-

Liliana Filipa Araújo Ringler

BSc in Biotechnology Engineering

Structural and functional characterization of the human malectin glycan-binding pocket

Dissertation to obtain a Master of Science Degree in Biochemistry for Health

Supervisor: Benedita Andrade Pinheiro, UCIBIO, FCT/UNL

Co-supervisor: Maria Angelina de Sá Palma, UCIBIO, FCT/UNL

December 2020

Liliana Filipa Araújo Ringler

BSc in Biotechnology Engineering

Structural and functional characterization of the human malectin glycan-binding pocket

Dissertação para obtenção do Grau de Mestre em

Bioquímica para a Saúde

Orientador: Benedita Andrade Pinheiro, Researcher, FCT/UNL

Co-orientador: Maria Angelina de Sá Palma, Researcher, FCT/UNL

Júri:

President: Prof. Doutora Teresa Catarino

Arguer: Doutor Jorge da Silva Dias

Vowel: Doutora Benedita Andrade Pinheiro

Faculty of Science and Technology – UNL

December 2020

2020

Structural and functional characterization of the human malectin glycan-binding pocket

Acknowledgments

Este trabalho não seria possível sem a ajuda e o apoio de muitos. Antes de mais quero agradecer à minha orientadora **Doutora Benedita Pinheiro** e à minha coorientadora **Doutora Angelina Palma** pelo apoio durante todo o projeto da tese, pela orientação e pelos conhecimentos transmitidos ao longo de todo este trabalho.

À **professora Doutora Maria João Romão**, líder do grupo de investigação XTAL da FCT/NOVA e Diretora da Unidade de Investigação UCIBIO (Unidade de Ciências Biomoleculares Aplicadas), por permitir desenvolver este projeto no seu grupo de investigação e também pela forma como lidera este grande grupo que me acolheu.

Não posso deixar de agradecer também à **Doutora Marcia Correia** pelo acompanhamento e pelos conhecimentos transmitidos durante toda a parte laboratorial. À **Filipa Trovão** que me ajudou nos ensaios de *microarrays*, e à **Patrícia Borges** que foi a minha companhia durante as tardes mais longas no laboratório. A todo o **grupo XTAL (e GlycoLab)** que me recebeu num ambiente propício para o meu desenvolvimento científico, e do qual levo muitos conceitos profissionais.

À **Fundação para a Ciência e Tecnologia - Ministério da Ciência, Tecnologia e Ensino Superior** pelo financiamento no âmbito do projecto PTDC/BIA-MIB/31730/2017 e da Unidade de Investigação UCIBIO (UID/Multi/04378/2019); e pela bolsa de investigação de 3 meses no âmbito da Escola de Verão UCIBIO, em colaboração com a Direção Geral do Ensino Superior (DGES). Ao **professor Doutor Eurico Cabrita** pela coordenação de todo o procedimento concursal.

À **Professora Doutora Teresa Catarino**, pela coordenação do mestrado de Bioquímica para a Saúde. E às instituições (ITQB-NOVA e FCT-NOVA) que desenvolvem este curso.

Quero agradecer à minha **tia Di** que sempre me apoiou na vida académica, com genuíno interesse neste projeto e no meu sucesso. Obrigada por tudo Di!

À minha **Nancy** que foi a minha companhia por muitas horas (mais do que o habitual), que me deu muitos mimos e me confortou umas quantas vezes.

Às **duas Estrelinhas** lá em cima, avô sei que estás muito orgulhoso! Aos meus **avós Artur e Constância**. A toda a minha família, aos meus primos pequeninos que me fazem acreditar num mundo melhor. Adoro-vos a todos!

Ao **José** por estar sempre do meu lado, por ouvir as minhas reclamações e não me deixar ir abaixo, nem que seja através de uma brincadeira tola que me faça sorrir. Obrigada, contigo do meu lado tudo é mais fácil. Ao senhor **Vítor** e à dona **Carmita** por torcem muito por mim.

Tenho de agradecer também à minha irmã **Deby** e ao meu sobrinho **Biel**, que por muitas vezes foram o motivo do meu sorriso, a minha alegria. Tenho muitas saudades vossas! Voltem logo! Amo-vos.

E ao meu maior apoio, ao meu pilar, aos **meus pais**. Sem eles do meu lado, eu não seria capaz. Obrigada mãe por todos os "abracinhos" quando mais precisei, por todos os mimos e por todo o amor. E obrigado pai pelo exemplo de força que és, por todos os valores transmitidos, pela atenção em cada conversa e pelo excelente pai que és. Amo-vos muito!

Abstract

The N-glycosylation pathway has several chaperones, enzymes and complexes dedicated to protein quality control and correct folding of proteins. Malectin is a membrane-anchored endoplasmic reticulum (ER) protein, which recognizes a Glc-2-*N*-glycan intermediate of the early steps of this pathway in nascent proteins. The high conservation of malectin in animals reveals the importance of this lectin in the ER, and recent studies point to its involvement in the ER protein quality control. The X-ray crystal structure of malectin-ligand complex revealed the molecular interaction with the D1 arm of the di-glucosylated high mannose N-glycan (Glc2Man3).

In this work, we aimed to characterize the interaction of malectin binding-pocket with its natural ligand. Therefore, to understand the molecular determinants for this highly specific interaction, we designed five specific mutants (E102A, Y104F, Y131F, D201A and Y199AY200A). Also, looking for a new binding specificity to O-glucosylation epidermal growth factor like repeats, we designed a mutation of a residue that sits in the bottom of the binding-pocket (S80Q). Structural stability studies of the malectin mutants were carried out using Differential Scanning Calorimetry analysis and their effects in the interaction with glucosylated ligands were accessed by a first screening in a polysaccharide microarray. We also analysed the binding of the wild-type malectin and the specifically designed mutants to the Glc2Man3 using the NGL-based oligosaccharide microarray. Comparison of the six proteins revealed the importance of the residues that establish interactions between the protein and the glycan. The objective of developing a protein with a new specificity was not achieved during this framework. However, the process led to the confirmation that the malectin binding pocket has adapted to perfectly accommodate the two glucose rings from the di-glucosylated N-glycan.

An important outcome from this study was to clarify the importance of several residues in the malectin specificity and, therefore, to provide more information that will contribute to the clarification of the function of this protein involved in the N-glycosylation pathway.

Keywords: Malectin, Endoplasmic reticulum, N-glycosylation, Glycobiology, carbohydrate microarrays.

Resumo

A via da N-glicosilação tem vários '*chaperones'*, enzimas e complexos dedicados ao controlo de qualidade e ao enrolamento correto das proteínas. A malectina é uma proteína presente na membrana do reticulo endoplasmático, que reconhece nas proteínas nascentes o intermediário Glc-2-N-glicano, existente nas primeiras etapas desta via. Estudos recentes indicam que a malectina está envolvida no processo de controlo de qualidade do reticulo endoplasmático. Outro aspeto que reforça a importância desta lectina na via da N-glicosilação é a sua forte presença no reino animal. A estrutura de raios-X da malectina em complexo com o seu ligando revelou as interações moleculares com o 'D1 arm' di-glucosilado do N-glicano (Glc2Man3).

O presente trabalho teve como principal objetivo caracterizar as interações do sítio de ligação da malectina com o seu ligando natural, pelo que, para identificarmos os determinantes moleculares desta interação altamente específica, desenhámos cinco mutantes específicos (E102A, Y104F, Y131F, D201A e Y199AY200A). O segundo objetivo deste trabalho foi desenhar uma nova proteína específica para a o-glucosilação presente em domínios tipo fatores de crescimento epidérmico (EGF-like). Nesse intuito realizámos uma mutação num resíduo localizado no fundo do sítio de ligação (S80Q). Os estudos estruturais de estabilidade dos mutantes foram desenvolvidos através da análise de Calorimetria Diferencial de Varrimento e os seus efeitos na interação com ligandos glucosilados foram obtidos através de um *screnning* de *microarrays* de polissacarídeos. Recorrendo à tecnologia de *microarrays* de oligossacarídeos NGL realizámos a análise da interação da malectina nativa e dos mutantes com o Glc2Man3. A comparação das proteínas revelou a importância dos resíduos que estabelecem interações entre a malectina e o glicano. Durante este trabalho não foi possível alcançar o segundo objetivo, no entanto, os resultados permitiram confirmar que o sítio de ligação da malectina adapta-se de forma a acomodar os dois anéis de glucose do N-glicano di-glucosilado.

Os resultados obtidos neste estudo permitiram clarificar a importância de diversos resíduos que conferem a especificidade da malectina e, assim, contribuir com informação que irá auxiliar no esclarecimento da função desta proteína envolvida na via da N-glicosilação.

Palavras-chave: Malectina, Reticulo endoplasmático, N-glicosilação, Glicobiologia, Microarrays de hidratos de carbono.

Table of contents

Index of Figures

Index of Tables

Abbreviation and Symbols

ALG	Asparagine-linked glycosylation
C. thermocellum	*Clostridium thermocellum*
CNX	Calnexin
CRT	Calreticulin
Dol-P	Dolichol-Phosphate
ERAD	Endoplasmic reticulum associated degradation
GBPs	Glycan binding proteins
GH	Glycosyl hydrolase
GI	Glucosidase I
GII	Glucosidase II
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IMAC	Immobilized metal affinity chromatography
NGL	Neoglycolipid
Ni-NTA	Nickel-nitriloacetic acid
NECD	Notch extracellular domain
NICD	Notch intracellular domain
NRR	Negative control region
OST	Oligosaccharyltransferase
UGT1	Uridine diphosphate glucuronosyltransferase 1
Z. galactanivorans	*Zobellia galactanivorans*

Chapter 1 – Introduction and Objectives

In this chapter, I will make a brief introduction of the glycobiology and the protein under study. Also, I will describe the microarray technique, used on the protein-ligand interaction analyses of the human malectin.

1.1 Glycosylation – An overview

Carbohydrates are the biomolecules most abundant on Earth. These are divided into three major sizes classes i) monosaccharides – a single polyhydroxy aldehyde; ii) oligosaccharides – chains of monosaccharide units, or residues linked by glycosidic bonds; and iii) polysaccharides – more than 20 monosaccharide units. These carbohydrates can act as signals when attached to proteins or lipids and are named as glycoconjugates. These glycoconjugates determine the intracellular location or metabolic fate of the bound molecule, revealing that glycans contribute to numerous biological functions in animal systems (Cox & Nelson, 2004).

When proteins are modified by the addition of carbohydrates components, they form glycoproteins. This post-translation modification is called glycosylation and can have a big microheterogeneity (Moremen et al., 2012). Glycosylation is a dynamic process with a highly regulated mechanism that plays a critical role in determining the protein structure, function and stability (J. Arey, 2012). In eukaryotes, this process takes place in the endoplasmic reticulum (ER) and Golgi apparatus, starting through the recognition of short amino acid sequences on glycan-acceptor polypeptide (sequons) by glycosyltransferases and subsequent steps of sugar transfer by transferases and cleavage by glycosidases. These series of steps are controlled by substrate availability, enzyme activity and location within the organelles, and levels of gene transcription. Protein glycosylation is a ubiquitous post-translational modification found in all domains of life and can be classified according to the type of linkages (Moremen et al., 2012; Reily et al., 2019). In this work, we will focus on the two major types of additions: the N-linked glycans and the O-linked glycans.

A few years ago, it was estimated that most sequons in proteins are glycosylated and more than half of all proteins are glycoproteins. By combining this information with the discovery that the human genome encodes close to 50,000 proteins, the importance of post-translational modifications is emphasized. The glycan structures in newly synthesized glycoproteins are crucial for the secretion of proteins, due to their influence on protein folding, on the supply of ligands for lectin chaperones, on the contribution to the monitoring of quality

control in the ER and mediation of transit and protein target selection along the secretory pathway. Interestingly, most human proteins contain N-glycan structures. (Apweiler et al., 1999).

1.1.1 N-glycosylation pathway

In the beginning, it was not known why the cells went through the entire process of synthesising the oligosaccharide, to then be cleaved and build it up again with different sugars. Taking to account the final product, it seemed like an unnecessary process. However, it was realised that for the glycoprotein folding to be correct, the N-glycan must have different configurations at each stage of the pathway.

The first phase of the N-glycosylation pathway occurs in the ER, where several steps to provide proper protein folding and quality control are taking place. The second phase is an intracellular process of transport and enzymatic targeting. After extensive modification in the Golgi, the third phase occurs with the arrival of the mature protein in the targeted extracellular space. The mature glycoprotein carries N-glycans that depend on the expressed glycosylation genes and physiological state of the cell type in which the glycoprotein is made. There are three major classes of N-glycans, that share a common core sequence and are classified in oligomannose, complex or hybrid (**Figure 1.1**; Helenius & Aebi, 2004; Stanley P, Taniguchi N, 2017).

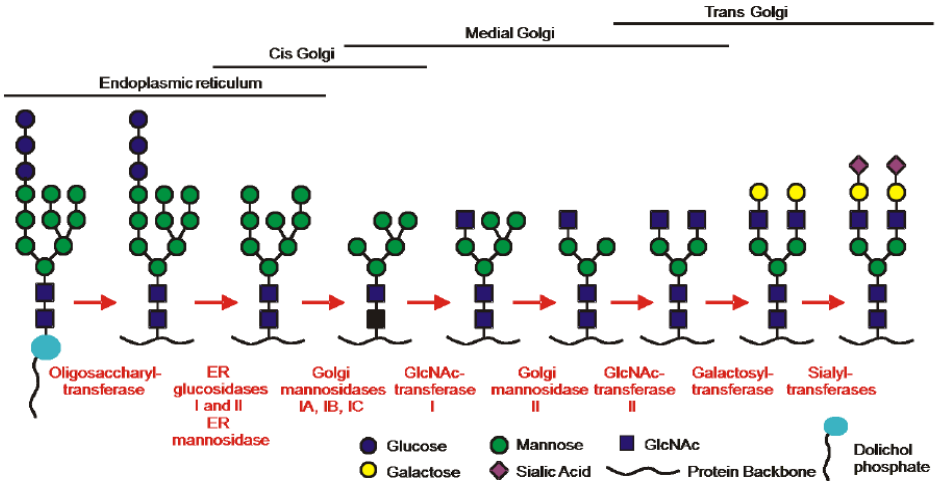


Figure 1.1 – Overall representation of the biosynthesis of N-linked glycans. The oligosaccharide assembled on Dolichol-phosphate is transferred by oligosaccharyltransferase to Asn in selected Asn-X-Ser/Thr sequons. Glycosidases and Glycosyltransferases process the N-glycan in the lumen of the ER, and that continues in the Golgi in a species-, cell type-, protein-, and even site-specific manner. Adapted from Haltiwanger, 2007.

N-glycan biosynthesis

At this stage, the N-glycan structures consist of 14 carbohydrate residues: three glucose (Glu), nine mannoses (Man), and two N-acetylglucosamines (GlcNac). For a long time, it was not clear why this structure is first built up in the asparagine pathway, to then to be trimmed in the N-glycosylation pathway. It was necessary for several studies and new findings to understand that its role is crucial to provide the regulation of the protein folding.

This first step of the N-linked core oligosaccharide biosynthesis is initiated at the cytoplasm of the endoplasmic reticulum with the assembling on the lipid carrier dolichol phosphate. The tetradecasasaccharide substrate is build up by the addition of GlcNac-phosphate to dolichol-phosphate (Dol-P). Through a series of glycosyltransferases encoded by asparagine-linked glycosylation (ALG) genes a second GlcNac residue and five mannoses are attached and after translocation into the ER lumen by RFT1 four additional mannose and three glucose residues are added (Breitling & Aebi, 2013; J. Helenius et al., 2002). The terminal α-1,2 linked glucose residue is the trigger for the recognition by the oligosaccharyltransferase (OST). The OST transfer the N-glycan structure in a bloc to be covalently attached to a protein by a GlcNacβ1-Asn linkage.

Glycoprotein Folding

OST is a multi-subunit protein complex in the ER membrane that catalyses the transfer of the oligosaccharide from Dol-P-P to Asn-X-Ser/Thr by an N-glycosidic bond (A. Helenius & Aebi, 2004; Stanley P, Taniguchi N, 2017). Since OST scans the emerging sequence for glycosylation consensus sequences, its active site must be located close to the exit from translocon pore, but at the same time accessible to luminal chaperones and folding enzymes (Kowarik et al., 2002).

The folding begins with the passage of the polypeptide chain through the translocon complex (Kowarik et al., 2002). Close to the OST is glucosidase I (GI), which is a type II membrane protein of the glycosyl hydrolase (GH) family 63. This α1,2 exo-glucosidase acts immediately after the addition of the oligosaccharide onto the nascent polypeptide chain removing co-translationally the terminal glucose of $Glc_3Man_9GlcNac_2$ (Aebi et al., 2010; A. Helenius & Aebi, 2004). The structure generated by GI, $Glc_2Man_9GlcNac_2$, is also trimmed by glucosidase II (GII), which removes the two remaining glucose residues (Trombetta, 2003). GII is an α1,3 exo-glucosidase, belongs to a luminal member of the GH family 31, and contains a catalytic α-subunit and a regulatory β-subunit (Aebi et al., 2010; A. Helenius & Aebi, 2004). The regulation of GII function is not fully understood. Still, the trimming by this soluble

heterodimer is a carefully regulated process which provides cells with a sophisticated mechanism to modulate the entry of nascent and newly synthesised glycoproteins to the Calnexin/Calreticulin cycle (Cnx/Crt cycle; Deprez et al., 2005).

Calnexin (CNX) is a type I transmembrane protein and calreticulin (CRT) is its soluble homolog. Both are composed of a globular carbohydrate-binding domain, which folds into a leguminous (L)-type lectin-like β-sandwich structure with a proline-rich P-domain that recruits the lectin-associated oxidoreductase ERp57 (Aebi et al., 2010). CNX and ERp57 associate with nascent glycoproteins in a complex to assists in di-sulphide bond formation, which is essential for proper folding (A. Helenius & Aebi, 2004; Lederkremer, 2009). The mono-glucosylated core glycan binds to Cnx/Crt cycle preventing aggregation and export of the incompletely folded chains from the ER. Upon the release of the folding polypeptide from the cycle, GII removes the innermost glucose (A. Helenius & Aebi, 2004). The end of this step allows the release of the glycoprotein to leave the ER unless recognition by UDP-Glc: glycoprotein glucosyltransferase (UGT1 or UGGT).

Glycoprotein refolding attempts and targeting to ERAD

Throughout the process, the glycans serve as sorting signals to reflect the folding status of the protein and consequently their fate.

UGT1 is a member of the glycosyltransferase family 24 and has a dual function of "inspecting" polypeptides that display $Man_9GlcNac_2$ oligosaccharides and to re-glucosylating exclusively the terminal mannose in polypeptides that have an incomplete fold. The re-glucosylated high mannose promotes the rebind of glycoprotein to Cnx/Crt cycle, for proper folding or degradation (Aebi et al., 2010; A. Helenius & Aebi, 2004).

The crucial decision is when to abandon refolding attempts through the Cnx/Crt cycle and to target the glycoprotein to ER-associated degradation (ERAD). During the off phase, N-glycans are exposed to the ER-resident α1,2-mannosidases that can remove the mannose residue to provide interruption of folding attempts and ERAD mechanism (A. Helenius & Aebi, 2004; Lederkremer, 2009).

1.1.2 O-glycosylation pathway

Although it is also a protein modification process, O-glycosylation is quite different from N-glycosylation. The entire process occurs in Golgi, and the O-glycan biosynthesis does not require a lipid-linked oligosaccharide precursor.

The initiating event is the addition of a GalNac residue from UDP-GalNac to a serine or threonine by a polypeptide GalNac transferase (GalNAcT), resulting in an O-linked oligosaccharide (Varki et al., 2009). The amino acid sequence of the protein that serves as a substrate has not been identified yet but is clear that it is the regulatory mechanism (Brockhausen & Stanley, 2015). The O-GalNAc glycan is usually extended to form one of four common structures, and each core can subsequently be extended to give a mature linear or branched glycan. The class of glycoproteins carrying the most significant number of O-GalNAc glycans is mucins, and abnormal mucin O-GalNAc glycans are associated with many diseases (Brockhausen & Stanley, 2015).

1.1.3 Epidermal Growth Factor (EGF) repeats

The glycosylation process can also modulate interactions between proteins (Varki, 2017), and the density of glycans can influenced the interactions between glycans and glycan-binding proteins. This glycosylation-dependent binding is found in some growth factor receptors (Haltiwanger et al., 2017), as it seems to happen in epidermal growth factors (EGF).

EGF is a small growth factor (~40 amino acids) that are characterized by six conserved cysteine residues that form three disulphide bonds (Haltiwanger et al., 2017; **Figure 1.3.A**). Most of the times, they are found repeated in secreted proteins and the extracellular domain of transmembrane proteins to provide cell motility, proliferation, differentiation, and survival (Haltom & Jafar-Nejad, 2015). EGF repeats are found in signalling receptors, cell adhesion molecules, plasma proteins, and extracellular matrix, with functions that involve protein trafficking and mediation of protein-protein interactions (Haltom & Jafar-Nejad, 2015). A good example of these interactions is cell-cell communication via Notch.

Notch signalling pathway

In multi-cellular organisms, neighbouring cells need to communicate with each other to ensure proper cell fate decisions and differentiation. Notch signalling is a cell-cell signalling pathway that is required for proper development and homeostasis (Haltom & Jafar-Nejad, 2015). The process of signalling occurs when the transmembrane ligand from one cell binds to the transmembrane Notch receptor from the neighbouring cell.

Notch receptors and ligands are large, single-pass transmembrane proteins, and its underlying molecular structure is well-conserved from flies to humans (Takeuchi & Haltiwanger, 2014). Therefore, several studies to understand the signalling pathway were

developed in *Drosophila*, which has only one Notch receptor and two ligands (Delta and Serrate; Bray, 2016). Unlike flies, mammals have 4 Notch receptors (Notch1-4) and five canonical ligands: Jagged1, Jagged2, Delta-like (DLL) 1, DLL3 and DLL4 (Haltom & Jafar-Nejad, 2015). The Notch receptor consist of a large extracellular domain (NECD), a transmembrane region, and a large intracellular domain (NICD; Takeuchi & Haltiwanger, 2014). The majority of the NECD receptors have up to 36 EGF repeats (Haltom & Jafar-Nejad, 2015).

For signalling to occur, the ligand-mediated activation induces the unfolding of the juxta-membrane negative control region (NRR) to provide accessibility to ADAM10. This protease and γ-secretase acts on cleavage in the members of the Notch family receptor, S2 and S3, respectively, releasing the Notch intracellular domain (NICD) into the cytoplasm. Once released, the NICD enters the nucleus where it can promote transcription of the target genes (**Figure 1.2**; Haltom & Jafar-Nejad 2015; Bray, 2016).

Mutation of the transmembrane Notch receptor and other components of this pathway could lead to congenital disorders and other diseases (Takeuchi & Haltiwanger, 2014). Due to the function of the Notch, activating or inactivating mutations play a role in tumour suppressor or tumour promoter (Takeuchi & Haltiwanger, 2014). That is the reason it became a molecular target for a variety of cancers.

The glycosylation process of the EGF occurs in the ER and Golgi apparatus, where the glycosyltransferases progressively act to add glycans to the NECD (Takeuchi & Haltiwanger, 2014). In this process, a subset of EGF repeats is covered by a variety of specific O-glycans defined by the consensus sequences (Stanley & Okajima, 2010). Three unusual types of O-linked glycosylation (O-Fucose, O-Glucose and, more recently, O-GlcNac) were found on EGF repeats and play an essential role in the function of Notch receptors. The consensus site for O-linked fucose and O-linked glucose is mostly present in Notch1 (Moloney et al., 2000b). Some studies reveal that O-linked fucose and its elongation are crucial for Notch signalling activation by a modulatory step that occurs before the proteolytic cleavages of NICD (Moloney, Panin, et al., 2000a). However, the scope of this project focuses only on the modification with O-glucose glycans, which also showed to be essential for Notch activity in mice and flies (Fernandez-Valdivia et al., 2011).

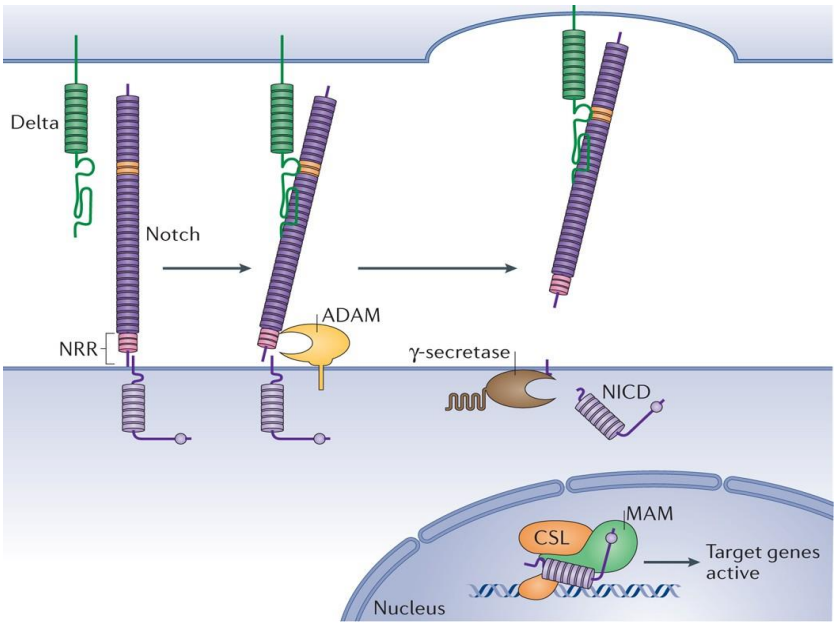


Figure 1.2 – Ligand binding leads to exposure of the cleavage site in Notch. Summary of core pathway: when canonical Notch ligands (green) bind to Notch receptors (purple; pink indicates the negative regulatory region (NRR)) on the adjacent cell surface, they prompt two proteolytic cleavage events, the first by ADAM10 and the second by γ-secretase, releasing the Notch intracellular domain (NICD). In the nucleus, NICD interacts with other proteins to promote gene transcription. Taken from Bray 2016.

O-Glucosylation on EGF repeats

The first modification on the EGF repeats by the addition of the O-glucose was discovered in bovine blood coagulation factors VII and IX (Hase et al., 1988). Later the same type of linkage was found on human factors VII and IX, protein Z (H. Nishimura et al., 1989), thrombospondin (H. Nishimura et al., 1992), and murine fetal antigen-1/delta like protein (FA-1-DLK; Krogh et al., 1997). This modification of the NECD EGF repeats occurs when O-glucose is added to serine residues of the reviewed consensus sequence C^1-X-S-X-P/A-C^2. However, this glycosylation is also site- and cell type-dependent (Rana et al., 2011; Takeuchi et al., 2012). In Drosophila Notch, 18 out of 36 EGF repeats contain the consensus sequence for O-glucose, and all were found to be O-glycosylated (Leonardi et al., 2011). Whereas in mammalian Notch1, the efficiency of O-glycosylation of some consensus sequences is affected by the folding stage of the EGF repeats and the composition of the amino acids surrounding (Rana et al., 2011; Takeuchi et al., 2012).

The Notch1 O-glucose glycans were mostly found as a linear trisaccharide by the addition of two α1–3 linked xylose residues, to form xylose-xylose-glucose-O-EGF (Moloney et al., 2000b). The protein O-glucosyltransferase 1 (POGLUT1) is the mammalian homolog of Rumi, the transferase capable of adding glucose to serine residues in Drosophila Notch

(Acar et al., 2008). In human Notch, the elongation of this type of linkages is regulated by three enzymes: glucosidase xylosyltransferases (GXYLT)1 and GXYLT2 add the first xylose (Sethi et al., 2010), while xyloside xylosyltransferase (XXYLT1) adds the second (Sethi et al., 2012; **Figure 1.3.B**). Leet and coworkers suggest that this uncommon terminal of xylose residues plays a reasonably specific role in negatively regulating the Notch signalling(Lee et al., 2013). Also, Notch signalling is affected in a temperature-sensitive manner by the loss of Rumi (Acar et al., 2008) which is in agreement with the suggestion that glucose residues on Notch contribute to maintaining robust signalling as the temperature increase (Leonardi et al., 2011).

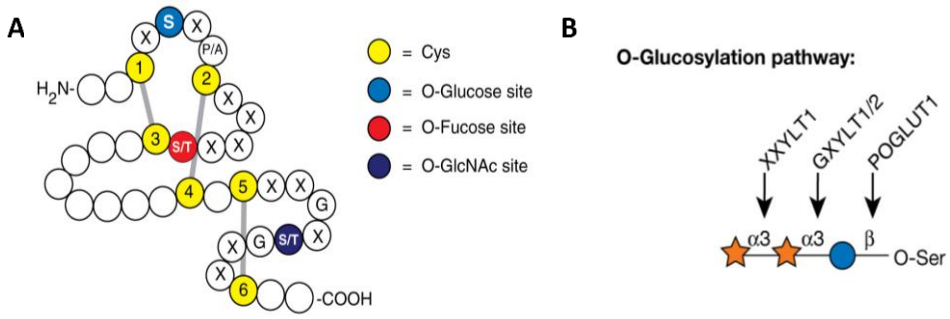


Figure 1.3 - Modifications of epidermal growth factor (EGF)-like repeats. (*A*) EGF-like repeats can be modified by O-fucose, O-glucose, and O-GlcNAc. A schematic representation of an EGF-like repeat is shown: (*yellow*) conserved cysteine residues; (*grey lines between the cysteines*) disulfide-bonding pattern. The modification sites for O-glucose (*blue*), O-fucose (*red*), and *O*-GlcNAc (*dark blue*) are shown in the context of the putative consensus sequences for each. S, serine; T, threonine; P, proline; A, alanine; G, glycine; X, any amino acid. (*B*) EGF-like repeat O-glucose and the enzymes responsible for the addition of each sugar. Adapted from Haltiwanger et al., 2017.

In 2011, the Haltiwanger group revealed that Rumi uses UDP-glucose and UDP-Xyl as a donor substrate. They also suggest that the efficient transfer of xylose requires the presence of a second serine in the consensus sequence, C^1-X-S-S-P-C^2 (Takeuchi et al., 2011). However, recently, a structural study of POGLUT1/Rumi shown that this protein can assume two local conformational stages and that the type of O-glycosylation depends only on the concentration of substrate available (UDP-glucose and UDP-xylose) in the ER (Li et al., 2017). Also, it was found a new consensus sequence that contains an Asn at the Xb position, C^1-X-S-N-P-C^2, demonstrating that POGLUT1 has the potential to xylosylate many more EGF-like domains than previously thought (Li et al., 2017).

1.2 Discovery of animal lectins

Lectins are carbohydrate-binding proteins. They were first discovered more than 100 years ago in plants, but only in 1982 were reported the first direct evidence for a mammalian lectin. At that time, several studies and discoveries were addressed and the concept of vertebrate lectins that could recognize specific endogenous ligand had become firmly established. The studies also uncover the discovery of several soluble lectins circulating in the blood plasma of various species with varied carbohydrate-binding specificities. Most recently, several lectins have been discovered within the ER-Golgi pathway itself. After the animal lectin discovery, they were classified according to their higher carbohydrate sequences interaction. Only with evolution of methodologies to study the proteins, a more consistent classification based on amino acid sequence homology and evolutionary relationship emerge (Brewer & Dam, 2000). Currently, several categories and families of lectins have emerged.

Several structures of animal lectins with their cognate ligands have been revealed, leading to some principles that have emerged. First, the binding sites are of relatively low affinity and are mostly found in shallow indentations on the surface of the proteins. Second, selectivity is mostly achieved through hydrogen bonds with the saccharide and by van der Waals' packing of the hydrophobic face of monosaccharide rings against aromatic amino acid side chains. Additional contacts between the saccharide and the protein (like water-network or divalent cations) can provide further selectivity. Finally, the actual region of contact between the saccharide and the polypeptide typically involves only one to three monosaccharide residues (Brewer & Dam, 2000).

1.3 Malectin – A lectin involved in the N-glycosylation pathway

The N-glycosylation pathway is a very dynamic and multi-stage process. Over the years, several hypotheses have been proposed (Hirschberg & Snider, 1987; Hubbard & Ivatt, 1981; Kornfeld & Kornfeld, 1985) for understanding the overall pathway, but there are still details to clarify. The more recently discovered malectin protein may help to define the early steps of N-glycosylation but it is still subject to many studies and hypotheses of function.

In 2008, Schallus and coworkers identified a new type I membrane-anchored ER protein which they called “Malectin” due to the first nuclear magnetic resonance (NMR) screening results showed binding to maltose and related oligosaccharides. The protein was first identified in the *Xenopus laevis* pancreas but became more surprising when it was found

to be widely distributed in various tissues and well conserved in animals. The evidence points to be a vital lectin in providing proper protein folding and, playing a role in N-glycosylation quality control. Also, structural and ligand-recognition studies demonstrated that malectin is a novel carbohydrate-binding protein with a remarkable and selective binding to Glc_2-*N*-glycan ER intermediate (Mamidi & Surolia, 2015; Schallus et al., 2008).

1.3.1 A novel carbohydrate-binding protein

The discovery of malectin gene in *X. laevis* pancreas led to new other relevant findings of this protein. First, it was observed that malectin mRNA is expressed in several tissues throughout development. It was detected in *X. laevis* oocyte and in tissues like heart, kidney, lung, intestine, muscle, pancreas, and stomach (Schallus et al., 2008). The database search indicated that malectin protein was annotated to a hypothetical human gene KIAA0152, and the bioinformatics tool suggested homology between these two malectin proteins (Schallus et al., 2008).

Malectin is an ER lumen-resident lectin conserved in the animal kingdom that has a highly selective binding to Glc2Man9GlcNac2, an intermediate of N-glycosylation. The malectin specificity to di-glucosylated glycans is very tolerant towards the linkage of the glucosidic bonds- although the highest binding event occurs with the recognition of Glcα1-3Glc (nigerose). Microarrays analysis showed that malectin also interacts with Glcα1-4Glc (maltose) and Glcα1-6Glc (isomaltose) linkages (Palma et al., 2010). Therefore, malectin allows other glucosidic linkages that change the glycan form, placing the polar OH groups of the second glucose in different spatial positions concerning the first glucose residue (Schallus et al., 2008 & 2010).

1.3.2 Malectin structure and ligand-bound conformation

Malectin structure

Protein domain databases such as SMART (Letunic et al., 2006) and Pfam (Finn et al., 2006) predicted malectin as a type I membrane protein with an N-terminal signal peptide (AA 1-26), a C-terminal transmembrane helix (AA 255-274) and a highly conserved globular domain of ~190 residues (Schallus et al., 2008). In some organisms, like the human, this globular domain precedes a highly charged sequence that spans eight contiguous glutamate residues (**Figure 1.4**; Palma *et al.*, 2010). After this prediction, Schallus and coworkers assigned the entire *Xenopus leavis* malectin structure in its free and ligand-bound form

(Schallus et al., 2008). The structure is composed of the predicted globular segment (AA 28-201), two elongated β-sheets packed onto each other and three short α-helices that flank the core β-sandwich (Schallus et al., 2008). Four loops (L1-L4) located at only one side of the domain are the carbohydrate-binding region (G62-G68, T86-N90, E114-A118, and Y185-N187; Schallus et al., 2010).

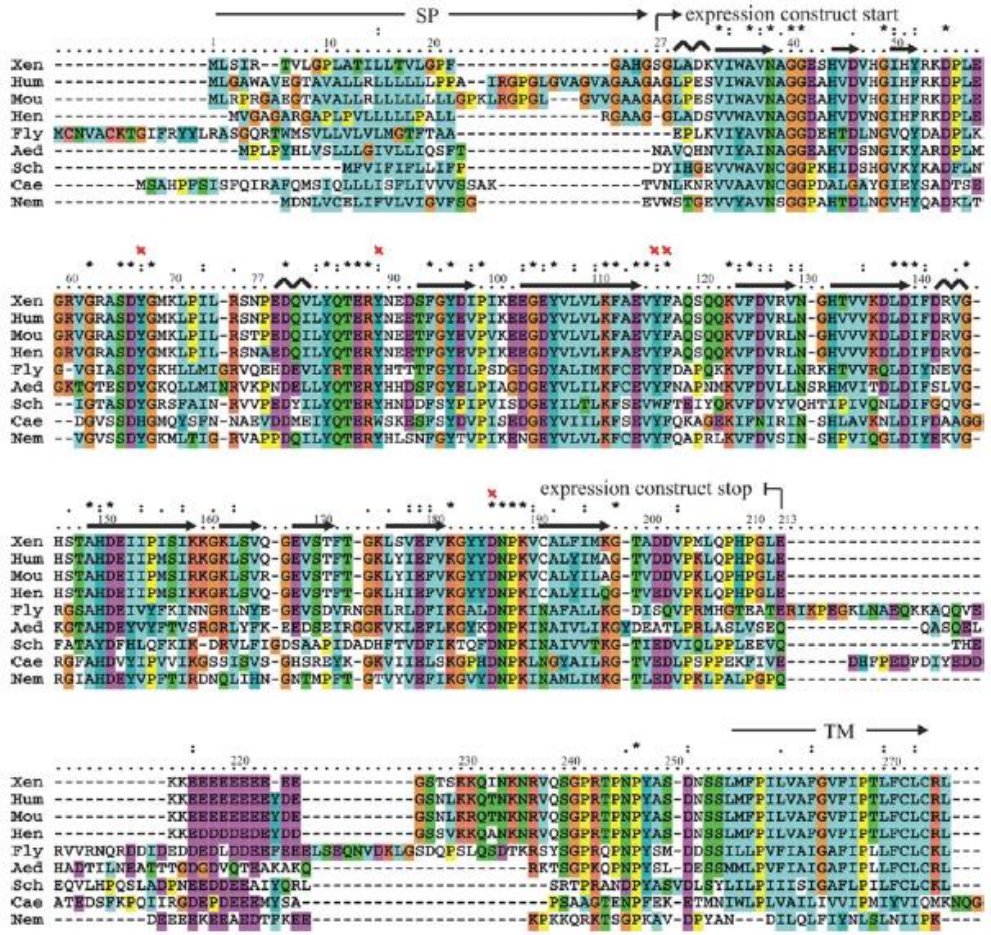


Figure 1.4 – Sequence alignment of malectin proteins in animals. Malectin proteins are composed of an N-terminal signal peptide (SP, AA 1-26), a C-terminal transmembrane helix (TM; AA 255-274) and a highly conserved central part of 190 residues followed by an acidic, glutamate-rich region. Taken from Schallus *et al*., 2008.

Most recently, Pinheiro *et al.* determined the X-ray structure of human malectin bound to the pentasaccharide Glc2Man3 (at 2.0 Å), and obtained three other 3D-structures of malectin complexed with Glc4 (at 1.4 Å), nigerose (1.6 Å) and maltose (2.1 Å) to obtain precise structure insight (Pinheiro el at., unpublished). All complexes have a strong structural relationship to the NMR malectin structure assigned by Schallus, with the characteristic paired β-sandwich domain formed of antiparallel β-strands evidenced by a rmsd of 1.7 Å for 166 alpha carbons. However, the superposition of the NMR structures and human malectin

crystallography structure revealed a few structural differences (**Figure 1.5**). These new structures uncover four short α-helices and a considerable divergence in loops 1 and 2 compared with the loops reported in *Xenopus leavis* malectin structure. The α-helices 1 (Asp81-Leu86) and 2 (Gln96-Thr101) are located between β-strands 4 and 5 (Arg78-Ser80 and Glu102-Glu105, respectively), which form a small antiparallel β-sheet. The α-helix 3 stretches between β-strands 9 (Val149-Leu153) and 10 (His165-Arg174), and the α-helix 4 precedes the C-terminal. Interestingly, the most relevant differences arise around the binding pocket, more specifically in loop 1 (Leu72-Asp81) and loop 2 (Glu102-Thr108). Loop 1 is part of a longer stretch that includes β-strand 4 and α-helices 1. The position of this first loop diverges by more than 11 Å from the corresponding loop in the NMR structure, and appears to force loop 2 to be closer to the ligand by almost 4 Å (**Figure 1.5.A, B**). Also, a structural atom was identified, and the anomalous electron density maps were attributed to the presence of a magnesium atom coordinated in an octahedral geometry by the Asn53, Ala54, Leu98, Thr101 and Cys206 residues.

