Development and characterisation of recognition molecules for the specific identification of non-human glycans

A thesis submitted to the National University of Ireland, Galway for the degree of Doctor of Philosophy (Ph.D)

By

Shashank Sharma, B.Tech, MSc,

Glycoscience Group,
National Centre for Biomedical Engineering Science,
National University of Ireland, Galway

January 2015

Supervisor: Prof. Lokesh Joshi

Contents

Declaration

I certify that this thesis has not been previously submitted as an exercise for a degree at National University of Ireland, or at any other university, and I further declare that the work embodied in it is my own.

Shashank Sharma

Acknowledgements

I would like to acknowledge first and foremost Prof. Lokesh Joshi, my PhD advisor. His belief, positive attitude and scientific guidance helped me to complete my research work. Prof. Joshi gave me all the independence to lead this project and explore the different aspect of PhD life.

I would like to thank Dr. Marian Kane for her immense support, patience and guidance not only in the experimental part but taking time out to help me improve my scientific writing skills. My GRC committee had Prof. Samali as one of the member, I would like to thank him for reviewing my work each year, positively criticizing and guiding me during the research. I also thank my research collaborator Prof. Leif Eriksson, who always took out his time in Sweden and helping me to carry out my research.

I spent my graduate life at a great learning and helpful environment. A special thanks goes to all members of the Glycoscience Group, who helped me working towards the success of my research. I thank Dr. Michelle Kilcoyne and Dr. Iain Shaw for their extended support in thesis corrections and scientific guidance. I would like to specially thank Dr. Jared Gerlach for his prompt help right from the beginning of my PhD till the thesis completion.

I would like to thank Science Foundation Ireland for funding my research.

I would also like to take this opportunity to thank my parents and family who have contributed in all my aspects of life. Last but never been the least, a big thanks to my lovely wife Satbir who supported me in every possible ways to continue my research, I owe her this research thesis.

Abstract

Glycosylation is an enzymatic process through which glycans attach to proteins, lipids and other organic molecules with huge structural diversity. This is a common post-translational modification found on proteins and cell surfaces and has significant effects on cellular functions, such as, cell adhesion, migration, differentiation, inflammation, immunity, structural folding and bimolecular interactions. Glycans of non-human origin, that are not naturally synthesised or expressed in humans but are introduced into the body through dietary intake or through biological therapeutic drugs that have been derived in non-human production systems have the potential to elicit an immune response. These responses can vary from minor inflammation to anaphylaxis.

There is a constant need for the identification and characterisation of specific recognition molecules that can be incorporated into cost effective and high-throughput methods to specifically detect and quantify non-human glycans. These assays may have utility in development and quality control of recombinant biopharmaceuticals that express these molecules. The development and characterisation of recognition molecules to two important non-human glycans: Galactosyl-α-(1, 3)-Galactose (Gal-α-(1, 3)-Gal) and *N*-glycolylneuraminic acid (Neu5Gc) are described here.

Gal-α-(1, 3)-Gal is immunogenic in humans and is directly involved in hyper-acute rejection during xenotransplantation. Recombinant therapeutic molecules generated in non-human cell culture systems such as CHO, NS0 and SP2/0 cell lines can potentially contain Gal-α-(1, 3)-Gal. The structural basis of the interaction of the anti-Gal-α-(1, 3)-Gal epitope with a previously developed scFv recombinant antibody fragment was analysed and key binding regions and properties among the amino acid residues critical in the recognition and binding of scFvs' against the target carbohydrate epitopes were determined.

Like Gal-α-(1, 3)-Gal, Neu5Gc can also be added to recombinant biopharmaceutical molecules when using CHO, NS0 and SP2/0 cell lines. There is also evidence that Neu5Gc-containing glycoconjugates occur frequently in cancer patients, and that tumour associated Neu5Gc facilitate tumour progression and inflammation. There

are very few recognition molecules produced with high specificity for Neu5Gc that can be used to assay for this molecule.

Aptamers are DNA or RNA oligonucleotides which have the capacity to bind to a variety of molecules with high affinity and specificity. Here I describe the development of DNA aptamers that show high affinity and specificity allowing for discrimination between the non-human Neu5Gc from the human Neu5Ac form. These molecules will have great application in biopharmaceutical production as part of a quality control workflow in allowing for the identification of potentially harmful molecules.

List of abbreviations

Å:	Angstrom
AA:	Amino acid
BLAST:	Basic Local Alignment Search Tool
BSA:	Bovine Serum Albumin
CDR:	Complementarity Determining Region
DNA:	Deoxyribonucleic Acid
EDC:	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
ELISA:	Enzyme Linked Immunosorbent Assay
ELLA:	Enzyme Linked Lectin Assay
ER:	Endoplasmic Reticulum
FASTA:	Fast Alignment
FR:	Framework region
Fuc:	Fucose
Gal:	Galactose
GalNAc:	*N*-acetylgalactosamine
GB/VI:	Generalised Born / Volume Integral
Glc:	Glucose
GlcNAc:	*N*-acetlyglucosamine
GS-I-B4:	*Griffonia simplicifolia* I B4 Isolectin
GT:	Glycosyltransferases
HPLC:	High Performance Liquid Chromatography
HRP:	Horseradish Peroxidase
K:	Kelvin
K_a:	Association constant
kcal/Mol:	Kilo Calorie per Mole
K_d:	Dissociation constant
kDa:	Kilo Dalton
Lac:	Lactose
Man:	Mannose
MD:	Molecular Dynamics
MES:	2-(N-morpholino)ethanesulfonic acid buffer

mM:	Milli Molar
MMFF:	Merck Molecular Force Field
MOE:	Molecular Operating Environment
NaOAc:	Sodium Acetate
Neu5Ac:	*N*-acetylneuraminic acid
Neu5Gc:	*N*-glycolyneurminic acid
NHS:	*N*-Hydroxysuccinimide
nM:	Nano Molar
NPT:	Isobaric/Isothermal
PAA:	Polyacrylic Acid
PBS:	Phosphate Buffer Saline
PCR:	Polymerase Chain Reaction
pM	Pico Molar
PTM:	Post Translational Modification
RMSD:	Root Mean Square Deviation
RNA:	Ribonucleic Acid
RT:	Room Temperature
s:	Second
scFv:	Single chain variable region
SE:	Standard Error
SELEX:	Systematic Evolution of Ligands by Exponential enrichment
SEM:	Standard Error of Mean
SOC:	Super Optimal broth with Catabolite repression
sp-bt:	Spacer-Biotin
SPR:	Surface Plasmon Resonance
ssDNA:	Single Stranded DNA
TEMED:	N,N,N',N'-Tetramethylethylenediamine
V_H:	Heavy Chain
V_L:	Light Chain
X-gal:	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside
Xyl:	Xylose
µM:	Micro Molar

List of figures

List of tables

Table of Contents

Chapter-1: Introduction

1.1. Carbohydrates

Carbohydrates have been shown to be ubiquitous and are present as the major component of the cell surface and the extra cellular matrix. These carbohydrates coat the cell surface and act as a barrier on the cell membrane where they provide the first point of cell-to-cell communication and the cellular recognition process (Varki, 1993, Osborn et al., 2004). The interaction of cell surface carbohydrate with a protein ligand aids in the immuno and molecular recognition process, where they can act as disease determinants and can be used as biomarkers to target such diseases (Sharon and Lis, 1993, Svarovsky and Joshi, 2008, Rudd et al., 2001). The carbohydrate part of many glycoproteins and glycoconjugates possesses receptors that bind to viruses and play an important role in viral infections (Olofsson and Bergstrom, 2005). In summary, the biological roles of carbohydrates in humans vary from development, growth, and function to survival (Varki et al., 2009a, Varki, 1993).

In the central dogma, DNA is transcribed into RNA, and then translated with a mRNA based template into Protein: DNA- RNA- Protein, which results in a limited diversity of proteins (Crick, 1970). However, DNA does not encode for carbohydrates, and there is massive diversity of carbohydrate structures present on the cell surface. The synthesis of complex carbohydrate structures from basic sugar building blocks is driven by a diverse pathways of enzymes. The structures can be linear or branched with a diversity of linkages and branching patterns within the carbohydrates, and this allows a huge amount of information in these dense structures (Laine, 1994, Schachter, 2001). The diversity and branching of these structures vary from cell to cell at different levels of organisation and therefore can play many different biological roles.

1.1.1. Carbohydrate diversity

Carbohydrates contain the basic elements of carbon, hydrogen and oxygen mostly in the ratio of $C_x(H_2O)_y$, where x and y can be different. Natural carbohydrates are generally built with simple carbohydrate (monosaccharide) $C_x(H_2O)_y$ where x could be three or more.

The carbohydrates are defined as polyhydroxyketones, polyhydroxyaldehyde and their derivatives, while the larger derivatives can be hydrolysed in to smaller units

(Varki and Sharon, 2009). The carbohydrates or the saccharides are of four different types: monosaccharides, disaccharides, oligosaccharides and polysaccharides.

Monosaccharides are simpler glycans that cannot be hydrolysed further such as Fucose (Fuc), Mannose (Man), Galactose (Gal), Glucose (Glu), etc. The disaccharides are the product of condensation of two monosaccharides such as Lactose (Lac) made up of Glu and Gal. Oligosaccharides are formed from 3 to 9 monosaccharides, while polysaccharides are sugars formed through more than 10 monosaccharides, such as cellulose and chitin. The nature of carbohydrate polymers or polysaccharides is governed by the positions of the monosaccharides linked with e other carbohydrate molecule, also their order of branching and the type of residues linked to them. It has been calculated that the number of possible hexasaccharide structures for mammalian glycans after permutations and combinations of their branching and linkage can be up to 10^{12}, however, not all of these structures exist in nature (Laine, 1994). Many saccharide structures differ only in the orientation of the hydroxyl groups (-OH); these monosaccharide structures can be classified into aldose or ketose forms, where an aldose consists of an aldehyde group (CH=O) formed through a C=0 bond (carbonyl group) at the end of the molecule. The ketose consists of one ketone group per molecule. The Fisher projection introduced by Herman Emil Fisher in 1891 is a representation of the three-dimensional saccharides into linear forms. The family tree of aldoses with their stereoisomers is shown in Figure 1.1.

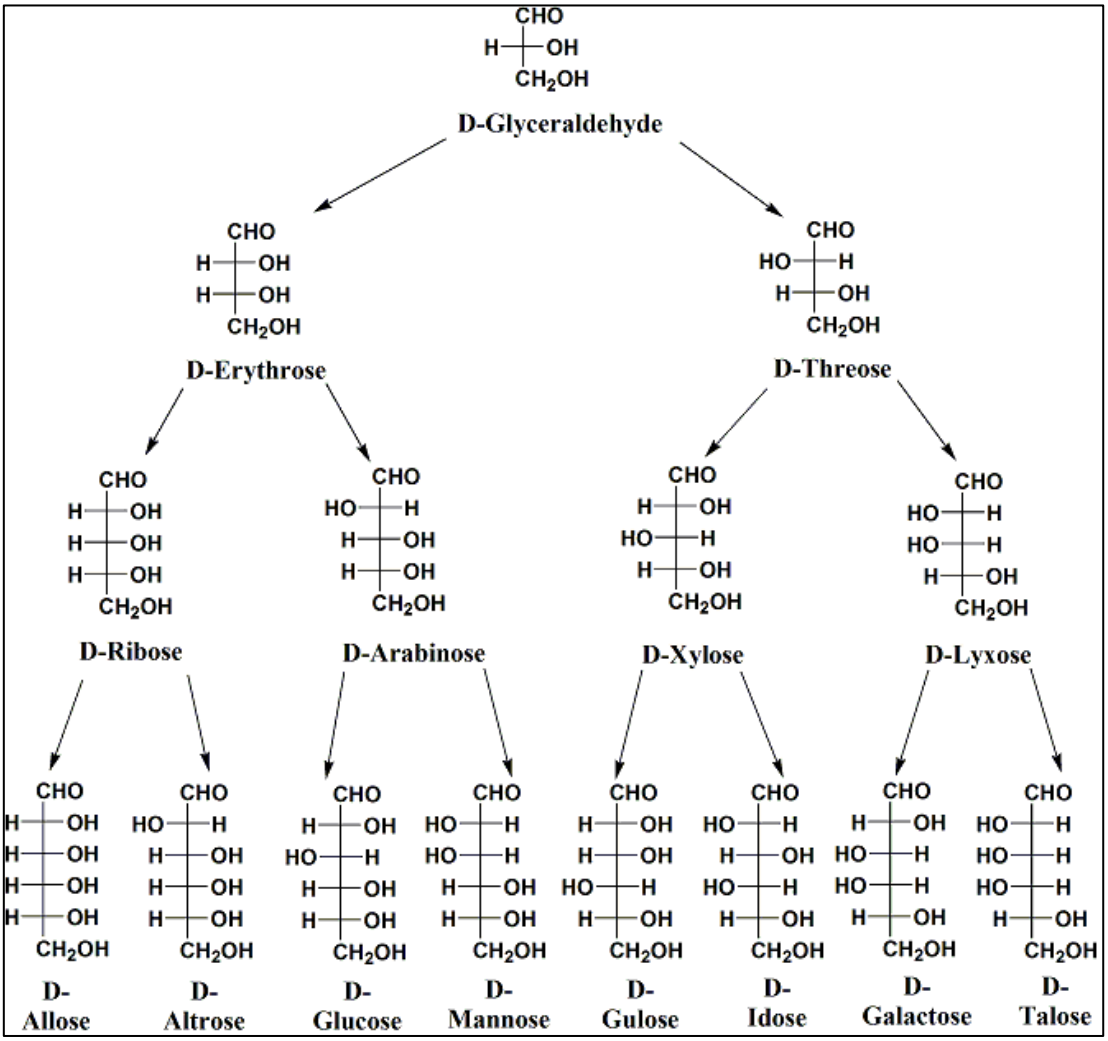


Figure 1.1: Family tree of D-aldoses ranging from tri carbon to hexa carbon monosaccharides

The straight-chain polyhydroxyl aldehydes and ketones are converted in to cyclic monosaccharide residues by the intramolecular cyclisation of linear saccharides (Figure 1.2). The linear D-Galactose cyclises into either the 5 carbon furanose (D-Galactofuranose) or the 6 carbon pyranose form (D-Galactopyranose). The monosaccharides can transform into both linear or branched form arise due to nucleophilic attack with the two possible sterochemical isoform as α and β form resulted in huge complexity of glycans. For example, the two possible orientations for both D-Galactofuranose and D-Galactopyranose leads to α and β form (Figure 1.2). These α and β form are anomers of each other. These anomers are epimers and they differ to each other in the configurational carbon at C-1 in aldoses, and at C-2 in case of ketoses (Figure 1.3). These cyclic monosaccharides are joined together to form a disaccharide through a glycosidic bond (Figure 1.4).

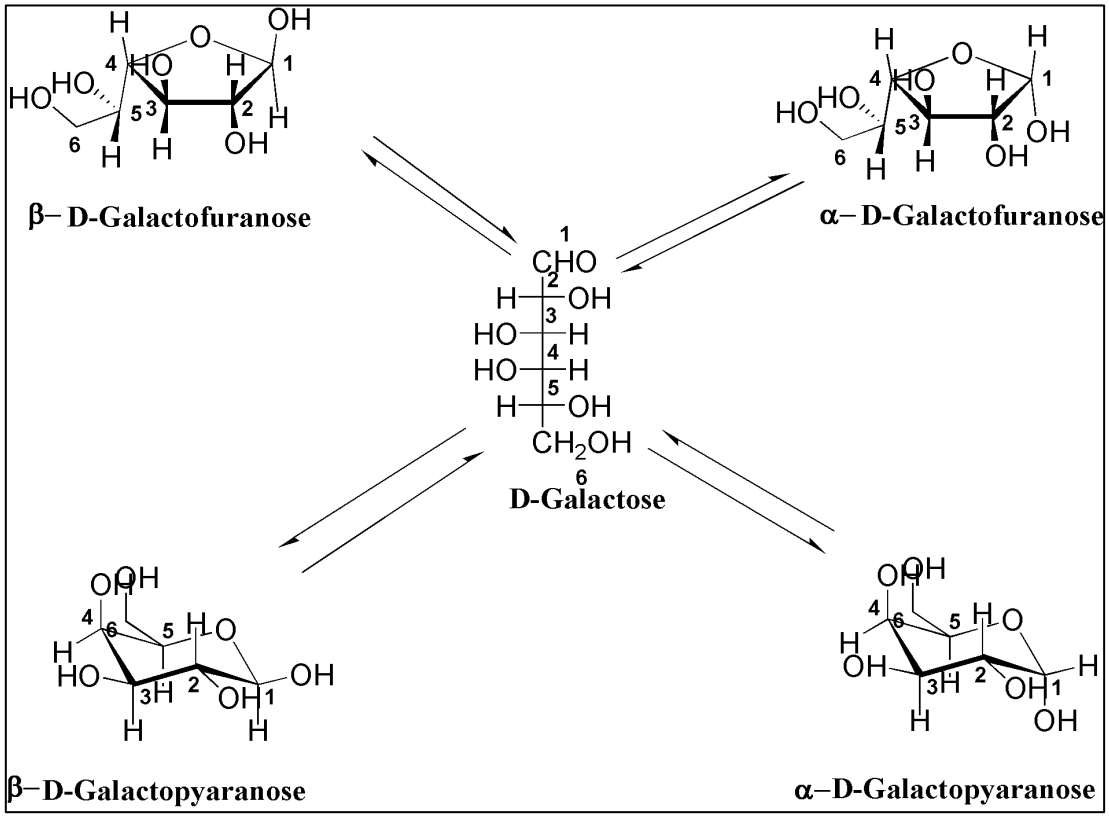


Figure 1.2: Cyclisation of aldo-D-Galactose to furanose and pyranose forms.

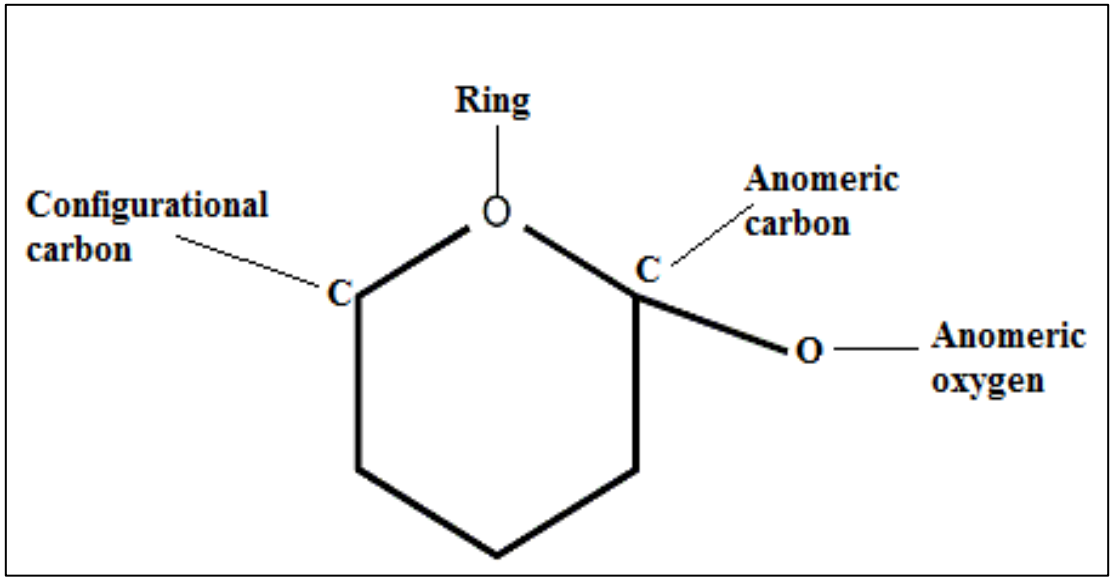


Figure 1.3: A pyranose ring showing configurational carbon, anomeric carbon and anomeric oxygen.

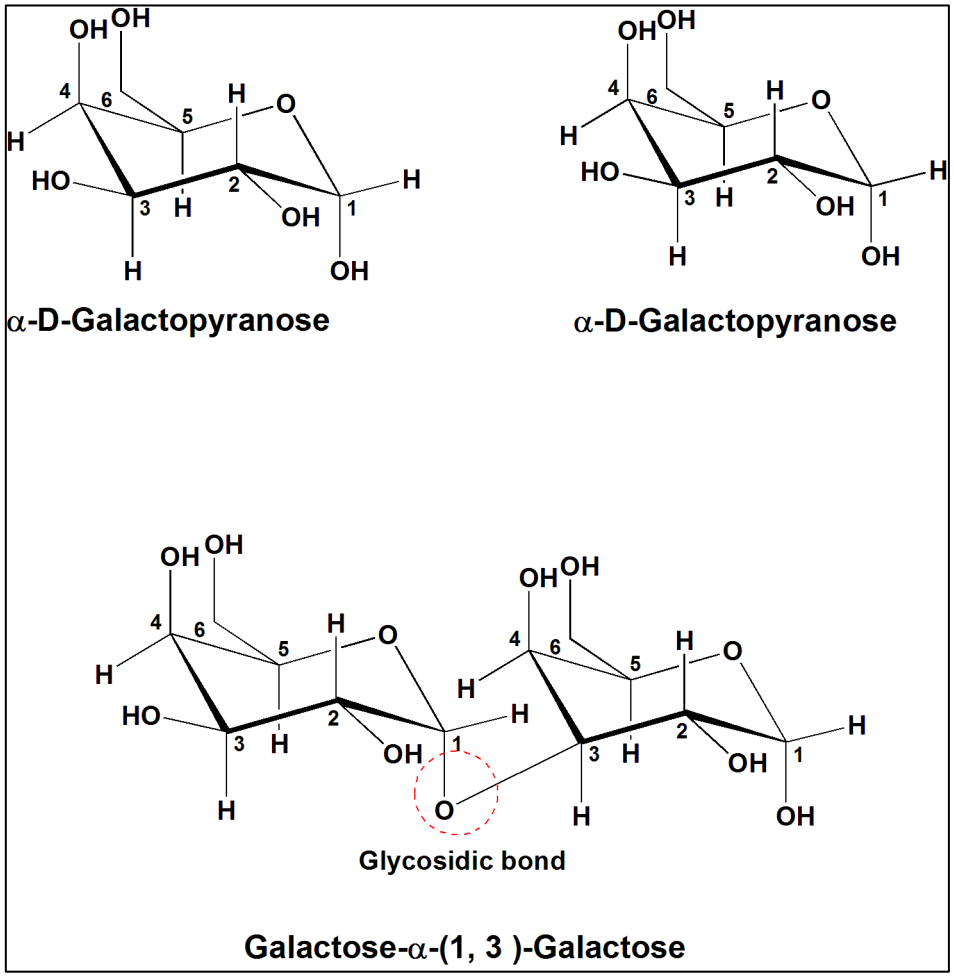


Figure 1.4: Linking of two separate monosaccharides α-D-Galactopyranose to form disaccharide Galactose- α-(1, 3)-Galactose

1.2. Post-translational modifications

In eukaryotic cells, the majority of proteins are modified during or soon after translation called as post-translational modifications (PTMs). PTMs present on proteins and cell surfaces have to role to play in cell adhesion, migration, differentiation, inflammation, immunity and structurally, folding, and bimolecular interactions (Walsh et al., 2005, Freeze and Schachter, 2009, Boscher et al., 2011, Seo and Lee, 2004). There are more than 250 different types of PTMs found on proteins such as acetylation, sulphation, phosphorylation, glycosylation, etc.

1.2.1. Glycosylation

One of the most prevalent PTM that is focussed in this thesis is glycosylation; it is an enzymatic process through which oligosaccharide structures (glycans) attach to

proteins, and lipids. Usually, glycosylation of proteins occurs in the secretory pathway and is carried out in discrete biosynthetic steps divided between the endoplasmic reticulum (ER) and Golgi apparatus (Varki et al., 1999, Sesma et al., 2009). The secretory pathway in the cells, starting with the ER and ending at the *trans*-Golgi, delivers properly folded and glycosylated proteins to the cell surface which is important in cell to cell communication and the development and homeostasis of all multi-cellular organisms (Dennis et al., 2009, Boscher et al., 2011, Gerlach et al., 2012, Hulsmeier et al., 2011). Unlike nucleic acids and proteins, glycan structures are neither template driven, nor the genes determine them. They follow a branching pattern, where they could be linear or branched in structure.

1.2.2. The *N*-linked glycosylation process

N-linked glycosylation is the most common and frequent form of PTMs in eukaryotes and vital in the process of proper protein folding and cell interactions (Gerlach et al., 2012, Helenius and Aebi, 2001). Since *N*-linked glycosylation starts while the polypeptide chain is still being folded in the ER and continues in the Golgi after the protein is folded, it is a co-translational as well as post-translational modification. The initial synthesis steps of *N*-linked oligosaccharide occur in the ER that are common to all eukaryotic cells. However, the specific glycan chains (glycoforms) found in different organisms arise due to differences in the oligosaccharide processing steps that take place in the Golgi complex. *N*-linked attachment of glycans requires the consensus amino acid sequence Asn-X-Ser/Thr to be contained within nascent proteins. Herein the *N*-linked glycosylation becomes different from *O*-glycosylation where the latter doesn't require fully formed recognition domain. In this consensus sequence, the amino acid, X, following Asn can be occupied by any amino acid except Pro. Transfer of the pre-formed fourteen-sugar structure ($GlcNAc_2Man_9Glc_3$) *en bloc* to the appropriate asparagines of the protein initiates *N*-linked glycosylation in the ER (Ruddock and Molinari, 2006). All further steps of glycan processing in the ER are catalysed by different conserved glucosyltransferases and glycosidases that help in correct folding of the newly synthesised proteins. During this process two molecular chaperon (proteins) that participate in the ER quality control, one soluble and one membrane bound, calreticulin and calnexin, assess the state of protein folding prior to allowing further

glycosylation modification. Both calreticulin and calnexin are carbohydrate-binding proteins (lectins) and their chaperoning mechanism relies on their ability to bind to the oligosaccharide moiety in the ER and hence assist in protein folding by associating with monoglucosylated oligosaccharides on the nascent protein. After removal of first glucose through Glucosidase I and second and third glucose residues from the oligosaccharide chains through Glucosidase II, UDP-glucose:glycoprotein glucosyltransferase (UGGT) acts as a folding sensor by interacting with both the glycan and the backbone of the protein being folded.

If the protein is still unfolded, iterative reglucosylation by UGGT and deglucosylation by glucosidase II occur ensuring the correct folding of the target protein. If it still fails to fold properly, the target protein is sent for degradation with additional chaperones such as EDEM and OS9 (Chen et al., 2011). Removal of the last glucose signifies proper protein folding and thereafter exit to Golgi. After the removal of glucose residues in ER, α-(1, 2) mannosidase removes one mannose to generate $GlcNAc_2Man_8$ structure (Avezov et al., 2008). Glycoproteins bearing $GlcNAc_2Man_8$ structure are then transported to Golgi complex where, further trimming of glycan residues take place followed by addition of monosaccharides, which is carried out by sugar- and linkage- specific glycosyltransferases. In mammalian cells, *N*-linked oligosaccharide chains often possess terminal sialic acid residues and penultimate galactose residues at the non-reducing end (Figure 1.5). Specific enzymes, such as sialyltransferases and galactosyltransferases, carry out these reactions respectively.

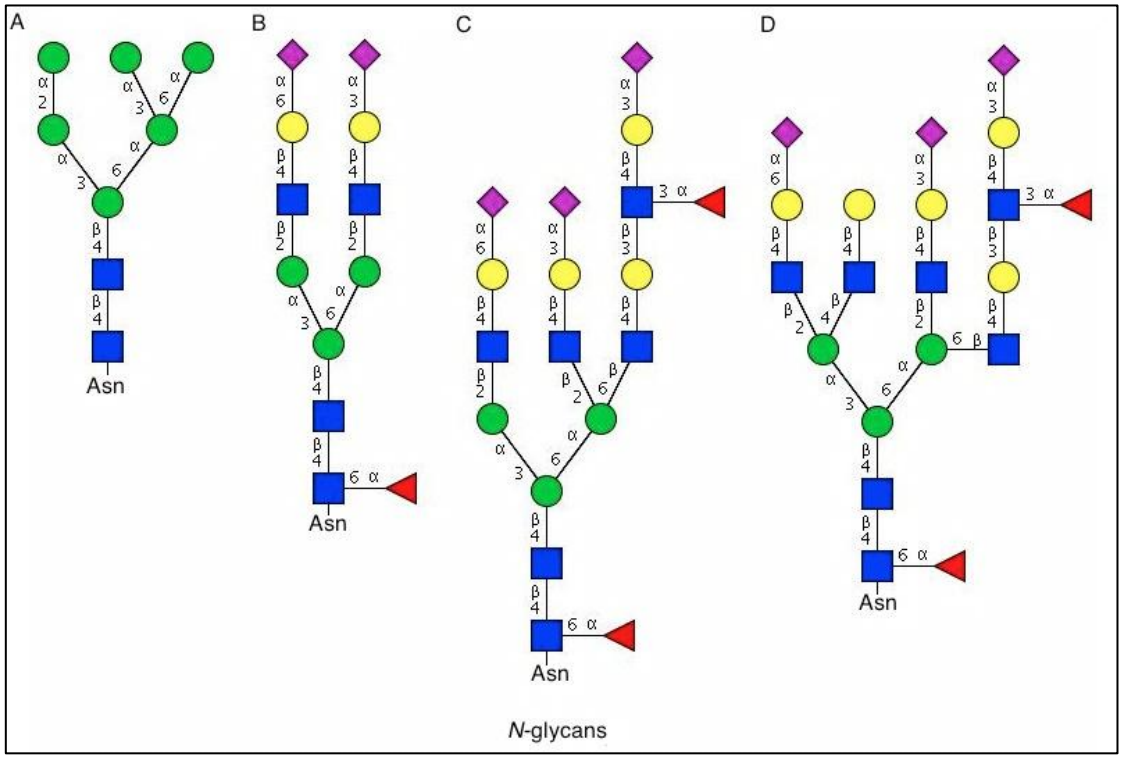
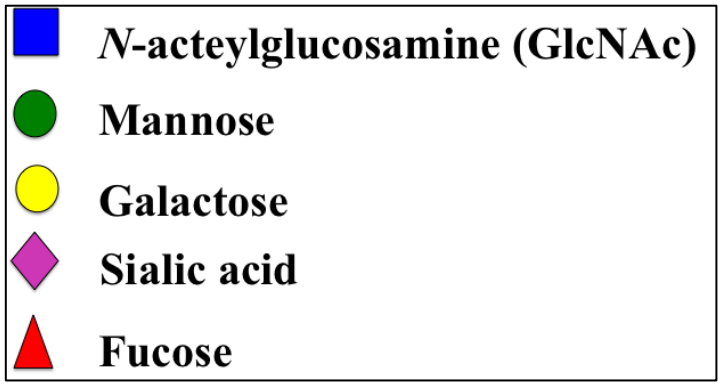


Figure 1.5: Some common *N*-glycan structures linked with Asn: (A) High-mannose, (B) Bi-antennary complex with core fucosylation and terminal sialylation, (C) tri-antennary complex demonstrating both core and distal fucosylation, and (D) tetra-antennary complex with similar features to C. All the structures were made on Glycoworkbench (Ceroni et al., 2008) using the Consortium for Functional Glycomics (CFG) (www.functionalglycomics.org, 2014) notations for linkages and monosaccharide symbols. Symbolic representation of common monosaccharides is shown below:



1.2.3. The *O*-linked glycosylation process

The glycan addition to the hydroxyl group of serine or threonine amino acid residues is referred to as *O*-linked glycosylation (Stanley, 2011, Brockhausen et al., 2009, Peter-Katalinic, 2005). *O*-linked glycosylation is a post-translational modification, where *O*-glycosylation (initiation and maturation) of fully folded proteins occurs only in the Golgi complex. Each monosaccharide is added in a sequential manner to generate linear or branched oligosaccharide structures.

There are different types of *O*-linked glycosylation in eukaryotic cells but among them mucin type glycosylation is the most common in mammals, where addition of *N*-acetyl-galactosamine (GalNAc) is to serine/threonine residue on the protein backbone (Figure 1.6) (Brockhausen et al., 2009, Clausen and Bennett, 1996). This core GalNAc residue is attaches on amino acid residues by 24 different *N*-acetyl-α-galactosaminyltransferases (*O*-GalNAc transferases) and then further extended by galactosyltranferases to form higher-order *O*-linked structures (Dube et al., 2006).

There are total 8 core structures in *O*-GalNAc glycans, of which four are the common glycan cores, while four other are additional designated cores. The four common cores (core 1-4) are present in mucins and glycoproteins are shown in Figure 1.6: A-D. The core 1 and core 2 structures are more common in serum glycoproteins and mucins, core 3 and core 4 are present in secreted mucins of salivary glands, colon and bronchi respectively, while core 5-8 only rarely occur in human tissues and mucins. The simplest of *O*-GalNAc glycan is the Tn antigen (GalNAc-Ser/Thr), while the most common form is the core 1 or T antigen, which forms the core of many longer structures. Both the Tn and T antigen are often sialylated and antigenic in nature (Brockhausen et al., 2009). The O-GlcNAc modification where serine or threonine hydroxyl moieties are modified by β-linked *N*-acetylglucosamine was discovered in 1984 (Torres and Hart, 1984). This modification is present on residues of both nuclear and cytosolic proteins and is highly abundant in eukaryotes (Wells et al., 2001). Additional *O*-linked glycosylation modifications include *O*-Glu, *O*-Man and *O*-Fuc (Figure 1.6: E-G) are unusual forms of *O*-linked glycosylation where Glucose, Mannose and Fucose are directly attached to proteins through an *O*-linkage (Luther and Haltiwanger, 2009).

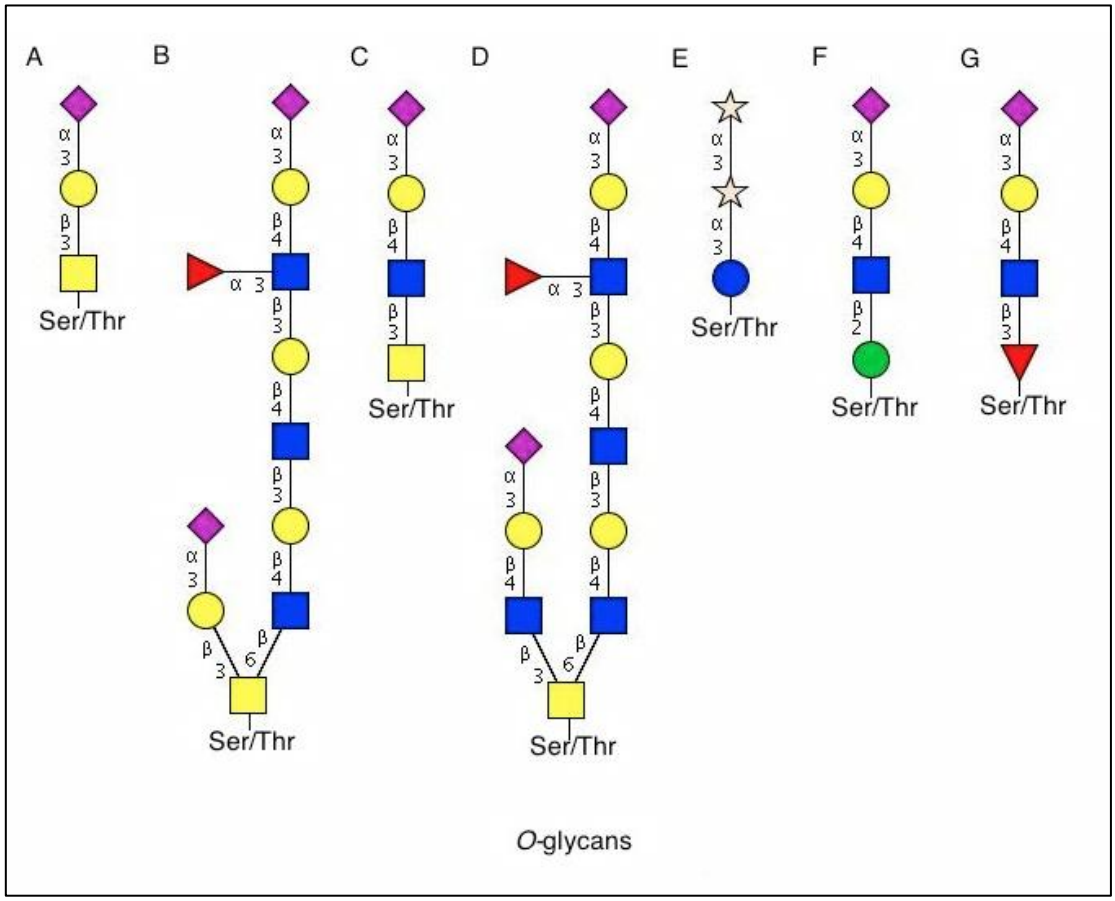
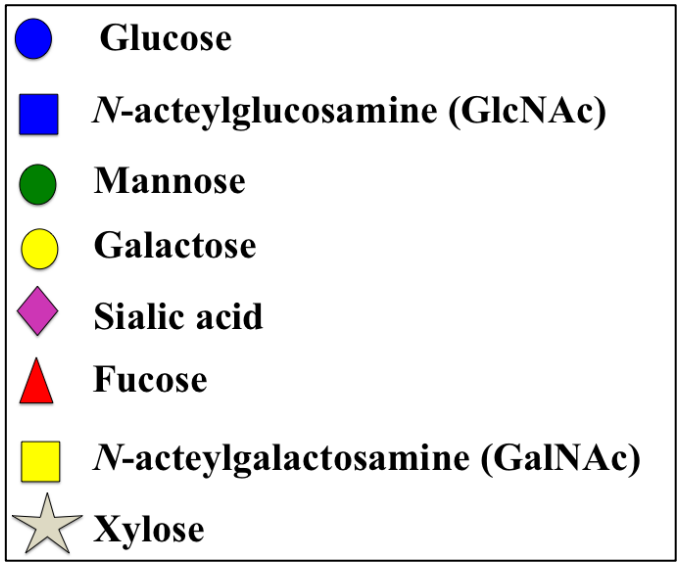


Figure 1.6: Examples of *O*-glycan structures possessing core structures. (A) core 1, (B) core 2, (C) core 3, (D) core 4, (E) *O*-Glu, (F) *O*-Man and (G) *O*-Fuc. All the structures were made on Glycoworkbench (Ceroni et al., 2008) using CFG pattern.



1.2.4. Glycolipids

Glycolipids are composed of hydrophilic glycans attached through the glycosidic linkage to the hydroxyl group of a hydrophobic ceramide, a lipid moiety composed of both a waxy core and long fatty acid chains. Ceramide is synthesised in the smooth ER and transported into Golgi for the synthesis of glycolipids (Maccioni et al., 2002). Within the Golgi complex, enzymes as glycosyltransferases, sugar transporters and synthesis intermediates as donor sugar nucleotide (all associated with Golgi membrane structures) act to assemble complete glycolipids. Glycolipids play an important role in pathways of cellular physiology and the alteration of sugar arrangement during glycosylation in these pathways may be beneficial to cancer cells. The incomplete synthesis of the normally-expressed glycan disialyl Lewis^{a} leads to increased expression of sialyl Lewis^{a}, the over expression which increases cancer cell adhesion and motility, and is associated with increased metastasis (Durrant et al., 2012, Kannagi, 2003). The most common glycolipids are glycosphingolipids, and a majority of these are exclusively localised in the external surface of the plasma membrane (Thompson and Tillack, 1985), whereas few glycolipids are incorporated into the membrane of the Golgi apparatus (Fleischer, 1977, Lannert et al., 1998). Gangliosides are other important glycolipids that also contain sialic residues. Gangliosides are glycosphingolipids with additional *N*-acetylneuraminic acid (Neu5Ac) or *N*-glycolylneuraminic acid (Neu5Gc) on their carbohydrate surface. Ernst Klenk in 1942 coined the term gangliosides to the lipids isolated from ganglion cells of the brain (Ledeen, 1966). They have been shown to be over-expressed in tumours of neuroectodermal origin and in malignant melanoma (Hamilton et al., 1993). Gangliosides play a prominent role in cell-cell interaction, immune recognition (Hakomori, 1981) and cell surface recognition (Schnaar, 1991).

1.3. Roles of glycans

Glycosylation is an important process of the ER's protein quality control system which helps in identifying improperly folded proteins for their systematic recycling. *N*- and *O*-oligosaccharide variants on glycoproteins (glycoforms) can alter protein activity or function. The presence of glycan moieties on glycoproteins are involved in a wide array of functions including developmental, biological and pathological processes ranging from increasing the solubility and stability of proteins to extending

their circulatory half-life in the serum and also have roles in most chronic and infectious diseases (Varki et al., 2009a).

1.3.1. Glycans in cancer

Like normal cells, tumour cells demonstrate variations in glycosylation, which reflect states of activation, growth, and development. Glycosylation represents one of the most common PTM in current cancer research (Hakomori, 1985, Hakomori, 2002), and aberrant glycosylation in glycoproteins and glycosphingolipids is a result of initial oncogenic transformation, as well as a key event in induction of invasion and metastasis (Lau and Dennis, 2008).

The famous Warren-Glick phenomenon of increasing branching in *N*-glycans has been reviewed as a molecular basis for tumourigenesis and metastasis (Kobata, 1996, Kobata, 1998). This phenomenon was first observed in sugar chains of plasma glycoproteins in malignant cells (Meezan et al., 1969). The studies following the discovery of this phenomenon confirmed that the expression of tri- and tetra-antennary complex type *N*-linked sugar chains occurred in malignant cells (Ogata et al., 1976, Warren et al., 1978).

During cancer progression and metastasis, the cellular glycosylation machinery is significantly altered leading to the aberrant expression of glycan epitopes (also called Tumour Associated Carbohydrate Antigens or TACAs) and glycan binding proteins that are either on the surface of the cells or secreted into the environment (Ohtsubo and Marth, 2006). As glycans are involved in cell-cell and cell-environment interaction, adhesion and migration, these changes alter tumour cell behaviour and the response of the physiological niche to the cells (Zhao et al., 2008). In general, cancer cells frequently display glycan epitopes at different levels or with fundamentally different structures than those observed on normal cells (Dube and Bertozzi, 2005). A few examples of these tumour associated glyco-epitopes are aberrant α-(1, 3) fucosylation, and over expression of sialyl Lewis^{a} and sialyl Lewis^{x} (Miyoshi et al., 2008, Heimburg-Molinaro et al., 2011). Also, altered sialylation on the sialyl-Tn antigen plays an important role in the development of the malignant phenotype in gastric carcinoma (Pinho et al., 2007). It has been shown that Lewis^{y} (Le^{y}) is a biomarker of breast cancer as this glycan is associated with CD44

glycoprotein over-expression which is pro-oncogenic by nature (Lin et al., 2010). All of the cancer-associated glycan structures mentioned above are shown in Figure 1.7. These Lewis glycans are terminations of larger structures with different cores and linkages. Sialyl Lewis^{a}, sialyl Lewis^{x} and Le^{y} are part of either *N*-linked or *O*-linked oligosaccharides. A few glycosphingolipids: GM2, GM3, GB3cer and Globo H and glangliosides: GD2 and GD3 are also TACA antigens and are highly expressed in cancer cells due to altered glycosyltransferase and glycohydrolase activity (Daniotti et al., 2013). A non-human sialic acid, *N*-glycolylneuraminic acid (Neu5Gc), has also been shown to exist in human carcinoma; this specific glycan and its biological importance are discussed in detail in section 1.4.2.

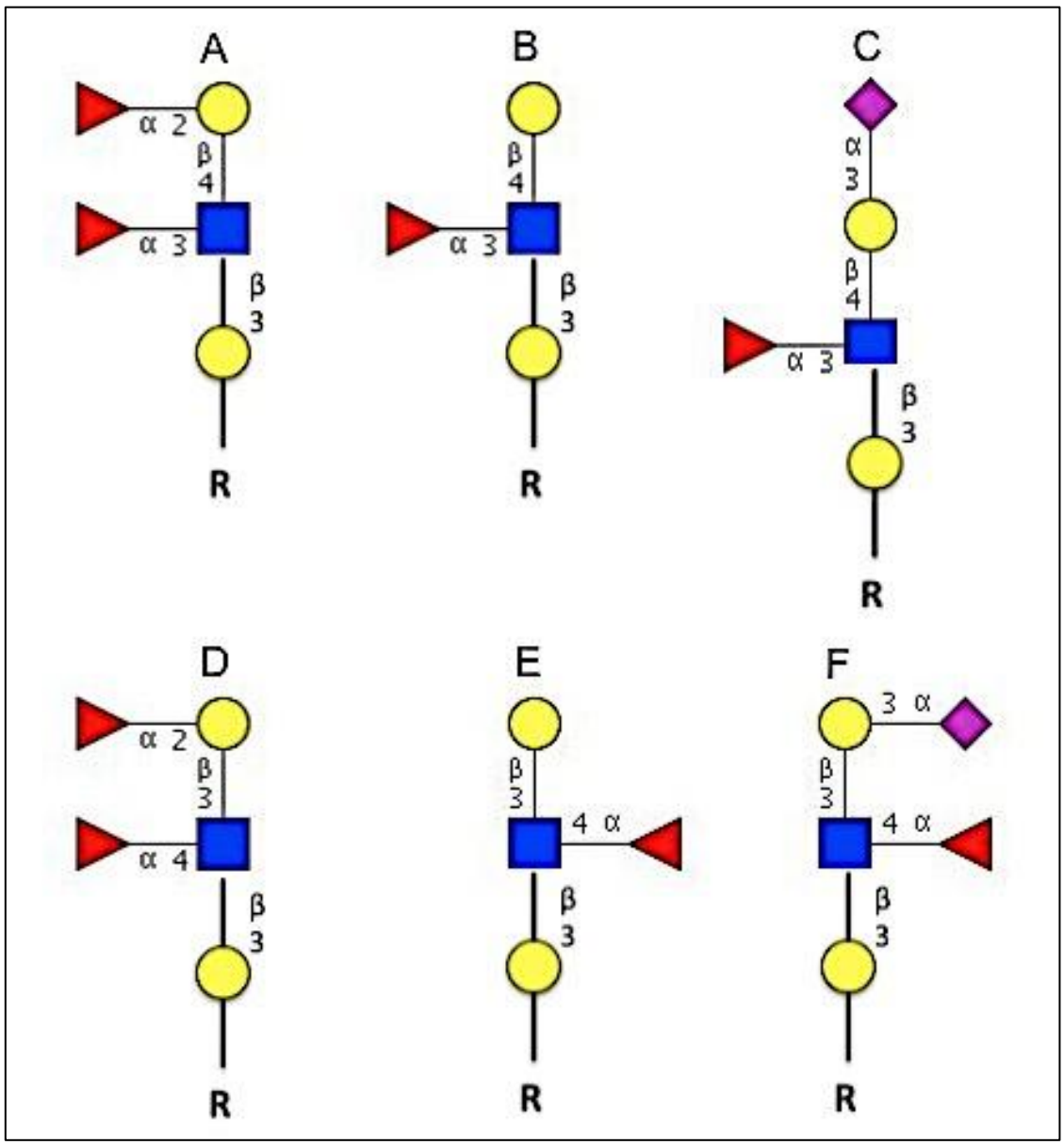
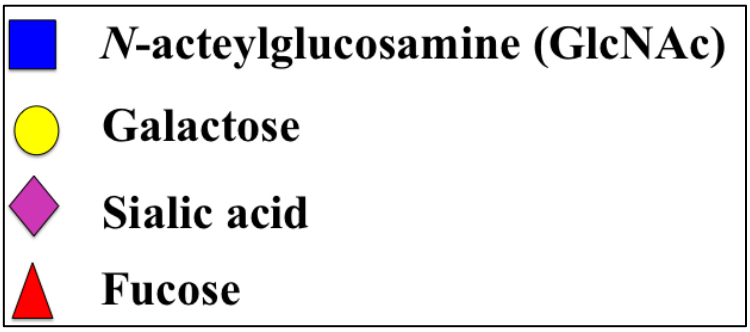


Figure 1.7: Lewis antigens associated with cancer. (A) Lewis^{y}, (B) Lewis^{x} , (C) Sialyl Lewis^{x} , (D) Lewis^{b}, (E) Lewis^{a}, and (F) Sialyl Lewis^{a}. All the structures were made on Glycoworkbench (Ceroni et al., 2008) using the CFG abbreviations for linkages and CFG symbols for monosaccharides.



1.3.2. Glycans in congenital diseases

Congenital disorders of glycosylation (CDGs) are diseases caused by defects in the biosynthesis of glycoconjugates, first reported in 1980 (Jaeken et al., 1980). These diseases were earlier called carbohydrate deficient glycoprotein syndromes (Jaeken and Carchon, 1993) but later, to include all glycoconjugate related diseases, they were renamed CDGs (Aebi et al., 1999). In all, there are 48 CDGs reviewed to date, in which 45 are diseases caused by glycoprotein glycosylation including disorders involving *N*-linked glycosylation, *O*-linked glycosylation and disorders with both *N* and *O*-linked glycosylation defects (Jaeken and van den Heuvel, 2014). They are classified into two groups based on the improper synthesis of glycans in the ER or Cytosol (Group 1) or aberrant attachment of the glycans to the glycoproteins and lipids in the ER and Golgi (Group 2) (Grunewald et al., 2002), respectively.

1.3.3. Glycans in bacterial pathogenesis

The Gram-negative pathogenic bacterium *Helicobacter pylori* (*H. pylori*) expresses adhesins: Blood group Ag-binding adhesin (Kelm et al.) and the sialic acid-binding adhesin (SabA), for recognition of carbohydrate structures in the host mucosa (Moran et al., 2011). These specific adhesins mediate adhesion of bacteria to the host tissue cells (Klemm and Schembri, 2000). The *H. pylori* adhesin BabA binds to the Lewis^{b} carbohydrate determinant (Ilver et al., 1998) and plays an important role in disease development (Gerhard et al., 1999, Aspholm-Hurtig et al., 2004). The

involvement of glycans in pathogenesis has also been elucidated in *Escherichia coli* infection in human urinary tract infection where the bacterial adhesin FimH binds to Man present on the human bladder epithelium (Krogfelt et al., 1990, Hull et al., 1981). Also, the *E. coli* PapG adhesin binds to Gal-α-(1, 4)-Gal of the host causing kidney and urinary tract infections in humans (Roberts et al., 1994, Lund et al., 1987).

1.3.4. Glycans in therapeutics

Glycosylation is found on multiple classes of protein therapeutics including, antibodies, protein hormones, growth factors, cytokines and vaccines (Walsh and Jefferis, 2006). Many of the therapeutics currently on the market are glycoprotein based where the activity and functionality of proteins is influenced by the glycans, often with individual isoforms (glycoforms) behaving differently in biological roles. Glycosylation is known to greatly influence the efficacy, half-life and potential immunogenicity of recombinant biopharmaceutical proteins (Kilcoyne and Joshi, 2007, Gerlach et al., 2011). Sialylation of glycoproteins plays an important role as the presence of sialic acid prevents exposure of Gal residues thereby protecting the protein molecule from capture and subsequent degradation. The sialylation on protein therapeutics improves protein stability and make them more resistant towards proteolytic degradation or solubility (Bork et al., 2009). Thus, the biopharmaceutical industry is required by the US Food and Drug Administration (FDA) and the European Medicines Agency (EMEA) to structurally characterise the glycans of recombinant therapeutic products throughout clinical trials and production (Fernandes, 2004, Ucakturk, 2012). A list of several glycoproteins therapeutics available on the market is mentioned in Table 1.1.

Table 1.1: The list of FDA approved glycan based therapeutics, mainly glycoprotein therapeutics: monoclonal antibodies, cytokines, hormone, clotting factors and enzymes. This table is modified from the retrieved information (Ghaderi et al., 2012, Shriver et al., 2004).

Drug name	**Disease**	**Manufacturer**
Lovenox	Thrombosis	Aventis
Avastin	Lung and bowl cancer	Genentech

Aransep	Anaemia	Amgen
Fragmin	Thrombosis	Pfizer
Actemera	Rheumatiod arthirits	Genentech
Sepragel	Anti-adhesive	Genzyme
Herceptin	Breast cancer	Genentech
Prolia	Postmenopausal osteoporisis	Amgen
Humira	Chrohn's disease, ulcerative colitis (Anti imflammatory)	Abbot laboratories
Yervoy	Metastatic melanoma	Bristol-Myers Squibb
Follisitism	Follicle development and maturation during pregnancy	Merck
Thyrogen	Hypothyroidism	Genzyme
Aransep	Anaemia	Amgen
NeoRecormon	Anaemia	Hoffman-La Roche
Kogenate	Haemophilia A	Bayer
Advate	Haemophilia A	Baxter
Elaprase	Hunter syndrome	Shire Pharmaceuticals
Cerezyme	Gaucher disease	Genzyme

1.3.5. Glycans in immune responses

The immune system recognises self and non-self and eliminates non-self from the host. Glycans are ubiquitous on the cell surface and represent the first interface between the host and the pathogen to facilitate both innate and adaptive immune recognition processes. The glycan chains of glycoconjugates and glycoproteins on the cell surface possess a high structural diversity. Alteration of *O*-linked oligosaccharides are associated with the development of T-cell activation, CD3, one of the cell surface glycoproteins present on T-cells has a majority of *O*-linked core structures and their altered sialoglycoprotein structures have played a prominent role

in T cell activation (Andersson et al., 1978, Piller et al., 1988). Alteration of *N*-linked glycosylation has been shown to affect immune responses such as evasion of immune of response in AIDS (acquired immunodeficiency syndrome) (Reitter et al., 1998), and in immunogenic evasion of influenza virus (Wang et al., 2009, Vigerust et al., 2007).

Expression of glycan binding proteins such as galectins in immune cells play multiple roles in innate and adaptive immunity (Marth and Grewal, 2008). The activated T and B cells, inflammatory macrophages, and natural killer (NK) cells differentially up-regulate these galectins thereby validating the expression of glycans in immune responses (Rabinovich et al., 2007, Koopman et al., 2003).

1.3.6. Non-human glycans

Non-human glycans are those, which are not naturally synthesised or expressed in humans. Many glycans from non-human mammals, invertebrates, plants, yeast and bacteria have different monosaccharide constituents and oligosaccharide structural organisation than those of humans. Glycan structures from plants lack sialic acids and Galactose on their oligo chains (Gomord and Faye, 2004). Also, in yeast, insects and some plants there is not even a penultimate Gal structure available, as these oligosaccharides are mainly build with high mannose contents (Trimble et al., 1983, Lerouge et al., 1998, Wojchowski et al., 1986). Sialic acid plays an important role in cell recognition and the glycans without sialic acid residues are immunogenic and rejected by the human system. Glycoproteins and glycoconjugates having non-human glycans such as core xylose present in plants and core α-(1, 3)-Fucose present in both insects and plants possess IgE subtypes, which has been reported to cause anaphylaxis in humans (Bardor et al., 2003). The production of recombinant therapeutics in either plant, insect or yeast expression systems contaminate the therapeutic with non-human glycans causing immunogenicity and reduces the efficacy of drug (Sethuraman and Stadheim, 2006). As yeast expression systems are used in the production of therapeutics and enzymes, however due to the presence of high mannose contents confers short half life of drugs and also causes immunogenicity in humans (Gerngross, 2004). Some mammals express Neu5Gc (found in non-human primates, cattle, etc) or Gal-α-(1, 3)-Gal (found in pigs, cow, sheep, etc.) epitopes on their cell surface, whereas humans do not naturally express

these epitopes. On introduction, these glycan epitopes can generate an immune response in humans due to the presence of natural antibodies. Within the scope of the work contained in this document, the non-human Gal-α-(1, 3)-Gal and Neu5Gc determinants were considered.

1.4. Galactosyl-α-(1, 3)-Galactose (Gal-α-(1, 3)-Gal)

Gal-α-(1, 3)-Gal (or α-galactosyl epitope) is present naturally in the form of Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc. α-(1, 3)-galactosyltransferase [α-(1, 3)-GT] synthesises α-galactosyl epitope in the golgi apparatus of non-primate mammals (Betteridge and Watkins, 1983). The synthesis reaction resulting in the α-galactosyl epitope is shown below.

Gal-β-(1, 4)-GlcNAc + UDP-Gal

⇓

Gal- α-(1, 3)-Gal-β-(1, 4)-GlcNAc-R + UDP

The Gal-α-(1, 3)-Gal epitope is a terminal glycan antigen not normally produced by humans due to the absence of α-(1, 3)-GT enzyme. However the GGTA1 gene that encodes for theα-(1, 3)-GT enzyme is present in humans, but during evolution, the GGTA1 gene became inactive as ascribed to two possible theories. One theory is that selective pressure for the suppression of α-Gal epitope. Another theory is linked to the presence of natural anti-α-Gal antibodies in humans causing an autoimmune response (Galili and Swanson, 1991, Koike et al., 2002). It has been estimated that nearly 1% of the total circulating I_gGs in the humans are anti- α-Gal antibodies (Galili et al., 1985), which have been found to interact with the Gal-α-(1, 3)-Gal epitope both on glycosphingolipidis and on glycoproteins (Galili et al., 1984). The different biological roles of the Gal-α-(1, 3)-Gal epitope are outlined in Table 1.2.

Table 1.2: Sources to and consequences of Gal-α-(1, 3)-Gal epitope exposure in humans and bio-therapeutics

Possible exposure routes to the Gal-α-(1, 3)-Gal epitope in humans	**References**
Major obstacle to successful xenotransplantation of porcine tissues in humans	(Joziasse and Oriol, 1999, Galili, 2005)
Contamination of CHO, SP2/0 and NS0 cell lines required for recombinant therapeutics with Gal-α-(1, 3)-Gal epitope. Also contamination in case of Cetuximab (Erbitux)	(Pointreau et al., 2012, Bosques et al., 2010)
Initiate host pathogen interaction: *Trypanosoma cruzi* and American *Leishmania* (parasites) causing Chagas disease and American cutaneous Leishmaniasis.	(Avila et al., 1989, Towbin et al., 1987)
Initialisation of fungal infection as extracellular vesicles of *Paracoccidioides brasiliensis* carries highly immunogenic α-linked galactopyranosyl (α-Gal) epitopes.	(Vallejo et al., 2011)

1.5. Sialic acids

Sialic acids are a class of nine-carbon, acidic monosaccharides first discovered seven decades ago (Blix et al., 1955, Saito, 1956). Today, over 50 types of sialic acids have been reported. Those found in nature usually appear to have α-(2, 3), α-(2, 6) or α-(2, 8) linkages to Gal or *N*-acetylgalactosamine (GalNAc) residues on N and O linked glycoproteins, and free glycans. The most common are terminal linkages to the C-3 (α-2, 3) and C-6 (α-2, 6) linked of the Galactose and to the C-6 (α-2, 6) positions linkage of the GalNAc residues (Varki et al., 2009b). Also, there are internal linkages where sialic acids are attached to other sialic acid residues via C-8 (α-2, 8) linkages to form poly sialic acid, this is a structure encountered frequently in neuronal tissues (Ledeen, 1984, Sato et al., 2002). Sialic acids are frequently present on cell surfaces and are involved in broad range of biological processes, including molecular and cellular interactions, immune response, and host pathogen interaction (Varki et al., 2009b). Also, sialic acid epitopes bind to sialic acid biding lectins, namely siglecs and selectins (for more detailed information about lectins please refer section 1.7.1),

initiating cell-to-cell communication in the inflammation and immune response (Kelm and Schauer, 1997). The predominant form of sialic acid found on human glycoproteins and glycolipids is Neu5Ac, whereas in most other mammals except humans, Neu5Gc is also widely expressed in addition to Neu5Ac. Neu5Ac and Neu5Gc are synthesised from the ManNAc and ManNGc, respectively. The pathway for Neu5Ac and Neu5Gc synthesis is shown in (Figure 1.8):

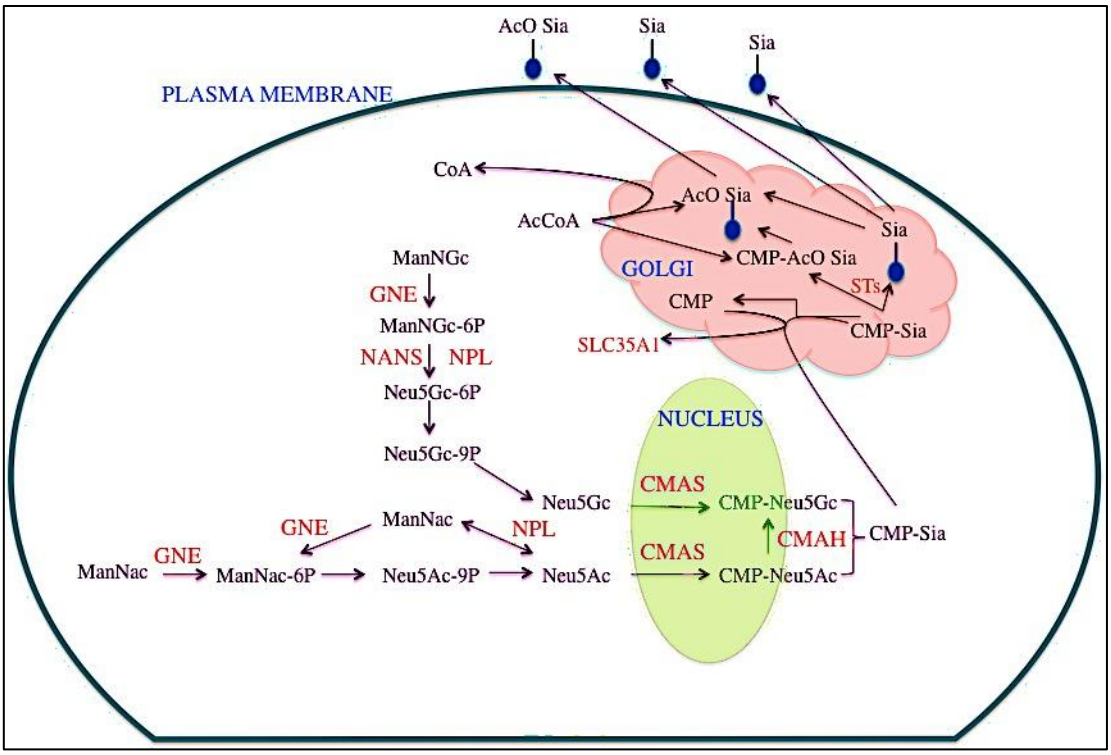


Figure 1.8: Sialic acid pathways in the Golgi apparatus and nucleus. The text in red are the genes involved in the synthesis of precursors of sialic acids. The image is modified from a version by Varki, Cummings et al. (2009) (Varki et al., 2009b).

1.5.1. *N*-acetylneuramin*ic* acid (Neu5Ac)

Neu5Ac is the most common sialic acid found in glycoconjugates in all mammals. Neu5Ac is an essential component of brain gangliosides and also polysialic acid that modifies neural cell adhesion molecule (NCAM). Both of these play important role in neuronal growth, cell-cell interactions and memory formation, thus making Neu5Ac bound to gangliosides and glycoproteins an essential sugar for brain development and cognition (Wang, 2009, Svennerholm et al., 1989, Zuber et al., 1992). It is also important in therapeutics as most protein therapeutics have conventional α-2, 3 and α-2, 6 sialic acid linkages whereas polysialic acids are

mainly chains of Neu5Ac with an α-2, 8 linkages on peptides and protein therapeutics have been shown to improve drug half life, retention of their activity *in vivo* and to reduce antigenicity and immunogenicity of these therapeutics (Gregoriadis et al., 2005, Werle and Bernkop-Schnurch, 2006). Neu5Ac has a major role in developing influenza virus infections in humans, hemagglutinin: a viral surface glycoprotein and a sialic acid receptor binding protein of influenza virus that recognises and binds to Neu5Ac on erythrocytes in case of influenza virus A and B, whereas in influenza C infection hemagglutinin recognises 9-*O*-acetyl-*N*-acetylneuraminic acid (9-*O*-Ac-NeuAc) and initiate viral infection in humans (Higa et al., 1985, Kelm et al., 1992). Neu5Ac also provides a good substrate for bacteria *Vibrio vulnificus* for invasion and colonisation of enteropathogenic bacteria in the human gut (Jeong et al., 2009).

1.5.2. *N*-glycolylneuraminic acid (Neu5Gc)

Neu5Gc is synthesised by the hydroxylation of its precursor sialic acid: Neu5Ac. The enzyme, Cytidine Monophosphate *N*-acetylneuraminic acid Hydroxylase (CMAH) catalyses Neu5Ac to form Neu5Gc (Figure 1.9). Neu5Gc is a widely expressed non-human sialic acid widely expressed in mammals except humans.

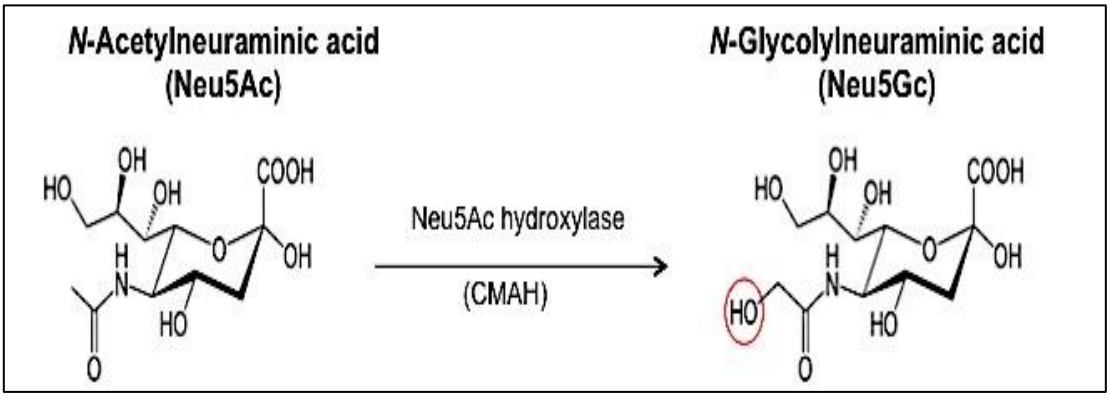


Figure 1.9: Illustration shows two different sialic acids and their conversion of Neu5Ac to Neu5Gc with the enzyme Neu5Ac hydroxylase.

1.5.2.1. Loss of Neu5Gc in humans and incorporation of Neu5Gc into humans

It is theorised that during human evolution a frameshift mutation due to a deletion of 92 base pairs of single strectch of sequences of a gene that corresponds to the gene 6 in mice resulted in the inactivation of single CMAH gene (Brinkman-Van der Linden

et al., 2000, Varki, 2001, Hedlund et al., 2007). Due to the inactivation of this gene, CMAH enzyme production ceased and humans were not able to synthesise Neu5Gc from Neu5Ac. However, traces of Neu5Gc in humans have been reported in many studies. Despite being antigenic to humans, Neu5Gc is found in humans as human embryonic stem (Aspholm-Hurtig et al.) cells have been shown to express Neu5Gc when grown under standard condition (Martin et al., 2005). These cells can take up Neu5Gc from animal media and convert it into CMP-Neu5Gc thereby incorporating it using similar sialyltransferases and transporters in the Golgi as CMP-Neu5Ac (Varki, 2001, Tangvoranuntakul et al., 2003). Meat from sheep, pigs and cattle (red meat) and the dairy products from bovine sources contain Neu5Gc, which on dietary accumulation could allow incorporation of Neu5Gc into human glycoproteins and glycolipids (Tangvoranuntakul et al., 2003, Hedlund et al., 2007). The *in vitro* incorporation mechanism of glycoprotein bound Neu5Gc into human carcinoma cells has been demonstrated to occur via micropinocytosis and delivered through lysosomal sialadases and transporters in to the cytosolic compartment for conversion to CMP-Neu5Gc (Bardor et al., 2005).

1.5.2.2. Neu5Gc immunogenicity in humans

Humans typically have detectable levels of anti-Neu5Gc antibodies which, on interacting with exogenous Neu5Gc present in excess amounts on glycoproteins and glycolipids, causes an immune response (Nguyen et al., 2005). Various pathogens have specificity to Neu5Gc over Neu5Ac. These pathogens could bind with incorporated Neu5Gc to initiate infection in humans, such as the subtilase cytotoxin secreted by Shiga toxigenic *E. coli* (STEC) (Byres et al., 2008). This toxin has been reported to bind to different α-(2, 3)-Neu5Gc glycoconjugates when presented to it on glycan arrays (Byres et al., 2008). The interaction of STEC with exogenous Neu5Gc in humans may cause haemolytic-uremic syndrome in the gastrointestinal tract and kidneys in humans (Lofling et al., 2009). The presence of Neu5Gc is a significant hindrance to xenotransplantation as pigs also incorporate Neu5Gc. Organs transplanted from pigs to humans may be rejected due to the presence of natural anti-Neu5Gc antibodies in humans (Padler-Karavani and Varki, 2011). This Neu5Gc motif can contaminate embryo development cells as animal by-products (possessing Neu5Gc) are required to derive and culture these cells (Lanctot et al., 2007).

1.5.2.3. Neu5Gc in cancer

Although Neu5Gc is not synthesised in humans, there is evidence that Neu5Gc-containing glycoconjugates occur frequently in the sera of cancer patients and on the surface of cancerous tissues (Taylor et al., 2010). Following studies on a Neu5Gc-deficient mouse model system, it has been proposed that the interaction of tumour-associated Neu5Gc with naturally occurring anti-Neu5Gc antibodies is involved in tumour progression via chronic inflammation (Padler-Karavani et al., 2011). Recently, it has been confirmed from the same research group that Neu5Gc from red meat promotes inflammation and cancer progression (Samraj et al., 2014). Neu5Gc with in ganglioside GM3 represents the Hanganutziu-Deicher antigen HD-3 and has been reported as a colon tumour-associated antigen in humans (Higashi et al., 1985).

1.5.2.4. Neu5Gc in bio-therapeutics

Recombinant proteins normally require glycosylation to maintain half-life, reduce required dosage and fully exert proper biological function (Sinclair and Elliott, 2005). The glycosylation pattern comes from the organism where bio-therapeutic proteins are expressed. For example, Chinese Hamster Ovary (CHO) and recombinant human erythropoietin produced within these contain at least 1% Neu5Gc of the total sialic acid content (Noguchi et al., 1996). During glycosylation of expressed proteins, the contamination of Neu5Gc is transferred into the recombinant glycoprotein therapeutics from the expression system (e.g. erythropoietin) (Hokke et al., 1990, Noguchi et al., 1996). Recent work has shown that the contamination of bio-therapeutics with this non-human epitope can result in an immunogenic response in humans, as e.g. as occurred with Erbitux (Cetuximab), a recombinant monoclonal antibody preparation produced in a murine myeloma cell line, and Vectibix (Panitumumab) a fully human antibody produced in CHO cells (Ghaderi et al., 2010). The likelihood of Neu5Gc as a contaminant in different FDA approved bio-therapeutics for human dosage, prepared in different mammalian expression system has recently been extensively reviewed (Ghaderi et al., 2012).

1.6. Detection approaches for glycans

Cell surface glycan structures are quantified using enzyme treatment and through released glycan estimation. However, analysis of cleaved glycans presents a major

analytical challenge, due to their inherent complexity, lack of optical properties useful for detection, and the existence of various isoforms (both position and linkage). In addition, populations of almost all glycoproteins consist of a heterogeneous collection of differently glycosylated variants, so the released glycan pool contains a range of structures. Methods which are currently in use include high performance liquid chromatography (HPLC), mass spectroscopy (MS), tandem MS (MS/MS), LC–MS, capillary electrophoresis (CE), high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC–PAD), lectin and antibody arrays, and binding assays, each of which has its own merits and limitations (Seo and Lee, 2004, Dalpathado and Desaire, 2008).

1.6.1. Chromatography and Mass spectrometry

Chromatography is the most conventional analytical method of separating and quantifying glycans in the complex biological samples. Thin layer chromatography has been in use to analyse glycan profiles (Lippiello and Mankin, 1971, Schneider et al., 1993) and further enhancement has been done in this method to improve resolution and for more accurate analyses using high performance thin layer chromatography (HPTLC) (Yao and Rastetter, 1985, Zlatkis and Kaiser, 2011). The advancement in chromatography continued when paper chromatography was replaced with liquid chromatography and then with the advanced version of liquid chromatography as High Performance Liquid Chromatography (high pressure) technique was implemented for separation and quantification of glycans (Harz et al., 1990). In the last two decades further modification and improvement in HPLC techniques include high-pH anion exchange chromatography (HPAEC) (Bruggink et al., 2005), ultra performance liquid chromatography (UPLC) (Bones et al., 2010), and hydrophilic interaction chromatography (HILIC) (Ahn et al., 2010) for efficient quantification and separation of the glycans. Databases have been developed that contain analytical parameters for the relative quantification of *N* and *O*-linked glycans (Royle et al., 2008, Royle et al., 2002). GlycoBase is a detailed database with an analytical tool autoGU identifies glycan residues for interpretation and assignment of HPLC glycan profiles (Campbell et al., 2008).

Mass spectrometry is another analytical technique to measure mass to charge ratio of glycans released from glycoconjugates. MS of glycoproteins has witnessed

considerable advancements in recent years, mainly as a necessary-but-peripheral addition to the proteomics field and aimed primarily at elucidating the structural complexity and functionality of the oligosaccharide components of molecules. Several MS methods are currently employed in conjunction with chromatographic techniques for oligosaccharide analysis, and include graphitised carbon-liquid chromatography (LC)/electrospray ionisation (ESI) MS detecting deprotonated molecules in the negative ion mode provided acceptable quantitation. Also, detailed analyses of tryptic glycopeptides employing either nano LC/ESI MS/MS or MALDI/MS has been shown to give reliable site-specific or subclass-specific glycan profiles (Wada et al., 2007).

1.6.2. Lectin and antibody arrays

Lectins are glycan-binding proteins, which bind specifically and reversibly, found in the plants, animals, and in microorganisms (Drickamer and Taylor, 1993, Rudiger and Gabius, 2001, Lis and Sharon, 1986). Lectins are used to detect glycans present on glycoconjugates in tissues and on cell surfaces (Sharon and Lis, 1989, Hebert, 2000). Lectin array-based glycan profiling has given promising results with a range of sample types, including glycopeptides, glycoproteins, live mammalian cell-surface glycome, formalin-embedded tissue sections and bacteria (Hsu and Mahal, 2006, Landemarre et al., 2013, Zhou et al., 2011). In most lectin microarray-based experiments, the analytes (target glycans or glycoconjugates) are fluorescently labelled and these targets incubated to lectin microarray features (spots) are later detected using standard laboratory laser scanners originally developed for DNA microarray applications. A ratiometric two-colour approach has also been described in an effort to improve the ability to do relative quantification between two different samples on lectin arrays (Pilobello et al., 2007). Antibody arrays have also been used for high-throughput analysis of glycan variation on native proteins (Chen et al., 2007). However, the lectin and antibodies arrays have limitations in glycan binding analyses as these arrays lack information about exact nature of glycans structures and their binding linkages sites. Also for high diversity glycans, it is difficult for lectins and antibodies to determine the sites of altered glycans (Haab, 2010).

1.6.3. Binding assays

ELISA (Enzyme Linked Immunosorbent Assays), used with antibodies raised against carbohydrates (Knuchel et al., 1992), and the ELLA (Enzyme Linked Lectin Assays), used with lectins (Lambre et al., 1991), are the most robust and convenient assays for the detection and quantification of free and bound glycans in biological samples (Gull et al., 2007, Gornik and Lauc, 2007). These assays are still widely in use and much progress has been made in the development of high-throughput versions of these assays, as in multiplex suspension arrays used to profile anti-carbohydrate antibodies (Pochechueva et al., 2011). Flow cytometry-based binding assays are also used to profile both cell surface glycans and free glycans which are immobilised onto beads and detected using both lectins and antibodies as probes (Yamamoto et al., 2005). Electrophoresis and Western blots are two other conventional approaches to detect glycans based on their sise and charges using both lectin and antibodies as probes (Tabares et al., 2006).

1.6.4. Structural and *in silico* glycan binding interactions

Protein glycosylation is not homogeneous. Furthermore, the glycosylation motifs present affect the physiochemical and biological properties of the glycoprotein and glycolipids. Therefore, structural knowledge of glycosylation is important for both understanding normal physiological processes and many disease states and for effective treatment of a range of infectious and chronic diseases. X-ray crystallography and NMR are used to ‘solve’ many structures of lectins, glycans and their complexes. These two techniques give detailed information about hydrogen bonding, hydrophobic sites, and van der Waals interactions, contribute to binding energies of lectin-glycan interaction (Weis and Drickamer, 1996). However, there are many newly-discovered or selected antibodies and single chain variable fragments (scFv) against glycans whose structure have not been elucidated and, for those molecules, alternative approaches such as molecular modelling and docking can be employed to help elucidate the nature of their interaction with carbohydrate targets (Imberty and Perez, 1994, Neumann et al., 2004).

1.7. Glycan recognition molecules

Alteration in glycosylation associated to many diseases and changes in immune responses. The evidence are now emerging to link such aberrations to disease properties (Goulabchand et al., 2014, Campbell et al., 2001). The availability of sensitive and specific reagents and tools for the detection of glycans at an early stage of diseases is of great importance. Glycan binding molecules could be used for detection, diagnosis and prognosis, and validation of therapeutic and disease management.

1.7.1. Lectins

Lectin carbohydrate recognition domains (CRD) (Drickamer, 1988) reversibly and specifically interact with carbohydrates (Gabius et al., 2004). Lectins have been used for the characterisation of cell surface glycans and glycoproteins, based on their selectivity for certain glycan structures and linkages. They are involved in diverse biological processes including cell-to-cell communication (Gabius, 1988), clearance of glycoproteins in the circulatory system (Ashwell and Morell, 1977), blood group typing, bacterial identification, and cell selection. They are considered to be a very large and heterogeneous group of proteins (Goldstein and Poretz, 1986) and are divided into different classes based on their amino acid sequences, folding, and other biochemical properties. Lectins with reported affinity and specificity against sialic acids and α-galactose epitopes are mentioned in Table 1.3.

Table 1.3: Lectins with their specificities against Neu5Ac, Neu5Gc and α-Galactose epitope.

Lectins name	**Abbreviation**	**Specificity**	**References**
Sambucus nigra	SNA -I, EBL	NeuAc-α-(2, 6)-Gal, NeuAc-α-(2, 6)-GalNAc	(Taatjes et al., 1988)
Maackia amurensis	MAA, MAL	NeuAc-α-(2, 3)-Gal, NeuAc-α-(2, 3)-Gal-β-(1,4)-GlcNAc/Glc	(Sata et al., 1989)
Limax flavus	LFA	NeuAc, NeuGc	(Miller et al., 1982)
Limulus polyphemus	LPA	NeuAc, GalNAc, GlcNAc	(Roche and Monsigny, 1974)
Bandeiraea simplicifolia (Lectin I) *Griffonia simplicifolia* (Isolectin I-B_4)	GS I, GSL I, BS I, BSL I, GS-I-B4	α-Gal, α-GalNAc	(Goldstein et al., 1981) (McCoy Jr et al., 1983)
Euonymus europaeus	EEA	Gal-α-(1, 3)-Gal, Gal-α-(1, 3)-[Fuc-α-(1,2)-Gal]	(Petryniak et al., 1980)
Maclura pomifera	MPA, MPL	α-Gal, α-GalNAc, Gal-β-(1, 3)-GlcNAc	(Sarkar et al., 1981)
Vigna radiata (Mung Bean lectin)	MBL-I	β-Gal	(Suseelan et al., 1997)
Vigna radiata (Mung Bean lectin)	MBL-II	β-Gal	(Suseelan et al., 1997)
Marasmius oreades agglutin (Mushroom lectin)	MOA	Gal-α-(1, 3)-Gal and Gal-α-(1, 3)-Gal –β-(1, 4)-GlcNAc	(Winter et al., 2002)

1.7.2. Glycan binding antibodies

Alterations in glycosylation patterns have been associated with multiple disease conditions, and with altered glycosylation, anti-glycan antibodies levels present in the serum may also vary with disease onset and progression. Within the immune system, antibodies are generated against pathogens including foreign glycan epitopes and therefore, these glycan-binding antibodies present in the serum could be used as biomarkers. Circulating antibodies within the serum can readily be isolated from the animals, for example polyclonal I_gY antibodies against Neu5Gc were isolated from the chicken (Diaz et al., 2005) for use as a diagnostic and therapeutic purposes.

1.7.3 Glycan binding peptides and scFv

A number of specific glycan-binding peptides have been described recently, prepared by screening phage-displayed peptide libraries against the targets of interest (Table 4). Phage Display is a well-established technique for the selection and production of peptides and scFvs (Smith, 1985).

In cancer glycobiology, the use of phage display libraries to identify peptide sequences that bind the T antigen was described nearly two decades ago (Peletskaya et al., 1997). Peptide libraries, now commercially available, have become a suitable method for selecting peptides against glycans with required binding properties (Frank, 2002, Tothill, 2009, Kriplani and Kay, 2005). scFvs are another class of peptides selected through phage display techniques. To produce scFvs, the DNA encoding the antigen-binding variable domain is cloned into a DNA phagemid vector and expressed in the phage as a single chain molecule in which the V_L and V_H domains of the antibody are joined by a flexible polypeptide linker (Gram et al., 1992). They have a number of advantages over traditional antibodies for use as a diagnostic tools (Hagemeyer et al., 2009). Their reduced sise (typically about 25-35 kDa, compared to monoclonal antibodies at 150 kDa), make them more suitable for the delivery of chemotherapeutic agents in tumour treatment (Lyu et al., 2008), for use as parasite therapeutics (Vukovic et al., 2002, Yoshida et al., 2003) and possibly for tumour vaccines (Ladjemi et al., 2011). The list of antibodies, peptides, and scFvs against Gal-α-(1,3)-Gal and Neu5Gc is detailed in Table 1.4.

Table 1.4: List of antibodies, peptides, and scFv against Gal-α-(1, 3)-Gal and Neu5Gc

Type	Target	Source	Reference
M86 monoclonal antibody	Galactosyl-α-(1, 3)-Galactose	Mouse	(Galili et al., 1998)
I_gY polyclonal antibody	Galactosyl-α-(1, 3)-Galactose	Chicken	(Bouhours et al., 1998)
Chimeric I_gE scFv	Galactosyl-α-(1, 3)-Galactose	Mouse	(Plum et al., 2011)
scFv	Galactosyl-α-(1, 3)-Galactose	Chicken	(Cunningham et al., 2012)
I_gY polyclonal	Neu5Gc	Chicken	(Diaz et al., 2005)
scFv	Neu5Gc	Chicken	(Donohoe et al., 2011)

1.7.4. Glycan binding aptamers

Aptamers are oligonucleotides that can specifically bind proteins or small molecules in the same way as antibodies and, in some cases, can do so with nanomolar to picomolar affinity against their target molecules (Keefe et al., 2010). Their stability and chemical flexibility have given aptamers significant attention as a new source of potentially high affinity and specific recognition elements for use in biosensor and assay development, diagnostics and also as possible therapeutic agents (Ni et al., 2011).

Aptamers have been traditionally selected using an *in vitro* selection process, known as SELEX (Systematic Evolution of Ligands by EXponential enrichment) that was simultaneously developed by Gold and Szostak (Ellington and Szostak, 1990, Tuerk and Gold, 1990). In the SELEX process (Figure 1.10), highly specific nucleic acid aptamers are isolated from a diverse pool of randomised nucleic acids containing more than 10^{12} different sequences. This involves an iterative process of binding, partitioning, washing, elution and amplification by polymerase chain reaction (PCR). Because a randomised RNA or DNA library can provide a vast number of three-

dimensional structures to different kind of targets, aptamers have been generated against a wide variety of target such as ions, glycans, small compounds and proteins. With increasing stringency of selection in each round of SELEX, the specificity and affinity of the aptamers become mature. The physical factors contributing to aptamer affinity and specificity against a target are the electrostatic and hydrophobic interactions between the molecules. Hydrogen bonding and the target shape complementarity are other features help in aptamer recognition. Also, highly conformational flexibility between aptamer and target could be the important feature to limit aptamer affinity and specificity (Eaton et al., 1995). The structural basis for ligand recognition by aptamers is through the use of its purine rich loops to bind against ligands. Also, these purine loops engage in noncanonical base pairing interactions with each other to facilitate H-bond acceptors and donors and the proper surfaces for ligand interaction (Wilson and Szostak, 1999).

