Carlos David Lourenço Lima

Licenciatura em Bioquímica

Deciphering the structural features of glycan-lectin interactions in the immune system

Dissertação para obtenção do Grau de Mestre em Bioquímica

Orientador: Doutora Filipa Margarida Barradas de Morais Marcelo, Investigadora Auxiliar, FCT-NOVA

Júri:

Presidente: Professor Doutor Pedro António de Brito Tavares
Arguente: Doutora Ana Luísa Moreira de Carvalho
Vogal: Doutora Filipa Margarida Barradas de Morais Marcelo

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

LOMBADA

Deciphering the structural features of glycan-lectin interactions in the immune system

Carlos Lima

2020

Carlos David Lourenço Lima

Licenciatura em Bioquímica

Deciphering the structural features of glycan-lectin interactions in the immune system

Dissertação para obtenção do Grau de Mestre em Bioquímica

Orientador: Doutora Filipa Margarida Barradas de Morais Marcelo, Investigadora Auxiliar, FCT-NOVA

Júri:

Presidente:
Arguente: Doutora Ana Luísa Moreira de Carvalho
Vogal: Doutora Filipa Margarida Barradas de Morais Marcelo

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Deciphering the structural features of glycan-lectin interactions in the immune system

Agradecimentos

Este ano foi recheado de altos e baixos, até porque houve uma pandemia mundial, mas não podia deixar de agradecer a toda a gente que tornou este ano melhor e memorável. Em primeiro lugar, quero agradecer à Doutora Filipa Marcelo por me ter aceite no grupo dela, me ter entregue um projeto que adorei do início ao fim e por me ter ensinado mais do que alguma vez esperei. Quero também agradecer ao Professor Eurico Cabrita por se ter disponibilizado e me ter encaminhado pelo que foi certamente a escolha certa para a minha tese de mestrado.

Quero agradecer também à Doutora Ana Luísa Carvalho e à Filipa Trovão do grupo X-tal na FCT-NOVA, ao Professor Jesús Jiménez-Barbero e à Doutora Ana Gimeno do CIC bioGUNE e ao Professor Francisco Corzana da Universidad de la Rioja pela colaboração que permitiu, não só a complementação dos resultados para esta tese, mas também o alcance de novos conhecimentos sobre o tema.

Agradeço também a todos os outros membros do grupo, primeiramente à Ana Diniz pela paciência e disponibilidade que teve para me ensinar ao longo deste ano, mas também pela simpatia e pelas conversas que me puseram à vontade e me fizeram sentir integrado desde o primeiro dia, à Helena Coelho por todas o conhecimento e pelos momentos de convívio e à Ana Sofia Grosso pela boa disposição.

Gostaria também de agradecer aos Doutores Jorge Dias, Ana Sofia Ferreira e Aldino Viegas, tal como aos alunos de Doutoramento Micael Silva e Sara Félix por toda a ajuda que me deram no laboratório e no tratamento de dados, mas também pelo ambiente de trabalho fantástico que criaram. Não podia deixar de agradecer aos meus colegas alunos de Mestrado Rodrigo Francisco, Débora Nunes e Philip O'Toole por este ano que, apesar de difícil, passámos unidos e cheios de bons momentos.

Quero agradecer a todos meus amigos que me ajudaram ao longo deste ano, em particular à Joana Dias por ter estado sempre presente e disponível para me ouvir, à Delfina Pereira que apesar de estar longe esteve sempre perto, ao Francisco Meda pelas noites de ramen, à Eva Carrasqueira que por muito tempo que estejamos sem falar, está sempre presente quando preciso, à Melissa Gama pelas conversas e disparates, aos meus amigos da Guarda por me perdoarem pelo tempo que não volto, à malta do Rodízio pelos momentos de desabafo e gargalhadas, ao Luís, à Adriana, à Filipa, ao Pedro, à Asiyah, à Soromenho, ao Padilha e ao Zé pela longa amizade e conselhos sábios e aos meus colegas de casa Thomas e Bruno por me aturarem nos bons e maus momentos, pelas conversas infindáveis e pelas fantásticas refeições.

Por fim, quero agradecer à minha família que sempre me acompanhou, mas especialmente ao meu irmão que esperava sempre que voltasse a casa e neste momento está ele mesmo a percorrer o caminho dele muito perto de mim e aos meus pais por perceberem todos os fins-de-semana seguidos que eu não ia a casa, por confiarem sempre em mim e por me terem dado a melhor educação e as melhores condições possíveis, um sincero obrigado!!

Abstract

This thesis is focused on the study of the molecular recognition of two cancer associated glycan epitopes, the LacdiNAc (GalNAcβ1-4GlcNAc) and the LacNAc (Galβ1-4GlcNAc) structures by two immune-related lectins, the human macrophage galactose-type lectin (MGL) and human galectin-3 (Gal-3) through a multidisciplinary approach that combines NMR spectroscopy, isothermal titration calorimetry (ITC), molecular dynamics (MD) and X-ray crystallography.

In the case of the MGL, the K_D of MGL/LacdiNAc complex is at low μM range (5.3 μM) and similar that obtain for the MGL/α-Me-GalNAc complex, while MGL/LacNAc complex is at the mM range (higher than 1 mM). STD-NMR binding studies demonstrate that MGL preferentially binds LacdiNAc and LacNAc structures, through the non-reducing end of the disaccharides. The $^{1}H,^{15}N$-HSQC based titrations show that both ligands perturb the same region of MGL, however, the magnitude of this perturbation is much higher in the case of LacdiNAc than LacNAc. Furthermore, significant differences in the chemical shift perturbations of Y236, H286 and D294 residues were observed. MD models show that these differences are intimately related with the specific engagement of the NHAc group of the GalNAc unit of LacdiNAc.

With respect of Gal-3 complexes both ligands bind Gal-3 with similar affinity (30-40 μM). Nevertheless, the $^{1}H,^{15}N$-HSQC based titration shows that additional residues are perturbed in β-strand S3 in the case of Gal-3/LacdiNAc complex. X-Ray structure of the complex Gal-3/LacdiNAc confirms modifications at this region of the protein showing that the LacdiNAc induces an alternate conformation for the side chain of R144. The R144B conformation, absent in the case of the complex with LacNAc, mediates a hydrogen-bond with the N-acetyl group of GalNAc of LacdiNAc, and increase the residence of a water molecule (W2) that in turn establish a hydrogen-bond network involving the N-acetyl group of GalNAc, the glycosidic bond of LacdiNAc and the R144B.

Keywords: Carbohydrate-lectin interactions; NMR Spectroscopy; MGL; Gal-3; Cancer.

Resumo

Esta tese está focada no estudo do reconhecimento molecular de dois epítopos de açúcares que estão associados ao cancro, o LacdiNAc (GalNAcβ1-4GlcNAc) e o LacNAc (Galβ1-4GlcNAc) por duas lectinas do sistema imunitário, a *human macrophage galactose-type lectin* (MGL) e a *human Galectin-3* (Gal-3) através de uma estratégia multidisciplinar que combina espetroscopia de RMN, calorimetria de titulação isotérmica (ITC), dinâmica molecular (MD) e cristalografia de raios-X.

No caso da MGL, o K_D do complexo MGL/LacdiNAc é na ordem dos baixos μM (5.3 μM), semelhante ao obtido para o complexo MGL/α-Me-GalNAc, enquanto que o complexo MGL/LacNAc é na ordem do mM (acima de 1mM). Os estudos de interação por STD-RMN demonstram que a MGL liga preferencialmente ao LacdiNAc e LacNAc através da unidade não redutora dos dissacáridos. As titulações baseadas em $^{1}H,^{15}N$-HSQC mostram que ambos os ligandos perturbam a mesma região da MGL, no entanto, a magnitude da perturbação é muito superior no caso do LacdiNAc que no do LacNAc. Além disso, foram observadas diferenças significativas nas perturbações de desvio químico de Y236, H286 e D294. Os modelos de MD atribuem essas diferenças à interação específica do grupo NHAc da unidade de GalNAc do LacdiNAc.

No que respeita aos complexos da Gal-3, ambos os ligandos têm afinidades semelhantes (30-40 μM). Ainda assim, as titulações baseadas em $^{1}H,^{15}N$-HSQC evidenciam que há resíduos adicionais perturbados na *β-strand* S3 no caso do complexo Gal-3/LacdiNAc. A estrutura de raios-X do complexo Gal-3/LacdiNAc confirma as modificações nesta região da proteína, mostrando que o LacdiNAc induz uma conformação alternada da cadeia lateral da R144. A conformação R144B, ausente no caso do complexo com LacNAc, media uma ponte de hidrogénio com o grupo N-acetil do GalNAc do LacdiNAc, e aumenta a residência de uma molécula de água (W2) que estabelece uma rede de pontes de hidrogénio que envolve o grupo N-acetil do GalNAc, a ligação glicosídica do LacdiNAc e a R144B.

Palavras-chave: Interações glicano-lectina; Espetroscopia de RMN; MGL; Gal-3; Cancro.

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

$\Delta\delta_{Comb}$ – Combined chemical shift difference

$\Delta\delta_{max}$ - Maximum combined chemical shift difference

Å - Ångstrom

Ala (A) - Alanine

Arg (R) - Arginine

ASGR - Asialoglycoprotein receptor

Asn (N) - Asparagine

Asp (D) - Aspartic Acid

CRD - Carbohydrate recognition domain

CSP - Chemical shift perturbation

Cys (C) - Cysteine

Da - Dalton

DC - Dendritic cell

DTT - Dithiothreitol

E. coli - *Escherichia coli*

ER - Endoplasmatic reticulum

Fuc - Fucose

Gal - Galactose

Gal-1 - Galectin-1

Gal-3 - Galectin-3

GalNAc - N-acetylgalactosamine

Glc - Glucose

GlcNAc - N-acetylglucosamine

Gln (Q) - Glutamine

Glu (E) - Glutamic acid

Gly (G) - Glycine

GT - Glycosyltransferase

HEPES - 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

His (H) - Histidine

HSQC - Heteronuclear single quantum coherence

Hz - Hertz

I_0 - Signal intensity in the absence of any enhancement

Ile (I) - Isoleucine

IPTG - Isopropyl β-D-1-thiogalactopyranoside

I_{sat} - Signal intensity in the presence of saturation

I_{STD} - Difference of intensity signal in the presence and absence of saturation

ITC - Isothermal titration calorimetry

K - Kelvin

K_A - Association constant

K_D - Dissociation constant

kex - Chemical exchange

k_{off} - Inverse proportion to the lifetime of the bound complex before dissociation

k_{on} - Probability of the free protein and ligand binding to form a complex

LB - Luria-Bertani

Leu (L) - Leucine

Lys (K) - Lysine

Man - Mannose

MD - Molecular dynamics

Met (M) - Methionine

MGL - Macrophage Galactose-type Lectin

Neu5Ac - N-acetylneuraminic acid

NHAc - N-acetyl

NMR - Nuclear magnetic resonance

NOE - Nuclear Overhauser effect

NOESY - Nuclear Overhauser effect spectroscopy

OD - Optical density

PBS - Phosphate-Buffered Saline

PDB – Protein data bank

PEG - Polyethylene glycol

Phe (F) - Phenylalanine

ppm - Parts per million

Pro (P) - Proline

ROESY - Rotating frame Overhauser effect spectroscopy

rpm - Rotations per minute

Ser (S) - Serine

STD - Saturation transfer difference

T - Temperature

T2 – Transversal relaxation time

Thr (T) - Threonine

TR - Transfer

Trp (W) - Tryptophan

TSP - 2,2,3,3-tetradeutero-3-trimethylsilylpropionic acid

Tyr (Y) - Tyrosine

Val (V) - Valine

ΔG - Variation in Gibbs free energy

ΔH - Variation in enthalpy

ΔS - Variation in entropy

1. Introduction

1.1. Protein Glycosylation

Glycans are ubiquitous. Every living cell is covered with a glycans' layer called glycocalyx (Figure 1.1) that is involved in numerous biological processes, such as cell signalling, immune regulation, and cell transportation[1].

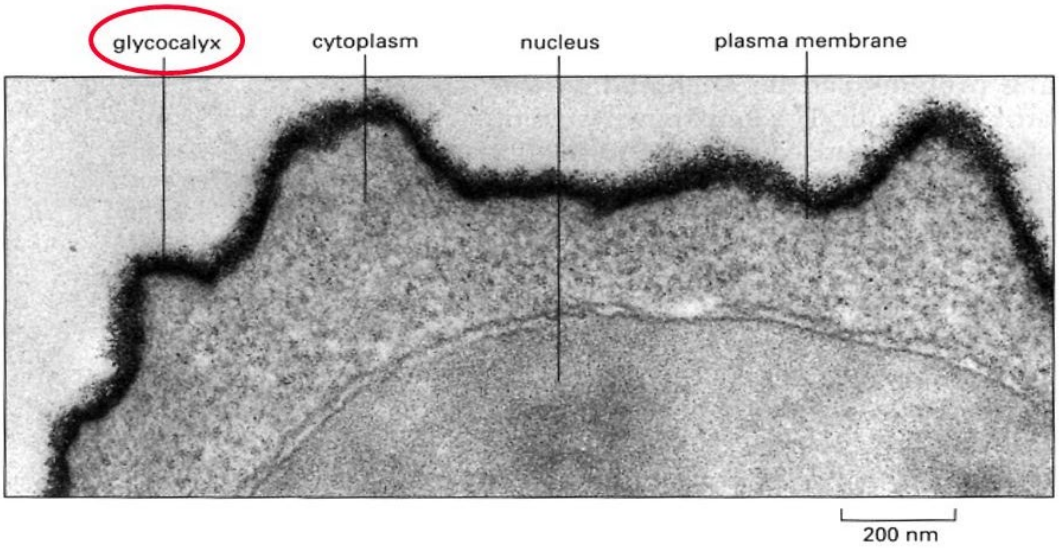

Figure 1.1 *- Electron micrograph of a human lymphocyte (Ruthenium Red staining)*[2].

Along with glycolipids, glycoproteins are major constituents of the glycocalyx. In glycoproteins, glycans are covalently attached to the protein through a process called glycosylation. Modifying the protein structure through glycosylation not only alters the protein's stability and solubility but can also change the function (bioactivity) of the protein. Glycans can be attached to proteins through a nitrogen (N-glycosylation) or through an oxygen atom (O-glycosylation)[3].

N-glycosylation is initiated in the endoplasmatic reticulum (ER) and ends in the Golgi apparatus and it consists of the β-1N linkage of an N-acetylglucosamine (GlcNAc) to the nitrogen atom of an asparagine residue at the consensus sequence motif of Asn-X-Ser/Thr, where X can be any amino acid except a proline[3,4]. O-glycosylation occurs on amino acids with functional hydroxyl groups, mostly serines and threonines, but also tyrosines. These residues are linked with N-acetylgalactosamine (GalNAc) and GlcNAc, and despite the absence of a specific consensus sequence, this type of glycosylation usually occurs in proline-rich sequences. The successive glycan units are as attached since the protein crosses the Golgi compartments. The process of glycosylation is catalysed by specific enzymes named glycosyltransferases (GTs). Most of GTs are associated with ER and Golgi membranes and many of them are expressed in

a tissue-specific regulated form. Importantly, the specificity of a determined GT is dictated by the donor and acceptor structures[5]. Therefore, the specificity, regulation and localization of GTs are essential to maintain the integrity of glycans on proteins.

Aberrant glycosylation is a well-established feature in cancer, strongly contributing to tumour progression and metastasis[6–8]. Indeed, tumour-associated glycans are considered key targets for the development of new diagnostic and therapeutic alternatives to fight cancer[6,9].

Mucins are glycoproteins that are heavily O-glycosylated, to a point in which a major part of the glycoprotein's weight consists of these O-linked oligosaccharides. The role of mucins is to control the interactions at the surface of the cell, including adhesion to the other cells. Overexpression of mucins in cancerous cells is known for many years, expressing this glycoprotein, and secreting it to the extracellular domain. Mucins in cancerous cells carry abnormal glycan epitopes. The modifications in the glycosylation pattern and the overexpression yield an alteration in the cell adhesion, causing it to detach from the primary tumour, migrate, and triggering metastasis[6].

Deregulation in the N-glycosylation process of proteins is also a hallmark of cancer. In fact, several glycoproteins carrying alteration on the N-glycans branches are actually used as biomarkers for diagnosis in many cancers[3,6]. When it comes to changes in motifs of N-glycans, some discern from others, such as the case of LacNAc (Galβ1 $\rightarrow$ 4GlcNAc) and LacdiNAc (GalNAcβ1 $\rightarrow$ 4GlcNAc). Even though LacNAc is present in healthy human cells and has physiological functions as a motif for ABO blood group antigens, it also plays a role in cancer due to its overexpression as poly-LacNAc structures, facilitating cancer cell-cell interactions and the formation of metastasis through interactions with lectins[10–12]. On the other hand, the terminal LacdiNAc is not usually present in human cells, however, it is specifically expressed in cancer cells, and it is strongly associated with cancer progression and immune invasion. It is also known that this disaccharide is highly expressed in parasitic helminths, such as schistosomes[13]. Figure 1.2 displays the structures of LacdiNAc and LacNAc epitopes and their location in nature. The specificity of the expression of this glycan in cancer and parasites might indicate that it is a potential diagnostic marker[13–16].

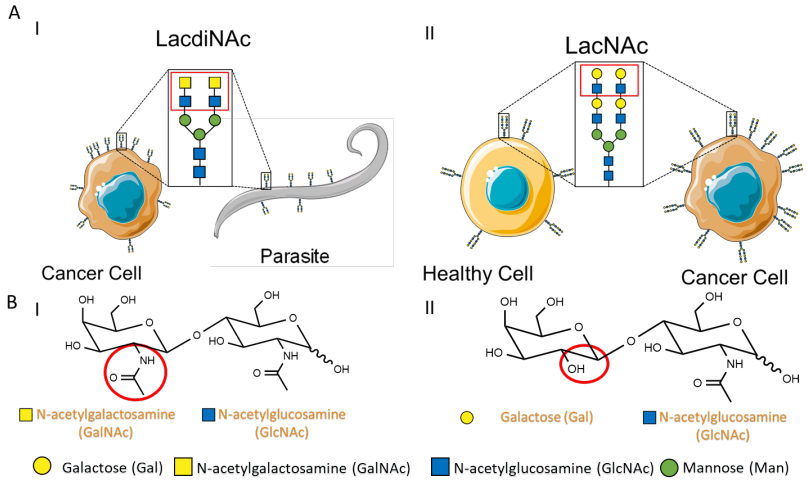

Figure 1.2 *- Schematic representation of N-glycans in different biological contexts.* ***A)*** *glycan structures containing* ***I:*** *LacdiNAc epitopes is a specific glycan epitope present in cancer cells and parasites;* ***II:*** *LacNAc epitope is present in healthy cells, however, it's overexpressed in cancer cells.* ***B)*** *Chemical structure and glycan nomenclature for* ***I:*** *LacdiNAc and* ***II:*** *LacNAc.*

1.2. Glycan-protein interactions

The glycome (sugar language or glycocode), emerges as the 3rd language of Life after genome and proteome[17]. The recognition of glycans by specific receptors translates the glycocode into key biological signals in physiological and pathological events (Figure 1.3)[18]. Unravelling the molecular recognition of glycans is imperial to understand the function of glycans by grasping cell processes and opens the door to develop alternative therapeutics based on glycans.

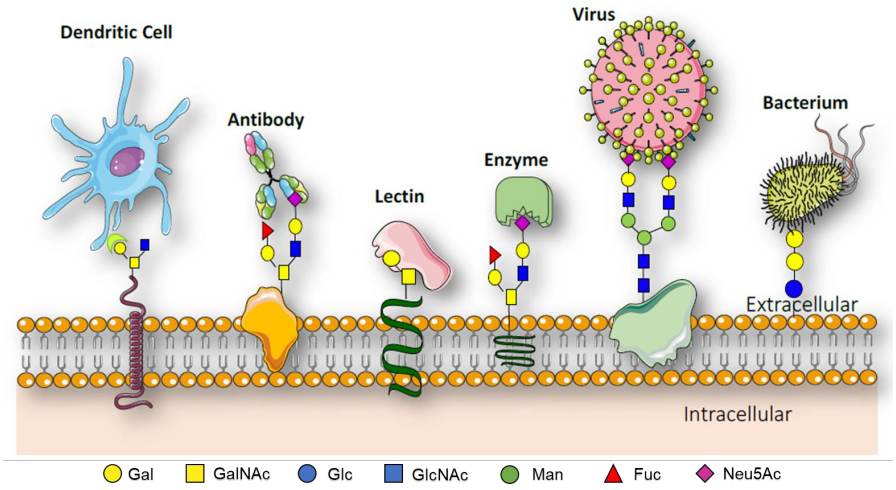

Figure 1.3 *- Examples of recognition glycans epitopes by distinct receptors. GalNAc: N-acetylgalactosamine; GlcNAc: N-acetylglucosamine; Neu5Ac: N-acetylneuraminic Acid; Gal: Galactose; Fuc: Fucose; Man: Mannose and Glu: Glucose.*

Most of the receptors that bind carbohydrates are very specific towards a limited amount of glycans and usually have individually low/moderate affinities (for 1:1 complexes) however, high affinities in a cellular context due to multivalence. Several and concerted weak interactions, such as electrostatic, hydrogen bonding, van der Waals interactions, stacking between the hydrophobic sugar faces and the aromatic and aliphatic amino acid chains and salt bridges with ions in more specific situations, take place in glycan-protein complexes[19,20]. All these interactions contribute favourably to the enthalpy of the system. Nevertheless, the complexation of a glycan with a protein is also characterized by a reduction in the conformational flexibility of the interacting carbohydrate units, through the decrease of the rotational and translational motions, which leads to a negative entropic contribution. To compensate this entropic penalty, a favourable entropy term, usually attributed to the release of water molecules from the binding site to the bulk solvent, balance the overall entropy of the system and strongly contributes to the complex formation.

Polar amino acids of proteins can form hydrogen bonds with the hydroxyl groups of carbohydrate units. The different positioning of the hydroxyl groups on each carbohydrate moiety makes hydrogen bonds the primary responsible for protein specificity. Also, the large amount of hydroxyl groups on carbohydrates and their ability to be both donors and receptors of electrons enables the formation of hydrogen bond interactions, which is advantageous to a high affinity complex[21]. There are three different types of hydrogen bonds: cooperative hydrogen bonds, bidentate hydrogen bonds and hydrogen bonds networks. The first one is the result of a hydroxyl group functioning both as a donor and acceptor of electrons. The second one happens when two adjacent hydroxyl groups from the carbohydrate with different atoms of the same aspartic or glutamic acid residue; The last one is formed by combining the previous two types of hydrogen bonds. This is possible since generally the carbohydrate-recognition domain of proteins have many residues with planar polar side chains that have at least two groups able to form all three types of hydrogen bonds[22]. Van der Waals contacts are also one of the most important interactions in a glycan-protein complex, occurring with nonpolar aromatic and aliphatic residues from the protein's binding site and the apolar face (C-H bond) of the carbohydrate unit[22]. Aromatic residues enable the formation of hydrophobic interactions, in particular, CH-π interactions, since the C-H bonds from glycans usually point towards aromatic systems on the protein binding site[23,24]. Even though CH-π interactions don't contribute much to the complex specificity, they are crucial for its stability contributing positively to the enthalpic term, by establishing non-conventional hydrogen bonds between the glycan C-H groups and the aromatic ring, and also to the entropic term, by shielding the apolar surfaces from the bulk water[23–25]. Both hydrogen bonds and CH-π bonds are the major contributors to the system enthalpy (Figure 1.4).

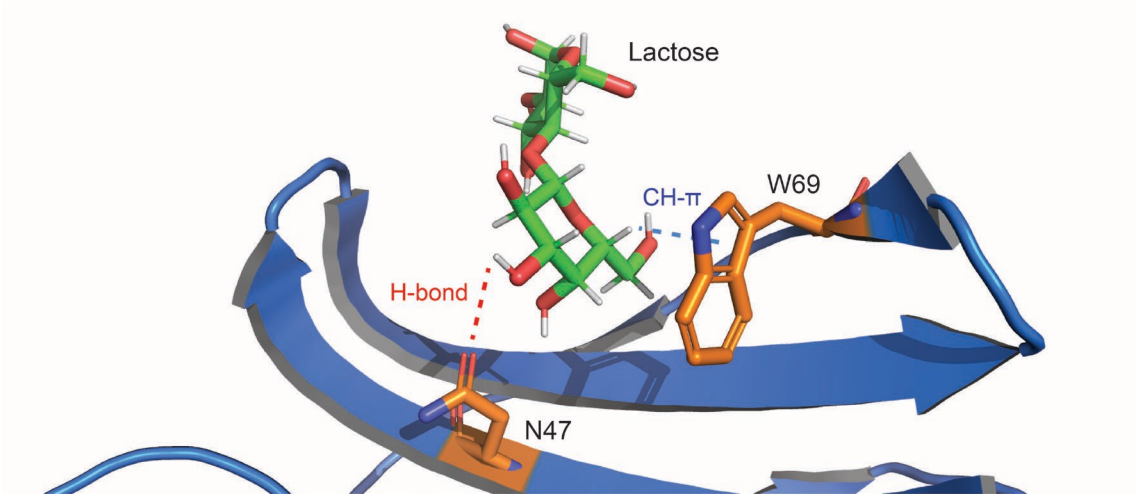

Figure 1.4 *- Representation of interactions that contribute positively to the enthalpy: Gal-1/Lactose complex (Protein Data Bank (PDB) code: 4GA9) close up where is displayed a lactose forming a hydrogen bond with the oxygen of the side chain of N47 (dashed red line) and a CH-π bond with the five-member aromatic ring of W69 (dashed blue line).*

The solvent also plays a role in the formation of these complexes, since it allows extra inter and intramolecular interactions that stabilize the complex due to the ability to be both a receptor and an acceptor of electrons[26]. Furthermore, water molecules found in the protein's binding site can also balance the entropy of the system, as previously discussed[27].

In a cellular context, the high affinity of protein-carbohydrate complexes is achieved based on multivalence effects (multiple binding epitopes interacting with multiple receptors). Multivalence effects include oligomerization of the receptors and/or multivalent presentation of the sugars, which allows ligand cross-linking.

Multivalency is the sum of individual different interactions that increase the binding affinity. These create modes of multivalent binding such as, the chelate effect, the receptor clustering, concentration effects, secondary binding sites, and the polyelectrolyte effect, which are inaccessible to monovalent ligands[26].

1.3. Lectins – Proteins that translate the Glycocode

Lectins are proteins that possess the ability to specifically recognize glycans, making them crucial to translate the glycocode into function. Lectins do not display enzymatic activity, instead lectins have functions such as regulation of cell adhesion, homeostasis and protein modulation. Several lectins play an important role in modulation of immune responses[9,28]. Lectins

are present in plants, animals and microorganisms, also each tissue expresses different lectins, as these can recognize different glycans and have different functions[28,29].

The recognition and binding to glycans, usually with high specificity, is the ability of the carbohydrate recognition domain (CRD) of the lectin. Lectins are capable to differentiate specific structures among a vast variety of complex structures, including ones that have the same minimal carbohydrate motif. These glycan-binding proteins also present high avidity towards specific sugars due to the recognition of another sugar unit in the glycan structure or to carbohydrate multiplicity presentation[30].

Lectins differ in their structure, function, glycan motif specificity and tissue where they are expressed, among others, which makes it difficult to establish classification criteria. For example, there are different classification criteria between plant and animal lectins. The animal lectin classification divides lectins into 15 different groups, as shown in the Figure 1.5. These categories are defined by their roles, location and structure. Each different class of these lectins is still divided into several subgroups, depending again on various factors, demonstrating the variety of lectins present in animal cells[29,30].

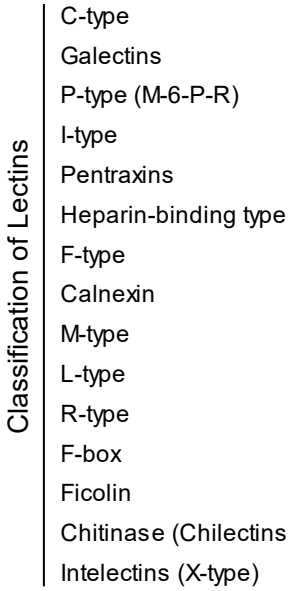

Figure 1.5 *- Classification of animal lectins*[30]*.*

1.3.1. Human Macrophage Galactose-Type Lectin (MGL)

Human Macrophage Galactose-type Lectin (MGL, CLEC10A, or CD301) belongs to a large family of lectins, the C-type lectins. This category of lectins is expressed in different types of leukocytes and is known for being responsible for multiple immune responses and the

regulation of homeostasis. C-type lectins bind to their sugars in a Ca^{2+} dependent manner, and this feature consists of the main distinction from other lectin classes[29,31].

C-type lectins englobe a huge and varied number of lectins. Thus, in order to group these proteins by structure and function, Drickamer in 1993[31] organized them into 7 different groups, a classification that was updated over the years, until it reached the present 17 groups of C-type lectins, as displayed in Figure 1.6. In particular, MGL belongs to the asialoglycoprotein and DC receptors group, more specifically, to the asialoglycoprotein receptor (ASGR) subgroup[31].

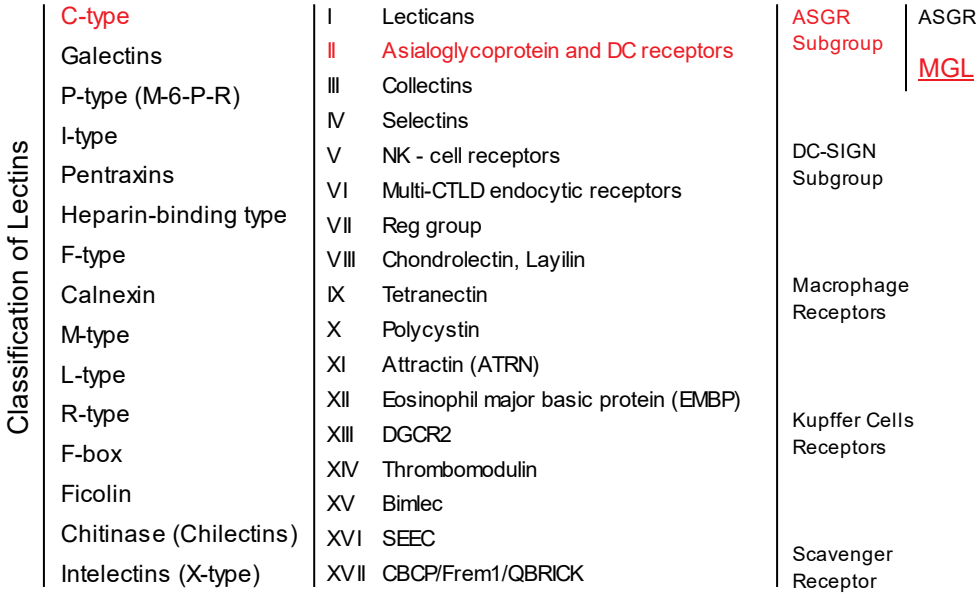

Figure 1.6 *- Classification of MGL accordingly to structure in the C-type lectins group.*

This attribution was done based on the common features between the structure of MGL and ASGR. Structurally, MGL (entry Q8IUN9-1 from UNIPROT), is composed of 316 amino acids, which are divided into a cytoplasmic domain (1-39), a transmembrane domain (40-60) and an extracellular domain composed of a tail region (61-180), and a calcium-dependent carbohydrate recognition domain (181-316, Figure 1.7-B)[32]. Additionally, the extracellular domain of MGL is presented as a trimer stabilized by the helical-coil tail region (Figure 1.7-A)[33,34]. The molecular weight of full length MGL is 35.4 kDa, while the molecular weight of the extracellular domain is 28.8 kDa and of its CRD is 15.6 kDa. Until nowadays, no X-ray high resolution structure was reported. Therefore, a 3D homology model was proposed based on the X-Ray structure of ASGR (Figure 1.7-C)[35,36].

