

**Development of a method to study glycoproteins modified with  
ribitol-5-phosphate**

**Angelo Lopez**

**PhD**

**University of York**

**Department of Chemistry**

**October 2023**

## Abstract

Alpha-dystroglycan ( $\alpha$ DG) is a heavily glycosylated protein that is part of the dystrophin-glycoprotein complex present on the membrane of mammalian cells. The function of this complex is to give rigidity to a tissue through binding laminin in the extracellular matrix.  $\alpha$ DG has a unique O-mannosyl glycan, essential for its function, consisting of a phosphorylated trisaccharide core M3 elongated with two ribitol-5-phosphate (Rbo5P) residues. The terminal Rbo5P acts as a linker to which a long chain of alternating xylose and a glucuronic acid residue, named matriglycan, is attached. Rbo5P is currently found in mammalian cells uniquely on  $\alpha$ DG. Enzymes responsible for the synthesis of the alditol donor CDP-Rbo and for the Rbo5P transfer onto the O-mannosyl glycans have been identified.

The driving force of the research presented here is the hypothesis that other mammalian proteins might also have glycans with Rbo5P and share a similar function to  $\alpha$ DG. This work aimed to identify novel Rbo5P glycoconjugates through chemical tagging of Rbo5P residues and subsequent enrichment of tagged glycoproteins from cultured cells. A method for tagging, enriching and analysing by mass spectrometry glycoproteins modified with Rbo5P has been developed and tested. The method consists of oxidising glycoproteins bearing Rbo5P with sodium periodate to generate aldehydes, which can then be conjugated with aminoxy-biotin. The biotin tag was exploited for the enrichment of Rbo5P glycoconjugates from mammalian cell lysates using streptavidin pulldown. Proteins recovered this way were then digested into peptides and analysed by liquid chromatography coupled to mass spectrometry. Two putative Rbo5P-containing glycoproteins were identified using this method. Moreover, the chemical tagging to recombinant  $\alpha$ DG revealed a new glycosylation site for the Rbo5P modification.

## **Author's Declaration**

I declare that this Thesis is a presentation of original work, and I am the sole author. I received guidance and support in the process of writing this Thesis from my supervisors, Prof Jane Thomas-Oates and Dr Lianne Willems. All the experiments described in this Thesis were performed by the author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references.

## **Acknowledgements**

I want to thank first my supervisors Prof Jane Thomas-Oates and Dr Lianne Willems. They have helped me grow not only as a scientist but also as a person. I also want to thank Prof Daniel Ungar for his feedback throughout my PhD.

I would like to thank Ed Bergström and Dr Adam Dowle for their technical support. Thank you to everyone in JTO and Willems' groups as well as in the Chemical Biology group for the fun shared in the last four years. Special thanks go to Karl Heaton. He has been a teacher, a friend and, when requested, a therapist!

My family, despite the distance, is always just a phone call away. Thanks to my parents for their support and for teaching me the right values. I want to thank my friends for laughing with me when I needed it the most. Finally, thank you, Marco, for your smiles, which always let me know everything will be alright.

# Table of contents

## Chapter 1 Introduction

1.1 Glycosylation .....	8
1.2 O-mannosylation .....	11
1.3 O-mannosylated proteins .....	17
1.4 $\alpha$ -Dystroglycan .....	19
1.4.1 Dystroglycanopathies .....	22
1.5 Glycoproteomics .....	23
1.5.1 Glycoprotein and glycopeptide enrichment .....	23
1.5.2 Metabolic oligosaccharide engineering (MOE) .....	24
1.5.3 Exploiting chemical properties of carbohydrates for enrichment .....	26
1.5.4 Mass spectrometric analysis .....	28
1.6 Research aims .....	33

## Chapter 2 Method development for tagging, enriching, and analysing glycoproteins modified with ribitol-5-phosphate

2.1 Introduction .....	34
2.2 Aims .....	37
2.3 Materials and Methods .....	38
2.3.1 Sodium periodate and biotinylation .....	38
2.3.2 In-gel digestion .....	38
2.3.3 Streptavidin pulldown .....	39
2.3.4 Western blotting .....	39
2.3.5 Mass spectrometric analysis .....	39
2.3.6 Data analysis .....	40
2.3.7 Mass spectra annotation .....	41
2.4 Results .....	42
2.4.1 Sodium periodate oxidises NeuAc on glycoproteins .....	42
2.4.2 Aminoxy-biotin conjugates to oxidised NeuAc on glycoproteins .....	49
2.4.3 Sodium periodate treatment and biotin conjugation are selective for NeuAc .....	54
2.5 Conclusions .....	56

**Chapter 3** Development of a method for the overexpression and chemical tagging of recombinant  $\alpha$ DG

3.1 Introduction.....	57
3.2 Aims .....	59
3.3 Materials and Methods.....	60
3.3.1 Plasmid cloning .....	60
3.3.1.1 p $\alpha$ DG-MycFLAG .....	60
3.3.1.2 p $\alpha$ DG-S654A-MycFLAG .....	60
3.3.1.3 pFLAG $\alpha$ DGh and pFLAG $\alpha$ DGr .....	60
3.3.2 Cell culture and lysate preparation .....	61
3.3.3 Transfection and purifications of constructs .....	61
3.3.4 Western blotting .....	62
3.3.5 Bioorthogonal ligation with aminooxy-biotin .....	63
3.3.6 In-gel digestion .....	63
3.3.7 Mass spectrometric analysis .....	63
3.3.8 Data analysis.....	65
3.4 Results .....	65
3.4.1 Expression of endogenous glycosylated $\alpha$ DG in mammalian cells .....	65
3.4.2 Expression of recombinant $\alpha$ DG in mammalian cells.....	66
3.4.2.1 Cloned p $\alpha$ DG-MycFLAG is transfected in HEK293T and HEK293F cells.....	66
3.4.2.2 Cloned p $\alpha$ DG-S654A-MycFLAG is transfected in HEK293F cells.....	72
3.4.2.3 FLAG-tagged $\alpha$ DG is successfully expressed in HEK293T cells.....	75
3.4.3 Analysis of recombinant $\alpha$ DG by mass spectrometry .....	78
3.4.4 Rbo5P is detected in recombinant $\alpha$ DG .....	83
3.5 Conclusions.....	91

**Chapter 4** Towards the identification of novel proteins modified with ribitol-5-phosphate

4.1 Introduction .....	92
4.2 Aims.....	95
3.3 Materials and Methods .....	96
4.3.1 Bioinformatic search .....	96
4.3.2 Cell lysate preparation .....	96
4.3.3 Sialidase treatment .....	97

4.3.4 Western blotting .....	97
4.3.5 Bioorthogonal ligation with aminoxy-biotin .....	97
4.3.6 Streptavidin pulldown .....	98
4.3.7 In-gel digestion .....	98
4.3.8 Mass spectrometric analysis .....	98
4.3.9 Data analysis.....	98
4.3.10 Sequence alignment.....	99
4.4 Results .....	100
4.4.1 Bioinformatic search of published proteomic data .....	100
4.4.2 Tagging Rbo5P in cell lysates.....	102
4.5 Conclusions .....	111
<b>Chapter 5</b> Discussion and future research .....	112
5.1 Rbo5P-bearing glycopeptides are not identified by untargeted strategies .....	112
5.2 Chemical tagging of glycoproteins allows their selective enrichment and mass spectrometric identification.....	113
5.3 Rbo5P is identified from full-length $\alpha$ DG.....	114
5.4 Putative Rbo5P-containing proteins are identified in cell lysate after oxidation and biotinylation.....	116
5.5 Candidates for follow-up studies and future recommendations .....	117
<b>List of abbreviations</b> .....	119
<b>References</b> .....	123

# Chapter 1: Introduction

## 1.1 Glycosylation

The central dogma of molecular biology states that genetic information flows in one direction from DNA to RNA to proteins (Crick, 1958). In an organism, proteins represent the genome's phenotype which is subject to change in time and space (Lamond et al., 2012). The greatest challenge of biologists nowadays is identifying as-yet uncharacterised proteins and understanding their function and interactions. The identification of a protein in a biological sample is only the first step towards its full characterisation. To study the biological activity of a protein or its role within the cell, consideration should also be given to a protein's post-translation modifications (PTMs). More than 500 different PTMs are currently known to covalently modify amino acids in proteins (Keenan et al., 2021). These modifications affect protein characteristics such as enzyme function (Ryšlavá et al., 2013), protein folding (del Monte and Agnetti, 2014) and molecular trafficking (Karve and Cheema, 2011). Phosphorylation of serine or threonine residues and acetylation of lysine residues are among the most experimentally observed modifications of amino acids (Khoury et al., 2011). This does not necessarily imply that these PTMs are more abundant than others but could also suggest that certain PTMs are easier to study and identify. Glycosylation is an example of a PTM which is predicted to affect about 20% of proteins (Khoury et al., 2011; Z. Li et al., 2021), yet is challenging to elucidate experimentally. Glycosylation refers to a covalent bond, formed during or after translation of the protein, between a mono/oligo/polysaccharide, or glycan, and an amino acid. Glycosylation has a structural function on membrane and secreted proteins since it protects them from protease activity (Arike and Hansson, 2016; Corfield, 2015), as well as on newly synthesised proteins helping the correct folding to take place and maintaining their conformation and solubility (Shental-Bechor and Levy, 2008). Glycans are also used to modulate interactions between proteins, acting as on/off switches, for example in the case of growth factor receptors where glycosylation 'activates' them and enables them to bind their target (Cole, 2010). Moreover, glycoproteins on the cell surface mediate cell recognition and signalling (Rudd et al., 2001),

The two most common types of glycosylation are *N*-linked and *O*-linked, referring to the type of bond by which the glycan is attached to the protein. *N*-glycosylation is characterised by the bond between the glycan and the amidic nitrogen of an asparagine residue; the *N*-glycan has a common core structure (two *N*-acetylglucosamine (GlcNAc) and three mannose (Man) residues) which can be further modified and extended in a variety of different ways. In eukaryotes, *N*-glycan biosynthesis is initiated in the endoplasmic reticulum (ER), where the first steps of processing occur, followed by further elongation and branching of the glycans in the Golgi apparatus. More

specifically, a precursor made of three glucose (Glc), nine Man and two GlcNAc residues (Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub>) is built on a lipid-like molecule (dolichol-phosphate (Dol-P)) on the cytosolic side of the ER membrane. Glycosyltransferases (GTs) are responsible for the precursor synthesis using an activated monosaccharide donor (sugar nucleotide) to create a new covalent bond between the monosaccharide and either a lipid (the first GlcNAc added to Dol-P) or another monosaccharide (Schenk et al., 2001). The glycan precursor Man<sub>5</sub>GlcNAc<sub>2</sub> is flipped into the ER lumen, extended to have Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub> and transferred ‘en-bloc’ from the Dol-P-linked donor onto a protein by the oligosaccharyltransferase, after which it is further processed in the lumen of the ER by glycosidases (GHs), which remove monosaccharides from the glycan. The newly synthesised glycoprotein undergoes a quality control cycle to ensure the correct folding and eventually moves to the Golgi, where other enzymes modify the *N*-glycan (Herscovics, 1999).

O-glycosylation is characterised by the bond between the glycan and the hydroxyl group of a serine, threonine or, less often, tyrosine or hydroxyproline residue. O-glycans are directly synthesised on the target protein by GTs, which create a covalent glycosidic bond between a monosaccharide and the protein, using a nucleotide-activated sugar as the donor; once added to the protein, the monosaccharide can act in turn as an acceptor and be further extended by GTs making more complex glycans. In O-glycoproteins, the first monosaccharide of a glycan commonly defines its glycosylation class. O-GalNAcylation (O-linked *N*-acetylgalactosamine) initiates in the Golgi and involves the synthesis of O-GalNAc glycans characterised by different core structures (Figure 1.1), which can be extended with galactose (Gal), GlcNAc, fucose (Fuc), or sialic acid (NeuAc).

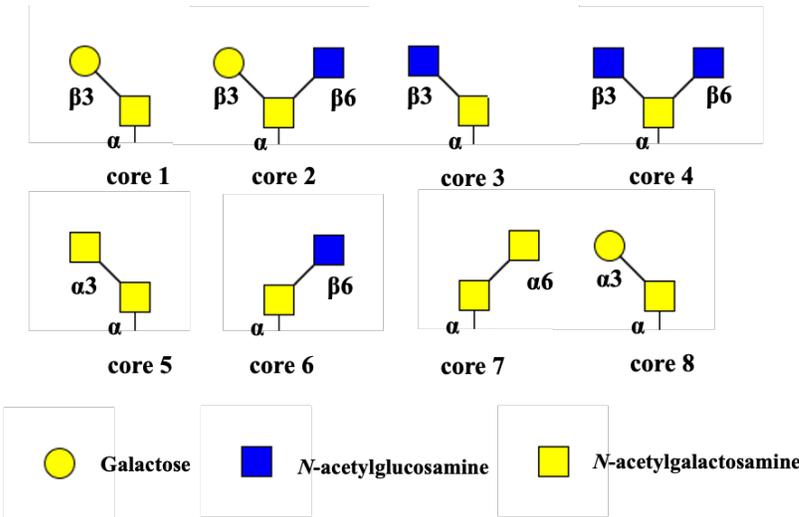


Figure 1.1/ O-GalNAc glycan core structures

The glycans are often found on proteins collectively named mucins. One feature that all mucins share is the presence of a domain rich in Ser and Thr residues, which are the sites of O-

GalNAcylation (Davis et al., 1986). Mucins are heavily glycosylated proteins present at epithelial surfaces of the body, which they protect from external agents or pathogens. An example is the gastrointestinal tract, where mucins form two layers of mucus gel used to prevent colonisation of the epithelium by enteric bacteria (Hansson and Johansson, 2010). The function is achieved thanks to the O-glycans creating a viscous environment that traps microbes. O-xylosylation (O-linked xylose (Xyl)) initiates in the Golgi and is found in proteoglycans, a family of glycoproteins bearing linear polysaccharides named glycosaminoglycans (GAG). GAGs such as heparan sulfate, chondroitin sulfate and dermatan sulfate all share the same core structure made of a tetrasaccharide bound to Ser or Thr through O-linked Xyl (Briggs and Hohenester, 2018; Fransson, 1968). Differences among GAG structures lie in the identity of the repeating disaccharide units that constitute the polysaccharide (Pedersen et al., 2000). The units consist of a hexosamine and a uronic acid sugar, which can be modified by sulfation in various positions. GAGs are essential in animals for their role in cell adhesion, growth factor signalling, and inflammatory response (Nicole et al., 2000; Pilia et al., 1996). O-fucosylation (O-linked Fuc) starts in the ER and occurs on proteins that contain an epidermal growth factor (EGF) or a thrombospondin-type repeat (TSR) sequence (Luo et al., 2006). The O-linked Fuc covalently bound to Ser or Thr can be elongated to a tetrasaccharide in EGF or a disaccharide in TSR sequences, and a consensus motif on the protein is required for the modification to take place (Hofsteenge et al., 2001; Wang and Spellman, 1998). An example of EGF protein O-fucosylated is Notch, a cell-surface receptor involved in mammal embryonic development (Weinmaster and Kintner, 2003). The O-fucosylation in Notch regulates its signalling pathway and is necessary for the ability of the protein to bind its ligands (Rampal et al., 2007). EGF sequences like the one found in Notch are also O-glycosylated, and the glycans, O-linked Glc either single or elongated to a trisaccharide, occur at a consensus sequence (Acar et al., 2008; Sethi et al., 2010). Similarly to O-fucosylation, the lack of O-glycosylation due to a mutation in the O-glycosyltransferase in a protein like Notch causes development defects (Fernandez-Valdivia et al., 2011). O-galactosylation (O-linked Gal) affects hydroxylysine on proteins in the ER and is found within collagen domains. Collagens have characteristic repeating repeats of glycine followed by two amino acids, more commonly proline and lysine (Spiro, 1967). O-galactosyl glycans on collagens are further modified with a single unit of Glc and regulate the protein folding and the size of the collagen fibrils (Tang et al., 2020).

Unlike the types of glycosylation described so far, O-GlcNAcylation does not occur in the ER or the Golgi but in the nucleus and the cytoplasm. In most cases, the modification consists of a single O-linked GlcNAc to Ser and Thr residues, not further elongated. This type of glycosylation is highly dynamic since it can be removed and installed multiple times on the same protein, similar to

phosphorylation. For this reason, it was discovered later compared to other types of O-glycosylation (Torres and Hart, 1984). There is an interplay between O-GlcNAcylation and phosphorylation, as the site of a protein can be alternatively modified with either of the two modifications in its lifetime. Moreover, O-GlcNAcylation can prevent or increase the phosphorylation of a specific protein (Liu et al., 2004). Most proteins exhibiting O-GlcNAc glycans are part of the RNA processing machinery, such as RNA polymerase II and its associated transcription factors, but also nuclear pore proteins (Comer and Hart, 2001).

The characterisation of glycans and glycoproteins, which is part of the glycobiology field of study, is of great interest but also challenging. One of the main challenges is that information about the structure of a glycan is not encoded in the glycoprotein's gene but instead depends on variables such as the levels of glycosyltransferases and availability of glycosyl donors, so it cannot be predicted directly from the gene sequence. The structural characterisation of glycans is further hampered by their heterogeneous nature (see Section 1.4.1). In the case of *N*-glycosylation, the attachment site for the *N*-glycan can be predicted. In fact, Asn residues can only carry an *N*-glycan if they are in the amino acid consensus sequence Asn–X–Ser/ Thr (where X can be any amino acid except proline) (Marshall, 1974). Nevertheless, the presence of a consensus sequence in a protein is not a guarantee of *N*-glycosylation. On the other hand, O-glycosylation does not have a consensus sequence and does not present a unique core structure, making it even more variable and unpredictable. Despite these challenges, much progress has been made in the characterisation of glycan structures and functions. This is especially true for *N*-glycans and O-GalNAc (often called mucin-type) glycans. Among the relatively less well understood classes of glycans is O-mannosylation, a form of glycosylation that has gained increasing attention over the past decade because of its unique structures and its biomedical relevance (Sheikh et al., 2017; Taniguchi-Ikeda et al., 2016).

## 1.2 O-mannosylation

Among the O-glycosylation classes, O-mannosylation represents a relatively poorly studied class in mammals. The discovery of O-mannosyl glycans in yeast (Sentandreu and Northcote, 1969) paved the way for the characterisation of the O-mannosylation pathway in higher eukaryotes. In yeast, the protein O-mannosyltransferases (PMT1-4) are responsible for transferring Man from the activated donor dolichol-phosphate-mannose (Dol-P-Man) to a Ser or Thr residue of an acceptor protein (Strahl-Bolsinger and Tanner, 1991). The glycosylation starts in the ER lumen and continues in the Golgi, where multiple units of Man are added to the O-linked Man using GDP-Man as the donor, forming a linear glycan. In higher eukaryotes, a protein homologue to PMTs was identified first in

the mutated *Drosophila melanogaster* with a rotated abdomen (Martin-Blanco and Garcia-Bellido, 1996). The same protein was later found to have homologues in mammals (Jurado et al., 1999), which led to the identification of mammalian protein O-mannosyltransferases (POMT1 and POMT2). Like in yeast, the classical pathway of O-mannosylation in mammals starts in the ER, where POMT1 and POMT2 are in charge of transferring Man from Dol-P-Man to the protein. The pathway regulated by POMT1 and POMT2 selectively modifies a few proteins, including  $\alpha$ -dystroglycan ( $\alpha$ DG) and an uncharacterised membrane protein (KIAA1549) (Vester-Christensen et al., 2013). Two other pathways for the O-mannosylation of proteins exist. One depends on a homologous family of protein O-mannosyltransferases encoded by *TMTCI-4* genes (transmembrane and tetratricopeptide repeat containing) (Larsen et al., 2017). Substrates for this pathway are the protein superfamily of cadherins and protocadherins, which are responsible for cell-cell adhesion (Brasch et al., 2012) and fulfil this function through their O-mannosyl glycans (Lommel et al., 2013). The last pathway depends on the *TMEM260* gene, which encodes a protein with structural similarities to O-mannosyltransferases (Larsen et al., 2023). The corresponding gene product, the enzyme TME260, is responsible for the O-mannosylation of plexins, which are transmembrane protein receptors for semaphorins, a group of extracellular proteins responsible for neurite growth and changes to the cytoskeleton outside the neuronal system (Perälä et al., 2012).

The O-mannosyl modification is conserved from fungi to mammals as it takes its first step in the same location (ER lumen) using the same activated donor (Dol-P-Man). Nevertheless, in mammals, the O-mannosylation pathway diverges from the fungi one as mature O-mannosyl glycans are not linear chains of Man units. In fact, after the addition of the O-linked mannose by POMT1 and POMT2, the glycan is further elongated in mammals, giving three possible structures (Figure 1.2) named core m1, m2 and m3 (Praisman and Wells, 2014). After transport of the O-mannosylated protein from the ER into the Golgi, protein O-linked mannose  $\beta$ -1,2-*N*-acetylglucosaminyltransferase 1 (POMGnT1) transfers GlcNAc from the activated donor uridine diphosphate (UDP)-GlcNAc to the O-Man, generating core m1 (GlcNAc $\beta$ 1-2Man). The enzyme MGAT5B ( $\beta$ -1,6-*N*-acetylglucosaminyltransferase) catalyses the formation of the GlcNAc $\beta$ 1-6Man linkage after the activity of POMGnT1, generating core m2 (GlcNAc $\beta$ 1-6 (GlcNAc $\beta$ 1-2)Man). The GlcNAc moieties of cores m1 and m2 can each be modified by the enzyme B4GALT ( $\beta$ -1,4-galactosyltransferase) with Gal in a  $\beta$ -1,4 linkage. The resulting O-mannosyl glycans are named core M1 and M2 (Praisman and Wells, 2014) and are found on  $\alpha$ DG, where they can be further capped with NeuAc in  $\alpha$ -2,3 linkages (Stalnaker et al., 2010). Cores m1 and m2 can also be found extended from GlcNAc to form structures such as the Lewis X (Gal $\beta$ 1-4(Fuc $\alpha$ 1-3)GlcNAc-) and the HNK-1 epitopes (HSO<sub>3</sub>-3GlcA $\beta$ 1-3Gal $\beta$ 1-4GlcNAc-, where GlcA stands for glucuronic acid).

These glycans are found in the brain on a protein named phosphacan and are essential for brain development (Morise et al., 2014a; Yaji et al., 2015). Lastly, among the O-mannosylated proteins in mammals, lecticans are modified with extended core m1 or m2 structures and are important for the structural stability of the brain tissue (Pacharra et al., 2013).

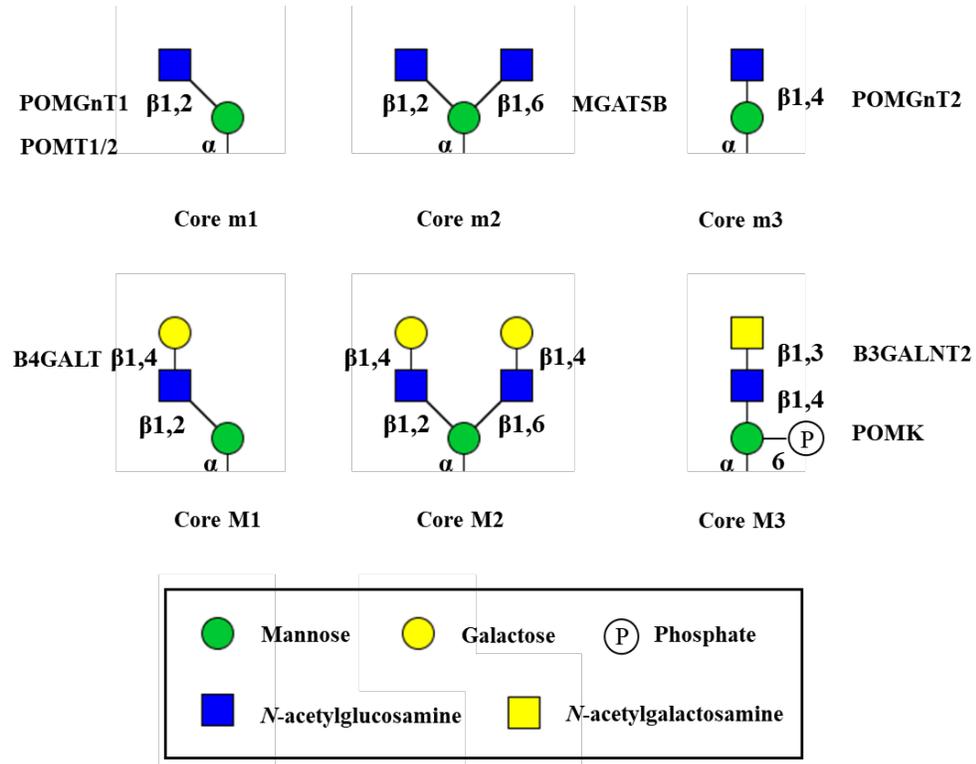


Figure 1.2/ Structures of the three core O-mannosyl glycans known in humans. Glycan biosynthesis starts in the ER with protein O-mannosyltransferase 1/2 (POMT1 and POMT2) and continues in the ER for core M3 while it proceeds in the Golgi for cores M1 and M2. Glycosidic linkage configuration and enzymes responsible for the biosynthesis of the glycans are indicated.

Differently from cores m1 and m2, the processing of core m3 proceeds in the ER (Yoshidamoriguchi et al., 2013). Here, the enzyme POMGnT2 (protein O-linked mannose  $\beta$ -1,4-N-acetylglucosaminyltransferase 2) catalyses the transfer of GlcNAc to O-linked mannose in  $\beta$ -1,4 linkage, generating the core m3 structure.  $\alpha$ DG is the only example known of a protein modified by POMGnT2 so far. POMGnT2 selectively acts on the O-mannosylated Thr residues in the “TPT” motif of  $\alpha$ DG (Halmo et al., 2017; Imae et al., 2021). Upstream of the modification site, a motif “RXR” (where X can be any amino acid) has been shown to be important but not essential for the enzyme activity (Halmo et al., 2017). Moreover, residues downstream of the “TPT” motif are involved in the enzyme recognition of the protein, but no consensus sequence was identified in this region (Imae et al., 2021). POMGnT2 causes the O-mannosylation pathway to split in two. It acts as

a gatekeeper enzyme since the O-mannosylation sites it modifies cannot be substrates for POMGnT1. The modification of these sites continues in the ER to form core M3 structures (Halmo et al., 2017). The O-mannosylated sites that POMGnT2 does not modify in the ER may proceed to be modified by the POMGnT1 enzyme when the nascent glycoprotein moves to the Golgi apparatus (Figure 1.3). Next, in the M3 pathway, the GlcNAc is modified by  $\beta$ 1,3-N-acetylgalactosaminyltransferase 2 (B3GALNT2), which adds GalNAc in  $\beta$ -1,3 linkage. The resulting glycan becomes the substrate for protein O-mannose kinase (POMK), which adds a phosphate to the oxygen on C6 of the mannose. The resulting glycan, GalNAc $\beta$ 1-3GlcNAc $\beta$ 1-4(phospho-6)Man, is referred to as core M3 (Praisman and Wells, 2014).

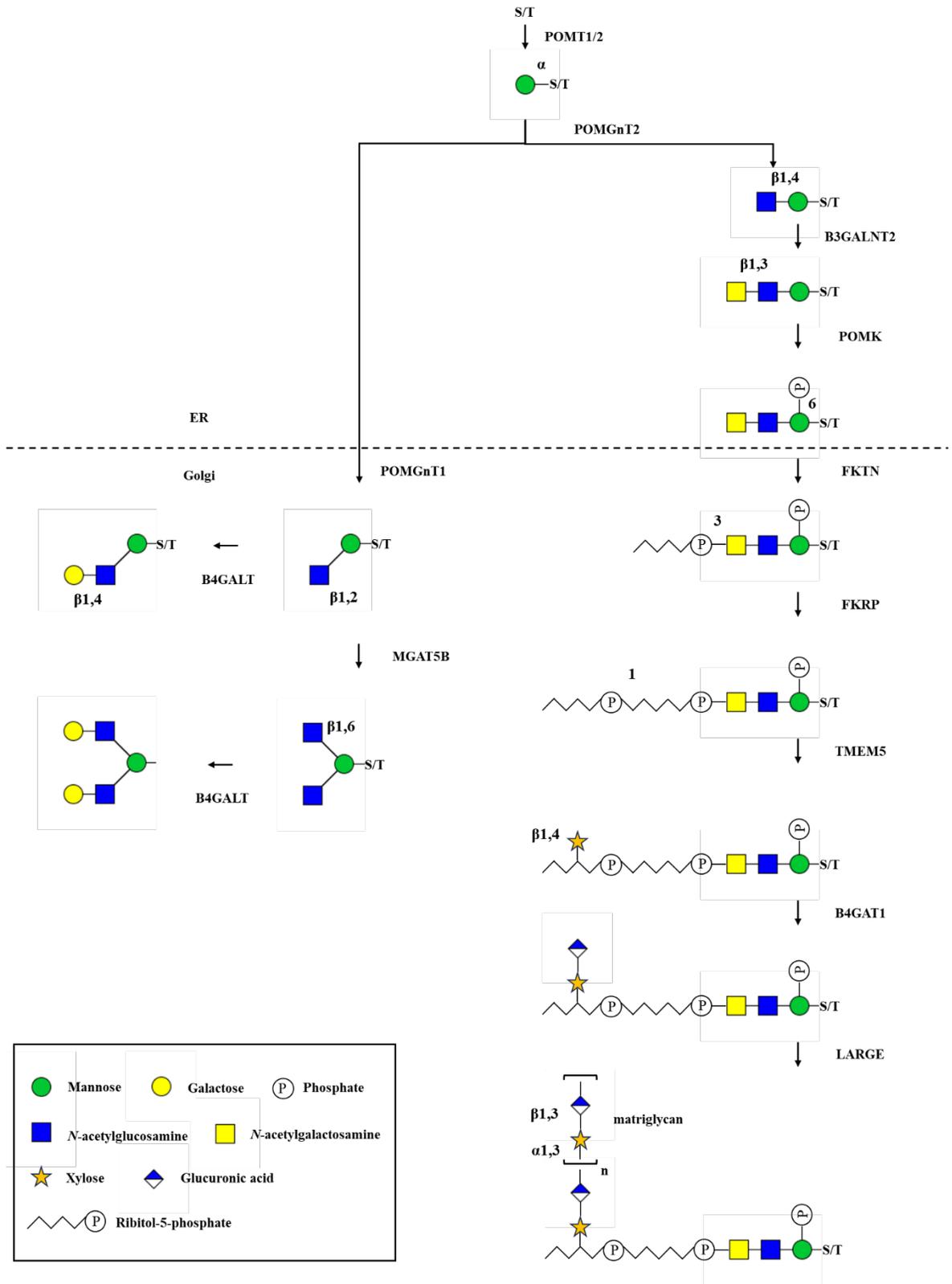


Figure 1.3/ The biosynthetic pathway of O-mannosyl glycans in mammals. *POMT1/2*, protein O-mannosyltransferase 1/2; *POMGnT1*, protein O-linked mannose O-1,2-N-acetylglucosaminyltransferase 1; *POMGnT2*, protein O-linked mannose O-1,4-N-acetylglucosaminyltransferase 2; *B3GALNT2*,  $\beta$ 1,3-N-acetylgalactosaminyltransferase 2; *POMK*, protein O-mannose kinase; *FKTN*, fukutin; *FKRP*, fukutin-related protein; *TMEM5*, ribitol  $\beta$ 1,4-xylosyltransferase; *B4GAT1*,  $\beta$ 1,4-glucuronyltransferase 1; *LARGE* like-acetylglucosaminyltransferase; *MGAT5B*,  $\beta$ -1,6-N-acetylglucosaminyltransferase; *B4GALT*  $\beta$ -1,4-galactosyltransferase.

Core M3 glycans have been found to date exclusively on  $\alpha$ DG and they can be further extended in the Golgi apparatus. The enzyme fukutin (FKTN) transfers ribitol-5-phosphate (Rbo5P) from CDP-ribitol (CDP-Rbo) to the 3-position of the GalNAc residue of the core M3 glycan while the enzyme fukutin-related protein (FKRP) adds a second Rbo5P to the 1-position of the first Rbo5P (Figure 1.4) (Gerin et al., 2016; Kanagawa et al., 2016). FKTN and FKRP share homology in their sequences but have different activities. In fact, FKRP is not able to add the first Rbo5P in place of FKTN and neither of the two enzymes can add extra Rbo5P moieties onto the glycan (Gerin et al., 2016). The crystal structure of FKRP with sugar donor and acceptor showed that the enzyme exists in solution as a tetramer (Kuwabara et al., 2020). One subunit recognises the phosphate group of Rbo5P, and a different subunit recognises the phosphate on the O-linked Man. These interactions are essential for enzyme activity and could explain why FKRP can form them exclusively on a peptide bearing a single unit of Rbo5P.

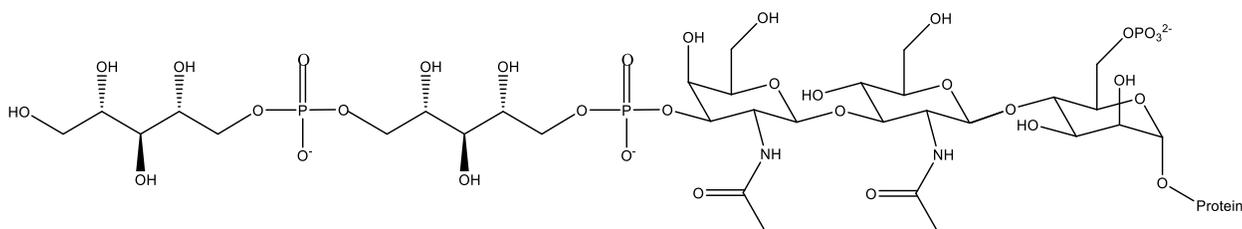


Figure 1.4/ Structure of the core M3 glycan, which is uniquely found in  $\alpha$ DG, extended with a tandem repeat of ribitol-5-phosphate (Rbo5P) which elongates the core m3 structure.

Rbo5P is an alditol that had until recently not been known to exist in humans, yet is commonly found in gram-positive bacteria as part of the teichoic acids present in the cell wall (Brown et al., 2013). In 2016, Rbo5P was found for the first time in mammalian cells by three different groups (Kanagawa et al., 2016; Praissman et al., 2016; Yagi et al., 2016). RboP was identified as part of the extended core M3 glycan structure on  $\alpha$ DG and was uniquely found in this mammalian protein. The tandem repeat of Rbo5P is still further processed in the Golgi by the enzyme ribitol  $\beta$ 1,4-xylosyltransferase (TMEM5), which adds xylose (Xyl) in  $\beta$ 1,4 linkage to the second Rbo5P, followed by the enzyme  $\beta$ 1,4-glucuronyltransferase 1 (B4GAT1) which adds GlcA in  $\beta$ 1,4 linkage to Xyl. The full extension of core M3 glycans is completed by the activity of the enzyme like-acetylglucosaminyltransferase (LARGE). LARGE is a bifunctional enzyme with Xyl and GlcA transfer activities (Inamori et al., 2012). In particular, the enzyme adds sequentially Xyl, to the existing GlcA on core M3, and then GlcA to the attached Xyl, forming the disaccharide GlcA $\alpha$ 1-3Xyl $\alpha$ 1-3. The disaccharide unit can be added multiple times forming an oligo/polysaccharide named matriglycan (Yoshida-Moriguchi and Campbell, 2014).  $\alpha$ DG is thus far known to be an exclusive substrate for LARGE while the enzyme's paralog, LARGE2, modifies both  $\alpha$ DG and proteoglycans such as glypican-4 (Inamori et al., 2016). In the case of glypican-4, the matriglycan is not added onto a core M3 glycan but is believed to be attached to an O-xylose glycan derived from the linkage region of a glycosaminoglycan structure (Inamori et al., 2016).

### 1.3 O-mannosylated proteins

The first examples of O-mannosyl glycans in mammals were discovered in the brains of rats (Finne et al., 1979), where this modification accounts for about 30% of the total O-linked glycans (Chai et al., 1999).  $\alpha$ DG was the first O-mannosylated protein to be identified (Chiba et al., 1997), and since then, the importance of its glycosylation in the brain has been demonstrated by several studies (Nguyen et al., 2013; Satz et al., 2010; Wright et al., 2012; reviewed in Nickolls and Bönnemann, 2018;). A few other O-mannosylated proteins were identified after the discovery of  $\alpha$ DG. First, IgG2 overexpressed in Chinese hamster ovary (CHO) cells was shown to be modified on a Ser with a single O-linked Man (Martinez et al., 2007). Then, the glycosylation of the signal transducer CD24 was studied, and core M1 and M2 glycans were identified (Bleckmann et al., 2009). At that point,  $\alpha$ DG was still the only O-mannosylated protein identified in the brain, but O-mannosyl glycans could still be identified in the brain of *DAG1* knockout mice (Stalnaker et al., 2011). After the Stalnaker study, other proteins were identified as O-mannosylated in the brain, such as receptor-type tyrosine-protein phosphatase zeta (PTPRZ1) (Dwyer et al., 2012; Morise et al., 2014; Trinidad et al., 2013), neurofascin 186 (Pacharra et al., 2012), and lecticans (Pacharra et al., 2013). A more targeted protocol was developed by the Clausen group with the aim of enriching and analysing

proteins modified with O-linked Man (Vester-Christensen et al., 2013). To achieve that, the breast cancer line MDA-MB-231, where POMGnT1 is inactive, was used to produce truncated O-glycan structures with O-linked Man. Proteins modified with such glycans could be enriched by lectin-weak affinity chromatography and analysed by mass spectrometry. Over 50 new O-mannosylated proteins, mostly belonging to the cadherin, protocadherins or plexin families, were identified. In the following years, the glycosylation of the proteins discovered was demonstrated to be unaffected by the activity of POMGnT1. In fact, cadherins, protocadherins and plexins O-mannosyl glycans are not extended to make core M1 or M2 structures (Larsen et al., 2023, 2017). A list of known O-mannosylated proteins is shown in Table 1.1.

<b>Protein name</b>	<b>Uniprot ID</b>
Cadherins	P12830,Q9UEJ6,Q9NYQ6,Q9HCU4,P55287 P55290,Q9H159, P19022,Q9H251,P22223, P55283,P55285,P55286,P55288, Q9WTR5,G1TL92
Desmocollin-2	Q02487
Desmoglein-2	Q14126
Dystroglycan	Q14118
Hepatocyte growth factor receptor	P08581
inter-alpha-trypsin inhibitor 5	Q8BJD1
Intercellular adhesion molecule 1	P05362
Lecticans	P55066, FIN2Y8, P81282
Macrophage-stimulating protein receptor	Q04912
Neurexin 3	Q6P9K9
Neurofascin 186	Q810U3
Plexins	Q9UIW2, O75051, O43157, O15031, Q9ULL4, Q9Y4D7, P70206, B2RXS4
Protein disulfide-isomerase	P07237
Protein disulfide-isomerase A3	P27773 P30101
Protocadherins	Q9Y5I4,Q9UN67,Q9Y5F2,Q9Y5E9,Q9Y5E7,Q9UN66,Q9Y5E1, Q14517,Q8BNA6,Q9QXA3,Q6V0I7,Q9Y5H3,Q9Y5G9,Q9Y5G3, Q9Y5G2,Q9UN70,Q9Y5F6,Q08174,Q9P2E7,Q9NPG4,Q96JQ0, O14917,O14917,Q9HCL0,Q95206,Q91XX5,Q91Y09,F8VPK8, E9PXQ7,Q91XY7,Q9HC56
Receptor-type tyrosine-protein phosphatase beta	P23467
Receptor-type tyrosine-protein phosphatase zeta	B9EKR1
Sinal transducer CD24	P24807
SUN domain-containing ossification factor	Q9UBS9
TGF-beta receptor type-1	P36897
UPF0606 protein KIAA1549	Q9HCM3

*Table 1.1/ List of known O-mannosylated proteins* (Abbott et al., 2008;Bartels et al., 2016; Bleckmann et al., 2009; Dwyer et al., 2012; Lommel et al., 2013; Pacharra et al., 2013, 2012; Sheikh et al., 2017; Stalnaker et al., 2010; Vester-Christensen et al., 2013; Winterhalter et al., 2013)

## 1.4 $\alpha$ -Dystroglycan

Dystroglycan was first identified in skeletal muscle tissue (Ervasti and Campbell, 1991) as part of the dystrophin-glycoprotein complex (DGC), a multimeric transmembrane protein complex connecting skeletal muscle cells to the basement membrane (Ibraghimov-Beskrovnaya et al., 1992). The dystroglycan gene (*DAG1*) is highly conserved in vertebrates and is transcribed into a single mRNA that translates into a proprotein (895 amino acids in humans) which is auto-proteolytically cleaved into two subunits,  $\alpha$ DG and  $\beta$ -dystroglycan ( $\beta$ DG) (Holt et al., 2000).  $\alpha$ DG is a highly glycosylated protein with both *N*- and *O*-glycans making the molecular weight (MW) of the protein rise from ~40 kDa (unmodified) to ~156 kDa in skeletal muscle and ~120 kDa in brain (Ibraghimov-Beskrovnaya et al., 1992). The structure of  $\alpha$ DG is characterised by three domains: an N-terminal domain (29-316), a central domain (317-485) and a C-terminal domain (486-653) (Figure 1.5). The N-terminal domain facilitates  $\alpha$ DG *O*-glycosylation, allowing the full extension of core M3 with matriglycan (Okuma et al., 2023), and is eventually removed from  $\alpha$ DG by the action of the protease furin (Singh et al., 2004). The central domain of  $\alpha$ DG is also referred to as a mucin-like domain because, similarly to mucins, it contains variable *O*-glycosylated tandem repeat domains rich in Pro, Thr, and Ser (Bansil and Turner, 2018). In this domain, all the *O*-linked glycans of  $\alpha$ DG are located. Sites with *O*-linked *N*-acetylhexosamine (HexNAc) were first identified in  $\alpha$ DG from human skeletal muscle (Nilsson et al., 2010) and later identified as *O*-GalNAc modification sites (Steentoft et al., 2011). The  $\alpha$ DG mucin-like domain also contains *O*-mannosyl glycans, which were first identified in bovine peripheral nerve (Chiba et al., 1997) and rabbit skeletal muscle (Sasaki et al., 1998) as core M1 capped with NeuAc. Later, core M2 and M3 glycans were identified in  $\alpha$ DG (Stalnaker et al., 2010; Yoshida-Moriguchi et al., 2010). Core M1 glycans are more abundant than cores M2 and M3 and are thought to assist the recruitment of core M3 GTs, in particular FKTN (Xiong et al., 2006).



The link between  $\alpha$ DG and the ECM is dependent on modification of the glycoprotein with core M3 glycans elaborated with matriglycan and plays a significant role in maintaining muscle membrane integrity. For this reason, non-functional  $\alpha$ DG, which is not appropriately glycosylated and unable to bind laminin, is responsible for a class of neuromuscular diseases that are collectively named dystroglycanopathies. Given the connection between core M3 glycans and dystroglycanopathies, several studies have focused on the analysis of core M3 biosynthetic pathways. Strikingly, in 2016 a novel glycosylation pathway responsible for the Rbo5P modification was discovered during attempts to characterise the structure of core M3 glycans extended with matriglycan (see section 1.2). The tandem Rbo5P repeat was identified in  $\alpha$ DG by mass spectrometry thanks to the expression of an  $\alpha$ DG-derived peptide carrying the site for core M3 glycosylation with an Fc (fragment crystallisable) tag used for enrichment of the peptide by immunoprecipitation (Kanagawa et al., 2016). After enrichment, hydrofluoric acid treatment was used to cleave the phosphodiester bonds of the Rbo5P linker, allowing the site of core M3 attachment to be identified. The presence of Rbo was determined by mass spectrometry through monosaccharide composition analysis of the glycans released from the enriched glycopeptide. A similar study, also expressing a recombinant glycopeptide carrying a site for core M3 glycosylation, was able to identify the glycan structure elongated with just one Rbo5P or glycerol phosphate (GroP), as well as a structure modified with the tandem Rbo5P linker and one GlcA-Xyl disaccharide (Yagi et al., 2016).

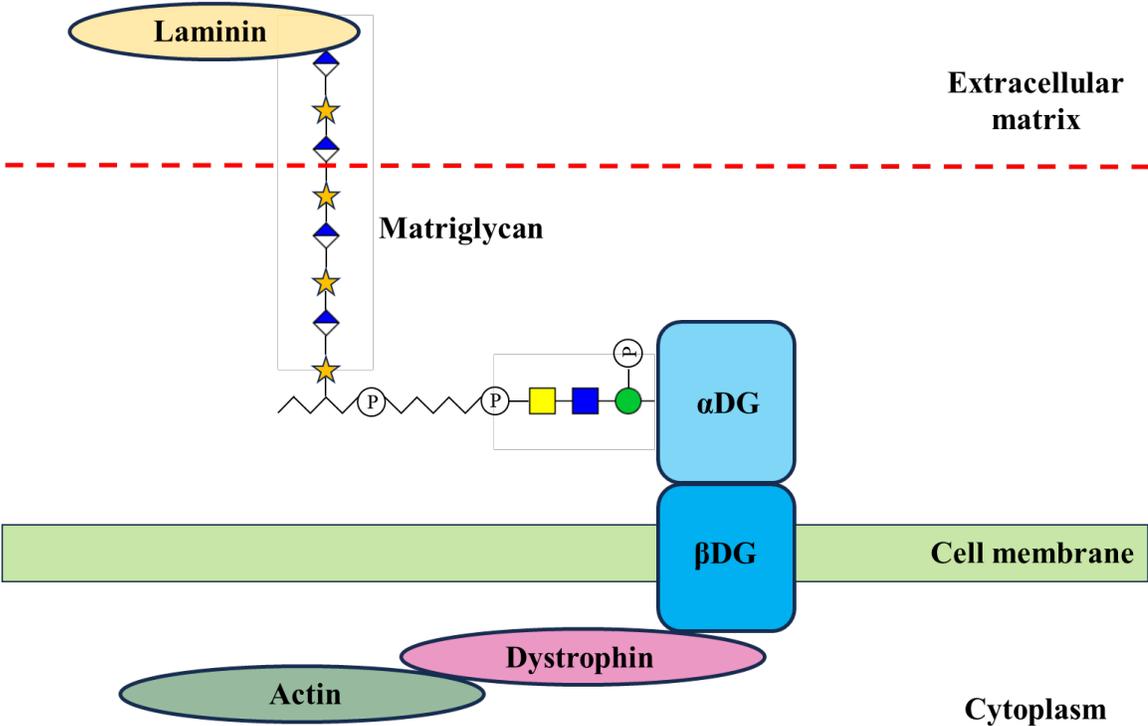


Figure 1.6 / Molecular organization of the dystrophin-glycoprotein complex (DGC). The glycosylated protein  $\alpha$ DG participates in the DGC and mediates interactions with the extracellular matrix (ECM) binding laminin through the polysaccharide matriglycan. Indirectly,  $\alpha$ DG binds actin inside the cell through a non-covalent bond with the transmembrane protein  $\beta$ DG, which in turn binds dystrophin.

