

eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

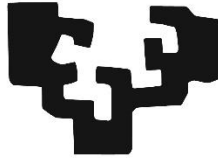
**NMR and molecular recognition: The interaction of human
galectin-4 with the histo blood group antigens and with
pathogen-associated molecules**

Jon Imanol Quintana García

Doctoral Thesis

2022

eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

**NMR and molecular recognition: The interaction of human
galectin-4 with the histo blood group antigens and with
pathogen-associated molecules**

Jon Imanol Quintana García

Doctoral Thesis

2022

Supervisors:

Prof. Dr. Jesús Jiménez Barbero

Dr. Ana Ardá Freire

University Tutor: Nuria Sotomayor Anduiza

**This doctoral thesis has been performed at the Center for Cooperative
Research in Biosciences (CIC bioGUNE)**

Acknowledgements

The work herein described was performed at the Center for Cooperative Research in Biosciences (CIC bioGUNE), under the supervision of Prof. Dr. Jesús Jiménez-Barbero and Dr. Ana Ardá Freire.

Firstly, I would like to thank both my supervisors. Thank you for the opportunity of working under your supervision for the last 4.5 years and developing my master's and doctoral thesis. I am very grateful for everything that I have learnt from you.

Secondly, I would like to thank to all my colleagues from the Chemical Glycobiology lab. Without your continuous support I would not be here.

During 2021, I performed a secondment at the Max Planck Institute for colloids and interfaces at Potsdam. I would like to thank Prof. Dr. Peter Seeberger for the opportunity of performing my secondment there. I am also very grateful for all the help and support of Dr. Martina Delbianco, Dr. Jose Angel Dangel-Flores during my secondment.

Finally, I would also like to thank to all the colleagues and collaborators at CIC bioGUNE.

Abbreviations

3'SLN 3'Sialyl N-acetyllactosamine

6'SLN 6'Sialyl N-acetyllactosamine

AGA Automated Glycan Assembly

BB Building Block

Bn Benzyl

BLI BioLayer Interferometry

Bz Benzoyl

CD Circular Dichroism

CRD Carbohydrate Recognition Domain

CryoEM Cryo Electron Microscopy

CSP Chemical Shift Perturbation

DCM Dichloromethane

Dectin-1 dendritic-cell-associated C-type lectin-1

DMAP 4-(Dimethylamino)pyridine

DOSY Diffusion Ordered Spectroscopy

DQ Double-Quantum

E. Coli *Escherichia Coli*

EA Ethyl acetate

ER Endoplasmic Reticulum

ESI ElectroSpray Ionization

EtOAc Ethyl acetate

EXSY Chemical Exchange Spectroscopy

Fuc Fucose

GalNAc N-Acetylgalactosamine

Gal Galactose

Gal-4 Galectin-4

Gal-4FL Full-Length Galectin-4
Gal-4N Galectin-4 N-terminal carbohydrate recognition domain
Gal-4C Galectin-4 C-terminal carbohydrate recognition domain
GAG GlycosAminoGlycan
Glc Glucose
GlcA Glucuronic Acid
GlcNAc N-Acetylglucosamine
GLS GLycoSphingolipid
GPI GlycosylPhosphatidylInositol
HB Hydrogen Bond
HBGA Histo Blood Group Antigens
HEK Human Embryonic Kidney
HPLC High Performance Liquid Chromatography
HSQC Heteronuclear Single Quantum Coherence
IdoA Iduronic Acid
INEPT Insensitive Nuclei Enhanced by Polarization Transfer
ITC Isothermal Titration Calorimetry
LacNAc N-Acetyllactosamine
LBP Lipopolysaccharide Binding Protein
LPS Lipopolysaccharide
MALDI Matrix-Assisted Laser Desorption/Ionization
Man Mannose
MeOH Methanol
MD Molecular Dynamics
MS Mass Spectrometry
MW Molecular Weight
NHAc N-Acetyl
NIS N-Iodosuccinimide

NMR Nuclear Magnetic Resonance **NOE** Nuclear Overhauser Effect

NOESY NOE Spectroscopy

PAMP Pathogen Associated Molecular Pattern

PDB Protein Data Bank

PRR Pattern-Recognition Receptor

ROESY Rotating frame Overhauser Effect Spectroscopy

SAXS/WAXS Small Angle X-ray Scattering/ Wide Angle X-ray Scattering

SDS-PAGE Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis

SNFG Symbol Nomenclature For Glycans

SPR Surface Plasmon Resonance

STD Saturation Transfer Difference

TBSCI tert-Butyldimethylsilyl chloride

THF Tetrahydrofuran

TLR Toll-Like Receptor

TMSOTf Trimethylsilyl trifluoromethanesulfonate

TOCSY Totally Correlated Spectroscopy

TROSY Transverse Relaxation-Optimized Spectroscopy

UV Ultraviolet

WT Wild Type

Xyl Xylose

ZQ Zero-Quantum

Abstract

Glycans are everywhere. Carbohydrates are essential biomolecules that are found ubiquitously in every living species. In humans, every cell is coated with a dense layer of carbohydrates known as the glycocalyx, which interacts with multiple kind of entities. These interactions are essential for cell adhesion, signalling events, host-pathogen interactions, cancer development, and many more.

Galectins are a family of β -galactoside-binding proteins that mediate in a broad variety of physiological functions because of their interaction with saccharides, participating in cell-cell interactions, immune response, and intracellular signaling, among others. Human galectin-4 (hGal-4) is a tandem-repeat protein with two carbohydrate recognition domains (CRD) united through a linker-peptide. Both up- and down-regulation of hGal-4 expression levels have been found in cancerous tissues, making this lectin an interesting drug target. The relevance of the molecular recognition of human blood group antigens by hGal-4 in biological processes has been reported. For instance, tumor cells are able to express hGal-4 to interact with red blood cells. Additionally, hGal-4 has shown to have bactericidal activity against bacteria mimicking blood group antigens. However, the essential features of the interaction events, including the knowledge of the dynamic properties of both the receptor and the ligand are yet to be ascertained.

To gain insights into the key elements of the recognition process, the interaction between the N- and C-terminal CRDs and the blood group A- and B-antigens have been studied by Nuclear Magnetic Resonance Spectroscopy (NMR). In particular, the analysis of STD-NMR, tr-NOESY, and tr-ROESY experiments, and ^1H - ^{15}N -HSQC based titrations, reporting on information from the ligand

and the lectin's point of view respectively, have permitted to shed light into the fine details of these glycan-lectin interactions. In addition, the interaction of the antigens with the full-length hGal-4 has also been analyzed. Production of two mutants of Gal-4 (Gal-4FL-H236R and Gal-4FL-H63R) that only contain one active CRD helped to decipher the impact of the structure of the full-length protein into the recognition events. STD-NMR, ^1H - ^{15}N HSQC based titrations and Isothermal Titration Calorimetry (ITC) measurements were performed to achieve such objectives. Finally, the combination of the NMR experimental data with those deduced from computational protocols have allowed explaining the selectivity of the lectin towards the distinct ligands.

Since multiple pathogens mimic blood group antigens, the interaction between a lipopolysaccharide from an *E. Coli* strain whose O-antigen mimics blood group antigens (*E. Coli* O55) and galectin-4 has been analysed. ^1H - ^{15}N HSQC of Gal-4 (both the single domains, the full-length and the mutants) in the presence and absence of LPS allowed revealing key features on the recognition of this multivalent endotoxin by Gal-4.

The main co-receptor of LPS, the glycoprotein CD14, has been reported to interact with Gal-4 through its N-glycans. Therefore, CD14 has been expressed in human embryonic kidney cells (HEK293F cells) to study the nature of these N-glycans and the interaction with Gal-4. The combination of NMR and mass spectrometry have allowed determining the glycoprofile at each glycosylation sites. Through point mutations of the detected N-glycosylation sites, the impact of each N-glycosite in the folding and stability of the protein has been deduced. Additionally, the interaction of CD14 with LPS has been characterized.

Finally, the last project involved β -glucans, which are polysaccharides produced by certain fungi and bacteria. These repeating-glycan structures, intrinsically heterogeneous, are known to interact with lectins of our immune

system and display different biotechnological and clinical applications. With the aim of producing β -glucans of well-defined structure, solid-phase-automated synthesis (also known as Automated Glycan Assembly) was explored. For that, two monosaccharide building blocks have been synthesized, and the synthesis of three different well-defined β -glucans has been pursued.

Contents

1	Introduction.....	1
1.1	Structural Aspects of Carbohydrates.....	2
1.2	Distribution of Glycans.....	8
1.3	Protein-carbohydrate interactions	15
1.4	Experimental techniques to characterize protein-glycan interactions.....	30
1.4.1	Nuclear Magnetic Resonance.....	32
1.4.2	Isothermal Titration Calorimetry	54
2	Objectives	58
3	Results and Discussion	60
3.1	The interaction between Galectin-4 and the histo blood group antigens ...	61
3.1.1	Introduction.....	61
3.1.2	The N-terminal domain of Galectin-4.....	66
3.1.3	The C-terminal domain of Galectin-4.....	83
3.1.4	Full-length Galectin-4.....	97
3.2	Towards understanding Galectin-4 interactions: the recognition of LPS and CD14	108
3.2.1.	Introduction.....	108
3.2.2.	The glycoprofile of CD14 as deduced by NMR	113
3.2.3.	Site-specific determination of the glycans by MS	116
3.2.4.	Glycosidase treatment to determine glycosylation profiles.	117
3.2.5	Towards a 3D-model of glycosylated CD14.....	119
3.2.6	Impact of the N-glycosylation on the stability of the protein	121
3.2.7	The interaction of CD14 with human Galectin-4.....	122
3.2.8	The interaction of Galectin-4 with LPS	125
3.2.9	The interaction of CD14 with the LPS.....	130
3.2.10	Conclusions.....	132
3.3	Synthesis of well-defined β -glucans using Automated Glycan Assembly	134
3.3.1	Introduction.....	134
3.3.2	Synthesis of the building blocks	138
3.3.3	Automated Glycan Assembly (AGA)	139

3.3.4	Deprotection and purification of the oligosaccharides.....	141
3.3.5	Conclusions.....	143
4	Materials and methods.....	144
4.1.	Protein expression and purification.....	145
4.1.1.	Cloning of the vectors.....	145
4.1.2.	Protein expression.....	145
4.2	Glycosidase treatment of CD14.....	149
4.3	Circular Dichroism (CD) measurements.....	149
4.4	Ligands.....	149
4.5	Molecular Dynamic (MD) Simulations.....	149
4.5.1	MD simulations of Gal-4N and Gal-4C in complex blood group antigens	150
4.5.2	MD simulations of CD14.....	150
4.6	Nuclear Magnetic Resonance (NMR).....	150
4.6.1	General information.....	150
4.6.2	¹ H- ¹⁵ N-HSQC based titrations.....	151
4.6.3	Saturation Transfer Difference (STD) NMR.....	151
4.6.4	trROESY spectra.....	152
4.6.5	Protein Backbone Resonance Assignment.....	152
4.6.6	Assignment of the ¹ H- ¹³ C-HSQC of the glycans of CD14.....	153
4.6.7	¹ H- ¹³ C-HSQC of the glycans of CD14 in the presence of lectins....	153
4.7	Isothermal Titration Calorimetry (ITC).....	153
4.8	Automated Glycan Assembly.....	154
4.8.1	General materials and methods.....	154
4.8.2	Building blocks.....	154
4.8.3	Preparation of stock solutions.....	155
4.8.4	Modules for automated synthesis.....	156
4.8.5	Post-AGA manipulations.....	160
4.8.6	Synthesis of the tetrasaccharide (11).....	161
4.8.7	Synthesis of the branched pentasaccharide (12).....	162
4.8.8	Synthesis of the linear octasaccharide (13).....	162
4.8.9	Experimental part of building block synthesis.....	163
5	General conclusions.....	170

6	References.....	174
7	Supporting Information.....	203
7.1	The interaction between Galectin-4 and histo blood group antigens	204
7.1.1	N-terminal domain of Galectin-4.....	204
7.1.2	C-terminal domain of Galectin-4	215
7.1.3	Full-length Galectin-4.....	224
7.2	Towards understanding Galectin-4 interactions: the recognition of LPS and CD14	226
7.3	Synthesis of well-defined β -glucans using Automated Glycan Assembly	230

1 Introduction

Carbohydrates are essential biomolecules that are found ubiquitously in every living species. The roles of carbohydrates, also known as sugars, glycans, or saccharides are multiple; they are the main energy source for humans and one of the major components of the cell wall of plants. However, despite these renowned functions, carbohydrates also have a role in recognition events related to health and disease that are of paramount relevance.

The interactions mediated through carbohydrates occur in multiple ways and with different kind of entities. Glycans can interact between themselves, for instance, in cell-cell recognition¹. However, the most studied and relevant systems involve protein-carbohydrate interactions. These interactions are essential for cell adhesion², signalling events³, host-pathogen interactions⁴, cancer development⁵, and many more.

In order to be able to comprehend the nature of these interactions mediated through carbohydrates, firstly, it is relevant to understand the chemical nature and distribution of these molecules.

1.1 Structural Aspects of Carbohydrates

Glycans, as other biologically relevant molecules, are composed of small building blocks named monosaccharides, which are combined to form large and highly diverse structures. Unlike for DNA or proteins, the combination of these building blocks is of higher chemical complexity for various reasons. Firstly, each monosaccharide exists in two epimers, alpha and beta (Fig 1). Secondly, the number of “pieces” to build complex glycans is over one hundred, significantly more in comparison to other biomolecules. Another relevant feature is that the same chemical structure can form rings of different sizes. Additionally, many monosaccharides have various functional groups that can be modified. Finally, each building block can be connected to another through different positions. All these factors, increase quantitatively the number of possible combinations to generate large structures, and subsequently increase the difficulty of understanding the role of glycans in nature.

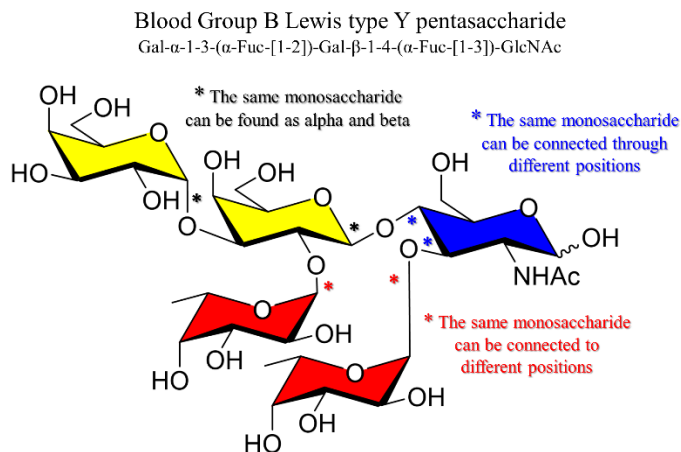


Figure 1. Chemical structure of the Blood Group B Lewis Type Y pentasaccharide antigen. Galactoses (Gal) are coloured in yellow, fucoses (Fuc) in red and the N-Acetylglucosamine (GlcNAc) in blue.

As mentioned before, each monosaccharide can be found as two different diastereomers in the anomeric position. The reason behind this fact is a phenomenon named mutarotation⁶. Although monosaccharides are mainly found as cyclic structures (hemiacetals) most of the time, in solution, these two epimers are in equilibrium with an open conformation (aldehyde) (Fig 2). There are various factors such as temperature, pH or concentration⁷ that affect the stability of each epimer, and therefore, the equilibrium and relative proportion of each species.

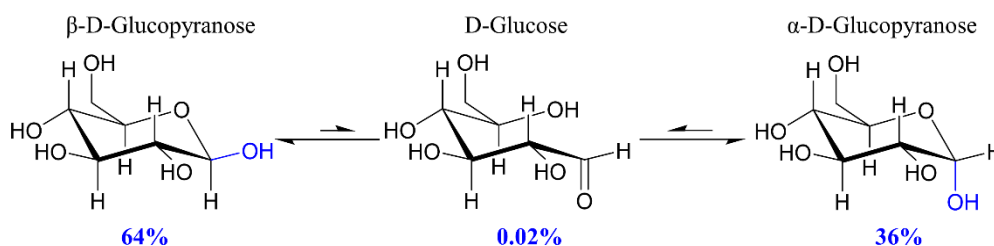


Figure 2. Mutarotation of D-Glucose and the percentage of each of species in water at 25°C.

Regarding the stability of each epimer, steric effects favour that the bulkier group adopts the equatorial position. However, in many occasions, the epimers with the hydroxyl group in an axial position are more populated than could be expected. The phenomenon behind this effect is known as the endo-anomeric effect⁸. When the hydroxyl group is axial, the anti-bonding orbital of the C1-O1 bond overlaps with one of the lone pairs of the endocyclic oxygen (Fig 3), whereas when it is equatorial, this interaction is not present. The hyperconjugative delocalization of the electron pair stabilizes the alpha anomer and makes this form energetically favourable. This explanation is supported by X-Ray crystallography data since, for α-D-glucopyranosides, the bond between O5 and C1 is shorter and the C1-O1 bond is longer compared to the β-anomers.

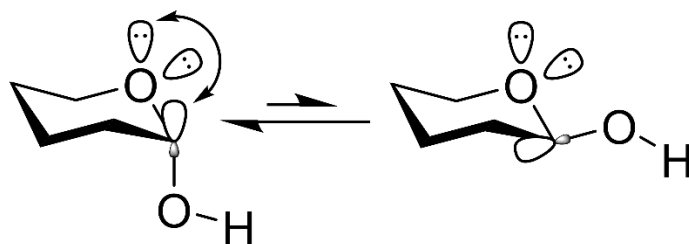


Figure 3. Schematic representation of the interaction between the lone pair of the endocyclic oxygen and the σ^* orbital of the C₁-O₁ linkage.

An additional complexity level arises from the linkage between monosaccharide units. This linkage, known as the glycosidic bond, is also of paramount relevance for the presentation of carbohydrates, since the type of linkage affects the orientation of each residue, which has a direct impact in the properties of the molecule. As an example of the relevance of this linkage, in nature, many linear glucose polymers are found with different glycosidic linkages, such as laminaran glucans (β 1-3 linked), cellulose (β 1-4 linked) or dextrans (β 1-6 linked). Although the chemical composition of these polymers in terms of monosaccharides is the same, their physical and chemical properties and functions are completely different⁹. The relevance of the glycosidic linkages has also been detected in recognition events, with many proteins only being able to recognize one specific type of linkage¹⁰.

The glycosidic linkage can be usually defined with two dihedral angles, Φ (H1-C1-O1-C_X) and Ψ (C1-O1-C_X-H_X). As shown in Fig 4, there are three possible canonical conformations for Φ . However, it has been shown that one of those conformations is the major one. For the *exo-syn* and *exo-anti* conformations, there is an overlap between the anti-bonding orbital of the O5-C1 bond and one of the lone pairs of the exocyclic oxygen, whereas for *non-exo* analogue, there is not overlap. Due to this overlap, those two conformers are more favoured

energetically than the third one. This effect is known as the exo-anomeric effect¹¹. In addition, steric factors also play a role in the conformational preferences, and since the first conformer has less steric hindrance than the second one, it is the most populated conformer.

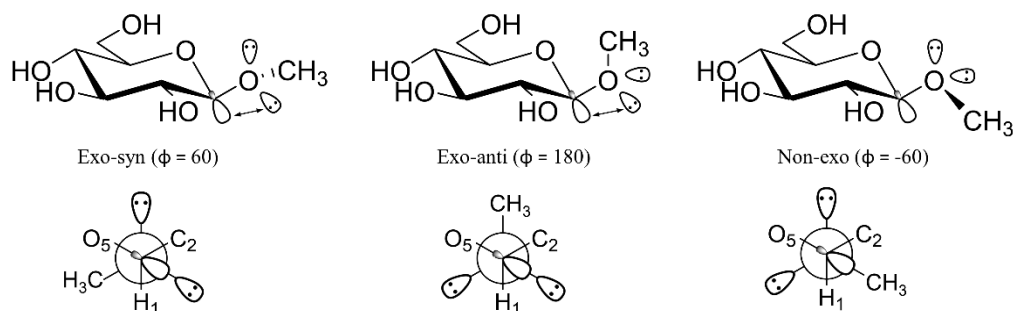


Figure 4. Possible torsion angles for the Φ dihedral angle in β -methylglucopyranose. Below, Newman projections for the C1-O5 bond.

In relation to the Ψ angle, steric interactions are the main factors to rule the predilection of the conformations around this bond.

The C6 hydroxymethyl group in pyranosides also displays conformational mobility¹². Since glycosidic linkages may also involve the oxygen at position O6, the conformation of the C6 hydroxymethyl can directly affect the presentation of the molecule. There are three possible conformations for the hydroxymethyl group, *gauche-gauche* (*gg*), *trans-gauche* (*tg*) and *gauche-trans* (*gt*) (Fig 5). However, in solution, for most pyranosides, a conformational equilibrium between two rotamers is present, as the orientation of the hydroxyl group at position 4 provides additional destabilizing interactions and restricts the motion for one of the three possible conformers. For instance, the population of *tg* for glucose or *gt* for galactose is almost negligible in solution, due to unfavourable interactions between O6 with the hydroxyl group at position 4.

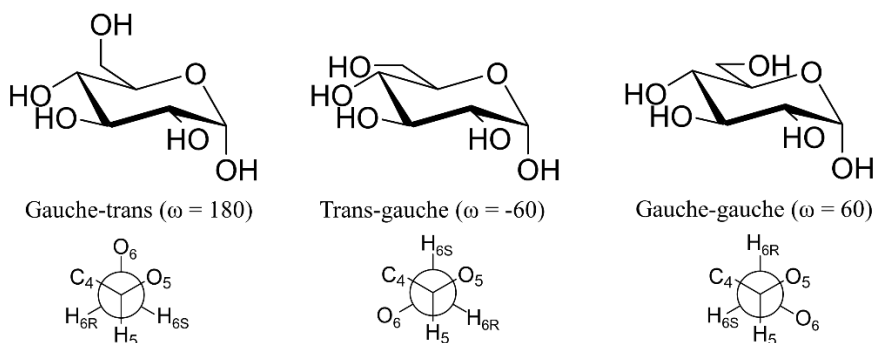


Figure 5. The three possible conformers for the hydroxymethyl group of pyranosides. ω is defined as the angle H5-C5-C6-O6. Below, Newman projections for the C5-C6 bond.

Even if in nature there are hundreds of different monosaccharides, most of the human glycome is composed by ten basic monosaccharides: D-mannose (Man), D-glucose (Glc), N-Acetyl-D-glucosamine (GlcNAc), L-fucose (Fuc), D-galactose (Gal), N-Acetyl-D-galactosamine (GalNAc), N-Acetylneuraminic acid (Neu5Ac), Xylose (Xyl), Iduronic Acid (IdoA), and Glucuronic Acid (GlcA) (Fig 6). Although the number of building blocks might seem small, the structures that these carbohydrates naturally form are multiple, since they can be connected through multiple positions. As a result, carbohydrates are one of the most complex, and yet extremely relevant biomolecules.

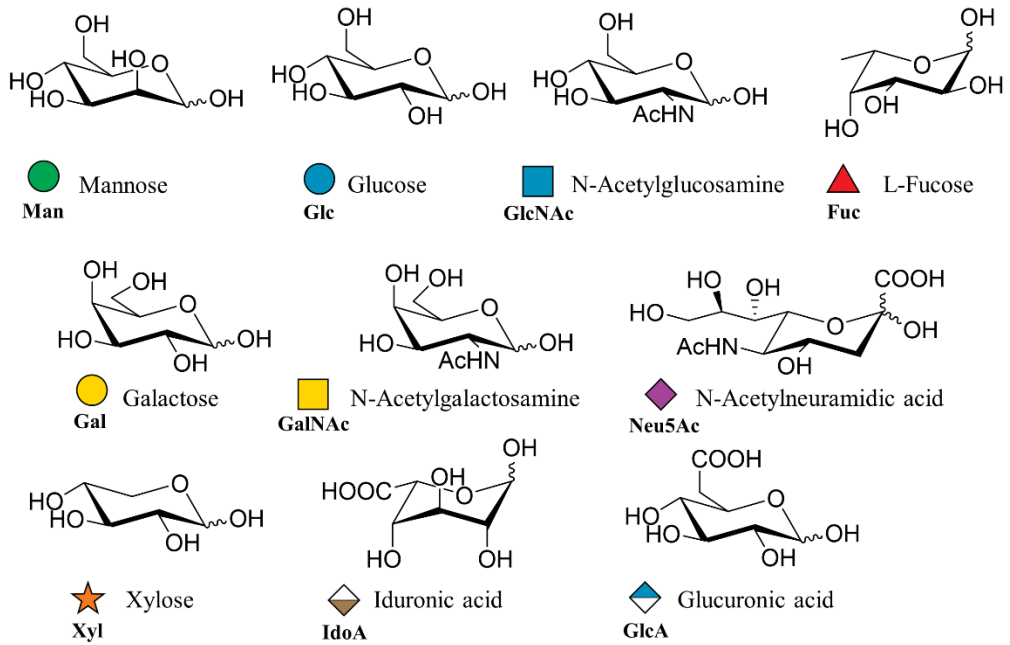


Figure 6. Main basic carbohydrates of the human glycome, their Symbolic Nomenclature for Glycans (SNFG), name, and chemical structure.

1.2 Distribution of Glycans

Glycans are everywhere. Every human cell is coated with a dense layer of carbohydrates known as the glycocalyx^{13,14}. This layer is formed by free glycans (polysaccharides) as well as by glycoconjugates (glycoproteins, proteoglycans, and glycolipids) (Fig 7). As the first layer of the cell, the roles of the glycocalyx are multiple¹⁵. It constitutes a physical barrier for any entity entering the cell, providing an important role in preventing pathogens to enter the cell. The glycocalyx is also relevant for the regulation of the immune system and the diffusion of membrane-proteins, as well as for the cell morphology, and diseases development.

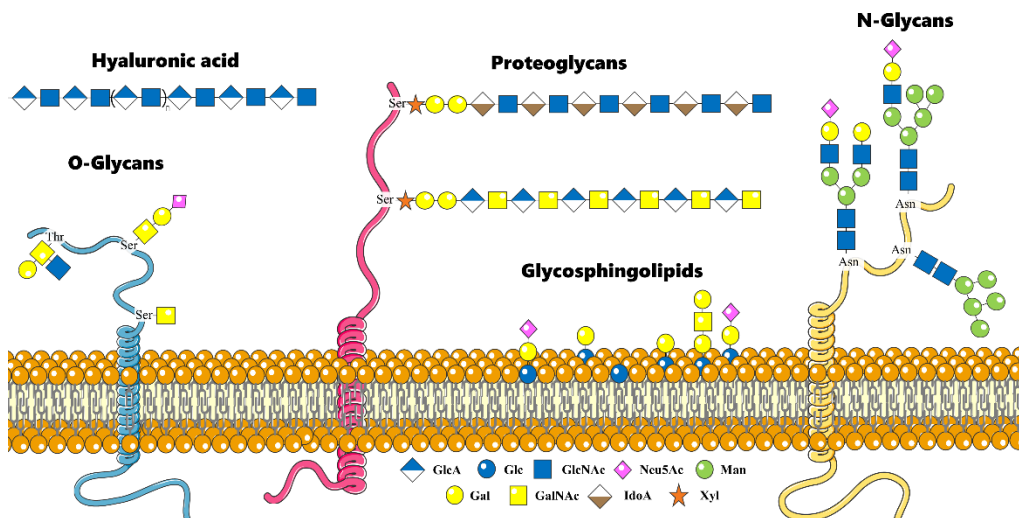


Figure 7. Different types of glycans found on the cell-surface.

One of the main components of the glycocalyx are glycoproteins, which may also be found in the cytoplasm and in the cell nucleus, as well as extracellularly. The glycosylation of proteins is one of the most common post-translational modifications and it is estimated that at least 50% of human proteins are

glycosylated¹⁶. Protein glycosylation is not a mere decoration, since it has been shown that it is relevant for the folding, stability, and function of proteins¹⁷. Therefore, understanding the mechanisms underlying the glycosylation of proteins is of huge relevance.

Most proteins are glycosylated at specific residues: asparagine (N-glycosylation), serine, threonine, and tyrosine (O-glycosylation), or tryptophan (C-glycosylation)¹⁸. Other proteins can be glycosylated through a glycosylphosphatidylinositol (GPI) anchor attached to the C-termini at different amino acids¹⁹. Although every kind of protein glycosylation is relevant, the most frequent ones are N- and O- glycosylation.

Protein N-glycosylation is a process that starts in the lumen of the endoplasmic reticulum (ER), in which a sugar containing 14 monosaccharides (Glc₃Man₉GlcNAc₂) is covalently attached to the side chain of an asparagine in an unfolded protein²⁰. This attachment is not aleatory, but rather specific. For the Glc₃Man₉GlcNAc₂ glycan to be attached to an asparagine residue, a specific amino acid sequence, Asn-X-Ser/Thr, in which X is not proline, is needed. Although this sequon is a minimal requirement for N-glycosylation, its presence does not obligatory imply that that particular Asn residue will be glycosylated.

Glycosylation is a very complex process that involves over 200 glycosyltransferase and glycosidase enzymes. Even if the event always starts with the attachment of the same sugar, the final glycosylation product is rather heterogeneous²¹. The starting Glc₃Man₉GlcNAc₂ structure is modified by various enzymes to yield multiple glycans in the same glycosylation site. This heterogeneity is defined as microheterogeneity or site heterogeneity. Nevertheless, there is a common pathway that is maintained in every human cell.

After the attachment of the initial tetradecasaccharide, glucosidases remove the external glucose units during protein folding (Fig 8). Before leaving the ER, mannosidases remove the mannose of the central arm. Then, the protein is translocated to the cis-Golgi, where $\text{Man}\alpha 1\text{-}2$ residues can be trimmed, obtaining $\text{Man}_{5\text{-}9}\text{GlcNAc}_2$. These type of N-glycans, defined as high-mannose constitute the first of the three canonical groups of N-glycans. Proteins can be secreted containing these high-mannose type N-glycans or follow their processing in the medial- and trans-Golgi. The maturation of the glycosylation can continue to obtain the complex and hybrid types of N-glycans. The modifications that the $\text{Man}_5\text{GlcNAc}_2$ N-Glycan can undergo are multiple, and as a result, hundreds of different structures can be built²².

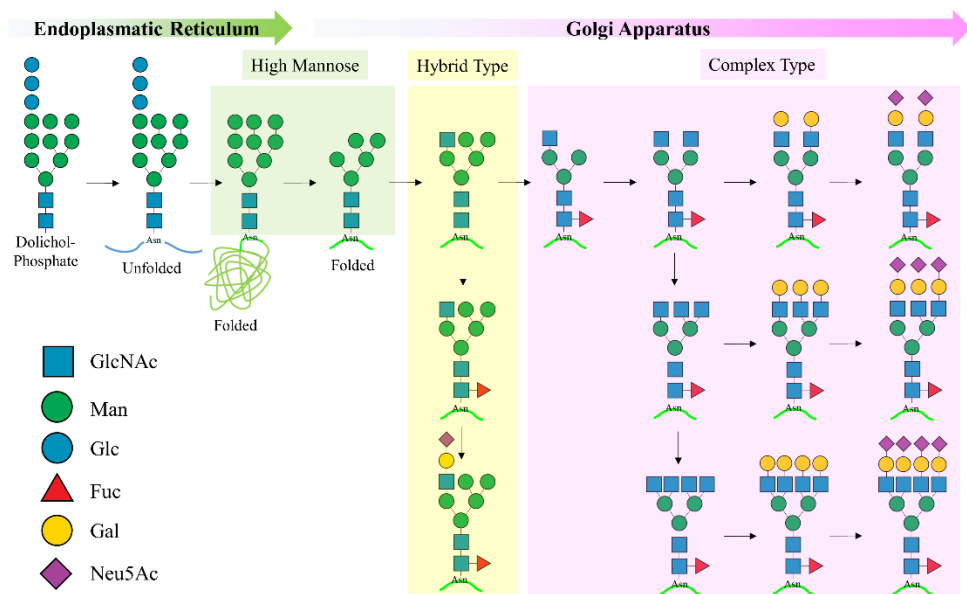


Figure 8. Common N-glycosylation pathway in humans. Adapted from ²²

Unlike for N-glycosylation, there is no consensus sequence for O-glycosylation, which makes it more difficult to study²³. The most common O-glycosylation is the mucin-type O-glycosylation, in which a GalNAc residue is

α 1-linked to the hydroxyl group of serine or a threonine residue²⁴. Subsequently, through various enzymes, the glycosidic part is elongated leading to multiple possible cores (Fig 9). Nevertheless, there are other types of O-glycosylation too, such as O-linked GlcNAc or O-linked Fuc, which are also fairly relevant^{25,26}.

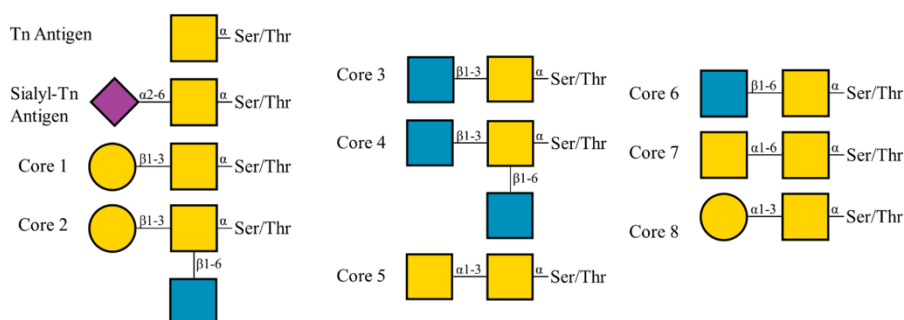


Figure 9. Mucin type O-glycosylation cores found in humans.

In order to transfer the first GalNAc to the target Ser or Thr, there are more than 20 GalNAc-transferases, which are expressed in a tissue-specific manner²⁷. Each transferase has its own specificity, and therefore, the mucin-type O-glycosylation event is very dependent on the tissue where the protein is expressed²⁸. Despite the fact that the O-glycosylation changes among tissues, a feature that hinders the site prediction, since O-glycosylation occurs after protein folding, mainly exposed Ser and Thr residues are affected.

Another type of relevant O-glycosylation is that found in proteoglycans. Proteoglycans are extremely glycosylated proteins, which have a highly negatively charged polysaccharide O-linked to a serine²⁹. The polysaccharide, known as glycosaminoglycan (GAG), is formed by repeating units of disaccharides (20 to 200 units) and can be classified in three types regarding their composition: heparin/heparan sulfate, chondroitin/dermatan sulfate, or

keratan sulfate (Fig 10). With the exception of hyaluronic acid, GAGs can be N- or O-sulfated during their biosynthesis. The anionic character of proteoglycans is very relevant from the structural and molecular recognition points of view, since due to their negative charge and viscoelasticity, they are able to attract cations and water molecules and neutralize pressure changes in tendons or in intervertebral disks³⁰. Additionally, they are also relevant to organize multiple interactions between proteins, directing many biological processes.

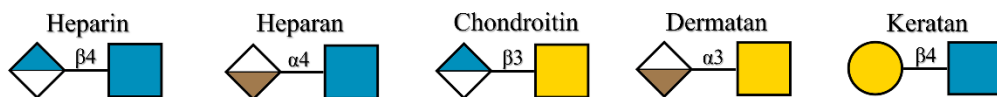


Figure 10. Repeating unit canonical disaccharides of the most common glycosaminoglycan types.

Despite its major importance, the factors that define the glycosylation of a protein are not completely understood. However, it is known that the protein sequence, the tissue in which is expressed and other factors, such as oxidative stress, pH, or temperature, modulate the outcome of the glycosylation^{18,31,32}.

Glycolipids are other glycoconjugates found in humans. Most human glycolipids are glycosphingolipids (GLS), which are constituted of a hydrophobic lipid moiety to which a hydrophilic carbohydrate is covalently attached (Fig 11)³³. The hydrophilic part of GLS on most occasions start with a glucose residue, although some may start with galactose. On the other side, the hydrophobic tail is usually a ceramide or, on some occasions, a sphingosine moiety. Regarding the glycosidic part, its composition is very heterogeneous, but most of the GLS can be subclassified in three groups, according to the common carbohydrate part: i) ganglio-series (GalNAc β 1-4Gal β 1-4Glc β 1-1'Cer), ii) lacto (GlcNAc β 1-3Gal β 1-4Glc β 1-1'Cer) and neolacto (Gal β 1-

Nevertheless, the glycosylation pattern of a cell may change not only due to genetic reasons. For example, upon cancer or inflammatory diseases, the expression level of glycosyltransferases are modified, altering the glycosylation on the cell surface. These alterations change how the cell interacts with its environment, what can contribute to the evolution of the disease. As a result, certain glycosylation patterns, like aberrant sialylation, can be tracked as biomarkers for certain diseases^{40,41}.

Carbohydrates are also present in pathogens like virus, bacteria, parasites, or fungi. The glycans on the surface of these entities are often the first interface with the host cell. As a result, targeting those carbohydrates can be useful to avoid infection. For instance, various vaccines based on the carbohydrates of the capsular polysaccharides of different bacteria have been developed. In these vaccines, fragments of the polysaccharides are conjugated to recombinant proteins to stimulate our immune system and thus generate memory antibodies against them⁴². As many viruses display a dense coat of glycans, another approach that has been proposed to battle pathogens is the use of glycan binding proteins^{43,44}. Fittingly, several glycan binding proteins have shown the ability of neutralising various viruses, including HIV, and therefore, these proteins are in the pipeline to be used to treat and prevent infections.

1.3 Protein-carbohydrate interactions

As the cell surface is surrounded by glycans, there are multiple entities that interact with them (Fig 12). For example, on the cell surface there are many glycan binding proteins that mediate cell-cell interactions⁴⁵. Some pathogens also interact with these glycans, as they can express glycan binding proteins on their surface to recognize and attach to the host cells. As pathogens also present glycans on their surface, host glycan binding proteins can also act as “bridges” between the host cell and the pathogen.

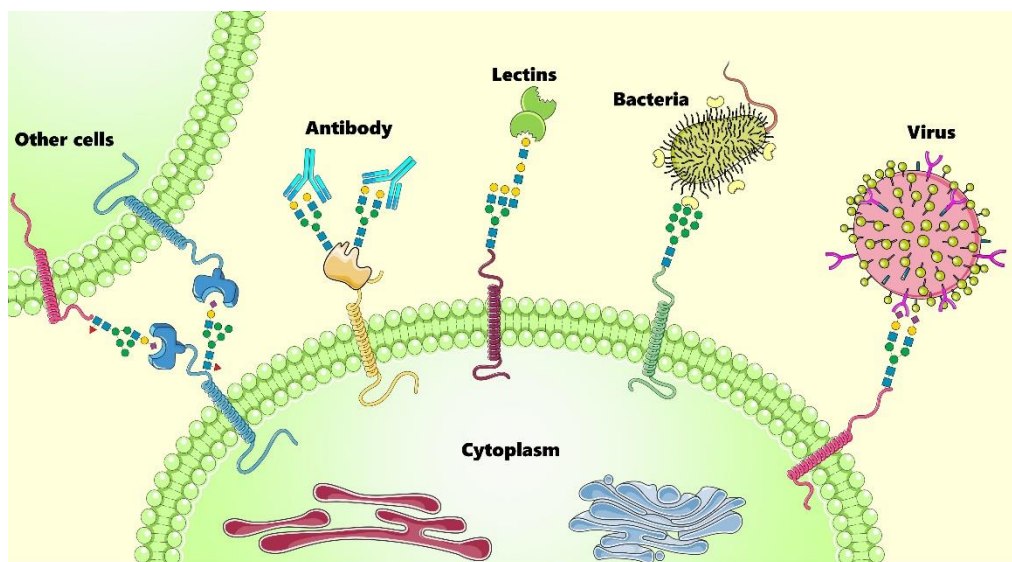


Figure 12. Typical interactions mediated through carbohydrates on the cell surface.

There are several types of interatomic forces that are involved in the binding events between proteins and carbohydrates (Fig 13). Carbohydrates are polyhydroxylated molecules that may form multiple hydrogen bonds (HB) with different amino acids of the protein (Fig 13-A). In principle, each hydroxyl group of the sugar can act as donor (with the hydrogen) and/or as acceptor

(twice, one with each electron pair of the oxygen) of hydrogen bonds. In addition, water molecules can also take part in the interaction by bridging the donors and acceptors at the protein and the sugar. These HB interactions may account for 1-2 kcal/mol.

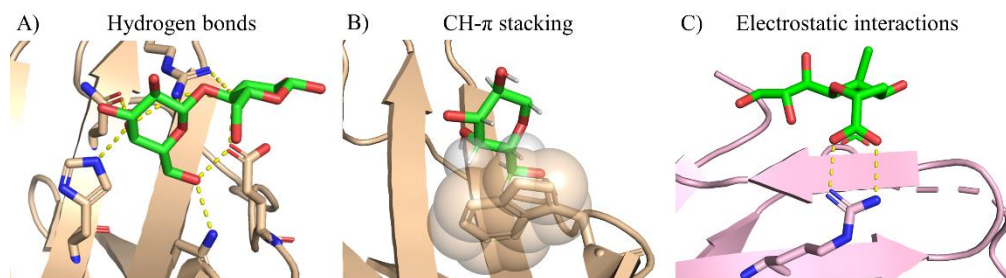


Figure 13. Typical interactions in protein-carbohydrate complexes. The structure used to depict interactions: A) and B) are taken from PDB 4YM0 (Galectin-4 complexed with lacto-N-tetraose). The structure for describing interaction C) is taken from PDB 2DF3 (Siglec-7 complexed with a sialylated glycan).

Hydrophobic interactions are also highly relevant. Even though carbohydrates are highly hydroxylated and polar compounds, they display amphiphilic nature, having less polar patches that can establish other types of interactions with the protein. These sites may participate in stacking interactions (CH- π stacking, ca. 1.5 kcal/mol) with the aromatic sidechains of tyrosine, phenylalanine, or tryptophan (Fig 13-B)^{46,47}. Aliphatic residues such as isoleucine or valine, with apolar sidechains, can also show weaker interactions (van der Waals) with non-polar regions of the sugars.

Electrostatic interactions can also stabilize protein-carbohydrate complexes. Sialic, glucuronic and iduronic acids, and sulfated carbohydrates can interact directly with positively charged residues of the protein or forming salt bridges (Fig 13-C), which may provide a strong stabilizing energy, beyond 2 kcal/mol.

Metal ions can also participate in these recognition events. For instance, in C-type lectins, in addition to maintaining the integrity of the binding site, Ca^{2+} ions are also necessary for sugar binding. The calcium cation located in the binding site normally exhibits two coordination bonds with two sugar hydroxyl groups and provides additional bridges to other four protein atoms.

Another relevant feature to take into consideration is the conformation of the carbohydrates. As stated before, glycans are rather flexible molecules. This feature can be, in many occasions, detrimental for the binding, due to high entropic penalties. Indeed, in certain systems, the glycan might need a specific conformation and presentation for the binding event to take place, which can result in high entropic penalties in flexible molecules. Thus, the design of synthetic molecules (glycomimetics) which already are preorganized for the binding may be a proper strategy to consider to target a biologically relevant sugar-binding protein⁴⁸.

Given all these structural and dynamic features, most protein-carbohydrate interactions are rather weak (K_D values in the mM- μ M range). In biological systems, however, this low-affinity binding is usually overcome through the engagement of simultaneous synergic interactions between the receptor and the ligand, a phenomenon known as multivalency.

Among the entities that interact with glycans, lectins are sugar binding proteins with no catalytic function (they are not enzymes) and do not provide a direct immune response (they are not antibodies). There are fourteen different types or group of lectins (Table 1) in the animal kingdom⁴⁹.

Table 1. Families of lectins found in animals. Taken from ⁴⁹

Animal Lectin Family	Typical carbohydrate ligands	Localization
Calnexin and Calreticulin	Glc ₁ Man ₉	Endoplasmic reticulum
M-type lectins	Man ₈	Endoplasmic reticulum
L-type lectins	Various	Endoplasmic reticulum, Golgi
P-type lectins	Man ₆ -phosphate	Secretory pathway
C-type lectins	Mannosides, galactosides, sialic acids and others	Membrane bound, extracellular
S-type lectins (Galectins)	β-galactosides	Cytosol, extracellular
I-type lectins (Siglecs)	Sialic acid	Membrane bound
R-type lectins	Various	Golgi, membrane bound
F-box lectins	GlcNAc ₂ of N-glycans	Cytoplasm
Fibrinogen-type lectin	GlcNAc, GalNAc	Membrane bound, extracellular
Chi-lectins	Chito-oligosaccharides	Extracellular
F-type lectins	Fucose terminating oligosaccharides	Extracellular
Intelectins	Galactose, galactofuranose, pentoses	Membrane bound, extracellular
Annexins	GAGs, heparin and heparin sulphate	Membrane bound

Three of the most relevant lectin families found in humans are C-type, I-type and S-type lectins. C-type (Ca²⁺-dependent) lectins are found both as transmembrane and as soluble proteins. Some of the lectins of this family, like DC-SIGN, Langerin, and MGL have key roles in pathogen recognition and have become targets in the field of drug discovery⁵⁰.

Within I-type lectins, the study of Sialic Acid Binding Immunoglobulin-Type Lectins (Siglecs) is nowadays a topic of major interest⁵¹. The Siglec family of transmembrane lectins is comprised by 15 members, which contain an N-terminal domain that recognizes sialic acids¹⁰. All siglecs (except for Siglec-4 and 6) are expressed in immune cells and help our immune system to distinguishing between self and non-self signals.

Galectins (earlier dubbed as S-type lectins) are β-galactoside binding proteins. This family is formed by 16 members, which are found ubiquitously in the human body⁵². These lectins are expressed in the cytoplasm and then secreted

through a non-classical pathway or transported to the nucleus. Through their ability of binding β -galactosides, galectins participate in cell-cell interactions and they are also involved in immune responses, inflammation and signalling events, among many others.

Galectins can be classified in three groups according to the organization of their carbohydrate recognition domains (CRDs) (Fig 14). The first group, the prototype galectins, contain two identical CRD that are linked through non-covalent interactions. Galectins 1, 2, 5, 7, 10, 11, 13, 14, 15, and 16 belong to the prototype group. The second group, which is formed by galectins 4, 6, 8, 9 and 12, are the tandem repeat galectins. They contain two CRDs that are united by a short linker peptide. Different to the prototype group, tandem-repeat galectins are heterodimers: the two CRDs have different amino acid sequences. Finally, Gal-3 is the only member of the chimera type galectins, which contains a single CRD united to a non-lectin N-terminal domain.

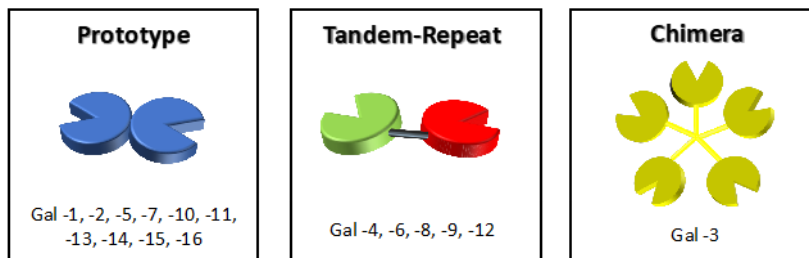


Figure 14. The three subgroup of galectins and the members of each group. Taken from ⁵³

The CRDs of galectins have a conserved bent β -sandwich structure (Fig 15), formed by six (S1 to S6) and five (F1 to F5) antiparallel strands, in which glycans are recognized in the concave side. Several amino acids located in the binding site are key to recognize β -galactosides and are conserved for all galectins.

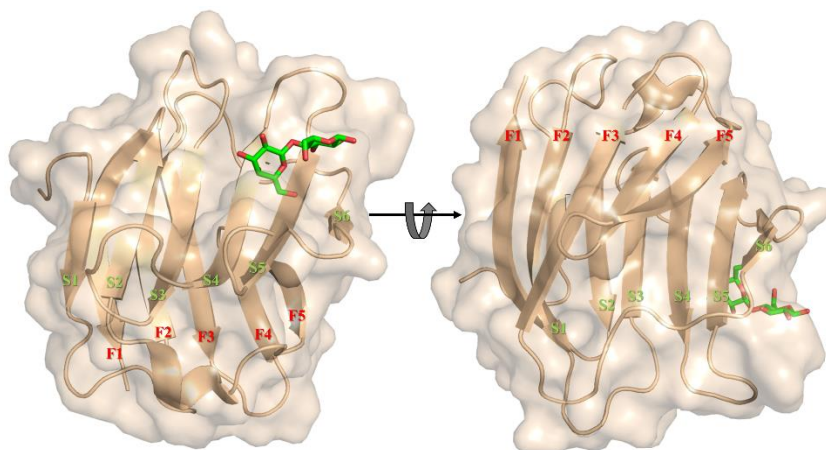


Figure 15. Top and bottom views of the X-Ray structure of *h*Galectin-4-N-terminal domain in complex with lactose (PDB 5DUV). The name of the strands are coloured in green (S strands) and red (F strands)

As shown in Fig 16, all galectins show a conserved tryptophan residue whose aromatic moiety interacts with H4, H5 and H6 of β -galactosides through CH- π stacking interactions. Located at the strand S4, there are histidine, asparagine and arginine residues that mediate key hydrogen bonding interactions with the hydroxyl group at position 4 of the bound galactose. At the strand S5, there is a conserved asparagine that interacts with the O6 hydroxymethyl. The glutamic acid at strand S6 is also conserved, and despite not interacting with the galactose, it is relevant for the interaction with a strategically positioned oxygen atom at its β 1-linked residue.

	21	29	44 46 48	59 61	68 71 73	79																			
hGal-1	VRGEVA-PDAKSFVNLGKDS-----NNLCLHFNPRFNAHGDANTI	V	C	N	S	K	D	G	G	A	W	G	T	E	Q	R	E	--AVFPFQP							
hGal-2	ITGSLA-DGTDGFVNLGQGT-----DKLNLHFNPRFSE----	S	T	I	V	C	N	S	L	D	G	S	N	W	G	T	E	Q	R	E	--DHLCPSP				
hGal-3	ILGTVK-FMANRIALDFQRG-----NDVAFHFNPRFENN--RRVIVC	N	T	K	L	D	N	N	W	G	T	E	R	Q	--SVFPFES										
hGal-4-Nter	IQQVAS-EHMKRFVNFVVGQ--DPGSDVAFHFNPRFDGWD---	K	V	V	F	N	T	L	Q	G	G	K	W	G	T	E	R	K	--RSMPPFK						
hGal-4-Cter	IKGYVP-PTGKSFAINFKVGS---SGDIALHFNPRMGNT---	V	V	N	S	L	L	N	G	S	W	G	T	E	R	K	I	--THNPFQP							
hGal-7	IRGLVP-PNARSFHVNLGCE--EQGSDALHFNPRLDTS---	E	V	V	N	S	K	Q	G	S	W	G	T	E	R	Q	--PGVFPQR								
hGal-8-Nter	IRGHVP-SDADRFQVDLQNGSSMKPRADVAFHFNPRFKRAG---	C	I	V	C	N	T	L	I	N	E	K	W	G	T	E	R	I	--YDTPFKR						
hGal-8-Cter	VKGEVN-ANAKSFNVDLLAGK---SKDIALHFNPRLNKA---	F	V	N	S	F	L	Q	S	W	G	T	E	R	N	I	--TSFPFSP								
hGal-9-Nter	VNGTVLSSSGTRFAVNFQTG---FSGNDIAFHFNPRFDGG---	Y	V	V	C	N	T	R	Q	N	G	S	W	G	T	E	R	K	--THMPFQK						
hGal-9-Cter	LSGTVL-PSAQRFHINLCSG-----NHIAFHFNPRFDENA---	V	V	R	N	T	Q	I	D	N	S	W	G	T	E	R	S	L	P	R	K	M	P	F	V

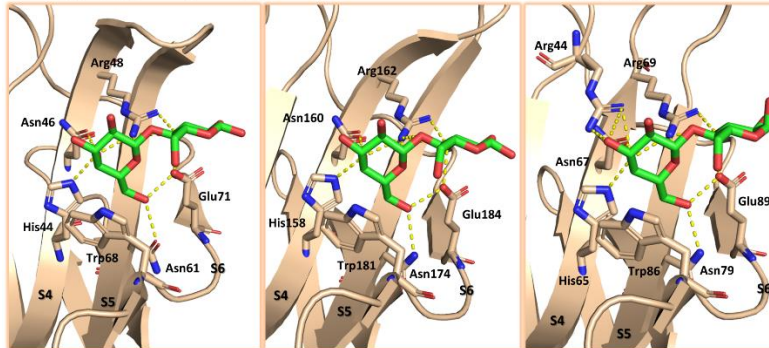





















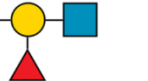















Figure 16. Top: Sequence alignment of the binding site of different galectins. Bottom: X-ray crystallographic structures of: Left: Galectin-1 (PDB 1GZW), center: Galectin-3 (PDB 1NN8), and right: Galectin-8-Nterminal (PDB 5T7S) in complex with lactose. The residues that participate in the binding event are displayed as sticks, while hydrogen bonds are displayed as yellow dashed lines. Taken from ⁵³

Additionally, there are significant differences in some amino acids located in the binding sites, which are responsible of dictating the specific binding preferences for each galectin (Table 2). For example, Arg44 located at the strand S3 of Gal-8N (See Fig 16-C) favours the interaction of this domain with sialylated β -galactosides⁵⁴.

Table 2. Galectins and their preferred ligands⁵⁵⁻⁶³. On the bottom the symbol representation of each ligand is noted.

Protein	Preferred Ligands
Galectin-1	
Galectin-3	 /  /       
Galectin-4 N-CRD	
Galectin-4 C-CRD	
Galectin-7	 /  /   
Galectin-8 N-CRD	
Galectin-8 C-CRD	
Galectin-9 N-CRD	
Galectin-9 C-CRD	 / 

LacNAc =  3'sialyl LacNAc =  LacdiNAc = 
 PolyLacNAc =     
 Blood Group Antigens =  Lacto-N-neotetraose =    

Due to their involvement in several diseases, galectins have been targeted for inhibitors development, for which different approaches have been used. Glycomimetics⁵³ have been employed, ranging from monovalent molecules with multiple chemical decorations to relative simple molecules displayed along multivalent scaffolds. One of the main drawbacks in developing mimetics for galectins is the difficulty of designing a molecule that is specific for just one galectin. However, there are various promising molecules that are fairly specific for Gal-1⁶⁴, Gal-3⁶⁵ and Gal-8⁶⁶.

In this Thesis, we have paid special attention to Galectin-4 (Gal-4). This lectin was first isolated in 1989 from rat intestine as a 17 kDa protein⁶⁷. Later, this protein was identified as a proteolytic fragment of a 36 kDa protein with two CRDs⁶⁸. Indeed, Galectin-4, which belongs to the tandem-repeat galectin group, is mainly expressed in the alimentary tract and it has been found intracellularly, extracellularly, and at the cell surface.

The two domains that recognize carbohydrates of *hGal-4* show a sequence identity of 36%. Their 3D structures in the presence of various ligands (Table 3) have been solved by X-ray crystallography, as well as the crystallographic structure of the apo N-terminal CRD.

Table 3. X-ray structures of the N- or C-terminal domains of Galectin-4 and its sugar complexes deposited on the PDB⁶⁹⁻⁷¹.

Protein – Ligand complex	PBD Code	Protein – Ligand complex	PBD Code
Gal-4C – Galactose derivative	6WAB	Gal-4C - Lacto-N-neotetraose	4YLZ
Gal-4C – Lactose	5CBL	Gal-4N - Apo	4XZP
Gal-4C – Lactose	4YM3	Gal-4N – Lactose	5DUV
Gal-4C – Lactose-3'-sulfate	4YM2	Gal-4N – Lactose-3'-sulfate	5DUW
Gal-4C – 2'Fucosyllactose	4YM1	Gal-4N – 2'Fucosyllactose	5DUX
Gal-4C – Lacto-N-tetraose	4YM0	Gal-4N - Glycerol	5DUU

Structurally, both domains have a similar molecular weight (16-17 kDa) and shape, with the typical β -jellyroll fold and shallow binding site of galectins (Fig 17-A). However, the charge distribution of the domains is different. Whereas the C-terminal domain (Gal-4C, amino acids 179-323) displays a positively charged surface, that of the N-terminal CRD (Gal-4N, amino acids 1-150) is more heterogeneous, although with a positively charged region in the binding site⁷⁰.

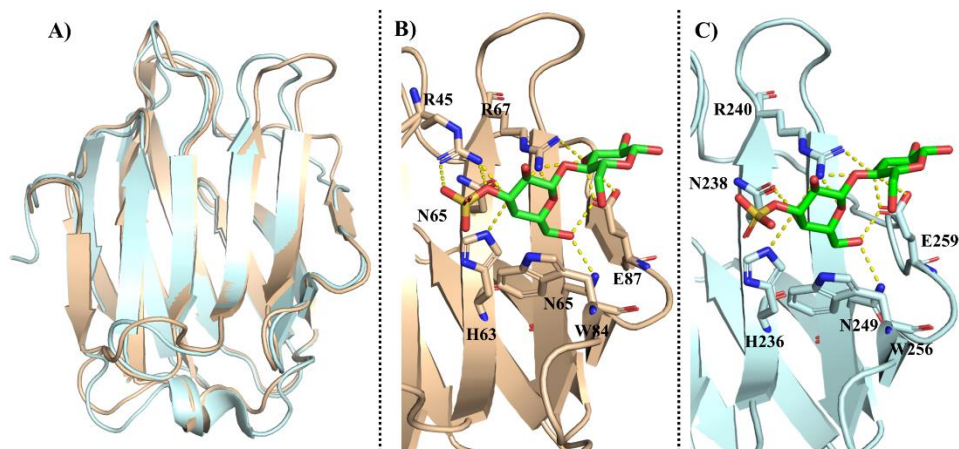


Figure 17. A) Superimposition of the X-ray crystallographic structures of Gal-4N (PDB 4XZP) and Gal-4C (PDB 4YM2). B) Structure of Gal-4N (5DUW) in complex with 3'sulfated lactose. C) Structure of Gal-4C in complex with 3'sulfated lactose (PBD 4YM2). In figures B and C, the residues of the lectin that participate in the binding are shown with sticks. The hydrogen bonds are depicted with yellow dots.

Interestingly, most of the amino acids at the binding sites of Gal-4N and Gal-4C are the same. Nevertheless, there are some relevant differences that are relevant for binding to specific ligands. In particular (Fig 17-B and C), residues His63/236, Asn65/238, Arg67/240, Asn77/249, Trp84/256 and Glu87/259 are conserved and provide the same intermolecular interactions. However, Gal-4N has an arginine residue at the strand S4 that directly interacts with a negatively charged residue (sulfate) at position Gal O3, whereas for Gal-4C, there is no arginine at that position.

Compared to other galectins, both domains of Gal-4 show relatively weak binding to lactose (both binding affinities are in the mM range), which is the minimum binding epitope for galectins. Notwithstanding, Gal-4 is able to exert various biological functions given its ability to bind β -galactosides. The binding preferences of each domain remain controversial, although it has been described that both domains have preferences for blood group antigens^{69,72-74}.

Moreover, the natural ligands of this lectin include N- and/or O-glycosylated proteins and glycosphingolipids^{75–80}.

In the intestinal tract, Gal-4 is able of cross-linking ligands. For instance, Gal-4 in lipid rafts participates in its stabilization by forming lattices with glycolipids and glycoproteins⁷⁵. Additionally, Gal-4 participates in apical protein trafficking as seen in Gal-4 knockdown HT-29 5M12 cancer cells^{76,77}. Whereas in cells containing Gal-4, the apical glycoproteins are effectively transported to the apical plasma membrane, in the absence of the lectin, some apical proteins, including mucin-1, could not be transported and are accumulated intracellularly.

Nevertheless, the mechanisms behind some roles of Gal-4, including its ability to promote intestinal wound healing⁸¹, remain unknown. Although the mechanism has not been elucidated yet, Gal-4 is able to bind to intestinal epithelial cells and to promote their healing.

Additionally, in neurons, Gal-4 regulates myelination^{78,80} and promotes the growth of axons⁷⁹. Indeed, myelination is a necessary process for the protection and function of the neurons. Oligodendrocytes are essential to deposit the layers of myelin around axons, which is disposed in sheaths. Between these sheaths, there are gaps, known as Nodes of Ranvier, in which there is no myelin. These nodes are relevant as the nerve impulses are speeded up there. Neurons are able to regulate the expression levels of Gal-4 to define these nodes, since Gal-4 repels myelin deposition by interacting with the glycoprotein contactin-1.

The interaction with another glycoprotein, the neural cell adhesion molecule L1, is also relevant for the growth of axons. During the elongation of axons, L1 plays a crucial role by producing intracellular signals to regulate axonal growth⁸². However, in order to achieve its proper function, L1 needs to be transported to axons. Interestingly, in Gal-4 knockdown neurons, the growth of the axon is retarded, an effect that can be reversed upon addition of exogenous

Gal-4. The interaction between L1 and Gal-4, which is mediated through the binding of Gal-4 to the LacNAc moieties on the N-glycans of L1, enables the transport of L1, likely by crosslinking it to other glycoproteins.

The role of Gal-4 in cancer is not completely understood since it plays contradictory roles; whereas in some tumours Gal-4 acts as a tumour suppressor, in others it promotes metastasis⁸³. Its expression also varies among cancers. In colon and acute myeloid leukaemia, its expression is downregulated^{84,85}. In contrast, in other type of cancers, as colorectal, pancreatic, and breast cancer, it is upregulated^{86,87,88}. Very intriguingly, Helwa *et al* have shown that colon cancer cells (KM20L2), which express high levels of Gal-4, interact with red blood cells (RBC) and that Gal-4 is accumulated at the interface between the tumour cells and the RBCs (Fig 18)⁸⁹. Since it is well known that Gal-4 is able to bind blood group antigens, these experimental data suggest that the interaction is mediated by crosslinking of the bivalent lectin. This phenomenon also occurred with pancreatic cancer derived cells with high expression of Gal-4 (ASPC-1 and HPAF-2)⁸⁹. However, in normal pancreatic cells, with low expression of Gal-4, no aggregation of RBC takes place. The data suggest that the interaction of Gal-4 with blood group antigens could promote metastasis. Therefore, targeting Gal-4 could help to reduce the development of the cancer.

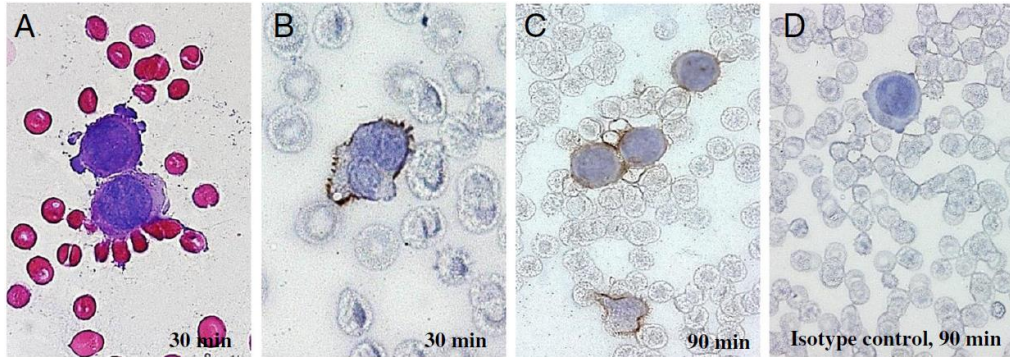


Figure 18. Interaction of tumour cells with high expression of Gal-4 (KM20L2 cells) and red blood cells. A) Hematoxylin and eosin stain. B) Immunohistochemical (IHC) staining of Gal-4 after 30 minutes of incubation. C) IHC staining of Gal-4 after 90 minutes. D) IgG isotype control. Taken from ⁸⁹.

Based on this same ability of recognizing blood group antigens, Gal-4 (as well as Gal-8) shows bactericidal activity^{90,91}. There are multiple pathogens that mimic human glycans to evade our immune system, some of them mimicking the blood group antigens⁹². The O-antigen of the *Escherichia Coli* O86 bacterium is very similar to the blood group B antigen. It has been reported that the C-terminal CRDs of Gal-4 and Gal-8 bind to this bacterial glycan and degrade its membrane, leading to the death of the pathogen (Fig 19). Bacteria killing occurs through glycan recognition, as the presence of lactose in the media avoids the bactericidal activity. Additionally, this activity is specific to these two lectins, since the presence of galectins 1 or 3 does not affect to the bacteria.