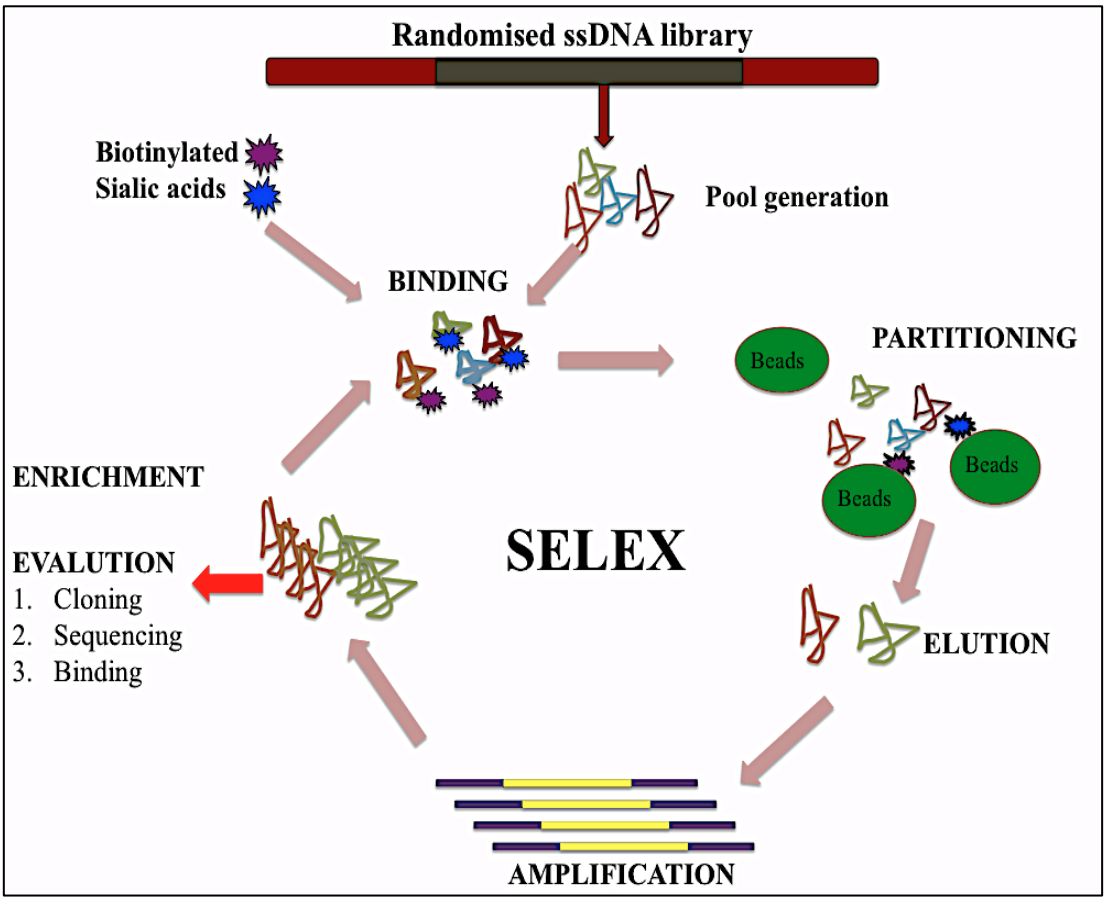


Figure 1.10: Illustration of SELEX process for the generation of ssDNA aptamers.

1.8. Review of aptamers reported against glycans

Aptamers have been employed in glycobiology recently to bind and differentiate between the complex glycan structures (Sun et al., 2010). Although they have been described against carbohydrate targets, revealing the feasibility of this approach for the generation of specific glycan receptors, there are no reports so far of their use as diagnostic reagents. The list of aptamers possessing affinity and specificity against glycans are included (Table 1.5).

Table 1.5: List of DNA aptamers reported against glycans.

Target	**Binding motif & Aptamer length**	**Affinity**	**Specificity**	**Reference**
5TRG1-4	70 mer library with 25 nucleotides variable region	18.6-34.4 nM	Tn antigen	(Ferreira et al., 2006)
GalNAc1-3	70 mer library with 25 nucleotides variable region	47.3-59.8 nM	*N*-Acetylgalactosamine	(Ferreira et al., 2006)
5TR1-4	70 mer library with 25 nucleotides variable region	47.3-59.8 nM	MUC1-5TR	(Ferreira et al., 2006)
Glycoprotein	90 mer library with 50 nucleotides variable region	6.2-321 nM	Fibrinogen Deglycosylated fibrinogen Periodated fibrinogen	(Li et al., 2008)
(1, 3)-β-D-glucan	95 mer library with 60 nucleotides	0.3 μM	Dextran: -- Mannan: --	(Low et al., 2009)

	variable region		Endotoxin: −− Laminarian: ++ Barley Glucan: ++ Curdlan: ++	
Sialyllactose	Three way junction with amino- modified thymidines, 60 nucleotides variable region	4.9 μM	Not determined	(Mehedi Masud et al., 2004)
Cellulose	G Rich Sequences streches, nd	Low μM	Cellulose: +++ Cellotetraose: 0.6uM Cellobiose:< 0.3uM Lactose: − Maltose: − Gentobiose:−	(Yang et al., 1998)
Chitin	103 nucleotides, 59 nucleotides variable region G Rich stem loops	n.d.	Chitin: ++ 52%, 73% Cellulose:+	(Fukusaki et al., 2000)
VEGF	77 nucleotides, 30 nucleotides variable region	0.1 nM	Heparin binding- $VEGF_{165}$++	(Ng et al., 2006)

Neu5Ac	76 nucleotides, 40 nucleotides variable region	Neu5Ac (1.35 nM) Neu5Gc (90 nM)	Neu5Ac++ Neu5Gc++	(Cho et al., 2013)
Neu5Gc	81 nucleotides, 40 nucleotides variable region	4.58 nM	Neu5Gc++ Neu5Ac: n.d.	(Gong et al., 2013)

* ++ indicates the level of binding affinity, -- indicates no binding to the target.

1.8.1. Aptamers in Therapeutics and Diagnostics

A number of aptamer-based therapeutics are now at various stages of preclinical and clinical trials. The global aptamers market is anticipated to reach $2.1 billion by 2018 (reasearchandmarkets.com, 2014). Pegaptanib (Macugen) (Figure 1.11) is an anti-angiogenic aptamer used for the treatment of neovascular age-related macular degeneration (AMD). The first RNA aptamer based drug targeted the vascular endothelial growth factor (VEGF), which blocks the action of VEGF resulting in the reduced growth of blood vessels for controlling leakage and swelling in the eye (Ng et al., 2006). This drug was approved and licensed to Eytech firm (Now OSI pharmaceuticals Inc. in USA) 2004, and in Europe 2005. Many other aptamer based drugs are in clinical trial stages and yet to get approval from FDA for commercialisation. A list of some of these aptamer-based drugs under clinical trial is presented in Table 1.6.

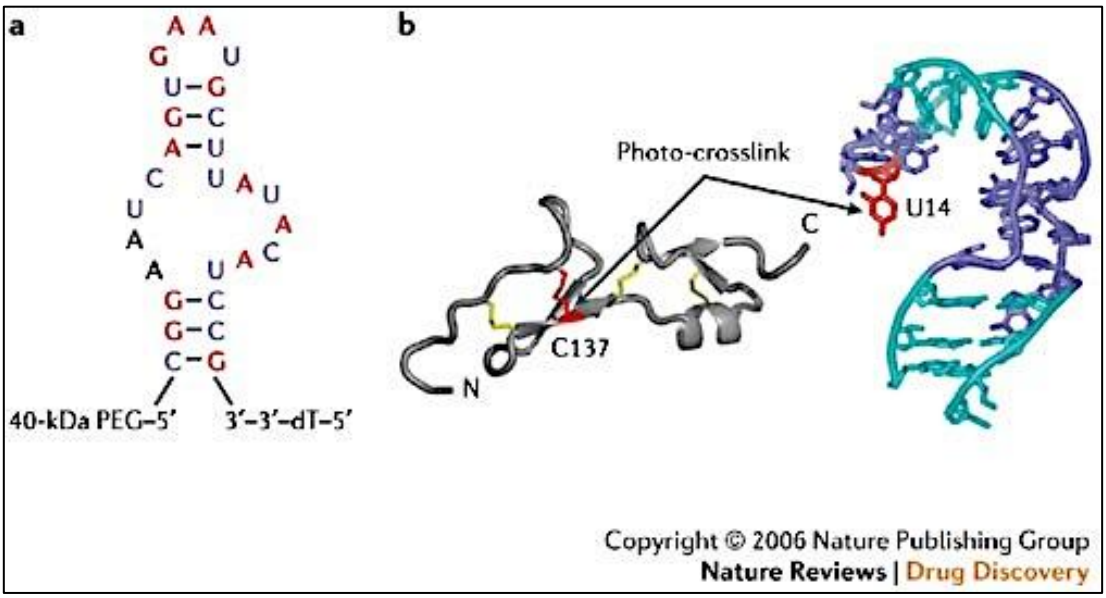


Figure 1.11: Pegaptanib (Macugen) – RNA based aptamer drug against $VEGF_{165}$ (Copyright license number- 3554740146757), (Ng et al., 2006).

Table 1.6: Aptamer based therapeutics under development

Drug name	**Target**	**Disease**	**Development Stage**	**Company**	**References**
AS1411	Solid and blood cancer cells	Solid and blood cancer	Phase IIb trial terminated	Antisoma	(Soundararajan et al., 2008)
NOX-E36	MCP-1 (CCL2) (Monocyte Chemoattractant Protein 1)	Type II diabetic complications, Diabetic Nephropathy	Phase Ib	Noxxon pharma AG	(Anders et al., 2007)
NOX-A12	SDF-1 (CXCL12) (Stromal Cell-Derived Factor-1)	Hematological and solid tumours	Phase I	Noxxon pharma AG	(Hoellenriegel et al., 2011)
NOX-H94	Hepcidin	Anemia of inflammation	Clinical trials	Noxxon pharma AG	(Schwoebel et al., 2011)

E10030	platelet-derived growth factor	age-related macular degeneration	Phase II	Ophthotech	(Green et al., 1996)
ARC1905	C5 component of the complement cascade	age-related macular degeneration	Phase I	Ophthotech	(Biesecker et al., 1999)
REG1	Factor IXa	Acute coronary syndrome, percutaneous coronary intervention	Phase IIb	Regado Biosciences	(Rusconi et al., 2002)
REG2	Venous thrombosis indications	Venous thrombosis indications prophylaxis	Phase I	Regado Biosciences	(Rusconi et al., 2004)

1.9. Intellectual Property in SELEX and Phage Display

Patent activity in the SELEX and phage display field has shown an exponential growth in the last five years according to the United States Patent and Trademark Office (USPTO). Aptamers and scFvs against several proteins and other targets involved in a broad range of disease causing pathogens and proteins have been patented for their use as biomarkers, therapeutics and in diagnostics. Archemix and Somalogic are the pioneers in aptamer development and they intend to transform medical practice through the development of aptamer-based, next generation disease screening, diagnosis, therapy selection and patient monitoring (Bock et al., 1992, Gold et al., 2003, Gold et al., 2002). However, the development of aptamers and scFvs in the field of glycobiology is very limited, which could be due to the complexity of glycan structural diversity. In addition, there are a few reports on the use of aptamers and phage display scFv molecules to develop carbohydrate biosensors in Table 1.7.

Table 1.7: Patents on application of aptamer and scFv as glycobiomimics and biosensors.

Patent number	**Description**	**References**	**Publication Date**
7,939,313	Biosensors for detecting macromolecules and other analytes	(Heyduk et al., 2011)	10-05-2011
7,927,547	Reagentless and reusable biosensors with tunable differential binding affinities and methods of making	(Anderson et al., 2011)	19-04-2011
7,910,523	Structure based and combinatorially selected oligonucleoside phosphorothioate and phosphorodithioate aptamer targeting AP-1 transcription factors	(Gorenstein et al., 2011)	22-03-2011
7,871,785	Use of secretor, Lewis and sialyl antigen levels in clinical samples as predictors of risk for disease	(Morrow et al., 2013)	22-05-2014
12/221,429	Aptamer based point of care test for glycated albumin	(Smith, 2013)	19/06/2014
WO 20071 28109 A1	Aptamers that recognise the carbohydrate *N*-acetylgalactosamine (GalNAc)	(Matos and Jean, 2007)	15-11-2007
7,915,387	Monoclonal antibody Sc104 and derivative thereof specifically binding to a sialyltetraosyl carbohydrate as a potential anti-tumour therapeutic agent	(Durrant and Parsons, 2011)	29-03- 2011
US 2011/0034 676 A1	Anti-Sialic acid antibody molecules	(Donohoe et al., 2011)	10-02-2011
7,674,605	Antibodies recognizing a carbohydrate containing epitope on CD-43 and CEA expressed on cancer cells and methods using same	(Lin et al., 2011)	09-03-2010
US 2010/ 0324271 A1	Phage-displaying single chain antibody capable of recognizing non-reduced mannose residue	(Yamaguchi et al., 2009)	23-10-2010
6,727,062	Identification of target structures E.G. in vivo selection method for a phage library	(Brodin et al., 2004)	27-04-2004

1.10. Scope of thesis

Knowledge about post-translational modifications (PTMs) of proteins, such as glycosylation, phosphorylation or sulphation, is proving increasingly critical for understanding the physiological state of cells, systems and organs and for relating genomic and proteomic measurements to a biological and functional context. Covalent attachment of oligosaccharides to the protein backbone, termed glycosylation, is the most common PTM on cell surface and secreted proteins and also the most complex.

Chapter 1 describes the importance of post-translational modification particularly glycosylation. The roles of glycans and their diversity are discussed in detail. Glycan recognition molecules mimicking lectins and antibodies are introduced and described against non human immunogenic glycans Neu5Gc and Gal-α-(1, 3)-Gal.

The wide range of random peptide libraries now available can be tapped for glycan-binding molecules. In addition to generating lectin mimics, this approach can also yield glycomimetic peptides, which could be used as standards in binding assays as an alternative to the often very costly oligosaccharides. These peptides also have therapeutic potential. Recently our lab has developed an scFv that binds and detects Gal-α-(1, 3)-Gal epitope in both direct and competitive ELISA.

Chapter 2 describes the *in silico* interaction of scFv against a range of glycans from monosaccharaides to trisaccharides. Structures of scFvs were modelled *in silico* based on the available crystal structures templates of closely related antibodies and scFvs. The glycan structures were modelled and docked with these scFv to understand the mechanism of scFv-glycan interaction.

Currently, antibodies and lectins are available as recognition tools for carbohydrates, but they have limitations in terms of costs incurred by the production and stability. Owing to their complex structure, antibodies are susceptible to degradation, aggregation, modification (e.g. oxidation or deamidation) and denaturation. The overlapping specificity and toxicity of lectins limit their potential applications. Thus, there is a pressing need for rapid, accurate and economic alternative recognition elements. Aptamers fulfill these requirements and have therefore emerged as a viable option. These are DNA based aptamers targeting against sialic acids. Although I was

more interested in immunogenic non-human glycan Neu5Gc, but I also developed aptamers against Neu5Ac as these two sialic acid structures are closely related and there is no molecule available that can detect Neu5Ac in the biological sample.

Chapter 3 describes the development of aptamers against two different sialic acids: Neu5Gc and Neu5Ac. The SELEX process was used to develop these aptamers, after enrichment of DNA pools against sialic acids; DNA pools were cloned and sequenced. The consensus sequences were obtained among all the DNA pools and their secondary structures were elucidated.

Chapter 4 describes the characterisation of the consensus sequences (aptamers) obtained against select sialic acids as discussed in **chapter-3**. This chapter involves the synthesis of consensus sequences, and their binding characterisation against both sialic acids. Different platforms were used to demonstrate binding affinities of the aptamers identified which had significant specific binding to Neu5Ac and Neu5Gc. The kinetic constants of aptamer binding affinity against the selected sialic acids were estimated. The binding specificities of aptamers against a range of additional monosaccharides were also tested. A competitive ELISA was developed to show specific inhibition of aptamer binding to respective sialic acids.

Chapter 5 The principal conclusions drawn from this research along with potential future research directions are discussed.

1.11. References

AEBI, M., HELENIUS, A., SCHENK, B., BARONE, R., FIUMARA, A., BERGER, E., HENNET, T., IMBACH, T., STUTZ, A. & BJURSELL, C. 1999. Carbohydrate-deficient glycoprotein syndromes become congenital disorders of glycosylation: an updated nomenclature for CDG. *Glycoconjugate journal,* 16**,** 669-671.

AHN, J., BONES, J., YU, Y. Q., RUDD, P. M. & GILAR, M. 2010. Separation of 2-aminobenzamide labeled glycans using hydrophilic interaction chromatography columns packed with 1.7 um sorbent. *Journal of Chromatography B,* 878**,** 403-408.

ANDERS, H.-J., BUCHNER, K., EULBERG, D., KLUSSMANN, S., KULKARNI, O., PURSCHKE, W. & SELVE, N. 2007. Spiegelmer NOX-E36 for the treatment of chronic inflammatory diseases: A new chemical entity that inhibits CCL2/MCP-1 signalling. *Naunyn-Schmiedebergs Archives of Pharmacology,* 375**,** 53-53.

ANDERSON, G. P., GOLDMAN, E. R., MAURO, J. M. & MEDINTZ, I. L. 2011. Reagentless and reusable biosensors with tunable differential binding affinities and methods of making. Google Patents. US 7927547 B2.

ANDERSSON, L. C., GAHMBERG, C. G., KIMURA, A. K. & WIGZELL, H. 1978. Activated human T lymphocytes display new surface glycoproteins. *Proceedings of the National Academy of Sciences,* 75**,** 3455-3458.

ASHWELL, G. & MORELL, A. G. 1977. Membrane glycoproteins and recognition phenomena. *Trends in Biochemical Sciences,* 2**,** 76-78.

ASPHOLM-HURTIG, M., DAILIDE, G., LAHMANN, M., KALIA, A., ILVER, D., ROCHE, N., VIKSTROM, S., SJOSTROM, R., LINDEN, S. & BACKSTROM, A. 2004. Functional adaptation of BabA, the H. pylori ABO blood group antigen binding adhesin. *Science,* 305**,** 519-522.

AVEZOV, E., FRENKEL, Z., EHRLICH, M., HERSCOVICS, A. & LEDERKREMER, G. Z. 2008. Endoplasmic reticulum (ER) mannosidase I is compartmentalized and required for N-glycan trimming to Man5-6GlcNAc2 in glycoprotein ER-associated degradation. *Molecular biology of the cell,* 19**,** 216-225.

AVILA, J. L., ROJAS, M. & GALILI, U. 1989. Immunogenic Gal-α-(1-3)-Gal carbohydrate epitopes are present on pathogenic American Trypanosoma and Leishmania. *The Journal of Immunology,* 142**,** 2828-2834.

BARDOR, M., FAVEEUW, C., FITCHETTE, A.-C., GILBERT, D. L., GALAS, L., TROTTEIN, F. O., FAYE, L. Ø. & LEROUGE, P. 2003. Immunoreactivity in mammals of two typical plant glyco-epitopes, core Œ± (1, 3)-fucose and core xylose. *Glycobiology,* 13**,** 427-434.

BARDOR, M., NGUYEN, D. H., DIAZ, S. & VARKI, A. 2005. Mechanism of uptake and incorporation of the non-human sialic acid N-glycolylneuraminic acid into human cells. *Journal of Biological Chemistry,* 280**,** 4228-4237.

BETTERIDGE, A. & WATKINS, W. M. 1983. Two α-3-d-Galactosyltransferases in Rabbit Stomach Mucosa with Different Acceptor Substrate Specificities. *European Journal of Biochemistry,* 132**,** 29-35.

BIESECKER, G., DIHEL, L., ENNEY, K. & BENDELE, R. 1999. Derivation of RNA aptamer inhibitors of human complement C5. *Immunopharmacology,* 42**,** 219-230.

BLIX, G., LINDBERG, E., ODIN, L. & WERNER, I. 1955. Sialic acids.

BOCK, L. C., GRIFFIN, L. C., LATHAM, J. A., VERMAAS, E. H. & TOOLE, J. J. 1992. Selection of single-stranded DNA molecules that bind and inhibit human thrombin.

BONES, J., MITTERMAYR, S., O'DONOGHUE, N., GUTTMAN, A. & RUDD, P. M. 2010. Ultra performance liquid chromatographic profiling of serum N-glycans for fast and efficient identification of cancer associated alterations in glycosylation. *Analytical chemistry,* 82**,** 10208-10215.

BORK, K., HORSTKORTE, R. D. & WEIDEMANN, W. 2009. Increasing the sialylation of therapeutic glycoproteins: the potential of the sialic acid biosynthetic pathway. *Journal of pharmaceutical sciences,* 98**,** 3499-3508.

BOSCHER, C., DENNIS, J. W. & NABI, I. R. 2011. Glycosylation, galectins and cellular signaling. *Current opinion in cell biology,* 23**,** 383-392.

BOSQUES, C. J., COLLINS, B. E., MEADOR, J. W., SARVAIYA, H., MURPHY, J. L., DELLORUSSO, G., BULIK, D. A., HSU, I. H., WASHBURN, N., SIPSEY, S. F., MYETTE, J. R., RAMAN, R., SHRIVER, Z., SASISEKHARAN, R. & VENKATARAMAN, G. 2010. Chinese hamster ovary cells can produce galactose-[alpha]-1,3-galactose antigens on proteins. *Nature Biotechnology,* 28**,** 1153-1156.

BOUHOURS, J.-F., RICHARD, C., RUVOEN, N., BARREAU, N., NAULET, J. & BOUHOURS, D. 1998. Characterization of a polyclonal anti-Gal-α-(1-3Gal) antibody from chicken. *Glycoconjugate journal,* 15**,** 93-99.

BRINKMAN-VAN DER LINDEN, E. C., SJOBERG, E. R., JUNEJA, L. R., CROCKER, P. R., VARKI, N. & VARKI, A. 2000. Loss of N-Glycolylneuraminic Acid in Human Evolution Implications for sialic acid recognition by Siglecs *Journal of Biological Chemistry,* 275**,** 8633-8640.

BROCKHAUSEN, I., SCHACHTER, H. & STANLEY, P. 2009. O-GalNAc glycans.

BRODIN, T., TORDSSON, J. & KARLSTROM, P. J. 2004. Identification of target structures E.G. in vivo selection method for a phage library. Google Patents. US 6727062 B1.

BRUGGINK, C., MAURER, R., HERRMANN, H., CAVALLI, S. & HOEFLER, F. 2005. Analysis of carbohydrates by anion exchange chromatography and mass spectrometry. *Journal of Chromatography A,* 1085**,** 104-109.

BYRES, E., PATON, A. W., PATON, J. C., LOFLING, J. C., SMITH, D. F., WILCE, M. C., TALBOT, U. M., CHONG, D. C., YU, H. & HUANG, S. 2008. Incorporation of a non-human glycan mediates human susceptibility to a bacterial toxin. *Nature,* 456**,** 648-652.

CAMPBELL, B. J., YU, L.-G. & RHODES, J. M. 2001. Altered glycosylation in inflammatory bowel disease: a possible role in cancer development. *Glycoconjugate journal,* 18**,** 851-858.

CAMPBELL, M. P., ROYLE, L., RADCLIFFE, C. M., DWEK, R. A. & RUDD, P. M. 2008. GlycoBase and autoGU: tools for HPLC-based glycan analysis. *Bioinformatics,* 24**,** 1214-1216.

CERONI, A., MAASS, K., GEYER, H., GEYER, R., DELL, A. & HASLAM, S. M. 2008. GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans,Ä†. *Journal of proteome research,* 7**,** 1650-1659.

CHEN, S., LAROCHE, T., HAMELINCK, D., BERGSMA, D., BRENNER, D., SIMEONE, D., BRAND, R. E. & HAAB, B. B. 2007. Multiplexed analysis of glycan variation on native proteins captured by antibody microarrays. *Nature methods,* 4**,** 437-444.

CHEN, Y., HU, D., YABE, R., TATENO, H., QIN, S.-Y., MATSUMOTO, N., HIRABAYASHI, J. & YAMAMOTO, K. 2011. Role of malectin in Glc2Man9GlcNAc2-dependent quality control of α1-antitrypsin. *Molecular biology of the cell,* 22**,** 3559-3570.

CHO, S., LEE, B. Ä., CHO, B. Ä., KIM, J. Ä. & KIM, B. Ä. 2013. In vitro selection of sialic acid specific RNA aptamer and its application to the rapid sensing of sialic acid modified sugars. *Biotechnology and bioengineering,* 110**,** 905-913.

CLAUSEN, H. & BENNETT, E. P. 1996. A family of UDP-GalNAc: polypeptide N-acetylgalactosaminyl-transferases control the initiation of mucin-type O-linked glycosylation. *Glycobiology,* 6**,** 635-646.

CRICK, F. 1970. Central dogma of molecular biology. *Nature,* 227**,** 561-563.

CUNNINGHAM, S., STARR, E., SHAW, I., GLAVIN, J., KANE, M. & JOSHI, L. 2012. Development of a Convenient Competitive ELISA for the Detection of the Free and Protein-Bound Nonhuman Galactosyl-α-(1,3)-Galactose Epitope Based on Highly Specific Chicken Single-Chain Antibody Variable-Region Fragments. *Analytical chemistry,* 85**,** 949-955.

DALPATHADO, D. S. & DESAIRE, H. 2008. Glycopeptide analysis by mass spectrometry. *Analyst,* 133**,** 731-738.

DANIOTTI, J. L., VILCAES, A. A., DEMICHELIS, V. T., RUGGIERO, F. M. & RODRIGUEZ-WALKER, M. 2013. Glycosylation of glycolipids in cancer: basis for development of novel therapeutic approaches. *Frontiers in oncology,* 3.

DENNIS, J. W., LAU, K. S., DEMETRIOU, M. & NABI, I. R. 2009. Adaptive Regulation at the Cell Surface by N-Glycosylation. *Traffic,* 10**,** 1569-1578.

DIAZ, S., GAGNEUX, P., MUCHMORE, E., TANGVORANUNTAKUL, P., VARKI, A. & VARKI, N. 2005. Methods for detecting and analyzing N-glycolylneuraminic acid (Neu5gc) in biological materials. Google Patents. WO 2005010485 A2.

DONOHOE, G. G., BYRNE, B., HEARTY, S. H. & O'KENNEDY, R. 2011. Anti-sialic acid antibody molecules. Google Patents. US 20110034676 A1.

DRICKAMER, K. 1988. Two distinct classes of carbohydrate-recognition domains in animal lectins. *J. biol. Chem,* 263**,** 9557-9560.

DRICKAMER, K. & TAYLOR, M. E. 1993. Biology of animal lectins. *Annual review of cell biology,* 9**,** 237-264.

DUBE, D. H. & BERTOZZI, C. R. 2005. Glycans in cancer and inflammation-potential for therapeutics and diagnostics. *Nature Reviews Drug Discovery,* 4**,** 477-488.

DUBE, D. H., PRESCHER, J. A., QUANG, C. N. & BERTOZZI, C. R. 2006. Probing mucin-type O-linked glycosylation in living animals. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 4819-4824.

DURRANT, L., NOBLE, P. & SPENDLOVE, I. 2012. Immunology in the clinic review series; focus on cancer: glycolipids as targets for tumour immunotherapy. *Clinical & Experimental Immunology,* 167**,** 206-215.

DURRANT, L. G. & PARSONS, T. 2011. Monoclonal antibody Sc104 and derivative thereof specifically binding to a sialyltetraosyl carbohydrate as a potential anti-tumor therapeutic agent. Google Patents. US 8343490 B2.

EATON, B. E., GOLD, L. & ZICHI, D. A. 1995. Let's get specific: the relationship between specificity and affinity. *Chemistry & biology,* 2**,** 633-638.

ELLINGTON, A. D. & SZOSTAK, J. W. 1990. In vitro selection of RNA molecules that bind specific ligands. *Nature,* 346**,** 818-22.

FERNANDES, D. 2004. Reducing risk in biopharmaceutical production by controlling glycosylation. *European Biopharmaceutical Review***,** 92-97.

FERREIRA, C., MATTHEWS, C. & MISSAILIDIS, S. 2006. DNA aptamers that bind to MUC1 tumour marker: design and characterization of MUC1-binding single-stranded DNA aptamers. *Tumor Biology,* 27**,** 289-301.

FLEISCHER, B. 1977. Localization of some glycolipid glycosylating enzymes in the Golgi apparatus of rat kidney. *Journal of supramolecular structure,* 7**,** 79-89.

FRANK, R. 2002. The SPOT-synthesis technique: synthetic peptide arrays on membrane supports-principles and applications. *Journal of immunological methods,* 267**,** 13-26.

FREEZE, H. & SCHACHTER, H. 2009. Genetic disorders of glycosylation.

FUKUSAKI, E.-I., KATO, T., MAEDA, H., KAWAZOE, N., ITO, Y., OKAZAWA, A., KAJIYAMA, S.-I. & KOBAYASHI, A. 2000. DNA aptamers that bind to chitin. *Bioorganic & medicinal chemistry letters,* 10**,** 423-425.

GABIUS, H.-J., SIEBERT, H.-C., ANDRE, S., JIMENEZ-BARBERO, J. & RUDIGER, H. 2004. Chemical biology of the sugar code. *ChemBioChem,* 5**,** 740-764.

GABIUS, H. Ä. 1988. Tumor lectinology: at the intersection of carbohydrate chemistry, biochemistry, cell biology, and oncology. *Angewandte Chemie International Edition in English,* 27**,** 1267-1276.

GALILI, U. 2005. The α-gal epitope and the anti-Gal antibody in xenotransplantation and in cancer immunotherapy. *Immunology and cell biology,* 83**,** 674-686.

GALILI, U., LATEMPLE, D. C. & RADIC, M. Z. 1998. A sensitive assay for measuring α-gal epitope expression on cells by a monoclonal anti-gal antibody. *Transplantation,* 65**,** 1129-1132.

GALILI, U., MACHER, B. A., BUEHLER, J. & SHOHET, S. B. 1985. Human natural anti-alpha-galactosyl IgG. II. The specific recognition of alpha (1----3)-linked galactose residues. *The Journal of experimental medicine,* 162**,** 573-582.

GALILI, U., RACHMILEWITZ, E., PELEG, A. & FLECHNER, I. 1984. A unique natural human IgG antibody with anti-alpha-galactosyl specificity. *The Journal of experimental medicine,* 160**,** 1519-1531.

GALILI, U. & SWANSON, K. 1991. Gene sequences suggest inactivation of alpha-1, 3-galactosyltransferase in catarrhines after the divergence of apes from monkeys. *Proceedings of the National Academy of Sciences,* 88**,** 7401-7404.

GERHARD, M., LEHN, N., NEUMAYER, N., BOREN, T., RAD, R., SCHEPP, W., MIEHLKE, S., CLASSEN, M. & PRINZ, C. 1999. Clinical relevance of the Helicobacter pylori gene for blood-group antigen-binding adhesin. *Proceedings of the National Academy of Sciences,* 96**,** 12778-12783.

GERLACH, J. Q., KILCOYNE, M., EATON, S., BHAVANANDAN, V. & JOSHI, L. 2011. Non-carbohydrate-mediated interaction of lectins with plant proteins. *The Molecular Immunology of Complex Carbohydrates-3,* 257-269. Springer. 1441978763

GERLACH, J. Q., SHARMA, S., LEISTER, K. J. & JOSHI, L. 2012. A Tight-Knit Group: Protein Glycosylation, Endoplasmic Reticulum Stress and the Unfolded Protein Response. *Endoplasmic Reticulum Stress in Health and Disease*, 23-39. Springer. 9400743505

GERNGROSS, T. U. 2004. Advances in the production of human therapeutic proteins in yeasts and filamentous fungi. *Nature biotechnology,* 22**,** 1409-1414.

GHADERI, D., TAYLOR, R. E., PADLER-KARAVANI, V., DIAZ, S. & VARKI, A. 2010. Implications of the presence of N-glycolylneuraminic acid in recombinant therapeutic glycoproteins. *Nature biotechnology,* 28**,** 863-867.

GHADERI, D., ZHANG, M., HURTADO-ZIOLA, N. & VARKI, A. 2012. Production platforms for biotherapeutic glycoproteins. Occurrence, impact, and challenges of non-human sialylation. *Biotechnology and Genetic Engineering Reviews,* 28**,** 147-176.

GOLD, L., ZICHI, D. A., JENISON, R. D. & SCHNEIDER, D. J. 2003. By contacting them with the target molecule; separating and amplifying using SELEX process. Google Patents. US 6569620 B1.

GOLD, L., ZICHI, D. A. & SMITH, J. D. 2002. Modified SELEX processes without purified protein. Google Patents. US 6376190 B1.

GOLDSTEIN, I., BLAKE, D., EBISU, S., WILLIAMS, T. & MURPHY, L. 1981. Carbohydrate binding studies on the Bandeiraea simplicifolia I isolectins. Lectins which are mono-, di-, tri-, and tetravalent for N-acetyl-D-galactosamine. *Journal of biological chemistry,* 256**,** 3890-3893.

GOLDSTEIN, I. J. & PORETZ, R. D. 1986. Isolation, Physicochemical Characterization, and Carbohydrate-Binding Specificity of Lectins. *The lectins, properties, functions, and applications in biology and medicine. Academic Press, Orlando, USA,* 33-247.

GOMORD, V. R. & FAYE, L. Ã. 2004. Posttranslational modification of therapeutic proteins in plants. *Current opinion in plant biology,* 7**,** 171-181.

GONG, S., REN, H.-L., TIAN, R.-Y., LIN, C., HU, P., LI, Y.-S., LIU, Z.-S., SONG, J., TANG, F. & ZHOU, Y. 2013. A novel analytical probe binding to a potential carcinogenic factor of N-glycolylneuraminic acid by SELEX. *Biosensors and Bioelectronics,* 49**,** 547-554.

GORENSTEIN, D. G., LEARY, J. & LUXON, B. A. 2011. Structure based and combinatorially selected oligonucleoside phosphorothioate and phosphorodithioate aptamer targeting AP-1 transcription factors. Google Patents. US 7910523 B2.

GORNIK, O. & LAUC, G. 2007. Enzyme linked lectin assay (ELLA) for direct analysis of transferrin sialylation in serum samples. *Clinical biochemistry,* 40**,** 718-723.

GOULABCHAND, R., VINCENT, T., BATTEUX, F. D. R., ELIAOU, J.-F. O. & GUILPAIN, P. 2014. Impact of autoantibody glycosylation in autoimmune diseases. *Autoimmunity reviews,* 13**,** 742-750.

GRAM, H., MARCONI, L.-A., BARBAS, C. F., COLLET, T. A., LERNER, R. A. & KANG, A. S. 1992. In vitro selection and affinity maturation of antibodies from a naive combinatorial immunoglobulin library. *Proceedings of the National Academy of Sciences,* 89**,** 3576-3580.

GREEN, L. S., JELLINEK, D., JENISON, R., OSTMAN, A., HELDIN, C.-H. & JANJIC, N. 1996. Inhibitory DNA ligands to platelet-derived growth factor B-chain. *Biochemistry,* 35**,** 14413-14424.

GREGORIADIS, G., JAIN, S., PAPAIOANNOU, I. & LAING, P. 2005. Improving the therapeutic efficacy of peptides and proteins: a role for polysialic acids. *International journal of pharmaceutics,* 300**,** 125-130.

GRUNEWALD, S., MATTHIJS, G. & JAEKEN, J. 2002. Congenital disorders of glycosylation: a review. *Pediatric Research,* 52**,** 618-624.

GULL, I., WIRTH, M. & GABOR, F. 2007. Development of a sensitive and reliable ELISA for quantification of wheat germ agglutinin. *Journal of immunological methods,* 318**,** 20-29.

HAAB, B. B. 2010. Antibody-lectin sandwich arrays for biomarker and glycobiology studies.

HAGEMEYER, C. E., VON ZUR MUHLEN, C., VON ELVERFELDT, D. & PETER, K. 2009. Single-chain antibodies as diagnostic tools and therapeutic agents. *Thromb Haemost,* 101**,** 1012-1019.

HAKOMORI, S. 2002. Glycosylation defining cancer malignancy: new wine in an old bottle. *Proceedings of the National Academy of Sciences,* 99**,** 10231-10233.

HAKOMORI, S.-I. 1981. Glycosphingolipids in cellular interaction, differentiation, and oncogenesis. *Annual review of biochemistry,* 50**,** 733-764.

HAKOMORI, S.-I. 1985. Aberrant glycosylation in cancer cell membranes as focused on glycolipids: overview and perspectives. *Cancer research,* 45**,** 2405-2414.

HAMILTON, W., HELLING, F., LLOYD, K. O. & LIVINGSTON, P. O. 1993. Ganglioside expression on human malignant melanoma assessed by quantitative immune thin-layer chromatography. *International journal of cancer,* 53**,** 566-573.

HARZ, H., BURGDORF, K. & HOLTJE, J.-V. 1990. Isolation and separation of the glycan strands from murein of *Escherichia coli* by reversed-phase high-performance liquid chromatography. *Analytical biochemistry,* 190**,** 120-128.

HEBERT, E. 2000. Endogenous lectins as cell surface transducers. *Bioscience reports,* 20**,** 213-237.

HEDLUND, M., TANGVORANUNTAKUL, P., TAKEMATSU, H., LONG, J. M., HOUSLEY, G. D., KOZUTSUMI, Y., SUZUKI, A., WYNSHAW-BORIS, A., RYAN, A. F. & GALLO, R. L. 2007. N-glycolylneuraminic acid deficiency in mice: implications for human biology and evolution. *Molecular and Cellular biology,* 27**,** 4340-4346.

HEIMBURG-MOLINARO, J., LUM, M., VIJAY, G., JAIN, M., ALMOGREN, A. & RITTENHOUSE-OLSON, K. 2011. Cancer vaccines and carbohydrate epitopes. *Vaccine,* 29**,** 8802-8826.

HELENIUS, A. & AEBI, M. 2001. Intracellular functions of N-linked glycans. *Science,* 291**,** 2364-2369.

HEYDUK, T., HEYDUK, E. & KNOLL, E. 2011. Biosensors for detecting macromolecules and other analytes. Google Patents. US 7939313 B2.

HIGA, H. H., ROGERS, G. N. & PAULSON, J. C. 1985. Influenza virus hemagglutinins differentiate between receptor determinants bearing *N*-acetyl-, *N*-glycollyl- and *N,O*-diacetyineuraminic acids. *Virology,* 144**,** 279-282.

HIGASHI, H., HIRABAYASHI, Y., FUKUI, Y., NAIKI, M., MATSUMOTO, M., UEDA, S. & KATO, S. 1985. Characterization of N-glycolylneuraminic acid-containing gangliosides as tumor-associated Hanganutziu-Deicher antigen in human colon cancer. *Cancer research,* 45**,** 3796-3802.

HOELLENRIEGEL, J., ZBORALSKI, D., ESTROV, Z., WIERDA, W. G., KEATING, M., KRUSCHINSKI, A. & BURGER, J. A. 2011. The Spiegelmer Nox-A12, a novel SDF-1 (CXCL12) inhibitor, and its effects on chronic lymphocytic leukemia (CLL) cell migration. *Blood,* 118**,** 1657-1658.

HOKKE, C. H., BERGWERFF, A. A., VAN DEDEM, G. W., VAN OOSTRUM, J., KAMERLING, J. P. & VLIEGENTHART, J. F. 1990. Sialylated carbohydrate chains of recombinant human glycoproteins expressed in Chinese hamster ovary cells contain traces of *N*-glycolylneuraminic acid. *FEBS letters,* 275**,** 9-14.

HSU, K.-L. & MAHAL, L. K. 2006. A lectin microarray approach for the rapid analysis of bacterial glycans. *Nature protocols,* 1**,** 543-549.

HULL, R. A., GILL, R. E., HSU, P., MINSHEW, B. H. & FALKOW, S. 1981. Construction and expression of recombinant plasmids encoding type 1 or D-mannose-resistant pili from a urinary tract infection Escherichia coli isolate. *Infection and immunity,* 33**,** 933-938.

HULSMEIER, A. J., WELTI, M. & HENNET, T. 2011. Glycoprotein maturation and the UPR. *METHODS IN ENZYMOLOGY, VOL 491: UNFOLDED PROTEIN RESPONSE AND CELLULAR STRESS, PT C,* 491**,** 163-182.

ILVER, D., ARNQVIST, A., OGREN, J., FRICK, I.-M., KERSULYTE, D., INCECIK, E. T., BERG, D. E., COVACCI, A., ENGSTRAND, L. & BOREN, T. 1998. Helicobacter pylori adhesin binding fucosylated histo-blood group antigens revealed by retagging. *Science,* 279**,** 373-377.

IMBERTY, A. & PEREZ, S. 1994. Molecular modelling of protein-carbohydrate interactions. Understanding the specificities of two legume lectins towards oligosaccharides. *Glycobiology,* 4**,** 351-366.

JAEKEN, J. & CARCHON, H. 1993. The carbohydrate-deficient glycoprotein syndromes: an overview. *Journal of inherited metabolic disease,* 16**,** 813-820.

JAEKEN, J. & VAN DEN HEUVEL, L. 2014. Congenital disorders of glycosylation. *Physician's Guide to the Diagnosis, Treatment, and Follow-Up of Inherited Metabolic Diseases*, 483-512. Springer. 3642403360

JAEKEN, J., VANDERSCHUEREN-LODEWEYCKX, M., CASAER, P., SNOECK, L., CORBEEL, L., EGGERMONT, E. & EECKELS, R. 1980. Familial psychomotor retardation with markedly fluctuating serum prolactin, FSH and GH levels, partial TBG-deficiency, increased serum arylsulphatase A and increased CSF protein: a new syndrome?: 90. *Pediatric Research,* 14**,** 179-179.

JEONG, H. G., OH, M. H., KIM, B. S., LEE, M. Y., HAN, H. J. & CHOI, S. H. 2009. The capability of catabolic utilization of *N*-acetylneuraminic acid, a sialic acid, is essential for *Vibrio vulnificus* pathogenesis. *Infection and immunity,* 77**,** 3209-3217.

JOZIASSE, D. & ORIOL, R. 1999. Xenotransplantation: the importance of the Gal α1,3Gal epitope in hyperacute vascular rejection. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease,* 1455**,** 403-418.

KANNAGI, R. 2003. Molecular mechanism for cancer-associated induction of sialyl Lewis X and sialyl Lewis A expression‚Äîthe Warburg effect revisited. *Glycoconjugate journal,* 20**,** 353-364.

KEEFE, A. D., PAI, S. & ELLINGTON, A. 2010. Aptamers as therapeutics. *Nature Reviews Drug Discovery,* 9**,** 537-550.

KELM, S., PAULSON, J. C., ROSE, U., BROSSMER, R., SCHMID, W., BANDGAR, B. P., SCHREINER, E., HARTMANN, M. & ZBIRAL, E. 1992. Use of sialic acid analogues to define functional groups involved in binding to the influenza virus hemagglutinin. *European Journal of Biochemistry,* 205**,** 147-153.

KELM, S. & SCHAUER, R. 1997. Sialic acids in molecular and cellular interactions. *International review of cytology,* 175**,** 137-240.

KILCOYNE, M. & JOSHI, L. 2007. Carbohydrates in therapeutics. *Cardiovascular & Hematological Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Cardiovascular & Hematological Agents),* 5**,** 186-197.

KLEMM, P. & SCHEMBRI, M. A. 2000. Bacterial adhesins: function and structure. *International journal of medical microbiology,* 290**,** 27-35.

KNUCHEL, M., ACKERMANN, M., MULLER, H. & KIHM, U. 1992. An ELISA for detection of antibodies against porcine epidemic diarrhoea virus (PEDV) based on the specific solubility of the viral surface glycoprotein. *Veterinary microbiology,* 32**,** 117-134.

KOBATA, A. 1996. c Cancer cells and metastasis: The Warren-Glick phenomenon-a molecular basis of tumorigenesis and metastasis. *New Comprehensive Biochemistry,* 30**,** 211-227.

KOBATA, A. 1998. A retrospective and prospective view of glycopathology. *Glycoconjugate journal,* 15**,** 323-331.

KOIKE, C., FUNG, J. J., GELLER, D. A., KANNAGI, R., LIBERT, T., LUPPI, P., NAKASHIMA, I., PROFOZICH, J., RUDERT, W. & SHARMA, S. B. 2002. Molecular basis of evolutionary loss of the α-1, 3-galactosyltransferase gene in higher primates. *Journal of biological chemistry,* 277**,** 10114-10120.

KOOPMAN, L. A., KOPCOW, H. D., RYBALOV, B., BOYSON, J. E., ORANGE, J. S., SCHATZ, F., MASCH, R., LOCKWOOD, C. J., SCHACHTER, A. D. & PARK, P. J. 2003. Human decidual natural killer cells are a unique NK cell subset with immunomodulatory potential. *The Journal of experimental medicine,* 198**,** 1201-1212.

KRIPLANI, U. & KAY, B. K. 2005. Selecting peptides for use in nanoscale materials using phage-displayed combinatorial peptide libraries. *Current opinion in biotechnology,* 16**,** 470-475.

KROGFELT, K. A., BERGMANS, H. & KLEMM, P. 1990. Direct evidence that the FimH protein is the mannose-specific adhesin of Escherichia coli type 1 fimbriae. *Infection and immunity,* 58**,** 1995-1998.

LADJEMI, M. Z., CHARDES, T., CORGNAC, S., GARAMBOIS, V., MORISSEAU, S., ROBERT, B., BASCOUL-MOLLEVI, C., AIT ARSA, I., JACOT, W. & POUGET, J.-P. 2011. Vaccination with human anti-trastuzumab anti-idiotype scFv reverses HER2 immunological tolerance and induces tumor immunity in MMTV. f. huHER2 (Fo5) mice. *Breast Cancer Res,* 13**,** R17.

LAINE, R. A. 1994. Invited Commentary: A calculation of all possible oligosaccharide isomers both branched and linear yields $1.05\sqrt{}$ ó 1012 structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology,* 4**,** 759-767.

LAMBRE, C. R., TERZIDIS, H. L. N., GREFFARD, A. & WEBSTER, R. G. 1991. An enzyme-linked lectin assay for sialidase. *Clinica chimica acta,* 198**,** 183-193.

LANCTOT, P. M., GAGE, F. H. & VARKI, A. P. 2007. The glycans of stem cells. *Current opinion in chemical biology,* 11**,** 373-380.

LANDEMARRE, L., CANCELLIERI, P. & DUVERGER, E. 2013. Cell surface lectin array: parameters affecting cell glycan signature. *Glycoconjugate journal,* 30**,** 195-203.

LANNERT, H., GORGAS, K., MEISNER, I., WIELAND, F. T. & JECKEL, D. 1998. Functional organization of the Golgi apparatus in glycosphingolipid biosynthesis Lactosylceramide and subsequent glycosphingolipids are formed in the lumen of the late Golgi. *Journal of biological chemistry,* 273**,** 2939-2946.

LAU, K. S. & DENNIS, J. W. 2008. N-Glycans in cancer progression. *Glycobiology,* 18**,** 750-760.

LEDEEN, R. 1966. The chemistry of gangliosides: a review. *Journal of the American Oil Chemists Society,* 43**,** 57-66.

LEDEEN, R. 1984. Biology of gangliosides: neuritogenic and neuronotrophic properties. *Journal of neuroscience research,* 12**,** 147-159.

LEROUGE, P., CABANES-MACHETEAU, M., RAYON, C., FISCHETTE-LAIN √ ©, A.-C., GOMORD, V. R. & FAYE, L. Ø. 1998. N-glycoprotein biosynthesis in plants: recent developments and future trends. *Plant molecular biology,* 38**,** 31-48.

LI, M., LIN, N., HUANG, Z., DU, L., ALTIER, C., FANG, H. & WANG, B. 2008. Selecting aptamers for a glycoprotein through the incorporation of the boronic acid moiety. *Journal of the American Chemical Society,* 130**,** 12636-12638.

LIN, S.-Y., LIN, L. & TSAI, Y.-Y. 2011. Antibodies recognizing a carbohydrate containing epitope on CD-43 and CEA expressed on cancer cells and methods using same. Google Patents. US 7982017 B2.

LIN, W.-M., KARSTEN, U., GOLETZ, S., CHENG, R.-C. & CAO, Y. 2010. Co-expression of CD173 (H2) and CD174 (Lewis Y) with CD44 suggests that fucosylated histo-blood group antigens are markers of breast cancer-initiating cells. *Virchows Archiv,* 456**,** 403-409.

LIPPIELLO, L. & MANKIN, H. J. 1971. Thin-layer chromatographic separation of the isomeric chondroitin sulfates, dermatan sulfate, and keratan sulfate. *Analytical biochemistry,* 39**,** 54-58.

LIS, H. & SHARON, N. 1986. Lectins as molecules and as tools. *Annual review of biochemistry,* 55**,** 35-67.

LOFLING, J. C., PATON, A. W., VARKI, N. M., PATON, J. C. & VARKI, A. 2009. A dietary non-human sialic acid may facilitate hemolytic-uremic syndrome. *Kidney international,* 76**,** 140-144.

LOW, S. Y., HILL, J. E. & PECCIA, J. 2009. DNA aptamers bind specifically and selectively to (1, 3)-α-d-glucans. *Biochemical and biophysical research communications,* 378**,** 701-705.

LUND, B. R., LINDBERG, F., MARKLUND, B.-I. & NORMARK, S. 1987. The PapG protein is the alpha-D-galactopyranosyl-(1, 4)-beta-D-galactopyranose-binding adhesin of uropathogenic Escherichia coli. *Proceedings of the National Academy of Sciences,* 84**,** 5898-5902.

LUTHER, K. B. & HALTIWANGER, R. S. 2009. Role of unusual *O*-glycans in intercellular signaling. *The international journal of biochemistry & cell biology,* 41**,** 1011-1024.

LYU, M., KURZROCK, R. & ROSENBLUM, M. G. 2008. The immunocytokine scFv23/TNF targeting HER-2/neu induces synergistic cytotoxic effects with 5-fluorouracil in TNF-resistant pancreatic cancer cell lines. *Biochemical pharmacology,* 75**,** 836-846.

MACCIONI, H. J., GIRAUDO, C. G. & DANIOTTI, J. L. 2002. Understanding the stepwise synthesis of glycolipids. *Neurochemical research,* 27**,** 629-636.

MARTH, J. D. & GREWAL, P. K. 2008. Mammalian glycosylation in immunity. *Nature Reviews Immunology,* 8**,** 874-887.

MARTIN, M. J., MUOTRI, A., GAGE, F. & VARKI, A. 2005. Human embryonic stem cells express an immunogenic nonhuman sialic acid. *Nature medicine,* 11**,** 228-232.

MATOS, F. C. S. & JEAN, G. 2007. Aptamers that recognize the carbohydrate n-acetylgalactosamine (galnac). Google Patents. WO 2007128109 A1.

MCCOY JR, J. P., VARANI, J. & GOLDSTEIN, I. J. 1983. Enzyme-linked lectin assay (ELLA): Use of alkaline phosphatase-conjugated *Griffonia simplicifolia* B isolectin for the detection of α-d-galactopyranosyl end groups. *Analytical biochemistry,* 130**,** 437-444.

MEEZAN, E., WU, H. C., BLACK, P. H. & ROBBINS, P. W. 1969. Comparative studies on the carbohydrate-containing membrane components of normal and virus-transformed mouse fibroblasts. II. Separation of glycoproteins and glycopeptides by sephadex chromatography. *Biochemistry,* 8**,** 2518-2524.

MEHEDI MASUD, M., KUWAHARA, M., OZAKI, H. & SAWAI, H. 2004. Sialyllactose-binding modified DNA aptamer bearing additional functionality by SELEX. *Bioorganic & medicinal chemistry,* 12**,** 1111-1120.

MILLER, R. L., COLLAWN, J. & FISH, W. W. 1982. Purification and macromolecular properties of a sialic acid-specific lectin from the slug Limax flavus. *Journal of biological chemistry,* 257**,** 7574-7580.

MIYOSHI, E., MORIWAKI, K. & NAKAGAWA, T. 2008. Biological function of fucosylation in cancer biology. *Journal of biochemistry,* 143**,** 725-729.

MORAN, A., GUPTA, A. & JOSHI, L. 2011. Sweet-talk: role of host glycosylation in bacterial pathogenesis of the gastrointestinal tract. *Gut***,** gut. 2010.212704.

MORROW, A. L., NEWBURG, D. S. & RUIZ-PALACIOS, G. M. 2013. Use of Secretor, Lewis and Sialyl Antigen Levels in Clinical Samples as Predictors of Risk for Disease. Google Patents. US 20090191551 A1.

NEUMANN, D., LEHR, C.-M., LENHOF, H.-P. & KOHLBACHER, O. 2004. Computational modeling of the sugar-lectin interaction. *Advanced drug delivery reviews,* 56**,** 437-457.

NG, E. W., SHIMA, D. T., CALIAS, P., CUNNINGHAM, E. T., GUYER, D. R. & ADAMIS, A. P. 2006. Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. *Nature Reviews Drug Discovery,* 5**,** 123-132.

NGUYEN, D. H., TANGVORANUNTAKUL, P. & VARKI, A. 2005. Effects of natural human antibodies against a nonhuman sialic acid that metabolically incorporates into activated and malignant immune cells. *The Journal of Immunology,* 175**,** 228-236.

NI, X., CASTANARES, M., MUKHERJEE, A. & LUPOLD, S. E. 2011. Nucleic acid aptamers: clinical applications and promising new horizons. *Current medicinal chemistry,* 18**,** 4206.

NOGUCHI, A., MUKURIA, C., SUZUKI, E. & NAIKI, M. 1996. Failure of human immunoresponse to N-glycolylneuraminic acid epitope contained in recombinant human erythropoietin. *Nephron,* 72**,** 599-603.

OGATA, S.-I., MURAMATSU, T. & KOBATA, A. 1976. New structural characteristic of the large glycopeptides from transformed cells.

OHTSUBO, K. & MARTH, J. D. 2006. Glycosylation in cellular mechanisms of health and disease. *Cell,* 126**,** 855-867.

OLOFSSON, S. & BERGSTROM, T. 2005. Glycoconjugate glycans as viral receptors. *Annals of medicine,* 37**,** 154-172.

OSBORN, H. M., EVANS, P. G., GEMMELL, N. & OSBORNE, S. D. 2004. Carbohydrate¸Äêbased therapeutics. *Journal of pharmacy and pharmacology,* 56**,** 691-702.

PADLER-KARAVANI, V., HURTADO-ZIOLA, N., PU, M., YU, H., HUANG, S., MUTHANA, S., CHOKHAWALA, H. A., CAO, H., SECREST, P. & FRIEDMANN-MORVINSKI, D. 2011. Human xeno-autoantibodies against a non-human sialic acid serve as novel serum biomarkers and immunotherapeutics in cancer. *Cancer research,* 71**,** 3352-3363.

PADLER-KARAVANI, V. & VARKI, A. 2011. Potential impact of the non-human sialic acid *N*-glycolylneuraminic acid on transplant rejection risk. *Xenotransplantation,* 18**,** 1-5.

PELETSKAYA, E. N., GLINSKY, V. V., GLINSKY, G. V., DEUTSCHER, S. L. & QUINN, T. P. 1997. Characterization of peptides that bind the tumor-associated Thomsen-Friedenreich antigen selected from bacteriophage display libraries. *J Mol Biol,* 270**,** 374-384.

PETER-KATALINIC, J. 2005. Methods in Enzymology: O-Glycosylation of Proteins. *Methods in enzymology,* 405**,** 139-171.

PETRYNIAK, J., PETRYNIAK, B., WASNIOWSKA, K. & KROTKIEWSKI, H. 1980. Isolation and immunochemical characterization of the Euonymus europaeus lectin receptor from the major sialoglycoprotein of human O erythrocytes. *European Journal of Biochemistry,* 105**,** 335-341.

PILLER, F., PILLER, V., FOX, R. I. & FUKUDA, M. 1988. Human T-lymphocyte activation is associated with changes in O-glycan biosynthesis. *Journal of biological chemistry,* 263**,** 15146-15150.

PILOBELLO, K. T., SLAWEK, D. E. & MAHAL, L. K. 2007. A ratiometric lectin microarray approach to analysis of the dynamic mammalian glycome. *Proceedings of the National Academy of Sciences,* 104**,** 11534-11539.

PINHO, S., MARCOS, N. T., FERREIRA, B., CARVALHO, A. S., OLIVEIRA, M. J., SANTOS-SILVA, F., HARDUIN-LEPERS, A. & REIS, C. A. 2007. Biological significance of cancer-associated sialyl-Tn antigen: modulation of malignant phenotype in gastric carcinoma cells. *Cancer letters,* 249**,** 157-170.

PLUM, M., MICHEL, Y., WALLACH, K., RAIBER, T., BLANK, S., BANTLEON, F. I., DIETHERS, A., GREUNKE, K., BRAREN, I. & HACKL, T. 2011. Close-up of the Immunogenic α-(1-3)-Galactose Epitope as Defined by a Monoclonal Chimeric Immunoglobulin E and Human Serum Using Saturation Transfer Difference (STD) NMR. *Journal of Biological Chemistry,* 286**,** 43103-43111.

POCHECHUEVA, T., CHINAREV, A., SPENGLER, M., KORCHAGINA, E., HEINZELMANN-SCHWARZ, V., BOVIN, N. & RIEBEN, R. 2011. Multiplex suspension array for human anti-carbohydrate antibody profiling. *Analyst,* 136**,** 560-569.

POINTREAU, Y., COMMINS, S. P., CALAIS, G., WATIER, H. & PLATTS-MILLS, T. A. 2012. Fatal Infusion Reactions to Cetuximab: Role of Immunoglobulin E-Mediated Anaphylaxis. *Journal of Clinical Oncology,* 30**,** 334-334.

RABINOVICH, G. A., TOSCANO, M. A., JACKSON, S. S. & VASTA, G. R. 2007. Functions of cell surface galectin-glycoprotein lattices. *Current opinion in structural biology,* 17**,** 513-520.

REASEARCHANDMARKETS.COM. 2014. Aptamer technology market. *researchandmarkets.com* [Online]. [Accessed 17th November 2014].

REITTER, J. N., MEANS, R. E. & DESROSIERS, R. C. 1998. A role for carbohydrates in immune evasion in AIDS. *Nature medicine,* 4**,** 679-684.

ROBERTS, J. A., MARKLUND, B.-I., ILVER, D., HASLAM, D., KAACK, M. B., BASKIN, G., LOUIS, M., MOLLBY, R., WINBERG, J. & NORMARK, S. 1994. The Gal-α-(1-4)-Gal-specific tip adhesin of Escherichia coli P-fimbriae is needed for pyelonephritis to occur in the normal urinary tract. *Proceedings of the National Academy of Sciences,* 91**,** 11889-11893.

ROCHE, A.-C. & MONSIGNY, M. 1974. Purification and properties of limulin: A lectin (agglutinin) from hemolymph of *Limulus polyphemus*. *Biochimica et Biophysica Acta (BBA)-Protein Structure,* 371**,** 242-254.

ROYLE, L., CAMPBELL, M. P., RADCLIFFE, C. M., WHITE, D. M., HARVEY, D. J., ABRAHAMS, J. L., KIM, Y.-G., HENRY, G. W., SHADICK, N. A. & WEINBLATT, M. E. 2008. HPLC-based analysis of serum N-glycans on a 96-well plate platform with dedicated database software. *Analytical biochemistry,* 376**,** 1-12.

ROYLE, L., MATTU, T. S., HART, E., LANGRIDGE, J. I., MERRY, A. H., MURPHY, N., HARVEY, D. J., DWEK, R. A. & RUDD, P. M. 2002. An analytical and structural database provides a strategy for sequencing O-

glycans from microgram quantities of glycoproteins. *Analytical biochemistry,* 304**,** 70-90.

RUDD, P. M., ELLIOTT, T., CRESSWELL, P., WILSON, I. A. & DWEK, R. A. 2001. Glycosylation and the immune system. *Science,* 291**,** 2370-2376.

RUDDOCK, L. W. & MOLINARI, M. 2006. N-glycan processing in ER quality control. *Journal of cell science,* 119**,** 4373-4380.

RUDIGER, H. & GABIUS, H.-J. 2001. Plant lectins: occurrence, biochemistry, functions and applications. *Glycoconjugate journal,* 18**,** 589-613.

RUSCONI, C. P., ROBERTS, J. D., PITOC, G. A., NIMJEE, S. M., WHITE, R. R., QUICK, G., SCARDINO, E., FAY, W. P. & SULLENGER, B. A. 2004. Antidote-mediated control of an anticoagulant aptamer in vivo. *Nature biotechnology,* 22**,** 1423-1428.

RUSCONI, C. P., SCARDINO, E., LAYZER, J., PITOC, G. A., ORTEL, T. L., MONROE, D. & SULLENGER, B. A. 2002. RNA aptamers as reversible antagonists of coagulation factor IXa. *Nature,* 419**,** 90-94.

SAITO, Y. 1956. An improved method for the preparation of sialic acid or acetylneuraminic acid.

SAMRAJ, A. N., PEARCE, O. M., LAUBLI, H., CRITTENDEN, A. N., BERGFELD, A. K., BANDA, K., GREGG, C. J., BINGMAN, A. E., SECREST, P. & DIAZ, S. L. 2014. A red meat-derived glycan promotes inflammation and cancer progression. *Proceedings of the National Academy of Sciences***,** 201417508.

SARKAR, M., WU, A. M. & KABAT, E. A. 1981. Immunochemical studies on the carbohydrate specificity of *Maclura pomifera* lectin. *Archives of biochemistry and biophysics,* 209**,** 204-218.

SATA, T., LACKIE, P., TAATJES, D., PEUMANS, W. & ROTH, J. 1989. Detection of the Neu5 Ac (alpha 2, 3) Gal (beta 1, 4) GlcNAc sequence with the leukoagglutinin from Maackia amurensis: light and electron microscopic demonstration of differential tissue expression of terminal sialic acid in alpha 2, 3-and alpha 2, 6-linkage. *Journal of Histochemistry & Cytochemistry,* 37**,** 1577-1588.

SATO, C., MATSUDA, T. & KITAJIMA, K. 2002. Neuronal differentiation-dependent expression of the disialic acid epitope on CD166 and its involvement in neurite formation in Neuro2A cells. *Journal of biological chemistry,* 277**,** 45299-45305.

SCHACHTER, H. 2001. The clinical relevance of glycobiology. *Journal of Clinical Investigation,* 108**,** 1579-1582.

SCHNAAR, R. L. 1991. Glycosphingolipids in cell surface recognition. *Glycobiology,* 1**,** 477-485.

SCHNEIDER, P., RALTON, J., MCCONVILLE, M. & FERGUSON, M. 1993. Analysis of the neutral glycan fractions of glycosyl-phosphatidylinositols by thin-layer chromatography. *Analytical biochemistry,* 210**,** 106-112.

SCHWOEBEL, D. F., SIMONE, S., CHRISTIAN, M., WERNER, P., DIRK, Z., JOHN, T., DIRK, E., MARTIN, H. & SVEN, K. Podium# 91. NOX-H94, in vitro and in vivo characterization of a Hepidin bolcking speigelmer. Fourth Congress of the International BioIron Society (IBIS) Biennial World Meeting (BioIron 2011) May 22-26, 2011 Sheraton Vancouver Wall Centre Hotel Vancouver, BC, Canada. American Journal of Hematology, 2011. E47.

SEO, J. & LEE, K.-J. 2004. Post-translational modifications and their biological functions: proteomic analysis and systematic approaches. *Journal of biochemistry and molecular biology,* 37**,** 35-44.

SESMA, J. I., ESTHER, C. R., KREDA, S. M., JONES, L., O'NEAL, W., NISHIHARA, S., NICHOLAS, R. A. & LAZAROWSKI, E. R. 2009. Endoplasmic reticulum/golgi nucleotide sugar transporters contribute to the cellular release of UDP-sugar signaling molecules. *Journal of biological chemistry,* 284**,** 12572-12583.

SETHURAMAN, N. & STADHEIM, T. A. 2006. Challenges in therapeutic glycoprotein production. *Current opinion in biotechnology,* 17**,** 341-346.

SHARON, N. & LIS, H. 1989. Lectins as cell recognition molecules. *Science,* 246**,** 227-234.

SHARON, N. & LIS, H. 1993. Carbohydrates in cell recognition. *Scientific American,* 268**,** 82-89.

SHRIVER, Z., RAGURAM, S. & SASISEKHARAN, R. 2004. Glycomics: a pathway to a class of new and improved therapeutics. *Nature Reviews Drug Discovery,* 3**,** 863-873.

SINCLAIR, A. M. & ELLIOTT, S. 2005. Glycoengineering: the effect of glycosylation on the properties of therapeutic proteins. *Journal of pharmaceutical sciences,* 94**,** 1626-1635.

SMITH, G. P. 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science,* 228**,** 1315-1317.

SMITH, H. J. 2013. Aptamer Based Point-of-Care Test for Glycated Albumin. Google Patents. US 20090042237 A1.

SOUNDARARAJAN, S., CHEN, W., SPICER, E. K., COURTENAY-LUCK, N. & FERNANDES, D. J. 2008. The nucleolin targeting aptamer AS1411 destabilizes Bcl-2 messenger RNA in human breast cancer cells. *Cancer research,* 68**,** 2358-2365.

STANLEY, P. 2011. Golgi glycosylation. *Cold Spring Harbor perspectives in biology,* 3**,** a005199.

SUN, W., DU, L. & LI, M. 2010. Aptamer-based carbohydrate recognition. *Current pharmaceutical design,* 16**,** 2269-2278.

SUSEELAN, K., BHATIA, C. & MITRA, R. 1997. Characteristics of two major lectins from mungbean (Vigna radiata) seeds. *Plant foods for human nutrition,* 50**,** 211-222.

SVAROVSKY, S. A. & JOSHI, L. 2008. Biocombinatorial selection of carbohydrate binding agents of therapeutic significance. *Current drug discovery technologies,* 5**,** 20-28.

SVENNERHOLM, L., BOSTROM, K., FREDMAN, P., MANSSON, J.-E., ROSENGREN, B. & RYNMARK, B.-M. 1989. Human brain gangliosides: developmental changes from early fetal stage to advanced age. *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism,* 1005**,** 109-117.

TAATJES, D., ROTH, J. R., PEUMANS, W. & GOLDSTEIN, I. J. 1988. Elderberry bark lectin-gold techniques for the detection of Neu5Ac (α-2, 6) Gal/GalNAc sequences: applications and limitations. *The Histochemical journal,* 20**,** 478-490.

TABARES, G., RADCLIFFE, C. M., BARRABES, S., RAMIREZ, M., ALEIXANDRE, R. N., HOESEL, W., DWEK, R. A., RUDD, P. M., PERACAULA, R. & DE LLORENS, R. 2006. Different glycan structures in prostate-specific antigen from prostate cancer sera in relation to seminal plasma PSA. *Glycobiology,* 16**,** 132-145.

TANGVORANUNTAKUL, P., GAGNEUX, P., DIAZ, S., BARDOR, M., VARKI, N., VARKI, A. & MUCHMORE, E. 2003. Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid. *Proceedings of the National Academy of Sciences,* 100**,** 12045-12050.

TAYLOR, R. E., GREGG, C. J., PADLER-KARAVANI, V., GHADERI, D., YU, H., HUANG, S., SORENSEN, R. U., CHEN, X., INOSTROZA, J. & NIZET, V. 2010. Novel mechanism for the generation of human xeno-autoantibodies

against the nonhuman sialic acid N-glycolylneuraminic acid. *The Journal of experimental medicine,* 207**,** 1637-1646.

THOMPSON, T. & TILLACK, T. W. 1985. Organization of glycosphingolipids in bilayers and plasma membranes of mammalian cells. *Annual review of biophysics and biophysical chemistry,* 14**,** 361-386.

TORRES, C.-R. & HART, G. W. 1984. Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. *Journal of biological chemistry,* 259**,** 3308-3317.

TOTHILL, I. E. Biosensors for cancer markers diagnosis. Seminars in cell & developmental biology, 2009. Elsevier, 55-62.

TOWBIN, H., ROSENFELDER, G., WIESLANDER, J., AVILA, J., ROJAS, M., SZARFMAN, A., ESSER, K., NOWACK, H. & TIMPL, R. 1987. Circulating antibodies to mouse laminin in Chagas disease, American cutaneous leishmaniasis, and normal individuals recognize terminal galactosyl α-(1-3)-galactose epitopes. *The Journal of experimental medicine,* 166**,** 419-432.

TRIMBLE, R., MALEY, F. & CHU, F. 1983. GlycoProtein biosynthesis in yeast. protein conformation affects processing of high mannose oligosaccharides on carboxypeptidase Y and invertase. *Journal of biological chemistry,* 258**,** 2562-2567.

TUERK, C. & GOLD, L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science,* 249**,** 505-10.

UCAKTURK, E. 2012. Analysis of glycoforms on the glycosylation site and the glycans in monoclonal antibody biopharmaceuticals. *Journal of separation science,* 35**,** 341-350.

VALLEJO, M. C., MATSUO, A. L., GANIKO, L., MEDEIROS, L. C. S., MIRANDA, K., SILVA, L. S., FREYMULLER-HAAPALAINEN, E., SINIGAGLIA-COIMBRA, R., ALMEIDA, I. C. & PUCCIA, R. 2011. The Pathogenic Fungus Paracoccidioides brasiliensis Exports Extracellular Vesicles Containing Highly Immunogenic α-Galactosyl Epitopes. *Eukaryotic cell,* 10**,** 343-351.

VARKI, A. 1993. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology,* 3**,** 97-130.

VARKI, A. 2001. Loss of N-*glycolylneuraminic* acid in humans: Mechanisms, consequences, and implications for hominid evolution. *American journal of physical anthropology,* 116**,** 54-69.

VARKI, A., CUMMINGS, R., ESKO, J., FREEZE, H., HART, G. & MARTH, J. 1999. N-glycans. *In:* VARKI A, C. R., ESKO J, ET AL. (ed.) *Essentials of Glycobiology*, Chapter 7. New York: Cold Spring Harbor Laboratory Press. 10: 0-87969-559-5

VARKI, A., CUMMINGS, R. D., ESKO, J. D., FREEZE, H. H., STANLEY, P., BERTOZZI, C. R., HART, G. W., ETZLER, M. E. & LOWE, J. B. 2009a. Biological roles of glycans. *In:* VARKI A, C. R., ESKO JD, ET AL. (ed.) *Essentials of Glycobiology*, 2, Chapter 6. Cold Spring Harbor Laboratory Press. 13: 9780879697709

VARKI, A., CUMMINGS, R. D., ESKO, J. D., FREEZE, H. H., STANLEY, P., BERTOZZI, C. R., HART, G. W., ETZLER, M. E. & SCHAUER, R. 2009b. Sialic acids. *In:* VARKI A, C. R., ESKO JD, ET AL. (ed.) *Essentials of Glycobiology*, 2, Chapter 14. New York: Cold Spring Harbor Laboratory Press. 13: 9780879697709

VARKI, A. & SHARON, N. 2009. Historical Background and Overview. *In:* VARKI A, C. R., ESKO JD, ET AL. (ed.) *Essentials of Glycobiology*, 2, Chapter 1. New York: Cold Spring Harbor Laboratory Press. 13: 9780879697709

VIGERUST, D. J., ULETT, K. B., BOYD, K. L., MADSEN, J., HAWGOOD, S. & MCCULLERS, J. A. 2007. N-linked glycosylation attenuates H3N2 influenza viruses. *Journal of virology,* 81**,** 8593-8600.

VUKOVIC, P., CHEN, K., QIN LIU, X., FOLEY, M., BOYD, A., KASLOW, D. & GOOD, M. F. 2002. Single-chain antibodies produced by phage display against the C-terminal 19 kDa region of merozoite surface protein-1 of Plasmodium yoelii reduce parasite growth following challenge. *Vaccine,* 20**,** 2826-2835.

WADA, Y., AZADI, P., COSTELLO, C. E., DELL, A., DWEK, R. A., GEYER, H., GEYER, R., KAKEHI, K., KARLSSON, N. G. & KATO, K. 2007. Comparison of the methods for profiling glycoprotein glycans-HUPO Human Disease Glycomics/Proteome Initiative multi-institutional study. *Glycobiology,* 17**,** 411-422.

WALSH, C. T., GARNEAU-TSODIKOVA, S. & GATTO, G. J. 2005. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angewandte Chemie International Edition,* 44**,** 7342-7372.

WALSH, G. & JEFFERIS, R. 2006. Post-translational modifications in the context of therapeutic proteins. *Nature biotechnology,* 24**,** 1241-1252.

WANG, B. 2009. Sialic acid is an essential nutrient for brain development and cognition. *Annual review of nutrition,* 29**,** 177-222.

WANG, C.-C., CHEN, J.-R., TSENG, Y.-C., HSU, C.-H., HUNG, Y.-F., CHEN, S.-W., CHEN, C.-M., KHOO, K.-H., CHENG, T.-J. & CHENG, Y.-S. E. 2009. Glycans on influenza hemagglutinin affect receptor binding and immune response. *Proceedings of the National Academy of Sciences,* 106**,** 18137-18142.

WARREN, L., BUCK, C. A. & TUSZYNSKI, G. P. 1978. Glycopeptide changes and malignant transformation a possible role for cabohydrate in malignant behavior. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer,* 516**,** 97-127.

WEIS, W. I. & DRICKAMER, K. 1996. Structural basis of lectin-carbohydrate recognition. *Annual review of biochemistry,* 65**,** 441-473.

WELLS, L., VOSSELLER, K. & HART, G. W. 2001. Glycosylation of nucleocytoplasmic proteins: signal transduction and O-GlcNAc. *Science,* 291**,** 2376-2378.

WERLE, M. & BERNKOP-SCHNURCH, A. 2006. Strategies to improve plasma half life time of peptide and protein drugs. *Amino acids,* 30**,** 351-367.

WILSON, D. S. & SZOSTAK, J. W. 1999. In vitro selection of functional nucleic acids. *Annual review of biochemistry,* 68**,** 611-647.

WINTER, H. C., MOSTAFAPOUR, K. & GOLDSTEIN, I. J. 2002. The mushroom Marasmius oreades lectin is a blood group type B agglutinin that recognizes the Gal-α-(1, 3)-Gal and Gal-α-(1, 3)-Gal –β-(1, 4)-GlcNAc porcine xenotransplantation epitopes with high affinity. *Journal of biological chemistry,* 277**,** 14996-15001.

WOJCHOWSKI, D. M., PARSONS, P., NORDIN, J. H. & KUNKEL, J. G. 1986. Processing of pro-vitellogenin in insect fat body: a role for high-mannose oligosaccharide. *Developmental biology,* 116**,** 422-430.

http://WWW.FUNCTIONALGLYCOMICS.ORG. 2014. [Accessed 25 November].

YAMAGUCHI, Y., NAKATA, M., SAKAI, K., SHIMIZU, Y., TAKASAKI, A., CHIBA, T., KUSADA, Y., SHIMIZU, N. & TAKAYANAGI, A. 2009. Phage-displaying single-chain antibody capable of recognizing non-reduced mannose residue. Google Patents. WO 2007132917 A1.

YAMAMOTO, K., ITO, S., YASUKAWA, F., KONAMI, Y. & MATSUMOTO, N. 2005. Measurement of the carbohydrate-binding specificity of lectins by a multiplexed bead-based flow cytometric assay. *Analytical biochemistry,* 336**,** 28-38.

YANG, Q., GOLDSTEIN, I. J., MEI, H.-Y. & ENGELKE, D. R. 1998. DNA ligands that bind tightly and selectively to cellobiose. *Proceedings of the National Academy of Sciences,* 95**,** 5462-5467.

YAO, J. K. & RASTETTER, G. M. 1985. Microanalysis of complex tissue lipids by high-performance thin-layer chromatography. *Analytical biochemistry,* 150**,** 111-116.

YOSHIDA, S., KOBAYASHI, T., MATSUOKA, H., SEKI, C., GOSNELL, W. L., CHANG, S. P. & ISHII, A. 2003. T-cell activation and cytokine production via a bispecific single-chain antibody fragment targeted to blood-stage malaria parasites. *Blood,* 101**,** 2300-2306.

ZHAO, Y. Ä., TAKAHASHI, M., GU, J. Ä., MIYOSHI, E., MATSUMOTO, A., KITAZUME, S. & TANIGUCHI, N. 2008. Functional roles of N-glycans in cell signaling and cell adhesion in cancer. *Cancer science,* 99**,** 1304-1310.

ZHOU, S.-M., CHENG, L., GUO, S.-J., ZHU, H. & TAO, S.-C. 2011. Lectin microarrays: a powerful tool for glycan-based biomarker discovery. *Combinatorial chemistry & high throughput screening,* 14**,** 711-719.

ZLATKIS, A. & KAISER, R. E. 2011. *HPTLC-high performance thin-layer chromatography,* Elsevier.

ZUBER, C., LACKIE, P. M., CATTERALL, W. & ROTH, J. 1992. Polysialic acid is associated with sodium channels and the neural cell adhesion molecule N-CAM in adult rat brain. *Journal of biological chemistry,* 267**,** 9965-9971.

Chapter-2: Structural analyses of Gal-α-(1, 3)-Gal-scFv interactions

2.1. Introduction

The non-human carbohydrate Galactosyl-α-(1, 3)-Galactose (Gal-α-(1, 3)-Gal) is immunogenic in humans and is of major clinical significance (Macher and Galili, 2008). Antibodies against this epitope are naturally present in human serum (Galili et al., 1985). This epitope is directly involved in hyperacute rejection during xenotransplantation (Joziasse and Oriol, 1999, Galili, 2005) and is present on a number of biopharmaceuticals produced in Chinese Hamster Ovary (CHO) and the mouse SP2/0 and NS0 cell lines. Biopharmaceutical products with this epitope can result in undesirable immunogenic response in patients, as seen in the case of Cetuximab, a recombinant monoclonal antibody (mAb) used in the treatment of metastatic colorectal cancer (Pointreau et al., 2012, Bosques et al., 2010). The presence of the Gal-α-(1, 3)-Gal epitope in beef and pork (red meat) causes delayed anaphylaxis in humans (Commins and Platts-Mills, 2013). The presence of anti-Gal-α-(1, 3)-Gal IgEs causing anaphylaxis was shown in the serum of patients treated with Cetuximab (Chung et al., 2008). In addition, the Gal-α-(1, 3)-Gal motif present on the surface of some pathogenic organisms appears to play an important role in host-pathogen interactions as is evident from the protozoan parasites *Trypanosoma cruzi* and *Leishmania* species, which cause Chagas disease and cutaneous leishmaniasis, respectively (Towbin et al., 1987, Avila et al., 1989), and the fungus, *Paracoccidioides brasiliensis*, responsible for paracoccidioidomycosis disease. (Vallejo et al., 2011).

The occurrence of natural anti-Gal-α-(1, 3)-Gal serum antibodies has also been exploited in cancer immunotherapy. Modification of autologous tumour cells to express the epitope on their surface and subsequent immunisation of the patient with the cells promotes up-take of the vaccine by the antigen-presenting cells through *in vivo* opsonisation by the anti-Gal-α-(1, 3)-Gal antibodies to give a protective anti-tumour response (Galili et al., 2003).

Despite the importance of the Gal-α-(1, 3)-Gal epitope, relatively few methods have been described in the literature for its determination (Higgins, 2010). The presence of Gal-α-(1, 3)-Gal can be detected by analysis of enzymatically released Gal monosaccharide or altered oligosaccharide using HPLC and capillary electrophoresis with laser-induced fluorescence (CE-LIF) detection methods (Maeda et al., 2012,

Yagi et al., 2012). However, these methods require release of the oligosaccharides or Gal monosaccharide using specific enzymes followed by purification and labeling of the carbohydrate prior to analysis.

Lectins are non-enzymatic proteins of non-immune origin, which bind specifically to certain carbohydrate motifs. Enzyme-linked lectin assays (ELLA) have been employed to detect intact Gal-α-(1, 3)-Gal present on glycoproteins. The lectins most frequently used are the B4 isolectin of *Griffonia simplicifolia* I (GS-I-B4), which binds to any α-linked terminal Gal regardless of linkage position (McCoy Jr et al., 1983, Goldstein and Winter, 1999), and *Marasmius oreades* agglutinin (MOA) and *Euonymus europaeus* agglutinin (EEA), both of which also have broad binding specificities with α-Gal terminal structures (Teneberg et al., 2003).

The generation of high-quality antibodies targeting specific carbohydrate motifs presents a significant challenge because of the generally low immunogenicity of carbohydrates. Natural human IgG and IgM antibodies against Gal-α-(1, 3)-Gal have been reported (Galili et al., 1984, Sandrin et al., 1993) and monoclonal antibodies (mAbs) (IgM, IgG1 and IgG3) were produced following immunisation of *α*-(1, 3)-galactosyltransferase gene knockout mice with rabbit red blood cells (Nozawa et al., 2001, Milland et al., 2007). A chicken polyclonal IgY (Bouhours et al., 1998) and a mouse mAb IgM (M86) (Galili et al., 1998) with specificity for Gal-α-(1, 3)-Gal have also been produced. The M86 antibody and GS-I-B4 have been used to detect and quantify Gal-α-(1, 3)-Gal in soft tissues and cells (Naso et al., 2011, Naso et al., 2012, Z Konakci et al., 2005, Galili et al., 1998) . However, no convenient enzyme-linked immunosorbent assay (ELISA) for biopharmaceutical products has been described to date using any of these antibodies. The ELISA format is a widely used diagnostic and quantitative tool in clinical and industrial laboratories.

The antigen-binding (Galili et al., 1998) domain of an antibody can be expressed in phage as a single chain molecule in which the light (V_L) and heavy chain (V_H) domains of the antibody are joined by a flexible polypeptide linker to form a single chain variable fragment (scFv) (Griffiths et al., 1993). scFvs have a number of advantages over traditional antibodies as they are easier and cheaper to recombinantly produce and purify and have higher affinities, usually nM (Charlton et al., 2001, Rahbarizadeh et al., 2004) with few scFvs displaying affinities in pM and

fM (Boder et al., 2000, Hanes et al., 2000) and are used as a diagnostic tools (Hagemeyer et al., 2009). Due to their reduced sise compared to monoclonal antibodies, this makes them more suitable as chemotherapeutic agents in tumour treatment (Lyu et al., 2008), for use as parasite therapeutics (Vukovic et al., 2002, Yoshida et al., 2003) and possibly as tumour vaccines (Ladjemi et al., 2011). Recently, single-chain antibody fragments (scFvs) from an immunised chicken-derived phage displayed scFv library have been described that show high specificity for Gal-α-(1, 3)-Gal (Cunningham et al., 2012). Two of the three characterised scFvs demonstrated exclusive and specific binding to the Gal-α-(1, 3)-Gal structure, and the scFv detection limit in competitive assay was estimated at 3.9 ng/mL, which is equal to 10 nM. This is higher than the detection limits of many lectins, which is mainly in the μg/mL range in similar binding assays (Tateno et al., 2007). These scFvs, as used in the competitive assay, will provide a convenient method for screening of recombinant therapeutics, with a benefit for pharmaceutical industry and increased safety of pateints (Cunningham et al., 2012).

Carbohydrate-specific antibodies tend to exhibit lower affinities for their targets usually in μM, and antibodies are most difficult to develop against carbohydrates due to the independent response of T-cells against carbohydrates (Heimburg-Molinaro and Rittenhouse-Olson, 2009). Antibodies that bind proteins and other small molecules have higher binding affinities, in the nM and pM range. Most lectin-glycan interaction also show lower binding affinities than protein binding antibodies; generally in the order of k_D 10^{-3} to 10^{-7} M (Hirabayashi, 2008). To assist in engineering of improved affinity and specificity of carbohydrate recognition molecules in the future, detailed examination of the interaction between carbohydrates and known binding molecules is needed. Crystal structures of α-Gal terminating glycans in complex with various lectins have been reported widely and reviewed (Yuriev et al., 2009). Two different modes of lectin-carbohydrate interaction were identified - groove type binding as occurs with MOA (Grahn et al., 2009), and end-on insertion binding as described for GS-I-B4 (Tempel et al., 2002). No X-ray crystal structures have been described for anti-Gal-α-(1, 3)-Gal antibodies in complex with the target, but a mouse scFv in complex with a trisaccharide containing α–Gal-(α-D-Gal-(1, 2)-[α-D-Abe(l, 3)]-α-D-Manp1-OMe, where Abe is Abequose (the immuno-dominant sugar of the Salmonella O antigen) has been

reported (PDB ID- 1MFA) (Zdanov et al., 1994). This Abequose was buried deep in the scFv pocket, whereas Gal and Man contacted the protein surface of the scFv. In addition, a saturation-transfer difference (STD) nuclear magnetic resonance (NMR) study of human and mouse chimeric anti-Gal-α-(1, 3)-Gal IgE and affinity-purified human serum polyclonal anti-Gal-α-(1, 3)-Gal antibodies interaction with Gal-α-(1, 3)-Gal-OMe has provided insight into the structural recognition (Plum et al., 2011). The low number of crystal structures reported for carbohydrate-recognition molecules are due to the difficulty of crystallisation of a very polar complex in native conditions and has also been attributed to the weak interaction between proteins and carbohydrate (DeMarco and Woods, 2008).