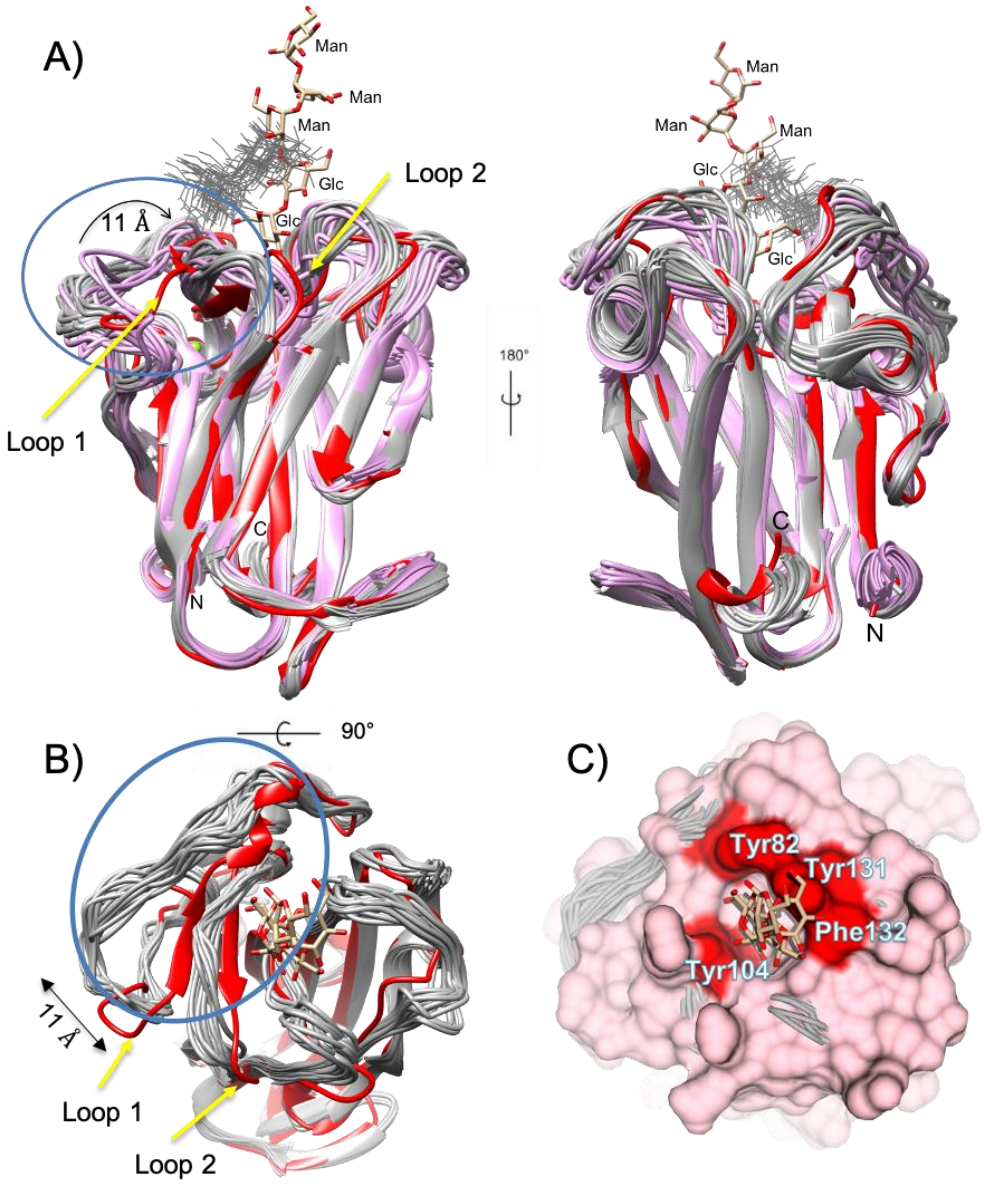


Figure 1.5 – Superposition of X-ray structure of human malectin-Glc2Man complex (in red ribbon) and NMR structures of *X. leavis* malectin isolated (in purple ribbon) and bound to nigerose (in gray ribbon). The magnesium ion identified in the X-ray structure is depicted as green sphere. The Glc2Man3 ligand is shown in stick and the nigerose models are represented in wire. (*A*) The conserved fold of the three structures, with the difference in loop 1 locations highlighted by a blue circle. *(B)* The top view of the binding site, which highlights the proximity of loops 1 and 2 (pointed by yellow arrows) with the ligand in the human malectin structure. *(C)* The binding pocket surface with the aromatic residues Y82, Y104, Y131 and F132 shown in red. Taken from Pinheiro *et al.*, unpublished.

When searching for similar structure, the malectin arrangement showed that its luminal part is quite like carbohydrate-binding modules (CBMs) of bacterial glycosylhydrolases. As already mentioned by Schallus, binding via the β-sheet connecting loops is untypical for β-sandwich animal lectins, it only occurs in the CBM families 35

(CBM35) and 6 (CBM6). The search by Palma and colleagues, displayed two structures of enzymes from these CBM families that match with human malectin (*Clostridium thermocellum* with a rmsd of 2.59 and *Zobellia galactanivorans* with a rmsd of 2.83, respectively), both showed the typical β-sandwich structure. The structure from *C. thermocellum* has diverse binding specificities due to the variable loop site (VLS) and, the structure from *Z. galactanivorans* has two binding sites, an VLS and a concave face site (CFS) with a more restricted binding profile. The homology with these two structures and the fact that malectin exhibits aromatic residues facing the solvent in both β-sheets led the authors to suggest that one or both faces of the β-sandwich can constitute an extended interface for recognition of the N-glycan chain. The two putative sites, identified as site A and B, showed some conserved residues, and the site B was found to be like one of the ligand binding sites present in CBM6 family. This was the first inkling that there might be an additional recognition of the ligand. Two tyrosines (Y199 and Y200) in B site were noticed to be close to the binding pocket, so it was proposed that these aromatic rings are probably involved in the recognition of other arms of the N-glycan.

Malectin binding event

As stated in the previous section, the carbohydrate-binding region contains four long loops. In 2010, when Schallus and colleagues assigned in the xenopus malectin four aromatic rings located in the binding cleft interacting with the ligand: three tyrosines and one phenylalanine. Using a mutational strategy and recording ^{1}H-NOESY spectra for the free and maltose-bound forms they could observe that Y67 and Y89 contacted the terminal non-reducing glucose of maltose, F117 contacted the reducing glucose and Y116 interacted with both glucoses. The residues Y67, Y89, Y116 and F117 correspond respectively to Y82, Y104, Y131 and F132 in the human malectin 3D-structure. Since these residues are present in the loops region, it was decided to search further, and using again mutations to inspect the role of each loop residue in the interaction with the carbohydrate ligand, it was determined that the more critical residues to the binding event were in loop 3 (E114V115Y116, E129V130Y131 in the human malectin structure; Schallus et al., 2010).

The newly revealed X-ray structure of human malectin bound to Glc2Man3 provides more structural insights for understanding this interaction (**Figure 1.6**). First, the X-ray structure revealed a binding-pocket with a well-defined structure that accommodates two glucose residues, with a tight binding to the first glucose ring which is accommodated in a "hand-glove" fit. It was also observed that the residues S80, E102, K138, D201 and N202 (S65, E87, K123, D186 and N187 from *X. leavis* structure) promote direct hydrogen bonds to

the first glucose unit. The interactions between ligand and S80, K138 and N202 were for the first time detected. And it was possible to see that D201 makes direct hydrogen bonds with both glucose residues. In addition to the hydrogen bond to the hydroxyl group of carbon 2 of the first glucose unit, D201 also establishes a hydrogen bond with the hydroxyl of carbon 3 in the second glucose unit. The importance of E87 residue (E102 residue from X-ray structure) that sits at the bottom of the pocket is reinforced. This residue forms hydrogen contacts with two hydroxyl groups from carbon 3 and 4 of the first glucose unit. Both residues (D201 and E102) showed to be essential to binding event, and together direct the ligand in the right orientation for establishing the π/C-H stacking from the tyrosine 104 with the first glucose unit and tyrosine 131 with the second glucose unit. In addition, the OH from tyrosine 131 makes a water-mediated hydrogen bond with the second glucose and the α1,3 conformation favours a higher number of other contacts (Pinheiro et al., unpublished). Actually, the different conformation of nigerose (α1,3) and maltose (α1,4) provide an explanation for the difference in affinity shown by malectin towards these two different oligosaccharides. The reducing glucose moiety is differently oriented in the two structures: the maltose has an inversion of the A and B face of reducing residues in α1-4 glucosidic bound when compared to the α1-3 bond in nigerose. The X-ray structure shows this inversion places the methyl $C^{6'}$ group of maltose outside of the binding pocket, contrary to the nigerose complex, where it forms an additional water-mediated hydrogen bond to Y131. In these two complexes, the change in the glucose moiety leads to more affinity for nigerose, which as a dissociation constant of 26.3 ± 7 μM compared with 50 ± 0.5 μM for maltose (Schallus et al., 2010).

The paper by Pinheiro et al., highlighted the importance of the water network for specificity. Malectin complexed with G2M3 revealed a remarkable network of water-mediated hydrogens bonds between the protein and the oligosaccharide. The most conserved water-mediated interactions are of Glu129, Val130 and Lys204 with the carbon 4 hydroxyl group from the first glucose unit. Another conserved water identified in all X-ray structures mediates the interaction between the hydroxyl from carbon 6 (from the first glucose) and Tyr104. These interactions and the interaction between Asp201 with the glucose from the non-reducing end were suggested to behave like a lock in the binding pocket. The equatorial position of the second mannose of G2M3 favours the formation of an additional water-mediated hydrogen bond between carbon 6 hydroxyl from MAN4 and Tyr 131. While the most conserved water was suggested to have a pivotal role in the binding pocket organizations, it was proposed that the second most conserved water behaved like a lock between the pocket and the ligand.

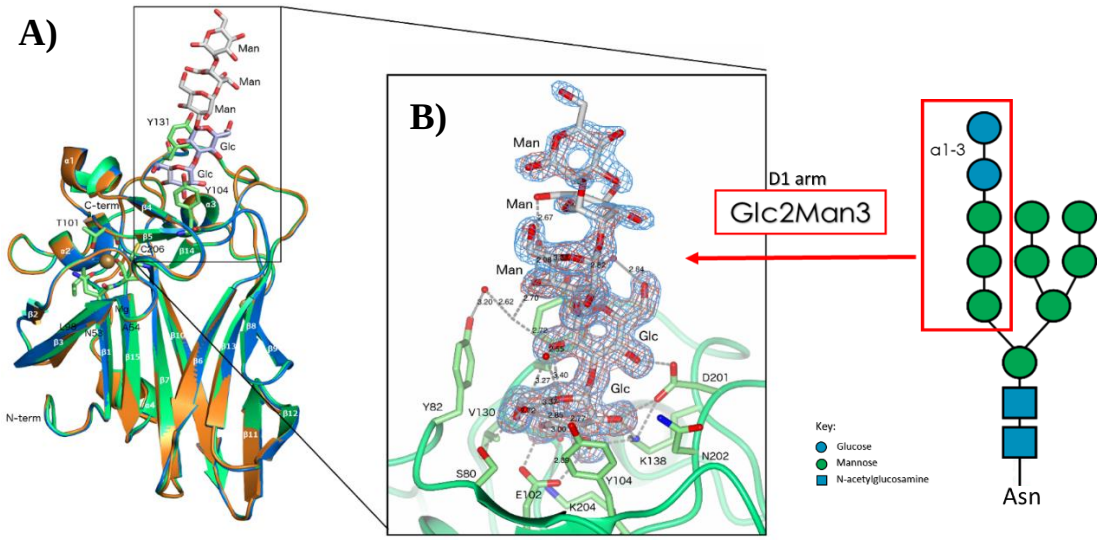


Figure 1.6 – Malectin binding pocket interactions with the Glc2Man9GlcNac2. *(A)* Ribbon representations of the overall structures of human Malectin obtained bound to Glc2Man3 (green), to Glc4 (orange) and Nigerose (blue). The oligosaccharide Glc2Man3 (white) are depicted as stick models in the binding pocket, showing that the orientation and position of the Glc-Glc (light blue) moiety in relation to the protein is conserved in all complexes. The two tyrosines residues that promotes π/C-H stacking contacts are also labelled and depicted as stick models (carbons in green). The magnesium atom is depicted as a green sphere. *(B)* Zoom in of the binding pocket stereo image. Malectin bounded to the di-glucosylated N-glycan D1arm (Glc2Man3) within its electron density, revealing the contacts with the protein residues (light grey). Adapted from Pinheiro et al., unpublished.

Malectin microarrays analysis showed that the intensity of the binding-signal with Glc2-N-glycan is substantially higher than whit nigerose, which is the capping sequence Glcα1-3Glc of the N-glycan intermediate. Therefore, the high-mannose N-glycan could have a strong cooperative effect, although no interaction whit Man9GlcNac2 was detectable (Schallus et al., 2008). These additional mannose residues have flexible conformation (Woods et al., 1998), which allows other contacts with the malectin surface and increases the affinity of binding. In contrast, the tri-glucosyl cap of the early step of N-glycosylation, Glcα1-2Glcα1-3Glc has a relatively rigid conformation (Petrescu et al., 1997) and the Glcα1-3Glc stays hidden until the trimming by GI. This could be the reason why malectin has a very low interaction with the Glc3-N-glycan, corresponding to the affinity to kojibiose (Schallus et al., 2008), which has a dissociation constant of 210 ± 6 μM obtained by isothermal titration calorimetry (Schallus et al., 2010). For different reasons, the interaction with the Glc1-N-glycan is also blocked, probably due to this N-glycan intermediate has an axial C-2 conformation on the mannose hydroxyl group. The conformation of the mannose residue of the Glcα1-3Man sequence instead of the glucose equatorial C-2 hydroxyl group hinders the stacking interaction of the sugar pyranose ring with Y116 and F117 in malectin.

Human malectin may interact with EGF-like repeats glycans

Moreover, in malectin microarrays analysis, Palma and colleagues noticed a weaker interaction with Xylα1-3Xylα1-3Glc. As mentioned previously, this trisaccharide is important in EGF repeats of mammalian notch which the glycosylation starts in the ER, where malectin is resident and could have a role in the folding of these proteins. Considering the location and similarities of the two sugars, they were able to model the trisaccharide in the electron density of the malectin-G2M3 complex (**Figure 1.7**). Except for the fact that it is a pentose, the xylose residues are similar to glucose residues and have the same equatorial arrangement. Therefore, the xylose residues fit perfectly except for the interaction with Ser80, which occurs with the extra -CH2OH group of glucose.

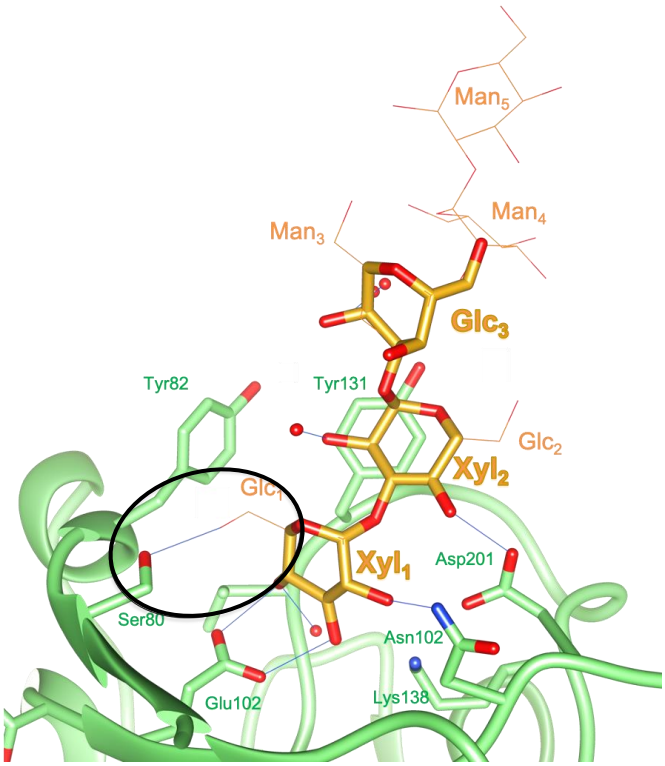


Figure 1.7 – Modelled structure of malectin binding pocket interactions with the Xyl2Glc (yellow and red sticks). The superimposition of the trisaccharide density fitting crystal structure of hMalectin with the hMalectin-Glc2Man3 complex shows the conservation of the H-bonding network (blue lines). Highlighted with a black circle is the unmodulated contact made between the CH_2OH group of the non-reducing glucose residue and the OH group of Ser80. Taken from Pinheiro et al., unpublished.

1.3.3 Malectin function on the Endoplasmic Reticulum

After the discovery of malectin, studies have been made with the intuit of discovering the function of this protein. Malectin is localised in the ER and this location is consistent with

the possible role in glycoprotein quality control (Galli et al, 2011; Schallus et al., 2008). Although it is consistent that malectin is involved in N-glycosylation, between GI and GII, the function in the pathway remains unclear. Two hypotheses have been proposed based on several experimental results.

Malectin recruit GII

After discovering malectin and its specificity, Schallus and coworkers suggested that this newly found protein may have the function of recruiting GII to the Glc2-N-glycan on the nascent polypeptide. Such propose is supported by the fact that GII is loosely associated whit the luminal side of the ER membrane (Brada & Dubach, 1984). So malectin, which has structural similarity to CBMs and the common aromatic sandwiching mode of binding, could act like CBMs and promote the association of the enzymes with the substrate. In this case, malectin interacts with the nascent polypeptide chain via the Glc2-N-glycan and is released only after the cleavage of the glucosidic bond by GII (Schallus et al., 2008). In contrast to GI, which acts immediately, the GII trimming is a regulated process and only occurs when a second glycan is present in the chain (Deprez et al., 2005). The delay in cleavage by the GII may also be due to this necessary recruitment mechanism operated by malectin. Therefore, malectin may function as a chaperone or may recruit chaperones to protect the nascent polypeptide against aggregation.

Malectin-ribophorin I complex in glycoprotein quality control

The first observation that triggered this hypothesis was a suggestion proposed by Galli and co-workers, which referred that malectin associates with newly synthesized glycoproteins, preferably with the misfolded conformers, interfering with the processing of N-linked oligosaccharides and substantially reducing their secretion (Galli et al., 2011). These studies, based on the augmentation of malectin expression during ER stress, confirmed that malectin selectively interacts with misfolded proteins (Chen et al., 2011; Qin et al., 2012). The malectin-misfolded glycoprotein interaction allows for its capture in a sugar-dependent manner and guides them into the ERAD pathway to prevent its secretion (Chen et al., 2011). The mechanisms by which malectin might act was suggested by Yamamoto group in 2011, proposing that malectin forms a complex with other chaperones to recognise misfolded glycoproteins (Qin et al., 2012).

The complex malectin-ribophorin I was predicted due to the ribophorin I interaction with newly synthesised incompletely folded glycoprotein (Wilson et al., 2005) and

the unspecific selective association of malectin with G2M9 glycan in folded or misfolded glycoprotein (Qin et al., 2012). Ribophorin-I is an ER-resident transmembrane protein that is associated with other OST subunits (Wilson et al., 2005). Malectin forms a stable complex with ribophorin-I which enhances the association with misfolded glycoproteins (Qin et al., 2012). Thereby, ribophorin-I selectively recognises misfolded proteins while malectin preferentially associates with it. Based on this, the hypothesis proposed is that misfolded proteins are trap by malectin-ribophorin I complex, and the interaction is maintained because of the hydrophobic properties of ribophorin I (**Figure 1.8**; Qin et al., 2012). Recently, Ramírez et al., reported a cryo-electron microscopy structure of human OST complexes, and the OST-B complex showed a short luminal stretch of malectin near to the luminal domain of ribophorin-I and the TM domain of an OST subunit (Ramírez et al., 2019).

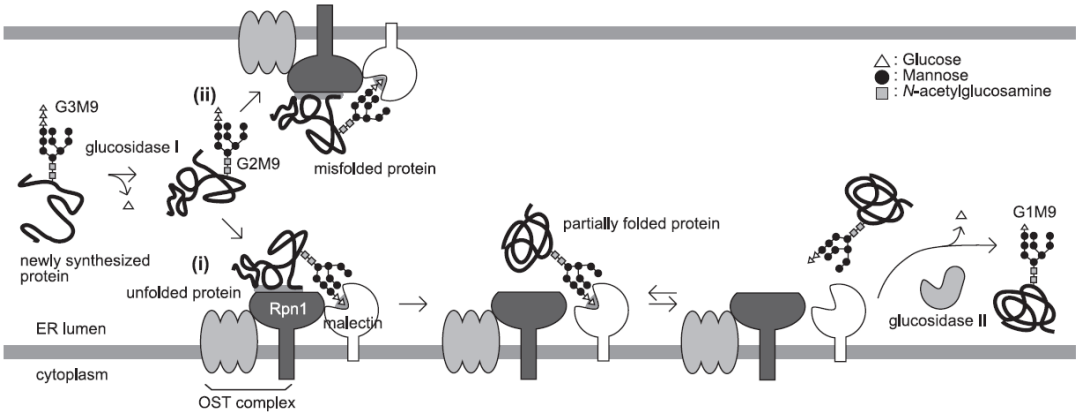


Figure 1.8 – The proposed role of malectin-ribophorin I complex in glycoprotein quality control. Malectin forms a complex with ribophorin I (Rpn1) to trap misfolded proteins after the trimming of glucosidase I. Despite no significant difference between folded and misfolded proteins, malectin specifically binds to G2M9 glycans. Whereas ribophorin I preferentially bind to *(i)* unfolded or *(ii)* misfolded glycoprotein, but the hydrophobic area of *(i)* unfolded proteins decrease and the polypeptide chain proceeds to proper folding. The opposite mechanism traps the *(ii)* misfolded protein and the release from the malectin-ribophorin I complex is difficulted. Adapted from Qin *et al*., 2012).

1.4 Carbohydrate Interactions Studies

As described in previous section, the glycans structure contain crucial information to biological activities. Protein recognition of glycans represents a major way in which this information is deciphered (Cummings et al., 2015). The structural basis by which a glycan-binding proteins (GBPs) binds with specificity and high affinity to a very limited number of glycans (or even a single glycan), among the many thousands that are produced by a cell,

underlying the need for physical techniques and high-throughput methods combined to structural studies (Rillahan & Paulson, 2011).

1.4.1 Carbohydrate microarrays technology

In 2002, carbohydrates microarrays (or glycan microarrays) emerged as a technique that allows researchers to simultaneously analyse unprecedented numbers of glycan-protein interactions *en masse* for various purposes (Rillahan & Paulson, 2011; Hirabayashi, 2003). Wang *et al.* described the fabrication of the technique, applied to detect carbohydrate-binding antibodies and to investigate the specificity of the antibodies for various carbohydrate structures (Wang et al., 2002). This technique was developed through the natural extension of the successful development of printed arrays and ELISA (Rillahan and Paulson 2011). The basis of the technique consists of fluorescence detection of GBP interaction with carbohydrates that are bound, covalently or noncovalently, to a solid surface in a spatially defined and miniaturised fashion (**Figure 1.9**; Paz & Seeberger, 2012).

This technology has made up for the lack of high-throughput methods for characterizing protein-carbohydrate interactions, addressing the three major requirements. First, a relatively small amount of samples can be analysed. Second, it allows parallel screening of many samples simultaneously on a single chip. Third, the presentation of carbohydrates on the surface creates a multivalent display that binds avidly and specifically to carbohydrate-binding proteins (Kiessling & Cairo, 2002). Its application depends on the glycan library, which is the most challenging part in carbohydrate microarrays. A practical approach to creating libraries is the fact that most GBPs have binding pockets that accommodate only a few monosaccharides residues. Thus, despite the numerous glycan structures, the information content relevant to most GBPs is contained in a very finite number of structural variations that occur at the tips of the complex glycan chain or are functional epitopes of small linear segments of a glycan polymer. Regardless of their source assembled libraries need to have a common functional group or property that allows them to be immobilized/arrayed on a compatible surface (Rillahan & Paulson, 2011).

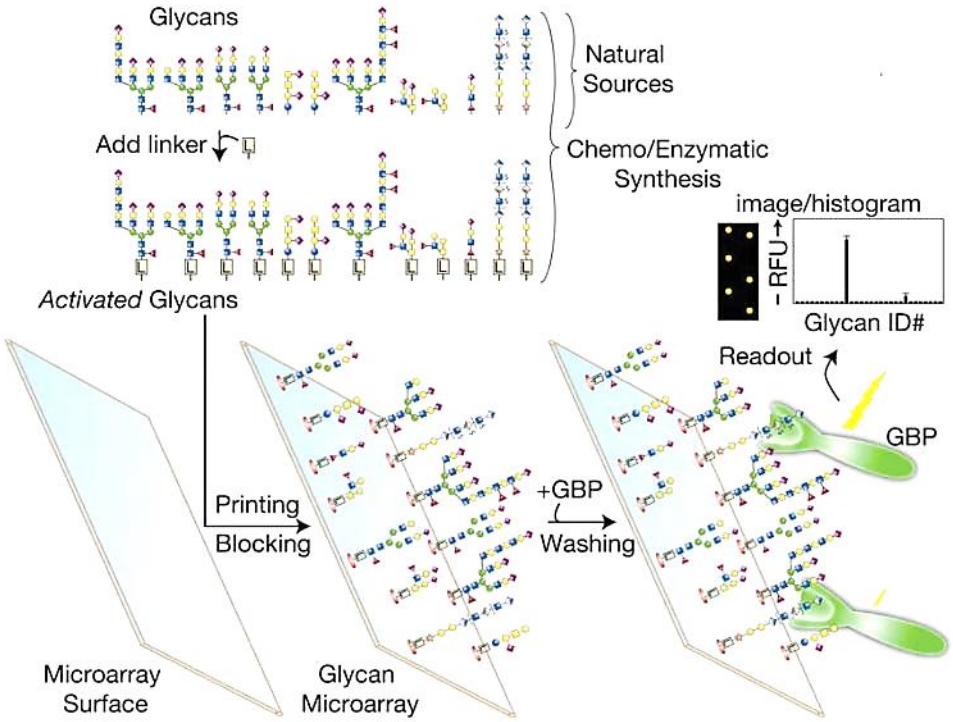


Figure 1.9 – Methods to produce and explore the carbohydrate microarrays, with GBP recognition of glycans. Taken from Cummings & Pierce, 2014.

Glycoarray construction

To assign the ligand of carbohydrate-binding protein and to be attached to an appropriate chip surface, it is essential to have pure sugars sequences (Feizi et al., 2003; Paz & Seeberger, 2012). As modern techniques analogous to those of molecular biology and recombinant protein expression cannot be employed to create oligosaccharides for microarrays, the first challenge was synthesize the carbohydrate epitopes that are important for specific recognition events. Currently, there are multiple strategies to access oligosaccharides for arraying, classified as synthetic approaches (chemical, enzymatic and combined chemoenzymatic synthesis) and isolation from natural sources.

The sources of natural glycans are infinite, comprising all the diversity of glycans that exist in the glycome, so the libraries of glycans isolated from natural sources would be sufficient to explore all the diversity of sugars. However, the isolation of pure compounds and verification of their structures is still a challenge. The isolation of glycoproteins and glycolipids provides a reducing sugar that can be used for introducing a functional group via a variety of efficient microscale conjugation strategies. The reducing end aldehyde of the monosaccharide serve as an electrophilic group for a chemoselective reaction with a number

of nucleophilic amine, hydrazide, or oxyamine containing reagents (Rillahan & Paulson, 2011).

Chemical or chemoenzymatic approaches to the synthesis of glycans have become an alternative to the isolation of natural glycans. But the diversity of carbohydrate structures that comprise linear or branched oligosaccharide chains, α or β anomeric configurations in the monosaccharide building blocks, and sugar rings linked via different carbon atoms, difficult the synthesis of oligosaccharides. Therefore, there was no routine systematic method for the synthesis of glycans of defined sequence. In chemical synthesis, multiple selective protection and deprotection steps are required for the hydroxyl groups on the glycosyl acceptor and donor, due to the similar reactivity of these groups (Feizi et al., 2003; Rillahan & Paulson, 2011). With the intention of automating the method, Seeberger and colleagues reengineered an existing apparatus used in automated peptide synthesis and created an automated solid-phase method for the glycan synthesis that includes selective protection and deprotection steps (Plante et al., 2001). An alternative approach to oligosaccharides synthesis is the OptiMer-based one-pot, in which an oligosaccharide of interest is generated by the sequential addition of building blocks (thioglycosides) that can be tuned in the presence of protecting groups. The OptiMer was a computer program developed to search the database of characterized glycosyl donors from building blocks and identify sets of monomer donor that would likely provide the best yield of the desired oligosaccharide (Zhang et al., 1999). In other hand, the enzymatic synthesis has the advantage that substrates are used in their natural form, without a need of protecting groups to direct the regio- or stereo-specificity of glycosidic bond formation. The glycosyltransferases products have higher yields than glycosidases, but their high cost and limited availability are the main drawbacks. To offset the disadvantages of these methods, the chemoenzymatic synthesis was developed conjugating both. In overall, the method uses an appropriate glycosyltransferase and nucleotide sugar to chemically synthesized mono- or oligosaccharide, with a linker containing an appropriate functional group (Rillahan & Paulson, 2011).

In addition to the synthesis of the glycan, the fabrication of the array requires that the functionality installed in the glycan library must be appropriate for the method of immobilization on the surface of the array. There are several immobilization methods involving covalent and noncovalent attachment. The covalent method depends on the synthetic strategy of the carbohydrate, and the coupling reaction used should be fast, specific, and high yielding. One of the methods widely employed in chemically synthesized oligosaccharides is the selective reaction between thiol and maleimide groups. Also, approaches to unprotected glycans were developed, involving direct derivatization, such as reductive amination, oxime

bond formation, and glycosylamine synthesis. In addition, to avoid prior modification, isolated sugars were directly attached on aminooxy- or hydrazide-containing surfaces (Paz & Seeberger, 2011). Unlike this methods, noncovalent immobilization techniques include nonspecific adsorption of free glycans to underivatized solid surfaces. The adsorption techniques involve hydrophobic forces and electrostatic interactions using nitrocellulose and oxidized polystyrene surfaces. Other noncovalent immobilization methods more suitable for microarrays containing monosaccharides or low molecular weight oligosaccharides have also been reported. Some involve the attachment of fluorous-tagged glycans, site-specific DNA-based glycan microarray and conjugation with a carrier molecule, such as a lipid chain, like in neoglycolipid probes (NGL; Paz & Seeberger, 2012; Rillahan & Paulson, 2011).

NGL-based oligosaccharide microarrays

In 1985, the NGL-technology emerged as a sensitive micro-method to study binding interactions of antibodies and other carbohydrate-recognizing proteins. This approach consists of making artificial glycolipids by linking oligosaccharides to amino-phospholipid tags (Palma et al., 2014; Tang et al., 1985). Inspired by Wang et al. method of non-covalently immobilizing polysaccharides and glycoproteins in nitrocellulose, Feizi and colleagues adapted their NGL technology to generate the first microarray system for complex oligosaccharides (Fukui et al., 2002). This technology was developed due to the need to use a microscale method that allows the detection and characterization of functional epitopes in the carbohydrate chains of glycoproteins. The construction of the array consists of conjugating the oligosaccharides by microscale reductive amination to an aminolipid, 1,2-dihexadecylsn-glycero-3-phosphoethanolamine (DHPE). The conjugation of the terminal aldehyde-reducing group of the oligosaccharide with the amino group of the lipid, provides an amphipathic property to the oligosaccharide. This property enables the immobilization of the converted lipid-linked probes (DH-NGLs) in clustered display on solid matrices (Feizi & Chai, 2004; Liu et al., 2009). However, the product of this reductive amination is a monosaccharide residue with a ring-opening at the reducing ends, which may affect the biological activities of short oligosaccharides. To overcome this limitation, in 2007, Yan Liu and co-workers were the pioneers of the NGL probes prepared via oxime ligation for microarray (Liu et al., 2007). This novel type of NGL with ring-closed monosaccharide cores, involved the conjugation with an aminooxy-functionalized DHPE (AOPE) via microscale oxime ligation (**Figure 1.10**). This technology allows the expansion of the library of probes generating ‘designer’ microarray (Palma et al., 2014).

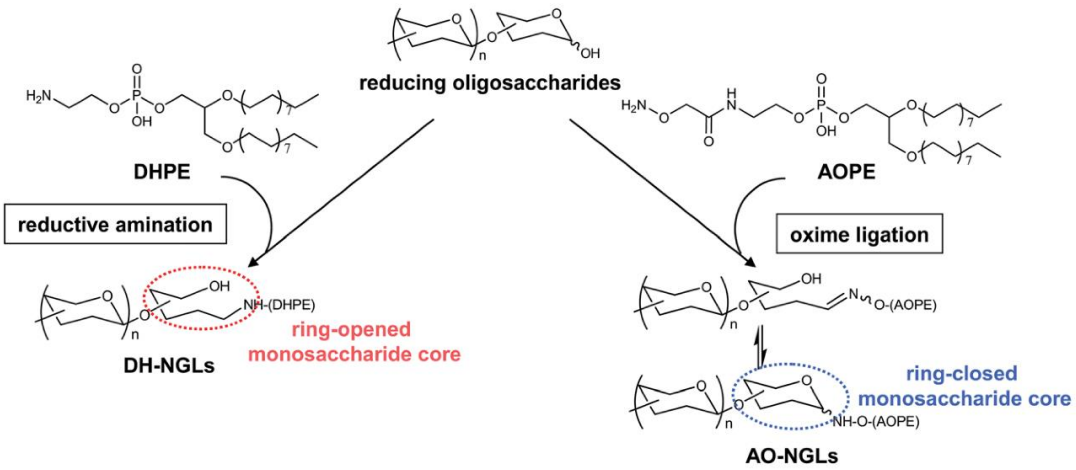


Figure 1.10 – Preparation of DH-NGL and AO-NGL probes from reducing oligosaccharides by reductive amination and oxime ligation, respectively. Taken from Feizi, 2019.

1.5 Thesis Objectives

The reported studies of Malectin indicate that this protein is a very important conserved lectin in animals involved in the quality control of glycoproteins. Thus, understanding its mechanisms of binding to the glycan ligand can have a potential value in human health. Previously, Malectin-ligand interactions were characterized by X-ray Crystallography (Pinheiro et al., unpublished) and the data revealed that some residues appear to be crucial for binding events. The current study will have as main objective to complement this work by defining the relative importance of these crucial amino acids in the Malectin-ligands interactions through specifically designed mutants and their carbohydrate microarray analyses. Meanwhile, encouraged by the prospect of the modelled human malectin in complex with Xyl2Glc and the weaker interactions with this trisaccharide observed in the microarrays, we venture the challenge of designing a protein capable of recognizing with more specificity this extended O-glucosylation present in the EGF repeats from some mammal protein.

Outputs of the Thesis:

3-month scholarship

Project entitled "Designing New Proteins for Human Health by Capitalising on the Unique glycan binding specificity of Malectin", UCIBIO summer school, 1st August – 31st October.

Oral communications

Designing New Proteins for Human Health by Capitalising on the Unique glycan binding specificity of Malectin, 13/02/2020, *Jornadas Intercalares de Mestrado*, FCT-NOVA, Lisbon.

Designing New Proteins for Human Health by Capitalising on the Unique glycan binding specificity of Malectin, 22/10/2020, *Symposium of the UCIBIO Summer school*, FCT-NOVA, Lisbon.

Chapter 2 – Materials and methods

Analytical scale (Highland, AE Adam) and precision scale (Pioneer™, OHAUS) were used for all buffer preparations and solutions. All protocols that include bacteria cell manipulation were performed in a sterile environment. All absorbance measures at 600 nm (OD600) were measured in a spectrophotometer (Thermo Scientific Evolution 201) using non-inoculated media as blank.

2.1 DNA, plasmid and primers design

The globular domain of malectin *H. sapiens* (AAs 42-228; Schallus et al., 2008) was previously cloned into pETM-10 vector of the M-series by Gunter Stier, EMBL (Palma et al., 2010). This vector contains an N-terminal His_6-Tag (Histidine affinity tag), a T7-Lac promoter and a kanamycin marker (Supplementary Figure 1).

The mutagenic primers are designed according to the desired mutation, which should be in the middle of the primer sequence. Also, the minimum value of GC content should be 40%, and the 3' primer termination with G or C bases promotes its stabilization and binding with the template. Also, the length of the primers in bases should be between 25 and 45, with a melting temperature (Tm) higher than 78 ºC. According to the aim of this project, the targets in the malectin sequence were selected (Supplementary table 1), and the mutagenic oligonucleotides were designed (Table 2.1). STAB VIDA services were requested to synthesize primers (see more details in Supplementary table 4).

Table 2.1 – Design of oligonucleotides used for mutations. The nucleotides encoding the mutated residue are highlighted in bold and the nucleotides that will promote the mutation are underlined.

Desired Mutation		*Primer Sequence (5'→3')*
S80Q	Fwd	TTCATGCCATAGTC**TTG**GGCTCGGCCCACCCG
	Rev	CGGGTGGGCCGAGCC**CAA**GACTATGGCATGAA
E102A	Fwd	TCTCCTCATTGTACC**GCG**CAGTTTGATACAGGATC
	Rev	GATCCTGTATCAAACT**GCG**CGGTACAATGAGGAGA
Y104F	Fwd	CAAAGGTCTCCTCAT**TGA**ACCGCTCAGTTTGATAC
	Rev	GTATCAAACTGAGCGG**TTC**AATGAGGAGACCTTTG
Y131F	Fwd	GCTGGGACTGTGCAA**AGA**AGACCTCTGCAAATTTC
	Rev	GAAATTTGCAGAGGTC**TTC**TTTGCACAGTCCCAGC
Y199AY200A	Fwd	TCTACATTGAGTTTGTCAAGGGG**GCCGCT**GACAATCC CAAGGTCTGTGCAC
	Rev	GTGCACAGACCTTGGGATTGTC**AGCGGC**CCCCTTGAC AAACTCAATGTAGA
D201A	Fwd	CACAGACCTTGGGATTG**GCA**TAGTACCCCTTGACA
	Rev	TGTCAAGGGGTACTAT**GCC**AATCCCAAGGTCTGTG

The two oligonucleotides (Forward and Reverse) have the same starting and ending positions on opposite strands of the plasmid DNA. All the primers were diluted in MiliQ-water to 1 µM.

2.2 Site-directed mutagenesis and DNA extraction

Mutagenesis refers to changes in the content of information in the DNA sequence of an organism or cell. In vitro mutagenesis, specific changes in the base sequence are customized and monitored to achieve the desired mutation for obtaining information of a component in a biological process or a determined amino acid on a target protein. For this project, we planned six local and precise alterations into the malectin sequence. The method used was the site-directed mutagenesis using double-stranded DNA templates with the selection of mutants with *Dpn*I. The *Dpn*I is a restriction endonuclease from the bacterium *Diplococcus pneumoniae*, which activity is blocked by overlapping CpG methylation. This methylation-sensitive allows the enzymatical digestion of the bacterially generated DNA used as a template for amplification without destabilized the DNA synthesized during the reaction in vitro. Therefore, the plasmids with the desired mutation will be the *DpnI*-resistant molecules. This circular mutagenesis method has high fidelity when the design of the primers and the choice of the DNA polymerase are appropriated.

All genes targets were modified by polymerase chain reaction (PCR) has described in table 2.2, using the MyCycler™ Thermal Cycler, BioRAD. The reaction mix includes 37 µL of MiliQ-water, 0.5 µL dNTPs, 1.5 µL dimethyl sulfoxide (DMSO), 5 µL of 10x reaction buffer, 3 µL of template DNA, 1.25 µL primer forward, 1.25 µL primer reverse and 0.5 µL of Nzyproof DNA polymerase.

Table 2.2 – PCR conditions used for target genes.

	Steps	**Temperature**	**Time**
1	**Initial denaturation**	95 ºC	3 min
2	**Amplification (18 cycles)**		
	Denaturation	95 ºC	1 min
	Annealing	*	1 min
	Extension	72 ºC	6 min
3	**Final extension**	72 ºC	10 min
	Hold	4 ºC	∞

(*) The annealing temperature was optimized for each primer set. Obtained results are disclosed in Chapter 3.

The PCR products were analysed by agarose gel (described in more details in the next Section). The positive results were digested with 1 µL of *Dpn*I (Scientific, Thermo Fisher) and incubated by 3 hours at 37 ºC. All enzymes were inactivated through a temperature shock at 80 ºC for 20 min.

2.3 Analysis by agarose gel electrophoresis

The agarose gel electrophoresis is one of several physical methods for determining the size of DNA. In this method, DNA is forced to migrate through a highly cross-linked agarose matrix in response to an electric current. The molecule will emigrate to the positive (*red*) pole due to the negative charge of the DNA. Also, the size and conformation of the DNA and the ionic strength of the running buffer can affect the migration rate through the gel. Therefore, the percentage of agarose on the gel depends on a few things: the size of the fragments we are looking for, and how good we need the separation of fragments to be. To analyse the results of PCR products, we used agarose gel of 0.8% (w/v), recommended for separating DNA fragments between 800 – 10000 bp (Green, 2012 – General Recommendations for DNA Electrophoresis, Thermo Scientific).

0.56 g of agarose (Sigma-Aldrich) was dissolved in 70 mL of 1x TAE buffer containing Tris-Acetate (242 g), 20 mM EDTA (18 g), Acetic acid (57.1 ml) pH 7,5 and heated in a microwave. After cooling, 2 µL of Invitrogen SYBR™ Safe DNA Gel Stain was added for visualization of DNA. The sample was prepared by added 1 µL of Invitrogen BlueJuice Gel Loading Buffer in 5 µL of DNA. 3 µL of 200 bp DNA Ladder (Bio-TEK, Omega) was used as a ladder and gel was ran at a constant voltage of 50 V for 50 min. After finished the run, the gel was visualized by Transilluminator (VWR).

2.4 DNA isolation, and sequencing

After transformation in *E.coli* DH5α as described in Section 2.6, isolated colonies were picked and inoculated into 10 mL LB medium with 50 µg/mL kanamycin overnight at 37 ºC in a shaker at 200 rpm (Orbital Shaker-Incubator ES-20, from Grant.bio). All samples were purified with NZYMiniprep Kit (MB010) from NZYTech®. The extraction and isolation of

the DNA were analysed by agarose gel electrophoresis (see Chapter 3). STAB VIDA sequencing service was used to confirm the protein sequence and desired mutation. For forward and reverse sequencing reaction, T7 forward primer (5'-TAATACGACTCACTATAGGG- 3') and T7 reverse primer (5' -GCTAGTTATTGCTCAGCGG- 3') were used.

