Characterisation of Novel Glycoprotein and Glycopolymer Ligands of Human C-type Lectins

by

Dipl.-Biochem. Anne Sophie Gleinich

A thesis submitted in the partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Medical Sciences

University of Warwick, Warwick Medical School

January 2019

ABSTRACT

Transmembrane proteins of the C-type lectin family are major carbohydrate-binding components of the immune system that are present in abundance on the surfaces of cells, such as macrophages and dendritic cells, and function as receptors. Previous screening data using human placental extracts suggested that the glycoprotein pregnancy-associated plasma protein-A1 (PAPP-A1) may be a natural ligand for the C-type lectin DC-SIGN. Interaction data using state of the art dip-and-read real-time biosensor technology show that PAPP-A1 interacts with DC-SIGN and DC-SIGNR with dissociation constants consistent with the physiologically relevant, low nanomolar range. PAPP-A1 is a large glycoprotein complex and a metalloproteinase that influences insulin-like growth factor (IGF) bioavailability by cleaving IGF-binding proteins (IGFBP) 4 and 5. PAPP-A1 is found in abundance in plasma during pregnancy in complex with its natural inhibitor, eosinophil major basic protein (proMBP). Moreover, studies have shown that in males and non-pregnant females, PAPP-A1 circulates at much lower levels and is positively associated with acute coronary syndromes. Notably, PAPP-A1 possesses a high density of glycosylation sites and subsequent glycoproteomic analysis of native PAPP-A1/proMBP, and also recombinant PAPP-A1 and proMBP, was carried out via nano-liquid chromatography mass spectrometry. This glycoproteomic study successfully characterised the N-linked oligosaccharide profile of these molecules, including N-glycan site assignment. Additional interaction analysis also showed that cutting-edge star-shaped glycopolymers bind to DC-SIGN with picomolar affinity. Overall, the work illustrates the importance of glycoproteins and glycopolymers as potential modifiers of the immune system and also as factors associated with understanding the mechanisms of vascular diseases such as pre-eclampsia and acute coronary syndromes. There is also emphasis on the power of new, combined biophysical technologies in uncovering novel glycomic networks.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am greatly indebted to my supervisors Dr Daniel Mitchell and Dr Harpal Randeva for being very supportive, abundantly helpful and offering invaluable assistance and excellent guidance throughout these four years.
Thank you!

I wish to acknowledge *The General Charities of the City of Coventry* for their generous and continuous support of my PhD studentship through an MRA PhD Fellowship.

Furthermore, I would like to deeply thank Professor Hendrik Lehnert, University of Lübeck, for additional project support.

I would also like to thank my viva examiners

Dr Robert B. Sim and

Dr Nicholas R. Waterfield

for letting my defence be an enjoyable moment, and for their very helpful comments and suggestions.

DECLARATION

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

Parts of this thesis were presented in the Upgrading presentation and report for this degree in October 2015.

The work presented (including data generated and data analysis) was carried out by the author except in the case of the laboratory work presented in chapter '4.5 Human native PAPP-A1/proMBP-complex'. However, part of the data analysis and the entire data interpretation was carried out by the author.

Parts of Chapter 5 of this thesis have been co-published by the author:

MITCHELL, D. A., ZHANG, Q., VOORHAAR, L., HADDLETON, D. M., HERATH, S., **GLEINICH, A. S.**, RANDEVA, H. S., CRISPIN, M., LEHNERT, H., WALLIS, R., PATTERSON, S. & BECER, C. R. 2017. Manipulation of cytokine secretion in human dendritic cells using glycopolymers with picomolar affinity for DC-SIGN. *Chemical Science*, 8, 6974-6980.

COPYRIGHTS AND PERMISSIONS

Figure 1.1 is reprinted with kind permission from Cold Spring Harbor Laboratory Press for figure 34.3 in Essentials of Glycobiology 3rd edition. Varki A, Cummings RD, Esko JD, et al. Cold Spring Harbor (NY) © 2015-2017 by The Consortium of Glycobiology Editors, La Jolla, California.

Figure 1.2 is reprinted with kind permission from Springer Nature through the Copyright Clearance Center, Rightslink® for FEINBERG, H., MITCHELL, D. A., DRICKAMER, K. & WEIS, W. I. 2001. Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR. Science, 294, 2163-6.

Figure 1.3 is reprinted with kind permission from Springer Nature through the Copyright Clearance Center, Rightslink® for VAN KOOYK, Y. & GEIJTENBEEK, T. B. H. 2003. DC-SIGN: Escape mechanism for pathogens. Nature Reviews Immunology, 3, 697-709.

Figure 1.4 is reprinted with kind permission from Springer Nature through the Copyright Clearance Center, Rightslink® for BROWN, G. D., WILLMENT, J. A. & WHITEHEAD, L. 2018. C-type lectins in immunity and homeostasis. Nat Rev Immunol, 18, 374-389.

Figure 1.5 is reprinted with kind permission from John Wiley and Sons through the Copyright Clearance Center, Rightslink® for MARTINEZ-POMARES, L. 2012. The mannose receptor. J Leukoc Biol, 92, 1177-86.

ABBREVIATIONS

$CaCl_2$	Calcium chloride
CD	Cluster of differentiation
CRD	Carbohydrate recognition domain
CLR	C-type lectin receptor (transmembrane protein of the C-type lectin family that functions as a receptor)
CTL	C-type lectin
CTLD	C-type lectin-like domain
DC	Dendritic cell
DC-SIGN	Dendritic cell-specific intracellular adhesion molecule [ICAM-3]-grabbing non-integrin 1
DC-SIGNR	Dendritic cell-specific intracellular adhesion molecule [ICAM-3]-grabbing non-integrin 1-Related (protein)
ddH2O	Double distilled water
EDTA	(Ethylenediamine)tetraacetic acid
Fuc	fucose
GalNAc	N-acetylgalactosamine
GlcNAc	N-acetylglucosamine
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
Hex	Hexose
HexNAc	N-acetylhexosamine
IGF	Insulin-like growth factor
IGFBP	Insulin-like growth factor-binding protein
k_{on}	Association rate
k_{off}	Dissociation rate
K_D	Dissociation constant

nLC-MS	Nano-liquid chromatography mass spectrometry
nLC-NSI-MS/MS	Nano-liquid chromatography Nano-spray ionisation Tandem mass spectrometry
$Man_9GlcNAc_2$	N-linked oligosaccharide of 9 mannose and 2 core GlcNAc residues
MMR	Macrophage mannose receptor
MS2 or MS^2	spectrum obtained in a tandem mass spectrometry experiment
MS/MS	Tandem mass spectrometry (experiment)
M_W	molecular weight
NaCl	Sodium chloride
NeuAc/ Neu5Ac	N-acetylneuraminic acid
NeuGc / Neu5Gc	N-glycolylneuraminic acid
N-glycan	N-linked glycan
nM	nanomolar
NSI	Nano-electrospray ionisation
PAPP-A/PAPP-A1	Pappalysin-1; Pregnancy-associated plasma protein A1
pM	picomolar
proMBP	Proform of the eosinophil major basic protein
PAGE	polyacrylamide gel electrophoresis
rpm	revolutions per minute
RU	response unit(s)
SDS	Sodium dodecyl sulphate
TRIS	2-Amino-2-(hydroxymethyl)-1,3-propanediol
vWF	von Willebrand Factor

1 THEORETICAL BACKGROUND

1.1 Glycobiology

Thirty years ago, the term *glycobiology* was coined by biochemist Raymond Dwek with the intention to reconnect the findings in the field of carbohydrate chemistry with basic biology as well as biochemistry (Blow, 2009).

Glycobiology aims to elucidate the structures and functional roles of covalent carbohydrate attachments to biomacromolecules within their biological context. The glycan modification of biomolecules such as proteins and lipids is referred to as glycosylation. The International Union of Pure and Applied Chemistry (IUPAC) defines glycans "as synonymous with polysaccharides" and, therefore, glycans are sequentially linked sugar moieties via glycosidic bonds (IUPAC). Glycans can be found unconjugated, but are usually enzymatically and site-specifically attached to a range of biomacromolecules, such as proteins, lipids and other organic molecules. Such molecules modified with glycoconjugates are then referred to as glycoproteins, proteoglycans or glycolipids. A very broad diversity of glycans exists, ranging from simple to complex structures, and they are present in all domains of life (Cummings and Pierce, 2014). Their polydisperse and microheterogenerous nature is intrinsic and derives from the lack of a single template mechanism, as it is established for protein and nucleic acid synthesis (Roseman, 2001).

Protein modification by glycosylation can occur as a co- as well as a post-translational event, depending on the nature of the protein and the cellular localisation of the enzymes involved. Glycosylation of proteins is mainly classified by the type of linkage to the polypeptide backbone: The major classes are N-linked and O-linked glycosylation (Taylor and Drickamer, 2011). It is has been found that nine amino acids are possible donors for modification via glycosylation, i.e. arginine (Arg/R), asparagine (Asn/N), serine (Ser/S), threonine (Thr/T), tyrosine (Tyr/Y), tryptophan (Trp/W), cysteine (Cys/C), hydroxylysine (Hyl) and hydroxyproline (Hyp) (Spiro, 2002, Stepper et al., 2011).

Cummings defines the *glycome* "as the repertoire of glycans expressed in a cell or organism" (Cummings and Pierce, 2014). Focusing on the mammalian glycome, the fundamental building blocks are glucose (Glc), N-acetylglucosamine (GlcNAc), galactose (Gal), N-acetylgalactosamine (GalNAc), mannose (Man), fucose (Fuc) and sialic acid (NeuAc). Additional carbohydrate moieties and derivatives such as xylose

and glucuronic acid can also be found in specific types of O-glycans and proteoglycan chains. The mammalian genome encodes for numerous enzymes, transporters and other accessory molecules that allow for the variety of glycan structures comprised within the glycome (Henrissat et al., 2015).

The approach of studying the relationship between structure and function of glycans has been termed as *glycomics* and even more specifically as *glycoproteomics* regarding protein glycosylation patterns and glycan-protein interactions (Raman et al., 2006).

Functional glycomics emphasises on the research of specificity and affinity between glycan binding proteins (GBP), such as lectins, and their glycan-conjugated ligands. It aims to achieve the characterisation of these interactions as well as illuminating their roles within the biological context, i.e. in cells and organisms, and targets the detailed function of each interaction partner, the GBP and the glycoconjugate respectively (Zaia, 2011).

To date, several functional roles have been proposed for glycans: They form extracellular matrices to provide tissue architecture and platforms for the regulation of cell and tissue interactions, and they serve as a marker for cellular identification. Glycans are involved in cellular processes including - but not limited to - the immune response, intracellular targeting and trafficking, intercellular recognition, as well as protein folding and stability (Varki, 1993). Particularly in the human body, studies have shown that glycans serve for communication purposes of self-tolerance, inflammation, blood clotting, and recognition of pathogens. Furthermore, it has been shown that metastasising cancer cells use glycans to evade the immune response (Alper, 2001). In summary, the structure, function, biosynthesis, variation and transformation of glycosylated molecules, cell organelles or cells are studied within the vast field of glycobiology (Alper, 2001, Taylor and Drickamer, 2011, Cummings and Pierce, 2014).

A comprehensive understanding of glycan structures is crucial for the insight into the functions that glycans serve. This will provide invaluable insight into the role of glycosylation in physiological and pathophysiological biologic events, since "one of the major goals of glycoscience is to develop therapeutics based on glycans" (Zaia, 2011).

This thesis emphasises the novel characterisation of human glycoprotein and new glycopolymer ligands that interact with human C-type lectins. The focus lies on the

human and other mammalian systems, which will be introduced and discussed in greater detail than plant, fungal, and microbial systems.

1.1.1 Glycosylation of proteins

The glycosylation of proteins can be found ubiquitously, in prokaryotes as wells as eukaryotes, inside the cell but also on the cellular surface, extracellular matrices and secreted glycoproteins (Roseman, 2001). In 1999, Apweiler concluded "that more than half of all proteins in nature will eventually be found to be glycoproteins" by a computerised analysis of a renowned protein sequence data bank (Apweiler et al., 1999).

The glycosylation of a protein extends not only its structure but also function: Glycan structures provide additional information on top of the already complex information content of the proteome. The glycan structures are generated enzymatically via *glycosyltransferases*, which create and extend oligosaccharide and polysaccharide chains by the formation of specific glycosidic linkages, and *glycosidases*, which trim and shape structures through the hydrolysis of specific glycosidic linkages (Bourne and Henrissat, 2001, Lairson et al., 2008).

Glycosylation can contribute to the protein folding itself and serves as a tool for surveillance of quality control in the rough endoplasmic reticulum. Moreover, the glycans stabilise polypeptides by maintaining the protein conformation and enhancing solubility of glycoproteins as well as marking them for transfer and selective targeting through the secretory pathway (Varki, 1993, Xu and Ng, 2015). From a more holistic point of view, the glycans of glycoproteins regulate cytosolic and nuclear functions, mediate and modulate immune responses including inflammatory reactions, autoimmunity and self-tolerance. Glycans mediate interactions with pathogens, including viruses, bacteria, fungi and helminths, and are involved in the development of cancer and tumour metastasis. Glycan structures are also involved in hormone action and homeostasis (Helenius and Aebi, 2004, Moremen et al., 2012).

1.1.1.1 N-linked glycosylation

Asparagine-linked glycans, referred to as N-linked glycans based on the One Letter Code for amino acids, have been comprehensively studied regarding their structural and functional aspects (Cummings, 2014). The consensus sequence N-X-S/T that guides N-linked glycosylation at the asparagine (N) in the polypeptide backbone, whereby X must not be proline (Spiro, 1970, Marshall, 1972). However, it has also been described that the N-X-S/T sequon is not strictly occupied by glycans and thus many N-X-S/T motifs in the human proteome display the unmodified amide side chain of asparagine (Petrescu et al., 2004).

The biosynthesis pathway for N-glycans is a co-translational as well as post-translational process, which mainly takes place at the endoplasmic reticulum (ER) and in some cases has been shown to start as early as a dozen amino acids after the ribosome at the nascent polypeptide (Hubbard and Ivatt, 1981). The biosynthesis of N-linked glycans can be roughly divided into three stages: First, the synthesis of the precursor polysaccharide that is linked to a lipid, namely a $Glc_3Man_9GlcNAc_2$ structure linked to dolichol. Secondly, the precursor $Glc_3Man_9GlcNAc_2$ is transferred from the lipid-linker to an asparagine in a sequon via the membrane-associated oligosaccharyl transferase (OST) protein complex in the ER lumen. Thirdly, the glycan is processed further, which can include trimming by glycosidases, extension through glycosyltransferases and potential modification by various enzymes. The processing of the glycan occurs from the ER throughout the Golgi complex (Helenius and Aebi, 2004, Weerapana and Imperiali, 2006, Taylor and Drickamer, 2011). Resultant to the common biosynthesis pathway of N-linked glycans, the protein-glycan linkage is always initiated at the reducing end by N-acetylglucosamine (GlcNAc) and all N-glycans contain a $(Man_3)GlcNAc_2$ pentasaccharide core structure (Kornfeld and Kornfeld, 1985).

Furthermore, the different stages of maturation of N-glycans serve specific purposes: The initial precursor $Glc_3Man_9GlcNAc_2$ is partially trimmed down to $Glc_1Man_9GlcNAc_2$, which interacts transiently and selectively with the calreticulin-calnexin chaperone complex. The trimming and subsequent interaction serves as a surveillance tool for the folding status of the protein and indicates quality control. Removal of all the terminal glucose residues down to $Man_9GlcNAc_2$, signals the

successful folding of the glycoprotein and initiates the transition into the Golgi complex for further processing of the N-linked glycans, which reflects the second stage of glycan maturation (Helenius and Aebi, 2004, Weerapana and Imperiali, 2006, Xu and Ng, 2015).

After modification of the glycans in the Golgi complex and *trans*-Golgi network, the glycoprotein is trafficked to its target site, which can be located in the cell, on the cell surface or in the extracellular space. N-glycan structures are typically classified into high-mannose, hybrid or complex types. The glycans belonging to the high-mannose class undergo processing by mannosidase enzymes (or, rarely, none at all) and usually contain between five and nine mannose residues. Additional processing - including addition of an N-acetylglucosamine residue or extension with galactose, N-acetylgalactosamine, fucose and sialic acid on one of the antennae - leads to the formation of a glycan that is classified as hybrid glycan. If hybrid glycans are processed even further, they can form multi-antennary, complex glycans that are extended with N-acetylglucosamine, galactose and sialic acid units. However, the classification of glycans is rather a guidance, due to the numerous possibilities of branching, extensions with various glycan moieties and anomeric linkages (Cummings, 2009, Taylor and Drickamer, 2011).

The mature glycans on the glycoprotein serve their individual roles, such as cell-cell and receptor-ligand interactions, triggers for immune responses and apoptosis, and underlying cause of the pathogenesis of various diseases (Helenius and Aebi, 2004, Moremen et al., 2012).

1.1.1.2 O-linked glycosylation

All types of O-linked glycosylation share the characteristic of being attached to a bio-macromolecule by an O-glycosidic linkage between the glycan and the conjugate. In the case of glycoproteins, this is often realised through formation of a covalent bond with the hydroxyl group present on serine (Ser), threonine (Thr), hydroxylysine (Hyl) or tyrosine (Tyr) (Moremen et al., 2012).

Apart from that, O-linked glycans are not attached to a specific consensus sequence at the polypeptide backbone, which makes the prediction of O-glycosylated sites more complex than the prediction of N-glycosylation. However, attempts have been

made that include more complex algorithms than solely the amino acid sequence of a given glycoprotein (Julenius et al., 2005, Steentoft et al., 2013).

O-glycosylation mainly takes place as a fully post-translational modification in the Golgi complex. O-linked glycans are extended stepwise and modifications are driven by specific enzymes. Several core structures have been elucidated so far: α-linked O-fucose, β-linked O-xylose, α-linked O-mannose, β-linked *N*-acetylglucosamine, α-linked *N*-acetylgalactosamine, α- or β-linked O-galactose, and α- or β-linked O-glucose glycans. However, the predominant type of O-linked glycans found in mammalian glycoproteins belongs to the mucin type, which is defined as an α-linked *N*-acetylgalactosamine (GalNAc) residue linked to the protein backbone (Van den Steen et al., 1998). The name originates from the high level of α-linked O-GalNAc glycosylation on mucins, which can be found on epithelial surfaces where they mainly bind water to provide a gel-like matrix (Julenius et al., 2005, Taylor and Drickamer, 2011).

Another major class of O-glycosylated proteins are proteoglycans. Proteoglycans serve mainly structural roles through their attached glycosaminoglycans (GAGs). The GAG core structure consists of an O-linked tetrasaccharide and is extended by a repetitive sequence of alternating linear disaccharide units. Further modification via sulfation and epimerisation leads to an extensive variety of possible GAG structures. GAG glycosylations can also be implemented in the transmembrane with or without a glycosylphosphatidiylinositol (GPI)-anchored lipid. Well-known proteoglycans are Heparin, Heparan sulfate, Chondroitin sulfate and Dermatan sulfate. However, GAGs can also be found as free complex polysaccharides, e.g. hyaluronic acid (Taylor and Drickamer, 2011, Moremen et al., 2012).

O-linked glycans are very heterogeneous and attachments of one single mono-saccharide up to high-mass polysaccharides of several kilodaltons, e.g. in mucous secretions, have been identified on glycoproteins (Rose, 1992).

Hart et al. discovered in the 1980s that β-linked O-GlcNAc can be attached on cytosolic and nuclear proteins and regulate cellular functions, including gene expression, as well as compete with the phosphorylation of selected Ser/Thr residues. The β-linked O-GlcNAc moieties are not extended but remain as a monosaccharide glycan modification (Hart et al., 2011). It has been shown that a

crosstalk between GlcNAc-glycosylation and phosphorylation exists, as both compete for the sites of modification (Butkinaree et al., 2010). This phenomenon has been described by some as the hexosamine signalling pathway.

The functional roles of O-linked glycans are just as diverse as their structures. Besides the aforementioned functions, O-glycans are involved in the formation of extracellular matrices and connective tissues in organs, in cell adhesion and trafficking, in immunological processes by recognition through cell-surface lectins, in apoptosis, antibody function, and antigen receptor activation (Carraway and Hull, 1991, Underhill, 1992, McEver and Cummings, 1997, Zachara et al., 2015).

Furthermore, studies have shown that O-GalNAc glycans could be potential markers for cancer and chronic inflammation (Hollingsworth and Swanson, 2004, Dube and Bertozzi, 2005).

1.1.1.3 C-linked glycosylation

In the 1990s, another type of glycosylation was found to exist in polypeptide backbones through a C-glycosidic linkage between the C2 atom of the indole moiety of a tryptophan residue and a mannose moiety. For C-linked mannosylation, the consensus sequence Trp-X-X-Trp (W-X-X-W) has been identified and usually the first tryptophan (Trp/W) residue of the sequon is mannosylated (Hofsteenge et al., 1994, Loffler et al., 1996, Dutta et al., 2017).

Within the last decades, it became clear that C-mannosylation is a common post-translational modification in mammalian proteins and driven by an enzymatic C-mannosyltransferase complex in the ER lumen (Doucey et al., 1998, Maeda and Kinoshita, 2008).

The C-mannosylation has been observed in a variety proteins, which includes - but is not limited to - interleukin-12, human RNase 2, mucins, thrombospondins as well as on the complement proteins properdin, C6, C7, C8 and C9 (Hofsteenge et al., 1994, Doucey et al., 1999, Hofsteenge et al., 1999, Hartmann and Hofsteenge, 2000, Hofsteenge et al., 2001, Perez-Vilar et al., 2004).

Interestingly, it has been shown that C-mannosylation is not recognised by conventional mannose-binding lectins, such as the plant lectin concanavalin A or mannose-binding lectin MBL-C, which is a lectin in mouse serum (Nishikawa et al.,

2004). Considering the biological context of the C-mannosylated glycoproteins, involvement in regulation of enzyme activity and immune responses, roles in molecular recognition and signalling, and even functions related to protein folding have been hypothesised for α-linked C-mannosylation. However, the function(s) have not been successfully elucidated so far (Furmanek and Hofsteenge, 2000, Ihara et al., 2011).

1.2 Lectins

Lectins are one of two subcategories of non-immune glycan-binding proteins (GBP), which generally exclude glycan-binding antibodies. The second subcategory of GBP is named glycosaminoglycan-binding proteins (Taylor et al., 2015).

The term *lectin*, originating from the Latin *legere* 'to choose, to select', was coined by William C. Boyd in 1954 (Boyd, 1954). At the time, the term mainly referred to plant-derived proteins that cause agglutination and precipitation with specific blood group antigens that are based on carbohydrate moieties. Within the following decades, the definition of a lectin was refined to "a sugar-binding protein or glycoprotein of non-immune origin which agglutinates cells and/or precipitates glycoconjugates" (Boyd and Shapleigh, 1954, Goldstein et al., 1980). It became clear that lectins are common and important in all domains of life and beyond, including viruses, bacteria, fungi, plants, insects and animals (Ofek et al., 1977, Sharon and Lis, 1989, Lis and Sharon, 1991, Gabius, 1997, Singh et al., 2011, Van Damme, 2014, Xia et al., 2018).

On molecular and cellular levels, lectins recognise carbohydrate moieties non-covalently via specific carbohydrate recognition domains (CRDs). The structural characteristics of a CRD classifies the lectin proteins and can provide insight into their typical glycan-binding properties. However, the classification is very complex due to an overlapping nature, i.e. lectin proteins can carry multiple CRDs that belong to different classes. Therefore, a final and definite assignment is not feasible for all GBPs but can rather be examined within the particular biological context. At present, numerous different lectin families are established, which includes - but is not limited to - the Calnexin type, Chitinase-like lectins, C-type lectins, F-box lectins, Ficolins, F-type lectins, Galectins, Intelectins, I-type lectins including Siglecs, L-type lectins,

M-type lectins, P-type lectins, and R-type lectins (Taylor and Drickamer, 2011, Drickamer, 2014c, Drickamer, 2014b, Taylor et al., 2015). The terminology is even more complex since additional names exist for some classes, e.g. intelectins are also known as eglectins or X-lectins.

Briefly mentioned, other systems of classification can be used, such as an intracellular or extracellular localisation of the CRD. Intracellular lectins are found in luminal compartments, e.g. the proteins calnexin and calreticulin that bind via their CRD to the remaining glucose residues of a maturing N-linked glycan in the ER lumen. Extracellular lectins can be found as transmembrane proteins with the CRD or CRDs reaching into the extracellular space, or as secreted lectins (Drickamer, 2014c, Xu and Ng, 2015).

This thesis focuses on mammalian lectins which are integrated into the plasma membrane via a transmembrane region and possess one or more extracellular CRD. These lectins recognise glycan structures outside the cell and mediate various functions, such as cell adhesion and cell signalling, as well as the trafficking, targeting and clearance of glycoproteins. Furthermore, the involvement in immune homeostasis and responses are a major functional role of lectins (Drickamer, 2014c).

The main classes of mammalian lectins with extracellular CRDs are summarised below:

Lectin proteins are ubiquitously expressed in leguminous plants, which coined the term 'L-type' lectin. They were the first class of lectins to be discovered in the late 19th century (Van Damme, 2014). This class was established on the basis of the common tertiary structure of their CRDs and they have been found present in plant, fungal and animal lectins. Many L-type CRDs preferably interact with high-mannose-type oligosaccharides in a calcium-dependent manner, a good example being concanavalin A (Kamiya et al., 2008, Van Damme et al., 2008).

The 'immunoglobulin-type' (I-type) lectins bind carbohydrate ligands via immunoglobulin superfamily domains. The best characterised I-type lectin subgroup are the Siglecs (sialic-acid-binding immunoglobulin-like lectins), which are type 1 membrane proteins and bind to sialic acid moieties. Siglecs, as well as other members of I-type lectins, contain extracellular carbohydrate-recognition domains that are

homologous to the variable region of immunoglobin G (IgG) antibodies (Powell and Varki, 1995, Crocker et al., 2007, Varki et al., 2015).

Members of the P-type lectin family have the ability to recognise phosphorylated mannose residues, which is a crucial function in the context of lysosomal trafficking within the cells of higher eukaryotes. After arrival at the luminal side of the lysosome, the acidic pH triggers the dissociation of receptor and cargo and the receptors are trafficked back to the Golgi apparatus of the cell (Dahms and Hancock, 2002, Ghosh et al., 2003).

The R-type carbohydrate-recognition domain, named after the CRD present in the castor bean protein ricin, can be found in many glycosyltransferases, bacterial and fungal hydrolases and other animal, plant, and bacterial lectins. Remarkably, the R-type CRD is well conserved throughout the domains of life and can be accompanied by an additional subunit that acts a potent toxin (Rutenber et al., 1987, Cummings and Schnaar, 2015).

The main focus of this thesis lies on C-type lectin receptors and, therefore, they will be discussed comprehensively in the following section.
C-type lectins contain a carbohydrate recognition domain that binds glycan structures and most members recognise the glycans in a calcium-dependent manner. C-type lectins serve a wide range of functions in mammalians (Zelensky and Gready, 2005).

1.2.1 C-type lectins

The C-type lectins (CTL) bind glycans via their carbohydrate-recognition domain in a calcium-dependent manner, which coined the term “C-type”. The C-type CRDs share a homologous primary structure of 110-130 amino acids with two highly conserved disulfide bridges, and specific hydrophobic as well as polar residues. These interaction partners form a secondary structure of a double-loop, which is flexible and the key arrangement for the calcium-dependent interactions with the ligand, which forms the ternary complex between the protein, calcium ion and saccharide components. In detail, the structurally flexible second loop is involved in

the coordination of calcium-ions. The coordinated Ca^{2+} ions enable themselves the coordination bonds to hydroxyl groups of the glycan or other ligands. Furthermore, hydrophobic interactions can stabilise the binding between the glycosylated interaction partner and the C-type lectin, referred to as secondary interactions (Zelensky and Gready, 2005, Drickamer, 2014a, Cummings and McEver, 2015).
However, due to the structural and evolutionary flexibility of this particular domain fold, not all members of this protein superfamily require calcium for the interaction between the CRD and the ligand. Therefore, the distinction between the sugar-binding C-type CRDs and the broad family of C-type lectin-like domains (CTLDs) has been introduced to refer to the homologous domains. Some CTLDs that carry a C-type-like CRD bind neither glycan structure nor behave in a calcium-dependent manner. Therefore, the C-type lectin (like) domain superfamily has been created, which also encloses the traditional C-type lectins (Weis et al., 1998).

The superfamily of CTLDs has been categorised into 16 groups (see figure 1.1.). The categorisation is based on characteristics and the sequence alignment of the domain. The groups cover soluble C-type lectins, such as the mannose-binding lectin (MBL) or the lung surfactant proteins A (SP-A) and D (SP-D) which are both collectins (Figdor et al., 2002, Jakel et al., 2013). Other groups include membrane-integrated proteins with C-type lectin (like) domains. To date, about 100 proteins carrying one or more CTLDs have been identified in the human genome (Drickamer, 2014d, Cummings and McEver, 2015).

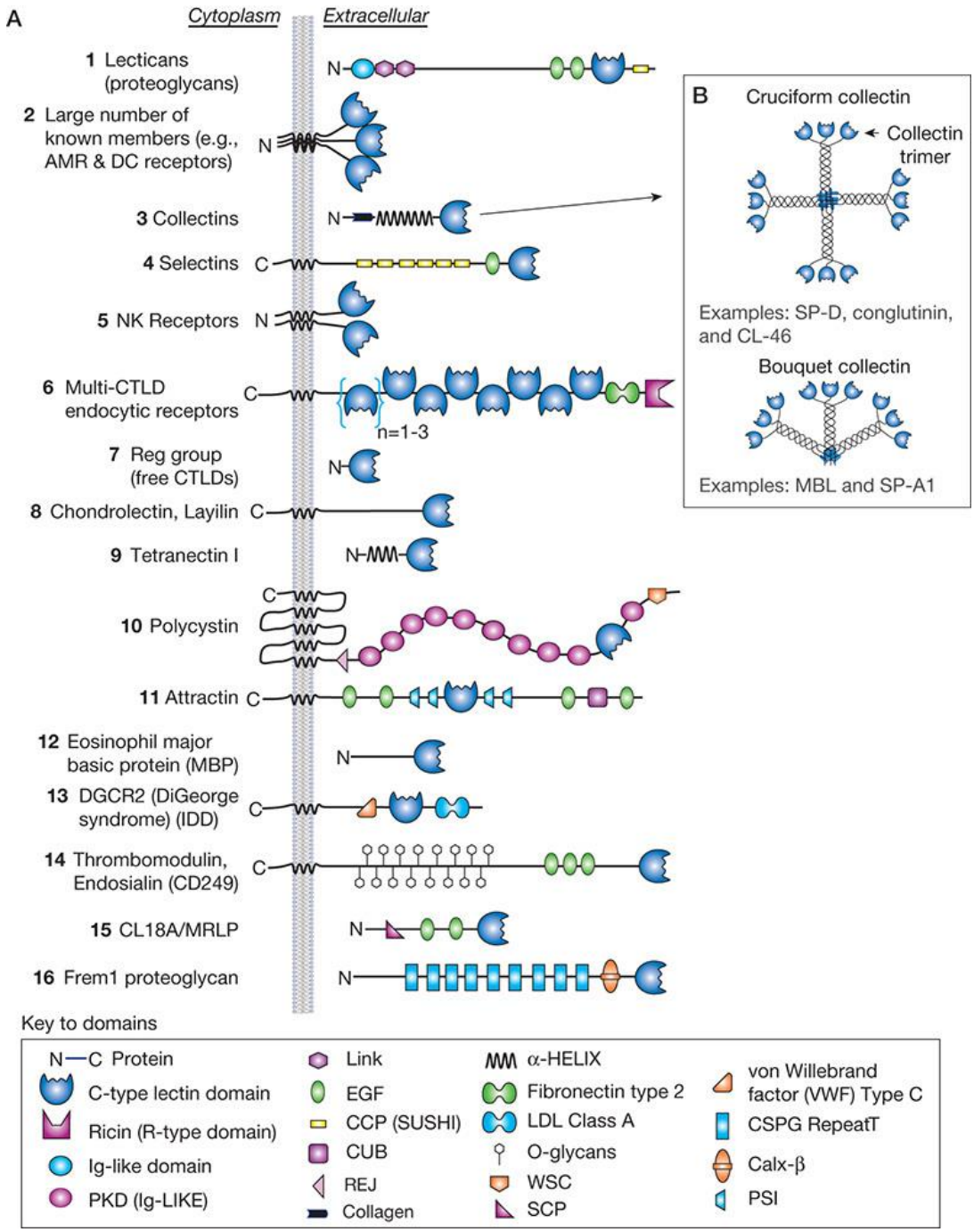


Figure 1.1: Different groups of C-type lectins (CTLs) and their domain structures.
(A) Sixteen groups are shown, defined by their phylogenetic relationships and domain structures. Some of the groups are soluble proteins and others are transmembrane proteins.
(B) The group 3 collectins form oligomeric structures, shown as cruciform and bouquet structures in the box. Each of the domains is named as indicated in the key. For abbreviations see the text.

(Figure and caption have been adapted from (Cummings and McEver, 2015) with kind permission from [Essentials of Glycobiology (Third Edition), Chapter 34, Figure 3, Cold Spring Harbor Laboratory Press, ©2017 The Consortium of Glycobiology Editors, La Jolla, California])

The focus of this thesis lies on animal lectins. Particularly, lectins expressed on human immune cells that bind to sugar moieties in a calcium-dependent manner. They are referred to as C-type lectin receptors (CLRs) and are assigned into the CTL group 2 for the Dendritic cell-specific intercellular adhesion molecule-grabbing non-integrin (DC-SIGN; CD209) and Dendritic cell-specific intercellular adhesion molecule-grabbing non-integrin-*related* protein (DC-SIGNR; CD299, L-SIGN), and the CTL group 6 for the Macrophage Mannose Receptor 1 (MMR; CD206), respectively.

Membrane-integrated C-type lectins are single-pass transmembrane proteins with the majority being type II transmembrane proteins, except for the Macrophage Mannose Receptor (MMR), Endo-180 and DEC-205, which are type I transmembrane receptors (Figdor et al., 2002). Type I proteins have their N-terminus on the ER luminal or extracellular side of the membrane. However, type II transmembrane proteins span the membrane with their N-terminus on the cytoplasmic side of the membrane (Curtis et al., 1992).

Typically, carbohydrate-protein interactions are selective but weak; for example, CRD-monosaccharide interactions of the C-type lectins DC-SIGN, DC-SIGNR, and MBL with mannose show dissociation constants (K_D) about one millimolar, respectively. This weak affinity is counterbalanced by nature through a multivalent presentation of CRDs. For example, DC-SIGN as well as DC-SIGNR form tetramers via parallel alpha-helical coiled-coil domains, which results in a supercluster effect and high-avidity binding platforms for multivalent ligands (Lee et al., 1992, Lee and Lee, 2000, Mitchell et al., 2001, Cummings and McEver, 2015). However, the single-polypeptide chain and non-oligomerising membrane protein MMR overcomes the weak affinity of a single CRD by the presentation of eight consecutive, tandem C-type CRDs that are intrinsic in the primary structure (Taylor et al., 1990). Both styles of arrangement ensure recognition of biologically relevant targets by the requirement of matching target density as well as configuration.

As mentioned, C-type lectins recognise glycans selectively and many can be broadly classified into either the mannose-binding subtype, e.g. DC-SIGN, DC-SIGNR, and MMR, or the galactose binding subtypes, e.g. the Asialoglycoprotein receptor (ASGPR) and the Macrophage Galactose-type Lectin (MGL; CD301) (Zelensky and Gready, 2005). The specificity for mannose-type sugars of some CLRs derives from the presence of the amino acid sequence Glu-Pro-Asn, also called the EPN motif, in

the primary structure of the calcium-binding site in the loop region of the CRD. The resulting secondary structure causes less steric hindrance for adjacent equatorial 3-hydroxyl and 4-hydroxyl groups, which can be found in D-mannose, N-acetyl-D-glucosamine, D-glucose, or L-fucose. On the other hand, the galactose binding CTL subtypes carry a Gln-Pro-Asp sequon (QPD motif) within their calcium-dependent CRD, which favours the axial 4-hydroxyl group sterically and involves ligands such as galactose and N-acetylgalactosamine residues (Drickamer, 1992, Sharon and Lis, 2007).

However, C-type lectins can bind ligands in several modes as it has been elucidated for DC-SIGN, DC-SINGR and MGL (Probert et al., 2013, Marcelo et al., 2014, Pederson et al., 2014). The putative diversity in binding modes of the ligand to the CRD as well as the importance of secondary interactions sites to certain CRDs broadens the specificity for some CTLs (Sharon and Lis, 2007, Drickamer and Taylor, 2015).

The recognition of ligands by animal lectins can initialise and concert various processes, which include but are not limited to adhesion events, intracellular signalling, involvement in homeostasis and inflammation, recognition of (damaged) self, and regulation of embryonic development (Gabius, 1997, Brown et al., 2018).
The functions of lectins can also to a large extent be assigned according to the lectin family. For example, members of the Calnexin family are involved in protein sorting in the endoplasmic reticulum, while members of the M-type lectins are involved in the ER-associated degradation of glycoproteins. L-type lectins have been found to be mainly involved in protein sorting in the endoplasmic reticulum and P-type lectins take part in protein sorting and trafficking, as well as ER-associated degradation of glycoproteins, and enzyme targeting. The extracellular lectin family of Galectins provides the crosslinking of glycans in the extracellular matrix. The I-type lectins, including their well characterised subfamily of Siglecs, has been found to contribute to cell adhesion. R-type lectins have been found in processes of enzyme targeting and glycoprotein hormone turnover. The members of the lectin family of F-box lectins is involved in the degradation of misfolded glycoproteins. Members of Ficolins, F-type lectins and Intelectins are associated with various processes in innate immunity such as complement activation. C-type lectins, as a final point, are involved in adhesion

events, endocytosis and clearance of glycoproteins, recognition of pathogens and regulation of the innate immunity (Drickamer, 2014c). Since specific CLRs are expressed on the cell surface of various but distinct types of immune cells, the CLRs can serve as markers for specific cell types or even their cellular states (Brown et al., 2018). However, the intracellular signalling pathways of CLRs subsequent to ligand binding events is still "the least well understood function of these receptors" (Drickamer and Taylor, 2015).

1.2.1.1 Dendritic cell-specific intercellular adhesion molecule 3-grabbing non-integrin 1 (DC-SIGN)

The Dendritic cell-specific intercellular adhesion molecule [ICAM-3]-grabbing non-integrin (DC-SIGN) is a member of the group 2 of the superfamily of CTLDs. It is also referred to as the surface marker CD209 or the C-type lectin domain family 4 member L (CLEC4L). It was first discovered in the early 1990s as an entry site for the human immunodeficiency virus via interaction with its glycosylated envelope protein gp120, and simply described as a gp120-binding C-type lectin (Curtis et al., 1992, Geijtenbeek et al., 2000b). It later became clear that the physiological relationship with the immune system is even more prominent: the actual nomenclature derives from DC-SIGN's characteristic to bind to the intercellular adhesion molecule 3 (ICAM-3), mainly expressed on lymphocytes. This interaction mediates an adhesion between Dendritic cells (DCs) and resting T cells, which express ICAM-3 on their cell surface. The interaction is an essential trigger for T cell proliferation (Geijtenbeek et al., 2000c). DC-SIGN itself is abundantly expressed on the surface of immature dendritic cells, and also on mature dendritic cells of dermal, lymphoid, and mucosal DCs, as well as on a subset of blood DCs. Furthermore DC-SIGN has been found to be expressed by hepatic sinusoidal endothelial cells, subpopulations of B cells, and on subsets of platelets. Moreover, DC-SIGN has been found on M2-type macrophages, such as interstitial and alveolar macrophages, as well as on maternal decidual macrophages and foetal-derived macrophages (Hofbauer cells) in the placenta (Geijtenbeek et al., 2000c, Engering et al., 2002c, Soilleux et al., 2002, Kammerer et al., 2003, Puig-Kroger et al., 2004, Chaipan et al., 2006, Lai et al., 2006).

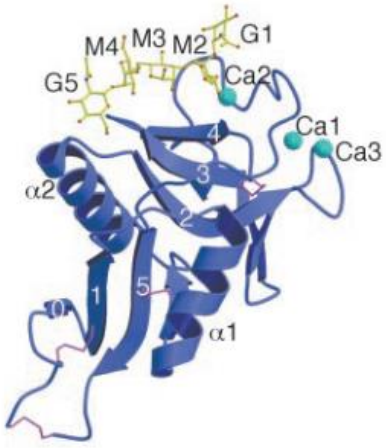


Figure 1.2: Crystal structure of the CRD from DC-SIGN (shown in blue) bound to a pentasaccharide (presented in ball-and-stick model in yellow; G for GlcNAc; M for mannose). Two of the three calcium binding sites (cyan, Ca1/3) are crucial for structural integrity of the CRD and one (Ca2) is important for coordination of the carbohydrate ligand via hydrogen bonds. Figure and caption have been adapted and altered with kind permission from (Feinberg et al., 2001).

The location of the gene that encodes for DC-SIGN has been identified on human chromosome 19p13

(Soilleux et al., 2000). The encoded information includes an extracellular, a transmembrane and a cytosolic domain within a total of 404 amino acids. DC-SIGN belongs to the majority of CLRs which are type II transmembrane proteins (Geijtenbeek et al., 2000c). The C-terminal, extracellular domain (depicted in figure 1.2) carries one calcium-dependent CRD with an EPN motif for recognition of carbohydrates, and shows a typical long-form C-type lectin fold. The CRD contains three calcium binding sites, whereby one site is important for coordination of the carbohydrate ligand via hydrogen bonds formed with its hydroxyl groups. The other calcium binding sites mainly ensure structural integrity of the C-type lectin fold (Feinberg et al., 2001, Feinberg et al., 2007).

The CRD region is preceded by a hydrophobic stretch of 15 amino acids within the extracellular domain. The following tandem-repeating neck region consists of 7.5 α-helical repeats, whereby each repeat is formed by 23 highly conserved amino acids.

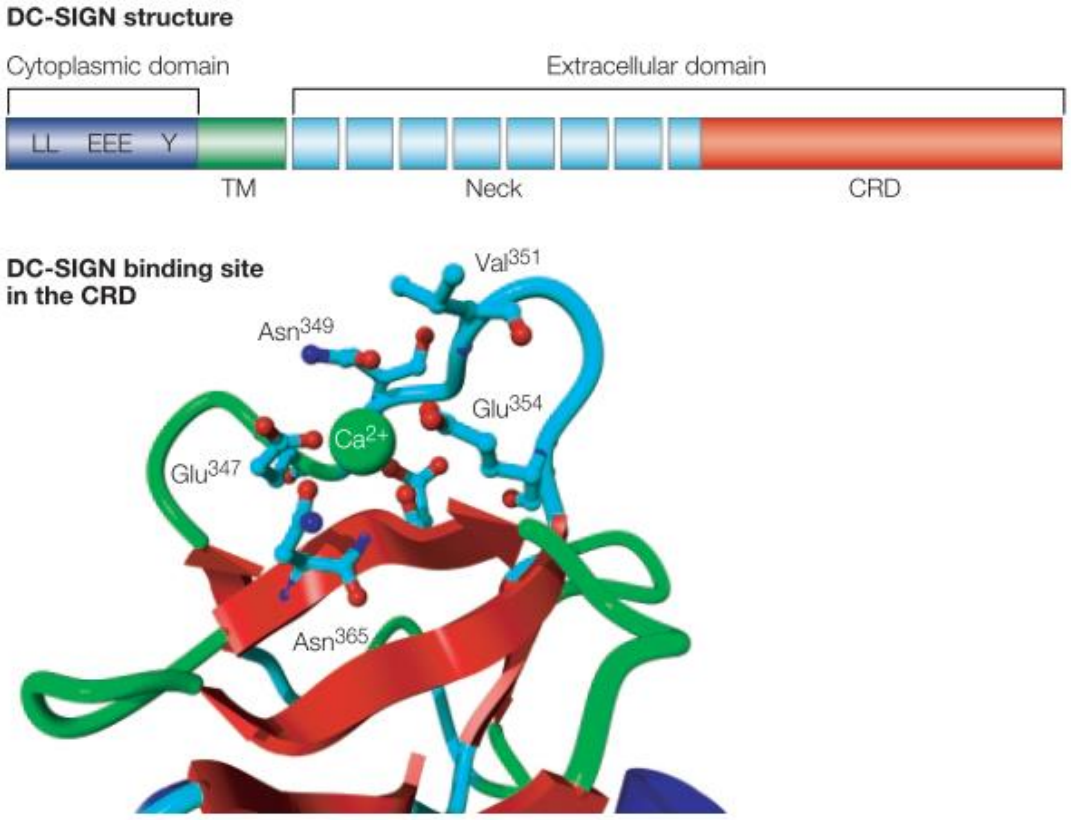


Figure 1.3: Schematic overview of the DC-SIGN structure. The carbohydrate recognition domain (CRD) of DC-SIGN is separated from the transmembrane (TM) region by a neck domain that consists of seven complete and one incomplete tandem repeat and facilitates oligomerisation. The cytoplasmic tail includes internalisation motifs, such as the di-leucine (LL) motif and the tri-acidic (EEE) clusters, and an incomplete immunoreceptor tyrosine-based activation motif (ITAM). Figure and caption have been adapted and altered with kind permission from (van Kooyk, 2003).

The repeats serve as an oligomerisation motif and form alpha-helical stalks that non-covalently homo-tetramerise. The tetramerisation provides a high-avidity binding platform resulting in a supercluster effect, which not only increases affinity via avidity but also specificity (Mitchell et al., 2001). The increase in specificity is realised by the defined spacing of the CRDs through tetramerisation. Moreover, the arrangement ensures that the CRDs are projected away from the cell surface for improved recognition of extracellular ligands (Svajger et al., 2010).

A hydrophobic transmembrane region with predicted alpha-helical character connects the extracellular domains of the CLR with the cytoplasmic tail (Geijtenbeek et al., 2000c). The N-terminal, cytoplasmic domain of DC-SIGN carries a di-leucine motif (LL), a tri-glutamate acidic cluster (EEE), as well as a tyrosine motif (YKSL) (The UniProt Consortium, 2004). These have been linked to intracellular signalling pathways as well as triggering antigen/pathogen internalisation for processing and subsequent presentation at the surface of the antigen presenting cell (APC) (Mellman, 1996, Engering et al., 2002a, Yon et al., 2009). In 2007, Hodges et al showed in a study using mutated DC-SIGN constructs that "[l]oss of signalling occur[ed] exclusively for the double mutant (with both the dileucine motif and the relevant tyrosine replaced by alanine), which suggested a critical function for the tyrosine residue in the DC-SIGN cytoplasmic tail in the transmission of intracellular signals." (Hodges et al., 2007).

It has become clear over the past decade that DC-SIGN modulates signalling pathways which are initiated by toll-like receptors (TLR), also referred to as crosstalk with other pattern recognition receptors (PRR). DC-SIGN has been shown to rely on the prior activation of the nuclear factor-κB (NF-κB) signalling pathway by other PRRs. NF-κB is a nuclear transcription factor and plays a prominent role in inducing gene expression within immunological responses. Prior to activation, DC-SIGN is bound by the adaptor protein Lymphocyte-specific protein 1 (LSP1). Furthermore, the proteins Kinase Suppressor of Ras 1 (KSR1), Connector Enhancer of KSR (CNK), and the kinase RAF1 form a KSR1-CNK-RAF1 triad which serves as a signalling complex (signalosome). The KSR1-CNK-RAF1 signalosome forms a complex with LSP1. Upon ligand binding at the CRD of DC-SIGN, the signalosome is activated and able to modulate the immunological response in either a pro- or anti-inflammatory direction.

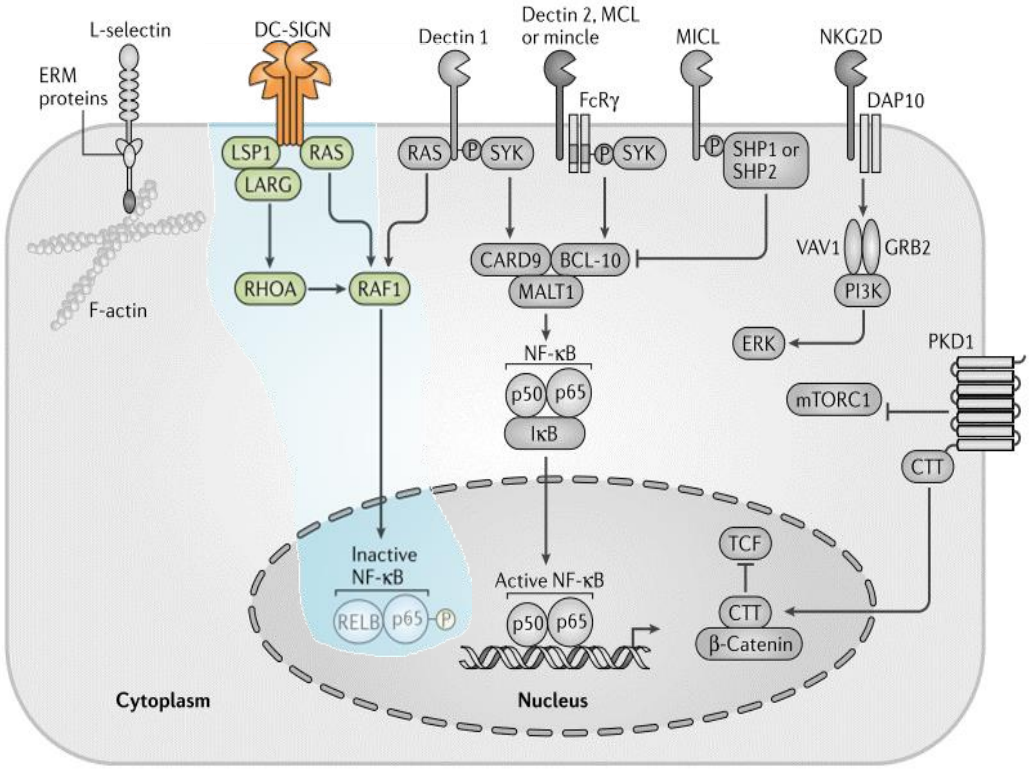


Figure 1.4: DC‑SIGN signalling pathway triggered by mannose-containing ligands (shown in the coloured section). DC-SIGN associates with the adaptor lymphocyte-specific protein 1 (LSP1) and can recruit the serine/threonine-protein kinase RAF1 signalosome, upon binding a mannose-containing ligand as well as parallel activation of other PRRs. DC-SIGN then requires the recruitment of Transforming protein RhoA (RhoA) and the Leukaemia-associated Rho guanine nucleotide exchange factor 12 (LARG) for activation of the RAF1 signalosome. The activation of the RAF1 signalosome results in phosphorylation and deacetylation of the p65 subunit of NF-κB in the nucleus. This leads to an increase in pro-inflammatory immune responses via cytokines IL-12, IL-23 and IL-6. Figure and caption have been adapted and altered with kind permission from (Brown, 2018).

Findings from the Gringhuis and Geijtenbeek groups showed that the modulation of the signalosome is dependent on the nature of the carbohydrate ligand that binds to the CRD of DC-SIGN. Recognition of mannose-containing glycan structures leads to activation of the RAF1 signalosome and results in a production of pro-inflammatory cytokines as shown in figure 1.4. Briefly, the serine/threonine-protein kinase RAF1 signalosome serves as a docking site for signalling proteins Transforming protein RhoA (RhoA) and the Leukaemia-associated Rho guanine nucleotide exchange factor 12 (LARG). Subsequently, the activation of the RAF1 signalosome results in phosphorylation and deacetylation of the p65 subunit of NF-κB (Gringhuis et al., 2007, Geijtenbeek and Gringhuis, 2009, Geijtenbeek and Gringhuis, 2016, Brown et

al., 2018). On the other hand, recognition of fucose-containing structures by DC-SIGN have not only been found to be independent of RAF1 activation but the KSR1-CNK-RAF1 signalosome is detached from DC-SIGN whilst LSP1 remains attached. The proteins Inhibitor of nuclear factor kappa-B kinase subunit epsilon (IKKε) and the Ubiquitin carboxyl-terminal hydrolase CYLD (CYLD) attach to LSP1. This complex has been termed the 'fucose signalosome' in the literature. After activation of CYLD and IKKε, an atypical member of the NF-κB protein family gets transported into the nucleus. The accumulating B-cell lymphoma 3 protein (Bcl3) forms complexes with other transcription factors that lead to upregulation of the expression of anti-inflammatory cytokine interleukin 10 (IL-10) as well as the suppression of pro-inflammatory cytokines IL-12, IL-23 and IL-6 (Gringhuis et al., 2009, Gringhuis et al., 2014, Geijtenbeek and Gringhuis, 2016). This shows that modulation of intracellular signalling by DC-SIGN is even more complex and suggests a high degree of plasticity in DC-SIGN signalling. The overall resultant production of cytokines by the cell reflects a tailored immune response to specific antigens or pathogens (Geijtenbeek and Gringhuis, 2009, Gringhuis et al., 2009).

DC-SIGN was identified through the investigation of the mechanisms that underpin the contact between dendritic cells (DCs) and resting T cells, leading to T cell activation. Studies highlighted DC-SIGN on DCs as the mediator for adhesion via its characteristic to bind to the Intercellular adhesion molecule 3 (ICAM-3) that is abundantly expressed on T cells. Initially, the β2 integrin receptors leukocyte adhesion protein 1 (LFA-1 or integrin α-L/β-2) and the integrin α-D/β-2 were thought to mediate the DC-T cell interaction. Since "DC interact with ICAM-3 through an integrin-independent mechanism that requires Ca^{2+}, [Geijtenbeek] designated this receptor DC-specific ICAM-3 grabbing nonintegrin (DC-SIGN)"(Geijtenbeek et al., 2000c). It was shown that DC-SIGN also binds to ICAM-2 but with a higher binding affinity than ICAM-3, and contributes to the tethering and rolling of DCs on surfaces such as vascular endothelium that express ICAM-2 (Geijtenbeek et al., 2000a). DC-SIGN interacts with N-linked high-mannose oligosaccharide structures with submillimolar affinity (Mitchell et al., 2001), preferentially with structures from $Man_5GlcNAc_2$ to $Man_9GlcNAc_2$. These glycans have been identified on ICAM-2, ICAM-3 and HIV gp120. The EPN motif in the CRD of DC-SIGN imparts a higher specificity for mannose-type sugars, which carry

adjacent equatorial 3-hydroxyl and 4-hydroxyl groups, e.g. mannose, N-acetylglucosamine, glucose, and L-fucose (Drickamer and Taylor, 2015). Through the use of glycan arrays, further endogenous ligands for DC-SIGN have been identified, and these are fucose-containing Lewis blood-group antigens such as $Lewis^X$, $Lewis^Y$, $Lewis^a$, and $Lewis^b$ carbohydrate epitopes, in addition to fucose-containing ABH Blood Group oligosaccharides, i.e. Blood Groups A and B (Guo et al., 2004).

The CRD of DC-SIGN binds the trisaccharide structure Manα1-3[Manα1–6]Man four times better than the monosaccharide mannose. A key characteristic for the recognition is the α-linkage of the core mannose, which is only found in high mannose oligosaccharides. This finding was elucidated by Feinberg et al. via X-ray crystal structure analysis of a pentasaccharide bound to the CRD of DC-SIGN and subsequent molecular modelling. A representation of the arrangement is shown in figure 1.2., where the annotation M2 marks the core mannose residue that interacts with the ligand-coordinating calcium ion. Furthermore, Feinberg et al. determined the precise interaction sites on the polypeptide chain and demonstrated that DC-SIGN interacts with distal components of the pentasaccharide structures also via additional interaction sites close to the CRD (Feinberg et al., 2001, Feinberg et al., 2007).

Similar carbohydrate structures have been found on the surfaces of various microorganisms, which are exogenous ligands for DC-SIGN. This includes bacteria, fungi, and viruses, e.g. mycobacterium tuberculosis, mycobacterium leprae, lactobacilli spp., helicobacter pylori, candida albicans, HIV‑1, measles virus, dengue virus, SARS coronavirus, and filoviruses. All of them have been found to exploit the modulating immune function of activated DC-SIGN to escape the immune surveillance of the host and mitigate neutralisation by the immune system (van Kooyk and Geijtenbeek, 2003, Guo et al., 2004, Geijtenbeek and Gringhuis, 2009).

Curtis et al. showed in 1992 that a hitherto undefined C-type lectin binds strongly to the HIV glycoprotein 120 (gp120), which is an envelope protein of the virus (Curtis et al., 1992). However, almost a decade later, this C-type lectin was "rediscovered" as

DC-SIGN and it was suggested that it might enhance the viral infection of T cells in *trans* via the exploitive nature of HIV (Geijtenbeek et al., 2000b). The high affinity interaction is facilitated between the CRD of DC-SIGN and the densely clustered high mannose glycan structures present on gp120 (Mizuochi et al., 1990, Feinberg et al., 2007). DC-SIGN has been a prominent target for HIV prophylaxis or therapeutic intervention since 2000, although unfortunately, this has not resulted in a commercially viable HIV treatment or prophylaxis option so far. However, the development of anti-viral agents with a higher affinity than HIV gp120 for the CRD of DC-SIGN have recently made significant progress with the synthesis and characterisation of mannose-containing glycopolymers (Becer et al., 2010).

Human milk glycans (HMGs), also referred to as human milk oligosaccharides (HMOs), have been defined as pathogen receptor decoys, modulators of neonatal immunity in the infant gut, regulators of host physiology, and act as prebiotics for the infant's gut microbiome (Bode, 2012). Noll et al. showed recently that DC-SIGN binds to human milk glycans – primarily to those which contain an α-linked fucose moiety, with an affinity that matches the physiological HMG concentrations. They suggested "that DC-SIGN may serve as an HMG receptor, which may have implications in infant immunity, physiology and development" via regulating gene expression and subsequent immune responses in the developing gut (Noll et al., 2016).

Moreover, it is still not fully understood why DC-SIGN-expressing cells are very prominent on both sides of the foetal-maternal interface during human pregnancy. Pregnancy can be interpreted as a highly immunologically privileged state in which the maternal immune system accepts the allogeneic embryo and subsequently the developing foetus without fully suppressing the maternal immune system. There is growing evidence that DC-SIGN plays an important role for the immune responses in human pregnancy, although the mechanisms underpinning this are not fully understood (Kammerer et al., 2003).

As a result of the structural and binding characteristics, human DC-SIGN acts as an adhesion molecule for the endogenous, N-glycosylated ICAM-2 and ICAM-3: DC-SIGN facilitates the tethering and rolling of DCs on surfaces that express ICAM-2, as well as contributing to the DC-T cell interaction and immunological synapse

through interactions with ICAM-3. This process helps to bridge the innate and adaptive immune responses and is crucial for the activation of T cells (Geijtenbeek et al., 2000c, Geijtenbeek et al., 2000a).

DC-SIGN also recognises carbohydrate structures on self-antigens as well as on foreign antigens, which is important for maintaining immune homeostasis (Geijtenbeek et al., 2004). When DC-SIGN binds foreign ligands, internalisation can be induced by activation of the di-leucine motif as well as the tri-glutamate acidic cluster in the cytoplasmic domain. The internalised ligand is transferred to proteolytic compartments and either degraded or processed for downstream antigen presentation via major histocompatibility complex (MHC) class II molecules on the surface of Dendritic cells (Engering et al., 2002b). Therefore, DC-SIGN also functions as an endocytic antigen receptor (Geijtenbeek et al., 2004).

It has been shown that the release of the ligand from the CRD is pH sensitive. Ligands bind to DC-SIGN at a physiological pH 7.4 and the CRD releases the cargo under more acidic conditions (pH 5.5) in lysosomal compartments, which was shown by Guo et al (2004). It also has been speculated that DC-SIGN might serve as a clearance receptor of glycoproteins in the human body, which are simply degraded in lysosomal compartments after uptake and internalisation (Guo et al., 2004). Initially, DC-SIGN was considered to be a scavenger receptor only. However, the knowledge that DC-SIGN can modulate immune responses after binding to glycan ligands via influencing cytokine production suggests that it is more sophisticated than a simple scavenger. As discussed above, mannose-containing ligands can promote expression of a mixture of pro-inflammatory cytokines, notably IL-6, and fucose-containing ligands suppress pro-inflammatory responses as well as driving anti-inflammatory cytokine production (Gringhuis et al., 2009). DC-SIGN can also passively modulate immune responses by antigen presentation on DCs, which will trigger downstream adaptive immune responses via T and B cells (Engering et al., 2002b).

In summary, DC-SIGN functions as a signalling, adhesion as well as antigen-uptake molecule (Svajger et al., 2010). Given the extent of the known functions, it is very likely that DC-SIGN binds more, yet unknown endogenous glycan ligands and provides more physiological functions.

1.2.1.2 Dendritic cell-specific intercellular adhesion molecule 3-grabbing non-integrin 1-related protein (DC-SIGNR)

Encoded adjacent to the DC-SIGN gene on the human chromosome 19p13, a very similar CLR receptor was identified and named Dendritic cell-specific intercellular adhesion molecule 3 [ICAM-3]-grabbing non-integrin 1-related protein (DC-SIGNR) (Soilleux et al., 2000). Since the DC-SIGNR receptor is mainly expressed on endothelial cells in the liver sinusoid, lymph nodes sinuses, and in capillary endothelial cells in term placenta, it is also referred to as L-SIGN (Liver/Lymph node-specific ICAM-3 grabbing non-integrin) (Bashirova et al., 2001, Pohlmann et al., 2001). However, DC-SIGNR has not been found on dendritic cells or macrophages (Figdor et al., 2002). Furthermore, it has been assigned the cluster of differentiation name of CD299 and the CTLD family member name CLEC4M.

DC-SIGN and DC-SIGNR share 73% identity within their nucleic acid sequence and even 77% identity of the primary protein structure. This results in a high level of homology in secondary and tertiary structure, i.e. both CLRs are type II transmembrane proteins with similar domains (Soilleux et al., 2000). However, DC-SIGNR's cytoplasmic domain only possesses a di-leucine motif and no strategic tyrosine motif. The transmembrane domain is followed by an extended neck domain that also facilitates tetramerisation via interactions of hydrophobic amino acid residues (Drickamer and Taylor, 2015). The research group around Kurt Drickamer showed that DC-SIGN and DC-SIGNR neck regions as well as CRDs differ considerably: DC-SIGNR presents a higher degree of neck length polymorphism, with potential for up to nine tandem repeats. The C-terminal CRD of DC-SIGNR also shows calcium-dependent mannose-binding specificity (Soilleux et al., 2000), but binds with comparatively very weak affinity to fucosylated glycan structures, such as Lewis antigens, compared with DC-SIGN (Mitchell et al., 2001, Guo et al., 2004, Feinberg et al., 2009).

Past studies by Guo et al. showed that the CRD of DC-SIGN can release ligands in a pH sensitive manner (Guo et al., 2004). However, DC-SIGNR remained to be investigated in full for the ability to facilitate endocytic functions. In 2013, it was found that DC-SIGNR binds and internalises the human blood serum glycoprotein

von Willebrand factor and influences its *in vivo* blood plasma levels. This suggest a potential role in von Willebrand factor bioavailability and clearance (Rydz et al., 2013). Furthermore, recent studies using heteronuclear NMR spectroscopy showed that DC-SIGNR does indeed release glycan ligands at a pH of 4.2, which indicates a potential role as an endocytic receptor for the clearance of senescent glycoproteins (Probert et al., 2014).

DC-SIGNR binds to endogenous ICAM-3, which suggested roles in cell adhesion and in immune responses and trafficking by interacting with resting T cells (Bashirova et al., 2001, Yabe et al., 2010). A study illuminating the interaction between DC-SIGNR and the respiratory syncytial virus (RSV) also revealed that DC-SIGNR is involved in cell signalling via phosphorylation of kinases within signalling cascades for activation of immune responses (Johnson et al., 2012).

Furthermore, DC-SIGNR has been found to bind to mannose-containing ligands on pathogens including HIV, the severe acute respiratory syndrome coronavirus (SARS), the influenza virus, as well as the dengue virus (Chan et al., 2006, Boily-Larouche et al., 2009, Dejnirattisai et al., 2011, Londrigan et al., 2011, Johnson et al., 2012).

DC-SIGNR also recognises glycan structures which are associated with cellular damage or cell death. These indicator molecules are referred to as damage-associated molecular patterns (DAMPs) (Neumann et al., 2014).

Given the substantial number of mannose-rich biomolecules, neither the range nor extent of exogenous and endogenous ligands for DC-SIGNR is likely to have been fully described to date. The identification and characterisation of further ligands could therefore provide more insight into the mechanism of binding as well as the physiological roles of DC-SIGNR.

1.2.1.3 Macrophage Mannose Receptor (MMR)

The macrophage mannose receptor (MMR), or macrophage C-type lectin (MCL), is another member of the C-type lectin superfamily and belongs to the CTL group VI, which are multi-CTLD endocytic receptors (see figure 1.1. on page 43). It has been assigned the cluster of differentiation number CD206 and the C-type lectin domain family 13 member D (CLEC13D).

The MMR is a type I transmembrane CLR with multiple tandem extracellular domains (Martinez-Pomares, 2012). The N-terminal domain is a cysteine-rich (CR) domain that mainly recognises carbohydrates with terminal sulfated glycans, such as SO_4-(3)-galactose or SO_4-(3/4)-N-acetylgalactosamine (Fiete et al., 1998, Leteux et al., 2000). Interestingly, this domain is an R-type CRD (Napper et al., 2006). The CR domain is followed by a domain "which resembles the type II repeats of fibronectin" (FNII) (Taylor et al., 1992). The FNII domain is involved in cellular interactions as well as recognition and internalisation of antigens such as collagen (Martinez-Pomares et al., 2006, Napper et al., 2006). Following these domains, and less commonly amongst the CLRs, the MMR contains a series of eight C-type CRDs in tandem within a single polypeptide that has been shown to increase avidity to overcome the weak affinity of single CRDs (Taylor et al., 1990, Taylor and Drickamer, 1993). CRDs 1-3 contribute only weakly to carbohydrate binding, CRD 4-5 together have a high affinity in the low nanomolar range for binding mannose/ fucose/GlcNAc-containing glycans, and CRDs 6-8 contribute to ligand-binding (Taylor et al., 1992, Taylor and Drickamer, 1993). The CRDs recognise mainly high mannose oligosaccharides, but recognition of antigens with single mannose moieties is evident as well, in addition to glycans with terminal fucose and N-acetyl-glucosamine residues (Taylor et al., 2015). The MMR is anchored via a transmembrane domain. Interestingly, the MMR can also undergo proteolytic cleavage from its transmembrane region, which results in the release of a soluble form (Jordens et al., 1999).

Figure 1.5 depicts the variety of exogenous ligands, such as viruses, fungi and bacteria, allergens, as well as endogenous ligands via the CRDs that bind to MMR (Martinez-Pomares, 2012).

MMR has been extensively studied in immature monocyte-derived DCs and macrophages, where it is abundantly expressed (Sallusto et al., 1995). Furthermore, it can be found on nonvascular endothelium, e.g. liver sinusoidal endothelial cells (LSECs) (Malovic et al., 2007).

The MMR is a multifunctional receptor: It mediates the "clearance of many potentially antigenic proteins from the circulation" (Figdor et al., 2002); it is a defence mechanism against microorganisms and capable of driving their endocytosis directly (Lee and Lee, 2000). The receptor is involved in antigen uptake and processing within adaptive immune responses (Sallusto et al., 1995, Weis et al., 1998). Collagen is released on sites of inflammation and since MMR's characteristic is to adhere to collagen, a role in augmenting immune responses has been proposed (Napper et al., 2006). Overall, the macrophage mannose receptor has been considered to provide a major link between the cellular innate and adaptive immune systems (Taylor et al., 2005).

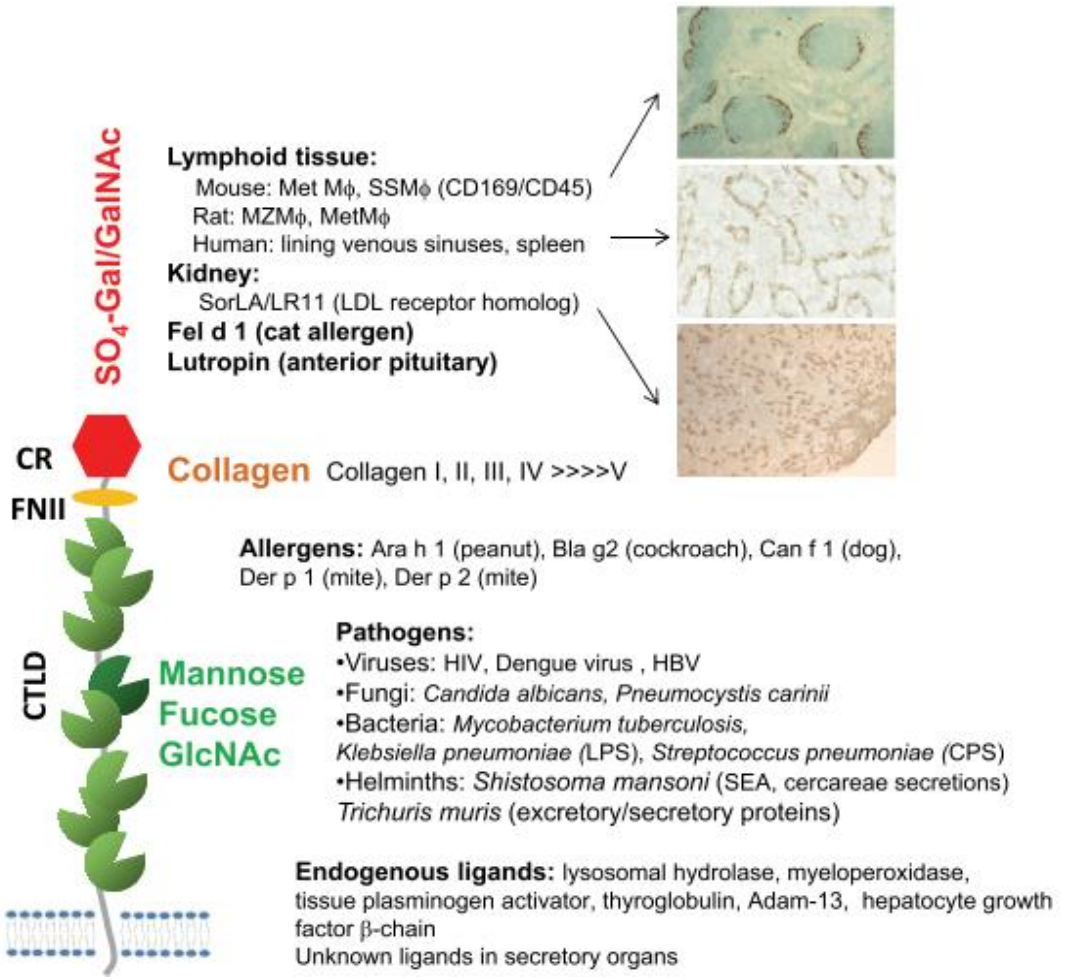


Figure 1.5: Structural properties of MMR. MMR is a type I membrane molecule with three types of extracellular domains. Through the CR domain, MR binds sulfated glycans that can be found in lymphoid tissues and kidney, as well as the major cat allergen Fel d 1 and lutropin. The FNII domain binds collagens, and the CTLDs bind endogenous and exogenous molecules, including allergens and microbial products.
Abbreviations: MΦ - macrophage; HBV - hepatitis B virus; CPS - capsular polysaccharide; SEA - secreted egg antigen; Adam-13 - a disintegrin and metalloprotease 13
Figure and caption have been adapted and altered with kind permission from (Martinez-Pomares, 2012).

1.3 Thesis Aims

Affinity chromatography using DC-SIGN immobilised on Sepharose resin, was used to screen for potential ligands from a solubilised placental extract in a prior research project (Ilyas, 2010).

Preliminary analysis of material selectively eluted from DC-SIGN-Sepharose (referenced against non-specific column binding on non-immune IgG-Sepharose) suggested PAPP-A1 (pregnancy-associated plasma protein A1) as a potential DC-SIGN ligand (figure 1.6). The identity of PAPP-A1 was determined via peptide mass spectrometry of bands excised from the SDS-PAGE gel and also validated via Western blotting.

Bioinformatic analysis revealed that the PAPP-A1 polypeptide chain possesses 14 potential N-glycosylation sites and is likely to be secreted. In its native form, PAPP-A1 is found most commonly in the plasma of pregnant women as a secreted 2:2 heterotetrameric complex with the proform of eosinophil major basic protein (pro-MBP). The level of potential N-glycosylation in this complex makes the PAPP-A1/pro-MBP complex an attractive candidate as a soluble, natural ligand for DC-SIGN.

When this discovery was made in 2010, techniques were not available to measure reliably the interaction kinetics with native glycoproteins that are only available as purified reagents in low amounts (typically microgram quantities). The same was the case for comprehensive glycan analysis of scarce, native human glycoproteins.

However, with the development of highly sensitive and sample-efficient analytical methods such as Biolayer Interferometry and nano-LC/MS, these questions can now be approached.

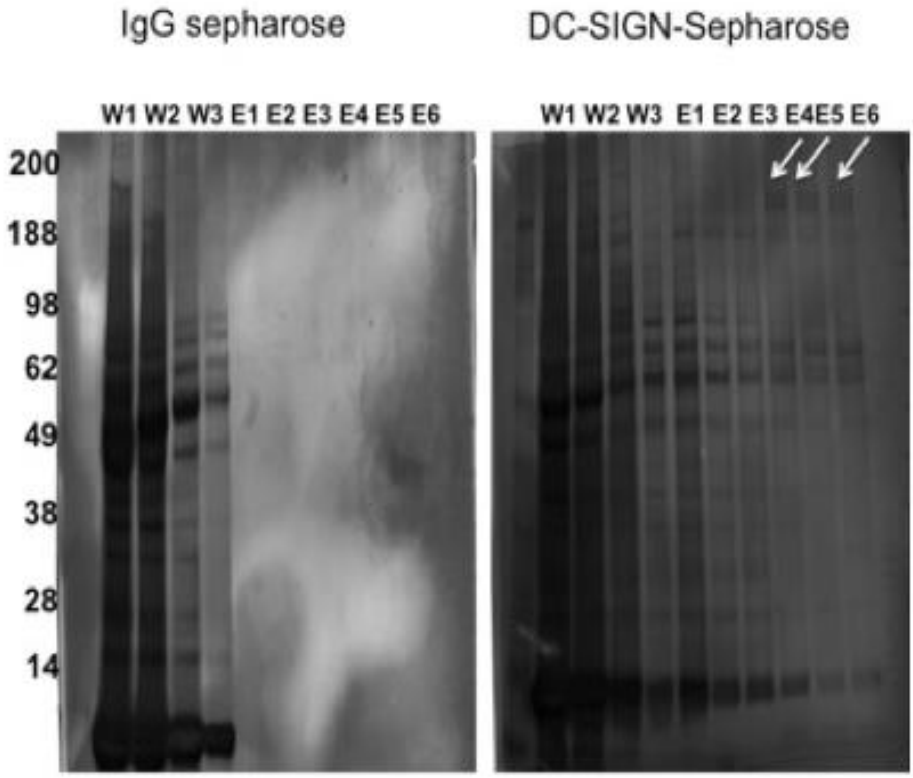


Figure 1.6: Placental ligands eluted from DC-SIGN-Sepharose. IgG-Sepharose was used as negative control. DC-SIGN-exclusive bands are visible in the late elution fractions (arrows) and represent good candidate ligands. Figure and caption have been adapted and altered with kind permission from (Ilyas, 2010).

2 MATERIALS, INSTRUMENTS AND METHODS

2.1 Materials for interaction studies

2.1.1 Chemical reagents

Table 1: Chemical reagents

Reagent	Product Code	Manufacturer
2-Mercaptoethanol	161-0710	Bio-Rad Laboratories Ltd., UK
Ampicillin sodium salt	A9518	Sigma-Aldrich Company Ltd., UK
Bovine serum albumin	A2153	Sigma-Aldrich Company Ltd., UK
Calcium chloride dihydrate	31307	Sigma-Aldrich Company Ltd., UK
D-(-)-Fructose	F0127	Sigma-Aldrich Company Ltd., UK
D-(+)-Galactose	G0625	Sigma-Aldrich Company Ltd., UK
D(+)-Glucose Anhydrous	G/0500/53	Fisher Scientific, Thermo Fisher Scientific, UK
D-(+)-Mannose	M2069	Sigma-Aldrich Company Ltd., UK
DL-Dithiothreiol (DTT)	D0632	Sigma-Aldrich Company Ltd., UK
Ethanol	32221	Sigma-Aldrich Company Ltd., UK
Ethylene-bis(oxyethylenenitrilo)tetraacetic acid (EGTA)	E3889	Sigma-Aldrich Company Ltd., UK
Ethylenediaminetetraacetic acid (EDTA)	25,235-2	Sigma-Aldrich Company Ltd., UK
Sulfo-N-hydroxysuccinimide (NHS)-LC-Biotin, EZ-linked	21335	Thermo Scientific, Thermo Fisher Scientific, UK
Glycerol 99 %	158920010	Acros Organics, Thermo Fisher Scientific, UK
Glycine	G/0800/60	Fluka Honeywell, Honeywell Specialty Chemicals Seelze GmbH, Germany
Glycogen from bovine liver, Type IX	G0885	Sigma-Aldrich Company Ltd., UK
Guanidine-HCl	G3272	Sigma-Aldrich Company Ltd., UK
HEPES sodium	215001000	Acros Organics, Thermo Fisher Scientific, UK

Table 1: Chemical reagents (continued)

Reagent	Product Code	Manufacturer
HEPES	H4034	Sigma-Aldrich Company Ltd., UK
Isopropyl β-D-thiogalactoside (IPTG)	I6758	Sigma-Aldrich Company Ltd., UK
Lysogeny broth (LB)	L3022	Sigma-Aldrich Company Ltd., UK
Lysozyme	62971	Sigma-Aldrich Company Ltd., UK
Magnesium chloride	M/0600/53	Fisher Scientific, Thermo Fisher Scientific, UK
N-(3-Dimethylaminopropyl)-N′-ethylcarbodiimide hydrochloride (EDAC)	E7750	Sigma-Aldrich Company Ltd., UK
N-Hydroxysuccinimide (NHS)	130672	Sigma-Aldrich Company Ltd., UK
Sodium Chloride	S5886	Sigma-Aldrich Company Ltd., UK
Sodium deoxycholate	D6750	Sigma-Aldrich Company Ltd., UK
Sodium hydroxide	S8045	Sigma-Aldrich Company Ltd., UK
TEMED	T-9281	Sigma-Aldrich Company Ltd., UK
Tris(hydroxymethyl)aminomethane (TRIS)	BP152-1	Fisher Scientific, Thermo Fisher Scientific, UK
Triton X-100	X100	Sigma-Aldrich Company Ltd., UK
Tween-20	P1379	Sigma-Aldrich Company Ltd., UK
Urea	ZU10001	Invitrogen, Thermo Fisher Scientific, UK

2.1.2 Materials and reagents

Table 2: Materials, consumables and other reagents

Reagent/Material/Consumable	Product Code/ size	Manufacturer
1-Step Ultra TMB-Blotting Solution	37574	Thermo Scientific, Thermo Fisher Scientific, UK
Biosensor, Amine Reactive Second Generation (AR2G)	18-5092	Pall ForteBio Europe, UK
Biosensor, Streptavidin (SA)	18-5019	Pall ForteBio Europe, UK
Deoxyribonuclease I	18047-019	Invitrogen, Thermo Fisher Scientific, UK

Table 2: Materials, consumables and other reagents (continued)

Reagent/Material/Consumable	Product Code/ size		Manufacturer
Dialysis Device, Slide-A-Lyzer MINI, 10K MWCO, 0.1 mL	69570		Thermo Scientific, Thermo Fisher Scientific, UK
Dialysis Device, Slide-A-Lyzer MINI, 3.5K MWCO, 0.5 mL	88400		Thermo Scientific, Thermo Fisher Scientific, UK
Dialysis membrane, VISKING, 12K-14K MWCO	DTV.12000.11		Medicell Membranes Ltd., UK
Dialysis membrane, VISKING, 3.5K MWCO	DTV.03500.04		Medicell Membranes Ltd., UK
Eppendorf reaction tubes	0.5/1.5 mL		various
Halt Protease and Phosphatase Inhibitor Cocktail (100X)	78440		Thermo Scientific, Thermo Fisher Scientific, UK
InstantBlue Protein Stain	ISB1L		Expedeon Ltd., UK
Membrane filter Durapore, pore size 0.45 µm	HVLP04700		Millipore (U.K.) Ltd., UK
NuPAGE Novex 10% Bis-Tris Midi Protein Gels, 26 well	WG1203BOX		Invitrogen, Thermo Fisher Scientific, UK
Needles, single use	orange	25G × 5⁄8″ (0.5 × 16 mm)	Becton, Dickinson U.K. Limited, UK
	blue	23G × 1″ (0.6 × 25 mm)	Becton, Dickinson U.K. Limited, UK
	green	21G × 1 1⁄2″ (0.8 × 40 mm)	Becton, Dickinson U.K. Limited, UK
Pipettes tips	20/200/1000 µL		various
Poly-Prep Chromatography Columns	731-1550		Bio-Rad Laboratories Ltd., UK
Precast Gels, SDS 12 Well 10x10, SDS-PAGE	NXG41212K		Expedeon Ltd., UK
Proteoloc Protease Inhibitor Cocktail	44202		Expedeon Ltd., UK
ProteOn 1 M Ethanolamine HCl	1762450		Bio-Rad Laboratories Ltd., UK
ProteOn EDAC	10004488		Bio-Rad Laboratories Ltd., UK
ProteON sulfo-NHS	10004489		Bio-Rad Laboratories Ltd., UK

Table 2: Materials, consumables and other reagents (continued)

Reagent/Material/Consumable	Product Code/ size	Manufacturer
RunBlue LDS Sample Buffer 4X	NXB31010	Expedeon Ltd., UK
RunBlue Tris Glycine SDS Run Buffer	NXB82500	Expedeon Ltd., UK
Sepharose 6B, bead-formed agarose-based gel filtration matrix	17011001	GE Life Sciences, GE Medical Systems Ltd, UK
Serological pipettes	5/10/25 mL	Greiner Bio-One Ltd., UK
Syringe	1/2/10/50 mL	Becton, Dickinson U.K. Limited, UK
Syringe filter Minisart, pore size 0.2 µm	16534	Sartorius UK Ltd., UK

2.1.3 Buffers and solutions

All buffers and solutions are prepared in double distilled, filtered water (ddH_2O).