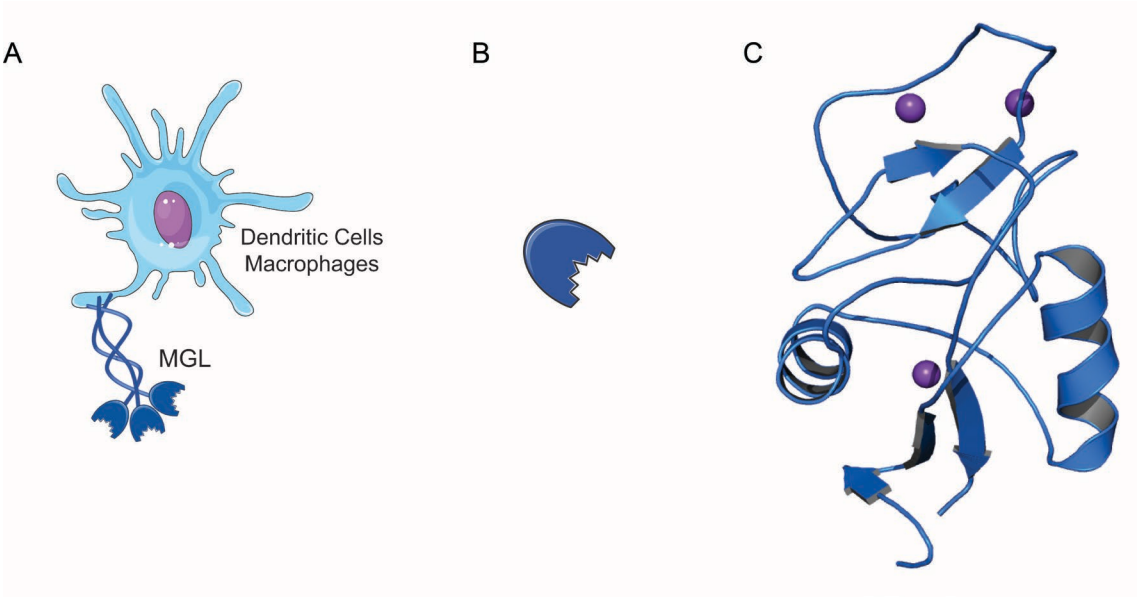

Figure 1.7 *- Location and structure of MGL* ***A)*** *MGL is expressed in dendritic cells and macrophages in the form of a trimer stabilized by the helical-coil tail region;* ***B)*** *schematic representation of the carbohydrate recognition domain (CRD) of MGL;* ***C)*** *3D homology model structure of MGL-CRD with the three Ca^{2+} ions presented as purple spheres created from X-Ray structure of ASGR (PDB code: 1DV8)*[35].

Based on the homology model, the MGL-CRD domain contains glutamine-proline-aspartic acid (Gln-Pro-Asp (QPD)) motif in the long loop region and the tryptophan-asparagine-aspartic acid (Trp-Asn-Asp (WND)) motif in the β4-strand which are responsible for Ca^{2+} coordination and ligand binding and characteristic from C-type lectins[29,36]. In terms of ligand recognition MGL presents high specificity towards galactose (Gal) and N-acetyl galactosamine (α- and β-GalNAc) epitopes[29,37]. Remarkable the availability and orientation of the equatorial/axial OH-3 and OH-4 groups of Gal/GalNAc residues are essential for Ca^{2+} interaction, and where extensions in OH-3 and OH-4 groups preclude MGL binding[29]. Nevertheless, additional contacts established between the 2-acetamido group (NHAc) and protein seems to determine the MGL binding preference in 100 times more for GalNAc (K_D ca. 10μM) over Gal (K_D ca. 1000μM)[36,37].

From a functional perspective, the human MGL is exclusively expressed in cells from the immune system, such as dendritic cells and macrophages[38,39] and plays a role in modulating immune responses in cancer and infections. It has tremendous physiological importance since it is able to activate dendritic cells for undertaking subsequent T cell downregulation or inducing apoptosis of these cells, even though this protein has a protective role in preventing tissue damage in inflammations and auto-immune diseases[38,40,41].

In cancer, MGL is the only lectin present on immune cells able to bind α-/β-GalNAc motif exclusively found in the N- and O-glycans chains of glycoproteins (e.g. mucins) expressed in tumour cells, such as the Tn (αGalNAc-Ser/Thr) and sialylated form of Tn-antigen (Neu5NAcα(2→6) GalNAcα1→O-Ser/Thr), as well as the LacdiNAc (GalNAcβ1-3GlcNAc) motif.

In the particular case of LacdiNAc epitope, studies have revealed that its expression *in vitro* is high on colon cancer cells, and it promotes increased tumour growth and metastasis *in vivo*[42]. Furthermore, other investigations have demonstrated the co-localization of tumour-associated MGL motifs and MGL expressing cells in colorectal cancer[7,43]. In fact, the interaction between tumour glycans and MGL seems to modulate immune responses, balancing tolerance vs. immunity. On the one hand, MGL binds Tn-structures in mucin glycoproteins (e.g. mucin-1, MUC1) and on self-glycoproteins (CD43/CD45) inducing suppressive anti-tumour immune responses[40,43,44]. On the other hand, MGL is capable to uptake Tn-derived peptide structures for antigen presentation inducing T cell responses and antibody production, thus prompting anti-tumour immunity[8]. Therefore, MGL is a potential target for cancer-related therapies[29,45].

1.3.2. Human Galectin-3 (Gal-3)

Galectins are associated with a variety of biological processes, including as immune functions, tissue organization, among others since these proteins are expressed in all types of cells and tissues. This family of lectins binds β-galactoside motifs through its carbohydrate recognition domain (CRD). Until nowadays, 15 different galectins were identified in vertebrates, and they were divided into three groups: prototype, which consists of monomers or homodimers with one CRD per subunit (Galectin-1, -2, -5, -7, -10, -11, -13 and -14); tandem-repeat, heterodimers containing two CRD connected by a linker peptide (Galectin-4, -6, -8, -9 and -12); and chimera-type, consisting solely of Galectin-3 (Gal-3) (Figure 1.8), due to being the only one that has two distinct domains, a flexible N-terminal domain and a CRD in C-terminal[46–48].

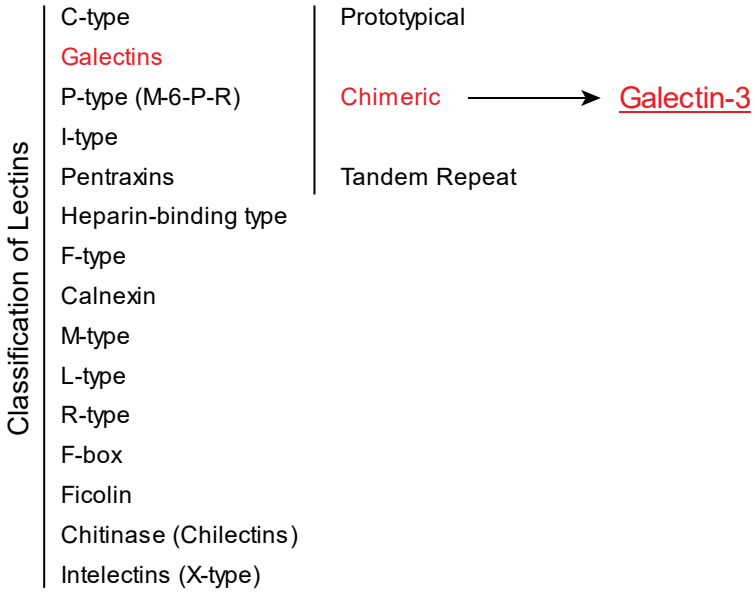

Figure 1.8 *- Classification of Galectin-3.*

Gal-3 is a very well-studied protein that is responsible for multiple biological functions. It is present in many different types of tissues and cell compartments, having distinct roles depending on its location, for example, in the cytosol, it interacts with the endosomal sorting complex required for transport; in the nucleus, it's involved in nuclear splicing of mRNA and gene expression; extracellularly, it modulates the interactions between cells such as cell fibrogenesis differentiation and inflammation processes. Due to its involvement in several biological functions, Gal-3 is directly implicated in obesity, cardiovascular and autoimmune diseases, as well in cancer, where it is overexpressed to control immune regulation, promote tumour cell adhesion, and it is strongly associated to metastasis. This makes Gal-3 a possible target for therapies, especially when it comes to cancer treatment[47,49].

When it comes to the structure, Gal-3 is composed of an unfolded N-terminal tail of approximately 115 amino acids responsible for multimerization of Gal-3 into a pentamer in presence of multivalent glycans (Figure 1.9-A) and a CRD consisting of approximately 135 amino acids responsible for sugar binding (Figure 1.9-B). Specifically, the N-terminal tail of Gal-3 is compared to collagen α1 (II) due to the homology of the sequence, that in the case of the lectin domain consists of a tandem repeat of 9 amino acids: Pro-Gly-Ala-Tyr-Pro-Gly-X-X-X (X cannot be charged or large side-chain hydrophobic residues). The CRD is displayed in two sets of five and six β-sheets arranged in a β-sandwich (Figure 1.9-C) and contains a Asp-Trp-Gly-Arg (NWGR) motif that is responsible for the binding of β-galactoside derivatives[46,48,50,51].

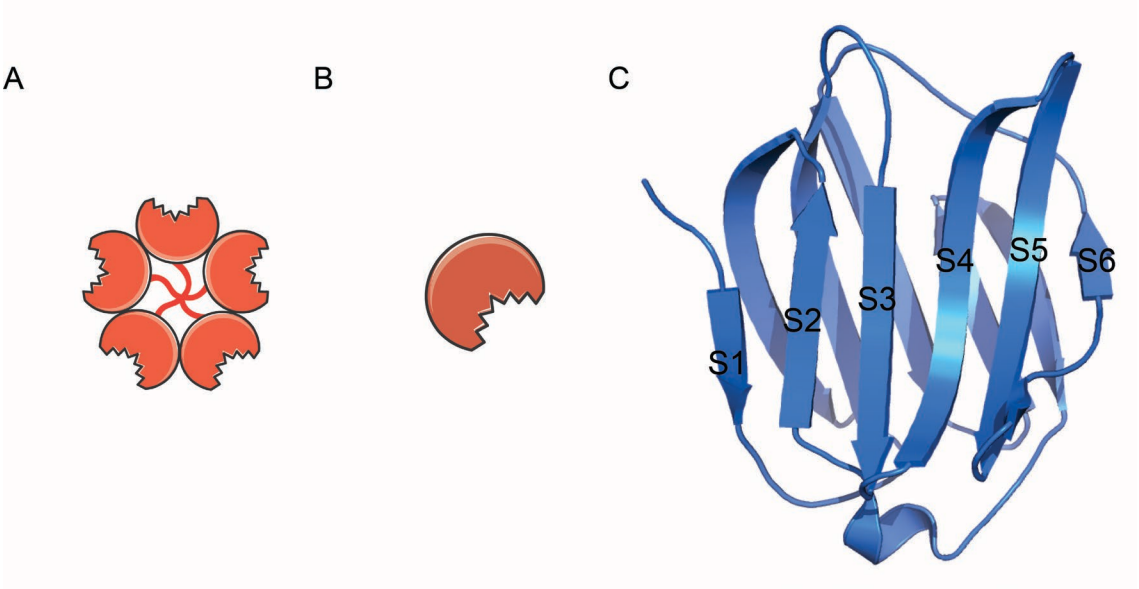

Figure 1.9 *- Representation of* ***A)*** *Gal-3 in the form of a pentamer due to the multimerization through the unfolded N-terminal tails of the protein;* ***B)*** *Gal-3 Carbohydrate Recognition Domain (CRD);* ***C)*** *X-ray Crystallography structure of Gal-3 CRD with labels for β-strands 1-6 (PDB code: 1KJL)[52]. The β-strands 1-6 constitutes the face of Gal-3 responsible for β-galactoside recognition.*

Even though Gal-3 CRD binds preferably to galactose units, it has a higher affinity towards glycans that have terminal galactose than just to the monosaccharide, for example, the disaccharides: lactose and LacNAc (N-acetyllactosamine), although LacNAc has a higher affinity than lactose[53]. The second sugar unit (GlcNAc in LacNAc and Glc in lactose) increases the stabilization of the complexes through the involvement of interactions with amino acids present between β-strands S4-S6[51].

1.4. Methodologies to investigate glycan-protein interactions

The investigation of the molecular determinants of protein-carbohydrate complexes can be unravelled by distinct techniques. The most used techniques to obtain structural information of glycan-protein complexes are X-ray crystallography and NMR spectroscopy. X-ray crystallography uses crystallized molecules and complexes to retrieve extremely precise structural data, however, depends on the molecules and complexes stability, and mostly on its ability to form crystals. Indeed, due to the intrinsic flexibility of glycans, especially in the case of more complex glycans (high number of sugar units), it is often difficult to obtain crystals due to the lower resolution of the glycans in some of these complexes. On the other hand, NMR spectroscopy in combination with molecular modelling offers the advantage to obtain information about dynamics of the glycans in its free and bound states and allows to investigate the structure, dynamics of the complex in solution (near to physiological conditions)[54,55]. Biophysical techniques can be also employed to complement the structural information obtained by X-Ray and/or NMR. In particular, isothermal titration calorimetry (ITC) measures are very useful and of paramount importance to obtain not only the affinity constants but also to infer the thermodynamic parameters[54,56]. Furthermore, due to the diversity of glycans present in nature, glycan microarrays technique has revolutionized the analysis of the specificity of several glycan binding proteins (lectins, antibodies), allowing to do a high-throughput screen of large amount of glycan's libraries[56–58]. In conclusion, the molecular recognition of glycans by their specific receptors is an intricate process, and the best way to do it is to combine all these techniques to guarantee a better chance to comprehend these mechanisms behind the translation of the glycocode[59].

Throughout this master thesis, a combination of some of the techniques was employed to obtain information on the glycan-protein complexes under study. Nevertheless, one technique was deeply used and had a prominent role during all this thesis: the NMR spectroscopy. Therefore, the following section of the introduction provides an overview of NMR binding methods in the context of the study of glycan-protein interactions.

1.4.1. NMR and glycan-protein complexes

The binding of a glycan to a glycan-binding receptor (lectin, antibody) relies on a biomolecular association reaction that usually follows a one-site binding model:

$$\mathrm{P} + \mathrm{L} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \mathrm{PL} \quad (1)$$

Equation (1) depends on two kinetic constants, k_{on} and k_{off}, where the bimolecular constant k_{on} represents the probability of the formation of the [PL] complex while the unimolecular constant k_{off} is the inverse of the lifetime of the bound complex prior to its dissociation[60].

The equilibrium dissociation constant can thus be expressed by equation (2):

$$\mathrm{K_D} = \frac{1}{\mathrm{K_A}} = \frac{[\mathrm{P}][\mathrm{L}]}{[\mathrm{PL}]} = \frac{k_{off}}{k_{on}} \quad (2)$$

The dissociation constant (K_D) is the inverse of the association constant (K_A), which means that the higher the affinity of a complex, lower is the K_D value. This depends on the two kinetic constants of the complex[60].

In the free states both glycan and receptor have their intrinsic NMR parameters (e.g. chemical shifts, relaxation rates, translational diffusion coefficients) (Figure 1.10). However, when we mix both in solution, their binding drives an exchange process that can toggle both sets of glycan and receptor between the free and bound states. In fact, in this process, several NMR parameters of the glycan and the receptor will be modulated by the interaction of the glycan to the receptor (Figure 1.10). The rate of the exchange between the free and bound states (slow, intermediate, and fast), will also influence the detection and evolution of the signals of the glycan and receptor free/bound forms. This exchange process is strongly dependent on the affinity of the complex and can be modulated by the temperature.

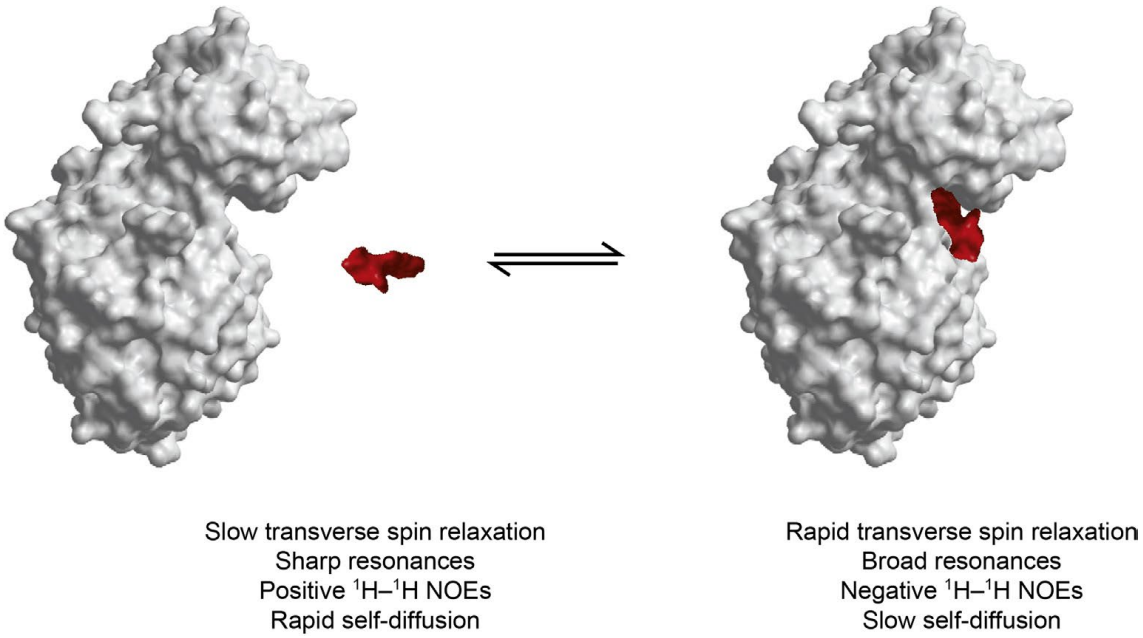

Figure 1.10 *- The difference of NMR parameters between the ligand in free state and when it is bound to a macromolecular receptor protein*[60]*.*

NMR methods can be divided into those detecting the receptor resonances (protein viewpoint) and those detecting ligand resonances (ligand viewpoint). To notice that the choice of the NMR binding method will be strongly dependent of the properties of the system under study, but in general, the protein detection methods are more adequate to probe moderate to high affinity complexes, while NMR binding methods based on ligand observation are more sensitive for moderate to low affinity complexes.

1.4.1.1. Protein viewpoint

The objective of this type of techniques is to observe the protein signals and the differences caused by the presence of the ligand, usually by detecting chemical shifts or through variations of linewidth of protein signals. Although very useful, these methods have some restrictions, such as the size of the protein, which should have at maximum 35 kDa, and it requires the isotopic labelling of the protein with ^{13}C and ^{15}N, due to the fact that these are the isotopes that produce NMR signal[20,59,61].

Out of all the protein-based NMR experiments, the $^{1}H,^{15}N$-HSQC is the most common experiment used due to its ability to produce a signal for each N-H bond. This is a heteronuclear 2D experiment that requires labelling the protein with ^{15}N. Since every amino acid, except proline, have one backbone N-H, it is possible to obtain a signal for almost every residue of the protein

which makes this experiment considered as the fingerprint of the protein. The signals are dispersed accordingly to the chemical environment of each N-H pair, depending on the vicinity and the folding of the protein[59,61].

It is important to understand that the chemical environment is what defines the position of the signal in the spectrum, which means that if something interferes with the chemical environment of some amino acid residues in a protein, thus it is expected differences for those amino acids in the spectrum, called chemical shifts. There are multiple reasons able to induce alterations on the chemical environment, such as alteration in pH and/or temperature, or the binding of a ligand. The formation of a protein-ligand complex depends on certain amino acids that are directly involved in the binding, this significantly alters the chemical environment of these specific residues, which can be observed through the $^{1}H,^{15}N$-HSQC spectrum of the protein. With the assignment of the protein backbone in hand, it is then possible to follow the chemical shift perturbation (CSP) caused by the presence of ligand for every protein residue[59,61].

The signals of the protein along the $^{1}H,^{15}N$-HSQC titration experiment will shift differently from the free to the bound state depending on the exchange rate between the free and bound state. It is possible to discriminate three exchange regimes (Figure 1.11)[61–63]: 1) the fast exchange regime on the chemical shift time scale ($k_{ex} >> \Delta\nu$), usually associated to high values of k_{off} and typically common of low affinity complexes (high K_D), and where it is possible to gradually follow the signal from free to bound state, appearing just an average cross peak along the gradual addition of the ligand during the titration; 2) the slow exchange regime on the chemical shift time scale ($k_{ex} << \Delta\nu$), often associated to low values of k_{off} and typically found on high affinity complexes (low K_D), where it is possible to observe, with the gradual addition of ligand, the progressive loss of intensity of the signal corresponding to the free state and, simultaneously, the gain of intensity of the signal corresponding to the bound state; and finally 3) the intermediate exchange regime on the chemical shift time scale ($k_{ex} \cong \Delta\nu$) , typically observed for intermediate affinity complexes, and where a progressive shift and disappearance of the signal is observed in simultaneous during the titration.

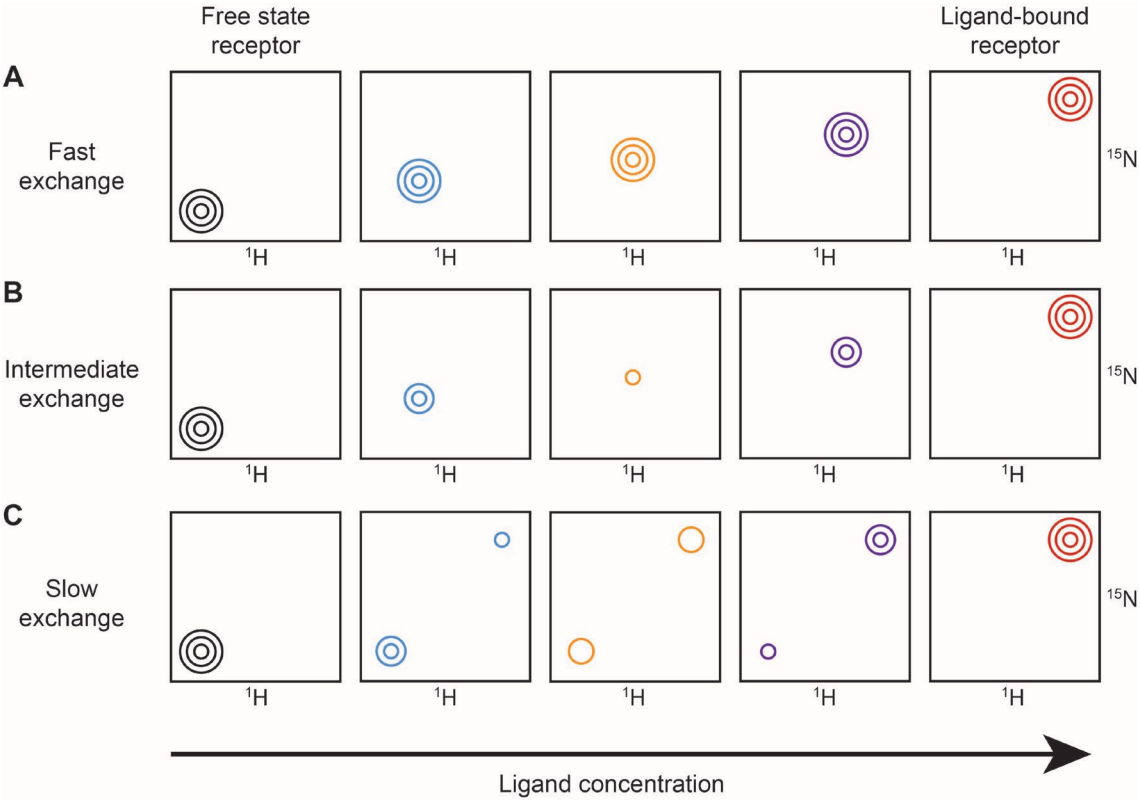

Figure 1.11 *- Representation of the three types of exchange in ^{1}H,^{15}N-HSQC titrations.* ***A.*** *Fast exchange;* ***B.*** *Intermediate exchange;* ***C.*** *Slow exchange.*

To analyse of the CSP caused by the presence of the ligand it is necessary to use an equation to weight the relative chemical shifts from the different nuclei. Usually, a combined chemical shift ($\Delta\delta_{Comb}$) caused by the ligand is calculated using the following equation[61,64,65]:

$$\Delta\delta_{Comb} = \sqrt{\Delta\delta_H{}^2 \times (0.102 \times \Delta\delta_N)^2} \quad (3)$$

In equation (3), $\Delta\delta_H$ represents the shift in the ^{1}H dimension, $\Delta\delta_N$ represents the shift in the ^{15}N dimension and 0.102 is the gyromagnetic ratio between ^{1}H and ^{15}N, which is the factor considered to balance the weight of the shift in the ^{15}N dimension[61,64,65].

Additionally, from the analysis of ^{1}H,^{15}N-HSQC-based titration it is possible to estimate the K_D, accordingly to the following equation[64,65]:

$$\Delta\delta_{Comb} = \Delta\delta_{max} \frac{(K_D + [L]_0 + [P]_0) - \sqrt{(K_D + [L]_0 + [P]_0)^2 - 4 \times ([L]_0 \times [P]_0)}}{2 \times [P]_0} \quad (4)$$

In the case of fast exchange regimes, it is possible to track a gradual shift upon each addition of ligand, so it is possible to determine the K_D of the complex by correlating the shift with the ligand concentration. In this equation, $\Delta\delta_{Comb}$ is the combined chemical shift defined by the

previous equation, $\Delta\delta_{max}$ is the maximum chemical shift between the free and bound signals, K_D is the dissociation constant, and $[L]_0$ and $[P]_0$ are the concentrations of ligand and protein, respectively. It is also possible to determine the K_D values in case of a slow exchange regime, where $\Delta\delta_{Comb}$ and $\Delta\delta_{max}$ are substituted by the volume of the bound signal and the maximum volume of the bound signal, respectively.

1.4.1.2. Ligand viewpoint

NMR binding methodologies based on ligand detection are particularly useful for moderate to high affinity complexes. The principle is that when a ligand is in fast and intermediate exchange with the receptor, it will be transiently adopting the characteristics of the receptor while bound. This includes the slow Brownian motion in solution thus, the ligand will experience shorter transverse relaxation time (T2) correlated with the longer correlation time of the receptor, which can preclude the detection of the ligand signals if long residence times of the ligand complexed with the receptor (typically observed for high affinity complexes) take place. Importantly, the ligand-based methods rely on direct observation of the free ligand in solution. The ligand-based NMR experiments[59] normally exploit: 1) differential Brownian motion of the ligand in the free vs bound form which leads to altered relaxation parameters and diffusion coefficients; 2) transfer of magnetization processes based on nuclear Overhauser effect (NOE) phenomenon. Briefly, the NOE phenomenon is a manifestation of cross-relaxation between two nuclear spins that are close to each other in space (typically < 5 Å).

The most popular methods based on the NOE effect are: 1) the saturation transfer difference (STD) NMR and 2) the transferred NOESY experiments.

1.4.1.2.1. Saturation transfer difference (STD) NMR

.The STD-NMR experiment is based on the transfer of saturation from the receptor to a bound ligand (through NOE effect) (Figure 1.12) [59,66–68]. From the experimental point of view, two ^{1}H-NMR spectra are recorded: 1) the on-resonance spectrum, where it is selectively saturated the protein. The frequency of saturation is normally set in the aromatic (around 7 ppm) or aliphatic region (0.5 to -1 ppm) of the protein. The choice of the frequency of saturation should be set far away from any ligand resonance; and 2) the off-resonance experiment considered the reference spectrum that will be used for comparison with the on-resonance spectrum. The off-resonance spectrum is recorded under the same conditions of the on-resonance experiment, but where the

frequency of saturation is set far away from any resonances (ligand and receptor) present in the system (typically 100 ppm). The off-resonance experiment yields a spectrum identical to the on-resonance except that the resonance intensities (I_0) are not affected by saturation. In the on-resonance experiment, only the resonances of the ligands that will receive saturation from the protein will experience a reduction in their resonance intensities (I_{sat}). With these two spectra in hand, a difference spectrum, called the STD spectrum ($I_{STD} = I_0 - I_{sat}$) is obtained and yields only those resonances that experienced saturation in the on-resonance experiment, the receptor resonances, and the ones from the binding ligands. Furthermore, the magnitude of STD intensities can be quantified as a fractional intensity (I_{STD}/I_0). The protons of the ligand with the highest STD intensity correspond to those that are closer to the protein protons. Therefore, it is possible to create an epitope mapping of the ligand, considering as 100% the proton with the highest STD intensity and calculating the others accordingly. STD enhancements are more effective in deuterated samples those only the non-exchangeable protons of the amino acids of the protein contribute to the transfer of saturation. There are two major restrictions to this technique: 1) the receptor must be large enough to experience self-diffusion properly and be able to transfer it to the ligand; and 2) it requires the ligand to be in fast exchange with the receptor (moderate to low affinity complex).