Thus, *in silico* molecular modelling is being increasingly used to study antibody-carbohydrate interactions (Kearns-Jonker et al., 2007, Agostino et al., 2009b). Homology models of the binding sites of natural human and a panel of mouse anti-Gal-α-(1, 3)-Gal epitope mAbs were described based on the available mouse scFv-trisaccharide crystal structure 1MFA. The human anti-Gal-α-(1, 3)-Gal immunoglobulin has been modelled in complex with the trisaccharide antigen of the Gal-α-(1, 3)-Gal epitope where important amino acid contact residues Gal-α-(1, 3)-Gal epitope were identified and confirmed by mutation studies (Kearns-Jonker et al., 2007). Similarities in binding site structures of their *in silico* modelled anti-Gal-α-(1, 3)-Gal antibody with the crystal structure 1MFA were also highlighted. The mouse mAb study was expanded from the protein and carbohydrate perspective (Agostino et al., 2009b) and site mapping of the antibody-carbohydrate complexes was used to examine van der Waals and hydrogen bonding which highlighted the role of the terminal α-linked Gal in xenoreactive antibody recognition (Agostino et al., 2009b, Agostino et al., 2010).

In this work, binding affinities of all three scFvs with the Gal-α-(1, 3)-Gal disaccharide were determined using surface plasmon resonance (SPR) and *in silico* approaches were used to determine key sites (active sites) on the scFvs and the target structures that were involved in recognition and scFv-glycan interaction. This *in silico* modelling approach was used to explore the molecular basis for the high specificity and affinity characteristics of the highly specific and selective anti-Gal-α-(1, 3)-Gal scFvs (Cunningham et al., 2012). The amino acids involved in the active

site maps obtained for the anti-Gal-α-(1, 3)-Gal scFvs would provide providing structural insights into Gal-α-(1, 3)-Gal recognition by the scFvs and identified a pool of amino acids crucial for Gal-α-(1, 3)-Gal recognition.

2.2. Materials and Methods

2.2.1. Materials

Biacore X100 instrument from GE Healthcare Bio-Sciences AB, Uppsala, Sweden was used. CM 5 Dextran sensor chip was obtained from GE Healthcare Bio-Sciences AB, Uppsala, Sweden. 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) / *N*-Hydroxysuccinimide (NHS), 1M Ehanolamine-HCl, 50 mM NAOH all were purchased from GE Healthcare Bio-Sciences AB, Uppsala, Sweden. Gal-α-(1, 3)-Gal-BSA was purchased from Dextra, UK. Eppendorf tubes and plastic vials were purchased from GE Healthcare Bio-Sciences AB, Uppsala, Sweden. Bovine Serum Albumin (BSA) was purchased from Sigma, UK. 10 mM sodium acetate pH 4.0, 10 mM sodium acetate pH 5.0, 10 mM 2-(*N*-morpholino) ethanesulfonic acid (MES) buffer pH 6.0, Phosphate Buffer Saline (PBS) pH 7.4 and 50 mM Sodium carbonate buffer pH 9.6 are used for pH scouting of immbolised ligand. M86 mouse monoclonal antibody (mAb) was purchased from Enzo Lifesciences, UK. Bradford-Protein Assay Kit for scFv quantification was purchased from Pierce Biotechnology, Inc. (Thermo Scientifc, UK). HP Pro computer terminal with - AMD Athlon II X3 440 Processor 3.00 Ghz, Installed RAM- 3.00 GB, 64-bit operating system was purchased from HP, Ireland. Molecular operating environment (MOE v. 2010.10) was licensed from Chemical Computing Group, Canada.

2.2.2. ScFv preparation and quantification

The three anti-Gal-α-(1, 3)-Gal scFvs A4, G12 and A11 were prepared as previously described (Cunningham et al., 2012). The scFv concentrations were quantified using a protein quantification method (Bradford, 1976). 5µL of each unknown scFv sample was pipetted into the appropriate microplate wells, then 250µL of the Coomassie reagent was added to each well and mixed for 30 s. After incubating the plate for 10 minutes at room temperature (RT), absorbance of each well was measured at 595nm. The blank replicates average was subtracted from the scFv sample average and the standard curve were plotted for the BSA standard vs. scFv sample to calculate scFv concentrations.

2.2.3. Surface Plasmon Resonance (SPR) analysis

The interaction affinity of both scFvs' and M86 monoclonal antibody were determined using a Biacore X100 SPR instrument. The temperature of the SPR machine was kept constant at 37 °C throughout the experiment. To determine the optimal pH of the ligand Gal-α-(1, 3)-Gal-BSA for immobilisation on to sensor chip, pH scouting of buffers was done, where five different buffers, of 55 µL each of : 10 mM sodium acetate pH 4.0, 10 mM sodium acetate pH 5.0, 10 mM 2-(*N*-morpholino) ethanesulfonic acid (MES) buffer pH 6.0, Phosphate Buffer Saline (PBS) pH 7.4 and 50 mM Sodium carbonate buffer pH 9.6 were tested. Out of these buffers, 10 mM sodium acetate pH 4.0 had the highest response unit (Table 2.1) and was selected for ligand immobilisation.

Table 2.1: Immobilisation of ligand (Gal-α-(1, 3)-Gal-BSA) pH scouting result on Biacore SPR X 100.

Buffer	**pH**	**Ligand**	**Response Unit (RU)**
50 mM Sodium Carbonate (Na_2CO_3)	9.6	Gal-α-(1, 3)-Gal-BSA	-1902.5
Phosphate Buffer Saline (PBS)	7.4	Gal-α-(1, 3)-Gal-BSA	-15.9
10 mM MES	6	Gal-α-(1, 3)-Gal-BSA	-896.3
10 mM Sodium Acetate (NaOAc)	5	Gal-α-(1, 3)-Gal-BSA	6162.9
10 mM Sodium Acetate (NaOAc)	4	Gal-α-(1, 3)-Gal-BSA	14363.2

The ligand Gal-α-(1, 3)-Gal-BSA conjugate (Dextra, UK) was covalently coupled to a CM 5 dextran sensor chip (GE Healthcare Bio-Sciences AB, Sweden) which consists of carboxymethylated dextran covalently attached to a gold surface. Prior to immobilisation, a response unit (RU) of 1500 was set on each flowcell for maximum immobilisation of the ligand onto the chip surface. The coupling of the ligand on the CM5 chip was done using the standard EDC/NHS coupling procedure and capping by ethanolamine. The chip surface was initialised by injecting 30 µL of 50 mM

NaOH and then activated using 85 µL each of EDC and NHS. The ligand Gal-α-(1, 3)-Gal-BSA was immobilised in flow cell 2, while BSA was immobilised in flow cell 1 as reference. After immobilisation the ligands were blocked using 126 µL of 1 M 2-ethanolamine-HCl solution for 500 s to block the active carboxyl groups remaining on the sensor chip. For the kinetic analyses, the binding affinities were measured at 37 °C with a linear increase of 2X concentrations of the scFvs (A4, G12, and A11) and M86 mouse mAb prepared in binding buffer (PBS buffer) pH 7.4 from 0 to 400 nM, with a flow rate of 20 µL/min for 300 s used throughout. The association phase was monitored for 100 s and the dissociation phase for 300 s. Sensor surfaces were regenerated after each binding cycle by subsequent injections of 10 mM glycine-HCl, pH 2.5 for 60 sec. After each cycle, the sensor chip was washed with 50 µL of washing buffer (PBS pH 7.4) at a flowrate of 25 µL/min for 120 s. Binding responses were normalised by subtracting the response generated from identical injection conditions over a flow cell containing only bovine serum albumin (BSA) in PBS buffer pH 7.4 (flow cell 1), and double-referenced by subtracting the response from phosphate buffered saline (PBS), pH 7.4, injections. Data analysis and calculation for the dissociation constant at equilibrium K_D were performed using BIAevaluation software (v3.2, GE Healthcare Bio-Sciences AB, Sweden) and Scrubber2 (BioLogic Software, Australia). Binding assays for each scFv were done in triplicate.

2.2.4. ScFv sequence analyses

Similarity searches of all three scFv sequences were performed using the BLAST online program, where scFv nucleotide sequences homology search (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) and translated amino acid sequence homology (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) were searched using these web links. *Gallus gallus* database (Taxon Id- 9801) was searched within the BLAST N program. Giving priority to query cover of the scFv sequences, followed by identity and then expect (E) value, nucleotide similarity was searched within the result output file. The default E value is 10 for the BLAST program, however, the lower the E-value, or the closer it is to zero, the more "significant" the matching sequence is.

Multiple sequence alignment of scFvs A4, G12 and A11 was performed using ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) (Larkin et al., 2007) with default conditions in the program. Complementarity determining regions (CDRs) across both scFvs were identified in MOE (2010.10) software (Chemical Computing Group, Montreal, Canada) and confirmed later by locating flanking conserved regions as listed on http://www.bioinf.org.uk/abs.

2.2.5. Molecular modelling and molecular dynamics simulations of the scFvs

ScFvs A4 and G12 were modelled using MOE Antibody modelling suite (MOE 2010.10), (Chemical Computing Group, Montreal, Canada). Templates were searched based on their structural similarity score in the Protein Data Bank (PDB) for the V_L and V_H chains of scFvs A4, G12 and A11 (Berman et al., 2000). To model the framework and CDRs of scFvs A4 and G12, multiple templates were used for assigning loops in both V_L and V_H regions. For the framework and CDR template search, BLOSUM62 substitution matrix (default in MOE) was used for alignment (gap start penalty set at 7; gap extend penalty set at 1). A threading approach was used with multiple templates searched for both framework and CDRs (Contreras-Moreira et al., 2003, Majumdar et al., 2011). In the threading approach, templates were searched and selected based on the sequence similarity to the scFvs. The obtained multiple templates were then aligned to the individual scFv aminoacids and the core of the query (scFv) structure was constructed based on the alignment. After construction of the core model, non-conserved loops were built connecting the secondary structure elements (SSE) and finally the model was refined using energy minimisation method. The cut-off structure score as calculated in MOE for framework and CDR was kept at default 45 (maximum 100). The final templates used to predict the homology models had PDB with PDB IDs as follows: for scFv A4; 3GCA, 3GCD (Teplyakov et al., 2009) and 1Q1J (Stanfield et al., 2004) and for G12; 8FAB (Strong et al., 1991) and 3BN9 (Farady et al., 2008). scFv models were protonated at pH 7, 310K and 100 mM Na^{++} concentration in the Generalised Born electrostatic model prior to energy minimisation, and the final models were minimised at an Root-Mean-Square-Deviation (RMSD) gradient of 0.5 kcal/mol Å to achieve stable models (Labute, 2008).

The Merck Molecular Force Field (MMFF94X) was employed to generate the models, and Generalised Born / Volume Integral (GB/VI) scoring was used to score all generated models (Halgren, 1996). The highest scored models were tested using Ramachandran plots, the outliers were constrained and the models were further energy minimised. Molecular dynamics simulations were performed for both scFvs using an NPT statistical ensemble, maintaining temperature at 310K throughout the simulations. The system was initially heated at 310K for 100 ps followed by 1,000 ps simulation in the presence of water molecules to check the stability of the generated scFv models. All bonds were constrained and water molecules were considered as rigid bodies during the simulations.

2.2.6. Carbohydrate modelling and docking

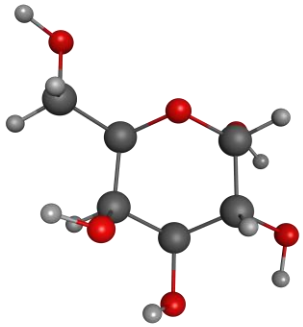
Seven different monosaccharide, disaccharide and a trisaccharide structures (Gal-α-(1, 3)-Gal, Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc, α-D-Gal, α-L-Fuc, β-D-Xyl, Gal-α-(1, 2)-Gal and Gal-β-(1, 4)-Gal) were selected (Figure 2.1) using the MOE carbohydrate builder suite and the GLYCAM carbohydrate builder (http://www.glycam.com). These structures were energy minimised to attain stable conformations (Figure 2.1). Carbohydrates were docked flexibly into the CDRs and the active sites of the scFvs which were predicted using the MOE Site Finder algorithm (MOE v. 2010.10) (Del Carpio et al., 1993, Hendlich et al., 1997). Triangle matcher placement methodology was applied to generate poses by aligning ligand triplets of atoms on triplets of alpha spheres in a systematic manner (Rarey et al., 1996). The different conformations were refined for carbohydrate-scFv interactions using the MM/GBVI algorithm (Wojciechowski and Lesyng, 2004). The final scoring for all conformations were calculated using the London dG scoring function (Muegge and Martin, 1999) and the free energy of binding of the ligand for a given pose was estimated as implemented in MOE (MOE, 2010.10).

The functional form is a sum of terms:

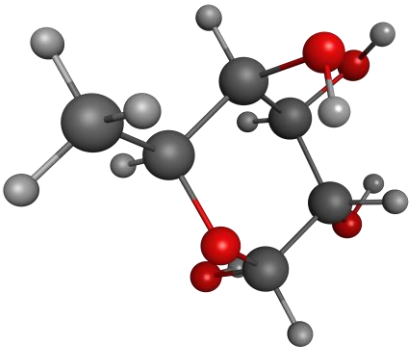
$$\Delta G = c + E_{flex} + \sum_{h-bonds} c_{HB} f_{HB} + \sum_{m-lig} c_M f_M + \sum_{atoms\ i} \Delta D_i$$

where c represents the average gain/loss of rotational and translational entropy; E_{flex} is the energy due to the loss of flexibility of the ligand (calculated from ligand topology only); f_{HB} measures geometric imperfections of hydrogen bonds and

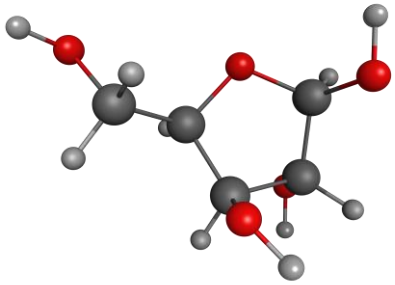
takes a value in [0,1]; c_{HB} is the energy of an ideal hydrogen bond; f_M measures geometric imperfections of metal ligations and takes a value in [0,1]; c_M is the energy of an ideal metal ligation; and D_i is the desolvation energy of atom i.



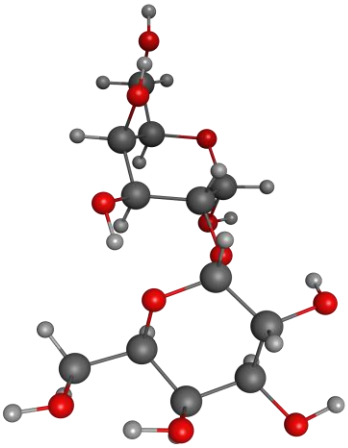
A



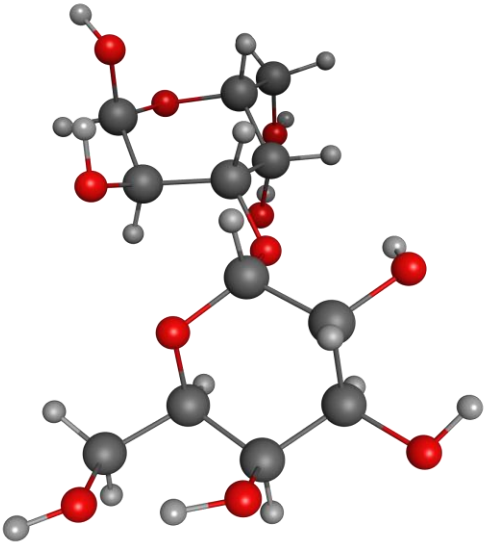
B



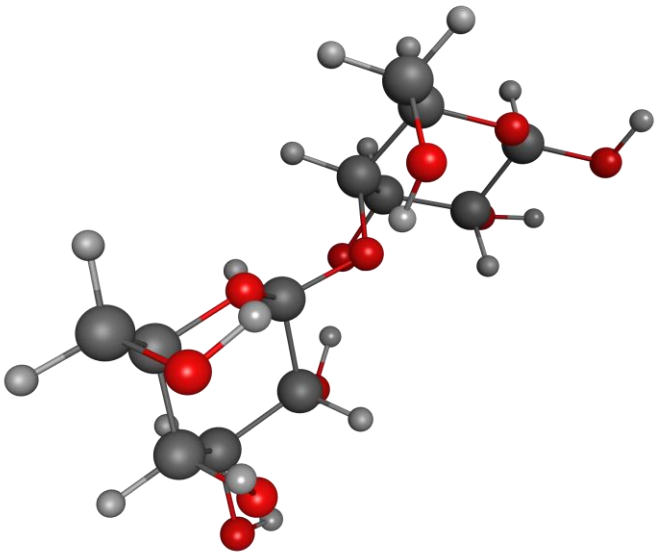
C



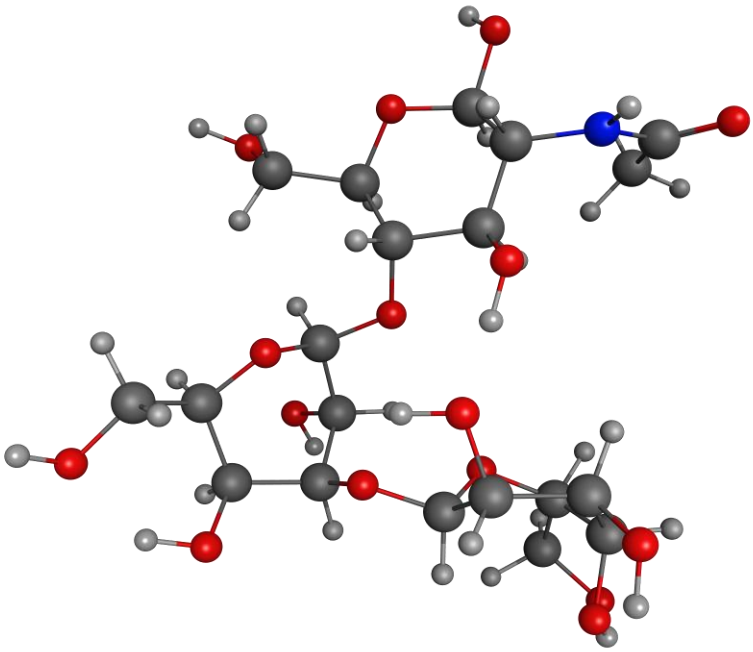
D



E



F



G

Figure 2.1: Glycan structures, which were drawn using Carbohydrate Builder of MOE. These structures were further energy minimised to attain stable conformation. (A) α-D-Galactose, (B) α-L-Fucose, (C) β-D-Xylose, (D) Gal-α-(1, 2)-Gal, (E) Gal-α-(1, 3)-Gal, (F) Gal-β-(1, 4)-Gal (G) Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc

2.2.7. Carbohydrate-scFv interaction analyses

The top 30 scoring poses for each glycan docking against scFv were considered for binding evaluation. Individual poses of scFvs docked with carbohydrates were visualised in the LigX module of MOE for interpretation of the scFv-carbohydrate interaction. Amino acid residues within predicted active sites at a distance of 4.5 Å from the ligand were displayed as visible. Interactions of scFv amino acid residues with the ligand (contact points) were analysed on the basis of the different types of interactions which may contribute to overall binding; side chain hydrogen bonds (donor or acceptor), backbone hydrogen bonds (donor or acceptor), ionic interactions and surface interactions. The van der Waals and electrostatic interactions were considered at a distance of < 4.5 Å from the carbohydrate ligand. The most potent of each of these interactions based on the lower contact distance and energy in each category, if present, were considered.

2.3. Results

2.3.1. ScFv and antibody interactions against Gal-α-(1, 3)-Gal analysed by SPR

As only ED_{50} values were available from previous ELISA assays (Cunningham et al., 2012) but K_D (binding affinity) values were required to validate their binding and verify the accuracy of *in silico* binding affinities. The binding of anti- Gal-α-(1, 3)-Gal scFvs, A4, G12 and A11 and the commercial mAb M86 to immobilised Gal-α-(1, 3)-Gal was examined on SPR. The initial concentration of scFv measured through Bradford assay for scFv A4, G12 and A11 were 110, 150 and 126 μg/mL, respectively. Therfore the calculated molar concentrations of scFv A4 (28.43 kDa), G12 (28.55 kDa) and A11 (28.9 kDa) were 3.87, 5.25, and 4.36 μM, respectively. The resulting kinetic data with detailed SPR kinetic association and dissociation rates of scFvs (0 – 400 nM) against Gal-α-(1, 3)-Gal at 37 °C are shown in (Table 2.2 and Figure 2.2). The binding affinities for scFv-G12 and scFv-A4 at 37^{o}C were in the range of 10^{-8} M. scFv-G12 binds against Gal-α-(1, 3)-Gal with affinities $2.27*10^{-8}$ M – $1.21*10^{-8}$ M (SE < 5%), and scFv-A4 having an affinity range $1.11*10^{-7}$ M – $6.5*10^{-8}$ M (SE < 5%) (Table 2.1). Both scFv-G12 and scFv-A4 binding affinities are in accordance to ELISA results against Gal-α-(1, 3)-Gal reported previously (Cunningham et al., 2012). scFv A11 gave higher variability across replicates resulting in the failure to determine the binding affinity, due to the standard errors of kinetic rates exceeding threshold of acceptance. Previously it was reported that scFv A11 gave a less sensitive standard curve (ED_{50} = 1225 μ g/mL, 3.58 mM) than scFv G12 and A4 (0.46 and 0.91 mM), respectively. Using SPR, M86 antibody failed to show binding against Gal-α-(1, 3)-Gal within the concentration range of 0- 400 nM as no binding curve could be generated before using M86 antibody with the same concentration range as that of scFV A4, G12 and A11 (Cunningham et al., 2012).

The affinity of A4 against Gal-α-(1, 3)-Gal target was approximately four times lower than G12 (mean K_D 9.1 x 10^{-8} M and 1.8 x 10^{-8} M for scFvs A4 and G12, respectively). The difference in affinities of the scFvs for Gal-α-(1, 3)-Gal was due to a difference in their association rates, with G12 showing four times higher association constants (K_a) than A4. The dissociation constants (off rate, K_d) for both scFvs were very similar. These differences were consistent with the lower ED_{50} for the Gal-α-(1, 3)-Gal disaccharide given by G12 in competitive ELISA compared to

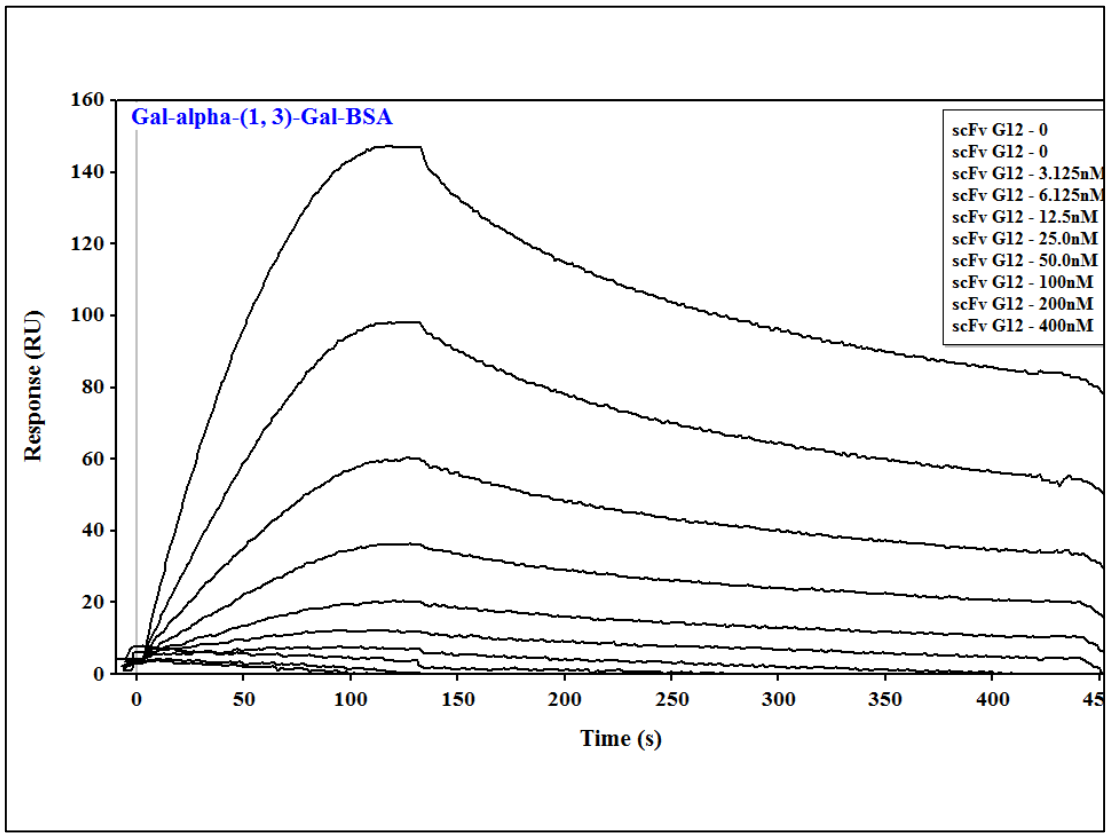
scFv A4 (Cunningham et al., 2012). The kinetic rate trends were consistent and statistically significant for both scFv A4 and G12 demonstrating the reproducibility of the scFvs on the SPR platform.

Table 2.2: Kinetic rates of scFv association and dissociation against with Gal-α-(1, 3)-Gal-BSA determined by SPR analysis.

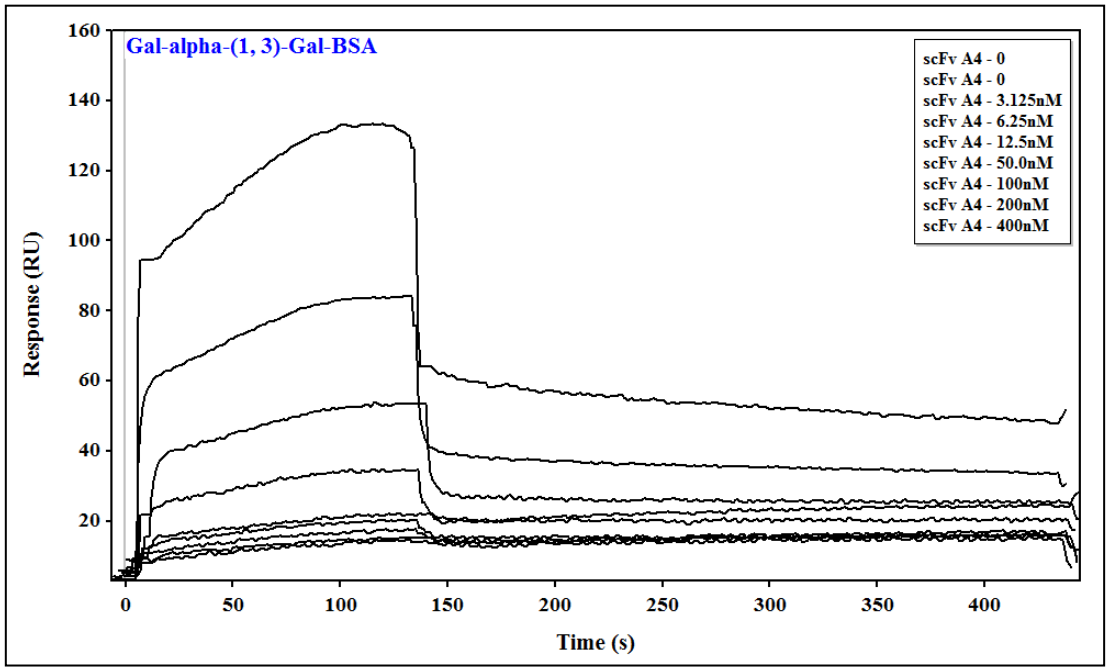
	K_a ($M^{-1}s^{-1}$)	**SE (K_a)**	**K_d (s^{-1})**	**SE (K_d)**	**K_D (M)**
scFv-G12	$1.16*10^5$	$1.6*10^3$	$1.4*10^{-3}$	$3.2*10^{-5}$	$1.21*10^{-8}$
	$8.67*10^4$	$3.3*10^3$	$1.9*10^{-3}$	$4.7*10^{-5}$	$2.27*10^{-8}$
	$9.44*10^4$	$1.4*10^3$	$1.8*10^{-3}$	$3.2*10^{-5}$	$1.92*10^{-8}$
scFv-A4	$3.07*10^4$	$3.5*10^2$	$2.0*10^{-3}$	$2.0*10^{-5}$	$6.51*10^{-8}$
	$1.21*10^4$	$2.5*10^2$	$1.3*10^{-3}$	$2.8*10^{-5}$	$1.11*10^{-7}$
	$3.01*10^4$	$7.8*10^2$	$2.9*10^{-3}$	$2.5*10^{-5}$	$9.6*10^{-8}$
scFv-A11	$3.07*10^5$	$9.4*10^3$	$4.7*10^{-3}$	$1.3*10^{-4}$	$1.54*10^{-8}$
	$3.58*10^5$	$5.8*10^5$	$3.1*10^{-3}$	$2.8*10^{-3}$	$8.72*10^{-9}$
	$3.39*10^7$	$8.6*10^6$	$3.38*10^{-4}$	$1.3*10^{-3}$	$1.14*10^{-11}$

K_a, association rate; K_d, dissociation rate; SE, standard error of K_a and K_d. SE (Standard Error) values less than 5% of their corresponding rates were deemed significant. K_D as binding affinity is calculated as the ratio of K_d/K_a. All these experiments were done in triplicate.

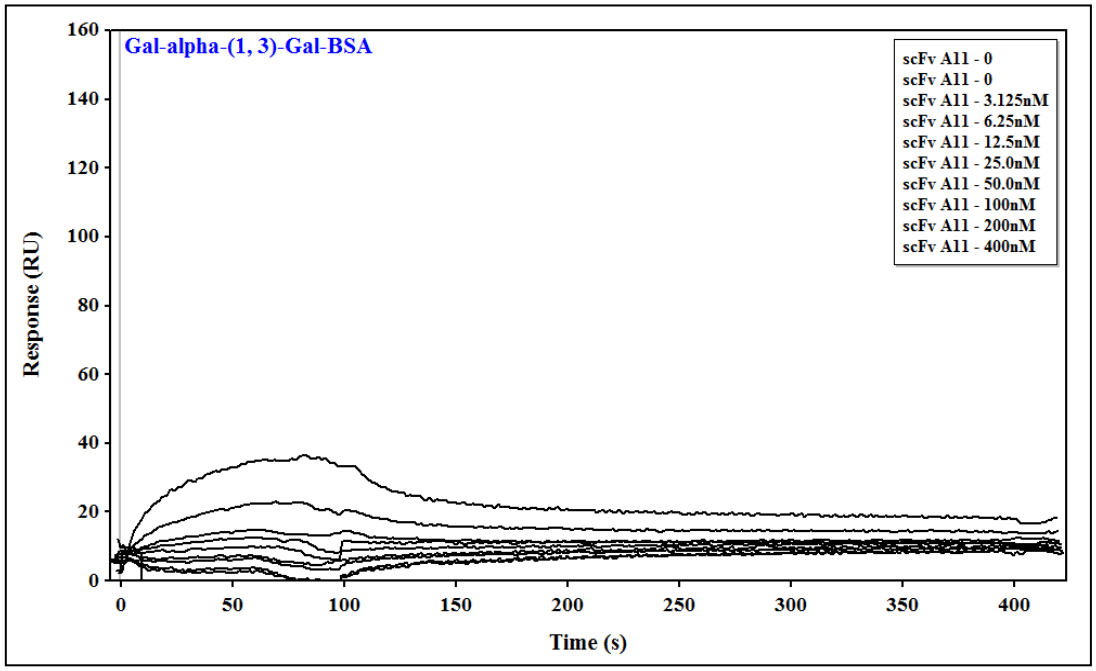
A


Gal-alpha-(1, 3)-Gal-BSA
scFv G12 - 0
scFv G12 - 0
scFv G12 - 3.125nM
scFv G12 - 6.125nM
scFv G12 - 12.5nM
scFv G12 - 25.0nM
scFv G12 - 50.0nM
scFv G12 - 100nM
scFv G12 - 200nM
scFv G12 - 400nM
Response (RU)
Time (s)
0
20
40
60
80
100
120
140
160
0
50
100
150
200
250
300
350
400

B


Gal-alpha-(1, 3)-Gal-BSA
scFv A4 - 0
scFv A4 - 0
scFv A4 - 3.125nM
scFv A4 - 6.25nM
scFv A4 - 12.5nM
scFv A4 - 50.0nM
scFv A4 - 100nM
scFv A4 - 200nM
scFv A4 - 400nM
Response (RU)
Time (s)
0
20
40
60
80
100
120
140
160
0
50
100
150
200
250
300
350
400

C



D

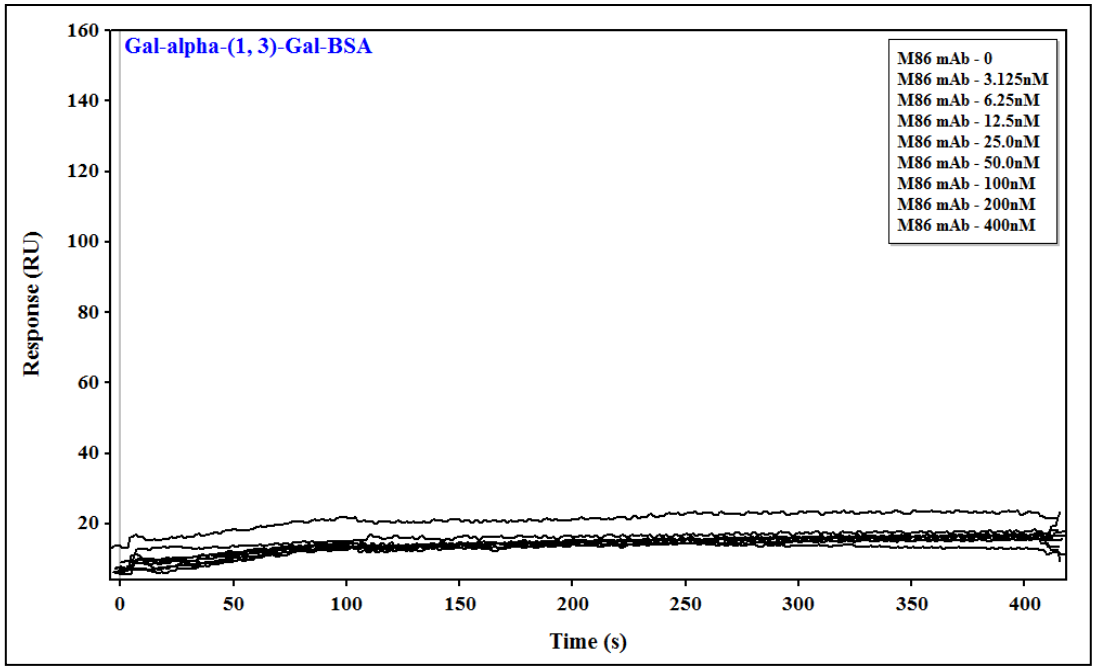


Figure 2.2: Binding curves of Gal-α-(1, 3)-Gal-BSA against (A) G12, (B) A4, (C) A11, & (D) M86 mouse mAb at a concentration range from 0 – 400 nM. Binding responses were normalised by subtracting the response generated from identical injection conditions over a flow cell containing only BSA (flow cell 1), and double-referenced by subtracting a response from PBS injections.

2.3.2. ScFv sequence comparison

The sequences of the V_L and V_H regions of scFvs A4 and G12 were aligned and a number of differences in the amino acid sequence across both the framework and the CDR regions were apparent (Figure 2.3). The CDRs are short amino acid sequences found in the variable regions of antibodies and are important in determining antibody specificities.

The length of CDR1 and CDR3 in V_L differed across the all three clones, with scFv-A11 having an extended CDR1, while scFv-G12 has an extended CDR3 (Figure 2.4). CDR2 in V_L contains 2 substitution sites, 1 of which is unique to scFv-A11 (Asn to Asp), while the second differs across all three scFv (Asn or Gln or Lys). Within the V_H, CDR1 was conserved across the scFv clones. CDR 2 of the V_H chain in scFv-G12 contained two substitutions (Asn to Tyr and an Arg to Ser), which are not shared with scFv-A4 and scFv-A11 where they are conserved as Tyr and Ser, respectively (Figure 2.4). The overall sequence alignment for scFv-A4:A11 was 93% identity, whereas sequence alignment for scFv-A4:G12 and scFv-A11:G12 were both having 89% identity to each other.

(A) scFv A4

L T Q P S S V S T N P G G T V K I T C S G G N G N Y G W Y Q Q K S P G S A P V T V I Y S N
N K R P S D I P S R F S G S K S G S T A T L T I T G V Q V D D E A V Y F C G A Y D N T Y V G
V F G A G T T L T V L **G Q S S R S S S G G G S S G G G G S** A V T L D E S G G G L Q T P G G
G L S L V C K A S G F T F S S Y S M Q W V R Q T P G K G L E F V A G I G Y S D S Y T Y F G
P A V K G R A T I S R D N G Q N T V R L Q L N N L R A E D T A T Y Y C A R S A D T I Y G C
T H P W C S A D N I D A W G H G T E V I V S S T S G Q A G Q

(B) scFv G12

L T Q P S S V S A N P G E T V K I T C S G G S Y H Y G W Y Q Q K S P G S A P V T V I Y S N
N Q R P S G I P S R F S G S T S D S T G T L T I T G V Q A D D E A V Y F C G S Y D S S N T Y
A G I F G A G T T L T V L **G Q S S R S S S G G G S S G G G G S** A V T L D E S G G G F Q T P
G G A L S L V C K A S G F T F S S Y S M Q W V R Q A P G K G L E F V A G I G N S D R Y T Y
F G P A V K G R A T I S R D N G Q S T L R L Q L N N L R A E D T A T Y F C A R S G D S G N
G C T H P W C S A D N I N A W G H G T E V I V S S T S G Q A G Q

(C) scFv A11

L T Q P S S V S A N P G E T V K I T C S G G G S Y G G S Y Y Y G W Y Q Q K S P G S A P V T
V I Y S N D N R P S D I P S R F S G S T S G S T S T L T I T G V Q V D D E A V Y Y C G T Y D
S S Y V G I F G A G T A L T V L **G Q S S R S S G G G S S G G G G S** A V T L D E S G G G L Q
T P G G G L S L V C K A S G F T F S S Y S M Q W V R Q T P G K G L E F V A G I G Y S D S Y
T Y F G P A V K G R A T I S R D N G Q N T V R L Q L N N L R A E D T A T Y Y C A R S A D T
I Y G C T H P W C S A D N I D A W G H G T E V I V S S T S G Q A G Q

V_L = variable region of light chain

red = **V_L** CDR1

purple = **V_L** CDR2

pink = **V_L** CDR3

V_H = variable region of heavy chain

blue = **V_H** CDR1

green = **V_H** CDR2

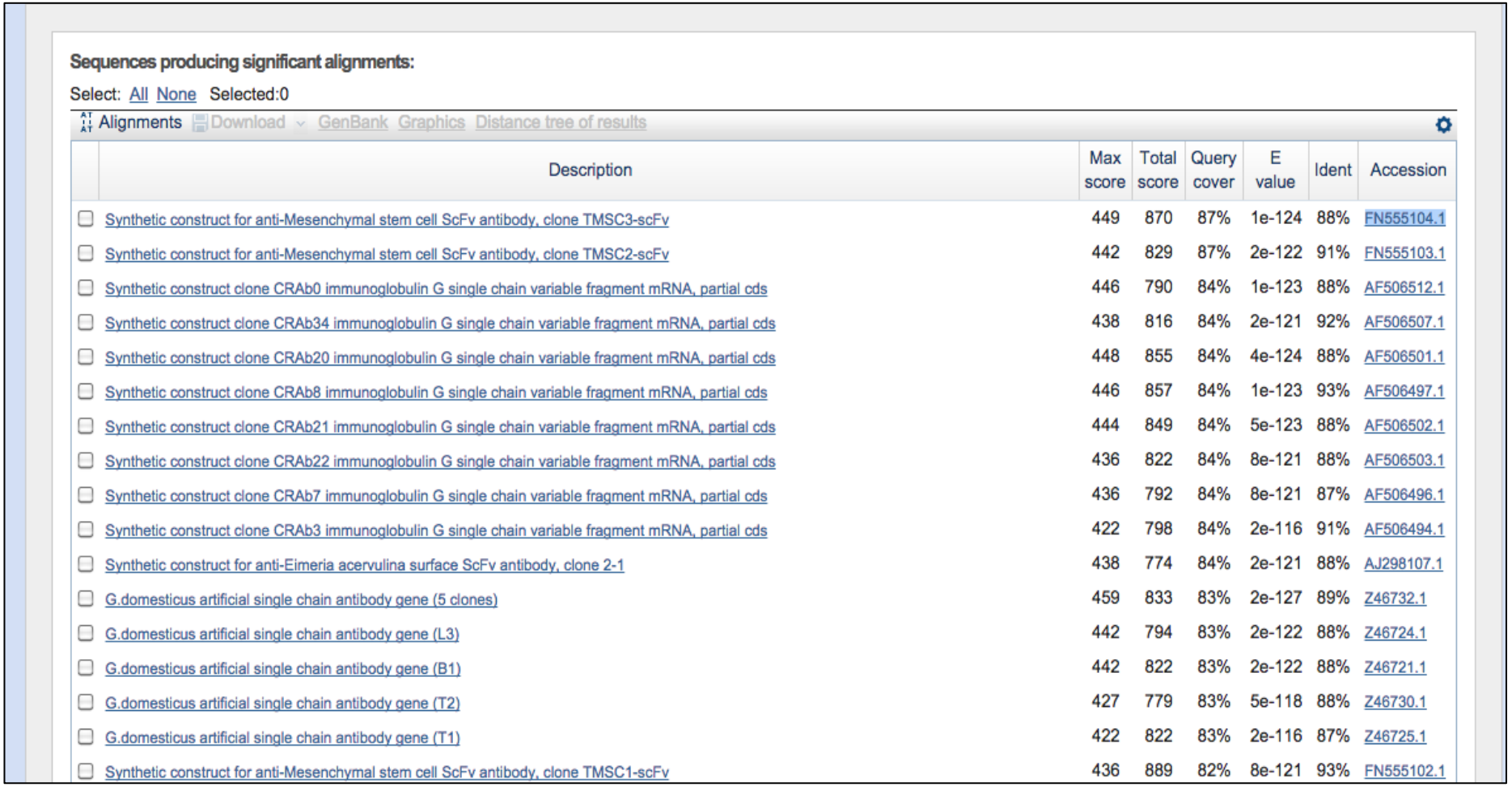
brown = **V_H** CDR3

Figure 2.3: Detailed translated scFv sequences. **(A)** scFv A4, **(B)** scFv G12 and **(C)** scFv A11. The sequence in bold black with underline is linker sequence. The light chain (**V_L**), and heavy chain (**V_H**) of scFvs are before and after linker sequence respectively. The sequences coloured in red, purple and pink are V_L: CDR1, CDR2, and CDR3 respectively, while the sequences coloured in blue, green and dark green are V_H: CDR1, CDR2, and CDR3 respectively.

Figure 2.4: Alignment of the amino acid translated scFv: A4, G12 and A11 sequences. The scFv full light and heavy chain sequences were aligned in CLUSTAL X software using default settings. Residues, AVFPMILW (Red) are Small (small+ hydrophobic (including aromatic)), DE (Blue) are acidic, RK (Magenta) are basic, STYHCNGQ (Green) are Hydroxyl + sulfhydryl + amine + G, and others (Grey) are unusual amino/imino acids. An * (asterisk) indicates positions which have a single, fully conserved residue. A : (colon) indicates conservation between groups of strongly similar properties. A . (period) indicates conservation between groups of weakly similar properties.

2.3.3. Molecular modelling of scFvs

The full nucleotide sequences of scFvs A4, G12 and A11 had 88 - 91% sequence identity with a query cover of 87% to chicken-derived scFvs by comparison with the chicken database of the NCBI using the BLAST-N program (Figure 2.5). However, no tertiary structural information has yet been reported for matched chicken scFv sequences. The scFv amino acid sequences were directly searched in PDB and the highest identity score hits reported were 2KH2, a solution NMR-derived structure of a mouse scFv-interleukin complex (Wilkinson et al., 2009), and 1H8N, an X-ray crystal structure of an mouse scFv-ampicillin complex (Jung et al., 2001), with E values of $1.04677 * 10^{-40}$ and $2.60291 * 10^{-35}$, respectively. Both 2KH2 and 1H8N were then used separately as templates for the generation of models for each of the scFvs using MOE antibody modeller suite criteria for the model generation, but the predicted structures failed the quality test (score < 45) for CDR modelling.



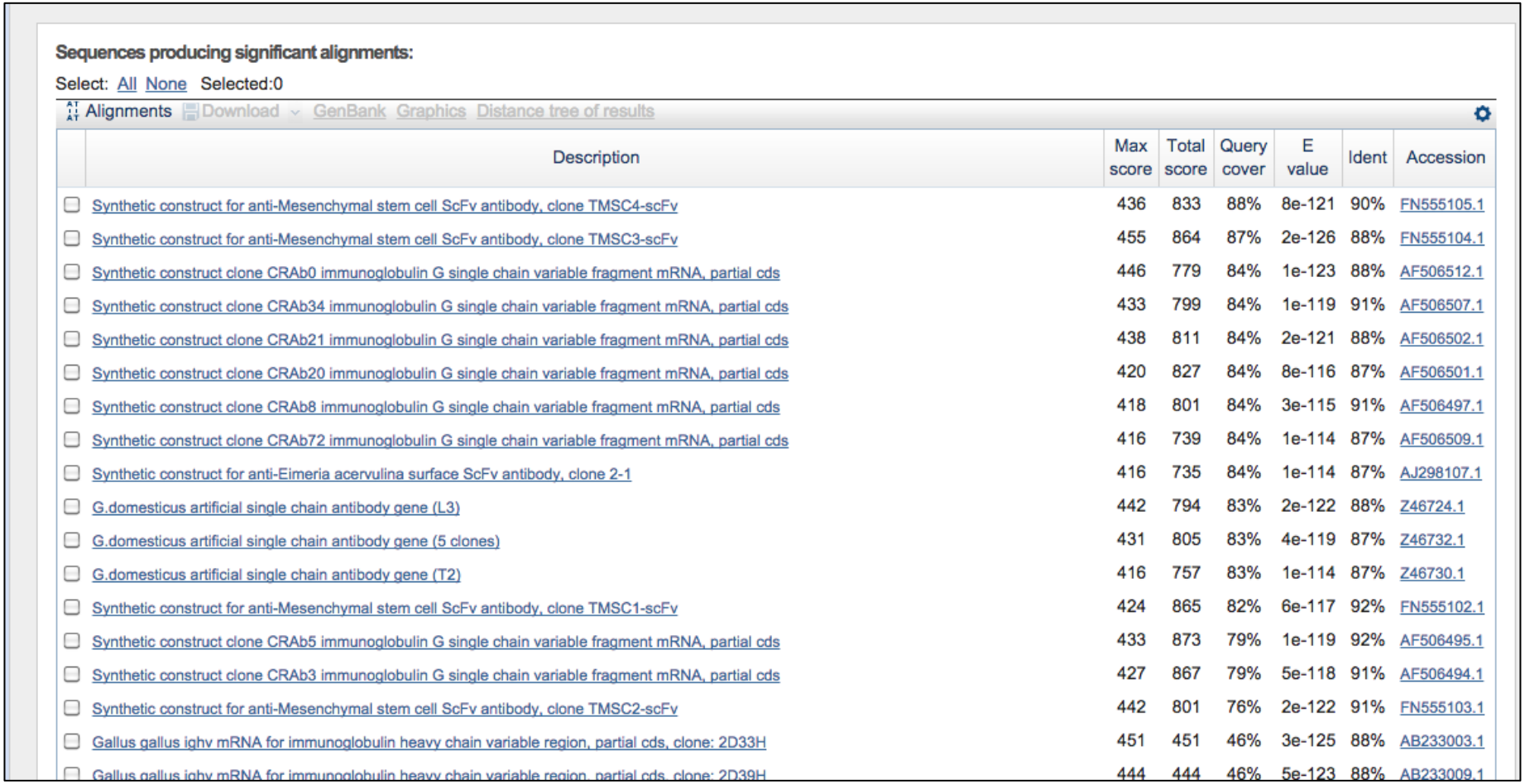
Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
Synthetic construct for anti-Mesenchymal stem cell ScFv antibody, clone TMSC3-scFv	449	870	87%	1e-124	88%	FN555104.1
Synthetic construct for anti-Mesenchymal stem cell ScFv antibody, clone TMSC2-scFv	442	829	87%	2e-122	91%	FN555103.1
Synthetic construct clone CRAb0 immunoglobulin G single chain variable fragment mRNA, partial cds	446	790	84%	1e-123	88%	AF506512.1
Synthetic construct clone CRAb34 immunoglobulin G single chain variable fragment mRNA, partial cds	438	816	84%	2e-121	92%	AF506507.1
Synthetic construct clone CRAb20 immunoglobulin G single chain variable fragment mRNA, partial cds	448	855	84%	4e-124	88%	AF506501.1
Synthetic construct clone CRAb8 immunoglobulin G single chain variable fragment mRNA, partial cds	446	857	84%	1e-123	93%	AF506497.1
Synthetic construct clone CRAb21 immunoglobulin G single chain variable fragment mRNA, partial cds	444	849	84%	5e-123	88%	AF506502.1
Synthetic construct clone CRAb22 immunoglobulin G single chain variable fragment mRNA, partial cds	436	822	84%	8e-121	88%	AF506503.1
Synthetic construct clone CRAb7 immunoglobulin G single chain variable fragment mRNA, partial cds	436	792	84%	8e-121	87%	AF506496.1
Synthetic construct clone CRAb3 immunoglobulin G single chain variable fragment mRNA, partial cds	422	798	84%	2e-116	91%	AF506494.1
Synthetic construct for anti-Eimeria acervulina surface ScFv antibody, clone 2-1	438	774	84%	2e-121	88%	AJ298107.1
G.domesticus artificial single chain antibody gene (5 clones)	459	833	83%	2e-127	89%	Z46732.1
G.domesticus artificial single chain antibody gene (L3)	442	794	83%	2e-122	88%	Z46724.1
G.domesticus artificial single chain antibody gene (B1)	442	822	83%	2e-122	88%	Z46721.1
G.domesticus artificial single chain antibody gene (T2)	427	779	83%	5e-118	88%	Z46730.1
G.domesticus artificial single chain antibody gene (T1)	422	822	83%	2e-116	87%	Z46725.1
Synthetic construct for anti-Mesenchymal stem cell ScFv antibody, clone TMSC1-scFv	436	889	82%	8e-121	93%	FN555102.1

Figure 2.5 A: Screenshot of BLAST analysis of scFv A4 sequences against nucleotide database of chicken derived scFvs in *Gallus gallus* database of BLAST N.



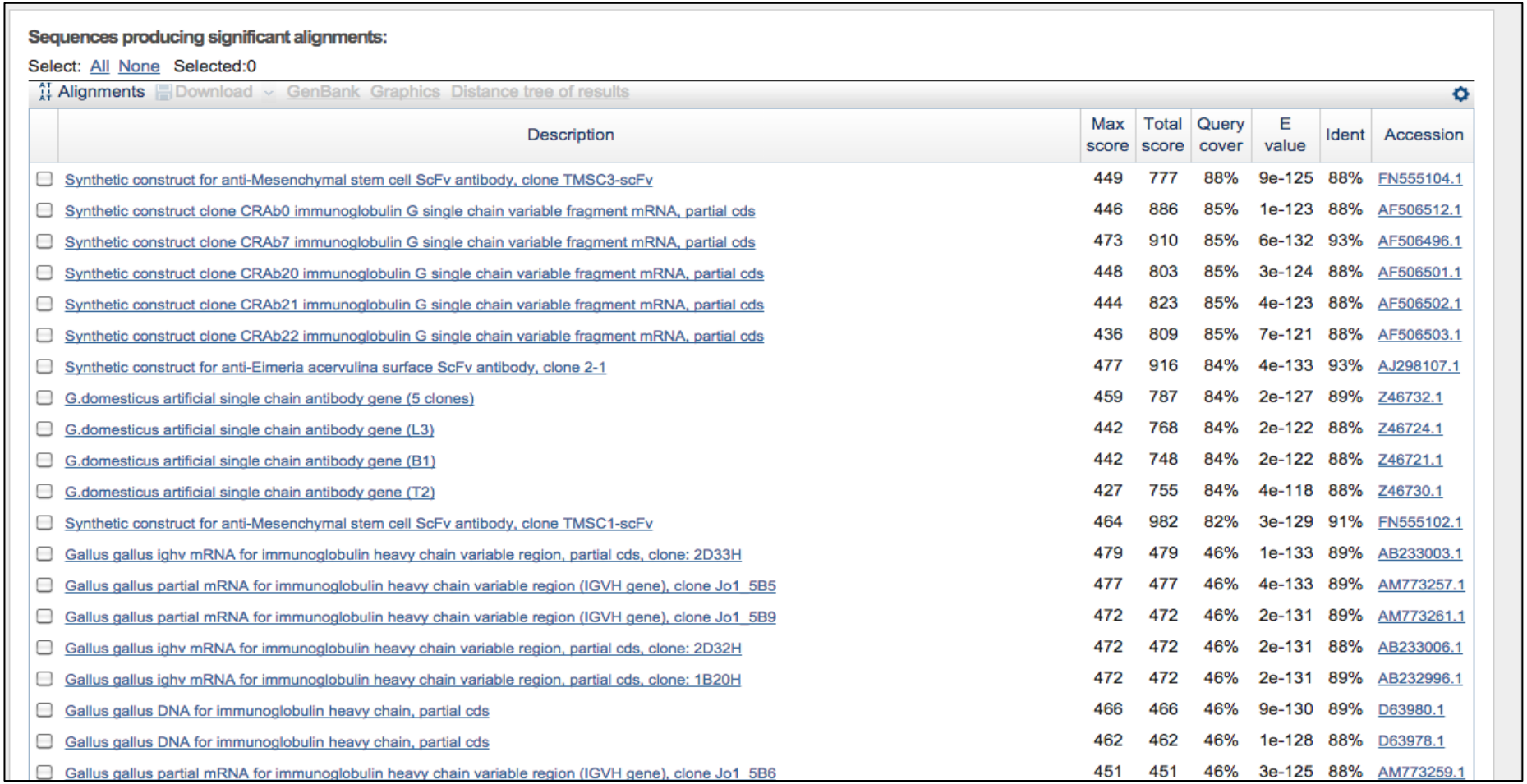
Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
Synthetic construct for anti-Mesenchymal stem cell ScFv antibody, clone TMSC4-scFv	436	833	88%	8e-121	90%	FN555105.1
Synthetic construct for anti-Mesenchymal stem cell ScFv antibody, clone TMSC3-scFv	455	864	87%	2e-126	88%	FN555104.1
Synthetic construct clone CRAb0 immunoglobulin G single chain variable fragment mRNA, partial cds	446	779	84%	1e-123	88%	AF506512.1
Synthetic construct clone CRAb34 immunoglobulin G single chain variable fragment mRNA, partial cds	433	799	84%	1e-119	91%	AF506507.1
Synthetic construct clone CRAb21 immunoglobulin G single chain variable fragment mRNA, partial cds	438	811	84%	2e-121	88%	AF506502.1
Synthetic construct clone CRAb20 immunoglobulin G single chain variable fragment mRNA, partial cds	420	827	84%	8e-116	87%	AF506501.1
Synthetic construct clone CRAb8 immunoglobulin G single chain variable fragment mRNA, partial cds	418	801	84%	3e-115	91%	AF506497.1
Synthetic construct clone CRAb72 immunoglobulin G single chain variable fragment mRNA, partial cds	416	739	84%	1e-114	87%	AF506509.1
Synthetic construct for anti-Eimeria acervulina surface ScFv antibody, clone 2-1	416	735	84%	1e-114	87%	AJ298107.1
G.domesticus artificial single chain antibody gene (L3)	442	794	83%	2e-122	88%	Z46724.1
G.domesticus artificial single chain antibody gene (5 clones)	431	805	83%	4e-119	87%	Z46732.1
G.domesticus artificial single chain antibody gene (T2)	416	757	83%	1e-114	87%	Z46730.1
Synthetic construct for anti-Mesenchymal stem cell ScFv antibody, clone TMSC1-scFv	424	865	82%	6e-117	92%	FN555102.1
Synthetic construct clone CRAb5 immunoglobulin G single chain variable fragment mRNA, partial cds	433	873	79%	1e-119	92%	AF506495.1
Synthetic construct clone CRAb3 immunoglobulin G single chain variable fragment mRNA, partial cds	427	867	79%	5e-118	91%	AF506494.1
Synthetic construct for anti-Mesenchymal stem cell ScFv antibody, clone TMSC2-scFv	442	801	76%	2e-122	91%	FN555103.1
Gallus gallus ighv mRNA for immunoglobulin heavy chain variable region, partial cds, clone: 2D33H	451	451	46%	3e-125	88%	AB233003.1
Gallus gallus ighv mRNA for immunoglobulin heavy chain variable region, partial cds, clone: 2D39H	444	444	46%	5e-123	88%	AB233009.1

Figure 2.5 B: Screenshot of BLAST analysis of scFv G12 sequences against nucleotide database of chicken derived scFvs in *Gallus gallus* database of BLAST N.



Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
Synthetic construct for anti-Mesenchymal stem cell ScFv antibody, clone TMSC3-scFv	449	777	88%	9e-125	88%	FN555104.1
Synthetic construct clone CRAb0 immunoglobulin G single chain variable fragment mRNA, partial cds	446	886	85%	1e-123	88%	AF506512.1
Synthetic construct clone CRAb7 immunoglobulin G single chain variable fragment mRNA, partial cds	473	910	85%	6e-132	93%	AF506496.1
Synthetic construct clone CRAb20 immunoglobulin G single chain variable fragment mRNA, partial cds	448	803	85%	3e-124	88%	AF506501.1
Synthetic construct clone CRAb21 immunoglobulin G single chain variable fragment mRNA, partial cds	444	823	85%	4e-123	88%	AF506502.1
Synthetic construct clone CRAb22 immunoglobulin G single chain variable fragment mRNA, partial cds	436	809	85%	7e-121	88%	AF506503.1
Synthetic construct for anti-Eimeria acervulina surface ScFv antibody, clone 2-1	477	916	84%	4e-133	93%	AJ298107.1
G.domesticus artificial single chain antibody gene (5 clones)	459	787	84%	2e-127	89%	Z46732.1
G.domesticus artificial single chain antibody gene (L3)	442	768	84%	2e-122	88%	Z46724.1
G.domesticus artificial single chain antibody gene (B1)	442	748	84%	2e-122	88%	Z46721.1
G.domesticus artificial single chain antibody gene (T2)	427	755	84%	4e-118	88%	Z46730.1
Synthetic construct for anti-Mesenchymal stem cell ScFv antibody, clone TMSC1-scFv	464	982	82%	3e-129	91%	FN555102.1
Gallus gallus ighv mRNA for immunoglobulin heavy chain variable region, partial cds, clone: 2D33H	479	479	46%	1e-133	89%	AB233003.1
Gallus gallus partial mRNA for immunoglobulin heavy chain variable region (IGVH gene), clone Jo1_5B5	477	477	46%	4e-133	89%	AM773257.1
Gallus gallus partial mRNA for immunoglobulin heavy chain variable region (IGVH gene), clone Jo1_5B9	472	472	46%	2e-131	89%	AM773261.1
Gallus gallus ighv mRNA for immunoglobulin heavy chain variable region, partial cds, clone: 2D32H	472	472	46%	2e-131	88%	AB233006.1
Gallus gallus ighv mRNA for immunoglobulin heavy chain variable region, partial cds, clone: 1B20H	472	472	46%	2e-131	89%	AB232996.1
Gallus gallus DNA for immunoglobulin heavy chain, partial cds	466	466	46%	9e-130	89%	D63980.1
Gallus gallus DNA for immunoglobulin heavy chain, partial cds	462	462	46%	1e-128	88%	D63978.1
Gallus gallus partial mRNA for immunoglobulin heavy chain variable region (IGVH gene), clone Jo1_5B6	451	451	46%	3e-125	88%	AM773259.1

Figure 2.5 C: Screenshot of BLAST analysis of scFv A11 sequences against nucleotide database of chicken derived scFvs in *Gallus gallus* database of BLAST N.

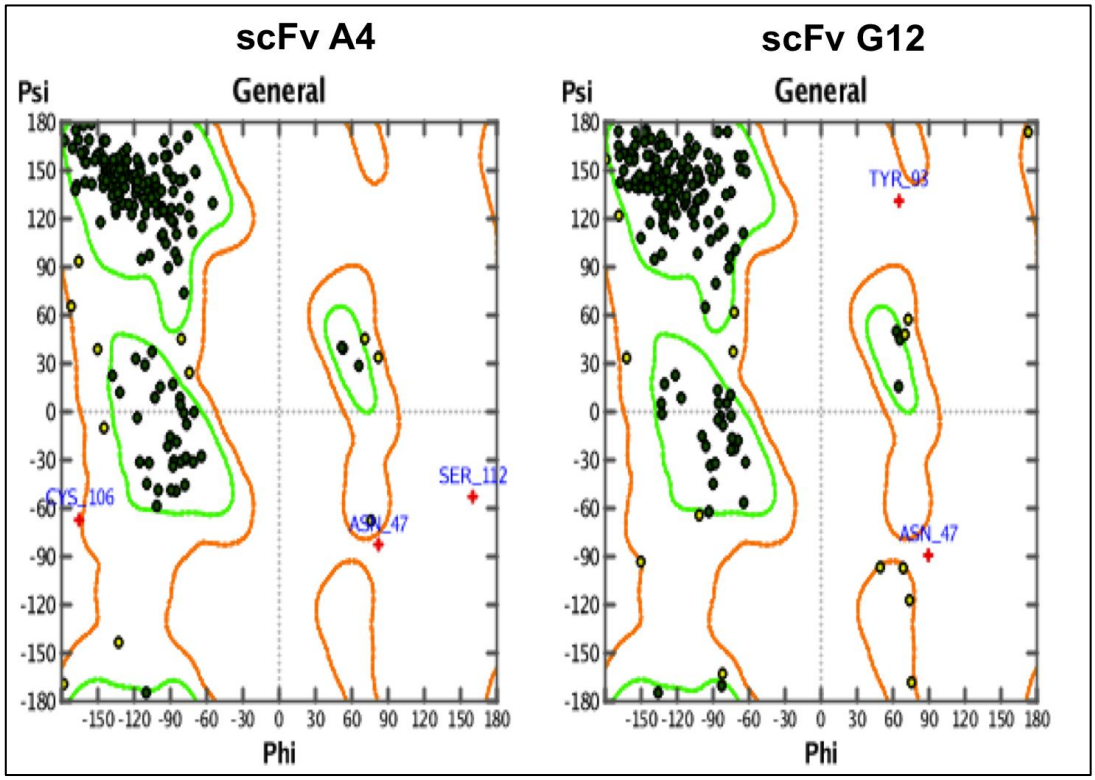
A threading approach was then implemented using multiple templates for both framework and CDRs (Contreras-Moreira et al., 2003, Majumdar et al., 2011). Within the MOE antibody modeller suite, three framework templates were identified on the basis of amino acid sequence similarity with low E value scores (high confidence), with PDB IDs: 8FAB (E value- 6.63838E-29), 3G6A (E value-1.46659E-28) and 3GCD (E value-1.46659E-28), for framework modelling of scFvs A4 and G12. The template 3G6D is the X-ray crystal structure of human anti-interleukin-13 mAb, 8FAB is the crystal structure of Fab fragment from the human anti-myeloma IgG, 1Q1J is the crystal structure of human anti-HIV mAb, and 3BN9 is the crystal structure of human Fab-protease complex. Similarly, templates were selected for CDR modelling and only those templates that showed sequence similarity equal to or greater than 55% to their respective CDRs were considered (Table 2.3). Individual loops were then modelled and those that gave a quality score of 65 or higher based on C-α root mean square deviation (RMSD) with a resolution of 0.3 - 0.5 Å were clustered together. The candidate loop with the highest score within a cluster was selected as representative of that cluster. The structural scores of the loops generated from selected frameworks and CDR templates were in the range of 65.9 – 98.8 (Table 2.3).

Table 2.3: PDB Templates used for Homology Modelling. A threading approach was being used using multiple separate templates to model both framework and CDRs of all three scFvs.

	Framework	**CDR1**	**CDR2**	**CDR3**
scFv-A4:				
V_L-	3G6D.L (66.9)*	3G6D.L (69.3)	1Q1J.L (82.2)	3G6D.L (65.9)
V_H-	3G6A.H (83.2)	3G6A.H (83.9)	3G6A.H (84.0)	3G6A.H (83.7)
scFv-G12:				
V_L-	8FAB.C (93.1)	8FAB.C (94.6)	1Q1J.L (82.2)	8FAB.C (96.1)
V_H-	8FAB.D (98.8)	3BN9.F (93.5)	8FAB.D (94.8)	8FAB.D (96.6)
scFv-A11:				
V_L-	8FAB.A (93.1)	3GBM.L (90.4)	3H42.L (94.6)	3GCA.L (85.2)
V_H-	8FAB.D (98.8)	3BN9.F (93.5)	3GCD.H (70.5)	1ZA3.B (50.4)

* Multiple templates with their PDB ID used for homology modelling of Framework, CDR1, CDR2 and CDR3 of light and heavy chain of scFv-A4, scFv-G12 and scFv A11. The parenthesis scores are structural scores of loop compared to their respective template. The template 3G6D is the X-ray crystal structure of human anti-interleukin-13 monoclonal antibody; 8FAB is the crystal structure of Fab fragment from the human myeloma immunoglobin I_gG. 1Q1J is the crystal structure of anti HIV human monoclonal antibody, and 3BN9 is the crystal structure of human Fab-protease complex.

In grafting the loops for the framework and CDR regions together, structures were energy minimised to relieve strained geometry and bring the molecules as a potential energy model, where the forces acting on each atom of the scFv are zero and this conformation of the molecule is considered to be the most stable. Thirty poses were generated considering appropriate numbers of poses to analyse model for each scFv, from which the final model was selected based on several parameters; the RMSD on each intermediate model, the electrostatic solvation energy, a knowledge-based residue packing quality function and an estimation of the effective atomic contact energies of each models. Protein geometries of the final models after energy minimisation were verified using Ramachandran plots, which showed that A4 had three amino acids and G12 two amino acids in the disallowed regions (Tyr 93 and Asn 47) (Figure 2.6). Two of three amino acid outliers in A4 were Cys 106 and Ser 112 from V_H CDR3 while the third residue was Asn 47, the same as G12. The final models of both scFvs were simulated in the presence of water molecules for 1,000 ps using the isobaric-isothermal (NPT) ensemble algorithm and the time-energy plot analysed. An instant increase in energy was observed when the temperature was brought to 310K and thereafter both scFvs had stable energies throughout the course of the simulation, confirming the reliability of the scFv models (Figure 2.7). The homology models of both scFvs with labelled CDRs and framework regions and predicted binding sites are shown in (Figure 2.8).



scFv A11

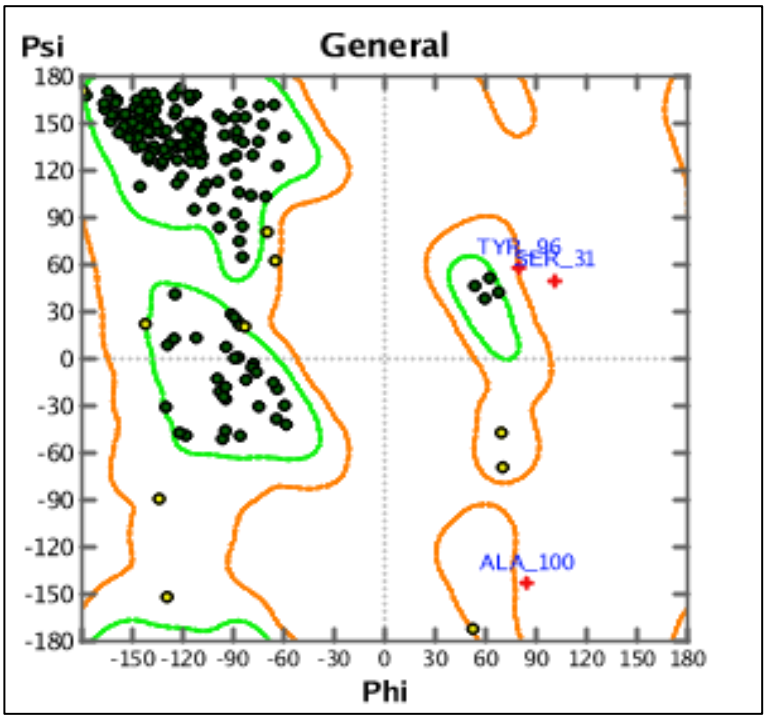


Figure 2.6: Ramachandran Plots of scFv A4 and G12. Ramachandran plot of scFv-A4, scFv-G12 and scFv-A11. Green zone indicates the core region, Yellow zone indicates allowed region, and total area outside these two regions are deemed as the disallowed region.

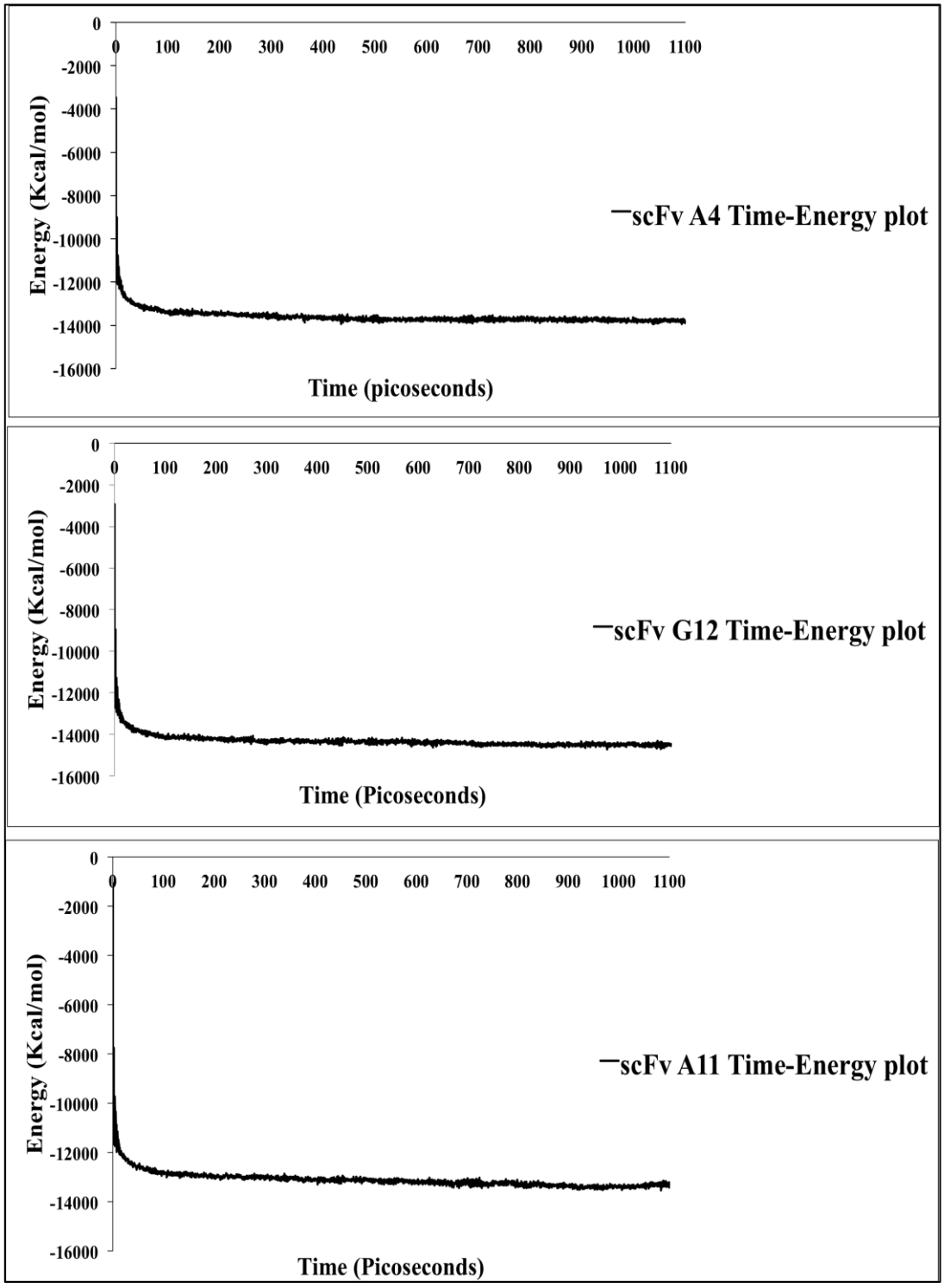
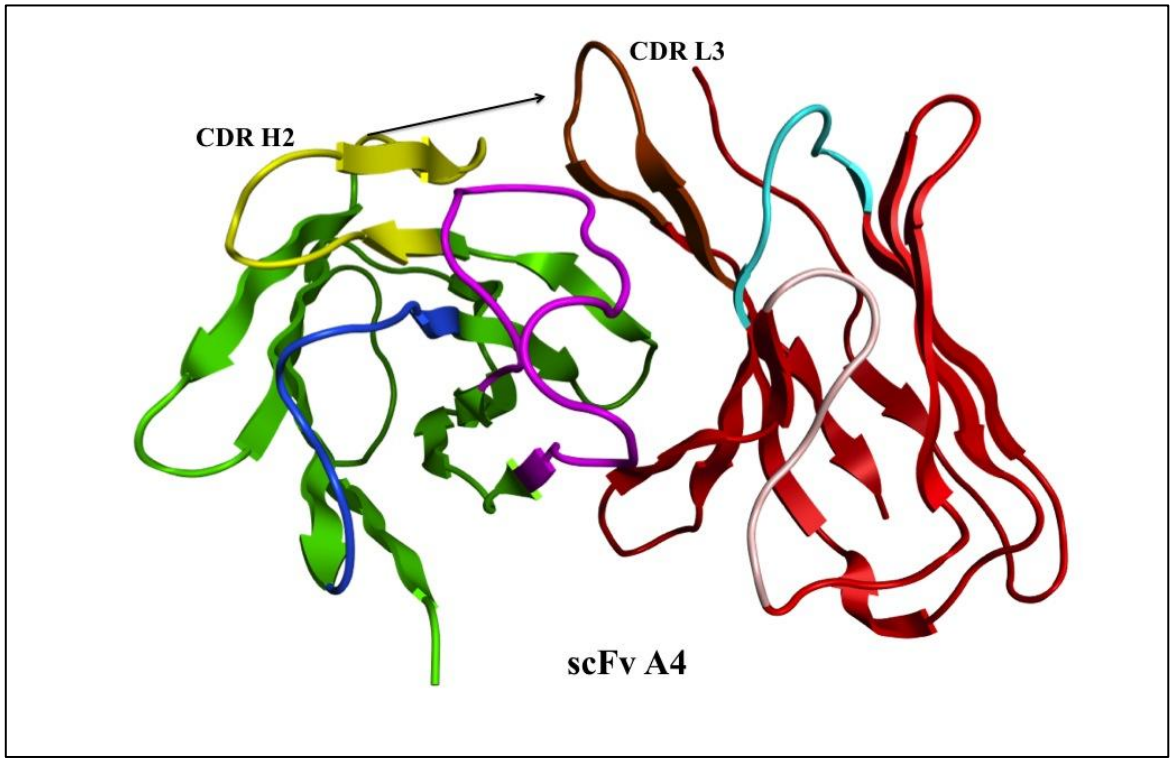
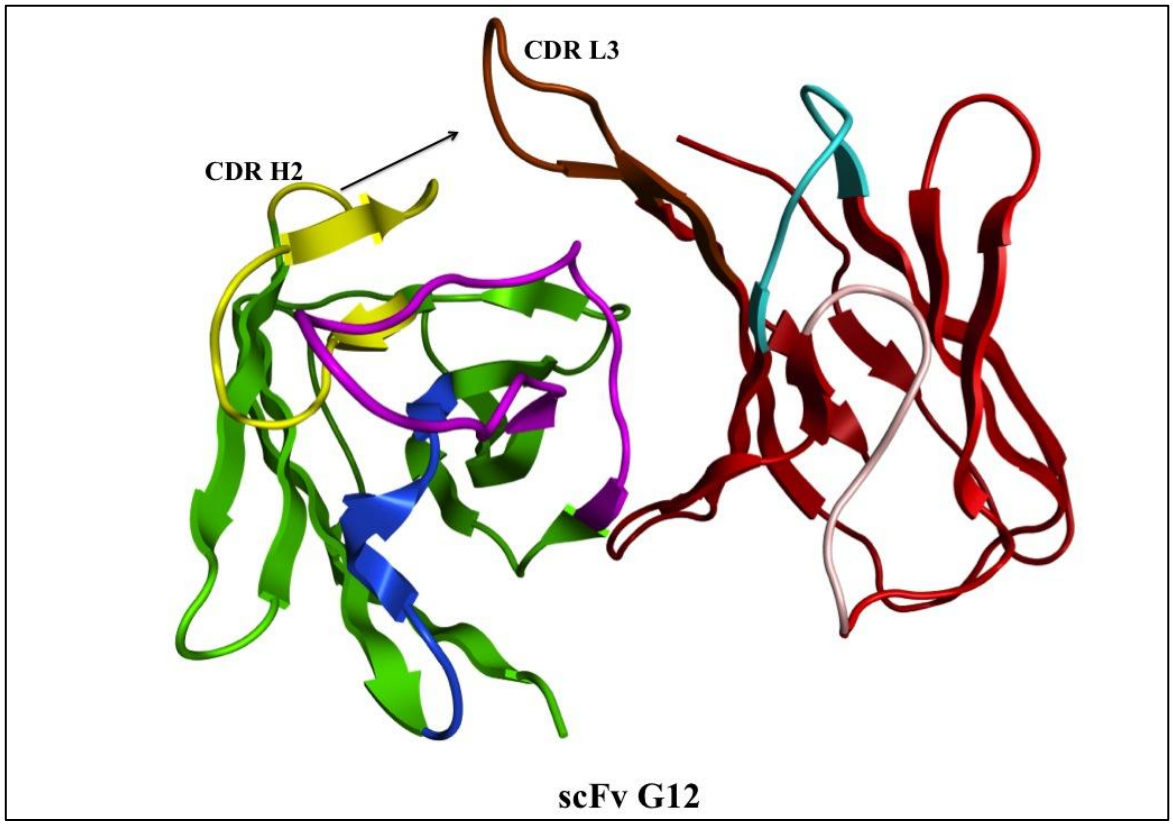


Figure 2.7: Molecular dynamics simulation of scFv A4, G12 and A11. Molecular dynamics simulations were performed for all three scFv (A) A4, (B) G12 and (C) A11 using an NPT statistical ensemble, keeping temperature constant at 310K throughout the simulations. The system was initially heated at 310K for 100 ps followed by 1,000 ps simulation in the presence of water molecule.


CDR L3
CDR H2
scFv A4


CDR L3
CDR H2
scFv G12

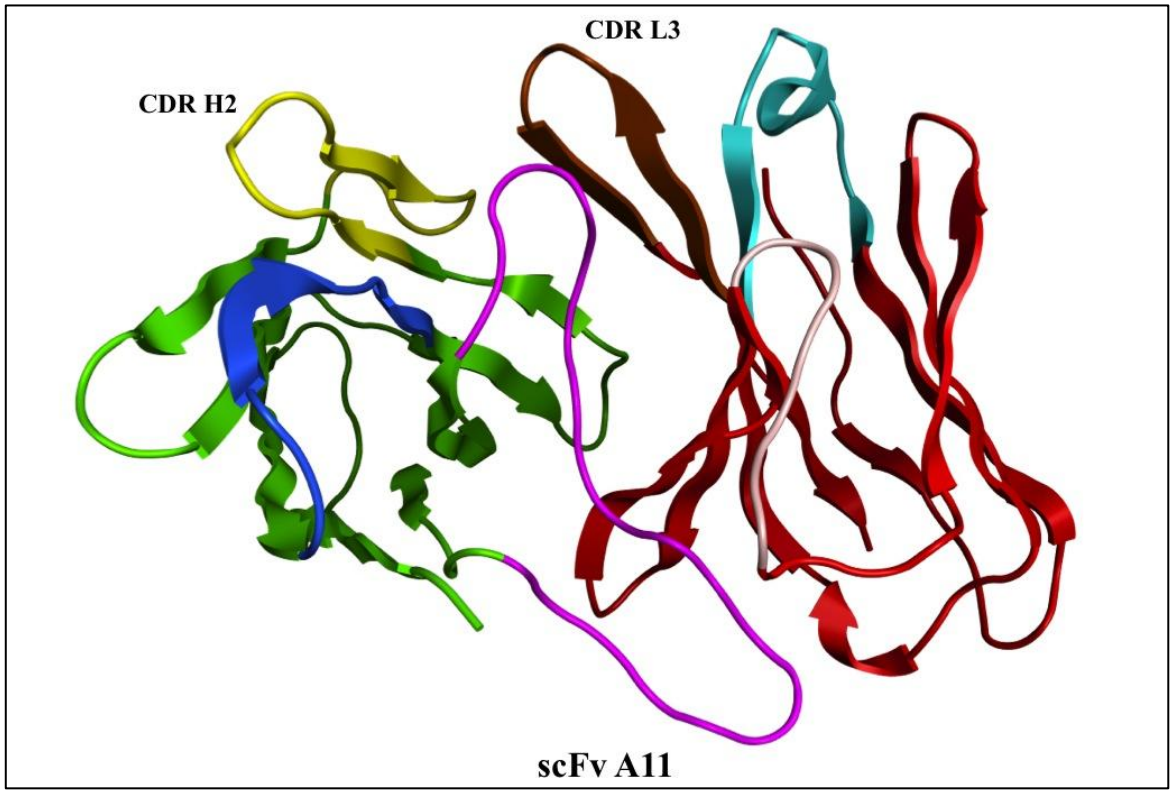


Figure 2.8: Predicted 3D structures of scFvs A4, G12 and A11. Homology scFv models constructed on basis of similarity scores with templates retrieved from PDB; heavy chain (V_H) represented in green, with light chain (V_L) represented in red. In heavy chain; blue, yellow and magenta represented CDR H1, CDR H2, and CDR H3 respectively while in light chain; cyan, pink and brown represented CDR L1, CDR L2 and CDR L3 respectively. Arrow indicates the binding pocket in between CDR L3 and CDR H2 in scFv A4 and G12.

2.3.4. *in silico* analysis of scFv binding to carbohydrates

The structures of seven carbohydrates, Gal-α-(1, 3)-Gal, Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc, α-D-Gal, α-L-Fuc, β-D-Xyl, Gal-α-(1, 2)-Gal and Gal-β-(1, 4)-Gal, were drawn using the MOE carbohydrate builder suite and the GLYCAM server. Two of these, Gal-α-(1, 3)-Gal and Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc, comprised the target ligand while the rest of the structures should not bind to the scFvs. The structures Gal-α-(1, 2)-Gal and Gal-β-(1, 4)-Gal were used to compare scFv binding with Gal-α-(1, 3)-Gal, whereas the monosaccharide structures α-D-Gal, α-L-Fuc, β-D-Xyl, were used to differentiate between mono and di-saccharides. The binding sites of the modelled scFvs were predicted using the Site-Finder package, based on three-dimensional alpha complexes, comprising molecular surface, solvent accessibility,

voids and envelopes within the receptor (Edelsbrunner et al., 1995). All carbohydrate structures were then docked flexibly into the scFv binding sites in the presence of water molecules, providing accessibility to carbohydrates onto the receptors. Binding energies for carbohydrate-scFv interactions were calculated (Table 2.4). Although, there was no cutoff value for binding of glycans against scFv, the binding energy values are realtively compared between the glycans to validate strong binding. The lower (calculated more negative energy) the binding energy value, higher the binding affinity, comparing binding energies across the other disaccharide structures tested. Of the seven carbohydrates examined, the lowest binding energy was found for Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc, followed by Gal-α-(1, 3)-Gal with both scFvs (Table 2.4). Both scFvs A4 and G12 showed lower binding energies with the Gal-α-(1, 3)-Gal structure than with either the Gal-α-(1, 2)-Gal or Gal-β-(1, 4)-Gal structures, however in scFv A11 binding against glycans Gal-β-(1, 4)-Gal showed lower binding energy than Gal-α-(1, 3)-Gal. The *in silico* binding energy pattern of scFv A11 did not correlate with scFv A4 and G12 binding, this difference was also validated with the SPR binding results of scFv against Gal-α-(1, 3)-Gal (Table 2.2). When the binding energies of the two scFvs A4 and G12 with either the trisaccharide Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc or the disaccharide Gal-α-(1, 3)-Gal were considered, the binding energies of the G12 complexes were lower (or higher binding affinity) than the A4 complexes, which agrees with affinity parameters derived from SPR and reported ELISA data for these scFvs (Cunningham et al., 2012). Similarly *in silico* binding affinities of scFv All (disaccharide: -18.0 kcal/Mol, trisaccharide: -22.2 kcal/Mol) were smaller (less negative) than that of scFv G12 (disaccharide: -31.6 kcal/Mol, trisaccharide: -34.4 kcal/Mol) and scFv A4 (disaccharide: -25.3 kcal/Mol, trisaccharide: -31.4 kcal/Mol).