### 1.4.1 Dystroglycanopathies

Dystroglycanopathies are a series of genetic diseases that cause muscular dystrophy, brain malformations and eye abnormalities. These diseases are related to the aberrant glycosylation of  $\alpha$ DG, which is found to be hypoglycosylated in patients affected by dystroglycanopathies. In particular, the  $\alpha$ DG core M3 glycan, if not fully extended with matriglycan, is unable to fulfil its function of binding to the ECM. Disruption of the glycosylation pathway can be caused by mutations in the  $\alpha$ DG gene itself or in any of the genes responsible for the enzymes involved in the biosynthesis of the core M3 glycans. A disease is classified as a primary dystroglycanopathy if a mutation occurs in the gene *DAG1* which encodes dystroglycan, and secondary if it is caused by defects in glycosylation enzymes required for the O-mannosylation of  $\alpha$ DG. Last, mutations of enzymes involved in the synthesis of sugar donors used to synthesise O-mannosyl glycans are responsible for tertiary dystroglycanopathies. A *DAG1* mutation in the N-terminal domain of  $\alpha$ DG affects the length of the matriglycan causing limb-girdle muscular dystrophy (LGMD) and cognitive impairment (Geis et al., 2013; Hara et al., 2011), and is an example of primary dystroglycanopathy. Mutations in the genes encoding POMT1/2 (Akasaka-Manyu et al., 2004), POMGnT2, B3GALNT2 and POMK (Jae et al., 2013) cause Walker-Warburg syndrome (WWS). WWS is an example of a secondary dystroglycanopathy and is a disorder characterised by congenital muscular dystrophy and brain malformations. Other examples are muscle-eye-brain disease (MEB) caused by a mutation in the gene encoding POMGnT1 (Yoshida et al., 2001), and congenital muscular dystrophy type 1D (MDC1D) caused by mutations in the *LARGE* gene. Notably, the enzymes responsible for transferring Rbo5P onto core M3, FKTN and FKRP, can also be inactive due to mutations and this similarly leads to dystroglycanopathies (Kanagawa and Toda, 2017). In particular, Fukuyama-type congenital muscular dystrophy (FCMD) results from mutations in FKTN and leads to congenital muscular dystrophy and eye abnormalities (Kobayashi, et al., 1998), while a mutation in FKRP is the cause of MDC1C (M. Brockington et al., 2001) and a form of LGMD named LGMD21 (Martin Brockington et al., 2001). Lastly, malfunctioning of the enzymes involved indirectly in the synthesis of core M3 glycans is responsible for tertiary

dystroglycanopathies. An example is the isoprenoid synthase domain-containing protein (ISPD), which synthesizes CDP-Rbo, and has been reported to be mutated in subjects affected by WWS (Willer et al., 2012).

## 1.5 Glycoproteomics

The importance of glycosylation is evident when considering all the processes it regulates in an organism. Moreover, altered glycosylation is the cause of various diseases like cancer (Rodrigues et al., 2018), infection (Bhat et al., 2019) and neurological disorders (Paprocka et al., 2021), as well as the dystroglycanopathy examples discussed above (section 1.3.1). Research aimed at generating a better understanding of the roles of glycosylation has pushed the development of methods for the analysis of the glycoproteome, which translates into the identification of glycoproteins, glycosylation sites and glycan compositions. Glycoproteomics has the objective of studying the glycoproteins present in an organism at a certain time, place and under specific biological conditions. A typical glycoproteomic experiment requires the isolation of proteins from cells or tissue, the enrichment of glycoproteins, their digestion into peptides and the analysis of these peptides and glycopeptides by liquid chromatography (LC) coupled to mass spectrometry (LC-MS/MS). Examples of enrichment strategies and mass spectrometric analyses are provided below.

### 1.5.1 Glycoprotein and glycopeptide enrichment

Studying glycosylation represents a challenge because the information about the structure of a glycan present on a glycoprotein is not written in the gene encoding the glycoprotein. In fact, glycan structures are diverse, often cell-type specific, and can change in response to stimuli. The specific glycosylation site cannot be predicted for O-glycans. While it can be predicted for *N*-glycans, the presence of a consensus sequence in a protein does not guarantee the presence of a glycan. The result is that only a fraction of the *N*-glycan sequons or the Ser and Thr residues in a given glycoprotein are glycosylated, a characteristic of glycosylation called macroheterogeneity. Glycoproteomic studies are also hampered by microheterogeneity of glycosylation, which reflects the presence of different glycan structures on the same glycosylation site at different times or on different copies of the same protein at the same time (Stanley and Sudo, 1981). Finally, the poor ionisation efficiency of glycosylated peptides compared to their non-glycosylated counterparts makes detection by mass spectrometry difficult. To overcome these issues, a glycoproteomic experiment usually requires an enrichment step aimed at recovering a defined set of glycoproteins

from a complex sample and separation of peptides and glycopeptides following glycoprotein digestion.

Glycoprotein enrichment can be achieved by a variety of different methods. The first is lectin affinity purification which exploits the interaction of a glycoprotein with a lectin. Lectins are carbohydrate-binding proteins with specific affinities towards certain glycan structures or more often substructures; lectins can be immobilized on a solid support, such as agarose, and used to enrich glycoproteins. Concanavalin A (ConA) is an example of a lectin which recognises terminal Man and Glc (Becker et al., 1975) and has an affinity for *N*-glycans (Morelle et al., 2006). ConA affinity purification has been used to study O-mannosylation in mammalian cells. Specifically, a stable gene knockout of POMGnT1 prevented the elongation of O-linked Man into core M1 and M2 structures allowing mono-O-mannosylated glycopeptides to be enriched by ConA affinity protocols (Vester-Christensen et al., 2013). A similar approach has been used to study O-GalNAcylation (Steentoft et al., 2011). For this purpose, cell lines underwent a stable gene inactivation of *cosmc*, which encodes a protein required for the elongation of O-linked GalNAc, and the resulting mono O-GalNAc glycopeptides were enriched using *Vicia villosa* lectin (VVL), a lectin specific for terminal O-GalNAc (Tollefsen and Kornfeld, 1983).

## 1.5.2 Metabolic oligosaccharide engineering (MOE)

An alternative method of enrichment involves a chemical biology approach termed metabolic oligosaccharide engineering (MOE). MOE is a technique in which unnatural tags are introduced into cellular glycans with the aim to enable the labelling, enrichment and eventually identification of glycoproteins and glycopeptides, for example by mass spectrometry. MOE involves the use of an unnatural monosaccharide, which is fed to cells and incorporated into glycans by the cell's own glycosylation machinery. The unnatural monosaccharide carries a chemical reporter that can be conjugated with a complementary reporter group (or probe) after its incorporation into a glycan. The probe is used to visualise or enrich all the glycoproteins containing the unnatural monosaccharide, after which the enriched glycoproteins can be digested into peptides and analysed by mass spectrometry. The reaction between the chemical reporter and the probe is referred to as a bioorthogonal chemical reaction and has the characteristics of being selective, non-toxic and efficient under mild, aqueous conditions. This is important since the reaction happens in a biological context in the presence of many other proteins and other molecules that can possibly interfere with the reaction.

The different functional groups introduced on a monosaccharide can be exploited to tag glycoproteins by making use of their unique chemical reactivity. Aldehydes and ketones can be

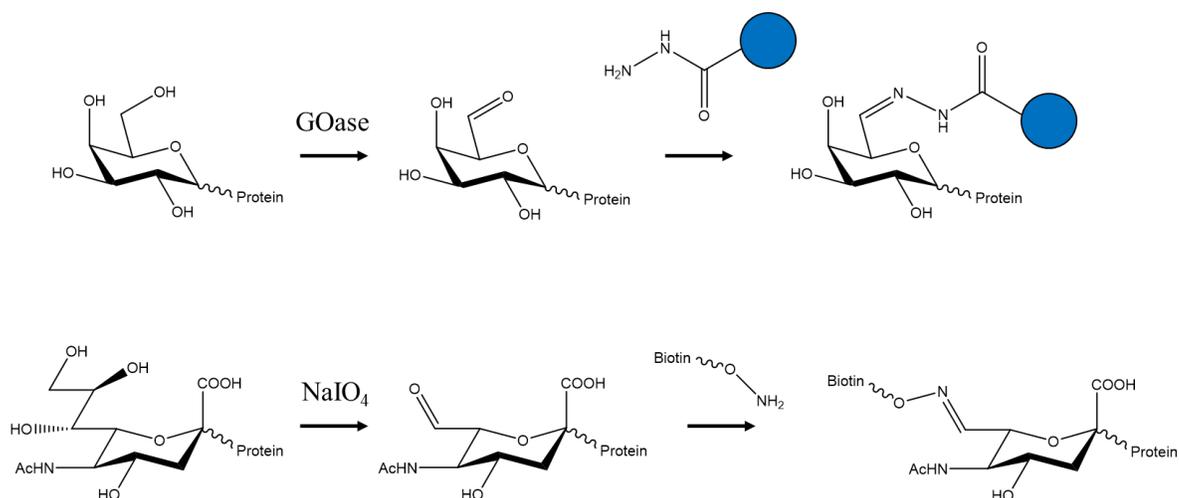
conjugated with a probe modified with an aminoxy or hydrazide by oxime and hydrazone ligation, respectively (Kölmel and Kool, 2017). Alkynes can react with an azide probe by a Cu(I)-catalyzed [3 + 2] cycloaddition (Rostovtsev et al., 2002), and the reverse sequence, reacting an azido-monosaccharide with an alkyne probe, is also used. Moreover, azides can react with triarylphosphine probes via Staudinger ligation (Saxon and Bertozzi, 2000) or with cycloalkyne probes via a strain-promoted [3+2] cycloaddition reaction (Agard et al., 2004). The resulting labelled glycoproteins can be detected and enriched by exploiting the conjugated probe. The detection is direct when fluorescent probes are used, and Cy5 is an example of this (Chang et al., 2007). In the case of probes based on biotin or FLAG-peptide (with sequence DYKDDDDK), the detection is indirect because these tags need to be visualised using a secondary reagent like fluorescein-5-isothiocyanate (FITC) avidin (Saxon and Bertozzi, 2000) and FITC-anti-FLAG antibody (Kiick et al., 2002). The same affinity probes can also be exploited to purify glycoproteins from a complex largely unlabelled background, using streptavidin or anti-FLAG agarose to enrich biotinylated or FLAG-tagged glycoproteins respectively.

Based on the type of glycosylation under investigation, a specific type of unnatural monosaccharide must be synthesised. The monosaccharide is functionalised and per-O-acetylated to allow its passage through the cell membrane into the cell. An example of a functionalised monosaccharide used as a chemical reporter is Ac<sub>4</sub>GalNAz, the azido-tagged version of GalNAc, which has been fed to cells (Hang et al., 2003) and animals (Dube et al., 2006) to label various glycans. Ac<sub>4</sub>GalNAz is metabolised by the cell through the GalNAc salvage pathway which converts it to the activated UDP-sugar, which is then used as a donor by GalNAc transferases. Not every functionalised monosaccharide can be used as a substrate by the native enzymes present in the cell. For example, the alkyne version of GlcNAc, Ac<sub>4</sub>GlcNAIk, is not efficiently incorporated into glycoproteins (Batt et al., 2017). Nevertheless, the Ac<sub>4</sub>GlcNAIk can enter the metabolic pathways of a cell if the enzyme involved in the biosynthesis of UDP-GalNAc is mutated (Cioce et al., 2021).

Despite the successes of these strategies, the design of novel MOE strategies is still challenging. First, if the non-native monosaccharide is unavailable commercially, it must be synthesised. Second, *in vivo* or *in vitro* experiments must be performed to show that the modified monosaccharides can be processed by the enzymes that act on the natural substrate, despite the presence of a new functional group. Last, the efficiency with which the unnatural substrate is incorporated into glycans must be taken into account; low incorporation efficiency could result in a low level of labelling as in the case of azido-fucose (Okeley et al., 2013).

### 1.5.3 Exploiting chemical properties of carbohydrates for enrichment

A different strategy for the enrichment of glycoproteins involves directly modifying native glycans *in vivo* or *in vitro* to generate chemical reporters which can be covalently linked to an enrichment/affinity probe. From a chemical perspective, the potential of hydroxyl groups to be oxidized to ketones or aldehydes offers an attractive way to generate reactive groups amenable to further modification. Selective oxidation of specific hydroxyl groups can be achieved either enzymatically or chemically. In the first case, the enzyme galactose oxidase (GOase) is used to oxidise terminal Gal and GalNAc residues in glycoconjugates at C6, generating an aldehyde on the glycoproteins (Figure 1.7). The method was used to study collagen O-glycosylation by conjugating the aldehyde-modified glycoprotein with hydrazide beads through the formation of hydrazone bonds (Taga et al., 2012). Glycoproteins enriched this way were then released from the beads under acidic conditions and analysed by mass spectrometry. The oxidation of certain monosaccharides in glycoconjugates can also be achieved chemically. Specifically, monosaccharides such as Man, Gal, Fuc, and NeuAc with vicinal cis diols are susceptible to oxidation with sodium periodate, which results in the cleavage of the carbon-carbon bond between the two hydroxyl groups and concomitant formation of an aldehyde group. This approach was reported for the first time applied to *N*-glycosylated proteins which were pulled from a cell lysate after oxidation, exploiting the reactivity of the aldehyde groups produced on oxidation with the hydrazide functional group, which had been immobilized on a solid support (Zhang et al., 2003). NeuAc oxidation requires a concentration of sodium periodate (1 mM) that is lower than that used to oxidise other monosaccharides (15 mM) (Walker, 2002). This characteristic has allowed the selective detection of sialylated proteins on the cell surface of Chinese hamster ovarian (CHO) cells by exploiting the conjugation of oxidized NeuAc with aminoxy-biotin (Zeng et al., 2009) (Figure 1.7).



*Figure 1.7 / Schematic diagram showing methods of monosaccharide oxidation and conjugation with probes. At the top, Gal on a glycoprotein is oxidised by GOase at C6 and is then conjugated with hydrazide-bearing beads by forming a hydrazone. At the bottom, NeuAc on a glycoprotein is oxidised at C7 with sodium periodate (NaIO<sub>4</sub>) and then conjugated with aminoxy-biotin by forming an oxime.*

Other studies used the selective oxidation of NeuAc to enrich sialylated glycoproteins and identify them by mass spectrometry. A strategy used to achieve that involved the capture of glycoproteins onto hydrazide beads followed by their digestion with a protease (Nilsson et al., 2009). Sialylated peptides, still attached to the beads, were released under mild acid conditions which break the bond between NeuAc and the penultimate monosaccharide in the glycan, resulting in the release of the desialylated glycopeptides from the beads.

Rbo5P, similarly to NeuAc, has vicinal diols which are susceptible to periodate oxidation. Indeed, sodium periodate treatment has been proven to effectively generate an aldehyde on  $\alpha$ DG glycopeptides modified with core M3 structures (Praisman et al., 2016) (Figure 1.8). In that study, a recombinant version of the  $\alpha$ DG peptide carrying the core M3 glycan extended with matriglycan was expressed in cells, purified, and treated with sodium periodate. The treatment removed the matriglycan and reduced the mass of the glycopeptide allowing its detection by mass spectrometry.

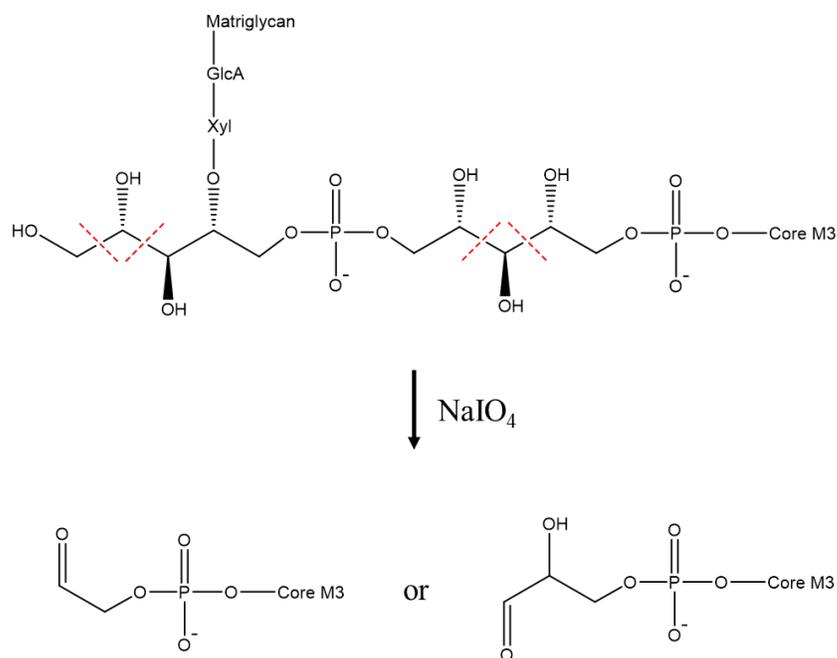


Figure 1.8 / Schematic diagram of Rbo5P oxidation in a core M3 extended glycan. The oxidation with sodium periodate affects either the bond between C-2 and C-3 or that between C-3 and C-4 of Rbo5P, generating two different products.

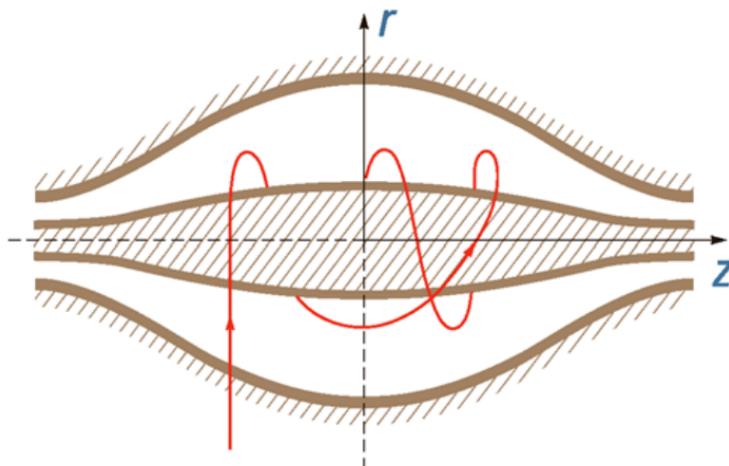
The oxidation of Rbo5P after sodium periodate treatment could also theoretically be exploited to generate an aldehyde that could act as a chemical reporter on Rbo5P-containing glycoconjugates. The aldehyde, susceptible to conjugation with an enrichment probe, would allow the enrichment of proteins modified with Rbo5P.

### 1.5.4 Mass spectrometric analysis

A glycoproteomics experiment can be designed in different ways. One approach consists of the digestion of enriched glycoproteins into peptides and glycopeptides, which are then analysed by LC-MS/MS. This digest and analyse approach is called bottom-up (glyco)proteomics since it identifies proteins present in a sample by matching generally only some of its constituent peptides. To enable detection, a sample in the liquid phase eluting from an LC column must enter the mass spectrometer as gas phase ions in order to be susceptible to the fields used in the mass analysers to separate the ions on the basis of their mass to charge ratios ( $m/z$ ). Electrospray ionization (ESI) is an example of a soft ionisation technique that transfers ionised species from solutions into the gas phase. The sample enters an ESI needle to which a voltage is applied. The charged needle is responsible for attracting or repelling the analytes in solution which are also charged. Analytes

attracted this way are removed while the repelled ones stay in solution. The liquid passing through the needle is nebulized, often with the help of a stream of  $N_2$ , into fine droplets containing the charged analytes. The ionisation chamber may be heated, and/or be furnished with a counter flow of  $N_2$  gas so that the solvent from the droplets evaporates over time and the electrical charge density on the surface of the droplet increases. Once the Rayleigh limit (the limit for the maximum density of charges in a droplet) is reached, charged analytes are expelled from the droplet and enter the gas phase. This model for ESI ion desolvation is called the ion evaporation model (IEM) (Nguyen and Fenn, 2007). An alternative model, the charged residue model (CRM), proposes a different mechanism for ion desolvation. In the CRM, once the Rayleigh limit is reached charge repulsion means the droplet explodes into smaller droplets and the cycle of solvent evaporation, droplet shrinkage and explosion repeats until all charged analytes have desolvated and reached the gas phase (Fernandez De La Mora, 2000). While IEM better explains the ionisation of small molecules, CRM describes the behaviour in the ionization chamber of larger molecules (above 1000 Da) (Wilm, 2011)

Gas phase ions are transferred from the ionisation chamber to the mass analyser where they are analysed based on their  $m/z$ . An example of a mass analyser that is typically used for (glycol)proteomics applications is the Orbitrap (Figure 1.9). The reason the Orbitrap family of instruments are such powerful instruments in proteomics and glycoproteomics are their versatility, excellent detection limits, speed (allowing excellent proteomic penetration) and their excellent mass resolution (Makarov, 2000; Michalski et al., 2012). The Orbitrap mass analyser consists of a central electrode in the shape of a spindle elongated along the axial axis  $z$ , and an outer electrode kept at ground potential used to detect image current (Makarov, 2000). A small packet of ions with different  $m/z$  enter the Orbitrap simultaneously and are subjected to an axial and a radial electric field. The radial field pushes ions into a circular movement around the central electrode while the axial field induces an oscillatory movement along the  $z$ -axis. The oscillation is achieved by applying an axial electric field which is zero at the center of the electrode but progressively increases in opposite directions along the  $z$  axis. The result of combining radial, axial and rotational movement of an ion is a stable trajectory, around the central electrode, in the shape of a spiral. Ions with different  $m/z$  also oscillate around the central electrode with different frequencies. Each component of an ion's movement has a frequency depending on the ion's  $m/z$  and can be detected as image current by the outer electrode. Since only the axial frequency is independent of the initial energy and position of the ions, this is used to determine the ion's  $m/z$  (Scigelova et al., 2011).



*Figure 1.9/ Schematic of the Orbitrap. In red an example of a stable ion trajectory. The axial z and radial r axes are shown. (Adapted from LTQ-Orbitrap XL manual, Thermo Scientific)*

An example of a mass spectrometer equipped with an Orbitrap analyser is the Orbitrap Fusion™ Tribrid™. The mass spectrometric analyses performed for the work described in this Thesis were achieved by LC-MS/MS using an Orbitrap Fusion™ Tribrid™, therefore the set-up of the instrument is described in more detail here (Figure 1.10). The analysis of peptides or glycopeptides by mass spectrometry involves a sequence of events that are set up to take place in the mass spectrometer. First, the intensities and  $m/z$  values of all the ions injected into the Orbitrap are measured and presented as a mass spectrum. The  $m/z$  of a peptide detected in the mass spectrum is characteristic of the peptide's composition, but a mass spectrum generated using a soft ionization technique, such as ESI, rarely contains fragment ions that provide information on the amino acid sequence of the peptide; therefore the peptide is isolated on the basis of its  $m/z$ , and fragmented, typically following ion acceleration and collision with molecules of a neutral gas, often nitrogen, in a process known as collision induced dissociation (CID). Careful choice of the CID energy input means that the majority of the product ions produced are so-called sequence ions that provide information on the amino acid sequence of the peptide. The product ions that result from these collisions are typically mass analysed in either the linear ion trap or in the Orbitrap to generate a product ion spectrum. The sequence of scanning to produce a mass spectrum, collision induced dissociation and product ion scanning is repeated sequentially throughout the LC run. The peptides selected for fragmentation are most commonly the ions in the mass spectrum with the highest intensities. This approach is known as data-dependent acquisition (DDA) since it chooses for CID

and product ion scanning the ions giving the most intense signals and is generally used when an unbiased identification of peptides is required.

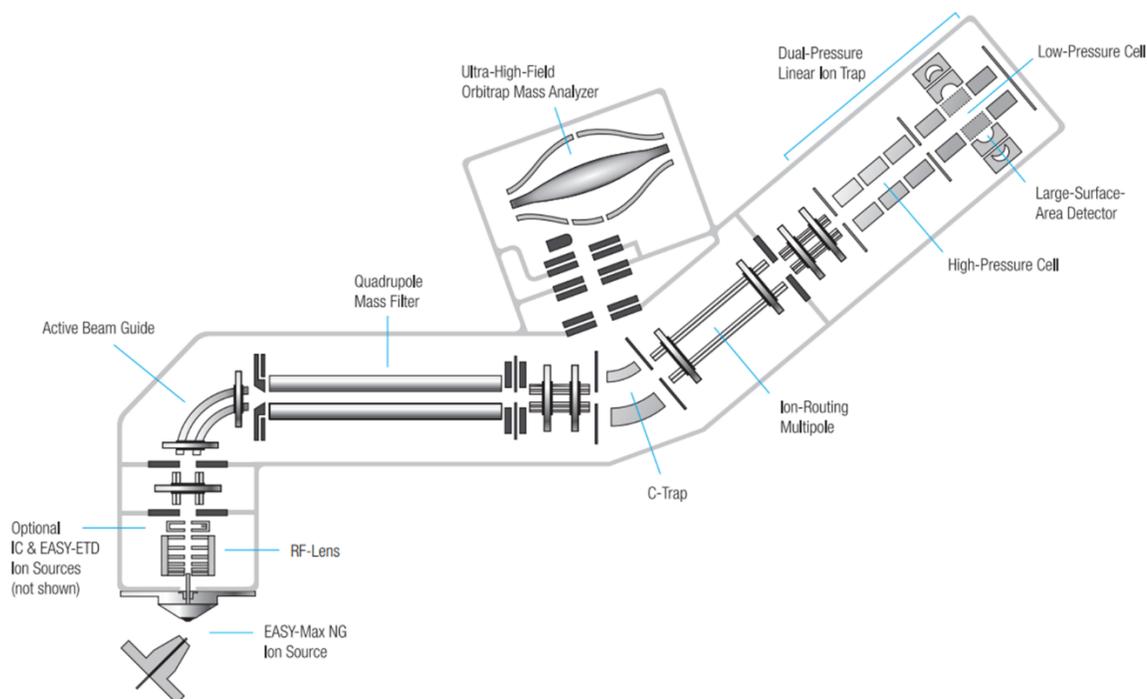


Figure 1.10/ Schematic view of the Orbitrap Fusion™ Tribrid™ (adapted from Orbitrap Fusion™ Tribrid™ manual, Thermo Scientific)

CID fragmentation in the Orbitrap Fusion™ Tribrid™ can be set up to take place in different parts of the instrument with different outcomes. In the ion-routing multipole, ions, previously accelerated, collide with an inert gas, converting a proportion of their translational energy to vibrational energy which causes fragmentation. This experimental arrangement has been called higher energy collision-induced dissociation (HCD) by Thermo. It is a CID experiment using a slightly higher energy input than the alternative arrangement (see below), and fragments the peptide bond generating b- and y- type sequence ions (Figure 1.11). A lower energy input regime that the manufacturer refers to as CID, takes place in the linear ion trap of the Orbitrap Fusion™ Tribrid™. The disadvantage of the latter arrangement is that fragmentation efficiency is poor and so sequence ion intensities are compromised. While the collision energy put into the precursor ions is necessary to generate peptide product ions, one of the additional consequences is that several PTMs are bound to the peptide by bonds weaker than the peptide bond. These weaker peptide-PTM bonds (including

peptide-glycan bonds) fragment more readily than the peptide bonds. The result is that, while the composition of a PTM can still be derived from the difference between the  $m/z$  of the unmodified peptide and the  $m/z$  detected for the modified peptide in the mass spectrum, the PTM's location on the peptide remains ambiguous. In the linear ion trap, electron transfer dissociation (ETD) can be set up to take place. ETD uses fluoranthene radical anions to transfer an electron to positively charged peptide ions, which can be mass selected for a product ion experiment. Electron transfer to multiply protonated peptide precursors reduces the charge of the latter and generates an unstable radical cation (Syka et al., 2004). Importantly, the radical cation fragments without energy randomization, which avoids the loss of the PTM, and the modified peptide undergoes fragmentation typically at the N-C $\alpha$  bond of the peptide giving c- and z- type peptide sequence ions (Figure 1.11). ETD fragmentation by providing peptide sequence ions that avoid PTM loss makes it a very useful technique for analysing the site of modification with PTMs like glycosylation.

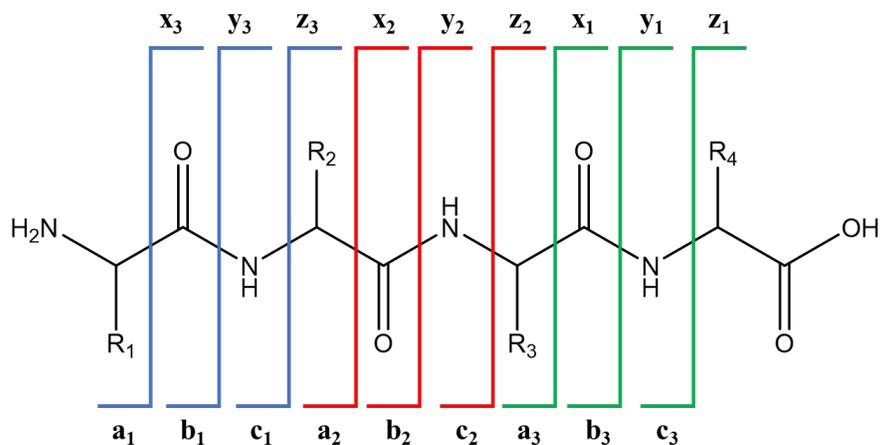


Figure 1.11/ Fragmentation of a peptide and resulting fragments. The ions are classed as either a, b or c if the charge is retained on the N-terminal fragment. The ions are classed as either x, y or z if the charge is retained on the C-terminal fragment (Johnson et al., 1987). A subscript indicates the number of amino acid residues in the fragment.

## 1.6 . Research aims

The recent discovery of the alditol Rbo5P in mammalian cells is fascinating, especially considering that the donor for Rbo5P (CDP-Rbo) was not among the monosaccharide glycosylation donors known in mammals (Ohtsubo and Marth, 2006). It is even more interesting to notice that only one protein,  $\alpha$ DG, is known to be modified with Rbo5P in mammalian cells and that Rbo5P has not been found in any other glycans than the core M3 glycan. The fact that overexpression of a truncated form of  $\alpha$ DG was required in all studies that identified Rbo5P in this glycan prompts the question of whether there are other glycoproteins in addition to  $\alpha$ DG that have Rbo5P moieties in their glycans but have so far evaded detection. Another question that remains unanswered is that of the presence of Rbo5P (and the structure of the tandem Rbo5P linker) in endogenous  $\alpha$ DG, which has not yet been demonstrated.

The overall aim of the research presented in this Thesis was therefore to develop and test a method for the enrichment and analysis of glycoproteins modified with Rbo5P, which would facilitate the glycoproteomic analysis of endogenous  $\alpha$ DG from mammalian cell lysates and the potential identification of novel Rbo5P-containing glycoconjugates.

Among the strategies available for the enrichment of glycoproteins, lectin affinity purification was not considered as an option since lectins that recognise Rbo5P selectively are not known. Metabolic oligosaccharide engineering of Rbo5P, on the other hand, would first require the synthesis of Rbo5P functionalised analogues to be fed to cells, and their inclusion into the metabolism of natural Rbo5P would not be guaranteed. For these reasons, the method chosen for the enrichment of glycoproteins bearing Rbo5P involves the direct modification of native glycoconjugates to generate a chemical reporter that can be conjugated with an enrichment tag. In chapter two, the description of the method is laid out. It consists of the chemical modification of proteins by periodate, which affects selectively Rbo5P and NeuAc residues in glycans, followed by bioorthogonal ligation with biotin. The biotinylated protein can then be enriched by streptavidin pulldown. In chapter two, the results are presented of testing the method on a sialylated model protein to establish the efficacy of protein tagging and enrichment, and detection of tagged sialylated peptides using mass spectrometric analysis. Chapter three describes the application of the method to  $\alpha$ DG which is recombinantly expressed in mammalian cells, to assess whether the procedure effectively labels Rbo5P residues and enables mass spectrometric identification of Rbo5P-containing glycopeptides. Lastly, chapter four describes the application of the method to mammalian cell lysates with the purpose of identifying novel proteins modified with Rbo5P and firmly establishing the presence of Rbo5P in endogenous  $\alpha$ DG.

## **Chapter 2: Method development for tagging, enriching, and analysing glycoproteins modified with ribitol-5-phosphate**

### **2.1 Introduction**

The recent discovery of Rbo5P in mammalian glycans raises the question of whether any other mammalian glycoproteins are modified with this alditol and, if so, how these can be identified. In order to answer this question, a powerful approach would be to use the tools of glycoproteomics for the selective enrichment of proteins modified with Rbo5P and their mass spectrometric identification (Chapter 1). The enrichment by MOE is an example of an enrichment method often used in a glycoproteomic experiment. MOE consists of metabolic labelling coupled with bioorthogonal reactions and has been used to selectively tag monosaccharides on glycoproteins in living cells and even whole organisms. Different functional groups have been used as chemical reporters for monosaccharides, such as aldehyde/ketone, alkyne, and azido groups. In particular, azido versions of GlcNAc (Vocadlo et al., 2003), GalNAc (Hang et al., 2003), and NeuAc (Saxon and Bertozzi, 2000) have been used successfully to label glycoproteins in living cells. MOE could be used to label and analyse proteins bearing Rbo5P but to achieve this, the synthesis of the modified alditol is required. There is no report describing the synthesis of Rbo5P with aldehyde/ketone, alkyne, or azido groups. Moreover, there is no evidence that these modified alditols would be tolerated by the endogenous metabolism of cells and animals.

The bioorthogonal reactions between the functional group and the chemical reporter allow the introduction of a tag on a glycoconjugate that can be used for both enrichment and detection purposes. These reactions can be exploited in MOE strategies but also in approaches where the chemical reporter is generated directly on the glycoconjugate. Specifically, a monosaccharide such as NeuAc with vicinal cis diols is susceptible to oxidation with sodium periodate to generate an aldehyde (Collins et al., 1997; Gahmberg and Andefsson, 1977). The mechanism of periodate oxidation of cis 1,2-diols is shown in Figure 2.1.

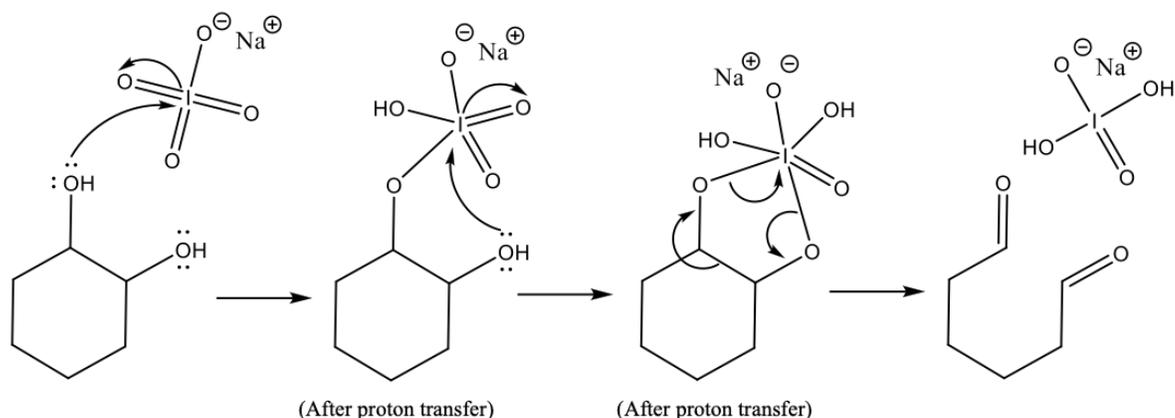


Figure 2.1/ Mechanism of sodium periodate oxidation of *cis* 1,2-diols

The oxidation modifies the native sialylated glycoprotein and allows an enrichment/affinity probe to be conjugated to the aldehyde. Exploiting the chemical properties of NeuAc has been used to label sialylated proteins on the surface of mammalian cells (Zeng et al., 2009). The method used 1mM sodium periodate for the oxidation step and aminoxy-biotin for conjugating oxidised sialylated proteins. Proteins labelled this way can also be digested into peptides and identified by mass spectrometry. Streptavidin-based enrichment allows biotinylated proteins to be recovered from a complex mixture, such as a cell lysate and eluted. The high-affinity bond between biotin and streptavidin (Stayton et al., 1999) requires harsh denaturing conditions to be broken, such as boiling in denaturing buffer. As a result, the eluted biotinylated proteins contain salts and detergent, which must be removed before protease digestion and mass spectrometric analysis. An alternative is performing protease digestion while biotinylated proteins are attached to the beads (Ramya et al., 2013). The drawback is that biotinylated peptides are not recovered after on-bead digestion; therefore, the information on the biotinylated site is lost. More recently, cleavable biotin tags for the conjugation with proteins have been used (Li et al., 2021), with the advantage of releasing proteins from streptavidin beads without using denaturing conditions.

The conditions used for tagging sialylated proteins (Zeng et al., 2009) can be, in theory, used also for the oxidation and biotinylation of proteins modified with Rbo5P since the alditol has *cis* vicinal diols. In the literature, the oxidation of Rbo5P-containing glycoconjugates was reported (Praisman et al., 2016). In Praissman's work, a truncated version of  $\alpha$ DG (amino acids 313-340) was overexpressed in mammalian cells. The sequence contains the "TPT" motif bearing core M3 glycans extended with Rbo5P and matriglycan. The mass spectrometric analysis of the glycopeptide was preceded by its oxidation with sodium periodate. The purpose was to remove the

matriglycan since its elevated molecular weight would hamper the detection of the glycopeptides by mass spectrometry. The analysis showed that site T317 of  $\alpha$ DG has a core M3 glycan with a cleaved Rbo5P between C2 and C3 or C3 and C4, as shown in Figure 1.8. Based on these observations, it would be possible to generate aldehydes on proteins modified with Rbo5P and successively conjugate them with aminoxy-biotin, similarly to the NeuAc.

## 2.2 Aims

The oxidation of monosaccharides with vicinal cis diols has been effectively used to introduce an aldehyde group on glycoproteins which allows their conjugation with probes and purification from complex mixtures. This strategy has, however, not yet been used for the labelling and enrichment of ribitol-containing glycoproteins. This chapter describes the development of a method for oxidising, tagging, enriching and analysing the standard protein fetuin. In the absence of a standard protein modified with Rbo5P, the sialylated protein fetuin was used since the conditions required for the oxidation of NeuAc and Rbo5P were expected to be the same. Fetuin was oxidised with sodium periodate, conjugated with aminoxy-biotin, purified using streptavidin agarose beads, and digested into peptides which were then analysed by mass spectrometry. The identification of fetuin after enrichment could be achieved by digestion with a protease and analysis of the peptides released from the beads. Nevertheless, the peptides carrying biotinylated NeuAc, still attached to the streptavidin agarose beads, would not be identified. For this reason, biotinylated fetuin was first recovered after enrichment and then digested into peptides. This method, when applied to glycoproteins modified with Rbo5P, will allow the analysis of the specific glycopeptides carrying Rbo5P.

The first objective of the experiments presented here was to achieve successful enrichment of biotinylated fetuin and mass spectrometric identification of the sialylated and biotinylated glycopeptides. Furthermore, since the product ion spectrum of such a glycopeptide would be expected to give information on the fragmentation of a biotinylated glycan, the second objective was to establish a mass spectrometric marker for the presence of biotin that can be used for future identification of labelled glycans. The marker will be helpful for the interpretation of data from glycoproteins modified with Rbo5P but with unknown glycan structures.

## 2.3 Materials and Methods

### 2.3.1 Periodate oxidation and biotinylation

Bovine fetuin (Sigma) was reconstituted in 1X phosphate-buffered saline (PBS, Gibco) pH 7.4. Protein was treated with 1 mM sodium periodate (Sigma) for 0.5 h, 2 mM for 0.5 h, or 2 mM for 1.5 h in 1X PBS (pH 7.4) at 4 °C using an end-to-end rotator in the dark. 100 µg of protein was used for each reaction in a final volume of 500 µL. The reaction was quenched by the addition of glycerol to a final concentration of 1 mM. Samples were concentrated in 30 kDa molecular weight cut-off (MWCO) centrifugal concentrators (Vivaspin), washed three times with 1X PBS pH 7.4, and were ready for digestion and mass spectrometry analysis.

For biotinylation, samples prepared as above were washed twice with 1X PBS pH 6.7 using 30 kDa MWCO concentrators. Aminoxy-biotin (Biotium) was added to a final concentration of 100 µM and reactions were carried out in 1X PBS pH 6.7 plus 10 mM aniline for 3 hours on an end-to-end rotator at 4°C. Samples were washed three times with PBS pH 6.7 using 30 kDa MWCO concentrators to remove the excess biotin and stored at -20 °C for subsequent analysis by western blotting, digestion, and mass spectrometry. Desialylated fetuin was obtained by reconstituting fetuin in 50 mM sodium acetate pH 5.5 and incubating it with sialidase from *Arthrobacter ureafaciens* (Sigma) overnight. Desialylated fetuin was then buffer exchanged with PBS pH 7.4 using a 30 kDa MWCO concentrator and treated with sodium periodate and biotin as described above.

### 2.3.2 In-gel digestion

Proteins were mixed with Laemmli buffer (final concentration 62.5 mM Tris-HCl pH 6.8, 1% lithiumdodecyl sulphate (LDS), 10% glycerol, and 0.005% bromophenol blue, Bio-rad), supplemented with 50 mM dithiothreitol (DTT) and separated on a 4-20% SDS-PAGE gradient gel (Bio-Rad) for 5 min at 200 V. Gel were stained with InstantBlue Coomassie (Abcam) for 1 h. Protein bands were excised and destained by repetitive washes with 100 mM ammonium bicarbonate (AMBIC) pH 8.0, followed by acetonitrile (ACN). Protein bands were treated with 100 µL of 10 mM DTT in 100 mM AMBIC for 45 min at 56 °C to reduce disulfide bonds. Then 100 µL of a solution of 55 mM iodoacetamide (IAM) in 100 mM AMBIC was added, and the samples were incubated at room temperature for 0.5 h in the dark to carbamidomethylate cysteines. Gel pieces were washed thrice with 100 µL of 50 mM AMBIC followed by 120 µL of ACN. Enzymatic digestion was carried out with 40 ng of trypsin (expressed in *Pichia pastoris* Proteomics Grade, Sigma) in 40 µL of 10 mM AMBIC pH 8.0. Protein bands were incubated in the trypsin solution at

4 °C for 2 h, and then 200 µL of 10 mM AMBIC pH 8.0 was added to cover the bands; the bands were incubated in the trypsin/AMBIC mixture for 18 h at 37 °C. The supernatant containing released peptides was collected. Gel pieces were dehydrated with ACN, and the supernatant was added to the previous aqueous one. Peptide mixtures were dried by using a centrifugal concentrator (SpeedVac) at room temperature concentration, then redissolved in 30 µL of 0.1% trifluoroacetic acid (TFA) for mass spectrometric analysis.

### **2.3.3 Streptavidin pulldown**

Biotinylated fetuin was incubated with 50 µL of streptavidin agarose slurry (Thermo) previously washed with 500 µL PBS pH 7.4 and incubated at room temperature using an end-over-end rotator for 0.5 h. After centrifugation at 500 x g, the supernatant was removed from the agarose and discarded, and the agarose was washed three times with 500 µL PBS pH 7.4. The agarose was then incubated with 2X Laemmli buffer for 10 min at 100 °C to release bound proteins. The supernatant was collected for western blot analysis.

### **2.3.4 Western blotting**

Samples were separated on a 4-20% SDS-PAGE gradient gel (Bio-Rad) for 40 min at 200 V. The gel was then transferred to a PVDF membrane (0.2 µm pore size) by wet transfer using 1 L Towbin buffer (25 mM Tris, 192 mM glycine, 10 % (v/v) methanol pH 8.3) at a constant voltage of 45 V for 1 h. The membrane was blocked in blocking buffer (50 mM Tris-HCl, 150 mM NaCl, 0.25% (w/v) BSA, and 0.05 % (v/v) NP-40, pH 7.4, for 30 min at room temperature, then incubated with HRP-conjugated anti-biotin (0.05 µg/mL, GeneTex) for 1 h at room temperature. The membrane was washed with 10 mL of blocking buffer thrice for 10 min. Proteins were detected by enhanced luminescence (GERPN2236 ECL kit from Sigma) using a G: BOX Chemi system (Syngene).