2.5 Preparation of competent cells

Cells ability to take up foreign DNA from its surrounding environment is named competence. The process of making competent cells creates temporary pores in a cells membrane for DNA to pass through (transformation). In this project, a chemical method was used to induce cell competence. The strain *E. coli* DH5α was used for molecular cloning, and *E. coli* BL21(DE3) was used as the host for overproduction of recombinant proteins.

Preparation of DH5α competent cells

In day 1, 50 µL of DH5α cells were uniformly spread in LB agar plates. In day 2, one colony of *E. coli* was selected from the plate and inoculated into 5 mL LB broth medium (previously autoclaved) and kept at 37 ºC, at 100 rpm, overnight (Orbital ShakerIncubator ES-20, Grand.bio). 1 mL of the pre-inoculum was added to 200 mL of super optimal broth (SOB) medium, incubated at 18 ºC, with 200 rpm, until the optical density at 600 nm (OD_{600nm}) reached 0.4-0.5. At this value of 0.5 represents the exponential phase of the microbial growth of bacterial cells. Cell culture was placed on ice for 15 min and was centrifuged at 2500 x g for 10 min at 4 ºC, in previously autoclaved cold centrifuge tubes. The pellet was resuspended in 80 mL of ice-cold transformation buffer (TB), it was kept on ice for 30 min and centrifuged as before. After the second centrifugation, the new pellet was gently resuspended in 18.6 mL of ice-cold TB and 1.4 mL of DMSO. After 15 min on ice, aliquots of 1.5 mL were prepared in sterile and ice-cold microcentrifuge tubes and stored immediately at -80 ºC.

Preparation of BL21 competent cells

In day 1, 50 µL of BL21(DE3) cells were uniformly spread in LB agar plates. In day 2, one colony of *E.coli* was selected from the plate and was inoculated into 5 mL LB medium (previously autoclaved) and kept at 37 ºC, at 200 rpm, overnight (Orbital ShakerIncubator ES-20, Grand.bio). 1 mL of the pre-inoculum was added to 200 mL of fresh LB medium (previously autoclaved) and added 20 mM $MgSO_4$ (filtrated). After that, it was incubated at

37 ºC, with 200 rpm, until it reached the OD_{600nm} 0.4-0.5. Cell culture was placed on ice for 15 min and was centrifuged at 4000 x g for 5 min at 4 ºC, in previously autoclaved cold centrifuge tubes. The pellet was resuspended in 60 mL of ice-cold transformation buffer 1 (TBF1), it was kept on ice for 90 min and centrifuged as before. After the second centrifugation, the new pellet was gently resuspended in 8 mL of ice-cold transformation buffer 2 (TBF2). Aliquots of 200 µL were prepared in sterile and ice-cold microcentrifuge tubes and stored immediately at -80 ºC.

2.6 Transformation

The transformation of competent cells DH5α and BL21(DE3) was performed in a sterile environment and used for cloning and expression, respectively. 1 µL of DNA was mixed with 50 µL of *E.coli* competent cells and incubated on ice for 30 min. Samples were placed in 42 ºC heat block for 1 min to perform heat shock and then incubated on ice for more 2 min. The cells were re-suspended in 1 mL of LB medium and the sample was incubated, to allow cell recovery. After 1 h at 37 ºC in the shaker at 200 rpm incubation (Orbital ShakerIncubator ES-20, from Grant.bio), the cell suspension was centrifuged (Mini Spin, Eppendorf) for 1 min at 5000 rpm and the 900 µL of supernatant was discarded. The cells were re-suspended in the remaining medium and spread with a sterilized glass rod on an LB-agar plate supplemented with kanamycin antibiotic at 50 µg/ml and incubated overnight at 37 ºC. The DNA plasmid has a gene that confers resistance to kanamycin and so, only cells that have the plasmid will be able to proliferate and form colonies. The LB-agar plates with colonies were stored at 4 ºC.

2.7 Expression with IPTG-induction

After transformation described in section 2.6 using the *E.coli* strain BL21(DE3), one colony of transformed cells was pre-inoculated into 10 mL LB medium supplemented with 50 µg/mL kanamycin, following incubation overnight at 37 ºC at 200 rpm (Orbital Shaker-Incubator ES-20, from Grant.bio). The overnight culture was transferred to 500 mL of media with 50 µg/mL kanamycin and incubated at 37 ºC at 110 rpm (Shaker IS-971R from Lab Companion). Cells were induced with 0.2 mM Isopropyl-β-D-thiogalactopyranoside (IPTG) when samples reached the OD_{600nm} of 0.6 ~ 0.8 and incubated overnight at 18 ºC at 110 rpm. The cells were harvested by centrifugation for 10 min at 5000 x g at 4 ºC (JA-10 rotor, Avanti J-26 XPI from Beckman Coulter).

2.8 Cell harvesting and lysis

The first step to disrupt the cell membrane was harvested the cells by centrifugation as described in the last section. The precipitate was recovered and stored at -20 ºC, a step that also facilitated cell rupture. This cell pellet was resuspended (10 mL/g wet cell pellet) in 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 10 mM Imidazole, 2 mM β-mercaptoethanol, protease inhibitors (SIGMAFAST™ Protease Inhibitor Tablets) to isolate the molecules of interest and keeping them in a stable environment. A sonication (UP100H, from Hielsher) step was used with 4-5 cycles of 30 s and pauses of 1 min, always on ice, to complete cell rupture and release cellular content. Centrifugation for 10 min at 2000 x g at 4 ºC (JA-10 rotor, Avanti J-26 XPI from Beckman Coulter) followed. Filtration with a 0.45 µm membrane of cellulose acetate (Whatman, GE) was performed to clarify the soluble fraction (supernatant) to avoid clogging the column in the purification step (see the next section in this chapter).

2.9 Purification of proteins by affinity chromatography

Purification of recombinant proteins can be carried out with different techniques: in batch, with gravity flow or centrifugation, and the system can be manual or automated. Tagged proteins are convenient to work with and make the purification process simpler. Tags enable recombinant proteins to be purified by affinity chromatography, which is designed to capture the tagged recombinant protein based on biorecognition of the tag. Separation principles in chromatographic purification, such as Gel filtration, Hydrophobic interaction, Ion exchange, Affinity and Reversed-phase, explore the differences in proteins specific properties.

Immobilized metal ion affinity chromatography (IMAC) is a method that uses metal ions immobilized in a chromatographic media. The system used to express malectin is a histidine tag which has a high selective affinity for Ni^{2+} and several other metal ions. So, the purification step was performed using an IMAC column by His GraviTrap™ containing precharged Ni Sepharose™ 6 Fast Flow (GE Healthcare), which has high protein binding capacity and low nickel ion (Ni^{2+}) leakage. His GraviTrap™ is a prepare, single-use gravity flow column designed for fast and straightforward purification. The protocol (Data File 11-0036-90 AA, GE Healthcare) was adapted to obtain the protein in the purest and more stable form, and for being a single purification step, two washes with different Imidazole concentrations were applied. Imidazole competes with proteins for binding to Ni Sepharose.

Equilibration was performed with buffer containing a low concentration of imidazole to reduce nonspecific binding of host cell proteins.

The purification on His GraviTrap starts with an equilibration step, following to sample application, washing and elution. First columns were washed with 10 mL of MiliQ-water to remove the ethanol. To equilibrate the column, 10 mL of the binding buffer containing 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 10 mM Imidazole, 2 mM β-mercaptoethanol, protease inhibitors (SIGMAFAST™ Protease Inhibitor Tablets) were used. Then, 10 mL of the sample was loaded onto the column, and the protein was retained by His-tag-Ni^{2+} binding. Following the wash step with 10 mL of equilibration buffer and after this 5 mL of 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 60 mM Imidazole, 2 mM β-mercaptoethanol, protease inhibitors (SIGMAFAST™ Protease Inhibitor Tablets) were applied. Finally, 2.5 mL of elution buffer with 20 mM Tris-HCl, pH 8.0, 250 mM NaCl, 250 mM Imidazole, 2 mM β-mercaptoethanol, protease inhibitors (SIGMAFAST™ Protease Inhibitor Tablets) were added to the column to recover the protein.

2.10 Desalting and protein quantification

Desalting or buffer-exchange is a well-proven, fast and straightforward method that quickly removes low molecular weight contaminants while transferring the sample to the desired buffer in one step. PD-10 Desalting Columns prepacked with Sephadex™ G-25 Medium were used with the gravity protocol. First, the LabMate PD-10 Buffer Reservoir was mounted, and the columns were placed in the PD-10 Desalting Workmate, followed by clean-up with 25 mL of MiliQ-water and equilibration with 25 mL of desalting buffer (20 mM Tris-HCl, 150 mM NaCl, 2 mM DTT). Then, 2.5 mL of the sample was loaded, and then eluted with 3.5 mL of desalting buffer.

The proteins concentration was determined with a SpectraMax®190 Absorbance Plate Reader from Molecular Devices at 280 nm. The desalting buffer (described above) was used as a blank sample, and the quantification calculated by applying the Lambert-Beer equation accordingly with the results in the following chapter.

2.11 Polyacrylamide gel electrophoresis SDS-PAGE analysis

In sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE), polyacrylamide gels are formed by the reaction of acrylamide and bis-acrylamide (N,N'-methylenebisacrylamide), resulting in a highly cross-linked gel matrix. Proteins move through the gel in response to an electric field. This movement is possible after denaturing the proteins to giving them a uniform negative charge, thus making it is possible to separate them based on the size as they migrate towards the positive electrode. SDS is a detergent that, along with boiling and a reducing agent (as β-mercaptoethanol), it disrupts the tertiary structure of the protein and coats to masks the intrinsic charges on the R-groups. This method was applied to analyse the separation of unfolded proteins based on their molecular weight.

The protein sample was prepared at 2 µg mixed with the loading buffer (4x Tris-HCl, 10 % SDS, 0.6 M DTT, 0.012 % bromophenol blue, 30 % glycerol), and boiling in a heat block at 100 ºC for 5 min. Running gel of 10% acrylamide and staking gel was prepared according to table 3 in the supplementary material. After polymerization, the gel was transferred to specific support, and 1x Running buffer (10x Running buffer (1 L): 30.3 g Tris base, 144 g Glycine, 10 g SDS) was added. Samples were loaded in each well of the gel after a fast spin and Marker II from NZYTech was used as a marker. The gels run at 100 mA and 120 W during 1h30 hour and are revealed by incubation with Bio-Safe Coomassie (Bio-Rad).

2.12 Differential Scanning Calorimetry (DSC)

In the 1960s, academic groups pioneered calorimetric methods to measure the fundamental thermodynamic driving forces behind the process of protein stability, folding and binding interactions. Differential scanning calorimetry (DSC) was the oldest technique developed in this field and can quantify kinetics and equilibrium. DSC also measures basic thermostability, or "susceptibility", of proteins to thermal denaturation. For simple proteins, a single peak of heat absorption is often seen in the scan, or 'thermogram', and, after correction for instrumental baseline, transition baseline and normalization for concentration, the peak can be integrated to provide a direct calorimetric measurement of the enthalpy for the process (ΔHcal) and a measure of the melting temperature (Tm; Johnson, 2013).

Usually, since proteins in solution are in equilibrium between folded (native) and unfolded (denatured) states, by heating the sample in a constant rate and, subsequently,

measuring the heat change associated with intrinsic thermal denaturation processes, we can determine the calorimetric enthalpy (ΔHcal) and the change in heat capacity (ΔCp) of the unfolding process Besides these thermodynamic proprieties, the DSC data also provides the Tm of the sample (maximum value). The Tm value correlates with the temperature at which the protein's unfolded and folded states are in equal concentrations and is an indicator of the protein thermal stability.

DSC technique was used to characterize the stability and the folding process of the six new proteins. The DSC assays were performed using a TA^{TM} Nano DSC apparatus. The melting temperature and thermodynamic parameters were obtained by fitting the thermograms in a NanoAnalyze^{TM}v3.10.0 Data Analysis software (TA Instruments).

For DSC studies, the proteins were prepared in 20 mMTris–HCl, pH 7.2, 150 mM NaCl, 2 mM DTT, and quantified as described in section 2.10. DSC measurements were used 700 µL of each sample in concentrations of 0.5 mg/mL, and the temperature range was 10 ºC to 100 ºC, with an increase of 0.5 ºC/min and 2.98 atm.

2.13 Carbohydrate microarray assay

Two Sets of carbohydrates were analyzed: Fungal, bacterial & plant PS Set 2, and 42-56 Set with NGL probes. The seven proteins were analyzed in each Set in parallel, using Malectin WT as a control based on Palma et al. 2010.

Sample preparation

The proteins were prepared to form a complex that contain the antibodies necessary for detection of interaction. To form the pre-complex, 2.4 µl of the primary antibody (mouse anti-his) was pre-incubated by gently mixing with 8 µl of the secondary antibody (biotinylated anti-mouse IgG) for 15 min. After centrifuging for 10 min at 1000 x g at 4 ºC, 10 µg/mL of protein was added to the antibodies, making a complex with the 1:3:3 ratio. The mixture was incubated for another 15 min. The blocking solution composed by 3% BSA (w/v) diluted in a HEPES saline buffer (HBS) pH 7.4 (which contained 5 mM HEPES, 150 mM NaCl and 5 mM $CaCl_2$) was added to a final volume of 80 µL.

Microarray slide preparation

First, the chosen glass slide is scanned at 532 nm using the GenePix® 4300A microarray scanner (Molecular Devices), to view the Cy3 fluorescence emitted by the printed probes spots. Then, the nitrocellulose membrane slide was placed carefully into a frame. The pads were wetted with 150 µL of Mili-Q water and removed by briefly inverting the frame. 150 µL of blocking solution was added onto each pad and incubated for 60 min. This step was essential to decrease the background noise in the analysis.

Carbohydrate microarray binding assay

After the incubation time with the blocking solution, the different protein-antibody complexes were added individually to each pad, without drying the glass slide. The binding remained in incubation for 90 min, followed by aspiration of each pad and four washes with 150 µL of HBS buffer. The Streptavidin was diluted in blocker solution (1:1000) and added into the pads to incubate for 40 min. Then, the pads were washed four times with HBS buffer and two times with MiliQ-water. The slide was removed from the frame and dried in the dark.

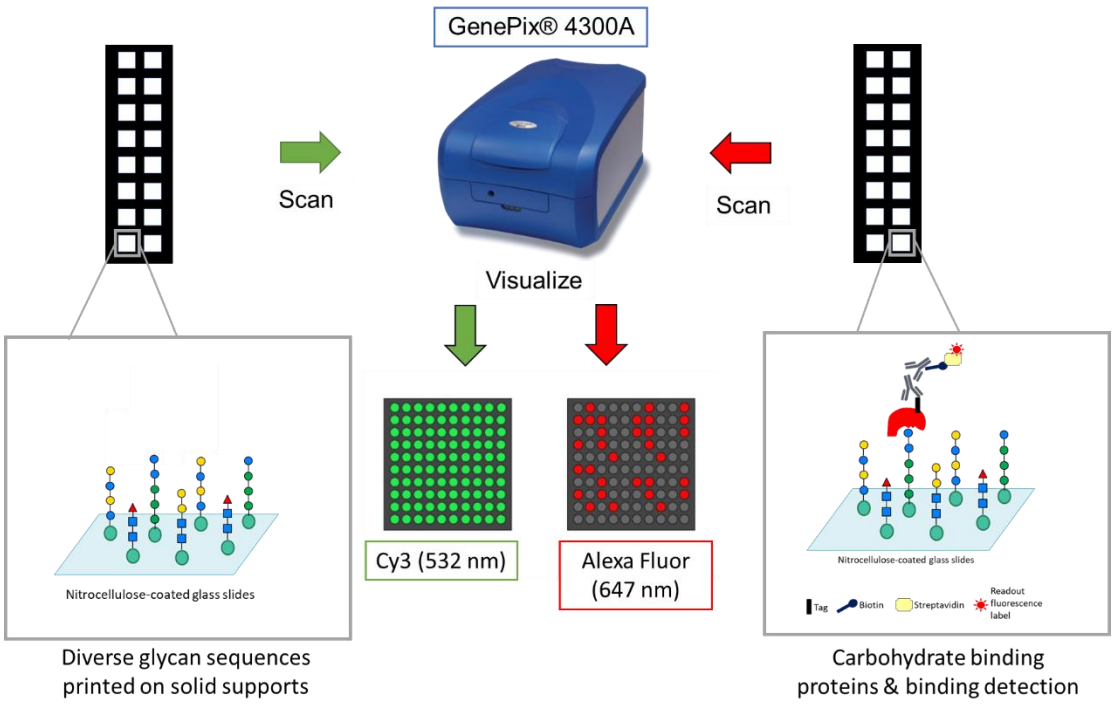


Figure 2.11 – Overview of the carbohydrate microarray binding of the protein-antibodies complex.

Carbohydrate microarray data analysis

The dried microarray slide was scanned for Alexa Fluor-647 (AF) at 647 nm using the GenePix® 4300A microarray. The detection and quantification of the binding event through AF fluorescence intensity on each spot was performed using the GenePixPro7 Software (Molecular Devices). This software used several parameters to perform the imaging analyses: laser wavelength, photomultiplier (PMT) gain and laser power (%). The laser wavelength used depends on the desired imaging, the Cy3 fluorescent dye can be excited using the 532 nm laser and the Alexa Fluor 647 dye is a bright, far-red–fluorescent dye with excitation ideally suited for the 647 nm laser; the PMT is used to detect the photons that are emitted by these two laser-excited fluorophores and must be adjusted to reduce imaging saturation but to acquire the total signal. In this scanner, the PMT inferior limited of 350 was used; The laser power can be reduced to fine-tuning focal plane adjustments, however reducing laser power also reduces signal-to-noise. Different laser powers were scanned to conveniently analyse the imaging. The images acquired were saved in the Tagged Image File Format (*.tif).

After Cy3 imaging acquisition, the construction of 16 blocks (a grid with *feature-indicator*) on each arrayer was performed to locates all the spots (8x8) present in every single block. The grids were saved as a GenePix settings files (.gps), and then adjusted to the Alexa Fluor 647 slide scan. The software calculates the Alexa Fluor 647 fluorescence intensity of the feature-indicator and identifies as background what is outside the grid spot. All the information about image acquisition and analysis, and the data extracted from each individual feature was saved as a single GenePix Results File (*.gpr). The results were processed and analysed in a microarray software developed at the Glycosciences Laboratory, Imperial College London (Stoll & Feizi, 2009).

Chapter 3 – Production of new proteins through the mutation of the lectin domain of malectin

3.1 Introductory remarks

In this chapter, the aim was to design new proteins that would allow the study of the interactions of human malectin with glycans. Based on the information already obtained with the malectin structure complexed with the ligand G2M3 (Pinheiro et al., unpublished), we defined three study objectives: 1) Characterization of the Malectin binding pocket interactions; 2) Role of two putative binding tyrosines in the interaction with the N-glycan; 3) Malectin modification for the recognition of O-glucosylation in EGF-repeats.

To fulfil these tasks, we projected several mutations:

Characterization of the Malectin binding pocket interactions

- Glutamate 102 makes two hydrogen bonds with two hydroxyl groups of the first glucose and Aspartate 201 with the first and second glucose residues of the G2M3 ligand. Therefore, we choose to mutate these two residues to alanine.
- Tyrosine 104 and 131 are involved in the water-network and made a π/C-H interaction with the pentasaccharide. To remove these water-mediated hydrogen bonds and to define the influence of the π/C-H interactions, we decided to mutate these selected residues to phenylalanine.

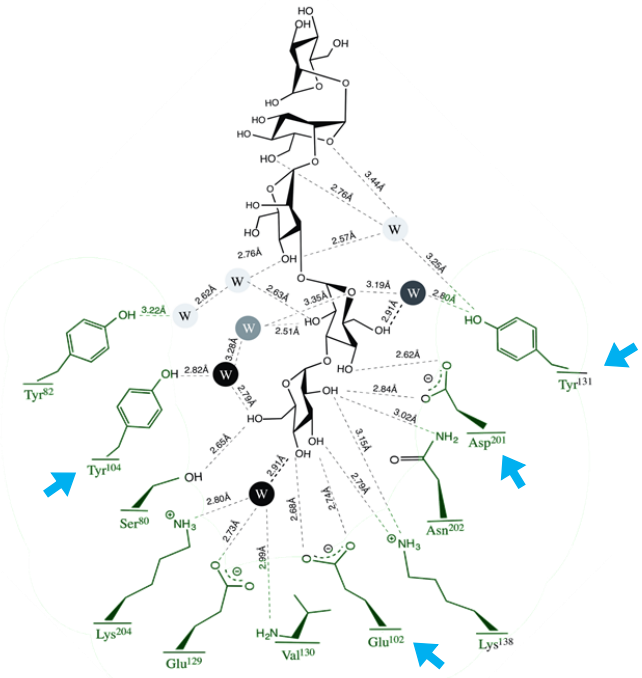


Figure 3.12 – Schematic representation of the malectin binding pocket and Glc2Man3 interactions. The pentasaccharide with two glucose and three mannose residues. The most conserved water-mediated hydrogen bonds are represented in black and the less conserved in light grey. The selected residues are highlighted with blue arrows. Adapted from Pinheiro et al., unpublished.

Role of two putative binding tyrosines in the interaction with the N-glycan

- A microarray binding screening showed more intense interaction of Malectin with the di-glucosylated N-glycan than with the pentasaccharide that mimics the N-glycan D1 arm. The analysis of Malectin (Pinheiro et al., unpublished) revealed the possibility of tyrosine 199 and 200, present in the loop close to the binding pocket residues, mediating the interaction with the remaining N-glycan. Therefore, we decided to perform a double-mutation of the two residues to alanine.

Malectin modification for the recognition of O-glucosylation in EGF-repeats

- The predicted model of hMalectin with the ligand Xylα1-3Xylα1-3Glc reveals that Ser80 has a weaker interaction with the ligand due to the absence of the OH'6 on first xylose residue (Pinheiro et al., unpublished). Thus, we decided to mutate Ser80 to glutamine, a bulkier residue with an available hydroxyl to interact with xylose. This modification has two purposes: to abolish or weaken the interaction with di-glucosylated ligands and to strengthen the interaction with Xylα1-3Xylα1-3Glc.

3.2 Results and discussion

As previously described, to study the effects of abolishing some interactions of malectin with its ligand, we decided to modify several target residues in the malectin globular domain. We obtained these mutations through site-directed mutagenesis and then purified the mutated proteins for interaction and stability studies.

3.2.1 Mutagenesis and sequencing

We mutated the amino acids on malectin using the Site-directed Mutagenesis technique. In this method, the DNA polymerase replicates both plasmid strands without displacing the oligonucleotide primers containing the desired mutation. Consequently, we designed for a mutation a pair of primers, with each primer complementary to opposite strands and with the desired mutation. The extension of this primers during the temperature cycling generates a mutated plasmid which will correspond to an intense band in the agarose gel. The

globular segment of malectin (amino acids [AA] 27-213; Schallus et al., 2008) is cloned into the pETM-10 vector, which has 5528 bp (Supplementary Figure 1).

The first PCR reactions contained in the reaction mixture 1.0 µL of the dNTPs solution and the electrophoresis revealed that there were no positive results since no band occurred at the expected size of 6000 bp (**Figure 3.2**).

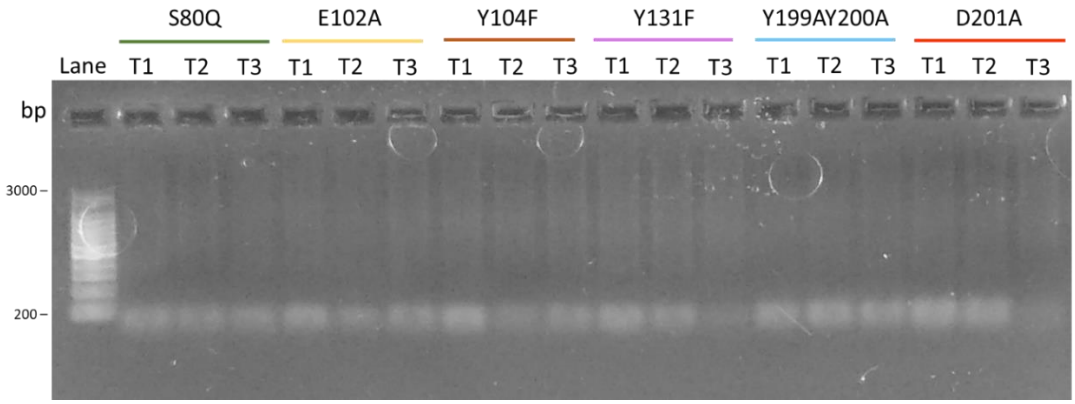


Figure 3.2 –Electrophoresis of the PCR reaction mixtures in an 0.8% (w/v) agarose gel. The three annealing temperatures are represented as T1 – 55 ºC; T2 – 57.4 ºC; T3 – 59 ºC. Each sample well was loaded with 5 µL of one PCR reaction mixture. The gels ran at a constant voltage of 50 V for 50 min and were visualized by Transilluminator (VWR). 200 bp DNA Ladder (Bio-TEK, Omega) was used as a molecular marker (Lanes).

The 10x reaction buffer has the ideal magnesium ion concentration for the polymerase used in this protocol. This Mg^{2+} acts as a cofactor for the enzyme that catalyses the phosphodiester bond formation between the 3'-OH of the primer oligonucleotide and the phosphate group of a dNTP. In addition, magnesium ions bind to dNTPs and reduces their availability for incorporation. So, therefore the free Mg^{2+} concentration can be affected by the dNTP concentration. If the dNTPs exceed optimal concentration it can inhibit PCR reaction (McPherson & Moller, 2006). In addition, lower concentrations of dNTPs decrease the frequency of mispair extension, as it affects the proofreading ability of exonuclease-proficient polymerases (Clayton & Branscomb, 1979; Cline et al., 1996). Thus, we decided to test a lower concentration of dNTPs, reducing the volume to 0.5 µL (as described in section 2.2). To test this new PCR reaction condition, two mutations with a single nucleotide exchange, E102A and Y104F, were selected. From the band location in the agarose gel (above 3000 bp; **Figure 3.3**), we concluded that the strong and intense bands achieved at annealing temperatures of 56 ºC corresponded to the vector containing the malectin gene with the target mutation.

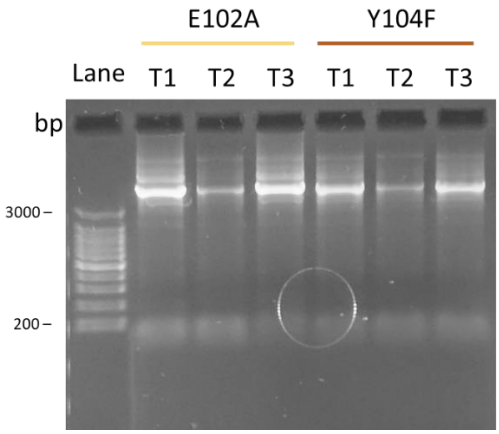


Figure 3.3 – Electrophoresis of the PCR mixtures for the malectin mutations E102A and Y104F in an 0.8% (w/v) agarose gel. The three annealing temperatures are represented as T1 – 56 ºC; T2 – 60.8 ºC; T3 – 64 ºC. Each sample well was loaded with 5 µL of one PCR reaction mixture. The gels ran at a constant voltage of 50 V for 50 min and were visualized by Transilluminator (VWR). 200 bp DNA Ladder (Bio-TEK, Omega) was used as a molecular marker (Lanes).

Although we had optimized the PCR reaction buffer, we still needed to adjust the annealing temperatures for the remaining mutant, which depends on the DNA sequence of the region to be mutated and of the primers used. Therefore, we tested several annealing temperatures in the PCR parameters for the other mutants. We analysed the PCR products (**Figure 3.4**) and established the annealing temperatures indicated in table 3.1, except for the double mutant Y199AY200A.

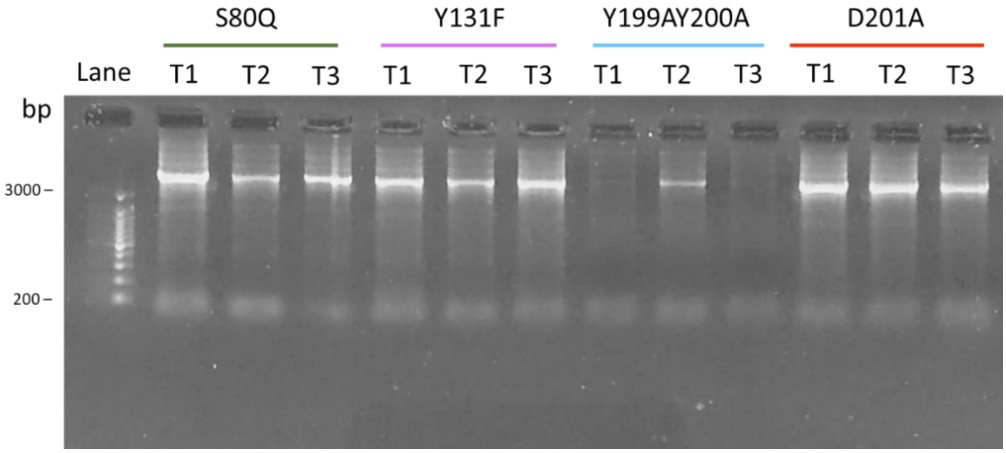


Figure 3.4 – Electrophoresis of the PCR mixtures for the malectin mutations S80Q, Y131F, Y199AY200A and D201A in an 0.8% (w/v) agarose gel. The three annealing temperatures are represented as T1 –55 ºC; T2 – 56.2 ºC; T3 – 57 ºC. Each sample well was loaded with 5 µL of one PCR reaction mixture. The gels ran at a constant voltage of 50 V for 50 min and were visualized by Transilluminator (VWR). 200 bp DNA Ladder (Bio-TEK, Omega) was used as a molecular marker (Lanes).

Table 3.1 – Optimized annealing temperatures for each mutation.

Mutation	Annealing Temperature (ºC)
E102A	56
D201A	57
Y104F	56
Y131F	57
S80Q	57

The mutation of tyrosine 199 and 200 to alanine required the modification of four nucleotides simultaneously. Two primers (Forward and Reverse) were designed with the double mutation, with a length of 51 nucleotides each (Supplementary table 4). Therefore, although the same mutagenesis technique was used (described in Section 3.2), further attempts were needed to reach the right annealing temperature for this PCR reaction. Thus, several temperatures were tested to allow primers to bind specifically to the template. Tested annealing temperatures are shown in table 3.2.

Table 3.2 – Annealing temperatures tested in the mutagenic PCR for obtaining the Y199AY200A mutated malectin.

Mutation	Annealing Temperature (ºC)
Y199AY200A	52
	53.2
	54
	55
	56.2
	57
	57.4
	59

We analysed the reaction mixtures of the attempted annealing temperatures, and an intense band was observed above the 3000 bp at temperature 52 ºC, indicating that these primers need a lower temperature to anneal to the sequence of the template plasmid. Then, all mutants proceeded to DNA sequencing to confirm the mutation.

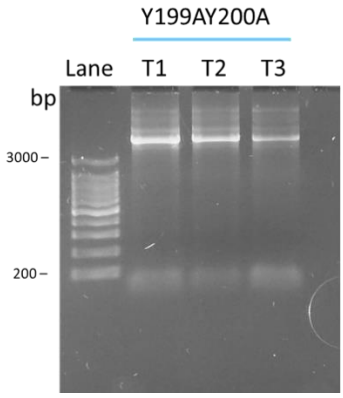


Figure 3.5 – Electrophoresis of the PCR mixtures for the malectin mutation Y199AY200A in an 0.8% (w/v) agarose gel. The three annealing temperatures are represented as T1 – 52 ºC; T2 – 53.2 ºC; T3 – 54 ºC. Each sample well was loaded with 5 µL of one PCR reaction mixture. The tested annealing temperature is indicated above the well. The gels ran at a constant voltage of 50 V for 50 min and were visualized by Transilluminator (VWR). 200 bp DNA Ladder (Bio-TEK, Omega) was used as a molecular marker (Lanes).

As described in Chapter 2, after the DNA analysis by agarose gel, *Dpn I* digestion, DH5α transformation and DNA extraction by miniprep, the DNAs containing the desired mutation were sent for Sanger sequencing at the STAB VIDA services (Supplementary table 3). We confirmed the presence of the desired mutations in the malectin gene sequences which were then translated to peptide sequences using the bioinformatic tool Transeq (table 3.3; (Madeira et al., 2019).

Table 3.3 – Protein sequences of the mutated malectin proteins. Mutated residues are highlighted in *red* and the *N*-terminal His-tag is underlined.

Protein	Protein Sequence
hMalectin	MKHHHHHHPMAGLPESVIWAVNAGGEAHVDVHGIHFRKDPLEGRVGRASDYGMKL PILRSNPEDQILYQTERYNEETFGYEVPIKEEGDYVLVLKFAEVYFAQSQQKVFDVRLNG HVVVKDLDIFDRVGHSTAHDEIIPMSIRKGKLSVQGEVSTFTGKLYIEFVKGYYDNPKVC ALYIMAGTVDDVPKLQPHPGLE
E102A	MKHHHHHHPMAGLPESVIWAVNAGGEAHVDVHGIHFRKDPLEGRVGRASDYGMKL PILRSNPEDQILYQTARYNEETFGYEVPIKEEGDYVLVLKFAEVYFAQSQQKVFDVRLNG HVVVKDLDIFDRVGHSTAHDEIIPMSIRKGKLSVQGEVSTFTGKLYIEFVKGYYDNPKVC ALYIMAGTVDDVPKLQPHPGLE
D201A	MKHHHHHHPMAGLPESVIWAVNAGGEAHVDVHGIHFRKDPLEGRVGRASDYGMKL PILRSNPEDQILYQTERYNEETFGYEVPIKEEGDYVLVLKFAEVYFAQSQQKVFDVRLNG HVVVKDLDIFDRVGHSTAHDEIIPMSIRKGKLSVQGEVSTFTGKLYIEFVKGYYANPKVC ALYIMAGTVDDVPKLQPHPGLE
Y104F	MKHHHHHHPMAGLPESVIWAVNAGGEAHVDVHGIHFRKDPLEGRVGRASDYGMKL PILRSNPEDQILYQTERFNEETFGYEVPIKEEGDYVLVLKFAEVYFAQSQQKVFDVRLNG HVVVKDLDIFDRVGHSTAHDEIIPMSIRKGKLSVQGEVSTFTGKLYIEFVKGYYDNPKVC ALYIMAGTVDDVPKLQPHPGLE

Y131F	MKHHHHHHPMAGLPESVIWAVNAGGEAHVDVHGIHFRKDPLEGRVGRASDYGMKL PILRSNPEDQILYQTERYNEETFGYEVPIKEEGDYVLVLKFAEVFFAQSQQKVFDVRLNG HVVVKDLDIFDRVGHSTAHDEIIPMSIRKGKLSVQGEVSTFTGKLYIEFVKGYYDNPKVC ALYIMAGTVDDVPKLQPHPGLE
Y199A200A	MKHHHHHHPMAGLPESVIWAVNAGGEAHVDVHGIHFRKDPLEGRVGRASDYGMKL PILRSNPEDQILYQTERYNEETFGYEVPIKEEGDYVLVLKFAEVYFAQSQQKVFDVRLNG HVVVKDLDIFDRVGHSTAHDEIIPMSIRKGKLSVQGEVSTFTGKLYIEFVKGAADNPKVC ALYIMAGTVDDVPKLQPHPGLE
S80Q	MKHHHHHHPMAGLPESVIWAVNAGGEAHVDVHGIHFRKDPLEGRVGRAQDYGMK LPILRSNPEDQILYQTERYNEETFGYEVPIKEEGDYVLVLKFAEVYFAQSQQKVFDVRLN GHVVVKDLDIFDRVGHSTAHDEIIPMSIRKGKLSVQGEVSTFTGKLYIEFVKGYYDNPKV CALYIMAGTVDDVPKLQPHPGLE

3.2.2 Malectin wild-type and mutant proteins expression and SDS-PAGE analysis

After successfully obtaining the malectin genes containing the planned mutation for this study, expression of the corresponding proteins was performed to carry out an interaction screening analysis using carbohydrate microarrays.

For the expression and production of the wild-type human malectin and corresponding mutants, we used the *E. coli* BL21 (DE3). A small-scale expression test was performed to verify whether the conditions of cell growth, protein induction and cell lysis of the malectin mutants could be the same as those of native Malectin. Thus, proteins expression was performed essentially as described in Palma et al., 2010, with the necessary adaptations (Chapter 2).

Small-scale expression tests were performed for each protein in order to verify the level of malectin present in the soluble fraction. So, after expression by IPTG-induction, cells were harvested and lysed, and the results were monitored using SDS-PAGE analysis.

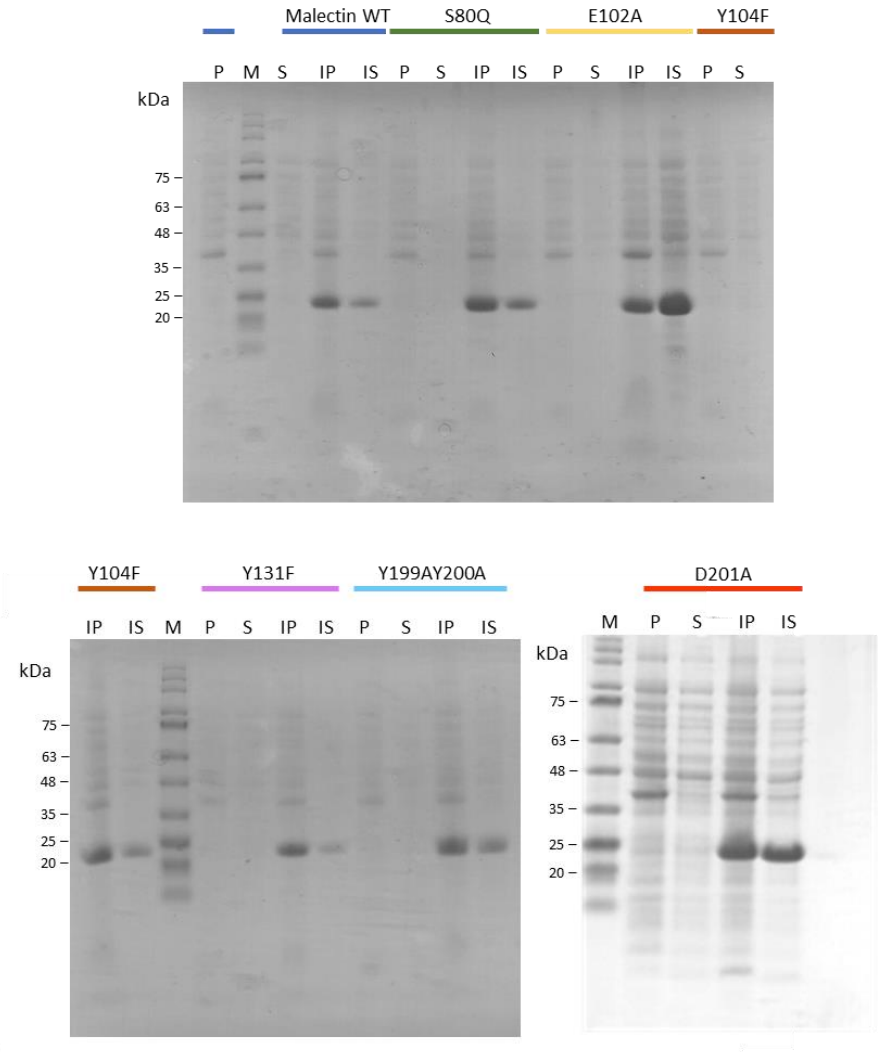


Figure 3.6 – SDS-PAGE (10% acrylamide) results showing expression levels of all proteins under study. The respective protein lanes are indicated above the well. P – *Pellet* (insoluble fraction); S – soluble fraction; IP – *Pellet* inducted by IPTG; IS – Soluble fraction induced by IPTG; M – Marker II from NZYTech®.

Figure 3.6 shows the results of the SDS-PAGE for small-scale solubilization tests. The 7 proteins were analysed, and with one exception (hMalectin E102A), all proteins came up with good expression. Similarly, to the WT malectin, a strong band in the insoluble fraction was observed in the SDS gel for the new mutants. However, we considered the quantity of protein present in the soluble fraction of the new mutants as sufficient for the following experiments. Thus, the expression protocol was approved to all proteins, however some optimizations were made to achieve a high level of expression of all proteins in the soluble fraction. The low yield in the soluble part may have occurred because the incubation step with the lysozyme described in the literature (Palma *et al.*, 2010) was not performed. This enzyme catalyses the hydrolysis of the β-(1→ 4) glycosidic linkage between N-acetylmuramic

acid and N-acetylglucosamine. This specific activity of lysozyme against peptidoglycan present in bacterial cell walls used in the correct concentrations helps to promotes cell lysis. Since the native protein was also retained in the insoluble fraction, so the cell lysis performed may not had been efficient. Hence, the freezing step at -80 ºC of the harvested cells was added to the protocol, and the number of sonication cycles were increased as described in Section 2.8. We performed another small-scale solubilization test, and the SDS-gels confirmed the increase in the level of expression on soluble fraction. So, we proceeded with this new protocol and the proteins were produced on a larger scale and then purified.