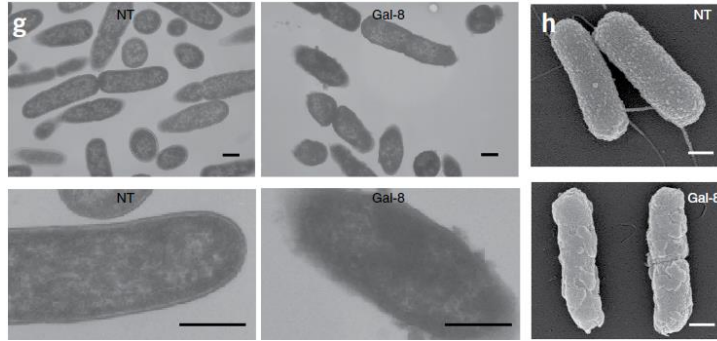


Figure 19. *Escherichia coli* O86 bacteria free and in the presence of Galectin-8. Taken from 90.

This capacity of galectins of killing pathogens is very interesting and could play an important role in our immune system. Since we hold antibodies against the blood group determinants that do not belong to our own blood group, we are protected against those pathogens that mimic another blood groups. However, if a pathogen mimics our own blood group, there is no specific antibody against it. In this gap of the immune system, galectins could protect us from the pathogens. Very recently, it has also been described that another tandem repeat galectin, Galectin-9, also displays antimicrobial activity against pathogens presenting blood group-like antigens⁹³.

Related to this bactericidal activity, it has been reported that Gal-4 interacts with CD14⁹⁴, a glycoprotein that detects bacteria through binding to their lipopolysaccharides (LPS)⁹⁵. The interaction of Gal-4 with CD14, which is mediated by the recognition of the glycans on CD14, induces significant changes on CD14⁺ monocytes. This interaction triggers the differentiation of monocytes into macrophage like cells (Fig 20), alters the expression of several proteins, and increases the cytokine levels. Although not exactly the same, the addition of Gal-4 to monocytes induced a similar effect to that produced by LPS.

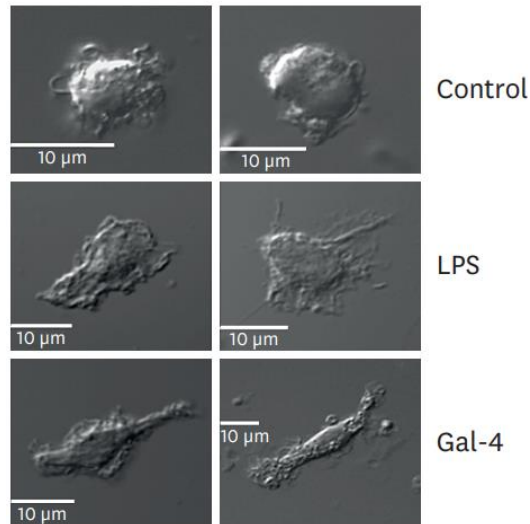


Figure 20. CD14⁺ monocytes in different conditions. Top panel: the control monocytes. Middle panel: after incubation with LPS. Bottom panel: after incubation with Gal-4. Taken from ⁹⁴.

Although the role of Galectin-4 in bacterial infection is interesting, the structural details of how the interaction takes place between Gal-4 and LPS, Gal-4 and CD14 and the interplay between these three entities is yet unknown.

1.4 Experimental techniques to characterize protein-glycan interactions

There are various techniques that can be applied in order to characterize protein-carbohydrate interactions. X-ray crystallography has been for decades the most employed biophysical technique for unravelling the structure of lectin-sugar complexes, as atomic resolution information of the complex can be obtained⁹⁶. However, the process of obtaining crystals can be very tedious and the proteins are not always crystallisable. Additionally, due to the intrinsic flexibility of glycans, crystallographic structures usually have erroneous conformations, linkages, or even wrong residues^{97,98}. Cryo electron microscopy (CryoEM) is also becoming an important structural biology technique⁹⁹. Indeed, there are already different complexes that have been evaluated through cryoEM¹⁰⁰.

Other biophysical techniques, as biolayer interferometry (BLI) or surface plasmon resonance (SPR) have also been widely employed to monitor protein-carbohydrate interactions^{101,102}. These techniques yield the kinetics and binding energy of the interaction and can be performed without any type of labelling. Another powerful label-free technique is isothermal titration calorimetry (ITC), which enables obtaining the thermodynamic profile of the binding event¹⁰³. These methods, which are extremely useful, provide key energy and thermodynamic data, but no direct information on the epitope and the paratope of the binding is obtained.

Nuclear Magnetic Resonance (NMR) has also been extensively used to study protein-carbohydrate interactions. Through NMR, the binding affinities and information on the epitopes can be obtained. Moreover, the 3D structure of the protein and/or the glycan ligand can also be deduced¹⁰⁴. Additionally, due to the intrinsic chemical properties of the glycans, NMR has been the technique of choice of many research groups to analyse the structure, conformation, and

dynamics of carbohydrates, as well as their interactions with biomolecular receptors, including proteins^{105,106}.

1.4.1 Nuclear Magnetic Resonance

1.4.1.1 Principles of Nuclear Magnetic Resonance

The physical phenomenon of Nuclear Magnetic Resonance is a consequence of the existence of nuclear spin (n)¹⁰⁷. Not every nuclei has a spin, since only the nuclei with odd atomic number or mass have it. These nuclei have a magnetic moment (μ) that is specific for each nuclei. The magnetic moment of a nucleus is proportional to its spin quantum number (I), Planck's constant (h), and the gyromagnetic ratio (δ), a property intrinsic of each nucleus (Eq 1)¹⁰⁸.

$$\mu = \frac{\gamma I h}{2\pi}$$

Equation 1

Upon the presence of a strong external magnetic field (B_0), the magnetic moment of the nuclei will try to align with this field¹⁰⁹. The possible orientations of the magnetic moment are limited and depend on the spin quantum number (I). The majority of the NMR-active nuclei (^1H , ^{13}C , ^{19}F , ^{31}P) show an I value of $\frac{1}{2}$. Quantum mechanics show that the possible orientations of a nucleus upon a magnetic field are $2I+1$. Thus, these nuclei may display 2 orientations, aligned or against the magnetic field. The spin aligned with the external magnetic field (α) is energetically more stable than that against the magnetic field (β), and therefore in equilibrium, this state is more populated (Fig 21).

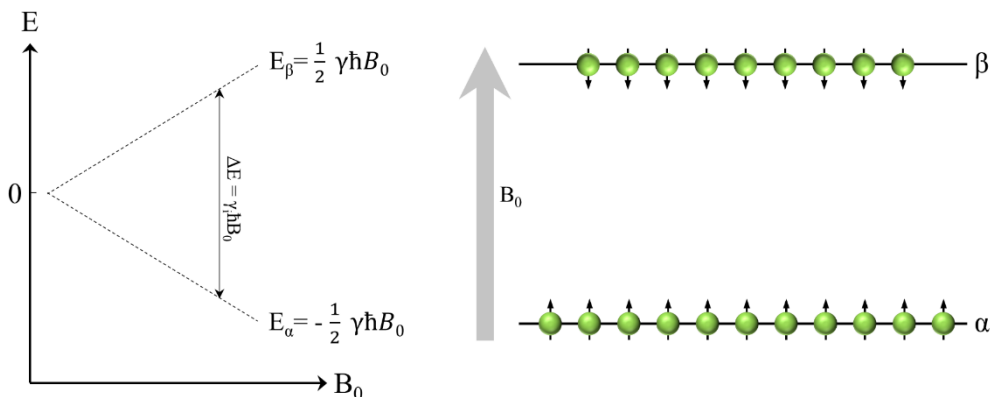


Figure 21. Left: Plot of the energy levels of α and β states over the external magnetic field (B_0). Right: Depiction of how the spins align with or against the external magnetic field.

The population of each state is described by a Boltzmann distribution, and the energy difference between these states determines the frequency of radiation required to promote a transition between them. This frequency is dubbed the Larmor precession frequency. This energy difference is proportional to the gyromagnetic ratio, Planck's constant and the external magnetic field (Fig 21 and Eq 2).

$$\Delta E = \gamma \hbar B_0$$

Equation 2

The electrons surrounding the nuclei also affect the Larmor frequency. They create a small magnetic field opposing the external magnetic field, known as shielding. Since the environment of the electrons around each nucleus is different, the diverse nuclei suffer a different shielding and thus display distinct Larmor frequencies. Since the differences in frequencies of each nuclei are very small, instead of the Larmor frequency of a nucleus, its chemical shift (δ) is used. The chemical shift for a given nucleus depends on its chemical environment, and can be calculated following equation 1.3.

$$\delta = \frac{\text{Frequency of the signal} - \text{Frequency of the reference}}{\text{Frequency of the spectrometer}} \times 10^6$$

Equation 3

Usually, the signal corresponding to the Larmor frequency of a nucleus is not a single peak. This is due to the presence of spin-spin coupling, known as J-coupling¹¹⁰. When a nucleus A is coupled to the nucleus B, A will feel the presence of the magnetic field of B, which perturbs its Larmor frequency. For instance, if A and B are protons, i.e. the only possible states are alpha or beta, the peak for A will split into two equal peaks, one corresponding to the α state of B and another one to the β state of B (Fig 22).

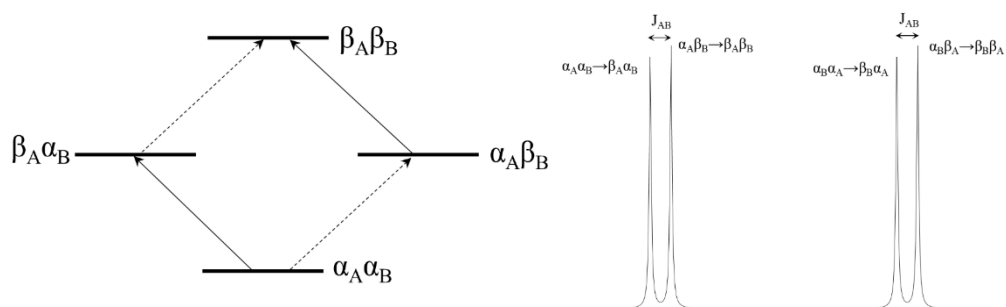


Figure 22. A) Energy levels and possible transitions of nuclei A and B when they are spin-spin coupled. B) Corresponding spectrum of the system of nuclei A and B.

The distance between the split signals is defined as J , the coupling constant. When two nuclei have a spin-spin coupling, the magnitude of the splitting is the same for both. For carbohydrates, the protons directly attached to a carbon usually have one or more vicinal protons (at 3 bonds of distance) and thus show one or more J-couplings (3J). In 1959, Martin Karplus discovered that the magnitude of this coupling constant depends on the dihedral angle between the protons, and defined the so-called Karplus equation (Eq 4)¹¹¹. Additionally, the

chemical nature of the substituents close to these nuclei affects the J-coupling values. This dihedral angle dependence is very useful for sugars.

$$J(\theta) = A \cos^2 \theta + B \cos \theta + C$$

Equation 4

Where θ is the dihedral angle and A, B and C are parameters which depend on the substituents of the molecule.

Another relevant NMR phenomenon is the Nuclear Overhauser Effect (NOE). Solomon described the NOE principles for the first time in 1955¹¹². However, it was not until 1965 that Anet and Bourn showed how the NOE could be used to determine the structure of small organic compounds¹¹³. This was possible as NOEs are strongly dependent on the distance between the two nuclei (Eq 5). The intensity of the NOE (I) is inversely proportional ($I \propto \frac{1}{r^6}$) to the distance between the two nuclei (r) and, as a result, it can be employed to obtain precise distances between them, enabling the derivation of the 3D-structure of molecules¹¹⁴.

$$I_{\text{NOE}} = \frac{1}{r^6} f(t_c)$$

Equation 5

Where f is a function that depends on the correlation time (t_c) of the molecule.

In order to understand the NOE, let's imagine two spins (A and B) that are close in space, showing the spin states shown in Fig 23-A. If spin B is selectively saturated using a specific radiofrequency, its α and β energy levels (α_B and β_B) will be equally populated (Fig 23-B). When the saturation stops, the spins will relax and the populations of the energy levels will come back to equilibrium.

This phenomenon can occur through different type of transitions. In fact, for this two spin system, relaxation can take place through two mechanisms, using either double-quantum (DQ) or zero-quantum (ZQ) transitions (Fig 23-C). In the DQ transition (Fig 23-D), the β states of both spins (A and B) will turn into α ($\beta_A\beta_B \rightarrow \alpha_A\alpha_B$). In the ZQ alternative, the β state of B will turn to α and the α of A will turn into β ($\alpha_A\beta_B \rightarrow \beta_A\alpha_B$) (Fig 23-E). These transitions generate changes in the populations of the α and β states of spin A, which are translated into variations in the intensity of the signal of A¹¹⁵. When DQ transitions predominate, since α_A state is now more populated than β_A compared to the starting equilibrium, the intensity of the signal for A will be higher, which is known as positive NOE. When ZQ transitions predominate, following a similar reasoning, negative NOEs take place. The probabilities of these transitions are related to the tumbling of the nuclei, which is related to the size of the molecules. In small molecules (MW<600 Da), which display fast motions with high frequencies, DQ double quantum transitions predominate, whereas in larger molecules, with slow motion and very low frequencies, zero quantum transitions are predominant.

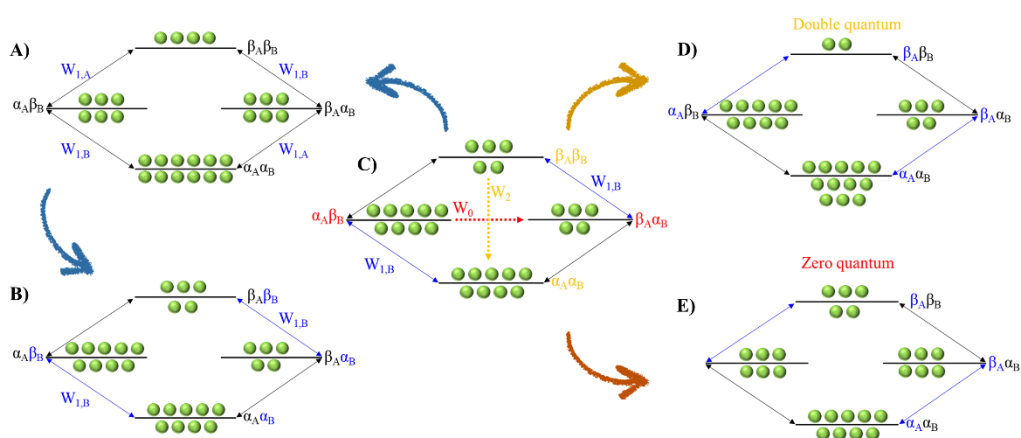


Figure 23. Spin states, energy levels and transitions of a system in which the nuclei A and B are close in space, conducting to the generation of NOE. A) The energy levels and its populations in equilibrium. B) The spin B is selectively saturated (blue transitions). C) When saturation is finished, the system will come back to equilibrium, for which there are 2 key

mechanisms: double-quantum (DQ, yellow) or zero-quantum (ZQ, red) transitions. D) Energy populations after the DQ mechanism. E) Energy populations after the ZQ mechanism.

1.4.1.2 NMR to study protein-carbohydrate interactions

The NMR methods employed to analyse interactions are classified into two groups. In ligand-based methods, changes in the NMR signals of the ligand (the glycan, in this Thesis) are observed, whereas in receptor-based methods changes in the NMR signals of the macromolecule (the lectin, in this Thesis) are monitored¹¹⁶.

1.4.1.2.1 Ligand-based NMR methods

Ligand-based NMR methods to characterize protein-carbohydrate interactions rely on the changes on the NMR properties of the glycan¹¹⁷. These variations can be observed through two strategies: i) taking advantage of the dramatic changes on the molecular motion of the ligand upon binding to the macromolecule or ii) through transfer of magnetization from the NMR signals of the macromolecule to those of the glycan.

For a small or medium size glycan, as depicted in Fig 24, its motional properties (Fast Brownian motion, slow relaxation, fast diffusion, and positive NOE values) are different to those of the receptor (Slow Brownian motion, fast relaxation, slow diffusion, and negative NOE values). However, when the glycan binds to the protein, its rotational motion properties change, and are similar to those of the large macromolecule.

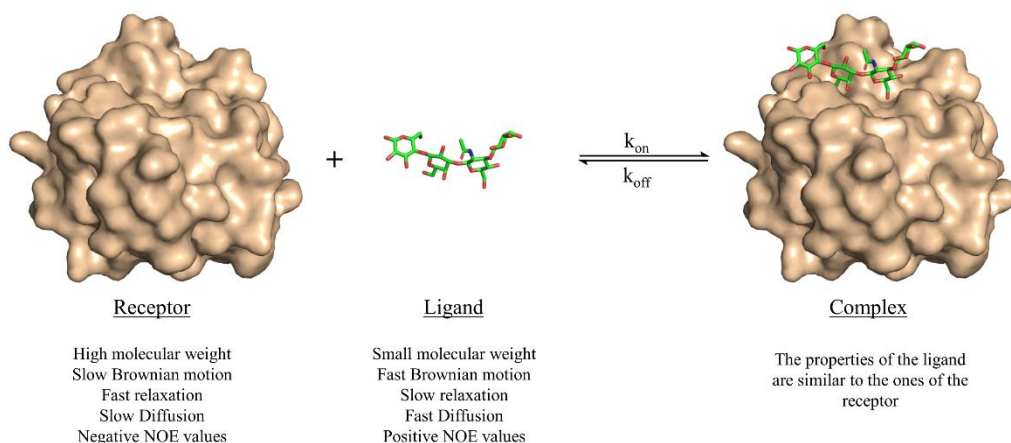


Figure 24. Motional properties of a receptor, a ligand, and the corresponding binary complex. The structures are taken from the PDB 4YM0.

i. The transferred NOE

One of the most significant NMR-related changes that carbohydrates may suffer upon protein binding is the change in the sign of the NOE. As explained above, the NOE intensity and sign depends on the tumbling rate, which is related to the molecular size (Chapter 1.4.1.1). For the free small sugars, the NOEs are positive or close to zero. However, when bound to the protein (using usually ca. 1:5 to 1:20 protein:glycan ratio) the rotational properties of the glycan are similar to those of the macromolecule, and the corresponding NOEs are negative (Fig 25). Therefore, the change in sign of the NOE can be exploited to characterize protein-sugar interactions^{118,119}.

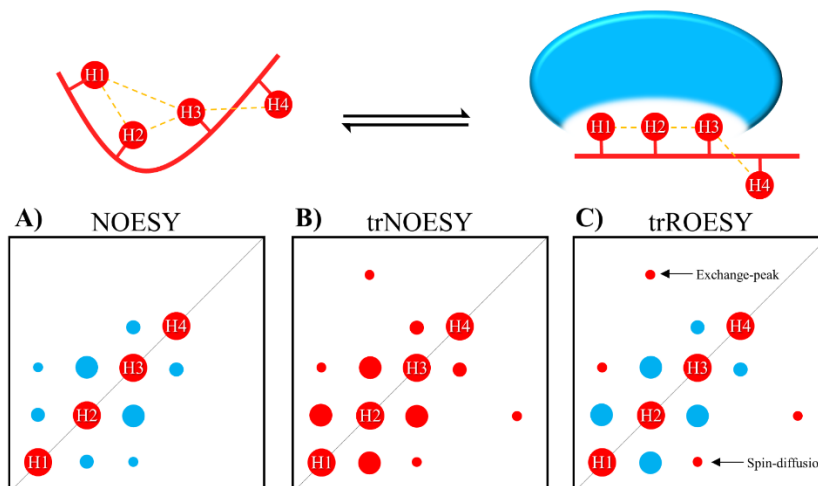


Figure 25. Top: A ligand and its receptor in equilibrium. Bottom: A) NOESY spectrum of the free ligand B) NOESY spectrum of the ligand bound to the protein (trNOESY). C) ROESY spectrum of the ligand in the presence of the protein (trROESY).

The signals obtained in the trNOESY spectrum are an average of the contribution of the free and bound states of the ligand. However, the absolute value of the NOEs for a large molecule are much more intense than those of the free molecule, and therefore, the contribution of the free ligand is minor, even if it is in excess. The election of the mixing time of the NOESY is also very relevant, as exemplified by Weimar and Peters when studying the interaction of α -Fuc-(1-6)- β -GlcNAc-OMe with *Aleuria aurantia* agglutinin¹²⁰. Whereas for the free disaccharide, NOEs are positive and small, for the complex, NOEs are negative and ca. ten times higher in absolute values (See the dots in Fig 26). Transferred NOESY (trNOESY) experiments are usually performed with short mixing times (ca. 100 ms). Under these conditions, the contribution of the free ligand is almost negligible, meaning that the obtained information basically describes the conformation of the bound ligand. In general, for small molecules, long mixing times are needed (600-1000 ms) to obtain optimum intensities.

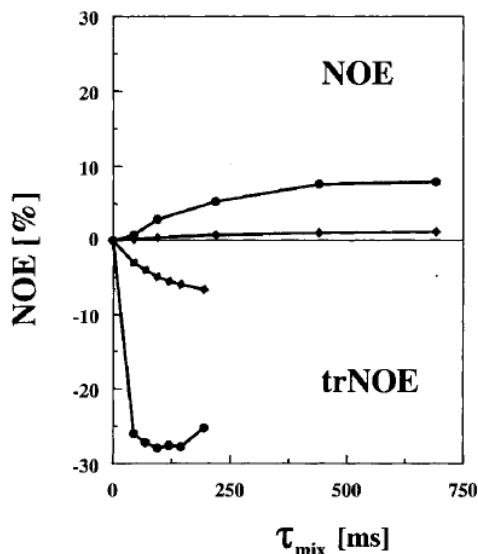


Figure 26. Plot of the NOE intensity at different mixing times. The NOE intensities of the α -Fuc-(1-6)- β -GlcNAc-OMe are shown. The round dots represent the NOEs between H6^R and H6^S of the GlcNAc. The diamonds represent the NOE between H1Fuc and H6^SGlcNAc. Top: NOEs for the free ligand. Bottom: NOEs for the ligand in the presence of *Aleuria aurantia* agglutinin. Taken from ¹¹⁹

However, NOEs for larger glycans are already negative, which can bias the analysis of the bound conformation. In order to avoid this issue, diverse approaches can be applied. For instance, very short mixing times (25-50 ms) can be used to minimize the contribution of the free state. Another alternative is to increase the temperature of the system, so that the free ligand tumbles faster. If available, the use of medium field NMR spectrometers can be advantageous, since for lower field magnets, slower tumbling rates are needed to generate negative NOEs, i.e. a close proton-proton contact could produce a negative NOE at 800 MHz and a positive one at 400 MHz. However, this has a cost in terms of sensitivity¹²¹.

One of the drawbacks of trNOESY experiments is the existence of spin-diffusion^{122,123}. Due to the fast transverse relaxation rate of large molecules, the magnetization that is transferred during the mixing time from an initial nucleus to another one close in space can be easily transferred from the second nucleus

to a third one, yielding cross peaks between the first and the third nucleus, even if they are not very close in space. These peaks are indistinguishable from the direct NOE in trNOESY experiments, which may lead to incorrect conclusions. Nevertheless, this three-spin effect can be distinguished in ROESY experiments. In ROESY experiments, which give rise to positive cross peaks for directly related proton pairs, these three-spin mediated signals are negative (Fig 25-C). Additionally, the existence of chemical exchange events between the free and bound species can also be assessed through ROESY experiments, since chemical exchange peaks are always negative (Fig 25-C)¹²⁴.

Thus, the conformation of the bound ligand can be elucidated through trNOESY experiments. Although no information of the protein is obtained, in some occasions, direct intermolecular NOEs between the protein and the ligand can also be observed, which can be used to decipher the location of the binding site^{125,126}.

ii. Diffusion Ordered Spectroscopy

Diffusion Ordered Spectroscopy (DOSY) experiments can also be used to monitor protein-carbohydrate interactions¹²⁷. DOSY experiments are pseudo-2D experiments, which relate the diffusion coefficient, in the Y axis, to the regular ¹H NMR spectrum (chemical shifts) in the X axis. Since the diffusion properties of a molecule depends on its molecular weight, size and shape, its diffusion coefficient will be different in the free and protein-bound states (Fig 27)^{128,129}. Thus, through DOSY, binding of ligands to proteins can be easily monitored.

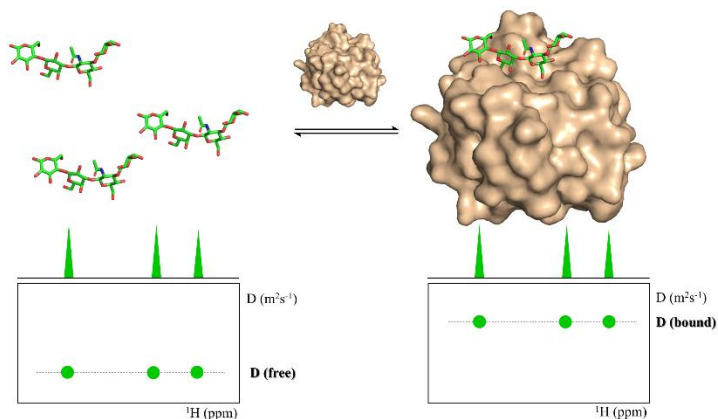


Figure 27. DOSY spectra of a free ligand (left) and of the ligand in the presence of its receptor (right).

iii. Saturation Transfer Difference-NMR

Saturation Transfer Difference (STD-NMR) is probably the most employed NMR techniques to study protein-ligand interactions¹³⁰. Through STD-NMR experiments, a clear picture of the “binding epitope” of the ligand towards the target receptor can be deduced. The binding epitope defines the region (protons in this context) of the ligand that is spatially close to the protein.

In STD-NMR, two different spectra are recorded. In the first one, a selective saturation is applied in a region devoid of any NMR signal, usually 100 ppm. This is the reference spectrum, defined as off-resonance. A second spectrum is recorded, in which only protons of the protein are selectively saturated. This spectrum is the on-resonance spectrum. As the protein is saturated, the magnetization is rapidly spread throughout the polypeptide chain protons. Due to its size, relaxation and spin-diffusion effects (i.e. NOE) are very efficient. Fittingly, if a given ligand binds to the saturated protein, the ligand will also receive this magnetization (Fig 28). As a result, the signals of those protons that are closer in space to the protein will suffer a decrease in their intensities. Thus, the on- and off- resonance spectra will be different. The subtraction of the on-

resonance spectrum to the off-resonance one will result in a spectrum in which only the signals of those protons that are close to the protein will be present. NMR signals corresponding to protons of non-binders or far from the protein protons will display the same intensity in both spectra and will be eliminated upon subtraction. This difference spectrum, defined as STD, contains the binding epitope of the ligand.

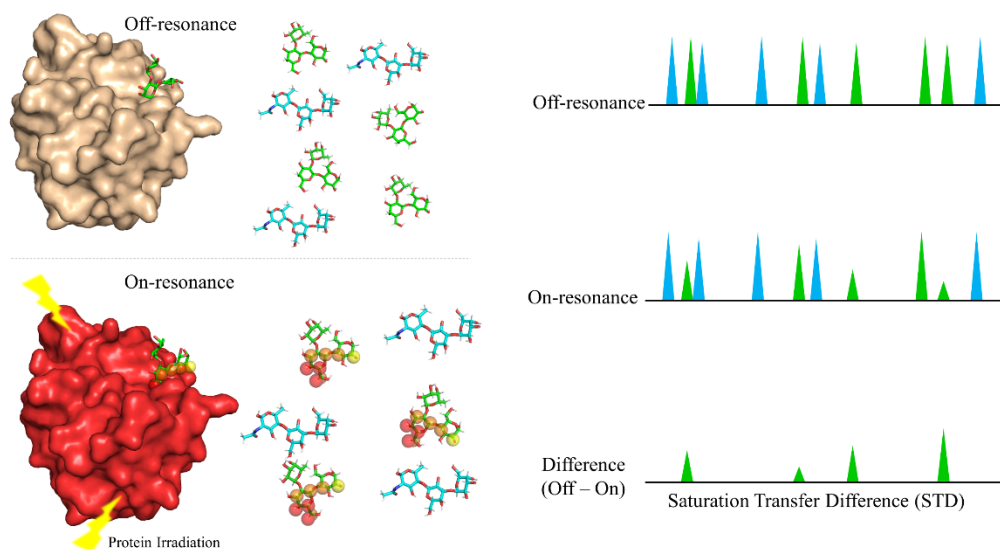


Figure 28. The STD NMR experiment for a mixture of ligands A (blue) and B (green). In the off-resonance spectrum, the protein is not saturated. In the on-resonance spectrum, the protein is saturated, and the magnetization is transferred to protons of the bound ligand, B. As a result, some of the protons of B suffer a decrease in intensity. The signals of A display the same intensity in both spectra. Thus, the difference STD NMR spectrum shows that only ligand B is bound and defines its binding epitope to the target receptor.

The difference between the off-resonance (I_0) and on-resonance (I_{on}) intensities is the STD (I_{STD}) intensity (Eq 6). Usually, the intensity for each proton ($I_{STD,i}$) is shown as the relative value compared to the proton that displays the most intense STD ($I_{STD,max}$) (Eq 7).

$$I_{STD} = I_0 - I_{on}$$

Equation 6

$$\text{Relative STD}_i = \left(\frac{I_{\text{STD},i}}{I_{\text{STDmax}}} \right)$$

Equation 7

STD-NMR displays multiple advantages. Firstly, there is no need of labelling the macromolecule. Additionally, there is no size limit of the macromolecule. Just a 5 μM concentration of the receptor can be enough to obtain successful STD-NMR experiments. Moreover, the affinity range of the systems for which STD experiments can be performed is fairly wide: $10^{-8} \text{ M} < k_D < 10^{-3} \text{ M}$. Since only signals from the binders arise in the difference spectrum, STD-NMR also identifies binders among different candidates. Due to all of these advantages STD-NMR is widely used in the drug discovery field.

Although STD-NMR does not provide direct information on the binding site of the protein, information of the amino acids surrounding the ligand can be obtained by analysing the results irradiating the protein signals at different frequencies. Saturation is usually performed either at the aromatic (7 to 9 ppm) or at the aliphatic (-1 to 1 ppm) region. Even if the magnetization is rapidly spread throughout the protein through spin diffusion, intensity differences between aromatic and aliphatic irradiations are usually present. These differences, although small, can indicate whether a certain proton is closer to aromatic or aliphatic residues¹³¹.

One of the drawbacks of STD-NMR is the overlap of numerous signals, especially for carbohydrates. Most of their ^1H -NMR signals appear between 4.5-3.5 ppm. In this narrow chemical shift range, many signals overlap, which hinders the precise analysis of the binding epitope. 2-Dimensional STD-NMR experiments have been developed in which the second dimension provides significant enhancement in the spectral dispersion. For instance, the synthesis

of ^{13}C labelled carbohydrates or carbohydrates with fluorine tags has enabled performing STD-HSQC and STD-TOCSY experiments^{132,133}.

1.4.1.3 Receptor-based NMR methods

In receptor-based methods, changes on the chemical shift of the macromolecule are observed. Usually, labelling with NMR active heteronuclei is necessary for this observation. In this Thesis, we have mainly performed $^1\text{H},^{15}\text{N}$ -HSQC-based titrations.

1.4.1.3.1 $^1\text{H},^{15}\text{N}$ -HSQC-based titrations

The use of ^{13}C (usually ^{13}C -labelled glucose) or ^{15}N (usually $^{15}\text{NH}_4\text{Cl}$) containing precursors allows the introduction of ^{13}C and/or ^{15}N atoms into recombinant proteins. The presence of ^{15}N nuclei in proteins permits obtaining ^1H - ^{15}N correlations: the magnetization is transferred from ^1H to ^{15}N through J-coupling, it evolves encoded in the chemical shift of ^{15}N and is finally transferred back to the ^{15}N -attached proton for detection, yielding a 2-D spectrum in which every proton attached to a ^{15}N provides a cross peak¹³⁴.

In order to obtain a spectrum in which the direct correlations between the ^1H and ^{15}N nuclei are observed, two different types of experiments have been performed in this Thesis: $^1\text{H},^{15}\text{N}$ -HSQC (Heteronuclear Single Quantum Coherence) for the smaller single domain proteins and the $^1\text{H},^{15}\text{N}$ -TROSY (Transverse relaxation optimized spectroscopy) variant for the larger two-domain proteins¹³⁵.

Due to the sensitivity of the chemical shift to the environment of each nucleus, the obtained spectrum is considered as the fingerprint of the protein, since it

reflects its 3-D structure¹³⁶. Additionally, the ¹H,¹⁵N-HSQC spectrum of a given protein is unique and different to those of others, and any modification on the protein is noticeable in the spectrum. For instance, although Gal-4N and Gal-9N are galectin domains with a very similar shape and a 60.1% of sequence similarity, they display very different ¹H,¹⁵N-HSQC spectra (Fig 29).

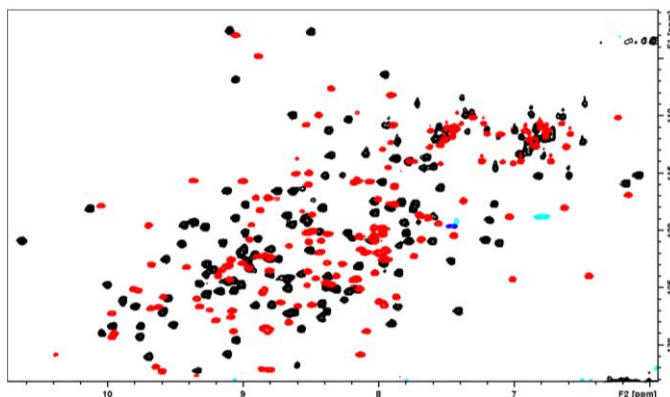
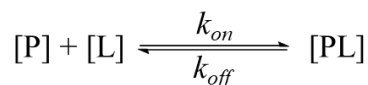


Figure 29. Superimposition of the ¹H,¹⁵N-HSQC spectra of Gal-4N (red) and Gal-9N (black).

This sensitivity of the chemical shift towards changes in the chemical environment of the nuclei can be exploited to monitor binding events^{106,137}. When a protein binds to a ligand, the ligand induces changes in the local environment of those protein nuclei that are close to it. As a result, the nearby protein nuclei suffer changes in their chemical shifts, defined as Chemical Shift Perturbations (CSP), which can be applied to monitor the binding events. The equilibrium of the system, i.e. the concentration of the protein (P), ligand (L), and protein-ligand complex (PL) is defined by equations 8 and 9.



Equation 8

$$K_A = \frac{1}{K_D} = \frac{[PL]}{[P] \cdot [L]}$$

Equation 9

Depending on the value of k_{off} , the system can be classified within the slow or fast exchange regime in the NMR chemical shift time scale¹³⁸. When the value of the k_{off} is slower than the difference (in Hz) between the frequency of a given nucleus in their free and bound states, the system is in the slow exchange regime. The corresponding result is shown in Fig 30-A. As the ligand is added, the HSQC cross peak corresponding to the bound protein will appear, whereas the intensity of the peak of the free protein decreases. Since the intensity of each peak is proportional to the population of each state and the protein and ligand concentrations of the mixture are known, the value of the dissociation constant can be calculated through fitting to equation 8.

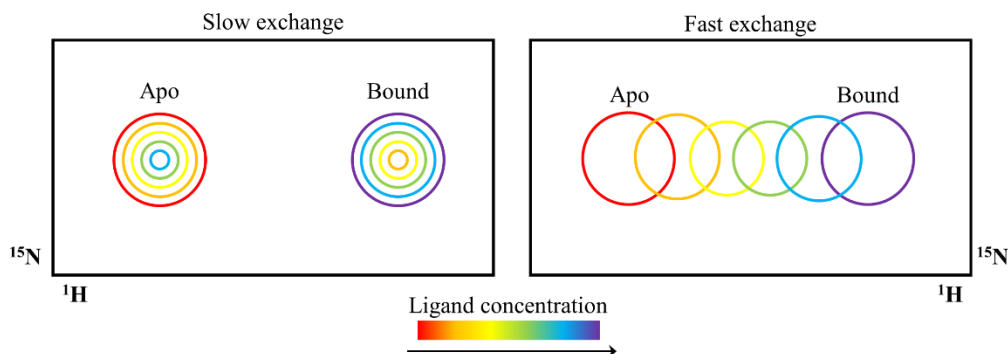


Figure 30. A) Example of a ^1H , ^{15}N -HSQC-based titration of a system in slow exchange in the chemical shift time scale. B) Example of a ^1H , ^{15}N -HSQC-based titration of a system in fast exchange.

Nevertheless, when the k_{off} is faster than the difference in frequency of the nucleus in the two states, the system is in the fast exchange-rate in the chemical shift time scale, and yields a different type of NMR spectrum. Under these conditions, the initial cross peak, corresponding to the apo state, smoothly shifts

towards the chemical shift of the bound state (Fig 30-B). In a fast-exchange system, the chemical shift of the observed peak corresponds to the weighted average of the two states (Eq 10). Since the sum of the fractions of free (f_{free}) and bound protein (f_{bound}) equals 1 (Eq 11, Eq 12 and Eq 13), equation 14 holds and can be used.

$$\delta_{\text{obs}} = \delta_{\text{free}} \cdot f_{\text{free}} + \delta_{\text{bound}} \cdot f_{\text{bound}}$$

Equation 10

$$f_{\text{free}} + f_{\text{bound}} = 1$$

Equation 11

$$f_{\text{bound}} = \frac{[\text{PL}]}{[\text{P}_T]} \quad f_{\text{free}} = \frac{[\text{P}_o]}{[\text{P}_t]}$$

Equations 12 and 13

$$\Delta\delta_{\text{obs}} = f_{\text{bound}} \cdot \Delta\delta_{\text{max}}$$

Equation 14

Since the bound fraction is related to K_D (Eq 9 and Eq 12), the equation 15 can be obtained, and during the titration, the total protein and ligand concentration are known, thus, the value of the k_D can be obtained.

$$\Delta\delta_{\text{obs}} = \Delta\delta_{\text{max}} \frac{\{([\text{P}]_t + [\text{L}]_t + K_D) - [([\text{P}]_t + [\text{L}]_t + K_D)^2 - 4[\text{P}]_t[\text{L}]_t]^{1/2}\}}{2[\text{P}]_t}$$

Equation 15

The general procedure and outcome of the $^1\text{H}, ^{15}\text{N}$ -based HSQC titration protocol is summarized in figure 31.

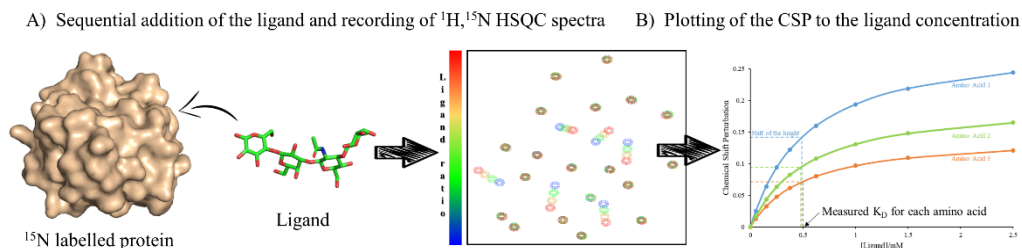


Figure 31. General process for a $^1\text{H}, ^{15}\text{N}$ -HSQC-based titration protocol. A) To a ^{15}N -labeled protein, different quantities of a certain ligand are added. After each addition a $^1\text{H}, ^{15}\text{N}$ -HSQC spectrum is recorded. B) From the titration data, the K_D can be obtained.

The changes that the ligand induces are not only useful to obtain the dissociation constant. The observed CSP that a given ligand generates in the backbone cross peaks are plotted for every residue. The obtained plot allows determining the binding site of the protein, as well as discovering secondary binding sites or conformational changes on the protein^{56,139}. However, in order to obtain this key information, the peaks in the $^1\text{H}, ^{15}\text{N}$ -HSQC spectrum need to be assigned.

1.4.1.3.2 Protein backbone NMR assignment

Generally, in order to achieve the assignment of the backbone $^1\text{H}, ^{15}\text{N}$ signals of a protein, 3-Dimensional NMR experiments are needed. In these experiments, through a series of INEPT-like sequences, the magnetization is sequentially transferred from the amide proton (the most sensitive nucleus) to the less sensitive nuclei (^{15}N , $^{13}\text{C}_{\text{CO}}$, $^{13}\text{C}_{\alpha}$, and/or $^{13}\text{C}_{\beta}$) through their J-couplings. Then, the chemical shifts of these heteronuclei evolve during the corresponding evolution times and then the magnetization is transferred back to the amide proton for detection, obtaining correlations between all these nuclei^{140–146}.

For example, in the HNCACB experiment, the magnetization is firstly transferred from the proton to the amide nitrogen, then to the C_{α} , and afterwards to the C_{β} . Finally, the magnetization is sequentially transferred back to the

amide proton (Fig 32-A)¹⁴⁵. As a result, a 3D spectrum is obtained in which, for each NH, correlations with the C_α and C_β within its residue and those of the previous one are obtained (Fig 32-B and C). In the complementary experiment, the HNcoCACB, the magnetization is not directly transferred from the ^{15}N to the C_α . Alternatively, it goes through the carbonyl group towards to the vicinal amino acid¹⁴⁶. As a result, only the correlations with the C_α and C_β frequencies of the previous amino acid are observed.

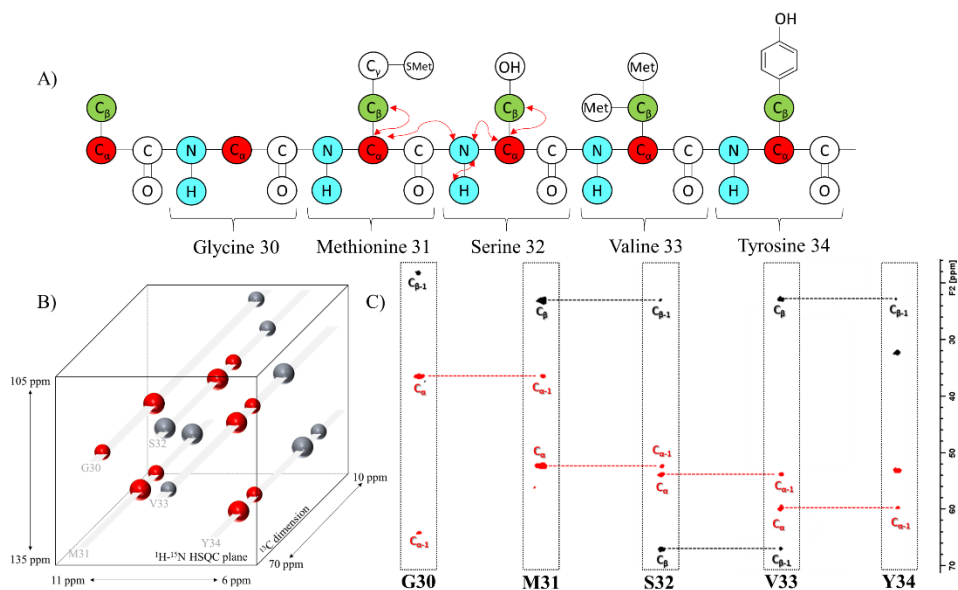


Figure 32. A) A five amino acid sequence from Gal-4N. The amide ^1H and ^{15}N are depicted in blue, the C_α in red and the C_β in green. The pathway of the magnetization transference in the HNCACB experiment is indicated with red arrows. To avoid confusion, only the transferences in the serine residue are noted. B) Simulation of the 3D experiment of the five amino acids. C) Strips of the HNCACB experiment of Gal-4N. The assignment and the correlation between each peak is noted.

HNCA and HNcoCA experiments are very similar to HNCACB and HNcoCACB, except that the magnetization is not transferred throughout the amino acid side chain further away than the C_α ^{140,142}. The consequence is that

only correlations with the corresponding C_α frequencies are observed, although the sensitivity is higher.

In the HNcaCO experiment, the magnetization is transferred from the proton to the ^{15}N and then to the $^{13}\text{C}_\alpha$. From here, it is transferred to the ^{13}C carbonyl and back to the proton, obtaining correlations for each NH with the ^{13}C of the carbonyls for its own and previous amino acid (Fig 33)¹⁴¹. In the complementary experiment, the HNC(O), as it does not pass through the C_α , only the correlation with the ^{13}C of the carbonyl of the previous amino acid is observed¹⁴⁰.

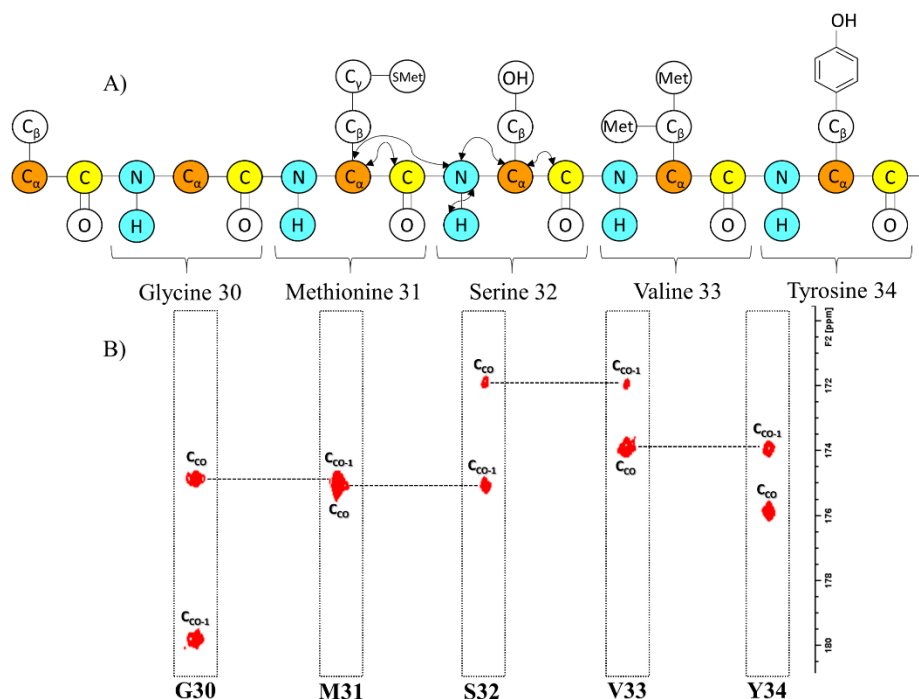


Figure 33. A) A five amino acid sequence from Gal-4N. The amide ^1H and ^{15}N are depicted in blue, the C_α in orange and the C_{CO} in yellow. The pathway of the magnetization transference in the HNcaCO experiment is indicated with black arrows. To avoid confusion only the transfereces in the serine residue are noted. B) Strips of the HNcaCO experiment of Gal-4N. The assignment and the correlation between each peak is noted.

These experiments yield the chemical shifts of the C_{α} , C_{β} and C_{CO} of the previous and own amino acids for each amide moiety, enabling the engagement of sequential amino acids. However, the presence of proline residues obstructs the sequential assignment process, as no information can be obtained given the lack of amide protons. Additionally, many cross peaks may also be “lost” (usually C_{β} s), due to relaxation events during the time periods required for efficient transfer of magnetization around the C-N and C-C bonds. These features may preclude the full assignment of the protein backbone nuclei.

The chemical shifts of the C_{α} , C_{β} and C_{CO} nuclei obtained in the aforementioned experiments are also of relevance for a successful protein backbone assignment. As mentioned in section 1.4.1.1, the chemical shift is very sensitive to the chemical environment of each nuclei, and therefore, it is a clear indicator of the chemical nature of the amino acid. In particular, C_{β} s display a broad chemical shift dispersion and help to distinguish between amino acid side chains (Fig 34). For instance, Trp and Tyr amino acids can be distinguished by the chemical shift of its C_{β} . There are certain amino acids whose chemical shifts are very distinct from others and therefore easy to identify. Specifically, the chemical shifts of C_{β} s of alanine, serine, and threonine, as well as the C_{α} s of glycine residues are very different to the others.

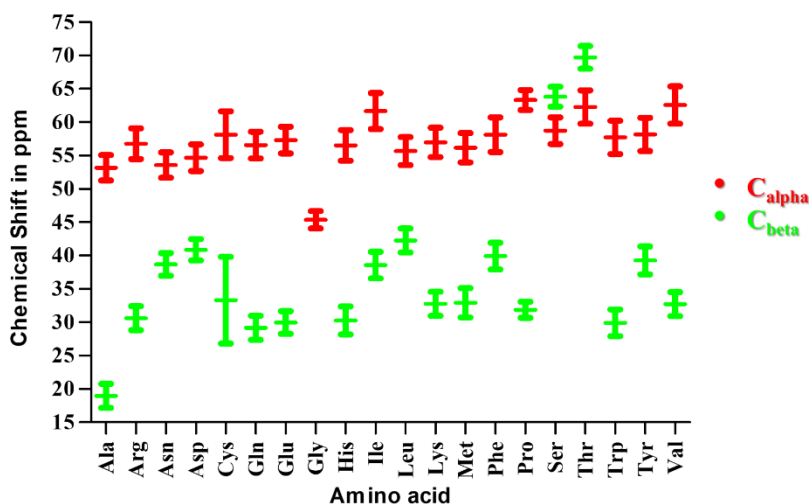


Figure 34. Plot of chemical shift values and their standard deviations for the C_α and C_β of each amino acid. Taken from the Biological Magnetic Resonance Data Bank (BRMB).

In order to facilitate the NMR assignment of a protein, various NMR approaches have been applied to identify the corresponding amino acid. For instance, there are NMR experiments in which only the peaks of a particular amino acid type is observed^{147,148}. Another alternative to identify certain amino acids is to add unlabelled or labelled precursors to the growing organism. In this manner, as the added precursor has a different isotope to the rest of the protein, those amino acids whose precursors have been added can be easily identified^{149,150}.

1.4.2 Isothermal Titration Calorimetry

Isothermal Titration Calorimetry (ITC) is a biophysical technique that allows obtaining the thermodynamic profile of interaction events^{151,152}. In ITC experiments, a molecule (receptor) is titrated with another molecule (ligand), and the heat exchange that the interaction releases (exothermic) or absorbs (endothermic) is measured. This heat is plotted against the ligand concentration (Fig 35) and the enthalpy (ΔH), stoichiometry (n) and dissociation constant (K_D) of the interaction can be obtained. Additionally, through the employment of the basic Eq. 16, the entropy (ΔS) of the interaction can also be estimated.

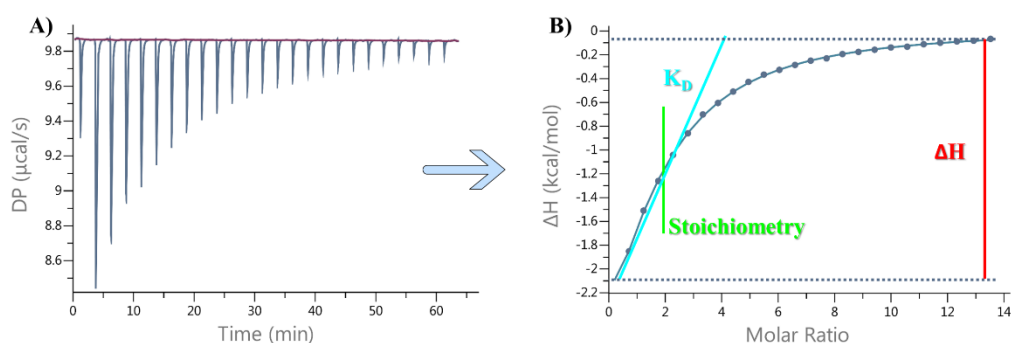


Figure 35. A) Example of an ITC thermogram. B) The binding parameters that can be obtained from the analysis of the thermogram.

$$\Delta G = -RT \ln K_A = \Delta H - T\Delta S$$

Equation 16

In the ITC instrument, two identical coin shaped cells are located in an adiabatic jacket (Fig 36)^{151,153}. The first cell is the reference and usually contains either water or the buffer in which the experiment is being performed. The second one is the sample cell, where the receptor is placed. These two cells are constantly maintained at the same temperature. Indeed, the temperature difference (ΔT)

between them is zero. When the ligand is added into the sample cell (usually 2-3 μL), heat will be released/absorbed as a result of the interaction. However, the ΔT between the cells will remain zero, as heating power will be applied to maintain it. The power applied to maintain the temperature constant can be converted into heat to measure the heat released by the interaction.

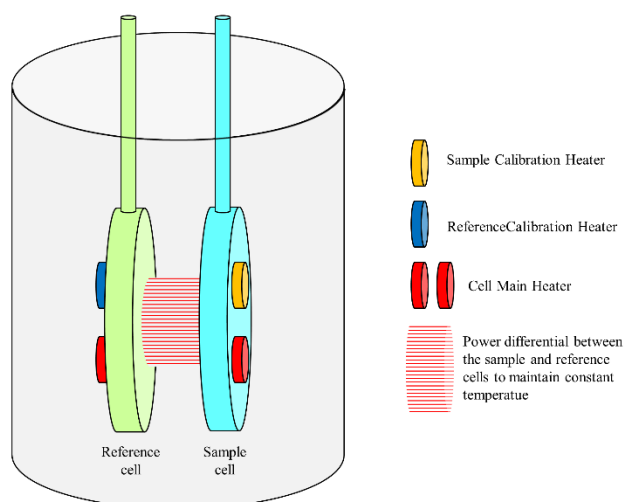


Figure 36. Basic scheme of an Isothermal Titration Calorimeter cell.

ITC allows measuring the interaction between many different types of entities, regardless of their sizes and without any kind of labelling. For instance, interactions of proteins with small ligands, glycoproteins or antibodies can be measured. As mentioned in chapter 1.3, interactions between sugars and proteins are usually in the mM- μM range. Through ITC, the thermodynamic profile of interactions in the nM to mM range can be obtained¹⁵⁴.

Other factors should be considered before applying ITC to estimate the strength of protein-carbohydrate interactions. As depicted in figure 35, the dissociation constant value is obtained from the slope of the titration curve. Therefore, to obtain precise k_D values, the curves of the titrations should be sigmoidal. The

value of c (Eq 17), a dimensionless constant, describes the practical window and the shape of the measurement¹⁵⁵.

$$c = n \frac{[P_T]}{K_D}$$

Equation 17

Where n is the stoichiometry of the system, $[P_T]$ the total protein concentration and K_D the dissociation constant.

In Fig 37, different ITC curves with different c values are simulated. For “extreme” K_D values - low nM (high c values) or mM (low c values) - the curves are far from sigmoidal and may lead to low quality estimations. In general, c -values between 1 and 1000 are needed to obtain reliable data, which means that the protein concentration ($[P_T]$) should be between one to one thousand times the corresponding K_D . For lectin-carbohydrate interactions, this requirement could be a bottleneck. Since most of these interactions are in the high μ M or mM range, large quantities (mgs) of protein and ligand may be required. Therefore, to acquire proper ITC curves, the affinity of the interaction, as well as the receptor and ligand concentration must be carefully considered.

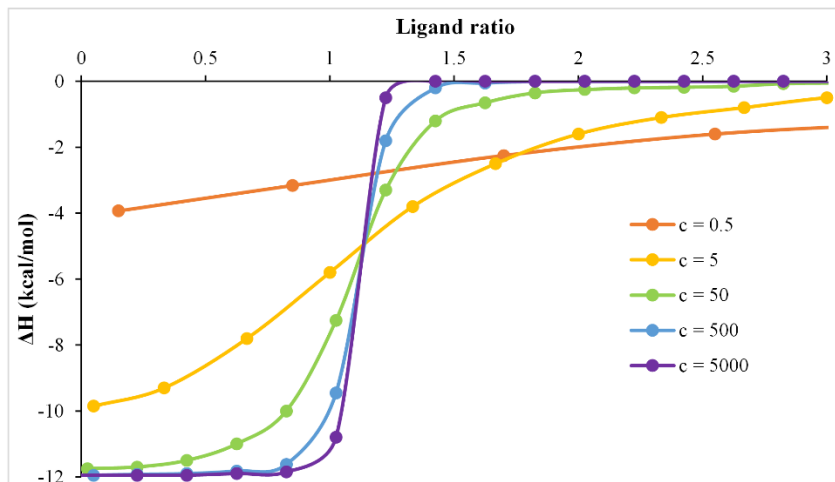


Figure 37. Simulation of diverse ITC thermograms with different c values.

2 Objectives

From the training point of view, the key objective of this Thesis has been to learn the principles of Nuclear Magnetic Resonance, and how NMR can be applied to characterize protein-carbohydrate interactions. To achieve this, a solid knowledge about the structural and conformational properties of glycans has been achieved. Additionally, in order to be able to study the interaction events, several proteins have been expressed and purified. Learning the techniques and strategies to obtain these proteins was also a main objective.

From the scientific point of view, a major target was deciphering how galectin-4 interacts with the blood group antigens, and how subtle chemical modifications on these antigens affects the binding events. Since the lectin has two binding sites, the first aim was to understand how each of the domains recognize these antigens, then to deduce how the full-length protein behaves.

Since the blood group antigens are mimicked by bacteria, the aim was to determine how galectin-4 interacts with the antigens mimics present on pathogens, addressing the impact of the multivalent presentation of blood group antigen-like glycans in multivalent scaffolds as lipopolysaccharides (LPS). Additionally, since galectin-4 has been reported to interact with the main co-receptor of LPS, CD14, this part of the Thesis has focused on studies towards deciphering the structural details that modulate the binding between the three entities: galectin-4, LPS, and CD14.

The last objective of the Thesis was learning how to synthesize well-defined β -glucans by employing a solid-phase synthesizer. For that, traditional chemical synthesis was employed to produce the building required blocks.

3 Results and Discussion

3.1 The interaction between Galectin-4 and the histo blood group antigens

3.1.1 Introduction

Galectin-4 (Gal-4) is one of the 16 members of the human galectin family, characterized by their participation in a myriad of biological phenomena, with important implications in immunity^{156–158}, inflammation, and cancer¹⁵⁹. Galectins exert their functions through their ability to bind β -galactoside-containing glycans, towards which the different galectins present different affinities and selectivities⁵². This family can be sub classified in three groups (see Figure in introduction) according to the organization of their carbohydrate recognition domains (CRD). Prototype galectins (-1, -2, -5, -7, -10, -11, -13, -14, -15 and -16) are homodimers with two identical CRDs united by non-covalent interactions, tandem-repeat galectins (-4, -8, -9, and -12) have two CRDs with different amino acid sequences that are covalently linked through a short linker peptide. Finally, galectin-3 is the only member of the chimera-type galectins, which has a C-terminal CRD with a non-lectin N-terminal tail.

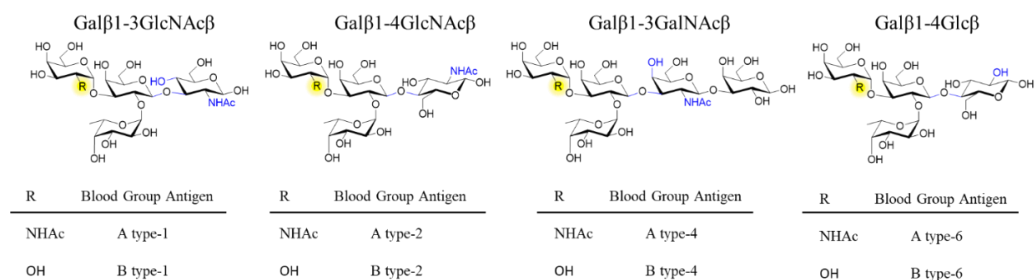
The physiological roles of Gal-4 have been reviewed⁸³, and include apical protein trafficking, intestinal wound healing, lipid raft stabilization, and bacterial pathogen fighting. In healthy individuals, Gal-4 is predominantly expressed in the epithelial cells along the alimentary tract although, as other members of the galectin family, its distribution in cancer is altered¹⁶⁰. Although its exact roles in the disease are not fully understood, the correlation between Gal-4 levels and malignancy has been reported for many cancers¹⁶¹. These evidences have fostered the use of galectins as targets for therapeutic intervention for several types of cancer^{162,163}.

Gal-4 belongs to the tandem repeat galectins subfamily, characterized by displaying two different carbohydrate recognition domains covalently linked

through a short peptide. The role of this peptide linker in the overall protein behavior is not yet clear, as well as its influence on the carbohydrate binding properties. In the case of Gal-4, the two CRDs show 49% similarity and, despite sharing certain glycan-binding preferences, they do not completely overlap. Previous studies on the binding preferences of full-length Gal-4^{60,74,164}, as well as for the two independent N- and C-terminal domains¹⁶⁰ have revealed their preference for the human histo blood group antigens (HBGA) A and B. Despite certain controversy⁷⁴, most studies agree on the fact that the N-terminal domain of Gal-4 (Gal-4N) recognizes A and B blood group antigens with less affinity than the C-terminal domain (Gal-4C). Nevertheless, significant discrepancies on the binding affinities were reported^{54,60,71}. X-ray crystallography analysis has provided structural details on how the two independent CRDs recognize different linear lactose derivatives, including 2'-fucosyl lactose (H-type 6 antigen)^{69,71}. However, no experimental data are yet available for the recognition of branched antigens, including the blood type antigens.

In particular, the ABO antigens are divided into six groups, depending on the peripheral core disaccharide structures (Scheme 1): type-1 (Gal β 1-3GlcNAc β), type-2 (Gal β 1-4GlcNAc β), type-3 (Gal β 1-3GalNAc α), type-4 (Gal β 1-3GalNAc β), type-5 (Gal β 1-3Glc β) and type-6 (Gal β 1-4Glc β), giving rise to a family of epitopes that are presented in different manners. This different presentation is known to influence their antigenicity¹⁶⁵⁻¹⁶⁷. Moreover, their distribution is also different. HBGA are mostly present as type-2 structures (N-acetyl lactosamine, Gal β 1-4GlcNAc β). However, HBGA are not only found on red blood cells and, in fact, most endothelial and epithelial cells cell surfaces, together with the secretions of the so-called AB-secreter genotype individuals, also contain HBGA. In fact, type-1 antigens are found on endodermal tissues, type-2 on both ecto- and endo-dermal tissues, while type -3 and -4 antigens are present on ecto or endodermally-derived tissues, including the salivary glands and kidneys. While the presence of type-5 in humans is controversial¹⁶⁸, the

type-6 has been much less studied, being reported to be in human milk and tissue samples¹⁶⁹.



Scheme 1. The blood group antigens employed in this study. The substitution at the position 2 in the non-reducing end (in yellow) determines whether the antigens belongs to group A or B, whereas their peripheral disaccharide core (in blue) determines the antigens type.

From a biological perspective, the significance of the specific association of HBGA with certain members of the galectin family seems intriguing. It has been suggested¹⁷⁰ that some galectins could fill an important gap in adaptive immunity by conferring protection against pathogens that use molecular mimicry for the infection, producing blood group-like antigenic glycan structures¹⁷¹. Thus, individuals of a given blood type, who do not produce antibodies against its own blood group antigens, would be then protected by the immune activity of galectins. In support of this hypothesis, Gal-4 has been demonstrated to be able to recognize and kill *Escherichia coli* (*E. Coli*) bacteria expressing the blood group B antigen⁹⁰. Interesting, a quick glance at the O-antigenic structures produced by *E. Coli*¹⁷² reveals that not only the O86 *E. Coli* serotype (that precisely express the B-antigen structure), but also other serotypes, incorporate O-antigens that share common structural elements with galectin ligands. Their chemical context compared to the endogenous epitopes is however different, for instance with respect to their neighboring residues and linkages. Considering the diversity of the ABO antigen presentation in terms of

peripheral glycan structures (types), both among endogenous glycans and host-mimicking pathogenic glycans, we have here explored the influence of this presentation in their recognition by Gal-4N and Gal-4C.

A library of commercially available HBGA with different presentations at the peripheral disaccharide core (Scheme 1) was selected to interrogate the binding preferences of Gal-4N and Gal-4C towards the possible chemical variations within this glycan antigenic family. In particular, we have focused on the A *versus* B preference, and on the chemical variations involving their core presentation, including the glycosidic linkage (β 1-3 *versus* β 1-4) as well as the nature of the first peripheral core monosaccharide (GlcNAc, GalNAc or Glc). We have performed a systematic study, based on experimental Nuclear Magnetic Resonance (NMR) data and Isothermal Titration Calorimetry (ITC) measurements, combined with computational chemistry tools, to provide an atomic-level rationalization of the observed binding preferences of Gal-4N and Gal-4C.

Once the binding affinities and preferences of the two isolated domains were assessed, we have also determined the behaviour of the full-length form (Gal-4FL), for which little was known. Rustiguel and co-workers⁷⁰ studied the structure of Gal-4FL using SAXS/WAXS, thermofluor assays and MD simulations and compared it with the separate domains. Their findings suggested that Gal-4FL folds as a compact structure and that there are some transient interactions between the CRDs. Nevertheless, it remained unclear whether there is crosstalk between the domains and if the full-length overall structure has an impact on the recognition events.

With the purpose of solving this issue, we have analysed the interaction between Gal-4FL and the blood group antigen A type-6 (best binder). Since both domains compete for the same ligands, two different lectins were produced to better understand the interaction. In these new proteins, the binding site of one

of the domains was disrupted,¹⁷³ whereas the other one was active, maintaining the overall structure of the protein, but only with one active binding site.