Column Washing Buffer pH 7.4

10 mM HEPES, 150 mM NaCl, 5 mM $CaCl_2$

Column Eluting Buffer pH 7.4

10 mM HEPES, 150 mM NaCl, 5 mM EDTA

Dialysis Buffer

25 mM TRIS, 1.25 M NaCl

EDTA stock solution pH 8.0

0.5 M EDTA

EGTA stock solution pH 8.0

1 M EGTA

Loading Buffer pH 8.0

25 mM TRIS, 1.25 M NaCl, 25 mM $CaCl_2$

Lysis Buffer pH 8.0

25 mM TRIS, 150 mM NaCl, 1 mM EDTA, 0.5% Triton X-100, 0.5 mg/mL lysozyme

Lysogeny broth medium

20 g Lysogeny broth medium (Sigma-Aldrich) in 1 L double distilled water, autoclaved

Magnesium chloride stock solution

1 M $MgCl_2$

Resolubilisation/Guanidine Denaturing Buffer pH 7.8

6 M Guanidine-HCl, 100 mM TRIS, 0.01% 2-Mercaptoethanol (freshly added)

Resuspension Buffer 1 pH 8.0

25 mM TRIS, 150 mM NaCl, 1 mM EDTA, 0.5% Triton X-100

Resuspension Buffer 2 pH 8.0

25 mM TRIS, 500 mM NaCl, 1 M Urea, 1 g/L Sodium Deoxycholate

Sample Buffer, reducing

LDS Sample Buffer 4X (Expedeon PLC), 25 mM DTT (w/v)

Sodium chloride stock solution

1 M NaCl

Sodium hydroxide solution

1M NaOH

TRIS buffer pH 8.0

50 mM TRIS

TRIS stock solutions

1 M TRIS pH 8.0

50 mM TRIS pH 8.0

TRIS buffered saline (TBS) pH 8.0

10 mM TRIS, 150 mM NaCl

TRIS buffered saline with Tween-20 (TBST)

10 mM TRIS, 150 mM NaCl, 0.1 % Tween-20

2.1.4 Commercial proteins and glycoproteins for interaction studies

The (glyco)proteins of interest were purchased from commercial sources as stated below.

Table 3: Commercial proteins and glycoproteins used in interaction studies

Glycoprotein	Specifications	Product Code	Company
Proform of Eosinophil Major Basic Protein (proMBP), recombinant, human	• Alternative names: Bone Marrow Proteoglycan, (BMPG), Proteoglycan 2 (PRG2), Eosinophil Granule Major Basic Protein (MBP or EMBP), Pregnancy-Associated Major Basic Protein • Species: human • Source species: Human Cells • Sequence: Leu17-Tyr222 • Fusion tag: C-terminal 6 His-tag	91-647/ PRSI91-647	ProSci Inc., USA (distributed via VWR International Ltd., UK)
Pregnancy-associated plasma protein-A1 (PAPP-A1), homodimeric form, recombinant, human	• Alternative names: Pregnancy-associated plasma protein-A1/ Pappalysin-1 (PAPP-A1) • Species: human • Sequence: aa 82-1214 • Source species: Mouse myeloma cell line, NS0-derived • Carrier-free • Fusion tag: C-terminal 10 His-tag	2487-ZN-020	R&D Systems, Bio-Techne Ltd., UK

Table 3: Commercial proteins and glycoproteins used in interaction studies (continued)

Glycoprotein	• **Specifications**	**Product Code**	**Company**
Pregnancy-associated plasma protein-A1 (PAPP-A1), heterotetrameric form (htPAPP-A), native, human	• consisting of PAPP-A1 and pro-MBP subunits (2:2 PAPP-A1/proMBP complex) • Sequence: human, • Source species: human • purified from pooled human retro-placental blood • material was tested and found negative for HBsAg, HCV, HIV-1 and 2 antibodies, HCV antibodies, and syphilis	8P64	HyTest Ltd., Finland
Macrophage Mannose Receptor (MMR), recombinant, human	• Alternative names: Macrophage Mannose Receptor 1 (MMR), CD206, Human mannose receptor (hMR), Mannose Receptor C type 1 (MRC1), CLEC13D (C-type lectin domain family 13, member D • Species: human • Source species: Mouse myeloma cell line, NS0-derived • Sequence: Leu19-Lys1383 with Thr399Ala and Leu407Phe • Carrier-free • Fusion tag: C-terminal 6 His-tag	2534-MR-050/CF	R&D Systems, Inc., UK

2.1.5 Plasmids for recombinant protein expression

The generation of soluble, recombinant lectin proteins (see Table 4) was carried out in the *Escherichia coli* strain BL21 DE3, which expresses the T7 RNA polymerase.

The plasmids used for transformation of the bacteria were either the established pT5T vectors from the Mitchell lab (Eisenberg et al., 1990, Mitchell et al., 2001) (see Figure 2.1), or novel plasmids produced by the Protein Expression Laboratory (PROTEX) at the University of Leicester.

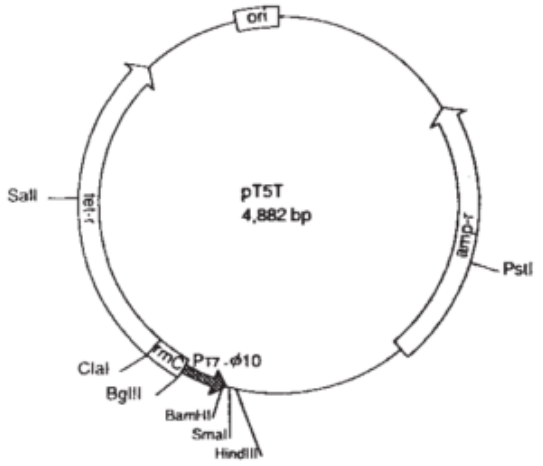


Figure 2.1: vector pT5T (adapted and altered from Eisenberg et al., 1990)

For generation of the novel bacterial expression vectors, the construct pLEICS-05 (see figure 2.2) was used, which is a member of the in-house vector family B and carries a resistance cassette for ampicillin as the antibiotic selection marker. Furthermore, the vector includes the Lac I operon which is required for an IPTG-induced expression of the protein of interest.

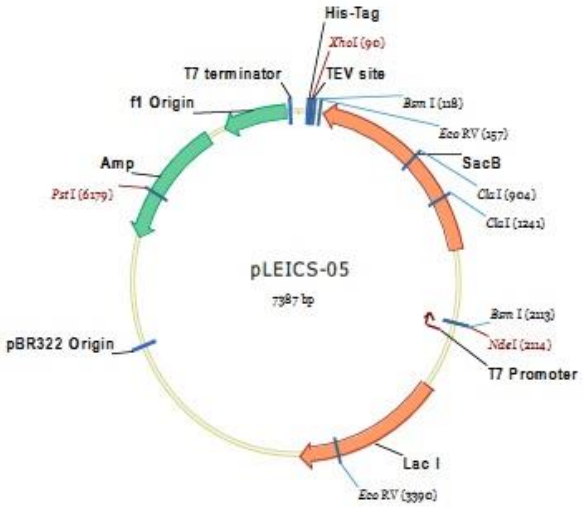


Figure 2.2: pLEICS-05 vector, family B (N-terminal genetic sequence AGGAGATATACATATG...; C-terminal genetic sequence GAAGTACAGG-TTCTCTCA...), provided by Protein Expression Laboratory (PROTEX), University of Leicester

The affinity chromatography, described in section 2.3.3.1, served as purification step as well as functionality test for the expressed lectin proteins.

Table 4: Recombinant lectin proteins expressed in transformed E. coli via the pLEICS-05 vector construct

Plasmid	Recombinant lectin	Encoded structures	Antibiotic resistance
DC020	DC-SIGN	Extracellular domain	Ampicillin
DSR034	DC-SIGNR	Extracellular domain	Ampicillin

2.2 Instruments and software for interaction studies

2.2.1 Instruments

Table 5: Laboratory apparatus and instruments

Instrument	Manufacturer
Balance CS200	OHAUS Europe GmbH, Switzerland
Balance PF-6001	Fisherbrand, Fisher Scientific, Thermo Fisher Scientific, UK
Balance PS-100	Fisherbrand, Fisher Scientific, Thermo Fisher Scientific, UK
BioLogic DuoFlow chromatography system, Fast protein liquid chromatography (FPLC) including BioLogic Maximizer, BioLogic QuadTec UV-Vis Detector, BioLogic BioFrac Fraction Collector	Bio-Rad Laboratories Ltd., UK
Bio-Silect SEC 250-5 Column	Bio-Rad Laboratories Ltd., UK
Block Heater, QBT4, Digital	Grant Instruments, UK
Centrifuge Micro Centaur	MSE (UK) Ltd, UK
Centrifuge Sorvall Evolution RC	Thermo Scientific, Thermo Fisher Scientific, UK
Centrifuge Sorvall Legend RT	Thermo Scientific, Thermo Fisher Scientific, UK
ChemiGenius2 Bio Imaging System	Syngene, Synoptics Ltd, UK
Incubator-shaker CERTOMAT BS-1	Sartorius Stedim UK Ltd., UK
NanoDrop ND-1000 UV-Vis Spectro-photometer	NanoDrop, Thermo Fisher Scientific, UK
pH meter electrode, Orion ROSS Ultra pH/ATC Triode	Thermo Scientific, Thermo Fisher Scientific, UK
pH meter, ORION STAR A211	Thermo Scientific, Thermo Fisher Scientific, UK
Pipet aid, PIPETBOY acu	INTEGRA Biosciences Ltd, UK
Pipettes, FINNPIPETTE 1-10/10-100/ 30-300/100-1000 µL	Thermo Scientific, Thermo Fisher Scientific, UK
Power supply, PowerEase 500	Invitrogen, Thermo Fisher Scientific, UK
Quartz cuvettes, Spectrosil Precision Cells (FAR UV Spectral Range)	Thermal Syndicate Ltd., UK
Roll mixer, Spiramix 5, MPR-455-020F/ DS507V	Thermo Scientific/Denley, Thermo Fisher Scientific, UK
Rotor for Sorvall Evolution RC, Aluminium Fixed Angle, SS-34	Thermo Scientific, Thermo Fisher Scientific, UK
Shake mixer, orbital shaker SSL1	Stuart, Cole-Parmer, UK
Sonicator 3000	Misonix Inc., USA

Table 5: Laboratory apparatus and instruments (continued)

Instrument	Manufacturer
Spectrophotometer 6505 UV/Vis	Jenway, Cole-Parmer, UK
Stirrer CB161	Stuart, Cole-Parmer, UK
Superose 12 FPLC, 17-5173-01	GE Life Sciences, GE Medical Systems Ltd, UK
Superose 6 FPLC, 17-5172-01	GE Life Sciences, GE Medical Systems Ltd, UK
Super-Q Plus water purification system with carbon filter, ion exchange and pore size filter Durapore 0.2 µm	Millipore (U.K.) Ltd., UK
UNO Q1 Column, #720-0001	Bio-Rad Laboratories Ltd., UK
Vortex mixer, VX100 Labnet	MO BIO Laboratories, Inc
Vortex-Genie 2	Scientific Industries, Inc., USA
XCell SureLock Mini-Cell	Invitrogen, Thermo Fisher Scientific, UK
XCell4 SureLock Midi-Cell	Invitrogen, Thermo Fisher Scientific, UK

Surface plasmon resonance (SPR) technique was performed with the ProteOn XPR36 Protein Interaction Array System (Bio-Rad Laboratories Ltd, UK). The instrument provides real-time label-free analysis of the specificity, affinity and kinetics of biomolecular interactions. The sensor chips provide six channels for ligand binding as well as perpendicular analyte flow-through that enables a simultaneous analysis of up to 36 interactions.

Bio-layer interferometry (BLI) was performed using the Octet RED96 instrument (Pall ForteBio Europe, UK), which provided the opportunity to run experiments with up to eight sensors simultaneously, via state-of-the-art dip-and-read protocol. Kinetic parameters and affinity analysis can be measured through a sequence of time-resolved analyte contact and subsequent wash steps.

2.2.2 Software

The ProteOn XPR36 System used the ProteOn Manager Software (Bio-Rad Laboratories Ltd., UK) for instrument control, experiment design, data collection, and data analysis.

The Octet RED 96 instrument obtained data with the Data Acquisition Software (version 9.0.0.49) and data analysis was performed using the Data Analysis Software (version 9.0.0.14), from Pall ForteBio Europe, respectively.

The ChemiGenius2 Bio Imaging System (Syngene, Synoptics Ltd, UK) used the GeneSnap 6.08.04 Software as well as GeneTools 4.00.00 Software for acquisition of gel pictures and analysis of data.

The BioLogic DuoFlow chromatography System (Bio-Rad Laboratories Ltd., UK) was controlled with the BioLogic DuoFlow Version 5.10 Build 2 software.

Software used for further data analysis and interpretation were OriginPro 2018b (OriginLab Corporation) as well as the Office 365 programmes (Microsoft).

Furthermore, EndNote X9 (Clarivate Analytics) was used for management of literature references.

2.3 Methods for interaction studies

2.3.1 Generation of transfected E. coli cultures

The generation of novel transformed *E. coli* was carried out in the laboratories of the collaborating research group of Prof Russell Wallis (University of Leicester, United Kingdom) following the procedure described by Furze (2013):

Prior to the transformation, cells of the *E. coli* strain BL21 DE3 were treated with calcium chloride to induce competence. The cells were grown in Lysogeny Broth (LB) medium to an OD_{590} between 0.6-0.8, decanted, and subsequently cooled on ice for 10 min, followed by centrifugation at 3500 rpm, 4°C for 10 minutes. The supernatant was discarded, and the pellet resuspended in 2.5 ml of ice cold 0.1 M $CaCl_2$ solution. The cells were incubated for 20 minutes (on ice) and centrifuged at 3500 rpm, 4°C for 10 minutes. The pellet was resuspended in 5 ml of ice cold 0.1 M $CaCl_2$ with 15% glycerol (v/v) and 100 µL aliquots prepared. The aliquots were frozen on dry-ice and stored at -80°C.

The frozen cell aliquots were thawed on ice, and transformed with the plasmid of interest (see section 2.1.5) by adding 2 µL of ligation reaction to each tube. A heat-shock at 42°C was applied for 60-90 s and the tubes incubated on ice for 15 min. To each tube, 300 µL of LB medium was added and the cultures incubated for 60-90 min at 37 °C whilst shaking until the culture reached an OD_{590}=0.6. 150 µL of the cell suspension was plated onto ampicillin agar plates via the streak plate method. The plates were incubated at 37 °C overnight and stored at 4 °C afterwards.

2.3.2 Generation and extraction of soluble lectins in the bacterial expression system E. coli

2.3.2.1 Generation of pre-cultures

The pre-cultures were set up by picking a single colony from the agar plate that contained ampicillin as selection factor. The colony was diluted in approximately 5 mL LB medium (containing 50 µg/mL ampicillin) in a 15 mL falcon tube and was incubated in the shaking incubator (150 rpm) for 3-4 hours at 37 °C. When the pre-culture reached an OD_{590}=0.8, glycerol was added to a final concentration of 15% (v/v) and the suspension was mixed thoroughly. The pre-cultures were frozen and stored at -80 °C.

2.3.2.2 Lysogeny Broth (LB) medium expression

Starter cultures were set up from the pre-culture by inoculating 200 mL LB medium (containing 50 µg/mL ampicillin as final concentration) with a small piece of the frozen pre-culture. The starter cultures were grown over night (approximately 15 h) in the incubator-shaker at 200 rpm and 37 °C.

The main cultures were set up by inoculating two 2 L Erlenmeyer baffle flasks with 50 mL starter culture each. Both Erlenmeyer flasks contained 1 L LB medium and 50 µg/mL freshly added antibiotic. The main cultures of the transfected *E. coli* were grown at 37 °C in the shaking incubator (200 rpm) until they reached an absorbance of at least 0.7 at 550 nm. The protein expression was induced by addition of isopropyl-

β-D-1-thiogalacto-pyranoside (IPTG, final concentration 100 µg/L) and the main cultures were incubated for another 150 min at 37 °C with shaking.

The bacteria were harvested by centrifugation at 3,440 x g for 30 min at 4 °C. The cells were washed in 40 mL Washing Buffer pH 8.0 and centrifuged again. The pellet was either processed immediately or frozen at -20 °C.

2.3.2.3 Extraction and purification of recombinant proteins

Either the fresh or thawed pellet was resuspended in approximately 25 mL Lysis buffer pH 8.0. After 20 min of incubation at room temperature, 5 mM $MgCl_2$ and 4 µL of DNAse I (277.3 U/µL) were added to the lysing cell suspension and incubated for 15 min at room temperature during shaking. Finally, the cells were completely lysed by sonication (10 bursts of 30 s duration each, approximately 70 W) with breaks for chilling. The suspension was treated with protease inhibitors and centrifuged at 20,000 x g for 20 min at 4 °C (applicable to all following centrifugation steps if not stated otherwise). The pellet containing the inclusion bodies was purified by resuspension with 40 mL Resuspension Buffer 1 pH 8.0, followed by a centrifugation step. For further refinement, the pellet was resuspended in 40 mL Resuspension Buffer 2 pH 8.0 and the suspension was centrifuged. The last purification step of the pellet containing the inclusion bodies was carried out by a wash with 50 mM TRIS buffer pH 8.0. The suspension was centrifuged to obtain a smooth, consistent pellet with inclusion bodies containing the protein of interest.

After discarding the supernatant, the inclusion bodies were solubilised in 15 mL 6 M Guanidine Denaturing Buffer pH 7.8 per litre of harvested medium at the beginning of the extraction. The mixture was gently rotated for 60 min at 4 °C, sonicated on ice for 30 sec, and 0.01% 2-mercaptoethanol added. The mixture was incubated at 4 °C on the roller shaker. After 30 min, the suspension was centrifuged at 50,000 x g for 40 min at 4 °C.

The supernatant was added drop-wise at 4 °C, through a 23-gauge needle into Loading Buffer to treble the volume whilst stirring. The diluted protein mixture was dialysed (11-14 kDa M_W cut-off) against 10 volumes of Dialysis Buffer with a minimum of 2 buffer changes over 48 h. After dialysis, the insoluble precipitate was

removed by centrifugation at 50,000 x g for 30 min at 4 °C. Protease inhibitor cocktail was added to the supernatant, which contained the refolded protein of interest, and was carefully transferred into a new vial.

2.3.3 Purification of lectins by chromatography

2.3.3.1 Affinity chromatography

Purification of the soluble recombinant C-type lectin receptors DC-SIGN and DC-SIGNR was performed by affinity chromatography (Axen et al., 1967). Given the affinity for carbohydrates of the carbohydrate recognition domain of the receptors, the purification was carried out using sugar-derivatised resins, specifically mannose-derivatised Sepharose 6B matrix. The mannose-Sepharose matrix was kindly provided by Prof Russell Wallis, University of Leicester.

Prior to further purification steps, $CaCl_2$ was added up to a final concentration of 25 mM to the protein solution, since the proteins of interest were C-type lectins (see

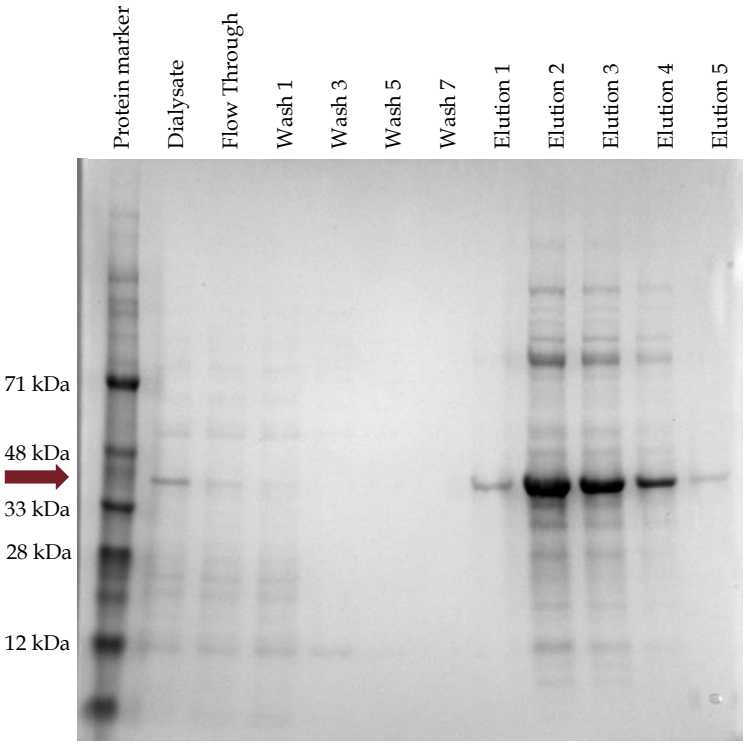


Figure 2.3: Representative SDS-PAGE gel after DC-SIGN purification via mannose-Sepharose affinity chromatography. The arrow shows the molecular weight that is expected for the monomeric recombinant DC-SIGN extracellular domain (38 kDa).

section '2.3.2.3 Extraction and purification of recombinant proteins'). Eight millilitres of mannose-Sepharose matrix were equilibrated with Column Washing Buffer, added to the protein solution (approximately 180 mL; derived from 2 L culture medium) and gently stirred at 4 °C overnight. The mannose-Sepharose matrix was loaded onto a disposable gravity column, washed initially with 10 mL Column Washing Buffer, followed by 5 fractions of 1 mL Column Washing Buffer, and eluted with Column Eluting Buffer (containing EDTA) in 0.5-1 mL fractions. The absorbance of the fractions at 280 nm was detected via spectrophotometry using either a quartz cuvette or the NanoDrop method.

The fractions were further analysed by SDS polyacrylamide gel electrophoresis (SDS PAGE) under denaturing and reducing conditions. A representative gel of the fractions is shown in figure 2.3.

2.3.3.2 Ion exchange chromatography and size exclusion Chromatography

Soluble recombinant lectins were purified via fast protein liquid chromatography (FPLC).

Size exclusion chromatography (SEC) was performed using either a GE Healthcare Superose 12 FPLC column or a Bio-Rad Bio-Silect SEC 250-5 column (flow rate 0.5 mL/min, fraction size 1 mL, buffer 10 mM TRIS 5 mM EDTA pH 7.8); where a 0.5 mL protein sample was applied to the column in isocratic flow mode within the BioLogic DuoFlow chromatography system.

After analysis of the obtained fractions via SDS PAGE gel, the pooled protein solution was loaded onto a Bio-Rad UNO Q anion-exchange column and run at a flow rate of 1 mL/min (fraction size 1 mL). The column was eluted using a 0 mM-500 mM NaCl gradient (buffer A 10 mM TRIS pH 7.8; buffer B 10 mM TRIS 1 M NaCl pH 7.8) over 20 column volumes.

2.3.4 Biotinylation of proteins

Sulfo-N-hydroxysuccinimide (sNHS) biotin solution was prepared as a 10 mM solution in ddH_2O at a neutral pH around 7.

The 10 mM biotin solution was added in a 5:1 molar ratio of biotin:protein to the protein solution (concentration range 0.3-3 mg/mL; buffer 10 mM HEPES, 150 mM NaCl) to start the biotinylation reaction.

The mixture was incubated on ice for 2 hrs and the reaction was stopped by adding 1 M ethanolamine pH 8.0 up to a final concentration of 100 mM. The biotinylated protein was subsequently desalted via overnight dialysis.

The recovered samples were stored at 4 °C and used for SPR experiments with NLC sensor chips (Neutravidin chips) or BLI experiments with SA (streptavidin) biosensors.

2.3.5 Surface Plasmon Resonance (SPR)

The technique of Surface Plasmon Resonance (SPR) is an optical method to detect changes of the refractive index of incident light on the surface of a sensor in relation to changes in mass in real time. In particular, the ligand is immobilized on the surface of the sensor and the analyte flows through the cell over the ligand. In the event of an interaction between the ligand and the analyte, a change of the mass on the surface occurs which shifts the refractive index. The difference between the sensorgrams of the refractive index over time is measured as response units (RU) (Lottspeich, 2006).

The SPR technique provides real-time, partially label-free analysis of the specificity, affinity and kinetics of biomolecular interactions. The sensor chips have 6 channels for ligand binding as well as perpendicular analyte flow-through that enables a simultaneous analysis of up to 36 interactions.

Recombinant soluble lectins were immobilised as ligands on sensor chips and purified native human glycoproteins, synthetic glycopolymers or other potential interaction partners served as analytes in the solution phase. Depending on the set-up of the experiment, a blank channel for monitoring of system drifts as well as a channel with a negative control for non-specific binding were included. The polymers were analysed as 'heterogeneous ligand' regarding fitting models.

2.3.5.1 GLM chip

The GLM sensor chip is an amine-coupling based chip which can be activated by reagents such as EDAC and sulfo-NHS to react specifically with free surface amines of proteins. The deactivation step is achieved using 1 M ethanolamine pH 8.0. After the activation of the surface, the amount of the injected ligand controls the density of ligand on the surface. The GLM chip has thickened surface layer of coupling matrix with a medium capacity (~12 kRU) and is ideal for probing both, protein-protein as well as protein-small molecule interactions.

2.3.5.2 NLC chip

The NLC sensor chip is designed for site-specific capturing of biotinylated targets via NeutrAvidin molecules which is a low nonspecific-binding streptavidin derivative. The chip is used for capturing biotinylated proteins or peptides and has the capacity to capture ~2 kRU of ligand. The NLC sensor chip is ideal for immobilizing ligands without amine-coupling but requires a modification of the ligand with biotin prior to immobilisation (see section 2.3.6).

2.3.6 Biolayer-Interferometry (BLI)

Bio-Layer Interferometry (BLI) is an optical analytical method, which is closely related to the SPR technique (see section '2.3.5 Surface Plasmon Resonance (SPR)'). It provides complementary data to the SPR approach since it is used to perform real-time interaction studies with a high sensitivity for biomolecular interactions.

The underlying physical phenomenon is the interference pattern of white light caused by reflection from a biosensor-internal surface (λ1) and the tip of the biosensor with the immobilised macromolecule of interest (λ2). The interaction between the immobilised ligand and the analyte during association results in a shift of the wavelength λ2. The resulting change in the interference pattern Δλ reveals information about the putative interaction between ligand and analyte (see figure 2.4 adapted from ForteBio Molecular Devices (2019a)). Furthermore, kinetic screenings via different concentrations of the analyte of interest can be performed and the reaction rates as well as reaction constant determined.

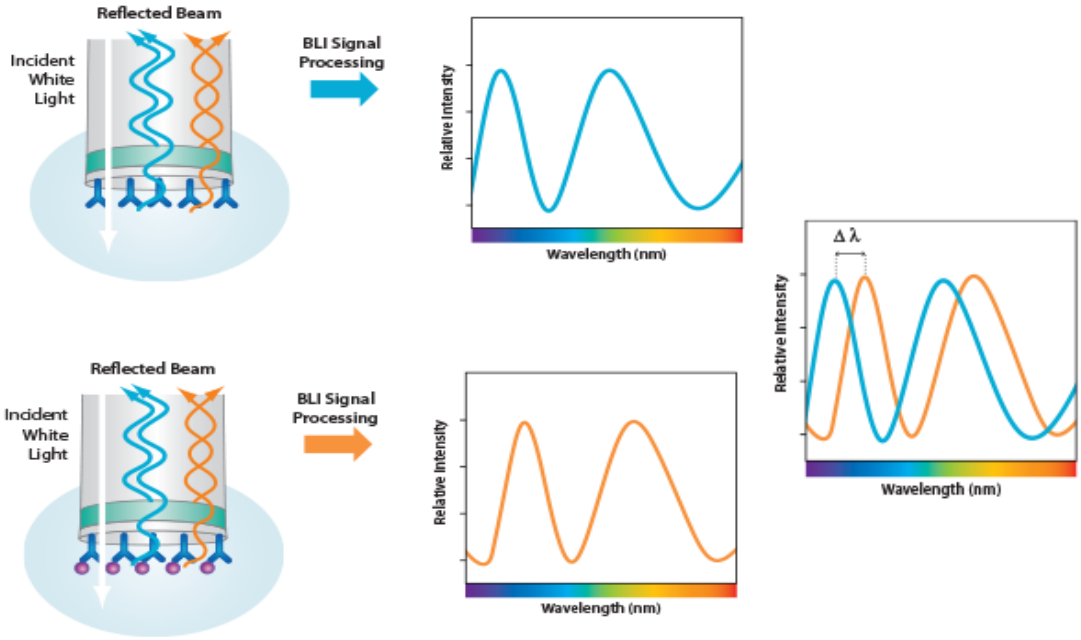


Figure 2.4: Scheme of the BLI technology, which measures a change of the reflected light's wavelength (Δλ) between the two biosensor surfaces. [Adapted from ForteBio Molecular Devices (2019a).]

The instrument used for the BLI experiments was the Octet RED96 system (Pall ForteBio Europe, UK), which allows for real-time analysis of the specificity, affinity and kinetics of biomolecular interactions. Access to the instrument was kindly provided by Prof Matthew Gibson, Department of Chemistry, University of Warwick.

Either the soluble lectins or the glycoproteins of interest were immobilised on biosensor probes from Pall ForteBio Europe. Two types of "Dip and Read" biosensors were used for the experiments: Streptavidin biosensors or amine-coupling biosensors.

Biosensors were hydrated in assay buffer (10 mM HEPES pH 7.4, 150 mM NaCl) for 10 minutes prior to the experiment, which included loading of the ligand of interest, setting the baseline, as well as association phase and dissociation phase of the analyte of interest, respectively.

Depending on the set-up of the experiment, a sensor with immobilised ligand but without analyte in the association phase (reference well) for monitoring of a system drift as well as a sensor with a negative control for non-specific binding were

included. All glycoproteins were categorised as 'heterogeneous analyte' for the kinetic fitting model.

2.3.6.1 Streptavidin (SA) sensors

The streptavidin (SA) biosensors (Pall ForteBio, UK) make use of the high affinity interaction between streptavidin and biotin for immobilisation of the biomacromolecule of interest on the tip of the biosensor. The sensor tip comes pre-coated with streptavidin and the biomacromolecule of interest is modified with an attachment of biotin units, referred to as biotinylation (see section '2.3.4. Biotinylation of Proteins').

After pre-hydration of the sensors in assay buffer for 10 minutes, the biosensor was dipped into the solution containing the biotinylated ligand for 120 s. The interaction between biotin and streptavidin is considered one of the strongest non-covalent interactions with a dissociation constant K_D between 10^{-16} and 10^{-14} M (Laitinen et al., 2006). However, any non-specifically bound molecule was removed through a washing step of 60 s.

The high affinity interaction between biotin and streptavidin, which immobilises the ligand on the tip of the biosensor, provides a stable baseline for interaction experiments including kinetic analyses with the analytes of interest.

2.3.6.2 Amine-coupling (AR2G) sensors

The Amine Reactive 2nd Generation (AR2G) biosensor type (Pall ForteBio Europe, UK) is an amine-coupling sensor, which requires activation of the surface via reagents such as EDC and sulfo-NHS to react specifically with free surface amines of biomacromolecules, especially proteins.

After pre-hydration of the sensors in assay buffer for at least 10 minutes and activation of the biosensor surface, the tip was dipped into the solution containing the ligand of interest for 120 s. The pH of the buffer used during immobilisation significantly influences the amount of protein binding to the biosensor surface. Therefore, a pH range was tested prior to the experiment. The pH can also affect the protein of interest and, thus, functionality was tested. In summary, a pH for the buffer

was chosen that facilitated a good binding response but did not influence the functionality of the immobilised ligand.

The deactivation of the surface was achieved by using 0.1 M ethanolamine at pH 8.0, also referred to as quenching.

The covalent binding between the biosensor tip and amines present in the protein of interest provides a stable baseline for interaction experiments and allows for kinetic analyses of the analytes of interest.

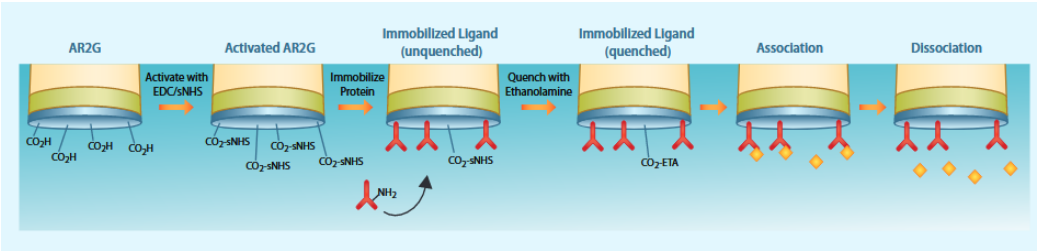


Figure 2.5: Schematic overview of the process for covalent immobilisation of the ligand of interest on AR2G biosensors as well as representation of the association and dissociation phase for interaction experiments (adapted from (ForteBio Molecular Devices, 2019b)).

2.4 Materials for mass spectrometry

2.4.1 Chemical reagents

Table 6: Chemical reagents

Reagent	Product Code	Manufacturer
2,5-Dihydroxybenzoic acid	149357	Sigma-Aldrich Corp., USA
Acetic acid	338826	Sigma-Aldrich Corp., USA
Acetone	650501-M	Sigma-Aldrich Corp., USA
Acetonitrile	34998-M	Sigma-Aldrich Corp., USA
Ammonium bicarbonate	40867	Honeywell Fluka, Honeywell International, Inc., USA
Chymotrypsin, Sequencing Grade	V1061	Promega Corp., USA
Dichloromethane	650463	Sigma-Aldrich Corp., USA
Dimethyl sulfoxide (DMSO)	34869	Sigma-Aldrich Corp., USA
DL-Dithiothreitol	43815	Sigma-Aldrich Corp., USA
Formic acid	56302	Honeywell Fluka, Honeywell International, Inc., USA
Iodoacetamide	I1149	Sigma-Aldrich Corp., USA
Iodomethane	I8507	Sigma-Aldrich Corp., USA
Isopropanol, 2-Propanol	650447	Sigma-Aldrich Corp., USA
Methanol	900688	Sigma-Aldrich Corp., USA
Nitrogen, Ultra High Purity 5.0 Grade	NI UHP300	-
PNGase F, glycerol-free, 500,000 units/ml, size 75,000 units	P0705L	New England Biolabs, Inc., USA
Sodium borohydride	213462	Sigma-Aldrich Corp., USA
Sodium Hydroxide, 50%(w/w) solution	3727-01	J.T.Baker, Avantor, Inc., USA
Trypsin, sequencing grade modified	V511A	Promega Corp., USA

2.4.2 Material and reagents

Table 7: Material, consumables and other reagents

Reagent/Material/Consumable	Product Code	Manufacturer
AG 50W-X8 Cation Exchange Resin (DOWEX), analytical grade, 100–200 mesh, hydrogen form, 500 g	1421441	Bio-Rad Laboratories, Inc., USA
Autosampler Inserts, 9mm, glass	C4010-627L	Thermo Scientific, Thermo Fisher Scientific Inc., USA
C18 SPE Cartridges, 1 mL, 100 mg	26030	Restek Corporation, USA
Calibrated Micropipettes, 1-5 µL	2-000-001	Drummond Scientific
Delicate Task Wipers, Premium 1 ply small	07-301- 005	Fisherbrand, Fisher Scientific Co. L.L.C., USA
Disposable Borosilicate Glass Pasteur Pipets, 14.6 cm/22.86 cm	13-678-20D/ 13-678-20B	Fisherbrand, Fisher Scientific Co. L.L.C., USA
Freeze dry glassware, Complete Fast-Freeze Flask	Various sizes	Labconco Corporation, USA
Hamilton GASTIGHT 1700 series syringe, volume 25 µL/500 µL	Model 1702N /1750N	Hamilton Company, USA
LTQ Velos ESI Positive Ion Calibration Solution	88323	Pierce, Thermo Fisher Scientific Inc., USA
National 9mm Screw Thread Cap, 9mm, Solid Top Cap	C4000-99	Thermo Scientific, Thermo Fisher Scientific Inc., USA
National Target DP Vials, 9mm, volume 300 µL, Plastic Screw Thread	C4000-11	Thermo Scientific, Thermo Fisher Scientific Inc., USA
Negative Ion Calibration Solution	88324	Pierce, Thermo Scientific, Thermo Fisher Scientific Inc., USA,
Opti-TOF plate, 384 well, 123 x 81 mm, 1 insert	P/N 1016491	Applied Biosystems Inc. USA/ AB SCIEX, USA

Table 7: Material, consumables and other reagents (continued)

Reagent/Material/Consumable	Product Code	Manufacturer
Pipettes tips 20/200/1000 µL	various	various
Poly-Prep Chromatography Columns	731-1550	Bio-Rad Laboratories, Inc., USA
PYREX 10 mL Sterile Glass Serological Pipets, plugged, sterile, disposable	7077-10N	Corning Inc., USA
Screwcap Glass Culture Tubes, Pyrex, 13 x 100mm Round Bottom	99449-13	Corning Inc., USA
Screwcaps with White Rubber Liner, Disposable Phenolic, 13-415 GPI Thread Finish	99999-13	Corning Inc., USA
Screw-Cap Microcentrifuge Tubes	89004-304	VWR Inter-national LLC, Avantor Inc., USA
Spatula, Disposable	smartSpatula macro/micro	LevGo, Inc., USA
Tube-O-DIALYZER, Micro, 4,000 Da MWCO	786-611	G-Biosciences, Geno Technology Inc., USA
Universal Indicator paper pH 1-11	37035	Sigma-Aldrich Corp., USA

2.4.3 Buffers and solutions

All buffers and solutions are prepared in double distilled water (ddH_2O) if not stated otherwise.

50 mM Ammonium bicarbonate pH 8.0-8.5

0.1 % Formic acid (v/v)

5 % Acetic acid (v/v)

10 % Acetic acid (v/v)

20 % Isopropanol in 5 % Acetic acid (v/v)

40 % Isopropanol in 5 % Acetic acid (v/v)

100 % Isopropanol (v/v)

Methanol:Acetic acid (9:1) (v/v)

2.4.4 Glycoproteins used for mass spectrometry

The glycoproteins of interest were purchased from commercial sources as stated below.

Table 8: Commercially available glycoproteins for analysis with mass spectrometry technique

Glycoprotein	Specifications	Product Code	Company
Proform of Eosinophil Major Basic Protein (proMBP), recombinant, human	• Alternative names: Bone Marrow Proteoglycan, (BMPG), Proteoglycan 2 (PRG2), Eosinophil Granule Major Basic Protein (MBP or EMBP), Pregnancy-Associated Major Basic Protein • Species: human • Source species: Human Cells • Sequence: Leu17-Tyr222	91-647/ PRSI91-647	ProSci Inc., USA (distributed via VWR International Ltd., UK)
Pregnancy-associated plasma protein-A1 (PAPP-A1), homodimeric form, recombinant, human	• Alternative names: Pregnancy-associated plasma protein-A1 /Pappalysin-1 (PAPP-A1) • Species: human • Sequence: aa 82-1214 • Source species: Mouse myeloma cell line, NS0-derived • Carrier-free	2487-ZN-020	R&D Systems, Bio-Techne Ltd., UK
Pregnancy-associated plasma protein-A1 (PAPP-A1), heterotetrameric form (htPAPP-A), native, human	• consisting of PAPP-A1 and pro-MBP subunits (PAPP-A1/proMBP complex) • Sequence: human, • Source species: human • purified from pooled human retro-placental blood • material was tested and found negative for HBsAg, HCV, HIV-1 and 2 antibodies, HCV antibodies, and syphilis	8P64	HyTest Ltd., Finland

2.5 Instruments and software for mass spectrometry

2.5.1 General instruments

Table 9: Laboratory apparatus and instruments

Instrument	Company
Blocks Heaters, various	VWR Scientific, VWR International LLC , USA
Freeze Dryer BenchTop, K Series	VirTis, SP Scientific, SP Industries, Inc., USA
LSE Mini Microcentrifuge 6765	Corning, Sigma-Aldrich Corp., USA
Nanopure Diamond water purifier with Hollow Fiber 0.2µm Final Filter	Barnstead Water Purification Systems Thermo Scientific, Thermo Fisher Scientific Inc., USA
Nitrogen evaporator system	Not applicable/ custom-made
Savant USV400 Universal SpeedVac Vacuum System with Savant SPD131DDA SpeedVac Concentrator	Thermo Scientific, Thermo Fisher Scientific Inc., USA
Vortex-Genie 2	Scientific Industries Inc., USA

2.5.2 Mass spectrometers and software for mass spectrometry

Two techniques of mass spectrometry were used within the scope of the structural analysis of the glycoproteins of interest.

Matrix-assisted laser desorption ionisation-time-of-flight mass spectrometry (MALDI-TOF MS) was performed on an AB SCIEX MALDI TOF/TOF 5800 (Applied Biosystems Inc., USA) mass spectrometer. All data was acquired in positive ion mode and the samples measured on an Opti-TOF sample plate. Data analysis was performed using AB SCIEX Analyst Software and Data Explorer.

Electrospray ionisation (ESI) mass spectrometry (MS) and tandem mass spectrometry was performed with a nano-electrospray ion (NSI) source at an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific Inc., USA) on-line to a prior nano-flow liquid chromatography (nanoLC) system, namely the UltiMate 3000 RSLCnano System (Thermo Scientific Thermo Fisher Scientific Inc., USA). The

nanoLC system used a pre-packed Acclim PepMap RSLC C18 column (75 µm x 15 cm, 3 µm particle size, Thermo Scientific, Thermo Fisher Scientific Inc., USA). The nanoLC and the MS systems were integrated via the Chromeleon Chromatography Data System (CDS) software from Thermo Scientific, Thermo Fisher Scientific Inc., USA.

Software used for further data analysis and interpretation were Xcalibur 3.0 (Thermo Scientific, Thermo Fisher Scientific Inc., USA), GlycoWorkbench 1.1 (Imperial College London, United Kingdom and University of Giessen, Germany), Byonic v2.3.5 (Protein Metrics Inc., USA). Furthermore, PeptideCutter (SIB Swiss Institute of Bioinformatics, Switzerland), ProteinProspector v5.22.1 (UCSF Mass Spectrometry Facility, University of California, San Francisco, USA), GlycoMod (SIB Swiss Institute of Bioinformatics, Switzerland) and NetNGlyc 1.0 (DTU Bioinformatics, Center for Biological Sequence Analysis, Technical University of Denmark, Denmark), which were available via the ExPASy tools page (http://www.expasy.org/tools).

2.6 Methods for mass spectrometry

Sample preparation, data acquisition, data analysis, as well as data interpretation were carried out according to the state-of-the-art procedures in the Analytical Service and Training Laboratory, Complex Carbohydrate Research Center at the University of Georgia, United States of America (Shajahan et al., 2017a, Shajahan et al., 2017b).

The commercially available glycoproteins had sufficient purity; certified through the data sheets that were provided by the manufacturers, and confirmed in-house by SDS-PAGE. Therefore, direct processing of the samples without additional purification steps was conducted.

2.6.1 Digestion of the glycoproteins

The analysis of the glycoproteins followed a bottom-up approach, i.e. the enzymatic digestion of the full-length glycoprotein into smaller glycopeptide units (Chait, 2006).

The glycoproteins of interest are known to form intra- and intermolecular disulfide bonds. Explicitly, PAPP-A1 homodimerises via one inter-chain disulfide bond and proMBP homodimerises via two inter-chain disulfide bonds. In addition for the

2:2 PAPP-A1/proMBP-complex, one PAPP-A1 unit links to another pro-MBP unit via two disulfide bonds. Therefore, the 2:2 PAPP-A1/proMBP-complex has been shown to contain a total of seven inter-polypeptide disulfide bonds (Overgaard et al., 2003).

a) In-solution digestion

For further analysis via Glycomics and Glycoproteomics, the glycoproteins required reduction of the disulfide bonds: Approximately 150 µg of the protein of interest was dissolved in 25 µL of 50 mM ammonium bicarbonate pH 7.5 and 25 µL of 25 mM dithiothreitol (DTT) added for reduction. The mixture was incubated at 45 ˚C for 45 min. This reaction was followed by a carbamidomethylation of the thiol groups with 25 µL of 90 mM iodoacetamide at room temperature for 45 min in the dark. If not broken apart via chemical reactions, intra- and intermolecular disulfide bonds limit the efficacy of the enzymatic digest of the glycoprotein into glycopeptides. Subsequently, the samples were dialysed in a 4 kDa cut-off dialyser tube against nanopure water at 4 ˚C overnight. After lyophilisation, the sample was redissolved in 50 mM ammonium bicarbonate buffer pH 7.5 and digested with at least one suitable protease that produces target peptides of ideal length, polarity and charge state for detection via mass spectrometry (Chen et al., 2009, Froehlich et al., 2013). Through analysis with the software tool PeptideCutter, trypsin was chosen as the protease for digestion of the recombinant proMBP. A sequential digest with the proteases trypsin and chymotrypsin was selected for the recombinant PAPP-A1. The tryptic digest was performed for 12-16 hours at 37 ˚C, whereas the proteolytic digest with chymotrypsin was incubated for 12-16 hours at room temperature. Each protease was used in a protease to glycoprotein molar ratio of 1:20. The proteolytic digestions were stopped by heat inactivation at 100 ˚C for 5 min and lyophilised afterwards. The prepared samples were used for glycan identification by Glycomics and Glycoproteomics.

b) In-gel digestion

This particular part of the research project was performed in the Analytical Service and Training Laboratory at the Complex Carbohydrate Research Center (University of Georgia, USA) as in-gel digest of the native 2:2 PAPP-A1/proMBP-complex. It was

carried out with assistance from the CCRC analysts Dr Ganapati Bhat and Dr Nitin Supekar.

The lyophilised sample was redissolved in 50 mM ammonium bicarbonate buffer pH 7.5 and loaded onto a 4-15% polyacrylamide gel electrophoresis (PAGE) gel under denaturing conditions. The gel was Coomassie-stained and the bands at 200 kDa as well as 75 kDa were isolated and processed separately. The gel bands were cut into pieces of approximately 1 mm^3, and de-stained by repeated washing with a buffer containing ammonium bicarbonate and aceto-nitrile. Reduction via DTT and alkylation by iodoacetamide of the thiol groups was performed in-gel. After removal of the denaturing chemicals through washes with ammonium bi-carbonate buffer, the 200 kDa and 75 kDa bands of the native 2:2 PAPP-A/proMBP-complex were digested with trypsin and chymotrypsin. The resulting peptides and glycopeptides were extracted with increasing proportions of acetonitrile acidified with formic acid. The proteases were heat inactivated at 100 ˚C for 5 min and subsequently lyophilised.

The prepared samples were used for glycomic and glycoproteomic analysis.

2.6.2 Glycoproteomics

After the proteolytic digest, the lyophilised glycopeptides were redissolved in nanopure water containing 0.1% formic acid and either directly analysed via nanoLC-MS/MS or stored at -30˚C until further analysis.

a) Recombinant proMBP/Recombinant PAPP-A1

For analysis, the sample was injected at 4 ˚C via the autosampler of the UltiMate 3000 RSLCnano System into a pre-loading C18 column that led the sample into the analytical nano-flow Acclim PepMap RSLC C18 column. The system was equilibrated in solvent A (100% water, 0.1% formic acid). The separation was realised by a 61 min long, linear elution gradient (table 10, figure 2.6) generated with an increase of 1-99% of solvent B (100% Acetonitrile, 0.08% formic acid) over a total of 72 min. The nano-flow liquid chromatography ensured the best separation of peptides and glycopeptides, including possible glycoforms, for on-line mass spectrometry (Gilar et al., 2011, Alley et al., 2013).

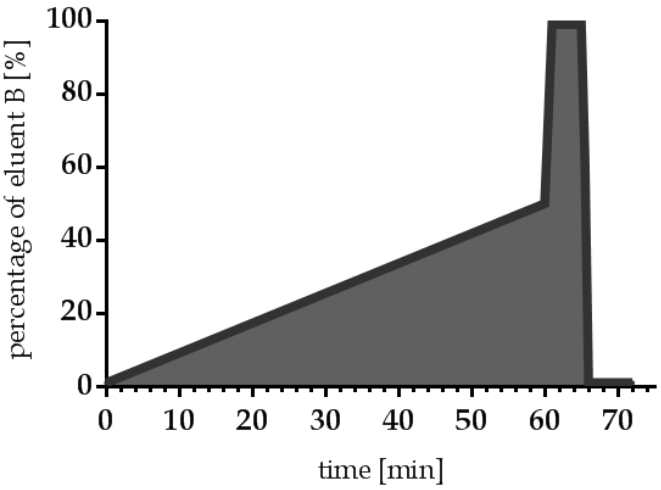


Figure 2.6: Elution gradient used for separation of the glycopeptides derived from recombinant proMBP and recombinant PAPP-A1, respectively. The final spike in solvent B ensures the elution of remains on the column.

Table 10: Elution gradient programme used on the Acclaim PepMap 100 C18 LC column for glycopeptides derived from recombinant proMBP and recombinant PAPP-A1, respectively.

Retention time [min]	**Flow rate [nl/min]**	**Percentage of solvent A (100% water, 0.08% formic acid) [%]**	**Percentage of solvent B (100% Acetonitrile, 0.08% formic acid) [%]**
0	300	99	1
0	300	99	1
60	300	50	50
61	300	1	99
65	300	1	99
65.5	300	37.5	62.5
66	300	99	1
72	300	99	1

The data acquisition of the glycopeptides was carried out on the Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific Inc., USA) with an Orbitrap as well as an Iontrap detector type available. The samples were sprayed in positive ion mode via a metal emitter tip using the method of nano-electrospray ionisation (NSI). NSI provides better sensitivity, lower sample consumption, as well as a higher compatibility with the method of nano-LC than conventional electrospray ionisation (ESI), due to a lower flow rate in comparison to ESI (Burlingame, 1996).

Full MS scans were acquired from m/z 200-2000 at a resolution of 120,000 of the Orbitrap analyser. The downstream method selected precursor ions at a time frame of 3 s for fragmentation using the criteria of monoisotopic precursor selection, charge states (included states 1-8, as well as undetermined), dynamic exclusion (exclusion duration of 60 s) and an intensity threshold of 50,000 A.U. The following decision tree of the method programme included, first, the highest charge state and, second, the most intense ion precursors for MS/MS analysis.
MS/MS scans were performed via fragmentation with the methods of collision-induced dissociation (CID) and higher-energy collisional dissociation (HCD). Fragment ions of HCD and CID were detected in the Orbitrap at 30,000 resolution.
A subsequent fragmentation was carried out using a HCD-product-triggered electron-transfer-dissociation (ETD) programme. The ETD method detected fragment ions via the Iontrap after detection of listed diagnostic ions, i.e. oxonium ions derived from glycan structures, produced via HCD MS/MS.
The complete workflow is depicted in figure 2.7.

The acquired raw spectra were searched against FASTA sequences of the glycoproteins using the software Byonic with the mass tolerance set as 10 ppm for both, full MS and MS/MS scans. For N-and O-glycopeptide analysis, carbamido-methylation as variable modification and oxidation of methionine as variable modification were used as search parameters. The assigned glycopeptides were initially filtered based on probability and further validated manually by evaluating glycan neutral loss pattern, oxonium ions, and the glycopeptide fragmentations.

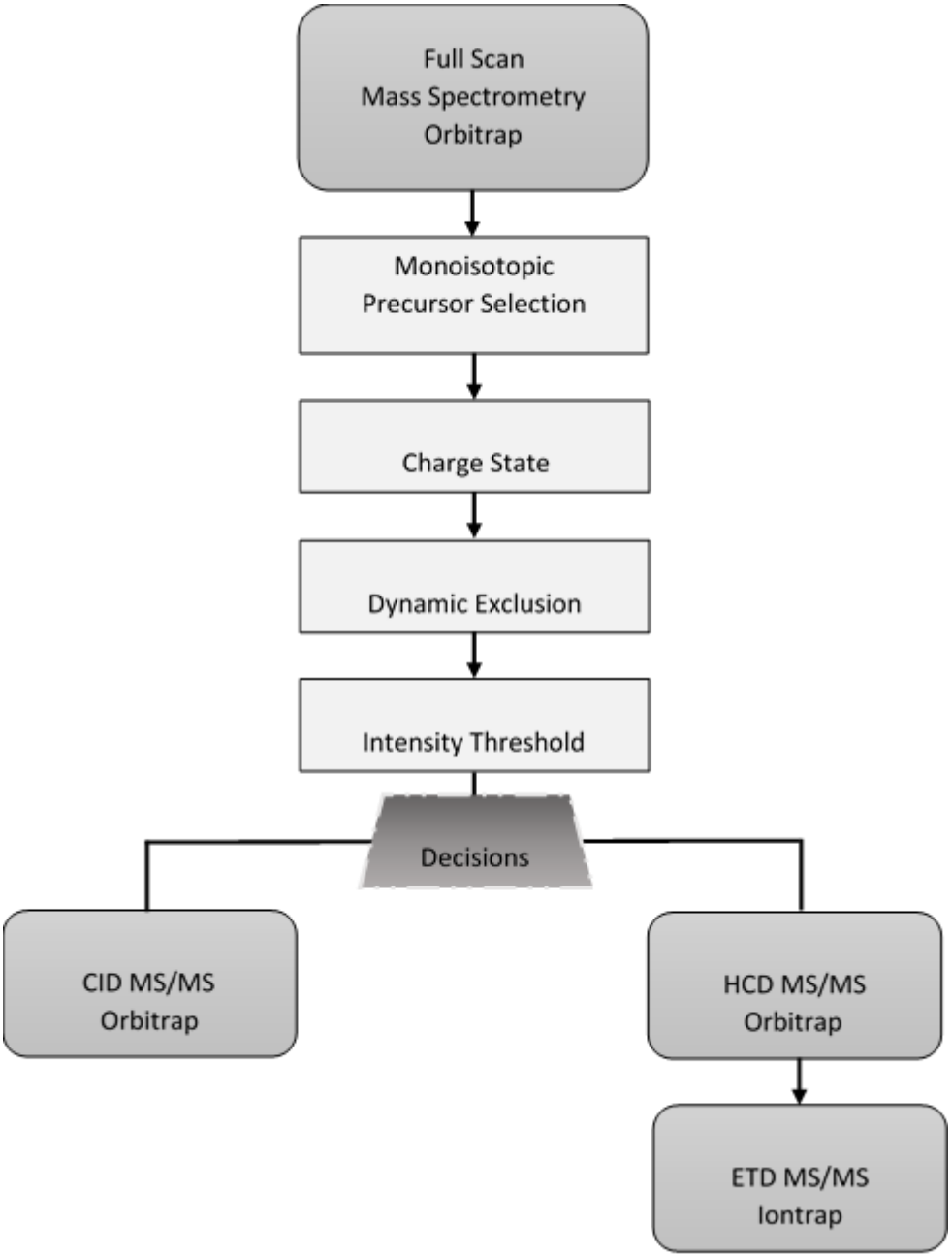


Figure 2.7: Schematic overview of the MS method used for on-line nano-LC-MS/MS analysis of the glycomic and glycoproteomic characteristics of the glycopeptides derived from recombinant proMBP and recombinant PAPP-A1, respectively.

b) Native PAPP-A1/proMBP-complex

The glycopeptide mix, derived from the native PAPP-A1/proMBP-complex, was analysed by the CCRC analysts Dr Ganapati Bhat and Dr Nitin Supekar in the Analytical Service and Training Laboratory at the Complex Carbohydrate Research Center (University of Georgia, USA).

The sample was injected via the autosampler of the UltiMate 3000 RSLCnano System at 4 ˚C. The pre-loading C18 column adjusted the sample size and loading speed to the on-line analytical nano-flow Acclim PepMap RSLC C18 column. The system was equilibrated in solvent A (100% water, 0.1% formic acid) and the separation of the sample was realised over a 55.5 min long elution gradient (table 11, Figure 2.8) generated through an increase by 1-99% of solvent B (100% Acetonitrile, 0.08% formic acid) over a total programme time of 72 min. The nano-LC method separated peptides and glycopeptides, including possible glycoforms, for on-line mass spectrometry (Gilar et al., 2011, Alley et al., 2013).

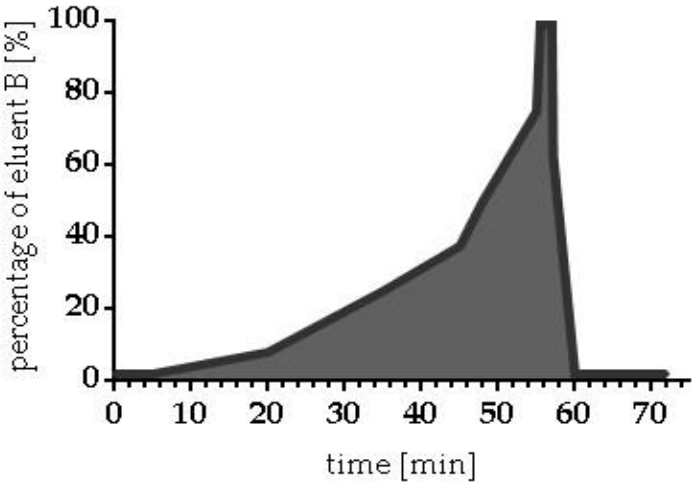


Figure 2.8: Elution gradient used for separation of the glycopeptides derived from the native proMBP/PAPP-A1-complex. The final spike of solvent B ensures the elution of remains from the column.

Table 11: Elution gradient programme used on the Acclaim PepMap 100 C18 LC column for analysis of the glycopeptides derived from the native proMBP/PAPP-A1-complex.

Retention time [min]	**Flow rate [nl/min]**	**Percentage of solvent A (100% water, 0.08% formic acid) [%]**	**Percentage of solvent B (100% Acetonitrile, 0.08% formic acid) [%]**
0	300	98	2
0	300	98	2
5	300	98	2
10	300	96	4
20	300	92	8
35	300	75	25
45	300	62.5	37.5
48	300	50	50
55	300	25	75
55.5	300	1	99
57	300	1	99
57.2	300	37.5	62.5
60	300	98	2
72	300	98	2

The data acquisition of the glycopeptides was carried out on the Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific Inc., USA) with an Orbitrap as well as an Iontrap analyser available. The samples were sprayed in positive ion mode via a metal emitter tip using the NSI method.

Full MS scans were acquired from m/z 200-1600 at 120,000 resolution via the Orbitrap analyser. Within a cycle time of 3 s, precursor ions were selected for fragmentation using the criteria of monoisotopic precursor selection, charge states (included states 2-8), dynamic exclusion (exclusion duration of 5 s) and an intensity threshold of 50,000 A.U. The following decision tree of the method programme included, first, the highest charge state and, second, the most intense precursor ions for MS/MS analysis.

MS/MS scans were performed via fragmentation of the precursor ions via the method of higher-energy collisional dissociation (HCD) as well as subsequent fragmentation using a HCD-triggered collision-induced dissociation (CID) programme. Both methods detected fragment ions via the Orbitrap analyser at a resolution of 15,000 after detection of listed diagnostic ions, i.e. oxonium ions derived from glycan structures, produced via HCD MS/MS.

A schematic overview of the workflow is presented in figure 2.9.

The acquired raw spectra were searched against FASTA sequences of the glycoproteins via Byonic with the mass tolerance set as 10 ppm for both, full MS and MS/MS scans. For N-and O-glycopeptide analysis, carbamidomethylation of cysteine as well as and oxidation of methionine were used as search parameters (as variable modification). Assigned glycopeptides were, first, software-filtered based on probability and, second, validated manually by evaluating glycan neutral loss pattern, oxonium ions, and the glycopeptide fragmentations.

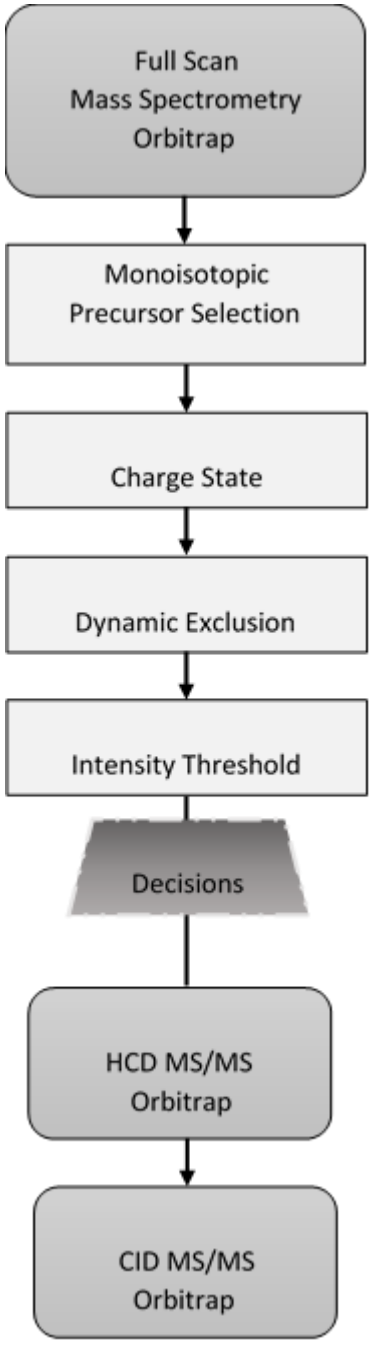


Figure 2.9: Schematic overview of the MS method used for on-line nano-LC-MS/MS analysis of the glycomic and glycoproteomic characteristics of the glycopeptides derived from the native proMBP/PAPP-A1-complex.

2.6.3 Glycomics

a) N-linked Glycans

N-linked glycans can be released enzymatically from glycoproteins, but more efficiently from glycopeptides considering the lower level of steric hindrance. The commercially available enzyme PNGase F, purified from *Flavobacterium meningosepticum*, is able to cleave the glycosidic bond that connects the asparagine to the first, core GlcNAc residue of the glycan. Hence, PNGase F is the enzyme of choice for mammalian-derived glycoproteins/peptides. However, PNGase F is not able to efficiently release N-glycans that contain fucose linked α-1→3 as well as for many N-linked oligosaccharides from plants, insects and microbes (Tretter et al., 1991). Given that the recombinant glycoproteins of interest were expressed in mammalian expression systems and the native PAPP-A1/proMBP-complex derived from human blood, PNGase F was the first choice.

The lyophilised fraction of the reduced and alkylated glycopeptides was re-dissolved in 48 µL GlycoBuffer 2 (50 mM Sodium Phosphate buffer pH 7.5) provided for the PNGase F reaction (from New England Biolabs). For release of N-glycans, 2 µL PNGase F (equals 1000 units) were added and incubated at 37˚C for 16-18 hrs. Afterwards, the mixture was cooled to room temperature and loaded onto pre-equilibrated C18 SPE cartridges. The cartridges were washed three times with 1 mL 5% acetic acid and the flow-through collected, which contained the hydrophilic, released N-glycans. The more hydrophobic de-glycosylated peptides as well as O-linked glycopeptides were eluted in fractions from the C18 SPE cartridges with 1 mL of 20%, 40%, 100% isopropanol in 5% acetic acid, respectively. The solvent was removed via evaporation by Nitrogen gas. The wash fraction containing the N-glycans and the elution fraction containing peptides and O-glycopeptides were lyophilised.

In the following step, the released N-glycans were permethylated, which replaces the hydrogen atoms of the hydroxyl groups with methyl groups. This derivatisation offers several advantages, such as reduced hydrophilicity that results in a more sensitive MS analysis, replacing negative charges that enable MS conducted completely in positive ionisation mode, enhancing cross-ring fragmentation for MS/MS analysis as well as providing structural information on the monosaccharide

sequence of the glycans since linkages do not become permethylated (Harvey, 2011, Alley et al., 2013).

The lyophilised N-glycans were dissolved in 200 µL anhydrous dimethyl sulfoxide (DMSO) and 300 µL of anhydrous sodium hydroxide in DMSO were added. After adding 150 µL iodomethane, the mixture was vortexed thoroughly for 10 min. The permethylation reaction was stopped by adding 2 mL nanopure water, which turns the sample opaque and indicates a successful permethylation. Excess iodomethane was volatilised by directing a stream of nitrogen gas through the sample that turned the sample transparent again. To extract the permethylated N-glycans, 2 mL dichloromethane were added. The mixture was vortexed vigorously and centrifuged briefly for separation of the layers. The aqueous layer was removed, and the dichloromethane layer washed again with 2 mL nanopure water. The procedure was repeated five times. Finally, the dichloromethane layer was removed carefully, transferred into a new vial and dried by a stream of nitrogen gas.

The permethylated N-glycans were re-dissolved in 20 µL methanol using a gastight Hamilton syringe and 2 µL of the sample were mixed with 2 µL of 2,5-dihydroxybenzoic acid matrix (10 mg/mL in 1:1 methanol/nanopure water). The mixture was incubated until solvents volatilised down to approximately 1 µL. With a micropipette, 0.5 µL of the mixture were placed on an Opti-TOF sample plate and air-dried. The spotted samples were analysed with an AB SCIEX MALDI TOF/TOF 5800 mass spectrometer (Applied Biosystems Inc., USA) with the laser intensity set to 60% around 4,000 A.U. The data was acquired in positive ion mode and data analysis performed using AB SCIEX Analyst Software and Data Explorer. Data was interpreted and assigned with the GlycoWorkbench 1.1 software.

b) O-linked Glycans

The O-linked glycopeptides as well as unglycosylated peptides were separated from the released N-glycans during the C18 solid-phase extraction. The O-linked glycopeptides were found in the elution fraction due to a higher hydrophobicity. To date, no de-glycosylation enzymes with a wide specificity for O-glycans are available and, therefore, chemical reactions (i.e. reductive β-elimination, ammonia-based nonreductive β-elimination and hydrazinolysis) are typically used for a comprehensive release (Alley and Novotny, 2013).

The alkaline-based β-elimination was the chemical method of choice for O-linked glycan release. The lyophilised O-glycopeptide fraction was re-dissolved in 250 µL of 50 mM sodium hydroxide and the pH checked for a basic pH via indicator paper. Sodium borohydride was prepared at 76 mg/mL in 50 mM sodium hydroxide and 250 µL of the solution added to the 250 µL of dissolved O-glycopeptides. The mixture was vortexed, incubated at 45 ˚C for 16-18 hrs and cooled down to room temperature afterwards. Carefully, 10% acetic acid solution was added dropwise to the sample for neutralisation of the pH, which stopped the alkaline-based β-elimination reaction, until no further formation of hydrogen gas was observable.

For improvement of the MS analysis of the sample, the removal of salts was required. This step was realised with ion-exchange chromatography using the AG 50W-X8 cation exchange resin (approximately 2 mL bed size volume) packed into a gravity column (10 mL reservoir volume). After equilibration of the resin with approximately 5 mL 5% acetic acid, the 500 µL sample was loaded onto the column and washed with 3 mL 5% acetic acid to elute the released O-glycans. The sample was lyophilised and re-dissolved in 500 µL of methanol:acetic acid (9:1) solution for removal of the remaining borates. The sample was dried under a stream of nitrogen gas to volatise the borates and the procedure repeated five times.

The following step was the permethylation of the released O-glycans, which follows the same protocol as permethylation of N-glycans (see section '2.6.3. a) N-linked Glycans'). In summary, the glycans were dissolved in 200 µL anhydrous DMSO and 300 µL of anhydrous sodium hydroxide in DMSO, followed by addition of 50 µL iodomethane. After vortexing for 10 min, the reaction was stopped by adding 2 mL nanopure water and excess iodomethane was volatilised with nitrogen gas. For extraction of the O-glycans, 2 mL dichloromethane were added and the mixture vortexed. The aqueous layer was removed, and the dichloromethane layer washed five more times with 2 mL nanopure water. Eventually, the dichloromethane layer was transferred into a new vial and dried by a stream of nitrogen gas.

The lyophilised, released O-glycans were dissolved in 20 µL methanol. Before spotting on a MS sample plate, 2 µL of the sample were mixed with 2 µL 2,5-dihydroxybenzoic acid matrix (10 mg/mL in 1:1 methanol/nanopure water). The

mixture was incubated until solvents volatilised down to approximately 1 µL. With a micropipette, 0.5 µL of the mixture were spotted on an Opti-TOF sample plate and air-dried. Analysis of the spotted samples was performed with an AB SCIEX MALDI TOF/TOF 5800 mass spectrometer (Applied Biosystems Inc., USA) with the laser intensity set to 60% around 4,000 A.U. The data was acquired in positive ion mode. The analysis of the data was realised by use of AB SCIEX Analyst Software and Data Explorer. Data was interpreted and assigned with the GlycoWorkbench 1.1 software.

3 Results and Discussion (I)

Biophysical Interaction studies between human C-type lectins and human glycoproteins

3.1 Background

Preliminary findings (Ilyas, 2010) raised the possibility that there could be a significant interaction between the human glycoprotein PAPP-A1 (a metzincin protease) and human C-type lectins. Establishing the kinetic characterisation of these interactions is an area of great interest, since quantitative data can indicate whether the interactions lie within the physiologically relevant concentration ranges.

Healthy, non-pregnant humans (male and female) typically have a blood plasma level of PAPP-A1 between 0 nM and at most 0.8 nM. However, in pregnant women PAPP-A1 plasma levels become detectable early in pregnancy and steadily rise during all three trimesters - with levels up to 85 nM at term (Gall and Halbert, 1972, Lin et al., 1974). The PAPP-A1 polypeptide circulates in plasma of pregnant women as a 2:2-complex with the propeptide of eosinophil major basic protein (proMBP), which appears to act as an inhibitor to the protease function of PAPP-A1.

Interestingly, PAPP-A1 has been found - without the inhibition of its enzymatic nature via proMBP - in patients suffering from acute coronary syndromes (Bayes-Genis et al., 2001). PAPP-A1 levels were determined to rise up to 3 nM in patients with a STEMI (ST-segment elevation myocardial infarction) heart attack.

A set of soluble, recombinant C-type lectin receptors with a similar binding specificity to that of DC-SIGN were included in the biophysical interaction study, such as DC-SIGNR and MMR. All have a preference for binding to mannose-type oligosaccharides but differ in the patterns of recognition and the structural organisation of CRDs.
Even though the mechanisms and principles of binding between C-type lectins and ligands have been widely investigated, much is still not known about the range of endogenous ligands for these immune-related receptors.

The biophysical technique used for real-time interaction analysis was the state-of-the-art technique of biolayer interferometry (BLI) using the ForteBio Octet RED96 system.

The native human PAPP-A1/proMBP-complex (purified from retroplacental blood) as well as the recombinant PAPP-A1 homodimer (human gene product derived from mouse myeloma cells) were only available in small amounts within the microgram range. This scarcity extended to the availability of commercial stocks; the advantages of the BLI platform are its high sensitivity and modest requirement of sample volumes. Therefore, it was possible to investigate and quantify protein-protein interactions that were previously too challenging to resolve via more popular techniques such as surface plasmon resonance (SPR).

The experiments were set up and designed in keeping with the manufacturers' technical notes - available through the official website www.fortebio.com.

The flow rate for all steps within the experiments presented in this chapter was set to 1000 rpm. Controls were included, such as blanks with assay buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 5 mM $CaCl_2$) to check for a system drift during the experiment, and a negative control, i.e. a channel with the highest analyte concentration within experiment but 5 mM ETDA instead of 5 mM $CaCl_2$, since the analysed C-type lectins are calcium-dependent for carbohydrate binding.

The control values were subtracted from the raw data prior to kinetic analysis. The processed data was analysed to yield sets of binding rates and constants.

3.2 Functionality tests for BLI studies

Biolayer interferometry is a recent technology with relatively little exposure to the field of glycobiology, therefore, functionality tests with an abundant ligand that is known to yield a clear interaction response with C-type lectins were conducted to validate the integrity of the method.

The functionality of the C-type lectins was confirmed during the purification of the lectins by chromatography as described in section '2.3.3.1 Affinity chromatography'.

3.2.1 Functionality of recombinant DC-SIGN

Biotinylated, recombinant extracellular domain of DC-SIGN was immobilised on SA (streptavidin) biosensors and glycogen - an abundant, soluble and robust ligand (Geurtsen et al., 2009) - was used as the analyte at a concentration of 0.01 mg/mL in assay buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 5 mM $CaCl_2$). The protocol for the functionality test involved a baseline step for 60 s in assay buffer, a contact time for association between ligand and analyte for 350 s, and a dissociation period of 450 s in assay buffer. A clearly positive binding response with an interferometric deflection level of 1.19 nm was observed, as shown in figure 3.1.

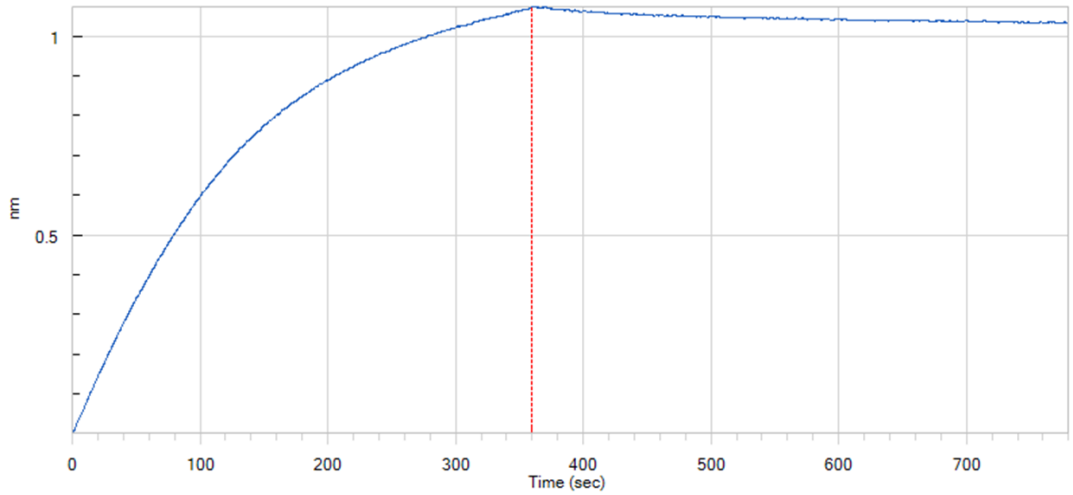


Figure 3.1: Representative sensorgram of immobilised DC-SIGN binding to glycogen at a concentration of 0.01 mg/mL. The shown binding event had a maximum binding response of 1.19 nm, which is a clear positive binding event.

3.2.2 Functionality of recombinant DC-SIGNR

The recombinant extracellular domain of DC-SIGNR was biotinylated and tested for functionality, and comparability to obtained BLI data to SPR data.

After immobilisation on SA (streptavidin) biosensors, the natural polymer glycogen was used as positive control at a concentration of 0.01 mg/mL in assay buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 5 mM $CaCl_2$).

The protocol for testing the functionality of DC-SIGNR included a baseline step (60 s) in assay buffer, a contact time for association for 350 s, and a dissociation period of 450 s in assay buffer. A clear binding response for glycogen to the extracellular domain of DC-SIGNR was observed, as shown in figure 3.2. This confirmed the integrity of the BLI detection and functionality of DC-SIGNR, although a comparatively lower overall interferometric response of 0.21 nm was observed. This is most likely due to gradually different preferences for spatial patterns of carbohydrate recognition between the CRDs of DC-SIGN and DC-SIGNR.

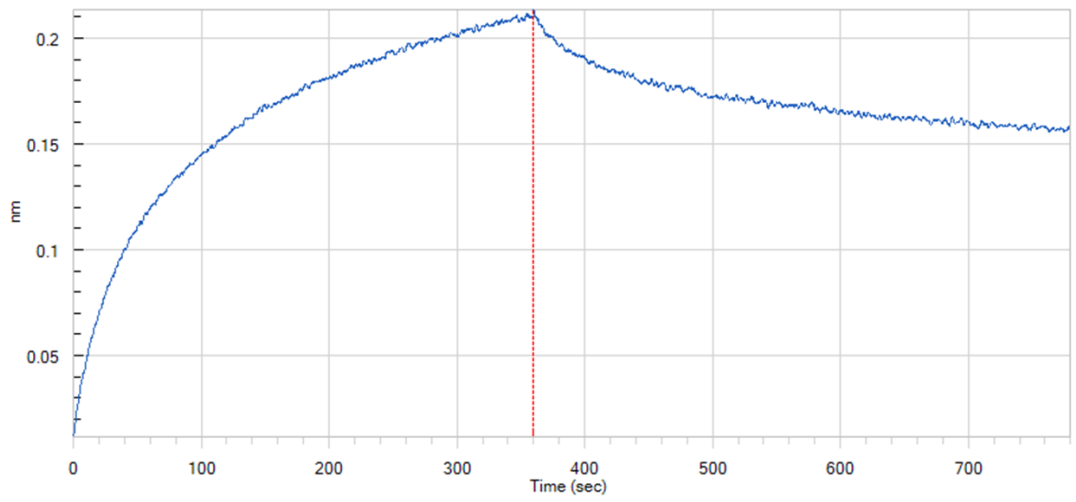


Figure 3.2: Representative sensorgram of immobilised DC-SIGNR binding to glycogen at a concentration of 0.01 mg/mL. The interaction reached a maximum binding response of 0.21 nm, which represents a significant binding event.