The proteins were purified in a single purification step by IMAC. The soluble fractions were loaded into a His GraviTrap™ (GE-Healthcare) column according to the manufacture's protocol, and buffers with increasing concentration of Imidazole were used (table 3.4). The molecular weight of native Malectin is ~ 20 kDa, so all fractions collected were analysed by SDS-PAGE to search for an intense band in that range. As expected, the protein eluted in the fraction containing the highest concentration of Imidazole, showing an intense band between 20 kDa and 25 kDa for all purified proteins (**Figure 3.7**). Each protein expressed with good yield and without significant degradation.

Table 3.4 – Buffers used in the IMAC purification of all proteins.

Buffer	Imidazole	Buffer composition
Equilibrium	10 mM	20 mM Tris-HCl, 150 mM NaCl, 2 mM β-mercaptoethanol, protease inhibitors, pH 8.0
Wash	60 mM	
Elution	250 mM	

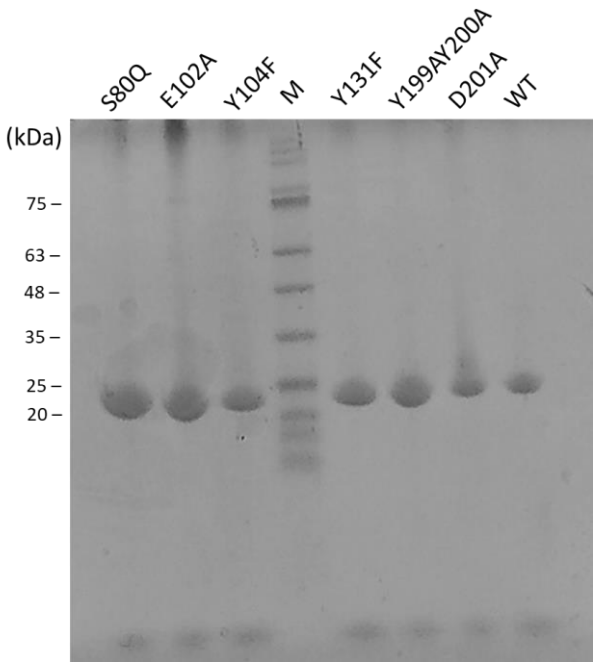


Figure 3.7 – SDS-PAGE (10% acrylamide) results showing purification levels of all proteins under study. Results of purification with His GraviTrap columns. Proteins were eluted with 2,5 mL of 20 mM Tris-HCl, 150 mM NaCl, 250 mM Imidazole, 2 mM β-mercaptoethanol, protease inhibitors, pH 8.0. In each lane are indicated the corresponded purified protein and all bands corresponds to 12 µL of the elution fraction. Marker II from NZYTech® was used (M).

After the purification, these same fractions were submitted to a desalting protocol as described in Section 2.10. We collected a final elution fraction of 3.5 mL for the WT malectin and mutant proteins. Before the DSC technique, we performed an SDS-PAGE analysis with 2µg/protein to check for major differences (**Figure 3.8**).

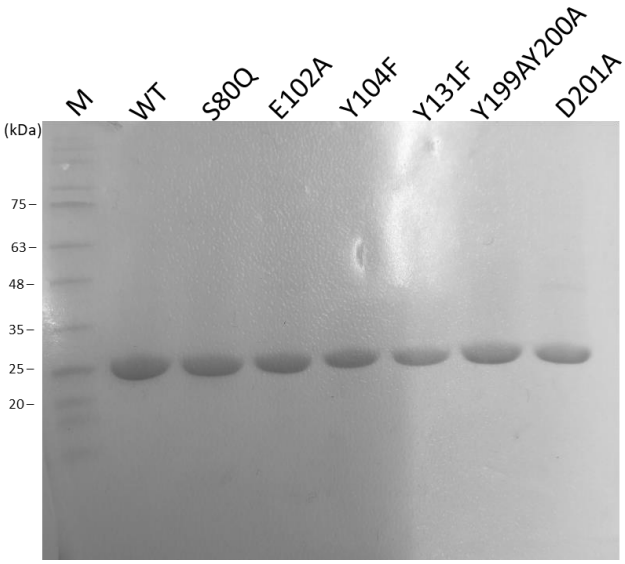


Figure 3.8 – SDS-PAGE (10% acrylamide) results after proteins desalting. Results of buffer-exchanges with PD-10 Desalting columns. The fraction containing the protein was eluted with 3.5 mL of 20 mM Tris-HCl, 150 mM NaCl, 2 mM DTT. In each lane are indicated the corresponded purified protein and all bands corresponds to 2 µg of protein. Gel was performed with a constant voltage of Marker II from NZYTech® was used (M).

Figure 3.8 shows an intense band in the expected range disclosing high purity for all proteins. To perform the following assays with a similar protein concentration, malectin and mutants were quantified as described in Section 2.10. The theoretical parameters were calculated with Expasy's bioinformatics tool ProtParm (Gasteiger et al., 2005) and the results are summarized in table 3.5.

Table 3.5 – Protein quantification after buffer-exchange. The absorbance median was used, and the MW and Epson value were calculated with Bioinformatical tools.

Protein	Absorbance	Molecular Weight	ε ($M^{-1}cm^{-1}$)	Concentration (µg/µL)
WT	0,211	22334,4	20400	4.620
E102A	0,369	22276,36	20400	8.059
D201A	0,449	22290,39	20400	10.140
Y104F	0,284	22318,4	18910	6.680
Y131F	0,209	25462	18910	5.628
Y199AY200A	0,233	22150,2	17420	5.925
S80Q	0,292	22375,45	20400	6.406

3.2.3 Differential Scanning Calorimetry data analysis

DSC studies were performed to determine the effect of the mutations in the malectin structure and stability. The two-state model was used to analyse the presence of discrete macrostates of native and unfolded protein conformation in our protein solutions.

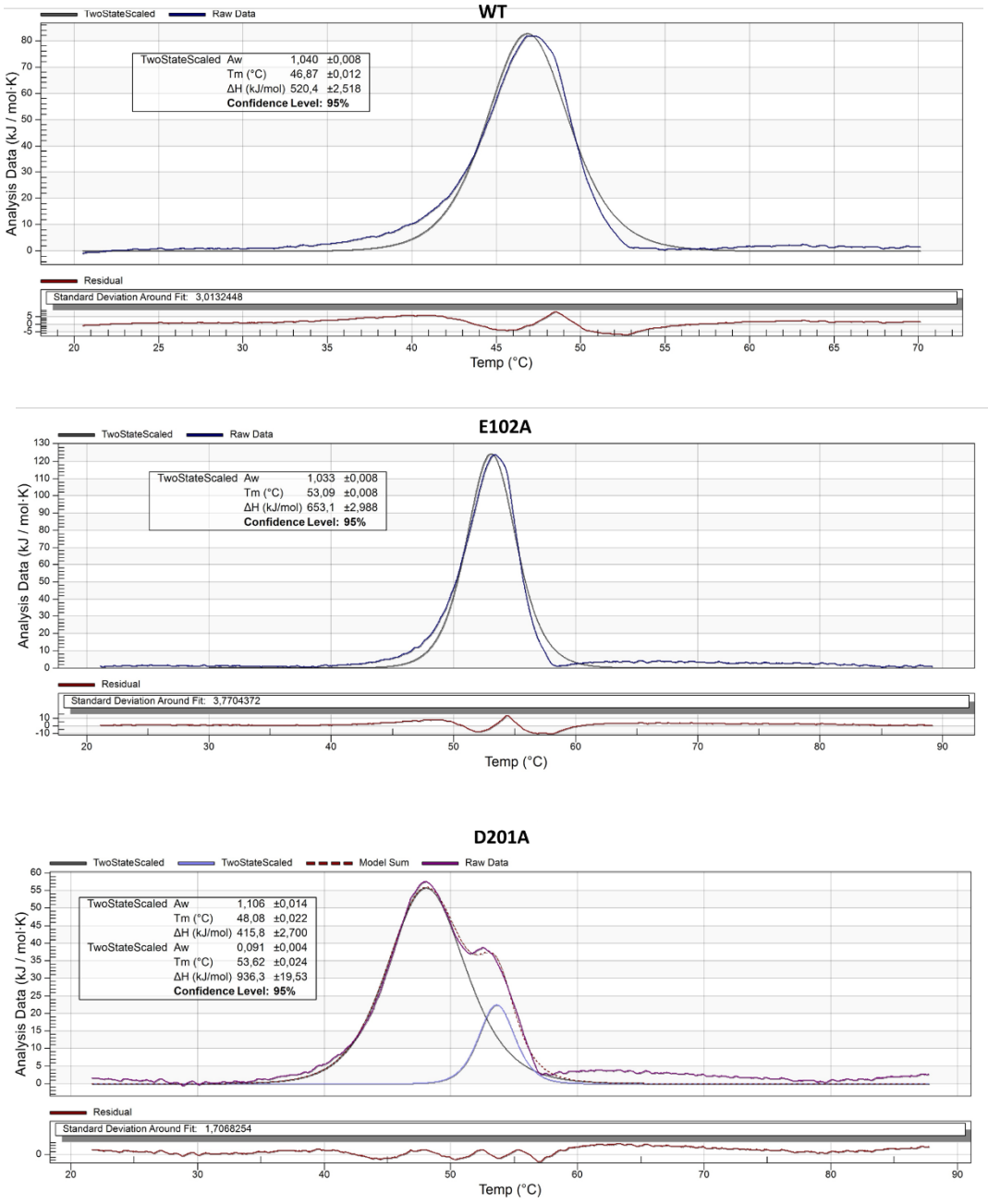


Figure 3.9 **(see legend on next page).**

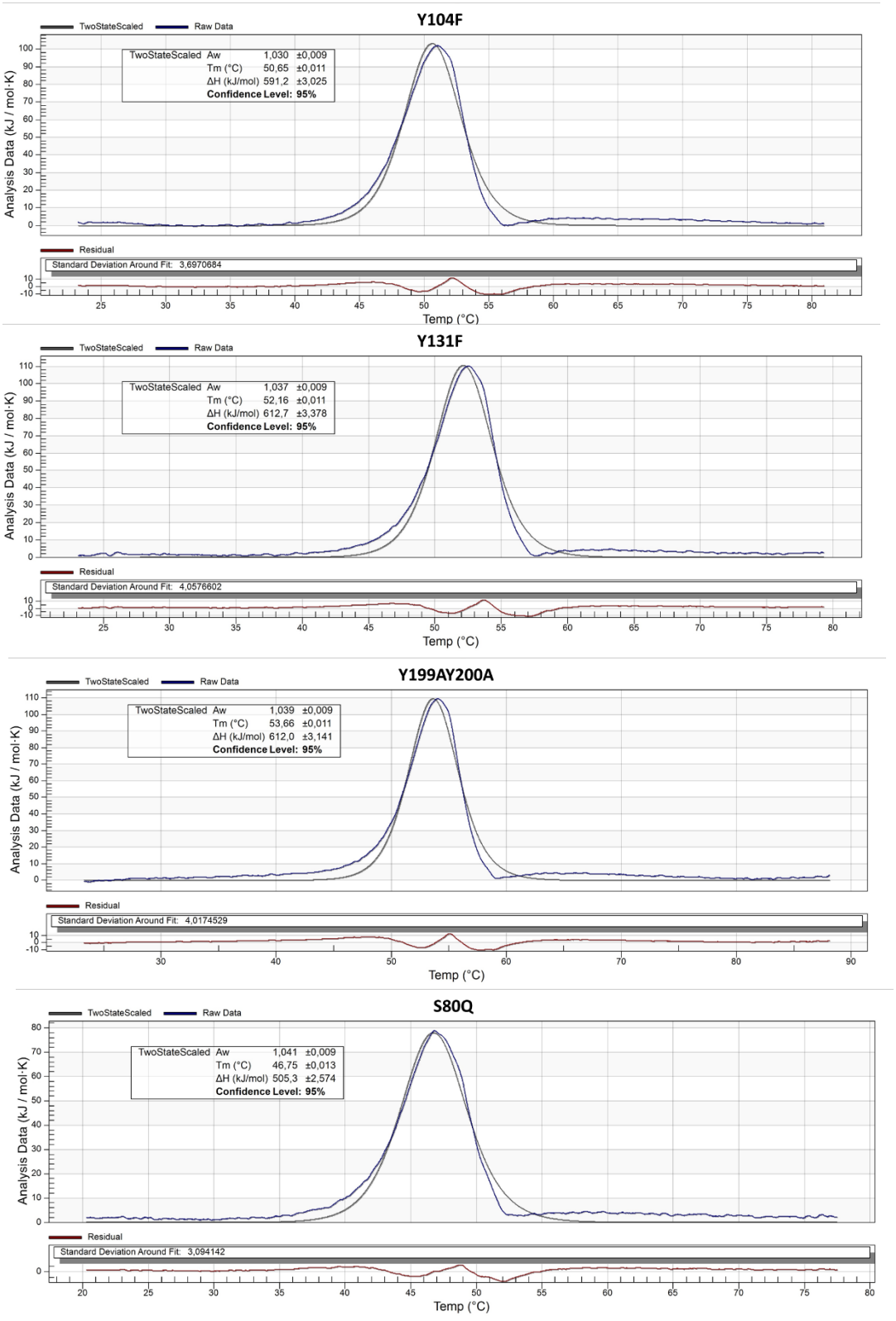


Figure 3.9 – DSC thermograms of human malectin and mutants E102A, D201A, Y104F, Y131F, Y199AY200A and S80Q. The proteins were analysed in 20 mMTris–HCl, 150 mM NaCl, 2 mM DTT and the data was fitted in a NanoAnalyze^{TM}v3.10.0 Data Analysis software. Each thermograms is composed by two graphs. The upper graph shows at least two curves: a curve of the energy absorbed (KJ/mol^{-1}K) during the temperature gradient (raw data) and a curve corresponding to the fitted two-state model used to calculate the thermodynamic values (two state scaled). The graph below shows the quality of fitting of the model to the raw data.

Apart from one protein, all thermograms showed a typical melting curve, with low absorption of energy that increases with the temperature increment and decrease after protein unfolding (**Figure 3.9**). The fits to the two-state model are good and allows to determine the thermodynamic parameters depicting the unfolding energetics.

The exception is the thermogram of the D201A protein, which unlike the other proteins, reveals two peaks. This second peak in the thermogram may be due to several factors. The two-state model assumes that the native and unfolded states are thermodynamically distinct, but between these states there are several unstable microstates with intermediate conformational properties. This second peak could be a high free energy barrier reflected by several microstates with partial-unfolded conformation, indicating that this protein has conformation heterogeneity and/or progressive melting of the structure (Ibarra-Molero et al., 2016). Besides interacting with the ligand through two direct hydrogen bonds, D201 makes a direct hydrogen bond (2.81 Å) to K138, which belongs to the loop containing Y131. The mutation of D201 to alanine breaks the interaction with K138, which in its turn can become free to establish interactions with other residues and make the protein loops adopt different conformations from the presented in malectin WT structure. Further studies are necessary to conclude about this second peak on D201A thermogram, however we decided to proceed to the interaction studies since we obtained an apparent more stable protein. Nevertheless, we will take in account the effects of these structural alterations when discussing the glycan-microarray interaction results.

The estimated Tm value (table 3.6), based on the DSC analysis, suggests that the denaturation of these mutated proteins begins in the range of melting transitions between of 40 to 55 ºC. The thermodynamic parameters of native malectin and mutated proteins were compared, and there was no decrease in the melting temperature, which indicates that the mutations produced in the native protein did not destabilize its structure. However, in the case of E102A, Y131F and Y199AY200A proteins, the estimated Tm value had a considerable increase when compared to malectin WT.

Table 3.6 – Melting temperature estimated by DSC.

Protein	WT	E102A	D201A	Y104F	Y131F	Y199AY200A	S80Q
Tm (ºC)	46.87	53.09	48.08	50.65	52.16	53.66	46.75

The increase in the melting temperature may be due to stabilization of the loops in the protein structure. In liganded human malectin structure, E102 is located inside the binding-pocket. The sidechain of this residue doesn´t make any contacts apart the two hydrogen bonds

with the sugar ligand. However, substitution of a bulky side chain with a carboxylic group by a hydrophobic methyl may make the binding-pocket conformation more stable and more compact, requiring more energy to denature. The same phenomenon happens with tyrosines 131 and 199-200, which are aromatic residues present in the large loops and exposed to the solvent. The heat capacity change of proteins during unfolding is originated by the exposure of hydrophobic groups in the denaturated state. This may be the reason why the mutation of these tyrosines to alanine resulted in an increase in the melting temperature of these mutated proteins. As previously, we decided to proceed the interaction studies with the mutants, with the awareness that these may have differences in the binding pocket structure that will probably affect their interaction with the ligands.

3.3 Conclusions

The first step in this project was to design and produce new proteins using the mutagenesis technique and, expression and purification of recombinant proteins. Despite the challenging task, the native Malectin was mutated at seven sites (six single mutations and one double mutation) and the DNA sequences were confirmed. The process of expression by induction with IPTG was identical for all proteins and the purification corresponded to a good yield.

The second part of this work was aimed at studying the stability of these proteins. All proteins had thermodynamic parameters that indicate that there was no loss of stability when compared to the native protein. However, some major structural changes may have occurred.

In conclusion, this part of the project was successfully achieved, thus we proceed with the analysis of interactions with glycans.

Chapter 4 – Carbohydrate microarrays analysis of the specific Malectin-ligand interactions

4.1 Introductory remarks

The aim of this chapter was to determine the importance of residues from the malectin binding-pocket and of the two putative binding residues for the specificity of malectin towards its ligand and analyse the specificity of the malectin mutants. To fulfil these aims, we analysed the malectin WT and mutants using the carbohydrate microarray technology, which is a powerful tool for evaluating glycan-mediated recognition events in a high-throughput manner. In this Chapter, the results of the initial screening analysis performed within the time frame of the Thesis will be presented and discussed. Although this analysis is still ongoing, the carbohydrate microarray results analysed confirmed the binding specificity of the malectin wild-type protein, and allowed initial assessment of the effect of inhibiting some interactions, and to evaluate structure-function relationships of individual amino acid residues.

First, the carbohydrate-binding of the WT malectin and the effect of the designed mutations in the interaction with glucose-based ligands were confirmed and accessed by performing a first screening in a polysaccharide microarray (in house designation of Fungal, bacterial & plant PS microarray Set 2). Among the polysaccharide samples there are 3 major groups: glucans and hemicelluloses, pectins and the glycans derived from bacteria or fungi (constitution of the printed microarray set is in Supplementary Table 7). Among these is the α1-6-glucan (dextran), which is specifically bound by malectin. Then, a second screening was performed in a sequence-diverse oligosaccharide microarray to gain further insight into the interactions that give malectin specificity. The oligosaccharide microarray is comprised of more than 900 neoglycolipid (NGL) probes with a substantial structural and chemical diversity (in house designation of microarray *Sets 42-56*; Supplementary table 8). This glycan microarray contains a wide range of mammalian type sequence present on glycoproteins (N-glycans and O-glycans), glycosaminoglycans and glycolipids; among these are the di-glucosylated high-mannose N-glycans that are specifically bound by malectin. The list of the diverse probes can be accessed through the link of the carbohydrate microarray facility at Imperial College London (https://glycosciences.med.ic.ac.uk/glycanLibraryList.html).

The glycan microarrays used for these studies were constructed using a noncontact arrayer robot with a spot delivery on glass slides coated with 16-pad nitrocellulose. The probes were immobilized noncovalently on the slide and in duplicate printing with two levels of concentration.

As processing of the complete full microarray data in *Sets 42-56* is still ongoing, but the major results will be presented and discussed in this Chapter. To do this, the microarray data

presentation will be focused on the α1-6-glucan (dextran) polysaccharide in the first screening and on the Glc2Man3 pentasaccharide that comprises the di-glucosylated N-glycan D1 arm in the second screening. Also, we analysed the binding intensity of malectin WT and S80Q with the trisaccharide that mimics the Notch 1 O-glucosylation, a probe that is printed in the microarray Sets 42-56. To systematize the information, the results are presented according to the three objectives outlined for this project: 1) Characterization of the Malectin binding pocket interactions; 2) Role of two putative binding tyrosine residues in the interaction with the di-glucosylated high mannose N-glycan; 3) Malectin modification for the recognition of O-glucosylation in EGF-repeats.

4.2 Characterization of the Malectin binding pocket interactions

4.2.1 Results and discussion

After production and analysis of the structural stability study, the wild type and E102A, D201A, Y104F and Y131F mutated human malectin proteins were prepared for testing in the microarrays assays as described in Section 2.13. These four proteins were designed to understand and characterize interactions that are crucial for the specificity of human malectin. The polysaccharide microarray screening analysis with the wild-type and malectin mutants are presented in **Figure 4.1**.

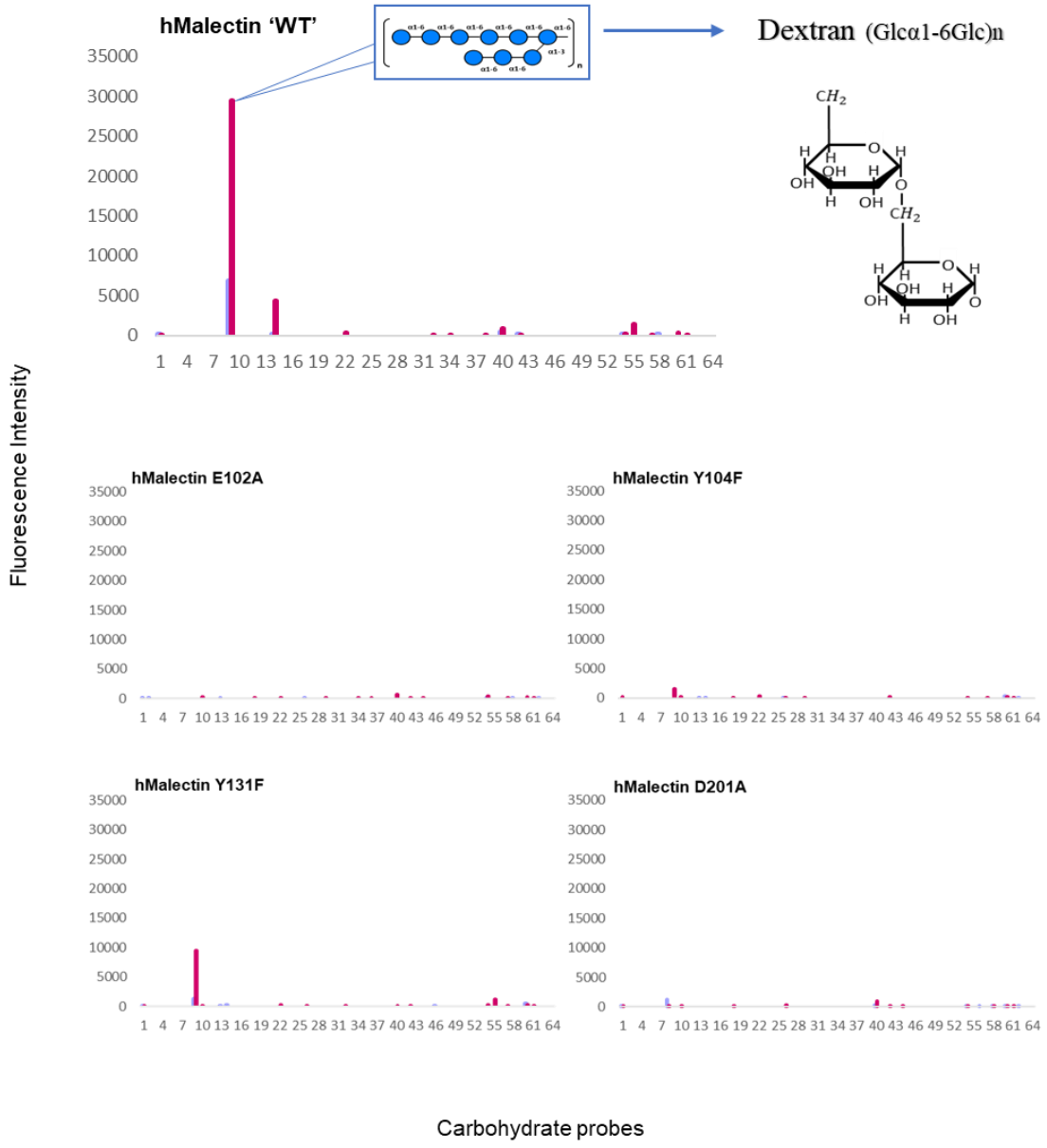


Figure 4.1 – Polysaccharide microarray analysis of the interactions of hMalectin WT and E102A, D201A, Y104F and Y131F proteins (*Fungal, bacterial & plant PS microarray Set 2*). The binding signals are depicted as means of fluorescence intensities of duplicate spots, elicited with 30 and 150 pg polysaccharide per spot (blue and purple bar, respectively).

Among the diverse repertoire of polysaccharides printed in the microarray, human malectin WT showed, as predicted a stronger binding signal to the dextran polysaccharide. This result is in accord with the capacity of human malectin bind to other glucose disaccharides such as isomaltose (Glcα1-6Glc), which is the main building block of dextran, a higher oligomer of glucose. Although the reported interaction with isomaltose was much lower when compared to the other glucose disaccharides (nigerose and maltose) that showed interaction (Schallus et al.,

2008), we expect the residues involved in the interaction to be the same. The X-ray structures of human-malectin complexed with maltose and nigerose showed that amino acid residues mediating the interaction are identical for the two ligands (Pinheiro et al., unpublished). In addition, the chemical shift perturbations in 1H15N-HSQC spectra of xenopus malectin with several glucose disaccharide, analogs of nigerose were also in similar areas (Schallus et al., 2010). None of the malectin mutants showed a change in the specificity to other types of polysaccharide sequences. We were anticipating, however, effects of the mutants on the binding to dextran, resulting in a decrease in the signal intensity of the interaction for this polysaccharide. **Figure 4.2**. shows a comparison of the fluorescence intensities obtained from the interaction of the mutants and WT malectin with dextran.

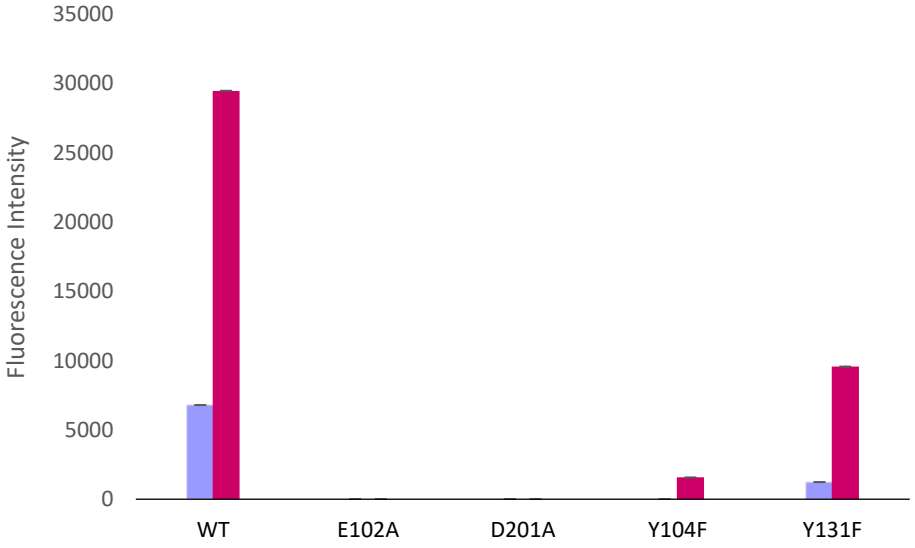


Figure 4.2 – Comparison of the interactions of human malectin WT, E102A, D201A, Y104F and Y131F the α1-6 glucan polysaccharide Dextran. The binding signals are depicted as means of fluorescence intensities of duplicate spots with error bars, elicited with 30 and 150 pg polysaccharide per spot (blue and purple bar, respectively).

The mutation of E102 and D201 completely abolished the interaction with the dextran, showing the importance of the two direct hydrogen bonds that each residue makes to the glucose disaccharide in the polysaccharide ends. Also, the mutation of the two tyrosine residues involved in the π/C-H stacking and in the water network altered the binding of malectin to dextran. The Y104F malectin protein showed a residual binding and the Y131F malectin protein showed three times less binding than malectin WT. Interestingly, the mutation Y104F has a greater effect in the interaction of human malectin with the dextran polysaccharide than the mutation Y131F. A possible explanation could be that the contribution of Y131 for the interaction of malectin with

disaccharides other than nigerose is diminished. The crystal structure of the human malectin complexed with nigerose showed the perfect stacking of the two glucose residues with Y104 and Y131. This was not observed in the structure of malectin complexed with maltose where the second glucose is more distant from Y131, which ultimately resulted in decreased specificity and affinity for this disaccharide. The Y131 is also expected to not stack well to the second glucose of the isomaltose and thus to have a smaller contribution to the interaction of malectin with this disaccharide and consequently with the dextran polysaccharide.

In the second screening in microarray *sets 42-56*, the human malectin WT showed its restricted binding to di-glucosylated high-mannose N-glycan probes (data being processed), in accord with the reported specificity (Schallus et al., 2008 & 2010; Palma et al., 2010), with a strong binding to the Glcα1-3Glcα1-3Manα1-3Manα1-2Man probe (Glc2Man3 pentasaccharide comprising the N-glycan D1 arm) (**Figure 4.3**). None of the malectin mutants showed a change in the glycan-binding specificity, but as predicted these exhibited different binding to the Glc2Man3 pentasaccharide probe (**Figure 4.3).**

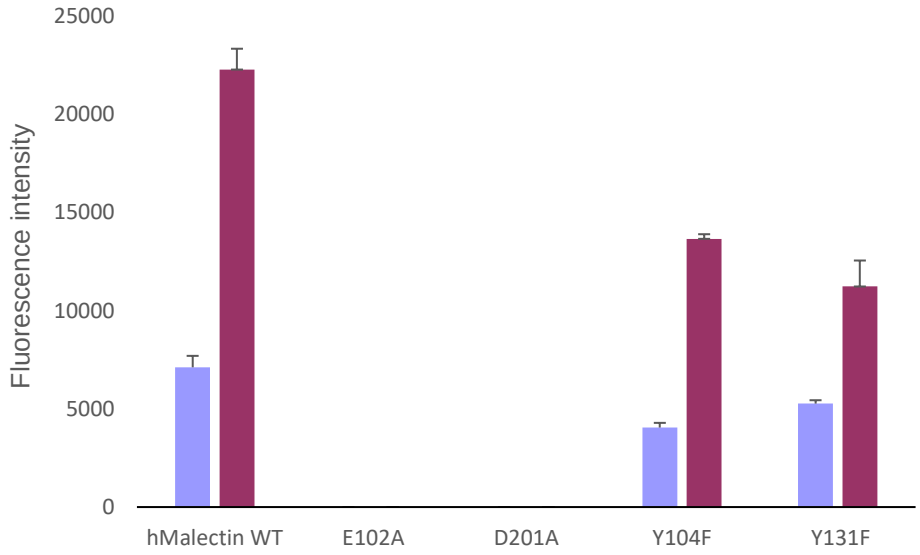


Figure 4.3 – Interactions of human malectin and designed mutants with Glc2Man3 pentasaccharide comprising the N-glycan D1 arm. The binding signals are depicted as means of fluorescence intensities of duplicate spots with error bars, elicited with 2 and 5 fmol **oligosaccharide** per spot (blue and purple bar, respectively).

As observed before in the microarray analysis using the polysaccharide set, the hydrogen bonds established by E102 and D201 revealed to be crucial for the protein-carbohydrate interaction since the mutation of these residues to alanine residues completely abolished binding. These two residues are thought to anchor and direct the glycan in the right orientation for the π/C-H stacking interaction with Y104 and Y131(Schallus et al., 2010; Pinheiro et al, unpublished). So

removing these hydrogen bonds most probably made the remaining interactions less favourable to occur. Contrary to what happened in the polysaccharide set, the mutation of the Y104 and Y131 to phenylalanine residues had less effect in the interaction with the G2M3 oligosaccharide. Since the G2M3 oligosaccharide is similar to the true malectin ligand, the di-glucosylated N-glycan, the binding pocket establishes several interactions that have probably evolved with the ligand. The effect of the mutation of one residue may be less impactful since there are still other important residues that are still making important interactions with the oligosaccharide.

In previous work from our group, the importance of the water network in the malectin-ligand interactions for the interaction with the G2M3 oligosaccharide was discovered. We hypothesize that removing the hydroxyl groups (tyrosines mutation to phenylalanines) would decrease the strength of the interaction of malectin with the G2M3 oligosaccharide. This may be the reason why a major effect was seen when mutating Y131, that establishes more water mediated hydrogen bonds with G2M3, than Y104. We believe that the mutation Y131F may be obliterating the three water-mediated hydrogen bonds with MAN3 and MAN4. However, we acknowledge that the mutation of these residues to phenylalanines may also be influencing the π/C-H stacking established with the two glucose residues from the reducing terminal (Pinheiro et al., unpublished), since tyrosine residues are more likely to participate as π/C-H acceptor over the phenylalanine residues, due to their greater electronegativity (Hudson et al., 2015).

4.3 Role of two putative binding tyrosines in the interaction with the N-glycan

4.3.1 Results and discussion

The specificity of human Malectin can go beyond binding site interactions. Two solvent exposed tyrosines (Y199 and Y200) located in loop 4 could interact with the mannose residues of the di-glucosylated N-glycan, thus increasing the specificity of this protein for this ligand. The hypothesis is that this can confer a third level of specificity, where the first level would be the binding to the Glcα1,3Glc disaccharide, the second level would be the water network until the second mannose and this third level would be the recognition of another N-glycan arm. Thus, these two tyrosine residues were mutated to alanine with the aim to test the interaction with the di-glucosylated fully mannosylated high N-glycan (Glc2Man9).

Despite the challenging task of designing this double mutant, producing the protein and assuring that the tertiary structure remained stable, we could perform the microarray assay as

described in Chapter 2. To observe the initial effects of this double mutation on the human malectin function, we performed the polysaccharide microarray screenings with the wild-type and Y199AY200A malectin, shown in **Figure 4.4**.

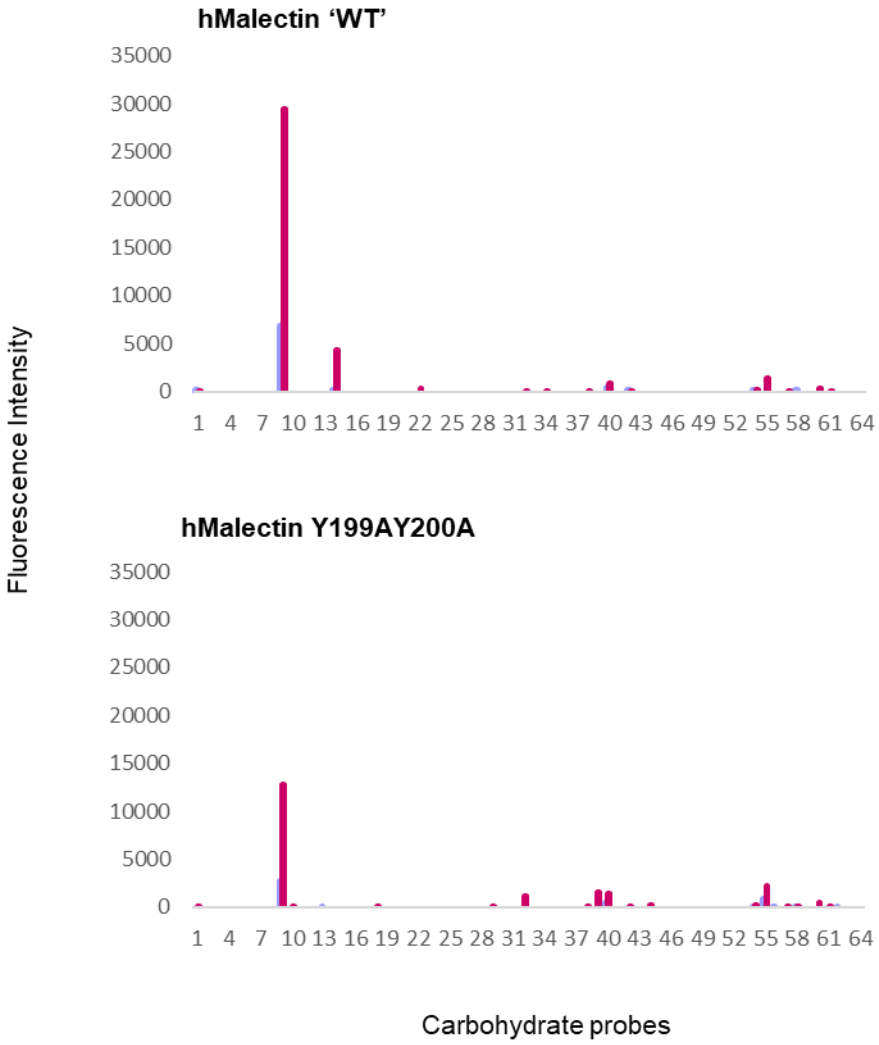


Figure 4.4 – Polysaccharide microarray analysis of the interaction of human malectin WT and Y199AY200A. The binding signals are depicted as means of fluorescence intensities of duplicate spots, elicited with 30 and 150 pg polysaccharide per spot (blue and purple bar, respectively).

The Y199 and Y200 residues are sited outside the binding pocket, and no change in binding specificity was observed for this mutant. Interestingly the mutation of these two residues slightly decreased the interaction with dextran. So, the results of this first carbohydrate microarray analysis were unexpected. Possible explanations could be that these residues participate in the interaction with dextran or that their mutation could perhaps alter the structure of the binding pocket. This latter hypothesis is supported by the data obtained by DSC, where these mutations

increased the Tm considerably (ΔTm≈6ºC). Nevertheless, we also analysed the human malectin Y100AY200A protein in the NGL microarray sets 42-56 to see the effect of the mutation in the interactions with oligosaccharides with di-glucosyl terminals. The comparison of the Y199AY200A mutant with the human malectin WT showed a consistent moderate decrease in the binding intensity with ligands that were supposed to interact mainly with malectin binding-pocket. Depicted in **Figure 4.5** as an example is the binding to the Glc2Man3 pentasaccharide.

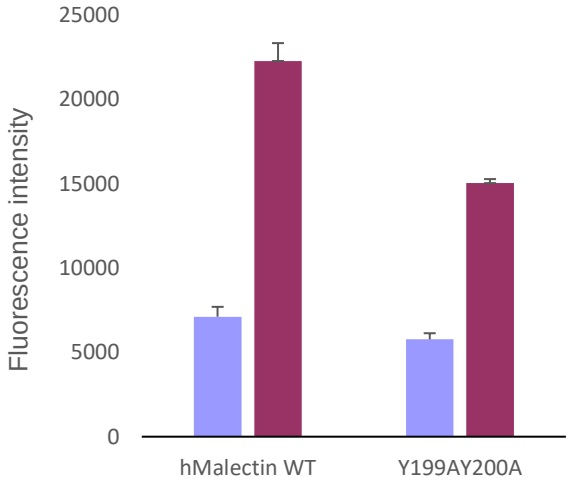


Figure 4.5 – Interactions of human malectin and Y199AY200A with Glc2Man3 pentasaccharide comprising the N-glycan D1 arm. The human malectin WT and Y199AY200A were analysed in the NGL microarray *Sets 42-56*. The binding signals are depicted as means of fluorescence intensities of duplicate spots with error bars, elicited with 2 and 5 fmol oligosaccharide per spot (blue and purple bar, respectively).

The increased Tm in the structural stability study and the microarray results indicate that this mutation may have repercussions for the protein-ligand interaction that were not initially predicted. Unfortunately, the di-glucosylated Glc2Man9 N-glycan was not printed on the microarray, and further analyses are required to conclude about the effect of this double mutation of the specificity and affinity of malectin to its ligands.

4.4 Malectin modification for the recognition of O-glucosylation in EGF repeats

4.4.1 Results and discussion

As described in Section 1.1.3, EGF repeats undergo the O-glucosylation process were the trisaccharide Xylα1-3Xylα1-3Glc is appended to a serine residue in a consensus sequence. The function of this type of glycosylation on EGF repeats is not clear, but it has been reported to occur in Notch1 mammals during the maturation process when passing through ER and Golgi. Furthermore, several studies have revealed that the dysfunction of this type of glycosylation can lead to numerous disorders or even to dead (see more details in Section 1.1.3). As this trisaccharide has a similar structure to the D1 arm of the malectin natural ligand, and occurs in the ER, we hinted to the possibility that malectin recognizes this type of glycosylation. Indeed, in an initial screening an interaction between human malectin and this trisaccharide was detected, and a model of this trisaccharide could be fitted into the density of the three first residues of the G2M3 pentasaccharide (Pinheiro et al., unpublished). When compared with the X-ray structure density, only the absence of an OH'6 from the capping xylose residue reveals that the majority of the interactions would maintain except the hydrogen bond with Ser80. So, we performed the mutation S80Q to reduce the distance between this residue and the first xylose residue. With this approach, we intended to design a new protein with a new specificity for this sugar present in EGF-repeats from Notch pathway.