### **2.3.5 Mass spectrometric analysis**

Samples obtained after in-gel digestion were loaded onto a nanoAcquity UPLC system (Waters) equipped with a nanoEase M/Z Symmetry 100 Å, C<sub>18</sub> 5 µm trap column (180 µm x 20 mm, Waters) and a PepMap, 2 µm, 100 Å, C<sub>18</sub> RSLC nanocapillary column (75 µm x 50 cm, Thermo). The trap wash solvent was 0.1% (v/v) aqueous trifluoroacetic acid, and the trapping flow rate was 10 µL/min. The trap was washed (to waste) for 5 min before switching flow to the capillary column. The separation used gradient elution of solvent A (H<sub>2</sub>O + 0.1% TFA) and solvent B (ACN + 0.1%

TFA). The flow rate for the capillary column was 330 nL/min. The linear multi-step gradient profile was: 1-5% B over 1 min, 5-30% B over 32 min, 30-90% B over 5 min, then proceeded to wash with 90% solvent B for 2 min. The column was returned to initial conditions and re-equilibrated for 15 min before subsequent injections. The nanoLC system was interfaced with an Orbitrap Fusion™ Tribrid™ mass spectrometer (Thermo Scientific) with an Easy-Spray ionisation source (Thermo). Positive ESI-MS and product ion mass spectra were acquired using the Xcalibur software (version 4.6, Thermo). The effluent from the column was fed into the mass spectrometer starting 5 min after the beginning of the LC gradient. Instrument source settings were: ion spray voltage 2,300 V; sweep gas 0 Arb; ion transfer tube temperature 275°C. ESI mass spectra were acquired in the Orbitrap with 120,000 resolution; scan range:  $m/z$  350-2000; AGC target,  $1e^6$ ; max fill time, 50 ms; data type, profile. HCD spectra were acquired in the Orbitrap specifying: quadrupole isolation, isolation window,  $m/z$  2; activation type, HCD; collision energy, 30%; scan range, normal; Orbitrap resolution, 30,000; first mass,  $m/z$  110; AGC target,  $5e^4$ ; max injection time, 60 ms; data type, profile. Data-dependent acquisition was performed in top speed mode using a 3 s cycle, with the most intense precursors selected. Dynamic exclusion was performed for 10 s post precursor selection and a minimum threshold for fragmentation was set at  $5e^4$ .

### 2.3.6 Data analysis

Mass spectrometric data acquired from the analysis of biotinylated digested fetuin was searched for glycopeptides through the Byonic software (Protein Metrics Inc). The search criteria were set as follows: KR (trypsin) cleavage sites;  $\leq 1$  missed cleavages and semi-specific cleavage allowed; error tolerances for the precursor of 3 ppm and 0.05 Da for fragment ions; carbamidomethyl cysteine as a fixed modification, oxidation of methionine (common 2), and pyro-Glu from glutamine (rare 1) as variable modifications (max common set to 3 and max rare set to 4); false discovery rate (FDR) was set to 1%; possible glycan composition was set as Hex<sub>5</sub>HexNAc<sub>4</sub>, Hex<sub>5</sub>HexNAc<sub>4</sub>NeuAc<sub>1</sub>, Hex<sub>5</sub>HexNAc<sub>4</sub>NeuAc<sub>2</sub>, Hex<sub>6</sub>HexNAc<sub>5</sub>, Hex<sub>6</sub>HexNAc<sub>5</sub>NeuAc<sub>1</sub>, Hex<sub>6</sub>HexNAc<sub>5</sub>NeuAc<sub>2</sub> and Hex<sub>6</sub>HexNAc<sub>5</sub>NeuAc<sub>3</sub> (on Asn), or Hex<sub>1</sub>HexNAc<sub>1</sub>NeuAc<sub>1</sub>, Hex<sub>1</sub>HexNAc<sub>1</sub>NeuAc<sub>2</sub> and Hex<sub>2</sub>HexNAc<sub>2</sub>NeuAc<sub>2</sub> (on Ser/Thr). Oxidation and biotinylation of NeuAc were implemented in the search by substituting the mass of NeuAc in the glycans with either 229.05 (for oxidised NeuAc at carbon 7) or 612.26 (for oxidised and biotinylated NeuAc).

### 2.3.7 Mass spectra annotation

Glycan fragments obtained after HCD fragmentation of glycopeptides and reported in the product ion spectra correspond to B-ions (Domon and Costello, 1988). These ions all have +1 charge and are represented here with conventional symbols for monosaccharides (Varky et al., 2009) added of the drawing  which represents their fragmented and charged state.



Figure 2.2 / HCD fragmentation on a glycan. In positive ion mode, the fragments result from protonation of the glycosidic bond, which is subsequently broken to generate the B-ion (oxonium ion). Alternatively, cleavage of the glycosidic bond is accompanied by a proton transfer which generates the Y-ion.

B ions, also referred to as oxonium ions, derived from monosaccharides like HexNAc and NeuAc are particularly intense in the product ion spectrum and can be used to distinguish the spectrum of a glycopeptide from that of a peptide. To show the effect of sodium periodate treatment on sialylated glycopeptides the oxonium ion for NeuAc was monitored. Unmodified NeuAc generates an oxonium ion (minus H<sub>2</sub>O) observed at *m/z* 274.09 while oxidised NeuAc would produce an equivalent signal at *m/z* 260.08 if carbon 8 is oxidised, or 230.07 if carbon 7 is oxidised (Figure 2.3).

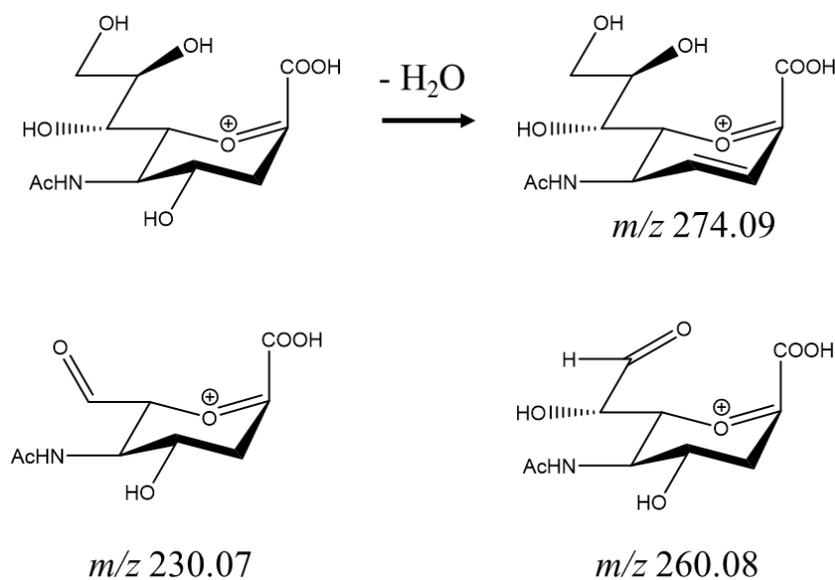


Figure 2.3 / Oxonium ions generated from monodehydrated NeuAc (*m/z* 274.09), NeuAc with oxidised carbon 7 (*m/z* 230.07) and NeuAc with oxidised carbon 8 (*m/z* 260.08)

All the product ion spectra from the sodium periodate-treated fetuin analysis were searched for the *m/z* values of unmodified or oxidised NeuAc. The results of the experiment performed using 1 mM sodium periodate incubated for 0.5 h are shown in Figure 2.4.

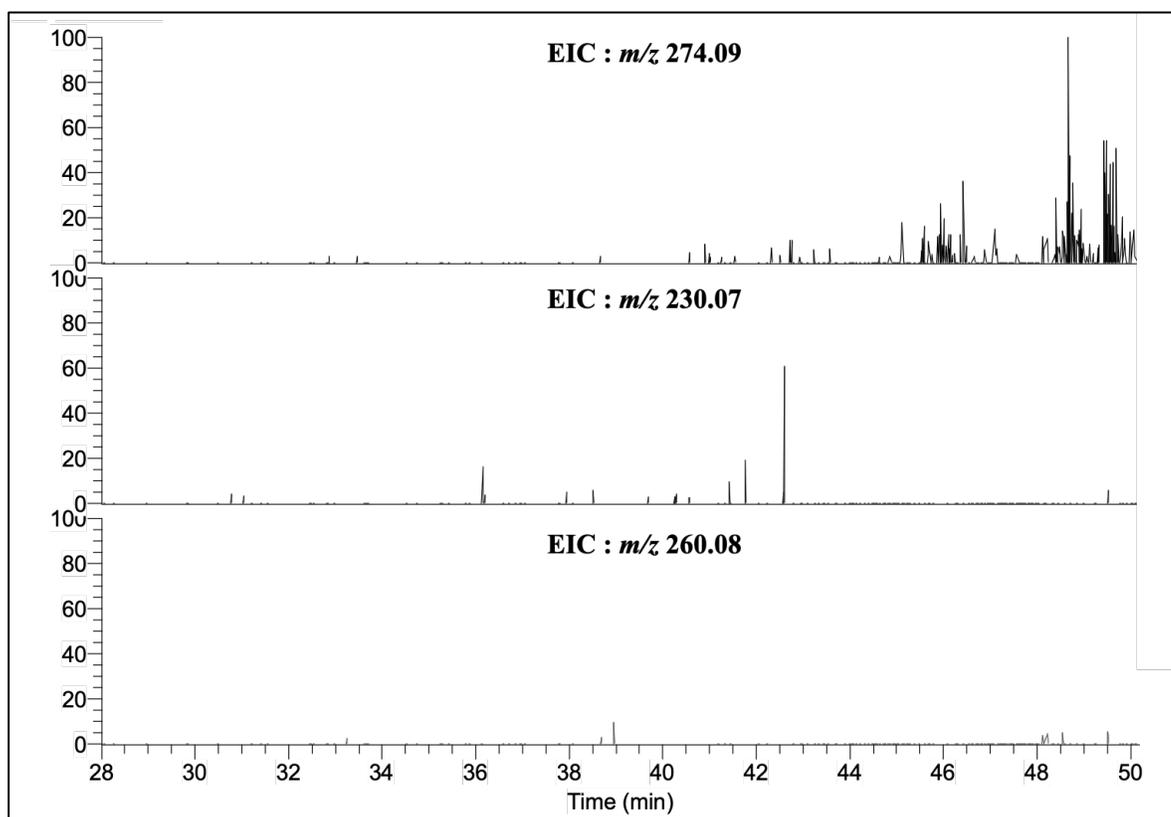


Figure 2.4 / Extracted ion chromatograms (EICs) derived from the product ion spectra of sodium periodate-treated fetuin (1 mM for 0.5 h). The  $m/z$  values used to generate the EICs correspond to the dehydrated oxonium ion for NeuAc ( $m/z$  274.09 unmodified, 230.07 oxidised at carbon 7, 260.08 oxidised at carbon 8). Relative abundance is set at the same scale for all chromatograms.

The extracted ion chromatograms (EICs) for  $m/z$  230.07 or 260.08 show some signals, but most come from the EIC for  $m/z$  274.09. An example of a glycopeptide observed to be affected by oxidation is RPTGEVYDIEIDTLETTCHVLDPTPLANCSVR bearing the glycan with composition Hex<sub>6</sub>HexNAc<sub>5</sub>NeuAc<sub>3</sub>. The mass spectrum with the glycopeptide's precursor is shown in Figure 2.5.

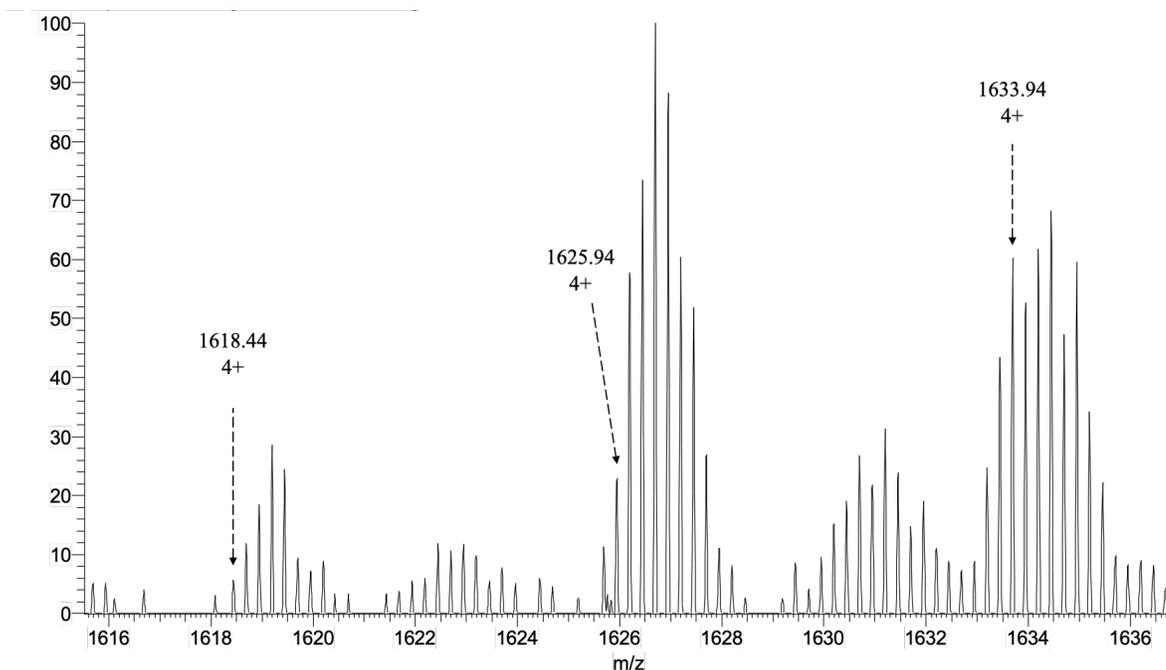


Figure 2.5 / Mass spectrum (MS1) from the analysis of bovine fetuin treated with sodium periodate (1 mM for 0.5 h) and digested with trypsin. The mass of the peak at  $m/z$  1633.94 with  $z=4$  corresponds to the monoisotopic mass calculated for the peptide RPTGEVYDIEIDTLETTCHVLDPTPLANCSVR bearing 5 HexNAc, 6 Hex, 3 NeuAc and 2 carbamidomethylated cysteines. The mass of the peaks at  $m/z$  1618.44 and 1625.94 correspond to the monoisotopic mass of the same glycopeptide having one unit of NeuAc oxidised at carbon 7 or 8, respectively.

The peptide RPTGEVYDIEIDTLETTCHVLDPTPLANCSVR is predicted to be *N*-glycosylated, and Hex<sub>5</sub>HexNAc<sub>6</sub>NeuAc<sub>3</sub> is one of the glycans reported to modify the site N99 (Windwarder and Altmann, 2014). In Figure 2.5, three different isotopic patterns can be interpreted as belonging to the glycopeptide non-oxidised ( $m/z$  1633.94) or oxidised ( $m/z$  1618.44 and 1625.94) at only one of the three units of NeuAc. Moreover, the oxidation can occur at either the carbon 7 ( $m/z$  1618.44) or the carbon 8 ( $m/z$  1625.94) of NeuAc. This is an example of how oxidation with sodium periodate generates a mixture of oxidised and non-oxidised species. The concentration of sodium periodate and incubation time of the reagent with the protein were questioned as factors determining the completeness of the reaction. At the same time, the freshness of the sodium periodate was also thought to play a role in the reaction. A new experiment was performed using a fresh batch of sodium periodate, and the mass spectrometric analysis was repeated. The results of the experiment performed using 1 mM of fresh sodium periodate incubated for 0.5 h are shown in Figure 2.6.

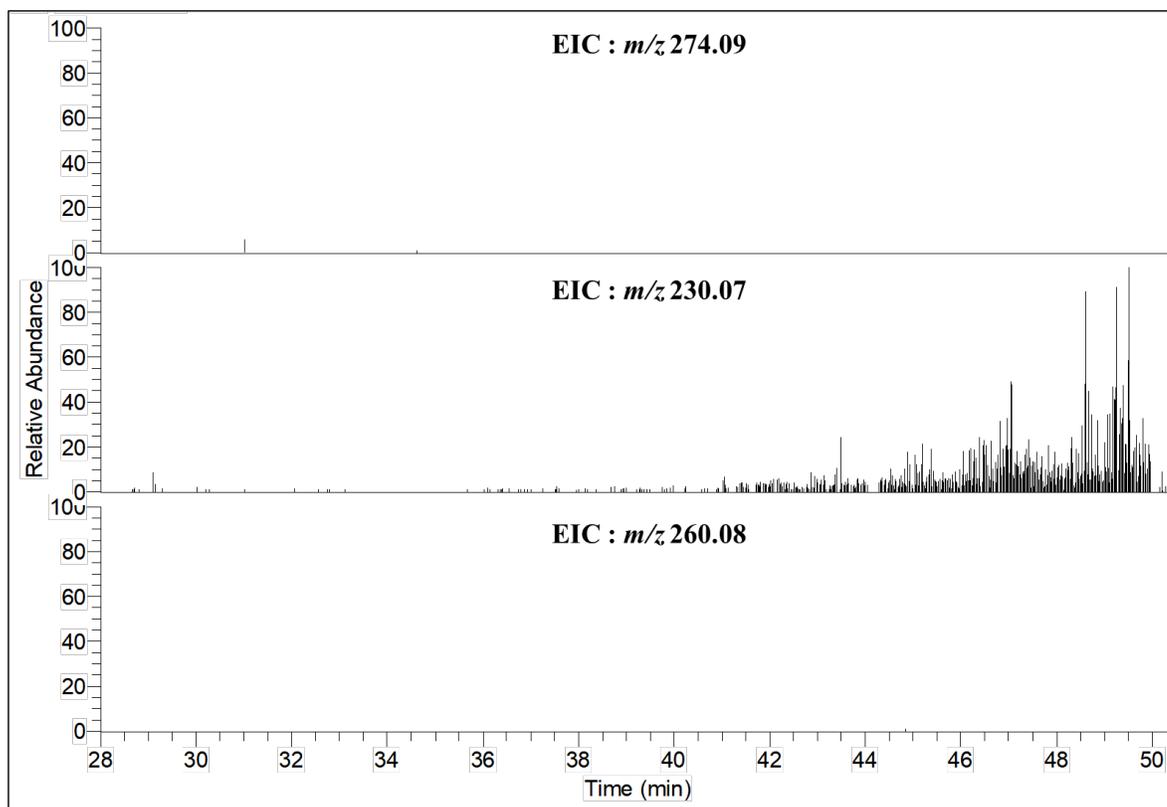


Figure 2.6 / Extracted ion chromatograms (EICs) derived from the product ion spectra of sodium periodate-treated fetuin (1 mM for 0.5 h, fresh batch of reagent). The  $m/z$  values used to generate the EICs correspond to the dehydrated oxonium ion for NeuAc ( $m/z$  274.09 unmodified, 230.07 oxidised at carbon 7, 260.08 oxidised at carbon 8). Relative abundance is set at the same scale for all chromatograms.

The EIC for  $m/z$  230.07 is the only among the three EICs to show a plethora of signals which suggest that the oxidation of NeuAc goes to completion and occurs uniquely between carbon 7 and carbon 8 of the monosaccharide, resulting in the cleavage of the bond between these two atoms and the generation of an aldehyde at position 7. The only parameter changed in this experiment compared to the one described above (Figure 2.4) was the freshness of the sodium periodate used, which suggests the importance of a fresh reagent for the success of the oxidation reaction.

In Figure 2.6, each of the product ion spectra that contributed to the EIC for  $m/z$  230.07 may be interpreted as the fragmentation spectrum of a sialylated fetuin glycopeptide after NeuAc oxidation. An example of such a spectrum is shown in Figure 2.7.

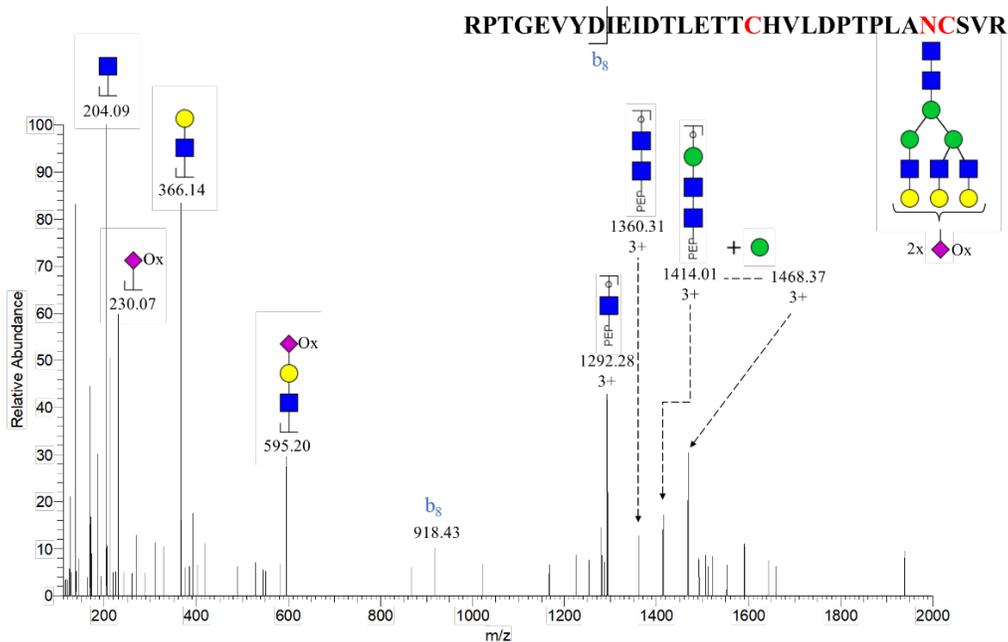


Figure 2.7 / HCD product ion spectrum of a tryptic glycopeptide from bovine fetuin treated with sodium periodate (1 mM for 0.5 h). The precursor ion  $[M+5H]^{5+}$   $m/z$  1224.32 has a monoisotopic mass corresponding to that calculated for the peptide RPTGEVYDIEIDTLETTCHVLDPTPLANCSVR bearing 5 HexNAc, 6 Hex, 2 NeuAc oxidised at carbon 7 and 2 carbamidomethylated cysteines ( $\Delta=8$  ppm). Ions corresponding in  $m/z$  to a glycan fragment bearing the peptide are represented with the tag “PEP”. Oxidised NeuAc is represented with the tag “Ox”. Product ions have a 1+ charge unless specified otherwise (see section 2.3.7).

The product ion spectrum shows peaks corresponding to fragments of the *N*-glycan attached to the peptide, fragments from the backbone of the peptide (b/y ions), and oxonium ions for GlcNAc and oxidised NeuAc. The oxidation of NeuAc at carbon 7 is reflected in the presence of the ions at  $m/z$  230.07 (NeuAc-Ox) and  $m/z$  595.20 (GlcNAc<sub>1</sub>Gal<sub>1</sub>NeuAc-Ox<sub>1</sub>).

The experiments performed with higher concentrations of sodium periodate and different incubation times showed no product ion spectra with signals for unmodified NeuAc and NeuAc oxidised at carbon 8 (data not shown). This result is similar to the one observed for the experiment performed with 1 mM sodium periodate incubated for 0.5 h (Figure 2.6). To compare the results from the different experiments, the EICs for  $m/z$  230.07 obtained from the samples treated in different conditions are shown in Figure 2.8.

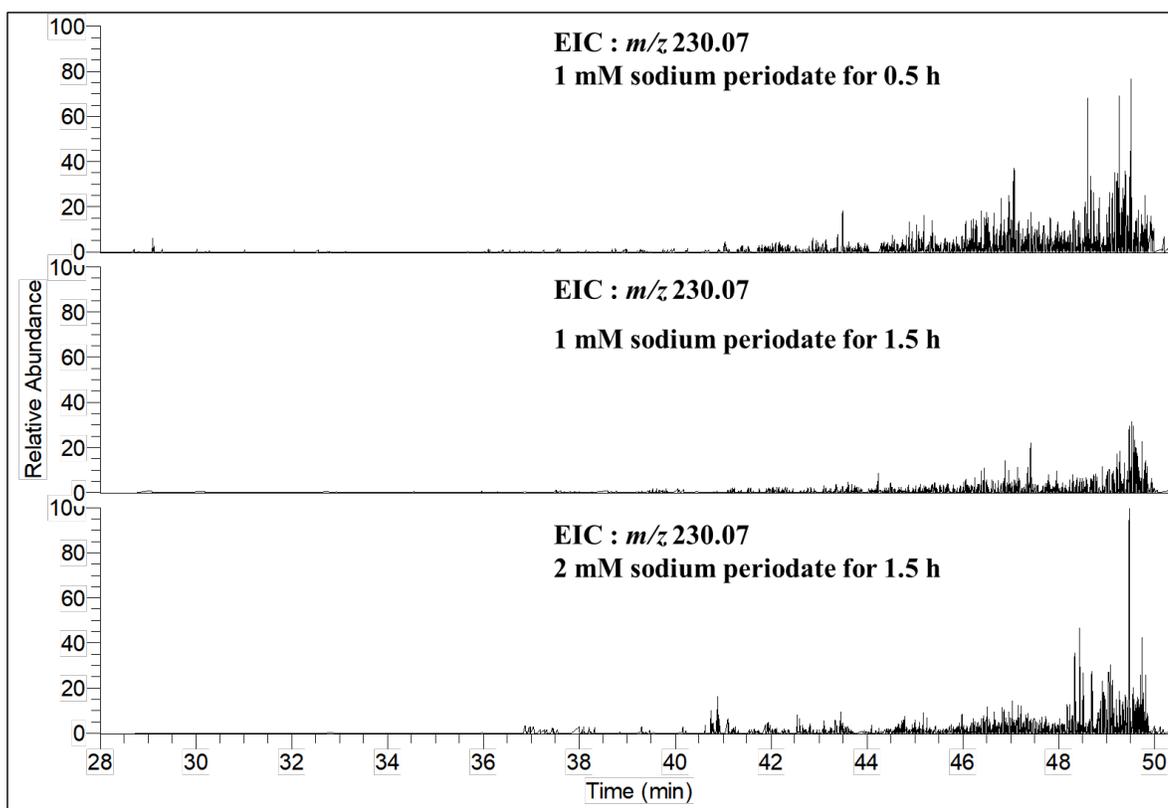


Figure 2.8 / EICs derived from the product ion spectra of a range of conditions for sodium periodate treatment of fetuin. The  $m/z$  used to generate the EICs refers to the oxonium ion for oxidised NeuAc oxidised at carbon 7 ( $m/z$  230.07). Relative abundance is set at the same scale for all chromatograms.

Product ion spectra containing signals for oxidised NeuAc were obtained from fetuin treated under all three conditions. Variability in terms of intensities of each spectrum is expected and still, the same oxidised sialylated glycopeptides were detected in all conditions. Since using a higher concentration of sodium periodate or higher incubation time did not improve the number of glycopeptides identified carrying oxidised NeuAc, the lowest concentration of sodium periodate (1 mM) and shortest time of incubation (0.5 h) were used for the following experiments. These conditions, used here for a single purified protein, were the same used for the oxidation performed on the proteins present on the surface of cells (Zeng et al., 2009). This suggests that the same conditions could be used for a more complex sample, like a cell lysate, which will be treated in

## 2.4.2 Aminoxy-biotin conjugates to oxidised NeuAc on glycoproteins

Oxidised NeuAc on fetuin was conjugated with aminoxy-biotin with the purpose of labelling the monosaccharide with a detectable tag that can also be retrieved by affinity purification. The resulting product is shown in Figure 2.9.

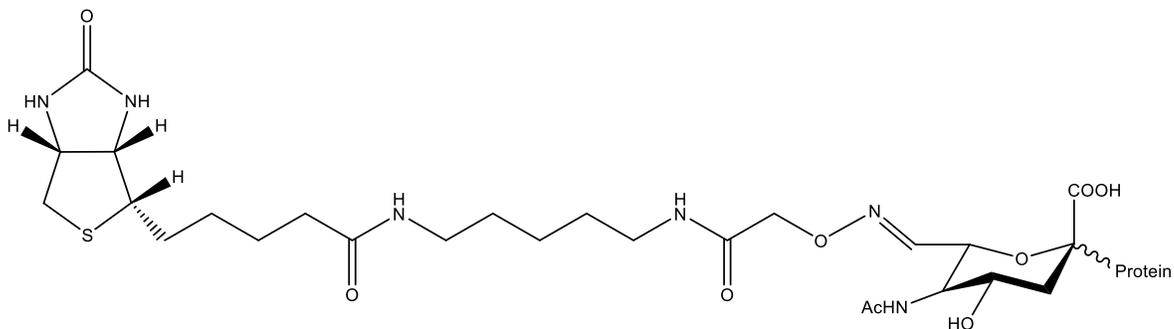
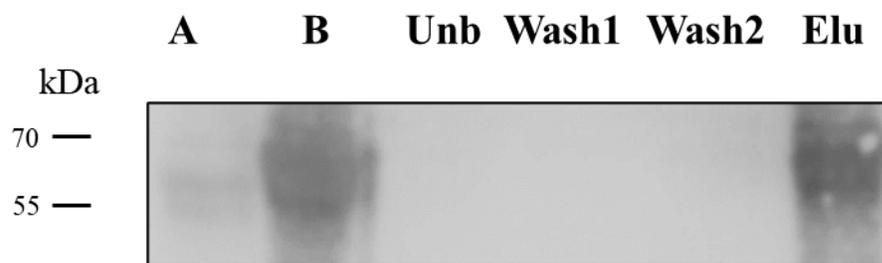


Figure 2.9 / Oxidised NeuAc conjugated with aminoxy-biotin

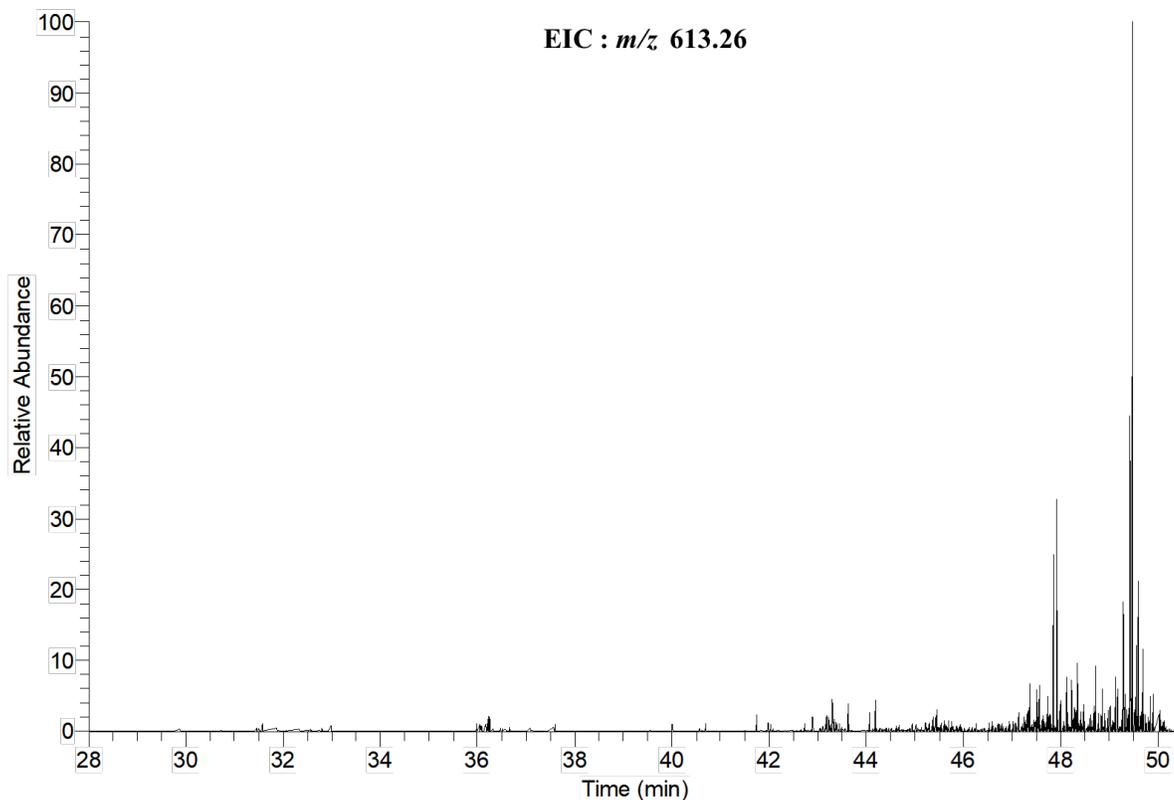
Biotin is commonly used to tag proteins because it can be efficiently purified from a complex mixture using streptavidin pulldown and can be detected by western blotting or microscopy using anti-biotin or HRP-streptavidin. In the case of this study, biotinylated fetuin is detected by western blotting as an intensely stained band at ~ 65 kDa using anti-biotin (Figure 2.10 lane-B). A negative control was performed to show that biotinylation depends on the presence of NeuAc. For this purpose, fetuin was desialylated with a sialidase before the sodium periodate treatment, biotinylated with aminoxy-biotin, and analysed by western blotting (Figure 2.10 lane-A). As expected, sialidase treatment prevented biotinylation as evidenced by the absence of a signal on western blot.

Biotinylated fetuin was incubated with streptavidin agarose to demonstrate that it could be retrieved using streptavidin pulldown. The flow-through recovered after incubation (Unb), the following washes of the agarose (Wash 1/2) and the proteins eventually eluted from the agarose (Elu) were analysed by western blotting and are shown in Figure 2.10. A band at ~ 65 kDa is detected in the lane with the Elu sample but is absent in the Unb and Wash 1/2 lanes. The result is indicative of the successful recover of biotinylated fetuin using streptavidin pulldown.



*Figure 2.10 / Western blot analysis of biotinylated bovine fetuin and its retrieval using streptavidin pulldown. After sodium periodate and biotinylation treatment, fetuin is detected by binding to HRP-conjugated anti-biotin (B) while it is not detected if it is incubated with sialidase before oxidation (A). Biotinylated fetuin was incubated with streptavidin agarose. Aliquots from the flow-through (Unb) washes (Wash1/2) and elution (Elu) are shown. HRP-conjugated anti-biotin was used as the antibody.*

Biotinylated fetuin obtained after streptavidin pulldown was then digested with trypsin and analysed by LC-MS/MS. While on-bead digestion would ease sample preparation, this would prevent direct detection and characterisation of the biotinylated glycopeptide. Therefore, proteins were first eluted before trypsin digestion. The predicted  $m/z$  value for oxidised and biotinylated NeuAc as an oxonium ion is 613.26. Therefore, all product ion spectra from the biotinylated fetuin analysis were searched for the presence of an ion at  $m/z$  613.26, and the results were plotted as an EIC for  $m/z$  613.26 (Figure 2.11).



*Figure 2.11 / EIC derived from the product ion spectra of sodium periodate and aminoxy-biotin-treated bovine fetuin. The  $m/z$  value used to generate the EIC refers to the oxonium ion for oxidised and biotinylated NeuAc (613.26).*

Several signals for modified peptides eluting at different retention times can be observed in the EIC for  $m/z$  613.26 and an example of a product ion spectrum is shown in Figure 2.12.

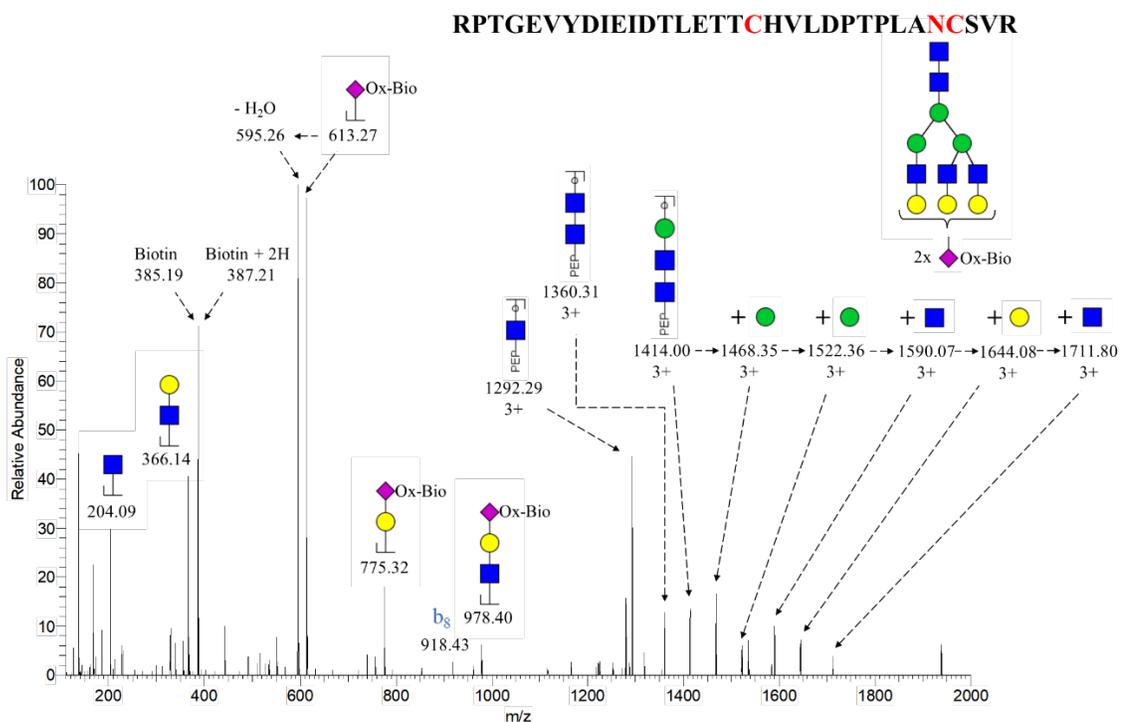


Figure 2.12 / HCD product ion spectrum of a tryptic glycopeptide from bovine fetuin treated with sodium periodate and aminoxy-biotin. The precursor ion  $[M+5H]^{5+}$  at  $m/z$  1378.01 has a monoisotopic mass corresponding to that calculated for the peptide RPTGEVYDIEIDTLETTCHVLDPTPLANCSVR plus 5 HexNAc, 6 Hex, 2 NeuAc oxidised and biotinylated at carbon 7, and 2 carbamidomethylated cysteines ( $\Delta=7$  ppm). Ions corresponding to the  $m/z$  of the peptide plus a glycan fragment are represented with the tag “PEP”. Oxidised and biotinylated NeuAc is represented with the tag “Ox-Bio”. Ions have a 1+ charge unless specified otherwise (see section 2.3.7).

The signal at  $m/z$  613.27 can be considered to be a marker for the presence of NeuAc in a glycopeptide after its oxidation and biotinylation. The monodehydrated oxonium ion at 595.26  $m/z$  is thus similarly a marker because it is presumably derived from the ion at  $m/z$  613.27. Additionally, two new intense ions appeared at  $m/z$  385.19 and 387.21. These correspond to biotin fragments formed by loss from biotinylated NeuAc-containing species (the ion at  $m/z$  385.19 = C<sub>17</sub>H<sub>29</sub>O<sub>4</sub>N<sub>4</sub>S<sub>1</sub>,  $\Delta=2$  ppm and  $m/z$  387.21 = C<sub>17</sub>H<sub>31</sub>O<sub>4</sub>N<sub>4</sub>S<sub>1</sub>,  $\Delta=1$  ppm).

The data obtained from the mass spectrometric analysis of biotinylated fetuin were searched, using bioinformatic software, for other glycopeptides belonging to fetuin. The list of the glycopeptides identified with the relative glycans is reported in Table 2.1.

Peptide sequence	Glycan composition
<sup>145</sup> LCPDCPLLAPL <b>N</b> DSR <sup>159</sup>	Hex <sub>5</sub> HexNAC <sub>4</sub> NeuAc-Bio <sub>1</sub>
	Hex <sub>5</sub> HexNAC <sub>4</sub> NeuAc-Bio <sub>2</sub>
	Hex <sub>5</sub> HexNAC <sub>4</sub> NeuAc-Ox <sub>1</sub> NeuAc-Bio <sub>1</sub>
	Hex <sub>6</sub> HexNAC <sub>5</sub> NeuAc-Bio <sub>2</sub>
	Hex <sub>6</sub> HexNAC <sub>5</sub> NeuAc-Ox <sub>1</sub> NeuAc-Bio <sub>1</sub>
<sup>72</sup> RPTGEVYDIEIDTLETTCHVLDPTPLA <b>N</b> CSVR <sup>103</sup>	Hex <sub>5</sub> HexNAC <sub>4</sub> NeuAc-Bio <sub>1</sub>
	Hex <sub>5</sub> HexNAC <sub>4</sub> NeuAc-Bio <sub>2</sub>
	Hex <sub>5</sub> HexNAC <sub>4</sub> NeuAc-Ox <sub>1</sub> NeuAc-Bio <sub>1</sub>
	Hex <sub>6</sub> HexNAC <sub>5</sub> NeuAc-Bio <sub>2</sub>
	Hex <sub>6</sub> HexNAC <sub>5</sub> NeuAc-Ox <sub>1</sub> NeuAc-Bio <sub>1</sub>
<sup>160</sup> VVHAVEVALATFNAESNGSYLQLVEISR <sup>176</sup>	Hex <sub>6</sub> HexNAC <sub>5</sub> NeuAc-Ox <sub>1</sub> NeuAc-Bio <sub>2</sub>
	Hex <sub>6</sub> HexNAC <sub>5</sub>

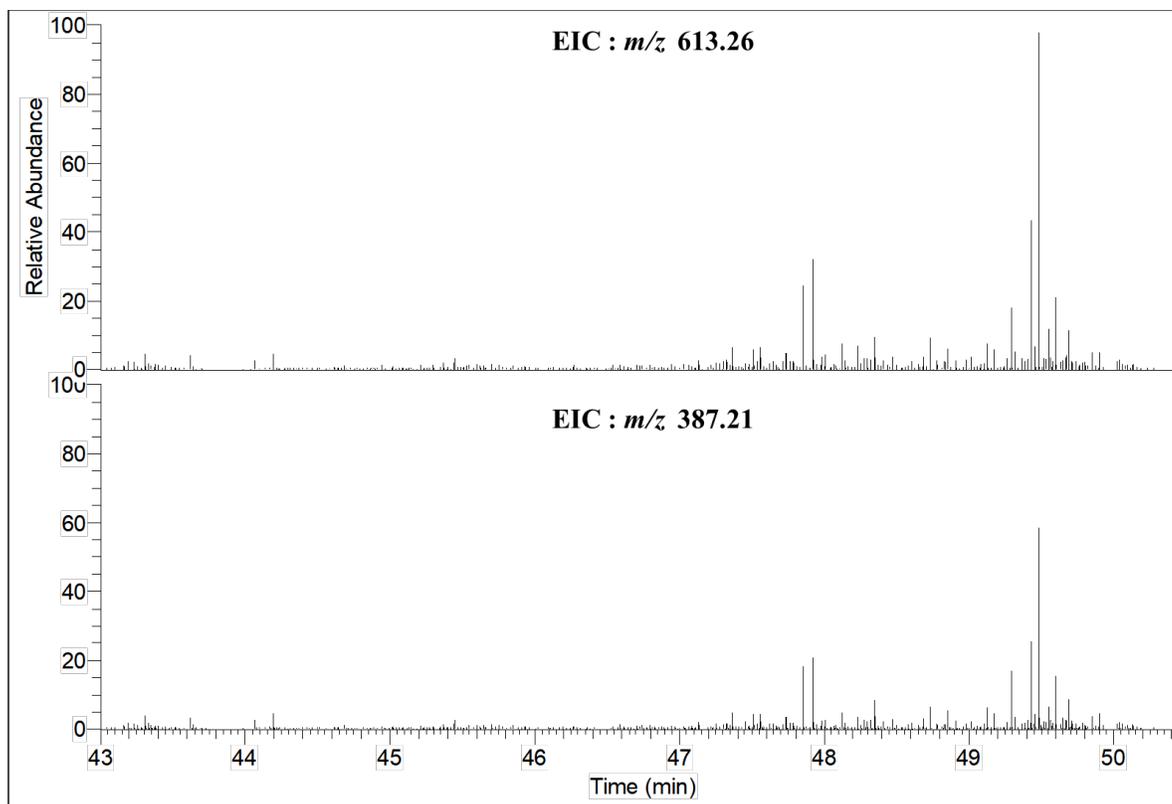
Table 2.1 / List of glycopeptides identified by LC-MS/MS analysis of the tryptic digest of oxidised (1 mM sodium periodate for 0.5 h) and biotinylated fetuin. N-glycosylation sites are in bold. NeuAc-Ox denotes NeuAc oxidised at carbon 7, while NeuAc-Bio denotes NeuAc oxidised and conjugated with biotin.

Bovine fetuin has three N-glycosylation sites: N99, N156, and N176. These are all observed in the data bearing N-glycans. Fetuin also has five O-glycosylation sites: S271, T280, S282, S296, and S341. Glycopeptides bearing O-glycans were not detected. This result can be explained if considering that four of the five O-glycosylation sites (S271, T280, S282, and S296) are all within a single tryptic glycopeptide with 61 amino acids. The predicted molecular weight of such peptide bearing biotinylated O-glycans is high and might have been outside the optimal *m/z* range of analysis used to acquire mass spectrometric data. Lastly, site S341 is expected to have a very low occupancy percentage (Windwarder and Altmann, 2014), which would explain why it was not detected. Table 2.1 lists different glycoforms for the same glycopeptide. The NeuAc units in these glycoforms are all oxidised but not all biotinylated. The result suggests that the biotinylation reaction does not go to completion. Nevertheless, at least one biotin unit is found in each

glycopeptide bearing NeuAc, which is enough to detect the marker for biotin. The presence of this marker will be important for analysing samples that are more complex than a single pure protein.

## 2.4.2 Sodium periodate treatment and biotin conjugation are selective for NeuAc

Selectivity is one of the required features for tagging NeuAc on glycoproteins if the purpose is to exploit the tagging for selective retrieval. In other words, no other monosaccharide on a glycopeptide should be affected by the treatment. The incorporation of biotin is evidenced in the product ion spectrum of a biotin containing-peptide by the ion at  $m/z$  387.21 while the presence of biotinylated NeuAc is marked by the ion at  $m/z$  613.26. Based on this information, the EICs for the two ions were compared (Figure 2.13). If biotinylation occurred on a monosaccharide other than NeuAc there would be product ion spectra of peptides containing the ion at  $m/z$  387.21 but not the ion at  $m/z$  613.26. This was, however, not the case, suggesting the oxidation reaction was selective for NeuAc over the other monosaccharides present in the fetuin sample.