3.3 Biophysical interaction studies between glycoproteins and DC-SIGN

3.3.1 Human native PAPP-A1/proMBP-complex and DC-SIGN

Two experimental protocols were tested, using streptavidin-binding as well as amine-coupling biosensors.

For SA (streptavidin) biosensors, the biotinylated, recombinant DC-SIGN was immobilised at the probe as ligand, and the human native PAPP-A1/proMBP-complex (M_W=500 kDa) was tested as analyte in a range of serial dilutions (100 nM-1 nM). The protocol of the assay included a baseline step for 60 s in assay buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 5 mM $CaCl_2$), a contact time for association between ligand and analyte for 600 s, and a dissociation period of 600 s in assay buffer. Binding was observed in a concentration-dependent manner, as depicted in figure 3.3.

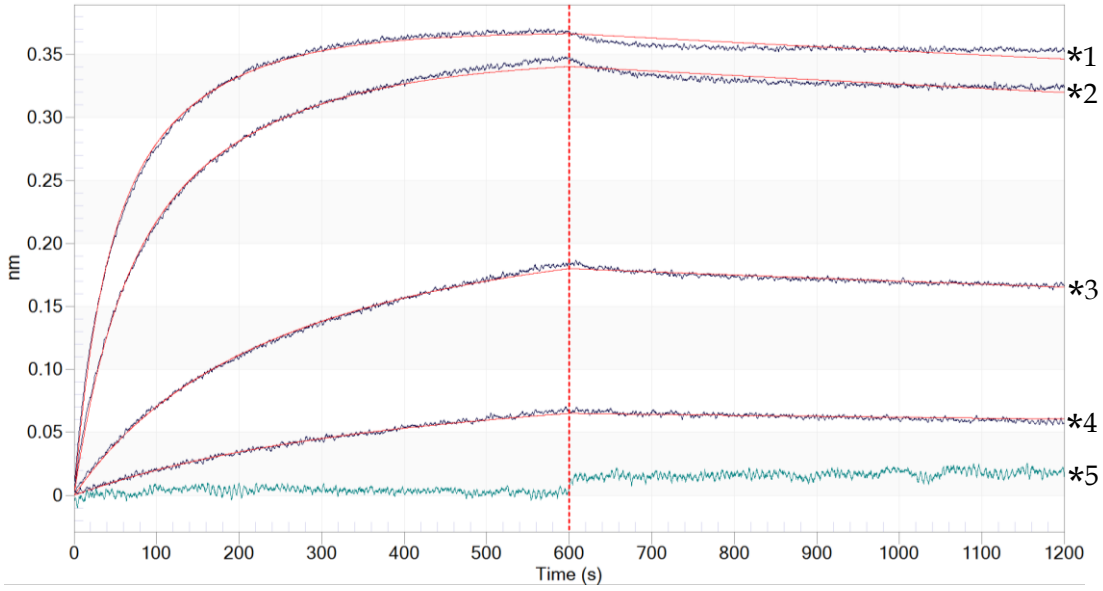


Figure 3.3: Representative sensorgram of the interaction between biotinylated DC-SIGN and the human native PAPP-A1/proMBP-complex via SA (streptavidin) biosensors. The sensorgram shows the shift in the wavelength of the reflected light, interpreted as binding response, over time during the association and dissociation phase, 600 s respectively.
The interaction with the human native PAPP-A1/proMBP-complex (hnPAPP-A1) is significant and clearly concentration-dependent. Annotations:
**1 75 nM hnPAPP-A1, *2 50 nM hnPAPP-A1, *3 25 nM hnPAPP-A1,*
**4 12.5 nM hnPAPP-A1, *5 100 nM hnPAPP-A1 in assay buffer with 5 mM EDTA*

Table 12 summarises the calculated kinetic parameters, whereby each run/experiment is based on n≥3.

The average dissociation constant was calculated as K_D=0.61 nM (6.14E-10 M) for the SA biosensors with immobilisation of the extracellular domain of DC-SIGN. One strinking detail of the interaction is the very slow off-rate (k_{off}), indicating that once formed complexes between DC-SIGN and PAPP-A1/proMBP are very stable in the presence of calcium.

Table 12: Overview of the kinetic parameters obtained from the binding studies between the immobilised extracellular domain of DC-SIGN and the human native PAPP-A1/proMBP-complex (hnPAPP-A1).

Experiment	k_{on} **(M^{-1} s^{-1})**	k_{off} **(s^{-1})**	K_D **(M)**
Run 1 soluble hnPAPP-A1	8.79E+05	6.76E-04	7.69E-10
Run 2 soluble hnPAPP-A1	8.77E+05	4.06E-04	4.62E-10
Run 3 soluble hnPAPP-A1	8.72E+05	5.32E-04	6.10E-10

For AR2G (amine-coupling) biosensors, the human native PAPP-A1/proMBP-complex was immobilised (0.02 mg/mL; loading buffer 10 mM HEPES pH 5.5, 150 mM NaCl) at the probe as ligand via amine-coupling (scarcity of PAPP-A1/proMBP material prohibited the biotinylation of this protein for use with SA probes). The soluble recombinant DC-SIGN extracellular domain ($M_W^{TETRAMER}$= 150 kDa) was tested as analyte in a serial dilution (100 nM-50 nM). Negative controls included a buffer blank as well as 100 nM analyte in EDTA-containing buffer. The protocol for the assay included a baseline step (60 s) in assay buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 5 mM $CaCl_2$), a contact time for association between ligand and analyte for 1200 s, and a dissociation period of 1200 s in assay buffer. Binding was observed in a concentration-dependent manner, which is shown in figure 3.4. The kinetic parameters derived from experiments (n≥3 per single-experiment K_D) are shown in table 13. The average dissociation constant was calculated as K_D=0.44 nM (4.46E-10 M) for the AR2G biosensors with immobilisation of the human native PAPP-A1/proMBP.

Table 13: Overview of the kinetic parameters obtained from the binding studies between the immobilised human native PAPP-A1/proMBP-complex and the soluble extracellular domain of DC-SIGN.

Experiment	**k_{on} (M^{-1} s^{-1})**	**k_{off} (s^{-1})**	**K_D (M)**
Run 1 soluble DC-SIGN	3.90E+05	1.45E-04	3.72E-10
Run 2 soluble DC-SIGN	3.96E+05	2.29E-04	5.79E-10
Run 3 soluble DC-SIGN	1.36E+06	5.23E-04	3.86E-10

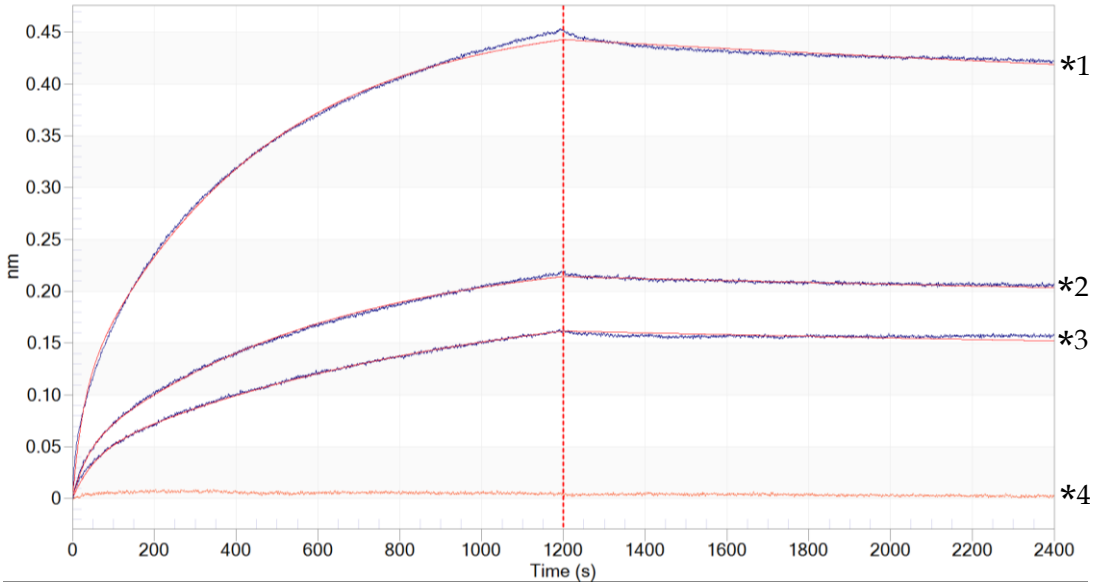


Figure 3.4: Representative sensorgram of soluble DC-SIGN binding to immobilised hnPAPP-A1. The graphs show a significant and concentration-dependent binding (dilution series from 100 nM to 50 nM of the soluble, extracellular domain of DC-SIGN). The EDTA reference was subtracted for analysis of kinetic parameters but is shown as a visual reference. Annotations:

*[*1] 100 nM DC-SIGN, [*2] 75 nM DC-SIGN, [*3] 50 nM DC-SIGN*

*[*4] 100 nM DC-SIGN in assay buffer with 5 mM EDTA*

3.3.2 Human recombinant PAPP-A1 and DC-SIGN

Homodimers of human recombinant PAPP-A1 (M_W=400 kDa) were tested for binding to the extracellular domain of DC-SIGN. Biotinylated DC-SIGN was immobilised on the tip of SA (streptavidin) biosensors.

The recombinant PAPP-A1 was prepared as solution-phase analyte in serial dilution (100 nM-1 nM). The protocol for the assay included a baseline step for 60 s in assay buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 5 mM $CaCl_2$), a contact time for association between ligand and analyte for 600 s, and a dissociation period of 840 s in assay buffer. As shown in figure 3.5, binding was observed in a concentration-dependent manner using a dilution series from 100 nM to 1 nM. Evaluation of n=6 experiments is presented in table 14. The analysis of the association and dissociation phase revealed an average dissociation constant K_D= 1.3 nM (1.307E-09 M).

Given that the recombinant PAPP-A1 was produced in murine NS0 myeloma cells, the N-glycosylation pattern was hypothesised to differ to the native PAPP-A1 form (human origin). The structural aspects of N-glycosylation were investigated in section '4.4 Recombinant human PAPP-A'.

Table 14: Overview of the kinetic parameters obtained from the binding studies between the recombinant PAPP-A1 (homodimer) and the extracellular domain of DC-SIGN.

Experiment	k_{on} **(M^{-1} s^{-1})**	k_{off} **(s^{-1})**	K_D **(M)**
Run 1 soluble recPAPP-A1	7.88E+05	5.31E-04	6.74E-10
Run 2 soluble recPAPP-A1	2.61E+05	5.06E-04	1.94E-09

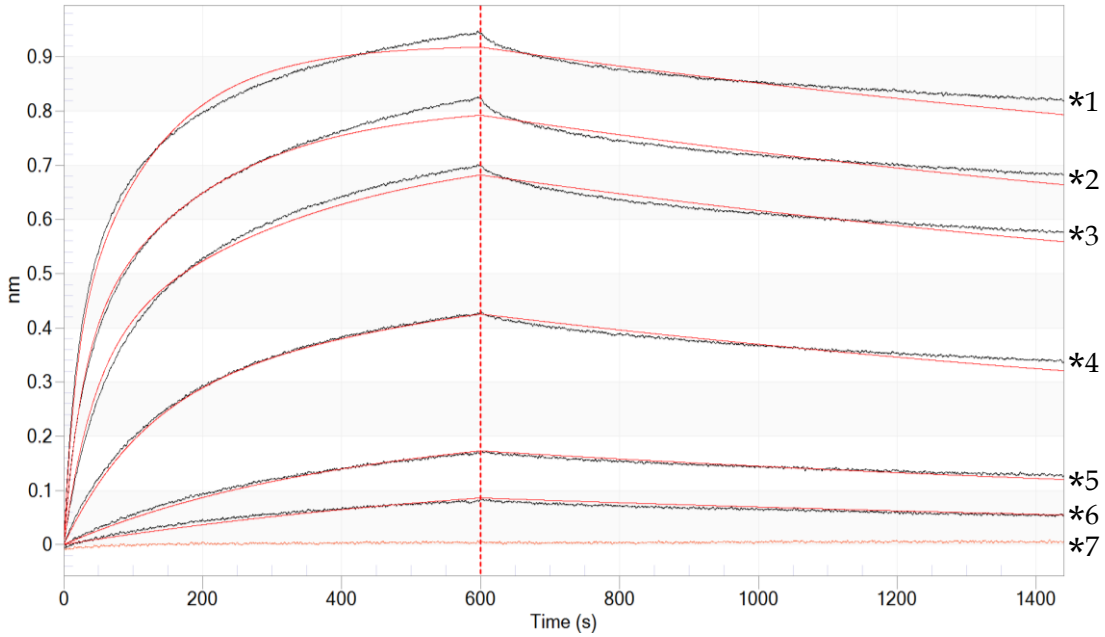


Figure 3.5: Representative sensorgram of human recombinant PAPP-A1 binding to the immobilised extracellular domain of DC-SIGN. The experiments revealed a significant and concentration-dependent binding (dilution series ranging from 100 nM to 1 nM) of recombinant PAPP-A1 (recPAPP-A1) to DC-SIGN. The EDTA reference was subtracted for analysis of kinetic parameters but is depicted for visual reference.
Annotations:

**1 80 nM recPAPP-A1, *2 50 nM recPAPP-A1, *3 30 nM recPAPP-A-1,*
**4 10 nM recPAPP-A1, *5 3 nM recPAPP-A1, *6 1 nM recPAPP-A-1,*
**7 80 nM recPAPP-A-1 in assay buffer with 5 mM EDTA*

3.4 Biophysical interaction studies between glycoproteins and DC-SIGNR

DC-SIGNR is closely related to DC-SIGN but with a more restricted binding specificity that does not favour fucosylated ligands. Furthermore, the disposition of the CRDs relative to the neck domain is different in DC-SIGNR compared to DC-SIGN (Feinberg et al., 2005, Feinberg et al., 2009).

For more than decade, DC-SIGNR has been considered for potential involvement in the clearance of endogenous glycoproteins (Soilleux et al., 2000, Figdor et al., 2002).

Rydz et al. (2013) showed the ability of DC-SIGNR to internalise the human plasma glycoprotein von Willebrand Factor (vWF) after binding. Since DC-SIGNR has a putative role in clearance and bioavailability of vWF, the blood serum levels of other human glycoproteins could be influenced by DC-SIGNR *in vivo*.

Therefore, it was of interest to elucidate the binding affinity of the human native PAPP-1/proMBP-complex as well as the human recombinant PAPP-A1 homodimer, respectively, to the recombinant, extracellular domain of DC-SIGNR.

3.4.1 Human Native PAPP-A1/proMBP-complex and DC-SIGNR

For the characterisation of the binding between human native PAPP-1/proMBP-complex and the extracellular domain of DC-SIGNR, two experimental protocols were tested, i.e. streptavidin-binding (SA) and amine-coupling (AR2G) biosensors.

Biotinylated, recombinant DC-SIGNR was immobilised at the probe of SA sensors as ligand. The human native PAPP-A1/proMBP-complex (M_W=500 kDa) was run in serial dilution (100 nM-1 nM) as analyte. The experiments were designed with a baseline step for 60 s in assay buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 5 mM $CaCl_2$), a contact time for association between ligand and analyte for 600 s, and a dissociation period of 600 s in assay buffer. The concentration-dependent binding, as shown in figure 3.6, was analysed. A significant binding response was obtained for concentrations higher than 10 nM. Table 15 shows the obtained kinetic parameters (n≥3 for each experiment). The average dissociation constant was calculated as K_D= 17.5 nM (1.751E-08 M) for the SA biosensors with immobilisation of the extracellular domain of DC-SIGNR.

Table 15: Overview of the kinetic parameters obtained from the binding studies between the human native PAPP-A1/proMBP-complex and the extracellular domain of DC-SIGNR.

Experiment	k_{on} **(M^{-1} s^{-1})**	k_{off} **(s^{-1})**	K_D **(M)**
Run 1 soluble hnPAPP-A1	4.52E+04	1.05E-04	2.33E-09
Run 2 soluble hnPAPP-A1	1.53E+05	7.33E-03	4.79E-08
Run 3 soluble hnPAPP-A1	1.83E+04	4.20E-05	2.30E-09

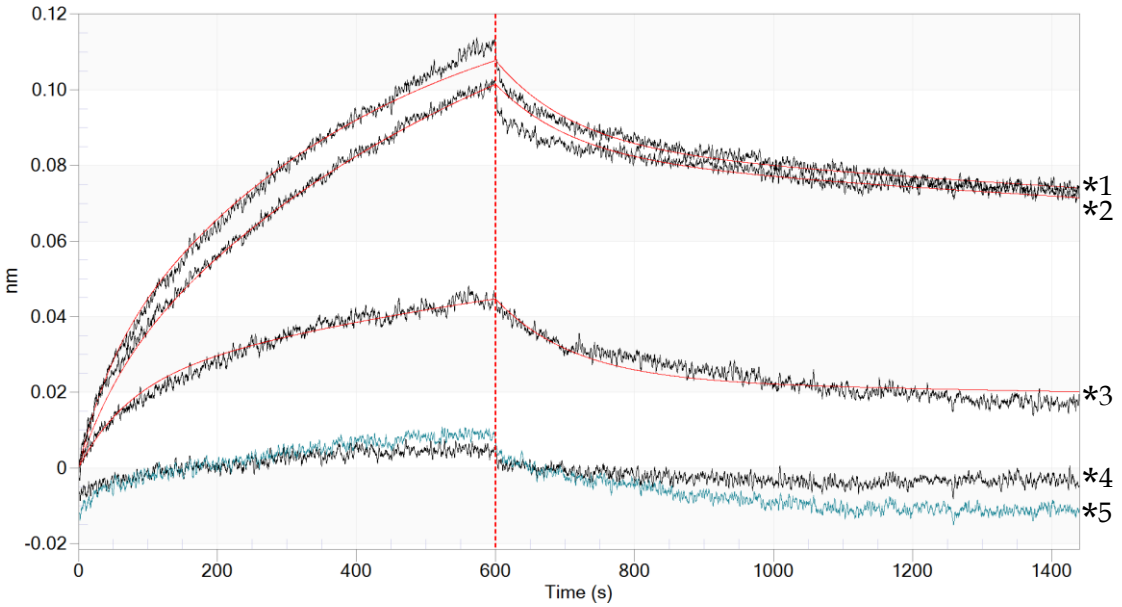


Figure 3.6: Representative sensorgram of the interaction between biotinylated DC-SIGNR and the human native PAPP-A1/proMBP-complex via SA (streptavidin) biosensors. The sensorgram shows the shift in the interference pattern of the reflected light, interpreted as binding response in nm, over time during the association phase (600 s) and dissociation phase (840 s). The interaction with the human native PAPP-A1/proMBP-complex (hnPAPP-A1) is significant and clearly concentration-dependent.
Annotations:
**1 80 nM hnPAPP-A1, *2 50 nM hnPAPP-A1, *3 30 nM hnPAPP-A1,*
**4 10 nM hnPAPP-A1, *5100 nM hnPAPP-A1 in assay buffer with 5 mM EDTA*

Interestingly, for the protocol using AR2G (amine-coupling) biosensors, it was not possible to detect significant binding nor obtain consistent kinetic parameters of the putative interaction between the immobilised human native PAPP-A1/proMBP-complex (0.02 mg/mL; loading buffer 10 mM HEPES pH 5.5, 150 mM NaCl) and the soluble recombinant DC-SIGNR extracellular domain. Most of the experiments indicated no consistent binding of soluble DC-SIGNR to the immobilised human native PAPP-A1/proMBP-complex, as shown in figure 3.7.

Only during one experiment, evidence of a single binding event was observed: The kinetic parameters of this single experiments are shown in table 16 with an apparent dissociation constant of K_D=2.3 nM (2.30E-09 M). However, the coarseness of the signal and lack of reproducibility makes the kinetic data highly speculative.

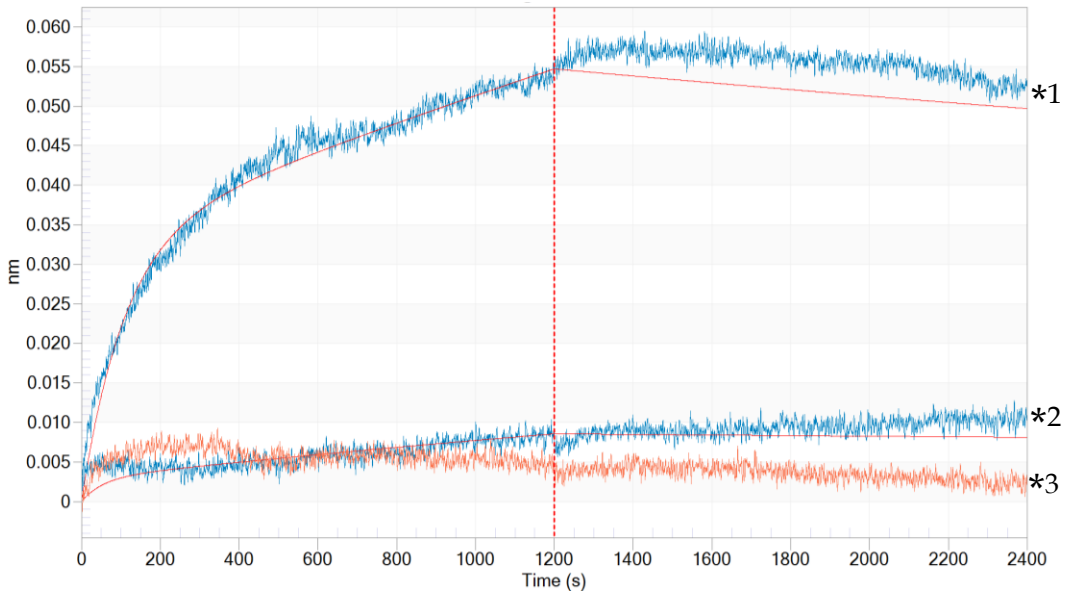


*Figure 3.7: Representative sensorgram of soluble DC-SIGNR binding to immobilised human native PAPP-A1/proMBP-complex. Only very high concentrations, i.e. more than 50 nM, enabled DC-SIGNR to bind to immobilised human native PAPP-A1/proMBP-complex. Overall, the binding response was very low and not significant. The EDTA reference was subtracted for analysis of kinetic parameters but is shown for visual reference. Annotations: *1 100 nM DC-SIGNR, *2 50 nM DC-SIGNR, *3 100 nM DC-SIGN in assay buffer with 5 mM EDTA*

Table 16: Overview of the kinetic parameters obtained from the binding studies between the immobilised human native PAPP-A1/proMBP-complex and the extracellular domain of DC-SIGNR.

Experiment	k_{on} **(M^{-1} s^{-1})**	k_{off} **(s^{-1})**	K_D **(M)**
Run 1 soluble DC-SIGNR	1.83E+04	4.20E-05	2.30E-09
Other runs	No binding		

3.4.2 Human recombinant PAPP-A1 and DC-SIGNR

The recombinant DC-SIGNR was investigated for potentially binding the human recombinant PAPP-A1 homodimer (M_W=400 kDa).

The biotinylated DC-SIGNR extracellular domain was immobilised on a streptavidin (SA) biosensor. The recombinant PAPP-A1 was set as soluble analyte in a serial dilution (80 nM-1 nM). The set-up of the assay included a baseline step for 60 s in assay buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 5 mM $CaCl_2$), a contact time for association between ligand and analyte for 720 s, and a dissociation period of 840 s in assay buffer. None of the experiments showed a convincing, significant binding response between DC-SIGNR and the recombinant PAPP-A1. Figure 3.8 shows clearly how weak the overall binding response was – even at higher concentrations of the analyte. However, the forced fitting suggests that any apparent binding response is nonspecific.

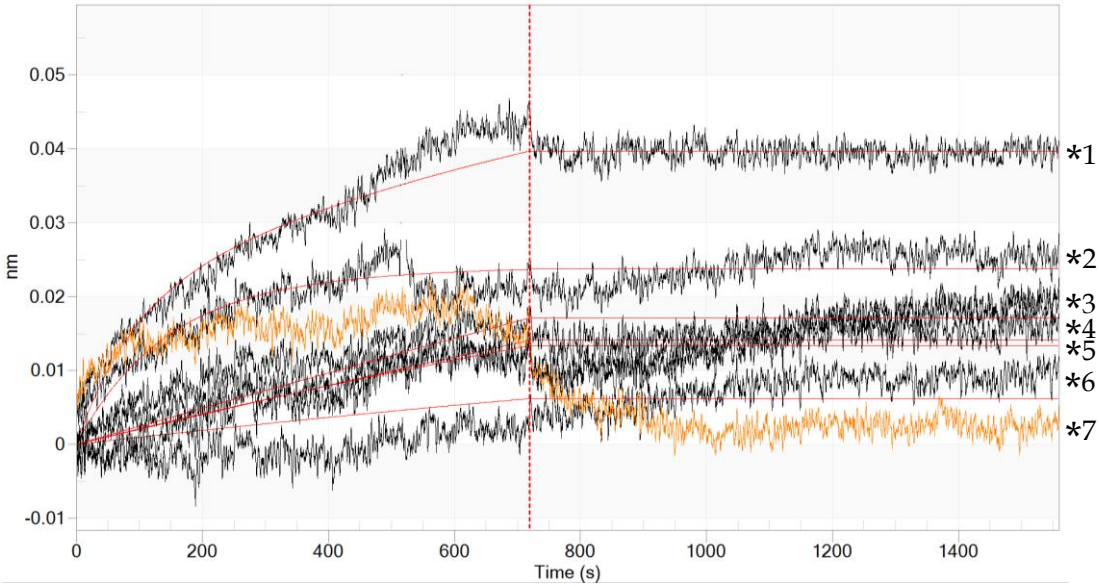


Figure 3.8: Representative sensorgram of the interaction between human recombinant PAPP-A1 and the immobilised extracellular domain of DC-SIGNR. The experiments (dilution serial from 80 nM to 1 nM) yielded a very low overall binding response, which was non-specific. The EDTA reference was subtracted for analysis of kinetic parameters but is depicted for visual reference. Annotations:

**1 80 nM recPAPP-A1, *2 50 nM recPAPP-A1, *3 30 nM recPAPP-A-1,*

**4 10 nM recPAPP-A1, *5 3 nM recPAPP-A1, *6 1 nM recPAPP-A-1,*

**7 80 nM recPAPP-A-1 in assay buffer with 5 mM EDTA*

3.5 Biophysical interaction studies between glycoproteins and MMR

The commercially available macrophage mannose receptor (MMR) was tested for interaction with the native human PAPP-A1/proMBP-complex. The MMR was chosen since it contains multiple CRDs of several CTLD categories that preferentially bind to mannose-type sugars and terminal mannose, fucose and N-acetyl-galactosamine residues. MMR is a major receptor on the surface of subpopulations of macrophages and dendritic cells.

Its structural characteristics make the MMR a very interesting C-type lectin for comparative investigation with regards to DC-SIGN and DC-SIGNR. Interestingly, DC-SIGN and MMR are often co-expressed on macrophages, and all three C-type lectins are strongly associated with immune responses. However, the extent of ligand range and subtle differential binding preferences between these receptors still remain unclear.

Availability of soluble recombinant full-length MMR protein was a challenge, although the BLI platform did allow for comparative molar amounts to be used in order to relate any ligand binding properties to those of DC-SIGN and DC-SIGNR.

3.5.1 Human Native PAPP-A1/proMBP-complex and MMR

For the investigation of a putative interaction between the human native PAPP-1/proMBP-complex and the recombinant macrophage mannose receptor, the method of amine-coupling (AR2G) biosensors was chosen.

The human native PAPP-1/proMBP-complex was immobilised (0.02 mg/mL; loading buffer 10 mM HEPES pH 5.5, 150 mM NaCl) at the probe of AR2G sensors as ligand. The recombinant MMR was run as analyte. The experiments were designed with a baseline step for 60 s, and association and a dissociation periods of 1200 s in assay buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 5 mM $CaCl_2$).

No consistent and significant binding response was detectable. Most runs showed only a very low level of interaction (see figure 3.9). Forced fitting of the data suggested that any apparent binding response is nonspecific.

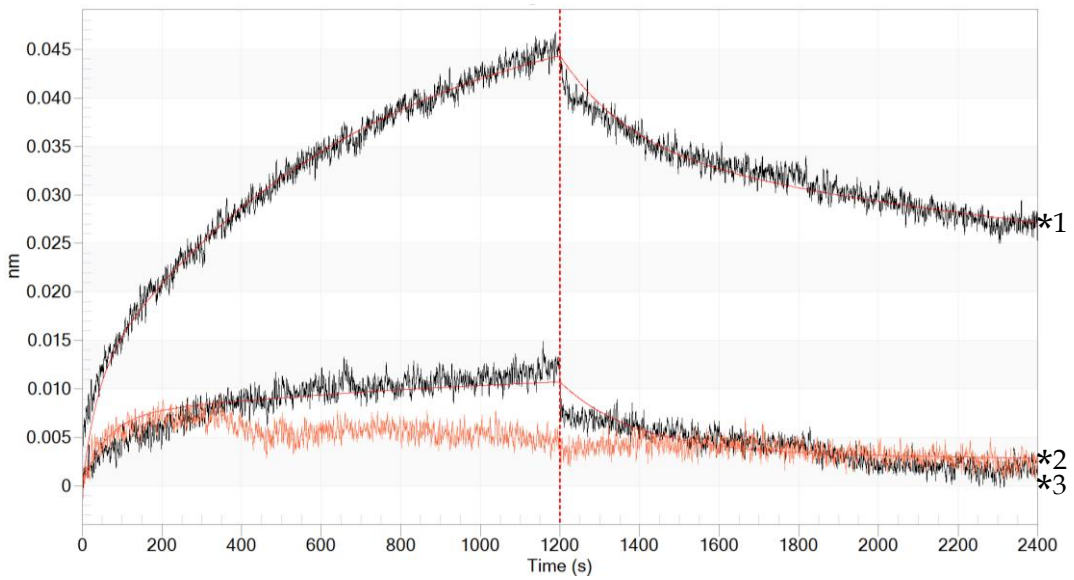


Figure 3.9: Representative sensorgram of soluble macrophage mannose receptor interacting with immobilised human native PAPP-A1/proMBP-complex. Only high concentrations (over 50 nM analyte) enabled the detection of a marginal binding response. The EDTA reference is depicted for visual reference.
Annotations:
*[*1] 100 nM MMR, [*2] 50 nM MMR,*
*[*3] 100 nM DC-SIGN in assay buffer with 5 mM EDTA*

3.6 Conclusion of the interaction studies between C-type lectins and human glycoproteins

The binding studies between human C-type lectins and glycoproteins, performed via cutting-edge biolayer interferometry, have opened up the feasibility of working with low abundance samples, which is especially important for scarce materials, such as purified mammalian glycoproteins.

Overall, it enabled the specific measurement of interaction between complex carbohydrates and the selected C-type lectins, as significant binding only occurred in the presence of calcium and was prevented in the presence of EDTA.

The known binding affinities for C-type lectins to endogenous or exogenous ligands lie within the millimolar to mid-micromolar range for the C-type lectins examined here: DC-SIGN, DC-SIGNR and MMR (Taylor et al., 1992, Mitchell et al., 2001). Furthermore, all three lectins are known to mediate cellular uptake of glycoproteins ligands (Taylor and Drickamer, 1993, Figdor et al., 2002, Rydz et al., 2013).

Within this study, the speculation that DC-SIGN binds to the human native PAPP-A1/proMBP-complex (hnPAPP-A1; derived from human retroplacental blood) has been confirmed. This finding represents a novel endogenous ligand for DC-SIGN. It is also important to note that this is a native, endogenous and *soluble* ligand, as hitherto most reported glycoprotein ligands for DC-SIGN have been membrane-bound.

DC-SIGN revealed a clearly positive binding response to hnPAPP-A1. DC-SIGN was found to bind within the low nanomolar range (K_D=0.61 nM) to soluble-phase hnPAPP-A1 as well as immobilised hnPAPP-A1 (K_D= 0.44 nM). The parity of the data for the reciprocal experiments is encouraging.

DC-SIGNR showed a lower binding response to the human native PAPP-A1/proMBP-complex. The dissociation constant for the interaction between DC-SIGNR and soluble hnPAPP-A1 lies in the low to medium nanomolar range with K_D=17.5 nM. Interestingly, the investigation of the interaction between immobilised

hnPAPP-A1 and DC-SIGNR did not provide a clear result and the binding signal was of lower quality and consistency compared to similar experiments with DC-SIGN. This discrepancy could arise from the marked difference in CRD clustering and steric availability between immobilised and solution-phase DC-SIGNR. Relative to the neck proportions of the molecule, the CRDs of DC-SIGNR project in a perpendicular fashion (Feinberg et al., 2009). In DC-SIGN, the CRDs face directly upwards from the neck. Therefore - in solution phase - the DC-SIGNR binding surface for PAPP-A1/proMBP would be greatly reduced, consisting possibly of only a single available CRD domain.

Similarly, the MMR was not found to yield a significant binding signal in the presence of hnPAPP-A1. It is likely that the tandem organisation of CRDs in MMR do not favour the glycan arrangements on the surface of the glycoprotein.

Figure 3.10 summarises a comparative sensorgram of all three C-type lectins as analytes with the human native PAPP-A1/proMBP-complex as immobilised ligand. The data demonstrates binding between DC-SIGN and hnPAPP-A1 with affinity values that lie well within the physiological range for *in vivo* interaction, especially considering the fact that hnPAPP-A1 plasma levels of 85 nM are found in pregnancy at term (Smith et al., 1979).

The positive interaction between DC-SIGN and recombinant PAPP-A1 yielded a dissociation constant K_D=1.3 nM. Even though the K_D still lies within the low nanomolar range, it is nearly 3-fold higher in comparison to the interaction with PAPP-A1/proMBP. DC-SIGNR did not show any binding to the recombinant sample of PAPP-A1. This suggests that the glycosylation pattern of recombinant PAPP-A1 is different to the native form, and that the presence of the proMBP polypeptide may enhance the interaction with DC-SIGN.

Notably, the immobilisation of the hnPAPP-A1 on the BLI probes gave rise to some discrepancy, especially for DC-SIGNR, as noted above. The immobilisation of the C-type lectin receptors, to some extent, mimics the physiological situation of the cell better: The C-type lectins used here are naturally anchored receptors within the cell membrane. The binding of the soluble PAPP-A1/proMBP-complex is possible with

all degrees of freedom. The reciprocal set-up via amine-coupling restricts the freedom significantly and might be the cause for steric hindrance of binding epitopes. DC-SIGN did not show a reduced binding affinity for the amine-coupling set-up of the experiment, which is an encouraging validation of the interaction and may point to a simple stoichiometry with the PAPP-A1/proMBP-complex, although further biophysical investigation would be required to establish this hypothesis.

The significant differences in binding responses for DC-SIGN compared with the other two C-type lectins raises the attractive hypothesis that PAPP-A1/proMBP is a specific ligand for DC-SIGN and that selectivity is achieved not through the intrinsic monosaccharide preference of the CRDs, but rather through the specific arrangement of the multiple CRDs within the fully-folded DC-SIGN molecule. Reciprocally, the glycosylation pattern of PAPP-A1/proMBP could provide the spatial display of oligosaccharides in an ideal conformation for recognition by DC-SIGN.

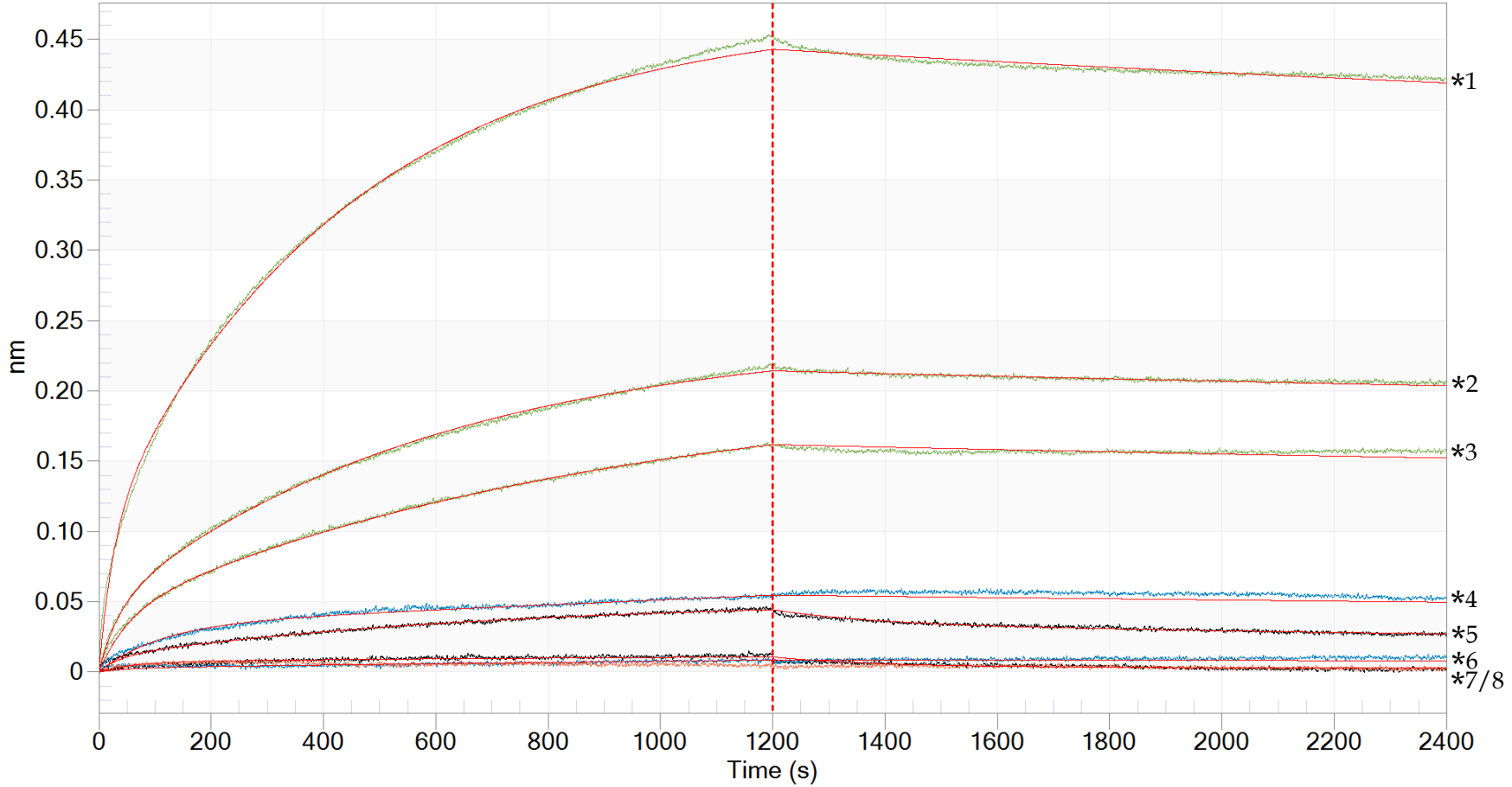


Figure 3.10: Comparative sensorgram of all three C-type lectins (DC-SIGN, DC-SIGNR, MMR) as analytes interacting with the human native PAPP- A1/proMBP-complex as immobilised ligand, respectively, at different concentrations (ranging from 100 nM to 50 nM). The negative control with EDTA confirms that the interaction is calcium specific. Annotations:

*[*1] 100 nM DC-SIGN, [*2] 75 nM DC-SIGN, [*3] 50 nM DC-SIGN, [*4] 100 nM DC-SIGNR, [*5] 100 nM MMR,*

*[*6] 50 nM DC-SIGNR, [*7] 50 nM MMR, [*8] 100 nM DC-SIGN in assay buffer with 5 mM EDTA*

4 Results and Discussion (II)
Structural Analyses of Complex Carbohydrate structures on Glycoproteins via Mass Spectrometry

4.1 Background

The biophysical interaction study discussed in section 3.3.1 established not only the novel interaction between DC-SIGN and the native human PAPP-A1/proMBP-complex but also demonstrated the dissociation constant K_D to be within the low nanomolar and, therefore, well within the physiological range of blood plasma levels of native human PAPP-A1/proMBP. Furthermore, the interactions are disrupted in the presence of EDTA, which chelates calcium ions. Therefore, the calcium-dependent interaction of glycans, present on the PAPP-A1/proMBP-complex, and the CRD of DC-SIGN has been described. Since this CRD has a binding preference for N-glycans of the mannose type, the structural analysis of the complex carbohydrate structures involved in this interaction is of great interest.

The native human PAPP-A1/proMBP-complex is abundant in the plasma of pregnant women, although the enzymatic domain of PAPP-A1 has been found to be physiologically inhibited by proMBP. Overgaard et al. have stated that the human PAPP-A1 carries glycan structures on 11 sites out of the 14 potential N-glycosylation sites, which follow the consensus sequence N–X–S/T (except P for X) (Overgaard et al., 2003).

Within this doctoral project, comprehensive analyses of the glycomic and glycoproteomic characteristics of PAPP-A1 samples were performed at the Analytical Service and Training Laboratory (Complex Carbohydrate Research Center, University of Georgia, USA), in collaboration with Dr Parastoo Azadi. Within this chapter, the glycomic and glycoproteomic characteristics of human recombinant proMBP, human recombinant PAPP-A1, and the human native PAPP-A1 derived from the PAPP-A1/proMBP-complex are described and analysed.

For schematic representation of glycoforms, annotations are simplified by using antennary structures with the symbol nomenclature established by the Consortium for Functional Glycomics (CFG) (The Consortium for Functional Glycomics, 2010).

4.2 Detailed example of glyco(proteo)mic MS data analysis and structural interpretation

The complex process of mass spectrometric data analysis and interpretation of glycan structures will be examined on one example in detail. However, the subsequent sections will be limited to describing the determinations reached through the process demonstrated within this specific section.

The MS data acquired were searched for modification by glycosylation via the software Byonic. The search for N-linked glycosylation is presented for proMBP (alternative name BMPG), which carries one sequon for N-glycosylation (figure 4.11. in section 4.3.1.1). The MS data set was searched against specific parameters, such as the coverage of peptides, the amino acids expected to be cut during selective proteolytic digest, the assumed number of missed enzymatic cleavages, the tolerance of mass error, the quality of single spectra, expected contaminants, expected modifications and maximal quantity, databases for reference, etc. Table 17 provides an overview of the selected parameters/rules used by Byonic and shows exemplary values. Figure 4.1 shows the visualised result found for an N-glycan search in proMBP MS data.

Table 17: Overview of selected parameters (rules) to search for in mass spectrometric data.

Rule	**Value**
Cleavage residues (amino acid one-letter-code)	RK
Digest cutter	C-terminal cutter
Peptide termini	Semi specific
Maximum number of missed cleavages	4
Precursor tolerance	10.0 ppm
Fragment tolerance (HCD)	10.0 ppm
Fragment tolerance (ETD)	10.0 ppm
Charges applied to charge-unassigned spectra	1,2,3,4
Precursor mass max.	10000.0
N-glycan search	common 2
Skip bad spectrum	true
Off by x isotopes	-2, -1, 0, +1, +2
Contaminants added	false
Decoys added	true

Table 17: Overview of selected parameters (rules) search for in mass spectrometric data. (continued)

Rule	Value
Modification searches	
Common modifications max.	5
Rare modifications max.	2
Carbamidomethyl / +57.021464 @ C	common 2
Oxidation / +15.994915 @ M	common 2
Glycan Databases	N-glycan 309 mammalian no sodium @ NGlycan | common 2

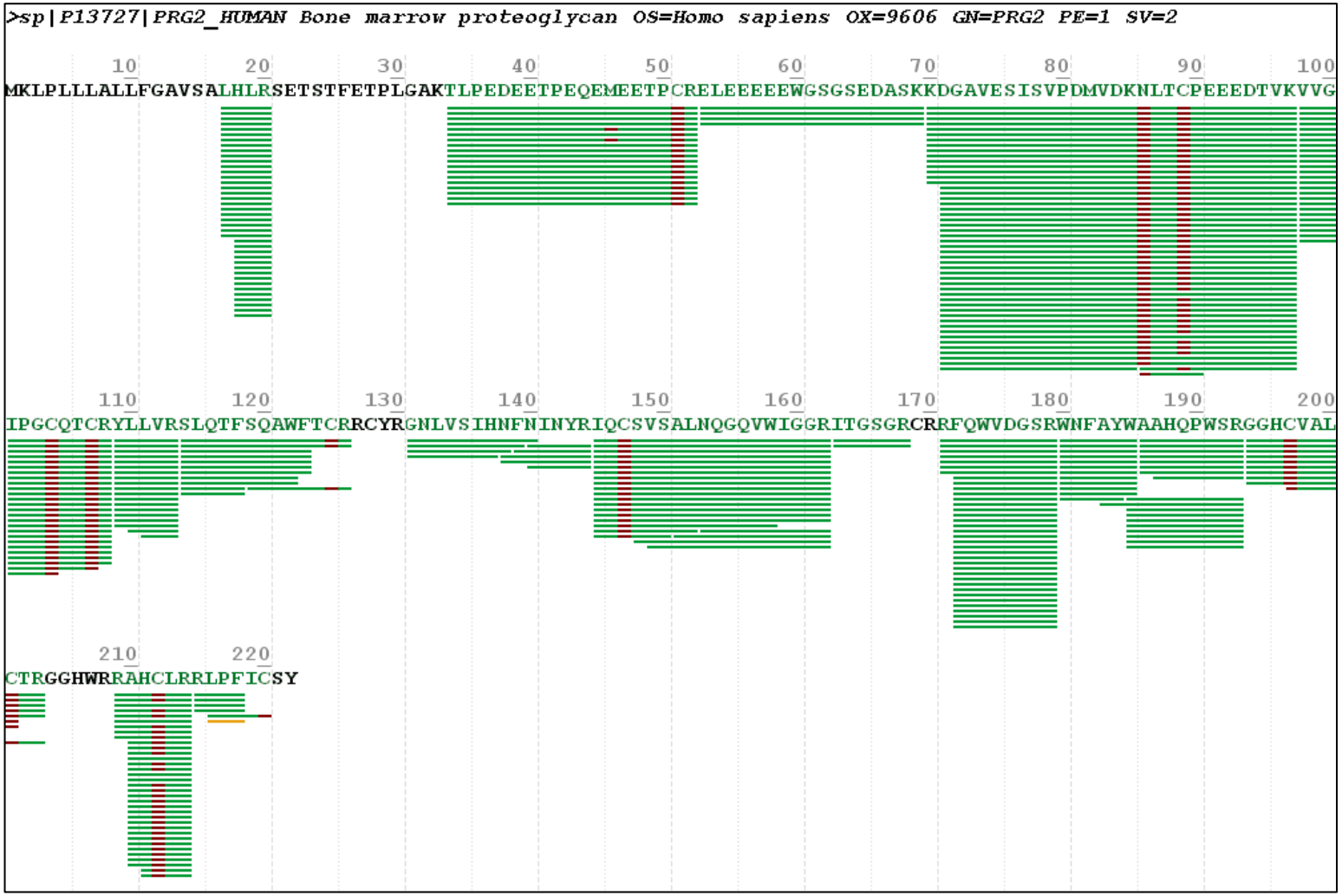


Figure 4.1: Visualised data of peptide coverage (81.1%) and N-glycan modifications retrieved from MS data analysis of proMBP (alternative name BMPG). The first line names the peptide sequence searched against (in FASTA format). The sequence of amino acids is numbered in blocks of ten. The amino acids/peptides depicted in black were not covered. The amino acids/peptides depicted in green were covered by the software search through MS data. The number of spectra/scans assigned is represented in green bars/rows below the sequence. The red highlights indicate a modification, e.g. carbamidomethylation of cysteine (C), oxidation of methionine (M) or glycosylation of an asparagine within a sequon (N-X-T/S, X not P).

The rows, depicted below the peptide sequence of proMBP/BMPG in figure 4.1., represent single spectra and were assigned by Byonic. The assignment of the programme is based on theoretical data that fragment ions of a given glycopeptide would present, i.e. diagnostic ions. If the data matches with the acquired data, the peaks are annotated by Byonic accordingly. A successfully annotated spectrum through Byonic's algorithm is shown representatively in figure 4.2.

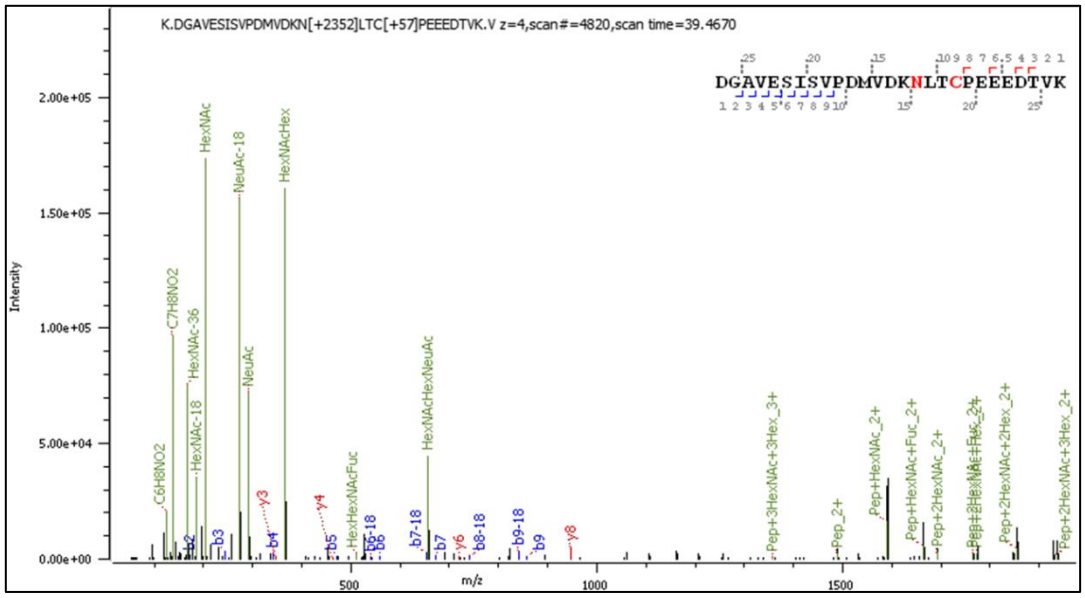


Figure 4.2: HCD MS² spectrum of the glycopeptide DGAVESISVPDMVDKNLT-CPEEEDTVK (aa 71-97) derived from proMBP as presented in the software Byonic. The spectrum shows scan number 4820, which derived from sample eluted at retention time point 39.4670 min from the nanoLC. The highlighted and annotated C within the peptide sequence shows that the cysteine is carbamdiomethylated, i.e. 57.021 Da heavier than an unmodified cysteine. The highlighted and annotated N represents a modification of the asparagine with a putative glycan structure of about 2352 Da. Numerically matched b- (in blue) and y-ions (in red) of the peptide are marked in the associated spectrum as well as peptide sequence. MS fragment ions matching typical diagnostic ions derived from glycan structures are highlighted and annotated in green within the spectrum. Abbreviations: b – b-ion, Fuc – fucose/pentose, Hex – hexose, HexNAc - N-acetylhexosamine, NeuAc – sialic acid or N-acetylneuraminic acid, Pep – peptide, y – y-ion.

The HCD fragmentation method causes low-energy bonds to cleave, such as in the backbone of peptide ions that yields b- and y-ions (Zubarev et al., 1998). Diagnostic ions can be derived from the peptide, e.g. the b- and y-ions of fragmented peptide species, or from the attached glycan, e.g. glycan oxonium ions. As a first step towards the verification of the diagnostic ions, the original HCD spectrum is investigated (shown in figure 4.3.). The pattern of the peaks is identical for both spectra (figure 4.2. and figure 4.3.).

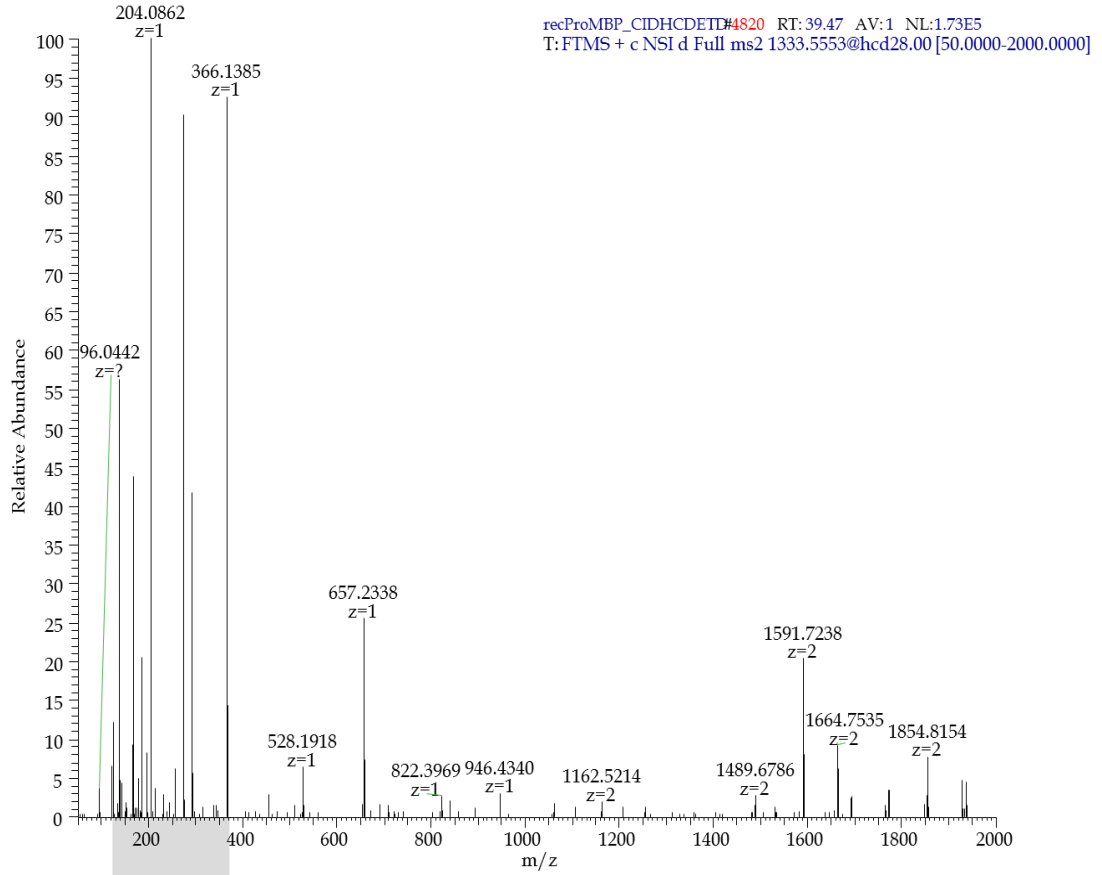


Figure 4.3: Original HCD MS² spectrum (m/z range 50-2000; tandem MS step of peak m/z 1333.5553 from the full MS) of the glycopeptide DGAVESISVPDMVDKNLT-CPEEEDTVK as presented in the software Xcalibur from 'Thermo Scientific LC-MS systems'. The spectrum shows scan number 4820 that derived from sample eluted at retention time (RT) point 39.47 min from the nanoLC.

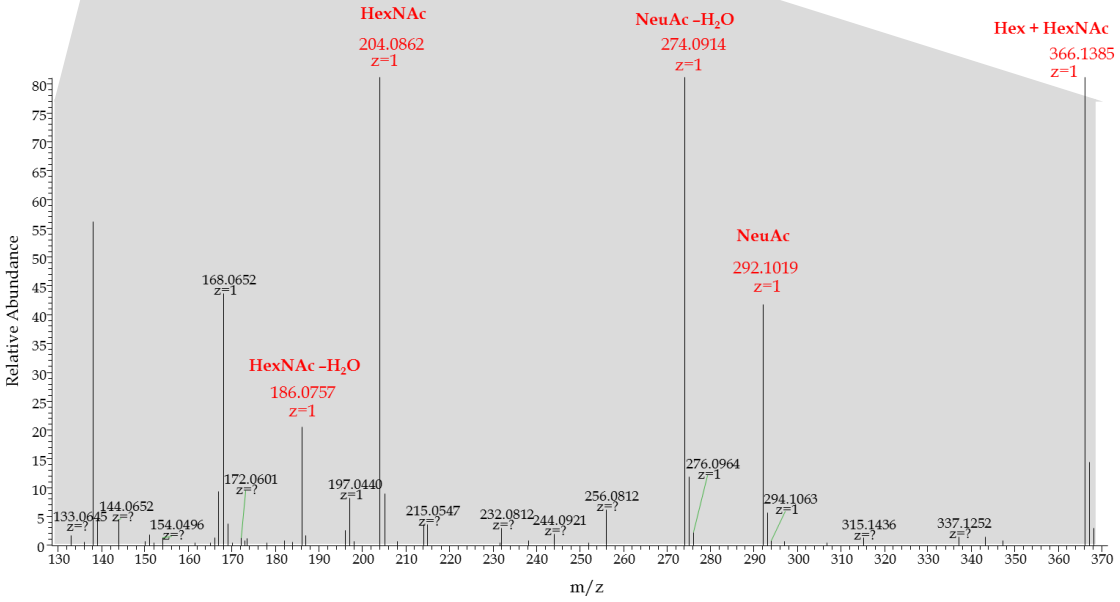


Figure 4.4: HCD MS² spectrum (m/z range 120-370) of MS scan (number 4820) of the glycopeptide DGAVESISVPDMVDKNLTCPEEEDTVK from proMBP. Diagnostic ions derived from glycans, i.e. oxonium ions, are highlighted in red and annotated with their m/z as well as charge state (z). Abbreviations: Hex – hexose, HexNAc - N-acetylhexosamine, NeuAc – sialic acid or N -acetylneuraminic acid, H_2O – water molecule

Figure 4.4. shows a lower m/z range (m/z 120-370) of HCD spectrum number 4820 (figure 4.3) enlarged. Diagnostic ions originating from glycan structures (i.e. oxonium ions from hexoses [Hex], sialic acid residues [NeuAc], and hexosamines [HexNAc]), are strong indicators that the MS spectrum includes information derived from a glycosylated peptide of the protein of interest. The original software labelling (top right corner of the spectrum in figure 4.3.) also reveals that the tandem MS scan was performed on an ion with m/z 1333.5553 that was produced during the full MS fragmentation step.

Biomacromolecules contain naturally occurring stable isotopes of elements, such as carbon and nitrogen. For example, the ^{13}C isotope of carbon occurs with a natural abundance of about one percent. Small, human N-glycan structures contain more than 50 carbon atoms, and any ^{13}C isotope present instead of a ^{12}C isotope adds 1 Da in mass. Ion signal peaks for ions of various isotopic compositions are detectable; for example, the enlarged m/z scale in figure 4.4. enables the observation of three peaks for one sialic acid fragment ion (m/z 291.0954 – 294.0964) versus the labelled average mass peak of m/z 292.1019. However, no ion signal peak exists at the average mass. This is also the cause of the rule "Off by x isotopes" being applied to the search in mass spectrometric data (see table 17). The software programmes Byonic, as well as Xcalibur, annotate peaks numerically with the average mass that takes into account all isotopic peaks. The annotation of the average mass but not the monoisotopic mass peak is evident throughout the following sections and is the cause of the difference between the annotated numeric values of peaks by some daltons from the calculated masses. The calculations performed with software tools required mainly monoisotopic mass values.

The larger the glycan structure and/or the glycopeptide, the more atoms can be naturally substituted with stable isotopes, and the more prevalent will be the numeric difference between the average mass and the monoisotopic mass.

For further analysis, the full mass spectrum was taken into account. It provided insight into the charge states of the glycopeptides detected. Since the spectra are always presented as relative abundance in relation to mass over charge (m/z), the information is helpful for identification of the numeric range of detection. Furthermore, the molecular mass of the original molecule can be re-calculated from m/z and results in a mass of 5334.2188 Da for m/z 1333.5547 with z=4. The HCD spectrum (figure 4.3.) shows the tandem fragmentation of precursor ion m/z 1333.5553 that is identifiable in the full mass spectrum in figure 4.5.

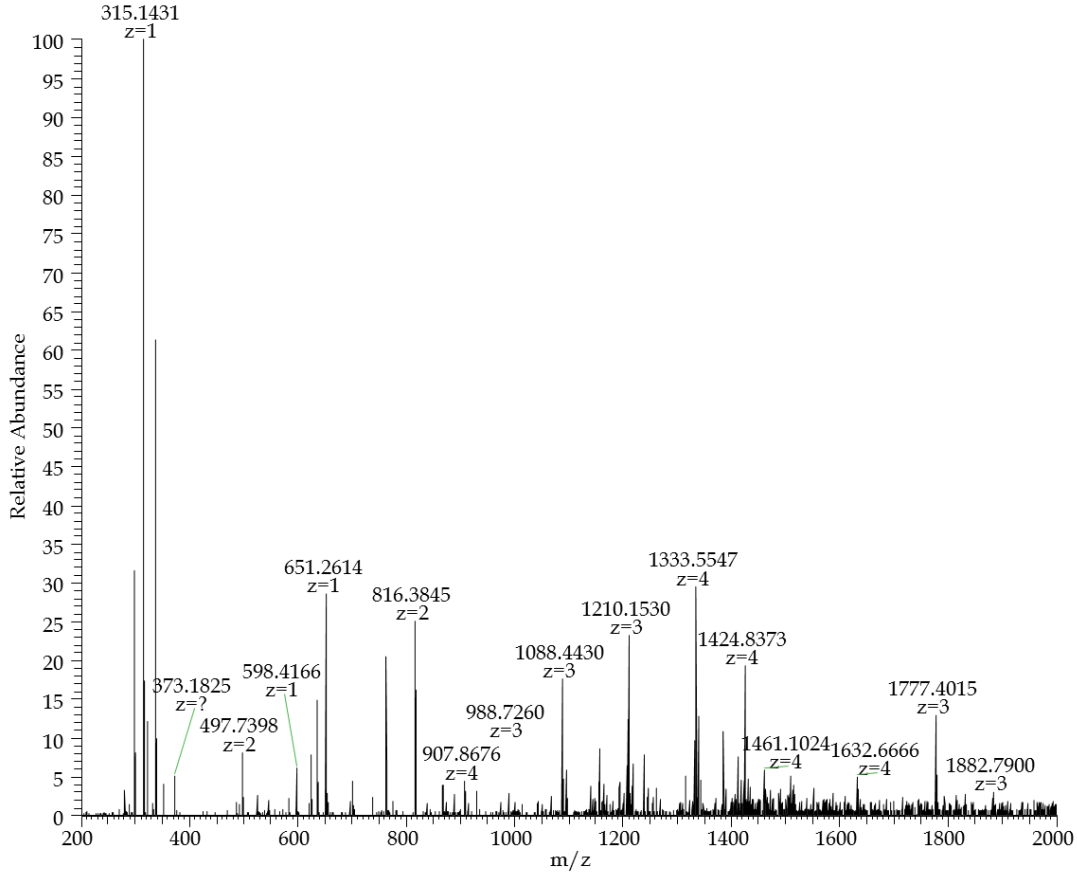


Figure 4.5: Full mass spectrum (m/z 200-2000, MS scan number 4815) of the full MS scan that derived from the eluted sample of (glyco)peptides of proMBP at retention time 39.43 min from the on-line nanoLC. The average mass peak at m/z 1333.5547 is clearly identifiable, has a charge state of z=4, and is surrounded by multiple other peaks that either derive from isotopic ions of the same molecule or fragment ions from other molecules. The determination of the molecular mass of the original molecule after charge deconvolution of peak m/z 1333.5547 results in a theoretical mass of 5334.2188 Da.

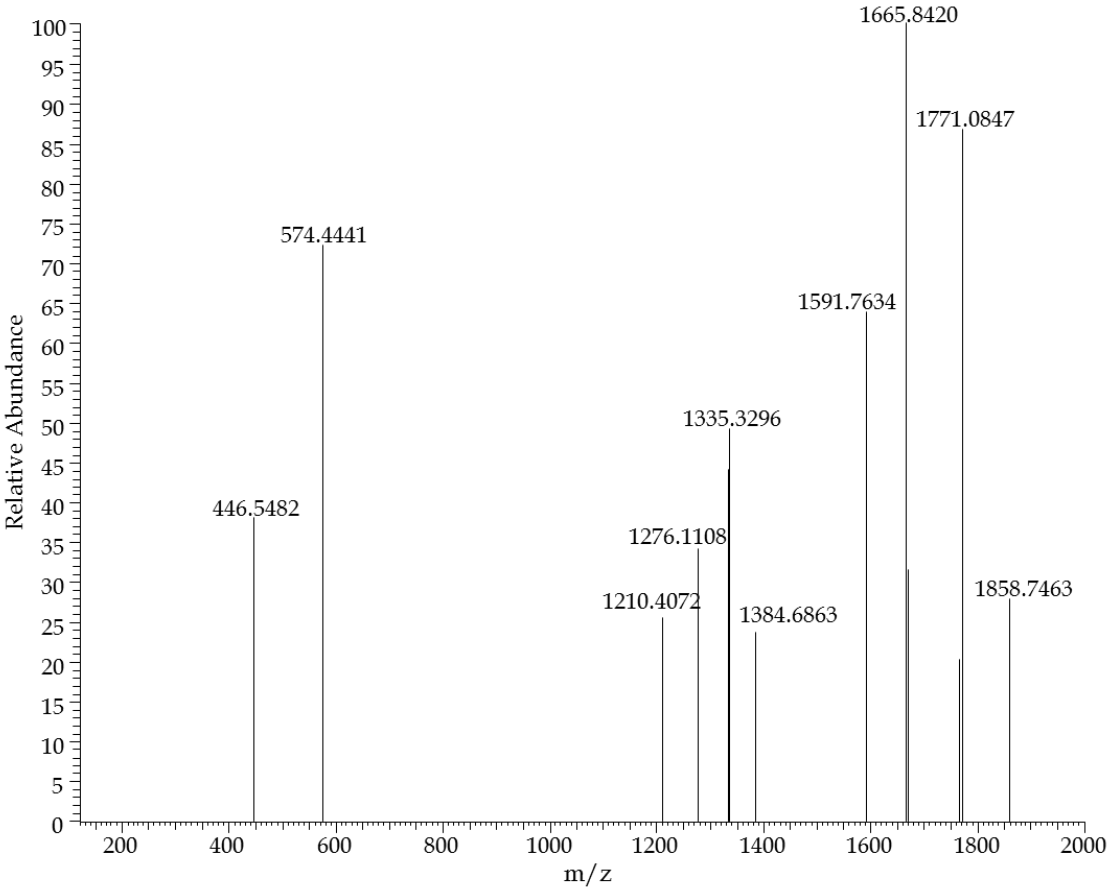


Figure 4.6: ETD MS^2 spectrum (m/z 120-2000) of precursor ion m/z 1333.5553 (full MS scan) derived from the glycopeptide DGAVESISVPDMVDKNLTCPEEEDTVK from proMBP. The retrieved m/z values of a fragmented glycopeptide reflect peptide ions with glycan modification. Ideally, the glycan remained unmodified throughout the ETD fragmentation process. However, the information is complex, since the glycan mass itself is unknown at this stage of the glycoproteomic MS data analysis.

The electron-transfer dissociation (ETD) fragmentation method targets the fragmentation of peptide bonds with higher energy than the HCD and CID methods. This causes the stronger amide bonds to cleave and yields c-ions and z-ions of the peptide with minimal fragmentation of the glycosylation (Zubarev et al., 1998). Empirically, not all glycopeptides fragment with the ETD method in a way that contributes easy-to-access information, since the m/z values reflect peptide ions that, ideally, contain an unmodified but still unidentified glycoform with an unknown mass. None of the analyses within the following sections was based on an ETD spectrum for this project. The ETD spectrum of the tandem fragmentation of the precursor ion m/z 1333.5553 derived from the full mass fragmentation is depicted in figure 4.6.

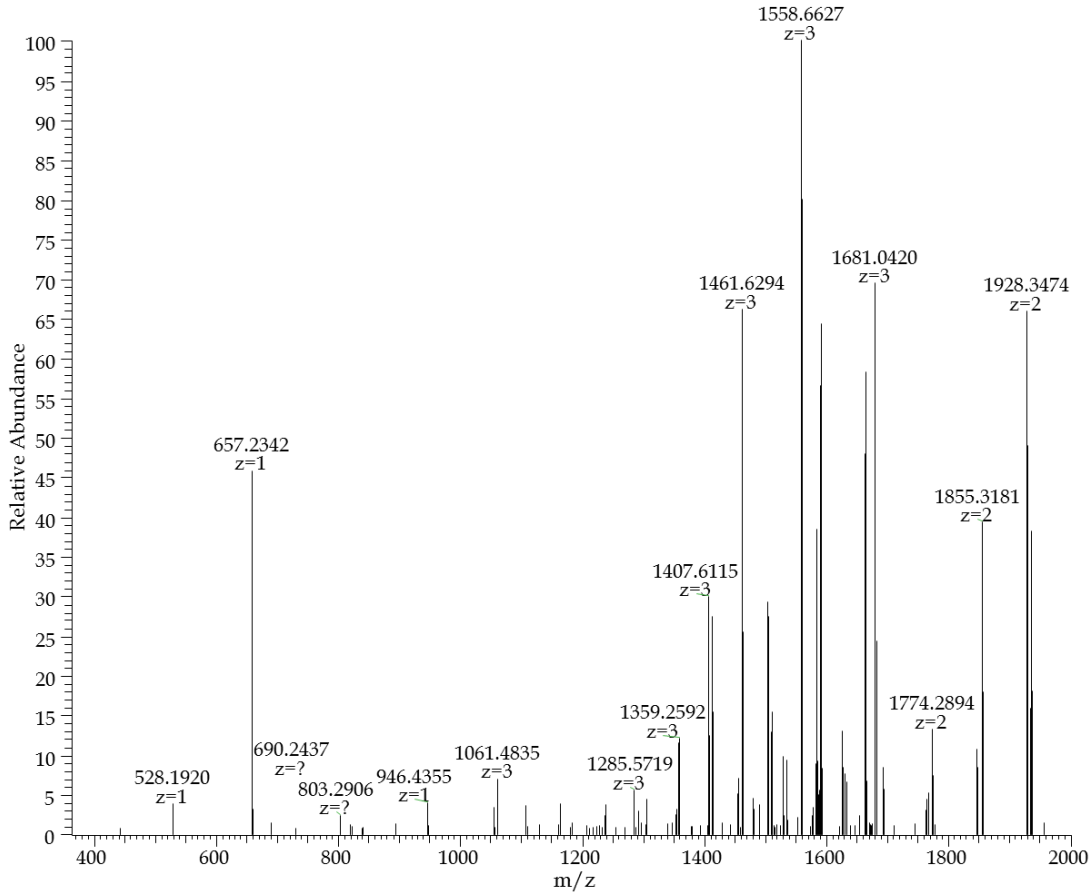


Figure 4.7: CID MS² spectrum (m/z 360-2000) of the precursor ion m/z 1333.5553 (full MS scan) derived from the glycopeptide DGAVESISVPDMVDKNLTCPEEEDTVK from proMBP. The ion peaks represent multiply charged ions as annotated accordingly. Figure 4.8. shows the same CID scan but with charge deconvolution applied, which simplifies the identification of glycoforms.

The method of CID fragmentation preferentially targets glycosidic bonds over peptide bonds in a precursor ion deriving from a glycopeptide and yields mainly glycan oxonium ions but also minimal b-ion and y-ion coverage. CID is a method using even lower energy than HCD fragmentation and realises stepwise loss of single carbohydrate residues that enables determination of glycoforms including their structural characterisation, but does not allow for linkage information (Mechref, 2012).

However, the pattern of the spectra in CID (figure 4.7.) and HCD (figure 4.3) are comparable.

The detailed analysis of this CID spectrum, deriving from a tandem fragmentation of the precursor ion at m/z 1333.5553 in the full MS fragmentation, is presented in a deconvoluted spectrum in figure 4.9.

Noteworthy, the term 'deconvolution' has several meanings within the field of mass spectrometry. Here, the practice of charge deconvolution is defined as follows: A mass spectrum containing ions with various charge states (z=1/2/3/4/etc.) is mathematically processed to retrieve only singly charged species (z=1). This does not only simplify the spectrum but also reveals the molecular mass (m) of the original molecule, when z=1 in m/z values (Mallet and Down, 2009).

Figure 4.8. shows the deconvoluted full MS spectrum: The deconvolution significantly simplifies the spectrum, as depicted in figure 4.8, in direct comparison with the original full MS spectrum (figure 4.5.). The average mass peaks at m/z 5040.0911, m/z 5185.1614, m/z 5331.1908, and m/z 5459.2920 have been identified as glycopeptides with different glycoforms. All of these glycopeptides were eluted from the nanoLC at the same time, fragmented in the full MS scan, and contributed to diagnostic ions identified. As already calculated on page 130, one glycoform of the original molecule species correlates closely with the calculated mass of 5334.2188 Da.

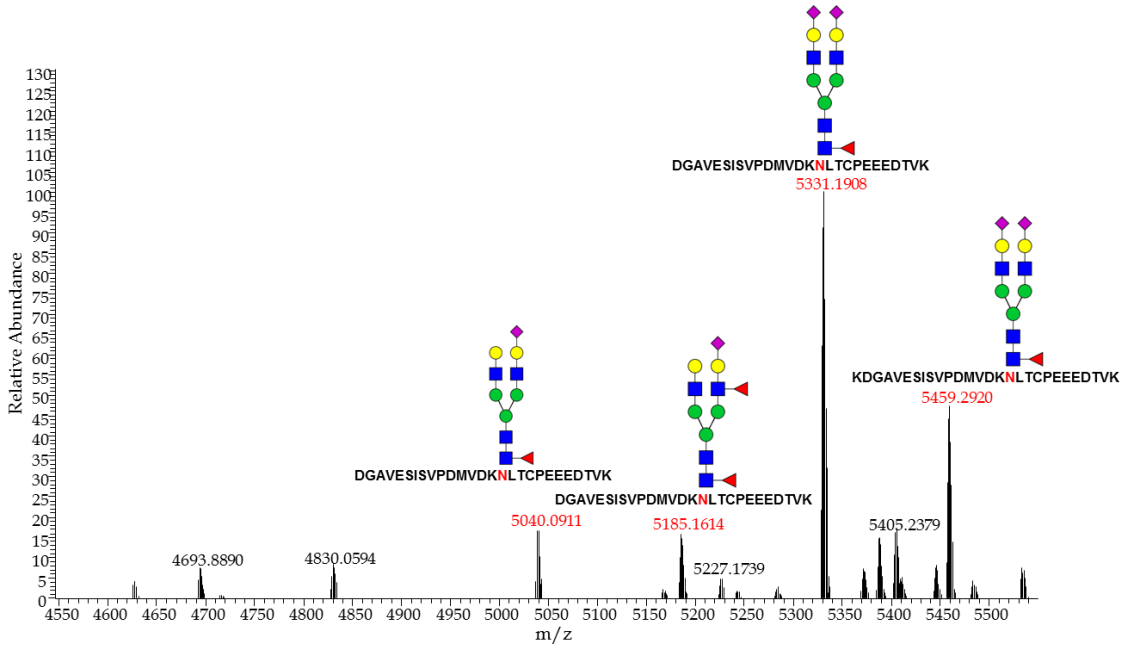


Figure 4.8: Deconvoluted full MS spectrum (m/z 4550-5540) derived from the glycopeptide DGAVESISVPDMVDKNLT-CPEEEDTVK from proMBP. Four glycoforms have been identified at m/z 5040.0911, m/z 5185.1614, m/z 5331.1908, and m/z 5459.2920. The detailed process of structural assignment by CID fragmentation is shown exemplary for m/z 5040.0911 in figure 4.9.

Symbols: GalNAc, GlcNAc, Man, Glc, Gal, Fuc, NeuAc

Figure 4.9. shows the exact process of matching the species at m/z 5040.0911 from the deconvoluted full MS spectrum in figure 4.8. with a specific glycoform. The CID spectra are ideal to show loss of neutral sugar moieties from the glycan. This process also enables the structural assignment of the bi-antennary glycoform that belongs to the complex type with core fucosylation and terminal sialyation.

Demonstrated peptide sequences carrying putative N-glycans are referred to as '(glyco)peptide' within the following sections. The sequences are presented in FASTA peptide format, e.g. XXXNX (with X being any amino acid). The sequences are also presented as (X)XXXNX(X) with the highlighted amino acid (N) depicting the asparagine within the N-glycan sequon, which was found to be glycosylated via analysis of glycan site attachment. The amino acids in brackets represent the sites after/before cleavage by either trypsin or chymotrypsin, respectively.