Through NMR and Isothermal Titration Calorimetry (ITC) experiments with Gal-4FL and its two mutants, it was demonstrated that both domains compete for the same ligands and that despite this competition, there is no cross-talk between the domains and that each of them acts independently. Additionally, the interaction is identical to that observed for the separate domains. Thus, the conclusions obtained for the free domains can be extrapolated to the full-length.

3.1.2 The N-terminal domain of Galectin-4

The interaction of Gal-4N with the A and B type-1, -2 and -6 tetrasaccharides, and with the A and B type-4 pentasaccharides (Scheme 1) was addressed by combining different NMR-based strategies. ¹H-Saturation Transfer Difference (STD) NMR experiments were employed to report on the glycan binding epitope^{174–176}. Additionally, ROESY experiments of ligand/Gal-4N mixtures revealed specific ligand ¹H resonances in slow exchange in NMR chemical shift timescale between the free and bound states. Fittingly, the analysis of their chemical shift differences ($\Delta\delta^1\text{H}$ free-bound) proved to be instrumental to provide a sort of ligand binding epitope information, complementary to that obtained by STD NMR experiments. Additional NMR experiments from the protein perspective, based on the observation of ¹⁵N-labeled Gal-4N, were used to estimate binding constant affinities (K_D) and determine the binding epitope of the lectin in each case, which were further quantitatively estimated by ITC. The synergic combination of these experimental data with molecular modelling procedures provided detailed 3D models of the complexes.

3.1.2.1 Nuclear Magnetic Resonance (NMR)

3.1.2.1.1 ¹H-STD NMR

¹H-STD-NMR spectra were acquired for mixtures of Gal-4N (50 μM) with a 50-fold excess of ligand. Protein aromatic irradiation afforded the best STD intensities and are thus these are those discussed henceforth. The STD NMR spectra were very similar among the different ligands with respect to both the observed glycan protons and their intensities. For all the ligands, all the protons of the central βGal residue except H2, together with H1 and H2 of the terminal $\alpha\text{Gal}/\alpha\text{GalNAc}$ residues showed the strongest relative STD NMR effects

(above 50%), (Fig 38 and SI). Severe proton overlapping precluded a full comparison among ligands, but the analysis of the key isolated ^1H -NMR resonance signals permitted to highlight evident differences.

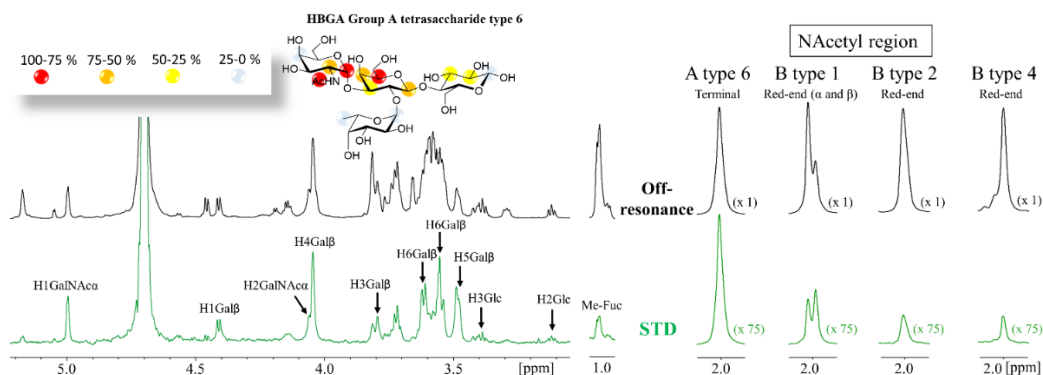


Figure 38. (A) ^1H -STD-NMR experiment for the complex formed by the A type-6 tetrasaccharide and Gal-4N (50:1 molar ratio). Top: the reference spectrum (black, off-resonance). Bottom: the STD-NMR spectrum (green, on-resonance at the aromatic region). The ^1H -NMR signals showing STD effect are annotated. The epitope mapping (relative STD) is shown the ligand structure. (B) Details of the off-resonance and STD-NMR spectra showing the NAc region obtained for the HBGA antigens in the presence of Gal-4N: From left to right: A type-6, B type-1, B type-2 and B type-4. Terminal refers to the terminal αGalNAc residue. Red-end refers to the reducing-end GlcNAc or GalNAc residues. The intensities of the STD spectra are 75 fold incremented with respect to the reference off-resonance spectra.

In particular, the comparison of the ^1H -STD NMR signals originating from the acetyl groups was highly informative. For all the A-type antigens, the Ac group at the terminal GalNAc residue showed very strong STD intensity (Fig 38-B). In contrast, the Ac groups at the reducing-end GlcNAc, present in type-1 (β 1-3) and type-2 (β 1-4) antigens, displayed much weaker intensities. However, these signals were always stronger for the β 1-3 than for the alternative β 1-4 linked antigens (Fig 38-B). This fact indicates that this NAc group is closer to protein aromatic residues (the STD on-resonance frequency irradiation) for the β 1-3 linked epitopes than for the analogous NAc for β 1-4 presentations.

The methyl group of the Fuc moieties showed weak STD intensities, always below 20%. Interestingly, they were systematically much weaker for the A than for the B antigens (see SI). These data strongly suggest that the Fuc residue is further away from the protein surface in the lectin complexes with the A antigens than in those with the B analogues.

For the longer pentasaccharides (type-4) there was a remarkable lack of STD effect at the reducing-end Gal residue (Fig 39) and very weak for the β GalNAc, clearly indicating that the binding epitope exclusively involves the terminal residues. This is a fairly interesting observation since the type-4 core structure (Scheme 1) contains the GalNAc β 1-3Gal moiety at the reducing end that could act as an additional binding epitope by positioning the GalNAc β - residue on the primary binding site (stacked on W84) increasing the apparent binding affinity. These data strongly suggest, however, that this is not the case.

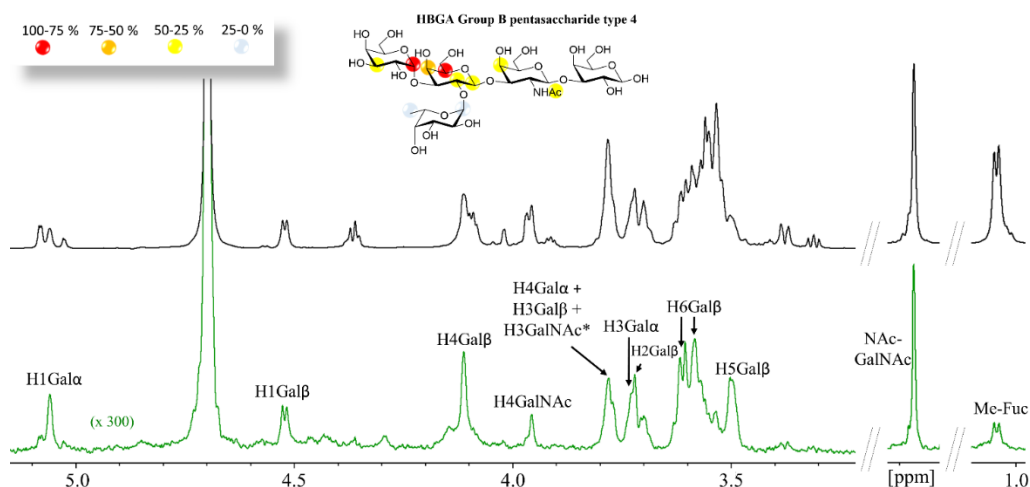


Figure 39. ^1H STD-NMR experiment for the complex formed by the B type-4 pentasaccharide and Gal-4N (50:1 molar ratio). Top: the reference spectrum (black, off-resonance). Bottom: the STD-NMR spectrum (green, on-resonance at the aromatic region). The ^1H -NMR signals showing STD effect are annotated. The epitope mapping (relative STD) is shown in the ligand structure. * overlapping proton signals.

3.1.2.1.2 trROESY NMR

trROESY spectra were acquired at 1:20 protein/ligand molar ratios¹⁷⁷. The observed ROESY cross peaks were very similar to those recorded for the free ligands, indicating that no significant conformational changes around the glycosidic torsion angles or at the pyranose chairs took place upon binding. Most interestingly, the trROESY analysis evidenced that some ligand protons were in slow chemical exchange between their free and bound states (Fig 40), thus revealing the corresponding chemical shifts in the bound states¹⁷⁸. Fittingly, those signals experiencing slow exchange in the NMR chemical shift timescale are expected to arise from protons that significantly change their chemical environment upon complex formation and that are highly affected by their proximity to the lectin surface. Remarkably, not all ligand protons showed chemical exchange cross peaks. The analysis of the chemical shift difference between the free and bound forms ($\Delta\delta_{(\text{free-bound})}$) thus afforded additional epitope mapping information from the ligand perspective. Moreover, the comparison of the behavior observed for the different ligands also permitted to deduce subtle differences related to their different binding geometries into the galectin binding site. Notably, for all ligands, all the protons of the central β Gal residue experienced drastic up-field shifts upon complex formation, with H4, H5 and one of the H6 standing out with more than 2 ppm difference between the free and bound forms (Fig 40 and 41). This effect is in full agreement with the expected key CH- π stacking interaction between the central β Gal residue and the indol side chain of W84 of the lectin¹⁷⁹. This key β Gal/Trp interaction is ubiquitous for all galectins^{53,180}.

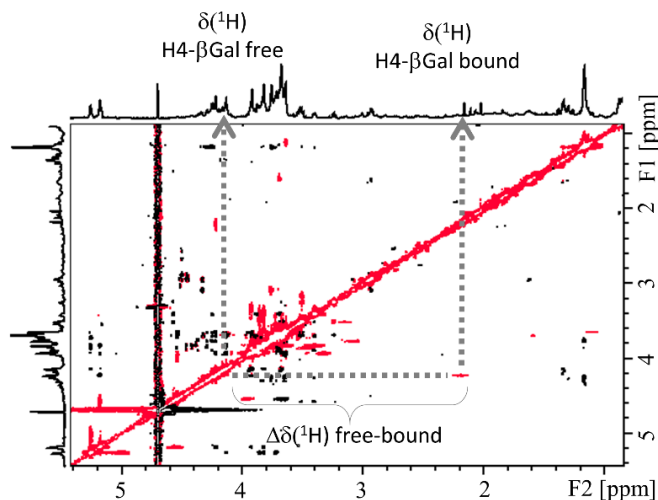


Figure 40. ^1H -NMR-based ligand chemical exchange (EXSY) analysis from trROESY experiments. trROESY spectrum of the B type-6 tetrasaccharide antigen in the presence of Gal-4N (molar ratio 20:1). The cross-peaks in black correspond to those arising from regular NOE effects in the rotating frame, while those in red arise from chemical exchange. Their chemical shifts at the different ^1H dimensions in the 2D spectrum correspond to those in the free (downfield) and bound (upfield) states.

Nevertheless, the profile is slightly but noticeable different for the A and the B antigens (Fig 41). For the B epitopes, H1 and H2 of the terminal αGal moiety experienced slow exchange, being up-field shifted upon binding. On the contrary, for the A-antigens, only H1 of the analogous αGalNAc residue displayed slow exchange, and a with reduced chemical shift perturbation, while H2 was not in slow exchange. The chemical shift difference free-bound for H4, H5 and H6' of the central βGal was also smaller for the A than for the B antigens (Fig 41), suggesting that the sugar-aromatic stacking interaction is stronger for the latter. Additionally, H1-Fuc (the only proton of the Fuc ring under slow exchange) is more affected in the B ($\Delta\delta_{(\text{free-bound})} = 0.2\text{-}0.3$ ppm) than in the A antigens (less than 0.2 ppm). These differences strongly suggest that the A and B antigens display a slightly different orientation in the Gal-4N binding site. Additionally, very few protons at the reducing-end residues showed weak chemical exchange effects, being these protons different among the different

peripheral core types (see SI). Finally, the type-4 oligosaccharides did not show any chemical exchange effect.

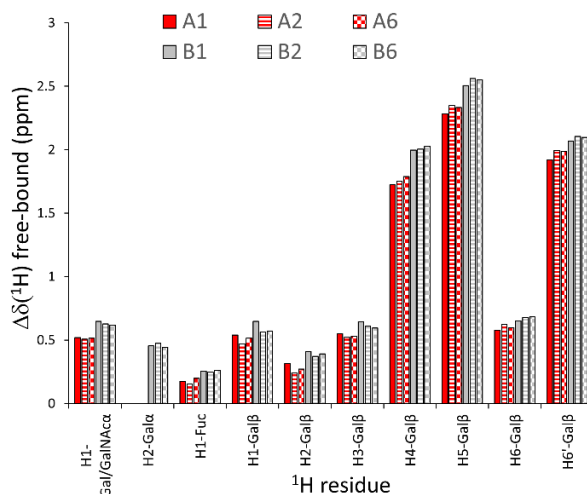


Figure 41. ^1H -NMR-based ligand chemical exchange (EXSY) analysis from trROESY experiments. Plot for the measured differences in chemical shifts between the free and bound states ($\Delta\delta(^1\text{H})$ free-bound) for those selected protons of the antigens that experience slow exchange in the ROESY spectrum in the presence of the lectin.

3.1.2.1.3 Chemical shift perturbations at Gal-4N

The molecular recognition event was also analyzed from the lectin's perspective. Thus, a chemical shift perturbation (CSP) analysis of the changes produced in the ^1H - ^{15}N HSQC spectra recorded for ^{15}N -labelled Gal-4N upon addition of the different ligands was carried out. This analysis required the NH backbone NMR assignment of the lectin signals, following well-established triple resonance 3D NMR methods (see Materials and methods and SI). The assignment protocol allowed assigning 80% of the NH backbone resonances (Fig 42). In addition to the ligands shown in Scheme 1, the corresponding non-fucosylated analogs and the H antigens (devoid of the terminal Gal/GalNAc moieties, see SI), as well as lactose, were also employed. Thus, a full

comparative analysis on the monosaccharide-specific effects on the lectin CSP was achieved. In addition, titration experiments afforded an estimation of the binding affinities, which are gathered in Table 4.

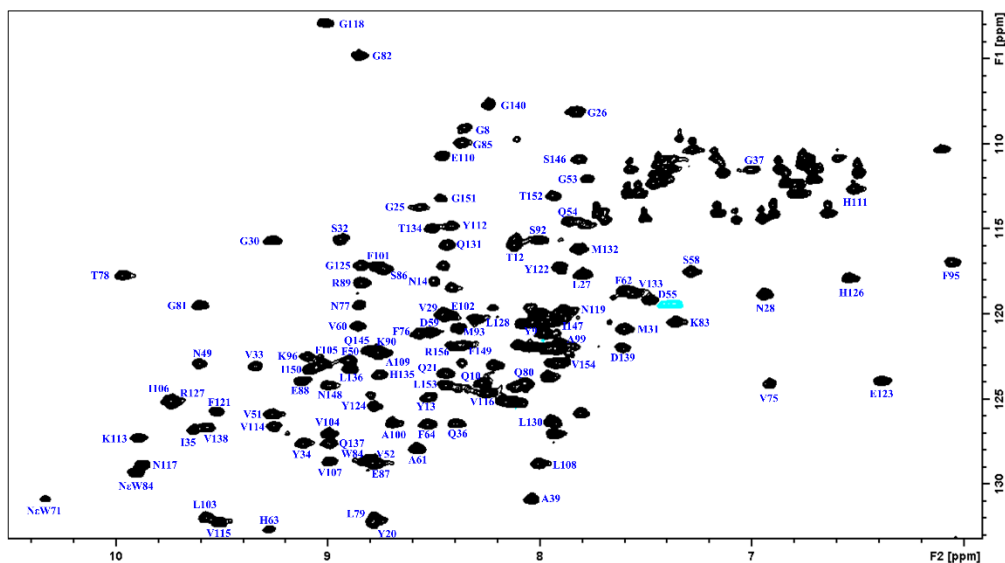


Figure 42. ^1H - ^{15}N HSQC of Gal-4C. The identity of the assigned peaks is noted next to each peak.

The observed CSP for the interaction with the trisaccharide 2'-fucosyllactose (H type-6) were basically identical to those for lactose, suggesting that the Fuc residue does not provide direct interactions with the lectin. In contrast, when the non-fucosylated B type-6 antigen was employed, additional CSP with respect to lactose were observed. These perturbations took place mainly at the S2 strand and at the side chain of W84, evidencing that the terminal αGal unit indeed interacts with this area of Gal-4N (SI).

The CSP measured for the A *versus* the B antigens showed significant differences for several residues, all located along the S face of Gal-4N (Fig 43). Some of these residues cluster in the S2 strand, such as V138, D139 and G140. These residues are remarkably differently affected between the A and B

antigens, both in magnitude and direction: they are more strongly perturbed in the presence of A, moving up-field in the ^1H dimension, than in B antigens (Fig 43-C). The analysis of the published X-ray crystallographic structures⁶⁹ showed that these amino acid residues are facing F47, located at the contiguous S3 strand. The observed trends suggest that the interaction with the A-antigens places residue F47 closer to these V138-G140 residues, a motion that does not take place when the B-antigens are bound (Fig 43-A).

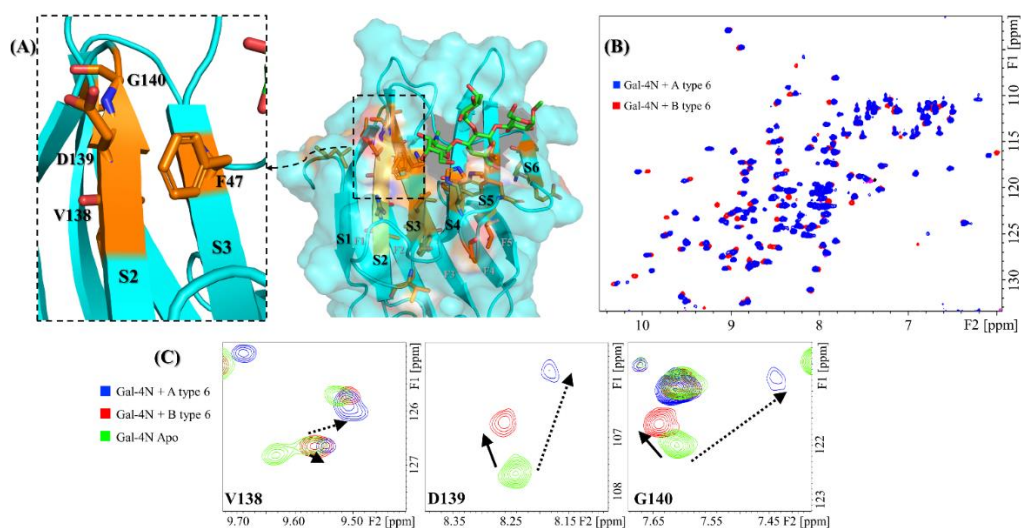


Figure 43. Protein backbone CSP analysis for the interaction of Gal-4N with A and B type-6 tetrasaccharide antigens. (A) Molecular model (MD simulation) for the complex of Gal-4N with A type-6 antigen. The residues that significantly differ in their chemical shift perturbation upon binding to group A or B antigens are highlighted in orange, as well as F47 (unknown resonance assignment). Left: zoom at the S2-S3 region showing residues F46, V138, D139 and G140. (B) Superimposition of the ^1H - ^{15}N HSQC spectra of Gal-4N in the presence of A type-6 (blue) and B type-6 antigens (red). (C) Expansion of ^1H - ^{15}N HSQC spectra at residues V138, D139 and G140: in green, Gal-4N apo; in blue, Gal-4N/A type-6 (12 eq.); and in red Gal-4N/B type-6 (10 eq.).

The different blood group types also displayed distinct CSP of the lectin cross peaks. The CSP between types-1 and -2 were basically identical, with only subtle differences at V75 (S5) and R89 (S6) (SI), while type-6 oligosaccharides presented different perturbations all along the S5 and S6 strands (SI), with the

strongest differences involving residues E87, R89 and K90. As for the type-4 antigens, there was a remarkable lack of perturbations at the S5 strand (Fig 44), more pronounced for the B than for the A antigen (SI).

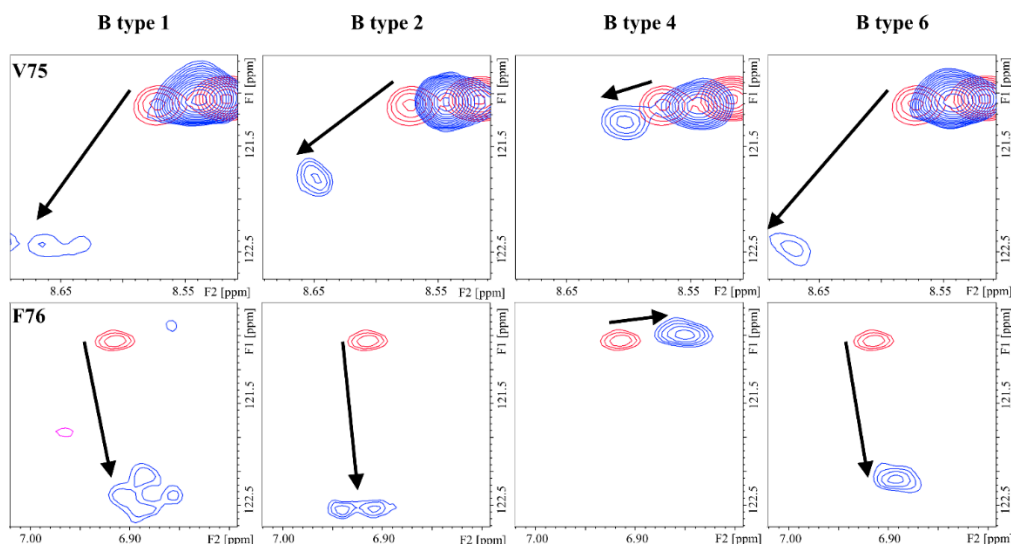


Figure 44. ^1H - ^{15}N HSQC: chemical shift perturbation of residues V75 and F76 of Gal-4N in the presence of the B antigens types -1, -2, -4 and -6.

In addition to the information yielded by the CSP analysis, titrations provided an estimation of the equilibrium dissociation constants of each ligand (K_D , Table 4). The NMR-based binding affinities spanned a 20-fold range between the best (B-type 6) and worse (A-type 4) binders, with K_D values in the μM -mM range. Lactose, the minimum binding epitope, also exhibited a very low, although measurable, affinity with a K_D of 1.6 mM, in agreement with previously reported data^{71,181}. Interestingly, the tetrasaccharides presented conserved trends in their binding preferences. Thus, for the same peripheral core structure, the B-type antigens were always better binders than the corresponding A-antigens. With respect to the core structures, the type-6 antigens are the best binders, followed by type-1 and -2, which are very similar,

irrespective of their terminal residues (whether they are A or B). Finally, the type-4 oligosaccharides were the worse binders, with the A type-4 almost as weak as lactose.

3.1.2.2 Isothermal Titration Calorimetry

In order to provide alternative and complementary estimations of the binding affinity data, ITC experiments were also carried out for the six HBGA tetrasaccharides (Table 4). The binding with the type-4 pentasaccharides was too weak to obtain reliable ITC data. The thermodynamic profile for all the interactions showed the typical enthalpy-driven binding process (Table 4). The best enthalpy values are always compensated by opposing entropy contributions, especially significant for the B type-6 saccharide. The significant favorable binding enthalpy (-12.3 kcal/mol) is accompanied by the largest entropy penalty (-6.4 kcal/mol). This binding profile constitutes a prototypical example of enthalpy-entropy compensation¹⁸² with a nearly perfect relationship between ΔH and $-T\Delta S$ (see Supporting Information). However, due to the moderate or low affinity interactions (all with K_D above 50 μM), the resulting thermodynamic parameters will not be further interpreted¹⁵⁵. Although the magnitudes of the ITC-derived K_D values were in average 1.5-fold lower than those determined by NMR, both techniques yielded conserved trends, reproducing the binding preferences described above, schematized in Figure 45.

Table 4. Equilibrium dissociation constants (K_D) determined for the interactions of the different A and B blood group antigens with Gal-4N by employing HSQC NMR-based titrations and ITC measurements, along with the thermodynamic parameters for the interactions, as determined by ITC.

Ligand	K_D	K_D	ΔG (kcal mol^{-1})	ΔH (kcal mol^{-1})	$-T\Delta S$ (kcal mol^{-1})
	(μM) (NMR)	(μM) (ITC)			

A type-1 (tetra)	440	267	-4.9	-5.9	1
A type -2 (tetra)	370	187	-5.1	-6.9	1.8
A type-4 (penta)	1380	-	-	-	-
A type-6 (tetra)	150	86	-5.6	-8.5	2.9
B type-1 (tetra)	200	109	-5.4	-7.7	2.3
B type-2 (tetra)	190	88	-5.5	-9.3	3.8
B type-4 (penta)	460	-	-	-	-
B type-6 (tetra)	60	51	-5.9	-12.3	6.4

The trend is fairly evident: The B group antigens are better binders than their corresponding A-analogues. Regarding the peripheral disaccharide cores, the type-6 antigens are preferentially recognized, while the type-4 ones display rather weak affinity. The type-1 and type-2 antigens display intermediate affinities.

In particular, for β 1-4 linked antigens, the substitution of Glc by GlcNAc at the reducing-end is detrimental for the binding event, as evidenced by the comparison between A type-2 *versus* A type-6, and B type-2 *versus* B type-6. In both cases, a two-fold increased affinity is observed for the type-6 linkages.

In contrast, the presentation of the β 1-3GlcNAc *versus* β 1-4GlcNAc epitopes at the peripheral core (type-1 *versus* type-2) does not basically influence the binding affinity, neither for the A nor the B tetrasaccharides.

Finally, the B antigens (with a terminal α Gal residue) always bind two-fold stronger than the A antigens (with a α GalNAc residue), irrespective of the peripheral core disaccharide.

The binding preferences of Gal-4N can be thus be categorized as shown in Figure 45, where the B type-6 is the best binder with a difference in affinity of 5-fold. Therefore, Gal-4N prefers α Gal moieties at the non-reducing end and Glc residues at the reducing end.

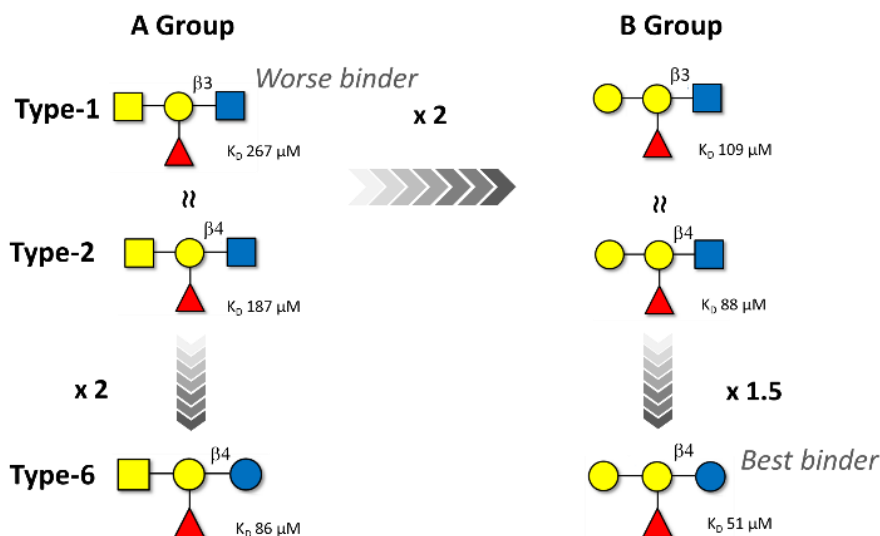


Figure 45. Schematic representation of the affinity trends for the binding of the A and B blood group antigen tetrasaccharides (types-1, -2 and -6) to Gal-4N as deduced by ITC. The NMR-based data follow the same trend although the estimated binding affinities are ca. 1.5 weaker. The type-4 oligosaccharides are not displayed since their affinities are even weaker than those of the type-1 analogues.

3.1.2.3 3D structural models for the binding complexes.

Once the experimental data were available, 3D structural models for the complexes were generated by atom-pair superimposition of the central β Gal pyranose of the ligands with the corresponding residue in the X-ray crystallographic structure of the reported complex of Gal-4N bound to lactose (PDB 5DUV)⁶⁹. Then, the generated complexes were submitted to 1 μs MD simulations using the ff14SB and GLYCAM06j-1 force fields as implemented in Amber18 (see materials and methods). Figure 46 gathers the superimposition of the complexes formed by Gal-4N with the A type-6 and the B type-6 antigens, as summary of the results. Fittingly, the subtle different orientations and presentations of the ligands at the lectin binding site permits explaining the observed differences found in the combined ligand-based and receptor-based NMR analysis and the ITC measurements.

As discussed above, the B antigens are better binders (and display better enthalpies) than the A analogues. In fact, according to MD simulations, only in the A-complex, the bulky Ac group at the terminal GalNAc pushes F47 (in S3) away, moving it towards S2. The lack of this Ac group in the B-complex allows for a better binding. These theoretical predictions are in full agreement with the experimental observations found in the CSP analysis (Fig 43). In fact, for all the A antigens, the α GalNAc residue is further away from the protein surface. For instance, in the type-6 antigen, H2- α GalNAc is on average at 5.3 Å from the aromatic ring of F47, whereas for the group B type-6, H2- α Gal is only 3.5 Å apart. This fact perfectly matches with the chemical exchange observations in the trROESY experiments in which this proton is up-field shifted only for the B antigens. The ring current effect of the F47 aromatic ring affects H2- α Gal, due to its close proximity and orientation (Fig 46). Regarding the central β Gal moiety, given the diverse presentation and fitting of the terminal residues, this β Gal residue in the A and B antigens is also differently positioned with respect to the amino acids that comprise the binding site (Fig 46). In fact, in the A-antigens, H4- β Gal is 0.5 Å further away from the imidazole ring of H63 than in the B-antigens. Additionally, H5- β Gal is also 0.4 Å further away from the indole ring of W84 in the A than in the B-antigens. These different geometries also correlate with the observed stronger chemical shift perturbations of these protons in the B than in A antigens. These geometry differences also impact both the strength of the β Gal/W84 CH- π interaction, as well as the OH4- β Gal/H63 hydrogen bond, key interactions in the molecular recognition of β Gal moieties by galectins.

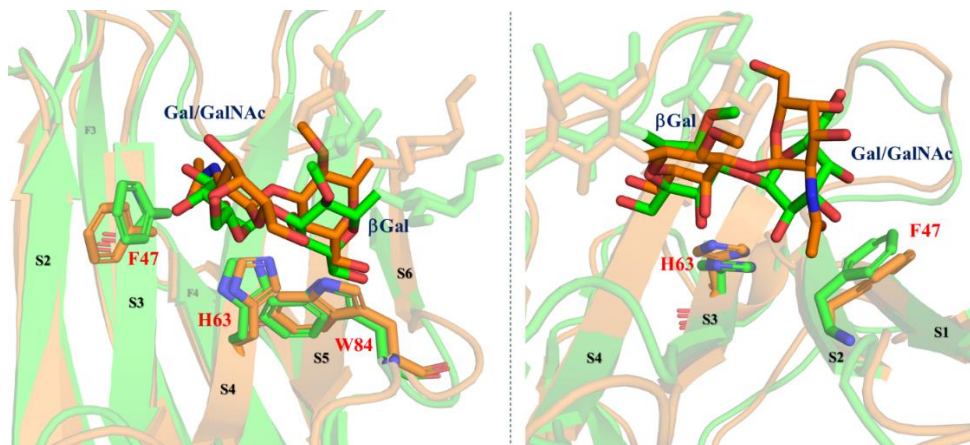


Figure 46. Different perspectives of the superimposition of the 3D models for the complexes Gal-4N/A type-6 (orange) and Gal-4N/B type-6 (green) according to MD simulations. Reducing-end and Fuc residues are faded. Key residues of the protein are highlighted.

With respect to the core-disaccharides, MD simulations show that, in the β 1-4 linked epitopes (types-2 and 6), positions 2 and 3 of the reducing-end pyranose ring are facing the lectin, while in β 1-3 linked analogues (types-1 and -4), these positions are in the opposite direction, exposed to the bulk solvent (Fig 47). These MD-based orientations are also in full agreement with the STD NMR data discussed above. For the β 1-4 linked epitopes, the MD simulations also predict hydrogen bonding interactions between OH3 of the reducing end sugar (Glc/GlcNAc) with the polar side chains of R65 and E87. Also, only in the type-6 antigens, an additional hydrogen bond is established between OH2-Glc and the carboxylate group of E87. This additional stabilizing interaction could provide the impetus for the additional affinity gain observed in the type-6 versus the type-2 cores. In type-1 antigens, which are β 1-3 linked, R65 and E87 establish hydrogen bonding with OH4-GlcNAc. Thus, despite the different orientation of the reducing-end pyranose with respect to β 1-4 linked antigens, type-1 and type-2 would be balanced in terms of establishing interactions, in agreement with their similar affinities. Finally, for type-4 antigens, the axial

disposition of the OH4-GalNAc β precludes the interactions with R65 and E87, in agreement with their reduced affinity.

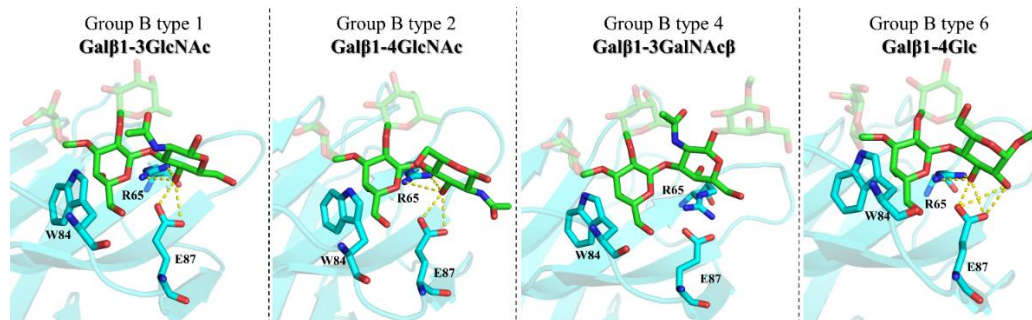


Figure 47. Molecular models of the complexes between Gal-4N and group B type -1, -2, -4 and -6 antigens according to MD simulations. Key residues are highlighted.

3.1.2.4 Conclusions

Herein the binding preferences of the N-terminal of Gal-4 towards the A and B blood group antigens has been examined at the molecular level. Different antigenic presentations in terms of peripheral core structure have been compared, as well as the A *versus* B preference. ITC experiments permitted to classify the ligands in terms of binding affinity, while different NMR strategies in combination with modelling protocols provided the structural rationale for the observed binding preferences. All the ligands bound Gal-4N with similar and low affinities, with K_D in 10^{-4} - 10^{-5} M range for the tetrasaccharides and a 7-fold difference between the best and worse ligands, while the pentasaccharides bound weaker in one order of magnitude (K_D in 10^{-3} - 10^{-4} M range) (Fig 48). Conserved binding trends were observed. For the same peripheral core structure, the B antigen was always a better binder than the A antigen, indicating that NAc group on the terminal α Gal residue was detrimental for binding. A similar effect was found for β 1-4linked antigens at the peripheral core, for which the presence of the NAc in type-2 antigens (β 1-4GlcNAc) was again detrimental for the binding with respect to type-6 antigens (β 1-4Glc). The type-1 and type-2 antigens, despite the different orientation of their reducing-end residues with respect to the protein, displayed similar affinities.

The experimental NMR observations, both from the ligand and lectin perspectives, could be nicely reconciled with the molecular models for the complexes obtained through MD simulations, where the slightly different chemical modifications among the ligands provoked subtle structural differences in the complexes, with diverse impact in the binding affinities. These data might help in understanding the binding preferences of Gal-4 towards the histo blood group antigens, especially with respect to their different presentation in distinct biological contexts (e.g. tissues) or from diverse origin (self *versus* pathogenic). In addition, it provides detailed atomic level

information and dynamics features about Gal-4N binding characteristics, relevant in the context of rational ligand design.

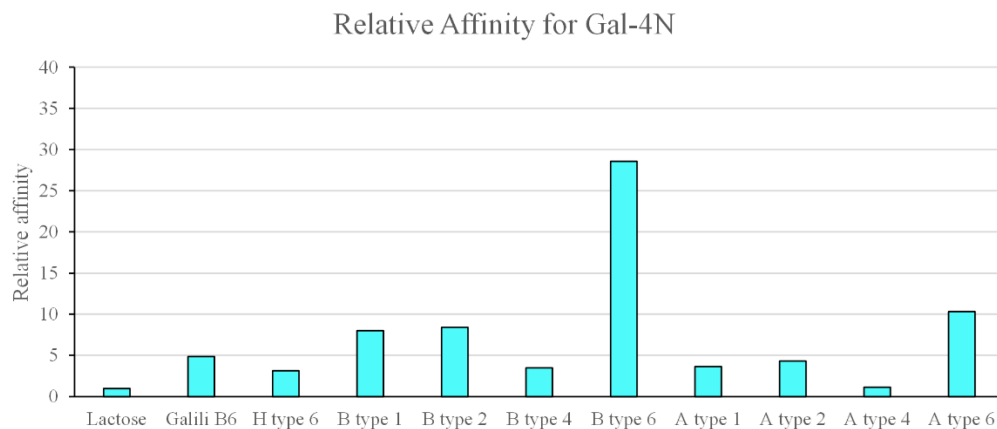


Figure 48. Relative binding affinities of different ligands for Gal-4N compared to lactose ($K_D = 1.6$ mM) as determined by ^1H - ^{15}N -HSQC based titrations.

3.1.3 The C-terminal domain of Galectin-4

The interaction between Gal-4C and the blood group A and B antigens of the different core types (Scheme 1) was analysed by using a combination of various NMR techniques, as described above for the N-terminus. The use of Saturation Transfer Difference (STD-NMR) was employed to characterize the epitope of the ligands^{130,131}. The use of ROESY experiments of the glycans in the presence of the lectin revealed specific ligand ¹H resonances in slow exchange in the NMR chemical shift timescale between the free and bound states¹⁸³. Furthermore, ¹H-¹⁵N HSQC based titrations were performed to calculate the dissociation constant¹³⁸ (K_D) for each ligand, in addition to analysing the perturbation that each ligand generated on the protein. The combination of all this data and molecular dynamic (MD) simulations provided a 3D view of the interaction of Gal-4C with each antigen.

3.1.3.1 Nuclear Magnetic Resonance

3.1.3.1.1 ¹H-STD-NMR

¹H-STD-NMR experiments were performed with a 1:50 protein:ligand ratio at 298 K, which allowed the identification of the binding epitope of each saccharide. The intensity and the profile of the STD-NMR spectra were very similar among the different ligands (See SI). Aromatic irradiation produced a much stronger STD response than that at the aliphatic region, and thus epitope mapping was carried out for the aromatic irradiation spectra. The analysis showed that, for all the ligands, no STD effect was observed for any of the protons of the Fuc and reducing-end residues, except for the Ac group whenever present. Basically, STD NMR signals were observed for protons of the central Gal residue and H1 and H2 of the terminal Gal/GalNAc residues. Intriguingly,

the STD NMR effect was particularly enhanced at the aromatic irradiation for some signals, namely, H1 and H2 of the terminal α Gal/ α GalNAc residue, H3 and H4 of the central Gal β and the Methyl of the NAc group of the terminal residue in type-1 and -4 antigens (Fig 49), suggesting that, when complexed to the lectin, these protons are close to aromatic residues. This is expected to be case for the H3 and H4 of Gal β , which according to X-ray crystallography stacks on top of the W256 indol side chain^{53,69}. On the contrary, the N-Acetyl group of GalNAc residue at the terminal end, showed a relative stronger STD at aliphatic irradiation (Fig 49), suggesting its proximity to an aliphatic residue.

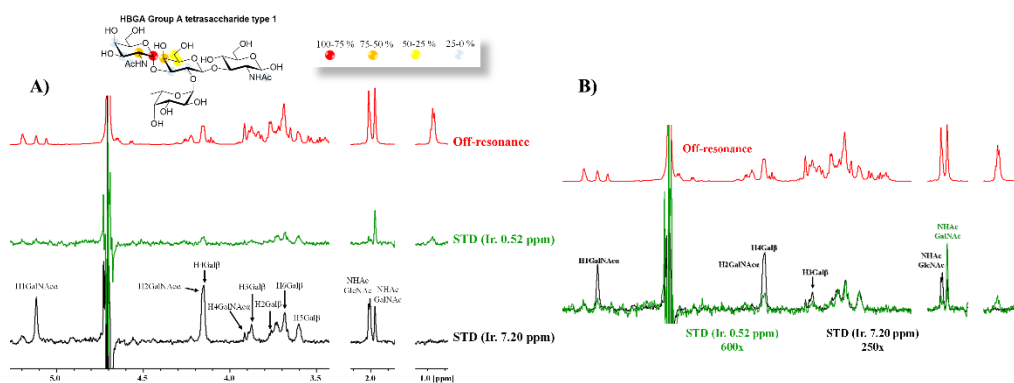


Figure 49. ¹H-STD-NMR experiment for the interaction between A type-1 and Gal-4C. A) Top: In red, the off-resonance (reference) spectra. The epitope mapping (relative STD) is shown in the ligand structure. Middle: In green, the STD-NMR spectrum upon aliphatic irradiation. Bottom: In black, the STD-NMR spectrum upon aromatic irradiation. ¹H-NMR signals showing STD effect are annotated. B) The same as in panel A), but the two STD spectra (aromatic and aliphatic irradiation in black and green respectively) are superimposed. The level of the spectrum with aliphatic irradiation is 2.4 times higher. The signals with more significant difference are annotated.

Remarkably, the acetyl region of tetrasaccharides revealed significant differences among the different core types. Whereas the N-Acetyl group of the terminal α GalNAc displayed the same intensity for the different core types presentations (always between 0.25 and 0.30 % of absolute STD), the N-Acetyl group of the reducing-end varied notably among types (Fig 50). In particular,

for the type-1 (Gal β 1-3GlcNAc) and type-4 (Gal β 1-3GalNAc) antigens, a strong STD effect was observed for the Me(Ac) at the reducing end with the aromatic irradiation, which nearly disappeared in the aliphatic irradiation. This difference between aromatic and aliphatic irradiation was less evident for the type-2 (Gal β 1-4GlcNAc) antigens. This trend suggests that for Gal β 1-3 linked antigens the N-Acetyl group at the reducing-end is closer to an aromatic group in comparison with that in the Gal β 1-4 linked antigens.

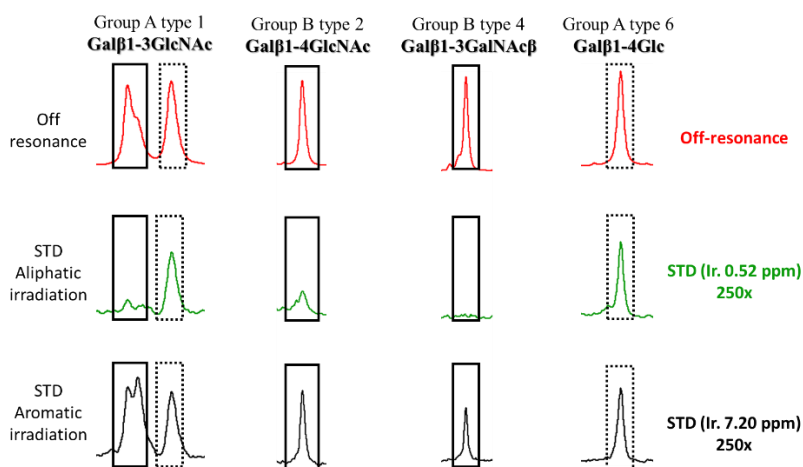


Figure 50. ^1H -STD-NMR experiments for the interaction of Gal-4C with the different antigen types. From left to right, STD-NMR experiments with A type-1, B type-2, B type-4 and A type-6. Group B Type-2 and Type-4 experiments are presented since their respected group A antigens present overlapped signals. The non-reducing terminal acetyl group (αGalNAc) are highlighted in dotted boxes, the acetyls at the reducing-end are highlighted in plain boxes. Top: In red, the off-resonance spectra. Middle: In green, the STD-NMR spectra with the aliphatic irradiation. Bottom: In black, the STD-NMR spectra with the aromatic irradiation.

3.1.3.1.2 trROESY experiments

trROESY experiments were performed with a 1:10 protein:ligand ratio at 298 K. The observed ROESY cross peaks for the bound carbohydrates were very similar to those detected for the free ligands, suggesting that no significant changes in the conformation of the ligand occurred upon binding. Fittingly,

some ligands displayed signals in slow exchange between the free and bound states (Fig 51-A). These signals are expected to arise from nuclei whose chemical environment is significantly different between the free and bound states, and the exchange rate between these two states is smaller than their resonance frequency difference. ($K_{ex}/\Delta\omega < 1$). Even if all the antigens are expected to bind with a similar affinity, only group A antigens displayed chemical exchange cross-peaks. H4, H5 and H6 of the central β -Gal showed the highest chemical shift difference between the free and bound states. Indeed, these protons are highly up field shifted in the bound state, correlating with the expected CH- π stacking with the tryptophan residue conserved among galectins at the binding site^{53,58,59}. The profile of the chemical shift difference between the free and bound states for these protons is nearly identical for all ligands, suggesting that the orientation of the β Gal residue with respect to the Trp residue at the binding site is very similar for these ligands. Interestingly, differences are found at specific protons of the monosaccharide at the reducing-end, which displayed slow exchange signals depending on the antigen type (Fig 51-B). Whereas type-2 and -4 antigens (the weaker ligands, see below) did not present any chemical exchange crosspeaks for the reducing-end residue protons, the proton H4 (of the reducing GlcNAc) of A type-1 and the protons H3 and H5 (of the reducing Glc) of A type-6 were in slow exchange, revealing that binding differences caused by the antigen type only affect the reducing end residue, whereas the rest of the ligand remains identical.

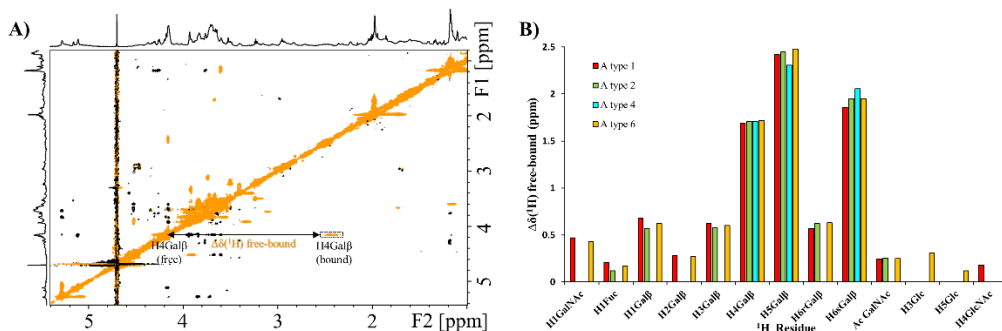


Figure 51. ¹H-NMR-based ligand chemical exchange (EXSY) analysis from trROESY experiments A) trROESY spectrum of the complex Gal-4C/A type-6 (1:10 ratio). The cross-peaks in black correspond to those arising from regular NOE effects in the rotating frame, whereas those in orange result from chemical exchange. (B) Plot for the measured differences in chemical shifts between the free and bound states ($\Delta\delta(^1\text{H})_{\text{free-bound}}$) for the protons of the antigens that suffer slow exchange in the ROESY spectrum in the presence of the lectin.

3.1.3.1.3 ¹H-¹⁵N based HSQC titrations of ¹⁵N labelled Gal-4C.

Titration of ¹⁵N labelled Gal-4C with the different blood group antigens (Scheme 1) allowed detecting the perturbation that the presence of the ligand produced on the lectin, in addition to the estimation of the dissociation constants (Table 5). In order to achieve this analysis, the NH backbone NMR assignment was required, for which triple resonance 3D NMR experiments were required. The assignment protocol allowed assigning 86% of the NH backbone resonances (Fig 52). In addition to the aforementioned eight tetra and pentasaccharides, titrations were also performed with the H antigen (devoid of the α Gal/GalNAc residue), a linear trisaccharide lacking the α 1-2 linked fucose (Galili type 6), and lactose. This analysis allowed pinpointing the impact of each monosaccharide residue in the binding events.

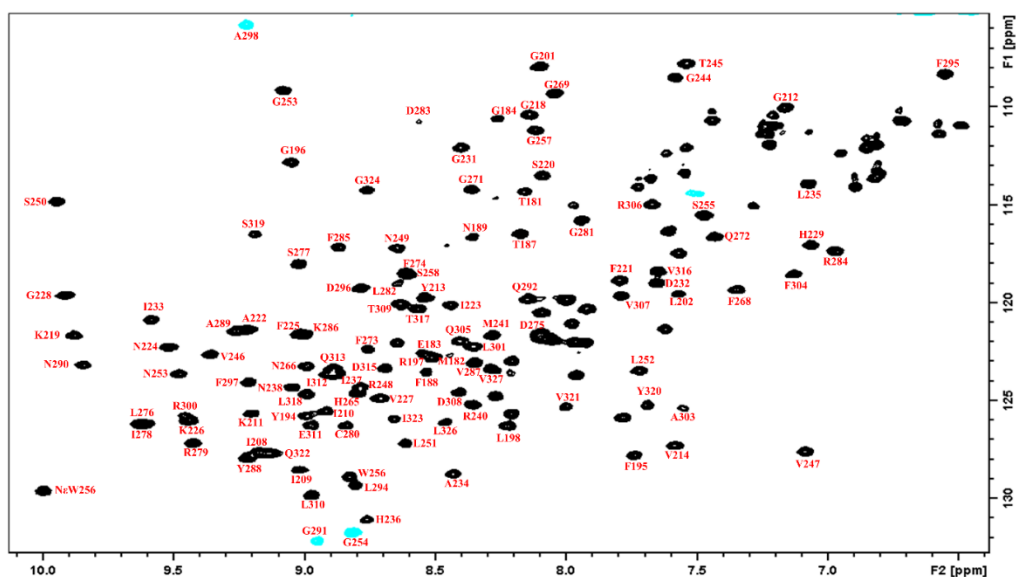


Figure 52. ^1H - ^{15}N HSQC of Gal-4C. The identity of the assigned peaks it noted next to each peak.

Titration with lactose revealed that the main residues affected by the presence of the ligand were located at strands S4 and S5, although other regions were also affected in a minor way. The spectrum of Gal-4C saturated with H type 6 is identical to that saturated with lactose (SI), suggesting that the fucose residue is far from the protein surface, in agreement with the STD NMR data. On the other hand, the presence of a galactose α 1-3 linked to the Gal β residue (Galili type 6) produced numerous new perturbations on the lectin, suggesting the proximity of this residue to the lectin surface. These perturbations are mainly in the strand S3, although residues at strands S4 and S5 also present differences (SI).

Comparison of the CSP for group A *versus* B antigens revealed significant differences in several residues (Fig 53-A). These amino acids are essentially located at strands S2 and S3, namely K219, S220, F221, A222 and Q313, in addition to the side chain of W256 (Fig 53 - B and C, and SI). These differences are conserved for every antigen type, again confirming that the same blood groups adopt the same binding modes regardless the core antigen presentation.

The titrations also revealed that the A group antigens have a two-fold better affinity for Gal-4C than the group B analogues, what could be related to additional interactions established by the N-Acetyl group, the only structural difference between A and B antigens. Blanchard and coworkers⁶⁹ have predicted, by MD simulations, the interaction between the side chain of S220 and this N-Acetyl group. However, in their model, the N-Acetyl group has a *cis* conformation. The presence of this energetically unfavourable conformation would provide a strong NOE signal between H2GalNAc and the methyl group, which is not present in our NMR spectra, thus discarding the existence of a major population of this conformation and therefore, the corresponding interaction. The different behaviour of the crosspeak corresponding to the side chain N_εH of W256 is noteworthy, since for titrations with the group A antigens, this ¹H peak shifts downfield 0.2 ppm more than in the presence of the B antigens (Fig 53-A). It is well known that proton downfield shifts are due to a loss of electron density around this nucleus (deshielding) that can take place, among other situations, when the proton acts as hydrogen bonding donor, which could be the case here. All these data suggest that the N-Acetyl group of the non-reducing end of group A antigens is oriented towards strands S2 and S3, providing intermolecular interactions with protein moieties at this site, but also with the distal W256. This fact is markedly different to the interaction with the B-antigens, which lack this moiety.

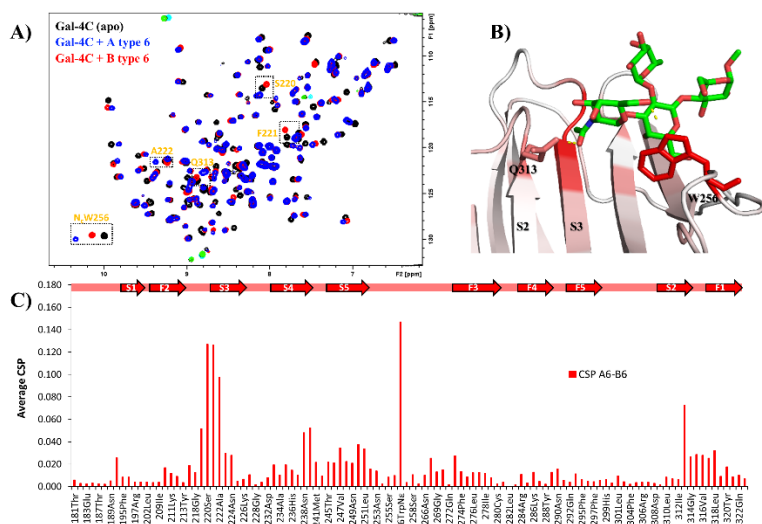


Figure 53. A) Superimposition of the ^1H - ^{15}N HSQC spectra of apo Gal-4C (black) and Gal-4C saturated with A type-6 (blue) and B type-6 (red). The peaks with strongest differences are underlined B) Molecular model for the complex of Gal-4C with A type-6. The lectin is colored according to the intensity of the CSP between Gal-4C saturated with A type-6 and B type-6. C) Difference of the Chemical Shift Perturbation of Gal-4C in the presence of A type-6 or B type-6.

The distinct types of antigens (Type -1, -2, -4 and -6) also yielded differences in CSP plots for Gal-4C (SI). For types -1, -2 and -6, the CSP were very similar, with the main differences located at strand S5, while the type-4 antigens displayed significant CSP differences, especially at strands S4 and S5. This disparity of the type-4 antigens, which are consistently the worse binders (see below) for this domain, was also observed for the N-terminal CRD of Gal-4¹⁸⁴.

In addition to the CSP plots, the ^1H - ^{15}N based HSQC titrations of Gal-4C allowed the estimation of the corresponding dissociation constants (Table 5)¹³⁸. Whereas the affinity for lactose (the minimum binding epitope) is in the millimolar range (K_D ca. 2 mM), the tetra and pentasaccharides were in the low-medium micromolar range. The analysis performed with the eight antigens yielded clear and consistent trends. The A-antigens always presented a 2 to 3 times affinity enhancement in comparison with their respective B-antigen type. Regarding the peripheral core, the type-6 was the best binder, but very similar

to the type-1. Then, the type-2 antigen is around two-fold worse binder than type -1 and -6. Finally, the type-4 antigens displayed slightly less affinity than the type-2 analogues.

Table 5. Dissociation constants (K_D) obtained for Gal-4C with the different blood group antigens by HSQC based titrations.

K_D [μ M]	Type-1	Type-2	Type-4	Type-6
Group A	86	170	185	63
Group B	190	300	458	170

3.1.3.2 3D Models of the antigens in complex with Gal-4C

Molecular Dynamics Simulations were carried out for Gal-4C (PDB: 4MY0) in bound to the eight blood group antigens. The Chemical Shift Perturbation analysis described above has revealed that the main difference in the interaction between group A and B antigens with Gal-4C takes place at strands S2 and S3, as well as at the side chain of W256 (Fig 53). Located at S2, there is a glutamine, Q313, for which MD simulations predict an hydrogen bond with the carbonyl group at the N-Acetyl group of the terminal α GalNAc and with the OH-2 group at terminal α Gal residues, which is much more persistent for group A antigens with respect to B-antigens (54% vs 9%) (Fig 54). During the simulations, the methyl group of the N-Acetyl moiety of the terminal α GalNAc residue is very close to S220, F221 and A222 in strand S3, residues that showed significant different behaviors in the $^1\text{H},^{15}\text{N}$ -HSQC titrations with A and B antigens. Additionally, the hydrogen bond between the hydroxymethyl group of the terminal GalNAc/Gal residue and the side chain of W256 appears to be more persistent along the simulation for blood group A than for B-antigen (80% vs 32%). This fits with the observed stronger downfield shift of the W256 side chain in the presence of group A antigens. Thus, the 3D models from MD

simulations for the complexes between Gal-4C domain with the different antigens are in full agreement with the experimental NMR data.

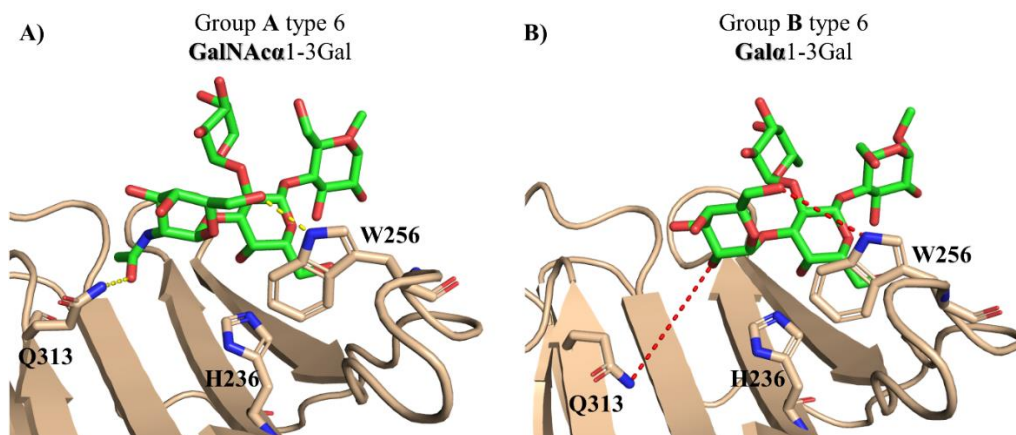


Figure 54. Models of Gal-4C in complex with A type-6 (A) and B type-6 (B). Key Q313, H236 and W256 are depicted with sticks. The feasible hydrogen bonds for A type-6 with Q313 and W256 are depicted with yellow lines. For B type-6 the interactions are depicted with red lines.

STD-NMR experiments showed that the aromatic irradiation is more effective than the aliphatic one (Fig 49). The models from MD simulations are in agreement with the observations, since the protons of α Gal and β Gal residues that are closer to W256 suffer an enormous increase of STD at aromatic irradiation. Fittingly, the N-Acetyl group of α GalNAc, which is the only proton that does not have a remarkable decrease of the absolute STD with aliphatic irradiation, is very close to the aliphatic A222, according to the MD predictions.

With respect to the situation at the reducing-end, MD simulations predict that, for the β 1-3 linked antigens (types -1 and -4), the methyl groups at the reducing-end residue are pointing towards the aromatic W256, whereas in type-2 antigens it is close to K261 (Fig 55). These orientations fit with the increased STD effect observed at type -1 and -4 antigens upon aromatic irradiation.

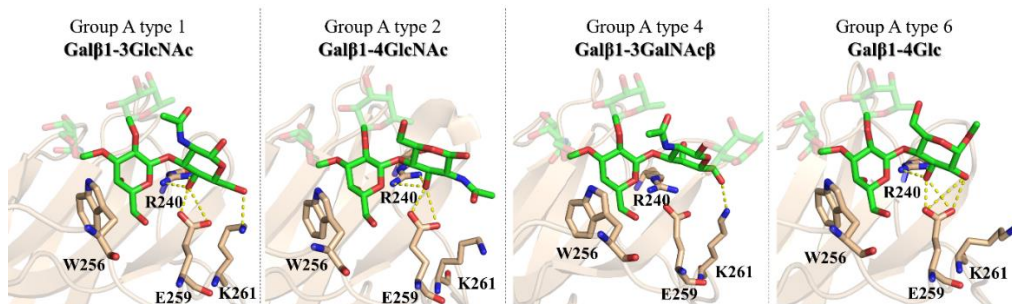


Figure 55. Molecular models for the complexes of Gal-4C and group A type -1, -2, -4, and -6 antigens according to MD simulations. Key residues are depicted as sticks and labelled.

Regarding the different affinities for the core-disaccharides, the C-terminal CRD of Galectin-4 showed a preference for type-6 antigens, a preference present also in the N-terminal CRD. MD simulations show that for β 1-4 linked antigens (types -2 and -6), positions 2 and 3 of the reducing-end residue (GlcNAc or Glc respectively) are pointing towards the lectin (Fig 55 - B and D). For type-6, both positions (OH-2 and OH-3) display hydrogen-bonding interactions with E259, while for type-2, the N-Acetyl group at position 2 does not interact with this residue, losing a hydrogen bond. At the same time, the presence of this NHAc group weakens the H-bond involving OH-3.

For β 1-3 linked antigens (type -1 and -4), positions 4 and 6 of the reducing-end residue (GlcNAc or GalNAc respectively) are pointing towards the lectin (Fig 55- A and D). In the case of type-1 antigens, the OH-4 has a similar orientation to that of the OH-3 in the β 1-4 linked antigens, resulting in similar interactions with the lectin. Additionally, the hydroxyl group at position 6 is also able to interact with K261 through hydrogen bonding interactions. For the type-4 antigens, the interaction of OH-6 with K261 side chains is also present; however, the axial disposition of the hydroxyl group at position 4 of the reducing-end GalNAc residue precludes any interaction between this hydroxyl group and the lectin, resulting in a lower affinity in comparison to the other antigens. These predicted interactions are in perfect agreement with the affinity

trends observed experimentally, where type-1 and type-6 antigens bind Gal-4C with similar affinities, stronger than the type-2 antigens, while type-4 are the worse binders. Overall, the affinity trends observed for Gal-4C are similar to those for Gal-4N, although with some minor differences.

3.1.3.3 Conclusions

Galectin-4 displays two different carbohydrate recognition domains (CRD) that bind to the same type of glycan antigens although with different affinities. In this chapter, the binding preferences of the C-terminal carbohydrate binding domain for the distinct blood group antigens with different core presentation have been addressed (Fig 56).

The STD and trROESY NMR experiments have provided information on the binding epitopes and binding modes for the different glycan antigens. STD NMR revealed similar epitopes for all the ligands, mainly involving the central β Gal and terminal α Gal/GalNAc residues, with no participation of the Fuc residue. The ROESY spectra, through chemical exchange crosspeaks, confirmed the consistent key stacking interaction of the central β Gal residue with the conserved Trp moiety at the lectin binding site for all the complexes, and anticipated important different binding modes for the diverse core types.

The ^1H - ^{15}N HSQC based titrations revealed clear and consistent tendencies in binding affinities, showing that Gal-4C has a preference for the blood group A antigens. Indeed, important differences were found in the lectin's HSQC spectra for the interactions with the A, B, and H-type antigens. These divergences affected residues at S3 and S4 strands, demonstrating that the terminal α Gal and α GalNAc interact differently with protein residues at that protein site.

The observed trends in affinity can be rationalised by the small differences predicted by MD simulations among the different complexes. The A *vs* B preference could be related to an additional hydrogen bond between the N-Acetyl group of the non-reducing end of group A antigens and the glutamine located at the strand S2 (Q313). Regarding the peripheral core, type-6 antigens were the best binders, although type-1 antigens showed a similar affinity, followed by type-2 and finally type-4 antigens, the worst binders. The lack of

hydrogen bond between E259 and OH-2 in type-2 with respect to type-6 antigens, could explain the worse affinity for the former. The axial disposition of the hydroxyl group at position 4 in type-4 antigens precludes its interaction with E259, which is present in the complexes with all the antigen types except for this one, causing a decrease in terms of affinity.

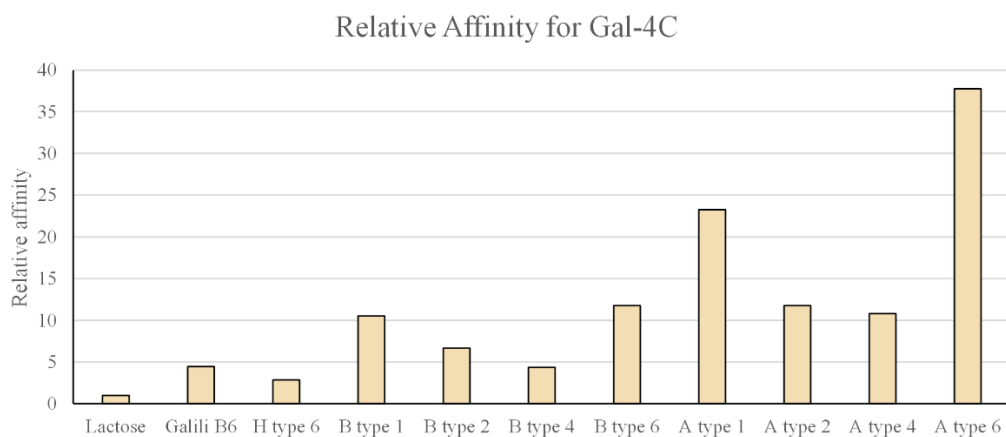


Figure 56. Relative binding affinities of different ligands for Gal-4C compared to lactose ($K_D = 2.0$ mM) as determined by ^1H - ^{15}N -HSQC based titrations.