To confirm the effect of this mutation, and as previously done for the other mutants we performed the analysis with the WT and S80Q malectin to the polysaccharide microarray, as shown in **Figure 4.6.**

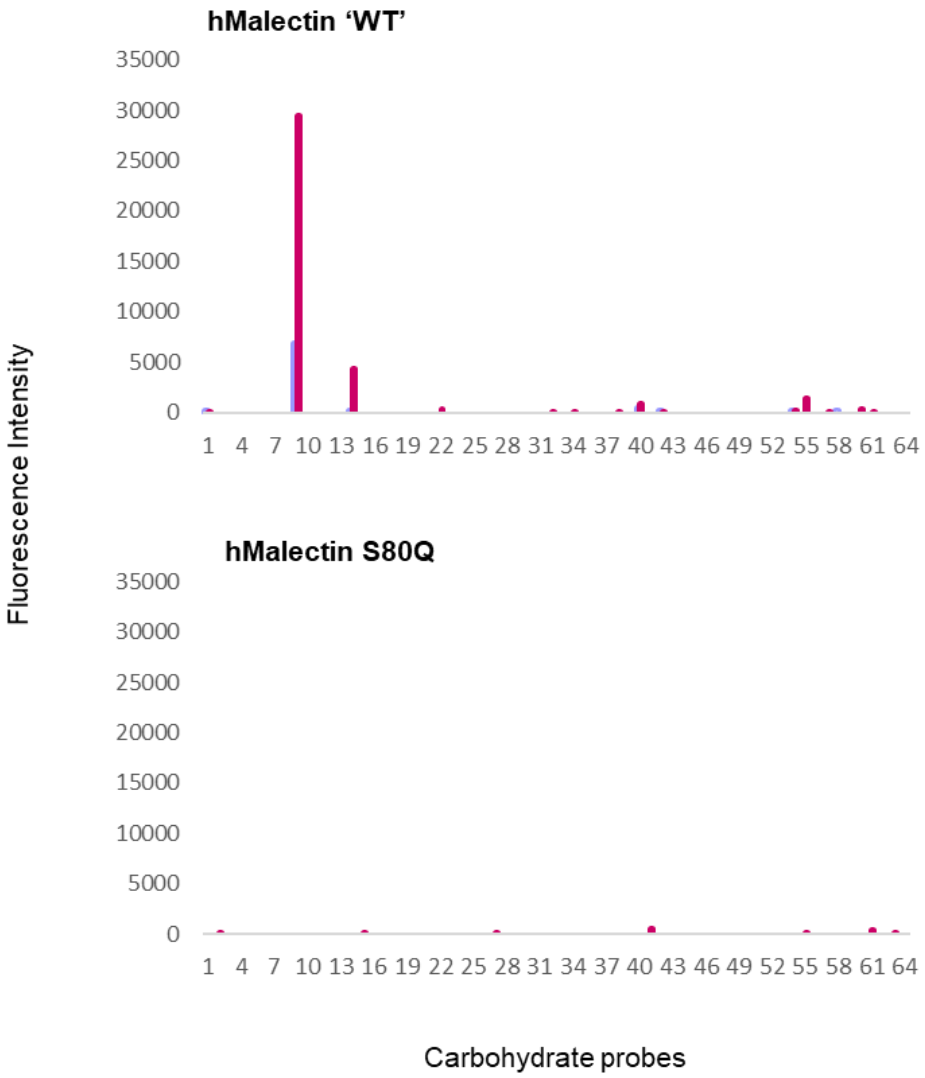


Figure 4.6 – Polysaccharide microarray analysis of the interaction of hMalectin WT and S80Q. The binding signals are depicted as means of fluorescence intensities of duplicate spots with error bars, elicited with 30 and 150 pg polysaccharide per spot (blue and purple bar, respectively).

The results of the polysaccharide microarray analysis showed that the mutation of this residue almost abolished the interaction with the dextran. Encouraged by the prospect of this result, which is the first step to modify the specificity of this protein, we performed the screening analysis in the NGL microarray *Sets 42-56*, and analysed the interaction with Glc2Man3 and Xyl2Glc1 oligosaccharides (**Figure 4.7**).

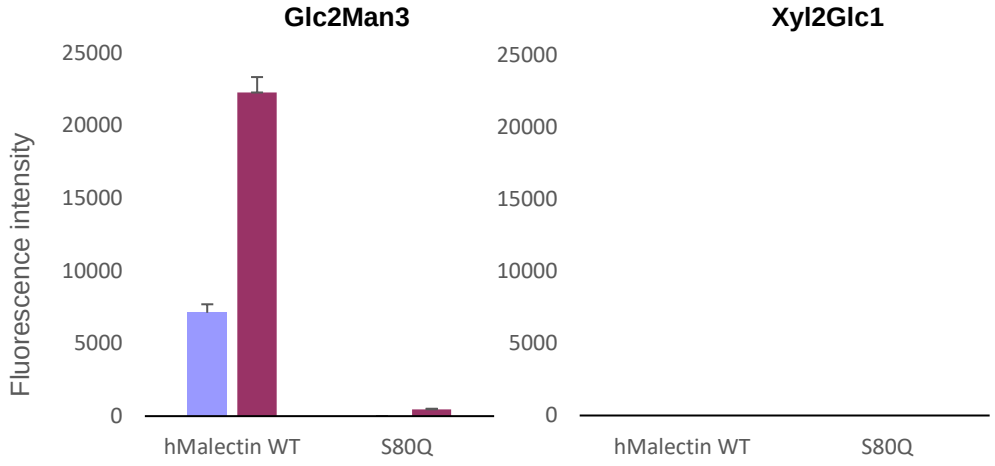


Figure 4.7 – Interactions of human malectin and S80Q with Glc2Man3 pentasaccharide comprising the N-glycan D1 arm and with Xyl2Glc1 from EGF-repeats glycosylation. The human malectin WT and S80Q were analysed in the NGL microarray *Sets 42-56*. The binding signals are depicted as means of fluorescence intensities of duplicate spots with error bars, elicited with 2 and 5 fmol oligosaccharide per spot (blue and purple bar, respectively).

The mutation to a bulkier residue in position 80, almost abolished the protein-ligand interactions. A residual binding was detected in the high level of Glc2Man3 concentration. This result suggests that the pocket has the right volume to accommodate the two glucose rings, and also supports the hypothesis of the binding-pocket "hand-glove" conformation (Pinheiro el al., in preparation). We think that the presence of a bulky amino acid decreases the volume of the binding pocket, forcing the disaccharide to be deviated from its normal position where it is not able to align perfectly with tyrosines 104 and 131. Although promising, we could not observe binding to the Xyl2Glc1 trisaccharide, with this mutant in the conditions of the analysis. However, the binding with the WT malectin could also not be reproduced in this analysis. Further analysis needs to be considered to evaluate the interaction with this type of oligosaccharides, and conclude or design other mutants for the creation of a new specificity.

4.5 Conclusions

The most relevant outcomes from the carbohydrate microarray analyses here performed were the inspection of the role of carbohydrate-interacting residues on the specificity of human malectin to N-glycan and understand the effect of the different mutation on the glycan-binding.

The specificity of this lectin was confirmed with the analysis of a variety of probes that included several glycans present in the early steps of the N-glycosylation and with the recognition of oligosaccharides with di-glucosyl terminals only.

Despite the high specificity towards to nigerose and Glcα1,3Glc containing oligosaccharides, the malectin protein recognizes also polysaccharides containing terminal di-glucosides. Since the NMR and X-ray revealed that malectin carbohydrate-interacting residues are similar for the different di-glucosylated ligands, it was possible to perform a microarray screening for making a first assessment of the effect of the designed mutations in the interaction with this unnatural ligand. The human malectin WT showed a stronger interaction with the α1-6 glucan (dextran) and the influence of the mutants was a decrease in the interaction with this polysaccharide. The mutation of the two residues (E102 and D201) that make hydrogen bonds with the glucose disaccharides resulted in the abolishment of the interaction with dextran. Comparing the influence of the two mutated tyrosine residues (Y104 and Y131), the mutation of Y131 showed less effect in the interaction of malectin with dextran. Since in this interaction the π/C-H stacking interaction of Y131 with the second glucose of the ligand has possibly a more reduced effect, mutating this residue to a phenylalanine had a lesser impact on its binding strength.

After having performed a microarray screening in the polysaccharide, we proceeded the study of the effect of the human malectin mutants on interaction with di-glucosylated ligands using an NGL-based oligosaccharide microarray. In our analysis focused on the G2M3 pentasaccharide, as already happened in the previous screening, the hydrogen bonds made by E102 and D201 emerged as crucial for the binding event. However, regarding the mutation of the two tyrosine residues (Y104 and Y131) involved in water-network, the analysis of effect on the interaction of malectin with Glc2Man3 had different outcome from what was previously observed. Although the greatest decrease in the fluorescence intensity was observed with the Y131F mutation, the effect on this mutant was similar to the repercussions on the interactions of mutating Y104 to phenylalanine. We suggest that this could be the result of the perturbation of the water-mediated interactions between tyrosine 131 and the first and second mannose of the pentasaccharide, underlining the importance of the water-mediated hydrogen bonds in the preference for the D1 arm of Glc2Man9GlcNac2.

The possibility of an extended binding site (the tyrosines 199 and 200 in loop 4) for the recognition of other portions of the N-glycan polypeptide chain could not be confirmed in the time frame of this thesis. And, given the results of structural stability and microarray analyses of this protein, further work is needed to understand the interference of this mutation in the binding pocket. Concerning to the protein with the new specificity, the mutation of Ser80 present at the

bottom of the pocket, to a bulkier residue did not showed conclusive results, further studies of the interaction of malectin with this type of glycans is required. However, with this mutation we could inhibit the malectin interaction to di-glucosylated glycans. This, in turn, allowed us to conclude that, as reported in X-ray structure of malectin complexed with the pentasaccharide, the pocket has the precise volume to accommodate the two glucose rings of Glc2Man3 reducing terminal.

Chapter 5 – Integrative conclusions and future perspectives

Malectin is a membrane-anchored ER protein, very conserved in animals. The role of this lectin has not yet been assigned however, several studies indicate that it is involved in the early steps of N-glycosylation interacting with Glc2Man9GlcNac2.

The reported structures of malectin highlighted residues in the binding pocket important for its specificity. In this project, using glycan microarray analysis, we analysed the interactions of malectin with glucosylated glycans, but in specific with the pentasaccharide Glc2Man3. To characterize the binding pocket interactions, we designed mutations to the protein sequence with the intuit of understanding the importance of the hydrogen bonds and of the water-network. Also, considering the information obtained from the malectin-bound X-ray structures, two mutation were designed 1) to search for an extended binding-site that interacts with other arm of the N-glycan and 2) to design a new protein with a new binding specificity (for O-glycan terminus from proteins containing EGF-repeats). In this way, we mutated E102A, D201A, Y104F, Y131F, Y199AY200A and S80Q. We were able to achieve the six mutations through site-directed mutagenesis process with specific primers designed for this purpose. Following protein production and purification, the structural stability studies performed using Differential Scanning Calorimetry (DSC) indicated that there was no loss of stability in the new mutants when compared to the native protein. However, there was a substantial increase in the melting temperature (Tm) for the E102A, Y131F and Y199AY200A mutated proteins (of 6.22 ºC, 5.29 ºC, 6.79 ºC, respectively). These increases in Tm suggest that some major structural changes may have occurred. The literature has already mentioned the high mobility of the four loops in the free malectin binding pocket. Therefore, to further comprehension we could perform DSC analysis on the malectin mutants in the ligand-bound form and more structural stability studies.

The specificity of the WT malectin and the effect of the mutations in the interaction with glucosylated ligands were determined by a first screening in a polysaccharide microarrays Set (Fungal, bacterial & plant PS Set 2). The malectin protein which interacted with dextran and the mutations had effect on the interaction with this polysaccharide. Afterwards, the proteins were analysed in an NGL-based oligosaccharide microarray to gain further insight into the interactions that give malectin specificity. The highly specificity of malectin was confirmed to α1,3 di-glucosylated N-glycans. The hydrogen bonds performed by S80, E102 and D201 emerged as crucial for the interactions of malectin with the Glc2Man3 pentasaccharide. It was also notable to observe the importance of the water-mediated hydrogen bonds and π/C-H staking, established by Y104 and Y131, to the interaction. Unfortunately, the role of the two tyrosines 199 and 200 could not been assign due to the lack of the di-glucosyl high-mannose N-glycan in the array. Also, contrary to what has been reported, no signs of interaction between malectin WT and the Xyl2Glc trisaccharide was found. Thereby, the attribution of a new specificity could not be explored

during this work. However, through this mutation we could see the consequence of decreasing the volume of the pocket, which led to the abrogation of the interaction of the human malectin with the pentasaccharide. This indicates that the malectin binding-pocket has the right size to accommodate the two glucose rings of the di-glucosylated *N*-glycan.

To fulfil the other two objectives proposed, future work will involve re-analysing human malectin WT and the two mutants in the glycan microarray with a novel protein batch or optimized assay conditions. In addition, consider mutating the S80 residue to a shorter residue and to determine the dissociation constants of the successful protein mutant with the Xyl2Glc trisaccharide. Nevertheless, the results of this project allowed us to confirm the important role of these residues in the binding event of human malectin. Thus, they constitute another step towards determining the function of this protein so important in the control of the maturation process of nascent proteins. In this work, the advantage of microarrays in allowing the analysis of several proteins in parallel was imperious. The evolution of this high-throughput technology will promote not only in the detection of carbohydrate-binding proteins and their ligands but also in the discoveries of clinically relevant biomarkers.

References

Acar, M., Jafar-Nejad, H., Takeuchi, H., Rajan, A., Ibrani, D., Rana, N. A., Pan, H., Haltiwanger, R. S., & Bellen, H. J. (2008). Rumi Is a CAP10 Domain Glycosyltransferase that Modifies Notch and Is Required for Notch Signaling. *Cell*. https://doi.org/10.1016/j.cell.2007.12.016

Aebi, M., Bernasconi, R., Clerc, S., & Molinari, M. (2010). N-glycan structures: recognition and processing in the ER. In *Trends in Biochemical Sciences*. https://doi.org/10.1016/j.tibs.2009.10.001

Apweiler, R., Hermjakob, H., & Sharon, N. (1999). On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophysica Acta - General Subjects*. https://doi.org/10.1016/S0304-4165(99)00165-8

BRADA, D., & DUBACH, U. C. (1984). Isolation of a homogeneous glucosidase II from pig kidney microsomes. *European Journal of Biochemistry*, *141*(1), 149–156. https://doi.org/10.1111/j.1432-1033.1984.tb08169.x

Bray, S. J. (2016). Notch signalling in context. In *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm.2016.94

Breitling, J., & Aebi, M. (2013). N-linked protein glycosylation in the endoplasmic reticulum. *Cold Spring Harbor Perspectives in Biology*. https://doi.org/10.1101/cshperspect.a013359

Brewer, C. F., & Dam, T. K. (2000). Essentials of Glycobiology, Edited by A. Varki, R. Cummings, J. Esko, H. Freeze, G. Hart, and J. Marth; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1999, 653 pp.; ISBN 0-87969-559-5 (cloth, $175), ISBN 0-87969-560-9 (printed hard co. *Carbohydrate Research*. https://doi.org/10.1016/s0008-6215(99)00334-1

Brockhausen, I., & Stanley, P. (2015). O-GalNAc Glycans. In *Essentials of Glycobiology*.

Chen, Y., Hu, D., Yabe, R., Tateno, H., Qin, S. Y., Matsumoto, N., Hirabayashi, J., & Yamamoto, K. (2011). Role of malectin in Glc 2Man 9GlcNAc 2-dependent quality control of α1-antitrypsin. *Molecular Biology of the Cell*, *22*(19), 3559–3570. https://doi.org/10.1091/mbc.E11-03-0201

Clayton, K., & Branscomb, W. (1979). *Error Induction Polymerases by Mutant and Wild Type*. *254*(6), 1902–1912.

Cline, J., Braman, J. C., & Hogrefe, H. H. (1996). PCR fidelity of Pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Research*. https://doi.org/10.1093/nar/24.18.3546

Cox, M. M., & Nelson, D. L. (2004). Chapter 7. Carboydrates and glycobiology. *Lehninger Principles of Biochemistry*, 238–272.

Cummings, R. D., Darvill, A. G., Etzler, M. E., & Hahn, M. G. (2015). Glycan-Recognizing Probes as Tools. *Essentials of Glycobiology*.

Cummings, R. D., & Pierce, J. M. (2014). The challenge and promise of glycomics. *Chemistry and Biology*, *21*(1), 1–15. https://doi.org/10.1016/j.chembiol.2013.12.010

Deprez, P., Gautschi, M., & Helenius, A. (2005). More than one glycan is needed for ER glucosidase II to allow entry of glycoproteins into the calnexin/calreticulin cycle. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2005.05.029

Feizi, T. (2019). Nanolithography of biointerfaces. *Faraday Discussions*, *219*, 262–275. https://doi.org/10.1039/c9fd00082h

Feizi, T., & Chai, W. (2004). *Oligosaccharide microarrays to decipher the glyco code*. *5*(July), 582–588.

Feizi, T., Fazio, F., Chai, W., Wong, C. H., Mamidi, A. S., & Surolia, A. (2003). Carbohydrate microarrays - A new set of technologies at the frontiers of glycomics. *Current Opinion in Structural Biology*, *33*(5), 1363–1384. https://doi.org/10.1080/07391102.2014.948070

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*. https://doi.org/10.1242/dev.060020

Finn, R. D., Mistry, J., Schuster-Böckler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S. R., Sonnhammer, E. L. L., & Bateman, A. (2006). Pfam: clans, web tools and services. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkj149

Fukui, S., Feizi, T., Galustian, C., Lawson, A. M., & Chai, W. (2002). *Oligosaccharide microarrays for high-throughput detection and specificity assignments of carbohydrate-protein interactions*. *20*(October), 1011–1017. https://doi.org/10.1038/nbt735

Galli, C., Bernasconi, R., Soldà, T., Calanca, V., & Molinari, M. (2011). Malectin participates in a backup glycoprotein quality control pathway in the mammalian ER. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0016304

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*. https://doi.org/10.1385/1-59259-890-0:571

Green, I. (2012). General Recommendations for DNA Electrophoresis. *Thermo Scientific*, 1–2. papers2://publication/uuid/AB6F6299-D314-4424-B164-14DCFDBBF243

Haltiwanger, R. S. (2007). Introduction to Glycobiology. Second Edition. By Maureen E Taylor and , Kurt Drickamer. Oxford and New York: Oxford University Press. $44.50 (paper). xix + 255 p; ill.; index. ISBN: 0-19-928278-1. 2006. . *The Quarterly Review of Biology*. https://doi.org/10.1086/523133

Haltiwanger, R. S., Wells, L., Freeze, H. H., & Stanley, P. (2017). Other Classes of Eukaryotic Glycans. *Essentials of Glycobiology, 3rd Edition*.

Haltom, A. R., & Jafar-Nejad, H. (2015). The multiple roles of epidermal growth factor repeat O-glycans in animal development. In *Glycobiology*. https://doi.org/10.1093/glycob/cwv052

Hase, S., Kawabata, S. ichiro, Nishimura, H., Takeya, H., Sueyoshi, T., Miyata, T., Iwanaga, S., Takao, T., Shimonishi, Y., & Ikenaka, T. (1988). A new trisaccharide sugar chain linked to a serine residue in bovine blood coagulation factors VII and IX. *Journal of Biochemistry*. https://doi.org/10.1093/oxfordjournals.jbchem.a122571

Helenius, A., & Aebi, M. (2004). Roles of N-linked glycans in the endoplasmic reticulum. In *Annual Review of Biochemistry*. https://doi.org/10.1146/annurev.biochem.73.011303.073752

Helenius, J., Ng, D. T. W., Marolda, C. L., Walter, P., Valvano, M. A., & Aebi, M. (2002). Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. *Nature*. https://doi.org/10.1038/415447a

Hirabayashi, J. (2003). Oligosaccharide microarrays for glycomics. *Trends in Biotechnology*, *21*(4), 141–143. https://doi.org/10.1016/S0167-7799(03)00002-7

Hirschberg, C. B., & Snider, M. D. (1987). Topography of Glycosylation in the Routh Endoplasmic Reticulum and Golgi Apparatus. *Ann. Rev. Biochem.*, *56*, 63–87.

Hubbard, S. C., & Ivatt, R. J. (1981). Synthesis and Processing of Asparagine-linked oligosaccharides. *Ann. Rev. Biochem.*, *50*, 555–583.

Hudson, K. L., Bartlett, G. J., Diehl, R. C., Agirre, J., Gallagher, T., Kiessling, L. L., & Woolfson, D. N. (2015). Carbohydrate-Aromatic Interactions in Proteins. *Journal of the American Chemical Society*, *137*(48), 15152–15160. https://doi.org/10.1021/jacs.5b08424

Ibarra-Molero, B., Naganathan, A. N., Sanchez-Ruiz, J. M., & Muñoz, V. (2016). Modern Analysis of Protein Folding by Differential Scanning Calorimetry. In *Methods in Enzymology*. https://doi.org/10.1016/bs.mie.2015.08.027

J., B. Arey (2012). The Role of Glycosylation in Receptor Signaling. In *Glycosylation*. https://doi.org/10.5772/50262

Johnson, C. M. (2013). Differential scanning calorimetry as a tool for protein folding and stability. In *Archives of Biochemistry and Biophysics*. https://doi.org/10.1016/j.abb.2012.09.008

Kiessling, L. L., & Cairo, C. W. (2002). *Hitting the sweet spot A method for producing large carbohydrate microarrays allows high-throughput , sensitive detection of carbohydrate – antibody T cells have been recruited to both produce and deliver vectors to*. *20*(March).

Kornfeld, R., & Kornfeld, S. (1985). Rosalind Kornfeld and Stuart Kornfeld. *Annual Review of Biochemistry*, *54*(1), 631–664.

Kowarik, M., Küng, S., Martoglio, B., & Helenius, A. (2002). Protein folding during cotranslational translocation in the endoplasmic reticulum. *Molecular Cell*. https://doi.org/10.1016/S1097-2765(02)00685-8

Krogh, T. N., Bachmann, E., Teisner, B., Skjødt, K., & Højrup, P. (1997). Glycosylation analysis and protein structure determination of murine fetal antigen 1 (mFA1). *European Journal of Biochemistry*, *3244*, 334–342.

Lederkremer, G. Z. (2009). Glycoprotein folding, quality control and ER-associated degradation. In *Current Opinion in Structural Biology*. https://doi.org/10.1016/j.sbi.2009.06.004

Lee, T. V., Sethi, M. K., Leonardi, J., Rana, N. A., Buettner, F. F. R., Haltiwanger, R. S., Bakker, H., & Jafar-Nejad, H. (2013). Negative Regulation of Notch Signaling by Xylose. *PLoS Genetics*, *9*(6). https://doi.org/10.1371/journal.pgen.1003547

Leonardi, J., Fernandez-Valdivia, R., Li, Y. D., Simcox, A. A., & Jafar-Nejad, H. (2011). Multiple

O-glucosylation sites on Notch function as a buffer against temperature-dependent loss of signaling. *Development*. https://doi.org/10.1242/dev.068361

Letunic, I., Copley, R. R., Pils, B., Pinkert, S., Schultz, J., & Bork, P. (2006). SMART 5: domains in the context of genomes and networks. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkj079

Li, Z., Fischer, M., Satkunarajah, M., Zhou, D., Withers, S. G., & Rini, J. M. (2017). Structural basis of Notch O-glucosylation and O-xylosylation by mammalian protein-O-glucosyltransferase 1 (POGLUT1). *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-00255-7

Liu, Y., Feizi, T., Campanero-Rhodes, M. A., Childs, R. A., Zhang, Y., Mulloy, B., Evans, P. G., Osborn, H. M. I., Otto, D., Crocker, P. R., & Chai, W. (2007). Neoglycolipid Probes Prepared via Oxime Ligation for Microarray Analysis of Oligosaccharide-Protein Interactions. *Chemistry and Biology*, *14*(7), 847–859. https://doi.org/10.1016/j.chembiol.2007.06.009

Liu, Y., Palma, A. S., & Feizi, T. (2009). *Carbohydrate microarrays : Key developments in glycobiology Carbohydrate microarrays : key developments in glycobiology*. *June*. https://doi.org/10.1515/BC.2009.071

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., & Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkz268

Mamidi, A. S., & Surolia, A. (2015). Hierarchical sampling for metastable conformers determines biomolecular recognition: The case of malectin and diglucosylated N-glycan interactions. *Journal of Biomolecular Structure and Dynamics*, *33*(6), 1363–1384. https://doi.org/10.1080/07391102.2014.948070

McPherson, M. J., & Moller, S. G. (2006). PCR: The basics. In *Vasa*.

Moloney, D. J., Panin, V. M., Johnston, S. H., Chen, J., Shao, L., Wilson, R., Wang, Y., Stanley, P., Irvine, K. D., Haltiwanger, R. S., & Vogt, T. F. (2000). Fringe is a glycosyltransferase that modifies Notch. *Nature*. https://doi.org/10.1038/35019000

Moloney, D. J., Shair, L. H., Lu, F. M., Xia, J., Locke, R., Matta, K. L., & Haltiwanger, R. S. (2000). Mammalian Notch1 is modified with two unusual forms of O-linked glycosylation

found on epidermal growth factor-like modules. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.275.13.9604

Moremen, K. W., Tiemeyer, M., & Nairn, A. V. (2012). Vertebrate protein glycosylation: Diversity, synthesis and function. In *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm3383

Nishimura, H., Kawabata, S., Kisiel, W., Hase, S., Ikenaka, T., Takao, T., Shimonishi, Y., & Iwanaga, S. (1989). Identification of a disaccharide (Xyl-Glc) and a trisaccharide (Xyl2-Glc) O-glycosidically linked to a serine residue in the first epidermal growth factor-like domain of human factors VII and IX and protein Z and bovine protein Z. *Journal of Biological Chemistry*. https://doi.org/10.1016/S0021-9258(19)47065-8

Nishimura, Hitoshi, Yamashita, S., Zeng, Z., Walz, D. A., & Iwanaga, S. (1992). Evidence for the existence of O-linked sugar chains consisting of glucose and xylose in bovine thrombospondin. *Journal of Biochemistry*. https://doi.org/10.1093/oxfordjournals.jbchem.a123780

Palma, A. S., Feizi, T., Childs, R. A., Chai, W., & Liu, Y. (2014). The neoglycolipid (NGL)-based oligosaccharide microarray system poised to decipher the meta-glycome. In *Current Opinion in Chemical Biology* (Vol. 18, Issue 1). https://doi.org/10.1016/j.cbpa.2014.01.007

Palma, A. S., Liu, Y., Muhle-Goll, C., Butters, T. D., Zhang, Y., Childs, R., Chai, W., & Feizi, T. (2010). Multifaceted approaches including neoglycolipid oligosaccharide microarrays to ligand discovery for malectin. *Methods in Enzymology*, *478*(C), 265–286. https://doi.org/10.1016/S0076-6879(10)78013-7

Paz, J. L. De, & Seeberger, P. H. (2012). Chapter 1 Recent Advances and Future Challenges. *Carbohydrate Microarrays: Methods and Protocols, Methods in Molecular Biology, Vol. 808*, *808*. https://doi.org/10.1007/978-1-61779-373-8

Petrescu, A. J., Butters, T. D., Reinkensmeier, G., Petrescu, S., Platt, F. M., Dwek, R. A., & Wormald, M. R. (1997). The solution NMR structure of glucosylated N-glycans involved in the early stages of glycoprotein biosynthesis and folding. *EMBO Journal*. https://doi.org/10.1093/emboj/16.14.4302

Pinheiro, A. Benedita, Yan Liu, Yoichi Takeda, Wengang Chai, Yukishige Ito, Maria João Romão, Ten Feizei, Ana Luísa Carvalho, Angelina Palma, (in preparation). Strutural Basis for Highly Selective Recognition of Di-Glucosylated N-glycan by Human Malectin in the

Endoplasmic Reticulum

Plante, O. J., Palmacci, E. R., & Seeberger, P. H. (2001). Automated solid-phase synthesis of oligosaccharides. *Science*, *291*(5508), 1523–1527. https://doi.org/10.1126/science.1057324

Qin, S. Y., Hu, D., Matsumoto, K., Takeda, K., Matsumoto, N., Yamaguchi, Y., & Yamamoto, K. (2012). Malectin forms a complex with ribophorin i for enhanced association with misfolded glycoproteins. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M112.394288

Ramírez, A. S., Kowal, J., & Locher, K. P. (2019). Cryo-electron microscopy structures of human oligosaccharyltransferase complexes OST-A and OST-B. *Science*, *366*(6471), 1372–1375. https://doi.org/10.1126/science.aaz3505

Rana, N. A., Nita-Lazar, A., Takeuchi, H., Kakuda, S., Luther, K. B., & Haltiwanger, R. S. (2011). O-glucose trisaccharide is present at high but variable stoichiometry at multiple sites on mouse Notch1. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M111.268243

Reily, C., Stewart, T. J., Renfrow, M. B., & Novak, J. (2019). Glycosylation in health and disease. In *Nature Reviews Nephrology*. https://doi.org/10.1038/s41581-019-0129-4

Rillahan, C. D., & Paulson, J. C. (2011). *Glycan Microarrays for Decoding the Glycome*. https://doi.org/10.1146/annurev-biochem-061809-152236

Schallus, T., Fehér, K., Sternberg, U., Rybin, V., & Muhle-Goll, C. (2010). Analysis of the specific interactions between the lectin domain of malectin and diglucosides. *Glycobiology*, *20*(8). https://doi.org/10.1093/glycob/cwq059

Schallus, T., Jaeckh, C., Fehér, K., Palma, A. S., Liu, Y., Simpson, J. C., Mackeen, M., Stier, G., Gibson, T. J., Feizi, T., Pieler, T., & Muhle-Goll, C. (2008b). Malectin: A novel carbohydrate-binding protein of the endoplasmic reticulum and a candidate player in the early steps of protein N-glycosylation. *Molecular Biology of the Cell*, *19*(8). https://doi.org/10.1091/mbc.E08-04-0354

Sethi, M. K., Buettner, F. F. R., Ashikov, A., Krylov, V. B., Takeuchi, H., Nifantiev, N. E., Haltiwanger, R. S., Gerardy-Schahn, R., & Bakker, H. (2012). Molecular cloning of a xylosyltransferase that transfers the second xylose to O-glucosylated epidermal growth factor repeats of notch. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M111.302406

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-glucosylated Notch epidermal growth factor repeats. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.C109.065409

Stanley P, Taniguchi N, A. M. (2017). N-glycans. In *Essentials of Glycobiology*. https://doi.org/10.1101/glycobiology.3e.009

Stanley, P., & Okajima, T. (2010). Roles of glycosylation in notch signaling. In *Current Topics in Developmental Biology*. https://doi.org/10.1016/S0070-2153(10)92004-8

Stoll, M., & Feizi, T. (2009). Software Tools for Storing, Processing and Displaying Carbohydrate Microarray Data. *Glyco-Bioinformatics – Bits 'n' Bytes of Sugars*, *4-8 October*, 123–140.

Takeuchi, H., Fernańdez-Valdivia, R. C., Caswell, D. S., Nita-Lazar, A., Rana, N. A., Garner, T. P., Weldeghiorghis, T. K., Macnaughtan, M. A., Jafar-Nejad, H., & Haltiwanger, R. S. (2011). Rumi functions as both a protein O-glucosyltransferase and a protein O-xylosyltransferase. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1109696108

Takeuchi, H., & Haltiwanger, R. S. (2014). Significance of glycosylation in Notch signaling. In *Biochemical and Biophysical Research Communications*. https://doi.org/10.1016/j.bbrc.2014.05.115

Takeuchi, H., Kantharia, J., Sethi, M. K., Bakker, H., & Haltiwanger, R. S. (2012). Site-specific O-glucosylation of the epidermal growth factor-like (EGF) repeats of notch: Efficiency of glycosylation is affected by proper folding and amino acid sequence of individual EGF repeats. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M112.401315

Tang, P. W., Gool, H. C., Hardy, M., Lee, Y. C., & Felzi, T. (1985). Novel approach to the study of the antigenicities and receptor functions of carbohydrate chains of glycoproteins. *Biochemical and Biophysical Research Communications*. https://doi.org/10.1016/0006-291X(85)91158-1

Trombetta, E. S. (2003). The contribution of N-glycans and their processing in the endoplasmic reticulum to glycoprotein biosynthesis. In *Glycobiology*. https://doi.org/10.1093/glycob/cwg075

Varki, A. (2017). Biological roles of glycans. *Glycobiology*, *27*(1), 3–49.

https://doi.org/10.1093/glycob/cww086

Varki, A., Freeze, H. H., & Gagneux, P. (2009). Evolution of Glycan Diversity. In *Essentials of Glycobiology*.

Wang, D., Liu, S., Trummer, B. J., Deng, C., & Wang, A. (2002). Carbohydrate microarrays for the recognition of cross-reactive molecular markers of microbes and host cells. *Nature Biotechnology*. https://doi.org/10.1038/nbt0302-275

Wilson, C. M., Kraft, C., Duggan, C., Ismail, N., Crawshaw, S. G., & High, S. (2005). Ribophorin I associates with a subset of membrane proteins after their integration at the Sec61 translocon. *Journal of Biological Chemistry*, *280*(6), 4195–4206. https://doi.org/10.1074/jbc.M410329200

Woods, R. J., Pathiaseril, A., Wormald, M. R., Edge, C. J., & Dwek, R. A. (1998). The high degree of internal flexibility observed for an oligomannose oligosaccharide does not alter the overall topology of the molecule. *European Journal of Biochemistry*. https://doi.org/10.1046/j.1432-1327.1998.2580372.x

Zhang, Z., Ollmann, I. R., Ye, X., Wischnat, R., Baasov, T., Wong, C., Scripps, T., Torrey, N., Road, P., & Jolla, L. (1999). *Programmable One-Pot Oligosaccharide Synthesis*. *6*, 734–753.

Supplementary material

Supplementary table 1 – Characteristics of hMalectin used in the previous investigation (Pinheiro et al., unpublished). The residues that were under study are highlighted in red and the molecular weight, the extinction coefficient and the concentration of the purified protein are depicted. Protein concentration was measured by UV measurements at 280 nm.

Recombinant Protein Sequence	Extinction Coefficient ($M^{-1}cm^{-1}$)	Molecular Weight	Concentration
MKHHHHHHPMAGLPESVIWAVNAGGEAHVDVHGIHFRKDPLEGRVGRASDYGMKLPILR SNPEDQILYQTERYNEETFGYEVPIKEEGDYVLVLKFAEVYFAQSQQKVFDVRLNGHVVVKDL DIFDRVGHSTAHDEIIPMSIRKGKLSVQGEVSTFTGKLYIEFVKGYYDNPKVCALYIMAGTVDD VPKLQPHPGLE	20400	22334.4	54 µM (1,2g/L)

Supplementary table 2 – DNA sequence for each protein provided by STABVIDA services.

	DNA Sequence
hMalectin	ctagcttctttcgggctttgttagcagccggatctcagtggtggtggtggtggtgctcgagtgcggccgcactagtggtacctactccaatcccggatgaggctgaagctttggtacatcatccactgtcccagcc atgatgtagagtgcacagaccttgggattgtcatagtacccccttgacaaactcaatgtagagtttccctgtgaaggtggacacctccccctggacactcagcttcccctttctgatgctcataggtataatttcat cgtgagctgtgctatgcccaacacgatcaaagatatccaagtccttcaccacgacgtggccattcaatcgtacatcaaatacctttttgctgggactgtgcaaagtagacctctgcaaatttcaagaccagcacg tagtccccctcctctttgatgggcacttcgtagccaaaggtctcctcattgtaccgctcagtttgatacaggatctggtcctcagggttggaacgcaggattggcagtttcatgccatagtctgaggctcggccca cccggccttccaaagggtccttgcggaagtggatcccgtgcacgtccacatgcgcctctccacccgcgttgaccgcccaaatgacgctctcgggcagcccggccatggggtgatggtgatggtgatgtttcatg gtatatctccttcttaaagttaaacaaaattatttctagaggggaattgttatccgctcacaattcccctatagtgagtcgtattaatttcgcgggatcgagatctcgatcctctacgccggacgcatcgtggccg gcatcaccggcgccacaggtgcggttgctggcgcctatatcgccgacatcaccgatggggaagatcgggctcgccacttcgggctcatgagcgcttgtttcggcgtgggtatggtggcaggccccgtggccgg gggactgttgggcgccatctccttgcatgcaccattccttgcggcggcggtgctcaacggcctcaacctactactgggctgcttcctaatgcaggagtcgcataagggagagcgtcgagatcccggacaccatc gaatggcgcaaaaccttcgcggtatggcatgatagcgcccggaagagaatcaattcagggtggtgaatgggaaaccagtaacgttatacgatgtcgcagagtatgccggtgtctcttatcaaaccgtttccc ccggggtgaaccaggccagccacgtttctgcgaaaacc

S80Q	ccctagcttctttcgggctttgttagcagccggatctcagtggtggtggtggtggtgctcgagtgcggccgcactagtggtacctactccaatcccggatgaggctgaagctttggtacatcatccactgtcccag ccatgatgtagagtgcacagaccttgggattgtcatagtacccccttgacaaactcaatgtagagtttccctgtgaaggtggacacctccccctggacactcagcttcccctttctgatgctcataggtataatttc atcgtgagctgtgctatgcccaacacgatcaaagatatccaagtccttcaccacgacgtggccattcaatcgtacatcaaataccttttgctgggactgtgcaaagtagacctctgcaaatttcaagaccagca cgtagtccccctcctctttgatgggcacttcgtagccaaaggtctcctcattgtaccgctcagtttgatacaggatctggtcctcagggttggaacgcaggattggcagtttcatgccatagtcttgggctcggcc cacccggccttccaaagggtccttgcggaagtggatcccgtgcacgtccacatgcgcctctccacccgcgttgaccgcccaaatgacgctctcgggcagcccggccatggggtgatggtgatggtgatgtttca tggtatatctccttcttaaagttaaacaaaattatttctagaggggaattgttatccgctcacaattcccctatagtgagtcgtattaatttcgcgggatcgagatctcgatcctctacgccggacgcatcgtggcc ggcatcaccggcgccacaggtgcggttgctggcgcctatatcgccgacatcaccgatggggaagatcgggctcgccacttcgggctcatgagcgcttgtttcggcgtgggtatggtggcaggccccgtggccg ggggactgttgggcgccatctccttgcatgcaccattccttgcggcggcggtgctcaacggcctcaacctactactgggctgcttcctaatgcaggagtcgcataagggagagcgtcgagatcccggacacca tcgaatggcgcaaaaccttcccggtatggcatgatagcgcccggaagagagtcattcagggtggtgaatgggaaaccagtaacgttatacgatgtcgcaaaatatgccg
E102A	ctcgctctttcgggctttgttagcagccggatctcagtggtggtggtggtggtgctcgagtgcggccgcactagtggtacctactccaatcccggatgaggctgaagctttggtacatcatccactgtcccagcca tgatgtagagtgcacagaccttgggattgtcatagtacccccttgacaaactcaatgtagagtttccctgtgaaggtggacacctccccctggacactcagcttcccctttctgatgctcataggtataatttcatc gtgagctgtgctatgcccaacacgatcaaagatatccaagtccttcaccacgacgtggccattcaatcgtacatcaaataccttttgctgggactgtgcaaagtagacctctgcaaatttcaagaccagcacgt agtccccctcctctttgatgggcacttcgtagccaaaggtctcctcattgtaccgcgcagtttgatacaggatctggtcctcagggttggaacgcaggattggcagtttcatgccatagtctgaggctcggcccac ccggccttccaaagggtccttgcggaagtggatcccgtgcacgtccacatgcgcctctccacccgcgttgaccgcccaaatgacgctctcgggcagcccggccatggggtgatggtgatggtgatgtttcatgg tatatctccttcttaaagttaaacaaaattatttctagaggggaattgttatccgctcacaattcccctatagtgagtcgtattaatttcgcgggatcgagatctcgatcctctacgccggacgcatcgtggccggc atcaccggcgccacaggtgcggttgctggcgcctatatcgccgacatcaccgatggggaagatcgggctcgccacttcgggctcatgagcgcttgtttcggcgtgggtatggtggcaggccccgtggccgggg gactgttgggcgccatctccttgcatgcaccattccttgcggcggcggtgctcaacggcctcaacctactactgggctgcttcctaatgcaggagtcgcataagggagagcgtccagatcccggacaccatcga atggggcaaaaccttcgcggtatggcatgaaaccccccggaaaaaatcaattcaggca
Y104F	tcgccttctttcgggctttgttagcagccggatctcagtggtggtggtggtggtgctcgagtgcggccgcactagtggtacctactccaatcccggatgaggctgaagctttggtacatcatccactgtcccagcc atgatgtagagtgcacagaccttgggattgtcatagtacccccttgacaaactcaatgtagagtttccctgtgaaggtggacacctccccctggacactcagcttcccctttctgatgctcataggtataatttcat cgtgagctgtgctatgcccaacacgatcaaagatatccaagtccttcaccacgacgtggccattcaatcgtacatcaaataccttttgctgggactgtgcaaagtagacctctgcaaatttcaagaccagcacg tagtccccctcctctttgatgggcacttcgtagccaaaggtctcctcattgaaccgctcagtttgatacaggatctggtcctcagggttggaacgcaggattggcagtttcatgccatagtctgaggctcggccca cccggccttccaaagggtccttgcggaagtggatcccgtgcacgtccacatgcgcctctccacccgcgttgaccgcccaaatgacgctctcgggcagcccggccatggggtgatggtgatggtgatgtttcatg gtatatctccttcttaaagttaaacaaaattatttctagaggggaattgttatccgctcacaattcccctatagtgagtcgtattaatttcgcgggatcgagatctcgatcctctacgccggacgcatcgtggccg gcatcaccggcgccacaggtgcggttgctggcgcctatatcgccgacatcaccgatggggaagatcgggctcgccacttcgggctcatgagcgcttgtttcggcgtgggtatggtggcaggccccgtggccgg gggactgttgggcgccatctccttgcatgcaccattccttgcggcggcggtgctcaacggcctcaacctactactgggctgcttcctaatgcaggagtcgcataagggaaagcgtcgagatcccggaccccatc gaaggga
Y131F	Cctcgcttctttcgggctttgttagcagccggatctcagtggtggtggtggtggtgctcgagtgcggccgcactagtggtacctactccaatcccggatgaggctgaagctttggtacatcatccactgtcccagc catgatgtagagtgcacagaccttgggattgtcatagtacccccttgacaaactcaatgtagagtttccctgtgaaggtggacacctccccctggacactcagcttcccctttctgatgctcataggtataatttca tcgtgagctgtgctatgcccaacacgatcaaagatatccaagtccttcaccacgacgtggccattcaatcgtacatcaaataccttttgctgggactgtgcaaagaagacctctgcaaatttcaagaccagcac gtagtccccctcctctttgatgggcacttcgtagccaaaggtctcctcattgtaccgctcagtttgatacaggatctggtcctcagggttggaacgcaggattggcagtttcatgccatagtctgaggctcggccc acccggccttccaaagggtccttgcggaagtggatcccgtgcacgtccacatgcgcctctccacccgcgttgaccgcccaaatgacgctctcgggcagcccggccatggggtgatggtgatggtgatgtttcat

	ggtatatctccttcttaaagttaaacaaaattatttctagaggggaattgttatccgctcacaattcccctatagtgagtcgtattaatttcgcgggatcgagatctcgatcctctacgccggacgcatcgtggcc ggcatcaccggcgccacaggtgcggttgctggcgcctatatcgccgacatcaccgatggggaagatcgggctcgccacttcgggctcatgagcgcttgtttcggcgtgggtatggtggcaggccccgtggccg ggggactgttgggcgccatctccttgcatgcaccattccttgcggcggcggtgctcaacggcctcaacctactactgggctgcttcctaatgcaggagtcgcataagggagagcgtcgagatcccggacacca tcgaatgggcgcaaaaccttccggtatggcatgatagcgcccggaagaga
Y199AY200A	Ctcagcttctttcgggctttgttagcagccggatctcagtggtggtggtggtggtgctcgagtgcggccgcactagtggtacctactccaatcccggatgaggctgaagctttggtacatcatccactgtcccag ccatgatgtagagtgcacagaccttgggattgtcagcggccccttgacaaactcaatgtagagtttccctgtgaaggtggacacctccccctggacactcagcttcccctttctgatgctcataggtataatttc atcgtgagctgtgctatgcccaacacgatcaaagatatccaagtccttcaccacgacgtggccattcaatcgtacatcaaataccttttgctgggactgtgcaaagtagacctctgcaaatttcaagaccagca cgtagtccccctcctctttgatgggcacttcgtagccaaaggtctcctcattgtaccgctcagtttgatacaggatctggtcctcagggttggaacgcaggattggcagtttcatgccatagtctgaggctcggcc cacccggccttccaaagggtccttgcggaagtggatcccgtgcacgtccacatgcgcctctccacccgcgttgaccgcccaaatgacgctctcgggcagcccggccatggggtgatggtgatggtgatgtttca tggtatatctccttcttaaagttaaacaaaattatttctagaggggaattgttatccgctcacaattcccctatagtgagtcgtattaatttcgcgggatcgagatctcgatcctctacgccggacgcatcgtggcc ggcatcaccggcgccacaggtgcggttgctggcgcctatatcgccgacatcaccgatggggaagatcgggctcgccacttcgggctcatgagcgcttgtttcggcgtgggtatggtggcaggccccgtggccg ggggactgttgggcgccatctccttgcatgcaccatttcttgcggcggcggtgctcaacggcctcaacctactactgggctgcttcctaatgcaggagtcgcataagggagagcgtcgagatcccggacaccat cgaatggggcaaaaccttccggtatggcatgatagcgcccggaagaaagtcaattcagggggggggaatgtgaaaccagtaag
D201A	Caggctttctttcgggctttgttagcagccggatctcagtggtggtggtggtggtgctcgagtgcggccgcactagtggtacctactccaatcccggatgaggctgaagctttggtacatcatccactgtcccag ccatgatgtagagtgcacagaccttgggattggcatagtacccctgacaaactcaatgtagagtttccctgtgaaggtggacacctccccctggacactcagcttcccctttctgatgctcataggtataatttc atcgtgagctgtgctatgcccaacacgatcaaagatatccaagtccttcaccacgacgtggccattcaatcgtacatcaaataccttttgctgggactgtgcaaagtagacctctgcaaatttcaagaccagca cgtagtccccctcctctttgatgggcacttcgtagccaaaggtctcctcattgtaccgctcagtttgatacaggatctggtcctcagggttggaacgcaggattggcagtttcatgccatagtctgaggctcggcc cacccggccttccaaagggtccttgcggaagtggatcccgtgcacgtccacatgcgcctctccacccgcgttgaccgcccaaatgacgctctcgggcagcccggccatggggtgatggtgatggtgatgtttca tggtatatctccttcttaaagttaaacaaaattatttctagaggggaattgttatccgctcacaattcccctatagtgagtcgtattaatttcgcgggatcgagatctcgatcctctacgccggacgcatcgtggcc ggcatcaccggcgccacaggtgcggttgctggcgcctatatcgccgacatcaccgatggggaagatcgggctcgccacttcgggctcatgagcgcttgtttcggcgtgggtatggtggcaggccccgtggccg ggggactgttgggcgccatctccttgcatgcaccattccttgcggcggcggtgctcaacggcctcaacctactactgggctgcttcctaatgcaggagtcgcataagggagagcgtctagatcccggacaccat cgaatggggcaaaaccttcccggaatggcatgaaagcccccggaaaaaatccatttat

Supplementary table 3 – Protein quantification and preparation for NGL-based microarray experiments. For the carbohydrate microarray assays the protein-antibody complex was used in a 1:3:3 ratio.