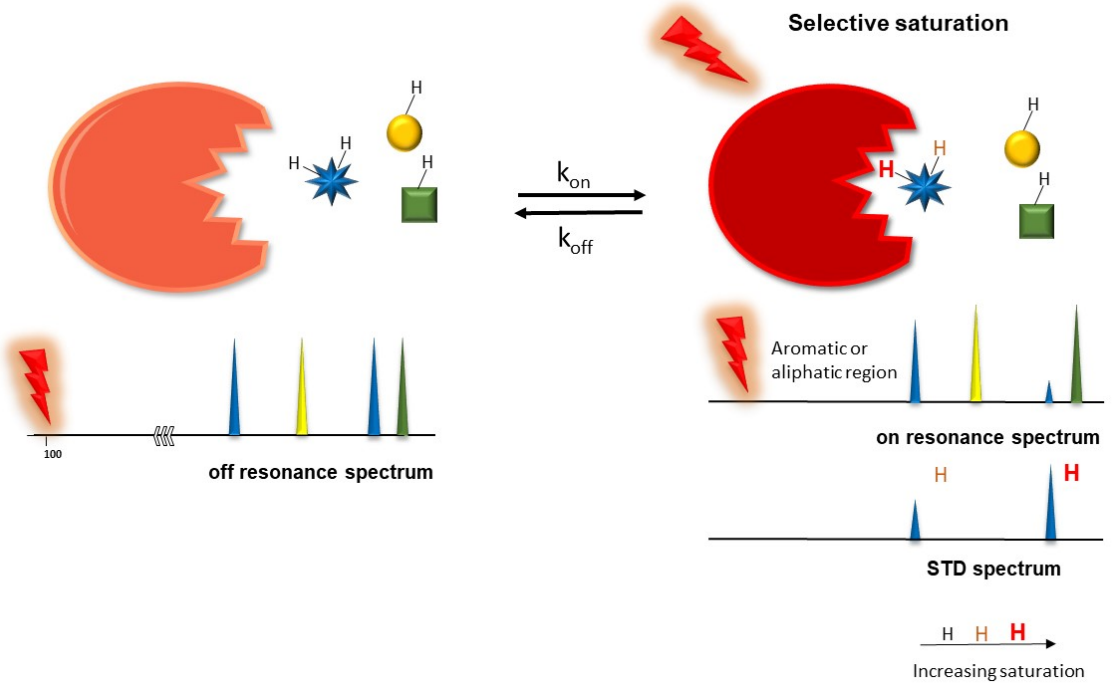

Figure 1.12 *- Schematic representation of the STD-NMR technique. The saturation of the protein is transferred to the ligand which causes differences in the on-resonance spectrum with respect to the off-resonance spectrum. These differences are accounted for in an STD spectrum to identify potential binders and to determine the STD-derived epitope mapping of the binder*[59].

1.4.1.2.2. Exchanged transferred NOE effects (transferred NOEs)

Transferred NOE (TR-NOE) experiments are based on the build of intramolecular NOEs between protons of the ligand while it is bound to the protein. These NOEs are further transferred to the free ligand state by chemical exchange, where they are detected. In bound state, the ligand adopts the molecular motion of the protein thus, the NOEs of the ligand generated in the bound state change from positive to negative (Figure 1.13). Once again, as STD-NMR, the transferred NOESY (TR-NOESY) technique relies on fast exchange between the free and bound states. In this way the strong and rapidly growing negative NOEs of the bound state dominate any weaker, slowly developing free ligand NOEs. Ligands with slow off-rate in the relaxation timescale will not produce detectable TR-NOE, since the relaxation of the ligand's signals will happen faster than the dissociation of the ligand from the protein, which precludes the detection. In a fast exchange regime in the chemical shift timescale, the TR-NOEs will appear at the same chemical shifts of the free ligand due to the excess of the ligand. By analysis of the TR-NOEs of the ligand, it is possible to infer the bioactive conformation of the ligand. This is especially important in a drug design and optimization process. Moreover, TR-NOESY is of special interest for glycans since they are intrinsic flexible molecules with multiple conformations in free state and commonly adopt some conformation upon binding.

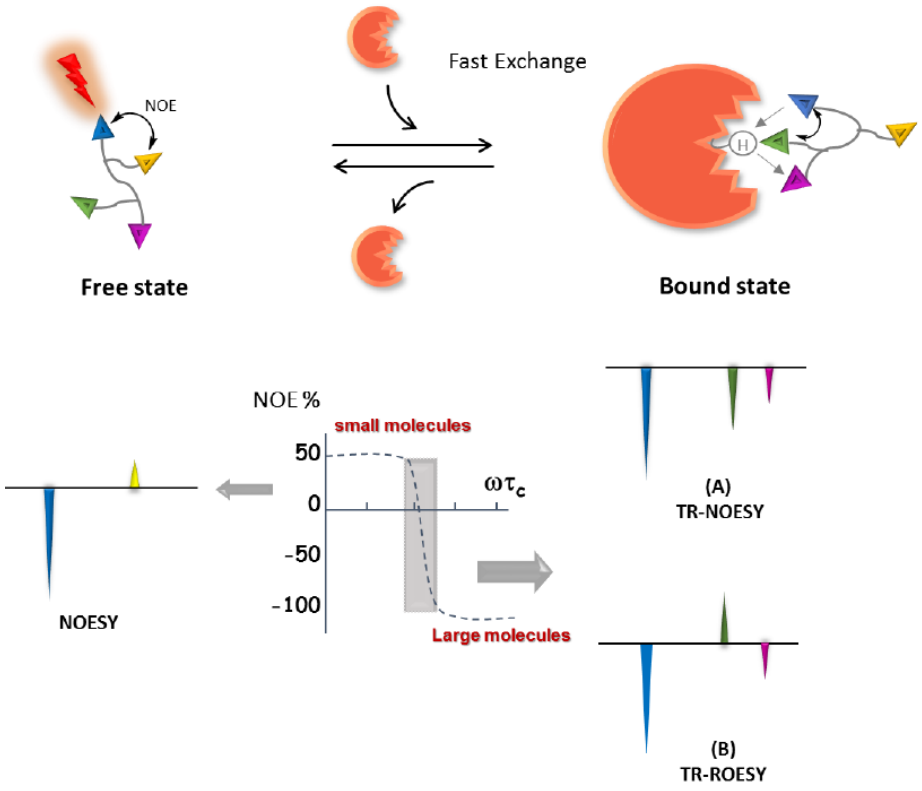

Figure 1.13 *- Schematic representation of the exchange-transferred NOE process for the protein-ligand analysis. When the ligand is free, since it is a small molecule, it has a positive NOE signal between its protons, while when it is bound to a macromolecule such as a protein, it gains the characteristics of a large*

molecule and has negative NOE signals (A). In the TR-ROESY experiment, the direct NOE effects are positive whereas the indirect are negative (B)[59].

An experimental drawback of TR-NOESY experiments concerns NOEs artifacts due to spin diffusion issues. In those cases, NOEs cross peaks appear between protons that are not close in space, however close to one or more protons within the ligand or at the receptor (indirect interactions). In case of suspicion of spin diffusion artifacts, transferred ROESY (TR-ROESY) experiments should be carried out. In the TR-ROESY spectrum, the direct NOE effects are positive while indirect NOEs are negative (Figure 1.13) allowing to discriminate between direct and indirect NOEs.

2. Objectives

The tumour-glycan/lectin interactions induce suppressive immune responses. By targeting these interactions with glycomimetics (small molecules) can in principle recover immune responses against cancer. In this perspective, the major goal of this thesis relies on the elucidation of the molecular determinants that govern the recognition of specific cancer-related glycan epitopes, the LacdiNAc and the LacNAc epitopes by two human lectins, the MGL and the Gal-3.

The study of glycan-protein interactions requires a multidisciplinary approach and the combination of different techniques. Therefore, the binding of the glycan epitopes LacdiNAc and LacNAc to MGL and Gal-3 lectins was scrutinized by the concerted application of distinct methodologies: NMR spectroscopy, X-Ray Crystallography, isothermal titration calorimetry (ITC) and molecular dynamics (MD) simulations.

3. Methods

3.1. Transformation of competent cells

The expression of recombinant proteins requires a transformation of competent cells to insert the wanted plasmid. The transformation for all proteins expressed in this work was the same, only differing the plasmid. It was added 1 µL of plasmid (0.5-1 µg in 20 µL) in 50 µL of *E. coli* BL21 (DE3) cells and incubated 15 minutes in ice. Then, the cells were incubated at 42ºC for 40 seconds and then back on ice for another 15 minutes, for the thermal shock. It was added 500 µL of sterile LB (Luria-Bertani) medium (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) and incubated for 1 hour at 37ºC in the shaker (Orbital Shaker-Incubator ES-20). The cells were recovered by centrifugation at 13500 rpm (Labogene scanspeed mini) for 1 minute. The excess of medium was removed and the pellet was resuspended in the remaining supernatant. The resulting cell suspension was spread in an LB-agar plate containing 100 µg/mL of ampicillin. The plate was incubated during approximately 16 hours at 37ºC.

3.2. Recombinant expression and purification of MGL

The expression and purification of MGL was based on reported work.[69] The gene of the carbohydrate recognition domain (CRD) of MGL (CLEC10A) was synthesized by NZYTech using a codon optimization strategy of *E. coli* expression. The synthesized region corresponds to residues C181-H316 (Appendix A) and it was subcloned into the expression vector pET21a by NZYTech. The protein was overexpressed in *E. coli* BL21(DE3) cells using LB medium supplemented with 100 µg/mL of ampicillin. After an OD between 0.6 and 0.8 was achieved, the induction of the expression of the protein, with 1 mM IPTG, was done at 25°C for approximately 16 hours. The cells were retrieved, by centrifugation during 15 minutes at 6000 rpm at 4°C (Rotor JA-10, Avanti J-26 XPI, Beckman Coulter), and resuspended in 50 mM Tris buffer pH 8.0, 150 mM NaCl and lysed by sonication with 10 cycles of ultrasounds that lasted 1 minute with 1 minute intervals between each one (UP100H/Tip MS7, Hielscher Ultrasonics; 50% cycle and 80% amplitude). MGL is expressed in inclusion bodies, therefore the lysate was centrifuged (Rotor f-34-6-38, Eppendorf centrifuge 5804R) for 15 minutes at 11000 rpm at 8°C, and the pellet was recovered. The inclusion bodies were washed with 50 mM Tris buffer pH 8.0, 150 mM NaCl, 0.5% TritonX-100, 10 mM EDTA twice and once with water. To recover the inclusion bodies between washes, they were centrifuged (Rotor f-34-6-38, Eppendorf centrifuge 5804R) for 15 minutes at 11000 rpm at 8°C. Then, the inclusion bodies were solubilized with 2 M NH_4OH for 3h30 and the refolding of the protein was done by dialysis against 10 mM Tris buffer pH 7.5, 75 mM NaCl, 20 mM $CaCl_2$ for approximately 16h. The refolded protein was centrifuged for 30 minutes at 19000 rpm at 8°C (Rotor JA-25.50, Avanti J-26 XPI, Beckman Coulter) and the supernatant was

recovered. MGL was purified by affinity chromatography using an α-Lactose agarose column (Sigma-Aldrich L7634) and it was eluted with 10 mM Tris buffer pH 7.5, 75 mM NaCl, 20 mM $CaCl_2$ with 150 mM of lactose. A desalting (HiTrap™ Desalting, GE Healthcare) step was performed, followed by dialysis against 10 mM Tris buffer pH 7.5, 75 mM NaCl, and 20 mM $CaCl_2$, in order to remove the lactose. The yield was approximately 4.3 mg/L of culture.

3.3. Recombinant expression and purification of ^{15}N labelled MGL

The expression procedure is similar to that used for the unlabelled MGL, however the protein was overexpressed in *E. coli* BL21(DE3) cells using $^{15}NH_4Cl$ as an exclusive nitrogen source in M9 minimal medium (50 mM Na_2HPO_4, 20 mM KH_2PO_4, 9 mM NaCl, 2.5 g/L $^{15}NH_4Cl$ as the sole nitrogen source, 4 g/L glucose, 2 mM $MgSO_4$, 10 mg/L thiamine HCl, 100 µM $CaCl_2$, 100 µM $FeSO_4$ and 100 µg/mL ampicillin) to isotopically label MGL with ^{15}N. The yield was approximately 2.6 mg/L of culture.

3.4. Recombinant expression and purification of Galectin-3

The expression and purification of this protein was based on reported work.[70] The gene of the carbohydrate recognition domain of Galectin-3 (LGALS3) was synthesized by NZYTech using a codon optimization strategy of *E. coli* expression. The synthesized region corresponds to residues L114-I250 (Appendix A) and it was subcloned into the expression vector pET21a by NZYTech. The protein was overexpressed in *E. coli* BL21(DE3) cells using LB medium (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) supplemented with 100 µg/mL of ampicillin. After an OD between 0.6 and 0.8 was achieved the induction of the expression of the protein, with 1 mM IPTG, was done at 37°C for approximately 4 hours. The cells were recovered by centrifugation at 6000 rpm for 15 minutes at 4°C (Rotor JA-10, Avanti J-26 XPI, Beckman Coulter) and resuspended 50 mM PBS buffer pH 6.9, 50 mM NaCl, 1 mM DTT, 0.1% NaN_3 and lysed by sonication with 10 cycles of ultrasounds that lasted 1 minute with 1 minute intervals between each one (UP100H/Tip MS7, Hielscher Ultrasonics; 50% cycle and 80% amplitude). The supernatant was retrieved after centrifuging for 30 minutes at 16000 rpm (Rotor JA25-50, Avanti J-26 XPI, Beckman Coulter) and Gal-3 was purified by affinity chromatography using an α-Lactose agarose column (Sigma-Aldrich L7634) and it was eluted with 50 mM PBS buffer pH 6.9, 50 mM NaCl, 1 mM DTT, 0.1% NaN_3 with 150 mM of lactose. A desalting (HiTrap™ Desalting, GE Healthcare)

was performed, followed by a dialysis against 50 mM PBS buffer pH 6.9, 50 mM NaCl, 1 mM DTT, 0.1% NaN_3, in order to remove the lactose. The yield was approximately 24.3 mg/L of culture.

3.5. Recombinant expression and purification of ^{15}N labelled Galectin-3

The expression procedure was similar to that for the unlabelled Gal-3, however the protein was overexpressed in *E. coli* BL21(DE3) cells using $^{15}NH_4Cl$ as an exclusive nitrogen source in M9 minimal medium (50 mM Na_2HPO_4, 20 mM KH_2PO_4, 9 mM NaCl, 2.5 g/L $^{15}NH_4Cl$ as the sole nitrogen source, 4 g/L glucose, 2 mM $MgSO_4$, 10 mg/L thiamine HCl, 100 µM $CaCl_2$, 100 µM $FeSO_4$ and 100 µg/mL ampicillin) to label Gal-3 CRD with ^{15}N.

3.6. Isothermal titration calorimetry (ITC) measurements

The ITC measurements were carried out in collaboration with the group of Prof. Jiménez-Barbero at CIC bioGUNE in Bilbao, Spain.

The ITC measurements were performed using MicroCal PEAQ-ITC calorimeter. Gal-3 in the presence of LacdiNAc and LacNAc were prepared in PBS 20 mM buffer containing 150 mM NaCl, and 10 mM β-mercaptoethanol at pH 6.5; while MGL-CRD in the presence of LacdiNAc/LacNAc were prepared in HEPES buffer 10 mM containing 20 mM $CaCl_2$ and 75 mM NaCl at pH 7.5. The concentration of the protein solution was set between 46–80 µM and that of the sugar stock between 1.5–15 mM. During the automated experiment, small amounts (3-5 uL) of the sugar solutions were titrated into a cell containing the protein solution and the heat dispersed was detected. Titrations were carried out in at least duplicate for all the ligands, and the errors are the standard deviations of the means of these replicates. The analysis of the curves was accomplished using the MicroCal Origin 7 software. The association constants and the thermodynamic parameters were obtained from the fit of the titration profile to a single-site binding model.

3.7. $^{1}H,^{15}N$ HSQC MGL titrations with LacdiNAc and LacNAc

The titration of ^{15}N-labelled MGL with LacdiNAc and LacNAc was monitored through the acquisition of $^{1}H,^{15}N$ HSQC spectra at 600 MHz (Bruker Avance III spectrometer equipped with a 5-mm inverse detection triple-resonance cryogenic probe head with z- gradients) and 293 K. The concentration of MGL-CRD was set to 200 µM in 10 mM Tris buffer, 75 mM NaCl, 20 mM $CaCl_2$ in H_2O:D_2O 90:10 at pH 7.5. The ligand was added to achieve the protein:ligand ratios of 1:0.25; 1:0.5; 1:1; 1:2 and 1:10 for LacdiNAc and 1:1; 1:5; 1:15; 1:35; 1:70; 1:100 and 1:150 for LacNAc. The pH was verified along the titration. The experiments were acquired at 293 K with 32 scans and 2048x128 points, with the spectral width being 9615.4 Hz centred at 2801 Hz for ^{1}H and 1946.2 Hz centred at 7175 Hz for ^{15}N. The central frequency for proton was set on the solvent signal (2801 Hz) and for nitrogen was set on the centre of the amide region (7175 Hz). The data was processed with Bruker TopSpin 4.0.9 and analysed with CcpNmr Analysis 2.4.2[71].

3.8. $^{1}H,^{15}N$ HSQC Gal-3 titration with LacdiNAc and LacNAc

The titration of ^{15}N-labelled Gal-3 with LacdiNAc and LacNAc was monitored through the acquisition of $^{1}H,^{15}N$ HSQC spectra at 600 MHz and 298 K. Two NMR tubes were prepared: 1) one with the protein containing an excess of the ligand in a protein:ligand ratio of 1:70 (tube 1) and 2) another with the protein in absence of the ligand (tube 2). Before titration, a $^{1}H,^{15}N$-HSQC spectrum is acquired of both tubes. The titration is performed by adding small volumes of tube 1 to tube 2 in order to achieve the protein:ligand ratios of 1:0.5; 1:1; 1:5; 1:15 1:35 and for LacdiNAc and 1:0.5; 1:1; 1:2; 1:5; 1:15 and 1:35 for LacNAc. The concentration of Gal-3 was set to 42 µM in 50 mM PBS buffer, 50 mM NaCl, 1 mM DTT, 0.1% NaN_3 in H_2O:D_2O 90:10 at pH 7.15 in both tubes. The experiments were acquired at 298 K with 32 scans and 2048x128 points, with the spectral width being 9615.4 Hz centred at 2791 Hz for ^{1}H and 1946.2 Hz centred at 7175 Hz for ^{15}N. The central frequency for proton was set on the solvent signal (2790 Hz) and for nitrogen was set on the centre of the amide region (7175 Hz). The data was processed with Bruker TopSpin 4.0.9 and analysed with CcpNmr Analysis 2.4.2[71].

3.9. Chemical Shift Perturbation (CSP) calculation

To determine the chemical shift perturbation (CSP) the following equation[64,65] was used

$$\Delta\delta_{Comb} = \sqrt{\Delta\delta_H{}^2 \times (0.102 \times \Delta\delta_N)^2} \quad (3)$$

In which $\Delta\delta_{Comb}$ is the combined chemical shift observed and the $\Delta\delta_H$ and $\Delta\delta_N$ are the chemical shifts of the amide hydrogen and nitrogen atoms in ppm, respectively, between the free and bound states of the proteins. The value of 0.102 is the ratio between the gyromagnetic constants of ^{15}N and ^{1}H. With the $\Delta\delta_{Comb}$ values obtained for each residue, a histogram that displays the $\Delta\delta_{Comb}$ for each amino acid was constructed. A cut-off line (0.06 ppm for MGL and 0.05 ppm for Gal-3) was used to distinguish the residues that strongly shifted due to the presence of the ligand.

3.10. K_D calculations

The K_D value was estimated by correlating the ligand concentration with the chemical shifts along the $^{1}H,^{15}N$-HSQC based titrations using the following equation[64,65]:

$$\Delta\delta_{Comb} = \Delta\delta_{max} \frac{(K_D + [L]_0 + [P]_0) - \sqrt{(K_D + [L]_0 + [P]_0)^2 - 4 \times ([L]_0 \times [P]_0)}}{2 \times [P]_0} \quad (4)$$

In which $\Delta\delta_{Comb}$ is the combined chemical shift, the $\Delta\delta_{max}$ is the maximum combined chemical shift, the K_D is the dissociation constant, the $[L]_0$ is the concentration of ligand and the $[P]_0$ is the concentration of protein.

Using equation (4), a calculated $\Delta\delta_{Comb}$ was determined for each step of the titration. It was also calculated a difference between the calculated $\Delta\delta_{Comb}$ and the $\Delta\delta_{Comb}$ obtained experimentally. The solver tool of excel was used to minimize this difference and estimate the K_D and $\Delta\delta_{max}$ values for each residue. In the case of the MGL-CRD/LacdiNAc titration, the volume of the bound signal for each concentration of ligand, and the maximum volume of the bound signal were used to substitute $\Delta\delta_{Comb}$ and $\Delta\delta_{max}$, respectively.

3.11. Carbohydrates

Carbohydrates were purchased from Carbosynth. The ^{1}H-NMR resonances of LacdiNAc and LacNAc were completely assigned through standard 2C-TOCSY (30 ms mixing time), 2D-NOESY (600 ms mixing time) and 2D-^{1}H,^{13}C-HSQC experiments. The concentration of the ligand was set to 2 mM for the heteronuclear experiments in D2O Tris(D11)-DCl buffer solutions with uncorrected pD 7.5. The assignment was accomplished at 298 K (Appendix B, Table A1). The resonance of 2,2,3,3-tetradeutero-3-trimethylsilylpropionic acid (TSP) was used as a chemical shift reference in the experiments (δ TSP = 0 ppm).

3.12. Saturation Transfer Difference (STD) NMR

For STD-NMR experiments, MGL-CRD was suspended in 10 mM perdeuterated Tris (D11)-DCl buffer (Sigma–Aldrich), 20 mM $CaCl_2$, 75 mM NaCl and 0.09% sodium azide at pH 7.5 in 100% D_2O. The experiments were performed with 450 μM of LacdiNAc/LacNAc and 22.5 μM of MGL CRD (ligand:protein molar ratio of 20:1) at 310 K and 293 K for LacdiNAc and LacNAc, respectively. Spectra were acquired with 1728 scans in a matrix with 64 K points in t2 in a spectral window of 12335.5 Hz centred at 2813.1 Hz. An excitation sculpting module with gradients was used to suppress the water proton signals. Selective saturation of the protein resonances (on resonance spectrum) was performed by irradiating at 7 ppm (aromatic residues) using a series of 40 Eburp2.1000-shaped 90° pulses (50 ms) for a total saturation time of 2 seconds. For the reference spectrum (off resonance), the samples were irradiated at 100 ppm. The STD spectra were obtained by subtracting the on-resonance spectrum to the off-resonance spectrum. Then, the percentages of absolute spin saturation (% STD) were calculated by comparing the intensity of the signals in the STD spectrum with the signal intensities of the reference spectrum (off resonance spectrum). Finally, the relative percentages of spin saturation (relative % STD) to calculate the STD-derived epitope map were estimated by setting to 100% the proton with the highest % STD intensity and calculating the others accordingly.

3.13. ROESY and TR-ROESY experiments

2D-ROESY experiments of LacdiNAc (3 mM) in the free state was recorded at 600 MHz (Bruker Avance III spectrometer equipped with a 5-mm inverse detection triple-resonance cryogenic probe head with z- gradients) with a mixing time of 300 ms at 293K. Also, 2D-ROESY experiments of LacdiNAc (500 µM) in presence of MGL-CRD (100 µM) were acquired at 800 MHz (Bruker Avance III 800 MHz spectrometer, equipped with a TCI cryoprobe) with a mixing time of 150 ms at 293K.

3.14. Molecular Dynamics (MD) Simulations

The MD simulations were carried out in collaboration with Dr. Francisco Corzana of Universidad de La Rioja, Spain. MD simulations were performed using the same software and force fields previously described in Diniz et al[37]. The 3D models of the MGL complexed with LacdiNAc and LacNAc were generated according to Marcelo et al[36]. The setup of MD used was the same as described previously[37], except for the time of the production step that was set to 2 µs instead of 100 ns.

3.15. X-Ray Crystallography

The crystallization procedure as well as the experiments and data treatment that followed were carried out in collaboration with Dr. Ana Luisa Carvalho and Filipa Trovão of the XTAL Macromolecular Crystallography Group of UCIBIO at FCT-NOVA.

3.15.1. Co-crystallization of Gal-3 in the presence of LacdiNAc

The co-crystallizations of the carbohydrate recognition domain of galectin-3 in complex with LacdiNAc were performed using the hanging drop diffusion method in a 24-well crystallization plate (Molecular Dimensions). The protein (Gal-3 in 50 mM sodium phosphate buffer (pH 6.8), 1 mM DTT, and 0.1% NaN_3) was previously incubated for 4 hours with 10 mM of LacdiNAc at 4 °C. The crystallization drops were set up with different concentrations of the incubated protein (10 to

20 mg/mL), different precipitant conditions (28% to 34% (w/v) polyethylene glycol (PEG) 4000, 100 mM Tris-HCl (pH 7.5), 100 mM $MgCl_2$ and 8 mM β-mercaptoethanol) and different proportions [(2 µL protein + 2 µL precipitant solution) and (3 µL protein + 2 µL precipitant solution)]. The crystals appeared after 48-72 hours and continued to grow for 4-5 days, after which the crystals were flash-cooled in liquid nitrogen, cryo-protected in a solution containing 34% (w/v) polyethylene glycol (PEG) 4000, 100 mM Tris-HCl (pH 7.5), 100 mM $MgCl_2$ and 8 mM β-mercaptoethanol, and 12% (v/v) glycerol.

3.15.2. X-ray diffraction data collection and structure refinement

Complete X-ray diffraction data of Gal-3/LacdiNAc crystals were collected at the BL13-XALOC beamline of the ALBA Synchrotron (Barcelona, Spain) with a radiation wavelength of 0.976 Å. The best crystal diffracted to a maximum resolution of 1.25 Å, which was grown with a protein concentration of 15 mg/mL against a reservoir containing 32% (w/v) polyethylene glycol (PEG) 4000, 100 mM Tris-HCl (pH 7.5), 100 mM $MgCl_2$ and 8 mM β-mercaptoethanol.

The diffraction data was processed using autoPROC software package[72] from the beamline pipeline. The structure was solved by molecular replacement using PhaserMR[73], with the unbound Gal-3 (PDB code: 3ZSL) as search model[74]. All subsequent refinement cycles were carried out in the Phenix platform[75], using program phenix.refine[76]. The ligand coordinate file was generated through eLBOW program[77] using the isomeric SMILES string from the compound. The resulting R_{work} and R_{free} were 0.17 and 0.19, respectively, with all residues in the allowed region of the Ramachandran plot (97.8% in the favoured region). All data collection and refinement statistics are summarized in Appendix C, Table A2.

4. Results and Discussion

4.1. Molecular recognition of LacdiNAc and LacNAc epitopes by immune-related lectins

The interaction of cancer-specific glycans by lectins plays an essential role in immune surveillance and tumour immune modulation[78]. In this perspective, the elucidation of the molecular determinants that govern the interaction of tumour glycan epitopes with immune-related lectins is crucial for the rational design of glycan-based cancer therapies, such as molecules (glycomimetics) able to block the aberrant interaction between tumour-glycans and lectins. By combining and integrating different biochemistry and biophysics techniques, like isothermal titration calorimetry (ITC), molecular dynamics, X-Ray crystallography and mainly NMR spectroscopy, it was characterized the recognition process of two key glycans epitopes, the LacdiNAc (GalNAcβ1 → 4GlcNAc) and the LacNAc (Galβ1 → 4GlcNAc) structures, by two lectins: the human macrophage galactose C-type lectin (MGL) and human galectin-3 (Gal-3) (Figure 4.1).

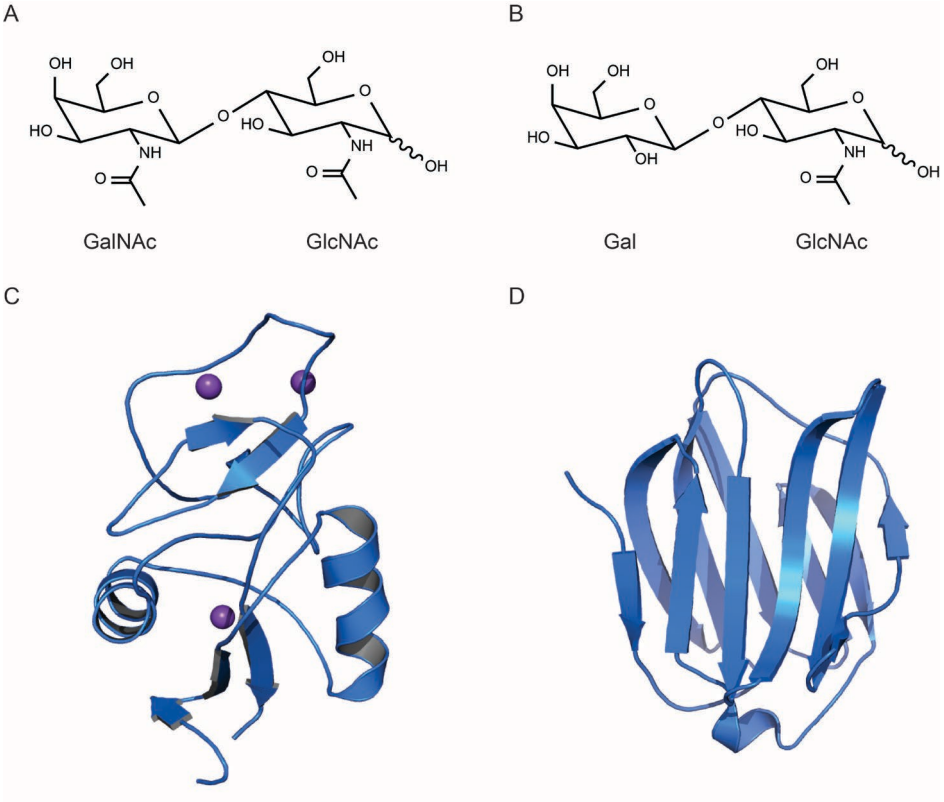

Figure 4.1 *- Representation of the glycan structures and 3D models of the lectins under study* ***A.*** *LacdiNAc chemical structure;* ***B.*** *LacNAc chemical structure;* ***C.*** *MGL CRD 3D structure[36];* ***D.*** *Gal-3 CRD 3D structure[52].*

4.2. Characterization of MGL/LacdiNAc complex

4.2.1. Isothermal titration calorimetry (ITC) experiments

To determine the equilibrium dissociation constant (K_D) of the MGL/LacdiNAc complex and to obtain the thermodynamic parameters behind the complex formation isothermal titration calorimetry (ITC) experiments were accomplished. For comparison reasons, ITC data for MGL/α-Me-GalNAc was also carried out. These experiments were performed in collaboration with the group of Prof. Jiménez-Barbero at CIC bioGUNE in Bilbao. The ITC experimental data and the summary of the estimated parameters are shown in Figure 4.2, and Table 4.1, respectively.