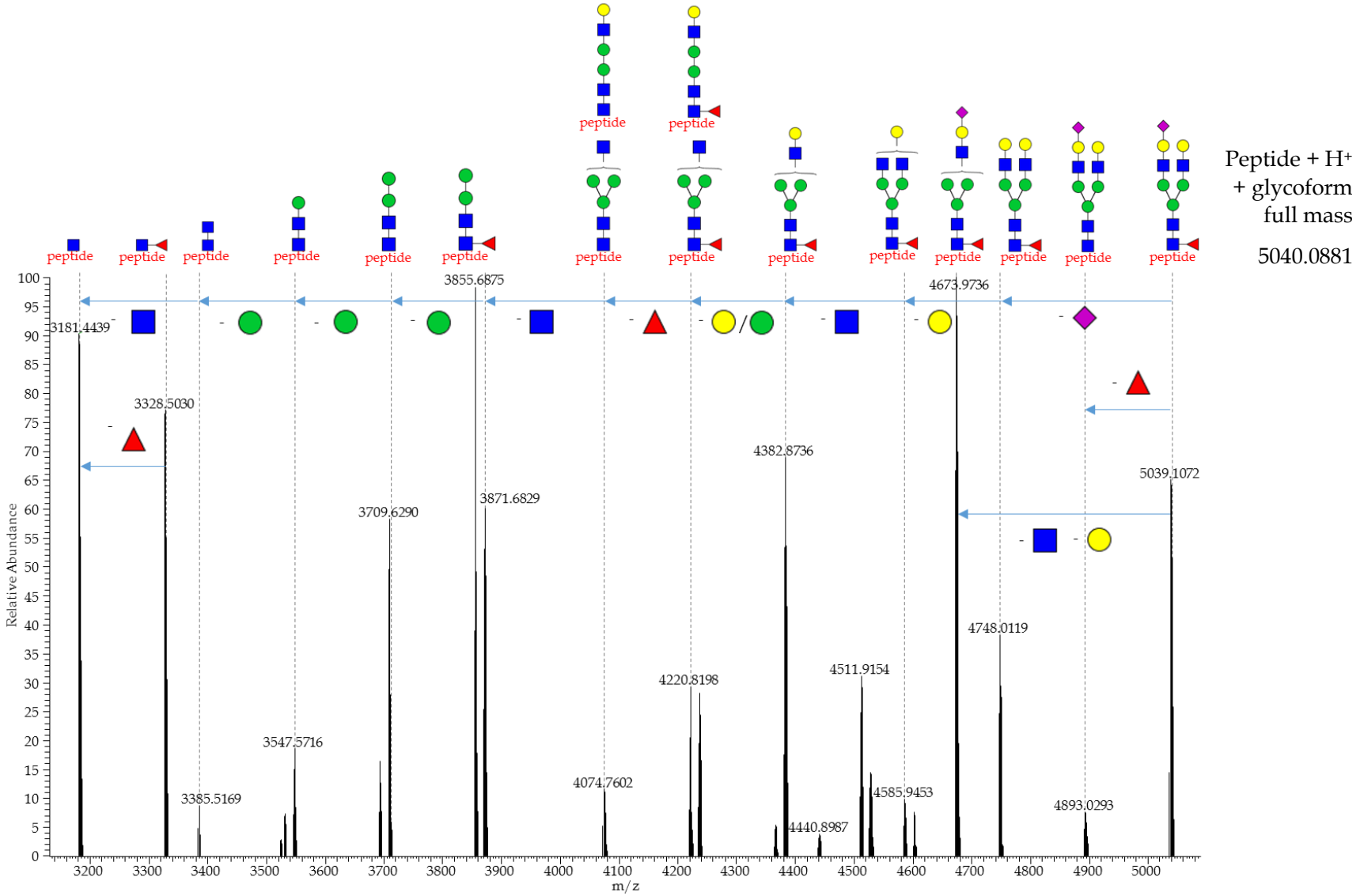


Figure 4.9: Deconvoluted CID MS² spectrum (m/z 3140-5080) of the precursor ion m/z 1333.5553. This spectrum illustrates the final steps for the structural assignment of one of the glycoforms attached to the asparagine (red) within the sequon in the peptide DGAVESISVPDMVDKNLTC(57.02146)PEEEDTVK of proMBP. The neutral loss of carbohydrate moieties is demonstrated.

Symbols: GalNAc, GlcNAc, Man, Glc, Gal, Fuc, NeuAc

4.3 Recombinant human proMBP

The sample investigated within this section is a recombinant preparation of the mature Proform of Eosinophil Major Basic Protein (proMBP; alternative names Bone Marrow Proteoglycan [BMPG], Proteoglycan 2 [PRG2]) that consists of 206 amino acids, since the signal peptide (aa 1-16) has been cleaved off before secretion by the eukaryotic expression system (expressed in human cells by the manufacturer). Remarkably, the propeptide (aa 17-105) sequence has acidic properties (isoelectric point [pI] around 4.0), whereas the pI of the sequence of the mature MBP lies at 11.0. Given that proMBP contains the sequences of the propeptide as well as the main chain, the pI is predicted and observed to lie around 6.1 (Wasmoen et al., 1988). The significantly different chemical characteristics underpin the foundation for diverse functions of proMBP and MBP, respectively. ProMBP is known to be a physiological proteinase inhibitor of PAPP-A1 during pregnancy (Overgaard et al., 2000). The full protein sequence for proMBP (protein ID P13727, FASTA format) shown in figure 4.10, was retrieved from the Uniprot database (The UniProt Consortium, 1990).

Since the amount of sample (50 µg) was neither sufficient to cover glycomic analysis nor analysis of the glycosaminoglycan attached, only glycoproteomic analysis for N-linked and O-linked glycans was carried out.

The study was conducted at the Analytical Services and Training Laboratory at the Complex Carbohydrate Research Center.

```
MKLPLLLALL FGAVSALHLR SETSTFETPL GAKTLPEDEE TPEQEMEETP
CRELEEEEEW GSGSEDASKK DGAVESISVP DMVDKNLTCP EEEDTVKVVG
IPGCQTCRYL LVRSLQTFSQ AWFTCRRCYR GNLVSIHNFN INYRIQCSVS
ALNQGQVWIG GRITGSGRCR RFQWVDGSRW NFAYWAAHQP WSRGGHCVAL
CTRGGHWRRA HCLRRLPFIC SY
```

Figure 4.10: FASTA peptide sequence of human proMBP. The sequence includes the signal peptide sequence (aa 1-16; highlighted in dark grey) and the propeptide sequence (aa 17-105; highlighted in light grey). The mature proMBP consists of 206 amino acids with 1 potential N-glycosylation site (sequon N-X-S/T, X not P, highlighted in yellow). Amino acids (T23, S24, T25, T34) that have been found to be O-glycosylated in past studies are highlighted in green. Furthermore, serine 62 (highlighted in magenta) has been found to be modified by a glycosaminoglycan (Oxvig et al., 1994).

4.3.1 Glycoproteomic analysis of N-linked glycans on human recombinant proMBP

The proMBP investigated for its N-linked glycosylation modifications had been recombinantly expressed in human cells (HEK293).

The proMBP sample was pure enough to be directly processed with trypsin for enzymatic digestion. The expected proteolytic cleavages were therefore amino acids with a positive charge, i.e. lysine (K) or arginine (R), from the C-terminal side, if not followed by a proline (Baird and Craik, 2013). An *in silico* digest with PeptideCutter predicted the best theoretical peptide coverage for mass spectrometry with trypsin. Peptides are required to have a length between 5-30 aa to lie within the detection limit.

The subsequent acquisition of MS data for glycoproteomic analysis of the (glyco)peptides was carried out with the nanoLC-MS/MS method described in section '2.6 Methods for Mass Spectrometry'. For the data analysis, the software programmes Xcalibur and Byonic were used. Byonic quantified the peptide coverage with 81.1% confidence with respect for search parameters set for N-glycan modification, carbamidomethylation of cysteines and oxidation of methionines. Most importantly, the peptide carrying the sequon NTL was well within the detectable range with a theoretical glycopeptide length of 12 amino acids.

Oxvig et al. (1994) showed in the past that an "N-linked glycan is bound to Asn-86" in proMBP considering the whole FASTA peptide sequence including the propeptide.

4.3.1.1 nLC-NSI-MS/MS analysis of N-linked glycoforms on glycopeptides from human recombinant proMBP

For analysis of N-glycosylation on peptides derived from the Proform of Eosinophil Major Basic Protein (proMBP), the method of nano-scale liquid chromatography (nanoLC) with on-line tandem mass spectrometry (MS/MS) was performed, as stated in section '2.6 Methods for Mass Spectrometry'.

Briefly, the sample derived from a tryptic digest of proMBP and the resulting peptides as well as glycopeptides were separated via nanoLC for better resolution, i.e. a significantly reduced amount of co-elution and co-ionisation of different glycopeptides for mass spectrometry. After ionisation by NSI, full MS scans were acquired, and precursor ions selected via various criteria. The selected precursors were then processed by tandem MS for parallel CID and HCD fragmentation. Diagnostic ions derived from glycan structures triggered further analysis via a subsequent fragmentation method, namely ETD. The glycoproteomic approach aimed to illuminate site-specific information of glycosylation, such as the site of the glycan attachment and prevalent glycoforms.

Since the proMBP was sourced from human cells, the protein was expected to be partially processed and the signal peptide not to be present. All forthcoming references to amino acids represent the mature FASTA peptide sequence of human proMBP without the signal peptide. The database Uniprot lists the FASTA peptide sequence of human proMBP with one N-glycosylation being present within the 206 amino acids long proMBP, as shown in figure 4.11. (The UniProt Consortium, 1990).

```
LHLRSETSTF ETPLGAKTLP EDEETPEQEM EETPCRELEE EEEWGSGSED
ASKKDGAVES ISVPDMVDKN LTCPEEEDTV KVVGIPGCQT CRYLLVRSLQ
TFSQAWFTCR RCYRGNLVSI HNFNINYRIQ CSVSALNQGQ VWIGGRITGS
GRCRRFQWVD GSRWNFAYWA AHQPWSRGGH CVALCTRGGH WRRAHCLRRL
PFICSY
```

Figure 4.11: FASTA format of the peptide sequence of human proMBP without the signal peptide. The mature proMBP consists of 206 amino acids, which includes the propeptide (aa 1-89, highlighted in grey). The propeptide carries one sequon (N-X-S/T, X not P, highlighted in yellow) that is a putative site for N-glycosylation.

a) Amino acid asparagine-70

The programme PeptideCutter was used to predict the potential cleavage sites around the sequon NLTC at asparagine 70 (N70) for mature human proMBP. This resulted in the (glyco)peptide (K)NLTCPEEEDTVK(V) for an *in silico* digestion with trypsin. However, two distinct glycopeptides, (K)DGAVESISVP-DMVDKNLTCPEEEDTVK(V) as well as (K)KDGAVESISVP-DMVDKNLTCPEEEDTVK(V) were identified that cover this putative N-glycan site. For both (glyco)peptides found, trypsin missed the cleavage at amino acid lysine 69 (L69). For a glycosylated asparagine 70, the steric hindrance could be the reason for the missed cleavage site by trypsin. Interestingly, proMBP carries two serial lysine residues at amino acid 53 (L53) and 54 (L54), which results in the two (glyco)peptides mentioned above that only differ by one amino acid.

The glycosylation site at asparagine 70 was found to be occupied with seven different glycoforms that are classified as either complex type with fucosylation and sialyation (figure 4.12), or bisected type with terminal sialic acid residues (figure 4.13 and figure 4.14). Annotated structures are based on the molecular weight, MS data, the N-glycan biosynthesis pathway and prevalent glycoforms in humans.

The data was gathered via on-line nanoLC-MS/MS. All figures show a full MS spectra.

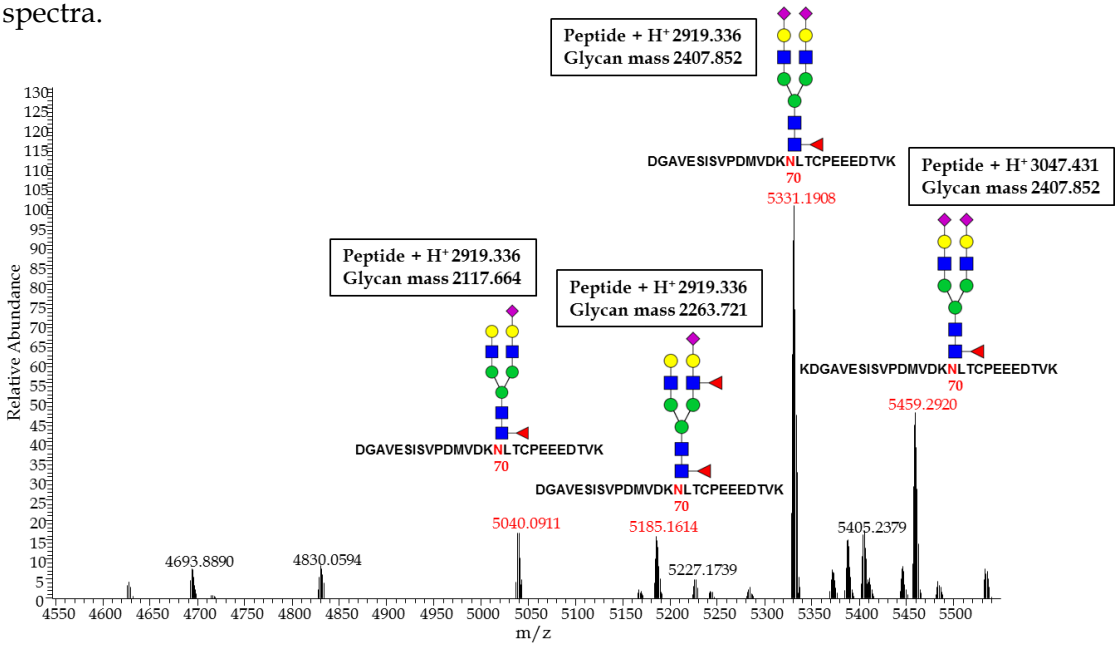


Figure 4.12: Deconvoluted and annotated full MS spectrum of glycopeptides (K)DGAVESISVPDMVDKNLTCPEEEDTVK(V) and (K)KDGAVESISVPDMVDKNLT-CPEEEDTVK(V) (m/z 4550-5550) derived from human recombinant proMBP. Bi-antennary glycans of the complex type with fucosylation and sialylation (m/z 5040.0911, m/z 5185.1614, m/z 5331.1908, m/z 542920) were found to be present on the asparagine 70 within the NTL sequon. Symbols: GalNAc, GlcNAc, Man, Glc, Gal, Fuc, NeuAc

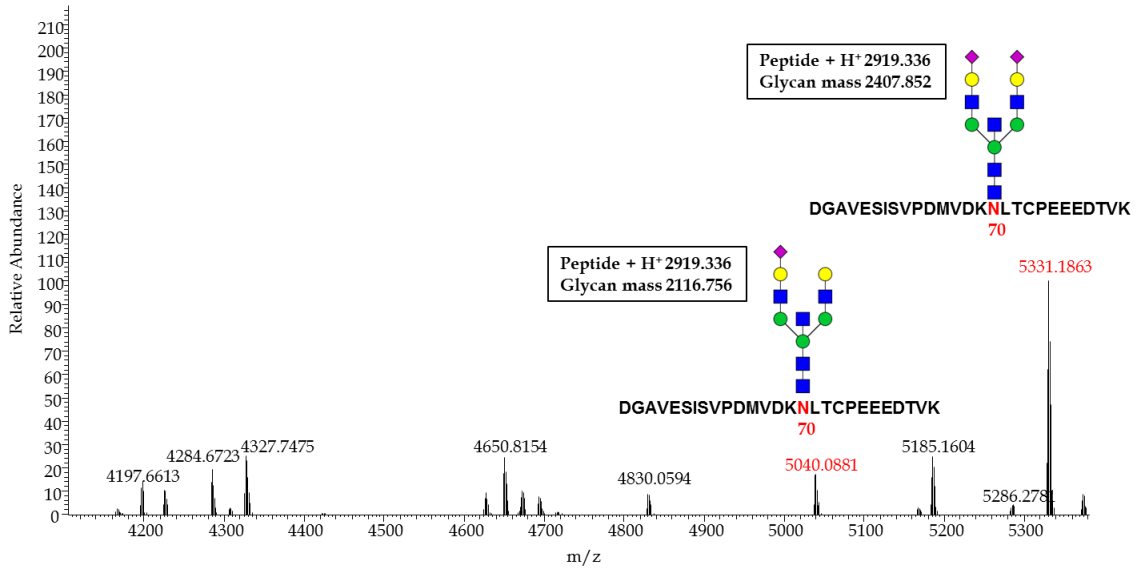
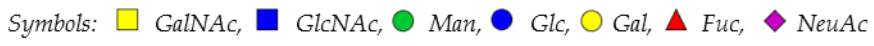


Figure 4.13: Deconvoluted and annotated full MS spectrum of the glycopeptide (K)DGAVESISVPDMVDKNLTCPEEEDTVK(V) (m/z 4100-5380) derived from human recombinant proMBP. Bi-antennary N-glycans of the bisected, complex type with fucosylation and sialylation (m/z 5040.0911, m/z 5331.1908) were found to be present on asparagine 70 of a subset of glycopeptides from proMBP.



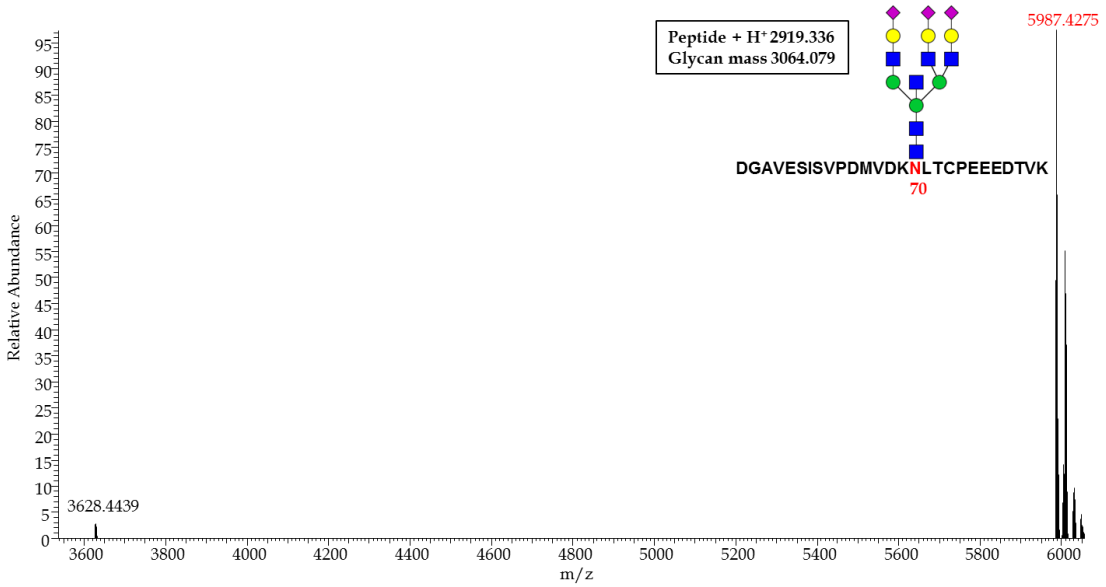


Figure 4.14: Deconvoluted and annotated full MS spectrum of the glycopeptide (K)DGAVESISVPDMVDKNLTCPEEEDTVK(V) (m/z 3575-6025) derived from human recombinant proMBP. A tri-antennary glycan of the bisected, complex type with terminal sialylation (m/z 5987.4275) was found as one of the glycoforms on asparagine 70 from proMBP.

4.3.2 Glycoproteomic analysis of O-linked glycans on human recombinant proMBP

O-linked glycans are known to be attached to peptides by amino acids with a hydroxyl group side chain, such as serine (S) or threonine (T) residues within the protein sequence. O-glycosylation does not occur on distinctive sequons as N-glycosylation does. However, domains rich in serines, threonines and prolines have been shown to be modified with O-linked glycans of the mucin-type. They tend to be shorter regarding the degree of glycan extension, as well as showing a lower degree of branching. Studies have shown that the majority of O-linked glycans in human blood plasma proteins are short mucin-type glycans classified as structural core 1 and core 2 (Hoffmann et al., 2016). Mucin-type glycans share a core N-acetylgalactosamine (GalNAc) residue linked to the peptide backbone and can be composed of GalNAc, GlcNAc, galactose, fucose and sialyl residues but not mannose or glucose residues (Brockhausen and Stanley, 2015).

All references to amino acids represent the mature FASTA peptide sequence of human proMBP without the signal peptide, since the proMBP was sourced from human cells. The database Uniprot lists the FASTA peptide sequence of human proMBP with four sites of known O-glycosylation, which are highlighted in figure 4.15 (The UniProt Consortium, 1990).

```
LHLRSETSTF ETPLGAKTLP EDEETPEQEM EETPCRELEE EEEWGSGSED
ASKKDGAVES ISVPDMVDKN LTCPEEEDTV KVVGIPGCQT CRYLLVRSLQ
TFSQAWFTCR RCYRGNLVSI HNFNINYRIQ CSVSALNQGQ VWIGGRITGS
GRCRRFQWVD GSRWNFAYWA AHQPWSRGGH CVALCTRGGH WRRAHCLRRL
PFICSY
```

Figure 4.15: FASTA format of the peptide sequence of human proMBP without the signal peptide. The mature proMBP is known to be fully or partially glycosylated on threonine 7, serine 8, threonine 9 and threonine 18 (aa positions with signal peptide would be threonine 23, serine 24, threonine 25 and threonine 34).

Oxvig et al. (1994) showed that proMBP contains O-linked glycans that have been localised to amino acid threonine 7 (partially substituted), serine 8 (fully substituted), threonine 9 (fully substituted), and threonine 18 (partially substituted). Furthermore, they identified an O-linked glycosaminoglycan that modifies serine 46. Their analysis revealed an enormous heterogeneity of the glycan profile present on proMBP.

4.3.2.1 nLC-NSI-MS/MS analysis of O-linked glycoforms on glycopeptides from human recombinant proMBP

For analysis of O-linked glycans on peptides derived from the Proform of Eosinophil Major Basic Protein (proMBP), the same methods were used as for N-linked glycan analysis: After the tryptic digest of proMBP, the (glyco)peptides were separated via on-line nanoLC, ionised by NSI, a full MS acquired and then precursor ions processed through tandem MS for parallel CID and HCD fragmentation. Diagnostic HCD-derived ions from glycan structures triggered further analysis via a subsequent ETD fragmentation.

a) Amino acids serine 5, threonine 7, serine 8, threonine 9, threonine 12

In the literature and databases, amino acids threonine-7, serine-8, threonine-9 (amino acid positions with signal peptide: threonine-23, serine-24, threonine-25) are known to be modified with O-glycans (The UniProt Consortium, 1990). The programme PeptideCutter predicted the glycopeptide (R)SETSTFETPLGAK(T) that contains theses amino acids in the sequence of the mature human proMBP. The glycopeptide was confirmed via b-ions and y-ions in tandem MS spectra. All serines and threonines within the sequence are highlighted, since the identified glycoforms could be located on any of these. A detailed evaluation of the exact site of glycosylation within a glycopeptide could be achieved via β-elimination by Michael addition with dithiothreitol (BEMAD), which releases O-glycans and simultaneously labels the site, i.e. serine or threonine residue (Shajahan et al., 2017a). However, this method was not performed within the scope of this study.

The glycopeptide (R)SETSTFETPLGAK(T) was found to be occupied with six different glycoforms of O-glycans of the mucin-type. Figure 4.16 shows the identified Tn antigen, i.e. a single GalNAc residue, at m/z 1570.7465 to be present on a subset of glycopeptides. Since no linkage information was obtainable with the MS methods performed, the following typification into core classes are unconfirmed; however, the structural composition is based on acquired MS data. Figure 4.17 depicts glycan structures of sialylated core 1 O-glycans (m/z 2972.2196, m/z 2681.1232), which form the most common O-GalNAc core. Figure 4.18 shows a glycopeptide at m/z 2519.0715, which could carry either a sialylated and extended core 1 or core 3

glycan, as both share the same glycopeptide mass. Figure 4.19 shows evidence of a glycopeptide with two concurrent O-glycosylations (m/z 3629.4503) of the types sialylated core 1, as well as sialylated extended core 3 O-GalNAc glycans.

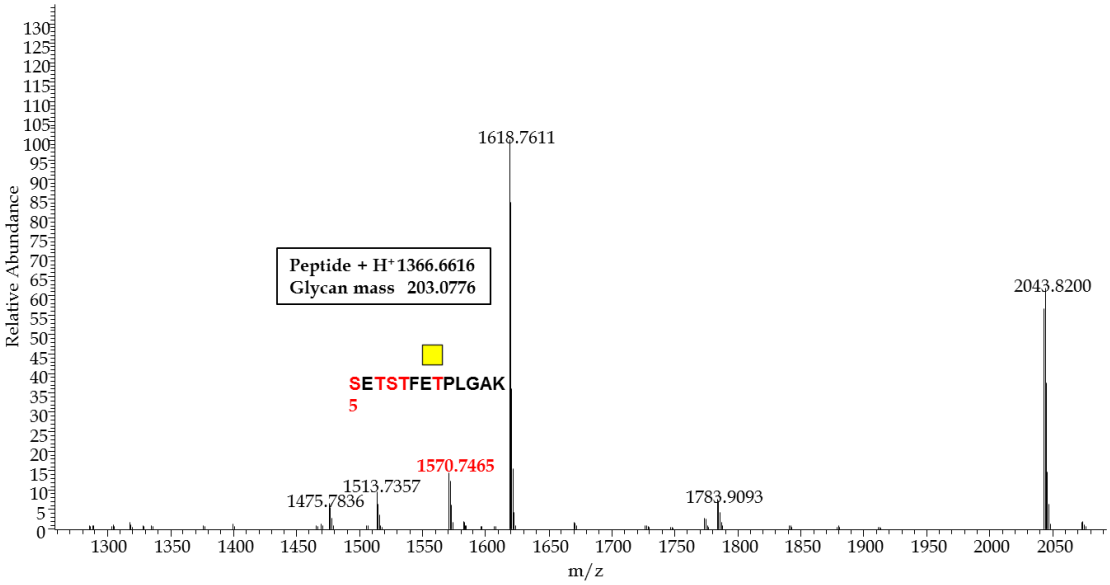


Figure 4.16: Deconvoluted and annotated full MS spectrum of the glycopeptide (R)SETSTFETPLGAK(T) (m/z 1250-2100) derived from human recombinant proMBP. The O-linked glycan (m/z 1570.7465) identified on the glycopeptide resembles the Tn antigen. Precise assignment of the glycan site was not possible, since several serines and threonines are present within this glycopeptide.

Symbols: GalNAc, GlcNAc, Man, Glc, Gal, Fuc, NeuAc

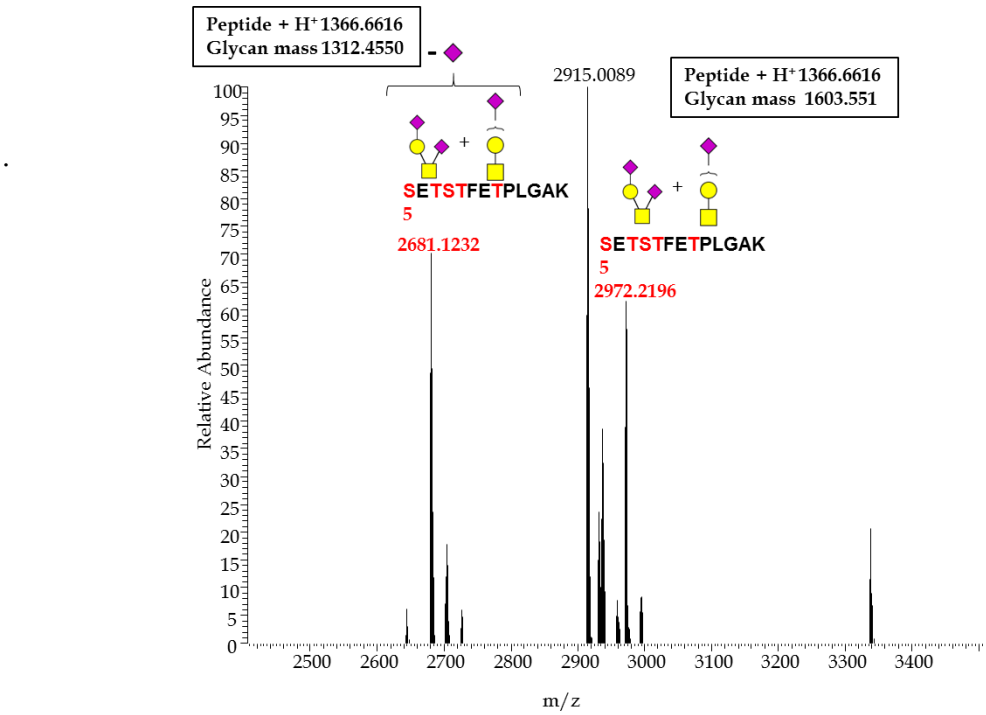


Figure 4.17: Deconvoluted and annotated full MS spectrum of the glycopeptide (R)SETSTFETPLGAK(T) (m/z 2420-3500) derived from human recombinant proMBP. The O-linked glycans at m/z 2681.1232 and m/z 2972.2196, are typified as of sialylated core 1 O-glycans.

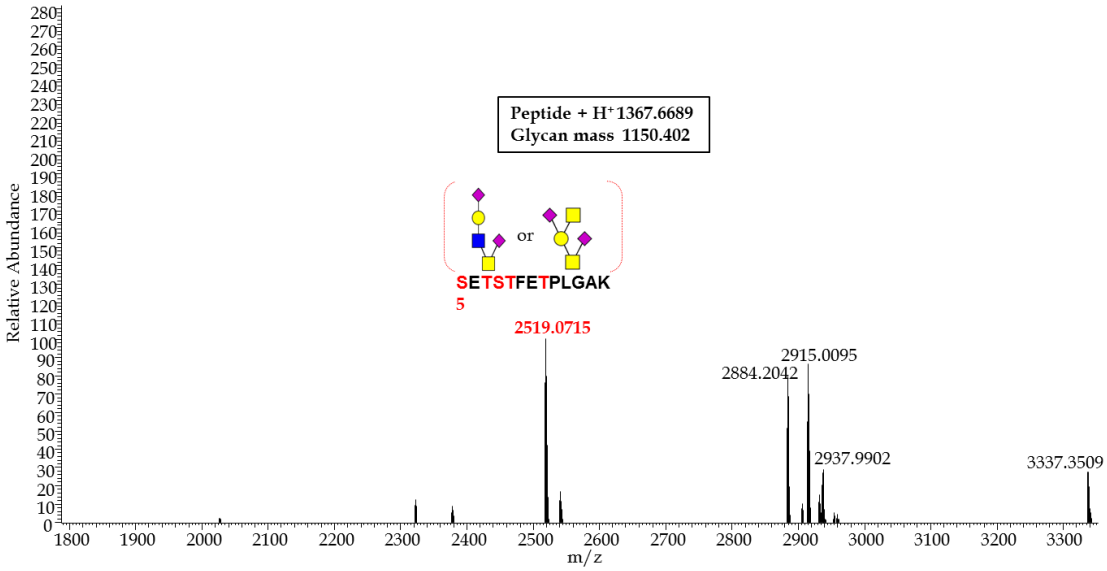


Figure 4.18: Deconvoluted and annotated full MS spectrum of the glycopeptide (R)SETSTFETPLGAK(T) (m/z 1800-3320) derived from human recombinant proMBP. The identified glycopeptide at m/z 2519.0715 matches with two putative, isobaric glycan structures of either a sialylated and extended core 1 type or a sialylated and extended core 3 type O-glycan.

Symbols: GalNAc, GlcNAc, Man, Glc, Gal, Fuc, NeuAc

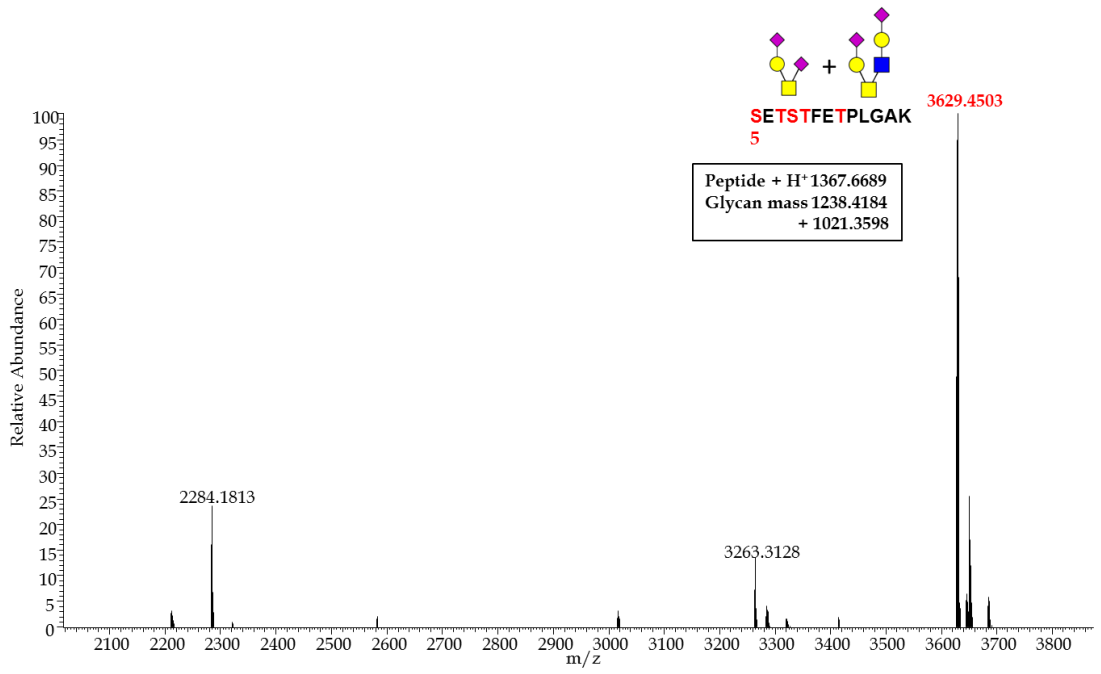


Figure 4.19: Deconvoluted and annotated full MS spectrum of the glycopeptide (R)SETSTFETPLGAK(T) (m/z 2020-3860) derived from human recombinant proMBP. The glycopeptide with two concurrent O-GalNAc glycoforms (m/z 3629.4503) carries possibly sialylated core 1, as well as sialylated and extended core 3 type glycans.

b) Amino acids threonine 18, threonine 25, threonine 33

From previous studies, the amino acid threonine-18 (positions with signal peptide: threo-nine-34) is known to be occupied with O-GalNAc glycosylation (The UniProt Consortium, 1990). The glycopeptide (E)TLPEDEETPEQEMEETPCR(K), containing the site of interest, was predicted through the programme PeptideCutter for the mature human proMBP and confirmed via b-ions and y-ions in tandem MS spectra. All three threonine residues in the sequence are highlighted because the identified glycoform could be located on any. The BEMAD method could also be used in future studies for a precise site mapping of the glycan. However, since threonine-25 and threonine-33 are directly followed by a proline, they are less likely to be modified with glycosylation.

The glycopeptide TLPEDEETPEQEMEETPCR showed one O-GalNAc glycoform at m/z 3634.4155 that is shown in figure 4.20. The glycoform was confirmed in several full MS as well as HCD tandem spectra (data not shown) within the MS data set and represents putatively an extended core 3 O-GalNAc with branching and sialyation.

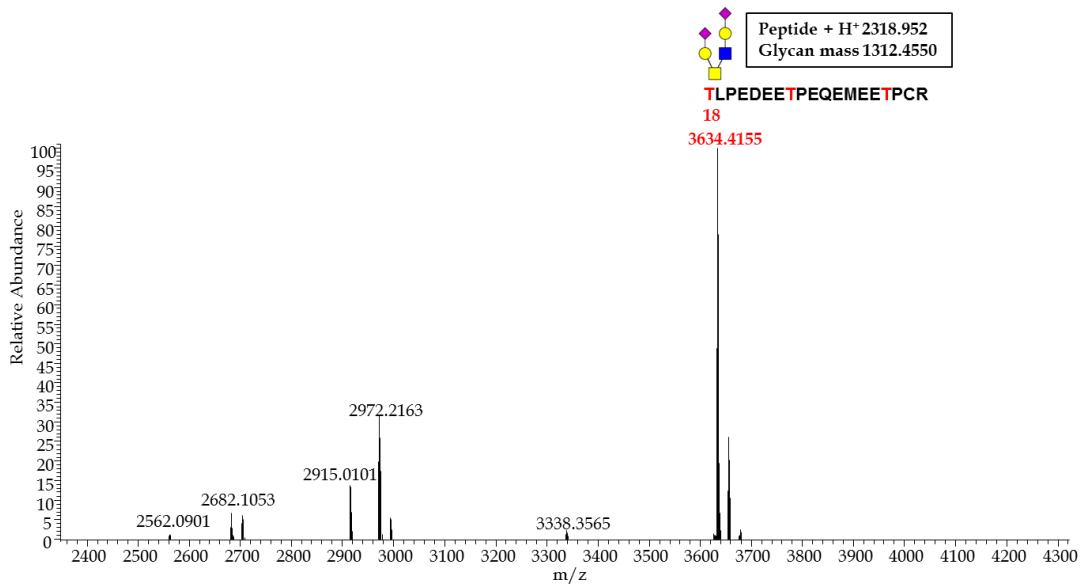


Figure 4.20: Deconvoluted and annotated full MS spectrum of the glycopeptide (E)TLPEDEETPEQEMEETPCR(K) (m/z 2360-4300) derived from human recombinant proMBP. The peak at m/z 3634.4155 matches the glycopeptide modified with an extended core 3 O-GalNAc glycan with branching and sialyation. The glycoform was confirmed through multiple spectra.

4.3.3 Glycoproteomic analysis of glycosamino-glycan-site on human recombinant proMBP

ProMBP, that is also referred to as Bone marrow proteoglycan (BMPG) or Proteoglycan 2 (PRG2), is itself a member of the CTL superfamily. Interestingly, it recognises glycans of the heparan sulfate glycosaminoglycan type as well as heparin, but not via its CTL ligand-binding domain (Swaminathan et al., 2005, Glerup et al., 2006). Moreover proMBP carries within its own protein sequence of the main chain, i.e. mature proMBP, a glycosaminoglycan (GAG) on serine-46 (amino acid 62 with signal peptide) of the heparan sulfate type. Glerup et al. (2006) demonstrated that proMBP binds to glycosaminoglycan-lined cell surfaces and, furthermore, that this interaction is independent of the GAG structure of proMBP.

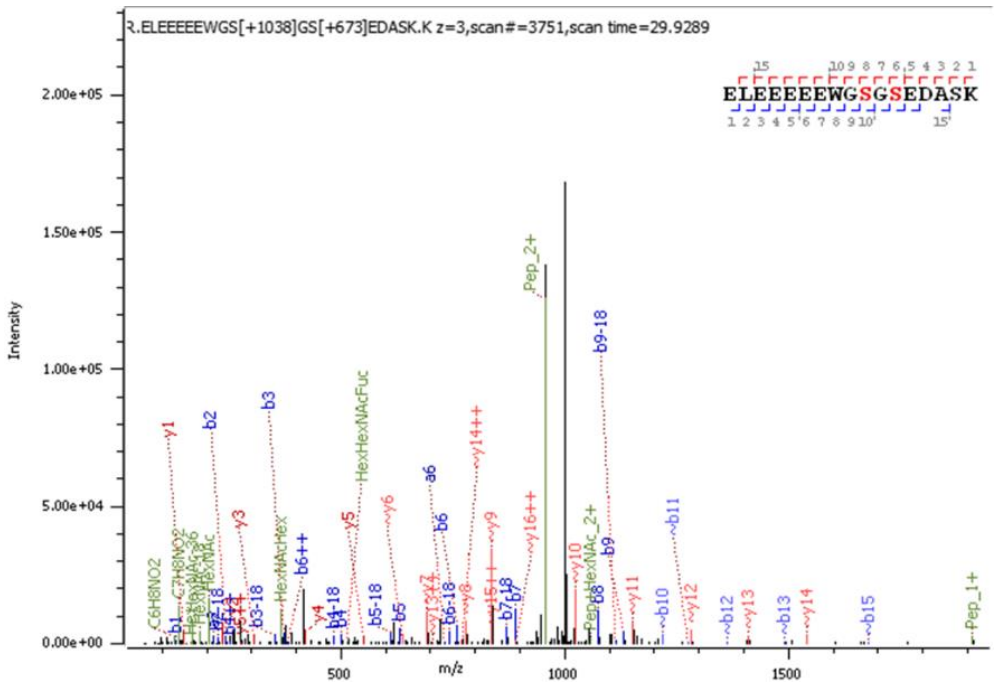


Figure 4.21: HCD MS² spectrum of the glycosaminoglycan detected on mature proMBP. HCD spectrum of the glycopeptide (R)ELEEEEEWGSGSEDASK(K) as presented in the software Byonic. The spectrum shows scan number 3751 that derived from sample eluted at retention time point 29.9289 min from the nanoLC. The highlighted and annotated S within the peptide sequence reflects a modification of the serine(s) with a putative glycan structure(s) of about 5133 Da (resulting from 1038Da and 673 Da, both z=3) in total. This aligns with the most prominent peak at m/z 1038, which was not annotated by Byonic. From past studies it is known that site S46 is occupied with a glycosaminoglycan of the heparan sulfate type (Glerup et al., 2006). Matched b- (in blue) and y-ions (in red) of the peptide as well as diagnostic ions derived from glycan structures (green) are labelled in the spectrum. Abbreviations: b – b-ion, Fuc – fucose/pentose, Hex – hexose, HexNAc - N-acetylhexosamine, NeuAc – sialic acid or N-acetylneuraminic acid, Pep – peptide, y – y-ion.

The predicted peptide, which derived from a tryptic digest *in silico* with PeptideCutter, was (R)ELEEEEEWGSGSEDASK(K). The software Byonic did detect a modification within this peptide as shown in figure 4.21.

A prominent peak at m/z 1038 with z=3 is dominating the HCD MS^2 spectrum (depicted in figure 4.21.) However, the software Byonic is not designed for identification of GAG-sites. Instead, diagnostic oxonium ions of N-acetylhexoseamine-hexose (HexNAcHex) residues were identified by the software and resulted in the prediction of glycosylation of the peptide. The peptide sequence itself was fully covered by b- and y-ions, which verified the presence of the peptide ELEEEEEWGSGSEDASK.

The deconvoluted full MS spectrum of this scan (presented in Xcalibur software) is shown in figure 4.22. The prominent peak at m/z 4944.8696 was identified as the glycopeptide with GAG-modification. The peptide mass was calculated with 1909.77 Da. This allows for a glycoform mass of approximately 3032 Da, considering monoisotopic mass differences. The glycan search of the monoisotopic mass of m/z 4942.87221 via the tool GlycoMod resulted in about 1400 possible glycan structures.

The complete study of glycosaminoglycan structures is even more complex than standard O-glycan analysis and requires several milligrams of the glycan. Due to the scope of this study and the availability of only 50 µg of glycosylated protein in total, the structural analysis of the GAG attached to proMBP was not attempted. However, preliminary evaluation of its size has been obtained.

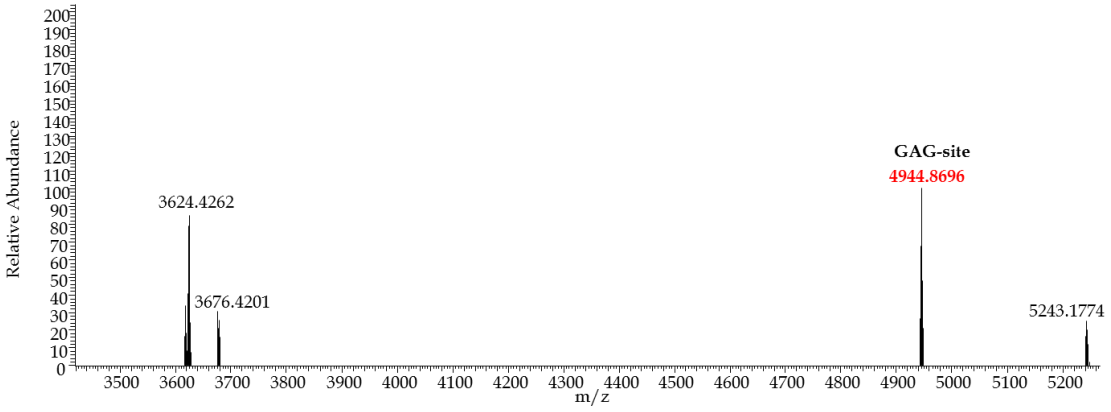


Figure 4.22: Deconvoluted and annotated full MS spectrum of the glycopeptide (R)ELEEEEEWGSGSEDASK(K) (m/z 3420-5260) derived from human recombinant proMBP. The peak at m/z 4944.8696 shows the glycopeptide modified with a glycosaminoglycan of unknown structure.

4.3.4 Conclusions from the glycoproteomic analysis of the recombinant human proMBP

As early as 1993, it became evident that proMBP is heavily glycosylated with complex carbohydrate structures of the O-linked glycan, N-linked glycan and a glycosaminoglycan (GAG) type (Shikata et al., 1993).

Within this part of the study, the mass spectrometric data obtained from glycopeptides derived from recombinantly expressed human proMBP was analysed via glycoproteomics. The findings fully confirm the glycosylation of the propeptide with N-linked glycoforms on asparagine-70, and O-linked glycans on two glycopeptides that include serine-5, threonine-7, serine-8, threonine-12, threonine-18, threonine-25 and threonine-33 (all amino acid positions without signal peptide). However, the exact site analysis for O-glycans would require an in-depth technique, such as the β-elimination by Michael addition with dithiothreitol (BEMAD), which releases the O-glycans and simultaneously labels the site of attachment for subsequent analysis. Furthermore, a modification of serine 46 was found, which presented diagnostic MS ions of glycans when analysed. It is without much doubt that this data represents the glycosaminoglycan, which is known to modify the serine-46 in past studies.

In summary, N-linked and O-linked glycans were detected and their structural characteristics analysed within this study, which is in keeping with earlier, preliminary information available in the database Uniprot as well as the study conducted by Oxvig et al. (1994).

The novel insight gained via this study is the composition of the N-linked and O-linked glycans attached to proMBP.

The sole N-glycan attachment site at asparagine 70 was found to be occupied with N-glycans of the complex type. The glycoforms were identified as bi- and tri-antennary structures with some of them presenting fucosyl residues at the core as well as terminal ends. About half of the glycoforms elucidated belong to the bisected category of the complex glycan type that carries a core GlcNAc residue, which is not extended. The glycans of the bisected complex type cover the core mannoses, which

are known to be recognised by DC-SIGN, with a GalNAc residue. Most importantly, no high mannose glycoform has been detected. The lack of high mannose type glycans and the underrepresentation of core mannose structures in N-glycans of proMBP indicate that an interaction between the CRD of DC-SIGN and the glycosylations of proMBP cannot be excluded; however, the overall interaction would be anticipated to be of a nature of lower affinity, since the CRD of DC-SIGN favours mannose-type sugars. On the other hand, the fucosylation of some N-linked glycoforms could provide the basis of a weaker interaction between proMBP and DC-SIGN.

In contrast, DC-SIGNR only recognises mannose residues. Therefore, no interaction between the CRD of DC-SIGNR and the N-linked glycans of proMBP would be expected.

Furthermore, all of the identified N-linked glycoforms contained galactose residues that were predominantly capped by sialic acid residues.

Excluding the GAG, the remaining O-glycans of proMBP belong to the mucin-type. They are found to be shorter regarding the degree of glycan extension and show a lower degree of branching than the identified N-glycans. This aligns with finding from studies about O-linked glycans in human blood plasma proteins. The Tn antigen and, partially extended and branched, core 1 and core 3 O-glycans were identified for O-linked glycans on proMBP.

Since the O-glycan analysis in this study did not show any O-glycans with Lewis-type glyco-epitopes present for recombinant proMBP, it seems unlikely that DC-Sign interacts with the O-glycans present on proMBP – presuming they are sterically available for interaction.

4.4 Recombinant human PAPP-A1

The sample of recombinant human PAPP-A1 was expressed in mouse myeloma cells (derived from the NS0 cell line), as stated by the manufacturer. The expressed protein sequence does not cover the full sequence of native PAPP-A1 (signal peptide aa 1-22; propeptide aa 23-80; mature chain aa 81-1627), but the amino acids 82/84-1214. The recombinant protein sequence corresponds to approximately three quarters of the native, mature PAPP-A1 sequence, and was confirmed prior to the glycoproteomic analysis through MS-derived proteomic data. In the interest of clarity and better

```
MRLWSWVLHL GLLSAALGCG LAERPRRARR DPRAGRPPRP AAGPATCATR
AARGRRASPP PPPPPGGAWE AVRVPRRRQQ REARGATEEP SPPSRALYFS
GRGEQLRLRA DLELPRDAFT LQVWLRAEGG QRSPAVITGL YDKCSYISRD
RGWVVGIHTI SDQDNKDPRY FFSLKTDRAR QVTTINAHRS YLPGQWVYLA
ATYDGQFMKL YVNGAQVATS GEQVGGIFSP LTQKCKVLML GGSALNHNYR
GYIEHFSLWK VARTQREILS DMETHGAHTA LPQLLLQENW DNVKHAWSPM
KDGSSPKVEF SNAHGFLLDT SLEPPLCGQT LCDNTEVIAS YNQLSSFRQP
KVVRYRVVNL YEDDHKNPTV TREQVDFQHH QLAEAFKQYN ISWELDVLEV
SNSSLRRRLI LANCDISKIG DENCDPECNH TLTGHDGGDC RHLRHPAFVK
KQHNGVCDMD CNYERFNFDG GECCDPEITN VTQTCFDPDS PHRAYLDVNE
LKNILKLDGS THLNIFFAKS SEEELAGVAT WPWDKEALMH LGGIVLNPSF
YGMPGHTHTM IHEIGHSLGL YHVFRGISEI QSCSDPCMET EPSFETGDLC
NDTNPAPKHK SCGDPGPGND TCGFHSFFNT PYNNFMSYAD DDCTDSFTPN
QVARMHCYLD LVYQGWQPSR KPAPVALAPQ VLGHTTDSVT LEWFPPIDGH
FFERELGSAC HLCLEGRILV QYASNASSPM PCSPSGHWSP REAEGHPDVE
QPCKSSVRTW SPNSAVNPHT VPPACPEPQG CYLELEFLYP LVPESLTIWV
TFVSTDWDSS GAVNDIKLLA VSGKNISLGP QNVFCDVPLT IRLWDVGEEV
YGIQIYTLDE HLEIDAAMLT STADTPLCLQ CKPLKYKVVR DPPLQMDVAS
ILHLNRKFVD MDLNLGSVYQ YWVITISGTE ESEPSPAVTY IHGSGYCGDG
IIQKDQGEQC DDMNKINGDG CSLFCRQEVS FNCIDEPSRC YFHDGDGVCE
EFEQKTSIKD CGVYTPQGFL DQWASNASVS HQDQQCPGWV IIGQPAASQV
CRTKVIDLSE GISQHAWYPC TISYPYSQLA QTTFWLRAYF SQPMVAAAVI
VHLVTDGTYY GDQKQETISV QLLDTKDQSH DLGLHVLSCR NNPLIIPVVH
DLSQPFYHSQ AVRVSFSSPL VAISGVALRS FDNFDPVTLS SCQRGETYSP
AEQSCVHFAC EKTD
```

Figure 4.23: FASTA peptide sequence of human recombinant PAPP-A1. The full sequence would consist of the signal peptide sequence (aa 1-22; highlighted in light grey), the propeptide sequence (aa 23-80; highlighted in dark grey) and the mature chain (aa 82/84-1214). Highlighted in magenta are aa82 and aa84. The expressed recombinant PAPP-A1 does neither hold the signal peptide nor the propeptide sequence. As per the manufacturer, the protein consists of 1133/1131 aa with 9 potential N-glycosylation sites (sequon N-X-S/T, X not P [valid sequons highlighted in yellow; sequons N-P-S/T highlighted in cyan]).

comparability between the data sets, the amino acids will be referred to within the numeric frame of the full mature chain, i.e. aa 81 set as the first amino acid of the protein sequence (aa 1). The database-retrieved and adapted sequence is shown in figure 4.23. (The UniProt Consortium, 2001), including highlighted sequences that represent the signal peptide (light grey), propeptide (dark grey), valid N-glycosylation sequons (N-X-S/T, X not P; yellow) and invalid sequons N-P-S/T (cyan).

The available amount of recombinant PAPP-A1 was limited at 20 µg, which restricted the structural analysis of the glycan structures to a glycoproteomic approach only.

This study was carried out at the Analytical Services and Training Laboratory at the Complex Carbohydrate Research Center of the University of Georgia, USA.

4.4.1 Glycoproteomic analysis of N-linked glycans on human recombinant PAPP-A1

The recombinant human PAPP-A1 was investigated for its N-linked glycan modifications.

The PAPP-A1 sample had sufficient purity (>90% purity as certified) for a direct and consecutive enzymatic digest with trypsin and chymotrypsin. The same enzymes were also chosen for the enzymatic digest of the native human PAPP-A1 (see section '4.5.2 Glycoproteomic analysis of N-linked glycans on human native PAPP-A1/proMBP-complex') to facilitate the mass spectrometric analysis. The proteolytic cleavages were expected according to the cleavage specificities of the enzymes: Lysine (K) and arginine (R) (from the C-terminal side, if not followed by a proline) for trypsin, and bulky hydrophobic amino acids (from C-terminus, if not followed by a proline), such as tyrosine, tryptophan, and phenylalanine, for chymotrypsin (Perona and Craik, 1995, Baird and Craik, 2013, Gráf et al., 2013). Per *in silico* digest, the proteases trypsin and chymotrypsin yielded the best theoretical peptide coverage for mass spectrometry with peptide lengths between 5-30 aa. Importantly, most peptides carrying a sequon were theoretically within this range, except for NISW (aa 310-313). For convenience of identification, the sequons of N-glycosylation sites will be presented in a peptide sequence of four amino acids.

In a study carried out in 2003, Overgaard et al. showed the occupation of 11 sequons out of 14 potential N-glycosylation sites for the human mature PAPP-A1, i.e. N322 (NSSL), N349 (NHTL), N400 (NVTQ), N521 (NDTN), N539 (NDTC), N645 (NASS), N745 (NISL), N946 (NASV), N1146 (NCSS), N1243 (NNSL), and N1439 (NVTV). The asparagine residues N310 (NISW), N1142 (NASL), and N1405 (NGSF) were found without glycosylation within their study (Overgaard et al., 2003). All data was acquired on basis of liquid chromatography and N-terminal sequence analysis of (glyco)peptides, but no mass spectrometric method was applied. Furthermore, no analysis of the types of carbohydrates attached to PAPP-A1 was undertaken at the time.

4.4.1.1 nLC-NSI-MS/MS analysis of N-linked glycoforms on glycopeptides from human recombinant PAPP-A1

For the glycoproteomic analysis of glycopeptides derived from recombinant PAPP-A1, the technique of nano-scale liquid chromatography with on-line tandem MS (nLC-MS/MS or nLC-MS^2; see section '2.6 Methods for Mass Spectrometry' for details) was the method of choice to determine site-specific information of glycosylation, such as the site of the glycan attachment and prevalent glycoforms. The nLC-MS/Ms is able to separate isobaric glycans, which have the same mass but different sugar compositions and/or conformations (Shajahan et al., 2017a). After ionisation via the NSI method, full MS scans were acquired, and precursor ions selected applying various criteria. The selected precursors were then processed by tandem MS for parallel CID and HCD fragmentation. Diagnostic ions derived from glycan structures triggered further analysis via a subsequent fragmentation method, namely ETD. For data analysis of MS data, the software programmes Xcalibur and Byonic were used. Byonic quantified the peptide coverage with 75.1% confidence for recombinant PAPP-A1, applying search parameters for N-glycan modification, carbamidomethylation of cysteines, and oxidation of methionines.

Since the peptide sequence of the recombinant human PAPP-A1 covers about 1133 amino acids (full sequence reference aa 82/84-1214), only N322 (NSSL), N349 (NHTL), N400 (NVTQ), N521 (NDTN), N539 (NDTC), N645 (NASS), N745 (NISL), and N946 (NASV) are the putative sites present. Out of these, N310 (NISW) would not be expected to be modified via glycosylation, according to Overgaard et al. (2003).

The two sequons with a proline (N-P-S/T) are not expected carry glycan structures, according to the existing criterion for N-linked glycosylation sequons.

The following subchapters present the glycoforms identified on the confirmed glycosylation sites. The spectra do not contain information about linkages between the sugar moieties, since linkage information was not part of the analysis and requires additional methods. The annotations in the HCD tandem (MS^2 or MS/MS) mass spectra illustrate the assigned glycan structure based on the mass, the pathway of synthesis of N- linked glycans, and the species-specific glycosylation pattern for mammals, i.e. mice.

a) Amino acid asparagine-349

The peptide sequence around the sequon NHTL at asparagine 349 (N349) was predicted via PeptideCutter for the recombinant PAPP-A1, which resulted in the (glyco)peptide (K)IGDENCDPECNHTLTGHDGGDCR(H) for a combined *in silico* digest with trypsin and chymotrypsin. This glycopeptide was confirmed and showed seven different glycoforms that are presented in an overview in figure 4.26. The glycan structure depicted in figure 4.24. at m/z 3848.4396 was identified as high mannose type, whereas the bi-antennary glycoforms at m/z 4050.5135 and m/z 4212.5654 belong to the hybrid type - since they contain one antenna with unsubstituted terminal mannose residues as well as another antenna with GlcNAc-substituted mannose residues. The annotated glycoforms at m/z 4577.6993, m/z 4942.8297, and m/z 5105.8859 represent bi-antennary complex glycans in figure 4.25. The additional glycoform m/z 4561.6986 in figure 4.26. is classified as a complex N-linked glycan with core fucosylation.

The data of the ion fragments was retrieved via on-line nano-LC-MS/MS. All figures show full MS spectra.

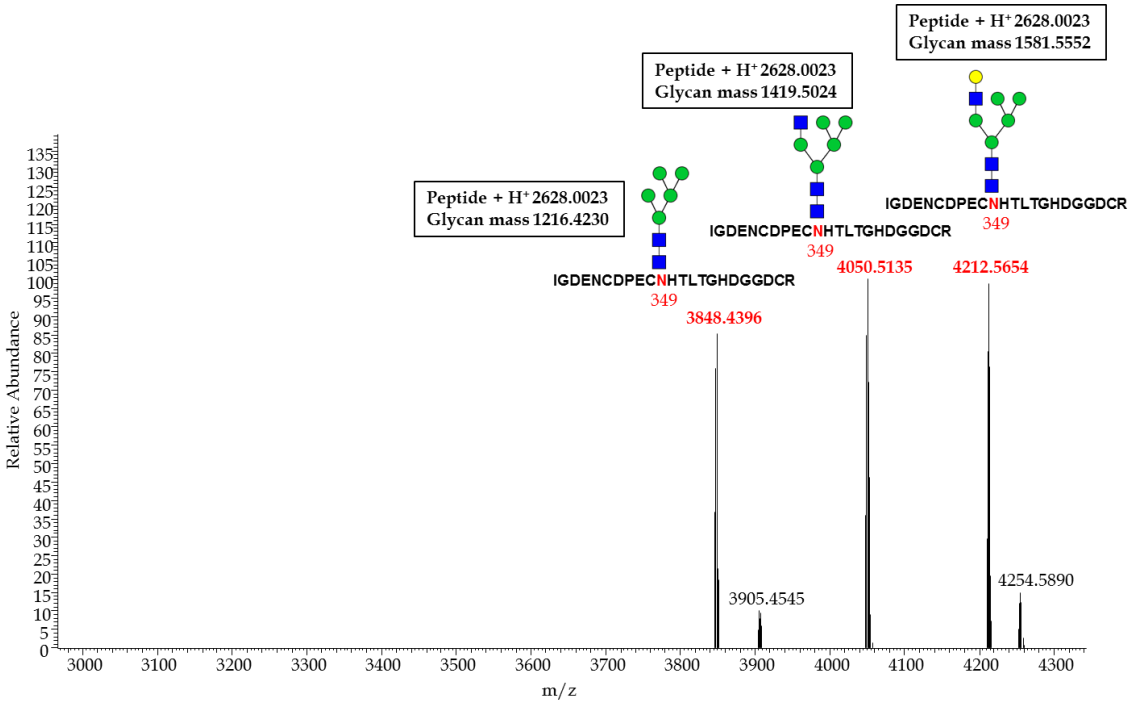


Figure 4.24: Deconvoluted and annotated full MS spectrum of glycopeptide (K)IGDENCDPECNHTLTGHDGGDCR(H) (m/z 2980-4340) derived from human recombinant PAPP-A1 via nanoLC-NSI-MS/MS. Glycoforms identified in this spectrum are high-mannose (m/z 3848.4396) or hybrid type (m/z 4050.5135, m/z 4212.5654).

Symbols: GalNAc, GlcNAc, Man, Glc, Gal, Fuc, NeuAc

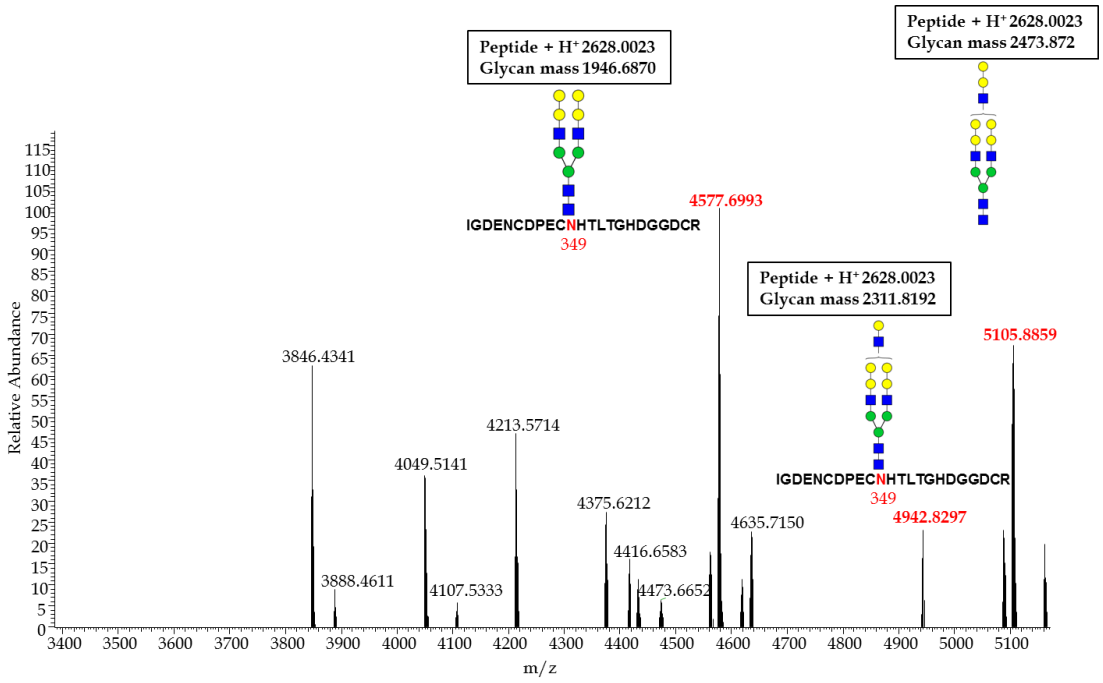


Figure 4.25: Deconvoluted and annotated full MS spectrum of glycopeptide (K)IGDENCDPECNHTLTGHDGGDCR(H) (m/z 3400-5160) derived from recombinant PAPP-A1 via nanoLC-NSI-MS/MS. Glycoforms identified in this spectrum (m/z 4577.6993, m/z 4942.8297, m/z 5105.8859) are classified as complex type.

Symbols: GalNAc, GlcNAc, Man, Glc, Gal, Fuc, NeuAc

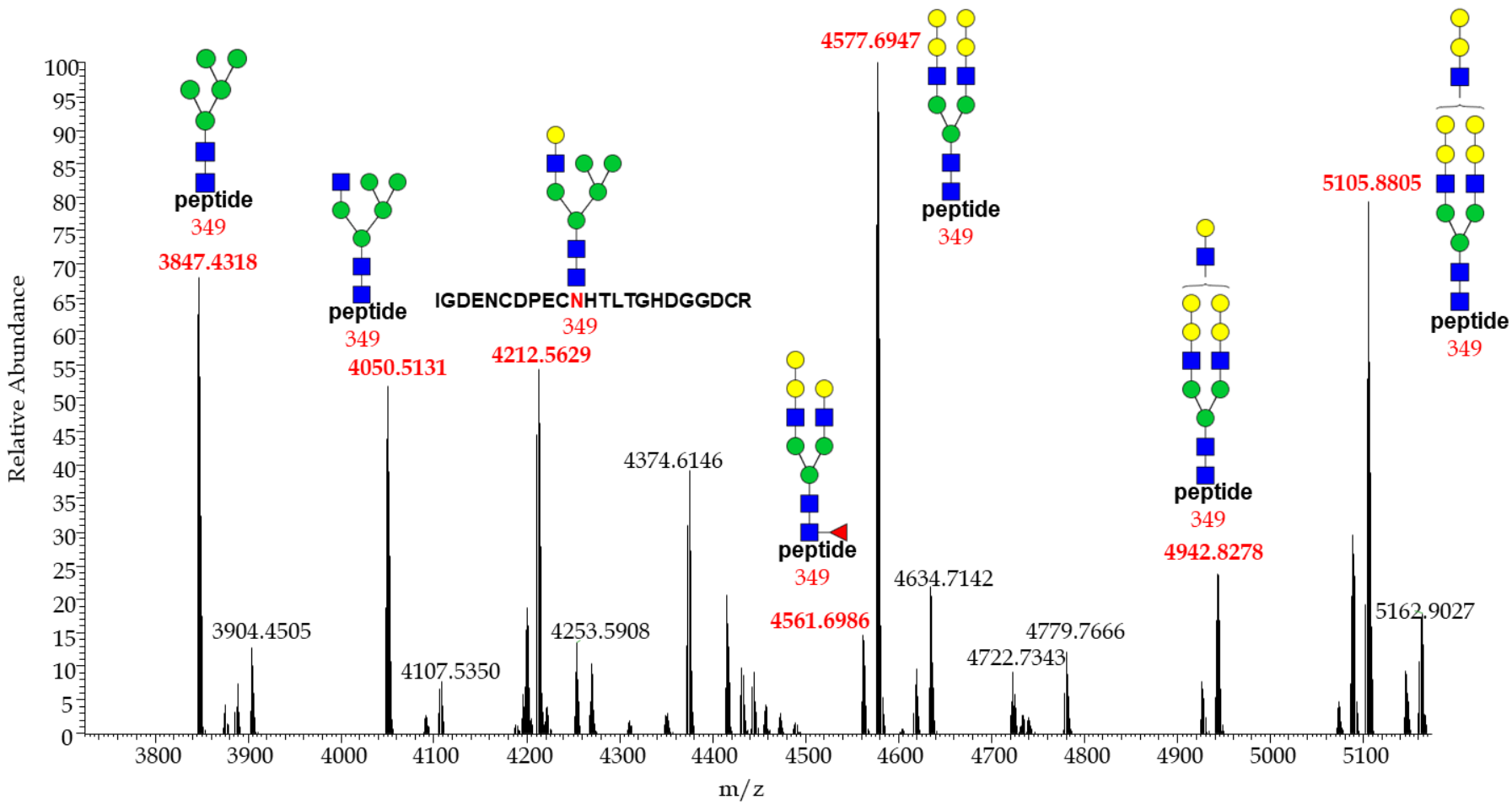


Figure 4.26: Deconvoluted and annotated full MS spectrum of the glycopeptide (K)IGDENCDPECNHTLTGHDGGDCR(H) (m/z 3720-5160) that provides an overview of the glycoforms identified at N349 for the recombinant PAPP-A1. All classes of glycoforms have been found, i.e. high-mannose type (m/z 3848.4396), hybrid type (m/z 4050.5135, m/z 4212.5654) or complex N-glycan type (m/z 4561.6986, m/z 4577.6993, m/z 4942.8297, m/z 5105.8859), respectively.

Symbols: GalNAc, GlcNAc, Man, Glc, Gal, Fuc, NeuAc

b) Amino acid asparagine-400

The predicted (glyco)peptide, around the sequon NVTQ at asparagine-400 for the recombinant human PAPP-A1, was (R)FNFDGGECCDPEIT-NVTQTCFDPDSPHR(A) according to the *in silico* enzymatic digest with PeptideCutter. The identified sequence confirmed the prediction. The determined glycoform at asparagine-400 is shown in figure 4.27. at m/z 5938.2722. The glycan structure is classified as tri-antennary complex type with a core fucosylation.

The data was obtained via on-line nano-LC-MS/MS and shows a full MS spectrum.

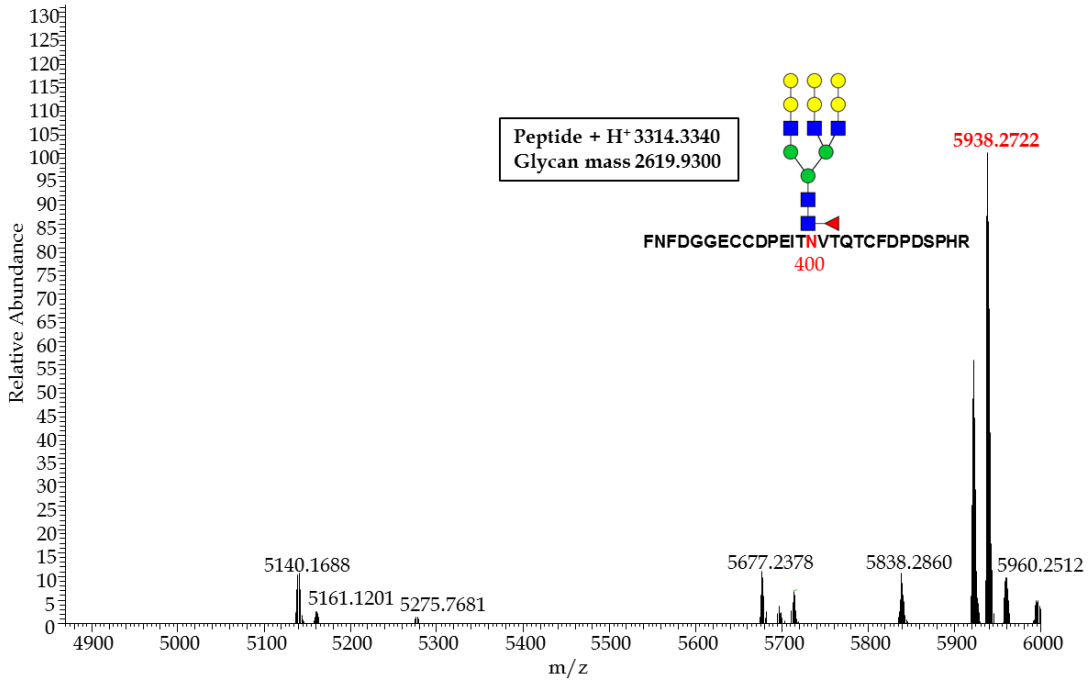


Figure 4.27: Deconvoluted and annotated full MS spectrum of the glycopeptide (R)FNFDGGECCDPEITNVTQTCFDPDSPHR(A) (m/z 4880- 6000) derived from human recombinant PAPP-A1. The glycoform found to occupy asparagine-400 is a glycan structure classified as fucosylated complex glycan (m/z 5938.2722).

Symbols: GalNAc, GlcNAc, Man, Glc, Gal, Fuc, NeuAc

c) Amino acid asparagine-521

The (glyco)peptides carrying the sequon NDTN (asparagine-521) for the recombinant human PAPP-A1 were predicted as (R)GISEIQSCSDPCMETEPSF(E) and (F)ETGDLCNDTNPAPK(H) as per *in silico* digestion with both, trypsin and chymotrypsin. However, the glycopeptide identified was (R)GISEIQSCS-DPCMETEPSFETGDLCNDTNPAPK(H). Very likely, the proximal proline caused a change in the secondary structure, that might have prevented chymotrypsin from cleavage before the amino acid phenylalanine-514 and caused the elongated peptide.

The glycopeptide was found to carry a tri-antennary carbohydrate structure of the complex glycan type, which is presented in figure 4.28. at m/z 6001.3410.

The data was retrieved via on-line nano-LC-MS/MS and shows a full MS spectrum.

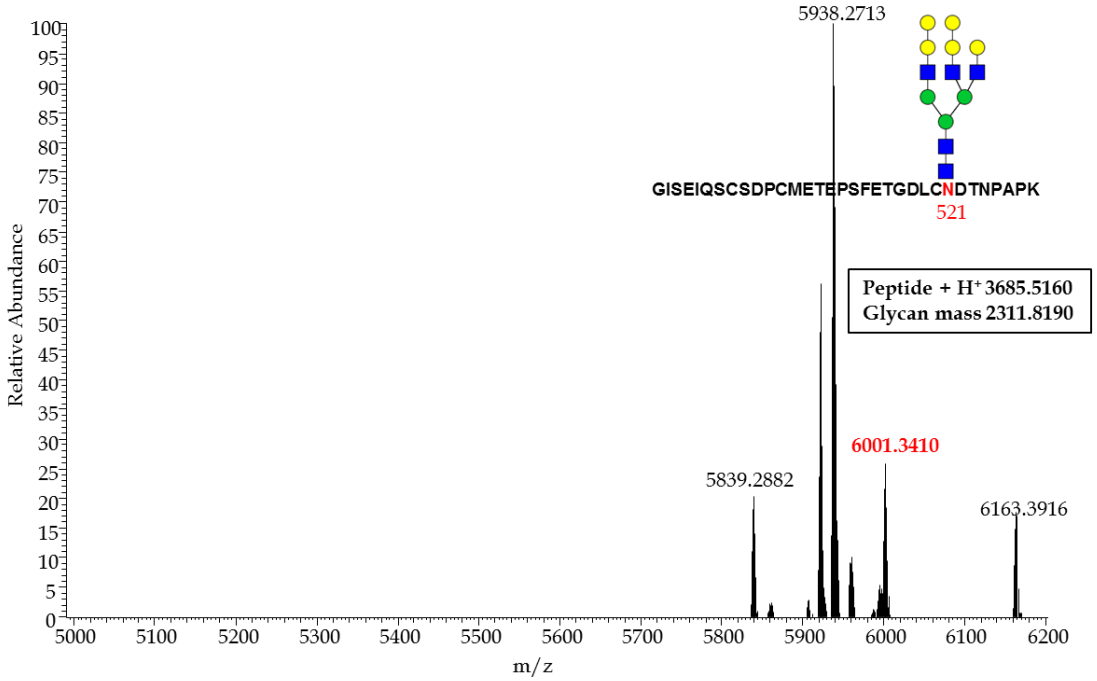


Figure 4.28: Deconvoluted and annotated full MS spectrum of the glycopeptide (R)GISEIQSCSDPCMETEPSFETGDLCNDTNPAPK(H) (m/z 5000-6200) obtained from human recombinant PAPP-A1 via nanoLC-NSI-MS/MS. The glycoform identified at m/z 6001.3410 is typified as tri-antennary complex glycan structure.

Symbols: *GalNAc,* *GlcNAc,* *Man,* *Glc,* *Gal,* *Fuc,* *NeuAc*

d) Amino acid asparagine-539

After an *in silico* digest with trypsin and chymotrypsin, the programme PeptideCutter predicted the (glyco)peptide (K)SCGDPGPGNDTCGF(H) around the sequon NDTC at asparagine-539 for the recombinant human PAPP-A1 protein sequence. The glycopeptide was confirmed and the glycosylation site at asparagine-539 found to be occupied with two different tri-antennary glycoforms that are both classified as complex type with core fucosylation (m/z 3900.3958, m/z 4062.4443), as annotated in figure 4.29.

The data was retrieved via on-line nano-LC-MS/MS and shows a full MS spectrum.

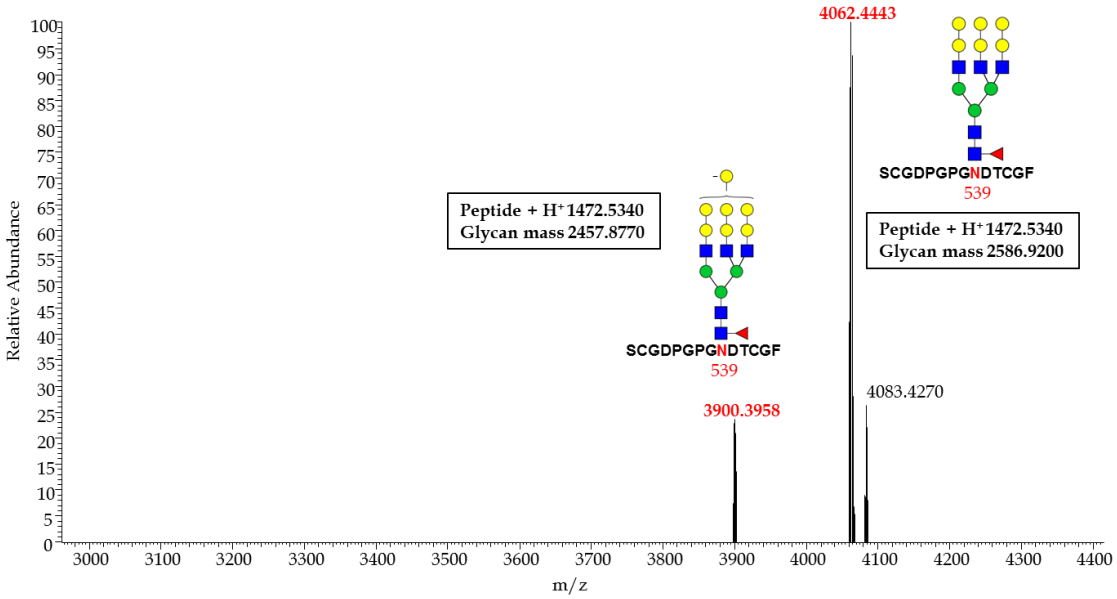


Figure 4.29: Deconvoluted and annotated full MS spectrum of the glycopeptide (K)SCGDPGPGNDTCGF(H) (m/z 2980-4400) derived from human recombinant PAPP-A1. The identified glycoforms at m/z 3900.3958 and m/z 4062.4443 represent complex glycan structures with core fucosylation.

Symbols: GalNAc, GlcNAc, Man, Glc, Gal, Fuc, NeuAc

e) Amino acid asparagine-745

The programme PeptideCutter was used to predict the potential cleavage sites around the sequon NISL at asparagine-745 (N745) for human recombinant PAPP-A1. For an *in silico* digest with trypsin and chymotrypsin, the resulting (glyco)peptide (K)NISLGPQNVF(C) was confirmed by MS data.

The glycosylation site at N745 was found to be occupied with two different glycoforms that are classified as complex type with core fucosylation at m/z 3531.4396 (bi- or tri-antennary glycan structure shown in figure 4.30.), and m/z 3710.5058 (tri-antennary glycan structure depicted in figure 4.31.). Interestingly, the glycoform at m/z 3710.5058 in figure 4.30 is the only glycan structure with a N-glycolylneuraminic acid (NeuGc) residue. NeuGc is an abundant glycan moiety in the murine glycosylation system and often appears in murine glycoproteins as terminal residue. However, NeuGc is almost non-detectable within adult humans.

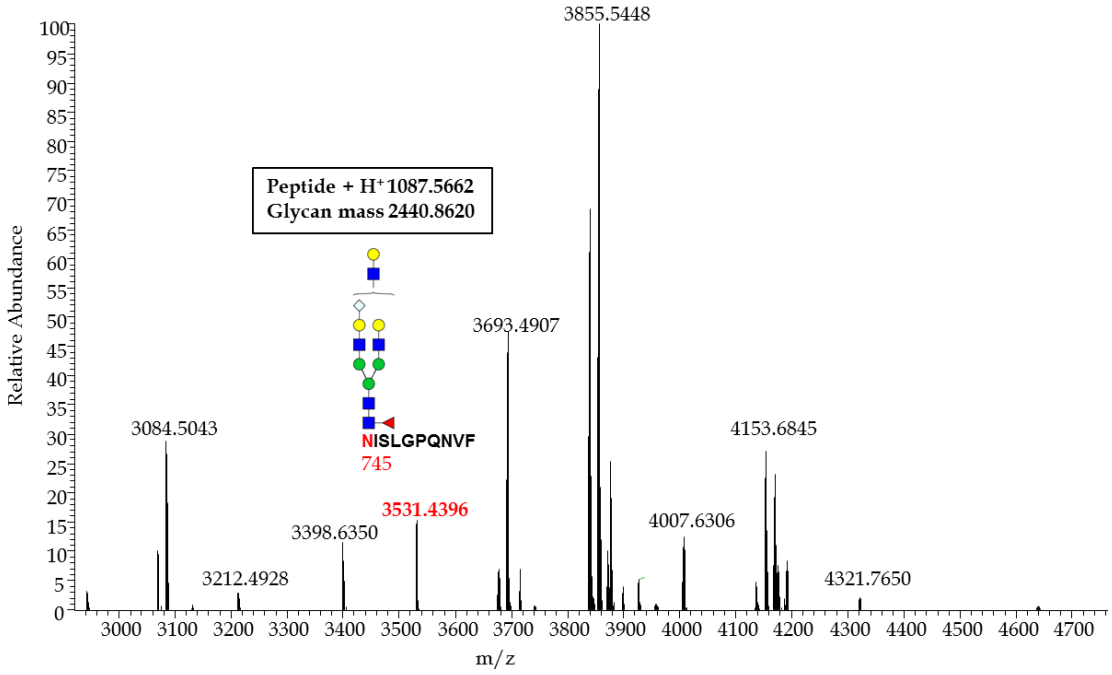


Figure 4.30: Deconvoluted and annotated full MS spectrum (m/z 2920-4760) of the glycopeptide (K)NISLGPQNVF(C) derived from human PAPP-A1 that was recombinantly expressed in a murine cell line. The glycoform identified at m/z 3531.4396 is classified as a bi-antennary, complex glycan with core fucosylation (m/z 4247.8412) but, uniquely within this study, carries a terminal N-glycolylneuraminic acid (NeuGc) residue.