3.1.4 Full-length Galectin-4

Galectin-4 contains two carbohydrate recognition domains (CRD) that recognize blood group antigens. In order to understand the recognition processes of this lectin, and whether there is any sort of communication between both domains, the interaction between the blood group A type-6 antigen and Galectin-4FL has been analysed. As described above, the separate domains exhibited affinities in the low μM range for this ligand, with the C-terminal CRD having a slightly better affinity in comparison with the N-terminal CRD.

3.1.4.1 Nuclear Magnetic Resonance

3.1.4.1.1 ^1H - ^{15}N HSQC based titrations

Gal-4FL displayed a crowded ^1H - ^{15}N -HSQC spectrum with many overlapped peaks. However, the careful analysis eas allowed the identification of several peaks of the separate domains (Fig 57). Interestingly, whereas for the N-domain most identified peaks remain unaffected with respect to the isolated Gal-4N protein, other peaks presented chemical shift differences. These peaks correspond mainly to amino acids located at the F1 to F3 strands, at the opposite face to the binding site, while the signals of residues at the binding site remained identical, suggesting that the binding site remains unaltered. For the C-terminal domain, many more peaks appear perturbed with respect to the isolated Gal-4C domain, in comparison with Gal-4N. However, the binding site remains unperturbed, while the significant changes correspond to residues on the F face (Fig 58, SI).

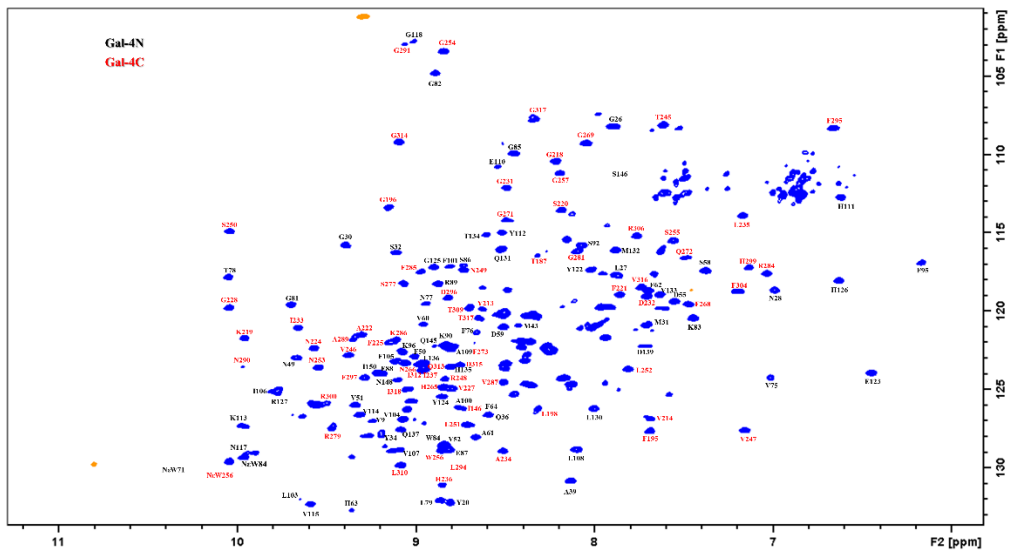


Figure 57. ^1H - ^{15}N HSQC of Gal-4FL. The identity of the assigned peaks is noted next to each peak.

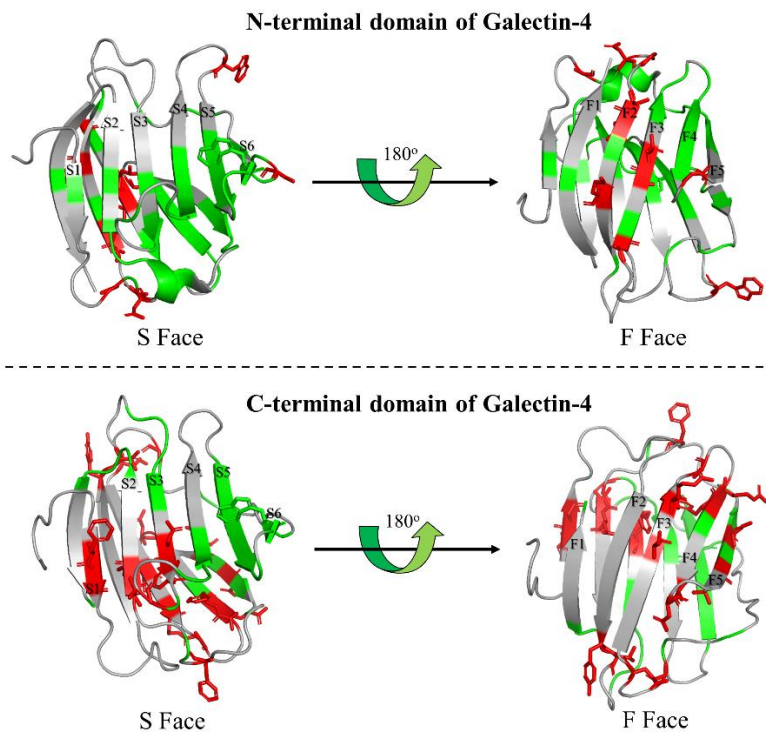


Figure 2.58. Top: Crystal structure of Gal-4N (PDB: 5DUV). Bottom: Crystal structure of Gal-4C (PDB: 4YM0). The protein is colored according to the intensity of the CSP between Gal-

4N/Gal-4C and Gal-4FL. Red: over than the standard deviation of CSP. Green: less than the standard deviation of CSP. The key tryptophan of the binding site is depicted with sticks. The name of each strand is annotated.

Titration with the blood group antigen A type-6 showed that both domains compete for the binding to the same ligand. Additionally, the chemical shift perturbation observed for Gal-4FL and the separate domains is identical, suggesting that Gal-4N and Gal-4C recognize glycans in the same way as in Gal-4FL (SI). The fact that both domains compete for the ligand in the same range of affinity, hinders the quantitative estimation of reliable dissociation constants for the isolated domains within the full-length form.

In order to decipher whether the structure of the full-length Gal-4 has an impact on the binding, two mutants of Gal-4FL were produced. In these mutants, one residue at the binding site of one of the domains was mutated so that it is not able to recognize the glycan, and only the other domain contributes to the binding. In these constructs (Gal-4FL-H63R and Gal-4FL-H236R), the key histidine of the binding site is replaced by an arginine. These mutations were designed based on galectin-13¹⁷³, for which this mutation has been reported to inhibit binding to β -galactosides while preserving the jelly roll structure of galectins.

Comparison of the ¹H,¹⁵N-HSQC of Gal-4FL-H63R and Gal-4FL-H236R with the wild type Gal-4FL revealed that there are little differences between the lectins, with these differences only affecting the mutated CRD, specifically to residues close to the mutation site, suggesting that the overall structure of the protein remains the same (SI). This fact was supported by circular dichroism (CD) experiments, which yielded identical CD spectra for the wild type and mutants of Gal-4FL (Fig 59), suggesting that the secondary structure between proteins remains fully conserved.

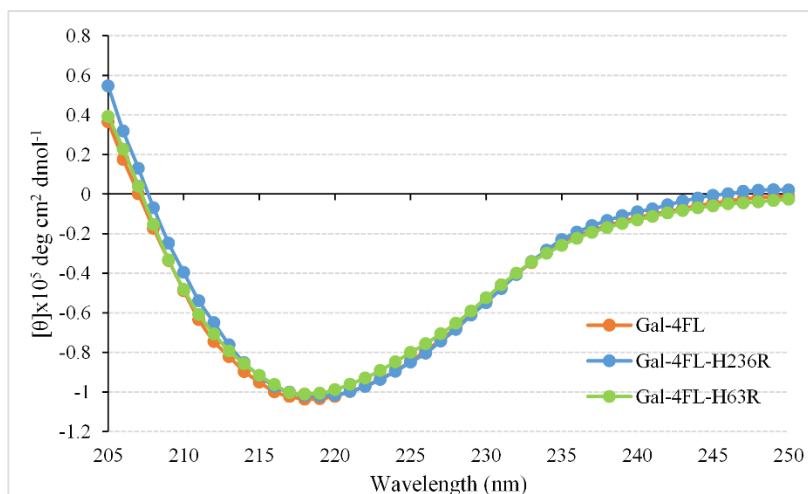


Figure 59. Circular Dichroism spectra of Gal-4FL, Gal-4FL-H236R, and Gal-4FL-H63R.

Titration of the mutants with the A type-6 antigen showed that when the histidine of the binding site (His63 in the N-terminal CRD and His236 in the C-terminal one) was replaced by an arginine, the binding to the glycan is completely abolished, as the peaks of the CRD carrying the mutation remained unaffected in the presence of the ligand (Fig. 60).

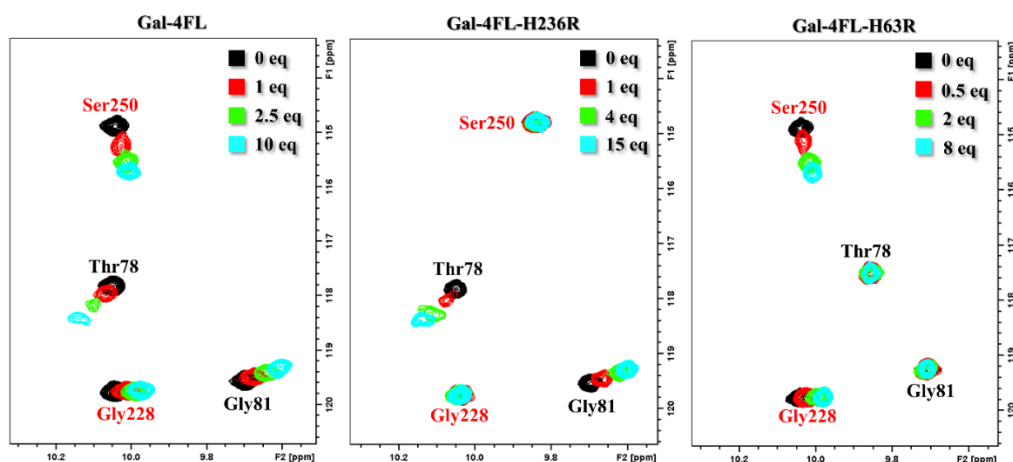


Figure 60. Region of the ^1H - ^{15}N HSQC spectra for the titrations of Gal-4FL and the mutants Gal-4FL-H63R and Gal-4FL-H236R with the A type-6 tetrasaccharide. In black, the apo form, an in red, green, and blue, upon increasing additions of the ligand. The names of the residues

located at the N-terminal CRD are written in black and the ones located at the C-terminal CRD in red. The legend on right-top of the ^1H - ^{15}N HSQC specifies the number of ligand equivalents.

The chemical shift perturbations (CSP) of the residues at the N terminal domain for the interaction of the proteins Gal-4FL, Gal-4FL-H236R and Gal-4N with the A type-6 antigen were identical (Fig 61). The same holds for the C-terminal domain, for which residues affected by ligand binding were the same for the three different proteins Gal-4FL, Gal-4FL-H63R and Gal-4C (Fig 62). Additionally, the titrations with the mutants enabled obtaining the dissociation constants and the comparison with those described above for the free domains. Gal-4FL-H63R yielded a K_D of 50 μM , close to that measured for the free domain (Gal-4C), 57 μM . The other mutant, Gal-4FL-H236R yielded a K_D of 153 μM , whereas that for Gal-4N was 154 μM .

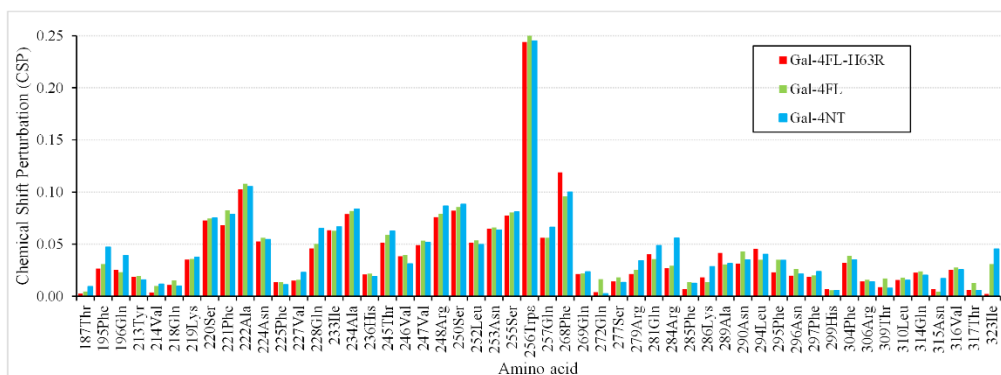


Figure 61. In red, the CSP of Gal-4FL-H236R in the presence of A type -6 (100 μM lectin, 1.5 mM ligand). Only residues of the N-terminal domain are shown. In green, the CSP of Gal-4FL in the presence of A type-6 (200 μM lectin, 2 mM ligand). Only residues of the N-terminal domain are shown. In blue, the CSP of Gal-4N in the presence of A type 6 (50 μM lectin, 0.6 mM ligand).

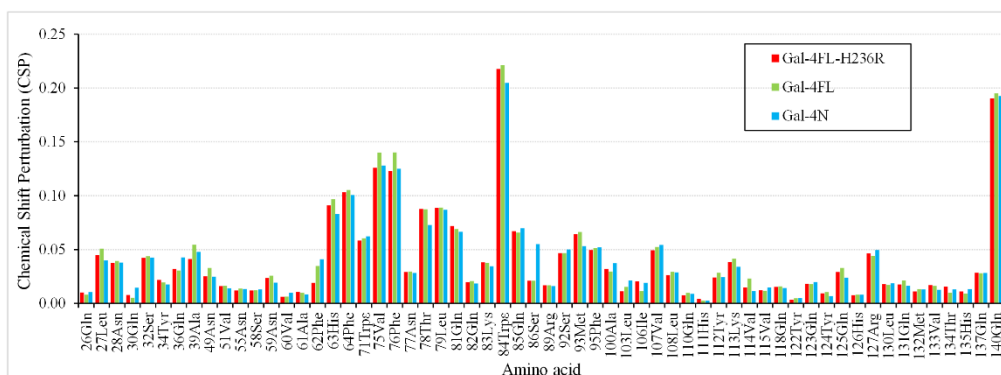


Figure 62. In red, the CSP of Gal-4FL-H236R in the presence of A type 6 (100 μ M lectin, 0.8 mM ligand). Only residues of the C-terminal domain are shown. In green, the CSP of Gal-4FL in the presence of A type 6 (200 μ M lectin, 2 mM ligand). Only residues of the C-terminal domain are shown. In blue, the CSP of Gal-4N in the presence of A type 6 (150 μ M lectin, 0.6 mM lectin).

2.1.4.1.2. ^1H -STD-NMR

The analysis followed the protocol described above for the separate domains. Thus, the recognition of A type-6 by Gal-4 was also analysed from the point of view of the ligand. ^1H -STD-NMR experiments were performed with a 1:50 protein:ligand ratio at 298 K with Gal-4N, Gal-4FL-H236R, Gal-4C and Gal-4FL-H63R. For Gal-4FL, with two binding sites, the protein:ligand ratio was 1:100, in order to keep the ratio protein-binding-sites:ligand. For all the lectins, H1, H2 and Me of the α GalNAc residue, and H4, H5 and H6 of β Gal showed the most intense STD signals (Figure 63, Table 6). The STD profiles were very similar, with subtle differences among the variants.

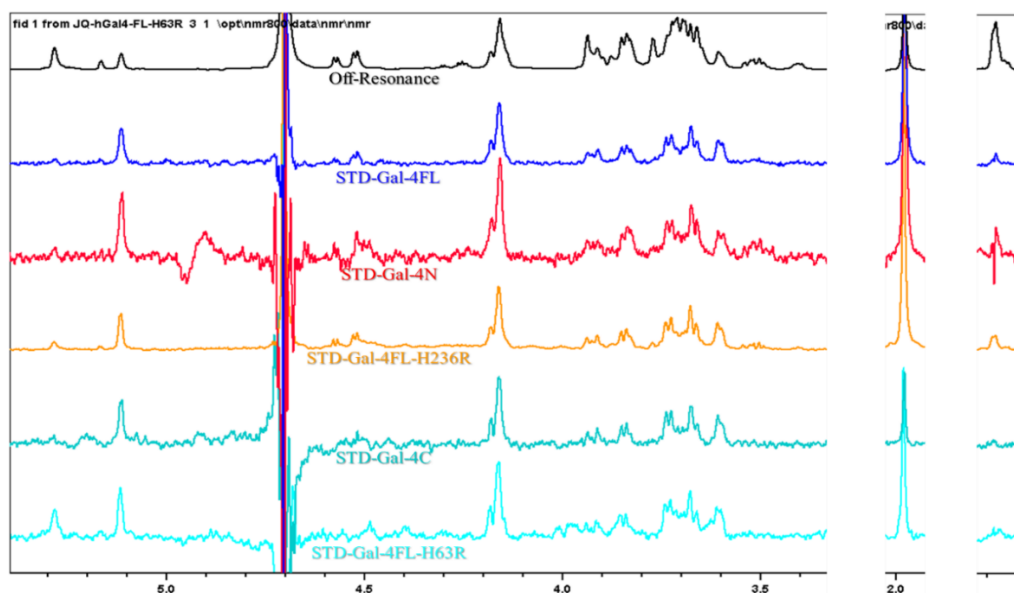


Figure 63. ^1H -STD-NMR spectra for the interaction of the different forms of Gal-4 with A type-6 tetrasaccharide. On top in black, reference spectrum. Below: in blue with Gal-4FL, in red with Gal-4N domain, in orange with Gal-4FL-H236R, in green with Gal-4C domain, and in cyan with Gal-4FL-H63R.

Gal-4N and Gal-4FL-H236R displayed identical STD profiles (Table 6), although the STD intensity was higher (3-4 times) for Gal-4FL-H236R, which can be explained by the large molecular weight of the protein. For Gal-4C, the STD profile and intensities were similar to that for Gal-4FL-H63R, with only a little difference in the N-Acetyl moiety. These spectra show that the binding epitope of the mutants is the same as in the separate domains.

Table 6. Relative STD effect (%) observed for the complexes of A type-6 with Gal-4N, Gal-4FL-H236R, Gal-4C, Gal-4FL-H63R and Gal-4FL.

Relative STD	Gal-4N	Gal-4FL-H236R	Gal-4C	Gal-4FL-H63R	Gal-4FL
H1Fuc	8%	12%	0%	43%*	8%
MeFuc	13%	11%	0%	0%	9%
H1GalNAc	85%	71%	100%	100%	83%
H2GalNAc	45%	42%	55%	57%	50%
NHAcGalNAc	100%	100%	31%	68%	100%

H1Gal β	25%	25%	0%	0%	24%
H4Gal β	39%	40%	49%	53%	47%
H5Gal β	36%	46%	57%	60%	53%
H1Glc	18%	22%	0%	0%	15%

2.1.4.2. Isothermal Titration Calorimetry (ITC)

Isothermal Titration Calorimetry experiments were performed for the interaction of the antigen with Gal-4FL as well as with the mutants and the separate domains. The analysis of the ITC data showed that Gal-4FL has more than one binding site. Indeed, the best-fit N value (number of bindings sites) was ca. 2, with a K_D of 71 μ M. Because the affinities of the two sites are very similar, the available fitting procedures are not able to provide a reliable good fitting for a two-sites model. For the isolated N- and C-domains, the obtained K_D were 98 and 26 μ M respectively.

The ITC analysis was also performed for the Gal-4FL mutants. Very interestingly, these mutants displayed basically identical thermodynamic parameters to those observed for the single domains (Table 7). This experimental evidence demonstrates that the domains act independently and that the conclusions obtained for the single domains can be completely extrapolated to the full-length Galectin-4. These results are in full agreement with the results obtained by NMR.

Table 7. Thermodynamic parameters obtained upon titration of the different Gal-4 mutants with A type 6.

	NMR	ITC		
	K_D (μ M)	K_D (μ M)	ΔH (kcal/mol)	$-T\Delta S$ (kcal/mol)
Gal-4FL		71	-7.1	1.5

Gal-4NT	154	98	-8.9	3.4
Gal-4FL-H236R	153*	93	-8.3	2.8
Gal-4CT	57	26	-8.3	2.0
Gal-4FL-H63R	50*	35	-7.9	1.8

*k_D's were measured at 35°C

2.1.4.3. Conclusions

Gal-4 is a tandem repeat protein. Thus, the behaviour of the full-length protein has been further analysed by different strategies and compared to the behavior of the separate domains described above. The ^1H - ^{15}N HSQC NMR spectrum of Gal-4FL showed that some residues have a different chemical shift in comparison with the separate domains, probably due to contacts between the two domains, or simply the effect of the presence of the linker unit. The analysis showed that the peaks of the C-terminal domain showed more differences than those of the N-terminal domain respect to the isolated moieties. Nevertheless, the observed shifts are mainly located at the opposite face to the binding site. The residues at the binding site show the same chemical shift both in the full-length and in the separate domains, strongly suggesting that the binding sites are unaffected.

Titration of the Gal-4FL with the group A type-6 antigen showed that both N- and C-terminal domains interact with this tetrasaccharide antigen, and the chemical shifts perturbations are very similar to those measured for the isolated domains. The mutations of the histidine moieties at the binding sites enabled obtaining two full-length lectins with the same structure of Gal-4FL, but with only one active binding site. STD NMR and ^1H - ^{15}N HSQC based titrations and ITC measurements showed that the mutants bind like the separate domains. This means that the conclusions described above for the separate domains (chapters 3.1.2 and 3.1.3) can be extrapolated to full-length Gal-4 (Fig 64). Basically, both domains compete for the same ligands although with different preferences, and despite being part of the same protein, they act independently.

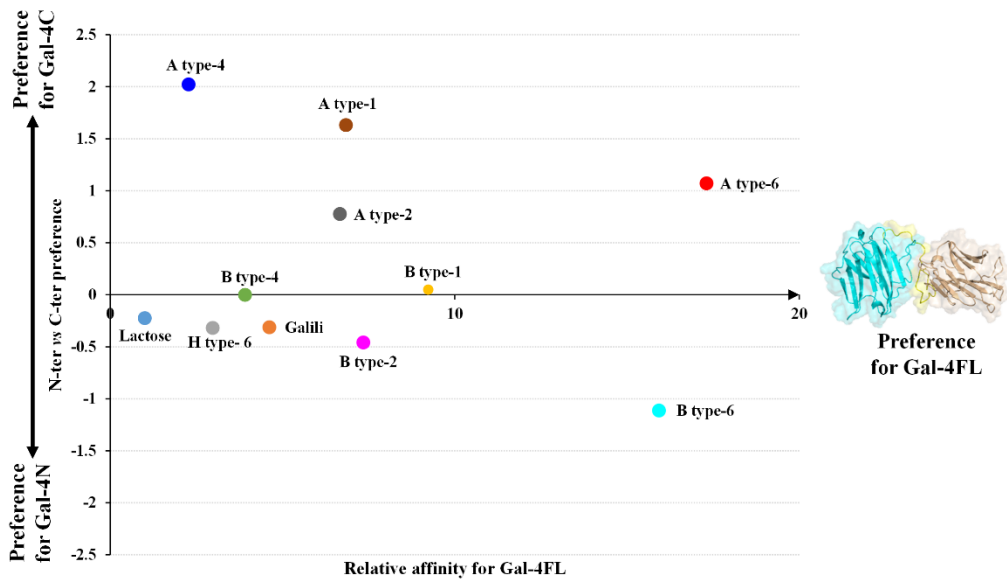


Figure 64. Summary of the binding preferences of Gal-4FL and the single domains for the different ligands tested by ^1H - ^{15}N -HSQC based titrations.

3.2 Towards understanding Galectin-4 interactions: the recognition of LPS and CD14

3.2.1. Introduction

Bacterial infections are one of the main health problems worldwide due to its high mortality rate and their increased resistance to antibiotic treatment¹⁸⁵. For both host-pathogen interactions and bacterial resistance to antibiotics, the nature of the cell envelope of the bacteria plays a key role¹⁸⁶. Diverse layers compose the cell envelope of bacteria, which helps them to protect from hostile environments¹⁸⁷. The main key element of the outer membrane of Gram-negative bacteria are the lipopolysaccharides (LPS), which are large glycolipids composed of three different parts (Fig 65): the lipid A, usually dubbed as endotoxin, the core oligosaccharide, and the O antigen^{188,189}.

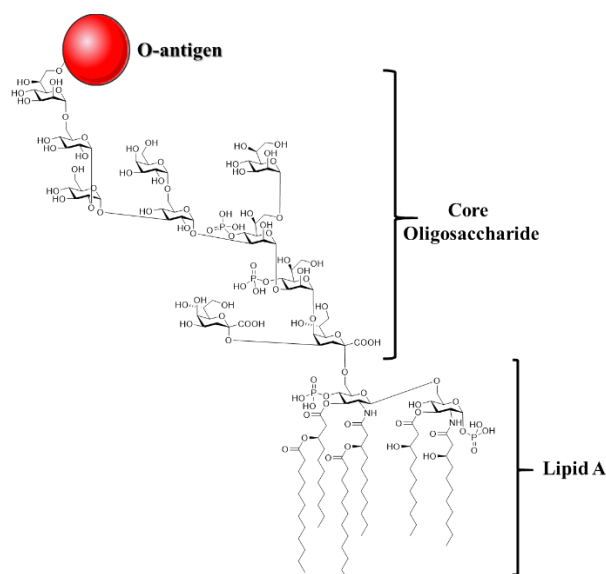


Figure 65. General Scheme of the chemical structure of the LPS from *E. Coli* K12¹⁸⁹.

Among the many bacterial components that stimulate the innate immune system, LPS are one of the strongest immunostimulating compounds, since its presence, even at picomolar levels, generates an immune response^{190,191}. The most relevant receptors that recognize LPS are the Toll-Like Receptors 2 and 4 (TLR2 and TLR4), whose LPS recognition activates the innate immune response^{191,192}. These receptors are expressed both in innate immune cells, as dendritic cells and macrophages, and in non-immune cells. It should be noted that, while TLR2 only recognizes LPS from non-enterobacteria, TLR4 is able to recognize LPS from all Gram-negative bacteria^{193,194}. For its activity, TLR4 has two co-receptors, MD-2 and CD14^{192,195,196}. The latter is a glycoprotein that is found both as a GPI-anchored protein and in a soluble form. Binding of CD14 to the endotoxin followed by the consequent shuttling of LPS to the TLR4/MD-2 complex greatly enhances the immune response (Fig 66) through activating different signalling pathways¹⁹⁷.

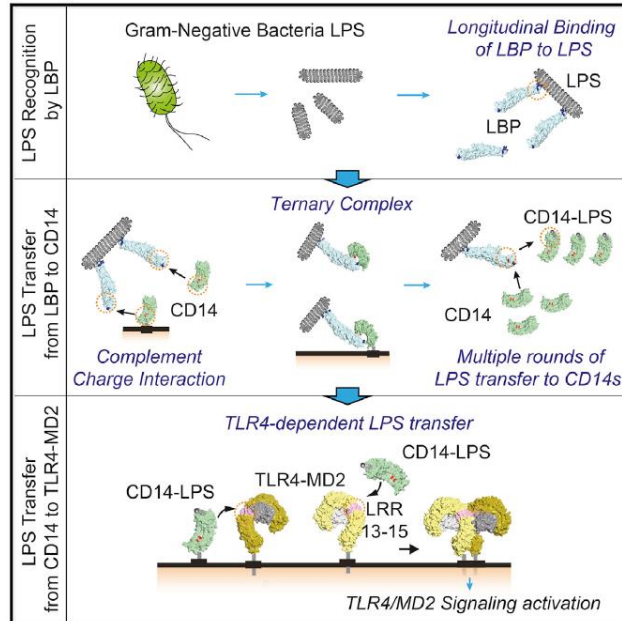


Figure 66. Model of the cellular response to bacteria LPS mediated by CD14, Lipopolysaccharide Binding Protein (LBP), TLR4, and MD-2. Taken from ¹⁹⁷

In this context of bacterial infections, it has been proposed that not only TLRs, but also galectins play a defensive role, since some of them have shown bactericidal activity^{90,91,93}. It is well documented that various bacteria contain blood group antigen-like carbohydrates in their O-antigens, which aim to deceive and evade the human immune system⁹². Whereas our immune system has specific antibodies against those blood group antigens that do not belong to our own blood group, it lacks of specific defence for pathogens that mimic our own ones. Multiple studies have shown that some galectins are able of recognizing blood group antigens^{53,56,58,59,184}. Additionally, some of them, specifically galectins 4, 8, and 9, which belong to the tandem-repeat galectin family, exhibit bactericidal activities towards bacteria that mimic these antigens^{90,93}. Intriguingly, prototype galectin-1 and chimeric galectin-3, which are also capable of binding blood group antigens, do not display any activity against these bacteria^{56,58,90}.

Very interestingly, Hong et al. showed that one of the galectins with bactericidal activity, human galectin-4, directly interacts with the main co-receptor of bacterial LPS, CD14, although not with TLR2 or TLR4⁹⁴. This interaction is able to modulate the morphology and induce functional changes of CD14⁺ monocytes. In the presence of Gal-4, CD14⁺ monocytes differentiate into macrophage-like cells. Additionally, the presence of Gal-4 alters the expression levels of cytokines and TLRs, generating a similar effect on cells to that of LPS. Although the impact of this interaction in the immune response to bacterial infection is yet to be ascertained, it has already been detected that Gal-4 has a relevant role in the immune response against bacteria in other species other than humans^{198,199}.

The interaction of Gal-4 with CD14 is likely based on the recognition of the N-glycans of CD14 by the lectin, since the presence of lactose decreased the effect of Gal-4 on CD14⁺ monocytes⁹⁴. This fact taunts the glycosylation of CD14 as an interesting target. There are four potential N-glycosylation sites in human CD14 according to the N-glycosylation consensus sequence: Asn37, Asn151, Asn282, and Asn323. Meng et al. have studied the impact of the glycosylations by inserting point mutations in every feasible N-glycosylation site²⁰⁰. In order to determine whether a potential site was glycosylated, the shape of the bands of CD14 mutants in the supernatants of insect cells transfected with CD14 was compared with the WT CD14 by Western blot experiments. Their analysis shows that mutations at positions Asn151 (N151) and Asn282 (N282) results in decreased height of the bands in the Western Blot, a sign of reduction of heterogeneity²⁰⁰. This data suggests that these positions are glycosylated since N-glycans are usually very heterogeneous. Moreover, the same authors studied the impact of the N-glycosylations on the stability and function of CD14, showing that N282 enhances protein stability, while N151 is relevant for the secretion of the protein and its interaction with LPS. The evidence that those two positions are glycosylated is supported by the analysis of the X-ray

structure of mouse CD14 (PDB 1WWL), which contains the four potential glycosylation sites and only those two are glycosylated²⁰¹. There is also a X-ray structure of human CD14, which has a very similar structure to mouse CD14²⁰². In order to crystallize CD14, Kelley *et al.* treated the protein with PNGase F, a glycosidase that cleaves every type of N-Glycan²⁰³. In the X-ray structure, no glycans were observed, however, the resolution of the crystal structure was not optimal (4.0 Å), while the SDS-PAGE of the protein showed a wide band, typical for glycoproteins. Thus, the presence of glycans cannot be discarded. Additionally, through mass spectroscopy studies, CD14 has been detected in human serum, with sites N151 and N282 found to be glycosylated²⁰⁴.

However, the role of N-glycans for the biological function of CD14 remains unclear regarding the immune response. Whereas Meng *et al.* claimed that the N-glycan at N151 is critical for the response to LPS, it has been reported that a shorter fragment of CD14 (amino acids 1-152), without N-glycans, is also able to recognize LPS²⁰⁵. Additionally, CD14 can also promote macrophage activation mediated by TLR2, for which CD14 glycans could be relevant. For instance ArtinM, a tetrameric lectin from *Artocarpus integer*, interacts with both CD14 and TLR2 by recognizing their N-glycans and subsequently activating macrophages²⁰⁶. These events induce an inflammatory profile in macrophages, which could be the result of a bridging between CD14 and TLR2. Therefore, the detailed analysis of the nature and interactions of the CD14 N-glycans could help understanding how CD14 exerts some of its functions. For instance, it has been shown that core fucosylation on CD14 is vital for the CD14-dependent TLR2 and TLR4 signalling response in macrophages^{207,208}. Moreover, diverse TLR agonists have been designed to enhance immune responses and, since CD14 mediates interactions with various TLRs, deciphering its glycoprofile could be key for developing new therapeutics²⁰⁹⁻

213.

Thus, galectin-4 and CD14 are able to interact with bacterial LPS and these interactions generate an immune response. Therefore, the study of the structural features that drive these molecular recognition events is of major interest. Our previous findings have allowed us determining how gal-4 recognizes the chemical structures present in blood group antigens¹⁸⁴. Many LPS present long oligosaccharides mimicking these structures in their O-antigen. Therefore, it is highly likely that Gal-4 bind them in a similar way. However, the impact of the multivalent presentation of LPS in the binding event is yet to be ascertained. Herein, we have focused on revealing the specific details of the binding events among the three partners: First, the interaction of Gal-4 to the LPS from *E. Coli* O55 by using NMR methods. Additionally, we have also defined the CD14 glycoprofile using NMR and mass spectrometry, the structural and conformational relevance of these glycans, and deciphered the essential structural motifs that participate in its binding to Gal-4. Finally, we have also unravelled key features of the encounter between LPS and CD14.

3.2.2. The glycoprofile of CD14 as deduced by NMR

CD14 is a glycoprotein with four potential N-glycosylation sites and one O-glycosylation site (Fig 67-A). Based on the work of Kelly *et al*,²⁰² CD14 (aa 19-335) was expressed, fused to the Fc of IgG1, including a TEV site between CD14 and the Fc and a His-tag at the C-termini (Fig 67-B). Although this construct yielded high expression levels of the protein (over 70 mg/L of culture), the yield of the cleavage of the Fc was very low, needing over 2 weeks of incubation with TEV protease at room temperature to yield of 2 mg/L of culture of cleaved CD14. As a result, a spacer was introduced between the CD14 and the cleavage site, which was substituted by a thrombin site (Fig 67-C). The final yield with this construct was 27.5 mg/L of culture of cleaved CD14 after overnight incubation at room temperature with thrombin.

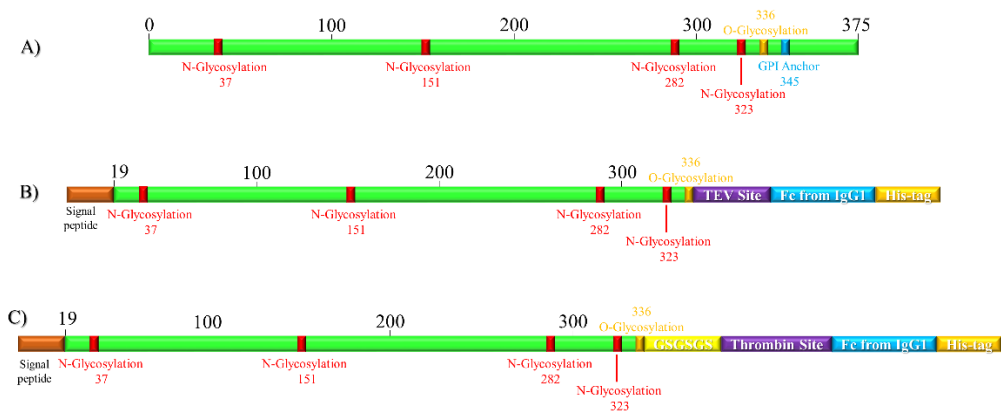


Figure 67. Sequence of human CD14 and the main constructs employed in this Thesis.

The glycoprofile of CD14 produced in HEK293F cells was determined using the NMR-based approach recently established in our lab^{214–216}. The addition of ¹³C labelled glucose (Glc) to the media allowed the ¹³C labelling of the glycans covalently attached to CD14. In order to identify those glycans, 3D H',CH NOESY-HSQC, H',CH TOCSY-HSQC, and [¹³C,¹³C]TOCSY-HSQC experiments were performed. The combination of these experiments allowed decoding the spin-systems for each monosaccharide residue, as well as the connectivity between them. As a result, the different glycoforms present in the intact CD14 were unambiguously revealed (Fig 68).

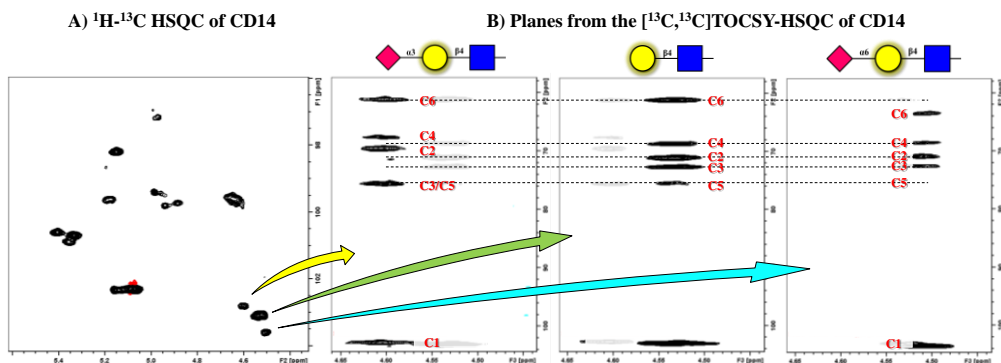


Figure 68. A) Anomeric region of the ^1H - ^{13}C -HSQC of CD14. B) Different planes from the 3D ^{13}C , ^{13}C TOCSY-HSQC of CD14.

The analysis of the data showed the presence of multiple mannose (Man) residues, both in high-Man and complex type N-glycans (Fig 69). Indeed, NMR signals for Man₉ and Man₅ (high-Man), as well as for the Man₃ core (complex type) were observable in the anomeric region of the ^1H - ^{13}C -HSQC spectrum. The Man₃ core detected had always attached β 1,2- linked GlcNAc residues. These GlcNAc residues presented β 1,4-linked galactose (Gal) units, conforming the LacNAc epitope. This epitope was also present with three different substitutions, although the unsubstituted terminal LacNAc was the most abundant glycoform. α 2,3- and α 2,6- linked sialic acid moieties were present, in a similar proportion between them. Finally, a small signal arising from LacNAc with an α 1,3- linked fucose (Fuc) to the GlcNAc (Lewis X epitope) was also detected. Interestingly, although it has been claimed that CD14 expressed in HEK293T cells carries core Fuc²⁰⁸, this epitope was not detected by NMR.

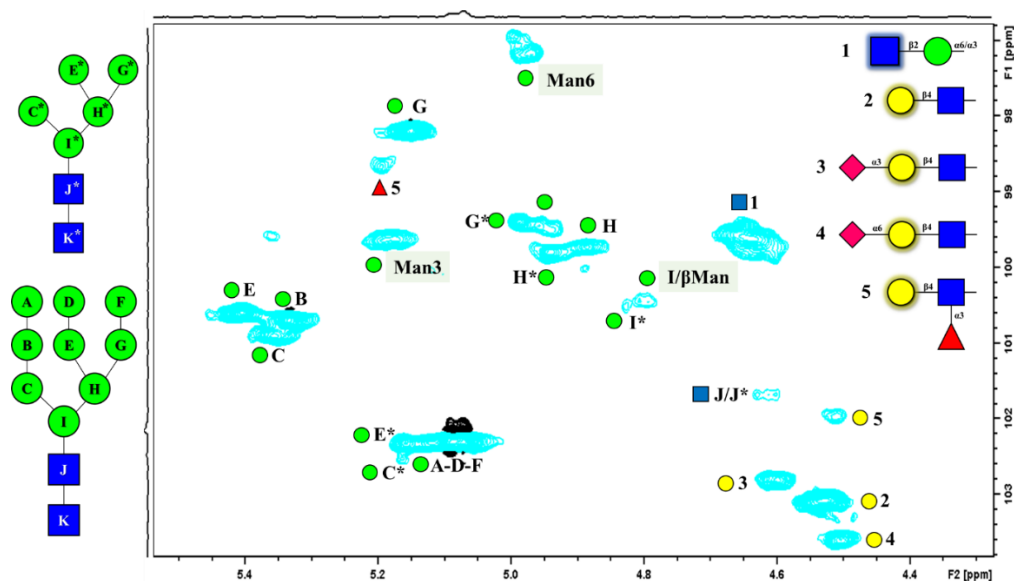


Figure 69. Anomeric region of the ^1H - ^{13}C HSQC NMR spectrum of glycosylated CD14. The identification of each peak is done with a SNFG symbol representing the corresponding monosaccharide and a letter/number relating it to a specific glycan epitope/structure. The mannoses with a light green background correspond to the Man_3 core of complex glycans.

3.2.3. Site-specific determination of the glycans by MS

The glycosylation pattern of CD14 was also determined using mass spectroscopy as described in the experimental section. Despite the usefulness of NMR to determine the glycan types present in a protein, no information of the site specificity of each glycan can be obtained. In contrast, MS has shown to be very efficient to determine the whether a potential glycosylation site is glycosylated, and can reveal the composition of those glycans (hexose, deoxyhexose, N-Acetylhexosamine, and sialic acids)²¹⁷. On the other hand, MS does not provide information of the type of linkage and branching of the sugar.

In this case, treatment of the samples with different enzymes was carried out to obtain different peptide fragments of CD14. The use of trypsin allowed identifying peptides containing the glycosylation at N151, whereas Glu-C protease allowed identifying the peptides containing the N282 glycosylation. Additionally, both enzymes also allowed identifying the glycans of the fused

Fc fragment of the IgG1 employed for purification purposes, yielding information in full agreement with the reported values for this glycoprotein²¹⁸.

For the non-exposed N282 site, only Man₉ (76.5%) and Man₈ (23.5%) glycans were detected. For the exposed N151 site, a large microheterogeneity was detected (Table 8). The major glycoform was Man₅ (26.8%). Nevertheless, multiple glycans containing LacNAc epitopes were also detected, along with mono sialylated glycans. Fucosylated glycan are also present, although in small quantities. Overall, complex type N-glycans are the most abundant N-glycans types in this site. Globally, these data are in agreement with the NMR results.

Table 2.8. Percentage of occupation of each glycoform at the position N151 of CD14 determined by mass spectroscopy. Glycoforms with less than 3% of occupancy are not shown.

Glycoform	Occupation (%)
HexNAc(2)Hex(5)	26.8
HexNAc(4)Hex(5)Neu5Ac(1)	11.4
HexNAc(4)Hex(4)Neu5Ac(1)	8.7
HexNAc(4)Hex(3)	5.5
HexNAc(5)Hex(4)	5.5
HexNAc(5)Hex(3)	5.0
HexNAc(4)Hex(4)	4.2
HexNAc(4)Hex(5)Fuc(1)Neu5Ac(1)	3.0

3.2.4. Glycosidase treatment to determine glycosylation profiles.

In order to determine the specific glycosylation site for the Man₉ and Man₅ N-glycans, the folded and ¹³C-labelled CD14 was treated with endo-H, a glycosidase able to cleave high-Man- and hybrid-type N-glycans²¹⁹, but not complex type N-glycans (Fig 70). Thus, ¹H-¹³C-HSQC NMR spectra of the glycoprotein were recorded before and after treatment with endo-H (Fig 70-

A,B). The obtained sample, containing CD14 and endo-H, was filtered using a 10 kDa cut-off filter to remove the cleaved glycans, and a new ^1H - ^{13}C -HSQC spectrum of the obtained glycans was also acquired (Fig 71-C). Treatment of CD14 with endo-H resulted in the specific cleavage of only the Man_5 glycans, indicating that the Man_9 glycan at N282 is not accessible to the human mannosidases. Therefore, the lack of accessibility of the site N282 conditions the observed glycosylation pattern.

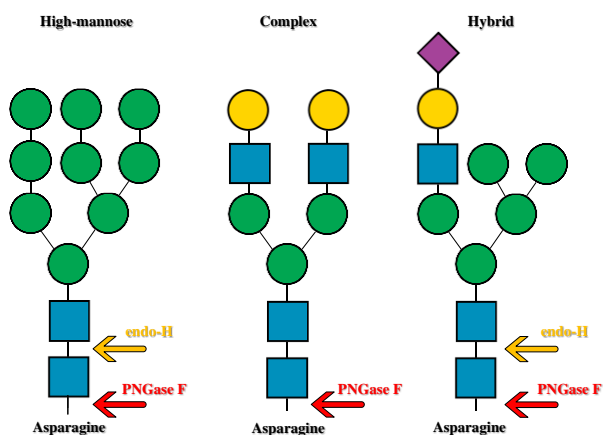


Figure 70. The three type of N-glycans and the cleavage point of glycosidases endo-H and PNGase F.

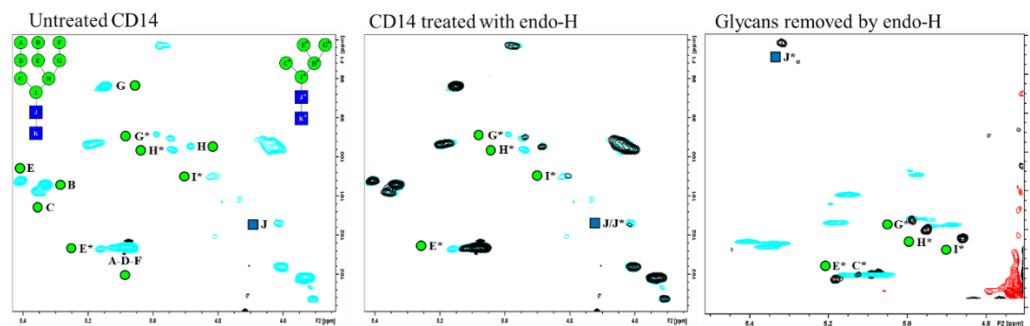


Figure 71. A) Anomeric region of the ^1H - ^{13}C -HSQC of CD14. B) Superimposition of the anomeric region of the ^1H - ^{13}C -HSQC of CD14 before (blue) and after (black) treatment with endo-H. C) Superimposition of the anomeric region of the ^1H - ^{13}C -HSQC of CD14 (blue) and of the glycans removed from CD14 by endo-H (black). Note that the spectral window is different to show peak corresponding to the reducing-end residue (J^*). The peaks corresponding to signals that could be affected by cleavage of endo-H are annotated.

3.2.5 Towards a 3D-model of glycosylated CD14

Using the X-ray crystallographic structure of CD14 deposited in the PDB (code 4GLP), models of glycosylated CD14 were built using the glycoprotein builder module of GLYCAM²⁰¹. It is well known that the first step of the N-glycosylation process involves the attachment of the tetradecasaccharide Glc₃Man₉GlcNAc₂ to the unfolded protein²²⁰. Later, the Glc residues of this glycan are trimmed during protein folding, yielding the initial Man₉GlcNAc₂, which can then be further modified^{18,220}. Therefore, in the generated model, the Man₉GlcNAc₂ N-glycan was first placed in both glycosylation sites, N151 and N282. Then, MD simulations of this glycoprotein were carried out, which showed that the N-glycan at position N151 is exposed to the solvent, whereas that at N282 is buried into the concave side of the protein (Fig 72-A). Indeed, the terminal Man residues of the α 3 and central arms, which are the first residues cleaved by human mannosidases upon protein maturation, display multiple hydrogen bonds with Glu and Asp residues at the acidic patch located in this side of the protein (Fig 72-B, SI), and are not solvent accessible¹⁸. This fact reinforces the idea that the lack of accessibility of the initial Man₉ glycoform in the N282 glycosylation site to glycosidase enzymes could be the reason behind the presence of the full Man₉ glycoform in CD14²¹⁴.

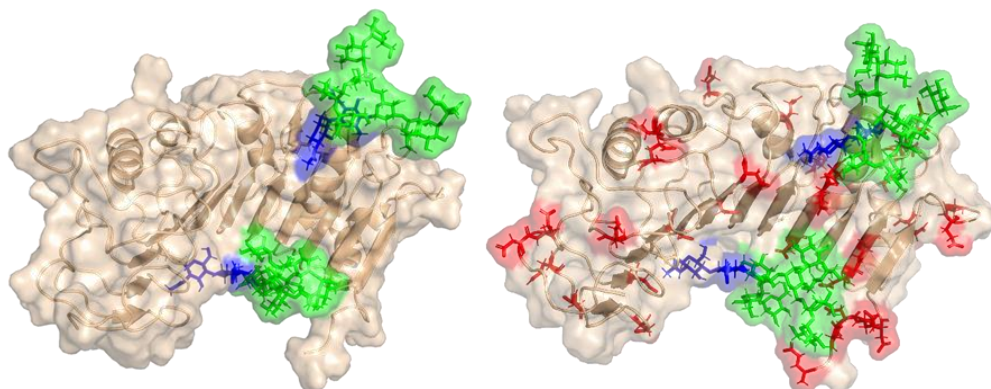


Figure 72. A) Our model of CD14 containing Man₉ N-glycans. B) The model showing the acidic residues depicted with sticks and colored in red.

As deduced by the mass spectrometry analysis, the major N-glycan type at N151 belongs to the complex type. Therefore, we also performed MD simulations of CD14 with Man₉ at Asn 282 and with complex N-glycans at Asn 151. This simulations show that the LacNAc/3'SLacNAc epitopes are fully exposed to the solvent (Fig 73).

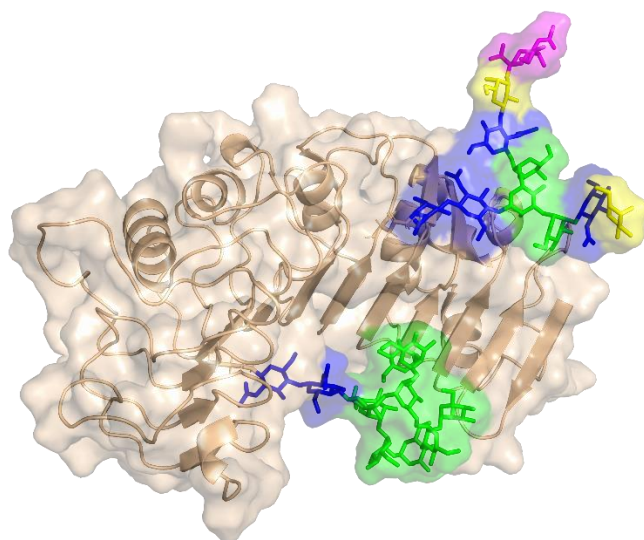


Figure 73. Our model of CD14 containing Man₉ N-glycans at position N282 and a complex type glycan at position N151.

3.2.6 Impact of the N-glycosylation on the stability of the protein

The impact of each N-glycosylation site on CD14 was then evaluated by generating two point mutations, which were inserted in two different constructs, N151Q and N282Q. The yield of protein on the supernatants of HEK293F cells transfected with CD14-Fc, CD14-Fc-N151Q, and CD14-Fc-N282Q was quantified by Western Blot analysis, using an anti-Fc antibody (Fig 74). Interestingly, the expression levels of the mutants showed drastic changes compared to those previously reported²⁰⁰. It has been reported that mutation at N151 dramatically decreased the secretion level of CD14, suggesting that it is employed as a secretion or sorting signal. However, the point mutation N151Q showed just a subtle decrease in protein expression (ca. 18 %). Moreover, these authors²⁰⁰ also reported that mutation at N282 did not affect the secretion levels of CD14. Nevertheless, our findings show that mutation at this site results in a dramatic decrease (over 80%) in the secreted protein.

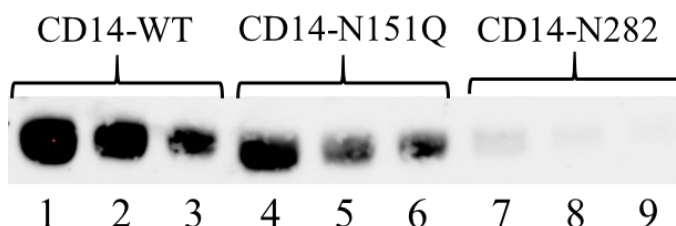


Figure 74. Western blot of HEK293F cells transfected with different proteins. Lines 1, 2, and 3 correspond to the cells transfected with CD14-Fc. Lines 4, 5, and 6 to CD14-FC-N151Q. Lines 7, 8, and 9 to CD14-Fc-N282Q. The volume of supernatant of HEK293F cells loaded was 20, 10 and 5 μ L for each protein respectively.

The obtained proteins from the supernatants were purified through affinity and gel filtration chromatography. Whereas the wild-type CD14 was mainly (97%) well folded, the mutants did not show the same behaviour (Fig 75 & Table 9). Intriguingly, despite not having a significant impact in the secretion level, the

N151Q mutation largely affected to the protein stability, since ca. one third of the protein was aggregated. This observation strongly suggests that the N151 N-glycan is key for the stability of the protein. In line of the observed decrease in the levels of secreted protein, the CD14-Fc-N282Q variant did not show proper folding and eluted completely aggregated, showing the key relevance of this N-glycosylation for the stability of CD14.

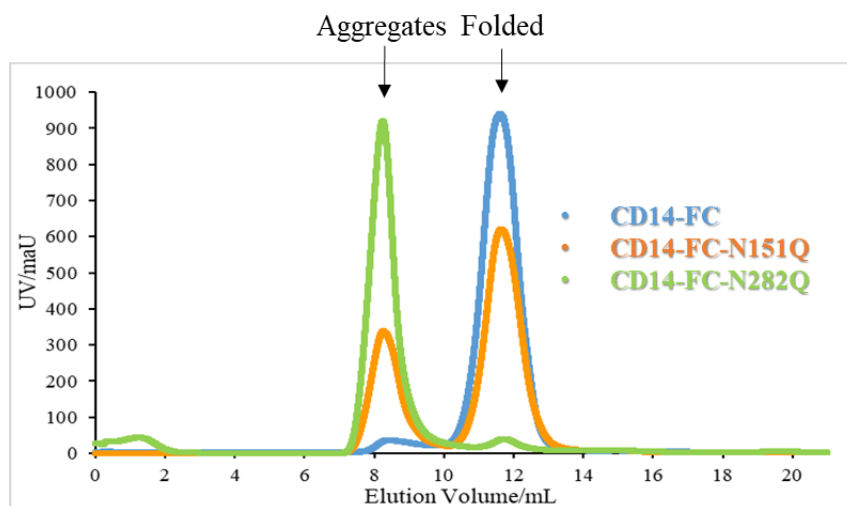


Figure 75. Profile of CD14 and its mutants in a Superdex 200 Increase 10/300 column. The profile of CD14-Fc is in blue, the one of CD14-Fc-N151Q in orange and the one of CD14-Fc-N282Q in green.

Table 2.9. Percentages of each of the folding states of CD14 and its mutants.

Percentage	CD14-WT	CD14-N151Q	CD14-N282Q
Aggregates (%)	3	32	95
Folded (%)	97	68	5

3.2.7 The interaction of CD14 with human Galectin-4

In addition to its role for the protein stability and folding, the N-glycans of CD14 could play a role in interaction events. Indeed, Hong et al. have reported that the presence of lactose decreased the effect that Gal-4 generates in CD14⁺ monocytes, suggesting that the CD14/Gal-4 interaction is on cargo of the N-glycans⁹⁴. Thus, the CD14/Gal-4 molecular recognition event was studied by NMR, both from the perspective of the CD14 glycans and of the galectin.

Blood group antigens are the preferred ligands for both domains of Gal-4^{69,184}. Although CD14 lacks these epitopes, it contains three β -galactosides: LacNAc (LN) and 3'- and 6'-sialyl-LacNAc (3'SLN and 6'SLN). Our previous findings showed that Gal-4 binds LacNAc in the mM range¹⁸⁴, however, there is some controversy regarding the binding of Gal-4 to sialylated epitopes^{61,221–223}. Thus, in order to identify the glycan epitopes on CD14 recognized by Gal-4, ¹H-¹³C-HSQC experiments were carried out for the ¹³C-labeled glycans of CD14 in the presence of Gal-4FL, Gal-4FL-H236R, and Gal-4FL-H63R and compared to those obtained for the apo glycoprotein. The addition of Gal-4FL induced chemical shift perturbations and intensity decrease for the HSQC cross peaks belonging to LN and 3'SLN. These facts suggest that the binding of the lectin to the glycoprotein is relatively weak (Fig 76-A). Gal-4FL-H236R, in which only the N-domain is active for sugar binding, induced similar changes in the glycoprotein, suggesting that the N-terminal domain binds both epitopes (Fig 76-B). However, Gal-4FL-H63R, in which only the C-terminal is active, exclusively perturbed the signals of LN, and to a less extent compared to the other galectin variants (Fig 76-C), indicating that 3'SLN is only recognized by the N-terminal domain. Note that signals arising from 6'SLN are not affected by the presence of any of the Gal-4 proteins. Since the epitopes affected by the presence of Gal-4 are only at the N151 glycosylation site, the glycan-mediated interaction between Gal-4 and CD14 takes place on the convex side of CD14.

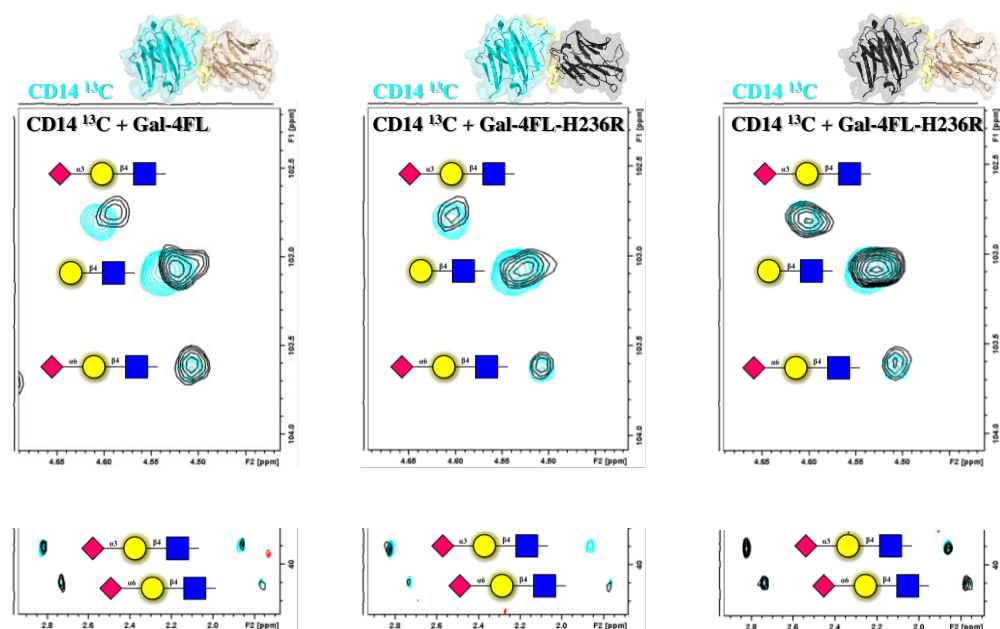


Figure 76. Zoom at specific regions of the ^1H - ^{13}C HSQC spectra of the ^{13}C labelled glycans of CD14 in light blue. The corresponding spectra in the presence of Gal-4FL (left, 1:2.8 CD14:lectin ratio), Gal-FL-H236R (center, 1:1.6 CD14:lectin ratio), and Gal-4FL-H63R (right, 1:2 CD14:lectin ratio) are superimposed in black. The top panels show the anomeric region, while the region of the H3 signals of sialic acids is displayed at the bottom panels.

The interaction was also observed from the point of view of the lectin. In particular, ^1H - ^{15}N -HSQC spectra of the ^{15}N labelled lectins (25 μM) were recorded, both in the free state and in the presence of 200 μM of CD14 (Fig 77). In this case, single domains of Gal-4 (Gal-4N and Gal-4C) were employed for better spectral quality in separate experiments, as well as Gal-4FL. For both Gal-4 isolated domains, a clear decrease of the signal intensities of many cross peaks was observed, which were not only localized in the binding sites, but throughout the entire protein (Fig 77). The average decrease in the signal intensities was more pronounced for the N-terminal domain. In particular, the Gal-4N spectrum lost an overall of 55% of the signal intensity, whereas for Gal-4C, ca. 35% of the average intensity was lost upon addition of CD14. The analysis of the same experiments carried out for full-length Gal-4 allowed

deducing a larger overall decrease in intensity, although the differences between the individual domains was smaller: ca. 63% decrease for the N-terminal CRD and 53% for the C-terminal one.

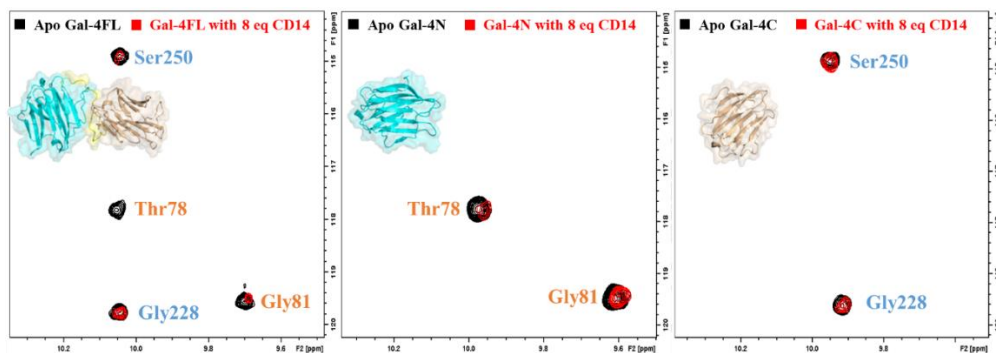


Figure 77. Key region in the ^1H - ^{15}N HSQC spectra of 25 μM of the galectin variants in the apo (black) form and in the presence of 200 μM of CD14 (red). From left to the right: Gal-4FL, Gal-4N and Gal-4C.

3.2.8 The interaction of Galectin-4 with LPS

The interaction of Gal-4 with LPS was analysed using the full LPS of *E. Coli* O55, whose saccharide repeating unit is composed by colitose, Gal, GalNAc, and GlcNAc (Fig 78), and thus mimics the blood group antigens¹⁷². Colitose (3-deoxy-L-fucose²²⁴) is a dideoxyhexose found in Gram-negative bacteria. This *E. Coli* O55 antigen is very similar to the H type 1 antigen (Fig 78). Indeed, according to our previous findings (see above), the lack the O3 hydroxyl group in colitose should not affect its recognition by Gal-4¹⁸⁴.

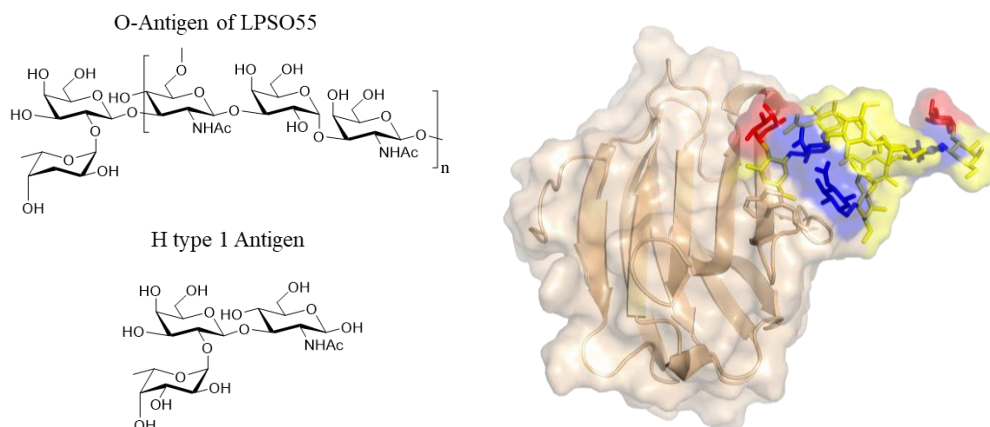


Figure 78. Left top: the O-antigen of the LPS from *E. Coli* O55. Left bottom: the Blood Group Antigen H type 1. Right: Model of Gal-4C bound to the *E. Coli* O55 O antigen. Fuc has been employed instead of colitose. Each monosaccharide has been colored according to its SNFG color (Fuc in red, GlcNAc in blue, and Gal/GalNAc in yellow).