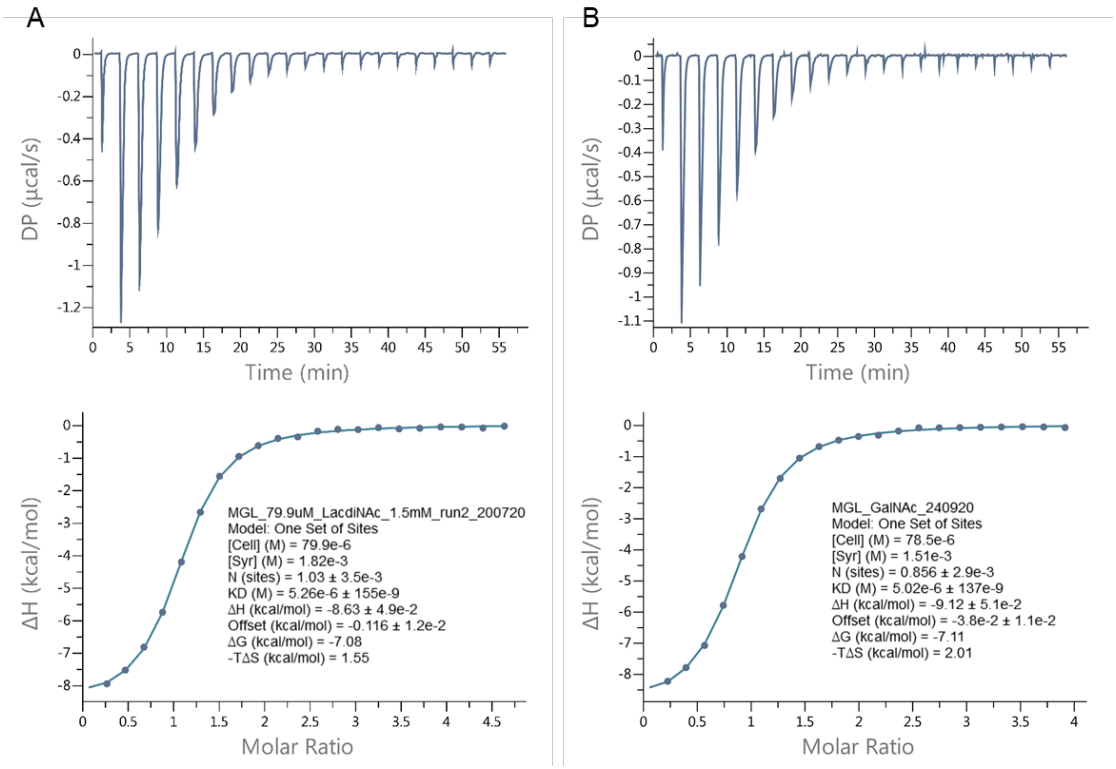

Figure 4.2 *- Representative ITC data for the titrations of MGL with LacdiNAc (A) and α-Me-GalNAc (B) in HEPES buffer (10 mM) containing 20 mM $CaCl_2$ and 75 mM NaCl at pH 7.5. Injections were performed every 150 s at* 293 K. *Top: raw data for the titration of glycan into protein. Bottom: binding isotherm of the calorimetric titration. The continuous line represents the least-squares fit of the data to a single-site binding model.*

Table 4.1 *- Thermodynamic parameters for the binding of LacdiNAc and α-Me-GalNAc to MGL determined by ITC experiments. Data fitted to a single-site binding model. Average values and standard deviation were indicated.*

	MGL/LacdiNAc	MGL/α-Me-GalNAc
K_D (μM)	5.3 ± 0.4	6.6 ± 1.5
ΔG (kcal/mol)	-7.08 ± 0.05	-7.04 ± 0.08
ΔH (kcal/mol)	-8.8 ± 0.4	-8.4 ± 0.8
-TΔS (kcal/mol)	1.7 ± 0.3	1.3 ± 0.7
N (sites)	0.99 ± 0.06	0.92 ± 0.12

The binding isotherms obtained for both complexes are well defined, which suggests that the complexes are stable and fit for a single-site binding model (N of 0.99 and 0.92 for LacdiNAc and α-Me-GalNAc, respectively). According to the ITC results, the K_D for both complexes is similar in the low μM range and compatible with moderate to high affinity complexes. The thermodynamics of the complexes are identical suggesting that GalNAc unit is the key unit for the recognition process. In both complexes, the small negative entropic contribution ($-T\Delta S > 0$) is compensated by the enthalpic factor. This enthalpy/entropy balance corresponds to that typically observed in glycan-protein complexes[22,53].

4.2.2. Saturation transfer-difference (STD) NMR experiments

With the ITC results in hand, the interaction of LacdiNAc by MGL was investigated by STD-NMR. The assignment of the ^{1}H-NMR proton of the ligand (Appendix B, Table A1) was firstly carried out by the interpretation of 2D NMR experiments (see experimental section). The assignment of ^{1}H-NMR proton is essential to identify which protons are mostly involved in the recognition. The STD-NMR experiments allowed to obtain information about the binding epitope of LacdiNAc to MGL. Figure 4.3 displays the off-resonance spectrum and the difference between the off and on-resonance spectrums, the called STD-NMR spectrum, of LacdiNAc in presence of MGL. To the resulting STD-NMR spectrum of the complex it was subtracted the STD-NMR spectrum of the protein and ligand alone (see experimental part).

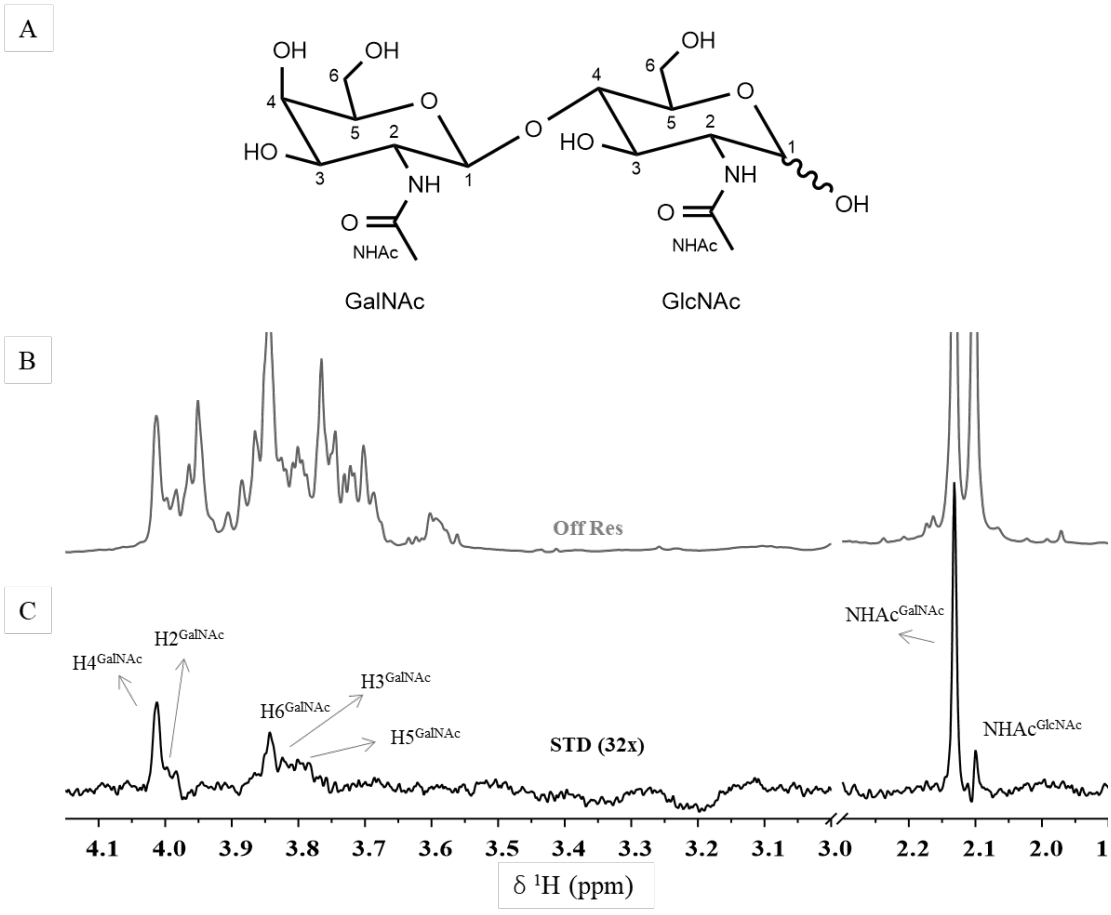

Figure 4.3 *- STD-NMR of LacdiNAc at 450 µM in the presence of 22,5 µM MGL obtained at 310 K and 600 MHz.* ***A.*** *LacdiNAc chemical structure with the protons numbering.* ***B.*** *Off resonance spectrum (Off Res)* ***C.*** *STD-NMR spectrum (STD). In the STD spectrum the key interacting protons are identified.*

The best STD-NMR spectrum was obtained at 310 K, which is compatible with the high affinity of the MGL/LacdiNAc complex estimated by ITC. By changing the temperature, it is possible to modulate the kinetics of the process. Thus, by increasing the temperature, the k_{off} of the complex increases, which favours the exchange between the bound and free ligand, and thus allowing to detect the transfer of magnetization from the protein to the ligand by analysing the intensities of the proton signals of the ligand in the free state.

Therefore, by comparing the intensities between the off and STD-NMR spectra, it was possible to derive the LacdiNAc epitope map (Figure 4.4). The H4 of the GalNAc unit is the one that presents the highest absolute STD percentage, thus it was set to 100%. The % STD of the other protons were then calculated accordingly, allowing to determine the STD-derived epitope map (Figure 4.4). Besides H4, the protons H3, H2, and the methyl group of the NHAc group of the GalNAc also present moderate STD intensities. In contrast, the protons of GlcNAc unit shows negligible STD enhancements. This result indicates that the GalNAc unit of LacdiNAc structure is the main interacting motif responsible for MGL binding, which agrees with previous studies that demonstrated the importance of the GalNAc unit in the recognition of GalNAc-containing ligands,

such as the blood group antigen type A, the Forsmann antigen and the GM2 type of structures, by MGL[37]. The STD-derived epitope map of α-Me-GalNAc in presence of the carbohydrate recognition domain of MGL was recently recorded by Coelho and Diniz in our group (Appendix D, Figure A1 unpublished results). Hence, an identical STD-derived epitope map is deduced by for GalNAc moiety either in LacdiNAc and α-Me-GalNAc structures, indicating a conserved binding mode for the interaction of both ligands to MGL.

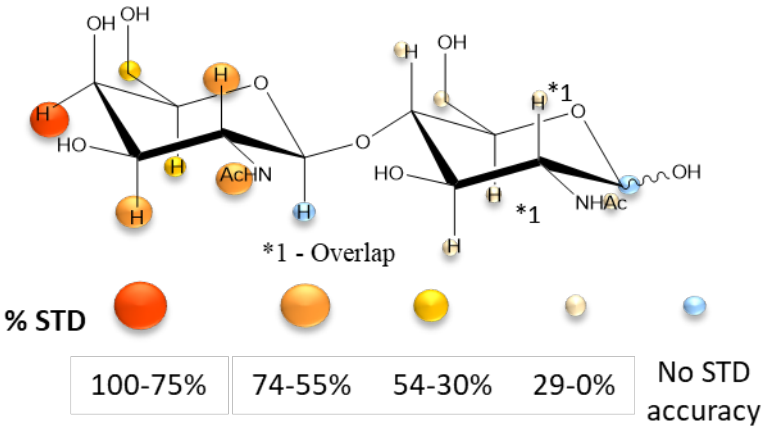

Figure 4.4 *- STD-derived epitope map obtained for LacdiNAc in presence of MGL.*

4.2.3. ^{1}H, ^{15}N-HSQC titration experiments

^{1}H,^{15}N-HSQC titration experiments of MGL with LacdiNAc were further performed to obtain structural information from the protein viewpoint (Appendix E, Figure A2). These experiments allow to determine the amino acids residues mostly perturbed in the presence of the ligand. Figure 4.5 displays the overlay of three selected ^{1}H,^{15}N-HSQC spectra of MGL in the absence and after addition of different concentrations of LacdiNAc. The differences in the spectra in the free and bound state account the influence of LacdiNAc in the chemical environment of MGL amino acids upon binding.

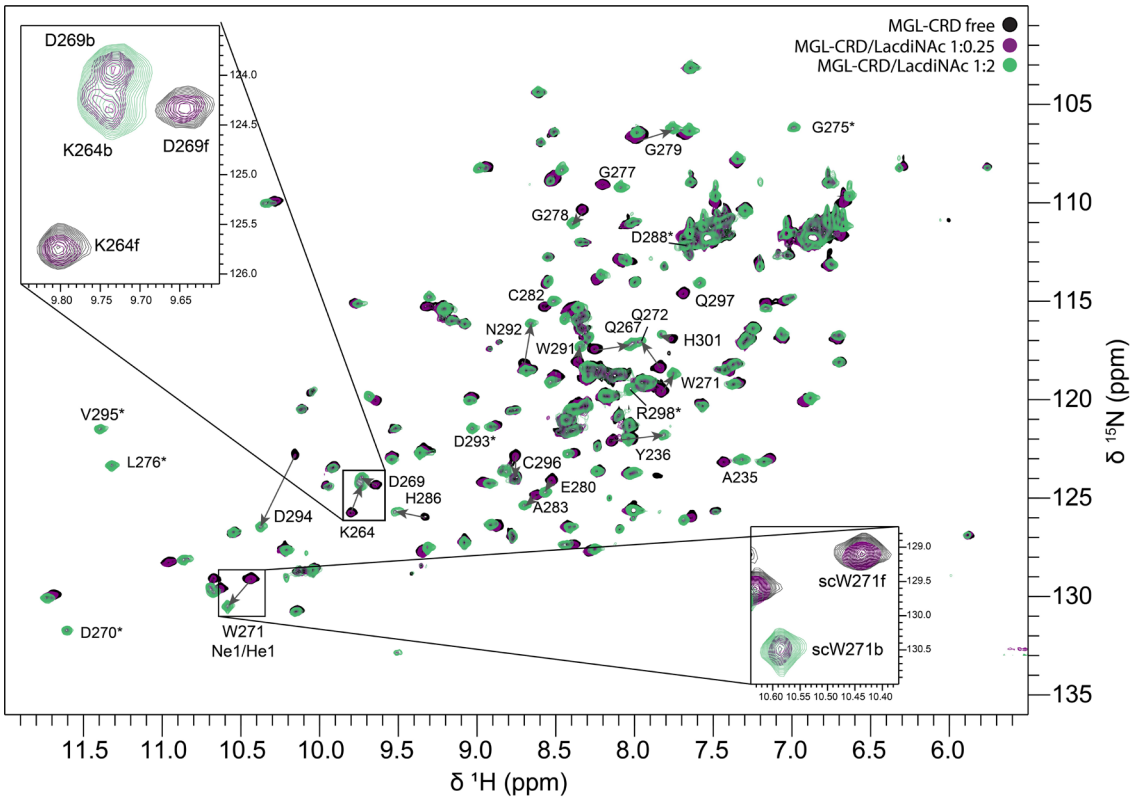

Figure 4.5 *- Overlay of the $^{1}H,^{15}N$-HSQC spectra of the titration of MGL with LacdiNAc: MGL CRD (200 μM) in apo state (black), MGL CRD/LacdiNAc complex in presence of 0.25 equivalent of LacdiNAc (purple) and MGL CRD/LacdiNAc complex in presence of 2 equivalent of LacdiNAc (green). The bottom right and top left of the spectra present the expansion region of the K264/D269 and W271 side chain cross-peaks, respectively. The residues that appear after addition of LacdiNAc are labelled with *.*

From the analysis of $^{1}H,^{15}N$-HSQC titration (Figure 4.5 and Appendix E, Figure A2) it is possible to observe that few resonances, absent on the free state, appear upon addition of LacdiNAc (residues labelled with *). These resonances correspond to residues D270, G275, L276, D288, D293, V295 located at the long loop L5, β-strand β4 and loops L6 and L7 (homology model Figure 4.6-B) and near to the putative binding site of GalNAc. This result indicates that part of the long loop L5 should adopt different conformations in free state with some of the residues exchanging in the intermediate regime in the NMR chemical shift time scale. In the presence of the ligand, this region of the loop stabilizes preferably adopting one major conformation. The appearance of some of the loop resonances along with the overall increase of signal intensities in the $^{1}H,^{15}N$-HSQC spectra highlight conformational changes in the protein upon ligand binding. To notice that this behaviour was previously reported in the case of the monosaccharide α-Me-GalNAc[37], which indicate that LacdiNAc and α-Me-GalNAc induces the same perturbations in the protein and might share the same binding mode to MGL.

In addition, the titration clearly shows a slow exchange regime between the free and bound state of MGL. This slow exchange regime is deduced by the observation of separate resonances for the free and bound form with the concomitant change in their intensities through the titration, directly reflecting the population between the two states (Figure 4.5 and Appendix E,

Figure A2). In Figure 4.5 the expansions show that the resonances in purple (MGL/LacdiNAc 1:0.25) appear at the two chemical shifts corresponding to the free and bound states. The assignment of the peaks of MGL/LacdiNAc complex was deduced by comparison with the $^{1}H,^{15}N$-HSQC of MGL/Me-GalNAc previously reported[37]. In addition, for the residues that are present in the free and bound state, it is possible to also account the chemical shift perturbation (CSP) induced by the ligand by comparing the $^{1}H,^{15}N$-HSQC in absence and presence of LacdiNAc. The CSP of each residue can thus be tracked by the calculation of the combined ^{1}H and ^{15}N chemical shift ($\Delta\delta_{Comb}$) (see experimental section) and plotted in a histogram. The histogram of the CSP for the MGL/LacdiNAc complex along with the map of the most affected residues (those that appear and strongly shift) in the 3D homology model of MGL-CRD is displayed in Figure 4.6.

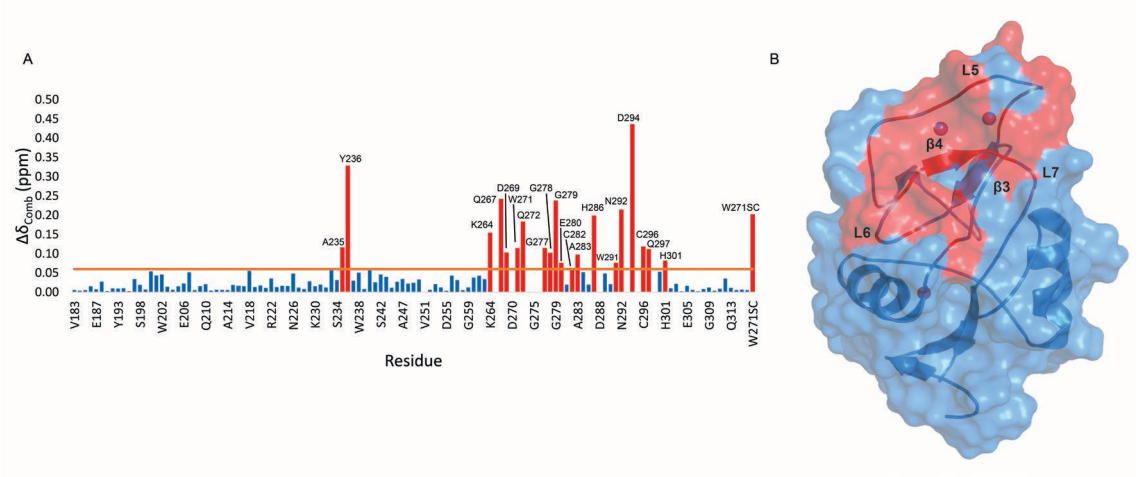

Figure 4.6 *- MGL backbone amide chemical shift variations upon LacdiNAc binding.* ***A.*** *Histogram of the $\Delta\delta_{Comb}$ for each amino acid of MGL/LacdiNAc complex in the presence of 2 equivalents of LacdiNAc. A cut-off line at 0.06 ppm (arbitrary threshold), displayed as an orange line, is considered to discriminate the residues that strongly shift in the presence of LacdiNAc. As red bars, there are displayed the residues with $\Delta\delta_{Comb}$* > 0.06 ppm. ***B.*** *Map of the residues perturbed upon LacdiNAc binding highlighted in 3D MGL-CRD homology model. The three Ca^{2+} present on the MGL-CRD structure are shown as spheres. The amino acids that are most affected in the presence of LacdiNAc (residues that appear and strongly shift) are displayed in red in the model.*

Observing the histogram and the map depicted in Figure 4.6 several MGL residues are strongly perturbed in the presence of the ligand. Residues from the long loop L5 are significantly perturbed (Y236 and residues from K264-E280) along with the residues of L6 (H286, D288) beta-strand β4 (N292-D294) and L7 (V295, Q297, H301). All these residues were previously identified to be affected upon addition of the simple monosaccharide, the α-methyl GalNAc[37], indicating a common binding mode for LacdiNAc and α-methyl GalNAc. The binding mode of α-Me-GalNAc was previously proposed by the group[36,37] and is depicted in Figure 4.7.

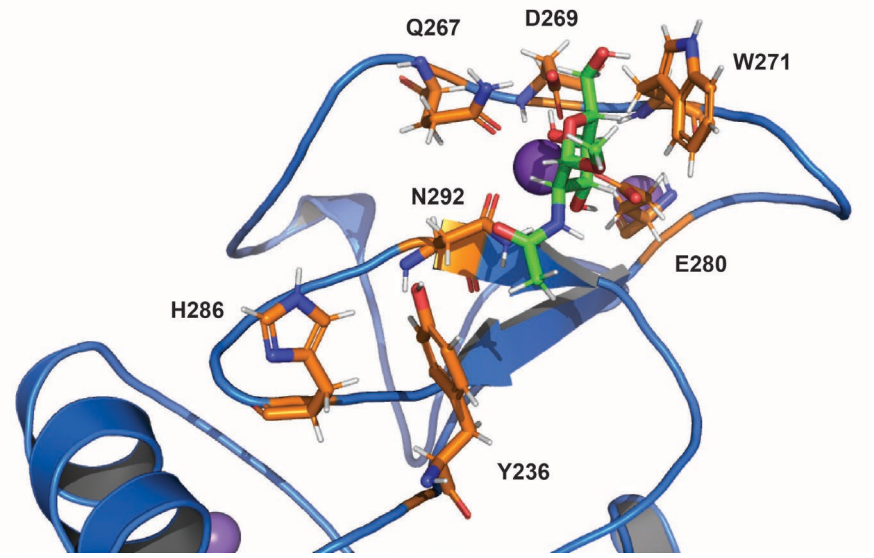

Figure 4.7 *- 3D view of the binding site of MGL/ α-Me-GalNAc complex obtained from molecular dynamics simulations showing the interaction of several residues (Y236, Q267, D269, W271, E280, H286, N292) with the ligand. Adapted from*[36].

Furthermore, as previously observed in the case of α-Me-GalNAc, LacdiNAc induces significant alterations in the protein structure upon ligand binding. To account the differences between the MGL/LacdiNAc and MGL/α-Me-GalNAc complexes the comparison of the corresponding $^{1}H,^{15}N$-HSQC spectra and the CSP calculation between the two MGL bound forms, were performed (Figure 4.8).

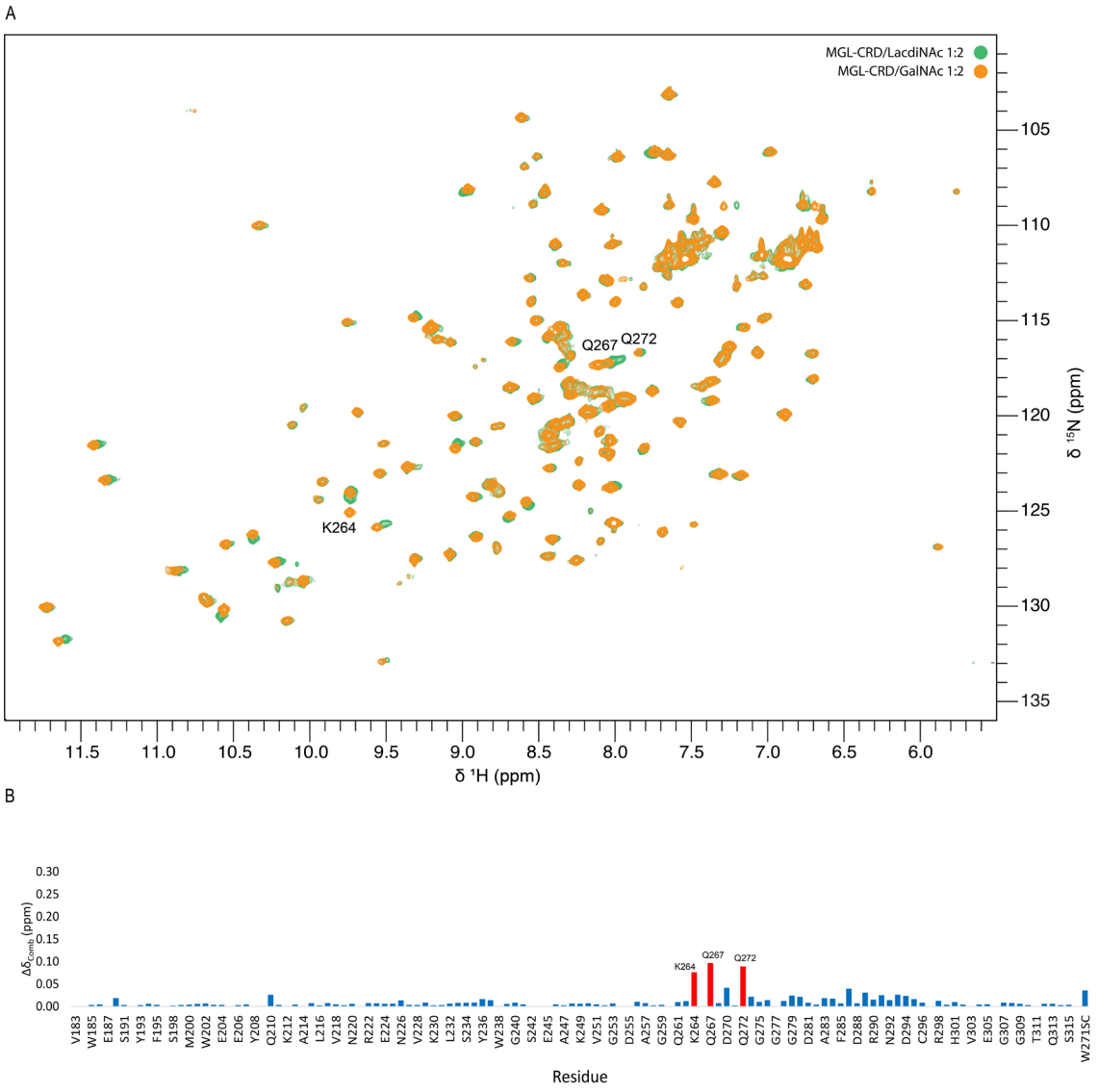

***Figure 4.8** - Differences of MGL/LacdiNAc and MGL/α-Me.GalNAc complexes inspected by NMR **A.** Overlay of the $^{1}H,^{15}N$ HSQC spectra of the MGL/LacdiNAc complex (green) with the MGL/GalNAc complex (orange). **B.** Histogram of the $^{1}H,^{15}N$ combined chemical shift ($\Delta\delta_{comb}$) between MGL/LacdiNAc and MGL/GalNAc complexes. The residues that show major differences ($\Delta\delta_{comb}$ > 0.05 ppm) are annotated and displayed as red bars.*

Small differences in a few amino acids can be appreciated between the two complexes in the CSP histogram (Figure 4.8-B), showing the sensitivity of some residues to the two GalNAc-containing structures.

Finally, from the analysis of the intensities of the $^{1}H,^{15}N$-HSQC spectra of MGL upon gradual addition of the LacdiNAc it is also possible to estimate the K_D of the complex. Along the $^{1}H,^{15}N$-HSQC titration there is a gradual increase in the volume of bound signals that can be correlated to the gradual addition of ligand, allowing to estimate the apparent K_D of the complex (K_{Dapp}) see experimental section). In the Figure 4.9 are presented the curves obtained for six residues. The selection of the residues was based on their involvement on GalNAc binding.

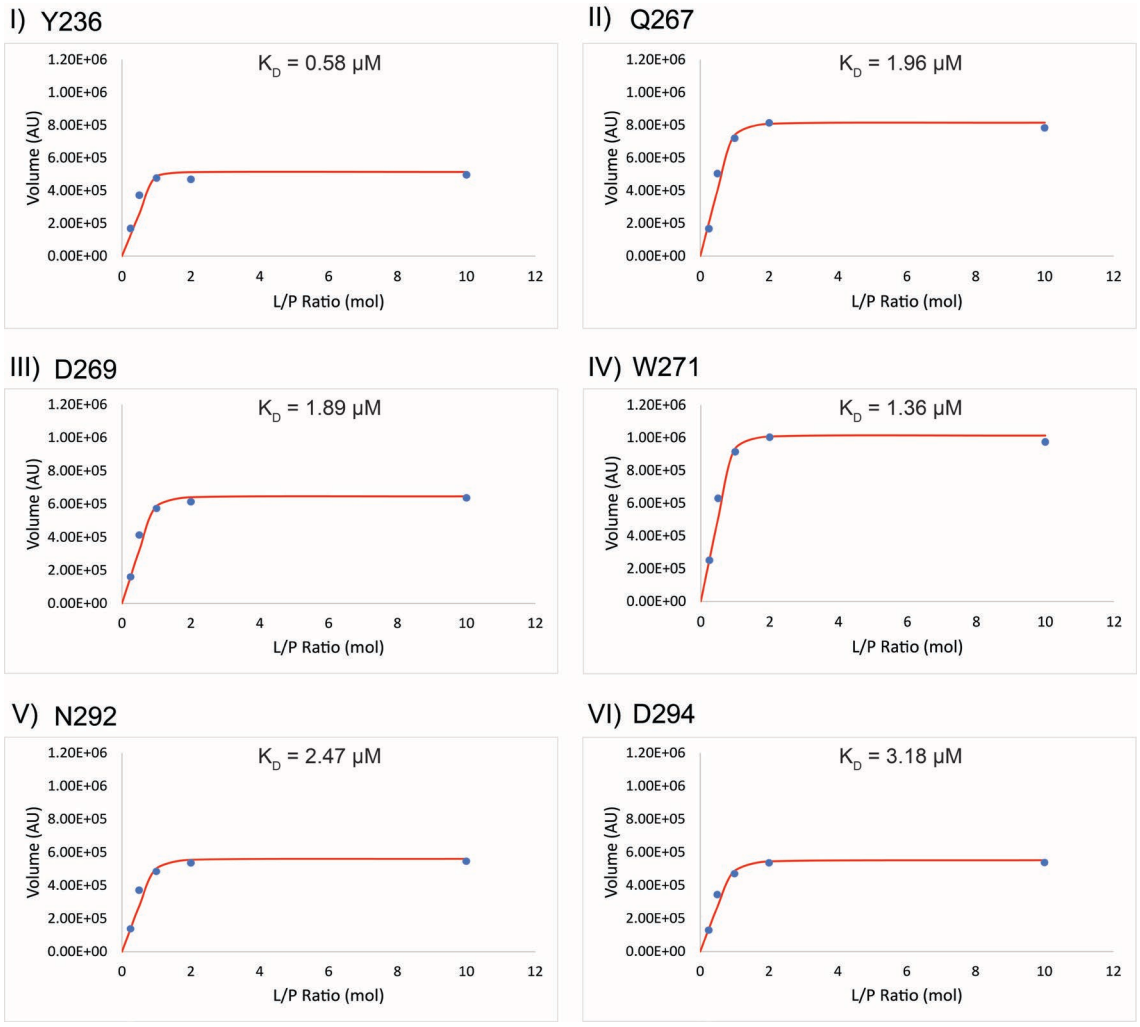

Figure 4.9 *– Fitting curves used for the estimation of the apparent K_D values for six residues: I) Y236; II) Q267; III) D269; IV) W271; V) N292; VI) D294.*

The estimated average of K_{Dapp} is of 1.9 ± 0.9 µM, a high affinity complex, which agrees with the previous result obtained by ITC.