Table 2.4: Estimated binding energy between scFv molecule and selected glycan structures derived from the predicted structures of scFv-glycan epitope complex.

	Binding Energy (kcal/Mol)		
Glycan/scFv	**scFv A4**	**scFv G12**	**scFv A11**
α-D-Galactose	**-16.6**	**-18.7**	**-13.0**
α-L-Fucose	**-10.2**	**-17.2**	**-10.3**
β-D-Xylose	**-15.7**	**-10.7**	**-11.1**
Gal-α-(1, 2)-Gal	**-19.6**	**-24.7**	**-15.8**
Gal-α-(1, 3)-Gal	**-25.3**	**-31.6**	**-18.0**
Gal-β-(1, 4)-Gal	**-21.7**	**-23.5**	**-19.9**
Gal-α-(1, 3)-Gal-β-(1, 4) GlcNAc	**-31.4**	**-34.4**	**-22.2**

* Binding energy scores of top pose of both scFvs' against selected glycan structures calculated using MM/GBVI algorithm. Values are a combined determination of van der Waals, electrostatic energies and other calculated interactions.

2.3.5. Identification of scFv amino acids residues in contact with ligand

The five lowest binding energy poses of each scFv complexes with the monosaccharide, α-D-Gal, the disaccharide, Gal-α-(1, 3)-Gal, and the trisaccharide, Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc, were examined manually to identify amino acid residues of the scFv that came in contact with the carbohydrates (Table 2.5). With both scFvs, fewer residues were deemed to make contact with the α-Gal than with the known binding against Gal-α-(1, 3)-Gal or Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc as expected. Also, there was no CDR involved in the binding of scFv A4 against α-Gal, however in scFv G12 binding V_L CDR1 shows contacts against α-D-Gal, but V_L CDR 1 has no role to play in binding against Gal-α-(1, 3)-Gal or Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc, therefore it could not be considered as a good binding contact against

scFv G12 (Table 2.5 A & B). Interestingly, a greater degree of contact was noted between the V_L and each carbohydrate structure with both scFvs and there was a good overlap between the residues involved in G12 and A4. The V_L hexapeptide region between residues 35-41 (K35, P36, G37, Q38, A39 and V41) had a amino acid contact region for both scFvs with Gal-α-(1, 3)-Gal and Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc and there were also a number of common amino acid contact between residues 75 and 99 against scFvs A4 and G12 (Table 2.6 A and B). More contact points were identified between α-D-Gal and Gal-α-(1, 3)-Gal in the A4 V_H compared to G12, while the opposite was true for Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc. Interaction of the A4 V_H with these carbohydrates was spread over three regions: residues 61-65 (G61, P62 and K65), 93-95 (V93 and Y95) and 121-124 (Q121, G122 and L124) (Table 2.5 B).

In scFv A4, V_L residue K35, G37, Q38 and V41 were the common residue contacts in all three α-D-Gal, Gal-α-(1, 3)-Gal, and Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc structures, whereas in V_H, residues V93, Q121 and L124 were the common contacts for all three carbohydrates (Table 2.5 A).

In scFv G12, α-D-Gal did not show similar binding contacts as with other carbohydrates except residue E3 in V_L, which was a common contact amino acid for all three-carbohydrate structures. Between Gal-α-(1, 3)-Gal, and Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc, Q3, K35, P36, Y93 and F97 in V_L were the common residues that bound to both carbohydrates, whereas in V_H CDR2 residues P62, A63, and K65 were the common binding contacts (Table 2.5 B).

In scFv A11 binding against Gal-α-(1, 3)-Gal active site does not have any CDR amino acids involved in binding, infact light chain framwork aminoacid Asp86, Tyr87, Tyr 88, Cys89 and Gly101, Gly102, Gly103, Thr104, Lys105 were involved which were positioned just before and after light chain CDR 2 amino acids repectively (Table 2.5 C).

Table 2.5: Active site amino acid residues of scFv, **(A) A4, and (B) G12** involved in binding against α-D-Galactose (monosaccharide) Gal-α-(1, 3)-Gal (disaccharide) and Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc (trisaccharide) determined by examination of the top 5 poses for proximity between the residue and the ligand using a cut-off of 4.5A. The CDR residues are labelled in red colour.

scFv A4	**Light chain**														**Heavy chain**									
Mono-	35	37	38	41	77										93	121	122	124						
	K	**G**	**Q**	**V**	**E**										**V**	**G**	**Q**	**L**						
Di-	35	36	37	38	41	75	79	88	89	90	91	92			57	58	59	61	62	65	93	103	121	124
	K	**P**	**G**	**Q**	**V**	**E**	**E**	**D**	**N**	**T**	**Y**	**V**			**Y**	**T**	**Y**	**G**	**P**	**K**	**V**	**Y**	**Q**	**L**
Tri-	3	5	34	35	36	37	38	39	41	53	55	57	77	79	34	39	41	93	95	121	122	124		
	E	**T**	**Q**	**K**	**P**	**G**	**Q**	**A**	**V**	**G**	**R**	**P**	**E**	**E**	**M**	**Q**	**P**	**V**	**Y**	**Q**	**G**	**L**		

scFv G12	**Light chain**															**Heavy chain**											
Mono-	3	48	49	50	51	52	56	99								43	44	47									
	E	**N**	**Q**	**R**	**P**	**S**	**Q**	**G**								**K**	**G**	**W**									
Di-	3	5	6	7	8	35	36	79	81	83	92	93	94	96	97	61	62	63	64	65	124						
	E	**T**	**Q**	**P**	**P**	**K**	**P**	**E**	**D**	**Y**	**T**	**Y**	**A**	**I**	**F**	**G**	**P**	**A**	**V**	**K**	**L**						
Tri-	3	9	11	35	36	38	39	67	93	97	98	99	102	104		42	43	44	49	50	56	62	63	65	66	67	87
	E	**S**	**S**	**K**	**P**	**Q**	**A**	**R**	**Y**	**F**	**G**	**G**	**K**	**T**		**G**	**K**	**G**	**A**	**G**	**R**	**P**	**A**	**K**	**G**	**R**	**R**

Table 2.5: Active site amino acid residues of scFv, **(C) A11** involved in binding against α-D-Galactose (monosaccharide) Gal-α-(1, 3)-Gal (disaccharide) and Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc (trisaccharide) determined by examination of the top 5 poses for proximity between the residue and the ligand using a cut-off of 4.5A. The CDR residues are labelled in red colour.

scFv A11	**Light chain**																**Heavy chain**					
Mono-	8	9	42	**53**	**56**												**2**	**3**	**5**	47		
	P	**S**	**G**	**S**	**D**												**V**	**Q**	**Q**	**W**		
Di-	5	6	7	8	9	42	51	53	87	88	89	101	102	103	104	105	40	41	42	43	44	45
	T	**Q**	**P**	**P**	**S**	**G**	**S**	**D**	**Y**	**Y**	**C**	**G**	**G**	**G**	**T**	**K**	**A**	**P**	**G**	**K**	**G**	**L**
Tri-	6	7	8	43	86	88	102	103	104	105							41	42	43	114	116	
	Q	**P**	**P**	**R**	**D**	**Y**	**G**	**G**	**T**	**K**							**P**	**G**	**K**	**D**	**I**	

2.3.6. Anatomy of binding sites for scFv A4 and G12 Gal-α-(1, 3)-Gal and Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc

As is evident from the SPR and *in silico* binding results, scFv A11 had the least binding affinity against Gal-α-(1, 3)-Gal and Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc compared to scFv A4 and G12. Also, the *in silico* results demonstrated the non-specific binding of scFv A11 against di-saccharides. Therefore, only scFv A4 and G12 were considered for further binding site analyses. The molecular surface of the binding site in both scFvs with Gal-α-(1, 3)-Gal and Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc showed that the majority of the area was hydrophobic (green), with many small patches of hydrophilic regions (red), and a few traces of mild polar region (white) (Figure 2.9). Most of the amino acid residues involved in binding were from the CDRs of both V_L and V_H. The amino acids of scFv-A4 that bound **Gal-α-(1, 3)-Gal** were from V_L CDR3 Y87 – V92 and V_H CDR2 Y57 - G66 and CDR3 Y104. The framework residue, Leu124, also directly interacted with the motif whereas the main pocket formed by CDRs from both V_L and V_H, detailed in (Figure 2.9; Table 2.6). Amino acids from scFv-G12 involved in binding include V_L CDR3 D88 - I96 and V_H CDR2 F60 - K65.

According to the models generated, the interactions of scFvs A4 and G12 with Gal-α-(1, 3)-Gal involved a number of the same residues in V_L and V_H CDRs. These included D88, N89, T90, and Y91 in V_L CDR3 of scFv A4, which correspond to D88, N91, T92, and Y93 in scFv G12 (because of the two extra serine residues at positions 89 and 90 in scFv G12), and amino acids F60 - K65 of V_H CDR2 in both scFv active sites. Also, the framework residue L124 present in the V_H chains of both scFv's is involved in both cases.

For scFv A4, the binding site was formed with heavy CDR2 (Y57, Y59, F60, P62, and K65) and light chain CDR3 (Y85, D88, N89, Y91 and V92) (Figure 2.10 A). Residues hydrogen binding interaction with ligand was at maximum distance of 2.5 Å from the ligand. Metal ligation interaction was found to be 2.6 Å from the ligand. Three water molecules were also involved in the hydrophilic region forming hydrogen bonding and two of three water molecules bridges hydrogen bonding with ligand and residue (F60 and Y104) (Figure 2.9 A).

For scFv G12, the heavy chain CDR2 (Y59 – K65) and light chain CDR3 (D88, S90, T92, V94, G95 and I96) were directly binding with Gal-α-(1, 3)-Gal (Figure 2.9 B). The ligand was fixed in to the hydrophobic pocket between the two CDRs of light and heavy chain. There were more hydrophilic regions than A4 and seven water molecules were directly interacting with the residue and ligand. Hydrogen bonding between the residue and ligand varies form 1.9 to 3.7 Å whereas the metal ligation distances were from 1.9 to 2.2 Å.

Also, in scFv G12 complexes with trisaccharide, Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc, only heavy chain residues were involved in the active site, CDR residues (P62, A63, V64, K65, G 66 and R67) and a few other non CDR residues (S85, R87, E89 and D90) (Figure 2.9 C). A single water molecule near the ligand was involved in binding K65 with the ligand. Solvent hydrogen bonding distance from the residue and the ligand was 1.7 Å and the residue hydrogen-bonding distances from the ligand varied from 1.9 to 3.8 Å.

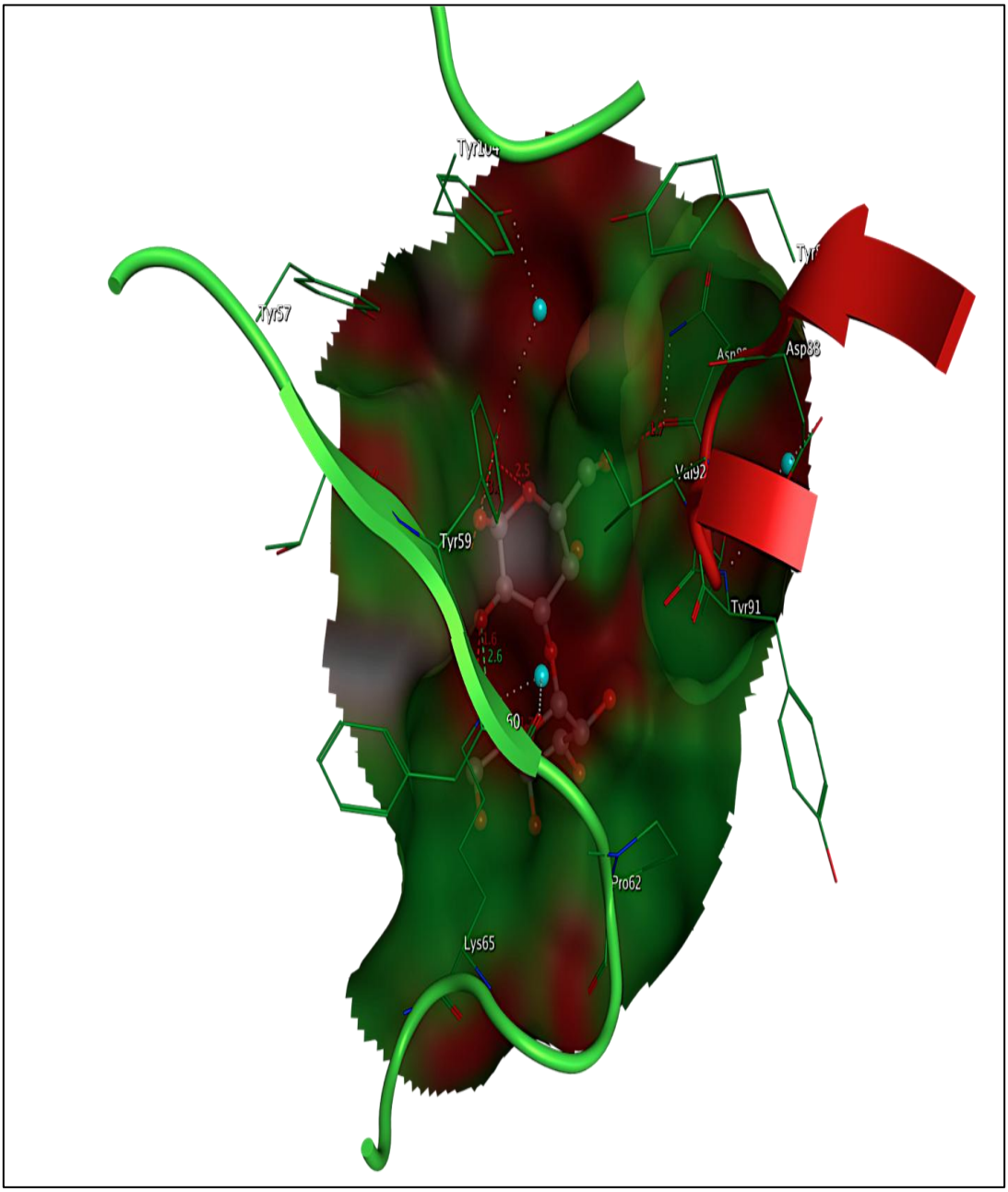


Figure 2.9 A: shows similar residues involved in Gal-α-(1, 3)-Gal binding against scFv A4. In the molecular surface area of active site, which was drawn at 6 Å from the ligand: green, red and white colour indicates hydrophobic region, hydrophilic region, and mild polar region respectively. The blue colour small sise dot balls were shown as water molecules. White, green and red dotted lines were indicated as solvent hydrogen bonding, residue hydrogen bonding, and metal ligation respectively.

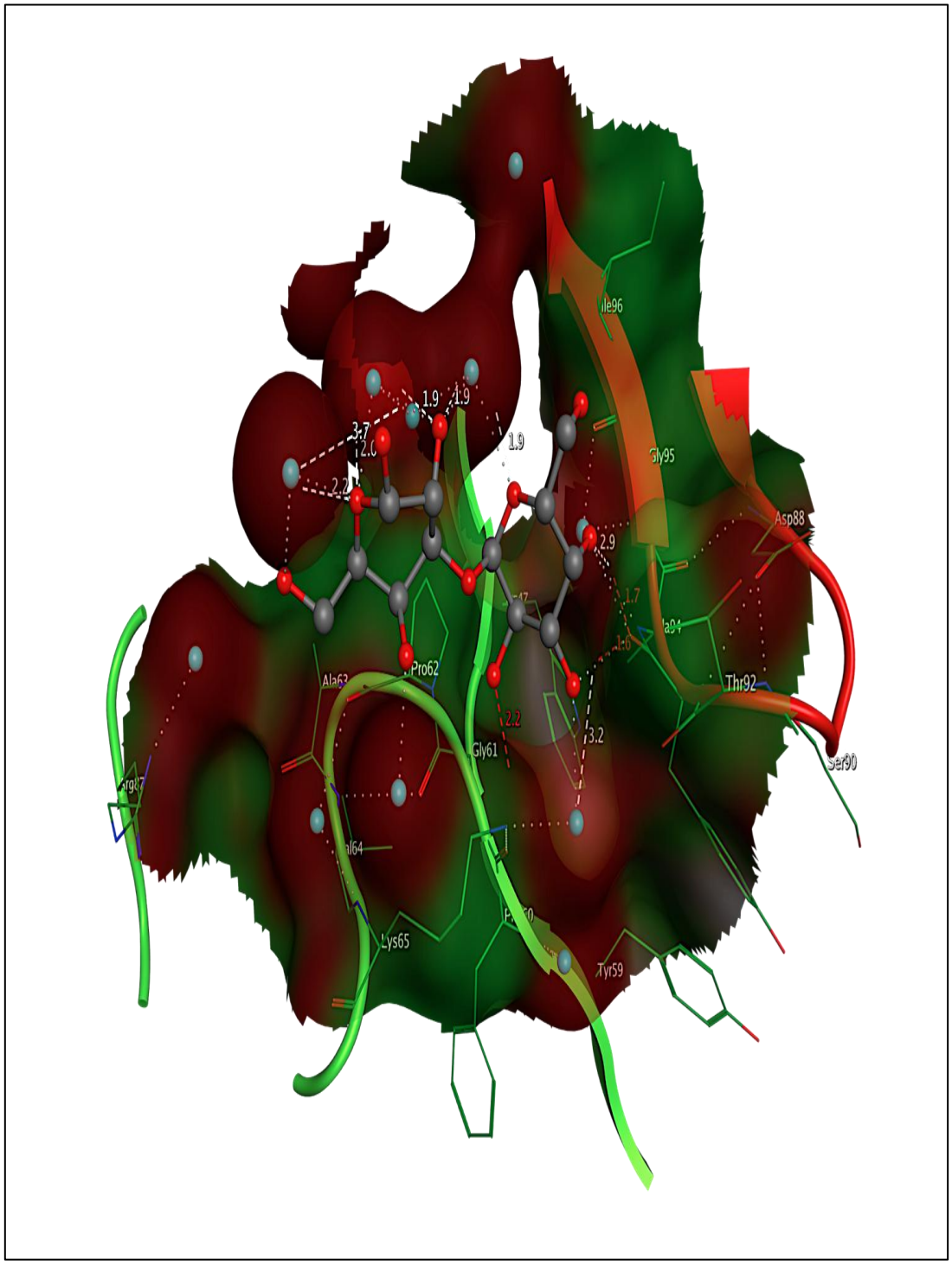


Figure 2.9 B: Similar residues involved in Gal-α-(1, 3)-Gal binding against scFv G12. In the molecular surface area of active site, which was drawn at 6 Å from the ligand: green, red and white colour indicates hydrophobic region, hydrophilic region, and mild polar region respectively. The blue colour small sise dot balls were shown as water molecules. White, green and red dotted lines were indicated as solvent hydrogen bonding, residue hydrogen bonding, and metal ligation respectively.

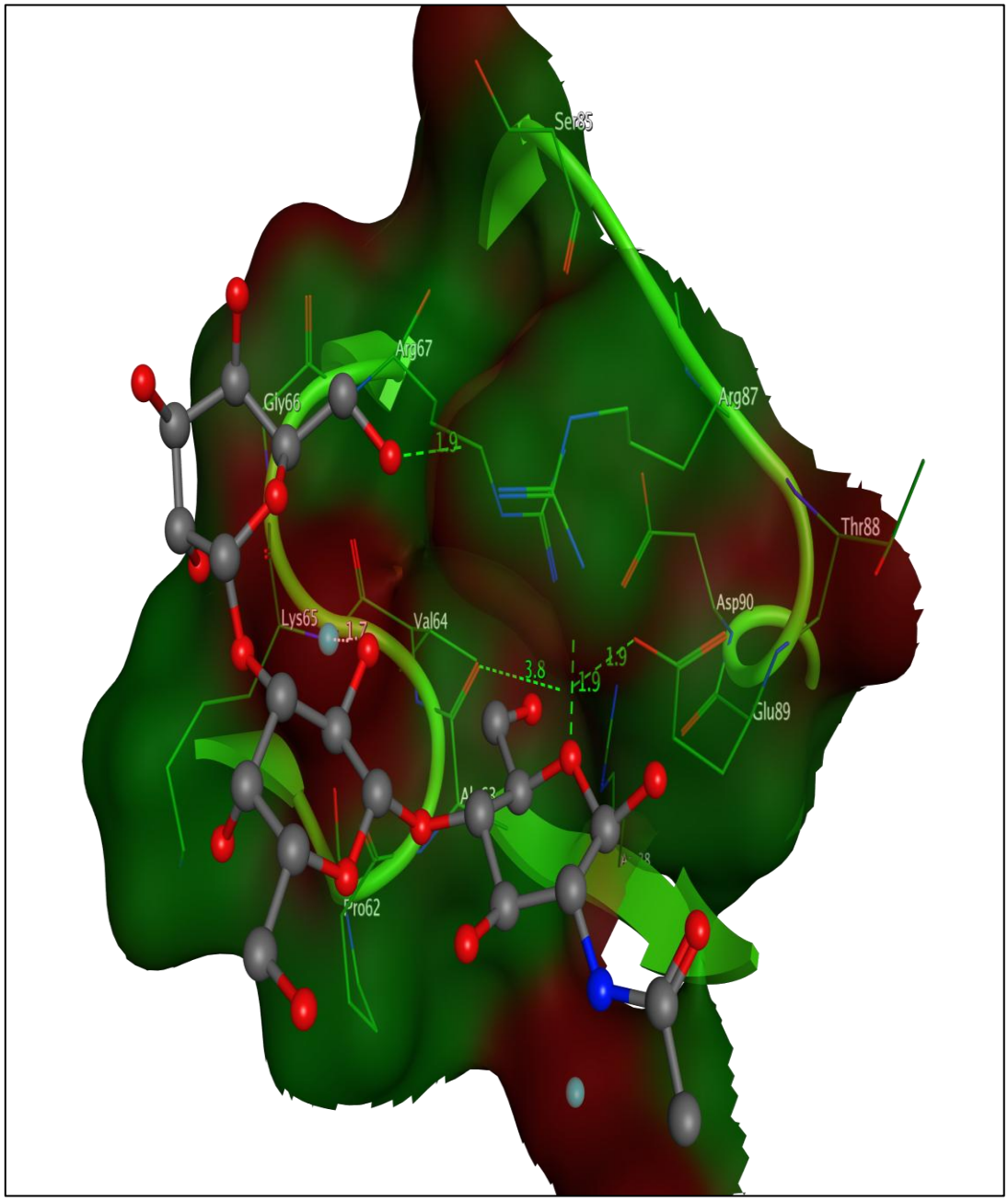
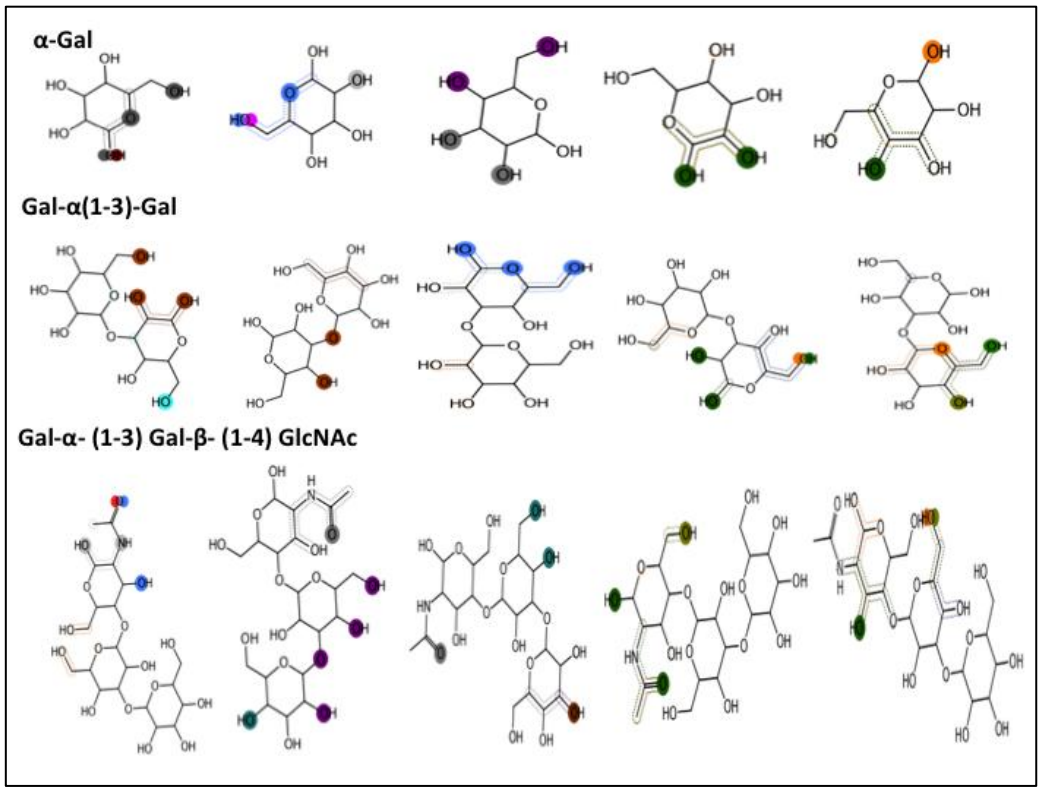


Figure 2.9 C: shows similar residues involved in Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc binding against scFv A4. In the molecular surface area of active site, which was drawn at 6 Å from the ligand: green, red and white colour indicates hydrophobic region, hydrophilic region, and mild polar region respectively. The blue colour small sise dot balls were shown as water molecules. White, green and red dotted lines were indicated as solvent hydrogen bonding, residue hydrogen bonding, and metal ligation respectively.

2.3.7. Epitope analyses of scFv A4 and G12 against Gal-α-(1, 3)-Gal and Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc

The atoms on the target carbohydrates that interacted with the scFvs in the same top five most stable were also explored in the docked complexes (Figure 2.10 A and B). C-6 of Gal at the non-reducing end of each molecule (or of the monosaccharide) had the greatest number of contacts with amino acids of the scFv A4 and G12 binding sites (Figure 2.10 A and B). A higher proportion of heavy atoms at the 6^{th} carbon position (C and O) in the ligand Gal-α-(1, 3)-Gal, contacted amino acid residues in the binding sites of their respective complexes than in the case of Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc. This was most likely because the monosaccharide fits well into the scFv pocket and can potentially interact with a relatively higher number of scFv residues. In complexes with Gal-α-(1, 3)-Gal and Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc, the sugar residue facing towards the binding pocket, i.e. the carbohydrate residue at the non-reducing end had more contact points than the other carbohydrate residues. This suggested that the ligand fitted end-on into the binding pocket and the *N*-acetyl group on the GlcNAc residue at the reducing end of Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc also made contact with both scFvs surface and this interaction could be responsible for the higher measured affinity for both scFvs to the trisaccharide compared to the other carbohydrates tested using SPR and previously reported ELISA (Cunningham et al., 2012).

(A) scFv A4



(B) scFv G12

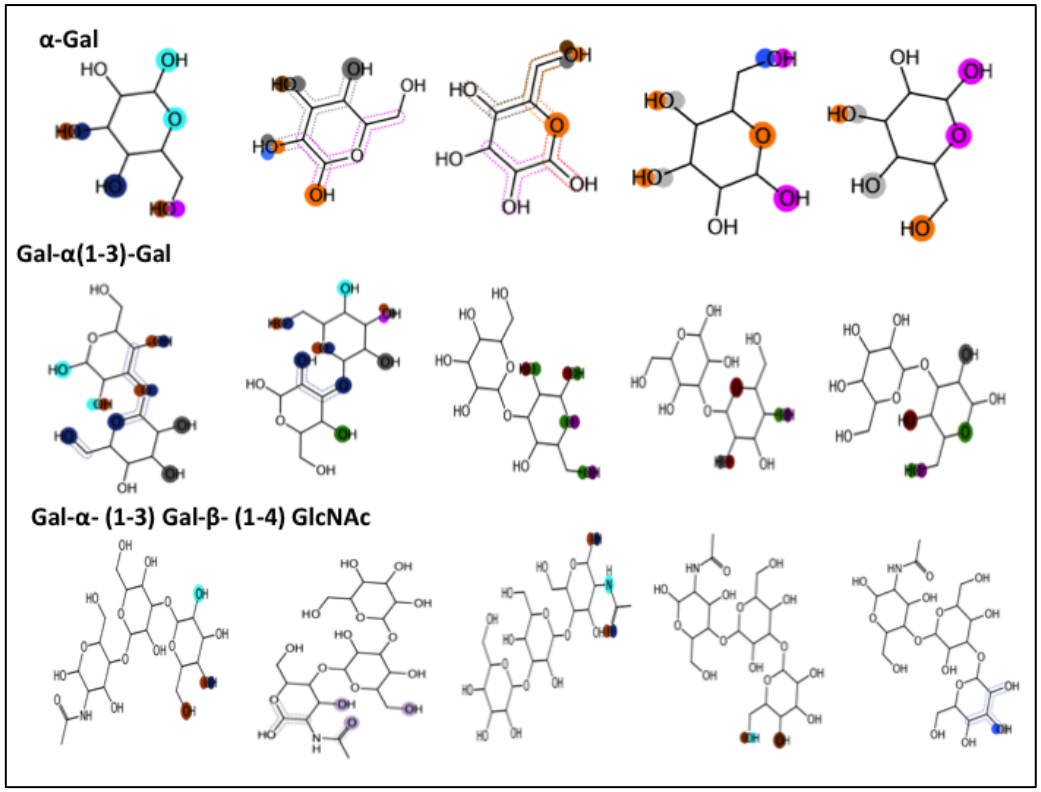


Figure 2.10: shows ligand binding analyses of α-Gal, Gal-α-(1, 3)Gal, and Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc respectively against (A) scFv A4 and (B) scFv G12.

2.3.8. Sequence and structural Comparison between scFv A4 and G12

The overall sequence alignment of scFvs similarities of V_L and V_H between the two scFvs showed that CDR3 of V_H had the greatest differences (Figure 2.11). ScFv G12 had 104 residues in the V_L region, two more amino acids than A4. The V_L CDR1 of G12 and A4 had three differences (Asn-Ser, Gly-Tyr and Asn-His) and G12 had an extended CDR3, with two additional Ser residues. There was only one difference in V_L CDR2 (Lys-Gln). Within V_H, CDR1 was conserved across both scFvs A4 and G12, CDR2 differed in two residues (Tyr-Asn and Ser-Arg), while V_H CDR3 had five substitutions from A4 to G12 (Ala-Gly, Thr-Ser, Ile-Gly, Tyr-Asn and Asp-Asn). Superimposition of V_L and V_H of both scFvs gave RMSD scores of 1.344 Å and 1.61 Å for V_L and V_H, respectively, which indicated a high level of structural similarity between the scFvs (Figure 2.11 A and B).

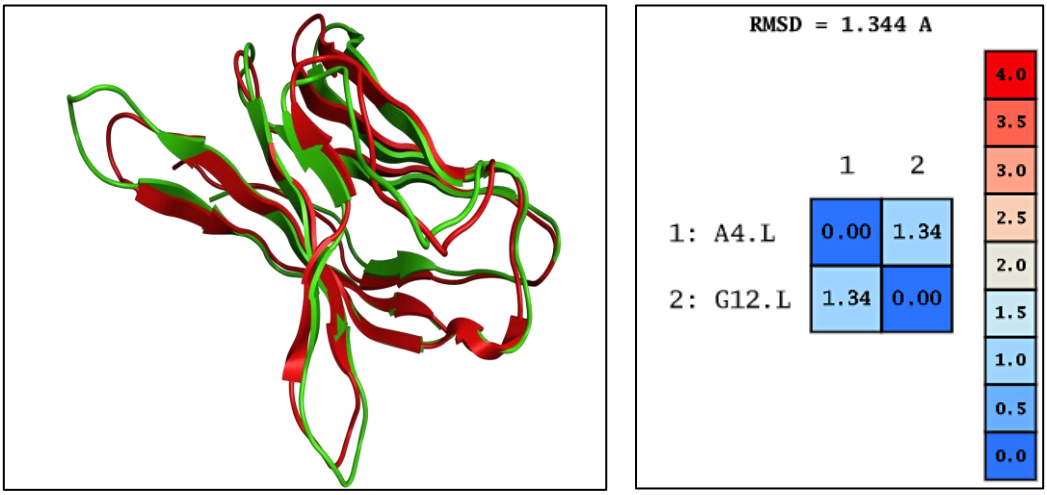


Figure 2.11 A: Light chain alignment of scFv A4 and G12. Red and green chains represent light chains of scFv A4 and G12 respectively.

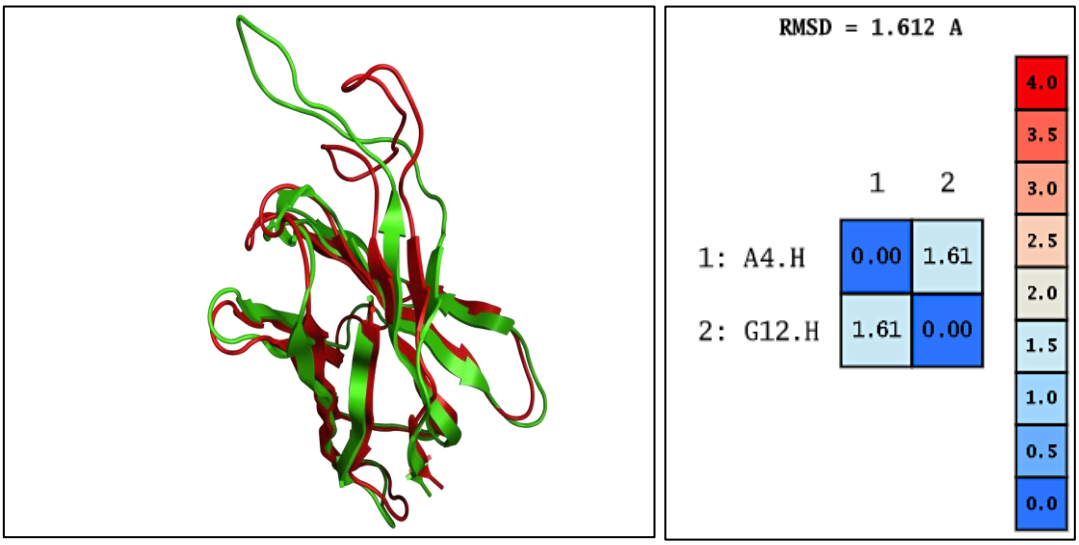


Figure 2.11 B: Heavy chain alignment of scFv A4 and G12. Red and green chains represent heavy chains of scFv A4 and G12 respectively.

2.3.9. scFv A4 and G12 sequence comparison with commerical M86 mAb and engineered scFv against Gal-α-(1, 3)-Gal

Chicken-derived scFv nucleotide sequences A4 and G12 seem to be conserved with sequence identity higher than 85% and few substitutions within the CDRs of V_L and V_H across A4 and G12. A sequence comparision of Chicken derived scFvs A4 and G12 were done against previously reported *Mus musculus* (mouse) anti-Gal mAbs M86, B1H12, B4H2 and Gal3 (Chen et al., 2000) and engineered scFv where light chain was developed using a non-human primate *Macaca mulatta* (Rhesus Macaque)

template and heavy chain from *Homo sapines* (humans) gene template (Kearns-Jonker et al., 2007). Comparison of the chicken-derived scFv sequences G12 and A4 with previously reported engineered scFv from human- (V_H) and monkey-derived (V_L) sequences (Kearns-Jonker et al., 2007) showed that the chicken-derived scFv G12 and A4 had longer sequences than the previous engineered scFvs, as the latter do not have CDR 3 region in either of their V_L and V_H chains (Figure 2.12). The framework regions (FR1 and FR2) at either end of CDR1 and CDR2, respectively, in the V_L of scFvs A4 and G12 showed higher similarity to the corresponding regions of V_L in non human primate-derived scFv sequences, whereas in relation to human-derived scFv V_H sequences the CDR1 (S30, S31 and Y32) were identical to A4 and G12 sequences (Figure 2.13 C). Also, the FR 1 and FR 2 showed high similarity at both ends of CDR 1 and CDR 2, respectively. The comparisons were also drawn from the reported mouse monoclonal M86 anti-Gal-α-(1, 3)-Gal mAbs (Chen et al., 2000). In heavy chain, V_H CDR 1 residues of chicken-derived scFvs G12 and A4 (G26, F27, T28, and F29) and in CDR 2 residues (V64, K65, and G66) were identical with the mouse mAbs (Figure 2.12 D). The framework regions (FR1 and FR2) at either end of CDR1 and CDR2 in scFv G12 and A4, respectively, showed high similarity across the antibody sequences (Figure 2.12 A, B, C, and D, in purple shaded aminoacid). While in the V_L, CDR1 of scFV A4 and G12 (S21 and G23) and CDR 2 (R48 and S50) had amino acid residues identical to that of mouse V_L mAbs (Figure 2.12 B).

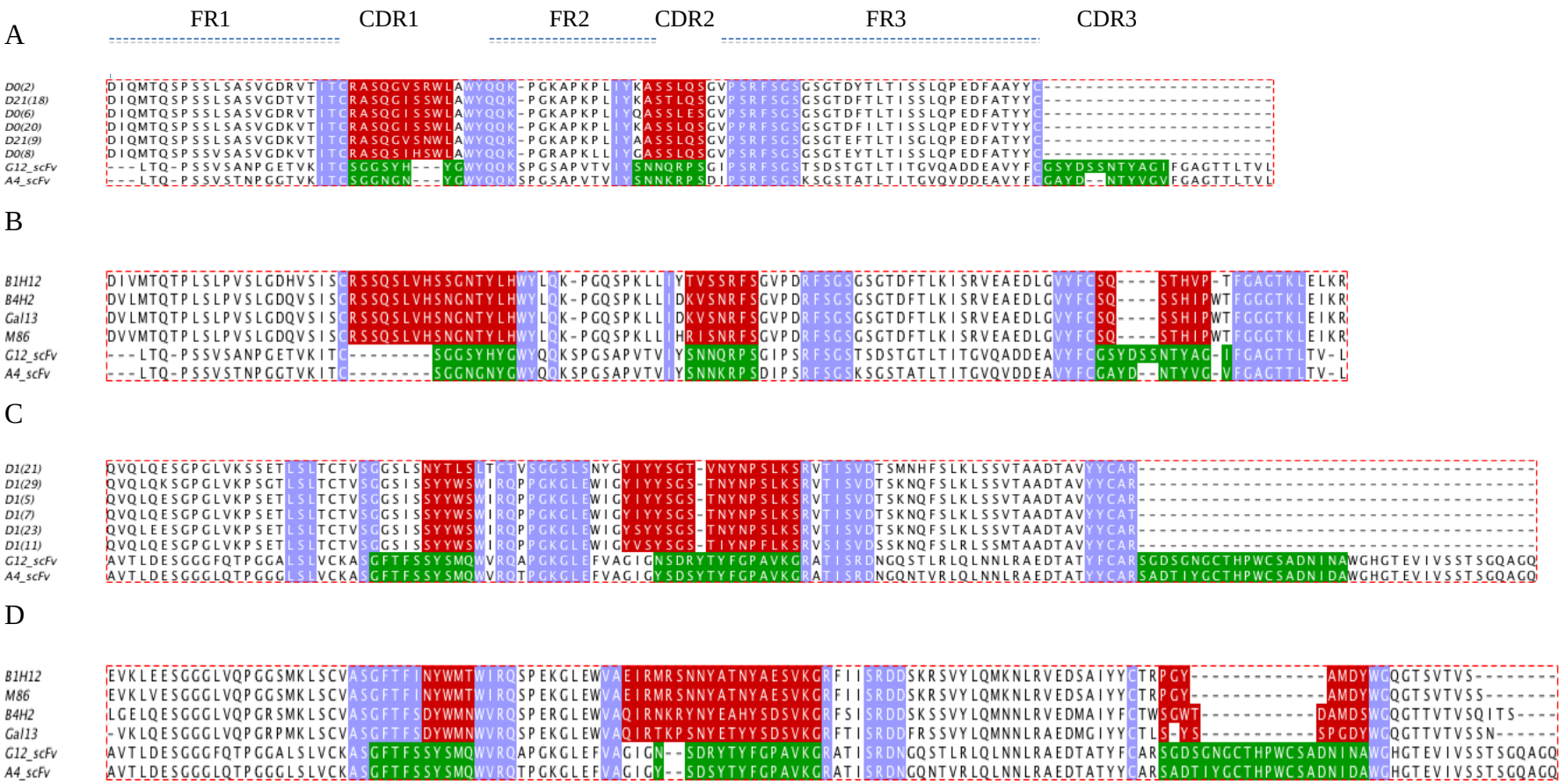


Figure 2.12: Alignment of chicken derived scFvs G12 and A4 against *Mus musculus* (mouse) antibodies, *Homo sapines* (humans) and non human primate *Macaca mulatta* (monkey*)* scFv. The amino acid sequence of the light chains of scFvs G12 and A4 aligned against light chain of (A) *Macaca mulatta* and (B) *Mus musculus*. The amino acid sequence of the heavy chains of chicken scFvs aligned against heavy chain of (C) *Homo sapiens* and (D) *Mus musculus*. The CDRs of the (A) *Macaca mulatta*, (B and D) *Mus musculus* and (C) *Homo sapiens* are shown in red while the CDRs of scFvs G12 and A4 are shown in green. The conserved amino acids in the flanking regions at either end of CDRs are shown in purple.

2.4. Discussion

The non-human carbohydrate epitope, Gal-α-(1, 3)-Gal is commonly found as a terminal motif on different glycoproteins in many species (Macher and Galili, 2008). It has a role in hyperacute rejection (Galili, 2005), delayed anaphylaxis (Commins and Platts-Mills, 2013) and biotherapeutic contamination (Bosques et al., 2010). Chickens are similar to humans in not being able to synthesise the Gal-α-(1, 3)-Gal motif and naturally generate antibodies against the epitope (Cunningham et al., 2012). Previously, an scFv that binds selectively against both bound and free Gal-α-(1, 3)Gal and also with Gal-α-(1, 3)-Gal epitope present on terminal of different glycoproteins has been reported (Cunningham et al., 2012). In this work, scFv characterisation and *in silico* approaches have been described to investigate the structural aspects of the interaction of these scFvs with Gal-α-(1, 3)-Gal.

The chicken-derived scFv sequences are conserved, with sequence similarity greater than 85% and have few amino acid substitutions within the CDRs of V_L and V_H across A4 and G12. Comparison of these scFv sequences with human and non-primate antibody sequences (Kearns-Jonker et al., 2007) showed that the CDR 3 region is missing in both V_L and V_H of the antibodies. V_H CDR 1 residues of the chicken scFvs (G, F, T, and F) (Figure 2.13) were identical to the mouse monoclonal anti-Gal-α-(1, 3)-Gal antibodies (Galili et al., 1998). The amino acids in the CDR 1 region of heavy chain in human and mouse monoclonal antibodies are similar and are important participants in the binding pockets of antibodies againt Gal-α-(1, 3)-Gal (Kearns-Jonker et al., 2007). The other two CDR regions of heavy and light chain of scFv A4 and G12 were not identical, nevertheless they showed high conservation when compared to previously known antibodies. It is known that sequences in the CDRs from different species can be highly variable (Ota and Nei, 1994). Interestingly, residues of the framework region outside CDRs of both light and heavy chain in chicken scFvs showed high similarity with mouse, human and non human primate antibody sequences, suggesting that the outer residues of CDR in the frame work region may play an important role in the binding site formation (Kunik et al., 2012).

Unlike protein-specific antibodies, carbohydrate-specific antibodies tend to exhibit low affinities and this is true for most lectin-carbohydrate interaction which exhibit affinities in the micromolar range (Brooks et al., 2010). SPR analysis demonstrate binding affinities of scFvs A4 and G12 with Gal-α-(1, 3)-Gal in the order of 10^8 M^{-1}, which was significantly higher than the commercially available M86 monoclonal antibody. Most lectin-glycan interactions also show lower binding affinities. Affinities of several lectins (GS-I-B4, *Maclura pomifera* agglutinin (MPA), *Pseudomonas aeruginosa* agglutinin (PA-IL) and EEA) binding with Gal-α-(1, 3)-Gal have been analysed using ELLA (Kirkeby and Moe, 2002). However the lectin receptor affinities were calculated with respect to the ligand, where K_D (nanograms) were in the range of 6.2 – 74.7 (approximately 10^{-8} – 10^{-9} M) therefore affinities of scFvs against Gal-α-(1, 3)-Gal in the order of 10^8 M^{-1} seem to be enough to detect this glycan epitope.

Carbohydrate-protein interactions are important and diversified biological interactions that are involved in molecular and cellular recognition, chronic diseases, immune responses. The investigation of these interactions at molecular level will not only help us in understanding and elucidating the biological role of carbohydrate–protein interaction but also modulating the activity of carbohydrates using scFvs to stimulate the biological environment.

Crystal structures of terminal α-linked Gal structures in complex with various lectins, including GS-I-B4, EEA, MPA and PA-IL (Yuriev et al., 2009) (Tempel et al., 2002, Greco et al., 2006, Natchiar et al., 2006), have demonstrated the importance of groove-type and end-on insertion mechanism of binding in lectin-carbohydrate interactions. These lectins interact primarily with the terminal α-linked Gal residue and are unable to distinguish between specific linkage positions (e.g., these lectins also bind to Gal-α-(1, 2)-R, or Gal-α-(1, 4)-R in addition to Gal-α-(1, 3)-R) (Mo et al., 1999). Thus, these crystal structures may not be representative of specific interactions with Gal-α-(1, 3)-Gal or may not be very comparable to Gal-α-(1, 3)-Gal complexes with whole antibodies or antibody fragments. No X-ray crystal structures have been described for anti-Gal-α-(1, 3)-Gal antibodies in complex with its known target. To our knowledge there are few reports available to map human and mouse antibody interaction with Gal-α-(1, 3)-Gal. However, there is no crystal structure information for chicken scFvs that could be exploited to explore the scFv contact

residues between the αGal-α-(1, 3)-Gal and the scFv. This study is the first to explore the interaction analyses of chicken scFvs against Gal-α-(1, 3)-Gal while making an antigen-antibody complex.

The structural basis of interaction of anti-Gal-α-(1, 3)-Gal scFv antibody fragments was suggested here based upon scFv sequence information. Models have been generated to define unique structural properties and specific sites relevant for optimal interaction of these molecules with their target epitope identified. For scFv modelling, a threading approach was used involving multiple templates for each of the scFv folds (Petrey et al., 2003, Contreras-Moreira et al., 2003). This approach has previously been used successfully to model antibody fragments (Hortiguela and Wall, 2013, Gurr et al., 2013, Xue et al., 2013). ScFv A4 and G12 share an overall sequence similarity of 89% to each other and the optimal templates that were selected in modelling of the framework and CDR regions were different for each scFv except for V_L CDR 2 template: 1Q1J (human monoclonal antibody template). However, both of these scFvs were found structurally similar with the overall RMSD of 1.34 Å and 1.69 Å of superposed light and heavy chains respectively for scFv A4 and G12 (Figure 2.11). For both scFvs, the models of the heavy chain CDR3s showed lowest scores while modelling, which is in agreement with the reports that 3D structures of V_H CDR3 are the most difficult to model using amino acid sequence relationships (Morea et al., 1998, Kuroda et al., 2008).

Docking experiments have demonstrated the affinity and specificity of scFvs for Gal-α-(1, 3)-Gal and similar glycans, including Gal-α-1-R structures. It has been discussed before that variation in molecular sise of ligand with differential number of heavy atoms changes the efficiency of ligand to bind with receptor (Reynolds et al., 2007). Within monosaccharides, α-Gal showed higher affinity than α-Fucose and β-Xylose in both scFv A4 and G12. Among the disaccharides tested it was evident that Gal-α-(1, 3)-Gal binds more favourably than other carbohydrates in both scFvs, which was also supported by ELISA results (Cunningham et al., 2012).

Different poses for scFv-glycan interaction revealed that CDR residues from both heavy and light chain are part of the active sites. CDR3 V_L and CDR2 V_H together formed a pocket and the ligand Gal-α-(1, 3)-Gal was placed well in the groove formed between the two chains CDRs. Similar poses of Gal-α-(1, 3)-Gal interaction

with CDRs were observed in both scFv A4 and G12 (Figure 2.9 A and B). The binding site for both scFvs against Gal-α-(1, 3)-Gal looks like a canyon shape in which many residues were bound to the discaccharide (Figure 2.9 A and B) whereas the binding site of both scFvs against Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc had a crater shape binding site where two trisaccharides bound perpendicular to the binding site (Figure 2.9 C) (Agostino et al., 2009a).

It was observed that water molecules also formed part of the binding site at a distance less that 3.5 Å from the ligand (Figure 2.10 A, B and C), playing an important contribution towards hydrogen bonding bridge between carbohydrates and the scFv residues (Tschampel and Woods, 2003). The carbohydrate-scFv interaction fingerprints were analysed for summarisingrizing the interactions with respect to carbohydrate. The C6 of α-Gal epitopes play an important role in binding to the scFvs, contributing to hydrophobic and van der Waals interactions and O6 to hydrogen bonding as reported in antibodies previously (Figure 2.10 A and B) (Agostino et al., 2010).

Taken all together the modelling studies show the recognition and interaction of reported scFv molecules against the Gal-α-(1, 3)-Gal motif. The key binding regions and properties among the scFv amino acid residues were determined, which could be critical in the recognition and binding to the target epitope. This work also demonstrates the potential of this approach for identifying glycan-binding scFvs and will be utilised to identify lectin mimics and their interaction against other carbohydrate epitopes.

2.5. References

AGOSTINO, M., JENE, C., BOYLE, T., RAMSLAND, P. A. & YURIEV, E. 2009a. Molecular docking of carbohydrate ligands to antibodies: structural validation against crystal structures. *Journal of chemical information and modelling,* 49**,** 2749-2760.

AGOSTINO, M., SANDRIN, M. S., THOMPSON, P. E., YURIEV, E. & RAMSLAND, P. A. 2009b. in silico analysis of antibody-carbohydrate interactions and its application to xenoreactive antibodies. *Molecular immunology,* 47**,** 233-246.

AGOSTINO, M., SANDRIN, M. S., THOMPSON, P. E., YURIEV, E. & RAMSLAND, P. A. 2010. Identification of preferred carbohydrate binding modes in xenoreactive antibodies by combining conformational filters and binding site maps. *Glycobiology,* 20**,** 724-735.

AVILA, J. L., ROJAS, M. & GALILI, U. 1989. Immunogenic Gal-α(1, 3)Gal carbohydrate epitopes are present on pathogenic American Trypanosoma and Leishmania. *The Journal of Immunology,* 142**,** 2828-2834.

BERMAN, H. M., WESTBROOK, J., FENG, Z., GILLILAND, G., BHAT, T., WEISSIG, H., SHINDYALOV, I. N. & BOURNE, P. E. 2000. The protein data bank. *Nucleic acids research,* 28**,** 235-242.

BODER, E. T., MIDELFORT, K. S. & WITTRUP, K. D. 2000. Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proceedings of the National Academy of Sciences,* 97**,** 10701-10705.

BOSQUES, C. J., COLLINS, B. E., MEADOR, J. W., SARVAIYA, H., MURPHY, J. L., DELLORUSSO, G., BULIK, D. A., HSU, I. H., WASHBURN, N., SIPSEY, S. F., MYETTE, J. R., RAMAN, R., SHRIVER, Z., SASISEKHARAN, R. & VENKATARAMAN, G. 2010. Chinese hamster ovary cells can produce galactose-[α]-1, 3-galactose antigens on proteins. *Nat Biotech,* 28**,** 1153-1156.

BOUHOURS, J.-F., RICHARD, C., RUVOEN, N., BARREAU, N., NAULET, J. & BOUHOURS, D. 1998. Characterisation of a polyclonal anti-Ga-α-(1-3)Gal antibody from chicken. *Glycoconjugate journal,* 15**,** 93-99.

BRADFORD, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry,* 72**,** 248-254.

BROOKS, C. L., SCHIETINGER, A., BORISOVA, S. N., KUFER, P., OKON, M., HIRAMA, T., MACKENZIE, C. R., WANG, L.-X., SCHREIBER, H. & EVANS, S. V. 2010. Antibody recognition of a unique tumour-specific glycopeptide antigen. *Proceedings of the National Academy of Sciences,* 107**,** 10056-10061.

CHARLTON, K., HARRIS, W. & PORTER, A. 2001. The isolation of super-sensitive anti-hapten antibodies from combinatorial antibody libraries derived from sheep. *Biosensors and Bioelectronics,* 16**,** 639-646.

CHEN, Z. C., RADIC, M. Z. & GALILI, U. 2000. Genes coding evolutionary novel anti-carbohydrate antibodies: studies on anti-Gal production in α-1, 3galactosyltransferase knock out mice. *Molecular immunology,* 37**,** 455-466.

CHUNG, C. H., MIRAKHUR, B., CHAN, E., LE, Q.-T., BERLIN, J., MORSE, M., MURPHY, B. A., SATINOVER, S. M., HOSEN, J. & MAURO, D. 2008. Cetuximab-induced anaphylaxis and IgE specific for galactose-α-1, 3-galactose. *New England Journal of Medicine,* 358**,** 1109-1117.

COMMINS, S. P. & PLATTS-MILLS, T. A. 2013. Delayed anaphylaxis to red meat in patients with IgE specific for galactose-α-1, 3-galactose (α-gal). *Current allergy and asthma reports,* 13**,** 72-77.

CONTRERAS-MOREIRA, B., FITZJOHN, P. W. & BATES, P. A. 2003. In silico protein recombination: enhancing template and sequence alignment selection for comparative protein modelling. *Journal of molecular biology,* 328**,** 593-608.

CUNNINGHAM, S., STARR, E., SHAW, I., GLAVIN, J., KANE, M. & JOSHI, L. 2012. Development of a Convenient Competitive ELISA for the Detection of the Free and Protein-Bound Nonhuman Galactosyl-α-(1,3)-Galactose Epitope Based on Highly Specific Chicken Single-Chain Antibody Variable-Region Fragments. *Analytical chemistry,* 85**,** 949-955.

DEL CARPIO, C. A., TAKAHASHI, Y. & SASAKI, S. 1993. A new approach to the automatic identification of candidates for ligand receptor sites in proteins:(I) search for pocket regions. *Journal of Molecular Graphics,* 11**,** 23-29.

DEMARCO, M. L. & WOODS, R. J. 2008. Structural glycobiology: a game of snakes and ladders. *Glycobiology,* 18**,** 426-440.

EDELSBRUNNER, H., FACELLO, M., FU, P. & LIANG, J. Measuring proteins and voids in proteins. System Sciences, 1995. Vol. V. Proceedings of the Twenty-Eighth Hawaii International Conference on, 1995. IEEE, 256-264.

FARADY, C. J., EGEA, P. F., SCHNEIDER, E. L., DARRAGH, M. R. & CRAIK, C. S. 2008. Structure of an Fab–protease complex reveals a highly specific non-canonical mechanism of inhibition. *Journal of molecular biology,* 380**,** 351-360.

GALILI, U. 2005. The α-gal epitope and the anti-Gal antibody in xenotransplantation and in cancer immunotherapy. *Immunology and cell biology,* 83**,** 674-686.

GALILI, U., CHEN, Z.-C. & DEGEEST, K. 2003. Expression of α-gal epitopes on ovarian carcinoma membranes to be used as a novel autologous tumour vaccine. *Gynecologic oncology,* 90**,** 100-108.

GALILI, U., LATEMPLE, D. C. & RADIC, M. Z. 1998. A SENSITIVE ASSAY FOR MEASURING [alpha]-GAL EPITOPE EXPRESSION ON CELLS BY A MONOCLONAL ANTI-GAL ANTIBODY 1. *Transplantation,* 65**,** 1129-1132.

GALILI, U., MACHER, B. A., BUEHLER, J. & SHOHET, S. B. 1985. Human natural anti-alpha-galactosyl IgG. II. The specific recognition of α-(1, 3)-linked galactose residues. *The Journal of experimental medicine,* 162**,** 573-582.

GALILI, U., RACHMILEWITZ, E., PELEG, A. & FLECHNER, I. 1984. A unique natural human IgG antibody with anti-alpha-galactosyl specificity. *The Journal of experimental medicine,* 160**,** 1519-1531.

GOLDSTEIN, I. & WINTER, H. 1999. The Griffonia simplicifolia I - B4 Isolectin. *In:* GALILI, U. & AVILA, J. L. (eds.) *α-Gal and Anti-Gal,* 32, 127-141. Springer US. 978-1-4613-7160-1

GRAHN, E. M., WINTER, H. C., TATENO, H., GOLDSTEIN, I. J. & KRENGEL, U. 2009. Structural Characterisation of a Lectin from the Mushroom *Marasmius oreades* in Complex with the Blood Group B Trisaccharide and Calcium. *Journal of molecular biology,* 390**,** 457-466.

GRECO, A., HO, J. G., LIN, S.-J., PALCIC, M. M., RUPNIK, M. & NG, K. K. 2006. Carbohydrate recognition by Clostridium difficile toxin A. *Nature structural & molecular biology,* 13**,** 460-461.

GRIFFITHS, A. D., MALMQVIST, M., MARKS, J., BYE, J., EMBLETON, M., MCCAFFERTY, J., BAIER, M., HOLLIGER, K. P., GORICK, B. & HUGHES-JONES, N. 1993. Human anti-self antibodies with high specificity from phage display libraries. *The EMBO journal,* 12**,** 725.

GURR, W., SHAW, M., HERZOG, R. I., LI, Y. & SHERWIN, R. 2013. Vaccination with Single Chain Antigen Receptors for Islet-Derived Peptides Presented on I-Ag7 Delays Diabetes in NOD Mice by Inducing Anergy in Self-Reactive T-Cells. *PloS one,* 8**,** e69464.

HAGEMEYER, C. E., VON ZUR MUHLEN, C., VON ELVERFELDT, D. & PETER, K. 2009. Single-chain antibodies as diagnostic tools and therapeutic agents. *Thromb Haemost,* 101**,** 1012-1019.

HALGREN, T. A. 1996. Merck molecular force field. I. Basis, form, scope, parameterisation, and performance of MMFF94. *Journal of computational chemistry,* 17**,** 490-519.

HANES, J., SCHAFFITZEL, C., KNAPPIK, A. & PLUCKTHUN, A. 2000. Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nature biotechnology,* 18**,** 1287-1292.

HEIMBURG-MOLINARO, J. & RITTENHOUSE-OLSON, K. 2009. Development and characterisation of antibodies to carbohydrate antigens. *Glycomics*, 341-357. Springer. 1588297748

HENDLICH, M., RIPPMANN, F. & BARNICKEL, G. 1997. LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *Journal of molecular graphics & modelling,* 15**,** 359.

HIGGINS, E. 2010. Carbohydrate analysis throughout the development of a protein therapeutic. *Glycoconjugate journal,* 27**,** 211-225.

HIRABAYASHI, J. 2008. Concept, strategy and realisation of lectin-based glycan profiling. *Journal of biochemistry,* 144**,** 139-147.

HORTIGUELA, M. J. & WALL, J. G. 2013. Improved Detection of Domoic Acid Using Covalently Immobilised Antibody Fragments. *Marine drugs,* 11**,** 881-895.

JOZIASSE, D. & ORIOL, R. 1999. Xenotransplantation: the importance of the Gal α-(1, 3)Gal epitope in hyperacute vascular rejection. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease,* 1455**,** 403-418.

JUNG, S., SPINELLI, S., SCHIMMELE, B., HONEGGER, A., PUGLIESE, L., CAMBILLAU, C. & PLUCKTHUN, A. 2001. The Importance of Framework Residues H6, H7 and H10 in Antibody Heavy Chains: Experimental Evidence for a New Structural Subclassification of Antibody VH Domains. *Journal of molecular biology,* 309**,** 701-716.

KEARNS-JONKER, M., BARTENEVA, N., MENCEL, R., HUSSAIN, N., SHULKIN, I., XU, A., YEW, M. & CRAMER, D. V. 2007. Use of molecular modelling and site-directed mutagenesis to define the structural basis for the immune response to carbohydrate xenoantigens. *BMC immunology,* 8**,** 3.

KIRKEBY, S. & MOE, D. 2002. Lectin interactions with α-galactosylated xenoantigens. *Xenotransplantation,* 9**,** 260-267.

KUNIK, V., PETERS, B. & OFRAN, Y. 2012. Structural consensus among antibodies defines the antigen binding site. *PLoS computational biology,* 8**,** e1002388.

KURODA, D., SHIRAI, H., KOBORI, M. & NAKAMURA, H. 2008. Structural classification of CDR-H3 revisited: A lesson in antibody modelling. *Proteins: Structure, Function, and Bioinformatics,* 73**,** 608-620.

LABUTE, P. 2008. Protonate3D: Assignment of ionisation states and hydrogen coordinates to macromolecular structures. *Proteins: Structure, Function, and Bioinformatics,* 75**,** 187-205.

LADJEMI, M. Z., CHARDES, T., CORGNAC, S., GARAMBOIS, V., MORISSEAU, S., ROBERT, B., BASCOUL-MOLLEVI, C., AIT ARSA, I., JACOT, W. & POUGET, J.-P. 2011. Vaccination with human anti-trastuzumab anti-idiotype scFv reverses HER2 immunological tolerance and induces tumour immunity in MMTV. f. huHER2 (Fo5) mice. *Breast Cancer Res,* 13**,** R17.

LARKIN, M., BLACKSHIELDS, G., BROWN, N., CHENNA, R., MCGETTIGAN, P., MCWILLIAM, H., VALENTIN, F., WALLACE, I., WILM, A. & LOPEZ, R. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics,* 23**,** 2947-2948.

LYU, M., KURZROCK, R. & ROSENBLUM, M. G. 2008. The immunocytokine scFv23/TNF targeting HER-2/neu induces synergistic cytotoxic effects with 5-fluorouracil in TNF-resistant pancreatic cancer cell lines. *Biochemical pharmacology,* 75**,** 836-846.

MACHER, B. A. & GALILI, U. 2008. The Galα-(1, 3)Galβ-(1, 4)GlcNAc-R (α-Gal) epitope: a carbohydrate of unique evolution and clinical relevance. *Biochimica et biophysica acta,* 1780**,** 75.

MAEDA, E., KITA, S., KINOSHITA, M., URAKAMI, K., HAYAKAWA, T. & KAKEHI, K. 2012. Analysis of nonhuman N-glycans as the minor constituents in recombinant monoclonal antibody pharmaceuticals. *Analytical chemistry,* 84**,** 2373-2379.

MAJUMDAR, R., RAILKAR, R. & DIGHE, R. R. 2011. Docking and free energy simulations to predict conformational domains involved in hCG-LH receptor interactions using recombinant antibodies. *Proteins: Structure, Function, and Bioinformatics,* 79**,** 3108-3122.

MCCOY JR, J. P., VARANI, J. & GOLDSTEIN, I. J. 1983. Enzyme-linked lectin assay (ELLA): Use of alkaline phosphatase-conjugated *Griffonia simplicifolia B4* isolectin for the detection of α-d-galactopyranosyl end groups. *Analytical biochemistry,* 130**,** 437-444.

MILLAND, J., YURIEV, E., XING, P.-X., MCKENZIE, I. F., RAMSLAND, P. A. & SANDRIN, M. S. 2007. Carbohydrate residues downstream of the terminal Galα-(1, 3) Gal epitope modulate the specificity of xenoreactive antibodies. *Immunology and cell biology,* 85**,** 623-632.

MO, H., RICE, K. G., EVERS, D. L., WINTER, H. C., PEUMANS, W. J., VAN DAMME, E. J. & GOLDSTEIN, I. J. 1999. Xanthosoma sagittifolium tubers contain a lectin with two different types of carbohydrate-binding sites. *Journal of Biological Chemistry,* 274**,** 33300-33305.

MOE, C. C. G. 2010.10. http://www.chemcomp.com.

MOREA, V., TRAMONTANO, A., RUSTICI, M., CHOTHIA, C. & LESK, A. M. 1998. Conformations of the third hypervariable region in the VH domain of immunoglobulins. *Journal of molecular biology,* 275**,** 269-294.

MUEGGE, I. & MARTIN, Y. C. 1999. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *Journal of medicinal chemistry,* 42**,** 791-804.

NASO, F., GANDAGLIA, A., COMACCHIO, G., SPINA, M. & GEROSA, G. 2012. AlphaGal Residue in Xenograft Heart Valve Bioprostheses and Tissue Engineered Construct. *QScience Proceedings*.

NASO, F., GANDAGLIA, A., IOP, L., SPINA, M. & GEROSA, G. 2011. First quantitative assay of alpha-Gal in soft tissues: presence and distribution of the epitope before and after cell removal from xenogeneic heart valves. *Acta Biomaterialia,* 7**,** 1728-1734.

NATCHIAR, S. K., SRINIVAS, O., MITRA, N., SUROLIA, A., JAYARAMAN, N. & VIJAYAN, M. 2006. Structural studies on peanut lectin complexed with disaccharides involving different linkages: further insights into the structure and interactions of the lectin. *Acta Crystallographica Section D: Biological Crystallography,* 62**,** 1413-1421.

NOZAWA, S., XING, P.-X., WU, G. D., GOCHI, E., KEARNS-JONKER, M., SWENSSON, J., STARNES, V. A., SANDRIN, M. S., MCKENZIE, I. F. & CRAMER, D. V. 2001. Characteristics of immunoglobin gene usage of the xenoantibody binding to Gal-α-(1, 3)-Gal target antigens in the Gal knockout mouse 1. *Transplantation,* 72**,** 147-155.

OTA, T. & NEI, M. 1994. Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. *Molecular Biology and Evolution,* 11**,** 469-482.

PETREY, D., XIANG, Z., TANG, C. L., XIE, L., GIMPELEV, M., MITROS, T., SOTO, C. S., GOLDSMITH- FISCHMAN, S., KERNYTSKY, A. & SCHLESSINGER, A. 2003. Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modelling. *Proteins: Structure, Function, and Bioinformatics,* 53**,** 430-435.

PLUM, M., MICHEL, Y., WALLACH, K., RAIBER, T., BLANK, S., BANTLEON, F. I., DIETHERS, A., GREUNKE, K., BRAREN, I. & HACKL, T. 2011. Close-up of the Immunogenic α-1, 3-Galactose Epitope as Defined by a Monoclonal Chimeric Immunoglobulin E and Human Serum Using Saturation Transfer Difference (STD) NMR. *Journal of Biological Chemistry,* 286**,** 43103-43111.

POINTREAU, Y., COMMINS, S. P., CALAIS, G., WATIER, H. & PLATTS-MILLS, T. A. 2012. Fatal Infusion Reactions to Cetuximab: Role of Immunoglobulin E‚ÄìMediated Anaphylaxis. *Journal of Clinical Oncology,* 30**,** 334-334.

RAHBARIZADEH, F., RASAEE, M., FOROUZANDEH MOGHADAM, M., ALLAMEH, A. & SADRODDINY, E. 2004. Production of novel recombinant single-domain antibodies against tandem repeat region of MUC1 mucin. *Hybridoma and hybridomics,* 23**,** 151-159.

RAREY, M., WEFING, S. & LENGAUER, T. 1996. Placement of medium-sised molecular fragments into active sites of proteins. *Journal of computer-aided molecular design,* 10**,** 41-54.

REYNOLDS, C. H., BEMBENEK, S. D. & TOUNGE, B. A. 2007. The role of molecular sise in ligand efficiency. *Bioorganic & medicinal chemistry letters,* 17**,** 4258-4261.

SANDRIN, M. S., VAUGHAN, H. A., DABKOWSKI, P. L. & MCKENZIE, I. 1993. Anti-pig IgM antibodies in human serum react predominantly with Gal (alpha 1-3) Gal epitopes. *Proceedings of the National Academy of Sciences,* 90**,** 11391-11395.

STANFIELD, R. L., GORNY, M. K., WILLIAMS, C., ZOLLA-PAZNER, S. & WILSON, I. A. 2004. Structural rationale for the broad neutralisation of HIV-1 by human monoclonal antibody 447-52D. *Structure,* 12**,** 193-204.

STRONG, R. K., CAMPBELL, R., ROSE, D. R., PETSKO, G. A., SHARON, J. & MARGOLIES, M. N. 1991. Three-dimensional structure of murine anti-p-azophenylarsonate Fab 36-71. 1. X-ray crystallography, site-directed mutagenesis, and modelling of the complex with hapten. *Biochemistry,* 30**,** 3739-3748.

TATENO, H., NAKAMURA-TSURUTA, S. & HIRABAYASHI, J. 2007. Frontal affinity chromatography: sugar-protein interactions. *Nature protocols,* 2**,** 2529-2537.

TEMPEL, W., TSCHAMPEL, S. & WOODS, R. J. 2002. The Xenograft Antigen Bound to Griffonia simplicifolia Lectin 1-B4 X-RAY CRYSTAL STRUCTURE OF THE COMPLEX AND MOLECULAR DYNAMICS CHARACTERISATION OF THE BINDING SITE. *Journal of Biological Chemistry,* 277**,** 6615-6621.

TENEBERG, S., ALSEN, B., ANGSTROM, J., WINTER, H. C. & GOLDSTEIN, I. J. 2003. Studies on Gal α-1, 3-binding proteins: comparison of the glycosphingolipid binding specificities of Marasmius oreades lectin and Euonymus europaeus lectin. *Glycobiology,* 13**,** 479-486.

TEPLYAKOV, A., OBMOLOVA, G., WU, S.-J., LUO, J., KANG, J., O'NEIL, K. & GILLILAND, G. L. 2009. Epitope mapping of anti-interleukin-13 neutralizing antibody CNTO607. *Journal of molecular biology*, 389**,** 115-123.

TOWBIN, H., ROSENFELDER, G., WIESLANDER, J., AVILA, J., ROJAS, M., SZARFMAN, A., ESSER, K., NOWACK, H. & TIMPL, R. 1987. Circulating antibodies to mouse laminin in Chagas disease, American cutaneous leishmaniasis, and normal individuals recognise terminal galactosyl (alpha 1-3)-galactose epitopes. *The Journal of experimental medicine*, 166**,** 419-432.

TSCHAMPEL, S. M. & WOODS, R. J. 2003. Quantifying the role of water in protein-carbohydrate interactions. *The Journal of Physical Chemistry A*, 107**,** 9175-9181.

VALLEJO, M. C., MATSUO, A. L., GANIKO, L., MEDEIROS, L. C. S., MIRANDA, K., SILVA, L. S., FREYM√ºLLER-HAAPALAINEN, E., SINIGAGLIA-COIMBRA, R., ALMEIDA, I. C. & PUCCIA, R. 2011. The Pathogenic Fungus Paracoccidioides brasiliensis Exports Extracellular Vesicles Containing Highly Immunogenic α-Galactosyl Epitopes. *Eukaryotic cell*, 10**,** 343-351.

VUKOVIC, P., CHEN, K., QIN LIU, X., FOLEY, M., BOYD, A., KASLOW, D. & GOOD, M. F. 2002. Single-chain antibodies produced by phage display against the C-terminal 19 kDa region of merozoite surface protein-1 of *Plasmodium yoelii* reduce parasite growth following challenge. *Vaccine*, 20**,** 2826-2835.

WILKINSON, I. C., HALL, C. J., VEVERKA, V., SHI, J. Y., MUSKETT, F. W., STEPHENS, P. E., TAYLOR, R. J., HENRY, A. J. & CARR, M. D. 2009. High Resolution NMR-based Model for the Structure of a scFv-IL-1 Complex POTENTIAL FOR NMR AS A KEY TOOL IN THERAPEUTIC ANTIBODY DESIGN AND DEVELOPMENT. *Journal of Biological Chemistry*, 284**,** 31928-31935.

WOJCIECHOWSKI, M. & LESYNG, B. 2004. Generalised Born model: Analysis, refinement, and applications to proteins. *The Journal of Physical Chemistry B*, 108**,** 18368-18376.

XUE, S., LI, H. Ä., ZHANG, J. Ä., LIU, J. Ä., HU, Z.-Q., GONG, A.-D., HUANG, T. & LIAO, Y.-C. 2013. A chicken single-chain antibody fused to alkaline phosphatase detects Aspergillus pathogens and their presence in natural samples by direct sandwich ELISA. *Analytical chemistry*.

YAGI, Y., KAKEHI, K., HAYAKAWA, T., OHYAMA, Y. & SUZUKI, S. 2012. Specific detection of N-glycolylneuraminic acid and Gal alpha1-3Gal epitopes of therapeutic antibodies by partial-filling capillary electrophoresis. *Analytical biochemistry*.

YOSHIDA, S., KOBAYASHI, T., MATSUOKA, H., SEKI, C., GOSNELL, W. L., CHANG, S. P. & ISHII, A. 2003. T-cell activation and cytokine production via a bispecific single-chain antibody fragment targeted to blood-stage malaria parasites. *Blood,* 101**,** 2300-2306.

YURIEV, E., AGOSTINO, M., FARRUGIA, W., CHRISTIANSEN, D., SANDRIN, M. S. & RAMSLAND, P. A. 2009. Structural biology of carbohydrate xenoantigens.

Z KONAKCI, K., BOHLE, B., BLUMER, R., HOETZENECKER, W., ROTH, G., MOSER, B., BOLTZ-NITULESCU, G., GORLITZER, M., KLEPETKO, W. & WOLNER, E. 2005. Alpha-Gal on bioprostheses: xenograft immune response in cardiac surgery. *European journal of clinical investigation,* 35**,** 17-23.

ZDANOV, A., LI, Y., BUNDLE, D. R., DENG, S.-J., MACKENZIE, C. R., NARANG, S. A., YOUNG, N. M. & CYGLER, M. 1994. Structure of a single-chain antibody variable domain (Fv) fragment complexed with a carbohydrate antigen at 1.7-A resolution. *Proceedings of the National Academy of Sciences,* 91**,** 6423-6427.

Chapter-3: Screening and selection of DNA aptamers against Neu5Gc and Neu5Ac

3.1. Introduction

As presented in detail in Chapter 1, sialic acids are a family of 9-carbon sugars that are present at the terminal ends of many glycans of glycoproteins and glycolipids. They act as a regulator of molecular and cellular interactions (Schauer, 2009), and also play a critical role in immune regulation (Varki, 2007). They act as a binding site for various pathogens including *Helicobacter pylori*, *Campylobacter jejuni*, mycoplasma, rotaviruses, SV40 virus and corona viruses (Varki, 2008). Aberrant sialylation of proteins and glycans on the cell surface is often reported in cancer cells, also, the over and under expression of sialylated glycans have been shown to induce metastasis of myeloma in humans (Lu et al., 2011, Samraj et al., 2014, Ma et al., 2014).

Correct sialylation of drugs is known to be important for drug stability, increasing half-life and pharmacodynamics of drug molecules *in vivo* (Gregoriadis et al., 2005, Kilcoyne and Joshi, 2007, Byrne et al., 2007, Varki, 2008). However, the sialylation of recombinant glycoprotein therapeutics expressed in non-human expression systems such as Chinese Hamster Ovary (CHO) cells may result in the addition of the Neu5Gc epitope (e.g. erythropoietin) (Hokke et al., 1990, Noguchi et al., 1996). Recent work has shown that the contamination of biotherapeutics with this epitope can result in an immunogenic response such as chronic inflammation and enhanced angiogenesis (Hedlund et al., 2008) as e.g. occurred with Erbitux (Cetuximab), a recombinant monoclonal antibody preparation produced in a murine myeloma cell line, and Vectibix (Panitumumab) a full human antibody produced in CHO cells (Ghaderi et al., 2010, Ghaderi et al., 2012).

The two most abundant forms of sialic acid in mammals are *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc). Neu5Ac is the predominant form in humans and can act as a ligand to the receptors promoting cell-cell adhesion in viral infection including influenza H1N1, H5N1 and H3N2 infection (Yao et al., 2008, Kumari et al., 2007). Also, the *Plasmodium falciparum* parasite which is one of the causative agents of malaria in humans recognises Neu5Ac on erythrocytes (Orlandi et al., 1992). In most other non-human mammals, Neu5Gc is also widely expressed in addition to Neu5Ac.

Normal human serum or other normal human tissues do not contain Neu5Gc on its surface and there are naturally occurring antibodies in humans, which on interaction with exogenous Neu5Gc creates an immune response. There is evidence that Neu5Gc-containing glycoconjugates occur frequently in the sera of cancer patients and on the surface of cancerous tissues (Taylor et al., 2010). Following studies on a Neu5Gc-deficient mouse model system, it has been proposed that the interaction of tumour-associated Neu5Gc with human anti-Neu5Gc antibodies is involved in tumour progression via chronic inflammation (Padler-Karavani et al., 2011, Malykh et al., 2001). This epitope has preference for pathogen binding such as subtilase cytotoxin (SubAB) produced in *E. coli*, which is highly toxic to eukaryotic cells causing hemolytic-uremic syndrome, which has a preference for α2-3-linked Neu5Gc present on the endothelium layer of host cells compared to α2-3-linked Neu5Ac (Lofling et al., 2009).

From the above, it can be seen that there is a need for a convenient method of analysis of Neu5Ac and Neu5Gc to facilitate studies aimed at understanding their different biological roles and actions and for characterisation of biotherapeutics. The current analysis methods for determination of Neu5Ac and Neu5Gc are High Performance Liquid Chromatography (HPLC) and Mass Spectrometry (MS) (Tousi et al., 2011, Zaia, 2008). However, these methods are time consuming, and complex sample purification is required prior to analysis. MS cannot be used for absolute quantification of glycans, also it would be difficult to analyse unknown glycans due to unavailability of quality standards (Sloth et al., 2003). The detection of oligosaccharides using HPLC and MS are in a template based sequential manner, where each linkage is digested excessively to convert them into simpler glycans, requiring much longer time to prepare samples. Also, HPLC and MS are complicated instruments that require personnel expertise to run these analytical machines and to interpret its data. Binding assays are preferred methods over conventional HPLC and MS as they give rapid results and are more convenient to carry out these assays than HPLC and MS with minimal directions and relatively easier to interpret results for analytical and diagnostic purposes. For binding assays, recognition molecules are required that can bind and specifically detect sialic acids.

Lectins are natural proteins originating primarily from plant and animal sources. Specific lectins exist which attach to or indicate the presence of sialic acids. Their specificities for sialic acid linkages were discussed in Chapter 1 (Varki, 1997). However, none of these lectins show an ability to specifically distinguish Neu5Ac from Neu5Gc. Siglecs (sialic-acid-binding immunoglobulin type lectins) are other natural molecules that recognise sialic acids present in mammals, which are involved in cell–cell interactions and signalling functions in the haemopoietic, immune and nervous systems (Crocker, 2002, O'Reilly and Paulson, 2009). The ancestral Siglecs 7 and 9 did bind preferentially to Neu5Gc, however, with the evolutionary loss of Neu5Gc, human siglecs now show more pronounced binding to Neu5Ac, as compared to chimpanzee siglecs which bind preferably to Neu5Gc (Sonnenburg et al., 2004, Angata and Varki, 2000). Lectins are one of the known glycan binding proteins that have been in use for many diagnostics application, however their broad specificities against sialic acids make them an ineffective reagent to distinguish two forms of sialic acid Neu5Ac and Neu5Gc.

Naturally occurring anti-Neu5Gc antibodies circulate at variable concentrations in the human serum within a population. However, the levels of Neu5Gc antibodies also vary within individuals (Padler-Karavani et al., 2008).

Polyclonal I_gY antibodies, raised in chickens against Neu5Gc were isolated and an ELISA assay was developed that was able to detect Neu5Gc in recombinant therapeutic glycoproteins (Diaz et al., 2009, Ghaderi et al., 2010). However, polyclonal antibodies have its own limitations: they can produce large amounts of non-specific antibodies, and they can bind to multiple epitopes on a single antigen.

Monoclonal antibodies are more advantageous than polyclonal antibody. They have more specific responses than polyclonal antibodies based on their homogeneity and specificity against single targets, but there are no reports to date of a monoclonal anti-Neu5Gc antibody. Also, the antibodies which are described binding to sialic acid do not show specificity to Neu5Ac and they do not discriminate between the two forms of sialic acid: Neu5Ac and Neu5Gc (Cho et al., 2013). There exists a need for the generation of specific affinity agents for Neu5Ac and Neu5Gc, which are small, stable, and deliverable (in biological systems) against these sialic acids for both diagnostic and potential downstream therapeutic applications.

Aptamers are small DNA or RNA oligonucleotides that can specifically bind target molecules with high specificity and affinity (Hermann and Patel, 2000, Stoltenburg et al., 2007). The use of aptamers instead as an alternative to lectins is desirable as many lectins from plant sources are toxic while aptamers are synthetic nucleic acid molecules that do not show any toxicity or immunogenicity in mammals (Brody and Gold, 2000, Min et al., 2011). The cost of production of most of the bulk lectins is much higher than synthetic aptamers and generally the isolation of lectin from natural sources leads to inconsistent activity and availability and differences in their purification that may not be optimal for clinical applications (Hsu et al., 2008). Aptamers have several advantages over antibodies, as aptamers are smaller in sise, they do not require an animal host for generation, and are easy to synthesise (Jayasena, 1999). Being a synthetic molecule, DNA aptamers are easily modifiable with robust structural restoration under ambient conditions (Kim et al., 2010, Kusser, 2000), and can be highly purified and scaled up for bulk production. A major advantage of the use of aptamers is their very high selectivity and specificity, with as low as sub-nanomolar dissociation constants reported as against VEGF and hemagglutinin of human influenza virus H3N2 (Jellinek et al., 1994, Ruckman et al., 1998, Misono and Kumar, 2005). The affinities of aptamers are equal to or better than lectins, which usually have an affinity in micromolar range, and bind non-specifically to closely related glycans (DeMarco and Woods, 2008, Brooks et al., 2010). These properties, combined with a relatively uncomplicated and automated synthetic procedure, make aptamers a highly potent and promising commercially viable reagent.

They are traditionally selected using the *in vitro* SELEX (Systematic Evolution of Ligands by EXponential enrichment) process, that was simultaneously developed by the Gold and Szostak laboratories (Tuerk and Gold, 1990, Ellington and Szostak, 1990). The SELEX process allow selection of specific nucleic acid aptamers from diverse pools of randomised nucleic acids containing typically greater than 10^{12} unique sequences, by successive rounds of selection and re-amplification. SELEX involves the enrichment of desired aptamers by incubating the aptamer pool or library with the target, separating bound from unbound aptamers and followed by enzymatic reamplification of the bound fraction, leading to successive enrichment of binders (recognition molecule).

Most early aptamers were RNA, because RNA could form more complex structures when folded and therefore was expected to be more likely to provide desired recognition agents. However, because of the prevalence of RNAses, RNA aptamers need to undergo modifications to provide the necessary stability for use with serum or other biological samples. DNA aptamers have become more popular recently possibly due to the easier selection process, with no reverse transcription step, and their relative stability, and good DNA binders have been identified.

DNA aptamers are stable, chemically flexible, and have gained significant attention for their uses as a source of recognition elements in biosensor and assay development, diagnostics and also as possible therapeutic agents (Song et al., 2008, Brody and Gold, 2000). They have been described against carbohydrate targets, including Sialyllactose (Mehedi Masud et al., 2004), Sialyl Lewis^{x} (Jeong et al., 2001), Cellobiose (Glc-β-(1→4)-Glc) (Yang et al., 1998), and dextran (repeating α-(1→6) Glucosidic bonds), confirming the feasibility of this approach for the generation of glycan specific receptors.

Here, the development of DNA aptamers against two sialic acids is described. These DNA aptamers, which bind with high affinity and specificity to Neu5Gc and Neu5Ac, were obtained using a SELEX approach.

3.2. Materials and Methods

3.2.1. Materials

Biotinylated spacer (sp-biotin), biotinylated Neu5Ac (Neu5Ac-sp-biotin) and biotinylated Neu5Gc (Neu5Gc-sp-biotin) were purchased from GlycoTech Inc. All carbohydrates were purchased from Glycotech (USA), Dextra (UK) and IsoSep (Sweden). MyOne streptavidin C1 Dynabeads was purchased from Life technologies, UK. Topo TA cloning kit was purchased from Invitrogen, UK, and streptavidin from *Streptomyces avidinii* were purchased from Sigma-Aldrich (St. Louis, MO, USA). Streptavidin was dissolved in 10 mM Tris-HCl buffer (pH 7.5). The Quant-iT OliGreen ssDNA Assay Kit was purchased from Invitrogen (Life Technologies). Taq polymerase, PCR buffer, 25 bp DNA ladder and dNTP were all obtained from Promega, UK (Thermo Scientific). Qiagen II DNA extraction kit was purchased from Qiagen, UK. 10% TBE-urea Gel 1.0mm * 10 well were purchased from, Invitrogen, UK. Luria-Bertani (LB) broth media, Isopropyl beta-D-thiogalactopyranoside (IPTG), 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-Gal) and Kanamycin were purchased from Sigma, UK. Nuclease free water was purchased from Promega, UK. All solutions were prepared with deionised water (ELGA water purification system, ELGA, London, UK).