The interaction between Gal-4 and LPSO55 was analysed by observing the CSP in the HSQC spectra of the ^{15}N labelled lectin. The addition of LPS to the NMR tube containing Gal-4FL (100 μM) induced precipitation of the lectin, resulting in a marked decrease of the cross peak intensities (Fig 79-A). Intriguingly, the intensity decrease (ca. 30%) was rather similar for every residue at both lectin domains. Further experiments were carried out to analyse the impact of the presence of the linker in the full-length construct. In particular, LPS was also added to an NMR tube containing a mixture of Gal-4N and Gal-4C (100 μM each), thus containing the same concentration of binding sites present in the experiment with Gal-4FL described above. Intriguingly, in this case, no precipitates appeared upon addition of the LPS. Additionally, a markedly different intensity decrease was perceived between both domains (Fig 79-B), in contrast to the observations for the full-length protein. Gal-4C lost an overall of 84% of the cross peak intensities, whereas Gal-4N only lost 46%, although above the observed values for Gal-4FL. Fittingly, the difference in intensity loss for the single domains matches with the relative affinity of each domain for the H type-1 antigens,¹⁸⁴ since the C-terminal domain is expected to bind around

two fold better to H type-1 compared to N-terminal domain. Indeed, we have shown (see chapter 3.1) that the binding preferences of the single domains can be extrapolated to FL Gal-4. However, since the single domains lack the ability for cross-linking, the observed differences between Gal-4FL and the single domains can be ascribed to the intrinsic heterodimeric nature of the full lectin.

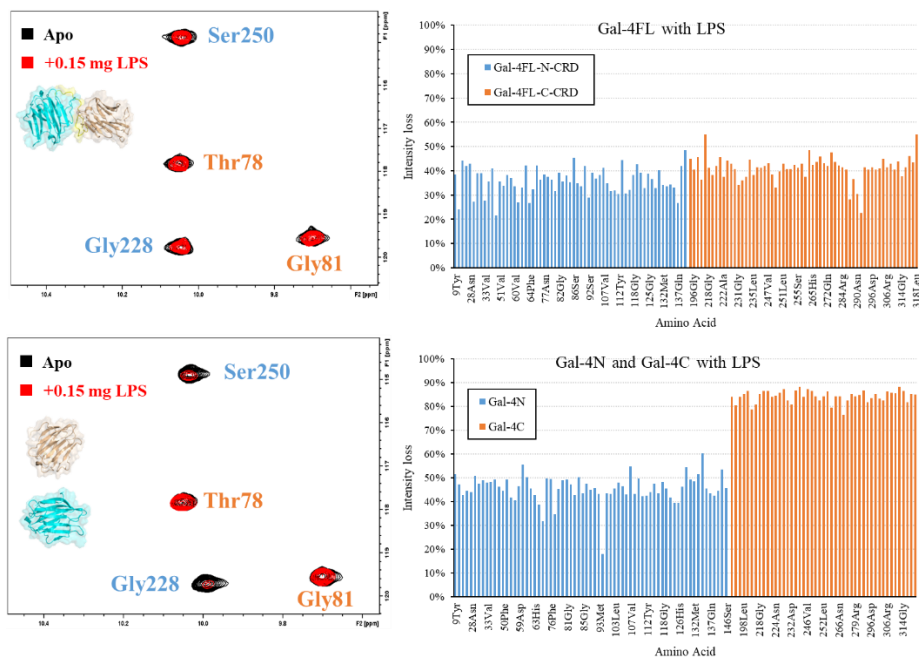


Figure 79. A) Superimposition of a key region of the ^1H - ^{15}N HSQC spectra recorded for apo Gal-4FL (black) and in the presence of 0.15 mg of LPS (red). B) Superimposition of a key region of the ^1H - ^{15}N HSQC spectra recorded for the isolated apo Gal-4N and Gal-4C (black) domains and in the presence of 0.15 mg of LPS (red). The loss of intensity of each amino acid is plotted at the right of each spectra pair. Residues from the N-terminal domain are plotted with blue bars, while those corresponding to the C-terminal are shown in orange.

The intensity loss of the single domains is proportional to the affinity of the particular lectin towards the LPS. Thus, the higher intensity loss of Gal-4C implies that this domain is better binder than Gal-4N for this LPS. However, for the full-length domain, the intensity loss is not affected in the same way as in the single domains, since different processes could be affecting to the observed signal intensity. It should be highlighted that the intensity loss

observed for the mixture of the single domains is larger than that for Gal-4FL. However, it does not directly mean that the single domains bind better the LPS than Gal-4FL.

Next, we employed Gal-4FL-H236R and Gal-4FL-H63R, for which the binding ability of one of the domains is abrogated. Strikingly, the obtained results were different to those described above for the single isolated domains. Whereas Gal-4C was clearly a better LPS binder than Gal-4N, the intensity decrease observed for Gal-4FL-H63R was similar (27% vs 29%) to that of Gal-4FL-H236R (Fig 80). Although the available amount of sugar-binding sites is lesser for the same concentration of LPS, the intensity loss is significantly smaller than that measured for the full-length lectin. Interestingly, in the heterodimers with single mutations, despite not being able to bind sugars, the cross peaks belonging to the mutated CRD displayed a similar decrease in intensity to those at the “active” CRD. Moreover, the addition of LPS to the NMR tubes containing the full-length mutants also induced the presence of precipitates, a phenomenon that was not present for the experiments carried out for the single domains.

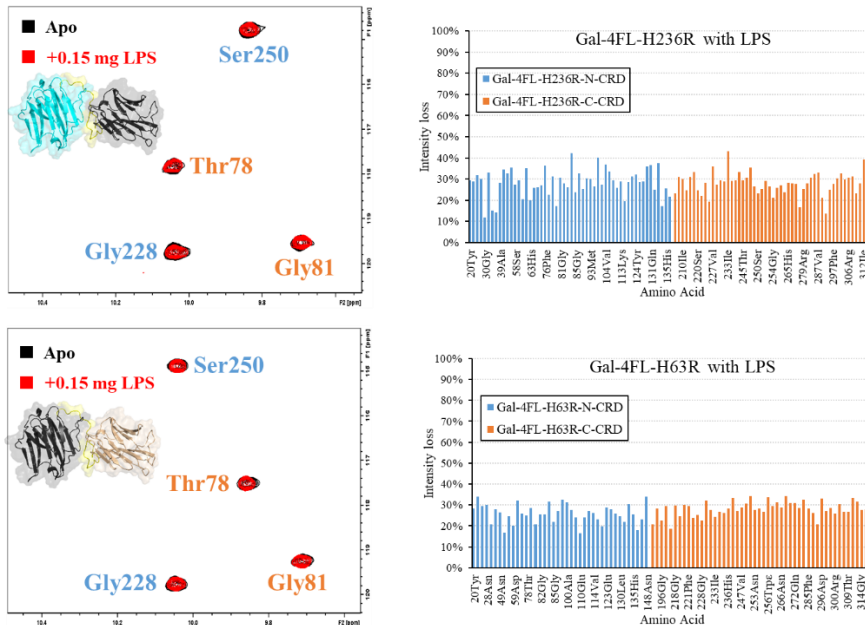


Figure 80. A) Superimposition of a key region of the ^1H - ^{15}N HSQC spectra recorded for apo Gal-4FL-H236R (black) and in the presence of 0.15 mg of LPS (red). B) Superimposition of a key region of the ^1H - ^{15}N HSQC spectra recorded for apo Gal-4FL-H63R (black) and in the presence of 0.15 mg of LPS (red). The loss of intensity of each amino acid is plotted at the right of each spectra pair. Residues from the N-terminal domain are plotted with blue bars, while those corresponding to the C-terminal are shown in orange.

It has been postulated that the linker region of galectins regulate the multimerization and lattice formation of galectins²²⁵. Thus, it is tempting to speculate that the full-length Gal-4 lectins, even with only one active domain, could cross-link ligands through dimerization (Fig 81). This phenomenon could then explain the precipitation and loss of intensity of the HSQC cross peaks of the non-active CRD, which only takes place in the full-length variants.

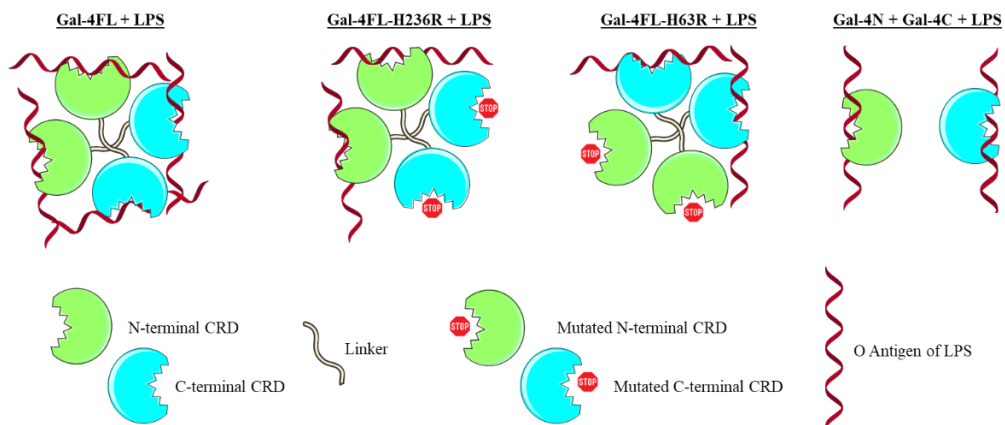


Figure 81. On the top: General scheme of the feasible interactions between LPS and Galectin-4 variants. On the bottom: Representation of each of the cartoons of the picture on top.

3.2.9 The interaction of CD14 with the LPS

To monitor this binding event, CD14 was incubated overnight at 37°C with LPS (1:2 CD14:LPS ratio) in the presence of 0.02 equivalents of LBP to foster the interaction^{190,197}. LBP is a protein capable of binding LPS micelles and to then transfer it to CD14, speeding up the interaction between LPS and CD14. The obtained mixture was then purified through gel-filtration chromatography to give three peaks (Fig 82-A). They were analysed through MALDI MS (SI) and SDS-PAGE electrophoresis (Fig 82-B). The volume of elution of the peaks was used to infer the apparent molecular weight of the eluted molecules using the proper standard (SI). Peak number 1 (diamond, 8.2 mL) eluted with an apparent molecular weight of 835 kDa, while no band appeared in the SDS-PAGE gel. Peak number 2 (triangle, 12.2 mL) had an apparent molecular weight of 158 kDa, providing a band in the SDS-PAGE with the same molecular weight as free CD14. Peak number 3 (square, 14.6 mL) eluted with an apparent molecular weight of 52 kDa, and produced a band corresponding to the molecular weight of CD14 in the SDS-PAGE gel.

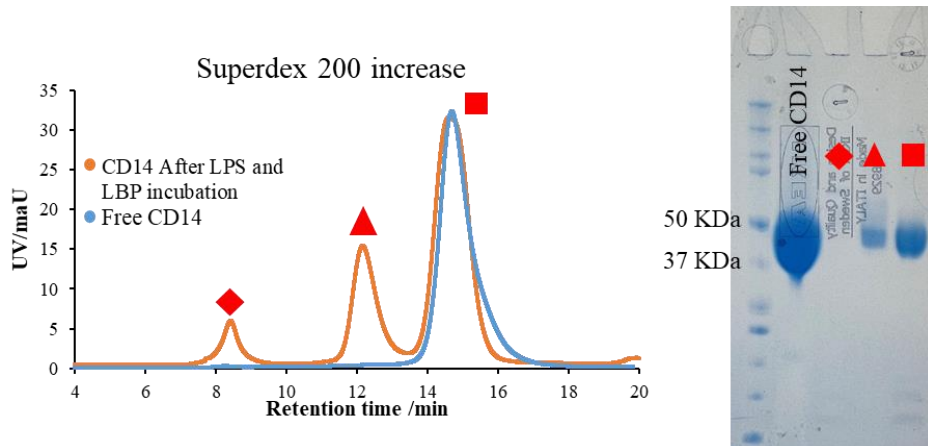


Figure 82. A) In orange, the profile of the CD14-LPS mixture left overnight at 37°C, characterized in a Superdex 200 Increase 10/300 column. In blue, the profile of free CD14. B) SDS-PAGE Gel of free CD14 (line 1) and of every peak eluted from the column. Peaks are labelled with symbols (diamond, triangle, square).

The MALDI analysis confirmed the presence of LPS in the three peaks, together with CD14 in peaks 2 and 3. Peak 3 eluted slightly faster than free CD14, suggesting the existence of a small increase in the size. Interestingly, the MALDI analysis confirmed the presence of both CD14 and LPS in this peak. Therefore, it is tempting to assume that a bimolecular complex has been generated. Although peak number 2 also contained both LPS and CD14, its apparent higher molecular weight can be accounted for the formation of higher order complexes (dimer or trimer) between CD14 and LPS. Further experiments will be required to confirm this hypothesis. Finally, peak number 1 only contained LPS, and its apparent weight is expected for the micelles produced by LPS.

3.2.10 Conclusions

The combination of NMR and mass spectrometry methods has allowed deciphering the glycoprofile of CD14. Whereas there are multiple glycoforms in the exposed N-glycosylation site (N151), only Man₉ and Man₈ glycans are present in the less exposed site (N282). The generated MD models show that the N-glycan at N282 is buried into the protein and interacts with an acidic patch in the concave side of the protein. The lack of spatial accessibility for glycosidases to trim this glycan precludes its maturation and eliminates the possibility of showing different glycans in this site.

Additionally, the effect of each CD14 N-glycan in the stability and ability of interacting with Gal-4 has also been analysed. The N-glycan at N282 is vital for the stability of the protein, as deduced by mutation experiments. Indeed, the mutation at this site reduced over 80% the yield of the soluble protein, and also avoided the proper folding of the protein. Interestingly, the positively charged patch that interacts with the glycan at N282 has previously been described to be relevant in the interaction with the lipopolysaccharide binding protein, the protein that catalyses the binding of LPS to CD14¹⁹⁷. Although there is no X-ray structure of the LBP-CD14 complex and only negative-stain TEM images are available, the N-glycans at N282 are likely located at the interface between CD14 and LBP, and could play a key role in this interaction. Therefore, this N-glycan is relevant for the LPS binding event.

The N-glycan at N151 is also relevant for the folding of the protein, although its mutation does not highly affect the secretion levels. Nevertheless, the mutated protein displays a larger aggregation tendency than the wild-type protein. Moreover, the glycans at this site are key for its interaction with Gal-4, which binds the LacNAc and 3'sialyl-LacNAc epitopes present at this site. Contrary to previous reports, the presence of core fucose could not be assessed either at N151 or at N282^{207,208}.

CD14 effectively interacts with LPS. Indeed, CD14 shows two different populations, both bound to LPS, although with different sizes, as deduced by gel filtration chromatography. In any case, further experiments should be performed to understand the nature of these complexes.

The investigations carried out in this chapter have shown that the O-antigen of the LPS from *E. Coli* O55, a Gram-negative bacteria, is recognized by the two domains of Galectin-4. Although the precise epitope of the LPS could not be deciphered, the $\text{Col}\alpha\text{-1,2-Gal}\beta\text{-1,3-GlcNAc}$ epitope is likely to be the recognized motif, based on our previous finding¹⁸⁴. The LPS was better bound by Gal-4C than by Gal-4N, as expected¹⁸⁴. This evidence was further confirmed by performing additional LPS-binding experiments with the mixture of Gal-4N and Gal-4C. Intriguingly, a different behaviour between the mutants of the full-length lectin and the single domains was observed. In fact, the mutants precipitated in the presence of LPS (as the WT heterodimer), while the single domains did not. The fact that all the full-length variants precipitate in the presence of the LPS strongly suggests that the full-length mutants crosslink different LPS moieties. It has been reported that the linkers of tandem-repeat galectins are responsible for galectin dimerization.²²⁵ Therefore, it is feasible that the full-length mutants with only one active site also generate dimers and crosslink their glycan ligands, even with only one active site (Fig 83).

3.3 Synthesis of well-defined β -glucans using Automated Glycan Assembly

This chapter gathers the results obtained during a short secondment (2 months) carried out in the Max Planck Institute of Colloids and Interfaces in Potsdam in collaboration with Dr. Martina Delbianco and Prof. Dr. Peter Seeberger. During this stay, two building blocks were synthesized to build larger oligosaccharides using the Automated Glycan Assembly (AGA).

3.3.1 Introduction

In the previous chapter, we highlighted the relevance of receptors like CD14 or lectins like galectin-4 in the immune response to bacterial pathogens. However, not only bacteria cause acute health issues in humans. Among the many infectious diseases caused by pathogens, fungal infection, also known as mycosis, is also a major cause of death (1.7 million deaths per year) and severe illness (over 150 million per year)²²⁶.

As previously mentioned, the response from the human immune system to pathogen infection consists of two components, the innate and the adaptive immune system. The first response to pathogens is carried out by the innate immune system, for which pattern-recognition receptors (PRRs) are key, since they are responsible of detecting those pathogens for the consequent response²²⁷. These receptors are a large family of proteins that recognize highly conserved pathogen-associated molecular patterns (PAMPs). The family of fungi, which include mushrooms, molds, and yeasts, holds a cell wall that contains many polysaccharides of different compositions (glucans, mannans, and chitin, among others)^{228,229}. Since this wall is the outer part of these pathogens, the polysaccharides that cover this wall directly interact with

multiple entities of the host cell, including PRRs, which recognize these glycans as PAMPs and generate an immune response²³⁰.

β -Glucans are one of the most common polysaccharides found in the cell wall of fungi^{228–230}. They have a backbone of β 1,3 linked glucose (Glc) units, which can also contain β 1,6 branched Glc units⁹. There are many human receptors that can interact with β -glucans. However, the main receptor towards glucans is the dendritic-cell-associated C-type lectin-1 or Dectin-1²³¹. Regardless of its name, the lectin is also expressed in other cell types as macrophages, monocytes, neutrophils, and T-cells, among others. Regarding its structure, Dectin-1 contains an extracellular C-type lectin domain, a transmembrane domain, and an ITAM-like motif (Immunoreceptor Tyrosine based Activation Motif) in its cytosolic tail (Fig 83-A)²³².

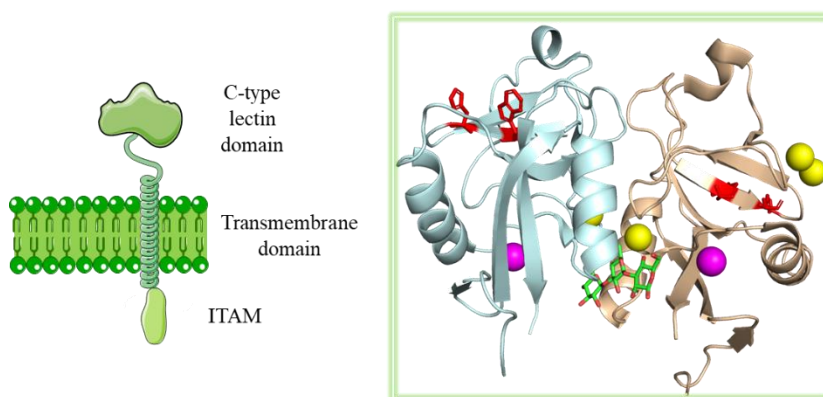


Figure 83. A) Schematic representation of Dectin-1 and its different domains. B) X-ray structure of mouse Dectin-1 in complex with a glucan trisaccharide (PDB 2CL8)²³³. The calcium atoms are depicted as purple spheres. The chlorine atoms are depicted as yellow spheres. The key Trp221 and His233 are depicted with sticks in red.

It is worth noting that the single C-type lectin domain of human Dectin-1 has not been purified yet, and it has been reported that it displays very low solubility²³⁴. Therefore, the studies carried out by the different research groups with this lectin have been performed with mouse Dectin-1. Despite being a C-type lectin and recognizing β -glucans through the C-type lectin domain, the

calcium cation is not located in the canonical binding site (Fig 83-B)²³³. The X-ray structure of mouse Dectin-1 in complex with a β -glucan trisaccharide, inserted through soaking, shows that the glycan is not located in the canonical binding site for C-type lectins, but in the interface between two proteins (Fig 83-B). Intriguingly, Trp221 or His223 (in red in Fig 83-B), located in the canonical binding site of C-type lectins, are vital for the recognition of β -glucans, suggesting that the interaction of Dectin-1 with glucans occurs through the canonical binding site and that the location of the ligand is an artefact of the soaking event²³⁵.

Although there is no precise structural knowledge on the binding to these glucans, it has been reported that Dectin-1 has preference for longer and branched β -glucans²³⁶⁻²³⁸. Yamaguchi and coworkers analysed the binding of linear hexasaccharide, hexadecasaccharide, and polysaccharide of β -glucans to Dectin-1, showing that the increase the length of the glycan improves the binding to Dectin-1²³⁶. Adams *et al.* showed that branching can have a huge impact in the binding to Dectin-1; for instance, a linear heptasaccharide displayed an IC_{50} of ca. 1 mM, whereas the branched analogue displayed an IC_{50} of 0.13 mM²³⁷. This effect was also present for the nonasaccharide (IC_{50} of 2.6 mM *versus* 29 μ M for the branched one); however, this was not the case for the octasaccharide (IC_{50} of 1.1 mM *versus* 1.3 mM). This contradictory data can be initially attributed to the position of the branching. In this case, the impact of the position of the branching point has not been unravelled, since branching can occur in multiple points, and the systematic effect of the branching in every position has not been analysed yet. For longer glycan chains, i.e. for polysaccharides, very low IC_{50} values at nM-pM levels have been reported^{239,240}.

As shown by the previous examples, the structural knowledge regarding the impact of the length and branching of β -glucans in the binding to Dectin-1 is far from being completely understood. Many research groups have studied the

impact of the length, linkage type, and branching of multiple different kinds of glycans, since it affects to the conformation and function of the molecule^{241–244}. Therefore, the synthesis of well-defined oligosaccharides structures has become a major topic of interest^{241,245–248}. For years, the use of solid phase synthesizers has enabled generating long and well-defined linear peptide or nucleic acid structures^{249,250}. However, the synthesis of carbohydrates is more complicated due to their intrinsic complexity, since they have multiple connection points and branches. The use of the Glyconeer[®], the first automated glycan synthesizer, has revolutionized the field of glycan synthesis, since it allows synthesizing long and well-defined glycans rapidly²⁵¹.

For the automated glycan assembly (AGA)²⁵², firstly a solid resin with a linker is placed in the reaction vessel. This resin has to be swollen before starting the synthesis. Then, the linker that has a nucleophile (the acceptor) will be incubated with the desired protected building block (BB), which bears a leaving group at the anomeric position (donor). The addition of a Lewis acid to the reaction vessel will activate the leaving group, and the nucleophile will attack the anomeric position of the BB. Since the anomeric position is activated, both α and β linkages can occur; however, different approaches can be applied to obtain the desired selectivity (the insertion of directing groups in the BB, among others). Once the first monosaccharide unit is attached, the reaction vessel will be washed to remove the exceeding reagents. Since the unreacted nucleophile may remain, after washing the vessel a capping step can be applied, so that the unreacted acceptor stops reacting. Afterwards, an orthogonal protecting group of the attached BB will be selectively removed and, in this way, the BB, which firstly worked as a donor, will also work as an acceptor. Through repeating the previous steps, the molecule can be elongated in the desired direction (linear or branching). Finally, the protected oligosaccharide should be removed from the resin and further purified to obtain the desired deprotected oligosaccharide.

In this context, we decided to design β 1,3 glucans with different lengths and branching points to interrogate their binding to Dectin-1. In order to achieve this, building blocks **1 (BB1)** and **2 (BB2)** were synthesized, with the aim of employing them in the Glyconeer[®] (Fig 84).

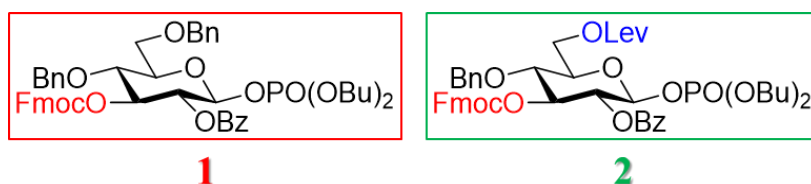
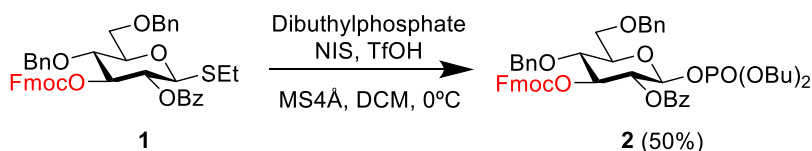


Figure 84. Structure of the building blocks synthesized for the Automated Glycan Assembly. The orthogonal protecting groups are depicted with red (Fmoc) and blue (Lev) colors

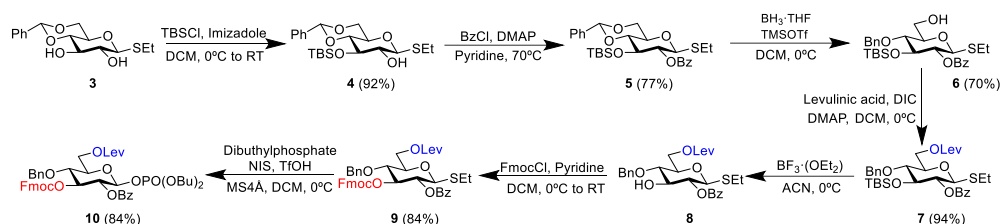
3.3.2 Synthesis of the building blocks

Both building blocks were synthesized following previously reported protocols²⁵³. The **BB1** was synthesized from its thioglycoside analogue (**1**), which was available in the lab from previous synthesis (Scheme 2). Conversion of the thioglycoside into a glycoside with a phosphate group as a leaving group was performed using dibutylphosphate, NIS, and triflic acid (TfOH) as a catalyst in DCM using molecular sieves. The reaction did not proceed with a complete conversion and the yield was moderate (50%).



Scheme 2. Synthesis of **BB1**

BB2 was synthesized from available commercial Ethyl 4,6-*O*-benzylidene-1-thio- β -D-glucopyranoside (**3**) following Scheme 3.



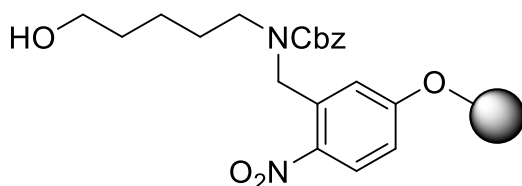
Scheme 3. Synthetic route for the preparation of **BB2**

The first step was the selective TBS protection of diol **3** at position 3, which was totally regioselective. Benzoylation of compound **4** was tested using three different conditions, BzCl, with pyridine in DCM with and without DMAP as cocatalyst at r.t. overnight, which none of them fully converted (2:1 starting material:product ratio). Only BzCl, with DMAP as cocatalyst at 70°C in pyridine, yielded complete conversion. The selective benzylidene reductive ring opening of compound **5** did not achieve complete conversion, always obtaining part of the starting product. The reaction was performed at 0°C. It has to be mentioned that the temperature cannot be increased since the use of higher temperatures resulted in complete hydrolysis of the benzylidene ring. Lev protection of hydroxyl group of **6** was completed in one hour at 0°C, with high yield. After that, TBS deprotection of **7** was performed, workup of this reaction afforded **8**, pure enough to be employed in the subsequent Fmoc protection without further purification, thus obtaining the protected compound **9**. Conversion of the thioglycoside into a glycoside with a phosphate group as a leaving group was performed as for **BB1**, obtaining higher yield than for the **BB1** (84% for **BB2** vs 50% for **BB1**).

3.3.3 Automated Glycan Assembly (AGA)

The automated syntheses were performed on a home-built synthesizer developed at the Max Planck Institute of Colloids and Interfaces²⁵⁴. A

photolabile linker (Scheme 4), previously synthesized at the Max Planck Institute of Colloids and Interfaces, attached to a Merrifield resin was employed for the AGA²⁵⁵.



Scheme 4. Chemical structure of the photolabile linker attached to a Merrifield resin (Black dot)

Employing **BB1** and **BB2**, the synthesis of three different oligosaccharides was attempted using AGA: a linear tetrasaccharide (**11**), a linear octasaccharide (**12**), and a branched pentasaccharide (**13**) (Fig 85). For the linear tetra- and octa-saccharides, **BB1** was employed. For the branched pentasaccharide, the **BB2** was employed for the branching point, and **BB1** for the other units. The modules employed for the synthesis of the oligosaccharides are described in the Materials and Methods section. Each module includes several actions. For each action, the number of cycles, the reagents and solvents employed, the temperature of the reaction, and the incubation times are also annotated.

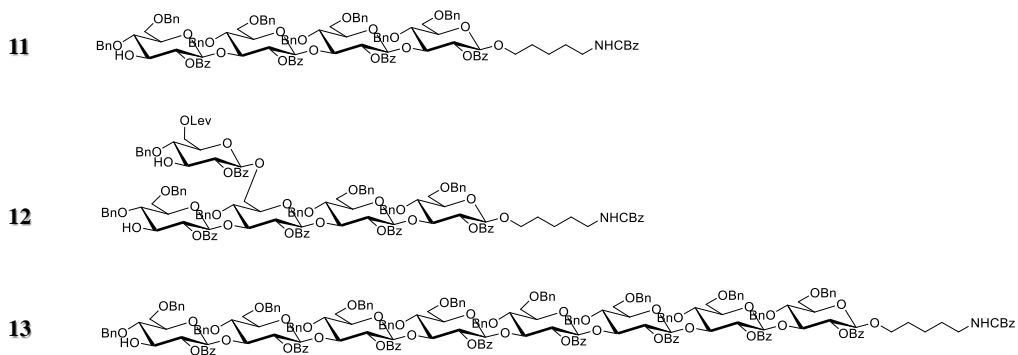


Figure 85. Structure of the oligosaccharides that were attempted to synthesize in the AGA

3.3.4 Deprotection and purification of the oligosaccharides

The first step of the purification process was the removal of the oligosaccharides from the resin. The cleavage of the resin was achieved using a Hg lamp to irradiate UV light. After removal from the resin, the efficiency of the AGA was measured through MALDI-MS. The linear tetrasaccharide provided a single peak corresponding to the fully protected tetrasaccharide, (ESI-HRMS) m/z 2044.777 $[M+Na^+]$ ($C_{121}H_{123}NNaO_{27}$ requires 2044.82) (See SI). The branched pentasaccharides provided two peaks at the MALDI, one corresponding to the capped dimer and another one corresponding to the non-capped trimer (SI). These results show that the **BB2** is not a good donor, since there is a peak corresponding to the capped dimer, meaning that there was unreacted dimer. Additionally, the lack of peaks corresponding to the tetramer or pentamer showed that **BB2** is not able to work as an acceptor, at least under the experimental conditions employed herein. For the linear octasaccharide, multiple deletion sequences were observed (SI), ranging from the dimer to the octamer; however, the peak arising from the octamer was very small, meaning that the yield was very low. Therefore, only the purification of the tetrasaccharide was carried out.

After the UV cleavage, the methanolysis of the tetrasaccharide was performed. Per each benzoyl group, 1.25 equivalents of NaOMe was employed. The

oligosaccharide was dissolved in THF, and NaOMe in MeOH was added, the mixture was stirred overnight. Next step of purification was the hydrogenation to remove benzyl groups. Per each benzyl group, 3 equivalents of Pd(OH)₂ in activated charcoal was employed. The oligosaccharide was dissolved in a mixture of EA:tert-butanol:H₂O (2:1:1) and stirred overnight in a H₂ atmosphere. The deprotected oligosaccharide was further purified through HPLC as described in the materials and methods section. Automated synthesis, global deprotection, and purification afforded the tetrasaccharide as a white solid (1.9 mg, 40% overall yield), (ESI-HRMS) m/z 752.3176 [M⁺H]⁺ (C₂₉H₅₄NO₂₁ requires 752.3183). The schematic procedure for the synthesis and purification of the tetrasaccharide can be seen in figure 86 and Table 10.

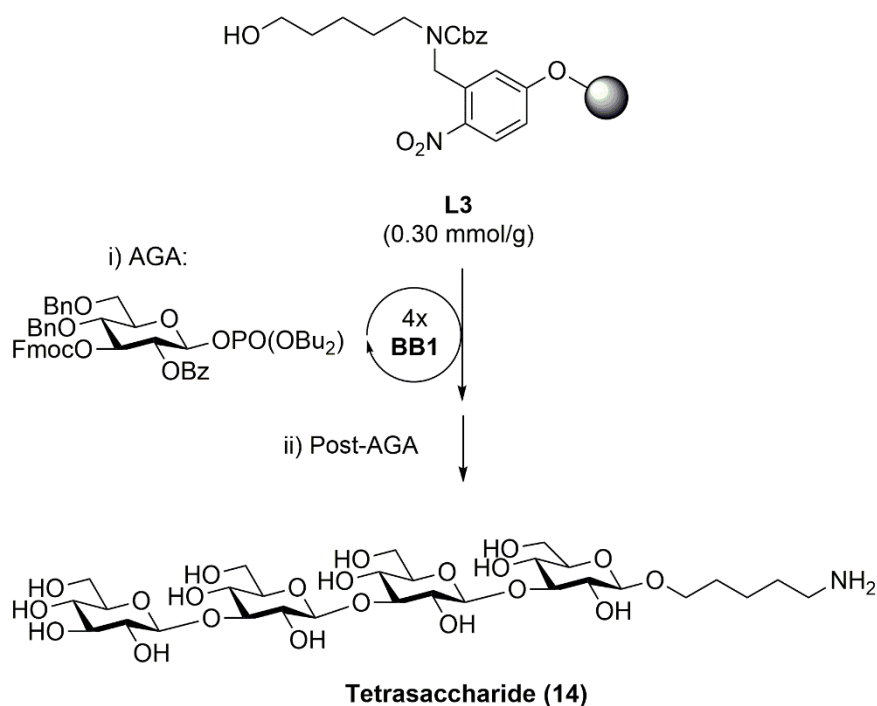


Figure 86. General procedure for the synthesis of tetrasaccharide 14

Table 10. Automation sequence employed for the synthesis of the tetrasaccharide and the process employed afterwards

Step	Modules	Notes
------	---------	-------

AGA	L3 (BB1) _{x4}	A (B, C, D, E) _{x4}	L3 swelling C: (BB1 , -20 °C for 5 min, 0 °C for 20 min)
Post-AGA		G, H, I, J	H: (16 h) I: (8 h) J: (Method A2, t _R = 13.4 min)

3.3.5 Conclusions

Two building blocks have been synthesized to be employed in the Automated Glycan Assembly to synthesize β -glucans. The use of Automated Glycan Assembly has enabled the fast synthesis of a linear tetrasaccharide with good yield (40%, 1.9 mg). However, although employing the same conditions, the linear octasaccharide was not obtained. In the synthesis of the branched pentasaccharides, **BB2** was not completely attached to the dimer, thus, increasing the number of equivalents or changing the conditions of the reaction should be tested. Additionally, **BB1** did not work as a donor when **BB2** was the acceptor. For attachment of **BB1**, both Fmoc and Lev were removed from the donor in the same step, so that both glycosylations took place simultaneously. However, the resulting molecule is not a good donor, since no **BB1** could be attached, neither at position 3 nor at 6. A possible solution to this issue could be the individual glycosylation of each position. Nevertheless, more experiments should be performed in order to obtain the desired oligosaccharides.

4 Materials and methods.

4.1. Protein expression and purification

The lectins employed in this thesis were expressed in *Escherichia coli* (*E. Coli*) bacterial cells, whereas the glycoproteins were expressed in human embryonic kidney cells (HEK293, specifically HEK293F).

4.1.1. Cloning of the vectors.

The amino acid sequence coding for the N-terminal CRD of Galectin-4 (aa 1-150) including a C-terminus His-tag and a thrombin-cleavage site was inserted into a pET29b(+) expression vector. The sequence coding the C-terminal CRD (aa 179-323) also included a C-terminus His-tag and a thrombin-cleavage site and was also inserted into a pET29b(+) expression vector.

The DNA fragment coding for the full-length *h*Galectin-4 (amino acids 1 – 323, including an additional C-terminal SUMO fusion protein and a His-Tag) was inserted into the pET28a-SUMO expression vector. Point mutations (H63R and H236R) were also inserted into two different pET28a-SUMO expression vectors. The vector for the expression of the sumo protease was a gift from Hideo Iwai (Addgene plasmid # 31122)

The amino acid sequence coding for the human CD14 (aa 1-336) was inserted into a PCDNA 3.4 vector. Two different point mutations (N151Q and N282Q) were also inserted into two different PCDNA 3.4 vectors. In order to improve the solubility of CD14 and its mutants, the proteins were fused to the Fc portion of the IgG1. The proteins also contained a His-tag at the C-terminus to facilitate the purification.

4.1.2. Protein expression

4.1.2.1. Expression of unlabelled proteins using E. Coli.

The vectors for the single CRD domains were amplified in *E. coli* DH5 α cells and afterwards transformed into *E. coli* BL21 cells. A single colony was inoculated into 200 mL Luria Broth (LB) medium containing 50 $\mu\text{g}\cdot\text{mL}^{-1}$ kanamycin and cultured at 37 °C overnight. A precise quantity of the cultured was then added to 2L of fresh LB containing kanamycin to achieve an initial optical density at 600 nm (OD600) of 0.1. Cells were grown at 37 °C until OD600 reached 0.6-0.8, subsequently, protein expression was induced by addition of 1 mM of isopropyl-1-thio- β -D-galactopyranoside (IPTG) and the culture was allowed to grow overnight at 20 °C^{71,256}.

For the full-length Gal-4 lectins, the transformation was into *E. coli Rosseta* (DE3) cells. The transformed cells were inoculated into 200 mL of LB medium containing 34 $\mu\text{g}\cdot\text{mL}^{-1}$ chloramphenicol and 30 $\mu\text{g}\cdot\text{mL}^{-1}$ kanamycin and cultured overnight at 37 °C. A precise quantity was transferred to 2L of fresh LB medium to obtain an initial OD600 of 0.1. Cells were grown at 37 °C until obtaining a OD600 of 0.6, consequently, protein expression was induced by adding 1 mM of IPTG and the culture was allowed to grow 24h at 25°C⁷⁰.

4.1.2.2. Expression of ¹⁵N or ¹³C and ¹⁵N labelled proteins using E. Coli.

The transformed cells were inoculated into 5 mL of LB containing the antibiotic and were grown 6h at 37°C. The small preculture was centrifuged and resuspended in 1 mL of M9 minimal medium, transferred to 200 mL of M9 medium and then incubated overnight at 37 °C. A precise quantity of the culture was then added to obtain a starting OD600 of 0.1 in 2L of fresh M9 medium containing 1 g/L ¹⁵NH₄Cl for the ¹⁵N-labeled samples, or 1 g/L ¹⁵NH₄Cl and 20% w/v U-¹³C-glucose for the ¹³C,¹⁵N-labeled samples²⁵⁷. Unlabelled amino

acids (L-Arginine, L-Phenylalanine, L-Valine or L-Lysine) were added in 1 g/L to M9 medium containing 1 g/L $^{15}\text{NH}_4\text{Cl}$ to selectively not label concrete amino acids of the protein¹⁴⁹. Upon reaching an OD600 of 0.6 protein expression was induced by addition of 1 mM of IPTG. *E. coli* BL21 cells were harvest after incubation overnight at 20°C. *E. coli Rosetta* (DE3) cells were harvest after 24h at 25°C.

4.1.2.3 Purification of proteins expressed using E Coli

The obtained culture pellets were resuspended in lysis buffer (Table 11) and sonicated at 4 °C. The crude extract was clarified by centrifugation at 35000 rpm for 1 h at 4 °C. The soluble fraction was purified by Ni-NTA affinity chromatography and further purified by size exclusion chromatography in a HiLoad 26/600 Superdex 75 column⁶⁹⁻⁷¹. Protein purity was checked by 4–12 % SDS-PAGE and by LC-MS.

Table 11. Lysis buffers employed for the purification of the proteins expressed in *E. Coli*.

Protein	Lysis Buffer
Gal-4N	50 mM sodium phosphate pH 8, 300 mM NaCl, 1 mM PMSF
Gal-4C	50 mM sodium phosphate pH 8, 600 mM NaCl, 1 mM PMSF
Gal-4FL, Gal-4FL-H63R, Gal-4FL-H236R	50 mM sodium phosphate pH 8, 600 mM NaCl, 1 mM PMSF, 1 mM DTT
Sumo Protease	20 mM Tris pH 8, 300 mM NaCl

For Gal-4N the C-terminal His-tag was removed by overnight incubation at 4 °C with thrombin (10 unit of thrombin per mg of protein). Afterwards, the

protein was loaded onto a 5 mL Ni-NTA column and the desired fragment collected in the wash and loaded in a HiLoad 26/600 Superdex 75 column.

4.1.2.4 Expression of glycoproteins using human embryonic kidney cells (HEK293F)

Every cell culture was grown in a Minitron Infors HT orbital shaker incubator using sterile culture flasks with cap with HEK293 FreeStyle media at 310 K, 70% humidity and 8% CO₂. Maintenance HEK293F cells were split every three days to obtain a cell concentration of $5 \cdot 10^5$ cells/mL (cell viability over 97%) into 200 mL flasks. Cells were split to a cell concentration of 10^6 cells/mL in 200 mL prior to protein transfection. The transient transfection of cells was achieved by adding 250 µg/L of culture of the desired DNA plasmid previously incubated 10 minutes with 250 µg/L of FectoPRO²¹⁵. To obtain the ¹³C labelled carbohydrates on the protein, 3 g/L of ¹³C labelled glucose was added to the culture after adding the plasmid with the FectoPRO²¹⁵.

4.1.2.5 Purification of proteins expressed using human embryonic kidney cells (HEK293F)

Six days after transfection cells were harvest at 10.000 rpm at room temperature and filtered through a 45-micron filter. The supernatant was loaded into a HiTrap Protein A HP column at 2 mL/min and the desired protein was eluted using 0.1M glycine pH 3²⁰². The eluted protein was dialysed against 20 mM tris pH 8, 300 mM NaCl and incubated with TEV protease for two weeks at room temperature. The mixture was loaded into a Ni-NTA affinity column and the protein without the Fc was eluted in the wash. The obtained product was further purified using a HiLoad 26/600 Superdex 200 column.

4.2 Glycosidase treatment of CD14

CD14 was treated with endo H in order to decipher its glycoprofile. CD14-Fc was incubated overnight at room temperature with 500 units of glycosidase per mg of CD14-Fc²⁵⁸.

4.3 Circular Dichroism (CD) measurements

CD measurements of Gal-4FL, Gal-4FL-H236R, Gal-4FL-H63R, CD14 were acquired using a CD spectrometer²⁵⁹. The thermal denaturation of the proteins was also measured using the CD spectrometer. The employed protein concentration was 2-4 μ M in phosphate-buffered saline (50 mM sodium phosphate, 150 mM NaCl, pH 7.4, with 1 mM DTT for the Gal-4 proteins).

4.4 Ligands

Lactose was purchased from Sigma. Glycans H type-6, B type-6 trisaccharide, A type -1, -2 and -6 tetrasaccharides, A type-4 pentasaccharide, B type -1, -2 and -6 tetrasaccharides and B type-4 pentasaccharide were purchased from Elicityl (references GLY031-3, GLY074-2, GLY035-1, GLY035-2, GLY035-3, GLY128, GLY038-1, GLY038-2, GLY038-3, GLY129). The LPS from *E. Coli* O55 was purchased from Sigma.

4.5 Molecular Dynamic (MD) Simulations

4.5.1 MD simulations of Gal-4N and Gal-4C in complex blood group antigens

The initial geometries of the complexes were built by pair-fitting of the central β Gal moiety of the corresponding ligand to the same moiety of the bound lactose in the X-ray structure (PDB 5DUV for Gal-4N and PDB 4YM3 for Gal-4C)^{69,71}. The glycan structures were built in the Glycam webpage²⁶⁰. MD simulations were run using Amber18 with the ff14SB force field for the protein and GLYCAM06j-1 for the carbohydrates. The complexes were prepared in explicit water (TIP3PBOX) and minimized in two steps before starting the simulations. MD simulations of 1 μ s were analysed using cpptraj.

4.5.2 MD simulations of CD14

Models of the glycosylated CD14 were produced using the glycoprotein builder module in the Glycam web and the deposited X-ray structure of the human CD14 (PDB 4GLP)²⁰². The models always contained a Man₉ oligosaccharide linked at asparagine 282 and different oligosaccharides attached to asparagine 151. MD simulations were run using Amber20 with the ff14SB force field for the protein and GLYCAM06j-1 for the carbohydrates²⁶¹. The complexes were prepared in explicit water (TIP3PBOX) and minimized in two steps before starting the simulations. MD simulations of 500 ns were analysed using cpptraj.

4.6 Nuclear Magnetic Resonance (NMR)

4.6.1 General information.

The total volume for the NMR samples was 500 μ L. The proteins were dissolved in phosphate-buffered saline (50 mM sodium phosphate, 150 mM

NaCl, pH 7.4, with 1 mM DTT for the full-length proteins), either in D₂O or 90:10 H₂O:D₂O depending on the NMR experiment. The pH was adjusted with the required amount of NaOH and HCl or NaOD and DCl.

4.6.2 ¹H-¹⁵N-HSQC based titrations

¹H-¹⁵N-HSQC spectra¹³⁴ were recorded using an 800 MHz Bruker spectrometer equipped with a cryoprobe or 600 MHz Bruker spectrometer with a triple-channel probe. The samples were prepared using 25-200 μM of the ¹⁵N labelled lectin in the corresponding buffer in 90:10 H₂O:D₂O. The experiments were acquired at 298 K. Six to nine points were recorded for each glycan and chemical shift perturbation (CSP) and dissociation constants (K_D) were then estimated. For the titration with CD14, the employed concentrations were 25 μM for the lectins and 200 μM for the glycoprotein. From the integrations of the peaks, CSP and K_D values were calculated using CcpNmr Analysis 2.4.2.

4.6.3 Saturation Transfer Difference (STD) NMR

All the STD NMR experiments were acquired using a 800 MHz Bruker spectrometer with a cryoprobe. The samples were prepared in the corresponding deuterated buffer. STD NMR spectra were acquired with 1024 scans, 2 s of saturation time using a train of 50 ms Gaussian-shaped pulses, and 3 s of relaxation delay²⁶². The spin-lock filter applied to remove the residual signals of the lectin was set at 40 ms. The protein:ligand ratio and the temperature was optimised for each lectin.

For the N-terminal CRD of Gal-4 the temperature was 288 K. 50 μM of the lectin with 50 equivalents of the ligand were employed except for lactose, for which a 1:70 ratio was used. The on-resonance frequency was set at the

aromatic region at 6.58 ppm, while the off-resonance frequency was set at 100 ppm.

For the C-terminal CRD of Gal-4, the temperature was 298 K. 50 μ M of the lectin with 50 equivalents of the ligand were employed. The on-resonance frequencies were set at 7.19 ppm (aromatic region) and at 0.52 ppm (aliphatic region), while the off-resonance frequency was set at 100 ppm.

The experiments with Gal-4FL, Gal-4FL-H63R and Gal-4FL-H236R were acquired at 298 K. The on-resonance frequency was set at 7.03 ppm, and the off-resonance, at 100 ppm. The experiments were performed using 25 μ M of Gal-4FL and 50 μ M of Gal-4FL-H63R and Gal-4FL-H236R, all with 50 equivalents of ligand.

4.6.4 trROESY spectra

ROESY NMR spectra for the glycans were acquired, in the 800 MHz Bruker spectrometer, in the presence of 50 μ M Gal-4N/Gal-4C using 1:20/1:10 protein:ligand ratios in the corresponding deuterated phosphate-buffered saline buffer, at 298 K^{177,263}

4.6.5 Protein Backbone Resonance Assignment

The experiments required for achieving the backbone resonance assignment of the N-terminal CRD of *h*Galectin-4 were carried out at 25 °C on a 800 MHz Bruker spectrometer equipped with a cryoprobe. In particular, 3D HNCO, HN(CA)CO, HN(CO)CACB and HNCACB experiments^{140–143,145,146} were acquired for the free Gal-4N containing the His-Tag and for Gal-4N without the His-Tag in the presence of 200 equivalents of lactose. For the C-terminal CRD

of *h*Galectin-4, the aforementioned experiments as well as 3D HN(CO)CA and HNCA experiments were recorded for the free protein containing the His-Tag. Additionally, 3D HNH-NOESY and HNH-TOCSY experiments were also recorded for this domain. The entire peak analysis provided the unambiguous identification of 80% of the expected NH signals for Gal-4N and 86% for Gal-4C. The spectra were processed with Bruker TopSpin 3.5.2 and analysed via CARANMR 1.9.1.4²⁶⁴.

4.6.6 Assignment of the ^1H - ^{13}C -HSQC of the glycans of CD14.

Assignment of the N-glycans of CD14 was performed using a sample of 400 μM of CD14 with ^{13}C labelled glycans. 3D H' ,CH NOESY-HSQC, H' ,CH TOCSY-HSQC, and [^{13}C , ^{13}C]TOCSY-HSQC experiments²¹⁴ were carried out using the 800 MHz Bruker spectrometer equipped with the cryoprobe at 310 K. The spectra were processed and analysed with Bruker TopSpin 3.5.2.

4.6.7 ^1H - ^{13}C -HSQC of the glycans of CD14 in the presence of lectins.

^1H - ^{13}C -HSQC spectra for ^{13}C -labelled CD14 were collected at 310 K. The spectra were collected with 24 scans and 1536 points in the F2 direction and 660 points in the F1 dimension. Different spectra were recorded for the glycoprotein both in the presence and in the absence of full-length Gal-4, full-length Gal-4-H63R and full-length Gal-4-H236R. The intensities of the peaks were measured using TopSpin 3.5.2.

4.7 Isothermal Titration Calorimetry (ITC)

Isothermal Titration Calorimetry experiments were performed using MicroCal PEAQ-ITC calorimeter. Samples containing 100-200 μM the lectins in phosphate-buffered saline buffer (50 mM sodium phosphate pH 7.4, 150 mM NaCl) were titrated with stocks of 3-10 mM in phosphate-buffered saline of the blood group antigens. During the automated experiment, small aliquots (2-3 μL) of the sugar stocks were added to the cell containing the target lectin. Curve fitting to a single binding site model was performed with MicroCal Origin 7 software²⁶⁵.

4.8 Automated Glycan Assembly

4.8.1 General materials and methods

The automated syntheses were performed on a home-built synthesizer developed at the Max Planck Institute of Colloids and Interfaces²⁶⁶. All solvents used were HPLC-grade. The solvents used for the building block, activator, TMSOTf and capping solutions were taken from an anhydrous solvent system (J.C. Meyer) for moisture-sensitive solutions. The building blocks were co-evaporated three times with toluene and dried for 1 h under high vacuum before use. Oven-heated, argon-flushed flasks were used to prepare all moisture-sensitive solutions. Activator, capping, deprotection, acidic wash and building block solutions were freshly prepared and kept under argon during the automation run. All yields of products obtained by AGA were calculated on the basis of resin loading. Resin loading was determined following previously established procedures²⁶⁷.

4.8.2 Building blocks

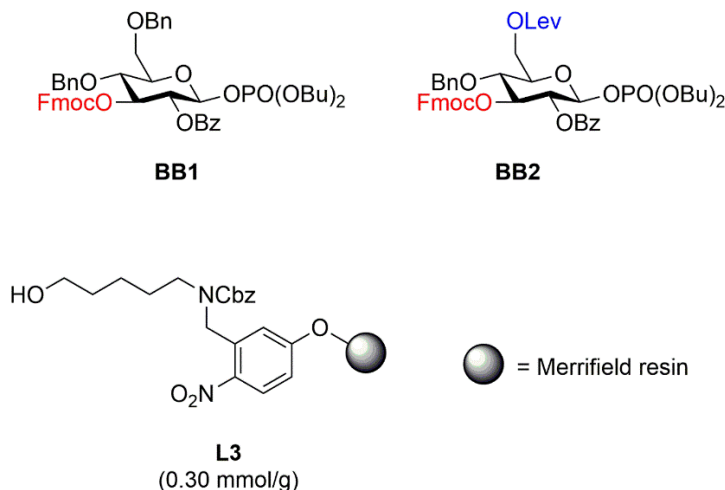


Figure 87. Chemical structure of the building blocks employed in this Thesis

4.8.3 Preparation of stock solutions

- **Building block solution:** Between 0.06 and 0.10 mmol of building block (depending on the BB, see Module C1 and C2) was dissolved in DCM (1 mL).
- **NIS/TfOH activator solution:** 1.35 g (6.0 mmol) of recrystallized NIS was dissolved in 40 mL of a 2:1 v/v mixture of anhydrous DCM and anhydrous dioxane. Then triflic acid (55 μ L, 0.6 mmol) was added. The solution is kept at 0 °C (ice bath) for the duration of the automation run.
- **Fmoc deprotection solution:** A solution of 20% piperidine in DMF (v/v) was prepared.
- **Lev deprotection solution:** Hydrazine acetate (550 mg, 5.97 mmol) was dissolved in pyridine/AcOH/H₂O (40mL, v/v, 32:8:2) and sonicated for 10 min.
- **TMSOTf solution:** TMSOTf (0.45 mL, 2.49 mmol) was added to DCM (40 mL).
- **Capping solution:** A solution of 10% acetic anhydride and 2% methanesulfonic acid in DCM (v/v) was prepared.

4.8.4 Modules for automated synthesis

Module A: Resin Preparation for Synthesis (20 min)

All automated syntheses were performed on 0.0125 mmol scale. Resin **L3** (45 mg) was placed in the reaction vessel and swollen in DCM for 20 min at room temperature prior to synthesis. During this time, all reagent lines needed for the synthesis were washed and primed. After the swelling, the resin was washed with DMF, THF, and DCM (three times each with 2 mL for 25 s).

Module B: Acidic Wash with TMSOTf Solution (20 min)

The resin was swollen in 2 mL DCM and the temperature of the reaction vessel was adjusted to -20 °C. Upon reaching the temperature, TMSOTf solution (1 mL) was added dropwise to the reaction vessel. After bubbling for 3 min, the acidic solution was drained and the resin was washed with 2 mL DCM for 25 s.

Table 12. Parameters of Module B

Action	Cycles	Solution	Amount	T (°C)	Incubation time
Cooling	-	-	-	-20	(15 min)*
Deliver	1	DCM	2 mL	-20	-
Deliver	1	TMSOTf solution	1 mL	-20	3 min
Wash	1	DCM	2 mL	-20	25 sec

*Time required to reach the desired temperature.

Module C: Phosphate Glycosylation (45 min)

The building block solution (0.06 mmol of BB1 or BB2 in 1 mL of DCM per glycosylation) was delivered to the reaction vessel. After the set temperature was reached, the reaction was started by dropwise addition of the TMSOTf

solution (1.0 mL, stoichiometric). After completion of the reaction, the solution was drained and the resin washed with DCM (six times, each with 2 mL for 25 s). The temperature of the reaction vessel was increased to 25 °C for the next module.

Table 13. Parameters of Module C

Action	Cycles	Solution	Amount	T (°C)	Incubation time
Cooling	-	-	-	T ₁	-
Deliver	1	BB solution	1 mL	T ₁	-
Deliver	1	TMSOTf solution	1 mL	T ₁	-
Reaction time (BB dependent)	1			T ₁ to T ₂	t ₁ to t ₂ min
Wash	1	DCM	2 mL	T ₂	5 sec
Heating	-	-	-	25	-
Wash	6	DCM	2 mL	> 0	25 sec

The AGA glycosylation conditions employed for the phosphate BB were previously reported²⁶⁸.

Table 14. Employed conditions for AGA glycosylation

BB	Equiv.	t ₁ (min)	T ₁ (°C)	t ₂ (min)	T ₂ (°C)
BB1	5	5	-35	40	-15
BB2	5	5	-35	40	-15

Module D: Capping (30 min)

The resin was washed with DMF (two times with 2 mL for 25 s) and the temperature of the reaction vessel was adjusted to 25 °C. 2 mL of Pyridine solution (10% in DMF) was delivered into the reaction vessel. After 1 min, the reaction solution was drained and the resin washed with DCM (three times with 3 mL for 25 s). 4 mL of capping solution was delivered into the reaction vessel. After 20 min, the reaction solution was drained and the resin washed with DCM (three times with 3 mL for 25 s).

Table 15. Parameters of Module D

Action	Cycles	Solution	Amount	T (°C)	Incubation time
Heating	-	-	-	25	(5 min)*
Wash	2	DMF	2 mL	25	25 sec
Deliver	1	10% Pyridine in DMF	2 mL	25	1 min
Wash	3	DCM	2 mL	25	25 sec
Deliver	1	Capping Solution	4 mL	25	20 min
Wash	3	DCM	2 mL	25	25 sec

*Time required to reach the desired temperature.

Module E: Fmoc Deprotection (10 min)

The resin was washed with DMF (three times with 2 mL for 25 s) and the temperature of the reaction vessel was adjusted to 25 °C. 2 mL of Fmoc deprotection solution was delivered to the reaction vessel and kept under Ar bubbling. After 5 min, the reaction solution was drained and the resin washed with DMF (three times with 3 mL for 25 s) and DCM (five times each with 2 mL for 25 s). The temperature of the reaction vessel was decreased to -20 °C for the next module.

Table 16. Parameters of Module E

Action	Cycles	Solution	Amount	T (°C)	Incubation time
Wash	3	DMF	2 mL	25	25 sec
Deliver	1	Fmoc depr. Solution 1	2 mL	25	5 min
Wash	1	DMF	2 mL		
Cooling	-	-	-	-20	-
Wash	3	DMF	2 mL	< 25	25 sec
Wash	5	DCM	2 mL	< 25	25 sec

Module F: OLev Deprotection (100 min)

The resin was washed with DCM (three times with 2 mL for 15 s) and the temperature of the reaction vessel was adjusted to 25 °C. Fresh DCM (1.3 mL) was delivered to the reaction vessel, followed by 0.8 mL of OLev deprotection solution. The reaction vessel and kept under Ar bubbling. After 30 min, the reaction solution was drained the whole process was repeated twice more. After that, the resin washed with DMF (three times with 2 mL for 15 s), THF (3 times each with 2 mL for 15 s) and DCM (3 times each with 2 mL for 15 s). The temperature of the reaction vessel was decreased to -20 °C for the next module.

Table 17. Parameters of Module F

Action	Cycles	Solution	Amount	T (°C)	Incubation time
Wash	3	DCM	2 mL	25	15 sec
Deliver	1	DCM	1.3 mL	25	-
Deliver	1	Lev depr. Solution 1	2 mL	25	5 min
Wash	3	DMF	2 mL	25	15 sec
Wash	3	THF	2 mL	25	15 sec

Wash	3	DCM	2 mL	25	15 sec
-------------	---	-----	------	----	--------

4.8.5 Post-AGA manipulations

Module G: Cleavage from Solid Support

The oligosaccharides were cleaved from the solid support using a continuous-flow photoreactor as described previously.²⁶⁹

Module H: Methanolysis

The cleaved oligosaccharide was suspended THF (4 mL). MeONa in MeOH (0.5 M, 0.4 mL) was added and the suspension was gently shaken at room temperature.

Module I: Hydrogenolysis at ambient pressure

The crude compound obtained from *Module H* was dissolved in 2 mL of EtOAc:*t*BuOH:H₂O (2:1:1). 100% by weight Pd/C (10%) was added and the reaction was stirred in a flask equipped with a H₂ balloon. The reaction progress was monitored to avoid undesired side products formation. Upon completion, the reaction was filtered and washed with EtOAc, *t*BuOH, ACN and H₂O (4 mL each). The filtrates were concentrated *in vacuo*.

Module J: Purification

The pure compound was analyzed using analytical HPLC (Agilent 1200 Series, **Method A1/A2**).

- **Method A1:** (Hypercarb column, ThermoFisher scientific, 150 x 4.6 mm, 3 μm) flow rate of 0.7 mL/min with H₂O (0.1% formic acid) and ACN as eluents [isocratic (5 min), linear gradient to 30% ACN (30 min), linear gradient to 100% ACN (5 min), isocratic 100% ACN (5 min)].

- **Method A2 (Prep):** (Hypercarb column, ThermoFisher scientific, 150 x 10 mm, 5 μ m), flow rate of 3 mL /min with H₂O (0.1% formic acid) and ACN as eluents [isocratic (5 min), linear gradient to 30% ACN (30 min), linear gradient to 100% ACN (5 min), isocratic 100% ACN (5 min)].

Following final purification, all deprotected products were lyophilized on a Christ Alpha 2-4 LD plus freeze dryer prior to characterization.

4.8.6 Synthesis of the tetrasaccharide (11)

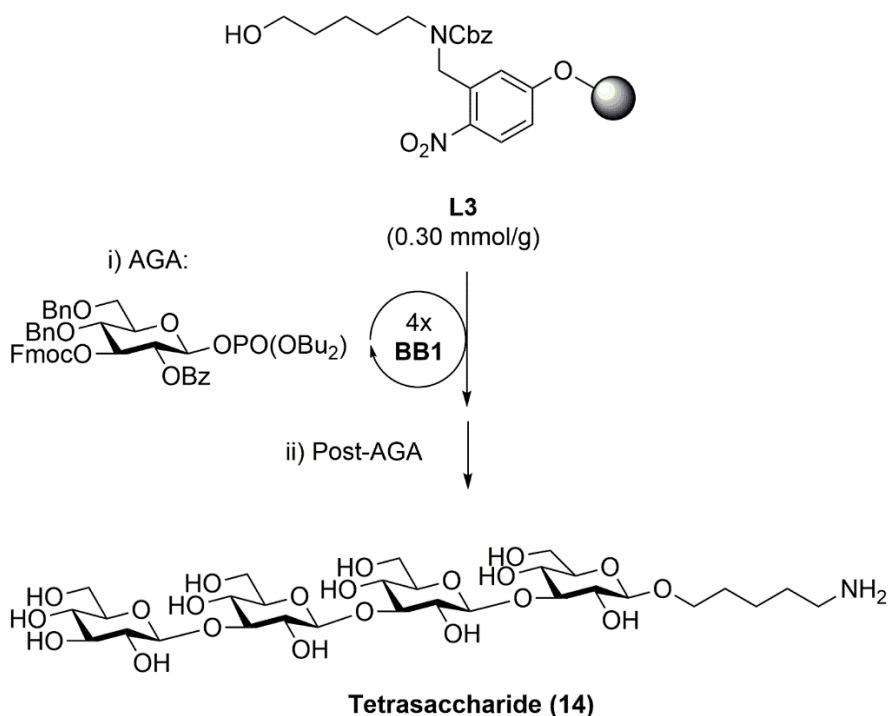


Figure 88. General procedure for the synthesis of tetrasaccharide 14

Table 18. Automation sequence employed for the synthesis of the tetrasaccharide and the process employed afterwards

Step	Modules	Notes
AGA	L3	A
		L3 swelling

	(BB1) _{x4}	(B, C, D, E) _{x4}	C: (BB1, -20 °C for 5 min, 0 °C for 20 min)
Post-AGA		G, H, I, J	H: (16 h) I: (8 h) J: (Method A2, t _R = 13.4 min)

4.8.7 Synthesis of the branched pentasaccharide (12)

Table 19. Automation sequence employed for the synthesis of the branched pentasaccharide

Step	Modules		Notes
AGA	L3	A	L3 swelling
	(BB1) _{x2}	(B, C, D, E) _{x2}	C: (BB1, -20 °C for 5 min, 0 °C for 20 min)
	(BB2) _{x1}	(B, C, D, E, F)	C: (BB2, -20 °C for 5 min, 0 °C for 20 min)
	(BB1)	(B, C*, D) _{x2}	C: (BB1, -20 °C for 5 min, 0 °C for 20 min)

*Double equivalents of BB1

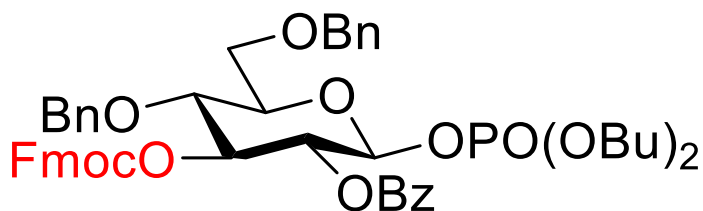
4.8.8 Synthesis of the linear octasaccharide (13)

Table 20. Automation sequence employed for the synthesis of the octasaccharide (13)

Step	Modules		Notes
AGA	L3	A	L3 swelling
	(BB1) _{x8}	(B, C, D, E) _{x8}	C: (BB1, -20 °C for 5 min, 0 °C for 20 min)

4.8.9 Experimental part of building block synthesis

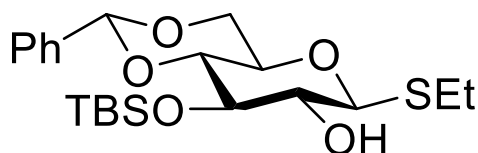
Dibutyl 2-O-benzoyl-4,6-di-O-benzyl-3-O-fluorenylmethoxycarbonyl-1-thio- β -D-glucopyranoside **2** or **BB1**



2

Thioglycoside **1** (2.00 g, 2.73 mmol) was co-evaporated three times with toluene. The residue and NIS (0.751 g, 3.34 mmol) were dissolved in DCM (55 mL, 50 mmol) under an Argon atmosphere and the solution was cooled down to 0°C. Dibutyl hydrogen phosphate (1.00 mL, 5.47 mmol) and triflic acid (16.90 μ L, 0.191 mmol) were added and the reaction was stirred at 0°C for 90 minutes. The reaction was quenched with pyridine, treated with Amberlite IR-120 and filtered through celite. After it was extracted with 10% aqueous $\text{Na}_2\text{S}_2\text{O}_3$ followed by saturated aqueous NaHCO_3 . The organic phase was dried over MgSO_4 and the solvent was removed in vacuo. The crude product was purified by column chromatography on silica gel (hexane/EtOAc = 3:1 to 1:1) to afford **2** or **BB1** (1.19g, 1.4 mmol, 50%). R_f : 0.35 (Hexane/EtOAc/DCM = 1:1:0.5).