4.2.4. Transferred ROESY experiments

TR-ROESY experiments of LacdiNAc in absence and presence of MGL were also carried out (Figure 4.10). Comparing both spectra, it is possible to appreciate exchange transfer cross-peaks between the free and bound state of LacdiNAc. By this experiment, it was also possible to assign the chemical shift of LacdiNAc in the bound state.

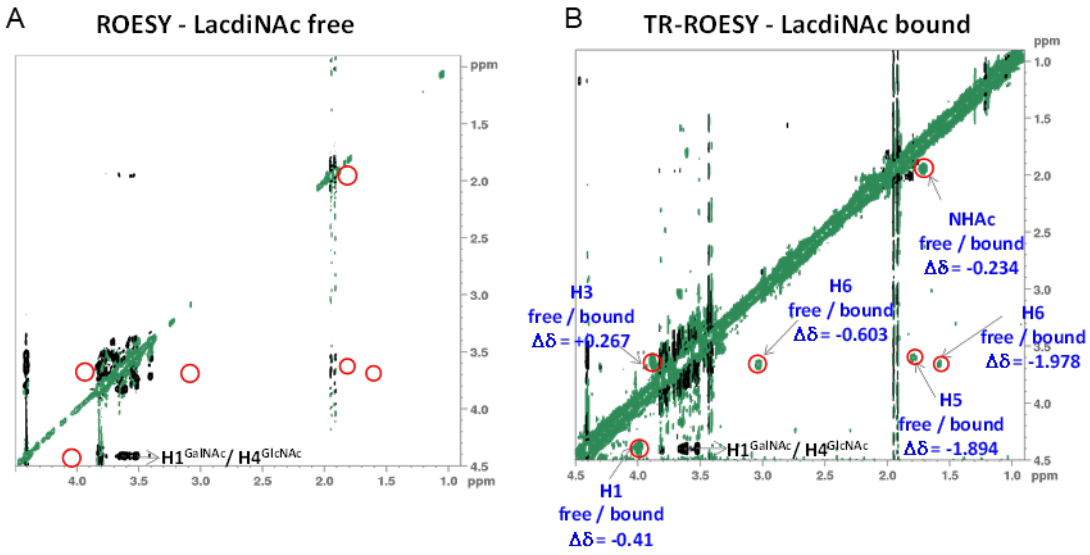

Figure 4.10 *- 2D-NMR ROESY spectra of LacdiNAc in free and bound to MGL. A) 2D ROESY of LacdiNAc (3 mM) at 600MHz and 293K (mixing time 300 ms) B) 2D TR-ROESY of 1:5 MGL/LacdiNAc at 800MHz and 293K (mixing time 150 ms).*

Curiously, the cross-peaks of H5, one of H6s and the methyl group of NHAc of LacdiNAc in the bound state show an upfield shift when compared to those observed in the free state. Specifically, H5 and one of the H6 of CH_2OH group of the GalNAc unit experience a large upfield shift of -1.89 ppm and -1.98 ppm, respectively. This observation contrasts with the downfield shift of H3 proton of +0.3 ppm. No exchange cross-peaks are detected for the GlcNAc unit of LacdiNAc, indicating once again that this unit should play a minor role in the recognition process. The upfield shift of H5 and one of H6 is characteristic of the proximity of these protons to an aromatic group. This result agrees with the binding pose proposed by Me-GalNAc to MGL (Figure 4.7). With these experiments, it is also possible to detect an interglycosidic NOE cross-peak between the H1 of GalNAc and H4 of GlcNAc in LacdiNAc in free and bound state. This NOE indicates that the conformer of LacdiNAc has the NHAc moieties of GalNAc and GlcNAc in opposite directions, and this conformation is prevalent either in free and bound state.

4.2.5. Molecular dynamics simulations

To further support these findings, it was performed microseconds molecular dynamics (μs-MD) simulations. These simulations were carried out in collaboration with Dr. Francisco Corzana of Universidad de La Rioja, Spain.

The MGL-CRD/LacdiNAc complex is stable for the 2 μs of the simulation. The MD model shows that the GlcNAc unit of the ligand is completely exposed to the solvent, which agrees with the STD-NMR results. Besides the coordination with Ca^{2+} at the binding site, there is a CH-π bond

between the H5, and one of the H6 of CH_2OH group of GalNAc unit and the aromatic five-membered ring of W271, and another CH-π between the N-acetyl group of GalNAc and Y236, which explains both the strong CSP of these residues (W271 and Y236), as well as the upfield shift of H5, H6 and methyl group of NHAc of LacdiNAc due to the ring current effect of (Y236 and W271). It was also detected an H-bond between H286 and the carbonyl of the N-acetyl group of GalNAc unit. Finally, it is also possible to confirm that the major conformer of LacdiNAc along the MD remained with NHAc groups of GalNAc and GlcNAc in opposite directions.

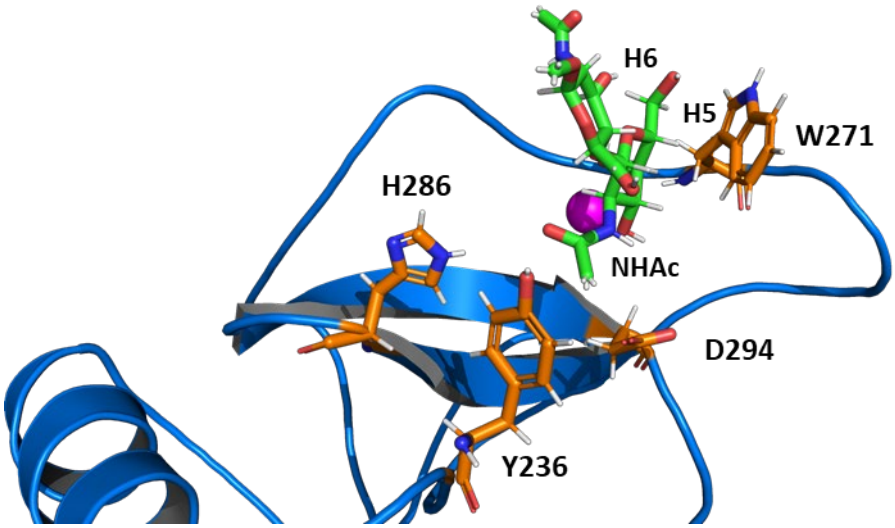

Figure 4.11 *- 3D view of the binding site of MGL/LacdiNAc complex derived from MD demonstrating the proximity of the side chain of W271 to H5 and one of the H6s protons of GalNAc unit, as well as the interaction of Y236 and H286 and D294 with the N-acetyl group of the GalNAc unit of LacdiNAc.*

4.3. Characterization of MGL/LacNAc Complex

4.3.1. Isothermal titration calorimetry (ITC) experiments

The MGL/LacNAc complex was analysed by ITC experiments. The obtained results and thermodynamic values are shown in Figure 4.12 and Table 4.2 respectively.

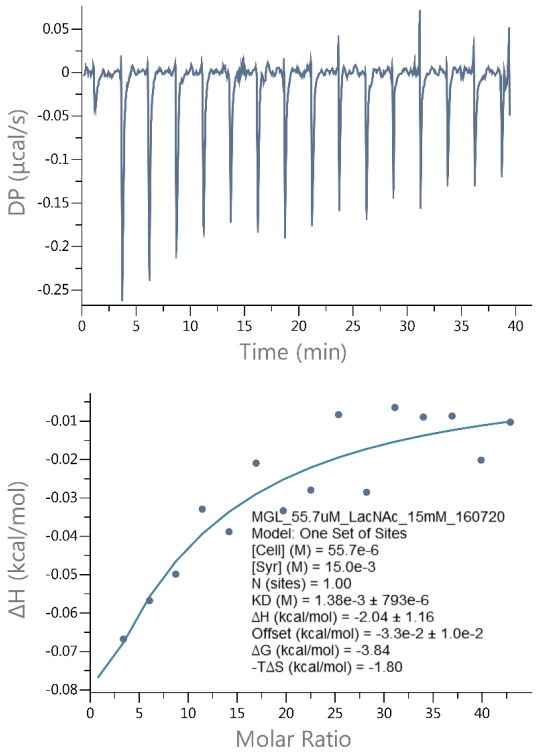

Figure 4.12 *- Representative ITC data for the titrations of MGL with LacNAc in HEPES buffer (10 mM) containing 20 mM $CaCl_2$ and 75 mM NaCl at pH 7.5. Injections were performed every 150 s at 293 K. Top: raw data for the titration of glycan into protein. Bottom: binding isotherm of the calorimetric titration shown on top. The continuous line represents the least-squares fit of the data to a single-site binding model.*

Table 4.2 *- Thermodynamic parameters for the binding of LacNAc to MGL as determined by ITC experiments. Data fitted to single-site binding model. Average values and standard deviation were indicated.*

	MGL/LacNAc
K_D (μM)	>1380
ΔG (kcal/mol)	-----
ΔH (kcal/mol)	-----
-TΔS (kcal/mol)	-----
N (sites)	1.00

The impossibility to obtain a good binding isotherm for MGL/LacNAc complex precluded to determine an accurate value of K_D and the thermodynamic values of this complex. ITC suggests a K_D for MGL/LacNAc over 1380 μM, indicating a low affinity complex.

4.3.2. Saturation transfer-difference (STD) NMR experiments

With the assignment of the ^{1}H-NMR spectrum of LacNAc (see experimental section and Table A1) it was possible to analyse the STD-NMR of LacNAc in presence of MGL. In the Figure 4.13, it is presented the off-resonance spectrum along with the spectrum of the difference between off and on-resonance spectra, the STD spectrum. The STD spectrum of the mixture takes into account (subtracted) the STD controls of the LacNAc in absence of the protein and the protein in absence of LacNAc.

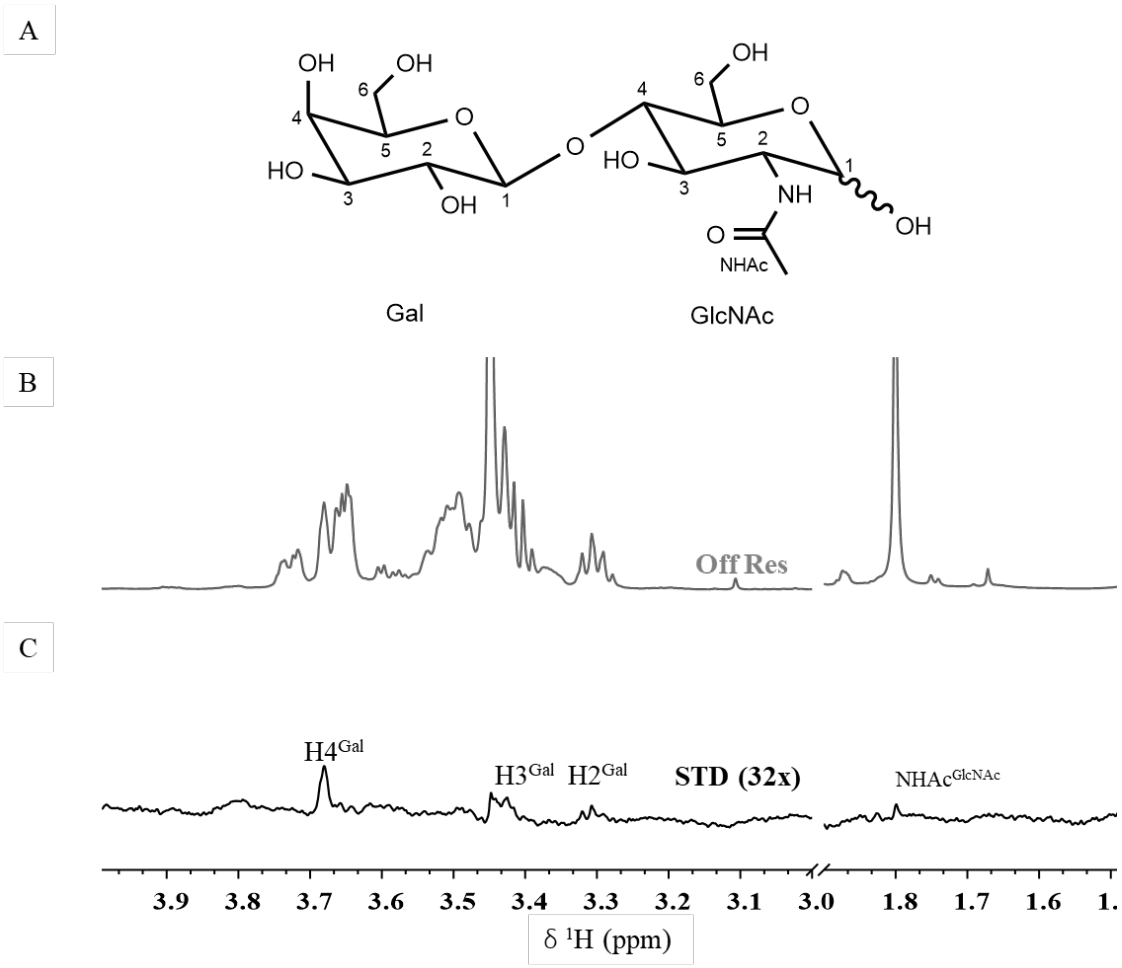

Figure 4.13 *- STD-NMR of LacNAc at 450 µM in the presence of 22,5 µM MGL obtained at 283 K and 600 MHz* ***A.*** *LacNAc chemical structure with the protons numbering.* ***B.*** *Off resonance spectrum (Off Res)* ***C.*** *STD-NMR spectrum (STD). In the STD spectrum the key interacting protons are identified.*

First, it is important to emphasize that best STD responses were obtained decreasing the temperature to 283 K. At 310, 298 or even 293 K negligible STD intensities of LacNAc are detected. This behaviour correlates very well if we assume a low affinity complex. Therefore, decreasing the temperature it is expected to increase the residence time of the ligand in the binding site (lowering the k_{off}), allowing the transference of saturation from the protein to the ligand. By observing the STD spectrum, the H4 of Gal is the proton that receives the most % of saturation from the protein, with H3 and H2 also showing some STD responses (Figure 4.14). No STD signals from the GlcNAc unit is detected suggesting the absence of any direct interaction between GlcNAc and MGL.

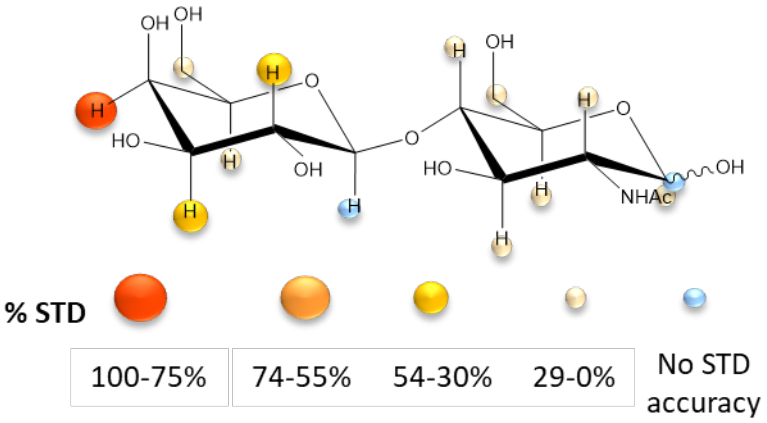

Figure 4.14 *- STD-derived epitope map obtained for LacNAc in presence of MGL.*

4.3.3. $^{1}H,^{15}N$-HSQC titration experiments

$^{1}H,^{15}N$-HSQC titration of MGL with LacNAc was performed. In the Figure 4.15, it is shown the overlay of four $^{1}H,^{15}N$-HSQC spectra, corresponding to: MGL in absence of LacNAc (black), MGL fully saturated with LacNAc (red, P:L 1:150), and two spectra in the intermediate points of the titration (blue, P:L 1:15 and purple, P:L 1:35).

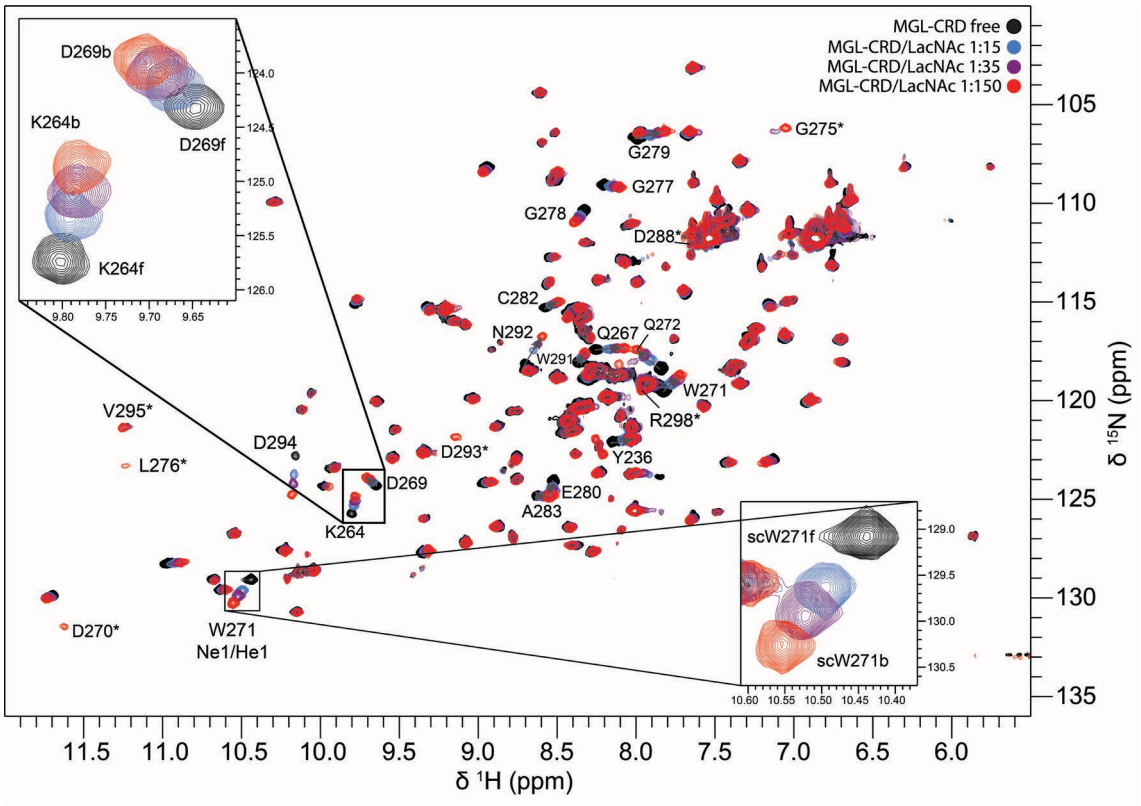

Figure 4.15 *- Overlay of the $^{1}H,^{15}N$-HSQC spectra of the titration of MGL-CRD with LacNAc: MGL (200 μM) in apo state (black), MGL/LacNAc complex in presence of 15 equivalents of LacNAc (blue), MGL/LacNAc complex in presence of 35 equivalents of LacNAc (purple) and MGL/LacNAc complex in presence of 150 equivalents of LacNAc (red). The bottom right and top left of the spectra present the expansion region of the K264/D269 and W271 side chain cross-peaks, respectively. The residues that appear in the presence of the ligand are labelled with *.*

In Figure 4.15 and Appendix E, Figure A3, it is possible to observe the fast exchange regime between MGL free and complexed state, a gradual shift of the signals from free to bound state can be followed during the ligand addition. Also, to displace completely MGL to the bound state a ratio of P:L of 1:150 was necessary, which is compatible with a low affinity complex and is in conformity with the previous ITC and STD-NMR results.

As in the case of LacdiNAc, several cross-peaks corresponding to the flexible part of the loop L5 appear (labelled with * on $^{1}H,^{15}N$-HSQC, Figure 4.15) indicating a stabilization of the dynamic loop in presence of LacNAc. Nevertheless, the intensity of some of those residues is lower in the case of LacNAc than LacdiNAc (e.g., Y236, K264, D270, L276, D294, etc). In addition, several residues of MGL also suffer chemical shift perturbation in the presence of LacNAc. CSP was determined through the analysis of the combined chemical shift ($\Delta\delta_{Comb}$) and shown in the histogram along with the map in the 3D homology model of the most affected residues of MGL in presence of LacNAc (Figure 4.16).

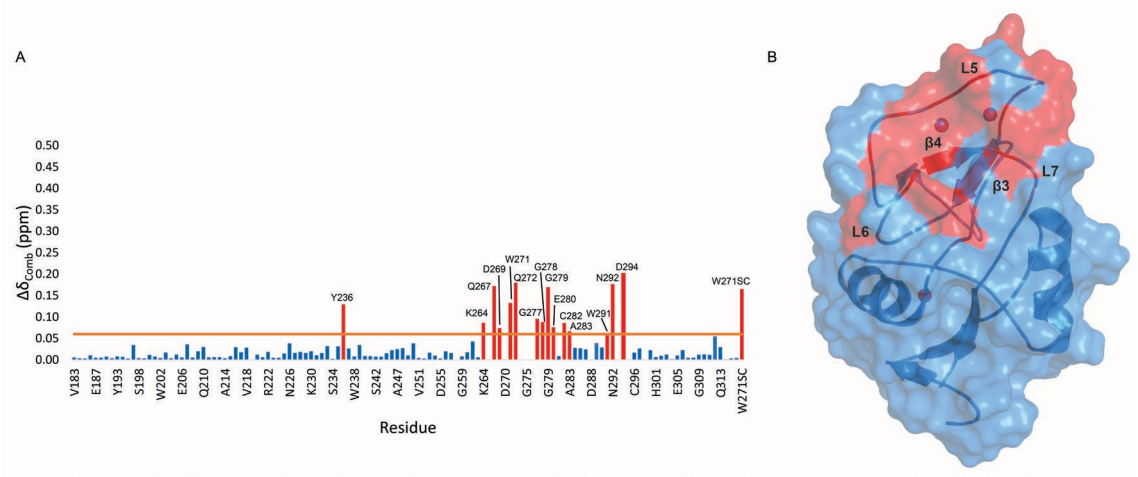

Figure 4.16 *- MGL backbone amide chemical shift variations upon LacNAc binding.* ***A.*** *Histogram of the $\Delta\delta_{Comb}$ for each amino acid of MGL/LacNAc complex in the presence of 150 equivalents of LacNAc. A cut-off line at 0.06 ppm (arbitrary threshold), displayed as an orange line, is considered to discriminate the residues that strongly shift in the presence of LacNAc, respectively. As red bars, there are displayed the residues with $\Delta\delta_{Comb}$* > 0.06 ppm. ***B.*** *Map of the residues perturbed upon LacNAc binding highlighted in 3D MGL homology model. The three Ca^{2+} present on the MGL structure are shown as spheres. The amino acids affected in the presence of LacNAc (residues that appear and strongly shift) are displayed in red in the model.*

As in the case of LacdiNAc, most of the residues that were perturbed in the presence of the ligand belong to the long loop L5 and beta-sheet β4 along with few residues of the loops L6 and L7, which indicates the same binding site to MGL for LacdiNAc and LacNAc.

Noteworthy, in comparison with LacdiNAc, the magnitude of the shifts of MGL residues in the presence of LacNAc is lower than with LacdiNAc, indicating a weaker perturbation in MGL structure in presence of LacNAc than LacdiNAc. This result corroborates with the lower affinity deduced by ITC measurements for LacNAc than LacdiNAc. MGL/LacNAc seems to be a weaker complex precluding the production of heat and becoming a not suitable complex to investigate by ITC technique. In this context, the apparent K_D was estimated based on the gradual shift of the amide resonances observed along the $^{1}H,^{15}N$-HSQC based titration of MGL with LacNAc (see experimental section). From the several residues that shifted in the presence of the LacNAc, six residues were selected based on their relevance for binding, and the corresponding fitting curves are presented in Figure 4.17.

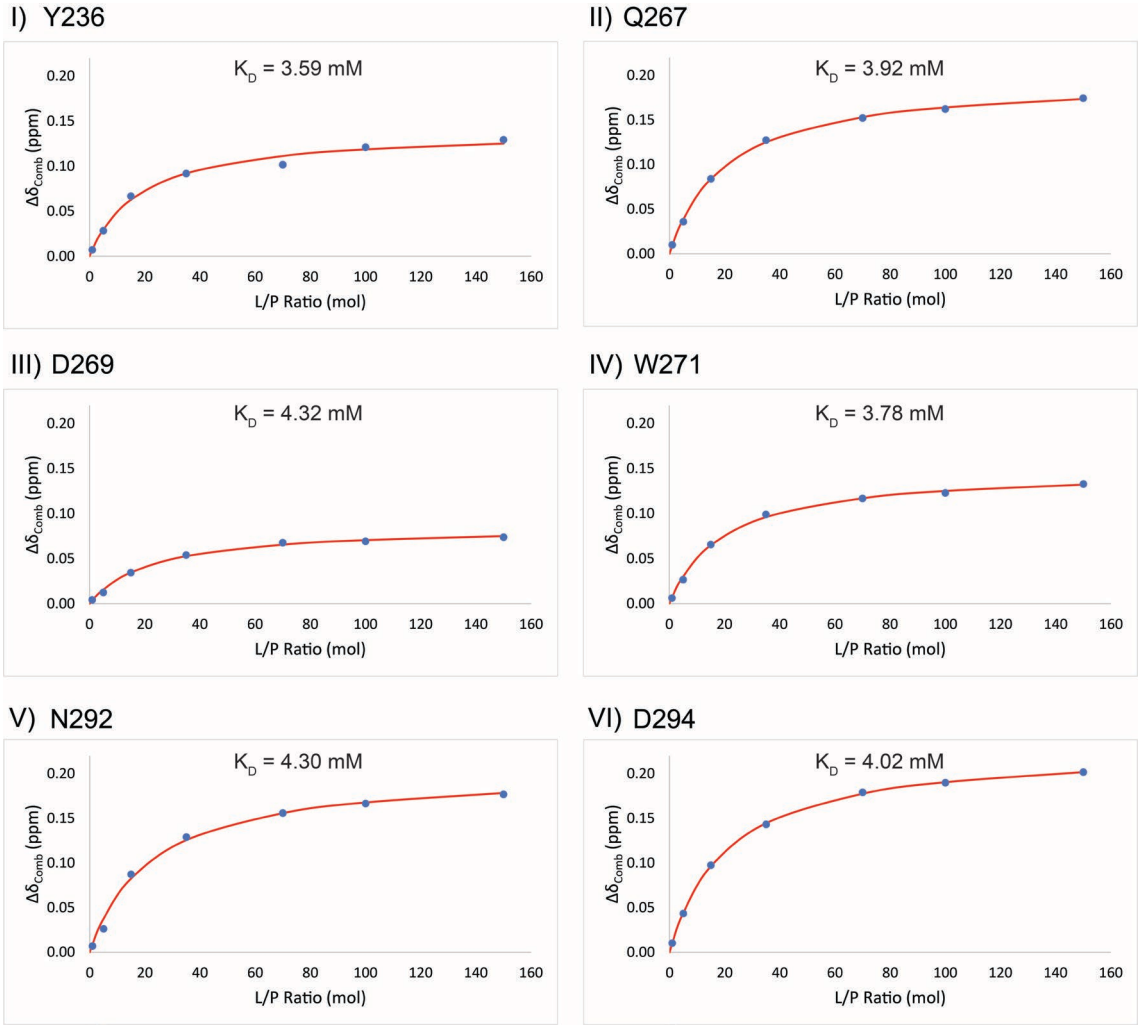

Figure 4.17 *- Fitting curves used for the estimation of K_{Dapp} values for six residues: I) Y236; II) Q267; III) D269; IV) W271; V) N292; VI) D294.*

The average of K_{Dapp} estimated for this complex by NMR was 4.0 ± 0.3 mM and agrees with the high P:L ratios required for saturation, as well as with the difficulty to obtain binding isotherms from ITC.

4.3.4. Molecular dynamics simulations

As in the case of the MGL/LacdiNAc complex, a 2 µs molecular dynamic simulation (µs-MD) for MGL/LacNAc complex was performed. This long dynamic simulation allows to observe that the complex was stable for only 1.8 µs in contrast to MGL/LacdiNAc complex. Nevertheless, the 3D model of the complex illustrated in Figure 4.18 shows the same binding site for LacNAc than LacdiNAc. Indeed, the key interaction between the Ca^{2+} and the two hydroxyl groups, OH-3 and OH-4, is conserved in both MGL/LacdiNAc and MGL/LacNAc complexes. Furthermore, the involvement of certain residues such as W271, Q267, D269 is also preserved in both complexes. However, besides these similarities, important differences arise in terms of MGL affinity and specificity towards LacdiNAc and LacNAc.

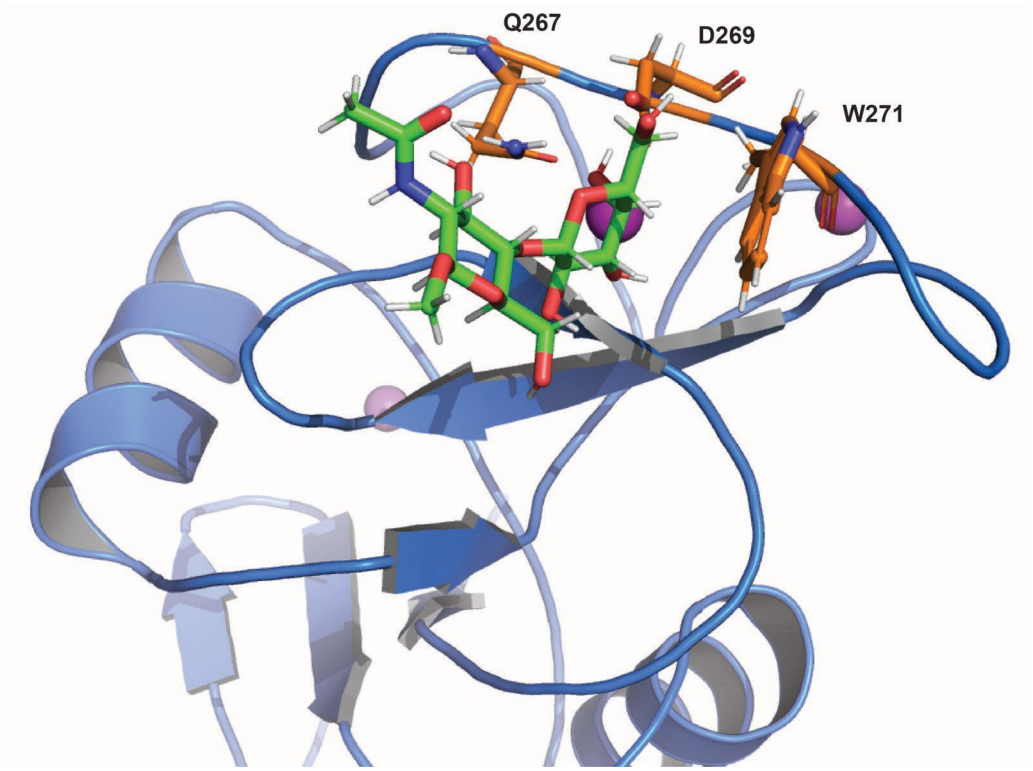

Figure 4.18 *- 3D view of the binding site of MGL/LacNAc complex derived from MD simulations illustrating the involvement of the residues Q267, D269 and W271 in the recognition of LacNAc.*

4.4. Comparison of MGL/LacdiNAc vs MGL/LacNAc complexes

Chemically the difference of LacdiNAc and LacNAc rely on the functional group at C2 of the non-reducing end of the disaccharide that transforms a GalNAc into a Gal unit from LacdiNAc to LacNAc (Figure 4.19).

