
Triantennary complex type					
2693.35^	$FcGlcNAc_2A_3G_3$	0.3%	3.5%	0.2%	3.3%
2897.45^	$FcGlcNAc_2A_3G_4$	0.3%	3.8%	0.2%	3.1%
2910.45	$GlcNAc_2A_3G_3S$*1	0.2%	2.4%	0.1%	1.3%

3054.52^	$FcGlcNAc_2A_3G_3S_1$	0.2%	2.7%	0.2%	2.3%
3084.54^	$GlcNAc_2A_3G_4S_1$	1.2%	13.3%	1.0%	13.2%
3101.55^	$FcGlcNAc_2A_3G_5$	0.6%	6.2%	0.1%	1.6%
3258.62^	$FcGlcNAc_2A_3G_4S_1$	0.4%	4.6%	0.3%	3.8%
3271.62	$GlcNAc_2A_3G_3S_1S'_1$	0.1%	1.6%	0.1%	1.2%
3288.64^	$GlcNAc_2A_3G_5S_1$	1.2%	13.7%	1.5%	21.0%
3301.63^	$GlcNAc_2A_3G_3S'_2$	0.1%	1.6%	0.1%	1.6%
3305.65^	$FcGlcNAc_2A_3G_6$	0.5%	5.9%	0.2%	2.6%
3445.71^	$GlcNAc_2A_3G_4S_2$	0.3%	3.0%	0.2%	2.1%
3462.72^	$FcGlcNAc_2A_3G_5S_1$	0.5%	5.7%	0.3%	4.4%
3475.72^	$FcGlcNAc_2A_3G_3S'_2$	0.8%	9.2%	0.8%	10.5%
3492.73^	$FcGlcNAc_2A_3G_5S'1$	0.6%	7.2%	0.5%	6.4%
3619.80^	$FcGlcNAc_2A_3G_4S_2$	0.1%	1.7%	0.1%	1.4%
3649.81	$FcGlcNAc_2A_3G_4S'_1S_1$	0.3%	3.0%	0.2%	2.5%
3679.82	$FcGlcNAc_2A_3G_4S'_2$	0.5%	5.3%	0.7%	9.3%
3806.88	$FcGlcNAc_2A_3G_3S2S'_1$	0.1%	1.0%	0.1%	1.4%
3836.89	$FcGlcNAc_2A_3G_3S1S'_2$	0.1%	1.4%	0.1%	2.0%
3866.90	$FcGlcNAc_2A_3G_3S'_3$	0.3%	3.2%	0.4%	4.9%
	Subtotal	**8.9%**		**7.2%**	
Hybrid and others					
1824.91	$GlcNAc_2A_1Man_5$	0.8%	13.1%	0.3%	4.6%
2029.01	$GlcNAc_2A_1Man_5G_1$	0.8%	12.6%	0.4%	5.2%
2216.09	$GlcNAc_2A_1Man_4G_1S'_1$	0.7%	11.2%	0.7%	9.8%
2420.19	$GlcNAc_2A_1Man_5G_1S'_1$	3.4%	55.3%	3.7%	51.8%
2418.21*	$FcGlcNAc_2A_2FcG_2$	0.3%	4.3%	0.1%	1.4%
3041.53*	$FcGlcNAc_2A_3Fc_2G_3$	0.2%	3.4%	0.1%	1.3%
	Subtotal	**6.2%**		**5.3%**	
	Total	**100.0%**		**100.0%**	

Note: A_1, A_2 and A_3 represent trimannosyl core with one, two and three GlcNAc sugar units (or antenna), respectively; whereas Fc, G, S, S' represent fucose, galactose, Neu5Ac and Neu5Gc sugar units, respectively; and B represents a bisecting GlcNAc sugar that is attached β1-4 to A trimannosyl core. Subscript of each sugar shows the number of sugar units that are attached. *Mass peaks of 2418.21 3041.53 are biantennary complex type N-glycans with a terminal Le^a epitope, however the presence of these structures requires further confirmation by western blot. ^N-glycan masses require further fragmentation by MALDI-TOF-TOF MS/MS to elucidate the structure.