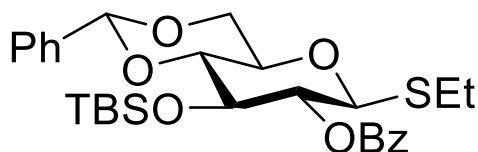
Ethyl 4,6-O-benzylidene-3-O-tert-butyldimethylsilyl-1-thio- β -D-glucopyranoside **4**



4

To a solution of compound **3** (10 g, 32 mmol) in anhydrous DCM (213 mL, 0.15 M) were added TBSCl (7.24 g, 48.0 mmol) and imidazole (4.36 g, 64 mmol) at 0°C, and the mixture was stirred for 18 h at room temperature. After the reaction mixture was quenched with MeOH and saturated with aqueous NaHCO₃, it was diluted with DCM and dried over MgSO₄, and the solvent was evaporated *in vacuo*. The crude product was purified by column chromatography on silica gel (hexane/EtOAc = 9:0.5 to 9:1) to afford **4** (12.6g, 29.4mmol, 92%). R_f: 0.50 (Hexane/EtOAc = 4:1).

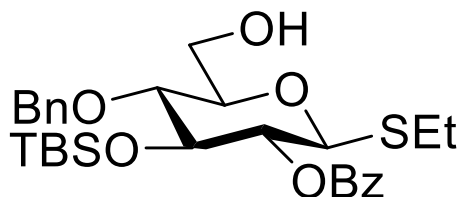
Ethyl 2-O-benzoyl-4,6-O-benzylidene-3-O-tert-butyldimethylsilyl-1-thio- β -D-glucopyranoside **5**



5

To a solution of compound **4** (1.99 g, 4.67 mmol) in pyridine (9.5 mL, 0.5 M) were added BzCl (1.63 mL, 14 mmol) and DMAP (0.11 g, 0.93 mmol), and the mixture was stirred for 12 h at 70°C. After the reaction mixture was quenched with MeOH and brine. Then diluted with DCM and dried over MgSO₄, and the solvent was evaporated *in vacuo*. The crude product was purified by column chromatography on silica gel (hexane/EtOAc = 9:1) to afford **5** (2.3g, 4.3mmol, 92%). R_f: 0.43 (Hexane/EtOAc = 4:1).

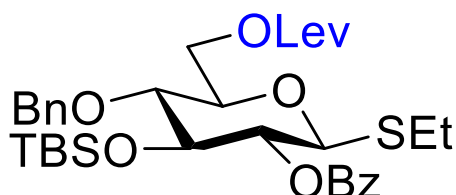
Ethyl 2-O-benzoyl-4-O-benzyl-3-O-tert-butyldimethylsilyl-1-thio-β-D-glucopyranoside 6



6

Compound 5 (9.80 g, 18.48 mmol) was co-evaporated with toluene and dissolved under an Argon atmosphere in DCM (123 mL, 0.15 M). To this solution were added $\text{BH}_3 \cdot \text{OEt}_2$ (8.82 ml, 92.4 mmol), and TMSOTf (0.333 ml, 1.85 mmol) successively at 0 °C. The mixture was stirred for 6 h at 0 °C. After the mixture was quenched with saturated aqueous NaHCO_3 it was diluted with DCM and dried over MgSO_4 . The combined organic layer was evaporated in vacuo. The crude compound was purified by column chromatography on silica gel (hexane/EtOAc = 9:1 to 4:1) to afford **6** (6.88g, 12.9mmol, 70%). R_f : 0.15 (Hexane/EtOAc = 4:1).

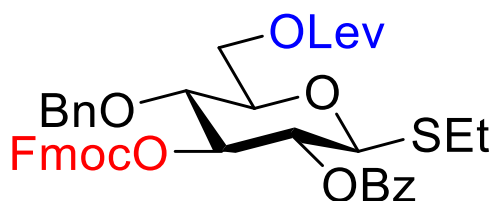
Ethyl 2-O-benzoyl-4-O-benzyl-3-O-tert-butyldimethylsilyl-6-O-levulinyl-1-thio- β -D-glucopyranoside 7



7

To a solution of compound 6 (7.31 g, 13.73 mmol) in anhydrous DCM (69 ml, 0.2 M) were added levulinic acid (4.23 mL, 41.18 mmol) and DIC (6.37 ml, 41.18 mmol), and a catalytic amount of DMAP (0.50 g, 4.12 mmol) at 0 °C, and the mixture was stirred for 1 h at 0 °C. After the mixture was quenched with saturated aqueous NaHCO₃ it was diluted with DCM, extracted with DCM, and the combined organic layer was dried over MgSO₄ and evaporated in vacuo. The crude compound was purified by column chromatography on silica gel (hexane/EtOAc = 3:1 to 1:1) to afford **7** (8.13g, 12.9 mmol, 94%). R_f: 0.30 (Hexane/EtOAc = 4:1).

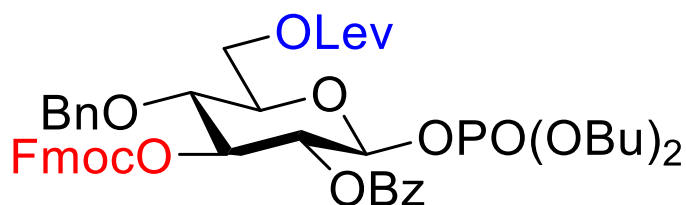
Ethyl 2-O-benzoyl-4-O-benzyl-3-O-fluorenylmethoxycarbonyl-6-Olevulinyl-1-thio-β-D-glucopyranoside 9



9

To a solution of 8 (8.00 g, 12.69 mmol) in anhydrous acetonitrile (158 ml, 0.08 M) was added boron trifluoride etherate ($\text{BF}_3 \cdot \text{OEt}_2$) (1.88 mL, 15.23 mmol) at 0 °C, and the mixture was stirred for 20 min. at 0 °C. After the mixture was quenched with saturated aqueous NaHCO_3 it was diluted with DCM, and the combined organic layer was dried over MgSO_4 and evaporated in vacuo. To a solution of this crude product in DCM (63.5 ml, 0.2 M) were added 9-fluorenylmethyl chloroformate (6.56 g, 25.38 mmol) and pyridine (6.15 mL, 76.14 mmol) successively at 0 °C, and the mixture was stirred 2 h at room temperature. After the mixture was quenched with saturated aqueous NaHCO_3 it was diluted with DCM and extracted with DCM, then the combined organic layer was washed with 1M aqueous HCl, dried over MgSO_4 and evaporated in vacuo. The crude product was purified by column chromatography on silica gel (hexane/EtOAc = 3:2) to afford **9** (7.85g, 10.66 mmol, 84% over two steps). R_f : 0.36 (Hexane/EtOAc = 4:1).

Dibutyl 2-O-benzoyl-4-O-benzyl-3-O-fluorenylmethoxycarbonyl-6-O-levulinyl-1-thio- β -D-glucopyranoside **10** or **BB2**



Thioglycoside **9** (2.54 g, 3.44 mmol) was co-evaporated twice with toluene. The residue and NIS (0.94 g, 4.19 mmol) were dissolved in DCM (69 mL, 0.05 M) under an Argon atmosphere and the solution was cooled to 0 °C. Dibutyl hydrogen phosphate (1.36 mL, 6.88 mmol) and triflic acid (21.2 μ L, 0.24 mmol) were added and the reaction was stirred at 0 °C for 90 minutes. After complete conversion of the starting material the reaction was quenched with pyridine, treated with Amberlite IR-120 and filtered through celite. After it was extracted with 10% aqueous $\text{Na}_2\text{S}_2\text{O}_3$ followed by saturated aqueous NaHCO_3 . The organic phase was dried over MgSO_4 and the solvent was removed in vacuo. The crude product was purified by silica gel flash column chromatograph (hexane/EtOAc = 3:1 to 1:1) to afford **10** or **BB2** (2.56g, 2.9 mmol, 84%). R_f : 0.20 (Hexane/EtOAc = 4:1).

5 General conclusions

1A.-The combination of Nuclear Magnetic Resonance (NMR), Isothermal Titration Calorimetry (ITC) and Molecular Dynamic (MD) simulations has provided key insights into the recognition events of blood group antigens by galectin-4. The use of the single carbohydrate recognition domains enabled understanding the binding preferences of each domain. Regarding the nature of the blood group, whereas the N-terminal domain prefers group B antigens (terminal α Gal), the C-terminal one prefers group A antigens (terminal α GalNAc).

1B.-Regarding the peripheral core presentation, both domains prefer type-6 antigens (Gal β 1-4Glc). Additionally, for both domains, the substitution of this lactose in the reducing-end by a LacNAc (type-2, Gal β 1-4GlcNAc) was detrimental for binding. Concerning the type of linkage (β 1,3GlcNAc vs β 1,4GlcNAc), while for the N-terminal domain no effect on the affinity was observed, the C-terminal domain preferred β 1,3 linked GlcNAc (type-1 vs type-2). Finally, for both domains, the type-4 (Gal β 1,3GalNAc) antigens were the worst binders, as a result of the loss of interaction of the hydroxyl group at position 4 of the GalNAc.

1C.-Overall, both domains displayed similar affinity; a fact that hindered the estimation of accurate dissociation constants within the full-length (FL) form for each domain. Thus, the production of two mutants of galectin-4 (Gal-4FL-H236R and Gal-4FL-H63R), with only one active site, enabled obtaining accurate dissociation constants for each domain in the FL form. In particular, the NMR and ITC data permitted demonstrating that both domains behave independently, even if they compete for the same ligands.

2.-NMR experiments using the LPS from *E. Coli* O55 and galectin-4 showed that the lectin binds exogenous glycans from pathogens. The structural conclusions obtained from the studies with the single domains are in agreement with those obtained for the LPS, as the C-terminal domain bound better to this

LPS than the N-terminal one. Thus, the previous conclusions can be employed to understand how galectin-4 binds pathogenic glycans mimicking blood group antigens. The observed precipitation of the galectin-4 mutants in the presence of LPS suggests that these proteins could crosslink different LPS molecules, despite only having one active site, probably through dimerization of galectin-4.

3.-The structural features of the interaction between galectin-4 and the glycoprotein CD14 have also been unravelled. As first step, the glycoprofile of CD14 has been determined. Whereas CD14 presents high mannose (Man₉ and Man₅) and complex type N-glycans, the interaction occurs through the LacNAc and 3'sialyl LacNAc epitopes present in the complex type glycans. Mass spectrometry (MS) enabled detecting two glycosylated sites in CD14 and determining the type of glycan present in each site. MS showed that the N282 glycosylation site only contains the Man₉ and Man₈. Using NMR, it has been shown that this glycan is not accessible to the solvent, as treatment with the endo-H glycosidase did not remove the glycan from the protein. The other glycosylation site, N151, is exposed to the solvent and presented glycan heterogeneity. The lack of accessibility of the glycan at N282 is determinant for the nature of the glycan at this site (Man₉, Man₈), since it cannot be modified by glycosidases. The N151 site, being exposed, it can be modified and thus presents multiple type of glycans.

Additionally, the relevance of each N-glycan in the structure and folding of the protein was revealed. Mutation of the exposed N-glycosylation site (N151) resulted in a small decrease in expression level (ca. 20%), although the protein showed larger tendency to aggregate (one third was aggregated). Mutation of the non-exposed site (N282) drastically decreased the protein expression levels (>80%) and induced improper folding of CD14 (>95% aggregated), highlighting the relevance of N-glycans in the folding and stability of the protein.

4.-In the quest for well-defined β -glucans as Dectin-1 ligands, two building blocks have been synthesized and the synthesis of a linear β 1,3-linked tetrasaccharide using a solid phase synthesizer was achieved.

6 References

- (1) Bucior, I.; Burger, M. M. Carbohydrate-Carbohydrate Interactions in Cell Recognition. *Curr. Opin. Struct. Biol.* **2004**, *14* (5), 631–637. <https://doi.org/10.1016/j.sbi.2004.08.006>.
- (2) Oppenheimer, S. B. Cell Surface Carbohydrates in Adhesion and Migration. *Integr. Comp. Biol.* **1978**, *18* (1), 13–23. <https://doi.org/10.1093/icb/18.1.13>.
- (3) Furukawa, K.; Ohkawa, Y.; Yamauchi, Y.; Hamamura, K.; Ohmi, Y.; Furukawa, K. Fine Tuning of Cell Signals by Glycosylation. *J. Biochem.* **2012**, *151* (6), 573–578. <https://doi.org/10.1093/jb/mvs043>.
- (4) Karlsson, K. A. Pathogen-Host Protein-Carbohydrate Interactions as the Basis of Important Infections. *Adv. Exp. Med. Biol.* **2001**, *491*, 431–443. https://doi.org/10.1007/978-1-4615-1267-7_28.
- (5) Nardy, A. F. F. R.; Freire-de-Lima, L.; Freire-de-Lima, C. G.; Morrot, A. The Sweet Side of Immune Evasion: Role of Glycans in the Mechanisms of Cancer Progression. *Front. Oncol.* **2016**, *6* (MAR), 1–7. <https://doi.org/10.3389/fonc.2016.00054>.
- (6) Ouellette, R. J.; Rawn, J. D. Carbohydrates. In *Principles of Organic Chemistry*; Elsevier, 2015; Vol. 17, pp 343–370. <https://doi.org/10.1016/B978-0-12-802444-7.00013-6>.
- (7) Kendrew, J. C.; Moelwyn-Hughes, E. A. The Kinetics of Mutarotation in Solution. *Proc. R. Soc. London. Ser. A. Math. Phys. Sci.* **1940**, *176* (966), 352–367. <https://doi.org/10.1098/rspa.1940.0094>.
- (8) Juaristi, E.; Cuevas, G. Recent Studies of the Anomeric Effect. *Tetrahedron* **1992**, *48* (24), 5019–5087. [https://doi.org/10.1016/S0040-4020\(01\)90118-8](https://doi.org/10.1016/S0040-4020(01)90118-8).
- (9) Synytsya, A.; Novak, M. Structural Analysis of Glucans. *Ann. Transl. Med.* **2014**, *2* (2), 1–14. <https://doi.org/10.3978/j.issn.2305-5839.2014.02.07>.
- (10) Büll, C.; Nason, R.; Sun, L.; Van Coillie, J.; Sørensen, D. M.; Moons, S. J.; Yang, Z.; Arbitman, S.; Fernandes, S. M.; Furukawa, S.; McBride, R.; Nycholat, C. M.; Adema, G. J.; Paulson, J. C.; Schnaar, R. L.; Boltje, T. J.; Clausen, H.; Narimatsu, Y. Probing the Binding Specificities of Human Siglecs by Cell-Based Glycan Arrays. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118* (17), 1–12. <https://doi.org/10.1073/pnas.2026102118>.
- (11) Tvaroška, I.; Bleha, T. Anomeric and Exo-Anomeric Effects in

- Carbohydrate Chemistry. *Adv. Carbohydr. Chem. Biochem.* **1989**, *47* (C), 45–123. [https://doi.org/10.1016/S0065-2318\(08\)60412-6](https://doi.org/10.1016/S0065-2318(08)60412-6).
- (12) Bock, K.; Duus, J. A Conformational Study of Hydroxymethyl Groups in Carbohydrates Investigated by ¹H Nmr Spectroscopy. *J. Carbohydr. Chem.* **1994**, *13* (4), 513–543. <https://doi.org/10.1080/07328309408011662>.
- (13) Luft, J. H. Fine Structures of Capillary and Endocapillary Layer as Revealed by Ruthenium Red. *Fed. Proc.* *25* (6), 1773–1783.
- (14) Reitsma, S.; Slaaf, D. W.; Vink, H.; Van Zandvoort, M. A. M. J.; Oude Egbrink, M. G. A. The Endothelial Glycocalyx: Composition, Functions, and Visualization. *Pflugers Arch. Eur. J. Physiol.* **2007**, *454* (3), 345–359. <https://doi.org/10.1007/s00424-007-0212-8>.
- (15) Möckl, L. The Emerging Role of the Mammalian Glycocalyx in Functional Membrane Organization and Immune System Regulation. *Front. Cell Dev. Biol.* **2020**, *8* (April), 1–14. <https://doi.org/10.3389/fcell.2020.00253>.
- (16) An, H. J.; Froehlich, J. W.; Lebrilla, C. B. Determination of Glycosylation Sites and Site-Specific Heterogeneity in Glycoproteins. *Curr. Opin. Chem. Biol.* **2009**, *13* (4), 421–426. <https://doi.org/10.1016/j.cbpa.2009.07.022>.
- (17) Varki, A. Biological Roles of Glycans. *Glycobiology* **2017**, *27* (1), 3–49. <https://doi.org/10.1093/glycob/cww086>.
- (18) Schjoldager, K. T.; Narimatsu, Y.; Joshi, H. J.; Clausen, H. Global View of Human Protein Glycosylation Pathways and Functions. *Nat. Rev. Mol. Cell Biol.* **2020**, *21* (12), 729–749. <https://doi.org/10.1038/s41580-020-00294-x>.
- (19) Kinoshita, T. Glycosylphosphatidylinositol (GPI) Anchors: Biochemistry and Cell Biology: Introduction to a Thematic Review Series. *J. Lipid Res.* **2016**, *57* (1), 4–5. <https://doi.org/10.1194/jlr.E065417>.
- (20) Aebi, M. N-Linked Protein Glycosylation in the ER. *Biochim. Biophys. Acta - Mol. Cell Res.* **2013**, *1833* (11), 2430–2437. <https://doi.org/10.1016/j.bbamcr.2013.04.001>.
- (21) Fisher, P.; Thomas-Oates, J.; Wood, A. J.; Ungar, D. The N-Glycosylation Processing Potential of the Mammalian Golgi Apparatus. *Front. Cell Dev. Biol.* **2019**, *7* (August), 1–11. <https://doi.org/10.3389/fcell.2019.00157>.

- (22) Cao, L.; Diedrich, J. K.; Ma, Y.; Wang, N.; Pauthner, M.; Park, S. K. R.; Delahunty, C. M.; McLellan, J. S.; Burton, D. R.; Yates, J. R.; Paulson, J. C. Global Site-Specific Analysis of Glycoprotein N-Glycan Processing. *Nat. Protoc.* **2018**, *13* (6), 1196–1212. <https://doi.org/10.1038/nprot.2018.024>.
- (23) Van Den Steen, P.; Rudd, P. M.; Dwek, R. A.; Opdenakker, G. Concepts and Principles of O-Linked Glycosylation. *Crit. Rev. Biochem. Mol. Biol.* **1998**, *33* (3), 151–208. <https://doi.org/10.1080/10409239891204198>.
- (24) Jensen, P. H.; Kolarich, D.; Packer, N. H. Mucin-Type O-Glycosylation - Putting the Pieces Together. *FEBS J.* **2010**, *277* (1), 81–94. <https://doi.org/10.1111/j.1742-4658.2009.07429.x>.
- (25) Holdener, B. C.; Haltiwanger, R. S. Protein O-Fucosylation: Structure and Function. *Curr. Opin. Struct. Biol.* **2019**, *56* (1), 78–86. <https://doi.org/10.1016/j.sbi.2018.12.005>.
- (26) Liu, C.; Li, J. O-GlcNAc: A Sweetheart of the Cell Cycle and DNA Damage Response. *Front. Endocrinol. (Lausanne)*. **2018**, *9* (July), 1–11. <https://doi.org/10.3389/fendo.2018.00415>.
- (27) Bennett, E. P.; Mandel, U.; Clausen, H.; Gerken, T. A.; Fritz, T. A.; Tabak, L. A. Control of Mucin-Type O-Glycosylation: A Classification of the Polypeptide GalNAc-Transferase Gene Family. *Glycobiology* **2012**, *22* (6), 736–756. <https://doi.org/10.1093/glycob/cwr182>.
- (28) Coelho, H.; Rivas, M. de las; Grosso, A. S.; Diniz, A.; Soares, C. O.; Francisco, R. A.; Dias, J. S.; Compañón, I.; Sun, L.; Narimatsu, Y.; Vakhrushev, S. Y.; Clausen, H.; Cabrita, E. J.; Jiménez-Barbero, J.; Corzana, F.; Hurtado-Guerrero, R.; Marcelo, F. Atomic and Specificity Details of Mucin 1 O -Glycosylation Process by Multiple Polypeptide GalNAc-Transferase Isoforms Unveiled by NMR and Molecular Modeling . *JACS Au* **2022**, *2* (3), 631–645. <https://doi.org/10.1021/jacsau.1c00529>.
- (29) Salinas-Marín, R.; Villanueva-Cabello, T. M.; Martínez-Duncker, I. Biology of Proteoglycans and Associated Glycosaminoglycans. In *Comprehensive Glycoscience*; Elsevier, 2021; pp 63–102. <https://doi.org/10.1016/B978-0-12-819475-1.00065-1>.
- (30) Walimbe, T.; Panitch, A. Proteoglycans in Biomedicine: Resurgence of an Underexploited Class of ECM Molecules. *Front. Pharmacol.* **2020**, *10* (January), 1–13. <https://doi.org/10.3389/fphar.2019.01661>.
- (31) Khoder-Agha, F.; Kietzmann, T. The Glyco-Redox Interplay: Principles and Consequences on the Role of Reactive Oxygen Species during

- Protein Glycosylation. *Redox Biol.* **2021**, *42*, 101888. <https://doi.org/10.1016/j.redox.2021.101888>.
- (32) Hennicke, J.; Reinhart, D.; Altmann, F.; Kunert, R. Impact of Temperature and PH on Recombinant Human IgM Quality Attributes and Productivity. *N. Biotechnol.* **2019**, *50* (August 2018), 20–26. <https://doi.org/10.1016/j.nbt.2019.01.001>.
- (33) Schnaar, R. L. Biosynthesis of Glycolipids. In *Comprehensive Glycoscience*; Elsevier, 2021; pp 117–126. <https://doi.org/10.1016/B978-0-12-409547-2.14949-2>.
- (34) Russo, D.; Capolupo, L.; Loomba, J. S.; Sticco, L.; D'Angelo, G. Glycosphingolipid Metabolism in Cell Fate Specification. *J. Cell Sci.* **2018**, *131* (24), 1–11. <https://doi.org/10.1242/jcs.219204>.
- (35) Breimer, M. E. Tissue Specificity of Glycosphingolipids as Expressed in Pancreas and Small Intestine of Blood Group A and B Human Individuals. *Arch. Biochem. Biophys.* **1984**, *228* (1), 71–85. [https://doi.org/10.1016/0003-9861\(84\)90048-1](https://doi.org/10.1016/0003-9861(84)90048-1).
- (36) Lingwood, C. A. Glycosphingolipid Functions. *Cold Spring Harb. Perspect. Biol.* **2011**, *3* (7), 1–26. <https://doi.org/10.1101/cshperspect.a004788>.
- (37) Flynn, R. A.; Pedram, K.; Malaker, S. A.; Batista, P. J.; Smith, B. A. H.; Johnson, A. G.; George, B. M.; Majzoub, K.; Villalta, P. W.; Carette, J. E.; Bertozzi, C. R. Small RNAs Are Modified with N-Glycans and Displayed on the Surface of Living Cells. *Cell* **2021**, *184* (12), 3109–3124.e22. <https://doi.org/10.1016/j.cell.2021.04.023>.
- (38) Péanne, R.; de Lonlay, P.; Foulquier, F.; Kornak, U.; Lefeber, D. J.; Morava, E.; Pérez, B.; Seta, N.; Thiel, C.; Van Schaftingen, E.; Matthijs, G.; Jaeken, J. Congenital Disorders of Glycosylation (CDG): Quo Vadis? *Eur. J. Med. Genet.* **2018**, *61* (11), 643–663. <https://doi.org/10.1016/j.ejmg.2017.10.012>.
- (39) Chang, I. J.; He, M.; Lam, C. T. Congenital Disorders of Glycosylation. *Ann. Transl. Med.* **2018**, *6* (24), 477–477. <https://doi.org/10.21037/atm.2018.10.45>.
- (40) Pietrobono, S.; Stecca, B. Aberrant Sialylation in Cancer: Biomarker and Potential Target for Therapeutic Intervention? *Cancers (Basel)*. **2021**, *13* (9). <https://doi.org/10.3390/cancers13092014>.
- (41) Mereiter, S.; Balmaña, M.; Campos, D.; Gomes, J.; Reis, C. A. Glycosylation in the Era of Cancer-Targeted Therapy: Where Are We Heading? *Cancer Cell* **2019**, *36* (1), 6–16.

<https://doi.org/10.1016/j.ccell.2019.06.006>.

- (42) Berti, F.; Romano, M. R.; Micoli, F.; Adamo, R. Carbohydrate Based Meningococcal Vaccines: Past and Present Overview. *Glycoconj. J.* **2021**, *38* (4), 401–409. <https://doi.org/10.1007/s10719-021-09990-y>.
- (43) Mazalovska, M.; Kouokam, J. C. Lectins as Promising Therapeutics for the Prevention and Treatment of HIV and Other Potential Coinfections. *Biomed Res. Int.* **2018**, *2018*. <https://doi.org/10.1155/2018/3750646>.
- (44) Li, Y.; Liu, D.; Wang, Y.; Su, W.; Liu, G.; Dong, W. The Importance of Glycans of Viral and Host Proteins in Enveloped Virus Infection. *Front. Immunol.* **2021**, *12* (April), 1–12. <https://doi.org/10.3389/fimmu.2021.638573>.
- (45) Sharon, N.; Lis, H. Lectins as Cell Recognition Molecules. *Science* (80-.). **1989**, *246* (4927), 227–234.
- (46) Asensio, J. L.; Ardá, A.; Cañada, F. J.; Jiménez-Barbero, J. Carbohydrate–Aromatic Interactions. *Acc. Chem. Res.* **2013**, *46* (4), 946–954. <https://doi.org/10.1021/ar300024d>.
- (47) Hudson, K. L.; Bartlett, G. J.; Diehl, R. C.; Agirre, J.; Gallagher, T.; Kiessling, L. L.; Woolfson, D. N. Carbohydrate-Aromatic Interactions in Proteins. *J. Am. Chem. Soc.* **2015**, *137* (48), 15152–15160. <https://doi.org/10.1021/jacs.5b08424>.
- (48) Claveria-Gimeno, R.; Vega, S.; Abian, O.; Velazquez-Campoy, A. A Look at Ligand Binding Thermodynamics in Drug Discovery. *Expert Opin. Drug Discov.* **2017**, *12* (4), 363–377. <https://doi.org/10.1080/17460441.2017.1297418>.
- (49) De Schutter, K.; Van Damme, E. J. M. Protein-Carbohydrate Interactions as Part of Plant Defense and Animal Immunity. *Molecules* **2015**, *20* (5), 9029–9053. <https://doi.org/10.3390/molecules20059029>.
- (50) Valverde, P.; Martínez, J. D.; Cañada, F. J.; Ardá, A.; Jiménez-Barbero, J. Molecular Recognition in C-Type Lectins: The Cases of DC-SIGN, Langerin, MGL, and L-Sectin. *ChemBioChem* **2020**, *21* (21), 2999–3025. <https://doi.org/10.1002/cbic.202000238>.
- (51) Lenza, M. P.; Atxabal, U.; Oyenarte, I.; Jiménez-Barbero, J.; Ereño-Orbea, J. Current Status on Therapeutic Molecules Targeting Siglec Receptors. *Cells* **2020**, *9* (12), 1–19. <https://doi.org/10.3390/cells9122691>.
- (52) Johannes, L.; Jacob, R.; Leffler, H. Galectins at a Glance. *J. Cell Sci.*

- 2018**, *131* (9), 1–9. <https://doi.org/10.1242/jcs.208884>.
- (53) Bertuzzi, S.; Quintana, J. I.; Ardá, A.; Gimeno, A.; Jiménez-Barbero, J. Targeting Galectins With Glycomimetics. *Front. Chem.* **2020**, *8* (August), 1–17. <https://doi.org/10.3389/fchem.2020.00593>.
- (54) Ideo, H.; Matsuzaka, T.; Nonaka, T.; Seko, A.; Yamashita, K. Galectin-8-N-Domain Recognition Mechanism for Sialylated and Sulfated Glycans. *J. Biol. Chem.* **2011**, *286* (13), 11346–11355. <https://doi.org/10.1074/jbc.M110.195925>.
- (55) Hirabayashi, J.; Hashidate, T.; Arata, Y.; Nishi, N.; Nakamura, T.; Hirashima, M.; Urashima, T.; Oka, T.; Futai, M.; Muller, W. E. G.; Yagi, F.; Kasai, K. I. Oligosaccharide Specificity of Galectins: A Search by Frontal Affinity Chromatography. *Biochim. Biophys. Acta - Gen. Subj.* **2002**, *1572* (2–3), 232–254. [https://doi.org/10.1016/S0304-4165\(02\)00311-2](https://doi.org/10.1016/S0304-4165(02)00311-2).
- (56) Bertuzzi, S.; Gimeno, A.; Núñez-Franco, R.; Bernardo-Seisdedos, G.; Delgado, S.; Jiménez-Osés, G.; Millet, O.; Jiménez-Barbero, J.; Ardá, A. Unravelling the Time Scale of Conformational Plasticity and Allostery in Glycan Recognition by Human Galectin-1. *Chem. - A Eur. J.* **2020**, *26* (67), 15643–15653. <https://doi.org/10.1002/chem.202003212>.
- (57) Lima, C. D. L.; Coelho, H.; Gimeno, A.; Trovão, F.; Diniz, A.; Dias, J. S.; Jiménez-Barbero, J.; Corzana, F.; Carvalho, A. L.; Cabrita, E. J.; Marcelo, F. Structural Insights into the Molecular Recognition Mechanism of the Cancer and Pathogenic Epitope, LacdiNAc by Immune-Related Lectins. *Chem. - A Eur. J.* **2021**, *27* (29), 7951–7958. <https://doi.org/10.1002/chem.202100800>.
- (58) Gimeno, A.; Delgado, S.; Valverde, P.; Bertuzzi, S.; Berbís, M. A.; Echavarren, J.; Lacetera, A.; Martín-Santamaría, S.; Surolia, A.; Cañada, F. J.; Jiménez-Barbero, J.; Ardá, A. Minimizing the Entropy Penalty for Ligand Binding: Lessons from the Molecular Recognition of the Histo Blood-Group Antigens by Human Galectin-3. *Angew. Chemie - Int. Ed.* **2019**, *58* (22), 7268–7272. <https://doi.org/10.1002/anie.201900723>.
- (59) Gómez-Redondo, M.; Delgado, S.; Núñez-Franco, R.; Jiménez-Osés, G.; Ardá, A.; Jiménez-Barbero, J.; Gimeno, A. The Two Domains of Human Galectin-8 Bind Sialyl- and Fucose-Containing Oligosaccharides in an Independent Manner. A 3D View by Using NMR. *RSC Chem. Biol.* **2021**, *2* (3), 932–941. <https://doi.org/10.1039/d1cb00051a>.
- (60) Ideo, H.; Seko, A.; Yamashita, K. Galectin-4 Binds to Sulfated Glycosphingolipids and Carcinoembryonic Antigen in Patches on the Cell Surface of Human Colon Adenocarcinoma Cells. *J. Biol. Chem.*

- 2005**, 280 (6), 4730–4737. <https://doi.org/10.1074/jbc.M410362200>.
- (61) Nielsen, M. I.; Stegmayr, J.; Grant, O. C.; Yang, Z.; Nilsson, U. J.; Boos, I.; Carlsson, M. C.; Woods, R. J.; Unverzagt, C.; Leffler, H.; Wandall, H. H. Galectin Binding to Cells and Glycoproteins with Genetically Modified Glycosylation Reveals Galectin–Glycan Specificities in a Natural Context. *J. Biol. Chem.* **2018**, 293 (52), 20249–20262. <https://doi.org/10.1074/jbc.RA118.004636>.
- (62) Horlacher, T.; Oberli, M. A.; Werz, D. B.; Kröck, L.; Bufali, S.; Mishra, R.; Sobek, J.; Simons, K.; Hirashima, M.; Niki, T.; Seeberger, P. H. Determination of Carbohydrate-Binding Preferences of Human Galectins with Carbohydrate Microarrays. *ChemBioChem* **2010**, 11 (11), 1563–1573. <https://doi.org/10.1002/cbic.201000020>.
- (63) Blixt, O.; Head, S.; Mondala, T.; Scanlan, C.; Huflejt, M. E.; Alvarez, R.; Bryan, M. C.; Fazio, F.; Calarese, D.; Stevens, J.; Razi, N.; Stevens, D. J.; Skehel, J. J.; Van Die, I.; Burton, D. R.; Wilson, I. A.; Cummings, R.; Bovin, N.; Wong, C. H.; Paulson, J. C. Printed Covalent Glycan Array for Ligand Profiling of Diverse Glycan Binding Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, 101 (49), 17033–17038. <https://doi.org/10.1073/pnas.0407902101>.
- (64) Dahlqvist, A.; Leffler, H.; Nilsson, U. J. C1-Galactopyranosyl Heterocycle Structure Guides Selectivity: Triazoles Prefer Galectin-1 and Oxazoles Prefer Galectin-3. *ACS Omega* **2019**, 4 (4), 7047–7053. <https://doi.org/10.1021/acsomega.9b00320>.
- (65) Bojarová, P.; Tavares, M. R.; Laaf, D.; Bumba, L.; Petrásková, L.; Konefał, R.; Bláhová, M.; Pelantová, H.; Elling, L.; Etrych, T.; Chytil, P.; Křen, V. Biocompatible Glyconanomaterials Based on HPMA-Copolymer for Specific Targeting of Galectin-3. *J. Nanobiotechnology* **2018**, 16 (1), 1–16. <https://doi.org/10.1186/s12951-018-0399-1>.
- (66) Hassan, M.; Baussière, F.; Guzelj, S.; Sundin, A. P.; Håkansson, M.; Kovačič, R.; Leffler, H.; Tomašič, T.; Anderluh, M.; Jakopin, Ž.; Nilsson, U. J. Structure-Guided Design of d -Galactal Derivatives with High Affinity and Selectivity for the Galectin-8 N-Terminal Domain. *ACS Med. Chem. Lett.* **2021**, 12 (11), 1745–1752. <https://doi.org/10.1021/acsmchemlett.1c00371>.
- (67) Leffler, H.; Masiarz, F. R.; Barondes, S. H. Soluble Lactose-Binding Vertebrate Lectins: A Growing Family. *Biochemistry* **1989**, 28 (23), 9222–9229. <https://doi.org/10.1021/bi00449a039>.
- (68) Oda, Y.; Herrmann, J.; Gitt, M. A.; Turck, C. W.; Burlingame, A. L.; Barondes, S. H.; Leffler, H. Soluble Lactose-Binding Lectin from Rat

- Intestine with Two Different Carbohydrate-Binding Domains in the Same Peptide Chain. *J. Biol. Chem.* **1993**, 268 (8), 5929–5939. [https://doi.org/10.1016/s0021-9258\(18\)53409-8](https://doi.org/10.1016/s0021-9258(18)53409-8).
- (69) Bum-Erdene, K.; Leffler, H.; Nilsson, U. J.; Blanchard, H. Structural Characterization of Human Galectin-4 C-Terminal Domain: Elucidating the Molecular Basis for Recognition of Glycosphingolipids, Sulfated Saccharides and Blood Group Antigens. *FEBS J.* **2015**, 282 (17), 3348–3367. <https://doi.org/10.1111/febs.13348>.
- (70) Rustiguel, J. K.; Soares, R. O. S.; Meisburger, S. P.; Davis, K. M.; Malzbender, K. L.; Ando, N.; Dias-Baruffi, M.; Nonato, M. C. Full-Length Model of the Human Galectin-4 and Insights into Dynamics of Inter-Domain Communication. *Sci. Rep.* **2016**, 6 (September), 1–13. <https://doi.org/10.1038/srep33633>.
- (71) Bum-Erdene, K.; Leffler, H.; Nilsson, U. J.; Blanchard, H. Structural Characterisation of Human Galectin-4 N-Terminal Carbohydrate Recognition Domain in Complex with Glycerol, Lactose, 3'-Sulfo-Lactose, and 2'-Fucosyllactose. *Sci. Rep.* **2016**, 6 (October 2015), 1–13. <https://doi.org/10.1038/srep20289>.
- (72) Knirel, Y. A.; Gabius, H. J.; Blixt, O.; Rapoport, E. M.; Khasbiullina, N. R.; Shilova, N. V.; Bovin, N. V. Human Tandem-Repeat-Type Galectins Bind Bacterial Non-BGal Polysaccharides. *Glycoconj. J.* **2014**, 31 (1), 7–12. <https://doi.org/10.1007/s10719-013-9497-3>.
- (73) Ideo, H.; Seko, A.; Ohkura, T.; Matta, K. L.; Yamashita, K. High-Affinity Binding of Recombinant Human Galectin-4 to SO 3 – → 3Gal β 1 → 3GalNAc Pyranoside. *Glycobiology* **2002**, 12 (3), 199–208.
- (74) Vokhmyanina, O. A.; Rapoport, E. M.; André, S.; Severov, V. V.; Ryzhov, I.; Pazynina, G. V.; Korchagina, E.; Gabius, H. J.; Bovin, N. V. Comparative Study of the Glycan Specificities of Cell-Bound Human Tandem-Repeat-Type Galectin-4,-8 and-9. *Glycobiology* **2012**, 22 (9), 1207–1217. <https://doi.org/10.1093/glycob/cws079>.
- (75) Danielsen, E. M.; Van Deurs, B. Galectin-4 and Small Intestinal Brush Border Enzymes Form Clusters. *Mol. Biol. Cell* **1997**, 8 (11), 2241–2251. <https://doi.org/10.1091/mbc.8.11.2241>.
- (76) Delacour, D.; Gouyer, V.; Zanetta, J. P.; Drobecq, H.; Leteurtre, E.; Grard, G.; Moreau-Hannedouche, O.; Maes, E.; Pons, A.; André, S.; Le Bivic, A.; Gabius, H. J.; Manninen, A.; Simons, K.; Huet, G. Galectin-4 and Sulfatides in Apical Membrane Trafficking in Enterocyte-like Cells. *J. Cell Biol.* **2005**, 169 (3), 491–501. <https://doi.org/10.1083/jcb.200407073>.

- (77) Stechly, L.; Morelle, W.; Dessein, A. F.; André, S.; Grard, G.; Trinel, D.; Dejonghe, M. J.; Leteurtre, E.; Drobecq, H.; Trugnan, G.; Gabius, H. J.; Huet, G. Galectin-4-Regulated Delivery of Glycoproteins to the Brush Border Membrane of Enterocyte-like Cells. *Traffic* **2009**, *10* (4), 438–450. <https://doi.org/10.1111/j.1600-0854.2009.00882.x>.
- (78) Stancic, M.; Slijepcevic, D.; Nomden, A.; Vos, M. J.; de Jonge, J. C.; Sikkema, A. H.; Gabius, H. J.; Hoekstra, D.; Baron, W. Galectin-4, a Novel Neuronal Regulator of Myelination. *Glia* **2012**, *60* (6), 919–935. <https://doi.org/10.1002/glia.22324>.
- (79) Velasco, S.; Díez-Revuelta, N.; Hernández-Iglesias, T.; Kaltner, H.; André, S.; Gabius, H. J.; Abad-Rodríguez, J. Neuronal Galectin-4 Is Required for Axon Growth and for the Organization of Axonal Membrane L1 Delivery and Clustering. *J. Neurochem.* **2013**, *125* (1), 49–62. <https://doi.org/10.1111/jnc.12148>.
- (80) Díez-Revuelta, N.; Higuero, A. M.; Velasco, S.; Peñas-De-La-Iglesia, M.; Gabius, H. J.; Abad-Rodríguez, J. Neurons Define Non-Myelinated Axon Segments by the Regulation of Galectin-4-Containing Axon Membrane Domains. *Sci. Rep.* **2017**, *7* (1), 1–13. <https://doi.org/10.1038/s41598-017-12295-6>.
- (81) Paclik, D.; Lohse, K.; Wiedenmann, B.; Dignass, A. U.; Sturm, A. Galectin-2 and -4, but Not Galectin-1, Promote Intestinal Epithelial Wound Healing in Vitro through a TGF-Beta-Independent Mechanism. *Inflamm. Bowel Dis.* **2008**, *14* (10), 1366–1372. <https://doi.org/10.1002/ibd.20499>.
- (82) Kamiguchi, H.; Lemmon, V. A Neuronal Form of the Cell Adhesion Molecule L1 Contains a Tyrosine- Based Signal Required for Sorting to the Axonal Growth Cone. *J. Neurosci.* **1998**, *18* (10), 3749–3756. <https://doi.org/10.1523/jneurosci.18-10-03749.1998>.
- (83) Cao, Z. Q.; Guo, X. L. The Role of Galectin-4 in Physiology and Diseases. *Protein Cell* **2016**, *7* (5), 314–324. <https://doi.org/10.1007/s13238-016-0262-9>.
- (84) Rechreche, H.; Mallo, G. V.; Montalto, G.; Dagorn, J. C.; Iovanna, J. L. Cloning and Expression of the MRNA of Human Galectin-4, an S-Type Lectin down-Regulated in Colorectal Cancer. *Eur. J. Biochem.* **1997**, *248* (1), 225–230. <https://doi.org/10.1111/j.1432-1033.1997.00225.x>.
- (85) El Leithy, A. A.; Helwa, R.; Assem, M. M.; Hassan, N. H. A. Expression Profiling of Cancer-Related Galectins in Acute Myeloid Leukemia. *Tumor Biol.* **2015**, *36* (10), 7929–7939. <https://doi.org/10.1007/s13277-015-3513-0>.

- (86) Makoto, W.; Ichiro, T.; Naoki, K.; Yuhki, Y.; Matsuo, E. I.; Susumu, I.; Masaki, M.; Nariaki, M.; Morito, M.; Osamu, N. Clinical Significance of Circulating Galectins as Colorectal Cancer Markers. *Oncol. Rep.* **2011**, *25* (5), 1217–1226. <https://doi.org/10.3892/or.2011.1198>.
- (87) Barrow, H.; Guo, X.; Wandall, H. H.; Pedersen, J. W.; Fu, B.; Zhao, Q.; Chen, C.; Rhodes, J. M.; Yu, L. G. Serum Galectin-2, -4, and -8 Are Greatly Increased in Colon and Breast Cancer Patients and Promote Cancer Cell Adhesion to Blood Vascular Endothelium. *Clin. Cancer Res.* **2011**, *17* (22), 7035–7046. <https://doi.org/10.1158/1078-0432.CCR-11-1462>.
- (88) Hu, D.; Ansari, D.; Zhou, Q.; Sasor, A.; Said Hilmersson, K.; Andersson, R. Galectin 4 Is a Biomarker for Early Recurrence and Death after Surgical Resection for Pancreatic Ductal Adenocarcinoma. *Scand. J. Gastroenterol.* **2019**, *54* (1), 95–100. <https://doi.org/10.1080/00365521.2018.1561937>.
- (89) Helwa, R.; Heller, A.; Knappskog, S.; Bauer, A. S. Tumor Cells Interact with Red Blood Cells via Galectin-4 - a Short Report. *Cell. Oncol.* **2017**, *40* (4), 401–409. <https://doi.org/10.1007/s13402-017-0317-9>.
- (90) Stowell, S. R.; Arthur, C. M.; Dias-Baruffi, M.; Rodrigues, L. C.; Gourdine, J. P.; Heimburg-Molinaro, J.; Ju, T.; Molinaro, R. J.; Rivera-Marrero, C.; Xia, B.; Smith, D. F.; Cummings, R. D. Innate Immune Lectins Kill Bacteria Expressing Blood Group Antigen. *Nat. Med.* **2010**, *16* (3), 295–301. <https://doi.org/10.1038/nm.2103>.
- (91) Arthur, C. M.; Cummings, R. D.; Stowell, S. R. Evaluation of the Bactericidal Activity of Galectins. *Methods Mol. Biol.* **2015**, *1207*, 421–430. https://doi.org/10.1007/978-1-4939-1396-1_27.
- (92) Croinin, T. O.; Clyne, M.; Drumm, B. Molecular Mimicry of Ferret Gastric Epithelial Blood Group Antigen A by *Helicobacter Mustelae*. *Gastroenterology* **1998**, *114* (4), 690–696. [https://doi.org/10.1016/S0016-5085\(98\)70582-7](https://doi.org/10.1016/S0016-5085(98)70582-7).
- (93) Blenda, A. V.; Kamili, N. A.; Wu, S.-C.; Abel, W. F.; Ayona, D.; Gerner-Smidt, C.; Ho, A. D.; Benian, G. M.; Cummings, R. D.; Arthur, C. M.; Stowell, S. R. Galectin-9 Recognizes and Exhibits Antimicrobial Activity toward Microbes Expressing Blood Group-like Antigens. *J. Biol. Chem.* **2022**, *298* (4), 101704. <https://doi.org/10.1016/j.jbc.2022.101704>.
- (94) Hong, S. H.; Shin, J. S.; Chung, H.; Park, C. G. Galectin-4 Interaction with CD14 Triggers the Differentiation of Monocytes into Macrophage-like Cells via the MAPK Signaling Pathway. *Immune Netw.* **2019**, *19* (3).

<https://doi.org/10.4110/in.2019.19.e17>.

- (95) Kitchens, R. L. Role of CD14 in Cellular Recognition of Bacterial Lipopolysaccharides. In *CD14 in the Inflammatory Response*; KARGER: Basel, 1999; Vol. 74, pp 61–82. <https://doi.org/10.1159/000058750>.
- (96) Rondeau, J. M.; Schreuder, H. *Protein Crystallography and Drug Discovery*; Elsevier Ltd, 2015. <https://doi.org/10.1016/B978-0-12-417205-0.00022-5>.
- (97) Lütteke, T.; Frank, M.; Von Der Lieth, C. W. Data Mining the Protein Data Bank: Automatic Detection and Assignment of Carbohydrate Structures. *Carbohydr. Res.* **2004**, 339 (5), 1015–1020. <https://doi.org/10.1016/j.carres.2003.09.038>.
- (98) Agirre, J. Strategies for Carbohydrate Model Building, Refinement and Validation. *Acta Crystallogr. Sect. D Struct. Biol.* **2017**, 73 (2), 171–186. <https://doi.org/10.1107/S2059798316016910>.
- (99) Lee, J. H.; Ozorowski, G.; Ward, A. B. Cryo-EM Structure of a Native, Fully Glycosylated, Cleaved HIV-1 Envelope Trimer. *Science (80-.)*. **2016**, 351 (6277), 1043–1048. <https://doi.org/10.1126/science.aad2450>.
- (100) Atanasova, M.; Bagdonas, H.; Agirre, J. Structural Glycobiology in the Age of Electron Cryo-Microscopy. *Curr. Opin. Struct. Biol.* **2020**, 62, 70–78. <https://doi.org/10.1016/j.sbi.2019.12.003>.
- (101) Penezic, A.; Deokar, G.; Vignaud, D.; Pichonat, E.; Happy, H.; Subramanian, P.; Gasparović, B.; Boukherroub, R.; Szunerits, S. Carbohydrate-Lectin Interaction on Graphene-Coated Surface Plasmon Resonance (SPR) Interfaces. *Plasmonics* **2014**, 9 (3), 677–683. <https://doi.org/10.1007/s11468-014-9686-3>.
- (102) Ji, Y.; Woods, R. J. Quantifying Weak Glycan-Protein Interactions Using a Biolayer Interferometry Competition Assay: Applications to ECL Lectin and X-31 Influenza Hemagglutinin. *Adv. Exp. Med. Biol.* **2018**, 1104, 259–273. https://doi.org/10.1007/978-981-13-2158-0_13.
- (103) Murthy, B. N.; Sinha, S.; Surolia, A.; Indi, S. S.; Jayaraman, N. SPR and ITC Determination of the Kinetics and the Thermodynamics of Bivalent versus Monovalent Sugar Ligand-Lectin Interactions. *Glycoconj. J.* **2008**, 25 (4), 313–321. <https://doi.org/10.1007/s10719-007-9076-6>.
- (104) Ardá, A.; Jiménez-Barbero, J. The Recognition of Glycans by Protein Receptors. Insights from NMR Spectroscopy. *Chem. Commun.* **2018**, 54 (38), 4761–4769. <https://doi.org/10.1039/c8cc01444b>.

- (105) Bubb, W. A. NMR Spectroscopy in the Study of Carbohydrates: Characterizing the Structural Complexity. *Concepts Magn. Reson. Part A Bridg. Educ. Res.* **2003**, *19* (1), 1–19. <https://doi.org/10.1002/cmr.a.10080>.
- (106) Marchetti, R.; Perez, S.; Arda, A.; Imberty, A.; Jimenez-Barbero, J.; Silipo, A.; Molinaro, A. “Rules of Engagement” of Protein–Glycoconjugate Interactions: A Molecular View Achievable by Using NMR Spectroscopy and Molecular Modeling. *ChemistryOpen* **2016**, *5* (4), 274–296. <https://doi.org/10.1002/open.201600024>.
- (107) Macomber, R. S.; Harbison, G. S. A Complete Introduction to Modern NMR Spectroscopy. *Physics Today*. 1999, pp 68–68. <https://doi.org/10.1063/1.882558>.
- (108) Rabi, I. I.; Zacharias, J. R.; Millman, S.; Kusch, P. A New Method of Measuring Nuclear Magnetic Moment. *Phys. Rev.* **1938**, *53* (4), 318–318. <https://doi.org/10.1103/PhysRev.53.318>.
- (109) Keeler, J. Understanding NMR Spectroscopy, 2nd Edition | Wiley. Wiley **2010**, 190–200.
- (110) Hahn, E. L.; Maxwell, D. E. Spin Echo Measurements of Nuclear Spin Coupling in Molecules. *Phys. Rev.* **1952**, *88* (5), 1070–1084. <https://doi.org/10.1103/PhysRev.88.1070>.
- (111) Karplus, M. Contact Electron-Spin Coupling of Nuclear Magnetic Moments. *J. Chem. Phys.* **1959**, *30* (1), 11–15. <https://doi.org/10.1063/1.1729860>.
- (112) Solomon, I. Relaxation Processes in a System of Two Spins. *Physical Review*. 1955, pp 559–565. <https://doi.org/10.1103/PhysRev.99.559>.
- (113) Anet, F. A. L.; Bourn, A. J. R. Nuclear Magnetic Resonance Spectral Assignments from Nuclear Overhauser Effects 1. *J. Am. Chem. Soc.* **1965**, *87* (22), 5250–5251. <https://doi.org/10.1021/ja00950a048>.
- (114) Neuhaus, D.; Williams, M. *The Nuclear Overhauser Effect in Structural and Conformational Analysis*; 1989; Vol. 97. [https://doi.org/10.1016/0022-2364\(92\)90256-7](https://doi.org/10.1016/0022-2364(92)90256-7).
- (115) Kaiser, R. Use of the Nuclear Overhauser Effect in the Analysis of High-Resolution Nuclear Magnetic Resonance Spectra. *J. Chem. Phys.* **1963**, *39* (10), 2435–2442. <https://doi.org/10.1063/1.1734045>.
- (116) Atxabal, U.; Gimeno, A.; Jiménez-Barbero, J. Nuclear Magnetic Resonance Techniques for the Study of Glycan Interactions. In *Comprehensive Glycoscience: Second Edition*; 2021; pp 329–345.

<https://doi.org/10.1016/B978-0-12-819475-1.00015-8>.

- (117) Unione, L.; Galante, S.; Díaz, D.; Cañada, F. J.; Jiménez-Barbero, J. NMR and Molecular Recognition. the Application of Ligand-Based NMR Methods to Monitor Molecular Interactions. *Medchemcomm* **2014**, *5* (9), 1280–1289. <https://doi.org/10.1039/c4md00138a>.
- (118) Ni, F.; Scheraga, H. A. Use of the Transferred Nuclear Overhauser Effect To Determine the Conformations of Ligands Bound to Proteins. *Acc. Chem. Res.* **1994**, *27* (9), 257–264. <https://doi.org/10.1021/ar00045a001>.
- (119) Peters, T. Transfer NOE Experiments for the Study of Carbohydrate-Protein Interactions. In *Carbohydrates in Chemistry and Biology*; Wiley-VCH Verlag GmbH: Weinheim, Germany, 2008; Vol. 2–4, pp 1003–1023. <https://doi.org/10.1002/9783527618255.ch35>.
- (120) Weimar, T.; Peters, T. Aleuria Aurantia Agglutinin Recognizes Multiple Conformations of A-L-Fuc-(1→6)-β-D-GlcNAc-OMe. *Angew. Chemie Int. Ed. English* **1994**, *33* (1), 88–91. <https://doi.org/10.1002/anie.199400881>.
- (121) Raghothama, S. REVIEWS NMR of Peptides. *J. Indian Inst. Sci.* **2010**, *90* (1), 90.
- (122) Nieto, P. M. The Use of NMR to Study Transient Carbohydrate-Protein Interactions. *Front. Mol. Biosci.* **2018**, *5* (APR), 1–7. <https://doi.org/10.3389/fmolb.2018.00033>.
- (123) Jayalakshmi, V.; Krishna, N. R. Complete Relaxation and Conformational Exchange Matrix (CORCEMA) Analysis of Intermolecular Saturation Transfer Effects in Reversibly Forming Ligand-Receptor Complexes. *J. Magn. Reson.* **2002**, *155* (1), 106–118. <https://doi.org/10.1006/jmre.2001.2499>.
- (124) Jeannerat, D.; Furrer, J. NMR Experiments for the Analysis of Mixtures: Beyond 1D 1H Spectra. *Comb. Chem. High Throughput Screen.* **2011**, *15* (1), 15–35. <https://doi.org/10.2174/138620712798280853>.
- (125) Asensio, J. L.; Cañada, F. J.; Bruix, M.; Rodriguez-Romero, A.; Jimenez-Barbero, J. The Interaction of Hevein with N-acetylglucosamine-containing Oligosaccharides: Solution Structure of Hevein Complexed to Chitobiose. *Eur. J. Biochem.* **1995**, *230* (2), 621–633. <https://doi.org/10.1111/j.1432-1033.1995.0621h.x>.
- (126) Valverde, P.; Delgado, S.; Martínez, J. D.; Vendeville, J. B.; Malassis, J.; Linclau, B.; Reichardt, N. C.; Cañada, F. J.; Jiménez-Barbero, J.; Ardá, A. Molecular Insights into DC-SIGN Binding to Self-Antigens: The Interaction with the Blood Group A/B Antigens. *ACS Chem. Biol.*

- 2019**, *14* (7), 1660–1671. <https://doi.org/10.1021/acscchembio.9b00458>.
- (127) del Carmen Fernandez-Alonso, M.; Diaz, D.; Alvaro Berbis, M.; Marcelo, F.; Canada, J.; Jimenez-Barbero, J. Protein-Carbohydrate Interactions Studied by NMR: From Molecular Recognition to Drug Design. *Curr. Protein Pept. Sci.* **2013**, *13* (8), 816–830. <https://doi.org/10.2174/138920312804871175>.
- (128) Morris, K. F.; Johnson, C. S. Diffusion-Ordered Two-Dimensional Nuclear Magnetic Resonance Spectroscopy. *J. Am. Chem. Soc.* **1992**, *114* (8), 3139–3141. <https://doi.org/10.1021/ja00034a071>.
- (129) Morris, K. F.; Johnson, C. S. Mobility Ordered 2D-NMR Spectroscopy. *J. Am. Chem. Soc.* **1992**, *114* (2), 776–777. <https://doi.org/10.1021/ja00028a063>.
- (130) Viegas, A.; Manso, J.; Nobrega, F. L.; Cabrita, E. J. Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *J. Chem. Educ.* **2011**, *88* (7), 990–994. <https://doi.org/10.1021/ed101169t>.
- (131) Monaco, S.; Tailford, L. E.; Juge, N.; Angulo, J. Differential Epitope Mapping by STD NMR Spectroscopy To Reveal the Nature of Protein–Ligand Contacts. *Angew. Chemie - Int. Ed.* **2017**, *56* (48), 15289–15293. <https://doi.org/10.1002/anie.201707682>.
- (132) Moure, M. J.; Gimeno, A.; Delgado, S.; Diercks, T.; Boons, G.; Jiménez-Barbero, J.; Ardá, A. Selective ¹³C-Labels on Repeating Glycan Oligomers to Reveal Protein Binding Epitopes through NMR: Polylactosamine Binding to Galectins. *Angew. Chemie* **2021**, *133* (34), 18925–18930. <https://doi.org/10.1002/ange.202106056>.
- (133) Martínez, J. D.; Infantino, A. S.; Valverde, P.; Diercks, T.; Delgado, S.; Reichardt, N. C.; Ardá, A.; Cañada, F. J.; Oscarson, S.; Jiménez-Barbero, J. The Interaction of Fluorinated Glycomimetics with Dc-Sign: Multiple Binding Modes Disentangled by the Combination of NMR Methods and MD Simulations. *Pharmaceuticals* **2020**, *13* (8), 1–18. <https://doi.org/10.3390/ph13080179>.
- (134) Bodenhausen, G.; Ruben, D. J. Natural Abundance Nitrogen-15 NMR by Enhanced Heteronuclear Spectroscopy. *Chem. Phys. Lett.* **1980**, *69* (1), 185–189. [https://doi.org/10.1016/0009-2614\(80\)80041-8](https://doi.org/10.1016/0009-2614(80)80041-8).
- (135) Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. Attenuated T2 Relaxation by Mutual Cancellation of Dipole-Dipole Coupling and Chemical Shift Anisotropy Indicates an Avenue to NMR Structures of Very Large Biological Macromolecules in Solution. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (23), 12366–12371.

<https://doi.org/10.1073/pnas.94.23.12366>.

- (136) Gronenborn, A. M.; Clore, G. M. Three-Dimensional Structures of Proteins in Solution by Nuclear Magnetic Resonance Spectroscopy. *Protein Seq. Data Anal.* **1989**, *2* (1), 1–8.
- (137) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering High-Affinity Ligands for Proteins_SAR by NMR (First Paper about FBS NMR). *Science (80-.)*. **1996**, *274*, 1531–1534.
- (138) Williamson, M. P. Using Chemical Shift Perturbation to Characterise Ligand Binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1–16. <https://doi.org/10.1016/j.pnmrs.2013.02.001>.
- (139) Aretz, J.; Baukman, H.; Shanina, E.; Hanske, J.; Wawrzinek, R.; Zapol'skii, V. A.; Seeberger, P. H.; Kaufmann, D. E.; Rademacher, C. Identification of Multiple Druggable Secondary Sites by Fragment Screening against DC-SIGN. *Angew. Chemie - Int. Ed.* **2017**, *56* (25), 7292–7296. <https://doi.org/10.1002/anie.201701943>.
- (140) Kay, L. E.; Ikura, M.; Tschudin, R.; Bax, A. Three-Dimensional Triple-Resonance NMR Spectroscopy of Isotopically Enriched Proteins. *J. Magn. Reson.* **1990**, *89*, 496–514.
- (141) Clubb, R. T.; Thanabal, V.; Wagner, G. A Constant-Time Three-Dimensional Triple-Resonance Pulse Scheme to Intrareidue ¹HN, ¹⁵N, and ¹³C' Chemical Shifts in ¹⁵N-¹³C-Labeled Proteins. *J. Magn. Reson.* **1992**, *97*, 213–217.
- (142) Bax, A.; Ikura, M. An Efficient 3D NMR Technique for Correlating the Proton and ¹⁵N Backbone Amide Resonances with the α -Carbon of the Preceding Residue in Uniformly ¹⁵N/¹³C Enriched Proteins. *J. Biomol. NMR* **1991**, *1* (1), 99–104. <https://doi.org/10.1007/BF01874573>.
- (143) Grzesiek, S.; Bax, A. An Efficient Experiment for Sequential Backbone Assignment of Medium-Sized Isotopically Enriched Proteins. *J. Magn. Reson.* **1992**, *99* (1), 201–207. [https://doi.org/10.1016/0022-2364\(92\)90169-8](https://doi.org/10.1016/0022-2364(92)90169-8).
- (144) Grzesiek, S.; Bax, A. Correlating Backbone Amide and Side Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance NMR. *J. Am. Chem. Soc.* **1992**, *114* (16), 6291–6293. <https://doi.org/10.1021/ja00042a003>.
- (145) Wittekind, M.; Mueller, L. HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins. *Journal of Magnetic Resonance, Series B.* 1993, pp 201–205.

<https://doi.org/10.1006/jmrb.1993.1033>.

- (146) Yamazaki, T.; Muhandiranv, D. R.; Kay, L. E.; Lee, W.; Arrowsmith, C. H. A Suite of Triple Resonance NMR Experiments for the Backbone Assignment of ^{15}N , ^{13}C , ^2H Labeled Proteins with High Sensitivity. *J. Am. Chem. Soc.* **1994**, *116* (26), 11655–11666. <https://doi.org/10.1021/ja00105a005>.
- (147) Schubert, M.; Labudde, D.; Leitner, D.; Oschkinat, H.; Schmieder, P. A Modified Strategy for Sequence Specific Assignment of Protein NMR Spectra Based on Amino Acid Type Selective Experiments. *J. Biomol. NMR* **2005**, *31* (2), 115–128. <https://doi.org/10.1007/s10858-004-8263-z>.
- (148) Lescop, E.; Rasia, R.; Brutscher, B. Hadamard Amino-Acid-Type Edited NMR Experiment for Fast Protein Resonance Assignment. *J. Am. Chem. Soc.* **2008**, *130* (15), 5014–5015. <https://doi.org/10.1021/ja800914h>.
- (149) Prasanna, C.; Dubey, A.; Atreya, H. S. *Amino Acid Selective Unlabeling in Protein NMR Spectroscopy*, 1st ed.; Elsevier Inc., 2015; Vol. 565. <https://doi.org/10.1016/bs.mie.2015.05.008>.
- (150) Rasia, R. M.; Brutscher, B.; Plevin, M. J. Selective Isotopic Unlabeling of Proteins Using Metabolic Precursors: Application to NMR Assignment of Intrinsically Disordered Proteins. *ChemBioChem* **2012**, *13* (5), 732–739. <https://doi.org/10.1002/cbic.201100678>.
- (151) Hawkes, C. M.; Janata, J. Isothermal Titration Calorimeter. *Anal. Lett.* **1973**, *6* (2), 163–169. <https://doi.org/10.1080/00032717308062192>.
- (152) C., J.; Murciano-Calles, J.; S., E.; Iglesias-Bexiga, M.; Luque, I.; Ruiz-Sanz, J. Isothermal Titration Calorimetry: Thermodynamic Analysis of the Binding Thermograms of Molecular Recognition Events by Using Equilibrium Models. In *Applications of Calorimetry in a Wide Context - Differential Scanning Calorimetry, Isothermal Titration Calorimetry and Microcalorimetry*; Elkordy, A. A., Ed.; InTech, 2013; p 474. <https://doi.org/10.5772/53311>.
- (153) Pierce, M. M.; Raman, C. S.; Nall, B. T. Isothermal Titration Calorimetry of Protein-Protein Interactions. *Methods A Companion to Methods Enzymol.* **1999**, *19* (2), 213–221. <https://doi.org/10.1006/meth.1999.0852>.
- (154) Dam, T. K.; Brewer, C. F. Thermodynamic Studies of Lectin-Carbohydrate Interactions by Isothermal Titration Calorimetry. *Chem. Rev.* **2002**, *102* (2), 387–429. <https://doi.org/10.1021/cr000401x>.
- (155) Turnbull, W. B.; Daranas, A. H. On the Value of c : Can Low Affinity

- Systems Be Studied by Isothermal Titration Calorimetry? *J. Am. Chem. Soc.* **2003**, *125* (48), 14859–14866. <https://doi.org/10.1021/ja036166s>.
- (156) Giovannone, N.; Smith, L. K.; Treanor, B.; Dimitroff, C. J. Galectin-Glycan Interactions as Regulators of B Cell Immunity. *Front. Immunol.* **2018**, *9*. <https://doi.org/10.3389/fimmu.2018.02839>.
- (157) Martínez Allo, V. C.; Toscano, M. A.; Pinto, N.; Rabinovich, G. A. Galectins: Key Players at the Frontiers of Innate and Adaptive Immunity. *Trends Glycosci. Glycotechnol.* **2018**, *30* (172), SE97–SE107. <https://doi.org/10.4052/tigg.1740.1SE>.
- (158) Sato, S.; St-Pierre, C.; Bhaumik, P.; Nieminen, J. Galectins in Innate Immunity: Dual Functions of Host Soluble β -Galactoside-Binding Lectins as Damage-Associated Molecular Patterns (DAMPs) and as Receptors for Pathogen-Associated Molecular Patterns (PAMPs). *Immunol. Rev.* **2009**, *230* (1), 172–187. <https://doi.org/10.1111/j.1600-065X.2009.00790.x>.
- (159) Elola, M. T.; Ferragut, F.; Méndez-Huergo, S. P.; Croci, D. O.; Bracalente, C.; Rabinovich, G. A. Galectins: Multitask Signaling Molecules Linking Fibroblast, Endothelial and Immune Cell Programs in the Tumor Microenvironment. *Cell. Immunol.* **2018**, *333*, 34–45. <https://doi.org/10.1016/j.cellimm.2018.03.008>.
- (160) Huflejt, M. E.; Leffler, H. Galectin-4 in Normal Tissues and Cancer. *Glycoconj. J.* **2003**, *20* (4), 247–255. <https://doi.org/10.1023/B:GLYC.0000025819.54723.a0>.
- (161) Tsai, C. H.; Tzeng, S. F.; Chao, T. K.; Tsai, C. Y.; Yang, Y. C.; Lee, M. T.; Hwang, J. J.; Chou, Y. C.; Tsai, M. H.; Cha, T. L.; Hsiao, P. W. Metastatic Progression of Prostate Cancer Is Mediated by Autonomous Binding of Galectin-4-O-Glycan to Cancer Cells. *Cancer Res.* **2016**, *76* (19), 5756–5767. <https://doi.org/10.1158/0008-5472.CAN-16-0641>.
- (162) Cagnoni, A. J.; Pérez Sáez, J. M.; Rabinovich, G. A.; Mariño, K. V. Turning-off Signaling by Siglecs, Selectins, and Galectins: Chemical Inhibition of Glycan-Dependent Interactions in Cancer. *Front. Oncol.* **2016**, *6* (MAY), 1–21. <https://doi.org/10.3389/fonc.2016.00109>.
- (163) Girard, A.; Magnani, J. L. Clinical Trials and Applications of Galectin Antagonists. *Trends Glycosci. Glycotechnol.* **2018**, *30* (172), SE211–SE220. <https://doi.org/10.4052/tigg.1744.1se>.
- (164) Ideo, H.; Seko, A.; Ohkura, T.; Matta, K. L.; Yamashita, K. High-Affinity Binding of Recombinant Human Galectin-4 to SO₃→3Gal β 1→3GalNAc. *Glycobiology* **2002**, *12* (3), 199–208. <https://doi.org/10.1093/glycob/12.3.199>.