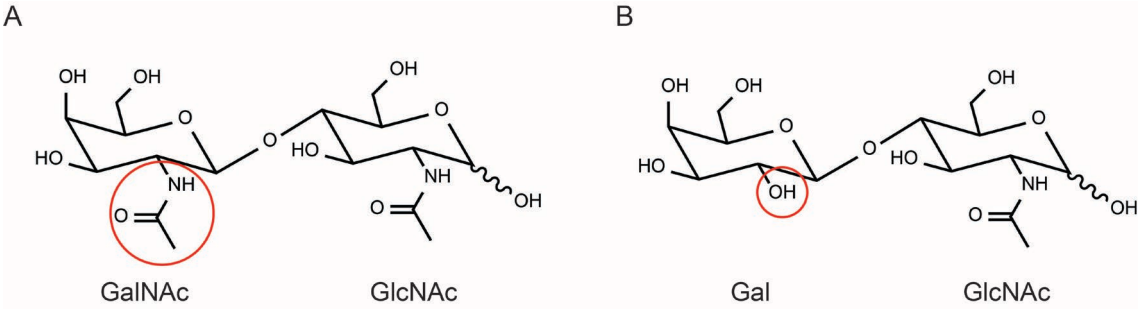

Figure 4.19 *- Representation of the chemical structures of* ***A.*** *LacdiNAc and* ***B.*** *LacNAc, evidencing the difference between the two functional groups at C2 of the non-reducing end of the disaccharides.*

From the chemical perspective, this alteration makes a substantial difference for MGL. The ITC results showed a significant variation in the terms of MGL affinity depending on the ligand context. MGL/LacdiNAc has a low K_D value of 5.3 ± 0.4 μM, while MGL/LacNAc has a K_D higher than 1380 μM. This difference in affinity is also reflected in the STD-NMR experiments. STD responses are strongly modulated by the rate of the exchange process between the free and bound state of the ligand. In this context, improved STD enhancements were obtained by changing the temperature of the experiments in opposite directions depending on the ligand context. Therefore, in the case of MGL/LacdiNAc, the STD-NMR experiment was acquired at 310 K (Figure 4.3) to facilitate the process of unbinding and to allow the detection; while in the case of MGL/LacNAc the experiment was recorded at 283 K (Figure 4.13) to increase the residence time of the ligand bound to MGL. Besides this difference in the exchange process, MGL recognizes both complexes through the non-reducing end of both disaccharides (GalNAc/Gal of LacdiNAc/LacNAc, respectively). In both complexes, the H4 proton, of GalNAc/Gal in LacdiNAc/LacNAc, respectively, receives the highest % of saturation from MGL. To notice that the GlcNAc residue present negligible STD response either for LacdiNAc and LacNAc (Figure 4.4 and 4.14). Thus, STD-NMR results clearly show that LacdiNAc and LacNAc present the same binding mode of interaction to MGL.

From the protein perspective the $^{1}H,^{15}N$-HSQC based titrations highlighted important differences in the recognition process of LacdiNAc and LacNAc by MGL. Along the $^{1}H,^{15}N$-HSQC titration it is possible to appreciate the difference in the exchange regime between the MGL free and ligand complexed state, depending on the employed ligand (Figure 4.5 and 4.15). For

MGL/LacdiNAc complex a slow exchange regime was detected (Figure 4.5), while in the case of MGL/LacNAc complex (Figure 4.15) it is observed a fast exchange regime on the chemical shift NMR timescale. These results suggest that LacdiNAc has longer residence times in the complexed form than LacNAc (k_{off} LacdiNAc << k_{off} LacNAc) and agree with the necessity to use different temperatures in the STD-NMR experiments. Since no major structural restraints are anticipated due to the additional N-acetyl group of GalNAc in LacdiNAc over the Gal unit in LacNAc, the k_{on} is expected to be limited by diffusion. Therefore, the k_{off} values for MGL/LacdiNAc and MGL/ LacNAc complexes are behind the differences in the exchange mechanism and explain the differences in the experimental K_D estimated by ITC. Close inspection of the $^{1}H,^{15}N$-HSQC of MGL/LacdiNAc and MGL/LacNAc complexes (Figure 4.20- A) allow to determine that both ligands induce perturbation in the same protein residues, which is compatible with the fact that they should binding in the same region of the protein. Remarkably, significant differences in the magnitude of the chemical shift perturbation of several residues can be identified for the two complexes. To better account these differences, a histogram of the CSP between the two MGL bound forms (MGL/LacdiNAc vs MGL/LacNAc) was accomplished and it is displayed in Figure 4.20-B.

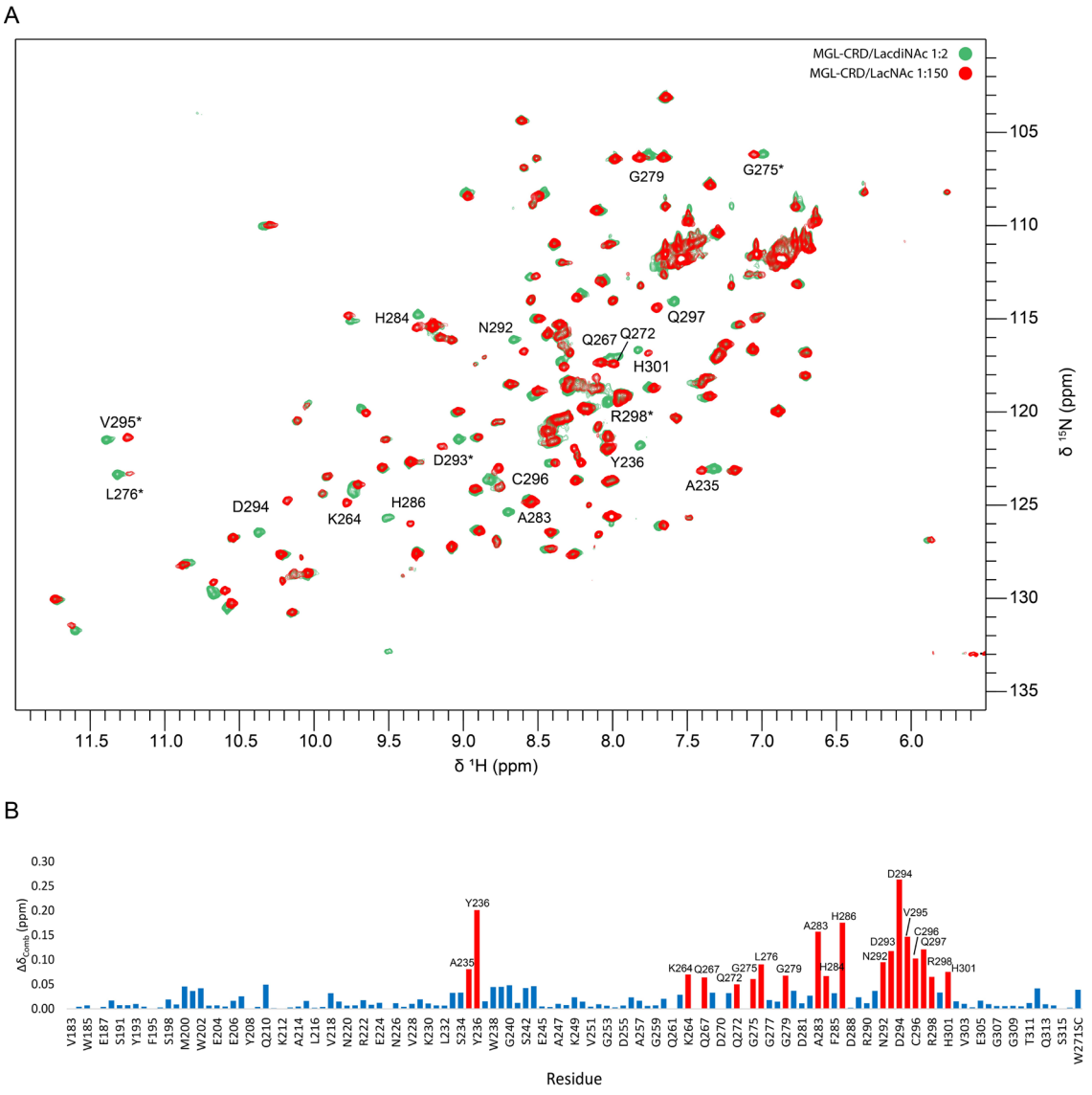

Figure 4.20 *- Differences of MGL/LacdiNAc and MGL/LacNAc complexes inspected by $^{1}H,^{15}N$-HSQC* ***A.*** *Overlay of the $^{1}H,^{15}N$ HSQC spectra of the MGL/LacdiNAc complex (green) with the MGL/LacNAc complex (red).* ***B.*** *Histogram of the $^{1}H,^{15}N$ combined chemical shift ($\Delta\delta_{comb}$) between MGL/LacdiNAc and MGL/LacNAc complexes. The residues that show major differences ($\Delta\delta_{comb}$ > 0.05 ppm) are annotated and displayed as red bars.*

The residues shown as red bars in the histogram correspond to those in which the CSP is higher than 0.05 ppm. These residues reflect the high sensitivity of MGL structure upon binding LacdiNAc or LacNAc. A high number of residues are differentially perturbed depending on the ligand context. In particular, the residues that show the highest CSP between the two bound forms are: Y236, H286 and D294. The specific perturbation of these residues should be related to the engagement of the NHAc group of GalNAc unit in LacdiNAc structure. To confirm this hypothesis molecular dynamics simulations were performed. Additionally, long times of simulations (2 μs) were used to evaluate the stability of the complexes.

The MD results show that MGL/LacdiNAc complex was stable along 2 μs of simulation, while the MGL/LacNAc complex was only stable for 1.8 μs. This result agrees with the difference in the exchange regime deduced in the $^{1}H,^{15}N$-HSQC titration experiments. Furthermore, according to the models (Figure 4.21), it is possible to confirm the proximity of Y236, H286 and D294 amino acids to the N-acetyl group of GalNAc in LacdiNAc structure. Specific interactions were detected involving the Y236 and H286 residues. A CH-π bond can be observed between Y236 and the methyl group of NHAc. Additionally, the H286 is participating in an H-bond with the carbonyl of NHAc of GalNAc in LacdiNAc. The interaction involving the NHAc group of GalNAc can thus explain the differences in the CSPs of these amino acids between the two complexes (Appendix E, Figure A4).

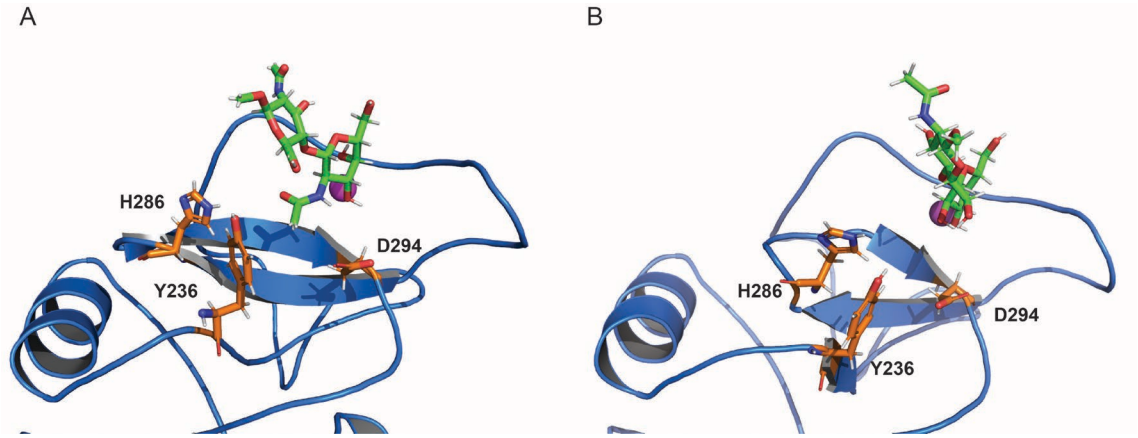

Figure 4.21 *- 3D view of the binding site of the complexes* ***A.*** *MGL/LacdiNAc and* ***B.*** *MGL/LacNAc. The proximity of the N-acetyl group of the GalNAc unit of LacdiNAc and the Y236, H286 and D294 residues is clearly highlighted in MGl/LacdiNAc complex.*

All these results point out that the N-acetyl group (NHAc) present in the GalNAc unit is behind the preference of MGL towards GalNAc-containing structures. Indeed, the additional interactions involving the side chains of Y236 and H286, not only stabilize the MGL/LacdiNAc complex but also turns it into a high affinity complex. Our results provide compelling evidence to explain why MGL has a high specificity and selectivity for GalNAc-containing structures.

4.5. Characterization of Galectin-3/LacdiNAc Complex. Comparison with Galectin-3/LacNAc complex

4.5.1. Isothermal titration calorimetry (ITC) experiments

As in the case of MGL complexes, binding information of Gal-3/LacdiNAc complex was accessed by ITC measurements. ITC data was also acquired for Gal-3/LacNAc for comparison purposes and values obtained are in agreement with those reported before[53]. Figure 4.22 and Table 4.3 shows the raw data/binding isotherm and the summary of the thermodynamic parameters for the complexes, respectively, deduced by the ITC experiments.

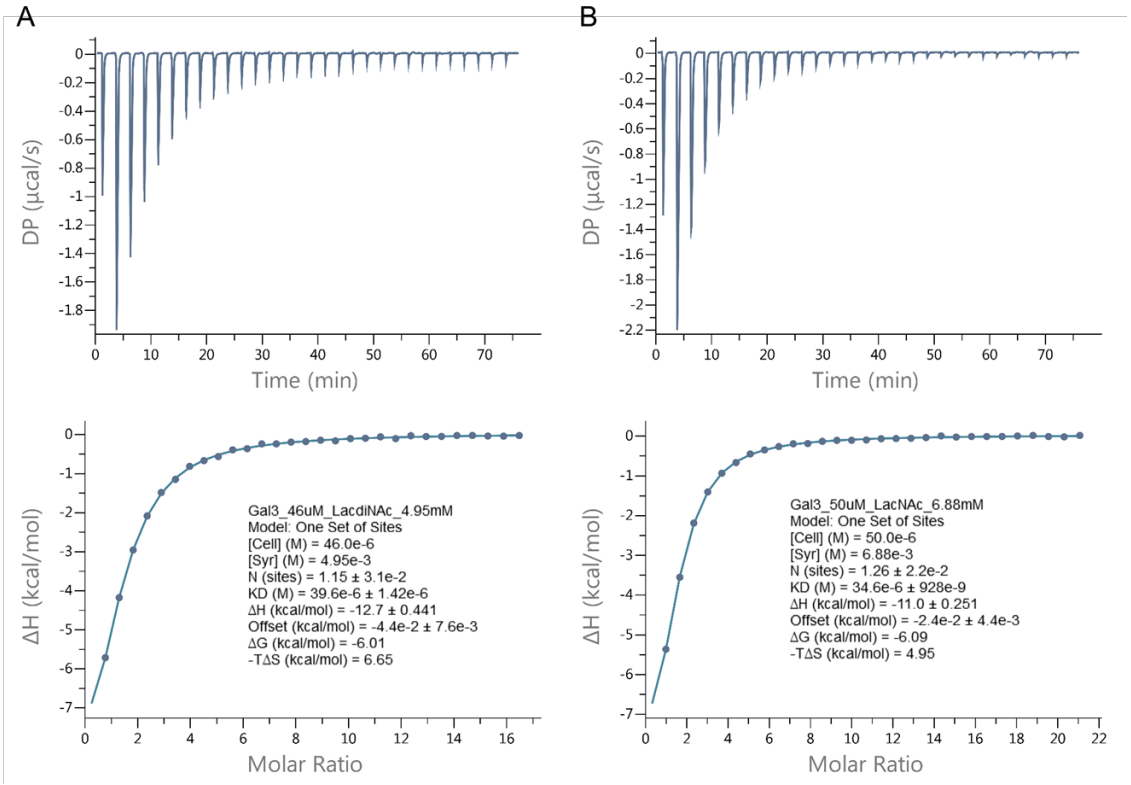

Figure 4.22 *- Representative ITC data for the titrations of Gal-3 (46 – 50 µM) with* ***A.*** *LacdiNAc (4.95mM); and* ***B.*** *LacNAc (6.88 mM) in phosphate buffer (20 mM) containing 150 mM NaCl, and 10 mM β-mercaptoethanol at pH 6.5. Injections were performed every 150 s at 298 K. Top: raw data for the titration of glycan into protein. Bottom: binding isotherm of the calorimetric titration shown on top. The continuous line represents the least-squares fit of the data to a single-site binding model.*

Table 4.3 *- Thermodynamic parameters for the binding of LacdiNAc and LacNAc to Gal-3 as determined by ITC experiments. Data fitted to single-site binding model. Average values and standard deviation were indicated.*

	Gal-3/LacdiNAc	Gal-3/LacNAc
K_D (μM)	38.1 ± 2.2	34.4 ± 0.3
ΔG (kcal/mol)	-6.04 ± 0.04	-6.09 ± 0.0
ΔH (kcal/mol)	-12.3 ± 0.6	-11.25 ± 0.4
-TΔS (kcal/mol)	6.24 ± 0.6	5.18 ± 0.3
N (sites)	1.17 ± 0.03	1.26 ± 0.01

The data obtained for both complexes fitted to a single-site binding model and the K_D estimated for both ligands are very similar. With respect to the thermodynamic values, LacdiNAc shows a slightly better enthalpy but worst entropy in comparison with LacNAc.

4.5.2. $^{1}H,^{15}N$-HSQC titration experiments

The molecular determinants of Gal-3/LacdiNAc recognition process were further inspected by $^{1}H,^{15}N$-HSQC based titration experiments. Figure 4.23 shows an overlay of four spectra of Gal-3 in the absence (black) and in presence of different equivalents of the LacdiNAc: Gal-3/LacdiNAc 1:1 (blue); Gal-3/LacdiNAc 1:5 (purple) and Gal-3/LacdiNAc 1:35 (green). For comparison purposes, the $^{1}H,^{15}N$-HSQC based titration of ^{15}N-labelled Gal-3 was also performed with LacNAc and Appendix E, Figure A6 and Figure A7 depicts the overlay of selected $^{1}H,^{15}N$-HSQC spectra recorded along with the titration of Gal-3 and LacNAc.

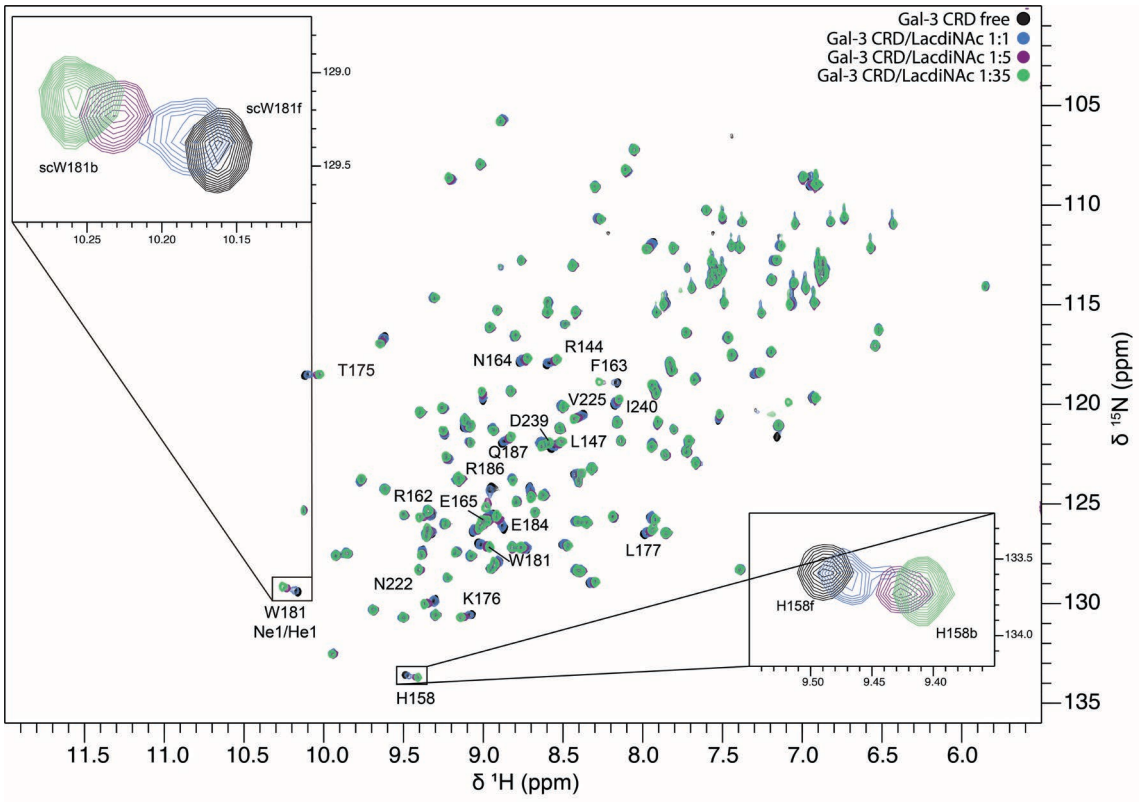

Figure 4.23 *- Overlay of the $^{1}H,^{15}N$-HSQC spectra of the titration of Gal-3 with LacdiNAc: Gal-3 (42 μM) in apo state (black), Gal-3 /LacdiNAc complex in presence of 1 equivalent of LacdiNAc (blue), Gal-3 /LacdiNAc complex in presence of 5 equivalents of LacdiNAc (purple) and Gal-3/LacdiNAc complex in presence of 35 equivalents of LacdiNAc (green). The bottom right and top left of the spectra present the expansion region of the H158 and W181 side chain cross-peaks, respectively.*

A clear progressive shift of the amide resonances of ^{15}N-labelled Gal-3 upon the addition of LacdiNAc is indicative of a fast exchange regime between the Gal-3 free and bound state at the NMR chemical shift scale.

To determine which residues were mostly affected by the presence of the LacdiNAc, it was elaborated a CSP histogram (see experimental section, Figure 4.24-A). The CSP histogram was also performed for Gal-3/LacNAc complex. The most affected residues deduced by CSP analysis of Gal-3/LacdiNAc and Gal-3/LacNAc were then mapped in the Gal-3 X-Ray structure (PDB code: 1KJL Gal-3 complexed with LacNAc)[52]. Figure 4.24-B highlights that the residues of the β-strands S4, S5 and S6 are both affected by the presence of LacdiNAc and LacNAc. However, in the case of Gal-3/LacdiNAc complex additional residues are affected in the β-strand S3. R144 and its neighbours N143 and I145 in the β-strand S3 are specifically affected in presence of LacdiNAc (Figure 4.24-C). Also, F163 at the loop connected to S4 and the D239 and the I240 on the loop connected to S2, are disturbed (Figure 4.24-C).

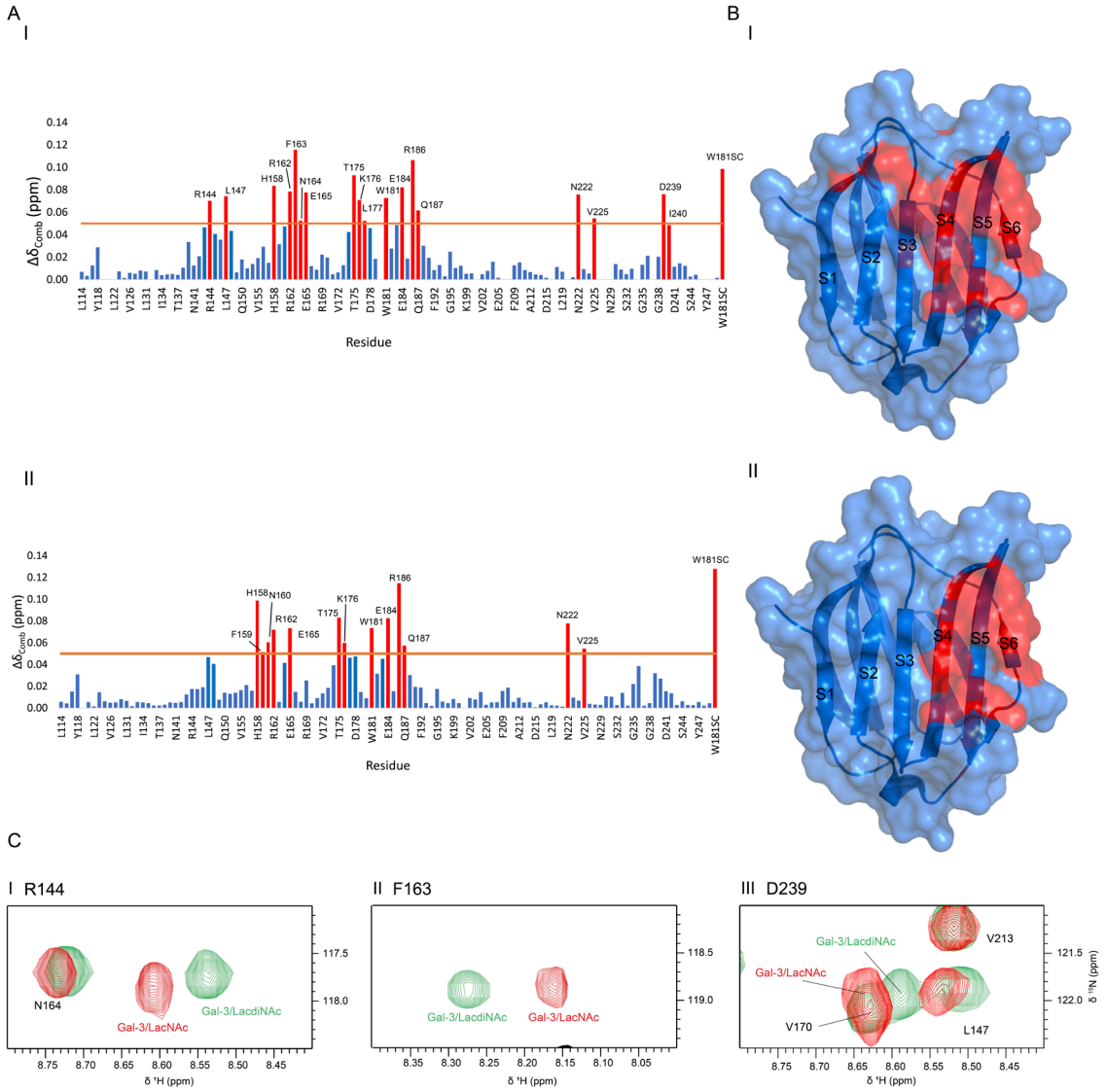

Figure 4.24 *- Gal-3 backbone amide chemical shift variations upon LacdiNAc and LacNAc binding.* ***A.*** *Histogram of the $\Delta\delta_{Comb}$ for each amino acid of Gal-3/LacdiNAc (Panel I) and Gal-3/LacNAc (Panel II) complexes in the presence of 35 equivalents of LacdiNAc and LacNAc, respectively. A cut-off line at 0.05 ppm (arbitrary threshold) displayed as orange is considered to discriminate the residues that strongly shift in the presence of the ligands. The residues with $\Delta\delta_{Comb}$ > 0.05 ppm are displayed as red bars.* ***B.*** *Map of the residues perturbed upon LacdiNAc (Panel I) and LacNAc (Panel II) binding highlighted in the 3D X-Ray structure of Gal-3 (PDB ID:1KJL). The amino acids most affected in the presence of the ligands ($\Delta\delta_{Comb}$ > 0.05 ppm) are displayed in red in the structure.* ***C.*** *Overlay of regions of the $^{1}H,^{15}N$ HSQC spectra of Gal-3 /LacdiNAc complex (green) and Gal-3/LacNAc complex (red), of signals of the amino acids with higher differences between the two bound states. I) R144; II) F163; III) D239.*

4.5.3. X-ray crystallography data

The Gal-3/LacdiNAc complex was further studied by X-Ray crystallography in collaboration with Dr. Ana Luisa Carvalho of the X-Ray crystallography group at UCIBIO unit of FCT-NOVA. The structure of Gal-3/LacdiNAc complex was compared with the corresponding Gal-3/LacNAc complex (Figure 4.25).