Symbols: GalNAc, GlcNAc, Man, Glc, Gal, Fuc, NeuAc, NeuGc

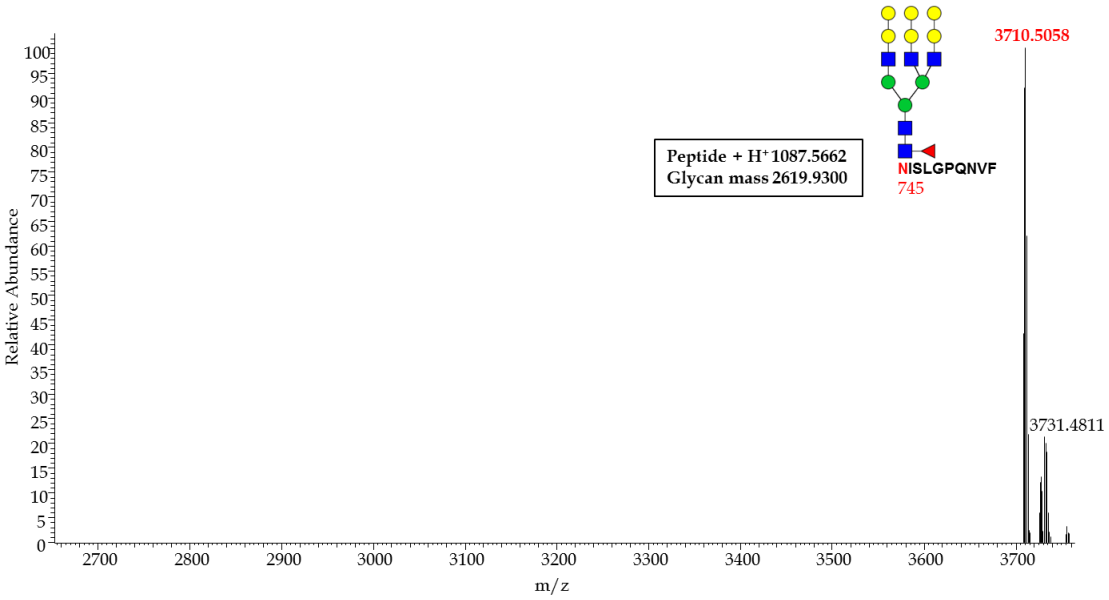


Figure 4.31: Deconvoluted and annotated full MS spectrum (m/z 2660-3760) of the glycopeptide (K)NISLGPQNVF(C) derived from recombinant PAPP-A1. The identified glycoform at m/z 3710.5058 in this spectrum is classified as tri-antennary complex glycan structure with a fucosyl residue attached to the N-glycan core.

Symbols: GalNAc, GlcNAc, Man, Glc, Gal, Fuc, NeuAc

f) Amino acid asparagine-946

The (glyco)peptide (W)ASNASVSHQDQQCPGW(V) was predicted after a combined *in silico* digest with trypsin and chymotrypsin with the programme PeptideCutter for the area around the sequon NASV at asparagine 946 (N946) for human recombinant PAPP-A1. The detected glycopeptide had the amino acid sequence (Y)TPQGFLDQWASNASVSHQDQ(Q), which represents two missed cleavages by chymotrypsin before phenylalanine-939 as well as tryptophan-943. However, an unpredicted cleavage occurred between glutamine-954 and glutamine-955.

The glycosylation site at N946 was found to be occupied with a tetra-antennary glycoform of the complex type at m/z 5467.1779 with fucosylation of the glycan core, as depicted in figure 4.32.

The data was obtained via on-line nano-LC-MS/MS. The figure shows a full MS spectrum.

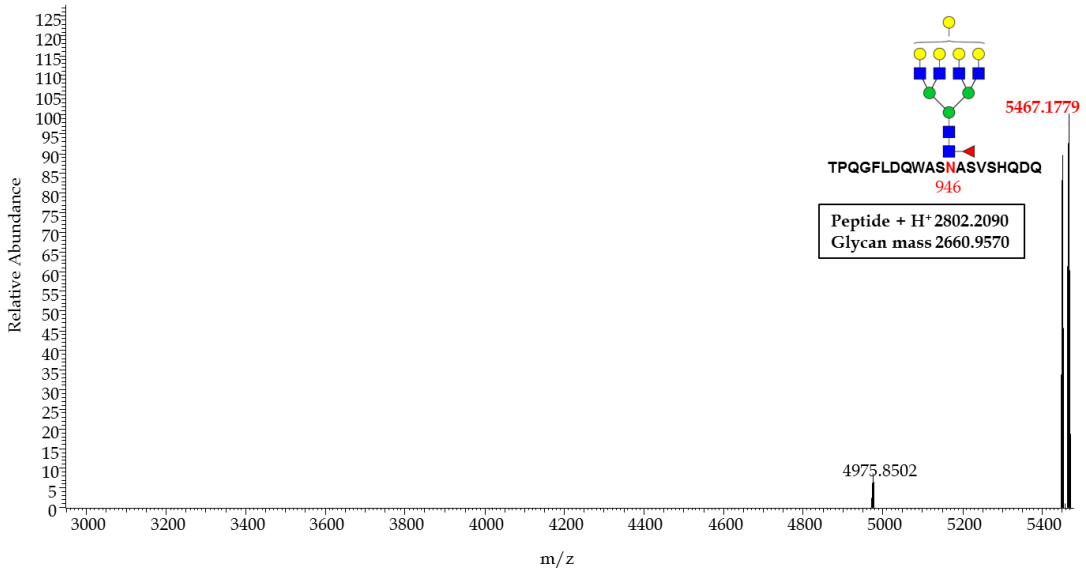


Figure 4.32: Deconvoluted and annotated full MS spectrum (m/z 2950-5475) of the glycopeptide (Y)TPQGFLDQWASNASVSHQDQ(Q) derived from human recombinant PAPP-A1 via nanoLC-NSI-MS/MS. The identified glycoform in this spectrum is classified as a tetra-antennary complex glycan with fucosylation of the glycan core (m/z 5467.1779).

Symbols: GalNAc, GlcNAc, Man, Glc, Gal, Fuc, NeuAc

4.4.1.2 Conclusions from the glycoproteomic analysis of the human recombinant PAPP-A1

The glycoproteomic site analysis of the recombinant PAPP-A1 showed that the asparagine residues N349 (NHTL), N400 (NVTQ), N521 (NDTN), N539 (NDTC), N745 (NISL), and N946 (NASV) were glycosylated. The peptides carrying the sequons N310 (NISW), N322 (NSSL) and N645 (NASS) were not covered and no assumption can be made. As expected, the two sequons with a proline (N-P-S/T) did not carry glycan structures, which is in line with the existing criterion for N-linked glycosylation sequons.

The analysis and characterisation of the glycan structures revealed that the recombinant PAPP-A1 is modified with glycoforms of the high-mannose, hybrid and complex type, respectively. Only asparagine-349 was partially occupied with glycoforms of the high-mannose and hybrid type. All other sequons revealed a modification with complex type glycans. No bisected complex carbohydrate structures were identified. Overall, the steric availability of mannose core structures (no bisection with a single GlcNAc) as well as a partial occupation with high-mannose glycoforms prefigure an interaction between the N-linked glycans attached to the investigated sample of human recombinant PAPP-A1 and the CRD of DC-SIGN. The carbohydrate recognition domain of DC-SIGNR might recognise the recombinant PAPP-A1 either with low affinity or not at all, given the low accessibility of mannose residues for binding.

Most sequons showed only a single glycoform attached. A summary of the findings is gathered in table 18.

One glycoform on asparagine-745 showed a sugar moiety that is present in abundance in mice but not adult humans. It is noteworthy that mice have a similar, but not identical glycosylation pattern to that of humans. In particular, N-glycolyl-neuraminic acid (NeuGc) is a common glycan moiety in murine glycoproteins. However, NeuGc is present only in traces within fully developed humans. It triggers an immune response when presented in abundance to the human immune system in postnatal phases (Varki, 2001). Remarkably, no terminal sialic acids residues (NeuAc) were identified in any of the glycoforms present on this recombinant human PAPP-A1 sample expressed in murine cells.

In keeping with past studies, no O-linked glycosylation was detected.

Table 18: Summary of the glycoproteomic analysis of N-linked glycans identified in recombinantly expressed PAPP-A1 (amino acid sequence(s) 82/84-1214). Out of the 9 N-glycan sequons (N-X-S/T, X not P; asparagine putatively linked to glycan is highlighted in red [N]), 6 have been found glycosylated including (partially fucosylated) glycoforms of the high-mannose, hybrid, and the complex type, respectively. However, 3 putative glycopeptides (sequons around N310, N322, N645) have not been covered by data.

‡ predicted peptide sequence due to PeptideCutter via tryptic and chymotryptic digest in silico

Amino Acid/ Glycosylation site	Glycosylation sequon	FASTA peptide sequence	Type of N-glycosylation
310	NISW	NISW‡	peptide not covered
322	NSSL	ELDVLEVSNSSLR‡	peptide not covered
349	NHTL	IGDENCDPECNHTLTGHDGGDCR	High-mannose, Hybrid, Complex [fucosylated]
400	NVTQ	FNFDGGECCDPEITNVTQTCFDPDSPHR	Complex [fucosylated]
521	NDTN	GISEIQSCSDPCMETEPSFETGDLCNDTNPAPK	Complex
539	NDTC	SCGDPGPGNDTCGF	Complex [fucosylated]
645	NASS	ASNASSPMPCSPSGHW‡	peptide not covered
745	NISL	NISLGPQNVF	Complex [fucosylated]
946	NASV	TPQGFLDQWASNASVSHQDQQCPGW	Complex [fucosylated, N-Glycolylneuraminic acid}
1142	NASL	sequence not present in recombinant PAPP-A1 sample	
1146	NCSS		
243	NNSL		
1405	NGSF		
1439	NVTV		

4.5 Human native PAPP-A1/proMBP-complex

This part of the research project was carried out at the Analytical Services and Training Laboratory at the Complex Carbohydrate Research Center and partly supported by the National Institutes of Health grants 1S10OD018530 and P41GM10349010. The laboratory work was conducted by Dr Ganapati Bhat and Dr Nitin Supekar.

The proMBP/PAPP-A1-complex was separated via SDS-PAGE (4-15 % polyacrylamide gel) under reducing conditions by the chemical DTT (figure 4.33). From each of the three lanes carrying 66.67 µg protein, the 200 kDa band was excised, which corresponds to the PAPP-A1 monomer. The sample was processed with trypsin and chymotrypsin for proteolytic digestion. A part of the sample was further treated with PNGase F, which released the N-glycans from the PAPP-A1 peptides. Released N-glycans were permethylated and analysed by MALDI-ToF-MS and NSI-MS. The remaining part of the sample was analysed for glycoproteomics by nLC-MS/MS.

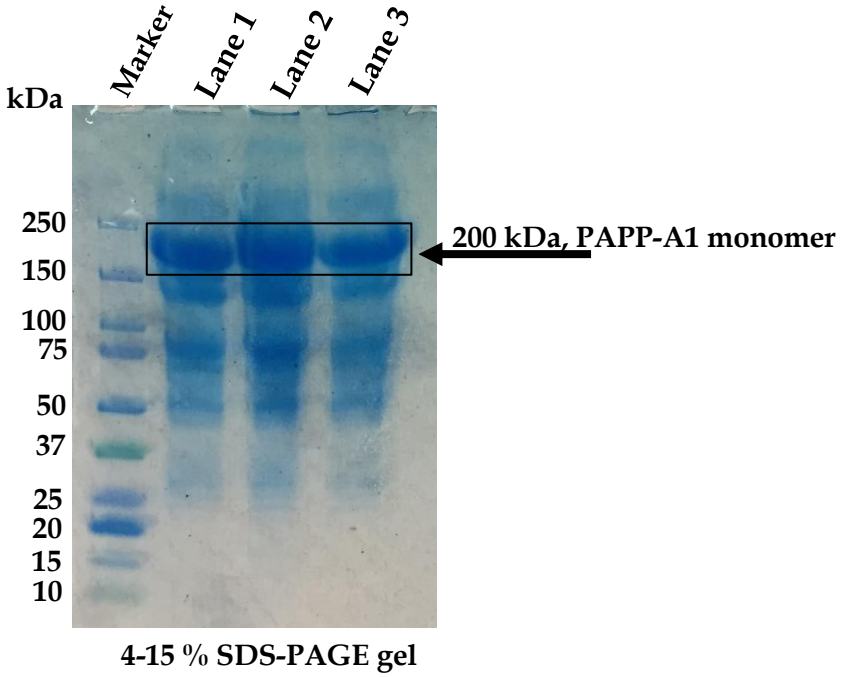


Figure 4.33: SDS-PAGE gel of the human native proMBP/PAPP-A1 complex. The gel ran under reducing conditions and lane 1-3 were loaded with 66.67 µg protein, respectively. The 200 kDa band of each lane was cut out and processed further for MS investigation of permethylated N-glycans. The marker lane served as reference for identification of the 200 kDa band.

The PAPP-A1 peptides after the PNGase F treatment were processed and analysed for O-glycopeptides. However, no O-glycans were detected. This finding is in agreement with the information available for PAPP-A1 (protein ID Q13219) from the established database Uniprot (The UniProt Consortium, 2001).

4.5.1 Glycomic analysis of the human native PAPP-A1/proMBP-complex

4.5.1.1 MALDI-ToF-MS analysis of released, per-methylated N-linked glycans

Matrix-assisted laser desorption ionisation-time-of-flight mass spectrometry data was acquired for permethylated N-glycans, which were released from native human proMBP/PAPP-A1-complex after extraction of the 200 kDa band from SDS-PAGE gel.

A representative MALDI-TOF spectrum of the permethylated N-glycans is shown in figure 4.34. The profile was obtained from the wash fraction of the C18 SPE cartridge as stated in section 2.3.1.3 a. The spectrum shows a very good signal-to-noise ratio (S/N) for the glycans, which is reflected in the clear distinction between signal peaks and the baseline. The baseline is caused by background ions that do not derive from the analytes, and is referred to as 'noise' (Krutchinsky and Chait, 2002). The spectrum was acquired with a maximum intensity of 2013.8 for the most abundant peak. Molecular ions were detected between m/z 1500-3900. Putative structures were based on the acquired data, the molecular weight, and on principles of the human N-glycan biosynthesis pathway (Kornfeld and Kornfeld, 1985). Through MALDI-ToF analysis, 13 different N-glycoforms were observed at m/z 1579.7441, 1783.8208, 1981.8928, 2028.9325, 2155.9700, 2185.9854, 2191.9919, 2390.0771, 2396.1822, 2792.2019, 2966.2622, 3602.4958 and 3776.5410 for human native PAPP-A1. The annotated data show that primarily high-mannose and also complex sialylated glycans with heterogeneous multi-antennary structures are present on human native PAPP-A1.

Since the MALDI technique transfers a higher degree of energy to the ions during the ionisation phase, false fragmentation of fragile glycan substituents can occur. This includes labile sialic acid and fucose residues and could be a reason for the observed abundance of high-mannose, non-fucosylated, non-sialylated N-glycans. All detected ions are singly charged sodium cationised ions, due to the nature of the MALDI-MS technique.

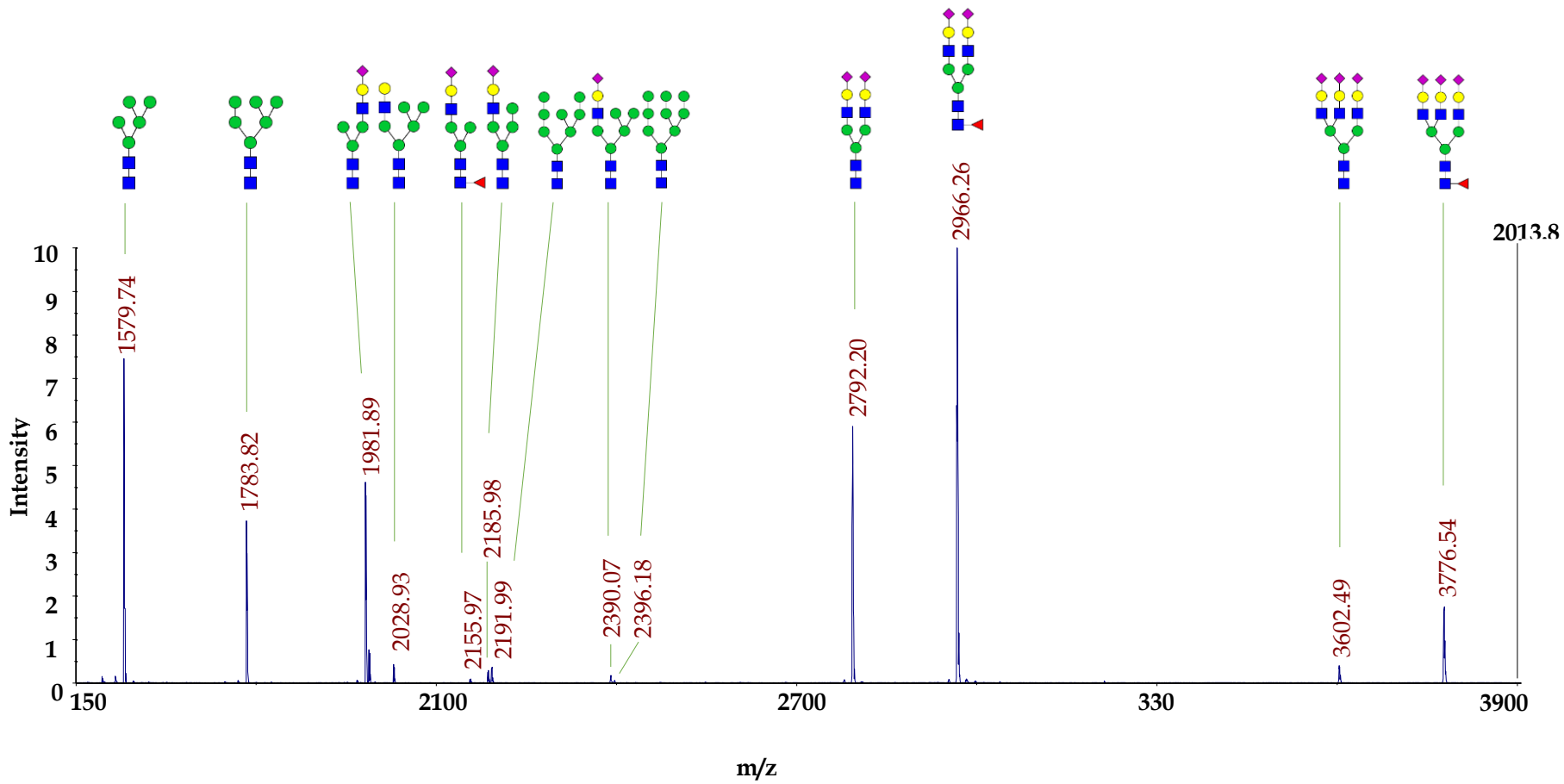


Figure 4.34: Annotated MALDI-TOF-MS spectrum (m/z 1500-3900) of permethylated N-glycans enzymatically released from the human native PAPP-A1, which was extracted from the 200 kDa band of a SDS-PAGE gel. The structures are based on the molecular weight, MS data and follow the principle of the N-glycan biosynthesis pathway in humans. Mainly bi- and tri-antennary high mannose and complex sialylated glycans are present on human native PAPP-A1.

Symbols: □ *GalNAc,* ■ *GlcNAc,* ● *Man,* ● *Glc,* ○ *Gal,* ▲ *Fuc,* ◆ *NeuAc*

4.5.1.2 NSI-MS analysis of released, permethylated N-linked Glycans

Nanospray ionisation (NSI)-mass spectrometry was used for the acquisition of data for released and permethylated N-glycans. These derived from the native human proMBP/PAPP-A1-complex after extraction of the 200 kDa band, which corresponds to the PAPP-A1 polypeptide monomer, from the reducing SDS-PAGE gel.

A representative NSI-MS spectrum of the permethylated N-glycans is shown in Figure 4.35. As stated in section 2.3.1.3 a, the profile was obtained from the wash fraction of the processing via the C18 SPE cartridge. The spectrum shows a good signal-to-noise ratio (S/N) for the glycans, based on the distinguishable analyte peaks. Since the NSI-MS technique is much more sensitive, more ions were visible in the overall spectrum. Molecular ions were detected between m/z 720–2000. Based on the molecular weight, the acquired data and principles of the human N-glycan biosynthesis pathway, putative structures were assigned.

The data complements the information revealed through MALDI-ToF-MS (see section '4.5.1.1 MALDI-ToF-MS analysis of released, per-methylated N-linked glycans'). Moreover, due to a lower amount of unintended fragmentation for the full MS data acquisition, more glycoforms were assignable. In detail, four more glycan structures were assigned, which totals in 17 detectable glycoforms via NSI-MS. The N-glycan peaks were observed at the m/z values of 801.3861, 903.4360, 1002.4863, 1025.9991, 1089.5307, 1104.5367, 1107.0372, 1206.0875, 1209.5851, 1407.6860, 1422.6904, 1487.7216, 1494.73013, 1582.2750 for ions with a charge state z=2. Three more N-glycans ionised with a charge state z=3 are resolved at m/z values of 1216.5883, 1274.9522, and 1332.9536 (depicted in figure 4.35). The charge states are caused by the comparatively slow evaporation of the droplet solvent during the NSI ionisation process. The additional glycoforms, which weren't evident in the MALDI-ToF-MS spectrum, carry more fragile glycan substituents, such as sialic acid and fucose residues. This finding is consistent with the literature as well as empirical data (Zaia, 2010, Shajahan et al., 2017a).

The analysis and interpretation of the MS data confirmed that high mannose and multi-antennary, sialylated complex glycans are mainly present on human native PAPP-A1. Notably, glycan species with and without additional fucosylation were

detectable, such as for the analyte ions with m/z values of 1407.6860, 1494.73013 and 1582.2750.

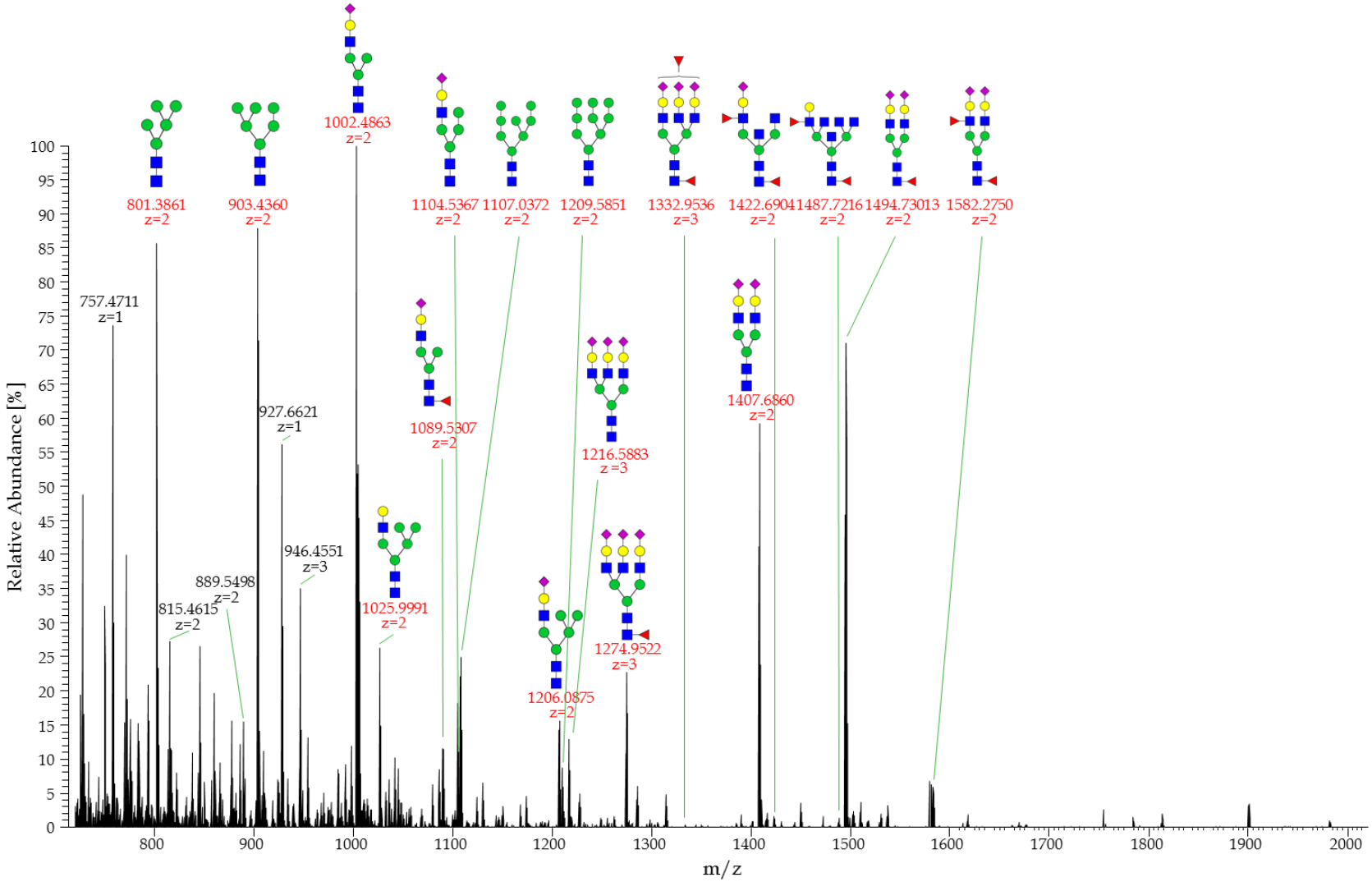


Figure 4.35: Annotated NSI-MS spectrum (m/z 720-2000) of permethylated N-glycans, which were enzymatically released from the human native PAPP-A1. PAPP-A1 was extracted from SDS-PAGE gel (200 kDa band). Annotated structures are based on the molecular weight, MS data and the human N-glycan biosynthesis pathway. High mannose and complex sialylated glycans with and without fucosylation are found to be present on human native PAPP-A1.

Symbols: □ *GalNAc,* ■ *GlcNAc,* ● *Man,* ● *Glc,* ○ *Gal,* ▲ *Fuc,* ◆ *NeuAc*

4.5.2 Glycoproteomic analysis of N-linked glycans on human native PAPP-A1/proMBP-complex

For the glycoproteomic approach, the technique of liquid chromatography tandem MS (LC-MS/MS) was the method of choice. MALDI-ToF was not applied since the energy is favourably absorbed by glycosidic bonds, which subsequently leads to a higher degree of glycosidic cleavages. The aim was to obtain site-specific information, such as the specific asparagine residues of the glycan attachment and also the prevalent glycoforms present at each site. Therefore, a MS technique that favours the cleavage of peptide bonds was required, such as NSI-MS. The site mapping was performed first, as the acquired data helped the following glycopeptide analysis. NSI-MS further allows for tandem MS (MS/MS or MS^2) fragmentation methods and can be refined through a prior liquid chromatography (LC) step, referred to as liquid chromatography-MS^n (LC-MS^n). The LC-NSI-MS/MS method is able to separate isobaric glycans, which have the same mass but different sugar compositions and/or conformations (Costello et al., 2007, Shajahan et al., 2017a).

As explicitly stated in '2.6 Methods for Mass Spectrometry', the fragmentation methods of higher-energy collisional dissociation (HCD) and subsequent, HCD-triggered collision–induced dissociation (CID) were used for analysis of (glyco)peptides derived from native human PAPP-A1. The database Uniprot lists the full FASTA peptide sequence of human PAPP-A1 with 1627 amino acids, including the signal peptide (aa 1-22) and the propeptide (aa 23-80) (The UniProt Consortium, 2001), as shown in figure 4.36.
The PAPP-A1/proMBP-complex was sourced from human retroplacental blood. Given the native state of the protein, the full human sequence and subsequent processing was expected, i.e. removal of the signal and propeptide. The remaining amino acid sequence of 1547 aa had been confirmed present - via proteomic analysis by mass spectrometry - prior to the glycan analysis.

All forthcoming references to amino acids represent the mature FASTA peptide sequence of human PAPP-A1 without the signal and propeptide. For convenience of

identification, the sequons of N-glycosylation sites will be presented in a peptide sequence of four amino acids.

```
MRLWSWVLHL GLLSAALGCG LAERPRRARR DPRAGRPPRP AAGPATCATR AARGRRASPP
PPPPPGGAWE AVRVPRRRQQ REARGATEEP SPPSRALYFS GRGEQLRLRA DLELPRDAFT
LQVWLRAEGG QRSPAVITGL YDKCSYISRD RGWVVGIHTI SDQDNKDPRY FFSLKTDRAR
QVTTINAHRS YLPGQWVYLA ATYDGQFMKL YVNGAQVATS GEQVGGIFSP LTQKCKVLML
GGSALNHNYR GYIEHFSLWK VARTQREILS DMETHGAHTA LPQLLLQENW DNVKHAWSPM
KDGSSPKVEF SNAHGFLLDT SLEPPLCGQT LCDNTEVIAS YNQLSSFRQP KVVRYRVVNL
YEDDHKNPTV TREQVDFQHH QLAEAFKQYN ISWELDVLEV SNSSLRRRLI LANCDISKIG
DENCDPECNH TLTGHDGGDC RHLRHPAFVK KQHNGVCDMD CNYERFNFDG GECCDPEITN
VTQTCFDPDS PHRAYLDVNE LKNILKLDGS THLNIFFAKS SEEELAGVAT WPWDKEALMH
LGGIVLNPSF YGMPGHTHTM IHEIGHSLGL YHVFRGISEI QSCSDPCMET EPSFETGDLC
NDTNPAPKHK SCGDPGPGND TCGFHSFFNT PYNNFMSYAD DDCTDSFTPN QVARMHCYLD
LVYQGWQPSR KPAPVALAPQ VLGHTTDSVT LEWFPPIDGH FFERELGSAC HLCLEGRILV
QYASNASSPM PCSPSGHWSP REAEGHPDVE QPCKSSVRTW SPNSAVNPHT VPPACPEPQG
CYLELEFLYP LVPESLTIWV TFVSTDWDSS GAVNDIKLLA VSGKNISLGP QNVFCDVPLT
IRLWDVGEEV YGIQIYTLDE HLEIDAAMLT STADTPLCLQ CKPLKYKVVR DPPLQMDVAS
ILHLNRKFVD MDLNLGSVYQ YWVITISGTE ESEPSPAVTY IHGSGYCGDG IIQKDQGEQC
DDMNKINGDG CSLFCRQEVS FNCIDEPSRC YFHDGDGVCE EFEQKTSIKD CGVYTPQGFL
DQWASNASVS HQDQQCPGWV IIGQPAASQV CRTKVIDLSE GISQHAWYPC TISYPYSQLA
QTTFWLRAYF SQPMVAAAVI VHLVTDGTYY GDQKQETISV QLLDTKDQSH DLGLHVLSCR
NNPLIIPVVH DLSQPFYHSQ AVRVSFSSPL VAISGVALRS FDNFDPVTLS SCQRGETYSP
AEQSCVHFAC EKTDCPELAV ENASLNCSSS DRYHGAQCTV SCRTGYVLQI RRDDELIKSQ
TGPSVTVTCT EGKWNKQVAC EPVDCSIPDH HQVYAASFSC PEGTTFGSQC SFQCRHPAQL
KGNNSLLTCM EDGLWSFPEA LCELMCLAPP PVPNADLQTA RCRENKHKVG SFCKYKCKPG
YHVPGSSRKS KKRAFKTQCT QDGSWQEGAC VPVTCDPPPP KFHGLYQCTN GFQFNSECRI
KCEDSDASQG LGSNVIHCRK DGTWNGSFHV CQEMQGQCSV PNELNSNLKL QCPDGYAIGS
ECATSCLDHN SESIILPMNV TVRDIPHWLN PTRVERVVCT AGLKWYPHPA LIHCVKGCEP
FMGDNYCDAI NNRAFCNYDG GDCCTSTVKT KKVTPFPMSC DLQGDCACRD PQAQEHSRKD
LRGYSHG
```

Figure 4.36: FASTA peptide sequence of human native PAPP-A1. The sequence includes the signal peptide sequence (aa 1-22; highlighted in light grey) and the propeptide sequence (aa 23-80; highlighted in dark grey). The mature PAPP-A1 protein consists of 1547 aa with 14 potential N-glycosylation sites (sequon N-X-S/T, X not P [valid sequons highlighted in yellow; sequons N-P-S/T are not glycosylated and highlighted in cyan]). Only 11 of the sites have been found to be occupied by Overgaard et al. (2003). The five complement control protein (CCP) modules, also called sushi domains, present in the native PAPP-A1 are highlighted in dark green (Sushi 1 aa 1213-1282, Sushi 2 aa 1283-1344, Sushi 3 aa 1345-1412, Sushi 4 aa 1413-1473, Sushi 5 aa 1476-1556): They fold in a beta-sandwich arrangement.

4.5.2.1 nLC-NSI-MS/MS analysis of N-linked glycoforms on glycopeptides from human native PAPP-A1

For the glycoproteomic analysis, an enzymatic digest was required to detect glycopeptides with the nanoLC-MS² technique. Per *in silico* digest, the proteases trypsin and chymotrypsin were chosen, since they yielded the best theoretical peptide coverage for mass spectrometry. Ideally, peptides are required to have a length between 5-30 aa to lie within the detection limit, and 113 of the 253 predicted peptides did match this criterion. More importantly, most peptides carrying an N-glycosylation sequon lay within this range, except for NISW (aa 310-313) and NGSF (aa 1385-1388).

Trypsin as well as chymotrypsin cleave before the substrate from the C-terminal side, i.e. the amino acid that served as proteolytic substrate is the last amino acid of the sequence in FASTA format. Trypsin cleaves from the C-terminus before amino acids with a positive charge, such as lysine (K) or arginine (R), if not followed by a proline. Chymotrypsin cleaves before C-terminal and bulky hydrophobic amino acids, e.g. tyrosine (Y), tryptophan (W), and phenylalanine (F). However, cleavage after smaller hydrophobic amino acids, such as methionine (M) and leucine (L), have also been described, but with slower rates of catalytic efficiency (Perona and Craik, 1995, Baird and Craik, 2013, Gráf et al., 2013).

Overgaard et al. (2003) published, on basis of liquid chromatography and N-terminal sequence analysis of (glyco)peptides, the occupation of 11 sequons within PAPP-A1. However, the publication clearly states that "[m]ass spectrometry of peptides substituted with carbohydrate was not attempted". The asparagine residues elucidated to carry N-linked glycosylation were N322 (NSSL), N349 (NHTL), N400 (NVTQ), N521 (NDTN), N539 (NDTC), N645 (NASS), N745 (NISL), N946 (NASV), N1146 (NCSS), N1243 (NNSL), and N1439 (NVTV) (Overgaard et al., 2003). This leaves the asparagine residues N310 (NISW), N1142 (NASL), and N1405 (NGSF) to be excluded from glycosylation within their study from 2003. Furthermore, no analysis of the types of carbohydrates attached to PAPP-A1 was undertaken at the time.

The following subchapters present the glycoforms identified on the confirmed glycosylation sites. The annotations in the full MS spectra illustrate the assigned

glycan structure based on the mass, the pathway of synthesis of N- linked glycans, and the species-specific glycosylation pattern for humans. The data was confirmed by HCD MS^2 data as well as CID MS^2 data (not shown).

The spectra do not contain information about linkages between the sugar moieties, since linkage information was not part of the analysis and requires additional methods. For a clearer arrangement of the annotated spectra, the peptide part of the glycoform is referred to as "peptide" in cases of shortage of space. Amino acids highlighted in red indicate a modification of the amino acid, i.e. glycan modifications for asparagine residues (N) or carbamidomethylation for cysteine residues (C).

a) Amino acid asparagine-349

The programme PeptideCutter was used to predict the potential cleavage sites around the sequon NHTL at asparagine-349 (N349) for mature human PAPP-A1. This resulted in the (glyco)peptide (K)IGDENCDPECNHTLTGHDGGDCR(H) for a combined *in silico* digestion with trypsin and chymotrypsin. The glycopeptide was identified and verified via b- and y-ions in the HCD MS^2 spectrum (data not shown). The peptide mass matched with all three cysteines being modified via carbamidomethylation (highlighted in red in figure 4.37).

Figure 4.37 shows the glycoforms that were determined for the glycosylation at site N349: Eights glycan structures were found in total - including three structures that are classified as high-mannose type (m/z 3845.4446, m/z 4007.4922, m/z 4169.5402), and five structures that belong to the hybrid type with sialylation (m/z 4177.5563, m/z 4339.6074, m/z 4501.6603) or without (m/z 4048.5121, m/z 4210.5662). The glycoforms possess either a bi-antennary or a tri-antennary structure.

The data was obtained via on-line nano-LC-MS/MS and figure 4.37 shows a deconvoluted full MS spectrum.

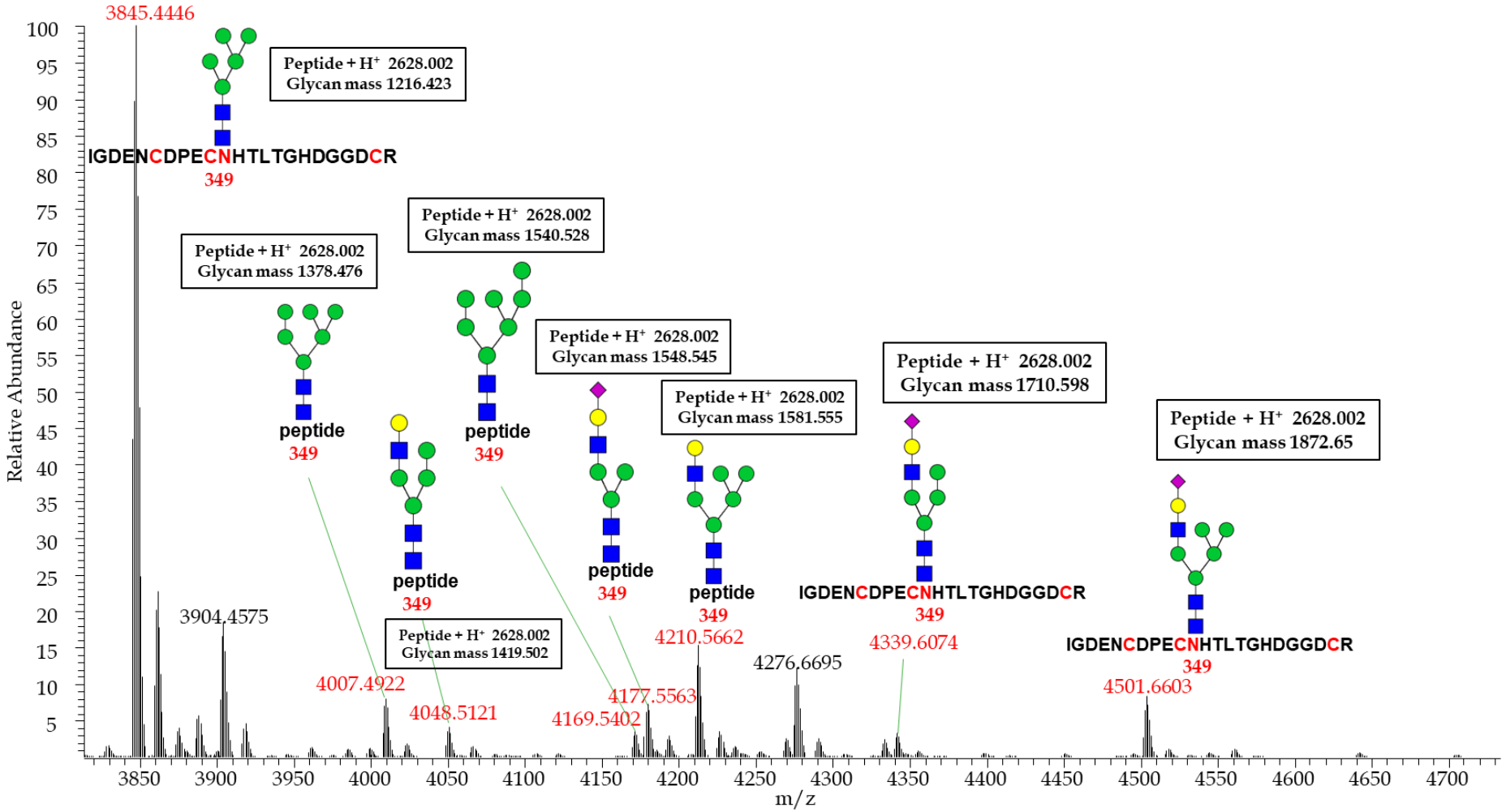


Figure 4.37: Deconvoluted and annotated full MS spectrum (m/z 3820-4730) of glycopeptide (K)IGDENCDPECNHTLTGHDGGDCR(H) derived from human native PAPP-A1 via nanoLC-NSI-MS/MS. Glycoforms identified and shown in this spectrum are glycan structures classified as either high-mannose type (m/z 3845.4446, m/z 4007.4922, m/z 4169.5402) or hybrid type (m/z 4048.5121, m/z 4177.5563, m/z 4210.5662, m/z 4339.6074, m/z 4501.6603).

Symbols: GalNAc, GlcNAc, Man, Glc, Gal, Fuc, NeuAc

b) Amino acid asparagine-400

The predicted (glyco)peptide, after *in silico* digestion with the programme PeptideCutter around the sequon NVTQ at asparagine-400 for the mature human PAPP-A1, was (R)FNFDGGECCDPEITNVTQTCFDPDSPHR(A). This finding matched with the identified sequence.

The glycosylation site at asparagine-400 was found to be occupied with two glycoforms of the bi-antennary complex glycan type (m/z 5520.1132, m/z 5666.1632), both carrying terminal sialyl moieties (figure 4.38.). The glycoform at m/z value of 5666.1632 was found to carry a core fucosylation.

The data was obtained via on-line nano-LC-MS/MS and the figure 4.38 shows a full MS spectrum.

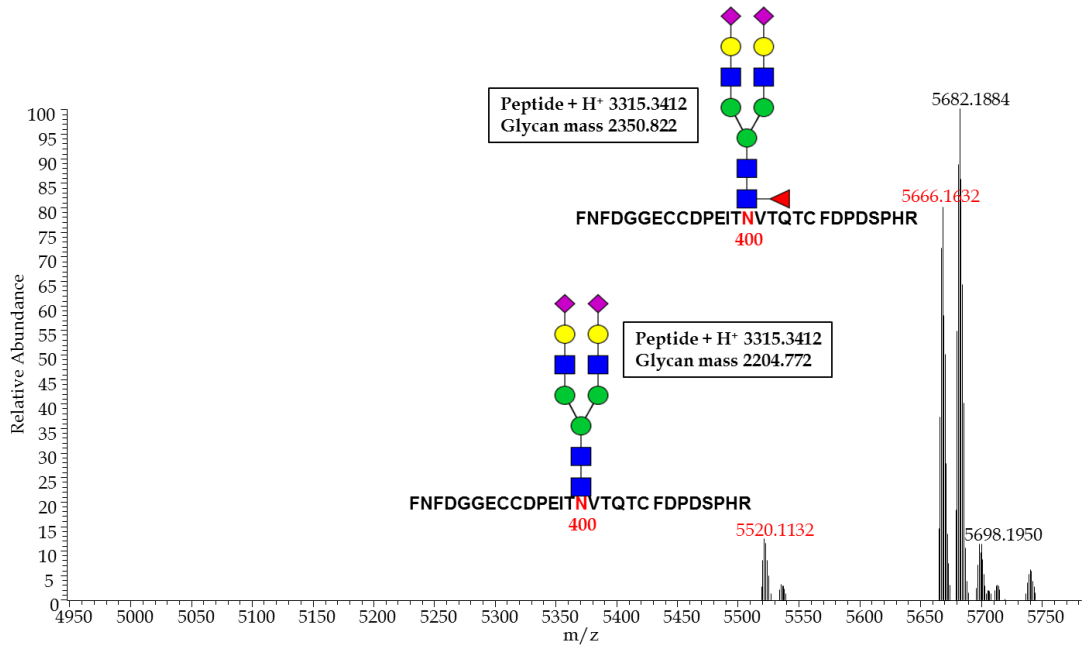


Figure 4.38: Deconvoluted and annotated full MS spectrum of the glycopeptide (R)FNFDGGECCDPEITNVTQTCFDPDSPHR(A) (m/z 4950-5780) derived from human native PAPP-A1 via nanoLC-NSI-MS/MS. Glycoforms identified in this spectrum are glycan structures classified as sialylated/fucosylated complex type (m/z 5520.1132, m/z 5666.1632).

Symbols: GalNAc, GlcNAc, Man, Glc, Gal, Fuc, NeuAc

c) Amino acid asparagine-521

The predicted (glyco)peptides around the sequon NDTN at asparagine-521 for the mature human PAPP-A1 were calculated as (R)GISEIQSCSDPCMETEPSF(E) and (F)ETGDLCNDTNPAPK(H) via an *in silico* digestion with both, trypsin and chymotrypsin. The glycopeptide found was (R)GISEIQSCSDPCMETEPSF-ETGDLCNDTNPAPK(H) for all glycopeptides. Considering the proline in proximity to the predicted cleavage site, a change in the secondary structure might have prevented chymotrypsin from cleavage before the amino acid phenylalanine-514, leading to the combined peptide species.

The glycopeptide was found to carry three glycoforms. Figure 4.39 depicts hybrid type glycans at m/z 5122.0259 (with fucosylation) and m/z 5154.0586 (without fucosylation), which were identified in a full MS spectrum from nano-LC-MS/MS. Furthermore, the glycoform found at m/z 5325.1279 is a complex glycan with fucosylation.

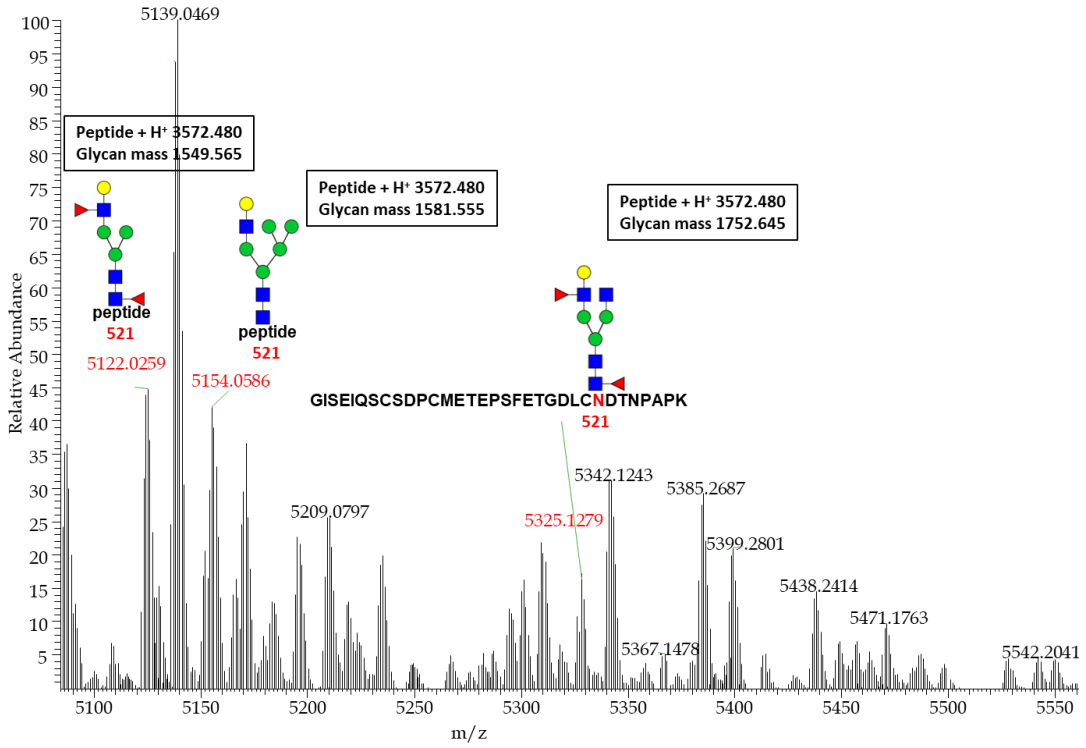


Figure 4.39: Deconvoluted and annotated full MS spectrum of the glycopeptide (R)GISEIQSCSDPCMETEPSFETGDLCNDTNPAPK(H) (m/z 4980-5510) derived from human native PAPP-A1 via nanoLC-NSI-MS/MS. The glycoforms identified at m/z 5122.0259 (with fucosylation) and m/z 5154.0586 classify as hybrid type and the glycoform at m/z 5325.1279 as a complex glycan with fucosylation.

Symbols: GalNAc, GlcNAc, Man, Glc, Gal, Fuc, NeuAc

d) Amino acid asparagine-539

After *in silico* digestion with trypsin and chymotrypsin, respectively, the programme PeptideCutter predicted the (glyco)peptide (K)SCGDPGPGNDTCGF(H) around the sequon NDTC at asparagine-539 (N539) for the mature human PAPP-A1. However, the identified sequence matched with (K)SCGDPGPGNDTCGFHSFFN(T), indicating that chymotrypsin did not cleave as predicted before the amino acid phenyl-alanine-547.

The glycosylation site at N539 was found to be occupied with two glycoforms of the high-mannose glycan type with m/z 3499.2822 and m/z 3661.335, as annotated in figure 4.40.

The data was retrieved from a full MS scan by on-line nano-LC-MS/MS and confirmed by HCD MS^2 data (not shown).

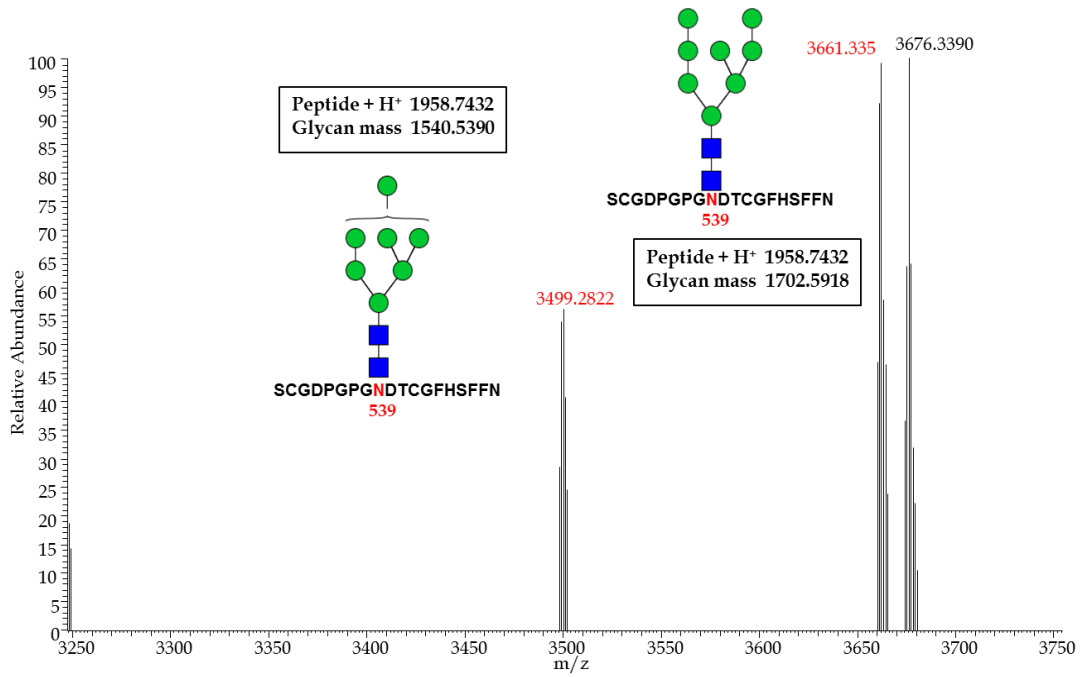


Figure 4.40: Deconvoluted and annotated full MS spectrum of the glycopeptide (K)SCGDPGPGNDTCGFHSFFN(T) (m/z 3250-3750) derived from human native PAPP-A1 via nanoLC-NSI-MS/MS. The identified glycoforms at m/z 3499.2822 and m/z 3661.335 are high-mannose glycan structures.

Symbols: GalNAc, GlcNAc, Man, Glc, Gal, Fuc, NeuAc

e) Amino acid asparagine-645

Via the programme PeptideCutter, the potential cleavage sites around the sequon NASS at asparagine-645 (N645) were predicted for mature human PAPP-A1. This resulted in the theoretical (glyco)peptide (Y)ASNASSPMPCSPSGHW(S) for a combined *in silico* digestion with trypsin and chymotrypsin. However, the identified glycopeptide had the aa sequence (Y)ASNASSPMPCSPSGHWSPR(E). The chymotrypsin did not cleave before the tryptophan-658 for the yielded glycopeptide. The peptide mass matched with a modification of the cysteine residue via carbamidomethylation.

The glycosylation site at N645 was found to be occupied with four glycoforms, of which two belong to the high-mannose type (m/z 3067.2395, m/z 3229.2937) and two are of the complex glycan type (m/z 4217.6449, m/z 4234.63650) with sialylation of the antennae (figure 4.41).

The data was retrieved via on-line nano-LC-MS/MS and the figure depicts a full MS spectrum.

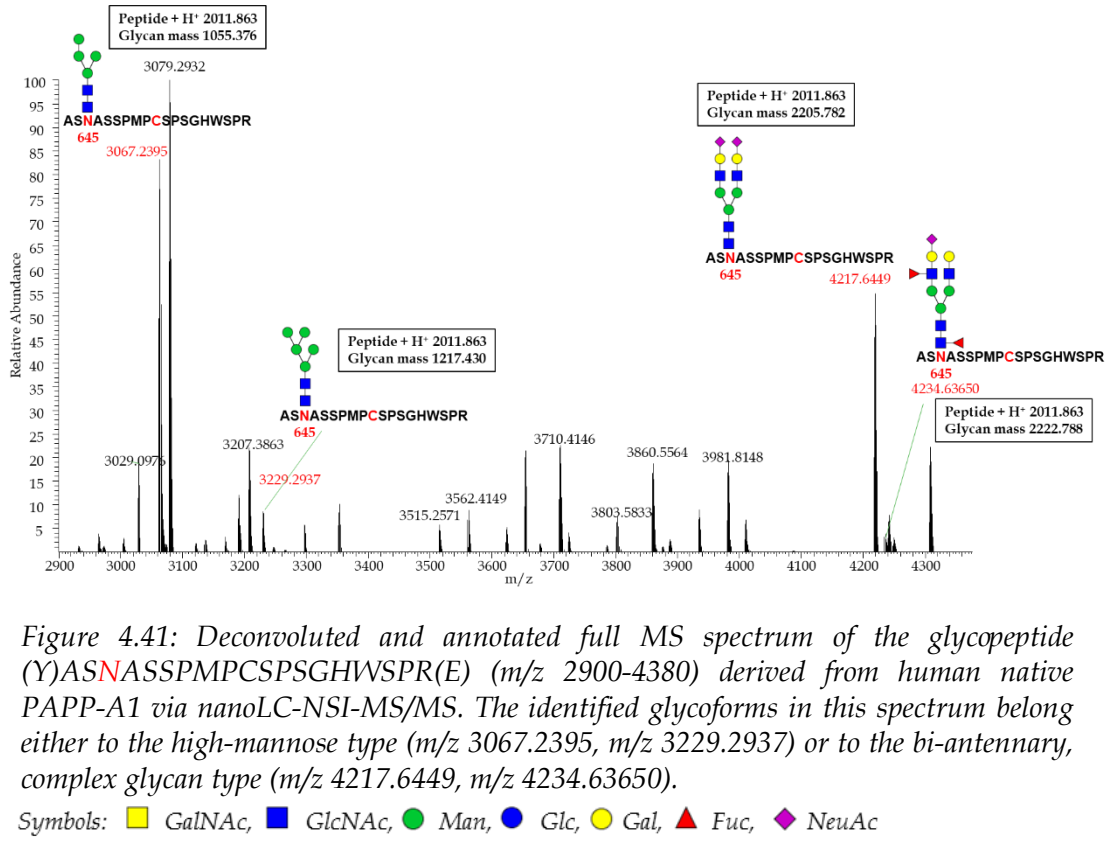


Figure 4.41: Deconvoluted and annotated full MS spectrum of the glycopeptide (Y)ASNASSPMPCSPSGHWSPR(E) (m/z 2900-4380) derived from human native PAPP-A1 via nanoLC-NSI-MS/MS. The identified glycoforms in this spectrum belong either to the high-mannose type (m/z 3067.2395, m/z 3229.2937) or to the bi-antennary, complex glycan type (m/z 4217.6449, m/z 4234.63650).

Symbols: GalNAc, GlcNAc, Man, Glc, Gal, Fuc, NeuAc

f) Amino acid asparagine-745

The predicted (glyco)peptide after *in silico* digestion with the programme PeptideCutter was (K)NISLGPQNVF(C) for the area around the sequon NISL at asparagine-745 (N745) for the mature human PAPP-A1. The identified and verified (glyco)peptide sequences were (K)NISLGPQNVF(C) and (K)NISLGPQNVF-CDVPLTIR(L), respectively. For the latter sequence, chymotrypsin did not cleave after the amino acid phenylalanine-754.

Four out of the five occupying glycoforms found at asparagine-745 were classified as complex type with terminal sialyl residues only (m/z 3949.606 in figure 4.44 and m/z 4247.8412 in figure 4.43), or core fucosylation as well as sialyl residues on terminal ends of the glycan structures (m/z 4095.656 in figure 4.44., m/z 4393.8912 in figure 4.43.). A glycan of the hybrid type with core fucosylation and terminal sialylation is shown in figure 4.42 at m/z 2783.176.

The data was obtained via on-line nano-LC-MS/MS. All figures show full MS spectra.

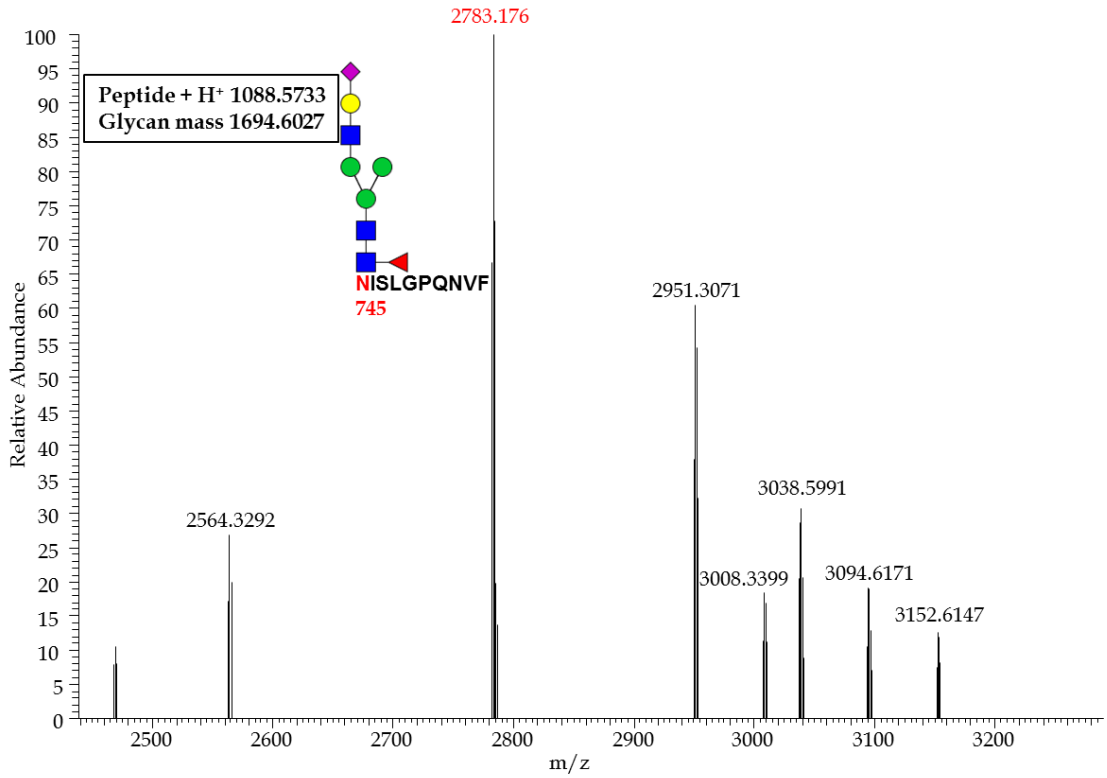


Figure 4.42: Deconvoluted and annotated full MS spectrum (m/z 2440-3290) of the glycopeptide (K)NISLGPQNVF(C) derived from human native PAPP-A1 via nanoLC-NSI-MS/MS. The identified glycoform at m/z 2783.176 in this spectrum (m/z 2440-3280) is classified as hybrid glycan with core fucosylation and sialylation on one terminal end.
Symbols: GalNAc, GlcNAc, Man, Glc, Gal, Fuc, NeuAc

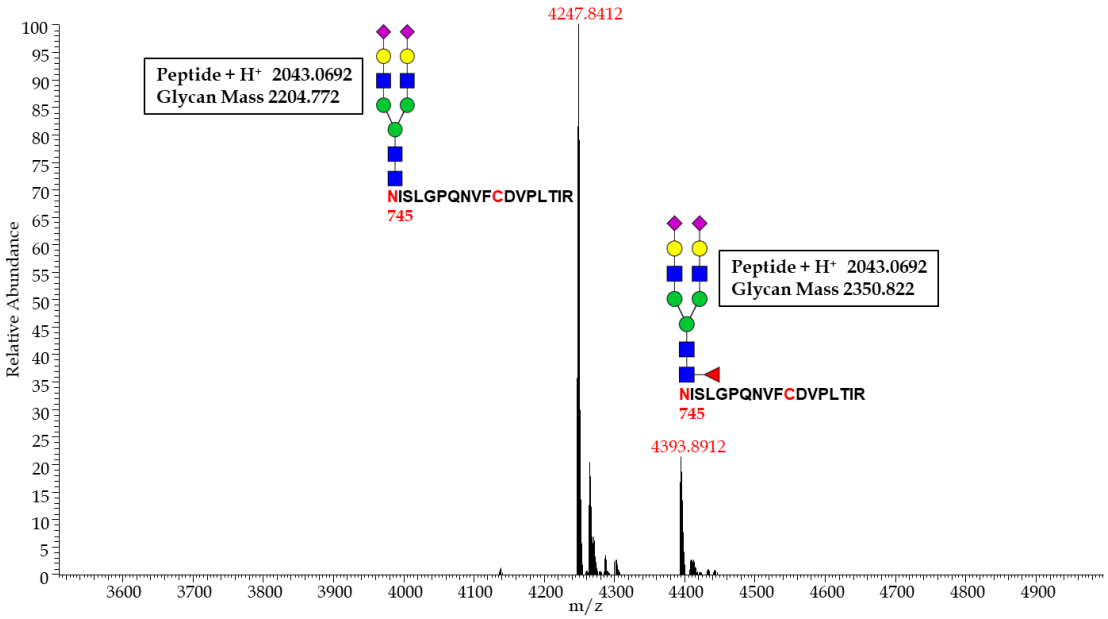


Figure 4.43: Deconvoluted and annotated full MS spectrum (m/z 3510-4990) of the glycopeptide (K)NISLGPQNVFCDVPLTIR(L) (cysteine-755 carbamidomethylated) derived from human native PAPP-A1 via nanoLC-NSI-MS/MS. The identified glycoforms in this spectrum (m/z 3510-5000) are classified as bi-antennary, sialylated complex glycans non-fucosylated (m/z 4247.8412) and with core fucosylation (m/z 4393.8912), respectively.

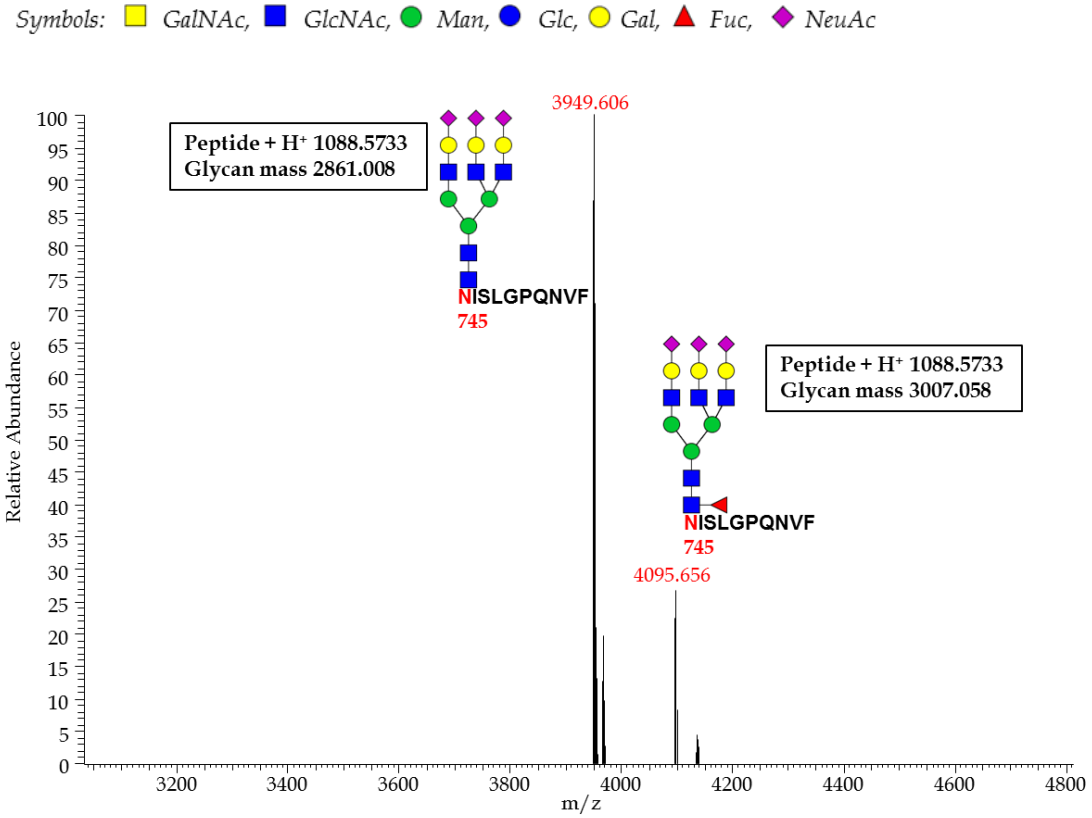


Figure 4.44: Deconvoluted and annotated full MS spectrum (m/z 3150-4800) of the glycopeptide (K)NISLGPQNVF(C) derived from human native PAPP-A1 via nanoLC-NSI-MS/MS. The identified glycoforms in this spectrum (m/z 3050-4800) are classified as tri-antennary complex glycans with terminal with core fucosylation (m/z 4095.656) and without fucose residues (m/z 3949.606), respectively.

g) Amino acid asparagine-946

Via the programme PeptideCutter, the potential cleavage sites around the sequon NASV at asparagine-946 (N946) were elucidated for mature human PAPP-A1. The (glyco)peptide (W)ASNASVSHQDQQCPGW(V) was predicted for a combined *in silico* digest with trypsin and chymotrypsin. The detected glycopeptide had the amino acid sequence (Y)TPQGFLDQWASNASVSHQDQQCPGW(V), which represents two missed cleavages by chymotrypsin before phenylalanine-939 as well as tryptophan-943. Furthermore, the cysteine-956 was found to be carbamidomethylated.

The glycosylation site at N946 was found to be occupied with a bi-antennary glycan structure of the complex type at m/z 5195.07209 with fucosylation of the core and sialylation of the antennae (figure 4.45.). The data was retrieved via on-line nano-LC-MS/MS and is depicts a full MS spectrum.

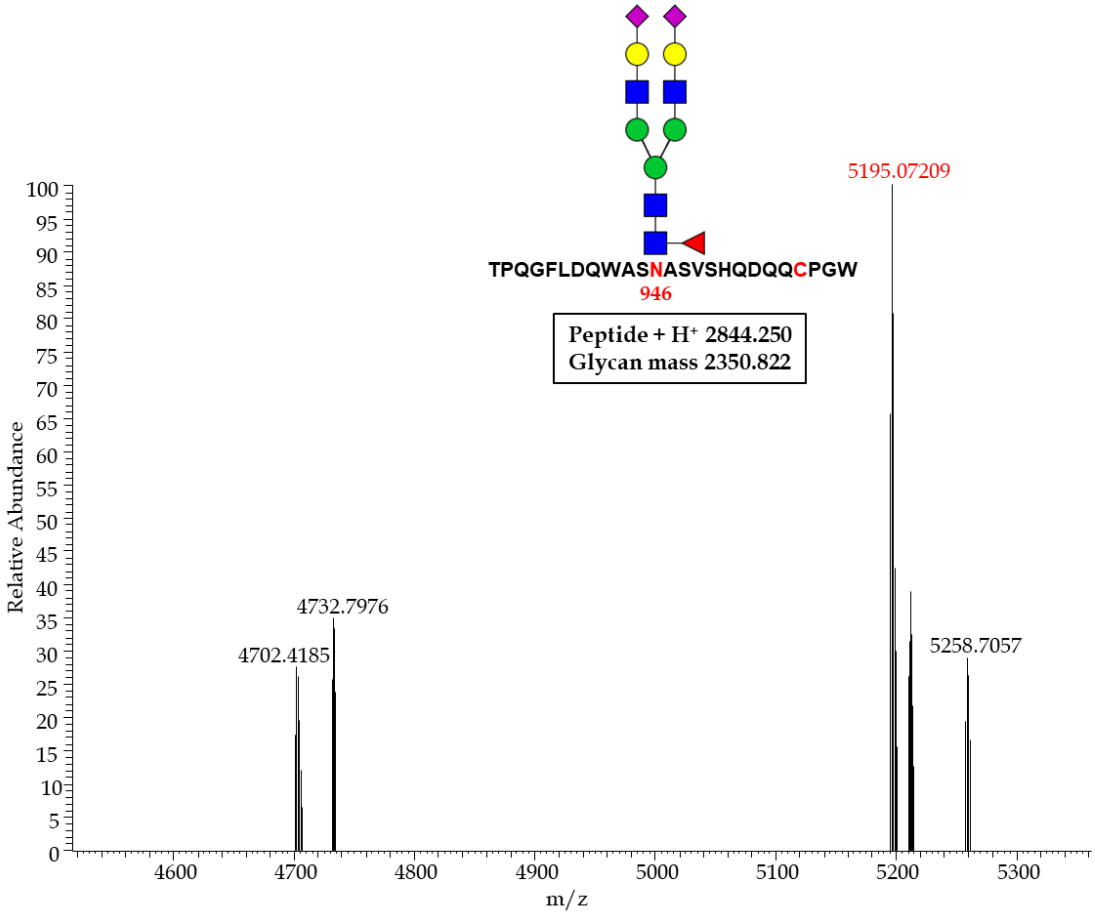


Figure 4.45: Deconvoluted and annotated full MS spectrum of the glycopeptide (Y)TPQGFLDQWASNASVSHQDQQCPGW(V) (carbamidomethylated cysteine-956) derived from human native PAPP-A1 via nanoLC-NSI-MS/MS. The identified glycoform in this spectrum (m/z 4520-5360) is classified as sialylated glycan of the complex type with core fucosylation (m/z 5195.07209).

Symbols: GalNAc, GlcNAc, Man, Glc, Gal, Fuc, NeuAc

h) Amino acids asparagine-1142 and asparagine-1146

The programme PeptideCutter was used to predict the potential cleavage sites around the sequon NASL at asparagine-1142 (N1142) and NCSS at asparagine-1146 (N1146) for mature human PAPP-A1. The (glyco)peptide (K)TDCPELAVE-NASLNCSSSDR(Y) was expected for a combined *in silico* digest with trypsin and chymotrypsin. The predicted (glyco)peptide sequence was confirmed and carries two N-glycosylation sequons: NASL and NCSS. Therefore, it is not possible to confidently assign the glycoforms on each site but determine the glycoforms that are found on the glycopeptide with the two N-glycosylation motifs.
The peptide mass matched with modifications (both are highlighted in red in figure 4.46) of both cysteine residues: One has been found to be modified via carb-amidomethylation and the other via propionamidylation. The propionamide is an artefact from the SDS-PAGE (see section 4.5).

The glycopeptide - carrying the glycosylation sites N1142 and N1146 - was found to be occupied with sialylated glycoforms of the complex type (figure 4.46). Complex, bi-antennary and sialylated glycan structures with core fucosylation (m/z 4589.7950) or without core fucosylation (m/z 4152.6417, m/z 4443.73715) were identified. Furthermore, tri-antennary glycans of the complex type with sialylation of the antennae (m/z 5099.9647) and additional core fucosylation (m/z 5246.0226), respectively, were assigned.

The data was obtained via on-line nano-LC-MS/MS and the figure shows a full MS spectrum.

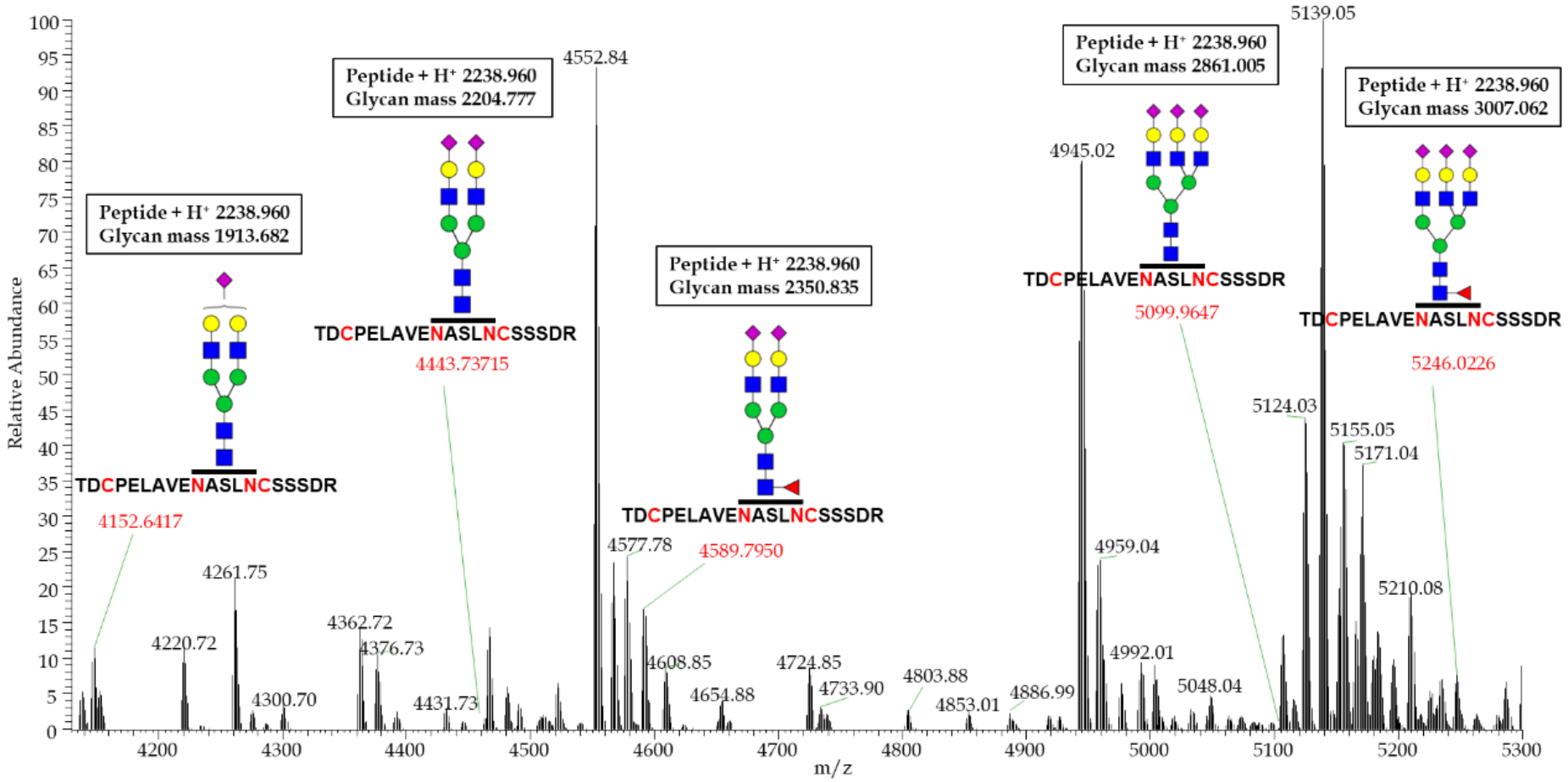


Figure 4.46: Deconvoluted and annotated full MS spectrum of the glycopeptide (K)TDCPELAVENASLNCSSSDR(Y) derived from human native PAPP-A1 via nanoLC-NSI-MS/MS. Both cysteine residues were identified as modified – one via carbamidomethylation and one via propionamidylation. The identified glycoforms in this spectrum (m/z 4130-5300) are classified as sialylated glycans of the complex type with fucosyl residues at the core (m/z 4589.7950, 5246.0226) or without (m/z 4152.6417, m/z 4443.73715, m/z 5099.9647).

Symbols: GalNAc, GlcNAc, Man, Glc, Gal, Fuc, NeuAc

i) Amino acid asparagine-1243

Via the programme PeptideCutter, the potential cleavage sites around the sequon NSLL at asparagine-1243 (N1243) were predicted for mature human PAPP-A1. The theoretical (glyco)peptide (K)GNNSLLTCMEDGLW(S) for a combined *in silico* enzymatic digest with trypsin and chymotrypsin was calculated. However, the identified glycopeptide had the aa sequences (K)GNNSLL(T). Interestingly, PeptideCutter did predict the cleavage site before amino acid threonine-148 for a chymotryptic digest only, but not for the combined digestion.

The glycosylation site at N1243 was found to be occupied with a bi-antennary complex glycan type that possesses terminal sialic acid residues as well as a core fucose (m/z 2968.14728), as depicted in the full MS spectrum in figure 4.47. The data was retrieved via on-line nano-LC-MS/MS.