		S80Q	E102A	Y104F	Y131F	Y199AY200A	D201A	WT
Protein Quantification	Ɛ ($M^{-1}cm^{-1)}$	20400	20400	18910	18910	17420	20400	20400
	MW	22375,45	22276,36	22318,4	25462	22150,2	22290,39	22334,4
	Abs	0,292	0,369	0,283	0,209	0,233	0,464	0,211
	l	0,05	0,05	0,05	0,05	0,05	0,05	0,05
	c (M)	0,00028627	0,000362	0,000299	0,000221	0,000267509	0,000455	0,000207
	c (µM)	286,3	361,8	299,3	221	267,5	455	207
	c*MW(µg/µl)	6,406	8,059	6,680	5,628	5,925	10,140	4,620
Volume (µl) for 10 µg/mL		0,6	0,5	0,6	0,7	0,7	0,4	0,9
Antibodies	Anti-Mouse Biotin (µL)	8	8	8	8	8	8	8
	Anti-His (µL)	2,4	2,4	2,4	2,4	2,4	2,4	2,4
Buffer	BSA in HBS (µL)	69,0	69,1	69,0	68,9	68,9	69,2	68,7

Supplementary table 4 - Oligonucleotide data sheet provided by STABVIDA.

Malectin_S80Q_F			
5’- ttc atg cca tag tct tgg gct cgg ccc acc cg-3’			
Yield	277,7 µg	**Ext. Coeff.**	318800
Vol.f.100	285 pmol/µL	**MW Calc.**	9728 g/mol
Length	32	**MW Found**	9746 g/mol
Tm	78 ºC	**Purific.**	HPLC
GC content	62,00 %	A:4,0 C:12,0 G:8,0 T/U:8,0	
Malectin_S80Q_F			
5’- ttc atg cca tag tct tgg gct cgg ccc acc cg-3’			
Yield	277,7 µg	**Ext. Coeff.**	318800
Vol.f.100	285 pmol/µL	**MW Calc.**	9728 g/mol
Length	32	**MW Found**	9746 g/mol
Tm	78 ºC	**Purific.**	HPLC
GC content	62,00 %	A:4,0 C:12,0 G:8,0 T/U:8,0	
Malectin_S80Q_F			
5’- ttc atg cca tag tct tgg gct cgg ccc acc cg-3’			
Yield	277,7 µg	**Ext. Coeff.**	318800
Vol.f.100	285 pmol/µL	**MW Calc.**	9728 g/mol
Length	32	**MW Found**	9746 g/mol

Tm	78 ºC	**Purific.**	HPLC
GC content	62,00 %	A:4,0 C:12,0 G:8,0 T/U:8,0	
Malectin_S80Q_F			
5’- ttc atg cca tag tct tgg gct cgg ccc acc cg-3’			
Yield	277,7 µg	**Ext. Coeff.**	318800
Vol.f.100 pmol/µL	285 µL	**MW Calc.**	9728 g/mol
Length	32	**MW Found**	9746 g/mol
Tm	78 ºC	**Purific.**	HPLC
GC content	62,00 %	A:4,0 C:12,0 G:8,0 T/U:8,0	
Malectin_S80Q_R			
5’- cgg gtg ggc cga gcc caa gac tat ggc atg aa-3’			
Yield	88,1 µg	**Ext. Coeff.**	360400
Vol.f.100 pmol/µL	89 µL	**MW Calc.**	9924 g/mol
Length	32	**MW Found**	9945 g/mol
Tm	78 ºC	**Purific.**	HPLC
GC content	62,00 %	A:8,0 C:8,0 G:12,0 T/U:4,0	
Malectin_E102A_F			
5’- tct cct cat tgt acc gcg cag ttt gat aca gga tc-3’			
Yield	350,9 µg	**Ext. Coeff.**	365900
Vol.f.100 pmol/µL	329 µL	**MW Calc.**	10673 g/mol
Length	35	**MW Found**	10694 g/mol

Tm	71 ºC	**Purific.**	HPLC
GC content	48,00 %	A:7,0 C:10,0 G:7,0 T/U:11,0	
Malectin_E102A_R			
5’- gat cct gta tca aac tgc gcg gta caa tga gga ga-3’			
Yield	141,0 µg	**Ext. Coeff.**	403300
Vol.f.100 pmol/µL	130 µL	**MW Calc.**	10829 g/mol
Length	35	**MW Found**	10850 g/mol
Tm	71 ºC	**Purific.**	HPLC
GC content	48,00 %	A:11,0 C:7,0 G:10,0 T/U:7,0	
Malectin_Y104F_F			
5’- caa agg tct cct cat tga acc gct cag ttt gat ac-3’			
Yield	360,2 µg	**Ext. Coeff.**	375800
Vol.f.100 pmol/µL	338 µL	**MW Calc.**	10666 g/mol
Length	35	**MW Found**	10686 g/mol
Tm	69 ºC	**Purific.**	HPLC
GC content	45,00 %	A:9,0 C:10,0 G:6,0 T/U:10,0	
Malectin_Y104F_R			
5’- gta tca aac tga gcg gtt caa tga gga gac ctt tg-3’			
Yield	300,5 µg	**Ext. Coeff.**	398800
Vol.f.100 pmol/µL	277 µL	**MW Calc.**	10835 g/mol
Length	35	**MW Found**	10855 g/mol

Tm	69 ºC	**Purific.**	HPLC
GC content	45,00 %	A:10,0 C:6,0 G:10,0 T/U:9,0	
Malectin_131F_F			
5’- gct ggg act gtg caa aga aga cct ctg caa att tc-3’			
Yield	219,2 µg	**Ext. Coeff.**	392900
Vol.f.100 pmol/µL	203 µL	**MW Calc.**	10780 g/mol
Length	35	**MW Found**	10799 g/mol
Tm	72 ºC	**Purific.**	HPLC
GC content	48,00 %	A:10,0 C:8,0 G:9,0 T/U:8,0	
Malectin_Y131F_R			
5’- gaa att tgc aga ggt ctt ctt tgc aca gtc cca gc-3’			
Yield	352,2 µg	**Ext. Coeff.**	376300
Vol.f.100 pmol/µL	328 µL	**MW Calc.**	10722 g/mol
Length	35	**MW Found**	10741 g/mol
Tm	72 ºC	**Purific.**	HPLC
GC content	48,00 %	A:8,0 C:9,0 G:8,0 T/U:10,0	
Malectin_Y199AY200A_F			
5’- tct aca ttg agt ttg tca agg ggg ccg ctg acavatc cca agg tct gtg cac-3’			
Yield	390,4 µg	**Ext. Coeff.**	550300
Vol.f.100 pmol/µL	249 µL	**MW Calc.**	15706 g/mol
Length	51	**MW Found**	15722 g/mol

Tm	84 ºC	**Purific.**	HPLC
GC content	52,00 %	A:11,0 C:13,0 G:14,0 T/U:13,0	
Malectin_Y199AY200A_R			
5’- gtg cac aga cct tgg gat tgt cag cgg ccc cct tga caa act caa tgt aga-3’			
Yield	115,1 µg	**Ext. Coeff.**	558500
Vol.f.100 pmol/µL	73 µL	**MW Calc.**	15684 g/mol
Length	51	**MW Found**	15698 g/mol
Tm	84 ºC	**Purific.**	HPLC
GC content	52,00 %	A:13,0 C:14,0 G:13,0 T/U:11,0	
Malectin_D201A_F			
5’- cac aga cct tgg gat tgg cat agt acc cct tga ca-3’			
Yield	146,9 µg	**Ext. Coeff.**	380800
Vol.f.100 pmol/µL	137 µL	**MW Calc.**	10716 g/mol
Length	35	**MW Found**	10733 g/mol
Tm	72 ºC	**Purific.**	HPLC
GC content	51,00 %	A:9,0 C:10,0 G:8,0 T/U:8,0	
Malectin_D201A_R			
5’- tgt caa ggg gta cta tgc caa tcc caa ggt ctg tg-3’			
Yield	330,9 µg	**Ext. Coeff.**	383000
Vol.f.100 pmol/µL	307 µL	**MW Calc.**	10787 g/mol
Length	35	**MW Found**	10791 g/mol

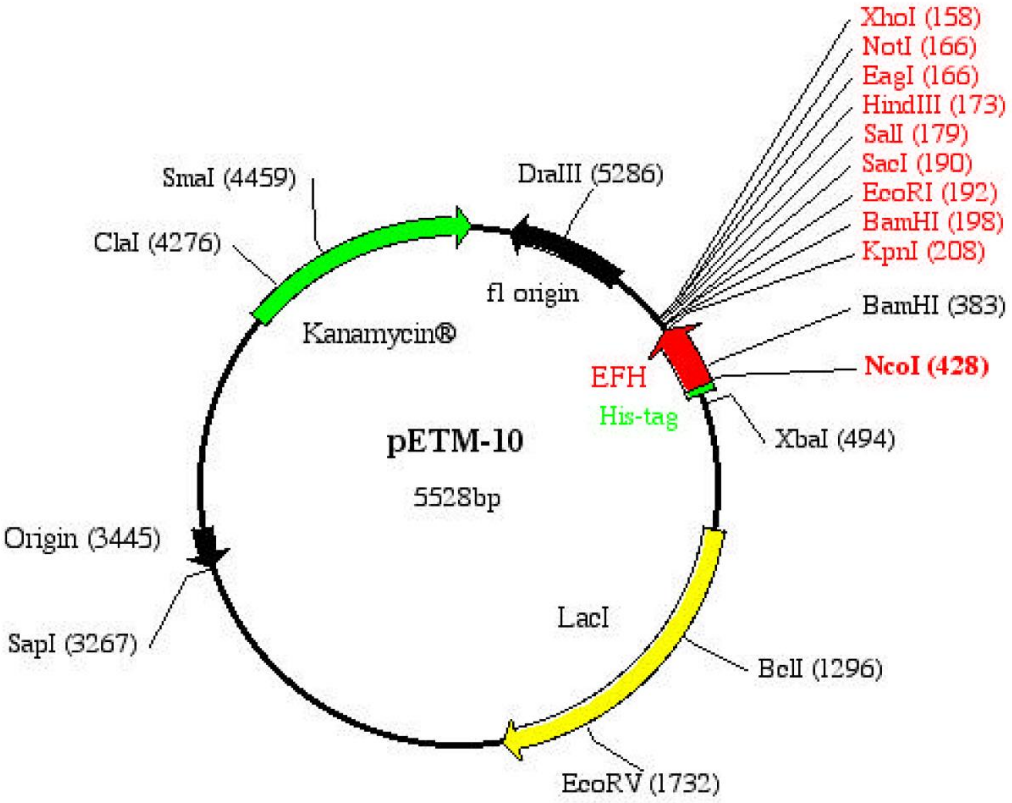
Tm	72 ºC	**Purific.**	HPLC
GC content	51,00 %	A:8,0 C:8,0 G:10,0 T/U:9,0	

Supplementary table 5 – Electrophoresis gel preparation.

Tris-Acetate-EDTA (TAE) buffer	Water
	40mM Tris Base, pH 8.3
	20 mM Acetic Acid
	1mM Na_2EDTA
0,8% agarose	

Supplementary table 6 – SDS-PAGE gel preparation.

10% Separating gel	Water
	3M Tris-HCL/SDS, pH 8.45
	30% acrylamide
	glycerol
	30% APS
	TEMED
4% Stacking Gel	Water
	3M Tris-HCl/SDS, pH 8.45
	30% acrylamide
	30% APS
	TEMED




```
     T7 promoter  -->            Lac operator      XbaI
CGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAA
GCTTTAATTATGCTGAGTGATATCCCCTTAACACTCGCCTATTGTTAAGGGGAGATCTTT

                        rbs                  His-tag
TAATTTTGTTTAACTTTAAGAAGGAGATATACCATGAAACATCACCATCACCATCACCCC
ATTAAAACAAATTGAAATTCTTCCTCTATATGGTACTTTGTAGTGGTAGTGGTAGTGGGG
                                  METLysHisHisHisHisHisHisPro

NcoI
ATGGCTGACACCGACACTGCCGAG..171bp..GCACTCTACGGGGAGAGCGATCTGTGA
TACCGACTGTGGCTGTGACGGCTC...EFH...CGTGAGATGCCCCTCTCGCTAGACACT
METAlaAspThrAspThrAlaGlu...57aa..AlaLeuTyrGlyGluSerAspLeu***

                                      NotI
       BamHI        SacI              EagI
 KpnI        EcoRI          SalI HindIII        XhoI  C-His-tag
GGTACCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCAC
CCATGGCCTAGGCTTAAGCTCGAGGCAGCTGTTCGAACGCCGGCGTGAGCTCGTGGTGGTG
                                                  HisHisHis
```

Supplementary Figure 1 - Schematic illustration and sequence of pETM-10 vector. Taken by EMBL Protein Expression and Purification Facility.

Supplementary information:

SOB (Super optimal broth) medium

2% peptone
0.5% Yeast extract
10 mM NaCl
2.5 mM KCL
Adjust the pH to 7.5 and autoclave
Then add 10 mM $MgCl_2$ and 10 mM $MgSO_4$.

TB (Transformation buffer)

250 M KCL
15 mM $CaCl_2$.2H2O
10 mM PIPES
Dissolve the components and adjust the pH to 6.7. Then add:
55 mM MnCl2
Filter with a 0.45 μm sterile filter and store at 4ºC.

Luria Bertani medium culture (1L)

10g Tryptone (Sigma-Aldrich®)
10g NaCl (Panreac®)
5g Yeast (NZYeTch®)
Required volume of distilled water.

TFBI (Transformation buffer I)

30 mM KOAc
100 mM RbCl
10 mM $CaCl_2$ + 50 mM MnCl215% glycerol
Adjust to pH 5.8 with acetic acid and filter (0.45 μm) to sterilize.

TFBII (Transformation buffer II)

10 mM MOPS
75 mM $CaCl_2$
10 mM RbCl
15% glycerol
Adjust pH to 6.5 with KOH and filter to sterilize.

Anti-Mouse IgG

(whole molecule) – Biotin antibody produced in goat – affinity isolated buffered aqueous solution
Product number: B7264
Batch number: SLBS5654
Brand: SIGMA
Protein content: 0.3 mg/mL

Anti-His antibody

Monoclonal Anti-polyHistidine antibody produced in mouse
Product number: H1029
Batch number: 097M4894V
Brand: SIGMA
Protein content: 2.2 mg/mL

Table 7 - Fungal, bacterial & plant PS Set 2. The first column indicates the probe position, and the second column depicts the name and/or the ligand structure. The dextran polysaccharide probe that interacted with malectin WT is highlighted in *light green*.

1	Empty
2	NSG-β-glucan
3	PGG-β-glucan
4	Mannan (S. cerevisiae)
5	Lichenan Mixed-linked β1-3, β1-4 glucose
6	Xylan (X1) Sample rich in xylans; Rha (3%), Fuc (2%), Ara (11%), Xyl (67%), Man (0%), Gal (6%), Glc (5%), Ur Ac (6%)
7	Xyloglucan (XG1) Sample rich in xyloglucans;Rha (2%), Fuc (5%), Ara (5%), Xyl (40%), Man (6%), Gal (13%), Glc (24%), Ur Ac (6%)
8	Xyloglucan (XG2) Sample rich in xyloglucans;Rha (2%), Fuc (6%), Ara (6%), Xyl (46%), Man (4%), Gal (14%), Glc (22%), Ur Ac (1%)
9	Dextran (170 kDa) (Amersham Biosciences)
10	Pullulan Megazyme
11	Lentinan
12	Grifolan
13	Pustulan U. papullosa (Calbiochem)
14	Barley glucan
15	Barley glucan Mixed-linked β1-3, β1-4 glucose (Medium viscosity, Megazyme)
16	Curdlan (50 mM NaOH)

17	Pullulan Megazyme
18	Oat β-glucan Mixed-linked β1-3, β1-4 glucose (Medium viscosity, Megazyme)
19	Xylan Elicityl
20	Xylan (X2) Sample rich in xylans; Rha (3%), Fuc (3%), Ara (12%), Xyl (73%), Man (0%), Gal (0%), Glc (4%), Ur Ac (5%)
21	Arabinoxylan (AX1) Dreche; Sample rich in arabinoxylans; KOH 0.1 M (DP 41); Ara (40%), Xyl (54%), Man (0%), Gal (3%), Glc (3%)
22	Arabinoxylan (AX2) Dreche; Sample rich in arabinoxylans; MW Et70 (DP24); Ara (25%), Xyl (46%), Man (1%), Gal (3%), Glc (25%)
23	Arabinogalactan (AG1) Sample rich in arabinogalactans (High MW, M23pp); Ara (5%), Man (29%), Gal (64%), Glc (1%)
24	Galactomannan (GM1) Sample rich in galactomannans (High MW); Ara (5%), Man (62%), Gal (31%), Glc (2%)
25	Galactomannan Guar (Megazyme)
26	Galactomannan Carob (low viscosity, Megazyme)
27	Galactomannan Guar (Galactose depleted, Megazyme)
28	Galactomannan (Hydrolysed, Megazyme)
29	Glucurono-XyloMannan Tremella fuciformi (Elicityl) (250 mM NaCl)
30	Galactomannan (GM2) Sample rich in galactomannans (High MW); Ara (2%), Man (87%), Gal (10%), Glc (2%)
31	Mannoprotein (MP1) Yeast cell wall; sequential extraction 4M KOH; Ara (1%), Xyl (1%), Man (85%), Glc (13%)
32	Mannoprotein (MP2) Yeast cell wall; sequential extraction 8M KOH; Ara (1%), Xyl (0%), Man (65%), Glc (35%)
33	Pectin-1
34	Pectin-2
35	Pectin-3
36	Pectin-4
37	Pectin-5

38	Pectin-6
39	Pectin-7
40	Pectin-8
41	Pectin-9
42	Pectin-10
43	Pectin-11
44	Pectin-12
45	Pectin-13
46	Glucurono-XyloMannan Tremella fuciformi (Elicityl) (250 mM NaCl)
47	Galactomannan Carob (low viscosity, Megazyme)
48	Oat β-glucan Mixed-linked β1-3, β1-4 glucose (Medium viscosity, Megazyme)
49	Mannan (C.albicans)
50	Mannan GP (A fumigatus)
51	H37Rv LAM BEI Number NR-14848; Purified Lipoarabinomannan (LAM), Strain H37Rv
52	Mycobacterium smegmatis LAM BEI Number NR-14849; Purified Lipoarabinomannan (LAM) from Mycobacterium smegmatis
53	H37Rv LM BEI Number NR-14850; H37Rv, Purified Lipomannan (LM)
54	H37Rv arabinogalactan BEI Number NR-14852; H37Rv, Purified arabinogalactan
55	H37Rv Native Protein BEI Number NR-14862; MPT32/Apa (Gene Rv1860), Purified Native Protein from Mycobacterium tuberculosis, Strain H37Rv
56	Glucurono-XyloMannan Tremella fuciformi(Elicityl)(250 mM NaCl)
57	Pectin-2
58	Pectin-10
59	Arabinoxylan (AX2) Dreche; Sample rich in arabinoxylans; MW Et70 (DP24); Ara (25%), Xyl (46%), Man (1%), Gal (3%), Glc (25%)
60	H37Rv arabinogalactan BEI Number NR-14852; H37Rv, Purified arabinogalactan
61	Glucurono-XyloMannan Tremella fuciformi (Elicityl) (250 mM NaCl)

62	CβG glucan (neutral) Cyclic β1-2 glucose
63	Cello-6-AO Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glc-AO
64	GN6-AO GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAc-AO

Supplementary table 8 - NGL probes Set (*42-56*).

Set 42	
1	Hanfland 1 Glcβ-Cer
2	Galactosylceramide Galβ-Cer
3	Lac Galβ-4Glc-DH
4	Lactocerebrosides Galβ-4Glcβ-Cer
5	GalNAcα-3Galβ-4Glc GalNAcα-3Galβ-4Glc-DH
6	Galα-4Galβ-4GlcNAc Galα-4Galβ-4GlcNAc-DH
7	Ceramide trihexoside Galα-4Galβ-4Glcβ-Cer
8	Globoside (P-antigen) alNAcβ-3Galα-4Galβ-4Glcβ-Cer
9	Forssmann glycolipid alNAcα-3GalNAcβ-3Galα-4Galβ-4Glcβ-Cer
10	B-Tri Galα-3Gal-DH | Fucα-2
11	NeuAcα-(3')Lac NeuAcα-3Galβ-4Glc-DH
12	NeuAcα-(6')Lac NeuAcα-6Galβ-4Glc-DH
13	NeuAcα-(3')LN NeuAcα-3Galβ-4GlcNAc-DH

14	NeuAcα-(6')LN NeuAcα-6Galβ-4GlcNAc-DH
15	Neu5,9Ac-(6')LN (9-OAc)NeuAcα-6Galβ-4GlcNAc-DH
16	GSC-210 SU-3GlcAβ-3Galβ-Cer42
17	GSC-150 SU-3Galβ-4Glcβ-C30 Fucα-3
18	SU(3')-Lea-Tri SU-3Galβ-3GlcNAc-DH Fucα-4
19	SU(3')-Lex-Tri SU-3Galβ-4GlcNAc-DH Fucα-3
20	GSC-209 GlcAβ-3Galβ-Cer42
21	GSC-432 (3-deoxy,3-carboxymethyl)Galβ-4Glcβ-C30
22	GSC-430 (3-deoxy,3-carboxymethyl)Galβ-3Glcβ-C30 Fucα-4
23	Glucocerebrosides Glcβ-Cer
24	Galactocerebrosides Galβ-Cer
25	Lac-AO Galβ-4Glc-AO
26	LacNAc-AO Galβ-4GlcNAc-AO
27	Lea-Tri Galβ-3GlcNAc-DH Fucα-4
28	Lea-Tri-AO Galβ-3GlcNAc-AO Fucα-4

29	Lex-Tri Galβ-4GlcNAc-DH | Fucα-3
30	Lex-Tri-AO Galβ-4GlcNAc-AO | Fucα-3
31	LNFP-III Galβ-4GlcNAcβ-3Galβ-4Glc-DH | Fucα-3
32	NeuAcα-(3')Lac-AO NeuAcα-3Galβ-4Glc-AO
33	NeuAcα-(6')Lac-AO NeuAcα-6Galβ-4Glc-AO
34	NeuAcβ-(3')Lac NeuAcβ-3Galβ-4Glc-DH
35	NeuAcβ-(3')Lac-AO NeuAcβ-3Galβ-4Glc-AO
36	NeuAcβ-(6')Lac NeuAcβ-6Galβ-4Glc-DH
37	NeuAcβ-(6')Lac-AO NeuAcβ-6Galβ-4Glc-AO
38	Neuα-(3')Lac Neuα-3Galβ-4Glc-DH
39	Neuα-(3')Lac-AO Neuα-3Galβ-4Glc-AO
40	Neuα-(6')Lac Neuα-6Galβ-4Glc-DH
41	Neuα-(6')Lac-AO Neuα-6Galβ-4Glc-AO
42	Neu4,5Ac-(3')Lac (4-OAc)NeuAcα-3Galβ-4Glc-DH
43	Neu4,5Ac-(3')Lac-AO (4-OAc)NeuAcα-3Galβ-4Glc-AO
44	Sulfatide SU-3Galβ-Cer
45	GSF-1 SU-3Galβ-C30

46	GSF-19 SU-6Glcβ-C30
47	GSC-260 3-deoxy,3-carboxymethyl-Galβ-4Glcβ-C30 Fucα-3
48	H-Di Fucα-2Gal-DH
49	A-Tri GalNAcα-3Gal-DH Fucα-2
50	SA(3')-Lea-Tri NeuAcα-3Galβ-3GlcNAc-DH Fucα-4
51	SU(3')-LN SU-3Galβ-4GlcNAc-DH
52	LacNAc Galβ-4GlcNAc-DH
53	SAA#1 NeuAcα-3Galβ-3GlcNAc-DH F-2
54	SAA#1-AO NeuAcα-3Galβ-3GlcNAc-AO
55	NeuAcα-(3')LN-AO NeuAcα-3Galβ-4GlcNAc-AO
56	Lex-Tri-(Me)AO Galβ-4GlcNAc-(Me)AO Fucα-3
57	LacNAc(1-3) Galβ-3GlcNAc-DH
58	LacNAc(1-3)-AO Galβ-3GlcNAc-AO
59	GSC-426 (3-deoxy,3-carboxymethyl)Galβ-C30
60	GSC-17 NeuAcα-3Galβ-4Glcβ-Cer36
61	GSC-511 (9-OAc)NeuAcα-3Galβ-4GlcNAcβ-C30 Fucα-3

62	GSC-513 (9-OAc)NeuAcα-3Galβ-3GlcNAcβ-C30 | Fucα-4
63	GSC-75 (4-deoxy)NeuAcα-3Galβ-4Glcβ-Cer36
64	GSC-76 (7-deoxy)NeuAcα-3Galβ-4Glcβ-Cer36
Set 43	
1	GSC-77 (8-deoxy)NeuAcα-3Galβ-4Glcβ-Cer36
2	GSC-51 (9-deoxy)NeuAcα-3Galβ-4Glcβ-Cer36
3	GSC-153 (4,8-deoxy)NeuAcα-3Galβ-4Glcβ-Cer36
4	GSC-78 (4-OMe)NeuAcα-3Galβ-4Glcβ-Cer36
5	GSC-79 (9-OMe)NeuAcα-3Galβ-4Glcβ-Cer36
6	GSC-61 NeuAcα-6Galβ-4Glcβ-Cer36
7	GSC-437 NeuAcα-8NeuAcα-8NeuAcα-3Galβ-4Glcβ-Cer36
8	GSC-144 KDNα-6Galβ-Cer36
9	GSC-197 KDNα-3Galβ-4Glcβ-Cer28
10	GSC-198 KDNα-3Galβ-4Glcβ-Cer34
11	GSC-96 NeuAcα-9NeuAcα-3Galβ-4Glcβ-Cer36
12	GSC-23 (C7)NeuAcα-3Galβ-4Glcβ-Cer36
13	GSC-24 (C8)NeuAcα-3Galβ-4Glcβ-Cer36
14	GSC-199 KDNα-3Galβ-4Glcβ-C30
15	GSC-229 NeuAcα-8NeuAcα-3Galβ-4Glcβ-Cer36

16	GSC-230 NeuAcα-8NeuAcα-3Galβ-Cer36
17	GSC-231 NeuAcα-8NeuAcα-6Galβ-Cer36
18	GSC-232 NeuAcα-8NeuAcα-6Glcβ-Cer36
19	GSC-234 NeuAcα-(S)-6Galβ-(S)-4Glcβ-Cer36
20	GSC-296 GlcAβ-3Galβ-4Glcβ-C30
21	GSC-353 SU-3GlcAβ-3Galβ-4Glcβ-C30
22	GSC-439 NeuAcα-8NeuAcα-8NeuAcα-6Galβ-Cer36
23	GSC-440 NeuAcα-3Galβ-4GlcNAcβ-C30 | Fucα-3
24	GSC-512 (4-OAc)NeuAcα-3Galβ-4GlcNAcβ-C30 | Fucα-3
25	GSC-9 NeuAcα-(S)-6Glcβ-Cer36
26	GSC-12 NeuAcα-(S)-6Galβ-4Glcβ-Cer36
27	GSC-13 NeuAcα-(S)-6Galβ-Cer36
28	GSC-16 NeuAcα-3Galβ-4Glcβ-Cer32
29	GSC-18 NeuAcα-3Galβ-4Glcβ-Cer42
30	GSC-27 NeuAcα-6Galβ-Cer36
31	LNFP-III Galβ-4GlcNAcβ-3Galβ-4Glc-DH | Fucα-3
32	GSC-40 NeuAcα-(S)-3Galβ-Cer42

33	GSC-50 (C8 diastereoisomer)NeuAcα-3Galβ-4Glcβ-Cer36
34	GSC-59 NeuAcα-6GlcNAcβ-Cer36
35	GSC-60 NeuAcα-6Glcβ-Cer36
36	GSC-62 NeuAcα-2Glcβ-Cer36
37	GSC-72 NeuAcα-(S)-6Galβ-(S)-Cer36
38	GSC-73 NeuAcα-(S)-6Galβ-4Glcβ-(S)-Cer36
39	GSC-95 NeuAcα-(S)-6GlcNAcβ-Cer36
40	GSC-160 SU-3Galβ-4Glcβ-Cer36 | Fucα-3
41	GSC-161 NeuAcα-3Galβ-4Glcβ-C30 | Fucα-3
42	GSC-162 NeuAcα-3Galβ-4Glcβ-Cer36 | Fucα-3
43	GSC-178 NeuAcα-3Galβ-4Glcβ-Cer34
44	GSC-187 NeuAcα-3Galβ-C29
45	PI-1 NeuAcα-3(6-NAc)Galβ-4GlcNAc-DH
46	PI-1-AO NeuAcα-3(6-NAc)Galβ-4GlcNAc-AO
47	PI-2 NeuAcα-3(6-NBz)Galβ-4GlcNAc-DH
48	PI-2-AO NeuAcα-3(6-NBz)Galβ-4GlcNAc-AO
49	SA(3')-Lea-Tri-AO NeuAcα-3Galβ-3GlcNAc-AO | Fucα-4

50	SA(3')-Lex-Tri-AO NeuAcα-3Galβ-4GlcNAc-AO | Fucα-3
51	NeuAcα-(6')LN-AO NeuAcα-6Galβ-4GlcNAc-AO
52	Globotri-AO Galα-4Galβ-4Glc-AO
53	B-Tri-AO Galα-3Gal-AO | Fucα-2
54	Fuc(3)-Lac-AO Galβ-4Glc-AO | Fucα-3
55	SA(3')-Lex-Tri NeuAcα-3Galβ-4GlcNAc-DH | Fucα-3
56	A-tri-AO GalNAcα-3Gal-AO | Fucα-2
57	A-Tetra-T1-AO GalNAcα-3Galβ-3GlcNAc-AO | Fucα-2
58	A-Tetra-T2-AO GalNAcα-3Galβ-4GlcNAc-AO | Fucα-2
59	B-Tetra-T1-AO Galα-3Galβ-3GlcNAc-AO | Fucα-2
60	B-Tetra-T2-AO Galα-3Galβ-4GlcNAc-AO | Fucα-2
61	B-penta-AO Galα-3Galβ-4Glc-AO | | Fucα-2 | Fucα-3
62	H-Tri-T2-AO Fucα-2Galβ-4GlcNAc-AO

63	Su(3')-Lea tri-AO SU-3Galβ-3GlcNAc-AO | Fucα-4
64	Su(3')-Lex tri-AO SU-3Galβ-4GlcNAc-AO | Fucα-3
Set 44	
1	NeuGcα-(3')LN NeuGcα-3Galβ-4GlcNAc-DH
2	Globo-H-Hexa Fucα-2Galβ-3GalNAcβ-3Galα-4Galβ-4GlcNAc-DH
3	Globo-A-Hepta GalNAcα-3Galβ-3GalNAcβ-3Galα-4Galβ-4GlcNAc-DH | Fucα-2
4	Globo-B-Hepta Galα-3Galβ-3GalNAcβ-3Galα-4Galβ-4GlcNAc-DH | Fucα-2
5	LNT Galβ-3GlcNAcβ-3Galβ-4Glc-DH
6	LNnT Galβ-4GlcNAcβ-3Galβ-4Glc-DH
7	Paragloboside Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer
8	B-like pentaosylceramide Galα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer
9	Klaus glycolipid Galβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer
10	LNFP-I Fucα-2Galβ-3GlcNAcβ-3Galβ-4Glc-DH
11	LNFP-II Galβ-3GlcNAcβ-3Galβ-4Glc-DH | Fucα-4
12	A-Hexa GalNAcα-3Galβ-3GlcNAcβ-3Galβ-4Glc-DH | Fucα-2
13	LNnDFH-I Fucα-2Galβ-4GlcNAcβ-3Galβ-4Glc-DH | Fucα-3

14	LNnDFH-II Galβ-4GlcNAcβ-3Galβ-4Glc-DH | | Fucα-3 Fucα-3
15	LNDFH-II Galβ-3GlcNAcβ-3Galβ-4Glc-DH | | Fucα-4 Fucα-3
16	LNnTFH-I Fucα-2Galβ-4GlcNAcβ-3Galβ-4Glc-DH | | Fucα-3 Fucα-2
17	LNTFH-I Fucα-2Galβ-3GlcNAcβ-3Galβ-4Glc-DH | | Fucα-4 Fucα-2
18	LSTa NeuAcα-3Galβ-3GlcNAcβ-3Galβ-4Glc-DH
19	Sialylparagloboside NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer
20	SA(3')-LNFP-II NeuAcα-3Galβ-3GlcNAcβ-3Galβ-4Glc-DH | Fucα-4
21	SA(3')-LNFP-III NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH | Fucα-3
22	GSC-64 NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36 | Fucα-3
23	SA(3/6)LNFP-I NeuAcα-3/6Galβ-3GlcNAcβ-3Galβ-4Glc-DH | Fucα-2
24	GSC-472 Neuα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36 | Fucα-3
25	DSLNT NeuAcα-3Galβ-3GlcNAcβ-3Galβ-4Glc-DH | NeuAcα-6
26	GSC-190 SU-3GlcAβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer42
27	GSC-208 SU-3GlcAβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-C30

28	SU(3')-LNFP-III SU-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH Fucα-3
29	SU(3',6)-LNFP-III SU-6 SU-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH Fucα-3
30	SU(6')-LNFP-III SU-6Galβ-4GlcNAcβ-3Galβ-4Glc-DH Fucα-3
31	LNFP-III Galβ-4GlcNAcβ-3Galβ-4Glc-DH Fucα-3
32	GSC-406 SU-6 Neuα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36 Fucα-3
33	GSC-268 deNAc SU-6 Neuα-3Galβ-4GlcNβ-3Galβ-4Glcβ-Cer36 Fucα-3
34	GSC-269 SU-6 NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36 Fucα-3
35	GSC-270 SU-6 SU-6 NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36 Fucα-3
36	GSC-189 GlcAβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer42
37	GSC-207 GlcAβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-C30
38	GSC-268 SU-6 NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36 Fucα-3