*Figure 2.13 / EICs derived from the product ion spectra of the sodium periodate (1 mM for 0.5 h) and biotin-treated bovine fetuin. The m/z values used to generate the EICs refer to the oxonium ion for oxidized and biotinylated NeuAc (m/z 613.26) and the fragment for biotin + 2H (m/z 387.21). Relative abundance is set at the same scale for all chromatograms.*

## 2.5 Conclusions

The purpose of the work presented in this chapter was to develop a method for the selective tagging, enrichment, and analysis of a sialylated protein. The same method can be used for glycoproteins modified with Rbo5P with the purpose of identifying new proteins modified with the alditol. The results show that incubation with 1 mM sodium periodate for 0.5 h is enough to oxidise NeuAc on a sialylated glycoprotein and generate an aldehyde on the monosaccharide. The oxidation occurs uniquely between carbon 7 and carbon 8 of the monosaccharide along with the cleavage of the bond between these two atoms. It is likely that the oxidation goes to completion as unmodified NeuAc was not detected in the HCD spectra after treatment. The oxidised sialylated protein was also successfully conjugated with aminoxy-biotin and retrieved with streptavidin agarose. The mass spectrometric data show that biotinylated and oxidised NeuAc has a diagnostic oxonium ion at  $m/z$  613.26. The results also show that biotin produces a diagnostic ion at  $m/z$  387.21 that appears in the product ion spectrum of glycopeptides after conjugation. This information is important for future experiments. In particular, the discovery of new glycoproteins modified with Rbo5P using the method described here relies on the detection of biotinylated and oxidised Rbo5P by mass spectrometry. Having a marker for the presence of biotin will aid future studies aimed at discovering new glycoproteins modified with Rbo5P, providing a way to detect the presence of biotin irrespective of the detection of the Rbo5P ion itself.

## Chapter 3: Development of a method for the overexpression and chemical tagging of recombinant $\alpha$ DG

### 3.1 Introduction

$\alpha$ DG is a heavily glycosylated protein, expressed as a polypeptide with a predicted molecular mass of ~72 kDa. The protease furin cleaves the N-terminus of  $\alpha$ DG from the protein after glycosylation is completed (Motoi Kanagawa et al., 2004; Singh et al., 2004). After furin cleavage, the remaining  $\alpha$ DG polypeptide backbone would have a predicted molecular mass of ~37 kDa; however, the apparent molecular weight of  $\alpha$ DG observed by western blotting ranges between ~120 and ~156 kDa depending on the type of tissue in which the protein is found (Barresi and Campbell, 2006).  $\alpha$ DG glycosylation has been intensively studied because of its relationship to congenital diseases named dystroglycanopathies (described in Chapter 1). In particular, disruption of the pathway responsible for the biosynthesis of  $\alpha$ DG O-mannosyl glycans is involved in developing muscular dystrophies and muscle-eye-brain diseases (Endo, 2015). The first O-mannosyl glycan identified in  $\alpha$ DG was in the protein obtained from bovine peripheral nerve and corresponds to a core M1 structure (Gal $\beta$ 1-4GlcNAc $\beta$ 1-2Man-O-Ser/Thr) capped with NeuAc (Chiba et al., 1997). The core M2 structure (Gal $\beta$ 1-4GlcNAc $\beta$ 1-6[Gal $\beta$ 1-4GlcNAc $\beta$ 1-2]Man-O-Ser/Thr) was first found in  $\alpha$ DG from rabbit skeletal muscle (Stalnaker et al., 2010). Lastly, the structure of the core M3 glycan (GalNAc $\beta$ 1-3GlcNAc $\beta$ 1-2[Phos6]Man-O-Ser/Thr) was described in recombinant  $\alpha$ DG expressed in HEK293 cells (Yoshida-Moriguchi et al., 2010) (see Figure 1.2).

$\alpha$ DG function relies on its O-mannosyl glycans and, in particular, on an oligo/polysaccharide made of alternating Xyl and GlcA named matriglycan ([Xyl $\alpha$ 1-3GlcA $\beta$ 1-3] $_n$ ) (Yoshida-Moriguchi and Campbell, 2014). Matriglycan is synthesised by the enzyme LARGE which presents two glycosyltransferase activities (Inamori et al., 2012). While it has been known for several years that  $\alpha$ DG glycosylation and function depend on LARGE (Barresi et al., 2004), for a long time the way matriglycan was attached to the O-mannosyl glycans of  $\alpha$ DG was unknown. In 2016, different groups identified the linker between core M3 glycan and matriglycan as a tandem repeat of Rbo5P (Kanagawa et al., 2016; Praissman et al., 2016; Yagi et al., 2016) by analysing the product of a recombinantly expressed peptide carrying the sites of glycosylation T317 and T319, at which matriglycan was known to be present. Furthermore, different glycoforms of the same peptide were identified as being modified with core M3 glycan and either one or two units of Rbo5P or a single unit of GroP.  $\alpha$ DG sites T317 and T319 are part of a  $^{317}\text{TPT}^{319}$  motif, which was shown to be essential for the core M3 glycosylation to take place. Another TPT motif is present in  $\alpha$ DG at

<sup>379</sup>TPT<sup>381</sup>, and the tryptic peptide it is found on (GAIQTPTLGPIQPTR) was shown to be modified with core M3 glycan (Yoshida-Moriguchi et al., 2010). The peptide could also potentially be modified with Rbo5P or GroP, but there is no evidence for this in the literature. It is interesting to notice that in preceding years several studies in the literature focused on the glycoproteomic analysis of peptides obtained from the digestion of endogenous (Nilsson et al., 2010; Stalnaker et al., 2010) or recombinant full-length  $\alpha$ DG (Harrison et al., 2012), but none of those studies was able to report the presence of Rbo5P or GroP. This observation suggests that glycopeptides modified with Rbo5P or Gro5P in  $\alpha$ DG are not very abundant and are difficult to detect using mass spectrometry. Specifically, Rbo5P-containing peptides from  $\alpha$ DG can be extended with matriglycan which has a range of different lengths. This modification increases the mass of the glycopeptide by several kDa, pushing the glycopeptide size outside the range of  $m/z$  usually scanned during a mass spectrometric analysis of peptides. Moreover, matriglycan heterogeneity contributes to splitting the signal of peptides bearing Rbo5P over a range of peaks, each of which is then harder to detect.

O-mannosyl glycans are not the only O-linked glycans present in  $\alpha$ DG. Glycoproteomic analysis of endogenous  $\alpha$ DG from humans (Nilsson et al., 2010) or rabbits (Stalnaker et al., 2010) and the analysis of recombinant  $\alpha$ DG from mice (Harrison et al., 2012) revealed several O-GalNAc glycan structures, in particular, core 1 (Gal $\beta$ 1-3GalNAc-O-Ser/Thr) and core 2 (Gal $\beta$ 1-3[GlcNAc $\beta$ 1-6]GalNAc-O-Ser/Thr) structures that can be extended or not. The different species used for  $\alpha$ DG analysis in these studies present a conserved region-specific glycosylation pattern. This translates into the same O-GalNAc and O-Man-initiated glycan structures on glycopeptides from homologous protein sequence stretches. The glycosylation sites among the different species can vary, and so can the relative abundance of each glycan (Toledo et al., 2012).

The glycosylation state of  $\alpha$ DG has been extensively studied, yet some questions remain unanswered. While three  $\alpha$ DG sites have been identified as modified with core M3 glycans (T317, T319 and T379) (Yagi et al., 2013; Yoshida-Moriguchi et al., 2010), only one (T317) has been found to be extended with Rbo5P or GroP (Kanagawa et al., 2016; Praissman et al., 2016; Yagi et al., 2016). Moreover, T313 is also the only  $\alpha$ DG site found to be extended with matriglycan. These observations raise the question of whether Rbo5P, GroP, and matriglycan can modify other  $\alpha$ DG sites. More importantly, the presence of Rbo5P and GroP has been demonstrated by analysing truncated versions of  $\alpha$ DG, while the mass spectrometric analyses of peptides obtained from full-length  $\alpha$ DG have not been able to replicate the findings (Harrison et al., 2012; Nilsson et al., 2010; Stalnaker et al., 2010). A question left to answer is then if there is a method to detect glycopeptides bearing Rbo5P from full-length  $\alpha$ DG. Such a method would also be useful to identify glycoproteins modified with Rbo5P other than  $\alpha$ DG.

## 3.2 Aims

The interest in understanding the function of  $\alpha$ DG and its relation to congenital muscular dystrophies has led to several studies of the glycosylation of  $\alpha$ DG. Because of those studies, the first glycan in mammalian cells to be modified with Rbo5P has been identified. The analysis leading to this discovery did not use the full-length  $\alpha$ DG protein but instead consisted of selective analysis of short recombinant peptides expected to carry the matriglycan. This observation highlights the challenge in characterising new proteins modified with Rbo5P in mammalian cells since there are no methods in the literature for the selective glycoproteomic analysis of Rbo5P glycoconjugates. One potential option for the identification and characterisation of novel proteins modified with Rbo5P is the method described in Chapter 2 for tagging and enriching proteins with NeuAc, which can in theory also be used for tagging and enriching proteins modified with Rbo5P. The only protein known to date to bear Rbo5P is  $\alpha$ DG; therefore, the isolation of recombinant full-length  $\alpha$ DG would facilitate the development of a method for the tagging and enrichment of Rbo5P-containing protein.

The objective of the work presented here is to identify glycopeptides modified with Rbo5P from  $\alpha$ DG. In this chapter, the design of constructs used for the overexpression of recombinant  $\alpha$ DG is described. After expression of the recombinant  $\alpha$ DG and purification using anti-FLAG pulldown, the protein will be digested into peptides and analysed by mass spectrometry to show that the expressed construct is indeed  $\alpha$ DG. The analysis will also make it possible to identify glycopeptides from  $\alpha$ DG and compare them to those reported in the literature. To test the method described in Chapter 2 on  $\alpha$ DG, the protein will be oxidised with sodium periodate, which is expected to cleave the tandem repeat of Rbo5P and release the matriglycan. Oxidised Rbo5P still attached to the protein should then be available for conjugation with biotin.  $\alpha$ DG that will have undergone this biotinylation, will be digested into peptides and analysed by mass spectrometry.

## 3.3 Materials and Methods

### 3.3.1 Plasmid cloning

#### 3.3.1.1 p $\alpha$ DG-MycFLAG

The plasmid p $\alpha$ DG-MycFLAG was amplified by PCR using Q5 polymerase with 5'AAACTGTA GCACCATCACCCCTGCAGAATATCACCCGGGGCAGCGGACCGACGCGTACGCG (forward) and 5'TGAGTTTCTGCTCGAGCGGCCGCGTACGCGTCCGGTCCGCTGCCCCGGGTGATATT CTGCA (reverse) primers in the presence of pDAG1-MycFLAG (RC230572 OriGene). The PCR product was analysed on a 0.8% agarose gel and extracted from a gel band. The purified plasmid was incubated with DnpI for 1 h at 37 °C, then transformed into Stellar competent cells by heat-shock method (as reported in the manufacturer's manual, ClonTech). Stellar cells were plated on LB agar + 25  $\mu$ g/mL kanamycin. Colonies were picked, grown overnight at 37 °C shaking at 300 rpm in LB medium containing 25  $\mu$ g/mL kanamycin, and their DNA was extracted (using QIAprep® Miniprep kit, Qiagen) and sequenced. Plasmid with the correct sequence was purified on a  $\mu$ g scale using an EndoFree Plasmid Maxi kit (Qiagen). p $\alpha$ DG-MycFLAG was then used for transfection of HEK293T or HEK293F cells.

#### 3.3.1.2 p $\alpha$ DG-S654A-MycFLAG

The plasmid p $\alpha$ DG-S654A-MycFLAG was obtained by PCR using Q5 polymerase with the non-homologous primers 5'CACCCGGGGCAGCGGACCGACG (forward) and 5'ATATTCTGCAG GGTGATGGTGC (reverse) in the presence of p $\alpha$ DG-MycFLAG. Alternatively, plasmid p $\alpha$ DG-S654A-MycFLAG was obtained by PCR using a QuickChange kit (Agilent) with the homologous primers 5'ATCACCCGGGGCAGCGGACCGACGCG (forward) and 5'CGCGTCCG TCCGCTGCCCCGGGTGAT (reverse). PCR products were eventually transformed into Stellar competent cells as described above. The DNA extracted was used for transfection of HEK293T or HEK293F cells.

#### 3.3.1.3 pFLAG $\alpha$ DGh and pFLAG $\alpha$ DGr

Plasmid pFLAG $\alpha$ DGh was cloned using an In-Fusion HD cloning kit (Clontech) with a linearised vector and two inserts. The linearised vector was the PCR product obtained from the amplification of the plasmid p $\alpha$ DG-MycFLAG in the presence of the primers 5'TAAACGGCCGGCCGCGGTC (forward) and 5'GGACTGAGCCATAACCCACAGAGAG reverse. The first insert, encoding for  $\alpha$ DG, was the PCR product obtained from the amplification of the plasmid p $\alpha$ DG-MycFLAG in the

presence of the primers 5'CGATGACAAGCACTGGCCCAGTGAACCCT (forward) and 5' TGACCGCGGCCGGCCGTTTGTAGCCCCGGGTGATATTCTGCA (reverse). The second insert, encoding for 3XFLAG, was custom-purchased. The In-Fusion product was transformed into Stellar competent cells as described above. The plasmid (vector pCMV1-FLAG) encoding N-terminally FLAG-tagged rabbit aDG (pFLAG $\alpha$ DGr) was kindly donated by the laboratory of Prof Paul Martin (Yoon et al., 2013). pFLAG $\alpha$ DGr was transformed into Stellar competent. Stellar cells were plated on LB agar + 100  $\mu$ g/mL ampicillin. Colonies were picked and grown overnight, and their DNA was extracted. pFLAG $\alpha$ DGh and pFLAG $\alpha$ DGr were used for transfection of HEK293T cells.

### **3.3.2 Cell culture and lysates preparation**

HEK293T and Panc-1 cells were grown in Dulbecco's modified Eagle's medium (DMEM, Gibco), while SKBR3 cells were grown in McCoy's 5A modified medium (Gibco). Both media were supplemented with 10% fetal bovine serum (FBS, heat-inactivated), 2 mM L-glutamine, 100 units/mL penicillin, and 100  $\mu$ g/mL streptomycin, and all cells were kept at 37°C in an incubator with 5% CO<sub>2</sub>. HEK293F cells were grown in suspension in FreeStyle 293 Expression medium (Gibco) and kept at 37°C in an incubator with 8% CO<sub>2</sub>. To obtain cell pellets, the medium was removed from the plates. Cells were washed with 1X PBS and harvested by scraping in 500  $\mu$ L of 1X PBS with a protease inhibitor cocktail (1 mM AEBSF, 20  $\mu$ M leupeptin, 1  $\mu$ M pepstatin A). Cell pellets were obtained after centrifugation for 5 min at 500 x g. Cell pellets were lysed in lysis buffer (250 mM NaCl, 50 mM Tris-HCl pH 8.0, 0.5% NP-40 and protease inhibitor cocktail) for 30 minutes on ice agitating every 5 min. The solution was centrifuged at 13000 x g for 30 min at 4°C, after which the supernatant was collected. Cell lysate concentration was measured with the Pierce BCA protein assay kit (Thermo Scientific) following the manufacturer's protocol.

### **3.3.3 Transfection and purification of constructs**

HEK293T cells were seeded in 10 cm cell culture dishes and grown to 90% confluency for transfection. The medium was changed 1 h before transfection. Plasmids pFLAG $\alpha$ DGh or pFLAG $\alpha$ DGr (10  $\mu$ g/cell culture dish) were diluted with Opti-MEM (1 mL/cell culture dish, Invitrogen) and incubated with X-tremeGENE HP DNA transfection reagent (30  $\mu$ L/cell culture dish, Roche) for 30 min at room temperature before being added dropwise to the cells. Cells transfected with pFLAG $\alpha$ DGh were grown in the presence of 50  $\mu$ M furin inhibitor (Calbiochem). Cell medium (10 mL/cell culture dish) was collected two days after transfection into 50 mL corning

tubes and centrifuged for 5 min at 500 x g, room temperature, to remove dead cells. For transfection with HEK293F, 30 mL ( $3 \times 10^7$  cells) of cells in suspension was used. The transfection reagent was prepared by incubating 1 mL of Opti-MEM with 30  $\mu$ g of DNA for 5 min at RT. Another 1 mL of Opti-MEM was incubated with 43.4  $\mu$ L of 293-Fectin (Gibco) for 5 min at RT. The two solutions were combined, incubated for 20 min at RT, and added to the cell culture. For the purification of the constructs FLAG $\alpha$ DGh and FLAG $\alpha$ DGr, slurry anti-FLAG M2 affinity gel (Sigma) was washed with Tris-buffered saline (TBS, 0.2M Tris-HCl, 1.5 M NaCl, pH 7.6) and incubated with medium from transfected cells (150  $\mu$ L gel/10 mL medium) on an end-to-end rotator overnight at 4 °C. After centrifugation at 500 x g, the supernatant was removed from the gel and discarded. The gel was washed three times with TBS. The gel was then incubated with 2X Laemmli sample buffer (62.5 mM Tris-HCl pH 6.8, 1% LDS, 10% glycerol, and 0.005% bromophenol blue, Bio-Rad) for 10 min at 70 °C to release bound proteins. The supernatant was collected for western blot analysis, bioorthogonal ligation with aminoxy-biotin and in-gel digestion.

### **3.3.4 Western blotting**

Cell lysates, media (concentrated 10 X with Vivaspin 30 kDa MWCO centrifugal concentrators) or purified constructs were loaded onto a 4-20% SDS-PAGE gradient gel, which was developed for 40 min at 200 V. The gel was then transferred to a PVDF membrane (0.2  $\mu$ m pore size) by wet transfer using 1 L Towbin buffer (25 mM Tris, 192 mM glycine, 10 % (v/v) methanol pH 8.3) at a constant voltage of 45 V for 1 h. The membrane was blocked in blocking buffer (50 mM Tris-HCl, 150 mM NaCl, 0.25% (w/v) BSA, and 0.05 % (v/v) NP-40, pH 7.4) for 30 min at room temperature, then incubated with primary antibodies I1H6 (1:1000, overnight at 4 °C, Sigma clone I1H6C4) or anti-FLAG (1:10000, 1.5 h at rt, Sigma) and secondary antibody anti-mouse HRP (1:4000, 1 h at rt, Thermo Fisher). Alternatively, the membrane was incubated with HRP-conjugated anti-FLAG antibody (HRP-66008, 1;10000 in blocking buffer, Proteintech) for 1.5 h at room temperature. The membrane was washed with 10 mL of blocking buffer three times for 10 min, and proteins were detected by enhanced luminescence (GERPN2236 ECL kit from Sigma) using a G: BOX Chemi system (Syngene).

### 3.3.5 Bioorthogonal ligation with aminooxy-biotin

Purified FLAG $\alpha$ DGh (obtained after transfection of HEK293T cells from one 10 cm cell culture dish) and FLAG $\alpha$ DGr (obtained after transfection of HEK293T cells from nine 10 cm cell culture dishes) were concentrated in 30 kDa MWCO centrifugal concentrators (Vivaspin) following the manufacturer's protocol and washed three times with 500  $\mu$ L of PBS pH 7.4. Proteins were treated with 1 mM sodium periodate (Sigma) for 0.5 h in PBS (pH 7.4) at 4 °C using an end-to-end rotator in the dark (final volume 4 mL). The reaction was quenched by adding glycerol to a final concentration of 1 mM. The samples were buffer exchanged three times with 500  $\mu$ L of PBS (pH 6.7) using 30 kDa MWCO centrifugal concentrators. Biotinylation was performed by adding aminooxy-biotin (Biotium) to the samples to a final concentration of 100  $\mu$ M in the presence of 10 mM aniline (final volume 1 mL). The reaction was carried out in PBS pH 6.7 for 3 hours on an end-to-end rotator at 4°C. Samples were washed thrice with PBS pH 6.7 using 30 kDa MWCO concentrators to remove the excess biotin and were ready for in-gel digestion.

### 3.3.6 In-gel digestion

Purified FLAG $\alpha$ DGh and FLAG $\alpha$ DGr (each obtained after transfection of HEK293T cells from one 10 cm cell culture dish) were supplemented with 50 mM dithiothreitol (DTT) and separated on a 4-20% SDS-PAGE gradient gel (Bio-Rad) for 5 min at 200 V. Gel were stained with InstantBlue Coomassie (Abcam) for 1 h. Protein bands were excised and in-gel digestion was performed as described in Paragraph 2.3.2. The proteolytic digestion was performed with trypsin (expressed in *Pichia pastoris* Proteomics Grade, Sigma) overnight at 37 °C or proteinase K (Invitrogen) for 1 h at 37 °C. FLAG $\alpha$ DGh and FLAG $\alpha$ DGr recovered after bioorthogonal ligation with aminooxy-biotin were mixed with Laemmli sample buffer, separated on a 4-20% SDS-PAGE gradient, and treated for in-gel digestion as described above. Trypsin was used for the proteolytic digestion of these samples.

### 3.3.7 Mass spectrometric analysis

Samples obtained after in-gel digestion were dissolved into 0.1% TFA and analysed by mass spectrometry as described in paragraph 2.3.5.

### 3.3.8 Data analysis

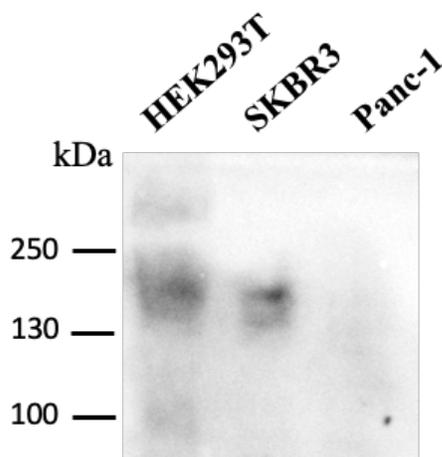
Automated database searches were performed with Byonic software (Protein Metrics Inc) using as reference database the *Homo sapiens* or *Oryctolagus cuniculus* (rabbit) proteomes (Uniprot). The search criteria were set as follows: KR (trypsin) or AVLIYWF (proteinase K) as cleavage sites;

≤ 3 missed cleavages and semi-specific cleavage allowed; error tolerances for the precursor of 3 ppm and 0.05 Da for fragment ions; carbamidomethyl cysteine as a fixed modification, oxidation of methionine (common 2), and pyro-Glu from glutamine (rare 1) as variable modifications (max common set to 3 and max rare set to 2); possible glycan composition set as Hex<sub>1</sub>, HexNAc<sub>2</sub>, Hex<sub>1</sub>HexNAc<sub>2</sub>, Hex<sub>2</sub>HexNAc<sub>1</sub>, Hex<sub>2</sub>HexNAc<sub>2</sub>, Hex<sub>1</sub>HexNAc<sub>1</sub>NeuAc<sub>1</sub>, Hex<sub>1</sub>HexNAc<sub>2</sub>NeuAc<sub>1</sub>, Hex<sub>1</sub>HexNAc<sub>1</sub>NeuAc<sub>2</sub>, Hex<sub>1</sub>HexNAc<sub>1</sub>NeuAc<sub>3</sub>, Hex<sub>2</sub>HexNAc<sub>1</sub>NeuAc<sub>1</sub>, Hex<sub>2</sub>HexNAc<sub>2</sub>NeuAc<sub>1</sub>, Hex<sub>2</sub>HexNAc<sub>2</sub>NeuAc<sub>2</sub>, Hex<sub>3</sub>HexNAc<sub>2</sub>NeuAc<sub>1</sub>, Hex<sub>2</sub>HexNAc<sub>2</sub>NeuAc<sub>1</sub>Fuc<sub>1</sub>, Hex<sub>1</sub>HexNAc<sub>1</sub>NeuAc<sub>1</sub>NeuGc<sub>1</sub> and HexNAc<sub>2</sub>NeuAc<sub>1</sub> (each set as rare 1, on Ser/Thr); false discovery rate (FDR) was set to 1%. Putative glycopeptides identified by Byonic were checked manually to confirm the assignment.

## 3.4 Results

### 3.4.1 Expression of endogenous glycosylated $\alpha$ DG in mammalian cells

Expression of  $\alpha$ DG in mammalian cells has been described in the literature, as summarised in the introduction, and was performed to characterise the glycosylation state of the protein.  $\alpha$ DG is also essential for this project so that it can be used to test whether the method devised and developed for tagging proteins modified with Rbo5P, described in Chapter 2, is effective with this protein. The method could be technically tested using a peptide from  $\alpha$ DG bearing Rbo5P. Nevertheless, the overarching purpose of the project presented here is to implement the method on protein mixtures (e.g. cell lysates) to identify novel Rbo5P glycoconjugates. For this reason, the full-length  $\alpha$ DG protein is considered a better sample for method testing than an  $\alpha$ DG peptide. A mammalian cell line was needed for the recombinant expression of  $\alpha$ DG. To be suitable for the purpose, a cell line should be able to express glycosylated  $\alpha$ DG with glycans bearing Rbo5P and matriglycan. Human embryonic kidney 293T (HEK293T) and human breast cancer SKBR3 cell lines have both been shown to express glycosylated  $\alpha$ DG (Kunz et al., 2005; Oppizzi et al., 2008). The expression of endogenous glycosylated  $\alpha$ DG in HEK293T and SKBR3 cells was tested before attempting the expression of recombinant  $\alpha$ DG in these cell lines. Cell lysates obtained from the two cell lines were analysed by western blotting using I1H6 as the primary antibody (Figure 3.1). The antibody I1H6 recognises matriglycan on  $\alpha$ DG and is used in western blot analysis to detect glycosylated  $\alpha$ DG (Ervasti and Campbell, 1993). Cell lysate from the human pancreatic cancer cell line Panc-1 was used as a negative control for the experiment. Panc-1 cells do not express at high levels the enzyme LARGE, responsible for the synthesis of matriglycan, therefore I1H6 does not detect  $\alpha$ DG in their lysate (Esser et al., 2013).



*Figure 3.1/ Western blot analysis of cell lysates obtained from HEK293T, SKBR3, and Panc-1. The antibodies used were anti-dystroglycan, IIH6, as primary and anti-mouse HRP as secondary. Equal amounts of proteins were loaded in each lane.*

A band over 130 kDa is observed in the lanes corresponding to HEK293T and SKBR3, while no band can be detected in the lane of Panc-1. The bands detected are consistent with glycosylated  $\alpha$ DG bearing matriglycan. At the same time, the absence of a similar band in the negative control shows the ability of the antibody to recognise glycosylated  $\alpha$ DG selectively. The result suggests that HEK293T and SKBR-3 could be used to express recombinant glycosylated  $\alpha$ DG. The transfection of HEK293T cells for the expression of recombinant  $\alpha$ DG has already been described in the literature (Kunz et al., 2005; Yoshida-Moriguchi et al., 2010). Therefore, the cell line was used in this study for transfection experiments.

### **3.4.2 Expression of recombinant $\alpha$ DG in mammalian cells**

#### **3.4.2.1 Cloned p $\alpha$ DG-MycFLAG is transfected in HEK293T and HEK293F cells**

The commercially available plasmids for the expression of full-length  $\alpha$ DG include the *DAG1* gene, which encodes for both  $\alpha$  and  $\beta$ DG and tags at the C-terminus of  $\beta$ DG.  $\alpha$ DG and  $\beta$ DG, despite being translated as a single polypeptide, are processed in cells to give two proteins that are non-covalently bound to each other. Considering these facts, the plasmids result in the expression of tagged  $\beta$ DG and untagged  $\alpha$ DG. The purpose here is instead to obtain tagged full-length  $\alpha$ DG, where the tag can be used for both detection and purification of the recombinant  $\alpha$ DG. For this reason, a new plasmid was cloned. A commercially available plasmid for the expression of human  $\alpha$  and  $\beta$ DG with a Myc and a FLAG-tag at the C-terminus of  $\beta$ DG was used for cloning and will be referred to as pDAG1-MycFLAG. Deleting the DNA sequence encoding for  $\beta$ DG from pDAG1-MycFLAG would result in a plasmid suitable for the expression of  $\alpha$ DG with C-terminal Myc- and FLAG- tags. For this scope, primers were designed, and a PCR-mediated DNA plasmid deletion was performed on pDAG1-MycFLAG. The PCR product was analysed on an agarose gel (Figure 3.2).

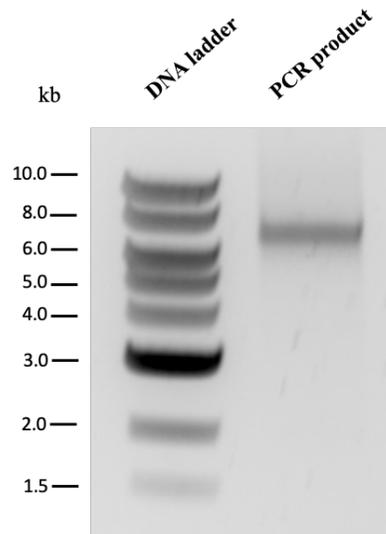


Figure 3.2/ Agarose gel electrophoresis of a PCR product. Plasmid *pDAG1-MycFLAG* was amplified by PCR using two primers to delete the DNA sequence encoding for  $\beta$ DG.

The PCR product lane shows a clear band between 6 and 8 kb. The plasmid after deletion is expected to be 6.8 kb, which is consistent with what was observed on the agarose gel. The DNA was purified from the gel band, treated with *Dpn1* to digest the methylated template, and transformed into *E. coli* competent cells. Colonies were picked for DNA sequencing and showed the correct sequence for  $\alpha$ DG with C-terminal Myc- and FLAG- tags. The plasmid obtained from the transformed cells will be called *p $\alpha$ DG-MycFLAG* hereafter. A schematic of the maps for the plasmids discussed so far is shown in Figure 3.3.

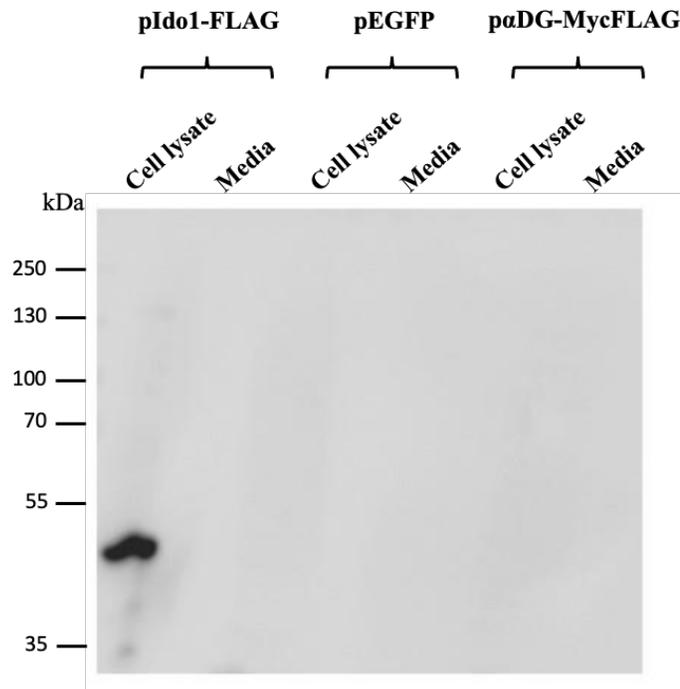


Figure 3.3/ Schematic of the maps for the plasmids *pDAG1-MycFLAG* and *p $\alpha$ DG-MycFLAG*. *SP* refers to the signal peptide.

HEK293T cells were transfected with p $\alpha$ DG-MycFLAG with the purpose of expressing  $\alpha$ DG with Myc- and FLAG- tags ( $\alpha$ DG-MycFLAG).  $\alpha$ DG-MycFLAG was expected to be released into the media of transfected cells (Kunz et al., 2005; Yoshida-Moriguchi et al., 2010). The reason for the presence of recombinant  $\alpha$ DG in the media instead of the cell membrane can be explained by considering that  $\alpha$ DG is non-covalently bound to  $\beta$ DG on the cell membrane. In cells where  $\alpha$ DG, but not  $\beta$ DG, is overexpressed, the recombinant protein lacks the substrate needed for anchoring to the cell membrane and is therefore released into the media.

Two and five days after transfection of HEK293T cells with p $\alpha$ DG-MycFLAG, cells and media were collected. Cells were lysed while the media was concentrated. Cell lysates and media were used for western blot analysis using anti-FLAG or I1H6 antibodies as primary antibodies. No signal was observed from the cell lysates or media using either of the two antibodies (data not shown). The result raised the question of whether the lack of detected recombinant  $\alpha$ DG was due to a problem with the transfection protocol. HEK293T cells were transfected with the plasmid pEGFP to address this concern. pEGFP encodes for the enhanced green fluorescent protein GFP, which can be detected in cells by fluorescent imaging. Two days after transfection, cells transfected with pEGFP appeared green under the microscope. The result suggests that the plasmid was transfected correctly into cells.

Since the efficacy of the transfection protocol was no longer in doubt, the anti-FLAG antibody's efficacy was tested. To achieve that, HEK293T cells were transfected with the plasmid pIdo1-FLAG. pIdo1-FLAG encodes for the FLAG-tagged protein Ido1 (Ido1-FLAG), which was used here to generate a positive control for the anti-FLAG antibody. Cells and media were collected two days after transfection of HEK293T cells and used for western blot analysis (Figure 3.4). The transfection with pEGFP or p $\alpha$ DG-MycFLAG was also performed in parallel.



*Figure 3.4/ Western blot analysis of cell lysates and media obtained from transfection of HEK293T cells with the plasmids pIdo1-FLAG, pEGFP or pαDG-MycFLAG. HRP anti-FLAG was used as the antibody.*

The anti-FLAG antibody detects a band at ~ 45 kDa in the cell lysate of HEK293T cells transfected with pIdo1-FLAG. Ido1-FLAG has a predicted molecular weight of 45 kDa, so the band detected in Figure 3.4 is evidence that the anti-FLAG antibody is able to detect a FLAG-tagged protein by western blot analysis. The figure also shows the absence of any signal in the cell lysate or media from HEK293T cells transfected with pαDG-MycFLAG. The results taken together suggest that αDG-MycFLAG is not detected in HEK293T after transfection with pαDG-MycFLAG and the cause is not linked to a problem with the transfection protocol or the ability to detect the FLAG-tag.

The lack of expression of recombinant αDG in HEK293T was hypothesised to be caused by an issue with either the cell line or the plasmid used for transfection. For this reason, a different cell line was tested and used for transfection. HEK293F is a human cell line that grows in suspension differently from HEK293T. HEK293F cells were transfected with either pαDG-MycFLAG or pEGFP and collected with the media after two and five days. As done previously, the transfection with pEGFP was used to assess the efficacy of the transfection by visually spotting green fluorescent cells. Cells were lysed after transfection, and the media was

concentrated. Both cell lysates and media were analysed by western blotting using I1H6 and anti-FLAG antibodies as primary antibodies. Results are shown in Figure 3.5 A and B.

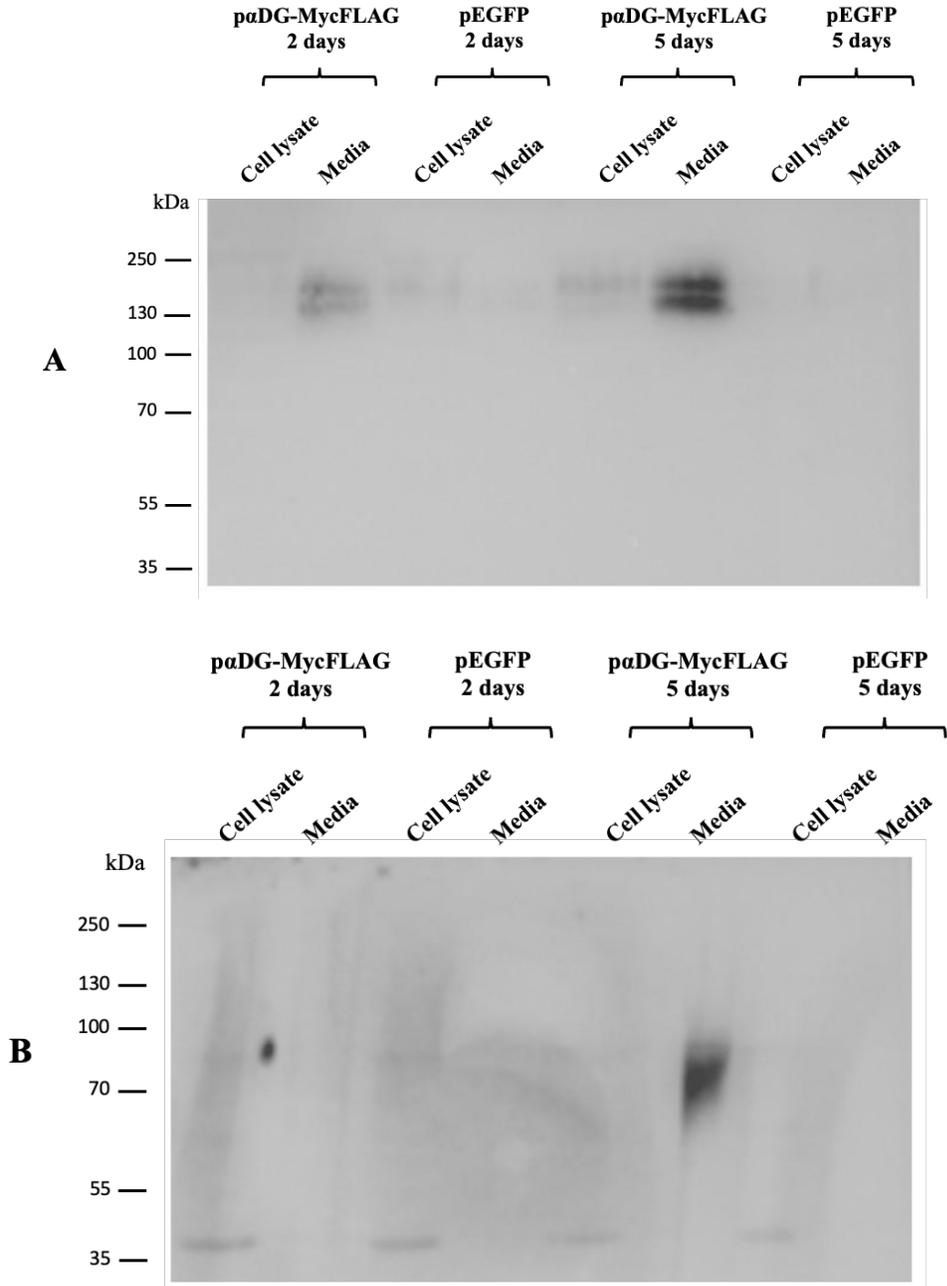
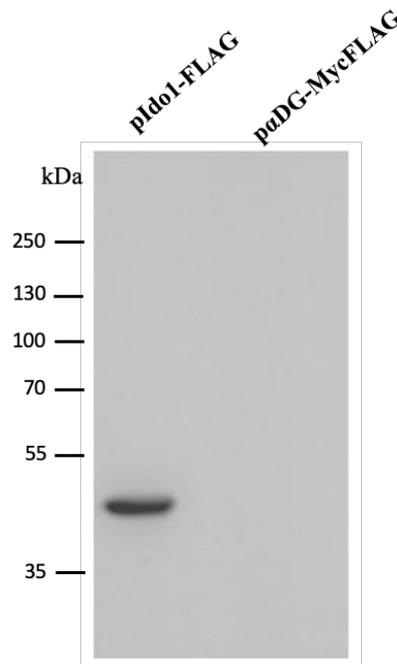


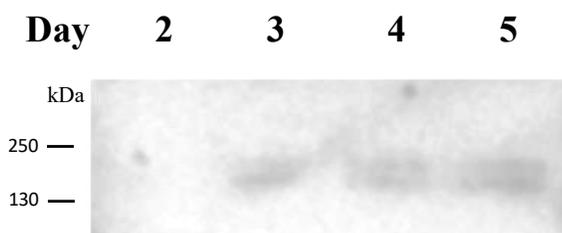
Figure 3.5/ Western blot analysis of cell lysates and media obtained from transfection of HEK293F cells with the plasmids pEGFP or pαDG-MycFLAG. Cells and media were collected two and five days after transfection. The antibodies used were I1H6 as primary and anti-mouse HRP as secondary (A) or HRP anti-FLAG (B).

In the western blot where I1H6 was used, bands between 130 and 250 kDa are observed in the media from cells transfected with p $\alpha$ DG-MycFLAG collected after two or five days. Fainter bands at the same molecular weight are observed in the cell lysate obtained after transfection with pEGFP. This signal can be associated with the endogenous  $\alpha$ DG. On the other hand, the signal from the media of cells transfected with p $\alpha$ DG-MycFLAG can be interpreted as the recombinant  $\alpha$ DG-MycFLAG. The result suggests that after transfection, HEK293F can express recombinant  $\alpha$ DG. In the western blot where anti-FLAG was used as the antibody, only one band between 70 and 100 kDa is observed in the media collected five days after transfection with p $\alpha$ DG-MycFLAG. The signal, which could, in theory, be associated with recombinant  $\alpha$ DG, does not appear at the same molecular weight as the signals observed in Figure 3.5 A. Moreover, the band between 70 and 100 kDa is not seen in the media collected two days after transfection with p $\alpha$ DG-MycFLAG. It was hypothesised that the signal detected by anti-FLAG could be non-specific; therefore, a different anti-FLAG antibody was tested. For the western blot analysis, the media of HEK293F cells transfected with  $\alpha$ DG-MycFLAG and the cell lysates containing Ido1-FLAG obtained from the previous experiment) were used (Figure 3.6).



*Figure 3.6/ Western blot analysis of cell lysates obtained from transfection of HEK293T cells with pIdo1-FLAG or media obtained five days after transfection of HEK293F with p $\alpha$ DG-MycFLAG. Anti-FLAG was used as the primary antibody, and anti-mouse HRP was used as the secondary antibody.*

The new anti-FLAG antibody detects Ido1-FLAG at the correct molecular weight but does not detect any signal in the media of HEK293F cells transfected with p $\alpha$ DG-MycFLAG. The result confirms that the signal observed in Figure 3.5 B is non-specific, and the FLAG-tag is not detected in transfected HEK293F expressing recombinant  $\alpha$ DG. The transfection of HEK293F with p $\alpha$ DG-MycFLAG was repeated a second time, and the media collected after two, three, four and five days was concentrated and used for western blot analysis (Figure 3.7).



*Figure 3.7/ Western blot analysis of media obtained from transfection of HEK293F cells with the plasmid p $\alpha$ DG-MycFLAG. Media was collected two, three, four and five days after transfection. The antibodies used were IIH6 as primary and anti-mouse HRP as secondary.*

The western blot shows bands between 130 and 250 kDa, similar to what was observed in the first experiment in the media collected after three, four and five days. The western blot analysis of the same samples using either one of the two anti-FLAG antibodies did not show any signal (data not shown). The results taken together confirm the reproducibility of the expression of  $\alpha$ DG in transfected HEK293F cells and the absence of a detectable FLAG-tag on the recombinant protein.

### **3.4.2.2 Cloned p $\alpha$ DG-S654A-MycFLAG is transfected in HEK293F cells**

The FLAG-tag was essential in the expression of recombinant  $\alpha$ DG because it would be used for its purification by immunoprecipitation. For this reason, the issue had to be addressed. The loss of the FLAG-tag from  $\alpha$ DG after expression was hypothesised. It is known that the gene *DAG1* encoding for  $\alpha$ DG and  $\beta$ DG translates into a single polypeptide, which undergoes autoproteolysis at Ser in position 654 to give the two different dystroglycan subunits (Holt et al., 2000). For the autoproteolysis to happen, an Ig-like domain upstream of S654 is required. Amino acids around the cleavage sites are not conserved between vertebrates and invertebrates, but similarities are found with the proteins that undergo autoproteolysis (Akhavan et al., 2008). For this reason, S654, being

part of a domain necessary for the autoproteolysis, cannot be excluded. The recombinant protein encoded by the plasmid p $\alpha$ DG-MycFLAG has the same first 654 amino acids encoded in *Dagl1*, corresponding to  $\alpha$ DG. This means the construct has the Ig-like domain required for the autoproteolysis and the S654 where the cleavage happens but lacks the amino acids downstream of S654, which could play a role in the autoproteolysis. Based on these considerations, it was hypothesised that the autoproteolysis could still happen. To prevent the cleavage and the loss of the Myc- and FLAG- tags, the cleavage site was mutated from Ser to Ala, as this mutation has been shown to prevent autoproteolysis in the dystroglycan polypeptide (Jayasinha et al., 2003). The cloning of the plasmid with the mutation S654A, named from now on p $\alpha$ DG-S654A-MycFLAG, started with the design of two non-homologous primers with the forward primer carrying the mutation (Figure 3.8).

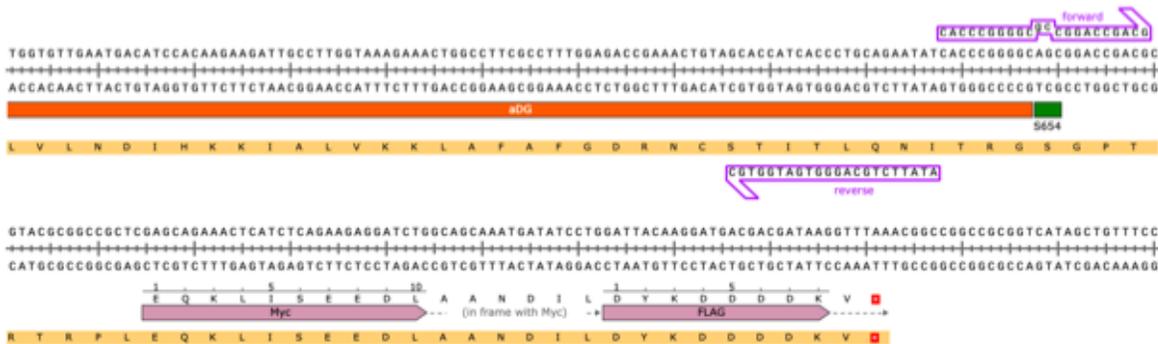


Figure 3.8/  $\alpha$ DG DNA sequence at the C-terminus in the plasmid p $\alpha$ DG-MycFLAG. Non-homologous forward and reverse primers designed for the mutation S654A are highlighted.