3.2.2. Random library and primers

The random DNA library (5'- GCGCGGATCCCGCGC-N30-GCGCGAAGCTTGCGC-3') contained a central randomised sequence of 30 nucleotides (N30) flanked by 15 nucleotide constant regions. The library was synthesised and purchased by Sigma-Genosys (Cambridge, UK) that was HPLC purified with a melting temperature of 89.6 °C. The set of primers used for the library amplification, were synthesised and HPLC purified by Eurofins MWG-Operon (Ebersberg, Germany). The reverse primer was: 5'-GCGCAAGCTTCGCGC-3' and the biotinylated (btn) forward primer, where biotinylation was at 5' end: 5'-[btn]GCGCGGATCCCGCGC-3'. These primers were HPLC purified and possess a melting temperature of 63 °C. The reason to use biotinylated forward primers is to allow biotinylation of one strand of the aptamer after amplification. The biotin molecule on the biotinylated aptamer strand could be

easily immobilised onto streptavidin-coated Dynabeads. The ssDNA library (10 nM) and forward and reverse primers (100 μM) were dissolved in nuclease free water and stored at -20 °C.

3.2.3. DNA amplification and conversion into ssDNA

Once resuspended, the synthesised DNA library (10 nM) was amplified by PCR using a PCR reaction mix including 0.2 μM dNTPs, 0.5 μM reverse primer, 0.5 μM forward primer, 5X Phusion buffer (Sigma), 0.2 μL Phusion DNA polymerase, 2 μL ssDNA template and nuclease free water to make a final volume 20 μL. The PCR conditions used were: initial denaturation of 3 min at 94 °C, then 25 cycles of 40 s at 94 °C, 40 s at 55 °C, and 20 s at 72 °C and final extension of 5 min at 72 °C with sample cooling at 4 °C for 15 minutes (Lou et al., 2009, Shangguan et al., 2006). The biotinylated forward primer was used in PCR reactions for the synthesis of biotinylated dsDNA. After amplification, the biotinylated dsDNA was incubated with streptavidin (1:10 molar ratio) for 1 h at room temperature (Pagratis, 1996). Non-biotinylated ssDNA strands were separated using urea-PAGE electrophoresis and extracted using a Phenol-Chloroform extraction of ssDNA (Sambrook and Russell, 2006).

3.2.4. Gel Preparation

2.5 % Agarose gel and 7 M urea gel with 8% Polyacrylamide gel were used for DNA detection. Agarose gels were used for ds DNA detection while urea gels were used for ssDNA detection.

3.2.4.1. Agarose gel preparation and electrophoresis

For preparation of 2.5% agarose gels, 0.75 gram of agarose, 30 mL of 1X TAE buffer (40mM Tris-acetate, 1mM EDTA pH 8.2), and 3 μL (1:10,000) of SYBR Safe DNA Gel Stain (Life Technologies). For Gel running, 1X TAE buffer was used as running buffer at 45V for one hour.

3.2.4.2. Urea Gel preparation and electrophoresis

7M urea/8% Polyacrylamide gels were prepared by combining 2.67 mL of Protogel (National Diagnostics, UK) with 4.2 gm of urea (Sigma, UK) and 2 mL of 5x TBE

buffer. Distilled water was added to a final volume 9.89 mL and filtered with a 0.45-μM filter membrane. After filtration, 100 μL of 10% (10 mg/100 mL) of Ammonium per sulphate (APS) with 10 μL of N,N,N',N'-Tetramethylethylenediamine (TEMED) were added in the gel mixture. The gel was run using 1 X TBE buffer (89mM Tris-HCl, 89mM Boric Acid and 2mM EDTA at pH 8.2) at 180 volts for 30 min.

3.2.5. DNA extraction using Qiagen kit

DNA bands were excised from the gel with a clean sharp scalpel. Excised gels were weighed and 6 volumes of QIAEX I buffer to 1 volume of gel was added in each gel tube. QIAEX II beads were resuspended and vortexed for 30 s, 10 μL of QIAEX II beads were added in each tube. The mixture was incubated for 10 min to solubilise the agarose and bind DNA, the mixture was vortexed every 2 min to keep QIAEX II beads in suspension. Further, the mixture samples were centrifuged 2000 *x* g for 30 s and supernatant was carefully removed. The pellets were initially washed with 500 μL of QIAEX I buffer and then washed twice with 500 μL of PE1 buffer. The pellets were air dried until they turned white. For elution of DNA, 20 μL of 10 mM Tris HCL pH 8.5 was added to resuspend the pellet by vortexing. DNA fragments in suspension were incubated for 5 min at room temperature. The suspension was centrifuged at 2000 *x* g for 30 s, and supernatant was carefully pipetted out in a clean tube containing supernatant with clean purified DNA.

3.2.6. Phenol-Chloroform Extraction of ssDNA

ssDNA bands from the urea gel separation were excised and diffused at twice volume of gel weight in 0.5 M Ammonium Acetate, 10 mM Magnesium Acetate, 1mM EDTA, 0.1% SDS with pH 8.0 (Diffusion buffer) at 50 °C for 30 min. The bands were vortexed and the ssDNA containing supernatant was removed. An equal volume of phenol/chloroform/isoamylol (25:24:1) was mixed with the ssDNA supernatant and centrifuged for 2 min at 2000 *x* g to recover the supernatant. To the supernatant, 2.5 volumes of ice-cold absolute ethanol and 0.1 volume of 3M NaCl was added for at least 2 hours at -20 °C to allow DNA precipitation The mixture was centrifuged at 16595 *x* g for 30 min at 4 °C and washed with 90% ethanol at 16595 *x* g for 10 min to recover the pellet after removing supernatant. The pellet was dried at

37 °C for 15 min and later suspended in 10mM Tris HCl buffer pH 7.5 and stored at -20 °C.

3.2.7. *in vitro* selection of aptamers

An indirect capture method was used for screening of aptamers against Neu5Gc-spacer-biotin (Neu5Gc-sp-biotin) and Neu5Ac-spacer-biotin (Neu5Ac-sp-biotin). Initially, a counter selection against spacer biotin (sp-biotin, where the sp-biotin is $HOCH_2(HOCH)_4CH_2NH$-Biotin) was carried out. The ssDNA pool of 400 ng in PBS + 5 mM $MgCl_2$, pH 7.4 (binding buffer) (Where PBS was prepared with 10 mM phosphate buffer, 2.7 mM potassium chloride and 137 mM sodium chloride, pH 7.4), was heated at 95 °C for 5 min, then quickly cooled down on ice for 10 min and subsequently incubated at room temperature for 10 min, prior to adding to the soluble target molecule. The binding reaction mixture (500 µL) contained binding buffer (pH 7.4), 400 ng ssDNA pools and 100 pmol of sp-biotin. The binding reaction mixture was incubated at 37 °C for 1 hr. After the binding reaction sp-biotin was immobilised on 250 µg of MyOne streptavidin C1 dynabeads at room temperature for 60 min. The beads coated with sp-biotin-aptamer complex were separated on a magnetic separator and unbound ssDNA was desalted and concentrated by phenol-chloroform extraction. This counter selected ssDNA was amplified by PCR and converted into ssDNA pool for aptamer selection against Neu5Gc-sp-biotin. The selection protocol was as described above except that the Neu5Gc-sp-biotin bound ssDNA was eluted twice using nuclease free water at 95 °C for 5 min (as in the negative selection only unbound ssDNA was used for successive selection rounds). The eluted ssDNA was desalted, amplified by PCR), converted into another ssDNA pool, and used for subsequent SELEX rounds. The stringency of binding conditions during aptamer (Table 3.1 and 3.2) selection was enhanced gradually by decreasing target (Neu5Gc) concentration (from 100 to 50 pmol) and interaction time (from 60 to 25 min), and increasing salt concentration in binding buffer (from 137 to 157 mM) and the number of washing steps (from 0 to 2). Two additional counter selections were done with Neu5Ac-sp-biotin and BSA/HSA after Rounds 5 and 7, respectively. Similar selection process was used against Neu5Ac-sp-biotin, except round 5, counter selection was done against Neu5Gc-sp-boitin. Detailed experimental conditions for aptamer selection against Neu5Gc-sp-biotin and Neu5Ac-sp-biotin are described in Tables 3.1 and 3.2 respectively.

Table 3.1: SELEX conditions during enrichment of aptamers against Neu5Gc-sp-biotin. An indirect capture method was used for screening the aptamer library against Neu5Gc-sp-biotin. The starting pool was a 60 bp synthetic DNA library pool with a 30 bp random variable region and flanking constant regions permitting re-amplification. In total, 10 rounds of solution-phase panning were carried out against Neu5Gc-spacer-biotin and bound aptamers were isolated utilising streptavidin-coated magnetic beads. Three rounds of negative selection were also performed to increase further specificity of DNA binders against Neu5Gc – through panning against the spacer-biotin alone, Neu5Ac-spacer-biotin, and against bovine and human serum albumin (BSA and HSA, respectively).

Rounds	**1st Negative round**	**1st Positive round**	**2nd positive round**	**3rd positive round**	**4th positive round**	**5th positive round**	**2nd negative round**	**6th positive round**	**7h positive round**	**3rd negative round**	**8th positive round**	**9th positive round**	**10th positive round**
Target sugar	**100 pmol Spacer-biotin**	**100 pmol (Neu5Gc-sp biotin)**	**100 pmol (Neu5Gc-sp biotin)**	**80 pmol (Neu5Gc-sp biotin)**	**80 pmol (Neu5Gc-sp biotin)**	**70 pmol (Neu5Gc-sp biotin)**	**80 pmol (Neu5Ac – spbiotin + sp biotin)**	**70 pmol (Neu5Gc-sp biotin)**	**60 pmol (Neu5Gc-sp biotin)**	**751 pmol of BSA + 746 pmol of HSA**	**50 pmol (Neu5Gc-sp biotin)**	**50 pmol (Neu5Gc-sp biotin)**	**50 pmol (Neu5Gc-sp biotin)**
ss DNA	**400 ng**	**400 ng**	**400 ng**	**400 ng**	**400 ng**	**400 ng**	**400 ng**	**400 ng**	**400 ng**	**400 ng**	**400 ng**	**400 ng**	**400 ng**
$MgCl_2$	**1 mM**	**1 mM**	**1 mM**	**1 mM**	**1 mM**	**1mM**	**1mM**	**1mM**	**2mM**	**1mM**	**2mM**	**5mM**	**5mM**
NaCl	**5 mM**	**5 mM**	**5 mM**	**5 mM**	**7.5 mM**	**10mM**	**10mM**	**10mM**	**12.5mM**	**15mM**	**15mM**	**17.5mM**	**20mM**
Incubation (37 °C)	**60 min**	**60 min**	**60 min**	**50 min**	**50 min**	**50 min**	**50 min**	**50 min**	**45 min**	**Overnight at 4°C**	**35 min**	**30 min**	**30 min**
Washing (PBS)	**No washing**	**1 washing**	**2 washing**	**2 washing**	**2 washing**	**2 washing**	**No washing**	**2 washing**	**2 washing**	**No washing**	**2 washing**	**2 washing**	**2 washing**
Elution	**No elution**	**2 elution**	**2 elution**	**2 elution**	**2 elution**	**2 elution**	**No elution**	**2 elution**	**2 elution**	**No elution**	**2 elution**	**2 elution**	**2 elution**

Table 3.2: SELEX conditions during enrichment of aptamers against Neu5Ac-sp-biotin. An indirect capture method was used for screening the aptamer library against Neu5Ac-sp-biotin. The starting pool was a 60 bp synthetic DNA library pool with a 30 bp random variable region and flanking constant regions permitting re-amplification. In total, 10 rounds of solution-phase panning were carried out against Neu5Ac-spacer-biotin and bound aptamers were isolated utilising streptavidin-coated magnetic beads. Three rounds of negative selection were also performed to increase further specificity of DNA binders against Neu5Ac – through panning against the spacer-biotin alone, Neu5Gc-spacer-biotin, and against bovine and human serum albumin (BSA and HSA, respectively).

Rounds	1st Negative round	1st Positive round	2nd positive round	3rd positive round	4th positive round	5th positive round	2nd negative round	6th positive round	7th positive round	3rd negative round	8th positive round	9th positive round	10th positive round
Target sugar	100 pmol Spacer-biotin	100 pmol Neu5Ac –sp biotin	90 pmol Neu5Ac – sp biotin	90 pmol Neu5Ac – sp biotin	80 pmol Neu5Ac – sp biotin	80 pmol Neu5Ac –sp biotin	100 pmol Neu5Gc –sp biotin	70 pmol Neu5Ac-sp biotin	60 pmol Neu5Ac-sp biotin	751 pmol of BSA + 746 pmol of HSA	50 pmol Neu5Ac -sp biotin	50 pmol Neu5Ac -sp biotin	50 pmol Neu5Ac -sp biotin
ss DNA	400 ng	400 ng	400 ng	400 ng	400 ng	400 ng	400 ng	400 ng	400 ng	400 ng	400 ng	400 ng	400 ng
$MgCl_2$	1 mM	1 mM	2 mM	3 mM	3 mM	4 mM	4 mM	4 mM	5 mM	1mM	2mM	5mM	5mM
NaCl	10mM	10mM	12.5 mM	15 mM	17.5 mM	20 mM	20 mM	20 mM	22.5 mM	22.5 mM	22.5 mM	25 mM	25 mM
Incubation (37 °C)	60 min	60 min	55 min	50 min	45 min	45 min	50 min	40 min	35 min	Overnight at 4°C	35 min	30 min	30 min
Washing (PBS)	No washing	1 washing	2 washing	2 washing	2 washing	2 washing	No washing	2 washing	2 washing	No washing	2 washing	2 washing	2 washing
Elution	No elution	2 elution	2 elution	2 elution	2 elution	2 elution	No elution	2 elution	2 elution	No elution	2 elution	2 elution	2 elution

3.2.8. Enrichment assay of DNA pools

The enrichment of Neu5Gc binding DNA aptamers over the course of selection was monitored using a fluorescent dye (OliGreen dye)-linked aptamer assay (Wochner and Glokler, 2007, Huang and Chang, 2008). DNA pools obtained after round 1-10 were examined. 100 pmol of sp-biotin or Neu5Gc-sp-biotin were immobilised on 250 µg of MyOne streptavidin C1 dynabeads in PBS (pH 7.4) at room temperature for 30 min. Carbohydrate coated dynabeads were washed twice using 500 µl PBS. ssDNA pool of each SELEX round (60 ng) prepared in 10 mM phosphate buffer pH 7.4, with 20 mM NaCl and 5 mM $MgCl_2$, (binding buffer) was denatured and renatured as described above and incubated with target coated dynabeads for 1 h at 37 °C. Beads were washed twice using 500 µL of binding buffer on magnetic separator for 2 minutes. Target bound ssDNA was eluted in 50 µL of nuclease free water by heating beads at 95 $^{\circ}$C for 5 min. Eluted ssDNA was incubated with 150 µl of OliGreen dye prepared in nuclease free 10 mM Tris-HCl buffer (pH 7.5) in flat-bottom black microtitre plates (Greiner bio-one, Belgium) and read using SpectraMax M5^e plate reader (Molecular Devices, Berkshire, UK) (excitation 480 nm, emission 520 nm)**.**

3.2.9. Cloning, and sequencing of enriched DNA pools

LB broth media (2g of LB media in 100 mL distilled water) was used for the growth of bacteria. This media also allows cells to recover from the shock of transformation. IPTG (2 mL of 100 mM), X-Gal (1.6 mL of 50 mg/mL) and Kanamycin (500 µL of 50 mg/mL) were added in 500 mL of LB media. Super optimally broth with catabolite repression (S.O.C) media (2% Tryptone, 0.5% Yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM Glucose was added and autoclaved. However, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM Glucose was autoclaved separately and then added to the mixture) is a nutrient rich growth media and used for higher transformation efficiency in plasmid. This media was stored at room temperature. Sterile distilled water was used in all preparation. For preparation of agar plates- 1.5 % agar was used in LB media.

Selected ssDNA pools from Neu5Gc and Neu5Ac panning were used as the template in a PCR reaction. The PCR primers were similar to those used in the panning

amplification; however, the forward primer used here is non-biotinylated. The PCR amplification conditions were similar to those used during panning amplification.

PCR amplified products were cloned into the pCR 4 TOPO vector (Invitrogen, UK) (Figure 3.1), a ligation reaction was prepared including, 4 µL of fresh PCR product, 1 µL of salt solution from cloning kit, 1 µL of pCR 4 TOPO vector. For control reaction without vector, 1 µL of nuclease free water was used instead of vector.

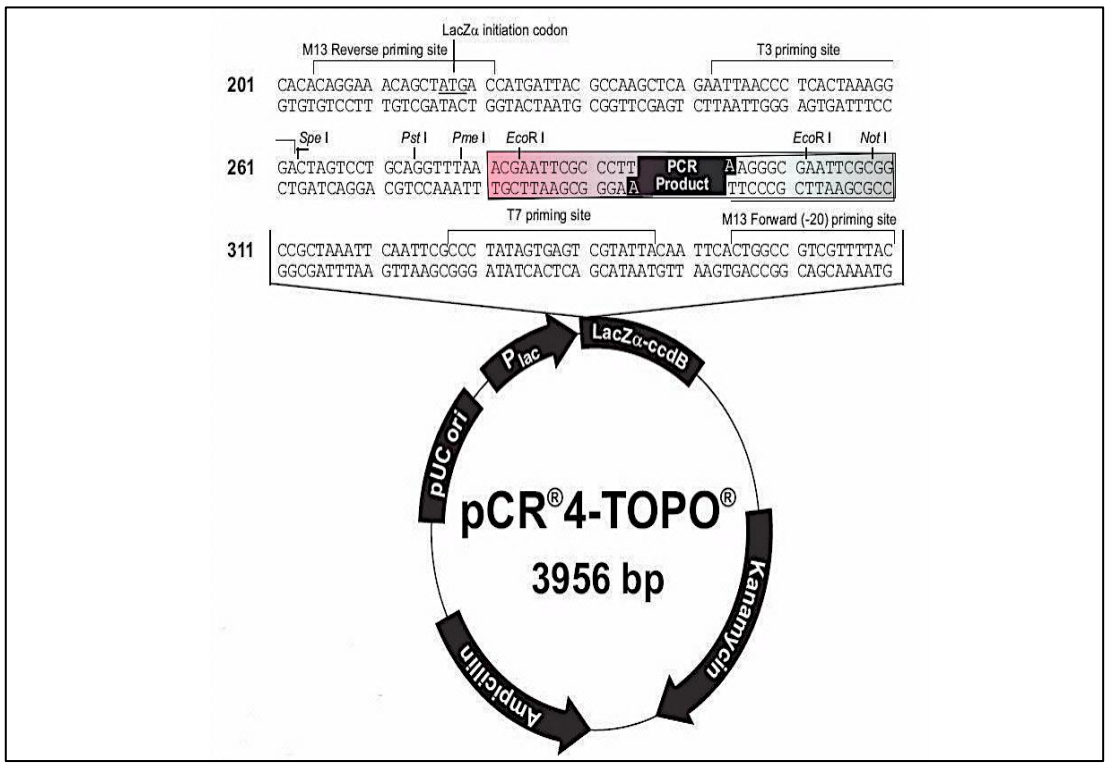


Figure 3.1: schematic representation of pCR 4-TOPO plasmid. This plasmid has a multiple cloning sites, genes for ampicillin and kanamycin antibiotic selection and it can be used for blue-white screening.

Ligation reactions were used to transform chemically competent One Shot® TOP10 *E. coli* cells (Invitrogen, UK), 50 µL of *E. coli* cells were thawed and 2 µL of ligation reaction was added into the cells and incubated on ice for 30 min. The cell and DNA mix was then heated in a water bath at 42 °C for 30 s and immediately put on ice for 2 min. 250 µL of SOC media was added to each reaction and at 37 °C for 1 hour with constant shaking and then were streaked onto LB-agar plates containing appropriate selection reagents at 37 °C overnight. The transformed clones were screened using blue-white screening.

Ligation of DNA into vector was determined using colony PCR using M13 forward and reverse primers (Sigma, UK). To confirm ligation of dsDNA, control ligation with water instead of dsDNA template was used. Both of these ligation reactions were then chemically transformed into *E.coli* TOP 10 cells and the inoculum (50 µL and 100 µL) was streaked onto agar plates for incubation at 37 °C overnight.

To confirm streaking and colony growth, three different conditions were used; for positive control, two plates for each round were streaked with 50 µL and 100 µL inoculum with kanamycin on agar plates. As a negative control, cells without any ligation reaction were subjected to heat shock under similar conditions as those with inserts and were streaked with 100 µL inoculum, with or without kanamycin. For PCR efficiency, heat shocked cells without DNA template amplicons ligation reactions were streaked with 100 µL inoculum on the agar plate.

The PCR amplification conditions used were 94 °C for 5 minutes, then 26 cycles of (94°C for 30 sec, 55°C for 30 sec and 72° for 30 sec) and then an extension of 72°C for 7 minutes with a final temperature of 4°C for 15 minutes. A single, well-defined colony from the selection plate was used as the template. The bacterial colonies having DNA insert were verified on agarose gel (2.5%) with a 100 bp DNA ladder (Sigma, UK). Positive transformants were regrown on an agar 96-well plate and sent for sequencing using microtitre plate sequencing (MTP) (MWG Eurofins, Germany).

3.2.10. Sequence extraction of sequenced DNA pools

All the sequences were extracted from the Neu5Gc and Neu5Ac binding DNA pool sequencing data file (FASTA files got from Eurofins Genomics) using the plasmid region information from either side of the DNA insert (Figure 3.1). A PERL programming script was run to search plasmid region 1 “CCGCGAATTCGCCCTT" and plasmid region 2 "AGGGCGAATTCGT". The DNA sequence found in between the two plasmid regions were extracted and only DNA sequence lengths from 58-62 bases were considered for further sequence analyses.

3.2.11. Sequence analyses of enriched and sequenced DNA pools

Multiple sequence alignment was carried out using ClustalW2 (Larkin et al., 2007). (http://www.ebi.ac.uk/Tools/msa/clustalw2/). The thirty nucleotides variable region

of aptamer was considered for aligning multiple sequences. For aligning multiple sequences, a Clustal W DNA weight matrix, with a gap open penalty of 10 and gap extension penalty of 0.20 with a gap distance of maximum 5 was used throughout the alignment. For two sequence alignment, a slow pairwise alignment option was used for higher sensitivity where the gap open penalty and gap extension penalty were kept at 10 and 0.1 respectively. Clustal w/ numbers format was used for output view of both multiple and two sequences alignment.

3.2.12. Secondary structure prediction

The lowest free energy secondary structures of candidate aptamers were predicted using the mfold program (http://mfold.rna.albany.edu/?q=mfold) (Zuker, 2003, Zuker, 2000). The input sequence was in FASTA format (text-based format for representing nucleotide sequences) (Figure 3.2).

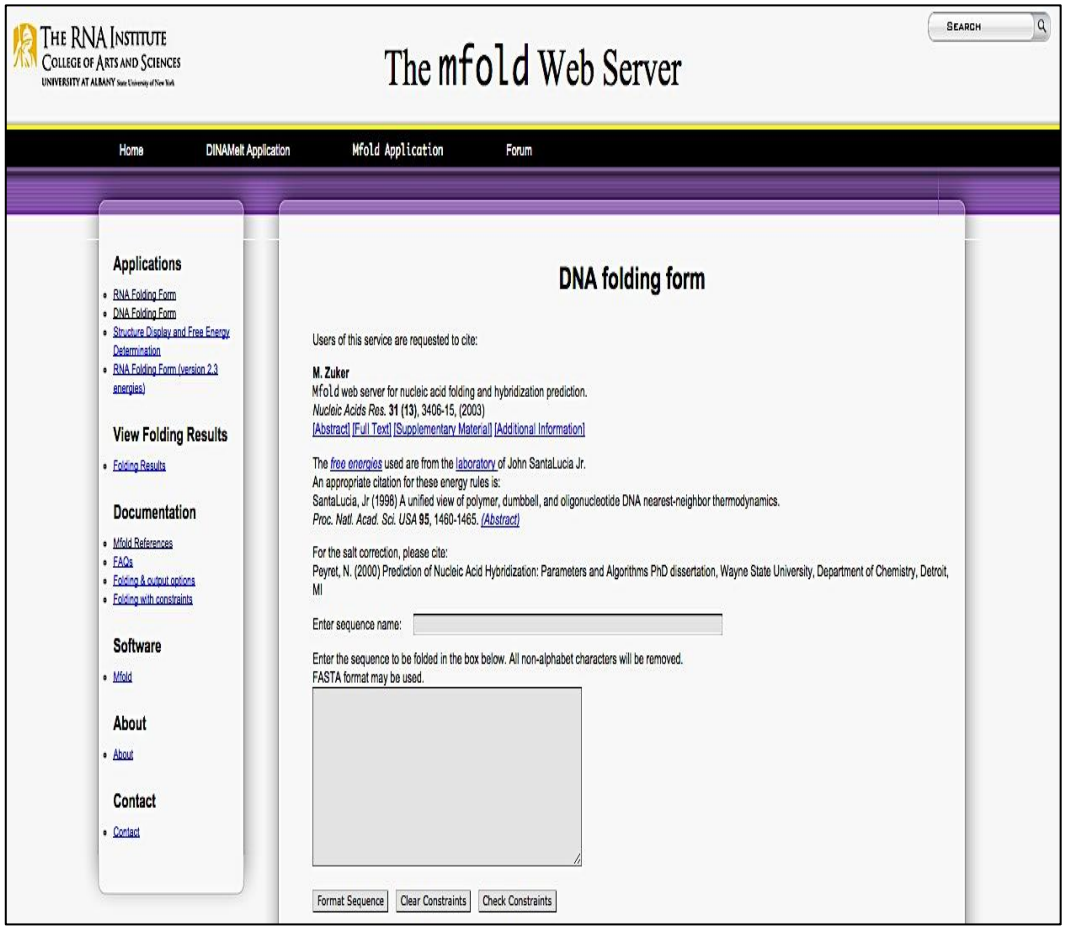


Figure 3.2: mfold web server for the prediction of DNA secondary structure.

To calculate the thermodynamics of aptamer folding, a free energy determination algorithm was used (SantaLucia, 1998) based on the equation-

$$\Delta G = \Delta H - T\Delta S$$

Where ΔG is the change in free energy, ΔH is the change in enthalpy, T is the temperature and ΔS is the change in entropy.

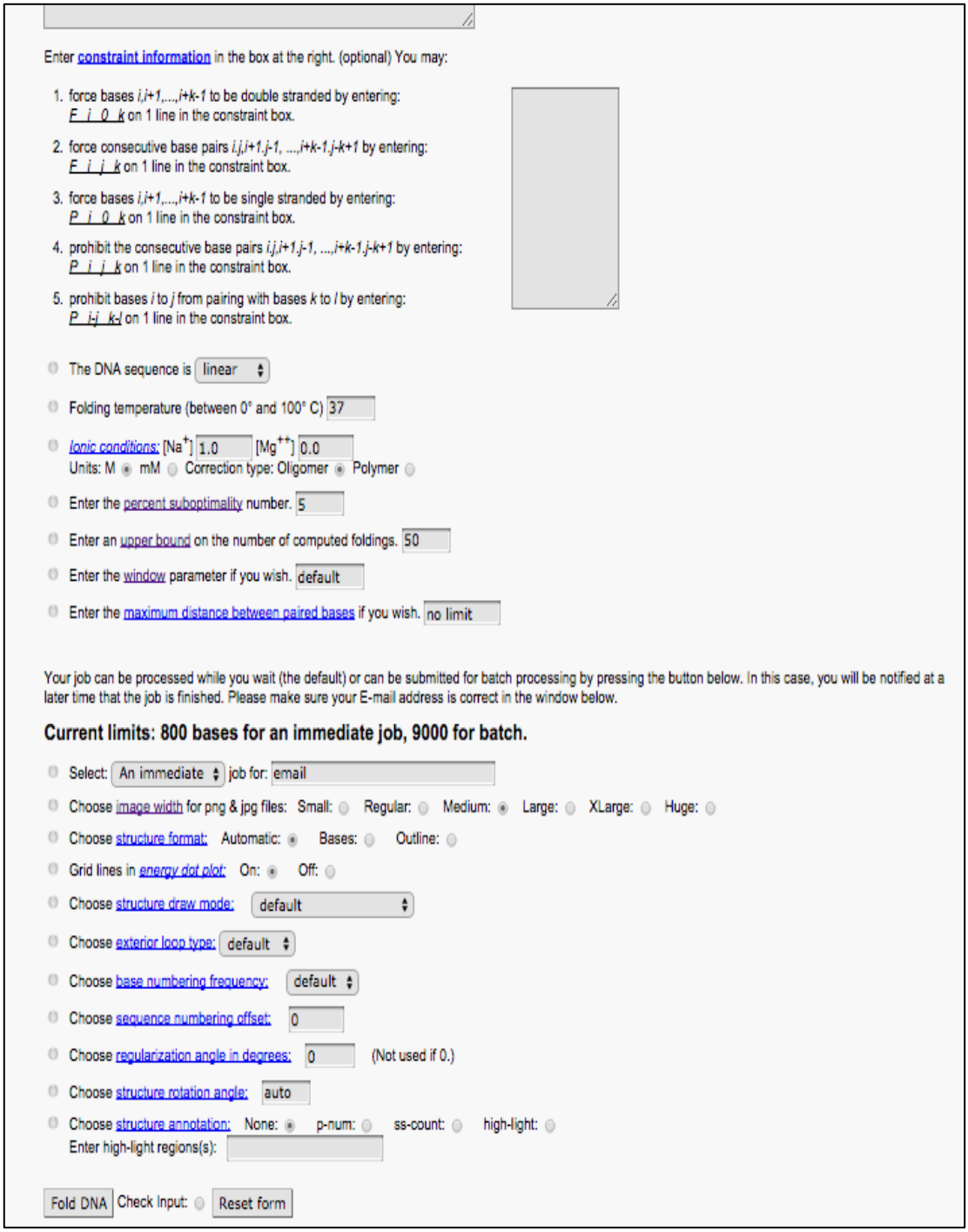


Figure 3.3: screen shot of mfold DNA folding constraint and output options.

The conditions required in the software are the nature of DNA sequence, which was considered to be linear. The folding temperature was kept constant at 37 °C as used in panning experiments. The ionic conditions used in the software were 157 mM Na^{+} and 5.0 mM Mg^{++} for Neu5Gc aptamer prediction and 162 mM Na^{+} and 5.0 mM Mg^{++} for Neu5Ac aptamer structure prediction. These are the ionic condition used in aptamer panning experiments and efforts were made to mimic the conditions used in the bench experiments.

The other conditions used were Percent suboptimality; to compute folding within the minimum free energy, Upper bound; to compute maximum number of secondary structures required within the prescribed energy requirement, Window parameter; to control the number of folding provided they are different from each other and the JPG structure format of the images of aptamer secondary structure.

3.2.13 Statistical analysis

The statistical analysis of data was performed using GraphPad Prism software for Mac versions 6.0c (GraphPad Software Inc., SanDiego, CA, USA). All data is represented as the mean of three replicate determinations. The results are shown as mean values with the standard error of mean (SEM) represented by error bars.

3.3. Results

3.3.1. ssDNA preparation

Prior to selection, the sise of ssDNA was confirmed by urea gel electrophoresis (60 bp ssDNA library) and the amplified ssDNA library (ds DNA) was again tested by agarose gel electrophoresis. The quality and quantity of ssDNA after each round was analysed using urea gel electrophoresis and Nanodrop absorbance reading, respectively. The initial ssDNA preparation involved 2-step quality and sise check of ssDNA as follows: 1) after PCR amplification and 2) after streptavidin incubation on urea gel for both ssDNA separation and sise confirmation (Figure 3.4). As this protocol required 5 hours of extra agarose gel run and DNA purification using Qiagen kit, and also due to requirement of multiple steps of gel analysis and purification, there was a considerable amount of ssDNA lost after each step. Therefore, this protocol was further optimised to require only one step to be used in which PCR amplified DNA was directly incubated with streptavidin and later DNA strands were separated on urea gel (Figure 3.5).

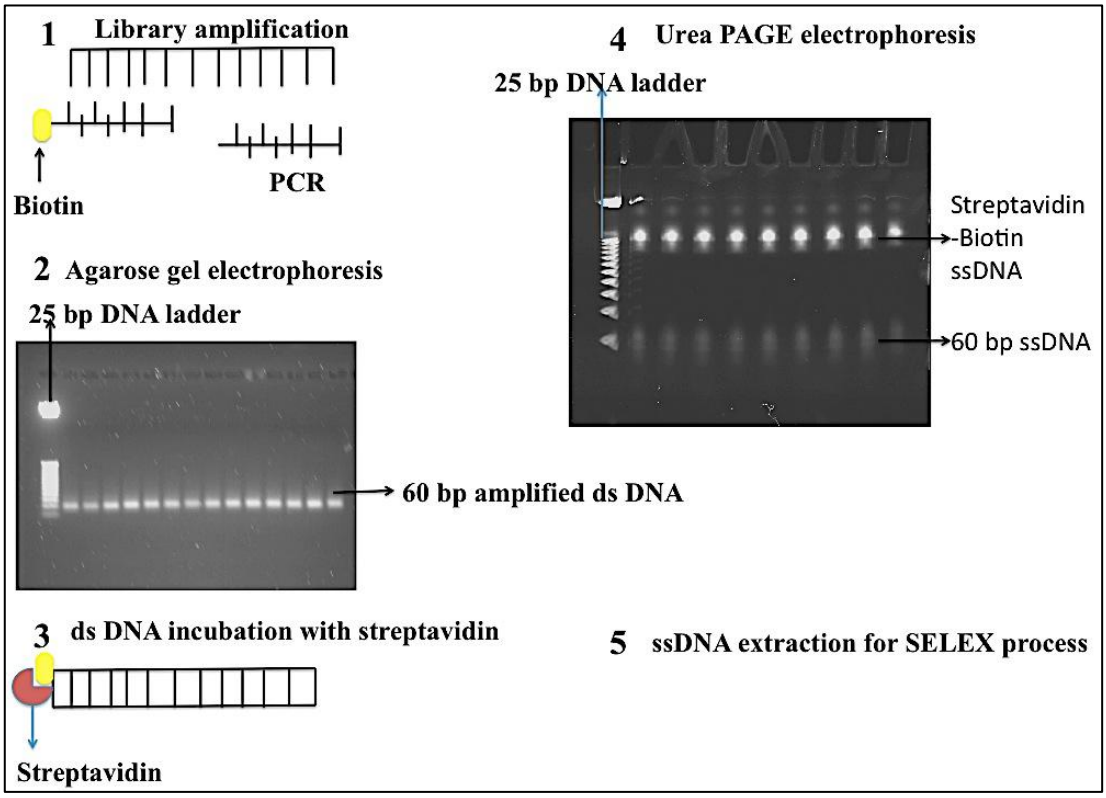


Figure 3.4: The five steps process involved for the amplification and preparation of ssDNA for subsequent selection. The first step involved the amplification of DNA where forward biotinylated primer makes one strand of amplified dsDNA biotinylated. The second step involved visualisation and extraction of biotinylated dsDNA on Agarose gel and extraction of biotinylated dsDNA using Qiagen kit method. The third step involved incubation of biotinylated dsDNA with streptavidin, the fourth step involved separation of ssDNA strand with streptavidin-bt-DNA strand on urea gel. The separated ssDNA was excised and extracted in the final step.

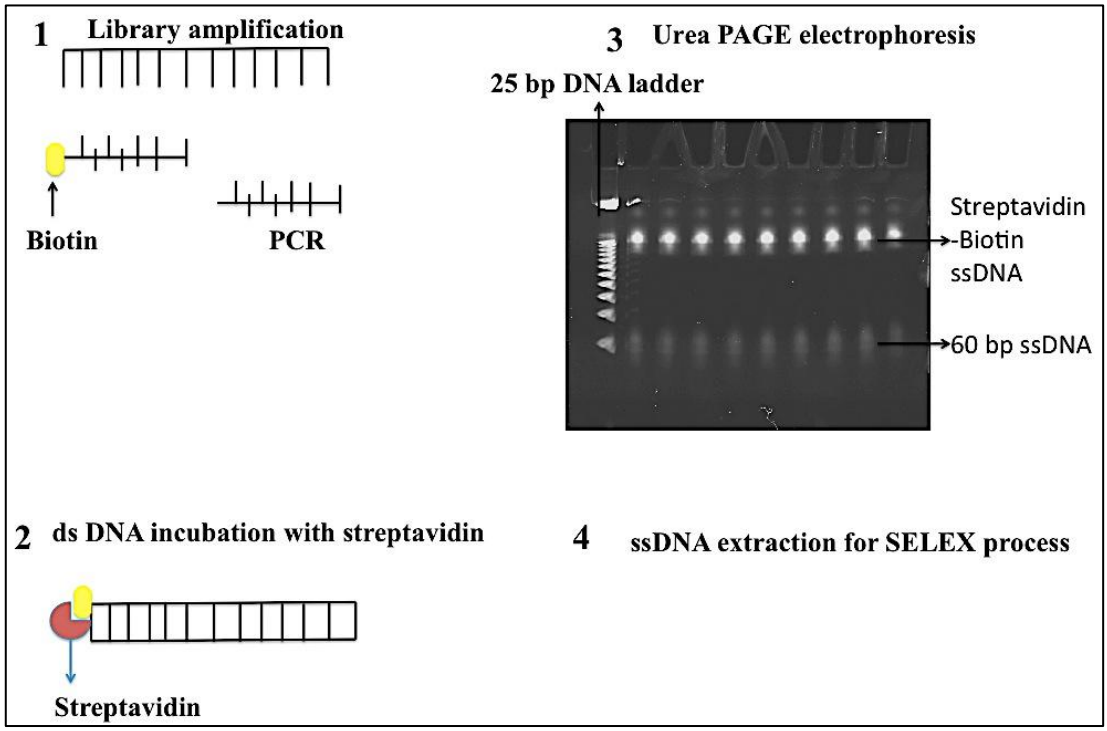


Figure 3.5: The preparation scheme requiring reduced steps in the process involved for the amplification and preparation of ssDNA as in Figure 3.4. The agarose gel electrophoresis step has been omitted and the amplified ds DNA library incubated directly with streptavidin and separated on urea gel.

For the 60 bases ssDNA extractions from gel, initially a commercial DNA extraction kit from Qiagen was used that gave less efficiency and resulted in the loss of ssDNA. The lower efficiency of DNA extraction was due to the small sise (60 bases) of the ssDNA. This method was later substituted with Phenol-Chloroform extraction (Sambrook and Russell, 2006), that gave higher yield and purity of the ssDNA (Table 3.3). To show the difference in efficiency two different gels of the same DNA were amplified, incubated with streptavidin and ran on urea gel (Figure 3.6- A & B). ssDNA from the urea gel (Figure 3.6A) was extracted using the Qiagen kit extraction method, while ssDNA from the urea gel (Figure 3.6 B) was extracted using the Phenol-Chloroform extraction method and their difference in quantity and quality of ssDNA was reported in Table 3.3.

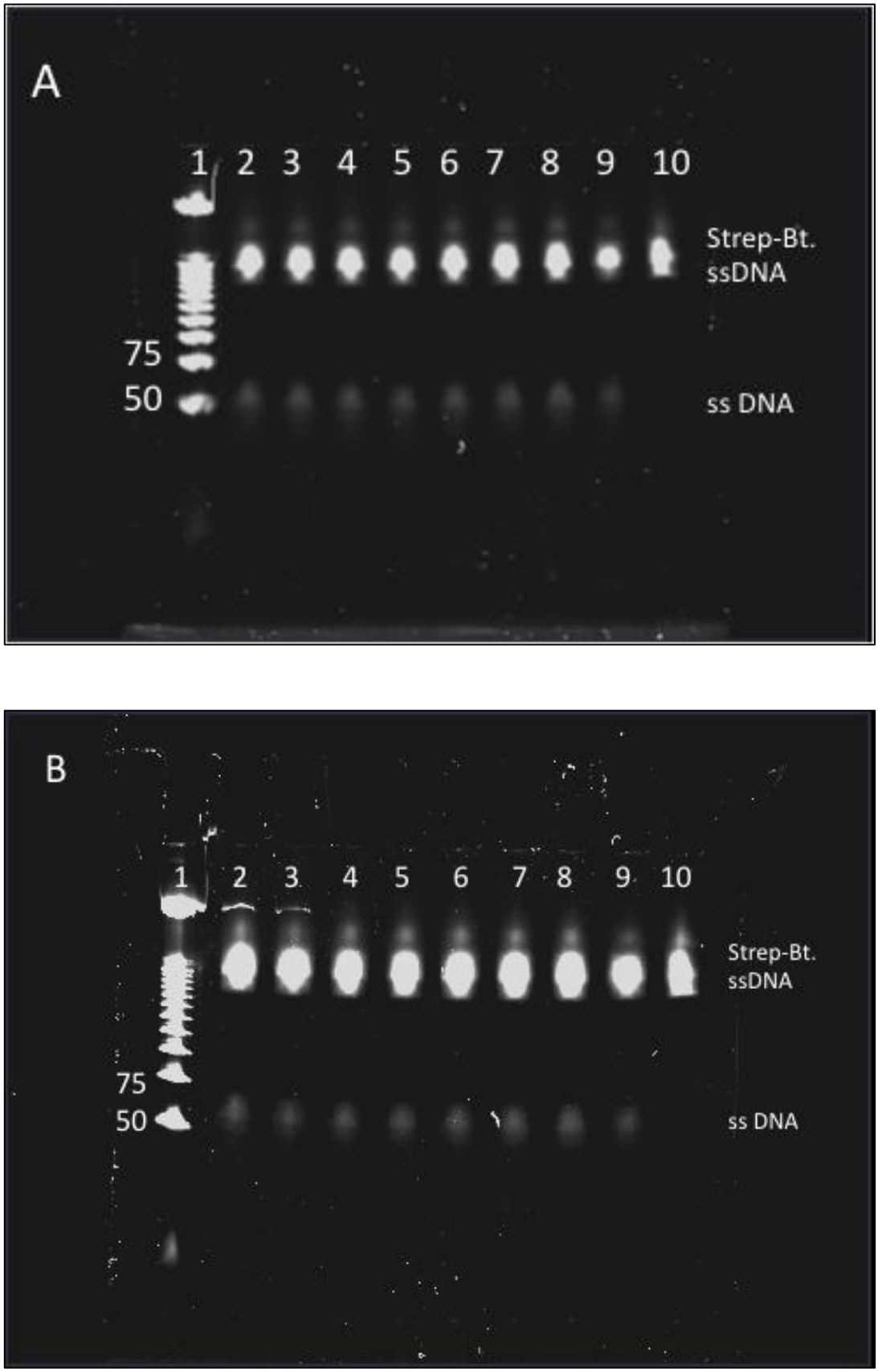


Figure 3.6: ssDNA separation from amplified ds DNA. (A) The separated ssDNA bands were excised and extracted using Qiagen kit method. (B) The separated

ssDNA were excised and extracted using Phenol-Chloroform extraction method. In figure 3.6 (A) and (B) lane 1 is a 25 bp ladder, lane 2-9 have ssDNA between 75 and 50 bp ladders, lane 10 is negative control (PCR amplicons without a ssDNA template) and the upper bright bands are streptavidin conjugated to single biotinylated DNA strand.

Table 3.3: Comparison of quantity and quality of ssDNA extraction using the two different methods, Qiagen kit extraction and Phenol-Chloroform extraction.

ss DNA extraction method	Concentration of ssDNA (ng/µL)	260/280	Volume of elution (µL)
Qiagen kit extraction	**2.8**	**1.21**	**100**
Phenol-Chloroform	**72.1**	**1.83**	**100**

***** ssDNA was quantified in ng/µL. 260/280 absorbance ratio was used to check the contamination in ssDNA elution, a ratio of 1.8 - 2.2 is considered to be without any contamination. Final volume of ssDNA elution was 100 µL.

3.3.2. SELEX cycles and enrichment for Neu5Gc and Neu5Ac binders

The SELEX approach was used to select DNA aptamers that bind to Neu5Gc and Neu5Ac from the same pool of randomised 60-mer DNA oligonucleotides. A total of ten positive and three negative selections were undertaken to isolate binders against each of the targets (Tables 3.1 and 3.2) respectively. The positive selections were performed against the appropriate sialic acid linked to biotin via a linker. The initial negative selection, in each case, involved incubation of the library with biotin linked to the spacer (Sp-biotin) to remove non-specific binders. The second negative selection after five rounds of positive selections was against Neu5Ac-sp-biotin when selecting for Neu5Gc binders and *vice versa* during Neu5Ac binder selection. The final negative selection after round seven was done against BSA and HSA to remove non-specific binders against these common components of binding assay systems. The stringency of the binding conditions in the positive selections was increased gradually during the course of selection. Enrichment of binding activity to the appropriate sialic acid target was assessed over the ten positive SELEX rounds for

each target using a fluorescent dye-linked aptamer assay. A gradual increase in binding of DNA molecules to Neu5Gc-sp-biotin was noted over the first 7 SELEX rounds (Figure 3.6). Although round 8 shows less binding than round 7, which could be due to an extra negative selection after round 7 (Figure 3.7), round 7 and 9 showed highest binding. A similar trend was noted on evaluation of Neu5Ac-sp-biotin binding enrichment, with rounds 8 and 10 showing highest binding (Figure 3.8).

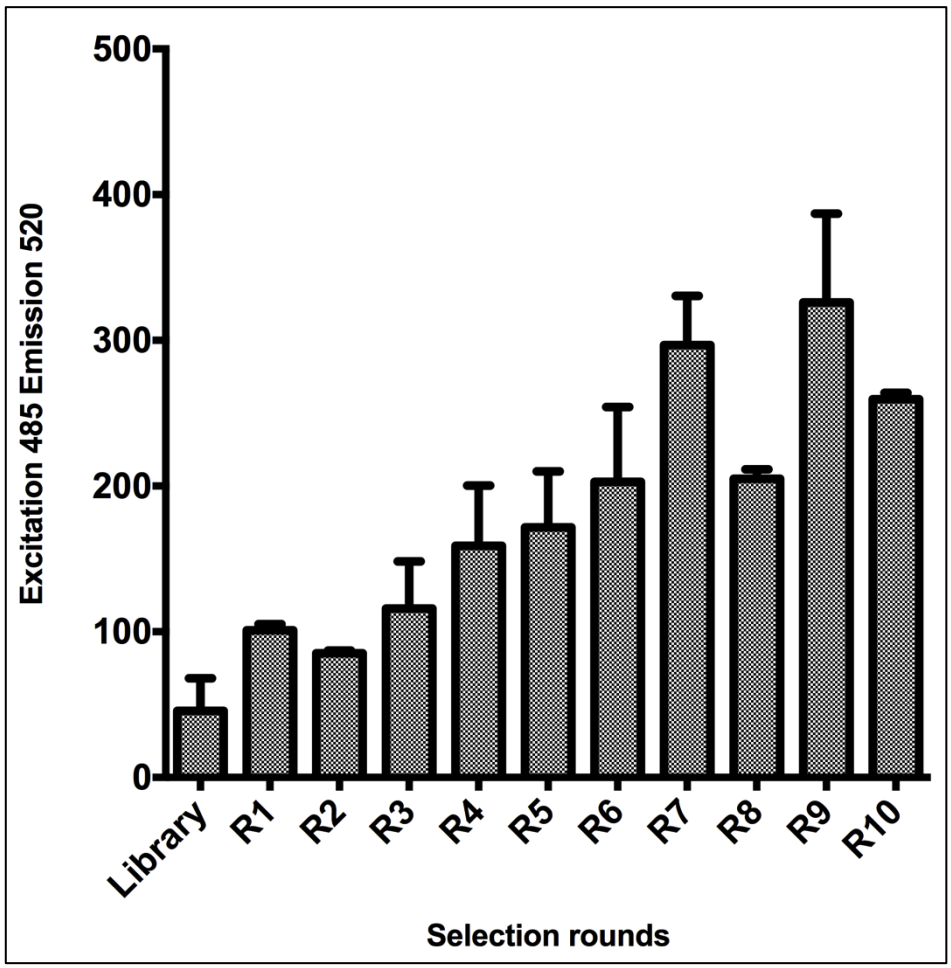


Figure 3.7: The SELEX enrichment of ssDNA pools against Neu5Gc-sp-biotin obtained after rounds 1-10 (x-axis). The values on y-axis represent relative fluorescence unit.

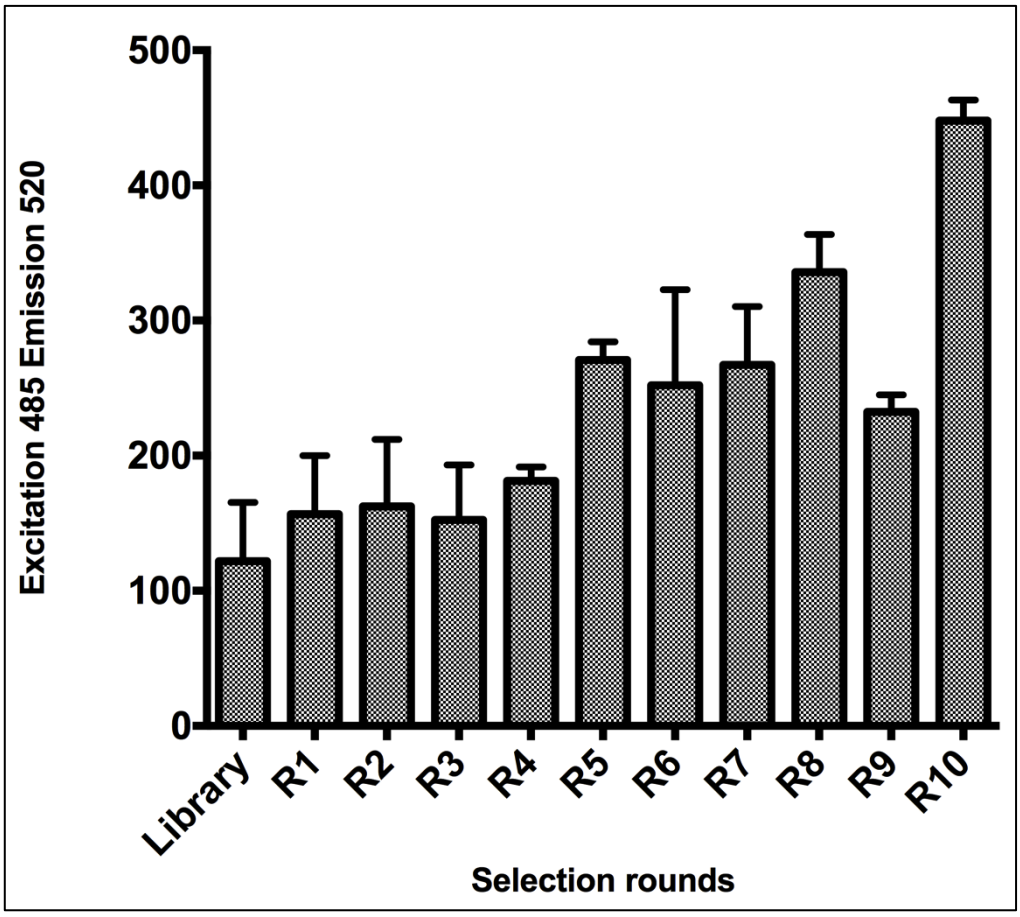


Figure 3.8: The SELEX enrichment of ssDNA pools against Neu5Ac-sp-biotin obtained after rounds 1-10 (x-axis). The values on y-axis represent relative fluorescence unit.

3.3.3. Cloning and sequencing of ssDNA

Following cloning of ssDNA from rounds 7 and 9 of Neu5Gc binding panning their sise and quality was checked on an agarose gel (Figure 3.9). The sise of ssDNA (60 bp) was indicated to be correct in comparison to the standard ladder. In addition, the control well (last well) did not show any bands thus confirming that template used for amplification from round 7 and 9 is right sise ssDNA insert.

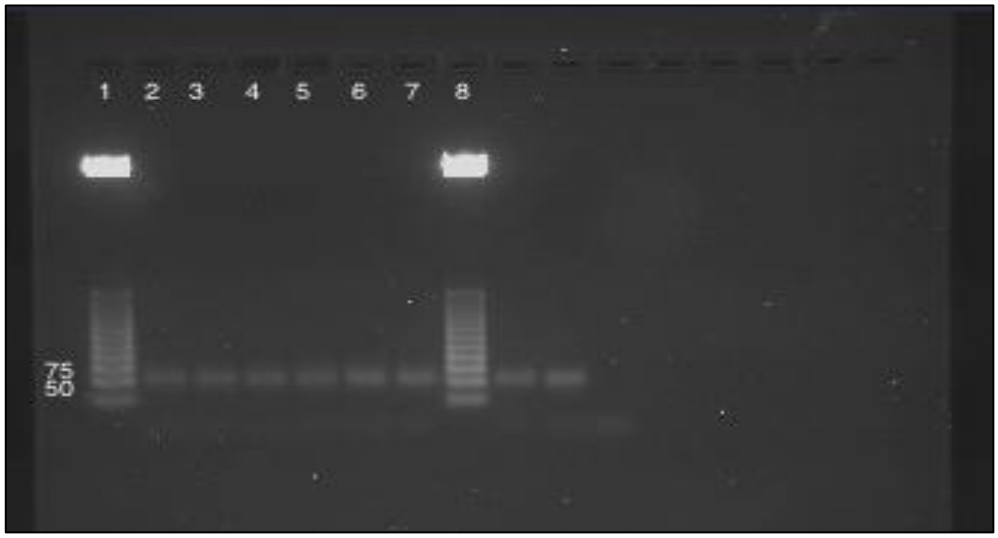


Figure 3.9: Agraose gel showing 60bp dsDNA after PCR amplification, Lane 1 and 8 are 25 bp DNA ladder

After confirmation of the ssDNA insert, PCR amplified dsDNA was ligated into pcr 4.0 TOPO vector. ssDNA was subsequently cloned into pCR4.0 TOPO ahead of sequencing of the ssDNA insert. Colony PCR showed that most of the colonies were of the right sise of about 225 bp (165 bp for M13 colony PCR primers + 60bp DNA insert). A small number of the colonies did show faint or no DNA bands lane, which suggested that the colonies do not have ssDNA insert. The colonies that demonstrated the correct sise insert (225 bp) on agarose gels were selected for sequencing (Figure 3.10).

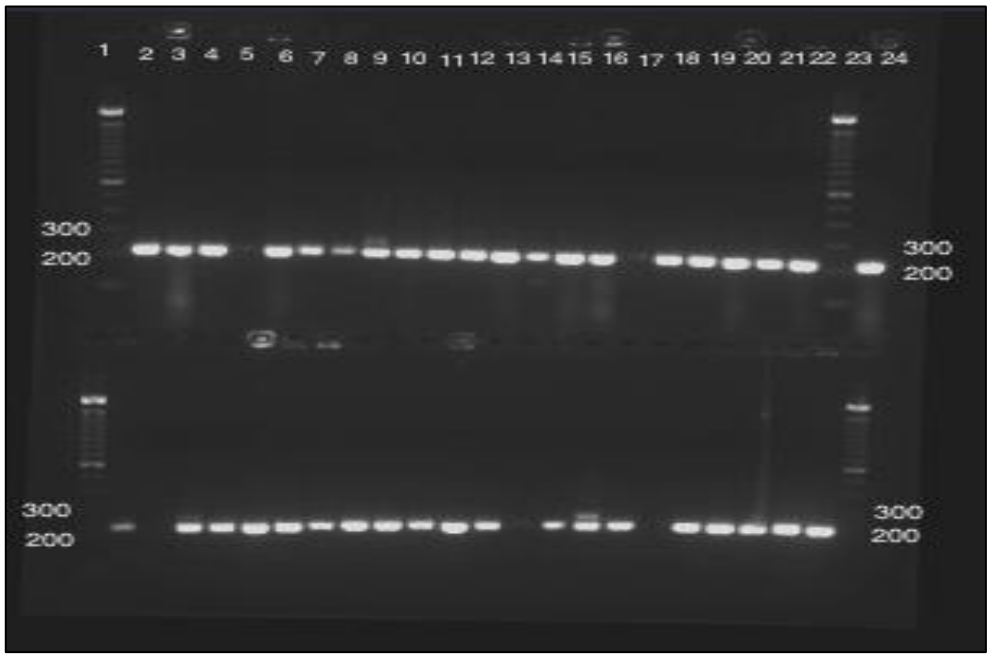


Figure 3.10: shows colony PCR of white colonies grown on the agar plate. 100 bp DNA ladders are on each end of the gel.

As a few picked colonies on the agarose gel (Figure 3.10) did not show any band, further testing of DNA insert in colonies was validated through PCR amplification and these amplified samples were again run on an agarose gel (Figure 3.11). The samples that were PCR amplified from the colony template showed DNA bands of the correct sise (lane 2-5), the samples loaded directly from colony PCR stock (PCR amplicons having white colony as DNA template) showed thick and bright DNA bands (lane 6-11). The PCR amplicons used for ligation reactions showed faint bands as expected (12-15). These DNA band intensity differences were due to the presence of higher amount of dsDNA in bacterial colony than in the DNA ligation reaction. The control wells lane (2,3 and 6,7) also showed a DNA band, and it was therefore concluded that blue colonies (control colonies) also had the DNA insert in them. The occurrence of blue colour could be due to the presence of small sise of DNA aptamer (60 bp) that was not able to disrupt the β-Galactosidase expression. The disruption of β-Galactosidase expression is generally due to shifting of the reading frame, that can introduce a stop codon and therefore removing the enzymatic activity required to cleave X-Gal and hence produce the blue colour. In the end, randomly isolated white colonies from different plates were picked and sent for sequencing.

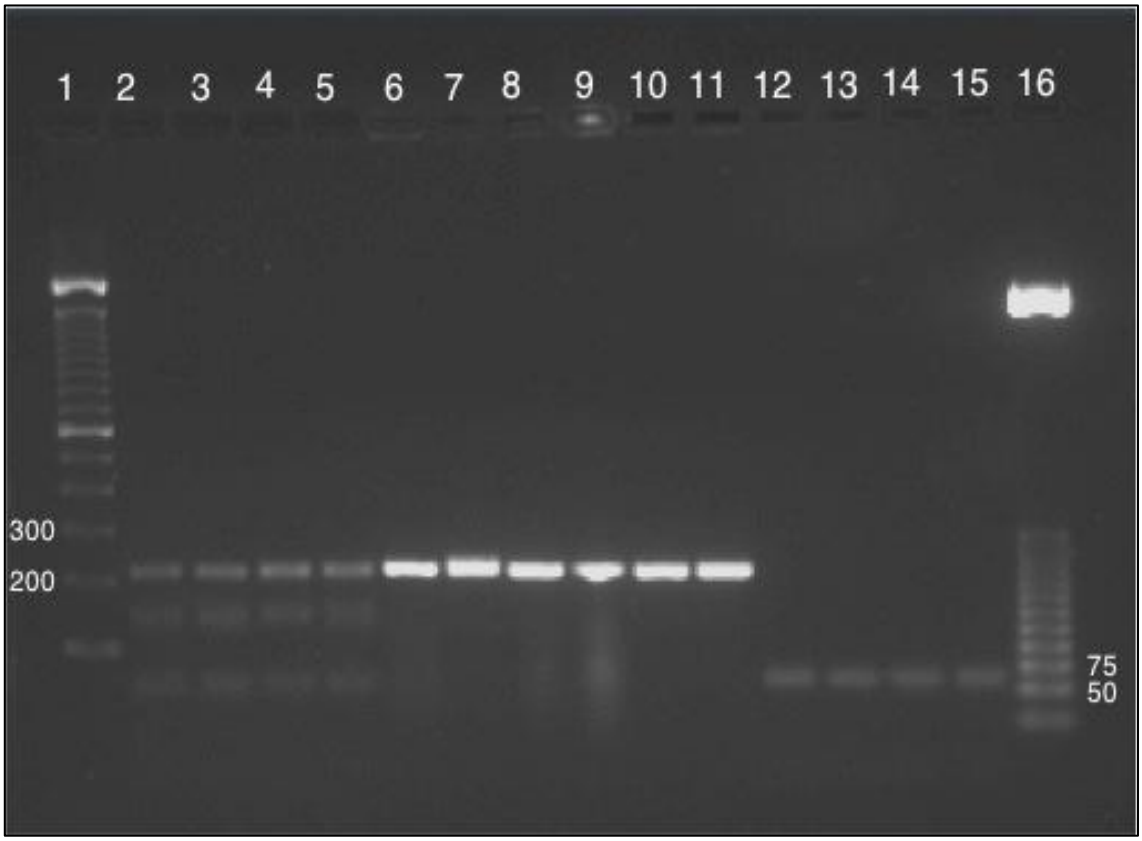


Figure 3.11: Testing of DNA insert in colonies after PCR amplification. Lane 1 is the 100 bp ladder, lane 2-5 samples amplified with colonies as their template for amplification, lane 6-11 samples are loaded directly from the colony PCR stock, lane 12-15 are actual PCR amplicons used ligation and transformation in *E.coli* (60 bp sise), lane 16 is 25 bp DNA ladder. Lane 2,3 and 6,7 are the control sample (Blue colonies picked from the plate). Lane 4, 8, 9, 12 and 13 are from round 7, whereas 5, 10, 11, 14 and 15 are from round 9.

3.3.4. Sequence analysis of enriched aptamers

As higher binding was demonstrated against the target Neu5Gc in SELEX rounds 7 and 9, these DNA pools were selected for cloning into pCR4.0 vector. A detailed sequence analysis was carried out and consensus was determined across the two rounds. A total of 218 sequences were extracted for round 7 and 250 sequences for round 9 clones were analysed. After aligning these sequences with ClustalW three consensus sequences occurred across the two rounds 7 and 9, accounting for the majority of sequences extracted (Table 3.4). The consensus sequences that have emerged from the 9th round of SELEX were anticipated to show higher specificity for Neu5Gc over other targets due to the additional positive selection rounds and the

negative selection rounds against non-target Neu5Ac, HSA and BSA included before positive round 9.

A total of 482 clones from the Neu5Gc output pools were sent for sequencing, of which 468 were found to be of full-length (58-62bp). Out of these 468 sequences, 218 and 250 sequences were extracted from rounds 7 and 9, respectively. Three consensus sequences emerged following detailed analysis, with two of the three sequences accounting for 98% of the sequences obtained (Table 3.4). For Neu5Ac, 189 clones were sent for sequencing, of which 168 sequences were full-length, with five consensus sequences being observed (Table 3.5).

Table 3.4: The consensus sequences of Neu5Gc binding aptamer obtained after sequencing round 7 and 9.

Neu5Gc	**Copies across round 7 and 9**	**Sequence per consensus**
Gc-101	**237/468**	**GCGCAAGCTTCGCGCCCGAACGCGAAGGTTGTGTGTATTGTTGCGCGCGGGATCCGCGC**
Gc-224	**221/468**	**GCGCGGATCCCGCGCGCAACAATACACACAACCTTCGCGTTCAGGCGCGAAGCTTGCGC**
Gc-239	**9/468**	**GCGCGGATCCCGCGCGCAACATAACCGACCCCAACCTTCGCCCGGCGCGAAGCTTGCGC**

***** Underlined sequences on the both ends of the aptamer sequences are the constant regions of 15 nucleotides length; the sequences inbetween the constant regions are the variable regions**.**

Table 3.5: The consensus sequences of Neu5Ac binding aptamer obtained after sequencing round 8 and 10.

Neu5Ac	Copies across round 8 and 10	Sequence per consensus
Ac_79	**42/168**	**GCGCAAGCTTCGCGCCAAGCCGCGAAGGTT GGGGTCTAAATGTGGGCGCGGGATCCGCG**
Ac_68	**28/168**	**GCGCAAGCTTCGCGCCCGAACGCGAAGGTT GTGTGTATTGTTGCGCGCGGGATCCGCG**
Ac_89	**25/168**	**GCGCAAGCTTCGCGCCGGGCTGCGAAGATT GGTGGTGTTGTTGGCGCGGGATCCGCGC**
Ac_94	**43/168**	**GCGCGGATCCCGGGCCAACAACACCACCAA TCTTCGCAGCCCGGCGCGAAGCTTGCGC**
Ac_93	**26/168**	**CGCGGATCCCGCGCGCAACAATACACACAA CCTTCGCGTTCGGGCGCGAAGCTTGCGC**

***** Underlined sequences on the both ends of aptamer sequences the constant regions of 15 nucleotides long, the sequences in between the constant regions are variable region**.**

3.3.5. Sequence motif search

The Neu5Gc binding consensus sequences Gc-224, 239 and 101 variable regions were aligned together (Figure 3.12). The constant regions of these sequences were omitted from this analysis as it was anticipated that the binding of the aptamer varies with the sequence change in their variable region. There were four sequence motifs observed after aligning The first motif is (G/C – C/G – A – A – C – A/G), the second A – C/A, the third G/C – A/T – C/T and the fourth TT respectively. On comparing Gc-224 and 239 they have a motif AACCTTCGCG, however the sample sise of Gc-239 (9 oligonucleotides) is smaller than Gc-224 (237 oligonucleotides) (Figure 3.10). In Neu5Ac aptamer motif alignment, it was difficult to predict an aptamer motif upon comparing all five aptamer species together (figure 3.13). However, aptamer Ac- 68, 89 and 93 share common motifs and aptamer Ac- 94 and 93 possess common

motifs. Also, sequences Gc-224 and Ac-79, which were selected for further characterisation also aligned and there are few motifs of 2 to 3 nucleotides long (Figure 3.14).

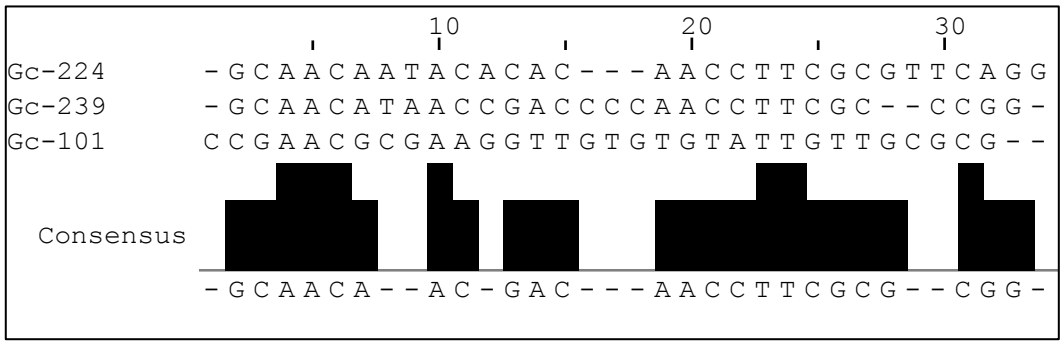


Figure 3.12: shows alignment of Neu5Gc binding sequence consensus, Gc-224, 239 and 101 variable regions.

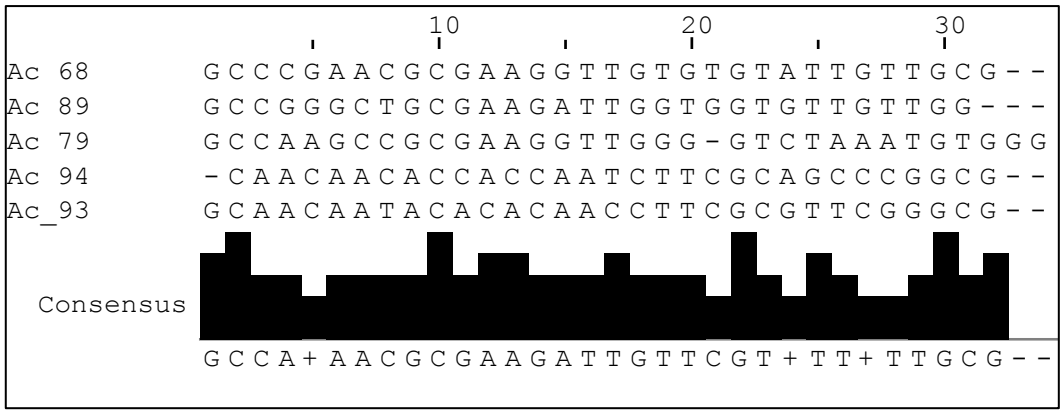


Figure 3.13: shows alignment of Neu5Ac binding sequence consensus, Ac-68, 89, 79, 94 and 93, variable regions.

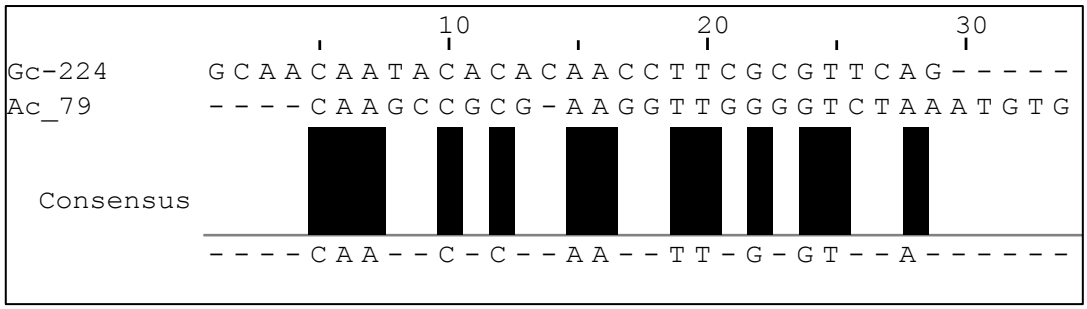


Figure 3.14: shows alignment of Neu5Gc binding sequence Gc-224 and Neu5Ac binding sequence Ac-79 variable regions.

3.3.6. Secondary structure prediction

The secondary structures of all aptamers against Neu5Gc and Neu5Ac were predicted after inputting the details of conditions used in the bench experiments. Neu5Gc binding aptamers Gc-224 and Gc-239 have similar structure, while Gc-101 is inverted than Gc-224 and Gc-239 (Figure 3.15). Neu5Ac binding sequences have different structure as their stem loop sequences have different nucleotides in each structure (Figure 3.16).

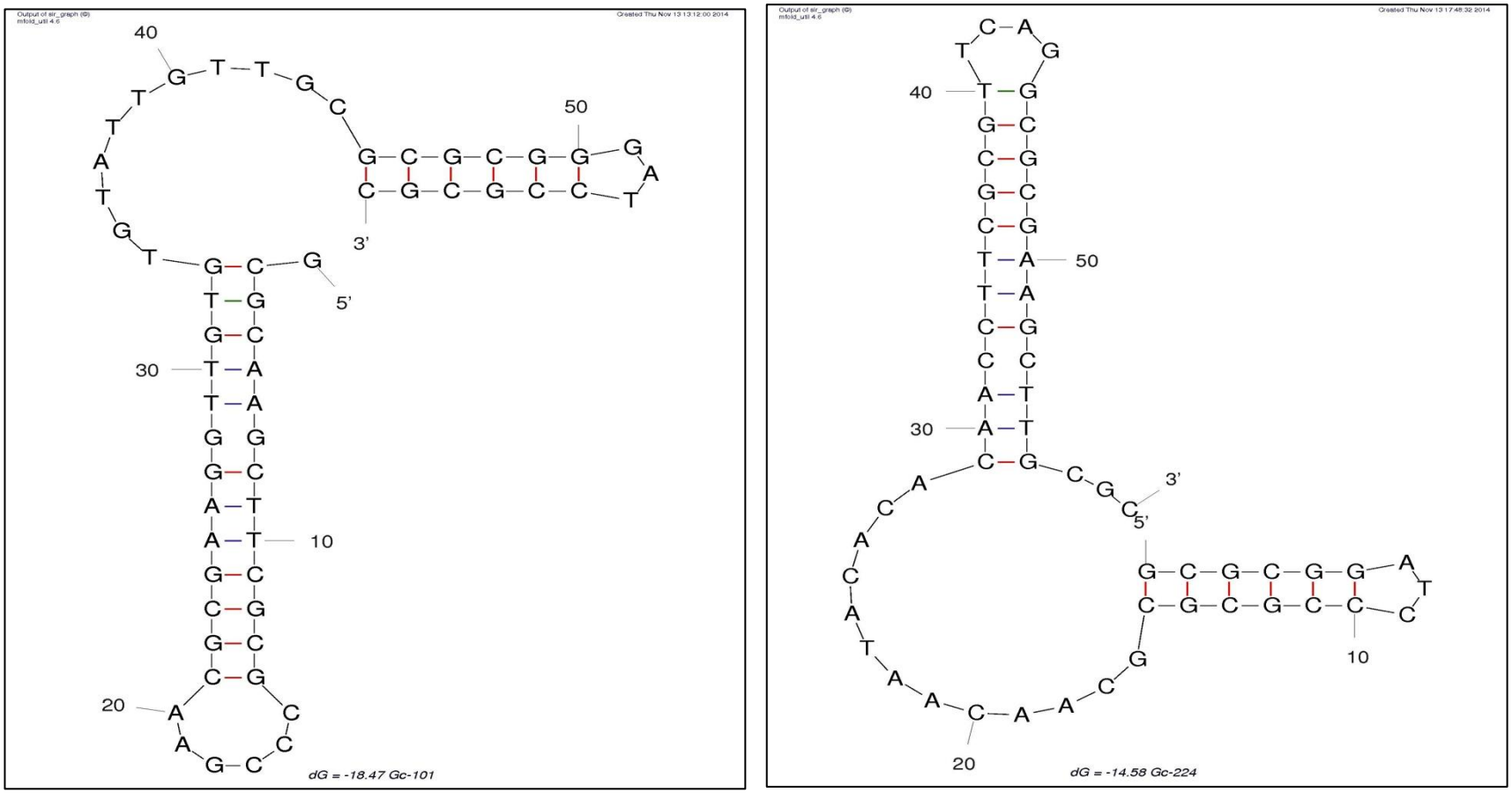


Figure 3.15: Secondary Structures of **(A)**: Gc-101, **(B)**: Gc-224. The secondary structures of ssDNA sequences selected against Neu5Gc were predicted using mfold DNA model software. The conditions required in the software are the nature of DNA sequence, which was considered to be linear. The folding temperature was kept constant at 37 °C as used in panning experiments. The ionic conditions used were 157 mM Na^{+} and 5.0 mM Mg^{++} for Neu5Gc aptamer prediction.

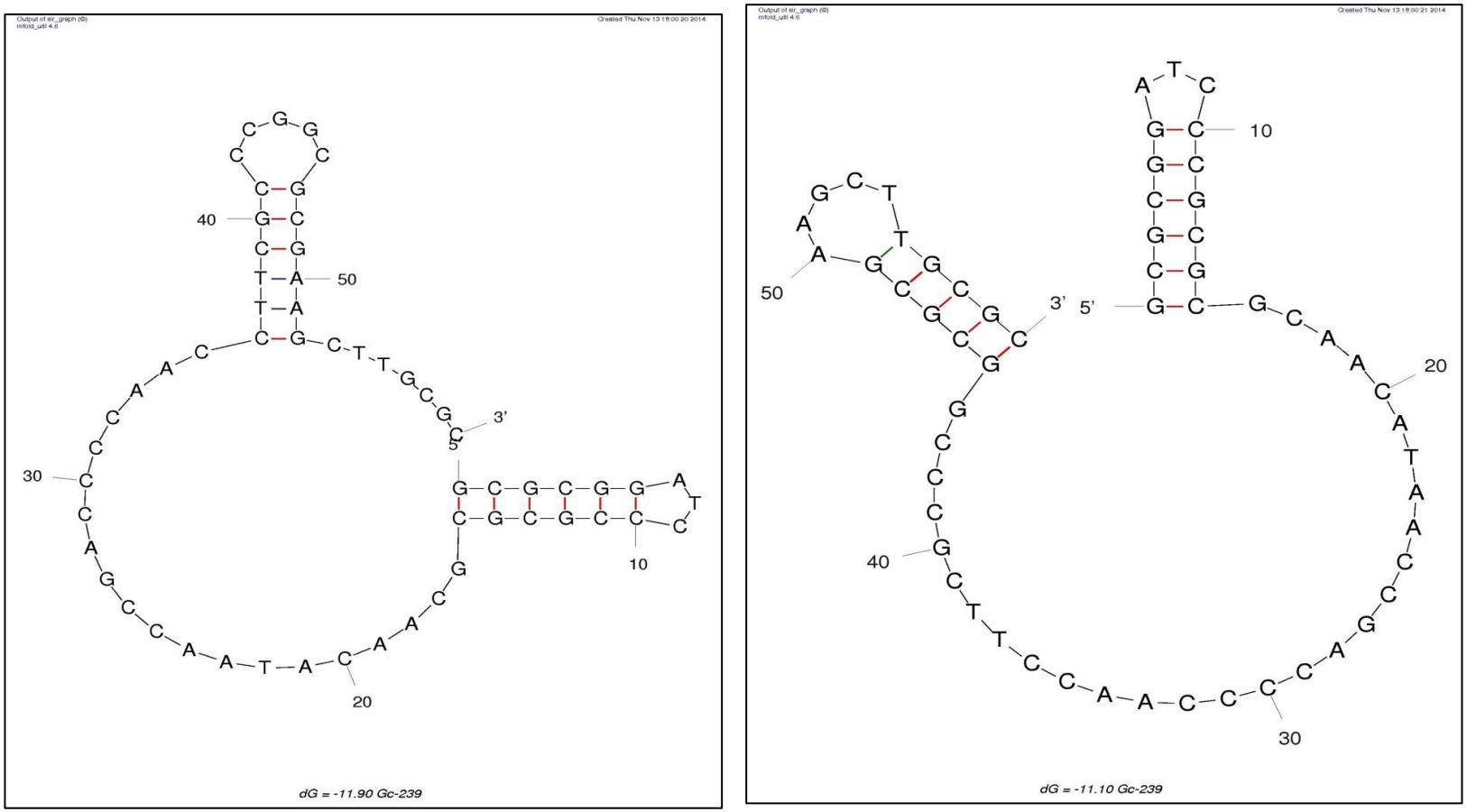


Figure 3.15 (C): Secondary Structures of (a) Gc-239 structure 1 and (b) Gc-239 structure 2. The secondary structures of ssDNA sequences selected against Neu5Gc were predicted using mfold DNA model software. The conditions required in the software are the nature of DNA sequence, which was considered to be linear. The folding temperature was kept constant at 37 °C as used in panning experiments. The ionic conditions used were 157 mM Na^{+} and 5.0 mM Mg^{++} for Neu5Gc aptamer prediction.

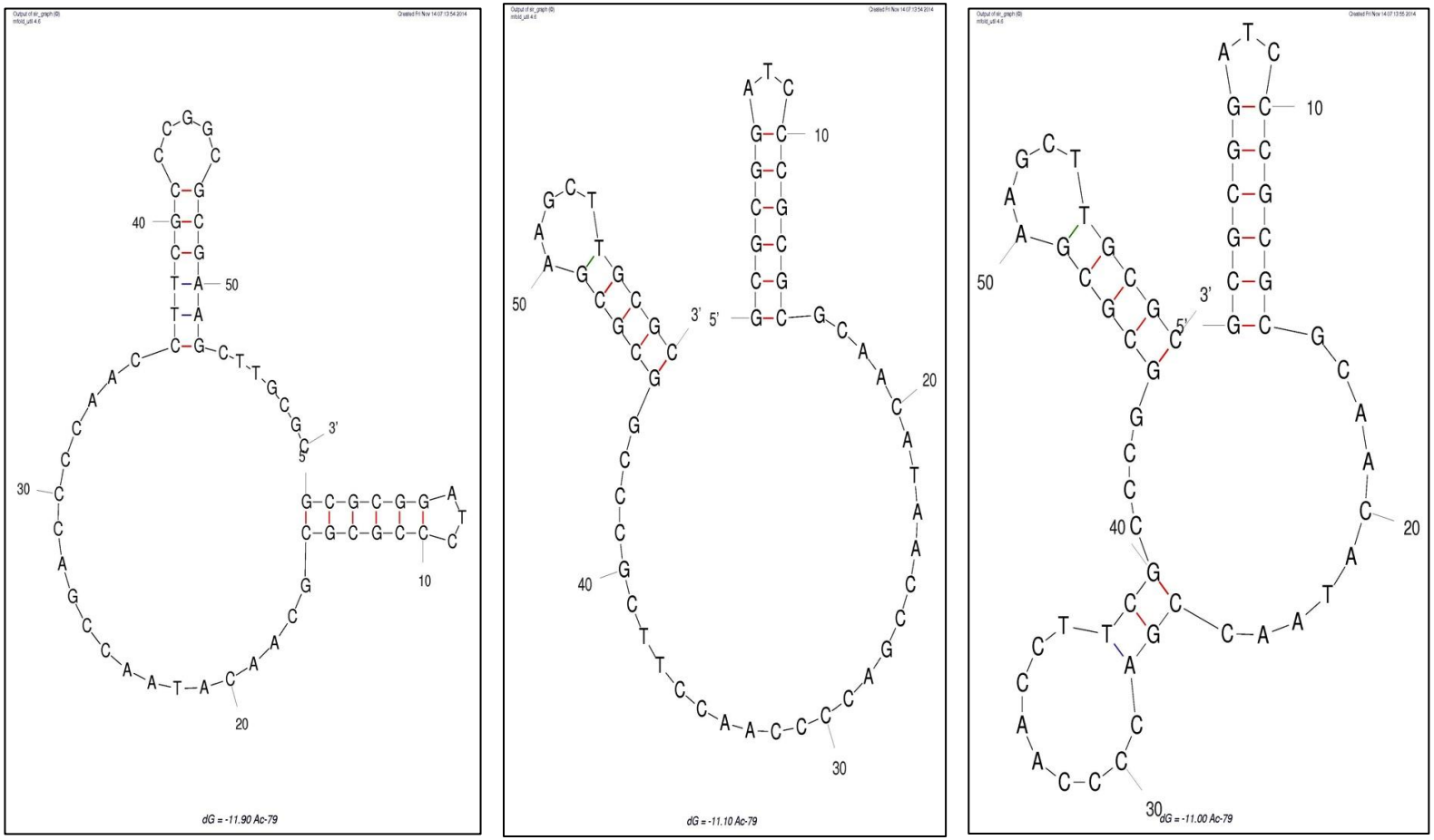


Figure 3.16 (A): Secondary Structures of (a) Ac-79 structure 1, (b) Ac-79 structure 2, (c) Ac-79 structure 3. The secondary structures of ssDNA sequences selected against Neu5Gc were predicted using mfold DNA model software. The conditions required in the software are the nature of DNA sequence, which was considered to be linear. The folding temperature was kept constant at 37 °C as used in panning experiments. The ionic conditions used were 157 mM Na^{+} and 5.0 mM Mg^{++} for Neu5Gc aptamer prediction.

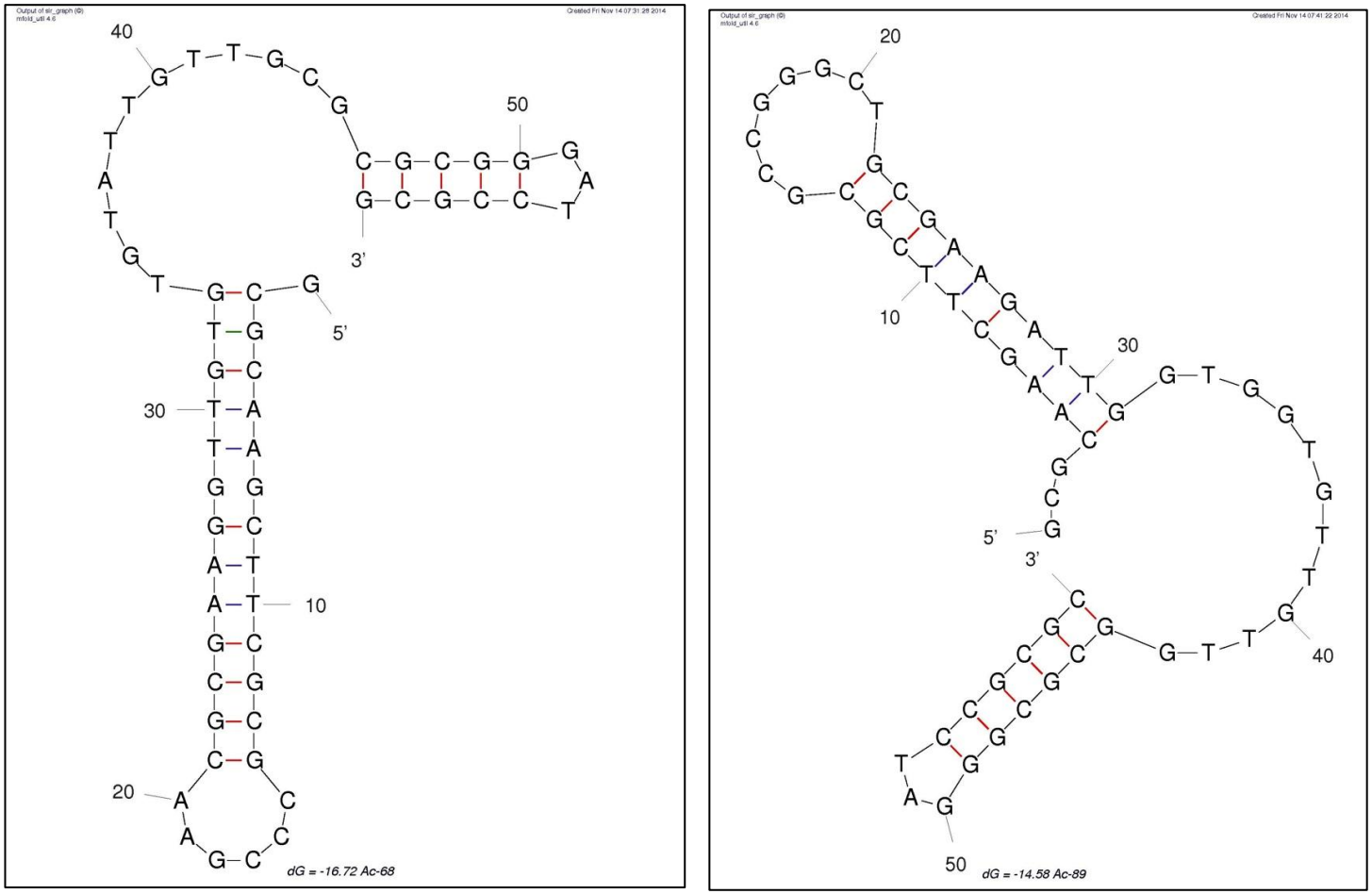


Figure 3.16: (B): Ac-68, **(C)**: Ac-89. The secondary structures of ssDNA sequences selected against Neu5Gc were predicted using mfold DNA model software. The conditions required in the software are the nature of DNA sequence, which was considered to be linear. The folding temperature was kept constant at 37 °C as used in panning experiments. The ionic conditions used were 157 mM Na^{+} and 5.0 mM Mg^{++} for Neu5Gc aptamer prediction.