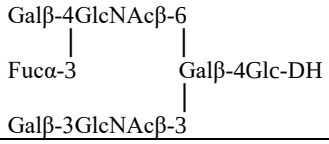
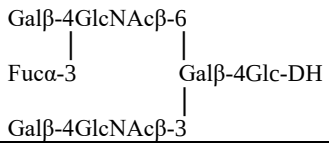
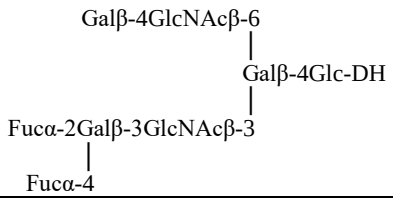
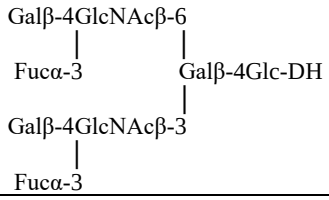
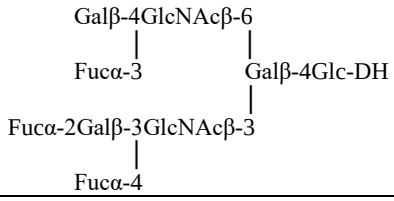
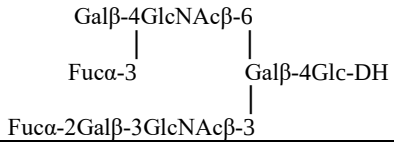
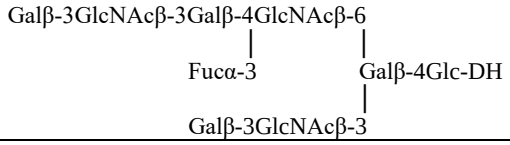
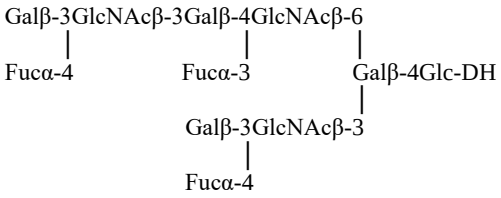
39	SA(6')-LNFP-VI NeuAcα-6Galβ-4GlcNAcβ-3Galβ-4Glc-DH Fucα-3
40	LNFP-III-AO Galβ-4GlcNAcβ-3Galβ-4Glc-AO Fucα-3
41	SU(3')-LNFP-II SU-3Galβ-3GlcNAcβ-4Galβ-4Glc-DH Fucα-4
42	LSTc NeuAcα-6Galβ-4GlcNAcβ-3Galβ-4Glc-DH
43	LSTb Galβ-3GlcNAcβ-3Galβ-4Glc-DH NeuAcα-6
44	LNnDFH-V Galβ-4GlcNAcβ-3Galβ-4Glc-DH Fucα-3 Fucα-2
45	LNDFH-I Fucα-2Galβ-3GlcNAcβ-3Galβ-4Glc-DH Fucα-4
46	GSC-97 NeuAcα-6Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36 Fucα-3
47	A-Hepta Fucα-4 GalNAcα-3Galβ-3GlcNAcβ-3Galβ-4Glc-DH Fucα-2
48	Leb-hexaosylceramide Fucα-2Galβ-3GlcNAcβ-3Galβ-4Glcβ-Cer Fucα-4
49	B-hexaosylceramide Galα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer Fucα-2
50	GSC-31 NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36
51	SU(3')-Tri SU-3Galβ-4GlcNAcβ-3Gal-DH
52	SU(6')-LNFP-II SU-6Galβ-3GlcNAcβ-3Galβ-4Glc-DH Fucα-4

53	Led-II pentaosylceramide Fucα-2Galβ-3GlcNAcβ-3Galβ-4Glcβ-CerA
54	Led-I pentaosylceramide Fucα-2Galβ-3GlcNAcβ-3Galβ-4Glcβ-CerB
55	DLNN GlcNAcβ-3Galβ-4Glc-DH
56	GSC-105 NeuAcα-3Galβ-4GlcNAcβ-3Galβ-Cer36 | Fucα-3
57	GSC-516B Neuα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36 | SU-6
58	GSC-177 NeuGcα-3Galβ-4GlcNAcβ-3Galβ-Cer36 | Fucα-3
59	GSC-396 NeuGcα-3Galβ-3GlcNAcβ-3Galβ-4Glcβ-C30
60	GSC-147 KDNα-3Galβ-3GlcNAcβ-3Galβ-4Glcβ-Cer36
61	GSC-341 KDNα-3Galβ-4GlcNAcβ-3Galβ-C30 | Fucα-3
62	GSC-149 KDNα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36 | Fucα-3
63	GSC-131 NeuAcα-3Galβ-4GlcNAcβ-3Galβ-Cer36 | Quvα-3
64	GSC-163 NeuAcα-3Galβ-4GlcNAcβ-3Galβ-Cer36 | Rhaα-3
Set 45	
1	GSC-191 GlcAβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36
2	GSC-192 SU-3GlcAβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36
3	GSC-225 (3-deoxy,3-carboxymethyl)Galβ-4GlcNAcβ-3Galβ-Cer36 | Fucα-3

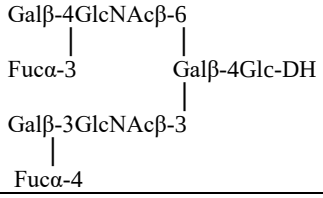
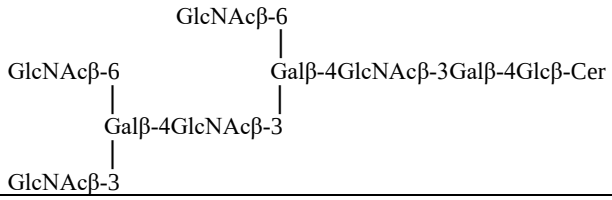
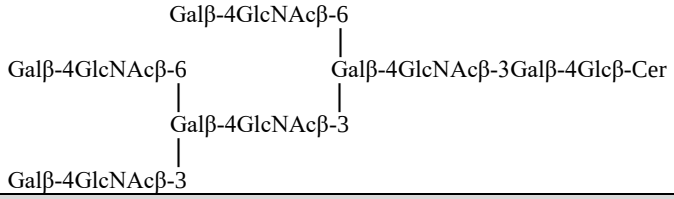
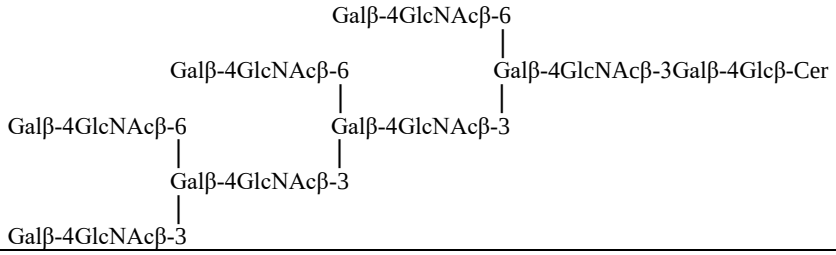
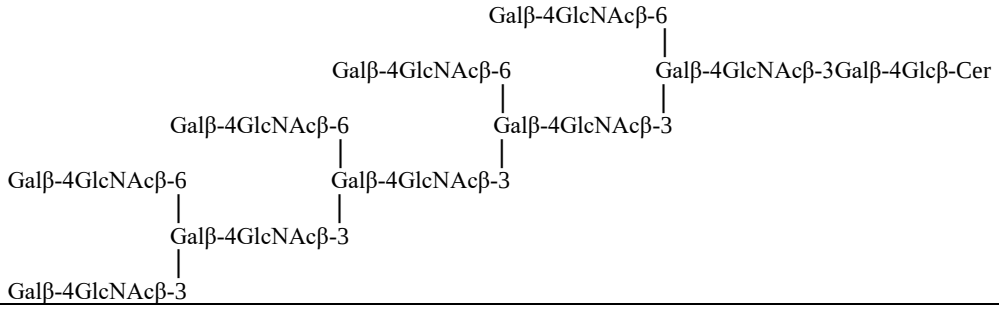
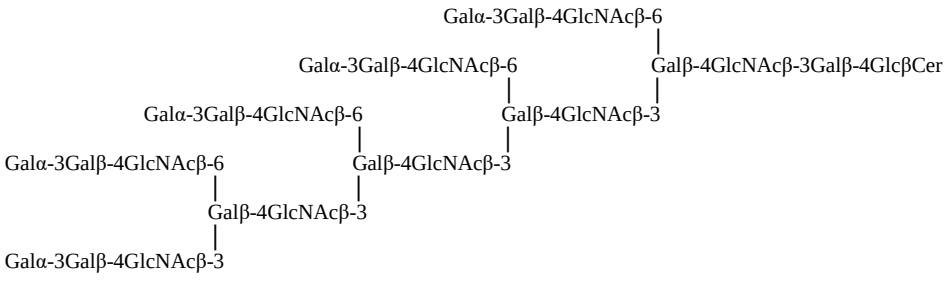
4	GSC-236 SU-3Galβ-4GlcNAcβ-3Galβ-C30 | Fucα-3
5	GSC-257 NeuAcα-3(4,6-deoxy)Galβ-4GlcNAcβ-3Galβ-Cer36 | Fucα-3
6	GSC-272 NeuAcα-3Galβ-3GlcNAcβ-3Galβ-4Glcβ-C30
7	GSC-273 NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-C30
8	GSC-311 KDNα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-C30 | Rhaα-3
9	GSC-314 KDNα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-C30 | Fucα-3
10	GSC-397 NeuGcα-6Galβ-3GlcNAcβ-3Galβ-4Glcβ-C30
11	GSC-479 NeuAcα-3Galβ-4GlcNAcβ-3Galβ-C30 | Fucα-3
12	GSC-533 NeuAcα-3Galβ-4GlcNβ-3Galβ-4Glcβ-Cer36 | Fucα-3
13	GSC-121 NeuAcα-3Galβ-4GlcNAcβ-3Galβ-Cer36 | (3-deoxy)Fucα-3
14	GSC-123 NeuAcα-3Galβ-4GlcNAcβ-3Galβ-Cer36 | (4-deoxy)Fucα-3
15	GSC-127 NeuAcα-3Galβ-4GlcNAcβ-3Galβ-Cer36 | (6-deoxy)Talα-3
16	GSC-133 NeuAcα-3Galβ-4GlcNAcβ-3Galβ-Cer36 | (2-OMe)Fucα-3
17	GSC-175 NeuAcα-3(4-deoxy)Galβ-4GlcNAcβ-3Galβ-Cer36 | Fucα-3

18	GSC-176 NeuAcα-3(6-deoxy)Galβ-4GlcNAcβ-3Galβ-Cer36 | Fucα-3
19	Leb tetra-AO Fucα-2Galβ-3GlcNAc-AO | Fucα-4
20	LNFP-I-AO Fucα-2Galβ-3GlcNAcβ-3Galβ-4Glc-AO
21	LNFP-II-AO Galβ-3GlcNAcβ-3Galβ-4Glc-AO | Fucα-4
22	LNDFH-I-AO Fucα-2Galβ-3GlcNAcβ-3Galβ-4Glc-AO | Fucα-4
23	LNnDFH-I-AO Fucα-2Galβ-4GlcNAcβ-3Galβ-4Glc-AO | Fucα-3
24	LNnFP-I-AO Fucα-2Galβ-4GlcNAcβ-3Galβ-4Glc-AO
25	Orsay-1-AO Galβ-4GlcNAcβ-6Gal-AO
26	Orsay-2-AO Galβ-4GlcNAcβ-3Gal-AO
27	Orsay-3-AO Galβ-3GlcNAcβ-6Gal-AO
28	Orsay-4-AO Galβ-3GlcNAcβ-3Gal-AO
29	LSTd NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH
30	pLNH Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH
31	LNFP-III Galβ-4GlcNAcβ-3Galβ-4Glc-DH | Fucα-3
32	pLNnH Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH
33	LNnH Galβ-4GlcNAcβ-6 | Galβ-4Glc-DH | Galβ-4GlcNAcβ-3

34	LNH Galβ-4GlcNAcβ-6 | Galβ-4Glc-DH | Galβ-3GlcNAcβ-3
35	iLNO Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-6 | Galβ-4Glc-DH | Galβ-3GlcNAcβ-3
36	I-octaosylceramide Galβ-4GlcNAcβ-6 | Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer | Galβ-4GlcNAcβ-3
37	B-like decaosylceramide Galα-3Galβ-4GlcNAcβ-6 | Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer | Galα-3Galβ-4GlcNAcβ-3
38	LND Galβ-4GlcNAcβ-6 | Galβ-4GlcNAcβ-6 | Galβ-3GlcNAcβ-3 Galβ-4Glc-DH | Galβ-3GlcNAcβ-3
39	B-like pentadecaosylceramide Galα-3Galβ-4GlcNAcβ-6 | Galα-3Galβ-4GlcNAcβ-6 Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer | | Galβ-4GlcNAcβ-3 | Galα-3Galβ-4GlcNAcβ-3
40	B-like eicosaosylceramide Galα-3Galβ-4GlcNAcβ-6 | Galα-3Galβ-4GlcNAcβ-6 Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer | | Galα-3Galβ-4GlcNAcβ-6 Galβ-4GlcNAcβ-3 | | Galβ-4GlcNAcβ-3 | Galα-3Galβ-4GlcNAcβ-3
41	pLNFH-IV Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH | Fucα-3
42	DFpLNH-II Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH | | Fucα-4 Fucα-3

43	MFLNH-III 
44	MFLNnH(a) 
45	DFLNH(c) 
46	DFLNnH 
47	TFLNH 
48	DFLNH(a) 
49	MFiLNO-IV 
50	TFiLNO 

51	MFLND Galβ-4GlcNAcβ-6 Fucα-3 Galβ-4GlcNAcβ-6 Galβ-3GlcNAcβ-3 Galβ-4Glc-DH Galβ-3GlcNAcβ-3
52	MSLNnH-I Galβ-4GlcNAcβ-6 Galβ-4Glc-DH NeuAcα-6Galβ-3GlcNAcβ-3
53	MFMSLNnH Galβ-4GlcNAcβ-6 Fucα-3 Galβ-4Glc-DH NeuAcα-6Galβ-3GlcNAcβ-3
54	MSMFLNH Galβ-4GlcNAcβ-6 Fucα-3 Galβ-4Glc-DH NeuAcα-3Galβ-3GlcNAcβ-3
55	DSLNnH NeuAcα-6Galβ-4GlcNAcβ-6 Galβ-4Glc-DH NeuAcα-6Galβ-4GlcNAcβ-3
56	GSC-219 SU-3GlcAβ-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36
57	GSC-216 GlcAβ-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer42
58	GSC-218 GlcAβ-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36
59	MSLNH NeuAcα-6Galβ-4GlcNAcβ-6 Galβ-4Glc-DH Galβ-3GlcNAcβ-3
60	C4U NeuAcα-3Galβ-4GlcNAcβ-3Galβ-3GlcNAc-DH SU-6 SU-6 SU-6
61	TFpLNH-I Fucα-2Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH Fucα-4 Fucα-3

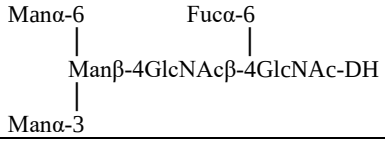
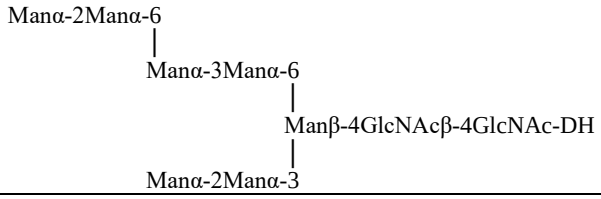
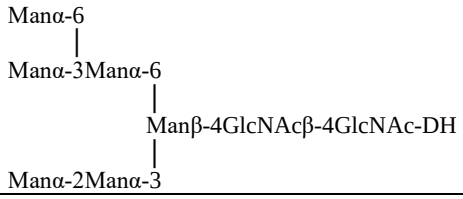
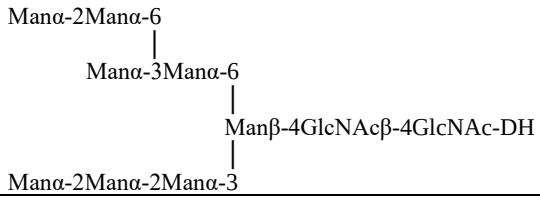
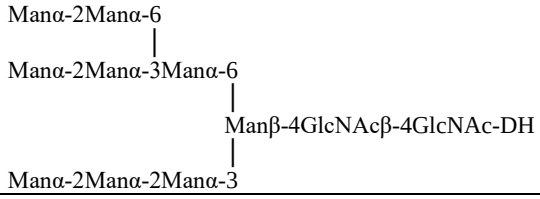
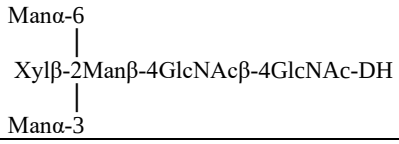
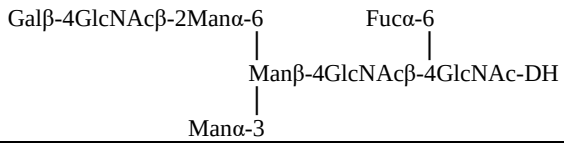
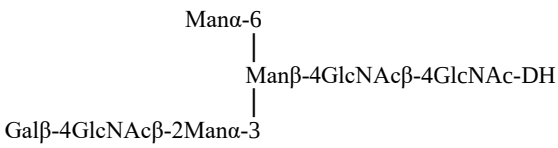
62	DFLNH(b) 
63	Nonaosylceramide 
64	I-dodecaosylceramide 
Set 46	
1	I-hexadecaosylceramide 
2	I-eicosaosylceramide 
3	B-like pentaeicosaosylceramide 

4	B-III dodecaosylceramide Galα-3Galβ-4GlcNAcβ-6 | Fucα-2 Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer Galα-3Galβ-4GlcNAcβ-3 | Fucα-2
5	B-IV tetradecaosylceramide Galα-3Galβ-4GlcNAcβ-6 | Fucα-2 Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer Galα-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3 | Fucα-2
6	GSC-217 SU-3GlcAβ-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer42
7	GSC-220 NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36 | Fucα-3 Fucα-3
8	GSC-221 NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36 | Fucα-3
9	LNnO Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH
10	Orsay-5-AO GlcNAcβ-6 | Gal-AO | Galβ-3GlcNAcβ-3
11	Orsay-6-AO Galβ-4GlcNAcβ-6 | Gal-AO | Galβ-3GlcNAcβ-3
12	Orsay-7-AO Galβ-4GlcNAcβ-6 | Gal-AO | Galβ-4GlcNAcβ-3
13	H2 (with H2+Fuc)* Fucα-2Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer
14	Ab (with Ab+Fuc)* GalNAcα-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer | Fucα-2

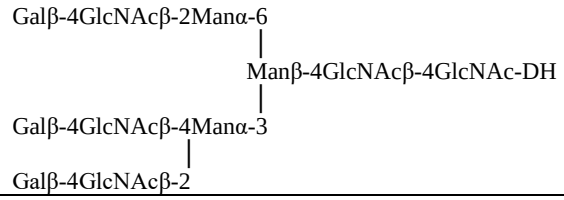
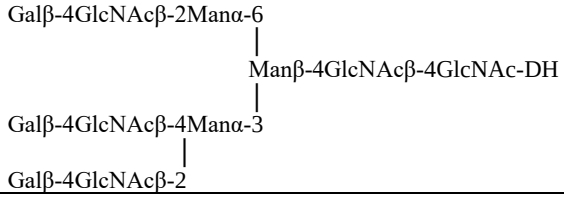
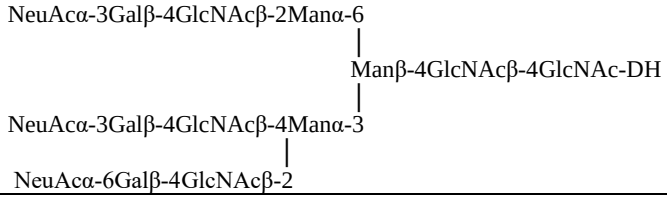
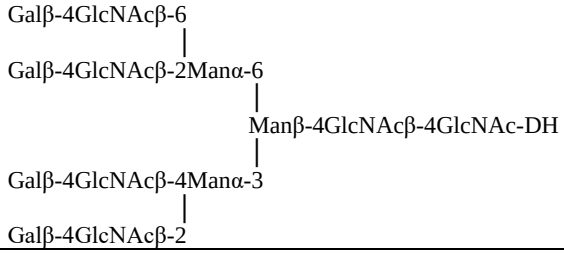
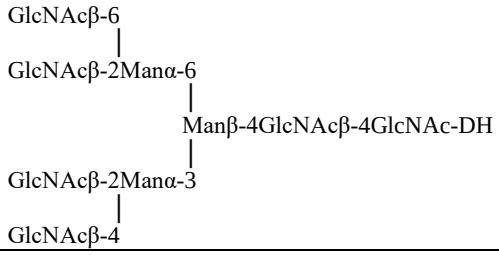
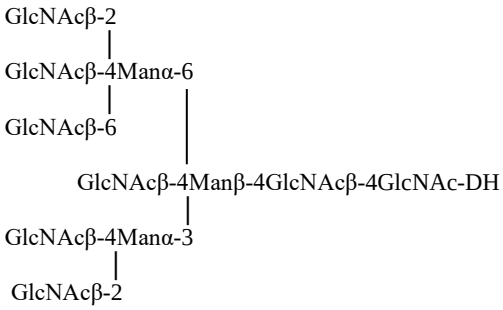
15	H3 (with H3-Fuc)* Fucα-2Galβ-4GlcNAcβ-6 Galβ-4GlcNAcβ-3Galβ-4Glc-Cer Fucα-2Galβ-4GlcNAcβ-3
16	Ad (with Ad+Fuc)* GalNAcα-3Galβ-4GlcNAcβ-6 Fucα-2 Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer GalNAcα-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3 Fucα-2
17	PSM-F1H2HN3-OY Fucα-2Galβ-4GlcNAcβ-6 Galβ-4GlcNAcβ-OY GlcNAcβ-3
18	LNnFP-I Fucα-2Galβ-4GlcNAcβ-3Galβ-4Glc-DH
19	A-Hexa-T1 GalNAcα-3Galβ-3GlcNAcβ-3Galβ-4Glc-DH Fucα-2
20	A-Hexa-T2 GalNAcα-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH Fucα-2
21	B-Hexa-T1 Galα-3Galβ-3GlcNAcβ-3Galβ-4Glc-DH Fucα-2
22	B-Hexa-T2 Galα-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH Fucα-2
23	MFLNH-I Galβ-4GlcNAcβ-6 Galβ-4Glc-DH Fucα-2Galβ-3GlcNAcβ-3
24	MSDFLNnH-AO Fucα-2Galβ-4GlcNAcβ-6 Fucα-3 Galβ-4Glc-AO NeuAcα-6Galβ-4GlcNAcβ-3
25	DFiLNO Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-6 Fucα-3 Galβ-4Glc-DH Fucα-2Galβ-3GlcNAcβ-3

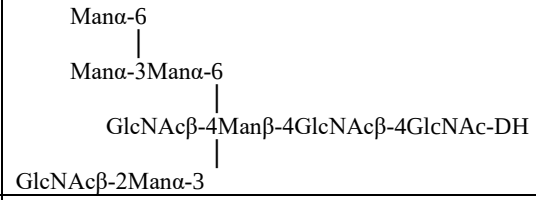
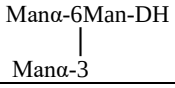
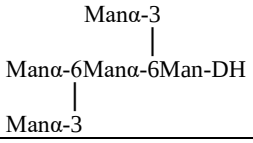
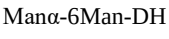
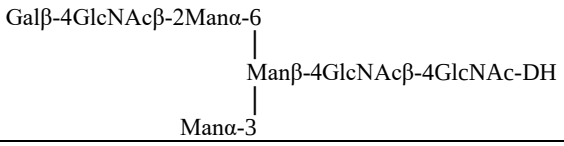
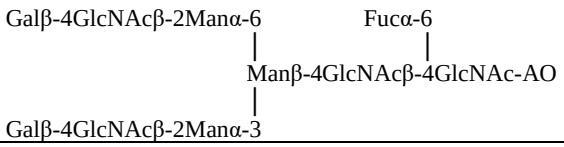
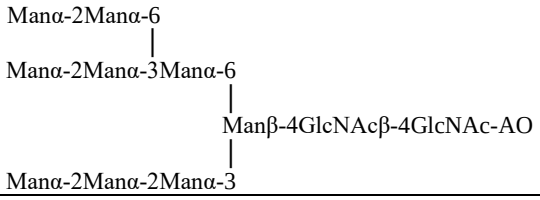
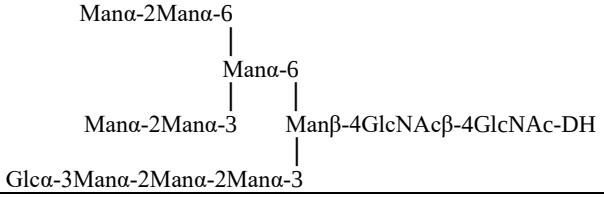
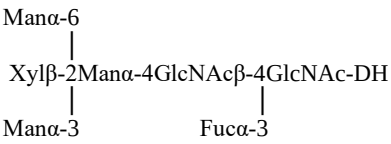
26	TFiLNO(1-2,2,3) Fucα-2Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-6 Fucα-3 Galβ-4Glc-DH Fucα-2Galβ-3GlcNAcβ-3
27	GSC-915 Fucα-2Galβ-4GlcNAcβ-6 Galβ-4GlcNAcβ-3Galβ-4Glc-DH Galβ-4GlcNAcβ-3
28	GSC-915-2 Fucα-2Galβ-4GlcNAcβ-6 Galβ-4GlcNAcβ-3Galβ-4Glc-DH GlcNAcβ-3
29	GSC-915-3 Fucα-2Galβ-4GlcNAcβ-6Galβ-4GlcNAcβ-3Galβ-4Glc-DH
30	GSC-915-4 (new) Galβ-4GlcNAcβ-6Galβ-4GlcNAcβ-3Galβ-4Glc-DH
31	LNFP-III Galβ-4GlcNAcβ-3Galβ-4Glc-DH Fucα-3
32	GSC-915-5 GlcNAcβ-6Galβ-4GlcNAcβ-3Galβ-4Glc-DH
33	NeuAcα-(3')LNnO NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH
34	NeuAcα-(6')LNnO (F1) NeuAcα-6Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH
35	NeuAcα-(6')LNnO (F2) NeuAcα-6Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH
36	pHGGs
37	GSC-915-AO Fucα-2Galβ-4GlcNAcβ-6 Galβ-4GlcNAcβ-3Galβ-4Glc-AO Galβ-4GlcNAcβ-3
38	TFiLNO(1-2,2,3)-AD Fucα-2Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-6 Fucα-3 Galβ-4Glc-AD Fucα-2Galβ-3GlcNAcβ-3

39	DFiLNO-AD Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-6 Fucα-3 Galβ-4Glc-AD Fucα-2Galβ-3GlcNAcβ-3
40	MFiLNO-IV-AD Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-6 Fucα-3 Galβ-4Glc-AD Galβ-3GlcNAcβ-3
41	iLNO Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-6 Galβ-4Glc-AD Galβ-3GlcNAcβ-3
42	pLNH-b Galβ-3GlcNAcβ-3Galβ-3GlcNAcβ-3Galβ-4Glc-DH
43	Man2aGN2 Manα-6Manβ-4GlcNAcβ-4GlcNAc-DH
44	Man3GN2 Manα-6 Manβ-4GlcNAcβ-4GlcNAc-DH Manα-3
45	Man4aGN2 Manα-3Manα-6 Manβ-4GlcNAcβ-4GlcNAc-DH Manα-3
46	Man4bGN2 Manα-6 Manα-3Manα-6 Manβ-4GlcNAcβ-4GlcNAc-DH
47	Man5GN2 Manα-6 Manα-3Manα-6 Manβ-4GlcNAcβ-4GlcNAc-DH Manα-3
48	Man7(D1)GN2 Manα-6 Manα-3Manα-6 Manβ-4GlcNAcβ-4GlcNAc-DH Manα-2Manα-2Manα-3

49	Man3FGN2 
50	Man7(D3)GN2 
51	Man6GN2 
52	Man8(D1D3)GN2 
53	Man9GN2 
54	Man3XylGN2 
55	N1 
56	N2 

57	NGA2 GlcNAcβ-2Manα-6 Manβ-4GlcNAcβ-4GlcNAc-DH GlcNAcβ-2Manα-3
58	NGA2B GlcNAcβ-2Manα-6 GlcNAcβ-4Manβ-4GlcNAcβ-4GlcNAc-DH GlcNAcβ-2Manα-3
59	NA2 Galβ-4GlcNAcβ-2Manα-6 Manβ-4GlcNAcβ-4GlcNAc-DH Galβ-4GlcNAcβ-2Manα-3
60	NGA2F GlcNAcβ-2Manα-6 Fucα-6 Manβ-4GlcNAcβ-4GlcNAc-DH GlcNAcβ-2Manα-3
61	NA2F Galβ-4GlcNAcβ-2Manα-6 Fucα-6 Manβ-4GlcNAcβ-4GlcNAc-DH Galβ-4GlcNAcβ-2Manα-3
62	NA2FB Galβ-4GlcNAcβ-2Manα-6 Fucα-6 GlcNAcβ-4Manβ-4GlcNAcβ-4GlcNAc-DH Galβ-4GlcNAcβ-2Manα-3
63	A2(2-6) NeuAcα-6Galβ-4GlcNAcβ-2Manα-6 Manβ-4GlcNAcβ-4GlcNAc-DH NeuAcα-6Galβ-4GlcNAcβ-2Manα-3
64	A2F(2-3) NeuAcα-3Galβ-4GlcNAcβ-2Manα-6 Fucα-6 Manβ-4GlcNAcβ-4GlcNAc-DH NeuAcα-3Galβ-4GlcNAcβ-2Manα-3
Set 47	
1	NGA3B GlcNAcβ-2Manα-6 GlcNAcβ-4Manβ-4GlcNAcβ-4GlcNAc-DH GlcNAcβ-4Manα-3 GlcNAcβ-2

2	NA3-Lex 
3	NA3 
4	A3 
5	NA4 
6	NGA4 
7	NGA5B 

8	GNMan5BGN2 
9	Man3(α3,α6) 
10	Man5(α3,α6) 
11	Man2(α6) 
12	N4 
13	NA2F-AO 
14	Man9GN2-AO 
15	Glc1Man9GN2 
16	Man3FXylGN2 

17	AGP-Bi-Ac2 NeuAcα-6Galβ-4GlcNAcβ-2Manα-6 | Manβ-4GlcNAcβ-4GlcNAc-DH | NeuAcα-6Galβ-4GlcNAcβ-2Manα-3
18	AGP-Bi-AcGc NeuGc(Ac)α-6Galβ-4GlcNAcβ-2Manα-6 | Manβ-4GlcNAcβ-4GlcNAc-DH | NeuAc(Gc)α-6Galβ-4GlcNAcβ-2Manα-3
19	AGP-Bi-Gc2 NeuGcα-6Galβ-4GlcNAcβ-2Manα-6 | Manβ-4GlcNAcβ-4GlcNAc-DH | NeuGcα-6Galβ-4GlcNAcβ-2Manα-3
20	Man1GN1 Manβ-4GlcNAc-DH
21	N3 GlcNAcβ-2Manα-6 Fucα-6 | | Manβ-4GlcNAcβ-4GlcNAc-DH | GlcNAcβ-2Manα-3
22	Man2GN1 Manα-3Manβ-4GlcNAc-DH
23	Fuc-GlcNAc Fucα-6GlcNAc-DH
24	Man2(α3) Manα-3Man-DH
25	Man2(α2) Manα-2Man-DH
26	Man7(D1)GN2-AO Manα-6 | Manα-3Manα-6 | Manβ-4GlcNAcβ-4GlcNAc-AO | Manα-2Manα-2Manα-3
27	Glc1Man9GN2-AO Manα-2Manα-6 | Manα-6 | | Manα-2Manα-3 Manβ-4GlcNAcβ-4GlcNAc-AO | Glcα-3Manα-2Manα-2Manα-3

28	Glc2Man7(D1)GN1-AO Manα-6 Manα-3Manα-6 Manβ-4GlcNAc-AO Glcα-3Glcα-3Manα-2Manα-2Manα-3
29	Glc3Man7(D1)GN1-AO Manα-6 Manα-3Manα-6 Manβ-4GlcNAc-AO Glcα-2Glcα-3Glcα-3Manα-2Manα-2Manα-3
30	GlcNac2Man3-AO GlcNAcβ-2Manα-6 Man-AO GlcNAcβ-2Manα-3
31	LNFP-III Galβ-4GlcNAcβ-3Galβ-4Glc-DH Fucα-3
32	Glc2Man-AO Glcα-3Glcα-3Man-AO
33	Glc2Man2-AO Glcα-3Glcα-3Manα-2Man-AO
34	Glc2Man3-AO Glcα-3Glcα-3Manα-2Manα-2Man-AO
35	P6-1 (GTP 4N(2,3)-4A+F) NeuAcα-3Galβ-4GlcNAcβ-6 NeuAcα-3Galβ-4GlcNAcβ-2Manα-6 Fucα-6 Manβ-4GlcNAcβ-4GlcNAc-DH NeuAcα-3Galβ-4GlcNAcβ-2Manα-3 NeuAcα-3Galβ-4GlcNAcβ-4
36	P7-2 (GTP 4N(2,3)-4A+1R+F) NeuAcα-3Galβ-4GlcNAcβ-6 NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-2Manα-6 Fucα-6 Manβ-4GlcNAcβ-4GlcNAc-DH NeuAcα-3Galβ-4GlcNAcβ-2Manα-3 NeuAcα-3Galβ-4GlcNAcβ-4

37	P8-1 (GTP 4N(2,3)-4A+2R+F) NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-6 NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-2Manα-6 Fucα-6 Manβ-4GlcNAcβ-4GlcNAc-DH NeuAcα-3Galβ-4GlcNAcβ-2Manα-3 NeuAcα-3Galβ-4GlcNAcβ-4
38	P22-1 (GTP 3N(2,3)-3A(2,6)+F) NeuAcα-3Galβ-4GlcNAcβ-6 NeuAcα-3Galβ-4GlcNAcβ-2Manα-6 Fucα-6 Manβ-4GlcNAcβ-4GlcNAc-DH NeuAcα-3Galβ-4GlcNAcβ-2Manα-3
39	(6P)Man5GN2.AO P-6Manα-6 Manα-6 Manα-3 Manβ-4GlcNAcβ-4GlcNAc-AO Manα-3 P-6Manα-6 Manα-6 Manα-3 Manβ-4GlcNAcβ-4GlcNAc-AO Manα-2Manα-3
40	(6P)Man6GN2.AO P-6Manα-6 Manα-6 Manα-3 Manβ-4GlcNAcβ-4GlcNAc-AO Manα-2Manα-3
41	Asialo-GM2 GalNAcβ-4Galβ-4Glcβ-Cer
42	Asialo-GM1-Tetra Galβ-3GalNAcβ-4Galβ-4Glc-DH
43	Asialo-GM1 Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer
44	GM4 NeuAcα-3Galβ-Cer
45	GM3 NeuAcα-3Galβ-4Glcβ-Cer
46	GM3(Gc) NeuGcα-3Galβ-4Glcβ-Cer

47	GM2 GalNAcβ-4Galβ-4Glcβ-Cer | NeuAcα-3
48	GM1b NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer
49	GM1(Gc) Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer | NeuGcα-3
50	GM1 Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer | NeuAcα-3
51	GD3 NeuAcα-8NeuAcα-3Galβ-4Glcβ-Cer
52	GD2 GalNAcβ-4Galβ-4Glcβ-Cer | NeuAcα-8NeuAcα-3
53	GD1a NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer | NeuAcα-3
54	GD1b Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer | NeuAcα-8NeuAcα-3
55	GalNAc-GD1a(Ac,Gc) GalNAcβ-4Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer | | NeuGcα-3 NeuAcα-3 GalNAcβ-4Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer | | NeuAcα-3 NeuGcα-3
56	GT1a NeuAcα-8NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer | NeuAcα-3
57	GT1b NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer | NeuAcα-8NeuAcα-3
58	GQ1b NeuAcα-8NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer | NeuAcα-8NeuAcα-3
59	SM3 SU-3Galβ-4Glcβ-Cer
60	Haematoside NeuAcα-3Galβ-4Glcβ-Cer

61	SM1a Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer SU-3
62	SB2 SU-3GalNAcβ-4Galβ-4Glcβ-Cer SU-3
63	SB1a SU-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer SU-3
64	GD1a-hexa NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glc-DH NeuAcα-3
Set 48	
1	GM1-penta Galβ-3GalNAcβ-4Galβ-4Glc-DH NeuAcα-3
2	GM1(Gc)-penta Galβ-3GalNAcβ-4Galβ-4Glc-DH NeuGcα-3
3	GD3-tetra NeuAcα-8NeuAcα-3Galβ-4Glc-DH
4	GD3-tetra-AO NeuAcα-8NeuAcα-3Galβ-4Glc-AO
5	GSC-68 NeuAcα-6Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer36
6	GSC-442 GalNAcβ-4Galβ-4Glcβ-Cer36 NeuAcα-6
7	GSC-155 Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer36 NeuAcα-6
8	GSC-118 NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer36 NeuAcα-6
9	GSC-107 NeuAcα-6Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer36 NeuAcα-6
10	GSC-193 GalNAcβ-4Galβ-4Glcβ-Cer36 KDNα-3

11	GSC-195 KDNα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer36 | KDNα-3
12	GSC-335 SU-6 | NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer36
13	GSC-576 GalNAcβ-4Galβ-3Glcβ-C30 | NeuAcα-3
14	GSC-108 GalNAcβ-4Galβ-4Glcβ-Cer36 | NeuAcα-3
15	GD1b Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer | NeuAcα-8NeuAcα-3
16	GM1b-DH NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glc-DH
17	GT1c-DH Galβ-3GalNAcβ-4Galβ-4Glc-DH | NeuAcα-8NeuAcα-8NeuAcα-3
18	Galβ-3GalNAc Galβ-3GalNAc-DH
19	Galβ-6GalNAc Galβ-6GalNAc-DH
20	GalNAcα-3GalNAc GalNAcα-3GalNAc-DH
21	A8/1 GlcNAcα-4Galβ-OX
22	A15/3 GlcNAcα-4Galβ-3Galβ-OX | Fucα-2
23	DST NeuAcα-3Galβ-3GalNAc-DH | NeuAcα-6
24	A15/1 SU-6GlcNAcβ-OY
25	A8/2 SU-6 | Fucα-3GlcNAcβ-OY

26	B13/a GlcAβ-3Galβ-3GlcNAcβ-OX
27	Man-Ser Manα-Ser-DH
28	Notch-1 Fucα-Thr-DH
29	Notch-2 GlcNAcβ-3Fucα-Thr
30	Notch-3 Galβ-4GlcNAcβ-3Fucα-Thr-DH
31	LNFP-III Galβ-4GlcNAcβ-3Galβ-4Glc-DH | Fucα-3
32	Man-Thr Manα-Thr-DH
33	GalNAc-Thr GalNAcα-Thr-DH
34	GalNAc-Ser GalNAcα-Ser-DH
35	Man-Thr-Succ Manα-Thr-Succ-DH
36	Man-Ser-Succ Manα-Ser-Succ-DH
37	Galβ-3GalNAc-AO Galβ-3GalNAc-AO
38	Galβ-6GalNAc-AO Galβ-6GalNAc-AO
39	GlcNAcβ1-2Fuc-AO GlcNAcβ-2Fuc-AO
40	GlcNAcβ1-4Fuc-AO GlcNAcβ-4Fuc-AO
41	GSC-488 NeuAcα-3Galβ-3GalNAcβ-C30
42	GSC-489 SU-6 | NeuAcα-3Galβ-3GalNAcβ-C30
43	GSC-490 NeuAcα-3Galβ-3GalNAcβ-C30 | NeuAcα-6

44	GSC-491 NeuAcα-3Galβ-3(6-deoxy-6-carboxymethyl)GalNAcβ-C30
45	BSM-Di-A1-AO NeuGcα-6GalNAc-AO
46	BSM-Di-A2-AO NeuAcα-6GalNAc-AO
47	DST-AO NeuAcα-3Galβ-3GalNAc-AO | NeuAcα-6
48	GlcNAcβ-3Fuc-AO GlcNAcβ-3Fuc-AO
49	GlcNAcβ-2Man-AO GlcNAcβ-2Man-AO
50	XylGlc-AO Xylα-3Glc-AO
51	Xyl2Glc-AO Xylα-3Xylα-3Glc-AO
52	SA2(α8) NeuAcα-8NeuAc-DH
53	SA3(α8) NeuAcα-8NeuAcα-8NeuAc-DH
54	SA4(α8) NeuAcα-8NeuAcα-8NeuAcα-8NeuAc-DH
55	SA5(α8)* NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAc-DH
56	SA6(α8)* NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAc-DH
57	SA7(α8)* NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAc-DH
58	SA8(α8)* NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAc-8NeuAcα-8NeuAc-DH
59	SA9(α8)* NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAc-8NeuAcα-8NeuAc-8NeuAcα-DH
60	SA10(α8)* NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAc-DH
61	Gal4 Galα-3Galβ-4Galα-3Gal-DH
62	Glc4(α6,α4,α4) Glcα-6Glcα-4Glcα-4Glc-DH

63	Xyl3Glc4 Xylα-6 Glcβ-4Glcβ-4Glcβ-4Glc-DH Xylα-6 Xylα-6
64	Ceramides Cer
Set 49	
1	Glc Glc-DH
2	Glc-AO Glc-AO
3	Gal Gal-DH
4	Gal-AO Gal-AO
5	Man Man-DH
6	Man-AO Man-AO
7	Fuc Fuc-DH
8	Fuc-AO Fuc-AO
9	Rha Rha-DH
10	Rha-AO Rha-AO
11	GN GlcNAc-DH
12	GN-AO GlcNAc-AO
13	GalNAc GalNAc-DH
14	GalNAc-AO GalNAc-AO

15	GalNAc-AO GalNAc-AO
16	SU-Tyr SU-Tyr-DH
17	NeuAc NeuAc-DH
18	NeuAc-AO NeuAc-AO
19	(6P)-Man P-6Man-DH
20	(6P)-Man-AO P-6Man-AO
21	SU-Cholesterol SU-Cholesterol
22	GSC-154 NeuAcα-3Galβ-4GlcNAcβ-6Galβ-4Glcβ-Cer36 | Fucα-3
23	GSC-441 NeuAcα-3Galβ-4GlcNAcβ-6GalNAcα-3Galβ-4Glcβ-C30
24	GSC-70 NeuAcα-6Galβ-6GalNAcβ-4Galβ-4Glcβ-Cer36
25	NeuGc NeuGc-DH
26	NeuGc-AO NeuGc-AO
27	Galα-6Glc-AO Galα-6Glc-AO
28	(6P)-Glc-AO P-6Glc-AO
29	(6P)-Fructose-AO P-6Fru-AO
30	GN-Asn GlcNAcβ-Asn-DH
31	LNFP-III Galβ-4GlcNAcβ-3Galβ-4Glc-DH | Fucα-3