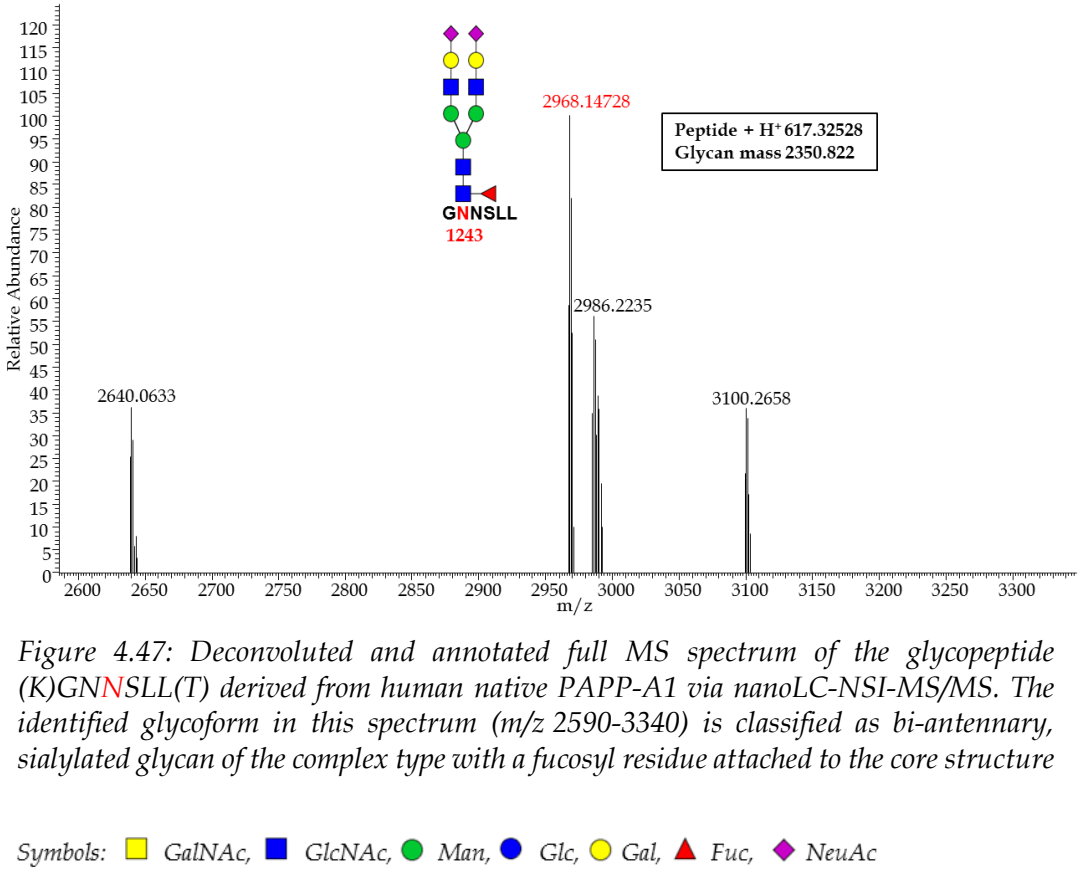


Figure 4.47: Deconvoluted and annotated full MS spectrum of the glycopeptide (K)GNNSLL(T) derived from human native PAPP-A1 via nanoLC-NSI-MS/MS. The identified glycoform in this spectrum (m/z 2590-3340) is classified as bi-antennary, sialylated glycan of the complex type with a fucosyl residue attached to the core structure

j) Amino acid asparagine-1439

The predicted (glyco)peptide after *in silico* digestion with the programme PeptideCutter was (Y)AIGSECATSCLDHNSESIILPMNVTVR(D) for the area around the sequon NVTV at asparagine-1439 (N1439) for the mature human PAPP-A1. The (glyco)peptide sequence found was (M)NVTVR(D) instead. Even though methionine has been reported as a rare cleavage site for chymotrypsin, PeptideCutter did not predict it for the amino acid methionine-1438.

The glycoform found at asparagine-1439 was classified as bi-antennary complex type with sialylation and core-fucosylation (m/z 2939.1787), as shown in the full MS spectrum in figure 4.48.

The data was obtained via on-line nano-LC-MS/MS.

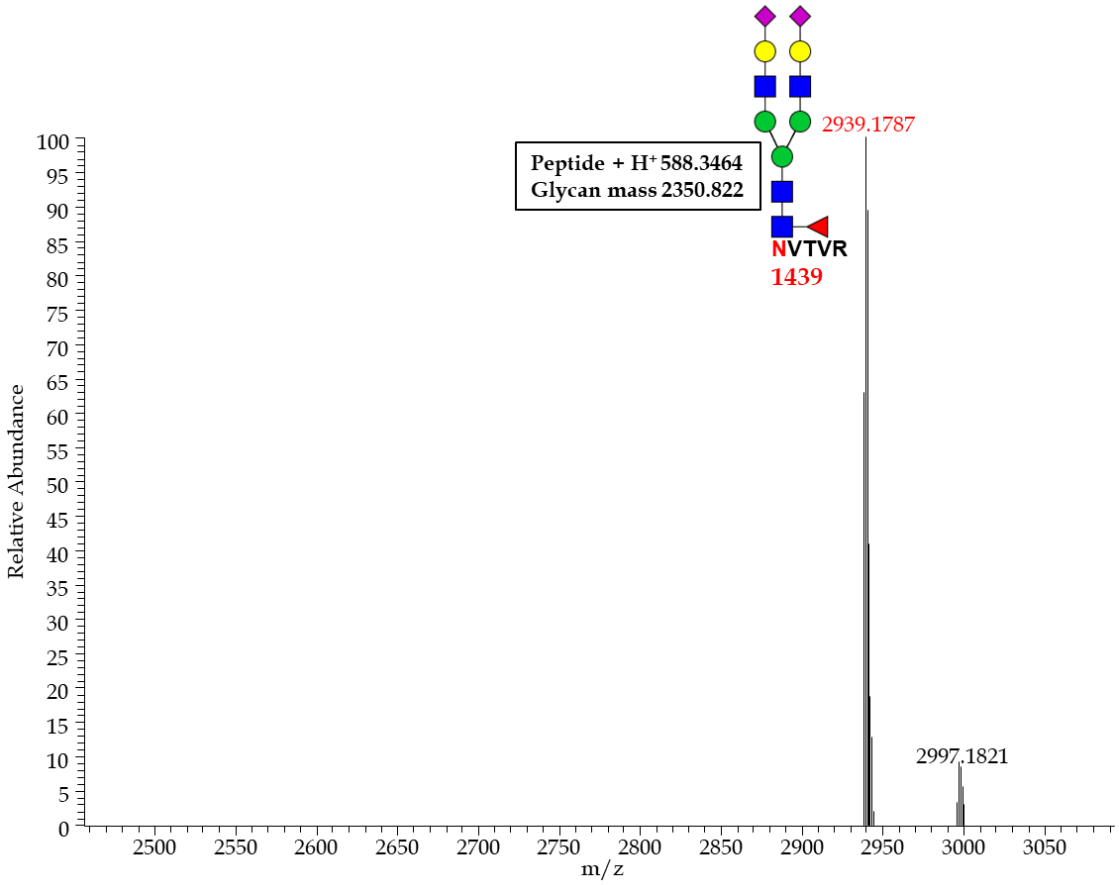


Figure 4.48: Deconvoluted and annotated full MS spectrum of the glycopeptide (M)NVTVR(D) derived from human native PAPP-A1 via nanoLC-NSI-MS/MS. The identified glycoform in this spectrum (m/z 2460-3090) is classified as bi-antennary, sialylated glycan of the complex type with fucosylation of the core (m/z 2939.1787).

Symbols: GalNAc, GlcNAc, Man, Glc, Gal, Fuc, NeuAc

4.5.3 Conclusions from the glycomic and glycoproteomic analyses of the human native PAPP-A1/proMBP-complex

The glycoproteomic site analysis of the glycopeptides derived from human native PAPP-A1 revealed that the asparagine residues N349 (NHTL), N400 (NVTQ), N521 (NDTN), N539 (NDTC), N645 (NASS), N745 (NISL), N946 (NASV), N1142 (NASL), N1146 (NCSS), N1243 (NNSL), and N1439 (NVTV) were modified by glycosylation. Hence, the asparagine residues N310 (NISW), N322 (NSSL), and N1405 (NGSF) were found without glycan modifications. As expected, all three sequons with a proline (N-P-S/T) did not carry glycan structures, which confirms the existing criterion for N-linked glycosylation sequons. In summary, 11 out of 14 putative N-glycosylation sites were found to be occupied with N-linked complex carbohydrate structures in this study.
The analysis and characterisation of the glycan structure revealed that PAPP-A1 is modified with glycoforms of the high-mannose, hybrid, and complex type, respectively; whereby most sequons showed multiple glycoforms attached. A summary of the findings is gathered in table 19.

It is important to point out that the glycoproteomic data obtained with nLC-NIS-MS/MS data (section 4.5.2.) matched with the glycomic data that was acquired via MALDI-ToF-MS and is presented in section 4.5.1. in this chapter.

The human native PAPP-A1 is primarily modified with high-mannose glycans but also complex sialylated glycans with heterogeneous multi-antennary structures. Bisection of the complex glycans is not prevalent and, therefore, the core mannose residues of the glycans theoretically without steric hindrance. An interaction between the glycan structures present on the native human PAPP-A1 and the CRD of DC-SIGN as well as DC-SIGNR is very likely, given their binding prevalence for mannose-type sugars.

No O-linked glycosylation was detected for this sample of human native PAPP-A1, which is in line with past studies aiming for elucidation of glycosylation of PAPP-A1.

Table 19: Summary of the glycoproteomic analysis of N-linked glycans identified on human native PAPP-A1. Out of the 14 N-glycan sequons, 11 have been found glycosylated including (partially fucosylated/sialylated) glycoforms of the high-mannose, hybrid, and complex type, respectively. The asparagine linked to the glycan structure (N) as well as carbamidomethylation/ propionamidylation of cysteine (C) are highlighted in red.

Amino Acid/ Glycosylation site	Glycosylation sequon	FASTA peptide sequence	Type of N-glycosylation
310	NISW	NISW	Not glycosylated
322	NSSL	ELDVLEVSNSSLR	Not glycosylated
349	NHTL	IGDENCDPECNHTLTGHDGGDCR	High-mannose, Hybrid [sialylated]
400	NVTQ	FNFDGGECCDPEITNVTQTCFDPDSPHR	Complex [sialylated, fucosylated]
521	NDTN	GISEIQSCSDPCMETEPSFETGDLCNDTNPAPK	Hybrid [fucosylated]
539	NDTC	SCGDPGPGNDTCGFHSFFN	High-mannose
645	NASS	ASNASSPMPCSPSGHWSPR	High-mannose, Complex [fucosylated, sialylated]
745	NISL	NISLGPQNVFCDVPLTIR NISLGPQNVF	Hybrid [fucosylated, sialylated], Complex [fucosylated, sialylated]
946	NASV	TPQGFLDQWASNASVSHQDQQCPGW	Complex [fucosylated, sialylated]
1142 1146	NASL NCSS	TDCPELAVENASLNCSSSDR TDCPELAVENASLNCSSSDR TDCPELAVENASLNCSSSDR	Complex [fucosylated, sialylated]
1243	NNSL	GNNSLL	Complex [fucosylated, sialylated]
1405	NGSF	DGTWNGSFHVCQEMQGQCSVPNELNSNLK	Not glycosylated
1439	NVTV	NVTVR	Complex [fucosylated, sialylated]

4.6 Comparison of the mass spectrometric analyses of complex-carbohydrate structures on recombinant proMBP, recombinant PAPP-A1 and native PAPP-A1 from the proMBP/ PAPP-A1-complex

The samples with the human protein sequences of recombinant proMBP, recombinant PAPP-A1, as well as native PAPP-A1 (derived from the human proMBP/PAPP-A1 complex) were investigated for their glycosylation patterns by the use of glyco(proteo)mic mass spectrometric approaches.

The sample of recombinant proMBP (expressed by human cells; cell line undisclosed) did not display any high-mannose type N-linked glycoforms. Furthermore, half of the glycoforms of the presented complex type N-glycans carried a bisecting GalNAc residue. This finding might indicate an exclusion of any high affinity and specific interactions with the CRDs of mannose-type binding C-type lectins, such as DC-SIGN and DC-SIGNR.

Since DC-SIGN is also known to bind fucosylated Lewis-type antigens, the O-glycosylation of the recombinant proMBP was also of interest. However, only O-glycans of the core 1 and 3 type (including extension and branching) were detected. Therefore, the O-glycans on proMBP are not likely to contribute any high affinity binding epitopes for the CRD of DC-SIGN.

The sample of human recombinant PAPP-A1, which was expressed in a murine myeloma cell line, was found to be modified with N-linked glycoforms of all types: high-mannose, hybrid and complex glycans. Unfortunately, not all of the putatively N-glycan-carrying peptide sequons were covered by MS data. Furthermore, no

bisected complex glycans were found and only a fractional occupation with high-mannose glycoforms was determined. Therefore, it is expected that DC-SIGN and DC-SIGNR binding to this preparation of recombinant PAPP-A1 will be reduced compared to native PAPP-A1, which has been shown to display multiple high-mannose N-glycans.

The discovery of a NeuGc residue in one glycoform, which is very rare in human adults but abundant in murine glycosylation patterns, reflects the importance of species-specific investigations within the field of glycobiology.

In line with past studies, no O-linked glycosylation was found.

The human native PAPP-A1 sample (derived from human retroplacental blood) was found to be post-translationally modified with a great range of high-mannose, hybrid and complex type glycan structures on 11 out of the 14 putative N-glycosylation sites. Overgaard et al. (2003) found asparagine-310, asparagine-1142 and asparagine-1405 were not glycosylated; whereby the analysis of MS data within this study revealed no glycosylation at asparagine-310, asparagine-322 and asparagine-1405 for the native PAPP-A1 sample. The prevalence of high-mannose structures as well as a great proportion of complex type N-glycans, provides, theoretically, numerous high affinity binding epitopes for the CRDs of DC-SIGN and DC-SIGNR in the presence of calcium.

In this study, no O-linked glycosylation was found attached to human native PAPP-A1, which confirms the findings of past studies.

The analysis of the MS data showed that the pattern of glycosylation between all three glycoproteins differed and supports the determination of varying binding affinities for these three glycoproteins and the extracellular domains of DC-SIGN and DC-SIGNR.

From the MS data as well as kinetic data (chapter 3), the mannose-type binding carbohydrate recognition domains of DC-SIGN and DC-SIGNR are very likely to show the highest affinity of interaction with the human native PAPP-A1 glycopolypeptides within the human PAPP-A1/proMBP-complex. Human native

PAPP-A1 presents the most mannose-binding epitopes within its pattern of glycosylation, from which the most likely binding contacts will be positioned for the formation of any physiological complex between DC-SIGN and PAPP-A1/proMBP.

5 Results and Discussion (III)

Biophysical Interaction Studies between Human C-Type Lectin DC-SIGN and Synthetic Glycopolymers

5.1 Background

Specific carbohydrate patterns are recognised by the CRDs of C-type lectin receptors (CLRs), depending on their sugar-binding specificity and the clustering of CRDs within oligomers. These carbohydrate patterns enable the host to differentiate between foreign structures, which could indicate contact with a harmful pathogen, and self/host structures via interactions with a range of lectins. The recognition of foreign structures via immune-related C-type lectins triggers immune responses, which can be of a pro-inflammatory or anti-inflammatory nature. Specifically designed ligands, such as synthetic glycopolymers, play not only a crucial role for a better understanding of the mechanisms and principles behind the recognition of ligands via CRDs, but also have the potential to lead towards the development of either anti-pathogen agents or immune-regulating drugs.

For the characterisation of interactions between synthetic, well-defined glyco-polymers and DC-SIGN, binding studies were carried out via surface plasmon resonance (SPR) spectroscopy.
The three types of mannose glycopolymers that were investigated had a defined shape and a conserved molecular weight of 25 kDa. The glycopolymer analytes for the binding studies were designed as either star-shaped - in the case of the 5-arm and 8-arm glycopolymers - or linear - as for the DP58 glycopolymer. The structures are depicted in figure 5.1.

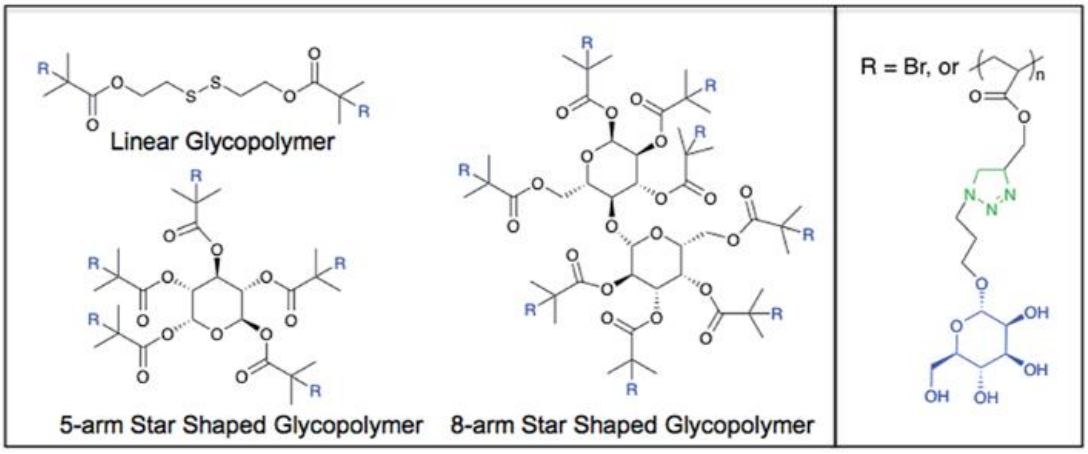


Figure 5.1: Representative chemical structures of the mannose glycopolymers. R= Br in the initiator structure (adapted from Mitchell et al. (2017) - Published by The Royal Society of Chemistry)

The glycopolymers were synthesised through living radical polymerisation and reflecting architectures of either a star of five or eight arms radiating from a multivalent initiator structure (5-arm, 8-arm) or a simple linear polymer of set degree of polymerisation (DP58, containing 58 mannose units).
The glycopolymer synthesis was carried out by the group of Dr Remzi Becer at the School of Engineering and Materials Science, Queen Mary University of London.

The data has been presented within the scope of a presentation as well as a report for the Upgrading from MPhil to PhD candidate in October 2015 for this specific degree at the University of Warwick.

A part of the study has been successfully published as a peer-reviewed scientific article in the journal of *Chemical Science* by the *Royal Society of Chemistry* [please see the appendix for (Mitchell et al., 2017)].

5.2 Biophysical interaction studies with synthetic glycopolymers and DC-SIGN

The interaction between the soluble, extracellular domain of the C-type lectin DC-SIGN and the glycopolymers 5-arm, 8-arm and DP58 were investigated. The concentration-dependent testing of the ligands was carried out in the presence of calcium.

The recombinant soluble DC-SIGN protein was immobilised (0.02 mg/mL; loading buffer 10 mM HEPES pH 5.0, 150 mM NaCl) as ligand on a 6-channel GLM sensor chip, which uses the chemistry of amine-coupling for binding. The experimental setup for the interaction study included a baseline step with assay buffer (10 mM HEPES pH 7.4, 150mM NaCl, 5mM$CaCl_2$, 0.01% (v/v) Tween-20, 0.01% (w/v) NaN_3) for 250 sec, an association step with the analyte for 900 s, and a dissociation step with assay buffer for 600 s. Each polymer was run (flow rate 25 µL/min; buffer 10 mM HEPES pH 7.4, 150mM NaCl, 5mM$CaCl_2$, 0.01% (v/v) Tween-20, 0.01% (w/v) NaN_3) in a 1:2-dilution series from 500 nM to 244 pM during the association step. The software programme ProteOn Manager was used for data presentation and analysis.

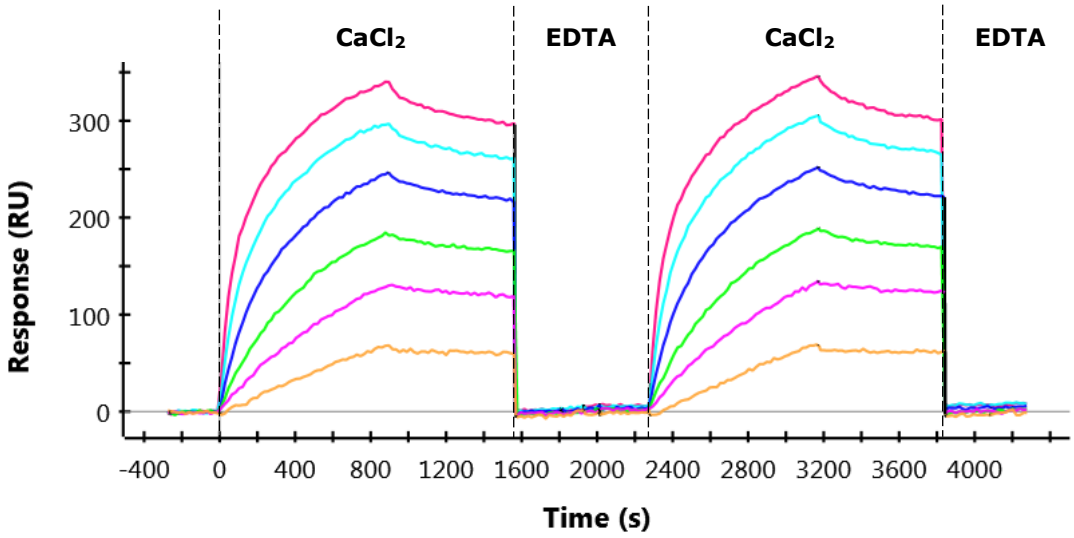


Figure 5.2: Representative sensorgram of the nature of the interaction between the mannose glycopolymers, depicted is DP58, and C-type lectin DC-SIGN. All interactions were specific, given that no signal was detectable in the absence of calcium and presence of the ion chelator EDTA. In fact, assay buffer with EDTA was used for regeneration of the GLM chip between the experiments.

The SPR approach allowed for the identification of the putative interactions between the CRD of DC-SIGN and the glycopolymers. Furthermore, binding kinetics of the interaction were determined. The data is presented in the following subchapters.
The specificity of the interactions between the CRD of DC-SIGN and the glycopolymers was tested and confirmed, since interaction signals (in response units [RU]) were only detectable in the presence of 5 mM calcium, but not 5 mM EDTA (see figure 5.2).

5.2.1 Star-shaped 5-arm glycopolymer

Interaction between the star-shaped 5-arm glycopolymer and the extracellular domain of DC-SIGN was investigated via surface plasmon resonance (SPR). The 5-arm glycopolymer showed a concentration-dependent binding to DC-SIGN within a serial dilution ranging from 500 nM to 0.244 nM. The representative sensorgrams, depicted in figure 5.5 and 5.6, were analysed and enabled the determination of the kinetic parameters of the interaction.

Binding kinetics of the interaction between the polymers and the human recombinant DC-SIGN extracellular domain were obtained via the analysis of association rate k_{on} and dissociation rate k_{off} of the detected interactions. The overview of the acquired data is presented in table 20.

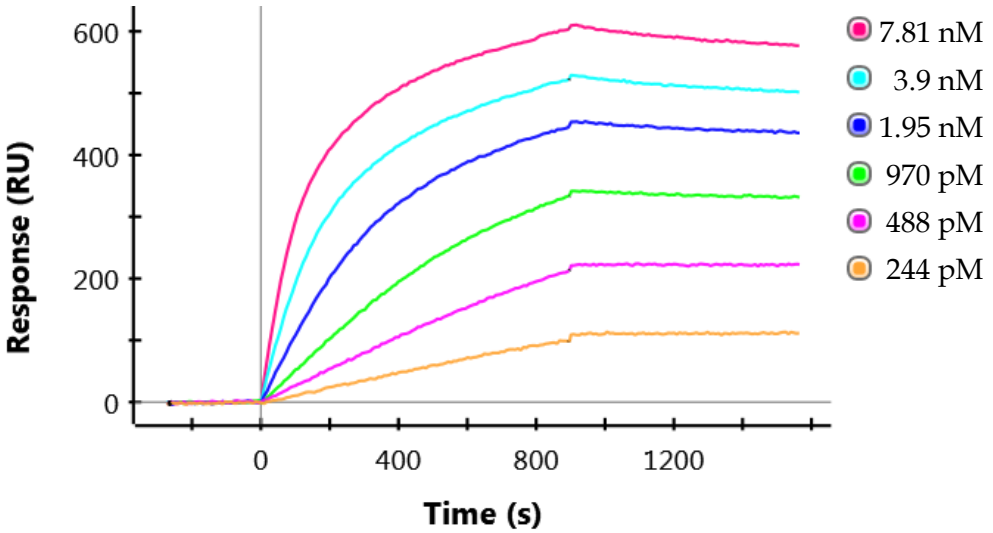


Figure 5.3: Representative sensorgram of the binding study between the 5-arm glycopolymer in the lower concentration range (between 0.244 nM and 7.81 nM) and DC-SIGN. Binding kinetics of the interaction between the polymers and the human recombinant DC-SIGN extracellular domain were derived from association and dissociation analysis of the detected interactions.

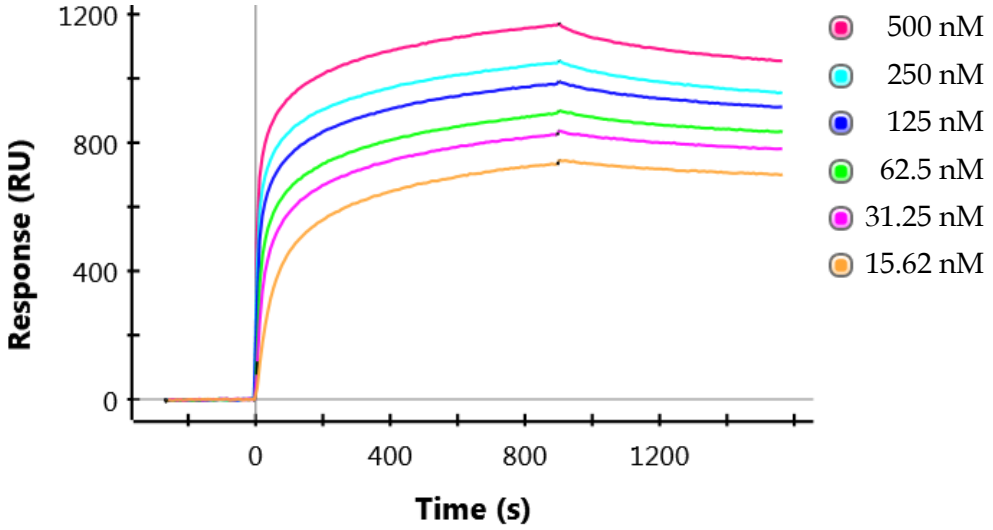


Figure 5.4: Representative sensorgram of the binding study between the 5-arm glycopolymer in the higher concentration range (between 15.62 nM and 500 nM) and soluble DC-SIGN. The association and dissociation phase allowed for detection and calculation of the kinetic parameters.

Table 20: Overview of the kinetic parameters obtained from the binding studies between the glycopolymer 5-arm and the extracellular domain of DC-SIGN.

Experiment	k_{on} **(M^{-1} s^{-1})**	k_{off} **(s^{-1})**	K_D **(M)**
5-arm polymer run1	1.01E6	7.33E-5	7.25E-11
5-arm polymer run2	1.01E6	7.04E-5	7.00E-11
5-arm polymer set2 run1	7.97E5	7.37E-5	9.25E-11
5-arm polymer set2 run2	7.93E5	7.01E-5	8.84E-11

The average dissociation constant of the 5-arm glycopolymers was calculated with K_D=80.8 pM. This reflects a comparatively very high affinity interaction, since most ligands interact with DC-SIGN within the nanomolar to submillimolar range.

5.2.2 Star-shaped 8-arm glycopolymer

The 8-arm star-shaped glycopolymer was investigated for its binding affinity to the extracellular domain of DC-SIGN.

The polymer with a measured molecular weight of 25 kDa showed binding to DC-SIGN (figure 5.7 and figure 5.8). The kinetic analysis was based on the acquired association rate k_{on} and dissociation rate k_{off} of the detected interaction, which is summarised in table 21.

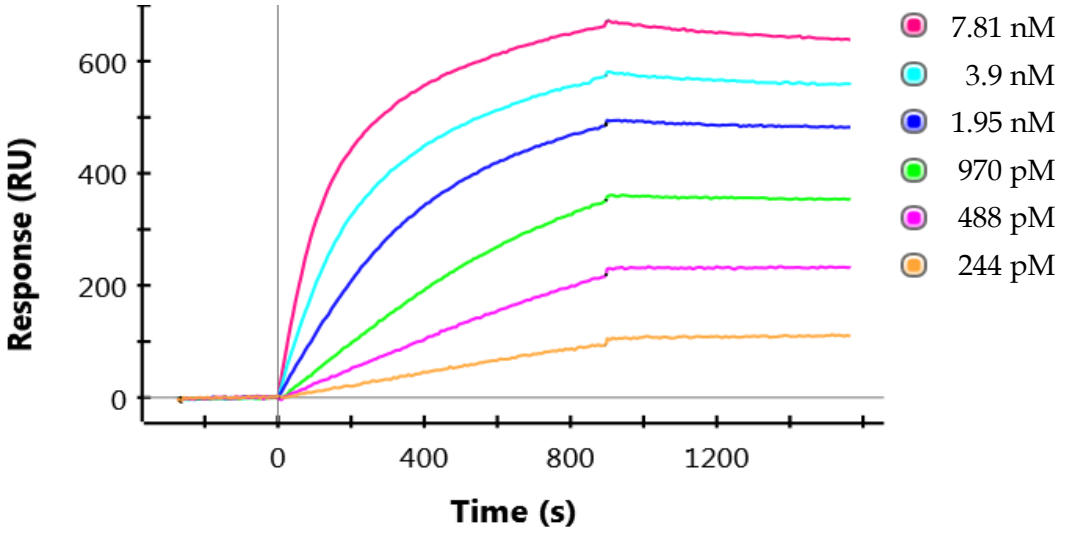


Figure 5.5: Representative SPR sensorgram of the interaction between DC-SIGN and the glycopolymer 8-arm within a concentration range between 0.244 nM and 7.81 nM (see legend for annotation). The kinetic analysis was based on the acquired data and showed a K_D^{8arm}=69.3 pM.

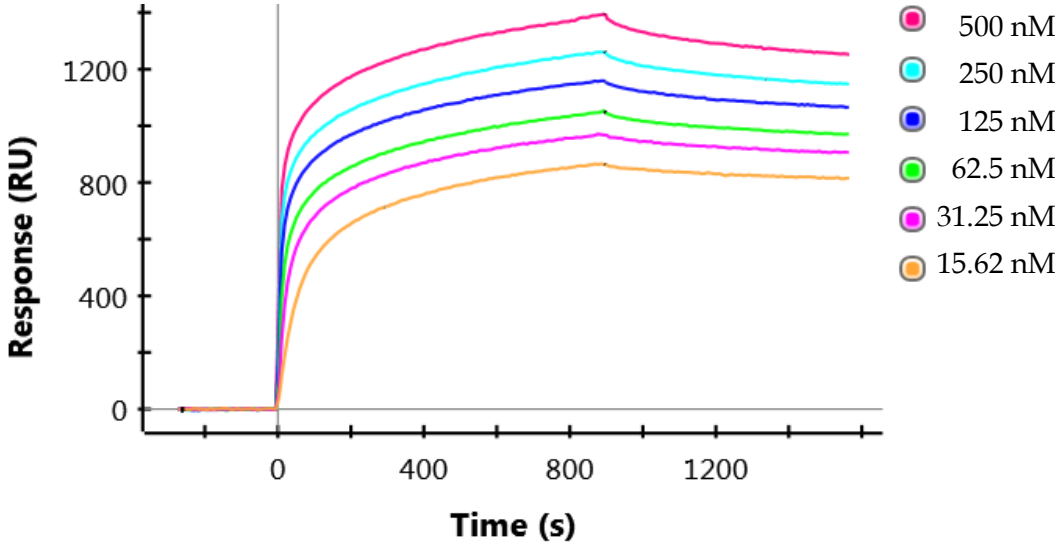


Figure 5.6: Representative SPR sensorgram of higher concentrations of the 8-arm glycopolymer (500 nM-15.62 nM) and soluble DC-SIGN (see legend for annotation). The graphs were analysed for kinetic parameters k_{on}, k_{off}, and resulting K_D.

Table 21: Overview of the kinetic parameters obtained from the binding studies between the 8-arm glycopolymer and the soluble polypeptide that encodes the extracellular domain of DC-SIGN.

Experiment	k_{on} **(M^{-1} s^{-1})**	k_{off} **(s^{-1})**	K_D **(M)**
8-arm polymer run1	9.27E5	6.03E-5	6.51E-11
8-arm polymer run2	9.29E5	5.18E-5	5.58E-11
8-arm polymer set2 run1	7.19E5	6.10E-5	8.47E-11
8-arm polymer set2 run2	7.27E5	5.19E-5	7.14E-11

Specifically, the average dissociation constants for the 8-arm glycopolymer was calculated as K_D=69.3 pM, which reveals a very high affinity interaction with DC-SIGN.

5.2.3 Linear DP58 glycopolymer

A glycopolymer with a degree of polymerisation (DP) value of 58 (DP58) of mannose units was tested as binding partner for DC-SIGN. DP58 has a measured molecular weight of 25 kDa and was able to bind to DC-SIGN with a very high affinity of K_D= 114 pM (figure 5.9 and figure 5.10). The calculated kinetic data is shown in table 22 and derived from the analysis of the association and dissociation phase of the detected interactions.

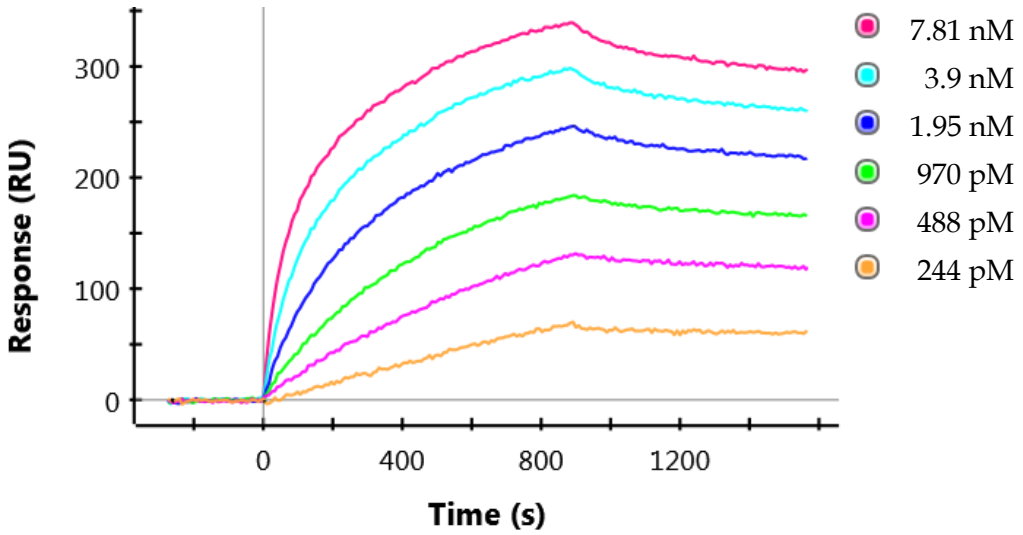


Figure 5.7: Representative SPR sensorgram of the interaction between DC-SIGN and the glycopolymer DP58 within a concentration range between 0.244 nM and 7.81 nM, as annotated in the legend. The kinetic analysis was based on the acquired data and revealed a K_D^{DP58}= 114 pM.

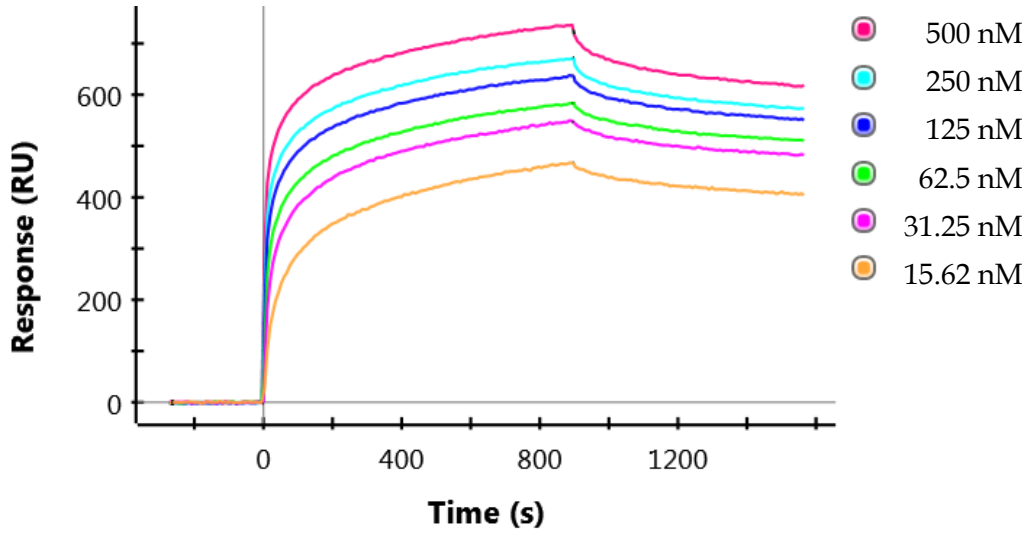


Figure 5.8: Representative SPR sensorgram of the interaction between DC-SIGN and the glycopolymer DP58 within a concentration range between 500 nM and 15.62 nM, as annotated in the legend. The kinetic analysis was based on the acquired data. The calculated average dissociation constant revealed a K_D^{DP58}= 114 pM.

Table 22: Overview of the kinetic parameters obtained from the binding studies between the glycopolymer DP58 and the soluble DC-SIGN, which consists of the extracellular domain.

Experiment	k_{on} **(M^{-1} s^{-1})**	k_{off} **(s^{-1})**	K_D **(M)**
DP58 polymer run1	1.24E6	1.34E-4	1.08E-10
DP58 polymer run2	1.27E6	1.37E-4	1.08E-10
DP58 polymer set2 run1	1.04E6	1.35E-4	1.30E-10
DP58 polymer set2 run2	9.52E5	1.04E-4	1.09E-10

The calculated average dissociation constant of the DP58 glycopolymer showed a very high affinity interaction with a K_D=114 pM.

5.3 Conclusion of the interaction studies between C-type lectins and synthetic glycopolymers

The binding studies, conducted via the biophysical method of SPR technique, clearly showed that the glycopolymers 5-arm, 8-arm and DP58 bound specifically to the CRD of DC-SIGN, i.e. only in the presence of calcium.

The tetramer of DC-SIGN is known to interact with most of its test ligands within the submillimolar range. However, the dissociation constants of the glycopolymers were determined to lie in the picomolar range (K_D^{5arm}=80.8 pM, K_D^{8arm}=69.3 pM, K_D^{DP58}=114 pM), which can be classified as very high affinity interaction. The star-shaped glycopolymers bound with an even higher affinity than the linear DP58 polymer; suggesting that the clustering of mannose units – realised by the star arrangement – increases the favourability of the interaction. Furthermore, these polymers are capable of disrupting interactions between DC-SIGN and other ligands, such as HIV gp120.

The principle would be particularly interesting for deeper study of immune-pathology and autoimmunity, where advanced glycopolymers could lead to the generation of drug-like materials for immune modulation via crosstalk with other pattern recognition receptors, which are present on dendritic cell (DC) and macrophage surfaces. The opportunity of triggering anti-inflammatory responses in immune cells expressing DC-SIGN has potential to lead to treat conditions, such as rheumatoid arthritis, or provide probes for the translational investigation of the processes of inflammation.

The same glycopolymer sets were used in dendritic cell analysis by Dr Shan Herath and Dr Steve Patterson at Imperial College London. The glycopolymers were shown to invoke elevated interleukin-10 and reduced interleukin-12p70 production, when co-incubated with LPS (Mitchell et al., 2017). This cytokine profile is a signature of the wound healing phenotype of DCs and macrophages.

These findings were part of a study that was published as a peer-reviewed scientific article (please see the appendix) in the journal of *Chemical Science* by the *Royal Society of Chemistry*:

MITCHELL, D. A., ZHANG, Q., VOORHAAR, L., HADDLETON, D. M., HERATH, S., GLEINICH, A. S., RANDEVA, H. S., CRISPIN, M., LEHNERT, H., WALLIS, R., PATTERSON, S. & BECER, C. R. 2017. Manipulation of cytokine secretion in human dendritic cells using glycopolymers with picomolar affinity for DC-SIGN. *Chemical Science,* 8**,** 6974-6980.

6 General Discussion and Future Work

Here, I would like to address the accomplished Thesis Aims set out in Chapter 1:

AIM ONE - To detect and quantify interactions between C-type lectins, such as DC-SIGN, DC-SIGNR and MMR, and candidate secreted human glycoproteins

Results described in Chapter 3, using biolayer interferometry, showed successful detection and quantification of calcium-dependent interactions between the human C-type lectin DC-SIGN and the candidate human glycoprotein PAPP-A1. For this interaction, the dissociation constant was determined as being in the low nanomolar range and, therefore, lies well within the physiologically-relevant range, which strengthens the hypothesis that PAPP-A1 serves as a novel, soluble endogenous ligand for DC-SIGN. Evidence of DC-SIGN binding to recombinant human PAPP-A1 was also shown, although the interaction was determined to be of lower affinity, suggesting that the glycosylation profile of this preparation is different to the native form.

Binding responses for two other C-type lectins, namely DC-SIGNR and MMR, were either comparatively low or inconclusive. Whilst further development of the BLI platform may reveal a level of PAPP-A1 binding to these molecules, the overall data support the hypothesis that PAPP-A1 is a much more specific ligand for DC-SIGN. This evidence of greater discrimination demonstrates the fundamental importance of higher-order lectin structure in defining physiologically relevant lectin-glycoprotein couplings.

AIM TWO - To characterise and assign the glycans of the PAPP-A1/proMBP-complex – a secreted ligand for the extracellular CRD of human DC-SIGN

The data presented in Chapter 4 demonstrate the ability of powerful and sensitive mass spectrometric methods, incorporating on-line nano-scale chromatography, to determine the detailed glycosylation patterns for human recombinant PAPP-A1,

human recombinant proMBP and human native PAPP-A1. N-glycosylation site assignment was also accomplished in these studies. The identified N-linked glycan types in these samples confirm the presence of complex carbohydrates favourable for DC-SIGN interactions and illustrate with great clarity the structural basis for measured differences in DC-SIGN binding affinities for native and recombinant forms of PAPP-A1. The data also represent pioneering work in understanding the structure and function of PAPP-A1 – an important protein in pregnancy and cardiovascular disease diagnostics.

AIM THREE - To characterise advanced synthetic glycoconjugate-ligand interactions with the C-type lectin DC-SIGN and relate them to natural ligands

The data presented in Chapter 5 show novel glycopolymer interactions with very high affinity for DC-SIGN via SPR. Remarkably, the dissociation constants were determined as lying in the picomolar range (K_D^{5arm}=80.8 pM, K_D^{8arm}=69.3 pM, K_D^{DP58}=114 pM), which - at the time of writing this thesis - are the highest affinity values obtained to date for DC-SIGN ligands. These glycopolymers will be particularly useful for further studies of mechanisms and principles behind the recognition of specific glycan arrangements via the multiple CRDs within the tetrameric DC-SIGN molecule arrangement.

Moreover, the glycopolymer sets were used in dendritic cell analysis by Dr Shan Herath and Dr Steve Patterson at Imperial College London. As the peer-review publication that I co-authored states, the glycopolymers triggered elevated interleukin-10 and reduced interleukin-12p70 production during dendritic cell activation with LPS (Mitchell et al., 2017)]. This cytokine profile is known to represent the phenotype of wound healing of the host via DCs and macrophages, raising the exciting possibility that glycopolymers of this kind could be adapted for therapeutic use in clinical scenarios such as post-surgical care and treatment of autoimmune disease.

PAPP-A1 (Pregnancy-associated Plasma Protein A1; also Pappalysin-1) inspired in-depth glycoproteomic investigation following its novel detection in an affinity chromatography screen using immobilised DC-SIGN and solubilised placental extract (Ilyas, 2010).

PAPP-A1 was originally discovered in the 1970s within studies to detect pregnancy-specific antigens in the maternal plasma. The protein is abundant in pregnant women – appearing early in pregnancy and peaking at term at concentrations of about 85 nM (Gall and Halbert, 1972, Lin et al., 1974). Non-pregnant, healthy males and females usually have a plasma level of PAPP-A1 between 0-0.8 nM.

PAPP-A1 itself is a metalloprotease of the metzincin superfamily with an intrinsic proteolytic function. The protease is secreted by fibroblasts from a wide range of tissues and is involved in the insulin-like growth factor (IGF) regulation cascade by cleaving IGF-binding proteins (IGFBP) 4 and 5 via its zinc-binding catalytic motif (Lawrence et al., 1999), which subsequently increases IGF bioavailability. These cleavage events are significant in aspects of healthcare - IGFBP4 and IGFBP5 are considered to inhibit cancer cell growth, wherein free IGF-1 can trigger oncogenic pathways. Therefore, PAPP-A1 has been described as a pro-oncogenic factor (Guo et al., 2018).

Within the last decade, free serum PAPP-A1 has been positively associated with elevated risk of acute coronary syndromes and complications associated with patients requiring haemodialysis in both, males and non-pregnant females (Bayes-Genis et al., 2001, Gururajan et al., 2012). In the context of cardiovascular diseases, PAPP-A1 is proteolytically fully active and, therefore, presumably involved in driving the IGF cascade, which appears to contribute to atherosclerosis cardiac hypertrophy.

It has been found that a decreased availability of IGF, resulting from a depletion of the enzymatic function of PAPP-A1 in gene-targeted PAPP-A1 knockout mice, is linked to a prolonged life span and, thus, slower aging (Conover and Oxvig, 2017). Within these pathways, it is reasoned that future research could promise a new approach to understanding the processes of aging, which includes the treatment and possibly partial prevention and attenuation of key age-related diseases such as cancer

and neurodegeneration. It is fascinating to contemplate the involvement of PAPP-A1 glycosylation and lectin interactions in this high-profile field of study.

Conversely, the proteolytic site of PAPP-A1 is almost exclusively inhibited by formation of a complex with the proform of eosinophil major basic protein (proMBP) during pregnancy (Oxvig et al., 1993). This complex was identified as a heterotetramer, i.e. 2:2-PAPP-A1/proMBP-complex, covalently bound via disulfide bonds (Christiansen et al., 2000, Overgaard et al., 2003). PAPP-A1 has been used for decades as a reliable biomarker marker in first-trimester screening for chromosomal abnormalities such as Down's Syndrome, preterm labour, intrauterine growth retardation, pre-eclampsia, and miscarriage/stillbirth. However, the underlying mechanisms behind the correlation of low plasma levels seen in pathologies of pregnancy are not fully understood (Kirkegaard et al., 2010). Considering that the proteolytic function of PAPP-A1 is believed to be inhibited by proMBP in the maternal form, it is still unclear why its plasma level increases dramatically (up to 100x) during pregnancy. The confirmation that PAPP-A1 binds to DC-SIGN, a receptor that is heavily expressed on both placental and uterine macrophages with known anti-inflammatory functions, raises the possibility that it is a component of establishing and maintaining a foeto-maternal immunological tolerance.

It is well-known that PAPP-A1 as well as proMBP are glycoproteins. The group of Claus Oxvig investigated the carbohydrate attachment sites around 15 years ago with the methods available. Such methods were insufficient for determination of the glycan structures at the time; however, they did identify glycan attachment in 11 out of the 13 potential N-linked glycosylation sites (Overgaard et al., 2003). Within this thesis, the occupation of these glycan attachment sites was confirmed and taken further with new insights into the precise nature of the PAPP-A1 and proMBP N-glycosylation patterns.

Shortly after the discovery of PAPP-A1, Bischof and colleagues showed that proteolytically active PAPP-A1 "markedly inhibits the haemolytic activity of complement" (Bischof, 1981). Since the PAPP-A1 gene also encodes for five complement control protein modules, a close association with the complement system is very feasible. However, pursuit of this association appeared to diminish

over the last two decades. It would therefore be of great interest to revive this research using the sensitive technologies now available, especially since the findings here illuminate a potential role for PAPP-A1 in immune regulation.

A holistic insight into the functional integration of PAPP-A1 throughout the human body – enzymatically active and inactive - is still pending. The discovery of the involvement of PAPP-A1 in the most diverse biological contexts has led to the development of exciting hypotheses above.

Importantly, the clearance mechanisms for PAPP-A1 are largely unknown and the failure to clear PAPP-A1 efficiently could accelerate life-threatening conditions such as cancer and cardiovascular disease. It is very tempting to hypothesise that the nature of the interaction with DC-SIGN, which is highly endocytic and expressed on resident M2 macrophages and hepatic sinusoidal endothelium, provides a mechanism for maintaining appropriate PAPP-A1 concentration levels in tissues and blood.

One detail of PAPP-A1 as a ligand for DC-SIGN is that it is a soluble, endogenous, secreted glycoprotein. This is significant, as the full DC-SIGN molecule possesses cytoplasmic motifs that are strongly associated with endocytosis into clathrin-coated pits that are well-suited to the cellular uptake and processing of soluble factors. In contrast, many biological DC-SIGN ligands that have been reported in the literature are membrane-bound molecules, such as viral envelope glycoproteins, human CEACAM1 (Bogoevska et al., 2006) and human CD11b/CD18 (van Gisbergen et al., 2005).

Reports have been published suggesting that a specific, di-sialylated glycoform of human IgG is able to interact with DC-SIGN (Anthony et al., 2008) and invoke the anti-inflammatory effects of therapeutic intravenous immunoglobulin (IVIG). However, this work has been carried out using mouse systems in combination with human IgG1 fragments and there are significant differences between the mouse and human DC-SIGN orthologues (Powlesland et al., 2006). Furthermore, measurement of binding affinities between purified human DC-SIGN and isolated glycoforms of human IgG1 showed that all the high-mannose IgG1 glycoforms bound to DC-SIGN, compared with none of the sialylated IgG1 glycoforms (Yu et al., 2013) . In addition, high field heteronuclear NMR has been used to investigate interactions between ^{13}C-labelled di-sialylated IgG1-Fc and ^{15}N-labelled DC-SIGN CRD without success (personal communication from Prof J Prestegard, Complex Carbohydrate Research Center, University of Georgia, USA). Therefore, this proposed ligand partner for DC-SIGN, including a subsequent biological role, is highly debatable.

However, the clear BLI binding data with encouraging affinity values - together with the determination of its high mannose N-glycans - makes human native PAPP-A1 a much more convincing endogenous, soluble DC-SIGN ligand worthy of further study.

In conclusion, through the combination of these three project themes with cutting-edge analytical and structural techniques, this thesis has made contributions to several new areas within the fields of glycobiology and immunology:

1. First time reporting of the high affinity interaction between the soluble human glycoprotein – the native PAPP-A1/proMBP-complex - and the human C-type lectin DC-SIGN within the physiologically relevant range, suggesting a viable biological connection between these molecules. This has also demonstrated pioneering use of dip-and-read BLI technology to measure human glycoprotein-lectin interactions in real time, with major implications in the future for quantitative evaluation of functional glycomics where sample scarcity has previously been a major challenge.

2. The determination of the glycosylation profile of the native PAPP-A1, which derived from human retroplacental blood. The reported N-glycan profile, illustrating both the specific glycan structures and their sites of attachment, is a novel contribution to the field of pregnancy biology and also provides crucial insights into the basis of the elucidated interaction with DC-SIGN. It is also the first report of comparative N-glycan profiling of recombinant and native human PAPP-A1 and this emphasises the importance of understanding the glycosylation state of glycoproteins used in the research laboratory as well as pharmaceutical industry.

3. The interaction studies with star-shaped glycopolymers and DC-SIGN have pushed the fields of polymer chemistry and advanced materials further into the domain of immunology, whilst convincingly achieving picomolar affinity for a significant human receptor using a relatively low molecular weight, soluble agent.

Future work

The results from this thesis highlight new territories within glycobiology and the immune system, in addition to potential implications in endocrinology and growth factor control. From here, a number of key future projects could be developed:

a) Cellular immunology work

Experiments to test PAPP-A1/proMBP interactions with DC-SIGN in cellular environments to assess the hypothesis further and apply translational approaches:

Strategies would include the investigation of PAPP-A1/proMBP binding to DC-SIGN+ cells using biotinylated hnPAPP-A1/proMBP and functional cell types, such as differentiated macrophages and DCs. Fluorophore-labelled streptavidin probes would facilitate PAPP-A1/proMBP detection via flow cytometry and also investigate phenomena such as glycoprotein uptake. Cytokine secretion and gene expression responses would also be possible to study the impact of PAPP-A1/proMBP on immune cells. Together with further glycopolymer investigations, including the analysis of fucose glycopolymers, it would also be valuable to use PAPP-A1 to probe the mannose and fucose signalosomes activated by cellular engagement of DC-SIGN.

Furthermore, an exciting possibility presents itself with PAPP-A1 being a soluble glycoprotein favoured by DC-SIGN: It is feasible to conjugate human PAPP-A1 to molecules of interest, such as vaccine preparations or siRNA oligonucleotides, in order to facilitate their putative uptake by human DCs without the risk of invoking immune responses or needing synthetic polymer components.

b) Further biophysical analysis

This would include deeper mass spectrometry analysis, including ion mobility MS to define full sugar-linkage of the glycoforms, and also complete the investigation of human proMBP, derived from the native complex.

Furthermore, opportunities for translational work could arise through characterisation of glycosylation patterns for PAPP-A1/proMBP, e.g. in mothers that develop pregnancy complications, such as pre-eclampsia and foetal growth restriction (via collaboration with Dr Manu Vatish, Nuffield Department of Obstetrics & Gynaecology, Oxford), and also in individuals at risk of acute coronary syndromes.

Additional structural work via techniques such as Cryo-Electron Microscopy could provide fundamental data on the nature of how the DC-SIGN/PAPP-A1-proMBP complex assembly is formed in three dimensions. Closely aligned with this would be ongoing collaborations with Prof Russell Wallis (University of Leicester, UK) involving co-crystallisation studies with the extracellular domain of DC-SIGN.

The feasibility of BLI as a technique for resolving lectin-glycoprotein interactions opens up opportunities for extracting crucial data from low abundance samples. In Coventry, we are fortunate to be near the Arden Tissue Bank based at the University Hospital Coventry & Warwickshire. This is a major source of a wide range of human tissue and blood samples collected from a broad range of patient groups. Together with the growing availability of native and recombinant human lectins in quantities amenable to the BLI platform, significant potential opens up to broaden the identification and characterisation of novel glycomic networks. This could incorporate other important human lectins such as the Siglec family and galectins, in addition to glycoproteins recovered from selected tissue and blood samples relating to specific disease states.

c) Exploring the Immune-Endocrine interface

Confirmation of PAPP-A1 as a feasible ligand for DC-SIGN raises the possibility of new links between growth factor endocrinology and the immune system. In addition to DC and macrophage studies above, it would be valuable to investigate whether PAPP-A1 connects fibroblast function in localised tissue growth and homeostasis with immune surveillance using fibroblast-DC/macrophage co-cultures. It is possible that DC-SIGN is involved in PAPP-A1 clearance, and that persistence of

PAPP-A1 in tissues and blood vessels has negative implications in atherosclerotic plaque and tumour development. The proximity of the Arden Tissue Bank and Pathology Services in Coventry also raises the possibility of visualising potential PAPP-A1 and DC-SIGN connections in cardiovascular disease and cancer tissues via immunohistochemistry.

7 REFERENCES

Book

LOTTSPEICH, F. 2006. *Bioanalytik,* Heidelberg, Germany, Spektrum Akademischer Verlag, Springer.

MALLET, A. I. & DOWN, S. 2009. *Dictionary of Mass Spectrometry,* Chichester ; Hoboken, NJ, John Wiley & Sons.

TAYLOR, M. E. & DRICKAMER, K. 2011. *Introduction to Glycobiology,* Oxford; New York, Oxford University Press.

Book Section

BAIRD, J. T. T. & CRAIK, C. S. 2013. Chapter 575 - Trypsin. *In:* RAWLINGS, N. D. & SALVESEN, G. (eds.) *Handbook of Proteolytic Enzymes (Third Edition).* Academic Press.

BROCKHAUSEN, I. & STANLEY, P. 2015. O-GalNAc Glycans. *In:* RD, VARKI, A., CUMMINGS, R. D., ESKO, J. D., STANLEY, P., HART, G. W., AEBI, M., DARVILL, A. G., KINOSHITA, T., PACKER, N. H., PRESTEGARD, J. H., SCHNAAR, R. L. & SEEBERGER, P. H. (eds.) *Essentials of Glycobiology.* Cold Spring Harbor (NY).

CUMMINGS, R. D. & MCEVER, R. P. 2015. C-Type Lectins. *In:* RD, VARKI, A., CUMMINGS, R. D., ESKO, J. D., STANLEY, P., HART, G. W., AEBI, M., DARVILL, A. G., KINOSHITA, T., PACKER, N. H., PRESTEGARD, J. H., SCHNAAR, R. L. & SEEBERGER, P. H. (eds.) *Essentials of Glycobiology.* Cold Spring Harbor (NY).

CUMMINGS, R. D. & SCHNAAR, R. L. 2015. R-Type Lectins. *In:* 3RD, VARKI, A., CUMMINGS, R. D., ESKO, J. D., STANLEY, P., HART, G. W., AEBI, M., DARVILL, A. G., KINOSHITA, T., PACKER, N. H., PRESTEGARD, J. H., SCHNAAR, R. L. & SEEBERGER, P. H. (eds.) *Essentials of Glycobiology.* Cold Spring Harbor (NY).

GRÁF, L., SZILÁGYI, L. & VENEKEI, I. 2013. Chapter 582 - Chymotrypsin. *In:* RAWLINGS, N. D. & SALVESEN, G. (eds.) *Handbook of Proteolytic Enzymes (Third Edition).* Academic Press.

HENRISSAT, B., SUROLIA, A. & STANLEY, P. 2015. A Genomic View of Glycobiology. *In:* RD, VARKI, A., CUMMINGS, R. D., ESKO, J. D., STANLEY, P., HART, G. W., AEBI, M., DARVILL, A. G., KINOSHITA, T., PACKER, N. H., PRESTEGARD, J. H., SCHNAAR, R. L. & SEEBERGER, P. H. (eds.) *Essentials of Glycobiology.* Cold Spring Harbor (NY).

SHARON, N. & LIS, H. 2007. Specificity And Affinity. *Lectins.* 2nd ed. ed. Dordrecht: Springer Netherlands.

TAYLOR, M. E., DRICKAMER, K., SCHNAAR, R. L., ETZLER, M. E. & VARKI, A. 2015. Discovery and Classification of Glycan-Binding Proteins. *In:* RD, VARKI, A., CUMMINGS, R. D., ESKO, J. D., STANLEY, P., HART, G. W., AEBI, M., DARVILL, A. G., KINOSHITA, T., PACKER, N. H., PRESTEGARD,

J. H., SCHNAAR, R. L. & SEEBERGER, P. H. (eds.) *Essentials of Glycobiology.* Cold Spring Harbor (NY).

VAN DAMME, E. J. M., LANNOO, N. & PEUMANS, W. J. 2008. Plant Lectins. *In:* KADER, J. C. & DELSENY, M. (eds.) *Advances in Botanical Research, Vol 48.* London: Academic Press Ltd-Elsevier Science Ltd.

VARKI, A., SCHNAAR, R. L. & CROCKER, P. R. 2015. I-Type Lectins. *In:* RD, VARKI, A., CUMMINGS, R. D., ESKO, J. D., STANLEY, P., HART, G. W., AEBI, M., DARVILL, A. G., KINOSHITA, T., PACKER, N. H., PRESTEGARD, J. H., SCHNAAR, R. L. & SEEBERGER, P. H. (eds.) *Essentials of Glycobiology.* Cold Spring Harbor (NY).

ZACHARA, N., AKIMOTO, Y. & HART, G. W. 2015. The O-GlcNAc Modification. *In:* RD, VARKI, A., CUMMINGS, R. D., ESKO, J. D., STANLEY, P., HART, G. W., AEBI, M., DARVILL, A. G., KINOSHITA, T., PACKER, N. H., PRESTEGARD, J. H., SCHNAAR, R. L. & SEEBERGER, P. H. (eds.) *Essentials of Glycobiology.* Cold Spring Harbor (NY).

Journal Article

ALLEY, W. R., JR., MANN, B. F. & NOVOTNY, M. V. 2013. High-sensitivity analytical approaches for the structural characterization of glycoproteins. *Chem Rev,* 113**,** 2668-732.

ALLEY, W. R., JR. & NOVOTNY, M. V. 2013. Structural glycomic analyses at high sensitivity: a decade of progress. *Annu Rev Anal Chem (Palo Alto Calif),* 6**,** 237-65.

ALPER, J. 2001. Searching for medicine's sweet spot. *Science,* 291**,** 2338-43.

ANTHONY, R. M., WERMELING, F., KARLSSON, M. C. & RAVETCH, J. V. 2008. Identification of a receptor required for the anti-inflammatory activity of IVIG. *Proc Natl Acad Sci U S A,* 105**,** 19571-8.

APWEILER, R., HERMJAKOB, H. & SHARON, N. 1999. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta,* 1473**,** 4-8.

AXEN, R., PORATH, J. & ERNBACK, S. 1967. Chemical coupling of peptides and proteins to polysaccharides by means of cyanogen halides. *Nature,* 214**,** 1302-4.

BASHIROVA, A. A., GEIJTENBEEK, T. B., VAN DUIJNHOVEN, G. C., VAN VLIET, S. J., EILERING, J. B., MARTIN, M. P., WU, L., MARTIN, T. D., VIEBIG, N., KNOLLE, P. A., KEWALRAMANI, V. N., VAN KOOYK, Y. & CARRINGTON, M. 2001. A dendritic cell-specific intercellular adhesion molecule 3-grabbing nonintegrin (DC-SIGN)-related protein is highly expressed on human liver sinusoidal endothelial cells and promotes HIV-1 infection. *J Exp Med,* 193**,** 671-8.

BAYES-GENIS, A., CONOVER, C. A., OVERGAARD, M. T., BAILEY, K. R., CHRISTIANSEN, M., HOLMES, D. R., JR., VIRMANI, R., OXVIG, C. & SCHWARTZ, R. S. 2001. Pregnancy-associated plasma protein A as a marker of acute coronary syndromes. *N Engl J Med,* 345**,** 1022-9.

BECER, C. R., GIBSON, M. I., GENG, J., ILYAS, R., WALLIS, R., MITCHELL, D. A. & HADDLETON, D. M. 2010. High-affinity glycopolymer binding to human DC-SIGN and disruption of DC-SIGN interactions with HIV envelope glycoprotein. *J Am Chem Soc,* 132**,** 15130-2.

BISCHOF, P. 1981. Pregnancy-associated plasma protein-A: an inhibitor of the complement system. *Placenta,* 2**,** 29-34.

BLOW, N. 2009. Glycobiology: A spoonful of sugar. *Nature,* 457**,** 617-20.
BODE, L. 2012. Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology,* 22**,** 1147-62.
BOGOEVSKA, V., HORST, A., KLAMPE, B., LUCKA, L., WAGENER, C. & NOLLAU, P. 2006. CEACAM1, an adhesion molecule of human granulocytes, is fucosylated by fucosyltransferase IX and interacts with DC-SIGN of dendritic cells via Lewis x residues. *Glycobiology,* 16**,** 197-209.
BOILY-LAROUCHE, G., ISCACHE, A. L., ZIJENAH, L. S., HUMPHREY, J. H., MOULAND, A. J., WARD, B. J. & ROGER, M. 2009. Functional genetic variants in DC-SIGNR are associated with mother-to-child transmission of HIV-1. *PLoS One,* 4**,** e7211.
BOURNE, Y. & HENRISSAT, B. 2001. Glycoside hydrolases and glycosyltransferases: families and functional modules. *Curr Opin Struct Biol,* 11**,** 593-600.
BOYD, W. C. 1954. CHAPTER 22: The Proteins of Immune Reactions. *Chemistry, Biological Activity, and Methods, Part B***,** 755-844.
BOYD, W. C. & SHAPLEIGH, E. 1954. Specific Precipitating Activity of Plant Agglutinins (Lectins). *Science,* 119**,** 419.
BROWN, G. D., WILLMENT, J. A. & WHITEHEAD, L. 2018. C-type lectins in immunity and homeostasis. *Nat Rev Immunol,* 18**,** 374-389.
BURLINGAME, A. L. 1996. Characterization of protein glycosylation by mass spectrometry. *Curr Opin Biotechnol,* 7**,** 4-10.
BUTKINAREE, C., PARK, K. & HART, G. W. 2010. O-linked beta-N-acetylglucosamine (O-GlcNAc): Extensive crosstalk with phosphorylation to regulate signaling and transcription in response to nutrients and stress. *Biochim Biophys Acta,* 1800**,** 96-106.
CARRAWAY, K. L. & HULL, S. R. 1991. Cell surface mucin-type glycoproteins and mucin-like domains. *Glycobiology,* 1**,** 131-8.
CHAIPAN, C., SOILLEUX, E. J., SIMPSON, P., HOFMANN, H., GRAMBERG, T., MARZI, A., GEIER, M., STEWART, E. A., EISEMANN, J., STEINKASSERER, A., SUZUKI-INOUE, K., FULLER, G. L., PEARCE, A. C., WATSON, S. P., HOXIE, J. A., BARIBAUD, F. & POEHLMANN, S. 2006. DC-SIGN and CLEC-2 mediate human immunodeficiency virus type 1 capture by platelets. *Journal of Virology,* 80**,** 8951-8960.
CHAIT, B. T. 2006. Chemistry. Mass spectrometry: bottom-up or top-down? *Science,* 314**,** 65-6.
CHAN, V. S., CHAN, K. Y., CHEN, Y., POON, L. L., CHEUNG, A. N., ZHENG, B., CHAN, K. H., MAK, W., NGAN, H. Y., XU, X., SCREATON, G., TAM, P. K., AUSTYN, J. M., CHAN, L. C., YIP, S. P., PEIRIS, M., KHOO, U. S. & LIN, C. L. 2006. Homozygous L-SIGN (CLEC4M) plays a protective role in SARS coronavirus infection. *Nat Genet,* 38**,** 38-46.
CHRISTIANSEN, M., JALIASHVILI, I., OVERGAARD, M. T., ENSINGER, C., OBRIST, P. & OXVIG, C. 2000. Quantification and characterization of pregnancy-associated complexes of angiotensinogen and the proform of eosinophil major basic protein in serum and amniotic fluid. *Clin Chem,* 46**,** 1099-105.
CONOVER, C. A. & OXVIG, C. 2017. PAPP-A: a promising therapeutic target for healthy longevity. *Aging Cell,* 16**,** 205-209.
COSTELLO, C. E., CONTADO-MILLER, J. M. & CIPOLLO, J. F. 2007. A glycomics platform for the analysis of permethylated oligosaccharide alditols. *J Am Soc Mass Spectrom,* 18**,** 1799-812.

CROCKER, P. R., PAULSON, J. C. & VARKI, A. 2007. Siglecs and their roles in the immune system. *Nat Rev Immunol,* 7**,** 255-66.

CUMMINGS, R. D. 2009. The repertoire of glycan determinants in the human glycome. *Mol Biosyst,* 5**,** 1087-104.

CUMMINGS, R. D. & PIERCE, J. M. 2014. The challenge and promise of glycomics. *Chem Biol,* 21**,** 1-15.

CURTIS, B. M., SCHARNOWSKE, S. & WATSON, A. J. 1992. Sequence and expression of a membrane-associated C-type lectin that exhibits CD4-independent binding of human immunodeficiency virus envelope glycoprotein gp120. *Proc Natl Acad Sci U S A,* 89**,** 8356-60.

DAHMS, N. M. & HANCOCK, M. K. 2002. P-type lectins. *Biochimica Et Biophysica Acta-General Subjects,* 1572**,** 317-340.

DEJNIRATTISAI, W., WEBB, A. I., CHAN, V., JUMNAINSONG, A., DAVIDSON, A., MONGKOLSAPAYA, J. & SCREATON, G. 2011. Lectin switching during dengue virus infection. *J Infect Dis,* 203**,** 1775-83.

DOUCEY, M. A., HESS, D., BLOMMERS, M. J. & HOFSTEENGE, J. 1999. Recombinant human interleukin-12 is the second example of a C-mannosylated protein. *Glycobiology,* 9**,** 435-41.

DOUCEY, M. A., HESS, D., CACAN, R. & HOFSTEENGE, J. 1998. Protein C-mannosylation is enzyme-catalysed and uses dolichyl-phosphate-mannose as a precursor. *Mol Biol Cell,* 9**,** 291-300.

DRICKAMER, K. 1992. Engineering galactose-binding activity into a C-type mannose-binding protein. *Nature,* 360**,** 183-6.

DRICKAMER, K. & TAYLOR, M. E. 2015. Recent insights into structures and functions of C-type lectins in the immune system. *Curr Opin Struct Biol,* 34**,** 26-34.

DUBE, D. H. & BERTOZZI, C. R. 2005. Glycans in cancer and inflammation--potential for therapeutics and diagnostics. *Nat Rev Drug Discov,* 4**,** 477-88.

DUTTA, D., MANDAL, C. & MANDAL, C. 2017. Unusual glycosylation of proteins: Beyond the universal sequon and other amino acids. *Biochim Biophys Acta Gen Subj,* 1861**,** 3096-3108.

EISENBERG, S. P., EVANS, R. J., AREND, W. P., VERDERBER, E., BREWER, M. T., HANNUM, C. H. & THOMPSON, R. C. 1990. Primary structure and functional expression from complementary DNA of a human interleukin-1 receptor antagonist. *Nature,* 343**,** 341-6.

ENGERING, A., GEIJTENBEEK, T. B. & VAN KOOYK, Y. 2002a. Immune escape through C-type lectins on dendritic cells. *Trends Immunol,* 23**,** 480-5.

ENGERING, A., GEIJTENBEEK, T. B., VAN VLIET, S. J., WIJERS, M., VAN LIEMPT, E., DEMAUREX, N., LANZAVECCHIA, A., FRANSEN, J., FIGDOR, C. G., PIGUET, V. & VAN KOOYK, Y. 2002b. The dendritic cell-specific adhesion receptor DC-SIGN internalizes antigen for presentation to T cells. *J Immunol,* 168**,** 2118-26.

ENGERING, A., VAN VLIET, S. J., GEIJTENBEEK, T. B. & VAN KOOYK, Y. 2002c. Subset of DC-SIGN(+) dendritic cells in human blood transmits HIV-1 to T lymphocytes. *Blood,* 100**,** 1780-6.

FEINBERG, H., CASTELLI, R., DRICKAMER, K., SEEBERGER, P. H. & WEIS, W. I. 2007. Multiple modes of binding enhance the affinity of DC-SIGN for high mannose N-linked glycans found on viral glycoproteins. *J Biol Chem,* 282**,** 4202-9.

FEINBERG, H., GUO, Y., MITCHELL, D. A., DRICKAMER, K. & WEIS, W. I. 2005. Extended neck regions stabilize tetramers of the receptors DC-SIGN and DC-SIGNR. *J Biol Chem,* 280**,** 1327-35.

FEINBERG, H., MITCHELL, D. A., DRICKAMER, K. & WEIS, W. I. 2001. Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR. *Science,* 294**,** 2163-6.

FEINBERG, H., TSO, C. K., TAYLOR, M. E., DRICKAMER, K. & WEIS, W. I. 2009. Segmented helical structure of the neck region of the glycan-binding receptor DC-SIGNR. *J Mol Biol,* 394**,** 613-20.

FIETE, D. J., BERANEK, M. C. & BAENZIGER, J. U. 1998. A cysteine-rich domain of the "mannose" receptor mediates GalNAc-4-SO4 binding. *Proc Natl Acad Sci U S A,* 95**,** 2089-93.

FIGDOR, C. G., VAN KOOYK, Y. & ADEMA, G. J. 2002. C-type lectin receptors on dendritic cells and Langerhans cells. *Nat Rev Immunol,* 2**,** 77-84.

FURMANEK, A. & HOFSTEENGE, J. 2000. Protein C-mannosylation: facts and questions. *Acta Biochim Pol,* 47**,** 781-9.

GABIUS, H. J. 1997. Animal lectins. *European Journal of Biochemistry,* 243**,** 543-576.

GALL, S. A. & HALBERT, S. P. 1972. Antigenic constituents in pregnancy plasma which are undetectable in normal non-pregnant female or male plasma. *Int Arch Allergy Appl Immunol,* 42**,** 503-15.

GEIJTENBEEK, T. B. & GRINGHUIS, S. I. 2009. Signalling through C-type lectin receptors: shaping immune responses. *Nature Reviews Immunology,* 9**,** 465-479.

GEIJTENBEEK, T. B. & GRINGHUIS, S. I. 2016. C-type lectin receptors in the control of T helper cell differentiation. *Nat Rev Immunol,* 16**,** 433-48.

GEIJTENBEEK, T. B., KROOSHOOP, D. J., BLEIJS, D. A., VAN VLIET, S. J., VAN DUIJNHOVEN, G. C., GRABOVSKY, V., ALON, R., FIGDOR, C. G. & VAN KOOYK, Y. 2000a. DC-SIGN-ICAM-2 interaction mediates dendritic cell trafficking. *Nat Immunol,* 1**,** 353-7.

GEIJTENBEEK, T. B., KWON, D. S., TORENSMA, R., VAN VLIET, S. J., VAN DUIJNHOVEN, G. C., MIDDEL, J., CORNELISSEN, I. L., NOTTET, H. S., KEWALRAMANI, V. N., LITTMAN, D. R., FIGDOR, C. G. & VAN KOOYK, Y. 2000b. DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell,* 100**,** 587-97.

GEIJTENBEEK, T. B., TORENSMA, R., VAN VLIET, S. J., VAN DUIJNHOVEN, G. C. F., ADEMA, G. J., VAN KOOYK, Y. & FIGDOR, C. G. 2000c. Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses. *Cell,* 100**,** 575-585.

GEIJTENBEEK, T. B., VAN VLIET, S. J., ENGERING, A., T HART, B. A. & VAN KOOYK, Y. 2004. Self- and nonself-recognition by C-type lectins on dendritic cells. *Annu Rev Immunol,* 22**,** 33-54.

GEURTSEN, J., CHEDAMMI, S., MESTERS, J., COT, M., DRIESSEN, N. N., SAMBOU, T., KAKUTANI, R., UMMELS, R., MAASKANT, J., TAKATA, H., BABA, O., TERASHIMA, T., BOVIN, N., VANDENBROUCKE-GRAULS, C. M., NIGOU, J., PUZO, G., LEMASSU, A., DAFFE, M. & APPELMELK, B. J. 2009. Identification of mycobacterial alpha-glucan as a novel ligand for DC-SIGN: involvement of mycobacterial capsular polysaccharides in host immune modulation. *J Immunol,* 183**,** 5221-31.

GHOSH, P., DAHMS, N. M. & KORNFELD, S. 2003. Mannose 6-phosphate receptors: new twists in the tale. *Nat Rev Mol Cell Biol,* 4**,** 202-12.

GILAR, M., YU, Y. Q., AHN, J., XIE, H., HAN, H., YING, W. & QIAN, X. 2011. Characterization of glycoprotein digests with hydrophilic interaction chromatography and mass spectrometry. *Anal Biochem,* 417**,** 80-8.

GLERUP, S., KLOVERPRIS, S. & OXVIG, C. 2006. The proform of the eosinophil major basic protein binds the cell surface through a site distinct from its C-type lectin ligand-binding region. *J Biol Chem,* 281**,** 31509-16.

GOLDSTEIN, I. J., HUGHES, R. C., MONSIGNY, M., OSAWA, T. & SHARON, N. 1980. What should be called a lectin? *Nature,* 285**,** 66.

GRINGHUIS, S. I., DEN DUNNEN, J., LITJENS, M., VAN DER VLIST, M. & GEIJTENBEEK, T. B. 2009. Carbohydrate-specific signaling through the DC-SIGN signalosome tailors immunity to Mycobacterium tuberculosis, HIV-1 and Helicobacter pylori. *Nat Immunol,* 10**,** 1081-8.

GRINGHUIS, S. I., DEN DUNNEN, J., LITJENS, M., VAN HET HOF, B., VAN KOOYK, Y. & GEIJTENBEEK, T. B. 2007. C-type lectin DC-SIGN modulates Toll-like receptor signaling via Raf-1 kinase-dependent acetylation of transcription factor NF-kappaB. *Immunity,* 26**,** 605-16.

GRINGHUIS, S. I., KAPTEIN, T. M., WEVERS, B. A., MESMAN, A. W. & GEIJTENBEEK, T. B. 2014. Fucose-specific DC-SIGN signalling directs T helper cell type-2 responses via IKKepsilon- and CYLD-dependent Bcl3 activation. *Nat Commun,* 5**,** 3898.

GUO, Y., BAO, Y., GUO, D. & YANG, W. 2018. Pregnancy-associated plasma protein a in cancer: expression, oncogenic functions and regulation. *Am J Cancer Res,* 8**,** 955-963.

GUO, Y., FEINBERG, H., CONROY, E., MITCHELL, D. A., ALVAREZ, R., BLIXT, O., TAYLOR, M. E., WEIS, W. I. & DRICKAMER, K. 2004. Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. *Nat Struct Mol Biol,* 11**,** 591-8.

GURURAJAN, P., GURUMURTHY, P., NAYAR, P., RAO, G. S., BABU, R. S., SARASABHARATI, A. & CHERIAN, K. M. 2012. Pregnancy associated plasma protein-A (PAPP-A) as an early marker for the diagnosis of acute coronary syndrome. *Indian Heart J,* 64**,** 141-5.

HART, G. W., SLAWSON, C., RAMIREZ-CORREA, G. & LAGERLOF, O. 2011. Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. *Annu Rev Biochem,* 80**,** 825-58.

HARTMANN, S. & HOFSTEENGE, J. 2000. Properdin, the positive regulator of complement, is highly C-mannosylated. *J Biol Chem,* 275**,** 28569-74.

HARVEY, D. J. 2011. Derivatization of carbohydrates for analysis by chromatography; electrophoresis and mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci,* 879**,** 1196-225.

HELENIUS, A. & AEBI, M. 2004. Roles of N-linked glycans in the endoplasmic reticulum. *Annu Rev Biochem,* 73**,** 1019-49.

HODGES, A., SHARROCKS, K., EDELMANN, M., BABAN, D., MORIS, A., SCHWARTZ, O., DRAKESMITH, H., DAVIES, K., KESSLER, B., MCMICHAEL, A. & SIMMONS, A. 2007. Activation of the lectin DC-SIGN induces an immature dendritic cell phenotype triggering Rho-GTPase activity required for HIV-1 replication. *Nat Immunol,* 8**,** 569-77.

HOFFMANN, M., MARX, K., REICHL, U., WUHRER, M. & RAPP, E. 2016. Site-specific O-Glycosylation Analysis of Human Blood Plasma Proteins. *Mol Cell Proteomics,* 15**,** 624-41.