A PCR using the two primers and the plasmid p $\alpha$ DG-MycFLAG as a template was performed, and the resulting product was transformed into *E.coli* competent cells. DNA was extracted and sequenced. The sequence showed missing base pairs at the extremities of the linear plasmid, which translated into an erroneous amino acid sequence compared to the predicted one. The cloning was therefore repeated using two new primers, which, differently from before, were homologous and both carrying the mutation (Figure 3.9)

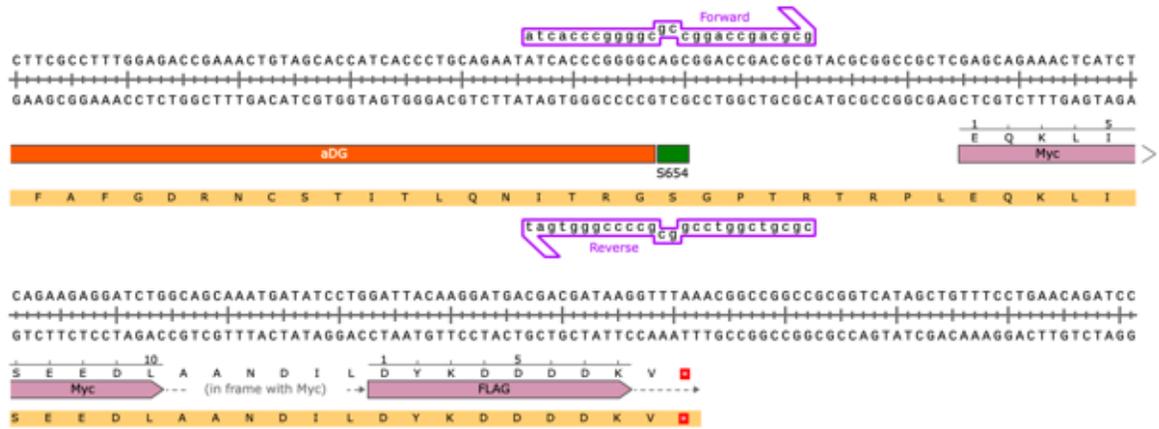


Figure 3.9/  $\alpha$ DG DNA sequence at the C-terminus in the plasmid  $paDG$ -MycFLAG. Homologous forward and reverse primers designed for the mutation S654A are highlighted.

This time, the plasmid  $paDG$ -S654A-MycFLAG had the correct sequence and could be used for transfection into mammalian cells. As described above, HEK293F cells were transfected with  $paDG$ -S654A-MycFLAG, and the media was collected after three, four, and five days. The media was concentrated, and the different time points were analysed by western blotting using IIIH6 as the primary antibody. Results are shown in Figure 3.10.

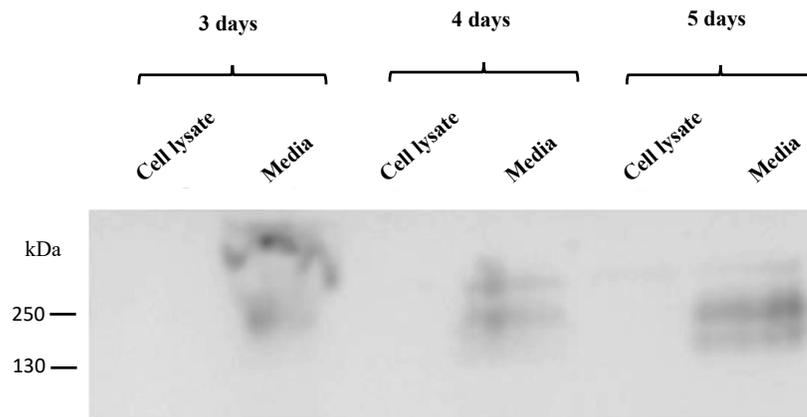


Figure 3.10/ Western blot analysis of cell lysates and media obtained from transfection of HEK293F cells with the plasmid  $paDG$ -S654A-MycFLAG. Cells and media were collected three, four and five days after transfection. The antibodies used were IIIh6 as primary and anti-mouse HRP as secondary.

Bands between 130 and 250 kDa and above 250 kDa are observed in the western blot. These bands are evidence of the expression of  $\alpha$ DG and appear at a higher molecular weight than the bands observed in Figure 3.5 A, where media from cells transfected with non-mutated  $\alpha$ DG was analysed. When anti-FLAG antibodies were used on the samples analysed in Figure 3.10, no bands were observed (data not shown). While the increase in molecular weight of the recombinant protein after mutation cannot be explained, overall, the result suggests that the S654A mutation did not help the detection of the FLAG-tag, which is still hypothesised to be absent on the recombinant  $\alpha$ DG.

### 3.4.2.3 FLAG-tagged $\alpha$ DG is successfully expressed in HEK293T cells

An issue with the tag's position in the construct could explain the absence of a detectable FLAG-tag from recombinant  $\alpha$ DG. The FLAG-tag, present in the plasmid at the C-terminus of the sequence encoding for  $\alpha$ DG, could be moved to the N-terminus. To test the hypothesis that the FLAG-tag position could affect its expression and detection, a new plasmid, named pFLAG $\alpha$ DGh, was cloned. pFLAG $\alpha$ DGh was designed to have the sequence for expressing a 3X FLAG-tag at the N-terminus of the  $\alpha$ DG gene. The new plasmid was assembled using an In-Fusion HD cloning kit with a linearised vector and two inserts. p $\alpha$ DG-MycFLAG was used as a template to PCR amplify the linearised vector and one of the inserts encoding for  $\alpha$ DG (30-653). The second insert, encoding for 3X FLAG-tag, was custom-purchased instead. The PCR products for the linearised vector and the insert encoding for  $\alpha$ DG were analysed on an agarose gel (Figure 3.11).

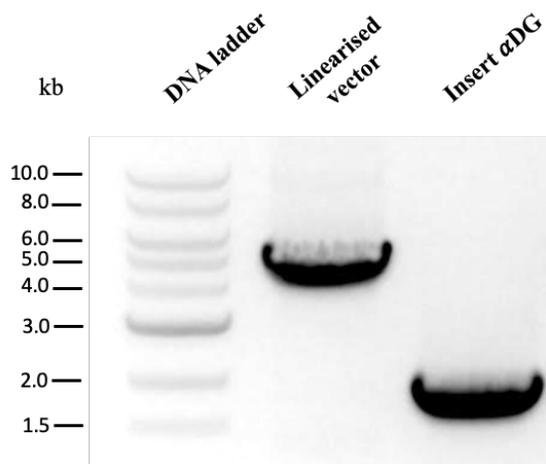
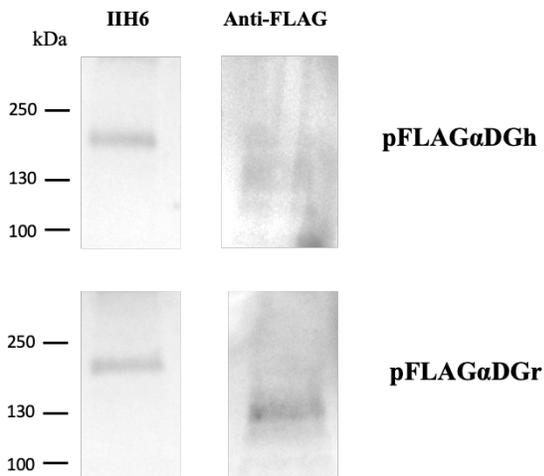


Figure 3.11/ Agarose gel electrophoresis of PCR products. Plasmid p $\alpha$ DG-MycFLAG was amplified by PCR using two sets of primers to obtain a linearised vector and an insert encoding for  $\alpha$ DG.

The bands in the agarose gel appear between 4 and 5 kb for the linearised vector and between 1.5 and 2 kb for the insert encoding for  $\alpha$ DG. The two PCR products are expected to be 4.9 kb (linearised vector) and 1.9 kb (insert  $\alpha$ DG), which is consistent with the bands observed in the PCR product lanes. Using the two PCR products and the insert encoding for 3X FLAG-tag, pFLAG $\alpha$ DGh was successfully cloned.

The transfection of pFLAG $\alpha$ DGh HEK293T cells requires the presence of a furin inhibitor to prevent the loss of the FLAG-tag. This is necessary because, in cells,  $\alpha$ DG is further processed by furin to cleave the first 312 amino acids of the protein. By deleting the furin cleavage site in  $\alpha$ DG, the recombinant protein keeps its N-terminus domain, as demonstrated in the work published by Yoon (Yoon et al., 2013). The plasmid used in that work for the expression of recombinant  $\alpha$ DG, named here pFLAG $\alpha$ DGr, presents the sequence for the expression of a 1X FLAG-tag at the N-terminus of the  $\alpha$ DG gene (from rabbit) and has a deletion of R312, the site of furin cleavage. pFLAG $\alpha$ DGr, donated by the lab of Dr Paul T. Martin, was used alongside pFLAG $\alpha$ DGh to transfect HEK293T cells and constituted a positive control for the expression of glycosylated recombinant  $\alpha$ DG. The cell media was collected two days after transfection, and western blot analysis was performed on the concentrated media (Figure 3.12)



*Figure 3.12/ Western blot analysis of concentrated media obtained two days after transfection of HEK293T cells with the plasmids pFLAG $\alpha$ DGh or pFLAG $\alpha$ DGr. The antibodies used were I1H6 as primary, anti-mouse HRP as secondary, or HRP anti-FLAG.*

IIH6 detects a narrow band between 130 and 250 kDa in the media of cells transfected with pFLAG $\alpha$ DGh, while the anti-FLAG detects a faint band around 130 kDa. Similar observation can be done for the western blots of samples obtained from the transfection with pFLAG $\alpha$ DGr. The molecular weights of the bands detected by IIH6 and anti-FLAG do not match. This result is unexpected since both signals should come from the same recombinant protein. Nevertheless, this was the first time a signal was detected by both the antibodies for the expression of recombinant  $\alpha$ DG. In particular, the presence of a signal from the FLAG-tag encouraged the idea that a recombinant protein could be purified with anti-FLAG affinity gel from the samples. The purification of the expected products FLAG $\alpha$ DGh and FLAG $\alpha$ DGr after transfection of HEK293T cells with pFLAG $\alpha$ DGh and pFLAG $\alpha$ DGr, respectively, was performed by incubating the media collected after transfection, with anti-FLAG agarose gel. Proteins bound to the gel were eluted and analysed by western blotting (Figure 3.13).

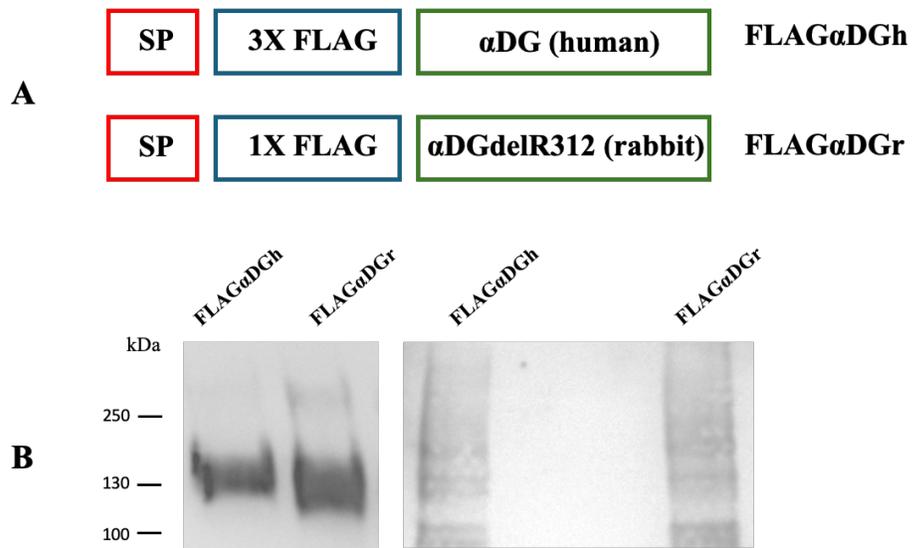


Figure 3.13 (A) Schematic of the constructs FLAG $\alpha$ DGh and FLAG $\alpha$ DGr. SP refers to the signal peptide. (B) Western blot analysis of the purified recombinant proteins FLAG $\alpha$ DGh and FLAG $\alpha$ DGr. The antibodies used were HRP anti-FLAG (left), IIH6 as primary and anti-mouse HRP as secondary (right). Proteins were retrieved from the medium of HEK293T cells and purified by anti-FLAG pull-down using anti-FLAG agarose gel. Proteins eluted from the anti-FLAG gel with Laemmli buffer were used for the western blot analysis.

A prominent band at the molecular weight of ~120-190 kDa is detected by anti-FLAG in the lanes of the samples expected to contain FLAG $\alpha$ DGh and FLAG $\alpha$ DGr. On the other hand, IIH6's signal in the same sample is faint and diffuse. While a match between the bands detected by IIH6 and anti-FLAG is not observed, the result shows that FLAG-tagged proteins are purified by anti-FLAG agarose gel. The broad bands at a high molecular weight observed with anti-FLAG are consistent with those expected for a heavily glycosylated protein like  $\alpha$ DG bearing Rbo5P and matriglycan. The similarity of the bands detected in the sample containing FLAG $\alpha$ DGh and the positive control FLAG $\alpha$ DGr encouraged the idea that the protein recovered was indeed recombinant glycosylated  $\alpha$ DG. The bands can then be interpreted as the purified glycosylated proteins FLAG $\alpha$ DGh and FLAG $\alpha$ DGr. Since the recombinant protein was needed to act as an authentic Rbo5P-containing protein for testing the methods for tagging and enrichment, the FLAG $\alpha$ DGh production had to be performed to a scale which made the experiment prohibitively expensive, especially for the use of furin inhibitor. For this reason, the experiments described in the rest of this chapter will only use FLAG $\alpha$ DGr, which could be produced at a lower cost.

### **3.4.3 Analysis of recombinant $\alpha$ DG by mass spectrometry**

FLAG $\alpha$ DGr construct was purified, reduced and carbamidomethylated, digested into peptides with trypsin, and analysed by mass spectrometry to confirm its identity as  $\alpha$ DG. The analysis involved the separation of peptides obtained from FLAG $\alpha$ DGr by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). Peptides were isolated and fragmented in the mass spectrometer to obtain product ion spectra, which were interpreted with the help of bioinformatic software. The identified peptides corresponded to  $\alpha$ DG peptides. Other peptides belonging to common contaminant proteins, such as human keratins, were identified. The mapped sequence of  $\alpha$ DG from the analysis is shown in Figure 3.14.

1	MRMSVGLSLL	LPLWGRTFLL	LLCVAVAQSH	WPSEPSEAVR	DWENQLEASM	50
51	HSVLSDLHEA	LPTVVGIPDG	TAVVGRSFRV	TIPTDLIGSS	GEVIKVSTAG	100
101	KEVLPSWLHW	DPQSHTLEGL	PLDTDKGVHY	ISVSAAQLDA	NGSHIPQTSS	150
151	VFSIEVYPED	HSEPQSVRAA	SPDLGEEAAS	ACAAEEPVTV	LTVILDADLT	200
201	KMTPKQRIDL	LHRMQSFSEV	ELHNMKLVVP	VNNRFLDMSA	FMAGPGNAKK	250
251	VVENGALLSW	KLGC SLNQNS	VPDIRGVEAP	AREGTMSAQL	GYPVVGWHIA	300
301	NKKPPLPKRI	RRQIHATPTP	VTAIGPPTTA	IQEPPSRIVP	TPTSPAIAPP	350
351	TETMAPVVRD	PVPGKPTVTT	RTRGAI IQTP	TLGPIQPTRV	SDAGTVVSGQ	400
401	IRATVTIPGY	VEPTAVATPP	TTTTKKPRVS	TPKPATPSTD	SSATTTTRRPT	450
451	KKPRTPRPVP	RVTTKAPITR	LETASPPTRI	RTTTSVPRG	GEPNQRPELK	500
501	NHIDRVDAWV	GTYFEVKIPS	DTFYDKEDTT	TDK LKLT LKL	REQQLVGEKS	550
551	WVQFNNSQL	MYGLPDSSHV	GKHEYFMHAT	DKGGLSAVDA	FEIHVHKRPQ	600
601	GDKAPARFKA	KFVGDPAPVV	NDIHKKIALV	KKLAFAGDR	NCSTVTLQNI	650
651	TRG					

Figure 3.14 / Mapped sequence of peptides identified in the mass spectrometric analysis of FLAG $\alpha$ DGr after trypsin digestion. Peptides identified are reported in red.

The peptides identified cover ~20% of the  $\alpha$ DG sequence and were found to be either glycosylated or not. It is interesting to determine whether any glycopeptides modified with Rbo5P or GroP were detected in this experiment since there are no literature reports of such glycopeptides detected by mass spectrometry after proteolytic digest of recombinant  $\alpha$ DG. Glycopeptides modified with Rbo5P were first reported in the literature in 2016 and were identified by selectively overexpressing the peptide containing the sites of matriglycan glycosylation (Kanagawa et al., 2016; Praissman et al., 2016). To reduce the molecular weight of such glycopeptides, allowing their identification by mass spectrometry, hydrofluoric acid was used to break the phosphodiester linkage between the two units of Rbo5P, thus releasing the matriglycan polysaccharide. The presence of matriglycan on glycopeptides modified with Rbo5P would hinder their identification, and that would explain why reports present in the literature before 2016, where hydrofluoric acid was not used, were not able to identify the peculiar glycosylation of  $\alpha$ DG. This logic, though, does not explain why glycopeptides modified with GroP were not identified earlier than 2016 since they are not extended with matriglycan. One explanation could be the low abundance of such glycopeptides, which were identified uniquely by overexpression of a peptide from  $\alpha$ DG (Yagi et al., 2016). Glycopeptides identified from the analysis of FLAG $\alpha$ DGr are listed in Table 3.1

Peptide sequence	Glycan composition
<sup>374</sup> G <u>A</u> I <u>I</u> Q <u>T</u> P <u>T</u> L <u>G</u> P <u>I</u> Q <u>P</u> T <u>R</u> V <u>S</u> D <u>A</u> G <u>T</u> V <u>V</u> S <u>G</u> Q <u>I</u> R <sup>402</sup>	Hex <sub>2</sub> HexNAc <sub>2</sub> NeuAc <sub>4</sub>
<sup>383</sup> G <u>P</u> I <u>Q</u> P <u>T</u> R <sup>389</sup>	HexNAc <sub>2</sub>
<sup>383</sup> G <u>P</u> I <u>Q</u> P <u>T</u> R <sup>389</sup>	Hex <sub>1</sub> HexNAc <sub>2</sub>
<sup>390</sup> V <u>S</u> D <u>A</u> G <u>T</u> V <u>V</u> S <u>G</u> Q <u>I</u> R <sup>403</sup>	Hex <sub>3</sub> HexNAc <sub>2</sub> NeuAc <sub>1</sub>
<sup>390</sup> V <u>S</u> D <u>A</u> G <u>T</u> V <u>V</u> S <u>G</u> Q <u>I</u> R <sup>402</sup>	Hex <sub>1</sub>
<sup>390</sup> V <u>S</u> D <u>A</u> G <u>T</u> V <u>V</u> S <u>G</u> Q <u>I</u> R <sup>402</sup>	Hex <sub>2</sub>
<sup>390</sup> V <u>S</u> D <u>A</u> G <u>T</u> V <u>V</u> S <u>G</u> Q <u>I</u> R <sup>402</sup>	Hex <sub>2</sub> HexNAc <sub>1</sub> NeuAc <sub>1</sub>
<sup>462</sup> V <u>T</u> T <u>K</u> A <u>P</u> I <u>T</u> R <sup>470</sup>	Hex <sub>1</sub> HexNAc <sub>1</sub> NeuAc <sub>2</sub>
<sup>462</sup> V <u>T</u> T <u>K</u> A <u>P</u> I <u>T</u> R <sup>470</sup>	Hex <sub>2</sub> HexNAc <sub>2</sub> NeuAc <sub>1</sub>
<sup>471</sup> L <u>E</u> T <u>A</u> S <u>P</u> P <u>T</u> R <sup>479</sup>	Hex <sub>2</sub> HexNAc <sub>2</sub>
<sup>471</sup> L <u>E</u> T <u>A</u> S <u>P</u> P <u>T</u> R <sup>479</sup>	Hex <sub>3</sub> HexNAc <sub>3</sub> NeuAc <sub>1</sub>
<sup>471</sup> L <u>E</u> T <u>A</u> S <u>P</u> P <u>T</u> R <sup>479</sup>	HexNAc <sub>2</sub>
<sup>471</sup> L <u>E</u> T <u>A</u> S <u>P</u> P <u>T</u> R <sup>479</sup>	Hex <sub>2</sub> HexNAc <sub>3</sub> NeuAc <sub>1</sub>
<sup>518</sup> I <u>P</u> S <u>D</u> T <u>F</u> Y <u>D</u> K <u>E</u> D <u>T</u> T <u>T</u> D <u>K</u> L <u>K</u> <sup>535</sup>	Hex <sub>1</sub> HexNAc <sub>1</sub> NeuAc <sub>2</sub>

Table 3.1 / List of the glycopeptides identified by LC-MS/MS from the tryptic digest of recombinant  $\alpha$ DG from rabbit (FLAG $\alpha$ DGr). Potential O-glycosylation sites are underlined. Hex denotes hexose, e.g., galactose (Gal), glucose (Glc), or mannose (Man). HexNAc denotes N-acetyl hexosamine, e.g., N-acetyl galactosamine (GalNAc) or N-acetyl glucosamine (GlcNAc). NeuAc denotes N-acetyl neuraminic acid.

The glycan composition for each glycopeptide was deduced from the  $m/z$  of the precursor ion and the product ion spectrum of each selected precursor ion. The structure of a glycan cannot be inferred from its composition. Therefore, looking at the data reported in Table 3.1, it is not possible to determine which type of O-glycosylation is present on each glycopeptide identified. To better interpret the data obtained, the potential glycosylation sites identified in this analysis were compared with those reported in the literature for rabbit  $\alpha$ DG (Stalnakier et al., 2010). In Stalnakier's study, sites S391, T395, S398, and S475 were reported as O-mannosylated with compositions Hex<sub>2</sub>HexNAc<sub>1</sub> (core M1), Hex<sub>2</sub>HexNAc<sub>1</sub>NeuAc<sub>1</sub> (core M1 capped with NeuAc) or Hex<sub>3</sub>HexNAc<sub>2</sub>NeuAc<sub>2</sub> (core M2 capped with two NeuAc). Based on this information, the glycan compositions associated with  $\alpha$ DG peptides identified in this study could be explained and correlated to specific glycan structures. For example, glycopeptide <sup>390</sup>VSDAGTVVSGQIR<sup>402</sup> presents three O-mannosylation sites and was identified with the glycan Hex<sub>2</sub>HexNAc<sub>1</sub>NeuAc<sub>1</sub> whose composition is consistent with core M1 capped with NeuAc. Other glycoforms of the peptide

were identified as having glycan composition of Hex<sub>3</sub>HexNAc<sub>2</sub>NeuAc<sub>1</sub>, consistent with core M2 capped with one NeuAc, and Hex<sub>1</sub> or Hex<sub>2</sub> which could be one or two O-linked mannose respectively. Stalnaker's work also reported sites T463, T464, T469, T473, S475 and S478 to be modified with core 1 (O-GalNAc-type) glycans. An example of an O-GalNAc peptide identified in this analysis and reported in Table 3.1 is <sup>462</sup>VTTKAPITR<sup>470</sup> with glycan composition Hex<sub>1</sub>HexNAc<sub>1</sub>NeuAc<sub>2</sub>, consistent with core 1 glycan capped with two NeuAc.

Peptides modified with core M3 glycans and/or extended with Rbo5P or GroP were not detected in the analysis presented here. Core M3 glycans were reported to modify sites T317, T319, T379 and T381 (Yagi et al., 2013; Yoshida-Moriguchi et al., 2010) in αDG from mice. The same sites were not detected as glycosylated in rabbit αDG (Stalnaker et al., 2010). Considering that sites T317, T319, T379 and T381 are conserved among αDG vertebrate homologues (Imae et al., 2021; Toledo et al., 2012), it is safe to assume that these sites could be modified with core M3 glycans in rabbit αDG. The peptide <sup>374</sup>GAIQTPTLGPIQPTRVSDAGTVVSGQIR<sup>402</sup>, which includes sites T379 and T381, was identified in this analysis as glycosylated with glycan composition Hex<sub>2</sub>HexNAc<sub>2</sub>NeuAc<sub>4</sub> which is not consistent with core M3 (Hex<sub>1</sub>HexNAc<sub>2</sub>Phos<sub>1</sub>).

The digestion of FLAGαDG<sub>r</sub> was performed using trypsin. Trypsin is widely used for peptide analysis because it cleaves proteins into peptides at an average size of about 700 to 1500 Da, which is ideal for mass spectrometric analysis. Moreover, by cleaving the C-terminal to K and R, trypsin generates peptides with a positive charge at the C-terminus, which is advantageous for the peptide fragmentation analysis (Huang et al., 2005). Due to steric hindrance, tryptic peptides are more challenging to obtain in the presence of glycans. Moreover, missed tryptic cleavages generate bigger peptides that are more difficult to analyse (Dodds et al., 2009). For these reasons, using proteases different from trypsin can help increase the chances to identify glycopeptides. Proteinase K is an example of such a protease. It cleaves the C-terminal of aliphatic and aromatic amino acids and has been used to generate glycopeptides for MS analysis (Zauner et al., 2010). Thus, FLAGαDG<sub>r</sub> was digested with proteinase K as an alternative to trypsin, and the obtained glycopeptides were analysed again by LC-MS/MS. The mapped sequence of αDG from analysis of the proteinase K digest is shown in Figure 3.15

1 MRMSVGLSLL LPLWGRTFLL LLCVAVAQSH **WPSEPSEAVR** **DWENQLEASM** 50  
 51 HSVLSDLHEA LPTVVGIPDG TAVVGRSFRV TIPTDLIGSS **GEVIKVSTAG** 100  
 101 **KEVLPS**WLHW DPQSHTLEGL PLDTDKGVHY ISVSAAQLDA NGSHIPQTSS 150  
 151 VFSIEVYPED HSEPQSVRAA SPDLGEEAAS **ACAAEEPVTV** **LTVILDADLT** 200  
 201 KMPKQRIDL LHRMQSFSEV ELHNMKLVV VNNRLFDMSA FMAGPGNAKK 250  
 251 VVENGALLSW KLGCSLNQNS VPDIRGVEAP AREGTMSAQL GYPVVGWHIA 300  
 301 NKKPPLPKRI RRQIHATPTP VTAIGPPTTA **IQEPPSRIVP** **TPTSPAIIAPP** 350  
 351 TETMAPVVRD PVPGKPTVTT RTRGAIQTP **TLGPIQPTRV** **SDAGTVVSGQ** 400  
 401 **IRATVTIPGY** VEPTAVATPP TTTTKKPRVS TPKPATPSTD **SSATTTTRRPT** 450  
 451 KKPRTPRPVP RVTTKAPITR LETASPTRI RTTTSVGPVPRG GEPNQRPELK 500  
 501 NHIDRVDAWV GTYF**EVKIPS** **DTFYDKEDTT** TDKLKLTLKL REQQLVGEKS 550  
 551 WVQFNSNSQL MYGLPDSSHV GKHEYFMHAT DKGGLS**AVDA** **FEIHVHKRPQ** 600  
 601 GDKAPARFKA KFGVDPAPVV NDIHKKIALV KKLAFAFGDR NCSTVTLQNI 650  
 651 TRG

Figure 3.15 / Mapped sequence of peptides identified in the mass spectrometry analysis of FLAG $\alpha$ DGr after proteinase K digestion. Peptides identified are reported in red.

The peptides identified cover ~17% of the  $\alpha$ DG sequence; therefore, proteinase K digestion in this context does not improve the overall sequence coverage of  $\alpha$ DG. Glycopeptides were also identified from the analysis of FLAG $\alpha$ DGr after proteinase K digestion and are listed in Table 3.2

Peptide sequence	Glycan composition
<sup>379</sup> <u>T</u> P <u>T</u> L <u>G</u> P <u>I</u> Q <u>P</u> T <u>R</u> V <u>S</u> D <u>A</u> <sup>393</sup>	Hex <sub>2</sub> HexNAc <sub>2</sub> NeuAc <sub>2</sub>
<sup>379</sup> <u>T</u> P <u>T</u> L <u>G</u> P <u>I</u> Q <u>P</u> T <u>R</u> V <u>S</u> D <u>A</u> <sup>393</sup>	Hex <sub>4</sub> HexNAc <sub>4</sub>
<sup>379</sup> <u>T</u> P <u>T</u> L <u>G</u> P <u>I</u> Q <u>P</u> T <u>R</u> V <u>S</u> D <u>A</u> <sup>393</sup>	Hex <sub>4</sub> HexNAc <sub>4</sub> NeuAc <sub>1</sub>
<sup>379</sup> <u>T</u> P <u>T</u> L <u>G</u> P <u>I</u> Q <u>P</u> T <u>R</u> V <u>S</u> D <u>A</u> <sup>393</sup>	Hex <sub>3</sub> HexNAc <sub>3</sub> NeuAc <sub>3</sub>
<sup>379</sup> <u>T</u> P <u>T</u> L <u>G</u> P <u>I</u> Q <u>P</u> T <u>R</u> V <u>S</u> D <u>A</u> <sup>393</sup>	Hex <sub>3</sub> HexNAc <sub>3</sub> NeuAc <sub>2</sub>
<sup>394</sup> G <u>T</u> V <u>V</u> S <u>G</u> Q <u>I</u> R <sup>402</sup>	Hex <sub>1</sub> HexNAc <sub>1</sub> NeuAc <sub>2</sub>
<sup>400</sup> Q <u>I</u> R <u>A</u> T <u>V</u> T <u>I</u> P <u>G</u> <sup>409</sup>	Hex <sub>1</sub> HexNAc <sub>1</sub> NeuAc <sub>2</sub>
<sup>401</sup> <u>I</u> R <u>A</u> T <u>V</u> T <u>I</u> P <u>G</u> <sup>409</sup>	Hex <sub>1</sub> HexNAc <sub>1</sub> NeuAc <sub>2</sub>
<sup>442</sup> <u>S</u> <u>A</u> T <u>T</u> T <u>R</u> R <u>P</u> T <sup>450</sup>	Hex <sub>3</sub> HexNAc <sub>3</sub>

Table 3.2 / List of the glycopeptides identified by LC-MS/MS from the proteinase K digest of recombinant  $\alpha$ DG from rabbit (FLAG $\alpha$ DGr). Potential O-glycosylation sites are underlined. Hex denotes hexose, e.g., galactose (Gal), glucose (Glc), or mannose (Man). HexNAc denotes N-acetyl

hexosamine, e.g., *N*-acetyl galactosamine (GalNAc) or *N*-acetyl glucosamine (GlcNAc). NeuAc denotes *N*-acetyl neuraminic acid.

New glycosylation sites were identified after proteinase K digestion compared to the data obtained after trypsin digestion. In particular, the sites T446 and T450 were reported in the literature to have O-GalNAc-type glycans (Stalnaker et al., 2010) and were found after proteinase K digestion in the peptide <sup>442</sup>SATTTRRPT<sup>450</sup>. The glycan Hex<sub>3</sub>HexNAc<sub>3</sub> associated with the peptide has a composition that can be explained, for example, by the co-presence of three core 1 (Hex<sub>1</sub>HexNAc<sub>1</sub>) glycans on the peptide. This observation suggests that sites other than T446 and T450 might be modified in the peptide <sup>442</sup>SATTTRRPT<sup>450</sup>.

Similarly to the experiment performed with trypsin, the digestion of FLAG $\alpha$ DGr with proteinase K did not result in detection by mass spectrometry of glycopeptides modified with Rbo5P or GroP. Rbo5P, differently from GroP, can be further modified with matriglycan polysaccharide, which, due to its size and probable heterogeneity, would not be expected to be detected under the conditions for glycopeptide analysis by mass spectrometry. In the literature, evidence is presented for a glycopeptide modified with core M3 extended with one or two units of Rbo5P or one unit of GroP without matriglycan (Yagi et al., 2016). It is worth pointing out that glycoforms of the peptide modified with Rbo5P without the matriglycan were identified in Yagi's work by expressing  $\alpha$ DG peptide and not full-length  $\alpha$ DG. It was recently reported that full-length  $\alpha$ DG, and in particular its N-terminal domain, was required for the extension of matriglycan by the activity of LARGE. Therefore, the glycosylation state of an overexpressed peptide from  $\alpha$ DG, lacking the  $\alpha$ DG N-terminus, is expected to differ from the glycosylation observed on a peptide obtained from full-length  $\alpha$ DG. This observation could explain why glycopeptides modified with Rbo5P were not found in the experiments described here.

### **3.4.4 Rbo5P is detected in recombinant $\alpha$ DG**

The method developed in Chapter 2 was designed to help identify glycopeptides modified with Rbo5P through a selective chemical tagging strategy. Treatment with sodium periodate under the condition used is expected to oxidise Rbo5P and generate an aldehyde on the alditol (Praisman et al., 2016). If Rbo5P is modified with matriglycan, the oxidation would release the polysaccharide, considerably improving the chances of detecting the glycopeptide by mass spectrometry. Oxidised Rbo5P on a glycopeptide would be susceptible to biotinylation if treated with aminoxy-biotin, and

the presence of the affinity tag on the modified alditol would help selectively purify proteins with Rbo5P. To test the efficacy of oxidation and biotinylation on a protein modified with Rbo5P, FLAG $\alpha$ DGr was first treated with sodium periodate and aminoxy-biotin, then digested with trypsin into peptides for mass spectrometric analysis. The treatment is also expected to biotinylate NeuAc units on the protein, which is not of interest in the context of the experiment. Treating FLAG $\alpha$ DGr with a sialidase would have prevented NeuAc from being modified. Nevertheless, the presence of biotinylated and sialylated peptides was not expected to hinder the detection of the glycopeptides bearing Rbo5P since the sample analysed consisted of a single protein. The mapped sequence of  $\alpha$ DG from this analysis is shown in Figure 3.16

1	MRMSVGLSLL	LPLWGRTFLL	LLCVAVAQSH	WPSEPSEAVR	DWENQLEASM	50
51	HSVLSDLHEA	LPTVVGIPDG	TAVVGRSFRV	TIPTDLIGSS	GEVIKVSTAG	100
101	KEVLPSWLHW	DPQSHTLEGL	PLDSDKGVHY	ISVSAQLDA	NGSHIPQTSS	150
151	VFSIEVYPED	HSEPQSVRAA	SPDLGEEAAS	ACAAEEPVTV	LTVILDADLT	200
201	KMTPKQRIDL	LHRMQSFSEV	ELHNMKLVVP	VNNRFLDMSA	FMAGPGNAKK	250
251	VVENGALLSW	KLGC SLNQNS	VPDIRGVEAP	AREGTMSAQL	GYPVVGWHIA	300
301	NKKPPLPKRI	RRQIHATPTP	VTAIGPPTTA	IQEPPSRIVP	TPTSPAIAAPP	350
351	TETMAPPVRD	PVPGKPTVTT	RTRGAIQTP	TLGPIQPTRV	SDAGTVVSGQ	400
401	IRATVTIPGY	VEPTAVATPP	TTTTKKPRVS	TPKPATPSTD	SSATTTTRPT	450
451	KKPRTPRPVP	RVTTKAPITR	LETASPTRI	RTTTSVPRG	GEPNQRPELK	500
501	NHIDRVD <del>AW</del>	GTYFEVKIPS	DTFYDKEDTT	TDK <del>L</del> KLTLKL	REQQLVGEKS	550
551	WVQFNSNSQL	MYGLPDSSHV	GKHEYFMHAT	DKGGLSAVDA	FEIHVHKRPQ	600
601	GDKAPARFKA	KFVGDPAPVV	NDIHKKIALV	KKLAFAFGDR	NCSTVTLQNI	650
651	TRG					

*Figure 3.16 / Mapped sequence of peptides identified in the mass spectrometry analysis of FLAG $\alpha$ DGr after oxidation, biotinylation and trypsin digestion. Peptides identified are reported in red.*

The peptides identified cover ~50% of the  $\alpha$ DG sequence. The glycopeptides identified from the analysis of FLAG $\alpha$ DGr after oxidation, biotinylation and trypsin digestion are listed in Table 3.3.

Peptide sequence	Glycan composition	
<sup>313</sup> QIHATPTPVTAIGPPTTAIQEPPSR <sup>337</sup>	HexNAc <sub>1</sub>	
<sup>360</sup> DPVPGKPTVTTTR <sup>371</sup>	Hex <sub>1</sub> HexNAc <sub>1</sub> NeuAc <sub>1</sub>	
<sup>360</sup> DPVPGKPTVTTTR <sup>371</sup>	Hex <sub>1</sub> HexNAc <sub>1</sub> NeuAc <sub>2</sub>	
<sup>361</sup> PVPGKPTVTTTR <sup>372</sup>	Hex <sub>1</sub> HexNAc <sub>1</sub> NeuAc <sub>2</sub>	
<sup>374</sup> GAIQTPTLGPIQPTR <sup>389</sup>	Hex <sub>3</sub>	
<sup>374</sup> GAIQTPTLGPIQPTR <sup>389</sup>	Hex <sub>1</sub> HexNAc <sub>1</sub>	
<sup>374</sup> GAIQTPTLGPIQPTR <sup>389</sup>	Hex <sub>2</sub> HexNAc <sub>1</sub>	
<sup>374</sup> GAIQTPTLGPIQPTR <sup>389</sup>	Hex <sub>2</sub> HexNAc <sub>2</sub>	
<sup>374</sup> GAIQTPTLGPIQPTR <sup>389</sup>	Hex <sub>2</sub> HexNAc <sub>1</sub> NeuAc <sub>2</sub>	
<sup>374</sup> GAIQTPTLGPIQPTR <sup>389</sup>	Hex <sub>3</sub> HexNAc <sub>2</sub> Phos <sub>1</sub>	
<sup>374</sup> GAIQTPTLGPIQPTR <sup>389</sup>	Hex <sub>3</sub> HexNAc <sub>3</sub> NeuAc <sub>2</sub> Phos <sub>1</sub>	
<sup>374</sup> GAIQTPTLGPIQPTR <sup>389</sup>	Hex <sub>3</sub> HexNAc <sub>2</sub> Phos <sub>1</sub> GroP <sub>1</sub>	
<sup>374</sup> GAIQTPTLGPIQPTR <sup>389</sup>	Hex <sub>3</sub> HexNAc <sub>2</sub> Phos <sub>1</sub> Rbo5P <sub>1</sub>	
<sup>374</sup> GAIQTPTLGPIQPTRVSDAGTVVSGQIR <sup>402</sup>	Hex <sub>2</sub> HexNAc <sub>2</sub>	
<sup>374</sup> GAIQTPTLGPIQPTRVSDAGTVVSGQIR <sup>402</sup>	Hex <sub>2</sub> HexNAc <sub>2</sub> NeuAc <sub>2</sub>	
<sup>374</sup> GAIQTPTLGPIQPTRVSDAGTVVSGQIR <sup>402</sup>	Hex <sub>2</sub> HexNAc <sub>2</sub> NeuAc <sub>4</sub>	*
<sup>374</sup> GAIQTPTLGPIQPTRVSDAGTVVSGQIR <sup>402</sup>	Hex <sub>3</sub> HexNAc <sub>3</sub> NeuAc <sub>2</sub>	
<sup>374</sup> GAIQTPTLGPIQPTRVSDAGTVVSGQIR <sup>402</sup>	Hex <sub>3</sub> HexNAc <sub>3</sub> NeuAc <sub>3</sub>	
<sup>374</sup> GAIQTPTLGPIQPTRVSDAGTVVSGQIR <sup>402</sup>	Hex <sub>3</sub> HexNAc <sub>3</sub> NeuAc <sub>4</sub>	
<sup>390</sup> VSDAGTVVSGQIR <sup>402</sup>	Hex <sub>1</sub>	*
<sup>390</sup> VSDAGTVVSGQIR <sup>402</sup>	Hex <sub>2</sub>	*
<sup>390</sup> VSDAGTVVSGQIR <sup>402</sup>	Hex <sub>2</sub> HexNAc <sub>1</sub>	
<sup>390</sup> VSDAGTVVSGQIR <sup>402</sup>	Hex <sub>3</sub> HexNAc <sub>1</sub>	
<sup>390</sup> VSDAGTVVSGQIR <sup>402</sup>	Hex <sub>1</sub> HexNAc <sub>2</sub> NeuAc <sub>1</sub>	
<sup>390</sup> VSDAGTVVSGQIR <sup>402</sup>	Hex <sub>2</sub> HexNAc <sub>1</sub> NeuAc <sub>1</sub>	*
<sup>390</sup> VSDAGTVVSGQIR <sup>402</sup>	Hex <sub>2</sub> HexNAc <sub>2</sub> NeuAc <sub>1</sub>	
<sup>390</sup> VSDAGTVVSGQIR <sup>402</sup>	Hex <sub>2</sub> HexNAc <sub>2</sub> NeuAc <sub>2</sub>	
<sup>390</sup> VSDAGTVVSGQIR <sup>402</sup>	Hex <sub>3</sub> HexNAc <sub>1</sub> NeuAc <sub>1</sub>	
<sup>429</sup> VSTPKPATPSTDSSATTTTR <sup>447</sup>	Hex <sub>2</sub> HexNAc <sub>2</sub>	
<sup>429</sup> VSTPKPATPSTDSSATTTTR <sup>447</sup>	Hex <sub>2</sub> HexNAc <sub>3</sub>	
<sup>429</sup> VSTPKPATPSTDSSATTTTR <sup>447</sup>	Hex <sub>3</sub> HexNAc <sub>3</sub>	
<sup>429</sup> VSTPKPATPSTDSSATTTTR <sup>447</sup>	Hex <sub>1</sub> HexNAc <sub>1</sub> NeuAc <sub>1</sub>	
<sup>429</sup> VSTPKPATPSTDSSATTTTR <sup>447</sup>	Hex <sub>1</sub> HexNAc <sub>1</sub> NeuAc <sub>2</sub>	

462	<u>V</u> <u>T</u> <u>T</u> <u>K</u> <u>A</u> <u>P</u> <u>I</u> <u>T</u> <u>R</u> <sup>470</sup>	Hex <sub>2</sub> HexNac <sub>2</sub>	
462	<u>V</u> <u>T</u> <u>T</u> <u>K</u> <u>A</u> <u>P</u> <u>I</u> <u>T</u> <u>R</u> <sup>470</sup>	Hex <sub>1</sub> HexNac <sub>1</sub> NeuAc <sub>2</sub>	*
462	<u>V</u> <u>T</u> <u>T</u> <u>K</u> <u>A</u> <u>P</u> <u>I</u> <u>T</u> <u>R</u> <sup>470</sup>	Hex <sub>2</sub> HexNac <sub>2</sub> NeuAc <sub>1</sub>	*
462	<u>V</u> <u>T</u> <u>T</u> <u>K</u> <u>A</u> <u>P</u> <u>I</u> <u>T</u> <u>R</u> <sup>470</sup>	Hex <sub>3</sub> HexNac <sub>3</sub> NeuAc <sub>1</sub>	
462	<u>V</u> <u>T</u> <u>T</u> <u>K</u> <u>A</u> <u>P</u> <u>I</u> <u>T</u> <u>R</u> <sup>470</sup>	Hex <sub>4</sub> HexNac <sub>4</sub> NeuAc <sub>1</sub>	
471	<u>L</u> <u>E</u> <u>T</u> <u>A</u> <u>S</u> <u>P</u> <u>P</u> <u>T</u> <u>R</u> <sup>479</sup>	HexNac <sub>2</sub>	*
471	<u>L</u> <u>E</u> <u>T</u> <u>A</u> <u>S</u> <u>P</u> <u>P</u> <u>T</u> <u>R</u> <sup>479</sup>	Hex <sub>1</sub> HexNac <sub>2</sub>	
471	<u>L</u> <u>E</u> <u>T</u> <u>A</u> <u>S</u> <u>P</u> <u>P</u> <u>T</u> <u>R</u> <sup>479</sup>	Hex <sub>1</sub> HexNac <sub>4</sub>	
471	<u>L</u> <u>E</u> <u>T</u> <u>A</u> <u>S</u> <u>P</u> <u>P</u> <u>T</u> <u>R</u> <sup>479</sup>	Hex <sub>2</sub> HexNac <sub>2</sub>	*
471	<u>L</u> <u>E</u> <u>T</u> <u>A</u> <u>S</u> <u>P</u> <u>P</u> <u>T</u> <u>R</u> <sup>479</sup>	Hex <sub>2</sub> HexNac <sub>3</sub>	
471	<u>L</u> <u>E</u> <u>T</u> <u>A</u> <u>S</u> <u>P</u> <u>P</u> <u>T</u> <u>R</u> <sup>479</sup>	Hex <sub>2</sub> HexNac <sub>4</sub>	
471	<u>L</u> <u>E</u> <u>T</u> <u>A</u> <u>S</u> <u>P</u> <u>P</u> <u>T</u> <u>R</u> <sup>479</sup>	Hex <sub>3</sub> HexNac <sub>3</sub>	
471	<u>L</u> <u>E</u> <u>T</u> <u>A</u> <u>S</u> <u>P</u> <u>P</u> <u>T</u> <u>R</u> <sup>479</sup>	Hex <sub>1</sub> HexNac <sub>1</sub> NeuAc <sub>2</sub>	
471	<u>L</u> <u>E</u> <u>T</u> <u>A</u> <u>S</u> <u>P</u> <u>P</u> <u>T</u> <u>R</u> <sup>479</sup>	Hex <sub>1</sub> HexNac <sub>2</sub> NeuAc <sub>3</sub>	
471	<u>L</u> <u>E</u> <u>T</u> <u>A</u> <u>S</u> <u>P</u> <u>P</u> <u>T</u> <u>R</u> <sup>479</sup>	Hex <sub>2</sub> HexNac <sub>2</sub> NeuAc <sub>1</sub>	
471	<u>L</u> <u>E</u> <u>T</u> <u>A</u> <u>S</u> <u>P</u> <u>P</u> <u>T</u> <u>R</u> <sup>479</sup>	Hex <sub>2</sub> HexNac <sub>2</sub> NeuAc <sub>2</sub>	
471	<u>L</u> <u>E</u> <u>T</u> <u>A</u> <u>S</u> <u>P</u> <u>P</u> <u>T</u> <u>R</u> <sup>479</sup>	Hex <sub>2</sub> HexNac <sub>3</sub> NeuAc <sub>1</sub>	*
471	<u>L</u> <u>E</u> <u>T</u> <u>A</u> <u>S</u> <u>P</u> <u>P</u> <u>T</u> <u>R</u> <sup>479</sup>	Hex <sub>3</sub> HexNac <sub>3</sub> NeuAc <sub>1</sub>	*
471	<u>L</u> <u>E</u> <u>T</u> <u>A</u> <u>S</u> <u>P</u> <u>P</u> <u>T</u> <u>R</u> <sup>479</sup>	Hex <sub>3</sub> HexNac <sub>3</sub> NeuAc <sub>2</sub>	
471	<u>L</u> <u>E</u> <u>T</u> <u>A</u> <u>S</u> <u>P</u> <u>P</u> <u>T</u> <u>R</u> <sup>479</sup>	Hex <sub>4</sub> HexNac <sub>4</sub> NeuAc <sub>2</sub>	
480	<u>I</u> <u>R</u> <u>T</u> <u>T</u> <u>T</u> <u>S</u> <u>G</u> <u>V</u> <u>P</u> <u>R</u> <sup>489</sup>	Hex <sub>2</sub> HexNac <sub>2</sub>	
482	<u>T</u> <u>T</u> <u>T</u> <u>S</u> <u>G</u> <u>V</u> <u>P</u> <u>R</u> <sup>489</sup>	Hex <sub>1</sub> HexNac <sub>1</sub> NeuAc <sub>1</sub>	
482	<u>T</u> <u>T</u> <u>T</u> <u>S</u> <u>G</u> <u>V</u> <u>P</u> <u>R</u> <sup>489</sup>	Hex <sub>2</sub> HexNac <sub>2</sub> NeuAc <sub>1</sub>	
518	<u>I</u> <u>P</u> <u>S</u> <u>D</u> <u>I</u> <u>F</u> <u>Y</u> <u>D</u> <u>K</u> <u>E</u> <u>D</u> <u>T</u> <u>T</u> <u>T</u> <u>D</u> <u>K</u> <u>L</u> <u>K</u> <sup>535</sup>	Hex <sub>2</sub> HexNac <sub>2</sub>	
518	<u>I</u> <u>P</u> <u>S</u> <u>D</u> <u>I</u> <u>F</u> <u>Y</u> <u>D</u> <u>K</u> <u>E</u> <u>D</u> <u>T</u> <u>T</u> <u>T</u> <u>D</u> <u>K</u> <u>L</u> <u>K</u> <sup>535</sup>	Hex <sub>1</sub> HexNac <sub>1</sub> NeuAc <sub>2</sub>	*
518	<u>I</u> <u>P</u> <u>S</u> <u>D</u> <u>I</u> <u>F</u> <u>Y</u> <u>D</u> <u>K</u> <u>E</u> <u>D</u> <u>T</u> <u>T</u> <u>T</u> <u>D</u> <u>K</u> <u>L</u> <u>K</u> <sup>535</sup>	Hex <sub>2</sub> HexNac <sub>2</sub> NeuAc <sub>2</sub>	

Table 3.3 / List of the glycopeptides identified by LC-MS/MS from the oxidation, biotinylation and tryptic digest of recombinant  $\alpha$ DG from rabbit (FLAG $\alpha$ DGr). Potential O-glycosylation sites are underlined. Hex denotes hexose, e.g., galactose (Gal), glucose (Glc), or mannose (Man). HexNac denotes N-acetyl hexosamine, e.g., N-acetyl galactosamine (GalNac) or N-acetyl glucosamine (GlcNac). NeuAc denotes N-acetyl neuraminic acid. Asterisks denote glycopeptides that were also identified after tryptic digestion of FLAG $\alpha$ DGr without prior oxidation and biotinylation.