32	GSC-284 GalNAcβ-6Galβ-4Glcβ-Cer36 NeuAcα-3
33	GSC-384 NeuAcα-3Galβ-4GlcNAcβ-4GalNAcβ-3Galβ-4Glcβ-C30 Fucα-3
34	GSC-446 NeuAcα-3Galβ-4GlcNAcβ-6GalNAcα-3Galβ-4Glc-C30
35	GSC-575 GalNAcβ-4Galβ-3Galβ-C30 NeuAcα-3
36	Glc(α6,α4,α4)-AO Glcα-6Glcα-4Glcα-4Glc-AO
37	O1-AO GlcNAcβ-6 Gal-AO GlcNAcβ-3
38	Rutinose-AO Rhaα-6Glc-AO
39	Hep-Di IS ΔUA-4GlcNS-DH SU-2 SU-6
40	Hep-Di-IS-AO ΔUA-4GlcNS-AO SU-2 SU-6
41	HS-S4-AO* GlcAβ-4GlcNAcα-4GlcAβ-4aMan-AO
42	HS-S6-AO* GlcAβ-4GlcNAcα-4GlcAβ-4GlcNAcα-4GlcAβ-4aMan-AO
43	HS-S8-AO* GlcAβ-4GlcNAcα-4GlcAβ-4GlcNAcα-4GlcAβ-4GlcNAcα-4GlcAβ
44	Hep-S4-AO* IdoAα-4GlcNSα-4IdoAα-4aMan-AO SU-2 SU-6 SU-2 SU-6
45	Hep-S6-AO* IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4aMan-AO SU-2 SU-6 SU-2 SU-6 SU-2 SU-6
46	Hep-S8-AO* IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4aMan-AO SU-2 SU-6 SU-2 SU-6 SU-2 SU-6 SU-2 SU-6

47	HS-S4a-AO
48	HS-S6a-AO
49	HS-S8a-AO
50	Hep-S8a
51	Hep-4-AO* SU-2 ΔUA-4GlcNSα-4IdoAα-4GlcNS-AO SU-6 SU-2 SU-6
52	Hep-6-AO* SU-2 ΔUA-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNS-AO SU-6 SU-2 SU-6 SU-2 SU-6
53	Hep-8-AO* SU-2 ΔUA-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNS-AO SU-6 SU-2 SU-6 SU-2 SU-6 SU-2 SU-6
54	Hep-10-AO* SU-2 ΔUA-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNS-AO SU-6 SU-2 SU-6 SU-2 SU-6 SU-2 SU-6 SU-2 SU-6
55	Hep-12-AO* SU-2 ΔUA-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNS-AO SU-6 SU-2 SU-6 SU-2 SU-6 SU-2 SU-6 SU-2 SU-6 SU-2 SU-6
56	Hep-14-AO* SU-2 ΔUA-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNS-AO SU-6 SU-2 SU-6 SU-2 SU-6 SU-2 SU-6 SU-2 SU-6 SU-2 SU-6 SU-2 SU-6
57	Hep-16-AO* SU-2 ΔUA-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNS-AO SU-6 SU-2 SU-6 SU-2 SU-6 SU-2 SU-6 SU-2 SU-6 SU-2 SU-6 SU-2 SU-6 SU-2 SU-6
58	CSA-3* GalNAcβ-4GlcAβ-3GalNAc-DH SU-4 SU-4
59	CSA-5* GalNAcβ-4GlcAα-3GalNAcβ-4GlcAα-3GalNAc-DH SU-4 SU-4 SU-4

60	CSB-3* GalNAcβ-4IdoAα-3GalNAc-DH | | SU-4 SU-4
61	CSB-5* GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAc-DH | | | SU-4 SU-4 SU-4
62	CSC-3* GalNAcβ-4GlcAβ-3GalNAc-DH | | SU-6 SU-6
63	CSC-5* GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-DH | | | SU-6 SU-6 SU-6
64	HA-12* ΔUA-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAc-DH
Set 50	
1	HA-S4* GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAc-DH
2	HA-S6* GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAc-DH
3	HA-S8* GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAc-DH
4	HA-S10* GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAc-DH
5	HA-S12* GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ3GlcNAc-DH
6	HA-S14* GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAc-DH
7	HA-S16* GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAc-DH
8	HA-S18* GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAc-DH
9	CSA-2* ΔUA-3GalNAc-DH | SU-4
10	CSA-4* ΔUA-3GalNAcβ-4GlcAβ-3GalNAc-DH | | SU-4 SU-4

11	CSA-6* ΔUA-3GalNAcβ-4GlcAα-3GalNAcβ-4GlcAα-3GalNAc-DH SU-4 SU-4 SU-4
12	CSA-8* ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-DH SU-4 SU-4 SU-4 SU-4
13	CSA-10* ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-DH SU-4 SU-4 SU-4 SU-4 SU-4
14	CSA-12* ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ3GalNAc-DH SU-4 SU-4 SU-4 SU-4 SU-4 SU-4
15	CSA-14* ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-DH SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4
16	CSA-16* ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-DH SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4
17	CSA-18* ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-DH SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4
18	CSA-20* ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-DH SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4
19	CSB-2* ΔUA-3GalNAc-DH SU-4
20	CSB-4* ΔUA-3GalNAcβ-4IdoAα-3GalNAc-DH SU-4 SU-4
21	CSB-6* ΔUA-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAc-DH SU-4 SU-4 SU-4
22	CSB-8* ΔUA-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAc-DH SU-4 SU-4 SU-4 SU-4
23	CSB-10 ΔUA-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAc-DH SU-4 SU-4 SU-4 SU-4 SU-4
24	CSB-12* ΔUA-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAc-DH SU-4 SU-4 SU-4 SU-4 SU-4 SU-4
25	CSB-14* ΔUA-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAc-DH SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4

26	CSB-16* ΔUA-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAc-DH SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4
27	CSB-18* ΔUA-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAc-DH SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4
28	CSB-20* ΔUA-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAc-DH SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4
29	CSC-2* ΔUA-3GalNAc-DH SU-6
30	CSC-4* ΔUA-3GalNAcβ-4GlcAβ-3GalNAc-DH SU-6 SU-6
31	LNFP-III Galβ-4GlcNAcβ-3Galβ-4Glc-DH Fucα-3
32	CSC-6* ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-DH SU-6 SU-6 SU-6
33	CSC-8* ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-DH SU-6 SU-6 SU-6 SU-6
34	CSC-10* ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-DH SU-6 SU-6 SU-6 SU-6 SU-6
35	CSC-12* ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ3GalNAc-DH SU-6 SU-6 SU-6 SU-6 SU-6 SU-6
36	CSC-14* ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-DH SU-6 SU-6 SU-6 SU-6 SU-6 SU-6 SU-6
37	CSC-16* ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-DH SU-6 SU-6 SU-6 SU-6 SU-6 SU-6 SU-6 SU-6
38	CSC-18* ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-DH SU-6 SU-6 SU-6 SU-6 SU-6 SU-6 SU-6 SU-6 SU-6
39	CSC-20* ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-DH SU-6 SU-6 SU-6 SU-6 SU-6 SU6 SU-6 SU-6 SU-6 SU-6
40	CSC-S4* GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-DH SU-6 SU-6

41	CSC-S6* GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-DH SU-6 SU-6 SU-6
42	CSC-S8* GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-DH SU-6 SU-6 SU-6 SU-6
43	CSC-S10* GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-DH SU-6 SU-6 SU-6 SU-6 SU-6
44	CSC-S12* GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ3GalNAc-DH SU-6 SU-6 SU-6 SU-6 SU-6 SU-6
45	CSC-S14* GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-DH SU-6 SU-6 SU-6 SU-6 SU-6 SU-6 SU-6
46	CSA-2-AO* (Nian) ΔUA-3GalNAc-AO SU-4
47	CSA-4-AO* (Nian) ΔUA-3GalNAcβ-4GlcAβ-3GalNAc-AO SU-4 SU-4
48	CSA-6-AO* (Nian) ΔUA-3GalNAcβ-4GlcAα-3GalNAcβ-4GlcAα-3GalNAc-AO SU-4 SU-4 SU-4
49	CSA-8-AO* (Nian) ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-AO SU-4 SU-4 SU-4 SU-4
50	CSA-10-AO* (Nian) ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-AO SU-4 SU-4 SU-4 SU-4 SU-4
51	CSA-12-AO* (Nian) ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-AO SU-4 SU-4 SU-4 SU-4 SU-4 SU-4
52	CSA-14-AO* (Nian) ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-AO SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4
53	CSA-16-AO* (Nian) ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-AO SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4
54	CSA-18-AO* (Nian) ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-AO SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4
55	CSA-20-AO* (Nian) ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-AO SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4

56	CSB-2-AO* (Nian) ΔUA-3GalNAc-AO SU-4
57	CSB-4-AO* (Nian) ΔUA-3GalNAcβ-4IdoAα-3GalNAc-AO SU-4 SU-4
58	CSB-6-AO* (Nian) ΔUA-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAc-AO SU-4 SU-4 SU-4
59	CSB-8-AO* (Nian) ΔUA-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAc-AO SU-4 SU-4 SU-4 SU-4
60	CSB-10-AO* (Nian) ΔUA-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAc-AO SU-4 SU-4 SU-4 SU-4 SU-4
61	CSB-12-AO* (Nian) ΔUA-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα3GalNAc-AO SU-4 SU-4 SU-4 SU-4 SU-4 SU-4
62	CSB-14-AO* (Nian) ΔUA-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAc-DH SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4
63	CSB-16-AO* (Nian) ΔUA-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAc-AO SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4
64	CSB-18-AO* (Nian) ΔUA-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAc-AO SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4
Set 51	
1	CSB-20-AO* (Nian) ΔUA-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAc-AO SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4 SU-4
2	CSC-2-AO* (Nian) ΔUA-3GalNAc-AO SU-6
3	CSC-4-AO* (Nian) ΔUA-3GalNAcβ-4GlcAβ-3GalNAc-AO SU-6 SU-6
4	CSC-6-AO* (Nian) ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-AO SU-6 SU-6 SU-6
5	CSC-8-AO* (Nian) ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-AO SU-6 SU-6 SU-6 SU-6

6	CSC-10-AO* (Nian) ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-AO SU-6 SU-6 SU-6 SU-6 SU-6
7	CSC-12-AO* (Nian) ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-AO SU-6 SU-6 SU-6 SU-6 SU-6 SU-6
8	CSC-14-AO* (Nian) ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-AO SU-6 SU-6 SU-6 SU-6 SU-6 SU-6 SU-6
9	CSC-16-AO* (Nian) ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-AO SU-6 SU-6 SU-6 SU-6 SU-6 SU6 SU-6 SU-6
10	CSC-18-AO* (Nian) ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-AO SU-6 SU-6 SU-6 SU-6 SU-6 SU-6 SU-6 SU-6 SU-6
11	CSC-20-AO* (Nian) ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-AO SU-6 SU-6 SU-6 SU-6 SU-6 SU-6 SU-6 SU-6 SU-6 SU-6
12	Hep-2-AO*(Nian)
13	Hep-4-AO*(Nian)
14	Hep-6-AO*(Nian)
15	Hep-8-AO*(Nian)
16	Hep-10-AO*(Nian)
17	Hep-12-AO*(Nian)
18	Hep-14-AO*(Nian)
19	Hep-16-AO*(Nian)
20	Hep-18-AO*(Nian)
21	Hep-20-AO*(Nian)
22	HS(deAc)-2-AO* (Nian
23	HS(deAc)-4-AO* (Nian)
24	HS(deAc)-6-AO* (Nian)
25	HS(deAc)-8-AO* (Nian)
26	HS(deAc)-10-AO* (Nian)
27	HS(deAc)-12-AO* (Nian)
28	KS-2-AO*[K'ase I, 23-27h]
29	KS-4-AO*[K'ase I, 23-27h]
30	KS-6-AO*[K'ase I, 23-27h]
31	LNFP-III Galβ-4GlcNAcβ-3Galβ-4Glc-DH Fucα-3
32	KS-8-AO*[K'ase I, 23-27h]

33	KS-10-AO*[K'ase I, 23-27h]
34	KS-12-AO*[K'ase I, 23-27h]
35	KS-14-AO*[K'ase I, 23-27h]
36	KS-16-AO*[K'ase I, 23-27h]
37	KS-18-AO*[K'ase I, 23-27h]
38	KS-20-AO*[K'ase I, 23-27h]
39	KS-2-AO*[K'ase I, 82h]
40	KS-4-AO*[K'ase I, 82h]
41	KS-6-AO*[K'ase I, 82h]
42	KS-8-AO*[K'ase I,82h]
43	KS-10-AO*[K'ase I,82h]
44	Carra-Tetra-1S aGalα-3Galβ-4aGalα-3Gal-DH | SU-4
45	Carra-Tetra-2S aGalα-3Galβ-4aGalα-3Gal-DH | | SU-4 SU-4
46	Carra-Hexa-3S aGalα-3Galβ-4aGalα-3Galβ-4aGalα-3Gal-DH | | | SU-4 SU-4 SU-4
47	Carra-Hexa-4S aGalα-3Galβ-4aGalα-3Galβ-4aGalα-3Gal-DH | | | | SU-4 SU-2 SU-4 SU-4
48	Carra-Octa-4S aGalα-3Galβ-4aGalα-3Galβ-4aGalα-3Galβ-4aGalα-3Gal-DH | | | | SU-4 SU-4 SU-4 SU-4
49	Alg-M4 ManAβ-4ManAβ-4ManAβ-4ManA-AO
50	Alg-M5 ManAβ-4ManAβ-4ManAβ-4ManAβ-4ManA-AO
51	Alg-G4 GulAα-4GulAα-4GulAα-4GulA-AO
52	Alg-G5 GulAα-4GulAα-4GulAα-4GulAα-4GulA-AO
53	K5(deAc) F2* GlcAβ-3GlcNα-4GlcAβ-3aMan-AO
54	K5(deAc) F3* GlcAβ-3GlcNAcα-4GlcAβ-3aMan-AO

55	K5(deAc) F4* GlcAβ-3GlcNAcα-4GlcAβ-3GlcNAcα-4GlcAβ-3aMan-AO
56	K5(deAc) F5* GlcAβ-3GlcNAcα-4GlcAβ-3GlcNAcα-4GlcAβ-3GlcNAcα-4GlcAβ-3aMan-AO
57	K5(deAc) F6* GlcAβ-3GlcNAcα-4GlcAβ-3GlcNAcα-4GlcAβ-3GlcNAcα-4GlcAβ-3GlcNAcα-4GlcAβ-3aMan-AO
58	K5(deAc) F7* GlcAβ-3GlcNAcα-4GlcAβ-3GlcNAcα-4GlcAβ-3GlcNAcα-4GlcAβ-3GlcNAcα-4GlcAβ-3GlcNAcα-4GlcAβ-3aMan-AO
59	Kojibiose Glcα-2Glc-DH
60	Nigerose Glcα-3Glc-DH
61	Malto-2 Glcα-4Glc-DH
62	Dext-2 Glcα-6Glc-DH
63	Lam-2 Glcβ-3Glc-DH
64	Cellobiose Glcβ-4Glc-DH
Set 52	
1	Cyano-2-AO Glcα-2Glc-AO
2	Cyano-3-AO Glcα-2Glcα-2Glc-AO
3	Cyano-4-AO Glcα-2Glcα-2Glcα-2Glc-AO
4	Cyano-5-AO Glcα-2Glcα-2Glcα-2Glcα-2Glc-AO
5	Cyano-6-AO* Glcα-2Glcα-2Glcα-2Glcα-2Glcα-2Glc-AO
6	Cyano-7-AO* Glcα-2Glcα-2Glcα-2Glcα-2Glcα-2Glcα-2Glc-AO
7	Cyano-8-AO* Glcα-2Glcα-2Glcα-2Glcα-2Glcα-2Glcα-2Glcα-2Glc-AO
8	Cyano-9-AO* Glcα-2Glcα-2Glcα-2Glcα-2Glcα-2Glcα-2Glcα-2Glcα-2Glc-AO
9	Nigerose-AO Glcα-3Glc-AO

10	Poria-3-AO Glcα-3Glcα-3Glc-AO
11	Poria-4-AO Glcα1-3Glcα1-3Glcα1-3Glc-AO
12	Poria-5-AO Glcα-3Glcα-3Glcα-3Glcα-3Glc-AO
13	Poria-6-AO Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glc-AO
14	Poria-7-AO Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glc-AO
15	Poria-8-AO* Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glc-AO
16	Poria-9-AO* Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glc-AO
17	Poria-10-AO* Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glα-3Glc-AO
18	Poria-11-AO* Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glα-3Glcα-3Glc-AO
19	Poria-12-AO* Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glα-3Glcα-3Glcα-3Glc-AO
20	Poria-13-AO* Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glα-3Glcα-3Glcα-3Glcα-3Glc-AO
21	Malto-2-AO Glcα-4Glc-AO
22	Malto-3-AO Glcα-4Glcα-4Glc-AO
23	Malto-4-AO Glcα-4Glcα-4Glcα-4Glc-AO
24	Malto-5-AO Glcα-4Glcα-4Glcα-4Glcα-4Glc-AO
25	Malto-6-AO Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glc-AO
26	Malto-7-AO Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glc-AO
27	Malto-8-AO* Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glc-AO
28	Malto-9-AO* Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glc-AO

29	Malto-10-AO* Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glc-AO
30	Malto-11-AO* Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glc-AO
31	LNFP-III Galβ-4GlcNAcβ-3Galβ-4Glc-DH | Fucα-3
32	Malto-12-AO* Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glc-AO
33	Malto-13-AO* Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glc-AO
34	Dext-2-AO Glcα-6Glc-AO
35	Dext-3-AO Glcα-6Glcα-6Glc-AO
36	Dext-4-AO Glcα-6Glcα-6Glcα-6Glc-AO
37	Dext-5-AO* Glcα-6Glcα-6Glcα-6Glcα-6Glc-AO
38	Dext-6-AO* Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glc-AO
39	Dext-7-AO Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glc-AO
40	Dext-8-AO* Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glc-AO
41	Dext-9-AO* Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glc-AO
42	Dext-10-AO* Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glc-AO
43	Dext-11-AO* Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glc-AO
44	Dext-12-AO* Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glc-AO
45	Dext-13-AO* Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glc-AO
46	Pano-3-AO Glcα-6Glcα-4Glc-AO
47	i-Pano-3-AO Glcα-4Glcα-6Glc-AO

48	Pullu-4-AO Glcα-6Glcα-4Glcα-4Glc-AO
49	Pullu-6-AO Glcα-4Glcα-4Glcα-6Glcα-4Glcα-4Glc-AO
50	Pullu-7-AO Glcα-6Glcα-4Glcα-4Glcα-6Glcα-4Glcα-4Glc-AO
51	CβG-2-AO Glcβ-2Glc-AO
52	CβG-3-AO Glcβ-2Glcβ-2Glc-AO
53	CβG-4-AO Glcβ-2Glcβ-2Glcβ-2Glc-AO
54	CβG-5-AO* Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glc-AO
55	CβG-6-AO* Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glc-AO
56	CβG-7-AO* Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glc-AO
57	CβG-8-AO* Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glc-AO
58	CβG-9-AO* Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glc-AO
59	CβG-10-AO* Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glc-AO
60	CβG-11-AO* Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glc-AO
61	CβG-12-AO* Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glc-AO
62	CβG-13-AO* Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glc-AO
63	Lam-2-AO Glcβ-3Glc-AO
64	Lam-3-AO Glcβ-3Glcβ-3Glc-AO
Set 53	
1	Lam-4-AO Glcβ-3Glcβ-3Glcβ-3Glc-AO
2	Lam-5-AO Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO

3	Lam-6-AO* Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO
4	Lam-7-AO Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO
5	Curd-8-AO* Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO
6	Curd-9-AO* Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO
7	Curd-10-AO* Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO
8	Curd-11-AO* Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO
9	Curd-12-AO* Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO
10	Curd-13-AO* Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO
11	HE-8-AO Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO
12	HE-9-AO Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO
13	HE-10-AO Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO
14	Cellobiose-AO Glcβ-4Glc-AO
15	Cello-3-AO Glcβ-4Glcβ-4Glc-AO
16	Cello-4-AO Glcβ-4Glcβ-4Glcβ-4Glc-AO
17	Cello-5-AO* Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glc-AO
18	Cello-6-AO* Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glc-AO
19	Cello-7-AO* Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glc-AO
20	Cello-8-AO* Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glc-AO
21	Cello-9-AO* Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glc-AO
22	Cello-10-AO*

	Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glc-AO
23	Cello-11-AO* Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glc-AO
24	Cello-12-AO* Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glc-AO
25	Cello-13-AO* Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glc-AO
26	Gentiobiose-AO Glcβ-6Glc-AO
27	Pust-3-AO Glcβ-6Glcβ-6Glc-AO
28	Pust-4-AO Glcβ-6Glcβ-6Glcβ-6Glc-AO
29	Pust-5-AO Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glc-AO
30	Pust-6-AO Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glc-AO
31	LNFP-III Galβ-4GlcNAcβ-3Galβ-4Glc-DH | Fucα-3
32	Pust-7-AO* Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glc-AO
33	Pust-8-AO* Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glc-AO
34	Pust-9-AO* Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glc-AO
35	Pust-10-AO* Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glc-AO
36	Pust-11-AO* Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glc-AO
37	Pust-15-AO* Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glc-AO
38	Pust-15a-AO* Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glc-AO
39	Barley-3-AO Glcβ-4Glcβ-3Glc-AO
40	Barley-3a-AO Glcβ-3Glcβ-4Glc-AO

41	Barley-3b-AO Glcβ-4Glcβ-3Glc-AO
42	Barley-4-AO Glcβ-4Glcβ-4Glcβ-3Glc-AO
43	Barley-4a-AO Glcβ-3Glcβ-4Glcβ-4Glc-AO
44	Barley-4b-AO Glcβ-4Glcβ-4Glcβ-3Glc-AO
45	Barley-4c-AO Glcβ-4Glcβ-3Glcβ-4Glc-AO
46	Barley-5-AO* Glcβ-4Glcβ-4Glcβ-4Glcβ-3Glc-AO
47	Barley-5a-AO Glcβ-3Glcβ-4Glcβ-6Glcβ-4Glc-AO
48	Barley-6-AO* (Glcβ-3/4Glcβ)2-4Glcβ-3Glc-AO
49	Barley-6a-AO Glcβ-3Glcβ-4Glcβ-6Glcβ-4Glcβ-3Glc-AO; and Glcβ-3Glcβ-4Glcβ-4Glcβ-4Glcβ-3Glc-AO (ratio ~1:1)
50	Barley-7-AO* Glcβ-4Glcβ-4Glcβ-3Glcβ-4Glcβ-4Glcβ-3Glc-AO
51	Barley-8-AO* (Glcβ-4/3Glcβ)3-4Glcβ-3Glc-AO
52	Barley-9-AO* (Glcβ-4/3Glcβ)4-3Glc-AO
53	Barley-10-AO* (Glcβ-4/3Glcβ)4-4Glcβ-3Glc-AO
54	Barley-11-AO* (Glcβ-4/3Glcβ)5-3Glc-AO
55	Barley-12-AO* (Glcβ-4/3Glcβ)5-4Glcβ-3Glc-AO
56	Barley-13-AO* (Glcβ-4/3Glcβ)6-3Glc-AO
57	Barley-14-AO* (Glcβ-4/3Glcβ)6-4Glcβ-3Glc-AO
58	Barley-15-AO* (Glcβ-4/3Glcβ)7-3Glc-AO
59	Barley-16-AO* (Glcβ-4/3Glcβ)7-4Glcβ-3Glc-AO

60	Grifo-3-AO* Glc3(β-3/β-6)-AO
61	Grifo-4-AO* Glc4(β-3/β-6)-AO
62	Grifo-5-AO* Glc5(β-3/β-6)-AO
63	Grifo-6-AO* Glc6(β-3/β-6)-AO
64	Grifo-7-AO* Glc7(β-3/β-6)-AO
Set 54	
1	Grifo-8-AO* Glc8(β-3/β-6)-AO
2	Grifo-9-AO* Glc9(β-3/β-6)-AO
3	Grifo-10-AO* Glc10(β-3/β-6)-AO
4	Grifo-11-AO* Glc11(β-3/β-6)-AO
5	Grifo-12-AO* Glc12(β-3/β-6)-AO
6	Grifo-13-AO* Glc13(β-3/β-6)-AO
7	Grifo-14-AO* Glc14(β-3/β-6)-AO
8	Grifo-15-AO* Glc15(β-3/β-6)-AO
9	Grifo-16-AO* Glc16(β-3/β-6)-AO
10	Lentin-2-AO Glc2(β-3/β-6)-AO
11	Lentin-3-AO Glc3(β-3/β-6)-AO
12	Lentin-4-AO* Glc4(β-3/β-6)-AO
13	Lentin-5-AO* Glc5(β-3/β-6)-AO
14	Lentin-6-AO* Glc6(β-3/β-6)-AO
15	Lentin-7-AO* Glc7(β-3/β-6)-AO
16	Lentin-8-AO* Glc8(β-3/β-6)-AO
17	Lentin-9-AO* Glc9(β-3/β-6)-AO
18	Lentin-10-AO* Glc10(β-3/β-6)-AO

19	Lentin-11-AO* Glc11(β-3/β-6)-AO
20	Lentin-12-AO* Glc12(β-3/β-6)-AO
21	Lentin-13-AO* Glc13(β-3/β-6)-AO
22	HE-9B7-AO Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO | Glcβ-6
23	HE-10B2-AO Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO | Glcβ-6
24	HE-10B3-AO Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO | Glcβ-6
25	HE-10B5-AO Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO | Glcβ-6
26	HE-10B7-AO Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO | Glcβ-6
27	Gu-6B1/3-AO Glcβ-3Glcα-3Glcβ-3Glc-AO | | Glcβ-6 Glcβ-6
28	HE-11B3/6-AO Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO | | Glcβ-6 Glcβ-6
29	Man-4(β4)-AO Manβ-4Manβ-4Manβ-4Man-AO
30	Man-5(β4)-AO Manβ-4Manβ-4Manβ-4Manβ-4Man-AO
31	LNFP-III Galβ-4GlcNAcβ-3Galβ-4Glc-DH | Fucα-3
32	Man-6(β4)-AO Manβ-4Manβ-4Manβ-4Manβ-4Manβ-4Man-AO
33	Man-7(β4)-AO Fraction
34	Man-8(β4)-AO Fraction

35	Man-9(β4)-AO Fraction
36	Man-10(β4)-AO Fraction
37	Man-11(β4)-AO Fraction
38	Man-12(β4)-AO Fraction
39	Man-13(β4)-AO Fraction
40	Gal-Mannan DP2-AO Galactomannan DP2.AO
41	Gal-Mannan DP3-AO Galactomannan DP3.AO
42	Gal-Mannan DP4-AO Galactomannan DP4.AO
43	Gal-Mannan DP5a-AO Galactomannan DP5a.AO
44	Gal-Mannan DP5b-AO Galactomannan DP5b.AO
45	Gal-Mannan DP6a-AO Galactomannan DP6a.AO
46	Gal-Mannan DP6b-AO Galactomannan DP6(DP5).AO (TLC-2) (Elicityl Man219)
47	Gal-Mannan DP7a-AO Di-galactosyl-mannopenta (DP7).AO (TLC-2) (Megazyme O-GGM5)
48	Gal-Mannan DP7b-AO Galactomannan DP7(DP6, DP8).AO (TLC-1) (Elicityl Man219)
49	Gal-Mannan DP8a-AO Di-galactosyl-mannopenta DP8(DP7).AO (TLC-1) (Megazyme O-GGM5)
50	Gal-Mannan DP8b-AO Galactomannan DP8b-AO
51	Gal-Mannan DP9-AO Galactomannan DP9-AO
52	Gal-Mannan DP10-AO Galactomannan DP10-AO
53	Gal-Mannan DP11-AO Galactomannan DP11-AO

54	Glc-Mannan DP3-AO Galactomannan DP12-AO
55	GN2-AO GlcNAcβ-4GlcNAc-AO
56	GN3-AO GlcNAcβ-4GlcNAcβ-4GlcNAc-AO
57	GN4-AO* GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAc-AO
58	GN5-AO* GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAc-AO
59	GN6-AO* GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAc-AO
60	GN7-AO* GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAc-AO
61	GN8-AO* GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAc-AO
62	GlcN4-AO (chitosan) GlcNβ-4GlcNβ-4GlcNβ-4GlcN-AO
63	GlcN5-AO (chitosan) GlcNβ-4GlcNβ-4GlcNβ-4GlcNβ-4GlcN-AO
64	GlcN6-AO (chitosan) GlcNβ-4GlcNβ-4GlcNβ-4GlcNβ-4GlcNβ-4GlcN-AO
Set 55	
1	Xyl-2(β4)-AO Fraction
2	Xyl-3(β4)-AO Fraction
3	Xyl-4(β4)-AO Fraction
4	Xyl-5(β4)-AO Xylβ-4Xylβ-4Xylβ-4Xylβ-4Xyl-AO
5	Xyl-5(β4)-AO Xylβ-4Xylβ-4Xylβ-4Xylβ-4Xyl-AO
6	Xyl-7(β4)-AO Fraction
7	Xyl-7(β4)-AO Fraction
8	Xyl-7(β4)-AO Fraction

9	Xyl-10(β4)-AO Fraction
10	Xyl-11(β4)-AO Fraction
11	Xyl-12(β4)-AO Fraction
12	Xyl-13(β4)-AO Fraction
13	Ara-2(α5)-AO Fraction
14	Ara-3(α5)-AO Fraction
15	Ara-4(α5)-AO Fraction
16	Ara-5(α5)-AO Fraction
17	Ara-6(α5)-AO Araα-5Araα-5Araα-5Araα-5Araα-5Ara-AO
18	Ara-7(α5)-AO Araα-5Araα-5Araα-5Araα-5Araα-5Araα-5Ara-AO
19	Ara-8(α5)-AO Arabino-octaose.AO (TLC-2) (Megazyme O-AOC)
20	Ara-9(α5)-AO Arabino-octaose DP9(DP8).AO (TLC-1) (Megazyme O-AOC)
21	Ara-5B3-AO Ara-3 | Araα-5Araα-5Araα-5Ara-AO
22	Ara-5B-AO Ara-3 | Araα-5Araα-5Araα-5Ara-AO | Araα-2 Ara-3 | Araα-5Araα-5Araα-5Araα-5Ara-AO
23	Ara-Xylan DP3-AO Araα-3 | Xylβ-4Xyl-AO
24	Ara-Xylan DP4a-AO Xylβ-4Xylβ-4Xyl-AO | Araα-2

25	Ara-Xylan DP4b-AO Xylβ-4Xylβ-4Xyl-AO | Araα-2 Araα-3 | Xylβ-4Xylβ-4Xyl-AO
26	Ara-Xylan DP5a-AO Araα-3 | Xylβ-4Xylβ-4Xylβ-4Xyl-AO
27	Ara-Xylan DP5b-AO Xylβ-4Xylβ-4Xylβ-4Xyl-AO | Araα-2 Araα-3 | Xylβ-4Xylβ-4Xylβ-4Xyl-AO
28	Ara-Xylan DP5c-AO Araα-3 | Xylβ-4Xylβ-4Xyl-AO | Araα-2
29	Ara-Xylan DP6-AO Araα-3 | Xylβ-4Xylβ-4Xylβ-4Xyl-AO | Araα-2
30	Xyl-Glucan DP7-AO Xylα-6 | Glcβ-4Glcβ-4Glcβ-4Glc-AO | | Xylα-6 Xylα-6
31	LNFP-III Galβ-4GlcNAcβ-3Galβ-4Glc-DH | Fucα-3
32	Xyl-Glucan DP8-AO Xylo-glucan Fb DP8(DP-7).AO (TLC-3) (Megazyme O-XGHON)
33	Xyl-Glucan DP9-AO Xylo-glucan Fb DP9.AO (TLC-1)(Megazyme O-XGHON)
34	FG-Xyl-Glucan DP7-AO XFG DP7-AO
35	FG-Xyl-Glucan DP9-AO XLFG DP10-AO
36	GalAra(1-3)-AO Galβ-3Ara-AO

37	GlcNAc2(1-6)-AO GlcNACβ-6GlcNAc-AO
38	Gal2GlcNAc(1-3)-AO Galα-3Galβ-3GlcNAc-AO
39	GalManNAc(1-4)-AO Galβ-4ManNAc-AO
40	Malto-3 Glcα-4Glcα-4Glc-DH
41	Malto-4 Glcα-4Glcα-4Glcα-4Glc-DH
42	Malto-5 Glcα-4Glcα-4Glcα-4Glcα-4Glc-DH
43	Malto-6 Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glc-DH
44	Malto-7 Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glc-DH
45	Dext-3 Glcα-6Glcα-6Glc-DH
46	Dext-4 Glcα-6Glcα-6Glcα-6Glc-DH
47	Dext-5* Glcα-6Glcα-6Glcα-6Glcα-6Glc-DH
48	Dext-6* Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glc-DH
49	Dext-7* Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glc-DH
50	Lam-3 Glcβ-3Glcβ-3Glc-DH
51	Lam-5 Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-DH
52	Lam-6* Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-DH
53	Lam-7 Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-DH
54	Man4(β4) Manβ-4Manβ-4Manβ-4Man-DH
55	Man6(β4) Manβ-4Manβ-4Manβ-4Manβ-4Manβ-4Man-DH

56	Xyl5(β4) Xylβ-4Xylβ-4Xylβ-4Xylβ-4Xyl-DH
57	Xyl6(β4) Xylβ-4Xylβ-4Xylβ-4Xylβ-4Xylβ-4Xyl-DH
58	Ara6(α5) Araα-5Araα-5Araα-5Araα-5Araα-5Ara-DH
59	Ara7(α5) Araα-5Araα-5Araα-5Araα-5Araα-5Araα-5Ara-DH
60	GN3 GlcNAcβ-4GlcNAcβ-4GlcNAc-DH
61	NSG-11-AO* Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO
62	Barley-5-AO* Glcβ-4Glcβ-4Glcβ-4Glcβ-3Glc-AO
63	Barley-6a-AO Glcβ-3Glcβ-4Glcβ-6Glcβ-4Glcβ-3Glc-AO; and Glcβ-3Glcβ-4Glcβ-4Glcβ-4Glcβ-3Glc-AO (ratio ~1:1)
64	Barley-14-AO* (Glcβ-4/3Glcβ)6-4Glcβ-3Glc-AO
Set 56	
1	Dext-4-AO Glcα-6Glcα-6Glcα-6Glc-AO
2	Barley-11-AO* (Glcβ-4/3Glcβ)5-3Glc-AO
3	Barley-14-AO* (Glcβ-4/3Glcβ)6-4Glcβ-3Glc-AO
4	Barley-15-AO* (Glcβ-4/3Glcβ)7-3Glc-AO
5	Lentin-6-AO* Glc6(β-3/β-6)-AO
6	HE-10B7-AO Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO Glcβ-6
7	Man-5(β4)-AO Manβ-4Manβ-4Manβ-4Manβ-4Man-AO
8	Gal-Mannan DP4-AO Galactomannan DP4.AO
9	Glc-Mannan DP3-AO Galactomannan DP12-AO

10	GN5-AO* GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAc-AO
11	Xyl-11(β4)-AO Fraction
12	Ara-3(α5)-AO Fraction
13	Ara-5B-AO Ara-3 | Araα-5Araα-5Araα-5Ara-AO | Araα-2 Ara-3 | Araα-5Araα-5Araα-5Araα-5Ara-AO
14	Ara-Xylan DP5a-AO Araα-3 | Xylβ-4Xylβ-4Xylβ-4Xyl-AO
15	FG-Xyl-Glucan DP7-AO XFG DP7-AO
16	Barley-5-AO* Glcβ-4Glcβ-4Glcβ-4Glcβ-3Glc-AO
17	Lea-Tri Galβ-3GlcNAc-DH | Fucα-4
18	NeuAcα-(3')Lac-AO NeuAcα-3Galβ-4Glc-AO
19	NeuAcα-(6')Lac-AO NeuAcα-6Galβ-4Glc-AO
20	Neu4,5Ac-(3')Lac-AO (4-OAc)NeuAcα-3Galβ-4Glc-AO
21	GSC-260 3-deoxy,3-carboxymethyl-Galβ-4Glcβ-C30 | Fucα-3
22	SU(3')-LN SU-3Galβ-4GlcNAc-DH
23	GSC-426 (3-deoxy,3-carboxymethyl)Galβ-C30
24	GSC-178 NeuAcα-3Galβ-4Glcβ-Cer34

25	B-Tetra-T1-AO Galα-3Galβ-3GlcNAc-AO | Fucα-2
26	H-Tri-T2-AO Fucα-2Galβ-4GlcNAc-AO
27	Globo-A-Hepta GalNAcα-3Galβ-3GalNAcβ-3Galα-4Galβ-4GlcNAc-DH | Fucα-2
28	SA(6')-LNFP-VI NeuAcα-6Galβ-4GlcNAcβ-3Galβ-4Glc-DH | Fucα-3
29	GSC-177 NeuGcα-3Galβ-4GlcNAcβ-3Galβ-Cer36 | Fucα-3
30	pLNH Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH
31	GSC-915-4 (new) Galβ-4GlcNAcβ-6Galβ-4GlcNAcβ-3Galβ-4Glc-DH
32	GSC-915-5 GlcNAcβ-6Galβ-4GlcNAcβ-3Galβ-4Glc-DH
33	NeuAcα-(6')LNnO (F1) NeuAcα-6Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH
34	Man3XylGN2 Manα-6 | Xylβ-2Manβ-4GlcNAcβ-4GlcNAc-DH | Manα-3
35	N2 Manα-6 | Manβ-4GlcNAcβ-4GlcNAc-DH | Galβ-4GlcNAcβ-2Manα-3
36	NGA2 GlcNAcβ-2Manα-6 | Manβ-4GlcNAcβ-4GlcNAc-DH | GlcNAcβ-2Manα-3
37	NGA2F GlcNAcβ-2Manα-6 Fucα-6 | | Manβ-4GlcNAcβ-4GlcNAc-DH | GlcNAcβ-2Manα-3

38	NA2F Galβ-4GlcNAcβ-2Manα-6 Fucα-6 | | Manβ-4GlcNAcβ-4GlcNAc-DH | Galβ-4GlcNAcβ-2Manα-3
39	NA2FB Galβ-4GlcNAcβ-2Manα-6 Fucα-6 | | GlcNAcβ-4Manβ-4GlcNAcβ-4GlcNAc-DH | Galβ-4GlcNAcβ-2Manα-3
40	AGP-Bi-AcGc NeuGc(Ac)α-6Galβ-4GlcNAcβ-2Manα-6 | Manβ-4GlcNAcβ-4GlcNAc-DH | NeuAc(Gc)α-6Galβ-4GlcNAcβ-2Manα-3
41	AGP-Bi-Gc2 NeuGcα-6Galβ-4GlcNAcβ-2Manα-6 | Manβ-4GlcNAcβ-4GlcNAc-DH | NeuGcα-6Galβ-4GlcNAcβ-2Manα-3
42	GM1-penta Galβ-3GalNAcβ-4Galβ-4Glc-DH | NeuAcα-3
43	BSM-Di-A2-AO NeuAcα-6GalNAc-AO
44	XylGlc-AO Xylα-3Glc-AO
45	SA4(α8) NeuAcα-8NeuAcα-8NeuAcα-8NeuAc-DH
46	Man Man-DH
47	HS-S8-AO* GlcAβ-4GlcNAcα-4GlcAβ-4GlcNAcα-4GlcAβ-4GlcNAcα-4GlcAβ-4aMan-AO
48	Hep-S4-AO* IdoAα-4GlcNSα-4IdoAα-4aMan-AO | | | | SU-2 SU-6 SU-2 SU-6
49	HA-S6* GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAcβ-4GlcAβ-3GlcNAc-DH
50	CSB-8* ΔUA-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAcβ-4IdoAα-3GalNAc-DH | | | | SU-4 SU-4 SU-4 SU-4

51	CSB-2-AO* (Nian) ΔUA-3GalNAc-AO | SU-4
52	CSC-20-AO* (Nian) ΔUA-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAcβ-4GlcAβ-3GalNAc-AO | | | | | | | | | | SU-6 SU-6 SU-6 SU-6 SU-6 SU-6 SU-6 SU-6 SU-6 SU-6
53	HS(deAc)-2-AO* (Nian)
54	HS(deAc)-4-AO* (Nian)
55	SA2(α2-9) NeuAcα-9NeuAc-DH
56	SA3(α2-9) NeuAcα-9NeuAcα-9NeuAc-DH NeuAcα-9NeuAcα-9NeuAcα-9NeuAc-DH
57	SA4(α2-9) NeuAcα-9NeuAcα-9NeuAcα-9NeuAc-DH
58	SA5(α2-9) NeuAcα-9NeuAcα-9NeuAcα-9NeuAcα-9NeuAc-DH
59	SA6(α2-9) NeuAcα-9NeuAcα-9NeuAcα-9NeuAcα-9NeuAcα-9NeuAc-DH
60	GSC-197 KDNα-3Galβ-4Glcβ-Cer28
61	LNFP-III Galβ-4GlcNAcβ-3Galβ-4Glc-DH | Fucα-3
62	PI-1 NeuAcα-3(6-NAc)Galβ-4GlcNAc-DH
63	PI-2 NeuAcα-3(6-NBz)Galβ-4GlcNAc-DH
64	Su(3')-Lex tri-AO SU-3Galβ-4GlcNAc-AO | Fucα-3