- (165) Tanaka, A.; Kimura, A.; Yamamoto, Y.; Uede, K.; Furukawa, F. Expression of Histo-Blood Group A Type 1, 2 and 3 Antigens in Normal Skin and Extramammary Paget's Disease. *ACTA Histochem. Cytochem.* **2008**, *41* (6), 165–171. <https://doi.org/10.1267/ahc.08021>.
- (166) Watkins, W. M.; Greenwell, P.; Yates, A. D.; Johnson, P. H. Regulation of Expression of Carbohydrate Blood Group Antigens. *Biochimie* **1988**, *70* (11), 1597–1611. [https://doi.org/10.1016/0300-9084\(88\)90295-7](https://doi.org/10.1016/0300-9084(88)90295-7).
- (167) Clausen, H.; Hakomori, S. ABH and Related Histo-Blood Group Antigens; Immunochemical Differences in Carrier Isotypes and Their Distribution1. *Vox Sang.* **1989**, *56* (1), 1–20. <https://doi.org/https://doi.org/10.1111/j.1423-0410.1989.tb03040.x>.
- (168) Meloncelli, P. J.; Lowary, T. L. Synthesis of ABO Histo-Blood Group Type I and II Antigens. *Carbohydr. Res.* **2010**, *345* (16), 2305–2322. <https://doi.org/https://doi.org/10.1016/j.carres.2010.08.012>.
- (169) Breimer, M. E.; Hansson, G. C.; Karlsson, K. A.; Leffler, H. Isolation and Partial Characterization of Blood Group A and H Active Glycosphingolipids of Rat Small Intestine. *J. Biol. Chem.* **1982**, *257* (2), 906–912.
- (170) Kamili, N. A.; Arthur, C. M.; Gerner-Smidt, C.; Tafesse, E.; Blenda, A.; Dias-Baruffi, M.; Stowell, S. R. Key Regulators of Galectin–Glycan Interactions. *Proteomics* **2016**, *16* (24), 3111–3125. <https://doi.org/10.1002/pmic.201600116>.
- (171) Arthur, C. M.; Patel, S. R.; Mener, A.; Kamili, N. A.; Fasano, R. M.; Meyer, E.; Winkler, A. M.; Sola-Visner, M.; Josephson, C. D.; Stowell, S. R. Innate Immunity against Molecular Mimicry: Examining Galectin-Mediated Antimicrobial Activity. *BioEssays* **2015**, *37* (12), 1327–1337. <https://doi.org/10.1002/bies.201500055>.
- (172) Stenutz, R.; Weintraub, A.; Widmalm, G. The Structures of Escherichia Coli O-Polysaccharide Antigens. *FEMS Microbiol. Rev.* **2006**, *30* (3), 382–403. <https://doi.org/10.1111/j.1574-6976.2006.00016.x>.
- (173) Su, J.; Wang, Y.; Si, Y.; Gao, J.; Song, C.; Cui, L.; Wu, R.; Tai, G.; Zhou, Y. Galectin-13, a Different Prototype Galectin, Does Not Bind β -Galacto-Sides and Forms Dimers via Intermolecular Disulfide Bridges between Cys-136 and Cys-138. *Sci. Rep.* **2018**, *8* (1), 1–13. <https://doi.org/10.1038/s41598-018-19465-0>.
- (174) Marchetti, R.; Perez, S.; Arda, A.; Imberty, A.; Jimenez-Barbero, J.; Silipo, A.; Molinaro, A. “Rules of Engagement” of Protein–Glycoconjugate Interactions: A Molecular View Achievable by Using NMR Spectroscopy and Molecular Modeling. *ChemistryOpen* **2016**, *5*

- (4), 274–296. <https://doi.org/10.1002/open.201600024>.
- (175) Wagstaff, J. L.; Taylor, S. L.; Howard, M. J. Recent Developments and Applications of Saturation Transfer Difference Nuclear Magnetic Resonance (STD NMR) Spectroscopy. *Mol. BioSyst.* **2013**, *9* (4), 571–577. <https://doi.org/10.1039/C2MB25395J>.
- (176) Bhunia, A.; Bhattacharjya, S.; Chatterjee, S. Applications of Saturation Transfer Difference NMR in Biological Systems. *Drug Discov. Today* **2012**, *17* (9), 505–513. <https://doi.org/https://doi.org/10.1016/j.drudis.2011.12.016>.
- (177) Asensio, J. L.; Cañada, F. J.; Jimenez-Barbero, J. Studies of the Bound Conformations of Methyl Alpha-Lactoside and Methyl Beta-Allolactoside to Ricin B Chain Using Transferred NOE Experiments in the Laboratory and Rotating Frames, Assisted by Molecular Mechanics and Dynamics Calculations. *Eur. J. Biochem.* **1995**, *233* (2), 618–630. https://doi.org/10.1111/j.1432-1033.1995.618_2.x.
- (178) Gimeno, A.; Delgado, S.; Valverde, P.; Bertuzzi, S.; Berbís, M. A.; Echavarren, J.; Lacetera, A.; Martín-Santamaría, S.; Surolia, A.; Cañada, F. J.; Jiménez-Barbero, J.; Ardá, A. Minimizing the Entropy Penalty for Ligand Binding: Lessons from the Molecular Recognition of the Histo Blood-Group Antigens by Human Galectin-3. *Angew. Chemie - Int. Ed.* **2019**, *58* (22), 7268–7272. <https://doi.org/10.1002/anie.201900723>.
- (179) Asensio, J. L.; Ardá, A.; Cañada, F. J.; Jiménez-Barbero, J. Carbohydrate-Aromatic Interactions. *Acc. Chem. Res.* **2013**, *46* (4). <https://doi.org/10.1021/ar300024d>.
- (180) Dings, R. P. M.; Miller, M. C.; Griffin, R. J.; Mayo, K. H. Galectins as Molecular Targets for Therapeutic Intervention. *Int. J. Mol. Sci.* **2018**, *19* (3), 1–22. <https://doi.org/10.3390/ijms19030905>.
- (181) Sindrewicz, P.; Li, X.; Yates, E. A.; Turnbull, J. E.; Lian, L.-Y.; Yu, L.-G. Intrinsic Tryptophan Fluorescence Spectroscopy Reliably Determines Galectin-Ligand Interactions. *Sci. Rep.* **2019**, *9* (1), 11851. <https://doi.org/10.1038/s41598-019-47658-8>.
- (182) Peccati, F.; Jiménez-Osés, G. Enthalpy–Entropy Compensation in Biomolecular Recognition: A Computational Perspective. *ACS Omega* **2021**, *6* (17), 11122–11130. <https://doi.org/10.1021/acsomega.1c00485>.
- (183) Asensio, J. L.; Cañada, F. J.; Jiménez-Barbero, J. Studies of the Bound Conformations of Methyl A-Lactoside and Methyl B-Allolactoside to Ricin B Chain Using Transferred NOE Experiments in the Laboratory and Rotating Frames, Assisted by Molecular Mechanics and Dynamics Calculations. *Eur. J. Biochem.* **1995**, *233* (2), 618–630.

https://doi.org/10.1111/j.1432-1033.1995.618_2.x.

- (184) Quintana, J. I.; Delgado, S.; Núñez-Franco, R.; Cañada, F. J.; Jiménez-Osés, G.; Jiménez-Barbero, J.; Ardá, A. Galectin-4 N-Terminal Domain: Binding Preferences Toward A and B Antigens With Different Peripheral Core Presentations. *Front. Chem.* **2021**, *9* (April). <https://doi.org/10.3389/fchem.2021.664097>.
- (185) Murray, C. J.; Ikuta, K. S.; Sharara, F.; Swetschinski, L.; Robles Aguilar, G.; Gray, A.; Han, C.; Bisignano, C.; Rao, P.; Wool, E.; Johnson, S. C.; Browne, A. J.; Chipeta, M. G.; Fell, F.; Hackett, S.; Haines-Woodhouse, G.; Kashef Hamadani, B. H.; Kumaran, E. A. P.; McManigal, B.; Agarwal, R.; Akech, S.; Albertson, S.; Amuasi, J.; Andrews, J.; Aravkin, A.; Ashley, E.; Bailey, F.; Baker, S.; Basnyat, B.; Bekker, A.; Bender, R.; Bethou, A.; Bielicki, J.; Boonkasidecha, S.; Bukosia, J.; Carneiro, C.; Castañeda-Orjuela, C.; Chansamouth, V.; Chaurasia, S.; Chiurchiù, S.; Chowdhury, F.; Cook, A. J.; Cooper, B.; Cressey, T. R.; Criollo-Mora, E.; Cunningham, M.; Darboe, S.; Day, N. P. J.; De Luca, M.; Dokova, K.; Dramowski, A.; Dunachie, S. J.; Eckmanns, T.; Eibach, D.; Emami, A.; Feasey, N.; Fisher-Pearson, N.; Forrest, K.; Garrett, D.; Gastmeier, P.; Giref, A. Z.; Greer, R. C.; Gupta, V.; Haller, S.; Haselbeck, A.; Hay, S. I.; Holm, M.; Hopkins, S.; Iregbu, K. C.; Jacobs, J.; Jarovsky, D.; Javanmardi, F.; Khorana, M.; Kisson, N.; Kobeissi, E.; Kostyanov, T.; Krapp, F.; Krumkamp, R.; Kumar, A.; Kyu, H. H.; Lim, C.; Limmathurotsakul, D.; Loftus, M. J.; Lunn, M.; Ma, J.; Mturi, N.; Munera-Huertas, T.; Musicha, P.; Mussi-Pinhata, M. M.; Nakamura, T.; Nanavati, R.; Nangia, S.; Newton, P.; Ngoun, C.; Novotney, A.; Nwakanma, D.; Obiero, C. W.; Olivares-Martinez, A.; Olliaro, P.; Ooko, E.; Ortiz-Brizuela, E.; Peleg, A. Y.; Perrone, C.; Plakkal, N.; Ponce-de-Leon, A.; Raad, M.; Ramdin, T.; Riddell, A.; Roberts, T.; Robotham, J. V.; Roca, A.; Rudd, K. E.; Russell, N.; Schnall, J.; Scott, J. A. G.; Shivamallappa, M.; Sifuentes-Osornio, J.; Steenkeste, N.; Stewardson, A. J.; Stoeva, T.; Tasak, N.; Thaiprakong, A.; Thwaites, G.; Turner, C.; Turner, P.; van Doorn, H. R.; Velaphi, S.; Vongpradith, A.; Vu, H.; Walsh, T.; Waner, S.; Wangrangsimakul, T.; Wozniak, T.; Zheng, P.; Sartorius, B.; Lopez, A. D.; Stergachis, A.; Moore, C.; Dolecek, C.; Naghavi, M. Global Burden of Bacterial Antimicrobial Resistance in 2019: A Systematic Analysis. *Lancet* **2022**, *399* (10325), 629–655. [https://doi.org/10.1016/S0140-6736\(21\)02724-0](https://doi.org/10.1016/S0140-6736(21)02724-0).
- (186) Delcour, A. H. Outer Membrane Permeability and Antibiotic Resistance. *Biochim. Biophys. Acta - Proteins Proteomics* **2009**, *1794* (5), 808–816. <https://doi.org/10.1016/j.bbapap.2008.11.005>.
- (187) Silhavy, T. J.; Kahne, D.; Walker, S. The Bacterial Cell Envelope. *Cold Spring Harb. Perspect. Biol.* **2010**, *2* (5), a000414–a000414.

<https://doi.org/10.1101/cshperspect.a000414>.

- (188) Raetz, C. R. H.; Whitfield, C. Lipopolysaccharide Endotoxins. *Annu. Rev. Biochem.* **2002**, *71* (1), 635–700. <https://doi.org/10.1146/annurev.biochem.71.110601.135414>.
- (189) Bertani, B.; Ruiz, N. Function and Biogenesis of Lipopolysaccharides. *EcoSal Plus* **2018**, *8* (1), 1–33. <https://doi.org/10.1128/ecosalplus.esp-0001-2018>.
- (190) Gioannini, T. L.; Teghanemt, A.; Zhang, D. S.; Coussens, N. P.; Dockstader, W.; Ramaswamy, S.; Weiss, J. P. Isolation of an Endotoxin-MD-2 Complex That Produces Toll-like Receptor 4-Dependent Cell Activation at Picomolar Concentrations. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (12), 4186–4191. <https://doi.org/10.1073/pnas.0306906101>.
- (191) Zamyatina, A.; Heine, H. Lipopolysaccharide Recognition in the Crossroads of TLR4 and Caspase-4/11 Mediated Inflammatory Pathways. *Front. Immunol.* **2020**, *11* (November), 1–22. <https://doi.org/10.3389/fimmu.2020.585146>.
- (192) Miyake, K. Innate Recognition of Lipopolysaccharide by Toll-like Receptor 4-MD-2. *Trends Microbiol.* **2004**, *12* (4), 186–192. <https://doi.org/10.1016/j.tim.2004.02.009>.
- (193) Takeda, K.; Akira, S. Toll-like Receptors in Innate Immunity. *Int. Immunol.* **2005**, *17* (1), 1–14. <https://doi.org/10.1093/intimm/dxh186>.
- (194) Kawasaki, T.; Kawai, T. Toll-like Receptor Signaling Pathways. *Front. Immunol.* **2014**, *5* (SEP), 1–8. <https://doi.org/10.3389/fimmu.2014.00461>.
- (195) Huber, R. G.; Berglund, N. A.; Kargas, V.; Marzinek, J. K.; Holdbrook, D. A.; Khalid, S.; Piggot, T. J.; Schmidtchen, A.; Bond, P. J. A Thermodynamic Funnel Drives Bacterial Lipopolysaccharide Transfer in the TLR4 Pathway. *Structure* **2018**, *26* (8), 1151–1161.e4. <https://doi.org/10.1016/j.str.2018.04.007>.
- (196) Zanoni, I.; Ostuni, R.; Marek, L. R.; Barresi, S.; Barbalat, R.; Barton, G. M.; Granucci, F.; Kagan, J. C. CD14 Controls the LPS-Induced Endocytosis of Toll-like Receptor 4. *Cell* **2011**, *147* (4), 868–880. <https://doi.org/10.1016/j.cell.2011.09.051>.
- (197) Ryu, J. K.; Kim, S. J.; Rah, S. H.; Kang, J. I.; Jung, H. E.; Lee, D.; Lee, H. K.; Lee, J. O.; Park, B. S.; Yoon, T. Y.; Kim, H. M. Reconstruction of LPS Transfer Cascade Reveals Structural Determinants within LBP, CD14, and TLR4-MD2 for Efficient LPS Recognition and Transfer. *Immunity* **2017**, *46* (1), 38–50.

<https://doi.org/10.1016/j.immuni.2016.11.007>.

- (198) Niu, J.; Liu, X.; Zhang, Z.; Huang, Y.; Tang, J.; Wang, B.; Lu, Y.; Cai, J.; Jian, J. A Tandem-Repeat Galectin-4 from Nile Tilapia (*Oreochromis Niloticus*) Is Involved in Immune Response to Bacterial Infection via Mediating Pathogen Recognition and Opsonization. *Mol. Immunol.* **2020**, *127* (July), 67–77. <https://doi.org/10.1016/j.molimm.2020.08.022>.
- (199) Arasu, A.; Kumaresan, V.; Ganesh, M. R.; Pasupuleti, M.; Arasu, M. V.; Al-Dhabi, N. A.; Arockiaraj, J. Bactericidal Activity of Fish Galectin 4 Derived Membrane-Binding Peptide Tagged with Oligotryptophan. *Dev. Comp. Immunol.* **2017**, *71*, 37–48. <https://doi.org/10.1016/j.dci.2017.01.019>.
- (200) Meng, J.; Parroche, P.; Golenbock, D. T.; McKnight, C. J. The Differential Impact of Disulfide Bonds and N-Linked Glycosylation on the Stability and Function of CD14. *J. Biol. Chem.* **2008**, *283* (6), 3376–3384. <https://doi.org/10.1074/jbc.M707640200>.
- (201) Kim, J. I.; Chang, J. L.; Mi, S. J.; Lee, C. H.; Paik, S. G.; Lee, H.; Lee, J. O. Crystal Structure of CD14 and Its Implications for Lipopolysaccharide Signaling. *J. Biol. Chem.* **2005**, *280* (12), 11347–11351. <https://doi.org/10.1074/jbc.M414607200>.
- (202) Kelley, S. L.; Lukk, T.; Nair, S. K.; Tapping, R. I. The Crystal Structure of Human Soluble CD14 Reveals a Bent Solenoid with a Hydrophobic Amino-Terminal Pocket. *J. Immunol.* **2013**, *190* (3), 1304–1311. <https://doi.org/10.4049/jimmunol.1202446>.
- (203) Tarentino, A. L.; Plummer, T. H. [4] Enzymatic Deglycosylation of Asparagine-Linked Glycans: Purification, Properties, and Specificity of Oligosaccharide-Cleaving Enzymes from *Flavobacterium Meningosepticum*. In *Methods in Enzymology*; 1994; pp 44–57. [https://doi.org/10.1016/0076-6879\(94\)30006-2](https://doi.org/10.1016/0076-6879(94)30006-2).
- (204) Liu, T.; Qian, W.-J.; Gritsenko, M. A.; Camp, D. G.; Monroe, M. E.; Moore, R. J.; Smith, R. D. Human Plasma N -Glycoproteome Analysis by Immunoaffinity Subtraction, Hydrazide Chemistry, and Mass Spectrometry. *J. Proteome Res.* **2005**, *4* (6), 2070–2080. <https://doi.org/10.1021/pr0502065>.
- (205) Majerle, A.; Kidrič, J.; Jerala, R. Expression and Refolding of Functional Fragments of the Human Lipopolysaccharide Receptor CD14 in *Escherichia Coli* and *Pichia Pastoris*. *Protein Expr. Purif.* **1999**, *17* (1), 96–104. <https://doi.org/10.1006/prep.1999.1109>.
- (206) Da Silva, T. A.; Zorzetto-Fernandes, A. L. V.; Cecílio, N. T.; Sardinha-Silva, A.; Fernandes, F. F.; Roque-Barreira, M. C. CD14 Is Critical for

TLR2-Mediated M1 Macrophage Activation Triggered by N-Glycan Recognition. *Sci. Rep.* **2017**, *7* (1), 1–14. <https://doi.org/10.1038/s41598-017-07397-0>.

- (207) Nakayama, K.; Wakamatsu, K.; Fujii, H.; Shinzaki, S.; Takamatsu, S.; Kitazume, S.; Kamada, Y.; Takehara, T.; Taniguchi, N.; Miyoshi, E. Core Fucose Is Essential Glycosylation for CD14-Dependent Toll-like Receptor 4 and Toll-like Receptor 2 Signalling in Macrophages. *J. Biochem.* **2019**, *165* (3), 227–237. <https://doi.org/10.1093/jb/mvy098>.
- (208) Iijima, J.; Kobayashi, S.; Kitazume, S.; Kizuka, Y.; Fujinawa, R.; Korekane, H.; Shibata, T.; Saitoh, S.-I.; Akashi-Takamura, S.; Miyake, K.; Miyoshi, E.; Taniguchi, N. Core Fucose Is Critical for CD14-Dependent Toll-like Receptor 4 Signaling. *Glycobiology* **2017**, *27* (11), 1006–1015. <https://doi.org/10.1093/glycob/cwx075>.
- (209) Baumann, C. L.; Aspalter, I. M.; Sharif, O.; Pichlmair, A.; Blüml, S.; Grebien, F.; Bruckner, M.; Pasierbek, P.; Aumayr, K.; Planyavsky, M.; Bennett, K. L.; Colinge, J.; Knapp, S.; Superti-Furga, G. CD14 Is a Coreceptor of Toll-like Receptors 7 and 9. *J. Exp. Med.* **2010**, *207* (12), 2689–2701. <https://doi.org/10.1084/jem.20101111>.
- (210) Adams, S. Toll-like Receptor Agonists in Cancer Therapy. *Immunotherapy* **2009**, *1* (6), 949–964. <https://doi.org/10.2217/imt.09.70>.
- (211) Farooq, M.; Batool, M.; Kim, M. S.; Choi, S. Toll-Like Receptors as a Therapeutic Target in the Era of Immunotherapies. *Front. Cell Dev. Biol.* **2021**, *9* (October), 1–15. <https://doi.org/10.1016/j.cell.2011.09.051>.
- (212) Martinsen, J. T.; Gunst, J. D.; Højen, J. F.; Tolstrup, M.; Søgaaard, O. S. The Use of Toll-Like Receptor Agonists in HIV-1 Cure Strategies. *Front. Immunol.* **2020**, *11* (June), 1–13. <https://doi.org/10.3389/fimmu.2020.01112>.
- (213) Zariri, A.; van der Ley, P. Biosynthetically Engineered Lipopolysaccharide as Vaccine Adjuvant. *Expert Rev. Vaccines* **2015**, *14* (6), 861–876. <https://doi.org/10.1586/14760584.2015.1026808>.
- (214) Unione, L.; Lenza, M. P.; Ardá, A.; Urquiza, P.; Laín, A.; Falcón-Pérez, J. M.; Jiménez-Barbero, J.; Millet, O. Glycoprofile Analysis of an Intact Glycoprotein As Inferred by NMR Spectroscopy. *ACS Cent. Sci.* **2019**, *5* (9), 1554–1561. <https://doi.org/10.1021/acscentsci.9b00540>.
- (215) Lenza, M. P.; Oyenarte, I.; Diercks, T.; Quintana, J. I.; Gimeno, A.; Coelho, H.; Diniz, A.; Peccati, F.; Delgado, S.; Bosch, A.; Valle, M.; Millet, O.; Abrescia, N. G. A.; Palazón, A.; Marcelo, F.; Jiménez-Osés, G.; Jiménez-Barbero, J.; Ardá, A.; Ereño-Orbea, J. Structural Characterization of N-Linked Glycans in the Receptor Binding Domain

- of the SARS-CoV-2 Spike Protein and Their Interactions with Human Lectins. *Angew. Chemie Int. Ed.* **2020**, *59* (52), 23763–23771. <https://doi.org/10.1002/anie.202011015>.
- (216) Hargett, A. A.; Marcella, A. M.; Yu, H.; Li, C.; Orwenyo, J.; Battistel, M. D.; Wang, L.-X.; Freedberg, D. I. Glycosylation States on Intact Proteins Determined by NMR Spectroscopy. *Molecules* **2021**, *26* (14), 4308. <https://doi.org/10.3390/molecules26144308>.
- (217) Cao, L.; Diedrich, J. K.; Ma, Y.; Wang, N.; Pauthner, M.; Park, S.-K. R.; Delahunty, C. M.; McLellan, J. S.; Burton, D. R.; Yates, J. R.; Paulson, J. C. Global Site-Specific Analysis of Glycoprotein N-Glycan Processing. *Nat. Protoc.* **2018**, *13* (6), 1196–1212. <https://doi.org/10.1038/nprot.2018.024>.
- (218) Shields, R. L.; Lai, J.; Keck, R.; O’Connell, L. Y.; Hong, K.; Meng, Y. G.; Weikert, S. H. A.; Presta, L. G. Lack of Fucose on Human IgG1 N-Linked Oligosaccharide Improves Binding to Human FcγRIII and Antibody-Dependent Cellular Toxicity. *J. Biol. Chem.* **2002**, *277* (30), 26733–26740. <https://doi.org/10.1074/jbc.M202069200>.
- (219) Trimble, R. B.; Tarentino, A. L. Identification of Distinct Endoglycosidase (Endo) Activities in *Flavobacterium meningosepticum*: Endo F1, Endo F2, and Endo F3. Endo F1 and Endo H Hydrolyze Only High Mannose and Hybrid Glycans. *J. Biol. Chem.* **1991**, *266* (3), 1646–1651. [https://doi.org/10.1016/S0021-9258\(18\)52343-7](https://doi.org/10.1016/S0021-9258(18)52343-7).
- (220) Roth, J.; Zuber, C.; Park, S.; Jang, I.; Lee, Y.; Kysela, K. G.; Fourn, V.; Santimaria, R.; Guhl, B.; Cho, J. W. Protein N-Glycosylation, Protein Folding, and Protein Quality Control. *Mol. Cells* **2010**, *30* (6), 497–506. <https://doi.org/10.1007/s10059-010-0159-z>.
- (221) Ideo, H.; Seko, A.; Ohkura, T.; Matta, K. L.; Yamashita, K. High-Affinity Binding of Recombinant Human Galectin-4 to SO₃-->3Gal 1->3GalNAc Pyranoside. *Glycobiology* **2002**, *12* (3), 199–208. <https://doi.org/10.1093/glycob/12.3.199>.
- (222) Vokhmyanina, O. A.; Rapoport, E. M.; André, S.; Severov, V. V.; Ryzhov, I.; Pazynina, G. V.; Korchagina, E.; Gabius, H.-J.; Bovin, N. V. Comparative Study of the Glycan Specificities of Cell-Bound Human Tandem-Repeat-Type Galectin-4, -8 and -9. *Glycobiology* **2012**, *22* (9), 1207–1217. <https://doi.org/10.1093/glycob/cws079>.
- (223) Ideo, H.; Seko, A.; Yamashita, K. Galectin-4 Binds to Sulfated Glycosphingolipids and Carcinoembryonic Antigen in Patches on the Cell Surface of Human Colon Adenocarcinoma Cells. *J. Biol. Chem.*

- 2005**, 280 (6), 4730–4737. <https://doi.org/10.1074/jbc.M410362200>.
- (224) Alam, J.; Beyer, N.; Liu, H. W. Biosynthesis of Colitose: Expression, Purification, and Mechanistic Characterization of GDP-4-Keto-6-Deoxy-D-Mannose-3-Dehydrase (ColD) and GDP-L-Colitose Synthase (ColC). *Biochemistry* **2004**, 43 (51), 16450–16460. <https://doi.org/10.1021/bi0483763>.
- (225) Earl, L. A.; Bi, S.; Baum, L. G. Galectin Multimerization and Lattice Formation Are Regulated by Linker Region Structure. *Glycobiology* **2011**, 21 (1), 6–12. <https://doi.org/10.1093/glycob/cwq144>.
- (226) Kainz, K.; Bauer, M. A.; Madeo, F.; Carmona-Gutierrez, D. Fungal Infections in Humans: The Silent Crisis. *Microb. Cell* **2020**, 7 (6), 143–145. <https://doi.org/10.15698/mic2020.06.718>.
- (227) Amarante-Mendes, G. P.; Adjemian, S.; Branco, L. M.; Zanetti, L. C.; Weinlich, R.; Bortoluci, K. R. Pattern Recognition Receptors and the Host Cell Death Molecular Machinery. *Front. Immunol.* **2018**, 9 (OCT), 1–19. <https://doi.org/10.3389/fimmu.2018.02379>.
- (228) Gow, N. A. R.; Latge, J. P.; Munro, C. A. The Fungal Cell Wall: Structure, Biosynthesis, and Function. *The Fungal Kingdom* **2017**, No. May, 267–292. <https://doi.org/10.1128/9781555819583.ch12>.
- (229) Garcia-Rubio, R.; de Oliveira, H. C.; Rivera, J.; Trevijano-Contador, N. The Fungal Cell Wall: Candida, Cryptococcus, and Aspergillus Species. *Front. Microbiol.* **2020**, 10 (January), 1–13. <https://doi.org/10.3389/fmicb.2019.02993>.
- (230) Hatinguais, R.; Willment, J. A.; Brown, G. D. PAMPs of the Fungal Cell Wall and Mammalian PRRs. *Curr. Top. Microbiol. Immunol.* **2020**, 425, 187–223. https://doi.org/10.1007/82_2020_201.
- (231) Brown, G. D.; Taylor, P. R.; Reid, D. M.; Willment, J. A.; Williams, D. L.; Martinez-Pomares, L.; Wong, S. Y. C.; Gordon, S. Dectin-1 Is A Major β -Glucan Receptor On Macrophages. *J. Exp. Med.* **2002**, 196 (3), 407–412. <https://doi.org/10.1084/jem.20020470>.
- (232) Brown, G. D. Dectin-1 : A Signalling Non-TLR Pattern-Recognition Receptor. *Nat. Rev. Immunol.* **2006**, 6 (1), 33–43. <https://doi.org/10.1038/nri1745>.
- (233) Brown, J.; O’Callaghan, C. A.; Marshall, A. S. J.; Gilbert, R. J. C.; Siebold, C.; Gordon, S.; Brown, G. D.; Jones, E. Y. Structure of the Fungal β -Glucan-Binding Immune Receptor Dectin-1: Implications for Function. *Protein Sci.* **2007**, 16 (6), 1042–1052. <https://doi.org/10.1110/ps.072791207>.

- (234) Dulal, H. P.; Nagae, M.; Ikeda, A.; Morita-Matsumoto, K.; Adachi, Y.; Ohno, N.; Yamaguchi, Y. Enhancement of Solubility and Yield of a β -Glucan Receptor Dectin-1 C-Type Lectin-like Domain in *Escherichia Coli* with a Solubility-Enhancement Tag. *Protein Expr. Purif.* **2016**, *123*, 97–104. <https://doi.org/10.1016/j.pep.2016.04.002>.
- (235) Adachi, Y.; Ishii, T.; Ikeda, Y.; Hoshino, A.; Tamura, H.; Aketagawa, J.; Tanaka, S.; Ohno, N. Characterization of β -Glucan Recognition Site on C-Type Lectin, Dectin 1. *Infect. Immun.* **2004**, *72* (7), 4159–4171. <https://doi.org/10.1128/IAI.72.7.4159-4171.2004>.
- (236) Hanashima, S.; Ikeda, A.; Tanaka, H.; Adachi, Y.; Ohno, N.; Takahashi, T.; Yamaguchi, Y. NMR Study of Short $\beta(1-3)$ -Glucans Provides Insights into the Structure and Interaction with Dectin-1. *Glycoconj. J.* **2014**, *31* (3), 199–207. <https://doi.org/10.1007/s10719-013-9510-x>.
- (237) Adams, E. L.; Rice, P. J.; Graves, B.; Ensley, H. E.; Yu, H.; Brown, G. D.; Gordon, S.; Monteiro, M. A.; Papp-Szabo, E.; Lowman, D. W.; Power, T. D.; Wempe, M. F.; Williams, D. L. Differential High-Affinity Interaction of Dectin-1 with Natural or Synthetic Glucans Is Dependent upon Primary Structure and Is Influenced by Polymer Chain Length and Side-Chain Branching. *J. Pharmacol. Exp. Ther.* **2008**, *325* (1), 115–123. <https://doi.org/10.1124/jpet.107.133124>.
- (238) Elder, M. J.; Webster, S. J.; Chee, R.; Williams, D. L.; Hill Gaston, J. S.; Goodall, J. C. β -Glucan Size Controls Dectin-1-Mediated Immune Responses in Human Dendritic Cells by Regulating IL-1 β Production. *Front. Immunol.* **2017**, *8* (JUL), 14–16. <https://doi.org/10.3389/fimmu.2017.00791>.
- (239) Palma, A. S.; Feizi, T.; Zhang, Y.; Stoll, M. S.; Lawson, A. M.; Díaz-Rodríguez, E.; Campanero-Rhodes, M. A.; Costa, J.; Gordon, S.; Brown, G. D.; Chai, W. Ligands for the β -Glucan Receptor, Dectin-1, Assigned Using “Designer” Microarrays of Oligosaccharide Probes (Neoglycolipids) Generated from Glucan Polysaccharides. *J. Biol. Chem.* **2006**, *281* (9), 5771–5779. <https://doi.org/10.1074/jbc.M511461200>.
- (240) Tada, R.; Adachi, Y.; Ishibashi, K.; Tsubaki, K.; Ohno, N. Binding Capacity of a Barley β -D-Glucan to the β -Glucan Recognition Molecule Dectin-1. *J. Agric. Food Chem.* **2008**, *56* (4), 1442–1450. <https://doi.org/10.1021/jf073221y>.
- (241) Delbianco, M.; Kononov, A.; Poveda, A.; Yu, Y.; Diercks, T.; Jiménez-Barbero, J.; Seeberger, P. H. Well-Defined Oligo- and Polysaccharides as Ideal Probes for Structural Studies. *J. Am. Chem. Soc.* **2018**, *140* (16), 5421–5426. <https://doi.org/10.1021/jacs.8b00254>.

- (242) Poveda, A.; Fittolani, G.; Seeberger, P. H.; Delbianco, M.; Jiménez-Barbero, J. The Flexibility of Oligosaccharides Unveiled Through Residual Dipolar Coupling Analysis. *Front. Mol. Biosci.* **2021**, *8*. <https://doi.org/10.3389/fmolb.2021.784318>.
- (243) Calloni, I.; Unione, L.; Jiménez-Osés, G.; Corzana, F.; Del Bino, L.; Corrado, A.; Pitirollo, O.; Colombo, C.; Lay, L.; Adamo, R.; Jiménez-Barbero, J. The Conformation of the Mannopyranosyl Phosphate Repeating Unit of the Capsular Polysaccharide of *Neisseria Meningitidis* Serogroup A and Its Carba-Mimetic. *European J. Org. Chem.* **2018**, *2018* (33), 4548–4555. <https://doi.org/10.1002/ejoc.201801003>.
- (244) Oldrini, D.; Bino, L.; Arda, A.; Carboni, F.; Henriques, P.; Angiolini, F.; Quintana, J. I.; Calloni, I.; Romano, M. R.; Berti, F.; Jimenez-Barbero, J.; Margarit, I.; Adamo, R. Structure-Guided Design of a Group B *Streptococcus* Type III Synthetic Glycan–Conjugate Vaccine. *Chem. – A Eur. J.* **2020**, *26* (31), 6944–6944. <https://doi.org/10.1002/chem.202002016>.
- (245) Seeberger, P. H.; Werz, D. B. Synthesis and Medical Applications of Oligosaccharides. *Nature* **2007**, *446* (7139), 1046–1051. <https://doi.org/10.1038/nature05819>.
- (246) Zhu, Y.; Tyrikos-Ergas, T.; Schiefelbein, K.; Grafmüller, A.; Seeberger, P. H.; Delbianco, M. Automated Access to Well-Defined Ionic Oligosaccharides. *Org. Biomol. Chem.* **2020**, *18* (7), 1349–1353. <https://doi.org/10.1039/d0ob00137f>.
- (247) Fittolani, G.; Tyrikos-Ergas, T.; Vargová, D.; Chaube, M. A.; Delbianco, M. Progress and Challenges in the Synthesis of Sequence Controlled Polysaccharides. *Beilstein J. Org. Chem.* **2021**, *17*, 1981–2025. <https://doi.org/10.3762/bjoc.17.129>.
- (248) Tyrikos-Ergas, T.; Sletten, E. T.; Huang, J. Y.; Seeberger, P. H.; Delbianco, M. On Resin Synthesis of Sulfated Oligosaccharides. *Chem. Sci.* **2022**, *13* (7), 2115–2120. <https://doi.org/10.1039/d1sc06063e>.
- (249) Sarac, I.; Meier, C. Efficient Automated Solid-Phase Synthesis of DNA and RNA 5'-Triphosphates. *Chem. - A Eur. J.* **2015**, *21* (46), 16421–16426. <https://doi.org/10.1002/chem.201502844>.
- (250) Palomo, J. M. Solid-Phase Peptide Synthesis: An Overview Focused on the Preparation of Biologically Relevant Peptides. *RSC Adv.* **2014**, *4* (62), 32658–32672. <https://doi.org/10.1039/c4ra02458c>.
- (251) Guberman, M.; Seeberger, P. H. Automated Glycan Assembly: A Perspective. *J. Am. Chem. Soc.* **2019**, *141* (14), 5581–5592. <https://doi.org/10.1021/jacs.9b00638>.

- (252) Hahm, H. S.; Schlegel, M. K.; Hurevich, M.; Eller, S.; Schuhmacher, F.; Hofmann, J.; Pagel, K.; Seeberger, P. H. Automated Glycan Assembly Using the Glycoener 2.1 Synthesizer. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (17), E3385–E3389. <https://doi.org/10.1073/pnas.1700141114>.
- (253) Weishaupt, M. W.; Hahm, H. S.; Geissner, A.; Seeberger, P. H. Automated Glycan Assembly of Branched β -(1,3)-Glucans to Identify Antibody Epitopes. *Chem. Commun.* **2017**, *53* (25), 3591–3594. <https://doi.org/10.1039/C7CC00520B>.
- (254) Eller, S.; Collot, M.; Yin, J.; Hahm, H. S.; Seeberger, P. H. Automated Solid-Phase Synthesis of Chondroitin Sulfate Glycosaminoglycans. *Angew. Chemie Int. Ed.* **2013**, *52* (22), 5858–5861. <https://doi.org/10.1002/anie.201210132>.
- (255) Le Mai Hoang, K.; Pardo-Vargas, A.; Zhu, Y.; Yu, Y.; Loria, M.; Delbianco, M.; Seeberger, P. H. Traceless Photolabile Linker Expedites the Chemical Synthesis of Complex Oligosaccharides by Automated Glycan Assembly. *J. Am. Chem. Soc.* **2019**, *141* (22), 9079–9086. <https://doi.org/10.1021/jacs.9b03769>.
- (256) Rustiguel, J. K.; Kumagai, P. S.; Dias-Baruffi, M.; Costa-Filho, A. J.; Nonato, M. C. Recombinant Expression, Purification and Preliminary Biophysical and Structural Studies of C-Terminal Carbohydrate Recognition Domain from Human Galectin-4. *Protein Expr. Purif.* **2016**, *118*, 39–48. <https://doi.org/10.1016/j.pep.2015.09.026>.
- (257) Hoopes, J. T.; Elberson, M. A.; Preston, R. J.; Reddy, P. T.; Kelman, Z. Protein Labeling in Escherichia Coli with 2H, 13C, and 15N; 2015; pp 27–44. <https://doi.org/10.1016/bs.mie.2015.08.023>.
- (258) Maley, F.; Trimble, R. B.; Tarentino, A. L.; Plummer, T. H. Characterization of Glycoproteins and Their Associated Oligosaccharides through the Use of Endoglycosidases. *Anal. Biochem.* **1989**, *180* (2), 195–204. [https://doi.org/10.1016/0003-2697\(89\)90115-2](https://doi.org/10.1016/0003-2697(89)90115-2).
- (259) Greenfield, N. J. Using Circular Dichroism Collected as a Function of Temperature to Determine the Thermodynamics of Protein Unfolding and Binding Interactions. *Nat. Protoc.* **2006**, *1* (6), 2527–2535. <https://doi.org/10.1038/nprot.2006.204>.
- (260) Woods Group. (2005-XXXX) GLYCAM Web. Complex Carbohydrate Research Center, University of Georgia, Athens, GA. ([Http://Glycam.Org](http://Glycam.Org)).
- (261) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A

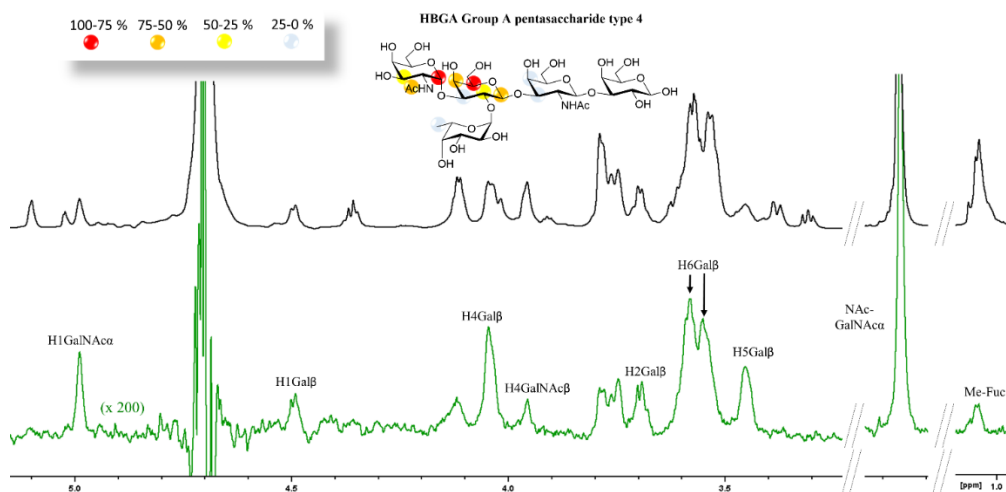
- Generalizable Biomolecular Force Field. Carbohydrates. *J. Comput. Chem.* **2008**, 29 (4), 622–655. <https://doi.org/10.1002/jcc.20820>.
- (262) Mayer, M.; Meyer, B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chemie Int. Ed.* **1999**, 38 (12), 1784–1788. [https://doi.org/10.1002/\(SICI\)1521-3773\(19990614\)38:12<1784::AID-ANIE1784>3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1521-3773(19990614)38:12<1784::AID-ANIE1784>3.0.CO;2-Q).
- (263) Cutting, B.; Ghose, R.; Bodenhausen, G. Tilt Angle Dependence of Cross-Relaxation in Off-Resonance ROESY. *J. Magn. Reson.* **1999**, 138 (2), 326–329. <https://doi.org/10.1006/jmre.1999.1734>.
- (264) Keller, R. *The Computer Aided Resonance Assignment Tutorial*; 2004.
- (265) MicroCal Origin 7 (OriginLab Corporation, Northampton, MA, United States).
- (266) Eller, S.; Collot, M.; Yin, J.; Hahm, H. S.; Seeberger, P. H. Automated Solid-Phase Synthesis of Chondroitin Sulfate Glycosaminoglycans. *Angew. Chemie Int. Ed.* **2013**, 52 (22), 5858–5861. <https://doi.org/10.1002/anie.201210132>.
- (267) Gude, M.; Ryf, J.; White, P. D. An Accurate Method for the Quantitation of Fmoc-Derivatized Solid Phase Supports. *Lett. Pept. Sci.* **2002**, 9 (4), 203–206. <https://doi.org/doi.org/10.1023/A:1024148619149>.
- (268) Guberman, M.; Bräutigam, M.; Seeberger, P. H. Automated Glycan Assembly of Lewis Type I and II Oligosaccharide Antigens. *Chem. Sci.* **2019**, 10 (21), 5634–5640. <https://doi.org/10.1039/C9SC00768G>.
- (269) Hurevich, M.; Kandasamy, J.; Ponnappa, B. M.; Collot, M.; Kopetzki, D.; McQuade, D. T.; Seeberger, P. H. Continuous Photochemical Cleavage of Linkers for Solid-Phase Synthesis. *Org. Lett.* **2014**, 16 (6), 1794–1797. <https://doi.org/10.1021/ol500530q>.

7 Supporting Information

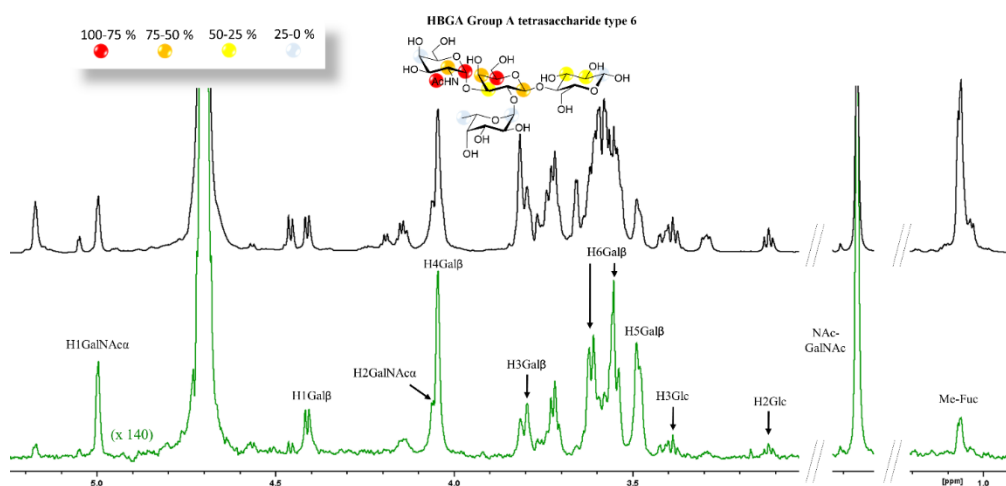
7.1 The interaction between Galectin-4 and histo blood group antigens

7.1.1 N-terminal domain of Galectin-4

7.1.1.1 Saturation Transfer Difference (STD-NMR)

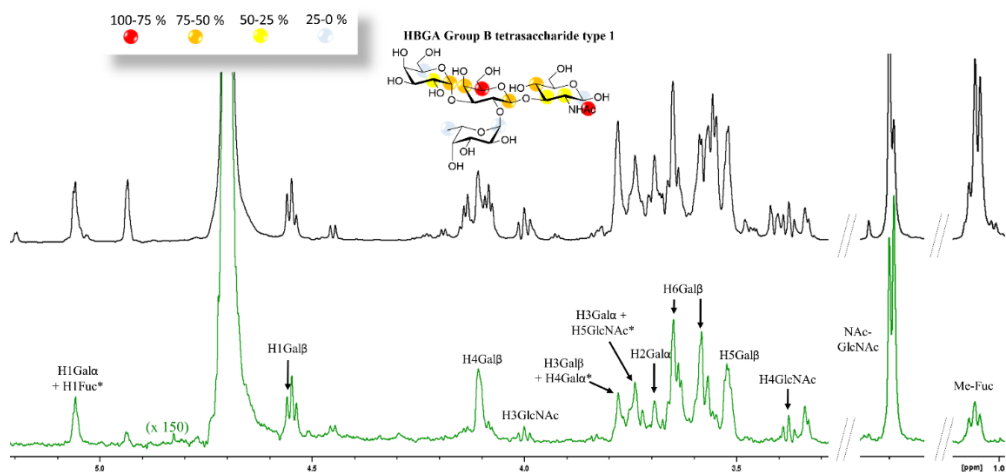


Supplementary Figure 1. ^1H STD-NMR for a sample of group A type-4 pentasaccharide and Gal-4N (50:1 molar ratio). Top: the reference spectrum (black, off-resonance). Bottom: the STD-NMR spectrum (green). The ^1H -NMR signals showing STD effect are annotated. The epitope mapping (relative STD) is shown in the ligand structure.

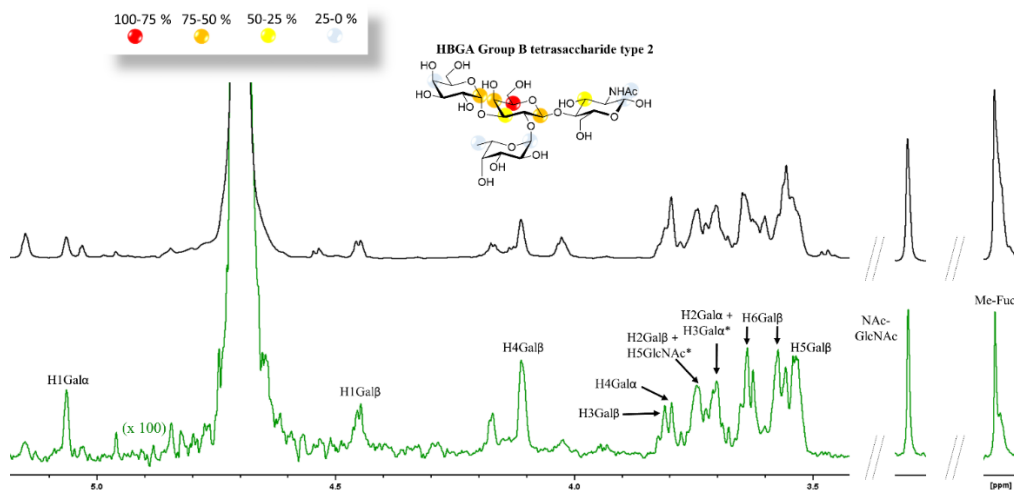


Supplementary Figure 2. ^1H STD-NMR for a sample of group A type-6 tetrasaccharide and Gal-4N (50:1 molar ratio). Top: the reference spectrum (black, off-resonance). Bottom: the

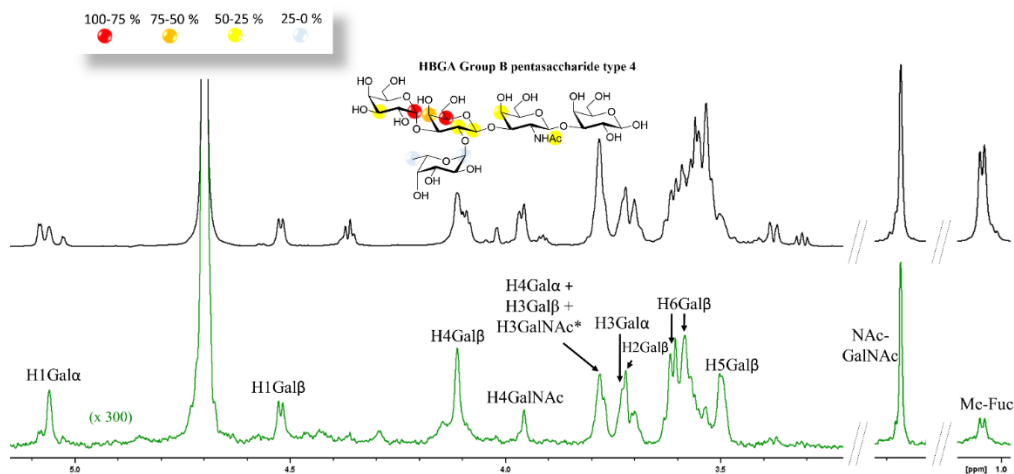
STD-NMR spectrum (green). The ^1H -NMR signals showing STD effect are annotated. The epitope mapping (relative STD) is shown in the ligand structure.



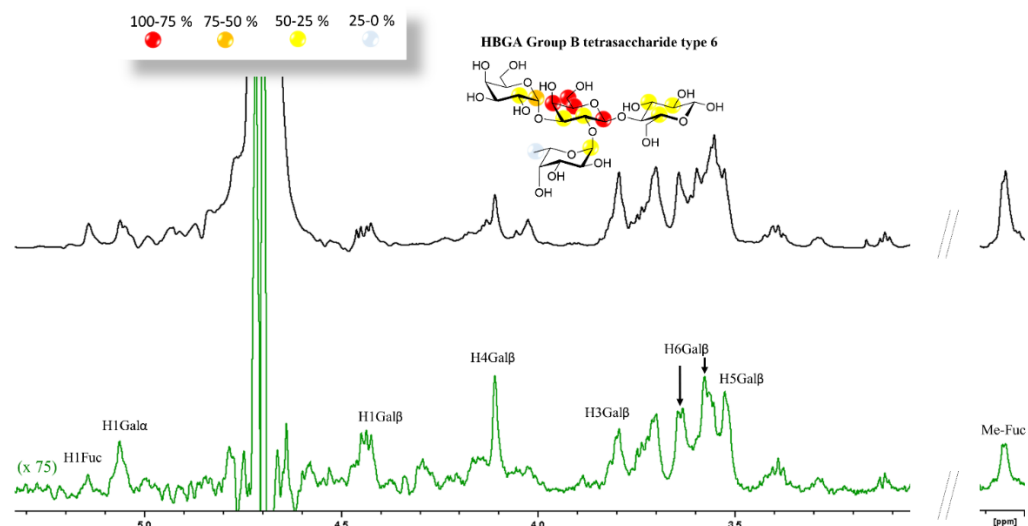
Supplementary Figure 3. ^1H STD-NMR for a sample of group B type-1 tetrasaccharide and Gal-4N (50:1 molar ratio). Top: the reference spectrum (black, off-resonance). Bottom: the STD-NMR spectrum (green). The ^1H -NMR signals showing STD effect are annotated. The epitope mapping (relative STD) is shown in the ligand structure. * overlapping resonances.



Supplementary Figure 4. ^1H STD-NMR for a sample of B type-2 tetrasaccharide and Gal-4N (50:1 molar ratio). Top: the reference spectrum (black, off-resonance). Bottom: the STD-NMR spectrum (green). The ^1H -NMR signals showing STD effect are annotated. The epitope mapping (relative STD) is shown in the ligand structure. * overlapping resonances.



Supplementary Figure 5. ^1H STD-NMR for a sample of group B type-4 pentasaccharide and Gal-4N (50:1 molar ratio). Top: the reference spectrum (black, off-resonance). Bottom: the STD-NMR spectrum (green). The ^1H -NMR signals showing STD effect are annotated. The epitope mapping (relative STD) is shown in the ligand structure. * overlapping resonances.



Supplementary Figure 6. ^1H STD-NMR for a sample of group B type-6 tetrasaccharide and Gal-4N (50:1 molar ratio). Top: the reference spectrum (black, off-resonance). Bottom: the STD-NMR spectrum (green). The ^1H -NMR signals showing STD effect are annotated. The epitope mapping (relative STD) is shown in the ligand structure.

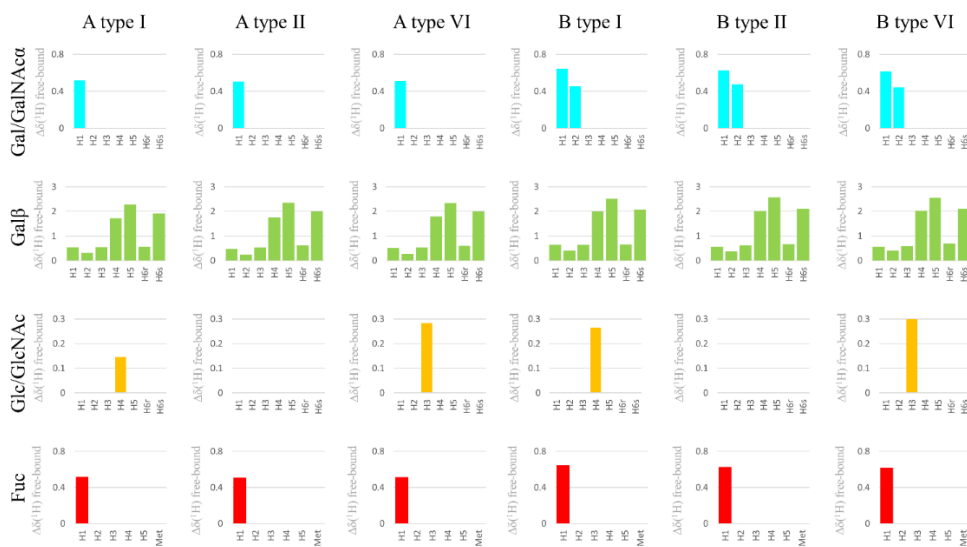
Supplementary Table 1. Relative STD-AF (Amplification factor) and STD percentage of each non overlapping proton of each ligand.

Ligand	^1H -residue	STD-AF	STD%
--------	-----------------------	--------	------

A type 4	H1GalNAc	0.0145	95%
	H3GalNAc	0.0043	28%
	NHAc GalNAc	0.0104	68%
	H1Gal β	0.0092	61%
	H2Gal β	0.0054	36%
	H3Gal β	0.0032	21%
	H4Gal β	0.0108	71%
	H5Gal β	0.0152	100%
	H3GalNAcb	0.0023	15%
	H4GalNAcb	0.0036	24%
	MetFuc	0.0012	18%
	A type 6	H1GalNAc	0.0129
H2GalNAc		0.0081	51%
H4GalNAc		0.0025	16%
NHAcGalNAc		0.0158	100%
H1Gal β		0.0091	58%
H3Gal β		0.0061	39%
H4Gal β		0.0095	60%
H5Gal β		0.0158	100%
H1Glc α		0.0033	21%
H1Glc β		0.0028	18%
H2Glc β		0.0048	30%
H2Glc α		0.0054	34%
H3Glc β		0.0050	32%
H1 Fuc		0.0020	13%
MetFuc		0.0016	10%
B type 1	H1Gal α	0.0053	56%
	H2Gal α	0.0033	35%
	H5Gal α	0.0022	23%
	H1Gal β	0.0066	70%
	H4Gal β	0.0070	74%
	H5Gal β	0.0094	100%
	H1GlcNAc	0.0015	16%
	H2GlcNAc	0.0027	29%
	H3GlcNAc	0.0036	38%
	H4GlcNAc	0.0050	53%
	NHAcGlcNAc	0.0133	141%
	NHAcGlcNAc	0.0055	59%
	Average NHAc	0.0094	100%

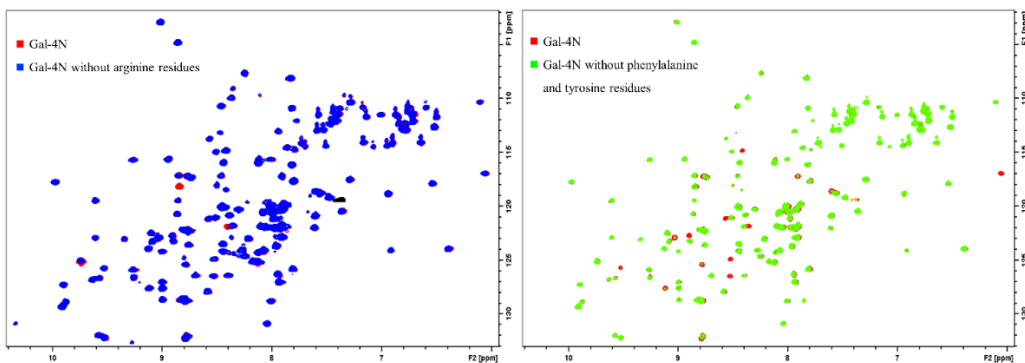
	H1 Fuc	0.0000	0%
	MetFuc	0.0013	14%
B type 2	H1Gal α	0.0108	73%
	H4Gal α	0.0027	18%
	H1Gal β	0.0088	59%
	H3Gal β	0.0054	36%
	H4Gal β	0.0081	55%
	H5Gal β	0.0148	100%
	H1GlcNAc	0.0035	24%
	H3GlcNAc	0.0054	36%
	NHAc GlcNAc	0.0035	9%
	H1 Fuc	0.0024	16%
	MetFuc	0.0023	16%
	B type 4	H1Gal α	0.0094
H3Gal α		0.0041	44%
H1Gal β		0.0046	49%
H2Gal β		0.0044	47%
H4Gal β		0.0064	68%
H5Gal β		0.0081	86%
H4GalNAc		0.0028	30%
NHAc		0.0034	36%
H1 Fuc		0.0020	21%
MetFuc		0.0009	10%
B type 6	H1Gal α	0.0276	74%
	H2Gal α	0.0142	38%
	H1Gal β	0.0298	80%
	H2Gal β	0.0129	34%
	H3Gal β	0.0108	29%
	H4Gal β	0.0307	82%
	H5Gal β	0.0374	100%
	H2Glc β	0.0135	36%
	H2Glc α	0.0146	39%
	H3Glc β	0.0172	46%
	H5Glc β	0.0163	44%
	H1 Fuc	0.0105	28%
	MetFuc	0.0074	20%

7.1.1.2 trROESY NMR

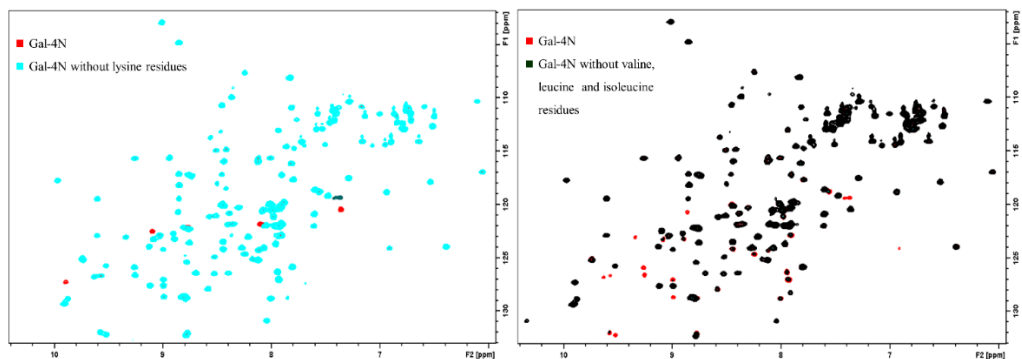


Supplementary Figure 7. Plot for the differences in chemical shifts between the free and bound states for every tetrasaccharide from trROESY experiments.

7.1.1.3 Protein Backbone Resonance Assignment

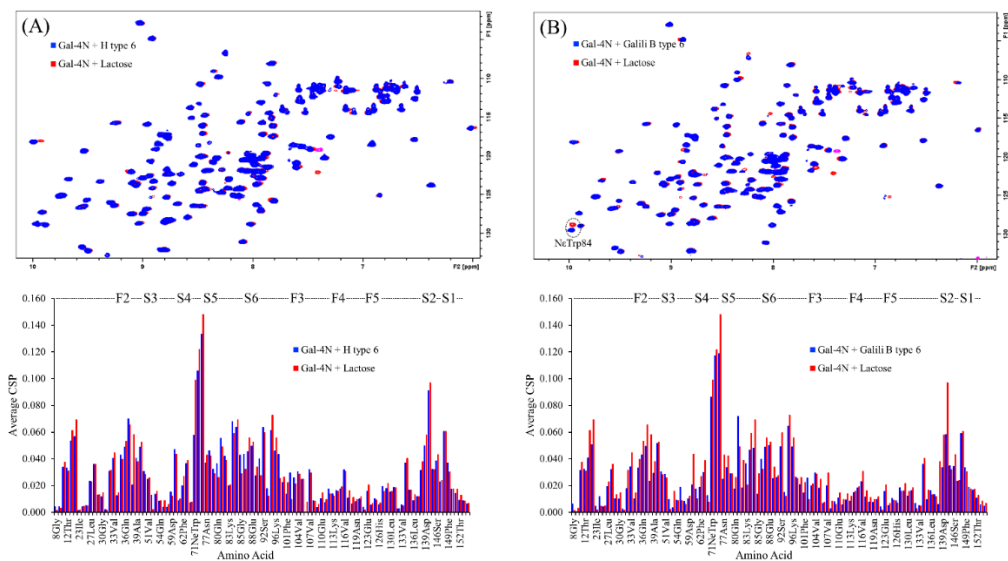


Supplementary Figure 8. Left: Superimposition of the ^1H - ^{15}N HSQC spectra of Gal-4N (red) and Gal-4N with reduced peak intensity for arginine residues (blue). Right: Superimposition of the ^1H - ^{15}N HSQC spectra of Gal-4N (red) and Gal-4N with reduced peak intensity for phenylalanine and tyrosine residues (green).