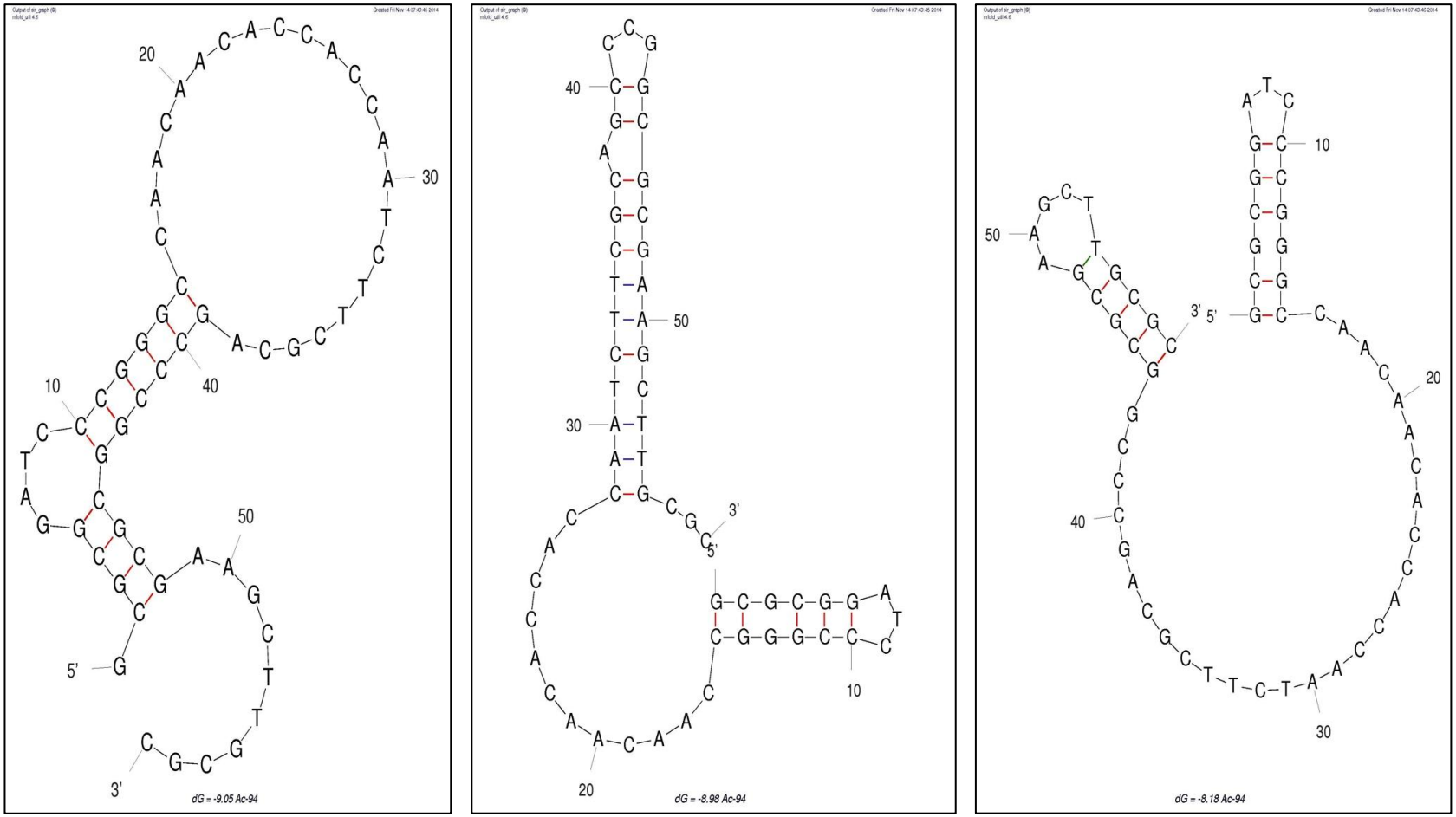


Figure 3.16 (D): Secondary Structures of (a) Ac-94 Structure 1, (b) Ac-94 structure 2, (c) Ac-94 structure 3. The secondary structures of ssDNA sequences selected against Neu5Gc were predicted using mfold DNA model software. The conditions required in the software are the nature of DNA sequence, which was considered to be linear. The folding temperature was kept constant at 37 °C as used in panning experiments. The ionic conditions used were 157 mM Na^{+} and 5.0 mM Mg^{++} for Neu5Gc aptamer prediction.

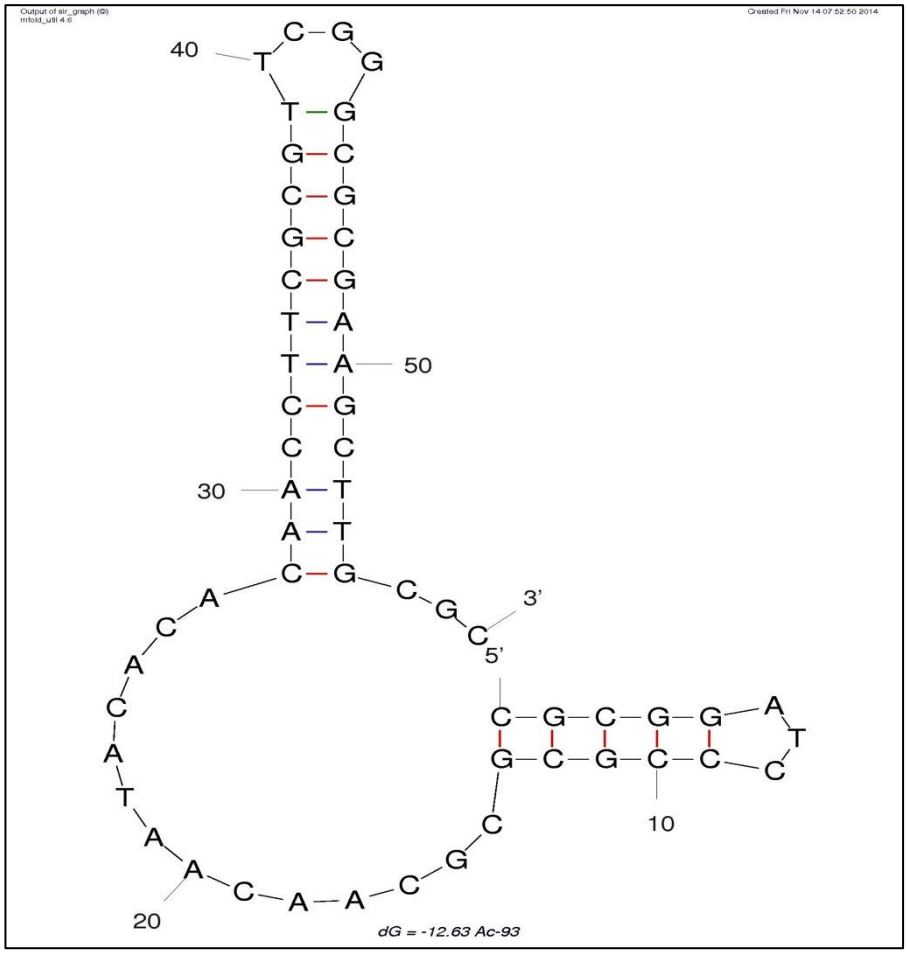


Figure 3.16 (E): Secondary Structures of Ac-93. The secondary structures of ssDNA sequences selected against Neu5Gc were predicted using mfold DNA model software. The conditions required in the software are the nature of DNA sequence, which was considered to be linear. The folding temperature was kept constant at 37 °C as used in panning experiments. The ionic conditions used were 157 mM Na^{+} and 5.0 mM Mg^{++} for Neu5Gc aptamer prediction.

3.4. Discussion

There is a considerable interest in Neu5Gc, commonly found as a terminal motif on glycoproteins and glycolipids in a range of non-human species. The ability to detect and the quantification of the motif on cells is important for their roles in bio therapeutic contamination (Ghaderi et al., 2010, Ghaderi et al., 2012), and in cancer progression (Padler-Karavani et al., 2011). This information has particularly highlighted the need for specific binding agents and a convenient analytical method for Neu5Gc. Recognition molecules can be used to develop cost effective and high-throughput methods that can be comparable to conventional carbohydrate analysis techniques such as mass spectrometry and chromatographic techniques (Shaw et al., 2008, Cunningham et al., 2012). Aptamer technology can offer a source of easily manufactured and reproducible glycan recognition molecules, which is exploited during this research.

Using SELEX, *in vitro* aptamer selection was performed against Neu5Ac and non-human sialic acid: Neu5Gc. The ssDNA library used to generate aptamers was 60 nucleotides long with a 30 base variable region, ensuring that it should neither be too small to generate enough diversity nor should be too long to reduce DNA product efficiency (Musheev and Krylov, 2006). For optimisation of the SELEX procedure, ssDNA preparation is an essential step to carry out successive rounds of amplification and selection. As it has been reported that frequent use of PCR and Gel electrophoresis reduces the ssDNA efficiency (Musheev and Krylov, 2006, Svobodova et al., 2012, Marimuthu et al., 2012, He et al., 2013), the initial ssDNA preparation method which involved 2-step quality and sise check of ssDNA (using both Agarose gel electrophoresis followed by PCR and then urea gel electrophoresis followed by PCR), was later optimised in to single step PCR amplification and ssDNA separation reducing the loss of ssDNA yield. Also, ssDNA extraction and purification methods were substituted and optimised that gave higher yield of ssDNA after subsequent rounds (Mallikaratchy et al., 2006).

The multiple numbers of positive selections of Neu5Gc-sp-biotin and Neu5Ac-sp-biotin were done to ensure that the strong binding DNA species should be retained and the low binding noise should be removed. Ten rounds of positive selections were

done against their respective targets a using biotin-streptavidin separation method based on the previous SELEX experiments (Gopinath, 2007, Svobodova et al., 2012, Marimuthu et al., 2012). 10 and 15 rounds of selections were done in the study of DNA aptamers selection against Neu5Gc and RNA aptamers selection against Neu5Ac, respectively (Cho et al., 2013, Gong et al., 2013).

As the two sialic acids, Neu5Gc and Neu5Ac are structurally similar, with the only difference being the oxygen atom at the 9^{th} carbon, the counter selection with the closely related sialic acid (For Neu5Gc binding aptamer selection: Neu5Ac counter selection was done, and vice versa) was done to ensure specificity against the desired target. In the earlier DNA aptamer report against Neu5Gc-BSA, no counter selection against Neu5Ac was performed, they only did counter selection against BSA (Gong et al., 2013), and no counter selection was reported in the RNA aptamer selection against Neu5Ac (Cho et al., 2013). The third counter selection with HSA and BSA was done to ensure that during binding assays, aptamers should not show non specific binding against these two common blocking agents. Also, sialic acid conjugates of BSA and HSA will be useful for immobilisation of sialic acids in the binding assays. Therefore, a counter selection with BSA/HSA was done to remove nonspecific binding of ssDNA molecules with these proteins. The quality and quantity of ssDNA after each round was analysed to ensure PCR contaminant carryover from each round should be the smallest.

The DNA pools demonstrating highest binding were rounds 7 and 9 against Neu5Gc-sp-bt, whereas in the previous study, the high binding plateau was observed from round 13 to 15 (Gong et al., 2013). For DNA aptamer selection against Neu5Ac, rounds 8 and 10 showed high binding and selected for cloning and sequencing, while in RNA aptamer selection against Neu5Ac round 10 showed highest binding and was selected for cloning and sequencing (Cho et al., 2013).

A total of 468 clones were obtained in Neu5Gc binding aptamer selection that were sent for sequencing, and three consensus sequences emerged, where two of the consensus account for more than 97% of the sequences. Previously only 36 clones were considered for sequencing and 13 different consensus sequences were obtained out of 36 cloned sequences (Gong et al., 2013). For Neu5Ac five different consensus

sequences were observed, however, after analysing their sequence motifs it shows that the entire consensus had 4-5 substitutions in their sequences and are highly conserved.

The ssDNA consensus from the alignment of cloned sequences obtained from this chapter will be synthesised commercially from MWG Eurofins and further used to characterise the recognition molecules by developing binding assays against Neu5Gc and Neu5Ac.

3.5. References

ANGATA, T. & VARKI, A. 2000. Siglec-7: a sialic acid-binding lectin of the immunoglobulin superfamily. *Glycobiology,* 10**,** 431-438.

BRODY, E. N. & GOLD, L. 2000. Aptamers as therapeutic and diagnostic agents. *Reviews in Molecular Biotechnology,* 74**,** 5-13.

BROOKS, C. L., SCHIETINGER, A., BORISOVA, S. N., KUFER, P., OKON, M., HIRAMA, T., MACKENZIE, C. R., WANG, L.-X., SCHREIBER, H. & EVANS, S. V. 2010. Antibody recognition of a unique tumour-specific glycopeptide antigen. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 10056-10061.

BYRNE, B., DONOHOE, G. G. & O,ÄÔKENNEDY, R. 2007. Sialic acids: carbohydrate moieties that influence the biological and physical properties of biopharmaceutical proteins and living cells. *Drug discovery today,* 12**,** 319-326.

CHO, S., LEE, B. Ä., CHO, B. Ä., KIM, J. Ä. & KIM, B. Ä. 2013. In vitro selection of sialic acid specific RNA aptamer and its application to the rapid sensing of sialic acid modified sugars. *Biotechnology and bioengineering,* 110**,** 905-913.

CROCKER, P. R. 2002. Siglecs: sialic-acid-binding immunoglobulin-like lectins in cell-cell interactions and signalling. *Current opinion in structural biology,* 12**,** 609-615.

CUNNINGHAM, S., STARR, E., SHAW, I., GLAVIN, J., KANE, M. & JOSHI, L. 2012. Development of a Convenient Competitive ELISA for the Detection of the Free and Protein-Bound Nonhuman Galactosyl-α-(1, 3)-Galactose Epitope Based on Highly Specific Chicken Single-Chain Antibody Variable-Region Fragments. *Analytical chemistry,* 85**,** 949-955.

DEMARCO, M. L. & WOODS, R. J. 2008. Structural glycobiology: a game of snakes and ladders. *Glycobiology,* 18**,** 426-440.

DIAZ, S. L., PADLER-KARAVANI, V., GHADERI, D., HURTADO-ZIOLA, N., YU, H., CHEN, X., BRINKMAN-VAN DER LINDEN, E. C., VARKI, A. & VARKI, N. M. 2009. Sensitive and specific detection of the non-human sialic Acid N-glycolylneuraminic acid in human tissues and biotherapeutic products. *PLoS One,* 4**,** e4241.

ELLINGTON, A. D. & SZOSTAK, J. W. 1990. In vitro selection of RNA molecules that bind specific ligands. *Nature,* 346**,** 818-822.

GHADERI, D., TAYLOR, R. E., PADLER-KARAVANI, V., DIAZ, S. & VARKI, A. 2010. Implications of the presence of N-glycolylneuraminic acid in recombinant therapeutic glycoproteins. *Nature biotechnology,* 28**,** 863-867.

GHADERI, D., ZHANG, M., HURTADO-ZIOLA, N. & VARKI, A. 2012. Production platforms for biotherapeutic glycoproteins. Occurrence, impact, and challenges of non-human sialylation. *Biotechnology and Genetic Engineering Reviews,* 28**,** 147-176.

GONG, S., REN, H.-L., TIAN, R.-Y., LIN, C., HU, P., LI, Y.-S., LIU, Z.-S., SONG, J., TANG, F. & ZHOU, Y. 2013. A novel analytical probe binding to a potential carcinogenic factor of N-glycolylneuraminic acid by SELEX. *Biosensors and Bioelectronics*.

GOPINATH, S. C. B. 2007. Methods developed for SELEX. *Analytical and bioanalytical chemistry,* 387**,** 171-182.

GREGORIADIS, G., JAIN, S., PAPAIOANNOU, I. & LAING, P. 2005. Improving the therapeutic efficacy of peptides and proteins: a role for polysialic acids. *International journal of pharmaceutics,* 300**,** 125-130.

HE, C.-Z., ZHANG, K.-H., WANG, T., WAN, Q.-S., HU, P.-P., HU, M.-D., HUANG, D.-Q. & LV, N.-H. 2013. Single-primer-limited amplification: A method to generate random single-stranded DNA sub-library for aptamer selection. *Analytical biochemistry,* 440**,** 63-70.

HEDLUND, M., PADLER-KARAVANI, V., VARKI, N. M. & VARKI, A. 2008. Evidence for a human-specific mechanism for diet and antibody-mediated inflammation in carcinoma progression. *Proceedings of the National Academy of Sciences,* 105**,** 18936-18941.

HERMANN, T. & PATEL, D. J. 2000. Adaptive recognition by nucleic acid aptamers. *Science,* 287**,** 820-825.

HOKKE, C. H., BERGWERFF, A. A., VAN DEDEM, G. W., VAN OOSTRUM, J., KAMERLING, J. P. & VLIEGENTHART, J. F. 1990. Sialylated carbohydrate chains of recombinant human glycoproteins expressed in Chinese hamster ovary cells contain traces of *N*-glycolylneuraminic acid. *FEBS letters,* 275**,** 9-14.

HSU, K.-L., GILDERSLEEVE, J. C. & MAHAL, L. K. 2008. A simple strategy for the creation of a recombinant lectin microarray. *Molecular BioSystems,* 4**,** 654-662.

HUANG, C.-C. & CHANG, H.-T. 2008. Aptamer-based fluorescence sensor for rapid detection of potassium ions in urine. *Chem. Commun.***,** 1461-1463.

JAYASENA, S. D. 1999. Aptamers: an emerging class of molecules that rival antibodies in diagnostics. *Clinical chemistry,* 45**,** 1628-1650.

JELLINEK, D., GREEN, L. S., BELL, C. & JANJIC, N. 1994. Inhibition of receptor binding by high-affinity RNA ligands to vascular endothelial growth factor. *Biochemistry,* 33**,** 10450-10456.

JEONG, S., EOM, T.-Y., KIM, S.-J., LEE, S.-W. & YU, J. 2001. *in vitro* Selection of the RNA Aptamer against the Sialyl Lewis X and Its Inhibition of the Cell Adhesion. *Biochemical and biophysical research communications,* 281**,** 237-243.

KILCOYNE, M. & JOSHI, L. 2007. Carbohydrates in therapeutics. *Cardiovascular & Hematological Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Cardiovascular & Hematological Agents),* 5**,** 186-197.

KIM, Y., DENNIS, D., MOREY, T., YANG, L. & TAN, W. 2010. Engineering dendritic aptamer assemblies as superior inhibitors of protein function. *Chemistry, an Asian journal,* 5**,** 56.

KUMARI, K., GULATI, S., SMITH, D. F., GULATI, U., CUMMINGS, R. D. & AIR, G. M. 2007. Receptor binding specificity of recent human H3N2 influenza viruses. *Virol J,* 4**,** 422X-4.

KUSSER, W. 2000. Chemically modified nucleic acid aptamers for in vitro selections: evolving evolution. *Reviews in Molecular Biotechnology,* 74**,** 27-38.

LARKIN, M., BLACKSHIELDS, G., BROWN, N., CHENNA, R., MCGETTIGAN, P. A., MCWILLIAM, H., VALENTIN, F., WALLACE, I. M., WILM, A., LOPEZ, R., THOMPPSON, J. D., GIBSON, T. J. & HIGGINS, D. G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics,* 23**,** 2947-2948.

LOFLING, J. C., PATON, A. W., VARKI, N. M., PATON, J. C. & VARKI, A. 2009. A dietary non-human sialic acid may facilitate hemolytic-uremic syndrome. *Kidney international,* 76**,** 140-144.

LOU, X., QIAN, J., XIAO, Y., VIEL, L., GERDON, A. E., LAGALLY, E. T., ATZBERGER, P., TARASOW, T. M., HEEGER, A. J. & SOH, H. T. 2009. Micromagnetic selection of aptamers in microfluidic channels. *Proceedings of the National Academy of Sciences,* 106**,** 2989-2994.

LU, D. Y., LU, T. R. & WU, H.-Y. 2011. Antimetastatic therapy targeting aberrant sialylation profiles in cancer cells. *Drugs and Therapy Studies,* 1**,** e12.

MA, H., ZHOU, H., SONG, X., SHI, S., ZHANG, J. & JIA, L. 2014. Modification of sialylation is associated with multidrug resistance in human acute myeloid leukemia. *Oncogene*.

MALLIKARATCHY, P., STAHELIN, R. V., CAO, Z., CHO, W. & TAN, W. 2006. Selection of DNA ligands for protein kinase C-δ. *Chem. Commun.***,** 3229-3231.

MALYKH, Y. N., SCHAUER, R. & SHAW, L. 2001. *N*-Glycolylneuraminic acid in human tumours. *Biochimie,* 83**,** 623-634.

MARIMUTHU, C., TANG, T.-H., TOMINAGA, J., TAN, S.-C. & GOPINATH, S. C. 2012. Single-stranded DNA (ssDNA) production in DNA aptamer generation. *Analyst,* 137**,** 1307-1315.

MEHEDI MASUD, M., KUWAHARA, M., OZAKI, H. & SAWAI, H. 2004. Sialyllactose-binding modified DNA aptamer bearing additional functionality by SELEX. *Bioorganic & medicinal chemistry,* 12**,** 1111-1120.

MIN, K., JO, H., SONG, K., CHO, M., CHUN, Y.-S., JON, S., KIM, W. J. & BAN, C. 2011. Dual-aptamer-based delivery vehicle of doxorubicin to both PSMA (+) and PSMA (-) prostate cancers. *Biomaterials,* 32**,** 2124-2132.

MISONO, T. S. & KUMAR, P. K. 2005. Selection of RNA aptamers against human influenza virus hemagglutinin using surface plasmon resonance. *Analytical biochemistry,* 342**,** 312-317.

MUSHEEV, M. U. & KRYLOV, S. N. 2006. Selection of aptamers by systematic evolution of ligands by exponential enrichment: addressing the polymerase chain reaction issue. *Analytica chimica acta,* 564**,** 91-96.

NOGUCHI, A., MUKURIA, C., SUZUKI, E. & NAIKI, M. 1996. Failure of human immunoresponse to N-glycolylneuraminic acid epitope contained in recombinant human erythropoietin. *Nephron,* 72**,** 599-603.

O'REILLY, M. K. & PAULSON, J. C. 2009. Siglecs as targets for therapy in immune-cell-mediated disease. *Trends in pharmacological sciences,* 30**,** 240-248.

ORLANDI, P. A., KLOTZ, F. W. & HAYNES, J. D. 1992. A malaria invasion receptor, the 175-kilodalton erythrocyte binding antigen of Plasmodium falciparum recognises the terminal Neu5Ac (alpha 2-3) Gal-sequences of glycophorin A. *The Journal of cell biology,* 116**,** 901-909.

PADLER-KARAVANI, V., HURTADO-ZIOLA, N., PU, M., YU, H., HUANG, S., MUTHANA, S., CHOKHAWALA, H. A., CAO, H., SECREST, P. & FRIEDMANN-MORVINSKI, D. 2011. Human xeno-autoantibodies against a non-human sialic acid serve as novel serum biomarkers and immunotherapeutics in cancer. *Cancer research,* 71**,** 3352-3363.

PADLER-KARAVANI, V., YU, H., CAO, H., CHOKHAWALA, H., KARP, F., VARKI, N., CHEN, X. & VARKI, A. 2008. Diversity in specificity, abundance, and

composition of anti-Neu5Gc antibodies in normal humans: potential implications for disease. *Glycobiology,* 18**,** 818-830.

PAGRATIS, N. C. 1996. Rapid preparation of single stranded DNA from PCR products by streptavidin induced electrophoretic mobility shift. *Nucleic acids research,* 24**,** 3645-3646.

RUCKMAN, J., GREEN, L. S., BEESON, J., WAUGH, S., GILLETTE, W. L., HENNINGER, D. D., CLAESSON-WELSH, L. & JANJIC, N. 1998. 2'-Fluoropyrimidine RNA-based aptamers to the 165-amino acid form of vascular endothelial growth factor (VEGF165) Inhibition of receptor binding and VEGF-induced vascular permeability through interactions requiring the exon 7-encoded domain. *Journal of Biological Chemistry,* 273**,** 20556-20567.

SAMBROOK, J. & RUSSELL, D. W. 2006. Isolation of DNA fragments from polyacrylamide gels by the crush and soak method. *Cold Spring Harb Protoc*.

SAMRAJ, A., LAUBLI, H., VARKI, N. & VARKI, A. 2014. Involvement of a Non-Human Sialic Acid in Human Cancer. *Name: Frontiers in Oncology* [Online]. doi: 10.3389/fonc.2014.00033 [Accessed 12 December].

SANTALUCIA, J. 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences,* 95**,** 1460-1465.

SCHAUER, R. 2009. Sialic acids as regulators of molecular and cellular interactions. *Current opinion in structural biology,* 19**,** 507-514.

SHANGGUAN, D., LI, Y., TANG, Z., CAO, Z. C., CHEN, H. W., MALLIKARATCHY, P., SEFAH, K., YANG, C. J. & TAN, W. 2006. Aptamers evolved from live cells as effective molecular probes for cancer study. *Proceedings of the National Academy of Sciences,* 103**,** 11838-11843.

SHAW, I., O'REILLY, A., CHARLETON, M. & KANE, M. 2008. Development of a high-affinity anti-domoic acid sheep scFv and its use in detection of the toxin in shellfish. *Analytical chemistry,* 80**,** 3205-3212.

SLOTH, J. J., LARSEN, E. H. & JULSHAMN, K. R. 2003. Determination of organoarsenic species in marine samples using gradient elution cation exchange HPLC-ICP-MS. *Journal of Analytical Atomic Spectrometry,* 18**,** 452-459.

SONG, S., WANG, L., LI, J., FAN, C. & ZHAO, J. 2008. Aptamer-based biosensors. *TrAC Trends in Analytical Chemistry,* 27**,** 108-117.

SONNENBURG, J. L., ALTHEIDE, T. K. & VARKI, A. 2004. A uniquely human consequence of domain-specific functional adaptation in a sialic acid binding receptor. *Glycobiology,* 14**,** 339-346.

STOLTENBURG, R., REINEMANN, C. & STREHLITZ, B. 2007. SELEX- a (r) evolutionary method to generate high-affinity nucleic acid ligands. *Biomolecular Engineering,* 24**,** 381-403.

SVOBODOVA, M., PINTO, A., NADAL, P. & O,ÄÔSULLIVAN, C. 2012. Comparison of different methods for generation of single-stranded DNA for SELEX processes. *Analytical and bioanalytical chemistry,* 404**,** 835-842.

TAYLOR, R. E., GREGG, C. J., PADLER-KARAVANI, V., GHADERI, D., YU, H., HUANG, S., SORENSEN, R. U., CHEN, X., INOSTROZA, J. & NISET, V. 2010. Novel mechanism for the generation of human xeno-autoantibodies against the nonhuman sialic acid N-glycolylneuraminic acid. *The Journal of experimental medicine,* 207**,** 1637-1646.

TOUSI, F., HANCOCK, W. S. & HINCAPIE, M. 2011. Technologies and strategies for glycoproteomics and glycomics and their application to clinical biomarker research. *Analytical Methods,* 3**,** 20-32.

TUERK, C. & GOLD, L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science,* 249**,** 505-510.

VARKI, A. 1997. Sialic acids as ligands in recognition phenomena. *The FASEB journal,* 11**,** 248-255.

VARKI, A. 2007. Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. *Nature,* 446**,** 1023-1029.

VARKI, A. 2008. Sialic acids in human health and disease. *Trends in molecular medicine,* 14**,** 351-360.

WOCHNER, A. & GLOKLER, J. 2007. Nonradioactive fluorescence microtiter plate assay monitoring aptamer selections. *BioTechniques,* 42**,** 578.

YANG, Q., GOLDSTEIN, I. J., MEI, H. Y. & ENGELKE, D. R. 1998. DNA ligands that bind tightly and selectively to cellobiose. *Proceedings of the National Academy of Sciences,* 95**,** 5462.

YAO, L., KORTEWEG, C., HSUEH, W. & GU, J. 2008. Avian influenza receptor expression in H5N1-infected and noninfected human tissues. *The FASEB journal,* 22**,** 733-740.

ZAIA, J. 2008. Mass spectrometry and the emerging field of glycomics. *Chemistry & biology,* 15**,** 881-892.

ZUKER, M. 2000. Calculating nucleic acid secondary structure. *Current opinion in structural biology,* 10**,** 303-310.

ZUKER, M. 2003. Mfold web server for nucleic acid folding and hybridisation prediction. *Nucleic acids research,* 31**,** 3406-3415.

Chapter-4: *In vitro* characterisation of DNA aptamers against Neu5Gc and Neu5Ac

4.1. Introduction

A variety of bio-analytical assays are available to analyse and measure free or bound sialic acid levels in the tissues or in the serum, including colourimetric (Skoza and Mohos, 1976) and fluorometric approaches (Shukla and Schauer, 1982). More recently chromatography-based methods have been used widely for the detection and analyses of both Neu5Gc and Neu5Ac which include reverse phase HPLC with ultraviolet detection (UV), where released sialic acid in the glycoconjugates was hydrolysed and benzoylated for their determination (Karamanos et al., 1990), or the released sialic acid was o-phenylenediamine-2HCl (OPD) labelled to yield stable fluorescent quinoxaline derivatives for use in HPLC with fluorescence detection (Anumula, 1995, Kawabata et al., 2000). High-pH anion-exchange chromatography with pulsed amperometric detection (HPAEC/PAD), a high throughput method, was used to detect hydrolysed sialic acid labelled with 1, 2-diamino-4, 5-methylenedioxybenzene (DMB) (Rohrer et al., 1998, Hurum and Rohrer, 2011).

Chromatography in combination with mass spectrometry (MS) has also been used to detect sialic acids, e.g. liquid chromatography (LC) with electrospray ionisation-mass spectrometry (ESI-MS) was used to detect DMB labelled sialic acid with high affinity and specificity (Klein et al., 1997, Morimoto et al., 2001). The other version involves direct DMB labeling followed by fluorescence chromatography. This is one of the most common ways to visualise and identify sialic acids (Shah et al., 2003).

Recently, a free sialic acid quantification method for human plasma based on robust quinoxalinone derivatisation has been reported (Wang et al., 2014). There has been a tremendous advancement in the field of MS based methods and conventional HPLC, Gas Chromatography (GC) and GC-MS based methods have been largely replaced by LC-MS-MS methods (Shah et al., 2000, Kind and Fiehn, 2010). However, both chromatographic and MS techniques have their own disadvantages over other bio-analytical assays. Many of the chromatographic techniques have limitations in detecting low abundance, hydrophobic, acidic/basic proteins and glycans (Garbis et al., 2005). Most of these chromatographic and MS methods are expensive, are time consuming to analyse samples, and they require great expertise and utmost precision to run these chromatographic instruments (Lee et al., 1990). Factors such as pressure flow, chromatographic column quality and the capacity to absorb UV also limit the

true analyses of quantitative determination of target in the complex biological samples (Stanczyk and Clarke, 2010, Brown, 2012).

Binding assays or immunoassays on the other hand are more convenient, easy to use, faster, and relatively inexpensive. Binding assays have been compared with chromatographic or MS based method and were found to have similar deviation of the target values compared to the conventional chromatographic and mass spectroscopy methods (Farrell et al., 2012, Janssen et al., 2012).

Lectin based binding assays involving glycan-binding lectins have been used for the characterisation of cell surface glycans and glycoproteins, based on their selectivity for certain glycan structures and linkages present in the biological samples. One example of this application is the enzyme linked lectin assay (ELLA) used to detect sialylation of transferrin in serum samples (Gornik and Lauc, 2007). Lectins such as SNA-I, MAL, MAH and MAA (Geisler and Jarvis, 2011) have been used for binding against sialic acids but these lectins do not differentiate between the different forms of sialic acids. C type lectins also show binding towards the tumour associated sialylated Tn antigen, however, they have affinity for both, α-(2,6)-Neu5Ac-Tn and α-(2,6)-Neu5Gc-Tn in the micro molar range (Mortezai et al., 2013).

Antibodies are another class of recognition molecules that can bind both free glycans and glycoconjugates with high affinity and specificity. A chicken polyclonal antibody based immunohistochemistry assay was developed that demonstrates selectivity between both α-(2, 3) or α-(2, 6) linked Neu5Gc on the glycoconjugates (Tangvoranuntakul et al., 2003). However, this assay did not report any binding and specificity against free sialic acids and cannot be used for the quantification of free Neu5Gc in the biological sample.

A DNA aptamer which specifically binds to Neu5Gc-BSA has been used in an ELISA assay, showing the potential of aptamers as glycan diagnostic reagents (Gong et al., 2013). However, the cross-reactivity of this assay towards the closely related Neu5Ac sugar was not described. An RNA aptamer against Neu5Ac has also been reported, but it shows high cross-reactivity with Neu5Gc (Cho et al., 2013). There is the need to develop an aptamer-based assay that can differentiate between Neu5Gc

and Neu5Ac and not cross react with closely related sialic acids present on glycoproteins or glycoconjugates.

This chapter describes the characterisation of Neu5Ac and Neu5Gc binding aptamers, identified in the previous chapter, using a variety of different binding assay platforms.

4.2. Materials and Methods

4.2.1. Materials

ssDNA aptamers obtained from SELEX were synthesised by Eurofins MWG (Germany). Biotinylated spacer (sp-biotin), biotinylated Neu5Ac (Neu5Ac-sp-biotin) and biotinylated Neu5Gc (Neu5Gc-sp-biotin) were purchased from GlycoTech Inc. All neo-glycoconjugates and carbohydrates were purchased from Dextra (UK) and IsoSep (Sweden). MyOne streptavidin C1 Dynabeads was purchased from Life Technologies, (UK). All solutions were prepared with deionised water purified by ELGA water purification system (ELGA, UK). Quant-iT^{TM} OliGreen ssDNA reagent was purchased from Molecular Probes, Invitrogen. Fetuin, asialofetuin from fetal calf serum and normal human serum (male) were purchased from Sigma-Aldrich UK. BSA was purchased from Sigma Aldrich, UK. Mouse serum was purchased from DakoCytomation (Denmark A/S). Normal human serum samples were collected with signed informed consents from the healthy donors. It was in agreement with the Research Ethics Committee, National University of Ireland, Galway for research purpose only.

4.2.2. Periodic acid treatment of BSA (Periodate BSA)

BSA was periodate treated by dissolving BSA (5g / 100ml) in periodic acid solution (10 mM periodic acid in 0.1 M sodium acetate at pH 4.5) solution and incubated at room temperature for 6 h. The periodate-treated BSA was dialysed against four changes of water over two days at 4 °C. Finally, the periodated BSA was lyophilised and stored at 4 °C.

4.2.3. ELISA-based binding assays

Neu5Gc-PAA, Neu5Ac-PAA conjugate and PAA were immobilised in individual wells of a NUNC Maxisorp plate at a concentration of 10 µg/mL overnight at 4 ^{o}C. Excess antigen was removed by washing with Phosphate Buffer Saline (PBS) with 0.05% Tween-20 (PBS-T) three times and any remaining free binding sites were blocked with 3% periodate-treated BSA for 60 min at room temperature with constant shaking. The plates were then washed again three times with PBS-T. Each selected aptamer was prepared in Binding Buffer (PBS, 5 mM $MgCl_2$, 25mM NaCl

at pH 7.4) and unfolded by heat denaturing at 95 ^{o}C for 5 min and cooled on ice for 5 min and then subsequently allowed to warm up at room temperature for 10 min prior to assay. The aptamers were then incubated with immobilised Neu5Gc-PAA, Neu5Ac-PAA and PAA for 90 min on a plate shaker at room temperature. Any unbound aptamer was removed by wasing three times with PBS-T, and bound aptamer was detected with Streptavidin-poly-horseradish peroxidase (SA-Poly-HRP, Pierce, Boston, MA, USA), 1: 8000 dilution supplemented with 0.1 % BSA, at room temperature for 1 h. After washing, colour was developed using 50 μL of one-step 3,3',5,5' Tetramethylbenzidine (TMB), (Sigma, UK) substrate at room temperature for 20 min. The colour development was stopped by the addition of 50 μL of 2 N H_2SO_4 and the optical density of each well was read at 450 nm (Higuchi et al., 2008, Ferreira et al., 2008).

4.2.4. Surface Plasmon Resonance (SPR) analysis

The interaction affinity of the DNA aptamer against target sugars was determined using SPR utilising a Reichert SR7500DC SPR instrument (Reichert Technologies Life Sciences, New York, USA). Carboxymethyl Dextran Hydrogel Surface Sensor Chip and pre-immobilised Streptavidin/NeutrAvidin Sensor Chip were purchased from Reichert Technologies Life Sciences (New York, USA). All the glass tubes for sample preparation, and incubation were also purchased from Reichert Technologies Life Sciences (New York, USA). Neu5Gc-PAA and Neu5Ac-PAA were purchased from Dextra, UK. EDC and NHS were purchased from Sigma, UK.

Flow cells were washed in a 1% TritonX-100 solution at a flow rate of 100 μL/min for 10 min. The prism surface was washed with an ethanol-wetted cotton 0.75 μL of immersion oil (1.515 refractive index matching fluid), spread onto the prism between the holes on either side. The sensor chip was then placed onto the prism surface and the clamp closed. To check correct placement of sensor chip, the detector scan data was checked and found to be properly overlapped with left (blue) and right (red) flowcell (Figure 4.1).

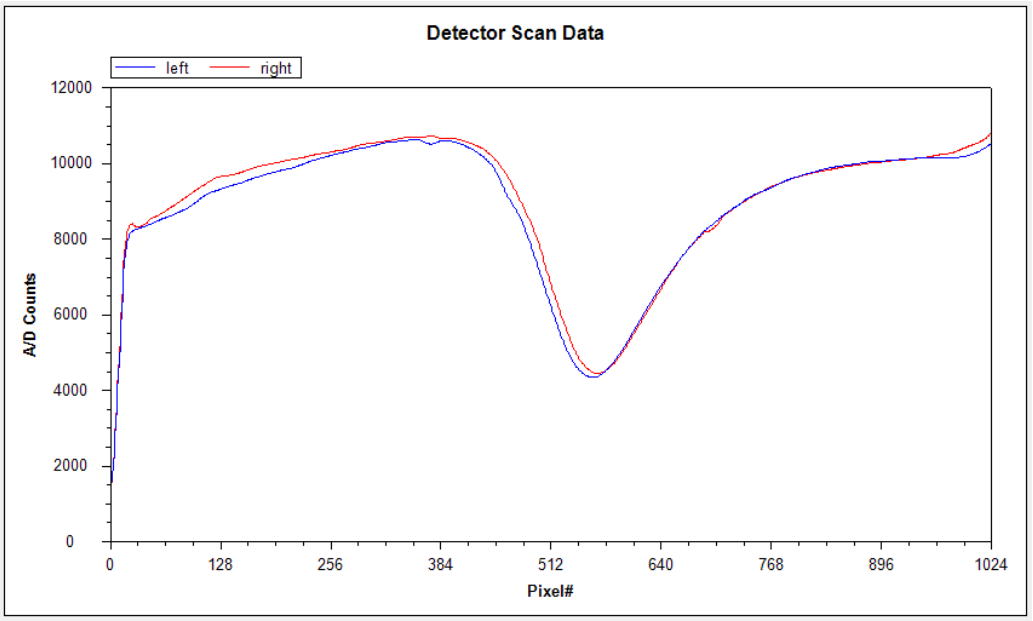


Figure 4.1: Detector scan data showing SPR minima. The SPR minima for left flow cell (blue) and right flow cell (red). The coherence of SPR minima of both channels indicates that pressure and flow rate is similar in both channels.

The temperature of the SPR machine was kept constant at 37 °C throughout all subsequent experiments. Buffers were prepared fresh daily, extensively degassed and filtered through 0.22 µM membrane which is essential to neglect any drift in the baseline. To optimise the pH of ligand for higher immobilisation efficiency onto the sensor chip surface, pH scouting for buffer was done. In pH scouting, two different pH buffers were used: 10 mM sodium acetate pH 4.0 (Figure 4.2A) and 10 mM sodium acetate pH 5.0 (Figure 4.2B). 10 mM sodium acetate pH 5.0 had a higher response unit (Figure 4.2B) and it was selected for ligand immobilisation.

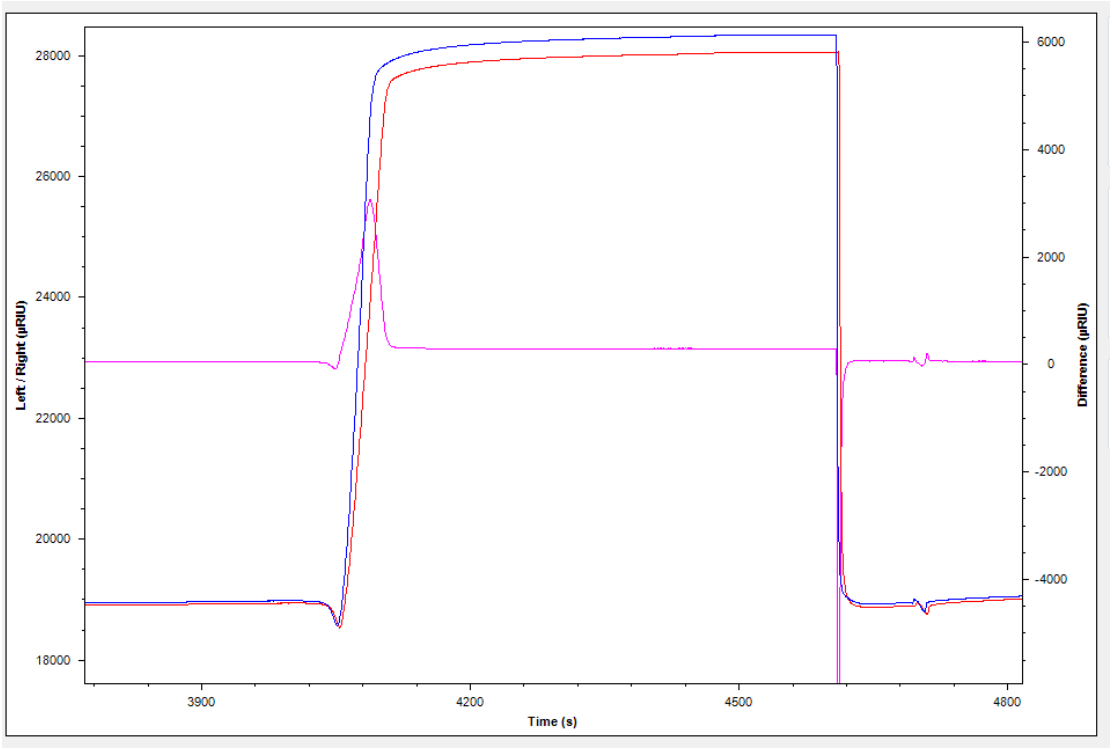


Figure 4.2 A: pH scouting of 10 mM sodium acetate buffer pH 4.0 with streptavidin ligand immobilised on to the amine coupled carboxymethylated sensor chips. Blue line (left flow cell) has a ligand immobilised, red line (right channel) is a blank flow cell, and the pink line is a difference of left and right flow cell. This buffer showed a response of 7 response units (pink line).

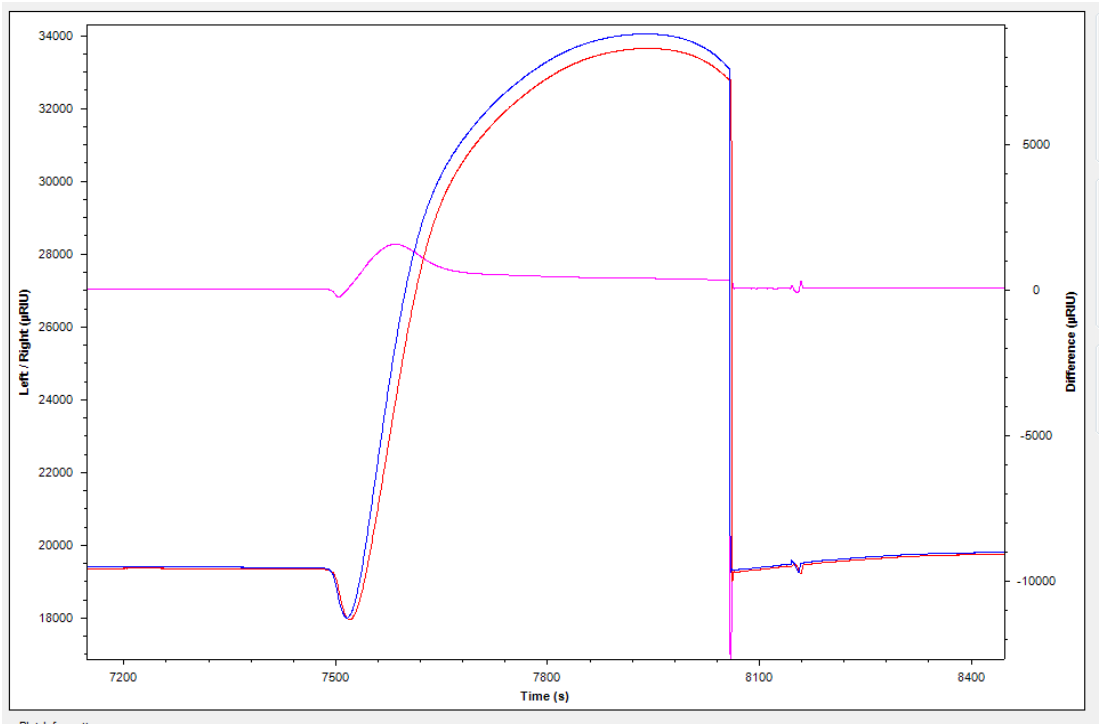


Figure 4.2 B: pH scouting of 10 mM sodium acetate buffer pH 5.0 with streptavidin ligand immobilised on to the amine coupled carboxymethylated sensor chips. Blue line (left flow cell) has ligand immobilised, red line (right channel) is a blank flow cell, and the pink line is a difference of left and right flow cell. This buffer showed a response of 874 response units (pink line).

Coupling of target molecules to the carboxymethylated dextran surface was carried out using a protocol provided by Reichert Technologies. 40 mg EDC and 10 mg NHS was diluted with 1 mL of water and injected over both channels for 10 minutes at a flow rate of 25 µ L/min. 100 µg/mL of ligand in 10 mM sodium acetate pH 5.0 was immobilised for 10 minutes at 20 µL/min over the left channel only. 1 M ethanolamine HCl pH 8.5 for blocking was injected over both channels for 10 minutes at 20 µ L/min. Sensor surfaces were regenerated after each binding cycle by subsequent injections of 10 mM glycine-HCl, pH 2.5 for 60 sec. After each cycle, the sensor chip was washed with 50 µL of washing buffer (PBS pH 7.4) at a flowrate of 25 µL for 30 s.

4.2.5. Bead-based fluorescence affinity and specificity analysis of aptamers

For the initial evaluation of the selected aptamers, a similar assay was performed as that in the ssDNA pool enrichment analyses against Neu5Gc and Neu5Ac. In the current assay, custom synthesised aptamers were used instead of the ssDNA pool, and for the eluted aptamer detection, both OliGreen dye and urea gel detection was used (Figure 4.3).

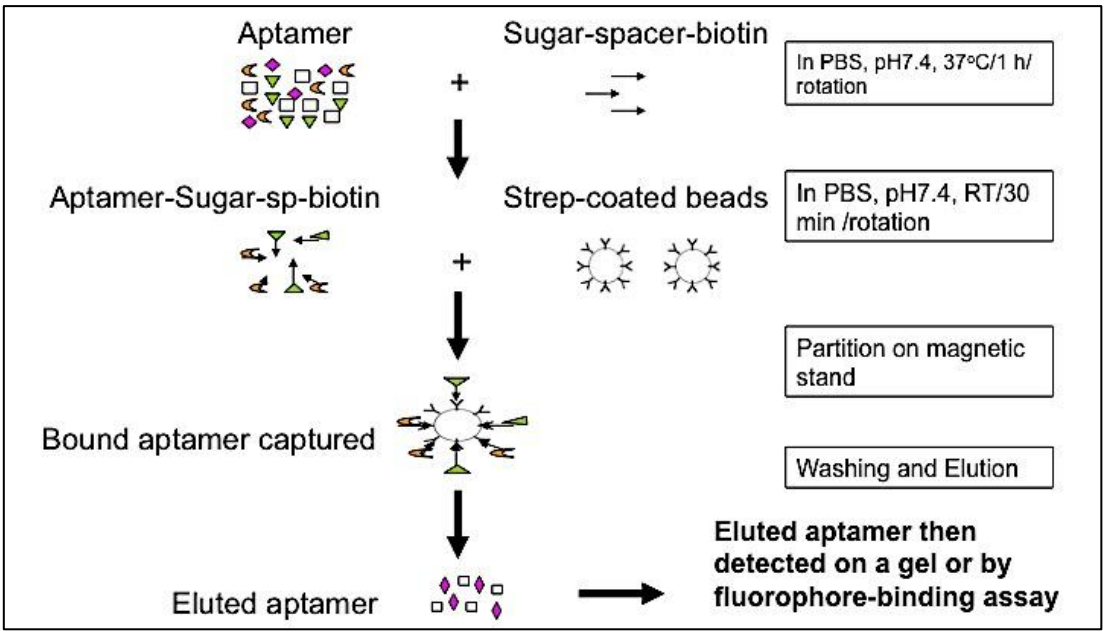


Figure 4.3: Pictorial representation of bead-based binding assay used for aptamer-biotinylated sugar interaction.

The target Neu5Gc-sp-biotin and control sp-biotin (2 µg/mL) were incubated with free ssDNA aptamer (500 nM) for 1 h at 37 °C on gentle rotation. Prior to incubation, each selected aptamer was prepared in binding buffer. The ssDNA was denatured at 95 °C for 5 min and subsequently cooled on ice for 5 min and later allowed to warm up at room temperature for 10 min. The ssDNA-biotinylated sugar complex was then immobilised onto streptavidin-coated Dynabeads for 30 min at room temperature on gentle rotation (approximately 40 rpm). The Dynabeads complex was washed 3 times with 200 µL of binding buffer. The bound ssDNA were then eluted using 50 µL of nuclease-free water at 95 °C for 5 min and fluorescence of ssDNA with OliGreen dye in TE buffer was measured at ex 485 nm and em 520 nm using a microtitre plate reader (SpectraMax Pro M5e, Molecular Devices).

4.2.6. Bead-based fluorescence aptamer binding inhibition assay

The target, Neu5Gc-sp-biotin (1 nmol or 2 μg/mL) was incubated with the selected aptamers (500nM) in the presence of free Neu5Gc ranging from 0 to 1 mg/mL in binding buffer for 1 h at 37 °C with gentle rotation. Prior to incubation, the aptamers were prepared in binding buffer for binding analysis by unfolding through heat denaturing at 95 °C for 5 min, cooling on ice for 10 min and then allowed to warm up to room temperature for 10 min. The aptamer-sugar complex was then incubated with 250 μg of streptavidin Dynabeads for 30 min at room temperature with gentle rotation. The Dynabead-aptamer-sugar complex was washed twice using 200 μL binding buffer and eluted using 50 μL of nuclease-free water at 95 °C for 5 min. Binding was determined using fluorescence by incubating the complex with OliGreen dye in TE buffer at ex 485nm and em 520 nm.

4.2.7. ssDNA detection on urea gel

7M urea/ 8% Polyacryl Amide were prepared by mixing, 2.67 mL of Protogel (National Diagnostics, UK) with 4.2 g of urea (Sigma, UK) and 2 mL of 5x TBE buffer and made to a volume of 9.89 mL with distilled water and then filtered with 0.45-μM filter membrane. After filtration, 100 μL of 10% (10 mg/100 mL) of ammonium per sulphate (APS) with 10 μL of N,N,N',N'-Tetramethylethylenediamine (TEMED) were added in the gel mixture.

The ssDNA was separated on the gel, where gel was run using 1 X TBE buffer (89mM Tris-HCl, 89 mM boric acid and 2 mM EDTA at pH 8.2) at 180 volts for 30 min. ssDNA was detected by immersing the gel in 50 mL of 1:10,000 volume of SYBRSafe DNA gel stain (Life Technologies, UK) in PBS buffer for 15 minutes with constant agitation. ssDNA was visualised on a ChemiDoc image analyser (BioRad, UK).

4.2.8. Biotinylation of glycoproteins

Bovine fetuin, bovine asialofetuin (Sigma) and normal mouse serum (Dako) were dissolved in 0.1M 2-*N*-morpholino- ethanesulfonic acid (MES) buffer, pH 4.7 (Thermo) to 10 mg/ml.

Glycoproteins and serum samples were conjugated with amino-PEG_3-biotin (Pierce) for aptamer binding assay according to the manual

(https://www.piercenet.com/instructions/2160750.pdf, 2013). 2 mg of protein were incubated with 60 µL amine-PEG_3-biotin (21 mg/ml) and 3ul EDC (19 mg/ml) in 0.5ml 0.1 M MES buffer, pH 4.9 (0.01 mole or 1.95 g of MES free acid in 100 mL of water) for 2 hours at room temperature. The modified proteins were desalted using an Amicon® ultra (Millipore, UK) column through exchange with binding buffer. Biotinylation of the protein was confirmed by ELISA using a commercially available biotinylated BSA (Thermo Scientific, UK) as control. The concentration of protein was quantified by Bradford protein estimation method (Peirce,Thermo Scientific, UK) (Bradford, 1976).

4.2.9. Confirmation of biotinylation of glycoprotein by ELISA

To confirm whether glycoproteins are biotinylated, different biotinylated glycoproteins (bt-Bovine fetuin, bt-Bovine asialofetuin, bt-Normal mouse serum and bt-Human serum) were immobilised in individual wells of a NUNC Maxisorp plate at a concentration of 100 µg/mL overnight at 4 ^{o}C. Excess antigen was removed by washing with Phosphate Buffer Saline with 0.05% Tween-20 (PBS-T) three times and any remaining free binding sites were blocked with 3% periodate-treated BSA for 60 min at room temperature with constant shaking. The plates were then washed again three times with PBS-T. The conjugated biotin was directly detected with streptavidin conjugated to Poly-HRP (SA-Poly-HRP, Pierce, Boston, MA, USA), 1: 8000 dilution supplemented with 0.1 % BSA, at room temperature for 1 h. The aptamers were then incubated with immobilised Neu5Gc-PAA, Neu5Ac-PAA and PAA for 90 min on a plate shaker at room temperature. Any unbound aptamer was removed by washing three times with PBS-T, and bound aptamer was detected with Streptavidin-poly-horseradish peroxidase (SA-Poly-HRP, Pierce, Boston, MA, USA), 1: 8000 dilution supplemented with 0.1 % BSA, at room temperature for 1 h. Again, washed with 200 µL of PBS-T three times and colour was developed using 50 µL of one-step 3,3',5,5' Tetramethylbenzidine (TMB), (Sigma, UK) substrate at room temperature for 20 min. The colour development was stopped by the addition of 2 N H_2SO_4 and the optical density of each well was read at 450 nm (Higuchi et al., 2008, Ferreira et al., 2008).

4.2.10. Statistical analysis and K_D calculation

The statistical analysis of data was performed using GraphPad Prism software for Mac version 6.0c (GraphPad Software Inc., SanDiego, CA, USA). All data represented as the average values of three replicate determinations. The results were expressed as mean values with standard error of mean (SEM). The K_D was calculated by fitting the data points by the non-linear regression analysis with the help of the following equation using GraphPad Prism software:

$$Y = B_{max} \times X/(K_d + X)$$

Where Y is specific binding, B_{max} is the maximum number of binding sites and X is concentration of the aptamer (Chen et al., 2009, Tang et al., 2007).

4.3. Results

4.3.1. Initial evaluation of selected Neu5Gc aptamers by fluorescent dye-linked aptamer assay

4.3.1.1. Neu5Gc

The three consensus sequences that emerged following selection against Neu5Gc aptamers Gc-224, Gc-101 and Gc-239 were synthesised for evaluation of their binding characteristics. Initially, they were tested at a single aptamer concentration in the same assay as employed during enrichment against spacer-biotin and Neu5Gc-spacer-biotin. Of the three sequences, aptamers Gc-224 and Gc-239 bound more strongly to Neu5Gc-spacer-biotin than to the spacer-biotin (Figure 4.4). Aptamer Gc-224 gave the highest binding response and hence was selected for further binding characterisation experiments.

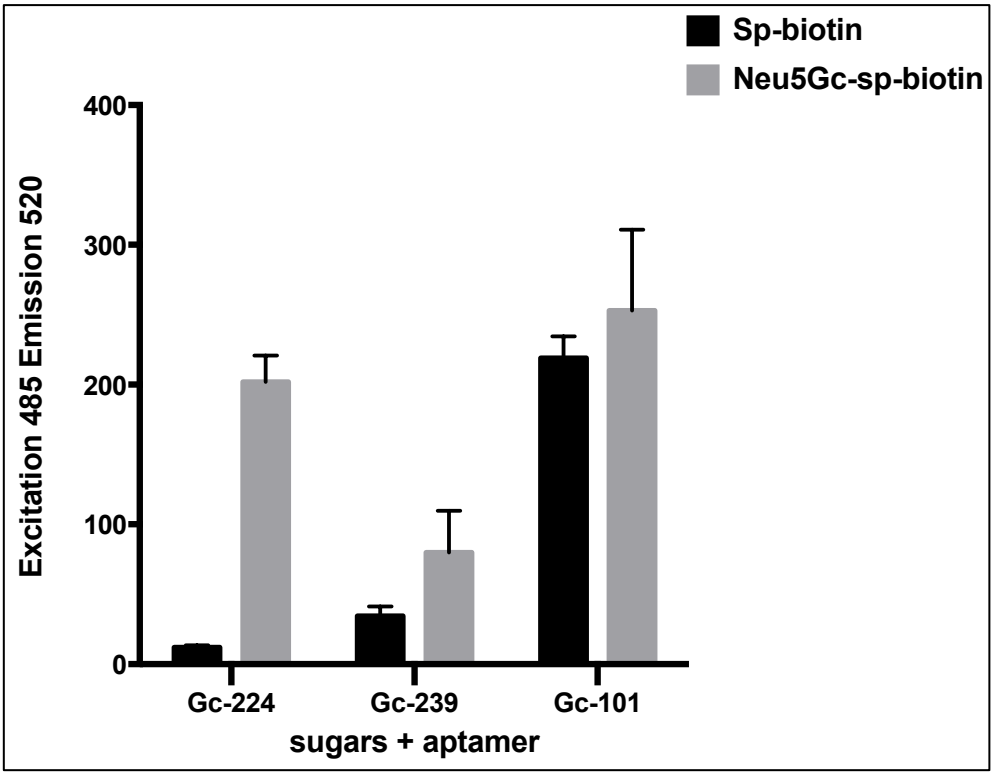


Figure 4.4: Specific binding profile of aptamer Gc-224, Gc-239 and Gc-101 against biotinylated Neu5Gc-sp-biotin and sp-biotin. 2 μM (1 nmol) of Neu5Gc-spacer-biotin and Spacer-biotin were immobilised on to the Dynabeads. 500 nM of aptamer was used for the binding reaction.

Aptamer Gc-224 was tested on urea gels to demonstrate its binding specificity to Neu5Gc-spacer-biotin, Neu5Ac-spacer-biotin and spacer-biotin. The aptamer Gc-224 showed a band in Neu5Gc-spacer-biotin lane, however there was no band in the Spacer-biotin and low intensity band in Neu5Ac-spacer-biotin (Figure 4.5). Aptamer Gc-224 demonstrated only negligible interaction with spacer-biotin relative to Neu5Gc-sp-biotin. Also, overall binding intensity of aptamer Gc-224 was higher for Neu5Gc-spacer-biotin than Neu5Ac-spacer-biotin. The traces of ladder image in the lane 7 are due to the spill over of DNA ladder solution from lane 8 into lane 7 (Figure 4.5).

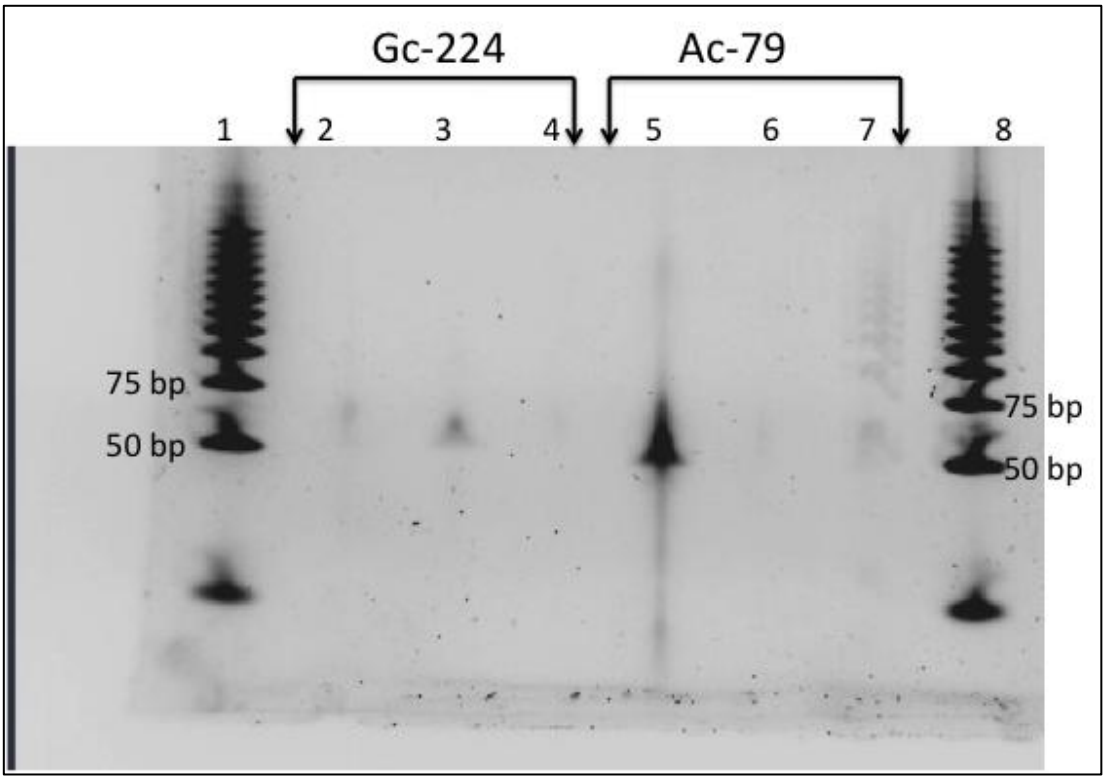


Figure 4.5: Eluted ssDNA aptamers Gc-224 and Ac-79 from the Dynabeads-based binding assay run on urea gel. Lanes 2 and 5, Neu5Ac-sp-biotin, lanes 3 and 6, Neu5Gc-sp-biotin, lanes 4 and 7, sp-biotin. Lanes 1 and 8 are 25bp DNA ladder is on either side of the gel. ssDNA bands of 60 bp sise are visible between 50 bp and 75 bp ladder.

4.3.1.2. Neu5Ac

Similarly, aptamer Ac-79 was tested on urea gels to demonstrate its binding specificity to Neu5Gc-sp-biotin, Neu5Ac-sp-biotin and Sp-biotin. Aptamer Ac-79

did not show any binding to Neu5Gc-spacer-biotin and Spacer-biotin. This aptamer was binding specifically to Neu5Ac-spacer-biotin over other immobilised glycans (Figure 4.5).

4.3.2. ELISA-based binding assay

Different variations of ELISA-based assays were tested to assess the binding of biotinylated aptamer Gc-224 against PAA and BSA conjugates of Neu5Gc. Initially, PAA conjugates were bound to ELISA plates and the PAA, Neu5Gc-PAA and Neu5Ac-PAA conjugates were incubated with and without aptamer Gc-224 (Figure 4.6). However, the background with PAA conjugate without aptamer was higher than the conjugates with aptamer. Also, there was not much difference in the aptamer binding against Neu5Gc-PAA and Neu5Ac-PAA (Figure 4.7). 3% milk and BSA blocking were tried to see if there is any binding difference due to blocking of the immobilised targets, however, the binding pattern observed looks similar and there was not any significant binding differences between Neu5Gc-PAA and PAA. Also, two different secondary antibodies, one from R&D systems and another from Thermo Scientific, were used in three different dilutions of 1:500, 1:1000 and 1:8000. Again, there was no difference in binding compared to the previous protocol (Figure 4.8). Finally, BSA and Neu5Gc-BSA were immobilised on the plate and incubated with the biotinylated Gc-224 aptamer (bt-Gc-224) (Figure 4.9). However, these results were similar to those of the PAA conjugates.

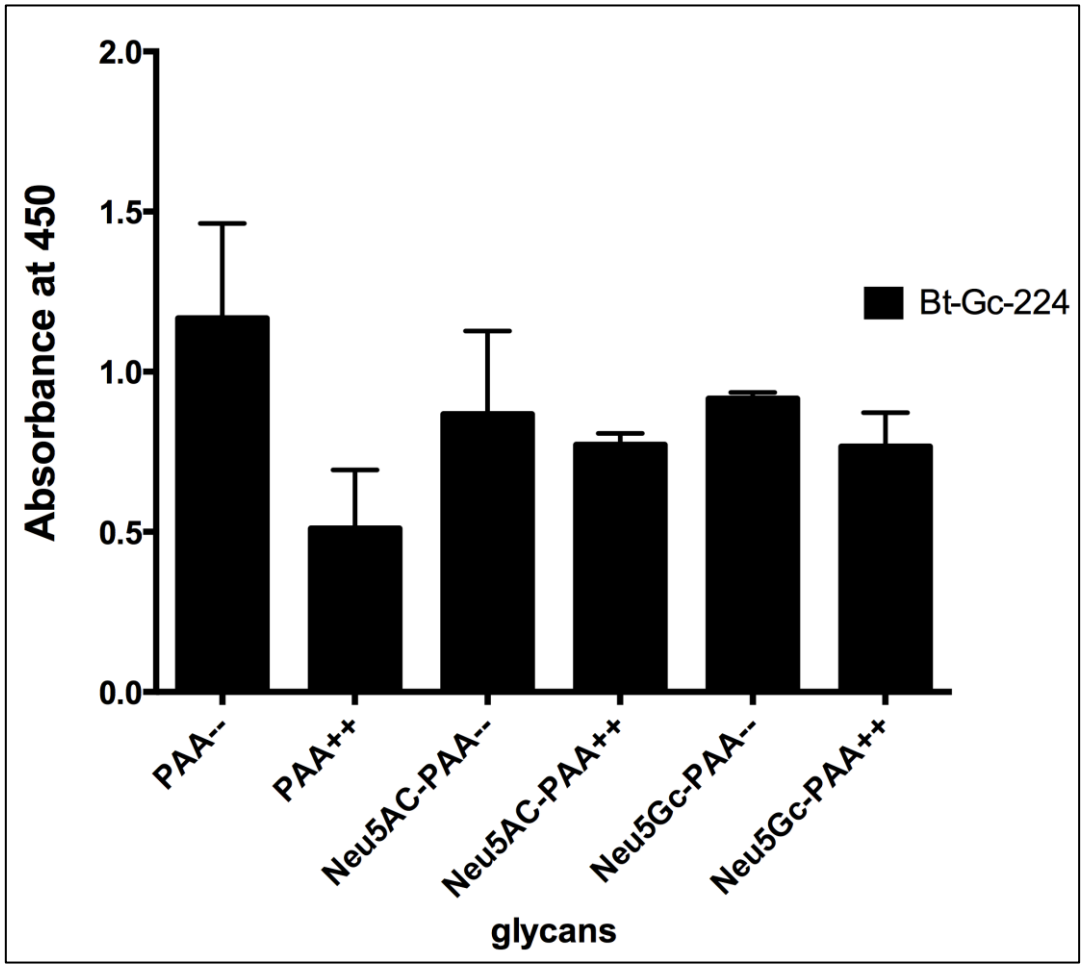


Figure 4.6: ELISA based analysis of biotinylated Gc-224 aptmer against PAA conjugates. (--) are the PAA conjuates incubated without aptamer. (++) are the PAA conjugates incubated with aptamer.

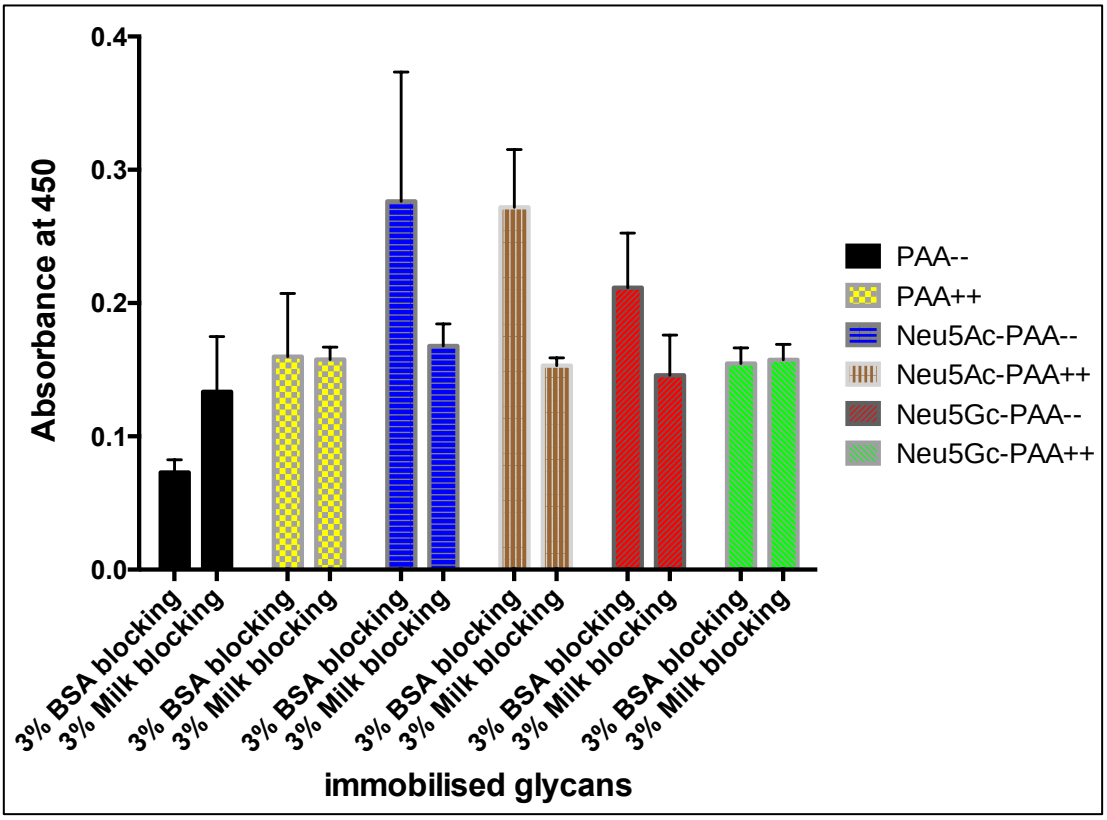


Figure 4.7: ELISA based analysis of biotinylated Gc-224 aptmer against PAA conjugates. Two different blocking conditions with 3% BSA blocking and 3% milk blocking were used in the assay. (--) are the PAA conjuates incubated without aptamer. (++) are the PAA conjugates incubated with aptamer.

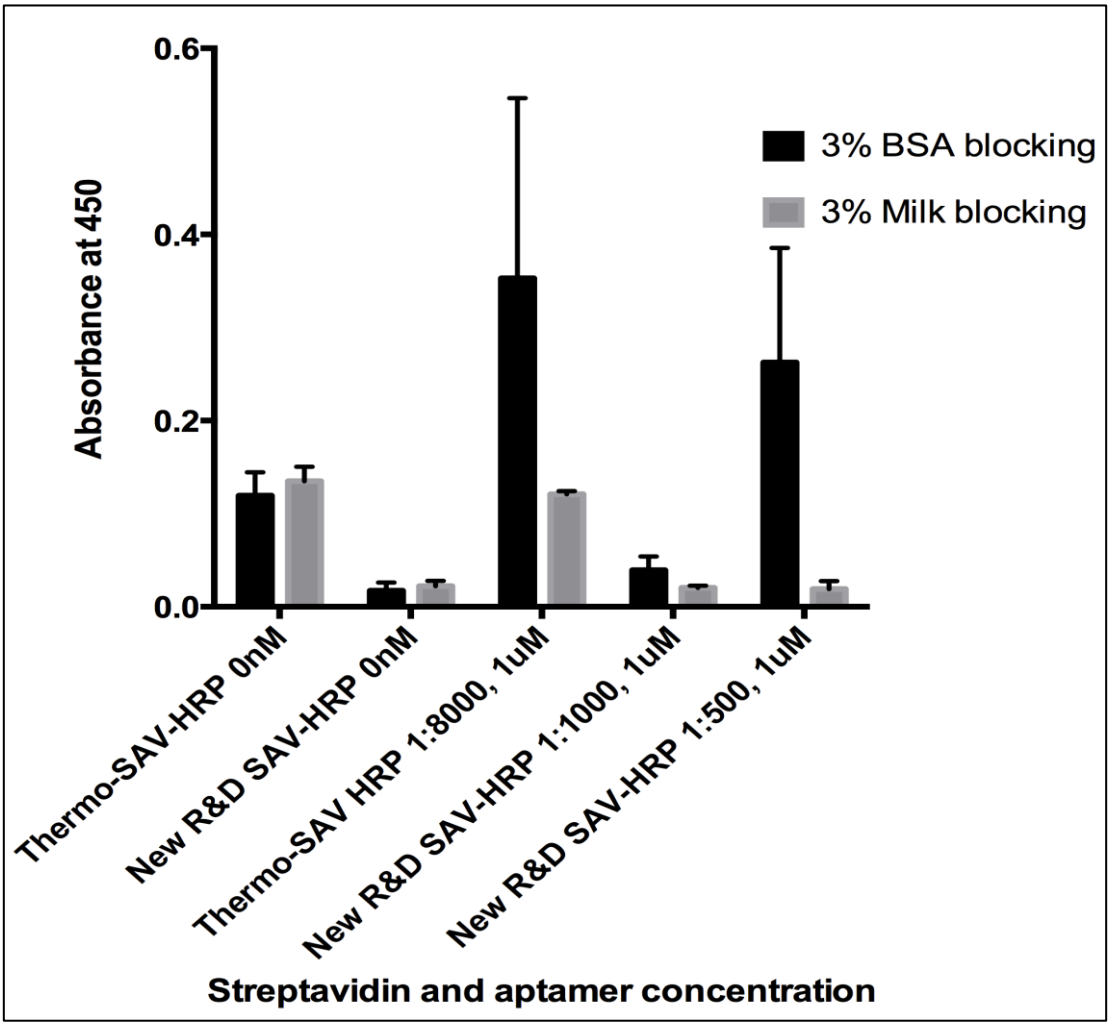


Figure 4.8: ELISA of biotinylated Gc-224 aptmer against Neu5Gc-PAA conjugates. Two different secondary antibody conjugated with HRP from Thermo Scientific streptavidin-HRP and R&D systems streptavidin were tested. Two concentrations of aptamer from 0 and 1000 nM (1 μM) were used against Neu5Gc-PAA.

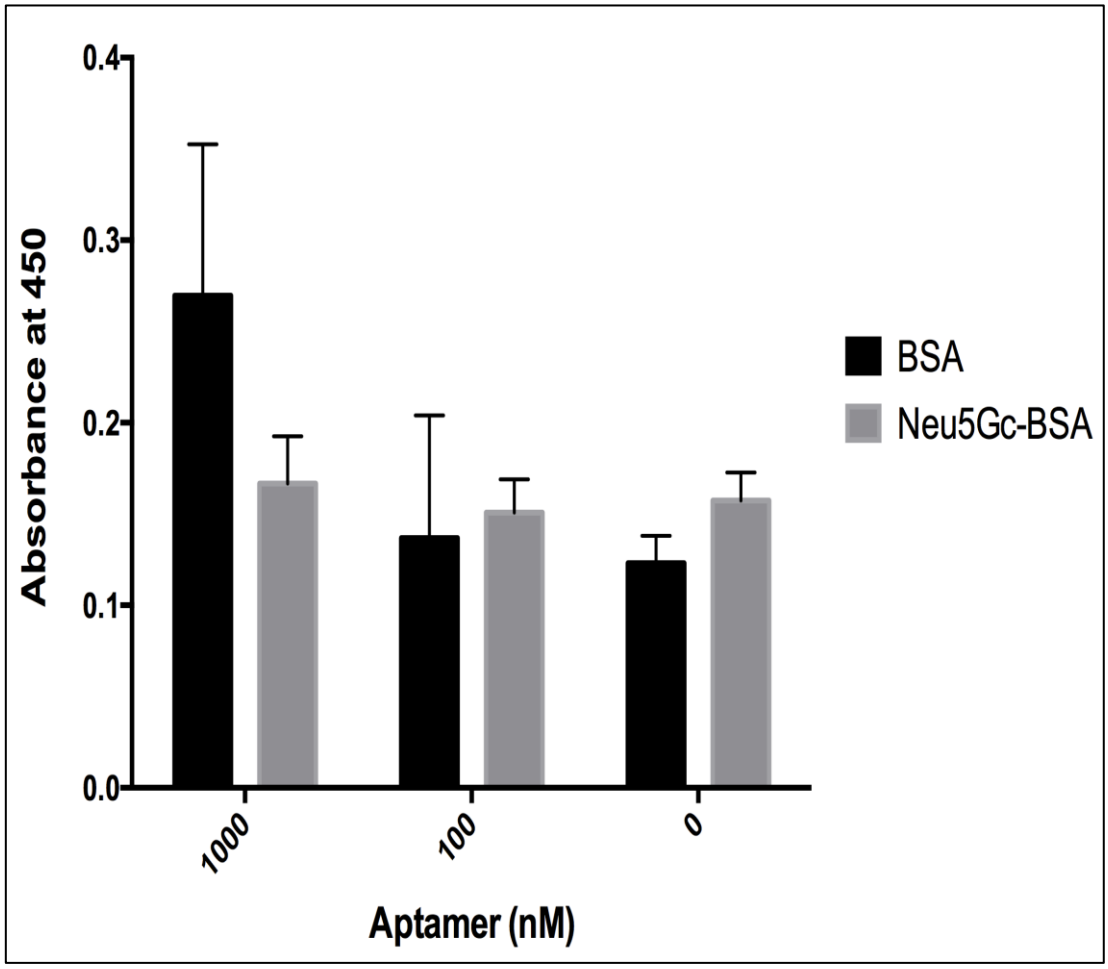
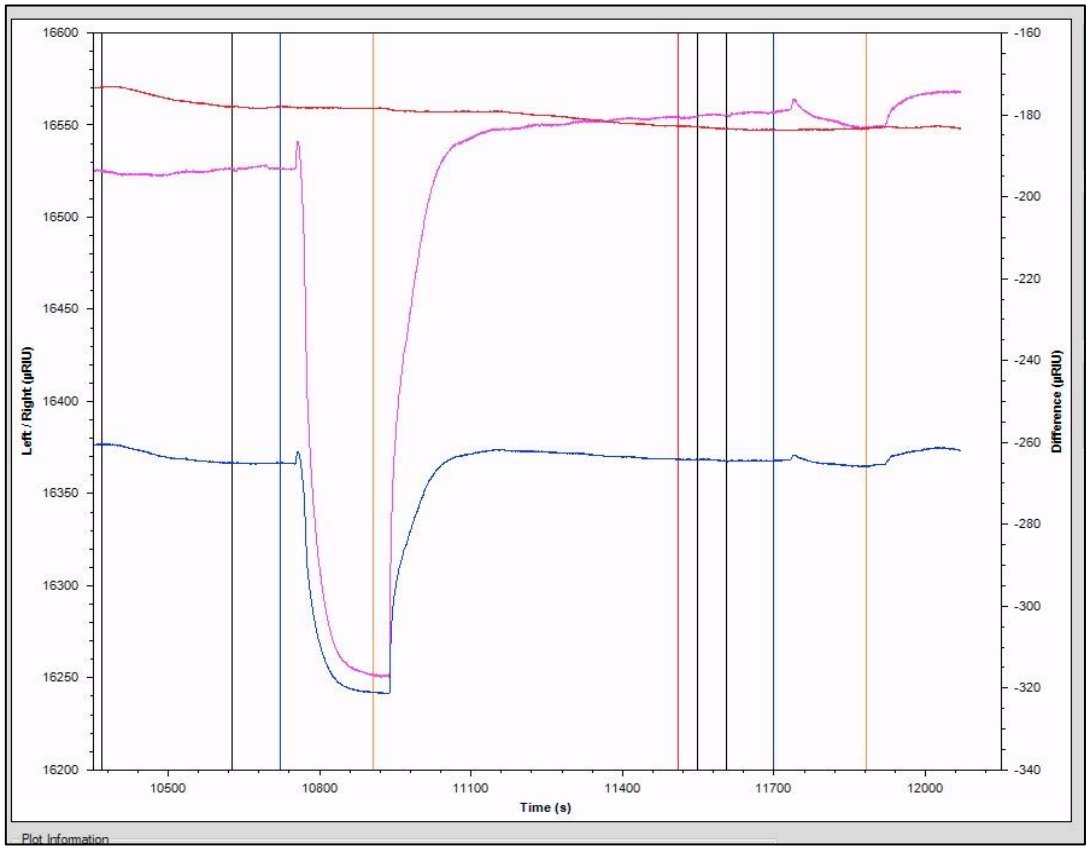


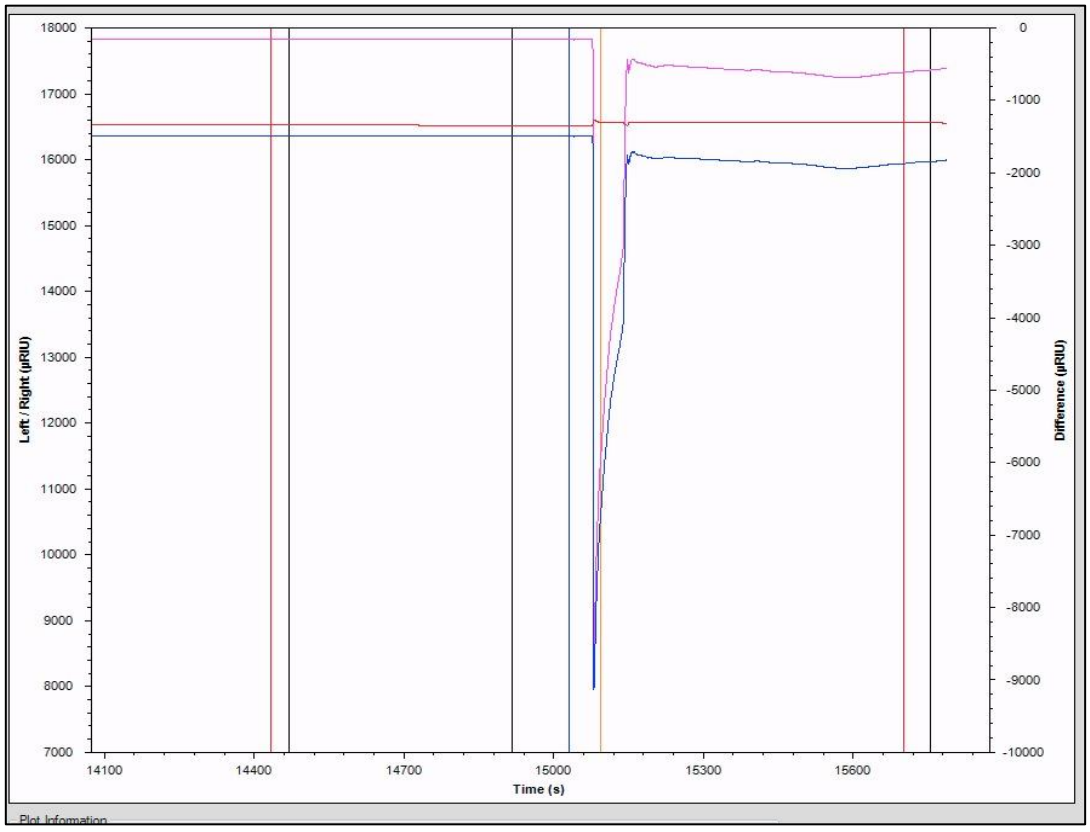
Figure 4.9: ELISA based analysis of biotinylated Gc-224 aptmer against BSA conjugates. Three different concentrations of aptamer from 0 to 1000 nM were used against Neu5Gc-BSA and BSA.