Most of the glycopeptides identified after digestion of FLAG $\alpha$ DGr (listed in Table 3.1) are also identified if FLAG $\alpha$ DGr is oxidised and biotinylated before being digested (listed in Table 3.3 with an asterisk). The oxidation/biotinylation treatment results in the identification of a more significant number of glycopeptides compared to the same analysis performed without the treatment. One possible explanation for the difference between the two results is that FLAG $\alpha$ DGr, after oxidation, would be more accessible to the proteolytic enzyme since it lost the matriglycan. Several glycopeptides from Table 3.3 present a glycan composition with one or more NeuAc units. The molecular weight observed for a NeuAc unit is consistent with the mass of the monosaccharide unmodified. The result suggests that the sialylated glycopeptides identified in the analysis were not detected as oxidised and/or biotinylated. Moreover, glycoforms presenting more than one unit of NeuAc were observed to be completely unmodified. This observation contrasts with the results presented in Chapter 2, where fetuin was oxidised and biotinylated. Glycopeptides derived from fetuin were identified as biotinylated and oxidised, while this is not the case for FLAG $\alpha$ DGr treated with the same method. Moreover, the  $m/z$  613.26, marker of biotinylated and oxidised NeuAc, was not detected in any of the product ion spectra obtained from the LC-MS/MS analysis of peptides from FLAG $\alpha$ DGr. One possible explanation is that the sodium periodate and aminoxy-biotin treatments did not completely affect FLAG $\alpha$ DGr. As a result, the sample had a mixture of biotinylated and non-biotinylated proteins. The glycopeptides obtained from the digestion of this sample were also a mixture of biotinylated and non-biotinylated species, the latter of which were exclusively detected by mass spectrometry.

One striking result is that after oxidation and biotinylation of FLAG $\alpha$ DGr, glycopeptides modified with Rbo5P and GroP were detected. In particular, the peptide <sup>374</sup>GAIIQTPTLGPIQPTR<sup>389</sup> was identified in this analysis modified with core M3 glycans extended with one unit of Rbo5P or GroP. The glycoforms of the peptide with core M3 glycans listed in Table 3.3 are reported for clarity in Table 3.4.

Peptide sequence	Glycan composition
GAI <u>IQ</u> TPTLGPI <u>Q</u> PTR	Hex <sub>3</sub> HexNAc <sub>2</sub> Phos <sub>1</sub>
	Hex <sub>3</sub> HexNAc <sub>3</sub> NeuAc <sub>2</sub> Phos <sub>1</sub>
	Hex <sub>3</sub> HexNAc <sub>2</sub> Phos <sub>1</sub> GroP <sub>1</sub>
	Hex <sub>3</sub> HexNAc <sub>2</sub> Phos <sub>1</sub> Rbo5P <sub>1</sub>

Table 3.4 / List of the glycoforms of the peptide GAIIQTPTLGPIQPTR identified by LC-MS/MS from the oxidation, biotinylation and tryptic digest of recombinant  $\alpha$ DG from rabbit (FLAG $\alpha$ DGr). Potential O-glycosylation sites are underlined. Hex denotes hexose, e.g., galactose (Gal), glucose

*(Glc)*, or mannose (*Man*). *HexNAc* denotes *N*-acetyl hexosamine, e.g., *N*-acetyl galactosamine (*GalNAc*) or *N*-acetyl glucosamine (*GlcNAc*). *NeuAc* denotes *N*-acetyl neuraminic acid. *Phos* denotes a phosphate group.

The identification of the glycoforms listed in Table 3.4 was achieved by manually interpreting the data, since the bioinformatic software search was unable to detect them. Ions originated from the fragmentation of glycans containing Rbo5P and GroP ( $m/z$  233.04 and  $m/z$  358.09, respectively) were used to search the product ion spectra of glycopeptides obtained from the mass spectrometric analysis.

As described earlier, sites T379 and T381 are expected to be O-mannosylated with core M3 glycans in rabbit  $\alpha$ DG by similarity with those identified on the same sites in mouse  $\alpha$ DG. While core M3 glycans are reported in the literature to be extended with Rbo5P or GroP, the only site where this was observed is T317. Moreover, the only evidence of Rbo5P and GroP on a peptide is from the analysis of overexpressed peptide from  $\alpha$ DG. These observations highlight that in this study, for the first time, Rbo5P and GroP were identified on a full-length  $\alpha$ DG and that sites T379 or T381 can carry an extended core M3 glycan.

Product ion spectra of the glycopeptide <sup>374</sup>GAIHQPTLGIQPTR<sup>389</sup> modified core M3 glycans extended with Rbo5P and GroP are shown in Figure 3.17 and Figure 3.18, respectively.

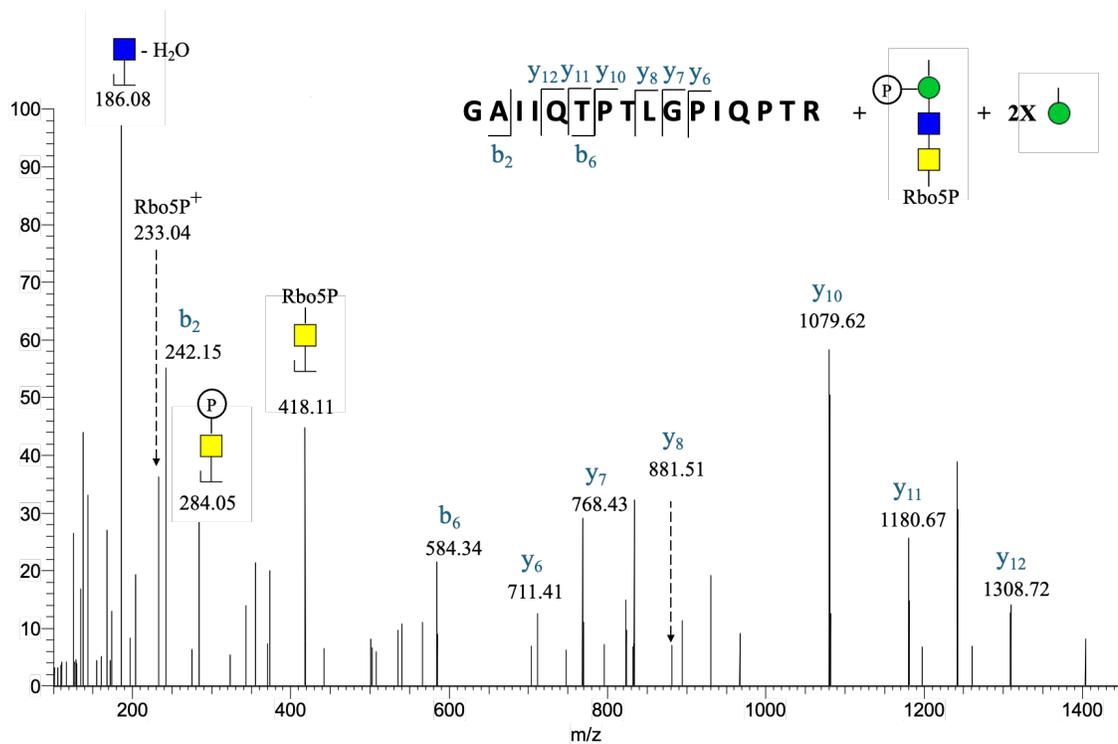


Figure 3.17 / HCD product ion spectrum of a tryptic glycopeptide from FLAGαDGr treated with sodium periodate and aminoxy-biotin oxidation. The precursor ion  $[M+3H]^{3+}$  at  $m/z$  950.43 has a monoisotopic mass corresponding to that calculated for the peptide G A I I Q T P T L G P I Q P T R bearing 2 HexNAc, 3 Hex, 1 Phos and 1 Rbo5P ( $\Delta=0$  ppm). Product ions have a 1+ charge (see section 2.3.7).

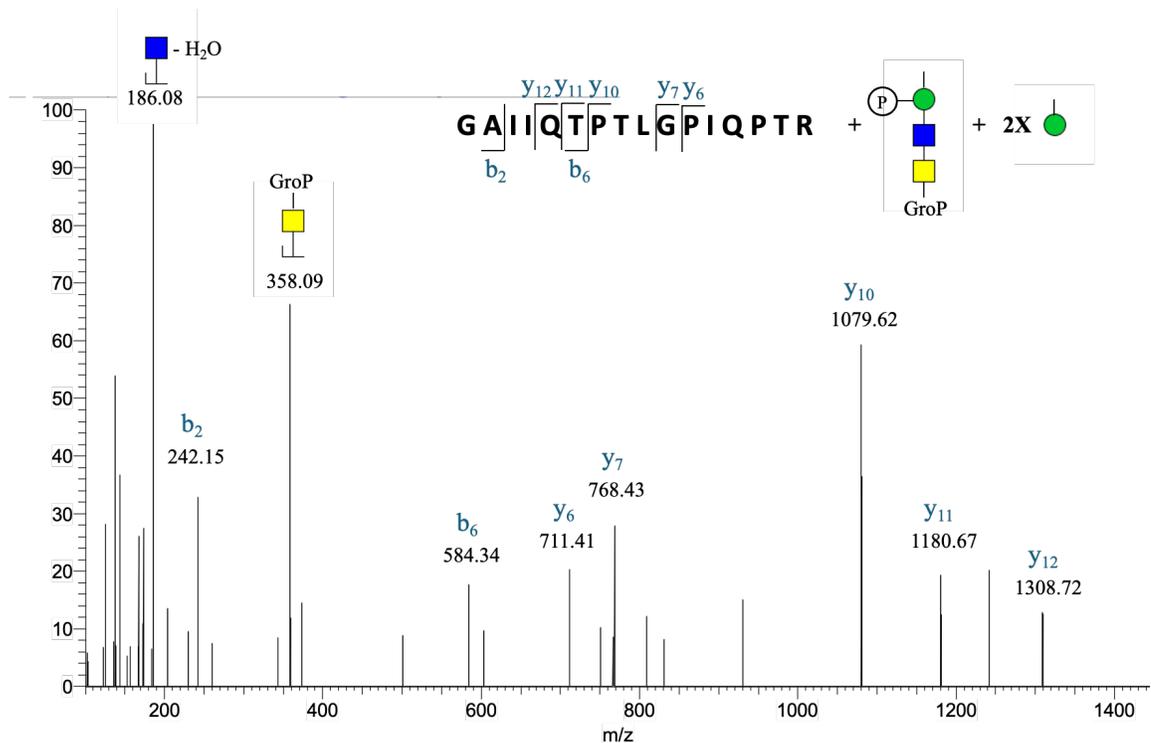


Figure 3.18 / HCD product ion spectrum of a tryptic glycopeptide from FLAG $\alpha$ DGr treated with sodium periodate and aminoxy-biotin oxidation. The precursor ion  $[M+3H]^{3+}$  at  $m/z$  930.42 has a monoisotopic mass corresponding to that calculated for the peptide GAIQTPTLGP IQPTR bearing 2 HexNAc, 3 Hex, 1 Phos and 1 GroP ( $\Delta=11$  ppm). Product ions have a 1+ charge (see section 2.3.7).

In the product ion spectrum shown in Figure 3.17, Rbo5P is expected to be oxidised and biotinylated since FLAG $\alpha$ DGr was treated with sodium periodate and aminoxy-biotin before tryptic digestion. Nevertheless, Rbo5P is not present in its oxidised form in the spectrum and biotin is not detected in any of the product ion spectra acquired in the analysis. Even if the biotinylation reaction did not go to completion, the fact that the treatment has led to the detection of RboP, which was not found in the previous analysis of untreated glycoprotein, suggests that the oxidation reaction has been effective. It would thus be reasonable to expect detection of oxidised RboP products. One possible explanation for the detection of Rbo5P despite the absence of any evidence of oxidation or biotinylation could be that the treatment did affect other glycosylation sites in  $\alpha$ DG, releasing the matriglycan, which made the rest of the protein more accessible for proteolytic digestion. Based on this hypothesis, the glycopeptide identified with Rbo5P would correspond to a glycoform unaffected by oxidation.

### 3.5 Conclusions

The purpose of the work presented in this chapter was to develop a procedure for the production of recombinant full-length  $\alpha$ DG and to apply the method developed in Chapter 2 for the oxidation and tagging of Rbo5P on the recombinant protein, known to be modified with Rbo5P. The results show that after oxidation and biotinylation of  $\alpha$ DG peptides carrying core M3 extended with Rbo5P or GroP were detected. Contrary to expectations, Rbo5P is not oxidised or biotinylated. The finding is promising since it opens up the opportunity to treat cell lysates with the same method for the detection of Rbo5P on endogenous  $\alpha$ DG and the potential identification of and identify novel proteins modified with Rbo5P.

Additionally, the data presented here have demonstrated the presence of Rbo5P and GroP on full-length  $\alpha$ DG for the first time, rather than the truncated protein used in previous reports. The results also suggest the presence of Rbo5P at a different glycosylation site, at which this alditol had not previously been detected. These findings contribute to the growing body of knowledge on the glycosylation of  $\alpha$ DG and will impact our understanding of dystroglycanopathies in which Rbo5P modification or extension of core M3 glycans is impaired.

## Chapter 4: Towards the identification of novel proteins modified with ribitol-5-phosphate

### 4.1 Introduction

$\alpha$ DG is the only mammalian protein known to date to be modified with Rbo5P. The discovery of Rbo5P in  $\alpha$ DG glycans is the result of an interest in  $\alpha$ DG glycosylation due to its connection to muscular dystrophy and brain malformation (Beltrán-Valero De Bernabé et al., 2002; Holzfeind et al., 2002). For a long time,  $\alpha$ DG was known for being part of the dystrophin-glycoprotein complex (Ervasti and Campbell, 1991) and for its function in binding laminin in the extracellular matrix (Ervasti and Campbell, 1993). A study of muscular dystrophy in mice showed for the first time that  $\alpha$ DG glycosylation was altered in the diseased animals (Grewal et al., 2001). Since then, the enzymes involved in the biosynthetic pathway of the O-mannosyl glycans on  $\alpha$ DG have been identified and mutations in them correlated with several diseases named dystroglycanopathies (Endo, 2015). One of these enzymes is LARGE, the mutation of which is responsible for the congenital muscular dystrophy type 1D (MCD1D). LARGE is a glycosyltransferase with dual activity:  $\alpha$ 3-xylosyltransferase activity and a  $\beta$ 3-glucuronyltransferase activity (Inamori et al., 2012). It is responsible for the synthesis of a polysaccharide composed of the repeating disaccharide Xyl $\alpha$ 1-3GlcA $\beta$ 1-3 (Inamori et al., 2012). When overexpressed, LARGE was found to increase the apparent molecular weight of  $\alpha$ DG on SDS-PAGE (Brockington et al., 2005) and to overcome the  $\alpha$ DG glycosylation defect in cells from individuals with MCD1D (Barresi et al., 2004). The linkage between the polysaccharide synthesised by LARGE, named matriglycan (Yoshida-Moriguchi and Campbell, 2014) and  $\alpha$ DG was unknown until 2016 when three different labs expressed the  $\alpha$ DG peptide carrying matriglycan and characterised its glycosylation by mass spectrometry (Kanagawa et al., 2016; Praissman et al., 2016; Yagi et al., 2016). The three studies showed for the first time that  $\alpha$ DG has a core M3 O-mannosyl glycan extended with a tandem repeat of Rbo5P. The second moiety of Rbo5P is modified with Xyl and GlcA added sequentially by two glycosyltransferases. The resulting glycan becomes the substrate for the LARGE enzymatic activity which initiates and extends the matriglycan polysaccharide.

The story of how Rbo5P was discovered in mammalian cells shows that it relied on the targeted analysis of a single glycoprotein,  $\alpha$ DG (Kanagawa et al., 2016; Praissman et al., 2016; Yagi et al., 2016). However, the search for novel mammalian proteins similarly modified with Rbo5P should rely on a method that specifically targets proteins modified with the alditol. A common

glycoproteomic strategy involves the enrichment of glycoproteins by exploiting their affinity for antibodies or lectins (Abbott and Pierce, 2010). The enrichment step is important because it reduces the complexity of the sample to analyse by mass spectrometry and reduces the background of non-glycosylated proteins, which can be more prominent than glycoproteins in the sample. Examples of lectins used for enrichment strategies are wheat germ agglutinin (WGA), which recognises glycopeptides modified with O-GlcNAc (Ma et al., 2013) and *Aleuria aurantia* lectin (AAL) which has an affinity for terminal Fuc (Bandini et al., 2016). Despite the existence of tens of lectins with different glycan specificities (Kobayashi et al., 2014), a glycan-binding protein specific to Rbo5P is unknown. Other strategies can be used to enrich glycoproteins when affinity purification using antibodies or lectins is not an option. The MOE strategy using unnatural monosaccharide analogues is a way to incorporate bioorthogonal functional groups into glycoproteins, allowing their subsequent conjugation with a tag and enrichment (see Chapters 1 and 2). Another strategy that involves tagging a glycoprotein with an enrichment probe is the chemical modification of glycans on glycoproteins by periodate oxidation (see Chapters 1 and 2).

Regardless of the method used, the enrichment of glycoproteins must be followed by their analysis and identification. A powerful technique for this purpose is the separation of peptides and glycopeptides, obtained from the digestion of proteins, by liquid chromatography and then their identification using tandem mass spectrometry, usually in an LC-MS/MS setup. This approach is referred to as bottom-up shotgun analysis (Zhang et al., 2013). In a typical shotgun experiment, peptides eluted from the column into the mass spectrometer are detected, isolated and fragmented to generate a product ion spectrum. This spectrum is given as input data to bioinformatic software like PEAKS or Byonic. A bioinformatic software uses a selection of protein sequences, for example, the annotated proteins reported for the organism under investigation, to generate virtual or '*in silico*' peptides. The *m/z* value of each peptide as well as the *m/z* of product ions generated from the peptide after '*in silico*' fragmentation are compared to the experimental *m/z* values obtained from mass spectrometric analysis. The result is a list of putative peptides identified and the names of the proteins of origin. Reducing the sample complexity is a way to increase the number of peptides identified. This is related to the fact that the mass spectrometer is usually set up so that the most intense peptide ions being detected in the mass spectrometer are selected for further fragmentation; therefore, to increase the number of (glyco)peptides analysed the sample complexity can be reduced first by enriching for the relevant subset of analytes and/or by introducing more than one step of liquid chromatography. An experiment with two chromatographic steps most usefully exploits two different stationary phases chemistries to separate analytes on the basis of different chemical

properties. An example is using a HILIC column separation (Zauner et al., 2011) as the first step and a C18 column as the second step inline with the mass spectrometer.

Despite all the different techniques available and improvements in mass spectrometric instrumentation and analysis, identifying glycoproteins remains challenging. Many glycoproteins are usually present in low abundance and in complex protein and glycoprotein mixtures. Moreover, each glycoprotein can exist in different glycoforms, which are characterized by a variety of possible glycan structures for a specific glycosylation site (microheterogeneity) and differences in the number of sites glycosylated (microheterogeneity). It is often necessary to devise a customised approach for different classes of glycoproteins to address the different challenges that can compromise the success of more generic analytical approaches.

## 4.2 Aims

The discovery of Rbo5P in  $\alpha$ DG O-mannosyl glycans opened a new chapter in the study of mammalian glycosylation. The existence of an otherwise un-used biosynthetic pathway that is specific for the synthesis of the Rbo5P sugar donor (CDP-Rbo) and the transfer of the alditol onto a glycan raises the question of whether there might be other proteins than  $\alpha$ DG that have this modification. These proteins that are so far unidentified as bearing this modification might have a similar function to  $\alpha$ DG and, therefore, be linked to muscular dystrophy and brain malformation. This chapter describes an investigation of raw mass spectrometric data available in public data repositories, with the purpose of determining whether these datasets contain any mass spectrometric evidence for glycopeptides carrying Rbo5P. The reason for re-analysing data is the relatively recent discovery of the alditol in mammalian cells. Studies published before the identification of Rbo5P on  $\alpha$ DG in 2016 would not have known to look for glycans containing Rbo5P when interpreting mass spectrometric data. Moreover, those carrying out other studies may not have thought to consider this modification if they were not working specifically in the  $\alpha$ DG/dystroglycan area.

This chapter also describes the application of a method for enriching proteins with Rbo5P from cells through chemical oxidation followed by ligation to a biotin-tagged oxime reagent. The method, development which is described in Chapter 2 and the results of testing it on  $\alpha$ DG described in Chapter 3, will in this chapter be used on a complex mixture, HEK293T cell lysates. Mass spectrometric analysis of glycoproteins enriched in this way have the potential to provide the identity of novel human proteins modified with Rbo5P.

## 4.3 Materials and Methods

### 4.3.1 Bioinformatic search

Raw mass spectrometric datasets deposited in the ProteomeXchange repository (<http://www.proteomexchange.org>) with the identifier PXD001468 (Chick et al., 2015) were used for automated database searching using the PEAKS software suite version X (Bioinformatics Solutions Inc., Waterloo, ON, Canada). The search criteria were set as follows: *Homo sapiens* proteome (Uniprot); trypsin as the enzyme with  $\leq 3$  missed cleavages and  $\leq 1$  non-specific cleavage; the error tolerances for the precursor of 5 ppm and 0.02 Da for fragment ions; carbamidomethyl cysteine as a fixed modification; oxidation of methionine and pyro-Glu from glutamine as variable modifications. The peptide match threshold ( $-10 \log P$ ) was set to 15. A PTM search was performed using the same criteria as above including as variable modifications the O-linked glycans of composition: Hex<sub>2</sub>HexNAc<sub>1</sub>, Hex<sub>3</sub>HexNAc<sub>2</sub>, Hex<sub>1</sub>HexNAc<sub>2</sub>Phos<sub>1</sub>, Hex<sub>1</sub>HexNAc<sub>2</sub>Phos<sub>1</sub>RboP<sub>1</sub>, Hex<sub>1</sub>HexNAc<sub>2</sub>Phos<sub>1</sub>RboP<sub>2</sub>, Hex<sub>1</sub>HexNAc<sub>2</sub>Phos<sub>1</sub>Rbo5P<sub>2</sub>Xyl<sub>1</sub>GlcA<sub>1</sub>, Hex<sub>1</sub>HexNAc<sub>2</sub>Phos<sub>1</sub>GroP<sub>1</sub>.

### 4.3.2 Cell lysate preparation

HEK293T cells were grown in T-75 cell culture flasks in Dulbecco's modified Eagle's medium (DMEM) with 10% fetal bovine serum (heat-inactivated), 2 mM L-glutamine, 100 units/mL penicillin, and 100  $\mu$ g/mL streptomycin and kept at 37°C in an incubator with 5% CO<sub>2</sub>. Cells at 90% confluence were washed with 1X PBS after removing the medium and incubated with 0.25% trypsin-EDTA until detachment from the bottom of the flask was observed. Cells were collected in a tube, centrifuged at 500 x g for 5 min, split and seeded into new flasks with fresh medium (10% confluence) and kept at 37 °C in an incubator with 5% CO<sub>2</sub>. To obtain cell pellets, the medium was removed from the flasks before cells were washed with 1X PBS and cells were harvested by scraping in 500  $\mu$ L of 1X PBS with protease inhibitor cocktail (1 mM AEBSF, 20  $\mu$ M leupeptin, 1  $\mu$ M pepstatin A). Cell pellets were obtained after centrifugation for 5 min at 500 x g. For cell lysate preparation, cell pellets were lysed in 250  $\mu$ L of lysis buffer (250 mM NaCl, 50 mM Tris-HCl pH 8.0, 0.5% NP-40 and protease inhibitor cocktail) per T-75 flask for 30 minutes on ice agitating every 5 min. The solution was centrifuged at 13000 x g for 30 min at 4°C, after which the supernatant was collected. Cell lysate concentration was measured with the Pierce BCA protein assay kit (Thermo Scientific) following the manufacturer's protocol.

### 4.3.3 Sialidase treatment

Fetuin (100 µg) was reconstituted in 50 mM sodium acetate pH 5.5 (100 µL) and incubated with 1 µL of  $\alpha(2\rightarrow3,6,8,9)$  sialidase solution (P0722, New England BioLabs) overnight at 37 °C. HEK293T lysate (8 mg) was first buffer exchanged with 50 mM sodium acetate pH 5.5 using 30 MWCO centrifugal concentrators (Vivaspin). The lysates, in a final volume of 4 mL, were then incubated with 20 µL of sialidase overnight at 37 °C.

### 4.3.4 Western blotting

Fetuin (100 µg) and HEK293T cell lysate (50 µg), treated or not with sialidase, were mixed with 2X Laemmli buffer (62.5 mM Tris-HCl pH 6.8, 1% LDS, 10% glycerol, and 0.005% bromophenol blue, Bio-Rad) supplemented with 50 mM (DTT) and separated on a 4-20% SDS-PAGE gradient gel (Bio-Rad) for 40 min at 200 V. The proteins on the gel were then transferred to a PVDF membrane (0.2 µm pore size) by wet transfer using 1 L Towbin buffer (25 mM Tris, 192 mM glycine, 10 % (v/v) methanol pH 8.3) at a constant voltage of 45 V for 1 h. The membrane was blocked in blocking buffer (50 mM Tris-HCl, 150 mM NaCl, 0.25% (w/v) BSA, and 0.05 % (v/v) NP-40, pH 7.4), made with BSA previously desialylated by oxidation with periodic acid (Glass II et al., 1981), for 30 min at room temperature, then incubated with biotin-conjugated *Maackia amurensis* lectin II (MAL B-1265, 3 µg/mL in blocking buffer, Vector Laboratories) overnight at 4 °C. The membrane was washed with 10 mL of blocking buffer three times for 10 min and incubated with anti-biotin antibody (GTX77581, 1:4000, GeneTex) for 1 h at room temperature. Proteins were detected by enhanced luminescence (GERPN2236 ECL kit from Sigma) using a G: BOX Chemi system (Syngene).

### 4.3.5 Bioorthogonal ligation with aminoxy-biotin

Labelled and control samples were prepared starting from 8 mg of protein, each obtained from the same lysate of HEK293T cells. Labelled sample was first buffer exchanged with 50 mM sodium acetate pH 5.5 using 30 kDa MWCO centrifugal concentrators (Vivaspin) to a final volume of 4 mL and then incubated with 20 µL of  $\alpha(2\rightarrow3,6,8,9)$  sialidase (New England BioLabs) overnight at 37 °C. The sample was then buffer exchanged with 1X PBS pH 7.4 using 30 kDa MWCO centrifugal concentrators and incubated for 0.5 h with sodium periodate (1 mM final concentration) in PBS (final volume 8 mL) at 4 °C using an end-to-end rotator in the dark. The sample was buffer exchanged with 1X PBS (pH 6.7) using 30 kDa MWCO centrifugal concentrators to a final volume

of 4 mL. Biotinylation was performed by adding aminoxy-biotin (Biotium) to the sample in 1X PBS pH 6.7 to a final concentration of 100  $\mu$ M in the presence of 10 mM aniline (final volume 6 mL). The reaction was carried out for 3 hours on an end-to-end rotator at 4°C. All buffer exchanges were performed by concentrating the sample to a volume of 2 mL, then adding 2 mL of fresh buffer and repeating the cycle three times.

Control sample was first buffer exchanged with 1X PBS pH 6.7 and then treated with aminoxy-biotin as above (without sialidase or sodium periodate treatment). Labelled and control samples were eventually washed thrice with PBS pH 6.7 using centrifugal concentrators to remove the excess biotin and were ready for streptavidin pulldown.

#### **4.3.6 Streptavidin pulldown**

Samples obtained after bioorthogonal ligation with aminoxy-biotin were incubated each with 100  $\mu$ L of streptavidin agarose slurry (Thermo) previously washed with 1 mL PBS pH 7.4 and incubated at room temperature using an end-over-end rotator for 0.5 h. After centrifugation at 500 x *g*, the supernatant was removed from the agarose and discarded, and the agarose was washed three times with 1 mL PBS pH 7.4. The agarose was then incubated with 0.1 mL 2X Laemmli buffer for 10 min at 100 °C to release bound proteins. The supernatant was collected for SDS-PAGE and in-gel digestion.

#### **4.3.7 In-gel digestion**

Samples obtained after streptavidin pulldown in 2X Laemmli buffer were supplemented with 50 mM dithiothreitol (DTT) and partially separated SDS-PAGE gradient gel. The protocol followed was then the same reported in section 2.3.2.

#### **4.3.8 Mass Spectrometric Analysis**

Samples obtained after in-gel digestion were analysed by LC-MS/MS as described in paragraph 2.3.5

#### **4.3.9 Data analysis**

Automated database searches were performed using the Byonic software (Protein Metrics Inc) using the *Homo sapiens* proteome (Uniprot) as the reference database. The search criteria were set as follows: KR (trypsin) cleavage sites;  $\leq 3$  missed cleavages and semi-specific cleavage allowed; error tolerances for the precursor of 3 ppm and 0.05 Da for fragment ions; carbamidomethyl cysteine as a fixed modification, oxidation of methionine (common 2), and pyro-Glu from

glutamine (rare 1) as variable modifications (max common set to 3 and max rare set to 2); possible glycan composition was set as Hex<sub>1</sub>, Hex<sub>2</sub>HexNAc<sub>1</sub> (core M1), Hex<sub>3</sub>HexNAc<sub>2</sub> (core M2), Hex<sub>1</sub>HexNAc<sub>2</sub>Phos<sub>1</sub> (core M3), Hex<sub>1</sub>HexNAc<sub>2</sub>Phos<sub>1</sub>Rbo5P<sub>1</sub>, Hex<sub>1</sub>HexNAc<sub>2</sub>Phos<sub>1</sub>Rbo5P<sub>2</sub>, Hex<sub>1</sub>HexNAc<sub>2</sub>Phos<sub>1</sub>Rbo5P<sub>2</sub>Xyl<sub>1</sub>GlcA<sub>1</sub>, Hex<sub>1</sub>HexNAc<sub>2</sub>Phos<sub>1</sub>GroP<sub>1</sub> (each set as rare 1, on Ser/Thr); false discovery rate (FDR) was set to 1%. Product ion spectra of putative peptide matches made by Byonic were interpreted manually to determine whether the match was accurate.

#### **4.3.10 Sequence alignment**

Vertebrata orthologues of the protein sequences for PTFRF, NRCAM, LAMP2 and RPN2 were searched using the online tool available at <https://www.ensembl.org/index.html>. The alignment file was downloaded in 'clustal' format and opened with the software SnapGene. Peptides containing the motif "TPT" from orthologues sequences were used to create the weblogo using the tool available at <https://weblogo.berkeley.edu/logo.cgi>.

## 4.4 Results

### 4.4.1 Bioinformatic search of published proteomic data

The study of glycoproteins represents a challenge because glycosylation is characterised by macroheterogeneity, which refers to the presence or absence of glycans at specific glycosylation sites, and microheterogeneity, which causes a specific glycosylation site on a protein to have different glycan structures on different copies of the same protein. Glycoprotein enrichment helps to reduce the complexity of the sample analysed and facilitates glycoprotein identification by mass spectrometry. In the case of glycoproteins modified with Rbo5P, which were the targets of this study, tools for enrichment selective for the alditol are not known. For this reason, the initial objective was to isolate the entire proteome of a cell and facilitate the detection of glycoproteins by reducing the complexity of the sample at the stage of the mass spectrometric analysis. In particular, the mass spectrometric analysis of glycopeptides can be preceded by multiple liquid chromatographic steps to allow a better separation of the analytes and increase the chances of retention, selection and fragmentation of the glycopeptides.

While a shotgun analysis would be a valid experiment to investigate a proteome in search of glycoproteins modified with Rbo5P, the analysis was not performed in this study. The reason is that shotgun analysis data sets have already been obtained on mammalian cell lysates and their raw mass spectrometric data sets are available on ProteomeXchange. These data sets may be interpreted using bioinformatic software, which gives as output a list of peptide matches and the name of the protein they originate from. The type of modifications of a peptide is one of the variables to input in the bioinformatic software and is under the user's control. Therefore, a published data set could not have been searched for a modification discovered after its publication, such as glycosylation with Rbo5P. This makes it interesting to re-analyse mass spectrometric data sets from before the discovery of Rbo5P in mammalian cells and to include in the search the newly discovered glycans with Rbo5P as possible modifications of the peptides. The data set used here for the re-analysis (Chick et al., 2015) was picked among others because its mass spectrometric analysis was performed following two steps of liquid chromatography and a high-resolution mass spectrometer (LTQ Orbitrap Elite, Thermo Scientific). In particular, proteins obtained from the mammalian cells (HEK293T) were fractionated using basic pH reversed-phase fractionation. Each fraction obtained in this way was acidified and the components separated on a C18 column. The raw mass spectrometric data contain the  $m/z$  measured for each peptide detected and its related product ion spectrum acquired with an Orbitrap analyser. Using this information, the data set was re-analysed for the purpose of this study, and a list of proteins was identified. Unfortunately, no peptide matches for

peptides modified with glycans containing Rbo5P were detected. O-mannosyl glycans, such as core M1 M2 and M3 found in  $\alpha$ DG, were also searched for but also yielded no matches.

The absence of product ion spectra consistent with O-mannosylated peptides in the examined data can be explained by comparing the methods that have been used in the literature to detect these species with the method used to generate the data examined here. O-mannosylated proteins in mammalian cells were identified in the literature using a method specifically designed to enrich peptides with O-linked Man (Vester-Christensen et al., 2013). The method exploited a stable gene knockout of the enzyme POMGnT1; this gene knockout prevents the elongation of O-linked Man into core M1 and M2 structures, so allowing mono O-mannosylated glycopeptides to be enriched using ConA, a lectin specific for terminal Man and Glc. The enrichment was followed by mass spectrometric analysis, which revealed the identity of  $\sim 50$  O-mannosylated proteins besides  $\alpha$ DG in mammalian cells. The strategy used in the work of Vester-Christensen was tailored to identify a specific class of O-glycoconjugates. On the other hand, the study used earlier for the re-analysis of mass spectrometric data (Chick et al., 2015) adopted an untargeted strategy since every protein, glycosylated or not, was obtained from cells and analysed. These observations suggest that a targeted analysis of the glycoproteins of interest is more effective than an untargeted one. O-mannosylated peptides might be less abundant and harder to detect compared to non-glycosylated peptides; therefore, they were identified when enriched but were not detected in more complex mixtures.

The absence of product ion spectra consistent with peptides modified with Rbo5P in the data examined here (Chick et al., 2015) can be explained by looking at how Rbo5P-containing peptides were identified in the literature. Among the studies that unveiled the presence of Rbo5P in the  $\alpha$ DG glycans by overexpressing the glycopeptide bearing core M3 O-glycans, one (Yagi et al., 2016) reported the existence of the glycoform bearing core M3 with a single unit of Rbo5P. This glycoform, cannot be detected even when peptides obtained from full-length  $\alpha$ DG are analysed (see Chapter 3) unless the protein is first oxidised with sodium periodate. This observation suggests that the specific glycoform could be even more difficult to detect in a mixture of proteins and glycoproteins. Rbo5P-containing-peptides were also identified with a tandem repeat of Rbo5P and are known to be further elongated with the polysaccharide matriglycan. The size of matriglycan considerably increases the mass of a glycopeptide, hindering its detection. The treatment of overexpressed  $\alpha$ DG glycopeptides bearing matriglycan with hydrofluoric acid (HF) (Kanagawa et al., 2016) allowed their detection by mass spectrometry because HF cleaves the phosphodiester linkages in Rbo5P that connect matriglycan to the core glycan. For this reason, it is not expected to detect peptides modified with Rbo5P extended with matriglycan under normal mass spectrometric

set up for peptide and glycopeptides analysis and this would explain why these glycoforms are not detected in the data re-analysed here.

The re-analysis of mass spectrometric data described here showed that O-mannosylated peptides and in particular peptides modified with Rbo5P could not be detected in the untargeted analysis of protein and glycoprotein mixtures from mammalian cells. The result suggests that a shotgun analysis might not be the best strategy to identify novel proteins modified with Rbo5P. For this reason, the idea of performing such an experiment for this study was abandoned. At the same time, it was clear that a strategy tailored for the enrichment of O-glycoconjugates bearing Rbo5P and their analysis would have higher chances of success.

#### **4.4.2 Tagging Rbo5P in cell lysates**

Identification of novel proteins modified with Rbo5P requires a method for enriching such proteins. No lectins are known to recognise ribitol and therefore a different approach must be used. The method described in Chapter 2 of this thesis can enrich by streptavidin pulldown proteins modified with NeuAc that have been previously oxidised and biotinylated. Like NeuAc, Rbo5P is susceptible to oxidation on treatment with sodium periodate (Praisman et al., 2016) and can so theoretically be conjugated with aminoxy-biotin as can NeuAc. This hypothesis was tested on  $\alpha$ DG and the experiments are described in Chapter 3. The data show a glycopeptide modified with Rbo5P in its unmodified state. While no oxidised or biotinylated Rbo5P was observed, the data suggest that the oxidation and labelling procedure was essential for the detection of (unmodified) RboP. There is a possibility that Rbo5P at a different location on the protein was oxidised or biotinylated, and thereby influenced the mass spectrometric analysis, but was not detected by mass spectrometry. Despite the absence of direct detection of biotinylated Rbo5P, there is no reason why the method tested in the previous Chapter can't be used on a complex sample such as a cell lysate. The goal here was to show that using this method,  $\alpha$ DG can be enriched from a cell lysate by exploiting the tagging of Rbo5P, and be subsequently identified by mass spectrometry.

$\alpha$ DG presents several sialylated glycopeptides that would be targeted by the method too. In order to show that the method works on Rbo5P, glycoproteins in the cell lysate must first be desialylated by enzymatic treatment. The efficacy of the sialidase treatment was tested on a standard sialylated protein, fetuin, and on cell lysates. Fetuin is heavily sialylated and the loss of NeuAc moieties from the protein after sialidase treatment results in a molecular mass shift that can be observed on SDS-PAGE, as it is described in the sialidase instruction manual. As expected, treatment of fetuin with

sialidase resulted in a noticeable downward shift of the protein on the gel (Figure 4.1-A). Sialidase treatment of HEK293T cell lysates resulted in a pattern of gel bands that was undistinguishable from that observed in the untreated sample, suggesting that sialidase treatment had no effect (Figure 4.1-A). It is possible, however, that sialylated proteins do not contribute significantly to the bands observed by Coomassie staining of whole cell lysates. Therefore, western blot analysis of the same samples was performed using a lectin specific to NeuAc to enable direct detection of sialylated proteins, as shown in Figure 4.1-B. A staggering difference in the level of NeuAc staining is observed for fetuin treated and untreated with sialidase, suggesting that the enzyme is effective. Some staining is still observed for fetuin treated with sialidase, which can be attributed to incomplete desialylation of the sample. A difference in staining for HEK293T cell lysates treated with sialidase versus the non-treated control is also observed. While some bands are detected in the untreated sample, no staining is observed in the desialylating sample. The results suggest that sialidase treatment is effective in desialylating pure proteins or complex mixtures of proteins such as those found in a cell lysate. The treatment does not necessarily go to completion as judged from the fetuin sample in Figure 4.1-B. Therefore, in all following experiments, some sialylated proteins can still be expected to remain after sialidase treatment.

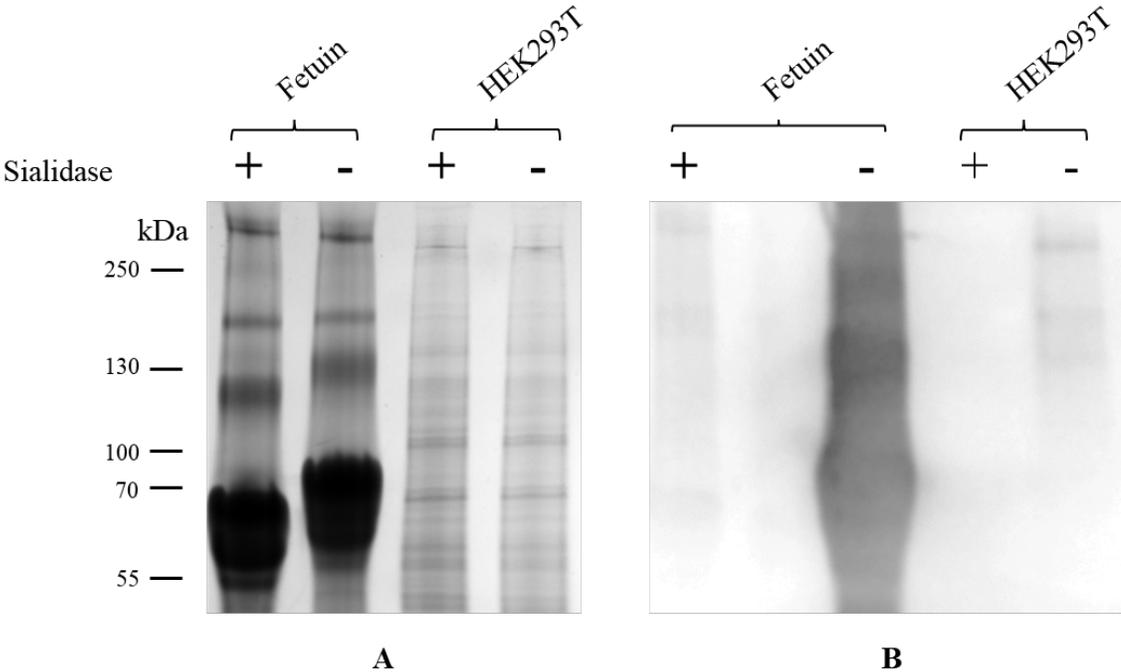


Figure 4.1 (A) Coomassie-stained SDS-PAGE gel obtained after loading fetuin or HEK293T cell lysate treated with sialidase or left untreated. (B) Western blot analysis of the same samples as in

panel A, detected using biotinylated *Maackia amurensis* lectin, and anti-biotin as secondary antibody.

Cell lysates from HEK293T cells were next subjected to the previously developed oxidation and aminoxy-labelling procedure with the purpose of biotinylating any proteins that bear Rbo5P. The workflow and samples prepared are shown in Figure 4.2. The “labelled” sample consists of a cell lysate first treated with sialidase to remove NeuAc residues, then oxidised and biotinylated to tag Rbo5P-bearing glycoproteins, and eventually incubated with streptavidin gel in order to recover biotinylated proteins so enriching for Rbo5P-bearing species. This sample is predicted to contain  $\alpha$ DG and other proteins modified with Rbo5P. It must be considered that several proteins might bind streptavidin non-specifically and be recovered with the proteins that are biotinylated, both naturally and artificially. To account for these proteins, a “control” sample was prepared and consisted of a cell lysate treated with aminoxy-biotin but not first treated with sialidase or sodium periodate.