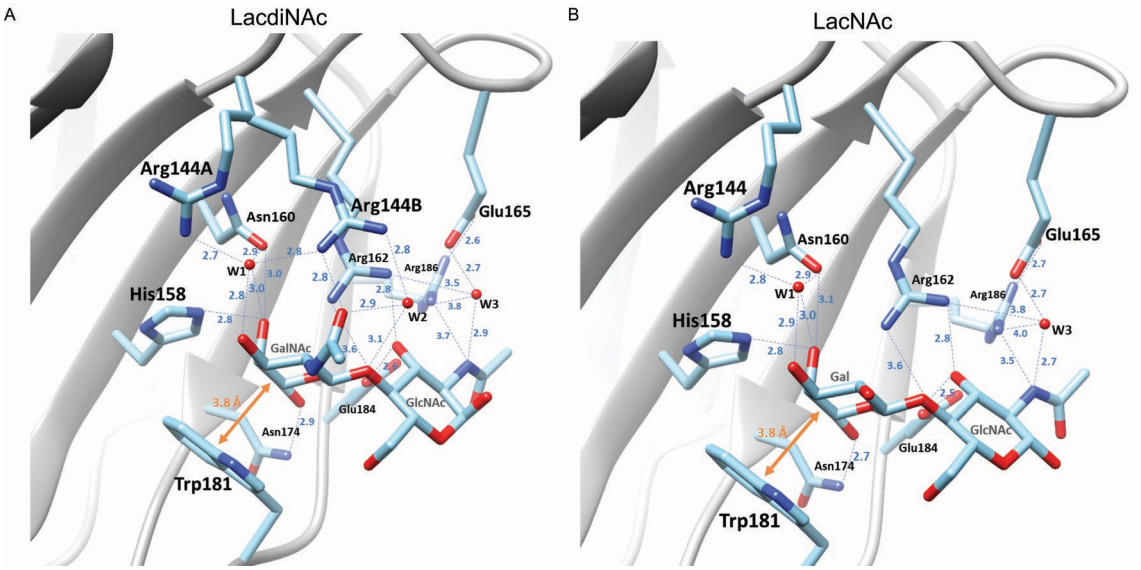

Figure 4.25 *- 3D view of the binding site of Gal-3 complexes deduced by X-Ray crystallography.* ***A.*** *Gal-3/LacdiNAc (PDB code: 7BE3; 1.25 Å).* ***B.*** *Gal-3 CRD/LacNAc (PDB code: 1KJL; 1.40 Å). The ligands and the amino acids of Gal-3 at the binding site are highlighted and displayed as sticks. Water molecules are shown as red spheres. The distances between the side chains of Gal-3 amino acids and the sugar units (GalNAc/Gal and GlcNAc) are also identified in the models.*

The key amino acids typically involved in the recognition of β-galactosides by galectin-3 are conserved in both complexes. As an example, the H158 residue, conserved among galectins, participates in hydrogen bonding with OH-4 of GalNAc of LacdiNAc and Gal in the case of LacNAc. Along with H158 there is the side chain of W181 that mediates a CH-π stacking interaction with H-4 and H-5 of Gal in LacNAc and GalNAc in LacdiNAc. Other common interaction is the hydrogen bond between the side chain R186 and the GlcNAc moiety present in both complexes. In both complexes, the GlcNAc unit is more solvent exposed than GalNAc/Gal in LacdiNAc and LacNAc.

Besides all these common features, it seems that the presence of LacdiNAc triggers an alternate conformation for the side chain of R144, between R144A (occ. 43%) and R144B (occ. 57%). The R144B conformation establishes an additional hydrogen-bond with the N-acetyl group of GalNAc of LacdiNAc, absent in the case of LacNAc. Furthermore, R144B increase the

residence of a water molecule, W2, also absent in the case of Gal-3/LacNAc complex. In turn, W2 mediate a hydrogen-bond network between the GalNAc N-acetyl group and the glycosidic bond of LacdiNAc and R144B. These additional contacts involving GalNAc of LacdiNAc might explain the slightly favoured enthalpy deduced by ITC measurements of LacdiNAc over LacNAc. In addition, the W2 detected on the structure deduced by X-Ray might explain the modest increase of the entropy penalty observed of LacdiNAc in comparison to LacNAc. The alternate conformation of R144 side chain observed for LacdiNAc correlates with the variations detected on the $^{1}H,^{15}N$-HSQC on the amide resonances of R144 and its neighbours N143, I145, D239, I240. In addition, the difference in the chemical shift of F163 can be tentatively explained by the alternate conformation of R169 side chain observed in the Gal-3/LacdiNAc model (Figure 4.26).

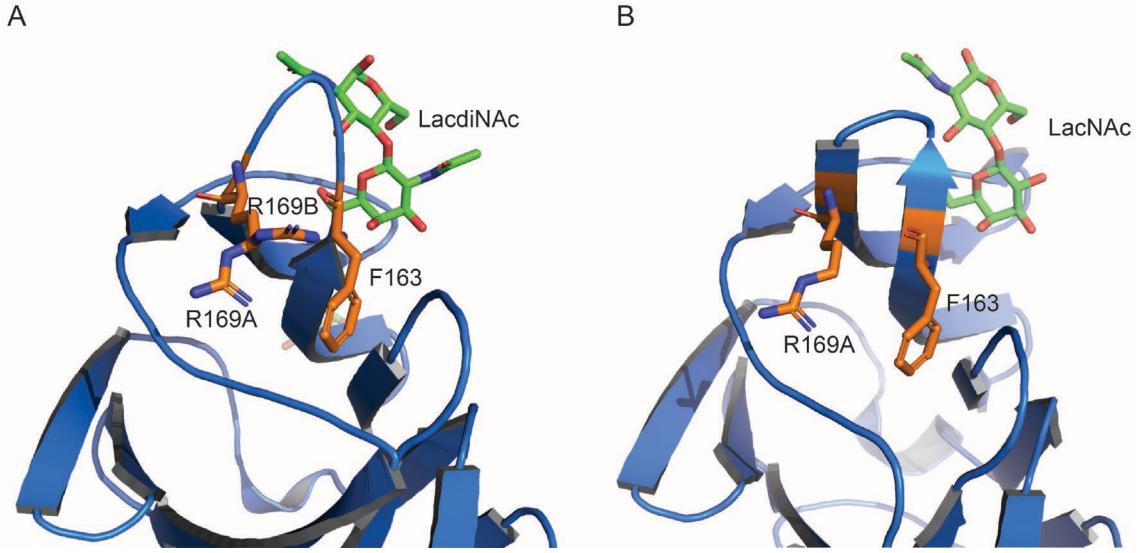

Figure 4.26 *- 3D view of the binding site obtained by X-Ray crystallography of the complexes* ***A.*** *Gal-3/LacdiNAc and* ***B.*** *Gal-3/LacNAc. R169 presents an additional side chain conformation in the Gal-3/LacdiNAc that can probably explain the alteration of chemical environment of F163 detected on the $^{1}H,^{15}N$-HSQC of Gal-3/LacdiNAc complex.*

Finally, in terms of affinity, Gal-3 does not discriminate LacNAc and LacdiNAc structures and only subtle differences between the two complexes were detected.

5. Conclusions and Future Work

Understanding the recognition process at an atomic level of specific glycan motifs by lectins enrolled in tumour immune surveillance opens the door for the development of glycan-based cancer therapies and new tools for diagnosis.

Herein, new structural insights of two cancer-related glycan structures, the LacdiNAc (GalNAcβ1-4GlcNAc) and the LacNAc (Galβ1-4GlcNAc) structures by two lectins, the human macrophage galactose-type lectin (MGL) and human galectin-3 (Gal-3) were unveiled, through a multidisciplinary approach that combines NMR spectroscopy, isothermal titration calorimetry (ITC), molecular dynamics (MD) and X-ray crystallography. The difference between the two structures relies on the functional group at C-2, an OH or NHAc, in the non-reducing end of the disaccharides that turns a Gal in LacNAc into a GalNAc in LacdiNAc. Besides both lectins recognize galactoside derivatives, the mechanism of binding of MGL and Gal-3 to LacdiNAc and LacNAc is significantly divergent and in line with the differences in their 3D architecture.

Our results highlight that MGL is a very selective and specific protein towards GalNAc-containing ligands. STD-NMR and MDs clearly show that MGL binds LacdiNAc and LacNAc through GalNAc and Gal units and where the GlcNAc moiety is exposed to the solvent. Nevertheless, the preference of MGL towards LacdiNAc rather than LacNAc is notable. The N-acetyl group of GalNAc unit in LacdiNAc performs additional interactions with the side chains of Y236 and H286, which in turn become the complex of MGL/LacdiNAc much more stable and of higher affinity complex (K_D of 5.3 ± 0.4 μM) than MGL/LacNAc counterpart (K_D > 1380 μM of 4.0 ± 0.3 mM according to NMR). Furthermore, the $^{1}H,^{15}N$-HSQC based titrations point out that the interactions involving NHAc moiety in LacdiNAc have a greater impact on the structure and dynamics of MGL than LacNAc.

On the other hand, Galectin-3 recognizes both LacdiNAc and LacNAc with similar affinity (38.1 and 34.4 μM, respectively). GalNAc and Gal unit in LacdiNAc and LacNAc ligands, respectively establish both the H-bond (OH-4 of GalNAc/Gal and His158) and CH-pi (H4 and H5 of GalNAc/Gal and W181) interactions conserved in all galectins. In addition, in both complexes Gal-3/LacdiNAc and Gal-3/LacNAc, the GlcNAc moiety establish similar interactions with Gal-3. The involvement of GlcNAc unit in Gal-3 is an issue that contrast to the MGL mechanism of binding. The major difference between Gal-3/LacdiNAc and Gal-3/LacNAc complexes relies on the alternate conformation of R144, between R144A (occ. 43%) and R144B (occ. 57%), only observed in the case of Gal-3/LacdiNAc complex. The R144B conformation establishes direct and water-mediated hydrogen-bonds with the N-acetyl group of GalNAc of LacdiNAc, which are absent in the case of LacNAc.

For future work, the interactions with these lectins could be studied with more complex sugars, resembling, even more, the structures that are present in healthy and cancerous cells. This way, we will be one step closer to better understand the recognition mechanism of these lectins and to exploit this knowledge in the creation of alternative tools for cancer diagnosis and anti-cancer therapies based on glycans.

6. References

(1) Ohtsubo, K.; Marth, J. D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* **2006**, *126* (5), 855–867. https://doi.org/10.1016/j.cell.2006.08.019.

(2) Schnaar, R. L.; Gerardy-Schahn, R.; Hildebrandt, H. Sialic Acids in the Brain: Gangliosides and Polysialic Acid in Nervous System Development, Stability, Disease, and Regeneration. *Physiol. Rev.* **2014**, *94* (2), 461–518. https://doi.org/10.1152/physrev.00033.2013.

(3) Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J. Glycosylation in Health and Disease. *Nat. Rev. Nephrol.* **2019**, *15* (6), 346–366. https://doi.org/10.1038/s41581-019-0129-4.

(4) Spiro, R. G. Protein Glycosylation: Nature, Distribution, Enzymatic Formation, and Disease Implications of Glycopeptide Bonds. *Glycobiology* **2002**, *12* (4), 43R-56R. https://doi.org/10.1093/glycob/12.4.43R.

(5) Rini, J. M.; Esko, J. D. Glycosyltransferases and Glycan-Processing Enzymes. In *Essentials of Glycobiology*; Ch. 6; Varki, A., Cummings, R. D., Esko, J., Eds.; Cold Spring Harbor Laboratory Press, **2015**. https://doi.org/10.1101/glycobiology.3e.006.

(6) Tuccillo, F. M.; de Laurentiis, A.; Palmieri, C.; Fiume, G.; Bonelli, P.; Borrelli, A.; Tassone, P.; Scala, I.; Buonaguro, F. M.; Quinto, I.; Scala, G. Aberrant Glycosylation as Biomarker for Cancer: Focus on CD43. *Biomed Res. Int.* **2014**, *2014*, 1–13. https://doi.org/10.1155/2014/742831.

(7) Saeland, E.; van Vliet, S. J.; Bäckström, M.; van den Berg, V. C. M.; Geijtenbeek, T. B. H.; Meijer, G. A.; van Kooyk, Y. The C-Type Lectin MGL Expressed by Dendritic Cells Detects Glycan Changes on MUC1 in Colon Carcinoma. *Cancer Immunol. Immunother.* **2007**, *56* (8), 1225–1236. https://doi.org/10.1007/s00262-006-0274-z.

(8) Freire, T.; Zhang, X.; Dériaud, E.; Ganneau, C.; Vichier-Guerre, S.; Azria, E.; Launay, O.; Lo-Man, R.; Bay, S.; Leclerc, C. Glycosidic Tn-Based Vaccines Targeting Dermal Dendritic Cells Favor Germinal Center B-Cell Development and Potent Antibody Response in the Absence of Adjuvant. *Blood* **2010**, *116* (18), 3526–3536. https://doi.org/10.1182/blood-2010-04-279133.

(9) Cornelissen, L.; Van Vliet, S. A Bitter Sweet Symphony: Immune Responses to Altered O-Glycan Epitopes in Cancer. *Biomolecules* **2016**, *6* (2), 26. https://doi.org/10.3390/biom6020026.

(10) Nonaka, M.; Fukuda, M. Expression and Function of Poly-N-Acetyllactosamine Type Glycans in Cancer. In *Glycosignals in Cancer: Mechanisms of Malignant Phenotypes*; Ch. 9; Furukawa, K., Fukuda, M., Eds.; Springer Japan: Tokyo, **2016**. https://doi.org/10.1007/978-4-431-55939-9_9.

(11) Elola, M. T.; Blidner, A. G.; Ferragut, F.; Bracalente, C.; Rabinovich, G. A. Assembly, Organization and Regulation of Cell-Surface Receptors by Lectin–Glycan Complexes. *Biochem. J.* **2015**, *469* (1), 1–16. https://doi.org/10.1042/BJ20150461.

(12) Yu, H.; Li, X.; Chen, M.; Zhang, F.; Liu, X.; Yu, J.; Zhong, Y.; Shu, J.; Chen, W.; Du, H.; Zhang, K.; Zhang, C.; Zhang, J.; Xie, H.; Li, Z. Integrated Glycome Strategy for Characterization of Aberrant LacNAc Contained N-Glycans Associated With Gastric Carcinoma. *Front. Oncol.* **2019**, *9* (636), 1–11. https://doi.org/10.3389/fonc.2019.00636.

(13) van den Berg, T. K.; Honing, H.; Franke, N.; van Remoortere, A.; Schiphorst, W. E. C. M.; Liu, F.-T.; Deelder, A. M.; Cummings, R. D.; Hokke, C. H.; van Die, I. LacdiNAc-Glycans Constitute a Parasite Pattern for Galectin-3-Mediated Immune Recognition. *J. Immunol.* **2004**, *173* (3), 1902–1907. https://doi.org/10.4049/jimmunol.173.3.1902.

(14) Haji-Ghassemi, O.; Gilbert, M.; Spence, J.; Schur, M. J.; Parker, M. J.; Jenkins, M. L.; Burke, J. E.; van Faassen, H.; Young, N. M.; Evans, S. V. Molecular Basis for Recognition of the Cancer Glycobiomarker, LacdiNAc (GalNAc[B1→4]GlcNAc), by Wisteria Floribunda Agglutinin. *J. Biol. Chem.* **2016**, *291* (46), 24085–24095. https://doi.org/10.1074/jbc.M116.750463.

(15) Hirano, K.; Matsuda, A.; Shirai, T.; Furukawa, K. Expression of LacdiNAc Groups on N-Glycans among Human Tumors Is Complex. *Biomed Res. Int.* **2014**, *2014*, 1–7. https://doi.org/10.1155/2014/981627.

(16) Kawar, Z. S.; Haslam, S. M.; Morris, H. R.; Dell, A.; Cummings, R. D. Novel Poly-GalNAcβ1–4GlcNAc (LacdiNAc) and Fucosylated Poly-LacdiNAc N-Glycans from Mammalian Cells Expressing B1,4- N-Acetylgalactosaminyltransferase and A1,3-Fucosyltransferase. *J. Biol. Chem.* **2005**, *280* (13), 12810–12819. https://doi.org/10.1074/jbc.M414273200.

(17) Weinbaum, S.; Tarbell, J. M.; Damiano, E. R. The Structure and Function of the Endothelial Glycocalyx Layer. *Annu. Rev. Biomed. Eng.* **2007**, *9* (1), 121–167. https://doi.org/10.1146/annurev.bioeng.9.060906.151959.

(18) Varki, A. Biological Roles of Glycans. *Glycobiology* **2017**, *27* (1), 3–49. https://doi.org/10.1093/glycob/cww086.

(19) Jiménez-Barbero, J.; L. Asensio, J.; Cuevas, G.; Canales, A.; Fernández-Alonso, M. C.; Javier Cañada, F. Conformational Insights on the Molecular Recognition Processes of Carbohydrate Molecules by Proteins and Enzymes: A 3D View by Using NMR. *Biocatal. Biotransformation* **2006**, *24* (1–2), 13–22. https://doi.org/10.1080/10242420600598103.

(20) 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.

(21) Zeng, X.; Andrade, C. A. S.; Oliveira, M. D. L.; Sun, X.-L. Carbohydrate–Protein Interactions and Their Biosensing Applications. *Anal. Bioanal. Chem.* **2012**, *402* (10), 3161–3176. https://doi.org/10.1007/s00216-011-5594-y.

(22) Quiocho, F. A. Protein-Carbohydrate Interactions: Basic Molecular Features. *Pure Appl. Chem.* **1989**, *61* (7), 1293–1306. https://doi.org/10.1351/pac198961071293.

(23) Spiwok, V. CH/π Interactions in Carbohydrate Recognition. *Molecules* **2017**, *22* (7), 1038. https://doi.org/10.3390/molecules22071038.

(24) 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.* **2012**, *13* (8), 816–830. https://doi.org/10.2174/138920312804871175.

(25) Ramírez-Gualito, K.; Alonso-Ríos, R.; Quiroz-García, B.; Rojas-Aguilar, A.; Díaz, D.; Jiménez-Barbero, J.; Cuevas, G. Enthalpic Nature of the CH/π Interaction Involved in the Recognition of Carbohydrates by Aromatic Compounds, Confirmed by a Novel Interplay of NMR, Calorimetry, and Theoretical Calculations. *J. Am. Chem. Soc.* **2009**, *131* (50), 18129–18138. https://doi.org/10.1021/ja903950t.

(26) Kiessling, L. L.; Young, T.; Mortell, K. H.; Gruber, T. D. Multivalency in Protein-Carbohydrate Recognition. In *Glycoscience*; B, F.-R., K, T., J, T., Eds.; Springer-Verlag Berlin Heidelberg: Berlin, Heidelberg, 2008; pp 2484–2523. https://doi.org/10-1007/978-3-540-30429-6_64.

(27) Roldós, V.; Cañada, F. J.; Jiménez-Barbero, J. Carbohydrate-Protein Interactions: A 3D View by NMR. *ChemBioChem* **2011**, *12* (7), 990–1005. https://doi.org/10.1002/cbic.201000705.

(28) Mason, C.; Tarr, A. Human Lectins and Their Roles in Viral Infections. *Molecules* **2015**, *20* (2), 2229–2271. https://doi.org/10.3390/molecules20022229.

(29) 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.

(30) Vasta, G. R.; Ahmed, H. *Animal Lectins*; Ch. 1; Vasta PhD, G. R., Ahmed PhD, H., Eds.; CRC Press, **2008**. https://doi.org/10.1201/9781420006971.

(31) Zelensky, A. N.; Gready, J. E. The C-Type Lectin-like Domain Superfamily. *FEBS J.* **2005**,

272 (24), 6179–6217. https://doi.org/10.1111/j.1742-4658.2005.05031.x.

(32) CLEC10A - C-type lectin domain family 10 member A - *Homo sapiens* (Human) - CLEC10A gene & protein https://www.uniprot.org/uniprot/Q8IUN9#Q8IUN9-1 (accessed Nov 27, 2020).

(33) Jégouzo, S. A.; Quintero-Martínez, A.; Ouyang, X.; dos Santos, Á.; Taylor, M. E.; Drickamer, K. Organization of the Extracellular Portion of the Macrophage Galactose Receptor: A Trimeric Cluster of Simple Binding Sites for N-Acetylgalactosamine. *Glycobiology* **2013**, *23* (7), 853–864. https://doi.org/10.1093/glycob/cwt022.

(34) Drickamer, K.; Taylor, M. E. Recent Insights into Structures and Functions of C-Type Lectins in the Immune System. *Curr. Opin. Struct. Biol.* **2015**, *34*, 26–34. https://doi.org/10.1016/j.sbi.2015.06.003.

(35) Meier, M.; Bider, M. D.; Malashkevich, V. N.; Spiess, M.; Burkhard, P. Crystal Structure of the Carbohydrate Recognition Domain of the H1 Subunit of the Asialoglycoprotein Receptor. *J. Mol. Biol.* **2000**, *300* (4), 857–865. https://doi.org/10.1006/jmbi.2000.3853.

(36) Marcelo, F.; Garcia-Martin, F.; Matsushita, T.; Sardinha, J.; Coelho, H.; Oude-Vrielink, A.; Koller, C.; André, S.; Cabrita, E. J.; Gabius, H.-J.; Nishimura, S.-I.; Jiménez-Barbero, J.; Cañada, F. J. Delineating Binding Modes of Gal/GalNAc and Structural Elements of the Molecular Recognition of Tumor-Associated Mucin Glycopeptides by the Human Macrophage Galactose-Type Lectin. *Chem. - A Eur. J.* **2014**, *20* (49), 16147–16155. https://doi.org/10.1002/chem.201404566.

(37) Diniz, A.; Coelho, H.; Dias, J. S.; Vliet, S. J.; Jiménez-Barbero, J.; Corzana, F.; Cabrita, E. J.; Marcelo, F. The Plasticity of the Carbohydrate Recognition Domain Dictates the Exquisite Mechanism of Binding of Human Macrophage Galactose-Type Lectin. *Chem. – A Eur. J.* **2019**, *25* (61), 13945–13955. https://doi.org/10.1002/chem.201902780.

(38) van Kooyk, Y.; Ilarregui, J. M.; van Vliet, S. J. Novel Insights into the Immunomodulatory Role of the Dendritic Cell and Macrophage-Expressed C-Type Lectin MGL. *Immunobiology* **2015**, *220* (2), 185–192. https://doi.org/10.1016/j.imbio.2014.10.002.

(39) van Vliet, S. J.; Saeland, E.; van Kooyk, Y. Sweet Preferences of MGL: Carbohydrate Specificity and Function. *Trends Immunol.* **2008**, *29* (2), 83–90. https://doi.org/10.1016/j.it.2007.10.010.

(40) van Vliet, S. J.; Gringhuis, S. I.; Geijtenbeek, T. B. H.; van Kooyk, Y. Regulation of Effector T Cells by Antigen-Presenting Cells via Interaction of the C-Type Lectin MGL with CD45. *Nat. Immunol.* **2006**, *7* (11), 1200–1208. https://doi.org/10.1038/ni1390.

(41) Zaal, A.; Li, R. J. E.; Lübbers, J.; Bruijns, S. C. M.; Kalay, H.; van Kooyk, Y.; van Vliet, S. J. Activation of the C-Type Lectin MGL by Terminal GalNAc Ligands Reduces the Glycolytic Activity of Human Dendritic Cells. *Front. Immunol.* **2020**, *11* (305), 1–13. https://doi.org/10.3389/fimmu.2020.00305.

(42) Huang, J.; Liang, J.-T.; Huang, H.-C.; Shen, T.-L.; Chen, H.-Y.; Lin, N.-Y.; Che, M.-I.; Lin, W.-C.; Huang, M.-C. β1,4-N-Acetylgalactosaminyltransferase III Enhances Malignant Phenotypes of Colon Cancer Cells. *Mol. Cancer Res.* **2007**, *5* (6), 543–552. https://doi.org/10.1158/1541-7786.MCR-06-0431.

(43) Lenos, K.; Goos, J. A. C. M.; Vuist, I. M.; den Uil, S. H.; Delis-van Diemen, P. M.; Belt, E. J. T.; Stockmann, H. B. A. C.; Bril, H.; de Wit, M.; Carvalho, B.; Giblett, S.; Pritchard, C. A.; Meijer, G. A.; van Kooyk, Y.; Fijneman, R. J. A.; van Vliet, S. J. MGL Ligand Expression Is Correlated to BRAF Mutation and Associated with Poor Survival of Stage III Colon Cancer Patients. *Oncotarget* **2015**, *6* (28), 26278–26290. https://doi.org/10.18632/oncotarget.4495.

(44) van Vliet, S. J.; Bay, S.; Vuist, I. M.; Kalay, H.; García-Vallejo, J. J.; Leclerc, C.; van Kooyk, Y. MGL Signaling Augments TLR2-Mediated Responses for Enhanced IL-10 and TNF-α Secretion. *J. Leukoc. Biol.* **2013**, *94* (2), 315–323. https://doi.org/10.1189/jlb.1012520.

(45) Marcelo, F.; Supekar, N.; Corzana, F.; Van Der Horst, J. C.; Vuist, I. M.; Live, D.; Boons, G. J. P. H.; Smith, D. F.; Van Vliet, S. J. Identification of a Secondary Binding Site in Human Macrophage Galactose-Type Lectin by Microarray Studies: Implications for the Molecular Recognition of Its Ligands. *J. Biol. Chem.* **2019**, *294* (4), 1300–1311. https://doi.org/10.1074/jbc.RA118.004957.

(46) Ludwig, A.-K.; Michalak, M.; Xiao, Q.; Gilles, U.; Medrano, F. J.; Ma, H.; FitzGerald, F. G.; Hasley, W. D.; Melendez-Davila, A.; Liu, M.; Rahimi, K.; Kostina, N. Y.; Rodriguez-Emmenegger, C.; Möller, M.; Lindner, I.; Kaltner, H.; Cudic, M.; Reusch, D.; Kopitz, J.; Romero, A.; Oscarson, S.; Klein, M. L.; Gabius, H.-J.; Percec, V. Design–Functionality Relationships for Adhesion/Growth-Regulatory Galectins. *Proc. Natl. Acad. Sci.* **2019**, *116* (8), 2837–2842. https://doi.org/10.1073/pnas.1813515116.

(47) 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.

(48) Halimi, H.; Rigato, A.; Byrne, D.; Ferracci, G.; Sebban-Kreuzer, C.; ElAntak, L.; Guerlesquin, F. Glycan Dependence of Galectin-3 Self-Association Properties. *PLoS One* **2014**, *9* (11), 1–9. https://doi.org/10.1371/journal.pone.0111836.

(49) Dong, R.; Zhang, M.; Hu, Q.; Zheng, S.; Soh, A.; Zheng, Y.; Yuan, H. Galectin-3 as a Novel Biomarker for Disease Diagnosis and a Target for Therapy (Review). *Int. J. Mol.*

Med. **2017**, *41* (2), 599–614. https://doi.org/10.3892/ijmm.2017.3311.

(50) Newlaczyl, A. U.; Yu, L. Galectin-3 – A Jack-of-All-Trades in Cancer. *Cancer Lett.* **2011**, *313* (2), 123–128. https://doi.org/10.1016/j.canlet.2011.09.003.

(51) Bumba, L.; Laaf, D.; Spiwok, V.; Elling, L.; Křen, V.; Bojarová, P. Poly-N-Acetyllactosamine Neo-Glycoproteins as Nanomolar Ligands of Human Galectin-3: Binding Kinetics and Modeling. *Int. J. Mol. Sci.* **2018**, *19* (2), 372. https://doi.org/10.3390/ijms19020372.

(52) Sörme, P.; Arnoux, P.; Kahl-Knutsson, B.; Leffler, H.; Rini, J. M.; Nilsson, U. J. Structural and Thermodynamic Studies on Cation−π Interactions in Lectin−Ligand Complexes: High-Affinity Galectin-3 Inhibitors through Fine-Tuning of an Arginine−Arene Interaction. *J. Am. Chem. Soc.* **2005**, *127* (6), 1737–1743. https://doi.org/10.1021/ja043475p.

(53) 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.

(54) Gimeno, A.; Reichardt, N.-C.; Cañada, F. J.; Perkams, L.; Unverzagt, C.; Jiménez-Barbero, J.; Ardá, A. NMR and Molecular Recognition of N-Glycans: Remote Modifications of the Saccharide Chain Modulate Binding Features. *ACS Chem. Biol.* **2017**, *12* (4), 1104–1112. https://doi.org/10.1021/acschembio.6b01116.

(55) Valverde, P.; Quintana, J. I.; Santos, J. I.; Ardá, A.; Jiménez-Barbero, J. Novel NMR Avenues to Explore the Conformation and Interactions of Glycans. *ACS Omega* **2019**, *4* (9), 13618–13630. https://doi.org/10.1021/acsomega.9b01901.

(56) Cummings, R. D.; Schnaar, R. L.; Esko, J. D.; Drickamer, K.; Taylor, M. E. Principles of Glycan Recognition. In *Essentials of Glycobiology*; Ch. 29; Varki, A., Cummings, R. D., Esko, J., Eds.; Cold Spring Harbor Laboratory Press, **2015**. https://doi.org/10.1101/glycobiology.3e.029.

(57) Gao, C.; Wei, M.; McKitrick, T. R.; McQuillan, A. M.; Heimburg-Molinaro, J.; Cummings, R. D. Glycan Microarrays as Chemical Tools for Identifying Glycan Recognition by Immune Proteins. *Front. Chem.* **2019**, *7* (833), 1–18. https://doi.org/10.3389/fchem.2019.00833.

(58) Rillahan, C. D.; Paulson, J. C. Glycan Microarrays for Decoding the Glycome. *Annu. Rev. Biochem.* **2011**, *80* (1), 797–823. https://doi.org/10.1146/annurev-biochem-061809-152236.

(59) Carvalho, A. L.; Santos-Silva, T.; Romao, M. J.; Cabrita, E. J.; Marcelo, F. Structural

Elucidation of Macromolecules. In *Essential Techniques for Medical and Life Scientists: A Guide to Contemporary Methods and Current Applications with the Protocols*; Ch. 2; BENTHAM SCIENCE PUBLISHERS, **2018**; https://doi.org/10.2174/9781681087092118010005.

(60) Claridge, T. D. W. Protein–Ligand Screening by NMR. In *High-Resolution NMR Techniques in Organic Chemistry*; Ch. 11; Elsevier, **2016**. https://doi.org/10.1016/B978-0-08-099986-9.00011-7.

(61) 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.

(62) Xu, D.; Li, B.; Gao, J.; Liu, Z.; Niu, X.; Nshogoza, G.; Zhang, J.; Wu, J.; Su, X.-C.; He, W.; Ma, R.; Yang, D.; Ruan, K. Ligand Proton Pseudocontact Shifts Determined from Paramagnetic Relaxation Dispersion in the Limit of NMR Intermediate Exchange. *J. Phys. Chem. Lett.* **2018**, *9* (12), 3361–3367. https://doi.org/10.1021/acs.jpclett.8b01443.

(63) Waudby, C. A.; Ouvry, M.; Davis, B.; Christodoulou, J. Two-Dimensional NMR Lineshape Analysis of Single, Multiple, Zero and Double Quantum Correlation Experiments. *J. Biomol. NMR* **2020**, *74* (1), 95–109. https://doi.org/10.1007/s10858-019-00297-7.

(64) Schumann, F. H.; Riepl, H.; Maurer, T.; Gronwald, W.; Neidig, K.-P.; Kalbitzer, H. R. Combined Chemical Shift Changes and Amino Acid Specific Chemical Shift Mapping of Protein–Protein Interactions. *J. Biomol. NMR* **2007**, *39* (4), 275–289. https://doi.org/10.1007/s10858-007-9197-z.

(65) Viegas, A.; Sardinha, J.; Freire, F.; Duarte, D. F.; Carvalho, A. L.; Fontes, C. M. G. A.; Romão, M. J.; Macedo, A. L.; Cabrita, E. J. Solution Structure, Dynamics and Binding Studies of a Family 11 Carbohydrate-Binding Module from *Clostridium Thermocellum* (CtCBM11). *Biochem. J.* **2013**, *451* (2), 289–300. https://doi.org/10.1042/BJ20120627.

(66) 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.