4.3.3. SPR analyses of glycan-aptamer interaction

Different approaches were taken to analyse aptamer-glycan interaction, where streptavidin was conjugated onto carboxymethylated dextran chips and Neu5Gc-sp-biotin was incubated with the streptavidin-coated chips, where streptavidin will bind to the biotinylated part of Neu5Gc-sp-biotin. On to these Neu5Gc-sp-biotin coated chips, unlabelled aptamers were bound to show binding responses of aptamers against Neu5Gc-sp-biotin. Alternatively, biotinylated aptamers Gc-224 (bt-Gc-224) were immobilised on to the streptavidin-coupled sensor and later free Neu5Gc interaction was measured against bt-Gc-224. In another approach, the Neu5Gc-PAA ligand was immobilised on to the carboxymethylated dextran chips and then unlabelled and biotinylated aptamers were bound to them.



A

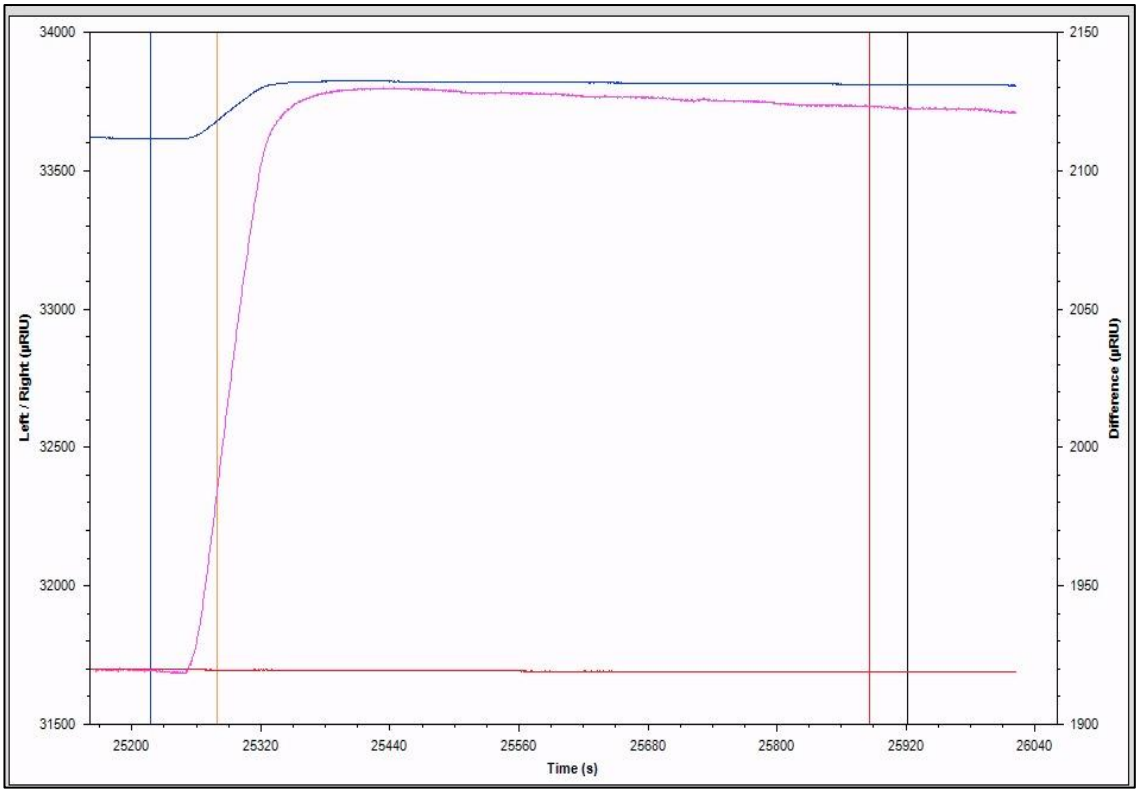


B

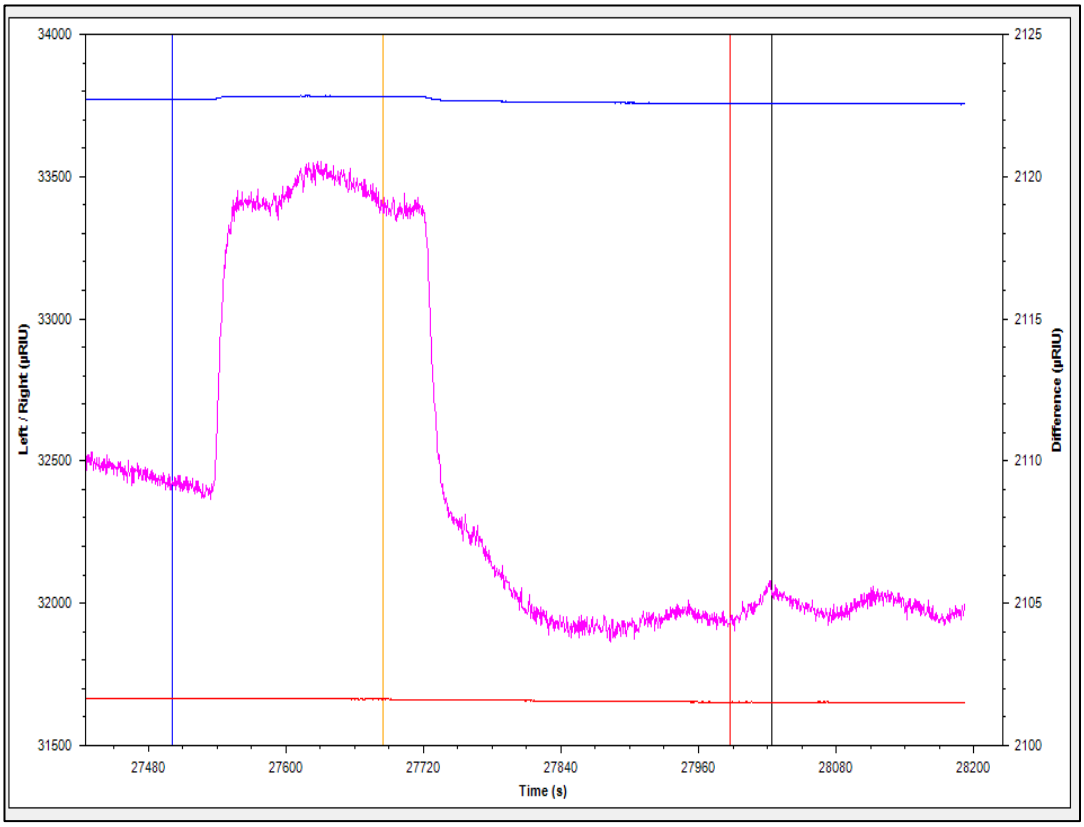
Figure 4.10: Response profile of aptamer glycan interaction on SPR. (A) Immobilisation of Neu5Gc-sp-biotin (100 µg/mL) on the streptavidin coated sensor

chip. Prior to immobilisation of NeuGc-sp-bt, streptavidin was immobilised onto carboxymethylated surface (B) Aptamer Gc-224 (1μM) in PBS buffer was flushed over biotinylated ligand at a flow rate of 20μL/min. Blue line (left flow cell) has ligand immobilised, red line (right channel) is a blank flow cell, and the pink line is a difference of left and right flow cell.

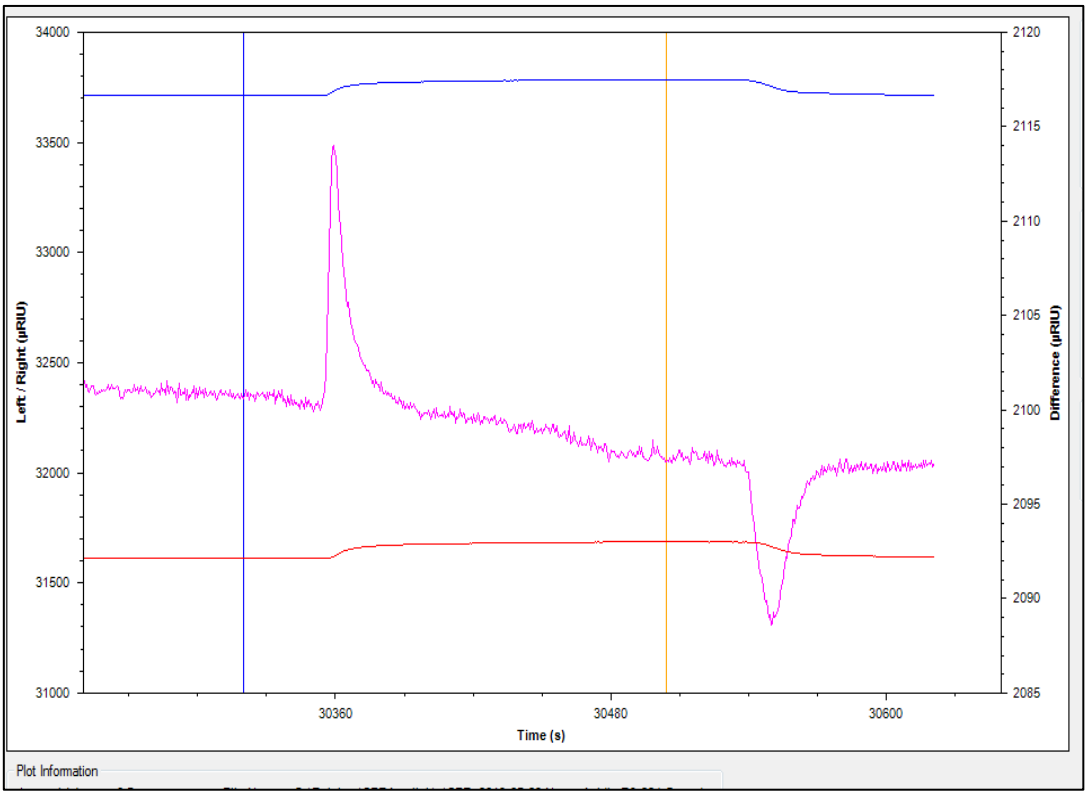
Prior to immobilisation of ligand Neu5Gc-sp-biotin on to the sensor chip, streptavidin was immobilised onto the carboxymethylated surface that gave a response of 321 RU. The ligand Neu5Gc-sp-biotin (100 μg/mL) was immobilised on the streptavidin coated sensor chip, which gave a response of 124 RU, demonstrating that Neu5Gc was immobilised on to the chip surface (Figure 4.10 A). However, when unlabelled aptamer Gc-224 (1 μM) in PBS buffer was flown over the aptamer, there was no binding reported between aptamer and Neu5Gc as the actual response unit was in negative values (Figure 4.10 B).



A



B



C

Figure 4.11: Response profile of aptamer glycan interaction on SPR. (A) Immobilisation of bt-Gc-224 (1 µM) on the streptavidin coated sensor chip. Prior to immobilisation of bt-Gc-224, streptavidin was immobilised onto carboxymethylated

surface. (B) Neu5Gc (100 μg/mL) and (C) (500 μg/mL) in PBS buffer was flown over biotinylated aptamer at a flow rate of 20 μL/min. Blue line (left flow cell) has ligand immobilised, red line (right channel) is a blank flow cell, and the pink line is a difference of left and right flow cell.

In the second experiment, biotinylated aptamers were immobilised onto the streptavidin-coupled surface, where four different injections of bt-Gc-224 (1μM) were flushed over streptavidin giving a response of 1170, 470, 322 and 205 RU respectively (Figure 4.11 A shows four bt-Gc-224 injections). Two concentration of free Neu5Gc (100 μg/mL and 500 μg/mL, respectively) were injected over bt-GC-224 to show aptamer-glycan interaction however, none of the glycan injections showed any binding as binding response were negative in both injections (Figure 4.11 B and C).

4.3.4. Evaluation of binding affinities of aptamers, Gc-224 and Ac-79 against Neu5Gc and Neu5Ac respectively, using bead-based fluorescence binding affinity method

4.3.4.1. Neu5Gc

A binding curve was established at varying concentrations of aptamer, Gc-224, from 0 to 625 nM and K_D values were estimated by linear regression. Binding intensity reached saturation at an aptamer concentration of approximately 500 nM and K_D was estimated as $2.83 * 10^7$ M^{-1} with a R^2 value of 0.95 (Figure 4.12).

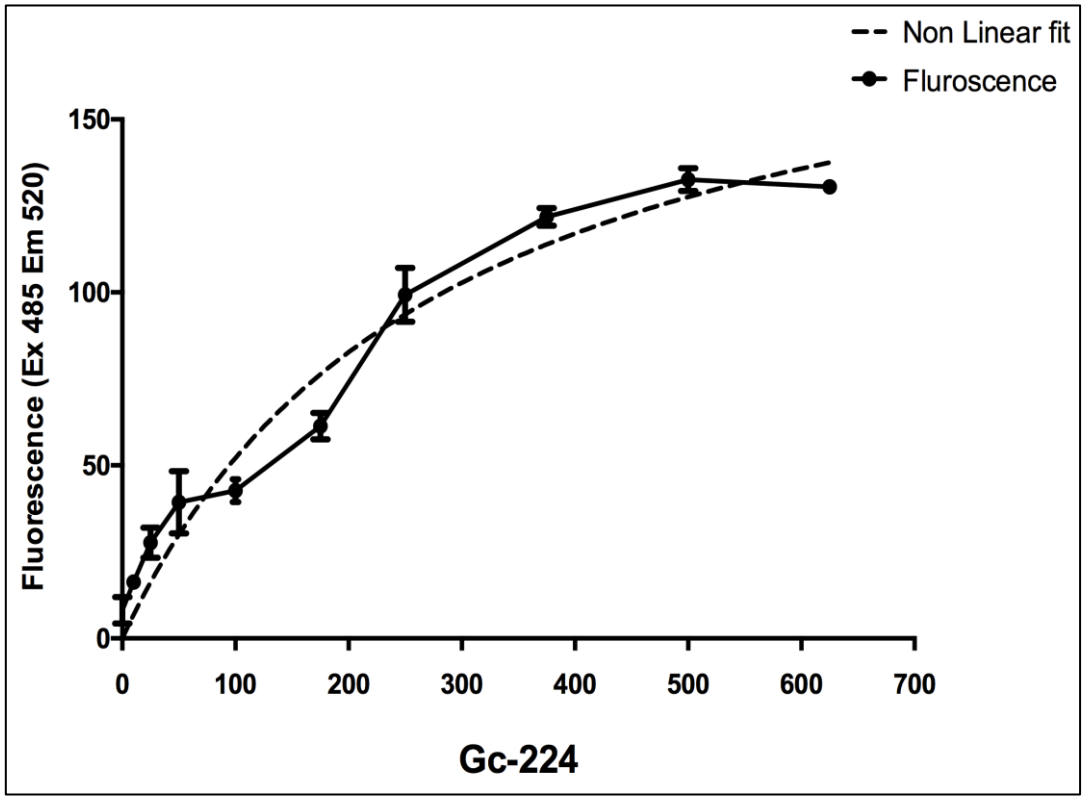


Figure 4.12: Affinity test of aptamer Gc-224 against Neu5Gc-sp-biotin. Differential concentrations ranging from 0- 625 nM were examined against 250 pmol of immobilised Neu5Gc-spacer-biotin.

4.3.4.2. Neu5Ac

For Neu5Ac binders, aptamer Ac-79 was selected for affinity analysis. A direct binding assay with differential concentrations of aptamer ranging from 0 to 625 nM were incubated with immobilised Neu5Ac-spacer-biotin. This aptamer exhibited binding saturation against Neu5Gc-spacer-biotin at a concentration of 375 nM. The dissociation constant of aptamer Ac-79 was $1.14* 10^{7}$ M^{-1} with a R^2 value of 0.96 (Figure 4.13).

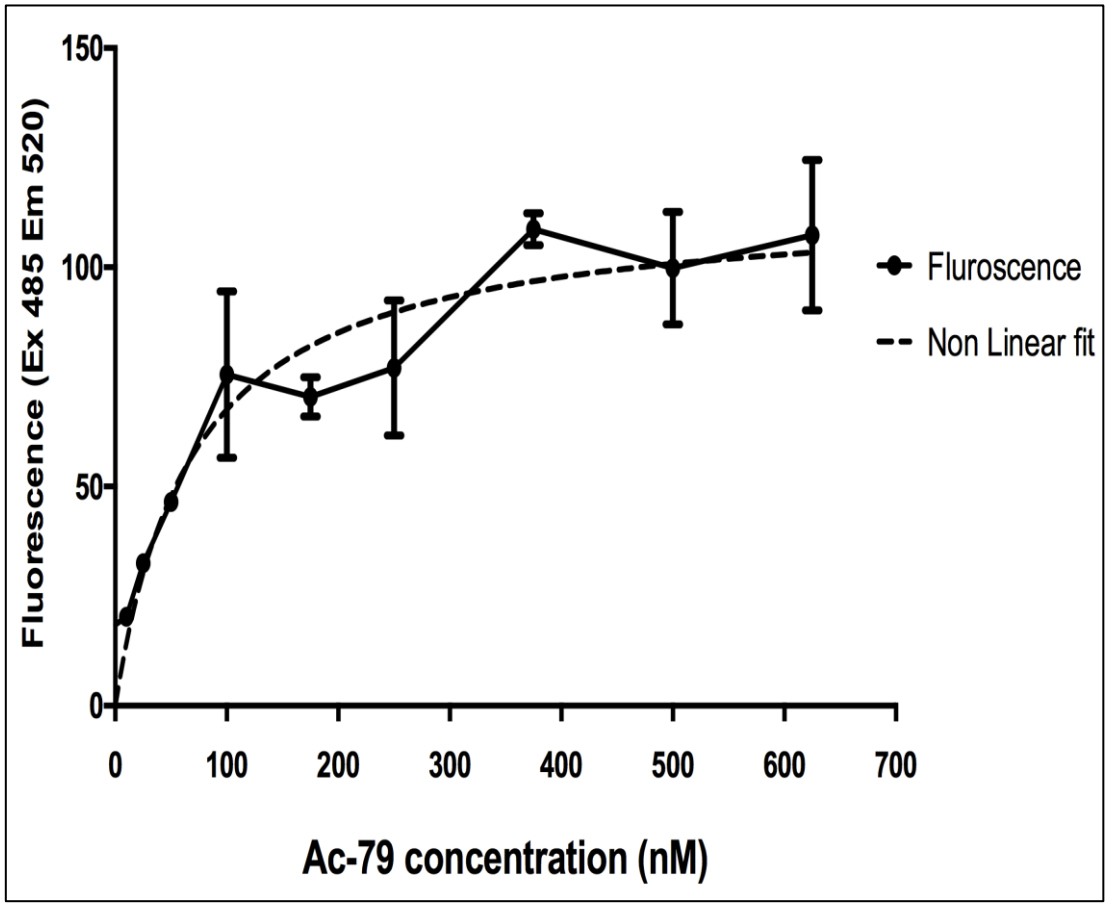


Figure 4.13: Affinity test of aptamer Ac-79 against Neu5Ac-sp-biotin. Differential concentrations ranging from 0- 625 nM were examined against 250 pmol of immobilised Neu5Ac-spacer-biotin immobilised on to the Dynabeads.

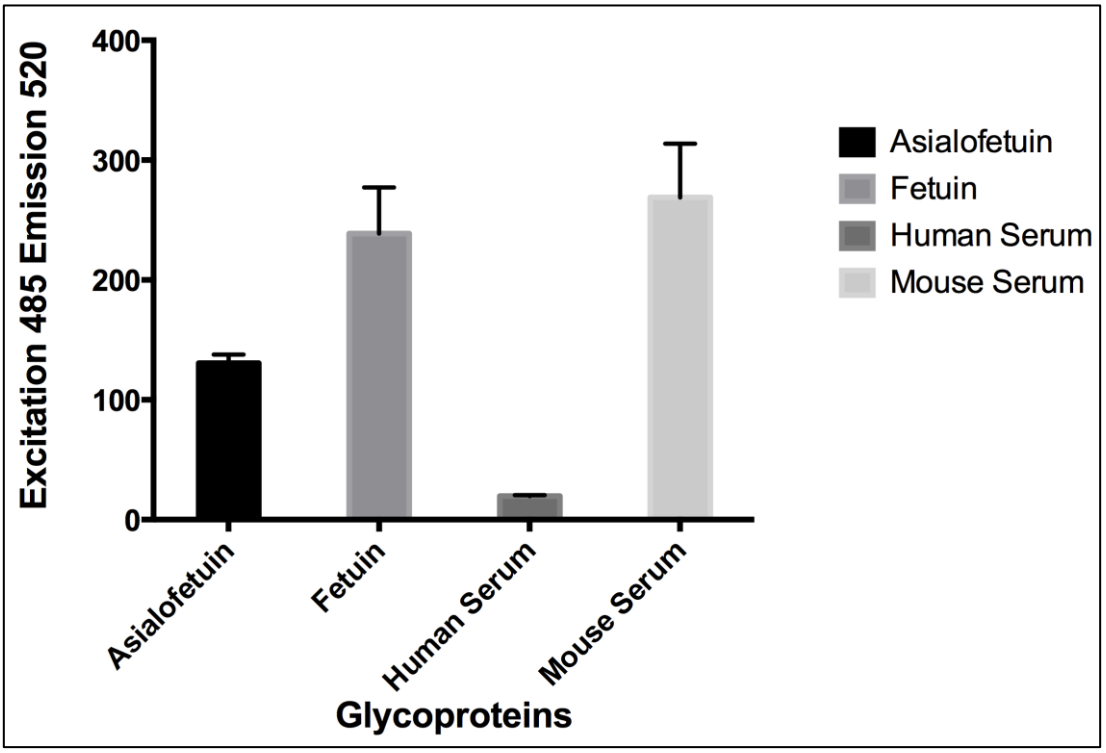


Figure 4.14: Detection of Neu5Gc in the glycoprotein sample using aptamer Gc-224. Four different glycoproteins viz: fetuin, asialofetuin, human serum (from human blood) and mouse serum were biotinylated before and then immobilised on to the streptavidin Dynabeads. Aptamer concentration used for the binding and detection of Neu5Gc is 500 nM.

4.3.5. Detection of Neu5Gc in glycoprotein samples

To determine whether aptamer Gc-224 could recognise Neu5Gc in its native state on glycoproteins, a panel of glycoproteins/glycoprotein pools were biotinylated, both bovine fetuin and bovine asialo-fetuin were biotinylated to facilitate immobilisation on the streptavidin-coated magnetic beads. Aliquots of mouse and human serum, representing complex glycoprotein pools, were similarly biotinylated and biotinylation confirmed by streptavidin-HRP based ELISA. Higher binding was observed with fetuin than asialo-fetuin, as expected (Figure 4.14), because the latter had more than 99.5% of the sialic acid moieties removed during preparation. Also, aptamer Gc-224 showed significantly higher binding to the biotinylated mouse serum protein pool than the human serum protein pool (Figure 4.14). Mouse serum glycoproteins would be expected to contain Neu5Gc, because mice can synthesise

this sialic acid, whereas none or very low levels would be expected to be present on human serum glycoproteins.

4.3.6. Development of an aptamer based competitive binding assay for Neu5Gc and Neu5Ac respectively by bead-based fluorescence binding inhibition method

4.3.6.1. Neu5Gc

A competitive binding assay was established with aptamer Gc-224 and free Neu5Gc as standard over the final assay concentration range of 10 ng/mL to 1 mg/mL. This gave a standard curve that was linear over the Neu5Gc concentration range 10 ng/mL to 10 μg/mL and had an IC_{50} value of 121 ng/mL (Figure 4.15). These results confirmed the specificity of the aptamer Gc-224-based competitive assay for Neu5Gc. When Neu5Ac was tested in the assay over the same concentration range, it showed only 30% inhibition of binding (B/B_0 = 70%), at 1 mg/mL and therefore the IC_{50} value was greater than 1mg/mL.

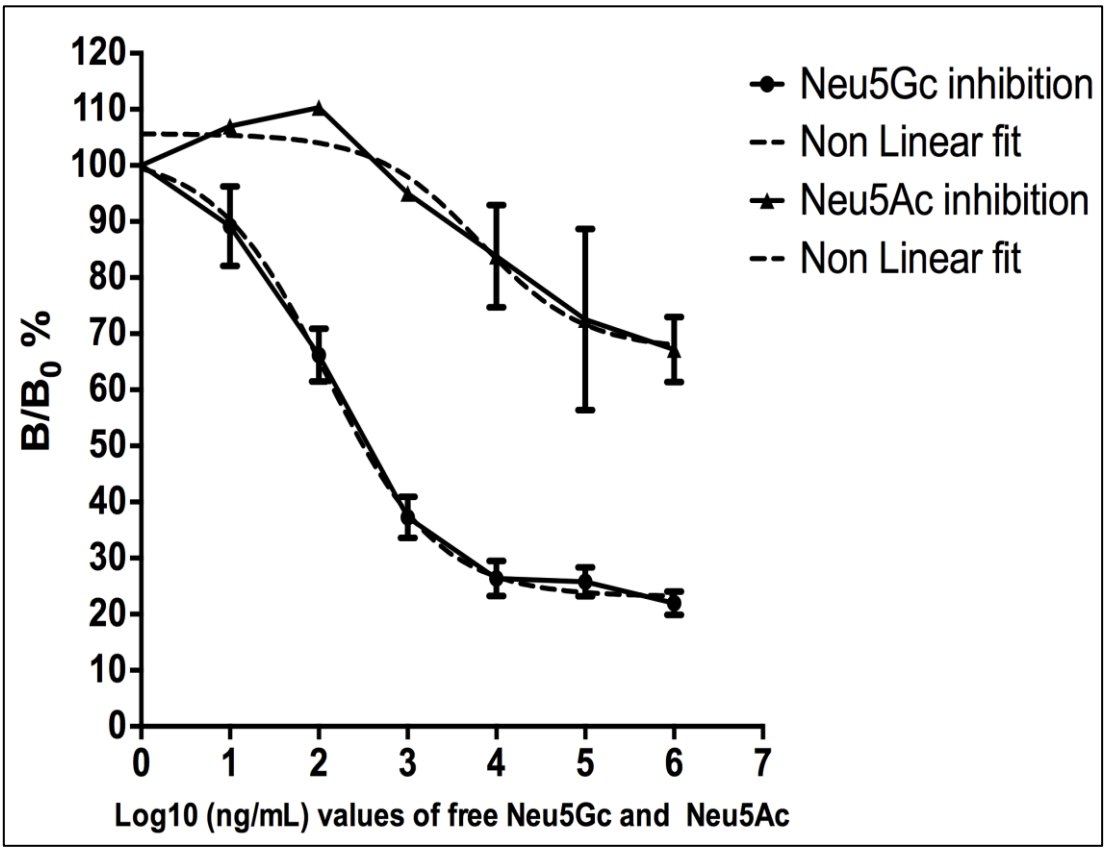


Figure 4.15: Binding inhibition profile of aptamer Gc-224 (500nM) against Neu5Gc–sp-biotin (250 pmol) using free Neu5Gc and free Neu5Ac, ranging from 0 - 1 mg/mL (represented as Log (10) ng/mL on x-axis).

4.3.6.2. Neu5Ac

A competitive assay was also developed for aptamer Ac-79 using free Neu5Ac ranging from 10 ng/mL – 1mg/mL. Free Neu5Ac at 1mg/mL inhibited 89% (B/B_0= 11%) of aptamer Ac-79 binding against Neu5Ac-spacer-biotin, and the estimates of 50 % (IC_{50}) of aptamer Ac-79 binding against Neu5Ac-sp-biotin was inhibited at 76 ng/mL (Figure 4.16).

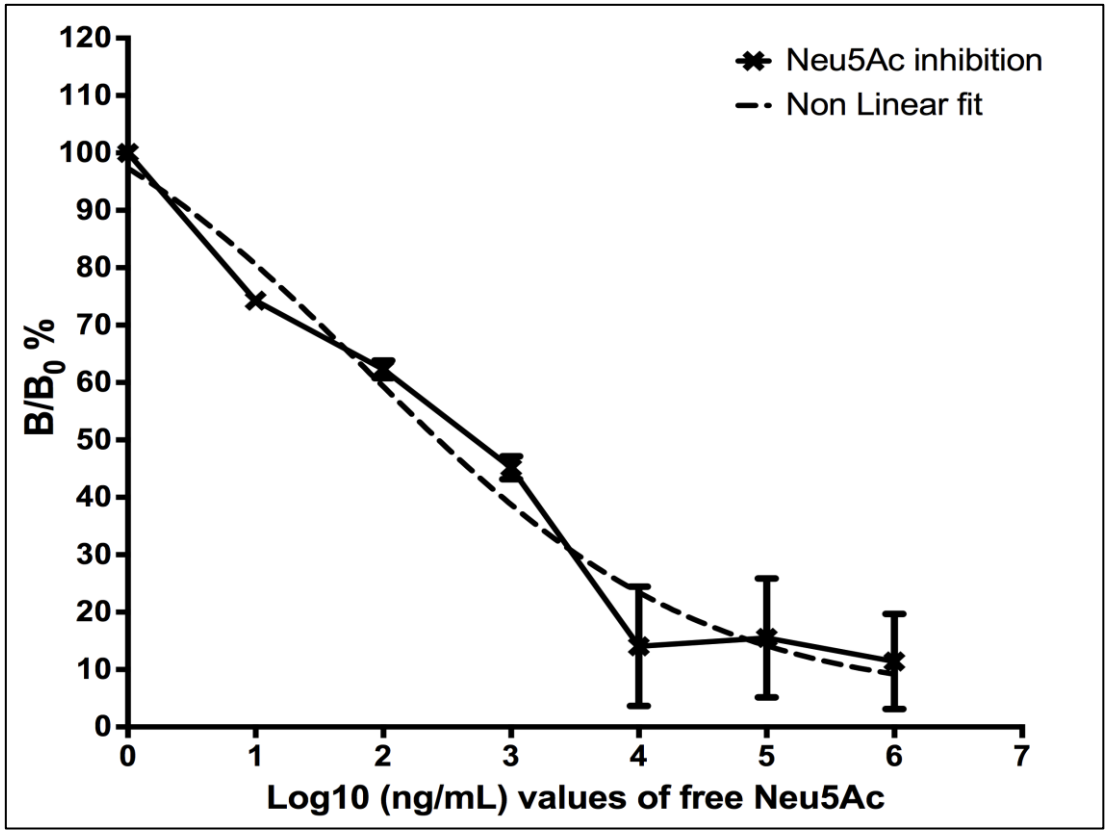


Figure 4.16: Binding inhibition profile of aptamer Ac-79 (500nM) against Neu5Ac–sp-biotin (250 p mol) using free Neu5Ac ranging from 0- 1 mg/mL (represented as Log (10) ng/mL on x-axis).

4.3.7. Evaluation of Gc-224 aptamer specificity

The target Neu5Gc-Sp-biotin (2 µg/ml) was immobilised onto magnetic Dynabeads and later aptamer Gc-224 was incubated without any inhibition and also along with equal concentration of 1 mg/mL of inhibitory free monosaccharides: Neu5Gc, Glc, Gal, Fuc, Man, and Xyl. The binding fluorescence intensity of Neu5Gc-Sp-biotin with aptamer Gc-224 without any inhibition was considered as standard 100% binding (Figure 4.17). Free Glc, Gal and Man showed more than 90% binding in comparison to standard binding, while Fuc and Xyl showed 71.7% and 75.7%

binding, respectively, effectively indicating that the aptamer was not substantially inhibited by these sugars. On the other hand, free Neu5Gc inhibition showed only 23.3% binding to the standard binding, which shows that free Neu5Gc at a concentration of 1mg/mL inhibited 76.6% of aptamer Gc-224 binding with Neu5Gc-sp-biotin (Figure 4.17).

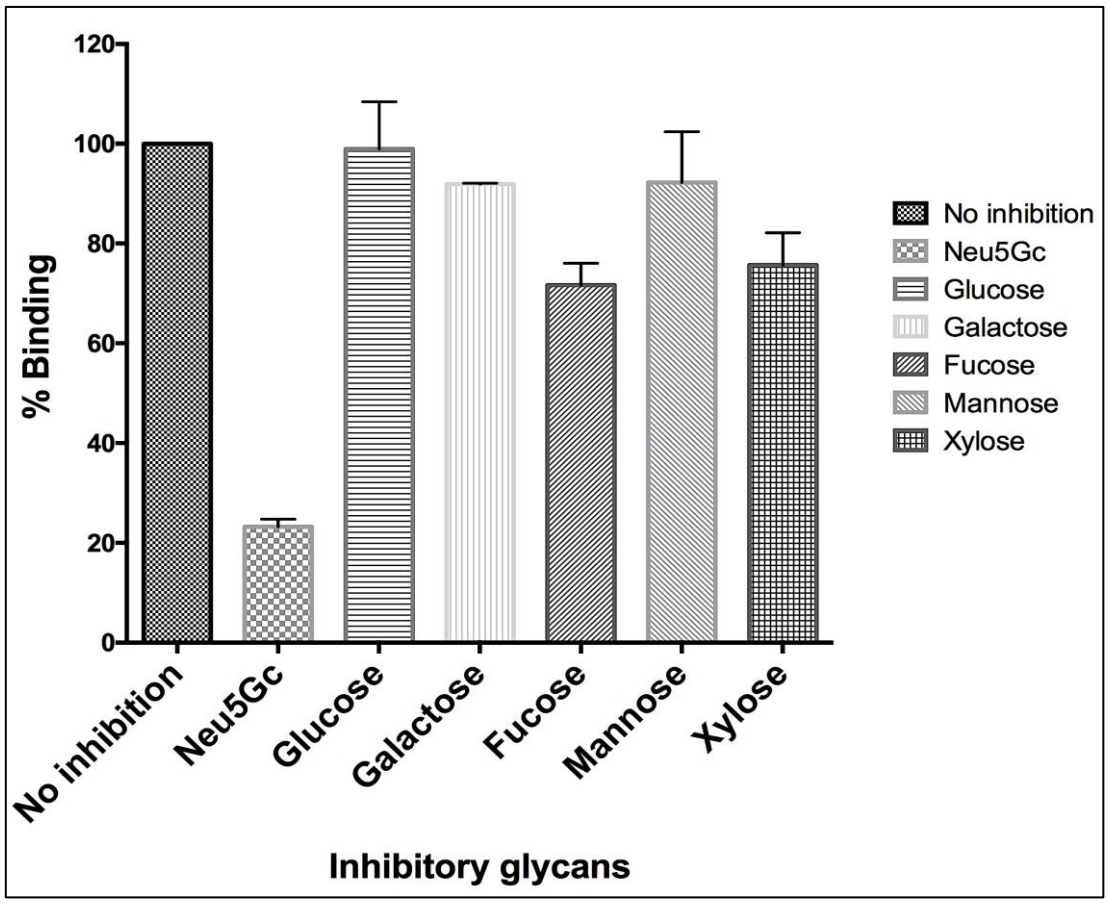


Figure 4.17: Aptamer Gc-224 specificity profile using inhibitory free monosaccharaides. Six different monosaccharaides ranging from Neu5Gc, Glu, Gal, Fuc, Man and Xyl were used as free inhibitory glycans at a concentration of 1 mg/mL. Aptamer 224 binding with Neu5Gc-sp-biotin without inhibition was considered as 100% binding.

4.4. Discussion

The ability to detect and quantify the Neu5Gc motifs on cells is important for determining their role in biotherapeutic contamination. The presence of this immunogenic motif on glycoprotein biopharmaceuticals has resulted in adverse reactions in patients, (Ghaderi et al., 2010, Ghaderi et al., 2012). Neu5Gc also has an important role in tumour progression (Padler-Karavani et al., 2011, Samraj et al., 2014). The above highlights the need for specific recognition reagents and a convenient analytical method for Neu5Gc detection and determination. Currently, there is no binding molecule reported that binds specifically against Neu5Ac.

Advancements in the fields of immunology, molecular biology and biotechnology, have led to the development of new analytical and diagnostics assays. In this work, several strategies were pursued to develop appropriate assays including biotinylated aptamers used in ELISA type assays, and unlabelled aptamers detected with a DNA-binding fluorescent dye in binding and inhibition assays.

Various assay formats were assessed and adapted for determining aptamer specificity and affinity. Microtitre plate-based approaches were used initially for specificity and affinity comparison with immobilised targets and either (i) direct fluorescence detection of bound nucleic acid or (ii) detection of fluorescently-labelled aptamers. A most desirable targeted assay format would be an ELISA type assay, which would be transferable to high-throughput methods. The choice of label and detection method for monitoring the binding event is a very important issue for aptamer assay development. Biotin labeling of aptamers is useful in a way that it can be detected using an avidin or streptavidin secondary detection step, such as with an HRP substrate. The most successful format was the bead based assay with subsequent fluorescence using OliGreen dye detection of the unlabelled aptamer (Wochner and Glokler, 2007).

Beads-based binding assay has been used for the detection of low molecular weight glycans; this assay has also demonstrated the detection of Neu5Ac using aptamer (Cho et al., 2013). For Neu5Ac binding aptamer; Ac-79 has shown binding affinity constant of $1.14 * 10^{7}\ M^{-1}$ in a bead based direct binding assay (Figure 4.13), and the IC_{50} value of 76 ng/mL of free Neu5Ac in the competitive bead based assay (Figure

4.16). An RNA aptamer has shown a dissociation constant of 1.5 nM against Neu5Ac in the epoxy bead based assay (Cho et al., 2013), which is considered to be good binder (Mairal et al., 2008, Friedman et al., 2008), however, this aptamer also shows dissociation constant of 90 nM against Neu5Gc, concluding that the published RNA aptamer has its limitation in specific selection to Neu5Ac. Aptamer Ac-79 however, has shown specificity against Neu5Ac in the bead-based assay followed by urea gel detection, in which Ac-79 shows DNA band (aptamer) in Neu5Ac-sp-biotin lane and no traces of DNA in Neu5Gc-sp-biotin (Figure 4.5). Aptamer Ac-79 is a potential aptamer that could detect Neu5Ac specifically in glycoconjugates, however more experiments need to be carried out to show higher specificities using a larger panel of sialic acids and other glycans.

A similar aptamer-binding bead based assay was developed for Neu5Gc, showing high affinity and specificity of aptamer Gc-224 against Neu5Gc. The aptamer Gc-224 can detect free Neu5Gc at the very low level of 121 ng/mL. The IC_{50} value of the previously reported Neu5Gc binding aptamer N8 in an ELISA assay was 127 ng/mL (Gong et al., 2013), higher than the aptamer Gc-224, (IC_{50} - 121 ng/mL) reported here. Also, aptamer N8, and five other aptamers reported in the previous chapter that had a different consensus sequence from Gc-224 did not show any specificity towards Neu5Ac. However, aptamer Gc-224 was not significantly inhibited with free Neu5Ac with only 30% inhibition using 1 mg/mL of free Neu5Ac. Inhibition at this very high concentration of free Neu5Ac is likely to be due to causes other than direct, specific recognition of sialic acid.

While possessing high affinity against immobilised Neu5Gc-sp-biotin, aptamer Gc-224 was also tested for its specificity against a range of monosaccharides in the competitive binding assay (Figure 4.17). Binding against the standard Neu5Gc-sp-biotin was considered 100 % binding and inhibitory effects competing glycans was compared to this binding. Free Neu5Gc showed 76.6 % of inhibition at 1 mg/mL where no other glycans inhibited binding more than 25%, indicating that the aptamer Gc-224 shows high specificity to Neu5Gc. Three different sialic acid binding aptamer studies have reported against Neu5Ac, Neu5Gc and Sialyl lactose (Mehedi Masud et al., 2004, Cho et al., 2013, Gong et al., 2013), but none of these studies had reported specificities against such variety of glycans in any form of assays.

This assay has also demonstrated the detection of Neu5Gc in glycoprotein samples. It was important to determine whether aptamer Gc-224 could bind Neu5Gc conjugated to glycoproteins and can recognise Neu5Gc present on glycoprotein in the direct assay format. The direct assay format could later be transferred to develop point of care diagnostics. The glycoproteins, bovine fetuin and bovine asialofetuin were selected because bovine fetuin has terminal Neu5Gc, and asialofetuin is chemically desialylated and contains no or only trace amounts of sialic acid. Aptamer Gc-224 showed higher binding with fetuin than asialo-fetuin (Figure 4.14), however it also showed low binding with asialo-fetuin as well and the reason is unknown, presuming it could be due to the high background. The aptamer Gc-224 showed significantly higher binding in mouse serum than human serum. The mouse serum contains Neu5Gc, and as expected, aptamer Gc-224 showed high binding to mouse serum demonstrating the ability of aptamer Gc-224 to detect the low levels of Neu5Gc present in the mouse serum. Thus the assay provides the option of either direct analysis of the Neu5Gc on a glycoprotein of interest or measurement of the free Neu5Gc in solution.

In conclusion, this chapter has examined different assay platforms to characterise aptamers against their respective targets. Robust platforms like ELISA did not show enough binding to detect aptamer-sialic acid interaction. The only assay that worked well was the bead-based fluorescent assay and subsequent assay development was done for both Neu5Ac and Neu5Gc aptamers, respectively. The reasons for failure of other assays could be a glycan presentation issue, as the enrichment of aptamers against glycan was done in similar bead-based assays that worked well against both aptamers. Another reason for assay failure could be the use of streptavidin as a secondary conjugate step in detection of the biotinylated aptamer. In the bead-based assay, DNA intercalating OliGreen dye was used that is highly sensitive to detect any DNA in the aptamer-sialic acid interaction (Wochner and Glokler, 2007). In ELISA assays, biotin labeling of the aptamer is a frequently used alternative because of the strong streptavidin/biotin interaction, but a secondary detection step using an avidin/streptavidin-HRP conjugate is necessary and only strong binding against target in an accessible presentation can be detected using this label (Platt et al., 2009). This could be due to inaccessibility of aptamers to bind against target glycans. Further optimisation in the presentation of glycans as well as aptamers on the

immobilisation surface is required and if needed affinity maturation methods could be used to enhance binding and specificity. These methods include aptamer sequence and structural modification both *in silico* and *in vitro,* focusing mainly on the variable region sequences or sequences involving the stem loop structures that could have higher binding affinity against the target (Nonaka et al., 2012, Soukup et al., 2001). Also, increasing the numbers of rounds of stringent SELEX selection could be employed to so as to generate fewer aptamers that will have increased affinity and specificity and thus perform better in different assay platforms.

4.5. References

ANUMULA, K. R. 1995. Rapid quantitative determination of sialic acids in glycoproteins by high-performance liquid chromatography with a sensitive fluorescence detection. *Analytical biochemistry,* 230**,** 24-30.

BRADFORD, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry,* 72**,** 248-254.

BROWN, P. 2012. *High Pressure Liquid Chromatography: Biochemical and Biomedical Applications,* Elsevier.

CHEN, F., HU, Y., LI, D., CHEN, H. & ZHANG, X.-L. 2009. CS-SELEX generates high-affinity ssDNA aptamers as molecular probes for hepatitis C virus envelope glycoprotein E2. *PLoS One,* 4**,** e8142.

CHO, S., LEE, B. Ä., CHO, B. Ä., KIM, J. Ä. & KIM, B. Ä. 2013. In vitro selection of sialic acid specific RNA aptamer and its application to the rapid sensing of sialic acid modified sugars. *Biotechnology and bioengineering,* 110**,** 905-913.

FARRELL, C.-J. L., MARTIN, S., MCWHINNEY, B., STRAUB, I., WILLIAMS, P. & HERRMANN, M. 2012. State-of-the-art vitamin D assays: a comparison of automated immunoassays with liquid chromatography‚Äìtandem mass spectrometry methods. *Clinical chemistry,* 58**,** 531-542.

FERREIRA, C., PAPAMICHAEL, K., GUILBAULT, G., SCHWARZACHER, T., GARIEPY, J. & MISSAILIDIS, S. 2008. DNA aptamers against the MUC1 tumour marker: design of aptamer-antibody sandwich ELISA for the early diagnosis of epithelial tumours. *Analytical and bioanalytical chemistry,* 390**,** 1039-1050.

FRIEDMAN, M., ORLOVA, A., JOHANSSON, E., ERIKSSON, T. L., HOIDEN-GUTHENBERG, I., TOLMACHEV, V., NILSSON, F. Y. & STAHL, S. 2008. Directed evolution to low nanomolar affinity of a tumour-targeting epidermal growth factor receptor-binding affibody molecule. *Journal of molecular biology,* 376**,** 1388-1402.

GARBIS, S., LUBEC, G. & FOUNTOULAKIS, M. 2005. Limitations of current proteomics technologies. *Journal of Chromatography A,* 1077**,** 1-18.

GEISLER, C. & JARVIS, D. L. 2011. Letter to the Glyco-Forum: Effective glycoanalysis with Maackia amurensis lectins requires a clear understanding of their binding specificities. *Glycobiology,* 21**,** 988-993.

GHADERI, D., TAYLOR, R. E., PADLER-KARAVANI, V., DIAZ, S. & VARKI, A. 2010. Implications of the presence of N-glycolylneuraminic acid in recombinant therapeutic glycoproteins. *Nature biotechnology,* 28**,** 863-867.

GHADERI, D., ZHANG, M., HURTADO-ZIOLA, N. & VARKI, A. 2012. Production platforms for biotherapeutic glycoproteins. Occurrence, impact, and challenges of non-human sialylation. *Biotechnology and Genetic Engineering Reviews,* 28**,** 147-176.

GONG, S., REN, H.-L., TIAN, R.-Y., LIN, C., HU, P., LI, Y.-S., LIU, Z.-S., SONG, J., TANG, F. & ZHOU, Y. 2013. A novel analytical probe binding to a potential carcinogenic factor of N-glycolylneuraminic acid by SELEX. *Biosensors and Bioelectronics*.

GORNIK, O. & LAUC, G. 2007. Enzyme linked lectin assay (ELLA) for direct analysis of transferrin sialylation in serum samples. *Clinical biochemistry,* 40**,** 718-723.

HIGUCHI, A., SIAO, Y.-D., YANG, S.-T., HSIEH, P.-V., FUKUSHIMA, H., CHANG, Y., RUAAN, R.-C. & CHEN, W.-Y. 2008. Preparation of a DNA Aptamer‚àí Pt Complex and Its Use in the Colourimetric Sensing of Thrombin and Anti-Thrombin Antibodies. *Analytical chemistry,* 80**,** 6580-6586.

HTTPS://http://WWW.PIERCENET.COM/INSTRUCTIONS/2160750.PDF. 2013. EZ-Link® Amine-PEGn-Biotin. [Accessed 28 October, 2013].

HURUM, D. C. & ROHRER, J. S. 2011. Five-minute glycoprotein sialic acid determination by high-performance anion exchange chromatography with pulsed amperometric detection. *Analytical biochemistry,* 419**,** 67-69.

JANSSEN, M. J., WIELDERS, J. P., BEKKER, C. C., BOESTEN, L. S., BUIJS, M. M., HEIJBOER, A. C., VAN DER HORST, F. A., LOUPATTY, F. J. & VAN DEN OUWELAND, J. M. 2012. Multicenter comparison study of current methods to measure 25-hydroxyvitamin D in serum. *Steroids*, 77**,** 1366-1372.

KARAMANOS, N. K., WIKSTROM, B., ANTONOPOULOS, C. A. & HJERPE, A. 1990. Determination of N-acetyl-and N-glycolylneuraminic acids in glycoconjugates by reversed-phase high-performance liquid chromatography with ultraviolet detection. *Journal of Chromatography A*, 503**,** 421-429.

KAWABATA, A., MORIMOTO, N., ODA, Y., KINOSHITA, M., KURODA, R. & KAKEHI, K. 2000. Determination of mucin in salivary glands using sialic acids as the marker by high-performance liquid chromatography with fluorometric detection. *Analytical biochemistry,* 283**,** 119-121.

KIND, T. & FIEHN, O. 2010. Advances in structure elucidation of small molecules using mass spectrometry. *Bioanalytical reviews,* 2**,** 23-60.

KLEIN, A., DIAZ, S., FERREIRA, I., LAMBLIN, G. V., ROUSSEL, P. & MANZI, A. E. 1997. New sialic acids from biological sources identified by a comprehensive

and sensitive approach: liquid chromatography-electrospray ionisation-mass spectrometry (LC-ESI-MS) of SIA quinoxalinones. *Glycobiology,* 7**,** 421-432.

LEE, K., LOGANATHAN, D., MERCHANT, Z. & LINHARDT, R. 1990. Carbohydrate analysis of glycoproteins A review. *Applied biochemistry and biotechnology,* 23**,** 53-80.

MAIRAL, T., OZALP, V. C., SANCHEZ, P. L., MIR, M., KATAKIS, I. & O'SULLIVAN, C. K. 2008. Aptamers: molecular tools for analytical applications. *Analytical and bioanalytical chemistry,* 390**,** 989-1007.

MEHEDI MASUD, M., KUWAHARA, M., OZAKI, H. & SAWAI, H. 2004. Sialyllactose-binding modified DNA aptamer bearing additional functionality by SELEX. *Bioorganic & medicinal chemistry,* 12**,** 1111-1120.

MORIMOTO, N., NAKANO, M., KINOSHITA, M., KAWABATA, A., MORITA, M., ODA, Y., KURODA, R. & KAKEHI, K. 2001. Specific distribution of sialic acids in animal tissues as examined by LC-ESI-MS after derivatisation with 1, 2-diamino-4, 5-methylenedioxybenzene. *Analytical chemistry,* 73**,** 5422-5428.

MORTEZAI, N., BEHNKEN, H. N., KURZE, A.-K., LUDEWIG, P., BUCK, F., MEYER, B. & WAGENER, C. 2013. Tumour-associated Neu5Ac-Tn and Neu5Gc-Tn antigens bind to C-type lectin CLEC10A (CD301, MGL). *Glycobiology,* 23**,** 844-852.

NONAKA, Y., YOSHIDA, W., ABE, K., FERRI, S., SCHULZE, H., BACHMANN, T. T. & IKEBUKURO, K. 2012. Affinity improvement of a VEGF aptamer by in silico maturation for a sensitive VEGF-detection system. *Analytical chemistry,* 85**,** 1132-1137.

PADLER-KARAVANI, V., HURTADO-ZIOLA, N., PU, M., YU, H., HUANG, S., MUTHANA, S., CHOKHAWALA, H. A., CAO, H., SECREST, P. & FRIEDMANN-MORVINSKI, D. 2011. Human xeno-autoantibodies against a non-human sialic acid serve as novel serum biomarkers and immunotherapeutics in cancer. *Cancer research,* 71**,** 3352-3363.

PLATT, M., ROWE, W., KNOWLES, J., DAY, P. J. & KELL, D. B. 2009. Analysis of aptamer sequence activity relationships. *Integrative Biology,* 1**,** 116-122.

ROHRER, J. S., THAYER, J., WEITZHANDLER, M. & AVDALOVIC, N. 1998. Analysis of the N-acetylneuraminic acid and N-glycolylneuraminic acid contents of glycoproteins by high-pH anion-exchange chromatography with pulsed amperometric detection (HPAEC/PAD). *Glycobiology,* 8**,** 35-43.

SAMRAJ, A. N., PEARCE, O. M., LAUBLI, H., CRITTENDEN, A. N., BERGFELD, A. K., BANDA, K., GREGG, C. J., BINGMAN, A. E., SECREST, P.

& DIAZ, S. L. 2014. A red meat-derived glycan promotes inflammation and cancer progression. *Proceedings of the National Academy of Sciences***,** 201417508.

SHAH, M. M., FUJIYAMA, K., FLYNN, C. R. & JOSHI, L. 2003. Sialylated endogenous glycoconjugates in plant cells. *Nature biotechnology,* 21**,** 1470-1471.

SHAH, V. P., MIDHA, K. K., FINDLAY, J. W., HILL, H. M., HULSE, J. D., MCGILVERAY, I. J., MCKAY, G., MILLER, K. J., PATNAIK, R. N. & POWELL, M. L. 2000. Bioanalytical method validation,Äîa revisit with a decade of progress. *Pharmaceutical research,* 17**,** 1551-1557.

SHUKLA, A. K. & SCHAUER, R. 1982. Fluorimetric determination of unsubstituted and 9 (8)-O-acetylated sialic acids in erythrocyte membranes. *Hoppe-Seyler's Zeitschrift fur physiologische Chemie,* 363**,** 255-262.

SKOZA, L. & MOHOS, S. 1976. Stable thiobarbituric acid chromophore with dimethyl sulphoxide. Application to sialic acid assay in analytical de-O-acetylation. *Biochem. J,* 159**,** 457-462.

SOUKUP, G. A., DEROSE, E. C., KOIZUMI, M. & BREAKER, R. R. 2001. Generating new ligand-binding RNAs by affinity maturation and disintegration of allosteric ribozymes. *Rna,* 7**,** 524-536.

STANCZYK, F. Z. & CLARKE, N. J. 2010. Advantages and challenges of mass spectrometry assays for steroid hormones. *The Journal of steroid biochemistry and molecular biology,* 121**,** 491-495.

TANG, Z., SHANGGUAN, D., WANG, K., SHI, H., SEFAH, K., MALLIKRATCHY, P., CHEN, H. W., LI, Y. & TAN, W. 2007. Selection of aptamers for molecular recognition and characterisation of cancer cells. *Analytical chemistry,* 79**,** 4900-4907.

TANGVORANUNTAKUL, P., GAGNEUX, P., DIAZ, S., BARDOR, M., VARKI, N., VARKI, A. & MUCHMORE, E. 2003. Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid. *Proceedings of the National Academy of Sciences,* 100**,** 12045-12050.

WANG, D., ZHOU, X., WANG, L., WANG, S. & SUN, X.-L. 2014. Quantification of free sialic acid in human plasma through a robust quinoxalinone derivatisation and LC-MS/MS using isotope-labeled standard calibration. *Journal of Chromatography B,* 944**,** 75-81.

WOCHNER, A. & GLOKLER, J. 2007. Nonradioactive fluorescence microtiter plate assay monitoring aptamer selections. *BioTechniques,* 42**,** 578.

Chapter-5: Conclusion and future perspectives

5.1. Conclusions

Protein glycosylation is a post-translational phenomenon that is involved in most physiological and disease processes. The importance of glycosylation and the presence of glycan motifs on cell surfaces is outlined in the beginning in chapter 1. Following from this, the importance of glycans in disease and recombinant bio therapeutics is introduced. Undesirable inclusion of non-human glycans, such as the Gal-α-(1, 3)-Gal motif and the sialic acid Neu5Gc, and the different detection methods for the identification of these glycan motifs with specific application for bio-therapeutics, are also introduced.

The high diversity and complexity of the primary structure of glycan chains together with their conformational plasticity make analysis of glycan-ligand interactions not simple to accomplish. Therefore, the structural complexity and lack of suitable glyco profiling tools are major challenges in glycoscience. Most of the known disease-associated glycobiomarkers were discovered individually using liquid chromatography and mass spectroscopy. Though valuable, there is room for improvement in these approaches for the discovery phase. Affinity molecules offer promise in the development of inexpensive, high-throughput methods that are complementary to traditional carbohydrate analysis techniques such as chromatography and mass spectrometry. Lectins are carbohydrate-binding proteins, which have been effectively used in many applications, however their broad and label-dependent specificities can make data interpretation with known structures difficult and their unambiguous use for analysis of unknowns impossible. There is also a critical need for innovative, rapid, and high-throughput (HTP) technologies to specifically detect and quantify non-human glycans.

These challenges were addressed through development and characterisation of aptamers and scFv that could mimic glycan binding lectins. The balance of this thesis document details the discovery methods used for novel scFvs and aptamers and applications relevant to the fields of biology, biotechnology, and diagnostics. The modelling and discovery of new molecules specifically for the identification of non human glycan structures has been the main objective of this thesis work, which is detailed in chapters 2, 3, and 4.

Chapter 2 of this thesis examines the recognition and interaction of reported chicken scFv molecules against the Gal-α-(1, 3)-Gal motif, using *in silico* approaches. From a structural point of view, a deep understanding of the fundamentals underlying the interaction between glycans and their recognition scFv is essential to rationalise and interpret experimental binding and selectivity data at atomic/molecular level. Furthermore this knowledge can assist in the design or structural optimisation of either binding partner and so improve performance of binding assays based on these interactions or provide novel leads for the structure of modulators of the binding event.

For protein (scFv) sequence and structural predictions, homology models of the scFvs were developed based on the available crystal structures of scFvs and antibodies, and their estimated binding affinity, which were correlating, with SPR affinities and previous ELISA results. The binding affinity of each scFv against selected glycans and the pattern observed show differential binding affinity across the ligands. The binding affinity pattern of the three scFvs against the ligands was similar in both *in silico* modelling and ELISA results.

The active sites of scFvs are predicted with in this thesis, which will be useful for mutagenesis of scFv sequences to enhance their affinities. The determination of key binding regions and properties among the amino acid residues, which could be critical in the recognition, and binding to the target epitope. *In silico* study demonstrates the potential of this approach for identifying glycan-binding scFvs and it could be utilised to identify lectin mimics and their interaction against other carbohydrate epitopes.

Chapter 3 of this thesis describes the aptamer development against sialic acids Neu5Gc, which is an important glycan motif in the context of cancer and also act as a contaminant in the production of recombinant therapeutics. I also developed aptamer against common sialic Neu5Ac to understand the binding mechanism or nucleotide preference of aptamer against these two closely related sialic acids.

SELEX procedure for the aptamer generation was optimised and standardised for use in carbohydrate-specific (sialic acids) aptamer selection. Enrichment analyses of all 10 rounds (pools) of selection against both sialic acids showed that round 7 & 9 and

round 8 & 10 demonstrated higher binding against Neu5Gc-sp-biotin and Neu5Ac-sp-biotin, respectively. The higher binding of a pool correlates with higher number of diverse ssDNA molecules binding against the select sialic acids. These specific pools were cloned, sequenced, and evaluated for the consensus sequences. The finding was that, for Neu5Gc, two consensus sequences accounted for approximately 98.0% of the total sequences synthesised, where as in Neu5Ac, the top two consensuses have more than half of the total sequences synthesised. The secondary structures of all synthesised aptamers were also predicted based on modelling to help better understand the stem-loop structure of these aptamers. These predictions suggested that the majority of binding motifs were in the loop structures.

The identification of aptamer sequences that bind glycan targets through the SELEX process represents the achievement of a key milestone in the development of an oligonucleotide microarray glycosignature platform. The sequences presented here could be included on the initial array and will also be manipulated by targeted and random mutations to generate a library of related aptamer sequences that will represent the feature space of the novel glycan-focused microarray platform.

Chapter 4 provides a detailed description of the characterisation and assay development of DNA aptamers against the sialic acid, non-human Neu5Gc and Neu5Ac, where different assay formats were developed for the aptamers generated (described in chapter 3) and these were used to evaluate and characterise the consensus sequences identified. Different assay formats were tested and the most successful assay showing higher binding affinity and specificity was the bead-based assay with subsequent fluorescence detection of the unlabelled aptamer. This system provided an option of either direct analysis of the Neu5Gc content of a glycoprotein of interest or the measurement of the free Neu5Gc in solution. Through the bead-based assay, high affinity and specificity of aptamer Gc-224 was shown against Neu5Gc. The aptamer Gc-224 could be used to detect free Neu5Gc at a very low level (mere ng of Neu5Gc per mL). The aptamer Gc-224 also detected Neu5Gc presence on glycoproteins such as those found in mouse serum and on bovine fetuin. Also, using this same assay, aptamer Ac-79 binds specifically to Neu5Ac and was able to detect free Neu5Ac at low concentrations (in the ng/mL range).

This to my knowledge is the first report of DNA aptamer binding assay that shows differentiation between Neu5Gc and Neu5Ac. By isolating aptamers that can distinguish between the two most common forms of sialic acid (Neu5Gc and Neu5Ac), it has been demonstrated that this aptamer could be used as both analytical and diagnostic tool for the detection of Neu5Gc and Neu5Ac in glycoprotein samples. Also, it has shown that oligonucleotide selection by the SELEX procedure is a valuable source of novel glycan recognition aptamers. These molecules will have great application in biopharmaceutical production as part of a quality control workflow in allowing for the identification of potentially harmful molecules.

5.2. Future perspectives

Following the description of valuable *in silico* scFv interaction modelling (chapter 2), it is obvious that the project needs to be continued and extended to include mutagenesis experiments. The active sites and the specific residues which have been identified from this project could be used as the starting point for the mutagenesis experiment to generate specific mutations in new libraries (Zhang et al., 2004). These mutated libraries would be anticipated to show higher affinity and enhanced specificity towards the Gal-α-(1, 3)-Gal epitope both as free in disaccharide form and in trisaccharide Gal-α-(1, 3)-Gal-β-(1, 4)-GlcNAc

Extending the work described in chapters 3 and 4, the aptamer development research against sialic acids, aptamer Gc-224 should be used in a competitive assay to detect Neu5Gc mediated inhibitions using biotherapeutics such as Eributux (cetuximab) that contains Neu5Gc. The detection and quantification of Neu5Gc contamination in recombinant bio therapeutics would be a commercial application for aptamers in the pharmaceutical industry. In total, 3 and 5 consensus sequences were synthesised for Neu5Gc and Neu5Ac, respectively. However, not all of these aptamers were fully tested and characterised against their respective sialic acids. Therefore, more complete aptamer sequence validation and characterisation must be done before these selected aptamers can be used in therapeutic production pipelines. In addition, the binding assays and their conditions developed and optimised in this thesis work pave the way for the use of oligonucleotide arrays as screening and glycosignaturing platforms.

The other aspect of future aptamer development is to discover new modalities for treatment that are fast, accurate, selective and cost-effective. Nucleotide-based aptamers have often been highlighted as ‘the new antibodies’ due to the extraordinarily high sensitivity and selectivity towards their targets as they are easy to produce and modify, are low cost, high specificity and affinity, have long shelf life and can bind a wide range of molecules including whole living cells (Ismail et al., 2013).

An approach that has huge potential is based on aptamer engineering. That is, to use tailored aptamer sequences which have their unique tertiary structure, which can

provide distinct and highly selective interaction against carbohydrate targets. The aim is to increase our understanding of what it is that makes a good aptamer, by systematically exploring the relationship between sequence, tertiary structure and ligand binding.

From this future work, the set of key oligosaccharides and carbohydrate targets identified as biomarkers and therapeutically important targets will be used as probes to generate and optimise aptamers using computational modelling (Eriksson et al., 2014) and NMR spectroscopy (Lebars et al., 2007). However, there is an enormous lack of structural data and very little knowledge about the specific interactions between the aptamers generated and their targets. The exploration of the fundamental details in secondary structure, tertiary fold and aptamer-ligand interactions will not only increase our basic understanding, but also enable the construction of tailored nucleotide sequences for a wide range of new targets. The optimised aptamers will be then tested through microarray screening techniques (Cho et al., 2006, Collett et al., 2005) to validate efficiency in binding to these specific targets.

The main objectives are suggested to be:

I. To utilise *de novo* structure prediction tools to generate tertiary models of these aptamers and their glycan complexes.
II. To verify these models of aptamers and aptamer-target-complexes by NMR.
III. To use a combination of directed synthesis, computational modelling and NMR studies, to explore the sequence-structure-binding relationship for a range of aptamer sequences closely related to those identified above; this will provide an understanding of the factors that drive the observed unique sequence – ligand matches.
IV. Finally, to develop an aptamer-based microarray prototypes for these specific glycans and test *in vitro* and on tissue samples.

With a detailed understanding of the underlying factors governing the tertiary structures of oligonucleotide aptamers and how to optimise interactions between sialic acids, and a validated and accurate predictive tool, this opens up for an entirely new area. It will become possible to generate target-specific aptamers for identification/diagnostics and treatment of a wide range of indications. In addition,

this will provide knowledge and fundamental understanding that opens for future projects and developments in terms of, e.g., biomarker-DNA interactions, analytic tools, and new treatment modalities.

5.3. References

CHO, E. J., COLLETT, J. R., SZAFRANSKA, A. E. & ELLINGTON, A. D. 2006. Optimisation of aptamer microarray technology for multiple protein targets. *Analytica chimica acta,* 564**,** 82-90.

COLLETT, J. R., CHO, E. J. & ELLINGTON, A. D. 2005. Production and processing of aptamer microarrays. *Methods,* 37**,** 4-15.

ERIKSSON, E. S., JOSHI, L., BILLETER, M. & ERIKSSON, L. A. 2014. De novo tertiary structure prediction using RNA123-benchmarking and application to Macugen. *Journal of molecular modelling,* 20**,** 1-18.

ISMAIL, S. I., ALSHAER, W., ABABNEH, N. & FATTAL, E. 2013. Aptamers: Promising Molecules for Cancer Stem Cell Targeting. *Journal of Molecular and Genetic Medicine,* 7.

LEBARS, I., RICHARD, T., DI PRIMO, C. & TOULME, J.-J. 2007. NMR structure of a kissing complex formed between the TAR RNA element of HIV-1 and a LNA-modified aptamer. *Nucleic acids research,* 35**,** 6103-6114.

ZHANG, M.-Y., SHU, Y., RUDOLPH, D., PRABAKARAN, P., LABRIJN, A. F., ZWICK, M. B., LAL, R. B. & DIMITROV, D. S. 2004. Improved breadth and potency of an HIV-1-neutralizing human single-chain antibody by random mutagenesis and sequential antigen panning. *Journal of molecular biology,* 335**,** 209-219.

Appendix-1: Selection and characterisation of DNA aptamers against Lewis^{b} (Le^{b})

Selection and characterisation of aptamers against Lewis^b (Le^b).

A1. Cloning and sequencing of aptamer pool against Le^b

For 5th, 8th, 10th, and 12th round of SELEX, the aptamer pools were amplified using unmodified forward primer and reverse primer as described above. These sequences were cloned into the pCR 4.0 TOPO vector and positive clones were selected for sequencing (Eurofins MWG Germany). A PERL script was used to extract the insert sequences in the pCR 4.0 Vector. The fetched sequences varied from 58-62 nucleotides and were aligned with ClustalW alignment program using the default settings to draw a consensus. Five consensus sequences were obtained with high copy number. (Table1)

Table1: Consensus sequences obtained after microtiter plate sequencing.

Consensus sequences	Copies across rounds 5, 8, 10 and 12	Sequences per consensus
R12-005	68	GCGCGGATCCCGCGCGGAAGACCAAGCTCCAAATTSTTACTGKGGGCGCGAAGCTTGCGCA
R12-063	61	GBGCGGATCCCGCGCGCAAACAGGATTAAGGGTAAGTCGTCAAAGGCGCGAAGCTTGCGCA
R12-GH04	25	GCGCAAGCTTCGCGCCCACAGTAAGAATTTGGAGCTTGGTCYTCCGCGCGGGATCCGCGCA
R10-ED12	13	GCGCGGATCCCGCGCGTAGTTTACACTATTGACTTTGCTATATTGGCGCGAAGCTTGCGCA
R12-GH12	14	GCGCAAGCTTCGCGCCAGCGAAGGTTGGGGTGTATTTATTGTTGGCGCGGGATCCGCGCA

A2. Binding analysis of aptamers against Le^b

The binding behaviour of the selected aptamers was evaluated using a 8 well strip Maxisorp NUNC plate based assay and biotinylated aptamers. Le^b antigen-BSA conjugate (Le^b -BSA) and BSA was immobilised on 8 well strip (10 µg/ml) overnight at 4 ^oC, washed with PBST (PBS+ 0.05% Tween-20) three times and blocked with 3% preiodate treated BSA for 60 minutes on shaker at room temperature, after which they were washed again three times with PBST. Each selected aptamer was prepared in Binding buffer (5mM $MgCl_2$ + 25mM NaCl + PBS) for binding analysis by unfolding by heat denaturing at 95^oC for 5 min and cooling on ice for 5 min and allowed to warm up at room temperature for 10 minutes. The aptamers were then incubated with immobilised Le^b-BSA for 90 minutes on

shaker at room temperature, unbound aptamer washed away three times with PBS with 0.05 % Tween-20, and bound aptamer detected with streptavidin conjugated to Poly-HRP (Fisher scientific) (1: 8000 dilution with 0.1 % BSA, room temperature for 1h). After washing, colour was developed using one-step TMB (Sigma; RT for 20 minutes). The reaction was stopped by addition of 1 M H_2SO_4 and wells were read at 450 nm. Three of five aptamers showed significant binding against Le^b (Figure 1).

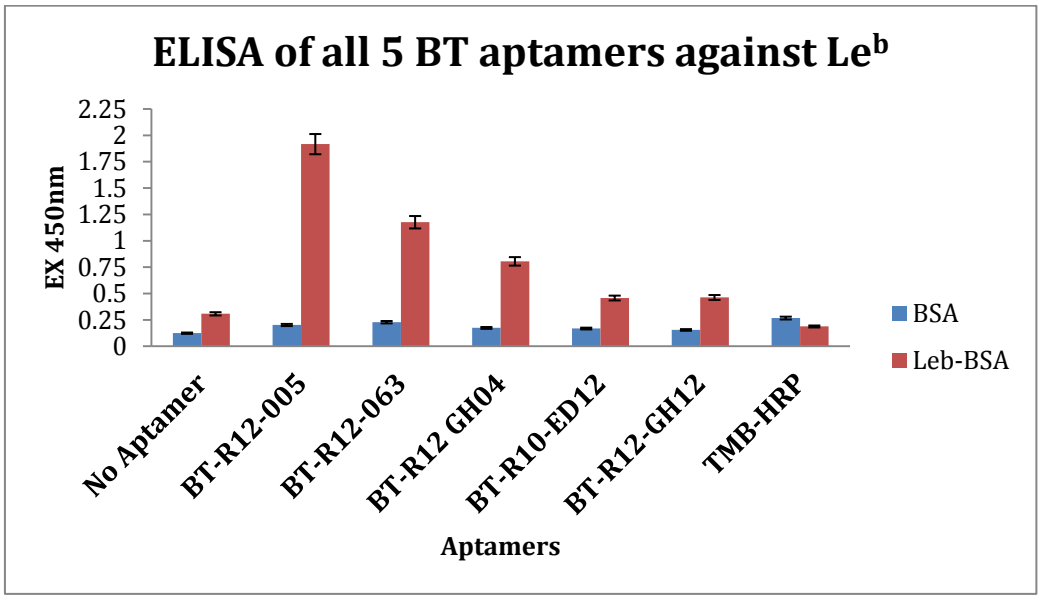


Figure 1: ELISA of all 5 biotinylated aptamers against Le^b

A3. Specificity of aptamer bt-R12-005 against glycans

The specificity analysis was done using same protocol mentioned above for binding analysis. Ten different neo-glycoconjugates: BSA conjugated and BSA (control) were used for analysis.

750 nM of aptamer used in each well. Le^b-3atom–BSA and Le^b-APD-BSA showed significantly higher binding than onther glycans, however Le^x-BSA showed moderate binding with bt-R12-005 (Figure 2)

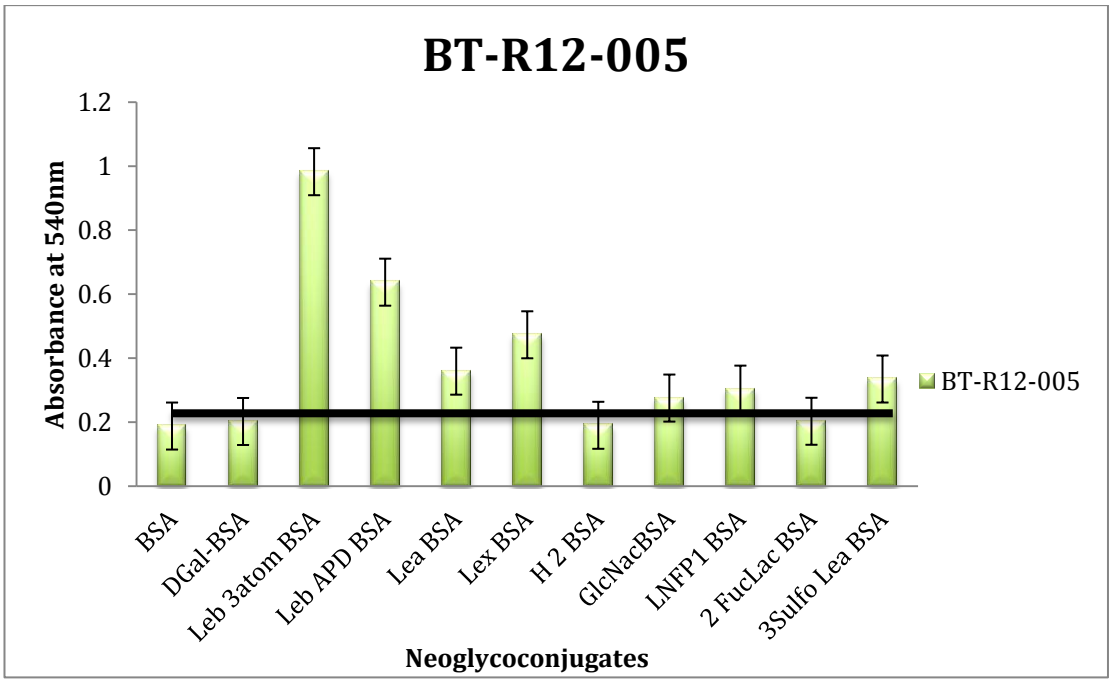


Figure 2: Specificity analysis based on ELISA of aptamer bt-R12-005 against different neo-glycoconjugates.

A4. Concentration dependent ELISA

Concentration dependent ELISA was done to calculate affinity constant of aptamer bt-R12-005 against Le^b. This assay was done using similar protocol mentioned above in binding analysis. Differential concentrations of aptamer ranging from 0 to 1000nM were used for analysis. Aptamer bt-R12-005 showed differential increase from 0 to 1uM, however saturation was not achieved till this concentration (Figure 3).

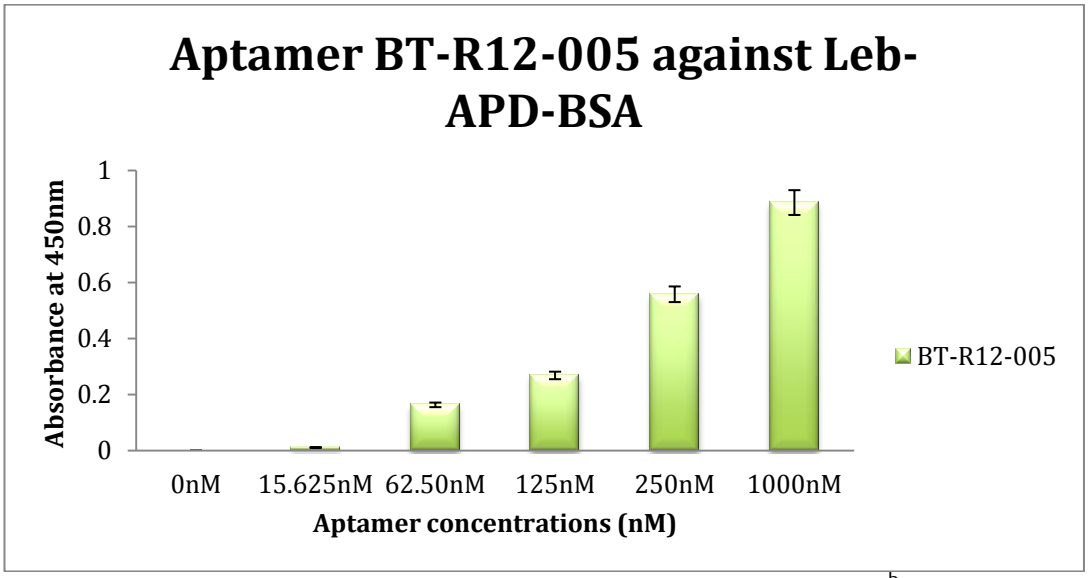


Figure 3: Concentration dependent ELISA of bt-R12-005 against Le^b-BSA

This experiment was repeated again with higher conentration till 10uM (Figure 4) but the intensity of absorbance was low. This might be because change in lot of Streptavidin-HRP, after that different Streptavidin-HRP: RandD DUSet Streptavidin-HRP, Neutravidn: HRP with different dilution were tried to achieve higher absorbance intensity, but the intensity was still low.

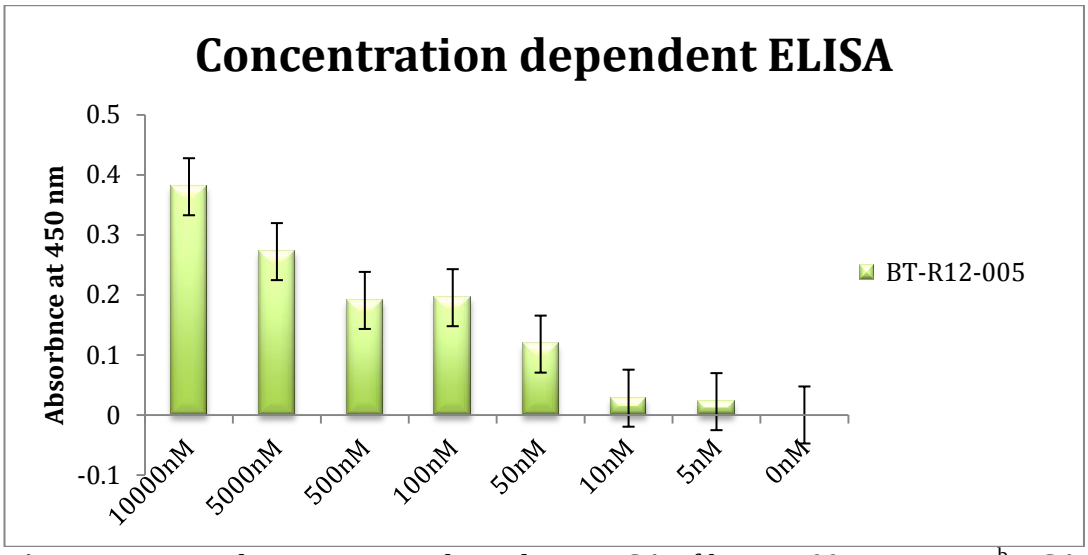


Figure 4: repeated concentration dependent ELISA of bt-R12-005 against Le^b-BSA

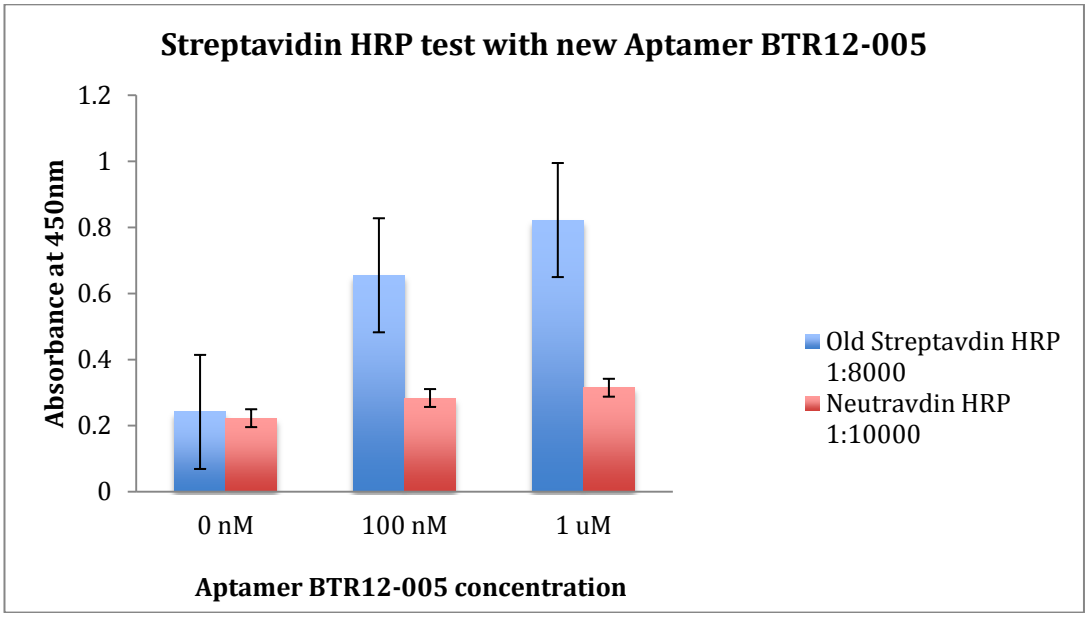


Figure 5: Streptavidin-HRP test with new vial of aptamer bt-R12-005

Later, new vial of aptamer bt-R12-005 was tried at different concentration using Invitrogen (old) Streptavidin-HRP and Neutravidin-HRP (Figure 5), but intensity was higher with old streptavidin-HRP, this showed that there was problem with aptamer stability, rather than the Streptavidin.

Appendix-2: A tightknit group: Protein glycosylation, endoplasmic reticulum stress and the unfolded protein response

A Tight-Knit Group: Protein Glycosylation, Endoplasmic Reticulum Stress and the Unfolded Protein Response

Jared Q Gerlach, Shashank Sharma, Kirk J Leister and Lokesh Joshi

Contents

Abstract

Although the dependence upon glycosylation for protein folding and function is understood to be a key part of protein maturation within the endoplasmic reticulum (ER), details concerning the interconnected nature of pathways associated with protein glycosylation, ER stress and the unfolded protein response are only now beginning to come to light. Changes in glycosylation may induce ER stress or may be induced by ER stress. It has been established that glycosylation and ER stress are essential in a variety of cellular processes and diseases. *N*-linked glycosylation within the ER is necessary for monitoring the state of protein folding and the state of glycosylation in the ER is a determinant for further processing of proteins in Golgi or destruction of improperly folded proteins in the ER associated degradation (ERAD) process. This chapter explores the interdepen-

L. Joshi (✉) · J. Q. Gerlach · S. Sharma
Glycoscience Group, National Centre for Biomedical Engineering Science (NCBES), National University of Ireland, Galway, Ireland
e-mail: Lokesh.Joshi@nuigalway.ie

K. J. Leister
Bristol-Myers Squibb, Syracuse, New York.

P. Agostinis, A. Samali (eds.), *Endoplasmic Reticulum Stress in Health and Disease,*
DOI 10.1007/978-94-007-4351-9_2,

dence of ER stress and glycosylation during protein quality control and cellular response to physiological stress.

Keywords

Golgi · Endoplasmic reticulum · *N*-glycan · *O*-glycan · Biopharmaceutical · Oligosaccharide · Lectin · Glycosidase · Chaperone · Glycosyltransferase · Calnexin · Calreticulin

Abbreviations

2dGlc	2-deoxy-Dglucose
Asn	Asparagine
BiP	immunoglobulin heavy chain-Binding Protein
CMPST	CMP-sialic acid transporter
DTT	Dithiothreitol
eIF2α	eukaryotic Initiation Factor-2α
ER	Endoplasmic Reticulum
ERAD	ER Associated Degradation
ERManI	ER α1–2 mannosidase I
EMC	ER membrane protein complex
GalNAc	N-acetyl-galactosamine
GDPFT	GDP-fucose transporter
Glc	Glucose
GlcNAc	N-actylglucosamine
GRP78	Glucose Regulated protein 78
HA	influenza hemagglutinin A
Man	Mannose
Man2C1	cytoplasmic α-mannosidase
MANEA	Golgi resident endomannosidase
PTMs	Post Translational Modifications
TM	Tunicamycin
UGGT	UDP-glucose:glycoprotein glucosyltransferase
UPR	Unfolded Protein Response
XBP1	XBP1

1 Introduction

In eukaryotic cells, the majority of proteins are modified during or soon after translation. These covalent modifications are collectively called post-translational modifications (PTMs) and serve to provide an additional level of regulation for proteins as well as to allow selective participation in multiple processes [1–3]. One of the

most prevalent PTMs is glycosylation which is the attachment of oligosaccharide structures (glycans).

With few exceptions, glycosylation of proteins occurs in the eukaryotic secretory pathway and is carried out in discrete biosynthetic steps divided between the endoplasmic reticulum (ER) and Golgi apparatus [4–9]. Unlike nucleic acids and proteins, glycan structures are not directly determined by genes, are not synthesized in a template-based manner, and may be linear or branched in structure. In some instances, glycans may be further modified by acetylation, sulfation or phosphorylation.