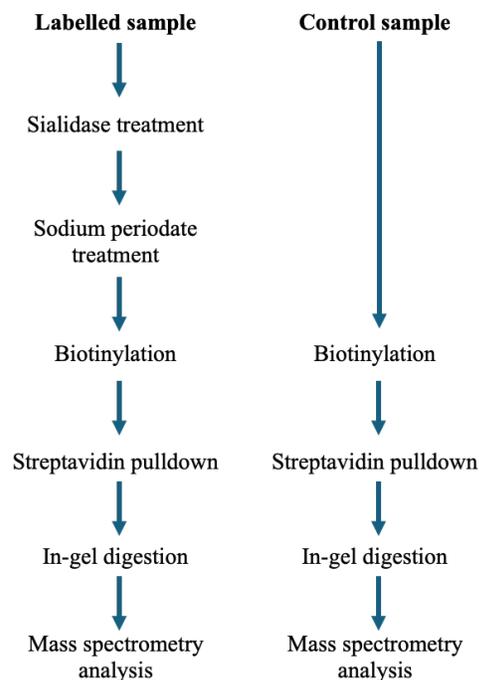


Figure 4.2 / Schematic representation of the workflow for selectively tagging and enriching proteins modified with Rbo5P from HEK293T cell lysates. Labelled and control samples are proteins from lysates of HEK293T cells, which were treated with or without sialidase and then oxidised or not

*with sodium periodate. All samples were treated with aminoxy-biotin to biotinylate susceptible structures and then incubated with streptavidin gel to recover biotinylated proteins. Samples enriched this way were partially separated on an SDS-PAGE gel which was stained with Coomassie. Peptides obtained by in-gel digestion were analysed by mass spectrometry.*

Proteins recovered after streptavidin pulldown with either the labelled or the control samples were digested into peptides and analysed by LC-MS/MS. Peptides were isolated and fragmented in the mass spectrometer to obtain product ion spectra which were interpreted with the help of bioinformatic software. Proteins identified following peptide matching of product ion mass spectra of the labelled and control samples were then inspected.  $\alpha$ DG was identified on matching of seven non-glycosylated peptides from the labelled sample, while it was not identified in the control sample. The result suggests that  $\alpha$ DG does not bind streptavidin non-specifically unless biotinylated. After treatment with sodium periodate and aminoxy-biotin  $\alpha$ DG could be biotinylated at the level of NeuAc and Rbo5P moieties., However, this is not what was observed in the experiments reported in Chapter 3 where the treatment of pure  $\alpha$ DG was described. In particular, glycopeptides modified with NeuAc and Rbo5P from  $\alpha$ DG were identified in their native state and were not identified in either oxidised or biotinylated form. Despite these earlier observations, the data presented here show that  $\alpha$ DG was in fact enriched after oxidation and aminoxy-biotin treatment, and therefore suggest that biotinylation has been successful.

The fact that  $\alpha$ DG was detected in the labelled sample suggests that  $\alpha$ DG is indeed oxidised and biotinylated at the level of Rbo5P since it is expected that most of the NeuAc has been removed by sialidase treatment. This claim, at first glance, is not consistent with the data shown in Chapter 3, where Rbo5P was identified as unoxidised and not biotinylated despite treatment of the sample with sodium periodate and aminoxy-biotin. One possible explanation is that one or more of the other  $\alpha$ DG glycosylation sites that bear Rbo5P are oxidised and biotinylated but not detected. Considering the fact that streptavidin pulldown was performed before the digestion of the protein, a single modification event would have been sufficient to result in the effective enrichment of the glycoprotein. It seems feasible that the modified (glyco)peptide itself has escaped detection. In other words, not detecting glycopeptides bearing biotinylated Rbo5P does not suggest that the method is not useful. Instead, evidence that the method is effective is provided by the detection of  $\alpha$ DG in desialylated cell lysates, in which there is a reasonable chance that enrichment has resulted from the modification of Rbo5P.

It is important to highlight the fact that, while  $\alpha$ DG was only identified in the labelled sample, several other proteins were identified in both the labelled and control samples. In theory, the protein list from the labelled sample contains a mixture of contaminant proteins (binding streptavidin non-specifically) and proteins modified with Rbo5P. The focus can then be moved to any other proteins identified in the labelled sample as they could be novel candidates for also carrying the Rbo5P modification. The protein list obtained from the mass spectrometric analysis of the labelled sample contains about 700 proteins. None of these proteins was directly observed via a peptide modified with Rbo5P. The heterogeneity of glycosylation and the abundance of each glycoform add extra complexity to the sample and explain the low likelihood of detecting a specific glycoform of any particular glycopeptide. About 500 of the 700 proteins were also identified in the control sample. The control sample is expected to recover proteins non-specifically binding streptavidin or proteins naturally biotinylated. Therefore, it would be tempting to strike from the labelled sample list all the proteins found in the control sample list. Nevertheless, it is not possible to exclude that a candidate for glycosylation with Rbo5P might be found in both labelled and control samples but with higher intensity in the former. Since the experiment was performed with a single biological replicate, directly comparing protein intensities between the two samples would not be statistically relevant. For these reasons, the 700 proteins identified in the labelled sample were all considered for the following data interpretation.

Finding candidates worth follow-on studies among those proteins is complicated, given the many targets available. Therefore, different criteria were used to extract a smaller subset of targets from the protein list. The criteria picked consider the glycosylation state of proteins modified with Rbo5P. As described in Chapter 1, Rbo5P has been found on O-mannosyl glycans, a type of glycosylation that takes its first step in the ER. Proteins carrying Rbo5P, other than  $\alpha$ DG, are likely to be also O-mannosylated, therefore they would need to translocate into the ER lumen. A signal peptide is required for a protein to enter the ER lumen (Blobel and Dobberstein, 1975a, 1975b). Hence, the proteins identified in the labelled sample can be sorted based on the presence of a signal peptide in their sequence. It is not surprising that all the sequences of known O-mannosylated proteins, reported in Table 1.1, are characterised by a signal peptide. The majority of O-mannosylated proteins, like cadherins, protocadherins, and plexins, are membrane proteins; therefore, they have transmembrane domains. To assess whether all O-mannosylated proteins known were characterised by a transmembrane domain, the sequences of proteins reported in Table 1.1 were submitted to the bioinformatic service DeepTMHMM for the prediction of transmembrane domains (Hallgren et al., 2022). Interestingly, 73 out of 76 proteins were found to have at least one transmembrane domain. Three proteins, protein disulfide isomerase, protein disulfide isomerase A3,

and cadherin 13, were an exception as they did not present a transmembrane domain. The result suggested that O-mannosylated proteins bearing Rbo5P will likely have a transmembrane domain. For this reason, the sequences of proteins identified in the labelled sample were first analysed with the bioinformatic service SignalP-6.0 for the prediction of signal peptides (Teufel et al., 2022) and then searched with DeepTMHMM for the prediction of transmembrane domains. The proteins identified with at least two peptides in the labelled sample having both a signal peptide and at least a transmembrane domain are reported in Table 4.1

Protein name	Uniprot ID	Gene	Peptides identified
Receptor-type tyrosine-protein phosphatase F	P10586	PTPRF	22
Basigin	P35613	BASI	22
Integrin beta-1	P05556	ITBI	14
Cation-independent mannose-6-phosphate receptor	P11717	MPRI	13
Prostaglandin F2 receptor negative regulator	Q9P2B2	FPRP	10
Dystroglycan	Q14118	DAG1	10
Calnexin	P27824	CALX	8
Nicastrin	Q92542	NICA	7
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	P04844	RPN2	7
Lysosome-associated membrane glycoprotein 2	P13473	LAMP2	5
Neuronal cell adhesion molecule	Q92823	NRCAM	4
Sortilin	Q99523	SORT	4
Lysosomal acid phosphatase	P11117	PPAL	4
CD166 antigen	Q13740	CD166	4
CD276 antigen	Q5ZPR3	CD276	4
Integrin alpha-2	P17301	ITA2	4
Lysosome-associated membrane glycoprotein 1	P11279	LAMP1	4
Transmembrane emp24 domain-containing protein 7	Q9Y3B3	TMED7	4
Neuroplastin	Q9Y639	NPTN	3
Transmembrane protein 109	Q9BVC6	TM109	3
Cation-dependent mannose-6-phosphate receptor	P20645	MPRD	3
Cell adhesion molecule 1	Q9BY67	CADM1	3
Junctional adhesion molecule A	Q9Y624	JAM1	2
Myelin protein zero-like protein 1	O95297	MPZL1	2
Leukocyte surface antigen CD47	Q08722	CD47	2
Membrane cofactor protein	P15529	MCP	2
Inactive tyrosine-protein kinase 7	Q13308	PTK7	2
Endothelial protein C receptor	Q9UNN8	EPCR	2

*Table 4.1 /List of putative proteins modified with Rbo5P. The list was obtained after mass spectrometric analysis of peptides obtained from the digestion of proteins from HEK293T cell lysates treated with sialidase, then sodium periodate, aminoxy-biotin, and enriched by streptavidin pulldown. Proteins all present a signal peptide and at least a transmembrane domain in their sequence.*

Other factors can be considered to discriminate which proteins in Table 4.1 will most likely be modified with Rbo5P. The first factor is the consideration that Rbo5P has (thus far) been identified exclusively on core M3 glycans. The enzyme POMGnT2, which adds a  $\beta$ -1,4 GlcNAc to the O-

linked mannose, is necessary for the synthesis of core M3 glycans. POMGnT2 was recently found to recognise a specific amino acid motif in  $\alpha$ DG, the “TPT” motif (Imae et al., 2021).  $\alpha$ DG has two “TPT” motifs modified with core M3 glycans, which are conserved among orthologues vertebrate genes. POMGnT2 also recognises in the substrate an “RXR” motif at the N-terminal side of the O-mannosylated residue, but the absence of such motif only affects the activity of the enzyme moderately, therefore, it is not considered essential (Halmo et al., 2017). On the other hand, the amino acids at the C-terminal side of the O-mannosylated residue contribute to binding POMGnT2 and their truncation results in a complete loss of activity (Imae et al., 2021). These amino acids are not conserved among vertebrate orthologues of  $\alpha$ DG; hence, it is not possible to define another motif essential for POMGnT2 activity except for the “TPT” motif. Analysis of the amino acid sequences of the proteins in Table 4.1 revealed that four of these proteins have a “TPT” motif, excluding  $\alpha$ DG. These proteins are receptor-type tyrosine-protein phosphatase F (PTPRF), dolichyl-diphosphooligosaccharide-protein glycosyltransferase subunit 2 (RPN2), lysosome-associated membrane glycoprotein 2 (LAMP2), and neuronal cell adhesion molecule (NRCAM).

It is possible that finding four proteins with the same triple amino acid motif as  $\alpha$ DG could be a coincidence. To better understand this result's significance, the human proteome was searched for the “TPT” sequence to assess how common this motif is. The tool “MOTIF”, available on <https://www.genome.jp>, allows the generation of a list of annotated human proteins that present the motif “TPT” in their sequence. The result showed that ~2000 out of ~20000 proteins searched have this same motif (data not shown). Considering that ~10% of human proteins have the “TPT” motif, finding four of these proteins in the list shown in Table 4.1 is what would be expected. Nevertheless, to further explore the possible significance of the “TPT” motif, a sequence alignment was performed to determine if these amino acids were conserved among orthologues of the proteins. While the “TPT” motif is essential for  $\alpha$ DG function and is therefore conserved it might not be conserved in other proteins if it is not essential. To determine if the motif is conserved across species, sequences of the proteins PTPRF, RPN2, LAMP2, and NRCAM were compared to the sequences of their respective vertebrate orthologues. The tool available on <https://www.ensembl.org/index.html> was used to compare vertebrate orthologues of a specific protein. The result is reported using the tool available on <https://weblogo.berkeley.edu/logo.cgi> and is shown in Figure 4.3.

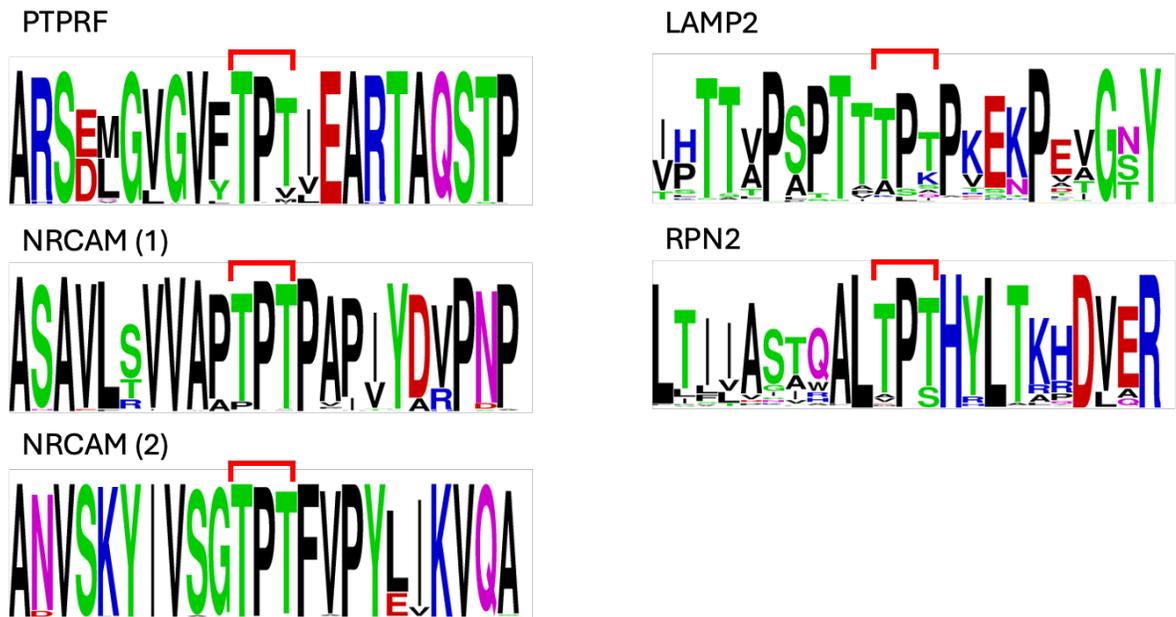


Figure 4.3 /Alignment results of amino acid sequences from protein receptor-type tyrosine-protein phosphatase F (PTPRF), neuronal cell adhesion molecule (NRCAM), dolichyl-diphosphooligosaccharide-protein glycosyltransferase subunit 2 (RPN2), and lysosome-associated membrane glycoprotein 2 (LAMP2). Vertebrate orthologues were used for the alignment, and the results are shown with WebLogo. The amino acid motif “TPT” is highlighted with a red bar in all sequences. NRCAM has two “TPT” motifs (labelled 1 and 2).

The image shows how common the presence of particular amino acids is in each position when comparing orthologous sequences. The larger the character of a specific amino acid is in a particular position the more that amino acid is conserved among species. In Figure 4.3, the “TPT” amino acid sequence is well conserved for proteins PTPRF, RPN2, and NRCAM, while it seems less conserved for LAMP2. The observation is encouraging but is not proof of the presence of Rbo5P or core M3 glycan on the proteins. Another detail to consider is that among the four proteins, only PTPRF and NRCAM were not identified in the control sample. The experiment performed here would need to be repeated with more biological replicates in order to show that these proteins are exclusively recovered in the labelled sample, which would give further proof of the proteins being Rbo5P-containing glycoconjugates.

An angle to further understand which of the four proteins might be of interest for follow-up experiments is their function. In fact, novel proteins modified with Rbo5P might share a similar function to  $\alpha$ DG and use their glycan to bind laminin in the extracellular matrix. NRCAM mediates cell-cell interaction in the brain and regulates neuronal projection (Hortsch, 1996). Although the binding of NRCAM with laminin is reported, the function of this molecular interaction is not clear (Grumet and Sakurai, 1996). RPN2 is part of the oligosaccharyltransferase complex responsible for the transfer of the *N*-glycan precursor from the Dol-P donor onto a protein (Kelleher et al., 1992). RPN2 can be excluded from the list of potential Rbo5P-containing proteins since it is an integral membrane protein of the ER (Crimaudo et al., 1987) and could not be glycosylated in the Golgi. LAMP2 is a membrane protein located in the lysosomal compartment, acting as a receptor for the degradation of proteins in the lysosome. Lastly, PTPRF is part of the protein tyrosine phosphatase subfamily and is involved in the regulation of the phosphorylation state of multiple proteins (Sarhan et al., 2016). PTPRF is also involved in the cell-extracellular matrix (ECM) interaction to promote neurite outgrowth (Chien and Ryu, 2013). From the functional point of view, PTPRF and NRCAM should be highlighted, given their role in the ECM, such as  $\alpha$ DG. It is interesting to speculate that PTPRF and NRCAM, in particular, could be recognised by POMGnT2, have a core M3 glycan and be modified with Rbo5P.

## 4.5 Conclusions

The work presented in this chapter aimed to identify novel human proteins modified with Rbo5P. Given the absence of a known method for selective enrichment of glycoproteins showing Rbo5P on their glycans, searching for such proteins started from the proteomic analysis of human cell lysates without prior enrichment. The mass spectrometric data that were re-analysed to include novel glycans found in mammalian cells did not identify any glycopeptides modified with O-mannosyl glycans or glycans with Rbo5P. The limits of glycoproteomics analysis and the challenges presented by the study of glycoproteins likely collide to make the identification of a specific glycopeptide with a particular glycan very difficult. A method for enriching proteins modified with Rbo5P, previously tested on standard proteins, was then used to reduce the complexity of the cell lysate sample and facilitate the analysis of the glycoproteins. A list of putative proteins likely to be modified with the alditol was obtained. Although glycopeptides with Rbo5P were not detected, the method's efficacy was validated by identifying  $\alpha$ DG among the proteins detected. Two proteins from the list have been identified as candidates for further studies. These proteins present features that are common to O-mannosylated proteins and peptides that are likely to be modified with a core M3 O-mannosyl glycan, which is also the scaffold for Rbo5P modification. Overexpression of these proteins will be necessary to show their glycosylation state.

## Chapter 5: Discussion and future research

Glycosylation is a common protein modification and is central to many biological processes. Aberrant glycosylation, often a result of changes to the activity or localisation of the enzymes responsible for glycan biosynthesis, may lead to non-functional glycoproteins and is the cause of many diseases. Understanding the complexity that characterises this post-translational modification is a challenge and requires the development and application of *ad hoc* strategies. Glycoproteomics deals with the problem of analytically studying glycoproteins, their glycosylation state and how this changes, for example, between healthy and diseased or otherwise compromised individuals. A peculiar type of glycans, Rbo5P-containing O-linked glycans, and the glycoproteins bearing them were the main focus of the research reported in this Thesis. Glycoproteins modified with Rbo5P are currently rare among glycoprotein types in mammalian cells, with the only example reported to date being  $\alpha$ DG (Kanagawa et al., 2016; Praissman et al., 2016; Yagi et al., 2016).  $\alpha$ DG is modified with Rbo5P-containing glycans which are responsible for its function of providing muscular rigidity to tissues (Nickolls and Bönnemann, 2018). The lack of functional enzymes or their mislocalisation are responsible for the Rbo5P modification in mammals and is one of the causes of a series of pathologies, the dystroglycanopathies, which present phenotypes ranging from muscular dystrophies to brain malformations. The Rbo5P modification has a dedicated biosynthetic pathway, which could also be expected to modify proteins other than  $\alpha$ DG, although other examples of glycoproteins modified by this pathway are currently unknown. The idea that other Rbo5P-modified proteins may exist is the force driving the work presented here, which had the overarching aim of identifying novel proteins modified with Rbo5P in mammalian cells.

### 5.1 Rbo5P-bearing glycopeptides are not identified by untargeted strategies

The research approach used here was intended to take the first step towards identifying novel Rbo5P-containing glycoconjugates by using an untargeted mass spectrometric (shotgun) analysis of proteins isolated from cell lysates. Data obtained by shotgun analysis of a mammalian cell proteome was already available in a public data repository and was re-analysed (Chapter 4) in light of the novel alditol modification discovered in mammals (Kanagawa et al., 2016; Praissman et al., 2016; Yagi et al., 2016). The data interpretation using bioinformatic software failed to identify detectable glycopeptides modified with Rbo5P. The result was not surprising, considering that even the analysis of pure  $\alpha$ DG did not lead to identifying Rbo5P (Harrison et al., 2012; Stalnaker et al.,

2010). The challenge still exists for the analysis of glycopeptides modified with Rbo5P since the alditol is known to be modified by the matriglycan (Yoshida-Moriguchi and Campbell, 2014), which is a large and, therefore, highly heterogeneous oligo- or polysaccharide; its presence would mean that not all Rbo5P residues are terminal/unextended. Unmodified  $\alpha$ DG peptides were detected in the shotgun analysis but not in high abundance, which would make the detection of a heterogeneous mixture of variably modified  $\alpha$ DG peptides even more difficult. Moreover, the experiment was carried out in the positive ion mode, so the phosphate-bearing modification, which would preferentially ionise as an anion, is likely to further compromise the mass spectrometric response in the positive mode. Further heterogeneity is introduced in the positive mode analysis of phosphoglycans by the tendency to form a range of sodium salts, compromising the ability to detect a single species. On the other hand, targeted strategies have been successfully used to identify Rbo5P-containing peptides from  $\alpha$ DG. In particular, the overexpression of glycopeptides modified with Rbo5P in mammalian cells allowed the characterisation of the novel alditol by mass spectrometry (Kanagawa et al., 2016; Praissman et al., 2016; Yagi et al., 2016). A new targeted strategy was needed for the identification of novel proteins modified with Rbo5P.

## **5.2 Chemical tagging of glycoproteins allows their selective enrichment and mass spectrometric identification.**

The novelty of the study presented here was to apply a method developed for tagging sialylated glycoconjugates (Zeng et al., 2009) to proteins with Rbo5P-containing glycans. The strategy involved the oxidation of NeuAc to generate an aldehyde, which then acts as a chemical reporter that can be conjugated with an enrichment tag. It is known that Rbo5P is susceptible to oxidation with sodium periodate (Praissman et al., 2016) as the method was used to remove matriglycan from the Rbo5P-bearing glycopeptide, which considerably increases its mass. Nevertheless, generating an aldehyde on Rbo5P was not exploited in the literature for tagging the alditol with a probe that could be used for enrichment.

The chemistry used in this work was already established, particularly the conditions for the reactions that involved the oxidation and the biotinylation of sialylated proteins (Bank et al., 2003; Ramya et al., 2013; Zeng et al., 2009). Still, testing the method showed the importance of a key factor in the success of the oxidation reaction. In fact, sodium periodate's freshness affects the oxidation efficiency. In Chapter 2, it was shown that in the same conditions, old reagent leads to mixture of oxidised and non-oxidised NeuAc on fetuin, with the oxidation products being mixed as aldehydes were generated on two different carbons of NeuAc. Using fresh sodium periodate instead

leads to the complete oxidation of sialylated peptides uniquely on the carbon 7 of NeuAc. The mass spectrometric analysis of oxidised and biotinylated glycopeptides bearing NeuAc presented a new challenge due to the need to identify the oxidised and biotinylated peptides. Studies that used tagging of proteins with biotin usually do not have the primary interest of determining the site of biotinylation, since identifying the proteins is the goal (Matta et al., 2019; Ramya et al., 2013). In the study presented here, biotin used to tag Rbo5P not only would mark the protein bearing the alditol but also the glycan modified with Rbo5P. There is an interest in knowing the identity of such glycan since it could be a core M3 glycan, as observed in  $\alpha$ DG, or a completely new glycan of unknown composition. For these reasons, biotinylated glycopeptides were recovered and analysed by mass spectrometry. An important step taken in this study was to identify a marker for biotinylated peptides. While examples of product ion spectra of biotinylated peptides are present in the literature (Nishino et al., 2022), the way biotin fragments in the collision cell of the mass spectrometer depends on the biotin tag used, and the linkage in which it is involved with the peptide. Aminoxy-biotin forms an oxime bond with the oxidised glycopeptide, regardless of the presence of NeuAc or Rbo5P. In Chapter 2, the specific marker for biotinylation with aminoxy-biotin was identified as the peak at  $m/z$  387.21. This result required extensive manual searching and interpretation of mass spectrometric data. The marker could then be used in Chapter 4 to manually search the data of cell lysates oxidised and biotinylated. Moreover, a peak for biotinylated and oxidised NeuAc was detected in the product ion spectra of the biotinylated sialylated peptides. The result suggested that an analogous marker ion characteristic of oxidised and biotinylated Rbo5P might be detected in the product ion spectrum of Rbo5P-containing peptides that had been oxidised and biotinylated in the same way. Therefore, such a marker ion could be used to interrogate the mass spectrometric data and facilitate the interpretation of the spectra obtained from Rbo5P-modified glycoproteins such as  $\alpha$ DG. Testing the chemical tagging formed the basis for the subsequent experiments and proved the efficacy of the method used.

### **5.3 Rbo5P is identified from full-length $\alpha$ DG**

The interest in  $\alpha$ DG glycosylation led to the expression of the recombinant protein (Harrison et al., 2012; Stalnakar et al., 2010) or the glycopeptide modified with matriglycan (Kanagawa et al., 2016; Praissman et al., 2016; Yagi et al., 2016), necessary for  $\alpha$ DG function in the ECM of skeletal muscle and brain (Goddeeris et al., 2013; Yoshida-Moriguchi and Campbell, 2014). The method for the enrichment of proteins modified with Rbo5P could have been tested on a recombinant glycopeptide from  $\alpha$ DG. Nevertheless, the method would eventually be used on intact proteins (Chapter 4) so

there was a reason to express a full-length  $\alpha$ DG instead of a single peptide. Having the ability to produce recombinant  $\alpha$ DG was also important to further characterise the glycosylation of the protein. Different studies have reported glycopeptide identified from the digestion into peptides of  $\alpha$ DG, but they were unable to identify peptides bearing Rbo5P or GroP (Harrison et al., 2012; Stalnaker et al., 2010), which so far have been characterised uniquely from recombinant peptides of  $\alpha$ DG. The expression of  $\alpha$ DG, described in Chapter 3, presented several challenges. The main issue was related to the inability to detect the FLAG-tag on the protein when the tag was designed to be at the C-terminus of the protein. In those conditions, glycosylated  $\alpha$ DG could be expressed and detected in the media of HEK293T transfected cells but the presence of the FLAG-tag on the recombinant proteins could not be demonstrated. It was not possible to prove that the auto-proteolysis of the recombinant protein was responsible for the loss of the FLAG-tag. This is in line with the fact that the region downstream of the auto-proteolytic site (S654), absent in the recombinant protein, is likely necessary for the auto-proteolytic event. This region does not have conserved residues among orthologue sequences but has a conserved secondary structure (Akhavan et al., 2008), which suggests its necessity to properly arrange the polypeptide DAG1 and allow autoproteolysis.

The successful expression of glycosylated  $\alpha$ DG with a FLAG-tag at the N-terminus allowed the characterization of  $\alpha$ DG glycosylation state. Characterising glycopeptides from  $\alpha$ DG represents a challenge because many glycans are found in a protein region with few tryptic sites, which makes difficult to generate peptides of adequate size for mass spectrometry analysis. This problem is common for mucin-like proteins (Rangel-Angarita and Malaker, 2021). Compared to studies where peptides from recombinant  $\alpha$ DG were characterised by mass spectrometry, the results presented in Chapter 3 showed some overlap. In particular, many of the known O-glycosylated peptides with either O-GalNac or O-Man were identified. While oxidising and biotinylating recombinant  $\alpha$ DG did not lead to the identification of biotinylated peptides, it increased the number of glycopeptides identified. The absence of detectable glycopeptides affected by oxidation and biotinylation does not imply their absence in the sample but instead suggests that these species might be more difficult to detect. The data interestingly revealed an unmodified Rbo5P-containing  $\alpha$ DG glycopeptide, which had not previously been reported as being modified with the alditol. The discovery was the result of extensive manual data search as bioinformatic software was unable to identify the glycopeptide bearing Rbo5P. The finding is interesting because, while  $\alpha$ DG is known to be modified with glycans bearing Rbo5P, only a single glycosylation site had been proven to bear such glycans. The result presented here suggests that more than one Rbo5P glycosylation site may exist on  $\alpha$ DG. It is possible that only one site was tagged with biotin while a second remained unmodified and was the

one detected mass spectrometrically. For the first time, the presence of Rbo5P on  $\alpha$ DG was shown starting from the analysis of the full-length recombinant protein instead of a truncated peptide known to bear matriglycan-extended core M3 glycans. Overall, the data provide new insight into the glycosylation of  $\alpha$ DG while not excluding that the approach of ribitol oxidation and biotinylation may be a viable method for enriching and detecting endogenous glycoproteins bearing Rbo5P.

#### **5.4 Putative Rbo5P-containing proteins are identified in cell lysate after oxidation and biotinylation**

$\alpha$ DG is not an abundant protein, as the results from the shotgun analyses showed (Chapter 4), and targeted strategies have been used in the literature to recover endogenous  $\alpha$ DG from cells or tissues (Kanagawa et al., 2009; Yamada et al., 1996). The strategies are characterised by pulldown protocols involving the incubation of lysates first with the lectin WGA, then laminin, a known interactor of  $\alpha$ DG. The results presented in Chapter 4 show that endogenous  $\alpha$ DG can be enriched from cell lysates using oxidation, biotinylation and streptavidin pulldown in the absence of NeuAc. The absence of  $\alpha$ DG in the control sample also shows that the protein cannot be easily recovered by streptavidin via non-specific binding. Because the only other monosaccharide that should be tagged under the conditions used is Rbo5P, the result also suggests that  $\alpha$ DG is biotinylated at the level of Rbo5P. Peptides bearing Rbo5P were, however, not identified directly in this experiment, and the same can be said for the detection of marker ions for biotinylated and oxidised Rbo5P or biotin itself. Nevertheless, the identification of  $\alpha$ DG validates the experiment for consideration of the other proteins identified together with  $\alpha$ DG following enrichment. The list of proteins to further investigate can be restricted to proteins that present three features: a signal peptide, a transmembrane domain and a TPT motif. The justification of the three conditions is based on the assumption that proteins modified with Rbo5P are also O-mannosylated, are modified with core M3 glycans, which happens on TPT motifs, and might be membrane proteins as the majority of known O-mannosylated proteins. Among the candidates that present the three features, PTPRF and NRCAM are worth further investigation.

## 5.5 Candidates for follow-up studies and future recommendations

PTPRF is part of the protein tyrosine phosphatase subfamily and is involved in the regulation of the phosphorylation state of multiple proteins (Sarhan et al., 2016). It is characterised by the intracellular subunit-P, which holds protein tyrosine activity, anchored to the cell membrane by a transmembrane peptide. Extracellularly, PTPRF presents the subunit-E, linked non-covalently to subunit-P, consisting of Ig-like domains and extracellular fibronectin type III (FNIII) domains both critical for ligand binding (Won et al., 2017; Woo et al., 2010). Interestingly, PTPRF is involved in the cell-extracellular matrix (ECM) interaction to promote neurite outgrowth (Chien and Ryu, 2013).  $\alpha$ DG also mediates the cell-ECM adhesion, but while it exploits this function by binding laminin, PTPRF binds chondroitin sulfate proteoglycans and heparan sulfate proteoglycans in the ECM. NRCAM is a neuronal cell adhesion molecule characterised by Ig-like and FNIII domains (Hortsch, 2000). It is found in neurons of both developing and adult nervous systems (Backer et al., 2002) and is involved in multiple processes in the nervous system, such as cell-cell adhesion, cell migration and neurite outgrowth (Hortsch, 1996). NRCAM has been shown to interact with proteins in the ECM (Grumet and Sakurai, 1996), such as neurocan, phosphacan, and laminin which is also a known interactor of  $\alpha$ DG (Yamada et al., 1996). Interestingly, PTPRF is also an interactor of NRCAM in the brain (Sakurai et al., 1997).

For the further investigation of the two putative Rbo5P-containing proteins, peptides from PTPRF and NRCAM bearing the TPT motif could be recombinantly expressed and characterised by mass spectrometry to show their glycosylation state. While the presence of Rbo5P is expected on core M3 glycans, it cannot be excluded that the alditol might be part of glycans with unknown composition. Therefore, the expression of full-length PTPRF and NRCAM is also advised. In case of discovering an alditol modifying the proteins, monosaccharide analysis of the glycans released from the proteins, followed by derivatisation and GC-MS analysis, can be used to identify whether ribitol is the alditol released from the purified proteins.

The results obtained in Chapter 4 offer many ideas for follow-up experiments. Applying the oxidation/biotinylation method to other cell lines or tissue extracts would be important to see if the results are reproducible or if additional putative proteins modified with Rbo5P can be identified. A subset of proteins instead of the whole cell lysate or tissue extract could be analysed to enhance the likelihood of detecting glycoforms with potentially low abundance. Therefore, an approach worth trying would be applying the method used in this study to proteins present on the cell surface. This experiment would involve the use of either purified cell membranes or intact cells as samples to be oxidised and biotinylated and eventually lysed and analysed by mass spectrometry.

The research reported in this Thesis offers new insights into  $\alpha$ DG glycosylation and a new approach to the analysis of Rbo5P-containing glycoconjugates. Applying this approach to endogenous protein mixtures opens the possibility of discovering novel proteins modified with Rbo5P, which would improve the knowledge of this rare modification in mammals.

## List of abbreviations

AAL	<i>Aleuria aurantia</i> lectin
AMBIC	Ammonium bicarbonate
B3GALNT2	$\beta$ 1,3- <i>N</i> -acetylgalactosaminyltransferase 2
B4GALT	$\beta$ -1,4-galactosyltransferase
B4GAT1	$\beta$ 1,4-glucuronyltransferase 1
CAN	Acetonitrile
CDP-Rbo	CDP-ribitol
CHO	Chinese hamster ovarian
CID	Collision induced dissociation
ConA	Concanavalin A
CRM	Charged residue model
DDA	Data-dependent acquisition
DGC	Dystrophin-glycoprotein complex
DMEM	Dulbecco's modified Eagle's medium
Dol-P	Dolichol-phosphate
DTT	Dithiothreitol
ECM	Extracellular matrix
EGF	Epidermal growth factor
EIC	Extracted ion chromatograms
ER	Endoplasmic reticulum
ESI	Electrospray ionization
ETD	Electron transfer dissociation
FBS	Fetal bovine serum

Fc	Fragment crystallizable
FCMD	Fukuyama-type congenital muscular dystrophy
FDR	False discovery rate
FITC	Fluorescein-5-isothiocyanate
Fuc	Fucose
FKRP	Fukutin-related protein
FKTN	Fukutin
GAG	Glycosaminoglycans
Gal	Galactose
GH	Glycosidase
Glc	Glucose
GlcA	Glucuronic acid
GlcNAc	<i>N</i> -acetylglucosamine
GOase	Galactose oxidase
GroP	Glycerol phosphate
GT	Glycosyltransferase
HCD	Higher energy collision-induced dissociation
HexNAc	<i>N</i> -acetylhexosamine
HF	Hydrofluoric acid
IAM	Iodoacetamide
IEM	Ion evaporation model
ISPD	Isoprenoid synthase domain-containing protein
LAMP2	Lysosome-associated membrane glycoprotein 2
LARGE	Like-acetylglucosaminyltransferase
LC	Liquid chromatography

LC-MS/MS	Liquid chromatography coupled to mass spectrometry
LDS	Lithiumdodecyl sulphate
LGMD	Limb-girdle muscular dystrophy
Man	Mannose
MCD1D	Congenital muscular dystrophy type 1D
ME1	NADP-dependent malic enzyme
MEB	Muscle-eye-brain disease
MGAT5B	$\beta$ -1,6-N-acetylglucosaminyltransferase
MOE	Metabolic oligosaccharide engineering
MS	Mass spectrometry
MW	Molecular weight
MWCO	Molecular weight cut-off
NeuAc	Neuraminic acid
NRCAM	Neuronal cell adhesion molecule
OGT	O-GlcNAc transferase
PBS	Phosphate-buffered saline
PMT	Protein O-mannosyltransferase (Yeast)
POMGnT1	Protein O-linked mannose O-1,2-N-acetylglucosaminyltransferase 1
POMGnT2	Protein O-linked mannose O-1,4-N-acetylglucosaminyltransferase 2
POMK	Protein O-mannose kinase
POMT1	Protein O-mannosyltransferase 1
POMT2	Protein O-mannosyltransferase 2
PTM	Post-translation modification
PTPRF	Receptor-type tyrosine-protein phosphatase F
PTPRZ1	Receptor-type tyrosine-protein phosphatase zeta

Rbo5P	Ribitol-5-phosphate
RPN2	Dolichyl-diphosphooligosaccharide-protein glycosyltransferase subunit 2
SP	Signal peptide
TBS	Tris-buffered saline
TFA	Trifluoroacetic acid
TMR	Transmembrane domain
TSR	Thrombospondin-type repeat
TMEM5	Ribitol $\beta$ 1,4-xylosyltransferase
WGA	Wheat germ agglutinin
WWS	Walker-Warburg syndrome
Xyl	Xylose
$\alpha$ DG	$\alpha$ -dystroglycan
$\beta$ DG	$\beta$ -dystroglycan

## References

- Abbott, K.L., Matthews, R.T., Pierce, M., 2008. Receptor Tyrosine Phosphatase  $\beta$  (RPTP $\beta$ ) Activity and Signaling Are Attenuated by Glycosylation and Subsequent Cell Surface Galectin-1 Binding. *J Biol Chem* 283, 33026.
- Abbott, K.L., Pierce, J.M., 2010. Lectin-Based Glycoproteomic Techniques for the Enrichment and Identification of Potential Biomarkers. *Methods Enzymol* 480, 461–476.
- Acar, M., Jafar-Nejad, H., Takeuchi, H., Rajan, A., Ibrani, D., Rana, N.A., Pan, H., Haltiwanger, R.S., Bellen, H.J., 2008. Rumi, a CAP10 domain protein, is a glycosyltransferase that modifies Notch and is required for Notch signaling. *Cell* 132, 247.
- Agard, N.J., Prescher, J.A., Bertozzi, C.R., 2004. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J Am Chem Soc* 126, 15046–15047.
- Akasaka-Manyá, K., Manyá, H., Endo, T., 2004. Mutations of the POMT1 gene found in patients with Walker-Warburg syndrome lead to a defect of protein O-mannosylation. *Biochem Biophys Res Commun* 325, 75–79.
- Akhavan, A., Crivelli, S.N., Singh, M., Lingappa, V.R., Muschler, J.L., 2008. SEA domain proteolysis determines the functional composition of dystroglycan. *The FASEB Journal* 22, 612–621.
- Arike, L., Hansson, G.C., 2016. The Densely O-Glycosylated MUC2 Mucin Protects the Intestine and Provides Food for the Commensal Bacteria. *J Mol Biol*.
- Bandini, G., Haserick, J.R., Motari, E., Ouologuem, D.T., Lourido, S., Roos, D.S., Costello, C.E., Robbins, P.W., Samuelson, J., 2016. O-fucosylated glycoproteins form assemblies in close proximity to the nuclear pore complexes of *Toxoplasma gondii*. *Proceedings of the National Academy of Sciences* 113, 11567–11572.
- Bank, P.A. De, Kellam, B., Kendall, D.A., Shakesheff, K.M., 2003. Surface Engineering of Living Myoblasts Via Selective Periodate Oxidation.
- Bansil, R., Turner, B.S., 2018. The biology of mucus: Composition, synthesis and organization. *Adv Drug Deliv Rev* 124, 3–15.
- Barresi, R., Campbell, K.P., 2006. Dystroglycan : from biosynthesis to pathogenesis of human disease. *J Cell Sci* 119, 199–207.
- Barresi, R., Michele, D.E., Kanagawa, M., Harper, H.A., Dovico, S.A., Satz, J.S., Moore, S.A., Zhang, W., Schachter, H., Dumanski, J.P., Cohn, R.D., Nishino, I., Campbell, K.P., 2004. LARGE can functionally bypass  $\alpha$ -dystroglycan glycosylation defects in distinct congenital muscular dystrophies. *Nature Medicine* 2004 10:7 10, 696–703.
- Bartels, M.F., Winterhalter, P.R., Yu, J., Liu, Y., Lommel, M., Möhrlein, F., Hu, H., Feizi, T., Westerlind, U., Ruppert, T., Strahl, S., 2016. Protein O-mannosylation in the murine brain: Occurrence of Mono-O-Mannosyl glycans and identification of new substrates. *PLoS One* 11, e0166119.

- Batt, A.R., Zaro, B.W., Navarro, M.X., Pratt, M.R., 2017. Metabolic chemical reporters of glycans exhibit cell-type selective metabolism and glycoprotein labeling HHS Public Access. *Chembiochem* 18, 1177–1182.
- Becker, J.W., Reeke, G.N., Wang, J.L., Cunningham, B.A., Edelman, G.M., 1975. The covalent and three dimensional structure of concanavalin A. III. Structure of the monomer and its interactions with metals and saccharides. *Journal of Biological Chemistry* 250, 1513–1524.
- Beltrán-Valero De Bernabé, D., Currier, S., Steinbrecher, A., Celli, J., Van Beusekom, E., Van Der Zwaag, B., Lya Kayserili, H., Merlini, L., Chitayat, D., Dobyns, W.B., Cormand, B., Lehesjoki, A.-E., Cruces, J., Voit, T., Walsh, C.A., Van Bokhoven, H., Brunner, H.G., 2002. Mutations in the O-Mannosyltransferase Gene POMT1 Give Rise to the Severe Neuronal Migration Disorder Walker-Warburg Syndrome, *Am. J. Hum. Genet.*
- Bhat, A.H., Maity, S., Giri, K., Ambatipudi, K., 2019. Protein glycosylation: Sweet or bitter for bacterial pathogens? *Crit Rev Microbiol* 45, 82–102.
- Bleckmann, C., Geyer, H., Lieberoth, A., Splittstoesser, F., Liu, Y., Feizi, T., Schachner, M., Kleene, R., Reinhold, V., Geyer, R., 2009. O-glycosylation pattern of CD24 from mouse brain. *Biol Chem* 390, 627–645.
- Blobel, G., Dobberstein, B., 1975a. Transfer of proteins across membranes. II. Reconstitution of functional rough microsomes from heterologous components. *Journal of Cell Biology* 67, 852–862.
- Blobel, G., Dobberstein, B., 1975b. Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *Journal of Cell Biology* 67, 835–851.
- Bowe, M.A., Deyst, K.A., Leszyk, J.D., Fallon, J.R., 1994. Identification and purification of an agrin receptor from torpedo postsynaptic membranes: A heteromeric complex related to the dystroglycans. *Neuron* 12, 1173–1180.
- Brasch, J., Harrison, O.J., Honig, B., Shapiro, L., 2012. Thinking outside the cell: How cadherins drive adhesion. *Trends Cell Biol.*
- Briggs, D.C., Hohenester, E., 2018. Structural Basis for the Initiation of Glycosaminoglycan Biosynthesis by Human Xylosyltransferase 1. *Structure.*
- Brockington, M., Blake, D.J., Prandini, P., Brown, S.C., Torelli, S., Benson, M.A., Ponting, C.P., Estournet, B., Romero, N.B., Mercuri, E., Voit, T., Sewry, C.A., Guicheney, P., Muntoni, F., 2001. Mutations in the Fukutin-Related Protein Gene (FKRP) Cause a Form of Congenital Muscular Dystrophy with Secondary Laminin  $\alpha$ 2 Deficiency and Abnormal Glycosylation of  $\alpha$ -Dystroglycan. *Am J Hum Genet* 69, 1198.
- Brockington, M., Torelli, S., Prandini, P., Boito, C., Dolatshad, N.F., Longman, C., Brown, S.C., Muntoni, F., 2005. Localization and functional analysis of the LARGE family of glycosyltransferases: significance for muscular dystrophy. *Hum Mol Genet* 14, 657–665.
- Brockington, Martin, Yuva, Y., Prandini, P., Brown, S.C., Torelli, S., Benson, M.A., Herrmann, R., Anderson, L.V.B., Bashir, R., Burgunder, J.M., Fallet, S., Romero, N., Fardeau, M., Straub, V.,

- Storey, G., Pollitt, C., Richard, I., Sewry, C.A., Bushby, K., Voit, T., Blake, D.J., Muntoni, F., 2001. Mutations in the fukutin-related protein gene (FKRP) identify limb girdle muscular dystrophy 2I as a milder allelic variant of congenital muscular dystrophy MDC1C. *Hum Mol Genet* 10, 2851–2859.
- Brown, S., Santa Maria Jr, J.P., Walker, S., 2013. Wall Teichoic Acids of Gram-Positive Bacteria PG: peptidoglycan WTA: wall teichoic acid MRSA: methicillin-resistant *Staphylococcus aureus*.
- Chai, W., Yuen, C.T., Kogelberg, H., Carruthers, R.A., Margolis, R.U., Feizi, T., Lawson, A.M., 1999. High prevalence of 2-mono- and 2,6-di-substituted manol-terminating sequences among O-glycans released from brain glycopeptides by reductive alkaline hydrolysis. *Eur J Biochem* 263, 879–888.
- Chang, P. V., Prescher, J.A., Hangauer, M.J., Bertozzi, C.R., 2007. Imaging cell surface glycans with bioorthogonal chemical reporters. *J Am Chem Soc* 129, 8400–8401.
- Chiba, A., Matsumura, K., Yamada, H., Inazu, T., Shimizu, T., Kusunoki, S., Kanazawa, I., Kobata, A., Endo, T., 1997. Structures of sialylated O-linked oligosaccharides of bovine peripheral nerve  $\alpha$ -dystroglycan. The role of a novel O-mannosyl-type oligosaccharide in the binding of  $\alpha$ -dystroglycan with laminin. *Journal of Biological Chemistry* 272, 2156–2162.
- Chick, J.M., Kolippakkam, D., Nusinow, D.P., Zhai, B., Rad, R., Huttlin, E.L., Gygi, S.P., Biotechnol, N., 2015. An ultra-tolerant database search reveals that a myriad of modified peptides contributes to unassigned spectra in shotgun proteomics HHS Public Access Author manuscript. *Nat Biotechnol* 33, 743–749.
- Chien, P.N., Ryu, S.E., 2013. Protein tyrosine phosphatase  $\sigma$  in proteoglycan-mediated neural regeneration regulation. *Mol Neurobiol* 47, 220–227.
- Cioce, A., Bineva-Todd, G., Agbay, A.J., Choi, J., Wood, T.M., Debets, M.F., Browne, W.M., Douglas, H.L., Roustan, C., Tastan, O.Y., Kjaer, S., Bush, J.T., Bertozzi, C.R., Schumann, B., 2021. Optimization of Metabolic Oligosaccharide Engineering with Ac4GalNAk and Ac4GlcNAk by an Engineered Pyrophosphorylase. *ACS Chem Biol* 16, 1961–1967.
- Cole, L.A., 2010. Biological functions of hCG and hCG-related molecules. *Reproductive Biology and Endocrinology*.
- Collins, B.E., Yang, L.J.S., Mukhopadhyay, G., Filbin, M.T., Kiso, M., Hasegawa, A., Schnaar, R.L., 1997. Sialic Acid Specificity of Myelin-associated Glycoprotein Binding. *Journal of Biological Chemistry* 272, 1248–1255.
- Comer, F.I., Hart, G.W., 2001. Reciprocity between O-GlcNAc and O-phosphate on the carboxyl terminal domain of RNA polymerase II. *Biochemistry* 40, 7845–7852.
- Corfield, A.P., 2015. Mucins: A biologically relevant glycan barrier in mucosal protection. *Biochim Biophys Acta Gen Subj* 1850, 236–252.
- Crick, F.H.C., 1958. On protein synthesis. *Symp Soc Exp Biol.* 12, 138–63.