HOFSTEENGE, J., BLOMMERS, M., HESS, D., FURMANEK, A. & MIROSHNICHENKO, O. 1999. The four terminal components of the

complement system are C-mannosylated on multiple tryptophan residues. *J Biol Chem,* 274**,** 32786-94.
HOFSTEENGE, J., HUWILER, K. G., MACEK, B., HESS, D., LAWLER, J., MOSHER, D. F. & PETER-KATALINIC, J. 2001. C-mannosylation and O-fucosylation of the thrombospondin type 1 module. *J Biol Chem,* 276**,** 6485-98.
HOFSTEENGE, J., MULLER, D. R., DE BEER, T., LOFFLER, A., RICHTER, W. J. & VLIEGENTHART, J. F. 1994. New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase Us. *Biochemistry,* 33**,** 13524-30.
HOLLINGSWORTH, M. A. & SWANSON, B. J. 2004. Mucins in cancer: protection and control of the cell surface. *Nat Rev Cancer,* 4**,** 45-60.
HUBBARD, S. C. & IVATT, R. J. 1981. Synthesis and processing of asparagine-linked oligosaccharides. *Annu Rev Biochem,* 50**,** 555-83.
IHARA, Y., INAI, Y. & IKEZAKI, M. 2011. Protein C-Mannosylation and Its Prospective Functions in the Cell. *Trends in Glycoscience and Glycotechnology,* 23**,** 1-13.
JAKEL, A., QASEEM, A. S., KISHORE, U. & SIM, R. B. 2013. Ligands and receptors of lung surfactant proteins SP-A and SP-D. *Front Biosci (Landmark Ed),* 18**,** 1129-40.
JOHNSON, T. R., MCLELLAN, J. S. & GRAHAM, B. S. 2012. Respiratory syncytial virus glycoprotein G interacts with DC-SIGN and L-SIGN to activate ERK1 and ERK2. *J Virol,* 86**,** 1339-47.
JORDENS, R., THOMPSON, A., AMONS, R. & KONING, F. 1999. Human dendritic cells shed a functional, soluble form of the mannose receptor. *Int Immunol,* 11**,** 1775-80.
JULENIUS, K., MOLGAARD, A., GUPTA, R. & BRUNAK, S. 2005. Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiology,* 15**,** 153-64.
KAMIYA, Y., KAMIYA, D., YAMAMOTO, K., NYFELER, B., HAURI, H. P. & KATO, K. 2008. Molecular basis of sugar recognition by the human L-type lectins ERGIC-53, VIPL, and VIP36. *J Biol Chem,* 283**,** 1857-61.
KAMMERER, U., EGGERT, A. O., KAPP, M., MCLELLAN, A. D., GEIJTENBEEK, T. B., DIETL, J., VAN KOOYK, Y. & KAMPGEN, E. 2003. Unique appearance of proliferating antigen-presenting cells expressing DC-SIGN (CD209) in the decidua of early human pregnancy. *Am J Pathol,* 162**,** 887-96.
KIRKEGAARD, I., ULDBJERG, N. & OXVIG, C. 2010. Biology of pregnancy-associated plasma protein-A in relation to prenatal diagnostics: an overview. *Acta Obstet Gynecol Scand,* 89**,** 1118-25.
KORNFELD, R. & KORNFELD, S. 1985. Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem,* 54**,** 631-64.
KRUTCHINSKY, A. N. & CHAIT, B. T. 2002. On the nature of the chemical noise in MALDI mass spectra. *J Am Soc Mass Spectrom,* 13**,** 129-34.
LAI, W. K., SUN, P. J., ZHANG, J., JENNINGS, A., LALOR, P. F., HUBSCHER, S., MCKEATING, J. A. & ADAMS, D. H. 2006. Expression of DC-SIGN and DC-SIGNR on human sinusoidal endothelium: a role for capturing hepatitis C virus particles. *Am J Pathol,* 169**,** 200-8.
LAIRSON, L. L., HENRISSAT, B., DAVIES, G. J. & WITHERS, S. G. 2008. Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem,* 77**,** 521-55.

LAITINEN, O. H., HYTONEN, V. P., NORDLUND, H. R. & KULOMAA, M. S. 2006. Genetically engineered avidins and streptavidins. *Cell Mol Life Sci,* 63**,** 2992-3017.

LAWRENCE, J. B., OXVIG, C., OVERGAARD, M. T., SOTTRUP-JENSEN, L., GLEICH, G. J., HAYS, L. G., YATES, J. R., 3RD & CONOVER, C. A. 1999. The insulin-like growth factor (IGF)-dependent IGF binding protein-4 protease secreted by human fibroblasts is pregnancy-associated plasma protein-A. *Proc Natl Acad Sci U S A,* 96**,** 3149-53.

LEE, R. T., ICHIKAWA, Y., KAWASAKI, T., DRICKAMER, K. & LEE, Y. C. 1992. Multivalent ligand binding by serum mannose-binding protein. *Arch Biochem Biophys,* 299**,** 129-36.

LEE, R. T. & LEE, Y. C. 2000. Affinity enhancement by multivalent lectin-carbohydrate interaction. *Glycoconj J,* 17**,** 543-51.

LETEUX, C., CHAI, W., LOVELESS, R. W., YUEN, C. T., UHLIN-HANSEN, L., COMBARNOUS, Y., JANKOVIC, M., MARIC, S. C., MISULOVIN, Z., NUSSENZWEIG, M. C. & FEIZI, T. 2000. The cysteine-rich domain of the macrophage mannose receptor is a multispecific lectin that recognizes chondroitin sulfates A and B and sulfated oligosaccharides of blood group Lewis(a) and Lewis(x) types in addition to the sulfated N-glycans of lutropin. *J Exp Med,* 191**,** 1117-26.

LIN, T. M., GALBERT, S. P., KIEFER, D., SPELLACY, W. N. & GALL, S. 1974. Characterization of four human pregnancy-associated plasma proteins. *Am J Obstet Gynecol,* 118**,** 223-36.

LIS, H. & SHARON, N. 1991. Lectin-carbohydrate interactions. *Current Opinion in Structural Biology,* 1**,** 741-749.

LOFFLER, A., DOUCEY, M. A., JANSSON, A. M., MULLER, D. R., DE BEER, T., HESS, D., MELDAL, M., RICHTER, W. J., VLIEGENTHART, J. F. & HOFSTEENGE, J. 1996. Spectroscopic and protein chemical analyses demonstrate the presence of C-mannosylated tryptophan in intact human RNase 2 and its isoforms. *Biochemistry,* 35**,** 12005-14.

LONDRIGAN, S. L., TURVILLE, S. G., TATE, M. D., DENG, Y. M., BROOKS, A. G. & READING, P. C. 2011. N-linked glycosylation facilitates sialic acid-independent attachment and entry of influenza A viruses into cells expressing DC-SIGN or L-SIGN. *J Virol,* 85**,** 2990-3000.

MAEDA, Y. & KINOSHITA, T. 2008. Dolichol-phosphate mannose synthase: structure, function and regulation. *Biochim Biophys Acta,* 1780**,** 861-8.

MALOVIC, I., SORENSEN, K. K., ELVEVOLD, K. H., NEDREDAL, G. I., PAULSEN, S., EROFEEV, A. V., SMEDSROD, B. H. & MCCOURT, P. A. 2007. The mannose receptor on murine liver sinusoidal endothelial cells is the main denatured collagen clearance receptor. *Hepatology,* 45**,** 1454-61.

MARCELO, F., GARCIA-MARTIN, F., MATSUSHITA, T., SARDINHA, J., COELHO, H., OUDE-VRIELINK, A., KOLLER, C., ANDRE, S., CABRITA, E. J., GABIUS, H. J., NISHIMURA, S., JIMENEZ-BARBERO, J. & CANADA, F. J. 2014. Delineating binding modes of Gal/GalNAc and structural elements of the molecular recognition of tumor-associated mucin glycopeptides by the human macrophage galactose-type lectin. *Chemistry,* 20**,** 16147-55.

MARSHALL, R. D. 1972. Glycoproteins. *Annu Rev Biochem,* 41**,** 673-702.

MARTINEZ-POMARES, L. 2012. The mannose receptor. *J Leukoc Biol,* 92**,** 1177-86.

MARTINEZ-POMARES, L., WIENKE, D., STILLION, R., MCKENZIE, E. J., ARNOLD, J. N., HARRIS, J., MCGREAL, E., SIM, R. B., ISACKE, C. M. &

GORDON, S. 2006. Carbohydrate-independent recognition of collagens by the macrophage mannose receptor. *Eur J Immunol,* 36**,** 1074-82.
MCEVER, R. P. & CUMMINGS, R. D. 1997. Perspectives series: cell adhesion in vascular biology. Role of PSGL-1 binding to selectins in leukocyte recruitment. *J Clin Invest,* 100**,** 485-91.
MECHREF, Y. 2012. Use of CID/ETD mass spectrometry to analyze glycopeptides. *Curr Protoc Protein Sci,* Chapter 12**,** Unit 12 11 1-11.
MELLMAN, I. 1996. Endocytosis and molecular sorting. *Annu Rev Cell Dev Biol,* 12**,** 575-625.
MITCHELL, D. A., FADDEN, A. J. & DRICKAMER, K. 2001. A novel mechanism of carbohydrate recognition by the C-type lectins DC-SIGN and DC-SIGNR. Subunit organization and binding to multivalent ligands. *J Biol Chem,* 276**,** 28939-45.
MITCHELL, D. A., ZHANG, Q., VOORHAAR, L., HADDLETON, D. M., HERATH, S., GLEINICH, A. S., RANDEVA, H. S., CRISPIN, M., LEHNERT, H., WALLIS, R., PATTERSON, S. & BECER, C. R. 2017. Manipulation of cytokine secretion in human dendritic cells using glycopolymers with picomolar affinity for DC-SIGN. *Chemical Science,* 8**,** 6974-6980.
MIZUOCHI, T., MATTHEWS, T. J., KATO, M., HAMAKO, J., TITANI, K., SOLOMON, J. & FEIZI, T. 1990. Diversity of oligosaccharide structures on the envelope glycoprotein gp 120 of human immunodeficiency virus 1 from the lymphoblastoid cell line H9. Presence of complex-type oligosaccharides with bisecting N-acetylglucosamine residues. *J Biol Chem,* 265**,** 8519-24.
MOREMEN, K. W., TIEMEYER, M. & NAIRN, A. V. 2012. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol,* 13**,** 448-62.
NAPPER, C. E., DRICKAMER, K. & TAYLOR, M. E. 2006. Collagen binding by the mannose receptor mediated through the fibronectin type II domain. *Biochem J,* 395**,** 579-86.
NEUMANN, K., CASTINEIRAS-VILARINO, M., HOCKENDORF, U., HANNESSCHLAGER, N., LEMEER, S., KUPKA, D., MEYERMANN, S., LECH, M., ANDERS, H. J., KUSTER, B., BUSCH, D. H., GEWIES, A., NAUMANN, R., GROSS, O. & RULAND, J. 2014. Clec12a is an inhibitory receptor for uric acid crystals that regulates inflammation in response to cell death. *Immunity,* 40**,** 389-99.
NISHIKAWA, T., KAJII, S., SATO, C., YASUKAWA, Z., KITAJIMA, K. & ISOBE, M. 2004. Alpha-C-mannosyltryptophan is not recognized by conventional mannose-binding lectins. *Bioorg Med Chem,* 12**,** 2343-8.
NOLL, A. J., YU, Y., LASANAJAK, Y., DUSKA-MCEWEN, G., BUCK, R. H., SMITH, D. F. & CUMMINGS, R. D. 2016. Human DC-SIGN binds specific human milk glycans. *Biochem J,* 473**,** 1343-53.
OFEK, I., MIRELMAN, D. & SHARON, N. 1977. Adherence of Escherichia coli to human mucosal cells mediated by mannose receptors. *Nature,* 265**,** 623-5.
OVERGAARD, M. T., HAANING, J., BOLDT, H. B., OLSEN, I. M., LAURSEN, L. S., CHRISTIANSEN, M., GLEICH, G. J., SOTTRUP-JENSEN, L., CONOVER, C. A. & OXVIG, C. 2000. Expression of recombinant human pregnancy-associated plasma protein-A and identification of the proform of eosinophil major basic protein as its physiological inhibitor. *J Biol Chem,* 275**,** 31128-33.
OVERGAARD, M. T., SORENSEN, E. S., STACHOWIAK, D., BOLDT, H. B., KRISTENSEN, L., SOTTRUP-JENSEN, L. & OXVIG, C. 2003. Complex of pregnancy-associated plasma protein-A and the proform of eosinophil major

basic protein. Disulfide structure and carbohydrate attachment. *J Biol Chem,* 278**,** 2106-17.

OXVIG, C., HAANING, J., HOJRUP, P. & SOTTRUP-JENSEN, L. 1994. Location and nature of carbohydrate groups in proform of human major basic protein isolated from pregnancy serum. *Biochem Mol Biol Int,* 33**,** 329-36.

OXVIG, C., SAND, O., KRISTENSEN, T., GLEICH, G. J. & SOTTRUP-JENSEN, L. 1993. Circulating human pregnancy-associated plasma protein-A is disulfide-bridged to the proform of eosinophil major basic protein. *J Biol Chem,* 268**,** 12243-6.

PEDERSON, K., MITCHELL, D. A. & PRESTEGARD, J. H. 2014. Structural Characterization of the DC-SIGN-Lewis(X) Complex. *Biochemistry,* 53**,** 5700-5709.

PEREZ-VILAR, J., RANDELL, S. H. & BOUCHER, R. C. 2004. C-Mannosylation of MUC5AC and MUC5B Cys subdomains. *Glycobiology,* 14**,** 325-37.

PERONA, J. J. & CRAIK, C. S. 1995. Structural basis of substrate specificity in the serine proteases. *Protein Sci,* 4**,** 337-60.

PETRESCU, A. J., MILAC, A. L., PETRESCU, S. M., DWEK, R. A. & WORMALD, M. R. 2004. Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology,* 14**,** 103-14.

POHLMANN, S., SOILLEUX, E. J., BARIBAUD, F., LESLIE, G. J., MORRIS, L. S., TROWSDALE, J., LEE, B., COLEMAN, N. & DOMS, R. W. 2001. DC-SIGNR, a DC-SIGN homologue expressed in endothelial cells, binds to human and simian immunodeficiency viruses and activates infection in trans. *Proc Natl Acad Sci U S A,* 98**,** 2670-5.

POWELL, L. D. & VARKI, A. 1995. I-type lectins. *J Biol Chem,* 270**,** 14243-6.

POWLESLAND, A. S., WARD, E. M., SADHU, S. K., GUO, Y., TAYLOR, M. E. & DRICKAMER, K. 2006. Widely divergent biochemical properties of the complete set of mouse DC-SIGN-related proteins. *J Biol Chem,* 281**,** 20440-9.

PROBERT, F., MITCHELL, D. A. & DIXON, A. M. 2014. NMR evidence for oligosaccharide release from the dendritic-cell specific intercellular adhesion molecule 3-grabbing non-integrin-related (CLEC4M) carbohydrate recognition domain at low pH. *Febs Journal,* 281**,** 3739-3750.

PROBERT, F., WHITTAKER, S. B., CRISPIN, M., MITCHELL, D. A. & DIXON, A. M. 2013. Solution NMR analyses of the C-type carbohydrate recognition domain of DC-SIGNR protein reveal different binding modes for HIV-derived oligosaccharides and smaller glycan fragments. *J Biol Chem,* 288**,** 22745-57.

PUIG-KROGER, A., SERRANO-GOMEZ, D., CAPARROS, E., DOMINGUEZ-SOTO, A., RELLOSO, M., COLMENARES, M., MARTINEZ-MUNOZ, L., LONGO, N., SANCHEZ-SANCHEZ, N., RINCON, M., RIVAS, L., SANCHEZ-MATEOS, P., FERNANDEZ-RUIZ, E. & CORBI, A. L. 2004. Regulated expression of the pathogen receptor dendritic cell-specific intercellular adhesion molecule 3 (ICAM-3)-grabbing nonintegrin in THP-1 human leukemic cells, monocytes, and macrophages. *J Biol Chem,* 279**,** 25680-8.

RAMAN, R., VENKATARAMAN, M., RAMAKRISHNAN, S., LANG, W., RAGURAM, S. & SASISEKHARAN, R. 2006. Advancing glycomics: implementation strategies at the consortium for functional glycomics. *Glycobiology,* 16**,** 82R-90R.

ROSE, M. C. 1992. Mucins: structure, function, and role in pulmonary diseases. *Am J Physiol,* 263**,** L413-29.

ROSEMAN, S. 2001. Reflections on glycobiology. *J Biol Chem,* 276**,** 41527-42.

RUTENBER, E., READY, M. & ROBERTUS, J. D. 1987. Structure and evolution of ricin B chain. *Nature,* 326**,** 624-6.

RYDZ, N., SWYSTUN, L. L., NOTLEY, C., PATERSON, A. D., RICHES, J. J., SPONAGLE, K., BOONYAWAT, B., MONTGOMERY, R. R., JAMES, P. D. & LILLICRAP, D. 2013. The C-type lectin receptor CLEC4M binds, internalizes, and clears von Willebrand factor and contributes to the variation in plasma von Willebrand factor levels. *Blood,* 121**,** 5228-5237.

SALLUSTO, F., CELLA, M., DANIELI, C. & LANZAVECCHIA, A. 1995. Dendritic cells use macropinocytosis and the mannose receptor to concentrate macromolecules in the major histocompatibility complex class II compartment: downregulation by cytokines and bacterial products. *J Exp Med,* 182**,** 389-400.

SHAJAHAN, A., HEISS, C., ISHIHARA, M. & AZADI, P. 2017a. Glycomic and glycoproteomic analysis of glycoproteins-a tutorial. *Anal Bioanal Chem,* 409**,** 4483-4505.

SHAJAHAN, A., SUPEKAR, N. T., HEISS, C., ISHIHARA, M. & AZADI, P. 2017b. Tool for Rapid Analysis of Glycopeptide by Permethylation via One-Pot Site Mapping and Glycan Analysis. *Anal Chem,* 89**,** 10734-10743.

SHARON, N. & LIS, H. 1989. Lectins as cell recognition molecules. *Science,* 246**,** 227-34.

SHIKATA, Y., HAYASHI, Y., YOSHIMATSU, K., OHYA, Y., SETO, T., FUKUSHIMA, K. & YOSHIDA, Y. 1993. Pro-major basic protein has three types of sugar chains at the pro-portion. *Biochim Biophys Acta,* 1163**,** 243-9.

SINGH, R. S., BHARI, R. & KAUR, H. P. 2011. Current trends of lectins from microfungi. *Crit Rev Biotechnol,* 31**,** 193-210.

SMITH, R., BISCHOF, P., HUGHES, G. & KLOPPER, A. 1979. Studies on pregnancy-associated plasma protein A in the third trimester of pregnancy. *Br J Obstet Gynaecol,* 86**,** 882-7.

SOILLEUX, E. J., BARTEN, R. & TROWSDALE, J. 2000. Cutting edge: DC-SIGN; a related gene, DC-SIGNR; and CD23 form a cluster on 19p13. *Journal of Immunology,* 165**,** 2937-2942.

SOILLEUX, E. J., MORRIS, L. S., LESLIE, G., CHEHIMI, J., LUO, Q., LEVRONEY, E., TROWSDALE, J., MONTANER, L. J., DOMS, R. W., WEISSMAN, D., COLEMAN, N. & LEE, B. 2002. Constitutive and induced expression of DC-SIGN on dendritic cell and macrophage subpopulations in situ and in vitro. *J Leukoc Biol,* 71**,** 445-57.

SPIRO, R. G. 1970. Glycoproteins. *Annu Rev Biochem,* 39**,** 599-638.

SPIRO, R. G. 2002. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology,* 12**,** 43R-56R.

STEENTOFT, C., VAKHRUSHEV, S. Y., JOSHI, H. J., KONG, Y., VESTER-CHRISTENSEN, M. B., SCHJOLDAGER, K. T., LAVRSEN, K., DABELSTEEN, S., PEDERSEN, N. B., MARCOS-SILVA, L., GUPTA, R., BENNETT, E. P., MANDEL, U., BRUNAK, S., WANDALL, H. H., LEVERY, S. B. & CLAUSEN, H. 2013. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J,* 32**,** 1478-88.

STEPPER, J., SHASTRI, S., LOO, T. S., PRESTON, J. C., NOVAK, P., MAN, P., MOORE, C. H., HAVLICEK, V., PATCHETT, M. L. & NORRIS, G. E. 2011. Cysteine S-glycosylation, a new post-translational modification found in glycopeptide bacteriocins. *FEBS Lett,* 585**,** 645-50.

SVAJGER, U., ANDERLUH, M., JERAS, M. & OBERMAJER, N. 2010. C-type lectin DC-SIGN: An adhesion, signalling and antigen-uptake molecule that guides dendritic cells in immunity. *Cellular Signalling,* 22**,** 1397-1405.

SWAMINATHAN, G. J., MYSZKA, D. G., KATSAMBA, P. S., OHNUKI, L. E., GLEICH, G. J. & ACHARYA, K. R. 2005. Eosinophil-granule major basic protein, a C-type lectin, binds heparin. *Biochemistry,* 44**,** 14152-8.

TAYLOR, M. E., BEZOUSKA, K. & DRICKAMER, K. 1992. Contribution to ligand binding by multiple carbohydrate-recognition domains in the macrophage mannose receptor. *J Biol Chem,* 267**,** 1719-26.

TAYLOR, M. E., CONARY, J. T., LENNARTZ, M. R., STAHL, P. D. & DRICKAMER, K. 1990. Primary structure of the mannose receptor contains multiple motifs resembling carbohydrate-recognition domains. *J Biol Chem,* 265**,** 12156-62.

TAYLOR, M. E. & DRICKAMER, K. 1993. Structural requirements for high affinity binding of complex ligands by the macrophage mannose receptor. *J Biol Chem,* 268**,** 399-404.

TAYLOR, P. R., GORDON, S. & MARTINEZ-POMARES, L. 2005. The mannose receptor: linking homeostasis and immunity through sugar recognition. *Trends Immunol,* 26**,** 104-10.

TRETTER, V., ALTMANN, F. & MARZ, L. 1991. Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase F cannot release glycans with fucose attached alpha 1----3 to the asparagine-linked N-acetylglucosamine residue. *Eur J Biochem,* 199**,** 647-52.

UNDERHILL, C. 1992. CD44: the hyaluronan receptor. *J Cell Sci,* 103 (Pt 2)**,** 293-8.

VAN DAMME, E. J. M. 2014. History of plant lectin research. *Methods In Molecular Biology (Clifton, N.J.),* 1200**,** 3-13.

VAN DEN STEEN, P., RUDD, P. M., DWEK, R. A. & OPDENAKKER, G. 1998. Concepts and principles of O-linked glycosylation. *Crit Rev Biochem Mol Biol,* 33**,** 151-208.

VAN GISBERGEN, K. P., SANCHEZ-HERNANDEZ, M., GEIJTENBEEK, T. B. & VAN KOOYK, Y. 2005. Neutrophils mediate immune modulation of dendritic cells through glycosylation-dependent interactions between Mac-1 and DC-SIGN. *J Exp Med,* 201**,** 1281-92.

VAN KOOYK, Y. & GEIJTENBEEK, T. B. H. 2003. DC-SIGN: Escape mechanism for pathogens. *Nature Reviews Immunology,* 3**,** 697-709.

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. *Am J Phys Anthropol,* Suppl 33**,** 54-69.

WASMOEN, T. L., BELL, M. P., LOEGERING, D. A., GLEICH, G. J., PRENDERGAST, F. G. & MCKEAN, D. J. 1988. Biochemical and amino acid sequence analysis of human eosinophil granule major basic protein. *J Biol Chem,* 263**,** 12559-63.

WEERAPANA, E. & IMPERIALI, B. 2006. Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology,* 16**,** 91R-101R.

WEIS, W. I., TAYLOR, M. E. & DRICKAMER, K. 1998. The C-type lectin superfamily in the immune system. *Immunological Reviews,* 163**,** 19-34.

XIA, X., YOU, M., RAO, X. J. & YU, X. Q. 2018. Insect C-type lectins in innate immunity. *Dev Comp Immunol,* 83**,** 70-79.

XU, C. & NG, D. T. 2015. Glycosylation-directed quality control of protein folding. *Nat Rev Mol Cell Biol,* 16**,** 742-52.

YABE, R., TATENO, H. & HIRABAYASHI, J. 2010. Frontal affinity chromatography analysis of constructs of DC-SIGN, DC-SIGNR and LSECtin extend evidence for affinity to agalactosylated N-glycans. *FEBS J,* 277**,** 4010-26.

YON, S. A., GRISHINA, G., SAMPSON, H. A. & SHREFFLER, W. G. 2009. The YKSL motif of DC-SIGN is Necessary for Ligand Internalization. *Journal of Allergy and Clinical Immunology,* 123**,** S126-S126.

YU, X., VASILJEVIC, S., MITCHELL, D. A., CRISPIN, M. & SCANLAN, C. N. 2013. Dissecting the molecular mechanism of IVIg therapy: the interaction between serum IgG and DC-SIGN is independent of antibody glycoform or Fc domain. *J Mol Biol,* 425**,** 1253-8.

ZAIA, J. 2010. Mass spectrometry and glycomics. *OMICS,* 14**,** 401-18.

ZAIA, J. 2011. At last, functional glycomics. *Nat Methods,* 8**,** 55-7.

ZELENSKY, A. N. & GREADY, J. E. 2005. The C-type lectin-like domain superfamily. *Febs Journal,* 272**,** 6179-6217.

ZUBAREV, R. A., KELLEHER, N. L. & MCLAFFERTY, F. W. 1998. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *Journal of the American Chemical Society,* 120**,** 3265-3266.

Thesis

FURZE, C. M. 2013. *PhD thesis 'Unravelling the mechanism of complement activation via the lectin pathway', University of Leicester, UK.*

ILYAS, R. 2010. *PhD thesis 'The role of functional glycomics in diabetes', University of Warwick, UK.*

Web Page

DRICKAMER, K. 2014a. *C-type lectins* [Online]. Division of Molecular Biosciences, Faculty of Natural Sciences, Imperial College London. Available: http://www.imperial.ac.uk/research/animallectins/ [Accessed 2018].

DRICKAMER, K. 2014b. *A genomics resource for animal lectins* [Online]. Division of Molecular Biosciences, Faculty of Natural Sciences, Imperial College London. Available: http://www.imperial.ac.uk/research/animallectins/ [Accessed 2018].

DRICKAMER, K. 2014c. *Part I : Structures and functions of animal lectins* [Online]. Division of Molecular Biosciences, Faculty of Natural Sciences, Imperial College London. Available: http://www.imperial.ac.uk/research/animallectins/ [Accessed 2018].

DRICKAMER, K. 2014d. *Sequence alignments for mammalian proteins containing CTLDs* [Online]. Division of Molecular Biosciences, Faculty of Natural Sciences, Imperial College London. Available: http://www.imperial.ac.uk/research/animallectins/ [Accessed 2018].

FORTEBIO MOLECULAR DEVICES. 2019a. *BLI Technology* [Online]. Available: https://www.fortebio.com/bli-technology.html [Accessed 2019].

FORTEBIO MOLECULAR DEVICES. 2019b. *Dip and Read Amine Reactive Second-Generation (AR2G) Biosensors* [Online]. Available: https://www.fortebio.com/documents/ForteBio_Tech_Note_26.pdf [Accessed 2019].

IUPAC. *Compendium of Chemical Terminology, 2nd ed. ("Gold Book")* [Online]. Oxford: Blackwell Scientific Publications. Available: http://goldbook.iupac.org/html/G/G02645.html [Accessed 2018].

THE CONSORTIUM FOR FUNCTIONAL GLYCOMICS. 2010. *Symbol and Text Nomenclature for Representation of Glycan Structure* [Online]. Available: http://www.functionalglycomics.org/static/consortium/Nomenclature.shtml [Accessed 2018].
THE UNIPROT CONSORTIUM. 1990. *PRG2_HUMAN* [Online]. Available: https://www.uniprot.org/uniprot/P13727 [Accessed 2018].
THE UNIPROT CONSORTIUM. 2001. *PAPP1_HUMAN* [Online]. Available: https://www.uniprot.org/uniprot/Q13219 [Accessed 2018].
THE UNIPROT CONSORTIUM. 2004. *CD209_HUMAN* [Online]. Available: https://www.uniprot.org/uniprot/Q9NNX6 [Accessed 2018].

APPENDIX

Polymer Chemistry

COMMUNICATION

View Article Online
View Journal | View Issue

CrossMark click for updates

Cite this: *Polym. Chem.*, 2016, **7**, 6293

Received 30th August 2016,
Accepted 30th September 2016

DOI: 10.1039/c6py01523a

www.rsc.org/polymers

Glyconanoparticles with controlled morphologies and their interactions with a dendritic cell lectin†

Gokhan Yilmaz,[a,b,c] Lea Messager,[d] Anne S. Gleinich,[e] Daniel A. Mitchell,[e] Giuseppe Battaglia*[d] and C. Remzi Becer*[c]

Well-defined amphiphilic block glycopolymers with equal mannose content have been self-assembled in aqueous solution to form glyconanoparticles with different morphologies. The size and shape of nanoparticles have significant effects on the interactions with the dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin (DC-SIGN; CD209), characterized using a surface plasmon resonance spectrometer (SPR).

Protein–carbohydrate interactions play an important role in many biological processes including cell interactions with immune systems, tumor metastasis, adhesion of infectious agents to host cells and many more.[1] Proteins involved in these signaling and processing pathways are known as lectins.[2] These carbohydrate binding proteins are wide-spread in nature and they differ from antibodies or enzymes entities.[3] For example, it has been shown that some viruses such as HIV express many carbohydrate entities at their surface, which enable them to bind to these lectins on the cell surface in the immune system. Therefore, one of the promising strategies for fighting against infectious diseases would be to design competing systems with higher lectin affinity than pathogens, thus preventing their adhesion.

Recent and elegant synthetic routes have allowed polymer chemists to prepare a wide range of glyco-polymers/particles that provide strong and selective recognition properties towards lectins.[4–8] In particular, amphiphilic block glycopolymers (GPs) have attracted a great attention in terms of their ability to form various types of glyconanoparticles. These amphiphilic GPs are generally composed of biocompatible, biodegradable hydrophobic polymer blocks covalently bonded to a biocompatible hydrophilic block.[9–11] Some recent studies have shown that amphiphilic block glycopolymers with different carbohydrate compositions were used to produce nanoparticles, such as micelles, nanospheres, core–shell nanoparticles, micelle-like nanoparticles, crew cut micelles, nanocapsules and polymersomes.[12–16] These studies facilitate the understanding and investigation of GP–lectin binding activities that are significantly influenced by GP architecture, valency, size, and density of binding elements. Furthermore, they offer a promising route for the creation of a broad variety of bioactive self-assembled glyco-nanostructures for biomedical applications such as drug delivery, biomaterials, bio- and nanotechnologies, and gene therapy.[17–24] In the last few years, the number of strategies devoted to prepare GPs *via* single electron transfer living radical polymerization (SET-LRP) polymerization technique has increased rapidly. Therefore, SET-LRP was used in this study to synthesize several types of amphiphilic block glycocopolymers bearing mannose moieties by using methyl acrylate (MA) and a glycomonomer (ManAc) and/or poly(ethylene glycol) (PEG) as a hydrophobic and a hydrophilic block, respectively. Mannose glycomonomer was prepared according to the procedure reported by Zhang *et al.*[25] Poly(ethylene glycol) 2-bromoisobutyrate (PEG-Br) initiator was used as an initiator to prepare amphiphilic triblock glycocopolymer. The self-assembly behaviors of these well-defined amphiphilic GPs were investigated in aqueous solution to obtain glyconanostructures with various sizes and morphologies. The resulting glyconanoparticles (GNPs) were further investigated for their binding affinity towards lectins, in particular with DC-SIGN, which is a C-type human lectin present on both macrophages and also dendritic cell subpopulations. DC-SIGN binds to microorganisms and host molecules by recognizing surfaces rich in mannose containing glycans through multivalent glycan–protein interactions and notably serves as a target molecule for several viruses such as HIV and hepatitis C virus.[26]

[a]*Department of Chemistry, University of Warwick, CV4 7AL Coventry, UK*
[b]*Department of Basic Sciences, Turkish Military Academy, Ankara, Turkey*
[c]*Polymer Chemistry Laboratory, School of Engineering and Materials Science, Queen Mary, University of London, E1 4NS London, UK. E-mail: r.becer@qmul.ac.uk*
[d]*Department of Chemistry, University College London, WC1H 0AJ London, UK*
[e]*Clinical Sciences Research Institute, Warwick Medical School, University of Warwick, CV2 2DX Coventry, UK*

† Electronic supplementary information (ESI) available: Synthesis and characterization of the polymers. See DOI: 10.1039/c6py01523a

Here, we report the synthesis of well-defined amphiphilic GPs using the SET-LRP technique, their spontaneous self-assembly in aqueous solution, and binding studies of the resulting glyconanoparticles with DC-SIGN.

A series of well-defined amphiphilic block GPs were prepared *via* SET-LRP using EBiB or PEG-Br initiator. The molar ratio of [MA]/[ManAc] has been varied to yield amphiphilic block copolymers with increasing ratios between their hydrophobic and hydrophilic moieties. One-pot polymerization approach has been employed to ensure well-defined block copolymer formation. Initially, MA was polymerized up to high conversion and subsequently; the chain extension with mannose glycomonomer (DP = 15, dissolved and degassed in 1 mL of DMSO) to form the second block was carried out (ESI, Fig. S7†). The monomer conversions were followed by ^{1}H NMR, which showed the disappearance of vinyl groups apparent between 5.8–6.4 ppm. The purified glycopolymers were further analyzed by GPC and ^{1}H NMR (ESI, Fig. S9†). Based on these results, it is evident that the copolymerization of MA and ManAc has been achieved with a good control as indicated by low polydispersity indices and increasing molecular weight of the block copolymer. Analysis of all P($(MA)_m$-*b*-$(ManAc)_n$) polymers (**P1**, **P2**, **P3**, and **P4**) are presented in Table 1. High chain end fidelity of P((PEG)-*b*-(MA)) has been confirmed by chain extension with MA prior to the synthesis of triblock glycocopolymer, **P5**, using the same approach in the presence of PEG-Br initiator. GPC analysis (ESI, Fig. S10†) revealed a clear baseline shift of elution traces after each monomer addition with increasing molecular weight and final PDI of 1.17. No significant tailing or shoulders have been observed in the GPC throughout the polymerization. Moreover, the measured $M_{n,GPC}$ and the $M_{n,theo}$ calculated from the monomer conversions are in good agreement.

Nanoparticles of these amphiphilic GPs having PEG block and/or carbohydrate bearing chains as hydrophilic block and MA as hydrophobic one were prepared using a nanoprecipitation method an the detailed conditions are summarized in the ESI.† Slow-injection of water droplets into glycopolymer in DMF solution ensures to obtain thermodynamically stable self-assembled structures. The morphology and size of the resulting nanoparticles were systematically analyzed by TEM and DLS.

Amphiphilic block copolymers **P1**, **P2** and **P3** yield well-dispersed nanoparticles of around 25 nm. While **P1** and **P2** generated micelles with a regular spherical shape, **P3** created non-uniform micelles due to their non-regular spherical shape according to TEM imaging and DLS measurements. These features are typical of spherical micelles. Further TEM analysis was conducted on each suspension measuring the average diameter of more than 50 single particles. The majority of the particles display an average diameter of 20 nm, which is slightly smaller than the values obtained by DLS, thus providing the dry structure of these particles. Furthermore, TEM imaging shows that most of the particles are not clustered, which is promising, as micelles generally tend to cluster together. Moreover, it was also observed that **P2** and **P3** yielded small amount of larger aggregates, which could potentially be vesicles. Increasing the hydrophobic content of the copolymer revealed the formation of self-assembled nanoparticles with larger hydrodynamic diameter. Indeed, the TEM image of **P4** indicated the formation of vesicular structures and DLS measurements showed a narrow size distribution centered at 380 nm.

According to TEM and DLS results, sizes of glyconanoparticles are increasing with increasing of polymer length (ESI, Table S2†). Depending on the increasing MA fraction in the polymer, the existence region of nanoparticles has become broader. This is possibly due to dependency on composition, molecular geometry, relative block lengths of the constitutive copolymers, and the preparation methods on the formation of glyconanoparticles. Another attempt was conducted on **P4** to obtain a higher number of vesicular particles, decreasing the injection rate of the aqueous solution.

The self-assembly of amphiphilic triblock GP, **P5**, has lower hydrophilic fraction than **P1–P4**, P($(MA)_m$-*b*-$(ManAc)_n$), due to the increasing molecular weight ratio of the hydrophobic block. According to TEM measurement, the majority of nanoparticles were worm-like micelles but with a small fraction of spherical micelles (Fig. 1). The length of these worm-like micelles is in the range of 45 to 60 nm and their width is 24 nm, which is in agreement with unimer chain scaling. Further increase of molar ratio of hydrophobic block (for **P4** and **P5**) allowed fabricating different aggregates such as vesicles and worm-like micelles.

Table 1 Summary of monomer conversions, number average molecular weight (M_n) and molecular weight distributions ($Đ$) of SET-LRP polymerization of all amphiphilic block glycocopolymers

Run	Polymer	[MA] : [ManAc] : [I][a]	Conv.[b] (%)		$M_{n,NMR}$ [c] (g mol^{-1})	$M_{n,GPC}$ [d] (g mol^{-1})	$Đ$
			MA	ManAc			
P1	P($(MA)_{76}$-*b*-$(ManAc)_{15}$)	80 : 15 : 1	96	98	11 850	12 400	1.16
P2	P($(MA)_{103}$-*b*-$(ManAc)_{15}$)	105 : 15 : 1	97	99	14 150	14 800	1.18
P3	P($(MA)_{130}$-*b*-$(ManAc)_{15}$)	135 : 15 : 1	96	99	16 400	17 600	1.17
P4	P($(MA)_{206}$-*b*-$(ManAc)_{15}$)	210 : 15 : 1	97	98	22 800	26 900	1.20
P5	P($(PEG)_{45}$-*b*-$(MA)_{172}$-*b*-$(ManAc)_{15}$)	180 : 15 : 1	96	99	23 800	24 600	1.17

[a] Initial molar ratio of monomers to initiator. [b] Conversions obtained from ^{1}H NMR analysis. [c] Calculated according to equation $M_{n,NMR}$ = ([M]$_0$/[I] × NMR conv.% × MW of MA and ManAc) + MW of initiator. [d] Determined by GPC (DMF as an eluent and relative to PMMA standards).

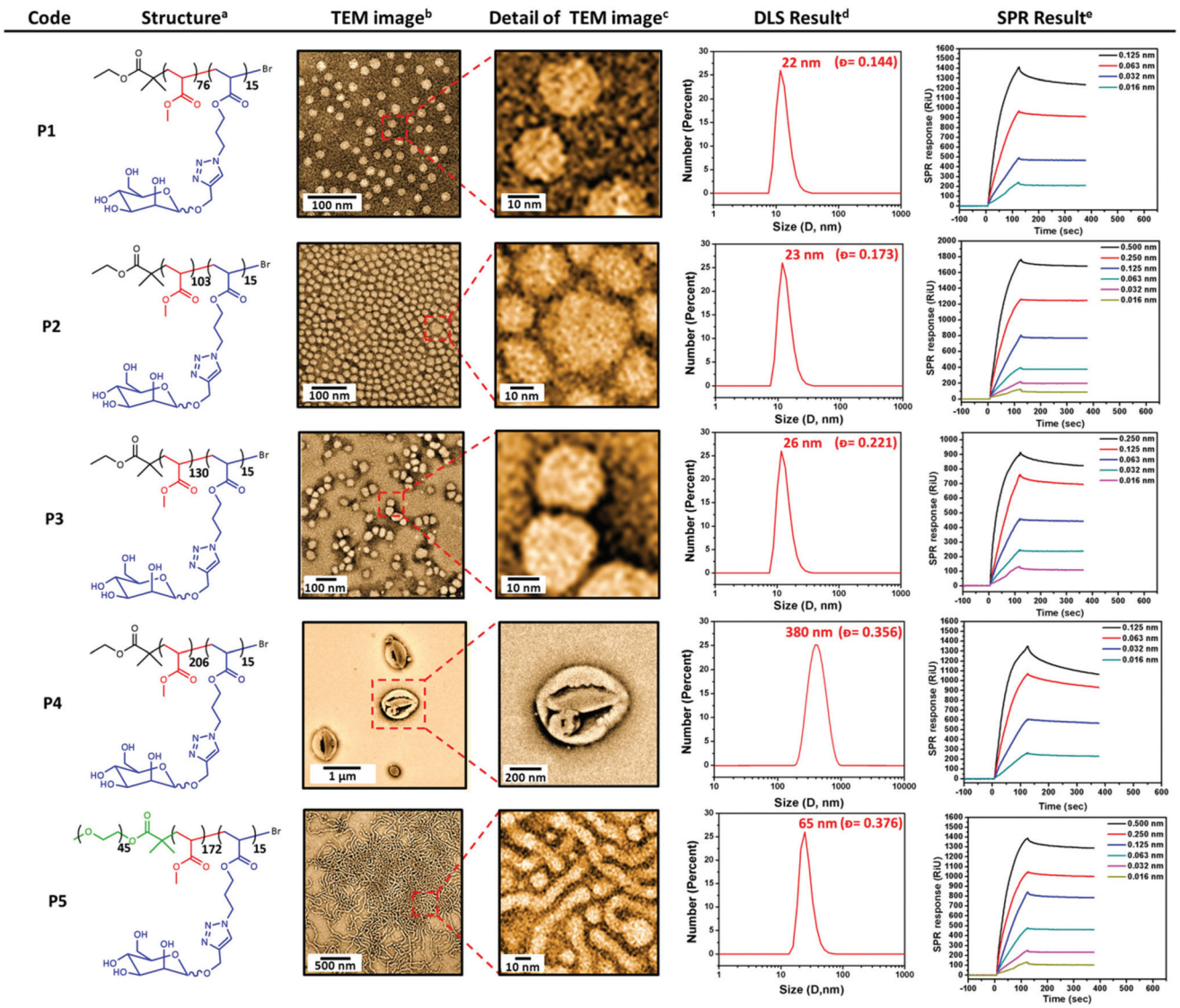


Fig. 1 Characterization of glyconanoparticles and their respective binding with DC-SIGN. [a] Chemical structures of polymers; [b] TEM images of glyconanoparticles in selected solvent condition; [c] zoomed in TEM images; [d] size of glyconanoparticles *via* DLS [e] DC-SIGN binding of glyconanoparticles measured by SPR.

The biological activity of these nanoparticles towards lectins was measured *via* SPR. DC-SIGN, which is a human lectin on cell surfaces that preferentially binds to N-linked high-mannose oligosaccharides, was employed for binding assays. DC-SIGN plays a critical role in human immunodeficiency virus (HIV) trafficking ability to interact with the highly mannosylated gp120 glycoprotein present on the envelope of HIV. In general, all glyconanoparticles showed strong binding signal with DC-SIGN (Fig. 1).

As expected, the binding of nanoparticles to DC-SIGN decreased at lower nanoparticle concentrations. All glyconanoparticles exhibited relatively high binding values at higher concentrations. Therefore, in order to compare the binding levels of these glyconanoparticles, the nanoparticle concentration of all nanoparticle solutions were kept constant to determine the effect of particle size on DC-SIGN binding (Fig. 2). According to the SPR data, vesicle structured nanoparticles **P4** had a much higher affinity in comparison to other GNPs. Surprisingly, **P1** showed the second highest binding level in this series after **P4**. This may be due to the mannose groups on the surface of **P1** glyconanoparticles being more accessible in order to mediate strong interactions with DC-SIGN. Another interesting result is that the worm-like micelles **P5** showed higher level of interaction than **P2** and **P3**. The possible explanation for this can be that the formation of worm-like micelles in clustered structures may enhance the binding capability of nanoparticles with the natural oligomeric structure of DC-SIGN. As a result, the small increase in the size of a nano-

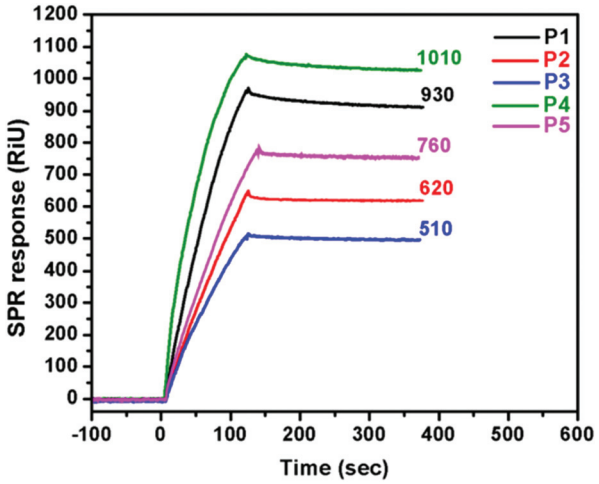


Fig. 2 SPR sensorgrams showing the binding of glyconanoparticles with DC-SIGN at the same nanoparticle concentration (0.063 nM).

particle showed a significant decrease on binding with DC-SIGN.

Conclusions

In summary, several types of amphiphilic block co-glycopolymers with optimal molecular weights and relatively narrow molecular weight distributions were synthesized *via* SET-LRP to prepare different glycopolymer nanostructures. These synthesized amphiphilic glycopolymers with the same number of mannose units self-assembled in water to generate glyconanoparticles with different morphologies such as spherical and worm-like micelles as well as spherical vesicles. The shape and size of these self-assembled glyconanoparticles were characterized *via* TEM and DLS. Finally, the interaction of these glyconanoparticles of different size and shape with DC-SIGN was monitored by SPR. The SPR results indicate that the size and shape of nanoparticles have significant effect on the binding level with DC-SIGN. According to the binding performance of these glycopolymer nanostructures, they can be potentially utilized in biomedical applications such as cell-targeted drug delivery and inhibition of viral infection such as HIV.

Acknowledgements

This work is supported by Turkish Armed Forces, and by the European Commission (EU-ITN EuroSequences Proposal No: 642083). A. S. G is supported by a studentship from The General Charity of the City of Coventry.

References

1 M. Ambrosi, N. R. Cameron and B. G. Davis, *Biomol. Chem.*, 2005, **3**, 1593.
2 C. R. Bertozzi and L. L. Kiessling, *Science*, 2001, **291**, 2357.
3 J. Arnaud, A. Audfray and A. Imberty, *Chem. Soc. Rev.*, 2013, **42**, 4798.
4 L. Wang, G. R. Williams, H. I. Nie, J. Quan and L. M. Zhu, *Polym. Chem.*, 2014, **5**, 3009.
5 Q. Zhang, L. Su, J. Collins, R. Wallis, D. Mitchell, D. M. Haddleton and C. R. Becer, *J. Am. Chem. Soc.*, 2014, **2**, 4325.
6 G. Yilmaz and C. R. Becer, *Polym. Chem.*, 2015, **6**, 5503.
7 W. Wang, D. Chance, V. Mossine and T. Mawhinney, *Glycoconjugate J.*, 2014, **31**, 133.
8 G. Yilmaz and C. R. Becer, *Eur. Polym. J.*, 2013, **49**, 3046.
9 A. Muñoz-Bonilla, O. León, V. Bordegé, M. Sánchez-Chaves and M. Fernández-García, *J. Polym. Sci., Part A: Polym. Chem.*, 2013, **51**, 1337.
10 M. Alvárez-Paino, R. Juan-Rodríguez, R. Cuervo-Rodríguez, A. Muñoz-Bonilla and M. Fernández-García, *J. Colloid Interface Sci.*, 2014, **417**, 336.
11 F. Suriano, R. Pratt, J. P. K. Tan, N. Wiradharma, A. Nelson, Y. Y. Yang, P. Dubois and J. L. Hedrick, *Biomaterials*, 2010, **31**, 2637.
12 A. Peyret, J. F. Trant, C. V. Bonduelle, K. Ferji, N. Jain, S. Lecommandoux and E. R. Gillies, *Polym. Chem.*, 2015, **6**, 7902.
13 S. Vandewalle, S. Wallyn, S. Chattopadhyay, C. R. Becer and F. Du Prez, *Eur. Polym. J.*, 2015, **69**, 490.
14 I. Kurtulus, G. Yilmaz, M. Ucuncu, M. Emrullahoglu, C. R. Becer and V. Bulmus, *Polym. Chem.*, 2014, **5**, 1593.
15 K. Godula and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2012, **134**, 15732.
16 S. F. Lou, L. Wang, G. R. Williams, H. Nie, J. Quan and L. Zhu, *Colloids Surf., B*, 2014, **113**, 368.
17 K. Babiuch, R. Wyrwa, K. Wagner, T. Seemann, S. Hoeppener, C. R. Becer, R. Linke, M. Gottschaldt, J. R. Weisser, M. Schnabelrauch and U. S. Schubert, *Biomacromolecules*, 2011, **12**, 681.
18 Y. Luo, L. Liu, X. Wang, H. Shi, W. Lv and J. Li, *J. Soft Matter*, 2012, **8**, 1634.
19 C. Y. Zhang, H. C. Yeh, M. T. Kuroki and T. H. Wang, *Nat. Mater.*, 2005, **4**, 826.
20 T. G. Iversen, T. Skotland and K. Sandvig, *Nano Today*, 2011, **6**, 176.
21 D. W. Lim, Y. I. Yeom and T. G. Park, *Bioconjugate Chem.*, 2000, **11**, 688.
22 E. Mahon, T. Aastrup and T. Barboiu, *Chem. Commun.*, 2010, **46**, 5491.
23 E. Mahon, T. Aastrup and T. Barboiu, *Chem. Commun.*, 2010, **46**, 2441.
24 A. Dag, J. Zhao and M. H. Stenzel, *ACS Macro Lett.*, 2015, **4**, 579.
25 Q. Zhang, J. Collins, A. Anastasaki, R. Wallis, D. A. Mitchell, C. R. Becer and D. M. Haddleton, *Angew. Chem., Int. Ed.*, 2013, **52**, 4435.
26 T. B. H. Geijtenbeek, D. S. Kwon, R. Torensma, S. J. van Vliet, G. C. F. van Duijnhoven, J. Middel, I. L. M. H. A. Cornelissen, H. S. L. M. Nottet, V. N. KewalRamani, D. R. Littman, C. G. Figdor and Y. van Kooyk, *Cell*, 2000, **100**, 587.

Article

pubs.acs.org/Biomac

Specific and Differential Binding of *N*-Acetylgalactosamine Glycopolymers to the Human Macrophage Galactose Lectin and Asialoglycoprotein Receptor

Joji Tanaka,[†] Anne S. Gleinich,[‡] Qiang Zhang,[†,¶] Richard Whitfield,[†] Kristian Kempe,[§,†] David M. Haddleton,[†,§] Thomas P. Davis,[§,†] Sébastien Perrier,[†,§] Daniel A. Mitchell,*[,‡] and Paul Wilson*[,†,§]

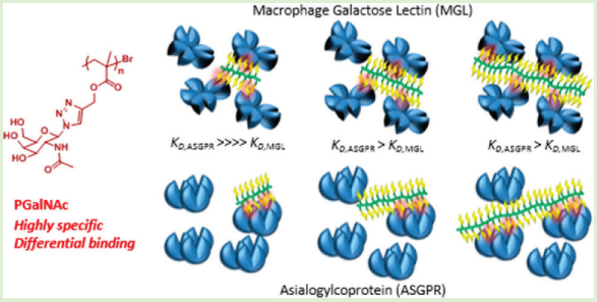
[†]Chemistry Department, University of Warwick, Library Road, CV4 7AL Coventry, United Kingdom

[‡]Clinical Sciences Research Institute, Warwick Medical School, University of Warwick, CV2 2DX Coventry, United Kingdom

[§]ARC Centre of Excellence in Convergent Bio-Nano Science and Technology, Monash Institute of Pharmaceutical Sciences, Monash University (Parkville Campus), 399 Royal Parade, Parkville, Victoria 3152, Australia

Supporting Information

ABSTRACT: A range of glycopolymers composed of *N*-acetylgalactosamine were prepared via sequential Cu(I)-mediated polymerization and alkyne−azide click (CuAAC). The resulting polymers were shown, via multichannel surface plasmon resonance, to interact specifically with human macrophage galactose lectin (MGL; CD301) with high affinity (K_D = 1.11 μM), but they did not bind to the mannose/fucose-selective human lectin dendritic-cell-specific intercellular adhesion molecule-3-grabbing nonintegrin (DC-SIGN; CD209). The effect of sugar ligand valency on the binding (so-called "glycoside cluster effect") of poly(*N*-acetylgalactosamine) to MGL was investigated by varying first the polymer chain length (DP: 100, 64, 40, 23, 12) and then the architecture (4- and 8-arm star glycopolymers). The chain length did not have a significant effect on the binding to MGL (K_D = 0.17−0.52 μM); however, when compared to a hepatic C-type lectin of a similar monosaccharide specificity, the asialoglycoprotein receptor (ASGPR), the binding affinity was more noticeably affected (K_D = 0.37− 6.65 μM). These data suggest that known differences in the specific configuration/orientation of the carbohydrate recognition domains of MGL and ASGPR are responsible for the differences in binding observed between the different polymers of varied chain length and architecture. In the future, this model has the potential to be employed for the development of tissue-selective delivery systems.



INTRODUCTION

Carbohydrate recognition controls many biological processes and is mediated through carbohydrate binding proteins termed lectins.[1] One of the most important classes of human lectins is the C-type lectin class (CTLs),[2−6] and these proteins mediate selective and specific carbohydrate recognition and subsequently drive important functions such as endocytosis.[7] A number of major C-type lectins are expressed on the surfaces of key immune cells such as macrophages and dendritic cells and have been identified as a potential platform for targeted drug delivery and synthetic vaccination.[8,9] In general, all CTLs share high sequence homology and tertiary structure, but the carbohydrate specificity can be generalized into two broad groups by tripeptide motifs in the carbohydrate recognition domain (CRD) that engages calcium chelation and H-bonding with the C3 and C4 hydroxyl groups of the target carbohydrate subcomponent. CTLs such as DC-SIGN and macrophage mannose receptor (MR; CD206) that contain the "Glu−Pro−Asn" (EPN) motif in the CRD prefer to bind pyranose units with C3/C4 hydroxyls at equatorial/equatorial positions; these include mannose, L-fucose, and *N*-acetylglucosamine (GlcNAc).[10] Physiologically, these mannose-selective CTLs are generally credited as pathogen pattern recognition receptors.[11] In contrast to the several members of EPN type CTLs in humans, there are a small set of CTLs with the "Gln−Pro−Asp" (QPD) motif in the CRD, and they have preferential binding to galactose-based sugars with the C3/C4 pyranosyl hydroxyls at an equatorial/axial position.[12] For the two major galactose-specific CTLs in the human body, ASGPR and MGL, where the structure and function are not fully understood, there have been no direct comparisons between their binding activities to synthetic ligands. ASGPR is a hepatic CTL largely

Received: February 13, 2017
Revised: March 28, 2017
Published: April 18, 2017

DOI: 10.1021/acs.biomac.7b00228
Biomacromolecules 2017, 18, 1624−1633

expressed on sinusoidal face of hepatocyte surfaces and is responsible for vascular homeostatic regulations[13] mediated by recognition of terminal galactose/GalNAc residues found on senescent desialylated glycan clusters found on circulating glycoproteins and platelets.[14] In contrast, MGL is an immunological CTL, expressed on the surface of macrophages and dendritic cells, implicated in stimulating T-cell signaling.[3]

Glycopolymers that interact with biological systems possess considerable potential as novel therapeutics and molecular probes. Recent discoveries have led to complex compositions and architectures, attracting interest in various fields for numerous applications such as biosensors,[15] drug-delivery components,[16] cryopreservation, synthetic vaccines,[9] immunodulators,[8] and cell culture matrices.[17] However, the majority of the literature focuses on interactions with plant derived lectins such as Concanavalin A (ConA),[18] Soybean Agglutinin (SBA),[19] Peanut Agglutinin (PNA)[20] and *Helix pomatia* Agglutinin (HPA).[21] In the context of human drug delivery and therapeutic applications, investigating human immunological C-type lectins such as MR, MGL, and DC-SIGN is of much higher importance and relevance.

Historically, the synthesis of well-defined glycopolymers has been a challenge because of the need for complex multistep syntheses. Recently, however, this has been alleviated by the development of methods whereby glycopolymers can be synthesized by direct polymerization of glycomonomers. For example, Davis et al. exploited chemoenzymatic reactions to synthesize (meth)acrylate-[22] and vinyl-ester-based[23] glycomonomers, which were subsequently polymerized by RAFT using dithiocarbonate and xanthate chain transfer agents, respectively. Cameron et al. polymerized α-GalNAc functionalized acrylamide by RAFT as a Tn-antigen mimic and coated gold nanoparticles as potential candidates for synthetic, carbohydrate-based anticancer vaccines.[24] Haddleton and co-workers, reported sequence controlled glycopolymers synthesized by single-electron transfer living radical polymerization (SET-LRP) from a library of glycomonomers.[25] Binding studies with DC-SIGN[26] and its competitive inhibition of the attachment of HIV glycoprotein gp120 was also reported. Finally, De Coen and co-workers used acetylated mannose to prepare a mannose−arcylamide monomer in a single step.[27]

Postpolymerization modification of functional polymer scaffolds is a complementary, efficient method for generating libraries of glycopolymers, with well-defined scaffolds accessible from commercially available or easily obtainable monomers. Glycidyl (meth)acrylate has been shown to be a versatile monomer, yielding scaffolds capable of undergoing thio-epoxy ring opening in the presence of thio-sugars,[28] as well as sequential nucleophilic ring opening, by sodium azide, and "click" with glycosyl alkynes.[29] Activated esters such as pentafluorophenyl and *N*-Hydroxysuccinimide (NHS) ester activated polymers are interesting precursor scaffolds; however, they are often hydrolytically unstable and do not necessary yield quantitative conversions.[30] Bertozzi reported a method involving the attachment of nonfunctionalized sugar to poly(acrylhydrazide) scaffolds via the reducing termini;[21] however, grafting efficiency is reduced for *N*-acetylated carbohydrates. "Click" chemistry is an attractive tool for glycopolymer synthesis because of the inherent efficiency and quantitative conversions associated with the reactions. In particular, copper(I)-catalyzed azide alkyne cycloaddition (CuAAC) has been employed over the past decade to generate glycopolymers.[31] Glycosyl azides/alkynes have been efficiently coupled to well-defined alkyne/azide-functional scaffolds, which can be synthesized by various radical polymerization protocols including RAFT,[32,33] ATRP,[34] and cobalt-mediated CCTP.[35]

Herein we report the synthesis of *N*-acetyl galactosamine glycopolymers and their specific/differential binding to recombinant human MGL and ASPGR lectins. The difference in binding with respect to the degree of polymerization (DP_n) and the molecular architecture is also reported.

EXPERIMENTAL SECTION

Materials and Methods. *N*-(Ethyl)-2-pyridylmethanimine and trimethylsilyl propargyl methacrylate were synthesized according to previously published work.[36] Copper(I) bromide (98%, Sigma-Aldrich) was treated with acetic acid and ethanol and dried under vacuum prior to use. Amberlite IR120 (Sigma-Aldrich) was treated with 1 M NaOH, deionized water and ethanol prior to use. 2-Chloro-1,3-dimethylimdazolinium chloride (DMC) (>97%, Fluka), *N*-acetyl-D-glucosamine GlcNAc (>98%, Alfa Aesar), GalNAc (99.6%, Dextra), triethylamine (TEA) (Fisher scientific, > 99%) were used as received. All the other reagents and solvents were obtained from Sigma-Aldrich at the highest purity available and used without any further purification. Dialysis tubing (1KDa MWCO) was supplied by Spectrum Laboratories. Soluble recombinant, tetrameric extracellular domains of DC-SIGN and DC-SIGNR were prepared as previously described.[37] Soluble recombinant Langerin trimeric extracellular domain was obtained from Elicityl SA (Grenoble, France). Soluble recombinant trimeric MGL extracellular domain and trimeric ASGPR H1 extracellular domain proteins were obtained from R&D Systems Inc. (Minneapolis, MN). All ^{1}H NMR and ^{13}C NMR spectra were recorded on Bruker HD 300 or 400 MHz spectrometers. The chemical shifts are reported in ppm with respect to the residual peaks of the deutrated solvents used as internal standards and ACDLABS software was used to analyze the data obtained. Infrared absorption spectra were recorded on Bruker VECTOR-22 FT-IR spectrometer using a Golden Gate diamond attenuated total reflection cell and OPUS software was used to analyze the data. Mass spectra were recorded on Bruker Esquire 2000 using ESI. SEC analysis was conducted on Agilent 1260 Infinity Multi-Detector GPC Systems in DMF (1.06 g·L^{-1} LiBr), calibrated with narrow PMMA standards (200−4.7 × 10^5 g mol^{-1}). The GPC data obtained were analyzed using Agilent Technologies GPC/SEC software. Melting point was measured using an Optimelt MPA100 system (Stanford Research Systems), and the data was analyzed with Meltview v.1.108.

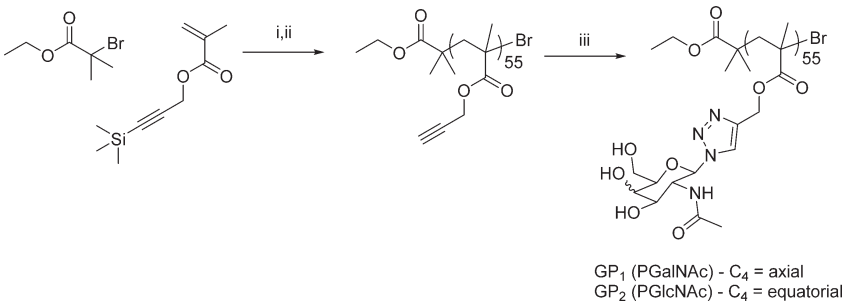
Synthesis. Representative spectra of the sugar azides and glycopolymers synthesized can be found in the Supporting Information.

Azido Sugars. 2-Chloro-1,3-dimethylimdazolinium chloride (4.6 g, 27.1 mmol, 3 equiv) was added to a mixture of H_2O (45 mL), sodium azide (5.9 g, 90.4 mmol, 10 equiv), triethylamine (9.2 g, 90.4 mmol, 10 equiv) and D-*N*-acetylgalactosamine/D-*N*-acetylglucosamine (2 g, 9.0 mmol, 1 equiv). The reaction was left stirring at 0 °C for 1 h and left to stir overnight at ambient room temperature. The reaction mixture was concentrated under reduced pressure. Ethanol (100 mL) was added and passed through a short column of Amberlite IR-120. The filtrate was concentrated under reduced pressure. H_2O (10 mL) was added and washed with dichloromethane (3 × 15 mL). The aqueous layer was collected and freeze-dried. The resulting yellow solid was redissolved in methanol (10 mL) and precipitated into dicholoromethane (100 mL) to yield the pure products β-*N*-acetylgalactosamine azide (1.2 g, 4.9 mmol, 54%) and β-*N*-acetylglucosamine azide (1.0 g, 4.1 mmol, 45%) that were spectroscopically equivalent to those previously reported in the literature.[38]

General Procedure for ATRP of Trimethylsilyl Propargyl Methacrylate (Linear). *N*-(Ethyl)-2-pyridylmethanimine (34.2 mg, 255.0 μmol, 2 equiv), ethyl α-bromoisobutyrate (24.8 mg, 127.0 μmol, 1 equiv), trimethylsilyl propargyl methacrylate (2.0 g, 10.2 mmol, 80 equiv), and toluene (2 mL) were charged to a dry Schlenk tube. The tube was sealed and subjected to seven freeze−pump−thaw cycles. The resulting degassed mixture was transferred via cannula under

DOI: 10.1021/acs.biomac.7b00228
Biomacromolecules 2017, 18, 1624−1633

Scheme 1. Synthesis of PGalNAc Prepared by Sequential Cu(I)-Mediated Polymerization and Azide−alkyne Click[a]



[a](i) CuBr, *N*-(ethyl)-2-pyridylmethanimine, toluene, 70 °C; (ii) TBAF, AcOH, THF, −20 °C; (iii) CuBr, bipy, sugar azide, DMSO, RT.

nitrogen into a second Schlenk tube, previously evacuated and filled with nitrogen, containing Cu(I)Br (18.3 mg, 127.0 μmol, 1 equiv) and a magnetic stirrer. The reaction mixture was stirred at 70 °C for 6 h. The conversion was measured by integrating the monomeric OCH_2 (4.74 ppm) and emerging polymeric OCH_2 (4.60 ppm) and comparing it against the vinylic signals (6.16 and 5.58 ppm). The polymerization was stopped at 70% conversion and terminated by diluting reaction mixture with 10 mL of toluene before bubbling air for 2 h. The terminated mixture was passed through a short column of alumina eluting with THF. The solvent was removed under pressure and redissolved in THF (10 mL) prior to precipitation in petroleum ether (200 mL). The precipitate was separated by centrifugation and decanting the solvent to yield poly(trimethylsilyl propargyl methacrylate) (1.1 g) as white solid.

For investigating the effect of chain length ($DP_{n,th}$ = 10, 20, 40, 60, 100), a benzyl functional initiator was employed for more accurate determination of the $DP_{n,NMR}$ using the aromatic protons (7.28−7.38 ppm). The amount of monomer and solvent used was kept constant, and the amount of initiator used was in accordance to eq 1. The amount of copper and ligand used was adjusted with respect to the initiator with an absolute ratio of [initiator]/[copper]/[ligand] = 1:1:2 for each polymerization.

$$DP_{Theory} = \frac{[monomer]}{[initiator]} \times conversion \quad (1)$$

ATRP of Trimethylsilyl Propargyl Methacrylate (4-Arm Star). General procedure followed using a pentaerythritol based tetrainitiator.[39] The initiator concentration was calculated according to eq 1 by taking each bromine group as an effective initiator to calculate effective initiator concentration to calculate target DP_n per arm. A ratio of $[initiator]_{eff}$/[copper]/ligand] = 1:1:2 was employed.

ATRP of Trimethylsilyl Propargyl Methacrylate (8-Arm Star). The general procedure used a lactose-based initiator[40] under more dilute conditions (25 wt % monomer) to limit the potential of star−star coupling. The initiator concentration was calculated according to eq 1 taking each bromine group as an effective initiator to calculate effective initiator concentration to calculate target DP_n per arm. A ratio of $[initiator]_{eff}$/[copper]/[ligand] = 1:1:2 ratio was employed. The reaction was terminated at 49% conversion, and the target DP_n based on eq 1 was used to estimate the final DP as the lactose core protons were obscured and could not be used to determine $DP_{n,NMR}$.

General Procedure for Deprotection. Poly(trimethylsilyl propargyl methacrylate) (1.0 g) and acetic acid (460.0 mg, 1.5 eq mol/mol with respect to the alkyne-trimethylsilyl groups) were dissolved in THF (60 mL) and sealed. The solution was bubbled with nitrogen for 20 min and then cooled to −20 °C. A solution of 1 M tetrabutylammonium fluoride hydrate (5 wt % water) in THF (2.3 mL, 1.5 eq mol/mol with respect to the alkyne-trimethylsilyl groups) was added slowly into the mixture via syringe. The resulting mixture was stirred at −20 °C for 30 min before being warmed to ambient temperature for 6 h. The resulting mixture was passed through a short column of silica, using THF as an eluent. The resulting filtrate was concentrated under reduced pressure, and the polymer was precipitated in water (150 mL). The precipitate was separated by centrifugation and decanting the water to yield poly(propargyl methacrylate) (834 mg) as white solid.

General Procedure for the Synthesis N-Acetylated Glycopolymers. Poly(propargyl methacrylate), sugar azide (1.2 eq mol/mol with respect to the alkyne groups), 2,2′-bipyridine (0.2 eq mol/mol with respect to the alkyne groups), and DMSO (5 mL) were charged into a dry Schlenk tube. The tube was sealed and subjected to seven freeze−pump−thaw cycles. The resulting degassed mixture was transferred under nitrogen into a second Schlenk tube, previously evacuated and filled with nitrogen, containing Cu(I)Br (1 eq mol/mol with respect to the alkyne groups) and a magnetic stirrer. The reaction mixture was stirred at room temperature for 36 h. The glycopolymer was precipitated into THF. The precipitate was then dissolved in water and treated with cuprisorb until the solution turned colorless. The polymer solution was then dialyzed using 1KDa MWCO dialysis tubing in deionized water for 2 days, changing the water at least twice a day. The solution was then freeze-dried overnight to yield N-acetylated glycopolymers as white solids.

Surface Plasmon Resonance (Amine Coupling Protein Surface Immobilization). Sensorgrams were recording using the ProteOn XPR36 instrument (BioRad Laboratories). Lectins were immobilized on BioRad ProteOn GLC sensor chips via amine coupling at pH 5.0 using surfaces activated with sulfo-*N*-hydroxysuccinimide. After blocking amine coupling sites with 1 M ethanolamine pH 8.0, chip flowcells were equilibrated in running buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM $CaCl_2$, 0.01% NaN_3, 0.005% Tween-20). Glycopolymer analytes were prepared in running buffer and flowed over the lectins at 25 °C and a flow rate of 25 μL/min with analyte association times of 300−900 s.

Surface Plasmon Resonance (His-Tag Capture Protein Surface Immobilization). Lectins were immobilized on BioRad ProteOn HTE sensor chips via Histidine tag-capture, using 10 mM nickel sulfate to activate the surface. Samples were run straight after lectin immobilization with running buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM $CaCl_2$, 0.01% NaN_3, 0.005% Tween-20) at 25 °C and a flow rate of 25 μL/min for association time of 300 s. Running buffer was flowed immediately afterward for dissociation phase of 375 s. The sensor chip was regenerated by complete removal of the lectins with 300 mM ethylene glycol-bis(β-aminoethyl ether)-*N*,*N*,*N*′,*N*′-tetraacetic acid (EGTA), in order to recoat the surface with new lectins.

Numerical Fit. The models (eq 2, 3) were fitted numerically using MATLAB, standard errors were gained for each parameter by confidence bound of each fit. Response at equilibrium (R_{eq}) is measured as the average response over the last 20 s of the association phase (response as the sensorgram curve platos). R_{max} and K_D are determined simultaneously by fitting eq 2:

$$R_{eq} = \frac{R_{max}[pol]}{K_D + [pol]} \quad (2)$$

DOI: 10.1021/acs.biomac.7b00228
Biomacromolecules 2017, 18, 1624−1633

To obtain dissociation rate, a decay equation (eq 3) was fitted to the dissociation phase (buffer flow). The response unit at the start of the dissociation phase was taken as the initial response unit (R_0):

$$R_t = R_0 \exp(-k_{off} t) \quad (3)$$

Association rate (k_{on}) was obtained from K_D and k_{off} according to the relationship (eq 4):

$$k_{on} = \frac{k_{off}}{K_D} \quad (4)$$

RESULTS AND DISCUSSION

Synthesis of the N-Acetylated Glycopolymers. Reversible deactivation radical polymerization (RDRP) is a popular method for the preparation of glycopolymers. This study was initiated by the synthesis and polymerization of a protected alkyne-functional monomer, trimethylsilyl propargyl methacrylate (TMS-PgMA) in accordance with previous reports (Scheme 1).[34] Atom transfer radical polymerization (ATRP) of TMS-PgMA was employed using ethyl α-bromoisobutyrate (EBiB) as initiator and Cu(I)Br/*N*-(ethyl)-2-pyridylmethanimine (NEPI) as catalyst in toluene (50 wt %) at 70 °C ([TMS-PgMA]: [EBiB]: [Cu(I)Br]: [NEPI] = [80]: [1]: [1]: [2]). The polymerization was stopped before reaching full conversion to reduce the occurrence of bimolecular termination (ρ = 68%, $DP_{n,NMR}$ = 69, $M_{n,NMR}$ = 13600 g mol^{-1}, $Đ$ = 1.30, Table 1, entry 1). Conversions were determined by monitoring the disappearance of vinylic protons at δ = 6.11−5.56 ppm, relative to residual $CHCl_3$ (7.26 ppm) and the degree of polymerization (DP_n) and number-average molecular weight ($M_{n,NMR}$) were determined by comparing the methylene signal present in the initiator end group (δ = 4.09 ppm) with a methylene signal in the polymer side chain at purification (δ = 4.60 ppm, Figure S1). The trimethylsilyl protecting groups were cleaved using tetra-*N*-butylammonium fluoride (TBAF) and acetic acid to afford an alkyne-functional polymer scaffold ($M_{n,NMR}$ = 8700 g mol^{-1}, $Đ$ = 1.40, Table 1, entry 2) as indicated by the disappearance of the TMS signal in ^{1}H NMR (δ = 0.2 ppm) which coincided with the appearance of a signal for the alkyne proton (δ = 2.5 ppm, Figure S2). Anomerically pure β-glycosyl azides of *N*-acetyl galactosamine (GalNAc) and *N*-acetyl glucosamine (GlcNAc), prepared via the one-step procedure of S. Shoda et al.[41−43] (Figure S3−6), were then "clicked" onto the pendant alkyne groups according to a literature procedure[28] using Cu(I)Br and bypyridine to furnish GalNAc (GP1, Table 1, entry 3) and GlcNAc (GP2, Table 1, entry 4) functional glycopolymers from a single polymer scaffold. The success of the cycloaddition reaction was confirmed by FT-IR, with the disappearance of the alkyne peak (ν(C−H): 3286 cm^{-1} and ν(C≡C): 2129 cm^{-1}, Figure S7) and ^{1}H NMR with the appearance of a signal corresponding to the triazole hydrogen (δ = 8.30 ppm, Figure S8−9).

Table 1. Molecular Weight Data for the Synthesized Polymer Scaffold and Glycopolymers

	polymer	$M_{n,th}$ (g mol^{-1})	$M_{n,NMR}$ (g mol^{-1})	$M_{n,SEC}$[c] (g mol^{-1})	$Đ$[c]
1	P(TMSPgMA)	10900[a]	13600[b]	15400	1.30
2	P(PgMA)	6900[d]	8700[d]	9700	1.40
3	GP1	20300[d]	25600[d]	22600	1.25
4	GP2	20300[d]	25600[d]	24300	1.26

[a]Determined by $MW_{monomer}$ × conversion × $[M]_0/[I]_0$ + $MW_{initiator}$. [b]As calculated by NMR of the α-end group to determine DP_n. [c]Determined by size-exclusion chromatography (with DMF as the eluent) using conventional calibration obtained with poly(methyl methacrylate) standards (Figure S10). [d]M_n based DP_n = 69 for the P(TMSPgMA) scaffold.

Investigating Binding Interactions of MGL with *N*-Acetylated Glycopolymers (GP1−GP2). MGL has only been reported to be expressed in subtypes of macrophages and dendritic cells as a scavenger receptor that facilitates antigen presentation of nonself-antigens.[44] MGL expression is implicated in cancer progression and binds to GalNAc based tumor-associated carbohydrate (Tn-antigen) on the mucin molecule MUC1[45] expressed by malignant cells.[46] Binding to MGL is reported to drive Th2-mediated responses via the MHC class II processing and presentation pathway,[47] as opposed to tumorcidal Th1 responses.[48] Although the engagement of MGL by ligands per se may not direct a specific response, it may influence and impact responses via coengagement of other receptors.[49] Therefore, from a bottom-up approach, it was necessary to investigate whether our polymers could also interact with MGL as well as closely related CTLs. The association and binding of glycopolymers to the target lectins was investigated using multichannel surface plasmon resonance (SPR).

Soluble, recombinant MGL extracellular portion (CRD domain and coiled-coil neck) was immobilized on the chip surface through EDC-NHS mediated amine coupling with carboxylic acid coated alginate on the SPR sensor chip.[50] Extracellular portions of the closely related immunological CTLs: DC-SIGN, DC-SIGNR and Langerin were also immobilized onto chips for parallel investigations to assess their interactions with the *N*-acetylated glycopolymers (**GP1** and **GP2**). The glycopolymers (range: 1.25−40 μM in serial 2:1 dilution) were flowed over, in parallel, across each lectin-bearing channel, followed immediately by buffer alone (see Supporting Information for full details). The sensorgrams (Figure 1) show a clear concentration-dependent association of PGalNAc **GP1** to MGL with slow dissociation. GP1 did not bind to the other lectins (DC-SIGN, DC-SIGNR, and Langerin) inferring high selectivity for MGL. In contrast, PGlcNAc, **GP2** was observed to bind to DC-SIGN and other related EPN-type lectins, but did not bind to MGL (Figure 1, Table 2). The extent of binding can be quantified using the SPR software via a two-state model for glycopolymer−lectin interactions.[26,50] From this fitting, two K_D values are obtained. For the purpose of this investigation a simple Langmuir-type isotherm fitting of each interaction was employed as a parsimonious model to quantify the glycopolymer-lectin interaction as a single dissociation equilibrium constant (K_D, eq 1). Individual rate constants were obtained separately; dissociate rate (k_{off}) from the dissociation phase (eq 3) and an apparent association rate, as the quotient of K_D and k_{off}.

Investigating the Effect of Chain Length and Polymer Architecture on Binding Affinity. Typically, plant- and human-derived lectins exist as multimeric assemblies, and as a result, individual CTL−sugar interactions are weak. Multimeric, or multivalent sugar ligands often confer much stronger binding than monomeric and short oligomeric equivalents. This phenomenon is widely referred to as the "glycoside cluster effect".[51] One of the advantages of using RDRP protocols is the ability to control polymer chain length and architecture. Thus, it is possible to manipulate ligand multivalency by varying the

DOI: 10.1021/acs.biomac.7b00228
Biomacromolecules 2017, 18, 1624−1633

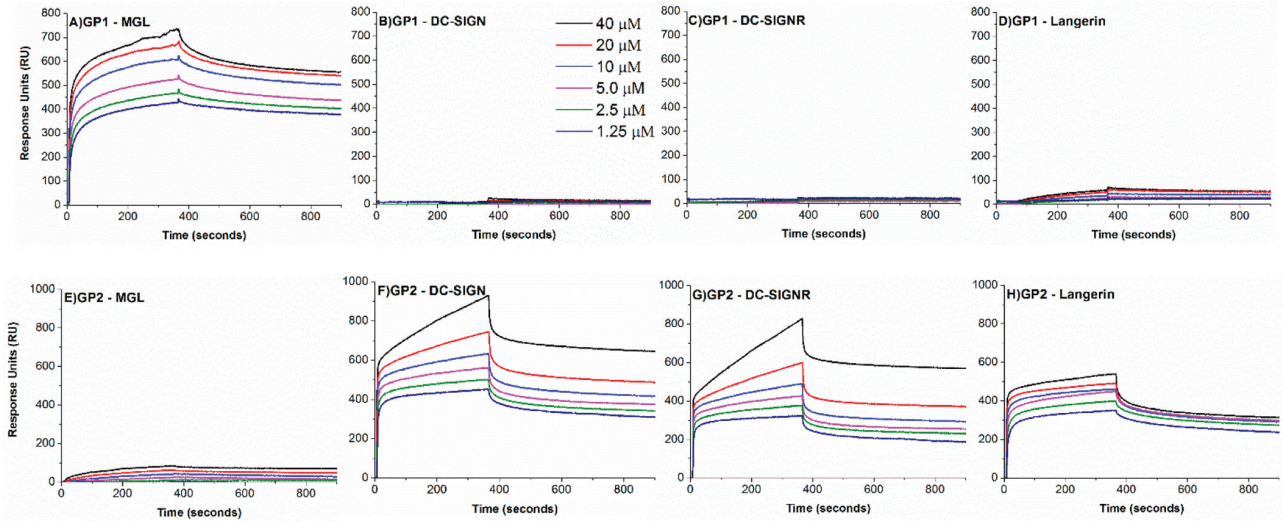


Figure 1. SPR sensorgrams for the binding of glycopolymers **GP1** (A–D) and **GP2** (E–H) with lectins MGL, DC-SIGN, DC-SIGNR, and Langerin. **GP1**/**GP2** were flowed at 0.25 μLs^{1-} for 375 s with varying concentration range to determine association. Buffer was flowed between t = 375 and 900 s to establish the dissociation phase.