Glycosylation is also a fundamental part of the ER's protein quality control system, which sorts improperly folded proteins for systematic recycling. The secretory pathway, beginning with the ER and ending at the *trans*-Golgi, is charged with the delivery of properly folded and glycosylated proteins to the cell surface and this activity is vital in the development and homeostasis of all eukaryotes as well as cell-to-cell communication within multi-cellular organisms [3, 5]. Glycan moieties on glycoproteins are also involved in a wide array of functions ranging from increasing the solubility and stability of proteins to extending their circulatory half-life in the serum and also have roles in most chronic and infectious diseases [10]. It is therefore unsurprising that glycosylation pathways are profoundly influenced by ER stress [11]. Situations, which induce ER stress contribute to the altered regulation of pathways associated with the unfolded protein response (UPR), endoplasmic reticulum associated protein degradation (ERAD) and protein secretion as well as underlying levels of transcription and translation [12, 13].

The purpose of this chapter is to introduce the process of glycosylation and explore some of the recent studies and findings instrumental in connecting observed changes in glycosylation with respect to ER stress, ERAD and the UPR.

2 The *N*-linked Glycosylation Process

Oligosaccharides attached through the amine of asparagine (Asn) within a polypeptide chain are termed *N*-glycans. Because *N*-linked glycosylation begins while the polypeptide chain is still being folded in the ER and continues in the Golgi after the protein is folded, it is a co-translation as well as post-translational modification. The initial steps of *N*-linked oligosaccharide biosynthesis in the ER are similar in all eukaryotic cells. However, there are distinct differences in the processing and elongation of *N*-glycans in the Golgi complex which lead to organism-specific oligosaccharide chains (glycoforms).

N-linked glycosylation requires the consensus amino acid sequence Asn-X-Ser/Thr within nascent polypeptide chains for attachment where X can be any amino acid except proline. Transfer of the fourteen-sugar structure consisting of glucose (Glc), mannose (Man) and *N*-acetylglucosamine (GlcNAc) as a complete, preformed oligosaccharide unit ($Glc_3Man_9GlcNAc_2$) from an ER-membrane lipid do-

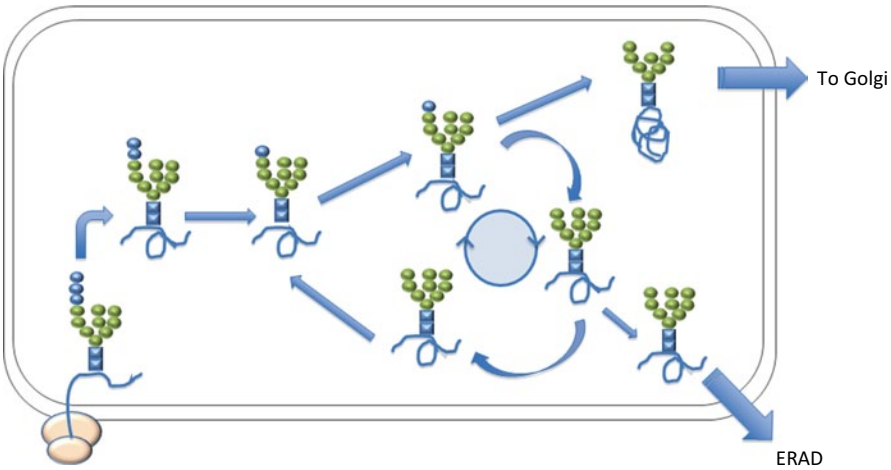


Fig. 1 N-Linked glycosylation process in the ER. GlcNAc; Man; Glc; calnexin and calreticulin mediated protein folding and reglucosylation by UGGT

nor to the sequon marks the initiation of *N*-linked glycosylation [9, 14]. Further steps of *N*-glycan processing in the ER (Fig. 1) are catalyzed by specific glycosylhydrolases (glycosidases) and glycosyltransferases. Glucosidase I removes the first Glc residue (furthest from the reducing end of the oligosaccharide) and glucosidase II removes the second Glc residues from the oligosaccharide chain. During this process two chaperon proteins, the soluble calreticulin and the membrane-bound calnexin, play a critical role in assessing the state of protein folding prior to allowing further glycosylation modification. Both calreticulin and calnexin are carbohydrate-binding proteins (lectins) and their chaperoning mechanism relies on their ability to specifically recognize the monoglucosylated oligosaccharide on the nascent protein while it remains within the ER. Normally, removal of the third and last Glc by glucosidase II is a signal of proper protein folding and exit to Golgi, however, further mechanisms ensure that misfolded proteins are not exported. For example, UDP-glucose:glycoprotein glucosyltransferase (UGGT) acts as a folding sensor by interacting with both the Glc-free *N*-glycan and the backbone of the protein being folded. If the protein is still unfolded after the removal of the last Glc, UGGT re-glucosylates the oligosaccharide chain and allows the protein another attempt at folding by associating with calreticulin and calnexin. In most cases, this cycle of deglycosylation by glucosidase II and reglycosylation by UGGT continues until the target protein is correctly folded. If it fails to fold, the target protein is sent to ERAD after association with additional chaperones such as EDEM and OS9 [15].

As properly folded proteins are prepared for export to Golgi, ER α1–2 mannosidase I (ERManI) removes one Man to generate the $Man_8GlcNAc_2$ structure, thus ensuring that the oligosaccharide can no longer be glucosylated [16]. Upon arrival in the Golgi, glycoproteins bearing the $Man_8GlcNAc_2$ structure are then further trimmed by glycosidases prior to the addition of monosaccharides which extend

the oligosaccharide chains. These sequential monosaccharide additions are carried out by sugar- and linkage-specific glycosyltransferases. In mammalian cells, *N*-linked oligosaccharide chains often possess penultimate galactose and terminal sialic acid residues at the non-reducing end (distal to the polypeptide to which they are attached). These reactions are carried out by galactosyltransferases and sialyltranferases, respectively.

3 The *O*-linked Glycosylation Process

Addition of glycans to the hydroxyl group of serine or threonine amino acid residues is known as *O*-linked glycosylation [8, 17, 18] and is thought to be an exclusively post-translational modification. Most *O*-glycosylation (initiation and maturation) of fully-folded proteins occurs entirely within the Golgi complex. Each monosaccharide is added in a sequential manner to generate linear or branched oligosaccharide structures. There are different types of *O*-linked glycosylation in eukaryotic cells but among them, addition of *N*-acetyl-galactosamine (GalNAc) to serine/threonine is the most common type in mammals which is commonly known as mucin type glycosylation [17].

4 Protein Glycosylation in Relation to ER Stress and the UPR

Together, the *N*- and *O*-linked glycosylation processes generate a very large repertoire of physiologically responsive glyco-epitopes. Even though protein folding and the influence of glycosylation during ER quality control is considered to be one of the better-understood cellular processes [11, 19, 20], elongation of *N*- and assembly of *O*-glycans in the Golgi is less completely understood and continues to complicate efforts to fully understand the complex ER/UPR/Glycosylation relationship. Only recently have the first findings concerning the effects of Golgi glycosylation on ER stress and the UPR been reported [21]. ER stress inducers that are routinely used to directly alter protein folding, such as the reducing agent dithiothreitol (DTT) or genetic mutation and resulting amino acid substitution, will ultimately alter glycosylation within the ER or suppress export to Golgi causing prolonged ER stress and activation of the UPR [22, 23]. Alternatively, agents that either block transfer of *N*-glycans to nascent proteins or inhibit completion of glycosylation, such as tunicamycin ™ derived from fungus or plant alkaloids which block glycosyltransferases, will lead to a surplus of unfolded proteins, thus creating prolonged ER stress, activation of the UPR and ultimately cell death [24, 25]. From these examples, it is clear that glycosylation will be affected in either situation and it is this duality that fogs the understanding of the specific mechanisms of glycosylation in relation to the UPR and how the occurrences of individual glycoforms are influenced by the onset and persistence of ER stress and *vice versa*.

The current state of understanding with respect to ER stress and UPR cell biology has recently been reviewed in detail (see Chap. 3). ER stress and initiation of the UPR are defined not only by the accumulation of unfolded proteins within the ER itself, but also by a surplus of incompletely glycosylated secreted proteins which are exported to expected and unexpected targets, a change in Ca^{2+}levels, occupation of resident chaperone proteins, and phosphorylation state of UPR signaling proteins [26, 27]. As these changes may ultimately manifest themselves as differential expressions of surface glycosylation, they may therefore be central to a variety of physiological phenomena [5, 22].

ER stress is associated with a wide variety of conditions and pathologies [28]. The list includes dystroglycanopathies [29], protein mis-localization and mis-trafficking [30, 31], microvascularization and tumorigenesis [32], Epstein-Barr virus replication [33], cardiovascular disease [34], diabetes [35], kidney disease [36, 37] and neuro-degenerative disorders [38], among others. It is the evidence supporting the association of ER stress with several human diseases, especially diabetes [35, 39] and cancer [40], that is largely driving interest in understanding ER stress and related mechanisms.

5 Glycosylation, ER Stress and the Biopharmaceutical Industry

The complex relationship between glycosylation, ER stress and the UPR is also an important focus of the biopharmaceutical products industry [41–45]. Many commercial therapeutic products are glycoproteins. Excessive levels of ectopic protein expression, like those encountered in large-scale biopharmaceutical production, can lead to ER stress and undesirable consequences for host cells, ultimately affecting yield and quality of recombinant products [43, 45]. Central to revealing more about ER stress, glycosylation and biopharmaceutical quality control, new analytical technologies and instrumentation have begun to enhance the ability of scientists and engineers to probe deeper into the cell's biochemistry. The enhanced measurement of genetic expression, transcription and translational controls, protein covalent modifications and secretion, could allow fine-tuning of metabolic pathways and bioenergetics to limit stress and help ensure complete glycosylation.

Regulatory agencies have increased pressures requiring companies to better understand and more tightly control bioprocesses that produce commercial biotherapeutics. As a result, the biopharmaceutical industry has increased its efforts to explore pathways which impact glycoprotein quality using systems biology approach combining transcriptomic, proteomic, glycomic and metabolomic measurements. A more thorough understanding of how inputs affect output parameters in both upstream and downstream process steps is necessary. Conventional bioreactor experiments exploring the impacts of soluble gas ratios and changing nutrients have already lead to clues about the cellular biochemical pathways affected and how these in turn influence the quality and quantity of the biotherapeutic product.

6 Recent Examinations of Glycosylation with Respect to ER Stress and the UPR

At the genetic level, there is significant overlap of key players in ER stress/UPR and glycosylation machinery [46–48]. RNA-based expression studies have been aimed at establishing links to individual secretory pathway genes for specific characteristics induced by ER stress. Glycosylation pathway genes ('glycogenes') account for approximately 5 % of the total expressed genome and ER-stress related expression changes of these are generally moderate across studies. Although the percentage and fold changes may appear small, their impact is widespread because the majority of the eukaryotic proteome is glycosylated. Altered expression of glycogenes has structural and functional implications on a large number of carrier proteins as well as lipids. Several expression studies [49–54] conducted over the past few years have aimed at establishing links to individual secretory pathway genes for specific characteristics induced by ER stress. Some more recent examples of these transcriptomic studies have examined gene expression changes associated with ER stress in fungi [55, 56], plants [57–61] and animals or cultured animal cells [45, 62]. Variations in the chemical treatments used to induce ER stress, such as thapsigargin, tunicamycin [50, 54], salubrinal [33], antimycin, buformin, metformin, phenformin, rotenone, versipelostatin [63] have produced different responses reflecting the slightly different mechanism(s) of inducing ER stress. Finally, individual cell lines may also demonstrate different responses to a single UPR modulator. For example, similar thapsigargin treatment of human medullablastoma cell lines [54] and lymphoblastoid cell lines [33] produced differing glycogene regulation (Table 1). All of these factors have contributed to a poor consensus about particular responses among existing data.

Within the network of genes related to ER stress and the UPR, there is still a number of uncertainties with respect to the necessity of individual components. A recent study of yeast by Bircham *et al.* [64] has essentially suggested that the UPR is non-essential within a range of non-optimal conditions because a wide margin of compensation mechanisms can still function to maintain homeostasis in the absence of a functioning UPR network. The ER membrane protein complex genes (*EMC*) associated with the UPR are believed to be part of the early chaperone-driven protein folding functions of the ER and four of these have been previously shown to produce UPR activation when deleted. To evaluate the necessity of *EMC* genes 1 through 6 and additional genes in the UPR, an automated confocal microscopy protocol was developed to monitor changes in localization of GFP-labeled plasma membrane proteins as a result of UPR gene deletion and these results were compared to microarray analyses. Not surprisingly, the authors discovered that yeast cells were hypersensitive to TM and DTT treatment as a consequence of removing the additional homeostatic regulation afforded by the UPR, but the mutant cells still had the ability to recover homeostasis at low doses of these ER stress agents. Specifically pertaining to glycosylation, one of the most interesting outcomes of the study was the observation that GFP-modified carboxypeptidase Y produced in *EMC1–6* double mutant cells experiencing significant ER stress had no detectable changes in glycosylation.

Table 1 Differentially regulated glycogenes involved in ER stress. Six different transcriptomic studies relevant to ER stress were retrieved from Gene Expression Omnimus (GEO, NCBI) and analyzed. Negative (−) sign in fold change columns indicates down regulated genes and blank columns are non differentially regulated glycogenes

Symbol	Description	GeneBank	Fold change							
			GSE8562 (Gomez *et al.* 2007)	GSE21979 (Pereira *et al.* 2010)	GSE2980 (Bridges *et al.* 2006)	GSE22097 (Kennedy *et al.* 2010)	GSE19519, (Dombroski *et al.* 2010)		GSE 31447, (Taylor *et al.* 2011)	
			XBP vs Control	Thapsigagrin vs Control	Exon4 mutant vs WT	hyperglycemic vs nondiabetic	Tunicamycin vs DMSO	Thapsigargin vs DMSO	Thapsigargin vs DMSO	Salubrinal vs DMSO
CALR	Calreticulin	NM_004343	1.26		−1.20	1.91	1.91	1.22		
CANX	Calnexin	NM_001746	–	1.57		−1.45	1.55	1.20	1.28	1.29
HPRT1	Hypoxanthine phosphoribosyltransferase 1	NM_000194			1.62		−1.26			
HSPA5	Heat shock 70 kDa protein 5 (glucose-regulated protein, 78 kDa)	NM_005347		8.94		1.77	2.99	1.98		
OS9	Osteosarcoma amplified 9, endoplasmic reticulum lectin	NM_006812		1.82	−1.15	1.69	1.77	1.13		
RPN1	Ribophorin I	NM_002950	1.75	1.46	1.14		1.82	1.19		
USP14	Ubiquitin specific peptidase 14 (tRNA-guanine transglycosylase)	NM_005151				−1.55	1.11	1.20	1.27	

The carboxypeptidase Y expressed by the mutant yeast did, however, demonstrate a lower surface localization efficiency and a higher proportion of the tagged protein was improperly localized to intracellular membranes [64]. Because yeast naturally do not have the ability to assemble hybrid or complex *N*-glycan structures, it may be possible that similar UPR-gene deletions in higher eukaryotes would produce vastly different results with respect to glycosylation of proteins which normally display more advanced glycan structures.

In contrast, a complete and functional UPR network has been suggested to be essential for the proper induction of apoptosis in some studies, including in the case of breast cancer treatment in a mouse model, again with the inhibitor of glycosylation, TM [32]. As a potential cancer treatment mechanism, TM induction of the UPR via ER stress was shown to effectively cause tumor cell death [32]. Findings like these have turned attention toward anti-tumor strategies incorporating chemical modulation of ER stress and continued activation of the UPR in order to induce apoptosis [25, 32, 48, 65, 66, 67]. It is important to note that chemical agents which may alter glycosylation and induce ER stress, such as TM, have a more global influence on the cellular physiology and that the observations made by researchers may or may not be directly due to ER stress events.

Global feedback mechanisms controlling pathways directly associated with the building and transfer of the oligosaccharide structures as well as those governing translation of proteins which are being glycosylated work in concert. The already complex relationship between glycosylation and protein expression levels is further complicated during ER stress. When populations of oligosaccharide structures demonstrate changes in the presence of ER stress inducing chemicals, including DTT, deoxynorjirimycin, Brefeldin A or TM, defining whether the origin of change is a product of translational feedback or posttranslational feedback requires that both systems and even those which feed into them (*e.g.* glycolysis and the pentose-phosphate pathway) be monitored for their own respective changes [62, 68]. Glycoform inventory is affected not only by the control of glycosyltransporters, glycosyltransferases and glycosylhydrolases, it is also affected by the availability of substrate nucleotide sugars, and nascent polypeptides [5, 7, 69, 70]. For example, total oligosaccharide populations during ER stress may show that particular oligosaccharide structures across the cell population decrease or increase, but determining the specific origin of these changes often is not as straightforward as just being the result of up or down regulation of glycosyltransferase or glycosidase expression. The application of a systems biology approach to studies involving glycosylation, ER stress and the UPR will be necessary to fully address these issues.

Influences external to the secretory pathway are crucial to the ER, UPR, and glycosylation networks. Glycolysis, the breakdown of glucose for use as an energy source or material for anabolic processes, occurs in the cytoplasm of eukaryotic cells. Yet, treatment of immortalized cells with the glycolysis inhibitor 2-deoxy-D-glucose (2dGlc) directly impacts glycosylation and ER stress [25]. Autophagy is a cellular survival mechanism which is induced along with cell death (apoptosis) during extreme stress. In their work on autophagy and ER stress induced by 2dGlc, Xi and colleagues demonstrated that an increase in exogenous mannose had the ability to abolish autophagy and also reduce cell stress [25].

In general, ER stress results in hypoglycosylation of which observed effects include a decrease in glyco-diversity and an increase in the relative number of high-mannose structures on secreted proteins. As a result of interest in ER stress-related hypoglycosylation, ER-resident glycosidases and lectins are a popular subject of study [23, 71, 72]. There has been particular interest in identifying further details associated with the final sorting steps which ultimately decide the fate of a newly-translated and folded proteins and whether they will be exported or degraded.

The relationship between ERAD and the UPR as a result of oligosaccharide trimming by ERManI has recently been examined [23]. ERManI, as introduced at the beginning of this chapter, is responsible for trimming high mannose core *N*-linked oligosaccharide structures on newly-synthesized proteins which reside in the ER for extended lengths of time, as in situations of stalled protein export to the Golgi. HEK293 cells were transfected with a gene for α1-antitrypsin containing a mutation which causes the protein to misfold. Through pulse-chase experiments, the authors discovered that ERManI expression was stabilized during ER stress and the resulting UPR, allowing a greater percentage of misfolded α1-antitrypsin protein in the ER to enter the ERAD system for disposal and in turn reducing ER stress [64].

During ERAD, misfolded proteins are transported to the cytoplasm and tagged with ubiquitin. Proteins are then systematically deglycosylated by oligosaccharide-specific cytoplasmic amidases and glycosidases [73]. One consequence of excessive ERAD activity in response to ER stress is a cytoplasmic accumulation of neutral oligosaccharides containing glucose (these are collectively termed free oligosaccharides, or FOS) [74, 75]. By the action of select glycosidases, constituent monosaccharides (Glc, Man, GlcNAc) accumulate as well, and these directly affect both processing of precursor sugars and assembly of oligosaccharides destined for ER-localized attachment to proteins [68]. The cytoplasmic α-mannosidase, Man2C1, was shown to be a key component of feedback mechanisms related to the breakdown or accumulation of free oligosaccharide chains resulting from ERAD activity in response to ER stress [73]. Through the use of HeLa cells overexpressing Man2C1, the authors demonstrated that the regulation of such glycosidase activity is tightly connected to the *N*-linked glycosylation process. Underglycosylation of GFP-tagged Null Hong Kong α-antitrypsin protein and β1-integrin was observed in conjunction with a 2.5-fold increase in transformed HeLa cells overexpressing Man 2C1 [73].

Malectin is an ER-resident lectin discovered in an African frog and later found to be highly conserved across metazoans [15, 76]. Because of its co-localization with calnexin, malectin has been hypothesized as a part of a ‘backup’ quality control mechanism which recognizes carbohydrates structures and operates during ER stress. *In vitro*, malectin has been shown to bind to di-glucosyl oligossacharides with high affinity; however the exact role of the lectin is not yet understood. HEK293 cells did not show differential expression of calnexin as a result of increased malectin expression and there was no change in the processing of influenza hemagglutinin A (HA) in cells expressing high levels of the lectin [76]. Interestingly, the amount of HA/malectin association in the ER was affected by folding efficiency of the viral protein as fully oxidized, rapidly folding molecules showed a delayed association with malectin in comparison to partially-oxidized counterparts. Malectin was also

shown to preferentially associate with HA molecules sensitive to deglycosylation with *endo*-glycosidase H, thus further reinforcing that it is involved in recognition of specific carbohydrate structures within the ER [76].

Other mechanisms may lead to an accumulation of underglycosylated or un-glycosylated proteins in the cell. The proteasome is a protein complex dedicated to the degradation of ubiquinated proteins during ERAD which operates in the cytoplasm. Prolonged ER stress chemically induced by the application of the proteasome inhibitor MG132 to cultured cells resulted in mis-localization of immunoglobulins and type I transmembrane protein, US2, and the observation of a high percentage of these proteins lacking *N*-glycans [30].

New evidence points to the existence of additional mechanisms of protein degradation related to quality control that operate outside of the ER. In metazoans, the Golgi resident endomannosidase (MANEA) acts to remove the terminal residues (Glc-α1–2−Man α1−) attached to incomplete *N*-glycans. This mechanism is separate from the ER-resident exo-mannosidase, ERMan1, that normally operates after the removal of the final Glc. The removal of the terminal Glc-α1–2−Man α1− allows glycoproteins to escape the calnexin/calreticulin cycle and can allow even glycoproteins which are improperly folded access to Golgi-resident glycosylation processing and elongation [77]. A MANEA knockout Chinese hamster ovary (CHO-K1) cell line was transfected with an HA-tagged version of the gene for the mannosidase (HA-MANEA) and the localization of the enzyme traced to the *medial*-Golgi. HA-MANEA expression restored Golgi processing in CHO-K1 cells and reduced the overall level of FOS. However, the authors were able to detect a relative increase in FOS without Glc upon the introduction of the plant-derived ER glucosidase inhibitor *N*-butyldeoxynorjirimycin, effectively showing that a mechanism for degradation of incompletely glycosylated proteins does exist outside of the ER. These observations and the implication that a second mechanism of quality control-related degradation is active in the Golgi also point to the existence of considerable antero/retrograde trafficking of glucosylated glycoproteins between the ER and Golgi during conventional calreticulin/calnexin cycling [77].

Until recently, research aimed at ER stress and the UPR had largely ignored the effects of glycosylation carried out exclusively in Golgi. Two transporters responsible for the import and localization of nucleotide sugar substrates critical for the assembly of elongated *N*- and *O*-glycans in the Golgi were silenced (87–95%) by short interfering RNA (siRNA) in HeLa cells to study the effects on protein accumulation and resulting stress responses [21]. The study targeted a GDP-fucose transporter (GDPFT) vital for the attachment of fucose (Fuc) and a CMP-sialic acid transporter (CMPST) necessary for transfer of acetylated sialic acid (Neu5Ac) to glycans inside the Golgi stacks. As a result of the reduced completion of glycans, proteins accumulated in the Golgi which ultimately led to activation of the ER stress response (introduced by a slowdown of traffic from ER to Golgi) and the induction of UPR as indicated by phosphorylation of the eukaryotic initiation factor 2 (eIF2α), splicing of X-box binding protein one (XBP1) mRNA and BiP/GRP78 upregulation. These effects were similar to the those seen with TM treatment of HeLa cells without siRNA silencing of CMPST and GDPFT [21]. These results further

call to attention the interconnected nature of all parts of the glycosylation and secretory process in relation to ER stress and the UPR.

The complete elimination of glycosylation by mutation of glycosylation attachment residues may also induce the UPR. E-cadherin is a cell-surface glycoprotein with four reported glycosylation sites. The elimination of one of these glycosylation sites through a site-directed mutation resulted in the protein being directed to ERAD [78] most likely due to the continued exposure of hydrophobic domains near the sequon normally controlled by the presence of the hydrophilic oligosaccharide at Asn 633. The necessity of *N*-glycosylation during the folding process has been demonstrated in a variety of additional studies [31, 79–82]. Not all glycoproteins require intact glycosylation for proper folding or to move through the secretory pathway, however. One recent study of the maturation of sodium/potassium ATPase subunits in canine kidney cells included the removal of all three glycosylation sites from the β1 subunit had no effect on the localization of the protein to the plasma membrane [83]. On the other hand, the authors demonstrated that its counterpart, the β2 subunit, required glycosylation and chaperone-mediated folding for proper maturation and localization. It is important to note that while proteins may be folded and directed to the proper location, they may have lost function simply because the glycosylation is a required part of the protein's functional apparatus such as in cell communication or adhesion [3].

7 Conclusions

Cellular functions which depend upon glycosylation (and their associated pathways) and their relationship to ER stress and protein folding are just beginning to be unraveled. Work to understand the intertwined nature of the ER/UPR/Glycosylation systems has been ongoing for some time, but individual studies still tend to focus on a single part of the network irrespective of the rest, often simply as a result of experimental necessity. It is now clear that glycosylation, which is among the most important PTMs, is closely associated with ER stress, ERAD, the UPR, autophagy and apoptosis. Considerable momentum for these areas of research has been generated by the realization that ER stress and the UPR are potential therapeutic targets to fight disease [40, 48, 67]. These promising areas of research will benefit from cross-validating several experimental approaches. By bringing systems biology approaches to these fields of research, integration of information about responses across all of the connected portions of the secretory pathway and the UPR will finally become a reality.

Acknowledgments The authors would like to acknowledge the Alimentary Glycosciences Research Cluster (AGRC) funded by Science Foundation Ireland (grant number 08/SRC/B1393) and the European Union FP7-funded GlycoHIT (Glycomics by High throughput Integrated Technologies) consortium (EU FP7 260600).

References

1. Walsh CT, Garneau-Tsodikova S, Gatto GJ Jr. (2005) Protein posttranslational modifications: the chemistry of proteome diversifications. Angew Chem Int Ed Engl 44(45):7342–7372
2. Freeze HH, Schachter H (2009) Genetic disorders of glycosylation. In: Varki et al (ed) Essentials of glycobiology, 2nd edn. Cold Spring Harbor Press, New York
3. Boscher C, Dennis JW, Nabi IR (2011) Glycosylation, galectins and cellular signaling. Curr Opin Cell Biol 23(4):383–392
4. Hulsmeier AJ, Welti M, Hennet T (2011) Glycoprotein maturation and the UPR. Methods enzymol. 491:163–182
5. Dennis JW, Lau KS, Demetriou M, Nabi IR (2009) Adaptive regulation at the cell surface by N-glycosylation. Traffic 10(11):1569–1578
6. Nilsson T, Au CE, Bergeron JJ (2009) Sorting out glycosylation enzymes in the Golgi apparatus. FEBS Lett 583(23):3764–3769
7. Sesma JI, Esther CR Jr, SM Kreda, Jones L, Neal WO, Nishihara S, Nicholas RA, Lazarowski ER (2009) Endoplasmic reticulum/golgi nucleotide sugar transporters contribute to the cellular release of UDP-sugar signaling molecules. J Biol Chem 284(18):12572–12583
8. Stanley P (2011) Golgi glycosylation. Cold Spring Harb Perspect Biol 3(4):1–13
9. Stanley P, Schachter H, Taniguchi N (2009) N-Glycans. In: Varki A et al (ed) Essentials of glycobiology, 2nd edn. Cold Spring Harbor Press, New York
10. Varki A, Lowe JB (2009) Biological Roles of Glycans. In: Varki A et al (ed) Essentials of glycobiology. 2nd edn. Cold Spring Harbor Press, New York
11. Ruddock LW, Molinari M (2006) N-glycan processing .In: ER quality control. J Cell Sci 119(Pt 21):4373–4380
12. Zhang K, Kaufman RJ (2006) The unfolded protein response: a stress signaling pathway critical for health and disease. Neurology 66(2 Suppl 1):S102–109
13. Chakrabarti A, Chen AW, Varner JD (2011) A review of the mammalian unfolded protein response. Biotechnol Bioeng 108(12):2777–2793
14. Pearse BR, Hebert DN (2010) Lectin chaperones help direct the maturation of glycoproteins in the endoplasmic reticulum. Biochim Biophys Acta 1803(6):684–693
15. Chen Y, Hu D, Yabe R, Tateno H, Qin SY, Matsumoto N, Hirabayashi J, Yamamoto K (2011) Role of malectin in Glc(2)Man(9)GlcNAc(2)-dependent quality control of alpha1-antitrypsin. Mol Biol Cell 22(19):3559–3570
16. Avezov E, Frenkel Z, Ehrlich M, Herscovics A, Lederkremer GZ (2008) Endoplasmic reticulum (ER) mannosidase I is compartmentalized and required for N-glycan trimming to Man5–6GlcNAc2 in glycoprotein ER-associated degradation. Mol Biol Cell 19(1):216–225
17. Brockhausen I, Schachter H, Stanley P (2009) OGalNAc glycans. In: Varki et al (ed) Essentials of glycobiology, 2nd edn. Cold Spring Harbor Press, New York
18. Peter-Katalinic J (2005) Methods in enzymology: O-glycosylation of proteins. Methods enzymol 405:139–171
19. Lederkremer GZ (2009) Glycoprotein folding, quality control and ER-associated degradation. Curr Opin Struct Biol 19(5):515–523
20. Yamamoto K (2009) Intracellular lectins involved in folding and transport in the endoplasmic reticulum. Biol Pharm Bull 32(5):767–773
21. Xu YX, Liu L, Caffaro CE Hirschberg CB (2010) Inhibition of Golgi apparatus glycosylation causes endoplasmic reticulum stress and decreased protein synthesis. J Biol Chem 285(32):24600–24608
22. Kimata Y, Ishiwata-Kimata Y, Yamada S, Kohno K (2006) Yeast unfolded protein response pathway regulates expression of genes for anti-oxidative stress and for cell surface proteins. Genes Cells 11(1):59–69

23. Termine DJ, Moremen KW, Sifers RN (2009) The mammalian UPR boosts glycoprotein ERAD by suppressing the proteolytic downregulation of ER mannosidase I. J Cell Sci 122(Pt 7):976–984
24. Hsu JL, Chiang PC, Guh JH (2009) Tunicamycin induces resistance to camptothecin and etoposide in human hepatocellular carcinoma cells: role of cell-cycle arrest and GRP 78. Naunyn Schmiedebergs Arch Pharmacol 380(5):373–382
25. Xi H, Kurtoglu M, Liu H, Wangpaichitr M, You M, Liu X, Savaraj N, Lampidis TJ (2010) 2-Deoxy-D-glucose activates autophagy via endoplasmic reticulum stress rather than ATP depletion. Cancer Chemother Pharmacol 67(4):899–910
26. Groenendyk J, Sreenivasaiah PK, Kim H do, Agellon LB, Michalak M (2010) Biology of endoplasmic reticulum stress in the heart. Circ Res 107(10):1185–1197
27. Michalak M, Groenendyk J, Szabo E, Gold LI, Opas M (2009) Calreticulin, a multi-process calcium-buffering chaperone of the endoplasmic reticulum. Biochem J 417(3):651–666
28. Yoshida H (2007) ER stress and diseases. FEBS J 274(3):630–658
29. Lin YY, White RJ, Torelli S, Cirak S, Muntoni F, Stemple DL (2011) Zebrafish Fukutin family proteins link the unfolded protein response with dystroglycanopathies. Hum Mol Genet 20(9):1763–1775
30. Drori A, Misaghi S, Haimovich J, Messerle M, Tirosh B (2010) Prolonged endoplasmic reticulum stress promotes mislocalization of immunoglobulins to the cytoplasm. Mol Immunol 47(9):1719–1727
31. Romero-Fernandez W, Borroto-Escuela DO, Perez Alea M, Garcia-Mesa Y, Garriga P (2011) Altered trafficking and unfolded protein response induction as a result of M3 muscarinic receptor impaired N-glycosylation. Glycobiology 21(12):1663–1672
32. Banerjee A, Lang JY, Hung MC, Sengupta K, Banerjee SK, Baksi K, Banerjee DK (2011) Unfolded protein response is required in nu/nu mice microvasculature for treating breast tumor with tunicamycin. J Biol Chem 286(33):29127–29138
33. Taylor GM, Raghuwanshi SK, Rowe DT, Wadowsky R, Rosendorff A (2011) ER-stress causes Epstein-Barr virus lytic replication. Blood 118(20):5528–5539
34. Glembotski CC (2008) The role of the unfolded protein response in the heart. J Mol Cell Cardiol 44(3):453–459
35. Schnell S (2009) A model of the unfolded protein response: pancreatic beta-cell as a case study. Cell Physiol Biochem 23(4–6):233–244
36. Adam J, Bollee G, Fougeray S, Noel LH, Antignac C, Knebelman B, Pallet N (2011) Endoplasmic reticulum stress in umod-related kidney disease: a human pathologic study. Am J Kidney Dis
37. Cybulsky AV (2010) Endoplasmic reticulum stress in proteinuric kidney disease. Kidney Int 77(3):187–193
38. Doyle KM, Kennedy D, Gorman AM, Gupta S, Healy SJ, Samali A (2011) Unfolded proteins and endoplasmic reticulum stress in neurodegenerative disorders. J Cell Mol Med 15(10):2025–2039
39. Jeschke MG, Boehning D (2011) Endoplasmic reticulum stress and insulin resistance post-trauma: similarities to type 2 diabetes. J Cell Mol Med 16(3):437–444
40. Boelens J, Lust S, Offner F, Bracke ME, Vanhoecke BW (2007) Review. The endoplasmic reticulum: a target for new anticancer drugs. In Vivo 21(2):215–226
41. Lefloch F, Tessier B, Chenuet S, Guillaume JM, Cans P, Goergen JL, Marc A (2006) Related effects of cell adaptation to serum-free conditions on murine EPO production and glycosylation by CHO cells. Cytotechnology 52(1):39–53
42. Ohya T, Hayashi T, Kiyama E, Nishii H, Miki H, Kobayashi K, Honda K, Omasa T, Ohtake H (2008) Improved production of recombinant human antithrombin III in Chinese hamster ovary cells by ATF4 overexpression. Biotechnol Bioeng 100(2):317–324

43. Omasa T, Takami T, Ohya T, Kiyama E, Hayashi T, Nishii H, Miki H, Kobayashi K, Honda K, Ohtake H (2008) Overexpression of GADD34 enhances production of recombinant human antithrombin III in Chinese hamster ovary cells. J Biosci Bioeng 106(6):568–573
44. Droogenbroeck B Van, Cao J, Stadlmann J, Altmann F, Colanesi S, Hillmer S, Robinson DG, Lerberge E Van, Terryn N, Montagu M Van, Liang M, Depicker A, Jaeger G De (2007) Aberrant localization and underglycosylation of highly accumulating single-chain Fv-Fc antibodies in transgenic Arabidopsis seeds. Proc Natl Acad Sci U S A 104(4):1430–1435
45. Young CL, Yuraszeck T, Robinson AS (2011) Decreased secretion and unfolded protein response upregulation. Methods Enzymol. 491:235–260
46. Arvas M, Pakula T, Lanthaler K, Saloheimo M, Valkonen M, Suortti T, Robson G, Penttila M (2006) Common features and interesting differences in transcriptional responses to secretion stress in the fungi Trichoderma reesei and Saccharomyces cerevisiae. BMC Genomics 7:32
47. Jorgensen TR, Goosen T, Hondel CA, Ram AF, Iversen JJ (2009) Transcriptomic comparison of Aspergillus niger growing on two different sugars reveals coordinated regulation of the secretory pathway. BMC Genomics 10:44
48. Saito S, Tomida A (2011) Use of chemical genomics in assessment of the UPR. Methods Enzymol 491:327–341
49. Bridges JP, Xu Y, Na CL, Wong HR, Weaver TE (2006) Adaptation and increased susceptibility to infection associated with constitutive expression of misfolded SP-C. J Cell Biol 172(3):395
50. Dombroski BA, Nayak RR, Ewens KG, Ankener W, Cheung VG, Spielman RS (2010) Gene expression and genetic variation in response to endoplasmic reticulum stress in human cells. Am J Hum Genet 86(5):719–729
51. Gomez BP, Riggins RB, Shajahan AN, Klimach U, Wang A, Crawford AC, Zhu Y, Zwart A, Wang M, Clarke R (2007) Human X-box binding protein-1 confers both estrogen independence and antiestrogen resistance in breast cancer cell lines. FASEB 21(14):4013
52. Kennedy J, Katsuta H, Jung MH, Marselli L, Goldfine AB, Balis UJ, Sgroi D, Bonner-Weir S, Weir GC (2010) Protective unfolded protein response in human pancreatic beta cells transplanted into mice. PloS One. 5(6):e11211
53. Taylor GM, Raghuwanshi SK, Rowe DT, Wadowsky RM, Rosendorff A (2011) Endoplasmic reticulum stress causes EBV lytic replication. Blood 118(20):5528–5539
54. Pereira ER, Liao N, Neale GA, Hendershot LM (2010) Transcriptional and post-transcriptional regulation of proangiogenic factors by the unfolded protein response. PloS One 5(9):e12521
55. Graf A, Gasser B, Dragosits M, Sauer M, Leparc GG, Tuchler T, Kreil DP, Mattanovich D (2008) Novel insights into the unfolded protein response using Pichia pastoris specific DNA microarrays. BMC Genomics 9:390
56. Geysens, S, G Whyteside, D B Archer (2009) Genomics of protein folding in the endoplasmic reticulum, secretion stress and glycosylation in the aspergilli. Fungal Genetics and Biology. 46(Supp. 1):S 121–S 140
57. Liu JX and Howell SH (2011) Endoplasmic reticulum protein quality control and its relationship to environmental stress responses in plants. Plant Cell 22(9):2930–2942
58. Lu DP, Christopher DA (2008) Endoplasmic reticulum stress activates the expression of a sub-group of protein disulfide isomerase genes and AtbZIP60 modulates the response in Arabidopsis thaliana. Mol Genet Genomics 280(3):199–210
59. Urade R (2009) The endoplasmic reticulum stress signaling pathways in plants. Biofactors. 35(4):326–331
60. Wakasa Y, Yasuda H, Oono Y, Kawakatsu T, Hirose S, Takahashi H, Hayashi S, Yang L, Takaiwa F (2011) Expression of ER quality control-related genes in response to changes in BiP1 levels in developing rice endosperm. Plant J. 65(5):675–689

61. Williams B, Kabbage M, Britt R, Dickman MB (2010) AtBAG7, an Arabidopsis Bcl-2-associated athanogene, resides in the endoplasmic reticulum and is involved in the unfolded protein response. Proc Natl Acad Sci U S A. 107(13):6088–6093
62. Shang J, Gao N, Kaufman RJ, Ron D, Harding HP, Lehrman MA (2007) Translation attenuation by PERK balances ER glycoprotein synthesis with lipid-linked oligosaccharide flux. J Cell Biol 176(5):605–616
63. Haga N, Saito S, Tsukumo Y, Sakurai J, Furuno A, Tsuruo T, Tomida A (2010) Mitochondria regulate the unfolded protein response leading to cancer cell survival under glucose deprivation conditions. Cancer Sci 101(5):1125–1132
64. Bircham PW, Maass DR, Roberts CA, Kiew PY, Low YS, Yegambaram M, Matthews J, Jack CA, Atkinson PH (2011) Secretory pathway genes assessed by high-throughput microscopy and synthetic genetic array analysis. Mol Biosyst 7(9):2589–2598
65. Chiang PC, Hsu JL, Yeh TC, Pan SL, Guh JH (2008) Elucidation of susceptible factors to endoplasmic reticulum stress-mediated anticancer activity in human hepatocellular carcinoma. Naunyn schmiedebergs Arch Pharmacol 377(2):167–1677
66. King FW, Fong S, Griffin C, Shoemaker M, Staub R, Zhang YL, Cohen I, Shtivelman E (2009) Timosaponin AIII is preferentially cytotoxic to tumor cells through inhibition of mTOR and induction of ER stress. PLoS One. 4(9):e7283
67. Ling YH, Li T, Perez-Soler R, Haigentz M Jr (2009) Activation of ER stress and inhibition of EGFR N-glycosylation by tunicamycin enhances susceptibility of human non-small cell lung cancer cells to erlotinib. Cancer Chemother Pharmacol 64(3):539–548
68. Freeze HH, Elbein AD (2009) Glycosylation Precursors
69. Lavery GG, Walker EA, Turan N, Rogoff D, Ryder JW, Shelton JM, Richardson JA, Falciani F, White PC, Stewart PM, Parker KL, McMillan DR (2008) Deletion of hexose-6-phosphate dehydrogenase activates the unfolded protein response pathway and induces skeletal myopathy. J Biol Chem 283(13):8453–8461
70. Tan SX, Teo M, Lam YT, Dawes IW, Perrone GG (2009) Cu, Zn superoxide dismutase and NADP(H) homeostasis are required for tolerance of endoplasmic reticulum stress in Saccharomyces cerevisiae. Mol Biol Cell 20(5):1493–1508
71. Foulquier F, Harduin-Lepers A, Duvet S, Marchal I, Mir AM, Delannoy P, Chirat F, Cacan R (2002) The unfolded protein response in a dolichyl phosphate mannose-deficient Chinese hamster ovary cell line points out the key role of a demannosylation step in the quality-control mechanism of N-glycoproteins. Biochem J 362(Pt 2):491–498
72. Miyake K, Nagai K (2009) Inhibition of alpha-mannosidase attenuates endoplasmic reticulum stress-induced neuronal cell death. Neurotoxicol 30(1):144–150
73. Bernon C, Carre Y, Kuokkanen E, Slomianny MC, Mir AM, Krzewinski F, Cacan R, Heikinheimo P, Morelle, Michalski JC, Foulquier F, Duvet S (2011) Overexpression of Man2C1 leads to protein underglycosylation and upregulation of endoplasmic reticulum-associated degradation pathway. Glycobiology. 21(3):363–375
74. Alonzi DS, Neville DC, Lachmann RH, Dwek RA, Butters TD (2008) Glucosylated free oligosaccharides are biomarkers of endoplasmic- reticulum alpha-glucosidase inhibition. Biochem J 409(2):571–580
75. Hirayama H, Seino J, Kitajima T, Jigami Y, Suzuki T (2010) Free oligosaccharides to monitor glycoprotein endoplasmic reticulum-associated degradation in Saccharomyces cerevisiae. J Biol Chem 285(16):12390–12404
76. Galli C, Bernasconi R, Solda T, Calanca V, Molinari M (2011) Malectin participates in a backup glycoprotein quality control pathway in the mammalian ER. PLoS One. 6(1):e16304
77. Kukushkin NV, Alonzi DS, Dwek RA, Butters TD (2011) Demonstration that endoplasmic reticulum-associated degradation of glycoproteins can occur downstream of processing by endomannosidase. Biochem J 438(1):133–142

78. Zhou F, Su J, Fu L, Yang Y, Zhang L, Wang L, Zhao H, Zhang D, Li Z, Zha X (2008) Unglycosylation at Asn-633 made extracellular domain of E-cadherin folded incorrectly and arrested in endoplasmic reticulum, then sequentially degraded by ERAD. Glycoconj J 25(8):727–740
79. Glozman R, Okiyoneda T, Mulvihill CM, Rini JM, Barriere H, Lukacs GL (2009) N-glycans are direct determinants of CFTR folding and stability in secretory and endocytic membrane traffic. J Cell Biol 184(6):847–862
80. Chan CP, Mak TY, Chin KT, Ng IO, Jin DY (2010) N-linked glycosylation is required for optimal proteolytic activation of membrane-bound transcription factor CREB-H. J Cell Sci 123(Pt 9):1438–1448
81. Maattanen P, Gehring K, Bergeron JJ, Thomas DY (2010) Protein quality control in the ER: the recognition of misfolded proteins. Semin Cell Dev Biol 21(5):500–511
82. Nakajima M, Koga T, Sakai H, Yamanaka H, Fujiwara R, Yokoi T (2010) N-Glycosylation plays a role in protein folding of human UGT1A 9. Biochem Pharmacol 79(8):1165–1172
83. Tokhtaeva E, Sachs G, Vagin O (2010) Diverse pathways for maturation of the Na, K-ATPase beta1 and beta2 subunits in the endoplasmic reticulum of Madin-Darby canine kidney cells. J Biol Chem 285(50):39289–39302

http://www.springer.com/978-94-007-4350-2

Endoplasmic Reticulum Stress in Health and Disease
Agostinis, P.; Afshin, S. (Eds.)
2012, VIII, 452 p., Hardcover
ISBN: 978-94-007-4350-2

Appendix-3: Neuronal glycosylation differentials in normal, injured and chondroitinase-treated environments

Biochemical and Biophysical Research Communications 420 (2012) 616–622

Contents lists available at SciVerse ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Neuronal glycosylation differentials in normal, injured and chondroitinase-treated environments

Michelle Kilcoyne [a], Shashank Sharma [a], Niamh McDevitt [b], Claire O'Leary [b], Lokesh Joshi [a], Siobhán S. McMahon [b,*]

[a] Glycoscience Group, National Centre for Biomedical Engineering Science, National University of Ireland, Galway, Ireland
[b] Anatomy, School of Medicine, National University of Ireland, Galway, Ireland

ARTICLE INFO

Article history:
Received 23 February 2012
Available online 20 March 2012

Keywords:
Spinal cord injury
Lectins
Glycosylation
Chondroitin sulphate proteoglycans
Neurons
In silico analysis

ABSTRACT

Glycosylation is found ubiquitously throughout the central nervous system (CNS). Chondroitin sulphate proteoglycans (CSPGs) are a group of molecules heavily substituted with glycosaminoglycans (GAGs) and are found in the extracellular matrix (ECM) and cell surfaces. Upon CNS injury, a glial scar is formed, which is inhibitory for axon regeneration. Several CSPGs are up-regulated within the glial scar, including NG2, and these CSPGs are key inhibitory molecules of axonal regeneration. Treatment with chondroitinase ABC (ChABC) can neutralise the inhibitory nature of NG2. A gene expression dataset was mined *in silico* to verify differentially regulated glycosylation-related genes in neurons after spinal cord injury and identify potential targets for further investigation. To establish the glycosylation differential of neurons that grow in a healthy, inhibitory and ChABC-treated environment, we established an indirect co-culture system where PC12 neurons were grown with primary astrocytes, Neu7 astrocytes (which over-express NG2) and Neu7 astrocytes treated with ChABC. After 1, 4 and 8 days culture, lectin cytochemistry of the neurons was performed using five fluorescently-labelled lectins (ECA MAA, PNA, SNA-I and WFA). Usually α-(2,6)-linked sialylation scarcely occurs in the CNS but this motif was observed on the neurons in the injured environment only at day 8. Treatment with ChABC was successful in returning neuronal glycosylation to normal conditions at all timepoints for MAA, PNA and SNA-I staining, and by day 8 in the case of WFA. This study demonstrated neuronal cell surface glycosylation changes in an inhibitory environment and indicated a return to normal glycosylation after treatment with ChABC, which may be promising for identifying potential therapies for neuronal regeneration strategies.

1. Introduction

Complex carbohydrates are found ubiquitously throughout the CNS and are involved in many developmental and functional processes in the nervous system, including cell–cell and cell–ECM interactions, adhesion and axonal guidance and neuronal migration [1,2]. *In vivo*, they interact with lectins, non-enzymatic carbohydrate-binding proteins of non-immune origin that precipitate glycoproteins or polysaccharides and agglutinate cells [3,4]. The expression of carbohydrates and their corresponding lectins is differentially regulated both temporally and spatially in the developing CNS [1,2,5] and cell surface glycosylation is known to be altered during cell differentiation [6] and disease states such as cancer [7].

After CNS injury, a glial scar is formed at the injury site which creates an inhibitory environment for axonal regeneration and remyelination [8]. This scar contains several growth-inhibitory compounds including myelin-associated glycoprotein (MAG, siglec-4), Nogo, semaphorines and CSPGs, of which NG2, neurocan and versican are the major components [9,10]. CSPGs are a group of molecules consisting of a protein core heavily substituted with covalently attached GAGs. CSPGs are normally found in the ECM and on cell surfaces, playing roles in barrier formation and axonal guidance [8].

Potential therapies for repair and regeneration after CNS injury include the manipulation or removal of GAGs from the injury site [8,11] by treatment with ChABC [12]. Numerous studies have demonstrated that ChABC treatment promotes functional recovery and axon regeneration [11–13]. Perineuronal nets (PNNs) are composed of CSPGs and hyaluronic acid and encase neurons in the spinal cord. As ChABC digests GAGs, it has been hypothesised that digestion of the PNNs contributes to the recovery of plasticity, improving functional recovery after peripheral nerve injury [14]. In addition, dermatan sulphate disaccharide, one possible product of ChABC degradation of CS, promoted neurite outgrowth in immortalised rat pheochromocytoma PC12 cells and primary cultures of hippocampal neurons and promoted neuronal survival *ex vivo* and *in vivo* [15].

* Corresponding author. Fax: +353 91 494520.
E-mail address: siobhan.mcmahon@nuigalway.ie (S.S. McMahon).

0006-291X/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bbrc.2012.03.047

Given the importance of glycosylation in the CNS, there has been little attention to glycosylation changes of neurons in the injured environment or to the effect of ChABC treatment upon neuronal glycosylation. A multidisciplinary approach was taken to address these questions. Initially, *in silico* mining of a publicly available gene expression dataset was done to verify that glycosylation-related genes were differentially regulated after spinal cord injury and to identify potential targets for further investigation. An *in vitro* model with PC12 cells was then used to examine glycosylation expression in simulated normal, injured and ChABC-treated environments.

The PC12 cell line differentiates into neuron-like cells upon treatment with neurotrophins [16], and has been extensively used as a model for studying neuronal functions and responses including neurite outgrowth [17,18] and neuro-protective effects [19,20]. PC12 cells were co-cultured with various astrocytes [17,20] (normal primary astrocytes, Neu7 astrocytes and Neu7 astrocytes treated with ChABC) to model the different environments [21]. The Neu7 astrocytic cell line is an inhibitory cell line that has been engineered to overproduce NG2, versican and the CS-56 antigen [22] and is used to mimic the inhibitory environment which occurs following spinal cord injury [21–23]. The glycosylation expression profile of the PC12 cell surface was examined at intervals using lectin cytochemistry. Lectins bind specifically to distinct carbohydrate moieties, but may also contain one or more non-carbohydrate ligand sites [24]. Plant lectins have long been used as an analytical tool in tissue and cell histochemistry and particular carbohydrate motifs have been associated with otherwise indistinguishable cell types and stages of cell differentiation [4,6,25].

We present *in silico* mining and lectin profiling results of a healthy, injured and ChABC-treated model to profile cell surface glycosylation of neurons and attribute any glycosylation changes to the presence and removal of one class of glial scar inhibitory molecule, GAGs on CSPGs.

2. Materials and methods

2.1. Materials

Culture trays, transwells and cell culture plastics were from BD Falcon. Lectins were purchased from EY Labs (CA, USA). ProLong Gold antifade was from Invitrogen (Biosciences, Dublin, Ireland). PC12 cells were from ECACC (Salisbury, UK). All other reagents were from Sigma Aldrich Co. (Dublin, Ireland) unless otherwise indicated, and were of the highest grade available.

2.2. Cell cultures and environmental models

PC12 cells were cultured on poly-L-lysine (PLL; 10 μg/μL for 3 h) coated cover slips in 12-well trays in Dulbecco's modified eagle's medium (D-MEM, high glucose with L-glutamine) supplemented with 10% horse serum, 5% foetal bovine serum (FBS) and 1% penicillin and streptomycin (P/S) at 37 °C in a 5% humidified CO_2 atmosphere. The medium was supplemented with nerve growth factor (50 ng/mL) at intervals of three days for PC12 cell differentiation.

Primary cerebral astrocytes were obtained from P2 Sprague Dawley rat pups, purified and cultured as previously described [21]. Neu7 astrocytes were cultured in D-MEM supplemented with 10% horse serum, 1% L-glutamine and 1% P/S.

For 'normal' condition simulation, PC12 cells were co-cultured with primary astrocytes. PC12 cells were seeded at 5000 cells for 8 days *in vitro* (DIV), 10,000 cells for 4 DIV, and 50,000 cells for 1 DIV per PLL coated transwell in a 12-well tray. Astrocytes were seeded at the same density as above onto sterile cover slips in a 12-well tray. The PC12 cells were grown on the transwell to allow CSPGs secreted from the astrocytes grown in the same well to enter the media and interact with the PC12 cells, but not allow the two cell types to interact. For 'injured' condition, PC12 cells were grown with Neu7 cells. The 'treated' model consisted of PC12 cells grown with Neu7 cells with media treated with 0.1 unit/mL ChABC every 2 days, i.e. treated day 0 and every two days thereafter.

2.3. Lectin cytochemistry

Lectin cytochemistry at room temperature was performed on PC12 cells after growth at 1, 4 and 8 DIV as follows. Cells were fixed with 4% para-formaldehyde for 10 min and washed four times in 10 mM Tris–HCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$, pH 7.4 (TBS). Cells were blocked with 2% periodate-treated [26] bovine serum albumin (BSA) in TBS for 30 min. The cells were washed four times in TBS and then incubated with fluorescein isothiocyanate (FITC)-labelled lectins (Table 1) for 1 h in the dark at the following concentrations: SNA-I, MAA, PNA and ECA at 20 μg/mL and WFA at 10 μg/mL in TBS. Inhibitory controls were also carried out in parallel by preincubation of the lectins with 100 mM concentrations of the appropriate haptenic carbohydrates in TBS for 1 h prior to cell staining as follows: SNA-I, MAA and PNA were prepared in lactose, ECA in galactose (Gal) and WFA in *N*-acetylgalactosamine (GalNAc) and staining was carried out in the presence of the sugar. The cells were washed twice in TBS, counterstained with 1 μg/mL DAPI in TBS for 5 min, washed four times in TBS and mounted on glass slides with a drop of ProLong Gold antifade.

2.4. Image and statistical analysis

Cells were imaged on an Olympus IX81 fluorescent microscope using Perkin-Elmer Volocity® image acquisition software. Observed intensity of staining for PC12 cells were tabulated using a scale of no binding (−), slight binding (+), moderate binding (++), intense binding (+++) and very intense binding (++++).

2.5. In silico data mining

A gene expression dataset of the response of rat motor neurons 0, 2, 7, 21 and 60 days after spinal cord injury (GeneChip rat genome 230 3.0 array, Affymetrix Inc., Santa Clara, CA) was downloaded from the Gene Expression Omnibus (accession number GSE19701, www.ncbi.nlm.nih.gov/geo/) [27]. Gene expression

Table 1
Lectins and their corresponding carbohydrate binding specificity.

Abbreviation	Lectin	Origin	Binding specificity
ECA	*Erythrina cristigalli* agglutinin	*Erythrina cristagalli* (coral tree)	Terminal Gal-β-(1→4)-GlcNAc/Glc (LacNAc/Lac), GalNAc, Gal
PNA	Peanut agglutinin	*Arachis hypogaea* (peanut)	β-Gal, Gal-β-(1→3)-GalNAc, (T-antigen), Lac
WFA	*Wisteria floribunda* agglutinin	*Wisteria floribunda* (Japanese wisteria)	GalNAc, lactose, Gal, chondroitin sulphate
MAA	*Maackia amurensis* agglutinin	*Maackia amurensis*	Neu-α-(2→3)-Gal-β-(1→4)-Glc(NAc), Gal-3-SO_4
SNA-I	*Sambucus nigra* lectin-I	*Sambucus nigra* (elderberry)	Neu-α-(2→6)-Gal(NAc)

analyses were performed using GeneSpring 11.5 (Agilent Technologies, Cork, Ireland) and further detailed in supplementary data.

3. Results

3.1. In silico analysis

In silico mining was done on a study of gene expression response of rat motor neurons following spinal cord injury [27]. After statistical and gene expression analyses, 6790 out of 31,099 probes were identified as having a 1.2-fold change above or below control day 0. From these probes, glycosylation-related genes were extracted and genes associated with the sialic acid pathway, glycosaminoglycans and chondroitin sulphate proteoglycans, galactose and lectins, including galectins, were differentially regulated and their products have been reported to have altered expression after spinal cord injury [9,28–30] (see supplementary data). Neuronally expressed β-galactoside α-2,3-sialyltransferase (ST) genes were both up- and down-regulated after injury, with two of the initially up-regulated post-injury genes down-regulated again at 21 days post-injury. Only one α-2,6- and one α-2,8-ST were extracted and both were down-regulated at all timepoints post-injury (Fig. 1A). *N*-acetylgalactosaminyltransferases (GalNAcTs) and galactosyltransferases (GalTs) were both significantly up- and down-regulated post-injury, with one GalNAcT going from up- to down-regulated at 7 days post-injury and returning to up-regulated by 21 days. One GalT (a β-1,3GalT) went from initial down-regulation at day 2 to subsequent up-regulation at all later timepoints and conversely, the only significant α-1,3GalT (AI178222) went from initial up-regulation to down-regulation from 7 days onwards (Fig. 1B).

C-type lectin family genes were all up-regulated at all timepoints, confirming the presence of carbohydrate recognition domains on cell surface. Galectins are β-galactoside binding lectins [1,31]. The majority of extracted galectin genes were up-regulated at all timepoints following injury except for a soluble galectin-3 (NM_031832), which was down-regulated at day 7 post-injury but up-regulated at all other timepoints, and a galectin-related protein (NM_057187), which was down-regulated at all timepoints but up-regulated at day 7 post-injury (Fig. S1, supplementary).

3.2. In vitro profiling

The cell surface glycosylation changes of differentiated PC12 cells were analysed after 1, 4 and 8 DIV using fluorescently-labelled lectins selected to detect sialic acid-, galactose- and chondroitin sulphate-related motifs. The intensity of staining was recorded and tabulated (Table 2). All lectin cytochemistry was also carried out in parallel in the presence of haptenic sugars. A reduction in binding intensity was noted in these cases (not shown), demonstrating that lectin binding was carbohydrate-mediated [25].

Slight staining of PC12 cells with ECA was observed at 1 and 8 days in the 'healthy' primary astrocyte co-culture (Fig. 2A and G, respectively), but increased staining intensity was observed at

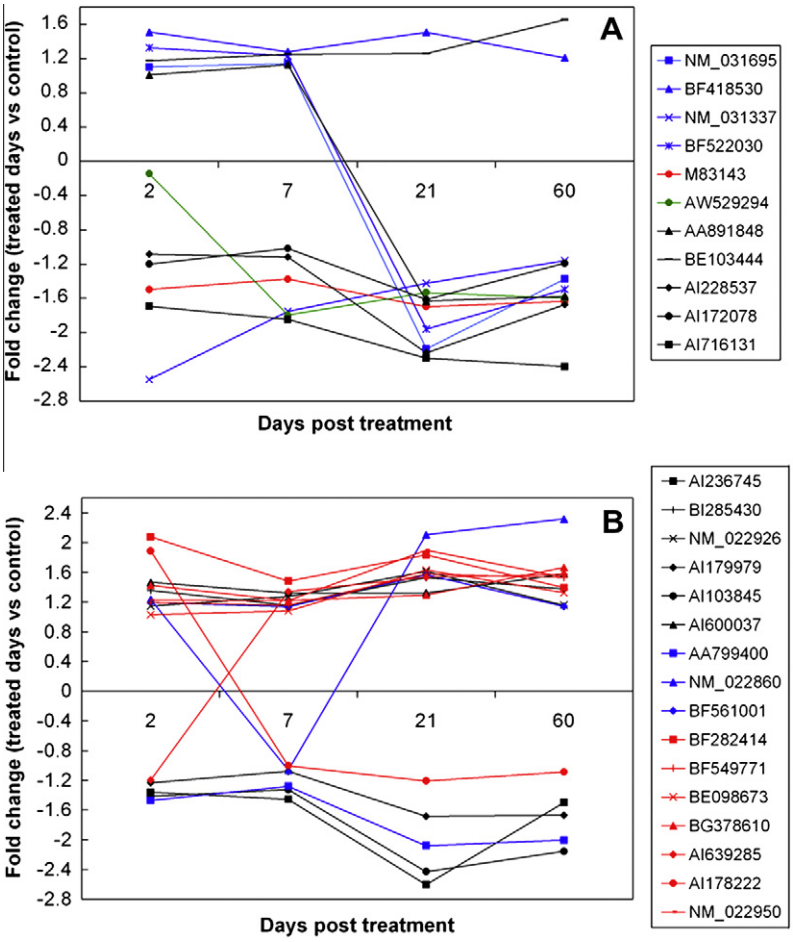


Fig. 1. Up- and down-regulated glycosylation-related genes compared to control day 0. (A) Sialic acid pathway related genes where the accession numbers of β-galactoside α-2,3-STs are blue, α-2,6-ST is red, α-2,8-ST is green and others which include sialic acid transporters are black. (B) Galactose-related genes where the accession numbers of polypeptide GalNAcTs are black, GalNAcTs are blue and GalTs are red. See supplementary data for full assignment of GenBank accession numbers (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Table 2
Intensity of lectin binding to PC12 cells. 'Normal' condition was simulated by PC12/primary astrocyte co-culture, 'injured' condition by PC12/Neu7 astrocyte co-culture and 'treated' condition by treatment of media of PC12/Neu7 astrocyte co-culture with ChABC.

Day	Lectin	Normal	Injured	Treated
1	ECA	+	++	++
4		+++	+	++
8		+	+++	+++
1	PNA	+	+	+
4		++++	+	+++
8		++	+++	++
1	WFA	++	±	+
4		+	+	++
8		+++	++	+++
1	MAA	+	+	+
4		+	++	+
8		++++	++	+++
1	SNA-I	–	–	–
4		–	–	–
8		–	++	–

– No binding; + slight binding; ++ moderate binding; +++ intense binding; ++++ very intense binding.

4 days (Fig. 2D and Table 2). Cells grown in the 'inhibitory' environment (Fig. 2B) showed more intense staining than cells growing in a normal astrocyte environment at 1 DIV, and maximum staining intensity was reached at 8 DIV (Fig. 2H), delayed compared to the 'healthy' environment. The ChABC-treated group showed no change in intensity of staining over time. At 1 DIV in the 'treated' environment, the staining intensity was moderate and highest intensity was observed on the cell surface, at 'caps' on the cells (Fig. 2C). At 4 DIV, the intensity remained moderate but was dispersed throughout the cell and was increased by 8 DIV, comparable to 4 DIV in the normal environment (Table 2 and Fig. 2F and I).

PNA staining of PC12 cells was observed under all conditions, at all timepoints, differing in intensity and staining dispersal and was similar to that of ECA for normal condition. Cells grown in 'normal' and 'treated' conditions stained slightly and very slightly, respectively, at 1 DIV. The intensity increased by day 4 and decreased again to moderate binding by day 8. Day 4 of 'normal' co-culture cells appeared to have maximal binding at the cell surface, especially where cells touched one another (Fig.S2 and Table 2). In contrast, cells grown in 'injured' conditions only reached maximal intensity by day 8.

In the 'normal' co-culture group, WFA bound with moderate intensity to PC12 cells at 1 DIV, with greatest intensity observed where cells touched one another. Binding intensity decreased to slight binding at day 4 and increased to very intense at 8 DIV, in common with both injured and treated environments (Fig. S3 and Table 2). The binding intensity variation of the injured and treated environments were similar, with very slight intensity

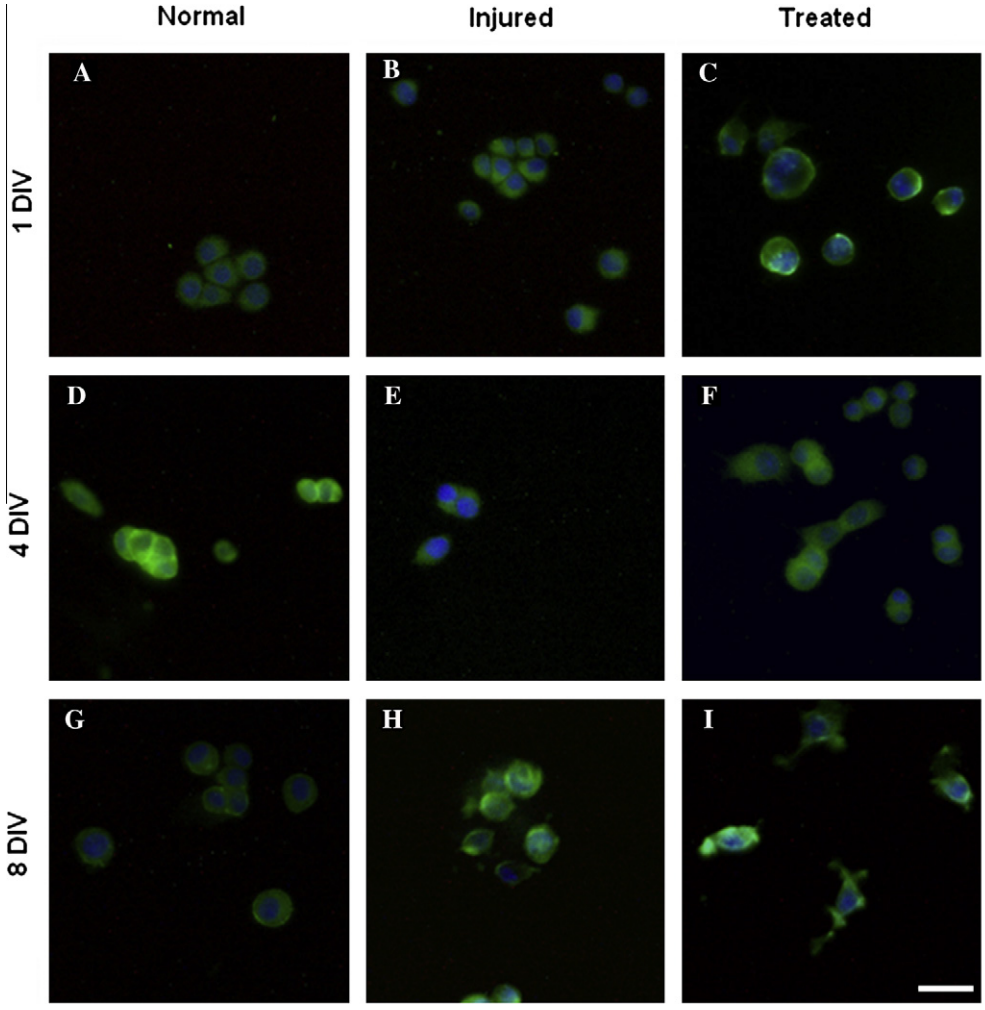


Fig. 2. Photomicrographs of ECA–FITC stained PC12 cells at 1 (A–C), 4 (D–F) and 8 DIV (G–I) co-cultured with primary astrocytes ('normal'), Neu7 cells ('injured') and Neu7 cells treated with ChABC ('treated'), respectively. Scale bar = 30 μm.

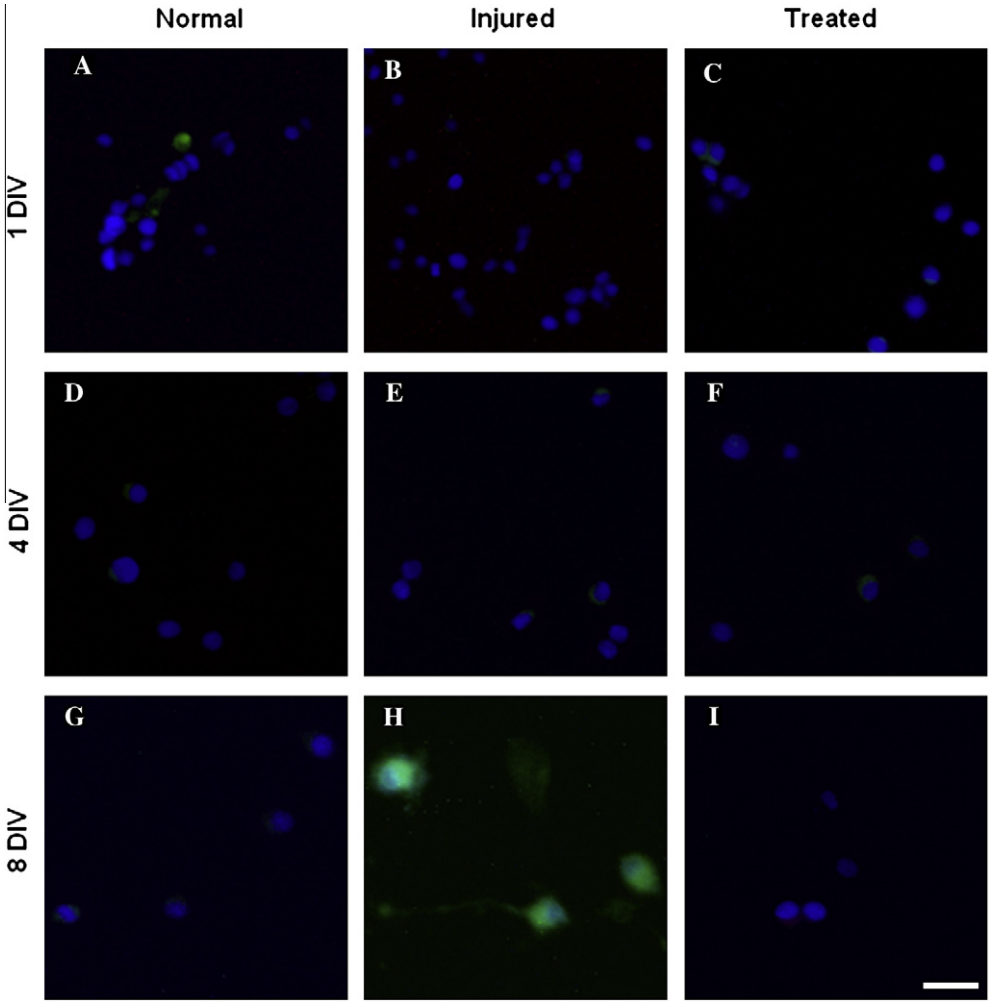


Fig. 3. Photomicrographs of SNA-I-FITC stained PC12 cells at 1 (A–C), 4 (D–F) and 8 DIV (G–I) co-cultured with primary astrocytes ('normal'), Neu7 cells ('injured') and Neu7 cells treated with ChABC ('treated'), respectively. Scale bar = 30 μm.

binding of WFA at 1 DIV, slightly increased binding at day 4 and greater intensity observed where cells touched one another in the 'injured' environment (Fig. S3E), and moderate to intense binding at day 8. The slight binding intensity observed at 4 DIV in the 'injured' co-culture environment was most apparent at the point where the cells touched or at the cell surface.

MAA stained PC12 cells under all conditions at each timepoint, and staining intensity increased over time reaching greatest intensity at 8 DIV for all environments (Fig. S4 and Table 2).

SNA-I did not stain the PC12 cells at any timepoint or condition (Fig. 3A–G and I), except at 8 DIV (Fig. 3H and Table 2) in the 'injured' co-culture group where cell bodies and neurites were moderately stained.

4. Discussion

The retrieved genes from the *in silico* analysis verified that the expression of glycosylation-related genes was altered at various timepoints post-injury and identified STs, GalNAcTs, GalTs and galectins as potential targets for differential expression in injury conditions. The *in vitro* model used to simulate normal, injured and ChABC-treated environments suggested that neuronal glycosylation changes occurred in these conditions.

The lectin ECA binds with greatest affinity to unsialylated terminal *N*-acetyllactosamine (LacNAc) structures, which are a major component of glycoprotein *N*-linked oligosaccharides and glycolipids (Table 1). Zhang, et al. [28] found that ECA binding increased after brain injury in mice, correlating to our observations at 8 DIV in the 'injured' environment. ECA staining where the cells touch may indicate a role for this carbohydrate motif in cell–cell adhesion. Galectins are involved in cell–cell interactions, adhesion, differentiation, apoptosis and axonal guidance [1,31], and bind to motifs elucidated by ECA. The majority of galectins with altered expression post-injury were up-regulated at all timepoints. Expression of the potential galectin receptor evidenced by ECA binding seemed to be time-dependent, and reached a maximum in the 'healthy' culture at day 4, which may in turn reflect temporal expression of galectins in the CNS. However, expression of this receptor appeared to be delayed in the injured environment as maximum expression was at day 8, in common with the treated environment.

In common with ECA, PNA also stained at areas of cell–cell contact and staining intensity was temporally regulated, except for the treated environment which was more similar to the healthy condition rather than the injured (Table 2). In the *in silico* analysis, all significant β-GalTs were up-regulated by day 7 post-injury, which may correlate with the intense PNA and ECA staining of the PC12 cells at 8 DIV.

WFA lectin is commonly used as a marker for PNNs [11,32]. In all three culture conditions maximal WFA intensity was noted at the latest timepoint, and was most intense where the cells

touched, which may indicate a role for the elucidated motif in adhesion. However, by day 7 post-injury, only one β1,4GalNAcT (BF561001) was up-regulated in the *in silico* analysis but the majority of CSPG-related significant genes were up-regulated at all timepoints post-injury (supplementary data).

Sialic acids are the mainly terminal residues of complex *N*- and *O*-linked oligosaccharides of glycoproteins and glycolipids and comprise polysialic acid in an α-(2,8)-linkage on neural cell adhesion molecule (NCAM). Sialylated motifs on the cell surfaces of vertebrates are involved in cell–cell communication, development and adhesion [33], interacting with siglecs and galectins [31]. The lectins MAA and SNA-I have binding specificity for α-(2,3)- and α-(2,6)-linked sialic acid, respectively, and MAA is also known to bind to Gal-3-SO_4 [34] (Table 1). In the brain, the occurrence of α-(2,3)-linked sialic acid is predominant with little to no α-(2,6)-linkage expected [2]. SNA-I binding was observed in the 'inhibitory' environment at 8 DIV, but was not seen in the ChABC-treated condition, where cell glycosylation was comparable to the 'normal' environment. The altered expression of significant α-2,6- and α-2,3-STs post-injury did not correlate with staining intensities, and interestingly, expression of the α-2,8-ST, relevant to PSA synthesis, was down-regulated at all timepoints. Electrical signalling in neurons, skeletal muscle cells and cardiomycetes is modulated by the sialic acid content of particular isoforms of ion channels [35]. Altered or aberrant sialic acid expression could impact neuron polarisation [35], which may be consistent with altered excitability of neurons post-injury [33]. Exposure of α-(2,6)-linked sialic acid and binding to SNA-I has been observed on apoptotic and necrotic cells [36], and α-(2,6)-sialylation has been identified as blocking binding to galectins, hence functioning as a biological 'off switch' [31]. Interestingly, the expression of α-2,6-ST was down-regulated at all timepoints post-injury while galectin-1 expression, which has been associated with pathogenesis in the injured spinal cord [30], was up-regulated.

ChABC treatment promotes functional recovery and reduces the inhibitory effect of CSPGs but anatomical regeneration post-treatment is limited [28]. This may be due in part to the potentially immunogenic carbohydrate 'stub' structures created by ChABC action [13]. The NG2 protein core has also been suggested to have an inhibitory effect on axonal growth [9]. However, in the model system, treatment with ChABC appeared to be successful in returning the neuronal glycosylation to normal conditions at all timepoints in the case of MAA, PNA and SNA-I staining, and by day 8 in the case of WFA.

A multidisciplinary approach allowed the *in silico* verification of differentially regulated glycosylation-related genes in neurons and targeting of altered carbohydrate motifs. The model system demonstrated neuronal cell surface glycosylation changes in an inhibitory environment and may be the first indication of the occurrence of abnormal sialylation in an injured environment. The limited number of cell types in a controlled environment can help attribute any changes to specific molecules and their degradation products alone, and help fit this information into a wider injury picture. In addition, this model indicated a return to normal neuronal glycosylation after treatment with ChABC which may be promising for identifying points of intervention or potential therapies for neuronal regeneration strategies.

Acknowledgments

The authors thank Dr. J.Q. Gerlach (NUI Galway) for helpful discussions, and Dr. J. Rogers and Prof. J. Fawcett (University of Cambridge) for their kind gift of the Neu7 astrocyte cell line. This work was supported by the Health Research Board of Ireland (SMcM), and grants 08/SRC/B1393 from Science Foundation Ireland for Alimentary Glycoscience Research Cluster (SS and LJ) and 260600 from the EU FP7 program for GlycoHIT (MK and LJ).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bbrc.2012.03.047.

References

[1] T.M. Jessell, M.A. Hynes, J. Dodd, Carbohydrates and carbohydrate-binding proteins in the nervous system, Annu. Rev. Neurosci. 13 (1990) 227–255.
[2] R. Kleene, M. Schachner, Glycans and neural cell interactions, Nat. Rev. Neurosci. 5 (2004) 195–208.
[3] I.J. Goldstein, R.C. Hughes, M. Monsigny, T. Osawa, N. Sharon, What should be called a lectin? Nature 285 (1980) 66.
[4] N. Sharon, H. Lis, Lectins, Kluwer Academic Publishers, Dordrecht, 2003.
[5] J. Tenne-Brown, A.C. Puche, B. Key, Expression of galectin-1 in the mouse olfactory system, Int. J. Dev. Biol. 42 (1998) 791–799.
[6] M.C. Dodla, A. Young, A. Venable, K. Hasneen, R.R. Rao, D.W. Machacek, S.L. Stice, Differing lectin binding profiles among human embryonic stem cells and derivatives aid in the isolation of neural progenitor cells, PLoS One 6 (2011) e23266.
[7] S.D. Szajda, A. Jankowska, K. Zwierz, Carbohydrate markers in colon carcinoma, Dis. Markers 25 (2008) 233–242.
[8] S.A. Busch, J. Silver, The role of extracellular matrix in CNS regeneration, Curr. Opin. Neurobiol. 17 (2007) 120–127.
[9] A.M. Tan, W. Zhang, J.M. Levine, NG2: a component of the glial scar that inhibits axon growth, J. Anat. 207 (2005) 717–725.
[10] R.A. Asher, D.A. Morgenstern, M.C. Shearer, K.H. Adcock, P. Pesheva, J.W. Fawcett, Versican is upregulated in CNS injury and is a product of oligodendrocyte lineage cells, J. Neurosci. 22 (2002) 2225–2236.
[11] J.M. Massey, C.H. Hubscher, M.R. Wagoner, J.A. Decker, J. Amps, J. Silver, S.M. Onifer, Chondroitinase ABC digestion of the perineuronal net promotes functional collateral sprouting in the cuneate nucleus after cervical spinal cord injury, J. Neurosci. 26 (2006) 4406–4414.
[12] E.J. Bradbury, L.D.F. Moon, R.J. Popat, V.R. King, G.S. Bennett, P.N. Patel, J.W. Fawcett, S.B. McMahon, Chondroitinase ABC promotes functional recovery after spinal cord injury, Nature 416 (2002) 636–640.
[13] D. Crespo, R.A. Asher, R. Lin, K.E. Rhodes, J.W. Fawcett, How does chondroitinase promote functional recovery in the damaged CNS? Exp. Neurol. 206 (2007) 159–171.
[14] C.M. Galtrey, R.A. Asher, F. Nothias, J.W. Fawcett, Promoting plasticity in the spinal cord with chondroitinase improves functional recovery after peripheral nerve repair, Brain 130 (2007) 926–939.
[15] A. Rolls, H. Avidan, L. Cahalon, H. Schori, S. Bakalash, V. Litvak, S. Lev, O. Lider, M. Schwartz, A disaccharide derived from chondroitin sulphate proteoglycan promotes central nervous system repair in rats and mice, Eur. J. Neurosci. 20 (2004) 1973–1983.
[16] L.A. Greene, A.S. Tischler, Establishment of a noradrenergic clonal line of rat adrenal pheochromocytoma cells which respond to nerve growth factor, Proc. Natl. Acad. Sci. U. S. A. 73 (1976) 2424–2428.
[17] J.L. Anderl, S. Redpath, A.J. Ball, A neuronal and astrocyte co-culture assay for high content analysis of neurotoxicity, J. Vis. Exp. 27 (2009).
[18] M.R. Andrews, S. Czvitkovich, E. Dassie, C.F. Vogelaar, A. Faissner, B. Blits, F.H. Gage, C. Ffrench-Constant, J.W. Fawcett, A9 integrin promotes neurite outgrowth on tenascin-C and enhances sensory axon regeneration, J. Neurosci. 29 (2009) 5546–5557.
[19] S. Yu, M. Liu, X. Gu, F. Ding, Neuroprotective effects of salidroside in the PC12 cell model exposed to hypoglycemia and serum limitation, Cell. Mol. Neurobiol. 28 (2008) 1067–1078.
[20] R.B. Tjalkens, X. Liu, B. Mohl, T. Wright, J.A. Moreno, D.L. Carbone, S. Safe, The peroxisome proliferator-activated receptor-γ agonist 1,1-bis(3′-indolyl)-1-(p-trifluoromethylphenyl)methane suppresses manganese-induced production of nitric oxide in astrocytes and inhibits apoptosis in cocultured PC12 cells, J. Neurosci. Res. 86 (2008) 618–629.
[21] E.M. Donnelly, P.M. Strappe, L.M. McGinley, N.N. Madigan, E. Geurts, G.E. Rooney, A.J. Windebank, J. Fraher, P. Dockery, T. O'Brien, S.S. McMahon, Lentiviral vector-mediated knockdown of the neuroglycan 2 proteoglycan or expression of neurotrophin-3 promotes neurite outgrowth in a cell culture model of the glial scar, J. Gene Med. 12 (2010) 863–872.
[22] P.S. Fidler, K. Schuette, R.A. Asher, A. Dobbertin, S.R. Thornton, Y. Calle-Patino, E. Muir, J.M. Levine, H.M. Geller, J.H. Rogers, A. Faissner, J.W. Fawcett, Comparing astrocytic cell lines that are inhibitory or permissive for axon growth: the major axon-inhibitory proteoglycan is NG2, J. Neurosci. 19 (1999) 8778–8788.
[23] W.-L. Kuan, C.B. Hurelbrink, R.A. Barker, Increased capacity for axonal outgrowth using xenogenic tissue in vitro and in a rodent model of Parkinson's disease, Xenotransplantation 13 (2006) 233–247.
[24] S.H. Barondes, Bifunctional properties of lectins: lectins redefined, Trends Biochem. Sci. 13 (1988) 480–482.
[25] J.Q. Gerlach, M. Kilcoyne, S. Eaton, V. Bhavanandan, L. Joshi, Non-carbohydrate-mediated interaction of lectins with plant proteins, in: A.M. Wu (Ed.), The

Molecular Immunology of Complex Carbohydrates-3, Springer, New York, pp. 257–269.
[26] W.F. Glass, R.C. Briggs, L.S. Hnilica, Use of lectins for detection of electrophoretically separated glycoproteins transferred onto nitrocellulose sheets, Anal. Biochem. 115 (1981) 219–224.
[27] J. Ryge, O. Winther, J. Wienecke, A. Sandelin, A.-C. Westerdahl, H. Hultborn, O. Kiehn, Transcriptional regulation of gene expression clusters in motor neurons following spinal cord injury, BMC Genomics 11 (2010) 365.
[28] H. Zhang, T. Muramatsu, A. Murase, S. Yuasa, K. Uchimura, K. Kadomatsu, *N*-Acetylglucosamine 6-*O*-sulfotransferase-1 is required for brain keratan sulfate biosynthesis and glial scar formation after brain injury, Glycobiology 16 (2006) 702–710.
[29] Y. Zhang, M. Ghadiri-Sani, X. Zhang, P.M. Richardson, J. Yeh, X. Bo, Induced expression of polysialic acid in the spinal cord promotes regeneration of sensory axons, Mol. Cell. Neurosci. 35 (2007) 109–119.
[30] D. Kurihara, M. Ueno, T. Tanaka, T. Yamashita, Expression of galectin-1 in immune cells and glial cells after spinal cord injury, Neurosci. Res. 66 (2010) 265–270.
[31] Y. Zhuo, S.L. Bellis, Emerging role of α2,6-sialic acid as a negative regulator of galectin binding and function, J. Biol. Chem. 286 (2011) 5935–5941.
[32] J. Ajmo, A. Eakin, M. Hamel, P. Gottschall, Discordant localization of WFA reactivity and brevican/ADAMTS-derived fragment in rodent brain, BMC Neurosci. 9 (2008) 14.
[33] E. Repnikova, K. Koles, M. Nakamura, J. Pitts, H. Li, A. Ambavane, M.J. Zoran, V.M. Panin, Sialyltransferase regulates nervous system function in *Drosophila*, J. Neurosci. 30 (2010) 6466–6476.
[34] X. Bai, J.R. Brown, A. Varki, J.D. Esko, Enhanced 3-*O*-sulfation of galactose in Asn-linked glycans and *Maackia amurenesis* lectin binding in a new Chinese hamster ovary cell line, Glycobiology 11 (2001) 621–632.
[35] T.A. Schwetz, S.A. Norring, E.S. Bennett, *N*-glycans modulate Kv1.5 gating but have no effect on Kv1.4 gating, Biochim. Biophys. Acta 1798 (2010) 367–375.
[36] N. Malagolini, M. Chiricolo, M. Marini, F. Dall'Olio, Exposure of α2,6-sialylated lactosaminic chains marks apoptotic and necrotic death in different cell types, Glycobiology 19 (2009) 172–181.