- Crimaudo, C., Hortsch, M., Gausepohl, H., & Meyer, D. I., 1987. Human ribophorins I and II: the primary structure and membrane topology of two highly conserved rough endoplasmic reticulum-specific glycoproteins. *The EMBO Journal*, 6(1), 75-82.
- Davis, C.G., Elhammerq, A., Russell, D.W., Schneider, W.J., Kornfeldll, S., Brown, M.S., Goldstein, J.L., 1986. Deletion of Clustered O-Linked Carbohydrates Does Not Impair Function of Low-Density Lipoprotein Receptor in Transfected Fibroblasts. *THE JOURNAL OF BIOLOGICAL CHEMISTRY* 261, 2828–2838.
- del Monte, F., Agnetti, G., 2014. Protein Post-Translational Modifications and Misfolding: New Concepts in Heart Failure. *Proteomics Clin Appl* 8, 534.
- Dodds, E.D., Seipert, R.R., Clowers, B.H., German, J.B., Lebrilla, C.B., 2009. Analytical Performance of Immobilized Pronase for Glycopeptide Footprinting and Implications for Surpassing Reductionist Glycoproteomics. *J Proteome Res* 8, 502–512.
- Domon, B., Costello, C.E., 1988. A Systematic Nomenclature for Carbohydrate Fragmentations in FAB-MS/MS Spectra of Glycoconjugates, *Glycoconjugate J.*
- Dube, D.H., Prescher, J.A., Quang, C.M., Bertozzi, C.R., 2006. Probing mucin-type O-linked glycosylation in living animals. *Proc Natl Acad Sci U S A* 103, 4819–4824.
- Durbeej, M., Henry, M.D., Ferletta, M., Campbell, K.P., Ekblom, P., 1998. Distribution of Dystroglycan in Normal Adult Mouse Tissues, *The Journal of Histochemistry & Cytochemistry (J Histochem Cytochem)*.
- Dwyer, C.A., Baker, E., Hu, H., Matthews, R.T., 2012. RPTPζ/phosphacan is abnormally glycosylated in a model of muscle-eye-brain disease lacking functional POMGnT1. *Neuroscience* 220, 47.
- Endo, T., 2015. Glycobiology of α-dystroglycan and muscular dystrophy. *J Biochem* 157, 1–12.
- Ervasti, J.M., Campbell, K.P., 1991. Membrane Organization of the Dystrophin-Glycoprotein. *Cell* 66, 1121–1131.
- Ervasti, J.M., Campbell, K.P., 1993. A role for the dystrophin-glycoprotein complex as a transmembrane linker between laminin and actin. *Journal of Cell Biology* 122, 809–823.
- Esser, A.K., Miller, M.R., Huang, Q., Meier, M.M., De Bernabé, D.B.V., Stipp, C.S., Campbell, K.P., Lynch, C.F., Smith, B.J., Cohen, M.B., Henry, M.D., 2013. Loss of LARGE2 Disrupts Functional Glycosylation of α-Dystroglycan in Prostate Cancer. *J Biol Chem* 288, 2132.
- Fernandez De La Mora, J., 2000. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Anal Chim Acta* 406, 93–104.
- Fernandez-Valdivia, R., Takeuchi, H., Samarghandi, A., Lopez, M., Leonardi, J., Haltiwanger, R.S., Jafar-Nejad, H., 2011. Regulation of mammalian Notch signaling and embryonic development by the protein O-glucosyltransferase Rumi. *Development* 138, 1925–1934.
- Finne, J., Krusius, T., Margolis, R.K., Margolis, R.U., 1979. Novel Mannitol-containing Oligosaccharides Obtained by Mild Alkaline Borohydride Treatment of a Chondroitin Sulfate Proteoglycan from Brain. *J Biol Chem* 254, 10295–10300.

- Fransson, L.-A., 1968. Structure of dermatan sulfate: IV. Glycopeptides from the carbohydrate-protein linkage region of pig skin dermatan sulfate. *Biochim Biophys Acta* 156, 311–316.
- Gahmberg, C.G., Andefsson, L.C., 1977. Selective Radioactive Labeling of Cell Surface Sialoglycoproteins by Periodate-tritiated Borohydride\*. *Journal of Biological Chemistry* 252, 5888–5894.
- Geis, T., Marquard, K., Rödl, T., Reihle, C., Schirmer, S., Von Kalle, T., Bornemann, A., Hehr, U., Blankenburg, M., 2013. Homozygous dystroglycan mutation associated with a novel muscle-eye-brain disease-like phenotype with multicystic leucodystrophy. *Neurogenetics* 14, 205–213.
- Gerin, I., Ury, B., Breloy, I., Bouchet-Seraphin, C., Bolsée, J., Halbout, M., Graff, J., Vertommen, D., Muccioli, G.G., Seta, N., Cuisset, J.M., Dabaj, I., Quijano-Roy, S., Grahn, A., Van Schaftingen, E., Bommer, G.T., 2016. ISPD produces CDP-ribitol used by FKTN and FKRP to transfer ribitol phosphate onto  $\alpha$ -dystroglycan. *Nat Commun* 7.
- Glass II, W.F., Briggs, R.C., Hnilica, L.S., 1981. Use of Lectins for Detection of Electrophoretically Separated Glycoproteins Transferred onto Nitrocellulose Sheets', *ANALYTICAL BIOCHEMISTRY*.
- Goddeeris, M.M., Wu, B., Venzke, D., Yoshida-Moriguchi, T., Saito, F., Matsumura, K., Moore, S.A., Campbell, K.P., 2013. LARGE glycans on dystroglycan function as a tunable matrix scaffold to prevent dystrophy. *Nature*.
- Grewal, P.K., Holzfeind, P.J., Bittner, R.E., Hewitt, J.E., 2001. Mutant glycosyltransferase and altered glycosylation of  $\alpha$ -dystroglycan in the myodystrophy mouse. *Nature Genetics* 28:2 28, 151–154.
- Grumet, M., Sakurai, T., 1996. Heterophilic interactions of the neural cell adhesion molecules Ng-CAM and Nr-CAM with neural receptors and extracellular matrix proteins. *Seminars in neurosciences* 8, 379–389.
- Hallgren, J., Tsigirgos, K.D., Damgaard Pedersen, M., Juan, J., Armenteros, A., Marcatili, P., Nielsen, H., Krogh, A., Winther, O., 2022. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *bioRxiv* 2022.04.08.487609.
- Halmó, S.M., Singh, D., Patel, S., Wang, S., Edlin, M., Boons, G.J., Moremen, K.W., Live, D., Wells, L., 2017. Protein O-linked mannanose  $\beta$ -1,4-N-acetylglucosaminyltransferase 2 (POMGNT2) is a gatekeeper enzyme for functional glycosylation of  $\alpha$ -dystroglycan. *Journal of Biological Chemistry* 292, 2101–2109.
- Hang, H.C., Yu, C., Kato, D.L., Bertozzi, C.R., 2003. A metabolic labeling approach toward proteomic analysis of mucin-type O-linked glycosylation. *Proc Natl Acad Sci U S A* 100, 14846.
- Hansson, G.C., Johansson, M.E. V, 2010. Gut Microbes The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Gut Microbes* 1, 51–54.
- Hara, Y., Balci-Hayta, B., Yoshida-Moriguchi, T., Kanagawa, M., Bernabé, D.B.-V. de, Gündeşli, H., Willer, T., Satz, J.S., Crawford, R.W., Burden, S.J., Kunz, S., Oldstone, M.B.A., Accardi, A., Talim,

- B., Muntoni, F., Topaloglu, H., Dinçer, P., Campbell, K.P., 2011. A Dystroglycan Mutation Associated with Limb-Girdle Muscular Dystrophy. *N Engl J Med* 364, 939.
- Harrison, R., Hitchen, P.G., Panico, M., Morris, H.R., Mekhail, D., Pleass, R.J., Dell, A., Hewitt, J.E., Haslam, S.M., 2012. Glycoproteomic characterization of recombinant mouse -dystroglycan. *Glycobiology* 22, 662–675.
- Henry, M.D., Campbell, K.P., 1998. A Role for Dystroglycan in Basement Membrane Assembly. *Cell* 95, 859–870.
- Herscovics, A., 1999. Importance of glycosidases in mammalian glycoprotein biosynthesis. *Biochim Biophys Acta* 1473, 96–107.
- 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. *Journal of Biological Chemistry* 276, 6485–6498.
- Holmberg, J., Durbeej, M., 2013. Laminin-211 in skeletal muscle function. *Cell Adh Migr*.
- Holt, K.H., Crosbie, R.H., Venzke, D.P., Campbell, K.P., 2000. Biosynthesis of dystroglycan: processing of a precursor propeptide. *FEBS Journal* 468, 79–83.
- Holzfeind, P.J., Grewal, P.K., Reitsamer, H.A., Kechvar, J., Lassmann, H., Hoeger, H., Hewitt, J.E., Bittner, R.E., 2002. Skeletal, cardiac and tongue muscle pathology, defective retinal transmission, and neuronal migration defects in the *Largemyd* mouse defines a natural model for glycosylation-deficient muscle-eye-brain disorders. *Hum Mol Genet* 11, 2673–2687.
- Hortsch, M., 2000. Structural and Functional Evolution of the L1 Family: Are Four Adhesion Molecules Better Than One? *Molecular and Cellular Neuroscience* 15, 1–10.
- Huang, Y., Triscari, J.M., Tseng, G.C., Pasa-Tolic, L., Lipton, M.S., Smith, R.D., Wysocki, V.H., 2005. Statistical Characterization of the Charge State and Residue Dependence of Low Energy CID Peptide Dissociation Patterns. *Anal Chem* 77, 5800.
- Imae, R., Kuwabara, N., Many, H., Tanaka, T., Tsuyuguchi, M., Mizuno, M., Endo, T., Kato, R., 2021. The structure of POMGNT2 provides new insights into the mechanism to determine the functional O-mannosylation site on  $\alpha$ -dystroglycan. *Genes to Cells* 26, 485–494.
- Inamori, K., Yoshida-moriguchi, T., Hara, Y., Anderson, M.E., Yu, L., Campbell, K.P., 2012. Dystroglycan Function Requires Xylosyl- and Glucuronyltransferase Activities of LARGE. *Science (1979)* 335, 93–97.
- Inamori, K.-I., Beedle, A.M., Beltrán-Valero De Bernabé, D., Wright, M.E., Campbell, K.P., 2016. LARGE2-dependent glycosylation confers laminin-binding ability on proteoglycans. *Glycobiology* 26, 1284–1296.
- Jae, L.T., Raaben, M., Riemersma, M., Van Beusekom, E., Blomen, V.A., Velds, A., Kerkhoven, R.M., Carette, J.E., Topaloglu, H., Meinecke, P., Wessels, M.W., Lefeber, D.J., Whelan, S.P., Van Bokhoven, H., Brummelkamp, T.R., 2013. Deciphering the Glycosylome of Dystroglycanopathies Using Haploid Screens for Lassa Virus Entry. *Science (1979)* 340, 479–483.

- Jayasinha, V., Nguyen, H.H., Xia, B., Kammesheidt, A., Hoyte, K., Martin, P.T., 2003. Inhibition of dystroglycan cleavage causes muscular dystrophy in transgenic mice. *Neuromuscular Disorders* 13, 365–375.
- Jiang, P., Du, W., Mancuso, A., Wellen, K.E., Yang, X., 2013. Reciprocal regulation of p53 and malic enzymes modulates metabolism and senescence. *Nature* 2013 493:7434 493, 689–693.
- Johnson, R.S., Martin, S.A., Biemann, K., Stults, J.T., Watson, J.T., 1987. Novel Fragmentation Process of Peptides by Collision-Induced Decomposition in a Tandem Mass Spectrometer: Differentiation of Leucine and Isoleucine. *Anal Chem* 59, 2621–2625.
- Kanagawa, M., Kobayashi, K., Tajiri, M., Manya, H., Kuga, A., Yamaguchi, Y., Akasaka-Manya, K., Furukawa, J. ichi, Mizuno, M., Kawakami, H., Shinohara, Y., Wada, Y., Endo, T., Toda, T., 2016a. Identification of a Post-translational Modification with Ribitol-Phosphate and Its Defect in Muscular Dystrophy. *Cell Rep* 14, 2209–2223.
- Kanagawa, M., Nishimoto, A., Chiyonobu, T., Takeda, Satoshi, Miyagoe-Suzuki, Y., Wang, F., Fujikake, N., Taniguchi, M., Lu, Z., Tachikawa, M., Nagai, Y., Tashiro, F., Miyazaki, J.I., Tajima, Y., Takeda, Shin'ichi, Endo, T., Kobayashi, K., Campbell, K.P., Toda, T., 2009. Residual laminin-binding activity and enhanced dystroglycan glycosylation by LARGE in novel model mice to dystroglycanopathy. *Hum Mol Genet* 18, 621–631.
- Kanagawa, M., Toda, T., 2017. Muscular Dystrophy with Ribitol-Phosphate Deficiency: A Novel Post-Translational Mechanism in Dystroglycanopathy. *J Neuromuscul Dis* 4, 259–267.
- Karve, T.M., Cheema, A.K., 2011. Small Changes Huge Impact: The Role of Protein Posttranslational Modifications in Cellular Homeostasis and Disease. *J Amino Acids* 2011, 1–13.
- Kazuhiro Kobayashi, Yutaka Nakahori, Masashi Miyake, Kiichiro Matsumura, Eri Kondo-Iida, Yoshiko Nomura, Masaya Segawa, Mieko Yoshioka, Kayoko Saito, Makiko Osawa, Kenzo Hamano, Youichi Sakakihara, Ikuya Nonaka, Yasuo Nakagome, Ichiro Kanazawa, Yusuke Nak, K.T.& T.T., 1998. An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. *Nature* 394, 388–392.
- Keenan, E.K., Zachman, D.K., Hirschey, M.D., 2021. Discovering the landscape of protein modifications. *Mol Cell* 81, 1868–1878.
- Kelleher, D.J., Kreibich, G., Gilmore, R., 1992. Oligosaccharyltransferase activity is associated with a protein complex composed of ribophorins I and II and a 48 kd protein. *Cell* 69, 55–65.
- Khoury, G.A., Baliban, R.C., Floudas, C.A., 2011. Proteome-wide post-translational modification statistics: Frequency analysis and curation of the swiss-prot database. *Sci Rep* 1, 1–5.
- Kiick, K.L., Saxon, E., Tirrell, D.A., Bertozzi, C.R., 2002. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc Natl Acad Sci U S A* 99, 19–24.
- Kobayashi, Y., Tateno, H., Ogawa, H., Yamamoto, K., Hirabayashi, J., 2014. Comprehensive list of lectins: Origins, natures, and carbohydrate specificities. *Methods in Molecular Biology* 1200, 555–577.

- Kolarich, D., Jensen, P.H., Altmann, F., Packer, N.H., 2012. Determination of site-specific glycan heterogeneity on glycoproteins. *Nat Protoc* 7, 1285–1298.
- Kölmel, D.K., Kool, E.T., 2017. Oximes and Hydrazones in Bioconjugation: Mechanism and Catalysis. *Chem Rev* 117, 10358–10376.
- Kunz, S., Rojek, J.M., Kanagawa, M., Spiropoulou, C.F., Barresi, R., Campbell, K.P., Oldstone, M.B.A., 2005. Posttranslational Modification of  $\alpha$ -Dystroglycan, the Cellular Receptor for Arenaviruses, by the Glycosyltransferase LARGE Is Critical for Virus Binding. *J Virol* 79, 14282.
- Kuwabara, N., Imae, R., Manya, H., Tanaka, T., Mizuno, M., Tsumoto, H., Kanagawa, M., Kobayashi, K., Toda, T., Senda, T., Endo, T., Kato, R., 2020. Crystal structures of fukutin-related protein (FKRP), a ribitol-phosphate transferase related to muscular dystrophy. *Nat Commun* 11.
- Lamond, A.I., Uhlen, M., Horning, S., Makarov, A., Robinson, C. V., Serrano, L., Hartl, F.U., Baumeister, W., Werenskiold, A.K., Andersen, J.S., Vorm, O., Linial, M., Aebersold, R., Mann, M., 2012. Advancing Cell Biology Through Proteomics in Space and Time (PROSPECTS). *Mol Cell Proteomics* 11.
- Larsen, I.S.B., Narimatsu, Y., Joshi, H.J., Siukstaite, L., Harrison, O.J., Brasch, J., Goodman, K.M., Hansen, L., Shapiro, L., Honig, B., Vakhrushev, S.Y., Clausen, H., Halim, A., 2017. Discovery of an O-mannosylation pathway selectively serving cadherins and protocadherins. *Proc Natl Acad Sci U S A* 114, 11163–11168.
- Larsen, I.S.B., Povolo, L., Zhou, L., Tian, W., Mygind, K.J., Hintze, J., Jiang, C., Hartill, V., Prescott, K., Johnson, C.A., Mullegama, S. V., McConkie-Rosell, A., McDonald, M., Hansen, L., Vakhrushev, S.Y., Schjoldager, K.T., Clausen, H., Worzfeld, T., Joshi, H.J., Halim, A., 2023. The SHDRA syndrome-associated gene TMEM260 encodes a protein-specific O-mannosyltransferase. *Proc Natl Acad Sci U S A* 120, e2302584120.
- Li, H., Frankenfield, A.M., Houston, R., Sekine, S., Hao, L., 2021. Thiol-Cleavable Biotin for Chemical and Enzymatic Biotinylation and Its Application to Mitochondrial TurboID Proteomics.
- Li, Z., Li, S., Luo, M., Jhong, J.-H., Li, W., Yao, L., Pang, Y., Wang, Z., Wang, R., Ma, R., Yu, J., Huang, Y., Zhu, X., Cheng, Q., Feng, H., Zhang, J., Wang, C., Bo-Kai Hsu, J., Chang, W.-C., Wei, F.-X., Huang, H.-D., Lee, T.-Y., 2021. dbPTM in 2022: an updated database for exploring regulatory networks and functional associations of protein post-translational modifications. *Nucleic Acids Res* 50.
- Liu, F., Iqbal, K., Grundke-Iqbal, I., Hart, G.W., Gong, C.-X., 2004. O-GlcNAcylation regulates phosphorylation of tau: A mechanism involved in Alzheimer's disease.
- Lommel, M., Winterhalter, P.R., Willer, T., Dahlhoff, M., Schneider, M.R., Bartels, M.F., Renner-Müller, I., Ruppert, T., Strahl, S., 2013. Protein O-mannosylation is crucial for E-cadherin-mediated cell adhesion. *PNAS* 110, 21024–21029.
- Luo, Y., Nita-Lazar, A., Haltiwanger, R.S., 2006. Two distinct pathways for O-fucosylation of epidermal growth factor-like or thrombospondin type 1 repeats. *Journal of Biological Chemistry* 281, 9385–9392.

- Ma, Z.Y., Skorobogatko, Y., Vosseller, K., 2013. Tandem lectin weak affinity chromatography for glycoprotein enrichment. *Methods in Molecular Biology* 951, 21–31.
- Makarov, A., 2000. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal Chem* 72, 1156–1162.
- Marshall, R., 1974. The nature and metabolism of the carbohydrate-peptide linkages of glycoproteins. *Biochem Soc Symp.* 40, 17–26.
- Martin-Blanco, E., Garcia-Bellido, A., 1996. Developmental Biology Mutations in the rotated abdomen locus affect muscle development and reveal an intrinsic asymmetry in *Drosophila* (helical symmetry/muscle development/segmental staggering). *Proc. Natl. Acad. Sci* 93, 6048–6052.
- Martinez, T., Pace, D., Brady, L., Gerhart, M., Balland, A., 2007. Characterization of a novel modification on IgG2 light chain: Evidence for the presence of O-linked mannosylation. *J Chromatogr A* 1156, 183–187.
- Matta, C., Boocock, D.J., Fellows, C.R., Miosge, N., Dixon, J.E., Liddell, S., Smith, J., Mobasheri, A., 2019. Molecular phenotyping of the surfaceome of migratory chondroprogenitors and mesenchymal stem cells using biotinylation, glycocapture and quantitative LC-MS/MS proteomic analysis. *Sci Rep* 9, 1–15.
- Michael Hortsch, 1996. The L1 Family of Neural Review Cell Adhesion Molecules: Old Proteins Performing New Tricks. *Neuron* 17, 587–593.
- Michalski, A., Damoc, E., Lange, O., Denisov, E., Nolting, D., Müller, M., Viner, R., Schwartz, J., Remes, P., Belford, M., Dunyach, J.J., Cox, J., Horning, S., Mann, M., Makarov, A., 2012. Ultra High Resolution Linear Ion Trap Orbitrap Mass Spectrometer (Orbitrap Elite) Facilitates Top Down LC MS/MS and Versatile Peptide Fragmentation Modes. *Mol Cell Proteomics* 11, 1–11.
- Morelle, W., Canis, K., Chirat, F., Faid, V., Michalski, J.C., 2006. The use of mass spectrometry for the proteomic analysis of glycosylation. *Proteomics* 6, 3993–4015.
- Morise, J., Kizuka, Y., Yabuno, K., Tonoyama, Y., Hashii, N., Kawasaki, N., Manya, H., Miyagoe-Suzuki, Y., Takeda, S., Endo, T., Maeda, N., Takematsu, H., Oka, S., 2014. Structural and biochemical characterization of O-mannose-linked human natural killer-1 glycan expressed on phosphacan in developing mouse brains. *Glycobiology* 24, 314–324.
- Motoi Kanagawa, Fumiaki Saito, Stefan Kunz, Takako Yoshida-Moriguchi, Rita Barresi, Yvonne M Kobayashi, John Muschler, Jan P Dumanski, Daniel E Michele, Michael B.A Oldstone, Kevin P Campbell, 2004. Molecular Recognition by LARGE Is Essential for Expression of Functional Dystroglycan. *Cell* 117, 953–964.
- Nguyen, H., Ostendorf, A.P., Satz, J.S., Westra, S., Ross-Barta, S.E., Campbell, K.P., Moore, S.A., 2013. Glial scaffold required for cerebellar granule cell migration is dependent on dystroglycan function as a receptor for basement membrane proteins. *Acta Neuropathol Commun* 1.
- Nguyen, S., Fenn, J.B., 2007. Gas-phase ions of solute species from charged droplets of solutions. *Proc Natl Acad Sci U S A* 104, 1111–1117.

- Nickolls, A.R., Bönnemann, C.G., 2018. The roles of dystroglycan in the nervous system: Insights from animal models of muscular dystrophy. *DMM Disease Models and Mechanisms* 11.
- Nicole, S., Davoine, C.S., Topaloglu, H., Cattolico, L., Barral, D., Beighton, P., Hamida, C. Ben, Hammouda, H., Cruaud, C., White, P.S., Samson, D., Urtizberea, J.A., Lehmann-Horn, F., Weissenbach, J., Hentati, F., Fontaine, B., 2000. Perlecan, the major proteoglycan of basement membranes, is altered in patients with Schwartz-Jampel syndrome (chondrodystrophic myotonia). *Nature Genetics* 26:4 26, 480–483.
- Nilsson, J., Rüetschi, U., Halim, A., Hesse, C., Carlsohn, E., Brinkmalm, G., Larson, G., 2009. Enrichment of glycopeptides for glycan structure and attachment site identification. *Nature Methods* 2009 6:11 6, 809–811.
- Nilsson, Johanna, Nilsson, Jonas, Larson, G., 2010. Characterization of site-specific O-glycan structures within the mucin-like domain of  $\alpha$ -dystroglycan from human skeletal muscle. *Glycobiology* 20, 1160–1169.
- Nishino, K., Yoshikawa, H., Motani, K., Kosako, H., 2022. Optimized Workflow for Enrichment and Identification of Biotinylated Peptides Using Tamavidin 2-REV for BioID and Cell Surface Proteomics. *J Proteome Res* 21, 2094–2103.
- Ohtsubo, K., Marth, J.D., 2006. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* 126, 855–867.
- Okeley, N.M., Alley, S.C., Anderson, M.E., Boursalian, T.E., Burke, P.J., Emmerton, K.M., Jeffrey, S.C., Klussman, K., Law, C.L., Sussman, D., Toki, B.E., Westendorf, L., Zeng, W., Zhang, X., Benjamin, D.R., Senter, P.D., 2013. Development of orally active inhibitors of protein and cellular fucosylation. *Proc Natl Acad Sci U S A* 110, 5404–5409.
- Okuma, H., Hord, J.M., Chandel, I., Venzke, D., Anderson, M.E., Walimbe, A.S., Joseph, S., Gastel, Z., Hara, Y., Saito, F., Matsumura, K., Campbell, K.P., 2023. N-terminal domain on dystroglycan enables LARGE1 to extend matriglycan on  $\alpha$ -dystroglycan and prevents muscular dystrophy. *Elife* 12, 82811.
- Oppizzi, M.L., Akhavan, A., Singh, M., Fata, J.E., Muschler, J.L., 2008. Nuclear translocation of  $\beta$ -dystroglycan reveals a distinctive trafficking pattern of autoproteolyzed mucins. *Traffic* 9, 2063–2072.
- Oxana Ibraghimov-Beskrovnaya, James M. Ervasti, Cynthia J. Leveille, Clive A. Slaughter, S.W.S. & K.P.C., 1992. Primary structure of dystrophin-associated glycoproteins linking dystrophin to the extracellular matrix. *Nature* 355, 696–702.
- Pacharra, S., Hanisch, F.G., Breloy, I., 2012. Neurofascin 186 is O-mannosylated within and outside of the mucin domain. *J Proteome Res* 11, 3955–3964.
- Pacharra, S., Hanisch, F.G., Mühlenhoff, M., Faissner, A., Rauch, U., Breloy, I., 2013. The lecticans of mammalian brain perineural net are O-mannosylated. *J Proteome Res* 12, 1764–1771.
- Paprocka, J., Jezela-Stanek, A., Tylki-Szymańska, A., Grunewald, S., 2021. Congenital Disorders of Glycosylation from a Neurological Perspective. *Brain Sci* 11, 1–25.

- Pedersen, L.C., Tsuchida, K., Kitagawa, H., Sugahara, K., Darden, T.A., Negishi, M., 2000. Heparan/Chondroitin Sulfate Biosynthesis STRUCTURE AND MECHANISM OF HUMAN GLUCURONYLTRANSFERASE I\*. *Journal of Biological Chemistry* 275, 34580–34585.
- Perälä, N., Sariola, H., Immonen, T., 2012. More than nervous: The emerging roles of plexins. *Differentiation*.
- Pérez Jurado Luis A., Coloma Antonio, Cruces Jesús, 1999. Identification of a Human Homolog of the *Drosophila* rotated abdomen Gene (POMT1) Encoding a Putative Protein O-Mannosyl-Transferase, and Assignment to Human Chromosome 9q34.1. *Genomics* 58, 171–180.
- Pilia, G., Hughes-Benzie, R.M., MacKenzie, A., Baybayan, P., Chen, E.Y., Huber, R., Neri, G., Cao, A., Forabosco, A., Schlessinger, D., 1996. Mutations in GPC3, a glypican gene, cause the Simpson-Golabi-Behmel overgrowth syndrome. *Nature Genetics* 1996 12:3 12, 241–247.
- Praissman, J.L., Wells, L., 2014. Mammalian O-mannosylation pathway: Glycan structures, enzymes, and protein Substrates. *Biochemistry* 53, 3066–3078.
- Praissman, J.L., Willer, T., Osman Sheikh, M., Toi, A., Chitayat, D., Lin, Y.Y., Lee, H., Stalnakar, S.H., Wang, S., Prabhakar, P.K., Nelson, S.F., Stemple, D.L., Moore, S.A., Moremen, K.W., Campbell, K.P., Wells, L., 2016. The functional O-mannose glycan on  $\alpha$ -dystroglycan contains a phosphoribitol primed for matriglycan addition. *Elife* 5, 1–28.
- Ramya, T.N.C., Weerapana, E., Cravatt, B.F., Paulson, J.C., 2013. Glycoproteomics enabled by tagging sialic acid- or galactose-terminated glycans. *Glycobiology* 23, 211–221.
- Rangel-Angarita, V., Malaker, S.A., 2021. Mucinomics as the Next Frontier of Mass Spectrometry. *ACS Chem Biol* 16, 1866–1883.
- Rodrigues, J.G., Balmaña, M., Macedo, J.A., Poças, J., Fernandes, Â., Cesar, J., De-Freitas-Junior, M., Pinho, S.S., Gomes, J., Magalhães, A., Gomes, C., Mereiter, S., Reis, C.A., 2018. Glycosylation in cancer: Selected roles in tumour progression, immune modulation and metastasis. *Cell Immunol* 333, 46–57.
- Rostovtsev, V. V., Green, L.G., Fokin, V. V., Sharpless, K.B., 2002. A stepwise Huisgen cycloaddition process: Copper(I)-catalyzed regioselective “ligation” of azides and terminal alkynes. *Angewandte Chemie - International Edition* 41, 2596–2599.
- Rudd, P.M., Elliott, T., Cresswell, P., Wilson, I.A., Dwek, R.A., 2001. Glycosylation and the Immune System. *Science* (1979) 291, 2370–2376.
- Ryšlavá, H., Doubnerová, V., Kavan, D., Vaněk, O., 2013. Effect of posttranslational modifications on enzyme function and assembly. *J Proteomics*.
- Sakurai, T., Lustig, M., Nativ, M., Hemperly, J.J., Schlessinger, J., Peles, E., Grumet, M., 1997. Induction of Neurite Outgrowth through Contactin and Nr-CAM by Extracellular Regions of Glial Receptor Tyrosine Phosphatase. *J Cell Biol* 136, 907–918.
- Sarhan, A.R., Patel, T.R., Creese, A.J., Tomlinson, M.G., Hellberg, C., Heath, J.K., Hotchin, N.A., Cunningham, D.L., 2016. Regulation of Platelet Derived Growth Factor Signaling by Leukocyte

Common Antigen-related (LAR) Protein Tyrosine Phosphatase: A Quantitative Phosphoproteomics Study. *Molecular & Cellular Proteomics* 15, 1823–1836.

- Sasaki, T., Yamada, H., Matsumura, K., Shimizu, T., Kobata, A., Endo, T., 1998. Detection of O-mannosyl glycans in rabbit skeletal muscle K-dystroglycan. *Biochim Biophys Acta* 1425, 599–606.
- Satz, J.S., Ostendorf, A.P., Hou, S., Turner, A., Kusano, H., Lee, J.C., Turk, R., Nguyen, H., Ross-Barta, S.E., Westra, S., Hoshi, T., Moore, S.A., Campbell, K.P., 2010. Distinct functions of glial and neuronal dystroglycan in the developing and adult mouse brain. *J Neurosci* 30, 14560–14572.
- Saxon, E., Bertozzi, C.R., 2000. Cell surface engineering by a modified Staudinger reaction. *Science* (1979) 287, 2007–2010.
- Schenk, B., Fernandez, F., Waechter, C.J., 2001. The ins(ide) and outs(ide) of dolichyl phosphate biosynthesis and recycling in the endoplasmic reticulum, *Glycobiology*.
- Schjoldager, K.T., Narimatsu, Y., Joshi, H.J., Clausen, H., 2020. Global view of human protein glycosylation pathways and functions. *Nat Rev Mol Cell Biol*.
- Scigelova, M., Hornshaw, M., Giannakopoulos, A., Makarov, A., 2011. Fourier transform mass spectrometry. *Molecular and Cellular Proteomics* 10, 1–19.
- Sentandreu, R., Northcote, D.H., 1969. Yeast Cell-Wall Synthesis. *Biochem. J* 115, 231.
- Sethi, M.K., Buettner, F.F.R., Krylov, V.B., Takeuchi, H., Nifantiev, N.E., Haltiwanger, R.S., Gerardy-Schahn, R., Bakker, H., 2010. Identification of Glycosyltransferase 8 Family Members as Xylosyltransferases Acting on O-Glycosylated Notch Epidermal Growth Factor Repeats. *J Biol Chem* 285, 1582.
- Sheikh, M.O., Halmo, S.M., Wells, L., 2017. Recent advancements in understanding mammalian O-mannosylation. *Glycobiology* 27, 806–819.
- Shental-Bechor, D., Levy, Y., 2008. Effect of glycosylation on protein folding: A close look at thermodynamic stabilization.
- Singh, J., Itahana, Y., Knight-Krajewski, S., Kanagawa, M., Campbell, K.P., Bissell, M.J., Muschler, J., 2004. Proteolytic Enzymes and Altered Glycosylation Modulate Dystroglycan Function in Carcinoma Cells. *Cancer Res* 64, 6152–6159.
- Smith, K.K., Strickland, S., 1981. Structural Components and Characteristics of Reichert's Membrane, an Extra-embryonic Basement Membrane\* 256, 46544661.
- Spiro, R.G., 1967. The Structure of the Disaccharide Unit of the Renal Glomerular Basement Membrane. *Journal of Biological Chemistry* 242, 4813–4823.
- Stalnaker, S.H., Aoki, K., Lim, J.M., Porterfield, M., Liu, M., Satz, J.S., Buskirk, S., Xiong, Y., Zhang, P., Campbell, K.P., Hu, H., Live, D., Tiemeyer, M., Wells, L., 2011. Glycomic Analyses of Mouse Models of Congenital Muscular Dystrophy. *J Biol Chem* 286, 21180.
- Stalnaker, S.H., Hashmi, S., Lim, J.M., Aoki, K., Porterfield, M., Gutierrez-Sanchez, G., Wheeler, J., Ervasti, J.M., Bergmann, C., Tiemeyer, M., Wells, L., 2010. Site Mapping and Characterization

- of O-Glycan Structures on  $\alpha$ -Dystroglycan Isolated from Rabbit Skeletal Muscle. *J Biol Chem* 285, 24882.
- Stanley, P., Sudo, T., 1981. Microheterogeneity among carbohydrate structures at the cell surface may be important in recognition phenomena. *Cell* 23, 763–769.
- Stayton, P.S., Freitag, S., Klumb, L.A., Chilkoti, A., Chu, V., Penzotti, J.E., To, R., Hyre, D., Le Trong, I., Lybrand, T.P., Stenkamp, R.E., 1999. Streptavidin-biotin binding energetics. *Biomol Eng* 16, 39–44.
- Steenftoft, C., Vakhrushev, S.Y., Vester-Christensen, M.B., Schjoldager, K.T.B.G., Kong, Y., Bennett, E.P., Mandel, U., Wandall, H., Levery, S.B., Clausen, H., 2011. Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat Methods* 8, 977–982.
- Strahl-Bolsinger, S., Tanner, W., 1991. Protein O-glycosylation in *Saccharomyces cerevisiae* Purification and characterization of the dolichyl-phosphate-D-mannose-protein O-D-mannosyltransferase. *Eur J Biochem* 196, 185–190.
- Sugiyama, J., Bowen, D.C., Hall, Z.W., 1994. Dystroglycan binds nerve and muscle agrin. *Neuron* 13, 103–115.
- Syka, J.E.P., Coon, J.J., Schroeder, M.J., Shabanowitz, J., Hunt, D.F., 2004. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *PNAS* 101, 9528–9533.
- Taga, Y., Kusubata, M., Ogawa-Goto, K., Hattori, S., 2012. Development of a Novel Method for Analyzing Collagen O-glycosylations by Hydrazide Chemistry. *Mol Cell Proteomics* 11.
- Tang, M., Wang, X., Gandhi, N.S., Foley, B.L., Burrage, K., Woods, R.J., Gu, Y.T., 2020. Effect of hydroxylysine-O-glycosylation on the structure of type I collagen molecule: A computational study. *Glycobiology* 30, 830.
- Taniguchi-Ikeda, M., Morioka, I., Iijima, K., Toda, T., 2016. Mechanistic aspects of the formation of  $\alpha$ -dystroglycan and therapeutic research for the treatment of  $\alpha$ -dystroglycanopathy: A review. *Mol Aspects Med* 51, 115–124.
- Toledo, A.G., Raducu, M., Nilsson, J., Halim, A., Grahn, A., 2012. O -Mannose and O – N -acetyl galactosamine glycosylation of mammalian  $\alpha$  -dystroglycan is conserved in a region-specific manner 22, 1413–1423.
- Tollefsen, S.E., Kornfeld, R., 1983. Isolation and characterization of lectins from *Vicia villosa*. Two distinct carbohydrate binding activities are present in seed extracts. *Journal of Biological Chemistry* 258, 5165–5171.
- Torres, C.-R., Hart, G.W., 1984. Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. *J Biol Chem* 259, 19–20.
- Trinidad, J.C., Schoepfer, R., Burlingame, A.L., Medzihradsky, K.F., 2013. N-and O-Glycosylation in the Murine Synaptosome. *Molecular and Cellular Proteomics* 12, 3474–3488.

- 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., 2017. *Essentials of glycobiology*. (3rd ed.) Cold Spring Harbor Laboratory Press.
- Vester-Christensen, M.B., Halim, A., Joshi, H.J., Steentoft, C., Bennett, E.P., Lavery, S.B., Vakhrushev, S.Y., Clausen, H., 2013. Mining the O-mannose glycoproteome reveals cadherins as major O-mannosylated glycoproteins. *Proceedings of the National Academy of Sciences* 110, 21018–21023.
- Vocadlo, D.J., Hang, H.C., Kim, E.J., Hanover, J.A., Bertozzi, C.R., 2003. A chemical approach for identifying O-GlcNAc-modified proteins in cells. *Proc Natl Acad Sci U S A* 100, 9116–9121.
- Walker, J.M., 2002. *Protein Protocols Edited by The Handbook SECOND EDITION*.
- Wang, Y., Spellman, M.W., 1998. Purification and characterization of a GDP-fucose:polypeptide fucosyltransferase from Chinese hamster ovary cells. *Journal of Biological Chemistry* 273, 8112–8118.
- Weinmaster, G., Kintner, C., 2003. Modulation of Notch Signaling during Somitogenesis. *Annu Rev Cell Dev Biol* 19, 367–395.
- Whitmore, T.E., Peterson, A., Holzman, T., Eastham, A., Amon, L., McIntosh, M., Ozinsky, A., Nelson, P.S., Martin, D.B., 2012. Integrative analysis of N-Linked human glycoproteomic data sets reveals ptpf ectodomain as a novel plasma biomarker candidate for prostate cancer. *J Proteome Res* 11, 2653–2665.
- Willer, T., Lee, H., Lommel, M., Yoshida-Moriguchi, T., De Bernabe, D.B.V., Venzke, D., Cirak, S., Schachter, H., Vajsar, J., Voit, T., Muntoni, F., Loder, A.S., Dobyns, W.B., Winder, T.L., Strahl, S., Mathews, K.D., Nelson, S.F., Moore, S.A., Campbell, K.P., 2012. ISPD loss-of-function mutations disrupt dystroglycan O-mannosylation and cause Walker-Warburg syndrome. *Nat Genet* 44, 575–580.
- Williamson, R.A., Henry, M.D., Daniels, K.J., Hrstka, R.F., Lee, J.C., Sunada, Y., Ibraghimov-Beskrovnaya, O., Campbell, K.P., 1997. Dystroglycan Is Essential for Early Embryonic Development: Disruption of Reichert's Membrane in Dag1-Null Mice. *Hum Mol Genet* 6, 831–841.
- Wilm, M., 2011. Principles of Electrospray Ionization. *Mol Cell Proteomics* 10.
- Windwarder, M., Altmann, F., 2014. Site-specific analysis of the O-glycosylation of bovine fetuin by electron-transfer dissociation mass spectrometry. *J Proteomics* 108, 258–268.
- Winterhalter, P.R., Lommel, M., Ruppert, T., Strahl, S., 2013. O-glycosylation of the non-canonical T-cadherin from rabbit skeletal muscle by single mannose residues. *FEBS Lett* 587, 3715–3721.
- Won, S.Y., Kim, C.Y., Kim, D., Ko, J., Um, J.W., Lee, S.B., Buck, M., Kim, E., Heo, W. Do, Lee, J.O., Kim, H.M., 2017. LAR-rptp clustering is modulated by competitive binding between synaptic adhesion partners and heparan sulfate. *Front Mol Neurosci* 10, 301013.
- Woo, J., Kim, S.-Y., Kim, H., Kim, E., 2010. Trans-synaptic Adhesions between Netrin-G Ligand-3 (NGL-3) and Receptor Tyrosine Phosphatases LAR, Protein-tyrosine Phosphatase (PTP), and

- PTP via Specific Domains Regulate Excitatory Synapse Formation \* □ S Seok-Kyu Kwon #1. *J Biol Chem* 285, 13966–13978.
- Wright, K.M., Lyon, K.A., Leung, H., Leahy, D.J., Ma, L., Ginty, D.D., 2012. Dystroglycan organizes axon guidance cue localization and axonal pathfinding. *Neuron* 76, 931–944.
- Xiong, H., Kobayashi, K., Tachikawa, M., Manya, H., Takeda, S., Chiyonobu, T., Fujikake, N., Wang, F., Nishimoto, A., Morris, G.E., Nagai, Y., Kanagawa, M., Endo, T., Toda, T., 2006. Molecular interaction between fukutin and POMGnT1 in the glycosylation pathway of  $\alpha$ -dystroglycan. *Biochem Biophys Res Commun* 350, 935–941.
- Yagi, H., Kuo, C.W., Obayashi, T., Ninagawa, S., Khoo, K.H., Kato, K., 2016. Direct mapping of additional modifications on phosphorylated o-glycans of  $\alpha$ -dystroglycan by mass spectrometry analysis in conjunction with knocking out of causative genes for dystroglycanopathy. *Molecular and Cellular Proteomics* 15, 3424–3434.
- Yagi, H., Nakagawa, N., Saito, T., Kiyonari, H., Abe, T., Toda, T., Wu, S.W., Khoo, K.H., Oka, S., Kato, K., 2013. AGO61-dependent GlcNAc modification primes the formation of functional glycans on  $\alpha$ -dystroglycan. *Sci Rep* 3.
- Yaji, S., Manya, H., Nakagawa, N., Takematsu, H., Endo, T., Kannagi, R., Yoshihara, T., Asano, M., Oka, S., 2015. Major glycan structure underlying expression of the Lewis X epitope in the developing brain is O-mannose-linked glycans on phosphacan/RPTP $\beta$ . *Glycobiology* 25, 376–385.
- Yamada, H., Chiba, A., Endo, T., Kobata, A., Anderson, V.B., Hori, H., Fukuta-ohi, H., Kanazawa, I., Campbell, I.P., Shimizu, T., Matsumura, K., 1996. Characterization of Dystroglycan Laminin Interaction in Peripheral Nerve. *J Neurochem* 1518–1524.
- Yao, Y., Norris, E.H., Strickland, S., 2015. The cellular origin of laminin determines its role in blood pressure regulation. *Cellular and Molecular Life Sciences* 72, 999–1008.
- Yoon, J.H., Xu, R., Martin, P., 2013. A Method to Produce and Purify Full-Length Recombinant Alpha Dystroglycan: Analysis of N- and O-Linked Monosaccharide Composition in CHO Cells with or without LARGE Overexpression. *PLoS Curr* 5.
- Yoshida, A., Kobayashi, K., Manya, H., Taniguchi, K., Kano, H., Mizuno, M., Inazu, T., Mitsushashi, H., Takahashi, S., Takeuchi, M., Herrmann, R., Straub, V., Talim, B., Voit, T., Topaloglu, H., Toda, T., Endo, T., 2001. Muscular Dystrophy and Neuronal Migration Disorder Caused by Mutations in a Glycosyltransferase, POMGnT1. *Dev Cell* 1, 717–724.
- Yoshida-Moriguchi, T., Campbell, K.P., 2014. Matriglycan: A novel polysaccharide that links dystroglycan to the basement membrane. *Glycobiology* 25, 702–713.
- Yoshida-moriguchi, T., Willer, T., Anderson, M.E., Venzke, D., Muntoni, F., Lee, H., Nelson, S.F., Yu, L., Kevin, P., 2013. SGK196 Is a Glycosylation-Specific O-Mannose Kinase Required for Dystroglycan Function. *Science* (1979) 341, 1–10.
- Yoshida-Moriguchi, T., Yu, L., Stalnaker, S., Sarah Davis, S.H., Kunz, S., Madson, M., Oldstone, M.B.A., Schachter, H., Wells, L., Campbell, K.P., 2010. O-Mannosyl phosphorylation of alpha-dystroglycan is required for laminin binding. *Science* (1979) 327, 88–92.

Zauner, G., Deelder, A.M., Wuhrer, M., 2011. Recent advances in hydrophilic interaction liquid chromatography (HILIC) for structural glycomics. *Electrophoresis* 32, 3456–3466.

Zauner, G., Koeleman, C.A.M., Deelder, A.M., Wuhrer, M., 2010. Protein glycosylation analysis by HILIC-LC-MS of Proteinase K-generated N- and O-glycopeptides. *J Sep Sci* 33, 903–910.

Zeng, Y., Ramya, T.N.C., Dirksen, A., Dawson, P.E., Paulson, J.C., 2009. High-efficiency labeling of sialylated glycoproteins on living cells. *Nat Methods* 6, 207–209.

Zhang, H., Li, X. jun, Martin, D.B., Aebersold, R., 2003. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* 21, 660–666.

Zhang, Y., Fonslow, B.R., Shan, B., Baek, M.C., Yates, J.R., 2013. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem Rev* 113, 2343.

I acknowledge that I received assistance from Grammarly to proofread thesis in line with the Policy on Transparency in Authorship in PGR Programmes.