Table 2. Binding Parameters of Glycopolymers GP1 and GP2 with Human Lectins Obtained from SPR, eq 2, and eq 3

pol	lectin	$k_{on,app}$ ($\times 10^2$ M^{-1} s^{-1})	k_{off}[b] ($\times 10^{-4}$ s^{-1})	K_D[a] (nM)
GP1	MGL	2.62	2.88	1106
	DC-SIGN			
	DC-DIGNR			
	Langerin			
GP2	MGL			
	DC-SIGN	2.49	3.59	1442
	DC-SIGNR	1.12	3.24	2907
	Langerin	8.58	4.95	577

[a] As determined by : $Req = \frac{Rmax[pol]}{K_D+[Pol]}$ (eq 1)

[b] As determined by: $R_t = R_0 \exp(-k_{off}t)$ (eq 2)

number and orientation of pendant sugar groups present within target glycopolymers. Furthermore, the possibility of forming well-defined block copolymers provides access to higher-ordered structures such as micelles, vesicles, and nanogels wherein the density of sugar ligands present can in intricately controlled by the monomer feed. To this end, De Coen and co-workers have investigated the potential of mannosylated nanogels to target MR expressed on dendritic cells and found that increased particle size and ligand (mannose) density resulted in enhanced receptor binding and internalization.[27]

To investigate the effect of polymer chain length and architecture on binding to MGL, a library of linear PGalNAc polymers (**GP3–GP7**) of varying chain length (DP$_n$ 12–100, Table S1, Figure S11) were prepared by ATRP employing a benzyl initiator[34] which enabled more accurate analysis of the number-average molecular weight by ^{1}H NMR by comparing the distinctive benzyl protons (7.28–7.38 ppm) against OCH_2-R (4.59 ppm) of the polymer side chain (Figure S12–S17). Likewise, alternative architectures were prepared using 4-arm (pentaerythritol-based) and 8-arm (lactose-based) initiators (Figure 2). Employing ATRP, the copper and ligand concentration was adjusted to account for each initiating group. The DP$_n$ of the 4-arm star PGalNAc (**GP8**, $M_{n,Th}$ = 45600 g·mol^{-1}, $Đ$ = 1.30, Table S1, Figure S11) was determined as 30 per arm according to the monomer conversion ($\rho \sim 60\%$) from an [M]/[I] ratio of 200 employed during P(TMSPgMA) scaffold synthesis. The core methylene protons (4.33 ppm) from the initiator were obscured, which prohibited accurate determination of DP$_n$ by ^{1}H NMR (Figure S18). Likewise the DP$_n$ of the 8-arm star PGalNAc (**GP9**, $M_{n,Th}$ = 68000 g·mol^{-1}, $Đ$ = 1.22, Table S1, Figure S11) was also determined by monomer conversion as the core lactose ring protons (3.84–5.86 ppm) could not be reliably integrated by ^{1}H NMR (Figure S19). Polymerization from the 8-arm initiator was carried out under more diluted conditions (25 wt %) with [M]/[I] ratio of 360 and was terminated at lower monomer conversion ($\rho \sim 49\%$) to reduce the occurrence of star–star coupling,[52,53] furnishing an 8-arm star PGalNAc with DP$_n$ of 22 per arm.

During the initial study (**GP1–2**), all lectins were covalently bound to SPR chips via established amine coupling chemistry.[26] Removal of bound polymers from covalently immobilized lectins, usually via chelation chemistry (EGTA)[54] or simple acid treatment (pH 3 glycine buffer),[26] limits the number of sequential experiments due to denaturation of the bound proteins. For a more reliable comparison of the PGalNAc library (**GP3–9**), fresh lectin was used for each SPR experiment to ensure the binding was not influenced by the previous experiment. This was carried out using Bio-Rad Laboratories ProteOn HTE Sensor chip, which contains *tris*-nitrilotriacetic acid, reported to have subnanomolar affinity toward His-tagged molecules.[55] Thus, His-tagged MGL was reversibly immobilized to a nickel-sulfate-activated SPR chip prior to analysis. Polymer-bound lectins were then rapidly removed using chelating agent EGTA before immobilization of a fresh batch of His-tagged MGL between each subsequent experiment.

The binding of PGalNAc to MGL was largely unchanged as a function of the polymer valency (Figure 3). Although increasing the linear chain length, from DP$_n$ = 12 to DP$_n$ = 100 (**GP3–7**) was found to lower the overall K_D of the polymer, the differences were not significant (517–169 nM,

DOI: 10.1021/acs.biomac.7b00228
Biomacromolecules 2017, 18, 1624–1633

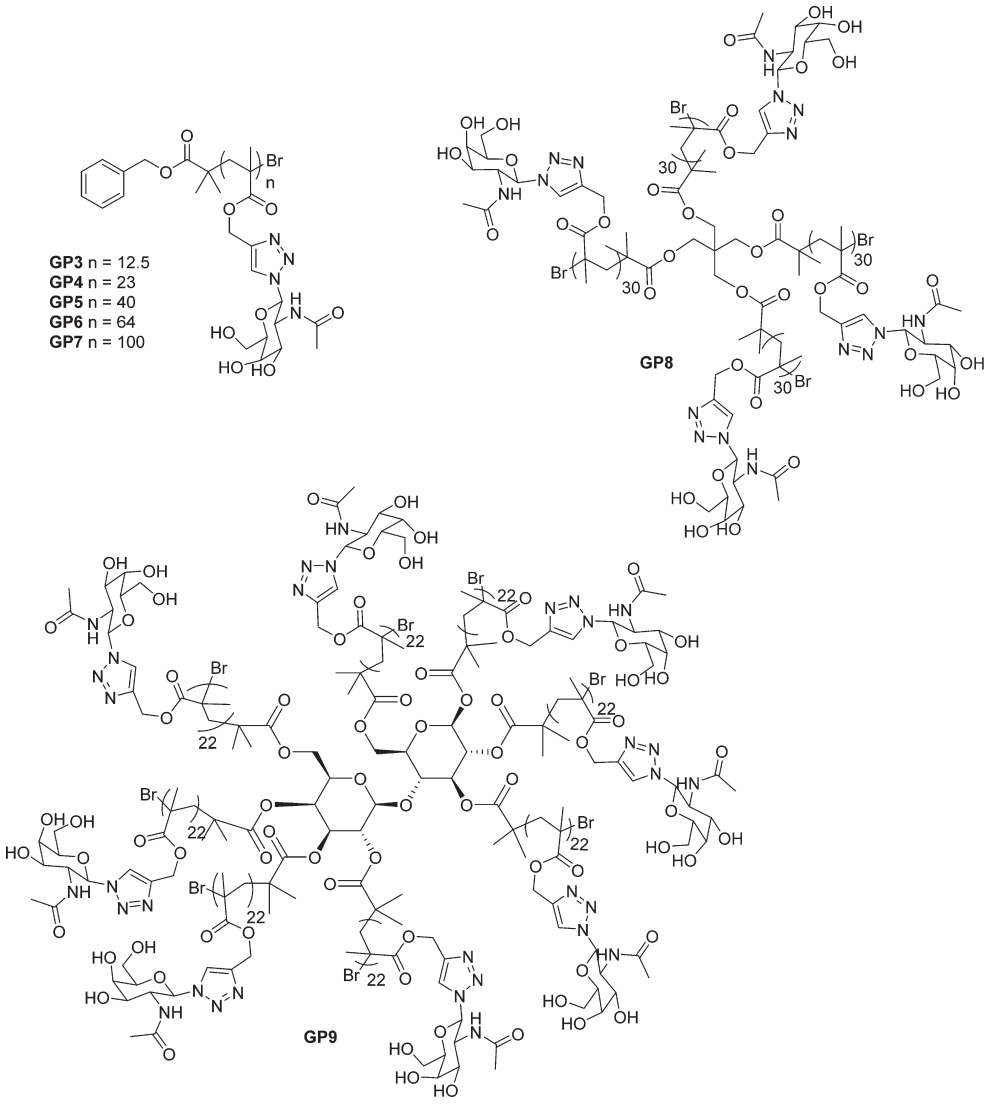


Figure 2. PGalNAc library synthesized by ATRP to investigate the effect of ligand valency on lectin binding.

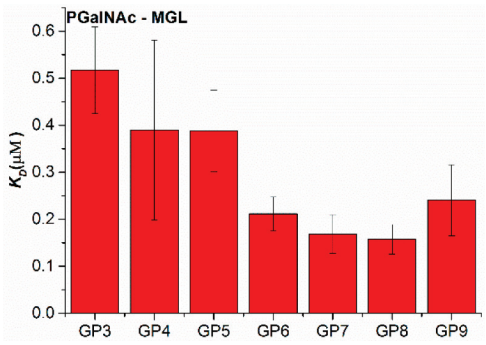


Figure 3. Binding constants of glycopolymers GP3-GP9 with MGL determined from SPR (Figure S20).

Figure 3, Table 3). The architecture of the PGalNAc polymers also had little effect on the binding to MGL with the 4-arm star (**GP8**) and 8-arm star PGalNAc (**GP9**), exhibiting comparable binding to the linear analogues (157 nM and 240 nM, respectively, Figure 3, Table 3). A similar observation was previously reported by Stenzel and co-workers, whereby the binding of glucose-containing 4-arm star glycopolymers to ConA, did not show a significant effect on binding compared to linear analogues.[18] As the association rate directly reflects the dissociation constant by our method, a similar trend was observed whereby PGalNAc appeared to bind marginally faster as a function of increasing valency (410−1870 $M^{-1}{\cdot}s^{-1}$, Table 3), whereas only minor changes in the dissociation rates were determined (1.66×10^{-4} s^{-1}−3.45×10^{-4} s^{-1}, Table 3) with no obvious trend with respect to valency.

To date, the structure of MGL has not been fully elucidated, but it has been reported to share similarities with DC-SIGN and mannose binding protein (MBP) where the CRDs are relatively distant and face away from one another. Such an orientation might be expected to be ideal for cross-linking via multivalent ligands.[56] Indeed, one of the immunological functions of MGL has been attributed to cross-linking of CD45 expressed on the surface of activated T cells, modulating their behavior.[57] Structurally, chemical cross-linking experiments have shown that MGL exists as a homotrimer.[56] Trimerization occurs in the neck domain and is independent

DOI: 10.1021/acs.biomac.7b00228
Biomacromolecules 2017, 18, 1624−1633

Table 3. Effect of Ligand Valency on the Binding of PGalNAc Glycopolymers to MGL and ASGPR

				MGL			ASGPR		
	DP_n[a]	M_n[a] (g·mol^{-1})	$Đ$[c]	K_D (nM)	$k_{on,app}$ (× 10^2 M^{-1}·s^{-1})	k_{off} (× 10^{-4}·s^{-1})	K_D (nM)	$k_{on,app}$ (× 10^2 M^{-1}·s^{-1})	k_{off} (× 10^{-4}·s^{-1})
GP3	12	4700	1.15	517	4.10	2.12	6653	0.35	2.33
GP4	23	8800	1.34	390	4.27	1.66	6652	0.39	2.58
GP5	40	15100	1.30	388	8.90	3.45	5535	0.71	3.94
GP6	64	24000	1.13	211	9.17	1.94	712	3.56	2.57
GP7	100	37300	1.21	169	15.70	2.64	372	12.00	4.47
GP8	30/arm[b]	45600[b]	1.30	157	18.70	2.94	341	10.20	3.47
GP9	22/arm[b]	68000[b]	1.22	240	13.50	3.26	560	8.23	4.61

[a]Determined by ^{1}H NMR via integration of the benzyl protons (7.28−7.38 ppm) of the chain end against OCH_2-R (4.59 ppm) of the polymer. [b]Approximated from data obtained for the polymer scaffold (Table S1). [c]Determined by size-exclusion chromatography (with DMF as the eluent) using conventional calibration obtained with poly(methyl methacrylate) standards (Figure S11). [d]Determined SPR whereby GP3−9 were flowed at 0.25 μLs-1 for 375 s to afford association before buffer was flowed (t = 375−900 s) to establish the dissociation phase (Figures S20,S21).

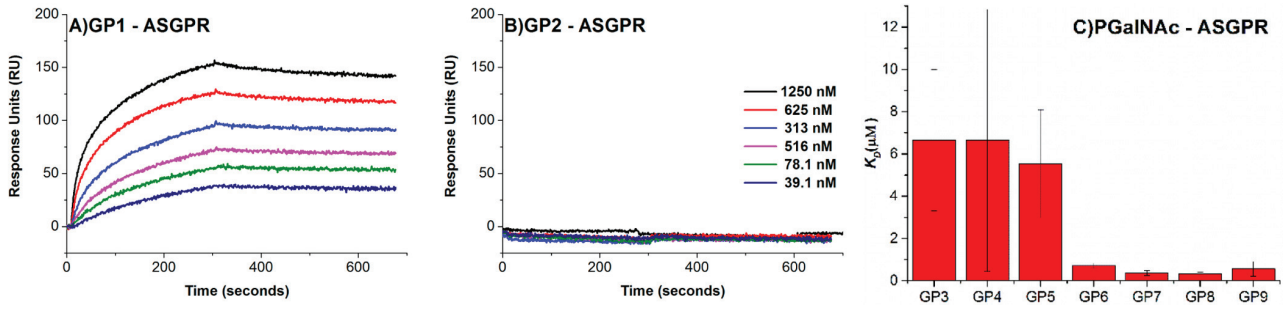


Figure 4. SPR sensorgrams for the binding of glycopolymers **GP1** (A) and **GP2** (B) with ASGPR. **GP1**/**GP2** were flowed at 0.25 μL s^{-1} for 300 s with varying concentration range to determine association. Buffer was flowed between t = 300 and 700 s to establish the dissociation phase. Binding constants of glycopolymers **GP3−GP9** with APSGR determined from SPR (Figure S21).

of the CRD, with the junction between the two domains being protease-sensitive. This indicates that the CRD consists of three discrete, distally orientated sugar binding sites that are potentially dynamic and capable of adjusting to compliment the glycan clusters presented to the lectin surface, similar to MBP.[56] Therefore, it is proposed that the three binding sites interact independently with the PGalNAc polymers employed during this investigation. On a molecular level, this translates to the polymers not having to traverse the multiple binding sites of a single MGL trimer to enhance binding.

Relative Binding of PGalNAc with ASGPR and MGL. In comparison to MGL, ASGPR was discovered much earlier,[58] and the structure and function has been more extensively explored.[59] Native human ASGPR consists of two different chains, H1 and H2, forming a hetero-oligomer.[60] Despite this, to date there are currently no crystal structures of the complete extracellular domain of the native H1/H2 heterooligomer, but the arrangement and orientation of the CRD are well-investigated, and optimal ASGPR ligands reported in the literature are selected on the basis of the spatial arrangement of Gal residues within the native tri/tetra-antennary *N*-glycans.[61] The recombinant ASGPR used during this investigation (CRD and neck domain of H1 subunit), is a self-assembled homotrimer,[62] the structure of which closely resembles the native hetero-oligomer lectin, in which the binding sites are in close proximity,[54] presenting an interface more suitable for clustered glycosides, typically at the termini of a single, branched oligosaccharide. From previous literature pertaining to monosaccharide selectivity,[63] it was hypothesized that APSGR would exhibit similar binding selectivity to MGL. Having demonstrated binding of PGalNAc (**GP1**, 1.62 μM), and a lack of binding of PGlcNAc (**GP2**) to ASGPR (Figure 4A,B), the effect of chain length and architecture was again investigated. The overall binding of PGalNAc (**GP3-GP9**) to ASGPR was found to be lower than MGL, particularly for short polymer chain lengths (**GP3−5**, Table 3). However, a more pronounced effect on binding with changing the valency of PGalNAc was observed.

At shorter chain lengths, the binding of linear PGalNAc (**GP3−5**) to ASGPR was comparable (**GP3−5,** 6.7−5.5 μM). When the chain length was increased further, the dissociation constant decreased much more significantly than previously observed for MGL (**GP6−7**, 712−372 nM). Similarly to MGL, no significant effect was observed upon changing the polymer architecture (**GP8−9,** 560−341 nM). Molecularly, this indicates that low-molecular-weight PGalNAc (**GP3−5**) share similar binding modes, as indicated by minimal changes in the K_D. ASGPR has been reported to bind markedly weaker to monovalent GalNAc ligands, and as such, ASGPR ligands are often based on a tri/tetra-antennary design to compliment the more rigid, proximal binding sites of the multivalent ASGPR oligomer.[64,65] When the chain length of the PGalNAc ligand is increased beyond DP_n = 64, the binding to ASGPR is markedly enhanced as a result of the "glycoside cluster effect" whereby the multivalent PGalNAc ligands are capable of binding to multiple binding sites of ASGPR, thus overcoming the individually weak interactions.

The differential binding of MGL and ASGPR was further exemplified using a GalNAc-BSA conjugate, where single GalNAc moieties are randomly exposed through conjugation to surface serine and threonine residues. Only binding to MGL, with its distally orientated, flexible CRD, was observed, and no binding was observed to the more rigid, proximal binding sites of ASGPR which requires more closely configured sugar ligands

DOI: 10.1021/acs.biomac.7b00228
Biomacromolecules 2017, 18, 1624−1633

(Figure 5). The PGalNAc polymers synthesized during this investigation were predicted to be long enough to span the

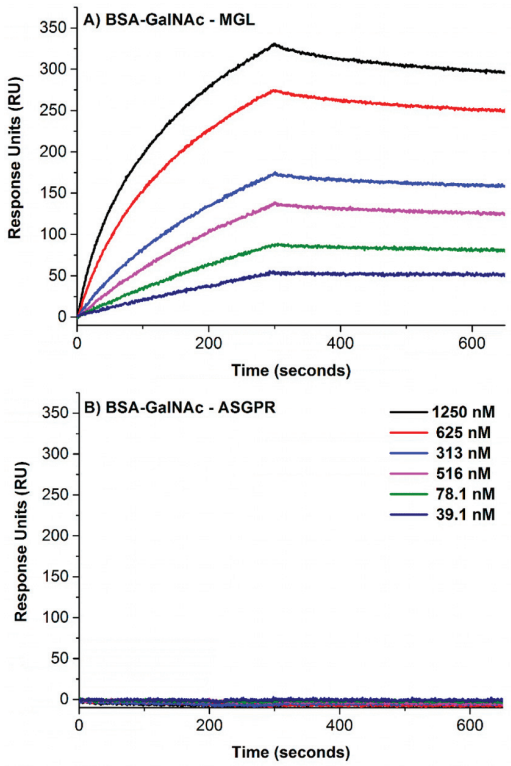


Figure 5. SPR sensorgrams for the binding of BSA-GalNAc conjugate to MGL (A) and ASGPR (B). BSA-GalNAc conjugate was flowed at 0.25 μL s^{-1} for 300 s with varying concentration range to determine association. Buffer was flowed between t = 300 and 700 s to establish the dissociation phase.

proximal binding sites of the ASGPR which are separated by approximately 15−25 Å.[66] With this in mind, the shortest PGalNAc (DP_n = 12.5, **GP3**) should be sufficiently long enough to bind to at least two of the binding sites based on a perfectly linear conformation, while **GP5** (DP_n = 40) should be sufficiently long enough to bind to all three binding sites of the trimer. However, the glycopolymers do not exist as ideal linear chains in solution, and as a result changes in the K_D were only observed at longer chain lengths (**GP6−GP7,** DP_n = 64, 100). Thus, the "glycoside cluster effect" was apparent at longer chain lengths as the conformation of the polymer in solution promotes binding at multiple sites of the lectin CRD. Furthermore, beyond a critical chain length, secondary interactions that result in cross-linking between ASGPR oligomers are possible which could contribute to the stronger binding interactions observed.

Bertozzi et al. previously described an analogous system in plant lectins; *Helix pomatia* agglutinin (HPA) and soybean agglutinin (SBA), where HPA (proximal binding sites) was found to interact strongly in a valence-independent manner and SBA (distant binding sites) to reversibly "bind and slide", leading to weaker interactions.[19] Conversely, this investigation has demonstrated that binding to ASGPR, with proximal binding sites, occurs in a valence-dependent manner, whereas MGL, with its flexible, distal binding sites binds independent of valence with affinity attributed a preference for multimolecular lectin binding via cross-linking through the glycopolymer scaffold. These contrasting conclusions serve to highlight the complexity of lectin/glycopolymer interactions which are strongly dependent upon the nature of lectin(s) and polymers under investigation.

CONCLUSIONS

A library of PGalNAc glycopolymers have been used for the first time to perform binding studies with human CTL's MGL and ASGPR. A polyalkyne scaffold prepared by Cu(I)-mediated radical polymerization was functionalized with azido-GalNAc and azido-GlcNAc respectively, and the binding and selectivity of the resulting polymers was assessed against target lectins MGL and ASGPR as well as related CTLs DC-SIGN, DC-SIGNR, and Langerin using SPR. PGalNAc was found to specifically bind to MGL and shows little affinity for related CTLs. Conversely, PGlcNAc exhibits little affinity for MGL but binds to the related CTLs with varying affinity. Increasing the ligand valency by increasing the DP_n of the PGalNAc resulted in only a small increase in affinity, which remained unchanged as a function of polymer architecture. This is attributed to the distal orientation and flexibility of the carbohydrate binding domain of MGL. ASGPR, which is known to have a more proximal orientated binding domain, exhibits comparable selectivity of PGalNAc and conforms in part to the "glycoside cluster effect". For example, increasing the GalNAc valency by increasing the DP_n of the PGalNAc exhibited a marked increase in affinity. The specific binding thresholds reported are opportunistic for further exploitation, with potential implications in tissue-selective drug delivery. For example, it is known that GalNAc based ligands can successfully target the liver via ASGPR in vivo without being obstructed by the immune system.[67] Therefore, ASGPR represents a promising entry for liver-specific drug and gene delivery using PGalNAc glycopolymers. Alternatively, selectively targeting the immune system using GalNAc-based systems represents a greater challenge due to competitive uptake by the hepatic system. It is hoped that the differential binding observed between MGL and ASGPR in this study can be exploited to systemically target macrophages and DCs via MGL without off-target uptake to liver via ASGPR. Due to their pivotal role in tumor immunology, macrophages and DCs could provide a set of clinically valuable targets.[68] Thus, targeting tumor environments via the carbohydrate binding properties of MGL could provide a favorable and selective therapeutic delivery strategy because of its higher expression in sites of solid tumors.[69]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biomac.7b00228.

Representative spectra of the sugar azides and glyocpolymers synthesized as well the SPR sensograms obtained (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: d.mitchell@warwick.ac.uk.
*E-mail: p.wilson.1@warwick.ac.uk.

ORCID

Richard Whitfield: 0000-0003-4787-2060

DOI: 10.1021/acs.biomac.7b00228
Biomacromolecules 2017, 18, 1624−1633

Kristian Kempe: 0000-0002-0136-9403
David M. Haddleton: 0000-0002-4965-0827
Thomas P. Davis: 0000-0003-2581-4986
Sébastien Perrier: 0000-0001-5055-9046
Paul Wilson: 0000-0002-9760-899X

Present Address

¶(Q.Z.) Nanjing University of Science and Technology, School of Environmental and Biological Engineering, Nanjing University, Luohan Alley, Xuanwu Qu, Nanjing Shi, Jiangsu Sheng, China, 210094.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from Engineering and Physical Sciences Research Council (EPSRC) under grant EP/F500378/1 through the Molecular Organisation and Assembly in Cells Doctoral Training Centre (MOAC-DTC). The authors also wish to acknowledge the facilities and personnel (K.K., D.M.H., T.P.D., S.P., P.W.) enabled by the Monash-Warwick Alliance. This work was carried out in conjunction with the Australian Research Council (ARC) Centre of Excellence in Convergent Bio-NanoScience and Technology (CE140100036). A.S.G. holds a Medical Research Award from the General Charity of the City of Coventry. R.W. is funded by Syngenta. K.K. gratefully acknowledges the award of a NHMRC-ARC Dementia Research Development Fellowship (APP1109945). T.P.D. gratefully acknowledges support from the ARC in the form of an Australian Laureate Fellowship. S.P. acknowledges a Royal Society Wolfson Merit Award (WM130055). P.W. thanks the Leverhulme Trust for the award of an Early Career Fellowship (ECF/2015-075).

■ REFERENCES

(1) Zanetta, J.-P.; Kuchler, S.; Lehmann, S.; Badache, A.; Maschke, S.; Thomas, D.; Dufourcq, P.; Vincendon, G. *Histochem. J.* **1992**, *24*, 791−804.
(2) Glavey, S. V.; Huynh, D.; Reagan, M. R.; Manier, S.; Moschetta, M.; Kawano, Y.; Roccaro, A. M.; Ghobrial, I. M.; Joshi, L.; O'Dwyer, M. E. *Blood Rev.* **2015**, *29*, 269−279.
(3) van Kooyk, Y.; Ilarregui, J. M.; van Vliet, S. J. *Immunobiology* **2015**, *220*, 185−192.
(4) Liu, F.; Patterson, R. J.; Wang, J. L. *Biochim. Biophys. Acta, Gen. Subj.* **2002**, *1572*, 263−273.
(5) Svajger, U.; Anderluh, M.; Jeras, M.; Obermajer, N. *Cell. Signalling* **2010**, *22*, 1397−1405.
(6) Zachara, N. E.; Hart, G. W. *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids* **2006**, *1761*, 599−617.
(7) Spiess, M. *Biochemistry* **1990**, *29*, 10009−10018.
(8) Ohta, M.; Ishida, A.; Toda, M.; Akita, K.; Inoue, M.; Yamashita, K.; Watanabe, M.; Murata, T.; Usui, T.; Nakada, H. *Biochem. Biophys. Res. Commun.* **2010**, *402*, 663−669.
(9) Sunasee, R.; Narain, R. *Macromol. Biosci.* **2013**, *13*, 9−27.
(10) Lee, R. T.; Hsu, T. L.; Huang, S. K.; Hsieh, S. L.; Wong, C. H.; Lee, Y. C. *Glycobiology* **2011**, *21*, 512−520.
(11) Cambi, A.; Koopman, M.; Figdor, C. G. *Cell. Microbiol.* **2005**, *7*, 481−488.
(12) Kolatkar, A. R.; Weis, W. I. *J. Biol. Chem.* **1996**, *271*, 6679−6685.
(13) Grozovsky, R.; Begonja, A. J.; Liu, K.; Visner, G.; Hartwig, J. H.; Falet, H.; Hoffmeister, K. M. *Nat. Med.* **2014**, *21*, 47−54.
(14) Hoffmeister, K. M. *J. Thromb. Haemostasis* **2011**, *9*, 35−43.
(15) Richards, S.-J.; Otten, L.; Gibson, M. I. *J. Mater. Chem. B* **2016**, *4*, 3046−3053.
(16) Song, E.; Manganiello, M. J.; Chow, Y.; Ghosn, B.; Convertine, A. J.; Stayton, P. S.; Schnapp, L. M.; Ratner, D. M. *Biomaterials* **2012**, *33*, 6889−6897.
(17) Hayward, A. S.; Eissa, A. M.; Maltman, D. J.; Sano, N.; Przyborski, S. A.; Cameron, N. R. *Biomacromolecules* **2013**, *14*, 4271−4277.
(18) Chen, Y.; Chen, G.; Stenzel, M. H. *Macromolecules* **2010**, *43*, 8109−8114.
(19) Godula, K.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2012**, *134*, 15732−15742.
(20) Lu, J.; Zhang, W.; Richards, S.-J.; Gibson, M. I.; Chen, G. *Polym. Chem.* **2014**, *5*, 2326−2332.
(21) Godula, K.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2010**, *132*, 9963−9965.
(22) Albertin, L.; Stenzel, M. H.; Barner-Kowollik, C.; Foster, L. J. R.; Davis, T. P. *Macromolecules* **2005**, *38*, 9075−9084.
(23) Albertin, L.; Kohlert, C.; Stenzel, M.; Foster, J. R.; Davis, T. P. *Biomacromolecules* **2004**, *5*, 255−260.
(24) Parry, A. L.; Clemson, N. a; Ellis, J.; Bernhard, S. S. R.; Davis, B. G.; Cameron, N. R. *J. Am. Chem. Soc.* **2013**, *135*, 9362−9365.
(25) Zhang, Q.; Wilson, P.; Anastasaki, A.; McHale, R.; Haddleton, D. M. *ACS Macro Lett.* **2014**, *3*, 491−495.
(26) Zhang, Q.; Collins, J.; Anastasaki, A.; Wallis, R.; Mitchell, D. A.; Becer, C. R.; Haddleton, D. M. *Angew. Chem., Int. Ed.* **2013**, *52*, 4435−4439.
(27) De Coen, R.; Vanparijs, N.; Risseeuw, M. D. P.; Lybaert, L.; Louage, B.; De Koker, S.; Kumar, V.; Grooten, J.; Taylor, L.; Ayres, N.; Van Calenbergh, S.; Nuhn, L.; De Geest, B. G. *Biomacromolecules* **2016**, *17*, 2479−2488.
(28) Zhang, Q.; Anastasaki, A.; Li, G.-Z.; Haddleton, A. J.; Wilson, P.; Haddleton, D. M. *Polym. Chem.* **2014**, *5*, 3876.
(29) Basuki, J. S.; Esser, L.; Duong, H. T. T.; Zhang, Q.; Wilson, P.; Whittaker, M. R.; Haddleton, D. M.; Boyer, C.; Davis, T. P. *Chem. Sci.* **2014**, *5*, 715.
(30) Das, A.; Theato, P. *Chem. Rev.* **2016**, *116*, 1434−1495.
(31) Slavin, S.; Burns, J.; Haddleton, D. M.; Becer, C. R. *Eur. Polym. J.* **2011**, *47*, 435−446.
(32) Krieg, A.; Becer, C. R.; Hoogenboom, R.; Schubert, U. S. *Macromol. Symp.* **2009**, *275*−*276*, 73−81.
(33) Semsarilar, M.; Ladmiral, V.; Perrier, S. *Macromolecules* **2010**, *43*, 1438−1443.
(34) Ladmiral, V.; Mantovani, G.; Clarkson, G. J.; Cauet, S.; Irwin, J. L.; Haddleton, D. M. *J. Am. Chem. Soc.* **2006**, *128*, 4823−4830.
(35) Nurmi, L.; Lindqvist, J.; Randev, R.; Syrett, J.; Haddleton, D. M. *Chem. Commun.* **2009**, 2727−2729.
(36) Haddleton, D. M.; Crossman, M. C.; Dana, B. H.; Duncalf, D. J.; Heming, A. M.; Kukulj, D.; Shooter, A. J. *Macromolecules* **1999**, *32*, 2110−2119.
(37) Mitchell, D. A.; Fadden, A. J.; Drickamer, K. *J. Biol. Chem.* **2001**, *276*, 28939−28945.
(38) Hong, S. Y.; Tobias, G.; Ballesteros, B.; El Oualid, F.; Errey, J. C.; Doores, K. J.; Kirkland, A. I.; Nellist, P. D.; Green, M. L. H.; Davis, B. G. *J. Am. Chem. Soc.* **2007**, *129*, 10966−10967.
(39) Jones, M.-C.; Ranger, M.; Leroux, J.-C. *Bioconjugate Chem.* **2003**, *14*, 774−781.
(40) Limer, A. J.; Rullay, A. K.; Miguel, V. S.; Peinado, C.; Keely, S.; Fitzpatrick, E.; Carrington, S. D.; Brayden, D.; Haddleton, D. M. *React. Funct. Polym.* **2006**, *66*, 51−64.
(41) Vinson, N.; Gou, Y.; Becer, C. R.; Haddleton, D. M.; Gibson, M. I. *Polym. Chem.* **2011**, *2*, 107.
(42) Tanaka, T.; Ishitani, H.; Miura, Y.; Oishi, K.; Takahashi, T.; Suzuki, T.; Shoda, S. I.; Kimura, Y. *ACS Macro Lett.* **2014**, *3*, 1074−1078.
(43) Tanaka, T.; Nagai, M.; Noguchi, M.; Kobayashi, A.; Shoda, S.-i. *Chem. Commun.* **2009**, 3378−3379.
(44) Napoletano, C.; Zizzari, I. G.; Rughetti, A.; Rahimi, H.; Irimura, T.; Clausen, H.; Wandall, H. H.; Belleudi, F.; Bellati, F.; Pierelli, L.; Frati, L.; Nuti, M. *Eur. J. Immunol.* **2012**, *42*, 936−945.

DOI: 10.1021/acs.biomac.7b00228
Biomacromolecules 2017, 18, 1624−1633

(45) Saeland, E.; Van Vliet, S. J.; Bäckström, M.; Van Den Berg, V. C. M.; Geijtenbeek, T. B. H.; Meijer, G. a.; Van Kooyk, Y. *Cancer Immunol. Immunother.* **2007**, *56*, 1225−1236.
(46) Hollingsworth, M. a.; Swanson, B. J. *Nat. Rev. Cancer* **2004**, *4*, 45−60.
(47) Napoletano, C.; Rughetti, A.; Agervig Tarp, M. P.; Coleman, J.; Bennett, E. P.; Picco, G.; Sale, P.; Denda-Nagai, K.; Irimura, T.; Mandel, U.; Clausen, H.; Frati, L.; Taylor-Papadimitriou, J.; Burchell, J.; Nuti, M. *Cancer Res.* **2007**, *67*, 8358−8367.
(48) Carlos, C. a.; Dong, H. F.; Howard, O. M. Z.; Oppenheim, J. J.; Hanisch, F.-G.; Finn, O. J. *J. Immunol.* **2005**, *175*, 1628−1635.
(49) Zizzari, I. G.; Napoletano, C.; Battisti, F.; Rahimi, H.; Caponnetto, S.; Pierelli, L.; Nuti, M.; Rughetti, A. *J. Immunol. Res.* **2015**, *2015*, 1−8.
(50) Becer, C. R.; Gibson, M. I.; Geng, J.; Ilyas, R.; Wallis, R.; Mitchell, D. a.; Haddleton, D. M. *J. Am. Chem. Soc.* **2010**, *132*, 15130−15132.
(51) Lundquist, J. J.; Toone, E. J. *Chem. Rev.* **2002**, *102*, 555−578.
(52) Barner-Kowollik, C.; Davis, T. P.; Stenzel, M. H. *Aust. J. Chem.* **2006**, *59*, 719−727.
(53) Whitfield, R.; Anastasaki, A.; Truong, N. P.; Wilson, P.; Kempe, K.; Burns, J. A.; Davis, T. P.; Haddleton, D. M. *Macromolecules* **2016**, *49*, 8914−8924.
(54) Venkatraman Girija, U.; Furze, C. M.; Gingras, A. R.; Yoshizaki, T.; Ohtani, K.; Marshall, J. E.; Wallis, a K.; Schwaeble, W. J.; El-Mezgueldi, M.; Mitchell, D. a; Moody, P. C. E.; Wakamiya, N.; Wallis, R. *BMC Biol.* **2015**, *13*, 27.
(55) Huang, Z.; Hwang, P.; Watson, D. S.; Cao, L.; Szoka, F. C. *Bioconjugate Chem.* **2009**, *20*, 1667−1672.
(56) Jégouzo, S. A.; Quintero-Martínez, A.; Ouyang, X.; Dos Santos, Á.; Taylor, M. E.; Drickamer, K. *Glycobiology* **2013**, *23*, 853−864.
(57) van Vliet, S. J.; Gringhuis, S. I.; Geijtenbeek, T. B. H.; van Kooyk, Y. *Nat. Immunol.* **2006**, *7*, 1200−1208.
(58) Ashwell, G.; Morell, A. G. *Adv. Enzym. Relat. Areas Mol. Biol.* **2006**, *41*, 99−128.
(59) Weigel, P. H.; Yik, J. H. N. *Biochim. Biophys. Acta, Gen. Subj.* **2002**, *1572*, 341−363.
(60) Bischoff, J.; Lodish, H. F. *J. Biol. Chem.* **1987**, *262*, 11825−11832.
(61) Khorev, O.; Stokmaier, D.; Schwardt, O.; Cutting, B.; Ernst, B. *Bioorg. Med. Chem.* **2008**, *16*, 5216−5231.
(62) Geffen, I.; Wessels, H. P.; Roth, J.; Shia, M. a; Spiess, M. *EMBO J.* **1989**, *8*, 2855−2861.
(63) van Vliet, S. J.; van Liempt, E.; Saeland, E.; Aarnoudse, C. A.; Appelmelk, B.; Irimura, T.; Geijtenbeek, T. B. H.; Blixt, O.; Alvarez, R.; van Die, I.; van Kooyk, Y. *Int. Immunol.* **2005**, *17*, 661−669.
(64) Biessen, E. A. L.; Beuting, D. M.; Roelen, H. C. P. F.; van de Marel, G. A.; Van Boom, J. H.; Van Berkel, T. J. C. *J. Med. Chem.* **1995**, *38*, 1538−1546.
(65) Prakash, T. P.; Yu, J.; Migawa, M. T.; Kinberger, G. A.; Wan, W. B.; Østergaard, M. E.; Carty, R. L.; Vasquez, G.; Low, A.; Chappell, A.; Schmidt, K.; Aghajan, M.; Crosby, J.; Murray, H. M.; Booten, S. L.; Hsiao, J.; Soriano, A.; Machemer, T.; Cauntay, P.; Burel, S. A.; Murray, S. F.; Gaus, H.; Graham, M. J.; Swayze, E. E.; Seth, P. P. *J. Med. Chem.* **2016**, *59*, 2718−2733.
(66) Meier, M.; Bider, M. D.; Malashkevich, V. N.; Spiess, M.; Burkhard, P. *J. Mol. Biol.* **2000**, *300*, 857−865.
(67) Dhande, Y. K.; Wagh, B. S.; Hall, B. C.; Sprouse, D.; Hackett, P. B.; Reineke, T. M. *Biomacromolecules* **2016**, *17*, 830.
(68) Solinas, G.; Schiarea, S.; Liguori, M.; Fabbri, M.; Pesce, S.; Zammataro, L.; Pasqualini, F.; Nebuloni, M.; Chiabrando, C.; Mantovani, a.; Allavena, P. *J. Immunol.* **2010**, *185*, 642−652.
(69) Allavena, P.; Chieppa, M.; Bianchi, G.; Solinas, G.; Fabbri, M.; Laskarin, G.; Mantovani, A. *Clin. Dev. Immunol.* **2010**, *2010*, 547179.

DOI: 10.1021/acs.biomac.7b00228
Biomacromolecules 2017, 18, 1624−1633

Chemical Science

EDGE ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Chem. Sci.*, 2017, **8**, 6974

Manipulation of cytokine secretion in human dendritic cells using glycopolymers with picomolar affinity for DC-SIGN†

Daniel A. Mitchell,*[ab] Qiang Zhang,[c] Lenny Voorhaar,[c] David M. Haddleton,[c] Shan Herath,[d] Anne S. Gleinich,[a] Harpal S. Randeva,[ab] Max Crispin,[e] Hendrik Lehnert,[f] Russell Wallis,[g] Steven Patterson[d] and C. Remzi Becer*[h]

The human C-type lectin DC-SIGN (CD209) is a significant receptor on the surface of dendritic cells (DCs) – crucial components of host defense that bridge the innate and adaptive immune systems. A range of linear glycopolymers, constructed *via* controlled radical polymerization techniques have been shown to interact with DC-SIGN with affinities in the physiologically active range. However, these first generation glycopolymers possess limited structural definition and their effects on DCs were not known. Here we report the development of star-shaped mannose glycopolymers with the aim of targeting the clustered domain arrangement of DC-SIGN and these were shown to bind with picomolar affinity. Increased secretion of IL-10 with simultaneous decrease in secreted IL-12p70 occurred in activated DCs incubated with star-shaped glycopolymers – a cytokine secretion pattern characteristic of wound-healing tissue environments. Incorporating stellar architecture into glycopolymer design could be key to developing selective and very high-affinity therapeutic materials with distinct immunomodulatory and tissue repair potential.

Received 5th April 2017
Accepted 11th August 2017

DOI: 10.1039/c7sc01515a

rsc.li/chemical-science

Introduction

The recognition of repeating molecular patterns is a major function of the human immune system, enabling it to interact with and interpret various biological structures. These include the surfaces of pathogens such as viruses, fungi and bacteria, in addition to host structures such as glycoproteins and apoptotic cells.[1,2] C-type lectins are a major class of pattern recognition molecules in humans that interact with complex carbohydrates such as microbial polysaccharides and oligosaccharides present on human and viral glycoproteins and glycolipids.[3] In particular, the C-type lectin DC-SIGN (dendritic cell-specific intercellular adhesion molecule-3 grabbing nonintegrin; CD209; CLEC4L) is significantly implicated in human disease through its interactions with viral carbohydrates on HIV and mycobacterial lipoarabinomannans (*e.g.* tuberculosis), in addition to its ability to transduce intracellular signaling events and influence dendritic cell responses.[4–7] DC-SIGN is also important in the proper responses to specific apoptotic cell uptake by dendritic cells (DCs).[2] DCs are especially important components of the human immune system owing to their roles as highly efficient surveillance entities and professional antigen presenting cells. DCs are the only cell type with the ability to activate naïve T lymphocytes, making them very powerful bridges between innate and adaptive immunity.[8] NMR analyses of DC-SIGN-oligosaccharide interactions in solution suggest discrete binding modes of the carbohydrate-recognition domain for distinct glycan types that in turn are thought to drive different cell signaling pathways.[9] Furthermore, the distribution of DCs (and DC-SIGN-positive cells such as selected macrophage subpopulations) within discrete regions of tissues of the human body makes them attractive targets for site-specific therapeutic intervention in a broad range of diseases with immune system involvement.[10,11]

Designed glycopolymers have been successfully generated for high affinity interactions with human DC-SIGN.[12–15] This has included the development of size- and sequence-controlled polymers, in addition to conformation-controlled polymers and glycoconjugates, collectively offering a diverse range of

[a]*Clinical Sciences Research Laboratories, University of Warwick, Coventry CV2 2DX, United Kingdom. E-mail: d.mitchell@warwick.ac.uk*
[b]*University Hospital Coventry, Warwickshire NHS Trust, Coventry CV2 2DX, United Kingdom*
[c]*Department of Chemistry, University of Warwick, Coventry CV4 7AL, United Kingdom*
[d]*Chelsea & Westminster Hospital, Imperial College School of Medicine, London SW10 9NH, United Kingdom*
[e]*Glycobiology Institute, University of Oxford, Oxford OX1 3QU, United Kingdom*
[f]*Universität zu Lübeck, 23562 Lübeck, Germany*
[g]*Department of Biochemistry, University of Leicester, Leicester LE1 9HN, United Kingdom*
[h]*School of Engineering and Materials Science, Queen Mary University, London E1 4NS, United Kingdom. E-mail: r.becer@qmul.ac.uk*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c7sc01515a

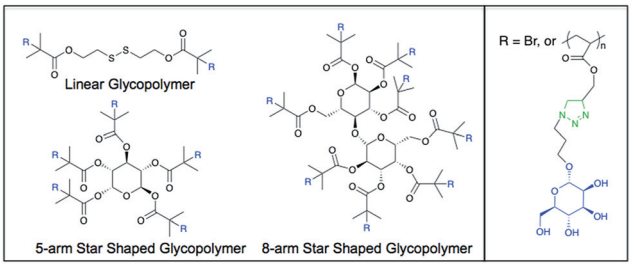


Fig. 1 Representative chemical structures of the mannose glycopolymers. R = Br in the initiator structure.

strategies for DC-SIGN targeting and exploitation. However, the concept of polymer architecture as a rational design strategy for DC-SIGN targeting with these classes of glycopolymer has yet to be employed and furthermore, the impact of advanced glycopolymers on dendritic cell function has yet to be investigated (Fig. 1). Knowledge of the tetrameric structure of DC-SIGN, wherein four mannoside-selective carbohydrate recognition domains (CRDs) are clustered at the top of a coiled-coil stalk projecting from the cell surface,[16,17] allows for the rational construction of novel glycopolymers for further affinity enhancement and potential biological impact. Understanding of the role of DC-SIGN stimulation in dendritic cell responses holds potential for efficient, high affinity ligands to serve as immunomodulators as well as antiviral agents. The main text of the article should appear here with headings as appropriate.

Materials and methods

Instrumentation

NMR spectra were recorded on a Bruker DPX-300 and DPX-400 using deuterated solvents. FTIR spectra were recorded on a Bruker Vector 22 FTIR spectrometer. For SEC a Varian PL 390-LC system was used, equipped with refractive index and viscosimetry detectors, using either $CHCl_3$ or DMF (LiBr 1 g l^{-1}) as the eluent (1 ml min^{-1}). Data analysis was performed using Cirrus software and molecular weights were determined relative to Polymer Labs pMMA standards. The LCST of the glycopolymers was determined by measuring the cloud point of a 5 mg ml^{-1} solution of the polymer in water with a Stanford Research Systems OptiMelt MPA100.

Synthesis of 2-methacrylic acid 3-trimethylsilanyl-prop-2-ynyl ester (TMSMA)

A flask was charged with 3-trimethylsilanyl-prop-2-yn-1-ol (25.0 g, 195 mmol) and triethylamine (41.0 ml, 294 mmol) in THF (250 ml) and cooled with dry ice. A solution of 2-methylacryloyl chloride (23.0 ml, 235 mmol) in THF (50 ml) was added drop wise. The mixture was stirred for 1 hour while cooling with dry ice and stirred overnight at room temperature, during which yellow coloured salts were formed. The mixture was purified on a basic Al_2O_3 column and the solvent was removed under reduced pressure, leaving behind a clear yellow liquid. The yield was 35.0 g (178 mmol, 92%) ^{1}H NMR ($CDCl_3$, 298 K, 300 MHz) δ (ppm) = 0.18 (s, 9H, $Si(CH_3)_3$); 1.96 (m, 3H, $CH_3C{=}CH_2$); 4.76 (s, 2H, OCH_2); 5.62 (m, 1H, C=CHH); 6.17 (m, 1H, C=CHH).

Typical atom transfer radical polymerization procedure

A Schlenk tube was charged with initiator (1.0 mmol), monomer (50 mmol), *N*-(ethyl)-2-pyridylmethanimine ligand (2.0 mmol) and toluene (same volume as monomer). Two different methods have been used to remove the oxygen from the solution, which have been found to be equally effective. The first method was a freeze–pump–thaw cycle, which was executed three times. The second method was bubbling the solution with nitrogen for at least 20 minutes. A second Schlenk tube was charged with Cu(I)Br (1.0 mmol) and a stirring bar and the oxygen was removed by evacuating and subsequently filling the tube with nitrogen three times. The solution was transferred to the second tube using a cannula and a vacuum to start the transfer. The solution was heated to 90 °C while being kept under nitrogen atmosphere for the entire duration of the reaction. Samples were taken periodically using a degassed syringe in case of the kinetic experiments. The polymerization was stopped by cooling the mixture to room temperature and exposing it to oxygen. The mixture was diluted with CH_2Cl_2 (approximately same volume as mixture) and passed through a basic Al_2O_3 column. The polymer was precipitated into a 5 : 1 vol/vol methanol/water mixture (pTMSMA). The solid was isolated by filtration and dried in a vacuum oven overnight. In most experiments with the star-shaped initiators the amount of toluene used was higher to prevent star–star coupling. This was either twice (for 5-arm star) or three times (8-arm star) the volume of the monomer.

Deprotection of pTMSMA

The pTMSMA (0.5 g) was dissolved in THF (35 ml) and glacial acetic acid (0.22 ml, 1.5 equiv. with respect to the TMS groups) was added. The solution was bubbled with nitrogen and cooled with dry ice for at least 15 minutes. A 0.2 M solution of TBAF in THF (50 ml, 4.0 equiv.) was added dropwise with a syringe while stirring the solution and cooling with dry ice. The dry ice was removed after 30 minutes and the mixture was stirred overnight at room temperature. The mixture was concentrated under reduced pressure, purified on a silica column and concentrated again. The polymer was precipitated in petroleum ether cooled with dry ice and acetone. The white powder was isolated by filtration and dried in a vacuum oven overnight. The deprotection was confirmed by the appearance of a peak at 2.5 ppm in the ^{1}H NMR and the disappearance of the 0.2 ppm peak.

Synthesis of α-mannose azide

Sodium azide (7.26 g, 111 mmol), D-(+)-mannose (2.00 g, 11.1 mmol) and triethylamine (15.5 ml, 111 mmol) were dissolved in water (40 ml) and cooled to 0 °C. 2-Chloro-1,3-dimethylimidazolinium chloride (5.61 g, 33.3 mmol) was added and the mixture was stirred for 1 hour at 0 °C. The solvent was removed under reduced pressure and EtOH (40 ml) was added. The solids were removed by filtration and the solution was purified on a long Amberlite IR-120 column, using EtOH as

the eluent. The mixture was checked with FTIR to confirm the removal of all sodium azide ($\nu = 2030\ cm^{-1}$). The solvent was removed under reduced pressure, water (30 ml) was added and the mixture was washed with dichloromethane (5 × 15 ml). The solvent was removed under reduced pressure, water (10 ml) was added and the solution was freeze-dried overnight to give α-mannose azide (1.59 g, 78 mmol, 70%) as an off-white solid. ^{1}H NMR (D_2O, 298 K, 400 MHz) δ (ppm) = 5.45 (1H), 3.95–3.60 (6H). ^{13}C NMR (D_2O, 298 K, 400 MHz) δ (ppm) = 89.72 (C1), 74.65, 69.86, 69.77, 66.40, 60.83 (C6). FT-IR (H_2O) ν (cm^{-1}) = 2112 (N_3).

Mannose azide clicking to propargyl polymers

A solution containing propargyl polymer (100 mg, 0.81 mmol propargyl groups), mannose azide (213 mg, 1.05 mmol) and CuBr (116 mg, 0.81 mmol) in DMSO (10 ml) was bubbled with nitrogen for 20 minutes. Ethyl ligand (216 mg, 1.61 mmol) was added and the solution was bubbled with nitrogen for several more minutes. The solution was subsequently stirred at ambient temperature for two days. The mixture was purified by dialysis (MWCO: 1000) against distilled water for two days, while changing the water at least four times. It was then concentrated under reduced pressure and freeze-dried overnight.

Glycopolymer synthesis and characterization

Before the click reactions could take place, the propargyl groups of the pTMSMA polymers need to be deprotected by removal of the TMS group. This was performed by reacting the polymer with CH_3COOH and TBAF in THF. The deprotection can be followed by ^{1}H NMR (Fig. S1†). The peak of the TMS group at 0.2 ppm disappears and a peak appears at 2.5 ppm for the propargylic proton. After removing the water from the reaction mixture, EtOH was added to filter the unreacted sodium azide from the mixture. However, some of the sodium azide was still soluble in EtOH, as is shown in the FTIR spectrum in Fig. S2.† As sodium azide can form explosive compounds when it is brought into contact with halogenated solvents, it was necessary to remove this from the mixture before continuing with subsequent purification steps. It was found that this last trace of sodium azide could be removed on an Amberlite IR-120 column using EtOH as the eluent. TEA was also removed from the mixture with this column. The remaining DMC was removed by washing the mixture in water with CH_2Cl_2. Mannose azide was attached to each of the different synthesized propargyl polymers using click chemistry. The reaction can be observed using ^{1}H NMR through the appearance of a triazole peak at 8.30 ppm (Fig. S3†). The peaks at 6.0 ppm and 5.4–4.3 ppm are from the mannose. The kinetic plot for the polymerization of TMSMA shows that it takes roughly 15 minutes for the polymerization to initiate (Fig. S5†). After that the conversion is linear up to at least 95%. The increase of M_n with conversion is almost linear within the margin of error. PDI values are all very close together around 1.3. The homopolymerization of TMSMA using the 8-arm initiator was performed in 75% solvent to avoid coupling of the star polymers. The first-order kinetic plot is linear up to 50% conversion. A short initiation period was observed that was also seen in the homopolymerization of TMSMA with the disulfide initiator. The increase in molecular weight is mostly linear with relation to the conversion (PDI ∼ 1.25). The theoretical molecular weight of the synthesized linear, 5 arm and 8 arm glycopolymers have been calculated as 35 700 g mol^{-1}, 37 100 g mol^{-1}, and 39 600 g mol^{-1}, respectively. The measured number average molecular weight of the linear, 5-arm, and 8-arm glycopolymers were found as 41 000 g mol^{-1}, 42 100 g mol^{-1}, and 45 600 g mol^{-1}, respectively.

Protein expression and surface plasmon resonance

Soluble recombinant DC-SIGN was generated in *E. coli* and purified *via* affinity chromatography using mannose–sepharose, as described previously and immobilized[16] on GLM sensor chips *via* amine coupling for use in the Bio-Rad ProteOn XPR36 instrument (Bio-Rad Laboratories, Hercules, USA).[12] Glycopolymer samples were dissolved in running buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 5 mM $CaCl_2$, 0.01% (v/v) Tween-20, 0.01% (w/v) NaN_3) and flowed over sensor chips at a flow rate of 25 μl ml^{-1} at 25 °C. Sensorgram data were collected and reference subtraction performed against a blank channel treated with amine coupling activators and subsequently blocked with 1 M ethanolamine. Kinetic parameters were determined using the Bia-evaluation software (GE Healthcare) utilizing heterogeneous ligand fitting models.

Blood samples

Peripheral blood mononuclear cells (PBMCs) were isolated from blood leukocyte cones (National Blood Transfusion Service, UK) by density gradient centrifugation with Lymphoprep™ 1077 (PAA Laboratories, GmbH, UK) followed by a 50% Percoll (Sigma, UK) gradient to obtain a high-density fraction enriched in lymphocytes and a low-density fraction enriched in monocytes. Fractions were aliquoted and stored in liquid N_2 until required.

Generation of dendritic cells

Monocyte derived dendritic cells (moDCs) were isolated from the monocyte rich fraction by magnetic bead separation using anti-CD14 coated magnetic beads (Miltenyi Biotec). Purified monocytes were cultured in complete medium (RPMI containing 10% FBS, 100 IU ml^{-1} penicillin, 0.1 mg ml^{-1} streptomycin, and 2 mM L-glutamine; Sigma-Aldrich) supplemented with GM-CSF (100 ng ml^{-1}) and IL-4 (50 ng mLl^{-1}; R&D Systems), this being replenished every other day. On day 7, DCs were harvested, washed twice and resuspended in serum-free RPMI-1640 prior to staining for flow cytometry.

Flow cytometry

Dendritic cells were harvested, washed and stained with either 5-arm polymer-FITC, 8-arm polymer-FITC, a linear polymer-FITC or gp120-FITC (generated from soluble recombinant gp120 produced in HEK cells as previously described) at the concentrations indicated in the results. To assess the ability of

the polymers to bind to the gp120 site, DCs were incubated with unlabeled polymer (at the concentrations indicated in the results) for 45 minutes at room temperature and then treated with gp120-FITC (1 μg ml^{-1}) for a further 30 minutes at room temperature. To determine binding of gp120 to $CD4^+$ cells, DCs and lymphocytes were firstly incubated with unlabeled gp120 at 1 μg ml^{-1} for 45 min at room temperature and then stained with anti-CD4-APC, anti-CD3-Percp, and gp120-FITC for a further 30 minutes. Following all the incubations, cells were washed and fixed with BD stabilizing fixative (BD Biosciences). Unstained DCs were used as a control. One hundred thousand cells were acquired within a live gate using a 3-laser configuration LSRII flow cytometer (BD Biosciences, USA), and analysed by FlowJo (Tree Star Inc, USA).

Cytokine assays

Sets of dendritic cell cultures, derived from four separate donors, were incubated with 1 μg ml^{-1} gp120 and also defined quantities of the three species of glycopolymer (8-arm, 5-arm and linear; at 1, 3 and 10 μg ml^{-1}). After 24 hours, one group of DC cultures was activated with 1 μg ml^{-1} lipopolysaccharide (LPS; Sigma, Poole, UK) and 20 ng ml^{-1} interferon gamma (IFN-γ), the other left unstimulated. Supernatants (100 μl) from the assorted cultures were recovered after a further 24 hours *via* centrifugation at 2000 rpm. Samples were diluted 1 : 1 in DMEM cell culture medium without supplements and duplicate measurements were made using Luminex bead immunoassay. A panel of nine cytokines was analyzed simultaneously using the commercial Bio-Rad Th1 and Th2 kit within the Bio-plex 200 Multiplex System in accordance with manufacturer instructions (Bio-Rad Laboratories, Hercules, USA). Quantitative readouts were obtained for the following cytokines: tumor necrosis factor alpha (TNF-α), IFN-γ, interleukins (IL)-2, IL-4, IL-5, IL-10, IL-12p70, IL-13, and granulocyte/monocyte-colony stimulating factor (GM-CSF).

Research using human material was carried out in accordance with institutional guidelines issued by the University of Warwick and Imperial College London, and with the principles expressed in the Declaration of Helsinki. Ethical approval was obtained from the Riverside Research Ethics Committee and informed written consent was obtained prior to patient blood collection. Data from human-derived material were processed anonymously.

Results and discussion

Carbohydrates in the form of oligosaccharides and polysaccharides represent a large, crucial and exciting family of molecules in the development of novel therapeutics. Uncovering the roles of complex carbohydrates in human physiology and disease coincides with the advancement of our understanding of carbohydrate-binding receptors such as DC-SIGN. Initially identified as a molecule exploited by HIV-1 in order to promote viral trafficking and within-host survival, it has since emerged recently that DC-SIGN is central to important physiological processes such as healthy responses during apoptotic cell clearance and foeto-maternal tolerance.[2,18] Studies into the impact of glycan ligands for DC-SIGN on immune cells indicate that this receptor is capable of driving anti-inflammatory and immunosuppressive cell signaling with major effects on gene expression and cellular responses.[6,7] Whilst undesirable in the context of certain infections, these responses are beneficial in the active processes of wound healing and the resolution of inflammatory episodes. It is becoming increasingly clear that these tissue repair and protection mechanisms are highly active and require advanced immune system involvement. Of significance is the association of molecules such as DC-SIGN with tissue environments such as placenta, uterus and gut, where tight control of tissue growth & stability and blood supply in the presence of abundant non-self material is essential.[11]

Understanding the structure and function of DC-SIGN allows us to design and prepare advanced synthetic ligands for this molecule and investigate opportunities for exploiting its anti-inflammatory properties. The strategy of generating defined star-shaped glycopolymers (GPs) to achieve enhanced binding to the tetrameric configuration of DC-SIGN has been very successful, showing increases in affinity exceeding an order of magnitude, in addition to demonstrating strong interactions with DCs. We speculate that the star-shaped architecture spans the four CRDs within the DC-SIGN tetramer, increasing the strength of the interaction and potentially influencing the consequences of DC-SIGN engagement. It is assumed that DC-SIGN can only transduce signals when clustered in multimeric complexes with multiple glycan ligands expressed on a large surface area, although small ligands such as Lewis-x oligosaccharides have been shown to drive signaling and gene expression responses in DCs *via* DC-SIGN engagement. Furthermore, NMR analyses of DC-SIGN–oligosaccharide interactions in solution suggest discrete binding modes of the carbohydrate-recognition domain for distinct glycan types that in turn are thought to drive different cell signaling pathways. Therefore, the consequences of DC-SIGN engagement by glycopolymers could be influenced and even tuned by variations in sugar composition, architecture and size.

Star-shaped GPs bind to recombinant DC-SIGN with subnanomolar affinity. Interaction analysis in real time *via* surface plasmon resonance (SPR) demonstrated distinct and strong binding between immobilized DC-SIGN and all of the GPs examined (Fig. 2), consistent with previous studies.[12] As expected, the star-shaped GPs showed greater affinity compared with the linear GP of equal molecular weight and dissociation rates (k_{off}) were markedly slow, in keeping with strong multivalent interactions with the oligomeric DC-SIGN protein. Apparent K_D values of 65.1 pM and 72.5 pM were determined for the star-shaped 5-arm and 8-arm glycopolymers, respectively (Table 1).

Star-shaped glycopolymers show improved and dose-dependent binding to human dendritic cells

To determine whether the polymers bound directly to dendritic cells (DCs), DCs were stained with FITC-labeled polymers at a range of polymer concentrations. Fig. 3A–C shows that all GPs

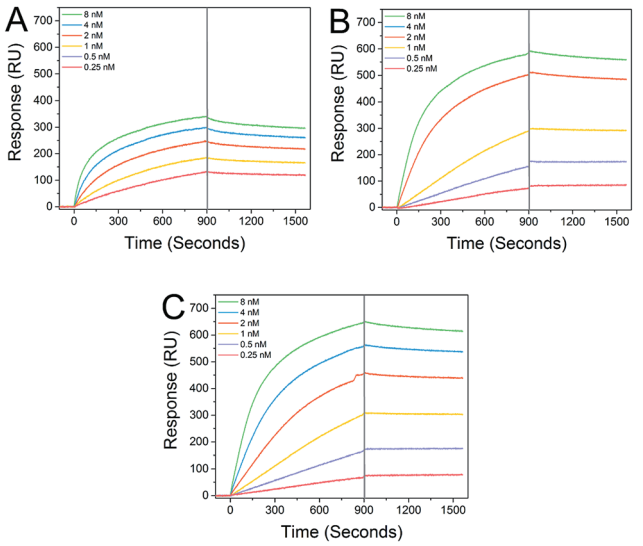


Fig. 2 Real-time analysis of GP interactions with DC-SIGN. SPR sensorgrams were collected at 25 °C and a flow rate of 25 μL min^{-1} and demonstrate clear interactions between linear glycopolymer **P1** (Panel A), 5-arm star glycopolymer **P2** (Panel B) and 8-arm star glycopolymer **P3** (Panel C) and immobilized DC-SIGN extracellular domain. The sensorgram curves illustrate slow dissociation, especially of the 5-arm and 8-arm GPs during the wash phase (indicated by the guideline at 900 s).

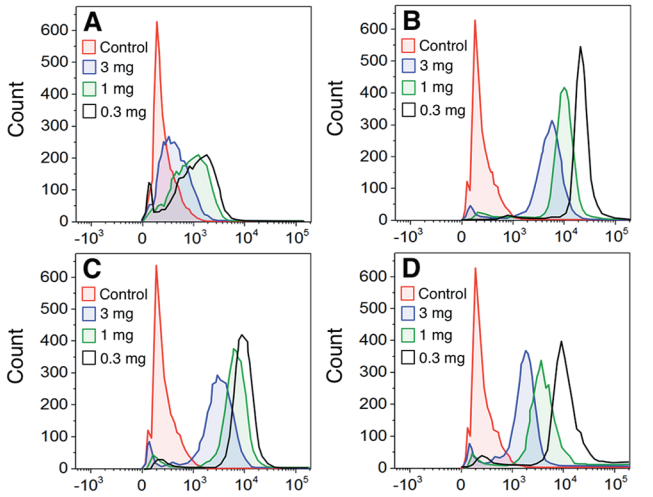


Fig. 3 Concentration-dependent glycopolymer binding to DCs. Dendritic cells were labeled with the indicated concentration of FITC-labeled glycopolymers/glycoprotein (Panel A – Linear **P1**; Panel B – 5-arm star **P2**; Panel C – 8-arm star **P3**; and Panel D – gp120) and analyzed by flow cytometry. The data are representative of one of 3 donors.

bound to DCs to some degree with the 5-arm and 8-arm glycopolymers binding better than the linear polymer. We also assessed the binding of gp120 to DCs and found that binding followed a similar profile to the star-shaped glycopolymers (Fig. 3D). In a direct comparison between the polymers and gp120, we found that the 5-arm and 8-arm GPs bound with a similar intensity as compared to the gp120 at the highest dose of 10 μg ml^{-1} (see ESI†).

In addition to assessing the intrinsic binding of GPs to DCs we also determined whether they exerted any measurable influence on the activation status of the treated DCs by analyzing their cell-surface markers. We found that following 24 hours of exposure to the polymers generated here, there was no significant change in DC expression of the canonical activation markers CD80, CD83, CD86 or HLA-DR when examining cells from three separate donors.

There is a likelihood that these mannose glycopolymers also interact with dendritic cell and macrophage subpopulations that express the mannose receptor (CD206), in addition to specialised endothelial cells that express receptors such as DC-SIGNR (CD299). Where the very high affinity interactions between the star-shaped glycopolymers and DC-SIGN could possess significant selectivity would be in discrete tissue spaces such as inflamed synovium associated with rheumatoid arthritis, where salient subpopulations of DC-SIGN-positive macrophages are upregulated.[19]

Table 1 Binding kinetics and apparent K_D values of the linear, 5-arm star-shaped, and 8-arm star-shaped glycopolymers

Run	Type	k_{on} (M^{-1} s^{-1})	k_{off} (s^{-1})	K_D (M)
P1	Linear	1.24×10^{6}	1.34×10^{-4}	1.08×10^{-10}
P2	5-Arm star	1.01×10^{6}	7.33×10^{-5}	7.25×10^{-11}
P3	8-Arm star	0.93×10^{6}	6.03×10^{-5}	6.51×10^{-11}

Competition of glycopolymers with gp120 for binding sites on dendritic cells due to cellular CD4 expression

To determine whether the polymers competed with the gp120 for the same cellular binding site, we incubated DCs with unlabeled polymers and then stained the cells with FITC-labeled gp120 (Fig. 4). Our results show that although there was a slight shift in binding when polymer was present, none of the polymers significantly abrogated gp120 binding.

To assess whether this was due to the gp120 using more than one binding partner on cell surfaces, we looked at determining the expression of the major gp120 ligand – CD4 – on both DCs and T cells. Our results show that gp120 is able to bind to CD4 molecules on both DCs and T cells – a site to which gp120 binding is not carbohydrate-dependent (Fig. 5). Consequently our results demonstrate that the polymers bind to separate sites on cells from those utilized and possibly preferred by gp120.

The flow cytometry studies involving HIV gp120 interactions with DCs raise a very important issue in the use of glycopolymers as agents for viral blockade. The presence of CD4 on moDCs (and also native peripheral DC populations) serves as a very effective adhesion route for HIV, such that polymer-based barrier preparations could be compromised *in vivo*. This would be consistent with *in vivo* findings that DC-SIGN is associated with parenteral but not mucosal trafficking of HIV.[20] However, it is yet to be known whether the presence of glycopolymers could affect HIV infection of $CD4^{+}$ T cells in trans and whether glycopolymers would be effective in blocking DC interactions with other dangerous glycosylated viruses that are known to interact

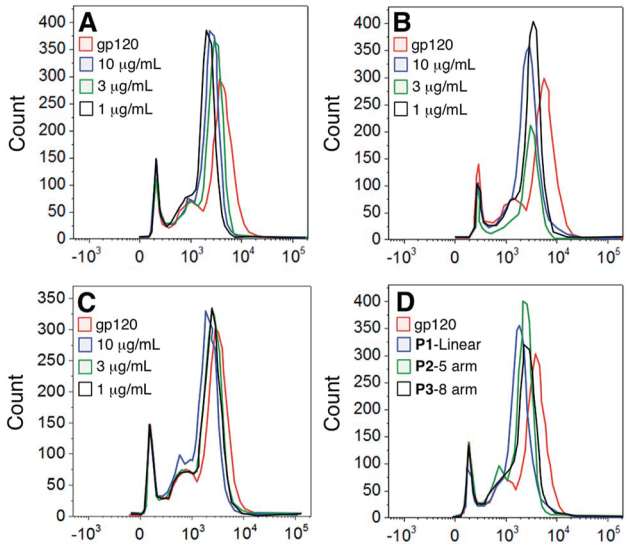


Fig. 4 Competition of polymer *vs.* gp120 binding to dendritic cells. Dendritic cells were firstly incubated with the indicated concentration of unlabeled polymer (Panel A – Linear **P1**; Panel B – 5-arm star **P2**; Panel C – 8-arm star **P3**; Panel D – summary of polymers at 10 μg ml^{-1}) and then probed with FITC gp120. Cells were analyzed by flow cytometry and the data represent one of 3 donors.

with DC-SIGN such as hepatitis C virus (HCV), Ebola and Dengue.

Glycopolymers modulate cytokine production by activated DCs

Cytokine analysis was carried out on supernatants recovered from DCs exposed to glycopolymers and consensus activation stimuli (LPS plus IFN-γ). Prior to supernatant harvesting, the DCs appeared healthy and normal in the presence of glycopolymers, suggesting that these materials were tolerated and not markedly toxic at the quantities used (1–10 μg ml^{-1}). Measured cytokine secretion levels were negligible in all of the DC cultures that were not stimulated with LPS and IFN-γ, suggesting that the GPs themselves do not invoke major secreted DC activation responses. In DC cultures activated with LPS and IFN-γ, key cytokines such as TNF-α were up-regulated in all cultures, as expected, and the exposure of DCs to star-shaped GPs led to significant changes in the secretion of two cytokines – IL-10 and IL-12p70. Levels of secreted IL-10 were significantly increased only in supernatants from activated DCs exposed to the 8-arm and 5-arm GPs, and only at the highest concentration used (10 μg ml^{-1}; Fig. 6A). Equivalent quantities of the linear GP did not invoke significant changes in IL-10 secretion. Levels of IL-12p70 were significantly reduced by all three species of GP in a dose dependent manner (Fig. 6B–D). Interestingly, the linear GP showed an effect in reducing IL-12p70 secretion (Fig. 6B). This suggests that whilst showing lower levels of binding to DC-SIGN and to DCs, linear GPs may still influence certain DC responses.

The changes in cytokine secretion by DCs invoked by the star-shaped polymers are consistent with those observed in mammalian wound healing. IL-10 is a potent immunosuppressive cytokine with broad effects on inflammation and its mediators.[21] IL-12p70 is a pro-inflammatory cytokine and a potent inhibitor of angiogenesis (blood vessel growth).[22] The combination of IL-10 increase and IL-12 decrease represents a rapid and effective means of reducing inflammation and restoring organized blood flow to a site of tissue injury. In general, qualitative changes in secretion of these two cytokines are in keeping with the consequences of DC-SIGN engagement by biologically-sourced mannosylated ligands. However, a key difference is that in the case of the glycopolymers, the chemical definition is very precise. In contrast, biological high mannose ligands such as gp120, yeast mannan and mycobacterial

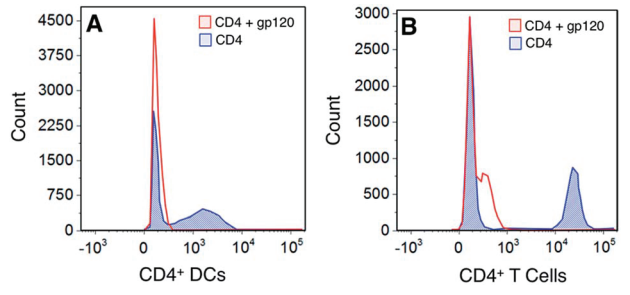


Fig. 5 Competition binding of gp120 against CD4. Dendritic cells and lymphocytes were first incubated with an unlabeled gp120 and then stained with an anti-CD4 antibody. Cells were analyzed by flow cytometry and the data represents one of 2 donors.

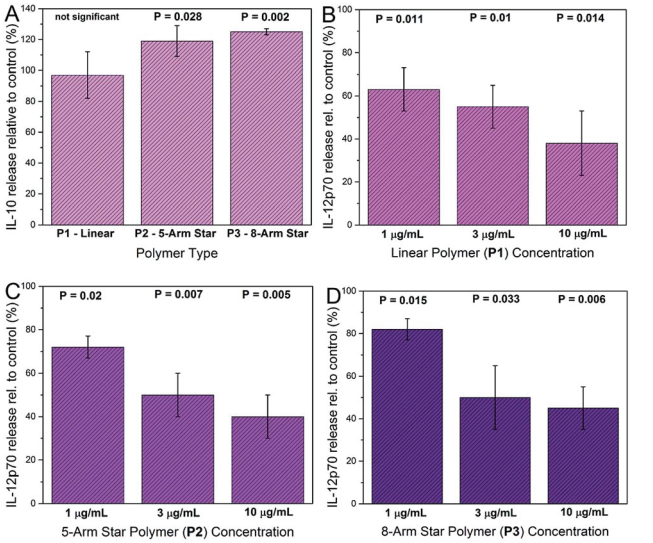


Fig. 6 Effects of glycopolymers on cytokine secretion by activated DCs. Cells were incubated with GPs before activation with LPS and IFN-γ and multiplex cytokine assay. IL-10 secretion was significantly increased in DC cultures from 4 donors incubated with 8-arm star **P3** and 5-arm star **P2** GPs at 10 μg ml^{-1} compared with controls, but not with linear GP **P1** (Panel A). IL-12p70 secretion was significantly decreased compared to controls in cells exposed to increasing concentrations of linear GP **P1** (Panel B), 5-arm star GP **P2** (Panel C) and 8-arm star GP **P3** (Panel D) (all *P* values calculated from paired *t*-tests).

lipoarabinomannan are notoriously heterogeneous.[23–25] Furthermore, the low cost of polymer precursors and the simple synthesis workflow make the glycopolymers described here an attractive and affordable means of generating abundant, water-soluble material with advanced immunological activity.

Conclusions

To summarize, we report the synthesis and characterization of highly water-soluble advanced glycopolymer sets of defined size and shape designed to interact profoundly with the key human lectin DC-SIGN. Biophysical studies demonstrate remarkable glycopolymer-DC-SIGN interactions with affinities in the pico-molar range and positive binding to human monocyte-derived dendritic cells. Star-shaped glycopolymers invoke significant changes in the production of two key cytokines, IL-10 and IL-12p70, when incubated with dendritic cells at pharmacologically realistic concentrations. Importantly, IL-10 levels rise and IL-12 levels fall in response to the glycopolymers and this cytokine release pattern is reminiscent of tissue environments associated with active wound healing, anti-inflammation, and also healthy pregnancy at the feto-maternal interface. In our experiments, however, the glycopolymers do not affect the ability of HIV gp120 to interact with monocyte-derived dendritic cells – previously a major direction for both DC-SIGN and glycopolymer research. We propose novel roles for glycopolymers of this kind, shifting the focus away from HIV prophylaxis towards the treatment of conditions that require immune modulation that promotes wound healing and inflammatory resolution. Examples of such conditions could include burns & chronic skin lesions, in addition to chronic inflammatory diseases with immunological involvement such as rheumatoid arthritis, multiple sclerosis and Crohn's disease.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was supported by the European Commission (EU-ITN EuroSequences Proposal No: 642083 to C. R. B.) and EPSRC (EP/P009018/1). ASG is supported by a Studentship Award from Coventry General Charities, UK (Registered Charity Number 216235).

Notes and references

1 Y. van Kooyk, A. Engering, A. N. Lekkerkerker, I. S. Ludwig and T. B. Geijtenbeek, *Curr. Opin. Immunol.*, 2004, **16**, 488–493.
2 S. K. Pathak, A. E. Skold, V. Mohanram, C. Persson, U. Johansson and A. L. Spetz, *J. Biol. Chem.*, 2012, **287**, 13731–13742.
3 M. E. Taylor and K. Drickamer, *Methods Enzymol.*, 2003, **363**, 3–16.
4 T. B. Geijtenbeek, D. S. Kwon, R. Torensma, S. J. van Vliet, G. C. van Duijnhoven, J. Middel, I. L. Cornelissen, H. S. Nottet, V. N. KewalRamani, D. R. Littman, C. G. Figdor and Y. van Kooyk, *Cell*, 2000, **100**, 587–597.
5 T. B. Geijtenbeek, S. J. Van Vliet, E. A. Koppel, M. Sanchez-Hernandez, C. M. Vandenbroucke-Grauls, B. Appelmelk and Y. Van Kooyk, *J. Exp. Med.*, 2003, **197**, 7–17.
6 S. I. Gringhuis, J. den Dunnen, M. Litjens, M. van der Vlist and T. B. Geijtenbeek, *Nat. Immunol.*, 2009, **10**, 1081–1088.
7 S. I. Gringhuis, J. den Dunnen, M. Litjens, B. van Het Hof, Y. van Kooyk and T. B. Geijtenbeek, *Immunity*, 2007, **26**, 605–616.
8 R. M. Steinman, *Cell*, 2000, **100**, 491–494.
9 K. Pederson, D. A. Mitchell and J. H. Prestegard, *Biochemistry*, 2014, **53**, 5700–5709.
10 E. J. Soilleux, L. S. Morris, B. Lee, S. Pohlmann, J. Trowsdale, R. W. Doms and N. Coleman, *J. Pathol.*, 2001, **195**, 586–592.
11 E. J. Soilleux, L. S. Morris, G. Leslie, J. Chehimi, Q. Luo, E. Levroney, J. Trowsdale, L. J. Montaner, R. W. Doms, D. Weissman, N. Coleman and B. Lee, *J. Leukocyte Biol.*, 2002, **71**, 445–457.
12 C. R. Becer, M. I. Gibson, J. Geng, R. Ilyas, R. Wallis, D. A. Mitchell and D. M. Haddleton, *J. Am. Chem. Soc.*, 2010, **132**, 15130–15132.
13 J. Huang, Q. Zhang, G. Z. Li, D. M. Haddleton, R. Wallis, D. Mitchell, A. Heise and C. R. Becer, *Macromol. Rapid Commun.*, 2013, **34**, 1542–1546.
14 Q. Zhang, J. Collins, A. Anastasaki, R. Wallis, D. A. Mitchell, C. R. Becer and D. M. Haddleton, *Angew. Chem.*, 2013, **52**, 4435–4439.
15 Q. Zhang, L. Su, J. Collins, G. Chen, R. Wallis, D. A. Mitchell, D. M. Haddleton and C. R. Becer, *J. Am. Chem. Soc.*, 2014, **136**, 4325–4332.
16 D. A. Mitchell, A. J. Fadden and K. Drickamer, *J. Biol. Chem.*, 2001, **276**, 28939–28945.
17 H. Feinberg, Y. Guo, D. A. Mitchell, K. Drickamer and W. I. Weis, *J. Biol. Chem.*, 2005, **280**, 1327–1335.
18 P. Hsu, B. Santner-Nanan, J. E. Dahlstrom, M. Fadia, A. Chandra, M. Peek and R. Nanan, *Am. J. Pathol.*, 2012, **181**, 2149–2160.
19 T. Sokka, E. Krishnan, A. Hakkinen and P. Hannonen, *Arthritis Rheum.*, 2003, **48**, 59–63.
20 M. P. Martin, M. M. Lederman, H. B. Hutcheson, J. J. Goedert, G. W. Nelson, Y. van Kooyk, R. Detels, S. Buchbinder, K. Hoots, D. Vlahov, S. J. O'Brien and M. Carrington, *J. Virol.*, 2004, **78**, 14053–14056.
21 A. King, S. Balaji, L. D. Le, T. M. Crombleholme and S. G. Keswani, *Adv. Wound Care*, 2014, **3**, 315–323.
22 A. Naldini and F. Carraro, *Curr. Drug Targets: Inflammation Allergy*, 2005, **4**, 3–8.
23 T. Mizuochi, M. W. Spellman, M. Larkin, J. Solomon, L. J. Basa and T. Feizi, *Biochem. J.*, 1988, **254**, 599–603.
24 C. E. Ballou, *Adv. Enzymol. Relat. Areas Mol. Biol.*, 1974, **40**, 239–270.
25 D. Chatterjee and K. H. Khoo, *Glycobiology*, 1998, **8**, 113–120.