(67) Haselhorst, T.; Lamerz, A.-C.; Itzstein, M. Von. Saturation Transfer Difference NMR Spectroscopy as a Technique to Investigate Protein-Carbohydrate Interactions in Solution. In *Glycomics*; Ch. 26; Humana Press: Totowa, NJ, **2009**; Vol. 534. https://doi.org/10.1007/978-1-59745-022-5_26.

(68) Venkitakrishnan, R. P.; Benard, O.; Max, M.; Markley, J. L.; Assadi-Porter, F. M. Use of NMR Saturation Transfer Difference Spectroscopy to Study Ligand Binding to Membrane

Proteins. In *Membrane Protein Structure and Dynamics*; Ch. 4; Vaidehi, N., Klein-Seetharaman, J., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, **2012**; Vol. 914, pp 47–63. https://doi.org/10.1007/978-1-62703-023-6_4.

(69) Yamamoto, K.; Ishida, C.; Shinohara, Y.; Hasegawa, Y.; Konami, Y.; Osawa, T.; Irimura, T. Interaction of Immobilized Recombinant Mouse C-Type Macrophage Lectin with Glycopeptides and Oligosaccharides. *Biochemistry* **1994**, *33* (26), 8159–8166. https://doi.org/10.1021/bi00192a021.

(70) Nkari, W. K.; Prestegard, J. H. NMR Resonance Assignments of Sparsely Labeled Proteins: Amide Proton Exchange Correlations in Native and Denatured States. *J. Am. Chem. Soc.* **2009**, *131* (14), 5344–5349. https://doi.org/10.1021/ja8100775.

(71) Pajon, A.; Ionides, J.; Diprose, J.; Fillon, J.; Fogh, R.; Ashton, A. W.; Berman, H.; Boucher, W.; Cygler, M.; Deleury, E.; Esnouf, R.; Janin, J.; Kim, R.; Krimm, I.; Lawson, C. L.; Oeuillet, E.; Poupon, A.; Raymond, S.; Stevens, T.; van Tilbeurgh, H.; Westbrook, J.; Wood, P.; Ulrich, E.; Vranken, W.; Xueli, L.; Laue, E.; Stuart, D. I.; Henrick, K. Design of a Data Model for Developing Laboratory Information Management and Analysis Systems for Protein Production. *Proteins Struct. Funct. Bioinforma.* **2004**, *58* (2), 278–284. https://doi.org/10.1002/prot.20303.

(72) Vonrhein, C.; Flensburg, C.; Keller, P.; Sharff, A.; Smart, O.; Paciorek, W.; Womack, T.; Bricogne, G. Data Processing and Analysis with the AutoPROC Toolbox. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67* (4), 293–302. https://doi.org/10.1107/S0907444911007773.

(73) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *J. Appl. Crystallogr.* **2007**, *40* (4), 658–674. https://doi.org/10.1107/S0021889807021206.

(74) Saraboji, K.; Håkansson, M.; Genheden, S.; Diehl, C.; Qvist, J.; Weininger, U.; Nilsson, U. J.; Leffler, H.; Ryde, U.; Akke, M.; Logan, D. T. The Carbohydrate-Binding Site in Galectin-3 Is Preorganized To Recognize a Sugarlike Framework of Oxygens: Ultra-High-Resolution Structures and Water Dynamics. *Biochemistry* **2012**, *51* (1), 296–306. https://doi.org/10.1021/bi201459p.

(75) Adams, P. D.; Afonine, P. V; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX : A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66* (2), 213–221. https://doi.org/10.1107/S0907444909052925.

(76) Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D. Towards Automated Crystallographic Structure Refinement with Phenix.Refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2012**, *68* (4), 352–367. https://doi.org/10.1107/S0907444912001308.

(77) Moriarty, N. W.; Grosse-Kunstleve, R. W.; Adams, P. D. Electronic Ligand Builder and Optimization Workbench (eLBOW): A Tool for Ligand Coordinate and Restraint Generation. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2009**, *65* (10), 1074–1080. https://doi.org/10.1107/S0907444909029436.

(78) Rodríguez, E.; Schetters, S. T. T.; van Kooyk, Y. The Tumour Glyco-Code as a Novel Immune Checkpoint for Immunotherapy. *Nat. Rev. Immunol.* **2018**, *18* (3), 204–211. https://doi.org/10.1038/nri.2018.3.

7. Appendix

Appendix A – Protein Sequences

The amino acid one-letter code sequence of MGL-CRD (C181-H316) with the starting methionine residue:

MCPVNWVEHQDSCYWFSHSGMSWAEAEKYCQLKNAHLVVINSREEQNFVQKYLGSAYTWM
GLSDPEGAWKWVDGTDYATGFQNWKPGQPDDWQGHGLGGGEDCAHFHPDGRWNDDVCQ
RPYHWVCEAGLGQTSQESH

The amino acid one-letter code sequence of Gal-3 CRD (L114-I250) with the starting methionine residue:

MLIVPYNLPLPGGVVPRMLITILGTVKPNANRIALDFQRGNDVAFHFNPRFNENNRRVIVCNTKL
DNNWGREERQSVFPFESGKPFKIQVLVEPDHFKVAVNDAHLLQYNHRVKKLNEISKLGISGDID
LTSASYTMI

Appendix B – Ligand Assignment

Table A1 *- ^{1}H-NMR assignments of LacdiNAc and LacNAc in D2O Tris(D11)-DCl buffer solutions with pD 7.5 at 298K.*

LacdiNAc			LacNAc		
	^{1}H	ppm		^{1}H	ppm
βGalNAc	H1	4.43	βGal	H1	4.38
	H2	3.84		H2	3.44
	H3	3.68		H3	3.56
	H4	3.84		H4	3.87
	H5	3.65		H5	3.62
	H6	3.70-3.62		H6	3.71 - 3.66
	NHAc	1.95	βGlcNAc	H1	4.61
βGlcNAc	H1	4.61		H2	3.60
	H2	3.62		H3	3.49
	H3	3.45		H4	3.62
	H4	3.56		H5	3.85
	H5	3.77		H6	3.91-3.75
	H6	3.55		NHAc	1.97
	NHAc	1.92	αGlcNAc	H1	5.09
αGlcNAc	H1	5.11		H2	3.79
	H2	3.76		H3	3.70
	H3	3.62		H4	3.62
	H4	3.56		H5	3.79
	H5	3.77		H6	3.91-3.75
	H6	3.70		NHAc	1.97
	NHAc	1.92			

Appendix C – X-Ray Crystallography Data Collection and Refinement Statistics

Table A2 *- Data collection and structure refinement parameters and statistics.*

Statistics for the highest-resolution shell are shown in parentheses.

Data collection and processing	
Space group	$P2_1\,2_1\,2_1$
Unit-cell parameters (Å)	a = 36.0, b = 58.0, c = 62.3
Resolution range (Å)	31.14 - 1.25 (1.29 - 1.25)
Matthews coefficient, V_M (Å^{3}Da^{-1})	2.33
Solvent content (%)	47.2
Protein molecules per asymmetric unit	1
Mosaicity (°)	0.17
$I/\sigma(I)$	20.7 (2.0)
Wilson B-factor (Å^2)	14.34
R_{merge}†	0.047 (0.999)
$R_{p.i.m.}$+	0.018 (0.382)
Half-dataset correlation CC1/2	1.00 (0.88)
Multiplicity	7.8 (7.7)
Total reflections	287616 (27673)
Unique reflections	36869 (3611)
Completeness (%)	99.85 (99.61)
Refinement statistics	
Number of reflections	36843
Number of protein atoms	1149
Number of ligand atoms	36
Number of solvent molecules	184
R_{work}‡	0.170
R_{free}§	0.195
R.m.s.d. bond lengths (Å)	0.005
R.m.s.d. bond angles (°)	0.85
Average B-factor (Å^2)	
Protein	16.1
LacdiNAc	
GalNAc	19.5
GlcNAc	30.9
Water molecules	28.9
Ramachandran plot	
Residues in favoured regions (%)	97.8
Residues in allowed regions (%)	2.2
Residues outliers (%)	0.0
PDB accession code	7BE3

† $R_{merge} = \frac{\sum_{hkl}\sum_{i=1}^{n}|I_i(hkl)-\bar{I}(hkl)|}{\sum_{hkl}\sum_{i=1}^{n}I_i(hkl)}$, where I is the observed intensity, and $\bar{I}$ is the statistically-weighted average intensity of multiple observations. + $R_{p.i.m.} = \frac{\sum_{hkl}\sqrt{1/(n-1)}\sum_{i=1}^{n}|I_i(hkl)-\bar{I}(hkl)|}{\sum_{hkl}\sum_{i=1}^{n}I_i(hkl)}$, a redundancy-independent version of R_{merge}. ‡ $R_{work} = \frac{\sum_{hkl}||F_{obs}(hkl)|-|F_{calc}(hkl)||}{\sum_{hkl}|F_{obs}(hkl)|}$, where $|F_{calc}|$ and $|F_{obs}|$ are the calculated and observed structure factor amplitudes, respectively. § R_{free} is calculated for a randomly chosen 5% of the reflections.

Appendix D – STD-NMR

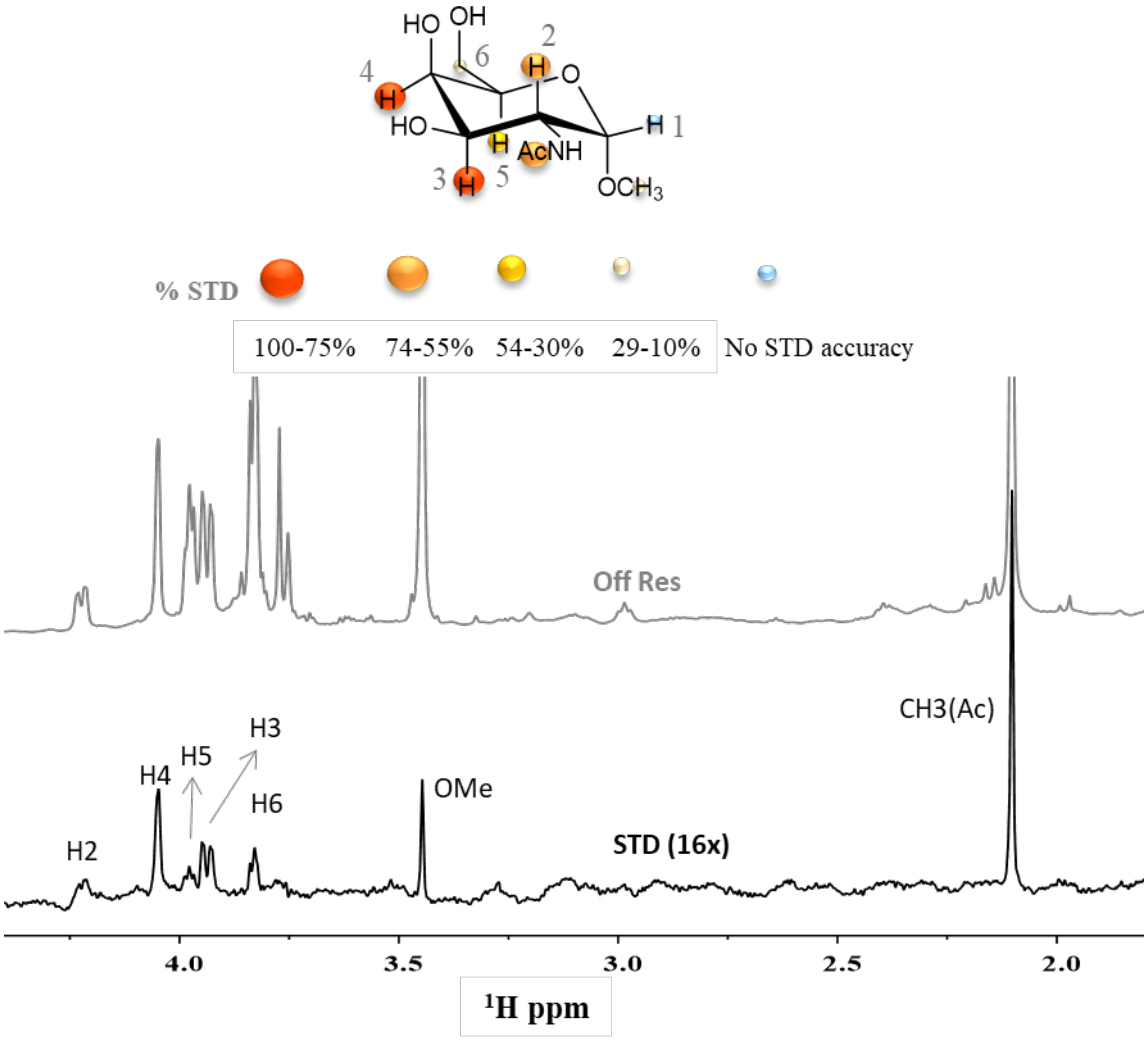

Figure A1 *- STD-NMR experiments for α-Me-GalNAc (1.1 mM) with MGL (30 μM) carried out at 600MHz, at 310K with irradiation at 7 ppm. The reference spectrum is displayed in gray color while the STD spectrum (STD) is displayed in black. The key proton resonances that have STD response are marked in the off-resonance spectrum. On top of the spectra is the epitope mapping obtained for α-Me-GalNAc with MGL.*

Appendix E – $^{1}H,^{15}N$-HSQC Titrations

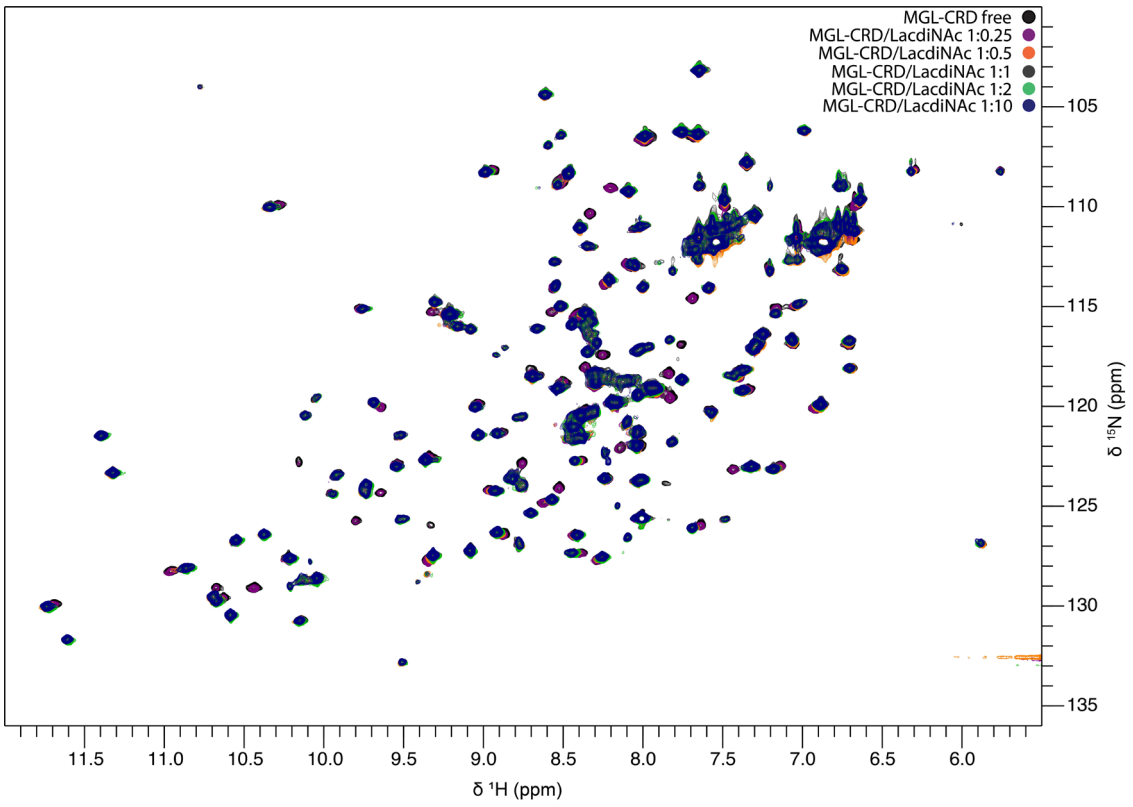

Figure A2 *- Overlay of the $^{1}H,^{15}N$-HSQC spectra of the titration of MGL-CRD (200 μM) with LacdiNAc at 293K.*

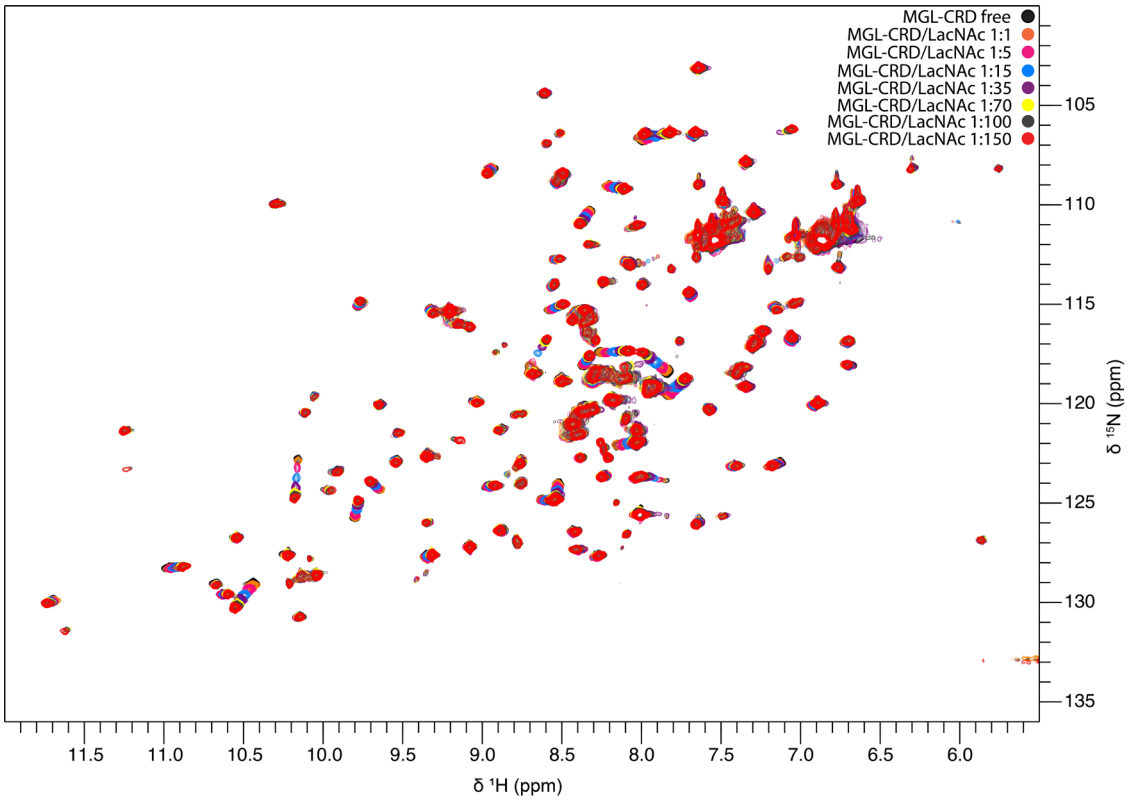

Figure A3 *- Overlay of the $^{1}H,^{15}N$-HSQC spectra of the titration of MGL-CRD (200 µM) with LacNAc at 293K.*

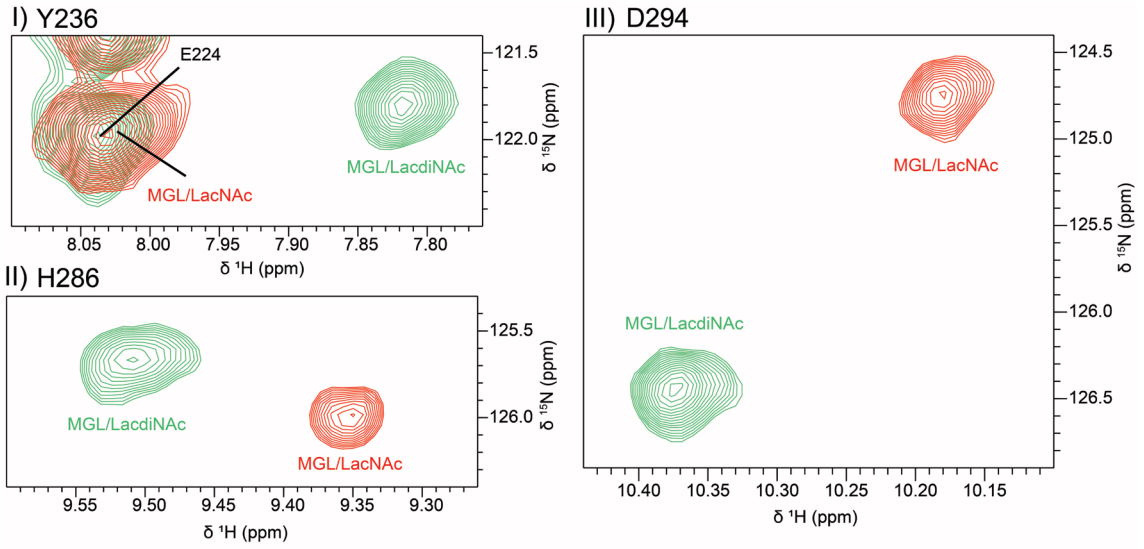

Figure A4 *- Overlay of regions of the $^{1}H,^{15}N$ HSQC spectra of MGL/LacdiNAc complex (green) and MGL/LacNAc complex (red), of signals of the amino acids with higher differences between the two bound states.*

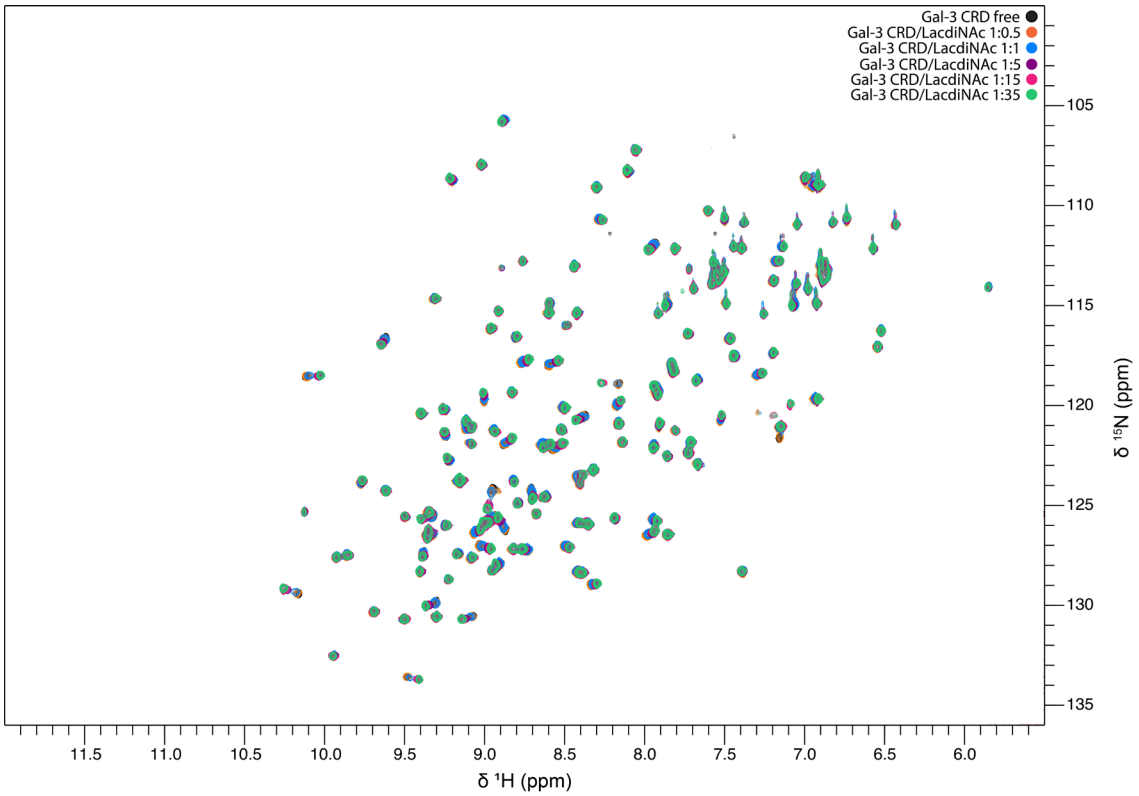

Figure A5 *- Overlay of the $^{1}H,^{15}N$-HSQC spectra of the titration of Gal-3 CRD (42 μM) with LacdiNAc at 298K.*

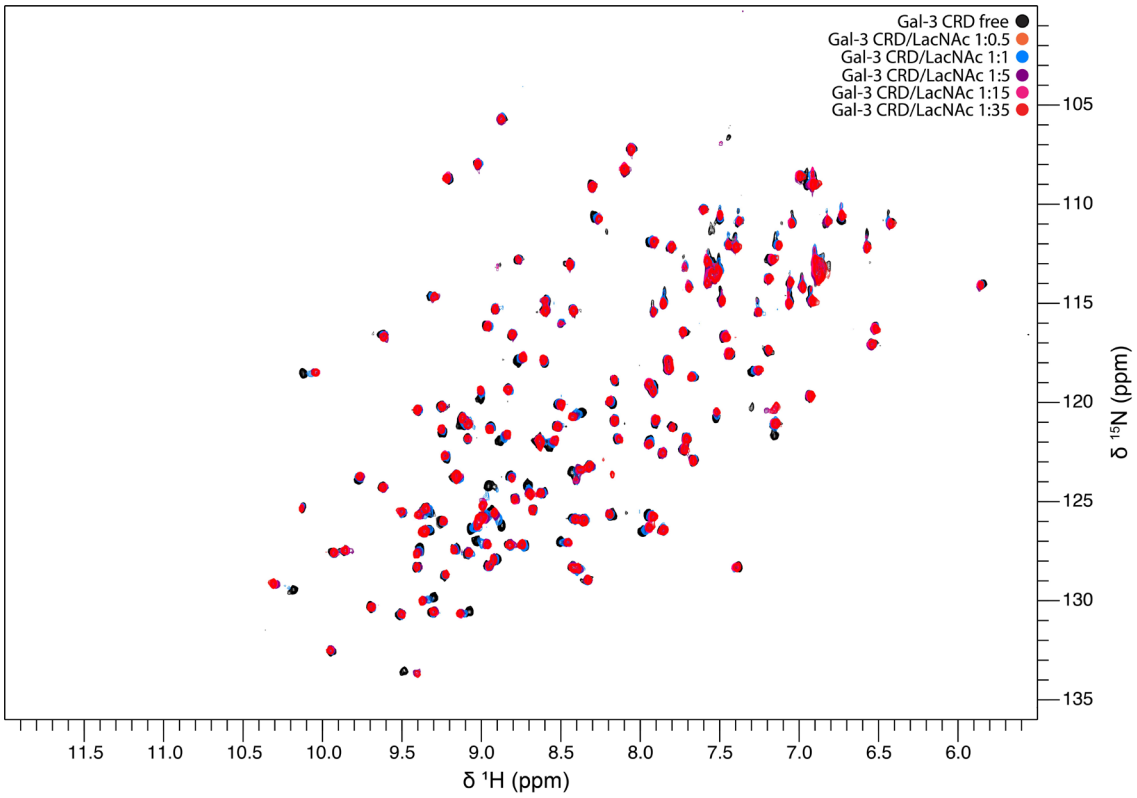

Figure A6 *- Overlay of the $^{1}H,^{15}N$-HSQC spectra of the titration of Gal-3 CRD (42 μM) with LacNAc at 298K.*

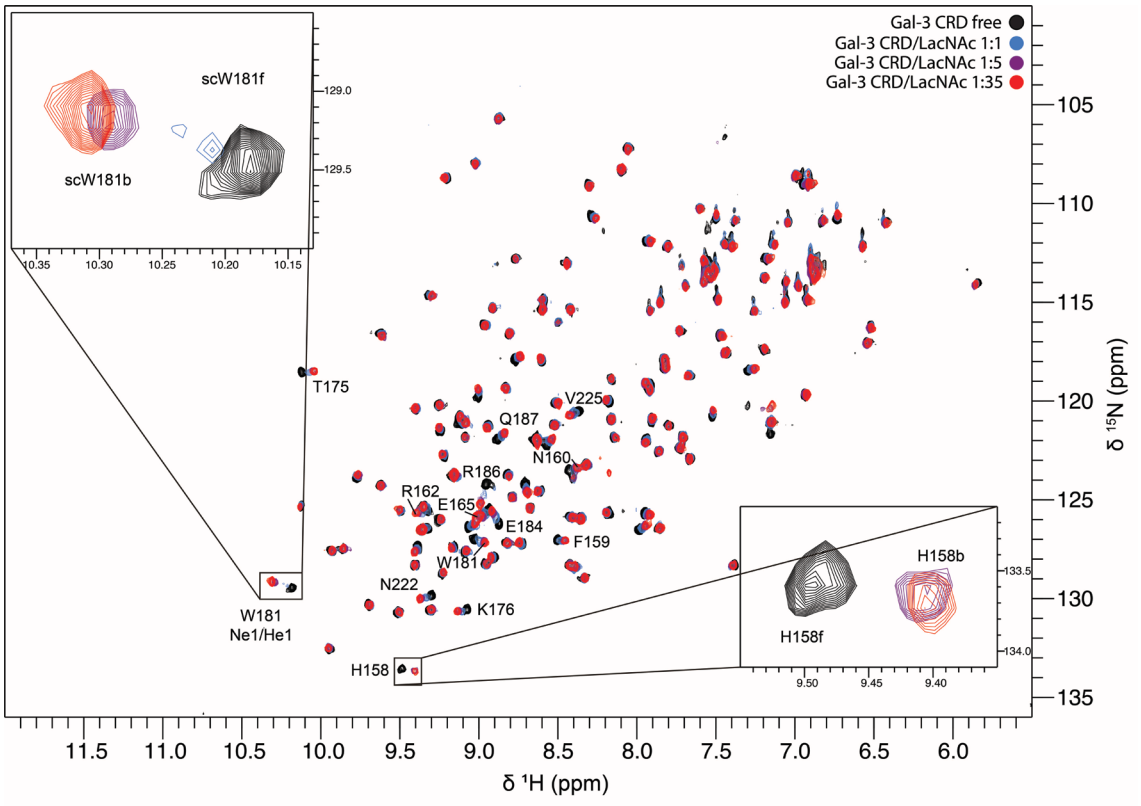

Figure A7 *- Overlay of the $^{1}H,^{15}N$-HSQC spectra of the titration of Gal-3 CRD with LacNAc: Gal-3 (42 µM) in apo state (black), Gal-3/LacNAc complex in presence of 1 equivalent of LacNAc (blue), Gal-3/LacNAc complex in presence of 5 equivalents of LacNAc (purple) and Gal-3/LacNAc complex in presence of 35 equivalents of LacNAc (red). The bottom right and top left of the spectra present the expansion region of the H158 and W181 side chain cross-peaks, respectively.*