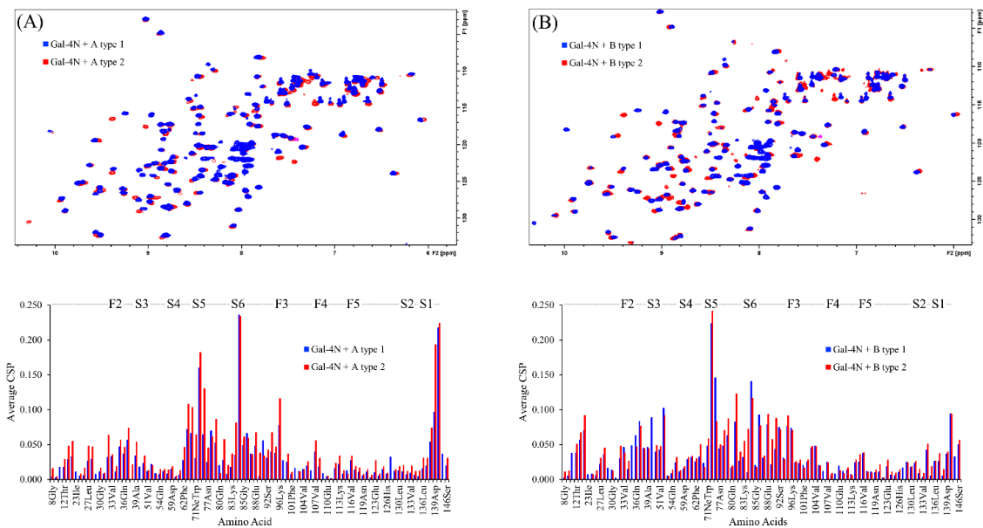


Supplementary Figure 9. Left: Superimposition of the ^1H - ^{15}N HSQC spectra of Gal-4N (red) and Gal-4N with reduced peak intensity for lysine residues (light blue). Right: Superimposition of the ^1H - ^{15}N HSQC spectra of Gal-4N (red) and Gal-4N with reduced peak intensity for valine, leucine and isoleucine residues (black).

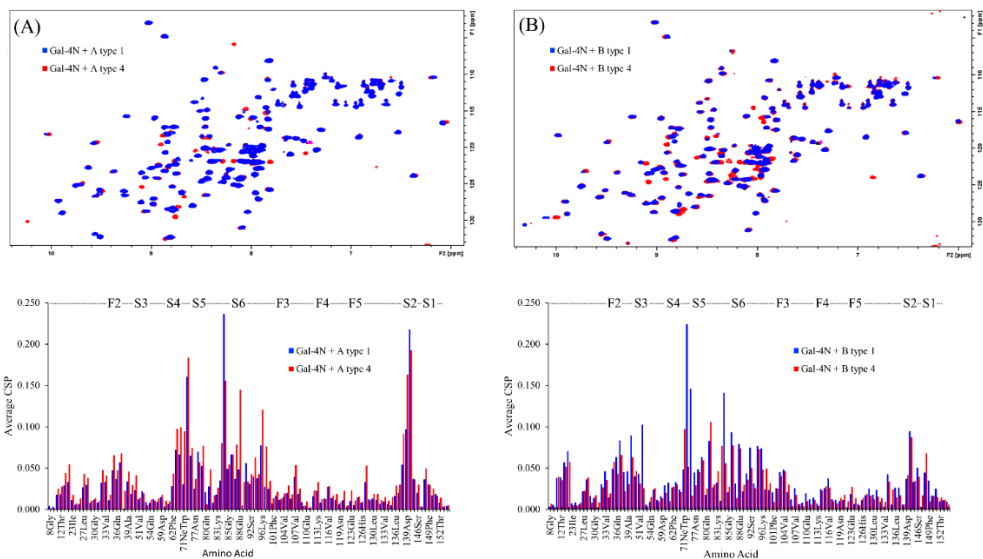
7.1.1.4 Chemical shift perturbations at Gal-4N



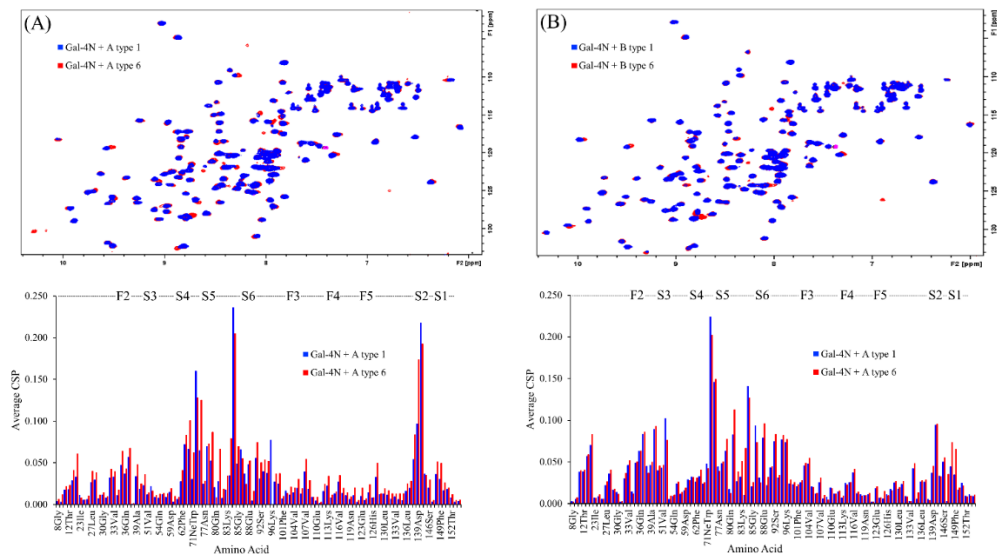
Supplementary Figure 10. A) Top: Superimposition of the ^1H - ^{15}N HSQC spectra of Gal-4N saturated with H type-6 (blue) and lactose (red). Bottom: Chemical Shift Perturbation of Gal-4N with H type-6 (blue) and lactose (red). B) Top: Superimposition of the ^1H - ^{15}N HSQC spectra of Gal-4N saturated with group B type-6 trisaccharide (blue) and lactose (red). Bottom: Chemical Shift Perturbation of Gal-4N with B type-6 trisaccharide (blue) and lactose (red).



Supplementary Figure 11. A) Top: Superimposition of the ^1H - ^{15}N HSQC spectra of Gal-4N saturated with A type-1 (blue) and A type-2 (red). Bottom: Chemical Shift Perturbation of Gal-4N with A type-1 (blue) and A type-2 (red). B) Top: Superimposition of the ^1H - ^{15}N HSQC spectra of Gal-4N saturated with B type-1 (blue) and B type-2 (red). Bottom: Chemical Shift Perturbation of Gal-4N with B type-1 (blue) and B type-2 (red).

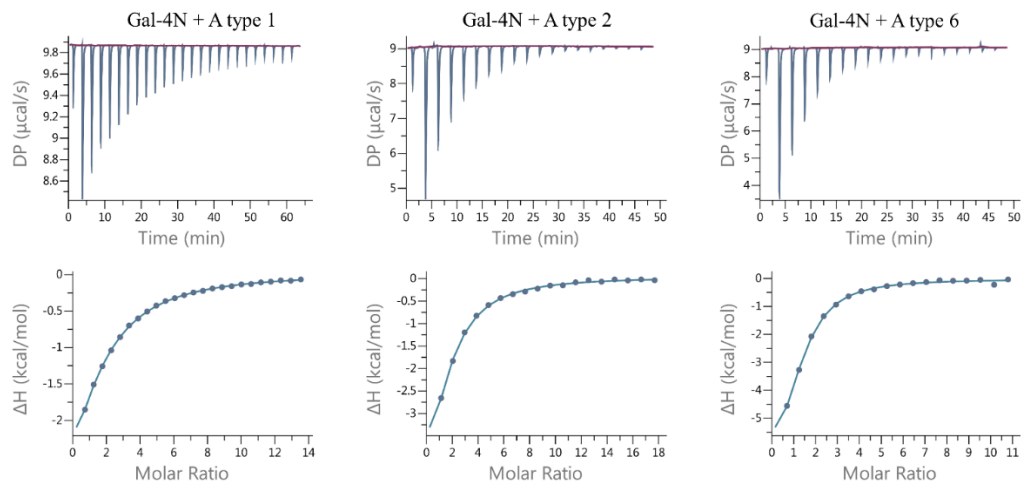


Supplementary Figure 12. A) Top: Superimposition of the ^1H - ^{15}N HSQC spectra of Gal-4N saturated with A type-1 (blue) and A type-6 (red). Bottom: Chemical Shift Perturbation of Gal-4N with A type-1 (blue) and A type-6 (red). B) Top: Superimposition of the ^1H - ^{15}N HSQC spectra of Gal-4N saturated with B type-1 (blue) and B type-6 (red). Bottom: Chemical Shift Perturbation of Gal-4N with B type-1 (blue) and B type-6 (red).

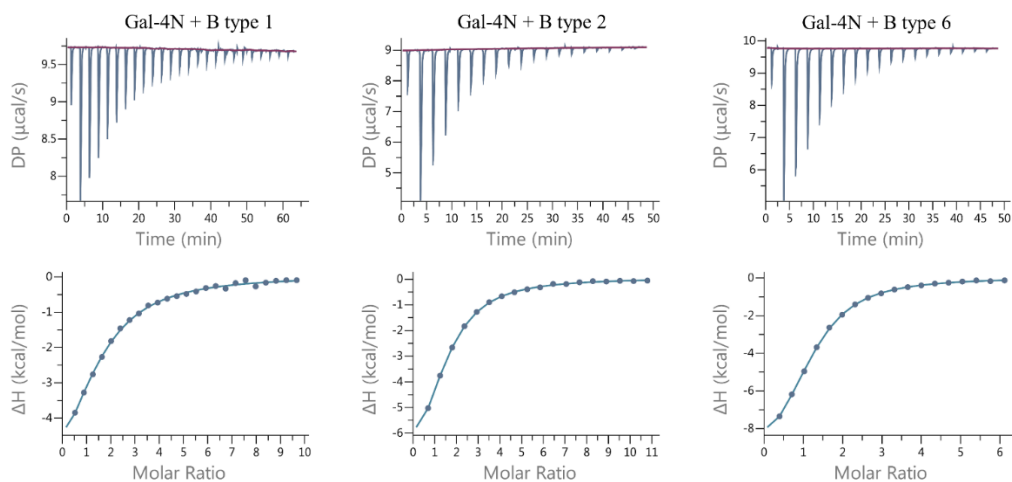


Supplementary Figure 13. A) Top: Superimposition of the ^1H - ^{15}N HSQC spectra of Gal-4N saturated with A type-1 (blue) and A type-6 (red). Bottom: Chemical Shift Perturbation of Gal-4N with A type-1 (blue) and A type-6 (red). B) Top: Superimposition of the ^1H - ^{15}N HSQC spectra of Gal-4N saturated with B type-1 (blue) and B type-6 (red). Bottom: Chemical Shift Perturbation of Gal-4N with B type-1 (blue) and B type-6 (red).

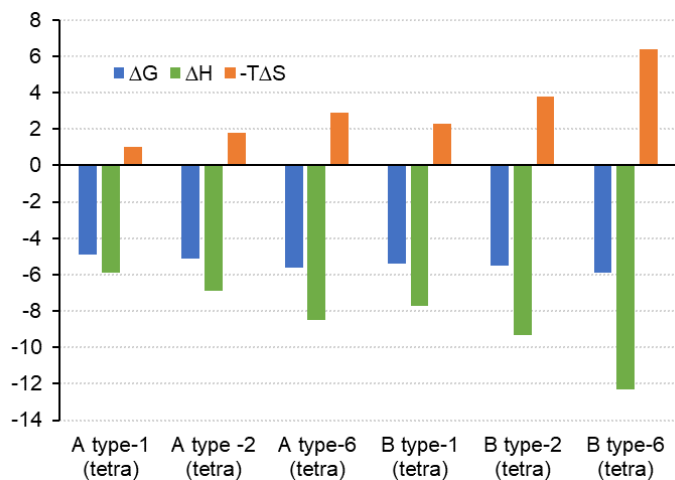
7.1.1.5 Isothermal titration calorimetry (ITC)



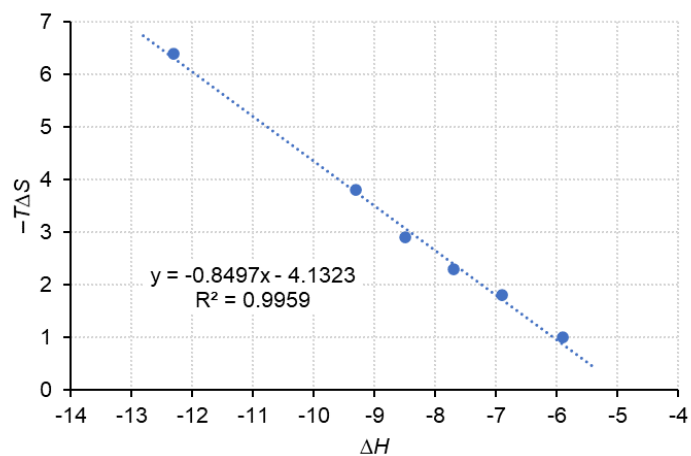
Supplementary Figure 14. Examples of titration profiles of titration of Gal-4N with group A tetrasaccharides. Top: Row data from the titration representing $\mu\text{cal/s}$ dispersed during time. Bottom: The enthalpy of binding (kcal/mol) for each injection plotted against the protein/ligand molar ratio. The continuous line represents the least-squares-fit of the data to a single-site binding model.



Supplementary Figure 15. Examples of titration profiles of titration of Gal-4N with group B tetrasaccharides. Top: Row data from the titration representing $\mu\text{cal/s}$ dispersed during time. Bottom: The enthalpy of binding (kcal/mol) for each injection plotted against the protein/ligand molar ratio. The continuous line represents the least-squares-fit of the data to a single-site binding model.



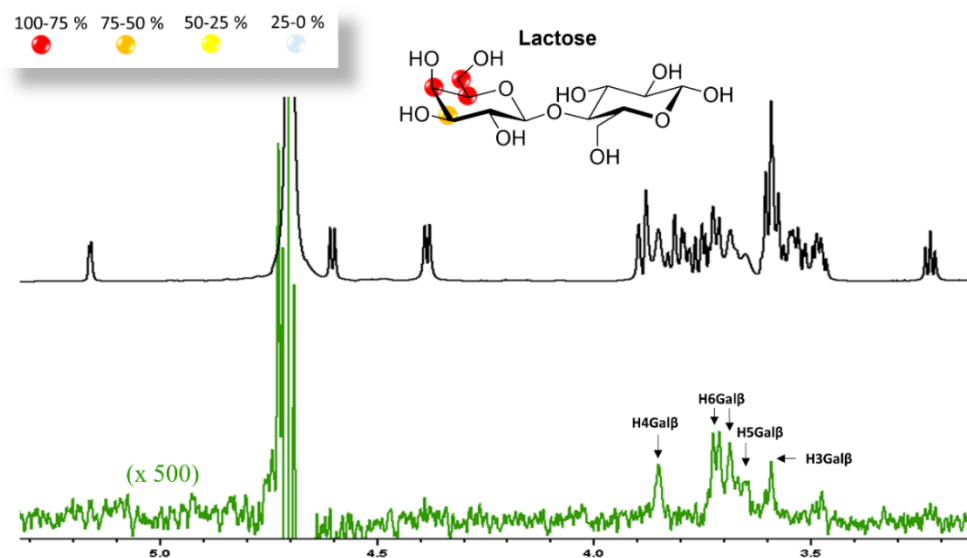
Supplementary Figure 16. Binding free energies (ΔG), enthalpies (ΔH) and entropies (ΔS), in kcal mol^{-1} , derived for ligands A/B-type-1/2/6 from Isothermal Titration Calorimetry (ITC) measurements.



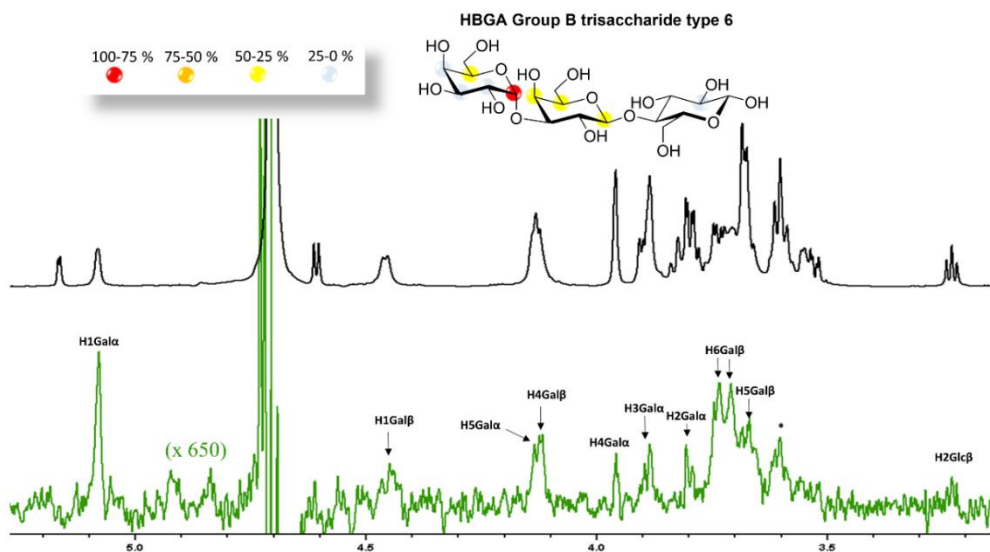
Supplementary Figure 17. Binding enthalpies (ΔH) versus entropies (ΔS), in kcal mol⁻¹, derived for ligands A/B-type-1/2/6 from Isothermal Titration Calorimetry (ITC) measurements. Note the nearly perfect linear relationship showing clear enthalpy/entropy compensation.

7.1.2 C-terminal domain of Galectin-4

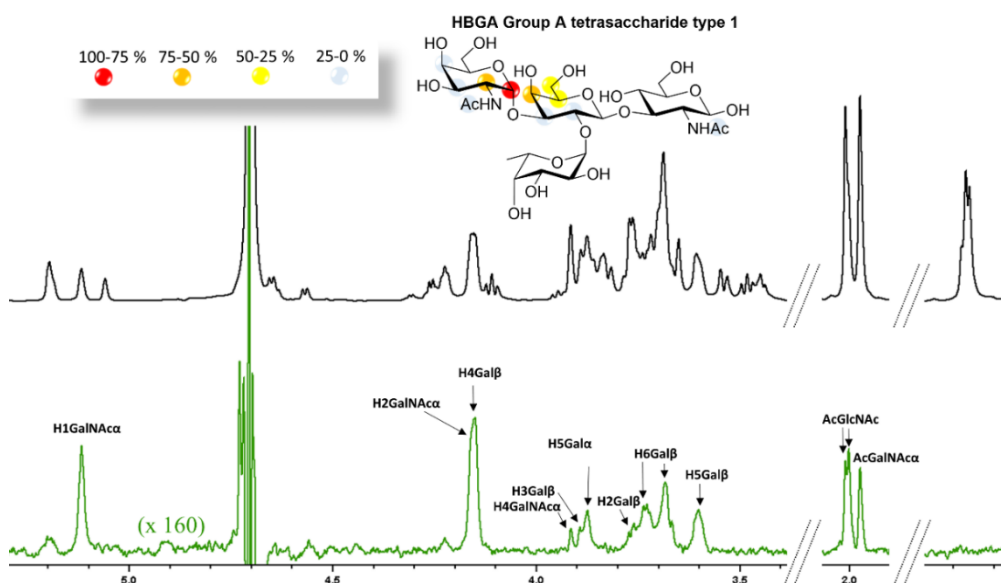
7.1.2.1 Saturation Transfer Difference (STD-NMR)



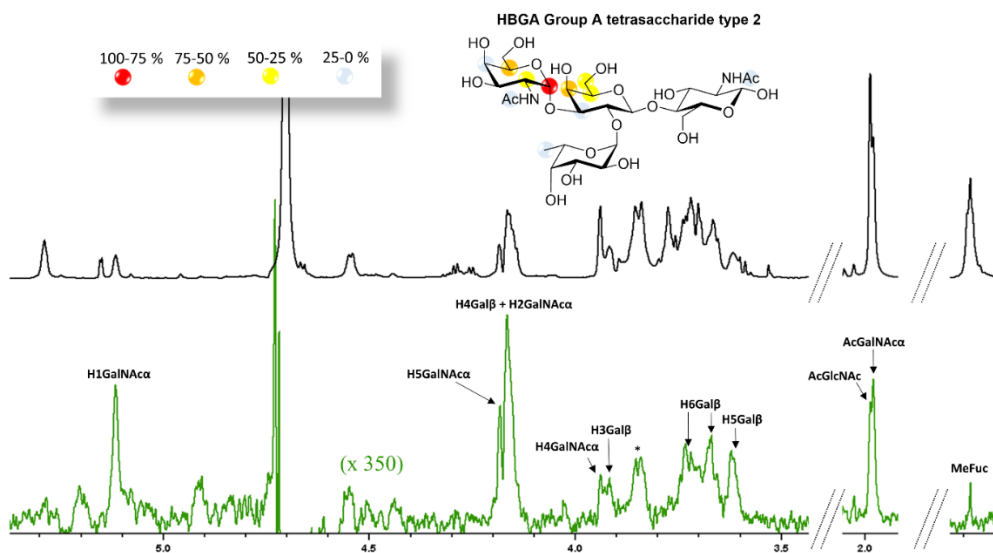
Supplementary Figure 18. ¹H-STD-NMR of the sample containing 50 μM of Gal-4C and 50 equivalents of lactose. In black the off-resonance spectrum (reference spectrum). In green the STD spectrum. The protons of the ligand that show STD effect are annotated. On top the structure of the ligand and the epitope mapping (relative STD) are shown.



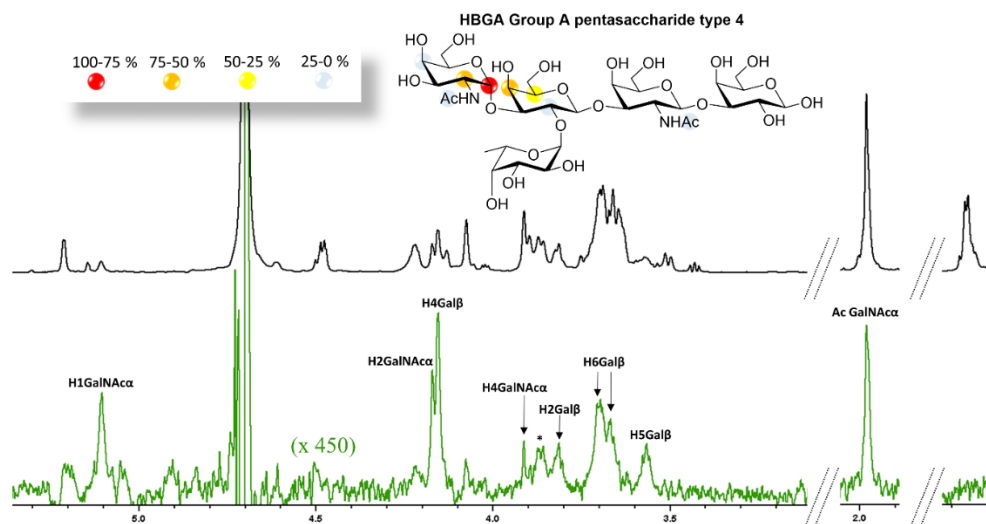
Supplementary Figure 19. ^1H -STD-NMR of the sample containing $50\ \mu\text{M}$ of Gal-4C and 50 equivalents of B type 6 linear trisaccharide. In black the off-resonance spectrum (reference spectrum). In green the STD spectrum. The protons of the ligand that show STD effect are annotated. On top the structure of the ligand and the epitope mapping (relative STD) are shown.



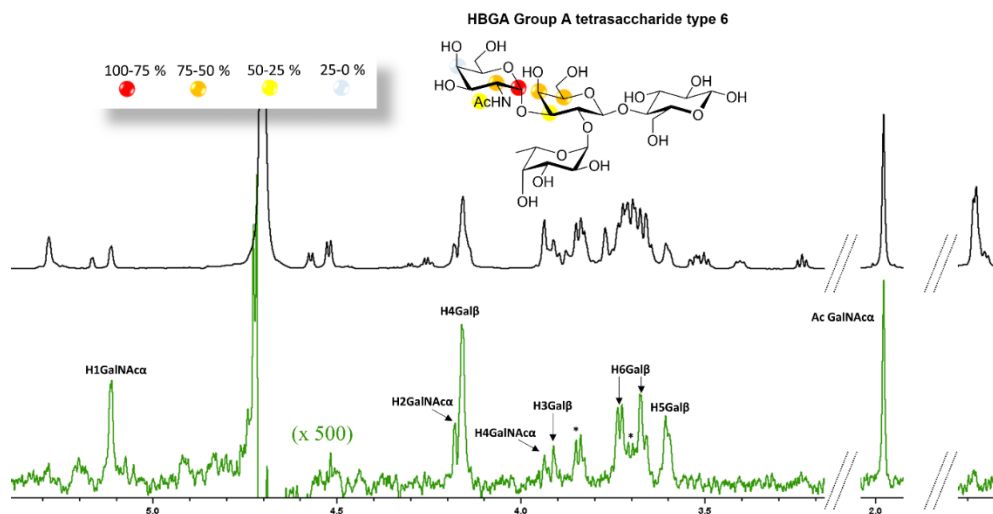
Supplementary Figure 20. ^1H -STD-NMR of the sample containing $50\ \mu\text{M}$ of Gal-4C and 50 equivalents of A type 1 tetrasaccharide. In black the off-resonance spectrum (reference spectrum). In green the STD spectrum. The protons of the ligand that show STD effect are annotated. On top the structure of the ligand and the epitope mapping (relative STD) are shown.



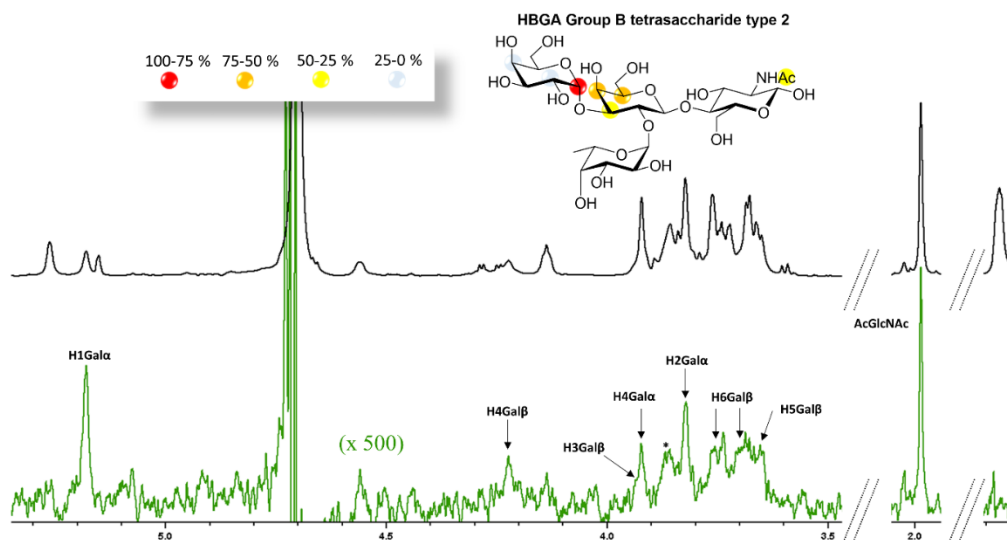
Supplementary Figure 21. ^1H -STD-NMR of the sample containing 50 μM of Gal-4C and 50 equivalents of A type 2 tetrasaccharide. In black the off-resonance spectrum (reference spectrum). In green the STD spectrum. The protons of the ligand that show STD effect are annotated. On top the structure of the ligand and the epitope mapping (relative STD) are shown.



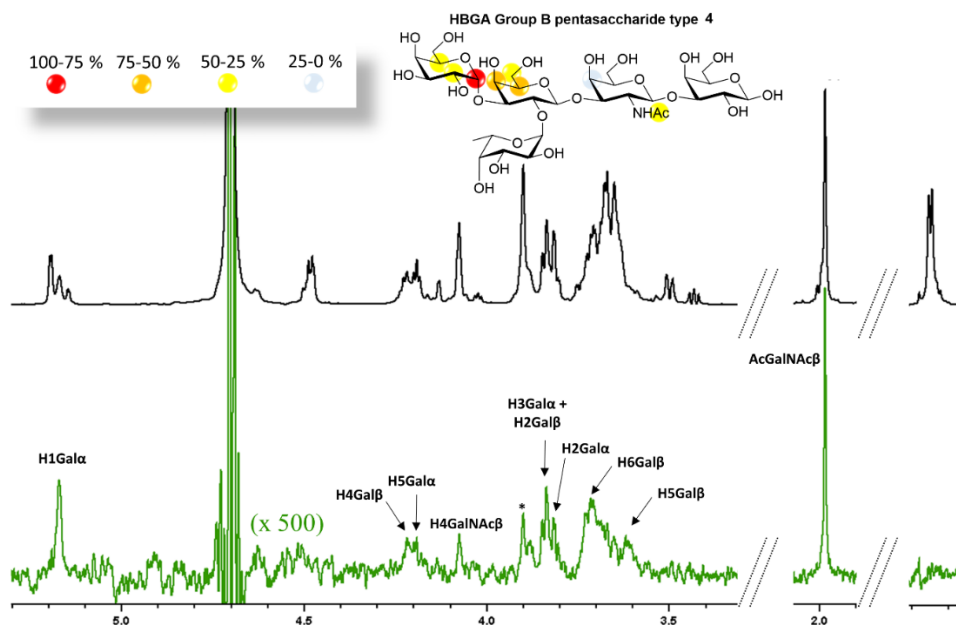
Supplementary Figure 22. ^1H -STD-NMR of the sample containing 50 μM of Gal-4C and 50 equivalents of A type 4 pentasaccharide. In black the off-resonance spectrum (reference spectrum). In green the STD spectrum. The protons of the ligand that show STD effect are annotated. On top the structure of the ligand and the epitope mapping (relative STD) are shown.



Supplementary Figure 23. ^1H -STD-NMR of the sample containing $50\ \mu\text{M}$ of Gal-4C and 50 equivalents of A type 6 tetrasaccharide. In black the off-resonance spectrum (reference spectrum). In green the STD spectrum. The protons of the ligand that show STD effect are annotated. On top the structure of the ligand and the epitope mapping (relative STD) are shown.



Supplementary Figure 24. ^1H -STD-NMR of the sample containing $50\ \mu\text{M}$ of Gal-4C and 50 equivalents of B type 2 tetrasaccharide. In black the off-resonance spectrum (reference spectrum). In green the STD spectrum. The protons of the ligand that show STD effect are annotated. On top the structure of the ligand and the epitope mapping (relative STD) are shown.



Supplementary Figure 25. ^1H -STD-NMR of the sample containing 50 μM of Gal-4C and 50 equivalents of B type 4 pentasaccharide. In black the off-resonance spectrum (reference spectrum). In green the STD spectrum. The protons of the ligand that show STD effect are annotated. On top the structure of the ligand and the epitope mapping (relative STD) are shown

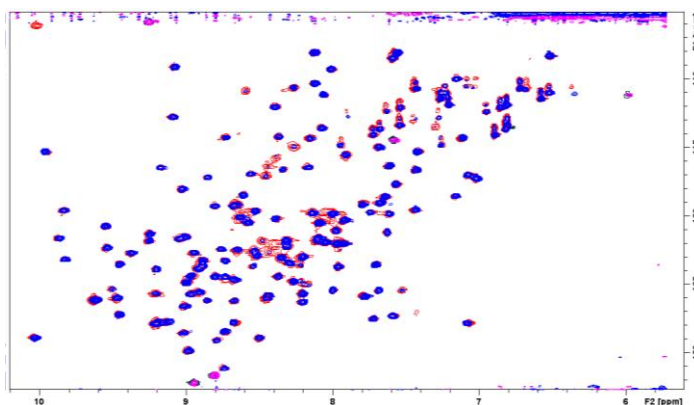
Supplementary Table 2. STD values for the complexes of Gal-4C and the different ligands upon aromatic irradiation

Ligand	^1H -residue	STD-AF	STD%
Lactose	H2Gal β	0.0012	40%
	H4Gal β	0.0024	80%
	H5Gal β	0.0028	93%
	H6Gal β	0.0030	100%
	H6Gal β	0.0027	90%
	H1Glc α	0.0017	57%
	H1Glc β	0.0009	30%
	H2Glc α	0.0008	27%
Galili B type 6	H2Glc β	0.0009	30%
	H1Gal α	0.0051	100%
	H2Gal α	0.0011	22%
	H3Gal α	0.0009	18%
	H4Gal α	0.0008	16%

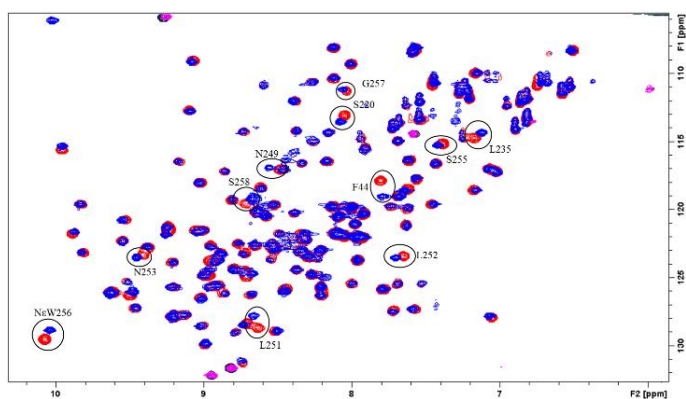
	H5Gal α	0.0013	25%
	H1Gal β	0.0019	37%
	H4Gal β	0.0020	39%
	H5Gal β	0.0024	47%
	H1Glc α	0.0011	22%
	H2Glc β	0.0013	25%
	H2Glc α	0.0010	20%
A type 1	H1GalNAc	0.0194	100%
	H2GalNAc	0.0104	54%
	H3GalNAc	0.0039	20%
	H4GalNAc	0.0019	10%
	NHAc GalNAc	0.0025	13%
	H2Gal β	0.0015	8%
	H3Gal β	0.0029	15%
	H4Gal β	0.0121	62%
	H5Gal β	0.0055	28%
	H6Gal β	0.0058	30%
	NHAc GlcNAc	0.0028	14%
	NHAc GlcNAc	0.0052	27%
	Average NHAc GlcNAc	0.0040	21%
A type 2	H1GalNAc	0.0162	100%
	H2GalNAc*	0.0067	41%
	H4GalNAc	0.0020	12%
	H5GalNAc	0.0095	59%
	NHAc	0.0030	19%
	H3Gal β	0.0033	20%
	H4Gal β *	0.0083	51%
	H5Gal β	0.0073	45%
	H6Gal β	0.0044	27%
	NHAc GlcNAc	0.0018	11%
A type 4	H1GalNAc α	0.0157	100%
	H2GalNAc α	0.0079	50%
	H4GalNAc α	0.0016	10%
	NHAc GalNAc α	0.0025	16%
	H2Gal β	0.0030	19%
	H4Gal β	0.0079	50%
	H5Gal β	0.0059	38%
	NHAc GalNAc	0.0018	11%

A type 6	H1GalNAc	0.0082	100%
	H2GalNAc	0.0043	52%
	H4GalNAc	0.0011	13%
	NHAc GalNAc	0.0025	30%
	H3Gal β	0.0024	29%
	H4Gal β	0.0043	52%
	H5Gal β	0.0047	57%
B type 2	H1Gal α	0.0102	100%
	H2Gal α	0.0021	21%
	H4Gal α	0.0016	16%
	H3Gal β	0.0026	25%
	H4Gal β	0.0056	55%
	H5Gal β	0.0052	51%
	NHAc GlcNAc	0.0027	26%
B type 4	H1Gal α	0.0062	100%
	H2Gal α	0.0016	26%
	H5Gal α	0.0018	29%
	H4Gal β	0.0044	71%
	H5Gal β	0.0032	52%
	H6Gal β	0.0025	40%
	H4GalNAc	0.0010	16%
	NHAc GalNAc	0.0026	42%

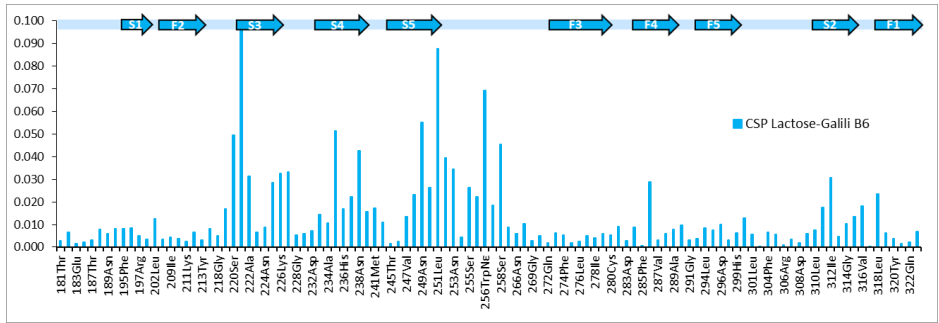
7.1.2.2 Chemical shift perturbations at Gal-4C



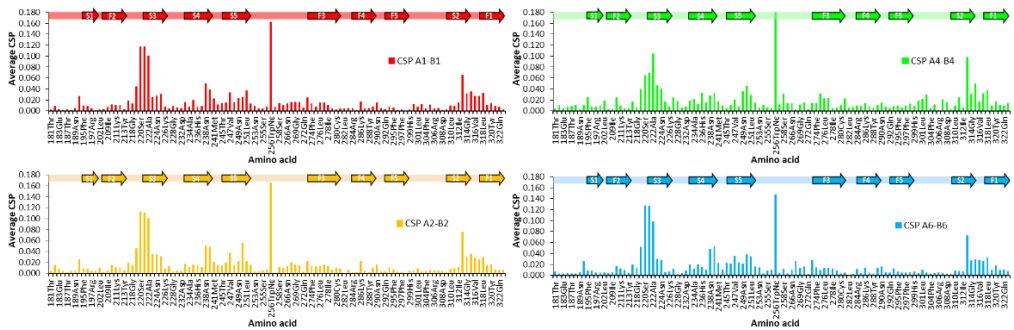
Supplementary Figure 3.26. Superimposition of the ^1H - ^{15}N HSQC spectra of Gal-4C in the presence of H type-6 (blue) and lactose (red).



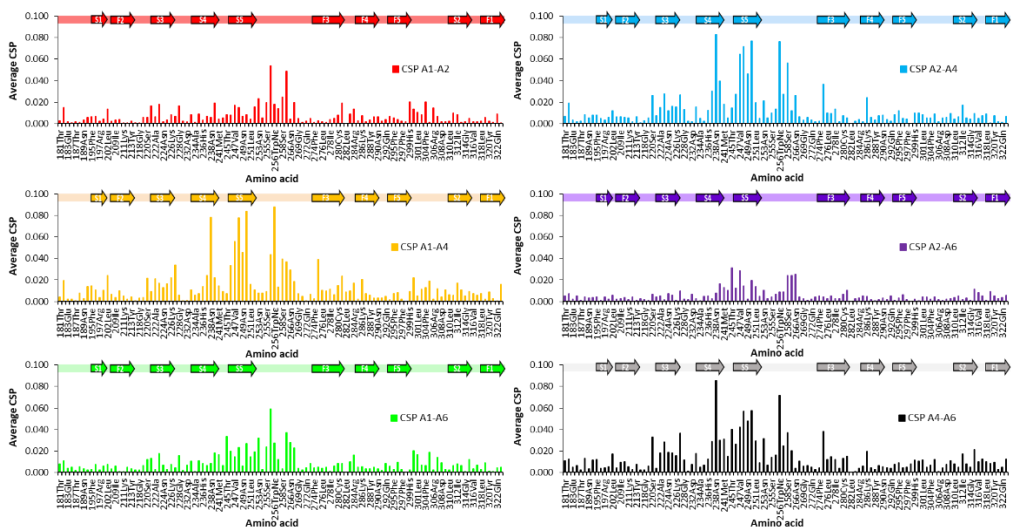
Supplementary Figure 3.27. Superimposition of the ^1H - ^{15}N HSQC spectra of Gal-4C in the presence of lactose (blue) and Galili linear trisaccharide (red).



Supplementary Figure 28. Chemical Shift Perturbation between Gal-4C in the presence of lactose and B type 6 linear trisaccharide.



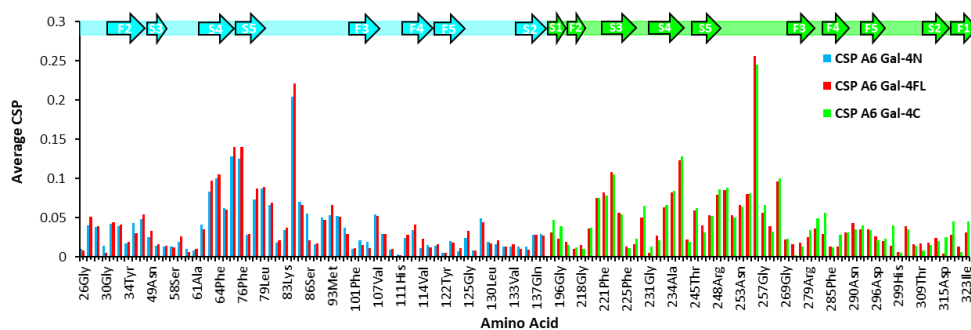
Supplementary Figure 29. Plots of the Chemical Shift Perturbation of Gal-4C in the presence of group A and group B oligosaccharides.



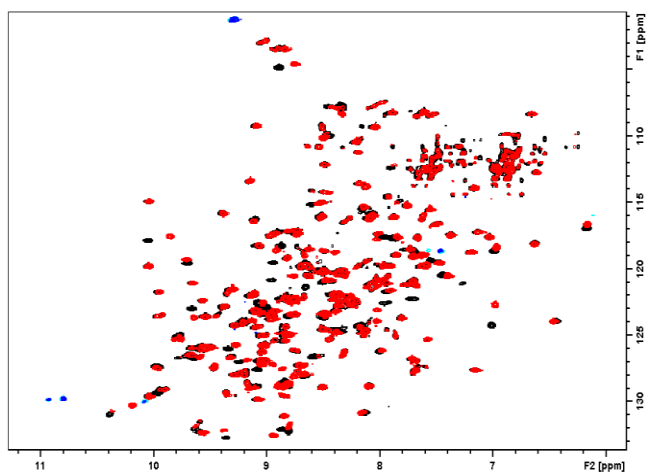
Supplementary Figure 30. Plots of the Chemical Shift Perturbation of Gal-4C in the presence of the different group A oligosaccharides.

7.1.3 Full-length Galectin-4

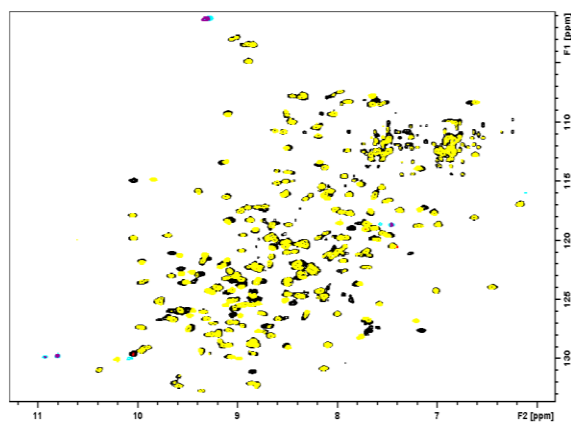
7.1.3.1 Chemical shift perturbations at Gal-4FL



Supplementary Figure 31. In red, the CSP of Gal-4FL in the presence of A type-6. In blue, the CSP of Gal-4N in the presence of A type-6. In green, the CSP of Gal-4C in the presence of A type-6.

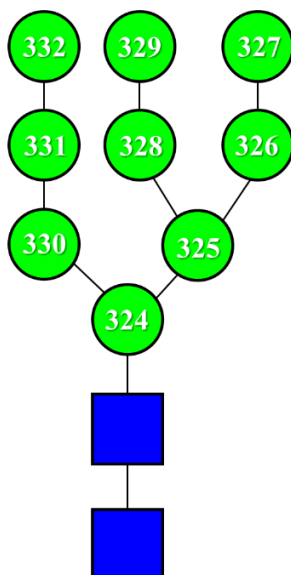


Supplementary Figure 32. In black the ¹H-¹⁵N-HSQC of Gal-4FL. In red the ¹H-¹⁵N-HSQC of Gal-4FL-H63R.



Supplementary Figure 3.33. In black the ^1H - ^{15}N -HSQC of Gal-4FL. In yellow the ^1H - ^{15}N -HSQC of Gal-4FL-H63R.

7.2 Towards understanding Galectin-4 interactions: the recognition of LPS and CD14

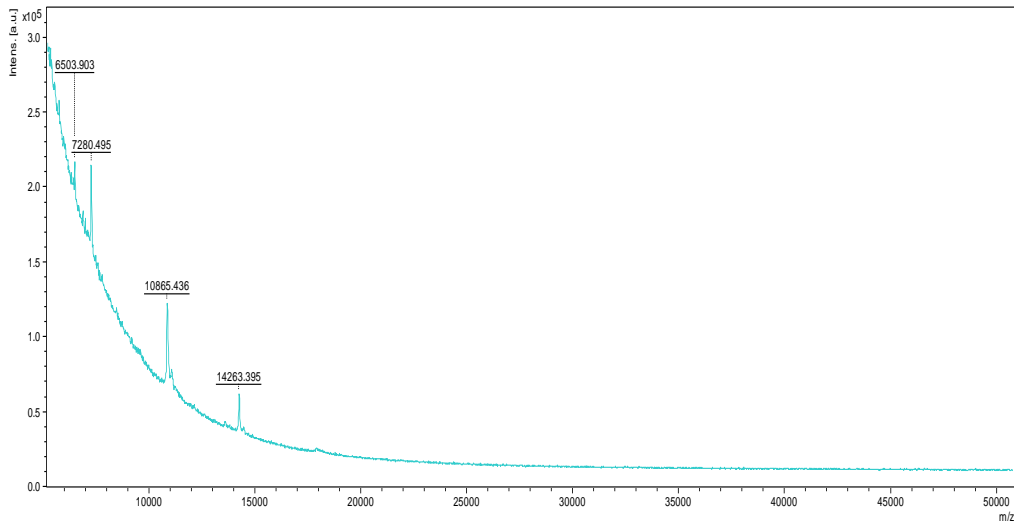


Supplementary Figure 34. Man₉ residue and the naming of each residue in the MD.

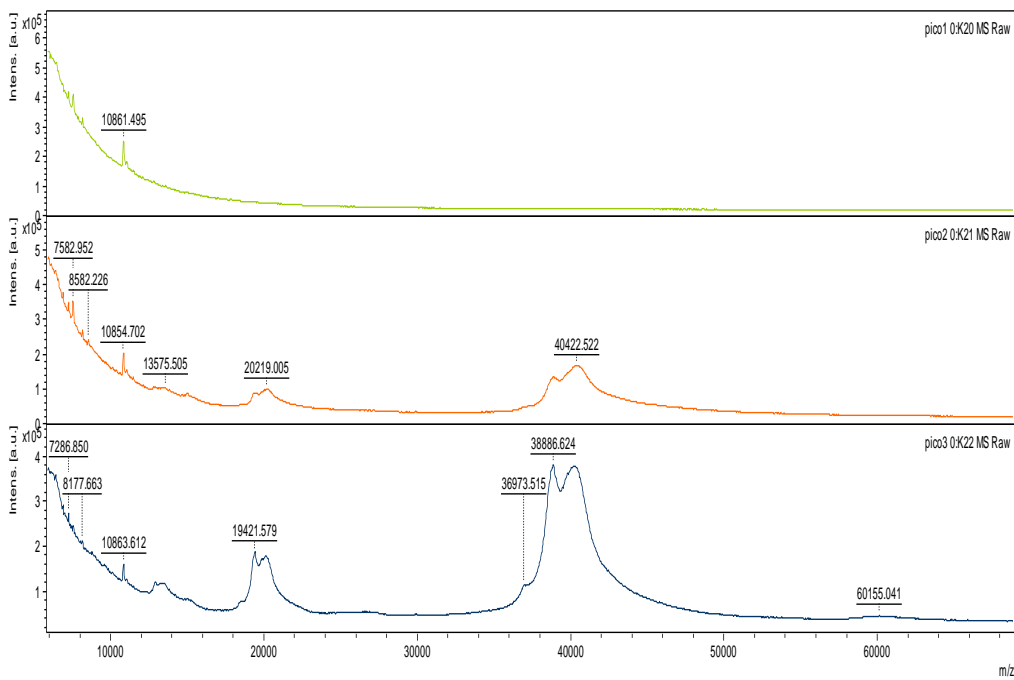
Supplementary Table 3. Predicted hydrogen bonding interactions in MD simulations between CD14 and the Man₉ residue at Asn282.

Acceptor	Donor	Fraction HB
GLU_33@OE2	0MA_329@H2O	15%
GLU_33@OE1	0MA_329@H2O	15%
GLU_33@OE1	0MA_329@H3O	11%
GLU_33@OE2	0MA_329@H3O	9%
GLU_31@OE1	0MA_329@H3O	9%
GLU_31@OE2	0MA_329@H3O	5%
GLU_99@OE1	0MA_329@H3O	7%
GLU_99@OE2	0MA_329@H3O	6%
GLU_31@OE2	0MA_329@H4O	19%
GLU_31@OE1	0MA_329@H4O	15%
ASP_3@OD1	0MA_329@H4O	8%
ASP_3@OD2	0MA_329@H4O	7%
GLU_99@OE2	0MA_329@H4O	6%
GLU_99@OE1	0MA_329@H4O	6%
VMA_325@O4	0MA_329@H6O	7%

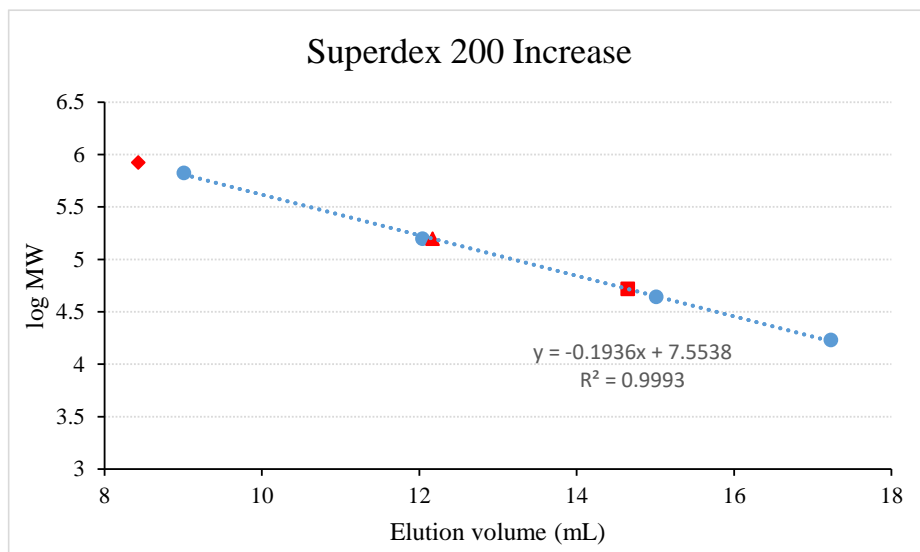
VMB_324@O2	0MA_329@H6O	5%
ASP_3@OD2	0MA_329@H6O	10%
ASP_4@OD2	0MA_332@H2O	6%
GLU_31@OE2	0MA_332@H2O	35%
GLU_31@OE1	0MA_332@H2O	34%
ASP_4@OD1	0MA_332@H2O	9%
0MA_332@O2	ARG_8@HH21	10%
GLU_31@OE2	0MA_332@H3O	14%
GLU_31@OE1	0MA_332@H3O	13%
ASP_4@OD2	0MA_332@H3O	19%
ASP_4@OD1	0MA_332@H3O	17%
ASP_3@OD1	0MA_332@H3O	7%
0MA_332@O3	ARG_8@HH21	11%
0MA_332@O3	ASP_4@H	9%
ASP_3@OD1	0MA_332@H4O	21%
ASP_3@OD2	0MA_332@H4O	11%
LEU_2@O	0MA_332@H4O	9%
GLU_95@OE2	2MA_330@H3O	33%
GLU_95@OE1	2MA_330@H3O	20%
2MA_330@O3	ARG_123@HH11	8%
2MA_330@O3	ARG_68@HH21	7%
GLU_95@OE2	2MA_330@H4O	21%
GLU_95@OE1	2MA_330@H4O	33%
VMB_324@O4	2MA_331@H6O	24%
GLU_95@OE2	2MA_331@H3O	17%
GLU_95@OE1	2MA_331@H3O	16%
ASP_3@OD2	2MA_326@H6O	8%
ASP_3@OD1	2MA_326@H6O	5%
ASP_3@OD2	0MA_327@H2O	5%



Supplementary Figure 3.35. MALDI-TOF profile of the LPS from E. Coli O55



Supplementary Figure 3.36. MALDI-TOF profiles of the different peaks obtained from the gel-filtration of the mixture of LPS from E. Coli O55 and CD14.



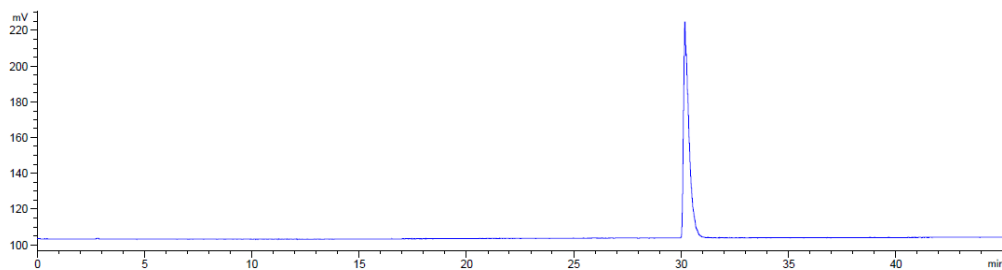
Supplementary Figure 3.37. Calibration curve obtained with an standard solution in a Superdex 200 Increase 10/300 column.

7.3 Synthesis of well-defined β -glucans using Automated Glycan Assembly

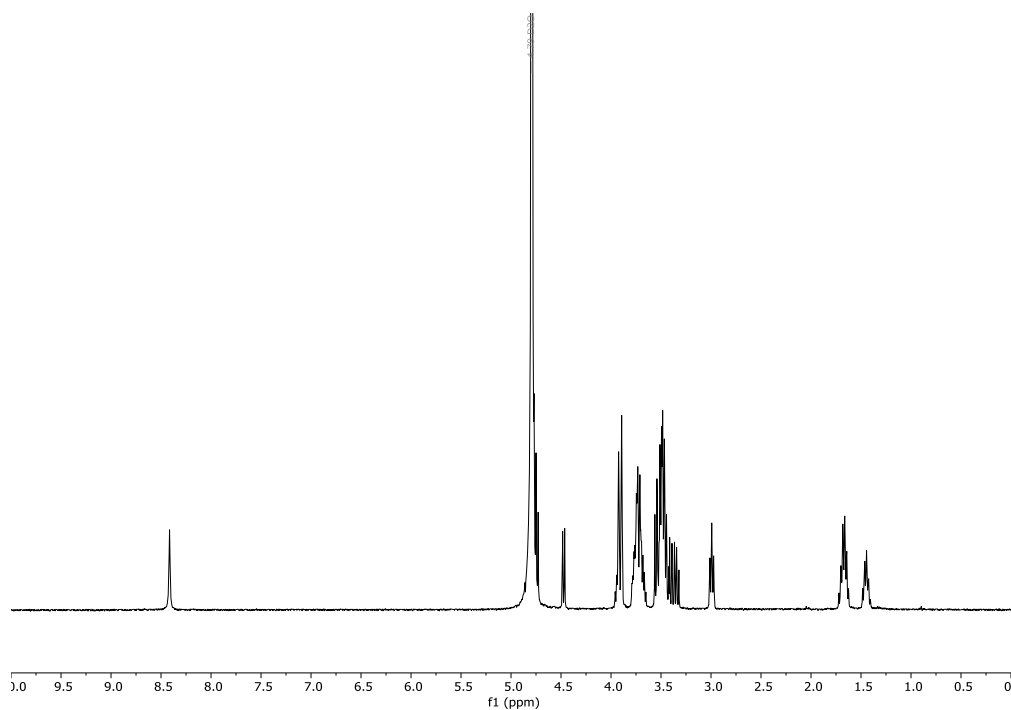
Synthesis of the branched tetrasaccharide (11)

Analytical data for tetrasaccharide **14**:

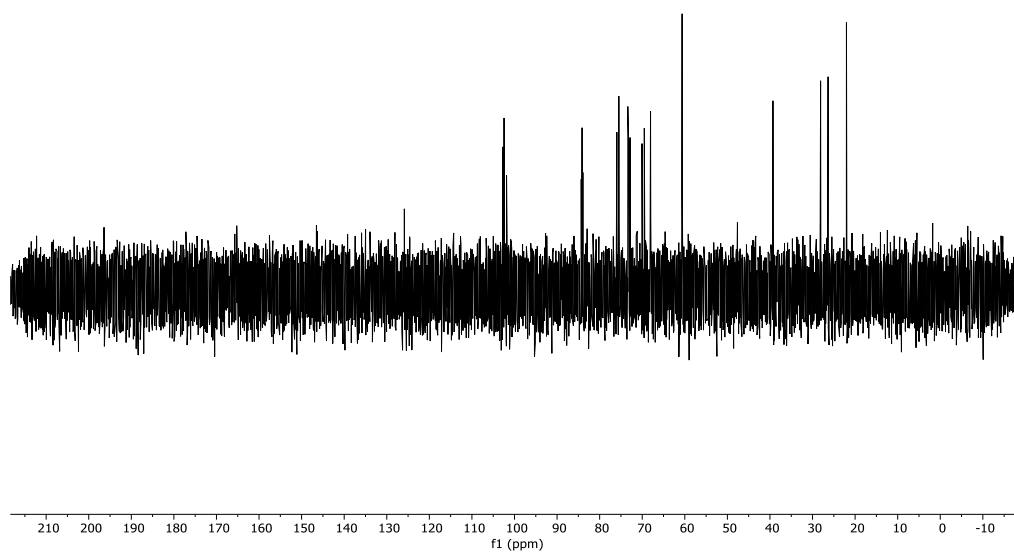
(ESI-HRMS) m/z 752.3176 $[M+Na]^+$ ($C_{29}H_{53}NO_{21}Na$ requires 774.3002).



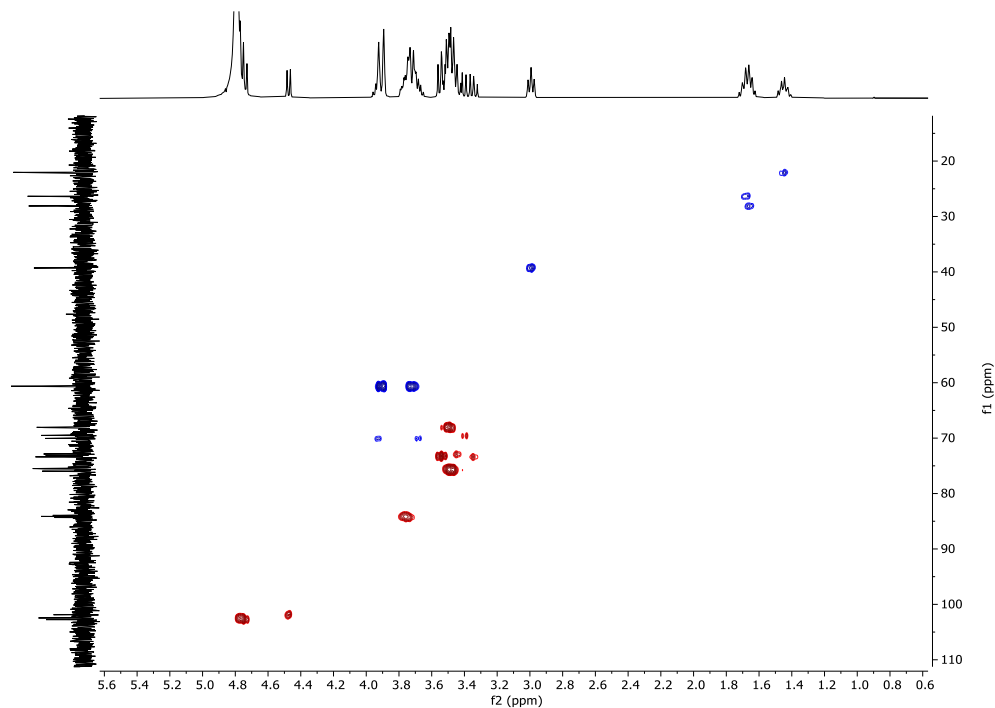
Supplementary Figure 38. RP-HPLC of the tetramer (ELSD trace, Method A1, $t_R = 13.4$ min)



Supplementary Figure 39. 1H -NMR of the linear tetrasaccharide (400 MHz, D_2O)

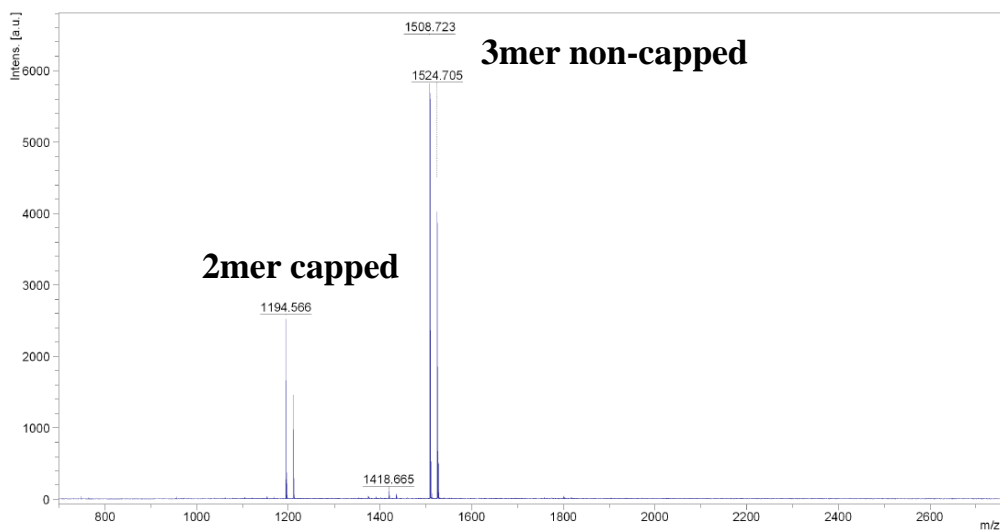


Supplementary Figure 40. ^{13}C -NMR of the linear tetrasaccharide (176 MHz, D_2O)



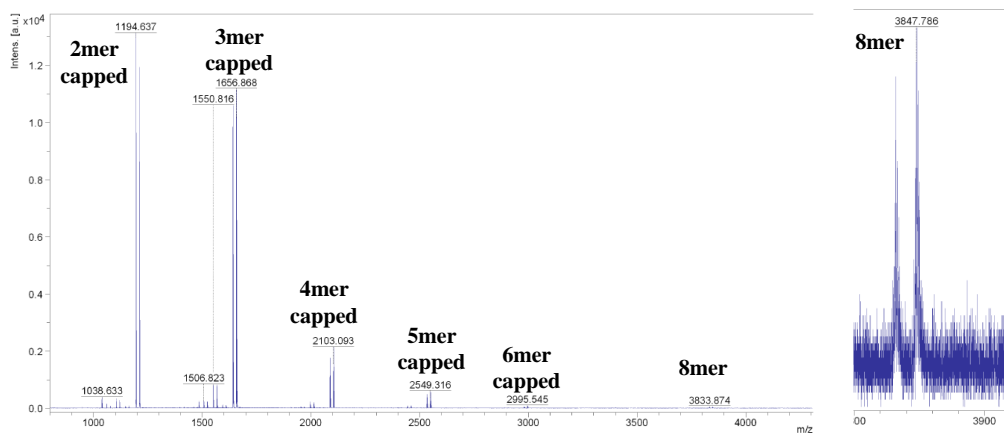
Supplementary Figure 41. ^1H - ^{13}C -HSQC of the linear tetrasaccharide

Synthesis of the branched pentasaccharide (12)



Supplementary Figure 42. MALDI-MS of the oligosaccharides cleaved from the resin in the synthesis of the branched pentasaccharide (12)

Synthesis of the linear octasaccharide (13)



Supplementary Figure 43. MALDI-MS of the oligosaccharides cleaved from the resin in the synthesis of the linear octasaccharide (13)